Proceedings SOR

Rupnik V. and L. Bogataj (Editors): The 1st Symposium on Operational Research, SOR'93. Proceedings. Ljubljana: Slovenian Society Informatika, Section for Operational Research, 1993, 310 pp.

Rupnik V. and M. Bogataj (Editors): The 2nd International Symposium on Operational Research in Slovenia, SOR'94. Proceedings. Ljubljana: Slovenian Society Informatika, Section for Operational Research, 1994, 275 pp.

Rupnik V. and M. Bogataj (Editors): The 3rd International Symposium on Operational Research in Slovenia, SOR'95. Proceedings. Ljubljana: Slovenian Society Informatika, Section for Operational Research, 1995, 175 pp.

Rupnik V., L. Zadnik Stirn and S. Drobne (Editors.): The 4th International Symposium on Operational Research in Slovenia, SOR'97. Proceedings. Ljubljana: Slovenian Society Informatika, Section for Operational Research, 1997, 366 pp. ISBN 961-6165-05-4.

Rupnik V., L. Zadnik Stirn and S. Drobne (Editors.): The 5th International Symposium on Operational Research SOR '99, Proceedings. Ljubljana: Slovenian Society Informatika, Section for Operational Research, 1999, 300 pp. ISBN 961-6165-08-9.

Lenart L., L. Zadnik Stirn and S. Drobne (Editors.): The 6th International Symposium on Operational Research SOR '01, Proceedings. Ljubljana: Slovenian Society Informatika, Section for Operational Research, 2001, 403 pp. ISBN 961-6165-12-7.

Zadnik Stirn L., M. Bastič and S. Drobne (Editors): The 7th International Symposium on Operational Research SOR'03, Proceedings. Ljubljana: Slovenian Society Informatika, Section for Operational Research, 2003, 424 pp. ISBN 961-6165-15-1.

Zadnik Stirn L. and S. Drobne (Editors): The 8th International Symposium on Operational Research SOR'05, Proceedings. Ljubljana: Slovenian Society Informatika, Section for Operational Research, 2005, 426 pp. ISBN 961-6165-20-8.

Zadnik Stirn L. and S. Drobne (Editors): The 9th International Symposium on Operational Research SOR'07, Proceedings. Ljubljana: Slovenian Society Informatika, Section for Operational Research, 2007, 460 pp. ISBN 978-961-6165-25-9.

Zadnik Stirn L., J. Žerovnik, S. Drobne and A. Lisec (Editors): The 10th International Symposium on Operational Research SOR'09, Proceedings. Ljubljana: Slovenian Society Informatika, Section for Operational Research, 2009, 604 pp. ISBN 978-961-6165-30-3.

Zadnik Stirn L., J. Žerovnik, J. Povh, S. Drobne and A. Lisec (Editors): The 11th International Symposium on Operational Research SOR'11, Proceedings. Ljubljana: Slovenian Society Informatika, Section for Operational Research, 2011, 358 pp. ISBN 978-961-6165-35-8.

Proceedings SOR'13

Proceedings of the 12th International Symposium on OPERATIONAL RESEARCH

SOR '13

Dolenjske Toplice, Slovenia
September 25-27, 2013

Edited by:
L. Zadnik Stirn • J. Žerovnik • J. Povh • S. Drobne • A. Lisec

SOR '13 Proceedings

The 12th International Symposium on Operational Research in Slovenia

Dolenjske Toplice, SLOVENIA, September 25 - 27, 2013

Edited by:

L. Zadnik Stirn, J. Žerovnik, J. Povh, S. Drobne and A. Lisec

Slovenian Society INFORMATIKA (SDI)
Section for Operational Research (SOR)

Proceedings of the 12th International Symposium on Operational Research
SOR'13 in Slovenia, Dolenjske Toplice, September 25 - 27, 2013.

Organiser : Slovenian Society Informatika – Section for Operational Research, SI 1000 Ljubljana, Litostrojska cesta 54, Slovenia (www.drustvo-informatika.si/sekcije/sor/)

First published in Slovenia in 2013 by Slovenian Society Informatika – Section for Operational Research, SI 1000 Ljubljana, Litostrojska cesta 54, Slovenia (www.drustvo-informatika.si/sekcije/sor/)

CIP - Kataložni zapis o publikaciji
Narodna in univerzitetna knjižnica, Ljubljana

519.8(082)
519.8:005.745(082)
519.81:519.233.3/.5(082)

INTERNATIONAL Symposium on Operational Research in Slovenia (12 ; 2013 ; Dolenjske Toplice)
SOR '13 proceedings / The 12th International Symposium on Operational Research in Slovenia, Dolenjske Toplice, Slovenia, September 25-27, 2013 ; [organiser] Slovenian Society Informatika (SDI), Section for Operational Research (SOR). - Ljubljana : Slovenian Society Informatika, Section for Operational Research, 2013

ISBN 978-961-6165-40-2
1. Slovensko društvo Informatika. Sekcija za operacijske raziskave
268742912

Proceedings of the 12th International Symposium on Operational Research in Slovenia (SOR'13) is cited in: ISI (Index to Scientific & Technical Proceedings on CD-ROM and ISI/ISTP&B online database), Current Mathematical Publications, Mathematical Review, MathSci, Zentralblatt für Mathematic / Mathematics Abstracts, MATH on STN International, CompactMath, INSPEC, Journal of Economic Literature

Technical editor : Samo Drobne
Designed by : Samo Drobne
Printed by : Statistical Office of the Republik of Slovenia, Ljubljana, Slovenia

The 12th International Symposium on Operational Research in Slovenia - SOR '13
Dolenjske Toplice, SLOVENIA, September 25 - 27, 2013

Program Committee:

L. Zadnik Stirn, University of Ljubljana, Biotechnical Faculty, Ljubljana, Slovenia, Chair
J. Žerovnik, University of Ljubljana, Faculty of Mechanical Engineering, Ljubljana, Slovenia, Chair
E. D. Andersen, MOSEK ApS, Copenhagen, Denmark
Z. Babić, University of Split, Faculty of Economics, Department for Quantitative Methods, Split, Croatia
M. Bastič, University of Maribor, Faculty of Business and Economics, Maribor, Slovenia
M. Bogataj, University of Ljubljana, Faculty of Maritime Studies and Transport, Portorož, Slovenia
V. Čančer, University of Maribor, Faculty of Business and Economics, Maribor, Slovenia
S. Drobne, University of Ljubljana, Faculty of Civil Engineering and Geodesy, Ljubljana, Slovenia
L. Ferbar, University of Ljubljana, Faculty of Economics, Ljubljana, Slovenia
W. Gutjahr, University of Vienna, Department of Statistics and Decision Support Systems, Vienna, Austria
H. W. Hamacher, University of Kaiserslautern, Department of Mathematics, Kaiserslautern, Germany
A. M.C.A. Koster, Lehrstuhl II für Mathematik, RWTH Aachen, Germany
J. Kušar, University of Ljubljana, Faculty of Mechanical Engineering, Ljubljana, Slovenia
L. Lenart, Institute Jožef Stefan, Ljubljana, Slovenia
A. Lisec, University of Ljubljana, Faculty of Civil Engineering and Geodesy, Ljubljana, Slovenia
Z. Lukač, University of Zagreb, Faculty of Economics and Business, Zagreb, Croatia
L. Neralić, University of Zagreb, Faculty of Economics & Business, Zagreb, Croatia
I. Pesek, University of Maribor, Faculty of Natural Sciences and Mathematics, Maribor, Slovenia
U. Pferschy, University of Graz, Department of Statistics and Operations Research, Graz, Austria
J. Povh, Faculty of Information Studies, Novo mesto, Slovenia
K. Šorić, University of Zagreb, Faculty of Economics & Business, Zagreb, Croatia
D. Škulj, University of Ljubljana, Faculty of Social Sciences, Ljubljana, Slovenia
P. Šparl, University of Maribor, Faculty of Organizational Sciences, Kranj, Slovenia
B. Zmazek, University of Maribor, Faculty of Natural Sciences and Mathematics, Maribor, Slovenia

Organizing Committee:

J. Povh, Faculty of Information Studies, Novo mesto, Slovenia, Chair
S. Drobne, University of Ljubljana, Faculty of Civil Engineering and Geodesy, Ljubljana, Slovenia
A. Lisec, University of Ljubljana, Faculty of Civil Engineering and Geodesy, Ljubljana, Slovenia
B. Pavlakovič, Faculty of Information Studies, Novo mesto, Slovenia
J. Gabrič, Faculty of Information Studies, Novo mesto, Slovenia
L. Zadnik Stirn, University of Ljubljana, Biotechnical Faculty, Ljubljana, Slovenia
J. Žerovnik, University of Ljubljana, Faculty of Mechanical Engineering, Ljubljana, Slovenia

*The 12th International Symposium on Operational Research in Slovenia - **SOR '13***
Dolenjske Toplice, SLOVENIA, September 25 - 27, 2013

Chairs:

Z. Babić, University of Split, Faculty of Economics, Split, Croatia
M. Bogataj, MEDIFAS, Šempeter pri Novi Gorici, Slovenia
S. Cabello, University of Ljubljana, Faculty of Mathematics and Physics, Ljubljana, Slovenia
V. Čančer, University of Maribor, Faculty of Economics and Business Maribor, Maribor, Slovenia
L. Ferbar, University of Ljubljana, Faculty of Economics, Ljubljana, Slovenia
D. Jukić, University of Josip Juraj Strossmayer in Osijek, Department of Mathematics, Osijek, Croatia
S. Klavžar, University of Ljubljana, Faculty of Mathematics and Physics, Ljubljana, Slovenia
A. Lisec, University of Ljubljana, Faculty of Civil Engineering and Geodesy, Ljubljana, Slovenia
S. Pivac, University of Split, Faculty of Economics, Split, Croatia
J. Povh, Faculty of Information Studies, Novo mesto, Slovenia
M. Pejić-Bach, University of Zagreb, Faculty Economics and Business, Zagreb, Croatia
R. Sotirov, Tilburg University, Department of Econometrics and Operational Research, Tilburg, Netherlands
D. Škulj, University of Ljubljana, Faculty of Social Sciences, Ljubljana, Slovenia
T. Trzaskalik, University of Economics in Katowice, Faculty of Informatics and Communication, Katowice, Poland
L. Zadnik Stirn, University of Ljubljana, Biotechical Faculty, Ljubljana, Slovenia
M. Zekić-Susac, University of Josip Juraj Strossmayer in Osijek, Faculty of Economics, Osijek, Croatia
J. Žerovnik, University of Ljubljana, Faculty of Mechanical Engineering, Ljubljana, Slovenia

Preface

This volume, Proceedings of The 12th International Symposium on Operations Research, called SOR'13, contains papers presented at SOR'13 (http://sor13.fis.unm.si/) that was organized by Slovenian Society INFORMATIKA (SDI), Section for Operations Research (SOR) and Faculty of Information Studies (FIS), Novo mesto, Slovenia, held in Dolenjske Toplice, Slovenia, from September 25 to September 27, 2013. The volume contains blindly reviewed papers or abstracts of talks presented at the symposium. The opening address at SOR'13 was given by Prof. Dr. L. Zadnik Stirn, the President of the Slovenian Section of Operations Research, Mr. Niko Schlamberger, the President of Slovenian Society INFORMATIKA, Prof. Dr. Janez Povh, the Dean of Faculty of Information Studies, Novo mesto, and presidents/representatives of a number of Operations Research Societies from abroad.

SOR'13 is the scientific event in the area of operations research, another one in the traditional series of the biannual international OR conferences, organized in Slovenia by SDI-SOR. It is a continuity of eleven previous symposia. The main objective of SOR'13 is to advance knowledge, interest and education in OR in Slovenia, in Europe and worldwide in order to build the intellectual and social capital that are essential in maintaining the identity of OR, especially at a time when interdisciplinary collaboration is proclaimed as significantly important in resolving problems facing the current challenging times. Further, by joining IFORS and EURO, the SDI-SOR agreed to work together with diverse disciplines, i.e. to balance the depth of theoretical knowledge in OR and the understanding of theory, methods and problems in other areas within and beyond OR. We believe that SOR'13 creates the advantage of these objectives, contributes to the quality and reputation of OR by presenting and exchanging new developments, opinions, experiences in the OR theory and practice.

SOR'13 was highlighted by a distinguished set of four keynote speakers. The first part of the Proceedings SOR'13 comprises invited papers/abstracts, presented by four outstanding scientists: Professor Dr. Dragan Jukić, University of Josip Juraj Strossmayer in Osijek, Department of Mathematics, Osijek, Croatia, Professor Dr. Sandi Klavžar, University of Ljubljana, Faculty of Mathematics and Physics, Ljubljana, Slovenia, Assoc. Prof. Dr. Renata Sotirov, Tilburg University, Department of Econometrics and Operational Research, Tilburg, The Netherlands, Dr. Michel Petitjean, MTi, INSERM UMR-S, University Paris, Paris, France. The second part of the Proceedings includes 56 papers written by 102 authors. Most of the authors of the contributed papers came from Slovenia (47), then from Croatia (24), Poland (6), Austria (4), France (4), Serbia (4), Slovakia (4), Switzerland (3), Bosnia and Herzegovina (2), Italy (2), Spain (2), Hungary (1), The Netherlands (1), Singapore (1) and United Arab Emirates (1). The papers published in the Proceedings are divided into sections: Plenary Lectures (4 contributions), Mathematical Programming and Optimization (14), Graphs and their Applications (10), Multiple Criteria Decision Making (8), Econometric Models and Statistics (5), Production and Inventory (3), Finance and Investments (6), Location and Transport (6), Creative core FIS - Simulations (5).

The Proceedings of the previous eleven International Symposia on Operations Research organized by the Slovenian Section of Operations Research are indexed in the following secondary and tertiary publications: Current Mathematical Publications, Mathematical Review, Zentralblatt fuer Mathematik/Mathematics Abstracts, MATH on STN International and CompactMath, INSPEC. The Proceedings SOR'13 are expected to be covered by the same bibliographic databases.

The success of the scientific events at SOR'13 and the present proceedings should be seen as a result of joint effort. On behalf of the organizers we would like to express our sincere thanks to all who have supported us in preparing the event. We would not have succeeded in attracting so many distinguished speakers from all over the world without the engagement and the advice of active members of the Slovenian Section of Operations Research. Many thanks to them. Further, we would like to express our deepest gratitude to prominent keynote speakers, to the members of the Program and Organizing Committees, to the referees who raised the quality of the SOR'13 by their useful suggestions, section's chairs, and to all the numerous people - far too many to be listed here individually - who helped in carrying out The 12th International Symposium on Operations Research SOR'13 and in putting together these Proceedings. Last but not least, we appreciate the authors' efforts in preparing and presenting the papers, which made The 12th Symposium on Operations Research SOR'13 successful

We would like to express a special gratitude to donators of the 12th International Symposium on Operational Research in Slovenia (SOR'13): Krka, tovarna zdravil, d.d., Novo mesto, Slovenia, and Terme Krka, d.o.o., Novo mesto, PE Dolenjske Toplice, Dolenjske Toplice, Slovenia.

Dolenjske Toplice, September 25, 2013

Lidija Zadnik Stirn
Janez Žerovnik
Janez Povh
Samo Drobne
Anka Lisec
(Editors)

Contents

Author index

The 12th International Symposium on
Operational Research in Slovenia

SOR '13

Dolenjske Toplice, SLOVENIA
September 25 - 27, 2013

Plenary Lectures

ON THE l_p-NORM ESTIMATION IN A QUASILINEAR REGRESSION MODEL

Dragan Jukić
Department of Mathematics
J.J. Strossmayer University of Osijek
Trg Ljudevita Gaja 6, HR-31 000 Osijek, Croatia

This talk will be based on my recently submitted manuscript on the l_p-norm $(1 \le p < \infty)$ estimation of the parameters in a quasilinear regression model of the form

$$g(t;\boldsymbol{\alpha}) = \varphi(f_0(t) + \alpha_1 f_1(t) + \cdots + \alpha_n f_n(t)),$$

where $\boldsymbol{\alpha} = (\alpha_1,...,\alpha_n)^T \in \mathbb{R}^n$ is an unknown vector parameter, $f_0, f_1,..., f_n$ are arbitrary fixed functions, and the function $\varphi : I \to \mathbb{R}$, with $I \subseteq \mathbb{R}$ being an interval (open, closed, half-open, bounded or unbounded), is continuous and strictly monotonic.

Many important model functions which often appear in applied research are quasilinear or can be parameterized as a quasilinear model. For example: When $\varphi(u) = \exp(u)$, we have exponential regression; when $\varphi(u) = u^a$, where $a \neq 0$ is given, we have power regression; when $\varphi(u) = 1/u$, we have hyperbolic regression.

The focus of this talk will be on the existence of the best l_p-norm estimator in a quasilinear regression model of the above form. I will review what is known about this problem and then present a theorem which guarantees the existence of the best l_p-norm estimator. From that theorem, which both extends and generalizes the previously known existence result, the existence of the best l_p-norm estimator for the whole class of nonlinear model functions follows immediately.

Two Applicable Network Families: Fibonacci Cubes and Sierpiński Graphs

Sandi Klavžar
University of Ljubljana, FMF, Jadranska 19, Ljubljana, Slovenia
University of Maribor, FNM, Koroška cesta 160, Maribor, Slovenia
Institute of Mathematics, Physics and Mechanics, Jadranska 19, SI-1111 Ljubljana

Abstract

Fibonacci cubes and Sierpiński graphs form families of networks/graphs that have appealing structural properties and are applicable in different contexts. Here these families are presented, some of their structural properties recalled, and their applicability indicated.

1 Introduction

A good indication that a (mathematical) model is important is that it has applications in different areas of science. This is certainly the case with Fibonacci cubes and with Sierpiński graphs, two infinite families of graphs of our main interest here.

Fibonacci cubes were introduced in [14] as a model for interconnection networks because they can emulate many hypercube algorithms as well as they can emulate other topologies, as for instance meshes. Later it turned out that Fibonacci cubes are applicable in theoretical chemistry [23, 34]. That also lead to the Fibonacci dimension of a graph [1, 30]. Moreover, Fibonacci cubes paved the way to several families of graphs such as Lucas cubes [25], Fibonacci (p, r)-cubes [27], and generalized Fibonacci cubes [15].

The introduction of Sierpiński graphs [19] was motivated by investigations of certain universal topological spaces (see the book [24] for more on these spaces) and the fact that for base 3 they are isomorphic to the Tower of Hanoi graphs (see the book [11] for more on the Tower of Hanoi). Even earlier, in computer science, the so-called WK-recursive networks were introduced in [3], see also [7]. WK-recursive networks are very similar to Sierpiński graphs—they can be obtained from Sierpiński graphs by adding a link (an open edge) to each of its extreme vertices. Hence for all practical purposes, WK-recursive networks and Sierpiński graphs can be considered as the same family of graphs. In addition, Sierpiński graph were independently studied in [28] and are also known in computer science as iterated complete graphs, cf. [4].

In the following two sections we will, respectively, formally introduce Fibonacci cubes and Sierpiński graphs, present some of their basic properties, and point to some of the main areas of research related to these graphs.

2 Fibonacci cubes

Let $B = \{0, 1\}$ and for $n \geq 1$ set

$$\mathcal{B}_n = \{b_1 b_2 \dots b_n : \ b_i \in B, 1 \leq i \leq n\}.$$

The *n-dimensional hypercube* Q_n, or *n-cube* for short, is the graph defined on the vertex set $\mathcal{B}_n$, vertices $b_1 b_2 \dots b_n$ and $b'_1 b'_2 \dots b'_n$ being adjacent if $b_i \neq b'_i$ holds for exactly one $i \in \{1, \dots, n\}$. Hypercubes form one of the most fundamental models in the design of parallel computers and interconnection networks, cf. [32, Chapter 7]. They possess numerous properties that are essential for network efficiency, such as recursive decomposition, many symmetries, low regularity, small diameter, hamiltonicity, and straightforward local routing. Consequently, actual machines based on hypercubes were implemented, see [32, p. 115] for a list of their implementations.

Clearly, the order of Q_n is 2^n. Therefore, Hsu [14] proposed Fibonacci cubes as an infinite family of graphs with similar properties as hypercubes, but with their order growing much slower. For $n \geq 1$ let

$$\mathcal{F}_n = \{b_1 b_2 \dots b_n \in \mathcal{B}_n : \ b_i \cdot b_{i+1} = 0, 1 \leq i \leq n-1\}.$$

The set $\mathcal{F}_n$ thus contains all binary strings of length n that contain no two consecutive ones. Then the *Fibonacci cube* Γ_n, $n \geq 1$, has $\mathcal{F}_n$ as the vertex set, two vertices being adjacent if they differ in exactly one coordinate. Therefore, Γ_n is obtained from Q_n by removing all vertices that contain at least two consecutive ones. See Fig. 1 for Γ_5.

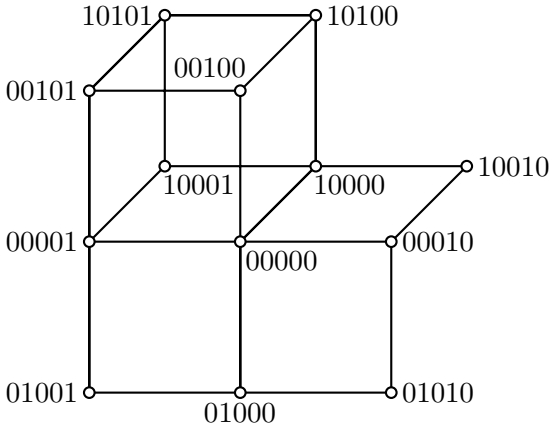


Figure 1: Fibonacci cube Γ_5

It is interesting to observe that for any $n \geq 1$, the Fibonacci cube Γ_n is isomorphic to $\kappa(\overline{P}_n)$, where $\overline{G}$ denotes the complement of a graph G and $\kappa(G)$ is the simplex graph of G. (The simplex graph has complete subgraphs of G as vertices, including the empty subgraph, where two vertices are adjacent if the two complete subgraphs differ in a single vertex.)

Fibonacci cubes can also be characterized in terms from chemical graph theory: it was proved in [23] that the resonance graph of an arbitrary fibonacene with n hexagons in exactly the Fibonacci cube Γ_n. (The resonance graphs are graphs that reflect the structure of perfect matching, while fibonacenes are hexagonal chains in which no three hexagons are linearly attached.) This characterization was extended in [34] by characterizing plane bipartite graphs whose resonance graphs are Fibonacci cubes.

For additional information on the structure of Fibonacci cubes see the survey [18]. To conclude this brief section we quickly present the variations and generalizations of Fibonacci cubes mentioned in the introduction.

Lucas cubes. These cubes are quite similar to Fibonacci cubes: The *Lucas cube* Λ_n ($n \geq 1$) is the subgraph of Q_n induced by the binary strings that do not contain two consecutive ones (so just as Fibonacci cubes) and, in addition, that do not contain 1 in both the first and the last coordinate. Lucas cubes thus form a symmetrization of Fibonacci cubes and have also found applications in theoretical chemistry, cf. [35].

Fibonacci (p, r)-cubes. The Fibonacci (p, r)-cube $\Gamma_n^{(p,r)}$ is the subgraph of Q_n induced on binary strings of length n in which there are at most r consecutive ones and at least p zeros between two substrings of ones. Note that $\Gamma_n = \Gamma_n^{(1,1)}$. Hence the Fibonacci (p, r)-cubes widely generalize Fibonacci cubes. At the same time they also generalize

some other interconnection networks from the literature, notably hypercubes and postal networks [31].

Generalized Fibonacci cubes. Let f be an arbitrary binary string and n a positive integer. Then the *generalized Fibonacci cube* $Q_n(f)$ is the graph obtained from Q_n by removing all the vertices that contain f as a factor. These cubes form a very wide generalization of Fibonacci cubes because the Fibonacci cube Γ_n is just the generalized Fibonacci graph $Q_n(11)$. Generalized Fibonacci cubes, among others, offer challenging problems in the area of combinatorics on words, see [16, 21].

3 Sierpiński graphs

The *Sierpiński graph* S_p^n, $p, n \geq 1$, is defined on the vertex set $\{1, \ldots, p\}^n$, two different vertices $u = (u_1, \ldots, u_n)$ and $v = (v_1, \ldots, v_n)$ being adjacent if and only if there exists an $h \in \{1, \ldots, n\}$ such that

(i) $u_t = v_t$, for $t = 1, \ldots, h-1$;
(ii) $u_h \neq v_h$; and
(iii) $u_t = v_h$ and $v_t = u_h$ for $t = h+1, \ldots, n$.

Abbreviations $\langle u_1 \ldots u_n \rangle$ and $u_1 \ldots u_n$ are used for the vertex $(u_1, \ldots, u_n)$ when appropriate. The Sierpiński graph S_3^4 together with the corresponding vertex labeling is shown on Fig. 2, while for S_5^3 see Fig. 3.

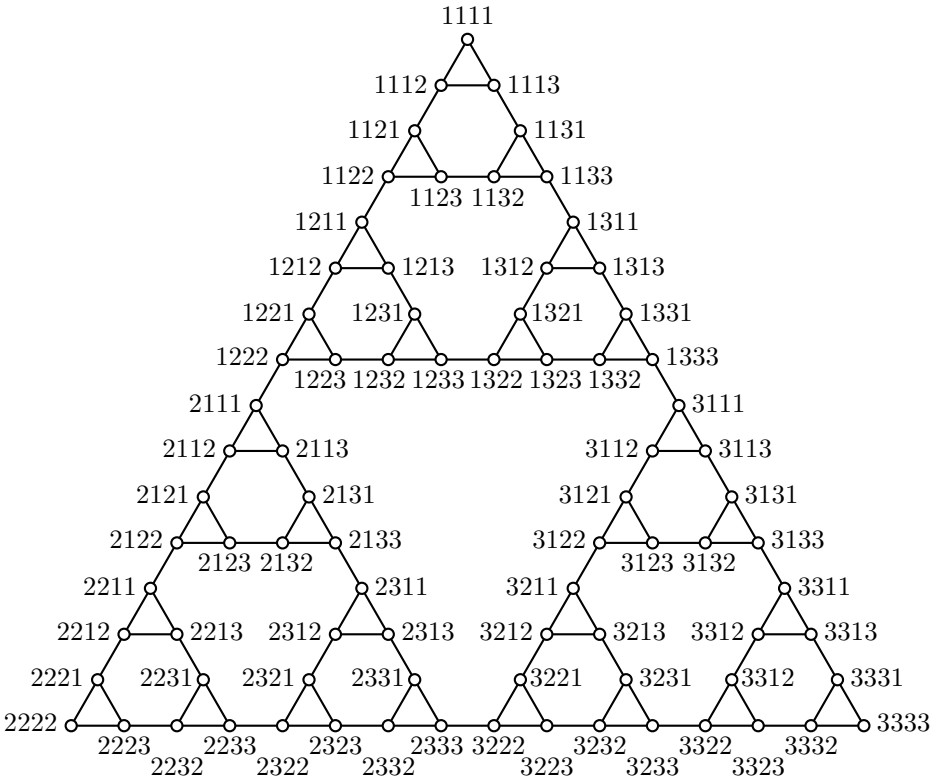


Figure 2: The Sierpiński graph S_3^4

A vertex of the form $\langle ii \ldots i \rangle$ of S_p^n is called an *extreme vertex*. S_p^n contains p^n vertices, out of which p are extreme. If $n \geq 2$, then for $i \in \{1, \ldots, p\}$ let iS_p^{n-1} be the subgraph of

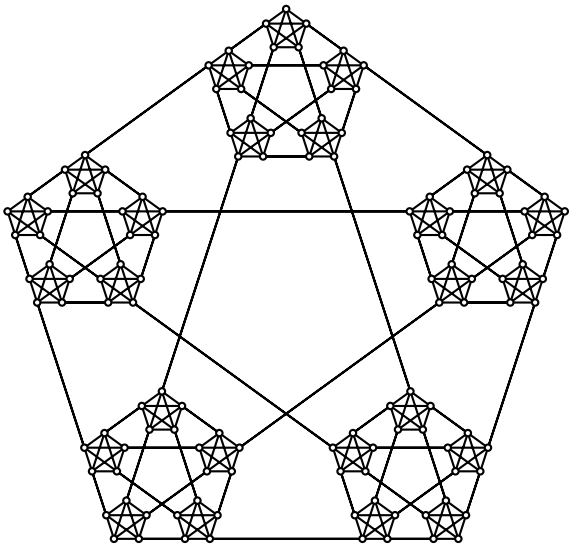


Figure 3: Sierpiński graph S_5^3

S_p^n induced by the vertices of the form $\langle iv_2 \dots v_n \rangle$. Then iS_p^{n-1} is isomorphic to S_p^{n-1}, cf. Figs. 2 and 3 again. This observation in particular implies that Sierpiński graphs have a fractal structure.

As already mentioned, the introduction of Sierpiński graph in [19] was in part motivated by the fact that S_3^n is isomorphic to the Tower of Hanoi graph with n discs. Hence, a shortest path in S_3^n between two vertices corresponds to an optimal solution to transfer discs between the corresponding regular states in the Tower of Hanoi puzzle. From this and from other reasons the metric structure of Sierpiński graphs have been extensively investigated. In the theory of the Tower of Hanoi it is known that there are at most two different shortest paths between any fixed pair of vertices. In [10] a formula is given that counts, for a given vertex v, the number of vertices u such that there are two shortest u, v-paths. The formula is expressed in terms of the celebrated Stern's diatomic sequence. Similarly, for a given almost-extreme vertex v, the set of vertices u is determined in [33] for which there exist two shortest u, v-paths. An *almost-extreme vertex* of S_p^n was introduced in [22] as a vertex that is either adjacent to an extreme vertex of S_p^n or is incident to an edge between two subgraphs of S_p^n isomorphic to S_p^{n-1}. For additional metric aspects of Sierpiński graphs see [12, 26].

The labeling of Sierpiński graphs is the key for structural studies and applications. As for the latter, we mention Romik's finite automaton for the so-called Tower of Hanoi ToH P2-problem [29]. From the structural studied, we point out that Sierpiński graphs appear to be interesting in coding theory, see [2, 8, 9, 20] and that different coloring problems were studied [5, 6, 13, 17].

References

[1] S. Cabello, D. Eppstein, S. Klavžar, The Fibonacci dimension of a graph, Electron. J. Combin. 18 (2011) Paper 55, 23 pp.

[2] P. Cull, I. Nelson, Error-correcting codes on the Towers of Hanoi graphs, Discrete Math. 208/209 (1999) 157–175.

[3] G. Della Vecchia, C. Sanges, A recursively scalable network VLSI implementation, Future Generation Comput. Syst. 4 (1988) 235–243.

[4] C. Frayer, S. Reddy, Perfect one error correcting codes and iterated complete graphs, Proceedings of the REU Program in Mathematics, NSF and Oregon State University, Corvallis, Oregon, 2002.

[5] H.-Y. Fu, $\{P_r\}$-free colorings of Sierpiński-like graphs, Ars Combin. 105 (2012) 513–524.

[6] H.-Y. Fu, D. Xie, Equitable $L(2,1)$-labelings of Sierpiński graphs, Australas. J. Combin 46 (2010) 147–156.

[7] J.-S. Fu, Hamiltonian connectivity of the WK-recursive network with faulty nodes, Inform. Sci. 178 (2008) 2573–2584.

[8] S. Gravier, S. Klavžar, M. Mollard, Codes and $L(2,1)$-labelings in Sierpiński graphs, Taiwanese J. Math. 9 (2005) 671–681.

[9] S. Gravier, M. Kovše, M. Mollard, J. Moncel, A. Parreau, New results on variants of covering codes in Sierpiński graphs, Des. Codes Cryptogr. 69 (2013) 181–188.

[10] A. M. Hinz, S. Klavžar, U. Milutinović, D. Parisse and C. Petr, Metric properties of the Tower of Hanoi graphs and Stern's diatomic sequence, European J. Combin. 26 (2005) 693–708.

[11] A. M. Hinz, S. Klavžar, U. Milutinović, C. Petr, *The Tower of Hanoi—Myths and Maths*, Springer, Basel, 2013.

[12] A. M. Hinz, D. Parisse, The average eccentricity of Sierpiński graphs, Graphs Combin. 28 (2012) 671–686.

[13] A. M. Hinz, D. Parisse, Coloring Hanoi and Sierpiński graphs, Discrete Math. 312 (2012) 1521–1535.

[14] W.-J. Hsu, Fibonacci cubes—a new interconnection technology, IEEE Trans. Parallel Distrib. Syst. 4 (1993) 3–12.

[15] A. Ilić, S. Klavžar, Y. Rho, Generalized Fibonacci cubes, Discrete Math. 312 (2012) 2–11.

[16] A. Ilić, S. Klavžar, Y. Rho, The index of a binary word, Theoret. Comput. Sci. 452 (2012) 100–106.

[17] M. Jakovac, S. Klavžar, Vertex-, edge-, and total-colorings of Sierpiński-like graphs, Discrete Math. 309 (2009) 1548–1556.

[18] S. Klavžar, Structure of Fibonacci cubes: a survey, J. Comb. Optim. 25 (2013) 505–522.

[19] S. Klavžar, U. Milutinović, Graphs $S(n,k)$ and a variant of the Tower of Hanoi problem, Czechoslovak Math. J. 47(122) (1997) 95–104.

[20] S. Klavžar, U. Milutinović, C. Petr, 1-perfect codes in Sierpiński graphs, Bull. Austral. Math. Soc. 66 (2002) 369–384.

[21] S. Klavžar, S. Shpectorov, Asymptotic number of isometric generalized Fibonacci cubes, European J. Combin. 33 (2012) 220–226.

[22] S. Klavžar, S. S. Zemljič, On distances in Sierpiński graphs: almost-extreme vertices and metric dimension, Appl. Anal. Discrete Math. 7 (2013) 72–82.

[23] S. Klavžar, P. Žigert, Fibonacci cubes are the resonance graphs of Fibonaccenes, Fibonacci Quart. 43 (2005) 269–276.

[24] S. Lipscomb, *Fractals and Universal Spaces in Dimension Theory*, Springer, Berlin, 2009.

[25] E. Munarini, C. Perelli Cippo, N. Zagaglia Salvi, On the Lucas cubes, Fibonacci Quart. 39 (2001) 12–21.

[26] D. Parisse, On some metric properties of the Sierpiński graphs S_k^n, Ars Combin. 90 (2009) 145–160.

[27] L. Ou, H. Zhang, Fibonacci (p, r)-cubes which are median graphs, Discrete Appl. Math. 161 (2013) 441–444.

[28] T. Pisanski, T. W. Tucker, Growth in repeated truncations of maps, Atti Sem. Mat. Fis. Univ. Modena 49 (2001) 167–176.

[29] D. Romik, Shortest paths in the Tower of Hanoi graph and finite automata, SIAM J. Discrete Math. 20 (2006) 610–622.

[30] A. Vesel, Fibonacci dimension of the resonance graphs of catacondensed benzenoid graphs, Discrete Appl. Math. 161 (2013) 2158–2168.

[31] J. Wu, The postal network: a recursive network for parameterized communication model, J. Supercomput. 19 (2001) 143–161.

[32] J.-M. Xu, Combinatorial Theory in Networks, Mathematics Monograph Series 26, Science Press, Beijing, 2013.

[33] B. Xue, L. Zuo, G. Wang, G. Li, Shortest paths in Sierpiňski graphs, Discrete Appl. Math., to appear.

[34] H. Zhang, L. Ou, H. Yao, Fibonacci-like cubes as Z-transformation graphs, Discrete Math. 309 (2009) 1284–1293.

[35] P. Žigert Pleteršek, M. Berlič, Resonance graphs of armchair nanotubes cyclic polypyrenes and amalgams of Lucas cubes, MATCH Commun. Math. Comput. Chem. 70 (2013) 533–543.

THE CHIRAL INDEX: APPLICATIONS TO MULTIVARIATE DISTRIBUTIONS AND TO 3D MOLECULAR GRAPHS

Michel Petitjean
MTi, INSERM UMR-S 973, University Paris 7
35 rue Hélène Brion, 75205 Paris Cedex 13, France.
petitjean.chiral@gmail.com
http://petitjeanmichel.free.fr/itoweb.petitjean.html

Abstract: We review the main properties of the chiral index. Its use as an asymmetry coefficient of multivariate probablity distributions is pointed out, and its application to measure the degree of chirality of rigid 3D molecular graphs is presented. Several extreme chirality sets are shown. Some open optimization problems are mentioned.

Keywords: chirality and symmetry measures, chiral index, asymmetry coefficient, colored mixture, colored Wasserstein distance, 3D molecular graphs.

1 INTRODUCTION

The historical definition of chirality is due to Lord Kelvin [5]: *I call any geometrical figure, or group of points, chiral, and say that it has chirality if its image in a plane mirror, ideally realized, cannot be brought to coincide with itself.* In other intuitive words, an object *identical* to one of its mirror images is achiral, i.e. not chiral: it has indirect symmetry. Despite what is believed since a long, the full mathematical definition of chirality does not relie on the existence of some oriented space. It is based on a general symmetry definition [17] and involves only basic group theory concepts [19]. In this paper we deal with a quantitative measure of the deviation from indirect symmetry. That problem goes back to the end of the 19^{th} century. It was of interest first for chemists and statisticians, but contributors from many fields are known (see [13] for a review). Although measuring the degree of asymmetry of the probability distribution of some random variable or vector is basically a geometric problem, the case of molecules is more complicated, even under assumption of a rigid model. To see this, we consider a simplified model of the molecule CHBrClF (bromochlorofluoromethane) with five ponctual atoms, four of them (H, F, Cl, Br) being the vertices of a regular tetrahedron with the fifth atom (C) at the center of the tetrahedron. Geometrically speaking we have an achiral object, but any chemist would say that this molecular object is chiral because a valid superposition of the molecule with any of its mirror image is expected to superpose an H atom with an H atom and so on with the four other atom types, and no valid superposition respecting these five constraints exists. The general situation for molecules is in fact more complicated because the labeling of the atoms does not depend only on their nature: it depends on the full molecular graph, where the punctual atoms are colored nodes, and the chemical bonds are colored edges. E.g., the graph of the water molecule H–O–H has three nodes and two edges. Such molecular graphs are of common use in chemistry [2, 6, 8]. The chiral index presented hereafter applies both to 3D molecular graphs and to multivariate distributions, discrete or continuous.

2 THE COLORED MIXTURE MODEL

A general process to define an indirect asymmetry coefficient of a multivariate distribution consists to consider a probability metric, and then to minimize the distance between the distribution and any of its indirect isometry image for all rotations and translations of that image.

The asymmetry coefficient is got via an adequate normalization of this minimized distance. Here, the L^2-Wasserstein distance D [3, 20] is considered: X_1 and X_2 being two random vectors in R^d, w being an element of the space W of their joint distributions and the quote denoting the transposition operator, then

$$D^2 = Inf_{\{w \in W\}} E[(X_1 - X_2)'(X_1 - X_2)] \tag{1}$$

In order to handle pairwise correspondences as required in chemistry, we first consider a probability space (C, A, P), where C ia a non empty set called the space of colors, A is a σ-algebra defined on C, and P is a probability measure. Then we define a mapping Φ from C on the space of probability distributions on (R^d, B), where B is the Borel σ-algebra of R^d. In other words, to each color $c \in C$ is associated a d-variate distribution $\tilde{P}_c = \Phi(c)$. The random variable (K, X) in the compound space $(C \times R^d, A \otimes B)$ is called a colored mixture [12, 14] because its distribution is viewed as a variant of the usual mixture distributions concept [4]. Then, considering a couple of random variables $(K_1, X_1), (K_2, X_2)$, the fundamental assumption of the colored mixture model is:

$$K_1 \overset{a.s.}{=} K_2 \tag{2}$$

It means that once a color is selected, we get two random vectors X_1 and X_2 which in general are not independant, and the set W_c of their joint distributions is a non empty subset of W introduced in eq. 1. The colored Wasserstein distance D_c is [12, 14]:

$$D_c^2 = Inf_{\{w \in W_c\}} E[(X_1 - X_2)'(X_1 - X_2)] \tag{3}$$

The case where C is of finite cardinality n is of interest. When $n = 1$, D_c and D coincide. For any n, when (a) the mixing distribution of K_1 (or K_2) is uniform, and (b) the mixed distributions are those of almost surely constant random vectors, D_c is the distance induced by the Frobenius norm, and this distance, minimized for some class of transformations of X_2 (e.g. linear, orthogonal, etc.), is the Procrustes distance [12]. This latter, with or without minimization for isometries of X_2, is called in the 3D case *RMS* or *RMSD* by many chemists and structural biologists. The colored mixture model is also a framework for defining shape complementarity and was used to define a geometric docking criterion when the expectation is replaced by a variance operator in the right member of eq. 3 [11, 14].

3 THE CHIRAL INDEX

The chiral index χ was introduced for finite sets in 1997 [7]. Then it was extended to weighted sets [10] before receiving its more general definition in 2002 for a colored mixture of finite inertia T [12], this inertia being referred to the marginal in R^d. The squared colored Wasserstein distance D_c^2 between a colored mixture and its image through any indirect isometry applied to its marginal in R^d (e.g. a mirror reflection), is minimized for all translations t and rotations R of the image, and then a normalization factor is applied so that $\chi \in [0; 1]$:

$$\chi = d \cdot [Inf_{\{R,t\}} D_c^2]/4T \tag{4}$$

The chiral index depends only on the distribution of the colored mixture and it is insensitive to isometries and scaling. It is null if and only if the distribution is indirect symmetric. The optimal translation is null for a centered distribution, and the optimal rotation is analytically known for $d = 2$ and $d = 3$ [12]. A direct symmetry index was defined for finite sets of points [9], but it cannot work for continuous distribution (see the discussion at the end of ref. [13]).

3.1 An asymmetry measure of multivariate distributions

When C is of cardinality 1, there is only one color and χ is an asymmetry coefficient of the distribution of the random vector associated to this unique color. In the unidimensional case, the chiral index of a distribution is expressable from the lower bound r_m of the correlation coefficient between two random variables following that distribution, taken over the space of their joint distributions:

$$\chi = (1 + r_m)/2 \tag{5}$$

Because r_m cannot be positive, in eq. 5 we have $\chi \in [0; 1/2]$. The chiral index should be compared with the skewness M_3, i.e. the reduced third order centered moment of the distribution. This latter is often presented as an asymmetry coefficient, and is such that $M_3^2 \leq M_4 - 1$, M_4 being the reduced fourth order centered moment [21, 23]. That inequality is itself a trivial consequence of equation A10 in [14] for a random vector G of null expectation:

$$Var(G'G) \geq E(GG'G) \cdot [E(GG')]^{-1} \cdot E(GG'G) \tag{6}$$

Unfortunately, the skewness can be null even for indirect symmetric distributions (see section 4.2 in [13]), although χ is null if and only if the distribution is achiral. Remark: an univariate *symmetric* distribution should be called achiral, because it has a mirror symmetry. An other advantage of χ over the skewness and its multivariate analogs is that χ is defined even when the third order moments do not exist.

From the convergence theorem section IV in [12], the sample chiral index is a consistent estimator of the chiral index of the parent distribution. Then, a class of open problems is to find simple asymptotic expressions of the distribution of the sample chiral index under hypothesis of interest for the experimentalist about the parent population, such as normality, uniformity, or else, in order to build symmetry tests.

In the case of a sample of n reals, r_m is got via correlating the ordered sample sorted in increasing order with the one sorted in decreasing order, and χ in eq. 5 is very easy to compute with a pocket calculator. Furthermore, χ offers simple expressions of the squared midranges or of the squared range lengths of the ordered sample (see section 2.9 in [13]).

Setting $d = 1$ and $n = 3$, and denoting by α the ratio of the lengths of the two adjacent segments defined by the three points, the chiral index is:

$$\chi = (1 - \alpha)^2/4(1 + \alpha + \alpha^2) \tag{7}$$

For this set, the chiral index satisfies to five properties:

1. χ is function of only the unique parameter of the set
2. χ is a continuous function of α
3. $\chi(1) = 0$
4. $\chi = 0 \Rightarrow \alpha = 1$
5. $\chi(\alpha) = \chi(1/\alpha)$ (invariance for scaling)

It has been emphasized in [16] that any safe chirality measure should first satisfy to the five properties above for this set, which is the simplest possible non trivial test set. By far it is not the case of many ones encoutered in the literature [13].

3.2 Colored sets and chemical graphs

The mechanism provided in section 2 permits to handle the constraints on pairwise correspondences (i.e. selecting permutations) between two sets of n points. When this constraint is relaxed, we are left to compute the Wasserstein distance between two uniform discrete distributions of n points, which needs to minimize the expectation in eq. 1 over the $n!$ pairwise correspondences. In the general case (e.g. continuous distributions), it is recalled that the constraints apply to a set of joint distributions. For molecules, the most used model is an undirected simple graph, where the nodes are colored by the Mendeleiev nature of the atoms and the edges are colored by the nature of the chemical bonds [6]. Molecular graphs are realized in R^3, and are assumed to be connected and rigid in the present framework.

In a molecular graph, a node x_2 is equivalent to a node x_1 when x_2 is the image of x_1 through a graph automorphism. The equivalence of all n nodes in a molecular graph does not mean that there are $n!$ automorphisms: e.g. consider a ring of 6 carbons with 6 single bonds such as in the cyclohexane squeleton, there are only 12 automorphisms, not 6!. For a general molecular graph, computing the chiral index needs to enumerate the permutations P associated to the graph automorphisms and to find the optimal rotation R for each permutation [7]. Let Y be the the array of n lines and d columns containing the coordinates of the n points, assumed to be centered, i.e. the mean of the n points is null. Q being an arbitrary negative determinant orthogonal matrix, the chiral index is:

$$\chi = \frac{d}{4Tr(Y'Y)} Min_{\{P,R\}}[Tr(Y - PYQ'R')'(Y - PYQ'R')] \tag{8}$$

For a molecular graph $d = 3$, and the optimal rotation R is known analytically [9].

4 SOME EXTREME CHIRALITY DISTRIBUTIONS

In eq. 4, a necessary condition to reach the upper bound $\chi = 1$ is to have the covariance matrix V proportional to the identity [12], i.e., σ being some positive real:

$$V = \sigma^2 I \tag{9}$$

Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$ be the eigenvalues of V and let us consider n equiprobable points with not two having the same color. The chiral index is [9]:

$$\chi = d\lambda_d / Tr(V) \tag{10}$$

In this situation, $\chi = 0$ iff the set is subdimensional and $\chi = 1$ iff eq. 9 is satisfied, which is the case for the regular simplex, the d-cube, etc. The most chiral triangles (i.e. sets of $n = 3$ points in the plane) have been computed [7]. When the 3 points have 3 different colors, it is equilateral. When 2 points have the same color and the last one has an other color, the squared side lengths ratios of the optimal triangle are $1 : 1 - \sqrt{6}/4 : 1 + \sqrt{6}/4$, and $\chi = 1 - \sqrt{2}/2$. When the 3 points have the same color, these ratios are $1 : 4 + \sqrt{15} : (5 + \sqrt{15})/2$ and $\chi = 1 - 2\sqrt{5}/5$. These three triangles are shown fig. 1. It can be checked from their cartesian coordinates given in [7] that they satisfy to the following property: each squared side length is proportional to three times a squared distance vertex-barycenter. That property appears also for the two triangles maximizing the direct symmetry index defined in [9]. It is symmetrical for all permutations of the 3 vertices only in the case of the equilateral triangle.

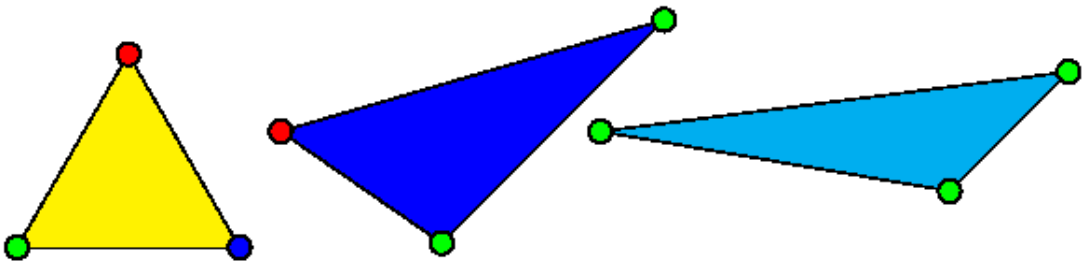


Figure 1: The maximal chirality triangles. From left to right, three different colors on vertices, two vertices with the same color, and three vertices with the same color.

We look now for the upper bound $\chi^*(d)$ of the chiral index in the case there is only one color, i.e. in the case of d-variate distributions (in fact, no need of color here). We get the following results for χ_1^*, χ_2^* and χ_d^* ($d \geq 1$) respectively from refs. [12], [1] and [18]:

$$\chi_1^* = 1/2 \tag{11}$$
$$\chi_2^* \in [1-1/\pi; 1-1/2\pi] \tag{12}$$
$$\chi_d^* \in [1/2; 1] \tag{13}$$

The Bernoulli distribution with parameter tending to 0 or to 1 has a chiral index tending to χ_1^* [12]. For $d \geq 2$, finding χ_d^* is an open problem. As mentioned in sect. 3.1, the sample chiral index is a consistent estimator of the parent population chiral index, so that χ_d^* can be seeked among samples of increasing size n. The case $d = 2$ is of interest. Defining $z \in C^n$, $z = x + \mathbf{i}y$, where x and y are the vectors in R^n of the marginals of the bidimensional sample, and P being the permutation matrix associated to their joint distribution matrix P/n, it is known that ithe optimal P is symmetric and the chiral index takes a simple expression [1]:

$$\chi = 1 - [Max_{\{P\}}|z'Pz|]/\|z\|^2 \tag{14}$$

Let Y be the matrix $[x|y]$, and μ_1 and μ_2 be the eigenvalues of $Y'PY$ ($\mu_1 \geq \mu_2$). Eq. 14 can be rewritten:

$$\chi = 1 - [Max_{\{P\}}(\mu_1 - \mu_2)]/Tr(Y'Y) \tag{15}$$

It was conjectured in [1] that $\chi_2^* = 1 - 1/\pi$ and a family of distributions in which the chiral index can be arbitrarily close to $1 - 1/\pi$ was exhibited.

The 3D molecular graph of an hydrocarbon designed by A. Schwartz [22] has, among several remarkable properties, a chiral index of 0.9824 and its carbon skeleton has $\chi = 1.0000$.

An attempt to define the closest achiral distribution to a given chiral one was done[15], but no satisfactory general approach to that problem is known.

References

[1] Coppersmith, D., Petitjean, M., 2005. *About the Optimal Density Associated to the Chiral Index of a Sample from a Bivariate Distribution*. Compt. Rend. Acad. Sci. Paris, Série I, **340**[8],599–604.

[2] Diudea, M.V., Petitjean, M., 2008. *Symmetry in Multi-Tori.*. Symmetry Cult. Sci. **19**[4], 285–305.

[3] Dobrushin, R.L., 1970. *Prescribing a system of random variables by conditional distributions*. Theor. Probab. Appl. **15**[3], 458–486.

[4] Everitt, B.S., Hand, D.J., 1981. *Finite Mixture Distributions*. Chap. 1, Chapman and Hall, London.

[5] Lord Kelvin, 1904. *Baltimore Lectures on Molecular Dynamics and the Wave Theory of Light*, Appendix H., chap. 22, footnote p. 619. C.J. Clay and Sons, Cambridge University Press Warehouse, London.

[6] Petitjean, M., 1992. *Applications of the Radius-Diameter Diagram to the Classification of Topological and Geometrical Shapes of Chemical Compounds*. J. Chem. Inf. Comput. Sci. **32**[4],331–337.

[7] Petitjean, M., 1997. *About Second Kind Continuous Chirality Measures. 1. Planar Sets*. J. Math. Chem. **22**[2-4],185–201.

[8] Petitjean, M., 1999. *Calcul de chiralité quantitative par la méthode des moindres carrés*. Compt. Rend. Acad. Sci. Paris, Série IIc, **2**[1],25–28.

[9] Petitjean, M., 1999. *On the Root Mean Square Quantitative Chirality and Quantitative Symmetry Measures*. J. Math. Phys. **40**[9],4587–4595.

[10] Petitjean, M., 2001. *Chiralité quantitative: le modèle des moindres carrés pondérés*. Compt. Rend. Acad. Sci. Paris, Série IIc, **4**[5],331–333.

[11] Petitjean, M., 2002. *Solving the Geometric Docking Problem for Planar and Spatial Sets*. Internet Electron. J. Mol. Des. **1**[4],185–192.

[12] Petitjean, M., 2002. *Chiral mixtures*. J. Math. Phys. **43**[8],4147–4157.

[13] Petitjean, M., 2003. *Chirality and Symmetry Measures: A Transdisciplinary Review*. Entropy **5**[3],271–312.

[14] Petitjean, M., 2004. *From Shape Similarity to Shape Complementarity: toward a Docking Theory*. J. Math. Chem. **35**[3],147–158.

[15] Petitjean, M., 2006. *À propos de la référence achirale*. Compt. Rend. Chim. **9**[10],1249–1251.

[16] Petitjean, M., 2006. *Minimal Symmetry, Random and Disorder*. Symmetry Cult. Sci. **17**[1-2], 197–205.

[17] Petitjean, M., 2007. *A Definition of Symmetry*. Symmetry Cult. Sci. **18**[2-3], 99–119.

[18] Petitjean, M., 2008. *About the Upper Bound of the Chiral Index of Multivariate Distributions*. AIP Conf. Proc. **1073**, 61–66.

[19] Petitjean, M., 2010. *Chirality in Metric Spaces*. Symmetry Cult. Sci. **21**[1-3], 27–36.

[20] Rachev, S.T., 1991. *Probability Metrics and the Stability of Stochastic Models*. Chap. 6, Wiley, New York.

[21] Rohatgi, V.K., Székely, G.J., 1989. Sharp Inequalities between Skewness and Kurtosis. *Stat. Prob. Lett.*, **8**, 297–299.

[22] Schwartz, A., Petitjean, M., 2008. *[6.6]Chiralane: A Remarkably Symmetric Chiral Molecule*. Symmetry Cult. Sci. **19**[4], 307–316.

[23] Wilkins, J.E. (1944). A Note on Skewness and Kurtosis. *Ann. Math. Stat.* **15**, 333–335.

WHY SEMIDEFINITE PROGRAMMING?

Associate Professor **Renata Sotirov**
Department of Econometrics & Operations Research
Tilburg University
Warandelaan 2
P.O. Box 90153
5000 LE Tilburg
The Netherlands
Telephone: +31 13 466 3178
Fax: +31 13 466 3280
Email: r.sotirov@uvt.nl
Homepage: https://stuwww.uvt.nl/~sotirovr

Semidefinite programming (SDP) is an extension of linear programming where the nonnegative vector variables are replaced by positive semidefinite matrix variables. SDP offers excellent possibilities for the design of very tight relaxations for several combinatorial optimization problems, and has diverse applications in eigenvalue optimization, control theory, robust optimization, engineering, etc. The roots of SDP trace back to the sixties of the previous century, but the interest has grown tremendously during the last twenty years. Nowadays, semidefinite programming is one of the most exciting areas in mathematical programming.

In this talk we provide motivation, background, and some latest developments in SDP. We also present relaxations and corresponding bounds for the maximum cut, the traveling salesman problem, the bandwidth problem in graphs, and the graph partition problem.

The 12th International Symposium on
Operational Research in Slovenia

SOR '13

Dolenjske Toplice, SLOVENIA
September 25 - 27, 2013

Section I:

Mathematical Programming and Optimization

ALGORITHM FOR STOCHASTIC GENERALIZED PRODUCTION-TRANSPORTATION PROBLEM WITH CONCAVE COSTS[1]

Marcin Anholcer
Poznań University of Economics, Faculty of Informatics and Electronic Economy
Al. Niepodległości 10, 61-875 Poznań, Poland
m.anholcer@ue.poznan.pl

Abstract: In the paper we present the Stochastic version of the Generalized Production-Transportation Problem with concave production costs. The solution algorithm, based on the branch and bound method, is presented. The subproblems are solved with a modified version of the Equalization Method.

Keywords: Generalized Transportation Problem, Stochastic Transportation Problem, Production-Transportation Problem, global optimization, concave costs, branch and bound, Equalization Method.

1 INTRODUCTION

Generalized Transportation Problem is a special case of the Generalized Flow Problem. The characteristic element of this kind of problems are the changes in amount of the delivered goods that occure during the transportation process.

Description and possible applications of the Generalized Flow Problem may be found e.g. in [1], some issues were also discussed in [6] and [14]. A polynomial method for the Generalized Flow Problem was presented in [32], while polynomial algorithms for the Generalized Circulation Problem may be found in [15]. Some interesting applications of generalized flows (like supplies of medical materials, food, pharmaceuticals and clothes) were analysed in [20].

The Generalized Transportation Problem was analysed e.g. in [4], [5] and [18]. In [21] a transportation problem with additional constraints of GTP type was considered. A special class of GTP was analysed also in [28]. In [3] the authors considered the application of GTP for modelling the distribution process where the complaints are involved. The influence of the complaints ratio on the structure of optimal network was analysed.

In the Generalized Production-Transportation Problem we assume that the commodity is delivered from factories to warehouses and additional production cost is included in the objective value. In this paper we are interested in the cases where the production costs are separable, concave functions (i.e. the production costs in chosen factory depend only on the production level in this factory and grow slower when the production grows). The problems of this kind with deterministic demand were considered e.g. in [17], [19], [25], [26], [30], [31] and [34] (linear concave transportation problems), [10], [11] and [27] (linear plant location problems), [7], [9], [13], [35] (other concave network problems). The solution algorithms applied in most cases were some kind of branch and bound method. Interesting exceptions are works [30] and [31], where authors provided an algorithm that is polynomial in the number of destinations (what is crucial, as the number of destinations is usually much bigger than the number of sources). Unfortunately, none of the mentioned treated the generalized version of the problem.

In the stochastic version of the problem, we assume that the demand is not deterministic, but we know the distribution of demand of every destination point. If the

[1] The research was part of the project „Nonlinear optimization in chosen economical applications". The project was financed by the National Science Center grant awarded on the basis of a decision number DEC-2011/01/D/HS4/03543.

delivery exceeds the demand, then additional surplus cost is imposed. If the delivery is too low, then the additional shortage cost is involved. Our goal is to minimize the sum of all the deterministic costs (in our case, the transportation and production costs) and the expected value of the additional costs imposed at the destination points. The Stochastic Transportation Problem (no production costs) was analysed e.g. in [8], [22], [24], [29] and [33]. The Stochastic Generalized Transportation Problem was considered in [2] and [23]. The only paper known to the author where stochastic demand and concave costs were considered simultaneously, is [16]. There is no such a work about the generalized version of the problem, which will be considered in the remainder of this paper. We will use some ideas from [16], but also from [12], where the algorithm for general nonconvex problem was provided.

In next section the problem is defined. In section 3, the solution method was described. Section 4 contains brief conclusions from the research.

2 PROBLEM FORMULATION

In the ordinary Generalized Transportation Problem, uniform good is transported from m supply points to n destination points. During the transportation process, the amount delivered to the demand point j from supply point i is equal to $r_{ij}x_{ij}$, where x_{ij} is the amount of good that leaves the supply point i and r_{ij} is the respective *reduction ratio*, corresponding with the change of the good. The unit transportation costs c_{ij} are constant, the demand b_j of every demand point j has to be satisfied and the supply a_i of each supply point i cannot be exceeded. Thus, the model has the following form:

$$\min f(x) = \sum_{i=1}^{m} c_{ij} x_{ij},$$

s.t.

$$\sum_{i=1}^{m} r_{ij} x_{ij} = b_j,\ j = 1,...,n, \qquad (1)$$

$$\sum_{j=1}^{n} x_{ij} \le a_i, i = 1,...,m,$$

$$x_{ij} \ge 0, i = 1,...,m,\ j = 1,...,n.$$

In the case of production – transportation problem, additional (production) costs are imposed on the total amount of good that leaves the factory. In this paper we assume that the production cost at every factory i is a concave function g_i.

In the stochastic version the demands b_j are not deterministic, but are given as the continuous random variables X_j with density functions φ_j. The unit surplus cost $s_j^{(1)}$ and the unit shortage cost $s_j^{(2)}$ are defined for every destination point j. The function of expected extra cost for destination j takes the form

$$f_j(x_j) = s_j^{(1)} \int_0^{x_j} (x_j - t)\varphi_j(t)dt + s_j^{(2)} \int_{x_j}^{\infty} (t - x_j)\varphi_j(t)dt. \qquad (2)$$

After some basic transformations (see e.g. [8]), the latter one takes the form

$$f_j(x_j) = s_j^{(2)}\left(E(X_j) - x_j\right) + (s_j^{(1)} + s_j^{(2)}) \int_0^{x_j} \Phi_j(t)dt, \qquad (3)$$

where Φ_j is the cumulative distribution function of the demand at destination j.

Finally, the SGPTP has the following form:

$$\min f(x,y) = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} + \sum_{j=1}^{n} f_j(x_j) + \sum_{i=1}^{m} g_i(y_i),$$

s.t.

$$\sum_{i=1}^{m} r_{ij}x_{ij} = x_j,\ j = 1,...,n, \qquad (4)$$

$$\sum_{j=1}^{n} x_{ij} = y_i \le a_i, i = 1,...,m,$$

$$x_{ij} \ge 0, i = 1,...,m,\ j = 1,...,n.$$

It is straightforward to see that the first two derivatives of the expected cost functions are

$$f_j'(x_j) = -s_j^{(2)} + (s_j^{(1)} + s_j^{(2)})\Phi_j(x_j) \qquad (5)$$

and

$$f_j''(x_j) = (s_j^{(1)} + s_j^{(2)})\varphi_j(x_j), \qquad (6)$$

so each function f_j is twice differentiable and convex. Recall also that we assume concavity of the functions g_i. This means that the problem SGPTP (4) is a hard global optimization problem. In next section we will present a solution method, based on the branch and bound approach.

3 SOLUTION METHOD

Observe that if we exchange the functions g_i to their linear lower estimators $\hat{g}_i$, then we will obtain a convex problem equivalent to the Stochastic Generalized Transportation Problem with the objective function $\hat{f}(x,y) \le f(x,y)$. This allows us to use the ideas presented in [12] and [16].

Let us consider an *m*-dimensional rectangle $R(l,u) = \bigotimes_{i=1}^{m}[l_i, u_i]$, where $0 \le l_i < u_i \le a_i$ for *i* = 1, …, *m*. Every rectangle can be divided into smaller ones by subdivision of any of the intervals $[l_i, u_i]$.

Let us define the linear functions $\hat{g}_i$ in such a way, that for every *i*,

$$\hat{g}_i(l_i) = g_i(l_i) \quad \text{and} \quad \hat{g}_i(u_i) = g_i(u_i). \qquad (7)$$

Obviously, exactly one such function $\hat{g}_i$ exists for every *i*. Now observe, that from the concavity of g_i it follows that for every *i* and every $l_i \le y_i \le u_i$, $\hat{g}_i(y_i) \le g_i(y_i)$. Let us consider the following problem (we will denote it SGTP(*l*, *u*)).

$$\min \hat{f}(x,y) = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} + \sum_{j=1}^{n} f_j(x_j) + \sum_{i=1}^{m} \hat{g}_i(y_i),$$

s.t.

$$\sum_{i=1}^{m} r_{ij}x_{ij} = x_j,\ j = 1,...,n, \qquad (8)$$

$$\sum_{j=1}^{n} x_{ij} = y_i, l_i \le y_i \le u_i, i = 1,...,m,$$

$$x_{ij} \ge 0, i = 1,...,m,\ j = 1,...,n.$$

This is a problem similar to SGTP, with additional constraints imposed on variables y_i. It means that SGTP(l, u) (8) can be solved with Equalization Method (see [2]) modified in such a way, that the total amount of good transported from any source must fulfil the additional conditions $l_i \le y_i \le u_i$. Observe that for every feasible solution (x, y) we have $\hat{f}(x, y) \le f(x, y)$, so also the optimal solution (x^*, y^*) of (8) satisfies this condition. Moreover, $\hat{f}(x^*, y^*) = f(x^*, y^*)$ if and only if (x^*, y^*) is also the optimum of (4) with additional constraints $l_i \le y_i \le u_i$. This means that the value of $\hat{f}(x, y)$ is a good lower bound on the value of $f(x, y)$.

Now we must establish also the branching method. The idea is as follows. After choosing the rectangle with lowest value of $\hat{f}(x^*, y^*)$, we check the differences $g_i(y_i^*) - \hat{g}_i(y_i^*)$, choose k highest ones and divide the rectangle into 2^k smaller rectangles by exchanging, for chosen values of i, the constraints $l_i \le y_i \le u_i$ to the constraints $l_i \le y_i \le y_i^*$ and $y_i^* \le y_i \le u_i$. There are many possible choices of k. The most popular (see e.g. [12], [16], [17]) is to choose always $k = 1$, i.e., the rectangle is always subdivided into two new rectangles and the search tree is a binary tree. Another way is to choose all the indices k, for which $g_i(y_i^*) - \hat{g}_i(y_i^*) > 0$ (or rather $g_i(y_i^*) - \hat{g}_i(y_i^*) > \varepsilon$ for some accuracy level ε). This method of choice leads us to much flatter search tree with possibly many vertices of large degrees, what may cause some difficulties with handling with many subsets at the same moment ($k = m$ in many cases). Finally, one can choose some compromise – e.g. choose in every step all the indices for which $g_i(y_i^*) - \hat{g}_i(y_i^*) > 0$, but not more than some fraction of m, not more than some constant, or only such, for which the difference exceeds some fraction (say half) of the biggest one.

All the above considerations lead us to the following algorithm for the concave SGPTP (4).

Algorithm 1

1. (Initialization) Let $l_i = 0$ and $u_i = a_i$ for every i. Let the only active subset of solutions be the one corresponding with the rectangle $R(l, u)$. Solve the SGTP(l, u), remember the optimal value of the objective function $\hat{f}(x^*, y^*)$. Set $f^* = f(x^*, y^*)$.
2. (Choosing the promising subset) Choose the active subset corresponding with the rectangle $R(l, u)$ with smallest value of $\hat{f}(x^*, y^*)$. If $f^* - \hat{f}(x^*, y^*) < \varepsilon$ for predefined accuracy level ε, then STOP. Obtained solution is optimal. Otherwise go to step 3.
3. (Dividing the promising subset) Choose k indices i with highest values of $g_i(y_i^*) - \hat{g}_i(y_i^*)$ and subdivide $R(l, u)$ by setting for all chosen i either $l_i = y_i^*$ or $u_i = y_i^*$. For each created rectangle $R(l, u)$ find the optimal solution (x^*, y^*) of SGTP(l, u) and if $f^* > f(x^*, y^*)$, then set $f^* = f(x^*, y^*)$.
4. (Closing subsets) Close all the subsets, for which $\hat{f}(x^*, y^*) > f^*$ and go back to step 2.

4 CONCLUSIONS

As far as the author knows, in this paper the concave Stochastic Generalized Production-Transportation Problem was described and analysed for the first time. The provided solution

algorithm, although based on the branch and bound method, is effective, as it can be seen in the table 1. Solution times in milliseconds are presented for randomly generated problems of size $m \times n$ with uniform or exponential distribution of demand and power functions of production costs.

Table 1. Solution times in milliseconds: average, standard deviation, minimum, maximum.

Problem	U(10×200)	U(20×200)	U(50×200)	Exp(10×200)	Exp(20×200)	Exp(50×200)
AVG	58.7	229.8	274.5	135.4	438.4	1066.9
STD	30.4	128.9	107.8	82.6	366.5	415.5
MIN	15.0	62.0	62.0	31.0	142.0	422.0
MAX	172.0	968.0	624.0	609.0	2496.0	2621.0

References

[1] Ahuja R.K., Magnanti T.L., Orlin J.B., 1993. Network Flows. Theory, Algorithms and Applications, Prentice Hall, 846 pp.

[2] Anholcer M., 2013. Algorithm for Stochastic Generalized Transportation Problem, Operations Research and Decisions, accepted.

[3] Anholcer M., Kawa A., 2012. Optimization of Supply Chain via Reduction of Complaints Ratio, Lecture Notes in Computer Science, Volume 7327/2012, pp.622-628.

[4] Balas E.,1966. The Dual Method for the Generalized Transportation Problem, Management Science, Vol. 12, No. 7, Series A, Sciences, pp. 555-568.

[5] Balas E., Ivanescu P.L., 1964. On the Generalized Transportation Problem, Management Science, Vol. 11, No. 1, Series A, Sciences, pp. 188-202.

[6] Bazaraa M.S., Jarvis J.J., Sherali H.D., 2010. Linear Programming and Network Flows, Fourth Edition, John Wiley & Sons Inc., 748 pp.

[7] Cao B., 1992. Transportation Problem with Nonlinear Side Constraints – a Branch and Bound Approach, ZOR – Methods and Models of Operations Research 36, pp. 185-197.

[8] Cooper L., LeBlanc L., 1977. Stochastic Transportation Problems and other Network Related Convex Problems, Naval Research Logistics Quarterly 24, pp. 327-337.

[9] Daeninck G., Smeers Y. 1977. Using Shortest Paths in Some Transshipment Problems with Concave Costs, Mathematical Programming 12, pp. 18-25.

[10] Davis P.S., Ray T.L., 1969. A Branch-Bound Algorithm for the Capacitated Facilities Location Problem, Naval research Logistics Quarterly, 16, pp. 331-344.

[11] Efroymson M. A., Ray T. L., 1966. A Branch-Bound Algorithm for Plant Location, Operations Research 14, pp. 361-368.

[12] Falk J.E., Soland R.M., 1969. An Algorithm for Separable Nonconvex Programming Problems, Management Science Vol. 15, No. 9, pp. 550-569.

[13] Erickson R.E., Monma C.L., Veinott A.F., Jr., 1987. Send-and-Split Method for Minimum - Concave-Cost Network Flows, Mathematics of Operations Research 12 (4), pp. 634-664.

[14] Glover F., Klingman D., Napier A., 1972. Basic Dual Feasible Solutions for a Class of Generalized Networks, Operations Research, Vol. 20, No. 1, pp. 126-136.

[15] Goldberg A.V., Plotkin S.A., Tardos E., 1988. Combinatorial Algorithms for the Generalized Circulation Problem, SFCS'88 Proceedings of the 29th Annual Symposium on Foundations of Computer Science, pp. 432-443.

[16] Holmberg K., Tuy H., 1999. A Production-Transportation Problem with Stochastic Demand and Concave Production Costs, Mathematical Programming 85, pp. 157-179.

[17] Kuno T., Utsunomiya T., 2000. A Lagrangian Based Branch-and-Bound Algorithm for Production-Transportation Problems, Journal of Global Optimization 18, pp. 59-73.

[18] Lourie J.R., 1964. Topology and Computation of the Generalized Transportation Problem, Management Science, Vol. 11, No. 1, Series A, Sciences, pp. 177-187.

[19] Nagai H., Kuno T., 2005. A Simplicial Branch-and-Bound Algorithm for Production-Transportation Problems with Inseparable Concave Production Costs, Journal of the Operations Research Society of Japan, Vol. 48, No. 2, pp. 97-110.

[20] Nagurney A., Yu M., Masoumi A. H., Nagurney L. S., 2013. Networks Against Time. Supply Chain Analytics for Perishable Products, Springer Briefs in Optimization, Springer.

[21] Pandian P., Anuradha D., 2011. Floating Point Method for Solving Transportation Problems with Additional Constraints, International Mathematical Forum, Vol. 6, mo. 40, pp. 1983-1992.

[22] Qi L., 1985. Forest Iteration Method for Stochastic Transportation Problem, Mathematical Programming Study 25, pp. 142-163.

[23] Qi L., 1987. The A-Forest Iteration Method for the Stochastic Generalized Transportation Problem, Mathematics of Operations Research Vol. 12, No. 1, pp. 1-21.

[24] Romeijn H. E., Sargut F. Z., 2012. The Stochastic Transportation Problem with Single Sourcing, European Journal of Operational Research 214(2), pp. 262-272.

[25] Sikora W., 1991. Branch and Bound Method for a Transportation-Production Problem with Concave Quadratic Cost Function (Metoda podziału i ograniczeń dla problemu transportowo-produkcyjnego z wklęsłą, kwadratową funkcją kosztów), Statistical Review (Przegląd Statystyczny) R. XXXVIII (3-4), pp. 335-348.

[26] Sikora W.,1993. Models and Methods of Optimal Distribution of Goods (Modele i metody optymalnej dystrybucji dóbr), Zeszyty naukowe - seria II, Prace doktorskie i habilitacyjne, Akademia Ekonomiczna w Poznaniu, Poznań (in Polish).

[27] Soland R.M., 1974. Optimal Facility Location with Concave Costs, Operations Research 22 (2), pp. 373-382.

[28] Srinivasan S., Thompson G.L., 1973. An Algorithm for Assigning Uses to Sources in a Special Class of Transportation Problems, Operations Research 21, pp. 284-295.

[29] Szwarc W., 1964. The Transportation Problem with Stochastic Demand, Management Science, Vol. 11, No. 1, pp. 33-50.

[30] Tuy H., Ghannadan S., Migdalas A., Värbrand P., 1993. Strongly Polynomial Algorithm for a Production-Transportation Problem with Concave Production Costs, Optimization 27, pp. 205-227.

[31] Tuy H., Ghannadan S., Migdalas A., Värbrand P., 1996. A Strongly Polynomial Algorithm for a Concave Production - Transportation Problem with a Fixed Number of Nonlinear Variables, Mathematical Programming 72, pp. 229-258.

[32] Wayne K.D., 2002. A Polynomial Combinatorial Algorithm for Generalized Minimum Cost Flow, Mathematics of Operations Research, Vol. 27, No. 3, pp. 445-459.

[33] Williams A.C., 1963. A Stochastic Transportation Problem, Operations Research 11, pp. 759-770.

[34] Youssef M.A., Mahmoud M. M., 1994. An Iterative Procedure for Solving the Uncapacitated Production-Distribution Problem under Concave Cost Function, International Journal of Operations & Production Management, Vol. 16, No. 3, pp. 18-27.

[35] Zangwill W.I., 1968. Minimum Concave Cost Flows in Certain Networks, Management Science 14 (7), pp. 429-450.

SIMULATION MODELING FOR PROCESS PERFORMANCE MANAGEMENT IN HIGHER EDUCATION: A CASE STUDY OF COLLABORATION IMPROVEMENT

Vesna Bosilj Vukšić and **Mirjana Pejić Bach**
Faculty of Economics & Business – Zagreb, University of Zagreb, Department of Informatics
Trg J.F. Kennedyja 6, HR-10000 Zagreb, Croatia
{vbosilj, mpejic}@efzg.hr

Katarina Tomičić-Pupek
Faculty of organization and informatics, University of Zagreb
Pavlinska 2, HR-42000 Varaždin, Croatia
ktomicic@foi.hr

Abstract: Process performance management (PPM) has become one of the most important management tools in profit organizations. However, non-profit organizations also started to benefit from PPM with the goal of efficiency improvement. Goal of the paper is to investigate usefulness of embedding simulation modeling approach for process performance management on the case study of collaboration improvement in higher education. Case study methodology has been used, and simulation modeling for PPM at University of Zagreb, Croatia with the purpose of collaboration improvement has been presented.

Keywords: process performance management, higher education, simulation, collaboration

1 INTRODUCTION

During a decade process performance measurement (PPM) has been very popular research topic, but until recently the focus has been on the profit organizations. Managing and measuring performance in public sector organizations is a growing phenomenon worldwide [22,3,10]. For two decades, worldwide Higher Education Institutions (HEIs) have been under increasing pressure to become more efficient for the services they provide [21,6,1]. According to Lam et al. [15] the degree of performance excellence that an organization can achieve greatly depends on the efficiency of business processes. Therefore these authors suggest quantitative methodologies to be used for supporting the business process improvement. Business process improvement efforts involve changes in people, processes and technology over time. As these changes happen over time, simulation appears to be a suitable process modeling method.

Goal of the paper is to investigate the usage of simulation modelling (SM) as a tool for PPM. For that purpose, simulation modelling was applied for process improvement in collaboration procedure example at the University of Zagreb, Croatia.

2 CASE STUDY OF COLLABORATION IMPROVEMENT

2.1 Process performance management

According to Neely et al. [17] a PPMS is a balanced and dynamic system that enables support of decision-making process by gathering, elaborating and analyzing information. It uses different measures and perspectives in order to give a holistic view of the organization. Kueng [14] defines a PPMS as an information system which: (1) gathers performance relevant data through a set of indicators; (2) compares the current values against historical or planned values, and (3) disseminates the results to the process actors and managers. Many firms have developed a wide variety of performance indicators which they review

periodically while some have very complex and sophisticated PPMSs that allow them to track what is happening in real time.

2.2 Characteristics of performance measurement and management in HEIs

A literature review on PPM implementation in the public sector highlights the factors driving performance in HEIs. A study conducted by Educause [9] showed that HEIs have invested heavily in business process change and redesign projects. These projects were driven mainly by budget shortages, information technology implementation and external requirements for improved efficiency and effectiveness [7,12]. Since expenditure on administration of HEIs is typically about 30% of that allocated to academic activities, [6] set up a data envelopment analysis (DEA) framework to identify good management practices leading to efficient administrative services in UK universities. This study demonstrated the problems in defining the unit of assessment and the relationship between inputs and outputs. To promote HEIs operating performance, performance measurement indicators (PMIs) are needed. Chen et al. [8] analyzed the literature and employed the established PMIs to identify the important key performance indicators (KPIs). As a result of this study, 78 PMIs were developed and were categorized in 18 measurement dimensions. The authors recommend that universities use these indicators to measure its operating performance.

2.3 Simulation modeling in PPM

Business processes simulation creates an added value in understanding, analyzing, and designing processes by introducing dynamic aspects [5]. It enables migration from a static towards a dynamic process model [2]. Nowadays most business process modeling tools include simulation capabilities, but in addition, there are some tools that are especially designed for more demanding simulation projects [16].

Many authors examined and described the development and implementation of simulation models in order to analyze the existing business processes and to predict the performance of new designs. Numerous advantages are listed: by using simulation it is possible to predict the effects of changes and the duration of the processes and bottlenecks and to thereby avoid bad decisions [19]; a what-if analysis can be conducted in order to assess various scenarios performance [4]; by running the simulation through time it is possible to gauge how changes at an operational level can lead to the meeting of strategic goals over time [11]. However, limitations to use simulation modelling in public sector are also discussed, such as: problem definition issues, socio-political issues and multi-perspective issues [20]; a lack of clear vision and support from top management [18]; the resistance to change [13].

3 CASE STUDY OF COLLABORATION IMPROVEMENT

3.1 Methodology

Process models in the following example are modeled as sequential iteration models, where the activities are repeated one after the other depending of inputs to the activities, outputs of activities, probabilistic rules, and available resources. Process models are modeled in accordance with BPMN 2.0 and they may consist of flow objects like activities and gateways, data objects, connecting objects, swim lanes and artifacts. Each activity is described by several parameters: name, duration, resources need to conduct activity, availability of the

resources assigned to activities, time required for resource to conduct activity, inputs of predecessors. Our case study is described in detail in the following sections.

3.2 1st Step:Inititation

An illustration of embedding simulation modeling approach for reengineering collaboration in higher education is given for the process of nomination and selection of themes for final thesis for undergraduate and graduate students. The process is modeled with IBM Websphere Business Modeler version 7.0.0.4 and in accordance with BPMN 2.0. The resources availability (working schedule) is also defined, but is not shown here. The process model shown in figure 1 is used to perform simulation of the process.

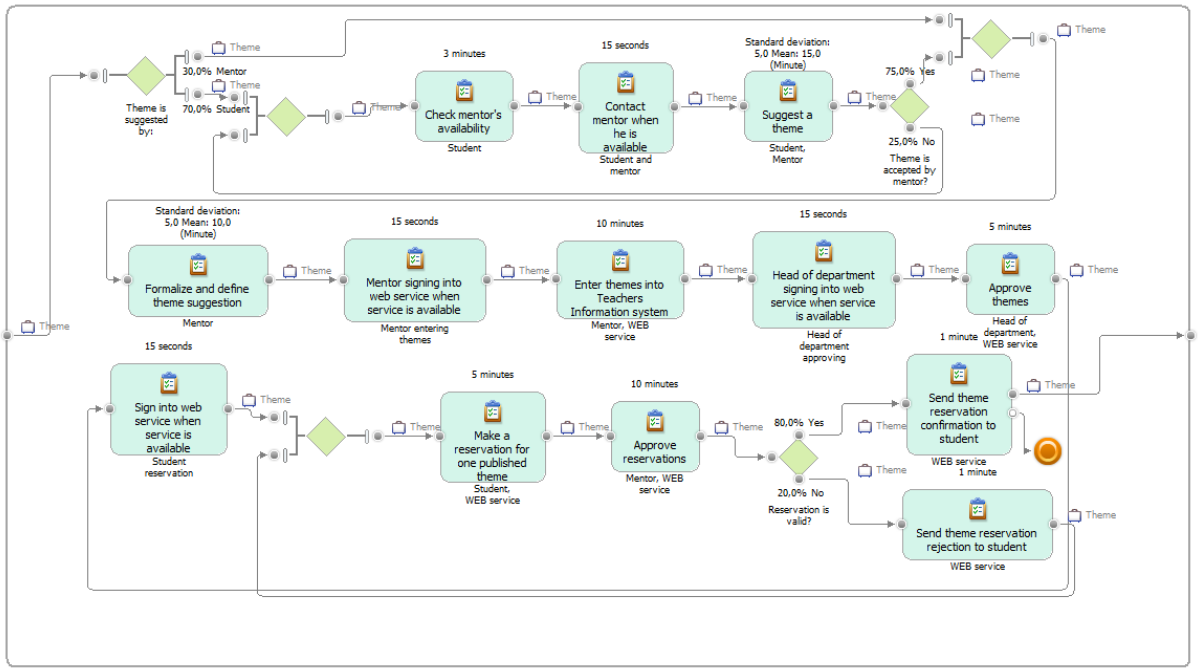


Figure 1: ***As-Is*** Process model of nomination and selection of themes for final thesis for undergraduate and graduate students

3.3 2nd Step: Analysis

Parameters that were used in the simulation of the process have been set according to real values gathered by a brief assessment based on student questioning and authors real experience. For simulation purpose in this case study the following simulation parameters were used: (i) Number of simulation instances: the simulation was ran over 40 instances representing 40 student inquiries for themes; (ii) Frequency of instances: 50% of instances appear every 0,5 days, 20% of instances start up every 1 day, in 15% of all cases instance appearance is every 0,75 days, and in 5% of cases the instances are triggered every 2 days.

The simulation of the As Is process showed the average duration of 40 days 7 hours 14 minutes with duration standard deviation of 8 days 7 hours 36 minutes 20 seconds 33 milliseconds. When the simulation results are compared with real data, an acceptable statistical error is under 5%, i.e. due to the lack of comparable real data the statistical significance is assessed based on real experience.

The analysis showed that the duration of the process is a feasible improvement opportunity. Other functional requirements of end-user (i.e. teachers or lecturers) imply the need of following the early stages of theme suggestion and formalizing, as well as the need of

tracking all communication of student and mentor. These changes are implemented in the To Be process model discussed in section 5.3.

3.4 3rd Step: Re-engineering

Based on the analysis of the As Is process that showed that the duration of the process is a possible improvement opportunity and which identified other functional requirements of end-user new reengineered process model (i.e. the To Be process) was developed (figure 2).

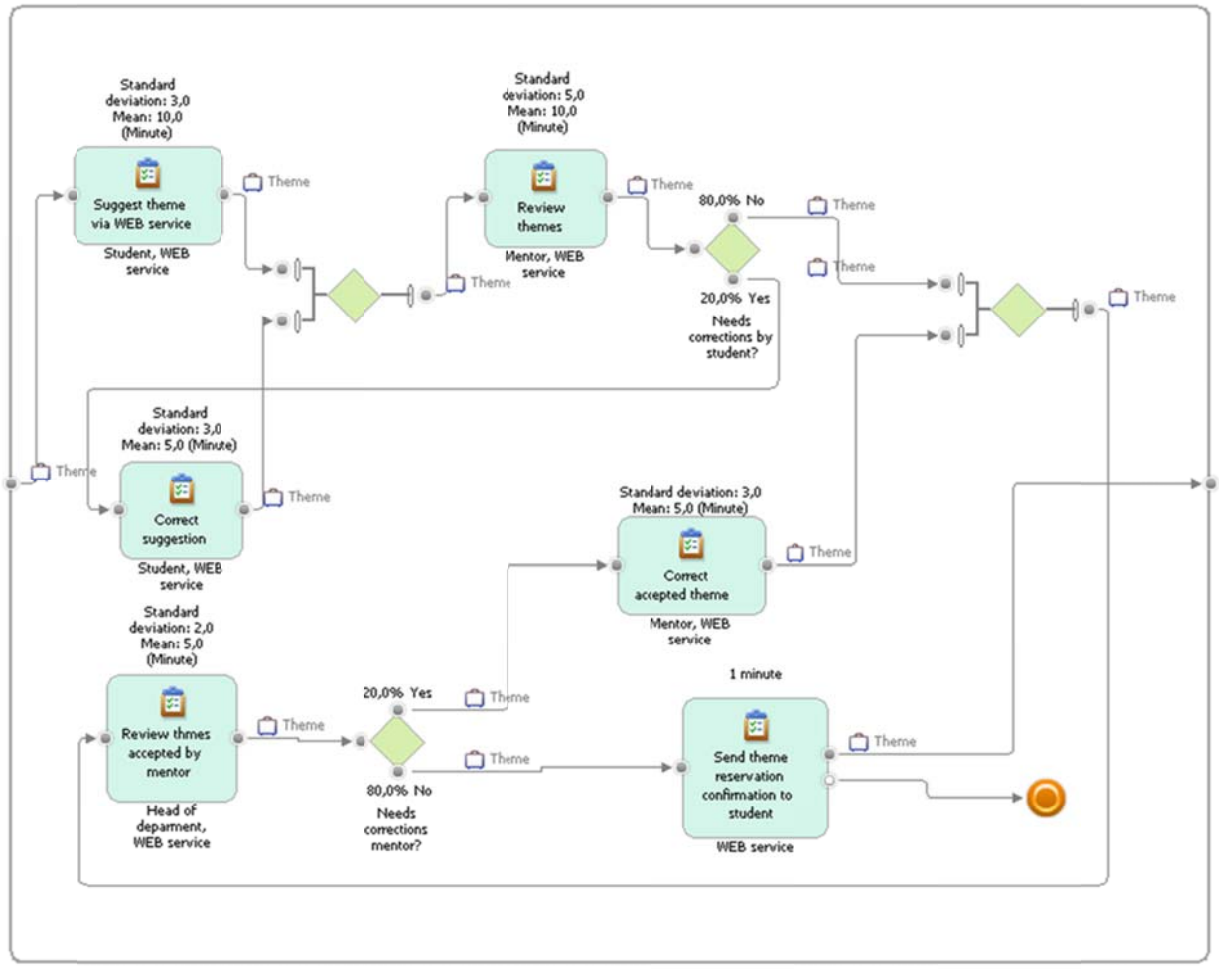


Figure 2: ***To-Be*** Process model of nomination and selection of themes for final thesis for undergraduate and graduate students

3.5 4th Step: Implementation

Like in real implementations a grace period for total transition must be well planned and the „new way of doing business“ should be introduced progressively. Simulation results listed in table 1 show the graduate improvement of the discussed process. The graduate implementation consists of progressive transition from the As Is process handling the themes as shown in figure 1, then with introduction of the new process variant in 20% of all instances (i.e. 20% of 40 instances), then in the next stage with 40% of all instances (i.e. 40% of 40 instance), and last with a total transition into the To Be process variant. Simulation parameters for the simulation of the To Be process are the same as in the simulation of the As Is process.

Table 1: Simulation results for progressive introduction of the new process model

	Simulation start time	*Current end time*	*Instanc. creat.*	*Instanc. complet.*	*Average duration*	*Duration st.dev.*
Process As Is	02.11.2012. 00:00:00	25.12.2012. 15:15:00	40	40	40 d 7 h 14 min	8 d 7 h 36 min
Instance To Be 20%	03.11.2012. 00:00:00	25.12.2012. 13:39:00	40	40	28 d 10 h 4 min	16 d 3 h
Instance To Be 40%	01.11.2012. 00:00:00	25.12.2012. 12:03:00	40	40	24 d 15 h 37 min	22 d 12 h 35 min
Instance To Be 100%	03.11.2012. 00:00:46	26.11.2012. 09:22:23	40	40	22 h 33 min	20 h 38 min

3.6 5th Step: Evaluation

Simulation results for progressive introduction of the new process model shown in table 1 need to be evaluated when a decision over accepting the To Be process is made. The key performance indicators relevant for this case study may be process/activity duration and duration standard deviation. As earlier stated duration standard deviation can be significant to analyze, assess or predict the stability of the process. Two conclusions regarding duration and its standard deviation are (1) if the ratio of duration and its standard deviation is small, predictions for resources allocation are more precise and (2) if the ratio is tending towards 1 then the organization is close to guessing when allocating resources to activities and possible threat of resources waste is greater. However, data in table 1 show that two conclusions do not necessary apply in general.

4 CONCLUSIONS

Overall conclusion of this case study should illustrate the significance of pondering KPI's and a need of conducting a detailed analysis of relevant KPIs corroborated by objective data. Implementation of a new scenario of the discussed process of nomination and selection of themes for final thesis for undergraduate and graduate students has a relatively short duration and thereby reduces overall average duration of the process. At the same time the relatively short duration of this process case influences the duration standard deviation in such a way that it enhances the duration standard deviation. This is natural because there are more process scenarios with durations which fall within the greater range. In our discussed case the reduction of average duration has a greater significance over enhancement of duration standard deviation as this was stated during the analysis stage as an important feasible improvement opportunity in combination with other functional requirements of end-users which lead to changes in the To Be process model. According to our results, simulation modelling has been proved as valuable method in PPM initiatives in HEIs.

References

[1] Abdous, M., (2011). Towards a framework for business process reengineering in higher education, Journal of Higher Education Policy and Management, Vol. 33, No.4, pp. 427-433.

[2] Aguilar, M., Rautert, T., Alexander, P., (1999). Business process simulation: A fundamental step supporting process centered management. Proceedings of 1999 Winter Simulation Conference, Phoenix, Arizona, pp. 1383-1392.

[3] Arnaboldi, M., Azzone, G., (2010). Constructing performance measurement in the public sector, Critial Perspectives on Accounting, 21(2010), pp. 266-282.

[4] Bertolini, M., Bevilacqua, M., Ciarapica F.E., Giacchetta, G., (2011). Business process reengineering in healthcare management: a case study, Business Process Management Journal, Vol.17, No.1, pp. 42-66.

[5] Bosilj Vukšić, V., Štemberger Indihar, M., Jaklič, J., Kovačič, A., (2002). Assessment of E-Business Transformation Using Simulation Modeling, Simulation, Vol. 78, No. 12, pp. 731-744.

[6] Casu, B., Thanassoulis, E., (2006). Evaluating cost efficiency in central administrative services in UK universities, Omega – The International Journal of Management Science, 34(2006), pp. 417-426.

[7] Chae, B., Poole, M., (2005). Enterprise system development in higher education. Journal of Cases of information technology, Vol.7, No.2, pp.82-101.

[8] Chen, S.H., Wang, H.H., Yang, K.J. (2009). Establishment and application of performance measure indicators for universities, The TQM Magazine, Vol.21, No. 3, pp. 220-235.

[9] Educause (2005). Good enough! IT investment and business process performance in higher education. Available at: http://www.educause.edu/library/resources/good-enough-it-investment-and-business-process-performance-higher-education, downloaded May 29th 2013.

[10] Goh, S.C., (2012). Making performance measurement systems more effective in public sector organizations, Measuring Business Excellence, Vol.16, No.1, pp.31-42.

[11] Greasley, A., (2006). Using process mapping and business process simulation to support a process-based approach to change in a public sector organization, Technovation, 26(2006), pp. 95-103.

[12] Harmon, P., (2007). Business Process Change: a Guide for Business Managers and BPM and Six Sigma Professionals. Morgan Kaufmann Publishers, Burlington, USA

[13] Jarvis, P., (2001). Universities and corporate universities, Kogan Page Limited.

[14] Kueng, P., (2000). "Process performance measurement system: a tool to support process-based organizations", Total Quality Management, Vol. 11, No. 1, pp. 67-85.

[15] Lam, C.Y., Ip, W.H., Lau, C.W., (2009). A business process activity model and performance measurement using a time series ARIMA intervention analysis, Expert Systems with Applications, 36 (2009), pp. 6986-6994.

[16] Mahal, A., (2010). Business Process Management, Basics and Beyond: How Work Gets Done, Technics Publications, New Yersey.

[17] Neely, A., Adams, C. and Kennerley, M., (2002). The Performance Prism: The Scorecard for Measuring and Managing Stakeholder Relationship, London: Prentice Hall.

[18] Parijat, U., Saeed, J., Pranab, K.D., (2011). Factors influencing ERP implementation in Indian manufacturing organizations: A study of micro, small and medium-scale enterprises, Journal of Enterprise Information Management, Vol. 24, No.2, pp. 130-145.

[19] Peček, B., Kovačič, A., (2011). Business Process Management: Use of Simulation in the Public Sector, Ekonomska istraživanja, Vol. 24, No. 1, pp. 95-106.

[20] Popovič, A. and Jaklič, J., (2004). 'Problematika simuliranja poslovnih procesov', In: Management in informatika: Zbornik posvetovanja DSI 2004, pp. 166-172.

[21] Seng, D., Churilov, L., (2003), Business Process-Oriented Information Support for a Higher Education Enterprise, 7th Pacifc Asia Conference on Information Systems, 10-13 July 2003, Adelaide, South Australia, pp. 1055-1074.

[22] Sole, F., (2009). A management model and factors driving performance in public organizations, Measuring Business Excellence, Vol.13, No.4, pp. 3-11.

OPTIMIZATIONS OF FREE POLYNOMIALS

Kristijan Cafuta
University of Ljubljana, Faculty of Electrical Engineering
Tržaška cesta 25, 1000 Ljubljana, Slovenia, E-mail: kristijan.cafuta@fe.uni-lj.si
Igor Klep
The University of Auckland, Department of Mathematics
Private Bag 92019, Auckland 1142, New Zealand, E-mail: igor.klep@auckland.ac.nz
Janez Povh
Faculty of Information Studies in Novo Mesto
Novi trg 5, 8000 Novo mesto, Slovenia, E-mail: janez.povh@fis.unm.si

Abstract: In this paper we present algorithms and their implementations in the computational algorithms package `NCSOStools` for extraction of the global eigenvalue optimizers and extraction of eigenvalue optimizers over the free ball. They are based on free noncommutative analogs of the classical Gram matrix method, which allows us to use semidefinite programming, solution to a truncated free noncommutative moment problem via flat extensions and the Gelfand-Naimark-Segal (GNS) construction.

Keywords: noncommutative polynomial, sum of squares, semidefinite programming, Matlab toolbox, free positivity, flat extension, GNS construction, NCSOStools.

1 INTRODUCTION

What makes (weighted) decompositions of a free polynomial as a sum of hermitian squares interesting are its many facets of applications. A nice survey on applications to control theory, systems engineering and optimization is given by Helton, McCullough, Oliveira, Putinar [7], applications to quantum physics are explained by Pironio, Navascués, Acín [14] who also consider computational aspects related so noncommutative sum of squares. Furthermore, the important Bessis-Moussa-Villani conjecture (BMV) from quantum statistical mechanics is tackled in [11] and by the authors in [3]. We have developed the freely available Matlab software package `NCSOStools` [4] as a consequence of this recent interest in free positivity and (weighted) sums of hermitian squares (*SOHS*). One of its features is the possibility to compute a global or constrained eigenvalue minimum of a symmetric free polynomial. We present the theoretical underpinning of an algorithm to extract the corresponding minimizers. The main ingredients are the noncommutative moment problem and its solution due to McCullough [13], and the Curto-Fialkow theory [6] of how flatness governs the truncated moment problem. Our results were motivated by the method of Henrion and Lasserre [9] for the commutative case.

2 PRELIMINARIES

2.1 Free polynomials and Sum of hermitian squares

Real linear combinations of words in letters $X_1, \ldots, X_n$, including the empty word 1, are denoted by $\mathbb{R}\langle \underline{X} \rangle$ and called *free polynomials*. We denote by $\underline{X}$ the n-tuple of letters $(X_1, \ldots, X_n)$. These free polynomials form a free algebra, which we equip with the *involution* $*$ that fixes $\mathbb{R}$ and letters $X_1, \ldots, X_n$ point-wise and thus reverses words. The subset of $\mathbb{R}\langle \underline{X} \rangle$ consisting of all symmetric free polynomials is denoted by $\operatorname{Sym}\mathbb{R}\langle \underline{X} \rangle := \{f \in \mathbb{R}\langle \underline{X} \rangle \mid f = f^*\}$. A free polynomial of the form g^*g is called a *hermitian square* and the set of all sums of hermitian squares is denoted by Σ^2. Clearly, $\Sigma^2 \subsetneq \operatorname{Sym}\mathbb{R}\langle \underline{X} \rangle$. The length of the longest word in $f \in \mathbb{R}\langle \underline{X} \rangle$ is the *degree* of f and is denoted by $\deg f$. The set of all words and free polynomials with degree $\leq d$ is denoted by $\langle \underline{X} \rangle_d$ and $\mathbb{R}\langle \underline{X} \rangle_d$, respectively. We can stack all words from $\langle \underline{X} \rangle_d$ using the

graded lexicographic order into a column vector W_d. The size of this vector is denoted by $\sigma(d)$. Every $f \in \mathbb{R}\langle \underline{X} \rangle_{2d}$ can be written (possibly nonuniquely) as $f = W_d^* G_f W_d$, where $G_f = G_f^*$ is called a *Gram matrix* for f. Testing whether a given free polynomial $f \in \mathbb{R}\langle \underline{X} \rangle$ is an element of Σ^2 can be done efficiently by using semidefinite programming and the Gram matrix method (the noncommutative version of the classical result for commuting variables) [10, 4]. For a related SOHS decomposition with commutators see [2, 1].

2.2 Hankel matrices, flatness and GNS construction

Definition 2.1. To each linear functional $L : \mathbb{R}\langle \underline{X} \rangle_{2d} \to \mathbb{R}$ we associate a matrix M_d (called a *free Hankel matrix*) indexed by words $u, v \in \langle \underline{X} \rangle_d$, with

$$(M_d)_{u,v} = L(u^* v). \tag{1}$$

If L is *positive*, i.e., $L(p^* p) \geq 0$ for all $p \in \mathbb{R}\langle \underline{X} \rangle_d$, then M_d is positive semidefinite.

Remark 2.2. Note that a matrix M indexed by words of length $\leq d$ satisfying the *free Hankel condition* $M_{u_1,v_1} = M_{u_2,v_2}$ if $u_1^* v_1 = u_2^* v_2$, yields a linear functional L on $\mathbb{R}\langle \underline{X} \rangle_{2d}$ as in (1). If M is positive semidefinite, then L is positive.

Definition 2.3. Let $A \in \mathbb{R}^{s \times s}$ be a symmetric matrix. A (symmetric) extension of A is a symmetric matrix $\tilde{A} \in \mathbb{R}^{(s+\ell)\times(s+\ell)}$ of the form

$$\tilde{A} = \begin{bmatrix} A & B \\ B^t & C \end{bmatrix}$$

for some $B \in \mathbb{R}^{s \times \ell}$ and $C \in \mathbb{R}^{\ell \times \ell}$. Such an extension is *flat* if $\operatorname{rank} A = \operatorname{rank} \tilde{A}$, or, equivalently, if $B = AZ$ and $C = Z^t A Z$ for some matrix Z.

The following is a solution to a free noncommutative moment problem in the truncated case. It resembles the classical results of Curto and Fialkow [6] in the commutative case.

Theorem 2.4. *Suppose $L : \mathbb{R}\langle \underline{X} \rangle_{2d+2} \to \mathbb{R}$ is positive and flat over $L|_{\mathbb{R}\langle \underline{X} \rangle_{2d}}$ (i.e. associated free Hankel matrix M_{d+1} is flat over M_d). Then there is an n-tuple $\underline{A}$ of symmetric matrices of size $s \leq \dim \mathbb{R}\langle \underline{X} \rangle_d$ and a vector v such that*

$$L(p^* q) = \langle p(\underline{A}) v, q(\underline{A}) v \rangle \tag{2}$$

for all $p, q \in \mathbb{R}\langle \underline{X} \rangle$ with $\deg p + \deg q \leq 2d$.

Remark 2.5. For the proof ([4]) we associate to L two positive semidefinite free Hankel matrices, M_{d+1} and its restriction M_d, where M_{d+1} is flat over M_d by an assumption, and then use the Gelfand-Naimark-Segal (GNS) construction.

2.3 Truncated quadratic modules

Given a subset $S \subseteq \operatorname{Sym} \mathbb{R}\langle \underline{X} \rangle$, we introduce

$$\begin{aligned} \Sigma^2_{S,d} &:= \Big\{ \sum_i h_i^* s_i h_i \mid h_i \in \mathbb{R}\langle \underline{X} \rangle,\ s_i \in S,\ \deg(h_i^* s h_i) \leq 2d \Big\}, \\ M_{S,d} &:= \Big\{ \sum_i h_i^* s_i h_i \mid h_i \in \mathbb{R}\langle \underline{X} \rangle,\ s_i \in S \cup \{1\},\ \deg(h_i^* s h_i) \leq 2d \Big\}, \end{aligned} \tag{3}$$

and call $M_{S,d}$ the *truncated quadratic module* generated by S. Note $M_{S,d} = \Sigma^2_d + \Sigma^2_{S,d} \subseteq \mathbb{R}\langle \underline{X} \rangle_{2d}$, where $\Sigma^2_d := M_{\varnothing,d}$ denotes the set of all sums of hermitian squares of free polynomials of degree

at most d. For example, if $S = \{1 - \sum_j X_j^2\}$ then $M_{S,d}$ contains exactly the polynomials f which have a *sum of hermitian squares decomposition over the ball*, i.e., can be written as

$$f = \sum_i g_i^* g_i + \sum_i h_i^* \big(1 - \sum_{j=1}^n X_j^2\big) h_i, \quad \text{where} \tag{4}$$

$$\deg(g_i) \le d, \quad \deg(h_i) \le d-1 \text{ for all } i.$$

We also call a decomposition of the form (4) a *sohs decomposition with weights*.
The truncated quadratic module generated by the generator for the free ball $\mathbb{B}$ is denoted by

$$M_{\mathbb{B},d} := \Big\{ \sum_i h_i^* s_i h_i \mid h_i \in \mathbb{R}\langle \underline{X}\rangle,\ s_i \in \{1 - \sum_j X_j^2,\ 1\},\ \deg(h_i^* s_i h_i) \le 2d \Big\} \subseteq \operatorname{Sym}\mathbb{R}\langle \underline{X}\rangle_{2d}. \tag{5}$$

3 GLOBAL EIGENVALUE OPTIMIZATION OF FREE POLYNOMIALS

In this section we use sums of hermitian squares and semidefinite programming to compute a global (eigenvalue) minimum of a symmetric free polynomial f and give an algorithm to extract the minimizers of f implemented in `NCSOStools` [4].

3.1 Eigenvalue optimization and SDP

Let $f \in \operatorname{Sym}\mathbb{R}\langle \underline{X}\rangle_{2d}$. We are interested in the smallest eigenvalue $f^\star \in \mathbb{R}$ of the polynomial f. That is,

$$f^\star = \inf \big\{ \langle f(\underline{A})v, v\rangle \mid \underline{A} \text{ an } n\text{-tuple of symmetric matrices, } v \text{ a unit vector} \big\}. \tag{6}$$

Hence $f^\star$ is the greatest lower bound on the eigenvalues that $f(\underline{A})$ can attain for n-tuples of symmetric matrices $\underline{A}$, i.e., $(f - f^\star)(\underline{A}) \succeq 0$ for all n-tuples of symmetric matrices $\underline{A}$, and $f^\star$ is the largest real number with this property. Given that a polynomial is positive semidefinite if and only if it is a sum of hermitian squares (the Helton-McCullough SOHS theorem [8, 13]), we can compute $f^\star$ conveniently with SDP. Let

$$\begin{array}{rcl} f^{\text{sohs}} & = & \sup \ \lambda \\ \text{s.t.} & & f - \lambda \ \in \ \Sigma^2. \end{array} \tag{$\text{SDP}_{\text{eig}-\min}$}$$

Then $f^{\text{sohs}} = f^\star$.

In general ($\text{SDP}_{\text{eig}-\min}$) does not satisfy the Slater condition. That is, there does not always exist a *strictly feasible* solution. Nevertheless ($\text{SDP}_{\text{eig}-\min}$) satisfies strong duality [10], i.e., its optimal value f^{sohs} coincides with the optimal value L_{sohs} of the dual SDP:

$$\begin{array}{rcl} L_{\text{sohs}} & = & \inf L(f) \\ \text{s.t.} & & L : \operatorname{Sym}\mathbb{R}\langle \underline{X}\rangle_{2d} \to \mathbb{R} \quad \text{is linear} \\ & & L(1) = 1 \\ & & L(p^*p) \ge 0 \quad \text{for all } p \in \mathbb{R}\langle \underline{X}\rangle_d. \end{array} \tag{$\text{DSDP}_{\text{eig}-\min}$}_d$$

3.2 The extraction of optimizers

In this subsection we investigate the attainability of $f^\star$ and explain how to extract the minimizers $(\underline{A}, v)$ for f from (6) if the lower bound $f^\star$ is attained. That is,

$$f^\star = \langle f(\underline{A})v, v\rangle. \tag{7}$$

Of course, in general f will not be bounded from below and even if f is bounded, the infimum $f^\star$ need not be attained [4].

In the sequel our main interest lies in the case where $f^\star$ is attained. We shall see below (see Corollary 3.2) that this happens if and only if the infimum $L_{\text{sohs}} = f^{\text{sohs}} = f^\star$ for ($\text{DSDP}_{\text{eig}-\min}$)$_{d+1}$ is attained.

Proposition 3.1. *Let $f \in \operatorname{Sym}\mathbb{R}\langle \underline{X}\rangle_{2d}$ be bounded from below. If the infimum L_{sohs} for $(\text{DSDP}_{\text{eig}-\min})_{d+1}$ is attained, then it is attained at a linear map L that is flat over its own restriction to $\mathbb{R}\langle \underline{X}\rangle_{2d}$.*

Proof. Let L be a minimizer for $(\text{DSDP}_{\text{eig}-\min})_{d+1}$. To it we associate M_{d+1} and its restriction M_d. Then

$$M_{d+1} = \begin{bmatrix} M_d & B \\ B^t & C \end{bmatrix}$$

for some B, C. Since M_{d+1} and M_d are positive semidefinite, $B = M_d Z$ and $C \succeq Z^t M_d Z$ for some Z. Now form a "new" M_{d+1}:

$$\tilde{M}_{d+1} = \begin{bmatrix} M_d & B \\ B^t & Z^t M_d Z \end{bmatrix} = \begin{bmatrix} I & Z \end{bmatrix}^t M_d \begin{bmatrix} I & Z \end{bmatrix}.$$

This matrix is obviously flat over M_d, positive semidefinite, and satisfies the free Hankel condition. So it yields a positive linear map $\tilde{L}$ on $\mathbb{R}\langle \underline{X}\rangle_{2d+2}$ flat over $\tilde{L}|_{\mathbb{R}\langle \underline{X}\rangle_{2d}} = L|_{\mathbb{R}\langle \underline{X}\rangle_{2d}}$. Moreover, $\tilde{L}(f) = L(f) = L_{\text{sohs}}$. □

From Proposition 3.1 and Theorem 2.4 we deduce

Corollary 3.2. *Let $f \in \mathbb{R}\langle \underline{X}\rangle_{2d}$. Then $f^\star$ is attained if and only if there is a feasible point L for $(\text{DSDP}_{\text{eig}-\min})_{d+1}$ satisfying $L(f) = f^\star$.*

For $f \in \operatorname{Sym}\mathbb{R}\langle \underline{X}\rangle_{2d}$ we can state the following algorithm for the extraction of optimizers.

Step 1: Solve $(\text{DSDP}_{\text{eig}-\min})_{d+1}$. If the problem is unbounded or the optimum is not attained, STOP. Otherwise let L denote an optimizer.

Step 2: To L we associate the positive semidefinite matrix $M_{d+1} = \begin{bmatrix} M_d & B \\ B^t & C \end{bmatrix}$. Modify M_{d+1}: $\tilde{M}_{d+1} = \begin{bmatrix} M_d & B \\ B^t & Z^t M_d Z \end{bmatrix}$, where Z satisfies $M_d Z = B$. This matrix yields a positive linear map $\tilde{L}$ on $\mathbb{R}\langle \underline{X}\rangle_{2d+2}$ which is flat over $\tilde{L}|_{\mathbb{R}\langle \underline{X}\rangle_{2d}} = L|_{\mathbb{R}\langle \underline{X}\rangle_{2d}}$. In particular, $\tilde{L}(f) = L(f) = f^\star$.

Step 3: Use the GNS construction on $\tilde{L}$ to compute symmetric matrices A_i and a vector v with $\tilde{L}(f) = f^\star = \langle f(\underline{A})v, v\rangle$.

Remark 3.3. We finish this section by emphasizing that the extraction of eigenvalue optimizers *always* works if the optimum for $(\text{DSDP}_{\text{eig}-\min})_{d+1}$ is attained. This is in sharp contrast with the commutative case; cf. [12].

4 EIGENVALUE OPTIMIZATION OF FREE POLYNOMIALS OVER THE FREE BALL

In this section we consider the eigenvalue optimization of free polynomials over the free ball. We can rephrase $f_\star^{\mathbb{B}}$, the greatest lower bound on the eigenvalues of $f \in \mathbb{R}\langle \underline{X}\rangle_{2d}$ over the ball $\mathbb{B}$, as follows:

$$\begin{array}{rcccl} f_\star^{\mathbb{B}} & = & f_{\text{sohs}}^{\mathbb{B}} & = & \sup \ \lambda \\ & & & \text{s.t.} & f - \lambda \in M_{\mathbb{B},d+1}. \end{array} \qquad (\text{PSDP}_{\text{eig}-\min}^{\mathbb{B}})$$

Verifying whether $f \in M_{\mathbb{B},d}$ is a semidefinite programming feasibility problem [5]:

Proposition 4.1. *Let $f = \sum_{w\in\langle \underline{X}\rangle_{2d}} f_w w$. Then $f \in M_{\mathbb{B},d}$ if and only if there exist positive semidefinite matrices H and G of order $\sigma(d)$ and $\sigma(d-1)$, respectively, such that for all $w \in \langle \underline{X}\rangle_{2d}$,*

$$f_w = \sum_{\substack{u,v\in\langle \underline{X}\rangle_d \\ u^*v=w}} H(u,v) + \sum_{\substack{u,v\in\langle \underline{X}\rangle_{d-1} \\ u^*v=w}} G(u,v) - \sum_{j=1}^{n} \sum_{\substack{u,v\in\langle \underline{X}\rangle_{d-1} \\ u^*X_j^2v=w}} G(u,v). \tag{8}$$

Remark 4.2. From Proposition 4.1 it follows how to construct the sohs decomposition with weights (4) for $f \in M_{\mathbb{B},d}$. First we solve the semidefinite feasibility problem in the variables $H \in \mathcal{S}^+_{\sigma(d)}$, $G \in \mathcal{S}^+_{\sigma(d-1)}$ subject to constraints (8), where $\mathcal{S}^+_k$ denotes the set of all real positive semidefinite $k \times k$ real matrices. Then we compute by Cholesky or eigenvalue decomposition vectors $H_i \in \mathbb{R}^{\sigma(d)}$ and $G_i \in \mathbb{R}^{\sigma(d-1)}$ such that $H = \sum_i H_i H_i^t$ and $G = \sum_i G_i G_i^t$. Polynomials h_i and g_i from (4) are computed as $h_i = H_i^t W_d$ and $g_i = G_i^t W_{d-1}$.

By Proposition 4.1, the problem $(\mathrm{PSDP}^{\mathbb{B}}_{\mathrm{eig-min}})$ is a SDP; it can be reformulated as

$$\begin{array}{rcl} f^{\mathbb{B}}_{\mathrm{sohs}} & = & \sup \;\; f_1 - \langle E_{1,1},\, H\rangle - \langle E_{1,1},\, G\rangle \\ \text{s.t.} & & f_w \;=\; \displaystyle\sum_{\substack{u,v\in\langle \underline{X}\rangle_{d+1} \\ u^*v=w}} H(u,v) + \sum_{\substack{u,v\in\langle \underline{X}\rangle_d \\ u^*v=w}} G(u,v) - \sum_{j=1}^{n} \sum_{\substack{u,v\in\langle \underline{X}\rangle_d \\ u^*X_j^2v=w}} G(u,v), \\ & & \quad \text{for all } 1 \neq w \in \langle \underline{X}\rangle_{2d+2}, \\ & & H \;\in\; \mathcal{S}^+_{\sigma(d+1)}, \;\; G \;\in\; \mathcal{S}^+_{\sigma(d)}. \end{array} \tag{PSDP'$^{\mathbb{B}}_{\mathrm{eig-min}}$}$$

The dual semidefinite program to $(\mathrm{PSDP}^{\mathbb{B}}_{\mathrm{eig-min}})$ and $(\mathrm{PSDP'}^{\mathbb{B}}_{\mathrm{eig-min}})$ is:

$$\begin{array}{rcl} L^{\mathbb{B}}_{\mathrm{sohs}} & = & \inf L(f) \\ \text{s.t.} & & L : \mathrm{Sym}\,\mathbb{R}\langle \underline{X}\rangle_{2d+2} \to \mathbb{R} \quad \text{is linear} \\ & & L(1) = 1 \\ & & L(q^*q) \geq 0 \quad \text{for all } q \in \mathbb{R}\langle \underline{X}\rangle_{d+1} \\ & & L(h^*(1 - \sum_j X_j^2)h) \geq 0 \quad \text{for all } \; h \in \mathbb{R}\langle \underline{X}\rangle_d. \end{array} \tag{DSDP$^{\mathbb{B}}_{\mathrm{eig-min}}$}$_{d+1}$$

Remark 4.3. Having Slater points for $(\mathrm{DSDP}^{\mathbb{B}}_{\mathrm{eig-min}})_{d+1}$ is important for the clean duality theory of SDP to kick in [15]. In particular, there is no duality gap, so $L^{\mathbb{B}}_{\mathrm{sohs}} = f^{\mathbb{B}}_{\mathrm{sohs}} (= f^{\mathbb{B}}_{\star})$. Since also the optimal value $f^{\mathbb{B}}_{\mathrm{sohs}} > -\infty$, $f^{\mathbb{B}}_{\mathrm{sohs}}$ is attained. More important for us and the extraction of optimizers is the fact that $L^{\mathbb{B}}_{\mathrm{sohs}}$ is attained, as we shall explain in Subsection 4.1.

4.1 The extraction of optimizers

In this subsection we establish the attainability of $f^{\mathbb{B}}_{\star}$ on $\mathbb{B}$, and explain how to extract the minimizers $(\underline{A}, \xi)$ for f (for more details see [5]).

Proposition 4.4. *$f \in \mathrm{Sym}\,\mathbb{R}\langle \underline{X}\rangle_{2d}$. There exists an n-tuple $\underline{A} \in \mathbb{B}(\sigma(d))$, and a unit vector $\xi \in \mathbb{R}^{\sigma(d)}$ such that*

$$f^{\mathbb{B}}_{\star} = \langle f(\underline{A})\xi, \xi\rangle. \tag{9}$$

In other words, the infimum in (6) *is really a minimum.*

Corollary 4.5. *$f \in \mathrm{Sym}\,\mathbb{R}\langle \underline{X}\rangle_{2d}$. Then there exist linear functionals*

$$L : \mathrm{Sym}\,\mathbb{R}\langle \underline{X}\rangle_{2d+2} \to \mathbb{R}$$

such that L is feasible for $(\mathrm{DSDP}^{\mathbb{B}}_{\mathrm{eig-min}})_{d+1}$, and we have

$$L(f) = f^{\mathbb{B}}_{\star}. \tag{10}$$

For $f \in \operatorname{Sym}\mathbb{R}\langle \underline{X}\rangle_{2d}$ we can state the algorithm implemented in `NCSOStools` showing how the optimizers $(\underline{A}, \xi)$ can be extracted from the solutions of the constructed SDPs.

Step 1: Solve $(\mathrm{DSDP}_{\mathrm{eig-min}})_{d+1}$. Let L denote an optimizer, i.e., $L(f) = f_\star^{\mathbb{B}}$.

Step 2: To L we associate the positive semidefinite matrix $H_L = \begin{bmatrix} H_{\check{L}} & B \\ B^t & C \end{bmatrix}$. Modify H_L:

$$H_{\hat{L}} = \begin{bmatrix} H_{\check{L}} & B \\ B^t & Z^t H_{\check{L}} Z \end{bmatrix},$$

where Z satisfies $H_{\check{L}} Z = B$. This matrix yields a flat positive linear map $\hat{L}$ on $\mathbb{R}\langle \underline{X}\rangle_{2d+2}$ satisfying $\tilde{L}|_{\mathbb{R}\langle \underline{X}\rangle_{2d}} = L|_{\mathbb{R}\langle \underline{X}\rangle_{2d}}$. In particular, $\tilde{L}(f) = L(f) = f_\star^{\mathbb{B}}$.

Step 3: Use the GNS construction on $\tilde{L}$ to compute symmetric matrices A_i and a unit vector ξ with $\tilde{L}(f) = f_\star^{\mathbb{B}} = \langle f(\underline{A})\xi, \xi\rangle$.

References

[1] S. Burgdorf, K. Cafuta, I. Klep, and J. Povh. Algorithmic aspects of sums of Hermitian squares of noncommutative polynomials. *Comput. Optim. Appl.*, 55(1):137–153, 2013.

[2] S. Burgdorf, K. Cafuta, I. Klep, and J. Povh. The tracial moment problem and trace-optimization of polynomials. *Math. Program.*, 137(1):557–578, 2013.

[3] K. Cafuta, I. Klep, and J. Povh. A note on the nonexistence of sum of squares certificates for the Bessis-Moussa-Villani conjecture. *J. Math. Phys.*, 51(8):083521, 10, 2010.

[4] K. Cafuta, I. Klep, and J. Povh. NCSOStools: a computer algebra system for symbolic and numerical computation with noncommutative polynomials. *Optim. Methods and Softw.*, 26(3):363–380, 2011. `http://ncsostools.fis.unm.si/`

[5] K. Cafuta, I. Klep, and J. Povh. Constrained polynomial optimization problems with noncommuting variables. *Siam Journal on Optimization*, 22(2):363–383, 2012.

[6] R. Curto and L. Fialkow. Solution of the truncated complex moment problem for flat data. *Mem. Amer. Math. Soc.*, 119(568):x+52, 1996.

[7] M. de Oliveira, J. Helton, S. McCullough, and M. Putinar. Engineering systems and free semi-algebraic geometry. In *Emerging applications of algebraic geometry*, volume 149 of *IMA Vol. Math. Appl.*, pages 17–61. Springer, New York, 2008.

[8] J. Helton. "Positive" noncommutative polynomials are sums of squares. *Ann. of Math. (2)*, 156(2):675–694, 2002.

[9] D. Henrion and J. B. Lasserre. Detecting global optimality and extracting solutions in GloptiPoly. In *Positive polynomials in control*, volume 312 of *Lecture Notes in Control and Inform. Sci.*, pages 293–310. Springer, Berlin, 2005.

[10] I. Klep and J. Povh. Semidefinite programming and sums of hermitian squares of noncommutative polynomials. *J. Pure Appl. Algebra*, 214:740–749, 2010.

[11] I. Klep and M. Schweighofer. Sums of Hermitian squares and the BMV conjecture. *J. Stat. Phys*, 133(4):739–760, 2008.

[12] J. B. Lassere. *Moments, positive polynomials and their applications*, volume 1 of *Imperial College Press Optimization Series*. Imperial College Press, London, 2009.

[13] S. McCullough. Factorization of operator-valued polynomials in several non-commuting variables. *Linear Algebra Appl.*, 326(1-3):193–203, 2001.

[14] S. Pironio, M. Navascués, and A. Acín. Convergent relaxations of polynomial optimization problems with noncommuting variables. *SIAM J. Optim.*, 20(5):2157–2180, 2010.

[15] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Rev.*, 38(1):49–95, 1996.

Computing the equity of a poker hand by Integer Linear Programming

Marcello Dalpasso* Giuseppe Lancia†

Abstract: We illustrate how Integer Linear programming techniques can be applied to the popular game of poker Texas Hold'em in order to evaluate the strength of a hand. In particular, we give models aimed at (i) minimizing the number of features that a player should look at when estimating his winning probability (called his *equity*); (ii) giving weights to such features so that the equity is approximated by the weighted sum of the selected features. We show that ten features or less are enough to estimate the equity of a hand with high precision.

Keywords: Poker Texas Hold'em; Integer Linear Programming; Equity.

1 Introduction

No-limit Texas Hold'em [2, 3] (NLTH) is a form of poker that has gained huge popularity in the past few years. This game is played with a full deck of 52 cards by two up to nine players. The game develops in four phases. In the first phase (called the *preflop*), each player is dealt two private cards (i.e., known only to him). These cards are called his *starting hand*. Then there is a round of betting in which some players fold, while others remain in play. Then three cards are turned face-up. These cards are called the *flop*, and they are community cards, i.e., they can be used by all players still in play. After another round of betting, if two or more players are still in play, a single community card, called the *turn*, is turned face-up. After another round of betting, another single community card is turned, called the *river*. A final round of betting follows. Each player still in play eventually computes the best 5-cards hand obtained by combining his starting cards with the five community cards (i.e., the best 5 cards out of 7). The owner of the highest-score hand wins the pot, and the pot is split in case of a tie.

The strength of all starting hands has been assessed in many books and by extensive use of computer programs. Since in NLTH the four suits have all the same value, the $\binom{52}{2}$ starting hands can be reduced to only 169 possibilities: 13 pairs, 78 non-pairs suited, and 78 non-pairs offsuit (i.e., of two different suits). It is now well-known that the best preflop starting hand is a pair of aces, followed by a pair of kings, while 72o (seven-deuce offsuit) is considered to be the worst starting hand. There are published tables of 169 entries listing the preflop strength of all starting hands, but similar tables are impractical for the $\binom{52}{2} \times \binom{50}{3} = 25,989,600$ combinations (h, f), where h is a starting hand, and f is a flop. The subject of our paper is to investigate some possible ways to compute the strength of a hand once a particular flop has been exposed.

*DEI – Dip. di Ingegneria dell'Informazione, University of Padova. marcello.dalpasso@unipd.it
†DIMI – Dip. di Matematica e Informatica, University of Udine. giuseppe.lancia@uniud.it

Equity of hand vs hand. When only two players remain in play, we say that they are *heads up*. The flop is a crucial point in a hand, when many times two players remain heads up. The flop is crucial since it is the time when most of the cards are exposed and the players try to estimate the potentiality of their hand with only two cards still to come.

Suppose a particular flop $f = \{c_1, c_2, c_3\}$ has been exposed (where the c_i are three specific cards), and that two players, A and B, are heads up. A holds $\{a_1, a_2\}$, while B holds $\{b_1, b_2\}$. What is the probability p_w that, after the turn c_4 and the river c_5 have been exposed, A will hold the winning hand? Let us call p_w such probability. It is easy to compute p_w by a simple algorithm: We loop over all possible values for $\{c_4, c_5\}$ (there are $\binom{45}{2} = 990$ such pairs). For each pair we record if A wins or if the hand is a tie. Let n_w be the number of wins for A and n_t the number of ties. Since in case of a tie the pot is split between the winners, for each tie we assign 1/2 win to A and 1/2 to B, so that $p_w = (n_w + \frac{1}{2}n_t)/990$. The probability for a hand to be the winner after all cards have been exposed, given some cards that have already been exposed, is called the hand's *equity*. A hand has a preflop equity, a flop equity, a turn equity and a river equity. In this paper we are particularly concerned with the flop equity, that is the most crucial (and difficult to estimate) in the course of the play. Many important decisions are based on good estimates of the flop equity of a hand. In particular, *commitment* decisions (i.e., decisions that put at stake all of our chips) can be based on the so called *pot-odds* which, in turn, are based on our equity. Loosely speaking, if the money that we can gain (i.e., the pot) is (or is expected to become) v_P, and it costs us v_u to play, and our equity is E, playing is profitable if $E > v_u/(v_u + v_P)$ and unprofitable otherwise.

Equity of hand vs range. In the game of poker it is very difficult (or, more likely, impossible) to pinpoint the opponent's starting hand to just one possibility. In practice, all one can do is to formulate an educated guess on a set of hands (the fewer, the better) that his opponent might be holding in a specific situation. In the poker jargon, such a set of hands is called a *range*.

We will consider ranges in which all hands have the same probability of being the actual hand that a player holds. That is, if R is a range, then each hand in R is equally likely and has probability $1/|R|$. More realistically, we could have defined a range in such a way that different hands in R could have different probabilities (for instance, in a particular situation a player could hold either AK (ace-king) or AQ (ace-queen), but, given that we know he his somewhat conservative, it could be more likely than he has AK than AQ). Generalizing our arguments to ranges in which each hand has its own probability is easy, and is delayed to the journal version of this extended abstract.

A range is a subset of all possible starting hands. Once the flop is exposed, the range is in fact a subset of $\binom{47}{2}$ starting hands, namely all hands that do not include any of the cards that we hold or that belong to the flop. Assume player A holds a hand h, player B holds one hand from a specific range R, and a flop f has been exposed. The equity of h vs R on this flop, denoted by $E(h, f, R)$ is the probability that, after turning two more cards (turn and river) the player A will be the winner.

Computing the equity of a hand vs a range with a computer is very easy, and there are many online sites that offer this service (see, e.g., [5]). It is just a matter of computing the average equity of h vs each hand $h' \in R$, as explained before. Of course this could be a demanding task (for instance, if R is the range "ATC" –any two cards– there are almost 1000 hand-vs-hand computations to be made). Needless to say, such computations are

impossible to be made mentally at the poker table.

Our approach The main goal of this paper is to show that mathematical programming techniques [4] can be applied to the game of poker to study the strength of a hand versus a range after the flop. Our objective is to come up with some relatively simple formulas that can potentially be computed mentally by a player at the table and give him his equity on any flop. In order to be simple, the formula has to rely on as few "features" of the flop texture as possible. Our approach can be applied to a given hand h versus a given range R. In the final, extended, version of this paper we will take some important cases for h such as strong pairs, strong non-pairs, suited connectors (e.g., 8-9 suited), and run them against some ranges taken from the literature. For space and time limitations, we will only consider two hands (namely JJ and AKs) in this abstract. We define a set of n binary features that each flop may or may not possess (e.g., "is there an ace on the flop?", "is the flop all of the same suit?", "is there a pair on the flop?", etc.), so that to each flop there corresponds a binary n-vector. Then, by using Integer Linear Programming, we both select a subset of "few" features to look at and assign weights to the selected features in such a way that the equity of each flop can be estimated with high precision by the weighted sum of the features possessed by each flop. In the remainder of the paper we will elaborate on this technique.

2 Ranges and features

The community of poker players regards David Sklansky as the person who laid the mathematical foundations of the game of poker. In one of his pioneering books [3], he suggested the division of all starting hands into nine groups (today known as *Sklansky groups* [1]), in such a way that all hands in the same group have roughly the same strength, and are stronger than all hands in groups that follow. For instance, group number 1 consists of the pairs AA, KK, QQ, JJ and of the unpaired suited hand AKs (we refer the reader to [3] for the remaining groups).

From Sklansky groups we have derived four ranges, namely (i) *Range ultra-strong (RUS)* (Sklansky group-1 hands); (ii) *Range strong (RS):* (groups 1 and 2); (iii) *Range medium-loose (RML):* (groups $1, \ldots, 5$); (iv) *Range any-two cards (ATC)*: (all possible starting hands). These ranges are meant to represent the possible holding of a player *based on the preflop action*, i.e., right before the flop is exposed. In poker, given specific preflop situations, experienced players are able to assign certain ranges to their opponents. Let us look at some examples: (i) If player A raises the pot, player B re-raises and then player C puts in a third raise, then it is very likely that C has a *very* strong hand, something like a pair of aces or kings. (ii) If, in the previous situation, player C does not raise, but still elects to call, then it is likely that he has a quite strong hand. (iii) If everybody folds, and just two players (the blinds) remain in play without any raise, then each of them can literally have any two cards.

The number of possible flops once a starting hand is known is $\binom{50}{3} = 19600$. Each flop has some peculiar characteristics. We want to characterize the flops by means of binary features. For instance, the flop [$K\heartsuit$ $Q\diamondsuit$ $8\clubsuit$] "has a King", "has two cards 10 or above", "has three different suits", etc. We have defined about a hundred different binary features. We will just briefly mention some of them here[1]. The goal of our model

[1]For space limitations we do not report all the features we used. They can be found at [6]

has then been to identify a small subset of these features which is sufficient to look at in order to compute the equity of any flop with high precision. Among the features we used we recall:

- The flop has $x = 3, 2, 1$ ranks	- It has $x = 3, 2, 1$ suits
- It allows for a straight/flush/both	- It has $x = 0, 1, \geq 2$ aces/kings
- It has $x = 0, 1, 2, 3$ "high" cards (Ten or above)	- It has $x = 0, 1, 2, 3$ "low" cards (≤ 8)
- It has no draws (such as 2 8 K in three suits)	. . .

Some features do not involve only the flop, but the flop *in combination* with the hand we hold. For instance "did we pair at least one of our starting cards?"; "do we have $x = 0, 1, 2$ overcards to the flop?", etc. Furthermore, some features are not "pure", but rather they are logical combinations ($\wedge, \vee, \neg$) of more elementary features. For instance, "do we have a backdoor straight (i.e., $\frac{3}{5}$ of a straight) and a backdoor flush and a pair?"; "Is it true that the flop is all of same suit and we do not have a card of that suit?" etc.

3 ILP for computing equities

In this section we describe our Integer Linear Programming (ILP) models for computing the equity of a certain hand h versus a certain range R. The actual hands and ranges utilized will be detailed in Section 4.

We will have two ILP models, one relative to the maximum error, and one to the average error of the estimated equity w.r.t. the true equity. Our models will be parametrized with parameters $\bar{N}$ and $\bar{E}$. More specifically, we can (i) fix $\bar{N}$ to be the maximum number of features that can be used, and then minimize the maximum (or the average) error, or (ii) fix $\bar{E}$ to be a threshold for the accepted maximum (respectively, average) error and minimize the number of features sufficient to obtain an error within the threshold.

There are $m = 19600$ flops $f^1, \ldots, f^m$. Each flop f^i does or does not possess each of n binary features F_j, and we denote by $F_j(f^i) \in \{0, 1\}$ the absence/presence of a particular feature in a given flop.

The first step consists in considering each flop f^i in turn, and computing the exact equity $e^i := E(h, f^i, R)$ of h vs R (we did this by a `C#` code that we developed, but the step can also be done by resorting to online sites that compute equities such as [5]). Moreover, we check all features on f^i, thus obtaining a binary vector $b^i = (b^i_1, \ldots, b^i_n)$ with $b^i_j = F_j(f^i)$. At the end, we have a matrix B of m rows and $n+1$ columns (the last column contains the equities $(e^1, \ldots, e^m)^T$). The matrix B is (part of) the constraint matrix of our linear programs, to be defined.

We introduce n binary variables y_j. Each variable can allow (if $y_j = 1$) or forbid (if $y_j = 0$) the use of the corresponding feature. Furthermore, we define real variables x_j, for $j = 1, \ldots, n$. Each variable x_j is the weight that we associate to a feature. The weight is meant to represent how much equity we gain (if $x_j > 0$) or we lose (if $x_j < 0$) when the feature F_j is present. Being a probability, it is $|x_j| \leq 1$. Let ϵ be a variable representing the maximum error (i.e., difference in absolute value between the estimated equity and the true equity of a flop). We obtain the following ILP for minimizing the maximum error, given a budget $\bar{N}$ on how many features can be used altogether:

$$\texttt{MAXERR}(\bar{N}) := \min \epsilon \tag{1}$$

s.t.

$$-y_j \leq x_j \leq y_j \qquad \forall j = 1, \ldots, n \tag{2}$$

$$\sum_{j=1}^{n} y_j \leq \bar{N} \tag{3}$$

$$e^i - \epsilon \leq \sum_{j=1}^{n} b^i_j x_j \leq e^i + \epsilon \qquad \forall i = 1, \ldots, m \tag{4}$$

$$\epsilon \geq 0,\ x_j \in \mathbb{R},\ y_j \in \{0, 1\} \qquad \forall j = 1, \ldots, n. \tag{5}$$

The objective function (1) calls for minimizing the error ϵ. Constraints (2) are "activation" constraints: when a variable y_j is 1, the weight x_j of the corresponding feature can be non-null. When $y_j = 0$ the weight x_j must be null. Constraint (3) says that we can use at most $\bar{N}$ features. Constraints (4) force that, for each flop f^i, the estimated equity differs from the true equity (in excess or in defect) by at most ϵ. The model has $2(n + m) + 1$ constraints and $2n + 1$ variables, n of which are integer.

From the above model, it is easy to derive a model in which we minimize the number of features sufficient to stay within a certain maximum error. We simply need to use the objective $\texttt{FEAT\_MAXERR}(\bar{E}) := \min \sum_{j=1}^{n} y_j$ under constraints (2) and (4) in which the variable ϵ has been replaced by the constant $\bar{E}$.

We now turn to the model for minimizing the average error. In this model we have real variables ϵ^i, for $i = 1, \ldots, m$, that represent the error in estimating the equity of each specific flop f^i. The model calls for the minimization of $\texttt{AVGERR}(\bar{N}) := \frac{1}{m} \sum_{i=1}^{m} \epsilon_i$ under constraints (2), (3) and

$$e^i - \epsilon_i \leq \sum_{j=1}^{n} b^i_j x_j \leq e^i + \epsilon_i \qquad \forall i = 1, \ldots, m \tag{6}$$

with variables $\epsilon_i \in \mathbb{R}^+$, $x_j \in \mathbb{R}$ and $y_j \in \{0, 1\}$. The model has $2(n + m) + 1$ constraints and $m+2n$ variables, n of which are integer. Again, we can obtain from the above a model for minimizing the number of features sufficient to stay within a certain average error. We simply need to use the objective $\texttt{FEAT\_AVGERR}(\bar{E}) := \min \sum_{j=1}^{n} y_j$ under constraints (2), (6) and with the constraint

$$\frac{1}{m} \sum_{i=1}^{m} \epsilon_i \leq \bar{E}. \tag{7}$$

4 Computational results and conclusions

For human players, computing the equity of a hand vs a range on the flop is more of an art than a science, requiring mathematical, analytical, and also psychological skills. Being able to estimate the equity with high accuracy is what separates skilled professional players from the rest. Yet, the estimates can never be too accurate due to the high number of factors in play. In this paper we have shown how ILP can be used to select relevant features of the flop and weights to assign to such features so that one can approximate the equity of a hand versus a range by a weighted sum with few addends.

hand	$\bar{N}$	Ranges			
		RUS	RS	RML	ATC
		avg (max)	avg (max)	avg (max)	avg (max)
JJ	20	1.4% (9.4%)	1.7% (10.3%)	1.7% (7.2%)	0.9% (3.9%)
	10	1.8% (11.5%)	2.3% (13.5%)	2.0% (10%)	1.7% (5.5%)
	5	2.5% (18%)	3.2% (20.5%)	3.3% (15.5%)	2.5% (8.5%)
AKs	20	1.8% (12%)	1.8% (9.5%)	2.5% (11.7%)	1.4% (12%)
	10	2.5% (13.8%)	2.5% (12.3%)	3.1% (12.8%)	2.0% (12%)
	5	4.8% (26%)	4.3% (22.5%)	4.5% (19.5%)	3.2% (17.5%)

Table 1: Results for JJ and AKs vs all ranges for average and maximum error.

This is the first time that Mathematical Programming techniques have been applied to the game of poker for estimating the strength of a hand. For space and time limitations, we have only considered two hands (namely, JJ representative of strong pairs, and AKs, representative of strong suited non-pairs). It is clear that our techniques can be applied to any hand versus any range, and in the full version of this paper we will take in consideration a larger number of starting hands.

In Table 1 we list our results. For a maximum number of features $\bar{N} = 5, 10, 20$ we have selected the best $\bar{N}$ features in order to optimize the average error and the maximum error. The minimization of the maximum error shows that there exist some flops that are very difficult to estimate with only few features. For instance, the error for JJ with 10 features is around 10% which is not too good. However, the method shows its strength when we consider the *average* error over all the 19600 flops. The average error is much more important, since although there may be a few flops for which the estimate can be quite off, on the vast majority of the flops the estimate is very accurate.

It is very surprising, and unknown prior to this article, that by looking at only five characteristics of a flop, it is possible, for instance, to estimate the equity of JJ vs any range with an average error of just around 3%. It is important to remark that an error of 3% is usually considered negligible. Poker is a very hard game of randomness and incomplete information, and humans look for simple rules of strategy. For instance, preflop, KQs vs 99 have equity of 46.5% and 53.5% respectively, but this matchup is considered a coin-flip (50-50). Similarly, books report that, if we have a 4/5 draw (flush or straight) on the flop, the probability of completing the draw is 1 in 3 while actually it is about 35% for the flush, and about 31% for the straight.

References

[1] `en.wikipedia.org/wiki/Texas_hold_'em_starting_hands#Sklansky_hand_groups`

[2] D. Sklansky, The Theory of Poker. Two Plus Two Publications. Henderson, NV, 1999.

[3] D. Sklansky and M. Malmuth, Hold 'em Poker for Advanced Players. Two Plus Two Publications. Henderson, NV, 1999.

[4] L. Wolsey and G. Nemhauser. Integer and Combinatorial Optimization. 1999. Wiley

[5] `www.cardplayer.com/poker-tools/odds-calculator/texas-holdem`

[6] `www.dei.unipd.it/~dalpasso/pokerFeatures.pdf`

THE IMPROVEMENT OF THE HOLT-WINTERS METHOD FOR INTERMITTENT DEMAND: A CASE OF OVERNIGHT STAYS OF TURISTS FOR SOME COMMUNITY IN REPUBLIC OF SLOVENIA

Liljana Ferbar Tratar[a] and **Ana Vehovec**[b]
[a]Faculty of Economics, University of Ljubljana, Kardeljeva pl. 17, 1000 Ljubljana, Slovenia
liljana.ferbar.tratar@ef.uni-lj.si
[b]DRI, upravljanje investicij, d.o.o., Kotnikova 40, 1000 Ljubljana

Abstract: Demand forecasting is used frequently in the world because of expedient source management and because the need for planning is becoming more important. Different methods of forecasting can be used, although exponential smoothing methods are most often used in practice because a lot of different products are forecasted and at the same time they are simple, fast and inexpensive. But Holt-Winters (HW) methods are not accurate enough for demand data showing too high a variation, often a property of real data. In this paper we propose an improved HW method and through results we demonstrate that a reduction in forecast error (MSE) can be reached.

Keywords: Demand forecasting, Holt-Winters method, Optimization.

1 INTRODUCTION

Exponential smoothing is used substantially throughout the world, because the method is simple, fast and inexpensive. It is particularly suitable for production planning and stock control, wherein forecasts are made with a large number of variables (stock accuracy forecasting is particularly important, because excessive forecasts lead to over-stocks and insufficient forecasting lead to stocks shortage) ([4]).

Exponential smoothing methods are a class of methods that produce forecasts with simple formulae, taking into account trend and seasonal effects of data. These procedures are widely used as forecasting techniques in inventory management and sales forecasting. Some papers ([5], [10]) have stimulated renewed interest in the technique, putting exponential smoothing procedures on sound theoretical ground by identifying and examining the underlying statistical models. Moreover, while exponential smoothing methods give reliable post-sample forecasts it would be worthwhile to develop procedures that would identify the most appropriate method ([7], [8] and [9]).

The HW method estimates three smoothing parameters - associated with level, trend and seasonal factors. The seasonal variation can be of either an additive or multiplicative form. The multiplicative version is used more widely and on average works better than the additive ([1]; of course, if a data series contains some values equal to zero, the multiplicative HW method could not be used). A problem which affects all exponential smoothing methods is the selection of smoothing parameters and initial values, so that forecasts would fit better into time series data ([3]). We estimate smoothing and initial parameters in HW methods by minimising the mean square error (MSE). The minimising problem is solved by using Solver (Microsoft Excel 2007).

The aim of the article is to expose the problem of demand forecasting involving data showing high variations. In this paper we present an improved HW method and we show that a reduction in forecast error (MSE) can be achieved. From the results obtained for real data we prove that the proposed method is more efficient than the ordinary HW method.

The remainder of the paper is organized as follows. We begin with the description of the Holt-Winters forecasting procedure and we present an improved Holt-Winters procedure (see Section 2). In Section 3, we present the calculations and results which allow us to

compare different forecasting methods. Finally, in Section 4, after the conclusions of our paper some further research steps are suggested.

2 THE HW PROCEDURE AND IMPROVED HW PROCEDURE

The Holt-Winters method of exponential smoothing involves trend and seasonality and is based on three smoothing equations: equation for level, equation for trend and equation for seasonality. The decision as to which method to use depends on time series characteristics: the additive method is used when the seasonal component is constant, the multiplicative method is used when the size of the seasonal component is proportional to the trend level ([2]). In other words: if a time series is presented on a chart, in case of additive seasonality a series exhibits constant seasonal fluctuations regardless of the variable level L_t; in case of multiplicative seasonality the size of seasonal fluctuations alters in dependence of total average of variable L_t.

2.1 Holt-Winters' additive procedure

The basic equations for the Holt-Winters additive method are:

Equation for level:

$$L_t = \alpha(Y_t - S_{t-s}) + (1-\alpha)(L_{t-1} + b_{t-1}) \tag{1}$$

Equation for trend:

$$b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1} \tag{2}$$

Equation for seasonality (seasonal index):

$$S_t = \gamma(Y_t - L_t) + (1-\gamma)S_{t-s} \tag{3}$$

Forecast for *m* period equals:

$$F_{t+m} = L_t + b_t m + S_{t-s+m} \tag{4}$$

where are L_t – estimation of variable in time *t*, Y_t – observed value, b_t – trend estimation of time series in time *t*, S_t – estimation of seasonality in time *t*, α, β, γ – smoothing parameters in the interval [0, 1], *m* – number of forecasted periods, *s* – duration of seasonality (for example, number of months or quarters in a year).

For initialization of the additive method initial values of variable L_t, trend estimation b_t and seasonality estimation S_t are needed. To determine initial estimates we need at least one whole data season (that is, *s* data). Initialization of variable L_s is calculated with the formula:

$$L_s = \frac{1}{s}(Y_1 + Y_2 + \cdots + Y_s) \tag{5}$$

For trend initialization it is more suitable if we use two whole seasons (that is, 2*s* data):

$$b_s = \frac{1}{s}\left(\frac{Y_{s+1}-Y_1}{s} + \frac{Y_{s+2}-Y_2}{s} + \cdots + \frac{Y_{s+s}-Y_s}{s}\right) \tag{6}$$

Seasonal indices are calculated as differences between observed value and variable estimation L_s:

$$S_1 = Y_1 - L_s,\ S_2 = Y_2 - L_s, \ldots,\ S_s = Y_s - L_s \tag{7}$$

The biggest advantages of the method are low costs, fast calculation and simplicity. Furthermore, the method is proved to be (regarding costs and calculation itself) comparable with more complex methods (for example Box-Jenkins); in some cases the results gained with the Holt-Winters were even better than more complex methods ([6]).

2.2 Improved Holt-Winters' procedure

The only difference between the additive HW and the improved HW method is in the equation for the calculation of level (1); all other equations – for seasonality (S_t), trend (b_t), forecast (F_{t+m}) and method initialization – remain the same as with the additive method (2 – 7). The improved HW method for level is given with the equation:

$$L_t = \alpha Y_t - S_{t-s} + (1-\alpha)(L_{t-1} + b_{t-1}) \quad (8)$$

With the improved HW method, unlike the additive HW method, the smoothing parameter is only attributed the observed value Y_t, and not seasonality S_{t-s}.

The improved HW method also belongs among exponential smoothing techniques, which assigns exponentially decreasing weights as the observation get older. In other words, recent observations are given relatively more weight in forecasting than the older observations. With this method smoothing parameters also adopt values for interval [0, 1]. The higher value of the smoothing parameter is the lower in smoothing.

3 FORECAST CALCULATIONS AND RESULTS

For research we used quarterly data of overnight stays of domestic and foreign tourists in the Republic of Slovenia between the years 2000 and 2009. We acquired data from the Statistical Office of the Republic of Slovenia (SI-STAT Data Portal – Economy – Tourism). We deal with 6 intermittent time series' for chosen Slovenian communities or municipalities, but in this chapter we will present only one in detail. At the end of the chapter results are presented for all time series' and conclusions are noted.

Table 1: Overnight stays of domestic guests (Lovrenc na Pohorju).

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Y_t	0	7	2	4	0	36	2	0	0	0	6	2	0	0	40
t	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Y_t	2	0	17	20	0	0	3	12	0	0	2	1	1	0	4
t	31	32	33	34	35	36	37	38	39	40					
Y_t	6	0	0	4	3	12	0	2	7	0					

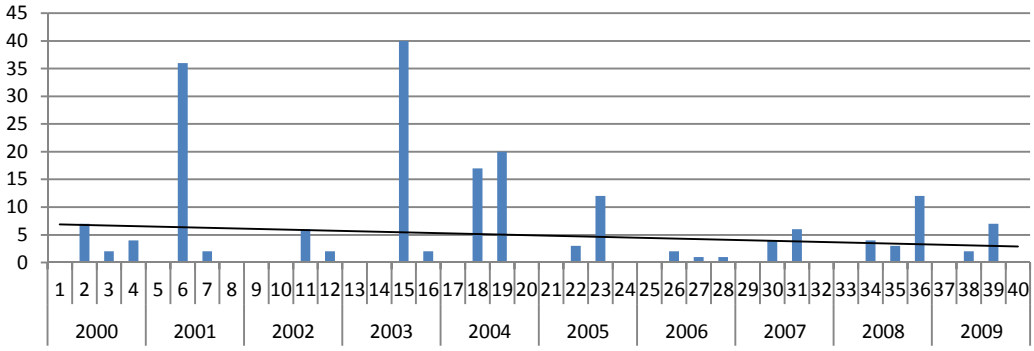


Figure 1: Overnight stays of domestic guests (Lovrenc na Pohorju).

The improved Holt-Winters method was also tested for examples taken from monography forecasting: methods and applications ([8]). The improved HW method was better than the additive HW method for all examples.

Table 1 and Figure 1 show the number of overnight stays of domestic guests in the community of Lovrenc na Pohorju between the years 2000 and 2009. It is obvious that this time series represents intermittent data (data with zeroes). From Figure 1 it is evident that comprehensive (random) fluctuations are present in the data.

We will present forecast calculations for Lovrenc na Pohorju. Forecasts are calculated by using the additive and improved HW method and the results are compared with each other. Regarding the Ferbar Tratar study [3] we also calculate forecasts with the methods where smoothing and initial parameters are estimated by minimising the mean square error (MSE).

In tables we use following notations: $s = 4$, $E^2 = (F_t - Y_t)^2$ and $MSE = \frac{1}{36}\sum_{t=5}^{40} E^2$. We use the first year (first four quarters) for initialization, which is afterwards used for calculation of estimates L_t, b_t and S_t. The following nine years (periods from 5 to 40) represent test series', used for minimization of *MSE*.

Table 2 and Table 3 show forecasted overnight stays of domestic guests in the community of Lovrenc na Pohorju between the years 2000 and 2009. In the first table forecasts are calculated with the additive HW method (AHW), where we estimated (only) smoothing parameters by minimising MSE. In the second table forecasts are calculated by using the AHW-init method, where smoothing and initial parameters are estimated by minimising MSE.

Table 2: Forecasts calculated with AHW method (Lovrenc na Pohorju).

Year	T	Y_t	L_t	b_t	S_t	F_t	E^2
2000	1	0			-3.25		
	2	7			3.75		
	3	2			-1.25		
	4	4	3.25	1.56	0.75		
2001	5	0	4.74	1.49	-3.37	1.56	2.44
	6	36	7.47	2.73	5.76	9.98	677.27
	7	3	9.87	2.40	-1.79	8.95	48.31
	8	0	11.65	1.78	-0.25	13.02	169.46
…	…	…	…	…	…	…	…
2009	37	0	0.79	1.02	-4.29	-4.00	16.04
	38	2	1.69	0.90	2.47	4.47	6.11
	39	7	2.72	1.03	2.00	4.39	6.83
	40	0	3.61	0.90	-1.04	2.70	7.30
						alpha =	**0.048**
						beta =	**1.000**
						gamma =	**0.081**
						MSE (5-40)=	**106.30**

Table 3: Forecasts calculated with AHW-init method (Lovrenc na Pohorju).

Year	T	Y_t	L_t	b_t	S_t	F_t	E^2
2000	1	0			-4.28		
	2	7			3.46		
	3	2			6.87		
	4	4	7.47	-0.19	-1.83		
2001	5	0	7.28	-0.19	-4.28	3.00	9.00
	6	36	7.09	-0.19	3.46	10.56	647.42
	7	3	6.91	-0.19	6.87	13.78	138.72
	8	0	6.72	-0.19	-1.83	4.89	23.90
…	…	…	…	…	…	…	…
2009	37	0	1.28	-0.19	-4.28	-3.00	9.00
	38	2	1.09	-0.19	3.46	4.56	6.53
	39	7	0.91	-0.19	6.87	7.78	0.60
	40	0	0.72	-0.19	-1.83	-1.11	1.23
						alpha =	**0.000**
						beta =	**1.000**
						gamma =	**0.000**
						MSE (5-40)=	**64.99**

The results obtained with the additive HW method (see Table 2 and 3) show that with the additional optimisation of initial values MSE is reduced by more than 38%.

In Table 4 forecasts are calculated with the improved HW method (IHW), where we estimated smoothing parameters by minimising MSE. In Table 5 forecasts are calculated by using the IHW-init method, where smoothing and initial parameters are estimated by minimising MSE.

Table 4: Forecasts calculated with IHW method (Lovrenc na Pohorju).

Year	t	Y_t	L_t	b_t	S_t	F_t	E^2
2000	1	0			-3.25		
	2	7			3.75		
	3	2			-1.25		
	4	4	3.25	1.56	0.75		
2001	5	0	7.57	2.62	-3.25	1.56	2.44
	6	36	9.09	2.20	3.75	13.94	486.81
	7	3	11.58	2.31	-1.25	10.03	64.52
	8	0	11.71	1.48	0.75	14.64	214.33
…	…	…	…	…	…	…	…
2009	37	0	2.40	1.41	-3.25	-0.44	0.20
	38	2	-0.13	-0.10	3.75	6.56	20.75
	39	7	1.77	0.66	-1.25	1.56	29.64
	40	0	1.43	0.28	0.75	3.56	12.64
						alpha =	**0.103**
						beta =	**0.383**
						gamma =	**0.000**
						MSE (5-40)=	**109.93**

Table 5: Forecasts calculated with IHW-init method (Lovrenc na Pohorju).

Year	T	Y_t	L_t	b_t	S_t	F_t	E^2
2000	1	0			-3.68		
	2	7			0.89		
	3	2			4.35		
	4	4	6.48	-1.42	-1.60		
2001	5	0	8.74	2.26	-3.68	1.37	1.88
	6	36	10.11	1.37	0.89	11.89	581.35
	7	3	7.13	-2.98	4.35	15.82	191.04
	8	0	5.76	-1.37	-1.60	2.55	6.49
…	…	…	…	…	…	…	…
2009	37	0	8.77	2.65	-3.68	-1.24	1.53
	38	2	10.53	1.76	0.89	3.34	1.79
	39	7	7.94	-2.59	4.35	6.79	0.04
	40	0	6.95	-0.99	-1.60	0.84	0.71
						alpha =	**0.000**
						beta =	**1.000**
						gamma =	**0.000**
						MSE (5-40)=	**57.37**

The results obtained with the improved HW method (see Table 4 and 5) show that with the additional optimisation of initial values, MSE is reduced by more than 47%. So, if we use the improved HW method with initial optimization instead of the additive HW method, MSE can be reduced by more than 46% (see also Table 6).

Table 6: Review of results for different community.

Community	MSE	Improvement (in %)				
		IHW/AHW	AHW-init/ AHW	IHW-init/ IHW	IHW-init/ AHW-init	IHW-init/ AHW
Komenda-AHW	5,699.55					
Komenda-AHW-init	3,605.81		36.74%			
Komenda-IHW	6,255.00	-8.88%				
Komenda-IHW-init	3,455.36			44.76%	4.17%	39.37%
Komenda-TUJ-AHW	42,076.54					
Komenda-TUJ-AHW-init	41,395.41		1.62%			
Komenda-TUJ-IHW	42,138.34	-0.15%				
Komenda-TUJ-IHW-init	40,730.58			3.34%	1.61%	3.20%
Logatec-TUJ-AHW	1,020,874.68					
Logatec-TUJ-AHW-init	987,429.14		3.28%			
Logatec-TUJ-IHW	943,746.05	7.56%				
Logatec-TUJ-IHW-init	823,322.34			12.76%	16.62%	19.35%
Lovrenc na Pohorju-AHW	106.30					
Lovrenc na Pohorju-AHW-init	64.99		38.86%			
Lovrenc na Pohorju-IHW	109.93	-3.30%				
Lovrenc na Pohorju-IHW-init	57.37			47.82%	11.73%	46.04%
Miren-Kostanjevica-AHW	11,731.89					
Miren-Kostanjevica-AHW-init	10,200.77		13.05%			
Miren-Kostanjevica-IHW	11,158.74	4.89%				
Miren-Kostanjevica-IHW-init	9,335.60			16.34%	8.48%	20.43%
Miren-Kostanjevica-TUJ-AHW	14,913.19					
Miren-Kostanjevica-TUJ-AHW-init	10,952.78		26.56%			
Miren-Kostanjevica-TUJ-IHW	13,645.80	8.50%				
Miren-Kostanjevica-TUJ-IHW-init	10,605.29			22.28%	3.17%	28.89%
					Average	**26.21%**

From the results for Lovrenc na Pohorju (see Table 6) we can see also that although the AHW method is better than the IHW method by 3.30%, the IHW-init is better than the AHW-init by 11.73%.

Table 6 shows the percentage of improvement of MSE, calculated by using the improved HW (init) method compared to the additive HW (init) method. We denote foreign guests with TUJ. It is obvious that if we treat the initial values for the level, trend and seasonal components as well as the three smoothing constants as decision variables, a considerable reduction in the MSE can be reached. The results show that on average with the additional optimisation of initial values the MSE is reduced on average by more than 24% for the improved HW method (and on average by more than 20% for the additive HW method). Finally, if we use the improved HW method with initial optimization instead of the additive HW method, MSE can be reduced on average by more than 26%.

4 CONCLUSION AND FURTHER RESEARCH

Demand forecasting is used throughout the world more often because of proper source management and the rising need to plan. Which method is going to be used depends on multiple factors: demanded comprehension of forecasts, further use of forecasts, and, of course, available data and price. One of the most commonly used forecasting techniques is exponential smoothing, which is relatively inexpensive, fast and simple and does not demand special software. There has been lot attention paid to the Holt-Winters forecasting procedure in recent years. Researchers discover new ways to improve the method itself, especially in dealing with more seasonal cycles and forecasting intervals.

The aim of this paper is to expose the problem of the forecasting of intermittent demand when data shows high variations. We propose an improved HW method and we show that a reduction in forecast error (MSE) can be achieved. From the results obtained for real data we prove that the proposed method is more efficient than the ordinary HW method, on average by more than 26%.

Because it is obvious from the given case that the improved HW method yields good results for data with significant fluctuations it would make sense to examine new methods for time series with multiplicative seasonal fluctuations and/or multiplicative trend. Because this exceeds the nature of this paper, this would be among our goals in the future.

References

[1] Bermúdez, J.D., Segura, J.V. and Vercher, E., 2006. A decision support system methodology for forecasting of time series based on soft computing. Computational Statistics & Data Analysis, 51, 177 – 191.

[2] Chatfield, C., 1978. The Holt-Winters Forecasting Procedure. Journal of the Royal Statistical Society, 27(3), 264–279.

[3] Ferbar Tratar, L., 2010. Joint optimisation of demand forecasting and stock control parameters. International Journal of Production Economics, 127(1), 173–179.

[4] Holt, C. C., 2004. Author's retrospective on 'Forecasting seasonals and trends by exponentially weighted moving averages'. Foresight: International Journal of Forecasting, 20(1), 11–13.

[5] Koehler, A.B., Snyder, R.D. and Ord, J.K., 2001. Forecasting models and prediction intervals for the multiplicative H–W method. Int. J. Forecasting, 17, 269–286.

[6] Makridakis, S. and Hibon, M., 1979. Accuracy of forecasting: An empirical investigation. Journal of the Royal Statistical Society, 142(2), 97–145.

[7] Makridakis, S. and Hibon, M., 2000. The M3-competition: results, conclusions and implications. Int. J. Forecasting, 16, 451–476.

[8] Makridakis, S., Wheelwright, S.C. and Hyndman, R.J., 1998. Forecasting : methods and applications, United States of America: John Wiley & Sons, Inc.

[9] Ord, J.K. (Ed.), 2001. Commentaries on the M3-competition. Int. J. Forecasting, 17, 537–584.

[10] Ord, J.K., Koehler, A.B. and Snyder, R.D., 1997. Estimation and prediction for a class of dynamic nonlinear statistical models. J. Am. Statist. Assoc., 92, 1621–1629.

ON A DECISION RULE SUPPORTED BY A FORECASTING STAGE BASED ON THE DECISION MAKER'S RISK AVERSION

Helena Gaspars-Wieloch
Poznan University of Economics
Al. Niepodleglosci 10, 61-875 Poznan, Poland
helena.gaspars@ue.poznan.pl

Abstract: The paper contains a description of a new approach (called the SF+AS method) that can be applied in the decision making under uncertainty when pure optimal strategies are sought-after. This procedure takes into consideration the level of decision maker's risk aversion and consists of two stages: the true scenario's forecasting (on the basis of the DM's preferences) and the appropriate alternative's selection by taking into account the payoffs of the true scenario appointed or the most probable scenarios.

Keywords: decision making under uncertainty, optimal pure strategy, true scenario's forecasting, decision maker's risk aversion, coefficients of pessimism and optimism

1 INTRODUCTION

The uncertainty is a consequence of the fact that we are not able to anticipate the future effectively. One may just forecast various phenomena and events, but in many cases it is extremely difficult to estimate the exact value of particular parameters (temperature, demand for a product, product prices etc.). When many future factors are not deterministic at the time of the decision, the decision maker has to choose the appropriate alternative (decision, strategy) on the basis of some scenarios (states of nature) predicted by experts, him- or herself. Let us add that the probability of these scenarios may be known (decision making under risk – DMUR) or not (decision making under uncertainty – DMUU), [4], [9], [19]. These two categories (risk and uncertainty) were formally integrated in economic theory by J. von Neumann and O. Morgenstern [16]. In this contribution we will focus on the second case which seems to be more frequent in realistic decision problems. The result of the choice made by the decision maker under uncertainty depends on two factors: which decision will be selected and which scenario will occur in the future. The consequence of any alternative is determined not just by the alternative itself but also by an external factor which is beyond the control of the decision maker. The DMUU may be presented with the aid of a profits' or payoffs' matrix (Tab. 1) where m is the number of mutually exclusive scenarios (let us denote them by $S_1, S_2, \ldots, S_m$), n signifies the number of decisions ($D_1, D_2, \ldots, D_n$) and a_{ij} is the profit connected with the scenario S_i and the alternative D_j. The goal of the DM consists in selecting this decision which maximizes the profit.

Table 1. Payoffs' matrix / decision table (general case)

Scenarios and Decisions	$\boldsymbol{D_1}$	$\boldsymbol{D_j}$	$\boldsymbol{D_n}$
$\boldsymbol{S_1}$	a_{11}	a_{1j}	a_{1n}
$\boldsymbol{S_i}$	a_{i1}	a_{ij}	a_{in}
$\boldsymbol{S_m}$	a_{m1}	a_{mj}	a_{mn}

Notice that sometimes the distribution of payoffs connected with particular alternatives is not discrete and then the profits for each decision D_j belong to an interval $[w_j, m_j]$, [7], [12]. In this contribution we will consider the scenarios' approach for DMUU which is characterized by a lower degree of uncertainty than the interval approach because only several values from this range are probable.

In the uncertainty case the decision maker may search an optimal pure strategy or an optimal mixed strategy. A pure strategy, in contradiction to a mixed strategy, is a solution assuming that the decision maker chooses and completely executes one and only one alternative. Meanwhile the mixed strategy allows the decision maker to select and perform a weighted combination of several accessible alternatives. The whole paper will focus on optimal pure strategy's searching.

We will also assume that each alternative is characterized by one criterion's value or by one synthetic aggregated value denoting the overall realization of all significant criteria.

The literature offers many procedures applied in DMUU, such as the Wald's criterion [21], [22], the maximax criterion described for example in [17], the Hurwicz's criterion [10], [11], the Savage's criterion [20], the Bayes' (Laplace's) criterion (see e.g. [19]), which for convenience, may be called "the classical decision rules" (CD rules), and many diverse extensions or hybrids of these methods (see e.g. [1], [2], [3], [6], [7], [8], [14], [15], [18]), which may be named "the extended decisions rules" (ED rules).

In all of them a measure precisely defined is computed for each alternative, which allows the decision-maker to choose in the last step the decision with the most preferable value of the applied index.

In the majority of existing methods the alternative is selected on the basis of the level of risk aversion declared by the decision maker. When he or she is adventurous, it is recommended to look at the highest payoffs assigned to each decision and to choose the alternative according to the maximax rule. When the DM represents a risk-averse behavior, it is suggested to compare the lowest profits (or the highest regrets) and to follow the Wald's rule (or the Savage's rule). Finally, when we deal with a moderate DM, the Hurwicz's approach is applied since it enables to assign a coefficient of pessimism (α) to the worst value and a coefficient of optimism (β=1-α) to the best value connected with particular strategies in order to obtain a weighted average for each alternative. It is worth emphasizing that usually the highest and the lowest profits of the decisions considered come from different states of nature. That means that a given scenario may be very optimistic from the point of view of one decision and simultaneously extremely bad with respect to an other alternative (see Tab. 2, scenarios S_2 and S_3). Hence, according to the nature of the existing methods the scenarios are very seldom considered as totally pessimistic or totally optimistic.

Table 2. Payoffs' matrix / decision table (example)

Scenarios and Decisions	$\boldsymbol{D_1}$	$\boldsymbol{D_2}$	$\boldsymbol{D_3}$
$\boldsymbol{S_1}$	5	4	3 (min)
$\boldsymbol{S_2}$	10 (max)	1 (min)	7 (max)
$\boldsymbol{S_3}$	0 (min)	8 (max)	5

Let us think over the following new question – is it possible to forecast the true state of nature on the basis of the decision maker's risk-aversion and to select the appropriate alternative taking into account not the whole payoffs' matrix (i.e. the whole set of possible scenarios) but only the scenario (or scenarios) meeting the DM's preferences?

The remainder of the paper is organized as follows. In Section 2 the author suggests and describes with the aid of a case study a new method enabling to forecast the true state of nature depending on the decision maker's attitude towards risk and to select in the second step the appropriate alternative. In this section the Reader will also find a formal presentation of the procedure. Conclusions are gathered in Section 3.

2 The SF+AS method – description and illustration

The method presented below (called the SF+AS method, *scenario's forecasting and alternative's selection*) appeals to a totally different concept than other procedures do. This time a given scenario will be treated as extremely pessimistic, moderately pessimistic, moderate, moderately optimistic or radically optimistic independently on the alternative.

Hence, the heart of the problem consists in applying a suitable tool enabling to determine correctly the status of each state of nature. This is the first step of the SF+AS method. Possible approaches may by diverse – here we will use the concept of dominance.

Let us analyze the following example. Table 1 presents a payoffs' matrix. Profits are given in million Euros and concern a period of one year. There are four decision makers (DM1, DM2, DM3, DM4). They dispose of four possible strategies (projects P1, P2, P3, P4) and they are aware of the fact that one out of four states of nature (S1, S2, S3, S4) will occur in the future, but they have no information about the likelihoods of particular scenarios. Each decision maker has a different attitude towards risk. The first one is a pessimist, his coefficient of optimism equals $\beta_1 = 0.1$, the second one is a moderate pessimist ($\beta_2 = 0.4$), the third one is a moderate optimist ($\beta_3 = 0.65$) and the last one is a radical optimist ($\beta_4 = 0.95$). Thus, each decision maker has a totally unlike opinion about the true state of nature, i.e. the scenario that will really happen. Notice that according to the concept of Pareto optimality [5] not a scenario enumerated in Table 3 is dominated by other scenarios (see Tab. 4–6). All of them are Pareto optimal, because each column of the Table 6 (representing the multicriteria comparison, i.e. the product of all orders) contains only zeros.

Table 3: Payoffs's matrix – Example.

Scenarios	***Alternatives***			
	P1	***P2***	***P3***	***P4***
S1	1	2	7	7
S2	5	4	1	6
S3	6	6	8	5
S4	10	3	9	5

Table 4: The order Q_1 (according to P1) and Q_2 (according to P2).

The order Q_1					***The order Q_2***				
Scenarios	***Scenarios***				***Scenarios***	***Scenarios***			
	S1	***S2***	***S3***	***S4***		***S1***	***S2***	***S3***	***S4***
S1	0	0	0	0	***S1***	0	0	0	0
S2	1	0	0	0	***S2***	1	0	0	1
S3	1	1	0	0	***S3***	1	1	0	1
S4	1	1	1	0	***S4***	1	0	0	0
(S_i is better than S_k, if according to the alternative P_1, $a_{i1} > a_{k1}$. $\forall S_i, S_k \in S : S_i Q_1 S_k \Leftrightarrow a_{i1} > a_{k1}$)					S_i is better than S_k, if according to the alternative P_2, $a_{i2} > a_{k2}$. $\forall S_i, S_k \in S : S_i Q_2 S_k \Leftrightarrow a_{i2} > a_{k2}$)				

But even if all scenarios are Pareto optimal, one can observe that the states S1 and S2 usually offer worse results than states S3 and S4 do. Therefore, we detect a possibility to work out a ranking of the states considered. Theoretically, there are many procedures allowing to generate this ranking. One may use the criterion of the sum of payoffs for each scenario, or the criterion of the sum of regrets (see the Savage's rule), or the criterion of the sum of utility functions [13]. Here we will apply the sum of "dominance cases" within each

alternative (Tab. 7). The scenario S1 is 4 times better than other events (for P3: $S1 \succ S2$, for P4: $S1 \succ S2$, $S1 \succ S3$ and $S1 \succ S4$). S2 is five times better than other scenarios (for P1: $S2 \succ S1$, for P2: $S2 \succ S1$, $S2 \succ S4$, for P4: $S2 \succ S3$, $S2 \succ S4$), S3 – seven times (for P1: $S3 \succ S1$, $S3 \succ S2$, for P2: $S3 \succ S1$, $S3 \succ S2$, $S3 \succ S4$, for P3: $S3 \succ S1$, $S3 \succ S2$) and S4 – seven times as well (for P1: $S4 \succ S1$, $S4 \succ S2$, $S4 \succ S3$, for P2: $S4 \succ S1$, for P3: $S4 \succ S1$, $S4 \succ S2$, $S4 \succ S3$).

Table 5: The order Q_3 (according to P3) and Q_4 (according to P4).

The order Q_3					***The order Q_4***				
Scenarios	***Scenarios***				***Scenarios***	***Scenarios***			
	S1	***S2***	***S3***	***S4***		***S1***	***S2***	***S3***	***S4***
S1	0	1	0	0	***S1***	0	1	1	1
S2	0	0	0	0	***S2***	0	0	1	1
S3	1	1	0	0	***S3***	0	0	0	0
S4	1	1	1	0	***S4***	0	0	0	0
(S_i is better than S_k, if according to the alternative P_3, $a_{i3} > a_{k3}$. $\forall S_i, S_k \in S : S_i Q_3 S_k \Leftrightarrow a_{i3} > a_{k3}$)					(S_i is better than S_k, if according to the alternative P_4, $a_{i4} > a_{k4}$. $\forall S_i, S_k \in S : S_i Q_4 S_k \Leftrightarrow a_{i4} > a_{k4}$)				

Table 6: Multicriteria comparison $W[Q_1, Q_2, Q_3, Q_4]$, i.e. the product of all orders

Scenarios	***Scenarios***			
	S1	***S2***	***S3***	***S4***
S1	0	0	0	0
S2	0	0	0	0
S3	0	0	0	0
S4	0	0	0	0

(S_i dominates S_k, if for all decision P_j, S_i is better than S_k)

Table 7: Payoffs' matrix and sum of "dominance cases" – Example.

Scenarios	***Alternatives***				***Sum of "dominance cases" (d_i)***	***Interval for β***
	P1	***P2***	***P3***	***P4***		
S1	1	2	7	7	4	[0.0, 0.25]
S2	5	4	1	6	5	]0.25, 0.50]
S3	6	6	8	5	7	]0.75,1.0]
S4	10	3	9	5	7	]0.75,1.0]

Now, having a ranking of states of nature (I place: S3 and S4, II place: S2, III place: S1), one may attempt to assign a suitable interval of values of the coefficient of optimism to each scenario. Obviously, higher the sum of "dominance cases" for a given scenario is, more optimistic this scenario should be. The width of the range (w) of each state of nature may be for instance defined in the following way:

$$w = \max\left\{\frac{1}{m}, \frac{1}{d_{\max} - d_{\min} + 1}\right\} \tag{1}$$

where m is the number of scenarios, d_{max} and d_{min} are the highest and the lowest number of "dominance cases" respectively. Such an approach allows to fit the width of the intervals to the overall number of scenarios and to the difference between the highest and the lowest number of "dominance cases". The extreme values (b_i and t_i) of a given interval, i.e. its endpoints, may be computed according to the Equations (2)–(4):

$$b_i = \max\left\{b \mid \left\{(b \mid w) \wedge \left(b \le \frac{d_i - d_{\min}}{d_{\max} - d_{\min}}\right) \wedge \left(b \in [0;1-w]\right)\right\}\right\} \quad (2)$$

$$t_i = \min\left\{t \mid \left\{(t \mid w) \wedge \left(t \ge \frac{d_i - d_{\min}}{d_{\max} - d_{\min}}\right) \wedge \left(t \in [w;1]\right)\right\}\right\} \quad (3)$$

$$t_i = b_i + w \quad (4)$$

where d_i is the number of "dominance cases" within the scenario i.
Additionally, let us assume that, apart from the interval(s) for the lowest number of "dominance cases", the intervals are left-open. One can observe two facts on the basis of the ranges set in Table 7:

- more than one state of nature may contain the same interval (see S3 and S4),
- the intervals do not have to cover the whole range of possible values for the parameter β (values bigger than 0.5 and not exceeding 0.75 do not occur).

According to the risk aversion declared by the decision makers, the state S1 may be the true state in DM1's opinion and S2 my be selected by DM2 as the true scenario. There are two states (S3 and S4) which correspond to the DM4's level of optimism and there is no scenario which can be directly assigned to the DM3's preferences. Therefore, it is recommended, in the DM4's case, to use for each decision an arithmetic average of the payoffs related to both states S3 and S4 (see Equation 5). On the other hand for the DM3 it is suggested to calculate a weighted average of the "nearest" scenarios, i.e. S3, S4 and S2 following the Equation (6):

$$A_{j,k}^{arit} = \frac{1}{p_k}\sum_{i=1}^{p_k} a_{ij} \quad (5)$$

$$A_{j,k}^{weig(e,f)} = \frac{\beta_k - t_e}{b_f - t_e} \cdot a_{fj} + \frac{b_f - \beta_k}{b_f - t_e} \cdot a_{ej} \quad (6)$$

where p_k is the number of scenarios suitable for the decision maker k, parameters e and f denote the scenarios which values of β are a little bit lower and a little bit higher than the parameter β_k. Parameters t_e and b_f signify the right endpoint of the interval e and the left endpoint of the interval f respectively. Finally a_{ej} and a_{fj} constitute the payoffs connected with the decision j and the scenarios e and f. Notice that if there are more than one state e or f (because of the occurrence of the same interval), then instead of a_{ej} or a_{fj} an arithmetic average of suitable payoffs is taken into consideration (see Equation 5).

Now, we can perform the second step of the SF+AS method which consists in selecting the appropriate alternative:

a) The DM1 makes his choice on the basis of the payoffs that can occur if the scenario S1 takes place: 1, 2, 7, 7. Hence he or she should select the project P3 or P4.
b) The DM2 ought to make the decision assuming that the payoffs related to the state S2 will occur (5, 4, 1 or 6). Thus, he or she should choose the project P4.
c) The DM4 disposes of four arithmetic averages calculated by means of the Equation (5) and the data coming from the scenarios S3 and S4:

$$A_{1,4}^{arit} = \frac{1}{2}(6+10) = 8 \quad A_{2,4}^{arit} = \frac{1}{2}(6+3) = 4.5 \quad A_{3,4}^{arit} = \frac{1}{2}(8+9) = 8.5 \quad A_{4,4}^{arit} = \frac{1}{2}(5+5) = 5$$

The results indicate that DM4 ought to be interested in the project P3.
d) The DM3 has to analyze the figures obtained after using the Equation (6) and the data concerning the scenarios S2, S3 and S4:

$$A_{1,3}^{weig(2,3+4)} = \frac{0.65-0.5}{0.75-0.5}\cdot 8 + \frac{0.75-0.65}{0.75-0.5}\cdot 5 = 6.8$$
$$A_{2,3}^{weig(2,3+4)} = \frac{0.65-0.5}{0.75-0.5}\cdot 4.5 + \frac{0.75-0.65}{0.75-0.5}\cdot 4 = 4.3$$
$$A_{3,3}^{weig(2,3+4)} = \frac{0.65-0.5}{0.75-0.5}\cdot 8.5 + \frac{0.75-0.65}{0.75-0.5}\cdot 1 = 5.5$$
$$A_{4,3}^{weig(2,3+4)} = \frac{0.65-0.5}{0.75-0.5}\cdot 5 + \frac{0.75-0.65}{0.75-0.5}\cdot 6 = 5.4$$

Hence, it will be recommended to select the project P1.

Notice that if the original Hurwicz's rule was used for the levels of β aforementioned, the following projects would be suggested: P4 for DM1 and DM2, P1 for DM3 and DM4.

After the illustration of the SF+AS method let us enumerate the steps of this procedure in the general case:

1) Calculate the sum of the "dominance cases" for each scenario (Equations 7 and 8).

$$d_{ij} = m - \max\{p(a_{ij})\} \qquad i = 1,...,m;\ j = 1,...,n \qquad (7)$$

$$d_i = \sum_{j=1}^{n} d_{ij} \qquad (8)$$

where d_{ij} denote the number of payoffs related to the alternative j which are worse than the payoff a_{ij}. The symbol m still signifies the number of scenarios and $p(a_{ij})$ is the position of the payoff a_{ij} in the non-increasing sequence of all profits connected with the decision j (when a_{ij} has the same value than other payoffs concerning a given alternative, then it is recommended to choose the farthest position of this payoff in the sequence – see Equation 7). d_i is the total number of "dominance cases" related to the state i.

2) Assign an interval for the coefficient of optimism to each scenario (Equations 1–4).
3) Find the set of values on the basis of which the DM will make the final choice:
   a) If the parameter β belongs to the interval assigned to exactly one scenario, then this set contains all payoffs connected with this state of nature.
   b) If β belongs to the interval assigned to more than one scenario, generate the set using the Equation 5 for each alternative.
   c) If β does not belong to any interval assigned to scenarios, compute the set using the Equation 6 for each alternative.
4) Choose the alternative which has the highest value in the set found in the step 3.

3 Conclusions

The new approach presented in the paper and called the SF+AS method can be applied in the decision making under uncertainty when pure optimal alternatives are sought-after. The procedure is designed for decision makers who are able to declare their coefficient of optimism (pessimism). In contradiction to existing decision rules this method contains an additional stage that precedes the searching of the optimal alternative and consists in forecasting the true state of nature on the basis of the DM's risk aversion. Such an approach signifies that the decision maker makes his or her choice by taking into consideration only the payoffs of the forecasted true scenario or the most probable (in his or her opinion) scenarios appointed in the first stage, and not the whole payoffs' matrix. Hence, in this procedure the status (pessimistic, moderate or optimistic) of a given state of nature does not vary depending on the alternative, but is fixed for all decisions.

References

[1] Basili, M., 2006. A rational decision rule with extreme events. Risk Analysis 26, pp.1721–1728.

[2] Basili, M., Chateauneuf, A., Fontini, F., 2008. Precautionary principle as a rule of choice with optimism on windfall gains and pessimism on catastrophic losses. Ecological Economics 67, pp.485–491.

[3] Basili, M., Zappia, C., 2010. Ambiguity and uncertainty in Ellsberg and Shackle. Cambridge Journal of Economics 34(3), pp. 449–474.

[4] Chronopoulos, M., De Reyck, B., Siddiqui, A., 2011. Optimal investment under operational flexibility, risk aversion, and uncertainty. European Journal of Operations Research 213, pp.221–237.

[5] Ehrgott, M., 2005. Multicriteria Optimization. Springer Berlin, Heidelberg.

[6] Ellsberg, D., 2001. Risk, Ambiguity and Decision. Garland Publishing, New York, USA.

[7] Gaspars-Wieloch, H., 2013. Modifications of the Hurwicz's decision rules. Central European Journal of Operations Research. DOI 10.1007/s10100-013-0302-y

[8] Ghirardato, P., Maccheroni, F., Marinacci, M., 2004. Differentiating ambiguity and ambiguity attitude. Journal of Economic Theory 118, pp.133–173.

[9] Groenewald, M.E., Pretorius, P.D., 2011. Comparison of Decision making under Uncertainty Investment Strategies with the Money Market. Journal of Financial Studies and Research, doi: 10.5171/2011.373376

[10] Hurwicz, L., 1952. A criterion for decision making under uncertainty. Technical Report 355. Cowles Commission.

[11] Hurwicz, L., 1951. The Generalized Bayes Minimax Principle: A Criterion for Decision Making Under Uncertainty. Cowles Commission. Discussion Paper Statistics 335.

[12] Huynh, V., Hu, C., Nakamori, Y., Kreinovich, V., 2009. On decision making under interval uncertainty: A new justification of Hurwicz optimism-pessimism approach and its use in group decision making. In: Proceedings of the 39th International Symposium on Multiple-Valued Logic ISMV/L/2009, Naha, Okinava, Japan.

[13] Kukula, K., 2000. Metoda unitaryzacji zerowanej. PWN, Warszawa.

[14] Marinacci, M., 2002. Probabilistic sophistication and multiple priors. Econometrica 70, pp.755–764.

[15] Nakamura, K., 1986. Preference relations on a set of fuzzy utilities as a basis for decision making. Fuzzy sets and systems 20, pp.147–162.

[16] Neumann, J., Morgenstern, O., 1944. Theory of Games and Economic Behavior. Princeton University Press.

[17] Pazek, K., Rozman, C., 2009. Decision making under conditions of uncertainty in agriculture: a case study of oil crops. Poljoprivreda (Osijek) 15(1), pp.45–50.

[18] Piasecki, K., 1990. Decyzje i wiarygodne prognozy, Akademia Ekonomiczna w Poznaniu.

[19] Render, B., Stair, R.M., Hanna, M.E., 2006. Quantitative Analysis for Management. Pearson Prentice Hall. Upper Saddle River. New Jersey

[20] Savage, L.J., 1961. The Foundations of Statistics reconsidered. In: Kyburg, H.E., Smokler H.E. (eds), Studies in Subjective Probability. New York: Wiley, pp.173–188.

[21] Wald, A., 1950. Basic ideas of a general theory of statistical decisions rules. In: Wald, A. (ed) Selected papers in Statistics and Probability. New York: McGraw-Hill, pp.656–668.

[22] Wald, A., 1950. Statistical decision functions. Wiley. New York.

HOW TO USE LINEAR PROGRAMMING FOR INFORMATION SYSTEM PERFORMANCES OPTIMIZATION

Marko Hell
University of Split, Faculty of Economics
Cvite Fiskovića 5, 21000 Split, Croatia
marko.hell@efst.hr

Abstract: The Balanced Scorecard (BSC) is a popular concept for performance measurement. The Linear programming (LP) is a mathematical technique for optimization of linear objective functions. The question is: "How to use LP for information system (IS) performances optimization (PO)?" Answer to this question is contained in this paper. The first step is a formalization of the IS and business performances relationship structure. The structure is designed in accordance withthe BSC concept. That will provide the application of LP for IS performances optimization.

Key words: linear optimization, information systems, performance management, balanced scorecard

1 INTRODUCTION

Strategic performance management is a relatively young field of managerial science. It deals with problems of effective strategy implementation and validation of its contribution to organization's success [4]. Dynamic environment of organisation changes in the process of implementation of the planned activities. Therefore, the ability to continuously adjust the strategic plan with the new conditions represents the prerequisite for the successful accomplishment of strategic objectives. Implementation of the strategic plan is usually based on the accomplishment of the planned activities. Each activity contributes to the accomplishment of a certain strategic objective of the organisation. Accomplishment of strategic objectives is measured by performances. By carrying out the activities, the organisation should, within a period of time in future, accomplish the transformation from the current value of performance (As is) to the future value of performance (To be). IT architecture is often assumed to follow the business strategy, to align IT with the business's strategic objectives [10]. In this context managers also need to estimate impact of new information technology (IT).

Balanced Scorecard methodology (BSC) is a popular concept of the balanced view of the organisation's performance [9]. It was originally developed by Kaplan and Norton and it aimed at enabling organisations to define their development strategies as well as to observe the success of the strategies' implementation [10]. Development of the BSC is based on the empirical experience of the large number of organisations in order to avoid disadvantages of measuring effectiveness only by financial indicators. Its implementation enables the process of strategic management not only to plan and organise but also to control the level of accomplishment of strategic objectives. In Strategic planning of information system methodology [1,2,3,5], BSC is suggested as a very powerful tool for measuring impact of new information technology on business performances [4]. The basic idea is included in the BSC concept for information system [8]. The paper provides guidelines for measuring the IS impact on the achievement of organization's business goals.

The proposed "BSC for IS" concept is similar to the classical BSC concept. The basic ideas for reshaping the BSC perspectives stem from the following [8]:

- The IS project works in favour of not just individual clients, but also of both the end user and the organization as a whole;
- The IS department should be perceived as internal rather than external service provider.

Accordingly, the perspectives for measuring the IS performances are the following:

- customer (end user) orientation ;
- business values;
- internal processes;

- readiness for the future.

The primary strategic objectives of the IS are divided into two types: objectives related to efficiency and objectives related to effectiveness. The efficiency-oriented objectives pertain to the processes. It is therefore necessary to consider them through the perspective of internal processes. The effectiveness-oriented objectives pertain to the users and therefore are analysed through the perspective of orientation towards the users and the perspective of business values. Recognizing the need for innovations and learning, the perspective of readiness for the future encompasses technologies and business opportunities, and challenges that will ensure stability of growth and development.

In this context, the paper we will show the original procedure used to enhance the BSC methodology in planning the optimal targets of IS performances value in order to maximise the organization's effectiveness.

2 FORMULATION OF IS PERFORMANCES RELATIONSHIP STRUCTURE

According to the defined mission, it is necessary to define the future course of development of the organisation, i.e. the vision of organization. This would mean that organisatio's vision sets the general guidelines which are to be followed in order to accomplish a higher quality mission. Implementation of the vision is formalized through development strategies of the organisation.

A badly formalised vision in the form of announcements may be transformed into descriptively and quantitatively determined set objectives[1](SO). For every SO it is necessary to determine strategy and activities the results of which are measured as level of accomplished of derived objectives (DO)[2]. This procedure requires forming judgments and strategies [7]. Activities are derived from the strategy and can be seen as the expansion of a descriptive part of the DO[3]. Numerical semantic elements of every objective in the context of this paper are observed as performance, i.e. measure of objective.

In this way, cause-consequence structure of impact between performances depends on the cause-consequence structure of the strategic objectives. Namely, it is to be expected that there are influences among certain activities in the real system. It means that undertaking one of IS development activities can influence the effect of the another business activity. Since every activity is undertaken with a precisely set objective, it can be concluded that the structure of all objectives is the same as the structure of all activities. A chain of interconnected objectives in the context of this paper are called the causes-consequences chain (CCC). Based on previous, it means that it is possible to establish a direct relationships among IS and all other business performances of an organization.

Possibility of processing a large number of relationships between performances demands using the table [7]. In this way, every row expresses the performance which makes a direct influence on performances in the column. Hence, every column expresses the performance on which a direct influence is made by performance in row. Depending on the existence of direct relationship among performaces, the elements in the table gain the values 1 or 0. If there is a direct influence, the value 1 is entered. If not, the value 0 is entered. Every cell in the table is supposed to be filled in this way.

According to the previous explanation, the set objectives l is determined and the derived objectives k is derived. The final set of performances can be presented by the following expression (1)

$$\tilde{C} = \{C_1, C_2, ..., C_n\}, \qquad n=k+l \tag{1}$$

A direct influence among performances may be presented in the strict form of the square matrix (2).

[1] Set strategic goals are derived from the vision which is why they are named set strategic goals.

[2] The name derived strategic goal results from the fact that they are derived from the set strategic goal. Detailed description of the method is available in [7]

[3] This results from the fact that every activity is undertaken with the particular goal (1:1). Unlike activities, more strategies can be accomplished through one activity (m:1).

$$SEP = \begin{bmatrix} 0 & c_{12} & c_{13} & \cdots & c_{1n} \\ c_{21} & 0 & c_{23} & \cdots & c_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ c_{n1} & c_{n2} & c_{n3} & \cdots & 0 \end{bmatrix}. \quad (2)$$

The order of the square matrix SEP (Structure of Enterprise Performances) represents a total number of the performances including IS performances. According to the previous explanation elements of the SEP matrix are $c_{ij} \in \{0,1\}$. Index of SEP elements indicates index of performances of observed direct relationship. In this way, formal prerequisites for optimization performances value are met.

3 LIMITATION OF PERFORMANCES VALUE INCREASES

The classic BSC concept, in the phase of planning the effects, includes the implementation of determined activities. However, in the real system, implementing the activities can depend on various limitations. That is why it is necessary to adjust an expected level of accomplishment of objectives to the potential limitations. The concept of the strategic management shown in the paper emphasises two types of limitations.

The classic type of limitation to accomplish the expected level of accomplishment of objectives is availability of resources for implementing activities by which these objectives can be achieved. But allocation of resource depends on structure of performances relationships. Based on the previous formalization, this suggests that we need to impose restrictions caused by the structure of performances relationships. It is a consequence of influences that occur between objectives. Achievement of the lower positioned IS objectives is a prerequisite for accomplishingthe effect of activities which are carried out as a purpose of their superior business objectives (figure 1)[4].

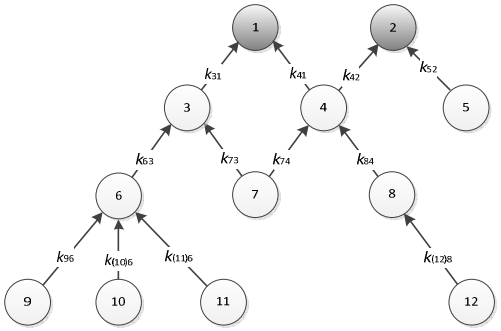


Figure 1: An example of a cause-effect performances relationship
(SOs are marked with dark colour; n=12; i=6)

Coefficients of influence between performances (of objectives) have been derived and defined by the expressions (3).

$$k_{ij} = \begin{cases} \dfrac{c_{ij}}{\sum_{i=1}^{n} c_{ij}} & if \quad \sum_{i=1}^{n} c_{ij} \neq 0 \\ 0 & if \quad \sum_{i=1}^{n} c_{ij} = 0 \end{cases}. \quad (3)$$

[4] It means: if we want positive change of accomplishment of objective 3, firstly we need to have positive change of accomplishment of objectives 9, 10, 11, 6 and 7.

Let the n number of all performances and i number of performances which are at the beginning of CCCs. This means that there is $n-i$ of calculated performances. This results in the existence of $n-i$ limitations, which can be defined by the system of inequality (4).

$$\begin{cases} 0 + k_{21}\cdot m_R C_2 + \cdots k_{i1}\cdot m_R C_{n-i} + \cdots + k_{n1}\cdot m_R C_n \geq 1\cdot m_R C_1 \\ 0 + \cdots k_{i2}\cdot m_R C_{n-i} + \cdots + k_{n2}\cdot m_R C_n \geq 1\cdot m_R C_2 \\ \vdots \qquad\qquad \vdots \qquad \vdots \\ 0 + \cdots + k_{n2}\cdot m_R C_n \geq 1\cdot m_R C_{n-i} \end{cases} . \tag{4}$$

Coefficients k_{ij} for $i \neq j$ have been calculated from equation (3). The system of inequalities (4) includes $n-i$ inequalities in which each inequality indicates one limitation to the calculated performance.

Defining objectives and their performances and determining “As is" and "To be" values sets the range of change of a performance. The defined range of performance enables the calculation of the relative change of performance. The relative change of performance of the objective C_j during the observed period of the strategic cycles is calculated according to the expression (5)

$$m_R C_j = \frac{mC_j - mC_j(0)}{mC_j(T) - mC_j(0)}, \qquad j=1,...,n \tag{5}$$

where: n stands for the number of the determined objectives, $m_R C_j$ stands for the relative value of performances of the objective C_j, $mC_j(0)$ is an initial value of the performance of the objective C_j, mC_j is a current value of the performance of the objective C_j at the end of the observed period and $mC_j(T)$ is the expected value of the performance of the objective C_j at the end of the strategic cycles with the time T. The relative change calculated in this way can occur in the segment [0,1].

Calculation done by using the relative value in the given concept imposes a prerequisite of inequality and maximum value of a performance for all strategic objectives i.e.

$$0 \leq m_R C_i \leq 1, \qquad \forall\ i=1,...,n. \tag{6}$$

In this way all the limitations have been included which enables the final determining of the optimal strategy (using an elaborated procedure).

The nature of each IS development activity indicates specific resources for its implementation. By determining the accompanied IS DOs, the measures and the range of changes are clerly defined. This is accomplished by implementation of the planned IS development and other business activities. This means that all necessary resources can be generated from activities and performances of DOs.

$$\begin{cases} r_{11}\cdot m_R C_1 + \cdots + r_{k1}\cdot m_R C_k \leq R_1 \\ r_{12}\cdot m_R C_1 \qquad\qquad r_{k2}\cdot m_R C_k \leq R_2 \\ \vdots \qquad\qquad\qquad \vdots \qquad \vdots \\ r_{1r}\cdot m_R C_1 + \qquad + r_{kr}\cdot m_R C_k \leq R_r \end{cases} . \tag{7}$$

Values r_{ij} indicate the required allocation of resources for 100% of accomplishment of k DOs which require the implementation of the observed activitiy. Every inequality in expression (7) indicates limitation caused by the availability of one particular resource (R_i). This defines and mathematically formalises the set of limitations over the total level of accomplishment of DO based on the availability of resources.

4 OPTIMISATION OF IS PERFORMANCES VALUE

The basic feature of the approach to the development of this model is that an organisation should be observed as a whole system. It means that the accomplishment of strategic objectives should not be

observed partially but in the context of accomplishment of set strategic objectives. Such an approach indicates that the maximum accomplishment of all DO is not always optimal. Determining the optimal level of accomplishment of strategic objectives represents a problem which can be solved by using linear programming.

Problem of linear programming can generally be the problem of maximum or the problem of minimum. The nature of the analysed problem belongs to the problem of maximum of the linear programming. Namely, taking into consideration the limitations caused by available resources and the determined structure of performances relationships, it is necessary to find the optimal level of accomplishment of derived strategic objectives in order to maximise the value of set SO performances.

l SO performances is determined. A function which requires a set maximum i.e. the function of an objective is defined by the expression (8)

$$Max\left(\frac{1}{l}\cdot\sum_{j=1}^{l} m_j C_j\right). \tag{8}$$

In this way the following elements have been determined:

- functions of performances of SO defined by the expression (8),
- limitations caused by the performances relationship structure defined by the expression (4),
- limitations caused by availability of IS resources defined by the expression (7),
- prerequisite of nonnegativity and maximum value of performaces defined by the expression (6),

The observed problem includes all required elements for implementation of the linear programming in order to define the optimal strategy. The result gained indicates the optimal values of DOs performances for maximum of value of performance of SOs. Sum of product of performances optimised values and r_i indicate total of *i* resource needed.

5 CONCLUSION

The paper leads us to conclude that the application of the linear programming within the classic concept of the BSC enables the optimisation of IS performances. Periodical repetition of the suggested procedure of the optimisation in the set discreet moments enhances the current method of management by implementing the strategy. The original algorithm shown in the paper and based on the matrix calculation by using the IT, enhances solving the economic problem of optimisation of IS performances due to the maximisation of accomplishment of the set strategic objectives.

References

[1] Brumec, J. 1996. A contribution to IS general taxonomy In: proceedings of the 7th International Conference Information Systems 96; Varaždin, Croatia, pp. 95.–105.

[2] Brumec, J. 1998 Strategic Planning of IS. Journal of Information and Organisational Sciences 23(2), pp.11.–26.

[3] Brumec, J. Dušak, V. 1999. The assessment of IS complexity based on genetic taxonomy, Evolution and Challenges in System Development, Kluwer Academic_Plenum Publ; New York,.

[4] Brumec, J. Vrček, N. 2002. Strategic Planning of Information Systems (SPIS)- A Survey of Methodology; CIT, Vol. 10, No. 3, pp. 225.-232.

[5] Brumec, J., Vrček, N. 1998. Structured and object-oriented methods in a complex IS. Journal of Information and Organisational Sciences 22 (2), pp. 45–59.

[6] De Waal, A.A. 2006. Strategic Performance Management: A Managerial and Behavioural Approach, Palgrave Macmillan, London

[7] Hell, M., Vidačić, S., Garača, Ž. 2009. Methodological Approach to Strategic Performance

Optimization; Management, Vol 2, pp.21.-42.

[8] Martinsons, M., Davidson, R., Tse, D. 1999. The Balanced Scoredcard: A foundation form the Strategic Management of Information Systems; Decision Support System, pp.71-88.

[9] Roest, P., 1997. The golden rules for implementing the balanced business scorecard; Information Management & Computer Security 5/5,; pp163.-165.

[10] Ross, J. W. 2003. Creating a Strategic IT Architecture Competency: Learning in Stages; MIT Sloan of Management Working Paper (Research Article) No.4314-03,

MODELING AND HANDLING UNCERTAIN UTILITY IN PUBLIC SERVICE SYSTEM DESIGN

Jaroslav Janáček
University of Žilina, Faculty of Management and Informatics,
Univerzitná 1, 010 26 Žilina, Slovak Republic,
jaroslav.janacek@fri.uniza.sk,

Abstract: The paper deals with the public service system design, in which the system optimal solution is searched for. The system utility function is represented by the sum of users' utilities, where an individual user's utility depends on the distance between the user and the nearest located service center. The user's utility is represented by a decreasing real function, which depends not only on the distance, but even on a parameter, value of which is not known exactly. To solve the public service design problem under the uncertainty, we use the theory of fuzzy sets and present the integer programming approach to the service system design.

Keywords: public service system, system utility, uncertainty, fuzzy sets

1 INTRODUCTION

The design of a public service system [3], [6], [7] includes a determination of center locations, from which the associated service is distributed to the users of the system. The service facilities must be concentrated to a limited number of centers due to economic and technological reasons. We assume here that the service is delivered to users from the nearest center along the shortest path on the transportation network, which covers the served area. Then, the public service system structure is determined by the deployment of limited number of service centers. In many approaches to the public service system design the associated objective in the standard formulation is to minimize the social costs, which are proportional to the distances between served objects and the nearest service centers. The user's utility in some public service systems is not proportional to the distance from the nearest service center. For example, the utility in emergency systems is almost constant, if the distance is small, and beyond some threshold it suddenly drops to zero. We model this user's utility by nonlinear function, where the threshold represents a parameter of the function. As each user can apply his or her attitude to the perceived utility followed from the service system, the uncertain value of the threshold may be considered as a fuzzy number. In the following sections, we introduce the utility function for an individual customer and formulate combinatorial model of the public service system design problem. Then, we give a transformation of the combinatorial model to a linear integer programming model with fuzzy coefficients in the objective function and suggest the necessary adjustment of the model for the Tanaka-Assai approach to be able to be used.

The series of linear integer problems is then solved by a special iterative process, which successively solves linear problems searching for the first feasible solution in each step. Advantages and disadvantages of the approach are studied and some results of the numerical experiments are presented in the concluding part of this paper to demonstrate efficiency of the approach, in the case, when a commercial software tool is used for obtaining final decisions on the service center deployment.

2 MODEL OF PERCEIVED USER'S UTILITY

The introduced model of the public service system utility for an individual user is based on taking into account the maximal utility contribution from the located service centers. The

utility contribution *u(t)* for a given service center depends on the time distance *t* between the user and the service center accordingly to the function described by (1). In the description the symbol t_{krit} represents some time-threshold (limit), where the utility contribution from the service center considerably drops, if the traveling time from the user to the service center reaches the limit. The positive shaping parameter *T* makes the decrease of the function steeper if it takes a value near to zero. The constant C_0 determines the maximal value of the contribution.

$$u(t) = w(t, t_{krit}) = \frac{C_0}{1 + e^{\frac{t - t_{krit}}{T}}} \tag{1}$$

If I_1 denotes the set of all located service centers in the public service system and t_{ij} denotes the travelling time from a user located at the place *j* to the service center location *i*, then the utility $U_j(I_1)$ of the system for the user *j* can be modeled by (2).

$$U_j(I_1) = \max\{u(t_{ij}) : i \in I_1\} \tag{2}$$

The public service system design problem with the system optimal utility for users is formulated as the task of service centers determination so that the sum of user utilities is maximal and the total number of located centers does not exceed a given number *p*. To describe the problems, we denote by *J* the set of user locations and by *I* the set of possible center locations. Let b_j denote the number of the users located at *j*. Then, the problem can be formulated in the following combinatorial form.

$$\max\{\sum_{j \in J} U_j(I_1) : I_1 \subset I, |I_1| \le p\} \tag{3}$$

3 PUBLIC SERVICE SYSTEM DESIGN PROBLEM

To formulate the public service system design problem with the system optimal utility on a discrete network, we use the above denotation of the set of users' locations by symbol *J* and a set of possible service center locations by symbol *I*.

At most *p* locations from *I* must be chosen so that the sum of users' utilities is maximum. The network time distance between a possible location *i* from *I* and user location *j* from *J* is denoted as t_{ij}. The decisions which determine the designed system can be modeled by further introduced decision variables. The variable $y_i \in \{0,1\}$ models the decision on service center location at place $i \in I$. The variable takes the value of 1 if a facility is located at *i* and it takes the value of 0 otherwise. In addition, the allocation variables $z_{ij} \in \{0,1\}$ for each $i \in I$ and $j \in J$ are introduced to assign a user location *j* to a possible service center location *i* ($z_{ij}=1$). Then the location-allocation model can be written as follows.

$$Maximize \sum_{i \in I} \sum_{j \in J} b_j u(t_{ij}) z_{ij} \tag{4}$$

$$Subject\ \ to \sum_{i \in I} z_{ij} = 1 \ \ for \ \ j \in J \tag{5}$$

$$z_{ij} \le y_i \ \ for\ i \in I,\ j \in J \tag{6}$$

$$\sum_{i \in I} y_i \le p \tag{7}$$

$$y_i \in \{0, 1\} \ \ for\ i \in I \tag{8}$$

$$z_{ij} \in \{0, 1\} \ \ for\ i \in I,\ j \in J \tag{9}$$

In this model, the objective function (4) gives the system utility value. The constraints (5) ensure that each user's location is assigned to exactly one of the possible service centers. The link-up constraints (6) assure that the users' locations are assigned only to the located service centers and constraint (7) limits the number of located facilities by *p*.

The problem (4)-(9) can be easily reformulated to the p-median problem, what is the task of determination of at most *p* network nodes as facility locations so that the sum of distances between each node and the nearest facility is minimal. Nevertheless the p-median problems associated with the above-mentioned service system designs are characterized by considerably big number of possible service center locations. To obtain good decisions on facility locations in a serviced area, a mathematical model of the problem can be formulated and some of mathematical programming methods can be used to obtain the optimal solution. The location-allocation model constitutes such mathematical programming problem, which resists to any attempt at fast solution. On the other side, it is known that large instances of the covering problem are easy to solve by common optimization software. The necessity of solving large instances of the p-median problem has led to the radius formulation [1], [2], [4], [5]. This approach avoids assigning the individual user location to some of located facilities and deals only with information, whether some facility is or is not located in a given radius from the user. The later approach leads to the model similar to the set covering problem, which is easily solvable even for large instances by a common optimization software tools.

4 UNCERTAINTY IN THE USER'S UTILITY FUNCTION

In this paper, we consider only the sort of uncertainty, which is connected with t_{krit} value in the utility function *u(t)* defined by (1). To describe the properties of *u(t)* regarding parameter t_{krit}, we use the denotation $u(t)=w(t, t_{krit})$. We describe the uncertain value of t_{krit} by the triangle fuzzy number defined by the membership function $\mu_{\boldsymbol{tkrit}}(s)$ depicted in Fig. 1.

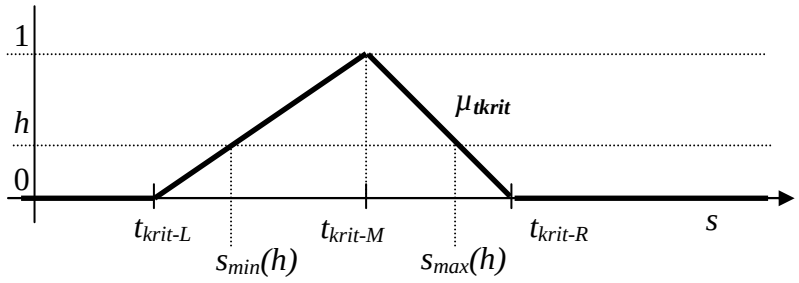


Figure 1: The membership function of a triangular fuzzy number $\boldsymbol{t}_{krit}$.

The membership function $\mu_{\boldsymbol{tkrit}}$ assigns a given value of *s* to the value $\mu_{\boldsymbol{tkrit}}(s)$, which expresses the power with which the value *s* belongs to the fuzzy number $\boldsymbol{t}_{\boldsymbol{krit}}$. The shape of triangle membership function defined on the universe *R* of real numbers is described by three values, which are denoted $t_{krit\text{-}L}$, $t_{krit\text{-}M}$ and $t_{krit\text{-}R}$ respectively (see Fig. 1).

Due to piece-wise linear form of the function $\mu_{\boldsymbol{tkrit}}$, the smallest and biggest values, which belong to the fuzzy number at a level of satisfaction *h*, can be determined by expressions $s_{min}(h)= t_{krit\text{-}L} + h(t_{krit\text{-}M} - t_{krit\text{-}L})$ and $s_{max}(h) = t_{krit\text{-}R} - h(t_{krit\text{-}R} - t_{krit\text{-}M})$ respectively.

The discussed user utility function $w(t, s)$ defined by (1) monotonously increases with increasing *s*. Assuming that $\boldsymbol{t}_{\boldsymbol{krit}}$ is a fuzzy number, the $\boldsymbol{u}(t) =w(t, \boldsymbol{t}_{\boldsymbol{krit}})$ is a fuzzy set defined on universe of real numbers for each fixed value of *t*. Due to monotonicity of the function $w(t, s)$, the fuzzy set $\boldsymbol{u}(t)$ is also the fuzzy number with nonlinear membership function $\mu_{\boldsymbol{u(t)}}$ defined accordingly to the "extension rule" [8] by (10).

$$\mu_{\mathbf{u}(t)}(v) = \max\{\mu_{tkrit}(s) : s \in R, v = w(t,s)\} \tag{10}$$

Furthermore, it follows from the monotonicity that $\mu_{\mathbf{u(t)}}(w(t, s)) = \mu_{\mathbf{tkrit}}(s)$. We can also determine the smallest and biggest values $u_{min}(t, h)$ and $u_{max}(t, h)$, which belong to the fuzzy number $\mathbf{u}(t)$ at a level of satisfaction h as $u_{min}(t, h) = w(t, s_{min}(h))$ and $u_{max}(t, h) = w(t, s_{max}(h))$ respectively. Accordingly to the α-cut concept [8], the objective function (4) is also a fuzzy number for fuzzy value $\mathbf{t_{krit}}$ and for given values of the variables $\mathbf{z}$. The value $U(\mathbf{z})$ can be expressed by (11).

$$U(\mathbf{z}) = \sum_{i\in I}\sum_{j\in J} b_j w(t_{ij}, \mathbf{t}_{krit}) z_{ij} = \sum_{i\in I}\sum_{j\in J} b_j \mathbf{u}(t_{ij}) z_{ij} \tag{11}$$

The smallest and biggest values $U_{min}(\mathbf{z},h)$ and $U_{max}(\mathbf{z},h)$, which belong to the fuzzy number $\mathbf{U}(\mathbf{z})$ at a level of satisfaction h, can be determined by expressions (12) and (13) respectively.

$$U_{\min}(\mathbf{z},h) = \sum_{i\in I}\sum_{j\in J} b_j w(t_{ij}, s_{\min}(h)) z_{ij} \tag{12}$$

$$U_{\max}(\mathbf{z},h) = \sum_{i\in I}\sum_{j\in J} b_j w(t_{ij}, s_{\max}(h)) z_{ij} \tag{13}$$

5 SYSTEM OPTIMAL DESIGN UNDER UNCERTAINY

The core of fuzzy approach to a general mathematical programming problem consists in determination of the highest level of satisfaction h, for which the associated constraints are satisfied and the objective function value belongs to a fuzzy set of satisfactorily big values of objective function. The fuzzy set of satisfactorily big values is usually constructed from two real values U^1 and U^2, where U^1 corresponds to the optimal objective function value for the least favorite case of the problem coefficients and U^2 corresponds to the optimal objective function value for the $_{krit\text{-}M}$ values or for the most favorite case of the problem coefficients. The membership function $\mu_{\mathbf{Ubig}}(U)$ is shown in Fig. 2.

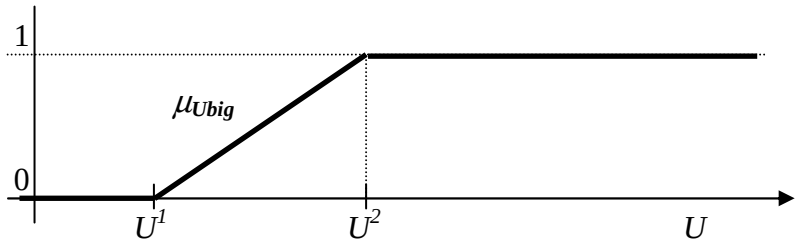


Figure 2: The membership function of a fuzzy set of sufficiently big values.

The constraint ensuring that the objective function value belongs to the sufficiently big values at a level of satisfaction h, follows.

$$U \geq (1-h)U^1 + hU^2 \tag{14}$$

Now, let us focus on the way, how the uncertainty influences the model (11), (5)-(9). We notice that the uncertainty influences only the objective function (11). This fuzzy constraint can be rearranged to the inequality (15), which expresses that the fuzzy value of the objective function belongs to the fuzzy set of big values.

$$\sum_{i\in I}\sum_{j\in J} b_j w(t_{ij}, s_{\max}(h)) z_{ij} \geq (1-h)U^1 + hU^2 \qquad (15)$$

Now we can formulate the associated problem as maximization of the level of satisfaction *h* subject to constraints (15), (5)-(9). Due to constraint (15), the problem is nonlinear and hard to solve. That is why; we use an iterative approach, which is known as the Tanako-Asai's method used in the fuzzy optimization [8].

The approach is based on a procedure, which searches only for a feasible solution of the problem formulated for a fixed value of *h.* If the feasible solution is found, then the value of *h* is increased, the associated model of the problem is reformulated and the searching process is repeated. In the opposite case, when no feasible solution exists, the next examined value of *h* is a bit lower. By a subsequent searching for feasible solutions for the increased or decreased values of *h* the optimal value can be estimated with an arbitrary precision ε.

We consider the following linear program for fixed value of *h*:

Maximize (13)

Subject to (15), (5) - (9).

As we denote the procedure *GetOpt(h),* which is able to solve the problem, we can implemented the Tanaka-Asai's method in accordance to the following steps, where ε is a demanded precision of the maximal level of satisfaction.

0. Set $h_{min} := 0$, $h_{max} := 1$.
1. Repeat the steps 2, 3 and 4 until $h_{max} - h_{min} < \varepsilon$ is met.
2. Set $h := (h_{max} + h_{min})/2$.
3. Apply procedure *GetOpt(h).*
4. If no solution $\boldsymbol{z}$ exists, set $h_{max} := h$, otherwise set $h_{max} := h$ and update the best found solution $\boldsymbol{z}^{best}$ and h^{best}.

6 PRELIMINARY NUMERICAL EXPERIMENTS

To reveal the properties of the suggested fuzzy approach to the public service system design and the impact of formalized uncertainty to deployment of the service centers, we performed the series of numerical experiments. The solved instances were derived from the real emergency health care system, which was originally designed for region of Zilina. This system covers demands of 315 communities - towns and villages spread over the region by 31 ambulance vehicles, where each of them represents one service center. These communities were considered as elements of the set *J* of users' locations and also as elements of the set *I* of possible service center locations. The time distances t_{ij} were computed from the road network distances for the average speed of 60 kilometer per hour.

The solved instances differ in values of t_{krit}, which takes values of =12, 14, 16, and 18 minutes in the utility contribution (1). In all experiments, the shaping parameter *T* was set to the value of 1 and the coefficient C_0 was set to the value of 10.

Each instance was solved for crisp value of t_{krit} and for further four cases with various level of uncertainty. The level of uncertainty is expressed by percentage of $t_{krit\text{-}R}$ regarding $t_{krit\text{-}M} = t_{krit}$ as 100 percent. We considered $t_{krit\text{-}R}$ as 110, 120, 130 and 140 percent of $t_{krit\text{-}M}$.

To solve the problems described by models (4)-(9) and (13), (15), (5) - (9), the optimization software FICO Xpress 7.3 (64-bit, release 2012) was used and the experiments were run on a PC equipped with the Intel® Core™ i7 3610 QM processor with the parameters: 2.3 GHz and 8 GB RAM.

To find characteristics of the public service system designed under uncertainty, the following parameters of the resulting system design were evaluated. For the fuzzy cases, the maximal level of satisfaction *h* at which the objective function value belongs to the suffi-

ciently high objective function values, was searched for. The associated optimal objective function (system utility) for the individual fuzzy and crisp cases is denoted as *objval*. The sum of time distances from user location to the nearest service center multiplied by number of users at the location is denoted by *wmed*. There is evaluated also the maximal distance *mxD* from a user to the nearest service center. To find influence of the uncertainty level to the resulting public service system design we evaluated also so called Hamming distances between the resulting vector ***y*** obtained for the crisp case and the vectors of location variables obtained for the fuzzy cases. This parameter is referred as *Hamming*. The label "*Ctime*" denotes the computation time in seconds, which was consummated by the respective computational process to obtain the associated solution. The results for individual instances are given in tables 1. – 4. Each table is organized accordingly to the scheme, where each column corresponds to one solved case. The cases distinguish in the level of anticipated uncertainty, which is described by *percentage*, which takes the values of 100, 110, 120, 130 and 140. The case with percentage 100 corresponds to the crisp case, where no fuzzy value is taken into account. The rows of the tables correspond with the above described parameters.

Table 1: Results of numerical experiments the instance with t_{krit} = 12

Percentage:	100	110	120	130	140
h	-	0.53	0.55	0.63	0.66
objval	67441	67839	68111	68441	68533
wmed	30378	29456	26055	25962	25962
mxD	20	20	26	20	20
Hamming	-	2	10	10	10
Ctime [s]	9	99	29	40	73

Table 2: Results of numerical experiments the instance with t_{krit} = 14

Percentage:	100	110	120	130	140
h	-	0.53	0.62	0.65	0.71
objval	68636	68769	68905	68938	68977
wmed	26150	26150	24782	30313	27252
mxD	20	20	17	17	17
Hamming	-	0	10	10	8
Ctime [s]	8	36	33	31	37

Table 3: Results of numerical experiments the instance with t_{krit} = 16

Percentage:	100	110	120	130	140
h	-	0.62	0.66	0.73	0.75
objval	69001	69057	69069	69080	69084
wmed	27252	28485	28485	28485	28485
mxD	17	17	17	17	17
Hamming	-	22	22	22	22
Ctime [s]	7	73	24	25	24

Table 4: Results of numerical experiments the instance with t_{krit}, = 18

Percentage:	100	110	120	130	140
h	-	0.65	0.73	0.75	0.8
objval	69090	69101	69105	69105	69105
wmed	26108	29772	29772	29772	29772
mxD	17	15	15	15	15
Hamming	-	6	6	6	6
Ctime [s]	6	24	23	24	24

It can be noticed that computational times are moderate for the size of the solved problem. The differences between the time referenced in the column denoted by ”100” and any value of time referenced in the other columns follow from the fact that the fuzzy cases are solved by the iterative process, which repeats the associated optimization several times.

If the entries of tables 1-4 are inspected in the order of increasing values of t_{krit}, it can be found that the values of parameters *wmed*, *mxD* and *Hamming* are constant with increasing level of fuzziness for the bigger values of t_{krit}. For the lower values of t_{krit}, the parameter *Hamming* tends to grow, while the parameters *wmed* and *mxD* tend to decrease or to stay constant with some random disturbance.

7 CONCLUSIONS

We suggested an approach to the public service system design, where user’s utility is modeled by non-linear function, which decreases with increasing time-distance of the user from the nearest located service center. In addition, we took into account some uncertainty connected with the utility perception by the individual users and we described the uncertainty by the triangle fuzzy number. The approach based on Tanaka-Asai’s method proved to be convenient for the optimal system design computation under assumption that the size of the problem does not exceed the size of a common region. The whole approach is represented by one program in the programming language *Mosel* and the design can be worked up using common commercial IP-solver.

The further research in this area will be aimed at usage of the radial formulation of the weighted p-median problem with the purpose to solve larger instances of the public service system design problem. The second branch of our research will be devoted to exploration of the cases, when also the value of the shaping parameter *T* is uncertain.

Acknowledgment: This work was supported by the research grants VEGA 1/0296/12 “Public Service Systems with Fair Access to Service” and APVV-0760-11 “Designing of Fair Service Systems on Transportation Networks”. We would also like to thank to „Centre of excellence for systems and services of intelligent transport“(ITMS 26220120050) for built up the infrastructure, which was used.

References

[1] Avella, P., Sassano, A., Vassil'ev, I., 2007. Computational study of large scale p-median problems. Mathematical Programming 109, pp. 89-114.

[2] Cornuéjols, G., Nemhauser, G.L., Wolsey, L.A., 1980. A canonical representation of simple plant location problems and its applications. SIAM Journal on Algebraic and Discrete Methods 1 (3), pp. 261-272.

[3] Current, J., Daskin, M., Schilling, D., 2002. Discrete network location models. In: Drezner Z. (ed) et al. Facility location. Applications and theory, Berlin, Springer, pp. 81-118.

[4] García, S., Labbé, M., Marín, A., 2011, Solving large p-median problems with a radius formulation. INFORMS Journal on Computing 23 (4) pp. 546-556.

[5] Janáček, J., 2008. Approximate Covering Models of Location Problems. In: Lecture Notes in Management Science. Proceedings of the 1st International Conference on Applied Operational Research ICAOR ´08, Vol. 1, Sept. 2008, Yerevan, Armenia, pp. 53-61.

[6] Janáček, J., Linda, B., Ritschelová, I., 2010. Optimization of Municipalities with Extended Competence Selection. Prager Economic Papers-Quarterly Journal of Economic Theory and Policy, Vol. 19, No 1, p. 21-34.

[7] Jánošíková, L., 2007. Emergency Medical Service Planning. Communications –Scientific Letters of the University of Žilina, Vol. 9, No 2, p. 64.

[8] Teodorovič, D., Vukadinovič, K., 1998. Traffic Control and Transport Planning: A Fuzzy Sets and Neural Networks Approach. Boston: Kluwer Academic Publishers, 387 p.

ZONE PARTITIONING PROBLEM WITH GIVEN PRICES AND NUMBER OF ZONES IN COUNTING ZONES TARIFF SYSTEM

Michal Kohání
University of Zilina, Faculty of Management Science and Informatics
Univerzitna 8215/1, 01026 Zilina, Slovakia
Michal.Kohani@fri.uniza.sk

Abstract: When designing a tariff system in regional public transportation, there are several approaches how to design it. One of various approaches is the zone tariff where the whole region is devided into the smalle sub-regions - tariff zones. We propose mathematical model of the zone partitioning problem with average deviation criterion. We perform a computational study using a universal optimization tool Xpress on the test data of selected regions with various problem sizes to study effectivenes of the model solution and the solving procedure.

Keywords: tariff planning, tariff zones design, IP solver, p-median problem.

1 INTRODUCTION

When the transport authorities plan the regional public transportation, one of the problems they deal with is the problem of the tariff and the ticket prices. As was mentioned in [4] and [9], there are various tariff types, such as *distance tariff*, *unit tariff* and *zone tariff*. In the distance tariff the price for travelling depends on the real length of the trip, in the unit tariff system the price is flat for all trips and is independent on the distance. In the *zone tariff* system the region is divided into smaller sub-regions (*tariff zones*) and the price for travelling depends on the origin zone, the destination zone and usually also on the number of travelled zones during the trip.

In the zone tariff systems there are two ways of determining the price. In the zone tariff with arbitrary prices, the prices depend on the pair of origin and destination zones and the number of travelled zones is not important, because prices are given for all pairs of zones separately and arbitrary. In the counting zone tariff system, the price of trip is calculated according to the origin and the destination zone of the trip and the number of travelled zones. On the contrary to the zone tariff with arbitrary prices, for all trips hold that passing the same number of zones must have the same price. Example of a counting zone tariff system in Bratislava region in Slovakia is in the Figure 1.

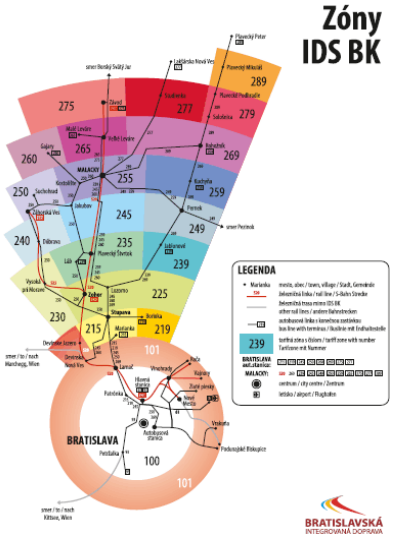


Figure 1: Example of the counting zones tariff system in Bratislava region *(www.bid.sk)*

Another important task in the planning process is how to design the zones and to fix the new fares. Hamacher and Schöbel in [3], Schöbel in [8] and Babel and Kellerer in [1] proposed approaches for the zone design problem with arbitrary prices. Hamacher and Schöbel in [4] and Schöbel in [9] mentioned the solving of the counting zones tariff system where the goal is to design the zones such that the new and the old price for most of the trips are as close as possible. They proposed three different objectives based on fair design, models for the fare problem and zone partitioning and three heuristic approaches. A note on fair fare rating was mentioned also by Paluch in [7]. Another approach was described by Müller, Haase and Klier in [6], where they formulated model and algorithm for revenue maximizing tariff zone planning.

This paper will be organizes as follows. In the section 2, we present the model of the zone partitioning with given prices and number of zones with the average deviation criterion for counting zones tariff system. To be able to calculate the optimal number of zones and fare prices, we will formulate the two stage algorithm in the section 3. In the chapter 4 we will present numerical experiments with three test examples of real networks to study computational demands of proposed model.

2 MATHEMATICAL MODEL OF THE ZONE PARTITIONING WITH GIVEN PRICES AND NUMBER OF ZONES

Let all stations in the network of public transport constitute the set of nodes I. The station i and j from set I are connected by the edge $(i,j) \in E$, if there is direct connection by public transport line between these two stations. Symbol E denotes the set of edges. The distance between stations i and j is denoted as d_{ij}. For each pair of stations i and j is c_{ij} the current price of travelling between these two stations. The number of passengers between stations i and j is b_{ij} (OD matrix).

If we want to calculate new price of the trip between nodes i and j in the counting zones tariff system, we need to calculate the number of zones crossed on this trip. The calculation of the number of crossed zones can be easily replaced by the calculation of crossed zone borders as was used in [4] and [6]. We assume that the station can be assigned only to one zone and then the border between zones is on the edge. We will introduce the binary variable w_{rs} for each existing edge $(r, s) \in E$, which is equal to 1 if stations r and s are in different zones and is equal to 0 otherwise. For calculation of the number of crossed borders we need to determine the used path for travelling between stations i and j. We introduce parameter a_{ij}^{rs}, where the used paths will be observed. a_{ij}^{rs} is equal to 1 if the edge (r,s) is used for travelling from station i to station j and 0 otherwise.

When we want to set a new price for travelling in such system, there are more possibilities how to do it. Hamacher and Schöbel in [4] and Schöbel in [8] proposed solution of fare problem with fixed zones to obtain new fares for trips with various travelled zones. In [5] and [6] a unit price for travelling per one zone was set. In this paper we define two different unit prices – price f_1 for travelling in the first zone and unit price f_2 for travelling in each additional zone. The final new price will be calculated as a sum of the basic price for the first zone and number of other travelled zones multiplied by the unit price for additional zones. This notion is more natural and often used also in distance tariff, where the average price per kilometre is higher for short trips. New price n_{ij}, determined by the number of crossed zones will be calculated according to this definition as follows (1):

$$n_{ij} = f_1 + \sum_{(r,s)\in V} f_2 a_{ij}^{rs} w_{rs} \tag{1}$$

Construction of the zone partitioning model was inspired by the model of the p-median problem. We introduce binary variables y_i, which represent the "fictional" centre of the zone. Variable y_i is equal to 1 if there is a centre of the zone in node i and 0 otherwise. For each pair of stations i and j we introduce variable z_{ij}. Variable z_{ij} is equal to 1 if the station j is assigned to the zone with centre in the node i and 0 otherwise. We expect to create at most p tariff zones.

When we want to suggest the objective function of the model, there are many possible ways. In [4] they proposed three different objectives based on fair design, in [6] authors formulated the criterion of revenue maximizing. In our model we will use the average deviation between current and new price for all passengers. According to the advices of experts in [9], in this paper we will use the average deviation between current and new price as a criterion in objective function. The current or fair price between stations i and j will be denoted by c_{ij}. The mathematical model of zone partitioning with fixed prices and number of zones (*Zone_part*) can be written in the form:

$$Minimize\ F = \frac{\sum_{i\in I}\sum_{j\in J}\left|c_{ij}-n_{ij}\right|b_{ij}}{\sum_{i\in I}\sum_{j\in J}b_{ij}} \tag{2}$$

$$subject\ to\ \sum_{i\in I} z_{ij} = 1,\ for\ j\in I \tag{3}$$

$$z_{ij} \le y_i,\ for\ i,j\in I \tag{4}$$

$$z_{ij} - z_{ik} \le w_{jk},\ for\ i\in I,(j,k)\in E \tag{5}$$

$$\sum_{i\in I} y_i \le p \tag{6}$$

$$z_{ij} \in \{0,1\},\ for\ i,j\in I \tag{7}$$

$$y_i \in \{0,1\},\ for\ i\in I \tag{8}$$

$$w_{ij} \in \{0,1\},\ for\ (i,j)\in E \tag{9}$$

Condition (3) ensures that each station will be assigned exactly to one zone. Condition (4) ensures that the station j will be assigned only to the existing centre of the zone. Condition (5) is coupling between variables for allocation of the station to the zone and variables for determining the zone border on the edge (j,k). Condition (6) ensures that we will create at most p tariff zones.

3 LINEARIZATION OF THE MODEL AND SOLUTION METHOD

This model will be solved using IP solver with exact methods, so we will obtain exact solution of the problem. Because the objective function (2) in this model is not a linear function, we need to modify this objective function to linear form. We introduce new variables u_{ij}, v_{ij}. Variables u_{ij} represent the calculated prices for travelling in case that new price is lower than current and variables v_{ij} represent the calculated prices for travelling in the opposite case. Then we can reformulate mathematical model *(Zone_part_lin)* to the linear form:

$$Minimize\ F = \frac{\sum_{i\in I}\sum_{j\in J}u_{ij}b_{ij} + \sum_{i\in I}\sum_{j\in J}v_{ij}b_{ij}}{\sum_{i\in I}\sum_{j\in J}b_{ij}} \tag{10}$$

$$subject\ to\ (3)-(9)$$

$$c_{ij} - n_{ij} = u_{ij} - v_{ij}, \; for \; i, j \in I \tag{11}$$

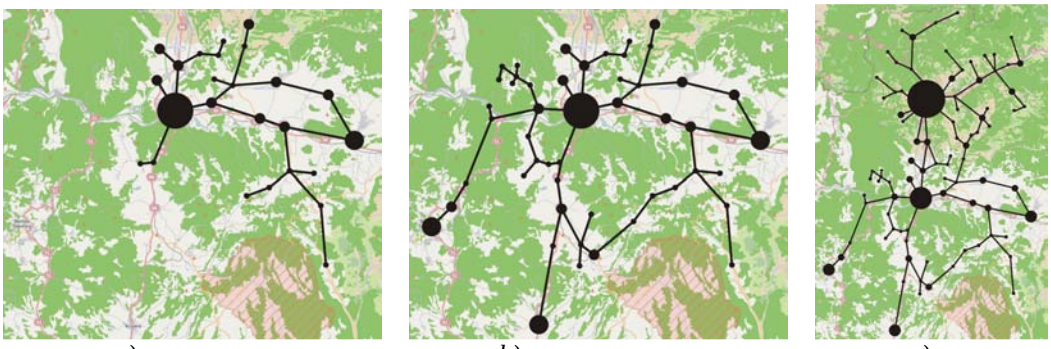
$$u_{ij} \geq 0, \; for \; i, j \in I \tag{12}$$

$$v_{ij} \geq 0, \; for \; i, j \in I \tag{13}$$

To determine the optimal values of parameters in the model, we can use a two-phase procedure. In the first phase we determine the optimal number of zones. In the second phase, for the given number of zones p, we repeatedly solve models with different settings of parameters f_1 and f_2. As the optimal we choose the solution of the model and parameters setting with the smallest value of the objective function.

4 NUMERICAL EXPERIMENTS

The goal of numerical experiments was to verify the possibilities of proposed model to find optimal zone partitioning with given values of number of zones and fare prices in the networks with different sizes. Numerical experiments were performed on the three data sets created from the real public transportation network in the Banska Bystrica Region in Slovak Republic. The stations in the networks are represented by the municipalities or part of municipalities. Networks have 25, 51 or 96 stations/municipalities respectively and are shown in the Figure 2. Black circles represent stations, the size of the circle represents approximate number of inhabitants and links represent existing connections of municipalities by public transportation.



a) *b)* *c)*

Figure 2: Test networks with a) 25 stations, b) 51 stations, c) 96 stations, Map source: *openstreetmap.org*

Current prices were calculated according to real prices depending on the distance for travelling by regional buses. The OD matrix was estimated using the gravity model as in [2], where the number of passengers between nodes i and j is calculated as follows:

$$\frac{b_i b_j}{d_{ij}},$$

where parameter b_i represents the number of inhabitants in the node i. To perform the computation we used the general optimization software tool FICO XPRESS 7.3 [10]. The experiments were performed on a personal computer equipped with Intel Core 2 Duo E6850 with parameters 3 GHz and 3.5 GB RAM.

In the computational study we focused on the second step of the proposed solving procedure. For selected values of parameter p we wanted to calculate optimal values of fare prices. According to the current fare prices, we set the values of parameter f_1 from 0.3 to 0.9 with step by 0.1 and values of parameter f_2 from 0.1 to 0.6 with step by 0.1 for all the experiments. Table 1 represents results for data set with 25 stations, in the Table 2 there are

results for data set with 51 stations and Table 3 represents results for the data set with 96 stations. In all tables the column *p_max* represent the value of parameter *p*, columns denoted as *F**, f_1* and f_2* represent optimal values of parameters and objective function for given *p* and the columns *Computational Time* represent values of total computational time (*Total*) of all instances with given *p* and minimal (*Min*) and maximal (*Max*) times from all instances in the set.

Table 1: Numerical experiments – data set with $|I| = 25$ stations

|I|=25				***Computational Time [s]***		
p_max	***F****	***f_1****	***f_2****	***Total***	***Min***	***Max***
4	4276.3	0.8	0.6	19.2	0.1	2.2
6	4124.3	0.8	0.5	17.9	0.1	1.4
8	4095.0	0.7	0.5	19.2	0.1	1.2
10	4054.4	0.7	0.4	21.9	0.1	2.2
13	3989.5	0.6	0.3	18.9	0.1	1.1
16	3999.2	0.6	0.2	18.7	0.1	1.5
18	4025.5	0.5	0.2	17.6	0.1	1.1
20	4030.9	0.5	0.1	19.2	0.1	1.7
22	4034.0	0.4	0.1	18.3	0.1	1.7

Table 2: Numerical experiments – data set with $|I| = 51$ stations

|I|=51				***Computational Time***		
p_max	***F****	***f_1****	***f_2****	***Total***	***Min***	***Max***
4	9780.9	0.9	0.6	147.9	0.4	12.6
6	9655.7	0.9	0.5	145.8	0.5	14.9
8	9628.7	0.9	0.5	148.5	0.5	11.5
10	9230.2	0.8	0.5	136.7	0.5	11.4
13	8455.2	0.8	0.4	137.8	0.5	8.8
16	8266.1	0.8	0.4	150.3	0.5	13.5
20	7859.5	0.7	0.4	140.7	0.5	12.6
25	7991.0	0.7	0.3	210.0	0.5	68.9
30	8382.9	0.7	0.2	164.0	0.5	19.0

Table 3: Numerical experiments – data set with $|I| = 96$ stations

|I|=96				***Computational Time***		
p_max	***F****	***f_1****	***f_2****	***Total***	***Min***	***Max***
4	36088.3	0.9	0.6	1449.0	1.7	361.9
6	34466.1	0.9	0.5	2065.8	1.9	611.4
8	34462.9	0.9	0.5	1368.2	1.9	189.9
10	33037.3	0.9	0.4	1689.2	1.7	374.1
13	34538.9	0.8	0.4	2576.2	2.0	1392.9
16	34538.9	0.8	0.4	1909.3	1.9	481.5
20	31839.5	0.7	0.4	3194.2	1.7	664.9
25	34542.5	0.7	0.3	1503.2	1.9	326.0
30	34544.7	0.6	0.2	1994.9	1.9	485.0

5 CONSLUSION

In the paper we described the mathematical model of the zone partitioning with given prices and number of zones with the average deviation criterion for counting zones tariff system and performed the numerical experiments on three different data sets with various size of the network. From the results of the numerical experiments we can see, that with the size of the problem computational times increase rapidly.
In the future we want to focus on the first step of the proposed solving procedure, incorporate dynamics of demand and prices into the model.

Acknowledgements
This work was supported by the research grants VEGA 1/0339/13 "Advanced microscopic modelling and complex data sources for designing spatially large public service systems", APVV-0760-11 "Designing of Fair Service Systems on Transportation Networks" and the institutional grant of Faculty of Management Science and Informatics (FRI ZU). We would like to thank to "*Centre of excellence for systems and services of intelligent transport*" (ITMS 26220120050) for built-up infrastructure, which was used.

References

[1] Babel, L., Kellerer, H., 2001. *Design of tariff zones in public transportation systems: Theoretical and practical results.* Technical report, Faculty of Economics, University of Graz, Austria

[2] Cenek, P., Janacek, J., Janosikova, L., 2002. Location of Transportation Districts at Modeling of Tranportation Processes in a Region. *Communications - Scientific Letters of the University of Zilina 4*, No. 1-2, pp. 5-9.

[3] Hamacher, H. W., Schöbel, A. 1995. On fair zone design in public transportation, *Computer – Aided Transit Scheduling, 430. Lecture Notes in Economics and Mathematical Systems*, Springer, Berlin Germany, pp. 8-22

[4] Hamacher, H. W., Schöbel, A., 2004. Design of Zone Tariff Systems in Public Transportation. *Operations Research 52*, pp. 897-908.

[5] Kohání, M., 2012. Exact approach to the tariff zones design problem in public transport, In: *Mathematical methods in economics: proceedings of the 30th international conference*, Silesian University in Opava, Karviná, pp. 426-431.

[6] Müller, S., Haase, K.., Klier, M., 2013. Revenue Maximizing Tariff Zone Planning in Public Transport, In: *Book of abstracts of XXVI EURO - INFORMS Joint International Conference*, Rome, pp. 335.

[7] Palúch, S., 2013. On a fair fare rating on a bus line. In: *Communications: scientific letters of the University of Žilina 15*, No. 1, pp. 25-28.

[8] Schöbel, A., 1994. Fair zone design in public transportation networks. In: *Operations Research Proceedings 1994*, Springer Verlag, Berlin, Germany, pp. 191-196.

[9] Schöbel, A., 2006. *Optimization in Public Transportation: Stop Location, Delay Management and Tariff Zone Design in a Public Transportation Network*, Springer, 2006.

[10] *XPRESS-Mosel User guide*. Fair Isaac Corporation, Birmingham, 2012.

DETERMINATION OF THE PORT ATTRACTIVENESS USING MIXED INTEGER LINEAR PROGRAMMING METHOD

Tomaz Kramberger*, **Tea Vizinger***, **Marko Intihar*** and **Anthony Chin****
*University of Maribor, Faculty of logistics, Mariborska 7, 3000 Celje, Slovenia
**National University of Singapore, 21 Lower Kent Ridge Road, Singapore 119077
tomaz.kramberger@fl.uni-mb.si

Abstract: In this paper we present well know Port Choice Problem as a Mixed Integer Linear Programing (MILP) problem. Using MILP we comapare North Adriatic ports with North European ports and determine their relative attractiveness for Bavarian shippers in case of importing containerized goods from far east. The results of the model show that despite better geographical position of North Adriatic ports Bavarian shippers are attracted to Nort European ports. The model also shows that land transport costs and subjective preferance rate play a large role in Port Choice.

Keywords: Mixed Integer Linear Programming, Port choice, Port attractiveness.

1 INTRODUCTION

Inter-port competition has encountered the enormous increase of volume in last few years [8]. In particular the port choice is important in multiple-port regions such as north Adriatic and north Mediterranean region. It has been recognized that the decision to route cargo through a port lies ultimately with the shippers [15]. The previous studies have identified and examined the factors that have influence on the port choice. Most of factors are good described in several papers as it follows, for instance cargo source, port facilities, delivery distance, port location and operating cost. Although the influences of these factors on port choice were explored in depth, the conclusions were different. Many researchers assumed that the port choice is the matter of minimizing the total operation cost, while the other claimed that the port choice is made from hinterland perspective.

With respect to the mentioned factors some mathematical models for port choice were proposed. Some of them use the linear programming technique to determine the optimal location of the port [7], the others proposed the weight factor analysis to integrate quantitative data with qualitative rating [14]. Lately the authors used also the fuzzy approach to solve the port selection problem [4]. But in general, no matter on the basis of two or more factors, they considered the problem of port choice as a multiple criteria decision-making problem.

In contrary to all these proposed approaches, we consider the port choice problem as a discrete optimization problem. We have modelled it as the connected weighted graph to minimize its total weight. The solution is a kind of trade off between the overall operating cost and preference factor developed in the previous papers [5].

The paper is organized as follows. In the next section we present brief literature review to justify the importance of Port Choice Problem. In third section we present the model and its solving. The data needed for modelling are presented in the fourth section. In the following section we present the computational results. The discussion of the results is done in conclusions.

2 LITERATURE REVIEW

Seaport researches have a long and rather interesting history. A structured review on methodological issues since 1980s can be observed in [18]. On the field of research into the Port Choice problem one can find many papers by various authors. The importance of this problem is evident by the fact that Sargent dealt with it already in 1938 [11] (he claimed: cargo tends to seek the shortest route to access the sea).

A number of mathematical programming models have been developed in order to minimize the total operation cost by selecting an appropriate port as the most favourable one to call. Therefore the port choice problem is often considered as a Multiple Criteria Decision Making problem (MCDM). But as explained in previous paragraph, the shipping carriers not only aim to minimize the cost but also take into account other criteria such as the volume of containers, port facility, port location, port operation efficiency and other conditions [6]. Chou [3] made a comparative study of models for port choice. He compared the Stackelberg model for port choice [19, 20], the Equilibrium model for port choice [2, 21] and fuzzy MCDM model for port choice [4]. The results show that these three models cannot be used to explain the actual port choices of carriers and shippers well. Thus Chou propose Analytic Hierarchy Process (AHP) model for the container port choice [5]. The results show that this AHP model seems to be promising. On the other hand Tran [16] studied port selection on liner routes from a logistics perspective. Paper introduces a non-linear model and heuristic model to minimize overall cost in cargo's journey, not only the seaside cost. The most important claim is that without taking into account of inland transport, we cannot fully understand the benefit of the direct call pattern on liner services. Weldman et al [17] introduce the demand choice function of a port's services to support the economic and financial evaluation of port investment projects. The outcomes of the linear regression model tests allow them to state that the location of a port is a key factor to explain the observed container port choice.

3 MODEL DESCRIPTION

3.1 Integration of subjectivity into the model

People usually do not behave in ways consistent with axiomatic rules, often their own. This often leads to violations of optimality. As we can see from the litearature review above mathematical programming models concearning cost does not explain the actual port choices of carriers and shippers well. The other factors, such as port facility, port location, port operation efficiency and other, are at least equaly important. These factors will be declared subjective. Their influence on the decision will be quantitatively defined as a preferance rate (PR). In the present section we wil describe integration of **subjectivity** into the mathematical model .

Let us define some notations first. Let O_i, $i = 1,2,...,I$ be the departure port, D_j, $j = 1,2,...,J$ the destination port, C_l, $l = 1,2,...,L$ point of consumption and S_k, $k = 1,2,...,K$ source point. Hence the path of moving goods from source to consumption point can be divide into three consisting parts, an edge $S_k O_i = x_{ki}$ between source point and departure port, the edge $O_i D_j = x_{ij}$ between departure and destination port and the edge $D_j C_l = x_{jl}$ between destination port and consumption point. In general, different shippers can use different departing

ports O_i. The situation we got, can be modeled as a graph as we can see on the figure 1 (heavy bolded line).

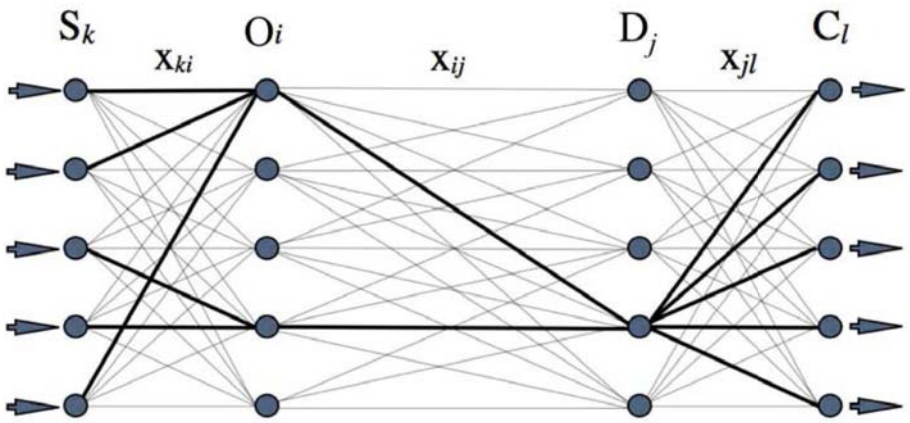


Figure 1: The required situation.

Generally speaking from the shippers point of view the most effective port is the port, which causes the lowest costs. If we look at the picture 1 we see that the costs for moving goods from S_k to C_l are the sum of land transport cost to move goods along the edge x_{ki}, the costs of maritime transport along x_{ij} and land transport cost along the edge x_{jl}. Therefore the costs of different parts of transport process can be express as a sum of weights $w_{x_{ki}}$, $w_{x_{ij}}$ or $w_{x_{jl}}$ assigned to certain edge respectively. The cost of this situation (see figure 1) can be mathematical expressed by the

$$W = \sum_i \left(\sum_k w_{x_{ki}} + w_{x_{ij}} \right) + \sum_l w_{x_{jl}} \tag{1}$$

When we study research which has already been done on this field we see that the costs are not the only criterion in the process for decision making [15, 1, 3, 5, 4]. One of promising criterion is so called preference rate gained with AHP method used by Chou in [5]. Let PR_{D_j} be the preference rate for j-th destination port. Now we can assume that PR_{D_j} has an impact on the weight of every edge connected to the port D_j, but the question that we must resolve is how much. Let we say that a certain percentage of weight is influenced by the performance rate. The first step is now to deduct the weights of the edges. For instance the edge D_jC_l we deduct by the PR_{D_j}. We get the following expression:

$$w'_{D_jC_l} \rightarrow \frac{1}{PR_{D_j}} \cdot w_{D_jC_l} \tag{2}$$

We can do the same with the weight of S_kO_i $\left(w'_{S_kO_i} \rightarrow \left(PR_{O_i} \right)^{-1} \cdot w_{S_kO_i} \right)$, but with the

weight of O_iD_j the picture is different. Preference rates of both ports have the influence on this weight. Because of that we simply calculate the average of the rates by

$$\overline{PR_{O_iD_j}} = \sqrt{PR_{O_i} \cdot PR_{D_j}} \tag{3}$$

At that point we still do not know what percentage of weight is influenced by the preference and which is not. We can simply write the equation $w_{D_jC_l} \cdot \left(1 - \frac{p}{100}\right) + \frac{1}{PR_{D_j}} \cdot w_{D_jC_l} \cdot \frac{p}{100}$, which tell us that the preference rate PR_{D_j} has the impact on p percent of the weight $w_{D_jC_l}$. After that maneuvers we can write a new deducted objective function as follows:

$$W' = \sum_i \left(\sum_k w'_{S_kO_i} + w'_{O_iD_j} \right) + \sum_l w'_{D_jC_l} \tag{4}$$

In order to choose the most effective port from the view of shippers we choose the destination port D_j for which the value W' of the objective function is minimal.

3.2 Solving the model

The problem defined in subsection 3.1 is at first glance similar to well known Hub and Spoke concept pioneered by Delta Airlines back method in 1955 [10]. The hub location problem has been studied a lot since O'Kelly [19] formulated the single allocation hub and spoke model as a quadratic integer program. Skorin-Kapov and O'Kelly [13] considered the uncapacitated p-hub median problem and developed linear programming formulations of both single and multiple allocation models. We formulate the Mixed Integer Linear Programm as follows.

The objective function consists of sum of all used edges x_{ki} beetwen production points and departing ports, edges x_{ij} between departing and destination ports and edges x_{jl} between destination ports and consumer points, all multiplied by their weights defined in section 2. As seen from the figure 1 the direct connections between production points and destination ports or consumer points are not allowed. All the paths from production points to consumer points need to go through two hubs, namely departing and destination ports. We can write the objective function to be minimized as:

$$B_{COST}J_{OPT} = \sum_{k=1}^{K}\sum_{i=1}^{I} x_{ki} \cdot w_{ki} + \sum_{i=1}^{I}\sum_{j=1}^{J} x_{ij} \cdot w_{ij} + \sum_{j=1}^{J}\sum_{l=1}^{L} x_{jl} \cdot w_{jl} \tag{5}$$

The value of cost J_{OPT} gives us the total cost of the solution which is the cheapest according to several constraints. We have three sets of constraints: for production points, for ports (departing and destination) and for consumption points. Production points constraints are formulated as

$$\sum_{i=1}^{I} x_{ki} \geq \frac{sp_{s_k}}{\sum_{k=1}^{K} sp_{s_k}} \qquad k = 1,2,...,K \tag{6}$$

where the left side is the flow from each S_k to all O_i and is grater or equal than the

supply sp into the S_k divided by the sum of all supplies. The constraints for departing ports are described as difference of incoming and outgoing flow at the port O_i which has to be greater or equal than zero.

$$\sum_{k=1}^{K} x_{ki} - \sum_{j=1}^{J} x_{ij} \geq 0 \qquad i = 1,2,...,I \tag{7}$$

Similar as for departure ports are constraints for destination ports that represent the difference between incoming and outgoing flow at the port D_j. Here aditional constraints ensure that only one port is selected at a time, so the sum $\sum_{i=1}^{I} x_{ij}$ is binary.

$$\sum_{i=1}^{I} x_{ij} - \sum_{l=1}^{L} x_{jl} \geq 0 \qquad j = 1,2,...,J \tag{8}$$

$$\sum_{i=1}^{I} x_{ij} = \begin{cases} 1\,; & \text{if there is a connection to } D_j \\ 0\,; & \text{otherwise} \end{cases} \tag{9}$$

The constraints for the consumer points cs are similar as for production points. On the left is the flow from D_j to all C_l, which is grater or equal than the demand in the C_l divided by the sum of all demands.

$$\sum_{j=1}^{J} x_{jl} \geq \frac{cs_{c_l}}{\sum_{l=1}^{L} cs_{c_l}} \qquad l = 1,2,...,L \tag{10}$$

4 REAL DATA MODELING

For real data modeling we have done some asumptions, about the vessel, departing and destination ports. Since the final result is expresed as the total weight of conected weighted graph, the input parameters are all weights and therefore real values are not important. Important are ratios between determinant parameters. The total weight is expressed with unit-less number. The port of choice is the port minimum total weight.

Shipping cost. The model is capable to simulate cost for many different types of vessel, but we chose Panamax size type of vessel, with GRT of 50350 tones, capacity of 4200 TEU and cruising speed of 21 knots.

Departing ports. Eventhough the model is capable to handle more ports on departure and destination side, we have chosen five of very frequently ports uniformly distributed over the South East Asia and East Asia. The Port of *Singapore, Honk Kong, Busan, Kaohsiung* and *Port Klang* are the ports, which are often used for transporting goods in Europe.

Destination ports. For the destination port we have chosen five ports uniformly distributed over the North Adriatic and three ports in Northern Europe. The candidate ports are *Koper, Rijeka, Trieste, Venezia, Ravenna, Rotterdam, Hamburg and Bremerhaven*.

Production points. Virtual production points are uniformly distributed over the South East Asia and East Asia.

Consumption points. For consumption point we have chosen four big consumption centers uniformly distributed over the Bavaria, i.e. *Regensburg, München, Ingolstadt, Nürnberg.*

Sailing time. We have calculated sailing time using online distance calculator searates and expressed them in days. We took in account the most common cruising speed for this kind of vessel of 21 knots.

Preference rate. We have calculated the preference rate using the analytic hierarchy process, presented by Saaty as explained in AHP [12]. We have ranked our five departing and eight destination ports according to eleven different criterions, which are the essence of criterions explained in papers [5, 1, 8] or [9, 15] and others. To get the data we have made a survey of several logistics providers, shippers, shipping lines and reatilers. We used a separate questionnaire for departing and destination side.

5 COMPUTATIONAL RESULTS

In order to calculate the results upon the model presented above we have build simple application of the model using Optimization Modeling Software LINGO 14.0. AHP analysis was done by using Matlab interactive environment for numerical computation. The results are shown on the table 1.

Table 1: The results, the port of choice is marked with number 1.

	Koper	Rijeka	Trieste	Venezia	Ravenna	Rotterdam	Hamburg	Bremerhaven
$C_{port}(ds)$	34033	35814	37164	35630	34095	43052	35900	36350
rank	1	4	7	3	2	8	5	6
	√	-	-	-	-	-	-	-
PR_{D_j}	0.097	0.095	0.106	0.101	0.1	0.168	0.167	0.166
rank	7	8	4	5	6	1	2	3
	-	-	-	-	-	√	-	-
$w'_{D_jC_l}$	3727	3861	3375	3795	4388	2701	2533	2652
rank	5	7	4	6	8	3	1	2
	-	-	-	-	-	-	√	-
$w_{O_iD_j}$	560	639.6	535	587	780	641	665	639.7
rank	2	4	1	3	8	6	7	5
	-	-	√	-	-	-	-	-
$ST^{day}_{O_iD_j}$	21	24	20	22	30	24	25	24
rank	2	4	1	3	5	6	8	6
	-	-	√	-	-	-	-	-

From the results we can se that the port of choice is really the trade off between the overall operating cost and other involved factors. For instance the winner, port of Hamburg, has enormously greater cost than many other competitors, but also grater preference rate. The sailing time is also the longets but the cost of land transport and PR compensate this disadvantage.

6 CONCLUSION

Graph in figure 1 represents the Port Choice problem as a discrete optimization problem by combining the subjective and objective factors. Mixed Integer Linear Program on the basis of these factors, calculates the optimal Port of Choice according to the given constraints.

The results obtained from the model can be an excellent base for a variety of policy decisions for port authorities, state or other decision-makers. The results show that despite the favorable geographical position of the North Adriatic ports North European ports win due to higher Preferance Rate and also because of good land transport connections. Conclusion for the North Adriatic Port Authorities should therefore be that they need to put a lot effort to increase the preferance of chossing North Adriatic port to Bavarian costumers, but this is already the matter for next papers.

References

[1] Bloningen, B., Wilson, W., 2006. International trade, transportation, networks and port choice. Tech. rep., Institute for Water Resources at the U.S. Army Corps of Engineers.

[2] Chang, W., 1974. Forecasting the imported/export cargoes demand split and the throughput of Taichung port. Master's thesis, National Taiwan University, Taiwan.

[3] Chou, C., 2005. A comparative study of models for port choice. In: Proceedings of the Eastern Asia Society for Transportation Studies 5, pp. 608–616.

[4] Chou, C., 2007. A fuzzy mcdm method for solving marine transhipment container port selection problem. Applied Mathematics and Computation 186, pp. 435–444.

[5] Chou, C., 2010. Ahp model for the container port choice in the multiple-port region. Journal of Marine Science Technology 18 (2), pp. 221–232.

[6] Chou, C., Kuo, F., Gou, R., Tsai, C., Wong, C., Tsou, M., 2010. Application of a combined fuzzy multiple criteria decision making and optimization programming model to the container transportation demand split. Applied Soft Computing 10, pp. 1080–1086.

[7] Dahlberg, M., May, J., 1980. Linear programming for sitting of energy facilities. Journal of Energy Engineering, pp. 5–14.

[8] Garcia-Alonso, Sanchez-Soriano, J., 2009. Port selection from a hinterland perspective. Marine Economics and logistics 11 (3), pp. 260–269.

[9] Mangan, J., L. C., Gardner, B., 2002. Modelling port/ferry choice in roro freight transportation. International Journal of Transport Management 1, pp. 15–28.

[10] O'Kelly, M. E., 1987. A quadratic integer program for the location of interacting hub facilities. European Journal of Operational Research 32, pp. 393-404.

[11] Sargent, A., 1938. Seaports and Hinterlands. Adam and Charles Black, London.

[12] Saaty, T. L., 1980. The Analytic Hierarchy Process: Planning, priority setting, resource allocation. McGraw-Hill, Bellingham, WA, 287 p.

[13] Skorin-Kapov, D., Skorin-Kapov, J., O'Kelly, M., 1996. Tight linear programming relaxations of uncapacitated P-hub median problems. European Journal of Operational Research 94, pp. 582-593.

[14] Spohrer, G., Kmak, T., 1984. Qualitative analysis used in evaluating alternative plant location scenarios. Industrial Engineering, pp. 52–56.

[15] Tongzon, J., 2009. Port choice and freight forwarders. Transportation Research Part E 45, pp. 186–195.

[16] Tran, N., 2011. Studying port selection on liner routes: An approach from logistics perspective. Research in Transportation Economics 32, pp. 39-53.

[17] Veldman, S., Garsia-Alonso, L., Vallejo-Pinto, J., September 2011. Determinants of container port choice in Spain. Maritime Policy and Management 38 (5), pp. 509–522.

[18] Woo, S., Pettit, J., Kwak, D., Beresford, A., 2011. Seaport research: A structured literature

review on methodological seaport research: A structured literature review on methodological issues since the 1980s. Transportation Research Part A 45, pp. 667–685.
[19] Yang, Z., 1995. Stackelberg equilibrium analysis of container cargo behaviour. Journal of the Eastern Asia Society for Transportation Studies 1, pp. 249–261.
[20] Yang, Z., 1996. An application of stackelberg problem to international container movement. In: Proceeding of the 1st JSPS-NUS Seminar on Integrated Engineering. pp. 125–134.
[21] Zong-Hwa-Consultant, 1978. Forecasting the volumes of cargoes and the ship-berth demand of Keelung port. Project report, Zong-Hwa Consultant Co.

A CONTINUOUS OPTIMIZATION APPROACH FOR FINANCIAL PORTFOLIO SELECTION UNDER DISCRETE ASSET CHOICE CONSTRAINTS

Mahdi Moeini
Braunschweig University of Technology, IBR, Algorithms Group,
Mühlenpfordtstr. 23, 38106 Braunschweig, Germany
moeini@ibr.cs.tu-bs.de

Abstract: In this paper we consider a generalization of the Markowitz's Mean-Variance model under linear transaction costs and cardinality constraints. The cardinality constraints are used to limit the number of assets in the optimal portfolio. The generalized model is formulated as a mixed integer quadratic programming (MIP) problem. The purpose of this paper is to investigate a continuous approach based on difference of convex functions (DC) programming for solving the MIP model. The preliminary comparative results of the proposed approach versus CPLEX are presented.

Keywords: portfolio selection problem, mixed integer programming, DC programming.

1 INTRODUCTION

Let us suppose that we are given a certain amount of money to invest. The investment must be done in a given set of assets or stocks. Each way of diversifying this amount of money between the given assets is called a *portfolio* [3]. The objective is to find a way to invest the money in the best possible way, which is called the optimal portfolio. This problem is known as the *portfolio selection problem* and it has been widely studied. Particularly, Markowitz [11] was one of the first researchers who provided a quantitative framework for finding the optimal portfolio. Markowitz [11] introduced the famous Mean-Variance (MV) model. The MV model is based on the expected return and the variance of returns between the assets [3]. The variance of returns is defined as the risk and, in this context; the objective of the portfolio selection problem consists of finding the set of portfolios offering the minimum level of risk for a given level of return. In order to find such portfolios, Markowitz proposes a convex quadratic programming (QP) model that is the MV model. This model has been widely used in practical applications. In spite of this fact, the standard MV model suffers from several inconveniences, for example, the MV model does not contain some practical constraints such as cardinality constraints, threshold constraints, or transaction costs functions. In fact, while an investor purchases or sells a stock, an extra charge will be made as the *transaction costs*. These costs must be taken into account in order to have realistic portfolio optimization models. There are different forms of the transaction costs functions: linear, piece-wise linear, step-wise linear functions, etc. The *cardinality constraints* limit the number of assets the optimal portfolio. The standard MV model is generalized by introducing these constraints [1-3]. The new model will be a mixed integer program (MIP) that is no more a convex programming problem. Due to the hardness of solving the MIP models, one needs to use local approaches that provide high quality solutions.

In this paper, we focus on solving the problem of portfolio selection under cardinality constraints in the presence of linear transaction costs that are proportional to the amount of the transactions. As the solution approach, a local deterministic method based on difference of convex functions (DC) programming and DC Algorithms (DCA) is used. This approach has been firstly introduced by Pham Dinh Tao in their preliminary form in 1985. They have been extensively developed since 1994 by Le Thi Hoai An and Pham Dinh Tao (see e.g. [7, 8, 12]). Due to successful application of the DC Algorithms for solving many large-scale mixed 0-1 programs (see, e.g., [4, 6, 8, 9]), a DC algorithm is developed for solving the

generalized MV model. For testing the efficiency of proposed algorithm, we compare it with the results of the standard solver CPLEX.

The paper is organized as follows. After the introduction, we present in Section 2 the model of the portfolio selection problem under cardinality constraints and linear transaction costs functions. Section 3 deals with DC programming, the reformulation of the proposed model in term of a DC program, and a special realization of DC algorithms to the underlying portfolio selection problem. Section 4 is devoted to the experimental results and some conclusions are reported in Section 5.

2 PORTFOLIO SELECTION PROBLEM UNDER CARDINALITY CONSTRAINTS

First of all, let us remind the famous Markowitz's Mean-Variance model for the portfolio selection problem [3, 11]. Let n be the number of available stocks, r_i be the mean return of stock i (for $i = 1,....,n$). $R \in \Re$ is the expected level of portfolio return and Q is the variance-covariance matrix computed by using the historical returns of the assets. The decision variable x_j is the proportional of the capital to be invested in the stock j. Using these notations, the standard Markowitz's Mean-Variance model is:

$$(P_{MV}): \quad \min\left\{ x^t Q x : x^t r \geq R, \sum_{j=1}^{n} x_j = 1, x_j \geq 0 \right\}.$$

This formulation is a simple convex quadratic program for which efficient algorithms are available. In this MV model, one minimizes the risk (i.e., $x^t Q x$) by ensuring the minimum level of portfolio return R.

In this paper, we study the generalized MV model by introducing realistic terms into the model. Particularly, we introduce the transaction costs and the cardinality constraints. The transaction costs are the amount of money that must be paid after each transaction (either purchasing or selling any stock). We suppose that the transaction costs are linear functions proportional to the amount of transactions. Furthermore, the cardinality constraints are introduced into the model to control the number of stocks representing the optimal portfolio. In order to define the cardinality constraints, we need to define the binary variables z_j (for $j = 1,....,n$). We define $z_j = 1$ if and only if the stock j is included in the optimal portfolio and $x_j \in [a_j, b_j]$, (where $0 \leq a_j \leq b_j \leq 1$ are lower and bounds, respectively), otherwise, z_j will be equal to 0. Furthermore, we are going to use the following complementary notations:

- $c_b, c_s \in \Re^n$: the transactions costs vectors for purchasing and selling stocks, respectively. We suppose that the transaction costs are proportional to the amount of the transactions;
- $x_b, x_s \in \Re^n$: vectors of the purchasing and selling variables, respectively;
- $P \in \Re^n$: the current holding portfolio of the investor;
- $\bar{x} \in \Re^n$: the benchmark portfolio;
- $z \in \Re^n$: the vector of binary variables, that are used for formulating the cardinality constraints;
- $card$: the cardinality parameter defining the number of the stocks in the final portfolio.

The generalized model is as follows:

(P_{card}):

$$\min \quad (x - \bar{x})^t Q (x - \bar{x}) \qquad (1)$$

Subject to:

$$(x-\bar{x})^t r-(c_b^t x_b+c_s^t x_s)\geq R, \tag{2}$$

$$P+x_b-x_s=x, \tag{3}$$

$$\sum_{j=1}^{n} x_j=1, \tag{4}$$

$$\sum_{j=1}^{n} z_j=card, \tag{5}$$

$$a_j z_j\leq x_j\leq b_j z_j : j=1,....,n, \tag{6}$$

$$z_j\in\{0,1\}: j=1,....,n,\ x_b, x_s\geq 0. \tag{7}$$

By solving this problem, one minimizes the total risk associated with the portfolio to change the current position P to the optimal portfolio x^* by purchasing (x_b) some stocks or selling (x_s) them (constraint (3)). $\bar{x}\in\Re^n$ represents the benchmark portfolio that can be ignored by taking it equal to zero. It has no crucial role in our model. The current situation of the portfolio is defined by P, that can be taken equal to zero, as well. The total amount of paid transaction costs are computed by $(c_b^t x_b+c_s^t x_s)$. The model ensures that the optimal portfolio has an expected level of return denoted by R after subtracting the transactions costs (constraint (2)). The constraint (4) means that the all amount of wealth must be invested in the stocks. The cardinality and bounding constraints are ensured by (5) and (6). The remaining constraints say which variables are continuous or binary.

It is well known that (P_{card}) is a Mixed Integer Program (MIP) that is an NP-hard problem. Due to this fact, one cannot use exact methods for solving this problem; particularly, if the dimension of the problem (i.e., n) is large. In the literature, different alternative methods have been proposed for solving the variants of MV model under cardinality constraints (see e.g., [2,3,5,9]). In this paper, we investigate a solution approach based on DC programming and DC algorithms for solving (P_{card}).

Before introducing the DC formulation of (P_{card}), a brief introduction to DC programming and DC algorithms is given in the following section.

3 SOLUTION METHOD VIA DC PROGRAMMING AND DC ALGORITHMS

3.1 DC Programming: A Short Introduction

In this section, we review some of the main definitions and properties of DC programming and DC Algorithms (DCA); where, *DC* stands for *difference of convex functions*.

Consider the following primal DC program

$$(P_{dc}):\quad \beta_p:=\inf\{F(x):=g(x)-h(x): x\in\Re^n\},$$

where g and h are convex and differentiable functions. $F(.)$ is a *DC function*, g and h are *DC components* of $F(.)$, and $g-h$ is called a *DC decomposition* of $F(.)$.

Let C be a nonempty closed convex set and χ_C be the indicator function of C, i.e., $\chi_C(x)=0$ if $x\in C$ and $+\infty$ otherwise. Then, one can transform the constrained problem

$$\inf\{g(x)-h(x): x\in C\},$$

into the following unconstrained DC program

$$\inf\{f(x):=\varphi(x)-h(x): x\in\Re^n\},$$

where $\varphi(x)$ is a convex function defined by $\varphi(x) := g(x) + \chi_C(x)$. Hence, without loss of generality, we suppose that the primal DC program is unconstrained and in the form of (P_{dc}). For any convex function g, its conjugate is defined by $g^*(y) := \sup\{\langle x, y\rangle - g(x) : x \in \Re^n\}$ and the dual program of (P_{dc}) is defined as follows

$$(D_{dc}): \quad \beta_d := \inf\{h^*(y) - g^*(y) : y \in \Re^n\},$$

One can prove that $\beta_p = \beta_d$ [12].

For a convex function θ and $x_0 \in dom\,\theta := \{x \in \Re^n : \theta(x_0) < +\infty\}$, the subdifferential of θ at x_0 is denoted by $\partial\theta(x_0)$ and is defined by

$$\partial\theta(x_0) := \{y \in \Re^n : \theta(x) \geq \theta(x_0) + \langle x - x_0, y\rangle, \forall x \in \Re^n\}.$$

We note that $\partial\theta(x_0)$ is a closed convex set in $\Re^n$ and is a generalization of the concept of derivative.

For the primal DC program (P_{dc}) and $x^* \in \Re^n$, the *necessary* local optimality condition is described as follows

$$\partial h(x^*) \subset \partial g(x^*).$$

We are now ready to present the main scheme of the DC Algorithms (DCA) [12] that are used for solving the DC programming problems. The DC Algorithms (DCA) are based on local optimality conditions and duality in DC programming, and consist of constructing two sequences $\{x^l\}$ and $\{y^l\}$. The elements of these sequences are trial solutions for the primal and dual programs, respectively. In fact, $\{x^{l+1}\}$ and $\{y^{l+1}\}$ are solutions of the following convex primal program (P_l) and dual program (D_{l+1}), respectively:

$$(P_l): \quad \inf\{g(x) - h(x^l) - \langle x - x^l, y^l\rangle : x \in \Re^n\},$$

$$(D_{l+1}): \quad \inf\{h^*(y) - g^*(y^l) - \langle y - y^l, x^{l+1}\rangle : y \in \Re^n\}.$$

One must note that, (P_l) and (D_{l+1}) are convexifications of (P_{dc}) and (D_{dc}), respectively, in which h and g^* are replaced by their corresponding affine minorizations. By using this approach, the solution sets of (P_{dc}) and (D_{dc}) are $\partial g^*(y^l)$ and $\partial h(x^{l+1})$, respectively. To sum up, in an iterative scheme, DCA takes the following simple form

$$y^l \in \partial h(x^l); \qquad x^{l+1} \in \partial g^*(y^l).$$

One can prove that the sequences $\{g(x^l) - h(x^l)\}$ and $\{h^*(y^l) - g^*(y^l)\}$ are decreasing, and $\{x^l\}$ (respectively, $\{y^l\}$) converges to a primal feasible solution (respectively, a dual feasible solution) satisfying the local optimality conditions. More details, on convergence properties and theoretical basis of the DCA, can be found in [12].

3.2 Reformulation of the problem

The model (P_{card}) is not in the form of a DC program. In order to reformulate (P_{card}), we use an exact penalty result presented in [10]. The process consists of formulating (P_{card}) in the form of a convex-concave minimization problem with linear constraints which is consequently a DC program. In order to simplify the notations, let us define

$$A := \left\{ \begin{array}{l} (x, x_b, x_s, z) \in \Re^{3n}_+ \times [0,1]^n : \sum_{j=1}^{n} x_j = 1, (x-\bar{x})^t r - (c_b^t x_b + c_s^t x_s) \geq R, P + x_b - x_s = x, \\ \sum_{j=1}^{n} z_j = card, a_j z_j \leq x_j \leq b_j z_j : j = 1,....,n \end{array} \right\}.$$

Using this notation, the (P_{card}) is transformed to

$$\min\left\{(x-\bar{x})^t Q(x-\bar{x}) : (x, x_b, x_s, z) \in A, z_j \in \{0,1\} : \forall j\right\}. \quad (8)$$

Define the penalty function $\alpha(.)$ by $\alpha(x, x_b, x_s, z) := \sum_{j=1}^{n} z_j(1-z_j)$. Clearly, $\alpha(.)$ is a concave function with nonnegative values on A and the feasible solutions' set of (8) can be written as

$$\left\{(x, x_b, x_s, z) \in A, z_j \in \{0,1\} : \forall j\right\} = \left\{(x, x_b, x_s, z) \in A, \alpha(x, x_b, x_s, z) \leq 0\right\}.$$

Consequently, (8) can be written as

$$\min\left\{(x-\bar{x})^t Q(x-\bar{x}) : (x, x_b, x_s, z) \in A, \alpha(x, x_b, x_s, z) \leq 0\right\}. \quad (9)$$

Since $(x-\bar{x})^t Q(x-\bar{x})$ is convex and A is a bounded polyhedral convex set, according to [10], there is $\theta_0 \geq 0$ such that for any $\theta > \theta_0$, the program (9) is equivalent to

$$(P_{card} - DC): \quad \min\left\{F := (x-\bar{x})^t Q(x-\bar{x}) + \theta\alpha(x, x_b, x_s, z) : (x, x_b, x_s, z) \in A\right\}. \quad (10)$$

The function F is convex in variables x, x_b, x_s and concave in variables z. Hence, the objective function of $(P_{card} - DC)$ is a DC function. A natural DC formulation of the problem $(P_{card} - DC)$ is

$$g(x, x_b, x_s, z) := (x-\bar{x})^t Q(x-\bar{x}) + \chi_A(x, x_b, x_s, z) \text{ and } h(x, x_b, x_s, z) := \theta\sum_{j=1}^{n} z_j(z_j - 1),$$

where χ_A is the indicator function over A, i.e., $\chi_A(x, x_b, x_s, z) = 0$ if $(x, x_b, x_s, z) \in A$, and $+\infty$, otherwise.

3.3 A DC algorithm for solving $(P_{card} - DC)$

According to the general framework of DC algorithms, we first need computing a point in the subdifferential of the function h defined by $h(x, x_b, x_s, z) := \theta\sum_{j=1}^{n} z_j(z_j - 1)$. This is done by:

$$(u^k, u_b^k, u_s^k, v^k) \in \partial h(x^k, x_b^k, x_s^k, z^k) \iff u^k = u_b^k = u_s^k = 0, \quad v^k = \theta(2z^k - 1). \quad (11)$$

Secondly, in order to compute $(x^{k+1}, x_b^{k+1}, x_s^{k+1}, z^{k+1}) \in \partial g^*(u^k, u_b^k, u_s^k, v^k)$, we need to solve the following *convex* quadratic program:

$$\min\left\{(x-\bar{x})^t Q(x-\bar{x}) - \left\langle (u^k, u_b^k, u_s^k, v^k), (x, x_b, x_s, z)\right\rangle : (x, x_b, x_s, z) \in A\right\}. \quad (12)$$

To sum up, the DC algorithm for solving $(P_{card} - DC)$ can be described as follows:

DC Algorithm for solving $(P_{card} - DC)$

1) **Initialization**: Let ε be a sufficiently small positive number, let $(x^0, x_b^0, x_s^0, z^0) \in \Re^{3n}_+ \times [0,1]^n$, and set $k = 0$;
2) **Iterations**: For $k = 0,1,2,....$, set $u^k = u_b^k = u_s^k = 0$, $v^k = \theta(2z^k - 1)$, and solve (12).
3) **Stopping criterion**: If $\left\|(x^{k+1}, x_b^{k+1}, x_s^{k+1}, z^{k+1}) - (x^k, x_b^k, x_s^k, z^k)\right\| \leq \varepsilon$, then stop, $(x^{k+1}, x_b^{k+1}, x_s^{k+1}, z^{k+1})$ is a solution, otherwise set $k \leftarrow k+1$ and go to the Step 2.

4 COMPUTATIONAL EXPERIMENTS AND RESULTS

The algorithm has been tested on two benchmark data sets that have been already used in [2, 3, 5]. These data sets correspond to weekly prices coming from the indices: *Hang Seng* in Hong Kong and *Dax 100* in Germany. The number n of different assets is 31 and 85, respectively. We suppose that $a_j = 0.05$ and $b_j = 1.0$ for both indices. Furthermore, θ is set to be 2.0, ε is equal to 10^{-6}, $P_j = 0$ and $\bar{x}_j = 1/n$ (for $j = 1,...,n$), $c_b, c_s = 0.1\%$ of transaction (buying/selling), and finally the value of R is chosen in a way to get feasible models. We have tested DCA and the standard IP solver IBM CPLEX for different values of the cardinality parameter *card* . A time limit of *1200* seconds has been set on the IP solver IBM CPLEX. In order to find a good initial solution for DCA, we first solve the relaxed problem of (P_{card}). The solution may not be integer, hence we round up each nonzero value to get an integer point.

In Tables 1 and 2, we give the results for two considered data sets. In these tables, the number of iterations of DCA, the computing time in seconds (CPU), and the solution values (Optimal Val.) obtained by each of the methods are presented.

Table 1: The results for the index Hang Seng in Hong Kong.

card	CPLEX		DC Algorithm (DCA)		
	Optimal Val.	CPU(s.)	Optimal Val.	CPU(s.)	Iterations
5	0.000080	4.031	0.000110	0.094	3
6	0.000062	10.297	0.000095	0.094	4
7	0.000052	29.500	0.000084	0.110	4
8	0.000043	54.485	0.000084	0.110	4
9	0.000038	107.860	0.000051	0.093	4
10	0.000033	154.546	0.000044	0.109	4
11	0.000029	140.562	0.000042	0.125	4
12	0.000026	48.235	0.000027	0.094	4
13	0.000022	21.141	0.000025	0.110	4
14	0.000020	9.906	0.000024	0.109	4
15	0.000018	3.094	0.000023	0.094	4

Table 2: The results for the index DAX 100 in Germany.

card	CPLEX		DC Algorithm (DCA)		
	Optimal Val.	CPU(s.)	Optimal Val.	CPU(s.)	Iterations
5	0.000071	1201.969	0.000114	0.343	4
6	0.000057	1201.157	0.000078	0.344	4
7	0.000050	1201.422	0.000072	0.360	4
8	0.000041	1201.297	0.000060	0.375	4
9	0.000037	1202.016	0.000056	0.344	4
10	0.000030	1201.500	0.000101	0.359	4
11	0.000029	1201.281	0.000068	0.360	4
12	0.000027	1201.282	0.000083	0.344	4
13	0.000026	1201.343	0.000050	0.359	4
14	0.000021	1201.110	0.000041	0.375	4
15	0.000020	1200.938	0.000038	0.359	4

The computational results show that DCA gives a good approximation of the optimal solution within a very short time. The running time is less than 1 second and the number of iterations is at most 4. It is interesting that the most of the values provided by DCA are exact until 4 or 5 digits after the point. When we compare the computational time that Cplex needs to find the solutions and the CPU time of the DCA, the achievements of the algorithm become more interesting.

5 CONCLUSIONS

In this paper, a new approach for solving the portfolio selection problem has been presented. Instead of the standard Markowitz Mean-Variance (MV) model, we have used an extension including the cardinality and bounding constraints. Furthermore, the extended model takes into account the linear transaction costs functions. The extended portfolio selection model is nonconvex and, consequently, very difficult to solve by existing algorithms. We have transformed the model to a DC program and developed a deterministic approach based on DC programming and DC algorithms (DCA). Preliminary numerical simulations show the efficiency of the proposed approach and its inexpensiveness in comparison to the standard IP solver of CPLEX. The good results make it possible to extend the work to larger dimensions and combining the DC algorithm with exact approaches in order to have a guarantee on the quality of the solutions. The work in these directions is currently in progress.

References

[1] Bartholomew-Biggs M., 2005. Nonlinear Optimization with Financial Applications, Kluwer Academic Publishers, First edition.

[2] Chang T.J., N. Meade, J.E. Beasley and, Y.M. Sharaiha, 2000. Heuristics for cardinality constrained portfolio optimization, Computers and Operations Research, Vol. 27, pp. 1271-1302.

[3] Fernandez A., Y.M. Gomez, 2007. Portfolio selection using neural networks, Computers & Operations Research, Vol. 34, pp. 1177-1191.

[4] Harrington J.E., B.F. Hobbs, J.S. Pang, A. Liu, G. Roch, 2005. Collusive game solutions via optimisation, Mathematical Programming Ser. B, Vol. 104, No. 1-2, pp. 407-43.

[5] Jobst N., M. Horniman, C. Lucas, G. Mitra, 2001. Computational aspects of alternative portfolio selection models in the presence of discrete asset choice constraints, Quantitative Finance, Vol. 1, pp. 1-13.

[6] Kroeller A., Moeini, M., Schmidt C., 2013. A Novel Efficient Approach for Solving the Art Gallery Problem, WALCOM 2013, LNCS 7748, pp. 5–16.

[7] Le Thi H.A. and T. Pham Dinh, 2001. A continuous approach for globally solving linearly constrained quadratic zero-one programming problems, Optimization, Vol. 50, No. 1-2, pp. 93-120.

[8] Le Thi H.A. and T. Pham Dinh, 2005. The DC (difference of convex functions) Programming and DCA revisited with DC models of real world non convex optimization problems, Annals of Operations Research, Vol. 133, pp. 23-46.

[9] Le Thi, H.A., Moeini, M., Pham Dinh, T., 2009. Portfolio Selection under Downside Risk Measures and Cardinality Constraints based on DC Programming and DCA. Computational Management Science 6(4), pp. 477–501.

[10] Le Thi H.A., T. Pham Dinh, V.N. Huynh, 2005. Exact Penalty Techniques in DC Programming, Research Report, LMI, National Institute for Applied Sciences - Rouen, France.

[11] Markowitz Harry M., 1952. Portfolio Selection, Journal of Finance, Vol. 7, No. 1, pp. 77-91.

[12] Pham Dinh T. and H.A. Le Thi, 1998, DC optimization algorithms for solving the trust region subproblem, SIAM J. Optimization, Vol. 8, pp. 476-505.

EFFICIENT CALCULATION OF BOUNDARY SOLUTIONS OF LINEAR INTERVAL DIFFERENTIAL INCLUSIONS

Damjan Škulj

University of Ljubljana, Faculty of Social Sciences

Kardeljeva pl. 5, Ljubljana, Slovenia

damjan.skulj@fdv.uni-lj.si, tel: +386 1 5805 289

Abstract. We propose a powerful new method for numerical estimation of the boundary solutions of linear differential inclusions. It combines a classical uniform grid method, which is generally computationally very expensive, with a much more efficient adaptive grid method. We provide an algorithm and demonstrate the method on a numerical example.

Key words. interval matrix differential equation, interval matrix, numerical solution

1 Introduction

To model uncertainty in parameters corresponding to *linear dynamical systems*, convex sets of matrices are often used. They have been successfully applied, for instance, in modelling of discrete time Markov chains with uncertain parameters (see [2, 3, 6, 7]). In the present article we apply similar methodology for continuous time models. The problem is a special case of *differential inclusions* used for modelling general uncertain systems [5]. We will propose a method for effective computation of differential inclusions where the multivalued maps are induced by convex sets of matrices.

The structure of the article is the following. In the next section we give an exact formulation of the problem. Then in Section 3 we propose two numerical methods. The first one is a slightly improved uniform grid method, which in general requires a large number of optimisation steps, and the second one is an adaptive grid method that vastly reduces the required number of optimisations. To make them functional, both methods then have to be combined, resulting in the algorithm presented in Subsection 3.3. We finish with a numerical example in Section 4.

2 Linear differential inclusions

Differential inclusions (see e.g. [5])

$$\dot{x} \in F(x), \tag{1}$$

are generalisations of a differential equations, where F is a set-valued map. Mostly, we are concerned with the sets of all possible solutions satisfying some initial conditions. Differential inclusions are mainly used in the theory of *dynamical systems* to model uncertainty in parameters of *differential equations*. Specifically we will restrict to the family of *linear differential inclusion* of the form

$$\dot{x} \in \mathcal{Q}x := \{Qx \colon Q \in \mathcal{Q}\}, \tag{2}$$

where $\mathcal{Q}$ stands for a set of matrices. Moreover, we require that $\mathcal{Q}x$ is of the form $[\underline{x}, \overline{x}] = \{x \colon \underline{x} \leq x \leq \overline{x}\}$, where $\underline{x} \leq \overline{x}$ and the inequality is componentwise, i.e. $\underline{x}_i \leq \overline{x}_i$. This convention will be adopted throughout the article. A property that ensures the above requirement is that the set of matrices has *separately specified rows* (see [6]). Such sets of matrices are sometimes called *rectangular sets of matrices* (see e.g. [1]). That is, a set of matrices is said to have separately specified rows if $\mathcal{Q} = \prod_{i=1}^{n} \mathcal{Q}_i$ where $\mathcal{Q}_i$ are sets of matrix rows.

The main purpose of the present article is to provide numerical methods for computation of the sets of solutions for general interval valued linear differential inclusions. Let $\mathcal{X}$ denote the set of all possible solutions x of differential inclusion (2) satisfying $x(0) = x_0$, where x_0 is a given initial vector. That is

$$\mathcal{X} = \{x \colon \mathbb{R}_+ \to \mathbb{R}^n \colon x(0) = x_0, \dot{x}(t) \in \mathcal{Q}x(t), \forall t \geq 0\}.$$

Moreover, let $\mathcal{X}(t) = \{x(t) \colon x \in \mathcal{X}\}$ denote the set of solutions corresponding to a specific time point. Thus, $\mathcal{X}(t)$ is a subset of $\mathbb{R}^n$. Moreover, it is of the form of an interval vector $\mathcal{X}(t) = [\underline{x}(t), \overline{x}(t)]$. We will call $\underline{x}(t)$ and $\overline{x}(t)$ the *minimal* and *maximal* solutions of linear differential inclusion (2) satisfying the initial condition, as themselves are also solutions of this differential inclusion.

The minimal and maximal solutions satisfy the minimal and maximal differential equations

$$\dot{\underline{x}} = \min_{Q \in \mathcal{Q}} Q\underline{x} \tag{3}$$

and

$$\dot{\overline{x}} = \max_{Q \in \mathcal{Q}} Q\overline{x}. \tag{4}$$

The calculation of the above minima and maxima is done through linear programming. Therefore at every time t we have a maximizing (or minimising) matrix $Q(t) \in \mathcal{Q}$ such that $Q(t)\overline{x}(t) = \max_{Q \in \mathcal{Q}} Q\overline{x}(t)$. Though, the matrix $Q(t)$ is in general unknown until $\overline{x}(t)$ is known. Besides, there is usually no analytical way to find $Q(t)$ and $\overline{x}(t)$ directly, whence numerical methods are needed.

3 Uniform and adaptive grid methods

Our goal is development of numerical methods for solving equations (3) and (4). Actually, we have obvious symmetry between the equations, and therefore we will only consider the equation (4) corresponding to the maximal solution.

A common approach to computation of sets of solutions of differential inclusions is to divide the time interval of interest, say $[0, T]$ into small enough subintervals, where the value of $Q(t)\overline{x}(t)$ is approximately constant. The solution at point t_{i+1} is then approximated by

$$x(t_{i+1}) = x(t_i) + (t_{i+1} - t_i)Q(t_i)\overline{x}(t_i). \tag{5}$$

At each step an optimisation problem has to be solved to find the maximising matrix $Q(t_i)$ from $\mathcal{Q}$.

The convergence to the exact solution using this approach, which belongs to the family of *discretisation methods* (see e.g. [4]), is in general slow. Therefore, the number of optimisation problems needed to be solved is in general very large.

To estimate the errors of approximations, we use vector and matrix norms. Thus, let $\|x\|$ denote any norm in $\mathbb{R}^n$, and $\|Q\|$ the corresponding operator norm of the matrix Q. For a set of matrices we define $\|\mathcal{Q}\| = \max_{Q\in\mathcal{Q}} \|Q\|$.

3.1 An improved uniform grid approximation

Clearly, the solutions of equation (2) are all continuous, including the minimal and maximal one. Therefore we may expect that $Q(t_i)\overline{x}(t)$ is close to $Q(t)\overline{x}(t)$ for $t \in [t_i, t_{i+1}]$, and then $\tilde{x}(t) = e^{(t-t_i)Q(t_i)}x(t_i)$ for $t \in [t_i, t_{i+1}]$ is also an approximation of the solution of eq. (4). Moreover, as $Q(t_i) \in \mathcal{Q}$, this approximation is also itself a solution of eq. (2), which may not be the case with (5). It is also possible to analytically estimate the error of the approximation with $\tilde{x}$, which for some $t \in \mathbb{R}_+$ is bounded from above by

$$g(t) = \frac{K}{2MN}(e^{2Mt} - 1) + e^{2Mt}E_0, \tag{6}$$

where $M = \|\mathcal{Q}\|, N$ is the size of partition, E_0 the error of initial estimate and

$$K = 7M^2T\|x_0\|e^{MT},$$

where T is the length of the interval where approximation is calculated. The error of the approximation is thus $O\left(\frac{1}{N}\right)$, but still N has to be very large in general to achieve a prescribed accuracy.

3.2 An adaptive grid approximation

The main drawback of the uniform grid approximation method is its computational cost. To overcome this problem we present another method that significantly reduces the number of points where maximizing matrix needs to be calculated. We exploit the fact that the set of matrices contains a finite number of extreme points, which suggest that the maximizing matrix function $Q(t)$ is piecewise constant. That is that

the matrix maximizing expression $Q\overline{x}(t)$ will likely maximize this expression on some interval following t.

Now suppose that we have an approximation of $\overline{x}(t)$ and that we know that $Q(t)\overline{x}(t') = \max_{Q\in\mathcal{Q}} Q\overline{x}(t')$ for every $t \leq t' \leq T$. Then clearly $\overline{x}(T)$ is equal to $e^{(T-t)Q(t)}\overline{x}(t)$. The method is applicable even if $Q(t)$ is only approximately constant on given interval. More precisely, the following holds. For simplicity assume that the interval of interest is $I = [0, T]$, and set $\tilde{Q} = Q(0)$. Now denote

$$x_n(t) = \sum_{k=0}^{n} \frac{(t\tilde{Q})^k}{k!}\overline{x}(0). \tag{7}$$

Note that $x_n(t)$ converge to $e^{t\tilde{Q}}\overline{x}(0)$. If $Q(t)$ is approximately constant on I then $x_n(t)$, for large n, is approximately equal to the solution of eq. (4). There is an elegant way to estimate the error of this approximation, as follows. Let e_N be a constant satisfying

$$\max_{Q\in\mathcal{Q}} Qx_n(T) - \tilde{Q}x_n(T) \leq Me_N,$$

for every $0 \leq n \leq N$, where $M = \|\mathcal{Q}\|$. Further we have that

$$e^{t\tilde{Q}}\overline{x}(0) - x_N(t) \leq \frac{(TM)^{N+1}}{(N+1)!}e^{TM}\|\overline{x}(0)\| =: \hat{e}_N.$$

Clearly $\hat{e}_N$ can become arbitrarily small for large enough N. Now the error of approximation of $\overline{x}(t)$ with $e^{t\tilde{Q}}\overline{x}(0)$ is bounded with

$$(2\hat{e}_N + e_N)(e^{Mt} - 1) + E_0e^{Mt}, \tag{8}$$

where E_0 is the error of initial estimate. So if e_N can be kept sufficiently small for large enough N, then the error can be kept within prescribed bounds.

3.3 Combining the uniform grid and the adaptive grid methods

Both, the uniform and the adaptive grid methods have advantages and disadvantages. The advantage of the former is its universal applicability, but suffers from high computational cost; and while the latter one is computationally very efficient, it has limited applicability. With a proper combination of both methods we propose a powerful method that makes use of the adaptive grid method on the intervals where this is possible and to bridge between those intervals uses the uniform grid method.

We propose an algorithm that determines the intervals where adaptive and uniform grid methods respectively are more suitable and calculates the maximal solution with required precision. The main steps of the algorithm are described in Algorithm 1.

The basic concern of the algorithm is to keep the error within required bounds. This is a non-trivial task, because it is impossible to know in advance how many steps will be required to complete the calculations. The error bounds (6) and (8) suggest that the error can be bounded by a suitable exponential function.

Algorithm 1 Finding the maximal solution.

procedure MAXIMALSOLUTION($x(0), \mathcal{Q}, T, E_{\max}$)
  $t_{\text{start}} \leftarrow 0$ ▷ start of the interval
  $t_{\text{end}} \leftarrow T$ ▷ end of the interval
  while $t_{\text{start}} < T$ **do**
    if $t_{\text{end}} - t_{\text{start}} > D$ **then** ▷ D is a given constant
      if ApplicableAdaptiveGrid($t_{\text{start}}, t_{\text{end}}$) **then**
        $[x(t_{\text{end}}), E(t_{\text{end}})] \leftarrow$ AdaptiveGrid($t_{\text{start}}, t_{\text{end}}$)
        ▷ new solution and error estimate
        $t_{\text{start}} \leftarrow t_{\text{end}}$
        $t_{\text{end}} \leftarrow T$
      else
        $t_{\text{end}} \leftarrow \frac{t_{\text{start}}+t_{\text{end}}}{2}$
      end if
    else
      $[x(t_{\text{end}}), E(t_{\text{end}})] \leftarrow$ UniformGrid($t_{\text{start}}, t_{\text{end}}$)
      ▷ new solution and error estimate
      $t_{\text{start}} \leftarrow t_{\text{end}}$
      $t_{\text{end}} \leftarrow T$
    end if
  end while
end procedure

Given an interval, say $[0, T]$ we thus require that the error at time $t \in [0, T]$ is below $\alpha e^{\beta M t}$, where α and β are suitable constants. The equations (6) and (8) suggest that the value of β must be at least equal to 1, and α is then calculated as $E_{\max} e^{-\beta M t}$, where $E_{\max}$ is the required maximal error.

The next thing to decide is when the interval $[t_{\text{start}}, t_{\text{end}}]$ is short enough to go with the uniform grid method. A reasonable criterion would be that the number of the optimisation steps required by the uniform method is smaller than a multiple (usually between 2 and 5 times) of the number of optimisation steps needed to test the applicability of the adaptive grid method.

4 Example

We now report the results of a simple numerical simulation. Let

$$Q = \begin{bmatrix} -0.7 & 0.3 & 0.4 \\ 0.2 & -0.9 & 0.7 \\ 0.5 & 0.5 & -0.1 \end{bmatrix} \text{ and } x_0 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

Consider the set $\mathcal{Q}$ of all matrices with row sums equal 0 between $Q - 0.1E$ and $Q + 0.1E$ where E is the matrix of ones. Clearly, $\mathcal{Q}$ is a convex set of matrices. We have that $\|\mathcal{Q}\| = 1.75$. We will estimate the boundary solution satisfying (4) and $\overline{x}(0) = x_0$ on the interval $[0, 1]$ with maximal allowed error 0.01.

With the uniform grid method, according to eq. (6), we would need approximately 160 000 steps. We have actually run a simulation with the combined method implemented, which took 355 optimisation steps, including those needed to test the applicability of the adaptive grid method. The resulting upper bound is

$$\overline{x}(1) = [0.7435\ 0.6165\ 0.1767]^T$$

and the lower bound is

$$\underline{x}(1) = [0.5087\ 0.3528\ -0.1802]^T.$$

A more detailed analysis of the performance of the algorithm shows that the uniform grid method was applied on those intervals:

$$[0, 0.0039],\ [0.0078, 0.0117],\ [0.0117, 0.0136],\ [0.0897, 0.0933]$$

$$[0.2562, 0.2620],\ [0.2620, 0.2678],\ [0.2792, 0.2848],$$

whose total length is 0.03. On all other intervals, whose total number is 9, the adaptive grid method was used. The degree of the reduction of computational requirements was similar for various randomly generated examples, and surprisingly it is not substantially affected by the number of extreme points of the set of matrices.

References

[1] M. Akian and S. Gaubert. Spectral theorem for convex monotone homogeneous maps, and ergodic control. *Nonlinear Analysis: Theory, Methods & Applications*, 52(2):637 – 679, 2003.

[2] R. J. Crossman and D. Škulj. Imprecise Markov chains with absorption. *International Journal of Approximate Reasoning*, 51:1085–1099, 2010.

[3] G. de Cooman, F. Hermans, and E. Quaeghebeur. Imprecise Markov chains and their limit behavior. *Probability in the Engineering and Informational Sciences*, 23(4):597–635, 2009.

[4] A. Dontchev and F. Lempio. Difference methods for differential inclusions: a survey. *SIAM Rev.*, 34(2):263–294, June 1992.

[5] G. V. Smirnov. *Introduction to the theory of differential inclusions*. Crm Proceedings & Lecture Notes. American Mathematical Society, 2002.

[6] D. Škulj. Discrete time Markov chains with interval probabilities. *International Journal of Approximate Reasoning*, 50(8):1314–1329, 2009.

[7] D. Škulj and R. Hable. Coefficients of ergodicity for Markov chains with uncertain parameters. *Metrika*, 76(1):107–133, 2013.

TABU SEARCH FOR A SINGLE MACHINE SCHEDULING PROBLEM WITH DISCRETELY CONTROLLABLE RELEASE DATES

Thevenin S. [a], Zufferey N. [a], and Widmer M. [b]

[a] Faculty of Economics and Social Sciences, HEC - University of Geneva, Uni-Mail, Bd du Pont-d'Arve 40, 1211 Geneva 4, Switzerland.
[b] University of Fribourg - DIUF, Decision Support & Operations Research, Bd de Pérolles 90, 1700 Fribourg, Switzerland.
simon.thevenin@unige.ch nicolas.zufferey-hec@unige.ch marino.widmer@unifr.ch

Abstract: We are interested in a single machine scheduling problem, where each job must either be scheduled within a given time window or rejected. The objective to minimize is the sum of tardiness penalties, release dates reduction costs (earliness penalties), and setup costs. We also take into account sequence dependent setup times. To tackle such a problem, a greedy heuristic and a tabu search are proposed. Due to time window constraints, feasibility has to be maintained after each move of tabu search, and we compare four repairing methods.
Keywords: scheduling, earliness penalties, abandon costs.

1 INTRODUCTION

When the production capacity of a company is overloaded, all received orders cannot be performed on time. It then makes sense to reject some of them. Following customer requirements, a *due date* corresponds to the date at which an order has to be delivered. Late deliveries lead to customers dissatisfaction, which is modeled by a tardiness penalty. Such tardiness penalties are quadratic functions depending on the completion time of the job. The *deadline* corresponds to the point in time where the dissatisfaction associated with the rejection of the order, modeled by a *rejection penalty*, is equal to the dissatisfaction of delivering late. In other words, it is preferable to reject the order to allow the client to get its goods by another supplier.

Usually, according to the scheduling terminology, no job can be scheduled before its associated *release date*. It often corresponds to the date at which all necessary raw materials are ready to be used. In contrast, we consider here the situation where release dates can be reduced (but remain integer). This incurs a cost, modeled by an earliness penalty, which is a quadratic function depending on the starting time of the job. Obviously, there is a lower bound and no job can start before its *available date*. Two situations, where the use of controllable release dates is relevant, are identified below.

1. As explained in [9], it may be profitable for the manufacturer and its suppliers to cooperate. In some cases, a supplier can allow to deliver raw materials earlier, which reduces the release dates. In counter part, the manufacturer will pay a higher price, which creates a win-win situation.
2. Production systems are often slowed down by a single bottleneck machine. In a flow shop environment, each job has to pass through a predefined sequence of machines. Release dates on the bottleneck machine can be reduced by speeding up the jobs preceding the bottleneck stage. This can be done by assigning more resources to these tasks (gas, electricity, human resources,...). A possible application is in the steel industry, where metal has to be heated up before to be rolled [5].

We moreover consider sequence depend setup times and costs between jobs of different families. They correspond to the time and costs (salaries and materials) associated with machine tunings between two successive jobs.

The considered problem (P) can be formally stated as follows. A set of n jobs is given, a subset of these jobs have to be selected and scheduled on a single machine which can handle only one job at a time. For each job j, the following data are given: a processing time p_j, an available date $\bar{r_j}$, a release date r_j, a due date d_j, a deadline $\bar{d_j}$, and a rejection penalty u_j. Let C_j and B_j respectively denote the completion time and the starting time of job j. In a feasible solution, each accepted (i.e., not rejected) job j satisfies $C_j \leq \bar{d_j}$ and $B_j \geq \bar{r_j}$. The earliness and tardiness penalties are respectively given in Equations (1) and (2), where w and w' are integer parameters.

$$E_j(B_j) = \begin{cases} w_j \cdot (r_j - B_j)^2 & \text{if } B_j < r_j \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

$$T_j(C_j) = \begin{cases} w'_j \cdot (C_j - d_j)^2 & \text{if } C_j > d_j \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Between two consecutive jobs j and j' of different families F and F', a setup time $s_{FF'}$ must be performed and a setup cost $c_{FF'}$ is incurred. Preemptions are not allowed and it is possible to insert idle time in the schedule. The objective function to minimize is the sum of the three following components: (1) the setup costs $c_{j,j'}$ between every successively performed jobs j and j'; (2) the rejection penalties u_j associated with each rejected job j; (3) the earliness and tardiness penalties $E_j + T_j$ for all accepted jobs j.

Note that the basic problem of scheduling jobs on a single machine to minimize setup costs is equivalent to the traveling salesman problem, which is NP-hard [7], and thus (P) is NP-hard too. As a consequence, heuristics are necessary to solve large size instances of (P). In [11], a greedy algorithm and a tabu search are proposed for the same problem with regular (i.e., non decreasing) cost functions instead of earliness and tardiness penalties. Using non regular cost functions, as it is the case here, implies however several modifications of the methods. The paper is organized as follows: a literature review is given in the next section, a greedy heuristic and a tabu search approach are proposed in Section 3, whereas Section 4 presents the performed experiments. Finally, a conclusion ends up the paper.

2 LITERATURE REVIEW

The range of problems consisting in selecting a subset of given jobs, and schedule them to minimize rejections and some other costs, are called *order acceptance and scheduling problems* (OAP). It has been studied in various scheduling environments, and a review is given in [10]. Such problems are particularly relevant in make-to-order production systems [15].

A problem related to (P) is studied in [6] and [14]. It consists in a single machine scheduling problem with release dates, deadlines, and sequence dependant setup times. The objective is to maximize the sum of the gains associated with each performed job, minus a weighted tardiness penalty. The authors propose a MILP (mixed integer linear programming) formulation, which is able to solve instances with up to 15 jobs, as well as constructive and local search heuristics. The local search method works in two steps: accept the orders first, then find a good sequence. The same problem is studied in [1], where the authors state that making simultaneously sequencing and order accepting decisions improves the results. Their approach consists

in a tabu search with *Swap* moves (i.e., exchange the position of two jobs). Note that in their version of *Swap*, it is allowed to exchange a performed job with a rejected one.

Earliness and tardiness penalties have captured a lot of attention due to their correspondence with the just in time paradigm. In [13] is mentioned that the use of quadratic tardiness functions is appropriated to model customers dissatisfaction. In [12] is studied the single machine scheduling problem consisting in minimizing quadratic earliness and quadratic tardiness penalties. The authors emphasize that quadratic penalties avoid situations in which only a few jobs contribute to the objective function. On the contrary to most scheduling objective functions, the one considered in this paper is not regular since earliness penalties are decreasing functions of the completion times. When objective functions are regular, most algorithms solving a single machine scheduling problem consist in finding an ordered sequence of jobs. From such a sequence, a schedule is easily built by starting each job as early as possible. In case of non regular cost functions, the insertion of idle times may decrease the costs. Therefore, building an optimal schedule when a production sequence is given is not as easy, and can be time-consuming. There exist a *timing algorithm* able to compute the optimal starting time of each job in $O(nlog(n))$ for the single machine scheduling problem where the function to minimize is the sum of linear earliness and tardiness penalties (e.g., [2]). In [8] is proposed a $O(n^2)$ timing procedure for the problem with quadratic tardiness penalties. In [4] is proposed a *dynamic programming* timing procedure able to browse all neighbors of a solution defined by the move *Swap* in $O(n^3log(n))$.

3 A GREEDY HEURISTIC AND A TABU SEARCH FOR (P)

In this section, heuristics are proposed for (P). Subsection 3.1 presents the used timing algorithm. The greedy algorithm and the tabu search approaches are respectively described in Subsections 3.2 and 3.3. Subsection 3.4 gives repairing procedures allowing to maintain feasibility for both proposed methods.

3.1 Timing algorithm

To solve (P), a solution s is modeled by an ordered sequences of job $\sigma(s)$, and a set of rejected jobs $\Omega(s)$. Given such a solution representation, a timing procedure computes the starting and ending times of each job of $\sigma(s)$, such that the objective function is minimized. We will adapt the timing procedure proposed in [4], which is particularly efficient for local search algorithms.

To take into account available dates and deadlines constraints, we set $E_j(t) = \infty$ if $t < \bar{r_j}$ and $T_j(t) = \infty$ if $t > \bar{d_j}$, for each job j of $\sigma(s)$. Therefore, an unfeasible solution would give an infinite cost. As the sequence of jobs is given, setup times associated with jobs of $\sigma(s)$ can be included in the processing times.

3.2 Greedy algorithm

A greedy procedure is a constructive heuristic. Starting from an empty solution, it builds a complete solution one step at a time. At each step, it performs the decision optimizing the objective. In line with the results found in [11], the first phase of the method consists in sorting the jobs by increasing slack time ($\bar{d_j} - \bar{r_j} - p_j$), where ties are broken by decreasing rejection penalties u_j (if there remain ties, they are broken randomly). In a second phase, jobs are taken one by one in the previously defined order, and inserted in the schedule at the position minimizing the costs.

Note that a job is rejected if it is better than inserting it. The insertions are enforced, that is, other jobs can be deleted to maintain feasibility. This last point will be clarified in Subsection 3.4.

3.3 Tabu search

Tabu search [3] is a local search metaheuristic. Starting from an initial solution s, at each iteration, it generates a neighbor solution s' from the current solution s. The set $N(s)$ of neighbor solutions of s is obtained by performing moves on s, which are slight modifications of the solution structure. To avoid cycling, a tabu list forbids to perform the reverse of recently performed moves. Basically, at each iteration, the best non tabu move is performed. Four types of moves are proposed for (P): *Add* takes a rejected job and inserts it in the schedule; *Drop* takes an accepted job and removes it from the schedule; *Reinsert* takes an accepted job, drops it from its current position, and inserts it elsewhere; *Swap*, exchanges the position of two jobs in $\sigma(s)$.

Note that all moves are enforced by using repairing procedures described in Subsection 3.4. We designed five different tabu structures. The first forbids to add a dropped job during t_1 iterations. The second forbids to remove an added job during t_2 iterations. The third forbids to move a job which has been added, reinserted or swapped, during t_3 iterations. The fourth forbids to move a job j between its two previous neighbors during t_4 iterations, if j has been reinserted or swapped. The cost function associated with each job is constant over the interval $[r_j, d_j]$. This induces plateaus in the search space. To escape quickly from such plateaus, a tabu status is associated with the cost of the most recently visited solutions during t_5 iterations: it is forbidden to visit a solution whose cost is tabu.

3.4 Repairing procedures

Adding a job may lead to an unfeasible solution due to available dates and deadlines constraints. To maintain feasibility, a repairing procedure must delete some jobs, and the choice of those jobs is a crucial point in local search methods for OAP. Note that a reinsert move can be performed by a drop move followed by an add move, and a swap move consists of two drops followed by two adds. As dropping a job cannot lead to unfeasible solutions, we only need a repairing procedure for the move *Add*. Assuming that job j is inserted at position p, we propose to use the three following methods.

Repairing procedure $\mathcal{R}_1$. Remove randomly a job adjacent to position p until the insertion of j is possible. Deleting jobs which are adjacent to the insertion position reduces the shifting of other jobs, which is expensive with quadratic penalties.

Repairing procedure $\mathcal{R}_2$. Let j' and j'' be two jobs such that j' is at the left of p, and j'' at its right. Jobs j' and j'' are said to be *blocking* if by shifting j' (resp. j'') as most as possible towards the left (resp. right), the insertion of j is still not possible. $\mathcal{R}_2$ deletes one of the closest blocking job to position p (ties are broken randomly) until the insertion of j is possible. These blocking jobs are likely to be associated with large earliness and tardiness penalties, and dropping them should not be expensive.

Repairing procedure $\mathcal{R}_3$. While the solution is not feasible, the job whose removal leads to the minimum cost is deleted.

4 EXPERIMENTS

To generate a set of instances for (P), two critical values are used: the number n of jobs, and a parameter α which controls the interval of time in which release dates and due dates are generated. More precisely, a value $Start$ is chosen large enough, and End is equal to $Start + \alpha \sum_j p_j$. Then, r_j is chosen in the interval $[Start, End]$, and d_j in $[r_j + p_j, End]$. Basically, methods are likely to reject more jobs in instances having small values for α. n is chosen in the set $\{25, 50, 100, 200\}$, and α in $\{0.5, 1, 2\}$. We generated one instance for each pair (n, α). The weights w_j and w'_j are randomly chosen in the set $\{1, 2, 3, 4, 5\}$. $\bar{d}_j$ and $\bar{r}_j$ are chosen is such that $T_j(\bar{d}_j) = E_j(\bar{r}_j) = u_j$. p_j is an integer randomly chosen in the interval $[50, 100]$. As observed in realistic situations, the rejection penalty u_j is related to the processing time: $u_j = \beta \cdot p_j$, where β is an integer randomly picked in the interval $[50, 200]$. The number of job families is chosen randomly between 10 and 20, setup times and costs are likely to be related in realistic situations, therefore $s_{FF'}$ is chosen in $[50, 200]$ and $c_{FF'} = \lfloor \gamma \cdot s_{FF'} \rfloor$, where γ is chosen in the interval $[0.5, 2]$. Note that the $c_{FF'}$'s and the $s_{FF'}$'s satisfy the triangle inequality.

Five methods are compared. *Greedy* refers to the method proposed in Section 3.2. A preliminary study showed that, for *Greedy*, it is better to use the timing procedure $\mathcal{R}_2$ to compute the costs associated with each position. $Tabu_i$ is the tabu search approach as described in Section 3.3, using repairing procedure $\mathcal{R}_i$. Parameters $(t_1, t_2, t_3, t_4, t_5)$ are set to $(80, 60, 90, 180, 30)$ for $n \in \{50, 100, 200\}$, and to $(20, 20, 15, 25, 10)$ for $n = 25$. Five different runs where performed for each method on each instance. Average results are presented in Table 1, where the column *Best* reports the best result found by any of the proposed methods for the considered instance. In each cell is indicated the percentage gap between the average result obtained by the concerned method and the *Best*.

Table 1: Comparison of the proposed methods

n	α	*Best*	*Greedy*	$Tabu_1$	$Tabu_2$	$Tabu_3$
25	0.5	115361	**0.00**	0.13	**0.00**	**0.00**
25	1	28602	6.70	**0.00**	7.39	**0.00**
25	2	149134	**0.00**	**0.00**	**0.00**	**0.00**
50	0.5	237414	0.89	2.03	3.62	**0.00**
50	1	148237	12.88	13.12	11.39	**4.22**
50	2	38899	32.40	2.72	2.55	**0.00**
100	0.5	550950	5.36	3.33	4.69	**1.11**
100	1	339100	23.90	12.77	10.95	**3.54**
100	2	31706	176.18	57.88	**42.38**	204.47
200	0.5	934898	22.17	**0.94**	1.43	10.66
200	1	473244	68.62	**3.03**	4.62	84.63
200	2	42397	302.82	**11.44**	12.45	1586.88
Average			28.70	10.22	**9.22**	23.70

The results clearly show the superiority of tabu search over *Greedy*, as the gap obtained by the best tabu search is 9.22%, versus 28.70% for Greedy. Tabu search with repairing procedure $\mathcal{R}_3$ obtains the best results for 8 instances over 12, however the results obtained for large instances are very bad. This is not surprising as $\mathcal{R}_3$ is efficient but very slow. The running time of $\mathcal{R}_3$ depends on the number of accepted jobs in the current solution, which is large for instances generated with large values for n and α. When the number of accepted jobs is large, $Tabu_3$ performs a small number of iterations, and the results are not good. This explains the very bad performance of $Tabu_3$ on the instance having $n = 200$ and $\alpha = 2$. $\mathcal{R}_2$ is slightly better

than $\mathcal{R}_1$: their respective average gaps are 9.22% and 10.22%. We would thus advise the use of repairing procedure $\mathcal{R}_3$ for small instances, and $\mathcal{R}_2$ for larger ones.

5 CONCLUSION

We propose a tabu search and a greedy algorithm to tackle an order acceptance and scheduling problem with controllable release dates and quadratic earliness and tardiness penalties. The proposed tabu search method is efficient, but cannot be applied to large instances due to the lack of speed of the timing procedure. Future works include to propose a way to speed up the neighborhood evaluation, and to propose hybrid metaheuristics for the problem.

References

[1] B. Cesaret, C. Oğuz, and F. S. Salman. A tabu search algorithm for order acceptance and scheduling. *Computers & Operations Research*, 39(6):1197 – 1205, 2012.

[2] M. R. Garey, R. E. Tarjan, and G. T. Wilfong. One-processor scheduling with symmetric earliness and tardiness penalties. *Mathematics of Operations Research*, 13(2):330–348, 1988.

[3] F. Glover. Tabu search - part I. *ORSA Journal on Computing*, 1:190–205, 1989.

[4] Y. Hendel and F. Sourd. An improved earliness-tardiness timing algorithm. *Computers & Operations Research*, 34(10):2931–2938, 2007.

[5] A. Janiak and W. Janiak. Single-processor scheduling problem with dynamic models of task release dates. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 41(2):264 –271, 2011.

[6] C. Oğuz, S. F. Salman, and Z. B. Yalçin. Order acceptance and scheduling decisions in make-to-order systems. *International Journal of Production Economics*, 125(1):200–211, 2010.

[7] C. H. Papadimitriou. The euclidean travelling salesman problem is np-complete. *Theoretical Computer Science*, 4(3):237 – 244, 1977.

[8] J. Schaller. Single machine scheduling with early and quadratic tardy penalties. *Computers & Industrial Engineering*, 46(3):511 – 532, 2004.

[9] N. V. Shakhlevich and V. A. Strusevich. Single machine scheduling with controllable release and processing parameters. *Discrete Applied Mathematics*, 154(15):2178 – 2199, 2006. International Symposium on Combinatorial Optimization CO'02.

[10] S. A. Slotnick. Order acceptance and scheduling: A taxonomy and review. *European Journal of Operational Research*, 212(1):1 – 11, 2011.

[11] S. Thevenin, N. Zufferey, and M. Widmer. Tabu search for a single machine scheduling problem with rejected jobs, setups and deadlines. In *9th International Conference of Modeling, Optimization and SIMulation (MOSIM 2012)*, Bordeaux (France), June 6-8 2012.

[12] J. Valente, M. Moreira, A. Singh, and R. Alves. Genetic algorithms for single machine scheduling with quadratic earliness and tardiness costs. *The International Journal of Advanced Manufacturing Technology*, 54:251–265, 2011.

[13] J. M. S. Valente and J. F. Gonçalves. A genetic algorithm approach for the single machine scheduling problem with linear earliness and quadratic tardiness penalties. *Computers & Operations Research*, 36:2707–2715, 2009.

[14] Z. B. Yalçin, C. Oğuz, and S. F. Salman. Order acceptance and scheduling decisions in make-to-order systems. In *Proceedings of the 3rd Multidisciplinary International Conference on Scheduling: Theory and Application (MISTA 2007)*, pages 80–87, Paris, France, August 2007.

[15] M. Zorzini, D. Corti, and A. Pozzetti. Due date (DD) quotation and capacity planning in make-to-order companies: Results from an empirical analysis. *International Journal of Production Economics*, 112(2):919 – 933, 2008.

A THRESHOLD FOR RETURNING USABLE LEFTOVERS BACK ON STOCK WHEN SOLVING ONE-DIMENSIONAL CUTTING STOCK PROBLEM WITH USABLE LEFTOVER

Luka Tomat, Mirko Gradišar and **Mitja Štiglic**
University of Ljubljana, Faculty of Economics
Kardeljeva ploščad 17, SI-1000 Ljubljana, Slovenia
{luka.tomat,miro.gradisar, mitja.stiglic}@ef.uni-lj.si

Abstract: For solving the one-dimensional cutting stock problem with usable leftover (1DCSPUL) exists many methods none of which consider the prevention of too many usable leftovers (UL) being returned back on stock after several successive instances. If the UL are longer than or equal to the threshold *t* they are returned back on stock to meet the future orders. Since the amount of UL on stock mostly depends on *t* we proposed a heuristic algorithm to determine optimal threshold *t* and the optimal number of UL in stock. The results show the effectiveness of the proposed method.

Keywords: inventory management, cost prevention, cutting, usable leftovers, optimization, simulation, heuristics.

1 INTRODUCTION

The one-dimensional cutting stock problem (1DCSP) occurs in many fields, for example in steel [1], paper [2], textile [3] and wood [4] industries. It is usually demarcated as cutting longer objects into shorter ones, which are required in an order [5]. To satisfy the order there are various possibilities that are called cutting plans. They diverge in the production of trim-loss. Decreasing the trim-loss is one of the the main objective in solving the 1DCSP [6].

Often other objectives must be taken into account as well. When items must be cut into an exactly required number of pieces, the outcome can be a high quantity of leftovers in stock after several consecutive instances. The leftovers that are returned to stock must be longer than or equal to some threshold *t*. They are termed usable leftovers (UL) due to their capability to be used again to fulfill the future order. Leftovers that are shorter represent the trim-loss. Such a cutting problem is called the 1DCSPUL [7].

The main issue in solving 1DCSPUL is the formulation of an objective function. The problem occurs if the trim-loss reduction is the only considered criteria since the bars would be cut to the threshold *t* and returned back on stock but the UL would unlimitedly accumulate in the stock. Such a situation would result in a high logistics and warehousing cost and therefore should be prevented.

According to the literature there is no method that would efficiently determine the threshold *t* with aim to prevent the excessive accumulation of UL in stock. The purpose of this paper is therefore to propose a method for solving the 1DCSPUL so that UL can be better controlled.

The paper has 5 sections. Section 2 defines the problem. In Section 3, the solution to the problem is developed. Section 4 presents the results. Finally, in Section 5, the conclusion is presented.

2 PROBLEM DEFINITION

To satisfy an order a definite number of bars, that are always adequate to fill the order, are available in stock. They can be of standard and nonstandard lengths. Nonstandard lengths are UL from previous orders. The order has to be satisfied in such manner that the trim-loss size and the amount of UL are minimized. Due to dependence of satisfying the next order on the

UL from previous instances, the minimization should not be restricted to a single order, but should be extended to a sequence of orders. Similar approach can be found in [8], but the study did not take the possibility of controlling the amount of UL in stock into account. When sequence begins there are only standard lengths in stock. They are considered as integers. We have used the following notation:

r = number of orders in the sequence.
l_{si} = item lengths in s-th order; $i = 1, \ldots, n_s$.
p_{si} = required number of pieces of l_{si}.
L_{sj} = bar lengths in s-th order; $j = 1, \ldots, m$.
δ_{sj} = leftover of L_{sj}.
x_{sij} = number of pieces of l_{si} having been cut from L_{sj}.
t = threshold for the trim-loss. Leftover that is larger than or equal to t is UL. Leftover that is smaller than t is trim-loss and is considered waste.
f = factor for which a cost of trim-loss is greater than the cost of the difference between the UL produced and used.

The 1DCSPUL is formulated as follows:

$$\min \sum_{s=1}^{r} \sum_{j=1}^{m} (f \cdot \delta_{sj} \cdot (w_{sj} + z_{sj}) + \delta_{sj} \cdot u_{sj} - L_{sj} \cdot z_{sj} - (L_{sj} - \delta_{sj}) \cdot v_{sj}) \quad (1)$$

s.t.

$$L_{sj} = \delta_{s-1j} \text{ if } u_{s-1j} = 1 \vee v_{s-1j} = 1, \forall j; L_{sj} = L_{1j} \text{ otherwise} \quad (2)$$

$$\delta_{sj} = L_{sj} - \sum_{i=1}^{n_s} l_{si} \cdot x_{sij} \quad \forall j \quad (3)$$

$$\sum_{j=1}^{m} x_{sij} = p_{si} \quad \forall i \quad (4)$$

$$x_{sij} \geq 0, \text{ integer} \quad \forall i, j \quad (5)$$

$$\delta_{sj} \geq 0 \quad \forall j \quad (6)$$

For the above model, the following functions are used:

$u_{sj} = 1$ if $\delta_{sj} < L_{sj} \wedge \delta_{sj} \geq t \wedge L_{sj} = L_{1j}, \forall j$; $u_{sj} = 0$ otherwise
$v_{sj} = 1$ if $\delta_{sj} < L_{sj} \wedge \delta_{sj} \geq t \wedge L_{sj} < L_{1j}, \forall j$; $v_{sj} = 0$ otherwise
$w_{sj} = 1$ if $\delta_{sj} < L_{sj} \wedge \delta_{sj} < t \wedge L_{sj} = L_{1j}, \forall j$; $w_{sj} = 0$ otherwise
$z_{sj} = 1$ if $\delta_{sj} < L_{sj} \wedge \delta_{sj} < t \wedge L_{sj} < L_{1j}, \forall j$; $z_{sj} = 0$ otherwise

The above formulation represents the minimization of trim-loss and the difference between the UL used and produced in r consecutive orders. There are no UL from previous orders in stock at the beginning of the sequence ($s = 1$).

3 SOLUTION DEVELOPMENT

With respect to the objective function the amount of UL in stock partially depends on difference between the UL used and produced as a result of an optimization method of a particular order, but mostly on the threshold *t*. Lower *t* would result in higher amount of UL. According to constraint (2) the UL could represent a great share of the stock, which would lower the stock-to-order length ratio and increase the size of the trim-loss since there would be less possible solutions available. In a way a presented paper continues the research published in [9], where more detailed explanation of the abovementioned ratio can be found. Also in the case of higher *t* the trim-loss increases due increased trim-loss on the account of the leftovers that could be used to satisfy future orders. Therefore the main problem is to determine the threshold t_m that would deliver minimal value of the objective function.

In addition to above explained problem we introduce a heuristic algorithm termed TOP (Figure 1) for determining threshold t_{opt} and the corresponding number of UL in stock U_{opt}. Thus the value of objective function is minimal or close to the minimum. The algorithm can be applied with the use of any existing method for solving the 1DCSPUL. In our case [3] was used. Presented

$h_0 \leftarrow$ maxint (maxint is the maximal integer value) (h_0 is the initial value of the criteria function)
$h_1 \leftarrow 0$ (h_1 is the initial value of the criteria function of the first experiment)
set e (e is the initial value of the threshold t and must be low enough that a small increase in t decreases the value of the criteria function)
$t \leftarrow e$
set r (r is the number of problem instances in the experiment. r must be high enough to reach the point where the amounts of UL used and produced in a single instance are equal.)
$g \leftarrow 0$ (counter of experiments)
while $h_{g-1} > h_g$ (Experiments are repeated while the value of the criteria function decreases. When it begins to grow, its minimum has been found at h_{g-1} with the threshold t_{opt}.)
    $g \leftarrow g + 1$
    empty stock of nonstandard bars (UL)
    set stock of standard bars
    repeat r times
        generate new order
        run algorithm for solving 1DCSPUL and save the results
        generate new shipment of standard bars and add them to stock
        add UL to stock
    end repeat
    calculate the value of criteria function h_g
    calculate the number of UL in stock (U_g)
    $t \leftarrow t + \Delta$ (Δ is the step for which the threshold t is increased in each experiment)
end while
$t_{opt} = e + (g - 1) \cdot \Delta$
$U_{opt} = U_{g-1}$

Figure 1: The algorithm TOP for selection of t_{opt} and U_{opt}.

t_{opt} and U_{opt} are the result of a sequence of *r* randomly generated orders that are satisfied with item in stock that consists of standard lengths and UL from previous orders that are the non-standard lengths. The t_{opt} and U_{opt} do not provide a minimum of the objective function but the value that is nearly optimal. Proximity to optimal value rest on two factors, the first being the method selected for solving the 1DCPSUL and the second being Δ, which has to be such of a size that the too long computation time is disallowed.

The literature does not provide any specific information about which size of instances may be solved in reasonable time. To verify if the problem can be solved exactly we performed an experiment using the C-CUT algorithm [10]. We were raising the number of

order items at usual *r* and lower *r* and observed whether the optimal solution was reached in a reasonable time. The algorithm could not find an optimal solution if the number of order items was equal to or higher than 21 at usual *r* and equal to or higher than 37 at lower *r*. Therefore it is possible to conclude that a comparable relationship would also be the case when using modern algorithms, which currently enable to process up to 100 order items at usual *r*. Thus the number of order items at lower *r* would be a bit lower than 200. The presented algorithm is able to process approximately up to 900 orders items.

4 RESULTS

To demonstrate the introduced algorithm we analyzed four cases with different ratios between the average bar and item lengths (Table 1). For orders generation we used the problem generator CUTGEN1 [11]. In stock there are two standard lengths (1.000 and 1.100) each consisting of 100 pieces. To highlight the significance of the UL we have set *f* to 2.

Table 1: Parameters for order generation.

	Case 1	Case 2	Case 3	Case 4
Number of different items	20	20	20	20
Interval in which each item is situated	[5, 83]	[6, 146]	[8, 209]	[11, 335]
Number of pieces	125	102	79	34
Number of consecutive orders	30	30	30	30

Instead to generate UL randomly we have used the method of simulation in order to obtain information about the real quantity of UL in stock. With respect to parameter values Δ is set to 5 in Case 1 and to 10 in Cases 2, 3 and 4.

From the results of a proposed algorithm, which are presented in Table 2, it is possible to conclude that the TOP succeeded in finding t_{opt} and U_{opt}.

Table 2: Results of TOP.

Case 1	*t*	20	25	30	35	40	45
	Number of UL in stock	3	3	3	3	3	3
	Value of objective function	1,782	1,745	1,745	907	907	1,750
Case 2	*t*	20	30	40			
	Number of UL in stock	1	1	1			
	Value of objective function	2,857	2,461	2,485			
Case 3	*t*	80	90	100			
	Number of UL in stock	14	1	2			
	Value of objective function	11,380	8,764	9,129			
Case 4	*t*	120	130	140	150	160	
	Number of UL in stock	18	8	6	4	5	
	Value of objective function	18,667	18,295	17,822	16,320	17,527	

Accordingly to the low value of f U_{opt} is relatively low and varies from 1 to 4. t_{opt} diverges from 35, where a minimal value of the objective function is 907, to 150, with a minimum of the objective function 16,320. Increasing value of the objective function from Case 1 to Case 4 can be attributed to decreasing ratio between average bar length and average order length, which makes the cutting problem more difficult to solve.

5 CONCLUSION

We proposed a new for finding optimal threshold t and the optimal number of UL in stock when solving 1DCSPUL. We have described the algorithm into details and tested an introduced method in four cases with different ratio between average bar length and average order length. The method succeeded in prevention of an increased number of UL on account of higher trim-loss in future orders and thus the increasing inventory costs are avoided.

The parameter f is in general dependent on warehouse economics and is not a decision variable. Researchers that are going to conduct further studies in the field of the 1DCSPUL should make testing with the different values of f and observe what impact it would have on the results.

References

[1] Cui, Y., Gu, T., Hu, W., 2009. A cutting-and-inventory control problem in the manufacturing industry of stainless steel wares. Omega, 37(4), pp. 864–875.

[2] Chauhan, S.S., Martel, A., D'amour, S., 2008. Roll assortment optimization in a paper mill: An integer programming approach. Computers & Operations Research, 35(2), pp. 614–627.

[3] Gradišar, M., Jesenko J., Resinovič, G., 1997. Optimization of roll cutting in clothing industry. Computers & Operations Research, 24(10), pp. 945–953.

[4] Venkateswarlu, P., 2001. The trim loss problem in wooden container manufacturing company. Journal of Manufacturing Systems, 20, pp. 166–176.

[5] Gilmore, P.C., Gomory, R.E., 1961. A linear programming approach to the cutting stock-problem. Operations Resarch, 9(4), pp. 849–859.

[6] Gradišar, M., Resinovič, G., Jesenko, J., Kljajić, M. 1999. A sequential heuristic procedure for one-dimensional cutting. European Journal of Operational Research, 114(3), pp. 557–568.

[7] Cherri, A.C., Areales, M.N., Yanasse, H.H., 2009. The one-dimensional cutting stock problem with usable leftover – A heuristic approach. European Journal of Operational Research, 196(3), pp. 897–908.

[8] Cherri, A., Arenales, M.N., Yanasse, H.H., 2013. The usable leftover one-dimensional cutting stock problem-a priority-in-use heuristic. International Transactions in Operational Research, 20(2), pp. 189-199.

[9] Gradišar, M., Erjavec, J., Tomat, L., 2011. One-dimensional cutting stock optimization with usable leftover: a case of low stock-to-order ratio. International Journal of Decision Support System Technology, 3(1), pp. 54-66.

[10] Trkman, P., 2008. Cutting stock process optimization in consecutive time periods (doctoral dissertation). Ljubljana: Ekonomska fakulteta, 161 p.

[11] Gau, T., Wäscher, G., 1995. CUTGEN1: A problem generator for the standard one-dimensional cutting stock problem. European Journal of Operational Research, 84(3), pp. 572–579.

The 12th International Symposium on
Operational Research in Slovenia

SOR '13

Dolenjske Toplice, SLOVENIA
September 25 - 27, 2013

Section II:

Graphs and Their Applications

Mathematical models of discrete acyclic decision processes

Drago Bokal*†‡[1] Tadej Kolmanič⊤, Andreja Smole†[2], Sabina Šmigoc*

* Univerza v Mariboru, Fakulteta za naravoslovje in matematiko,
Koroška cesta 160, 2000 Maribor, Slovenija
† Cosylab d.d., Teslova 30, 1000 Ljubljana, Slovenija
‡ Inštitut za matematiko, fiziko in mehaniko, Jadranska cesta 19, 1000 Ljubljana, Slovenija
⊤ Center odličnosti za biosenzoriko, instrumentacijo in procesno kontrolo,
Tovarniška 26, SI-5270 Ajdovščina, Slovenija
email: drago.bokal@uni-mb.si, tadej.kolmanic@cobik.si, andreja.smole@cosylab.com,
sabina.smigoc@gmail.com

Abstract

We investigate a new niche of applications of operations research: mathematical models of elementary small-scale decision processes applicable to a large quantity of users. With expansion of mobile or embedded devices hosting applications supporting such processes, we expect growing interest in this research direction. Our model formalizes discrete acyclic decision processes as an acyclic digraph equipped with data acquisition, utility evaluation, feasibility, and decision functions at each vertex. We establish conditions on the model and user preferences that allow users to find optimal feasible solutions with no backtracking.

Key words: decision support system, decision process model, acyclic decision process, decision tree, decision digraph.

1 Introduction

Technological development of the past decades has brought up new challenges to operations research community. If the original applications mid-way of the previous century stemmed out of massive scale military applications, growing availability of computing power enabled operations research and decision support system applications to down-scale in resource-complexity from governmental to corporate and small business world, and on the other hand up-scale in model complexity due to the growing availability of computing power. Recent developments in ubiquitous computing [1], embedded computing [12, 13], and internet of things [2] show that, for instance in a cell phone, each person can have at her disposal the computing power that was not available in personal computers a decade ago. This is generating the opportunity and need for operations research community to address personal-scale optimization problems that yield sufficient benefit to the individuals involved to generate interest, yet through massive deployment in personal applications bring justification to costly model development and deployment. Some research in this direction has already been reported in [3, 6, 10, 11].

2 Mathematical model

We address models, applicable to everyday human decisions. They involve several choices, each between several discrete alternatives. Model follows the multi-step decision process developed by Kersten and Szpankowicz in [9], who model agent's decision process as a series of transformations of the world, consisting of the agent, other individuals, and environment data,

[1]D. Bokal was funded through Slovenian Research Agency basic research projects J6-3600 and research programme P1-0297.

[2]A. Smole is funded through the grant Strengthening the R&D departments in SME by Ministry of Economic Development and Technology, Republic of Slovenia, and European Union Funds – European Social Fund.

but is considerably simpler. Contrary to their model that attempts to encompass complete decision process with its context, we focus on stepwise decision process of an individual agent, whose choices depend on gradually increasing information availability. Moreover, information in our model does not change once it is made available. This assumption allows us to consider only acyclic decision processes: with proper structuring of the availability of information, backtracking is not required to re-evaluate the past decisions. The assumption is satisfied in several contexts, where the user follows a sequence of decisions to select the best course of action, which is only executed after all the decisions have been reached on the basis of required information. Further simplification assumes that all the transformations of the world are under control of the agent, i.e. during the decision process, the environment is only involved in the decisions through its constant data, and no other individuals are involved in the decisions.

We model discrete decision processes, assuming that the agent is deciding between a finite list of possible actions. Each action has a certain utility for the agent, which depends on the initially unavailable information about the environment. The finiteness of solution space implies that at each step in the decision process, the agent eliminates some possible solutions either because the new information has made them infeasible, or because they can be proven suboptimal. During the decision process, the list of feasible candidates for optimal solution is either decreasing, leading to termination when this set reduces to a singleton, or insufficient information may available in the environment to select the optimal feasible solution. If this occurs in the decision process, the designer of the process may try to elicit information from the agent that would render some of the remaining solutions either infeasible or inferior to others. Properties that discriminate between solutions can usually be used for this purpose.

Our model shares certain distant similarity with the widely studied multi-attribute utility models initially proposed by Huber in [7] which later evolved into technological applications, such as the decision support system DEX [4, 5]. They both gradually apply the information obtained about the environment or user preferences to reach the final decision, a solution with greatest utility, and both model the decision process as a tree. However, the multi-attribute utility models represent attributes as leaves of the tree, taking them as initial data that is used to gradually compute the utility of certain values of attributes, leading to the final choice of the values of the attributes of the final solution. Internal vertices of the tree therefore represent intermediate utility calculations. In our model, the model is not utility based, but is process based: the user gradually assembles the information required to determine either the feasibility or optimality of various possible solutions that need not share the same set of attributes. Therefore, the structure of our model represents more the classical decision trees used in data classification [8], but with data being acquired during a walk in the tree.

Following the above discussion, we model the set of states in a discrete acyclic decision process as vertices V in an acyclic directed graph D rooted at an initial vertex v_0, in which arcs A represent possible decisions, made by the agent. At each vertex $v \in V$, the agent needs to decide into which of the successors she will move. The data about the environment is represented by a vector $x \in (\mathbb{R} \cup \{\cdot\})^N$. The components of this vector are initially unavailable, having the value $\cdot$. During the decision process, at each vertex $v \in V$, some new components $\beta(v) \subset \{1, \ldots, N\}$ are revealed, so that the component-reduced vector $x/\beta(v)$ changes values from all-$\cdot$ to $\psi(v)/\beta(v)$. Stipulating that $\cdot$ acts as 0 under addition, we denote $x := x + \psi(v)$.

At each vertex $v \in V$, the environment data available consists of the union of all the x-components, acquired on the traversed path $v_0 P v$, i.e. the components $\beta(P) := \bigcup_{u \in V(P)} \beta(u)$. However, for the decision process to be well defined, we need to understand which data is available at v regardless of the path P the agent used to traverse the decision tree from v_0 to v. The set of these components is defined as $B(v) := \bigcap_{P = v_0 P v} \beta(P)$. We further define $\psi(P) := x/\beta(P)$ and $\Psi(v) := x/B(v)$ to be the vectors of actual data values collected along the path P and the vector of data certainly available at vertex v.

Sinks of the digraph D are vertices with no outgoing edges. These vertices represent the solutions among which we need to choose during the decision process. We denote their set by S. For each vertex $v \in V$, we denote by $\Sigma(v) \subseteq S$ the set of sinks, reachable from v by a directed path in D. For every $v \in V$ we denote the set of feasible solutions from vertex v: $\Lambda(v, x) \subseteq S$.

Each solution $s \in S$ has defined a utility function μ_s that translates the environment data x into the utility of s for the agent. The utility function need not use all the components of x; we may assume that it uses at most the components with indices in $B(s)$: these are components certainly available when the agent chooses s as the final solution.

At each vertex v, a discrete decision needs to be taken among the successors of v. This decision is modelled by a function δ_v that needs to be consistent with the digraph, i.e. $v\delta_v(\Psi(v)) \in A$ for every $v \in V$.

For the purpose of studying optimality of acquired solution, we introduce the following conditions:

1. Consistency of transitions: codomain of the function δ_v has to be a subset of $A(v)$. For every y and v we have: $v\delta_v(y) \in A$.

2. Consistence with utility functions: at each step, if a solution $s \in S$ is present in $\Sigma(v)$ but not in $\Sigma(\delta_v(\Psi(v)))$, then either s is infeasible or s is suboptimal for $\Psi(v)$.

3. Convergence: for each pair $s, s' \in S$, at some point in the decision process, one of them becomes suboptimal or infeasible.

3 The algorithm and its correctness

With Algorithm 1, we find an optimal solution in the decision digraph. Optimal solution is the most suitable solution for the agent, according to her answers during the algorithm implementation. At the preparation of the digraph, we have to carefully choose questions for every vertex: at each step the number of feasible solutions has to reduce. At the same time, questions have to be clearly stated, so that we prevent unwanted deviations from optimal solution because of agent's potential misunderstanding of the questions.

The algorithm is equivalent to the evaluation of decision trees (cf. [8]), but is adapted to acyclic digraphs.

Theorem 3.1 *Assuming the conditions* $1. - 3.$*, in every step* i *of the algorithm 1, an optimal solution is always in a set of reachable solutions of vertex* u_i*, and the algorithm stops in this optimal solution.*

Proof. We assume that conditions $1. - 3.$ are satisfied, and proove the theorem by mathematical induction on the number of steps denoted by i: for $i = 0$ the theorem holds, as we are at the root vertex v_0, from where all the solutions are reachable, so is the optimal solution. We assume, that theorem holds at step i. Because of the consistency of transitions (condition 1.), we can move to vertex $\delta_{u_i}(y_i)$, so the algorithm correctly follows the arcs of the digraph.

Let s^* be an optimal solution. The induction hypotesis implies $s^* \in \Sigma(u_i) \cap \Lambda(u_i, x_i)$. Suppose $s^* \notin \Sigma(u_{i+1}) \cap \Lambda(u_{i+1}, x_{i+1})$. Then either $s^* \notin \Sigma(u_{i+1})$, implying s^* is not reachable, or $s^* \notin \Lambda(u_{i+1}, x_{i+1})$, implying s^* is not feasible. Second condition is in contradiction with optimality of s^*, because an optimal solution is always feasible. So there is $s^* \notin \Sigma(u_{i+1})$. Because of the consistence with utility functions follows, because $s \in \Sigma(u_i)$, that s^* is whether infeasible (again contradiction) whether is suboptimal (also contradiction). So we conclude that at each step, s^* is reachable and feasible.

Algorithm 1 Finding the optimal solution

```
//We set value of step i to 0.
i := 0
//We start in an initial vertex, u_i represents a vertex, in which we are located in step i.
u_i := v_0
//Components of vector x are, at the beginning, all ·.
x := (·, ·, ..., ·)
//Decision process takes place while vertex u_i is not a sink.
while u_i ∉ S do
  //In every step i of the decision process, we change values of vector x_{i-1} with ψ(u_i).
  x_i := x_{i-1} + ψ(u_i)
  //We reduce vector x to vector of data values certainly available at vertex u_i.
  y_i := x/B(u_i) := Ψ(u_i)
  //Function δ_v determines a successor u_{i+1} to vertex u_i. Decision is based on the data
  values in vector y_i.
  u_{i+1} := δ_v(y_i)
  //We increase step i by 1, meaning agent moves into the next vertex.
  i++
  //We end while loop.
end while
//Returns vertex u_i which represents a sink and an optimal solution.
return u_i
```

We further need to prove that algorithm stops in an optimal solution. Suppose it does not. Because the graph is finite, the algorithm stops in one of the vertices, from which there are multiple reachable solutions. This is in contradiction with convergence: during the execution of the process, one of the solutions of each pair should have become suboptimal or infeasible. □

4 Elimination of suboptimal solutions

At a given vertex $v \in V$, the suboptimality of a vertex is easily verified whenever there are two vertices $s, s' \subseteq S$ with available all required data, i.e. $B(s), B(s') \subseteq B(v)$. However, using ideas from branch-and-bound technique, suboptimality can be verified also if only some of the components $B(s), B(s')$ are available in $B(v)$. For each such vertex, the function μ_s is optimized over the subspace of all unavailable components, yielding an upper and lower bound for μ_s. If the corresponding intervals for s and s' are disjoint, one of the solutions is suboptimal.

The suboptimality condition can be used to aid the agent in the decision process, letting her choose only among those successors of a given vertex v that cannot be proven suboptimal. If the tree has a certain structure that is yet being investigated, then the choice of successors can be fully automated, at least at certain vertices.

With functions ν^+ and ν^- we determine, in every vertex, boundary values of intervals that represents utilities. Supremum represents an upper bound of the interval: $\nu^+(s_j, x) := \sup_{y \in \mathbb{R}^N} \mu_s((x/B(u_i)) + y/(B(s_j) \setminus B(u_i)))$. Infimum represents the lower bound of an interval: $\nu^-(s_j, x) := \inf_{y \in \mathbb{R}^N} \mu_s((x/B(u_i)) + y/(B(s_j) \setminus B(u_i)))$. Using these intervals, we can determine which solution is suboptimal. Then we can narrow a set of sensible choices in $A(s)$, as in some of them, we don't learn any new useful information that would influence on the selection of final

solution: $u \in \Sigma(u_{i-1}) \setminus \Sigma(u_i) \iff ((u \in \Lambda(u_{i-1}, x_{i-1}) \setminus \Lambda(u_i, x_i)) \vee (\exists u' \in \Sigma(u_i) : \nu^-(u', u_i) \geq \nu^+(u, u_i)))$. If in $A(s)$ only one choice remains in $A(v)$, then we can proceed to the next vertex.

The idea of applying the suboptimality verifications is sketched in Algorithm 2.

Algorithm 2 Elimination of suboptimal solutions

//Number i is the number of current step in outer loop of Algorithm 1.
//Set S_i is the set of solutions, reachable from u_i, and k is its cardinality.
//For every element $j \in S_i$ calculate upper M_j and lower m_j bound of utility.
//Value M is the largest lower bound of utility.
//K is the set of indices of all feasible solutions, for which upper bound is lower than the largest lower bound.
$K := \{j \in \{1, \ldots, k\} | M_j \geq M\} \cap \Lambda(u_i, x_i)$
//We find a successor, to which it is reasonable to move.
//This is a successor from which all solutions in K are reachable.
//If there are more such successors, then we choose the one among them with smallest set of reachable solutions.

Problem 4.1 *How can we adapt a mathematical model and Algorithm 2, so that Algorithm 2 will meet a condition of consistency with utility functions?*

Consideration: Assume that algorithm would meet a condition of consistency with utility functions. Let there be a solution $s \in \Sigma(u_i)$, but $s \notin \Sigma(u_{i+1})$. Then $s \notin K$, which means that whether $s \notin \Lambda(u_i, x_i)$ whether $M_s < M$. If $s \notin \Lambda(u_i, x_i)$, then s is infeasible. If $M_s < M$, then s is suboptimal because of some other solution. If this other solution is feasible, then we can discard s, as does Algorithm 2. If this other solution is infeasible because of the information, that we will acquire later, then we can't discard s, since it can become optimal when s becomes infeasible. This problem indicates, that is good to structure trees in a way, that space of feasible solutions is limited first, data acquisition, that determine utilities comes later. In such case, from a certain step further, no solution becomes infeasible due to it's properties, rather because of agent's preferences.

5 Discussion and further research

The approach adopted by our models is somewhat different from the classical optimization approaches of operations research paradigm. We focus on the user, who is following the process of steps gradually expressing his preferences and contributes data required to evaluate the final solutions of the decision process. This agent is following a predefined decision digraph like in a depth-first search, but unlike computer, the user does not have the patience to visit or evaluate all candidate solutions. Therefore our approach is that we assume the user will only follow the decision process till she reaches the first solution, and this will be the solution implemented. We need to understand the conditions, under which this elementary algorithm performed by the user will indeed reach an optimal solution. These conditions will then need to be considered by the designer of the model.

For further research, we intend to investigate possibilities of blending our decision-based model with utility based multi-attribute decision model [7, 4], considering to take the best out of both worlds. The synergy between the two models can result in methodology or even algorithms for generation of user-based decision models from known multi-attribute models. We intend to investigate the inverse direction, too: from analysis of agents' steps using decision-based models, relevant information on user preferences or utility functions can likely be derived. Another possible research direction presents itself by integrating the decision based model of

a single agent with distributed models [3]. These can be applicable to decision systems in distributed environments, and are applicable to groups of agents involved, for instance, in some social network.

References

[1] G. D. Abowd, E. D. Mynatt, Charting past, present, and future research in ubiquitous computing, *ACM Trans. Comput.-Hum. Interact.* 7 (2001), 29–58.

[2] L. Atzori, A. Iera, G. Morabito, The Internet of Things: A survey, *Comput. Networks* 54 (2010), 2787–2805.

[3] J. P. Barthélemy, R. Bisdorff, G. Coppin: Human centered processes and decision support systems, *Eur. J. Oper. Res.* 136 (2002), 233–252.

[4] M. Bohanec, V. Rajkovič, Multi-Attribute Decision Modeling: Industrial Applications of DEX, *Informatica (Lj.)* 23 (1999), 487–491.

[5] M. Bohanec, B. Zupan, V. Rajkovič, Applications of Qualitative Multi-Attribute Decision Models in Health Care, *Int. J. Med. Inform.* 58 (2000), 191–205.

[6] G. Coppin, A. Skrzyniarz: Human-centered processes: individual and distributed decision support, *IEEE Intell. Syst.* 18 (2003), 27–33.

[7] G. P. Huber: Multi-attribute utility models: a review of field and field-like studies, *Manage. Sci.* 20 (1974), 1393–1402.

[8] S. R. Savafian, D. Szpakowicz, A Survey of Decision Tree Classifier Methodology, *IEEE Trans. Systems, Man, & Cybernetics* 21 (1991), 660–674.

[9] G. E. Kersten, S. Szpakowicz, Decision making and decision aiding: defining the process, its representations, and support, *Group Decis. Negot.* 3 (1994), 237–261.

[10] M. Öztürk, A. Tsoukiás, Modelling uncertain positive and negative reasons in decision aiding, *Decis. Support Syst.* 43 (2007), 1512–1526.

[11] A. Tsoukiás, Decision Support From decision theory to decision aiding methodology, *Eur. J. Oper. Res.* 187 (2008), 138–161.

[12] M. Wolf, Computers as components, Principles of Embedded Computing System Design, third edition, Morgan Kaufmann Publishers, Waltham, 2012.

[13] W. Wolf, Cyber-physical systems, *IEEE Computer* 42 (2009), 88–89.

Stackelberg Shortest Path Tree Game, Revisited[1]

Sergio Cabello
Department of Mathematics, FMF, University of Ljubljana, Slovenia
sergio.cabello@fmf.uni-lj.si

Abstract: Let $G(V, E)$ be a directed graph with n vertices and m edges. The edges E of G are divided into two types: E_F and E_P. Each edge of E_F has a fixed price. The edges of E_P are the priceable edges and their price is not fixed a priori. Let r be a vertex of G. For an assignment of prices to the edges of E_P, the revenue is given by the following procedure: select a shortest path tree T from r with respect to the prices (a tree of cheapest paths); the revenue is the sum, over all priceable edges e, of the product of the price of e and the number of vertices below e in T.

Assuming that $k = |E_P| \geq 2$ is a constant, we provide a data structure whose construction takes $O(m + n \log^{k-1} n)$ time and with the property that, when we assign prices to the edges of E_P, the revenue can be computed in $(\log^{k-1} n)$. Using our data structure, we save almost a linear factor when computing the optimal strategy in the Stackelberg shortest paths tree game of [D. Bilò and L. Gualà and G. Proietti and P. Widmayer. Computational aspects of a 2-Player Stackelberg shortest paths tree game. Proc. WINE 2008].

Keywords: pricing networks, Stackelberg model, shortest paths, orthogonal range searching.

1 Introduction

A *Stackelberg game* is an extensive game with two players and perfect information in which the first player, the *leader*, chooses her action and then the second player, the *follower*, informed of the leader's choice, chooses her action. In a *Stackelberg pricing game in networks*, the leader owns a subset of the edges in a network and has to choose the price of those edges to maximize its revenue. The other edges of the network have a price already fixed. The follower chooses a subnetwork of minimum price with a prescribed property, like for example being a spanning tree or spanning two vertices. The revenue of the leader is determined by the prices of the edges that the follower uses in its chosen subnetwork, possibly combined with the amount of use of each edge.

Stackelberg network pricing games were first studied by Labbé et al [7] when the follower is interested in a cheapest path connecting two given vertices. They showed that even such "simple" problem is NP-hard when the number of priceable edges is not bounded. There has been much follow up research; we refer the reader to the overview by van Hoesel [10]. The case when the follower is interested in a cheapest spanning tree was introduced by Cardinal et al. [6]. Bilò et al. [2] considered the case when the follower is interested in a shortest path tree from a prespecified root r and the revenue of a priceable edge is the product of its price and the number of times such edge is used by paths from r in the tree. This is the model we will consider. We next provide the formal model in detail and explain our contribution.

The shortest path tree game. We next provide a description of the Stackelberg shortest path tree game. In fact, we present it as an optimization problem, which we denote by STACKSPT. The input consists of the following data:

- A directed graph $G = (V, E)$ with n vertices and m edges.

[1]This work has been partially financed by the Slovenian Research Agency, program P1-0297, project J1-4106, and within the EUROCORES Programme EUROGIGA (project GReGAS) of the European Science Foundation. Full version available at http://arxiv.org/abs/1207.2317.

- A partition of the edges E into $E_F \cup E_P$. The edges of E_P are the ***priceable*** edges and the edges of E_F are the ***fixed-cost*** edges.
- A root $r \in V(G)$.
- A ***demand*** function $\phi : V(G) \to \mathbb{R}_{\geq 0}$, where $\phi(v)$ tells the demand of vertex v.
- A ***cost*** function $c : E_F \to \mathbb{R}_{>0}$ fixing the price of the edges in E_F.

An example is given in Figure 1. A feasible solution is given by a ***price function*** $p : E_P \to \mathbb{R}_{>0}$. The cost function c and the price function p define a weight function $w_p : E \to \mathbb{R}_{\geq 0}$ over all edges by setting $w_p(e) = p(e)$ if $e \in E_P$ and $w_p(e) = c(e)$ if $e \in E_F$. This weight function defines shortest paths in G. (In fact, they should be called cheapest paths in this context.)

For a price function p and a path π, the revenue per unit along π is

$$\rho_u(\pi, p) \ := \sum_{e \in E_P \cap E(\pi)} p(e).$$

Note that only priceable edges contribute to the revenue. Let T be a subtree of G containing paths from r to all vertices. For any vertex $v \in V(G)$, let $T[r, v]$ denote the path in T from r to v. The ***revenue*** given by T is

$$\rho(T, p) \ := \sum_{v \in V(G)} \phi(v) \cdot \rho_u(T[r, v], p).$$

We would like to tell that the revenue given by the price function p is $\rho(T, p)$, where T is a shortest path tree from r with respect to w_p. However, there may be different shortest path trees T with different revenues. In such case, T is taken as the shortest path tree that maximizes the revenue. Although this assumption may seem counterintuitive at first glance, it forces the existence of a maximum and avoids the technicality of attaining revenues arbitrarily close to a value that is not attainable. Thus, the revenue of a price function p is defined as

$$\rho(p) \ := \ \max\{\rho(T, p) \mid T \text{ a shortest path tree in } G \text{ with respect to } w_p\}. \tag{1}$$

As an optimization problem, STACKSPT consists of finding a price function p such that the revenue $\rho(p)$ is maximized.

From the point of view of game theory, the leader chooses the price function p and the follower chooses a tree T containing paths from r to all vertices. The payoff of the leader is $\rho(T, p)$. The payoff of the follower is the sum, over all vertices v of G, of the distance in T from r to v. Among trees T with the same payoff for the follower, she maximizes the revenue $\rho(T, p)$. Thus, the follower uses a lexicographic order where, as primary criteria, lengths are minimized, and, as secondary criteria, revenue is maximized.

Our result and comparison. We assume henceforth that $k := |E_P| \geq 2$ is a constant. For $k = 1$, STACKSPT can be solved in $O(m + n \log n)$ time as discussed by Bilò et al [2].

We describe a data structure that can be constructed in $O(m + n \log^{k-1} n)$ time and with the property that, given a price function p, the revenue $\rho(p)$ can be computed in $O(\log^{k-1} n)$ time. Bilò et al. [2] show how to find an optimal price function p by evaluating the revenue of $O(n^k)$ price functions[2]. Combined with our data structure, we can then find an optimal price function in $O(m + n^k \log^{k-1} n)$ time.

[2]They only discuss the case when the demand function ϕ is identically 1. However, their discussion can be easily adapted to more general demand functions.

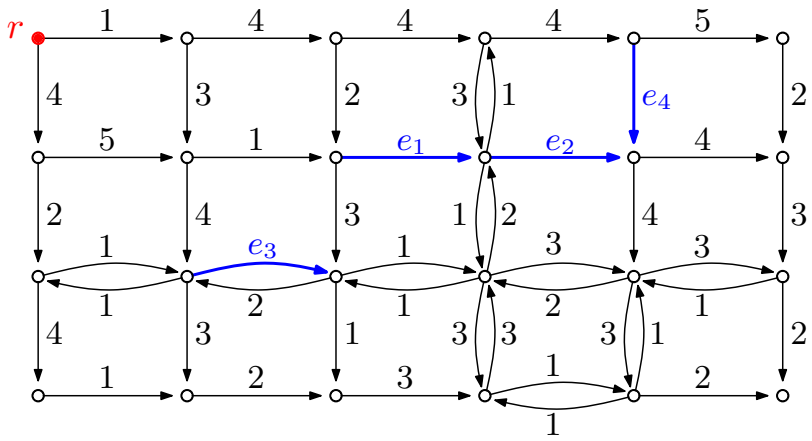


Figure 1: An example of a Stackelberg shortest path tree game. We assume that each vertex has unit demand.

Our result matches the result of Bilò et al. [2] for the case $k = 2$. For $k \geq 3$, the algorithm of Bilò et al. uses $O(n^k(m + n \log n))$ time. A previous algorithm by van Hoesel et al. [11] to compute the optimal solution in a more general Stackelberg pricing problem, where paths from different sources have to be considered, reduces StackSPT to $O(n^{4^k})$ linear programs of constant size.

The large dependency on k is unavoidable because the problem is NP-hard for unbounded k. Briest et al. [4] provide an approximation algorithm for more general Stackelberg network pricing games. When it is specialized to StackSPT, it provides a $O(\log n)$-approximation.

Our data structure is based on three main ideas:

- A careful rule to break ties when there are multiple shortest path trees. With this rule, we can easily split the vertices into groups that use the same priceable edges.
- Using a smaller network, of size $O(k^2)$, such that, for a given price function, we can find out the structure of the priceable edges in the shortest path tree of the network. This idea is similar to the *shortest paths graph model* of Bouhtou et al. [3].
- Mapping each vertex of the network to a point in Euclidean k-dimensional space in such a way that the vertices that use a certain subset of the priceable edges can be identified as a subset of points in a certain octant. This allows us to use efficient data structures for range searching. Similar ideas have been used for graphs of bounded treewidth; see [1, 5, 8] and [9, Chapter 4].

Notation. We use $e_1, e_2, \ldots, e_k$ to denote the edges of E_P. The enumeration of the edges is fixed; in fact we will use it to break ties. For a subset of vertices $U \subseteq V(G)$ we use the notation $\phi(U) := \sum_{u \in U} \phi(u)$. For a subset of priceable edges $F \subseteq E_P$ we use the notation $p(F) := \sum_{e \in F} p(e) = \sum_{e \in F \cap E_F} p(e)$.

2 Breaking Ties

Evaluating the revenue of a price function is easier in a *generic case*, when there is a unique shortest path from r to each vertex of $V(G)$. In contrast, in the *degenerate case*, there is at least one vertex v with two distinct shortest paths from r to v. Unfortunately, the price functions

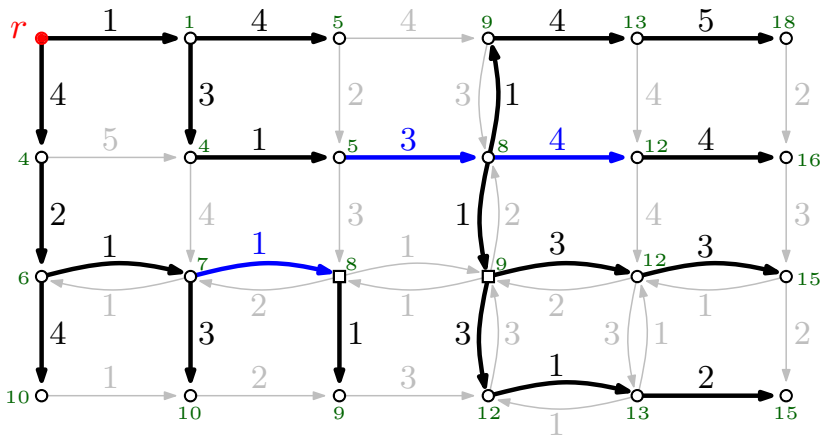


Figure 2: A $\widehat{w}_p$-shortest path tree for the price function $p(e_1) = 3$, $p(e_2) = p(e_4) = 4$, $p(e_3) = 1$ in the network of Figure 1. The values in the vertices are the distance from r. Note that there are some vertices, like for example the two that are marked with squares, for which there are different shortest paths using different priceable edges, so we have to select shortest paths maximizing revenue. The revenue given by this tree, if each vertex has unit demand, is $p(e_1) \cdot 10 + (p(e_1) + p(e_2)) \cdot 2 + p(e_3) \cdot 2 = 46$ units.

defining the optimum are degenerate. This is easy to see because, in a generic case, a slight increase in the price function leads to a slight increase in the revenue.

In our approach, we will count how many vertices use a given sequence of priceable edges. For this to work, we need a systematic way to break ties, that is, a rule to select, among the shortest path trees that give the same revenue, one. We actually do not go that far, and only care about the priceable edges on the paths of the tree.

We first discuss how to break ties among shortest paths, and then discuss how to break ties among shortest path trees. Essentially, we compare paths lexicographically according to the following: firstly, we compare paths by length; secondly, if they have the same length, we compare them by revenue; finally, if they have the same length and revenue, we compare the priceable edges on the path lexicographically, giving preference to priceable edges of larger index. Mathematically this is handled assigning a triple $\widehat{w}_p(\pi) \in \mathbb{R}^3$ to each path. We say that a path π is $\widehat{w}_p$-shorter than a path π' if and only if $\widehat{w}_p(\pi) \preceq \widehat{w}_p(\pi')$, where $\preceq$ denotes the lexicographic order. Details are provided in the full version.

The weights $\widehat{w}_p$ can be used to define $\widehat{w}_p$-shortest paths:

$$\pi \text{ from } u \text{ to } v \text{ is } \widehat{w}_p\text{-shortest} \iff \forall \text{ paths } \pi' \text{ from } u \text{ to } v : \ \widehat{w}_p(\pi) \preceq \widehat{w}_p(\pi').$$

A tree T is a ***$\widehat{w}_p$-shortest path tree*** (from r) if it contains a $\widehat{w}_p$-shortest path from r to each vertex. See Figure 2 for an example. A $\widehat{w}_p$-shortest path tree can be computed be computed in $O(m + n \log n)$ time using Dijkstra's algorithm with the weights $\widehat{w}_p$ and lexicographic comparison. (Here we need that k is a constant, which implies that two $\widehat{w}_p$-lengths can be compared in constant time. For general k, the running time of Dijkstra's algorithm may get an additional dependence on k, depending on the model of computation.) Note that there may be several $\widehat{w}_p$-shortest path trees because of different shortest paths without priceable edges. We next argue that $\widehat{w}_p$-shortest path trees give the revenue of the price function.

Lemma 1. *If T be a $\widehat{w}_p$-shortest path tree, then $\rho(T, p) = \rho(p)$.*

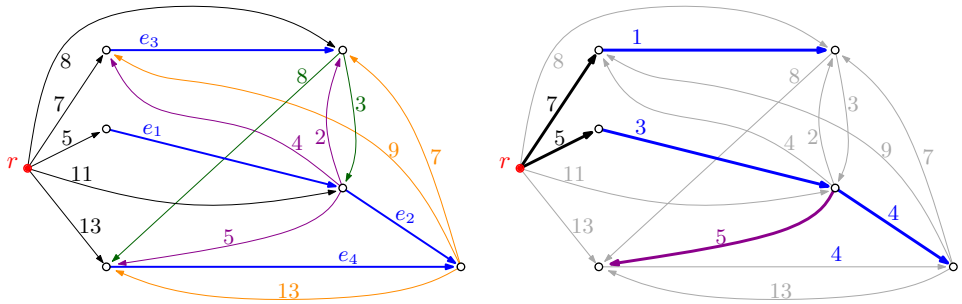


Figure 3: Left: The model graph for the network of Figure 1. Edges with infinite weight, like for example $r \to t_4$ or $t_1 \to t_2$, are not drawn. Right: the $\widehat{w}_p$-shortest path tree in the model for the price function of Figure 2: $p(e_1) = 3$, $p(e_2) = p(e_4) = 4$, and $p(e_3) = 1$.

3 Reduced trees and sequences of priceable edges

Consider a price function p. Let T be a $\widehat{w}_p$-shortest path tree from r. The $\widehat{w}_p$***-reduced tree*** RT is obtained from T by contracting all the fixed-cost edges $E_F \cap E(T)$. The resulting graph is a tree with edge set $E_P \cap E(T)$. When considering RT, we disregard the prices p and the orientation of the edges, and consider it as a rooted, unweighted, undirected graph with distinct labels $e_1, \ldots, e_k$ on its edges. In general, we will use RH to denote the reduced graph obtained from a graph H by contracting all non-priceable edges. The $\widehat{w}_p$-reduced tree for the example of Figure 2 contains the edges e_1 and e_3 adjacent to r and the edge e_2 below e_1.

We first show that the $\widehat{w}_p$-reduced trees are independent of the $\widehat{w}_p$-shortest path tree that is used. A useful consequence of this is that any two $\widehat{w}_p$-shortest path trees have the same subset of priceable edges.

Lemma 2. *If T and T' are $\widehat{w}_p$-shortest path trees, then $RT = RT'$.*

We have to compute the $\widehat{w}_p$-reduced tree for several different prices. We next provide a data structure to compute such reduced trees without looking at the whole graph each time. For this we use the *model graph* $\tilde{G} = \tilde{G}(G, E_P, c, r)$, defined as follows. The vertex set of $\tilde{G}$ consists of r and the endpoints of the priceable edges. Thus $V(\tilde{G}) = \{r\} \cup \{s_1, t_1, \ldots, s_k, t_k\}$. In $\tilde{G}$, we have edges from r to any other vertex. Furthermore, for each priceable edges e_i and e_j, $i \neq j$, we have an edge from t_i to s_j and to t_j. Finally, we have the edges $e_1, \ldots, e_k$ themselves. Each edge $u \to v$ in $E(\tilde{G})$ gets weight equal to the distance between u and v in $G - E_P$. This finishes the description of the model graph $\tilde{G}$. See Figure 3, left, for an example. This construction is similar to and inspired by the *shortest paths graph model* of Bouhtou et al. [3]. $\widehat{w}_p$-reduced trees in the model graph correspond to $\widehat{w}_p$-reduced trees in the original graph. This is the key observation to obtain the following result.

Lemma 3. *In $O(m + n \log n)$ time we can construct a data structure with the property that, for any given price function p, we can compute in $O(1)$ time the $\widehat{w}_p$-reduced tree RT.*

4 Data structure for computing the revenue

Consider a price function p and let T be a $\widehat{w}_p$-shortest path tree. For each edge $e_i \in E_P$, let $V_T(e_i, p)$ be the set of vertices with the property that e_i is the last edge of E_P used by $T[r, v]$. It may be that $V_T(e_i, p) = \emptyset$. In particular this happens when e_i does not appear in the shortest

path tree T. One can argue that $V_T(e_i, p)$ is independent of the choice of T, so we just denote it by $V(e_i, p)$.

Lemma 4. *Let p be a price function, let R be its $\widehat{w}_p$-reduced tree, and let $\sigma(e_i, R)$ be the sequence of priceable edges in the path from the root to e_i in R. The revenue given by p is*

$$\rho(p) \;=\; \sum_{e_i \in E(R)} p(\sigma(e_i, R)) \cdot \phi(V(e_i, p)).$$

Our objective is to compute $\phi(V(e_i, p))$ efficiently using data structures for orthogonal range searching. In orthogonal range searching we preprocess a weighted set of points in $\mathbb{R}^d$ such that the sum of the weights of the points inside a query rectangle can be retrieved efficiently. We use the data structure of Willard [12]. The key idea is to map each vertex of G to a point whose coordinates are described by graph distances. We omit the details.

Lemma 5. *Assume that $k \geq 2$ is a constant. In time $O(m + n \log^{k-1} n)$ we can construct a data structure with the following property: given a price function p we can obtain $\phi(V(e_i, p))$ in $O(\log^{k-1} n)$ time.*

Theorem 6. *Assume that $k \geq 2$ is a constant. Consider an instance to $StackSPT$ with n vertices, m edges, and k priceable edges. In time $O(m + n \log^{k-1} n)$ we can construct a data structure with the following property: given a price function p, the revenue $\rho(p)$ can be obtained in $O(\log^{k-1} n)$ time.*

Corollary 7. *Let $k \geq 2$ be a constant. The problem* STACKSPT *with n vertices, m edges, and k priceable edges can be solved in $O(m + n^k \log^{k-1} n)$ time.*

References

[1] B. Ben-Moshe, B. K. Bhattacharya, and Q. Shi. Efficient algorithms for the weighted 2-center problem in a cactus graph. In *ISAAC 2005*, volume 3827 of *LNCS*, pages 693–703, 2005.

[2] D. Bilò, L. Gualà, G. Proietti, and P. Widmayer. Computational aspects of a 2-player Stackelberg shortest paths tree game. In *WINE 2008*, volume 5385 of *LNCS*, pages 251–262, 2008.

[3] M. Bouhtou, S. P. M. van Hoesel, A. F. van der Kraaij, and J.-L. Lutton. Tariff optimization in networks. *INFORMS Journal on Computing*, 19(3):458–469, 2007.

[4] P. Briest, M. Hoefer, and P. Krysta. Stackelberg network pricing games. *Algorithmica*, 62(3-4):733–753, 2012.

[5] S. Cabello and C. Knauer. Algorithms for graphs of bounded treewidth via orthogonal range searching. *Computational Geometry: Theory and Applications*, 42(9):815–824, 2009.

[6] J. Cardinal, E. D. Demaine, S. Fiorini, G. Joret, S. Langerman, I. Newman, and O. Weimann. The Stackelberg minimum spanning tree game. *Algorithmica*, 59(2):129–144, 2011.

[7] M. Labbé, P. Marcotte, and G. Savard. A bilevel model of taxation and its application to optimal highway pricing. *Management Science*, 44:1608–1622, 1998.

[8] Q. Shi. Single facility location problems in partial k-trees, 2005. Poster at MITACS, Canada.

[9] Q. Shi. *Efficient Algorithms for Network Center/Covering Location Optimization Problems*. PhD thesis, School of Computing Science, Simon Fraser University, 2008.

[10] S. P. M. van Hoesel. An overview of Stackelberg pricing in networks. *European Journal of Operational Research*, 189(3):1393–1402, 2008.

[11] S. P. M. van Hoesel, A. F. van der Kraaij, C. Mannino, M. Bouhtou, and G. Oriolo. Polynomial cases of the tarification problem. Research Memoranda RM03063, METEOR, 2003.

[12] D. E. Willard. New data structures for orthogonal range queries. *SIAM J. Comput.*, 14:232–253, 1985.

PRACTICAL PLACEMENT OF TRAINEE TEACHERS TO SCHOOLS

Katarína Cechlárová[a], Tamás Fleiner[b] and Eva Potpinková[a]
Institute of Mathematics, Faculty of Science, P.J.Šafárik University, Jesenná 5, 040 01 Košice, Slovakia
katarina.cechlarova@upjs.sk, eva.potpinkova@student.upjs.sk
Department of Computer Science and Information Theory, Budapest University of Technology and Economics, Magyar tudósok körútja 2, H-1117 Budapest, Hungary
fleiner@cs.bme.hu

Abstract. Several countries successfully use centralized matching schemes for assigning students to colleges or newly-qualified graduates to their first career. In this paper we explore the computational aspects of a possible similar scheme for assigning trainee teachers to schools. The special feature of this model is that each teacher specializes in two subjects that have to be taught in the same school. We show that the model becomes intractable even under several strict restrictions concerning the total number of subjects and the number of acceptable schools each teacher is allowed to list.

Keywords: assignment of students, bipartite matching, algorithm, NP-completeness

1 INTRODUCTION

The traditional study of teachers-to-be in Slovakia involves the specialization of each student in two subjects, e.g. Mathematics and Physics, Chemistry and Biology, Slovak language and English etc. In addition to the study of the various topics of these subjects, principles of Pedagogics and Psychology, each curriculum contains a practical placement in a real school several times during the study. Students might try to find suitable schools by themselves, but to ensure the quality of such a placement, the faculties require that in each school a student is supervised by a qualified and experienced teacher who is approved by the faculty for taking this responsibility. Hence it is often the case that the faculty provides a list of such schools and the students may choose from the list.

The assignment is often performed on a first-come-first-served basis. However, not all schools provide supervisors for all subjects, or they may not have enough classes to accept several students for a particular subject. This might be a serious problem, as a student is usually required to follow both his/her subjects in the same school (even if each subject is supervised by a different teacher, placement at two different schools might be infeasible for example because of the school time table and time commitment required for travelling). So it might happen that for some unlucky students no place remains, or they might be forced to go to a school that is located neither in the town of their residence nor of the faculty, thus increasing their costs above an acceptable level.

The aims of this paper is to study the computational complexity of the trainee teachers assigning problem. We propose efficient algorithms that allocate all applicants to acceptable schools or decide that such an allocation is impossible for several special cases of the problem, as follows: (i) if there are altogether only 2 specialization subjects, or (ii) if there are 3 subjects but each school can accept at most 1 students for each subject (irrespectively of her other specialization), or, (iii) without the restriction on the number of specialization subjects, if each applicant is allowed to list at most two acceptable schools and each school has at most one place for each specialization. By contrast, we show that the problem to decide whether a full assignment exists is NP-complete if there are 3 subjects and schools may have capacity 2 in one of its subjects, or if there are 4 subjects and each school has capacity at most 1 in each subject.

2 RELATED WORK

The classical problems of combinatorial optimization like the maximum cardinality bipartite matching problem, assignment problem, or flow problem have successfully been applied to various variants of manpower allocation problems (see e.g. applications reviewed [3], Chapter 12). Practical situations have lead also to some NP-complete variants [9]. Recently, a lot of attention has been attracted by several large-scale centralized allocation schemes used for assigning pupils to public schools in Boston and New York [1], [2], assigning graduates of medical schools to their first jobs in hospitals in the USA [13], [14], university applicants to study places in Hungary [5] etc. In such schemes, the applicants as well as schools, in addition to simply stating acceptability, are also required to order the other side of the 'market' according to their preferences. For an overview of other applications, various models and their computational complexity, the reader is adviced to consult the recently published monograph by David Manlove [12] or the comprehensive web page containing a decription of matching practices for various levels of education in many European countries [15].

Of the models studied so far the closest to our situation are the so called hospital-residents problem with couples: members of a married couple wish to go to a pair of geographically close hospitals [8], or even refuse to be separated and insist on going to the same institution [11]. Another case is the Scottish scheme for medical students that have to be assigned to two training units (medical and surgical one), however, these two assignments have to be allocated to two differents semesters [10]. Our model differs from all ones presented so far due to the applicants specialization, the necessity to teach both subjects in the same school and schools allowed to have different capacities for different subjects.

3 DEFINITION

An instance *J* of the Teachers Assignment Problem, TAP for short, involves a set *A* of applicants, a set *S* of schools and a set *P* of subjects. For ease of exposition, elements of the set *P* will sometimes be referred to by letters like *M*, *F*, *I* or *B*, to remind of real subjects taught at schools, like Mathematics, Physics, Informatics or Biology etc.

Each applicant $a \in A$ is characterized by a pair of different subjects $\boldsymbol{p}(a) = \{p_1(a), p_2(a)\} \subseteq P$. Sometimes we shall also say that a particular applicant is of type *MF*, *MB*, or *IB*, etc.

Each school $s \in S$ has a certain capacity for each subject, the vector of capacities will be $\boldsymbol{c}(s) = \left(c_1(s), \dots, c_{|P|}(s)\right) \in \mathbb{N}^{|P|}$, an entry of $\boldsymbol{c}(s)$ will also be referred to as a *partial capacity* of school *s*. Here, $c_p(s)$ is the maximum number of students whose specialization involves subject p that school s is able to accept. Again, we shall sometimes write $c_M(s), c_I(s)$ etc.

A school *s* is compatible with applicant *a* if $c_p(s) \geq 1$ for both subjects $p \in \boldsymbol{p}(a)$. We suppose that each applicant *a* provides a list $S(a)$ of *acceptable* schools, i.e. schools to which he/she willing to go. An *assignment* $\mathcal{M}$ is a subset of $A \times S$ such that each applicant $a \in A$ is a member of at most one pair in $\mathcal{M}$. We shall write $\mathcal{M}(a) = s$ if $(a, s) \in \mathcal{M}$ and say that applicant *a* is *assigned* (to school *s*); if there is no such school, applicant *a* is *unassigned*. The set of applicants assigned to a school *s* will be denoted by $\mathcal{M}(s) = \{a \in A; (a, s) \in \mathcal{M}\}$. We shall also denote by $\mathcal{M}_p(s)$ the set of applicants assigned to *s* whose specialization includes subject *p* and by $\mathcal{M}_{p,r}(s)$ the set of applicants assigned to *s* whose specialization is exactly the pair $\{p, r\}$.

More precisely,

$$\mathcal{M}_p(s) = \{a \in A; (a,s) \in \mathcal{M} \ \& \ p \in \boldsymbol{p}(a)\}$$

and

$$\mathcal{M}_{p,r}(s) = \{a \in A; (a,s) \in \mathcal{M} \ \& \ \{p,r\} = \boldsymbol{p}(a)\}.$$

An assignment $\mathcal{M}$ is *feasible* if $\mathcal{M}(a) \in S(a)$ for each $a \in A$ and $|\mathcal{M}_p(s)| \leq c_p(s)$ for each school *s* and each subject *p*.

Example. Suppose there are 3 subjects *M,F* and *I*, four applicants $a_1$of type *IF*, a_2 of type *MF* and a_3, a_4 of type *MI*. There are two schools s_1, s_2 with $c_M(s_1) = 1, c_F(s_1) = c_I(s_1) = 2$ and $c_M(s_2) = 2, c_F(s_2) = c_I(s_2) = 1$. Both schools are acceptable for all applicants.

Here it is possible to assign all applicants, namely $\mathcal{M}(a_1) = \mathcal{M}(a_3) = s_1$ and $\mathcal{M}(a_2) = \mathcal{M}(a_4) = s_2$. However, suppose that applicant a_1 arrives first and he/she chooses s_2. This leaves no place in s_2 for the remaining applicants a_2, a_3 and a_4. Further, since they all have *M* as one specialization subject and $c_M(s_1) = 1$, at most one of them can be accepted to s_1.

This shows that in situations when all applicants could get a place, an unsuitable order of arrivals may leave half of them unassigned.

FULL-TAP denotes the problem to decide, given an instance *J* of TAP, whether a full feasible assignment exists, i.e. such that leaves no students unassigned. In the following section we explore the computational complexity of several special cases of FULL-TAP.

4 COMPUTATIONAL COMPLEXITY

Theorem 1 FULL-TAP *is solvable in polynomial time in each of the following cases:*

(i) $|P| = 2$;

(ii) $|P| = 3$ *and no partial capacity of a school exceeds 1;*

(iii) $|P|$ *is arbitrary, but each applicant is allowed to list at most two acceptable schools and all partial capacities are at most 1.*

Proof. For case *(i)* it suffices to realize that all applicants are essentially equivalent and a school with partial capacities c_1 and c_2 can admit at most $c = \min\{c_1, c_2\}$ students. Hence FULL-TAP reduces to the classical bipartite *b*-matching problem that can be solved in polynomial time by any well-known algorithm [3].

Similarly, in case *(ii)* each school can admit at most one applicant, so FULL-TAP is equivalent to the simple maximum cardinality bipartite matching problem, again solvable in polynomial time.

In case *(iii)* let us proceed in the following way. In the first phase we deal with applicants that list an incompatible school or a school that does not have enough capacity for both specialization subjects. Such schools can be removed from their lists. If we get some applicants with empty lists, FULL-TAP is clearly insolvable. Otherwise, if the list of an applicant contains only one school (let us call these applicants *spoiled*), to get a full assignment, he/she must be assigned to that particular school. This, however, decreases the respective partial capacities of the school involved and new spoiled applicants can emerge. If, in this first phase we are not able to place all spoiled applicants, no full matching exists; otherwise we continue with the second phase with the partial capacities reduced accordingly. (It is easy to see that the first phase can be performed in polynomial time.)

The obtained *canonical* FULL-TAP instance *J* has $|S(a)| = 2$ for each $a \in A$. Let us denote $S(a_i) = \{s_i^1, s_i^2\}$ and introduce a boolean variable x_i for each applicant a_i with the

following interpretation: if x_i is TRUE, we shall say that a_i is assigned to school s_i^1; if x_i is FALSE, we say that a_i is assigned to school s_i^2. Now create a boolean formula *B(J)* in the following way. For each pair of applicants a_i, a_j whose specialization involves at least one common subject and for each school $s \in S(a_i) \cap S(a_j)$ we create a clause $C_{i,j,s}$ as follows:

- if $s = s_i^1$ and $s = s_j^1$ then $C_{i,j,s} = \overline{x_i} + \overline{x_j}$;
- if $s = s_i^1$ and $s = s_j^2$ then $C_{i,j,s} = \overline{x_i} + x_j$;
- if $s = s_i^2$ and $s = s_j^1$ then $C_{i,j,s} = x_i + \overline{x_j}$;
- if $s = s_i^2$ and $s = s_j^2$ then $C_{i,j,s} = x_i + x_j$.

Clause $C_{i,j,s}$ ensures that a_i and a_j do not both occupy the only place for their common subject at school *s*. Formula *B(J)* is then the conjuction of clauses $C_{i,j,s}$ for all triples a_i, a_j, s as described above. It is easy to see that *B(J)* is solvable if and only if a full assignment for *J* exists (remember, we assume that *J* is canonical). *B(J)* is a boolean formula in conjunctive normal form and since each clause contains just two literals, its satisfiability can be decided in polynomial time [7]. This concludes that case *(iii)* is also polynomially solvable. ■

Let us remark here that the computational complexity of the case with acceptable sets of cardinality 2 but with arbitrary partial capacities of schools is still open.

In the following theorem we shall use as the starting known NP-complete problem 3-dimensional matching, *3DM* in brief (see [7], problem SP1). An instance of *3DM* contains three disjoint sets *U, V* and *W*, all of cardinality *n*, and a set of triples $\mathcal{T} \subseteq U \times V \times W$. The question is whether there exists a perfect matching, i.e. a subset $\mathcal{N} \subseteq \mathcal{T}$ such that $|\mathcal{N}| = n$ and $\mathcal{N}$ covers all elements of $U \cup V \cup W$. We shall use the NP-complete restriction of *3DM* to such instances where no element occurs in more than 3 triples in $\mathcal{T}$.

Theorem 2 FULL-TAP *is NP-complete even when* $|S(a)| \leq 3$ *and*
(i) $|P| = 3$ *and no partial capacity of a school exceeds 2; or*
(ii) $|P| = 4$ *and no partial capacity of a school exceeds 1.*

Proof. For case *(i)*, given an instance $J = (U, V, W, \mathcal{T})$ of *3DM*, we construct an instance J' of TAP with 3 subjects (say *M, F* and *I*) and $c_M(s) = 2, c_F(s) = c_I(s) = 1$ for each school.

For each triple $t \in \mathcal{T}$ we create a school s_t. For each $z \in U \times V \times W$ let $\mathcal{T}_z$ be the set of triples in $\mathcal{T}$ containing z and $\ell_z = |\mathcal{T}_z|$. For each $u \in U$ we create applicants $a_u^1, a_u^2, \ldots, a_u^{\ell_u - 1}$, each of type *IF*; their set will be denoted by A_u. For each $v \in V$ we create an applicant a_v of type *MI* and for each $w \in W$ an applicant a_w of type *MF*. For each applicant corresponding to an element $z \in U \times V \times W$, acceptable schools are those that correspond to triples in $\mathcal{T}_z$.

Suppose that the *3DM* instance *J* has a perfect matching $\mathcal{N} \subseteq \mathcal{T}$. We assign each applicant in J' to an acceptable school so that the capacity of no school in no subject will be exceeded.

For each $t = (u, v, w) \in \mathcal{N}$ we assign to school s_t applicants a_v and a_w. For each $u \in U$ there are $\ell_u - 1$ triples $t \in \mathcal{T} \setminus \mathcal{N}$ containing *u*, so to the corresponding schools we assign applicants $a_u^1, a_u^2, \ldots, a_u^{\ell_u - 1}$. It is easy to see that each applicant is assigned to an acceptable school and that the defined assignment obeys all capacities.

Conversely suppose that there exists a full feasible assignment $\mathcal{M}$. Let $S_{\mathcal{N}}$ be the set of schools to which two applicants are assigned in $\mathcal{M}$ and let $\mathcal{N} \subseteq \mathcal{T}$ be the set of corresponding triples. By the construction, if $s_t \in S_{\mathcal{N}}$ and $t = (u, v, w)$ then the assigned applicants are a_v and a_w. Clearly, for two different schools in $S_{\mathcal{N}}$ these two applicants are

different and so also any two different triples in $\mathcal{N}$ differ in their elements from *V* and *W*. It remains to show that if $t, t' \in \mathcal{N}$ are different then their corresponding elements from *U* are also different.

To get a contradiction, suppose that some element $u \in U$ belongs to at least two different triples $t, t' \in \mathcal{N}$. Notice that the only acceptable schools for the $\ell_u - 1$ applicants of the set A_u are the ℓ_u schools s_t for $t \in \mathcal{T}_u$. If two different schools $s_t, s_t{}'$ belong to $S_{\mathcal{N}}$ then the number of schools that have enough capacity for $\ell_u - 1$ applicants in A_u and are acceptable for them is at most $\ell_u - 2$. This is a contradiction with the assumption that $\mathcal{M}$ is a full assignment.

The proof for case *(i)* can easily be modified for *(ii)* by making the following changes:

- The set of subjects is *M,F,I,B;*
- each school *s* has $c_M(s) = c_F(s) = c_I(s) = c_B(s) = 1$;
- for each $v \in V$ the type of applicant a_v is *MF*;
- for each $w \in W$ the type of applicant a_w is *IB*;
- for each $u \in U$ contained in ℓ_u triples in $\mathcal{T}$ there are $\ell_u - 1$ applicants of type *MI* and $\ell_u - 1$ applicants of type *FB*.

The acceptability is defined in the same way according to the structure of $\mathcal{T}$ and the rest of the proof is analogous. ■

5 CONCLUSIONS AND OPEN QUESTIONS

In the quest for a possible centralized matching scheme the presented intractability results are pessimistic. Still, some other computational techniques could be employed, e.g. integer programming formulations. One should also see whether the complexity status of the problem changes if the students are not allowed to express acceptability, i.e. if each student were required to go to any school that provides both subjects of his/her specialization and has a free place for each.

The existing extensive literature on matchings and many existing schemes call for exploring other possible approaches. One can imagine that students, in addition to expressing acceptability, could be allowed to list the acceptable schools in order of their preference and/or the schools might also be given the right to order students. Then some other criteria for the obtained matching might be considered: Pareto optimality (from the viewpoint of students, see [4]) or stability (introduced by Gale and Shapley [6]).

6 ACKNOWLEDGEMENT

The authors thank David Manlove for several constructive remarks to the earlier version of the paper. This work was supported by VEGA grants 1/0410/11 and 1/0479/12 (Cechlárová, Potpinková) and by OTKA CK80124 and the ELTE-MTA Egerváry Research Group (Fleiner). We also gratefully acknowledge the support of the Operational Program "Education and Research" funded by the European Social Fund, grant "Education at UPJŠ – Heading towards Excellent European Universities", ITMS project code: 26110230056.

References

[1] A.Abdulkadiroglu, P. A. Pathak, A. E. Roth, *The Boston public school match*, American Economic Review 95(2), 368–371, (2005).

[2] A.Abdulkadiroglu, P. A. Pathak, A. E. Roth, *The New York City high school match*, American Economic Review 95(2), 364–367, (2005).

[3] R. K. Ahuja, T. L. Magnanti, J. B. Orlin, *Network flows: theory, algorithms, and applications*, Prentice Hall, 1993.
[4] D. Abraham, K. Cechlárová, D. Manlove and K. Mehlhorn, *Pareto optimality in house allocation problems*, Lecture Notes in Computer Science (LNCS) 3341, 3-15 (2004).
[5] P. Biró, T. Fleiner, R.W. Irving, D.F. Manlove, *The college admissions problem with lower and common quotas*, Theoret. Comput. Sci. 411 (2010), no. 34-36, 3136-3153.
[6] D. Gale, L.S. Shapley, *College Admissions and the Stability of Marriage*, Amer. Math. Monthly 69, no. 1, 9-15, (1962).
[7] M.R. Garey, D.S. Johnson, *Computers and Intractability*, Freeman, San Francisco (1979).
[8] D. Gusfield and R. W. Irving, *The Stable Marriage Problem: Structure and Algorithms*, Foundations od Computing, MIT Press, Cambridge (1989).
[9] A. Hefner, P. Kleinschmidt, *A constrained matching problem*, Annals of Operations Research 57: 135-145 (1995).
[10] R. W. Irving, *Matching medical students to pairs of hospitals: a new variation on an old theme*, Lecture Notes in Computer Science (LNCS) 1461, 381-392 (1998).
[11] E. McDermid, D.F. Manlove, *Keeping partners together: Algorithmic results for the Hospitals / Residents problem with couples*, J. Comb. Optim. 19 (3) : 279-303, 2010.
[12] D.F. Manlove, *Algorithmics of Matching Under Preferences*, World Scientific, 2013.
[13] A.E. Roth, *The evolution of the labor market for medical interns and residents: a case study in game theory*, Journal of Political Economy 6(4) (1984), 991-1016.
[14] http://www.nrmp.org
[15] http://www.matching-in-practice.eu

Network Formation with Nodewise Decay

Banchongsan Charoensook
b.charoensook at alhosnu.ae
Department of Business Administration
ALHOSN University, UAE

Abstract : This paper develops a model of noncooperative network formation. Link formation is two-sided. Information flow is two-way and imperfect. The paper is built upon Bala and Goyal [1]. A unique assumption is that the value of information decays as it flows through each agent, and the decay is increasing and concave in the number of his links. Thus, an agent may choose to avoid accessing an agent who possesses many links since he is aware of the increasing decay incurred at this agent. This avoidance leads to two particular results in the analysis of Nash networks: (1) Nash networks are not always connected; (2) Nash networks do not exist under some parameters. Since disconnectedness is reminiscent of a common feature of real-world network, the model may explain why real-world networks may exhibit this feature even when there is no heterogeneity among agents. Discussion on this insight is provided.

Keywords : Social Networks, Network Formation, Nash Network, Game Theory

I Introduction

This paper presents a model of network formation game that is built upon the two-way flow model of Bala and Goyal [1], henceforth BG. A unique assumption is that *an increase in link establishment damages the quality of information that flows in a network*. Each agent knows that whenever he establishes a link with another agent both of them transmit information less efficiently than before, causing a decline in the value of information that flow through them. This decline is, therefore, a disbenefit not only to the themselves but also other agents in the network. Put differently, on top of link formation cost, there are additional disbenefits associated with link formation. This paper aims to understand how this assumption may affect link-formation decision of agents and hence the shape of equilibrium networks. To this end I characterize the shapes of equilibrium networks and analyze why they differ from those of other models in the literature. Finally, using the analyses the paper discusses how the model may explain some features of real-world networks.

I argue that my assumption is realistic and hence worth studying, particularly in the context of information network. Consider a firm in which employees' task is to communicate with each other. In this network, there may be a center-like agent whose role is to collect and distribute information of other agents. Such agent is important because how much the information is loss depends on his communicating performance that is likely to decline as there are more contacts between him and other agents. Consequently, if the information loss is too high, an agent may avoid contacting the center by contacting another agent or staying completely disconnected from this network. The fact that the center finds more difficulties in transmitting information as he has more links is a form of network congestion and the fact that other agents may avoid contacting the center can be considered a form of *congestion avoidance*. However, how this realism affects agents' strategic linking decision has not been investigated in the literature in strategic network formation to my knowledge. My attempt to address this uninvestigated issue is thus the central contribution of this paper.

With this situation in mind, I address this network congestion issue by making the following modification to the two-way flow model of BG. First, in a network g I let the decay

factor be *nodewise*: as information is transmitted through agent i, a fraction of information equal to $1 - \sigma(i; g)$ is loss. Second, $\sigma(i; g)$ is decreasing and strictly concave on the amount of i's links. The strict concavity is assumed to reflect the realism that i faces increasing difficulties at an increasing rate in transmitting information as more agents contact him. This assumption entails a particular link-formation behavior, in that an agent may face a tradeoff between forming a link with an agent who has many links yet more difficulties to transmit the information and forming a link with an agent who has less links and less difficulties to do so.

Besides these two assumptions I retain all assumptions of two-way flow of BG, which are briefly described here for unfamiliar readers. Specifically, the original setting of BG is as follows. Each agent possesses a unique private piece of information that is nonrival. He can choose to sponsor costly links to any agents without their agreements. All links together form the network. If there is a link or a series of links between two agents (called chain), they are obliged to share their private informations. Thus, the decision of agent to form a link represents his decision to make his private information available to other agents in exchange of receiving their informations, and concurrently his willingness to be an information transmitting device. The decay factor is assumed to be geometric and linkwise: each link causes a fraction of information loss equal to $1 - \sigma$, where σ is constant.

Based on the observation from the main results, two insights on the structure of real-world networks can be learnt. First, when network congestion is present, an equilibrium network may be fragmented, consisting of subnetworks disconnected from each other. Second, with network congestion, moving from a smaller network to a larger one (a network with more agents) does not imply that the moving agent will improve his payoffs. The intuition is that agents in a larger network may be more congested (having more links), causing information to flow better in a smaller network. This may explain why real-world networks often consist of fragmented communities of notably different sizes. For example, in a friendship network, some students may prefer to keep their friendship within a small group rather than joining the crowd because they enjoy a stronger friendship that provides a higher benefit flow. These insights can be observed in my first proposition, which finds that no Nash network is connected under some restriction on the decay parameter. This disconnectedness is a sharp contrast to the result in the original model of BG that all Nash networks are connected.

My paper contributes to the literature in network formation. This literature is pioneered by the work of Jackson. and Wolinsky [9]. Their model assumes that two agents must share a mutual consent in order that a link is established. A seminal work that contrasts to this model is that of BG, in which one-sided link formation is assumed. These seminal works raise a question as to how certain realisms, when incorporated as assumptions, influence the shape of equilibrium networks. Most literature in this strand questions the role of agent's heterogeneity and/or link as a major cause of inefficiency in benefit flow or both [1]. Among such vast literature, the model of [3] has in mind a situation similar to mine, in that managing too many links simultaneously leads to information congestion. It assumes that the cost of link maintenance increases in relation to the quantity of informations received. Hence, accessing an agent does not damage the quality of information flow at the accessed agent. My model differs in that network congestion is reflected directly in the increasing information loss in both the agent being accessed and the accessing agent. This allows us to better observe the effects of congestion avoidance. Besides this difference, [3] assumes that information sharing is not two-way, in that the agent who forms a link does not share his information with his partner.

II The Model

Let $N = \{1, ..., n\}$ be a set of agents and let i and j be typical members of this set. Each agent possesses a nonrival unique private piece of information that is valuable both to himself and anyone who has an entry to it. There are two ways in which a pair of agents can have an entry to each other's information: there is a pairwise link between i and j, or a chain such that the two ends are i and j.

Link establishment and individual's strategy. Link establishment is costly and one-sided. A strategy of i is $g_i = \{g_{ij} : j \in N, j \neq i\}$, where $g_{ij} = 1$ if i forms a link with j and $g_{ij} = 0$ otherwise. If $g_{ij} = 1$, I say that i accesses j. Since all links form the network, I write $g = \{g_i : i \in N\}$ to represent both a strategy profile and a network. Naturally I define $\mathcal{G}$ as the set of all g to represent both strategy space and the set of all possible networks.

Network representation. In this paper a node depicts an agent, and an arrow from node i to node j represents that i forms a link with j. If all arrows are removed, the modification merely represents who has a link with who. Such modification is called network closure and is denoted by $\bar{g} \equiv \{\bar{g}_{ij} : i, j \in N, i \neq j\}$, where $\bar{g}_{ij} = 1$ if $g_{ij} = 1$ or $g_{ji} = 1$ or both, and $\bar{g}_{ij} = 0$ otherwise. A network closure also illustrates how information flows among agents.

Information flow. Information of j flows to i directly through a link between i and j, regardless to who sponsors it. Alternatively, information of j can also flow to i through a series of link called *chain*. Formally, an $ij-$chain is a sequence of distinct agents $j_0, ..., j_m$ such that $\bar{g}_{j_l, j_{l+1}} = 1$ for $l = 0, ..., m-1$ and $j_0 = i$ and $j_m = j$, and is denoted by $\bar{P}_{ij}$. In this case, I say that i *observes* j.

Value of information. Information decay is node-wise. That is, whenever information arrives or is sent to an agent i, a decay of information is incurred. The percentage rate of information that remains is $\sigma(i; g)$. Let the value of a piece of information when there is no decay be 1. Naturally if information of j flows through a chain between i and j, the value of information of j that i receives is $V(\bar{P}_{ij}) = \prod_{k \in \bar{P}_{ij}} \sigma(i; g)$, where $k \in \bar{P}_{ij}$ represents that an agent k is a part of the chain $\bar{P}_{ij}$.

Costs and benefits If i accesses j, then i pays $c_{i,j}$. If i observes j through multiple $ij-$chains, naturally i chooses to obtain $j's$ information through an *optimal chain*. Formally, an optimal $ij-$chain is $\bar{P}^*_{ij}$ such that $V(\bar{P}^*_{ij}) \geq V(\bar{P}'_{ij})$ for every existing $\bar{P}'_{ij}$ in the network. If an $ij-$chain exists, the value of j's information that i receives from an optimal $ij-$chain is $\bar{V}_{ij}(g) = V(\bar{P}^*_{ij})$. If an $ij-$chain does not exist, I set $\bar{V}_{ij}(g) = 0$. For i's own information, naturally I set $\bar{V}_{ii}(g) = 1$ if he has no link at all. If he has a link I set $\bar{V}_{ii}(g) = \sigma(i; g)$, reflecting the fact that he incurs some loss in his own information.

Payoffs. the payoff of player i from the strategy profile g is:

$$\Pi_i(g) = \sum_{j \in N} \bar{V}_{ij}(g) - c\mu_i(g)$$

where $\mu_i(g)$ is the amount of links that i establishes. I remark that the first term on the right-hand side is the total value of information that i receives in g or *the total benefit of* i in g and is denoted by $B_i(g) = \sum_{j \in N} \bar{V}_{ij}(g)$.

Network-related Notations. Recall from the above that a chain between i and j is a sequence of distinct players $j_0, ..., j_m$ such that $\bar{g}_{j_l, j_{l+1}} = 1$ for $l = 0, ..., m-1$ and $j_0 = i$ and $j_m = j$, a path is defined similarly except that $g_{j_l, j_{l+1}} = 1$ instead of $\bar{g}_{j_l, j_{l+1}} = 1$. A cycle is defined in the same fashion as a chain, except that $j_0 = i$ and $j_m = i$ and all other players in the sequence are distinct. I use these notations to define the following terms. A network is *connected* if there is a chain for every distinct $i, j \in N$. A subnetwork of g is a network g' such that $g' \subset g$. A *component* of g is a maximal connected subgraph of g. A component is said to be minimal if it contains no circle. A component is a *line* if it is minimal, and contains exactly two agents that have only one link and every other agent has exactly two links.

Nash Network. Let g_{-i} denote a strategy profile of all agents except i, ie., $g_i \cup g_{-i} = g$. A best response of an agent i is g_i such that $\Pi_i(g_i \cup g_{-i}) \geq \Pi_i(g_i' \cup g_{-i})$ for every g_i' that is a strategy of i. A strategy profile or a network g is Nash if every agent plays his best response.

II.1 Assumptions on decay

My key assumption is that the decay factor $\sigma(i; g)$ depends solely on the number of i's links. Let $\bar{\mu}(i; g) \equiv |j \in N : \bar{g}_{ij} = 1|$ be the amount of i's links.

Assumption (Concave Decreasing Decay)**.** *Let* $\varsigma : \mathbb{N} \to [0, 1]$ *be a function such that:*

1. ς_x *be a value at* $x \in \mathbb{N}$
2. $\varsigma_1 = 1$
3. *there exists* $K > 1$ *such that* $\sigma_x = 0$ *for all* $x > K$*. Moreover, for* $x \leq K$ ς *is decreasing and strictly concave.*

Throughout this paper I assume that $\sigma(i; g) = \varsigma_{\bar{\mu}(i;g)}$ *for all* $i \in N$ *and* $\bar{\mu}(i; g) > 0$.

I now elaborate on these assumptions. First, $\sigma(i; g) = \varsigma_{\bar{\mu}(i;g)}$ implies that an agent's decay factor depends solely on the number of his links. Moreover, two agents have the same decay factor if they have the same amount of links. That is, agent homogeneity is assumed. Second, $\varsigma_1 = 1$ implies that perfect information transmission between two agents occur only if both of them have links with no other agents but themselves. Third, that ς is strictly concave and decreasing implies that the decline in decay factor increases at an increasing rate. Put informally, I assume that agents find that the difficulties in transmitting information increases at an increasing rate as they have more links. While there is no theoretical support, I believe that this assumption can be justified by the following scenarios. Suppose that an agent stores all pieces of information in one place, then due to the limitedness of space the chance that multiple pieces of information get mixed up, causing more difficulties in communicating accurately is likely to increase at an increasing rate. Another example is when each piece of information is very similar to one another, then the chance that an agent does not know which is which is also likely to increase at an increasing rate. Finally, the existence of K in the last part implies that the decay factor reaches zero at a certain point and remain there, rather than becoming negative.

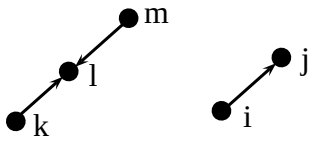


Figure 1: A Nash network with five agents for $c = 2\varsigma_2 < \frac{1}{2}$

III Main Result

For $\varsigma_2 \leq \frac{1}{2}$, I find the following result [2].

Proposition 1. *If $\varsigma_2 \leq \frac{1}{2}$, every non-empty component in Nash network is a two-agent line or three-agent line such that the center agent receives two links.*

Figure 1 demonstrates a Nash network as described in Proposition 1. Contrary to the above result, if $\varsigma_2 > \frac{1}{2}$, Nash network does not always exists. Readers are recommended to refer to Example 2 in my working paper [4] for the illustration of nonexistence and Proposition 2, 3 and 4 for partial characterization of Nash network for $\varsigma_2 > \frac{1}{2}$.

IV Discussions

This section points out two particular features of equilibrium networks in my model. I provide intuitions that explain why they arise. Finally I discuss how these intuitions may explain some features of real-world networks.

IV.1 Network congestion may lead equilibrium networks to be disconnected

The first observation is the fact that all Nash networks for $\varsigma_2 \leq \frac{1}{2}$ are disconnected (Proposition 1). The intuition can be summarized as follows. While establishing a link to an agent is a way to access a component, it also increases the congestion at the agent being accessed. This congestion may cause much loss in the information transmitted via the agent. When such congestion, or inefficiency in information transmission, is sufficiently high, an agent may be better off avoiding the congestion altogether and remain disconnected from the component.

How does this observation help understand real-world phenomena? My model may serve as a hypothesis that explains why empirical evidences find that real-world networks are often disconnected [3]. For example, if a society is considered as a network in which information is exchanged among agents, it is likely that the society is fragmented into small communities if agents find that avoiding connection to each other is a way to reduce inefficiency in information flow.

IV.2 Connecting to a larger component does not imply a higher gain

My second observation is that a smaller component may provide a higher gain to their members than a larger one. It comes from the fact that many Nash networks in Proposition 1

consist of components whose sizes (measured by the number of agents) are not equal. Consider, for example, the equilibrium network in Figure 1. Observe that i chooses to access an isolated agent j rather than someone in the larger component. If i accesses j, $\sigma(j;g) = \varsigma_1$. If i accesses someone in the larger component, $\sigma(j;g)$ is at most ς_2. Hence, if ς_2 is sufficiently lower than ς_1, then entering a larger component gives i a lower gain.

This observation may explain why there are agents who prefer to reside in a smaller community rather than a larger one in a real-world social network. When a link is a source of inefficiency, a smaller community that has less connections may provide a higher benefit to the participating members such that they do not want to join a crowded community. In other words, agents may face a tradeoff between quantity of information and quality of information when network congestion is present. While a larger community may have more information, the quality of information may be deterred if agents possess too many connections. A friendship network among students may serve as an example of this hypothesis. Some students may choose to maintain their friendships within a smaller group and avoid contacting the crowd because they enjoy a stronger tie of friendship [4].

V Conclusion

This paper provides a stylized model with two key assumptions. First, link can be formed without a mutual consent between agents. Second, link addition increases the congestion, or more information loss, at the accessed agent and the agent who accesses. The model allows us to see how an agent may avoid accessing other agents due to increasing congestion. The two key assumptions lead to equilibrium networks that are disconnected. Moreover, nonexistence of equilibrium network in pure strategies arises under some parameters. These two features are different from the results in the original setting of [1] from which this model is developed.

Finally, I remark that while it is hard to make generalization from my simplified model, the link-formation behavior of agents in equilibrium neworks may provide some insights to common features of real-world networks. First, the disconnectedness found in equilibrium networks root in that adding a link to bridge two components results in the increasing congestion at the accessed agent and the agent who accesses. As such the payoff of an accessing agent may not improve even though the link gives an entry to more information. This result may explain why real-world networks are often disconnected. Second, for the same reason accessing a big component that has more agents (and more information) also does not guarantee a payoff improvement. This may explain why some agents choose to be disconnected from the major component in real-world networks.

---

[1] See, for instance, [7], and [8] for agent heterogeneity. For link inefficiency see [5] [2]

[2] Indeed, if $\varsigma_2 \leq \frac{1}{2}$, Nash network always exists and can be fully characterized according to different levels of c. Readers are recommended to refer to the proposition 1 and its proof in my working paper [4].

[3] For instance, sociologists have long observe that a common feature of friendship networks is that there are agents who are social isolates, disconnecting themselves from the principal component. Also [10] gives a surprising remark that several online social networks contain isolated communities and singletons - agents who completely have no links.

[4] Indeed, there is a vast literature on the behavior of 'social isolates' especially in adolescent social networks. For an introduction see, for instance, [6].

References

[1] Bala, V., Goyal, S., 2000. A Noncooperative Model of Network Formation. Econometrica 68(5), pp. 1181–1230.

[2] Bala, V., Goyal, S., 2000. A Strategic Analysis of Network Reliability. Rev Econ Design 5(3), pp. 205–228.

[3] Caffarelli, F., 2009. Networks with Decreasing Returns to Linking. Economics Working Papers ECO2004/18, Bank of Italy, http://ideas.repec.org/p/eui/euiwps/eco2004-18.html

[4] Charoensook, B., 2012. Networks Formation with Productivity as Decay. MPRA Working Paper, http://mpra.ub.uni-muenchen.de/37099/

[5] Deroian, F., 2009. Endogenous Link Strength in Directed Communication Networks. Math Soc Sci 57(1), pp. 110–116.

[6] Ennett, T., Bauman, K., 2000. Adolescent Social Networks: Friendship Cliques, Social Isolates, and Drug Use Risk. In: Hansen, W., Giles, S., Fearnow-Kenney, M. (eds.), Increasing Prevention Effectiveness, Tanglewood Research Inc., pp. 83–92.

[7] Galleotti, A., Goyal, S., Kamphorst, J., 2006. Network Formation with Heterogeneous Players. Game Econ Behav 54(2), pp. 353–372.

[8] Haller, H., Sarangi, S., 2005. Nash networks with Heterogeneous Links. Math Soc Sci 50(2), pp. 181–201.

[9] Jackson, M., Wolinsky, A., 1996. A Strategic Model of Social and Economic Networks. J Econ Theory 71(1), pp. 44–74.

[10] Kumar, R., Novak, J., Tomkins, A., 2010. Structure and Evolution of Online Social Network. In: Philip, Y., Jiawei, H., Faloutsos, C. (eds.), Link Mining: Models, Algorithms, and Applications, Springer, pp. 611–617.

DIFFERENT GRAPH INVARIANTS AND HEXAGONAL GRAPHS

Rija Erveš[a,b] and **Petra Šparl**[a,c]
[a]Institute of mathematics, physics and mechanics
Jadranska 19, 1000 Ljubljana, Slovenia
[b]University of Maribor, Faculty of Civil Engineering
Smetanova 17, 2000 Maribor, Slovenia
[c]University of Maribor, Faculty of Organizational Sciences
Kidričeva 55a, 4000 Kranj, Slovenia
rija.erves@um.si, petra.sparl@fov.uni-mb.si

Abstract: Three similar graph invariants will be presented. More precisely, minimum vertex *k*-path cover of *G*, denoted by $\psi_k(G)$, dissociation number of *G*, denoted by *diss(G)*, and maximum induced matching of *G*, denoted by $\upsilon_k(G)$. This paper concentraces on maximum induced matching problem in special subset of planar graphs, called *hexagonal graphs*. Tight lower bound on maximum induced matching in hexagonal graphs is given.

Keywords: Matching, maximum induced matching, hexagonal graph.

1 INTRODUCTION

Let *G* be a graph and *k* be a positive integer. Then $S \subseteq V(G)$ is the *vertex k-path cover* of *G* if every path on *k* vertices in *G* contains a vertex from *S*. Let denote by $\psi_k(G)$ the cardinality of a minimum vertex *k*-path cover in *G*. This graph invariant was recently introduced in [1], where the motivation for this problem arises in ensuring data integrity communication in wireless sensor networks using the *k*-generalized Canvas Scheme [11]. Determining ψ_k k for *k=2* was shown to be NP-hard problem in general and polynomial only for some special sets of graphs (for details we refer to [1, 6] and references there).
A special case of a *k*-path vertex cover problem is finding a graph invariant $\psi_3(G)$, which corresponds to the concept of *dissociation number* of a graph, defined as follows. A subset of vertices in a graph *G* is called *dissociation set* if it induces a subgraph with maximum degree 1, i.e. edges and isolated vertices. The number of vertices in a maximum cardinality dissociation set in *G* is called the dissociation number of *G* and is denoted *diss(G)*. It is not diffucult to see that $\psi_3(G) = |V(G)| - diss(\mathrm{G})$. The problem of computing *diss(G)* has been introduced by Yannakakis [18], who also proved it to be NP-hard in the class of bipartite graphs. For the survey on results regarding the dissociation number problem we refer to [12] and references there.

The third graph invariant, which will be discussed in this paper, arises from the matching concept and is very similar to dissociation number. Let *G=(V,E)* be a simple connected graph. A set of edges $M \subseteq E(G)$ is a *matching* or *an independent edge set* if no two edges of *M* share a common vertex. Matchings have been researched extensively for many years. In this paper we will consider *induced matching*, which is a matching in which no two edges in the matching have a third edge in the graph connecting them. A well known problem is a problem of finding a maximum induced matching of a given graph *G* or shortly MIM. The size of a maximum induced matching of *G* is denoted by

$$\upsilon(G) = \max\{|\mathrm{M}| \mid \mathrm{M} \subseteq E(G) \text{ is an induced matching of } G\}.$$

Stockmeyer and Vazirani [13] introduced MIM as a variant of the maximum matching problem and motivated MIM as the ”risk-free” marriage problem. Induced matchings have stimulated a great deal of interest in the discrete mathematics community, since finding large induced matchings is a subtask of finding a strong edge colouring (i.e. a proper colouring of

the edges such that no edge is adjacent to two edges of the same colour) using a small number of colours. For a brief survey of applications of this type of colouring and some open questions, we refer the reader to [8, 17]. Stockmeyer and Vazirani [13] and Cameron in [2] showed that MIM is NP-hard in general and it remains NP-hard even when the input graph is bipartite. On the other hand, MIM has been shown to be solvable in polynomial time for several graph classes [3, 4, 5, 7, 9, 10, 19].

We will discuss MIM for special subset of planar graphs, called *hexagonal graphs*, which are induced subgraphs of triangular lattice. We take a combinatorial approach to the problem, establishing tight lower bound on the size of maximum induced matching in an arbitrary hexagonal graph.

2 PRELIMINARIES

A simple *graph* is determined by *G = (V,E)*, where *V = V (G)* is the *vertex set* and *E=E(G)* is the set of (unordered) pairs of vertices, called *edges*. For edge {*u,v*} we will use a short notation *uv* and call vertices *u* and *v endpoints* of edge *uv*. A path on n vertices will be denoted by P_n. We say that a graph is *connected* if there is a path between each pair of vertices, and is *disconnected* otherwise. As we already mentioned, we will discuss the MIM problem in hexagonal graphs. Graph *G* is called a *hexagonal graph* if it is induced on the subset of vertices of the triangular lattice. Hexagonal graphs arises within the problem of frequency assignment in cellular networks. For a more detailed explanation of the problem and a survey of existing results on the topic, we refer the reader to [14, 16] and references there. An example of a hexagonal graph is presented in Figure 1.

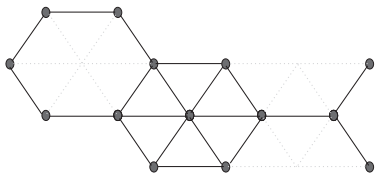


Figure 1: An example of a hexagonal graph.

More precisely, we will derive tight lower bound for the size of maximum induced matching of a connected hexagonal graph *G* with respect to the number of vertices of *G*.
In the continuation of this section some notations, definitions and some partial results are given. For an arbitrary edge $e = uv \subseteq E(G)$ the following notations will be used: $N_G(e)$ and $N_G[e]$ for the *open* and *closed neighborhood* of edge $e \subseteq E(G)$ in graph G, respectively. Edge degree of edge $e \subseteq E(G)$ will be denoted by $d_G(e) = |N_G(uv)| = |N_G(u) \cup N_G(v)|$-2. Further, *G(e)* denotes a subgraph of *G* induced on vertices $V(G)\backslash N_G[e]$, while isolated vertices in *G(e)* will be denoted by $I_G(e) = \{w \in V(G(e)) | d_{G(e)}(w) = 0\}$.

Let *G* be a connected hexagonal graph. We want to find an induced matching *M* of *G*. Note that an induced matching of a graph *G* actually divides the set of vertices *V(G)* into two subsets, such that endpoints of edges from *M* are in the first set, let say *S*, all the other vertices are in the second set, let say *P=V(G)\S*, called the set of *protectors*. Therefore, the induced matching can be also discussed as special bicolouring *c:V(G)*→{white,black}, which assigns white colour to vertices of S and black colour to *protectors*, i.e. vertices of *P*.

Let suppose that edge $e = uv \subseteq E(G)$ belongs to the induced matching *M* of *G*, which means that vertices u and v belong to S and are assigned white colour. Note that in this case all vertices in the open neighborhood of the edge e must be protectors and therefore coloured black. Moreover, all isolated vertices in G(e) (a subgraph of *G* induced on vertices $V(G)\backslash N_G[e]$, denoted by $I_G(e)$, must be assigned black colours too. Therefore, the inclusion of an

edge e to the induced matching M of G, contributes $d_G(e) + |I_G(e)|$ black vertices to the set of protectors P.

It turned out that only three connected hexagonal graphs, denoted by H_1, H_2 and H_3 in Figure 2, have different property regarding the minimal possible number of $d_G(e) + |I_G(e)|$ for an arbitrary edge $e \subseteq E(G)$. Namely, only for these three hexagonal graphs the following equation holds:

$$\min\left\{d_{H_i}(e) + \left|I_{H_i}(e)\right| \;\middle|\; e \in E(H_i)\right\} = 5, i = 1,2,3.$$

While for all other hexagonal graphs the result is the following (the proof is given in [15]).

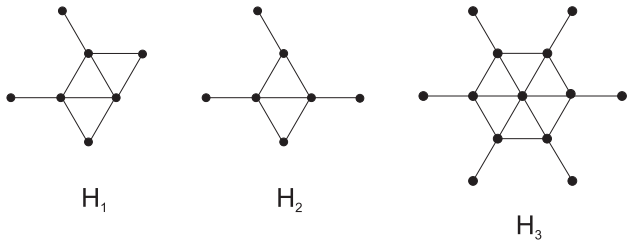


Figure 2: Graphs H_1, H_2 and H_3.

Lemma 1 Let G be a connected hexagonal graph with $n \geq 2$ vertices which is not isomorphic to graphs H_1, H_2 or H_3 from Figure 2. Then for a graph G it holds

$$\min\left\{d_G(e) + |I_G(e)| \;\middle|\; e \in E(G)\right\} \leq 4.$$

3 THE MAIN RESULT

Let G be an arbitrary connected hexagonal graph with $|V(G)| \geq 2$. The following procedure presents a bicolouring $c:V(G)\rightarrow$\{white,black\}, which assigns white colours to vertices of S and black colours to vertices of P, called protectors, such that $V(G) = S \cup P$, where S is the endpoints set of edges of an induced matching in G.

Procedure 2 *Let G be a connected hexagonal graph such that $|V(G)| \geq 2$ and let H_1, H_2 and H_3 be graphs depicted in Figure* **Napaka! Vira sklicevanja ni bilo mogoče najti.**.

Step 1 If graph G is isomorphic to graph H_1 or to grap H_2, then colour two adjacent vertices white and other five vertices black.

Step 2 If graph G is isomorphic to graph H_3, then colour six vertices white and seven vertices black such that each white vertex is a neighbor of exactly one white vertex.

Step 3 If graph G is not isomorphic to any of the graphs H_1, H_2 and H_3, then do what follows.

Step 3a If there exist one, choose an edge $e \in E(G)$ with minimal possible number $d_G(e) + |I_G(e)| \leq 4$ so that the subgraph $G(e) \setminus I_G(e)$ is either a connected or an empty graph.

Step 3b Otherwise, choose an edge $e \in E(G)$ with minimal possible number $d_G(e) + |I_G(e)| \leq 4$ so that the subgraph $G(e) \setminus I_G(e)$ is not a connected graph.

Colour the endpoints of the edge e white and vertices of $N_G(e) \cup I_G(e)$ black.

For every connected uncoloured component G_i of the subgraph $G(e)\backslash I_G(e)$ go to the Step 1 ($G \rightarrow G_i$). Repeat with white-black colouring of the remaining uncoloured components after Step 3 until such components do not exist.

Note that Step 3 of the procedure is divided into substeps (3a) and (3b). At first it looks like that every connected hexagonal graph *G*, such that $G \ncong H_1, H_2, H_3$, belongs to Step 3a, but actually this is not the case. Namely, Figure 3 represents an example of a hexagonal graph G_1 that belongs to Step 3b. More precisely, for every edge e, such that $d_{G_1}(e) + \left|I_{G_1}(e)\right| \leq 4$, the subgraph $G_1(e)\backslash I_{G_1}(e)$ is disconnected.

It turned out that the following results hold (proofs are given in [15]).

Proposition 3 *Let G be a connected hexagonal graph colored by Procedure 2. Then vertices that were assigned white colour correspond to endpoints of edges of an induced matching of G and* $\upsilon(G) \geq \frac{|S|}{2}$.

Lemma 4 *For each connected hexagonal graph, which is not isomorphic to any of graphs* H_1, H_2 and H_3 *from Figure 2, at most one connected component, obtained during the realization of Procedure 2, can be isomorphic either to* H_1 *or to* H_2.

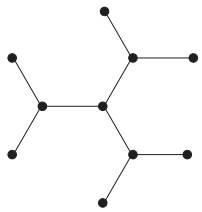


Figure 3: Graph G_1: $|V(G_1)| = 10, \upsilon(G_1) = 3$.

Lemma 5 *Let G be a connected hexagonal graph and M an induced matching of G. Further, let S be the set of endpoints of edges of M. For the set of protectors P = V (G)\S it holds*

$$|P| \leq 2|S| + 1.$$

The bound of inequality in Lemma 5 is tight. Namely, the example of the connected hexagonal graph G_2 with $|V(G_2)| = 25$, depicted on Figure 4, attains the maximal possible number of protectors depending on the number of white edges, $|P| = 17 = 2 \cdot 8 + 1 = 2|S| + 1$.

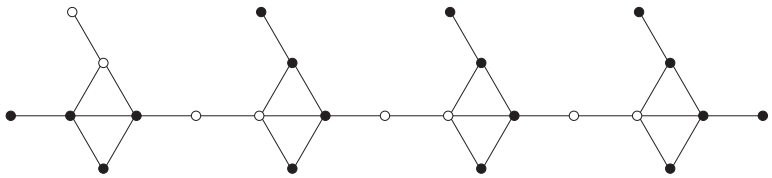


Figure 4: Graph G_2: $\upsilon(G_2)$=4.

Using Proposition 3 and Lemma 5 it can be proved that for an arbitrary hexagonal graph the following theorem holds.

Theorem 6 *Let G be a connected hexagonal graph with* $n \geq 2$ *vertices. Then*

$$\upsilon(G) \geq \left\lceil \frac{n-1}{6} \right\rceil.$$

Bound of Theorem 6 is tight. Namely, for each $n \geq 2$ there exists a hexagonal graph G, obtained by connecting several components of graph H_1, with $\upsilon(G)$ equal to the lower bound of Theorem 6. For example, figure 4 represents graph G_2, which is obtained by connecting four components of graph H_1, such that n = 25 and $\upsilon(G) \geq \left\lceil \frac{n-1}{6} \right\rceil = 4$.

4 CONCLUSIONS

If we are interested in the number of vertices in the set P, i.e. the number of protectors, the problem is very similar to the problem of finding a graph invariant 3, where we are looking for the minimal cardinality set of protectors P, needed to destroy every path of order 3. This means that vertices of the set $S = V(G) \backslash P$, called a dissociation set induces a subgraph with maximum degree 1, i.e. edges and isolated vertices, while in our problem set S consists only of isolated edges. Since determining ψ_k, for $k \geq 2$ was shown to be NP-hard problem in general and polynomial only for some special sets of graphs, it would be interesting to examine 3 or even k of hexagonal graphs.

References

[1] B. Brešar, F. Kardoš, J. Katreniš and G. Semanišin, Minimum k-path vertex cover. Discrete Appl. Math., 159, iss. 12 (2011), 1189–1195.

[2] K. Cameron: Induced matchings. Discrete Appl. Math., 24 (1989), 97–102.

[3] W. Duckworth, D. F. Manlove and M. Zito: On the approximability of the maximum induced matching problem. J. Discrete Alg., 3, vol 1 (2005), 79–91.

[4] G. Fricke and R. Laskar: Strong matchings on trees. Congr. Numer. 89 (1992) 239–243.

[5] M. C. Golumbic and M. Lewenstein: New results on induced matchings. Discrete Appl. Math. 101 (2000), 157–165.

[6] M. Jakovac and A. Taranenko: On the k-path vertex cover of some graph products. Discrete Math. 313, iss. 1 (2013), 94–100.

[7] I. Kanj, M. J. Pelsmajer, M. Schaefer and G. Xia: On the induced matching problem. J. Comp. Syst. Sci. 77(2011), 1058–1070.

[8] M. Mahdian: The strong chromatic index of graphs, M.Sc. Thesis. Department of Computer Science, University of Toronto, 2000.

[9] C. M. Krishnamurthy and R. Sritharan: Maximum induced matching problem on hhd-free graphs. Discrete Appl. Math. 160, iss.3 (2012), 224–230.

[10] H. Moser and S. Sikdar: The parameterized complexity of the induced matching problem in planar graphs. Proceedings of the 2007 International Frontiers of Algorithmics Workshop, in: Lect. Notes Comp. Sci., Springer-Verlag, Berlin, 2007.

[11] M. Novotny: Design and Analysis of a Generalized Canvas Protocol. Information Security Theory and Practice. Security and Privacy of Pervasive Systems and Smart Devices, LNCS 6033 (2010), 106–121.

[12] Y. Orlovich, A. Dolguib, G. Finkec, V. Gordond, F. Wernere: The complexity of dissociation set problems in graphs. Discrete Appl. Math. 159 (13) (2011) 13521366.

[13] J. Stockmeyer and V. V. Vazirani: NP-completeness of some generalization of the maximum matching problem. Inform. Process. Lett. 15 (1982), 14–19.

[14] P. Šparl, R. Witkowski and J. Žerovnik: 1-local 7/5-competitive algorithm for multicoloring hexagonal graphs. Algorithmica 64, no.4 (2012), 564–583.

[15] R. Erveš and P. Šparl: Maximum induced matching of hexagonal graphs. Submited.

[16] I. Sau Walls, P. Šparl, and J. Žerovnik: Simpler multicoloring of triangle-free hexagonal graphs. Discrete Math. 312, is. 1, 181–187.

[17] D. B. West: Strong edge-coloring, Open Problems - Strong edge coloring. http://www.math.uiuc.edu/˜west/openp/strongedge.html, accessed November 7th, 2012.

[18] M. Yannakakis: Node-deletion problems on bipartite graphs. SIAM J. Computing 10 (1981) 310327.

[19] M. Zito: Maximum Induced Matching on Regular Graphs and Trees. In Proceedings of WG'99: the 25th International Workshop on Graph-Theoretic Concepts in Computer Science, volume 1665 of Lect. Notes Comp. Sci., Springer-Verlag, Berlin (1999), 89–100.

FAULT DIAMETER OF CARTESIAN GRAPH BUNDLES

Rija Erveš[a,b] and **Janez Žerovnik**[a,c]
[a]Institute of mathematics, physics and mechanics
Jadranska 19, 1000 Ljubljana, Slovenia
[b]University of Maribor, Faculty of Civil Engineering
Smetanova 17, 2000 Maribor, Slovenia
[c]University of Ljubljana, Faculty of Mechanical Engineering
Aškerčeva 6, 1000 Ljubljana, Slovenia
rija.erves@um.si, janez.zerovnik@imfm.uni-lj.si

Abstract: The mixed fault diameter of a graph G, $D_{(a,b)}(G)$, is the maximal diameter among all subgraphs of G obtained by deleting any of its a vertices and b edges. Special cases are the (vertex) fault diameter $D^V_a(G) = D_{(a,0)}(G)$ and the edge fault diameter $D^E_a(G) = D_{(0,a)}(G)$. Let G be a Cartesian graph bundle with fibre F over the base graph B, and let $0 < a < \kappa(F)$, and $0 < b < \kappa(B)$. We recall some results on fault diameters and in particular state without proof the new result that $D^V_{a+b+1}(G) \leq D^V_a(F) + D^V_b(B)$ if $D_{(a-1,1)}(F) \leq D^V_a(F)$ and $D_{(b-1,1)}(B) \leq D^V_b(B)$ hold.

Keywords: vertex fault diameter, mixed connectivity, mixed fault diameter, Cartesian graph bundle, Cartesian graph product, interconnection network, fault tolerance.

1 INTRODUCTION

In the design of large interconnection networks several factors have to be taken into account. A usual constraint is that each processor can be connected to a limited number of other processors and that the delays in communication must not be too long. Extensively studied network topologies in this context include graph products and bundles. For example meshes, tori, hypercubes and some of their generalizations are Cartesian products. It is less known that some other well-known interconnection network topologies are Cartesian graph bundles, for example twisted hypercubes [9, 12] and multiplicative circulant graphs [21].

Furthermore, an interconnection network should be fault tolerant, because practical communication networks are exposed to failures of network components. Both failures of nodes and failures of connections between them happen and it is desirable that a network is robust in the sense that a limited number of failures does not break down the whole system. A lot of work has been done on various aspects of network fault tolerance, see for example the survey [8] and the more recent papers [16, 22, 25]. In particular the fault diameter with faulty vertices, which was first studied in [17], and the edge fault diameter have been determined for many important networks recently [1–4, 10, 11, 18, 23]. Usually either only edge faults or only vertex faults are considered, while the case when both edges and vertices may be faulty is studied rarely.

In recent work on fault diameter of Cartesian graph products and bundles [1–4], analogous results were found for both fault diameter and edge fault diameter. However, the proofs for vertex and edge faults are independent, and our effort to see how results in one case may imply the others was not successful. A natural question is whether it is possible to design a uniform theory that covers simultaneous faults of vertices and edges. Some basic results on edge, vertex and mixed fault diameters for general graphs appear in [5]. Mixed connectivity which generalizes both vertex and edge connectivity, and some basic observations for any connected graph are given in [13]. We are not aware of any earlier work on mixed connectivity. A closely related notion is the connectivity pairs of a graph [7] but the claimed proof of generalized Menger's theorem is not valid as showed in [19].

The concept of fault diameter of Cartesian product graphs was first described in [17], but the upper bound was wrong, as shown by Xu, Xu and Hou who provided a small counter

example and corrected the mistake [23]. More precisely, denote by $D^V_a(G)$ the fault diameter of a graph G, a maximum diameter among all subgraphs of G obtained by deleting any of its a vertices, and $G\Box H$ the Cartesian product of graphs G and H. Xu, Xu and Hou proved [23]

$$D^V_{a+b+1}(G\Box H) \leq D^V_a(G) + D^V_b(H) + 1,$$

while the claimed bound in [17] was $D^V_a(G) + D^V_b(H)$. (Our notation here slightly differs from notation used in [17, 23].) The result was later generalized to graph bundles in [1] and generalized graph products (as defined by [8]) in [24].

In most cases of Cartesian graph bundles the bound can indeed be improved to the one claimed in [17]. Methods used involve the theory of mixed connectivity and recent results on mixed fault diameters of Cartesian graph bundles [5, 13–15].

2 MIXED CONNECTIVITY AND MIXED FAULT DIAMETER

A graph is *connected* if there is a path between each pair of vertices, and is *disconnected* otherwise. The *connectivity* (or *vertex connectivity*) $\kappa(G)$ of a connected graph G, other than a complete graph, is the smallest number of vertices whose removal disconnects G. For complete graph is $\kappa(K_n) = n - 1$. We say that G is *k-connected* (or *k-vertex connected*) for any $k \leq \kappa(G)$. The *edge connectivity* $\lambda(G)$ of a connected graph G, is the smallest number of edges whose removal disconnects G. A graph G is said to be *k-edge connected* for any $k \leq \lambda(G)$. It is well-known that $\kappa(G) \leq \lambda(G) \leq \delta_G$, where δ_G is the smallest vertex degree of G. Thus if a graph G is k-connected, then it is also k-edge connected. The reverse does not hold in general.

The *mixed connectivity* generalizes both vertex and edge connectivity [13, 14]. Note that the definition used in [14] and here slightly differs from the definition used in a previous work [13].

Definition 1 *Let G be any connected graph. A graph G is* (p,q)+connected, *if G remains connected after removal of any p vertices and any q edges.*

Any connected graph G is $(0,0)$+connected, $(p,0)$+connected for any $p < \kappa(G)$ and $(0,q)$+connected for any $q < \lambda(G)$. In our notation $(i,0)$+connected is the same as $(i+1)$-connected, i.e. the graph remains connected after removal of any i vertices. Similarly, $(0,j)$+connected means $(j+1)$-edge connected, i.e. the graph remains connected after removal of any j edges. Clearly, if G is a (p,q)+connected graph, then G is (p',q')+connected for any $p' \leq p$ and any $q' \leq q$. Furthermore, for any connected graph G with $k < \kappa(G)$ faulty vertices, at least k edges are not working. Roughly speaking, a graph G remains connected if any faulty vertex in G is replaced with a faulty edge. It is known [13, 14] that if a graph G is (p,q)+connected and $p > 0$, then G is $(p-1,q+1)$+connected. Hence for $p > 0$ we have a chain of implications: (p,q)+connected $\rightarrow$ $(p-1,q+1)$+connected $\rightarrow$ … $\rightarrow$ $(1,p+q-1)$+connected $\rightarrow$ $(0,p+q)$+connected, which generalizes the well-known proposition that any k-connected graph is also k-edge connected. Therefore, a graph G is (p,q)+connected if and only if $p < \kappa(G)$ and $p+q < \lambda(G)$. If for a graph G $\kappa(G) = \lambda(G) = k$, then G is (i,j)+connected exactly when $i + j < k$. However, if $2 \leq \kappa(G) < \lambda(G)$, the question whether G is (i,j)+connected for $1 \leq i < \kappa(G) \leq i + j < \lambda(G)$ is not trivial. It is interesting to note that in general the knowledge of $\kappa(G)$ and $\lambda(G)$ is not enough to decide whether G is (i,j)+connected [14].

The *distance* between vertices x and y, is the length of a shortest path between x and y in G. The *diameter* of a connected graph G, $D(G)$, is the maximum distance between any two vertices in G. The *a-fault diameter* (or *a-vertex fault diameter*) of a graph G, $D^V_a(G)$, is the maximum diameter among all subgraphs of G obtained by deleting any of its a vertices. The *a-edge fault diameter* of G, $D^E_a(G)$, is the maximum diameter among all subgraphs of G obtained by deleting any of its a edges. In particular, $D^E_0(G) = D^V_0(G) = D(G)$, the diameter of G. It is known [5] that for any connected graph G the inequalities below hold.

$$D(G) = D^E_0(G) \le D^E_1(G) \le D^E_2(G) \le \ldots \le D^E_{\lambda(G)-1}(G) < \infty.$$
$$D(G) = D^V_0(G) \le D^V_1(G) \le D^V_2(G) \le \ldots \le D^V_{\kappa(G)-1}(G) < \infty.$$

Definition 2 *Let G be a (p,q)+connected graph. The* (p,q)-mixed fault diameter *of G is*

$$D_{(p,q)}(G) = \max\{D(G \setminus X) \mid X = X_E \cup X_V, X_E \subseteq E(G), X_V \subseteq V(G), |X_V| = p, |X_E| = q\}.$$

The mixed fault diameter $D_{(p,q)}(G)$ is the largest diameter among the diameters of all subgraphs obtained from G by deleting any p vertices and any q edges, hence $D_{(0,0)}(G) = D(G)$, $D_{(0,a)}(G) = D^E_a(G)$ and $D_{(a,0)}(G) = D^V_a(G)$. In previous work [5] on vertex, edge and mixed fault diameters of connected graphs the following theorem has been proved.

Theorem 3 [5] *Let G be (p,q)+connected graph and $p > 0$.*
If $q > 0$, then

$$D^E_{p+q}(G) = D_{(0,p+q)}(G) \le D_{(1,p+q-1)}(G) \le \ldots \le D_{(p,q)}(G).$$

If $q = 0$, then

$$D^E_p(G) = D_{(0,p)}(G) \le D_{(1,p-1)}(G) \le \ldots \le D_{(p-1,1)}(G) \le D^V_p(G) + 1.$$

Note that for $(p+1)$-connected graph G, and $p > 0$, we have either $D_{(p-1,1)}(G) \le D^V_p(G)$ or $D_{(p-1,1)}(G) = D^V_p(G) + 1$. For example, complete graphs, complete bipartite graphs, and cycles are graphs with $D_{(p-1,1)}(G) = D^V_p(G) + 1$ for all meaningful of values of p. More examples of both types of graphs can be found in [5].

3 FAULT DIAMETERS OF CARTESIAN GRAPH BUNDLES

Cartesian graph bundles are a generalization of Cartesian graph products, first studied in [20]. Let G_1 and G_2 be graphs. The *Cartesian product* of graphs G_1 and G_2, $G = G_1 \square G_2$, is defined on the vertex set $V(G_1) \times V(G_2)$. Vertices (u_1,v_1) and (u_2,v_2) are adjacent if either $u_1u_2 \in E(G_1)$ and $v_1 = v_2$ or $v_1v_2 \in E(G_2)$ and $u_1 = u_2$.

Let B and F be graphs. A graph G is the *Cartesian graph bundle with fibre F over the base graph B* if there is a graph map $p : G \to B$, such that for each vertex $v \in V(B)$, $p^{-1}(\{v\})$ is isomorphic to F, and for each edge $e = uv \in E(B)$, $p^{-1}(\{e\})$ is isomorphic to $F \square K_2$.

In recent work on fault diameter of Cartesian graph products and bundles [1–4], analogous results were found for both fault diameter and edge fault diameter.

Theorem 4 [1] *Let F and B be k_F-connected and k_B-connected graphs respectively, $0 \le a < k_F$, $0 \le b < k_B$, and G a Cartesian bundle with fibre F over the base graph B. Then*

$$D^V_{a+b+1}(G) \le D^V_a(F) + D^V_b(B) + 1.$$

Theorem 5 [4] *Let F and B be k_F-edge connected and k_B-edge connected graphs respectively, $0 \le a < k_F$, $0 \le b < k_B$, and G a Cartesian bundle with fibre F over the base graph B. Then*

$$D^E_{a+b+1}(G) \le D^E_a(F) + D^E_b(B) + 1.$$

Before writing theorems on bounds for the mixed fault diameter we recall a theorem on mixed connectivity.

Theorem 6 [13] *Let G be a Cartesian graph bundle with fibre F over the base graph B, graph F be (p_F,q_F)+connected and graph B be (p_B,q_B)+connected. Then Cartesian graph bundle G is (p_F+p_B+1,q_F+q_B)+connected.*

In recent work [14, 15], an upper bound for the mixed fault diameter of Cartesian graph bundles, $D_{(p+1,q)}(G)$, in terms of mixed fault diameter of the fibre and diameter of the base graph and in terms of diameter of the fibre and mixed fault diameter of the base graph, respectively, is given.

Theorem 7 [14] *Let G be a Cartesian graph bundle with fibre F over the base graph B, where graph F is (p,q)+connected, $p + q > 0$, and B is a connected graph with diameter $D(B) > 1$. Then we have:*
if $q > 0$, then

$$D_{(p+1,q)}(G) \le D_{(p,q)}(F) + D(B),$$

if $q = 0$, then

$$D^V_{p+1}(G) \le \max\{D^V_p(F), D_{(p-1,1)}(F)\} + D(B).$$

Theorem 8 [15] *Let G be a Cartesian graph bundle with fibre F over the base graph B, graph F be a connected graph with diameter $D(F) > 1$, and graph B be (p,q)+connected, $p + q > 0$. Then we have:*
if $q > 0$, then

$$D_{(p+1,q)}(G) \le D(F) + D_{(p,q)}(B),$$

if $q = 0$, then

$$D^V_{p+1}(G) \le D(F) + \max\{D^V_p(B), D_{(p-1,1)}(B)\}.$$

Theorems 7 and 8 improve results 4 and 5 for $a > 0$, $b = 0$, and $a = 0$, $b > 0$, respectively. However, results in [14] address only the number of faults given by the connectivity of the fibre (plus one vertex), while the connectivity of the graph bundle can be much higher when the connectivity of the base graph is substantial, and results in [15] address only the number of faults given by the connectivity of the base graph (plus one vertex), while the connectivity of the graph bundle can be much higher when the connectivity of the fibre is substantial. An upper bound for the mixed fault diameter that would take into account both types of faults remains to be an interesting open research problem.

In the case when $a = b = 0$ the fault diameter is determined exactly [14]. Let graphs F and B be connected graphs with diameters $D(F) > 1$ and $D(B) > 1$, and let G be a Cartesian graph bundle with fibre F over the base graph B. Then

$$D^V_1(G) = D^E_1(G) = D(G) = D(F) + D(B).$$

In other words, the diameter of a nontrivial Cartesian graph bundle does not change when one element is faulty.

4 IMPROVED UPPER BOUND FOR VERTEX FAULT DIAMETER OF CARTESIAN GRAPH BUNDLES

Theorem 9 *Let G be a Cartesian graph bundle with fibre F over the base graph B, graphs F and B be k_F-connected and k_B-connected respectively, and let $0 < a < k_F$, $0 < b < k_B$. If for fault diameters of graphs F and B, $D_{(a-1,1)}(F) \leq D^V_a(F)$ and $D_{(b-1,1)}(B) \leq D^V_b(B)$ hold then*

$$D^V_{a+b+1}(G) \leq D^V_a(F) + D^V_b(B).$$

The proof is omitted due to space limitations and will appear elsewhere. Theorem 9 improves Theorem 4 on the class of Cartesian graph bundles for which both, the fiber and the base graph, are at least 2-connected. Theorem 9 also improves result of [23] on the Cartesian graph products with at least 2-connected factors. The next example shows that the bound of Theorem 9 is tight.

Example 10 *Let $F = B = K_4 \setminus \{e\}$. Then graph F is 2-connected and $D^E_1(F) = D^V_1(F) = 2$. The vertex fault diameter of Cartesian graph product $F \square F$ on Fig. 1 is*

$$D^V_3(F \square F) = D^V_1(F) + D^V_1(F) = 4.$$

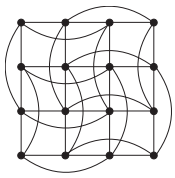


Figure 1: Cartesian graph product of two factors $K_4 \setminus \{e\}$.

Example 11 *Cycle C_4 is 2-connected graph and $D^E_1(C_4) = D^V_1(C_4) + 1 = 3$. The vertex fault diameter of Cartesian graph bundle G with fibre C_4 over base graph C_4 on Fig. 2 is*

$$D^V_3(G) = D^V_1(C_4) + D^V_1(C_4) + 1 = 5.$$

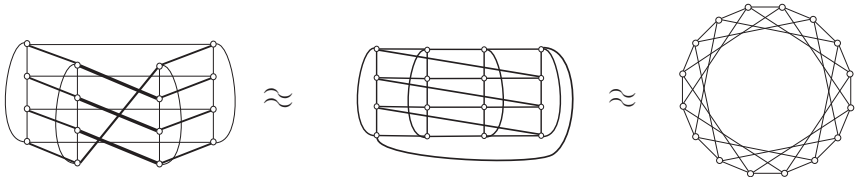


Figure 2: Twisted torus: Cartesian graph bundle with fibre C_4 over base C_4.

It is interesting to note that graph bundles also appear as computer topologies. A well known example is the twisted torus on Fig. 2. Cartesian graph bundle with fibre C_4 over base C_4 is the ILLIAC IV architecture [6], a famous supercomputer that inspired some modern multicomputer architectures. It may be interesting to note that the original design was a graph bundle with fibre C_8 over base C_8, but due to high cost a smaller version was build [26].

References

[1] I. Banič, J. Žerovnik, Fault-diameter of Cartesian graph bundles, Inform. Process. Lett. 100 (2006) 47–51.

[2] I. Banič, J. Žerovnik, Edge fault-diameter of Cartesian product of graphs, LNCS 4474 (2007) 234–245.

[3] I. Banič, J. Žerovnik, Fault-diameter of Cartesian product of graphs, Adv. Appl. Math. 40 (2008) 98–106.

[4] I. Banič, R. Erveš, J. Žerovnik, The edge fault-diameter of Cartesian graph bundles, Eur. J. Combin. 30 (2009) 1054–1061.

[5] I. Banič, R. Erveš, J. Žerovnik, Edge, vertex and mixed fault diameters, Adv. Appl. Math. 43 (2009) 231–238.

[6] G. H. Barnes, R. M. Brown, M. Kato, D. J. Kuck, D. L. Slotnick, R. A. Stokes, The ILLIAC IV Computer, IEEE Transactions on Computers, pp. 746–757, August, 1968.

[7] L. W. Beineke, F. Harary, The connectivity function of a graph, Mathematika 14 (1967) 197–202.

[8] J. C. Bermond, N. Honobono, C. Peyrat, Large Fault-tolerant Interconnection Networks, Graph. Combin. 5 (1989) 107–123.

[9] P. Cull, S. M. Larson, On generalized twisted cubes, Inform. Process. Lett. 55 (1995) 53–55.

[10] K. Day, A. Al-Ayyoub, Minimal fault diameter for highly resilient product networks, IEEE Trans. Parallel. Distr. Syst. 11 (2000) 926–930.

[11] D. Z. Du, D. F. Hsu, Y. D. Lyuu, On the diameter vulnerability of Kautz digraphs, Discrete Math. 151 (2000) 81–85.

[12] K. Efe, A variation on the hypercube with lower diameter, IEEE Trans. Comput. 40 (1991) 1312–1316.

[13] R. Erveš, J. Žerovnik, Mixed connectivity of Cartesian graph products and bundles, to appear in Ars Combinatoria, arXiv:1002.2508v1 [math.CO] (2010).

[14] R. Erveš, J. Žerovnik, Mixed fault diameter of Cartesian graph bundles, Discrete Appl. Math. 161 (2013) 1726–1733.

[15] R. Erveš, J. Žerovnik, Mixed fault diameters of Cartesian graph bundles II, to appear in Ars Math. Contemp.

[16] C. H. Hung, L. H. Hsu, T. Y. Sung, On the Construction of Combined *k*-Fault-Tolerant Hamiltonian Graphs, Networks 37 (2001) 165–170.

[17] M. Krishnamoorthy, B. Krishnamurty, Fault diameter of interconnection networks, Comput. Math. Appl. 13 (1987) 577–582.

[18] S. C. Liaw, G. J. Chang, F. Cao, D. F. Hsu, Fault-tolerant routing in circulant networks and cycle prefix networks, Ann. Comb. 2 (1998) 165–172.

[19] W. Mader, Connectivity and edge-connectivity infinite graphs, Surveys in Combinatorics, London Math. Soc. Lecture Notes 38 (1979), 66–95.

[20] T. Pisanski, J. Shawe-Taylor, J. Vrabec, Edge-colorability of graph bundles, J. Comb. Theory Ser. B 35 (1983) 12–19.

[21] I. Stojmenović, Multiplicative circulant networks: Topological properties and communication algorithms, Discrete Appl. Math. 77 (1997) 281–305.

[22] C. M. Sun, C. N. Hung, H. M. Huang, L. H. Hsu, Y. D. Jou, Hamiltonian Laceability of Faulty Hypercubes, Journal of Interconnection Networks 8 (2007) 133–145.

[23] M. Xu, J.-M. Xu, X.-M. Hou, Fault diameter of Cartesian product graphs, Inform. Process. Lett. 93 (2005) 245–248.

[24] J.-M. Xu, C. Yang, Fault diameter of product graphs, Information Processing Letters 102 (2007) 226–228.

[25] J. H. Yin, J. S. Li, G. L. Chen, C. Zhong, On the Fault-Tolerant Diameter and Wide diameter of ω-Connected Graphs, Networks 45 (2005) 88–94.

[26] http://www.computermuseum.li/Testpage/Illiac-IV-1960

THE RELIABILITY HOSOYA-WIENER POLYNOMIAL

Darja Rupnik Poklukar and **Janez Žerovnik**
University of Ljubljana, Faculty of Mechanical Engineering
Aškerčeva 6, SI-1000 Ljubljana, Slovenia
darja.rupnik@fs.uni-lj.si, janez.zerovnik@fs.uni-lj.si

Abstract: Assuming that, in a communication network, the weights of the edges quantify the volume or the quality of the information transmitted by the nodes, the strength of a path, called the reliability of the path can be calculated as the product of the weights of the edges belonging to the paths. Considering only the most reliable path between each pair of nodes, it is shown that some of the well-known relations of the Hosoya-Wiener polynomial to the Wiener number generalize to weighted graphs.

Keywords: reliability Wiener number, reliability Hosoya-Wiener polynomial

1 Introduction

In the design of large interconnection networks several factors have to be taken into account. Optimal design is important both to achieve good performance and to reduce the cost of construction and maintenance. Practical communication networks are exposed to failures of network components. Both failures of nodes and failures of connections between them happen and it is desirable that a network is robust in the sense that a limited number of failures does not break down the whole system.

Communication networks are generally modeled by weighted digraphs. The weight associated with each edge is taken to be the probability of that edge being operational. A *reliability measure* on such a network is then the expected value (in probabilistic sense) of the connectivity of the graph.

Many topological indices have been defined and several of them have found applications as a means for modeling chemical and physical properties of molecules and also in communication networks [3, 4, 8, 9, 2].

The Wiener number of a graph is defined as the sum of distances between all pairs of vertices. In more than 60 years after H. Wiener discovered remarkable correlation between the value $W(G)$ of the molecular graph G and some chemical properties of the molecule [10], the Wiener number and related graph invariants have been very extensively studied. In the last 20 years, a remarkably large number of modifications and extensions of Wiener number was put forward [11, 3, 4]. The Hosoya-Wiener polynomial has the property that its first derivative evaluated at $x = 1$ equals the Wiener number.

In this paper we give some new definitions for weighted graphs: the reliability Wiener number, considering the most reliable path between each pair of nodes, and the reliability Hosoya-Wiener polynomial. We show some properties and relations between them.

2 Definitions

A *weighted graph* $G = (V, E, p, \lambda)$ is a combinatorial object consisting of an arbitrary set $V = V(G)$ of *vertices*, a set $E = E(G)$ of ordered pairs $\{u, v\} = uv = e$ of distinct vertices of G called *edges*, and two *weighting functions*, p and λ. The weight function $p : E(G) \mapsto [0, 1]$ is interpreted as the probability of edges being operational. (That is, $1 - p(e)$ is the probability that edge $e \in E(G)$ has failed.) The distance function $\lambda : E(G) \mapsto \mathbb{R}^+$ assigns positive real numbers (lengths) to edges. We assume that the vertices are completely reliable and that all edge failures are statistically independent.

Note: Alternatively, we can consider the complete graph and model non existing vertices by setting $p(e) = 0$.

The order and size of G are $n = |V(G)|$ and $m = |E(G)|$, respectively.

A *path* P between u and v is a sequence of distinct vertices $u = v_i, v_{i+1}, \ldots, v_{k-1}, v_k = v$ such that each pair $v_l v_{l+1}$ is connected by an edge. The *length* of the path P is the sum of the lengths of its edges,

$$\ell(P) = \sum_{l=i}^{k-1} \lambda(v_l, v_{l+1}).$$

The *distance* $d_G(u, v)$, or simpler $d(u, v)$, between vertices u and v in graph G is the length of a shortest path between u and v. If there is no such path, we write $d(u, v) = \infty$. The *diameter* of a graph G is the maximal distance in G, $\mathcal{D}(G) = \max_{u,v \in V(G)} d_G(u, v)$.

We can also define the *reliability* of path P with

$$p(P) = \prod_{l=i}^{k-1} p(v_l, v_{l+1}).$$

In the special case when all edges have distance 1, $\ell(P)$ is the number of edges in P. Of course, several routes from one vertex to another can exist. The maximum reliability between two vertices is reached using the path with maximum reliability. In [7], the notion of *reliability of a graph* was introduced by a version of Wiener number where they considered the most reliable path between each pair of vertices. Following this idea, we define:

For two vertices $u, v \in V(G)$ denote with $P_{\overrightarrow{uv}}$ the set of all directed paths from u to v. The weight of the most reliable path from u to v can be called the *reliability of* (u, v):

$$F_{\overrightarrow{uv}} = \max_{P \in P_{\overrightarrow{uv}}} \{\, p(P) \,\}.$$

We set $F_{\overrightarrow{uu}} = 0$ for all $u \in V(G)$.

Define

$$\begin{aligned}
R^+(u) &= \sum_{v \in V(G)} F_{\overrightarrow{uv}} && \text{the weighted out-reliability of vertex } u,\\
R^-(u) &= \sum_{v \in V(G)} F_{\overrightarrow{vu}} && \text{the weighted in-reliability of vertex } u,\\
W_{R^+}(G) &= \sum_{u \in V(G)} R^+(u) && \text{the out-reliability Wiener number of } G,
\end{aligned}$$

$$W_{R^-}(G) \;=\; \sum_{u\in V(G)} R^-(u) \qquad \text{the in-reliability Wiener number of } G.$$

Obviously, in the case of a graph G, $R^-(u) = R^+(u) =: R(u)$ and $W_{R^-}(G) = W_{R^+}(G)$, so we can define the *reliability Wiener number* by

$$W_R(G,\lambda,p) = \frac{1}{2}\sum_{u\in V(G)} R(u) = \frac{1}{2}\sum_{u\in V(G)}\sum_{v\in V(G)} F_{\overrightarrow{uv}}. \tag{1}$$

The reliability Wiener number of G is a measure of the capacity of the vertices of G of transmitting information in a reliable form, where the information is transmitted through the most reliable path. As suggested in [7], the problem of finding $F_{\overrightarrow{uv}}$ can be solved by using Dijkstra's algorithm on a weighted digraph $G' = (V, E, -\ln p, \lambda)$.

3 Properties of reliability Wiener number

We can show some properties of reliability Wiener number, defined by (1):

(a) If all weights p are equal to 1 (i.e. all edges are working without possibility to fail) and the graph is complete, then $W_R(G)$ is equal to $\frac{1}{2}n(n-1) = \binom{n}{2}$ and this is obviously the upper bound on the reliability Wiener number:

(b) $W_R(G) \le \binom{n}{2} \le W(G)$, the original Wiener number, defined as the sum of all distances in graph.

(c) If $G = P_n$ is a path with n vertices $v_1,\ v_2,\ ...v_n$ and $p_i = p(v_i, v_{i+1}),\ i = 1, ...n-1$, then

$$W_R(P_n) = \sum_{i=1}^{n-1} p_i + \sum_{i=1}^{n-2} p_i\, p_{i+1} + \sum_{i=1}^{n-3} p_i\, p_{i+1}\, p_{i+2} + \cdots + p_1\, p_2\, \cdots p_{n-1}.$$

(d) If $G = P_n$ is a path with n vertices and all link probabilities are equal, say $p(e) = p_0$ for all $e \in E(G)$, where $0 < p_0 < 1$ is a constant, then

$$W_R(P_n) = (n-1)\,\frac{p_0}{1-p_0} - \left(\frac{p_0}{1-p_0}\right)^2 (1-p_0^{n-1}).$$

Proof:

$$\begin{aligned} W_R(P_n) &= \sum_{i=1}^{n-1} p_0 + \sum_{i=1}^{n-2} p_0^2 + \sum_{i=1}^{n-3} p_0^3 + \cdots + p_0^{n-1} \\ &= (n-1)p_0 + (n-2)p_0^2 + (n-3)p_0^3 + \cdots + p_0^{n-1} \\ &= \sum_{k=1}^{n-1} k\, p_0^{n-k} = \sum_{l=1}^{n-1}\sum_{k=1}^{n-l} p_0^k = \sum_{l=1}^{n-1} p_0 \frac{1-p_0^{n-l}}{1-p_0} \\ &= (n-1)\,\frac{p_0}{1-p_0} - \left(\frac{p_0}{1-p_0}\right)^2 (1-p_0^{n-1}). \end{aligned}$$

4 The reliability Hosoya-Wiener polynomial

A notion closely related to Wiener number is the Hosoya-Wiener polynomial of a graph G which is defined as

$$H(\lambda; x) = H(G, \lambda; x) = \sum_{u,v \in V(G)} x^{d(u,v)}.$$

This definition slightly differs from the definition used by Hosoya [6]:

$$\hat{H}(\lambda; x) = \hat{H}(G, \lambda; x) = \sum_{u,v \in V(G); u \neq v} x^{d(u,v)}. \tag{2}$$

Obviously, $H(\lambda; x) = \hat{H}(\lambda; x) + |V(G)|$.

We define the reliability Hosoya-Wiener polynomial for connected graphs as follows:

$$\hat{H}_R(G, \lambda, p\,; x) = \frac{1}{2} \sum_{u,v \in V(G); u \neq v} \frac{F_{\overrightarrow{uv}}}{d(u,v)}\, x^{d(u,v)}. \tag{3}$$

Note: $\hat{H}_R(G, \lambda, p\,; x)$ may not be a polynomial if edge lengths are allowed to be arbitrary real numbers. Obviously, if natural numbers are used for edge lengths, the function $\hat{H}_R(G, \lambda, p\,; x)$ is a polynomial. Hence, with appropriate scaling factor, we can always consider $\hat{H}_R(G, \lambda, p\,; x)$ to be a polynomial, for any model using rational edge lengths.

The Hosoya-Wiener polynomial has many interesting properties [6, 3, 9], perhaps the most interesting of them is that its derivative at 1 equals the Wiener number. In this work we generalize some of this results to reliability Hosoya-Wiener polynomial, summarized below.

Theorem 1

(a) $\hat{H}_R(G, \lambda, p\,; 0) = 0$,

(b) $\hat{H}_R(G, \lambda, p\,; 1) = \frac{1}{2} \sum_{u,v \in V(G); u \neq v} \frac{F_{\overrightarrow{uv}}}{d(u,v)}$,

(c) $\hat{H}'_R(G, \lambda, p\,; 1) = W_R(G)$.

Proof: Since $\hat{H}_R(G, \lambda, p\,; x)$ has been defined for connected graphs, $0 < d(u,v) < \infty$, $u \neq v$, (a) and (b) are obvious. Clearly

$$\hat{H}'_R(G, \lambda, p\,; x) = \frac{1}{2} \sum_{u,v \in V(G); u \neq v} F_{\overrightarrow{uv}}\, x^{d(u,v)-1},$$

which is equal to $W_R(G)$ if evaluated at $x = 1$.

References

[1] J. M. Aldous, R. J. Wilson, Graphs and Applications: An introductory Approach, Springer, Berlin, 2000.

[2] R. Erveš, D. Rupnik Poklukar, J. Žerovnik, On vulnerability measures of networks. In: Z. Babić (Ed.). 14th International Conference on Operational Research, Trogir, Croatia, vol. 4 (2013) 318-333.

[3] I. Gutman and B. Furtula (Eds.) Distance in Molecular Graphs - Theory, Univ. Kragujevac, Kragujevac, 2012.

[4] I. Gutman and B. Furtula (Eds.) Distance in Molecular Graphs - Applications, Univ. Kragujevac, Kragujevac, 2012.

[5] I.Gutman, S.Klavžar, M.Petkovšek and P.Žigert, On Hosoya polynomials of benzenoid graphs, MATCH Commun. Math. Comput. Chem. 43 (2001) 49-66.

[6] H.Hosoya, On some counting polynomials in chemistry, Discrete Appl. Math. 19 (1988) 239-257.

[7] J.A.Rodríguez-Velázquez, A.Kamišalić, J.Domingo-Ferrer, On reliability indices of communication networks, Comput. Math. Appl. 58 (2009) 1433-1440.

[8] B.Zmazek and J.Žerovnik, Computing the weighted Wiener and Szeged number on weighted cactus graphs in linear time, Croat. Chem. Acta 76 (2003) 137-143.

[9] B. Zmazek, J. Žerovnik, On generalization of the Hosoya-Wiener polynomial, MATCH Commun. Math. Comput. Chem. 55 (2006) 359–362.

[10] H.Wiener, Structural determination of paraffin boiling points, J. Amer. Chem. Soc. **69** (1947) 17-20.

[11] Discrete Mathematics 1997, vol. 80, Special issue on the Wiener index.

ON THE STRUCTURE OF LUCAS CUBES

Andrej Taranenko
University of Maribor, Faculty of Natural Sciences and Mathematics
Koroška cesta 160, SI-2000 Maribor, Slovenia
andrej.taranenko@uni-mb.si

Abstract: Lucas cubes are induced subgraphs of hypercubes obtained by excluding from the hypercube's vertex set all binary strings with two consecutive ones, as well as with one in the first and the last position. They are closely related to Fibonacci cubes. It is well known, that a Lucas cube of order n consists of two Fibonacci cubes of order $n-1$ and $n-3$ with additional edges between them. We characterize Lucas cubes based on peripheral expansions of a unique convex subgraph of an appropriate Fibonacci cube. This serves as the foundation for a recognition algorithm of Lucas cubes that runs in linear time.

Keywords. Lucas cubes, characterization, recognition algorithm.

1 INTRODUCTION

This is a shortened version of the paper [12]. The results here are stated without proofs, for the full version which includes the proof we refer the reader to [12].

Graphs based on binary strings are used as models for interconnection networks. Hypercube, being a popular interconnection scheme for multicomputers, has often been replaced by other models with similar properties comparable to those of hypercubes, where the number of vertices and edges in these alternative models does not increase as rapidly.

As such a model Fibonacci cubes have been defined in [3, 4], followed by extended Fibonacci cubes [16] and Lucas cubes [1]. Various studies have been made on the structure and different properties of Fibonacci cubes, Lucas cubes as well as Fibonacci-like cubes, see [6, 8, 9, 11, 13], to name a few. We also refer to the extensive survey on Fibonacci cubes by Klavžar [7]. Both, Fibonacci and Lucas cubes, also appear in connection with resonance graphs in chemistry [13, 17].

A *Fibonacci string* of length n is a binary string $b_0b_1 \dots b_{n-1}$ such that $b_i \cdot b_{i+1} = 0$, for $i = 0, 1, \dots n-2$. Equivalently, it is a binary string of length n without two consecutive ones.

The *Fibonacci cube* Γ_n of order n has the Fibonacci strings as vertices, with two vertices being adjacent whenever they differ in exactly one coordinate. We also set $\Gamma_0 = K_1$. See Fig. 1 for Fibonacci cubes of order n, for $n = 0, 1, \dots, 5$, with appropriate Fibonacci strings assigned to the vertices (these are omitted for Γ_5 for the clarity of the figure).

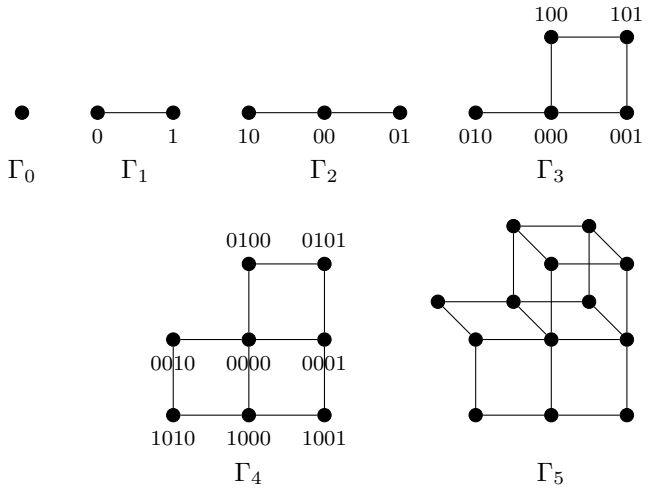


Figure 1: Fibonacci cubes Γ_n, for $n = 0, 1, \dots, 5$.

A *Lucas string* of length n is a binary string $b_0b_1 \dots b_{n-1}$ such that $b_i \cdot b_{i+1} = 0$, for $i = 0, 1, \dots, n-1$, where indices are computed modulo n. In other words, Lucas string is a binary string without two consecutive ones and without one both in the first and last position.

The *Lucas cube* Λ_n of order n is the graph with Lucas strings as vertices, again, with two vertices being adjacent whenever they differ in exactly one coordinate. We set $\Lambda_0 = K_1$. See Fig. 2 for Lucas cubes of order n, for $n = 0, 1, \ldots, 5$, with appropriate Lucas strings assigned to the vertices (again, these are omitted for Λ_5 for the clarity of the figure). Note, that vertices of Λ_n can be obtained from the vertices of Fibonacci cubes Γ_{n-1} and Γ_{n-3} as follows: $V(\Lambda_n) = 0V(\Gamma_{n-1}) \cup 10V(\Gamma_{n-3})0$ [11].

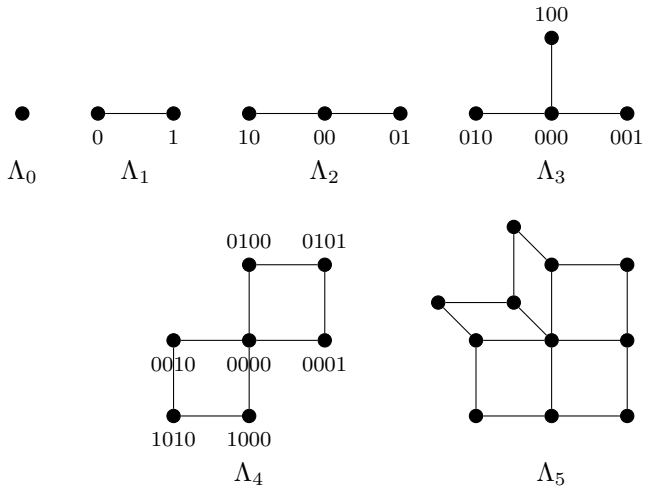


Figure 2: Lucas cubes Λ_n, for $n = 0, 1, \ldots, 5$.

It is known that Fibonacci cubes and Lucas cubes are median graphs [6]. For median graphs several efficient recognition algorithms are known. To mention the first polynomial algorithm with complexity $O(nm)$ presented in [5] and the much more advanced $O((m \log n)^{1.41})$ algorithm stated in [2]. This raises the natural question whether a faster recognition algorithm for special families of median graphs exists. For Fibonacci cubes this question was partially answered by presenting a recognition algorithm with the complexity $O(m \log n)$ [13]. Even more, Vesel recently answered the problem of existence of a linear recognition algorithm for Fibonacci cubes, presented in [7]. For Lucas cubes no recognition algorithm exists to our knowledge and is presented in this paper.

The paper is structured as follows. The next section contains some basic definitions and results concerning median graphs and Lucas cubes. In Section 3 several structural properties and a characterization of Lucas cubes are given. Using these results in the final Section 4 a recognition algorithm for Lucas cubes is presented. The presented algorithm correctly recognizes a Lucas cube in linear time.

2 PRELIMINARIES

The *hypercube* Q_n of order n is the graph with the vertex set containing all binary strings of length n, where two vertices are adjacent whenever the two strings differ in exactly one position. Isometric graphs of hypercubes are called *partial cubes*.

The *Fibonacci numbers* form a sequence of positive integers F_n, where $F_0 = 0$, $F_1 = 1$ and for $n \geq 2$ satisfy the recurrence $F_n = F_{n-1} + F_{n-2}$.

Since a Lucas string is a binary string without two consecutive ones and without one both in the first and last position, we can say that a Lucas string contains no two consecutive ones in a circular manner. Moreover, in the rest of the paper, when dealing with Lucas strings, we will always compute indices modulo n, even when not explicitly stated.

For a triple of vertices u, v and w of a given graph G, a vertex x of G is a *median* of u, v and w if x lies simultaneously on shortest paths joining u and v, u and w, and v and w, respectively. If G is connected and every triple of vertices admits a unique median, then G is a *median graph*. It is well known that median graphs are partial cubes (cf. [2]).

Let G be a connected graph with $e = xy$ and $f = uv$ two edges in G. We say that e is in relation Θ to f if $d(x, u) + d(y, v) \neq d(x, v) + d(y, u)$. Θ is reflexive and symmetric, but need not be transitive. We denote its transitive closure by Θ^*. It was proved in [15] that G is a partial cube if and only if G is bipartite and $\Theta = \Theta^*$.

For $X \subseteq V(G)$ we denote the subgraph of G induced by the set X with $G[X]$.

A subgraph H of a graph G is called *convex*, if it is connected and if any shortest path of G between two vertices of H is completely in H.

Let H be a fixed subgraph of a graph G, $H \subseteq G$. The *peripheral expansion* $\mathrm{pe}(G; H)$ of G with respect to H is the graph obtained from the disjoint union of G and an isomorphic copy of H, in which every vertex of the copy of H is joined by an edge with the corresponding vertex of H.

For an edge ab of graph G we define:

- $W_{ab} = \{w \in V(G) \mid d(a,w) < d(b,w)\}$
- $W_{ba} = \{w \in V(G) \mid d(b,w) < d(a,w)\}$
- $F_{ab} = \{xy \in E(G) \mid x \in W_{ab} \text{ and } y \in W_{ba}\}$
- $U_{ab} = \{w \in W_{ab} \mid w \text{ is the end vertex of an edge in } F_{ab}\}$
- $U_{ba} = \{w \in W_{ba} \mid w \text{ is the end vertex of an edge in } F_{ab}\}$

Let ab be an edge of a median graph G for which $U_{ab} = W_{ab}$. Then $G[W_{ab}]$ is called a *peripheral subgraph* of G. A Θ-class E of a median graph G is called *peripheral*, if at least one of $G[W_{ab}]$ and $G[W_{ba}]$ is peripheral for $ab \in E$. E is *internal* if it is not peripheral.

We will need the following well known lemma, cf. [2].

Lemma 2.1. *Let $e = ab$ be an edge of a connected bipartite graph G. Then*

(i) $F_{ab} = \{f | f \in E(G), e\Theta f\}$,

(ii) $G \backslash F_{ab}$ *has exactly two components.*

The next theorem characterizes median graphs.

Theorem 2.2. *[10] Let ab be an edge of a connected, bipartite graph G. Then G is a median graph if and only if the following three conditions are satisfied:*

(i) $G[U_{ab}]$ *is convex in* $G[W_{ab}]$ *and* $G[U_{ba}]$ *in* $G[W_{ba}]$.

(ii) F_{ab} *is a matching defining an isomorphism between* $G[U_{ab}]$ *and* $G[U_{ba}]$.

(iii) $G[W_{ab}]$ *and* $G[W_{ba}]$ *are median graphs.*

3 CHARACTERIZATION

Before we present the theorem that characterizes Lucas cubes, we will state some properties concerning this family of graphs.

Proposition 3.1. *[9] Let E be a Θ-class of the Lucas cube Λ_n. Then $|E| = F_{n-1}$.*

Let 0^p denote a binary string of length $p \geq 0$ with all bits equal to zero, similarly 1^p is a binary string of length p with all bits equal to 1. Let x and y be two arbitrary binary strings, we write xy for the concatenation of x and y.

Obviously the vertex 0^n, has exactly n neighbours in Λ_n, moreover it is the only vertex of degree n in Λ_n [8].

Proposition 3.2. *[12] All neighbours of the vertex 0^n in the Lucas cube Λ_n of order $n \geq 3$ are of degree $n-2$.*

While it is known [11] that a Lucas cubes of order n is composed from two Fibonacci cubes of order $n-1$ and $n-3$ with some additional edges between the two, the next proposition characterizes, how these two Fibonacci cubes are induced.

Proposition 3.3. *[12] Let $a = 0^n \in V(\Lambda_n)$ and $ab \in E(\Lambda_n)$. Then the following hold:*

(i) $\Lambda_n[W_{ab}]$ *is isomorphic to* Γ_{n-1}.

(ii) $\Lambda_n[W_{ba}]$ *is isomorphic to* Γ_{n-3}.

Proposition 3.4. *[12] Let E be a Θ-class of the Lucas cube Λ_n. Then E is peripheral.*

From Theorem 2.2 and Propositions 3.3 and 3.4 we immediately obtain the following corollary.

Corollary 3.5. *[12] Let $a = 0^n \in V(\Lambda_n)$ and $ab \in E(\Lambda_n)$. Then the graph $\Lambda_n[U_{ab}]$ is isomorphic to Γ_{n-3}.*

Let H be a subgraph of a graph G. Then ∂H is the set of all edges xy of G with $x \in H$ and $y \notin H$.

Theorem 3.6. *[12] There exist exactly one convex subgraph $H \subseteq \Gamma_n$ isomorphic to Γ_{n-2}, such that the peripheral expansion $\mathrm{pe}(\Gamma_n; H)$ is isomorphic the Lucas cube Λ_{n+1}.*

The next theorem characterizes Lucas cubes.

Theorem 3.7. *[12] Let G be a connected bipartite graph, $a \in V(G)$ of degree n and $ab \in E(G)$. G is isomorphic to Λ_n if and only if the following conditions are upheld:*

(1) All the neighbours of a are of degree $n-2$.

(2) $G[U_{ab}]$ is convex in $G[W_{ab}]$.

(3) F_{ab} defines an isomorphism between $G[U_{ab}]$ and $G[U_{ba}]$.

(4) $G[W_{ab}]$ ($G[W_{ba}]$) is isomorphic to Γ_{n-1} (Γ_{n-3}).

(5) $W_{ba} = U_{ba}$.

Note that condition (1) of Theorem 3.7 is necessary, since b has been arbitrarily chosen. There exists a graph $G = \mathrm{pe}(\Gamma_n; H)$ such that H is a convex subgraph of Γ_n isomorphic to Γ_{n-2} where G is not isomorphic to Λ_{n+1}. See Fig. 3 for an example.

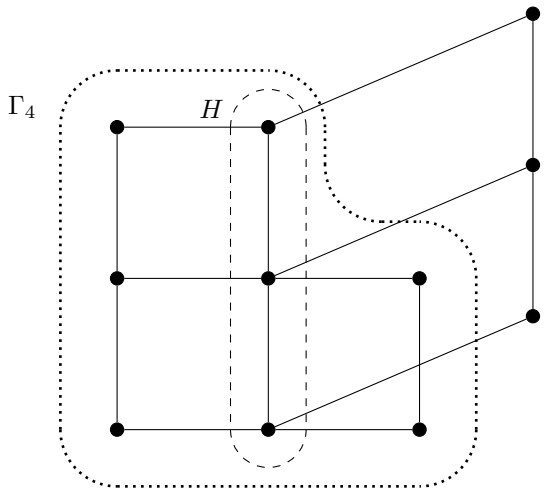


Figure 3: A peripheral expansion $\mathrm{pe}(\Gamma_4; H)$ where H is a convex subgraph isomorphic to Γ_2

4 RECOGNITION ALGORITHM

Theorem 3.7 is also good from algorithmic point of view since it serves as the basis for the algorithm presented with procedure LUCAS for recognition of Lucas cubes. Procedure `Fibonacci(G, n, uv)` used in procedure LUCAS is from Vesel [14] and returns ACCEPT if a given graph G is isomorphic to a Fibonacci cube of order n, and REJECT otherwise. The input parameter uv represents an edge of the input graph G with one end-vertex of degree n. Moreover, it runs in $O(|E(\Gamma_n)|)$ time.

Before calling the procedure LUCAS some preprocessing of the input graph is required. It is well known that $|V(\Gamma_h)| = F_{h+2}$. A given graph G, with $n = |V(G)|$ and $m = |E(G)|$, is examined only if $n = F_{h+1} + F_{h-1}$ for some $h \geq 1$ and it is bipartite, otherwise graph is rejected. The computed value of h is also an input parameter of the procedure LUCAS. The input graph fulfilling these conditions is declared isomorphic to a Lucas cube of order h if procedure LUCAS terminates without encountering REJECT statement.

Procedure LUCAS(G, h)

Result: ACCEPT if G is isomorphic to Λ_h, REJECT otherwise

begin
  if *G is K_1 or G is K_2* **then** ACCEPT

  Find $e = uv \in E(G)$ such that $d(u) = h$ and $d(v) = h - 2$
  if *e does not exist* **then** REJECT

  Compute sets W_{uv}, W_{vu}, U_{uv}, U_{vu} and F_{uv}
  if *F_{uv} is not a matching defining an isomorphism between $G[U_{uv}]$ and $G[U_{vu}]$* **then** REJECT

  if *$G[U_{uv}]$ is not convex in $G[W_{uv}]$* **then** REJECT

  if $U_{vu} \neq W_{vu}$ **then** REJECT

  $u' \in N_{W_{uv}}(u)$
  $v' \in N_{W_{vu}}(v)$
  if `Fibonacci(`$G[W_{uv}]$, $h-1$, uu'`)` *returns REJECT* **then** REJECT

  if `Fibonacci(`$G[W_{vu}]$, $h-3$, vv'`)` *returns REJECT* **then** REJECT

  Return ACCEPT
end

In the presented procedure we denote by $N_S(v)$ the set of all neighbours of a vertex v, where neighbours are from the set S .

Since the procedure LUCAS returns ACCEPT if and only if the conditions from Theorem 3.7 are satisfied, the following theorem follows immediately.

Theorem 4.1. *[12] Algorithm LUCAS correctly recognizes a Lucas cube.*

Theorem 4.2. *[12] Algorithm LUCAS runs in $O(|m|)$ time to successfully recognize a Lucas cube Λ_h.*

Acknowledgements

This work has been financed by ARRS Slovenia under the grant P1-0297 and within the EUROCORES Programme EUROGIGA (project GReGAS) of the European Science Foundation. The author is also with Institute of Mathematics, Physics and Mechanics, Ljubljana.

References

[1] E. Dedó, D. Torri, and N. Z. Salvi. The observability of the fibonacci and the lucas cubes. *Discrete Math.*, 255:55–63, 2002.

[2] R. Hammack, W. Imrich, and S. Klavžar. *Handbook of Product Graphs, second edition.* CRC Press, Boca Raton, 2011.

[3] W. J. Hsu. Fibonacci cubes – a new interconnection topology. *IEEE Trans. Parallel Distr. Systems*, 4:3–12, 1993.

[4] W. J. Hsu, C. V. Page, and J. S. Liu. Fibonacci cubes – a class of self-similar graphs. *Fibonacci Quart.*, 31:65–72, 1993.

[5] P. K. Jha and G. Slutzki. Convex-expansions algorithms for recognition and isometric embeddings of median graphs. *Ars Combin.*, 34:75–92, 1992.

[6] S. Klavžar. On median nature and enumerative properties of fibonacci-like cubes. *Discrete Math.*, 299:145–153, 2005.

[7] S. Klavžar. Structure of fibonacci cubes: a survey. *J. Comb. Optim.*, 25:505–522, 2013.

[8] S. Klavžar, M. Mollard, and M. Petkovšek. The degree sequence of fibonacci and lucas cubes. *Discrete Math.*, 311:1310–1322, 2011.

[9] S. Klavžar and I. Peterin. Edge-counting vectors, fibonacci cubes, and fibonacci triangle. *Publ. Math. Debrecen*, 71/3-4:267–278, 2007.

[10] H. M. Mulder. The structure of median graphs. *Discrete Math.*, 24:197–204, 1978.

[11] E. Munarini, C. P. Cippo, and N. Z. Salvi. On the lucas cubes. *Fibonacci Quart.*, 39:12–21, 2001.

[12] A. Taranenko. A new characterization and a recognition algorithm of lucas cubes. *submitted*, 2012.

[13] A. Taranenko and A. Vesel. Fast recognition of fibonacci cubes. *Algorithmica*, 49:81–93, 2007.

[14] A. Vesel. Linear recognition and embedding of fibonacci cubes. submitted, 2012.

[15] P. M. Winkler. Isometric embeddings in products of complete graphs. *Discrete Appl. Math.*, 7:221–225, 1984.

[16] J. Wu. Extended fibonacci cubes. *IEEE Trans. Parallel Distr. Systems*, 8:3–9, 1997.

[17] P. Žigert and M. Berlič. Lucas cubes and resonance graphs of cyclic polyphenanthrenes. *MATCH Commun. Math. Comput. Chem.*, 68:79–90, 2012.

EFFICIENT RECOGNITION OF FIBONACCI CUBES

Aleksander Vesel
Faculty of Natural Sciences and Mathematics, University of Maribor
Koroška cesta 160, SI-2000 Maribor, Slovenia
vesel@uni-mb.si

Abstract: Fibonacci strings are binary strings that contain no two consecutive 1s. The Fibonacci cube Γ_h is the subgraph of the hypercube of dimension h induced by the Fibonacci strings. These graphs are applicable as interconnection networks and in theoretical chemistry and lead to the Fibonacci dimension of a graph. We discuss efficient recognition algorithms for Fibonacci cubes.

Keywords: Fibonacci cube, partial cube, recognition algorithm.

1 INTRODUCTION

Hypercube is a popular interconnection scheme for multicomputers. Routing in a hypercube is a simple function of the Hamming distance between two nodes. That is, the message is successively sent along the connection corresponding to the bit position in both binary representations of the nodes with different values. The Fibonacci cube is a communication network that possesses many suitable properties which are important in network design and application. Its major advantage is that it uses fewer links than the comparable hypercube, while its size does not increase as fast as the hypercube's. In other words, they allow more alternatives to build networks of various sizes. Note also that the Fibonacci cube can emulate many hypercube algorithms. Moreover, they emulate other topologies, such as trees, rings and meshes very efficiently and can therefore find applications in fault-tolerant computing [6].

Fibonacci cubes with their extensive range of properties have appealed much attention in recent years and have been extensively investigated. Their structural and enumerative properties were studied in [1]. Very intriguing aspect is given by the fact that Fibonacci cubes are precisely the resonance graphs of fibonaccenes [2]. Beside the obvious consequence that Fibonacci cubes are median graphs, this property also induces a simple algorithm which recognizes in $O(m\,n)$ time whether a given graph on n vertices and m edges is a Fibonacci cube [4].

2 PRELIMINARIES

The *hypercube* of order $\boldsymbol{d}$, denoted by $\boldsymbol{Q_d}$, is the graph $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$ where the vertex set $\boldsymbol{V}(\boldsymbol{G})$ is the set of all binary strings $\boldsymbol{b_{n-1}} \ldots \boldsymbol{b_1 b_o}$. Two vertices $\boldsymbol{x}, \boldsymbol{y} \boldsymbol{\epsilon} \boldsymbol{V}(\boldsymbol{G})$ are adjacent in $\boldsymbol{Q_d}$, if and only if their Hamming distance equals one. Hypercubes $\boldsymbol{Q_1}$, $\boldsymbol{Q_2}$ and $\boldsymbol{Q_3}$ are depicted in Fig. 1.

A subgraph $\boldsymbol{H}$ of a graph $\boldsymbol{G}$ is isometric if $\boldsymbol{d_{H(u,v)}} = \boldsymbol{d_{G(u,v)}}$ for any pair of vertices $\boldsymbol{u}$ and $\boldsymbol{v}$ from $\boldsymbol{H}$. Isometric subgraphs of hypercubes are called *partial cubes*.

Let $\boldsymbol{G}$ be a connected graph and $\boldsymbol{e} = \boldsymbol{xy}$, $\boldsymbol{f} = \boldsymbol{uv}$ be two edges of $\boldsymbol{G}$. We say $\boldsymbol{e}$ is in relation $\boldsymbol{\Theta}$ to $\boldsymbol{f}$ if $\boldsymbol{d}(\boldsymbol{x}, \boldsymbol{u}) + \boldsymbol{d}(\boldsymbol{y}, \boldsymbol{v}) \neq \boldsymbol{d}(\boldsymbol{x}, \boldsymbol{v}) + \boldsymbol{d}(\boldsymbol{y}, \boldsymbol{u})$. $\boldsymbol{\Theta}$ is reflexive and symmetric, but need not be transitive. We denote its transitive closure by $\boldsymbol{\Theta}^*$. It well known that G is a partial cube if and only if $\boldsymbol{G}$ is bipartite and $\boldsymbol{\Theta}^* = \boldsymbol{\Theta}$.

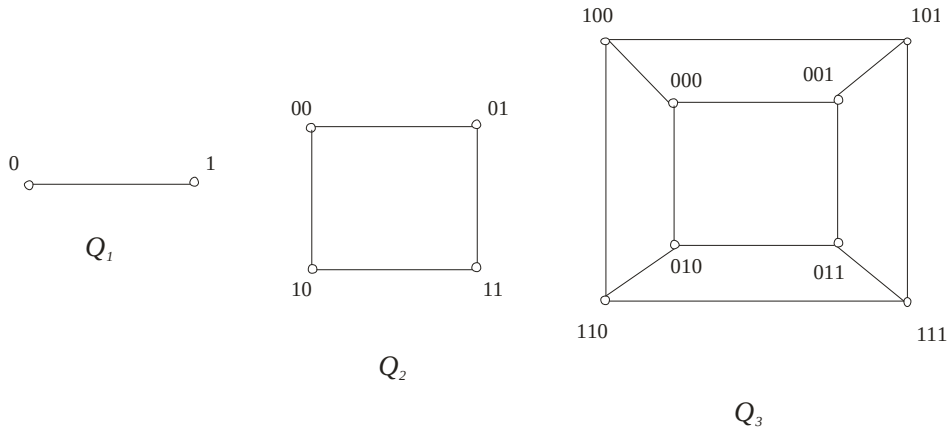


Figure 1: Hypercubes.

The *Fibonacci numbers* sequence of positive integers $\boldsymbol{F_n}$, where $\boldsymbol{F_0 = 0}$, $\boldsymbol{F_1 = 1}$ and for $\boldsymbol{n \geq 0}$ satisfy the recurrence $\boldsymbol{F_{n+2} = F_{n+1} + F_n}$. It is known that any natural number can be uniquely represented as a sum of Fibonacci numbers (Zeckendorf's Theorem). Assume that $\boldsymbol{i}$ is a positive integer such that $\boldsymbol{i \leq F_{n+2} - 1}$. Let $\boldsymbol{F(i) := b_{n-1} \ldots b_0}$ denote the order-$\boldsymbol{n}$ Fibonacci string of $\boldsymbol{i}$, where $\boldsymbol{i = \sum_{j=0}^{n-1} b_j F_{j+2}}$ and $\boldsymbol{b_j}$ is either 0 or 1, $0 \leq \boldsymbol{j \leq n - 1}$ with the condition $\boldsymbol{b_j b_{j+1} = 0}$.

The *Fibonacci cube* $\boldsymbol{\Gamma_h}$ is for $\boldsymbol{h > 0}$ defined as follows. The vertex set of $\boldsymbol{\Gamma_h}$ is the set $\boldsymbol{V = \{0, 1, \ldots, F_{h+2} - 1\}}$. Two vertices $\boldsymbol{x, y \epsilon V}$ are adjacent in $\boldsymbol{\Gamma_h}$ if and only if $\boldsymbol{H(F(x), F(y)) = 1}$. In other words, the vertices of $\boldsymbol{\Gamma_h}$ can be labeled with all binary strings $\boldsymbol{b_{n-1} \ldots b_1 b_o}$ containing no two consecutive ones; two vertices are adjacent if and only if their labels differ in precisely one bit. Fibonacci cubes $\boldsymbol{\Gamma_1}$, $\boldsymbol{\Gamma_2}$ and $\boldsymbol{\Gamma_3}$ are depicted in Fig. 1.

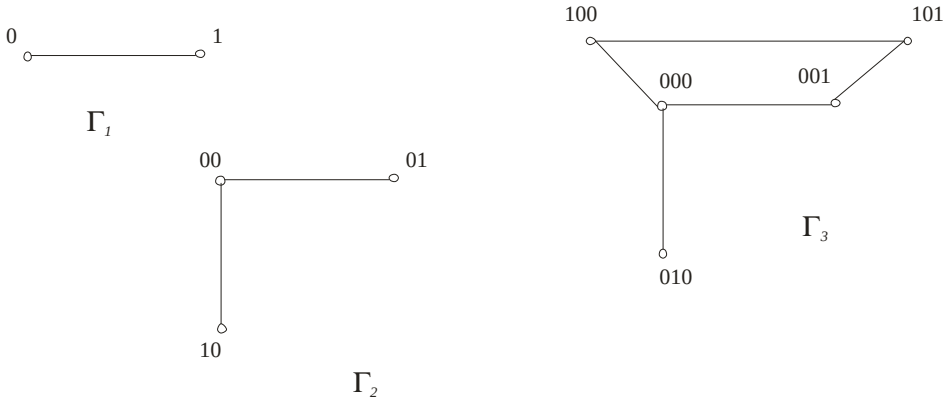


Figure 2: Fibonacci cubes.

For an edge $\boldsymbol{ab}$ in $\boldsymbol{E(G)}$ we write:

$\boldsymbol{W_{ab}} = \{\boldsymbol{w} \in \boldsymbol{V(G)} \mid \boldsymbol{d(a,w) < d(b,w)}\}$,
$\boldsymbol{W_{ba}} = \{\boldsymbol{w} \in \boldsymbol{V(G)} \mid \boldsymbol{d(b,w) < d(a,w)}\}$,
$\boldsymbol{F_{ab}} = \{\boldsymbol{xy} \mid \boldsymbol{xy}$ edge in $\boldsymbol{E(G)}$ with $\boldsymbol{x} \in \boldsymbol{W_{ab}}$, and $\in \boldsymbol{W_{ba}}\}$,
$\boldsymbol{U_{ab}} = \{\boldsymbol{w} \in \boldsymbol{W_{ab}} \mid \boldsymbol{w}$ is the end vertex of an edge in $\boldsymbol{F_{ab}}\}$,
$\boldsymbol{U_{ba}} = \{\boldsymbol{w} \in \boldsymbol{W_{ba}} \mid \boldsymbol{w}$ is the end vertex of an edge in $\boldsymbol{F_{ab}}\}$.

The above sets are illustrated in Fig. 3.

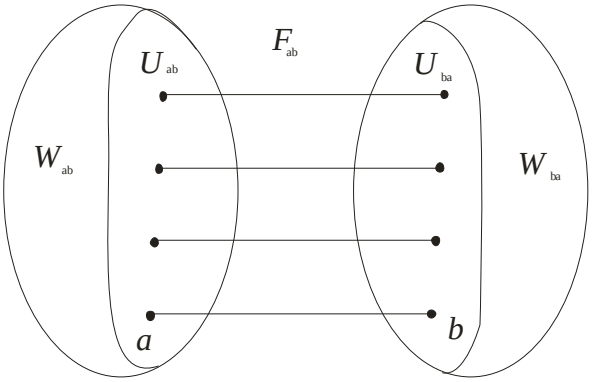


Figure 3: Important sets.

3 CHARACTERIZATION

Let $\boldsymbol{G}$ be a graph. For $\boldsymbol{X} \subseteq \boldsymbol{V}(\boldsymbol{G})$, let $\boldsymbol{G}[\boldsymbol{X}]$ denote the subgraph of $\boldsymbol{G}$ induced by the set $\boldsymbol{X}$. The following theorem is presented in [3].

Theorem 1. *Let* $\boldsymbol{ab}$ *be an edge of a connected, bipartite graph* $\boldsymbol{G}$ *such that* $\boldsymbol{d}(\boldsymbol{a}) = \boldsymbol{h}$ *and* $\boldsymbol{d}(\boldsymbol{b}) = \boldsymbol{h} - \boldsymbol{1}$, $\boldsymbol{h} \geq \boldsymbol{3}$. *Then* $\boldsymbol{G}$ *is isomorphic to* $\boldsymbol{\Gamma_h}$ *if and only if the following conditions hold:*

- $\boldsymbol{G}[\boldsymbol{U_{ab}}]$ *is convex in* $\boldsymbol{G}[\boldsymbol{W_{ab}}]$.
- $\boldsymbol{F_{ab}}$ *is a matching inducing an isomorphism between* $\boldsymbol{G}[\boldsymbol{U_{ab}}]$ *and* $\boldsymbol{G}[\boldsymbol{U_{ba}}]$
- $\boldsymbol{G}[\boldsymbol{W_{ba}}] = \boldsymbol{U_{ba}}$.
- $\boldsymbol{G}[\boldsymbol{W_{ab}}]$ *is isomorphic to* $\boldsymbol{\Gamma_{h-1}}$.
- $\boldsymbol{G}[\boldsymbol{W_{ba}}]$ *is isomorphic to* $\boldsymbol{\Gamma_{h-2}}$.

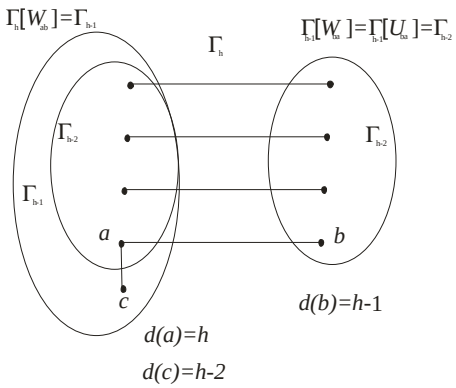


Figure 4: Recursive structure and important elements of Fibonacci cubes.

The theorem reflects a recursive structure of $\boldsymbol{\Gamma_h}$ which consists of two fundamentals subgraphs: $\boldsymbol{\Gamma_{h-1}}$ and $\boldsymbol{\Gamma_{h-2}}$. This subgraphs are joined with edges that comprise a $\boldsymbol{\Theta}$-class in $\boldsymbol{\Gamma_h}$. This recursive structure is depicted in Fig. 4. This characterization of Fibonacci cubes leads to the recognition algorithm presented in [3]. The algorithm finds in a given graph $\boldsymbol{G}$ an edge $\boldsymbol{ab}$ with endvertices of degree $\boldsymbol{h}$ and $\boldsymbol{h-1}$. The set $\boldsymbol{F_{ab}}$ decomposes $\boldsymbol{G}$ into two subgraphs induced by $\boldsymbol{W_{ab}}$ and $\boldsymbol{W_{ba}}$, respectively. The algorithm then checks the conditions of Theorem 1. The most consuming part of the algorithm is to assure whether $\boldsymbol{G}[\boldsymbol{W_{ab}}]$ and $\boldsymbol{G}[\boldsymbol{W_{ba}}]$ are isomorphic $\boldsymbol{\Gamma_{h-1}}$ and $\boldsymbol{\Gamma_{h-2}}$, which is done by two recursive calls of the

algorithm. Fig. 5 illustrate the action of the algorithm for the graph isomorphic to $\boldsymbol{\Gamma_5}$. We can see that the recursive calls for $\boldsymbol{\Gamma_4}$ and $\boldsymbol{\Gamma_3}$ are first performed. But when the algorithm is applied for $\boldsymbol{\Gamma_4}$, the recursive call for $\boldsymbol{\Gamma_3}$ is executed again. The example shows that the recursive algorithm revisits the same graph over and over again. The result is suboptimal $\boldsymbol{O(m \log n)}$ running time of the algorithm.

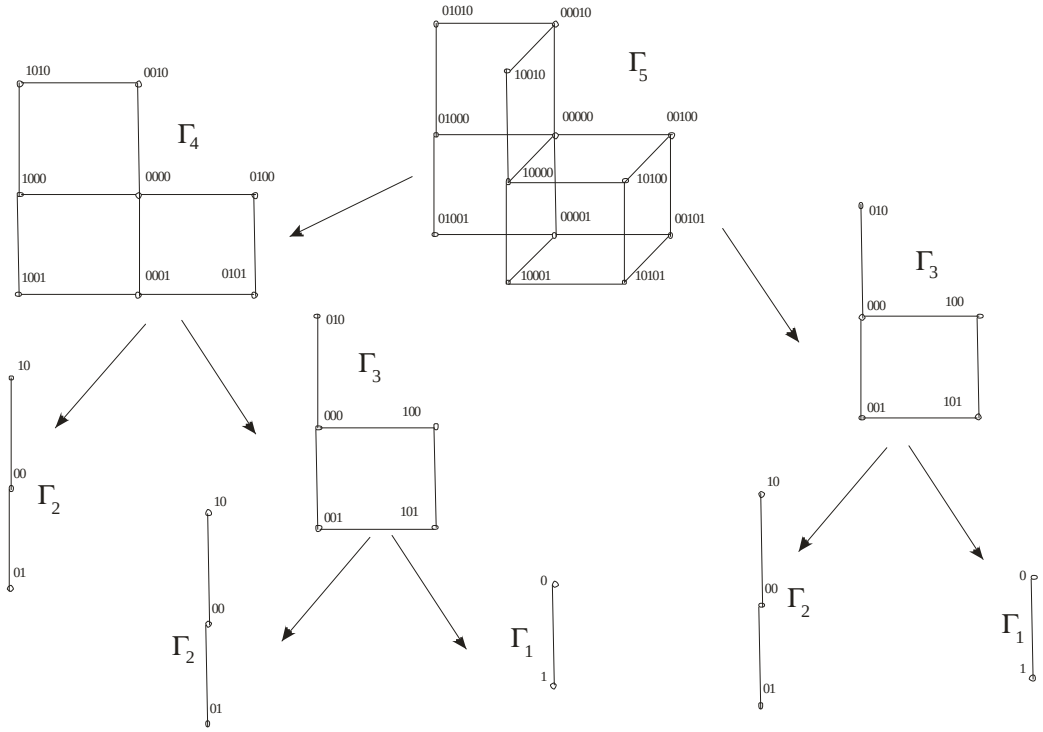


Figure 5: Decomposition of Γ_5.

In order to find a recognition algorithm with a better time bound, the following characterization has been found in [5] which refers to only one of both fundamentals subgraphs.

Theorem 2. *Let* $\boldsymbol{ab}$ *be an edge of a connected, bipartite graph* $\boldsymbol{G}$ *such that* $\boldsymbol{d(a) = h}$ *and* $\boldsymbol{d(b) = h - 1}$, $\boldsymbol{h \geq 3}$. *Then* $\boldsymbol{G}$ *is isomorphic to* $\boldsymbol{\Gamma_h}$ *if and only if the following conditions hold:*

- $\boldsymbol{G[U_{ab}]}$ *is convex in* $\boldsymbol{G[W_{ab}]}$.
- $\boldsymbol{F_{ab}}$ *is a matching inducing an isomorphism between* $\boldsymbol{G[U_{ab}]}$ *and* $\boldsymbol{G[U_{ba}]}$
- $\boldsymbol{G[W_{ba}] = U_{ba}}$.
- $\boldsymbol{G[W_{ab}]}$ *is isomorphic to* $\boldsymbol{\Gamma_{h-1}}$.
- $\boldsymbol{G}$ *admits exactly one vertex* $\boldsymbol{c \in N(a) \setminus (U_{ab} \cup U_{ba})}$ *such that* $\boldsymbol{d(c) = h - 2}$.
- $\boldsymbol{|W_{ca}| = F_{h-1}}$.
- $\boldsymbol{|U_{ab}| = F_h}$.

4 ALGORITHM

Theorem 2 is the basis for the following algorithm [5].

Procedure FIBONACCI(G, h, ab);
begin
1. **if** $|V(G)| \neq F_{h+2}$ **then** REJECT.
2. **if** $h = 1$ **and** G is K_2 or $h = 2$ **and** G is a path of length 2 **then** ACCEPT.
3. Find an edge ab, such that $d(a) = h$ and $d(b) = h - 1$.
4. **if** ab is not found **then** REJECT.
5. Find the sets $W_{ab}, W_{ba}, U_{ab}, U_{ba}$ and F_{ab}.
6. Find $c \in N(a) \setminus (U_{ab} \cup U_{ba})$ such that $d(c) = h - 2$.
7. **if** c is found **then** find the set W_{ca} else REJECT.
8. Verify that
   8.1. F_{ab} is a matching defining an isomorphism between $G[U_{ab}]$ and $G[U_{ba}]$
   8.2. $G[U_{ab}]$ is convex in $G[W_{ab}]$.
   8.3. $U_{ba} = W_{ba}$.
   8.4. $|W_{ca}| = F_{h-1}$.
   8.5. $|U_{ab}| = F_h$.
   8.6. FIBONACCI($G[W_{ab}], h - 1, ac$) returns ACCEPT.
9. **if** the foregoing conditions are fulfilled **then** ACCEPT **else** REJECT.

end.

We can prove [5] the following

Theorem 3. FIBONACCI(G, h, ab) *correctly recognizes a Fibonacci cube in* $O(m)$ *time.*

References

[1] W. J. Hsu, Fibonacci cubes - a new interconnection topology. IEEE Trans. Parallel Distr. Systems, 4 (1993), 3–12.

[2] S. Klavžar and P. Žigert, Fibonacci cubes are the resonance graphs of fibonaccenes, Fibonacci Quart., 43 (2005), 269–276.

[3] A. Taranenko, A. Vesel, Fast recognition of Fibonacci cubes, Algorithmica 49 (2007) 81–93.

[4] A. Vesel, Characterization of resonance graphs of catacondensed hexagonal graphs, MATCH Commun. Math. Comput. Chem., 53 (2005), 195–208.

[5] A. Vesel, Linear recognition and embedding of Fibonacci cubes, submitted.

[6] H.C. Wasserman and S.A. Ghozati, Generalized linear recursive networks: topological and routing properties. Computers & Electrical Engineering, 29 (2003), 121–134.

Fibonacci and Lucas cubes in chemical graph theory

Petra Žigert Pleteršek

Faculty of Chemistry and Chemical Engineering, University of Maribor, Slovenia

e-mail: petra.zigert@um.si

Abstract

Several classes of graphs based on Fibonacci strings were introduced in the last 10 years as models for interconnection networks, among them Fibonacci and Lucas cubes. The vertex set of a Fibonacci cube is the set of all binary strings of length n without consecutive 1's and in the case of a Lucas cubes we also forbid 1 in the first and the last bit. Two vertices of the Fibonacci or Lucas cube are adjacent if their strings differ in exactly one bit.

Benzenoid hydrocarbons are a graph model for aromatic hydrocarbons composed of benzen rings. Tubulene type structures are known as carbon nanotubes and were discovered around 20 years ago. They attend a lot of interest recently due to their unique structure which explains their unusual properties such as conductivity and strength. The aromaticity of such molecules is described by a Kekulé structure (t.i. perfect matching) and the interaction between them is depicted by a resonance graph.

It turns out that the resonance graphs of a certain type of benzenoid graphs are isomorphic to the Fibonacci cubes and the resonance graphs of some carbon nanotubes are closely connected to the Lucas cubes.

The 12th International Symposium on
Operational Research in Slovenia

SOR '13

Dolenjske Toplice, SLOVENIA
September 25 - 27, 2013

Section III:

Multiple Criteria Decision Making

VOLUME DISCOUNTS IN MULTIPRODUCT SUPPLIER SELECTION PROBLEM - MULTI-CRITERIA APPROACH

Zoran Babić
Faculty of Economics, Cvite Fiskovića 5, 21000 Split,
E-mail: babic@efst.hr

Tunjo Perić
Faculty of Economics and Business, Kennedyev trg 6, 10000 Zagreb,
E-mail: tperic@efzg.hr

Abstract
This paper deals with a concrete problem of flour purchase by a company that manufactures bakery products in a multiproduct situation where suppliers offer the discounts of the money volume of business in a particular period of time. The selection process is driven by the price, reliability and quality of particular vendors and subject to their capacity constraints. This problem has been solved using an integration of multi-objective methods and mixed integer programming to define the optimum quantities among the selected suppliers.

Keywords: vendor selection, volume discounts, multiproduct, AHP, fuzzy programming

1 INTRODUCTION

Identifying vendors with the lowest item price in a given industry becomes a major challenge for purchasing managers, especially when vendors offer multiple products and volume-based discount pricing schedules. In traditional quantity discount pricing schedules, price breaks that are function of the order quantity existed for each product, irrespective of the total magnitude of business the buyer contracts with the vendor over a given period of time. When every vendor offers the variety of products, vendors are finding it more meaningful to give discounts on the total value of multiproduct orders placed by a given buyer. In this environment, the supplier induces the buyer into making large purchases by offering discounts on the total value of sales volume, not on the quantity or variety of products purchased over a given period of time.

In this paper the authors present the integrated model which will have all of the possible issues of vendor selection problem in one hybrid model. The paper will show the construction of the model for volume discount case in multiproduct situation and the proposed methodology will be tested on the concrete example of vendor selection by a bakery. The final optimal solution has been found by the use of fuzzy multi-criteria programming approach.

The model combines number of methods used in operational researches. The first of them, analytic hierarchy process (AHP) is used to determine the coefficient weights of complex criteria functions (quality and reliability). Coefficients determined in this way will further be aggregated by Simple Additive Weighting method (SAW) or as it is recently called Weighted Sum Model (WSM) in order to present the coefficients of the objective functions in the fuzzy multi-criteria programming model providing the final selection and the quantity supplied from a particular vendor. The constraints in the multiple objective programming model are the total demand and the limitations of supplier capacities.

2 VENDOR SELECTION IN A BAKERY

Supplier selection and determination of quantities supplied by the selected suppliers is a multi-criteria problem. One of the most important issues in vendor selection is the choice of criteria for their evaluation. Which criteria will be chosen by the decision maker depends on

the kind of problem to be solved. Criteria which will be chosen for evaluation of flour vendors in this paper are: flour purchasing costs (C1), flour quality (C2), and supplier reliability (C3).

This supplier selection problem with the first objective function (flour purchasing costs) have been solved in the paper [1], so for this time we will introduce two more objective functions and we will concentrate on multi-criteria problem and solve it using the fuzzy programming approach.

Flour quality criterion important for bread production is expressed by the set of sub-criteria and the data for these criteria are presented in Table 1. The potential vendors supply the data on flour quality that they have to maintain throughout the contract period (Criterion C2). It is to be noted that the quality of flour depends on the wheat sort and quality and on technology used in flour production. In Table 1 quality indicators for the first type of flour (Type 550) are presented. Of course for the remaining three types of flour there exists the similar data.

Table 1: Quality indicators for flour Type 550

Quality indicators C2	Criteria type	Vendor			
		1	2	3	4
General characteristics of flour (A1)					
Moisture in % (B1)	min	13.53	13.27	13.49	13.33
Ash in % (B2)	min	0.57	0.549	0.53	0.486
Acidity level in ml/100 grams (B3)	min	1.5	1.5	1.6	1.8
Wet gluten in % (B4)	max	26.7	25.8	25.1	24.0
Farinograph (A2)					
Water absorption in % (B5)	max	60.8	59.8	58.5	61.1
Degree of mellowness in FU (B6)	min	70	65	85	60
Extensigraph (A3)					
Energy u cm^2 (B7)	max	81	104	87.2	107.3
Elasticity in mm (B8)	max<190	137	162	180	165
Resistance in EU (B9)	max	395	280	235	350
Amylograph (A4)					
Peak viscosity in AU (B10)	max	1054	860	1275	1325

Table 2: Vendor reliability indicators

Reliability indicators C3	Criteria type	Vendor			
		1	2	3	4
Financial stability, indebtedness and liquidity (A5)					
Coverage of fixed assets and stocks by capital and long term resources, (B11)	max	1.12	0.88	0.87	0.92
Share of capital in source of funds in %, (B12)	max	49.36	23.6	48.92	49.69
Indebtedness factor, number of years (B13)	min	7	19	13	19
Total assets turnover coefficient (B14)	max	0.65	0.49	0.52	0.35
General liquidity coefficient (B15)	max	7.17	1.19	1.07	0.75
Short term receivables collection period, in days (B16)	min	86	101	102	58
Performance indicators (A6)					
Coefficient of total revenue and expenditure ratio (B17)	max	1.06	1.03	1.03	1.02
Share of profit in total income in % (B18)	max	4.81	1.85	2.66	1.02
Share of profit in assets in % (B19)	max	3.14	0.91	1.39	1.01
Profit per employee in m.u. (B20)	max	60538	21189	12370	15446

When contracting flour supply, it is important to find reliable vendors, i.e. those that are assumed with a high degree of certainty that will not get into financial difficulties which could result in supply discontinuation. To evaluate vendor reliability the criteria of their solvency, financial stability, indebtedness, liquidity, and financial performance can be used and they are presented in Table 2. The vendors also should supply data on their reliability s_{kj} - (Criterion C3). We have to note that s_{kj} are the same for all types of flour and depend only on selected vendor. They are presented in Table 2.

A large number of vendor reliability sub-criteria and quality indicators for all types of flour (for each vendor) will make the decision making difficult. It would be hard to adequately evaluate vendors through these two set of criteria without support of experts and application of quantitative methods. That is the reason why in this paper authors made an aggregating model for this problem using analytic hierarchy process (AHP) [3], and simple additive weighting method (SAW) [5, page 6].

The data for the first objective function and all of the constraints are the same as in the previously mentioned paper so we will concentrate only on the two remaining objective functions - quality and reliability.

3 APPLICATION OF AHP AND WSM METHOD

The AHP method requires formation of hierarchical structure of goals and criteria. Considering the data from the Tables 1, 2, a hierarchical structure of goals and criteria for vendor selection is formed. The hierarchical structure (from Expert Choice software) is shown in the Figure 1.

The hierarchical structure of goals and criteria in this example consists of four levels as shown in the Figure 1. Level 1 represents the main goal - vendor selection. Level 2 represents three main criteria for vendor selection – Costs, Quality and Reliability, Level 3 represents sub-criteria for quality and reliability criteria, and Level 4 represents ten quality sub-criteria (B1-B10) and ten reliability sub-criteria (B11-B20).

After decomposition of the problem and formation of the hierarchical structure of goals and criteria, criteria and sub-criteria are compared pair-wise on their levels. This determines the relative priority of each element in the hierarchy. Pair-wise comparisons are carried out from the Level 2 to the Level 4 using the Saaty's scale.

All the comparisons were done by the owner of the bakery (DM), who is an expert in the area of management and flour technology. The DM has chosen three main criteria with several sub-criteria for flour quality and vendor reliability. After deciding on the selected criteria, the DM prioritized these criteria by using AHP method.

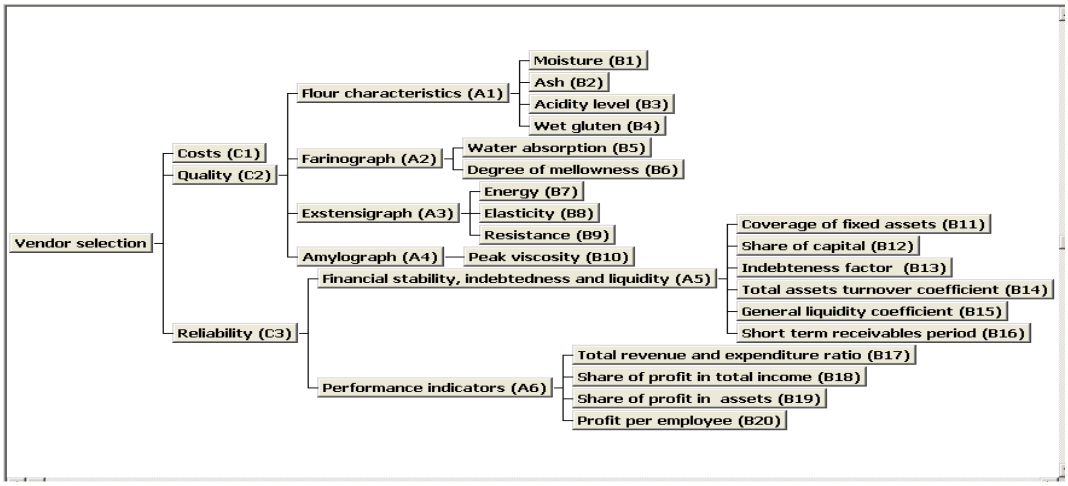


Figure 1: Hierarchical structure of the vendor selection problem

These priorities are used for obtaining the coefficients for second (F_{ij}) and third (S_j) objective functions in multi-criteria programming model. Using the data from Table 1 (for all types of flour) and Table 2, WSM method was used to determine the quality and reliability coefficients for each vendor. The data from Table 1 are first normalized by linear scale transformation. At the left side of the Table 3 there are original indicators from Table 1 (f_{kj}), and at the right side of that table the normalized coefficients are presented. In that way all normalised performance values are dimensionless, and we treat all criteria as benefit criteria - more is better. In the middle of the table are the weights of the sub-criteria obtained by AHP. In this notation WSM formula looks like: $F_{ij} = \sum_{k=1}^{k_1} w_{2k}\ f_{kj}'(P_i)$, i = 1,…, m; j = 1,..., n, where k_1 is the number of quality sub-criteria and w_{2k} are their ponders obtained by AHP method. In that way the coefficients F_{ij} are obtained and again normalized so that they also sum up to one. That is presented in the last two rows of Table 3.

Table 3: WSM method for quality indicators (flour Type 550)

C2		V1	V2	V3	V4	w_{2k}	V1	V2	V3	V4
		f_{kj} (P_1)					f_{kj}' (P_1)			
B1	min	13.53	**13.27**	13.49	13.33	0.03275	0.9808	1	0.9837	0.9955
B2	min	0.57	0.549	0.53	**0.486**	0.01475	0.8526	0.8852	0.9170	1
B3	min	**1.5**	**1.5**	1.6	1.8	0.00688	1	1	0.9375	0.8333
B4	max	**26.7**	25.8	25.1	24.0	0.07063	1	0.9663	0.9401	0.8989
B5	max	60.8	59.8	58.5	**61.1**	0.09375	0.9951	0.9787	0.9574	1
B6	min	70	65	85	**60**	0.28125	0.8571	0.9231	0.7059	1
B7	max	81	104	87.2	**107.3**	0.225	0.7549	0.9692	0.8127	1
B8	max	137	162	**180**	165	0.075	0.7611	0.9	1	0.9167
B9	max	**395**	280	235	350	0.075	1	0.7089	0.5949	0.8861
B10	max	1054	860	1275	**1325**	0.125	0.7955	0.6491	0.9623	1
				F_{1j}			0.8579	0.8922	0.8296	0.9768
				F_{1j} - normalized			**0.2412**	**0.2509**	**0.2333**	**0.2746**

Table 4: Quality (F_{ij}) and reliability (S_j) indicators for objective functions

Vendor	F_{1j} flour Type 550	F_{2j} flour Type 850	F_{3j} flour Type1100	F_{4j} flour Type 1150	S_j
V_1	0.2412	0.2513	0.2134	0.1835	0.4105
V_2	0.2509	0.2487	0.3105	/	0.186
V_3	0.2333	0.2335	0.2198	0.3014	0.2018
V_4	0.2746	0.2689	0.2768	0.2503	0.2017

In the same way the coefficients for the remaining three types of flours and the reliability coefficients S_j are obtained. All the coefficients for the quality and reliability objective functions are presented in Table 4.

4 FUZZY MULTI-CRITERIA PROGRAMMING MODEL

With all these data multi-criteria model for determining the vendors and supply quotas for each vendor in the multiproduct case with volume discounts can be formulated. This model can be solved by the use of mixed integer and fuzzy multi-criteria programming in the similar way as it can be seen in [1] and [2]. The difference is that now the authors present multiproduct situation and volume and not quantity discounts. Three objective functions are:

$$\min Z_1 = \sum_{j\in J}\sum_{r\in R_j}\left(1-d_{jr}\right)\cdot v_{jr}\ ,\max Z_2 = \sum_{i\in I}\sum_{j\in J_i}F_{ij}\cdot x_{ij}\ ,\ \max Z_3 = \sum_{i\in I}\sum_{j\in J_i}S_j\cdot x_{ij}$$

where d_{jr} are the percent of discount associated with bracket r of vendor j's cost function, v_{jr} are the volume of business (*money amount*) awarded to vendor V_j in discount bracket r, and x_{ij} are the units of item P_i to purchase from vendor V_j

All the constraints for this case study are the same as in the paper [1], and finally model has 39 variables and 32 constraints. To solve this multi-criteria model in this paper the fuzzy linear programming approach has been used. First with mixed integer linear programming software (Excel Solver) marginal solutions and the values of objective functions were obtained and they are shown in the Table 5.

Table 5: Payoff table for marginal solutions

	Z 1	**Z 2**	**Z 3**
X1*	**1511329.035**	1711.275	1884.780
X2*	1566040.400	**1881.150**	1477,100
X3*	1540552.741	1670.513	**2224.741**

The next step is to formulate fuzzy programming goal function. According to the paper [5] membership functions for the three objective functions are:

$$\mu_{f_1}(X)=\frac{1566040.4-f_1(X)}{54711.365},\quad \mu_{f_2}(X)=\frac{f_2(X)-1670.513}{210.6373},\quad \mu_{f_3}(X)=\frac{f_3(X)-1477.1}{747.641}.$$

With the weights obtained with AHP method fuzzy multi-objective linear programming model is:

$$Max\ f = w_1\cdot\lambda_1 + w_2\cdot\lambda_2 + w_3\cdot\lambda_3 = 0.429\cdot\lambda_1 + 0.429\cdot\lambda_2 + 0.143\cdot\lambda_3$$
$$\lambda_1 \le \mu_{f_1}(X),\ \lambda_2 \le \mu_{f_2}(X),\ \lambda_3 \le \mu_{f_3}(X),\ \lambda_i \in [0,1],\ X\in S$$

Solving this model with mixed integer programming software (Excel Premium Solver) the following result presented in Table 6 is obtained.

Table 6: Final optimal solution

Variables	Values	Variables	Values	Variables	Values
x_{11}	1633.45	x_{41}	500	v_{33}	0
x_{12}	2000	x_{43}	0	v_{41}	0
x_{13}	0	x_{44}	500	v_{42}	0
x_{14}	366.55			v_{43}	300000
x_{21}	740.01	v_{11}	0		
x_{22}	759.99	v_{12}	0	λ_1	0.950875
x_{23}	0	v_{13}	717796.37	λ_2	0.346823
x_{24}	0	v_{21}	0	λ_3	0.646829
x_{31}	0	v_{22}	0		
x_{32}	500	v_{23}	650000	z*	0.64856
x_{33}	0	v_{31}	0		
x_{34}	0	v_{32}	0		

From this Table it can be seen that first vendor delivers first, second and fourth type of flour in quantity of x_{11} = 1633.45 t, x_{21} = 740.01 t and x_{41} = 500 t. Because the unit price for this vendor and those types of flour is c_{11} = 220.1, c_{21} = 200.2 and c_{41} = 420.25 that mean that the buyer's cost for that purchase is v_{13} = 717796.37 Euros, and this amount of money falls in the third price level with the discount of 10% so it means that the buyer will pay 646016,69 Euros. Of course $v_{11} = v_{12} = 0$.

For the second vendor we have x_{12} = 2000, x_{22} = 759.99, x_{32} = 500, x_{42} = 0 (second vendor doesn't have fourth type of flour) and because the total amount of money for these quantities is v_{23} = 650000 (> 400000) falls in the third price bracket, the buyer has the maximum discount of 8% and will pay 598000 Euros to the second vendor.

Vendor 3 is not certificated well along the evaluation procedure, and should then receive no orders from the buyer, and of course there are no purchasing costs.

Fourth vendor delivers 366.55 t of the first type of flour (x_{14} = 366.55), and 500 t of the fourth type (x_{44} = 500). Value for these shipment is v_{43} = 300 000 Euros and the buyer is again in the third price bracket for this vendor and has the discount of 10%, which means he would pay 270000 Euros.

Optimal value of the first objective function, the total purchase cost (with these discounts), is 1514016.69 Euros. In Table 7 the distribution of the purchasing costs in this optimal solution are presented. It can be seen that the buyer saved 153779.63 Euros (1667796.32 - 1514016.69) due to discounts offered by the vendors.

Table 7: Distribution of purchasing costs in optimal solution

	V1	**V2**	**V3**	**V4**	**Total**
Purchasing costs (without discount)	717796.33	650000	0	300000	**1667796.33**
Discount	10%	8%	0%	10%	
Purchasing costs (with discount)	646016.69	598000	0	270000	**1514016.69**

5 CONCLUSION

Solving the concrete example by application of the proposed methodology we can make a number of conclusions presenting the advantages of using the proposed methodology in solving the problem of vendor selection and determination of order quotas with volume discounts. AHP and SAW method allow efficient reducing of complex criteria functions into simple criteria functions.

When solving the multi-criteria model the use of fuzzy technique proves to be very efficient. The developed model, verified in this paper on the real case study, has a general value because it can be successfully used in solving similar practical problems dependent on numerous qualitative and quantitative criteria.

References

[1] Babić, Z., T. Perić (2011): *Volume discounts in multiproduct supplier selection problem*, Proceedings of the 11th International Symposium on Operations Research, SOR '11, Dolenjske Toplice, Slovenia, p. 199-204.

[2] Perić, T., Z. Babić (2011): *Quantity Discounts in Supplier Selection Problem by Use of Fuzzy Multi-criteria Programming*, Croatian Operational Research Review, Vol 2, 49-58.

[3] Saaty, T.L. (2001): *Decision Making for Leaders. The Analytic Hierarchy Process for Decision in a Complex World*, RWS Publications, Pitsburgh USA.

[4] Triantaphyllou, E. (2000): Multi-Criteria Decision Making Methods: A Comparative Study, Kluwer Academic Publishers.

[5] Zimmermann, H.J. (1978): Fuzzy programming and linear programming with several objective functions. *Fuzzy Sets and System*, 1, 45-55.

A model for object evaluation based on users' comments/evaluations

Drago Bokal
University of Maribor, FNM, Koroška 160, SI-2000 Maribor
and
Polona Pavlič
University of Maribor, FNM, Koroška 160, SI-2000 Maribor
and
Janez Žerovnik
Institute of Mathematics, Physics and Mechanics, Jadranska 19, SI-1111 Ljubljana, and
University of Ljubljana, FME, Aškerčeva 6, SI-1000 Ljubljana, Slovenia.

Abstract

A proposal of an abstract model for objects evaluation is given and explained in more detail on the example where the users' comments of some service are considered.

1 Introduction

Sofware as a service (SaaS) and cloud computing, which is a more general infrastructure technology that facilitates this type of software delivery and pricing, are becoming new platforms for enterprise and personal computing [2]. Cloud computing is a paradigm to deliver ondemand resources (e.g., infrastructure, platform, software, etc.) to customers similar to other utilities like water, electricity and gas [3].

Traditionally, small and medium enterprises (SMEs) had to make high capital investment for IT infrastructure, skilled developers and system administrators, which results in a high cost of ownership. Cloud computing aims to deliver a network of virtual services so that users can access them from anywhere in the world on subscription at competitive costs depending on their Quality of Service (QoS) requirements. Therefore, SMEs have no longer to invest large capital outlays in hardware to deploy their service or human expense to operate it. However, with the growth of public Cloud offerings, for customers it has become increasingly difficult to decide which provider can fulfill their QoS requirements. Similar services are offered at different prices and performance levels with different sets of features. While one provider might be cheap for storage services, they may be expensive for computation. Therefore, given the diversity of Cloud service offerings, an important challenge for customers is to discover who are the optimal Cloud providers that can satisfy their requirements. In this context, the Cloud Service Measurement Index Consortium (CSMIC)[7] has identified metrics that are combined in the form of the Service Measurement Index (SMI), offering comparative evaluation of Cloud services. These measurement indices can be used by customers to compare different Cloud services. The features of interest include performance and cost, but also usability, security and privacy, assurance, agility, accountability, etc. that can in turn be defined more precisely when a particular type offering is regarded.

A challenge is how to rank the Cloud services based on (some of) these attributes. Deciding which service matches best with all functional and nonfunctional requirements of a user is a decision problem[5], more precisely a problem of Multi-Criteria Decision-Making (MCDM), a classical topic in operational research [1].

The quality of features that contribute to the evaluation used for ranking of the services is often not easy to measure. In this work we are interested in measuring services based on users' evaluation. User experience (UX) evaluation means investigating how a person feels about using a system (product, service, non-commercial item, or a combination of them). It

is non-trivial to evaluate user experience and come up with solid results, since user experience is subjective, context-dependent and dynamic over time [4]. In this work, it is assumed that we are collecting a sequence of evaluations given by users of the services. Clearly, simple aggregation, for example taking the average score, can be severely misleading due to various reasons, including attacks by service providers who might artificially trigger a population of very friendly users. Therefore, some mechanisms are needed which would minimize such effects. In the model discussed here, services are evaluated based on the users comments, but at the same time users and their particular comments/evaluations are evaluated aiming to exclude unuseful or clearly biased comments, and at the same time giving higher weight to evaluations of services given by provably experienced and objective users.

In this short paper we outline a proposal of an abstract model for objects evaluation that has been applied to cloud service ranking. The model is explained in more detail on the example where the users' comments of some (abstract) service are considered. This can be a part of a larger model where similar approach would be used also for evaluation of the users, and based on these, the objects commented may be evaluated. More details will be given in in the full paper (see also [6] for a preliminary version).

2 Evaluation model

Here we are interested in the following task. Assume we are given a series of evaluations, grades, assesments, or any other series from which user's opinion about the quality or ranking can be derived. The sequence may dynamically extend over time. The model is intended to be independent of particular application. From motivation above it follows that possible objects include services of cloud computing providers, that can be basic, i.e. performing computation, lending storage, running application, but also complex combining several basic services. Even the evaluations themselves can be evaluated regarding truthfulness, usefulness, etc. A trivial example is that there may be elements in the sequence that are result of an errorneus use of the system. These can in most cases be filtered easily, but also among syntactically sound comments there are comments that seem to be very useful and comments that provide very little information. Sometimes user's reputation as an experienced user of certain service or his/her previous record of useful comments may increase confidence that the current comment has to be taken seriously. Last but not least, when the evaluation/ranking is published, there may be interest by the sellers/providers to bias the results by attacking the system with numerous artifical comments.

A formal description of our model follows below. We start with introducing some general notions of the abstract model for evaluation of abstract objects. The general model may be used in various ways, in particular we are motivated by a possible application that would collect users' comments about certain services, and would provide (dynamicaly changing as more comments arrive) various evaluations and/or rankings of comments, users, basic services, complex services, service providers.

Notation

Attribute	Symbol
Set of evalution categories	$\mathbb{K}$
Possible marks	$\mathbb{O} \subseteq \{\cdot, -1, \ldots, 5\} \subseteq \mathbb{Z}$
Set of evaluated objects	$\mathcal{O}$
Evaluated object	$o \in \mathcal{O}$, $o = (O_o, \Omega_o, \omega_o)$
Particular mark for object $a \in O_o$	$(a^{(1)}, \ldots, a^{(k)}) \in \mathbb{O}^{\mathbb{K}}$
All marks for object $o \in \mathcal{O}$	$O_o \in (\mathbb{O}^{\mathbb{K}})^{\mathbb{N}}$
Aggregated mark for object $o \in \mathcal{O}$	$\Omega_o \in \mathbb{R}^{\mathbb{K}}$
Final mark for object $o \in \mathcal{O}$	$\omega_o \in \mathbb{R}$
Aggregation operator	$f : (\mathbb{O}^{\mathbb{K}})^{\mathbb{N}} \mapsto \mathbb{R}^{\mathbb{K}}$
Summarizing operator	$g : \mathbb{R}^{\mathbb{K}} \mapsto \mathbb{R}$

Remarks:

- The set of possible marks is assumed to be an ordinal set, in most cases $\{1, \ldots, 5\}$, or $\{-1, 0, 1\}$. In addition we allow mark "$\cdot$", meaning there was no mark given.
- Evaluation categories are assumed to be related to specific components of the grade or evaluation submitted.
- Evaluated object is represented by an ordered triple, where the first element in the triple is a series of submitted marks, the second element are the aggregated marks for components, and the third element is the summarized mark of the object.
- We have two operators, f and g: f gives the aggregated mark based on the sequence of marks, and g provides the summarized mark based on aggregated marks of the components.

3 First example: evaluation of users' comments

In this example, we can precise the meaning of the abstract sets of the general model.

1. $\mathbb{K} = \{(\text{vr}), (\text{up}), (\text{kr})\}$. *Trustworthiness* (Verodostojnost), (vr), is used to label comments that should be ignored due to identified (or very likely) false information they give: value 0 indicates false and value 1 indicates trustworthy comments. Applicability (Uporabnost), (up), with values -1, 1 measures applicability or usefulness of comments. Credibility (Kredibilnost), (kr), of the user is assumed to be available from a similar model for evaluation of users can have values from the set $\{1, 2, 3, 4, 5\}$.
2. $\mathbb{O} = \{-1, 0, 1, 2, 3, 4, 5\}$ similarly as above.
3. Aggregation operator f:
   - For category (vr)

     $$\Omega_o^{(\text{vr})} := \begin{cases} 0 & ; \quad |\{i \mid a_i^{(\text{vr})} = 0 \text{ in } a_i^{(\text{kr})} \geq c_{2,(\text{kr})}\}| > c_{1,(\text{kr})} \\ 1 & ; \quad \text{otherwise.} \end{cases}$$

     Here $a_i^{(\text{vr})}$ stands for the users mark of the comment, and $a_i^{(\text{kr})}$ for credibility of the user. Comment if not trustworthy (has value 0) if there is enough ($c_{1,(\text{kr})}$) well evaluated users (i.e. having mark at least $c_{2,(\text{kr})}$) who marked the comment as not trustworthy.

- For category (up):

$$\Omega_o^{(\mathrm{up})} := \begin{cases} \frac{1}{|O_o|}\sum_{i\in O_o} a_i^{(\mathrm{up})} & ; \text{ comment has less than } c_{1,(\mathrm{up})} \text{ marks,} \\ \frac{1}{|O_o|}\sum_{i\in O_o} a_i^{(\mathrm{up})} & ; \text{ comment is ranked according to } \frac{1}{|O_o|}\sum_{i\in O_o} a_i^{(\mathrm{up})} \text{ within the first } c_{2,(\mathrm{up})} \\ & \quad \text{most applicable comments of the} \\ & \quad \text{service with at least } c_{1,(\mathrm{up})} \text{ marks.} \\ -1 & ; \text{ comment is not ranked within first } c_{2,(\mathrm{up})} \\ & \quad \text{most applicable comments of the} \\ & \quad \text{service with at least } c_{1,(\mathrm{up})} \text{ marks.} \end{cases}$$

  First, until at least $c_{1,(\mathrm{up})}$ marks are available, a comment is evaluated by simply taking the average mark. Then, after at least $c_{1,(\mathrm{up})}$ marks are available, very low marks of a comment are deleted thus enabling newer comments to emerge. In other words, comments that are not very useful are forgotten.

- Summarizing operator g: A comment is of interest, if it is both trustworthy and applicable (useful). Hence, here we simply multiply the two aggregated marks.

$$g(\Omega_o^{(\mathrm{vr})}, \Omega_o^{(\mathrm{up})}) := \Omega_o^{(\mathrm{vr})} \cdot \Omega_o^{(\mathrm{up})}. \tag{1}$$

  As $\Omega_o^{(\mathrm{vr})}$ for a comment that is not trustworthy, this excludes all such comments.

Parameters of the mode - evaluation of comments

Parameter	Notation
Minimal number of marks for a stable comment	$c_{1,(\mathrm{up})} \in \mathbb{N}$
Number of comments regarded as applicable	$c_{2,(\mathrm{up})} \in \mathbb{N}$
Number of credible users that evaluated the comment as not trustworthy	$c_{1,(\mathrm{kr})} \in \mathbb{N}$
Lower bound for credibility of a user	$c_{2,(\mathrm{kr})} \in \mathbb{N}$

When evaluating the comments, a comment can be in two stages. The first stage is when the comment is fresh and the applicability of the comment is not yet clear. Later, in the second phase, the number of marks is sufficient to have a relatively stable evaluation of the applicability.

Parameter $c_{1,(\mathrm{up})} \in \mathbb{N}$ is the minimal number of marks needed to consider the evaluation of the comment as stable. For example, if the comments appear as the response of users of a website, we would expect to reach this number in a week or two. On the other hand, for the results to be statisticaly sound, the value should not be much less than 100.

Parameter $c_{2,(\mathrm{up})} \in \mathbb{N}$ is the number of comments that are believed to be useful for the users (applicable). The comments that are within the most $c_{2,(\mathrm{up})} \in \mathbb{N}$ applicable are evaluated based on the proportion of marks indicating they are applicable, all other comments are labeled as not useful (their applicability is 0).

When presenting the evaluations, in particular in a dynamic environment, it seems to be advantageous to also provide the results in the same way. Namely, the fresh comments (still in stage 1) and the comments with stable evaluation should be ranked in two separated lists. The most interesting most probably are the first (winning) comments from both lists.

Parameter $c_{1,(\mathrm{kr})} \in \mathbb{N}$ gives a lower bound on the number of users needed to label a comment as not trustworthy. Here the total number of users that evaluated the comment is not important.

Finaly, parameter $c_{2,(\mathrm{kr})} \in \mathbb{N}$ is a lower bound for credibility of a user.

A similar model can be used for evaluating the credibility of users. As it is used in th previous example, we give a short outline. In the example, we assume that the users comment certain service.

Second example: credibility of users

1. $\mathbb{K} = \{(\mathrm{up}), (\mathrm{km}), (\mathrm{oc})\}$. (up) is the frequency of usage of the service, values from $\{1, 2, 3, 4, 5\}$ are computed as follows. Five time windows are defined in advance. For each service, there is a predefined number of events $n^{(\mathrm{up})}$ given. According to this predefined table, the user is given his/her value based on the history of usage of the service. (km) with values from $\{1, 2, 3, 4, 5\}$ is based on the marks that other users have given to the previous comments of the user. (oc) measures how synchronized are the users marks with other users' evaluations.

   The aggregated evaluations are of the form $(a_i^{(\mathrm{up})}, a_i^{(\mathrm{km})}, \cdot)$ if it is a users comment, and $(a_i^{(\mathrm{up})}, \cdot, a_i^{(\mathrm{oc})})$, when a user evaluates a comment.

2. $\mathbb{O} = \{\cdot, 1, 2, 3, 4, 5\}$, as above.

3. Aggregation operator f:

   - For frequency of usage, (up) is the average of windows sizes, attached to the comment: $\Omega_o^{(\mathrm{up})} := \frac{1}{|O_o|} \sum_{i \in O_o} t[a_i^{(\mathrm{up})}]$.
   - For (km) it is the average:
     $\Omega_o^{(\mathrm{km})} := \frac{1}{|O_o|} \sum_{i \in O_o} a_i^{(\mathrm{km})}$.
   - For (oc), the difference to the average mark. Let $a_i^{(\mathrm{oc})} := a_i^{(\mathrm{vr})} \cdot a_i^{(\mathrm{up})} - \Omega_o^{(\mathrm{vr})} \cdot \Omega_o^{(\mathrm{up})}$. Then $\Omega_o^{(\mathrm{oc})} := 5 \cdot \frac{1}{|O_o|} \sum_{i \in O_o} a_i^{(\mathrm{oc})}$
   - Summarizing operator g: If the user is a frequent user, then his marks are (likely to be) important. Otherwise, it may be better to use the average.
     $g(\Omega_o^{(\mathrm{up})}, \Omega_o^{(\mathrm{km})}, \Omega_o^{(\mathrm{oc})}) := \alpha \Omega_o^{(\mathrm{km})} + (1 - \alpha)\Omega_o^{(\mathrm{oc})}$, where $\alpha = \dfrac{r}{N}$, r rang according to $\Omega_o^{(\mathrm{up})}$ and N the number of all users.

Parameters of the model - evaluation of users

Parameter	**Notation**
Number of usages of a service	$n^{(\mathrm{up})} \in \mathbb{N}$
Time window, in which the number of usages of a service occured	$t[i] \in \mathbb{N}$

Clearly, the users that use a service very rarely will likely not provide very useful information. As some services are naturally used more frequently than others, therefore a constant is defined for each service.

Prameter $n^{(\mathrm{up})} \in \mathbb{N}$ provides a bound for usages of the service to regards the user as experienced user.

Parameter $t[i] \in \mathbb{N}$ gives time within a user filed $n^{(\mathrm{up})}$ usages of a service.

4 Conclusion

A model for evaluation of abstract objects based on a sequence of evaluations is proposed. The model is illustrated by elaborating an example, where a sequence of comments is given in which users both comment some (abstract) service(s) and give comments about other users' comments. The motivation for this study is a possible application providing support for users of services within the cloud computing paradigm.

The very same model was applied also to basic and complex cloud services, as well as to service providers. For instance, we used evaluation categories measuring user satisfaction with service responsiveness, quality, reliability, GUI usability, functionality, and security for basic services. For operators, we used weighted averaging of the grades asigned by users, where the weights corresponded to user's own credibility. For complex services, we averaged evaluations of basic services, together with their supportive services, such as migration, integration and backup reliability. Similarly, we used list of services and of user support as basic evaluation categories of the service provider, together with the list of all (complex) services the provider is offering, and used weighted average as grade aggregation operators, again the weights being the credibility of users giving the grades.

The aim of this generic model is to provide an unifying, widely applicable model and methodology that will allow for development of common evaluation tools, study of interesting phenomena within its diverse applications, and develop mechanisms using aggregation operators that would allow for comparison of rankings even if obtained from different evaluation data sources. Achieving this goal, it will become possible to develop a multi criteria decision support system that will combine data from diverse sources and help users choose the optimal service and provider for their needs.

5 Acknowledgements

The first two authors were supported by grant SINTESIS, funded by European regional development fund and by the Ministry of education, science, and sport of Republic of Slovenia. The third author was supported in part by ARRS, Research Agency of Slovenia.

References

[1] J. Cochrane, M. Zeleny, Multiple Criteria Decision Making, Univ. of South Carolina Pr., 1973.

[2] M. Cusumano, Cloud computing and SaaS as new computing platforms, Communications of the ACM 53 (4) (2010) 2729.

[3] S. K. Garg, S. Versteeg, and R. Buyyaa, A framework for ranking of cloud computing services, Future Generation Computer Systems 29 (2013) 1012 1023.

[4] E. L-C. Law, V. Roto, M. Hassenzahl, A. P.o.S. Vermeeren, and J. Kort, Understanding, Scoping and Defining User EXperience: a Survey Approach, in Proceedings of Human Factors in Computing Systems, CHI09, pp. 719-728. http://doi.acm.org/10.1145/1518701.1518813

[5] I. Toma, D. Roman, D. Fensel, B. Sapkota, and J.M. Gomez, A Multi-criteria Service Ranking Approach Based on Non-functional Properties Rules Evaluation, Lecture Notes in Computer Science 4749 (2007) 435-441.

[6] delovno gradivo (in Slovene).

[7] Cloud Service Measurement Index Consortium (CSMIC), SMI framework. URL: http://beta-www.cloudcommons.com/servicemeasurementindex.

CONVERGENCE OF AUTONOMOUS GROUP DECISION-MAKING PROCEDURES: APPLICATION TO RANKING AND SORTING

Andrej Bregar
Informatika d.d., Vetrinjska ulica 2, 2000 Maribor
andrej.bregar@informatika.si

Abstract
Algorithms and metrics that enable the convergence of automated and autonomous group consensus seeking procedures are defined in this paper. They are applied to both most relevant decision-making problematics: (1.) ranking of alternatives and (2.) sorting of alternatives into arbitrary many ordered categories. Introduced metrics assess the majority opinion, determine the direction of the group and identify the most discordant decision-maker. Proposed algorithms adjust preferential parameters of the most opposing group member with the purpose to iteratively unify opinions.

Keywords: Multi-criteria decision analysis, Decision support systems, Negotiations, Group decision-making, Consensus seeking, Ranking, Sorting, Preference aggregation and disaggregation

1 INTRODUCTION

Group decision-making methods and procedures may provide various levels of support to decision-makers. Several simple approaches aggregate preferential parameters of individual group members into a fictive compromise solution that is compensatory in nature and does not necessarily correspond to any opinion. Such methods include the original PROMETHEE for groups [2] and group AHP [14]. More efficient and commonly used approaches utilize robustness analysis and visualization techniques [6, 12] or incorporate the role of a human moderator [9, 10] to identify conflicts and facilitate the group in reaching an agreed upon solution. The most advanced methods, however, are able to autonomously asses divergence in judgements and suggest necessary actions to iteratively approach consensus by applying appropriate metrics and algorithms [7, 11, 13]. Some are supplemented with mechanisms to automatically adjust evaluations of decision-makers in order to overcome discrepancies in the problem solving team while assuring that the collective decision remains robust and does not violate personal constraints of individual group members [5].

It is essential for any method of the latter type to implement (1.) metrics that determine the levels of (dis)agreement of individual decision-makers with the direction to which the group as an integral entity is heading, (2.) robustness measures that ensure a reliable decision and prevent group members with firm judgements to conform to opinions of other colleagues or intelligent agents, and (3.) an algorithm to iteratively and autonomously adjust preferences of the most discordant group member in order to unify him with the majority opinion of the group. These mechanisms have already been introduced in relation with a dichotomic sorting aggregation-disaggregation procedure for consensus seeking [5], which has been shown by simulation and case based studies to perform efficiently [3, 4]. However, dichotomic sorting is only a special localized case of more general sorting into an arbitrary number of categories [15, 16]. Furthermore, out of three basic decision-making problematics – ranking, sorting, and choice – the most widely used is ranking. The goal of the paper is thus to define original algorithms and metrics for convergent autonomous group consensus seeking based on the general problematics of sorting and ranking.

The rest of the paper is organized as follows. Section 2 presents the general autonomous consensus seeking procedure, which is based on the aggregation-disaggregation paradigm, and to which the introduced methodological solutions apply. Section 3 defines metrics and algorithms for the unification of decision-makers' preferences with regard to the problematic of sorting alternatives into an arbitrary number of ordered categories. In Section

4, the case of ranking is analogously addressed. Section 5 finally concludes the paper with a resume and some directions for further work.

2 GENARAL AUTONOMOUS GROUP CONSENSUS SEEKING PROCEDURE

The general group decision-making procedure is presented on Figure 1. It is based on the aggregation-disaggregation analysis [5], so it can automatically and autonomously converge towards a consensual solution. The depicted procedure is independent of both the decision-making problematic and the preference model. It can therefore be applied to ranking and sorting on one hand, as well as to different types of methods, such as the multi-attribute utility function, outranking or Analytic Hierarchy Process (AHP).

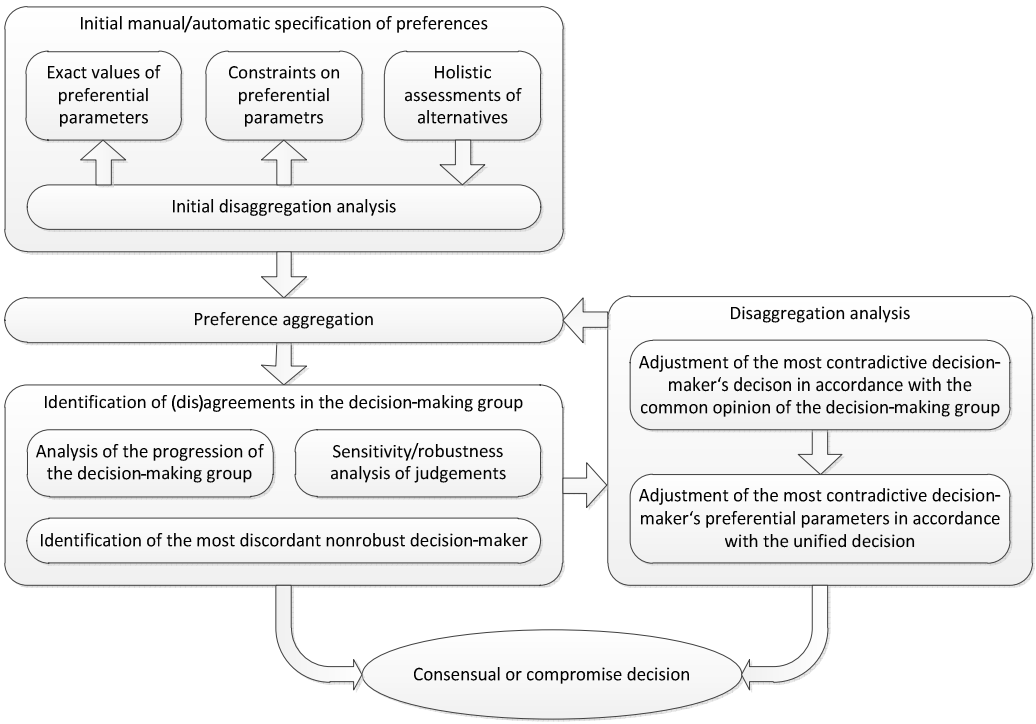


Figure 1: Aggregation-disaggregation based group consensus seeking procedure

A prerequisite for the operability of this procedure are appropriate metrics that determine the state of the decision-making group, and algorithms that iteratively and reasonably adjust preferences of all contradictive group members within the space of specified constraints. In this way, a mechanism for the unification of various discordant opinions can be provided. Several original metrics and algorithms are defined in the following two sections.

3 SORTING

Sorting refers to the assignment of a set of alternatives $A = \{a_1, \dots, a_m\}$ into q categories or classes, which are ordered from the least preferred one C^1 to the most preferred one C^q [16]. This means that alternative a_i is treated better than alternative a_j, if $a_i \in C^k$, $a_j \in C^l$ and $C^k > C^l$. There are no limitations to how many alternatives may be assigned to the same category. Two adjacent categories C^l and C^{l+1} are delimited with a referential profile b^l, which can be either a scalar or a vector of *n* criteria-wise values.

For the purpose of generalization, all formulas that are defined in Sections 3 and 4 are independent of the preferential model. Hence, they can be applied to all fundamental types of models, such as the multi-attribute utility function or pseudo-criterion based outranking. The evaluation of an alternative a_i is denoted with $u_i = u(a_i)$.

3.1 Concordance of the decision-maker and the decision-making group

Let C_i^{DM} denote the category into which the decision-maker *DM* sorts the alternative a_i. Let C_i^L denote the lowest (worst) and C_i^H the highest (best) category to which the alternative a_i is assigned by any decision-maker. C_i^{median} is hence the median of the $[C_i^L, C_i^H]$ interval, such that 50 percent of group members sort a_i into C_i^{median} or higher, and the other half into $C_i^{median-1}$ or lower. Thus:

$$\forall DM: C_i^{DM} \in [C_i^L, C_i^H],$$
$$C_i^{median} \in [C_i^L, C_i^H].$$

The degree of agreement of a single decision-maker *DM* with the opinion of the whole decision-making group with regard to the category to which the alternative a_i is assigned is obtained with the weighted distance metric. It considers the deviation of C_i^{DM} from C_i^{median}, and the distribution of memberships of a_i in different categories:

$$d_i^{DM} = \frac{abs\left(ord(C_i^{DM}) - ord\left(C_i^{median}\right)\right)}{abs\left(ord(C_i^H) - ord(C_i^L)\right)} \cdot \frac{card(C_i^{DM} + \{C_i^{<}\})}{card\{DM\}}.$$

The *ord* function returns the ordinal rank of a category. With *card*, sizes of two different sets are represented: (1.) the number of decision-makers that sort a_i into a certain category or a category subset, respectively, and (2.) the size of the set of decision-makers $\{DM\}$, i.e. the number of all group members. Finally, $\{C_i^{<}\}$ is the subset of categories that are more distant from the median category C_i^{median} than C_i^{DM}.

The agreement degree can now be calculated for *DM* with respect to a_i as

$$g_i^{DM} = 1 - d_i^{DM},$$

and the overall agreement of the decision-maker *DM* with the opinion of the whole group about the assignments of all alternatives from the set $A = \{a_1, \dots, a_m\}$ is

$$G^{DM} = \sum_{i=1}^{m} \frac{g_i}{m}.$$

3.2 Adjustment of the decision for the most contradictive group member

Let the set of all alternatives be divided with regard to the most contradictive decision-maker *DM* into two disjunctive subsets $A = A' \cup A''$, $A' \cap A'' = \emptyset$, so that $A' \subset A$ is a subset of alternatives that have to be reassigned by *DM* into different categories, while $A'' \subset A$ is a subset of alternatives that preserve their assessments. Alternatives belong to these subsets according to the following rules:

- $a_i \in A' \Leftrightarrow g_i^{DM} < 0.5$: (partial) disagreement of *DM* with the decision-making group about the assignment of the alternative a_i, i.e. less than half of group members sort a_i into the same category C_i^{DM} as *DM* does;
- $a_i \in A'' \Leftrightarrow g_i^{DM} \geq 0.5$: (partial) agreement of *DM* with the decision-making group about the assignment of the alternative a_i, i.e. more than half of group members sort a_i into the same category C_i^{DM} as *DM* does.

Alternatives are evaluated from the perspective of the decision-maker *DM* according to the set of preferential parameters P^{DM}, which is, for the purpose of generality of the paper, defined independently of the preference model, so that it is valid for various approaches,

such as the utility function, AHP or pseudo-criterion based outranking. Similarly, a general aggregation operator Θ is introduced to synthesize preferences into assessments:

$$\Theta\colon P^{DM} \rightarrow u^{DM}(a_i), \forall a_i \in A.$$

A prerequisite to ensure the convergence of individual opinions into a rational uniform decision is that lower and upper limits of parameter values are provided. Then, the algorithm for the automatic adjustment of preferential parameters can search in the constrained space:

$$\forall p \in P^{DM}\colon p_L^{DM} \leq p \leq p_H^{DM}.$$

The autonomous algorithm for the conformance of the most contradictive group member to collective judgements is specified below. In addition to aligning opinions it also aims at achieving robustness by maximizing distances between alternatives and lower and upper category limits b_i^{k-1} and b_i^k.

$conformation \leftarrow$ false
sort $\forall a_i \in A'\colon C_i^{new} \leftarrow C_i^{median}$
while $(\forall a_i \in A'\colon C_i^{new} \neq C_i^{DM}) \wedge (conformation =$ false)
try to derive new values of $P^{DM} = \{p\}$ by

$$\max \sum_{i=1}^{m} \left(u^{DM}(a_i) - b_i^{k-1}\right) + \left(b_i^k - u^{DM}(a_i)\right)$$

subject to

$$p_L^{DM} \leq p \leq p_H^{DM}, \forall p \in P^{DM}$$
$$\Theta\colon P^{DM} \rightarrow u^{DM}(a_i), \forall a_i \in A$$
$$u^{DM}(a_i) \geq b_i^{k-1}, \forall a_i \in A$$
$$u^{DM}(a_i) \leq b_i^k, \forall a_i \in A$$
$$k = \begin{cases} ord(C_i^{new}) & , a_i \in A' \\ ord(C_i^{DM}) & , a_i \in A'' \end{cases}$$

if it is possible to derive new values of P^{DM}
$conformation \leftarrow$ true
else
re-sort one alternative $a_i \in A'$ by
max G^{DM}
subject to $C_i^{new} \leftarrow \begin{cases} C_i^{new-1} & , ord(C_i^{DM}) < ord\left(C_i^{median}\right) \\ C_i^{new+1} & , ord(C_i^{DM}) > ord\left(C_i^{median}\right) \end{cases}$
end if
end while

4 RANKING

In ranking, alternatives are ordered from the best to the worst one(s) [15]. Three basic types of orders exist in multi-criteria decision analysis. In the complete order, all pairs of alternatives are in the relation of preference, which is denoted with $a_i \succ a_j$ and states that a_i is preferred to a_j. In the weak order, the decision-maker can also be indifferent between a_i and a_j. The indifference relation is denoted with $a_i \approx a_j$. Finally, the partial order introduces the incomparability relation $a_i ? a_j$. Two alternatives are incomparable, if their characteristics are so opposing that it is impossible and unreasonable to determine which of them is better.

4.1 Concordance of the decision-maker and the decision-making group

There exist two possibilities to identify the most discordant decision-maker. Firstly, the rank order of each observed decision-maker can be compared to the rank orders of all other group members. And secondly, the »average« rank order can be derived from all rank orders by applying an optimization algorithm. It is in both cases necessary to measure the distances between rank orders with appropriate metrics, which have been defined in the past [1, 8].

The first approach is chosen for the purpose of this research. Let the relation between a pair of alternatives $a_i \in A$ and $a_j \in A$ be denoted with $R_{ij} = R(a_i, a_j) \in \{\succ, \prec, \approx, ?\}$, where $\succ$ and $\prec$ represent preference, $\approx$ indifference, and ? incomparability. The distance between relations of two decision-makers DM_k and DM_l is then $d(R_{ij}^k, R_{ij}^l)$. Standard distances can be used [1]:

$$d(\succ, \prec) = 2 \cdot e, d(\succ, ?) = \frac{5}{3} \cdot e, d(\approx, ?) = \frac{4}{3} \cdot e, d(\approx, \prec) = e, \text{where } e \text{ is a constant.}$$

For each member of the decision-making group, the average normalized distance of his rank order to rank orders of other $o - 1$ decision-makers is calculated:

$$d(DM_k) = d_k = \frac{\sum_{\substack{l=1..o \\ l \neq k}} \sum_{i=1..m} \sum_{j=i+1..m} d(R_{ij}^k, R_{ij}^l) \cdot Z_{ij}^l}{e \cdot m \cdot (m-1) \cdot o \cdot (o-1)}.$$

The most discordant decision-maker is the one with the largest distance to others. Hence, the agreement degree of DM_k is $G_k = 1 - d_k$. In this formula, m and o denote the numbers of alternatives and decision-makers, respectively. Only $m \cdot (m-1)/2$ relations above the diagonal of the upper triangular matrix are considered, because all distances on the diagonal equal to 0, while all elements below the diagonal are the same as above it. The total distance is normalized with regard to the highest possible distance $2 \cdot e$ of a pair of alternatives. It is also weighted with a normalized frequency Z_{ij}^l / o of the relation in which alternatives a_i and a_j are for each compared decision-maker DM_l. In this way, the distance between rank orders of DM_k and DM_l depends on binary relations of the latter. Z_{ij}^l is obtained by counting types of relations for each pair of alternatives above the diagonal of the matrix, i.e. by determining the number of decision-makers for which alternatives a_i and a_j are in a certain relation:

$$Z_{ij}^{\succ} = card\{DM: R_{ij} = \succ\} = card\{DM: a_i \succ a_j\},$$
$$Z_{ij}^{\prec} = card\{DM: R_{ij} = \prec\} = card\{DM: a_i \prec a_j\},$$
$$Z_{ij}^{\approx} = card\{DM: R_{ij} = \approx\} = card\{DM: a_i \approx a_j\},$$
$$Z_{ij}^{?} = card\{DM: R_{ij} = ?\} = card\{DM: a_i ? a_j\}.$$

4.2 Adjustment of the decision for the most contradictive group member

1. Frequencies Z_{ij}^l are computed for each pair of alternatives and all decision-makers.
2. The decision-maker DM_k with the lowest agreement degree G_k is selected.
3. The rank order of DM_k is modified by maximizing the concordance of preference and indifference relations, and by minimizing the number of incomparability relations:

$$\forall a_i, a_j \in A: 1 \leq i \leq m, i < j:$$
$$R_{ij}^k = R_{ij}^{max} \Longrightarrow R_{ij}^k \text{ is preserved, } R_{ij}^{max} = ? \Longrightarrow R_{ij}^k \text{ is preserved.}$$

- Initially, the relation for each pair of alternatives conforms to the collective opinion maximally as is possible according to the table below, that is up to two levels.

- If the adjustment of preferential parameters is not possible by obeying the decision-maker's constraints, then the adjustment is relaxed for one level in each iteration, yet at most till the initial relation is reached.

R_{ij}^k	$\prec$	$\prec$	$\approx$	$\approx$	$\succ$	$\succ$
R_{ij}^{max}	$\succ$	$\approx$	$\succ$	$\prec$	$\approx$	$\prec$
Change of R_{ij}^k	$\succ$ or $\approx$	$\approx$	$\succ$	$\prec$	$\approx$	$\prec$ or $\approx$

4. Preferential parameters $P = \{p\}$ of DM_k are modified according to provided individual constraints on parameter values, and according to newly proposed relations.
5. If the adjustment is not possible, the algorithm returns to step 3 and relaxes the required relation change for one pair of alternatives, which is chosen in such a way to maximize the agreement degree G_k.

5 CONCLUSION

In this paper, mechanisms for the automated convergence of group consensus seeking procedures were defined. They were applied to two most common problematics of decision-making – ranking and sorting. Within the scope of future research work, the efficiency of the proposed general approaches will have to be evaluated, because a special case of dichotomic sorting was only systematically studied in the past. Additionally, robustness metrics for the general cases of ranking and sorting will be derived from the existent metrics for localized two-categorical sorting.

References

[1] Ben Khélifa, S., Martel, J.-M. A distance-based collective weak ordering. Group Decision and Negotiation [GDN], 10 (4), 317–329, 2001.

[2] Brans, J. P., Mareschal, B. PROMETHEE Methods. Multiple Criteria Decision Analysis: State of the Art Surveys. Springer, Boston, 163–196, 2005.

[3] Bregar, A. Efficiency of problem localization in group decision-making. Proceedings of 10th International Symposium on Operational Research in Slovenia, 139–149, 2009.

[4] Bregar, A. Outranking methods and their application in group decision-making: A case study. Uporabna informatika, 19 (2), 75–90, 2011.

[5] Bregar, A., Györkös, J., Jurič, M. B. Interactive aggregation/disaggregation dichotomic sorting proce-dure for group decision analysis based on the threshold model. Informatica, 19 (2), 161–190, 2008.

[6] Bregar, A., Györkös, J., Jurič, M. B. Robustness and visualization of decision models. Informatica, 33 (3), 385–395, 2009.

[7] Cai, F.-L., Liao, X., Wang, K.-L. An interactive sorting approach based on the assignment examples of multiple DM with different priorities. Annals of Operations Research, 197 (1), 87–108, 2012.

[8] Cook, W. D., Kress, M., Seiford, L. M. A general framework for distance-based consensus in ordinal ranking models. European Journal of Operational Research [EJOR], 96 (2), 392–397, 1996.

[9] Damart, S., Dias, L. C., Mousseau, V. Supporting groups in sorting decisions: Methodology and use of a multi-criteria aggregation/disaggregation DSS. Decision Support Systems, 43 (4), 1464–1475, 2007.

[10] Espinasse, B., Picolet, G., Chouraqui, E. Negotiation support systems: A multi-criteria and multi-agent approach. EJOR, 103 (2), 389–409, 1997.

[11] Herrera-Viedma, E., Herrera, F., Chiclana, F. A consensus model for multiperson decision making with different preference structures. IEEE Trans. on Systems, Man and Cybernetics, 32 (3), 394–402, 2002.

[12] Hodgkin, J., Belton, V., Koulouri, A. Supporting the intelligent MCDA user: A case study in multi-person multi-criteria decision support. EJOR, 160 (1), 172–189, 2005.

[13] Matsatsinis, N. F., Grigoroudis, E., Samaras, A. P. Aggregation and disaggregation of preferences for collective decision-making. GDN, 14 (3), 217–232, 2005.

[14] Saaty, T. L. Group decision making and the AHP. The Analytic Hierarchy Process: Applications and Studies. Springer, New York, 59–67, 1989.

[15] Roy, B. Multicriteria Methodology for Decision Aiding. Kluwer Academic Publishers, Dordrecht, 1996.

[16] Zopounidis, C., Doumpos, M. Multicriteria classification and sorting methods: A literature review. EJOR, 138 (2), 229–246, 2002.

THE MULTIPLE-CRITERIA MODEL BASED ON EXPLORATORY FACTOR ANALYSIS AND PRACTICAL EXPERIENCE: THE CASE OF HUMAN RESOURCE MANAGEMENT

Vesna Čančer and **Simona Šarotar Žižek**
University of Maribor, Faculty of Economics and Business
Razlagova 14, SI-2000 Maribor, Slovenia
{vesna.cancer,simona.sarotar-zizek}@uni-mb.si

Abstract: This paper develops the multiple-criteria model for the assessment of human resource management (HRM) in organizations considering the exploratory factor analysis results in problem structuring and the experts' judgments in measuring criteria's importance. The innovative aspect of this paper is that it gives solutions for group decision making with missing judgments about the criteria's importance. Application possibilities of the results of the multiple-criteria assessment of HRM are illustrated and discussed via a real-life case of organizations in Slovenia.

Keywords: exploratory factor analysis, group decision-making, human resource management, multiple-criteria model, weighting method.

1 INTRODUCTION

This paper uses an exploratory factor analysis (EFA) to structure the multiple-criteria model for the assessment of human resource management (HRM). It also brings solutions for measuring the local alternatives' values with respect to indicators and for criteria weighting in group processes.

A way to support group processes is the use of the group model by the web-based software called Web-HIPRE [9], where the value of the i^{th} alternative $v(X_i)$ is expressed as the weighted arithmetic mean of the aggregate alternatives' values obtained by decision makers [9]:

$$v(X_i) = \sum_{k=1}^{d} u_k v_k(X_i), \text{ for each } i = 1, 2, \ldots, n, \quad (1)$$

where u_k is the weight of the k^{th} decision maker, $v_k(X_i)$ is the value of the i^{th} alternative obtained by the k^{th} decision maker, and d is the number of decision makers. In multiple-criteria decision making based on preference elicitation is assumed that decision makers are able to express their judgments about all criteria sets that are structured in the hierarchy to obtain the aggregate alternatives' values. The innovative aspect of this paper is that it is not necessary that all participants express their judgments about the importance of all criteria. When in the respondents' answers are missing judgments about the importance of one or more criteria or even sets of them, the means of the individuals' points assigned to criteria's importance by the rest of participants are included in the model instead of (1) to obtain the aggregate value of the HRM measure.

The organization of this paper is as follows. The second section presents the selected methodological particularities used to develop the multiple-criteria model for the assessment of HRM. The development of the multiple-criteria model based on the EFA – objective statistical approach, completed by subjective group criteria weighting and measuring local alternatives' values, as well as the assessment of HRM is presented and illustrated via a practical application in organizations in Slovenia in the third section. The last section discusses the application possibilities of the presented solutions.

2 METHODOLOGICAL SOLUTIONS FOR THE DEVELOPMENT OF THE MULTIPLE-CRITERIA MODEL

2.1 Problem structuring by exploratory factor analysis

Factor analysis based on principal component analysis extraction method has already been used to reduce a large number of variables to a smaller number of factors for modelling purposes and to determine which sets of items should be grouped together in the multiple-criteria model [1]. Differently from this, EFA [3, 5, 6] was primarily used in our survey about HRM to explore the multiple-criteria model, to determine the number of constructs (i.e., factors, the first-level criteria) influencing the set of measures (i.e., indicators, the second-level criteria or attributes) of HRM, and to determine the strength of the relationship between each factor and each observed indicator. It was therefore used to select the "best" indicators of each factor.

2.2 Group criteria weighting

When assessing the HRM with respect to multiple criteria, the importance of criteria should be determined. In practical applications, decision makers often have difficulty defining the relative importance of criteria directly; thus, the criteria's importance can be expressed using several methods [2]. In this paper, special attention is given to the use of the SWING method [11] based on the interval scale, because experts in the field of HRM promote the use of this method. In SWING, a decision maker is first asked to assign 100 points to the most important criterion change from the worst criterion level to the best level and then to assign points (≤ 100, but ≥ 10) to reflect the importance of the criterion change from the worst criterion level to the best level relative to the most important criterion change [11].

Preference elicitation has traditionally been carried out in public meetings, exhibitions and workshops, as well as with questionnaires and interviews [9]. Let the individual judgments be expressed by using questionnaires. Further, let us allow missing judgments on the importance of one or more criteria.

Differently from obtaining the individual's weight of the j^{th} criterion, based on the judgments made on the interval scale, by normalization [2, 4, 11], we expressed the group weight of the j^{th} criterion, g_j, by the means of the individuals' judgments about the criteria's importance:

$$g_j = \frac{\left(\sum_{l=1}^{r_j} t_j\right) \Big/ r_j}{\sum_{j=1}^{m} \left(\left(\sum_{l=1}^{r_j} t_j\right) \Big/ r_j\right)}, \tag{2}$$

where t_j corresponds to the points given to the j^{th} criterion, m is the number of criteria and r_j is the number of the respondents that expressed the judgments about the j^{th} criterion importance. When the criteria are structured in two levels (which is the case in the practical example examined in this paper), the weight of the s^{th} attribute of the j^{th} criterion, g_{js}, is expressed as:

$$g_{js} = \frac{\left(\sum_{l=1}^{r_{js}} t_{js}\right) \Big/ r_{js}}{\sum_{s=1}^{p_j} \left(\left(\sum_{l=1}^{r_{js}} t_{js}\right) \Big/ r_{js}\right)} \tag{3}$$

where t_{js} corresponds to the points given to the s^{th} attribute of the j^{th} criterion, p_j is the number of the j^{th} criterion sub-criteria and r_{js} is the number of the respondents that expressed their judgments about the importance of the s^{th} attribute of the j^{th} criterion.

When eliciting weights for the highest level criteria, it is important that the respondent is fully aware of the meaning of the criteria [8]. Therefore, a bottom-up approach is appropriate for use in which the weights are first elicited to the attributes on the lowest level.

2.3 Measuring local and global alternatives' values

Let the data about each attribute be obtained by a questionnaire and let the measurement scale used in the survey be the interval or the ratio one. When alternatives are organizations or even groups of them, data about HRM in each group can be obtained as the mean of the respondents' data for each attribute. When greater agreement with the statements means that the HRM is better, the local alternatives' values with respect to the attributes can be obtained by increasing value functions. The lower and upper bounds of value functions should be determined for each attribute: the lower bound is less than or equal to the lowest datum at the considered attribute whereas the upper bound is greater than or equal to the highest datum at the considered attribute.

The additive model (see, e.g., [2]) was used to obtain the level of HRM – namely, the aggregate value of the HRM measure in several groups of organizations. As the criteria in our model are structured in two levels, the alternatives' values with respect to the first level criteria were obtained by:

$$v_j(X_i) = \sum_{s=1}^{p_j} g_{js} v_{js}(X_i), \text{ for each } i = 1, 2, \ldots n, \tag{4}$$

where $v_j(X_i)$ is the value of the i^{th} alternative with respect to the j^{th} criterion and $v_{js}(X_i)$ is the local value of the i^{th} alternative with respect to the s^{th} attribute of the j^{th} criterion. The aggregate alternatives' values were obtained by:

$$v(X_i) = \sum_{j=1}^{m} g_j \left(\sum_{s=1}^{p_j} g_{js} v_{js}(X_i) \right), \text{ for each } i = 1, 2, \ldots, n. \tag{5}$$

3 THE DEVELOPMENT OF THE HRM MODEL

Considering the theoretical foundations of HRM compiled by Šarotar Žižek [10] and the answers obtained via the in-depth interviews with five academics in and senior managers in 15 organizations, the original questionnaire about HRM in organizations was built. It consists of 22 Likert-type statements (from 1 – *absolutely not agree* to 7 – *completely agree*) designed for managers to express their opinions about HRM. During April 2011 until June 2011, 320 fulfilled questionnaires were gathered from the managers in 2409 randomly selected organizations in Slovenia, of which 260 respondents were classified by industry [10].

Table 1: The results of the EFA for HRM in organizations in Slovenia.

Statement	***Cronbach's alpha***	***Communality***	***Factor loadings***		
With employees, we established a dialogue inside the regular communication.	0.848	0.709	0.802		
Employees are included in the strategic management process with their suggestions.		0.599	0.732		
In the organization we have a system to motivate employees.		0.744	0.814		
The organization established rewards associates.		0.585	0.703		
We appreciate and practice team work.		0.471	0.624		
We know the characteristics of the supply of jobs in our industry.	0.788	0.623		0.762	
We know the characteristics of the supply of jobs in our region.		0.629		0.775	
We know the dangers of "emigration" of our employees in other organizations.		0.645		0.743	
We are familiar with the working conditions offered by our competitors to their employees.		0.576		0.722	
The organization regularly uses tutoring - mentoring.	0.769	0.712			0.788
The organization regularly uses coaching.		0.771			0.851
In the organization we have the diversity management for employees.		0.565			0.703
Kaiser-Meyer-Olkin measure: 0.851 Cumulative percentage of explained variance: 63.578%					

By the EFA we tested the dimensionality of the constructs of HRM. The value of Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy presented in Table 1 indicates that factor analysis is appropriate, since KMO > 0.5 (KMO = 0.851). Table 1 shows that the three constructs explain the most variance for all variables, namely 63.578 %. All communalities that express the variance in observed indicators accounted for by common factors are greater than 0.4. The indicators are accordingly weighted on individual factors. This is also proved by factor loadings, which are all greater than 0.6. Table 1 illustrates a very clean factor structure in which convergent and discriminant validity are evident by the high loadings within factors, and no cross-loadings between factors. The Cronbach's alpha coefficients show adequate reliability for all three constructs – they are all greater than 0.6 (see, e.g., [3, 6, 7]). In this solution from the initial 22, we kept 12 indicators that are thus best influenced by the factors (Tables 1 and 2). Within the EFA we got three constructs, namely: basic approaches, new approaches, and employee orientated approaches (Table 2). Constructs are considered as factors, and statements are considered as indicators (attributes). Table 2 presents the sets of indicators that are influenced by the factors, and thus presents the structure of the multiple-criteria model for the assessment of HRM in organizations.

Following the criteria structure in Table 2 and the basics for the weights elicitation by the SWING method, we constructed a questionnaire to obtain the criteria's importance judgments. It was sent to 20 organizations in Slovenia with well-developed HRM. From March 2013 until May 2013, 11 fulfilled questionnaires were gathered from managers. The number of respondents that fulfilled the questionnaires about the criteria's importance is as follows: $r_j = 10$, $r_{1s} = 11$, $r_{2s} = 9$, $r_{3s} = 3$. Table 2 presents also the factor weights obtained by (2), and the indicator weights obtained by (3).

Table 2: The weights based on professional experience.

Factor		***Indicator***	
Name	***Weight***	***Name***	***Weight***
Basic approaches	$g_1 = 0.315$	With employees, we established a dialogue inside the regular communication.	$g_{11} = 0.234$
		Employees are included in the strategic management process with their suggestions.	$g_{12} = 0.191$
		In the organization we have a system to motivate employees.	$g_{13} = 0.191$
		The organization established rewards associates.	$g_{14} = 0.172$
		We appreciate and practice team work.	$g_{15} = 0.213$
New approaches	$g_2 = 0.358$	We know the characteristics of the supply of jobs in our industry.	$g_{21} = 0.285$
		We know the characteristics of the supply of jobs in our region.	$g_{22} = 0.256$
		We know the dangers of "emigration" of our employees in other organizations.	$g_{23} = 0.235$
		We are familiar with the working conditions offered by our competitors to their employees.	$g_{24} = 0.224$
Employee orientated approaches	$g_3 = 0.327$	The organization regularly uses tutoring - mentoring.	$g_{31} = 0.405$
		The organization regularly uses coaching.	$g_{32} = 0.300$
		In the organization we have the diversity management for employees.	$g_{33} = 0.295$

The hierarchy structure completed with weights that is presented in Table 2 was applied for the assessment of HRM in groups of organizations. An organization was classified in the proper group with respect to its industry. Thus we obtained seven alternatives (Table 3). Considering the means of the respondents' answers about the statements regarding HRM, the local values of alternatives – different groups of organizations in Slovenia – were measured by using increasing value functions. The lower and upper bounds of value functions were determined for each attribute: to differentiate between the values of alternatives, the lower bound is equal to the lowest mean at the considered attribute, and the upper bound is equal to the highest mean at the considered attribute. Table 3 presents the alternatives' values with respect to each factor obtained by (4) and the aggregate alternatives' values obtained by (5).

Table 3: The alternatives' values – the HRM measures for groups of organizations in Slovenia.

	X_1	X_2	X_3	X_4	X_5	X_6	X_7
$v_1(X_i)$	0.397	1.000	0.486	0.660	0.511	0.030	0.622
$v_2(X_i)$	0.141	0.689	0.468	0.343	0.242	0.441	1.000
$v_3(X_i)$	0.141	0.595	0.160	0.102	0.255	0.371	0.414
$v(X_i)$	0.222	0.756	0.373	0.364	0.331	0.288	0.689
Rank	7.	1.	3.	4.	5.	6.	2.

Symbols: $v_1(X_i)$ – the i^{th} alternative's value with respect to 'basic approaches'; $v_2(X_i)$ – the i^{th} alternative's value with respect to 'new approaches'; $v_3(X_i)$ – the i^{th} alternative's value with respect to 'employee orientated approaches; $v(X_i)$ – the i^{th} alternative's aggregate value – the human resource management measure; X_1 – the manufacturing organizations; X_2 – the real estate, renting and business activities organization,; X_3 – the construction organizations, X_4 – the other community, social and personal service activities organizations; X_5 – the wholesale and retail trade organizations; X_6 – the hotels and restaurants organizations; X_7 – the transport, storage and communication organizations.

In the presented application of assessing HRM in organizations it can be concluded that alternative X_2 – the real estate, renting and business activities organizations – has the highest aggregate value (Table 3). X_2 has also the highest value with respect to basic approaches and employee orientated approaches. It is followed by X_7 – the transport, storage and

communications organizations (Table 3), which has also the highest value with respect to new approaches. The lowest aggregate value is achieved by X_1 – the manufacturing organizations. Its main key failure factor is new approaches, and – although its values with respect to basic approaches $v_1(X_1)$ and employee orientated approaches $v_3(X_1)$ are the second worst ones, these two factors can be considered as the key failure ones, as well. Studying the local values of X_1 with respect to each factor and comparing them with the ones of X_2 for basic approaches and employee oriented approaches, and of X_7 for new approaches (Table 3), we can plan possible actions to improve HRM in X_1.

4 CONCLUSIONS

The approach presented in this paper is appropriate when we want to assess a multidimensional concept with respect to multiple criteria; when the data are obtained from respondents with questionnaires and measured on an interval and ratio scales; when the weights are determined by preference elicitation from several stakeholders. Because the presented solutions for criteria weighting allow missing judgments, they enable respondents to express their judgments about the criteria's importance for the sets of criteria for which they are experts. The advantages of this approach come into forefront when the criteria structure includes conflict criteria, decomposed into several attributes, as well; in such cases it is appropriate that respondents judge about their field of expertize.

References

[1] Begičević, N., Divjak, B., Hunjak, T. (2007): Prioritization of e-learning forms: a multicriteria Methodology. *Central European Journal of Operations Research*, 15: 405–419.

[2] Belton, V., Stewart, T. J. (2002): *Multiple Criteria Decision Analysis: An Integrated Approach*. Boston, Dordrecht, London: Kluwer Academic Publishers.

[3] Costello, A. B., Osborne, J. (2005): Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation*, 10(7): 1-9.

[4] Čančer, V. (2012): Criteria weighting by using the 5Ws & H technique. *Business Sysems Research Journal*, 3(2): 41-48. doi: 10.2478/v10305-012-0011-3.

[5] Čančer, V., Šarotar Žižek, S. (2013): The multiple criteria assessment of social responsibility in organizations. *Croatian operational research review*, 4: 200-210.

[6] DeCoster, J. (1998): *Overview of Factor Analysis*. http://www.stat-help.com/factor.pdf, accessed 4. 6. 2013.

[7] Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E. (2010): *Multivariate data analysis* (7th ed.). New York: Prentice-Hall International.

[8] Marttunen, M., Hämäläinen, R. P. (2008): The Decision Analysis Interview Approach in the Collaborative Management of a Large Regulated Water Course. *Environmental Management*, 42: 1026-1042.

[9] Mustajoki, J., Hämäläinen, R. P., Marttunen, M. (2004): Participatory multicriteria decision analysis with Web-HIPRE: a case of lake regulation policy. *Environmental Modelling & Software*, 19: 537-547.

[10] Šarotar Žižek, S. (2012): *Vpliv psihičnega dobrega počutja na temelju zadostne in potrebne osebne celovitosti zaposlenega na uspešnost organizacije*. (*Influence of psychical well-being on success of organization on a basis of requisite personal holism of an employee*, in Slovene only). Maribor: University of Maribor, Faculty of Economics and Business.

[11] Von Winterfeldt, D., Edwards, W. (1986): *Decision analysis and behavioral research*. Cambridge: Cambridge University Press.

JUDGEMENT ON SOME APPROACHES FOR DERIVING INTERVAL GROUP MATRICES IN ANALYTIC HIERARCHY PROCESS

Petra Grošelj, Lidija Zadnik Stirn
University of Ljubljana, Biotechnical Faculty
Jamnikarjeva 101, 1000 Ljubljana, Slovenia
petra.groselj@bf.uni-lj.si, lidija.zadnik@bf.uni-lj.si

Abstract: The paper discusses group analytic hierarchy process (AHP), a well known multiple criteria method with several decision makers involved in a decision process. In cases, when the decision makers are exposed to the subjectivity and/or the lack of information, the individual comparison matrices are aggregated into a joint interval comparison matrix. Four aggregation methods: MIN-MAX, MEDINT, ADEXTREME and GEOSTDINT are considered, compared and applied to the problem of management of the Pohorje area, with six scenarios and five decision makers.

Key words: multiple criteria decision making; group decision making; analytic hierarchy process; interval judgments; management of natural resources.

1 INTRODUCTION

Analytic hierarchy process (AHP) [11], is a well-known approach for handling multi-criteria decision making problems. AHP enables combining empirical data and subjective judgments and also intangible and immeasurable criteria. It is based on pairwise comparisons of criteria and alternatives which are hierarchically structured. The 1-9 ratio scale is used. All comparisons are gathered in a pairwise comparison matrix A. When several decision makers are involved, group AHP [12] replaces AHP which deals with only one decision maker.

In group AHP, the main problems are how to aggregate the individual comparison matrices into a group comparison matrix and how to calculate weights from such matrix [5].

The complexity and uncertainty of the decision problems, the subjectivity and the lack of information of decision makers can be hardly expressed with the exact values. Interval judgments can be more suitable in such cases. When dealing with interval comparison matrices in group AHP two main methodological problems emerge:

a) how to aggregate the individual point-valued judgments into the interval group comparison matrix,
b) how to determine (calculate) the weights from interval group comparison matrix.

In the paper, we focus on the first problem. Four methods for aggregation of individual comparison matrices into group interval judgment are presented. These approaches are:

1. MIN-MAX approach, which was already studied and applied by several authors, see for example [13]; intervals are constructed using minimum and maximum individual judgments,
2. MEDINT approach, which uses the median and thus pays considerable regard on intermediate individual judgments [4], which is not the case with MIN-MAX approach,
3. ADEXTREME approach, where all individual judgments have the impact on the bounds of the group interval, but not all have equal power [7],
4. GEOSTDINT approach, which uses weighted geometric mean and standard deviation of individual judgments [6].

Further, for deriving interval weights from interval group comparison matrix A^{group}, we propose the approach of separating A^{group} into two point-valued comparison matrices [13].

Finally, a numerical example to illustrate the presented methods for aggregating the individual comparison matrices into group interval judgment, and deriving the weights from interval comparison matrices, is presented. The application is based on NATREG project [10]

which deals with selecting the optimal strategy for management of Pohorje, the mountain area in NE part of Slovenia, which is under Natura 2000 protection. There are six scenarios and five decision makers considered. At the end, in conclusion, a brief comparison of the results obtained by discussed methods/approaches and some open methodological questions are tackled.

2 INTERVAL COMPARISON MATRICES IN AHP

Let $A=\left(\left[l_{ij},u_{ij}\right]\right)_{n\times n}$ be $n\times n$ interval comparison matrix, which is a reciprocal matrix, i.e., $l_{ij}=1/u_{ij}$ and $u_{ij}=1/l_{ij}$ for all *i,j=1,…,n*. For deriving interval weights from *A*, we can separate *A* into two point-valued reciprocal matrices $A_L=\left(a_{ij}^L\right)$ and $A_U=\left(a_{ij}^U\right)$ [9]:

$$a_{ij}^L=\begin{cases}l_{ij}, & i<j\\ 1, & i=j,\\ u_{ij}, & i>j\end{cases}\qquad a_{ij}^U=\begin{cases}u_{ij}, & i<j\\ 1, & i=j.\\ l_{ij}, & i>j\end{cases} \tag{1}$$

Then, geometric mean method [2] should be used for deriving weights from A_L and A_U. Further, the point-valued weights are combined to gain the interval weights of *A*:

$$w_i=\left[w_i^L,w_i^U\right]=\left[\min\left\{w_i^{A_L},w_i^{A_U}\right\},\max\left\{w_i^{A_L},w_i^{A_U}\right\}\right]. \tag{2}$$

Ranking of the interval weights is not always easy if interval weights overlap. If we assume that the interval weights are uniformly distributed, the probability - degree formula [3], [13], [14], [15] can be used for obtaining the probabilities in the matrix *P*, presenting the degrees of preference:

$$p_{ij}=P\left(\omega_i>\omega_j\right)=\frac{\max\left\{0,\omega_i^U-\omega_j^L\right\}-\max\left\{0,\omega_i^L-\omega_j^U\right\}}{(\omega_i^U-\omega_i^L)+(\omega_j^U-\omega_j^L)},\ i,j=1,\ldots,n,\ i\neq j \tag{3}$$

$$P=\begin{bmatrix}- & p_{12} & \cdots & p_{1n}\\ p_{21} & - & \cdots & p_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ p_{n1} & p_{n2} & \cdots & -\end{bmatrix}. \tag{4}$$

The preference ranking order is then provided using row-column elimination method [13].

3 APPROACHES FOR DERIVING INTERVAL GROUP MATRICES

Let $A^{(k)}=\left(a_{ij}^{(k)}\right)_{n\times n}$, *k=1,…,m* be point-valued individual comparison matrices of *m* decision makers. Aggregated interval group matrix $A^{group}=\left(\left[l_{ij}^{group},u_{ij}^{group}\right]\right)_{n\times n}$ can be obtained in several ways.

3.1 MIN-MAX approach

Group interval judgments can be constructed using minimum and maximum individual judgments for the bounds of the intervals [13], [1]:

$$l_{ij}=\min_{k\in\{1,\ldots,m\}}a_{ij}^{(k)}\ \text{and}\ u_{ij}=\max_{k\in\{1,\ldots,m\}}a_{ij}^{(k)}. \tag{5}$$

3.2 MEDINT approach

MEDINT approach [4] presumes that the degree of influence on the lower bound should be greater for smaller values, smaller for the values that are close to the median and zero for values that are greater than median. Similar, the upper bound of the interval should be influenced by all values that are greater or equal to the median.

$$l_{ij} = \prod_{k=1}^{m} \left(c_{ij}^{(k)} \right)^{u_k^L} \text{ and } u_{ij} = \prod_{k=1}^{m} \left(c_{ij}^{(k)} \right)^{u_k^U}, \tag{6}$$

where $c_i^{(k)}$ is the ith largest value from the set $\left\{ a_1^{(k)}, ..., a_m^{(k)} \right\}$ and the weighting vector $U = \left(u_1, ..., u_m \right)$ depends on m (m could be even or odd):

$$U_L^{odd} = \left(\underbrace{0, ..., 0}_{\frac{m-1}{2}}, \frac{1}{s_{\frac{m+1}{2}}}, \frac{2}{s_{\frac{m+1}{2}}}, ..., \frac{\frac{m-1}{2}}{s_{\frac{m+1}{2}}}, \frac{\frac{m+1}{2}}{s_{\frac{m+1}{2}}} \right) \text{ and } U_U^{odd} = \left(\frac{\frac{m+1}{2}}{s_{\frac{m+1}{2}}}, \frac{\frac{m-1}{2}}{s_{\frac{m+1}{2}}}, ..., \frac{2}{s_{\frac{m+1}{2}}}, \frac{1}{s_{\frac{m+1}{2}}}, \underbrace{0, ..., 0}_{\frac{m-1}{2}} \right) \tag{7}$$

where $\frac{m+1}{2}$ is the median of numbers $1, 2, ..., m$ and $s_{\frac{m+1}{2}} = \frac{(m+1)(m+3)}{8}$ is the sum of numbers from 1 to $\frac{m+1}{2}$ or

$$U_L^{even} = \left(\underbrace{0, ..., 0}_{\frac{m}{2}}, \frac{1}{s_{\frac{m}{2}}}, \frac{2}{s_{\frac{m}{2}}}, ..., \frac{\frac{m-2}{2}}{s_{\frac{m}{2}}}, \frac{\frac{m}{2}}{s_{\frac{m}{2}}} \right) \text{ and } U_U^{even} = \left(\frac{\frac{m}{2}}{s_{\frac{m}{2}}}, \frac{\frac{m-2}{2}}{s_{\frac{m}{2}}}, ..., \frac{2}{s_{\frac{m}{2}}}, \frac{1}{s_{\frac{m}{2}}}, \underbrace{0, ..., 0}_{\frac{m}{2}} \right) \tag{8}$$

where if m is an even number, then median of numbers $1, 2, ..., m$ is not an integer and $s_{\frac{m}{2}} = \frac{m(m+2)}{8}$ is the sum of numbers from 1 to $\frac{m}{2}$, which are smaller than median.

3.3 ADEXTREME approach

ADEXTREME approach [7] presumes that the smallest value has an influence one half and all other values together have an influence one half on the lower bound of the interval. The highest value has influence one half and all other values together have an influence one half on the upper bound:

$$l_{ij} = \left(\min_{k \in \{1, ..., m\}} a_{ij}^{(k)} \right)^{(1/2)} \prod_{\substack{k=1 \\ k \neq m}}^{t} \left(a_{ij}^{(k)} \right)^{1/(2t-2)} \text{ and } u_{ij} = \left(\max_{k \in \{1, ..., m\}} a_{ij}^{(k)} \right)^{(1/2)} \prod_{\substack{k=1 \\ k \neq m}}^{t} \left(a_{ij}^{(k)} \right)^{1/(2t-2)}. \tag{9}$$

3.4 GEOSTDINT approach

Weighted geometric mean method [12] is the main approach for aggregation of individual judgments into point-valued group judgments. The main statistic for measuring the dispersion of values around the geometric mean is the geometric standard deviation, which can help by making interval group judgments. Let

$$a_{ij}^{(GMM)} = \sqrt[m]{\prod_{k=1}^{m} a_{ij}^{(k)}}, \; i, j = 1, ..., n \tag{10}$$

be geometric mean of individual judgments. Then, $\ln a_{ij}^{(GMM)}$ is the arithmetic mean of the set $\left\{ \ln a_{ij}^{(1)}, ..., \ln a_{ij}^{(m)} \right\}$, whose standard deviation is equal to

$$\ln s_{ij}^{(GMM)} = \sqrt{\frac{\sum_{k=1}^{m}\left(\ln a_{ij}^{(k)} - \ln a_{ij}^{(GMM)}\right)^2}{m-1}}, \tag{11}$$

which reduces to the geometric standard deviation

$$s_{ij}^{(GMM)} = \exp\sqrt{\frac{\sum_{k=1}^{m}\left(\ln \frac{a_{ij}^{(k)}}{a_{ij}^{(GMM)}}\right)^2}{m-1}}. \tag{12}$$

The group interval weights are then defined as [6]:

$$l_{ij} = \frac{a_{ij}^{(GMM)}}{s_{ij}^{(GMM)}} \text{ and } u_{ij} = a_{ij}^{(GMM)} s_{ij}^{(GMM)}. \tag{13}$$

4 CASE STUDY

The NATREG project [10] has defined six strategic goals (scenarios), which can contribute to the realization of the vision Pohorje 2030 [8]. These six goals are:
1. High quality life of locals, 2. Preservation of nature and landscape, 3. Sustainable tourism and limited visit, 4. Environmental and consumer friendly usage of natural resources, 5. Environmental and consumer friendly mobility and good infrastructure, 6. Preserved cultural heritage and local tradition. We selected five experts who took part in the NATREG project. They pairwise compared the six strategic goals. Their comparison matrices, their weights, calculated by geometric mean method [2], and the consistency ratios are:

Expert 1

$$\begin{bmatrix} 1 & 1/3 & 2 & 1/3 & 1/2 & 1/3 \\ 3 & 1 & 1/2 & 1 & 1 & 1/2 \\ 1/2 & 2 & 1 & 1/3 & 1/2 & 1/2 \\ 3 & 1 & 3 & 1 & 1 & 1/2 \\ 2 & 1 & 2 & 1 & 1 & 1/2 \\ 3 & 2 & 2 & 2 & 2 & 1 \end{bmatrix}; \quad \text{Preferences} \begin{bmatrix} 0.094 \\ 0.152 \\ 0.115 \\ 0.196 \\ 0.164 \\ 0.278 \end{bmatrix}$$

CR=0.084

Expert 2

$$\begin{bmatrix} 1 & 3 & 1/5 & 1 & 1 & 1 \\ 1/3 & 1 & 1 & 1 & 1 & 1 \\ 5 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}; \quad \text{Preferences} \begin{bmatrix} 0.162 \\ 0.135 \\ 0.249 \\ 0.151 \\ 0.151 \\ 0.151 \end{bmatrix}$$

CR=0.098

Expert 3

$$\begin{bmatrix} 1 & 1 & 3 & 2 & 2 & 1 \\ 1 & 1 & 2 & 2 & 3 & 1 \\ 1/3 & 1/2 & 1 & 1 & 1 & 1/2 \\ 1/2 & 1/2 & 1 & 1 & 2 & 1/2 \\ 1/2 & 1/3 & 1 & 1/2 & 1 & 1/2 \\ 1 & 1 & 2 & 2 & 2 & 1 \end{bmatrix}; \quad \text{Preferences} \begin{bmatrix} 0.234 \\ 0.232 \\ 0.102 \\ 0.124 \\ 0.092 \\ 0.217 \end{bmatrix}$$

CR=0.012

Expert 4

$$\begin{bmatrix} 1 & 1/2 & 1/2 & 1 & 1/2 & 1 \\ 2 & 1 & 1 & 1 & 2 & 1 \\ 2 & 1 & 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 2 & 1/2 & 1/2 & 1 & 1 & 1/2 \\ 1 & 1 & 1 & 1 & 2 & 1 \end{bmatrix}; \quad \text{Preferences} \begin{bmatrix} 0.118 \\ 0.202 \\ 0.202 \\ 0.162 \\ 0.133 \\ 0.183 \end{bmatrix}$$

CR=0.031

Expert 5

$$\begin{bmatrix} 1 & 5 & 3 & 1 & 1 & 1 \\ 1/5 & 1 & 1/2 & 1/2 & 2 & 1/2 \\ 1/3 & 2 & 1 & 1/2 & 2 & 2 \\ 1 & 2 & 2 & 1 & 2 & 2 \\ 1 & 1/2 & 1/2 & 1/2 & 1 & 1/2 \\ 1 & 2 & 1/2 & 1/2 & 2 & 1 \end{bmatrix}; \quad \text{Preferences} \begin{bmatrix} 0.258 \\ 0.095 \\ 0.161 \\ 0.230 \\ 0.103 \\ 0.152 \end{bmatrix}$$

CR=0.094

The individual judgments have been aggregated using four interval group approaches. The results are presented in Table 1. The comparison of results is in Figure 1.

Results show that all approaches place one strategic goal on the first place – Preservation of cultural heritage and local tradition. After the first goal, the following goals are pursued - Environmental and consumer friendly usage of natural resources, High quality life of locals and Preserved nature and landscape. The ranking of these three goals is different and depends on the selected approach. With any selected approach, Sustainable tourism and limited visit, and Environmental and consumer friendly mobility and good infrastructure are ranked on the last two places.

Table 1: The ranks of six strategic goals of NATREG project using four interval group approaches

approach	ranking
MIN,MAX	$w_6 \overset{58.4\%}{\succ} w_4 \overset{50.9\%}{\succ} w_1 \overset{59.8\%}{\succ} w_2 \overset{65.6\%}{\succ} w_3 \overset{70.0\%}{\succ} w_5$
MEDINT	$w_6 \overset{70.5\%}{\succ} w_4 \overset{51.8\%}{\succ} w_1 \overset{58.8\%}{\succ} w_2 \overset{70.2\%}{\succ} w_3 \overset{79.9\%}{\succ} w_5$
ADEXTREME	$w_6 \overset{71.9\%}{\succ} w_4 \overset{100.0\%}{\succ} w_2 \overset{50.3\%}{\succ} w_1 \overset{56.4\%}{\succ} w_3 \overset{100\%}{\succ} w_5$
GEOSTDINT	$w_6 \overset{55.2\%}{\succ} w_1 \overset{56.4\%}{\succ} w_4 \overset{72.8\%}{\succ} w_2 \overset{65.8\%}{\succ} w_3 \overset{73.2\%}{\succ} w_5$

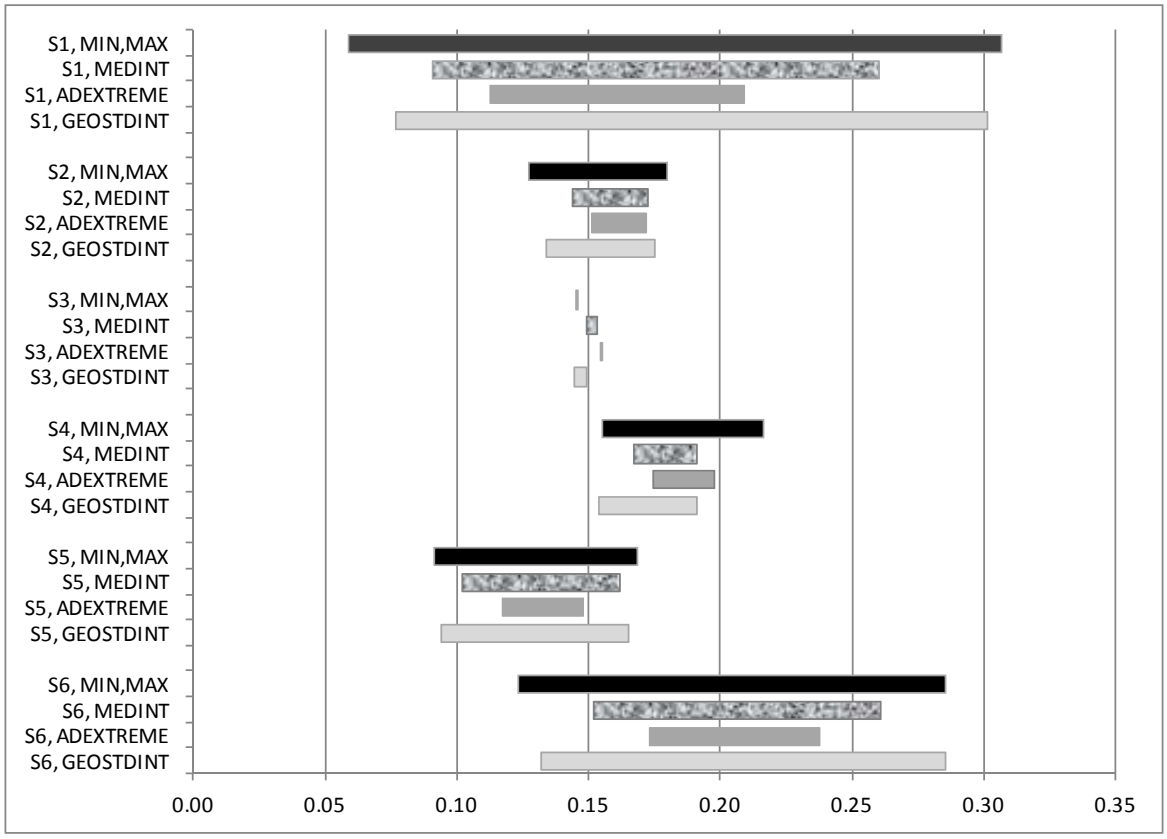


Figure 1: The comparison of the results of NATREG project with six strategic goals using four interval group approaches

5 CONCLUSION

Comparison of listed approaches shows that in our numerical example ADEXTREME performs the shortest intervals. In praxis, it is important, that the intervals of the resulting

aggregating group matrices are not “very” long. MIN-MAX approach uses only minimum and maximum value for creating intervals. When a value is outstanding the method could present very long intervals and the results would be questionable and of no practical use. In MEDINT approach degrees of influence differ between the individual judgments; practical examples can show the optimal rates between the degrees of influence. The intervals of GEOSTDINT approach are in our case of similar length as MIN-MAX intervals. Because GEOSTDINT approach uses geometric standard deviation, we assume, that it could be more convenient for the application in the cases where more than five decision makers are involved, but this statement still needs to be proved.

Note: the research was partly performed in the frame of project COOL, EU Iniciative WoodWisdom-Net 2, No 3211-11-000450.

REFERENCES

[1] Chandran, B., Golden, B. and Wasil, E., 2005. Linear programming models for estimating weights in the analytic hierarchy process. Comp. & Operations Research, 32(9), pp. 2235-2254.
[2] Crawford, G. and Williams, C., 1985. A note on the analysis of subjective judgment matrices. Journal of Mathematical Psychology, 29(4), pp. 387-405.
[3] Facchinetti, G., Ricci, R. G. and Muzzioli, S., 1998. Note on ranking fuzzy triangular numbers. International Journal of Intelligent Systems, 13(7), pp. 613-622.
[4] Grošelj, P. and Zadnik Stirn, L., 2011. Interval comparison matrices in group AHP. SOR '11 proceedings/ The 11th International Symposium on Operational Research in Slovenia 23. - 25. September 2011, Ljubljana, Slovenian Society Informatika (SDI), Section for Operational Research (SOR), pp. 143-148.
[5] Grošelj, P., Pezdevšek Malovrh, Š., Zadnik Stirn, L., 2011. Methods based on data envelopment analysis for deriving group priorities in analytic hierarchy process. Central European Journal of Operations Research, 19(3), 267-284.
[6] Grošelj, P. and Zadnik Stirn, L., 2013. Aggregation of individual judgments into group interval judgment in AHP, Submitted to Fuzzy sets and systems.
[7] Grošelj, P. and Zadnik Stirn, L., 2013. Estimating weights in group AHP using also interval comparison matrices. Submitted to Multiple Criteria Decision Making'13, Trzaskalik T. (eds.), University of Economics, Katowice.
[8] Hojnik, M., 2011. Vizija območja - Pohorje 2030. Projekt: NATREG. http://www.natreg.eu/pohorje/uploads/datoteke/1120VIZIJA2030OBMOCJA-POHORJE2030.pdf.
[9] Liu, F., 2009. Acceptable consistency analysis of interval reciprocal comparison matrices. Fuzzy Sets and Systems, 160(18), pp. 2686-2700.
[10] NATREG, 2011. NATREG - Managing Natural Assets and Protected Areas as Sustainable Regional Development Opportunities. In Danev, G. (ed). Zavod RS za varstvo narave, Ljubljana.
[11] Saaty, T. L., 1980. The Analytic Hierarchy Process. McGraw-Hill, New York.
[12] Saaty, T. L. and Peniwati, K., 2008. Group decision making: Drawing out and reconciling differences. RWS Publications, Pittsburgh, PA.
[13] Wang, Y.-M., Yang, J.-B. and Xu, D.-L., 2005. A two-stage logarithmic goal programming method for generating weights from interval comparison matrices. Fuzzy Sets and Systems, 152(3), pp. 475-498.
[14] Xu, Z. and Chen, J., 2008. Some models for deriving the priority weights from interval fuzzy preference relations. European Journal of Operational Research, 184(1), pp. 266-280.
[15] Xu, Z. S. and Da, Q. L., 2002. The uncertain OWA operator. International Journal of Intelligent Systems, 17(6), pp. 569-575.

CAPACITY PLANNING USING INTERACTIVE STOCHASTIC DYNAMIC PROGRAMMING

Maciej Nowak and **Tadeusz Trzaskalik**
University of Economics in Katowice, Faculty of Informatics and Communication
1 Maja 50, 40-287 Katowice, Poland
{maciej.nowak,tadeusz.trzaskalik}@ue.katowice.pl

Abstract: In the paper capacity planning problem is considered. A dynamic model of the problem is presented and the procedure combining Monte Carlo simulation, dynamic programming and interactive approach is proposed. The method makes it possible to take into account various risk factors that should be taken into account when capacity strategy is formulated. The results obtained during the whole process and also in each period can be analyzed. A numerical example is presented to show the applicability of our procedure.

Keywords: capacity planning, multiobjective stochastic dynamic programming, interactive approach, stochastic dominance.

1 INTRODUCTION

Capacity planning is fundamental for any organization. Insufficient production capacity means that the company is unable to meet the demand and loses potential revenues. On the other hand, excess capacity adds cost and results in lower productivity. Thus, determining facility size, with an objective of achieving high levels of utilization and a high return on investment, is crucial.

Capacity planning can be analyzed in various time horizons: long-range (greater than 1 year), intermediate-rage (3 to 18 month), and short-range (usually up to 3 months) [4]. In this paper strategic capacity decisions are considered. We will try to answer the question how to support the decision maker in making decisions affecting the long-term production capacity.

Forecast of demand is the starting point to any capacity decisions. In the real world, however, predictions, even professionally prepared, are always uncertain. Moreover, organizations usually try to define long-term strategies, as capacity decisions cannot be implemented quickly. Such strategies should define the sequence of actions in subsequent periods. As a result, capacity planning can be formulated as a dynamic decision making problem under risk.

A prerequisite for good decision-making is to define the objective. The overall goal for any organization is to improve it's productivity. Capacity decisions should contribute to it. However, it's not easy, or even possible, to construct a single criterion expressing how much a particular solution adds to the productivity challenge. As a result, when making capacity decisions, managers take into account multiple criteria, including market share, debt ratio, NPV, etc.

Various multiple criteria methods are proposed for capacity planning. In [3] a review of techniques dedicated for semiconductor manufacturing industry is presented. Most of these techniques can be adopted for other sectors. Multicriteria models for capacity planning are also proposed in [1, 2].

In this paper capacity planning is formulated as a multiobjective stochastic dynamic decision-making problem and an interactive procedure for solving it is proposed. The method combines Monte Carlo simulation, dynamic programming and interactive procedure.

2 CAPACITY PLANNING PROBLEM

Capacity is defined as a maximum level of value-added activity over a period of time that the operation can achieve under normal conditions [7]. For top management capacity decisions are of primarily importance, as they determine whether the organization will be able to meet the demand, and how effectively will it use it's resources. Such decisions cannot be made as isolated expenditures, but must be a part of a coordinated plan that will place the firm in an advantageous position. This means, that investments increasing organization's capacity should contribute in winning new customers and improving process flexibility, speed of delivery, quality, and so on.

Two characteristics of capacity: lead-time and economics of scale, must be taken into account when planning changes in capacity. As increasing capacity takes time, the decisions need to be made before demand levels can be estimated precisely. On the other hand, there is pressure to make a change in capacity big enough to exploit economies of scale. Thus, two questions must be answered: when to make a change and how large capacity increments should be.

Three generic strategies for timing capacity change can be considered. According to the first, capacity should lead demand. This means that there is always sufficient capacity to meet forecast demand. The second strategy assumes that capacity lags demand – the capacity is increased only if it can be fully utilized. In such case overtime or subcontracting can be used to accommodate excess demand. The last strategy is a mixture of these two: sometimes there is excess capacity and sometimes a shortage. Inventories are accumulated when the capacity exceeds demand, and used when demand is higher.

In addition to the decision on timing, the magnitude of capacity change must be determined. Larger increments provide economies of scale. However, there are also disadvantages, as organization will have substantial amounts of over-capacity for much of the period when demand is increasing, which results in higher unit costs. Thus, to make a good decision, a detailed analysis of investment costs, as well as production costs is needed.

In order to present a dynamic model of a capacity planning problem, let us assume that we consider a process, which consists of T periods (years). For $t \in \overline{1,T}$ we define:

$\mathbf{Y}_t$ – the set of all feasible states at the beginning of the period t,
$\mathbf{Y}_{T+1}$ – the set of all feasible states at the end of the process,
$\mathbf{X}_t(y_t)$ – the set of all feasible decisions for the period t and the state y_t,
$\mathbf{D}_t(y_t)$ – the set of all period realizations in the period t, defined as follows:

$$\mathbf{D}_t(y_t) = \{d_t(y_t, x_t) : y_t \in \mathbf{Y}_t, x_t \in \mathbf{X}_t(y_t)\} \tag{1}$$

$\Omega_t : \mathbf{D}_t \to \mathbf{Y}_{t+1}$ is a given transformation.

By $\mathbf{D}$ we denote the set of all process realizations, defined as follows:

$$\mathbf{D} = \{d = (d_1, \ldots, d_T) : \forall_{t \in \overline{1,T}}\ y_{t+1} = \Omega_t(y_t, x_t)\} \tag{2}$$

Let $d(y_t)$ be a partial realization for a given realization d, which begins at y_t. We have:

$$d(y_t) = (y_t, x_t, \ldots, y_T, x_T) \tag{3}$$

In our problem y_t is the level of capacity in period t, and x_t – the increment in capacity made in period t, which results in higher capacity in period $t + 1$. Thus for $t \in \overline{1,T}$ the transformation function is defined as follows:

$$y_{t+1} = \Omega_t(y_{t,} x_t) = y_t + x_t \quad \text{for } t \in \overline{1,T} \tag{4}$$

Let K be the number of criteria used to evaluate capacity strategies. Here we assume, that the results obtained when x_t volumes are added to the existing capacity y_t are uncertain. Thus, the evaluation of each period realization with respect to each criterion is represented by a random variable. Therefore, the evaluation of a partial realization $d(y_t) = (y_t, x_t, \ldots, y_T, x_T)$ is a random variable $\xi^{(k)}(d(y_t))$, which is a mixture of random variables representing evaluations of period realizations in periods $t, \ldots, T$. In our approach we assume that Monte Carlo simulation is used for generating distributions of random variables representing evaluations of alternatives with respect to criteria.

3 METHODOLOGY

The procedure we propose for solving the problem consists of two main steps. First, dynamic programming approach is used for identifying non-dominated solutions. Next, the problem is solved using interactive procedure INSDECM. Because of the lack of space we omit the general, formal description of the procedure, which can be found in [6].

3.1 Stochastic dominance rules

As the evaluations of process realizations are random, so a question arises how to compare the results obtained under various realizations. In our procedure we use stochastic dominance rules. Let us consider two partial period realizations $d_i(y_t)$ and $d_j(y_t)$, both beginning at y_t. $G_i^{(k)}(z)$ and $G_j^{(k)}(z)$ denote cumulative distribution functions representing the evaluations of $d_i(y_t)$ and $d_j(y_t)$ with respect to k-th criterion. We use FSD (*First Stochastic Dominance*) and SSD (*Second Stochastic Dominance*) for comparing probability distributions. The definitions are as follows:

$$G_i^{(k)}(z)\,\mathrm{FSD}\,G_j^{(k)}(z) \quad \Leftrightarrow \quad G_i^{(k)}(z) \neq G_j^{(k)}(z) \text{ and } H_1(z) = G_i^{(k)}(z) - G_j^{(k)}(z) \leq 0 \text{ for all } z \in \mathbf{R}$$

$$G_i^{(k)}(z)\,\mathrm{SSD}\,G_j^{(k)}(z) \quad \Leftrightarrow \quad G_i^{(k)}(z) \neq G_j^{(k)}(z) \text{ and } H_2(x) = \int_a^z H_1(q)dq \leq 0 \text{ for all } z \in \mathbf{R}$$

3.2 Identifying the non-dominated process realizations

We will say that partial process realization $d_i(y_t)$ dominates partial process realization $d_j(y_t)$ if the following condition is fulfilled:

$$\forall_{k \in \overline{1,K}} \quad G_i^{(k)}(z)\,\mathrm{FSD}\,G_j^{(k)}(z) \quad \vee \quad G_i^{(k)}(z)\,\mathrm{SSD}\,G_j^{(k)}(z) \tag{5}$$

Thus, we will assume that $d_i(y_t)$ dominates $d_j(y_t)$, if stochastic dominance relation can be identified for each criterion. In order to identify the set of non-dominated process realizations, we will use Bellman's principle of optimality.

Taking into account theorems presented in [8, 9, 10], we can use the following procedure for identifying non-dominated process realizations:

1. Start from the last period: $t = T$; for each feasible state y_T identify non-dominated realizations $d_T(y_T, x_T)$.
2. Go to the previous period: $t = t - 1$.
3. For each feasible state y_t, identify the set of non-dominated partial realizations, which begin at y_t.

4. If $t > 1$ – go to 2, otherwise: stop the procedure.

3.3 Dynamic INSDECM procedure

INSDECM [5] is devoted for problems with a finite number of feasible solutions and evaluations represented by random variables with known distributions. The detailed description of dynamic version of INSDECM is provided in [6]. Here we present just general description.

Each iteration consists of three phases: (1) presentation of the results to the decision maker, (2) collection of the preference information, (3) identification of the solutions satisfying the requirements specified by the decision maker in the second phase. The results are presented to the decision maker in a potency matrix. It consists of two rows grouping the best (optimistic) and the worst (pessimistic) values of distribution parameters chosen by the decision maker (expected value, median, quantilles, standard deviation, etc.). The decision maker is asked whether the pessimistic values of parameters are good enough. If the answer is *yes*, he or she is asked to select the final solution from the set of alternatives currently considered. Otherwise, the decision maker is asked to formulate a constraint that a satisfactiry alternative should satisfy. In the third phase, the set of alternatives satisfying this constraint are identified and next iteration starts. The process is continued until all pessimistic are accepted by the decision maker.

Due to the complexity of the problem (multiple criteria, multi-period process, random outcomes), dynamic version of INSDECM assumes, that the decision maker is able to define a hierarchy of the criteria. The results obtained under different process realizations are analyzed according to this hierarchy.

4 NUMERICAL EXAMPLE

In order to illustrate applicability of the procedure let us consider a company working on the capacity planning problem. The planning horizon is five years. The current capacity is 100 batches per week. Taking into account demand forecasts, the company concluded that at the end of the fifth year it's capacity should reach 150 batches per week. Due to technical reasons the capacity can be increased either by 25, or 50 units at one time. Thus, the company can either change capacity once by 50 batches, or twice by 25 units each time. The graph of the process is presented in Figure 1. The nodes in the graph represent states of the process at the beginning of each period and at the end of the process. Nodes 1, 4, 7, 10 and 13 represent states with the capacity is equal to 100 batches. Nodes 3, 6, 9, and 12 represent states with the capacity equal to 125 batches. Finally nodes 2, 5, 8, 11, and 14 represent states with capacity equal to 150 batches.

In state 1 representing the situation at the beginning of the first year, three decisions can be made: the company can increase the capacity in the first year by 50 units, increase it by 25 units, or resign of changing the capacity in the first year. The successive decisions lead to states 2, 3 and 4 respectively. If the capacity is changed by 50 units in the first year, it reaches the final capacity of 150 batches in the second year, and does not changes till the end of the process. If it is increased by 25 units, it must change the capacity once more in any of successive years. Finally if it does not change the capacity in the first year, it must change it later either once by 50 units, or in two phases.

We constructed a simulation model to analyze the results, that could be obtained for each period realization. In our model we considered the following risk factors: demand, product market price, investment cost, production cost (fix and variable). Expert opinions were used to assess the probability distribution for each factor. The main assumption of the

model is that the demand should be fully satisfied. If the current capacity is not enough to meet it, overtime and subcontracting is used. However, they are employed only if the variable cost is not higher than the market price.

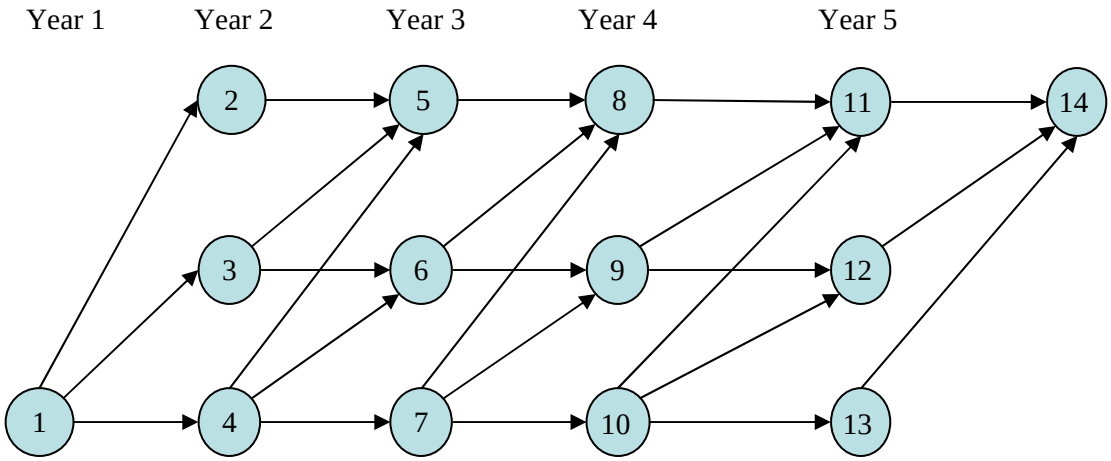


Figure 1: Graph of the process.

Three criteria are considered: profit margin (criterion f_1), customer service level (criterion f_2), and capacity utilization (criterion f_3). In our example the decision made in any state does not affect the results generated in this period, but only determines in what state the process will be in the next period. Table 1 summarizes the results of Monte Carlo simulation.

Table 1: Expected values of criteria functions for each state.

State	Profit margin	Customer service level	Capacity utilization	State	Profit margin	Customer service level	Capacity utilization
1	12997.50	97.73%	100.00%	8	15592.14	100.00%	100.00%
2	8944.37	100.00%	91.67%	9	10661.71	84.25%	100.00%
3	11074.66	97.73%	100.00%	10	13028.83	67.40%	100.00%
4	13019.73	78.18%	100.00%	11	16011.61	97.73%	100.00%
5	12197.04	100.00%	99.00%	12	10418.36	81.44%	100.00%
6	10880.03	90.49%	100.00%	13	13014.12	65.15%	100.00%
7	13030.42	72.39%	100.00%				

In the second step non-dominated process realizations are identified. The following process realizations are non-dominated:

1 – 2 – 5 – 8 – 11 – 14, 1 – 3 – 5 – 8 – 11 – 14, 1 – 3 – 6 – 8 – 11 – 14,
1 – 4 – 5 – 8 – 11 – 14, 1 – 4 – 6 – 8 – 11 – 14, 1 – 4 – 7 – 8 – 11 – 14.

Thus, in the last phase only solutions assuming that the capacity is changed in the first, second of third period are analyzed. The final solution is identified using INSDECM procedure. First, the decision maker is asked to define hierarchy of criteria. According to him the most important is criterion f_1, next are f_2 and f_3.

In the first iteration the dialog with the decision is conducted according to the following scenario:

1. The first criterion is considered. The decision maker specifies the data that he would like to analyze: probability, that the profit margin for the whole process will not be the less than 65000, and the probability that in any period the profit margin will not be less than 10000.

2. The potency matrix is presented to the decision maker (tab. 2)

Table 2: The potency matrix presented to the decision maker in iteration 1.

Distribution characteristic	Probability that the profit for the whole process will be not be less than 65000	Probability that for none period the profit will be less than 10000
Optimistic value	0.93	0.55
Pessimistic value	0.65	0.25

3. The decision maker specifies additional requirement: the probability that for none period the profit will be less than 10000 should be not less than 0.50.
4. The set of process realizations satisfying the constraint defined by the decision maker is identified and the procedure goes to the next iteration.

In next iterations criteria f_2 and f_3 are analyzed. At the end the set of process realizations satisfying decision maker's requirements is identified.

5 CONCLUSIONS

In this paper a procedure for capacity planning was proposed. Our method combines Monte Carlo simulation, dynamic programming and interactive approach. It assumes that the criteria are analyzed according to their importance. However, it is also possible to analyze the process period by period. The procedure can also be applied for other dynamic decision making problems under risk, such like project planning, project portfolio selection, or production planning.

References

[1] Chen, Y.-Y., Chen, T.-L., Liou, Ch.-D., 2013. Medium-term multi-plant capacity planning problems considering auxiliary tools for the semiconductor foundry. The International Journal of Advanced Manufacturing Technology, vol. 64, pp. 1213-1230.

[2] Cheng, L., Subrahmanian, E., Westerberg, A.W., 2004. Multi-objective Decisions on Capacity Planning and Production−Inventory Control under Uncertainty. Industrial & Engineering Chemistry Research, vol. 43, pp. 2192-2208.

[3] Geng, N., Jiang, Z., 2009. A review on strategic capacity planning for the semiconductor manufacturing industry. International Journal of Production Research, vol. 47, pp. 3639-3655.

[4] Heizer, J., Render, B., 2004. Operations Management, 7th ed., Pearson Education, Upper Saddle River.

[5] Nowak, M., 2006. INSDECM – An interactive procedure for discrete stochastic multicriteria decision making problems. European Journal of Operational Research, vol. 175, pp. 1413-1430.

[6] Nowak, M., Trzaskalik, T., 2012. Interactive procedure for a multiobjective stochastic discrete dynamic problem, Journal of Global Optimization, DOI: 10.1007/s10898-012-0019-9.

[7] Slack, N., Lewis, M., 2011. Operations Strategy, 3rd ed., Pearson Education, Harlow.

[8] Trzaskalik, T., 1992. Hierarchical Approach to Multi-Criteria Dynamic Programming. Information Systems and Operational Research INFOR, vol.20, pp. 132-142.

[9] Trzaskalik T, Sitarz S., 2002. Dynamic discrete programming with partially ordered criteria set. In: Trzaskalik, T., Michnik J. (eds.), Multiple Objective Programming and Goal Programming. Recent Developments, Phisica-Verlag, Heidelberg, pp. 186-195.

[10] Trzaskalik, T., Sitarz, S., 2007. Discrete Dynamic Programming with Outcomes in Random Variable Structures, European Journal of Operational Research, vol. 177, pp. 1535-1548.

UNIFIED PROCEDURE FOR BIPOLAR METHOD

Tadeusz Trzaskalik[1], Sebastian Sitarz[2], Cezary Dominiak[1]
[1]University of Economics in Katowice, Department of Operations Research,
ul. 1 Maja 50, 40-287 Katowice, Poland
(tadeusz.trzaskalik, cezary.dominiak)@ue.katowice.pl
[2] Institute of Mathematics, University of Silesia in Katowice,
ul. Bankowa 14, 40-007 Katowice, Poland
ssitarz@ux2.math.us.edu.pl

Abstract
Bipolar is one of the Multiple Criteria Decision Analysis (MCDA) methods, based on the concept of bipolar reference objectives, proposed by Ewa Konarzewska-Gubała. In Bipolar method decision alternatives are not compared directly to each other, but they are confronted to the two sets of reference objects: desirable and non-acceptable. Practical application of the method showed some its shortcomings. It may happen that a decision alternative can be evaluated as better than a desirable reference object and simultaneously as worse than a non-acceptable object. Also a case where reference sets are numerous needs some modifications. The aim of the paper is to formulate unified Bipolar procedure which contains classical Bipolar method as well as some modifications which help to overcome difficulties mentioned above.

Keywords: MCDA, BIPOLAR, reference sets, unified procedure

1 INTRODUCTION

One of the MCDA methods[3, 19] is Bipolar, proposed by Konarzewska-Gubała [7, 8, 9]. The essence of the analysis in Bipolar method consists in a fact that the decision alternatives are not compared directly to each other, but they are confronted to the two sets of reference objects: desirable (called "good") and non-acceptable (called "bad"). These two separate sets constitute bipolar reference system. It is assumed, that the decision maker applying Bipolar method in practice, on the base of her or his experience, gathered opinions and undertaken studies is able to create such a system. Many aspects of Bipolar approach have been described by the author of the method [9, 10]). Some improvements have been proposed (Dominiak [1, 2], Trzaskalik and Sitarz [17, 18]). The method has been used in applications (Jakubowicz [5], Jakubowicz and Konarzewska-Gubała [6], Dominiak [1,2], Konarzewska-Gubała [11]). Moreover, the method has also been applied to model multi-stage multi-criteria decision processes (Trzaskalik [16]). The Bipolar method belongs to a group of methods that involve reference objects while compare alternatives. Bipolar method belongs to a group of methods that involve reference objects while compare alternatives. Other examples of this approach include for example the Michałowski and Szapiro's bi-reference method [13] and the method developed by Skulimowski [15]. More recently learning methods, for instance DRSA, based on rough set methodology to derive classification rules has been developed (Greco, Matarazzo, Słowiński [4]).

When performing the procedure, some alternatives can be evaluated as better than "good" objects from the reference system. Such alternatives are named "overgood". Other alternatives can be evaluated as worse than "bad" objects from the reference system. Such alternatives are named "underbad". It may happen that some decision alternatives are overgood and underbad simultaneously. Trzaskalik and Sitarz [17, 18] described how to deal with underbad and overgood alternatives. Another problem arose when the set of reference objects was numerous. Some proposition were given by Dominiak [1,2].

The aim of the present paper is to describe unified Bipolar procedure, which contains classical Bipolar method as well as its modifications, described in previously published works.

The paper consists of six chapters. In Chapter 2 classical Bipolar procedure has been briefly described. In Chapter 3 modifications of reference sets and categories are presented. In Chapter 4 methods of supporting a decision maker when determining criterion weights and veto threshold values in the case the reference system is numerous are mentioned. In Chapter 5 a unified procedure of BIPOLAR method taking into account all the modifications is presented. The concluding remarks in Chapter 6 end the paper.

2 CLASSICAL BIPOLAR METHOD

It is assumed, that there are given: the set of decision alternatives $\mathcal{A} = \{\mathbf{a}^1, \mathbf{a}^2, ..., \mathbf{a}^m\}$ and the set of criteria functions $\mathcal{F} = \{f_1,\ldots,f_n\}$, where $f_k: \mathcal{A} \rightarrow \mathcal{K}_k$ for $k=1, \ldots, n$, and $\mathcal{K}_k$ is a cardinal, ordinal or binary scale. Criteria are defined in such a way that higher values are preferred to lower values. Description of remaining types of criteria is given by Konarzewska-Gubała [9]. For each criterion the decision maker establishes weight w_k of relative importance (it is assumed, that $\sum_{k=1}^{n} w_k = 1$ and $w_k \geq 0$ for each $k=1, \ldots,n$), equivalence threshold q_k and veto threshold v_k. The decision maker also establishes minimal criteria values concordance level s as the outranking threshold. It is assumed, that condition $0.5 \leq s \leq 1$ holds.

The decision maker establishes a bipolar reference system $\mathcal{R} = \mathcal{D} \cup \mathcal{Z}$, which consists of the set of „good" objects $\mathcal{D} = \{\mathbf{d}^1,\ldots, \mathbf{d}^d\}$ and the set of "bad" objects $\mathcal{Z} = \{\mathbf{z}^1,\ldots,\mathbf{z}^z\}$, where d and z denote the number of "good" and "bad" objects, respectively. It is assumed, that $\mathcal{D} \cap \mathcal{Z} = \varnothing$. The number of elements of the set $\mathcal{R}$ is equal to $d+z$. Elements of the set $\mathcal{R}$ are denoted as $\mathbf{r}^h$, $h=1,\ldots,r$. Values $f_k(\mathbf{r}^h)$ for k=1,...,n and $h=1,...,r$ are known. We assume, that holds condition

$$\forall_{k=1,\ldots,n} \forall_{\mathbf{d}\in\mathcal{D}} \forall_{\mathbf{z}\in\mathcal{Z}} \quad f_k(\mathbf{d}) \geq f_k(\mathbf{z}) \tag{1}$$

The Bipolar method consists of three phases. In the first phase decision alternatives are compared to reference objects and as a result outranking indicators and preference structure in the reference system are established. In the second phase position of each decision alterative with regard to bipolar reference system is established. In the third phase, on the basis of two mono-sortings of alternatives into specified categories and two partial preorders (mono-orders) introduced independently into the set of alternatives, the intersection of these two preorders, creating the bipolar partial preorder is obtained. In the phase I we can recognize the ideas of concordance and veto thresholds, introduced in Roy's Electre methodology [14], in the phase II - the idea of Merighi [12] algorithms of confrontation. The detailed description of the method (helpful in unified procedure description) can be found in [17].

3 MODIFICATIONS OF REFERENCE SETS AND CATEGORIES

M1. Modification of the reference set of "good" objects **[**18].

Let $\hat{\mathbf{f}}^{\mathcal{Z}}$ denote the ideal vector in the reference set of „bad" objects, hence

$$\hat{f}_k^{\mathcal{Z}} = \max \{ f_k(\mathbf{z}): \mathbf{z}\in\mathcal{Z}\}$$

We replace the set $\mathcal{D}$ by the set $\overline{\mathcal{D}} = \{ \overline{\mathbf{d}}^1, ..., \overline{\mathbf{d}}^d \}$, changing these evaluations, which are too low according to ideal solutions in the set $\mathcal{Z}$, that is

$$f_k\left(\overline{\mathbf{d}}\right) = \begin{cases} f_k\left(\mathbf{d}\right), & \text{if } f_k\left(\mathbf{d}\right) \geq \hat{f}_k^{\mathcal{Z}} \\ \hat{f}_k^{\mathcal{Z}}, & \text{if } f_k\left(\mathbf{d}\right) < \hat{f}_k^{\mathcal{Z}} \end{cases}$$

M2. Modification of the reference set of „bad" objects [18]

Let $\check{\mathbf{f}}^{\mathcal{D}}$ denote nadir vector in the reference set of „good" objects, hence

$$\check{f}_k^{\mathcal{D}} = \min \{ f_k(\mathbf{d}): \mathbf{d} \in D \}$$

We replace the set $\mathcal{Z}$ by the set $\overline{\mathcal{Z}} = \{ \overline{\mathbf{z}}^1, ..., \overline{\mathbf{z}}^z \}$, changing these evaluations, which are too high according to nadir solution in the set $\mathcal{D}$, that is

$$f_k\left(\overline{\mathbf{z}}\right) = \begin{cases} f_k\left(\mathbf{z}\right), & \text{if } f_k\left(\mathbf{z}\right) \leq \check{f}_k^{\mathcal{D}} \\ \check{f}_k^{\mathcal{D}}, & \text{if } f_k\left(\mathbf{z}\right) > \check{f}_k^{\mathcal{D}} \end{cases}$$

M3. Modification of categories [18]

In the source Bipolar method three categories of alternatives: B1, B2 and B3 are defined. Now an additional category B2' including all overgood and underbad alternatives is created.

4 MODIFICATIONS FOR NUMEROUS REFERENCE SETS

M4. Local preference function [1,2]

All the criteria are given on cardinal scales. It is assumed that the weight of considered criterion depends on the values of the "bad" reference objects. Functions of local preference describe that kind of dependence.

M5. Modification of position definition for an alternative in relation to the reference system [1, 2]

That modification refers to position the description for alternatives in relation to the reference system. We assume that the considered alternative outranks a reference set, if the number of objects outranked by that variant is greater than the number of objects from that set which outranked the considered alternative. Otherwise we assume that the reference set outranks the considered decision variant. As a measure of outranking we consider the ratio of the difference between these values to the number of elements of the reference set.

M6. Modification of criteria weights [1, 2]

Criteria weights are establish applying deciles distributions.

M7. Modification of veto thresholds [1, 2]

Veto thresholds are established applying deciles distributions.

5 UNIFIED PROCEDURE

To give a possibility for a decision maker to apply the modifications described above as well as the classical Bipolar approach (denoted as **C**) we propose a procedure, elaborated below. The block-scheme of the procedure is given in Figure 1.

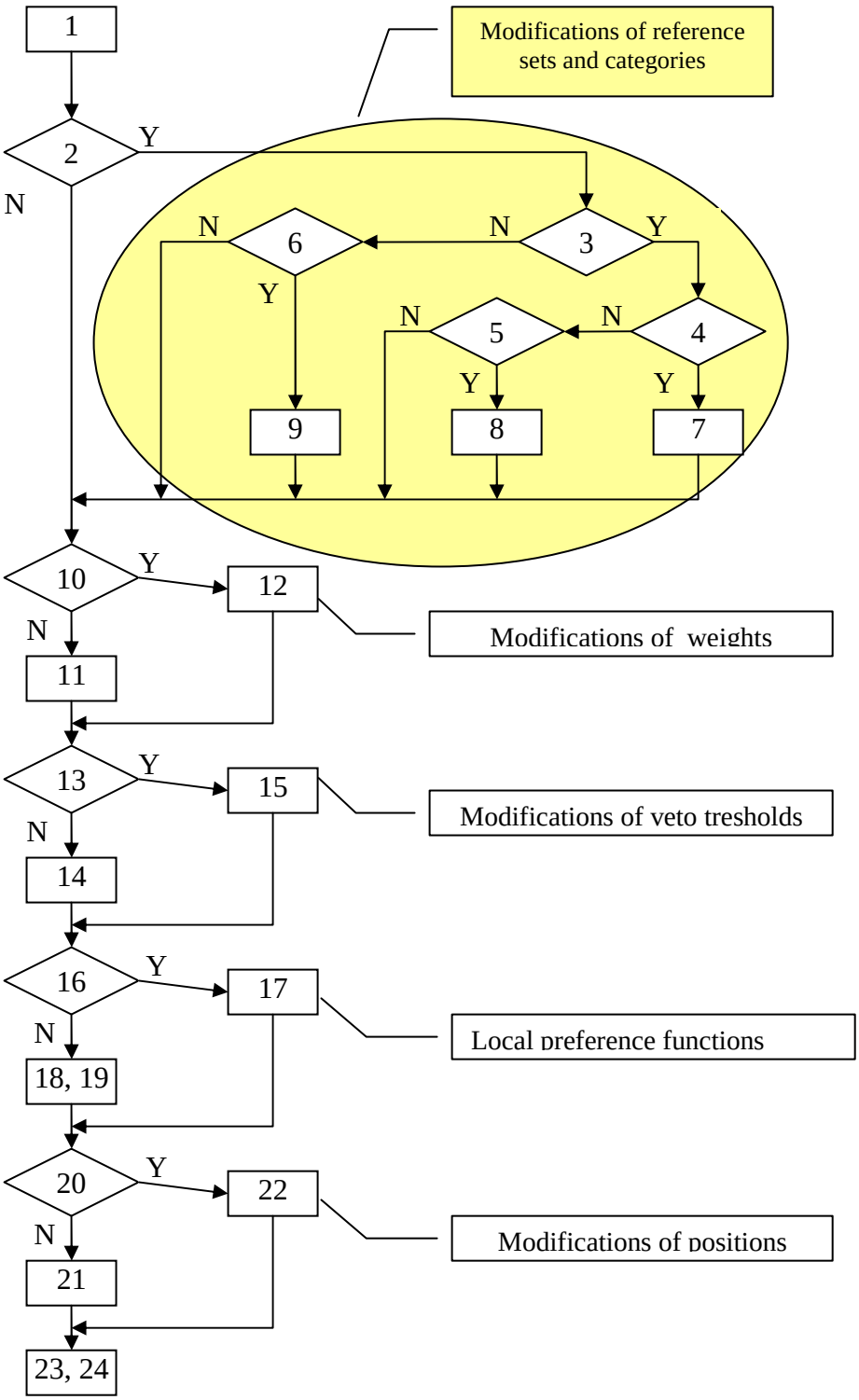


Figure 1: Block scheme of BIPOLAR unified procedure

The consecutive steps of the procedure can be described as follows:

Start

1. Establish sets $\mathcal{A}$, $\mathcal{D}$, $\mathcal{Z}$.
2. Is condition (1) fulfilled?
   Yes – go to 10. No – go to 3.
3. Do you want to modify reference sets?
   Yes – go to 4. No – go to 6.
4. Do you want to modify the set $\mathcal{D}$?
   Yes – go to 7. No – go to 5.
5. Do you want to modify set $\mathcal{Z}$?
   Yes - go to 8. No – go to 10.
6. Do you want to extend the set of Bipolar categories?
   Yes - go to 9. No - go to 10.
7. Modify the set $\mathcal{D}$ according to **M1** .
   Go to 10.
8. Modify the set $\mathcal{Z}$ according to **M2**.
   Go to 10.
9. Extend the set of categories in Bipolar ranking according to **M3**.
   Go to 10.
10. Do you want to apply the possibility of decision support for establishing weights?
    Yes – go to 12. No – go to 11.
11. Establish weights, k=1,...,n and concordance level s.
    Go to 13.
12. Establish weights according to **M6** and concordance level s.
    Go to 13.
13. Do you want to apply possibility of decision support for establishing veto thresholds?
    Yes – go to 15. No – go to 14.
14. Establish veto thresholds.
    Go to 16.
15. Establish veto thresholds according to **M7**.
    Go to 16.
16. Do you want to apply local preference functions to determine outranking coefficients?
    Yes – go to 17. No – go to 18.
17. Determine outranking coefficients according to **M4**.
    Go to 20.
18. Determine outranking coefficients according to **C**.
    Go to 20.
19. Determine preference structure according to **C.**
    Go to 20.
20. Do you want to apply modification of position definition for $\mathbf{a}^i$ in relation to $\mathcal{R}$?
    Yes – go to 22. No – go to 21.
21. Determine the position $\mathbf{a}^i$ in relation to $\mathcal{R}$ according to **C**.
    Go to 23.
22. Determine the position $\mathbf{a}^i$ in relation to $\mathcal{R}$ according to **M5**.
    Go to 23.
23. Perform mono-sortings and mono-rankings according to **C**.
    Go to 24.
24. Perform Bipolar-sorting and Bipolar-ranking according to **C**.
    Go to Stop.

Stop.

6 CONCLUDING REMARKS

Modifications of the source version of the Bipolar method allow both for rationalizing Bipolar incomparability of some alternatives and elaborating of a ranking. The unified procedure, described in the paper allows to incorporate all the modifications of the source version.

References

[1] Dominiak, C., 2006. Application of modified Bipolar method. In: T.Trzaskalik (ed.) Multicriteria Methods on Polish Financial Market, p.105-113, PWE (in Polish).

[2] Dominiak, C.,1997. Portfolio Selection Using the Idea of Reference Solution. In: G.Fandel, T.Gal (eds.) Multiple Criteria Decision Making. Springer-Verlag Berlin Heidelberg New York, p.593-602.

[3] Figueira, J., Greco, S., Ehrgott, M.(eds.), 2005. Multiple criteria decision analysis: States of the art surveys. Springer.

[4] Greco, S., Matarazzo, B., Słowinski, R., 2002. Rough set methodology for sorting problems in presence of multiple attributes and criteria. EJOR vol.138, p.247-259.

[5] Jakubowicz, S.,1987. Work Characteristics of a „Good" Physics Teacher on the Basis of His Lessons".RPBP.III.30.VI.4.6. The University of Wrocław (copied manuscript, in Polish).

[6] Jakubowicz, S., Konarzewska-Gubała, E., 1989. Work Characteristics of a Physics Teacher. University of Wrocław (copied manuscript, in Polish).

[7] Konarzewska-Gubała E., 1987. Multicriteria Decision Analysis with Bipolar Reference System: Theoretical Model and Computer Implementation". Archiwum Automatyk i Telemechaniki vol. 32, no 4, p.289-300.

[8] Konarzewska-Gubała E., 1989. BIPOLAR: Multiple Criteria Decision Aid Using Bipolar Reference System, LAMSADE, Cahier et Documents no 56, Paris.

[9] Konarzewska-Gubała, E., 1991. Multiple Criteria Decision Aid: System Bipolar. Scientific Works of the University of Economics in Wrocław, no 551 (in Polish).

[10] Konarzewska-Gubała, E., 1996. Supporting an effective performance appraisal system. Argumenta Oeconomica vol.1, p.123-125.

[11] Konarzewska-Gubała, E., 2002. Multiple Criteria Company Benchmarking Using the BIPOLAR Method. In T.Trzaskalik, J.Michnik (eds.) Multiple Objective and Goal Programming. Recent Developments. Physica-Verlag. Springer-Verlag Company, Heidelberg, New York, p.338-350.

[12] Merighi, D., 1980. Un modello di valutazione rispetto insiemi di riferimento assegnati. Ricerca Operativa no 13, p.31-52.

[13] Michałowski, W., Szapiro, T., 1992. A Bi-reference Procedure for Interactive Multiple Criteria Programming. Operations Research vol. 40, no 2, p. 247-258.

[14] Roy, B., 1985. Methodologie Multicritere d'Aide a la Decision. Economica, Paris.

[15] Skulimowski, A , 1996. Decision Support Systems Based on Reference Sets. AGH, Kraków.

[16] Trzaskalik T., 1987. Model of multistage multicriteria decision processes applying reference sets". In: Decision Models with Incomplete Information, Scientific Works of the University of Economics in Wrocław, no 413, p.73-93 (in Polish).

[17] Trzaskalik, T., Sitarz, S., 2007. Underbad and overgood alternatives in BIPOLAR method. Proceedings of the 9th International Symposium on Operational research SOR'07, Nova Gorica, September 26-28, 2007. p.159 – 165.

[18] Trzaskalik, T, Sitarz, S., 2012. How to Deal with Overgood and Underbad Alternatives in Bipolar Method. Proceedings of the 4th International Conference on Intelligent Decision Technologies (IDT´2012) , Intelligent Decision Technologies Smart Innovation, Systems and Technologies, Volume 16. p 345-354

[19] Tzeng G-H., Huang J-H. (2011) Multiple attribute decision making. Methods and applications. Taylor&Francis.

PERFORMANCE OF MACHINE LEARNING METHODS IN CLASSIFICATION MODELS WITH HIGH-DIMENSIONAL DATA

Marijana Zekić-Sušac, Sanja Pfeifer and **Nataša Šarlija**
University of Josip Juraj Strossmayer in Osijek, Faculty of Economics
Gajev trg 7, 31000 Osijek, Croatia
{marijana, pfeifer, natasa}@efos.hr

Abstract: The paper investigates the performance of four machine learning methods: artificial neural networks, classification trees, support vector machines, and k-nearest neighbour in classification type of problem by using a real dataset on entrepreneurial intentions of students. The aim is to find out which of the machine learning methods is more efficient in modelling high-dimensional data in the sense of the average classification rate obtained in a 10-fold cross-validation procedure. In addition, sensitivity and specificity is also observed. The results show that the accuracy of artificial neural networks is significantly higher than the accuracy of k-nearest neighbour, but the difference among other methods is not statistically significant.

Keywords: machine learning, support vector machines, artificial neural networks, CART classification trees, k-nearest neighbour, large-dimensional data, cross-validation

1 INTRODUCTION

Most research on dealing with high-dimensional data was focused on variable reduction methods in the pre-processing and in the post-processing stage of modelling. In some cases, pre-processing variable reduction methods based on t-test, Cronbach's alpha, chi-square, PCA or others do not give efficient results because while providing less information they yield lower accuracy of the model. Our previous research [16] shows that such situation exists in a real dataset collected in an international survey on entrepreneurship intentions, self-efficacy and identity. Based on proven instruments which measure certain attributes of students, a large number of input variables is used to provide a basis for finding an efficient model that will be able to classify students according to their entrepreneurial intentions. In previous investigations [16] it was found that non-linear machine learning methods such as ANNs could be efficient in the area of modeling entrepreneurial intentions of students. The purpose of this paper is to find out if other machine learning methods, such as support vector machines (SVMs), decision trees, and k-nearest neighbour (KNN) can outperform ANNs in classification type of problems with a large number of variables.

2 PREVIOUS RESEARCH

Research on entrepreneurial career choices of students mostly proposes a huge number of personal inputs that can interact on a variety of levels and directions. It has been presumed that students attitudes, values and career choices can be sufficiently well represented by the following groups of variables [5],[13]: (1) entrepreneurial intentions, (2) altruistic values and empathy, (3) subjective norms, 2006), (4) entrepreneurial self-efficacy, (4) allocentrism /idiocentrism, (5) prior family business exposure, (6) entrepreneurial outcome expectations (7) strength of entrepreneur identity aspiration, and (8) social entrepreneurship self-efficacy. Following such proven instruments for measuring entrepreneurial intentions, the constructed models could consist of hundreds of variables.

Methodology used for modeling entrepreneurial intentions was mostly focused to multiple regression and structural modelling [5]. Machine learning methods have not been investigated in this area, although they were frequently tested in other problem domains. ANNs outperformed discriminant analysis and other statistical methods in various problem

domains including financial prognosis, fraud detection, etc. [9]. SVMs were also compared to ANNs in financial failures, machine fault detection, medicine others [12], [14]. In addition to ANNs and SVMs, decision trees are a method that is frequently used in classification [7], as well as the KNN technique which has been used as an efficient classification technique in multivariate models [6].

3 METHODOLOGY

Artificial neural network (ANN) as a machine learning method has the ability to approximate any nonlinear mathematical function, which is useful especially when the relationship between the variables is not known or is complex [8]. It has been successfully used for both regression and classification type of problems in different areas [9]. Although there is a number of different types of ANNs, the multilayer perceptron (MLP) is the most common one that can use various algorithms to minimize the objective function. The input layer of an ANN consists of n input units with values $x_i \in R$, $i=1,2,...,n$, and randomly determined initial weights w_i usually from the interval [-1,1]. Each unit in the hidden (middle) layer receives the weighted sum of all x_i values as the input. The output of the hidden layer denoted as y_c is computed by summing the inputs multiplied with their weights, according to:

$$y_c = f\left(\sum_{i=1}^{n} w_i x_i\right) \tag{1}$$

where f is the activation function selected by the user (sigmoid, tangent hyperbolic, exponential, linear, step or other) [8]. The difference between the computed output y_c and the actual output y_a, is the local error ε which is computed at each learning iteration. The error ε is then used to adjust the weights of the input vector according to a learning rule, usually the Delta rule. The above process is repeated in a number of iterations (epochs), where the three different algorithm were tested to minimize the error: gradient descent, conjugate gradient descent, and Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm [4]. The number of hidden units varied from 2 to 20, and the training time is determined in an early-stopping procedure which iteratively trains and tests the network on a separate test sample in a number of cycles, and saves the network which produces the lowest error on the test sample.

Support vector machine (SVM) is a classification method based on the maximum margin hyperplane aimed to be used for non-linear mapping of the input vectors into a high-dimensional feature space [15]. It produces a binary classifier, so-called optimal separating hyperplanes, and results in a uniquely global optimum, high generalization performance, and does not suffer from a local optima problem [2]. The basic principle of learning in SVM can be described as follows. Suppose we are given a set of training data $x_i \in R^n$ with the desired output $y_i \in \{+1,-1\}$ corresponding with the two classes, and assume there is a a separating hyperplane with the target function $w \cdot x_i + b = 0$, where w is the weight vector, and b is a bias. We want to choose w and b to maximize the margin or distance between the parallel hyperplanes that are as far apart as possible while still separating the data. The non-negative Lagrange multipliers can be searched by solving the following optimization problem if the problem is nonlinear:

$$\text{Maximize } Q(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K\left(x_i x_j\right) \tag{2}$$

$$\text{subject to } \sum_{i=1}^{n} \alpha_i y_i = 0,\ 0 \le \alpha_i \le C,\ \text{i= i=1,2,...,n.} \tag{3}$$

where C is the nonnegative parameter chosen by users known as capacity. The final classification function is:

$$f(x) = \text{sgn}\left\{ \sum_{i=1}^{n} \alpha_i^* y_i K\left(x_i, x_j\right) + b^* \right\} \tag{4}$$

where K is a kernel function, which can be linear, sigmoid, RBF or polynomial.

SVM is able to select a small and most proper subset of data pairs (support vectors). Since its performance depends mostly on the choice of kernel function and hyper parameters, a cross-validation procedure is used as a successfull tool for adjusting those parameters [2]. Linear, polynomial, RBF, and exponential kernels were used, where gamma coefficient for polynomial and RBF kernel was 0.0625, degree was 3, coefficient varied from 0 to 0.1, C=10.

Decision tree i.e. classification tree is a machine learning method aimed to build a binary tree by splitting the input vectors at each node according to a function of a single input. CART steps were summarized in [10] as: (1) assign all objects to root node, (2) split each input variable at all possible split points, (3) for each split point, split the parent node into two child nodes by separating the objects with values lower and higher than the split point for the considered input variable, (4) select the variable and split point with the highest reduction of impurity, (5) perform the split of the parent node into the two child nodes according to the selected split point, (6) repeat steps 2–5, using each node as a new parent node, until the tree has maximum size, and (7) prune the tree back using cross-validation to select the right-sized tree. The evaluation function used in this research for splitting is the Gini index defined as [1]:

$$Gini(t) = 1 - \sum_i p_i^2 \tag{5}$$

where t is a current node and p_i is the probability of class i in t. The CART algorithm considers all possible splits in order to find the best one by Gini index. Prune of missclassification error was used as the stopping rule, with minimum n=5.

The aim of the KNN techique is to classify the outcome of a in input vector based on a selected number of its nearest neighbours. For a given input vector, the method estimates the outcome by finding k examples that are closest in distance to the input (i.e. its neighbours). For classification problems it uses a majority of voting. In estimating the model it is important to select the appropriate value of k. One way to select the optimal value of k is to use cross-validation (CV) procedure to smooth the k parameter, i.e. to find the value of k that is the optimal trade off [3]. In order to find the neighbours of a point, a distance metrics needs to be used. The most common is the Euclidean, while others possible metrics are Euclidean squared, City-block, and Chebychev distances. In this paper, the Euclidean distance is used according to [3].

The performance of all models on each validation sample is measured by the total classification rate (i.e. the proportion of correctly classified cases in the test set). The 10-fold cross-validation procedure (or leave k cases out, where k=1/10 of the total sample) is used in this paper because it produces no statistical bias of the result since each tested sample is not the member of the training set. Also, the classification rates of class 0 and class 1 were also observed in order to compute the sensitivity and specificity of the models. The sensitivity and specificity ratios were computed according to (Simon and Boring, 1990):

$$sensitivity = \frac{c_1}{(c_1 + d_0)}\ ,\quad specificity = \frac{c_0}{(c_0 + d_1)} \tag{6}$$

where c_0 is the number of students accurately predicted to have output 0, c_1 is the number of students accurately predicted to have output 1, d_0 is the number of false negatives (the number of students falsely predicted to have output 0), and d_1 is the number of false positives (the number of students falsely predicted to have output 1). The type I error (α *=1-specificity)* and type II error (β *=1-sensitivity)* were calculated in order to compare the cost of misclassification produced by each of the models, and to compute the likelihood ratios according to:

$$L_1 = \frac{sensitivity}{\alpha},\ L_0 = \frac{specificity}{\beta} \tag{7}$$

where L_1 is likelihood ratio for class 1, while L_0 is the likelihood ratio for class 0.

4 DATA

The dataset for this research was collected in an international survey on entrepreneurial intentions at the summer semester 2010 and 2012. It consisted of 443 regular students of business administration at the first year of study at University of Osijek, Croatia. The total number of 94 input variables was used based on proven instruments described in section 2. There were 48.76% of respondents with an intention to start a business, and 51.24% of them with no intention to start a business. For the purposes of ANNs training and testing, the total dataset is divided into three subsamples: train, test and validation subsample in the ANN models, while the SVM, CART and KNN models used the train and test sets together for analysis purposes and the validation sample for the final testing. The structure of samples is presented in Table 1.

Table 1: Sample structure used for the ANN, SVM, CART and KNN models

	ANN models		**SVM, CART, and KNN models**	
Subsample	**Total**	**%**	**Total**	**%**
Train	355	80.14	399	90.07
Test	44	9.93		
Validation	44	9.93	44	9.93
Total	443	100.00	443	100.00

For the purpose of testing the generalization ability of the models, 10 different datasets were randomly generated in the 10-fold CV procedure, each of them pursuing the same structure given in Table 1.

5 RESULTS

The results of the four models performed on 10 samples are presented in Table 2, where the classification rate of each method is expressed as the propotion of correctly classified cases in eac of the validation sample.

Table 2: Results of the 10-fold cross-validation procedure

Sample	*Total classification rate*			
	ANN	*CART*	*SVM*	*KNN*
1	0.7955	0.7273	0.7045	0.5909
2	0.6136	0.5909	0.5455	0.6136
3	0.7955	0.5909	0.7045	0.6364
4	0.7955	0.7727	0.6818	0.7273
5	0.7955	0.7045	0.6364	0.6364
6	0.7045	0.7045	0.7273	0.6136
7	0.7955	0.7045	0.7500	0.5227

8	0.8409	0.6591	0.6136	0.5682
9	0.8421	0.6818	0.6818	0.4773
10	0.8182	0.8409	0.7955	0.7045
Average classification rate	0.7797	0.6977	0.6841	0.6091
St.dev.of classification rates	0.0696	0.0758	0.0714	0.0756

It can be seen from Table 2 that the highest average classification rate was obtained by the ANN (0.7797), followed by the CART with the average classification rate of 0.6977. The lowest average rate was produced by the KNN (0.6091). The ANN also had the smallest standard deviation (0.0696), implying that this method is the most accurate and most stable accross 10 samples. It can be seen that KNN technique performed particularly low in most of the samples, while the ANN outperformed others in all samples except in sample 6 where the SVM was more accurate. Statistical significance of difference in the accuracy could be tested by the t-test of difference in proportion. The results of the t-test show that the p-value is significant on the 5% level only for the difference between the ANN and the KNN models (p=0.0430), while there is no statistically significant difference between the results of other models. In many situations, it is more important to correctly recognize one class of students – in our case the class of students with entrepreneurial intentions (class 1) than the class of students with no intention (class 0). Therefore, classification rates of class 1 and class 0 are further compared across methods and the sensitivity and specificity of each method is computed and presented in Table 3. The sensitivity and specificity ratios were computed according to [11], and the likelihood ratios L_1 and L_0 were also computed.

Table 3: The sensitivity and specificity of the best NN, CART, SVM, and KNN models.

Measure of efficiency	*NN model*	*DT model*	*SVM*	*KNN model*
Sensitivity	0.843889	0.721495	0.722853	0.635132
Specificity	0.690154	0.681263	0.654512	0.592231
Likelihood ratio L_1	2.930801	2.867496	2.144374	1.666607
Likelihood ratio L_0	0.230211	0.414052	0.422161	0.643737

The model with higher sensitivity ratio has a lower type I error in misclassifying a student with an actual positive entrepreneurial intention (class 1) into the class of students with no intention (class 0). Such error yields a greater loss for the society than the type II error, and it is more important to recognize more potential entrepreneurs than to misclassify those who have no entrepreneurial intention. Therefore, the most efficient model is the one that has highest sensitivity, and according to Table 4, it is the ANN model with the average sensitivity of 0.843889, and also the highest likelihood for recognizing class 1 (2.9308).

6 DISCUSSION AND CONCLUSION

The paper investigates the efficiency of machine learning methods in classification models with high-dimensional data, and finds out that the ANN method provides the most efficient model and outperforms other tested models according to criteria of classification accuracy, stability, sensitivity, and specificity. The reason for such domination of ANN could be found in its robustness and the ability to minimize the error in the iterative procedure of optimizing its parameters such as learning rate, while the other methods have predefined values of some input parameters. However, the accuracy of ANN is significantly higher only comparing to the accuracy of KNN model on the 0.05 level, while the difference between the ANNs and other tested models is not found to be statistically significant. It implies that the tested

machine learning methods have many similarities while dealing with a large number of input variables, and that further tests are necessary. Future research could be focused on testing some more methodological improvements in machine learning methods, such as SVM with hierarchical clustering, and others that will enable more thorough analysis of high-dimensional data in machine learning.

References

[1] Apte, C. and S. Weiss, 1997. Data Mining with Decision Trees and Decision Rules, Future Generation Computer Systems , Vol. 13, pp. 197-210.

[2] Behzad, M., Asghar, K., Eazi, M. and Palhang, M., 2009. Generalization performance of support vector machines and neural networks in runoff modeling, Expert Systems with Applications, Vol. 36, pp. 7624–7629.

[3] Bishop, C., 1995. Neural Networks for Pattern Recognition. University Press, Oxford, UK.

[4] Dai, Y-H., 2002. Convergence properties of the BFGS algorithm, SIAM Journal of Optimization, Vol. 13, No. 3, pp. 693-701.

[5] Krueger, N.F. JR., Reilly, M.D. and Carsrud, A.L., 2000. Competing Models of Entrepreneurial Intentions, Journal of Business Venturing, Vol. 15, pp. 411–432.

[6] Lee, J.H., Cha, G.H., Chung, C.W., 1999. A model for k-nearest neighbour query processing cost in multidimensional data space**,** Information Processing Letters, Volume 69, Issue 2, 29 pp. 69-76.

[7] Lee, S., 2010. Using data envelopment analysis and decision trees for efficiency analysis and recommendation of B2C controls, Decision Support Systems, Vol.49, pp. 486–497

[8] Masters, T., 1995. Advanced Algorithms for Neural Networks, A C++ Sourcebook, John Wiley & Sons, Inc., New York, USA.

[9] Paliwal, M. and Kumar U.A., 2009. Neural networks and statistical techniques: A review of applications, Expert Systems with Applications, Vol. 36, pp. 2–17.

[10] Questier, F., Put, R., Coomans, D., Walczak, B. and Vander Heyden Y., 2005. The use of CART and multivariate regression trees for supervised and unsupervised feature selection, Chemometrics and Intelligent Laboratory Systems, Vol. 76, pp. 45-54.

[11] Simon, D. and Boring J.R., 1990. Sensitivity, Specificity, and Predictive Value, In: Walker, H.K, Hall, W.D, Hurst J.W., (eds), Clinical Methods: The History, Physical, and Laboratory Examinations, 3rd edition, Butterworths, Boston, pp. 49-54.

[12] Shin, H.J., Eom, D.H. and Kim, S.S., 2005. One-class support vector machines - an application in machine fault detection and classification, Computers & Industrial Engineering, Vol. 48, pp. 395–408.

[13] Thompson, E.R., 2009. Individual entrepreneurial intent: Construct clarification and development of an internationally reliable metric. Entrepreneurship Theory and Practice, 33, pp. 669-694.

[14] Yeh, C.C, Chi, D.J., Hsu, M.F., 2010. A hybrid approach of DEA, rough set and support vector machines for business failure prediction, Expert Systems with Applications, Vol. 37, pp. 1535–1541.

[15] Yu, H., Yang, J. Han, J., 2003. Classifying Large Data Sets Using SVMs with Hierarchical Clustering, KDD '03 Proceedings of the ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM New York, NY, USA, pp. 306-315.

[16] Zekic-Susac, M., Sarlija, N., Pfeifer, S., 2013. Combining PCA Analysis and Artificial Neural Networks in Modelling Entrepreneurial Intentions of Students, Croatian Operational Research Review, Volume 4, pp. 306-317.

The 12th International Symposium on
Operational Research in Slovenia

SOR '13

Dolenjske Toplice, SLOVENIA
September 25 - 27, 2013

Section IV:

Econometric Models and Statistics

THE DETERMINANTS OF EXPORT PERFORMANCE IN FURNITURE MANUFACTURING: EVIDENCE FROM 26 EU COUNTRIES

Martina Basarac Sertić
Economic Research Division
Croatian Academy of Sciences and Arts
Strossmayerov trg 2, 10000 Zagreb, Croatia
mbasarac@hazu.hr

Abstract: The European Union furniture manufacturing is an assembling industry with high multiplier effect. However, the ongoing economic crisis has put the manufacturing industry under pressure. Therefore, this paper aims to provide analysis of indicators of export performance on the extensive dataset of the 26 European Union member states' furniture manufacturing. Hence, dynamic panel models are estimated by utilizing the "difference" and "system" generalised method of moments estimators for the 2000-2012 period. The results indicate significant impact of foreign demand and the real GDP growth rate on the success of the furniture sector and refocus attention on export performance.

Keywords: furniture manufacturing, export, "difference" GMM estimator, "system" GMM estimator European Union.

1 INTRODUCTION

Manufacturing is beyond doubt essential for the European Union (EU) economy. Namely, industry lies at the heart of the new growth model for the EU economy as outlined in the Europe 2020 Strategy [7]. However, manufacturing industries face a variety of significant challenges arising from the effects of the deep and prolongued global financial crisis and the slow economic recovery. Furthermore, as in previous deep recessions combined with a banking crisis, the crisis was preceded by a long period of rapid credit growth, low risk premiums, abundant availability of liquidity, strong leveraging, soaring asset prices and the development of bubbles in the real estate sector and within the EU, so some Member States became net lenders by a significant share of their GDP while other became large net borrowers [11]. These movements distorted the financial position of many European Union member states causing external imbalances [11]. Above mentioned is also reflected in the manufacturing industries where some sectors have been more deeply affected because they have been more vulnerable than others. Namely, industry, and in particular manufacturing, is bearing a disproportionate share of the burden of the crisis across all EU member states [11].

Therefore, in the broader sense, the area of proposed research will be the manufacturing industry. To be more accurate, a selected section – furniture manufacturing will be analysed.

The objective of this paper is to analyse the determinants of European Union's furniture manufacturing. More precisely, the paper investigates the effects of foreign demand, real effective exchange rate and economic growth on the export of the furniture manufacturing in European Union countries using the panel data analysis.

Individually, the member states have different comparative advantages in terms of forest resources, but form one of the strongest wood sectors in the world as a whole. Furthermore, although there is a large body of literature studying the export demand equations; the approach used in this paper expands existing knowledge on the export competitiveness in European Union economies in several ways. First, we include almost all EU member states in the analysis (Malta is exception due to data unavailability) during the

period between 2000 and 2012. Next, we do not analyse total export of goods and services, but furniture manufacturing export highlighting the significance of export performance on sectorial level. However, not many published papers assess the impact of income and price elasticity on the export competitiveness of the furniture sector. Thus, the analysis of the effect of the selected macroeconomic indicators on the exports of the European furniture sector will help us with finding the answer to the question of which macroeconomic policies form the background for the success of the European wood sector as a whole. Thus, the aim of this paper is to take a closer look at the potential determinants of the furniture sector's exports.

The remainder of the paper is structured as follows: In the next section we discuss relevant characteristics of the furniture sector in European Union member states. Section 3 offers brief literature review. Section 4 is dedicated to describing the data used and the method applied, as well as the reasons behind the choice of a linear dynamic panel model. Section 5 contains the concrete results of the econometric analysis and their interpretation. Finally, section 6 concludes and presents some limitations and possible paths of future research.

2 THE CHARACTERISTICS OF THE FURNITURE MANUFACTURING IN THE EUROPEAN UNION

We begin the analysis by examining the characteristics of the furniture manufacturing with special emphasis on the European Union. Generally, the furniture industry is an assembling, a traditionally labour-intensive and raw material oriented industry which includes craft businesses and large manufacturers.

According to the [6], the European (EU-27) furniture manufacturing included 130,000 enterprises and employed around 1.04 million people in 2010. In value added terms, Germany, Italy and the United Kingdom were the largest Member States in the furniture manufacturing sector, accounting for 22.5%, 16.8% and 10.8% of the EU total respectively [6]. Furthermore, according to the [8] the EU industry is faced with several competitiveness challenges: materials in the EU including wood based products and energy are among the highest priced in the world, labour costs are higher than in non-EU producers (e.g. China), the growing use of packed solutions and low international transport costs have facilitated imports of furniture and furniture components from third countries and the strength of the Euro has not favoured their exports. Hence, as a response to competitive pressure, furniture companies are undertaking a process of modernisation and restructuring as well as finding new business models.

3 LITERATURE REVIEW

The majority of studies analyzing the export performance of the European Union try to identify export demand equation based on aggregate estimation of trade elasticities. For example, [9] and [10] examined the determinants of export performance in euro area. These two studies used different methodologies, but reach the same empirical results. Specifically, they conclude that the real exchange rates and foreign demand to a large extent explain changes in exports for euro area countries.

Furthermore, [1] also investigates the impact of traditional determinants on manufactured goods and nonmanufactured goods and services across France, Germany, Italy and Spain during 2001-2004. According to their analysis, the real effective exchange rate appreciation adversely affected exports in selected countries. Moreover, global demand

contributed positively to exports. Other variables, like capacity utilization and trends contributed to a lesser extent. On the other hand, relative prices were insignificant.

However, studies using disaggregated data at industry level have so far focused mostly on import elasticities [10].

4 EMPIRICAL ANALYSIS

In this section we will examine the impact of potential determinants – the elasticity of income and price (approximated through foreign demand and the real exchange rate) of the furniture sector's real exports, based on panel analysis. Our paper builds on the work by Goldstein and Khan [13].

In our modelling we will employ data during the 2000-2012 period. The geographical coverage of this paper is as follows: EU27 countries were divided in two groups: 1) 15 so called «old» member states and – EU15; 2) 12 «new» member states which joined the European Union in May 2004 and in January 2007 – NMS15).

4.1 Dynamic linear panel data model

Observing the extensive research methodologies used in the empirical studies, we assess the impact of export determinants by using the first-differenced GMM (generalised method of moments) estimator proposed by Arellano and Bond [2] for dynamic panel data. This is because many economic relationships are dynamic in nature and one of the advantages of panel data is that they allow the researcher to better understand the dynamics of adjustment [4]. In so doing, "difference" GMM estimator proposed by Arellano and Bond [2] and "system" GMM estimator proposed by Arellano and Bover [3] and Blundell and Bond [5] are suited for the analysis of the small T, large N panels, characteristic to the data set in this paper.

Therefore, for the purposes of empirical testing, two linear dynamic panel data models are estimated. Furthermore, since there is no available data for all countries and all years of interest, an unbalanced panel model will be used to evaluate the appropriate models.

In so doing, the assumption is that the algebraic signs of foreign demand and the rate of GDP growth will be in line with economic theory, and that their increase will have a stimulating effect on the exports of the furniture manufacturing. When it comes to the impact of the real effective exchange rate on the furniture section's exports, the assumption is that the growth of the real effective exchange rate (depreciation of domestic currency) affects the increase in exports of furniture section. The assumption is that an increase in exchange rate (depreciation) has a positive effect on exports, since it makes them cheaper, whilst at the same time having a negative effect on imports, making them more expensive. The lagged value of a dependent one-period-lagged variable will be used as an instrumental variable.

Further, the models are tested using the Sargan test and Arellano-Bond test for zero autocorrelation in first-differenced errors (m_1 and m_2 tests).

4.2 Data description and sources

This subsection will explain the way of obtaining the variables included in the econometric analysis in great detail, and it will also highlight the specific characteristics of individual time series.

The export data of all 27 European Union member states were originally gathered using the UN Comtrade database in US dollars. The aforementioned database classifies products according to the Harmonized Commodity Description and Coding System managed

by the World Customs Organisation. In so doing, the UN Comtrade database only offers values on an annual basis. The exports of section 31 (according to NACE Rev. 2, it is the section of Furniture Manufacturing), fall in the category 94 according to the HS classification (Furniture, lighting, signs, prefabricated buildings). However, the summary category 94 is stripped of the values of subcategories 9405 (Lamps and lighting fittings, illuminated signs, ect.) and 9406 (Prefabricated buildings). Furthermore, the values of exports in dollars were translated into Euros. The annual values of exports in Euros were then deflated by the consumer prices of individual member states, assuming that all exports were agreed upon in Euros. Therefore, the analysis employed real values in order to exclude the effects of price changes. Finally, the real exports values were translated into indices with 2005 as the base year.

Foreign demand was approximated with the use of average GDP of the 27 EU member states. Values of the real effective exchange rate were taken from the Eurostat website. The analysis employed various values of the real exchange rate, deflated on the basis of the consumer price index. All values were recalculated into indices in the 2005 = 100 form. Eurostat served as a source of the variable of the real GDP growth rate.

5 RESULTS

To investigate the results robustness, two estimation procedures were employed: “difference” GMM estimator and “system” GMM estimator.

Table 1 contains the results of the impact assessment of the selected macroeconomic variables on the exports category 94, i.e. on the exports of the furniture manufactured in the "old" EU member states. The effects of foreign demand, the real effective exchange rate, and real GDP growth rate were examined. In “difference” GMM model there was no autocorrelation between the residuals of the first and second order. Furthermore, based on the Sargan test, the hypothesis that there is no correlation between the residuals and the instruments was accepted. The dependent lagged variable was statistically significant and had a positive algebraic sign. By examining the results of the evaluated panel model, it could be concluded that the estimated results confirmed the statistical significance of the foreign demand to stimulate the growth of exports of the manufactured furniture. On the other hand, the real effective exchange rate and the real GDP growth rate did not prove significant in the analysis.

In “system” GMM model diagnostic test (m_2 statistics) for estimated model are satisfying at 5% confidence level and therefore proposed model is well specified. The dependent lagged variable was statistically significant and had a positive algebraic sign. Furthermore, the results show that the variables foreign demand and real GDP growth rate are statistically significant with expected sign and estimated coefficients. Specifically, higher foreign demand and GDP growth rate lead to an increase in the furniture manufacturing’s real exports.

Table 1: The Results of the Dynamic Linear Panel Model – the ″Old″ EU member states

The impact on real export	Arellano-Bond	Arellano-Bover / Blundell-Bond
C	-1.073 (0.617)	-2.973 (0.388)
Lagged dependent variable	1.188* (0.000)	0.495* (0.000)
Foreign demand	0.054* (0.001)	0.966** (0.055)
Real effective exchange rate	0.033(0.949)	0.184 (0.772)
Real GDP growth rate	-0.011 (0.229)	0.019* (0.000)
Sargan test (p-value)	0.2656	0.756
First-order autocorr. (p-value)	0.2433	0.1911
Second-order autocorr. (p-value)	0.2613	0.2111

Number of observations	150	165
Number of groups	15	15

Source: Authors' calculations
Note: *, **, *** indicate statistical significance at 1%, 5% i 10%; p-values in parenthesis. The "difference" and "system" GMM models with robust standard errors are applied.

The results of the second estimated dynamic linear panel model are given in Table 2. There is no autocorrelation between second-order residual differences. Furthermore, based on the Sargan test, the hypothesis that there is no correlation between the residuals and the instruments is accepted. The dependent lagged variable is statistically significant and has a positive algebraic sign. Based on the analysis, it can be concluded that the export determinant for the furniture manufactured in the new EU member states is foreign demand. The influence of other variables included in the analysis did not appear statistically significant in the model of export competitiveness of the "new" EU member states. In "system" GMM model results are the same.

Table 2: The Results of the Dynamic Linear Panel Model – the ″New″ EU member states

The impact on real export	Arellano-Bond	Arellano-Bover / Blundell-Bond
C	-0.158 (0.513)	1.003 (0.689)
Lagged dependent variable	0.571* (0.004)	0.661* (0.000)
Foreign demand	0.387 (0.534)	-0.182 (0.740)
Real effective exchange rate	0.292 (0.257)	0.302 (0.205)
Real GDP growth rate	0.005*** (0.068)	0.007** (0.018)
Sargan test (p-value)	0.3563	0.9674
First-order autocorr. (p-value)	0.0203	0.0546
Second-order autocorr. (p-value)	0.1821	0.1966
Number of observations	110	121
Number of groups[1]	11	11

Source: Authors' calculations
Note: *, **, *** indicate statistical significance at 1%, 5% i 10%; p-values in parenthesis. The "difference" and "system" GMM models with robust standard errors are applied.

Results from this study suggest that foreign demand and real GDP growth rate are essential for export growth. In this regard, by implementing industrial policy strategies, industry as a whole can be a catalyst that can help create jobs and boost GDP.

6 CONCLUDING REMARKS

The European Union is the world's biggest trader of manufactured goods and services. Hence, in times of recession, it makes sense to rely on manufacturing industry to accelerate the economic recovery. Thus the paper starts by putting the furniture manufacturing into the context of European Union's export performance. We argue that EU furniture manufacturing is faced with several competitiveness challenges but at the same time furniture companies are undertaking a process of modernisation and restructuring.

With this in mind, the paper presents the results of an econometric analysis of the impact of potential determinants - the income and price elasticity of the furniture sector's real exports in the European Union member states – based on panel data analysis. A comparison between two estimation procedures ("difference" and "system" GMM) indicated that foreign demand and the growth rate of real GDP have the most influence on the increase

[1] Initially, 12 „new" EU member states were included in the analysis. However, real GDP growth rate variable for Malta is not available, so the analysis was applied on 11 countries.

in exports of the furniture manufactured in the "old" member states and the real GDP growth rate has the most influence on the increase in exports in "new" member states. However, the paper does not suggest any significant causal relationship between real exchange rate and export.

Additionally, econometric results provide further insight into the specific characteristic of the furniture manufacturing. Namely, foreign demand can help recover in the short run when internal demand is comparatively weak, but in the long-term, econonmic growth is only possible through openness and structural reforms that change the ability and incentives to adopt and develop new technologies [11]. Hence, the new industrial markets outside the European Union are crucial for the European competitiveness, particularly in the context of the economic recovery [11]. In that sense, this paper provides new empirical evidence for understanding the drivers of furniture manufacturing exports' in the post-crisis recession.

Although the potential determinants of export performance in furniture manufacturing presented in this paper prove significant in understanding furniture manufacturing competitiveness, they are not without limitations and can be further improved. In particular, the fact that a part of trade equation cannot be attributed to the traditional explanatory variable (real effective exchange rate) calls for prudence in the construction of the variable. However, aggregate indexes can be less effective than industry specific indexes in capturing changes in industry competitive conditions [12]. Furthermore, bearing in mind the limitations of the analysis, a number of extensions could be envisaged. First, extending the framework of the empirical analysis to other dynamic panel estimators, which could enhance the robustness of our empirical findings. Finally, we could also apply these techniques to estimate the model for each country individually and also for the central, eastern and southeastern european countries.

References

[1] Allard et al., 2005. Explaining Differences in External Sector Performance Among Large Euro Area Countires, IMF Country Report, No. 05/401.

[2] Arellano, M., Bond, S., 1991. Some test of specification for Panel data, Monte Carlo Evidence and Aplication to Employment Equations. Review of Economic Studies, 58, pp. 277-297.

[3] Arellano, M., Bover, O., 1995. Another look at the instrumental variable estimation of error-components models. Journal of Econometrics, 68, pp. 29–51.

[4] Baltagi, Badi H., 2005. Econometric Analysis of Panel Data. Third Edition. England: John Wiley & Sons, Ltd.

[5] Blundell, R., Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics, 87, pp. 115–143.

[6] Eurostat, http://epp.eurostat.ec.europa.eu.

[7] European Commission, 2010. An Integrated Industrial Policy for the Globalisation Era Putting Competitiveness and Sustainability at Centre Stage. Brussels, COM(2010) 614.

[8] European Commission, 2010. EU Manufacturing Industry: What are the Challenges and Opportunities for the Coming Years? 2nd high-level conference on industrial competitiveness, Brussels.

[9] European Commission, 2010. Quarterly Report on the Euro Area, 9(1), European Union.

[10] European Commission, 2010. Quarterly Report on the Euro Area, 9(2), European Union.

[11] European Commission, 2012. European Competitiveness Report: Reaping the benefits of globalization. Commission Staff Working Document, 299 final.

[12] Goldberg, L. S., 2004. Industry-Specific Exchange Rates for the United States. FRBNY Economic Policy Review.

[13] Goldstein, M., Khan, M. S., 1985. Income and Price Effects in Foreign Trade. In: R. W. Jones and P. B. Kenen (Eds.), Handbook in International Economics. Vol. II, Amsterdam: North-Holland, Chapter 20.

UNIT VALUE INDICES IN NATIONAL ACCOUNTS

Draženka Čizmić
Faculty of Economics & Business, University of Zagreb
Trg J. F. Kennedyja 6, Zagreb 10000, Croatia
dcizmic@efzg.hr

Abstract: In many countries unit values indices are used as a proxy to pure price or survey-based price indices. They are used as short-term indicators of inflation transmission, to measure changes in a country's terms of trade, to analyse the effect of exchange rates on import and export prices and as deflators for national accounts. Given that unit value indices are widely used it is important that compilers and users are fully aware of their properties, so that strategic decision to move to hybrid or establishment-based indices can be appropriately made.

Keywords: unit value index, bias, national accounts

1. INTRODUCTION

The System of National Accounts provides a framework within which an integrated set of price and volume measures can be compiled which are conceptually consistent and analytically useful. Unfortunately, it may sometimes happen, especially in the field of foreign trade statistics, that as a result of lack of information the data on which price and volume indices have to be calculated are not adequate for the purpose.[1]

Exports and imports are an important element of the national accounts that require careful treatment in the measurement of prices and volumes. This is especially true within an input-output framework that requires a consistent approach to deflation of exports and imports to be used. Transport costs are an important element of exports and imports. Imports and exports of products are recorded at border values. Total imports and exports are valued at the exporter's customs frontier (f.o.b.). Foreign transport and insurance services between the importer's and the exporter's frontiers should not be included in the value of goods, but recorded as services. However, it is not always possible to obtain f.o.b. values at the detailed product level and details of foreign trade are then shown valued at the importer's frontier. In this case, all transport and insurance services to the importer's frontier are included in the value of imports (c.i.f.).[2]

2. METHODS USED IN COMPILING THE PRICE INDICES FOR EXTERNAL TRADE

Export and import price indices are compiled by three general methods, the nature of which is largely dependent on the source data used. The first and predominant method uses unit value indices compiled from detailed import and export merchandise trade data derived from administrative customs documents. The second method is to compile price indices using data

[1] The primary objective is not simply to provide comprehensive measures of changes in prices and volumes for the main aggregates of the System but to assemble a set of interdependent measures which make it possible to carry out systematic and detailed analyses of inflation and economic growth and fluctuations.

[2] However, it is then necessary to apply a global adjustment within the supply and use table to correct imports from a cif valuation to the required fob basis. This adjustment requires deflation for the compilation of supply and use tables at constant prices. A suitable price index for the deflation of this cif/fob adjustment would need to take account of the price development of transport and insurance services for imported goods.

from surveyed establishments on the prices of representative items exported and imported. Price indices are costly to produce and represent a burden on respondents. Third, there is a hybrid approach that involves compiling establishment survey-based price indices for some product groups and customs-based unit value indices for others.

Unit value indices were advised for countries with a tight or medium budget and well-endowed countries were advised to base their external trade price indices on establishment survey data. The preference for price survey indices was due to bias in unit value indices mainly attributed to changes in the mix of the heterogeneous items recorded in customs documents, but was also attributed to the often poor quality of recorded data on quantities.

2.1 Unit value indices

It is sometimes the case that detailed price and quantity data for a group of closely related commodities are not available but information on the number of units is available in each period along with the value of the products in the shipment. In this case, the value of the products can be divided by the number of units and a unit value price is obtained for the period under consideration. If unit values for the product group can be calculated for two periods, then the ratio of the two unit values can be regarded as an approximate price index. This price index is known in the literature as a unit value price index or a Drobisch index.[3]

Index numbers are generally calculated in two stages. The first stage is the building block of price indices, the measurement of price changes of similar "elementary" items exported or imported by one or more institutional unit (elementary indices). At the next stage of aggregation weights are applied to the elementary indices, and weights are again applied to the resulting indices at higher stages of aggregation, until an overall index is derived.

A mayor problem with the Drobisch price index is that its axiomatic properties are not entirely satisfactory. In addition to not satisfying the invariance to changes in units test if the aggregation is over heterogeneous items[4], this index does not satisfy the identity test, which asks the index number to equal unity if the price vectors for the two periods under consideration remain the same. Unit value index does not satisfy the proportionality test, that is if all prices are multiplied by the positive number λ, then the new price index is λ.[5]

It is common knowledge that customs classes rarely contain only one product, thus the unit values suffer from composition effects, wherein the product composition of the unit value from a given customs class varies from period to period. This can cause the unit value price relative to change even if the prices of the component products have not. The unit value price index therefore tends to be biased. Unit value indices also fail to account for quality and characteristics changes, a difficulty which is associated with index numbers based on price surveys as well.

Products that are traded irregularly, have no quantities reported and display erratic month – to – month changes are usually excluded. Despite the exclusion, the coverage of unit value indices tends to be better than price indices.

Unit value indices in foreign trade are not amenable to the "normal" or usual interpretation of price indices. They differ from the latter by a number of reasons not only the formula but also concepts and data collection procedures. The difference between the two approaches to price measurement is hitherto not well understood.

---

[3] in honour of the German measurement economist who first introduced this type of index

[4] It is important to recognize that a Drobisch price index cannot be used over very heterogeneous items since the resulting index is not invariant to change in the units of measurement. Thus a unit value price index can only be used over products that are measured in the same units and are "reasonably" homogeneous.

[5] The unit value index only satisfies the proportionality test in the unlikely event that relative quantities do not change.

2.2 Unit value index bias

The following are grounds upon which unit value indices might be deemed unreliable: 1) Bias arises from compositional changes in quantities and quality mix of what is exported and imported. 2) For unique and complex goods, model pricing can be used in establishment – based surveys where the respondent is asked to price each period a product. This possibility is not open to unit value indices. 3) Methods for appropriately dealing with quality change, temporarily missing values, and seasonal goods can be employed with establishment – based surveys to an extent that is not possible with unit value indices. 4) The information on quantities in customs returns, and the related matter of choice of units in which the quantities are measured, has been found in practice to be seriously problematic. 5) With customs unions countries may simply have limited intra – area trade data to use. 6) An increasing proportion of trade is in services and by e – trade and not subject to customs documentation. 7) Unit value indices rely to a large extent on outlier detection and deletion. Given the stickiness of many price changes, such deletions run the risk of missing the large price catch – ups when they take place and understating inflation.

It is generally thought that constructing broader unit value prices (i.e., aggregating over more specific products to form unit value prices) will lead to a greater degree of bias in a unit value price index as compared to the underlying "true" index.

2.2 Evidence of unit value bias

Very few countries are able to provide both, a unit value index and a true price index on a regular basis. Germany is one of those countries which offer the opportunity to study the impact of the still not well understood methodological differences of the two tools of measuring the price development in export and import.

Silver (2008) compared unit value indices and price indices for both Germany and Japan for exports and imports. Unit value indices were found to seriously misrepresent price indices in the sense that discrepancies between unit value indices and price index were substantial; changes could not be relied upon to have the same sign; there was no evidence of long-run relationships between price index and unit value indices; and unit value indices were of little help in predicting price index.

Such discrepancies can be regarded as seriously misleading for economists. The discrepancy for individual months can be much larger than mean discrepancy, as reflected by an associated standard deviation of 1.0 percent and maximum of 7.3 percentage points for Germany's import month-on-month index changes. For about 25 percent of month-on-month comparisons the signs differed.

The values of exports and imports of Germany and Japan were deflated over the period from 1999 to 2005 by corresponding unit value indices and price indices and the results compared. For example, the volume of exports by Japan increased by 50 percent when a unit value deflator was used, but the increase was halved when a price index was used.

2.3 Unit value indices improvement

United Nations emphasized the need to stratify unit values to the (limited) extent possible and drew attention to doing so where possible by country of destination and size of batch. Stratification is also possible for shipments by/to (major) establishments to/from given countries. It will usually be the case that use of finer commodity classification to generate unit value prices that are then inserted into a bilateral index number formula will generate closer approximations to an underlying preferred index.

Large catch – up price changes may be deleted by automatic outlier detection routines, resulting in unit value indices that are unduly stable, and volatile prices changes, due to exchange rate fluctuations, may lead to unduly high dispersion parameter values, used in deletion routines, and deletion rates. Improved deletion routines are certainly advocated when unit value indices are used.

Superlative index number formulas (Fisher, Törnqvist and Walsh) make symmetric use of reference and current period quantity information, can be justified as providing a good approximation to a "true" index defined in economic theory. In particular, the Fisher index has good axiomatic properties.

2.4 Compilation of hybrid indices

Unit value indices are used by many countries and a move to price indices has resource consequences. One possibility is to identify whether there are particular commodity classes less prone to unit value bias and utilize unit value indices only for these sub aggregates in a hybrid overall index and price indices elsewhere. The extent to which unit value indices are included in a hybrid index depends on the resources of the country's statistical authority, the availability of alternative sources and the reliability of the unit value indices for the goods considered.[6]

However, it is the case for countries whose primary source of price change information is establishment – based price surveys, that unit value indices are exceptionally used for goods whose characteristics are considered to be homogeneous.

2.5 Move to establishment-based price surveys – The gradualist approach

A gradualist approach using hybrid indices has major resource benefits. There will be some "low – hanging fruit" establishments responsible for relatively high proportions of exports and imports some of which may be owned by the state and may have some reporting obligation. There will also be industries in which unit values indices are prima facie inadequate measures of price changes. Further, there may be industries which account for a substantial proportion of trade and the pay off of reliable data far outweighs the survey costs.

The gradualist approach requires as a first step a rigorous evaluation of each commodity group of the relative pay – off and cost of abandoning unit value indices. A potential problem with a gradualist approach is that longer – term changes in the index become problematic.

3. THE UNITED NATIONS PRICE INDICES FOR EXTERNAL TRADE

The Statistical Office of the United Nations Secretariat compiles the following indices relating to movements of prices of commodities entering into international trade: a) primary commodities: price index; b) non-ferrous base metals: price index; c) machinery and transport equipment: price index; d) manufactured goods exports: unit value indices and quantum index; e) fuel imports: unit value index and quantum index; f) total exports and imports: unit value index, quantum index and terms-of-trade index.

The unit value indices are estimates of the unit values of exports of manufactured goods from individual countries and groups of countries in any given period, relative to the unit

[6] For example, some oil – producing countries use unit value indices, but because detailed reliable data are readily available from the oil – producing establishments for this important sector, the unit value indices are complemented by survey – based price indices or price quotations from international markets.

values of those exports in a base year. Changes in the unit value indices could be considered to represent approximate price movements for world exports of manufactures. The unit value indices for each country are obtained mainly from national sources. Where unit value index numbers, or the national data necessary to compute them, are not available in any given period then estimates are made by the Statistical Office. The unit value indices for country groups are calculated according to the Paasche formula as current-period-weighted averages of indices for each of the countries included in the group.

The unit value indices are estimates of the unit values of fuel imports by individual developed countries and groups of developed countries in any given period, relative to the unit values of those imports in a base year. The unit value indices for country groups are calculated according to the Paasche formula.

The unit value indices are estimates of the unit values of total exports or imports from groups of countries in any given period, relative to the unit values of those exports or imports in a base year. The unit value indices of exports and imports for groups of the countries are calculated according to the Paasche formula.

4. EUROSTAT'S EXTERNAL TRADE INDICES

The primary source of data is the CN trade statistics supplied to Eurostat by the Member States. Since 1 January 1993, the date of abolition of the inner frontiers of the Union, statistics on trade between the Member States are no longer collected via customs declarations. Instead, monthly and recapitulative statistical declarations are transmitted directly by companies to the relevant national administrators.

Eurostat's unit value indices are calculated from the original data without aggregation over partners or products.[7] For most CN codes there is information on value, weight and sometimes a second, supplementary quantity unit, such as number of items. In this case two types of unit value are available.

Eurostat's method of dealing with wide-tailed distributions is to use the robust regression technique first described by Hinich and Talwar. The method starts from the observation that, whereas the level of unit values across partner countries may differ, changes in levels are very similar not only across partner countries but also across related products, compared with the background level of noise in unit value data.

Each month, the "isolated" monthly CN data for retained items are processed, block by block, to give Laspeyres and Paasche numerators and denominators for all the primary indices that are required. This information is stored, and used by a further stage of processing to produce index links at a higher level of product or zone aggregation.

Sets of indices are calculated for several product classifications. Higher levels of product class are found by aggregation of the numerators and denominators of the constituent indices.[8]

Laspeyres unit value and volume links for the EU are calculated by weighting the Laspeyres links for each individual reporting country by the value of trade for the previous year. An EU value link is found by combining the value links for individual reporters with the same weights. The Paasche links for the EU are found by division.

[7] One exception to the rule of no aggregation is where there is a change in the CN between two years.

[8] Sometimes a constituent index for a small country is missing for one month. Either its trade is zero, or its sample coverage ratio is judged too low to give a reliable unit value index. It has been found that it is not satisfactory to calculate the larger index simply by aggregating those constituent indices that happen to be available. Eurostat's solution is to estimate the level of the missing unit value index, and the index weight for the Paasche index.

5. CONCLUSION

In many countries unit values indices are used as a proxy to pure price or survey-based price indices. They are used as short-term indicators of inflation transmission, to measure changes in a country's terms of trade, to analyse the effect of exchange rates on import and export prices and as deflators for national accounts. In spite of their widespread use, price surveys are preferred due to bias in unit value indices mainly attributed to changes in the mix of the product groups or in the underlying products recorded in customs documents. A main advantage of the use of unit value indices is their coverage and relatively low resource cost.

Given that unit value indices are widely used it is important that compilers and users are fully aware of their properties, so that strategic decision to move to hybrid or establishment-based indices can be appropriately made. One possibility is to identify whether there are particular commodity classes less prone to unit value bias and utilize unit value indices only for these sub aggregates in a hybrid overall index and price indices elsewhere.

While a unit value index is basically resulting from foreign trade statistics as a kind of by – product, the compilation of a true price index is much more demanding. It requires special surveys addressing exporting and importing establishments as well as compliance with some principles of price statistics among which aiming at "pure price comparisons" is most prominent.

For the aggregation of homogeneous items, the unit value index is the best index and superlative index numbers biased, and for the aggregation of heterogeneous items, superlative index numbers are best index and unit value index numbers biased. The determination of whether or not an item is homogeneous is critical to the choice of index number formula, but in practice many items are broadly comparable, and neither a unit value nor a Fisher index is appropriate.

REFERENCES

(1) Diewert E., von der Lippe P.; Notes on Unit Value Bias; 2010; http://faculty.arts.ube.ca/ediewert/dp1008.pdf
(2) Export and Import Price Index Manual, Chapter 2, IMF, http://www.imf.org/external/np/sta/tegeipi/ch2.pdf
(3) Handbook on price and volume measures in national accounts, Eurostat, 2004, http://epp.Eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-41-01-543/EN/KS-41-01-543-EN.PDF
(4) Implementing unit value indices in the annual OECD international trade in commodity statistics (ITCS) database, 2011, http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=STD/TBS/WPTGS(2011)68
(5) Lippe von der P., Price indices and unit value indices in German foreign trade statistics, http://mpra.ub.uni-muenchen.de/5525/1/MPRA_paper_5525.pdf
(6) Methods used in compiling the United Nations price indexes for external trade, Volume II, 1991, http://unstats.un.org/unsd/publication/SeriesM_82vol2E.pdf
(7) Moore A., Jones J.; Import and export unit value indices, http://www.centralbank.org.bb
(8) Silver M., An Index Number Formula Problem: The Aggregation of Broadly Comparable Items, 2009, http://www.ottawagroup.org/Ottawa/ottawagroup.nsf/4a256353001af3ed4b256bb0012156
(9) Statistics on the trading of goods, User guide, Eurostat, http://edz.bib.uni-mannheim.de/www-edz/pdf/eurostat/98/CA-12-98-974-EN-1-EN.pdf

INTERNET BANKING USAGE IN SELECTED EUROPEAN COUNTRIES: MULTIPLE REGRESSION ANALYSIS APPROACH

Ksenija Dumičić, Anita Pavković and **Irena Palić**
Faculty of Economics and Business, University of Zagreb,
Trg J.F. Kennedy 6, HR-10000 Zagreb, Croatia
{kdumicic,apavkovic,ipalic}@efzg.hr

Abstract: Regression models using six regressors that impact Percentage of Internet users for Internet banking based on EUROSTAT data for EU27 and Croatia for 2011 are studied. Three regression models found to be useful for explanation of the regressand variable. An increase of "GDPpc in PPS" and "Share of GDP for education", as well as an increase of "Percentage of households with Internet access" and "Broadband penetration rate", influenced an increase of the regressand variable. Cluster analysis based on seven variables resulted with four clusters of countries.

Key words: Internet banking, European Union, GDP per capita in Purchasing Power Standards, Multiple linear regression analysis, OLS estimators, Cluster analysis

1 INTRODUCTION

After [11], Internet banking refers to the use of the Internet as a remote delivery channel for retail banking services. According to [9], Internet banking includes electronic transactions with a bank for payment etc., or for looking up account information. At the banks' side the Internet banking cuts business costs. At the customers' side, not only lower operational banking costs, but advancement of user-friendly information technology (IT) solutions encourage them towards Internet banking use. In the Internet Age an increased competition among banks has influenced the retail banking products and pricing. Since the Internet market has grown into a profitable competitive area for the banking industry, a key strategic issue for banks is adoption of Internet banking. IT improvements affect the retail distribution channels and the banking services' operating costs by reducing number of branches, etc.
The purpose of this paper is to investigate whether the Percentage of Internet users for Internet banking (according to EUROSTAT: Percentage of individuals aged 16 to 74 using the Internet for Internet banking within the last 3 months before the survey, see [9]), as the dependent variable $Y_{Int\text{-}B}$, is influenced by the following regressors: X_{GDPpc}- GDP per capita in PPS (EU 27=100); X_{ExpEdu}- Public expenditure on education as a percentage of GDP; X_{CSkill}- Individuals' level of computer skills as a percentage of people aged 16 to 74 using computers; $X_{AccessHH}$- Level of Internet access as a percentage of households with Internet access at home, $X_{IntSkill}$ - Individuals' level of Internet skills as percentage of the total number of individuals aged 16 to 74; and X_{BB}- Broadband penetration rate, which indicates the percentage of broadband connections *per capita*. Data for EU27 and Croatia for 2011 were taken from [9] and [7]. The research hypothesis is that concerning all variables under investigation, clusters of similar countries might be distinguished.
In [6] impact of Internet retailing are studied, seeking to break new ground by attempting to use the current literature to help predict future trends for online shopping. Security, personal and social influence on Internet use are investigated in many scientific papers. Some of them study technical IT solutions, from the banks' or from the customers' point of view. A majority of studies highlight that "security" is the biggest single concern for customers. "Push" and "pull" factors for explaining customer conversion to Internet banking using regression analysis is presented in [3]. The matters of consumer's trust in e-banking are investigated in [17] using regression analysis. According to [19] number of Internet banking users has not risen as rapidly as expected. In [5], applying structural equation modelling,

authors analyse customers' concerns about trust and security. Based on quantitative model which includes security, usability, personality and social influence, paper [20] investigates customers' perceptions influence on Internet banking use.
Many researches have been done for European countries. Adoption of Internet services in the EU Candidate Countries is described in [4]. Cases of Turkey and UK are studied in [16]. How IT development affects the way banks conduct their business in Estonia is investigated in [8]. In [2] competition between conventional 'brick and mortar' banks and pure Internet banks across European countries in the period 1995-2004 is studied with panel analysis. Paper [15] studies factors underlying the customers' decision to adopt Internet banking in Poland. Paper [14] elaborates the impact of Internet on the retail banking in Macedonia. Internet banking use in Central European transition countries is focused in [10], and [1] investigates the Balkans and Greek economy.

2 DATA EXPLORATION AND LINEAR REGRESSION ANALYSIS RESULTS

Key findings on *Information Society in European Countries* could be found at the websites [12] and [13]. Fig. 1 shows the trends of percentage of individuals using the Internet for Internet banking in EU27 and Croatia in the period 2004 to 2012.

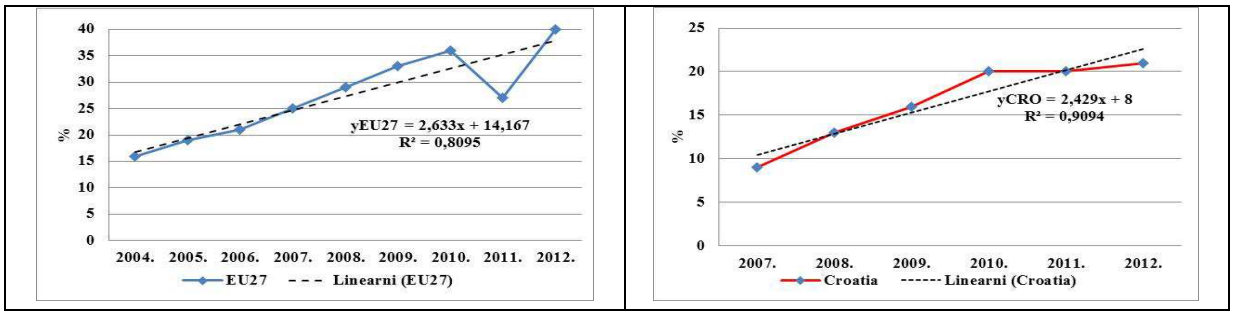


Figure 1: Linear trends for % of individuals using the Internet for Internet banking in EU27 and Croatia

Fig. 2 shows data for the same variable for each of 28 countries in 2011 and 2012. Only for UK data are not available for 2011, so the value 45% was imputed based on 2010.

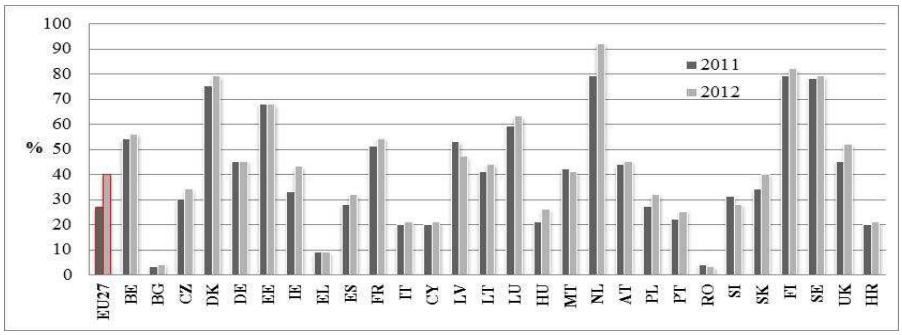


Figure 2: Percentage of Internet users for Internet banking in EU27 countries and Croatia in 2011 and 2012

The coefficients of variation for all variables show great data variability. The highest is $\mathbf{V}(Y_{Int\text{-}B})$=56.16%, and the lowest is $V(X_{AccessHH})$=18.80% (the minimum is 45% for Bulgaria, and the maximum 94% for The Netherlands. The distributions for all the variables are positively skewed, with the highest *skewness*(X_{GDPpc})=2.41, caused by Luxembourgh's outlier for „GDPpc in PPS" which is 271 (with the base EU27=100), see Fig. 3.

Correlation matrix for all seven variables shows that all the correlations are positive. The strongest positive correlation appears to arise for $Y_{Int\text{-}B}$ and $X_{AccessHH}$, with the correlation coefficient $r_{Y_{Int\text{-}B};X_{AccessHH}}$=0.87, and between $Y_{Int\text{-}B}$ and X_{BB}, with $r_{Y_{Int\text{-}B};X_{BB}}$ =0.78.

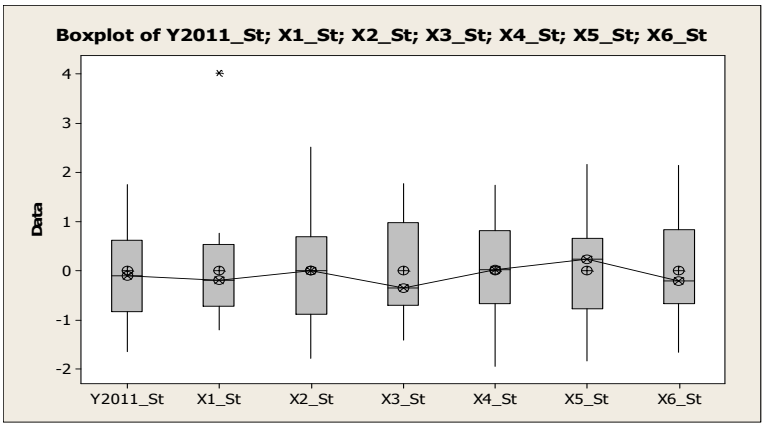


Figure 3: Box plot of standardized values for all seven variables* under study for n=28 countries in 2011

Further, cluster analysis using standardized values of all seven variables with Ward linkage and squared Euclidean distance resulted with four-cluster solution as the most appropriate (Tab. 1, Fig. 4).

Table 1: Clusters of countries based on standardized values for seven variables: $Y_{Int\text{-}B}$; X_{GDPpc}; X_{ExpEdu}; X_{CSkill}; $X_{AccessHH}$; $X_{IntSkill}$; X_{BB} using Ward linkage and squared Euclidean distance for EU27 and Croatia in 2011

Cluster no.	No. of countries	Countries
1	**10**	Belgium, United Kingdom, France, Germany, The Netherlands, Ireland, Austria, Denmark, Finland, Sweden
2	**5**	Bulgaria, Romania, Czech Republic, Poland, Slovakia
3	**12**	Estonia, Malta, Latvia, Lithuania, Cyprus, Greece, Croatia, Estonia, Italy, Hungary, Portugal, Slovenia
4	**1**	Luxembourg *(outlier for $X_{GDPpc\ in\ PPS}$ with z>3)*

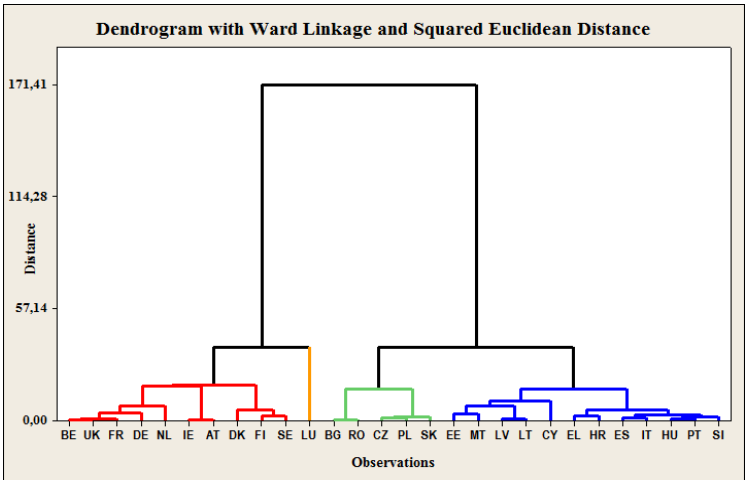


Figure 4: The dendrogram for the four-clusters solution

The linear regression model with parameters estimated based on ordinary least squares (OLS) for the sample would be:

$$\hat{y}_i = \hat{\beta}_0 + \sum_{j=1}^{K} \hat{\beta}_j \cdot x_{j,i}, \quad j = 1,2,...,K; \; i = 1,2,...,n \tag{1}$$

In this research n=28 countries and K=6 regressors, or j=1,2,…6. According to [18], *all possible regressions* analysis was applied.

Table 2: Part of the *All Possible Regressions Analysis* for $Y_{Int\text{-}B}$ = the dependent variable, n=28

	p-values for the regression coefficients										
	Economic level variables		IT development level variables								
var. #	X GDPpc	X ExpEdu	X CSkill	X AccessHH	X IntSkill	X BB	s.e. reg.	Adj R^2	R^2	Mallow's Cp	F-test p-value	Model #
1	.0059						19.64	.23	.2571	79.82	.0059	-
1		.0034					19.27	.26	.2849	75.94	.0034	-
1			.2590				22.23	.01	.0487	108.94	.2590	-
1				.0000			11.25	.75	.7563	10.05	.0000	**Model 2**
1					.0370		20.93	.12	.1568	93.83	.0370	-
1						.0000	14.35	.59	.6033	31.44	.0000	**Model 3**
2	.0035	.0021					16.51	.45	.4951	48.57	.0002	**Model 1**

With six regressors, among (2^6-1)=63 possible regression models, the vast majority of them were either not statistically significant, or with small value for R^2. In Tab. 2 for seven models the p-values for t-tests for parameters, coefficients of determination, regression standard errors, and Mallow's Cp indicators are given. Based on the predefined criteria for the model to be useful (statistical significance, R^2 to be at least 0.5, and filled regression model assumptions) only of the following three models might be accepted:

Model 1*: Multiple regression model with K=2 regressors*. The estimated model is:

$$\hat{y}_{Int-B,i} = -29.153 + 0.239 \cdot x_{GDPpc,i} + 8.345 \cdot x_{ExpEdu,i} \; .$$

Based on the t-test of significance for X_{GDPpc} (p-value=0.0035), and for X_{ExpEdu} (p-value=0.0021), each of two regressors happened to be statistically significant at 1% significance level. The whole multiple regression model based on overall F-test (p-value=0.0002) is also statistically significant at 1% significance level. The regression coefficient $\hat{\beta}_1$ shows that if X_{GDPpc} would increase by one (variable in PPS, EU27=100), without changing in X_{ExpEdu}, the regression value of Percentage of Internet users for Internet banking would increase by 0.239 percentage points. The regression coefficient $\hat{\beta}_2$ shows that if X_{ExpEdu} would increase by one (variable given as percentage of GDP), without changing X_{GDPpc}, the regression value of the variable Percentage of Internet users for Internet banking would increase by 8.345 percentage points. Diagnostic tests for Model 1 were conducted: the Jarqu-Bera normality test (p-value=0.8692), the White heteroskedasticity test (p-value=0.2705), Breusch-Godfrey test up to the second order serial correlation (p-value=0.4598), and multicollinearity examination with VIF=1.009< 5. Diagnostic indicates that none of the model assumptions is violated at 5% significance level. The coefficient of determination is R^2=0.4951, and the coefficient of variation for the regression is $\hat{V}$ =41.47%.

Model 2*: Simple linear regression model with regressor $X_{AccessHH.}$* The estimated model is:

$$\hat{y}_{Int-B,i} = -63.621 + 1.462 \cdot x_{AccessHH,i} \; .$$

The t-test of significance for $X_{AccessHH}$ (p-value<0.0001) shows that this variable is statistically significant at 1% significance level. The regression coefficient tells us that if the variable $X_{AccessHH}$ would increase by one percentage point (percentage of households with Internet access at home), the regression value of Percentage of Internet users for Internet banking would increase by 1.462 percentage points. The Jarqu-Bera normality test (p-value=0.1662), the White heteroskedasticity test (p-value= 0.2633), and Breusch-Godfrey second order serial correlation test (p-value=0.2537). No violations of the regression model assumptions were found at 5% significance level. The coefficient of determination is R^2=0.7563, and the regression coefficient of variation is $\hat{V}$ =28.25%.

Model 3*: Simple linear regression model with X_{BB} as the regressor.* The estimated model is:

$$\hat{y}_{Int-B,i} = -28.657 + 2.603 \cdot x_{BB,i} .$$

The t-test of significance of X_{BB} (p-value<0.0001) shows it is statistically significant for explaining the dependent variable at 1% significance level. The regression coefficient shows that if X_{BB} would increase by one percentage point (percentage of broadband connections *per capita),* the regression value of Percentage of Internet users for Internet banking would increase by 2.603 percentage points. The Jarqu-Bera normality test (p-value=0.2561), the White heteroskedasticity test (p-value=0.4128), and Breusch-Godfrey up to second order serial correlation test (p-value=0.1712), all show that no violations of the regression model assumptions were found at 5% significance level. The coefficient of determination R^2 indicates that 60.33% of the total variation is explained by the estimated linear regression model. The regression coefficient of variation is $\hat{V}$ =36.04%.

3 CONCLUSIONS

Using six explanatory variables for studying un impact on “Percentage of Internet users for Internet banking” in EU27 and Croatia in 2011, it was found that all correlations under study were positive, being weakly to moderately strong. Among all possible linear regression models only three of them were statistically and interpretatively acceptable explaining at least 50% of total sum of squares. The model with two regressors indicating economic development, “GDP *per capita* in PPS (EU-27=100)” and “Public expenditure on education as share of GDP in 2010“, has shown that their increase is resulting with a statistically significant increase of “Percentage of Internet users for Internet banking”. Two simple linear regression models have shown that an increase in variables indicating IT development, “Level of Internet access from home“ and “Broadband penetration rate”, is resulting with statistically significant increase of “Percentage of Internet users for Internet banking”. It is surprising that variables “Individuals' level of computer skills” and “Individuals' level of Internet skills” explained the regressand variable quite poorly, with coefficients of determination below 0.15. Cluster analysis conducted gave four-clusters solution with similar countries within each cluster. Croatia, which joined EU in 2013, happened to be in the cluster with Estonia, Malta, Latvia, Lithuania, Cyprus, Greece, Estonia, Italy, Hungary, Portugal, and Slovenia that are all similar with one another considering analysed variables. As expected, the most developed countries are geathered in the cluster of their own.

References:

[1] Apak, S., Atay, E., 2012. Eurozone Debt Crises Versus New Opportunities for Global Internet Banking Collaboration and Strategic Alliances in the EU and Balkan Countries. Procedia - Social and Behavioral Sciences, 8th International Strategic Management Conference, Vol. 58 (12), pp. 560–568.

[2] Arnaboldi, F., Claeys, P., 2008. Internet banking in Europe: a comparative analysis. Research Institute of Applied Economics, Working Abstract.
[3] Bughin, J., 2001. "e-push or e-pull? Laggards and first-movers in European on-line banking. Journal of Computer-Mediated Communication, Vol. 7 (1), pp. 15.
[4] Centeno, C., 2004. Adoption of Internet services in the Acceding and Candidate Countries, lessons from the Internet banking case. Telematics and Informatics, Vol. 21 (4), pp. 293-315.
[5] Chandio, F.H., Abbasi, M.S., Nizamani, H.A., Nizamani, Q.A., 2013. Online banking information systems acceptance: A structural equation modelling analysis. International Journal of Business Information Systems, Vol. 12 (2), pp. 177-193.
[6] Doherty, N.F., Ellis-Chadwick, F., 2010. Internet retailing: the past, the present and the future. International Journal of Retail & Distribution Management, Vol. 38 (11/12), pp. 943-965.
[7] Digital Agenda for Europe. https://ec.europa.eu/digital-agenda/en/download-data (July 15 2013).
[8] Eriksson, K., Kerem, K., Nilsson, D., 2008. The adoption of commercial innovations in the former Central and Eastern European markets: The case of internet banking in Estonia. International Journal of Bank Marketing, Vol. 26 (3), pp.154 – 169.
[9] EUROSTAT:
http://epp.eurostat.ec.europa.eu/tgm/table.do?tab=table&init=1&language=en&pcode=tin00099&plugin=1%5 (12 July 2013)
http://epp.eurostat.ec.europa.eu/tgm/table.do?tab=table&plugin=1&language=en&pcode=tec00114. (12 July 2013)
http://epp.eurostat.ec.europa.eu/tgm/table.do?tab=table&init=1&language=en&pcode=tsdsc510&plugin=1. (12 July 2013)
http://epp.eurostat.ec.europa.eu/tgm/table.do?tab=table&init=1&language=en&pcode=tsdsc460&plugin= 1. (12 July 2013)
http://epp.eurostat.ec.europa.eu/tgm/table.do?tab=table&init=1&language=en&pcode=tin00134&plugin=1. (12 July 2013)
http://epp.eurostat.ec.europa.eu/tgm/table.do?tab=table&init=1&plugin=1&language=en&pcode=tsdsc470. (12 July 2013)
[10] Havranek, T., Irsova, Z., 2013. Determinants of Bank Performance in Transition Countries: A Data Envelopment Analysis. Transition Studies Review, Vol. 20 (1), pp. 1-17.
[11] Hoa, C.T.B., Wub, D.D., 2009. Online banking performance evaluation using data envelopment analysis and principal component analysis. Computers & Operations Research, Vol. 36 (6), 1835–1842.
[12] ICT STATISTICS. http://www.itu.int/en/ITU-D/Statistics/Pages/default.aspx. (July 15 2013).
[13] Measuring the Information Society in 2012. http://www.itu.int/en/ITU-D/Statistics/Documents/publications/mis2012/MIS2012_without_Annex_4.pdf. (July 15 2013).
[14] Nenovski, T., Delova Jolevska, E., Andovski, I., 2012. Banking Services in Terms of Changing Environment: The Case of Macedonia. Procedia - Social and Behavioral Sciences, Vol. 44, XI International Conference, Service Sector in Terms of Changing Environment, Ohrid, p. 347–356.
[15] Polasik, M., Wisniewski, T.P., 2009. Empirical analysis of internet banking adoption in Poland. International Journal of Bank Marketing, Vol. 27 (1), pp. 32-52.
[16] Sayar, C., Wolfe, S., 2007. Internet banking market performance: Turkey versus the UK. International Journal of Bank Marketing, Vol. 25 (3), pp.122 – 141.
[17] Wong, D.H., Loch, C., Yap, K.B., Bak, R., 2009. To Trust or not to Trust: The Consumer's Dilemma with E-banking. Journal of Internet Business, Vol. 9 (6). 27 p.
[18] Woolridge, J. M., 2013. Introductory Econometrics: A Modern Approach. 5th Edt., South-Western College Publishing.
[19] Xue, M., Hitt, L., Chen, P., 2011. The Determinants and Outcomes of Internet Banking Adoption. Management Science, Vol. 57 (2), pp. 291-307.
[20] Yoon, H.S., Steege Barker, L.M., 2013. Development of a quantitative model of the impact of customers' personality and perceptions on Internet banking use. Computers in Human Behavior, Vol. 29 (3), pp. 1133-1141.

A SEMIPARAMETRIC APPROACH TO THE ANALYSIS OF YOUNG WOMEN'S PARTICIPATION IN THE LABOUR FORCE IN SERBIA

Kosovka Ognjenović
Institute of Economic Sciences
12 Zmaj Jovina, 11000 Belgrade, Serbia
kosovka.ognjenovic@ien.bg.ac.rs

Abstract: In this paper the participation of young women in the labour force in Serbia is discussed, following two econometric approaches most recently used in the analysis of the binary choice model. The parametric probit and the semiparametric single-index models are specified and estimated and a formal test for the selection of the appropriate model specification is conducted. Following the estimation outputs of the selected participation equation, an economic interpretation of the importance of the factors that determine the propensity to work of young women in Serbia is given.

Key words: labour force participation, semiparametric estimation, Serbia, young women.

1 INTRODUCTION

This paper examines the participation of young women in the labour force in Serbia following two econometric approaches most recently used in the analysis of the binary choice model. The parametric probit and the semiparametric single-index models are specified and estimated so that some insights into the behaviour of young women who participate in the labour force in Serbia are obtained. Additionally, a formal test for the selection of the appropriate model specification [7], i.e. for testing a parametric versus a semiparametric alternative, is conducted in order to choose the model that provides consistent estimates of the regression coefficients. Based on the selected participation equation, an economic interpretation of the importance of the factors that determine the propensity to work of young women in Serbia is given.

The use of both parametric and semiparametric econometric models in empirical studies is motivated by the relaxation of the assumptions about the error distribution that is enabled by an alternative semiparametric model specification. The inconsistency of estimators when the parametric econometric models are misspecified has induced the need for developing estimators for semiparametric models. These estimators can restore the assumptions about the error term, by applying corrections in the parametric models or by data rearrangement, and meet the requirements about consistency and asymptotic normality of estimates and reduced bias. In particular, these problems arise in the estimation of economic models of labour supply, when the standard procedures are applied for incomplete samples inducing the sample selection bias as it is explained in [5].

The subpopulation of young married women of the age 18-30 years is examined. The age interval of 15-30 years (or 18-30 years for young adults) is recommended in the policy documents [1] and [12] that guide national youth policies related to the issues of education, health and family planning, employment, social inclusion, etc. These policies are in line with the seminal policies that have been in place throughout the European Union member countries and with the requirements for the candidate countries. Position of young people in Serbia in terms of employment perspectives has been unfavourable for years, and the effects of the public policies aimed at increasing the employment chances of new entrants to the labour market are still low [10], [13]. This situation particularly affects young women and has broader social manifestations, such as entering the labour market at older ages, prolonged childhood and living with parents in the late twenties, low fertility rates, delayed decisions on family planning, etc. [14]. Some of these trends are common for many European countries

[2]. The economic aspects of these problems, caused by prolonged transition in Serbia, have still not examined with enough attention.

The paper contains the three main sections. After explaining the aim of the paper in the introductory part, section two provides a detailed explanation of the econometric methodology employed in the paper. Two forms of the binary choice model are derived, the parametric probit and the semiparametric single-index model, together with their estimators. An asymptotic test statistic for the selection of the supreme model is described. Section three explores a micro data set and provides estimates of the labour force participation equation and discussion of the results. The last section provides main conclusions.

2 ECONOMETRIC METHODOLOGY

The usual way to derive the parametric probit and the semiparametric single-index model is to start with the binary choice model as follows [4], [8]:

$$\begin{aligned} y &= 1 \quad \text{if } y^* = x'\beta + \varepsilon > 0 \\ &= 0 \quad \text{otherwise.} \end{aligned} \tag{1}$$

y is an indicator variable that takes two values (1,0) depending on the sign of the unobserved variable y^*, x and β are (*kx1*) and (*k+1)x1* vectors of explanatory variables and unknown regression coefficients, respectively, while ε is an error term.

The main difference between the parametric binary probit and the semi-parametric single-index model lies in the assumptions made about the error distribution.

If it is assumed that ε is identically and independently distributed and independent of the vector of explanatory variables x, i.e. $\varepsilon \sim N(\mu, \sigma^2)$, then the model (1) produces the parametric binary probit model of y that can be consistently estimated by the maximum likelihood (ML) method assuming the model (1) is correctly specified. The log-likelihood function for the parametric binary probit model has the form:

$$\log l_N(\beta) = \sum_{i=1}^{N} \{y_i \log \Phi(x_i'\beta) + (1 - y_i) \log[1 - \Phi(x_i'\beta)]\}, \; i=1,..,N, \tag{2}$$

where $\Phi(.)$ is the standard normal distribution function. The ML method relies on computing the vector $\hat{\beta}$ of β that maximises the log likelihood function (2). The estimates and their variances have desirable asymptotic properties [4].

If the distribution of ε is unknown, there are semiparametric methods that still make it possible to get the consistent estimates of the regression coefficients β. In that case the single-index form of the binary choice model (1) can be given as:

$$P(y = 1|x) = G(x'\beta), \tag{3}$$

where $G(.)$ is an unknown function, but $x'\beta$ is known up to the finite-dimensional coefficient $\beta \in \mathrm{B}$, $\mathrm{B} \in \Re^K$. The maximisation of the quasi log-likelihood function is the possible way to estimate β [9]:

$$\log l_N(\beta) = N^{-1} \sum_{i=1}^{N} \{y_i \log \Gamma_N(x_i'\beta) + (1 - y_i) \log[1 - \Gamma_N(x_i'\beta)]\}, \; i=1,..,N, \tag{4}$$

where $\Gamma_N(.)$ is the nonparametric Nadaraya-Watson kernel regression estimator of $G(.)$:

$$\Gamma_N(v_i) = \sum_{j \neq i}^{N} y_j K[(v_i - v_j)/h_N] / \sum_{j \neq i}^{N} K[(v_i - v_j)/h_N], \quad (5)$$

with $v_i = v(x_i'\beta)$. The bandwidth parameter h_N is defined as a nonstochastic window satisfying the following conditions: (*i*) $N^{-1/6} < h_N < N^{-1/8}$ and (*ii*) $\int v^2 K(v) dv = 0$. Under mild regularity conditions, it is shown that the quasi maximum likelihood (QML) estimator of the binary choice model has the properties of a consistent and asymptotically normally distributed estimator that attains the semiparametric efficiency bound [9].

A formal procedure for testing the results of alternative methods of the binary choice model estimation, i.e. the parametric probit model versus the semi-parametric alternative, is provided in [7]. The resulted test statistic is defined as follows:

$$T_N = \sqrt{h} \sum_{i=1}^{N} w[v(x_i, \hat{\beta}_N)]\{y_i - F[v(x_i, \hat{\beta}_N)]\} \times \{\hat{F}_{Ni}[v(x_i, \hat{\beta}_N)] - F[v(x_i, \hat{\beta}_N)]\}, \quad (6)$$

where $\hat{\beta}_N$ is the consistent estimate of β, $\hat{F}_{Ni}[v(x_i, \hat{\beta}_N)]$ is the kernel nonparametric estimator, while *h* is an optimal bandwidth from the kernel nonparametric regression. The weight function $w(.)$ is suggested to be 1 for the interval 95%-99% of $v(x_i, \hat{\beta}_N)$, *i*=1,…,*N*, and 0 otherwise. Under the null hypothesis T_N has asymptotically normal distribution with parameters zero and σ_T^2. A similar test with an empirical application is provided by [6].

3 DATA AND ESTIMATION RESULTS

3.1 Data

The data used in this paper come from the 2002 Living Standard Measurement Survey that was carried out by the Strategic Marketing and Media Research Institute and provided by the National Statistics Bureau. This survey contains data about labour market activities of the household members, but also some additional information, such as family composition and sources of households' incomes, that are not provided by the standard labour force surveys. For the purpose of the analysis presented in this paper a subsample of young married women aged 18-30 years is selected. Those young women who are engaged in any kind of self-employment or household activities are exempt from the analysis. The sample contains young married women who finished their education and who are capable of work. A total of 543 young married women is examined, out of which 236 are wage earners (as measured restrictively by positive working hours of those who work). The data are processed in [15].

3.2 Estimation results

Results of estimation of the parametric probit and semiparametric single-index models are presented in Table 1. The dependent variable is a binary choice variable that takes two values 1 and 0 representing young married women's decisions to participate in the labour force. A set of explanatory variables includes age (divided by 10), education in years, the number of small children of the preschool age in the family (zero to 6 years), percentage of married young women who live in urban areas, natural logarithm of the monthly husband's wage (divided by 1000), as well as natural logarithm of the monthly household income (excluding wages of employed members), including rents, remittances, social assistance and alike (divided by 1000).

Table 1: Estimation results for parametric probit and semiparametric single-index models

Variable	Parametric probit		Semiparametric single-index					
			Model 1 $h_N=0.35$		**Model 2** $h_N=0.38$		**Model 3** $h_N=0.41$	
	Coef.	**S.E.**	**Coef.**	**S.E.**	**Coef.**	**S.E.**	**Coef.**	**S.E.**
Age	1		1		1		1	
Education_y	0.0867[2]	0.0370	0.1231[1]	0.0404	0.1340[1]	0.0438	0.1453[1]	0.0489
Child_6years	-0.2634[2]	0.1047	-0.2943[2]	0.1157	-0.3139[2]	0.1270	-0.3342[2]	0.1400
Lnh_wage	0.0377[2]	0.0187	0.0532[3]	0.0287	0.0568[3]	0.0319	0.0601[3]	0.0358
Lnh_income	-0.0011	0.0732	0.0128	0.1030	0.0149	0.1092	0.0165	0.1161
P_urban	0.0070[2]	0.0030	0.0103[2]	0.0046	0.0108[2]	0.0049	0.0114[2]	0.0053
Intercept	-3.4241[1]	0.8123						
Log L	-333.92		-338.30		-338.57		-338.81	
LM test of normality $\chi^2(2)$	3.8397 (0.1466)							
Specif. test			**Test statistics**				**p-value**	
$T_{N,}\,h$=0.15			-0.0517				0.5164	
$T_{N,}\,h$=0.55			0.0399				0.4873	
$T_{N,}\,h$=2.00			0.1563				0.4506	
$T_{N,}\,h$=3.00			0.2316				0.4275	

Source: Author's calculation. p-values in brackets. ([1],[2],[3]) indicate statistical significance at the 1%, 5% and 10% levels, respectively.

The probit model is correctly specified as LM test reports. Heteroscedasticity corrected standard errors of the probit estimates are calculated by using the Huber-White sandwich estimator. The estimates and asymptotic standard errors of the semiparametric single-index model are calculated by using routines provided in [3]. The regression parameter of the continuous variable age (in years) with expected positive sign (0.86, p=0.00) is used as a normalization scale in the semiparametric approach; the intercept is excluded from the model. The same procedure is used in the estimation of the probit model to allow for comparisons with the single-index model parameters. Finally, a Gaussian kernel with different bandwidths is used for the QML estimator based upon the results of Monte Carlo simulations reported in [3] and the results of LM test for the normality assumption in the probit model. The probit estimates are used as initial values for semiparametric estimation. The results are reported for $h_N=0.35$, $h_N=0.38$ and $h_N=0.41$.

Comparison of the coefficient estimates of the two models underlies the differences of the methods used. Magnitudes and statistical significance of the coefficient estimates from the two models differ. Education and the presence of children of the preschool age are important for young married women's decision related to the participation in the labour force. Both coefficient estimates have interpretations that are in line with the economic theory of labour supply. The nonlinear term of age squared in years is exempt from the estimated participation equation due to the fact that only a subsample of young women is examined and that their propensity to work is expected to keep increasing up to a certain age. Husbands' log wages are significant factor of young women's decision to participate in the labour force, but, opposite to economic expectations, the coefficient estimate has a positive sign in the semi-parametric model, meaning that young married women are encouraged by the family members to take an active economic role in the society. Other sources of households' incomes are not important for their participation in the labour force, so that young married women from both wealthier and poorer families behave similarly. In general, all findings are in line with the assumption that young educated married women who live in urban areas, and

who are overrepresented in the sample, have more job opportunities than their counterparts in other areas.

The specification test [7] gives certain advantage to the parametric approach, indicating that the semiparametric single-index model is misspecified. However, several empirical studies confirmed better performances of the semiparametric estimator compared to a parametric alternative, but the results are obtained from the larger samples [7], [11]. Larger bandwidths are suggested in order to test the power of the specification test [11].

3.3 Discussion

The results show that education and the presence of children of the preschool age are important factors for young married women's participation in the labour force as it was expected. Comparison with the similar research for the sample of working age women in one of the European Union countries shows that husbands' wages are a significant limiting factor of the participation [11]. The research for Serbia indicates that husbands' wages positively contribute to the young women's participation in the labour force, while other sources of households' incomes are not important. The difference in these findings explains that the reservation wage is less valued by younger female participants in the labour market in Serbia.

These findings may have some practical implications to help us understand the frameworks of young women's behaviours regarding the choices between the participation in the labour force and some other possibilities, as for instance family planning or continuation of education. There are findings showing that women with children of the preschool age are encouraged by the family to be active participants in the labour force and that the role of women and men in rearing children is changing [2]. Institutionally, this is supported by existence of child care facilities and paid maternity leave. The level of education is a factor that strongly explains the young women's decisions to delay the birth of the first child in Serbia. The biggest difference is present between young women under thirty without primary education and with the university education [14]. The main problem of the labour market in Serbia is an insufficient dynamic of jobs creation for new entrants. Thereto, the position of young female participants is much worse than of their male counterparts [10], [13].

The estimated coefficients obtained by using two specifications of the labour force participation equations do not differ significantly. However, the standard errors for the coefficient estimates of the semiparametric single-index model are larger than those obtained for the estimates of the parametric probit model confirming that some efficiency loss occurs when the semiparametric approach is applied. The main difference between the results of the two approaches probably lies in the facts that the probit estimates are used as initial values for semiparametric estimation and that this estimation was done by using a small subsample of young married women, which was insufficient for satisfactory perturbation from initial values. This is the limitation of the research.

Due to the cumbersome procedure for calculating values of the specification tests given by [6] and [7] and certain arbitrariness in setting the bandwidth parameters for kernel estimators, one can choose the parametric approach if the models used are correctly specified. The proof about robustness of the estimated coefficients obtained by using two approaches requires further research on the extended samples encompassing a broader population of married women in Serbia and different time periods.

4 CONCLUDING REMARKS

This paper provides an analysis of young married women's decisions regarding the participation in the labour force in a transition economy. Two approaches are employed in

the estimation of econometric models, the standard parametric probit and the semiparametric single index models. Given that the participation in the labour force is a part of the overall analysis of the labour supply and that the coefficient estimates are obtained from models based upon different distributional assumptions, these models need to be compared in order to choose a reliable estimator that is consistent and asymptotically efficient. However, robustness of the coefficient estimates obtained by the two methods should be further tested due to a small subsample of young married women that is examined in this paper.

ACKNOWLEDGEMENT

The financial support of the Ministry of Education, Science and Technological Development of the Republic of Serbia through the research projects no. 47009 (*European integrations and social and economic changes in Serbian economy on the way to the EU*) and no. 179015 (*Challenges and prospects of structural changes in Serbia: Strategic directions for economic development and harmonization with EU requirements*) is gratefully acknowledged.

References

[1] Action Plan for Implementation of National Youth Strategy for the Period 2009-2014, Official Gazette of the Republic of Serbia, no. 7/09.

[2] Bobić, M., Vukelić, J., 2011, Second Demographic Transition De-blocked?, Sociologija, vol. 53, no. 2, pp. 149-176.

[3] De Luca, G., 2008, SNP and SML Estimation of Univariate and Bivariate Binary-Choice Models, Stata Journal, vol. 8, no. 2, pp. 190-220.

[4] Greene, W.H., 2012, Econometric Analysis: Seventh Edition, Upper Saddle River, New Jersey (USA), Prentice Hall International, Inc.

[5] Heckman, J.J., 1979, Sample Selection Bias as a Specification Error, Econometrica, vol. 47, no. 1, pp. 153-161.

[6] Horowitz, J.L., 1993, Semiparametric Estimation of a Work-Trip Mode Choice Model, Journal of Econometrics, vol. 58, no. 1-2, pp. 49-70.

[7] Horowitz, J.L., Härdle, W., 1994, Testing a Parametric Model Against a Semiparametric Alternative, Econometric Theory, vol. 10, no. 5, pp. 821-848.

[8] Jovičić, M., Dragutinović Mitrović, R., 2011, Econometric Methods and Models [Ekonometrijski metodi i modeli], Faculty of Economics, Belgrade, 332 p.

[9] Klein, R.W., Spady, R.H., 1993, An Efficient Semiparametric Estimator for Binary Response Models, Econometrica, vol. 61, no. 2, pp. 387-421.

[10] Krstić, G., Corbanese, V., 2009, In Search of More and Better Jobs for Young People of Serbia, ILO Subregional Office for Central and Eastern Europe, Policy Paper 2009/1.

[11] Martins, M.F.O., 2001, Parametric and Semiparametric Estimation of Sample Selection Models: An Empirical Application to the Female Labour Force in Portugal, Journal of Applied Econometrics, vol. 16, no. 1, pp. 23-39.

[12] National Youth Strategy, Official Gazette of the Republic of Serbia, no. 55/08.

[13] Ognjenović, K., 2012, Youth Employment Policies in Serbia: Framework, Interventions, Results, in: Zubović, J., Domazet, I. (eds.), New Challenges in Changing Labour Markets, Institute of Economic Sciences, Belgrade, pp. 59-74.

[14] Rašević, M., 2006, Odlaganje rađanja u optimalnoj dobi života – osnovna demografska cena 1990-ih u Srbiji, Demografski pregled, vol. 6, no. 21, pp. 1-4.

[15] StataCorp, 2009, Stata: Release 11, Statistical Software, College Station, Texas (USA), StataCorp LP.

ALGORITHMS OF ASSOCIATION AS A METHOD OF DATA MINING

Željko Račić and **Tamara Stražívuk**
Faculty of Economics, Banja Luka
zeljko.racic@efbl.org, bytamara@gmail.com

Abstract: Modern companies are more and more oriented towards the integration of their business activities and in general towards more comprehensive and complete overview of its business processes. Nevertheless, without the support of contemporary software applications and information – communication technology such process is not possible to convey. In order to succeed in business, modern companies have to direct right information to the right departments of the company at the right moment. Therefore, it is necessary to digitize up the processes in the organization and to make the organization "intelligent", and its human resources to the fullest extent, the workers of knowledge. Application of business intelligence and the use of its modern tools are necessary to obtain an advantage over the competition and stay at the market. For this purpose we need more skillful and sophisticated analyses of integrated data. Data mining and knowledge discovery in databases are new powerful technologies with a great potential to help companies to focus on the most important information in their databases. With proper use, high-quality data and necessary expertise, data mining definitely offers better solutions in marketing and decision making in business as well as in the optimization of technological processes and client services. System of business intelligence enables deep analyses of large amounts of data, and the possibility to observe the information from different views.

Key words: business intelligence, knowledge discovery in databases, data mining

1 INTRODUCTION

Information and communication technology is changing the ways in which people work and live, and is changing the organization and operation of modern enterprises. Those who fail to adapt to these changes - either individuals or businesses, will bring into question their existence and successful functioning in the new business and technological environment. Knowledge of some models and methods can fill a cup of prejudice, can keep us in one place, not allowing to look at the problem from another angle. Keynes (John Maynard Keynes) in 1936 defined the saying: The problem is not in the new ideas, but in severe abandonment of the old ones[1]! It is therefore necessary to know the possibilities offered by modern informational technologies, and the context and the business environment in which they operate in today's enterprises. It is, above all talked about the eternal present gap between the technology and business-oriented people, who so often have completely different visions of what constitutes an informational technology for one company and how to the full extent use its capabilities.

[1] Keynes, J.M: The General Theory of Employment Interest and Money, MacMillan&Co Ltd, London, 1964, pp 19.

1.1 Data mining and knowledge discovery

Data Mining and knowledge discovery in databases lately attracted considerable attention. Research in the field of data mining has all the characteristics of interdisciplinary research as it connects several disciplines, such as statistics, databases, artificial intelligence - pattern recognition, computer visualization, and others. The aim is to achieve a competitive advantage through the acquisition of profound knowledge that is stored in the large databases. Research in the field of databases, data warehouses, data mining and knowledge discovery in databases come with interesting solutions for the general population of computer users, IT professionals, managers of the business systems and other entities. Thus, for example, American Express achieved 10-15 percent increase in the use of credit cards in the United States as a result of the data mining and the use of the results to define marketing activities. In the process of data mining, it is necessary to take the following steps:

1. Sampling, i.e. taking part of the data, large enough to contain the necessary relevant information, and small enough for fast processing. If, for example, we have data on 20.000 consumers, in the search pattern can be selected only 15.000, and the remaining 5.000 can be used to assess individual models.
2. Exploration, i.e. the search for unanticipated trends and anomalies in order to improve the understanding of certain phenomena and others. During the research phase can for example by rotating the three-dimensional graph be discovered interesting properties of certain groups of consumers.
3. Modification, i.e. defining new variables, selection and transformation of variables for the model selection process. In the case of retail, in this phase, for example, is defined a new variable that divides consumers into low, medium and highly profitable customers.
4. Modeling is an automatic scanning of combination of data that reliably provide the desired result, for example, identifying the most profitable customer groups.
5. Assessment, i.e. evaluation of the usefulness and reliability of the results found in the data mining. In this phase, for example, are evaluated and compared some models, depending on how well they identify certain types of consumers.
6. Below is the interpretation of "excavated" forms, return to the steps 1-5 and visualization. In subsequent cycles, of the "digging" can, for example, be looked for connections between the identified consumer groups and sale in stores.
7. The final phase is carried out using the discovered knowledge in several forms:
   - Direct application in business
   - Turning knowledge into another system to take further action,
   -simple documenting and reporting the interested stakeholders.

The advantage over the competition, among others, is achieved through rapid response to market conditions, which can be achieved through faster and more flexible forms of recognition in the data that describes this situation. Depending on the kind of the goal, defined in the first phase of the entire process of knowledge discovery, we distinguish between two types of data search:

1. Verification (confirmatory analysis) of the pre-specified hypothesis, for example, "More than 80% of our sales were realized by consumers with young children."

2. Detection (research analysis), i.e. autonomously discovery of the new forms, such as, for example, "Identification of the most profitable customer groups."

Table 1 shows the four revolutionary steps that gave the opportunity of quick and precise answers that modern day business requires.

Table 1: Summary of four revolutionary step in data collection and processing

Period	Evolutionary steps	Business questions	Technology	Characteristics
1960.	Data collection	What is the total income in the last 5 years?	Computers, tapes, discs	Static delivery of historical data
1980.	Data access	How much was the sale of certain retail locations in the Banja Luka area in the past month?	Relational databases, SQL, ODBC	Dynamical delivery of historical data of one level
1990.	Data warehousing and decision support systems	How much was the sale of certain retail locations in the Banja Luka area in the past month? Explore (drill down) the locality of Banja Luka	OLAP, multidimensional database, data warehouse	Dynamical delivery of historical data with multiple levels
Today	Data mining	What can happen with the sale of the locality of Banja Luka in the next month? Why?	Advanced algorithms (Data Mining), multiprocessor computers, large databases	Predictable and proactive information delivery

1.2 Methods of data mining

Data mining methods are applied primarily in business. However, the data mining is applicable in other areas that have a large mass of data, based on which they want to disclose certain connections, regularities and legality (e.g. medicine, microbiology, genetics, mechanics, etc.).

There are a number of so-called the main and generally accepted methods, but also a whole range of methods from other fields that can not be assigned to any category. Given the function, data mining tools can be divided as follows [2]:

1. Classification - classifies data (entity) to one of several predefined classes (discriminant analysis, the method of branching, neural networks);
2. Regression - establishing the relationships with the help of predictor variables (linear and nonlinear regression, etc..)
3. Grouping - classification of data (entities) into one of several classes (clusters), where the class must be determined from the data - as opposed to the classification in which classes are pre-defined;

[2] Pyle, D., Pyle, D. 2003. Business Modelling and Data Mining, Morgan Kaufmann Publishers

4. compression, including visualization and exploratory data analysis;
5. modeling dependence (causal models, factor analysis);
6. associations (analysis of the consumer basket);
7. sequential analyzes (time series) and so on.

1.3 Algorithms of association 3

The results of the data processing with the help of algorithms of associations are associational rules. Associational rules indicate how often events occur together. Using transaction is possible to make a table that gives us the frequency of pairs (or larger number of elements) of certain elements in the transactions. From this table it is easy to make simple rules, such as:

R_1 = "Element 1 will appear along with the element 2 in 10% of all transactions",

where the 10% is a frequency measure (or measure of support) of appearing a pair of elements 1 and 2 in a set of transactions and presents the "importance", support or "significance" of rules.

If the frequency of occurrence of the element 1 in all transactions is 15%, and the element 2, 20%, then the ratio of the number of transactions in which both elements appear (i.e. the importance of the rules) according to the the number of transactions in which the element 1 is occurring (conditional part of the rules), we call a confidence of the rules. In this case, reliability of the rule R_1 is:

$$c(R_1) = \frac{10}{15} = 0{,}666.$$

It is easy to make an inverse rule:

R_2 =" Element 2 will appear along with the element 1 in 10% of all transactions ",

Although it is apparently the same rule, traits of R_1 and R_2 are different. Thus, the reliability of the rule is:

$$c(R_2) = \frac{10}{20} = 0{,}50.$$

Reliability of the Rule of 0.5 is equal to the claim that when in a transaction occurs the element 2, there is a 50% probability that the element 1 will also occur in the same transaction. At first glance it seems that the most reliable rules are those that are the best. The problem can occur when, for example element 1 occurs very frequently in transactions (for example, in 60% of transactions). In this case, the rule may have lower reliability than completely random choice. This shows that as a measure of good rules we need something a little better than the reliability. This measure is called the improvement, which tells us how a certain rule is better than random selection. Improvement was given with the following formula:

$$I(R_2) = \frac{p(conditions, consequences)}{p(conditions)\,p(consequences)}. \qquad (1)$$

[3] Data Mining Server, http://dms.irb.hr/

In our example,

$$I(R_2) = \frac{0,2}{0,2 \times 0,1} = 10,$$

while for the rule R_1, we have

$$I(R_1) = \frac{0,1}{0,1 \times 0,2} = 5.$$

When an improvement is more than 1, the rule is better than the random choice, when it is less than 1, it is worse. In our case R_2 is 10 times, and R_1 is 5 times better than the random selection. Generating the association rules is an iterative process. In essence it is very simple and comes down to a simple scheme:

1. Generate the table of frequency of occurrence of individual elements;
2. Generate the table of frequency of occurrence of two distinct elements; extract pairs from the table with the improvement of more than a pre-determined criteria;
3. Generate the table of frequency of occurrence of three distinct elements; aside from the table "triplets" with the improvement greater than a pre-determined criteria, and so on.

The best known algorithms to discover association rules are:

1. A priori algorithm;
2. tree of frequency samples;
3. method of the consumer basket..

The best known algorithm of association is **a priori algorithm**. The methodology of solving the problem by using a priori algorithm is performed in two stages. In the first phase, we find the frequency products or groups of products. In the second phase, based on the frequency product or group of products, we generate the association rules. The main drawback of this algorithm comes from its complexity and sensitivity to the growth of element analysis, which increases the number of combinations. There are reduction techniques for the analysis of a set of candidates, but despite that it does not completely solve the aforementioned problems. Popular methods to reduce the number of candidates entering the analysis using this algorithm are: a method of forming apparent variables, the method of grouping a set of products based on common characteristics, and the like. Because of all this, we are lead to the loss of precision of the analysis.

A very efficient algorithm for generating association rules is **a tree of frequency samples**. Passing through the base, with recorded transactions , it calculates the frequency of appearance of elements (e.g. product) contained in the database, and sorts them on the basis of frequency, ignoring non-frequency elements. After sorting elements according to the frequency of occurrence, and ignoring the low frequency elements, we access the construction of the tree of frequency samples.

Analysis of the consumer basket (often used as a synonym for the use of algorithms on the data from the Association of Retail), is a method of searching the data base, based on the discovery of association rules, which aims to discover the buying preferences of certain groups of products or product groups combined. Based on the findings resulting from the analysis, it is possible to give discounts, for example, on the product X if they buy the product Y (e.g. bread and milk), because they are usually bought together with increasing turnover of the ratio of goods, and therefore profit. Also, on the shelves can be put together

products for which the analysis has shown that are the most frequently purchased in pairs (e.g. healthy food) also working to increase the coefficient of the goods. Association algorithms can be used to form negational types: If the product X THEN NOT Y. Such an approach to the analysis provides the information about the reluctance to purchase in pairs.

Association algorithms can be used in the analysis of time series using a single model of transformation of time series (of which will be discussed later). In general, the association algorithms can be used in the implementation of customer profile analysis, aimed at identifying behavioral patterns and styles of purchase. Their use is determined by the defined objectives and analysis, with the possibility of creating creative analytical solutions.

Benefits of algorithm associations: association rules are simple and clear; method is intended for problems that are not predictive of classification type, i.e. there is no target attribute, it allows processing data in which we have a variable number of attributes; algorithms that generate association rules, are in principle, very simple.

2 CONCLUSION

The results obtained by the analysis of the data with processing association algorithms, are easy to interpret due to the fact that they show the frequency of occurrence of certain categories (product or group of commercial products) of buying in pairs. Based on the results of the data processing we can estimate the probability of the simultaneous purchase of the product in pairs or groups. To the association algorithms we can assign the time component, so that they can express intention of buying through a certain time period.

References

[1] Kenneth, C. Laudon, Jane P. Laudon. 2006. Management information systems: Managing the Digital firm, Pearson Prentice Hall.

[2] Keynes, J.M. 1964. The General Theory of Employment Interest and Money, MacMillan&Co Ltd, London.

[3] Liautaud, B. 2001. E-Business Intelligence: Turning Information into Knowledge into Profit, McGraw-Hill, New York.

[4] Pyle, D. 2003. Business Modelling and Data Mining, Morgan Kaufmann Publishers.

[5] Turban, E., McLean, E., Wetherbe, J. 2004. Information Technology for Management, John Wiley & Sons, Inc.

The 12th International Symposium on
Operational Research in Slovenia

SOR '13

Dolenjske Toplice, SLOVENIA
September 25 - 27, 2013

Section V:

Production and Inventory

STOCHASTIC QUEUING MODELS: A USEFUL TOOL FOR A CALL CENTRE PERFORMANCE OPTIMIZATION

Alenka Brezavšček and **Alenka Baggia**
University of Maribor, Faculty of Organizational Sciences
Kidričeva cesta 55a, SI-4000 Kranj, Slovenia
{alenka.brezavscek,alenka.baggia}@fov.uni-mb.si

Abstract:
Contemporary service oriented companies often offer call centre customer support. Since the call centre is usually the first contact of a customer with a company, the quality of a service and efficient performance of the call centre is of key importance to the company. An essential factor of the call centre efficiency is the optimal number of operators answering the customer's calls. Results of previous research show that stochastic queuing models can be used to analyse the efficiency of call centres. In the presented paper, a research was conducted on the case of Slovenian telecommunication provider's call centre, to demonstrate the usefulness of stochastic queuing models in optimization of call centre performance.

Keywords: call centre, stochastic queuing model, efficiency, optimization.

1 INTRODUCTION

The queuing theory is a special application of stochastic process theory (see e.g. [6], [9], [16], [17]). Its applications can be found in diverse areas: a) telecommunications [2], [8]; b) computer networks traffic studies [15]; d) road traffic studies [11], [14]; e) and others.

Many contributions in the literature prove that stochastic queuing models are also an useful and applicable tool to analyse the efficiency of call centres that are frequently used by organizations for marketing purposes and technical support for end users (see e.g. [4], [5], [7], [12], [13]). Relevance and usefulness of the results obtained with such an analysis depends on selection of an appropriate mathematical model, which is based on knowledge of the probability density functions of inter-arrival times (i.e. times between two successive incoming calls) and service times (i.e. call length), that are both random variables. These mathematical characteristics of the call centre can be obtained if accurate and complete data about the call centre operation are available. Contemporary technology enables automatic logging of all the events in the call centre, so data needed for the mathematical analysis of the call centre are usually available. However, lack of expert knowledge in practice prevents the companies from efficient usage of them for the call centre optimization. The number of operators in a shift is often based on the rule-of-thumb decision, and is frequently not an optimal solution.

Continuing prior research [3] conducted on the case of Slovenian telecommunication provider's call centre, we present a relatively simple usage of stochastic queuing models for planning an appropriate number of call centre operators.

2 METHODOLOGY

A typical queuing system comprises of one or more service units (i.e. servers), arrivals of customers demanding the service, and the service process. Whenever all the customers can not be served at once, queues are formed. This leads to costs (losses) due to waiting which increase with the number of customers in the queue. To decrease the waiting costs and raise the service level ensuring better system performance different improvements can be implemented. However, any improvement often comes with a certain investment leading to higher costs of the queuing system operation. Figure 1 shows that it is always possible to determine the optimal service level which ensures the total costs of queuing system performance are minimal.

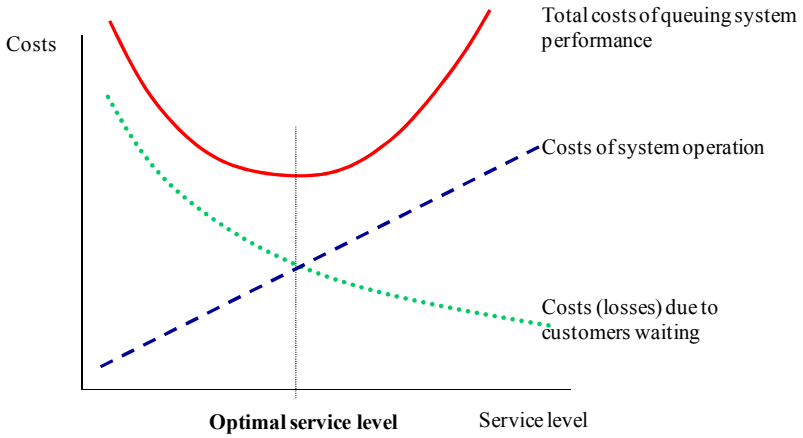


Figure 1: Costs of queuing system performance

To determine the optimal service level of a queuing system, different quantitative characteristics can be used. The values of these characteristics can be calculated using an appropriate mathematical model. Proper selection of the mathematical model is based on the following elements of the queuing system:

- *Arrival process:* Population of customers can be considered either limited (closed systems) or unlimited (open systems). Most mathematical models presume individual arrivals of customers and independent identically distributed inter-arrival times.
- *Service mechanism* is determined with the system capability, availability and probability density function of service times. Most of the mathematical models assume that service times are independent identically distributed random variables.
- *Queuing discipline* represents the way the queue is organised (e.g. First-In-First-Out (FIFO), Last-In-First-Out (LIFO), random selection of customers or selection based on customer priorities).

When there is only one sever, or there are a number of equivalent and parallel servers, the queuing system is called simple. Simple queuing models use the standard notation for describing the probability density function of inter-arrivals and service times:

M – a Poission process of the number of events (i.e. customer arrivals or end of services); exponential density function of times between two successive events.

G – a general distribution of times between two successive events (with a known mean and variance; e.g. normal density function).

D – a deterministic situation; times between two successive events are constant.

Notation *M/M/r* {infinity/infinity/FIFO} therefore describes the queuing system with *r* parallel servers, unlimited population, unlimited queue, FIFO queuing discipline, while both, inter-arrival and service times are distributed according to exponential density function (see e.g. [16]). For many types of simple queuing systems analytical solutions are available.

3 CALL CENTRE AS A QUEUING SYSTEM

The presented research was conducted on the case of a Slovenian telecommunication provider's call centre. The call centre is opened from 8:00AM till 12:00PM. It employs 8 full time operators while additional contractors are hired when needed. The schedule of operators is defined based on prior experiences. No analysis of the schedule has been performed. Customers are calling a single phone number. If at least one of the operators is available at the time of the call, he answers the call and serves the customer. If all of the operators are

busy the calling customer is not rejected but can wait for a free operator regardless the number of customers in the queue. The principal scheme of the call centre is presented in Figure 2.

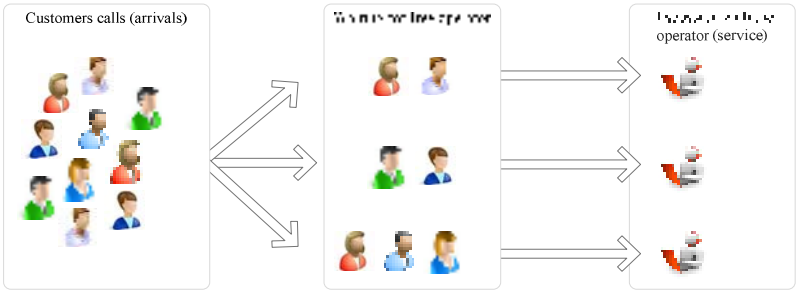


Figure 2: Call centre as a queuing system

The call centre under consideration can be treated as a simple queuing system, where the number of servers is determined with the number of active operators and the queuing discipline is FIFO.

The key element of the call centre efficiency analysis is the determination of the optimal number of active operators for different periods of the day. The number of calls in different time periods of the day on a typical working week was analysed. The working day of the call centre was divided into the following four periods: from 8:00 AM to 10:00 AM, from 10:00 AM to 1:00 PM, from 1:00 PM to 6:00 PM and from 6:00 PM to 12:00 PM. The results of analysis are presented in Figure 3.

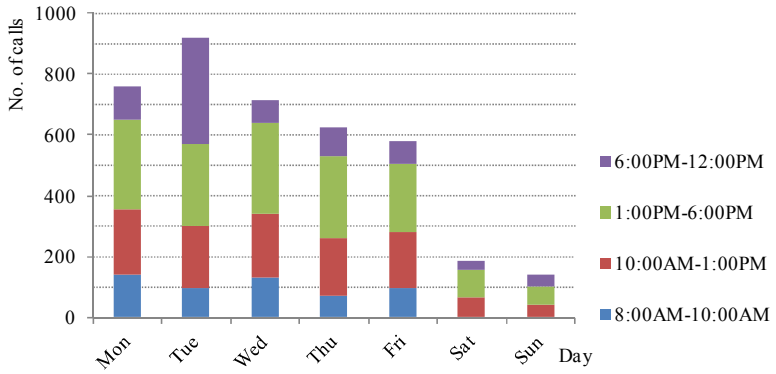


Figure 3: Number of calls in a typical working week by the time periods of the day.

The number of incoming calls is significantly lower during the weekend than during the working days. Therefore, the weekend days were excluded from further analysis. The number of calls in the particular period is similar for all working days. The exception is the last period on Tuesday. This deviation was caused by the unexpected downtime of one of the main provider's services. The third period (1:00PM to 6:00PM) is the most frequent period, while the morning and evening periods are less burdened.

3.1 Queuing model selection

To select the appropriate mathematical model, the distribution of inter-arrival times and the distribution of service times have to be analysed. Figure 4 shows the frequency distribution of inter-arrival times while Figure 5 shows the frequency distribution of service times in different periods of a typical working day.

We can conclude from Figure 4 that inter-arrivals times in all four periods fit the exponential density function. It can be seen from Figure 5 that probability density function of service time follows an asymmetric function (e.g. lognormal density function). When calls

shorter than one minute are omitted, we can assume that also the distribution of service times can be described by the exponential density function in all four periods of a working day. Since these short calls do not cause queues and therefore do not threaten the efficiency of the call centre, our assumption is justified.

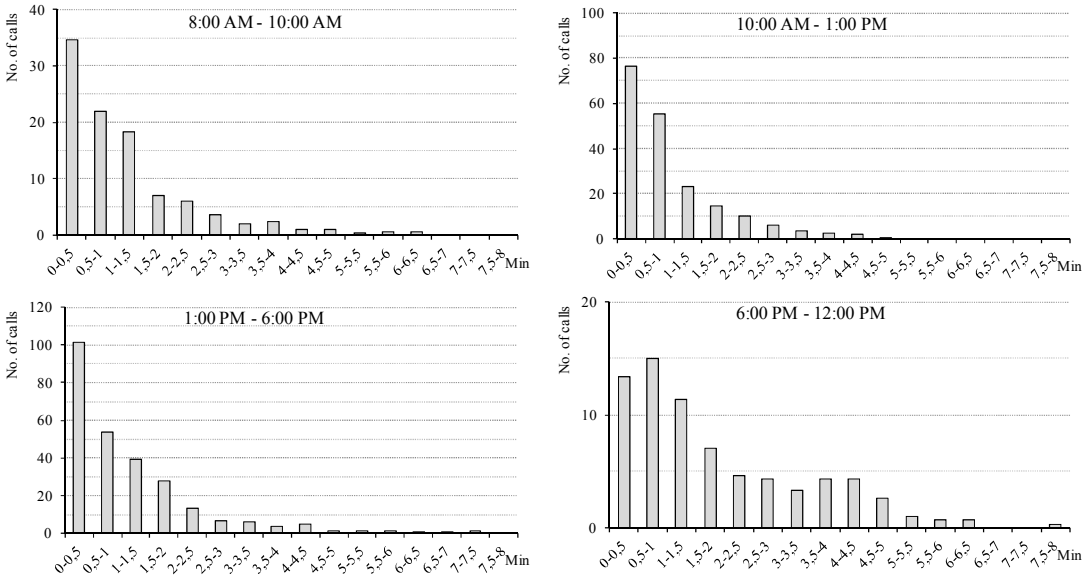


Figure 4: Frequency distribution of inter-arrival times in different periods of a typical working day.

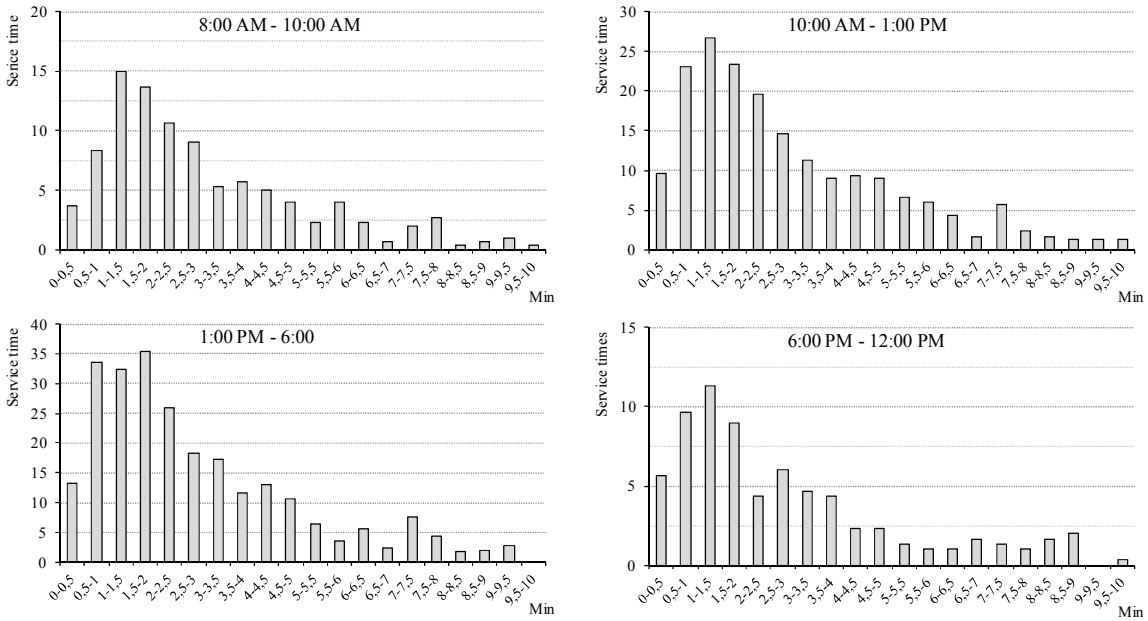


Figure 5: Frequency distribution of service times in different periods of a typical working day.

From the description of the call centre organization and from the analysis of arrival and service patterns we can conclude that the call centre under consideration can be described by the *M/M/r* {infinity/infinity/FIFO} queuing model.

3.2 Selection of optimization criterion

In the process of the call centre optimization, the expected waiting time was selected as a key performance criterion. Assuming the *M*/*M*/*r* {infinity/infinity/FIFO} queuing model the expected waiting time can be calculated according to the following equation (see e.g. [10]):

$$E(W_q) = \frac{1}{S} \frac{(r\rho)^r}{r!(1-\rho)^2 r\sigma} \tag{1}$$

The symbols in (1) denote:

r – the number of servers

α – the arrival rate; $1/\alpha$ is the expected time between two successive arrivals

σ – the service rate; $1/\sigma$ is the expected service time

ρ – the traffic intensity calculated as $\rho = \frac{\alpha}{r\sigma}$

S – the sum which can be calculated as follows:

$$S = 1 + r\rho + \frac{(r\rho)^2}{2!} + \cdots + \frac{(r\rho)^{r-1}}{(r-1)!} + \frac{(r\rho)^r}{r!} \cdot \frac{1}{1-\rho} \tag{2}$$

Equation (1) makes sense when S < ∞. This condition is met when ρ < 1. The condition ρ < 1 ensures that the steady state distribution exists. In such a case the infinite queues are not formed and the queuing system still operates after a long run.

The minimum number of servers $r_{\min}$ needed to satisfy the steady state condition is the lowest integer that fulfil the equation

$$r > \frac{\alpha}{\sigma} \tag{3}$$

3.3 Call centre optimisation

To optimize the call centre performance we set up the following requirement: *the expected waiting time in a particular time period should not be longer than 20 seconds*. The mathematical formulation of our requirement is: $E(W_q) \leq 20\,\text{sec} = 0,33\,\text{min}$.

Results of our analysis for a particular period of a working day are given in Table 1. From the field data the parameters α and σ were estimated (columns 1 and 2). The value $r_{\min}$ needed to satisfy the steady state condition was determined according to (3) and corresponding $E(W_q)$ was calculated using (1) and (2) (columns 4 and 5). The minimal number of servers r needed to fulfil our performance requirement was determined by iteration using (1) and (2) (columns 6 and 7).

Table 1: Results of call centre optimization.

Period	α [min^{-1}] (1)	σ [min^{-1}] (2)	α/σ (3)	$r_{\min}$ (4)	$E(W_q)$ [min] (5)	minimal r (6)	$E(W_q)$ [min] (7)
8:00AM - 10:00AM	0,847	0,336	2,25	3	4,34	5	0,16
10:00AM - 1:00PM	1,053	0,342	3,08	4	1,72	6	0,11
1:00PM - 6:00PM	0,877	0,356	2,46	3	3,57	5	0,14
6:00PM - 12:00PM	0,532	0,356	1,49	2	3,55	4	0,08

4 CONCLUSION

The usefulness of stochastic queuing models is demonstrated in the case of Slovenian telecommunication provider's call centre. We analysed the arrival and service patterns, and established that the call centre under consideration can be described by the *M*/*M*/*r* {infinity/infinity/FIFO} queuing model. The expected waiting time was selected as a key performance criterion in the process of the call centre optimization. The aim was to

determine the minimal number of servers in a particular period of a working day to ensure the expected waiting time should not exceed 20 seconds. This requirement will be fulfilled when in the period 10:00AM - 1:00PM at least six operators are available. In the periods 8:00AM - 10:00AM and 1:00PM - 6:00PM we need at least five operators, while in the last period (6:00PM - 12:00PM) four operators are enough.

Results obtained prove that stochastic queuing models represent an applicable and useful tool for a call centre performance optimization. Such models enable rather easily determination of an appropriate number of active operators regarding a specific key performance criterion. This is a preliminary condition to ensure the optimal service level and therefore minimal cost of queuing system performance.

Discrete event simulation is also a viable option for accurate performance modelling and subsequent decision support. Some authors (e.g. [1]) argue that, the analytical approach is not accurate enough, as it does not mimic randomness. In future research we will simulate the presented case with a discrete event simulation tool, where for describing the probability density function of service times an asymmetric function will be used.

References

[1] S. Akhtar and M. Latif, "Exploiting Simulation for Call Centre Optimization," in *Proceedings of the World Congress on Engineering*, London, 2010.

[2] A. Attahirusule, Queueing Theory for Telecommunications: Discrete Time Modelling of a Single Node System, Nev York: Springer, 2010.

[3] A. Brezavšček, A. Saje and J. Šumi, "Analiza učinkovitosti klicnega centra za uporabo stohastičnih modelov množične strežbe [Call centre efficiency analysis using stohastic queueing models]," in *Dnevi slovenske informatike*, Portorož, 2012.

[4] L. Brown, N. Gans, A. Mandelbaum, A. Sakov, S. Haipeng, S. Zeltyn and L. Zhao, "Statistical Analysis of a Telephone Call Center: A Queueing-Science Perspective," *Journal of the American Statistical Association,* vol. 100, no. 469, pp. 36-50, 2005.

[5] E. Chassioti, Queueing Models for Call Centres (PhD thesis), Lancaster: Lancaster University Management School, 2005.

[6] N. M. van Dijk and R. J. Bouchiere, Queuing Networks: A Fundamental Approach, New York: Springer, 2011.

[7] C. Dombacher, "Queueing Models for Call Centres," 2010. [Online]. Available: http://www.telecomm.at/documents/Queueing_Models_CC.pdf. [Accessed 12 june 2013].

[8] G. Giambene, Queuing theory and telecommunicaitons: networks and application, New York: Springer, 2005.

[9] D. Gross, J. F. Shorttle, J. M. Thompson and C. M. Harris, Fundamentals of Queuing Theory, Hoboken: Wiley Series in Probability and Statistics, 2008.

[10] A. Hudoklin, Stohastični procesi [Stohastic processes], Kranj: Moderna Organizacija, 2003.

[11] I. A. Ismail, G. S. Mokaddis, S. A. Metwally and M. K. Metry, "Optimal Treatment of Queuing Model for Highway," *Journal of Computations & Modelling,* vol. 1, pp. 61-71, 2011.

[12] G. Koole, "Call Center Mathematic: A scientific method for understanding and improving contact centers," 2007. [Online]. Available: http://www.academia.edu/542467/Call_center_mathematics. [Accessed 12 June 2013].

[13] G. Koole and A. Mandelbaum, "Queueing Models of Call Centers: An Introduction," 2001. [Online]. Available: http://www.columbia.edu/~ww2040/cc_review.pdf. [Accessed 17 June 2013].

[14] T. Raheja, "Modelling traffic congestion using queueing networks," *Sādhāna,* vol. 4, pp. 427-431, 2010.

[15] T. G. Robertazzi, Computer networks and systems: queueing theory and performance evaluation, New York: Springer, 2000.

[16] M. Tanner, Practical Queuing Analysis, London: The IBM McGraw-Hill Series, 1995.

[17] H. C. Tijms, A First Course in Stohastic Models, Southern Gate: Wiley, 2003.

DUAL SOURCING INVENTORY MODEL WITH AN UNRELIABLE SUPPLIER

Marko Jakšič
University of Ljubljana, Faculty of Economics, Kardeljeva ploščad 17, Ljubljana, Slovenia
School of Industrial Engineering, Eindhoven University of Technology, The Netherlands
marko.jaksic@ef.uni-lj.si

Abstract: We model a periodic review, single stage inventory system with non-stationary stochastic demand, where replenishment can occur either through a regular stochastically capacitated supply channel and/or an alternative uncapacitated supply channel with a longer fixed lead time. We derive the optimal dynamic programming formulation and we show some of the properties of the optimal policy by carrying out a numerical analysis. Additionally, our numerical results reveal several managerial insights.

Keywords: inventory, dual sourcing, stochastic models, dynamic programming, uncertain capacity

1 INTRODUCTION

Lead time reduction is one of the main goals when one wants to pursue a concept of a lean and agile modern supply chain. However, many companies that have actively embarked on the projects related to reducing the lead times were, at least in a short run, faced by the fact that their customer service performance suffered. This has forced the customers to search for alternative supply channels, through which they would improve the supply process reliability. In this paper we model the problem of a customer that is in a contractual supply agreement with a regular supplier that is working under make-to-stock principle, which results in a fast response to customer's ordering decisions and immediate delivery of the products available on stock. However, the supply availability is limited due to the supplier's on-hand stock availability. When the customer anticipates the supply shortage, he can rely on an alternative supplier, whose lead time is longer, but he is able to deliver the entire order with certainty. Such replenishment policy could be attributed to the supplier working under make-to-order principle.

While most of the multiple supplier research explores the trade-off between purchasing costs and indirect costs of holding safety inventory to cover against demand and supply variability, our focus lies more in further elaboration of the determinants of supplier's service. More precisely, we study the effect of capacity and lead time on supply reliability and the order allocation decision to suppliers, that is the decision between unreliable capacitated supplier with short lead time and reliable infinite capacity supplier with longer lead time.

We proceed with a review of the relevant literature on supply uncertainty models in a single-stage setting, where our interest lies in two research tracks, inventory models with random capacity and dual-sourcing models with unreliable suppliers. The way we model the supply availability is in line with the work of [1,2,3,4], where the random supply/production capacity determines a random upper bound on the supply availability in each period. Their research is mainly focused on establishing the structure of the optimal policy. For a finite horizon stationary inventory model they show that the optimal policy is of order-up-to type, where the optimal base-stock level is increased to account for possible, albeit uncertain, capacity shortfalls in future periods. A general assumption in capacitated inventory models is that the part of the order above the available supply capacity in a certain period is lost to the customer.

For a general review of the multiple supplier inventory models we refer the interested reader to [5]. The review is based on the important criteria for the supplier choice, mainly price and supplier service. A more focused review on multiple sourcing inventory models when supply components are uncertain by [6] reveals that most of these models consider uncertainty in supply lead time, supply yield, or supplier availability. In a deterministic lead time setting, several papers discuss the setting in which the lead times of the two suppliers differ by a fixed number of periods [7,8]. However, they all assume infinite supply capacity or at most a fixed capacity limit on one or both suppliers. For an identical lead time situation as ours, albeit uncapacitated, [9] derives the optimal inventory policies and parameters. In [10] a regular supply mode is governed by a base stock policy, whereas the faster capacitated emergency mode can be used in order to avoid stockouts.

However, when there is uncertainty in the supplier capacity, diversification through multiple sourcing has received very little attention. The exception to this are the papers by [11,12], where they study a single period problem with multiple capacitated suppliers and develop the optimal policy to assign orders to each supplier. Also, all of the capacitated multiple sourcing papers cited above assume identical lead time suppliers. Our paper makes a contribution to the literature by introducing a dual sourcing inventory model with a capacitated unreliable supplier and a reliable supplier with longer lead time.

The remainder of the paper is organized as follows. We present the model formulation in Section 2. In Section 3, we present the results of a numerical study to characterize the properties of the optimal policy, quantify the value of dual sourcing and provide the relevant managerial insights. Finally, we summarize our findings and suggest the possible extensions in Section 4.

2 MODEL DEFINITION

In this section, we give the notation and the model description. A regular, zero lead time, supply channel is capacitated, where the supply capacity is exogenous to the customer and the actual capacity realization is only revealed upon replenishment. An alternative supply channel is modeled as an uncapacitated with a fixed one period lead time. The demand and supply capacity are assumed to be stochastic non-stationary with known distributions in each time period, however, independent from period to period. In each period the customer places the order either to a regular, or to an alternative supplier, or both.

Table 1: Summary of notation.

T	:	number of periods in the finite planning horizon
c_h	:	inventory holding cost per unit per period
c_b	:	backorder cost per unit per period
α	:	discount factor ($0 \leq \alpha \leq 1$)
x_t	:	inventory position before ordering in period t
y_t	:	inventory position after ordering in period t
z_t	:	regular order size in period t
v_t	:	alternative order size in period t
d_t, D_t	:	actual realization and random variable denoting demand in period t
q_t, Q_t	:	actual realization and random variable denoting the available supply capacity of the regular supply channel in period t

Presuming that unmet demand is fully backordered, the goal is to find an optimal policy that would minimize the inventory holding costs and backorder costs over a finite planning horizon T. We intentionally do not consider any product unit price difference and

fixed ordering costs as we are primarily interested into the trade-off between capacity uncertainty associated with regular ordering and the delay in the replenishment of an alternative order. Any fixed costs would make the dual sourcing option less favorable, and the difference in the fixed costs related to any of the two ordering channels would result in a relative preference of one channel over the other. We give the relevant notation in Tab. 1.

We assume the following sequence of events. (1) At the start of the period t, the manager reviews inventory position before ordering x_t, where $x_t = \hat{x}_t + v_{t-1}$ is a sum of the on-hand stock and the delayed order from the previous period. (2) The regular order z_t and the delayed order v_t are placed and the inventory position is raised to inventory position after ordering y_t, $y_t = x_t + z_t + v_t$. (3) The delayed order from the previous period and the current period's order are replenished, and the inventory position is corrected according to the capacity availability $y_t - [z_t - q_t]^+ = x_t + \min(z_t, q_t) + v_t$. (4) At the end of the period, demand d_t is observed and satisfied through on-hand inventory; otherwise it is backordered. Inventory holding and backorder costs are incurred based on the end-of-period on-hand inventory, $\hat{x}_{t+1} = y_t - [z_t - q_t]^+ - v_t - d_t$. Correspondingly, the expected single period cost function $C_t(y_t, z_t) = \alpha E_{Q_t, D_t} \tilde{C}_t(\hat{x}_{t+1})$, where $\tilde{C}_t(\hat{x}_{t+1}) = c_h[\hat{x}_{t+1}]^+ + c_b[\hat{x}_{t+1}]^-$ is the regular loss function.

Correspondingly, the minimal discounted expected cost function that optimizes the cost over a finite planning horizon *T* from period *t* onward, starting in the initial state x_t, can be written as:

$$f_t(x_t) = \min_{z_t \geq 0, v_t \geq 0} \left\{ C_t(y_t - [z_t - Q_t]^+ - v_t - D_t) + \alpha E_{Q_t, D_t} f_{t+1}(y_t - [z_t - Q_t]^+ - D_t) \right\}, 1 \leq t \leq T \qquad (1)$$

and the ending condition is defined as $f_{T+1}(\cdot) \equiv 0$.

3 NUMERICAL RESULTS

In this section we present the results of the numerical analysis, which was carried out to characterize the properties of the optimal ordering policy, and to gain insights on the value of dual sourcing compared to sourcing from a single supplier. Numerical calculations were done by solving the dynamic programming formulation given in Eq. (1). We used the following set of input parameters: T=12, c_h=1, c_b=20, α=0.99, and a discrete uniform distribution to model stochastic demand and supply capacity. Throughout the experiments we have varied the utilization of the regular supply channel, $Util = (0, 0.5, 1, 2, \infty)$, defined as the ratio of available capacity at the unreliable supplier over the expected demand, and the coefficient of variation of demand, $CV_D = (0, 0.14, 0.37, 0.61)$, and supply capacity, $CV_Q = (0, 0.14, 0.37, 0.61)$.

To study the structure of the optimal policy, we have determined the optimal order sizes for orders made at a regular and an alternative supplier depending on the initial inventory position x_t. Additionally, in Fig. 1, we depict the inventory positions after ordering, given that either a regular order or alternative order is placed, or both. Looking at the x_t+z_t line, we observe that the ordering takes place in the manner of the base stock policy. For $x_t \leq y_t^z$ we order up to a base stock level y_t^z, while for $x_t > y_t^z$ no regular order is placed. Similarly, looking at x_t+v_t line, an alternative order is placed only if $x_t \leq y_t^v$. For

$y_t^z \le x_t < y_t^v$ the inventory position is increased to a base stock level y_t^v, while for $x_t < y_t^z$ the size of an alternative order depends on the anticipated supply shortage of a regular order. Apart from the effect of the shortage anticipation on the size of an alternative order, this result corresponds to the papers on Dual index policies [7,8].

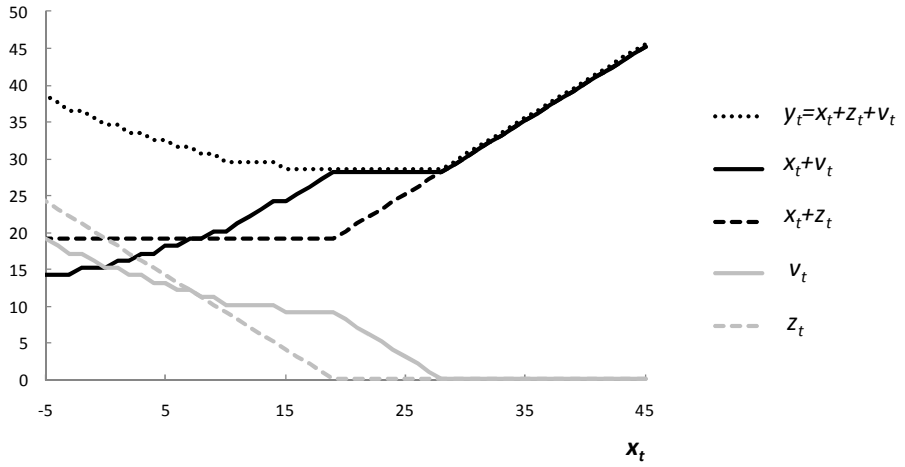


Figure 1: The optimal inventory position after ordering and the optimal regular and alternative order sizes.

The benefits of dual sourcing were addressed in two ways: first, the value of dual sourcing is quantified in relationship to sourcing only from the unreliable supplier, and secondly, we present the performance comparison of the dual sourcing to the two situations in which the two single sourcing options are optimal,

We define the *relative value* of dual sourcing, as a relative cost savings over the setting in which a reliable supply channel cannot be used ($v = 0$):

$$\%V_{DS} = \frac{f_t\{z\ge 0, v=0\} - f_t\{z\ge 0, v\ge 0\}}{f_t\{z\ge 0, v=0\}}. \tag{2}$$

Correspondingly, we also define the *absolute value* of dual sourcing:

$$\Delta V_{DS} = f_t\{z\ge 0, v=0\} - f_t\{z\ge 0, v\ge 0\}. \tag{3}$$

We present the results on the relative and absolute value of dual sourcing in Tab. 2. The utilization of the regular supply channel largely influences the extent of the savings that can be achieved through dual sourcing. The higher the utilization, the more benefits we obtain from sourcing from an alternative supplier. For overutilized system, $\%V_{DS}$ approaches 100%, while for reasonable utilizations the relative savings are still ranging from 20% to over 50%. Observe that when there is no demand uncertainty the system can be managed with zero cost solely through a reliable alternative supply channel. It is expected that the increase in the demand uncertainty (CV_D) will decrease the relative value of dual sourcing as demand variations prohibit the exact targeting of the optimal inventory level. When capacity uncertainty (CV_Q) increases, the probability of supply shortages at an unreliable supplier increases, therefore ΔV_{DS} also increases. However, such monotonic behavior does not hold for $\%V_{DS}$.

Table 2: The value of dual sourcing.

			$\%V_{DS}$				ΔV_{DS}			
Util	CV_Q	CV_D	0.00	0.14	0.37	0.61	0.00	0.14	0.37	0.61
∞			100.0	93.8	85.3	79.1	638.5	633.1	652.3	692.6
2	0.00		100.0	93.6	81.6	73.0	319.2	317.7	342.6	387.4
2	0.14		100.0	93.8	81.5	72.9	326.8	319.3	343.7	388.3
2	0.37		100.0	93.4	80.9	72.7	341.5	332.3	351.4	393.4
2	0.61		100.0	92.4	80.4	72.5	361.9	352.9	365.5	403.2
1	0.00		-	66.8	62.0	56.5	0.0	36.1	92.1	147.4
1	0.14		100.0	75.7	63.3	56.8	45.7	50.1	97.8	150.7
1	0.37		100.0	87.2	68.5	59.1	114.9	109.8	131.5	172.4
1	0.61		100.0	88.8	71.9	61.9	183.5	178.1	183.8	211.2
0.5	0.00		-	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.5	0.14		-	0.0	0.0	0.2	0.0	0.0	0.0	0.2
0.5	0.37		100.0	37.1	20.5	16.5	2.5	2.9	5.9	14.0
0.5	0.61		100.0	81.4	54.7	40.7	54.0	51.6	51.6	60.4

Next, we compare the performance of the dual sourcing model to two base cases: the case in which it is optimal to supply only from a regular "unreliable" supplier (*Util* = 0), and the case in which only reliable, longer lead time, supplier is optimal (*Util* = ∞). The results are presented in Fig. 2.

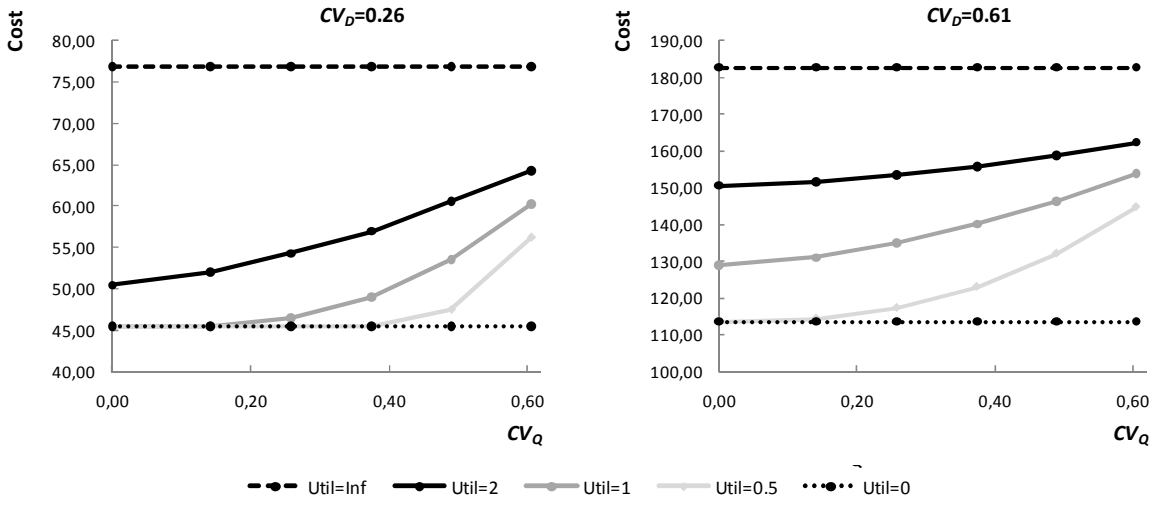


Figure 2: System costs for different regular supply channel utilizations.

The *Util* = 0 case represents the setting in which a regular supplier is not constrained by the capacity shortage, and becomes a fully reliable supplier. In this setting the system works with the lowest cost solely through the regular, zero lead time, supply channel. As the utilization of the regular supply channel increases the costs are increasing and are approaching the setting in which there is no available capacity at the regular supply channel (*Util* = ∞), and the supply is only done through the alternative supplier. The rate of the increase in cost depends also on the demand and capacity uncertainty, where the costs are more sensitive to the capacity uncertainty for low demand uncertainty values.

4 CONCLUSIONS

In this paper, we study a periodic review dual sourcing inventory model under stochastic demand and limited supply capacity, limiting the supply performance of a regular supply channel. There is an alternative, reliable, however longer lead time, supply channel, which effectively reduces the inventory costs incurred due to the capacity shortages of a regular supplier. Based on the model definition and inventory recursions we develop the dynamic programming formulation and proceed with a numerical analysis to characterize the structure of the optimal policy and provide insights into the value of dual sourcing. The optimal policy is characterized by the two base stock levels to which we increase the initial inventory position by placing a regular and/or an alternative order, however the policy is complicated by the fact that when both orders are placed, an alternative order size has to respond to anticipated supply shortage of a regular order. We characterize the setting in which the value of dual sourcing is the highest as that of a highly utilized regular supply channel, low demand uncertainty and generally high capacity uncertainty. A natural extension of this work would be to derive the structure of the optimal policy analytically by characterizing the optimal inventory positions to which the regular and alternative orders are placed, and by showing some additional properties of the optimal policy.

References

[1] Ciarallo, F. W., Akella R. and Morton T. E., 1994. A periodic review, production planning model with uncertain capacity and uncertain demand - optimality of extended myopic policies. Management Science, 40, pp. 320–332.

[2] Khang, D. B., O. Fujiwara. 2000. Optimality of myopic ordering policies for inventory model with stochastic supply. Operations Research, 48, pp. 181–184.

[3] Iida, T. 2002. A non-stationary periodic review production-inventory model with uncertain production capacity and uncertain demand. Eur. J. of Operational Research, 140, pp. 670–683.

[4] Jakšič, M., Fransoo J.C., Tan T., de Kok A. G., Rusjan B., 2011. Inventory management with advance capacity information. Naval Research Logistics, 58, pp. 355-369.

[5] Minner, S. 2003. Multiple-supplier inventory models in supply chain management: A review. Int. J. Production Economics, 81, pp. 265–279.

[6] Tajbakhsh, M. M., S. Zolfaghari, C. Lee. 2007. Supply uncertainty and diversification: A review. H. Jung, F. Chen, B. Seong, Eds. In Trends in Supply Chain Design and Management: Technologies and Methodologies. Springer Ltd., London, England.

[7] Fukuda. Y., 1964. Optimal policies for the inventory problem with negotiable leadtime. Management Science, 10 (4), pp. 690-708.

[8] Veeraraghavan, S., Sheller-Wolf A., Now or Later: A simple policy for effective dual sourcing in capacitated systems, Operations Research, 56 (4), pp. 850-864.

[9] Bulinskaya, E., 1964. Some results concerning optimal inventory policies. Theory Prob. Appli., 9, pp. 502-507.

[10] Vlachos, D., G. Tagaras. 2001. An inventory system with two supply modes and capacity constraints. International Journal Production Economics, 72, pp. 41–58..

[11] Dada, M., N. C. Petruzzi, L. B. Schwarz. 2007. A newsvendors procurement problem when suppliers are unreliable. Manufacturing & Service Operations Management, 9(1), pp. 9–32.

[12] Federgruen, A., N. Yang. 2009. Optimal supply diversification under general supply risks. Journal of Operations Research, 57(6), pp. 909–925.

DRM in digital publication: limiting buyers' (readers') personal freedoms or a solution to the problem of online piracy

Slavko Šimundić and **Danijel Barbarić**
University of Split, Faculty of Law
Domovinskog rata 8, 21000 SPLIT, CROATIA
slavko.simundic@pravst.hr ; danijel.barbaric@pravst.hr

Abstract: DRM is an acronym for the English expression „digital rights management" – managing digital rights. DRM technology is used by content givers, just as in internet shops, in order to control the way in which digital files procured from them are used and distributed. Many publishers today use DRM to protect their digital products from illegal use, especially e-books. DRM is in practice designed as a technological system, that is, protection from unauthorised copying and illegal e-book distribution.

Keywords: digital, rights, management, e-book, piracy

1 INTRODUCTION

To start with, it is necessary to define the acronym DRM. DRM is an acronym for the English expression „digital rights management" – managing digital rights. DRM is a term which denotes technologies through which access to information in the digital world is controlled by publishers, that is, the owners of intellectual property in order to limit the use of digital contents.[1] DRM is not the only technology nor is it the only philosophy. DRM is a wide range of technologies and standards.[2]

Managing digital rights is the term related to the group of legal technical mechanisms constructed with the aim of allowing owners' intellectual property a greater level of control over the distribution and use of their products in the digital environment, in a way that the consumer rights of users when they use computer goods or services are regulated once they buy products.[3]

Over the last few years, DRM has been known by many other names. The first generation of DRM was directed at protection, encryption and resolving the problem of illegal copying in the way that it locks contents and limits their distribution to those who have purchased them.[4] The second generation of DRM which came about in the last few years distances itself from protection and encryption, and increasingly deals with the very management of it. Even though DRM has existed for several years, there is still no standard single definition of DRM.[5]

Even though there is no universal definition of DRM, there are various other definitions. The American Publishers' Association (which has more than 200 000 members) describes DRM technology as the tools and process of protection of intellectual property during the trade of digital contents.[6]

---

[1] Prlja, Dragan; Reljanović, Mario; Ivanović, Zvonimir; Internet pravo; Institut za usporedno pravo, Beograd, 2012. p. 19.
[2] Coyle, Karen, The Technology of Rights: Digital Rights Management, web location http://www.kcoyle.net/drm_basics.pdf (date accessed:: 05.04.2013. god.)
[3] sl. v. Bates, J.Benjamin, Commentary: Value and Digital Rights Management – A social Economics Approach, Journal of Media Economics, volume 21, Issue 1, 2008. p. 56.
[4] Fetscherin, Marc; Digital Rights Management: What the Consumer Wants, Journal of Digital Asset Management, (2006), 2, p. 143.
[5] ibid.
[6] ibid. p. 144.

It is believed that DRM should be the abbreviation for: Digital Restrictions Management.[7]

Economists have acknowledged that computer goods and services are not traditionally private goods. Many believe that these goods have similar characteristics to those of public goods which are available to all without limitations.[8] Therefore, it can be concluded that these goods should be given protection. Given that computer goods are accessible to a greater number of people than things are, it is obvious that a greater and better form of protection is needed. It is understandable that such „better"ways of protection influence the product price which will then greater. Product demand will be less and will distort production and dissemination of computer and cultural goods to a certain extent.[9]

1.1 Subjects in DRM

There are three fundamental subjects involved in every DRM, and these are: the user, contents and rights. The user can be anyone e.g. publisher, publishing house, film studio, corporation, individual or person. Contents can be any digital content e.g. music, games, software or film.

Rights can be allowable, limited or obligatory which are approved or given to the user and which are not applicable to the contents.[10] Here it should be mentioned that some authors consider that a fourth form of rights exists in which the user may make a copy of the contents.[11] Limitations may be imposed on all these rights. They can be limited according to duration (how long these contents can be used – e.g. a week), according to number (how many times – e.g. a song can only be played up to five times) or according to the location of the device.[12]

1.2 Use of DRM

DRM technology is used by content givers, just as in internet shops, in order to control the way in which digital files procured from them are used and distributed. Internets shops sell and rent various contents to which DRM has been applied. A protected file is a file that DRM has been applied to. Once DRM is applied to a file, it can no longer be removed.[13] Even though sources state that DRM protection cannot be removed, this is not completely true. Individuals who have an excellent command of informational technology and systems can „crack" DRM protection. With the very appearance and development of computers, a minority of individuals decided to turn to fraudulent activities. Today, such persons are called hackers. If hackers can break into the most tightly secured military data bases in the most developed countries in the world, they can also „crack" DRM protection. Of course such people are few in number, but they do exist. For this reason, we should never cease to develop newer and securer forms of protection. As for„cracking" DRM protection , we do not need hackers who break into the most strictly protected (e.g. military) secrets. This can be done by various computer experts who are highly educated and have a great deal of

[7] Internet pravo, p. 24.
[8] Bates, p. 62.
[9] v. Bates, p. 67.
[10] ibid.
[11] ibid. (v. Rosenblatt, B., Tripper, B. and Mooney S. Digital Rights Management – Business and Technoogy. M&T Books, New York, 2002. p. 63.)
[12] ibid. p. 145.
13 http://windows.microsoft.com/hr-HR/windows-vista/Windows-Media-Player-DRM-frequently-asked-questions (date accessed: 26.10.2012.)

knowledge about information systems. Due to today's accessibility of the internet, all DRM protection can be „cracked“, even by minors. It is enough to know English, have access to the internet and seek ways of removing DRM on the computer. Many internet pages on the way this can be done exist. Various content givers to whom DRM is applied can determine the way protected files accessed from them can be used. Thus there are cases where the providers of various, for example music, contents can give approval for reproducing files on the computer or synchronising music contents with a portable device and more. The case is similar with publishing.

Content providers can chose one of three main rights to offer users. The user can be given only the right to reproducing, watching and printing contents and this is the 'softest' form. The user can only be offered to save and transmit contents, either from device to device or person to person. The user allowed to change contents has the most rights.[14]

Over the last few years, numerous television producers are demanding the use of DRM technology in order to control access to their programmes due to the increase in popularity of DVR devices, namely, the DVD (Digital Video Recorder).[15]

Technology that DRM is based on partly uses tracking systems which raises the issue of privacy[16]. Above all, the user must have independence when using a certain product.

Some authors claim that DRM is an excuse for preventing illegal activity with digital contents. There are those who through torrent illegally download books even though they do not purchase them because their intention was never to purchase the book. However, there are millions of people who are prepared to pay for e-books. This punishes and diminishes the rights of people after they purchase books.[17]

DRM in digital publishing is an issue for debate. Some believe that the use of DRM limits the personal rights of buyers, that is, readers. Others believe that DRM is a way to definitively resolve the problem of online piracy.

2 DRM AND PUBLISHING

The development of graphical technology and general access to the internet sets new directions in e-publishing. Constant market change does not ensure continuity of production for printed publications. The unknown factor in determining the amount of print is an important variable in graphical production overall. Books and journals are printed in smaller publishing places. Harmonisation and distribution increase the final cost of a production unit.[18]

DRM manages the author's digital rights by integrating the publisher and buyer of the publication positioned on the server.[19]

Many publishers today use DRM to protect their digital products from illegal use, especially e-books. DRM is in practice designed as a technological system, that is, protection from unauthorised copying and illegal e-book distribution.

---

[14] Fetscherin, Marc, p. 144.
[15] http://www.cis.hr/files/dokumenti/CIS-DOC-2011-02-003.pdf , p. 4., date accessed 03.04.2013
[16] Bates, p. 69.
[17] Turčić, Maja; Janković, Mario; Kako Digital Right Movement šteti e-knjigama, International scientific conference on printing & design 2013, web lokacija http://www.tiskarstvo.net/printing&design2013/ (date accessed 05.04.2013)
[18] Miljković, Petar; Žvorc, Dean; DRM – u grafičkoj produkciji e-knjige, International scientific conference on printing & design 2013, web lokacija http://www.tiskarstvo.net/printing&design2013/ (datum pristupa 04.04.2013.)
[19] ibid.

DRM mainly helps to protect publishers from unauthorised use. However, it partly causes difficulties to even honest users who legally buy e- books because the books cannot be lent to friends as can be done with regular „paper“books. This problem often emphasised by leaders in the battle against DRM. Users are „locked“in their own platform.

So that publishers can be sure that the contents they sell be read only by purchasers, they have gone for the solution which disadvantages the very purchasers for the contents they have bought. More precisely, the content they have purchased is only a licence which means that their purchased product is not owned by them but by the publisher and the services on these contents can be withdrawn at any time. DRM encrypts files by solely the purchaser of the licence having the key. Another way of protecting digital contents is by placing a seal to prevent multiplication of contents. There is a whole range of manufacturers of devices and software who are all trying to have their model and software for reading digital books applied. In the attempt to strengthen their position, a range of DRM models exists. Various manufacturers force DRM in the formats they support.[20]

As far as classical borrowing of books is concerned where the book which is my property is lent by me to a friend and it then becomes his or her property, we can suggest that the same be applied to e-books. An e-book which is my property because I bought it should be allowed to be transferred directly from my e-reader to a friend's e-reader. Thus, the book is no longer be in my e-reader, but in the e-reader of my friend. My book would be transferred into the possession of my friend just like a traditional or paper book. In this way we could implement the traditional method of borrowing books. Similarly, with synchronising two e-readers, my friend could return the book to me. This kind of software would equate the traditional book with the electronic one. Acquiring ownership should be allowed and not just of the licence. However, strict limits should be placed to avoid online piracy.

With advanced software solutions, this way of borrowing e-books can be even better. Software solutions can enable that on my friend's e-reader when the book was borrowed, the date and even the duration of licence can be determined e.g. that the book can be used only for one month or more or less. After expiry date, the book can be „locked“that is, it cannot be opened without the password-a. These advanced solutions could be useful for libraries.

DRM technologies allow the publisher to build in a code into electronic publication so that they can limit the use of digital contents to those who have purchased or are authorised by the publisher or in some other way.[21]

The case in 2009 where Amazon without warning erased the 1984 book „Animal farm“ by George Orwell from all its users' devices because it was selling the book without the necessary rights is interesting to note. In 2011, 43 e-books were erased without warning for a Norwegian woman, but the contents were returned and Amazon never attempted to explain.[22]

Fear of infringing authors’ rights in the digital world lead some publishers to believe that DRM was important for continuing to protect materials protected by copyright and it continues to be demanded from publishers afraid of piracy. However, today there is a fear that DRM has caused unexpected consequences in the battle of the e-book market. It is possible that the life of DRM is limited or will demand re- contemplation. DRM together with its advantages has its disadvantages. There are technical problems that need resolving. Due to limited technological user experiences and the complexity of DRM it is often not

[20] Turčić, Janković

[21] Smith, Kelvin; The Publishing Business, From p-books to e-books; AVA Publishing, Lausanne, Switzerland, September, 2012. page 155.

[22] Turčić, Janković

possible an e-book to be read by those who have in good faith purchased it. The reason for this is that not all DRM systems are compatible with various e-readers.

It is very unfair of online shops which in their buying and selling contract clauses include exemption from responsibility in the event lack of compatibility between the e-book and the e-reader. I believe this problem should be immediately resolved. It is the obligation of every e-book seller to clearly state which e-readers can be used with which e-books. In this way, the buyer would in advance know whether the e-book can be used with his/her e-reader.

Some authors believe that insisting on DRM has proved to be a mistake DRM has not reduced piracy, but has „locked" users in for example Amazon's platform (Amazon's walled garden), by which it has demonstrated its power over publishers.[23]

3 DRM AND OTHER FORMS OF PROTECTION

Various forms of DRM protection can be found in music. It is well known that Apple, Microsoft, RealNetworks and Sony all have the systems that allow the manufacturer of the contents to place limits upon the use of digital music in their systems. Of course, the correct mechanism and level of permission varies from system to system.[24]

The publisher and author decide how they will protect e-books. DRM is not only limited to one format. It can be applied to other formats. So, the protection of Adobe Digital Edition (ADE) is used for a certain purpose. In order to read a certain text, the user must firstly install ADE on the computer or mobile device and register with his /her own Adobe ID. This form of protection is rather complicated and therefore discourages even the mot expert of users.

Alternatives for Adobe Digital Editions could be a seal which is partially visible to users. This means that the ordered e-book can only be located with one user. The advantages of this so called weaker DRM protection is that there are no negative consequences in the process of reading e-books.

The third model presented in the USA by Amazon and Apple is where the contents are registered to one user. Such DRM policy allows the user to read e-books on may e-readers, but does not allow transferring the contents to devices registered with other users.. Such forms of protection are occurring unnoticeably and usually do not have any consequences for users.

In the long term, many experts expect that DRM will disappear and that the e-book market will follow the music market.. Music publisher abandoned DRM in the spring of 2009 after a long battle against file sharing. Experts believe that abandoning DRM-is necessary sooner or later because illegal contents will anyway be accessible and DRM will not be able to fulfil its function of protection.

The advantage of weaker DRM protection is that the owner is easily identifiable and forwarding e-books is not protected by law.[25] The publisher and authors should contemplate their stance on DRM. Apple and Amazon have successfully demonstrated that DRM can also be used towards the user for friendly purposes. Publishers should want to accept DRM systems so it is not frightening to the market by making it difficult to buy or use digital contents.[26]

[23] Smith, Kelvin; p. 155
[24] Bates, p. 59.
[25] Turning the Page, The Future of eBooks; PWC, http://www.pwc.com/en_GX/gx/entertainment-media/pdf/eBooks-Trends-Developments.pdf ; str./page. 14. (date accessed15.02.2013.)
[26] ibid. page. 31.

4 CONCLUSION

It is very difficult with certainty and the consensus of all interested participants to determine whether DRM in the end is good or bad. Some authors have dealt with DRM strongly criticise it, others however praise it and claim it is necessary. Certainly, digital products in the future will need protection, because they can be easily be copied and it is difficult to influence this. However, librarians can influence the development of DRM technology by participating in discussions in organisations and research arenas. Our professional duty is to secure participation in the development of technologies which will influence the future of reading and approach to information.[27]

Today's technological development and the weaknesses of DRM protection demand new approaches. Even though DRM offers good and purposeful protection, that today is not enough. There is also the opinion that DRM is aimed at protecting contents and that it does not prevent piracy or dividing users on a particular platform.[28] Certainly, the decision should be up to the professionals and users. Above all, librarians should be asked how they see the future of e-books and what they think would be the best solution.

Various electronic devices and also various e-readers like Amazon Kindle, Barnes and Noble Nook Tablet, Apple iPad, Kobo and so on have their serial numbers. By the serial number we know exactly that the e-reader is ours. It is necessary to find a way to link buying an e-book to an e-reader. If we cannot purchase a paper book without money, we should not be allowed to purchase an e-book without an e-reader linked to a computer. By buying an e-book, it becomes ours and we can do what we like with it. If we decide to lend it to a friend we can do so, but in that case it would be necessary to synchronise two readers and transfer the book from one e-reader to another e-reader, just like lending the paper version physically in person. Such a suggestion to a certain extent would equate the paper book with the e-book. However, it would be necessary to survey e-reader users. In this way we would have first hand knowledge about what people are happy with and what they would like or suggest.

Correct and appropriate decisions are often complex. Maybe publishers could try to bring in new technologies and help in resolving this problem, they could use optimum seeking methods or some other operational research method.

People of different profession should work together on this topic and try to solve this problem on common delectation.

References

[1] Bates, J.Benjamin, Commentary: Value and Digital Rights Management – A social Economics Approach, Journal of Media Economics, Volume 21, Issue 1, 2008

[2] Coyle, Karen, The Technology of Rights: Digital Rights Management, web lokacija http://www.kcoyle.net/drm_basics.pdf (datum pristupa 05.04.2013. god.)

[3] Fetscherin, Marc; Digital Rights Management: What the Consumer Wants, Journal of Digital Asset Management, (2006), 2

[4] http://windows.microsoft.com/hr-HR/windows-vista/Windows-Media-Player-DRM-frequently-asked-question (date accessed: 26.10.2012.)

[5] http://www.cis.hr/files/dokumenti/CIS-DOC-2011-02-003.pdf , (date accessed 03.04.2013.god.)

[6] http://www.cis.hr/www.edicija/LinkedDocuments/CCERT-PUBDOC-2007-10-207.pdf (date accessed 03.04.2013.god.)

[7] http://www.publishers.org/about/ (date accessed: 03.04.2013.god.)

[27] Coyle

[28] Turčić, Janković

[8] http://www.pwc.com/en_GX/gx/entertainment-media/pdf/eBooks-Trends-Developments.pdf (Turning the Page, The Future of eBooks; PWC,) (date accessed: 15.02.2013.god.)
[9] Klarić, Petar; Vedriš, Martin; Građansko pravo, Narodne novine, Zagreb, 2006. god.
[10] Miljković, Petar; Žvorc, Dean; DRM – u grafičkoj produkciji e-knjige, International scientific conference on printing & design 2013 , web lokacija http://www.tiskarstvo.net/printing&design2013/ (datum pristupa 04.04.2013.god.)
[11] Pravni leksikon, Leksikografski zavod Miroslav Krleža, Zagreb, 2007. god.
[12] Prlja, Dragan; Reljanović, Mario; Ivanović, Zvonimir; Internet pravo, Institut za uporedno pravo, Beograd, 2012. god.
[13] Rosenblatt, B., Tripper, B. and Mooney S. Digital Rights Management – Business and Technoogy. M&T Books, New York, 2002.
[14] Smith, Kelvin; The Publishing Business, From p-books to e-books; AVA Publishing, Lausanne, Switzerland, September, 2012.
[15] Turčić, Maja; Janković, Mario; Kako Digital Right Movement šteti e-knjigama, International scientific conference on printing & design 2013, web lokacija http://www.tiskarstvo.net/printing&design2013/ (datum pristupa 05.04.2013. god.)
[16] Vidaković-Mukić, Marta; Opći pravni rječnik, Narodne novine, Zagreb, 2006. god.
[17] Žvorc, Dean; Miljković, Petar; Upravljanje digitalnim pravima (DRM) – zaštita knjiga na internetu; 16th International conference of Printing Design and Graphic Communication, Blaž Baromić, 2012. god.

The 12th International Symposium on
Operational Research in Slovenia

SOR '13

Dolenjske Toplice, SLOVENIA
September 25 - 27, 2013

Section VI:

Finance and Investments

PENSIONS AND HOME OWNERSHIP IN THE WELFARE MIX FOR OLDER PERSONS

David Bogataj
European Faculty of Law, Nova Gorica, Slovenia, e-mail: dbogataj@actuary.si

Abstract: The paper draws on actuarial mathematics, examining the role of housing in the welfare mix for older persons. New concepts for asset-based welfare, where the housing owned by the occupant is part of the investment portfolio comprised of state pension, occupational defined contribution (private) pension and one's residential property, are examined. We present how the variance of pension income after retirement is reduced by using residential property as the 4th pillar of the pension system, as proposed in the EC Green Paper on Pensions (2010). The study is focusing on modelling the decumulation of the housing equity and the defined contribution private pension, incorporating insurance mechanisms for management of longevity. Here we propose a new model in which periodic payout that the beneficiary receives is the difference between the amount drawn and the annuity premium for longevity insurance. The paper shows how the drawing amount in the loan model ERS (reverse mortgage) is decreasing with the increasing interest rate, while the pension arising from defined contribution systems is increasing with the increasing interest rate. According to Markowitz's Portfolio Theory, these findings show that in combination of these products, the volatility of a combined pension cash flow from all pillars induced by volatile interest rates decreases, which improves the pension portfolio, i.e. where pensions from defined contribution systems exist.

Keywords: actuarial mathematics; equity release scheme; housing; longevity**;** reverse mortgage.

1 INTRODUCTION

In funded defined contribution pension systems, the amount of yearly pension depends on the accumulated amount in the individual retirement account and long-term interest rate at the moment of retirement. As presented by Shiller (http://www.irrationalexuberance.com, 2005; updated data), the long-term interest rates are highly volatile. Therefore, the yearly amount of one's pension is uncertain. Markowitz (1952) has shown that such volatility of returns can be mitigated by adding negatively correlated assets in the portfolio.

The paper examines the possibility of developing and implementing flexible Housing Equity Release Schemes (ERS) as a means of providing a more stable welfare provision for the elderly, adding housing wealth in a portfolio of pension instruments. This could provide a better welfare provision for older persons by stabilizing the total disbursement to the beneficiary, where the volatility depends on the volatility of the interest rate (funded pensions). The falling interest rates since 2009 have had adverse effects on funded pensions, and the austerity policies since 2010 have reinforced the inequalities, particularly among the 'income poor' but 'asset rich' older population. Pensions are often not sufficient to cover health expenses and other needs of older persons. Hence, the question of an optimal welfare mix for older people is significant.

Therefore, the author seeks to examine the following research questions: (a) How to ensure a more appropriate welfare mix for older people; (b) What role could Equity Release Schemes play; and (c) How can the financial industry develop attractive financial products to fit within the mix of other (private and public) welfare provisions, where the risks of poverty due to the falling interest rate can be mitigated to the benefit of older property owners? Using actuarial mathematics with life contingencies, the paper will present how reverse mortgage systems (ERS loan model) with the embedded insurance for longevity might improve the results of the senior housing provision and the satisfaction of inhabitants. Based on the presented findings and on the Portfolio Theory (Markowitz, 1952; Tu & Zhou, 2011), we can also show that the interest rate variation, which can reduce the income of older persons under the poverty line, has a significantly smaller impact on welfare of the elderly if

pension pillars are combined with the ERS loan model. Such financial product would be a novelty in the insurance and banking industry.

The paper first describes the existing models of ERS, then it introduces the model for mitigating the credit default risk using longevity insurance. A numerical example is also presented.

2 MODELLING ERS

2.1 ERS models

According to the clear Reifner's description, ERS transform fixed assets in owner occupied dwellings into liquid assets for private pensions. They thus enable a homeowner to access the wealth accumulated in the form of the home, while being able to continue to live in it. An illiquid asset becomes a source of liquidity, mainly for consumption needs. ERS can take two different forms: (a) Loan Model ERS, also known as reverse mortgage, provide a loan that will be repaid from the sale after the death of owner, and (b) Sale Model ERS, which involve an immediate sale of the property but provide for the right to remain in occupation and to use the cash price for income in retirement. ERS must therefore: (a) be a financial service; (b) be a source of liquidity for the future; (c) contain a strong entitlement to remain in occupation of the property; and (d) rely solely on the sale of the property for repayment/payment of the funds released to be used as a retirement pension. Payments take the form of a lump sum or periodic (monthly, yearly) income, and are either secured by means of a mortgage on the property or generated by an immediate sale. Under the Loan Model ERS, repayment is made from the proceeds of the sale of the property either after the death of the homeowner or when the property has become vacated for a longer time (see definition in Reifner et al., 2009).

2.2 Sale model

The sale model is very straightforward and is similar to the general annuity model (funded pensions – 2nd and 3rd pillars), as described in Gerber (1980). In the sale model, the whole value of the real estate is transferred to lifetime annuity at the moment of closing of the ERS contract. The value of the property is used for purchase of the lifetime annuity. The amount of yearly payout of the lifetime annuity therefore needs to cover the interest on the principal amount taken out and the yearly annuity paid to the beneficiary of the ERS, as is presented in:

$$\ddot{a}_{x+ps} = \sum_{j=0}^{110-x+ps} {}_j p_{x+ps} \cdot v^j = \sum_{j=0}^{110-x+ps} {}_j p_{x+ps} \cdot (1/(1+i))^j \qquad (1)$$

$\ddot{a}_{x+ps}$ is the present value of the prenumerando lifetime annuity of the amount 1 EUR for the person that is x years old (ps is the age correction conforming to the methodology of annuity mortality tables , ${}_j p_{x+ps}$ is the probability that the person that is x years old will survive the next j years, and v is the discounting factor ($\nu = 1/(1+i)$), where i is the annual interest rate. The amount of lifetime annuity is calculated as the annuity factor multiplied with the net value of real estate, which is calculated as the value of the real estate minus the cost associated with the transaction (valuation costs, taxes, costs of sale). Annuity factor $fr(x,i)$ is:

$$fr(x,i) = 1/\left\{(1+\gamma_2)\cdot \ddot{a}_{x+ps}\right\} = 1/\left\{(1+\gamma_2)\sum_{j=0}^{110-x+ps} {}_j p_{x+ps} \cdot (1/(1+i))^j\right\} \qquad (2)$$

where rate γ_2 represents the costs associated with the payout of the annuity that the insurance company charges for each payout in the period of annuity. The yearly amount of annuity R is calculated according to the value of real estate VN and annuity factor $fr(x,i)$:

$$R = fr(x,i)\cdot(VN-C) = (VN-C)/\left\{(1+\gamma_2)\sum_{j=0}^{110-x+ps} {}_j p_{x+ps}\cdot(1/(1+i))^j\right\} \quad (3)$$

2.3 Loan model

The loan model or "reverse mortgage" is a type of home loan that allows a borrower to open up a line of credit using their home as collateral. With the loan model the beneficiary draws liquid amounts in lump-sum or/and periodically from the value of the real estate in the form of loan secured by a mortgage on the real estate. With the part of this liquid amount that is drawn from the real estate the beneficiary purchases lifetime annuity in the form of a monthly premium. In this way the beneficiary insures his longevity so that if he lives longer than his life expectancy he will receive a lifetime annuity until his death. In the paper, we propose the ERS model with insurance for longevity, where the periodic payout that the beneficiary receives is the difference between the amount drawn and the annuity premium for longevity insurance. In this way, if the beneficiary survives the drawing period of ERS (n years), he receives a lifetime annuity that covers the disbursement to the beneficiary and the interest on the outstanding loan. This is a new scheme, proposed in Bogataj (2013). Generally, loan models allow the beneficiary to draw the value of the real estate in different ways: (a) In lump sum at the closing of the ERS contract; (b) In the form of line of credit so that he can draw it when necessary; (c) In uniform periodic amounts in the period of life expectancy. The maximum amount of loan (*MLA*) that can be drawn from the real estate is the value of the real estate *(VRE)* minus all the costs *(C)*, i.e. those associated with the closing of the ERS contract (C_1) and with the sale of the property after the death of the beneficiary (C_2).

$$MLA = VRE - C = VRE - C_1 - C_2 \quad (4)$$

A life annuity consists of a series of payments which are made while the beneficiary (of initial age x) lives. The present value of the life annuity due with yearly payments in the amount of 1 EUR is denoted by $\ddot{a}_{x+ps:\overline{n}|}$, where the following equation can be written:

$$\ddot{a}_{x+ps:\overline{n}|} = \sum_{j=0}^{n} {}_j p_{x+ps}\cdot v^j \quad (5)$$

The present value of the life annuity deferred for n years with yearly payments in the amount of 1 EUR is denoted by ${}_{n|}\ddot{a}_{x+ps}$, where the following equation can be written:

$${}_{n|}\ddot{a}_{x+ps} = {}_n p_x\cdot v^n\cdot\ddot{a}_{x+ps} = {}_n p_x\cdot v^n\cdot\sum_{j=0}^{110-(x+ps+n)} {}_j p_{x+ps+n}\cdot v^j \quad (6)$$

The premium rate for longevity insurance $prs(x,i,n)$ is:

$$prs(x,i,n) = (1+\gamma_2)\cdot {}_{n|}\ddot{a}_{x+ps} / (1+\gamma_1)\cdot\ddot{a}_{x+ps:\overline{n}|} \quad (7)$$

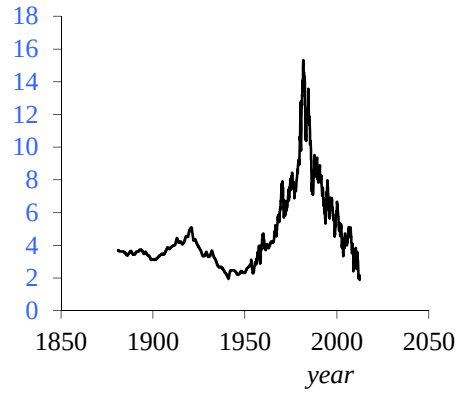
where γ_1 represents the rate of administration expenses that are charged against the policy in the period of premium payments and γ_2 represents the rate of administration expenses that are charged against the policy in the period of annuity payments. The yearly amount of premium *(PR)* is calculated as: $PR = prs(x,i,n)\cdot R$, (8)

where R is the annuity amount. In this case, the yearly amount *(YDA*) that the beneficiary can draw from the real estate is:

$$YDA = i \cdot MLA / \left[(1+i)^n - 1 \right] = YPA + \Pr = YPA + prs(x,i) \cdot (YPA + MLA \cdot i) \tag{9}$$

2.4 Mitigating credit default risk

The main risks concerning ERS that can cause credit default are the uncertain longevity of the owner occupier, the risk of increase in interest rates and depreciation in the value of the property. Deferred annuity as an insurance for longevity is already used in insurance industry, but not in combination with reverse mortgage, which is a novelty proposed here. Without an effective insurance for longevity, real estate cannot be used as the 4th pension pillar, because the equity release without a longevity insurance presents a great risk for the provider of reverse mortgage (bank) and also for the beneficiary. For the reverse mortgage provider, there is the risk that the value of the loan together with accrued interest will be greater than the value of real estate in case of the death of the beneficiary. For the beneficiary, there is the risk that he will live longer than the agreed period of drawing liquid amounts that is defined in the reverse mortgage loan contract. To avoid exposure to these risks, a safe reverse mortgage contract also needs to include an insurance for longevity. This insurance can be provided in three ways: through public finance so that the risk is socialized and the management of risk is assumed by the government (as is the case in the USA), through private sector with transfer of the risk to a joint stock insurance company or – a third way –through a mutual insurance company, as proposed in Bogataj (2013). The risk of longevity can be mitigated by the use of annuity insurance, but it is difficult to avoid the impact of the volatile interest rate. The volatility of the interest rate 1871 - 2012 is presented in Fig. 1.



Shiller, 2005+updates

Fig.1: Interest rate in % (1871 – 2012)

Table 1: Comparison of payout amounts according to private pension and loan model for a 65 years old man, mortality tables: *DAV1994R*, value of fund or home *VN=160,000* EUR Drawing period is 16 years.

i	$\ddot{a}_{65}$	$1.05 \cdot \ddot{a}_{65}$	$R=\frac{160,000}{1.005 \cdot \ddot{a}_{65}}$	*YPA*
2%	18.11	18.20	8,793	4,560
3%	16.31	16.39	9,762	3,960
4%	14.80	14.88	10,756	3,480
5%	13.53	13.59	11,771	3,000
6%	12.44	12.50	12,803	2,520
12%	8.56	8.60	19,128	780
16%	6.99	7.02	23,292	156

2.5 Portfolio of assets and the cash flow based on ERS and other pension schemes

From (3) it follows that *R* increases with increasing i . From (9) the yearly amount of the disbursement from the ERS to beneficiary *YPA* can be calculated:

$$\begin{aligned} & i \cdot MLA / \left[(1+i)^n - 1 \right] = YPA + prs(x,i_a) \cdot (YPA + MLA \cdot i) \\ & YPA = MLA \cdot i \left\{ 1 / \left[(1+i)^n - 1 \right] - prs(x,i_a) \right\} / (1 + prs(x,i_a)) \end{aligned} \tag{10}$$

YPA decreases with increasing positive *i* if

$$d\left\{i\left\{1/\left[(1+i)^n-1\right]-prs(x,i_a)\right\}\right\}\ /di<0\Rightarrow 1/\left[(1+i)^n-1\right]-prs(x,i_a)-i\cdot n(1+i)^{n-1}/\left[(1+i)^n-1\right]^2<0 \quad (11)$$

From (11) it follows that the sufficient condition is $(1+i)^{n-1}\left[1-(n-1)i\right]<1$. Therefore, in the loan model of ERS, where it is always $i>0$ and $n>1$, *YPA* always decreases with increasing positive i. The curves (3) and (10) are close to the linear curve, while at (3) correlation coefficient $\rho_{R,i}$ is always close to 1, and at (10) correlation coefficient $\rho_{YPA,i}$ is always close to -1. Therefore, we can expect that there exists an optimal portfolio which can substantially reduce the variance of portfolio (Markowitz, 1952).

3 A NUMERICAL EXAMPLE

3.1 Periodic liquid amounts drawn from the value of real estate

Based on equations (1)–(9), Table 2 presents the periodic liquid amounts drawn from the value of real estate, where part of the amount drawn is used for purchase of the longevity insurance. The example presents the ERS loan model for a 65-year old man whose property value is 160,000 EUR, the discounting factor $(v=1/(1+i))$ of the insurance company is (i=3.5%), the bank interest rate is also 3.5%, and mortality tables DAV 1994 R are used; embedded administrative cost of γ_1,γ_2 both equal 5% .

Table 2: Drawing periodic liquid amounts from the value of real estate

Year	*Closing costs*	*Amount drawn each year*	*A(y)+ B(y)+ D(y-1)*	*Interest amount 3.50%*	*Accumulated debt C(y)+D(y)*	*Annuity premium*	*Accounting costs*	*yearly disbursement equal to 288 EUR per month*
y	*A*	*B*	*C*	*D*	*E*			
1	3,200.00	6,524.98	9,724.98	340.37	10,065.35	2,948.98	120.00	3,456.00
2		6,524.98	16,590.33	580.66	17,170.99	2,948.98	120.00	3,456.00
3		6,524.98	23,695.97	829.36	24,525.33	2,948.98	120.00	3,456.00
4		6,524.98	31,050.31	1,086.76	32,137.07	2,948.98	120.00	3,456.00
5		6,524.98	38,662.05	1,353.17	40,015.22	2,948.98	120.00	3,456.00
6		6,524.98	46,540.20	1,628.91	48,169.11	2,948.98	120.00	3,456.00
7		6,524.98	54,694.08	1,914.29	56,608.38	2,948.98	120.00	3,456.00
8		6,524.98	63,133.36	2,209.67	65,343.02	2,948.98	120.00	3,456.00
9		6,524.98	71,868.00	2,515.38	74,383.38	2,948.98	120.00	3,456.00
10		6,524.98	80,908.36	2,831.79	83,740.15	2,948.98	120.00	3,456.00
11		6,524.98	90,265.13	3,159.28	93,424.41	2,948.98	120.00	3,456.00
12		6,524.98	99,949.39	3,498.23	103,447.62	2,948.98	120.00	3,456.00
13		6,524.98	109,972.60	3,849.04	113,821.64	2,948.98	120.00	3,456.00
14		6,524.98	120,346.62	4,212.13	124,558.75	2,948.98	120.00	3,456.00
15		6,524.98	131,083.73	4,587.93	135,671.66	2,948.98	120.00	3,456.00
16		6,524.98	142,196.63	4,976.88	147,173.52	2,948.98	120.00	3,456.00
	3,200.00			39,573.86		47,183.66	1,920.00	55,296.00
Selling cost of property at death of owner					12,826.48			
Realised property value at death of owner					160,000.00			

In the case represented in Table 2, the beneficiary owns the residential property that has a value of EUR 160,000, which does not grow over the time horizon of the owner's life span. All costs associated with entering into a reverse mortgage contract (brokerage fee, assessment fee, notary fee and other administrative costs) are 2% (EUR 3,200.00). The yearly amount drawn from the residential real estate equity equals EUR 6,524.98. This amount is then divided into three parts: EUR 120.00 covers the administrative fee for maintaining a reverse mortgage account with a financial institution, EUR 2,948.98 is used for purchasing the annuity premium that covers the annuities after exhausting all the equity in the residential real estate property (in case the real estate owner lives longer than expected). The yearly amount of EUR 3,456.00 is disbursed to the property owner who is staying in his property until the end of his life in any case. At the end of his life, the costs of refurbishing and selling the property are covered by the remaining EUR 12,826.48. After refurbishing, the house is sold for EUR 160,000.00, which was also the estimated price at the closing of contract.

3.2 Mitigating the risk of poverty at a volatile interest rate

The correlations between monthly funded pensions *R*, YPA and interest rate, and correlation coefficients ρ for this numerical example are the following:

$$R = 554 + 86.5\, i\,, \qquad \rho_{R,i} = 0.998;$$

$$YPA = 581 - 0.3\, R, \qquad \rho_{YPA,R} = -0.935$$

$$YPA = 406 - 27.5\, i; \qquad \rho_{YPA,i} = -0.972$$

Therefore by adding any amount drawn from ERS to the funded pension, the volatility is efficiently reduced, because sign ($\rho_{R,i}$) and sign ($\rho_{YPA,i}$) are different.

4 CONCLUSION

Equity Release Schemes transform fixed assets in owner occupied dwellings into liquid assets for private pensions. We have shown that the interest rate variation, which can reduce the income of older persons even below the poverty line, has a significantly smaller impact on welfare of the elderly if funded pensions, which have a positive covariance with the interest rate, are combined with the ERS loan model, where the correlation coefficient is negative. Because of the volatility of the interest rate, as the case presented in Fig. 1, volatility should be studied carefully and the proper combination of pensions and dynamics of ERS drawings has to be chosen to decrease the volatility of combined cash flows deriving from different pension pillars. Due to the negative correlation of cash flows from funded pensions with disbursements from ERS, the volatility induced by the volatile interest rate is reduced. Markowitz's Portfolio Theory should be considered here. Regarding their old-age welfare protection, young families should consider buying a home instead of renting one, in order to accumulate their assets. Therefore, governments should provide incentives to achieve this goal, e.g. in a way presented by Bogataj and Aver (2013), and should enable their citizens to buy properties based on mortgage financing.

References

[1] Bogataj, D., 2013. *Vlagaj v svoj dom, da boš dolgo živel in ti bo dobro na zemlji*, (MEORL, Ser.No. 13). Nova Gorica: MEDIFAS.

[2] Bogataj, D., Aver, B., 2013. Uvedba zakonskih podlag za načrte črpanja nepremičnega premoženja starostnikov. *PP, Prav. praksa (Ljubl.)*, 32/13, p 6-8.

[3] EUR opean Commission, 2010. *GREEN PAPER-towards adequate, sustainable and safe EUR opean pension systems*, Brussels, COM(2010)365 final.

[4] Gerber, H.U., 1980. *Life Insurance Mathematics*, Swiss Association of Actuaries, Springer –Verlag, Berlin, Heidelberg, New York.

[5] Markowitz, H., 1952. Portfolio Selection, *Journal of Finance*, American Finance Association, vol. 7(1), pages 77-91.

[6] Reifner, U.,2009. Clerc-Renaud, S., Pérez-Carrillo, E.F., Tiffe, A., Knobloch, M., *Study on Equity Release Schemes in the EU*, Institut für Finanzdienstleistungen e.V., Hamburg.

[7] Shiller, R.J., 2005. Irrational Exuberance, Princeton, New Jersey (updated data: http://www.irrationalexuberance.com).

[8] Tu, J. & Zhou, G., 2011. Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies, *Journal of Financial Economics*, Elsevier, vol. 99(1), pages 204-215.

THE ADAPTATION OF EXTENDED NET PRESENT VALUE THEORY AND SOLVENCY II IN RISK MANAGEMENT

David Bogataj*, Robert Vodopivec** and **Marija Bogataj****
* European Faculty of Law, Nova Gorica, Slovenia, e-mail: dbogataj@actuary.si
** MEDIFAS, Šempeter pri Gorici, Slovenia, e-mail: vodopivec.robert@siol.net, marija.bogataj@guest.arnes.si

Abstract: The focus of this paper is the risk management of total supply chains through identifying risk drivers that could appear simultaneously and mitigating supply chain risk. Any risk driver that is likely to disrupt the procurement, production, transportation, warehousing, delivery or financing of a good or service constitutes a realisation of supply chain risk. Risk drivers often appear simultaneously. As many cases from around the world show, disruptions to supply chains can be of low severity or catastrophic to corporations, global supply chains and even national and international economies. It is imperative, therefore, that an a priori assessment of risk drivers that pose risk to the global supply chain is undertaken and that contingency plans are developed at every level to monitor and mitigate these risks, even when they appear simultaneously. The main duty of a supply chain manager is to prevent the ruin of a supply chain exposed to risks. To avoid the ruin of a supply chain, we must ensure the availability of adequate funds. Therefore, the risk-mitigation approach advanced in our paper follows from our conviction that money is a stock of purchasing power of any activity cell in a global supply chain that could influence a perturbation of material flows—on many stages simultaneously—and not only financial flows in a supply chain. In the paper, we provide a method that is closely related to the Solvency II methods but appropriate for studying the long-term solvency of a supply chains. As the balance sheet assets under consideration are different from those of banks and insurance companies, the solvency method could not be adopted directly. This new approach is based on Material Requirements Planning (MRP) Theory, as developed by Grubbström and later extended by Bogataj and Grubbström, in which simultaneous perturbations in the timing of financial flows, information flows and flows of items can be better evaluated through Laplace transforms and the net present value (NPV) expression.

Keywords: Risk Management, Disruption risk, Supply chain, Solvency II, MRP Theory, Laplace Transforms.

1 INTRODUCTION

Finance is the lifeblood of any supply chain. Supply chain managers should not neglect the availability of financial resources. Many activities in the economy are being affected by the current economic downturn, and supply-chain financing is facing the same problems as other types of financing. These problems are caused by economic uncertainty. By increasing the distances between pairs of activity cells belonging to global supply chains, visibility has become lower, and vulnerability has increased. Supply chain risks have been characterised as circumstances in which "unexpected events might disrupt the flow of materials on their journey from initial suppliers to final customers". A sudden liquidity problem in an activity cell in a supply chain that would disrupt material flows could constitute such an event. These events continue to influence disruptions of material flows in some supply chains today. The main purpose of this paper is present a method for assessing risks and determining the amount of money that could mitigate a given risk using an approach similar to the Solvency II method in the insurance industry. In supply chain management, the risk of cascading failures of activity cells can cause a catastrophic failure, often referred to as systemic risk. Catastrophic failure is a sudden and total failure of a system, from which recovery is impossible. Recently, nearly all financial systems faced cascading system failures (i.e., systemic risk in finance). There are better known cascading failures of computer networks or electric-power transmission systems but few reported the cascading system failures in a

supply chain. The failure of one cell of activity in a supply chain can cause other cells of activities (i.e., its counterparties) to fail. For better supply chain management, financial and physical flows must be merged and studied dependently, especially when catastrophic risk is in question. Without proper formalisation of such supply networks, the mitigation of cascading risk can't be properly solved. EMRP Theory was found to solve the problem of the proper presentation of such problems. Some of the Solvency II directives, which are well described and available in the Directive of the European Parliament PE-CONS 3643/6/09, REV 6, have been modified and included.

2 EMRP THEORY AS A FRAMEWORK FOR RISK ANALYSIS IN GLOBAL SUPPLY NETWORKS

To better evaluate simultaneous perturbations of intensity of flows, perturbed delays and their cumulative impact on risk realisation EMRP model is used here, on the basis of Grubbström's basic MRP Theory (1998). And first extended by Bogataj, Grubbström and Bogataj (2011), while a detailed presentation and evaluation of simultaneous perturbations of various delays and their impact on the NPV of combined activities in a supply chain was presented in Bogataj and Grubbström (2012, 2013). The basic elements of MRP theory are the rectangular input and output matrices H and G, respectively, which have the same dimensions. We let m denote the number of processes (i.e., columns) and n denote the number of item types (and location, here rows, which are results of processes one stage earlier). If the j th process at location j is run on activity level P_j, the volume of required inputs of item i is $h_{ij}P_j$, and the volume of produced (transformed) outputs of item k is $g_{kj}P_j$. The total of all inputs may then be collected into the column vector HP, and the total of all outputs may be collected into the column vector GP, from which the net production is determined as (G – H)P. For the sake of simplicity, we assume that G=I. In MRP systems, lead times are essential ingredients that are often stochastic by nature, influencing losses (negative added values). They appear in activity cells and links between two activity cells. The volume $h_{ij}P_j$ of item i, previously having been a part of available inventory, is reserved at the end of the production of item i at time $(t-\tau_j-\tau_{ij})$ for the specific production $P_j(t)$ and thereby moved into work-in-process (allocated component stock, allocations). At time t, when this production is completed, the identity of the items of type i disappears; instead, the newly produced items $g_{kj}P_j(t)$ appear. Because of the stochastic nature, the production can have additional delays τ_j^d or τ_{ij}^d if delay appears in an activity cell or during the transportation of the procedure from i to j, respectively. τ_j^d and τ_{ij}^d could be random variables, influencing unforeseen future states of the supply chain added value or NPV. Consider an assembly system for which the components of the process j need to be sent from i at least $\tau_{ij}+\tau_{ij}^d$ time units before they must arrive to activity cell j and in place j at least $\tau_j+\tau_j^d$ time units before completion. Applying the time-translation theorem, the input requirements as transforms will be the following:

$$\begin{bmatrix} 0 & 0 & 0\,0 & \cdots & 0 \\ .... & ..... & h_{ij}e^{s(\tau_{ij}+\tau_{ij}^{d})} & \ddots & \vdots \\ h_{m1}e^{s(\tau_{m1}+\tau_{m1}^{d})} & h_{m2}e^{s(\tau_{m2}+\tau_{m2}^{d})} & & \cdots & 0 \end{bmatrix} \begin{bmatrix} e^{s(\tau_1+\tau_1{}^{d})} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & e^{s(\tau_m+\tau_m{}^{d})} \end{bmatrix} \tilde{\mathbf{P}}(s) = \tag{1}$$

$$= \mathbf{H}^{\omega\mathbf{d}}\tilde{\boldsymbol{\tau}}^{d}(s)\tilde{\mathbf{P}}(s) = \tilde{\mathbf{H}}^{\omega\mathbf{d}}(s)\tilde{\mathbf{P}}(s)$$

where $\tilde{\boldsymbol{\tau}}^{d}(s)$ is the perturbed lead-time matrix and $\tilde{\mathbf{H}}(s)^{\omega\mathbf{d}}$ is the generalised perturbed input matrix capturing the volumes of requirements and their advanced perturbed timing. This vector describes in a compact way all of the component volumes that must be in place for the production plan $\tilde{\mathbf{P}}(s)$ to be possible, as described in Bogataj and Grubbström (2013). The net production of such a system will conveniently be written as follows:

$$(\mathbf{I} - \mathbf{H}^{\omega\mathbf{d}}\tilde{\boldsymbol{\tau}}^{d}(s))\tilde{\mathbf{P}}(s) = (\mathbf{I} - \tilde{\mathbf{H}}^{\omega\mathbf{d}}(s))\tilde{\mathbf{P}}(s)^{\cdot} \tag{2}$$

Here, we may say that $\mathbf{I} - \tilde{\mathbf{H}}^{\omega\mathbf{d}}(s)$ is the stochastic generalised technology matrix. $\tilde{\mathbf{F}}(s)$ represents deliveries (i.e., exports) from the system. In his MRP model, Grubbström (1998) also introduced *cyclical processes*, repeating themselves in constant time intervals T_j, j = 1, 2, … , m. We may write the plan $\tilde{\mathbf{P}}(s)$ in the following way, using two new diagonal matrices $\tilde{\mathbf{t}}(s)$ and $\tilde{\mathbf{T}}(s)$,

$$\tilde{\mathbf{P}}(s) = \begin{bmatrix} e^{-st_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & e^{-st_m} \end{bmatrix} \begin{bmatrix} \left(1-e^{-sT_1}\right)^{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \left(1-e^{-sT_m}\right)^{-1} \end{bmatrix} \hat{\mathbf{P}} = \tilde{\mathbf{t}}(s)\tilde{\mathbf{T}}(s)\hat{\mathbf{P}}, \tag{3}$$

where $\hat{\mathbf{P}}$ is a vector of constants describing, for instance, the total amounts (i.e., batch sizes) to be produced in (or delivered by) each process during one of the periods T_j, j = 1, 2, … , m, and where t_j, j = 1, 2, … , m are the points in time when the first of each respective cycle starts. Matrix $\tilde{\mathbf{t}}(s)$ could also be considered perturbed as exposed to high-frequency small-severity risk. Further, we do not analyse perturbations in $\tilde{\mathbf{t}}(s)\tilde{\mathbf{T}}(s)$. In the case of demand disruptions, we can say that $\tilde{\mathbf{F}}(s) > \tilde{\mathbf{D}}(s) = \hat{\mathbf{D}}/s$, where $\hat{\mathbf{D}}$ is a stochastic vector with mean $\bar{\hat{\mathbf{D}}}$ influencing disruption risk if demand is under the critical value $\bar{\hat{\mathbf{D}}} - \boldsymbol{\Delta}$. If we wish to have 0.995 probability that demand is not falling under critical value, we have to put the following:

$$P\left(\left(\mathbf{I} - \mathbf{H}^{\omega\mathbf{d}}\right)\left\{\hat{P}_1/T_1, ..., \hat{P}_m/T_m\right\} > \bar{\hat{\mathbf{D}}} - \boldsymbol{\Delta}\right) > 0.995. \tag{4}$$

Let us introduce a price vector $\mathbf{p}$ as the following row vector: $\mathbf{p} = [p_1, p_2, ..., p_n]$ (5)

The NPV of the costs will be assumed here, as presented in Bogataj and Grubbström (2012, 2013). The chain could be exposed to operational risk because of low-severity, high-frequency perturbations of $\tilde{\mathbf{H}}^{\omega\mathbf{d}}(s)$ (also $\tilde{\mathbf{t}}(s)$) and disruption costs when demand or lead times exceed the critical value. Therefore, the overall NPV may be written as follows:

$$\mathrm{NPV} = \mathbf{p}\left(\mathbf{I} - \tilde{\mathbf{H}}^{\omega \mathbf{d}}(\rho)\right)\tilde{\mathbf{P}}(\rho) - \mathbf{K}\tilde{\mathbf{v}}(\rho). \tag{6}$$

The probability that NPV will be higher than a critical value $NPV(ctitical)$ at the critical demand should be higher than 0.995 if we follow Solvency II requirements:

$$NPV(ctitical) = \mathbf{p}(\left(\mathbf{I} - \tilde{\mathbf{H}}^{\omega \mathbf{d}}(\rho)\right)\tilde{\mathbf{P}}(\rho) - \Delta) - \mathbf{K}\tilde{\mathbf{v}}(\rho)). \tag{7}$$

Following Solvency II requirement, we put the demand that the quantitative requirements of amount q to be reserved (i.e., the amount of capital in combination with the additionally carefully reserved inventories of a total supply chain should have reservations to cover at least one year costs with probability 0.995) should hold. Where r is the effective interest rate in theperiod for which the reservation is made (according to Solvency II), it is the effective interest rate per year. Therefore, r is the effective interest rate per year.

$$q = (\mathbf{p}\tilde{\mathbf{H}}^{\omega \mathbf{d}}(\rho)\tilde{\mathbf{P}}(\rho) + \mathbf{K}\tilde{\mathbf{v}}(\rho))\frac{r(\rho)}{1 + r(\rho)}. \tag{8}$$

Shortening supply chains means not only moving manufacturing or sourcing closer to existing markets but also developing markets in the low-cost countries where manufacturing or sourcing takes place. Shorter supply chains influence $\tilde{\mathbf{H}}^{\omega \mathbf{d}}$, which implies more agility, more robustness against disruption, lower exchange rate risk and, in the long run, lower costs.

3 NUMERICAL EXAMPLES

Let us take a numerical example of the production part of the supply chain described in Bogataj and Grubbström (2011b). Activity cell D assembles 2 units of E and 1 unit of F; activity cell B demands 3 units of D for the production of 1 unit of B, and at the end, A demands 1 unit of B and 2 units of C for the production of 1 unit of A. The BOM of this example is presented in fig. 2 of Bogataj and Grubbström (2013) . The average production lead times $\bar{\tau}$ and values, which will not be exceeded with probability 0,005, $\tau(0,995)$ at nodes from A to F, are as follows: τ_A=3/4, τ_B=4/5, τ_C=3/5, τ_D=2/6, τ_E=2/3, τ_F=1/3

According to the BOM, we can determine the generalised input matrices using production and transportation averages of delays at *s* equal to continuous interest rate ρ as follows:

$$\mathbf{H}^{\omega}\tilde{\boldsymbol{\tau}}(s=\rho) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1e^{4\rho} & 0 & 0 & 0 & 0 & 0 \\ 2e^{3\rho} & 0 & 0 & 0 & 0 & 0 \\ 0 & 3e^{2\rho} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2e^{3\rho} & 0 & 0 \\ 0 & 0 & 0 & 1e^{1\rho} & 0 & 0 \end{bmatrix} \begin{bmatrix} e^{3\rho} & 0 & 0 & 0 & 0 & 0 \\ 0 & e^{4\rho} & 0 & 0 & 0 & 0 \\ 0 & 0 & e^{3\rho} & 0 & 0 & 0 \\ 0 & 0 & 0 & e^{2\rho} & 0 & 0 \\ 0 & 0 & 0 & 0 & e^{2\rho} & 0 \\ 0 & 0 & 0 & 0 & 0 & e^{1\rho} \end{bmatrix}$$

For extreme perturbations at each stage (exceeded with probability less than 0.005 at each activity cell and each link), we have the following:

$$\mathbf{H}^{\omega\mathbf{d}}\tilde{\boldsymbol{\tau}}(s=\rho)=\begin{bmatrix}0&0&0&0&0&0\\1e^{10\rho}&0&0&0&0&0\\2e^{9\rho}&0&0&0&0&0\\0&3e^{8\rho}&0&0&0&0\\0&0&0&2e^{12\rho}&0&0\\0&0&0&1e^{10\rho}&0&0\end{bmatrix}=\begin{bmatrix}0&0&0&0&0&0\\1.92&0&0&0&0&0\\3.59&0&0&0&0&0\\0&5.05&0&0&0&0\\0&0&0&4.36&0&0\\0&0&0&1.92&0&0\end{bmatrix}$$

Taking time spending distance in account, the longest path among the simple paths in the given graph is $\tau_E-\tau_{ED}-\tau_D-\tau_{DB}-\tau_B-\tau_{BA}-\tau_A$. The corresponding average total time-spending distance is 2+3+2+2+4+4+3=20 time units, and the variance at $z(0.005)=2.58$ equals 4.96:

$$\sigma^2=(\frac{1}{2.58})^2(1^2+1^2+4^2+1^2+3^2+2^2+1^2)=\frac{33}{6.656}=4.96$$

if delays are normally distributed and independent. If the data are from extreme value distributions, it would be different, but for the sake of simplicity, we suppose that the distribution of perturbed delay is normal.

On this longest path, $\tau_E^d+\tau_{ED}^d+\tau_D^d+\tau_{DB}^d+\tau_B^d+\tau_{BA}^d+\tau_A^d=3+6+6+3+7+6=31$

In this case, the probability that the time-spending distance from E to A will exceed 31 is equal to $\alpha(z=(31-20).\sqrt{4.96})=\alpha\left(z=24.5\right)<0.0001$. We can see that if the critical perturbation of delay is determined by individual activities in the supply chain at $\alpha=0.005$, the probability that total delay will exceed the sum of critical values is negligible. We set values for the price vector $\mathbf{p}=[560,38,25,34,14,15]$, setup cost parameters $\mathbf{K}=[200\quad 180\quad 210\quad 195\quad 175\quad 215]$, $\tilde{\mathbf{t}}(s)$ and $\tilde{\mathbf{T}}(s)$ of the same values, as in Bogataj and Grubbström, (20112, 2013), such that for $\tilde{\mathbf{P}}_0$ realisation $\tilde{\mathbf{P}}(s)$ and continuous interest rate $\rho=0.065$, we obtain the following:

$$\hat{\mathbf{P}}_0=\begin{bmatrix}100\\100\\200\\300\\600\\300\end{bmatrix},\quad \tilde{\mathbf{P}}_0(s)=\tilde{\mathbf{t}}(\rho)\tilde{\mathbf{T}}(\rho)\hat{\mathbf{P}}_0=\begin{bmatrix}38.4\\54.4\\168.4\\274.5\\849.9\\461.8\end{bmatrix}\qquad \tilde{\mathbf{v}}(\rho)=\begin{bmatrix}e^{-\rho t_1}/\left(1-e^{-\rho T_1}\right)\\\vdots\\e^{-\rho t_6}/\left(1-e^{-\rho T_6}\right)\end{bmatrix}=\begin{bmatrix}0.384\\0.544\\0.842\\0.915\\1.417\\1.539\end{bmatrix}$$

Using equation (6) for the net present value (*NPV*) of production activities, we can now calculate *NPV* with transportation time delays are included as follows:

$$\mathrm{NPV}=\mathbf{p}\left(\mathbf{I}-\tilde{\mathbf{H}}^{\omega}(\rho)\right)\tilde{\mathbf{P}}_0(\rho)-\mathbf{K}\tilde{\mathbf{v}}(\rho)=25{,}858$$

The value 25,858 does not include any transportation costs. Here we have got:

$$\mathbf{p}\tilde{\mathbf{H}}^{\omega}(\rho)\tilde{\mathbf{P}}_0(\rho)+\mathbf{K}\tilde{\mathbf{v}}(\rho)=30{,}082$$

Let us assume that the transportation costs are 10,000 such that the NPV of total costs is 40,082. Therefore, if we have solvency capital requirement q for the case of market

disruption in the next year according to (8), where at a continuous interest rate equal to 0.065, the effective interest rate $r(\rho)$ per year is 0.067, we can write the following:

$$q = (\mathbf{p}\tilde{\mathbf{H}}^{\omega}(\rho)\tilde{\mathbf{P}}(\rho) + \mathbf{K}\tilde{\mathbf{v}}(\rho) + 10000)\frac{r(\rho)}{1+r(\rho)} = 40082\frac{0.067}{1.067} = 2517$$

If the time delays appear as described, then the NPV including transportation costs *TC* is

$$\text{NPV} = \mathbf{p}\left(\mathbf{I} - \tilde{\mathbf{H}}^{\omega b}(\rho)\right)\tilde{\mathbf{P}}_0(\rho) - \mathbf{K}\tilde{\mathbf{v}}(\rho) - TC = 50045.01 - 38485.85 - 1000 = 10560.16$$

We can see that such a perturbed system has only 40.8 % of the NPV of the system that is not perturbed. However, a perturbed system that operates on time delay critical values also needs higher capital requirements q^d to have a 99.5% probability of surviving at least one year

$$q^d = (\mathbf{p}\tilde{\mathbf{H}}^{\omega d}(\rho)\tilde{\mathbf{P}}(\rho) + \mathbf{K}\tilde{\mathbf{v}}(\rho) + 1,000)\frac{r(\rho)}{1+r(\rho)} = (41{,}338 + 10{,}000)\frac{0.067}{1.067} = 3{,}224$$

4 CONCLUSION

We can see that such a system needs 28% higher solvency capital requirements (i.e., liquid assets + financial derivatives + borrowing capacity). It is clear that extended EMRP with an input-output matrix, which includes appropriate timing and straightforward presentation in a Laplace-transformed space, enables higher visibility of flows and functioning of studied supply chains and generally could show how perturbations in delays influences costs and added values. The NPV approach can also manifest that supply chains are viable as long as NPV is positive. During the recent financial crises, financial authorities, as regulators of the financial system, which should service the real economy, have enabled global supply chains to survive by drastically lowering the interest rate (0.25% in the USA, which is historically the lowest since introducing central banks in western economies) and adding liquidity to the system. These results can be subjected to further sensitivity analysis using either local methods (Borgonovo and Peccati, 2004) or global methods (Borgonovo and Peccati, 2009). Therefore, using EMRP, we can analyse simultaneous perturbations in timing and intensity of flows. The approach in frequency domain enables critical values to be better estimated in case of interactions of different risk drivers that are likely to disrupt the procurement, production, transportation, warehousing, delivery and financing of a good or service.

References

[1] Borgonovo, E. , Peccati, L. (2004) Sensitivity analysis in investment project evaluation Int J Prod Econ, 90 (1), p.17-25.

[2] Borgonovo, E. , Peccati, L. (2009) Financial management in inventory problems: Risk averse vs risk neutral policies . Int J Prod Econ 118 (1), p.233-242.

[3] Bogataj, M., Grubbström, R.W. (2012) On the representation of timing for different structures within MRP theory. Int J Prod Econ. 140 (2), p.749-755.

[4] Bogataj, M., Grubbström, R.W. (2013) Transportation delays in reverse logistics. Int J Prod Econ, 143 (2), p.395-402.

[5] Bogataj M, Grubbström RW, Bogataj L (2011) Efficient location of industrial activity cells in a global supply chain. Int J Prod Econ 133(1), p. 243–250.

[6] European parlament (2009), Directive of the European Parliament and of the Council on the taking-up and pursuit of the business of insurance and reinsurance (Solvency II), PE-CONS 3643/6/09, REV 6, SURE 15, ECOFIN 349, CODEC 693.

[7] Grubbström, R.W.,1998. A net present value approach to safety stocks in planned production. Int J Prod Econ, 56-57 p.213-229.

DISCOVERING FRAUD IN LEASING AGREEMENTS: DATA MINING APPROACH

Ivan Horvat
VB Leasing d.o.o.
Horvatova 82, HR-10000, Zagreb, Croatia
ivan.horvat@vbleasing.hr

Mirjana Pejić Bach
Faculty of Economics & Business – Zagreb, University of Zagreb, Department of Informatics
Trg J.F. Kennedyja 6, HR-10000 Zagreb, Croatia
mpejic@efzg.hr

Marjana Merkač Skok
Fakulteta za poslovne in komercijalne vede
Lava 5,Celje, Slovenia
Pavlinska 2, HR-42000 Varaždin, Croatia
marjana.merkac@fkpv.si

Abstract: Fraud attempts create large losses for financing subjects in modern economies. Leasing agreements have become more and more popular as a means of financing objects such as machinery and vehicles. The goal of the paper is to estimate the usability of the data mining approach in disovering fraud in leasing agreements. Real-world data from one of the Croatian leasing firms was used for creating tow models for fraud detection in leasing. The decision tree method was used for creating a classification model, and the CHAID algorithm was deployed.

Keywords: leasing, fraud, data mining, classification, decision tree, CHAID

1 INTRODUCTION

Leasing is a modern financing method developed in the U.S.A. in the 30s of the last century, and has been widely accepted and applied in the world from 1950s onwards. Leasing allows the user to use needed equipment or property for a required period of time, rather than to buy it. A leasing object is a movable or an immovable thing in accordance with the applicable rules governing property or other proprietary rights [6].

A leasing agreement becomes realized and active after being signed by a leasing company and a customer. There is no delay in activation or conditional activation of the agreement. There are two main ways in which a leasing agreement can be terminated: the expiration of the agreement and the premature termination. The circumstances that lead to an early termination can be devided into the circumstances caused by users of the lease (total loss, failure to pay monthly installments) and the circumstances caused by external influences (theft, total loss due to natural disasters).

If the agreement is terminated and the attempt to perpertrate fraud or deception is found, the damage for a leasing house is created. Therefore, risk management and using credit scoring are important levers for increasing the security of a leasing company. Advanced analytical methods of assessing the risk of fraud have proved successful in predicting one of the two possible outcomes of the agreements: a successful implementation and finalization of the agreement and an attempted fraud [4]. However, in previous studies, leasing has not been the subject of modeling knowledge discovery from databases, although the method is often used in practice. Therefore, the aim of the paper is to develop a model for detecting fraud in the lease, using actual data from a leasing company. To achieve the objective, knowledge discovery from databases was used and the decision tree method was applied [5].

2 METHODOLOGY

2.1. Data

The used database contains information on all leasing agreements and offers in the core system on the date of running the report. The number of active or completed agreements at the time of running the report was 25,000. In the same period a total of 561 agreements in which fraud was realized was found. In order to ensure the possibility of forming a decision tree model, the method of undersampling was used and 560 agreements with no fraud attempts were randomly selected from the total number of observed agreements.

Although the database contains more than a hundred variables, due to the confidentiality of data, selected variables are sufficiently general in character and do not disclose protected information about leasing customers, suppliers and employees, while at the same time they are specific enough to be important for the realization of the model. Table 1 contains the variables used in the discovery of knowledge from databases.

Table 1: The variables used in the discovery of knowledge from databases

Variable / Type of variable	Modalities*
Type of lease / Categorical	Finance lease (68.3%); Operating lease; (30.8%); Loans (0.9%)
Type of client / Categorical	Natural person (5.3%); Crafts (25.5%); Legal entity (69.2%)
Source of initial information / Categorical	Direct contact with the client (16.9%); E-business contract (0.4%); Contract concluded by dealers (57.3%); Other (0.4%); No answer (25.1%)
Object classification 1 / Categorical	GF1 = Passenger cars and light commercial vehicles (58.3%); GF2 = Commercial vehicles (21.8%); GF3 = Machinery and equipment (18.7%)
Object classification 2 / Categorical	More detailed object description. e.g. Construction equipment Industrial equipment; IT equipment; Trucks and towing trucks
New or used / Binomial	New (62.5%); Used (36.3%)
Company size / Categorical	Small (87.4%); Medium (5.5%); Large (1.8%); Natural persons (5.3%)
Fraud / Binomial	No fraud (50%); Fraud (50%)

* In cases when the sum is smaller than 100%, there were missing data.

2.2. Decision trees

Decision trees are a popular and widely accepted tool for classification and prediction, and their strength is reflected in the fact that they are easily understandable due to a graphical display [1]. A decision tree is a statistical method of pattern recognition which is used to solve problems with predictive nature while monitoring the learning process is needed. Predictive problems include forecasting values in the future, pattern recognition, regression of multiple features, the differential analysis, evaluation functions of more features and supervised learning. Decision trees are very efficient when dealing with large databases and when many variables should be taken into account [2].

The paper used the CHAID algorithm for trees to detect fraud in the leasing agreements, since this algorithm is suitable for classification problems where the variables have more than two modalities [3]. The paper uses the software package SPSS, ver. 19th, and two types of models have been developed: (i) Model A: the model with a simpler classification of leased assets (the variable Object classification 1) and (ii) Model B: the model with a complex classification of leasing involving facilities (the variable Object classification 2).

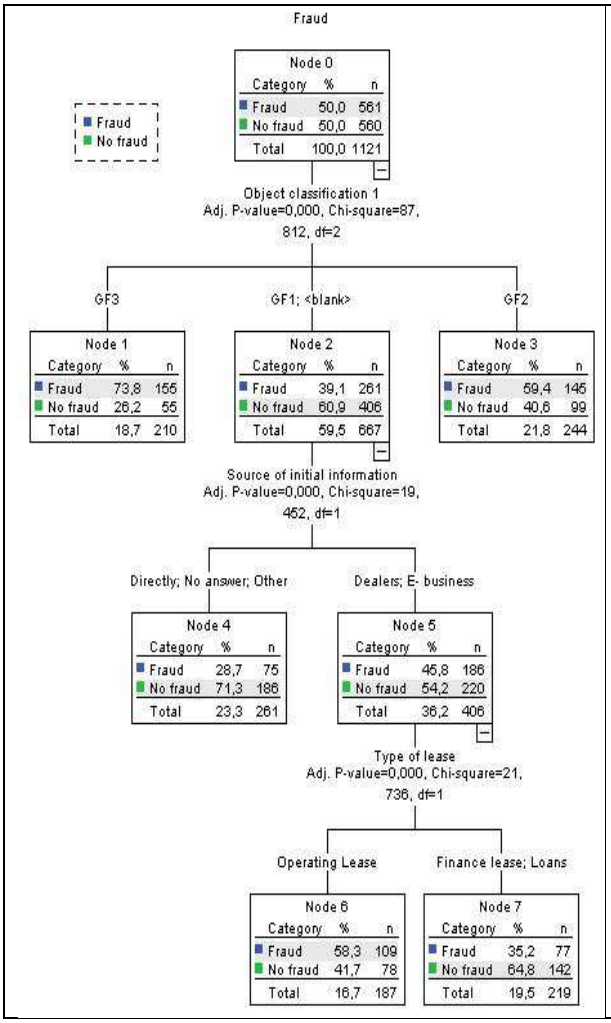
3 RESULTS

Table 2 shows the specification and the results of both models (Model A and Model B). The method used for growing both models is CHAID. The dependent variable is Fraud, while candidate independent variables are the same for both models which differ in object classifications variables. The cross validation approach has been used for validation of the model. The algorithm was applied with the following restrictions: the maximum tree depth (3 levels), the maximum cases in parent node (100 cases), and the maximum number of cases in child node (50 cases).

Table 2: Basic information on the specifications and model results

Specifications	Growing Method	CHAID
	Dependent Variable	Fraud
	Independent Variables	Type of lease, Type of client, Source of initial information, Object classification 1 (Model A), Object classification 2 (Model B), New or used, Company size
	Validation	Cross Validation
	Maximum Tree Depth	Both Model A and Model B (3 levels)
	Minimum Cases in Parent Node	Both Model A and Model B (100 Cases)
	Minimum Cases in Child Node	Both Model A and Model B (50 Cases)
Results	Independent Variables Included	Model A: Object classification 1, Source of initial information, Type of lease Model B: Object classification 2, Source of initial information, Type of lease
	Number of Nodes	Model A (8 Nodes); Model B (10 Nodes)
	Number of Terminal Nodes	Model A (5 Terminal Nodes); Model B (7 Terminal Nodes)
	Depth	Both Model A and Model B (3 levels)

Model A will be described in greater detail. The variable used for branching on the first level is Object 1, which is statistically significant with a level of 1% probability (P-value = 0.000). Second level nodes show branching variables Object 1 at three knots. Node 1 (node1) contains 210 data for which the average value of the variable Fraud is 0.738, which means that 73.8% of the agreements for which the subject of the agreement is GF3 resulted in fraud. Node 2 has 667 agreements for which the average value of the variable Fraud is 0.391, which means that 39.1% of the agreements for the GF1 and the unknown object contracting resulted in fraud. In the same way we interpret the third node. The variable for branching on the second level is Source of information, which is statistically significant with a probability level of 1% (p-value = 0.000). Third-level nodes show the branching variable Source of information on the two nodes. Node 4 shows the clients who come directly to the leasing company or or the source of initial information is not available. This node contains 261 agreements with the average value of 0.287, which means that 28.7% of the agreements resulted in fraud. Node 5 shows clients who are contracted through the dealer or the manufacturer, and via the Internet (only a small share). The average value of this node is 0.458, meaning that 45.8% of the agreements resulted in fraud. The variable used for branching on the third level is Type of leasing, which is statistically significant with a probability level of 1% (p-value = 0.000). Node 6 contains agreements of operating lease, where the average agreement value is 0.583, meaning that 58.3% of the agreements resulted in fraud. Node 7 includes financial leasing and loans, where the average agreement value is 0.352, meaning that 35.2% of the agreements resulted in fraud.




```
STRING pre_001 (A8).

/* Node 1 */.
DO IF (Object classification 1 EQ "GF3"). COMPUTE nod_001 = 1.
COMPUTE pre_001 = 'Fraud'. COMPUTE prb_001 = 0.738095.
END IF. EXECUTE.

/* Node 4 */.
DO IF (Object classification 1 NE "GF3" AND Object
classification 1 NE "GF2") AND (Source of initial information
EQ "Directly" OR Source of initial information EQ "No answer"
OR Source of initial information EQ "Other"). COMPUTE nod_001
= 4. COMPUTE pre_001 = 'No fraud'. COMPUTE prb_001 =
0.712644.
END IF. EXECUTE.

/* Node 6 */.
DO IF (Object classification 1 NE "GF3" AND Object
classification 1 NE "GF2") AND (Source of initial information
NE "Directly" AND Source of initial information NE "No answer"
AND Source of initial information NE "Other") AND (Type of
lease EQ "Operating Lease"). COMPUTE nod_001 = 6. COMPUTE
pre_001 = 'Fraud'. COMPUTE prb_001 = 0.582888. END IF.
EXECUTE.

/* Node 7 */.
DO IF (Object classification 1 NE "GF3" AND Object
classification 1 NE "GF2") AND (Source of initial information
NE "Directly" AND Source of initial information NE "No answer"
AND Source of initial information NE "Other") AND (Type of
lease NE "Operating Lease"). COMPUTE nod_001 = 7. COMPUTE
pre_001 = 'No'+ ' fraud'. COMPUTE prb_001 = 0.648402. END IF.
EXECUTE.

/* Node 3 */.
DO IF (Object classification 1 EQ "GF2"). COMPUTE nod_001 = 3.
COMPUTE pre_001 = 'Fraud'. COMPUTE prb_001 = 0.594262.
END IF. EXECUTE.
```

Figure 1: Decision tree generated with a more aggregate object classification (Object classification 1) and SQL code generated (Model A)

Model B will be described in greater detail in the following text. The variable used for branching on the first level is Object 2, which is statistically significant with a level of 1% probability (P-value = 0.000). Second level nodes are showing branching variables Object 2 at five knots. Node 1 (node1) contains 239 data for which the average value of the variable Fraud is 0.561, which means that 56.1% of the agreements for which the subject of the agreement is other equipment, trucks, busses and machines resulted in fraud. Node 2 has 151 agreements for which the average value of the variable Fraud is 0.728, which means that 72.8% of the agreements including a wide selection of equipment, machines and boats resulted in fraud. Node 3 has 450 agreements for which the average value of the variable Fraud is 0.420, which means that 42.0% of the agreements including passenger cars resulted in fraud. Node 4 has 63 agreements for which the average value of the variable Fraud is 0.889, which means that 88.9% of the agreements including farming machines, machines for processing plastics and cosmetic industry resulted in fraud. In the same way we interpret the fifth node. This node has 218 agreements for which the average value of the variable Fraud is 0.330, which means that 33.0% of the agreements including light commercial vehicles resulted in fraud. The variable for branching on the second level is Source of information, which is statistically significant with a probability level of 1% (p-value = 0.000). Third-level nodes show the branching variable Source of information on the two nodes. Node 6 shows the clients who come directly to the leasing company or the source of initial information is not available. This node contains 165 agreements with the average value of 0.297, which

means that 29.7% of the agreements resulted in fraud. Node 7 shows clients who are contracted through the dealer or manufacturer, and via the Internet (only a small share). The average value of this node is 0.491, meaning that 49.1% of the agreements resulted in fraud. The variable used for branching on the third level is Type of leasing, which is statistically significant with a probability level of 1% (p-value = 0.000). Node 8 contains 146 agreements of operating lease, where the average agreement value is 0.582, meaning that 58.2% of the agreements resulted in fraud. Node 9 includes financial leasing and loan and, contains 139 agreements where the average agreement value is 0.396, meaning that 39.6% of the agreements resulted in fraud.

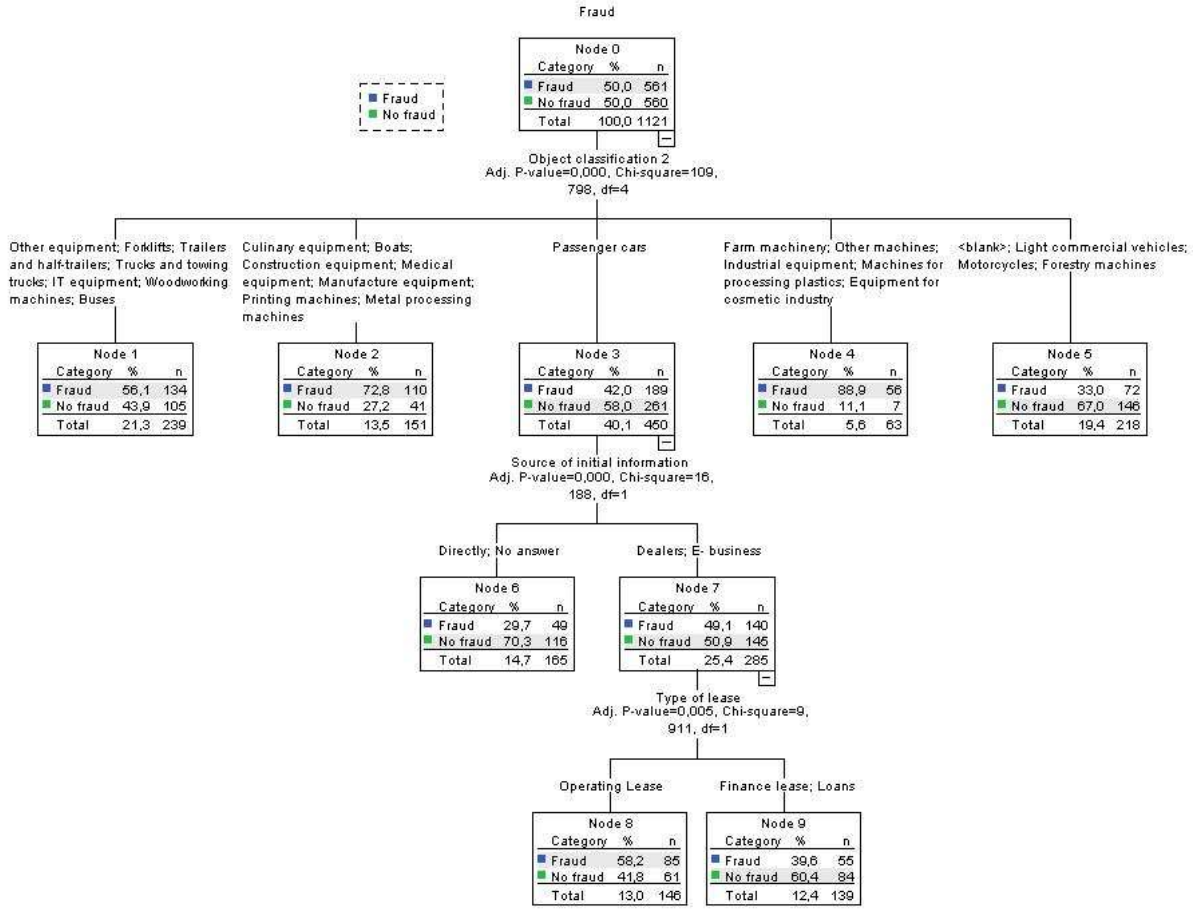


Figure 2: Decision tree generated with a more aggregate object classification (Object classification 1) (Model B)

Table 3: Classification matrixes for Model A and Model B

	Predicted					
	Fraud		No fraud		Percent Correct	
Observed	Model A	Model B	Model A	Model B	Model A	Model B
Fraud	409	385	152	176	72.9%	68.6%
No fraud	232	214	328	346	58.6%	61.8%
Overall Percentage	57.2%	53.4%	42.8%	46.6%	65.7%	65.2%

Table 3 presents classification matrixes for both Model A and Model B. Suprisingly, Model A is more accurate in predicting fraud, although it uses a more aggregate object classification. Comparison of these models leads to the conclusion that fraud is likely to

happen on Object1 - GF3 group, i.e. in the case of Model B – equipment and machinery. This is understandable since these objects of lease have greater value compared to other groups. The logic behind this is that if criminals are going to perpertrate fraud, they will try to maximize the effect. Models also show that firms should be more careful with agreements that come from dealers as there is a higher possibility of fraud. Implementing one of these models or one of their variations would create a good system for fraud detection and could create positive effects on business of a lease company. Implementation of such a solution should be made troughout the industry as a security standard.

4 CONCLUSION

Introduction of this model in the business would certainly show that certain frauds could be prevented and would indicate the leasing agreements which present a fraud risk. However, to make this project come to life, it would be necessary to develop software that would enable automated, quick and transparent retrieval of data from the system, processing according to the rules and displaying the results in multiple categories. It would be necessary to show already existing fraud events, fraud events that are emerging and potential fraud events so that for each of these categories an appropriate action could be taken.

The solution could be implemented into the current environment through the existing SQL-based applications by developing a separate module. In this case, it would be necessary to employ the original developers to integrate the module within the existing application to set up an alarm system. This is probably the best solution because the program would be incorporated into the existing central application enabling full access to all data in the core system, regardless of the period. According to similar projects, the estimated costs of the development of these modules would be at the level of approximately 15,000 EUR. This estimation is based on the market research conducted for the leasing firm used for the case study. Prevention of even a single case of fraud would prove the purposefulness of this project since instances of fraud in most cases involved expensive leasing objects. Prevention of fraud events results not only in savings connected with the value of lease agreements, but also results in a number of other positive externalities. The accounts receivable department has one less difficult case to handle, there is no need to pay the costs of interventions for finding fraud subjects of leasing and eventually significant legal costs and the costs of hiring legal services staff are avoided.

References

[1] Apté, C., Weiss, S., (1997). Data mining with decision trees and decision rules, Future Generation Computer Systems, Vol. 13, No. 2–3, pp. 197-210.

[2] Li, X-B. (2005)., A scalable decision tree system and its application in pattern recognition and intrusion detection, Decision Support Systems, Vol. 41, No. 1., pp.112-130.

[3] McCarty, J.A., Hastak, M., (2007). Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression, Journal of Business Research, Vol. 60, No. 6, pp.656-662.

[4] Ngai, E.W.T., Hu, Y., Wong, Y.H., Chen, Y., Sun, X., (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature, Decision Support Systems, Vol. 50, No. 3, pp. 559-569.

[5] Sinha, A.T., Zhao, H., (2008). Incorporating domain knowledge into data mining classifiers: An application in indirect lending, Decision Support Systems, Vol. 46, No. 1, pp. 287-299.

[6] Smith, C., Wakeman, L., (1985). Determinants of corporate leasing activity, Journal of Finance, Vol. 40, No. 3, pp. 895-911.

PRICE SENSITIVITY IN MULTI-LEVEL ASSEMBLY SYSTEMS: CASE STUDY OF SPANISH BABY FOOD COMPANY

Danijel Kovačić[a], **Eloy Hontoria**[b] and **Lorenzo Ros-McDonnell**[b]
[a] Danijel Kovačić s.p., Informacijsko svetovanje, Grič 4, SI-1000 Ljubljana, Slovenia
[b] GIO Universidad Politécnica de Cartagena, 30202 Cartagena, Spain
kovacic.danijel@gmail.com, eloy.hontoria@upct.es, lorenzo.ros@upct.es

Abstract: The current economic crisis, together with climate changes and exponential growth of the world population, is reflected in volatile prices of agricultural products. Prices are also strongly connected with the quality of agricultural products, which largely affect food production business in all aspects. Extended Material Requirements Planning (EMRP) Theory has proved to be capable of thoroughly analyzing entire supply chains. Price vector plays a crucial role in Net Present Value calculation, which gives us a strong background for financial evaluation of investment decisions. In this paper, we show the importance of ingredients' prices and quality in multi-level assembly systems on a real case study of a Spanish baby food company, using the principles of well-developed EMRP Theory.

Keywords: Extended MRP Theory (EMRP), Input-Output analysis, Net Present Value (NPV), simulations, multi-level assembly systems, food production.

1 INTRODUCTION

Material Requirements Planning (MRP) is well known from managing production processes, covering both production planning and inventory management [12]. It is well established in practice since most of the multi-level production systems are operated using MRP's obvious advantages. Moreover, strong technical background of MRP makes it an exceptionally good basis for deeper scientific research for which the term MRP Theory is established [5]. For the purpose of scientific observation, structures from the Bill Of Materials (BOM) can be conveniently captured within a pair of input and output matrices **H** and **G** [11]. To these structures, lead times can be assigned using Laplace transform theorems. This allows us to evaluate cash flows with the use of Net Present Value (NPV) calculation. Detailed review of the MRP Theory and its background can be found in [7].

Recently, MRP Theory was also recognized as a very useful method for studies of entire supply chains, covering not only production but also distribution, consumption and recycling processes [6]. These 4 subsystems create a closed loop with known structures where significant lead times can usually be expected (Figure 1).

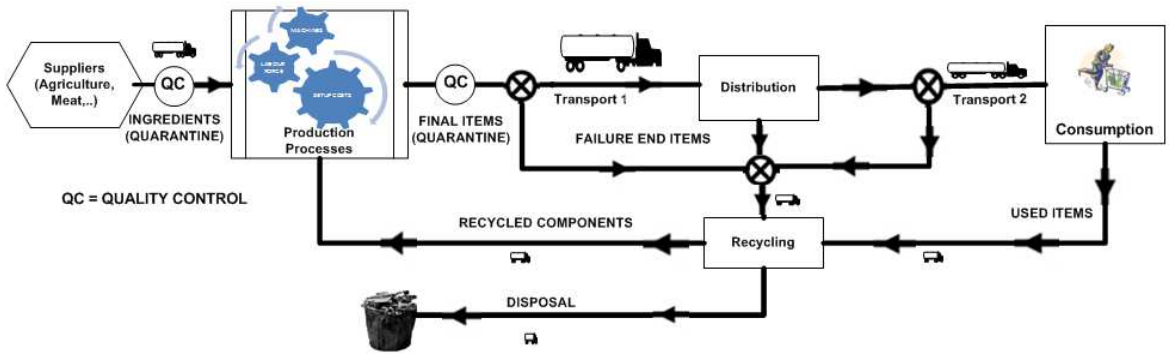


Figure 1: Involved subsystems: production, distribution, consumption and recycling.

Such systems can be scientifically researched using the so called Extended MRP (EMRP) Theory [2]. Detailed structures of input and output matrices **H** and **G** of such a complex system can be found in [8]. For proper modeling of arborescent subsystems (i.e. distribution or recycling), which are integral parts of global supply chains and cyclical processes inside them, generalized output matrix $\breve{\mathbf{G}}(s)$ is introduced [1]. Further, lead times appearing inside or between any pair of activity cells in the grid are recognized as an important factor determining economic viability of the system [3]. Additionally, many other technological and environmental parameters, such as energy and environmental taxes, can be introduced into the system, giving us strong analytical tool for various researches of supply chains [9].

A strong theoretical basis of the MRP Theory is also capable of solving real world problems. Practical application was presented for the first time by Grubbström in 1990 for analyzing production processes in a paper mill [4]. Recently, Extended MRP Theory was used for modeling production processes in a baby food company located in Spain, with special emphasis on residues of the production process and associated environmental costs [10]. This paper extends previously presented work with a further study related to price sensitivity of baby food jars. We show how EMRP Theory can be used in evaluating the risks of fluctuating prices of ingredients. EMRP Theory can help in decision making process when optimal balance between quality and price are being evaluated.

2 MODEL AND PRICE SENSITIVITY SIMULATIONS

In this paper, we further develop the previously presented model of a baby food company [10]. The company is located in Spain, and most of its final production is distributed to the domestic market. We are analyzing one of the company's many products: a 250 gram jar of baby meat food. Figure 1 presents a comprehensive structure of the product, together with all lead times where 98.9 % of total production is launched on the market at the retail price of 0.6155 €/jar. The remaining 1.1 % are residues which have to be disposed of at a cost of 0.145 €/jar.

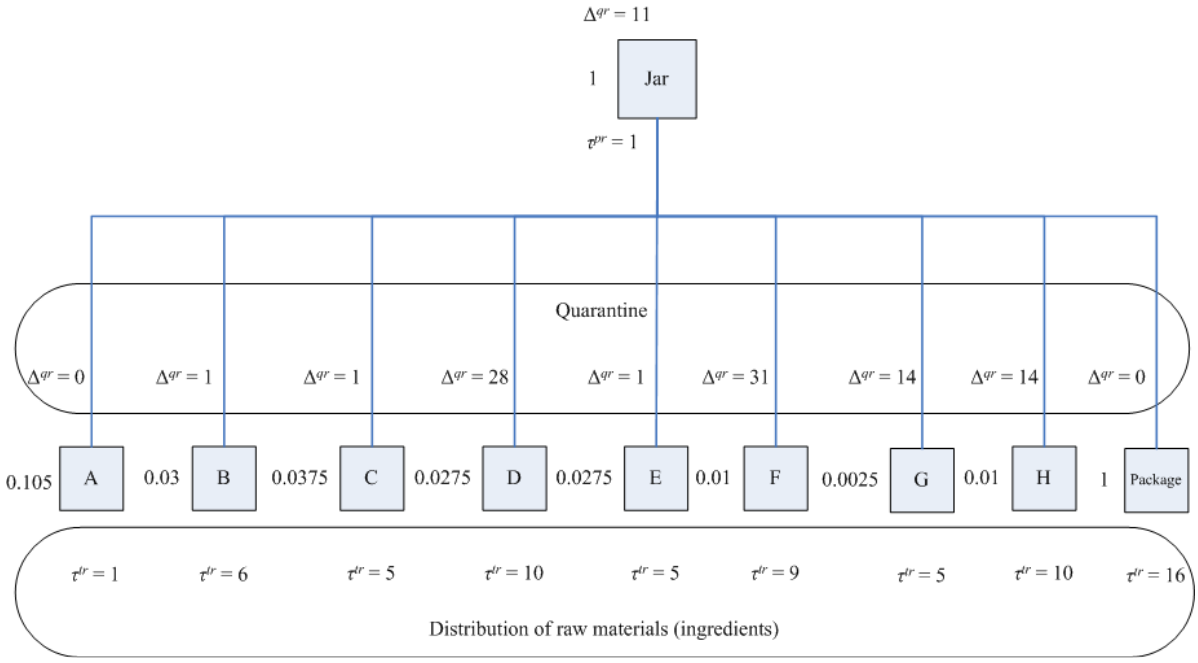


Figure 2: Structure (BOM) of the final product (jar of baby food) together with distribution, quarantine and production lead times.

According to the BOM presented in Figure 2 and the EMRP input-output matrix structure, generalized input and output matrices $\breve{\mathbf{H}}$ and $\breve{\mathbf{G}}$ for production and recycling subsystem can be written as:

$$\breve{\mathbf{H}} = \left[\begin{array}{c|c} \begin{matrix} 0.105e^{\rho(1+0+1)} \\ 0.03e^{\rho(6+1+1)} \\ 0.0375e^{\rho(5+1+1)} \\ 0.0275e^{\rho(10+28+1)} \\ 0.0275e^{\rho(5+1+1)} \\ 0.01e^{\rho(9+31+1)} \\ 0.0025e^{\rho(5+14+1)} \\ 0.01e^{\rho(10+14+1)} \\ 1e^{\rho(16+0+1)} \end{matrix} & \\ \hline & \end{array}\right] \qquad \breve{\mathbf{G}} = \left[\begin{array}{c|c} 0.989e^{-\rho*11} & \\ \hline & 0.011e^{-\rho*11} \end{array}\right] \tag{1}$$

Production takes place 3 times per month in batches of 152000 jars. It takes 1 day to complete 1 batch, with setup costs of 20000.00 €. Therefore, activity vector **P** can be written as:

$$\mathbf{P} = \begin{bmatrix} 152000 \\ 152000 \end{bmatrix} \tag{2}$$

and setup costs can be captured inside vector **K**:

$$\mathbf{K} = \left[-20000 \mid 0\right] \tag{3}$$

Prices of the final product, raw materials (from A to H, respectively) and environmental tax are captured inside price vector **p**:

$$\mathbf{p} = \left[0.6155 \;\; 0.006 \;\; 0.8434 \;\; 0.5 \;\; 6.9277 \;\; 0.6988 \;\; 0.6024 \;\; 1.8072 \;\; 0.7229 \;\; 0.0843 \mid -0.145\right] \tag{4}$$

Lengths of cycles *T* are known. This allows us to calculate the aggregate NPV of the system for an infinite number of cycles. Using initiation times *t* we can calculate given timings $\tilde{\mathbf{v}}(\rho)$ as:

$$\tilde{\mathbf{v}}(\rho)=\tilde{\mathbf{t}}(\rho)\tilde{\mathbf{T}}(\rho)=\left[\begin{array}{c|c} e^{-\rho t_1} & \\ \hline & e^{-\rho t_2} \end{array}\right]\left[\begin{array}{c|c} (1-e^{-\rho T_1})^{-1} & \\ \hline & (1-e^{-\rho T_2})^{-1} \end{array}\right]=$$
$$=\left[\begin{array}{c|c} e^{-41\rho} & \\ \hline & e^{-52\rho} \end{array}\right]\left[\begin{array}{c|c} (1-e^{-10\rho})^{-1} & \\ \hline & (1-e^{-10\rho})^{-1} \end{array}\right]= \quad (5)$$
$$=\left[1213.07 \mid 1211.98\right]$$

The company can borrow money at a 3.5 % interest rate (ρ = 0.035 per year). Overall NPV of the cyclical system with an infinite number of repeating cycles can be calculated as:

$$\mathrm{NPV}=\mathbf{p}\left(\breve{\mathbf{G}}(\rho)-\breve{\mathbf{H}}(\rho)\right)\tilde{\mathbf{P}}(\rho)-\hat{\mathbf{K}}\tilde{\mathbf{v}}(\rho)=18559553.20\ € \quad (6)$$

Initial NPV for meat baby food production is positive, which indicates that system is economically viable. The company can evaluate variation of different parameters through calculation of the NPV. In multi-level production systems prices of components can drastically change the NPV. This fact is especially dangerous in a food production business due to volatile prices of agricultural products. Since agricultural products usually have to be fresh when entering the production, their long-term storage is not possible. The management of the company should prepare relevant decision making strategies for situations where parameters in the business environment change. Figure 3 shows the effect of price change of ingredient D. We can see that NPV decreases to 0.00 € when the price for 1 unit of D increases to 11.18 €. Such a rapid increase of the price of element D would make production economically unviable. In such a case, the company would have to think about changing its final product's price. On the other hand, if an increase of the price of ingredients is expected to occur in the near future, the company can use the EMRP model to find additional business strategies. For example, when the price of the ingredient is expected to increase, the company could also increase produced batch of final products. This would result in larger inventories of jars which will be used to cover future demand. This strategy could be especially relevant if increase of prices of ingredients is expected to be short-term (or seasonal). In this case, optimal balance between the prices of ingredients, batch size, setup costs and inventory holding costs has to be found, which can be achieved using the EMRP Theory model and simulations.

Further, from Figure 4 we can see mutual impact of lead times and prices of ingredients on the NPV of the system. If the company decides to compensate higher quality of ingredients (which usually is reflected through higher prices) for shorter lead times, or vice versa, it can choose solutions which are shown on the area of graph in Figure 4. Not all solutions from the graph are feasible, but it can be clearly seen that the impact of prices on the NPV is incomparably greater than the impact of the lead times. Only a slight increase of ingredients' prices would compensate significant reduction of the lead times. However, this situation might change drastically in an economic environment with higher interest rates.

Figure 3: Impact of ingredient D price change on the NPV of the system.

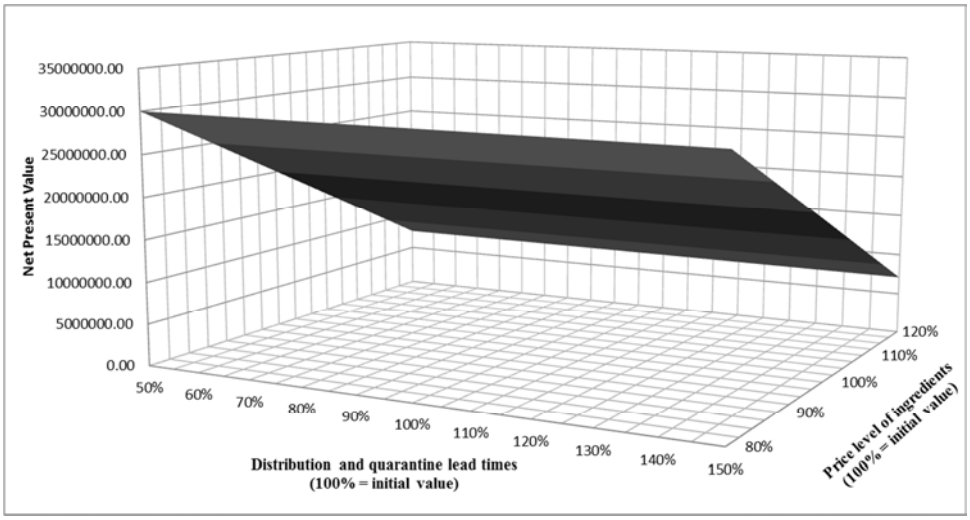


Figure 4: Impact of lead times and ingredients' price levels on the NPV of the system.

3 CONCLUSION

In this paper, we further research a Spanish baby food company's previous study by using the concepts of the Extended MRP Theory. Emphasis is given to the price component and its impact on the NPV of the whole system. We discuss some potential benefits of the EMRP Theory's approach in the company's decision making process.

References

[1] Bogataj, M., Grubbström, R.W., 2012. On the Representation of Timing for Different Structures within MRP Theory. International Journal of Production Economic, 140 (2), 749–755.

[2] Bogataj, M., Grubbström, R.W., 2013. Transportation delays in reverse logistics. International Journal of Production Economics, 143 (2), 395–402.

[3] Bogataj, M., Grubbström, R.W., Bogataj, L., 2011. Efficient location of industrial activity cells in a global supply chain. International Journal of Production Economics, 133 (1), 243–250.

[4] Grubbström, R.W., 1990. The distribution of an additive in a chemical process - an application of input-output theory. Engineering Costs and Production Economics, 19 (1-3), 333–340.
[5] Grubbström, R.W., 2007. Transform Methodology Applied to Some Inventory Problems. Zeitschrift für Betriebswirtschaft, 77(3), 297–324.
[6] Grubbström, R.W., Bogataj, M., Bogataj, L., 2007. A compact representation of distribution and reverse logistics in the value chain. Ekonomska fakulteta, Ljubljana.
[7] Grubbström, R.W., Tang, O., 2000. An Overview of Input-Output Analysis Applied to Production-Inventory Systems. Economic Systems Review, 12, 3-25.
[8] Kovačić, D., Bogataj, L., 2011. Multistage reverse logistics of assembly systems in extended MRP Theory consisting of all material flows. Central European Journal of Operations Research, 19(3), 337–357.
[9] Kovačić, D., Bogataj, M., 2013. Reverse logistics facility location using cyclical model of extended MRP theory. Central European Journal of Operations Research, 21(1), 41–57.
[10] Kovačić, D., Hontoria, E., Bogataj, M., Ros, L., 2012. Application of the Extended MRP Theory to a baby food company. In: Z. Babić (ed.), Croatian operational research review (14th International conference on operational research, Trogir, Croatia, September 26 - 28, 2012), 3, 41–51.
[11] Leontief, W., 1966. Input-output economics. New York, Oxford: Oxford University Press.
[12] Orlicky, J.A., 1975. Material Requirements Planning. McGraw-Hill, New York.

COMPARATIVE ANALYSIS OF ANNUAL REPORT DISCLOSURE QUALITY FOR SLOVENIAN AND CROATIAN LISTED COMPANIES

Snježana Pivac, Tina Vuko and **Marko Čular**
Faculty of Economics Split, University of Split
Cvite Fiskovića 5, 21000 Split, Croatia
snjezana.pivac@efst.hr, tina.vuko@efst.hr, marko.cular@efst.hr

Abstract: This paper analyzes disclosure quality of annual reports for Slovenian and Croatian listed companies, for the year 2011. Fourth and seventh EU directives require companies in both countries to provide specific disclosures through their annual reports. This paper examines the level of corporate disclosure in annual report of listing companies, by constructing appropriate disclosure quality index (DQI) and applying relevant statistical analysis. Based on conducted comparative analysis, it can be concluded that Slovenian companies have greater level of disclosure quality than Croatian.

Keywords: Annual report, Disclosure quality index, Transparency, Multi-criteria ranking, Binary logistic regression

1 INTRODUCTION

Disclosure of accurate, comprehensive and timely information is critical for the functioning of efficient capital market. The quality of information presented in annual report influences investors and other stakeholder decisions by mitigating information and incentive problems, as explained in agency theory [13]. The aim of annual report is to provide a fair review of development of the company's business and its position. Transparent presentation of information in annual report is especially important for listed companies. General consensus among financial economist is that a rich disclosure environment and low information asymmetry have many desirable consequences, like efficient allocation of resources, capital market development, market liquidity, decreased cost of capital, lower return volatility, and high analyst forecast accuracy [14]. Annual report is integrated report covering different aspects of company's financial and non-financial performance. Typically, the report consists of accounting policies, financial statements, chairman's letter, auditor's report and company's business vision for the future. While traditional business reporting model emphasized backward-looking, quantified, financial information, qualitative, forward-looking, non-financial information has generally been ignored [3]. However, such narrative sections of annual report increase the overall quality of corporate reporting and have considerable value to its users [7]. The aim of this paper is to investigate and compare the level of both mandatory and voluntary disclosures in annual reports of listed companies in Slovenia and Croatia. The remainder of the paper is structured as follows. Section two provides brief literature review on the issue. Third section discusses the institutional and economic background that is considered to be relevant for understanding potential differences in practice of annual reporting between Slovenia and Croatia. Section four describes the construction of disclosure index used for annual report disclosure quality assessment. The results of empirical research are provided in the next section. The paper ends with concluding remarks.

2 LITERATURE REVIEW

Researches about importance of annual reports, most often are related to the essential elements of annual reports. Annual report aims to communicate with users and on easy and

understandable way provide timely, reliable and relevant information on past, current and future organizational activities [6]. Research of Day and Woodward [11] has shown that if annual reports are easier to read, they have greater positive earnings. Moreover, longer annual reports lead to higher costs of information processing. Beattie and Jones [2] researched differences in graphic practice through various national accounting environments. Graphs communicate effectively through faster access to key financial indicators. Aljifri [1] creates the index of transparency reporting. The hypothesis that extent of the information varies among industries was confirmed. Mušura [17] follows reporting of Croatian listed companies and emphasizes the essential elements of annual report: management structure, auditors, shareholders' rights, code of corporate governance, business ethics, environmental management and social policy governance. Pivac and Čular [19] researched quality of annual report for Croatian listed companies in two different periods. Annual reports are of average quality, measured by quality index. They do not change significantly through the review period and a large number of key elements are missing. Many other authors emphasize the importance of the annual reports elements (Cohen [9]; Coy and Dixon [10]; Li [15]; Linsley and Shrives [16]; Santema and van de Rijt [20]). Important research of financial reporting for Slovenian and Croatian listed companies was conducted by Pervan [18] in which he states that Slovenian companies have a higher level of financial reporting. The research showed that the average level of voluntary financial reporting for the Croatian sample was almost three times lower than that in the Slovenian sample. The reasons for this difference and the backwardness of the Croatian companies are probably to be found in the overall business environment, particularly in the demand for financial information and the level of corporate governance in companies. It is important to highlight research of Garrod and Turk [12] where they point out that a company in Slovenia must present financial and non-financial elements in annual report.

3 INSTITUTIONAL AND ECONOMIC BACKGROUND

To better compare Slovenian and Croatian sample, it is useful to highlight macroeconomic indicators, the importance of the capital market, financial system, corporate governance and published information to external users, using annual report. Macroeconomic indicators show that Slovenian economy is more developed than Croatian [24]. Certainly both of them, regardless of EU membership, required a number of reforms for faster GDP growth, higher employment, better coverage of imports by exports and to decrease the growth of public debt. Looking at the Slovenian and Croatian financial systems, we conclude that both countries are similar, because of the dominating banking system over institutional investors. Compared with developed market economy countries, the capital markets in Slovenia and Croatia are still pretty undeveloped. But compared with Central and Eastern European transition countries, Slovenia and Croatia are in a good position, mainly due to the strengthening of institutional investors' role (Slovenian capital market is developed because they have significant role of the institutional investors). Observing the corporate governance, as an essential factor of development, especially in transition countries, we conclude that protection of shareholder rights is better in Slovenia than in Croatia [8]. Transparency and financial reporting are an important factor of corporate governance for listed companies. Analysis of the World Bank pointed out that Croatia is a leader in use of the International Financial Reporting Standards (IFRS) (operationally making body of accounting standards). They also point out that 'poor auditing' impairs the quality of reporting and that Croatian companies are not required to publish the entire annual report with detailed analysis [24]. Certainly, the entry of foreign investors and the positive effect of capital market increased transparency. On the other side, Slovenian companies give more disclosure of information

through the annual report, using statutory environment. In disclosure of information and transparency, voluntary reporting on Internet was more frequent in Slovenia [18]. Looking at legislation, both countries and listed companies use standards, issued by the IFRS. Looking at the official sites of Ljubljana and Zagreb Stock Exchange ([22]; [23]), we can conclude that published financial statements of Slovenian companies are more informative for users than Croatian companies. The conclusion is that Croatia lags behind Slovenia in terms of financial reporting and quality of the information provided. Croatian companies should certainly work on transparency and codes of corporate conduct.

4 DISCLOSURE QUALITY INDEX OF ANNUAL REPORT

Disclosure quality index of annual report creating (DQI) has five stages [19]. Based on the set of annual report (AR) elements, primary research is conducted (reference group from accounting and finance area) about AR elements importance. Primary research was conducted in order to ***(1) evaluate the significance of AR elements***. Score range is from 1 (AR element is not important) to 5 (AR element is extremely important). In order to get the weight, which will be used to calculate DQI, it is necessary to calculate ***(2) coefficient of AR elements importance (C.I.j).*** To make the process of creating DQI easier, it is necessary to create the amount of weight that will be from 1 (element is not significant to the AR quality) to 2 (element is extremely significant to the AR quality). The coefficient of AR elements importance (C.I.) is shown by the equation (1):

$$C.I. = \frac{\sum_{i=1}^{n} x_{ij}}{\frac{\max_{j}(\sum_{i=1}^{n} x_{ij})}{2}}, \qquad (1)$$

where is: $\sum_{i=1}^{n} x_{ij}$ the total score of the each elements importance; x_{ij} the experts assessments of the each elements importance (1-5); n number of experts (40); i an expert; j an element of AR; $C.I.$ coefficient of AR elements importance (for Croatian companies we observe 44 AR elements, while for Slovenian companies we observe 43 AR elements). The next step in calculating DQI, refers to the ***(3) assessment quality of AR (A.Q.j)***. To obtain A.Q.j it is necessary to know the individual persistence of AR elements (1-element exist in AR; 0-element does not exist in AR). To reach the DQI, it is necessary to calculate ***(4) overall quality of AR,*** which is the sum of the assessment quality of AR. Finally, ***(5) disclosure quality index of annual report (DQI)*** is defined by the following expression (2):

$$DQI = \frac{OVERALL\ \ QUALITY\ \ OF\ \ AR}{\max OVERALL\ \ QUALITY\ \ OF\ AR} \cdot 100. \qquad (2)$$

AR quality may be: poor quality AR (DQI 0-20), low quality AR (DQI 21-40), average quality AR (DQI 41-60), sufficient quality AR (DQI 61-80) and high quality AR (DQI 81-100).

5 EMPIRICAL RESEARCH

Empirical research includes randomly selected companies from Ljubljana (n=30) and Zagreb Stock Exchange (n=30) ([22]; [23]). For each of companies, we observed all the important

annual report elements. The observed elements are presented in Table 1. Also, Table 1 shows the importance of observed elements, as well as the persistence of the same in Slovenia and Croatia, reported in relative values (% companies of observed element). Using DQI, Table 2 shows that Slovenian companies have sufficient quality and high quality annual reports, which is not the case for Croatian companies, because it is only 10% of sufficient quality annual reports, while others have DQI less than 60.

Table 1: Important elements of annual report with the importance weights and the number of companies that have annual report element in Slovenia and Croatia

ELEM.*	***C.I.j***	% SLO	% CRO	**ELEM.***	***C.I.j***	% SLO	% CRO
1	1,64	100	53	23	1,46	97	17
2	1,36	100	70	24	1,26	93	23
3	1,48	100	43	25	1,36	67	10
4	1,02	57	30	26	1,36	63	7
5	1,46	100	67	27	1,38	60	13
6	1,58	90	10	28	1,48	63	13
7	1,48	70	13	29	1,46	77	10
8	2	93	17	30	1,52	57	7
9	1,22	70	33	31	1,5	80	3
10	1,18	*	10	32	1,6	97	67
11	1,18	97	63	33	1,6	90	23
12	1,22	90	63	34	1,7	97	63
13	1,24	100	23	35	1,8	97	83
14	1,34	83	10	36	2	100	93
15	1,48	87	13	37	2	100	93
16	1,34	73	3	38	2	100	97
17	1,54	70	3	39	2	93	97
18	1,52	77	0	40	2	100	83
19	1,6	83	0	41	1,68	87	83
20	1,4	87	47	42	2	100	70
21	1,64	87	20	43	1,66	90	40
22	1,44	50	7	44	1,16	100	60

***ELEMENTS:** 1. Executive summary and key financial indicators; 2. Company profile; 3. Business activities; 4. Short company review; 5. The most important achievements in the reporting year; 6. Position on the business market; 7. Report by the Supervisory Board; 8. Report by the Management Board; 9. Code of Corporate Governance; 10. Annual survey of the CCG; 11. Management Board members; 12. Members of the Supervisory Board; 13. The Authority of company bodies; 14. Organizational structure; 15. Expectations for future periods; 16. The mission and vision of the company; 17. Corporate strategy; 18. Relationship to stakeholder group – customers; 19. Relationship to stakeholder group – shareholders; 20. Major shareholders; 21. The report on the movement of companies shares; 22. Relationship to stakeholder group – suppliers; 23. Relation to employees; 24. Structure of employees; 25. Corporate social responsibility; 26. Company contributions to economic prosperity; 27. Environmental protection; 28. The quality management; 29. Business environment risk; 30. Competition risk; 31. Industry risk; 32. Liquidity risk; 33. Business risk; 34. The accounting policies; 35. Financial indicators; 36. Balance sheet; 37. Income statement; 38. Cash flow statement; 39. Statement of changes in equity; 40. Notes in financial statements; 41. Responsibility for the financial statements; 42. Independent Auditor's Report; 43. Events after the balance sheet date; 44. Contact information.

Source: Calculated according to data of selected listed companies

Table 2: Quality of annual report for Slovenian and Croatian listed companies

DQI	**Quality of Annual Report**	**SLOVENIA**		**CROATIA**	
		No.	**Percent**	**No.**	**Percent**
0 - 20	Poor quality AR	0	0	3	10%
21 - 40	Low quality AR	0	0	8	27%
41 - 60	Average quality AR	0	0	16	53%
61 - 80	Sufficient quality AR	10	33%	3	10%
81 - 100	High quality AR	20	67%	0	0
Total		30	100%	30	100%

Source: Calculated according to data of selected listed companies

Using Mann Whitney U-test, we analyze differences between Croatian and Slovenian listed companies in accordance with DQI and selected financial indicators as follow: ROA, ROE, Debt ratio, Coefficient of own funding and Net profit margin. Table 3 shows that there is a significant difference in ranks between Slovenian and Croatian listed companies, using DQI (rank of Slovenian companies is a higher) and debt ratio (Croatian companies are more indebted). Also, there is no significant difference in ranks between Slovenian and Croatian listed companies, using other test variables.

Table 3: Results of Mann Whitney U-test for Slovenian and Croatian listed companies

Test Variable	Mean Ranks		M-W U-test p-value	Test Variable	Mean Ranks		M-W U-test p-value
	SLO	CRO			SLO	CRO	
Disclosure quality index	45,25	15,75	**,000**	**Debt ratio**	23,90	37,10	**,003**
Return on assets (ROA)	27,45	33,55	,176	Coefficient of own funding	30,07	30,93	,848
Return on equity (ROE)	32,70	28,30	,329	Net profit margin	27,25	31,60	,327

Source: Calculated according to data of selected listed companies

Binary Logistic Regressions were estimated to find dependence of DQI and companies' financial success. Parameters were evaluated by iterative maximum-likelihood estimation (MLE). There were no significant odds ratios. Only for Slovenian companies' profitability odds ratios are significant at p-value 0.10, i.e. Slovenian companies with a high profitability have a greater probability of high DQI. Spearman correlation coefficients show that there are no significance correlations between DQI and selected financial indicators for companies in Slovenia and Croatia. Further the companies are ranked according to DQI and selected financial indicators by multi-criteria PROMETHEE method [4]; [5]; [21]. Table 4 shows matrix types of preference functions and criteria's weights for multi-criteria PROMETHEE II ranking method. Companies ranking has been provided according to the degree of DQI and selected financial indicators. It is visible that up to 20th percentile overcomes Croatian companies (67%). On the other hand, companies with the lower degree business success and DQI (80th-100th percentile) in majority belong again to Croatian companies (67%).

Table 4: Types of preference functions, weights and companies' ranking according to DQI and selected financial indicators by multicriteria PROMETHEE II method

	Criterions					
CRITERIA	ROA	ROE	DR	COF	NPM	DQI
Min/Max and Type	max**	max**	min*	min*	max**	max**
Indiference Treshold	0.05	0.01	0.50	0.50	0.50	0.50
Weight	0.14	0.14	0.14	0.14	0.14	0.16
	SLOVENIA			**CROATIA**		
Percentiles 0-20:	33%			67%		
Percentiles 80-100:	33%			67%		

*U-Shape preference function; **Gaussian preference function

Source: Calculated according to data of selected listed companies

6 CONCLUSION

In this paper disclosure quality index of annual report is created and appropriate financial indicators are selected for Slovenian and Croatian listed companies. The average DQI for Slovenian listed companies is 85, i.e. Slovenian companies have high quality AR. The average DQI for Croatian listed companies is 43. That means that Croatian companies have average quality AR. Multicriteria PROMETHEE ranking show that some Croatian companies with good financial indicators have lower DQI. It can be concluded that Slovenian companies are better than Croatian, evaluating the persistence and substantiality of individual elements inside the annual report.

References

[1] Aljifri, K., 2008. Annual report disclosure in a developing country: The case of the UAE, Advances in accounting, Incorporating Advances in International Accounting, No. 24, pp. 93-100.

[2] Beattie, V., Jones, M., J., 2001. A six-country comparison of the use of graphs in annual reports, The International Journal of Accounting, No. 36, pp. 195-222.

[3] Beattie, V., McInnes, W., Fearnley, S., 2004. A methodology for analysing and evaluating narratives in annual reports: a comprehensive descriptive profile and metrics for disclosure quality attributes, Accounting Forum 28(3), pp. 205-236.

[4] Brans, J.P. and Vincke P., 1985. A Preference Ranking Organisation Method for MCDM. Management Science, 31(6), pp. 647-656.

[5] Brans, J.P. and Mareschal B., 1989. The PROMETHEE Methods for MCDM, the PROMCALC, GAIA and Bankadviser Software, Working Paper STOO/224, Vrije Universiteit Brussel.

[6] Breton, G., 2009. Semiotics analysis of storytelling in the annual report, University du Quebec at Montreal.

[7] Chatterjee, B., Tooley, S., Fatseas, V., Brown, A., 2011. An Analysis of the Qualitative Characteristics of Management Commentary Reporting by New Zealand Companies, Australasian Accounting Business and Finance Journal, 5(4), pp. 43-64.

[8] Claessens, S., Djankov, S., Klingebiel, D., 2000. Stock markets in transitional economies, Available from: [http://www.ssrn.com/abstract=240703].

[9] Cohen, D., 2002. Quality of Financial Reporting Choice: Determinants and Economic Consequences, University of Texas at Dallas-School of management.

[10] Coy, D., Dixon, K. 2004. The public accountability index: crafting a parametric disclosure index for annual reports, The British Accounting Review, 36(1), pp. 79-106.

[11] Day, R., Woodward T., 2004. Disclosure of information about employees in the Directors' report of UK published financial statements: substantive or symbolic?, Accounting Forum No. 28, pp. 43-59.

[12] Garrod, N. and Turk, I., 1995. The development of accounting regulation in Slovenia, The European Accounting Review, No. 4, pp. 749-764.

[13] Healy, P., M. and Palepu, K., 2000. Information Asymmetry, Corporate Disclosure and the Capital Markets, A Review of the Empirical Disclosure Literature; Available from: [http://ssrn.com/abstract=258514 or http://dx.doi.org/10.2139/ssrn.258514].

[14] Kothari, S., P., Li, X., Short, J., E., 2009. The effect of disclosures by management, analysts, and financial press on the equity cost of capital: a study using content analysis, The Accounting Review 84, pp. 1639–1670.

[15] Li, F., 2008. Annual Report Readability, Current Earnings and Earnings Persistence, Journal of Accounting and Economies, No. 45, pp. 221.

[16] Linsley, P., M., Shrives, P., J., 2006. Risk reporting: A study of risk disclosures in the annual reports of UK companies, The British Accounting Review No. 38, pp. 387-404.

[17] Mušura, A., 2006. Izvještavanje o korporacijskoj društvenoj odgovornosti vodećih javnih hrvatskih tvrtki, Zagrebačka škola ekonomije i menadžmenta, Zagreb.

[18] Pervan, I., 2006. Dobrovoljno financijsko izvještavanje na Internetu: analiza prakse hrvatskih i slovenskih dioničkih društava koja kotiraju na burzama, Financijska teorija i praksa 30, pp. 1-27.

[19] Pivac, S., Čular, M., 2012. Indeks kvalitete godišnjih izvješća-studija slučaja odabranih listanih poduzeća na zagrebačkoj burzi, Matematički modeli u analizi razvoja hrvatskog financijskog tržišta, Ekonomski fakultet u Splitu, pp. 97-126.

[20] Santema, S., van de Rijt, J., 2001. Strategy Disclosure in Dutch Annual Reports, European Management Journal, No. 1, pp. 101-108.

[21] Tomić-Plazibat N., Aljinović Z. and Pivac S., 2010. Risk Assessment of Transition Economies by Multivariate and Multicriteria Approaches. Panoeconomicus, vol. 57, No. 3, pp. 283-302.

[22] http://www.zse.hr; [Zagreb Stock Exchange] (accessed May 2013).

[23] http://www.ljse.si; [Ljubljana Stock Exchange] (accessed May 2013).

[24] http://data.worldbank.org/ (accessed May 2013).

ON ILLIQUIDITY MEASURES ON EUROPEAN EMERGING STOCK MARKETS

Jelena Vidović
University of Split, The University Department of Professional Studies
Kopilica 5, 21000 Split, Croatia
jvidovic@oss.unist.hr

Tea Poklepović and **Zdravka Aljinović**
University of Split, Faculty of Economics
Cvite Fiskovića 5, 21000 Split, Croatia
{tea.poklepovic,zdravka.aljinovic}@efst.hr

Abstract: In the paper the problem of applicability and validity of two well known illiquidity measures, *ILLIQ* and *TURN*, on European emerging markets is observed. It is shown that these two measures are not appropriate for seven observed markets. The measures do not follow obligatory request that returns increase in illiquidity. Therefore, new illiquidity measure, named Relative Change in Volume (*RCV*) is proposed. All measures are tested and proposed using single stock approach.

Keywords: illiquidity measures, emerging markets, relative change in volume-*RCV*.

1 INTRODUCTION

Liquidity is in practice of portfolio investment an important attribute of stocks. Investor should be able to sell stock to meet his liquidity objectives without major trading costs. But despite its evident importance in practice the role of liquidity in capital markets is hardly reflected in academic research [3]. Especially, there is a lack of researches dealing with (il)liquidity on emerging markets. In this paper we investigate problem of illiquidity measures' validity observing stock returns and related traded volumes on selected Central and South-East European emerging markets. Our approach is based on observation of single stock liquidity while we have reason to believe that changes in traded volume can result in increase of stock return or decrease of stock return as suggested in Dey [6]. Emerging markets are thin what can be concluded from observing market capitalization and number of listed companies [8]. Common situation on these markets is absence of quality stocks to be traded with what makes a big pressure on the demand for stocks of good companies. According to Bekaert et al. [4], another problem is long non-trading periods associated with greater illiquidity effects. The majority of trading during the longer periods is reserved for few most interesting stocks.

Previous literature generally consists of two large groups of liquidity measures; those are trade based and order based measures. Trade based measures include trading value, trading volume, the number of trades (frequency) and the turnover ratio. These measures are attractive, as they can be easily calculated using available data on stock prices and traded volumes. According to Aitken et al. [1] these measures have wide acceptance particularly among market professionals. Order based measures are based on more detailed trading data like data from order book. Many authors have concluded that liquidity is easy to define but has proved to be difficult to measure. In general, empirical findings support assumption that expected returns are increasing in illiquidity. Fulfilling this assumption an illiquidity measure can be considered as valid measure. The question is whether these measures are valid on emerging markets since these markets are characterized by great illiquidity and by problem of illiquidity measurement.

Today on world stock markets two measures are the most popular and used: *ILLIQ* [2] and *TURN* [5], both from the group of trade based measures. Datar et al. [5] examined asset returns and liquidity by using a turnover ratio (*TURN*), defined as the number of shares traded divided by number of shares outstanding, as a proxy for liquidity. Authors founded that stock returns are strongly negatively related to their turnover rates confirming the notion that illiquid stocks provide higher average returns for non-financial firms from the NYSE. Amihud [2] examines the average ratio of the daily absolute return to the dollar trading volume on that day for the U.S. market. It can be interpreted as the daily price response associated with one dollar trading volume thus serving as a rough measure of price impact. Author found that stock returns are negatively related over time to contemporaneous unexpected illiquidity, suggesting that illiquidity affects more strongly firms with smaller market capitalization.

Through the literature inspection it can be seen that authors define liquidity in various ways and measure liquidity using different approaches. There is no consensus about the most appropriate measure.

The paper is organized as follows: after this introductory section the data and two selected illiquidity measures are defined. In the third part these two illiquidity measures are tested. Since these measures do not confirm the main validity assumption on observed emerging markets, in the next part of the paper the new measure – Relative Change in Volume - is proposed. At the end of the paper we bring the most important conclusions.

2 DATA AND ILLIQUIDITY MEASURES

Data for this study is obtained from REUTERS database and include information on stock returns and traded volumes for 12 stocks which are constituents of stock indices on seven observed markets. Selected markets are placed in Central and South-East Europe and include stock markets of Poland, Czech Republic, Hungary, Bulgaria, Romania, Croatia and Germany as a benchmark. Data consists of around 500 daily observations in period from the beginning of November 2009 to the end of October 2011. Some characteristics of observed markets are given in Table 1.

Table 1: Features of observed emerging markets and benchmarks

Exchange	Market capitalization value at the end of the month (EUR m)	N° of companies with listed shares	Turnover (EUR m)
Bucharest Stock Exchange	12.722,64	79,00	489,1
Bulgarian Stock Exchange	6.174,27	392,00	50,8
CEESEG - Budapest	16.773,56	52,00	3.427,90
CEESEG - Prague	29.927,35	27,00	3.867,80
Deutsche Börse	1.038.389,74	746,00	370.234,00
Istanbul Stock Exchange	190.880,78	265,00	91.404,80
NYSE Euronext	1.958.378,00	1.109,00	433.025,00
Warsaw Stock Exchange	122.158,45	808,00	16.123,80
Zagreb Stock Exchange	17.629,92	246,00	138,44

Source: Federation of European Stock Exchanges FESE, values on the March 31, 2012 and Zagreb Stock Exchange

In general all observed markets are thin compared to German stock market and New York Stock Exchange. Table 1 shows very clearly that emerging markets have negligible market capitalization, turnover and number of listed shares. Istanbul and Warsaw stock exchange have the best performances in the group of emerging markets, but still far behind

the benchmarks. Investor willing to invest in stocks from these markets is facing with variety of problems. The major problem is infrequent trading. The most common situation on these markets is a trade for a day or two followed by a short non trading period. This inconsistency in trading corresponds to jumps and falls in traded volumes what could make pressure on stock returns.

Daily data are employed for the calculation of daily fluctuations in stock returns and traded volumes. This gives us an opportunity to capture day by day variations in returns and traded volumes, and allows us examination of liquidity effects across a large number of stocks and countries.

In this research we use well known Amihud's proxy for illiquidity *ILLIQ* for each stock in the form as given in Ghysels and Pereira [7]:

$$ILLIQ_i = \frac{1}{I}\sum_{t=1}^{I} |R_{it}| / V_{it} P_{ti} \tag{1}$$

where R_{it} is the daily return on stock i on day t, V_{it} is the respective daily volume, P_{it} is the price of stock i on day t and I is the number of days for which data are available for stock i. In literature *ILLIQ* is often referred as measure of price impact (*PI*).

Daily return is calculated in continuous time:

$$R_{it} = \ln\left(P_{it} / P_{i,t-1}\right) \tag{2}$$

Turnover rate measure of liquidity *TURN* is employed from Datar at al. (1998):

$$TURN_i = \sum_{t=1}^{I} V_{it} \Big/ N_i \tag{3}$$

where N_i is the number of shares outstanding.

Applying these measures on observed emerging markets we found that they are not adequate, lead to inconsistent conclusions with no statistically significant relations between stock returns and illiquidity.

3 EMPIRICAL TESTING OF *ILLIQ* AND *TURN*

In this part of the analysis we test two most commonly used illiquidity measures, *ILLIQ* and *TURN*, previously defined by relations (1) and (3). We use Pearson correlation coefficient to determine the strength and direction of relation between return and two applied illiquidity measures. These measures are very easy to calculate from widely available data on stock returns, volume and the number of shares outstanding. Our findings in this analysis do not support the findings of Amihud [2] and Datar et al. [5]. When observing every stock individually we found that each stock does not react to proven illiquidity in the same direction and/or with the same strength. An illustrative example of such results is Croatian stock market. Table 2 shows calculated values of *TURN* and *ILLIQ* and their correlation with return based on series of daily data using single stock approach for 12 stocks with the highest weight in the Croatian stock index CROBEX.[1]

According to Table 2, the results of correlation analysis do not support the hypothesis that stock returns increase in illiquidity on Croatian Stock Market. Return and illiquidity correlation in case of *ILLIQ* is statistically insignificant and has not positive sign in all cases as expected by Amihud [2]. *TURN* gives better results indicating in some cases significant but week relation to stock returns. However, the direction of that relation is in most cases positive, meaning that stock returns increase in liquidity, which is opposite to conclusions of Datar et al. [5]. Results for stocks from Hungarian stock market through *ILLIQ* measure

[1] Results for other countries are expelled from the paper due to lack of space.

show negative but insignificant relation between illiquidity and stock return what does not support the findings of Amihud [2]. According to *TURN* most stocks from Hungarian stock market do not show strong relation between liquidity and stock returns, only in two isolated cases this relation is significant and negative, but weak. When observing data for Czech market *ILLIQ* measure confirms negative relation between stock returns and illiquidity, but the *TURN* as proxy for liquidity does not support this hypothesis giving significant correlations between stock returns and liquidity measure with positive and negative sign. In case of Poland, liquidity measures are not consistent relating the strength and direction of the relationship between return and liquidity measures. While *ILLIQ* indicates positive but insignificant relation between stock return and illiquidity, *TURN* shows positive relation between increase in liquidity and increase in stock return, which is opposite to conclusions of Datar et al. [5]. All stocks from Bulgarian stock market confirm Amihuds findings and show positive return illiquidity relationship between illiquidity (*ILLIQ*) and stock returns while stocks on Romania stock market do not show consistent pattern. According to *TURN* in some cases stocks show strong positive relationship between stock returns and *TURN* suggesting that increase in traded volumes should result in increase of stock returns.

Table 2: TURN, ILLIQ and correlations with return for Zagreb Stock Exchange

Croatia	*TURN*	Correlation between return and *TURN*	*ILLIQ*	Correlation between return and *ILLIQ*
HT	0,0003	-0,2145**	1,029E-09	-0,0682
ADGR	0,0002	0,0816	4,797E-08	-0,0072
PODR	0,0003	0,0713	2,229E-07	0,0104
ERNT	0,0004	0,1761**	2,977E-08	-0,0072
ZBB	0,0000	0,2351**	2,942E-07	-0,0173
KRAS	0,0002	0,2018**	2,642E-07	-0,0091
ATPL	0,0007	0,0289	2,424E-08	0,0272
KONCAR	0,0003	0,1012	1,511E-07	-0,0925
ATGR	0,0002	-0,1106*	6,530E-08	-0,0852
PTKM	0,0007	0,0962	6,298E-07	0,0889
ADPL	0,0007	0,0635	2,843E-07	0,0121
KNZM	0,0000	0,0515	1,893E-06	0,0820

**Correlation is significant at the 0.01 level; *Correlation is significant at the 0.05 level

For the greatest European market – German stock market, results are contrary. The smallest values of *ILLIQ* measure and the highest values of *TURN* measure among all observed markets, indicate liquid market. The same conclusion can be derived from Table 1 according to market capitalization data.

In general it can be concluded that this two widely accepted liquidity measures do not drive to equal and/or valid conclusions regarding stock illiquidity performances on observed emerging markets.

4 A PROPOSAL OF NEW ILLIQUIDITY MEASURE

This paper attempts to shed light on the relation between liquidity and asset returns using a proxy for liquidity that is different from the order based measures relying on bid-ask spread and is somewhat similar to the trade based measures like Amihud's *ILLIQ* or Datar's *TURN*. The new proposed measure is very easy to calculate from the data on traded volume and stock returns in observed period. Our measure of illiquidity attempts to take into account the pressure of big differences in volume on return. Stocks that do not trade continuously have a potential price pressure of any trade following a non trading interval [4].

We measure the relative change in volume in the following way. In the first step we calculate average trading volume *AVV* for each stock in the observed period:

$$AVV_i = \sum_{t=1}^{I} V_{it} \Big/ I \tag{4}$$

In the second step we calculate relative daily change in volume *RDCV* as the absolute difference between traded volume on day *t* and *t*-1 over average volume for each stock in observed period:

$$RDCV_{it} = \left| V_{it} - V_{i,t-1} \right| \Big/ AVV_i \,. \tag{5}$$

This ratio defines daily change of traded volume in respect to average traded volume of that stock for day *t*. *RDCV* measures daily illiquidity, and when it is calculated for the whole period it represents illiquidity measure of single stock – Relative Change in Volume *(RCV)*:

$$RCV_i = \sum_{t=1}^{I} RDCV_{it} \Big/ I \,. \tag{6}$$

Proposed illiquidity measure gives information about the stocks liquidity status. For example stocks that have compact trading volumes, i.e. which have small differences between *t* and *t*-1 volume in comparison to average volume in that period have illiquidity measure under 1. Stocks whose differences in daily traded volumes approach to the average traded volume in that period have illiquidity ratio up to 1. Last category consists of illiquid stocks with *RCV* above 1. These stocks may have price pressure related to huge differences in traded volumes which exceed the average daily volume in observed period. This illiquidity measure is appropriate for emerging markets while it captures the main problems on these markets such as infrequent trading and small number of good stocks to be traded with.

Table 3: RCV on Croatian stock market

Stock	Number of trading days	Expected return	Standard deviation	Relative change in volume (*RCV*)
HT	502	-0,00018	0,01029	0,54144
ADGR	502	-0,00046	0,01147	0,97502
PODR	502	-0,00018	0,01574	1,06118
ERNT	502	-0,00048	0,01508	0,79083
ZBB	488	-0,00035	0,01992	1,20395
KRAS	502	0,00116	0,01494	1,03168
ATPL	502	-0,00177	0,01522	0,55896
KONCAR	496	0,00021	0,01378	1,02341
ATGR	501	-0,00057	0,00966	1,14765
PTKM	495	0,00047	0,02231	0,93829
ADPL	497	0,00059	0,02070	0,73430
KNZM	456	0,00028	0,02208	1,14015

To show possible good properties of Relative Change in Volume (*RCV*) we employ *RCV* on Croatian stock market. From Table 3, the value of *RCV* suggests that the most liquid stock on Croatian stock market in observed period is HT, as can be arguably confirmed from practice and values of all other stock market indicators. Among all the others, it is also contributed by the largest number of trading days, negative daily return and small risk, measured by standard deviation. KNZM is illiquid stock. It has the *RCV* value of 1.14765, which is above 1. That is supported by the lowest number of trading days, high risk and positive daily expected return. Here it has to be emphasized that in cases of illiquid stocks we can see either small number of trading days or illiquidity caused by small daily volumes.

It can be seen that most of observed stocks follow this pattern. However, in some cases the results are inconsistent. Clearly, more serious econometric analysis has to be done to prove the validity of proposed illiquidity measure, primarily in sense of proving impact of illiquidity on stock returns.

5 CONCLUSION

In this paper the problem of illiquidity on emerging markets, using single stock approach, is addressed. Since empirical findings support the assumption that expected returns increase in illiquidity, fulfilling this assumption an illiquidity measure can be considered as valid. Therefore, two most commonly used illiquidity measures, *ILLIQ* and *TURN*, have been discussed, calculated and tested on the sample of seven stock markets. It is shown that this two widely accepted liquidity measures do not drive to equal and/or valid conclusions regarding stock illiquidity performances. Therefore a new illiquidity measure, Relative Change in Volume (*RCV*) is proposed. It has the ability to take into account the pressure of big differences in volume on return. Although it gives proper information about the stocks' (il)liquidity for most of observed stocks, in some cases the results are inconsistent. Hence, future research should be conducted to prove the validity of proposed illiquidity measure using more serious econometric analysis.

References

[1] Aitken, M., Comerton-Forde, C., 2003. How should liquidity be measured?, Pacific-Basin Finance Journal, Volume 11, Issue 1, pp. 45–59

[2] Amihud, Y., 2002. Illiquidity and stock returns: Cross section and time series effects, Journal of Financial Markets, Volume 5, Issue 1, pp. 31–56

[3] Amihud, Y., Mendelson, H., 1986. Asset pricing and the bid-ask spread, Journal of Financial Economics, Volume 17, Issue 2, pp. 223–249

[4] Bekaert, G., Harvey, C.R., Lundblad, C., 2007. Liquidity and Expected Returns: Lessons from Emerging Markets, Review of Financial Studies, Volume 20, Issue 6, pp. 1783-1831

[5] Datar, V.T., Naik, N.Y., Radcliffe, R., 1998. Liquidity and stock returns: An alternative test, Journal of Financial Markets, Volume 1, Issue 2, pp. 203–219

[6] Dey, M.K., 2005. Turnover and return in global stock markets, Emerging Markets Review, Volume 6, Issue 1, pp. 45–67

[7] Ghysels, E., Pereira, J.P., 2008. Liquidity and conditional portfolio choice: A nonparametric investigation, Journal of Empirical Finance, Volume 15, Issue 4, pp. 679–699

[8] Pagano, M., 1989. Trading Volume and Asset Liquidity, The Quarterly Journal of Economics, MIT Press, vol. 104(2), pp. 255-27

The 12th International Symposium on
Operational Research in Slovenia

SOR '13

Dolenjske Toplice, SLOVENIA
September 25 - 27, 2013

Section VII:

Location and Transport

EXTRACTING A TRANSIT GEOPOINT SET FROM ROUTING API

Karlo Bala
University of Novi Sad, Faculty of Philosophy
Dr Zorana Đinđića 2, 21000 Novi Sad, Serbia
kbalayu@gmail.com

Nebojša Gvozdenović[1] and **Nenad Mirkov**
University of Novi Sad, Faculty of Economics Subotica
Segedinski put 9-11,24000 Subotica, Serbia
{nebojsa.gvozdenovic, nenad.mirkov}@gmail.com

Abstract: In the paper we deal with the data describing road routes between many to many geographical points (geopoints) provided by a black box for routing. The black box takes two geographical points as an input, and outputs the travel time/distance of a route and the array of manoeuvre points. A single call of the black box incurs unit cost. Linear growth in the number of geographical points thus leads to the quadratic growth of costs. We propose a method for extracting a transit point set. The set is then used for determining suboptimal routes while generating nearly linear costs. We show experimentally that suboptimal routes do not deviate significantly from originally generated routes.

Keywords: transit-node, road routing, networks.

1 INTRODUCTION

Computing efficient route between two geopoints A and B on a road network is one of the most used algorithmic applications nowadays. Road networks are modelled as graphs, and computing a route is done via adding nodes A, B and edges representing their connections to nearest road junctions, and finally by determining a shortest path between A and B. The classic shortest path algorithms from graph theory include Dijkstra's algorithm for one to all nodes, or Floyd-Warshall for all to all nodes. However, for large road networks classical algorithms are too slow. Recently, much effort has been put in developing speedup techniques. These techniques are often based on a preprocessing that generates and stores information about potential subroutes. Recent work of Schultes [3] contains a detailed overview of both classic and contemporary shortest path algorithms used for generating routes in large road networks.

Finding efficient direction between two geographical points A and B in practice is often done by calling an application programming interface (API) that receives the coordinates of the origin and the destination, and outputs an information about driving (walking) time and distance accompanied with some description, e.g. with manoeuvre points on the route. Such APIs are based on efficient implementations of some shortest path routing algorithm that rely on up to date road data.

A practical problem of interest is to find all routes between given geopoints using APIs. A corresponding problem in graph networks is "many-to-many" shortest path problem considered e.g. in [2] and [3]. Using routing API as a black box that incurs unit costs, for a pair of points, leads to costs $n(n\text{-}1)$ for given n geopoints. In this paper we show how to use the manoeuvre points set of a good sample for a certain region to create a set of transit geopoints. The core idea is to use transit geopoints for routing of the satisfactory quality by generating only linear costs. However, the generation of the transit geopoints incurs initial costs which are negligible in most practical applications. Namely, the generation of routes between many to many geopoints is done frequently, even on daily bases. A typical example

[1] Supported by the Serbian Ministry of Education and Science (projects III 44006 and OI 174018)

is the generation of the time and distance matrices used for daily vehicle routing occurring in the planning of currier routes or home delivery routes for online shopping. In such a planning, the time and distance matrix, containing the routes` travel times, for a given set of geolocations need to be generated to be used as an input for a vehicle routing algorithm.

We first recall, in Section 2, the ideas of transit-node routing that were first presented in [2]. In Section 3 we describe our method inspired by the transit-node approach that we apply to routing APIs. We present results obtained via Bing maps APIs (see [4].) in Section 4 and derive conclusions in Section 5.

2 TRANSIT-NODE ROUTING IN LARGE NETWORKS

If two geopoints are close enough, there are several approaches for an efficient calculation of the shortest route between them. On the other hand, if two geopoints are far enough, a route between them would most probably pass through several `important` traffic junctions. These traffic junctions are recognized as *access* or *transit* nodes in road network graphs. The set of transit nodes is usually small enough compared to the number of nodes in a road network, and one can store the distances to/from all transit nodes into distance matrix.

If u and v are the first and the last transit node on a route between a given origin s and a destination t, the route travel distance is $d_{u,v}(s,t) = d(s,u) + d(u,v) + d(v,t)$, where $d(\cdot,\cdot)$ is the shortest path distance between two nodes. Obviously, minimizing $d_{u,v}(s,t)$ overall interesting pairs of transit nodes (u,v) can be done efficiently, if $d(s,u)$ and $d(v,t)$ are provided, by exploring the distance matrix entries of transit nodes for all interesting (u,v).

Several approaches are proposed in literature for computing the set of transit nodes, hierarchies of the transit nodes and the set of interesting transit nodes of a geopoint (see [3]).

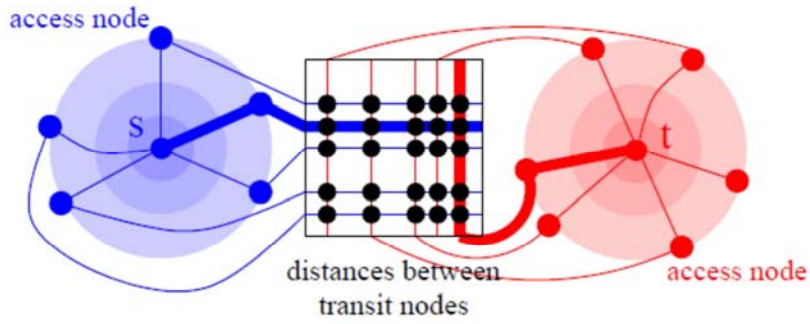


Figure 1: Schematic representation of transit-node routing from [3, p. 132].

3 TRANSIT GEOPOINTS SET

In this section, inspired by the transit-node routing idea, we describe an approach for extracting a set of transit geopoints from responses obtained via routing black box. A routing black box is a system that receives coordinates of two geopoints as an input, first called the origin and second the destination. It outputs a route, its travel time and its distance. The route is the sequence of manoeuvre geopoints, first being the origin and the last being the destination. The majority of manoeuvre points are traffic junctions. We can instruct black box either to return route with minimized travel time, or with minimized travel distance.

The steps in our procedure for extracting relevant transit geopoints and the corresponding time/distance matrix are:

- Create a sample of geopoints;

- Find route for each pair of the sample points and store all manoeuvre points;
- Count the number of appearances for each manoeuvre point;
- Put all manoeuvre points that appear more than k times (k can depend on region/sample) into the transit geopoints set T;
- Find a route for each pair of the transit geopoints and store the data about travel time/distance into time/distance matrix.

For a set of geopoints A, routes can be generated via transit geopoints set T in the following 2 steps:

1) For all $\boldsymbol{a} \in \boldsymbol{A}$:
   - Calculate air distances from a to all $\boldsymbol{t} \in \boldsymbol{T}$ and set r_a to the minimum;
   - Find all points from $\boldsymbol{A} \cup \boldsymbol{T}$ that are not further than $(1+\varepsilon)r_a$ and put them into set S_a;
   - Use the black box and get the routes from a to all geopoints in S_a;
   - Use the black box and get the routes from all geopoints in $S_a \backslash A$ to a.
2) For all $\boldsymbol{a}, \boldsymbol{b} \in \boldsymbol{A}$ find a route from a to b in the following way:
   - If $\boldsymbol{b} \in \boldsymbol{S_a}$ use the route generated in step 1);
   - If $\boldsymbol{b} \notin \boldsymbol{S_a}$ a best route from a to b via points from T.

4 COMPUTATIONAL RESULTS

As a test region we considered Vojvodina in Serbia. To create a sample, presented in Figure 1, we took 49 points that represent geolocated centres of the major municipalities in Vojvodina and some region border crossings. For the towns with more than 50,000 inhabitants we added several additional points on the borders of the towns.

Routing procedure for sample points resulted in 18789 manoeuvre points (multiplicities counted), and 711 after removing copies. We kept all 711 points in the transit geopoints set.

Figure 2: Sample points in Vojvodina

We applied procedures described in the previous section and obtained results for five sets of geopoints sets with 100, 150, 200, 250, and 300 geopoints. The results are presented in Table 1. The column named `Full’ contain some information about generating full set of routes by using black box. It is assumed that routes obtained in that way are optimal with respect to travel time. `Indicators’ serve to compare routes obtained via the proposed transit geopoints set approach with the optimal routes. The results for the transit geopoints set approach are given in the last 6 columns. Note that tests were performed for ε taking values from 0 to 1.25.

Several observations can be made from the results:

- Increase in ε leads to checking more routes through transit geopoints and thus to better results;
- Even a small sample gives up to 3.7% increase on average route times compared to optimal routes;
- Huge reductions in the number of black box calls while keeping the travel times and distances of the derived routes within satisfactory intervals.

5 CONCLUSIONS AND FUTURE WORK

We have presented an approach for generating time and distance routing matrices via black box (map API) for routing. It is inspired by recent transit-node set approach for road networks. We have demonstrated experimentally that even a rather small set of sample points gives up to 3.7% increase on average route times compared to optimal routes. It would be interesting to explore the idea further and see results on larger sample geopoint sets.
Another interesting direction of research is to explore additional possibilities of map APIs, e.g. via building a sparse road network from extracted manoeuvre points.

Table 1

|A|	Indicators	Full	R = 1+ε					
			1	1.25	1.5	1.75	2	2.25
100	Nr of black box calls full	9900	154	220	268	343	389	478
	Nr of black box calls/full	1	0.016	0.022	0.027	0.035	0.039	0.048
	total distance (km)	652371	675300	675197	675113	674992	674844	674736
	total time (min)	669812	692942	692727	692601	692453	692351	692029
	tot. dist./full	1	1.035	1.035	1.035	1.035	1.034	1.034
	tot. time/full	1	1.035	1.034	1.034	1.034	1.034	1.033
150	Nr of black box calls full	22350	329	455	564	697	789	980
	Nr of black box calls/full	1	0.015	0.020	0.025	0.031	0.035	0.044
	total distance (km)	1512824	1551378	1551152	1550887	1550675	1550324	1549869
	total time (min)	1546549	1584930	1584627	1584284	1583954	1583652	1583178
	tot. dist./full	1	1.025	1.025	1.025	1.025	1.025	1.024
	tot. time/full	1	1.025	1.025	1.024	1.024	1.024	1.024
200	Nr of black box calls full	39800	498	753	943	1194	1366	1705
	Nr of black box calls/full	1	0.013	0.019	0.024	0.030	0.034	0.043
	total distance (km)	2580208	2674717	2674012	2673559	2672967	2672387	2671699
	total time (min)	2582853	2678033	2677091	2676500	2675768	2675333	2674520
	tot. dist./full	1	1.037	1.036	1.036	1.036	1.036	1.035
	tot. time/full	1	1.037	1.036	1.036	1.036	1.036	1.035
250	Nr of black box calls full	62250	729	1142	1479	1898	2152	2748
	Nr of black box calls/full	1	0.012	0.018	0.024	0.030	0.035	0.044
	total distance (km)	4254446	4410903	4409554	4408219	4407226	4406332	4405162
	total time (min)	4220839	4368937	4367322	4366208	4365041	4364322	4362649
	tot. dist./full	1	1.037	1.036	1.036	1.036	1.036	1.035
	tot. time/full	1	1.035	1.035	1.034	1.034	1.034	1.034
300	Nr of black box calls full	89700	1015	1687	2200	2899	3318	4231
	Nr of black box calls/full	1	0.011	0.019	0.025	0.032	0.037	0.047
	total distance (km)	5947843	6162487	6160758	6159220	6157760	6155080	6152408
	total time (min)	5967372	6187844	6185276	6182719	6180209	6178287	6175286
	tot. dist./full	1	1.036	1.036	1.036	1.035	1.035	1.034
	tot. time/full	1	1.037	1.037	1.036	1.036	1.035	1.035

References

[1] Bast, H., Funke, S., Sanders, P., and Schultes D., 2007. Fast routing in road networks with transit nodes. *Science*, 316(5824):566.

[2] H. Bast, S. Funke, and D. Matijevic. TRANSIT—ultrafast shortestpath queries with linear-time preprocessing. In *9th DIMACS Implementation Challenge [1]*, 2006.

[3] Schultes D., 2008. Route Planning in Road Networks. PhD Thesis, Universität Fridericiana zu Karlsruhe

[4] http://www.microsoft.com/maps/developers/web.aspx

IMPACT OF POPULATION AGING ON MIGRATION TO REGIONAL CENTRES OF SLOVENIA

Samo Drobne* and Marija Bogataj**
* University of Ljubljana, Faculty of Civil and Geodetic Engineering, Ljubljana, Slovenia,
e-mail: samo.drobne@fgg.uni-lj.si
** MEDIFAS, Šempeter pri Gorici, Slovenia, e-mail: marija.bogataj@guest.arnes.si

Abstract: In this paper, the stickiness and attractiveness of Slovenian regional centres are analysed, particularly regarding the aging index as an indicator of the age structure of the municipalities. Migration flows between Slovene municipalities are studied in correlation with the aging index. A special attention has been given to the differences in the intensity of the flows just before the recession, in 2007, and four years later, in dependence of the aging index. It is obvious that a higher aging index in an origin and/or in a destination induce higher intensity of flows. This induction is stronger in regional centres and, furthermore, it is stronger after the crisis than it was before.

Keywords: population aging, aging index, migration, regional centres, recession, attractiveness, stickiness, SIM, Slovenia.

1 INTRODUCTION

Aging is one of the most serious challenges that Europe, particularly Southern Europe, Central and Eastern Europe, is facing in the 21st century. According to Kovács [9], by 2050 the number of older persons will be more than twice the number of children in most European countries. According to the UN Population Aging and Development [13], Europe will have more than 241 million people aged 60 and above by 2050. This will present 34% of European population (in 2012, it was 22%)! The most problematic area is, and will be, Southern Europe (especially the Mediterranean region), where this percentage will exceed 38% [13]. In Slovenia, by 2050 there will be 37% of those aged 60 and above [13]. According to Bogataj et al. [1], nearly one third of the housing stock needs to be transformed to homes and service facilities for the elderly. This could be achieved in two ways: (a1) segregation of seniors in senior cities and (a2) universality of cities including adaptability of central places and suburbs. In the near future, more than one in ten inhabitants will need long-term care which will bring together a range of medical and social services, which also means job opportunities for young persons. A careful introduction of universality and adaptability in the towns will allow for a greater mobility of the elderly that will enable them to stay in their homes longer and postpone reallocation to long-term care facilities [1].

In the European Union, Member States are responsible for the planning, funding and administration of health care and social protection systems. Local authorities and state governments should undertake research toward developing an appropriate array of community-based care services for the elderly. Moving toward consumer-centred services for the elderly would require appropriate infrastructure and a mix of changes in consumer and provider attitudes. New business practices and public policies are needed. New care processes and management structures will be introduced. According to Pogačnik et al. [10] and Zavodnik Lamovšek et al. [14], these activities should be organised in regional centres and the network of settlements in the region. To test this statement, we should first answer two questions: (b1) Is the community on NUTS 5 level (municipality in Slovenia) an appropriate level for such policies?, and (b2) Is NUTS 3 (regional) level or the networks of settlements with hubs in regional centres more appropriate? Achieving an appropriate array of community-based care services for the elderly requires research, time, and effort to integrate the elements of consumer-centred services of the aging population. In order to

achieve these objectives both the social system needs to be improved and the built environment needs to be adapted to the aging society.

When planning the infrastructure for the aging Slovenian population, the following questions are raised: (c1) What is the attractiveness of communities with older population and how sticky are the communities with older population in Slovenia?, and (c2) Are the attractiveness and stickiness stronger in the regional centres of Slovenia? If the answer to (c2) is positive, then regional centres would probably provide the best hubs for supply chains for the aging population.

To answer the questions posed above, we analysed the stickiness and attractiveness of regional centres and other local communities in Slovenia, particularly regarding the age structure in origins and destinations. For that purpose, we modified the general spatial interaction model to study the impact of population aging on migration flows. A special attention has been given to the differences in the intensity of flows just before the recession, in 2007, and four years later, in 2011.

2 PROBLEM

2.1 Aging index

A well-known indicator of the age structure is the aging index (also referred to as the elder–child ratio) that is defined as the number of people aged 65 and over per 100 youths under age 15. According to Gavrilov and Heuveline [7], in 2000 only a few countries (Germany, Greece, Italy, Bulgaria, and Japan) had more elderly than youth (aging index above 100). By 2030, however, the aging index is projected to exceed 100 in all developed countries, and the indices of several European countries and Japan are even expected to exceed 200.

In Slovenia, there were three municipalities with the aging index above 200 [11]: in 2007 (Kostel, Osilnica, Gornji Petrovci) and in 2011 (Osilnica, Kostel, Šalovci); however, the number of the municipalities where the aging index was higher than 100 increased from 136 in 2007 to 157 municipalities in 2011; see also Fig. 1.

2.2 Regional centres in Slovenia

The concept of Slovenia's urban system is defined in the Spatial Development Strategy of Slovenia [12]. The most important regional centres, or "urban centres of national significance", are [12]: Ljubljana, Maribor, conurbation Koper–Izola–Piran, Celje, Kranj, Novo mesto, Nova Gorica, Murska Sobota, Velenje, Postojna, Ptuj, and conurbations Slovenj Gradec–Ravne na Koroškem–Dravograd, Jesenice–Radovljica, Zagorje ob Savi–Trbovlje–Hrastnik, and Krško–Brežice–Sevnica (see Fig. 1). The concept of polycentric urban development emphasizes the improved (equal) accessibility to public services, i.e. administration, employment, services and knowledge, which are, in general, located in urban centres. The workplaces and economic activities in Slovenia are concentrated in the (wider) urban areas of Ljubljana, Maribor, Celje, coastal conurbation Koper–Izola–Piran, followed by Kranj, Novo mesto, Velenje, and Nova Gorica. According to [12], the most of workers commute to work in the eight aforementioned employment (regional) centres, followed by other "urban centres of national significance".

Considering the population aging in the regional centres of Slovenia (see Fig. 1), one cannot fail to observe that Maribor was the most critical regional centre, while Novo mesto had the most advantageous aging index of all, both before (in 2007) and in the recession (in 2011).

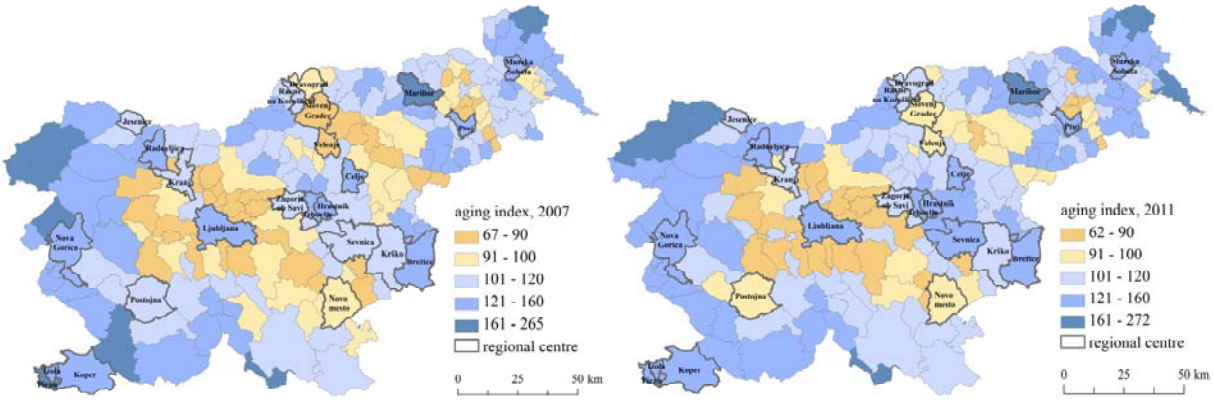


Figure 1: Aging index in the municipality in 2007 and 2011 (source: [11] and own calculation).

2.3 Spatial interaction model

Spatial interaction is a broad concept that describes movement over space. In the human sciences, the most relevant spatial interactions are defined by human migration, (daily or weekly) commuting, travelling to school, information flows, commodity flows, etc. Gravity models are the most common spatial interaction models used to analyse spatial interactions [6, 8]. Nevertheless, their application has been broadly criticised, namely, that it is not acceptable to simply replace the physical concept of "mass" with the social concept of "population". But, Cesario [2, 3] proved that "social" spatial interactions can be analysed using the general Spatial Interaction Model (SIM):

$$I_{ij} = k\, E_i\, A_j\, f(d_{ij}), \tag{1}$$

where I_{ij} is the interaction between origin i and destination j, k is the proportionality constant, E_i is emissivity in origin i, A_j is the attraction in destination j, and $f(d_{ij})$ is the function of the distance between origin i and destination j.

3 METHOD

The impact of attractiveness of regional centres on migration was studied in a modified SIM. For that purpose, we modified model (1) to

$$M_{ij} = k \cdot K(d(t))_{ij}^{\gamma} \prod_{s \in S} K(s)_i^{\alpha(s)} K(s)_j^{\beta(s)}, \tag{2}$$

where M_{ij} is the migration flow from origin municipality i to destination municipality j, k is the proportionality constant, $K(d(t))_{ij}$ is the coefficient of time-spending distance by car from the centre of municipality of origin i to the centre of municipality of destination j, and $K(s)_i$ and $K(s)_j$ are the coefficients of the analysed factor s in the municipality of origin i (of the factor of emissivity, also called the factor of *stickiness* [4]) and in the municipality of destination j (of the factor of *attractiveness*), respectively. The coefficient of the analysed factor is the proportion between the factor in the municipality and the factor at the state level. The variables analysed in model (2) are explained in Tab. 1. The impacts of stickiness of the origin, the impacts of attractiveness of the destination (i.e. regional centres or "urban centres of national importance"), and the impact of the time-spending distance between an origin and a destination on the interactions were analysed in the regression analysis using

regression coefficients γ, $\alpha(s)$, $\beta(s)$. The *stickiness* was measured by $\alpha(s)$, and the *attractiveness* was measured by $\beta(s)$.

Table 1: Variables analysed in model (2).

Sign in (2)	***Variable***
M_{ij}	migration flow from municipality of origin i to municipality of destination j (number of migrations in a year)
$K(d(t))_{ij}$	coefficient of the time-spending distance by car from the centre of the municipality of origin i to the centre of the municipality of destination j was defined as a quotient between the time-spending distance by car from the municipality of origin i to the municipality of destination j, $d(t)_{ij}$, and an average time-spending distance between all municipal centres in Slovenia, $\overline{d(t)}_{ij}$; $K(d(t))_{ij} = d(t)_{ij} / \overline{d(t)}_{ij}$; the time-spending distance between the centres of municipalities was estimated using a GIS model separately for each year, by taking into consideration the traffic situation in Slovenia
$K(P)_{\bullet}$	the coefficient of population in the municipality was defined as a quotient between the population in the municipality, $P_{\bullet}$, and the average population in the municipality in Slovenia, $\overline{P}_{SI}$; $K(P_{\bullet}) = P_{\bullet} / \overline{P}_{SI}$
$K(A)_{\bullet}$	the coefficient of aging was defined as a quotient between the aging index in the municipality, $A_{\bullet}$, and the aging index in Slovenia, A_{SI}; $K(A)_{\bullet} = A_{\bullet} / A_{SI}$; the aging index is the quotient between the population aged 65 or over and the population younger than 15

Note: $\bullet$ denotes the separate consideration of the variable in the municipality of origin i and in the municipality of destination j.

4 RESULTS

Comparing the regional centres of Slovenia, the aging index increased the most for Dravograd (+20.5%), Murska Sobota (+20%), Sevnica (+18.4%), and Ptuj (+15.4%), while it decreased the most for Koper/Capodistria (-9.1%). Tab. 2 shows the aging index in the regional centres of Slovenia in 2007 and 2011 and its relative change. Generally, the aging index in regional centres increased from 121.1 in 2007 to 125.7 in 2011. Tab. 3 shows the results of the regression analysis of migration flows in model (2) for 2007 and 2011. From the (standardized) regression coefficients it is obvious that higher aging index in an origin and higher aging index in a destination induce higher intensity of flows. This induction is stronger in regional centres and, also, it is stronger in the recession than it was before.

Table 2: Aging index in regional centres of Slovenia in 2007 and 2011 and their relative change (source: [11] and own calculation).

Regional centre	**Aging index in 2007**	**Aging index in 2011**	**Relative change of the aging index (2007–2011)**
Koper/Capodistria	139.4	126.7	-9.1%
Ljubljana	136.5	128.2	-6.1%
Postojna	104.8	99.0	-5.5%
Izola/Isola	141.7	134.8	-4.9%
Kranj	114.2	109.1	-4.5%
Nova Gorica	139.8	135.8	-2.9%
Piran/Pirano	160.4	156.8	-2.2%

Regional centre	Aging index in 2007	Aging index in 2011	Relative change of the aging index (2007–2011)
Krško	109.7	109.2	-0.5%
Brežice	137.0	138.5	1.1%
Radovljica	128.4	130.2	1.4%
Celje	131.3	133.7	1.8%
Jesenice	110.3	112.9	2.4%
Novo mesto	96.0	98.4	2.5%
Zagorje ob Savi	113.8	116.8	2.6%
Hrastnik	145.5	153.7	5.6%
Maribor	164.3	173.8	5.8%
Ravne na Koroškem	112.1	118.6	5.8%
Slovenj Gradec	89.9	98.0	9.0%
Trbovlje	145.4	158.8	9.2%
Velenje	83.5	92.9	11.3%
Ptuj	120.5	139.0	15.4%
Sevnica	106.7	126.3	18.4%
Murska Sobota	121.2	145.5	20.0%
Dravograd	99.0	119.3	20.5%

Table 3: The results of the regression analysis of migration flows (*M*) in model (2) to regional centres (urban centres of national significance) and to other municipalities in Slovenia in 2007 and 2011.

Regression Statistics for (2)				
Parameter	**Year 2007**		**Year 2011**	
	to regional centre	**to other municipalities**	**to regional centre**	**to other municipalities**
N	1,275	3,573	2,574	8,049
R^2	0.583	0.416	0.685	0.499
Adj. R^2	0.581	0.416	0.684	0.499

Regression Coefficients in (2)									
Para meter	**Symbol**	**Year 2007**				**Year 2011**			
		to regional centre		**to other municipalities**		**to regional centre**		**to other municipalities**	
		Unst. Coeff.	**St. Coeff.**	**Unst. Coeff.**	**St. Coeff.**	**Unst. Coeff.**	**St. Coeff.**	**Unst. Coeff.**	**St. Coeff.**
constant	k	0.832		1.137		1.134		1.567	
$d(t)_{ij}$	γ	-1.118	-0.674	-0.788	-0.636	-1.166	-0.526	-0.895	-0.617
$K(P)_i$	$\alpha(P)$	0.407	0.357	0.304	0.362	0.645	0.445	0.475	0.494
$K(P)_j$	$\beta(P)$	0.464	0.374	0.233	0.188	0.862	0.527	0.301	0.223
$K(A)_i$	$\alpha(A)$	0.845	0.155	0.448	0.101	1.098	0.170	0.597	0.124
$K(A)_j$	$\beta(A)$	0.456	0.068	0.241	0.053	0.706	0.084	0.468	0.099

Note: "Unst. Coeff." is the unstandardized regression coefficient; "St. Coeff." is the standardized regression coefficient; all P-values < 0.001.

5 CONCLUSIONS

It is predicted that in less then 40 years more than one third of the population in Slovenia will be older than 60 [13]. The future aging structure depends on today's aging index and migration. According to our results, the higher aging index in origin and in destination induces more intensive flows of migration. The flows to the regional centres of Slovenia are more intensive than those to local centres; therefore regional centres will probably provide the best hubs for the supply chains for the elderly. Therefore, access to properly equipped home and community-based services, including personal care for the elderly, will be needed,

not necessarily uniformly available across state, but rather available in central places of national importance with a network across the region. An access to appropriate services is essential to the quality of life for older people and should be included in spatial plans soon enough. The health and social care administration at the (yet to be shaped) regional levels, i.e. functional regions for the elderly, and the state government should undertake research toward developing an appropriate array of community-based care services for the elderly. Moving toward meaningful consumer-centred services for the elderly would require a mix of changes of public policies and supply network management structures. Achieving such changes requires research today – not only through the identification of regional centres, but also by predicting and evaluating the future functional areas for supply networks needed for older persons, which, however, is the topic of another paper [5].

Note: *The research was partly financed by the Slovenian Research Agency, research project 'The impact of recession on the interaction of regions in the global supply chain and land use', No. J5-4279-0792 2011-2014.*

References

[1] Bogataj, D., Temeljotov Salaj, A., Aver, B. (2012). Urban Growth in Ageing Societies. In: Michell, K., Bowen, P., Cattel, K. (Eds.). *Delivering Value to the Community*. Cape Town: Department of Construction Economics and Management, 437–446.

[2] Cesario, F. J. (1973). A Generalized Trip Distribution Model, Journal of Regional Science, 13: 233–247.

[3] Cesario, F. J. (1974). More on the Generalized Trip Distribution Model, Journal of Regional Science, 14: 389–397.

[4] Drobne, S., Bogataj, M. (2011). *Accessibility and Flow of Human Resources between Slovenian Regions* (MEORL, serial no. 11). Ljubljana: Faculty of Civil and Geodetic Engineering; Šempeter pri Gorici: MEDIFAS.

[5] Drobne, S., Bogataj, M. (2013). Evaluating Functional Regions for Servicing the Elderly. In Zadnik Stirn, L., Žerovnik, J., Povh, J., Drobne, S., Lisec, A. (2013). SOR '13 proceedings. Ljubljana: Slovenian Society Informatika, Section for Operational Research, here.

[6] Fotheringham, A. S., O'Kelly, M. E. (1989). Spatial Interaction Models: Formulations and Applications, Dordrecht, Kluwer Academic Publishers.

[7] Gavrilov, L. A., Heuveline, P. (2003). Aging of Population. In: Demeny, P., McNicoll, G. (Eds.). The Encyclopedia of Population. Macmillan, New York.

[8] Haynes, K. E., Fotheringham, S. (1984). Gravity and Spatial Interaction Models. SAGE publications, Inc.

[9] Kovács, Z. (2010). Challenges of Ageing in Villages and Cities: the Central European Experience. Department of Economic and Social Geography, University of Szeged, Szeged.

[10] Pogačnik, A., Zavodnik Lamovšek, A., Drobne, S. (2009). A Proposal for Dividing Slovenia into Provinces. Lex localis 7/ 4, 393–423.

[11] SORS (2013). Population - selected indicators, municipalities, Slovenia, half-yearly. Statistical Office of the Republic of Slovenia, Ljubljana. (http://pxweb.stat.si/pxweb/Dialog/varval.asp?ma=05C4008E&ti=&path=../Database/Demographics/05_population/10_Number_Population/20_05C40_Population_obcine/&lang=1, accessed: 25 March 2013).

[12] SPRS (2004). Spatial Development Strategy of Slovenia. Ministry of the Environment, Spatial Planning and Energy, Ljubljana. (http://www.mop.gov.si/fileadmin/mop.gov.si/pageuploads/publikacije/drugo/en/sprs_eng.pdf, accessed: 15 August 2012).

[13] UN (2012). Population Aging and Development 2012. United Nations. (http://www.un.org/esa/population/publications/2012PopAgeingDev_Chart/2012AgeingWallchart.html, accessed: 25 March 2013).

[14] Zavodnik Lamovšek, A., Drobne, S., Žaucer, T. (2008). Small and Medium-Size Towns as the Basis of Polycentric Urban Development, Geod. vestn. 52/2, 290–312.

EVALUATING FUNCTIONAL REGIONS FOR SERVICING THE ELDERLY

Samo Drobne* and Marija Bogataj**
* University of Ljubljana, Faculty of Civil and Geodetic Engineering, Ljubljana, Slovenia,
e-mail: samo.drobne@fgg.uni-lj.si
** MEDIFAS, Šempeter pri Gorici, Slovenia, e-mail: marija.bogataj@guest.arnes.si

Abstract: In this paper, we suggest a method to evaluate regions for servicing the elderly. In the case study, functional regions of Slovenia are evaluated by looking at the propensity to travel between regions and by attractiveness of the aging population in a municipality for commuting in functional regions. Functional regions were modelled using the Intramax method and the attractiveness of the aging population was analysed using the spatial interaction model.

Keywords: population aging, aging index, servicing the elderly, recession, functional region, Intramax, SIM, Slovenia.

1 INTRODUCTION

Aging is one of the most serious problems that most developed countries are facing in the 21st century [5,7]. According to [16], more developed regions in the world will have 32% of population aged 60 and above by 2050. In Europe, there will be 34% of population, and in Slovenia, by 2050 there will be 37% of those aged 60 and above [16]. Costs of aging (pensions, elder-care …) are mostly covered from gross earnings of labour; therefore there is a relationship between employment and aging expenses in a functional region.

In EU, Member States are responsible for the planning, funding and administration of health care and social protection systems. Local and regional authorities and state governments should undertake research toward developing an appropriate array of community-based care services for the elderly [4]. According to [12], these activities should be organised in regions, i.e. regional centres and the network of settlements in the region. The Spatial Development Strategy of Slovenia [15] defines regional centres of Slovenia; see Fig. 1. But, their gravitation areas are not clearly defined and may overlap.

Drobne and Bogataj [4] showed that the regional centres defined in [15] could provide hubs for supply chains for servicing the elderly. But, the pertinent questions emerge: What would be the most convenient regionalization of Slovenia considering the aging population? How to evaluate functional areas for supply networks needed for older persons?

2 PROBLEM

The aging index is defined as the number of people aged 65 and over per 100 youths under age 15. Officially, since July 2003, there are more elderly aged 65+ than youths 15- in Slovenia [13]. More on aging index in regional centres of Slovenia is in [4].

In Slovenia, there are 12 statistical regions of which the first version dates back to the mid-1970s. The first regionalization of statistical regions was supported by exhaustive gravity analysis of labour markets, education areas and supply markets in twelve regional centres [14]. However, labour and supply markets etc. are changing all the time, especially during crises. For that purpose, we evaluated the local labour markets in Slovenia, i.e. functional regions of Slovenia. A functional region is a region characterised by its agglomeration of activities and by its intra-regional transport infrastructure. The basic characteristic of a functional region is the integrated labour market, in which intra-regional commuting as well as intra-regional job search and search for labour demand is much more

intensive than the inter-regional counterparts [6]. Consequently, the border of a labour market region is a good approximation of the border of a functional region [11].

3 METHOD

The functional regions of Slovenia have been modelled by the Intramax method [8,9,10] using the Flowmap software [1]. The objective of the Intramax procedure is to maximise the proportion within the group interaction at each stage of the grouping process, while taking account of the variations in the row and column totals of the matrix.

Slovenia has been divided into sets of 2 to 30 functional regions for each analysed year between 2007 and 2010. The sets of regions have been evaluated following two characteristics: (C1) propensity to commute between functional regions, and (C2) concentration of the aging population in the region. (C1) was measured by the cumulative intra-regional interactions, and (C2) was estimated in the spatial interaction model by the measure of attractiveness of the aging population in urban centres of regions. ***If the urban centres of functional regions should provide the hubs for servicing the elderly in the region, then (i) cumulative intra-regional interactions should be relatively high and (ii) the high concentration of the elderly in a destination should induce high intra-regional commuting flows (that define functional regions).***

The impact of the aging population in a destination on commuting flows in functional regions was estimated in the spatial interaction model (SIM; [2,3]) that was modified to

$$C_{ij} = k \cdot K(d(t))_{ij}^{\gamma} \prod_{s \in S} K(s)_i^{\alpha(s)} K(s)_j^{\beta(s)} \quad \text{for } i \in MFR_g,\ j \in MFR_h \text{ and } MFR_g = MFR_h \tag{1}$$

where C_{ij} is the commuting flow from origin municipality i to destination municipality j, k is the proportionality constant, $K(d(t))_{ij}$ is the coefficient of the time-spending distance from the centre of municipality of origin i to the centre of municipality of destination j, $K(s)_i$ and $K(s)_j$ are the coefficients of the analysed factor s in the municipality of origin i and in the municipality of destination j, MFR_g is the set of municipalities in the functional region of origin g, and MFR_h denotes the set of municipalities in the functional region of destination h. The coefficient of the analysed factor is the proportion between the factor in the municipality and the factor at the state level. The variables analysed in model (1) are explained in Tab. 1. The impacts of *stickiness* in the origin, the impacts of *attractiveness* in the destination, and the impact of the time-spending distance between the origin and the destination on the interactions were analysed in the regression analysis using regression coefficients γ, $\alpha(s)$, $\beta(s)$. In our application, we focused most on the results regarding the stickiness of the aging population in origin i, which was measured by $\alpha(A)$, and on the attractiveness of the aging population in destination j, measured by $\beta(A)$.

Table 1: Variables analysed in model (1).

Sign in (1)	***Variable***
C_{ij}	commuting flow from municipality of origin i to municipality of destination j (number of commuters)
$K(d(t))_{ij}$	coefficient of the time-spending distance by car from the centre of the municipality of origin i to the centre of the municipality of destination j was defined as a quotient between the time-spending distance by car from the municipality of origin i to the municipality of destination j,

	$d(t)_{ij}$, and an average time-spending distance between all municipal centres in Slovenia, $\overline{d(t)}_{ij}$; $K(d(t))_{ij} = d(t)_{ij} / \overline{d(t)}_{ij}$; the time-spending distance between the centres of municipalities was estimated using a GIS model separately for each year, by taking into consideration the traffic situation in Slovenia
$K(P)_{\bullet}$	the coefficient of population in the municipality was defined as a quotient between the population in the municipality, $P_{\bullet}$, and the average population in the municipality in Slovenia, $\overline{P}_{SI}$; $K(P_{\bullet}) = P_{\bullet} / \overline{P}_{SI}$
$K(EMP)_{\bullet}$	the coefficient of employment in the municipality was defined as a quotient between the number of employed persons in the municipality, $EM_{\bullet}$, divided by the number of active population in the municipality, $AP_{\bullet}$, $EMP_{\bullet} = EM_{\bullet} / AP_{\bullet}$, and the number of employed persons in Slovenia, EM_{SI}, divided by the number of active population in Slovenia, AP_{SI}, $EMP_{SI} = EM_{SI} / AP_{SI}$; $K(EMP)_{\bullet} = EMP_{\bullet} / EMP_{SI}$
$K(GEAR)_{\bullet}$	the coefficient of gross earning per capita in the municipality was defined as a quotient between the gross earning per capita in the municipality, $GEAR_{\bullet}$, and the gross earning per capita in Slovenia, $GEAR_{SI}$; $K(A)_{\bullet} = A_{\bullet} / A_{SI}$
$K(UFSP)_{\bullet}$	the coefficient of useful floor space of dwellings per capita in the municipality was defined as a quotient between the useful floor space of dwellings per capita in the municipality, $UFSP_{\bullet}$, and the useful floor space of dwellings per capita in Slovenia, $UFSP_{SI}$; $K(UFSP)_{\bullet} = UFSP_{\bullet} / UFSP_{SI}$
$K(BUDG)_{\bullet}$	the coefficient of the budget of the municipality was defined as a quotient between the budget of the municipality, $BUDG_{\bullet}$, and the average budget of municipalities in Slovenia, $BUDG_{SI}$; $K(BUDG)_{\bullet} = BUDG_{\bullet} / BUDG_{SI}$
$K(APF)_{\bullet}$	the coefficient of the average price per m^2 of flat in the municipality was defined as a quotient between the average price per m^2 of flat in the municipality, $APF_{\bullet}$, and the average price per m^2 of flat in Slovenia, APF_{SI}; $K(APF)_{\bullet} = APF_{\bullet} / APF_{SI}$
$K(A)_{\bullet}$	the coefficient of aging was defined as a quotient between the aging index in the municipality, $A_{\bullet}$, and the aging index in Slovenia, A_{SI}; $K(A)_{\bullet} = A_{\bullet} / A_{SI}$; the aging index is the quotient between the population aged 65 or over and the population younger than 15

Note: $\bullet$ denotes the separate consideration of the variable in the municipality of origin i and in the municipality of destination j.

4 RESULTS

Fig. 1 shows the linear trend of the aging index (LTA) in Slovenian municipalities in 2000–2012 (for 18 new municipalities in 2007–2012). It is obvious that the linear trend of the aging index is positive for all regional centres. The most critical centres with a very high linear trend ($LTA > 3$) are: Murska Sobota, Maribor, Ptuj, Dravograd, Ravne na Koroškem, Slovenj Gradec, Velenje, Hrastnik, Trbovlje, Sevnica, Novo mesto, Radovljica, and Piran. Other regional centres with also a positive linear trend of the aging index ($0 < LTA \leq 3$) are: Celje, Krško, Brežice, Zagorje ob Savi, Ljubljana, Jesenice, Kranj, Postojna, Nova Gorica, Koper, and Izola. The lowest dynamics of the aging index is shown for Postojna ($LTA = 0.90$), while in Ljubljana ($LTA = 1.99$), Koper ($LTA = 1.53$) and Krško ($LTA = 1.85$) the dynamics of the aging population is somewhat higher.

Fig. 2 shows regression coefficients for the aging indexes, $\alpha(A)$ and $\beta(A)$, in (1) (left) and cumulative intra-regional commuting (right) in relation to sets of functional regions in Slovenia in 2007 and 2010. Considering the attractiveness of the aging population, the most

convenient regionalization would be in local maximums of $\beta(A)$. The results show that local maximums of $\beta(A)$ have not changed for delineation of Slovenia into 7 and 11 functional regions. Before the crisis (in 2007), the local maximum of $\beta(A)$ was also identified for 15 functional regions, but, in the crisis, the population in Postojna and in neighbouring municipalities is getting older much slower than that in the neighbouring municipalities. Considering the regionalization into a higher number of smaller regions, the regionalization into 29 and 30 regions are playing an important role before and in the crisis.

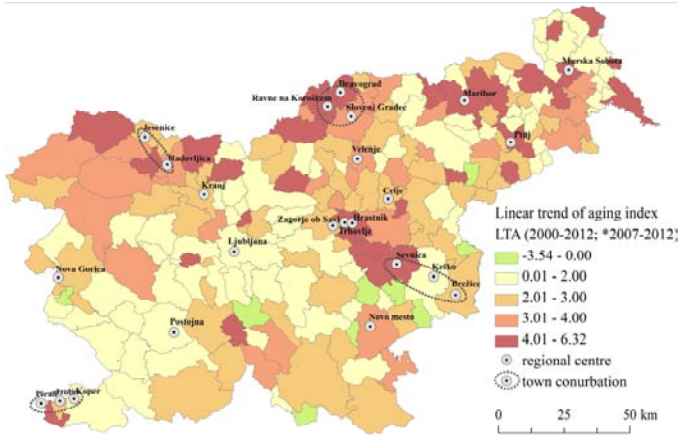


Figure 1: Linear trend of the aging index (*LTA*) in the municipality (2000–2012; * 2007–2012 for 18 new municipalities; regional centres are defined in [15]; source: [13] and own calculation).

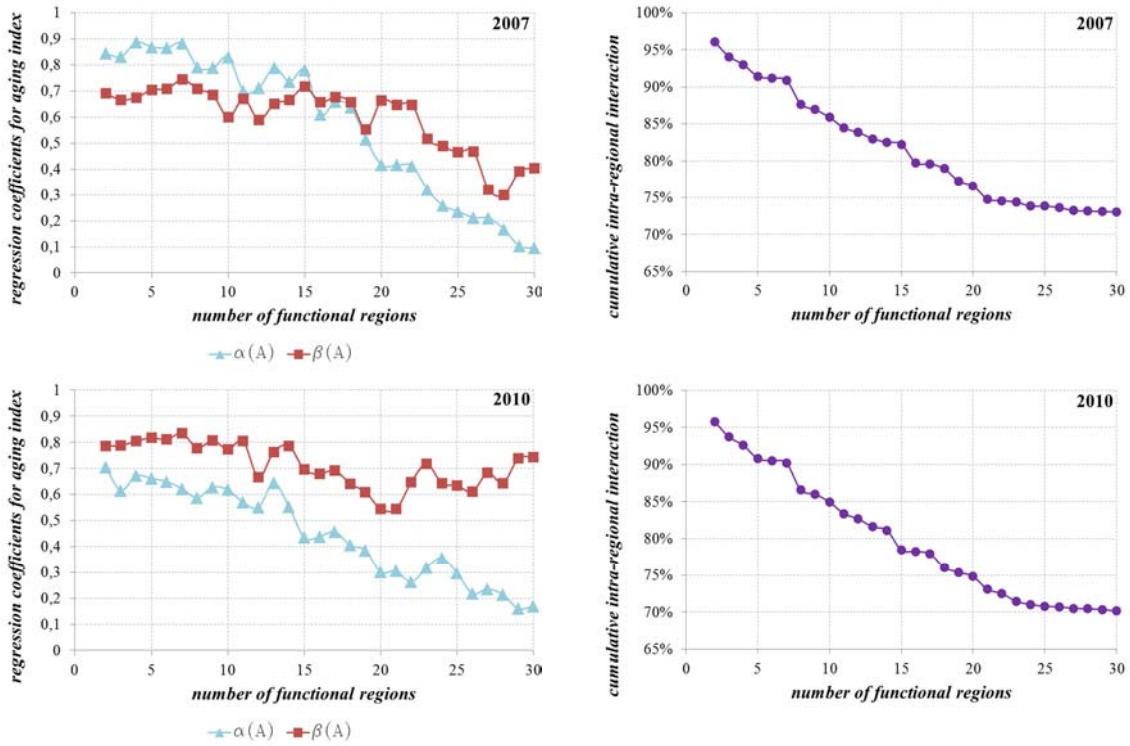


Figure 2: Regression coefficients for aging indexes in (1) (left) and cumulative intra-regional interactions (commuting) in relation to the sets of functional regions in Slovenia (right) in 2007 and 2010.

By comparing the attractiveness of the aging population in urban centres in relation to the commuting flows in the region and cumulative intra-regional commuting, we obtained the most convenient sets of functional regions for servicing the elderly, i.e. regionalization into 7

functional regions for the period of 2007–2010, and regionalization into 15 functional regions before the crisis (in 2007); however, that was changed in the crisis (2008–2010) when 14 functional regions have become more convenient for servicing the elderly. In the crisis, the flows in the functional region of Postojna became relatively less important than the flows in other functional regions. Hence, Postojna has been included in the functional region of Koper. Figs. 2 and 3 show 7 functional regions in 2007–2010 and 14 functional regions in 2008–2010.

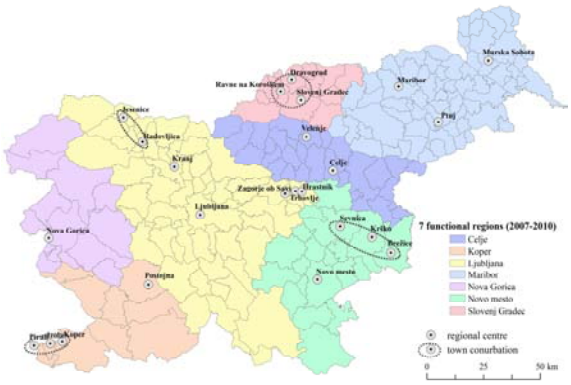


Figure 2: Seven functional regions for servicing the elderly in Slovenia in 2007–2010.

Figure 3: Fourteen functional regions for servicing the elderly in Slovenia in 2008–2010.

Tab. 2 shows the results of the regression analysis of commuting flows in model (1) for the most convenient and stable regionalization of Slovenia: delineation into 7 functional regions for servicing the elderly. It is obvious that the stickiness of the aging population continued to decrease; however, the impact of the aging index in the destination increased the most in the first year of recession and slightly decreased in 2009. It is also evident that the attractiveness of urban centres in the functional regions increased again from 2009 to 2010.

Table 2: The results of the regression analysis of commuting flows (C_{ij}) in model (1) in seven functional regions of Slovenia (see Fig. 2) in 2007–2010.

Parameter	Symbol	Year 2007	Year 2008	Year 2009	Year 2010
N		5039	5161	5171	5243
Adj. R^2		0.700	0.693	0.700	0.705
constant	k	1.458	1.392	1.431	1.681
$d(t)_{ii}$	γ	-2.099	-2.141	-2.123	-2.076
$K(P)_i$	$\alpha(P)$	0.631	0.637	0.650	0.630
$K(P)_j$	$\beta(P)$	0.774	0.741	0.757	0.755
$K(EMP)_i$	$\alpha(EMP)$	-0.357	-0.589	-0.475	-0.290
$K(EMP)_j$	$\beta(EMP)$	1.502	1.426	1.255	1.261
$K(GEAR)_i$	$\alpha(GEAR)$	-0.485	-0.592	-0.495	[-0.192]
$K(GEAR)_j$	$\beta(GEAR)$	0.601	0.623	0.708	0.914
$K(UFSP)_i$	$\alpha(UFSP)$	[0.215]	[0.016]	[-0.068]	[0.100]
$K(UFSP)_j$	$\beta(UFSP)$	0.369	[-0.075]	[-0.212]	-0.231
$K(BUDG)_i$	$\alpha(BUDG)$	0.604	0.558	0.889	0.783
$K(BUDG)_j$	$\beta(BUDG)$	0.605	0.687	0.764	0.851
$K(APF)_i$	$\alpha(APF)$	-0.472	-0.410	-0.303	-0.353
$K(APF)_j$	$\beta(APF)$	0.077	0.311	0.241	0.260
$K(A)_i$	$\alpha(A)$	0.883	0.854	0.650	0.622

$K(A)_j$	$\beta(A)$	0.744	0.899	0.819	0.834

Note: regression coefficients where P-values > 0.15 are in grey and in square bracket [ ]

5 CONCLUSIONS

It is predicted that there will be 37% of those aged 60 and above in Slovenia by 2050 [16]. The future aging structure depends on today's aging index and migration. According to [4], the flows to the regional centres of Slovenia are more intensive than those to local centres; therefore regional centres will probably provide the best hubs for the supply chains for the elderly. Our results for all sets of functional regions by year from 2007 to 2010 confirm the results published in [4] as well. In this paper we suggested a method to evaluate functional regions (and test regional centres) for supply networks needed for the elderly. The method has been tested for present-day data, but it can also be used on estimated data for the future.

Note: *The research was partly financed by the Slovenian Research Agency, research project 'The impact of recession on the interaction of regions in the global supply chain and land use', No. J5-4279-0792 2011-2014.*

References

[1] Breukelman, J., Brink, G., de Jong, T., Floor, H. (2009). Manual Flowmap 7.3. Faculty of Geographical Sciences, Utrecht University, The Netherlands. http://flowmap.geo.uu.nl, accessed 15 August 2011.

[2] Cesario, F. J. (1973). A Generalized Trip Distribution Model, Journal of Regional Science, 13: 233–247.

[3] Cesario, F. J. (1974). More on the Generalized Trip Distribution Model, Journal of Regional Science, 14: 389–397.

[4] Drobne, S., Bogataj, M. (2013). Impact of Population Aging on Migration to Regional Centres of Slovenia. In Zadnik Stirn, L., Žerovnik, J., Povh, J., Drobne, S., Lisec, A. (2013). SOR '13 proceedings. Ljubljana: Slovenian Society Informatika, Section for Operational Research, here.

[5] Gavrilov, L. A., Heuveline, P. (2003). Aging of Population. In: Demeny, P., McNicoll, G. (Eds.). The Encyclopedia of Population. Macmillan, New York.

[6] Karlsson, C., Olsson, M. (2006). The Identification of Functional Regions: Theory, Methods, and Applications, Ann Reg Sci, 40: 1–18.

[7] Kovács, Z. (2010). Challenges of Ageing in Villages and Cities: the Central European Experience. Department of Economic and Social Geography, University of Szeged, Szeged.

[8] Masser, I., Brown, P. J. B. (1975). Hierarchical aggregation procedures for interaction data. Environment and Planning A, 7(5): 509–523.

[9] Masser, I., Brown, P. J. B. (1977). Spatial representation and spatial interaction. Papers of the Regional Science Association 38, 71–92.

[10] Masser, I., Scheurwater, J. (1980). Functional regionalisation of spatial interaction data: an evaluation of some suggested strategies. Environment and Planning A, 12(12): 1357–1382.

[11] OECD (2002). Redefining Territories – The functional regions, Organisation for Economic Co-operation and Development, Paris, France.

[12] Pogačnik, A., Zavodnik Lamovšek, A., Drobne, S. (2009). A Proposal for Dividing Slovenia into Provinces. Lex localis 7/4, 393–423.

[13] SORS (2013a). Population – selected indicators, municipalities, Slovenia, half-yearly. Statistical Office of the Republic of Slovenia, Ljubljana. http://pxweb.stat.si/pxweb/Dialog/varval.asp?ma=05C4008E&ti=&path=../Database/Demographics/05_population/10_Number_Population/20_05C40_Population_obcine/&lang=1, accessed 25 March 2013.

[14] SORS (2013b). Administrative-territorial division, Statistical Office of the Republic of Slovenia, Ljubljana. http://www.stat.si/eng/tema_splosno_upravno.asp, accessed: 25 March 2013.

[15] SPRS (2004). Spatial Development Strategy of Slovenia. Ministry of the Environment, Spatial Planning and Energy, Ljubljana. http://www.mop.gov.si/fileadmin/mop.gov.si/pageuploads/publikacije/drugo/en/sprs_eng.pdf, accessed 15 August 2012.

[16] UN (2012). Population Aging and Development 2012. United Nations. http://www.un.org/esa/population/publications/2012PopAgeingDev_Chart/2012AgeingWallchart.html, accessed 25 March 2013.

ASSESSMENT METHODOLOGY OF THE RADIATION LOAD OF MULTILATERATION IN COMPARISON TO THE TRADITIONAL SECONDARY SURVEILLANCE RADAR FOR AN AREA CELL

Rainer Graf[1], **Michael Löffler**[2] **and Gerhard Navratil**[3]
[1] University of Applied Sciences Technikum Wien
Höchstädtplatz 6, 1200 Vienna, Austria
it10m027@technikum-wien.at
[2] Austro Control Corporation
michael.loeffler@austrocontrol.at
[3] Vienna University of Technology, Department for Geodesy and Geoinformation
Gusshausstr. 27-29/E120.2, 1040 Vienna, Austria
navratil@geoinfo.tuwien.ac.at

Abstract: This paper examines the radiation load of multilateration and compares it to the radiation load of a traditional secondary surveillance radar. One of the questions that need to be answered before implementing a new technical system is the influence on the environment. Since positioning systems typically use emitted signals, there is radiation load that may harm the local population around the transmitters. In order to determine the radiation load for multilateration, there scenarios were developed to determine maximum and average radiation load and compare them to the currently used systems. It is shown that multilateration has only a fraction of the radiation load of the currently used systems, which would be replaced by multilateration.

Keywords: Multilateration, Secondary Surveillance Radar, Radiation load, Air Traffic Management, Air Navigation Service Provider, Positioning.

1 INTRODUCTION

Global air traffic has been increasing continuously in the last decades. Doubling of passengers is predicted for the next 15 years [1]. Thus more aircrafts will be used and the currently busy airspace will be loaded even more. Air Navigation Service Providers (ANSPs) are responsible for the safety and efficiency of the air traffic. Conventional radar technology cannot keep up with the increasing demand. Therefore, many ANSPs search for new technologies, which can increase efficiency, minimize infrastructure costs and improve safety. A possible solution is multilateration (MLAT) [2].

One problem with multilateration is that there is no expertise about the radiation load of a countrywide multilateration system. During the technical licensing process only the radiation of a single station is determined and there are no models for the overlapping radiation load for multiple stations. The Austrian ANSP, Austro Control, is thus interested in this expertise. The main questions are:

- How could be a model for the assessment of the radiation load of multilateration for an area cell look like?
- How much lower or higher is the radiation load of multilateration in comparison to the traditional Secondary Surveillance Radar (SSR) for an area cell?

2 MULTILATERATION

2.1 Definition

Multilateration was designed in the early 1990's when the International Civil Aviation Organization (ICAO) developed the concept of Future Air Navigation System (FANS), which has been based on satellite and data link technology. This technology is a new

surveillance technology for the Air Navigation Service Providers (ANSPs) to control the increasing air traffic. The official definition is:

> *"Locating an object by computing the Time Difference of Arrival (TDOA) of a signal to three or more receivers."* (cf. [2], p. 1)

2.2 Working principle

Multilateration can use all available signals (A/C Radar, Mode S, Mode S Extended Squitter and ADS-B) to calculate a position. Multilateration uses receivers set to receive signals at a frequency of 1090 MHz. The remote units receive the signals (Pulses of A/C Code, in case of Mode S only the ID) permanently. It is necessary, that the signals can be assigned and temporally correlated between receiving stations, to calculate a position. If more information or updates are needed, the system must interrogate the missing information. If the system receives information or updates, it always knows which aircraft has sent this information.

2.3 Position calculation of Multilateration

A remote unit processes received data in the CPS (Central Processing System). Here, the important data is the ID of Mode S from an aircraft. The IDs get successively into the CPS, in which ID plus time are written into the database. Thus it is known, at which time, which signal has been received from which aircraft (due to its ID and the related time stamp) at a specific remote unit. The clocks of these remote units must be synchronized with each other (by reference pulse or time base systems). Finally, a list of remote units is built up.

Multilateration operates with the Time Difference of Arrival (TDOA) method and determines the position by intersection of spherical hyperboloids. The following steps explain the position calculation of MLAT in more detail:

1) Time of arrival at each remote unit (ID + Time)
2) Generate remote unit (RU) couples (TDOA hyperbola)
3) Positions times xTDOA

The physical basis consists of the nearly constant speed of propagation of the electromagnetic wave (300,000 km/s) in the air. By measurement of the signal propagation time between the time of request and the arrival of the response, a distance can be calculated. Formally, we can state (1):

$$s = v \cdot t \tag{1}$$

The technology uses TDOA between remote unit pairs to accurately determine a target position. The result of $T_a - T_b$ is the distance between "RU-a" and "RU-b", where the target may be located. The exposition of all points, which can provide $T_a - T_b$, is a curve in the shape of a hyperbola. All points on the hyperbola are possible positions of the target. A signal reception at two remote units is necessary to calculate a TDOA, because the TDOA is the difference in time of arrival between two remote units.

A third remote unit allows finding the 2D position of a target. This provides two additional solutions, $T_a - T_c$ and $T_b - T_c$ which both intersects the hyperbola formed by $T_a - T_b$. The point, at which all hyperbolas intersect, is the location of the target. However, the 2D positions cannot provide height information. Either a Mode C reply or an additional fourth remote unit is necessary to determine target height. The point, at which all the hyperbolas intersect, determines the position of the target. Then, the additional solutions provide the

range, direction and height of the target. Figure 1 shows the TDOA position with four remote units.

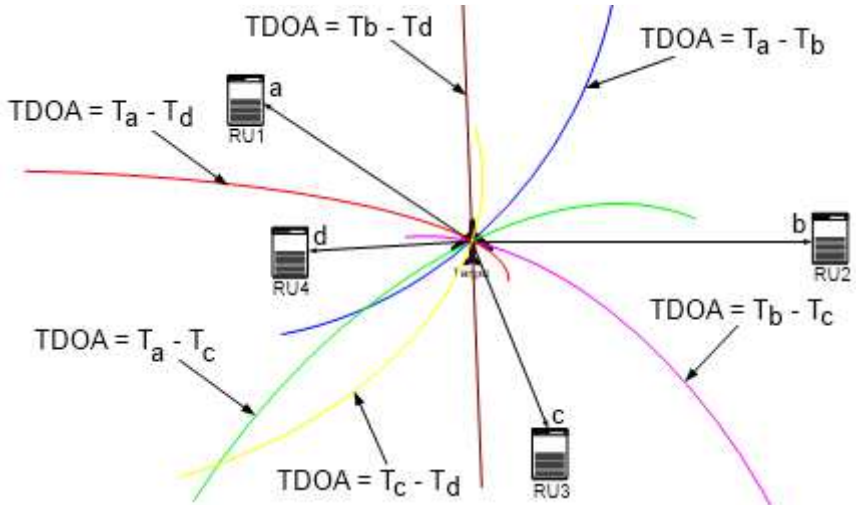


Figure 1: TDOA Position of with 4 RU's

3 RADIATION LOAD OF MLAT IN COMPARISION TO THE TRADITIONAL SSR FOR AN AREA CELL

All considerations and calculations are based on the Austrian MLAT system developed by Saab Sensis Corporation. The solution of Saab Sensis contains high-performance sensors and provides accurate and reliable WAM (Wide Area Multilateration) surveillance for Mode S and Mode A/C equipped aircrafts and ADS-B (Automatic Dependent Surveillance – Broadcast) surveillance for ADS-B equipped aircrafts. The Saab Sensis MDS equipment is used and certified to support multiple WAM and PRM (Precision Runway Monitoring) surveillance applications.

3.1 System design of the Austrian MLAT solution

The Austrian ANSP, Austro Control, divides the Austrian Flight Information Region (FIR) airspace into four WAM regions, each region covering one or more Terminal Maneuvering Areas (TMA) and Control Zones (CTR), a number of Control Areas (CTA), and a part of the FIRs that are bordering to the Austrian FIR. The four-system architecture is determined:

- Achieves optimal low-altitude coverage in all TMAs given the topography of Austria
- Ensures that the resulting WAM coverage is similar to the coverage of the existing Austrian SSRs
- Ensures that the size of each coverage volume and the number of remote units in each system is easily managed
- Regional system configurations can be changed without affecting the overall system
- System can be deployed regionally with overlapping project activities in each region
- Future expandability by the use of additional regional systems or expanding the coverage of existing systems

The four WAM regions are the Region Wien, Region Graz-Klagenfurt, Region Linz-Salzburg and Region Innsbruck. One MDS (Multistatic Dependent Surveillance) system per region provides WAM and ADS-B surveillance services for each of these four regions. Each

MDS system consists of a set of remote units and a CPS, with some sharing of remote units between systems. Figure 2 shows the regional WAM system architecture.

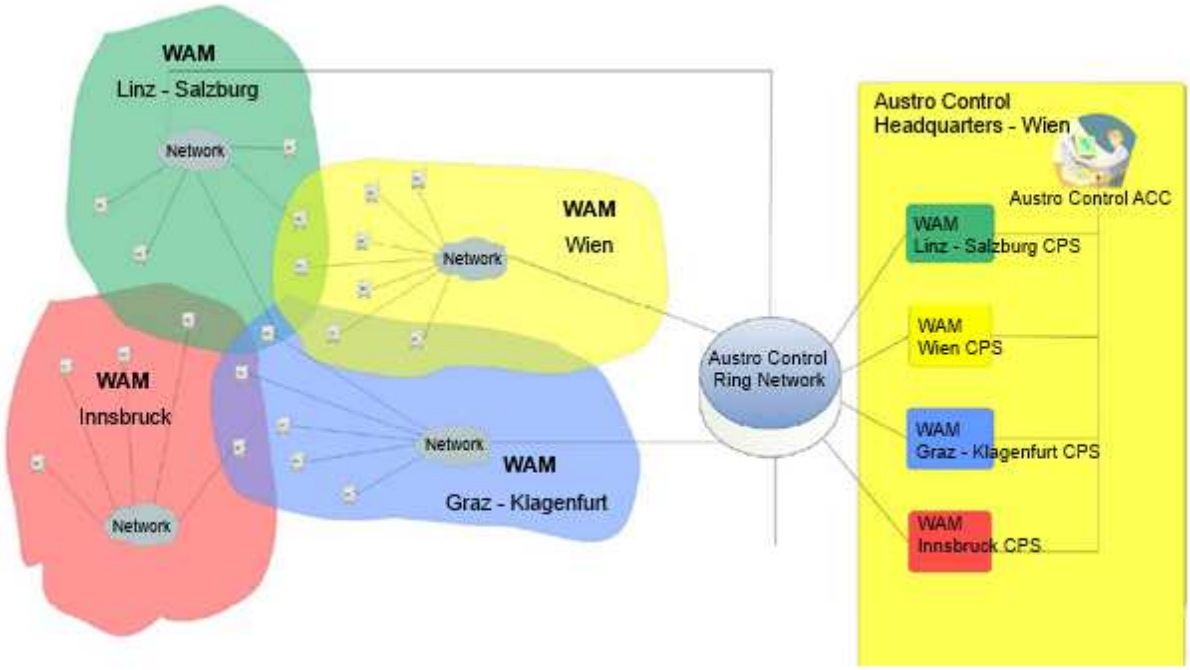


Figure 2: Regional WAM System Architecture – Operational View

3.2 Consideration for an area cell without topography data

For the assessment of the radiation load of multilateration in comparison to the traditional SSR for an area cell, five MLAT stations (Feichtberg, Meierhofberg, Oftering, Neumarkt and Sonntagberg) are defined as a whole system around Linz (City in Austria). The model should describe all the processes of the system in relation to the radiation load. Figure 3 shows a site plan in which all MLAT stations (red antennas) are illustrated.

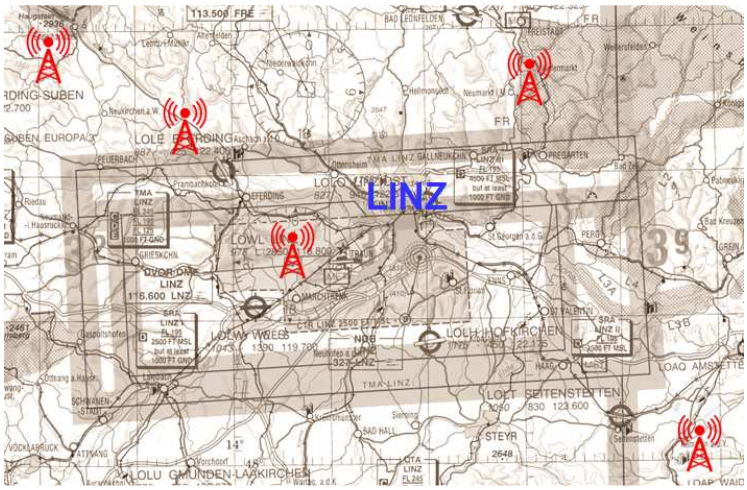


Figure 3: MLAT stations of the model

The first step was to create the basic conditions to build up a realistic model. The basic conditions are the area of the model, the place of the MLAT stations and the available data of a traditional radar as well as of the MLAT system in this area. For this reason, the city of

Linz was chosen to model the radiation load, because five active MLAT stations and one SSR are located around this city and all the required data were available.

The second step was defining the states of the MLAT system to allow the calculation of the radiation load. Therefore, three different states of the system were defined:

1) **Ideal typical case:** The ideal typical case is considered under the assumption that there is an active radar in the area of the MLAT system. The radar system is interrogating the transponders to provide A-Code and C-Code information, which can also be received and used by the MLAT system. Due to that, the Mode S transponders are also active 4.7 times per second (average) without getting interrogated by the radar. So, there are enough signals for the MLAT system to determine a high quality position within 4 seconds. Therefore, the multilateration system operates in the typical ideal case only passively and produces no radiation load. This means that in this case only the radiation load of the radar must be considered.

2) **Realistic case:** The realistic case uses the assumption that there is no active radar and that there are 4 aircrafts (respectively two in takeoff and landing) in the area of Linz as well as 8 more aircrafts (respectively two per MLAT station) in the fly-over phase. This means that the real scenario is based on 12 aircrafts (a high number of aircrafts in reality) in the entire system (around Linz) and no other Mode S radars are interrogating. Thus, in this case only the radiation load of the MLAT system must be considered. The radiation load of the MLAT system is caused by 6 interrogations per second, because at least every 4 seconds an A-Code and a C-Code update is necessary, which means that every 2 seconds one interrogation is performed. This results in 6 interrogations per second for the 12 aircrafts.

3) **Worst case:** The worst case is considered under the assumption that there is no active radar and no aircraft (flying object) in the airspace of Austria, which has to be interrogated, except for the area of Linz. This means that in this case only the radiation load by the MLAT system must be considered. The radiation load of the MLAT system is caused by the maximum possible interrogation rate per second.

The third step was to determine the basic data of the MLAT stations. The data contains the coordinates, the transmission power, the gain, the angle as well as the characteristic of each MLAT station. These collected data were used to calculate the transmission power of each single MLAT station as well as for the total radiation load of Linz. Another important data for the calculation are the distances between every MLAT station and the city center of Linz, because the radiation load decreases proportionally to the square of the radius. The next steps were the calculation of the transmission power of each single MLAT station and the calculation for the total radiation load for the city of Linz.

At the beginning of the calculation of the power flux density of each single MLAT station, the gain must be converted from the antenna gain (G) into the gain factor (g_S) with the following formula (2).

$$G = 10 \cdot \log(g_S) \rightarrow g_S = 10^{\frac{G}{10}} \quad (2)$$

For the calculation of the maximum power flux density of each station, the transmission power (P) and the antenna gain (g_S) as well as the distance from each station to the city of Linz (r^2) must be known. Consequently, the maximum power flux density of each station can be calculated with the following formula (3).

$$S = \frac{P \cdot g_S}{4 \cdot \pi \cdot r^2} \tag{3}$$

Depending on the manufacturer of the MLAT system, the maximum transmission power of a MLAT station depends on the respective type of the station and the maximum duty cycle (calculated over a Mode S interrogation). This procedure is performed individually for each single MLAT station.

The last step in this calculation is to divide the number of interrogations of the system on each MLAT station. Depending on the system state (Ideal typical-, Realistic-, Worst case), the system operates with various interrogations per second. Afterwards, these interrogations must be divided on the active MLAT stations. Finally, the resulting maximum power flux density of the city of Linz is calculated by the summation of all MLAT stations.

4 RESULTS & DISCUSSION

The global air traffic has been increasing continuously in the last few decades. Therefore, a new surveillance technology for the air navigation service providers is needed. Multilateration is a suitable approach as this technology increases the efficiency, minimizes the infrastructure costs and improves the safety of the system.

A new technology does not only lead to new challenges and experiences within the air navigation system, but also means changes for the staff. Especially, the changes in the work area of an air traffic controller can cause many problems: First of all, air traffic controllers often have an aversion against changes in the system and significant changes on the control screen are seldom accepted. The most important difference between multilateration and a traditional SSR is the higher update frequency (1 second vs. 4-12 seconds) on the display of the air traffic controller. Therefore, multilateration provides more accurate data than traditional SSRs do. The refresh time of 1 second seems to be very uncomfortable for the air traffic controller as changes on the screen (smooth moving items vs. jumping items) cannot be detected that easily. As a consequence, the refresh time is now adjusted to 2 seconds.

For the assessment of the radiation load of multilateration in comparison to the traditional SSR for an area cell, a model was designed to describe all the processes of the system. Therefore, three different states (ideal typical case, realistic case & worst case) of the system were defined and evaluated. The comparison of the radiation load of the multilateration system with the radiation load of the traditional SSR shows that the radiation load of the multilateration system is lower than the radiation load of a traditional SSR for all three states. For example, the radiation load of the multilateration system for the realistic case is 7 powers of ten lower than the radiation load of a traditional SSR. It is therefore obvious that the radiation load can be reduced dramatically with such a system. And the reduction of the radiation load can in turn reduce the impact on the population and the environment.

References

[1] Ohler Rainer (2011) Mobilitäts – und Kraftstoffstrategie der Bundesregierung. Presentation at „Auftakt zur Mobilitäts- und Kraftstoffstrategie der Bundesregierung“, Berlin, 09.06.2011.

[2] Multilateration & ADS-B Executive Reference Guide, [online].

[3] Available: http://www.multilateration.com/downloads/1/1/210.html (Access on 11.12.2011)

AN INTEGER PROGRAMMING MODEL FOR THE DYNAMIC LOCATION AND RELOCATION OF EMERGENCY VEHICLES: A CASE STUDY

Mahdi Moeini, Zied Jemai and **Evren Sahin**
Laboratoire Génie Industriel (LGI), Ecole Centrale Paris,
Grande Voie des Vignes, F-92 295, Châtenay-Malabry, France.
{mahdi.moeini, zied.jemai, evren.sahin}@ecp.fr

Abstract: In this paper, we address the dynamic Emergency Medical Service (EMS) systems. A dynamic location model is presented that tries to locate and relocate the ambulances. The proposed model controls the movements and locations of ambulances in order to provide a better coverage of the demand points under different fluctuation patterns that may happen during a given period of time. Some numerical experiments have been carried out by using some real-world data sets that have been collected through the French EMS system.

Keywords: integer programming, emergency medical service systems, location problem.

1 INTRODUCTION

Due to the crucial role of the Emergency Medical Services (EMS) in saving lives, numerous studies have been carried out in order to improve the quality of the EMS systems. In this context, different research directions have been taken into account. Some examples are adaptation of the service modes to the changes in the customer needs (such as home care services), personnel scheduling in the medical centres, location of the service centres, etc. In any case, two main objectives are saving the lives (by reducing the mortality in the emergency cases) and reducing the costs.

Among the EMS literature, the problem of locating the emergency service vehicles has attracted special attention during decades of research. The vehicle location problem in the context of EMS systems is dealing with locating the vehicles in some potential service sites in order to reduce the delay of covering the emergency service demands.
Each emergency service vehicle is completely equipped to all emergency facilities that medical team may need in their missions. Due to this fact, it is quite expensive to buy any of these emergency vehicles. Consequently, any emergency service has access to a limited number of emergency vehicles; hence, it is important to optimally locate them in order to improve the responsiveness of the system.

1.1 Literature review

The earliest EMS models have been introduced in 70s by Toregas et al. [7]. During decades the location problems of EMS vehicles became an active research area and numerous papers have been published on this topic. The published papers may be classified according to their nature: static, dynamic, or stochastic models.

Toregas et al. [7] introduced the Location Set Covering (LSC) model that minimizes the number of the necessary ambulances for covering all demand points. The LSC model can penalize the users of the model by its expensive solutions; because it may provide a necessary number of ambulances that is too larger than it would be. Furthermore, the LSC model is so rather basic and it does not permit location of more than one ambulance in a service centre.

Due to the limits of the LSC model, the Maximal Covering Location problem (MCLP) has been presented by Church et al. [3]. The MCLP model tries to maximize the covered population by taking into account a predefined number of ambulances.

The LSC and MCLP models are static models, in the sense that they do not take into account the possible fluctuations in the EMS system. In fact, when a call arrives to the call centre of the EMS service centre, it may need affectation of an ambulance. If such need is confirmed, an ambulance will be affected to cover the demand point. At this stage, the corresponding ambulance will be no longer available. Consequently, the static models must be solved from scratch for a smaller number of ambulances. This procedure is computationally expensive. At this point, one may use the dynamic models.

Another inconvenience of the LSC and MCLP models is due to the problem of simultaneous emergency calls. More precisely, it is possible to receive more than one emergency service demand call at the same time or in a very short time delay. Each of the service demand points must be covered; consequently, we may need to support the zones by more than one ambulance. The classical LSC and MCLP models are not able to provide such service.

In order to overcome the inconveniences of the static models, several approaches have been introduced in the literature. One approach consists of employing more than one ambulance to cover the simultaneous emergency demand calls.

The double standard model (DSM) [4] is an example of the models that use multiple ambulances in covering the demands. The DSM model is based on the assumption that all demands must be covered by an ambulance within r_2 minutes and a proportion α of the total demand must be covered within r_1 minutes. Obviously, $r_2 > r_1$.

The multi-coverage models try to handle the problem of the uncertainty in the demands. Stochastic programming is another approach that is used to take into account the uncertainty. In spite of the multi-coverage models, the stochastic programming models try to cover the uncertainty in a more explicit way. In the stochastic models, the origins of uncertainty are considered to be either the availability of the ambulances (vehicles) or the occurrence of the service requests at the demands points (see [1], [2], [6]).

Finally, due to the dynamic setting of the EMS systems, dynamic optimization models seem to be suitable choices in efficient covering of the EMS demand points. In this esprit, Gendreau et al. [5] has introduced a dynamic redeployment (relocation) problem (RP^t). This model is as an extension to the DSM [4]. The model (RP^t) maximizes the number of the demand points that are covered two times and minimizes the costs associated to the movements of the vehicles. The model contains a penalty parameter that takes all dynamic changes into account.

1.2 Results

In this paper, we are interested in proposing a new model that fits to the French EMS system. The new model is based on the RP^t (see [5]) and we show that (in the context of the French EMS system) the proposed model is more efficient than RP^t. In order to formalize the model, we introduce a new parameter into the RP^t model. We believe that the new parameter improves the ability of the model in covering the emergency demands. The parameter is computed by using different fluctuation patterns of emergency demands during a given period of time.

Once the model is built, we need to verify its abilities in covering the emergency demands. Hence, the presented model is compared to the RP^t in terms of the ability in covering the emergency service requests. The comparisons are based on the experiments that

have been carried out by using some real-world data sets. They have been collected through the French emergency medical service system. According to the numerical results, the proposed model provides a better coverage of the emergency demands.

The structure of the paper is as follows. The RP^t model [5] is reviewed and our dynamic model is presented in Section 2. The models are tested on real-world data sets that have been collected through the French EMS system. The computational experiments are reported in Section 3. Finally, the last section includes some conclusions.

2 DYNAMIC LOCATION AND RELOCATION MODEL

In this section, we present our dynamic location and relocation model. The proposed model can be considered as an extended version of the classical RP^t model that has been introduced by Gendreau et al. [5].

In the EMS' context, each service point covers some demand points (zones). One or more vehicles are associated to each service point and they are responsible to cover the demands. In some circumstances, one may need to cover a demand point (zone) by more than one vehicle. This is due to the curse of uncertainty and it is related to the potential of a point (zone) in producing more than one emergency requests during a specific period of time. This situation corresponds to reception of an emergency service request while the covering vehicle is busy because of giving service to another demand in the *same zone*. We will call these demands as *simultaneous* emergency service demands versus the *simple* emergency service requests (that correspond to the demands arriving during the availability of the vehicle). There will be a conflict if the two demand points (with emergency needs) are located in the same zone and are supposed to be covered by the same service point. We address this situation by introducing some parameters into the model. These parameters are computed by using the historical emergency demands' data of each zone.

2.1 The Dynamic Relocation Problem $\left(DRP^t\right)$

For the sake of completeness, we start by describing the classical RP^t model of Gendreau et al. [5]. In order to present the model, we will use the notations that are summarized in Table 1.

Table 1: Notations: parameters and decision variables.

---	***Description***
$i \in I := \{1,...,n\}$	i is a demand point and I is the set of all potential demand points.
$j \in J := \{1,...,m\}$	j is a service point (centre) and J denotes the set of all service centres.
$k \in K := \{1,...,\lvert K\rvert\}$	k is an ambulance and K is the set of all ambulances.
U_j	the maximum number of the ambulances that can be assigned to the service centre j.
d_i	denotes the mean density of the emergency demands at the point i.
d_i^1, d_i^2	mean density of the emergency service demands at the point i for a simple (d_i^1) or simultaneous (d_i^2) call.
r_1, r_2	the time thresholds to be respected in covering any demand point $(r_1 < r_2)$.
γ_{ij}	a binary parameter that denotes whether a demand point i is accessible from the service centre j in r_1 minutes.

δ_{ij}	a binary parameter that denotes whether a demand point i is accessible from the service centre j in r_2 minutes.
$\alpha \in [0,1]$	a real number indicating the proportion of all emergency service demands that must be covered in a given delay.
M^t_{jk}	a real valued parameter for controlling the relocations and movements of the vehicles at each period t; particularly, M^t_{jk} takes larger values in order to prevent long-distance travels of the vehicles.
$x_i^\lambda \in \{0,1\}$	to say whether the demand point i is covered λ times (for $\lambda \in \{1,2\}$).
$y_{jk} \in \{0,1\}$	to say whether the ambulance k is located in the service point j.

Associated to a given time t, there is a real valued penalty parameter M^t_{jk} that is incorporated into the model. This parameter plays an important role in the dynamic structure of the model and in the stability of the provided location plans throughout the day (or the operational period of the model). In fact, for a given t, the penalty parameter M^t_{jk} is associated to the relocation of ambulance k from its current position to service point $j \in J$. The value of this parameter is adjusted at any time t according to the different information regarding the vehicle k and the service point j. This information may contain frequent moves of the vehicle in the past, round trips, and relocations over long distances.

By using the presented notations, the model RP^t of Gendreau et al. [5] reads as follows:

$$\max \sum_{i=1}^{n} d_i x_i^2 - \sum_{j=1}^{m}\sum_{k=1}^{|K|} M^t_{jk} y_{jk} \tag{1}$$

$$\text{Subject to: } \sum_{j=1}^{m}\sum_{k=1}^{|K|} \delta_{ij} y_{jk} \geq 1 : \forall i \tag{2}$$

$$\sum_{i=1}^{n} d_i x_i^1 \geq \alpha \sum_{i=1}^{n} d_i \tag{3}$$

$$\sum_{j=1}^{m}\sum_{k=1}^{|K|} \gamma_{ij} y_{jk} \geq x_i^1 + x_i^2 : \forall i \tag{4}$$

$$x_i^1 \geq x_i^2 : \forall i \tag{5}$$

$$\sum_{j=1}^{m} y_{jk} = 1 : \forall k \tag{6}$$

$$\sum_{k=1}^{|K|} y_{jk} \leq U_j : \forall j \tag{7}$$

$$x_i^1, x_i^2 \in \{0,1\} : \forall i \text{ and } y_{jk} \in \{0,1\} : \forall j, \forall k. \tag{8}$$

In this model, the objective is to maximize the demand points that are covered two times and to minimize the costs related to the relocation of the vehicles. The constraint (2) ensures the absolute coverage of the demands within r_2 units (of time per minutes). The requirements related to the partial covering of the demands are expressed by the constraints (3) and (4). According to the constraint (3), $\alpha\%$ of all emergency demands is covered. The constraint (4) states that the number of vehicles waiting in r_1 units (of time per minutes) from the demand point i must be either at least equal to 1, if x_i^1 is equal to 1, or at least 2 if

$x_i^1 = x_i^2 = 1$. Constraints (5) say that any demand point i can be covered twice if and only if it is already covered at least once. According to the constraint (6), each ambulance must be assigned to a service centre. Finally, an upper bound is defined, by the constraints (7), on the number of the ambulances that can be assigned to a service point. Constraints (8) are the integrality constraints.

The model (RP^t) has been successfully applied in different countries. In spite of this fact, it can be changed in order to be casted into our case study in the context of the French EMS system. In fact, our case study has been carried out on Val-de-Marne (that is a county in France) where the intensity of emergency service demands is not high. In spite of this fact and in order to take into account the uncertainties, we need to provide the best possible coverage of the emergency demand points. To this aim, we introduce some new parameters in order to take into account different kinds of coverage. According to a given demand point, one may need to give more importance to the double coverage in comparison to another point. This fact is included explicitly in the new model. In order to present our model, we need to define two new parameters. We will use the parameters d_i^1 and d_i^2 for specifying, respectively, the mean occurrence number (intensity) of the all service requests and the mean occurrence number (intensity) of the simultaneous emergency demands at the demand point i. We summarize our Dynamic Relocation Problem (DRP^t) as follows:

$$\max \sum_{i=1}^{n} \left(d_i^1 x_i^1 + d_i^2 x_i^2 \right) - \sum_{j=1}^{m} \sum_{k=1}^{|K|} M_{jk}^t y_{jk} \tag{10}$$

$$\text{Subject to:} \quad \sum_{i=1}^{n} d_i^1 x_i^1 \geq \alpha \sum_{i=1}^{n} d_i^1 \tag{12}$$

and the constraints (2), (4) – (8).

In this model, the objective is to maximize the demand points that are covered and to minimize the costs related to the relocation of the vehicles. In order to cover the demand points, the model takes into account the weights associated to the two categories of service demands: d_i^1 and d_i^2 (see Table 1 for more precise definitions of d_i^1 and d_i^2). According to the constraint (12), $\alpha\%$ of all emergency demands is covered. The remaining constraints of the model are the same as the model (RP^t).

The differences between the models (RP^t) and (DRP^t) are essentially in the objective functions and also in the proportional coverage constraints. In a similar way to the (RP^t), the (DRP^t) maximizes the coverage of the emergency demands but the (DRP^t) model uses two parameters d_i^1 and d_i^2. Due to the randomness of the demands, there may happen some situations during which some simultaneous demands occur. By enforcing double coverage of the demand points, we can cover this kind of situations. To this aim, the parameter d_i^2 is used.

3 COMPUTATIONAL EXPERIMENTS

3.1 Data description

The models were used for the EMS system in Val-de-Marne, a county in France. The population of the county amounts to approximately 1.3 million inhabitants and it is divided into 47 quarters.

The EMS call centre of the county of Val-de-Marne receives more than 1000 calls per day, but just a small part of the calls require sending an EMS vehicle; that is, in general, between 20 and 30 calls per day. In our experiments, we suppose that each of the calls must be covered in less than 10 minutes and 8 ambulances are in use in the county. Furthermore, in our experiments, we suppose that 12 centres are in daily use (see Figure 1).

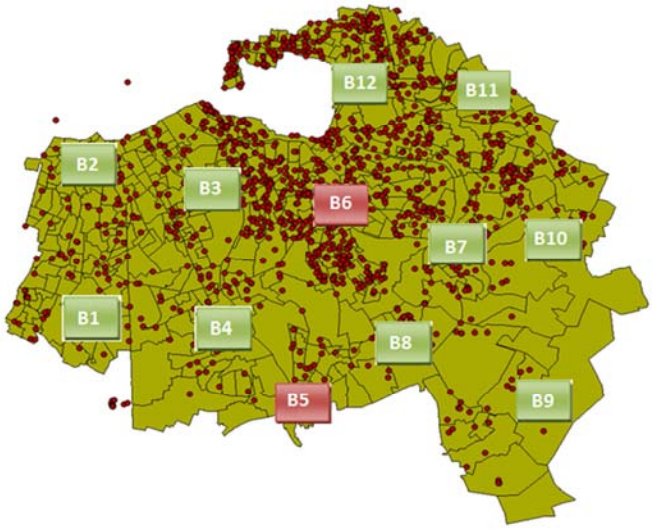


Figure 1: The 12 service centres (B1-B12) in the county of Val-de-Marne. B5 and B6 are currently operational and there is a plan to use the other centres.

For our computational experiments we used some recently collected data from the EMS system of the county. Data collection has been made possible by means of a new GPS localization system. It has been installed in the Hospital *Henri Mondor* that is located in the county of Val-de-Marne.

The standard solver IBM Cplex (version 12.2) has been used to solve the mathematical optimization models corresponding to the case study. Since the size of the models is not large, the models are solved in a quite short time, which is less than 2 seconds.

In order to compare the models (RP^t) and (DRP^t), we solved them under same conditions by using the same data sets. The performance of the models is measured by means of their ability in covering the EMS service demand points.

Different proportional covering percentages (i.e., α) have been taken into account. In fact, in our experiments, α varies from 90 % to 100 %.

3.2 Results

Figures 2 and 3 show the experiments that have been carried out on two consecutive time periods. The figures show the coverage proposed by the models (RP^t) and (DRP^t) (shown on the figures by RP and DRP, respectively).

Figure 2 shows the results for the starting time period. At this period the values of M_{jk}^t are all initialized by zero (for all j,k).

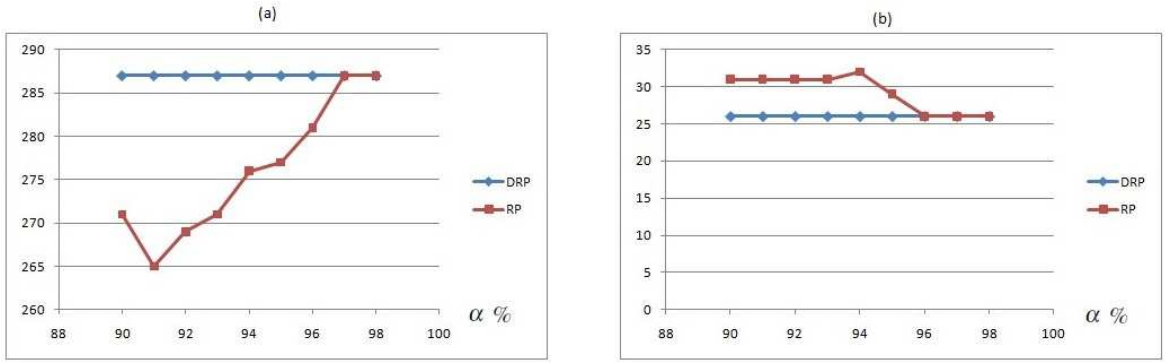


Figure 2: The number of the single (a) and double (b) covered demands (vertical axis) for different proportions of the total demands (horizontal axis).

Some observations on the results of the first period:

- For a given proportion of demand to be satisfied, the behaviour of the two models is significantly different. The value of the covered demands remains stable in the (DRP^t) model, but this value may decrease or increase in the (RP^t) model. This observation can be justified by reviewing the structure of the objective functions. In fact, the variables x_i^1 and x_i^2 (weighted by d_i^1 and d_i^2, respectively) are present in the objective function of (DRP^t), but this is not the case of (RP^t).
- According to Figure 2, the (DRP^t) presents an ambulance deployment policy with a better coverage in comparison to the (RP^t). The coverage includes all types of the calls, i.e., simultaneous demands as well as the non-simultaneous (i.e., simple) ones. In contrary to Fig. 2 (a), Fig. 2 (b) shows a better coverage provided by the (RP^t) model. In a similar way to the previous case, the difference is justified by the structure of the objective functions in the (DRP^t) and the (RP^t). The (RP^t) model includes only the x_i^2 (that is weighted by d_i^1), which privileges the double coverage.

In order to pass from the first period to the second period, we need to adjust the values of M_{jk}^t (where j indicates a service centre and k is a given EMS vehicle). The main issue is to reduce the movements of the vehicles. Based on this policy, the distance between service centres has been considered as the value of M_{jk}^t. Furthermore, we suppose that in the second period one of the vehicles is busy because of a mission. Hence, we must solve the optimization models with one vehicle less than the previous period, i.e., $k = 7$. The results are depicted in Figure 3.

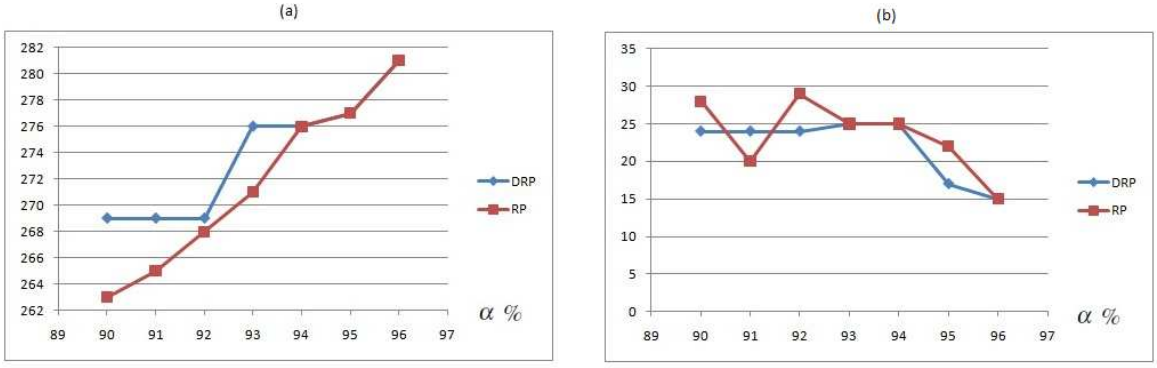


Figure 3: The number of the single (a) and double (b) covered demands for the Period 2. The vertical axis corresponds to the covered demand points and the horizontal axis shows the different values for α.

Some observations on the results of the second period:

- The results of the second period are significantly different from the results of the first period. We remember that the values of M_{jk}^{t} are adjusted in a way to reduce the total movements of the ambulances. Due to this fact, when we consider two (DRP^t) models corresponding to two different values of α, the corresponding objective functions of the models may be different. Indeed, *any similarity in the solutions of the first period may provide similar models for the second period*.
- Similarly to the first period, we observe that the (DRP^t) model provides solutions for which we have a better coverage of the *simple* demands. Furthermore, there is no more absolute superiority in the quantity of the *double* covered demands by the (RP^t) model in comparison to the (DRP^t) model.

4 CONCLUSIONS

In this paper, we presented a new dynamic location and relocation model in the context of the Emergency Medical Services. The model has been tested and verified on real-world data sets. According to the experiences, the model is solved efficiently for the studied cases. In spite of this fact, one will need some more efficient approaches for solving the large-scale programs. A set of experiments has been carried out to emphasise usefulness of the proposed model. To this aim, the model has been compared to one of the classical existing models. The numerical results show improvements in the coverage of the demands by using the introduced model.

Acknowledgments
This work has been supported by the French National Agency of Research (ANR) under the contract *Performance Optimization of SAMU* (ANR - POSAMU).

References

[1] Ball M.O., Lin L.F., 1993. A reliability model applied to emergency service vehicle location, Operations Research, Vol. 41, pp. 18-36.

[2] Beraldi P., Bruni M.E., Conforti D., 2004. Designing robust emergency medical service via stochastic programming, European journal of Operational Research, 158, 183-193.

[3] Church R. L., ReVelle C.S., 1974. The maximal covering location problem, Papers of Regional Science Association, Vol. 32, pp. 101-118.

[4] Gendreau M., Laporte G., Semet F., 1997. Solving an ambulance location model by tabu search, Location Science, Vol. 5, pp. 75-88.

[5] Gendreau M., Laporte G., Semet F., 2001. A dynamic model and parallel tabu search heuristic for real-time ambulance relocation, Parallel Computing, Vol. 27, pp. 1641-1653.

[6] ReVelle C. S. et K. Hogan., 1989. The maximum availability location problem, Transportation Science, Vol. 23, pp. 192-200.

[7] Toregas C., Swain R., ReVelle C.S., Bergman L., 1971. The location of emergency service facilities, Operations Research.

APPLICATION OF ANFIS IN THE VEHICLE TRACK APPROXIMATION

Polona Pavlovčič Prešeren, Bojan Stopar and **Oskar Sterle**
University of Ljubljana, Faculty of Civil and Geodetic Engineering, Jamova 2, Ljubljana, Slovenija
Jamova 2, SI-1000 Ljubljana, Slovenia
{polona.pavlovcic, bojan.stopar, oskar.sterle}@fgg.uni-lj.si

Abstract: Since adaptive neuro-fuzzy inference systems (ANFIS) are recognized as universal approximations, they became suitable tool for function construction from discrete data. In this paper, we discuss the problem of function approximation in 2D from discrete positional data in D96/TM coordinate system, gathered from GNSS-receiver along the track. In particular, we show the effectiveness of the proposed ANFIS method, as well as suitability in specific geodetic applications where continuous functions are needed for further processing. The experimental results confirmed that the proposed ANFIS method has potential in the vehicle track function approximation.

Key-words: Adaptive neuro-fuzzy inference systems (ANFIS), position data, GNSS (Global Navigation Satellite System), D96/TM coordinate system, track approximation, geodetic applications.

1 INTRODUCTION

Function approximation can be described as solving the problem, where finite set of data are the part of a continuous function, but only these finite set of data are known in the situation. From discrete data we have to re-construct function to determine the specific value in not-known data. Very often we use polynomial interpolation for function construction, for example Taylor expansion [1], when we have function values and *n* derivatives in one particular point, and Neville's algorithm for Lagrange polynomial [2], when we have functional values in tabular grid points for disposal [3]. Hermite polynomial is used when functional values in grid points are known as well as derivatives in those points [4]. The basic task of interpolation is to find the coefficients of the polynomial. Very often polynomial values oscillate around the function values, particularly near the end-points of the interval. Modern approach of function construction doesn't follow the same logic as the interpolation, when coefficients of the polynomial should be set. For example construction of neural networks (NN) [5] or adaptive network-based inference system (ANFIS) [6] follow well-constructed forms in the learning process, which could be further used in function computation at any point. When the structure in the training process is properly set and validated in the testing process, the same structure could be used anywhere in the definition area, sometimes also over the area.

ANFIS was originally proposed by Jang in 1993 [6], where explained a system model based on mathematical conventional tools, i. e. nonlinear function modeling, like differential equations [7], [8]. Several studies compared ANFIS and neural networks [9], [10]. ANFIS similar to NN constructs the input to the output data mapping, which is based on human knowledge, explained by the form of fuzzy if-then rules. Inputs go through the input membership functions and associated parameters, and through output membership functions and parameters to the output to set input/output map. Comparing to NN ANFIS has an advantage since neuro-fuzzy systems use prior knowledge, but NN start the training process from scratch. ANFIS allows us to use system of fuzzy-rules to approximate function and that means prior knowledge in the training process. Actually fuzzy-rules allow us to place different areas of treatment specifically in the initialization. ANFIS follows the approach of learning the rules (structure) and membership functions (parameters) from data. ANFIS applies two techniques in updating parameters. Since the process combines gradient descent

and the least-squares method, the approach is known hybrid learning method. It has been successfully implemented in several problem solving. Some authors [10] showed advantages of ANFIS over known traditional methods, especially in automatic searching of connections from input-output data and in complementing parameters without modifying the model structure.

In the geodetic tracking problems there are often only discrete data across the track for disposal. Very often we need also in-between well-defined values for further processing. In this paper we attempt to solve the situation, where discrete position data in the new Slovenia horizontal coordinate system D96/TM[1] are available. In this paper we explain, how to construct a continuous function using ANFIS model from those discrete data.

2 ADAPTIVE NETWORK-BASED FUZZY INFERENCE SYSTEMS

ANFIS structure is shown in Figure 1. It incorporates if-then fuzzy rules and provides tuning of membership function according to the known input-output data. ANFIS network consists of two parts: the first part is the antecedent and the second is the conclusion part. Both parts are connected to each other by rules to the network form. ANFIS structure actually follows five-layered structure and is often introduced as multi-layered NN.

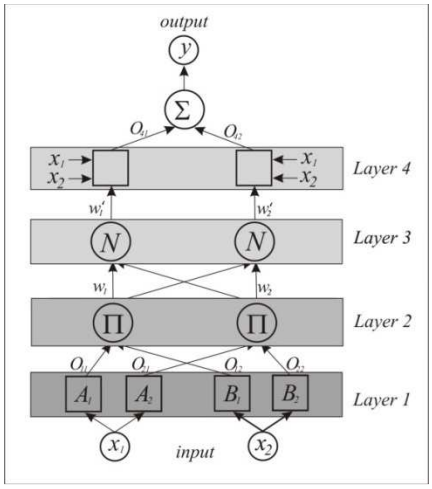


Figure 1: ANFIS multi-layer structure.

ANFIS implements a first order Sugeno-Takagi fuzzy system rules [11], [12], described as:

$$\textbf{rule 1}: if\ x\ is\ A_1\ and\ y\ is\ B_1, then\ f_1 = p_1x + q_1y + r_1$$
$$\textbf{rule 2}: if\ x\ is\ A_2\ and\ y\ is\ B_2, then\ f_2 = p_2x + q_2y + r_2$$

Layer 1*: *The input layer

$O_{l,i}$ is the output of the node *i* of the layer *l*. Every node *i* in the layer *l* is an 2daptiven ode with the node function, usually defined as bell-shaped functions [6]:

[1] The »new« horizontal coordinate system D96/TM is the Slovenian realization of ETRS89. The mean epoch of the three realized EUREF GPS campaigns in the Slovenian region was 1995.55 – that is why the name D96. The name D96/TM comes from: Geodetic Datum 1996, Transverse Mercator Projection. D96/TM is referred to the Geodetic Reference System 1980 (GRS-80), using Transverse Mercator projection. Horizontal coordinates are labeled: *e* for easting and *n* for northing.

$$\mu_{A_i}(x) = \frac{1}{1 + \left[\left(\frac{x - c_i}{a_i}\right)^2\right]^{b_i}} \quad (1)$$

where a_i, b_i, c_i are known as *premise parameters* and denote to adaptive nodes and *x* is the input node. The output nodes are defined as [6]:

$$O_{1,i} = \mu_{A_i}(x) \text{ for i = 1, 2}$$
$$O_{1,i} = \mu_{B_{i-2}}(x) \text{ for i = 3, 4} \quad (2)$$

and A_i or B_{i-2} are the linguistic labels associated with the specific node. Finally $O_{l,i}$ is a degree of membership for variable x to a fuzzy set, i. e. linguistical terms (A_1, A_2, B_1, B_2).

Layer 2: fuzzification layer

Every node of the second layer is a fixed node, known as *Prod*. Every node only multiplies the incoming signals and sends them out [6]:

$$O_{2,i} = w_i = \mu_{A_i}(x) \cdot \mu_{B_i}(y) \quad (3)$$

In this layer any other *T*-norm operator can be used, but it should perform as AND operator.

Layer 3: fuzzy-rule layer

Every node of the third layer is a fixed node, known as *Norm*. Each node calculates the ratio of *i*-th rule firing strength to the firing strenght of all rules [6]:

$$O_{3,i} = w_i' = \frac{w_i}{\sum w_i}, i = 1,2 \quad (4)$$

The outputs of this layer are known as *normalized firing strengths*.

Layer 4: output membership layer

In the fourth layer the nodes are adaptive, calculated as [6]:

$$O_{4,i} = w_i' \cdot f_i = w_i'(p_i x + q_i y + r_i), i = 1,2 \quad (5)$$

Where (p_i, q_i, r_i) is the parameter set of the i-th node. Nodes in this layer are called the *consequent parameters.*

Layer 5: defuzzication layer

In the fifth layer there is only one fixed node, which is labeled *Sum*. The node computes the overall output as the summation of the incoming signals [6]:

$$y = O_{5,i} = \sum w_i' \cdot f_i = \frac{\sum w_i' \cdot f_i}{\sum w_i'} \quad (6)$$

The ANFIS is trained by a hybrid algorithm; in the forward pass least-squares algorithm is employed to identify consequent parameters in the Layer 4. In the backward pass the errors are propagated backwards to update the premise parameters in the first layer using gradient descent algorithm. In such way minimization of the input-output data error is achieved.

ANFIS uses back-propagation or a combination of least squares estimation and back-propagation for membership function parameter estimation. The training process of the ANFIS can be stopped either when testing error is less than the pre-defined tolerance limit or when the number of learning iterations is reached.

3 EXPERIMENTAL RESULTS

In this section we present the results of experiments and the comparison and analysis of results between two different ANFIS structures. In a fuzzy inference system, basically there are three types of input space partitioning: grid, tree, and scattering partitioning. The first ANFIS structure used in this research was based on grid partitioning. Since it generates all the rules by enumerating all combinations of membership functions of the input data, the large amount of data occur and this can be time consuming. In this aspect we had to define two parameters, number of iterations as well as tolerance for early stopping. So in the next ANFIS structure we used different ANFIS utilization, i. e. subtractive clustering, for practical reason – to achieve faster training.

ANFIS approximation was demonstrated on the vehicle track, where coordinates (*e*, *n*) of 79 points were gathered in the national horizontal coordinate system of Slovenia, known as D96/TM. Positions were gathered using real-time-kinematic (RTK) GNSS method. To simplify the analysis and the explanation of the results, we show only graphical performance of two different approaches of ANFIS utilization.

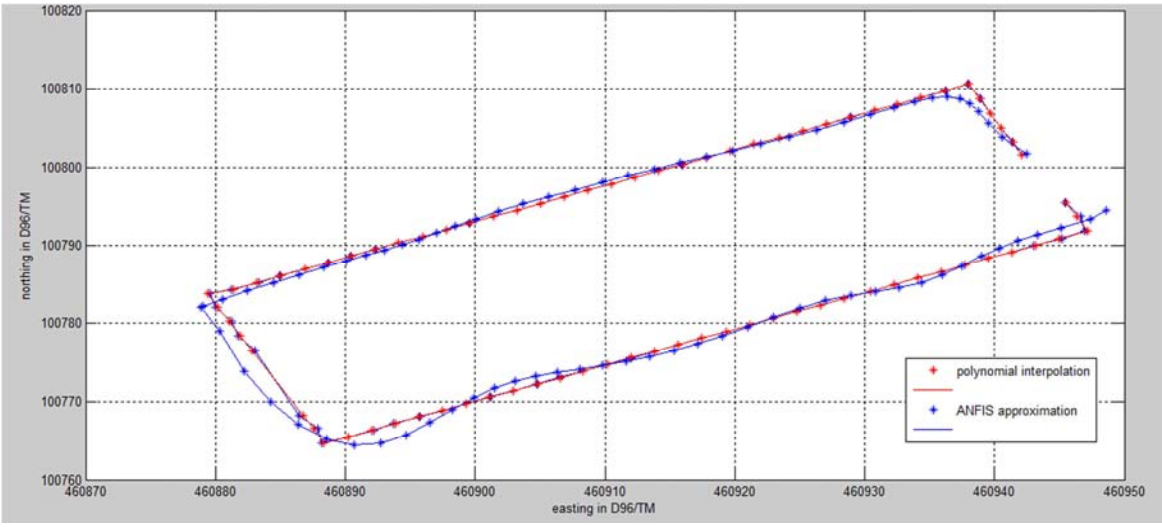


Figure 2: Time series of actual discrete data along the track and further continuous function generation using polynomial interpolation (red) and ANFIS performance (blue) using early stopping. Positions are given in D96/TM coordinate system.

The first structure of ANFIS, which uses grid, used 5 bell-shaped membership functions, while the output was linear. The ANFIS learning was set to maximum 1000 iterations. As seen from Figure 1 deviations in ANFIS approximation occur, especially near the end of the track and in situation, where the vehicle obviously changes direction of movement. The maximal error in the *e*-component was 2.48 m and in *n*-component 2.68 m. Since the ANFIS stucture used for training can be time-consuming and led us to deviations in range of several meters, different ANFIS structure was used in further data processing.

The second structure of ANFIS, which used scattering partitioning, performed faster since the function did not perform iterative optimization; only each cluster corresponded to the specific fuzzy rule. Function approximation was dependent on the different cluster radius as well as the number of rules, that both varied in our study. ANFIS predicted values were compared to a set of test-data and further evaluated in the process of ANFIS structure improvement process.

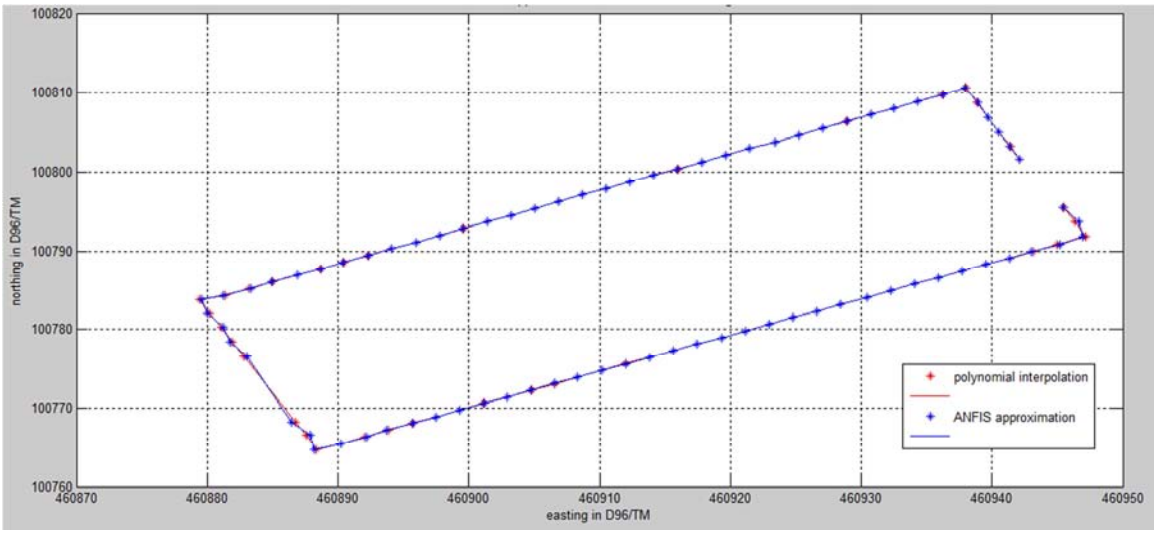


Figure 3: Time series of actual discrete data along the track and further continuous function generation using polynomial interpolation (red) and ANFIS performance (blue) – approximation with best set of fuzzy rules. Positions are given in D96/TM coordinate system.

ANFIS approximation was evaluated using equation (7) and further compared to given positions; approximation results were evaluated using root mean square error (*RMSE*) function as the error criteria:

$$RMSE = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (f_i - y_i)^2} \tag{7}$$

where f_i is the ANFIS output and y_i the known value. δ_{min} and δ_{max} are minimal and maximal values of differences. Smaller the *RMSE* is, better is the accuracy of approximation.

Table 1: Statistics for 1st ANFIS structure (graphical representation in Figure 2).

***e*-component**	
Minimal value of differences δ_{min}	-3.084 m
Maximal value of differences δ_{max}	2.478 m
Mean value of differences $\bar{\delta}$	0.025 m
RMSE	0.337 m

***n*-component**	
Minimal value of differences δ_{min}	-3.084 m
Maximal value of differences δ_{max}	2.684 m
Mean value of differences $\bar{\delta}$	0.025 m
RMSE	0.280 m

Table 2: Statistics for 2nd ANFIS structure (graphical representation in Figure 3).

***e*-component**	
Minimal value of differences δ_{min}	0.000 m
Maximal value of differences δ_{max}	0.000 m
Mean value of differences $\bar{\delta}$	0.000 m
RMSE	0.208 m

***n*-component**	
Minimal value of differences δ_{min}	-0.011 m
Maximal value of differences δ_{max}	0.012 m
Mean value of differences $\bar{\delta}$	0.000 m
RMSE	0.000 m

4 CONCLUSIONS

The main objective of this paper was to demonstrate that ANFIS can be addressed to successfully approximate continuous function from discrete position data along the vehicle track. Experimental results confirm that the method is effective and can be used as an alternative to the traditional polynomial interpolation.

References

[1] Wong, S. C., Barhorst, A. A., 2006. Polynomial interpolated Taylor series method for parameter identification of nonlinear dynamics system. Journal of Computation and Nonlinear Dynamics, Vol. 1, No. 3, 248 – 256.

[2] Schenewerk, M., 2003. A brief of basic GPS orbit interpolation strategies. GPS Solutions, Vol. 6, No. 4, 265 – 267.

[3] Neta, B., Sagovac, C. P, Danielson, D. A., Clynch, J. R., 1996. Fast interpolation for Global Positioning System (GPS) Satellite Orbits. In: Proc. AIAA/AAS Astrodynamics Specialist Conference, San Diego, CA, Paper Number AIAA 96-3658.

[4] Andrews, G. E., Askey, R., Roy, R., 1999. Hermite Polynomials. In *Special Functions*, Cambridge, England: Cambridge University Press, pp. 278-282.

[5] McCulloch, W., Pitts, W., 1943. A Logical Calculus of Ideas Immanent in Nervous Activity. Bulletin of Mathematical Biophysics, Vol 5, No. 4, 115–133.

[6] Jang, J. S. R., 1993. ANFIS: Adaptive network based fuzzy inference system, IEEE Transaction Systems, Man & Cybernetics, Vol. 23, No. 3, 665–683.

[7] Roger Jang, J. S., Sun, C. –T., 1995. Neuro-Fuzzy Modeling and Control. Proceedings of the IEEE, Vol. 83, No. 3, 378 – 406.

[8] Jang, R. J. S., Sun, C. –T., Mizutani, E., 1996. Neuro-Fuzzy and Soft Computing: a computational approach to learning and machine intelligence. Prentice-Hall.

[9] Yun, Z., Quan, Z., Caixin, S., Shaolin, S., Yuang, L., Yang, S., 2008. RBF neural network and ANFIS-based short-term load forecasting approach in real time price environment. IEEE Tran. Power Syst., col. 23, no. 3, 853-858.

[10] Zuperl, U., Cus, F., 2003. Optimization of cutting conditions during cutting by using neural networks, robotics and computer-integrated manufacturing, Vol. 19, 189-199.

[11] Takagi, T., Sugeno, M., 1985. Fuzzy identification of systems and its applications to modeling and control. IEEE Trans. Syst. Man Cybern, Vol. 15, 1, 116-132.

[12] Palit, A. K., Babuska, R., 2001. Efficient Training Algorithm for Takagi-Sugeno Type Neuro-Fuzzy Network. Proc. IEEE Int'l Conf. Fuzzy Systems, Vol. 3, 1367-1371.

The 12th International Symposium on
Operational Research in Slovenia

SOR '13

Dolenjske Toplice, SLOVENIA
September 25 - 27, 2013

Section VIII:

Creative core FIS - Simulations

AGENT APPROXIMATION MODELLING AND SIMULATION: MISSING PERSON INCIDENT CASE STUDY

Jernej Agrež and **Nadja Damij**
Faculty of information studies in Novo mesto
Ulica talcev 3, SI-8000, Novo mesto, Slovenia
{jernej.agrez,nadja.damij}@fis.unm.si

Abstract: Paper presents the agent approximation modelling and simulation approach developed on the basis of TAD methodology and within iGrafx simulation environment. The significant contribution of the paper is the presented approach, which is demonstrated within the context of the public safety and is further applied in regards to the missing person incident investigation. With developed approach we analyse such investigations and point out the critical success factors. Not only does the paper address the important answers, it also proposes new research questions that additionally contribute to the agent approximation approach under development.
Work supported by Creative Core FISNM-3330-13-500033 'Simulations' project funded by the European Union, The European Regional Development Fund. The operation is carried out within the framework of the Operational Programme for Strengthening Regional Development Potentials for the period 2007-2013, Development Priority 1: Competitiveness and research excellence, Priority Guideline 1.1: Improving the competitive skills and research excellence.

Keywords: agent approximation, modelling, simulation, process management, public safety, missing person incident.

1 INTRODUCTION

Agent-based modelling offers a way to model social systems that are composed of agents who interact with and influence each other, learn from their experiences, and adapt their behaviours so they are better suited to their environment [9]. As such it can be used as powerful tool for assessment of public safety scenarios.
In year 2012 there were “135 missing person incidents”[12] (MPI) reported and officially investigated in Slovenia. In the same year Ministry of interior issued basic guidelines used as a basis for a midterm “Police work and development plan for period 2013-2017”. As such the development plan highlights “cooperation between law enforcement authorities and local community” [13] that includes also sustaining public safety. Joint MPI investigation taskforce, composed out of law enforcement and actively participating local community, could be an important step forward, suggested also by ”Balanced scorecard model within law enforcement authority” [2]. We believe not only such a solution provides better MPI investigation results but also presents a solution that reaches beyond field measured or proven actions.

Taking in consideration that MPI is a very fragile part within the field of public safety, it is of great importance to be assessed and analysed in virtual environment, but within a real life based scenario. In this article we will present approximation to a multi-agent model that we will build with “Tabular Application Development (TAD) methodology” [3,4]. We will apply the model on the real MPI case, where public and private investigations interacted and created visible results. We will also present the translation of the model into decision making simulation that will provide us with critical success factors of the joint MPI investigation. Section 2 investigates process based complex modelling where we introduce TAD methodology as an option for complex modelling approximation. In the same section we present iGrafx as a tool for simulation of a complex environment. We continue with assessment of complex modelling within the field of public safety in section 3, and introduce missing person incident investigation practice in Slovenia. In this section we present also the

real missing person incident case modelling and simulation with agent approximation approach. In section 4 we conclude with the results and propose further research directions.

2 PROCESS BASED COMPLEX MODELLING

A complex system is a group of agents (individual interacting units), connected by some sort of communication, relation or any other type of interaction [5]. A usual way of defining the system would be through the isolation of its elements and then positioning them in hierarchical composition [16]. To be able to capture and understand a complex system we need to get an overview of every of its parts (agents) together with relations among them [18]. When modelling complex systems, relations are of the crucial importance.. If the model is set as a basic pyramid shaped hierarchy without any complex relations, it would fail to provide us with accurate answers that could represent real world state [16]. On the other hand roles of elements and relations can be switched.. Instead of an agent acting as a primary widget of the system connected to the other elements through the relation, a hierarchical model can be created where the relations could connect through elements. Such an approach enables us to define a relation model similar to "representation of the hierarchy of operation, sequences, tests and procedures" [15]. If such approach is further compared to the state of art, it presents a close match to the modelling hierarchy presented in "TAD methodology" [3]. This suggests that modelling of complex social environments could be conducted also with a business process based methodology.

2.1 TAD and complex modelling approximation

TAD represents simple concept for description of the organization using several tables [3]. Originally intended for information systems development and business process reengineering it can be adopted for modelling a complex system. First and second phase of TAD methodology includes framework, how to capture and map system functionality that includes following tables: Entity table, Activity table and Task table. TAD methodology enables a creation of a model with any number of agent approximations (AA) that can incorporate any number of activities. We define AA as a decision making individual, pair or group of people that is modelled on a basis of real life events. Therefore such a model precisely summarises process reality and maps it in digital environment. On the other hand, the reality based facts do not allow us to design true agents that would have the possibility of fully independent decision making. Their actions are limited by activities and decisions that are part of modelled process.

2.2 iGrafx and simulation of a complex environment

iGrafx is a business process modelling and simulation environment that is fully compatible with TAD methodology as well as with AA upgrade of the TAD. We choose iGrafx according to its possibilities to define attributes of a single included activity from time, resource, input/output and risk perspective. iGrafx also enables designing of any number of AA that can differ by their hierarchical structure, resource range, process inclusion intensity and complexness of their decision making logic. Based on process phase definition the process implementation can be investigated from a transition perspective, which consequently enables a possibility to determine importance level of different process patterns based on conducted activities or taken decisions. At the same time, the simulation environment itself supports such approach with the integrated pallet of useful statistical tools that evaluate relations among different variables within the process. For the purpose of this

paper the most important variables are transactions investigated firstly from the within the process perspective and secondly from the single activity perspective.
The proposed framework of the simulation is built from the main process layer that can be compared with a business process within the TAD methodology. Further on we develop parallel process layers that serve as AA. Agent approximations have the ability to run independently of the main process and can conduct their influence without taking any consequences. On the other hand they could be completely subjected to the main process, being unable to run without triggered input. Such range of flexibility creates a possibility to determine attributes of the agent approximations in a way that is a close approximation of reality as possible. Within the suggested framework, the ability to simulate simultaneously "as-is" as well as "to-be" state of the main process, with the same agent approximations is achieved. On the basis of the existing process completely new process scenario can be designed to simulate side effects of the main process as well as its comparison to another process, community, environment, etc.

3 PUBLIC SAFETY AND COMPLEX MODELLING

Public policy analysts note that decision making in western societies is mostly rational choice based, often involving cost effectiveness calculations, after carefully considering a variety of proposals [7]. Organizational orientation, based on the rationality and cost reduction is constantly present and is due to limited financial resources sometimes unavoidable. But even though, the overall work effectiveness does not necessary reaches the desired level. This is the reason why private contractors became important part of public safety mechanisms [17]. Public safety concerns variety of different fields, such as: public health incidents, social safety, industrial accidents, natural disasters [8], as well as law-enforcement and counter terrorism activities [17, 10]. The cooperation between the public and private organizations within the scope of public safety differs according to the area where they emerge. At the same time, when such relation is established, terms of cooperation are not necessary determined in details, as practiced in public safety crowdsourcing approach: America's Most Wanted [17]. Yet overall work effectiveness increases. And to be able to assess the increase and determine what the critical success factors of such cooperation are, complex modelling of public safety scenarios enable us to map the cooperation process as well as highlight the most important cooperation elements.

3.1 MPI investigation in Slovenia

A definition of a missing person is sometimes misinterpreted as person who is wanted by law enforcement authorities, on the basis of criminal or terrorism activity. The missing person incidents could be indirectly connected also with a law breaching background, but a disappearance of a missing person must not result of one's prosecution. If so, such person is not missing, but wanted by law. According to Police obligations and enforcement authority law in Slovenia, Police as a law enforcement authority is responsible for conducting a missing person search, if, due to circumstances, possibility exists that missing person is in need for help[2]. A missing person search warrant is regulated by "Instructions for arrest, missing person and missing things warrants", while operational tactics and methodology is defined in "Guidelines for Police work in MPI" [14].

3.1.1 MPI case study

The following real-life case is based on events that took place in May 2013. The entities involved in the MIP incidents were the following: victim, victims` family, police patrol, police call centre, individual police officer, Human rights ombudsman, Prosecution service, public, SAR responders and Distress call centre. Due to the privacy concerns no personal information that could reveal identity of people, who were involved in the case, will be revealed. The time scope of the incident was defined to be seven days according to most important activities that took place within the identified timeframe. Even though roots of the incident reaches far back in the past and consequences could be present long time in the future, we will not include them in the research scope due to indirect connection with the topic we present. MIP activities that remain in our research scope can be divided by days. Day 1: victim cleared internet history, temporary internet files, cookies and trash bin content with Ccleaner – software tool for PC optimization and cleaning. Day 2: victim left home between 8 AM and 3 PM. Victim had been seen in public the same day twice. First contact - 0,5 kilometre away from home and second contact around 4 PM and 1,8 kilometre away from home, heading approximately NNW direction. Day 3: victim had been identified on cash machine surveillance camera recording, approximately 40 kilometres away from home. New heading of victim`s movement - NW. Day 4: victim`s family asks for SAR responders to support the search. Day 5: victim had been seen 66 kilometres away from home. Heading remains the same – NW. Victim`s family made contact with Human rights ombudsman and Prosecution service due to their dissatisfaction with police work. Around 10 PM family received an e-mail from the victim, explaining that the victim is alive and expressing few thoughts about dissatisfaction in life. Day 6: Victim had been found by family approximately 80 kilometres away from home in a shelter for homeless people. Day 7: Victim makes contact with SAR responders and eventually they meet, discuss the situation and decide how to close the case in constructive way.

3.2 MPI modelling and simulation solution

We developed the model, beginning with indentification of the entities that will take role as agent aproximations. We adopted the entity table of TAD methodology and adjust it in form of matric system [1], to be able to present how each AA influences another. At this point we defined the following AA roles: influencing AA (the one that triggers influence), influenced AA (the one that changes behaviour under infleunce) and neutral AA (the one who influences only activity flow). Further more, we mapped MPI activity flow with the TAD Activity table, excluding any decision making. The activity flow consists of one beginning and one end, and inbetween all activities are lined up within the time order as it was in real situation. According to the fact that activity flow already took the place in the past, any additional decisin making would present deviation from captured reality. From the present perspective, the activity flow is inevitable and should not be treated differently. Adjustments to the Activity table that we condusted are following: We added a column that defines the time dimension of the process and at the same time we implemented coloring of the patterns, due to weakly defined business process – work process relations of the MPI. Coloring of the patterns together with tabular separation reveals work processes within weakly defined enviroment more clearly and makes it easier to understand process intuitively. Further more we added the agent aproximation matrix and model decision making agent aproximations in additional layers of the activity table. Agent aproximations were based on the decision making that follows the predefined protocols, law based directives, past experiences and knowledge, personal judgement, prejudice, emotional lability, or any other kind of influence.

Not every agent aproximation consists of all possible influences, but it is important to not neglect those that could importantly deviate their modelled behaviour. We adjusted the activity table in a way that one agent aproximation replaced several organizational departments and different kinds of influeces replaced the entities. If we previously defined interaction in agent aproximation matrix, we must now include AA as influence as well, but at the same time we must define such relation as activity in influencing AA, otherwise the influence is never realized but exists only as possibility. With the activity layer and AA layers we allready got the general insight of influences in directions: influencing AA – influenced AA and influencing/influenced/neutral AA – activity flow. If decision making of single AA would be elemental, hierarhical and would not interfere within the itself, the number of possible AA influences would evolve in a predictable manner. Such a process would be simple to analyse but in fact it is far from real life public safety processes. According to high complexitiy of AA influences, it is neccessary for us to develop a simulation that will incorporate and connect activity flow together with AA influences. For this purpose we translated the activity flow and AA tables into iGrafx proces diagrams. To be able to implement every AA into activity flow as an influence, we designed the whole MPI in single process layer, but created different starting points for each specific AA and for the activity flow. This is neccessary for the AA to run with independant trigger that send activation signal (transaction). Further more we created the influence connection poinsts, previously identified with TAD. For this purpose we used the attributes that are similar to the programming variables and can communicate data (information) and manage the flow of transactions through a process. For example, the attributes can set the duration of an activity based on the specifics of a transaction, control the flow of specific transactions through a decision output, or set global controls that can affect multiple transactions or activities [6]. Through AA influences that are simulated as yes/no or true/false decisions and through the activity attributes we designed process scenario that reflected real MPI investigation. To be able to analyse it we ran the simulation with predefined analytical queries based on trasaction count within the activity flow and decision making within every single AA that is supported by statistically based business process management approach Six sigma.

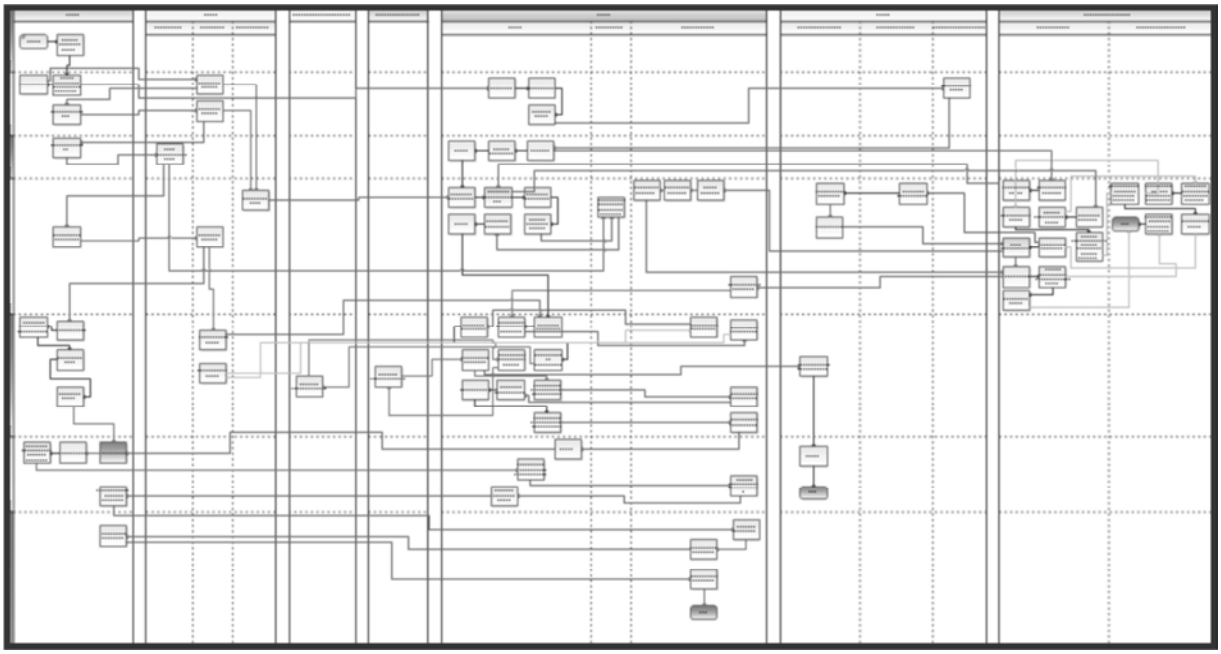


Figure 1: MPI investigation process diagram

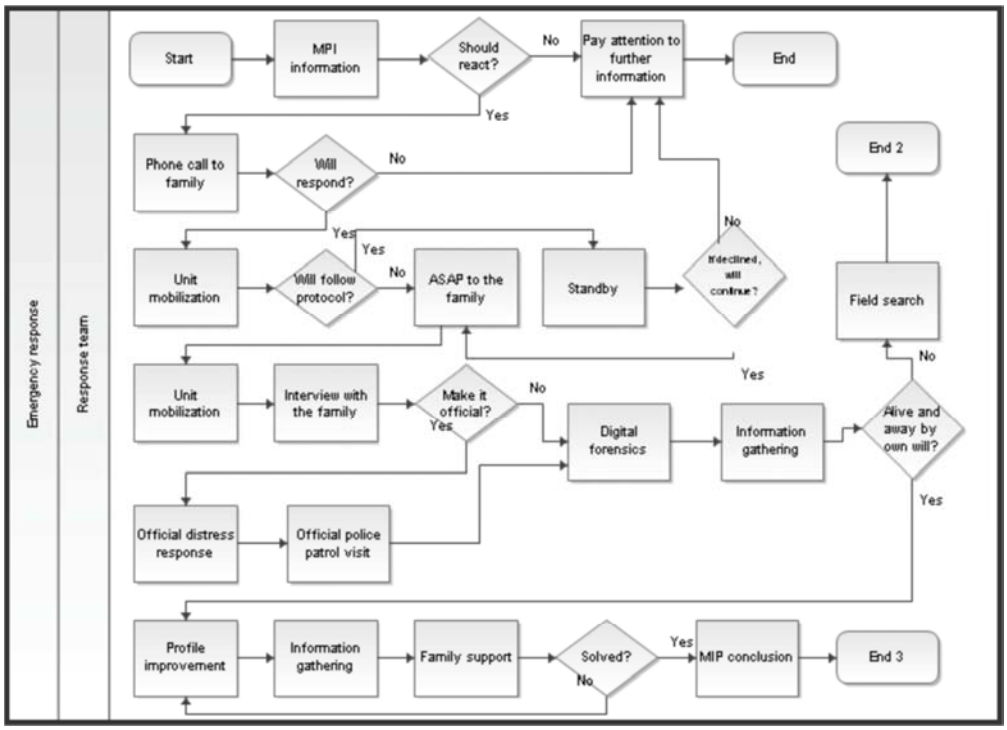


Figure 2: Example of an agent approximation

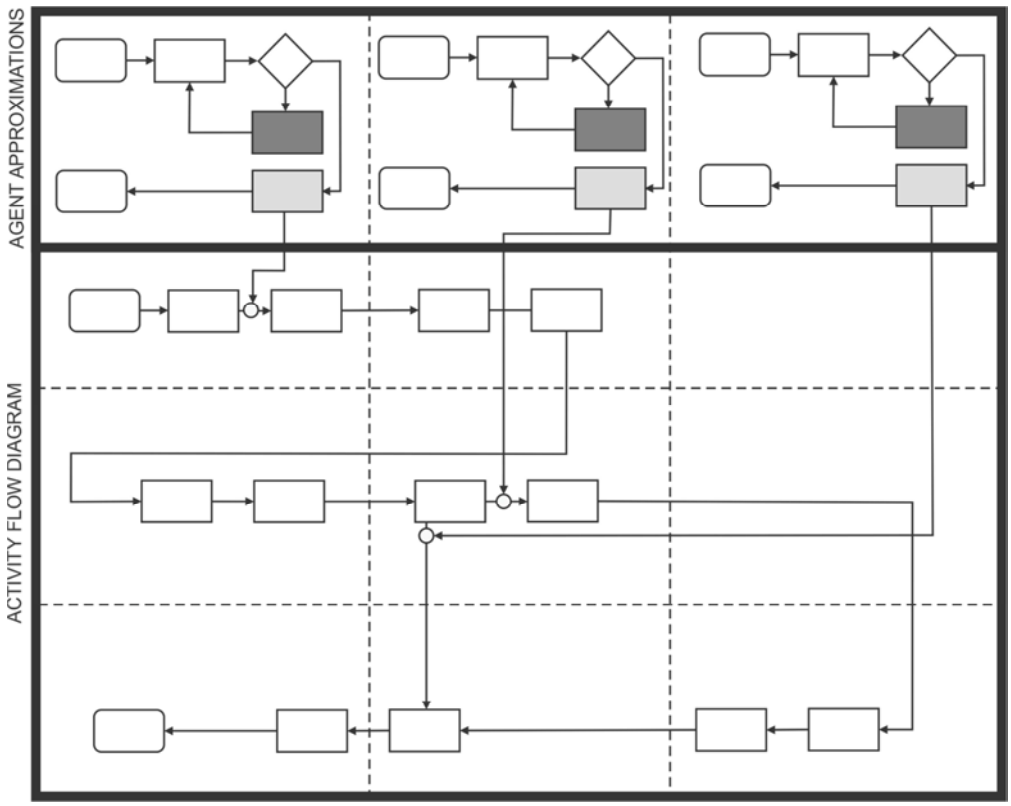


Figure 3: Agent approximation modeling concept

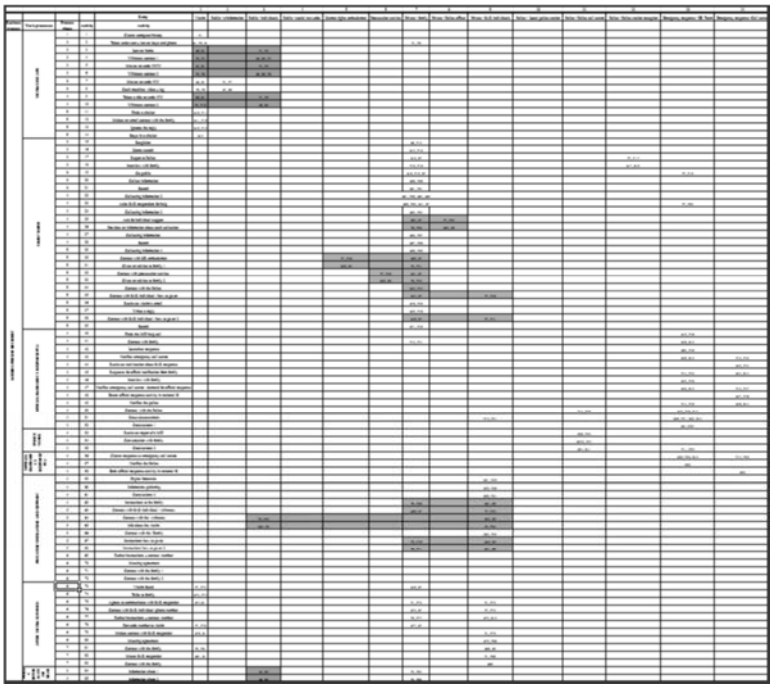


Figure 4: Colored process patterns within the TAD activity table

4 RESULTS AND CONCLUSION

A development of the proposed AA solution and real life MPI application provided us with the important answers concerning the critical success factor of the investigation. We were capable to determine the high importance of family`s initiative to do whatever is necessary to find the missing person. As well we detect the crucial interaction among community, family and private support which present: wide range of information, momentum sustaining, special knowledge and information. Such interaction presents the critical element for the process to be realized within the desired output. The absence of any of these three elements would consequently lead to alternative ending that can only be predicted as what-if scenario and would not necessarily be realized. The simulation enabled us with another important insight to a MPI: partial endings of the process that are still treated as successful output even though the process itself never reaches the end as it had happened in real case. Such partial endings are highly correlated with family`s priorities defining on how far the family wants to go into the investigation process. At the same time the results show us that if such partial ends exit, other entities involved in process could take partial end as the final end, and drop out of the activity flow, according to their regulations, practice, etc. Work of entities that are considered as public services presented 18% of all activities within the investigation. Remaining 81% of activities were conducted by two groups of entities; the major group with 61% of activities represents the people that were involved in investigation in completely private manner, and the 18% presents the entities that used their official public status to provide information and knowledge for private investigation. In fact they were acting against the law and could be prosecuted for abusing their position, even though they present one of the three critical success factors of the investigation.

Concerning agent approximation modelling and simulation we can clearly consider this research as a work in progress. It provided us with the approach on how to successfully use business process management methodologies for studying complex social environments

through the modelling and simulation. At the same time it revealed new unanswered questions such as: how to qualitatively evaluate AA influences, how to simulate process based learning, how to define and manipulate process event horizon, etc. The suggested approach presents a solid base for further process research within the modelling and simulation areas that can be used for purposes of public safety as well as in other complex social environments.

References

[1] Agrež, J., Damij, N., (2013): A layer-based, matric oriented business process simulation solution. Proceedings of the 7th European Computing Conference, WSEAS, Dubrovnik, pp. 167-172.

[2] Aristovnik, A., Cankar-Setnikar, S., Čadež, S., Keševljevič, A., Pečarič, M., Pevcin, P., Rakar, I., Seljak, J., Tomaževič, N., Mencinger, J., (2012): Establishment of a efficiency, effectiveness and quality measurement system within Slovene national Police department.

[3] Damij, T., (2000): An object-oriented methodology for information systems development and business process reengineering. Journal of Object - Oriented Programming, vol.: 13, no.: 4, pp.: 23-34.

[4] Damij, N., (2007): Business process modelling using diagrammatic and tabular techniques. Business Process Management Journal, vol.: 13, no.: 1, pp.: 70-90.

[5] Fichter, L. S., Pyle E.J., Whitmeyer S. J., (2010): Strategies and Rubrics for Teaching Chaos and Complex Systems Theories as Elaborating, Self-Organizing, and Fractionating Evolutionary Systems. Journal of Geoscience Education, vol.: 58, no.: 2, pp.: 65-85.

[6] iGrafx LLC, (2013): iGrafx 2013 tutorials. Retrieved from http://igrafx.com/landing/download.html?name=iGrafx%202013%20Tutorials&ao_f=0137&ao_d=d-0001&c_type=c.

[7] Hebenton B., Jou, S., Chang, Y., (2010): Developing Public Safety and Crime Indicators in Taiwan. Asian Criminology, vol.: 5, no.: 1, pp.: 45-67.

[8] Jin, J., Zhao, S., Hu, B., (2012): Defining the Safety Integrity Level of Public Safety Monitoring System Based on the Optimized Three-dimension Risk Matrix. Proceedia Engineering, vol.: 43, no.: 1, pp.: 119-124.

[9] Macal, C. M., North, M. J., (2010): Tutorial on agent-based modelling and simulation. Journal of Simulation, vol.: 4, no.: 1, pp.: 151-162.

[10] Plecas, D., McCormick, A. V., Levine, J., Neal, P., Cohen, I. M., (2011): Evidence-based solution to information sharing between law enforcement agencies. An International Journal of Police Strategies & Management, vol.: 34, no.: 1, pp.: 120-134.

[11] Police obligations and enforcement authority law. (ZNPPol). Ur. l. RS, št. 15/2013 (18.2.2013).

[12] Republic of Slovenia, Ministry of the interior, Police. Annual Police work report for year 2012. Retrieved from http://www.policija.si/images/stories/Statistika/LetnaPorocila/PDF/LetnoPorocilo2012.pdf

[13] Republic of Slovenia, Ministry of the interior, Police. A basic guidelines for preparation of a midterm Police work and development plan for period 2013-2017. Retrieved from http://www.mnz.gov.si/fileadmin/mnz.gov.si/pageuploads/DPDVN/Usmeritve/Temeljne_usmeritve_2013-2017.pdf.

[14] Republic of Slovenia, Ministry of the interior, Police (2013). Missing person. Retrieved from http://www.policija.si/index.php/en/pogosta-vpraanja/1176.

[15] Richardson, K. A., (2011): TECS: A Browser-Based Test Environment For Complex Systems. Emergence : Complexity and Organization, vol.: 13, no.: 1-2, pp.: 213-229.

[16] Seck, M. D., Honig, H. J., (2012): Multi-perspective modelling of complex phenomena. Computational and Mathematical Organization Theory, vol.: 18, no.: 1, pp.: 128-144.

[17] Simeone, M. J., (2008): Integrating Virtual Public-Private Partnerships into Local Law Enforcement for Enhanced Intelligence-Led Policing. Homeland Security Affairs, Proceedings of the 2008 Center for Homeland Defense and Security Annual Conference, Monterey.

[18] Vespignani, A., (2012): Modelling dynamical processes in complex socio-technical systems. Nature physics, vol.: 8, no.: 1, pp.: 32-39.

A KNN BASED ALGORITHM FOR TEXT CATEGORIZATION

Jože Bučar and **Janez Povh**
Laboratory of Data Technologies
Faculty of Information Studies, University of Novo mesto
Ulica talcev 3, SI-8000 Novo mesto, Slovenia
{joze.bucar,janez.povh}@fis.unm.si

Abstract: In the recent decade categorization of web texts has experienced increased attention. Huge amount of textual information available on the web emerged a need to find and obtain relevant information for strategically supported decisions. There are many machine learning algorithms dealing with text categorization and classification issues. In the paper the experiment has been conducted on the k-Nearest Neighbor (KNN) classifier. Because of its simplicity and effectiveness it is widely applied method in a field of machine learning and pattern recognition.

Keywords: KNN, text categorization, text classification, text mining, web mining, data mining

1 INTRODUCTION

Exponential growth of content available on the World Wide Web offers enormous collection of textual resources. The increasing interest has emerged rush to integrate new processes and features, which brings together scientists from various fields like computational linguistics, data mining, computer science, machine learning, graph theory, neural networks, sociology, and psychology. Automatic text categorization became a significant tool to utilize text information and contribute to more efficient work.

Classification techniques have been used extensively. They have been applied to filter and route emails, identify different languages, classify genre, and determine the degree of readability of a text. For that reason, data miners use various tools and a wide range of learning algorithms such as Naïve Bayes (NB) probabilistic classifiers [7], Centroid-Based Classifier (CB) [3], Decision Trees classifiers [16], Decision Rules [8], regression methods [19], Neural Network [15], KNN classifiers [10], [16], [20], Support Vector Machines (SVM) [5], [16], etc. KNN, as a lazy learning instance-based algorithm, is commonly used for text categorization, especially because of its simplicity and low error rate.

About 80% of the information created and used by enterprises is unstructured data located in content [4]. Unstructured data consists of information that doesn't fit neatly into rows and columns of a spread sheet or a table (e.g., unstructured text, audio, video data, and also likes). Unlocking this holds huge potential. Text categorization is essential in information retrieval and text mining; both industry and academia are aware of its advantages. Especially organizations related to business, sale, finance, etc., quickly realized the importance of additional information, which can be useful in providing structural, organizational, business solutions, and decision support. For that reason, we conclude that text categorization technology is fundamental method in retrieval of textual information, and can provide answers with important research value in the future.

The rest of the paper is organized as follows: Section 2 introduces basic concepts of traditional KNN classifier. An experiment of KNN based algorithm for text categorization and its evaluation of efficiency on predicted category are given in Section 3. Finally, Section 4 concludes the paper.

2 The KNN classifier

In data mining, the KNN is one of the most important non-parametric methods and supervised learning predictable algorithm for classifying objects [21]. The KNN algorithm is

amongst the fastest, simplest, and easy to conceptualize of all machine learning algorithms. Prediction of the test sample's category is based on the *k* training samples that are the nearest to the test sample; where *k* is positive integer and is usually small. We then assign the category of the test sample according to the category with the largest category probability. Euclidean distance is most commonly applied metric for continuous variables. The optimal selection of *k* depends on the data or can be selected by one of heuristic techniques. When classifying textual information, larger values of *k* in general reduce the effect of noise on the one hand, but it makes boundaries between classes less distinct on the other hand.

Fig. 1 shows the visual presentation of KNN classifier. The triangle represents the test sample and it should be classified either to the class of circles or squares. At this point let us assume that *k* is equal 3. Test sample is therefore assigned to the class circles, since there are 2 circles and only 1 square inside the inner circle. If *k* is equal 7, then it is assigned to the class of squares, because there are 4 squares and only 3 circles inside the outer circle.

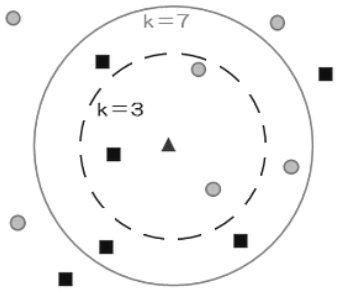


Figure 1: Example of KNN classification [11].

In order to classify unknown documents we must pre-process documents [2], [6], [21]. To Guo's [2] six sub-components of data pre-processing: document conversion function, word removal, word stemming, feature selection, dictionary construction, and feature weighting, we add transform cases, tokenization, and word lemmatization. The list of functionalities [2] is described and supplemented:

(1) Document converting – converts different types of documents to plain text format.
(2) Transform cases – transforms all characters in documents to either lowercase or uppercase, respectively, we usually transform to lowercase (e.g., KNN -- knn).
(3) Tokenization – splitting the text of a document into a sequence of tokens - words, phrases, symbols, or other meaningful elements (e.g., [Andy loves candy] -- [Andy] [loves] [candy]).
(4) Word removal – removes topic-neutral words such as articles (a, an, the), prepositions (in, of, at), conjunctions (and, or, nor), etc. from the documents.
(5) Word stemming – standardizes word's suffixes (e.g., labeling -- label).
(6) Word lemmatization – determines part of speech (POS) of a word, and applies different normalization rules for each POS.
(7) Feature selection – reduces the dimensionality of the data space by removing irrelevant or less relevant features. In our prototype, we choose information gain as a feature selection criterion.
(8) Dictionary construction – constructs a uniform dictionary, which is used as a reference for converting the text document to a vector of features. Each feature in the vector corresponds to a word in the dictionary.
(9) Feature weighting – assigns different weights to words in the dictionary. We use standard normalized term frequency-inverse document frequency (TFIDF) as the weighting function in our TC prototype system.

Yang [19] presented the procedure of KNN based algorithm to classify input document *X*. Assignation category to *X* document is based on KNN classifier, which ranks the document's neighbors among the training samples, and uses the *k* most similar class labeled neighbors. Let's assume that *N* is the number of training samples *(i = 1, 2,..., N)*, and *j* is the number of various training categories *(C_1, C_2,..., C_j)*. After pre-processing we get *m*-dimensional feature vector for each training sample.

1) The same *m*-dimensional text feature vector form *(X_1, X_2,..., X_j)* is assigned to document *X (j = 1, 2,..., m)*.

2) Similarity between training samples and document *X* is calculated. As an example we take *i*-th document d_i (d_{i1}, d_{i2},..., d_{im}) and calculate similarity between them *sim(X,d_i)*.

$$sim(X, d_i) = \frac{\sum_{j=1}^{m} X_j \cdot d_{ij}}{\sqrt{\left(\sum_{j=1}^{m} X_j\right)^2 \cdot \left(\sum_{j=1}^{m} d_{ij}\right)^2}} \tag{1}$$

3) Then choose *k* samples that are greater than *N* similarities of *sim(X,d_i)*, and treat them as collection of document *X*. We have to calculate the probability of *X* belong to each category respectively with following formula.

$$P(X, C_j) = \sum_{d_i \in KNN(X)} sim(X, d_i) \cdot \delta(d_i, C_j) \tag{2}$$

KNN(X) stands for a set of *k*-nearest neighbors of document *X*, and *$\delta(d_i,C_j)$* indicates a category attribute function for document d_i with respect to class C_j.

$$\delta(d_i, C_j) = \begin{cases} 1 & d_i \in C_j \\ 0 & d_i \notin C_j \end{cases} \tag{3}$$

4) Consequently, we assign document *X* the category which has the largest argument of *P(X,Cj)*.

$$\arg \max_{C_j} \left(P\left(X, C_j\right)\right) = \arg \max_{C_j} \left(\sum_{d_i \in KNN(X)} sim(X, d_i) \cdot \delta(d_i, C_j) \right) \tag{4}$$

3 Experiment and evaluation

We have conducted experiments based on collection of speeches of two well known American politicians; Barack Obama and Mitt Romney. The purpose was to classify category (in our case the author of the speech) of test samples (speeches) based on KNN algorithm.

To predict author of the speech we have to build two corpora; one for each candidate. All documents for training and testing require a pre-process step, which includes tasks of word removal, tokenization, word stemming, feature selection, and weighting. We apply given functions to clean both corpora. For example, transformation to upper and lowercases

is not significant at our analysis. We also use information gain as the feature of selection criterion and TFIDF as the weighting function.

Fig. 2 shows the simplified process, where allocation of category corresponds to given test sample. KNN classifier predicts category of the test sample according to the *k* training samples that are the nearest to the test sample. In order to estimate the statistical performance of the KNN algorithm, we use cross-validation.

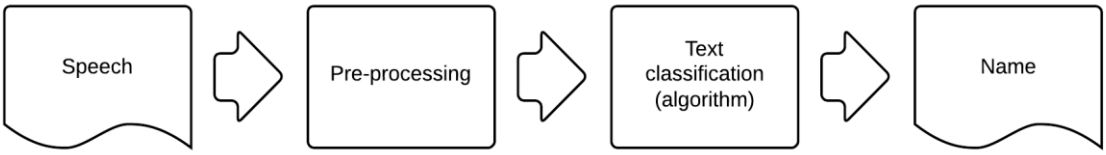


Figure 2: Creative predictive model.

When classifying textual information, it is necessary to evaluate results. Literature related to text and data mining most commonly defines three standard measures: *accuracy*, *precision*, and *recall*, to evaluate an algorithm's effectiveness on predicted category [1], [2], [9], [13], [17], [18], [21].

$$accuracy = \frac{true\ positive + true\ negative}{true\ positive + true\ negative + false\ positive + false\ negative} \quad (5)$$

$$precision = \frac{true\ positive}{true\ positive + false\ positive} \quad (6)$$

$$recall = \frac{true\ positive}{true\ positive + false\ negative} \quad (7)$$

3.1 Datasets for experiment

We use 161 textual documents; 104 speeches of Barak Obama written between 2002 and 2009, and 57 speeches of Mitt Romney written between 2004 and 2012. For research purposes, the textual data was collected and download from official politician's web sites [12], [14].

3.2 Evaluation

In our experiment, we use ten-fold cross validation method. The value of *k* in KNN algorithm includes all positive integers to value 10 *(k = 1, 2,..., 10)*. In order to evaluate the efficiency on predicted category of KNN algorithm, we calculated specified measures: *accuracy*, *precision*, and *recall*. Results of our analysis and evaluation are shown in Tab. 1.

Table 1: Experimental results and evaluation based on KNN algorithm for text categorization.

k	*Accuracy (%)*	*Class precision (%)*		*Class recall (%)*	
		Pred. Romney	*Pred. Obama*	*True Romney*	*True Obama*
1	76.00	62.86	85.71	77.19	75.00
2	71.00	55.21	93.85	92.98	58.65
3	**79.33**	**67.16**	**87.23**	**78.95**	**78.85**
4	74.00	58.33	89.61	85.96	66.35
5	**79.33**	**65.33**	**90.70**	**85.96**	**75.00**
6	76.67	60.71	92.21	89.47	68.27
7	78.50	64.38	88.64	82.46	75.00
8	71.33	55.06	88.89	85.96	61.54
9	74.67	59.49	87.80	82.46	69.23
10	73.00	56.98	89.33	85.96	64.42

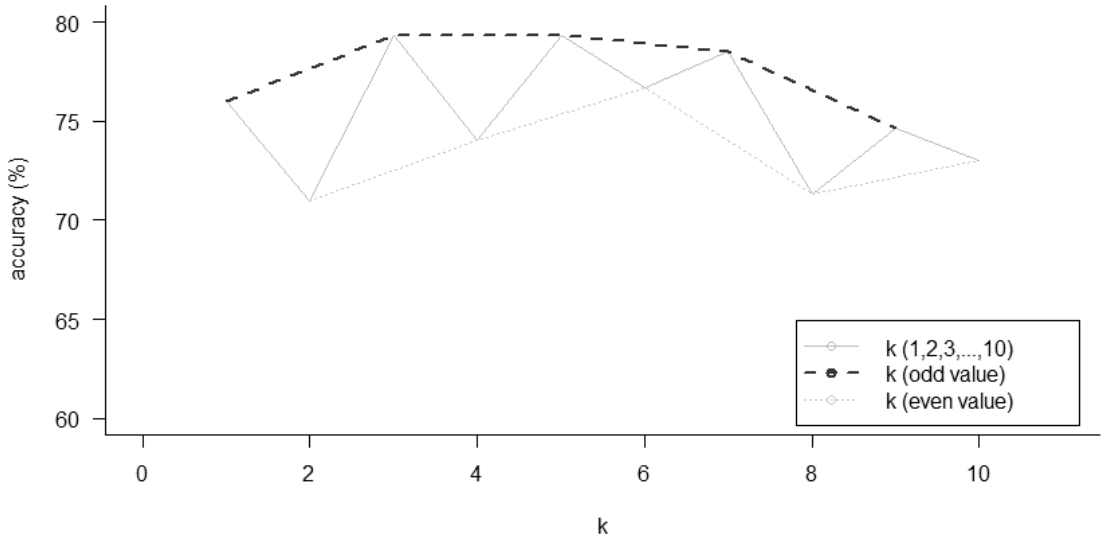


Figure 3: Accuracy of KNN classifier related to selection of k value.

As it is expected, odd values of *k* used in KNN classifier show better results if compared to even values (Fig. 3). When dealing with binary (two class) classification problems, it is helpful to choose odd number of *k* as this avoids tied votes. In general, the best result for our corpora is when *k* is equal 3 or 5, when accuracy is *79.33%* (Tab. 1). To conclude, the traditional KNN algorithm brings satisfactory results but far away from perfect. We assume that it can be outperformed by any of improved KNN or other text classification algorithms.

4 Conclusion

The amount of web content like customer feedback, competitor information, client emails, tweets, press releases, legal filings, product & engineering documents, etc., rapidly grow. In addition, humankind is still hungry of knowledge derived from retrieved information.

In this paper, we introduced one of the fastest, simplest, and most widely used methods for text categorization – KNN algorithm; present experimental application on collection of Barack Obama's and Mitt Romney's speeches, and evaluate obtained results. The results reported in this paper are satisfactory and not necessarily the best that can be achieved. Moreover, additional investigation and comparison of various classifiers via selected corpora is needed in order to evaluate performance of applied KNN algorithm.

Acknowledgments

Work supported by Creative Core FISNM-3330-13-500033 'Simulations' project funded by the European Union, The European Regional Development Fund. The operation is carried out within the framework of the Operational Programme for Strengthening Regional Development Potentials for the period 2007-2013, Development Priority 1: Competitiveness and research excellence, Priority Guideline 1.1: Improving the competitive skills and research excellence.

References

[1] Bishop, C. M., 2006. Pattern Recognition and Machine Learning. Springer, pp. 124-128.
[2] Guo, G., Wang, H., Bell, D., Bi, Y., Greer, K., 2006. Using KNN Model for Automatic Text Categorization. In Soft Computing, Vol.10, No.5, pp. 423-430.
[3] Han, E., Karypis, G., 2000. Centroid-Based Document Classification: Analysis and Experimental Results. In Proceedings of the 4th European Conference on Principles of Data Mining and Knowledge Discovery, pp. 424-431.
[4] IBM. Apply new analytics tools to reveal new opportunities. http://www.ibm.com/smarterplanet/us/en/business_analytics/article/it_business_intelligence.html
[5] Joachims, T., 2001. A Statistical Learning Model of Text Classification for Support Vector Machines. In Proceedings of SIGIR-01, 24th ACM International Conference on Research and Development in Information Retrieval, pp. 128-136.
[6] Kuang, Q., Zhao, L., 2009. A Practical GPU Based KNN Algorithm. In International Symposium on Computer Science and Computational Technology (ISCSCT), pp. 151-155.
[7] Lewis, D. D., 1998. Naïve (Bayes) at Forty: The Independent Assumption in Information Retrieval. In Proceedings of ECML-98, 10th European Conference on Machine Learning, pp. 4-15.
[8] Li, H., Yamanishi, K., 1999. Text Classification Using ESC-based Stochastic Decision Lists. In Proceedings of CIKM-99, 8th ACM International Conference on Information and Knowledge Management, pp. 122-130.
[9] Liu, B., 2006. Web Data Mining. Chapter Opinion Mining. Springer, pp. 459-526.
[10] Mitchell, T. M., 1996. Machine Learning. McGraw Hill, New York.
[11] Numerical Algorithms Group. K-nearest neighbor. http://www.nag-j.co.jp/nagdmc/knn.htm
[12] Obama speeches. http://obamaspeeches.com/
[13] Qi, X., Davison, B. D., 2009. Web Page Classification: Features and Algorithms. In ACM Computing Surveys (CSUR), Vol.41, No.2, pp. 1-12.
[14] Romney speeches. http://mittromneycentral.com/speeches.com/
[15] Ruiz, M. E., Srinivasan, P., 1999. Hierarchical Neural Networks for Text Categorization. In Proceedings of SIGIR-99, 22nd ACM International Information Retrieval, pp. 281-282.
[16] Sebastiani, F., 2002. Machine Learning in Automated Text Categorization. ACM Computing Surveys (CSUR), Vol.34, No.1, pp. 1-47.
[17] Sudha, L. R., Bhavani, R., 2012. Performance Comparison of SVM and KNN in Automatic Classification of Human Gait Patterns. Int. J. Comput, Vol.6, No.1, pp. 19-28.
[18] Witten, I. H., Eibe, F, Hall, M. A., 2013. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann Publishers, pp. 78-83.
[19] Yang, L., Dai, Q., Guo, Y., 2006. Study on KNN Text Categorization Algorithm. In Micro Computer Information, No.21, pp. 269-271.
[20] Yang, Y., Liu, X., 1999. A Re-examination of Text Categorization Methods. In Proceedings of SIGIR-99, 22nd ACM International Conference on Research and Development in Information Retrieval, pp. 42-49.
[21] Zhou, Y., Li, Y., Xia, S., 2009. An Improved KNN Text Classification Algorithm Based on Clustering. In Journal of Computers, Vol.4, No.3, pp. 230-237.

APPLICATION OF POLYNOMIAL APPROXIMATION HIERARCHY TO QUADRATIC ASSIGNMENT PROBLEM

Peter J.C. Dickinson
Department of Statistics and Operations Research, University of Vienna,
Oskar-Morgenstern-Platz 1, 1090 Wien, Austria, E-mail: peter.dickinson@univie.ac.at
Janez Povh
Faculty of Information Studies in Novo Mesto
Ulica talcev 3, 8000 Novo mesto, Slovenia, E-mail: janez.povh@fis.unm.si

Abstract: In the paper we demonstrate how to use a very general and powerful approximation hierarchy for general polynomial optimization problems to get strong and tractable lower bounds for the well-known Quadratic assignment problem. We show that the first members of this hierarchy give linear and semidefinite programming bounds comparable with the strongest bounds from the literature.
Keywords: real algebraic geometry; polynomial optimization; approximation hierarchy; Quadratic assignment problem
Mathematics Subject Classification (2010): 90C05; 90C25; 12D15; 14P10

1 INTRODUCTION

In *polynomials optimization* problems one wants to optimize an objective polynomial function over the feasible set defined by a set of polynomial equalities and inequalities (we call such a feasible set a *semialgebraic set*). Several NP-hard problems can be formulated in this form, e.g. testing matrix copositivity is equivalent to optimizing a homogeneous quadratic polynomial over the non-negative orthant (see [11, 4]); solving linear optimization problems with binary constraints [14]; solving nonconvex quadratic optimization problems (for example the Quadratic assignment problem, the Graph partitioning problem, the MAX-CUT problem) - see [9, 12]).

Polynomial optimization problems are in general very difficult, so it is a natural choice to look for tractable relaxations. These relaxations are typically obtained by convexification and simplification of the feasible set. De Klerk [3], Burer [2] and Eichfelder, Dickinson and Povh [8, 5] presented a way how to make the problem convex but the resulting convex sets are defined by set-semidefinite constraint which is difficult to verify (separation problems over such sets are still NP-hard).

Therefore simplification is needed. Most of the authors who approached NP-hard problems by set-semidefinite reformulations used approximation hierarchies based on moments and sums-of-squares to provide new lower or upper bounds for the optimal values of the original problems. Parrilo, de Klerk and Pasechnik [3] introduced two monotonic hierarchies of cones that approximate the cone of copositive matrices from the inside. One hierarchy consists of cones described by linear constraints and the other contains cones described by positive semidefinite constraints, see also [1] for alternative description of these cones.

Dickinson and Povh [6] represented new reformulation-approximation strategy based on a new Positivstellensatz from [7], which yield a hierarchy of linear or semidefinite programming problems with increasing lower bound for the original problem.

In this paper we simplify this construction and demonstrate it's contribution to the Quadratic assignment problem.

1.1 Contribution

The main contribution of this paper is translation of the very general polynomials approximation hierarchy from [6] to the well-known combinatorial optimization problem. We show that the Quadratic assignment problem satisfy the assumptions of the hierarchy and that the lower bounds implied by the hierarchy are at least as strong as the existing bounds from the literature.

1.2 Notation

Here is some notation that will be used in this paper. For the strictly positive integer m and the nonnegative integer t we define the following, where we shall exclude the m from the notation if it is equal to one:

$$\begin{aligned}
\mathbb{R}^m &: = \text{The set of real vectors of order } m;\\
\mathbb{R}^m_+ &:= \text{The set of non-negative real vectors of order } m;\\
\mathbb{R}^m_{++} &:= \text{The set of strictly positive real vectors of order } m;\\
\mathbb{N}^m &:= \text{The set of non-negative integer vectors of order } m;\\
\mathbb{N}^m_{=t} &:= \{\boldsymbol{\alpha} \in \mathbb{N}^m \mid \mathbf{e}^\mathsf{T}\boldsymbol{\alpha} = t\};\quad \mathbb{N}^m_{\leq t} := \{\boldsymbol{\alpha} \in \mathbb{N}^m \mid \mathbf{e}^\mathsf{T}\boldsymbol{\alpha} \leq t\};
\end{aligned}$$

where $\mathbf{e} \in \mathbb{R}^m$ is the all-ones vector. Also, for $i \in \{1, \ldots, m\}$, we define $\mathbf{e}_i \in \mathbb{R}^m$ to be the unit vector with i-th component equal to one and all other components equal to zero. For $\mathbf{e}_i$ and $\mathbf{e}$, the value of m will be apparent from the context. For $\mathbf{x} \in \mathbb{R}^m$ we refer to its i-th component via $(\mathbf{x})_i$. We define inner product of $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$ as $\langle \mathbf{x}, \mathbf{y} \rangle := \sum_{i=1}^m (\mathbf{x})_i (\mathbf{y})_i$ and consider the standard Euclidean norm $\|\mathbf{x}\|_2 = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$.

We note that $|\mathbb{N}^m_{=t}| = (m+t-1)!/\big(t!\ (m-1)!\big)$ and define the set

$$\mathbb{R}^{\mathbb{N}^m_{=t}} := \text{ The set of real vectors of order } |\mathbb{N}^m_{=t}|, \text{ indexed by elements in } \mathbb{N}^m_{=t},$$

The definitions of the inner product and Euclidean norm are then naturally extended for these spaces.

For $\mathbf{x} \in \mathbb{R}^n$, $\boldsymbol{\alpha} \in \mathbb{N}^n$ and $t \in \mathbb{N}$, we define $\mathbf{x}^{\boldsymbol{\alpha}} \in \mathbb{R}$ and $\mathbf{u}_t(\mathbf{x}) \in \mathbb{R}^{\mathbb{N}^n_{=t}}$ as follows (where $0^0 := 1$):

$$\mathbf{x}^{\boldsymbol{\alpha}} := \prod_{i=1}^n (\mathbf{x})_i^{(\boldsymbol{\alpha})_i}, \qquad \mathbf{u}_t(\mathbf{x}) := (\mathbf{x}^{\boldsymbol{\alpha}})_{\boldsymbol{\alpha} \in \mathbb{N}^n_{=t}}.$$

We let $\deg(f)$ denote the degree of a polynomial f and let $\mathbb{R}_t[\mathbf{x}]$ denote the set of *homogeneous* polynomials of degree t with real coefficients acting on $\mathbb{R}^n$. Note that for any $f \in \mathbb{R}_t[\mathbf{x}]$ there exists a unique $\mathbf{f} \in \mathbb{R}^{\mathbb{N}^n_{=t}}$ such that $f(\mathbf{x}) = \langle \mathbf{f}, \mathbf{u}_t(\mathbf{x}) \rangle = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^n_{=t}} (\mathbf{f})_{\boldsymbol{\alpha}} \mathbf{x}^{\boldsymbol{\alpha}}$. Using this fact, from now on we shall freely interchange between a function $f \in \mathbb{R}_t[\mathbf{x}]$ and a vector $\mathbf{f} \in \mathbb{R}^{\mathbb{N}^n_{=t}}$.

For a function $f \colon \mathbb{R}^n \to \mathbb{R}$ and a set $\mathcal{M} \subseteq \mathbb{R}$, we let $f^{-1}(\mathcal{M}) := \{\mathbf{x} \in \mathbb{R}^n \mid f(\mathbf{x}) \in \mathcal{M}\}$. For $\alpha \in \mathbb{R}$, we shall write $f^{-1}(\alpha)$ instead of $f^{-1}(\{\alpha\})$.

2 APPROXIMATION HIERARCHY FOR POLYNOMIAL OPTIMIZATION PROBLEMS

Dickinson and Povh [6] considered the following general polynomial optimization problem

$$\begin{array}{lll} \min\limits_{\mathbf{x}} & g_1(\mathbf{x}) & \\ \text{s.t.} & f(\mathbf{x}) \geq 0 & \forall f \in \mathcal{F} \\ & g_2(\mathbf{x}) = 1 & \\ & \mathbf{x} \in \mathbb{R}^n_+, & \end{array} \tag{P1}$$

under the following assumptions:

Assumption 2.1. $\mathcal{F} \cup \{g_2\} \subseteq \bigcup_{i \in \mathbb{N}_{\geq 1}} \mathbb{R}_i[\mathbf{x}]$ *and* $g_1 \in \{0\} \cup \bigcup_{i \in \mathbb{N}} \mathbb{R}_i[\mathbf{x}]$, *i.e. we assume that all polynomials from* (P1) *are homogeneous.*

Assumption 2.2. *Either* $g_1 = 0$ *or* $\deg(g_1) = \deg(g_2)$, *and we let* $d = \deg(g_2)$.

Assumption 2.3. $g_2(\mathbf{x}) \geq 0$ *for all* $\mathbf{x} \in \mathbb{R}^n_+ \cap \bigcap_{f \in \mathcal{F}} f^{-1}(\mathbb{R}_+)$.

Assumption 2.4. $g_1(\mathbf{x}) > 0$ *for all* $\mathbf{x} \in \mathbb{R}^n_+ \cap g_2^{-1}(0) \cap \bigcap_{f \in \mathcal{F}} f^{-1}(\mathbb{R}_+) \setminus \{\mathbf{0}\}$.

Mote that Assumption 2.3 can simply be enforced by adding g_2 to the set $\mathcal{F}$. This would currently be a redundant inequality in the problem, however it will come in useful later on. We also point out that the original result has not restricted $\mathcal{F}$ to be a finite set.

Problem (P1) can be reformulated and relaxed into the following primal dual pair of convex conic optimization problems with linear objective function.

Based on Positivstellensatz from [7] Dickinson and Povh suggested the following hierarchy of primal and dual linear programming problems which are relaxation of the pair (P10) and (**??**) and have optimal values converging to the optimal value of the original problem (P1):

$$\begin{array}{lll} \min\limits_{\mathbf{y},\mathbf{z}} & \langle \mathbf{g}_{1,r}, \mathbf{y} \rangle & \\ \text{s.t.} & \langle \mathbf{g}_{2,r}, \mathbf{y} \rangle = 1 & \\ & (\mathbf{z_m})_{\mathbf{p}} = (\mathbf{y})_{\mathbf{m+p}} & \text{for all } \mathbf{m} \in \mathbb{N}^n_{\leq r+d},\ \mathbf{p} \in \mathbb{N}^n_{=r+d-\mathbf{e}^\mathsf{T}\mathbf{m}} \\ & \mathbf{z_m} \in \mathcal{Y}_{r+d-\mathbf{e}^\mathsf{T}\mathbf{m}} & \text{for all } \mathbf{m} \in \mathbb{N}^n_{\leq r+d} \\ & \mathbf{y} \in \mathbb{R}^{\mathbb{N}^n_{=r+d}} & \end{array} \tag{P2$_r$}$$

$$\begin{array}{lll} \max\limits_{\lambda,\mathbf{f}} & \lambda & \\ \text{s.t.} & (\mathbf{g}_{1,r})_{\mathbf{q}} - \lambda(\mathbf{g}_{2,r})_{\mathbf{q}} = \sum\limits_{\substack{\mathbf{p} \in \mathbb{N}^n_{\leq r+d}: \\ (\mathbf{p})_i \leq (\mathbf{q})_i\ \forall i}} (\mathbf{f_p})_{\mathbf{q-p}} & \text{for all } \mathbf{q} \in \mathbb{N}^n_{=r+d} \\ & \mathbf{f_p} \in \mathcal{Y}^*_{r+d-\mathbf{e}^\mathsf{T}\mathbf{p}} & \text{for all } \mathbf{p} \in \mathbb{N}^n_{\leq r+d} \end{array} \tag{D2$_r$}$$

The $\mathbf{z_m}$ variables can easily be removed from (P2$_r$) and are only there to simplify the notation.

Cones $\mathcal{K}$ and $\mathcal{K}^*$ are defined by

$$\begin{aligned}
\mathcal{K} &:= \{\mathbf{f} \in \mathbb{R}^{\mathbb{N}^n_{=d}} \mid f(\mathbf{x}) \geq 0 \text{ for all } \mathbf{x} \in \mathcal{Y}\}. && (3)\\
\mathcal{K}^* &:= \text{conv}\{\mathbf{u}_d(\mathbf{x}) \mid \mathbf{x} \in \mathcal{Y}\}, \quad \text{where} && (4)\\
\mathcal{Y} &:= \{\mathbf{x} \in \mathbb{R}^n_+ \mid \mathbf{u}_i(\mathbf{x}) \in \mathcal{Y}_i \text{ for all } i \in \mathbb{N}\}, && (5)
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{Y}_i &:= \{\mathbf{y} \in \mathbb{R}^{\mathbb{N}^n_{=i}} \mid \langle \mathbf{f}, \mathbf{y}\rangle \geq 0 \text{ for all } f \in \mathcal{F} \text{ such that } \deg(f) = i\} && (6)\\
\mathcal{Y}^*_i &:= \text{cl cone}\{\mathbf{f} \in \mathbb{R}^{\mathbb{N}^n_{=i}} \mid f \in \mathcal{F},\ \deg(f) = i\}, && (7)
\end{aligned}$$

for $i = 1, \ldots$ and $\mathcal{Y}_0 = \mathcal{Y}^*_0 = \mathbb{R}_+$.

3 THE QUADRATIC ASSIGNMENT PROBLEM

The quadratic assignment problem (QAP) is a standard problem in location theory and is very famous because of its hardness. One of classical formulations is the following [10]:

$$(QAP) \qquad OPT_{QAP} = \min\{\langle X,\, AXB + C\rangle \colon X \text{ a permutation matrix}\}.$$

QAP is known to be very hard problem from a theoretical and practical point of view. We suggest the reader to read the comprehensive survey with results up to 2007 [10].

We can reformulate QAP into the following problem (see [13] and references therein):

$$OPT_{QAP} = \min\{\langle L,\, Y\rangle \colon Y = \begin{bmatrix}1\\ x\end{bmatrix} \cdot \begin{bmatrix}1\\ x\end{bmatrix}^T,\ \ x = \text{vec}(X),\ XX^T = I,\ X \in \mathbb{R}^{n\times n}_+\} \quad (8)$$

where $\text{vec}(X)$ is a column vector obtained from matrix X columnwise.

$$L = \begin{bmatrix} 0 & \frac{1}{2}c^T \\ \frac{1}{2}c & B \otimes A \end{bmatrix}.$$

We can add to the formulation (8) the following initially redundant constraints:

$$X\mathbf{e} = \mathbf{e}, \quad X^T\mathbf{e} = \mathbf{e}, \quad XX^T = I. \qquad (9)$$

QAP with (9) satisfy Assumptions 2.1 to 2.4. Indeed, if we write it as follows

$$\begin{aligned}
\min_{(x_0,X)\in\mathbb{R}_+\times\mathbb{R}^{n\times n}_+} \quad & \langle \hat{L}, Y\rangle \\
\text{s.t.} \quad & Y = \begin{bmatrix}x_0\\ x\end{bmatrix} \cdot \begin{bmatrix}x_0\\ x\end{bmatrix}^T \\
& X^TX - x_0^2 I = 0, \quad XX^T - x_0^2 I = 0 \\
& X\mathbf{e} - x_0\mathbf{e} = 0, \quad X^T\mathbf{e} - x_0\mathbf{e} = 0 \\
& x_0^2 = 1
\end{aligned} \qquad (P9)$$

then all polynomials are homogeneous and the objective function and the last constraint are of the same degree. Only Assumption 2.4 may not be satisfied. In this case we can

always add to L a matrix of all ones multiplied with sufficiently large number, since the sum of all entries in any feasible Y is always $(n+1)^2$. The resulting L is denoted by $\hat{L}$.

To consider hierarchies $\{(\mathrm{P2}_r), (\mathrm{D2}_r),\ r = 0, 1, \ldots\}$ we point out that $Y0 = \mathbb{R}_+$, $\mathcal{Y}_k = \mathbb{R}^{\mathbb{N}^{n^2+1}_{=k}}$ for $k \geq 3$, and

$$\mathcal{Y}_1 = \{\mathbf{y} \in \mathbb{R}^{\mathbb{N}^{n^2+1}_{=1}} \mid \sum_{i=1}^{n} (\mathbf{y})_{(k-1)n+i+1} - (\mathbf{y})_1 = \sum_{i=1}^{n} (\mathbf{y})_{(i-1)n+k+1} - (\mathbf{y})_1 = 0, \quad k = 1, \ldots, n\}$$

$$\mathcal{Y}_2 = \{\mathbf{y} \in \mathbb{R}^{\mathbb{N}^{n^2+1}_{=2}} \mid \sum_{i=1}^{n} Y^{ii} - (\mathbf{y})_1^2 I = 0, \ \ \mathrm{trace}(Y^{ij}) - (\mathbf{y})_1^2 \delta_{ij} = 0\}$$

For $\mathbf{y} \in \mathcal{Y}_2$ we used that every such vector can be represented by symmetric matrix Y with rows and columns labeled by elements from $\mathbb{N}^{n^2+1}_{=1}$ such that $Y_{\mathbf{p},\mathbf{q}} = (\mathbf{y})_{\mathbf{p}+\mathbf{q}}$. We also used that such matrix Y can be represented with the following block structure:

$$Y = \left[\begin{array}{c|ccc} Y^{00} & Y^{01} & \ldots & Y^{0n} \\ \hline Y^{10} & Y^{11} & \ldots & Y^{1n} \\ \vdots & \vdots & \ddots & \vdots \\ Y^{n0} & Y^{n1} & \ldots & Y^{nn} \end{array}\right] \tag{10}$$

where 0-th row corresponds to $(1, 0, 0 \ldots, 0) \in \mathbb{N}^{n^2+1}_{=1}$ and i-th block of rows refers to labels (vectors) $\mathbf{p} \in \mathbb{N}^{n^2+1}_{=1}$ having 1 on positions $(i-1)n+2, \ldots, in+1$, for $i = 1, \ldots, n$. Therefore for $r = 0$ the relaxation $(\mathrm{P2}_r)$ becomes

$$\begin{array}{ll} \min & \langle \hat{L}, Y \rangle \\ \text{s.t.} & \mathbf{y} \in \mathbb{R}_+^{\mathbb{N}^{n^2+1}_{=2}}, \quad Y_{\mathbf{p},\mathbf{q}} = (\mathbf{y})_{\mathbf{p}+\mathbf{q}}, \ \forall \mathbf{p}, \mathbf{q} \in \mathbb{N}^{n^2+1}_{=1} \\ & Y_{:,\mathbf{p}} \in \mathcal{Y}_1, \ \forall \mathbf{p} \in \mathbb{N}^{n^2+1}_{=1}, \quad Y^{00} = 1 \\ & \sum_{i=1}^{n} Y^{ii} - Y^{00} I = 0, \ \ \mathrm{trace}(Y^{ij}) - Y^{00} \delta_{ij} = 0 \end{array} \tag{P10}$$

This is a linear programming relaxation comparable with existing linear programming bounds (see [10]). We can strengthen it by adding a natural constraint that Y is positive semidefinite. In this case we get a bound which is equivalent to sthe strongest known semidefinite programming bound from the literature (more precisely: to bound $QAP_{\mathcal{K}:n^{0*}}$ from [13].

If we go further to $r = 1$ we obtain stronger bound but it complexity is very big and we are currently searching for a way to simplify this bound and reduce the complexity.

4 Conclusions

In this paper we showed how to use the new polynomial approximation hierarchy from [6] to get linear and semidefinite programming lower bound for the Quadratic assignment problem. This approach naturally yields lower bounds, comparable with the strongest existing lower bounds from the literature. We could go further and present bounds based on later members from the hierarchy but these bounds turn out to be too complex for

reasonable applications, therefore we will first try to simplify them using the problem specific structure. This is the main task for the ongoing research.

Acknowledgements

The second author wishes to thank to Slovenian research agency for support via program P1-0383 and project L74119 and to Creative Core FISNM-3330-13-500033 'Simulations' project funded by the European Union.

References

[1] I. M. Bomze and E. de Klerk. Solving standard quadratic optimization problems via linear, semidefinite and copositive programming. *Journal of Global Optimization*, 24(2):163–185, 2002.

[2] S. Burer. *Copositive programming*, volume 166 of *International Series in Operations Research & Management Science*, pages 201–218. Springer US, 2012.

[3] E. de Klerk and D. V. Pasechnik. Approximation of the stability number of a graph via copositive programming. *SIAM Journal on Optimization*, 12(4):875–892, 2002.

[4] P. J. C. Dickinson. *The Copositive Cone, the Completely Positive Cone and their Generalisations*. PhD thesis, University of Groningen, Groningen, The Netherlands, 2013.

[5] P. J. C. Dickinson, G. Eichfelder, and J. Povh. Erratum to the paper "On the set-semidefinite representation of nonconvex quadratic programs over arbitrary feasible sets" [Optim. Letters, 2012]. *Optimization Letters*, pages 1–11, 2013.

[6] P. J. C. Dickinson and J. Povh. A new tractable approximation hierarchy for general polynomial optimization problems. *Preprint, submitted. Available at* `http://www.optimization-online.org/DB_HTML/2013/06/3925.html`, 2013.

[7] P. J. C. Dickinson and J. Povh. On a generalization of Pólya's and Putinar-Vasilescu's positivstellensätze. *Preprint, submitted. Available at http://www.optimization-online.org/DB_HTML/2013/05/3879.html*, 2013.

[8] G. Eichfelder and J. Povh. On the set-semidefinite representation of nonconvex quadratic programs over arbitrary feasible sets. *Optimization Letters*, pages 1–16, 2013.

[9] M. Laurent and F. Rendl. Semidefinite programming and integer programming. In G. N. K. Aardal and R. Weismantel, editors, *Discrete Optimization*, volume 12 of *Handbooks in Operations Research and Management Science*, pages 393 – 514. Elsevier, 2005.

[10] E. M. Loiola, N. M. Maia de Abreu, P. O. Boaventura-Netto, P. Hahn, and T. Querido. A survey for the quadratic assignment problem. *European J. Oper. Res.*, 176(2):657–690, 2007.

[11] K. G. Murty and S. N. Kabadi. Some NP-complete problems in quadratic and nonlinear programming. *Mathematical Programming*, 39(2):117–129, 1987.

[12] J. Povh and F. Rendl. A copositive programming approach to graph partitioning. *SIAM Journal on Optimization*, 18(1):223–241, 2007.

[13] J. Povh and F. Rendl. Copositive and semidefinite relaxations of the quadratic assignment problem. *Discrete Optim.*, 6(3):231–241, 2009.

[14] L. Wolsey. *Integer Programming*. A Wiley-Interscience publication. Wiley, 1998.

AGENT BASED VULNERABILITY DISCOVERY MODEL

Andrej Dobrovoljc
University of Ljubljana, Faculty of Computer and Information Science
Tržaška cesta 25, SI-1000 Ljubljana, Slovenia
{andrej.dobrovoljc}@fis.unm.si

Abstract: Risk assessment of information systems largely depends on software vulnerabilities and interests of individuals for their detection. The question that we address is how to predict future vulnerability discoveries. Among important factors we can consider various changes in the user environment. Existing vulnerability discovery models (VDM) take them into account only partly. We have developed an agent based simulation model which considers additional relevant factors from the user environment that affect the discovery process.

Keywords: software vulnerability, vulnerability discovery model, risk assessment.

1 INTRODUCTION

When the software producers launch new IT solutions on the market (e.g. World Wide Web) they hope at first to reach the projected growth in the number of users. Nevertheless, they must also be aware of the risks. Namely, after a while the product can attract also individuals, who want to compromise it. The IT solution may represent the means of achieving their special goals. Their first step is to identify vulnerabilities and then to exploit them.

Every software product is vulnerable. Individuals with the sufficiently high level of technical expertise and ability to innovate constantly discover new vulnerabilities, what proves that there are many undiscovered ones. The associated risks should be eliminated as quickly as possible. For this purpose, it is necessary to assure appropriate level of resources to sanitize the software holes. It would be ideal, if we could accurately anticipate, when and how many vulnerabilities will be found in our product in the future.

Hackers (black and white ones) are more interested to discover them on the products, with the highest potential benefit for them. Generally, these are the most popular products on the market, what we measure with the number of their users [1].

Publicly available data on discovered vulnerabilities show that the successful ideas quickly diffuse within the hackers community and that we can find similar discoveries on many products. We can talk about the innovation diffusion phenomena. In this context, it is also important to mention, that one of the main sources from which the hackers learn, is patched program code [2, 3].

When the knowledge about the vulnerability and its potential exploit are publicly disclosed without the needed patches, it is extremely dangerous for the information system. There is simply no protection against such attacks. In order to mitigate the risk, the culture of responsible disclosure evolved. Individuals with positive intentions (ethical hackers) provide sensitive information on their discoveries at first to the software authors. By doing this, they give the authors a reasonable period of time (e.g. a week or a month) in order to eliminate product defects. Only then they disclose it to the public. The problem is that this time period can significantly vary and can be very short [4].

Some companies have taken a step further. They decided to buy knowledge about discovered vulnerabilities. By running so called *Bug Bounty Programs* they try to motivate ethical hackers to test and consequently help improve the safety of their products. Recently some new types of services appeared: *Bug Bounty as a service* and *Bug Bounty as a platform*. Both of them are types of outsourced bounty programs.

The risks are associated with various factors and not only with the number of users. The question that we address is how to predict future vulnerability discoveries taking into

account all these known factors. The answer to this question can help us assess the future risk and consequently to decide for most suitable measures for their mitigation.

In the next section is an overview of the work related to our research. It is followed by the section "Method", where we present our findings from publicly available data and the concept of the simulation model. We summarize our results in the "Results and Discussion" section. Conclusion and an overview of the future work are given in the last section.

2 RELATED WORK

Anticipating software vulnerability discoveries is not a completely new challenge. It became important in the last decade with the development of web applications and technologies. Vulnerability Discovery Model (VDM) shall predict the time and the frequency of future discoveries. This information allows the software producers to acquire needed resources in order to fix defects on time.

The basic idea for VDM models came from Software Reliability Models (SRM). They are used for discovering software bugs. It turned out that the nature of the software vulnerabilities is different from bugs. That is why SRM models are not suitable for detecting vulnerabilities [8]. There are several definitions of software vulnerabilities. In order to make clear distinction to the software bugs, we use the following definition in our research: *"Software vulnerability is an instance of a mistake in the specification, development, or configuration of software such that its execution can violate the security policy"* [8].

Several VDM models have been proposed in last years (Fig. 1). We separate them into two categories, *Time-based* and *Effort-based* models, according to their prediction approach. Effort-based models are difficult to realize due to the lack of data. Namely, the number of the product users is constantly changing and there are no accurate records about it [10].

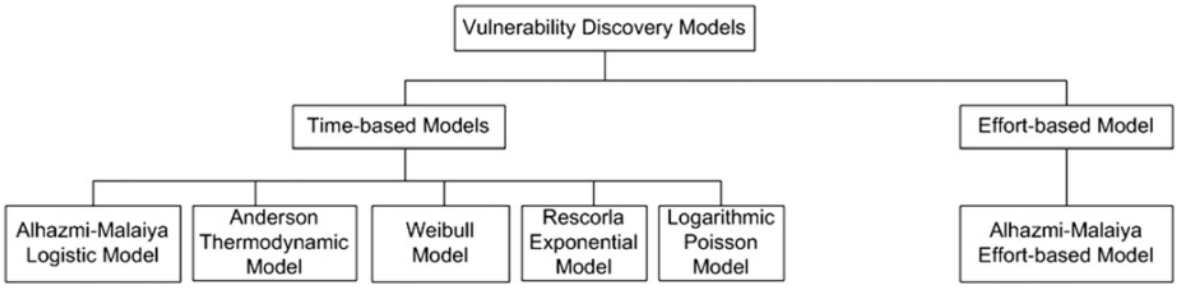


Figure 1: Existing Vulnerability Discovery Models (VDM) [10]

Among *Time-based* models the Alhazmi-Malaiya Logistic model (AML) is the most accurate one [2]. It is based on a logistic function and on a quite simple assumption. In an early phase, when the product enters the market, it has a few users. Simultaneously with increased popularity grows also the number of users. At the end of its life cycle their number declines. According to the findings of the authors vulnerability discoveries follow the same logistic function. They have proposed the metrics *Vulnerability density* (1), which is used to estimate the number of expected vulnerabilities in the observed product [1, 7].

$$V_{KD} = \frac{KnownVulNerabilities}{SourceLinesOfCode} \quad (1)$$

Proposed models only partly consider the factors that affect the discovery process. Besides the number of product users, we have to take into account also the vulnerability management processes of software producers and the learning process of hackers.

3 METHOD

The main goal of our research was to develop a VDM model to predict the vulnerability detection in software, taking into account some other factors:

- open innovation in the hackers learning process,
- innovation diffusion in the vulnerability discovery process,
- the influence of different vulnerability management strategies,
- the impact of software producers patching process on the disclosure process.

3.1 Vulnerability data

Existing VDM models were verified on the data from publicly available databases. There are two similar initiatives: CVE (Common Vulnerabilities and Exposures) database organized by MITRE and NVD (National Vulnerability Database) maintained by NIST. In both databases we can find only the dates of publication of vulnerabilities. The discovery dates are not available. Before introduction of responsible disclosure culture and the Bug Bounty programs these two days used to be the same. Publication date is now the time of vulnerability disclosure and regularly comes with some delay. Delay depends on (ethical) hackers patience for responsible disclosure or efficiency of producers patching process. Apparently, if we are strict in the definition of these models, we have to name them *Vulnerability Disclosure Models*.

In our study, we verified two different types of vulnerability management strategies at the producers side. The example of the product, which depends only on the responsible disclosure culture, is Apache HTTP server. Among well-known software producers, which run bounty programs (i.e. they buy vulnerability knowledge), we found Mozilla, Google, Facebook and PayPal. In this category we have chosen FireFox browser, because it has been held under this policy since 2004. Others started their programs in 2010 or later. Both products, Apache and FireFox, have wide bases of users and have been present on the market for a decade or more. These facts ensures suitable sample for analysis.

VDM models are presented as cumulative number of vulnerability discoveries. Empirical data for FireFox show linear growth, while on Apache server we can recognise multiple logistic functions (Fig. 2). Our goal is to find the reasons for the differences.

In the year 2006, it seemed as the discoveries on Apache had reached the saturation phase and no more vulnerabilities would be expected. Additional years brought new discoveries. Detailed analyses showed that another wave on the curve represents the bunch of similar vulnerabilities. More than 60% of them in the period from the middle of 2007 to the middle 2008 were XSS vulnerabilities. This type of vulnerability is very rare on this product otherwise. We hypothesize that it is the result of an innovative idea, born on some product within hackers society and applied on Apache server.

3.2 Conceptual model

Agent-based modelling and simulation (ABMS) of human systems proved to be useful tool in practice [6]. Therefore, we developed a VDM based on ABMS. We used Repast Simphony platform. In model validation phase we used publicly available data from NVD database.

Vulnerability discovery area is firmly connected with the human factor. Detailed description and the behaviour of the key agents involved in this ecosystem is given by [4]. Schneier and Miller [5,9] describe how big interest exists for such knowledge, how respected it is and how trading takes place in this society. The learning process in the global hacker

society is similar to the global research community, which was studied by [11]. Their results show that the dealing with the same things is a strong positive factor for open innovation.

We took into consideration all these findings and built the agent model with the following features and mechanisms:

Simulation environment: Simulation environment is a network where hackers $H=\{h_1,h_2,...h_n\}$ link to the products i.e. their producers $P=\{p_1,p_2,...p_m\}$ according to their market share $0 \leq ms \leq 100$. Hacker randomly $(s \leftarrow rand(100))$ selects the product among those with $ms > s$. Every hacker is always associated with exactly one product. Each product has its initial set of vulnerabilities $V=\{1,2,...v_k\}$, where v_k follows the equation (1), and the set of code patches with one randomly selected starting element $C=\{rand(v_k)\}$. Set of code patches represents sanitized vulnerabilities.

Responsible and public disclosure: Knowledge about vulnerabilities is very sensitive asset and represents the basis for interactions between hackers and vendors. When ethical hacker silently warns the producer about vulnerability, we speak about *responsible disclosure*. Let the set of all such vulnerabilities be K_{Resp}. The opposite of this is *public disclosure* of knowledge. This is the most critical knowledge, because the suitable patches are not available. Let the K_{Pub} is the set of all publicly disclosed unpatched vulnerabilities. Then we can define the knowledge of hacker as $K_H = K_{Pub} \cup C$ and the knowledge of producer as $K_P = K_{Pub} \cup K_{Resp} \cup C$.

Learning mechanism: In order to discover vulnerability, hacker needs an appropriate level of skills (initially $skill \leftarrow 0$). Skills can be achieved by studying patches of code. With each successful review (*learn*) of patched code, hacker increases his skills (Alg. 1).

Algorithm 1: Learning from existing patches of code

$learn \leftarrow rand(v_k), v_k \in V$
if $learn \in C$ **then**
  $skill \leftarrow skill + 1$
end if

Discovery mechanism: Hacker with the appropriate level of skills (*Slevel*) is able to discover vulnerability. He has one attempt to find it within selected product. If randomly selected vulnerability is still available, he discovers it. He always spends all his skills (Alg. 2). With successful discovery, he increases his innovation potential (initially $I \leftarrow \{\}$), which is calculated by the number of different products, where he discovered the vulnerability. If the hacker is not successful for a long time (*timeTicksInactive* > *agentPatience*), he can change the product. On transitions, he retains his innovation abilities. Learning and discovery mechanisms together represent the open innovation concept, because discoveries happen as a result of aggregated knowledge and skills.

Innovation mechanism: Hacker with an appropriate ability to innovate ($|I| \geq Ilevel > 1$) is able to discover completely new type of vulnerability that has not been known previously on this product (innovation diffusion from other product). As a result of innovation, the number of available vulnerabilities in the product is increased $(V \leftarrow V \cup \{v_{k+1}, ... v_l\})$. Hacker can use its ability to innovate when he has not been successful for a long time (*timeTicksInactive* > *agentPatience*).

Knowledge purchasing: The model provides the choice of two different vulnerability management strategies. If the producer buys the knowledge about vulnerabilities, hackers withdraw from public disclosures after bargain. If the producer does not buy knowledge, hackers publicly disclose knowledge after their patience period $(K_{Pub} \leftarrow K_{Pub} \cup \{try\})$.

Algorithm 2: Vulnerability discovery

```
if skill ≥ Slevel then
    try ← rand(v_k), v_k ∈ V
    if try ∈ {V \ K_P} then
        K_Resp ← K_Resp ∪ {try}
        I ← I ∪ {p_i}, p_i ∈ P
    end if
    skill ← 0
end if
```

Patching process: Software producer removes all known vulnerabilities ($K_{Resp} \cup K_{Pub}$) in a queue according to priorities. The first priority are publicly disclosed vulnerabilities K_{Pub}, because there is no protection against them. Patching process time *patchTime* is fixed for all vulnerabilities. After patching the vulnerability v_s is removed from the knowledge sets ($K_{Pub} \leftarrow K_{Pub} - v_s$ or $K_{Resp} \leftarrow K_{Resp} - v_s$) and published as patch ($C \leftarrow C \cup v_s$).

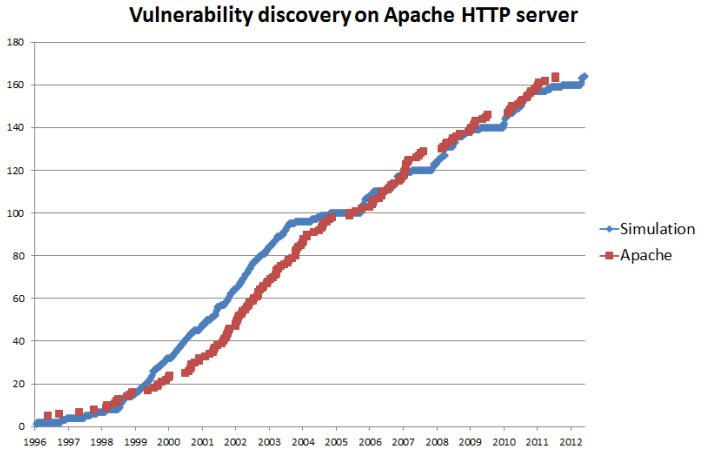


Figure 2: Simulation results compared to empirical data (cumulative number of publicly known vuln. K_H).

3.3 Validation

The simulation model presents the cumulative number of disclosed vulnerabilities (Fig. 2). Expected number of vulnerabilities in the product is determined by using vulnerability density metrics (1). The number of products, their market shares and strategy are empirical data. Some parameters (*VendorResponseTime*, *PatchingTime*, *HackersPatienceTime*) are based on discovery process description in [4] and were determined through the model calibration phase. The same was with the number of hackers, because we do not know exactly how many hackers participate in this society. Remaining parameters are required skills for hackers to discover, the needed innovation abilities to innovate and the number of vulnerabilities, born as a consequence of innovation. We gain their values through model calibration (Chrome browser was used for calibration purpose).

4 RESULTS AND DISCUSSION

We used goodness of fit analyses using chi-square (2) with $\alpha = 5\%$ where s_i are simulated and d_i empirical data. We compare data on quarterly bases. The null hypothesis is that the model fits the data, which is in case of $\chi^2 \leq \chi^2_{\text{critical}}$. Otherwise it will be rejected as a bad fit.

$$\chi^2 = \sum_{i=1}^{n} \frac{(d_i - s_i)^2}{s_i} \quad (2)$$

In case of Firefox, we performed a simulation using the buying knowledge strategy. Results fit well with the empirical data (χ^2 = 53.67 < $\chi^2_{critical}$ (5%) = 70.99, Degrees Of Freedom = 53). The simulation model produces the linear function, what proves that the producer controls the knowledge disclosure by its buying and patching processes and consequently can better control the risks.

The results for the Apache HTTP server with responsible disclosure strategy does not confirm the goodness of fit. The model needs further calibration. We confirmed the assumption that innovations give births to new vulnerabilities. They allow new discoveries, which manifest as additional logistic functions above the primary one (Fig. 2).

5 CONCLUSION AND FURTHER WORK

With the presented simulation model, we examined the impact of various factors on the discovery and disclosure of software vulnerabilities. We have shown that in this community we can recognize the concept of open innovation. In this society hackers participate with their discoveries and software providers with patches of code. All this knowledge is further used by agents on both sides.

Our further steps will be directed to the model calibration for responsible disclosure strategy and its validation. We need more empirical data in order to improve the accuracy of the model. Improved model will be used for proactive risk assessment.

References

[1] Alhazmi, O., Malaiya, Y., Ray, I.: Measuring, analyzing and predicting security vulnerabilities in software systems. Computers & Security 26(3), 219-228 (2007)

[2] Alhazmi, O., Malaiya, Y.: Measuring and Enhancing Prediction Capabilities of Vulnerability Discovery Models for Apache and IIS HTTP Servers. 2006 17th International Symposium on Software Reliability Engineering pp. 343-352 (2006)

[3] Arora, A., Telang, R.: Economics of software vulnerability disclosure. IEEE Security and Privacy Magazine 3(1), 20-25 (Jan 2005)

[4] Frei, S., Schatzmann, D., Plattner, B., Trammell, B.: Modelling the Security Ecosystem - The Dynamics of (In)Security, Workshop on the Economics of Information Security (WEIS) (2009)

[5] Miller, C.: The Legitimate Vulnerability Market Inside the Secretive World of 0-day Exploit Sales. In Sixth Workshop on the Economics of Information Security, pp. 1-10 (2007)

[6] North, M.J., Macal, C.M.: Managing Business Complexity: Discovering Strategic Solutions with Agent-Based Modeling and Simulation. Oxford University Press, Inc., New York, USA (2007)

[7] Ozment, A.: Improving Vulnerability Discovery Models Problems with Definitions and Assumptions Categories and Subject Descriptors (2007)

[8] Ozment, A.: Vulnerability discovery & software security (2007)

[9] Schneier, B.: The Vulnerabilities Market and the Future of Security. Forbes (2012)

[10] Woo, S.W., Joh, H., Alhazmi, O.H., Malaiya, Y.K.: Modeling vulnerability discovery process in Apache and IIS HTTP servers. Computers & Security 30(1), 50-62 (2011)

[11] Zou, G., Yilmaz, L.: Exploratory simulation of collective innovative behavior in global participatory science communities. Simulation Conference (WSC), Proceed. pp. 708-719 (2010)

MANAGEMENT OF BUSINESS PROCESSES IN HIGHLY DYNAMIC AND LOW-STRUCTURED SCENARIOS

dr. Grzegorz Majewski, prof. dr. Nadja Damij
Institute for process management
Faculty of Information Studies
Ulica talcev 3, Novo mesto
SLOVENIA
g.majewski@fis.unm.si, nadja.damij@fis.unm.si

Abstract: This paper presents a literature review and discusses various contemporary approaches to business process management in highly dynamic and low-structured scenarios. This paper introduces and describes the types of business processes that occur in low-structured and highly dynamic environments. Next section reviews most recent approaches existing in the business process management community. Following is a section on Process-Aware Information Systems (PAIS) from the perspective of how PAIS support business process management in highly dynamic and low-structured scenarios. The significant contribution of this paper is the combination of investigation into how business processes can be simulated and/or modeled in highly dynamic and low-structured scenarios. There is a comprehensive literature review provided on this subject paired with some information from the business environment. This work is supported by Creative Core FISNM-3330-13-500033 'Simulations' project funded by the European Union, The European Regional Development Fund. The operation is carried out within the framework of the Operational Programme for Strengthening Regional Development Potentials for the period 2007-2013, Development Priority 1: Competitiveness and research excellence, Priority Guideline 1.1: Improving the competitive skills and research excellence.

Key-Words: low-structured business processes, business process simulation, business process modeling, dynamic processes, ad hoc processes.

1 INTRODUCTION

In the modern economy Process-Aware Information Systems (PAISs) are broadly used to model and simulate all human activities and tasks. These range from established, well – structured, classical ones (e.g. management of the supply chain) to very dynamic, less - structured ones (e.g. emergency and crisis management, change management, Research and Development).

Every aspect of business processes, from smaller or greater degree, requires a certain amount of knowledge. There are a number of factors that influence both the degree as well as the amount and character of the knowledge required. These factors range from the background and experience of process stakeholders to the complexity of the problem domain.

Some business processes may be characterized as less structured than others. This is usually the case of knowledge-intensive business processes. Apart from that such processes are often happening in a highly dynamic environment. Some researchers introduce the concept of project types (PTs) with regards to BPM [2]. PT is a key concept, which describes development situations in situational method engineering. Together with a "complementary taxonomy of context types (CTs) can be used to differentiate multiple scenarios of BPM development" [2], p. 549.

These scenarios encompass the character of the business process itself, major stakeholders and contextual variables. [14] points out similarities between business modelling and software design. Author distinguishes that while software scenarios (more commonly referred to as use cases) typically involve one or more users interacting with the

software; business scenarios typically involve a mix of human-computer and human-human interactions. Apart from that business scenarios can be modelled in both "as-is" (existing business) and "to-be" (potential future business) forms. In the case of low-structure scenarios it is possible to state that there are more uncertainties, which are posing difficulties for the traditional process simulation or modelling (e.g. research processes related to the new product or service, provision of an artistic performance). Such scenarios are difficult to grasp by traditional BPM approach.

Flexible business processes, which can be easily adapted to certain challenges, unexpected and rapid changes or unforeseen failures, are one of the most important challenges faced by the modern companies [7].

Traditional approaches (to business process modelling or business process simulation) try to anticipate how the actual work or tasks are performed at the designated design time. In some approaches it is possible to manually change the process structure at run time. This may however not be enough in rapidly evolving contexts [4]. In this case the design-time specifications of all possible actions require an extensive manual effort from the process analyst. Moreover, there is a need to anticipate all potential problems and ways to handle them. Apart from that, it requires a process-specific knowledge, which may be not available at the design time. In short the design stage may produce the solutions that are obsolete although initially may be perceived as satisfactory. Highly dynamic and low structured processes may run differently each time and the recovery procedures may be dependent on the contextual information.

Due to the facts stated before such processes may not be completely captured by common business process models. This is a recent open research question in the Business Process Modelling (BPM) field. Researchers have wondered how to "tackle scenarios characterized by being very dynamic and subject to higher frequency of unexpected contingencies than classical scenarios" [4], p. 38. Such scenarios apart from the competitive business world may also take place in for example crisis or emergency management.

In the case of such process variations or divergence from common, well-structured, pre-defined models are mostly due to autonomous user decisions as well as a result of unpredictable events and environmental changes. These changes in the context of the business processes make the whole structure significantly less rigid. The flow of control may be implicitly determined by the decisions made by autonomous agents of change as well as by the contextual conditions. These may be also coupled with previously unforeseen alternative activities, decisions and process fragments. Moreover this low-structured environment may be constantly changing making the traditional simulation and business process modelling very difficult if not impossible. Therefore this may be also a situation where tasks are mainly discovered as the process unfolds. In the worst scenario, there are no pre-defined views of the knowledge-intensive business processes. Modelling and simulation of such knowledge-intensive processes poses a challenge as described in [10].

The overview of the contributions of different authors provided above offers a broad perspective on the existing research as well as potential trends in the field. It can be perceived as an intersection across a wide range of most challenging research topics in the operational research. Further sections will provide a deeper insight into the contemporary approaches to simulation and modelling as well as Process-Aware Information Systems.

2 CONTEMPORARY APPROACHES

Due to the increase in the demand for effective solutions, methodologies and tools (that could aid the simulation and modelling of knowledge-intensive business processes) there were a growing number of proposed approaches ([4] p. 38). These approaches try to

emphasize how to integrate data, rules, user decisions and control flow in order to support the specification, analysis and simulation of such processes. It is possible to distinguish some commonalities between these approaches such as: focus on object-centric processes or artefact-centric processes, the adaptive process management or dynamic process management. Some researchers (e.g.: [3]; [19]) propose adaptive processes, declarative processes, late binding and modelling as possible solutions.

From both the practical as well as contextual viewpoint it is necessary to reason how to integrate these aspects with traditional business process management. It is still a largely open issue that needs to be addressed by relevant research. It is expected that the results may ultimately reshape the entire process life-cycle.

[7] proposes a concept for dynamic and automated business process workflow re-scheduling. This mechanism also allows almost instant recovery from task failures. It consists of a multistep procedure, which among others includes the termination of failed activities, suspension of the workflow of business processes and the generation of a new complete business process definition as well as satisfactory business process resumption. The last feature is one of the most interesting ones. "After suspending the process instance a new process definition is generated based on the current state of the case" ([7] p. 4). In this approach in terms of business process input instead of the initial state, the current state of the system is used. This current state is derived by "starting with the initial state and retracting all effects of all currently executed or terminated activities" ([7], p. 4). In this perspective the current state reflects all previously unexpected effects of failing activities or processes.

The fact that processes support the work in the highly dynamic contexts is a reality due to the growing use of mobile devices ([4], p. 38). In this case it is possible to talk about highly dynamic processes. Such processes usually consist of a wide range of knowledge-intensive tasks. As the process develops these tasks and their sequence depends heavily on the specific context and environment. For example an execution of a given dynamic process may be dependent on which resources and in what volume are available in a given process stage. Apart from that the number of available options determines the outcome of such processes. In this view dynamic processes are very close to chaotic systems where the particular cases may have the same entry point, but slight differences in subsequent conditions may produce completely different outcomes. In other words dynamic processes may be totally unpredictable in how they unfold. This is due to the high number of tasks represented, their unstable nature and the intrinsic difficulty to model the whole knowledge of the domain of interest in the design stage.

Another approach is to augment the existing process models with error-handling capabilities. These capabilities may be available to the process designer at various stages of the design process ranging from the process analysis to process run-time [6].

In this approach disrupted processes can be for instance rolled back or completed on an alternative execution path (e.g. customer instead of being handled by an online customer care can be redirected to the call centre). The authors propose pre-defined and dedicated error-handling mechanism and the concept of "ad-hoc" processes. Such processes can in principle be used to provide for "flexible error handling in case of an exception" ([6], p. 5). Ad hoc processes can be defined as an activity or group of activities that have no predefined execution order. In this view only the actual performers decide on the actual execution flow.

During the execution of the business process, whenever an exception or deviation from pre-defined process model occurs, a new repair plan in dynamically generated by taking into account constraints posed by the process structure and by the addition or deletion of the actions taken from a pre-defined generic repair plan. This generic repair plan is defined by the process analyst at the design stage.

[18] proposes a set of structural process change patterns in order to support the handling of unforeseen exceptions, which allow the possible ad-hoc deviations from pre-modelled processes. These can be applied at run-time. The research was based on the analysis of real world process models.

These suggested change patterns can be applied at process type level or process instance level. Some of the patterns can be utilized to delay decision regarding exact control to run-time in order to better deal with the uncertainty. In other words this approach advocates the use of pre-defined change patterns that already exist in a given industry (automotive and healthcare in the case of [19]). In the next stage it identifies those processes that usually occur and those that have a high degree of uncertainty. In the latter case user may choose to delay the exact process simulation as close as to the run-time as possible in order to know as many variables from the real-world as possible.

Industry specialists propose a combination of active knowledge modelling and business process management as an approach to simulate or model knowledge-intensive processes [1].

Active Knowledge Modelling (AKM) aims to support human (so called knowledge workers) in performing creative, knowledge-intensive work. Such approach acknowledges that 80-90% of all work processes cannot be completely automated. Moreover such processes bring most important value to the companies and are responsible for their competitive advantage. In this perspective tasks and process models are owned and defined by those performing the actual work.

Processes, products and services should be designed and adapted in parallel. Apart from that most processes are unique and may require instead of a generic model to be represented as instances of task patterns. In order to tackle the low-level structure of such processes it is necessary to consider processes as emerging right from the work instead of being enforced by constraints. Another feature is the interactivity of the proposed process models. Users of the process models have greater influence on the sequence, decision points than in the traditional processes. Moreover exceptions and deviations from the standard process model are supposed to be a norm and nothing extraordinary.

3 PROCESS-AWARE INFORMATION SYSTEMS

The ultimate goal of the management of business processes in highly dynamic and low-structure scenarios is to be able to come with a solution, tool or technology that will aid the process designers to cope with the challenges described in the previous paragraphs. PAIS are believed to be one of the potential technology answers in this context.

[20] point out that the frequent changes in the business environment force the Enterprise Information Systems to provide “flexible support while still enforcing some degree of control” (p. 93). The authors state that there is an essential requirement for maintaining higher cohesiveness between real-world business processes and the functionality provided by the IS. PAIS should be more flexible than traditional ones.

Different process aspects (e.g. organizational, functional, control flow and information perspectives) need to be met by adaptive process management. Different process levels need to be addressed as well. PAIS are (as opposed to data- or function-centred information systems) characterized by a strict separation of process logic and application code. Most PAISs model process logic explicitly in terms of a process model. In this view PAISs provide the schema for process execution [19].

One of the most important features PAISs can offer is the degree to which they can deal with process change. PAISs can achieve that by the separation of concerns. Separation of concerns is a (computer science) design principle. It provides a principle for separating a

computer program into distinct sections. Each of these sections addresses a separate concern. A concern (in computer science) is a set of information that affects the code of a computer program. A computer program that embodies separation of concerns well has a high degree of modularity. Modularity enhances separation and can be achieved by information encapsulation (information hiding). Another way to achieve separation of concerns is by a way of layers (e.g. presentation layer, business logic layer, database layer).

[19] notes that although there are lots of benefits of PAIS it is necessary to introduce a PAIS in a way that it does not freeze existing business processes. PAIS need to be flexible so that they can capture real-world processes in an adequate way and in the same way they do not lead to a mismatch between the computerized business process and those that run in the real world. High quality PAISs should also have the abilities, which would allow authorized users to deviate from predefined business processes when required (e.g. when there is a need to adopt a dynamic change in the environment or when there is a need to deal with exceptions). Apart from that a sought-after feature is the ability to "evolve" scenarios in order to continually adapt the available process models to process optimizations.

4 CONCLUSIONS

One of the most important and interesting problems of the contemporary operations research is how to tackle business processes in low-structured, highly dynamic scenarios. These sorts of scenarios differ considerably from the traditional settings, in which business processes are usually modelled. Therefore the application of traditional process models poses a challenge. Knowledge intensive business processes belong to a group of processes that are characterized by low structure and high volatility. This paper presented a broad overview of the existing literature on the relevant topics. It presented a theoretical background on which kind of business processes can be characterized as those occurring in low-structured, highly dynamic scenarios. This was followed by an investigation into the contemporary approaches to business process simulation and modelling and a theoretical reasoning on how these approaches can be applied to this special kind of business process scenarios. After that Process-Aware Information Systems were presented as a potential technological solution to the modelling and simulation of business processes in these special scenarios. It is expected that the modelling and simulation of knowledge intensive business processes can also benefit from the approaches, methodologies, tools and technologies mentioned in this paper. Another contribution of added value of this paper is, that it presents the possible solutions to these problems by providing a comprehensive literature review of the subject as well as some insights from the business world of what are the possible solutions.

References:

[1] Active Knowledge Modelling Group, (2009), BPM for Knowledge-Intensive Processes, http://activeknowledgemodeling.com/2009/03/13/bpm-for-knowledge-intensive-processes/ (accessed 26.07.2013)

[2] Bucher, T. and Winter, R., (2009), Project types of business process management, Towards a scenario structure to enable situational method engineering for business process management, Business Process Management Journal, Vol. 15, No. 4, pp. 548-568

[3] Adams, M., Hofstede, A. ter., Edmond, D., van der Aalst, W., (2006), A Service-Oriented Implementation of Dynamic Flexibility in Workflows, in: Proc. Coopis'06, pp. 291-308.

[4] Ciccio, C. D., Marrella, A. and Russo, A., (2012), Knowledge-intensive Processes: An Overview of Contemporary Approaches, CEUR Workshop Proceedings, Vol. 861, pp. 33-47.

[5] Evanschitzky, H., Ahlert, D., Blaich, G. and Kenning, P. (2007), Knowledge management in knowledge-intensive service networks: a strategic management approach, *Management Decision*, Vol. 45 No. 2, pp. 265-83.

[6] Friedrich, G., Fugini, M., Mussi, E., Pernici, B., Tagni, G., (2010), Exception Handling for Repair in Service-Based Processes. IEEE Trans. on Soft. Eng. 36.

[7] Gajewski, M., Meyer, H., Momotko, M., Schuschel, H.,Weske, M., (2005), Dynamic Failure Recovery of Generated Workows. In: DEXA'05 (2005)

[8] Gregoriades, A. and Sutcliffe, A., (2008), A socio-technical approach to business process simulation, *Decision Support Systems*, Vol. 45, Issue 4, pp. 1017-1030.

[9] Larsen, J. (2001), Knowledge, human resources and social practice: the knowledge-intensive business service firm as a distributed knowledge system, *Service Industries Journal*, Vol. 21 No. 1, pp. 81-102.

[10] Majewski, G., Damij, N., (2013), Knowledge intensive processes as a challenge for business process simulation, Recent Advances in Information Science, Proceedigns of the WSEAS ECC, Dubrovnik 2013.

[11] McCormack, K., Willems, J., van den Bergh, J., Deschoolmeester, D., Willaert, P., Indihar, M., Štemberger, M. I., Škrinjar, R., Trkman, P., Ladeira, M. B., de Oliveira, M. P. V., Vukšić, V. B. and Vlahovic, N., (2009), A global investigation of key turning points in business process maturity, Business Process Management Journal, Vol. 15, No. 5, 2009, pp. 792-815.

[12] Milanović G., Lj., (2011), Understanding Process Performance Measurement Systems, Business Systems Research**,** Vol.2, No.2, pp. 1-56.

[13] Nurmi, R. (1998), Knowledge-intensive firms, *Business Horizons*, Vol. 41 No. 3, pp. 26-32.

[14] Rosenberg, D., (2010), Business Process Modeling with Structured Scenarios, Whitepaper ICONIX Software Engineering, Inc., http://www.iconixsw.com/Articles/BPRoadmapV3.pdf (accessed 30.07.2013).

[15] Seethamraju, R. and Marjanovic, O., (2009), Role of process knowledge in business process improvement methodology: a case study, Business Process Management Journal, Vol. 15, No. 6, pp. 920-936.

[16] Smith, K., (2000), *What is the "knowledge economy"? Knowledge-intensive industries and distributed knowledge bases*, Step Group, Oslo, May 2000

[17] Starbuck, W.H., (1992), Learning by knowledge-intensive firms, *Journal of Management Studies*, Vol. 3, No. 4, pp. 262-75.

[18] Vlahović, N., (2010), Reaching inter-institutional business processes in e-Society, Business Systems Research, Vol.01, No.1-2, pp. 1-50.

[19] Weber, B., Reichert, M., Rinderle-Ma, S., (2008), Change Patterns and Change Support Features - Enhancing Flexibility in Process-aware Information Systems. Data Knowl. Eng. 66 (2008).

[20] Weber, B., Wild, W., Lauer, M., Reichert, M., (2006), Improving Exception Handling by Discovering Change Dependencies in Adaptive Process Management Systems, In: BPI'06 (2006).

[21] Wheelan, E., Collings, D. G. and Donnellan, B., (2009), Managing talent in knowledge-intensive settings, *Journal of Knowledge Management*, Vol. 14, No. 3, pp. 486-504

[22] Willoughby, K. and Galvin, P. (2005), Inter-organizational collaboration, knowledge intensity, and the sources of innovation in the bioscience-technology industries, *Knowledge, Technology, and Policy*, Vol. 18 No. 3, pp. 56-73.

The 12th International Symposium on
Operational Research in Slovenia

SOR '13

Dolenjske Toplice, SLOVENIA
September 25 - 27, 2013

Appendix

Authors' addresses

Addresses of SOR'13 Authors

(The 12th International Symposium on OR in Slovenia, Dolenjske Toplice, SLOVENIA, September 25 – 27, 2013)

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
1.	Jernej	Agrež	Faculty of Information Studies in Novo mesto	Ulica talcev 3	8000	Novo mesto	Slovenia	jernej.agrez@ fis.unm.si
2.	Zdravka	Aljinović	University of Split, Faculty of Economics	Cvite Fiskovića 5	21000	Split	Croatia	zdravka.aljinovic@ efst.hr
3.	Marcin	Anholcer	Poznań University of Economics, Faculty of Informatics and Electronic Economy	Al. Niepodległości 10	61-875	Poznań	Poland	m.anholcer@ ue.poznan.pl
4.	Zoran	Babić	University of Split, Faculty of Economics	Cvite Fiskovića 5	21000	Split	Croatia	babic@efst.hr
5.	Alenka	Baggia	University of Maribor, Faculty of Organizational Sciences	Kidričeva cesta 55a	4000	Kranj	Slovenia	alenka.baggia@ fov.uni-mb.si
6.	Karlo	Bala	University of Novi Sad, Faculty of Philosophy	Dr. Zorana Đinđića 2	21000	Novi Sad	Serbia	kbalayu@gmail.com
7.	Danijel	Barbarić	University of Split, Faculty of Law	Domovinskog rata 8	21000	Split	Croatia	danijel.barbaric@ pravst.hr
8.	Martina	Basarac Sertić	Economic Research Division, Croatian Academy of Sciences and Arts	Strossmayerov trg 2	10000	Zagreb	Croatia	mbasarac@hazu.hr
9.	David	Bogataj	European Faculty of Law			Nova Gorica	Slovenia	dbogataj@actuary.si

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
10.	Marija	Bogataj	MEDIFAS			Šempeter pri Gorici	Slovenia	marija.bogataj@ guest.arnes.si
11.	Drago	Bokal	University of Maribor; Faculty of Natural Sciences and Mathematics	Koroška cesta 160	2000	Maribor	Slovenia	drago.bokal@uni-mb.si
12.	Vesna	Bosilj Vukšić	University of Zagreb, Faculty Economics and Business, Department of Informatics	Trg J.F. Kennedyja 6	10000	Zagreb	Croatia	vbosilj@efzg.hr
13.	Andrej	Bregar	Informatika d.d.	Vetrinjska ulica 2	2000	Maribor	Slovenia	andrej.bregar@ informatika.si
14.	Alenka	Brezavšček	University of Maribor, Faculty of Organizational Sciences	Kidričeva cesta 55a	4000	Kranj	Slovenia	alenka.brezavscek@ fov.uni-mb.si
15.	Jože	Bučar	Laboratory of Data Technologies, Faculty of Information Studies in Novo mesto	Ulica talcev 3	8000	Novo mesto	Slovenia	joze.bucar@ fis.unm.si
16.	Sergio	Cabello	Department of Mathematics, FMF, University of Ljubljana	Jadranska 19	1111	Ljubljana	Slovenia	sergio.cabello@ fmf.uni-lj.si
17.	Kristijan	Cafuta	University in Ljubljana, Faculty of Electrical Engineering	Tržaška 25	1000	Ljubljana	Slovenia	kristijan.cafuta@ fe.uni-lj.si
18.	Katarína	Cechlárová	Institute of Mathematics, Faculty of Science, P. J. Šafárik University	Jesenná 5	040 01	Košice	Slovakia	katarina.cechlarova@ upjs.sk
19.	Banchongsan	Charoensook	Department of Business Administration, ALHOSN University				United Arab Emirates	b.charoensook@ alhosnu.ae

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
20.	Anthony	Chin	National University of Singapore	21 Lower Kent Ridge Road	119077	Singapore	Singapore	
21.	Vesna	Čančer	University of Maribor, Faculty of Economics and Business	Razlagova 14	2000	Maribor	Slovenia	vesna.cancer@uni-mb.si
22.	Draženka	Čizmić	University of Zagreb, Faculty of Economics and Business	Trg J.F. Kennedyja 6	10000	Zagreb	Croatia	dcizmic@efzg.hr
23.	Marko	Čular	University of Split, Faculty of Economics	Cvita Fiskovića	21000	Split	Croatia	marko.cular@efst.hr
24.	Marcello	Dalpasso	Dip. di Ingegneria dell'Informazione, University of Padova				Italy	marcello.dalpasso@ unipd.it
25.	Nadja	Damij	Faculty of Information Studies in Novo mesto	Ulica talcev 3	8000	Novo mesto	Slovenia	nadja.damij@ fis.unm.si
26.	Peter J.C.	Dickinson	Department of Statistics and Operations Research, University of Vienna	Oskar-Morgenstern-Platz 1	1090	Wien	Austria	peter.dickinson@ univie.ac.at
27.	Andrej	Dobrovoljc	University of Ljubljana, Faculty of Computer and Information Science	Tržaška cesta 25	1000	Ljubljana	Slovenia	andrej.dobrovoljc@ fis.unm.si
28.	Cezary	Dominiak	University of Economics in Katowice, Department of Operations Research	Ul. 1 Maja 50	40-287	Katowicw	Poland	cezary.dominiak@ ue.katowice.pl
29.	Samo	Drobne	University of Ljubljana, Faculty of Civil and Geodetic Engineering	Jamova 2	1000	Ljubljana	Slovenia	samo.drobne@ fgg.uni-lj.si
30.	Ksenija	Dumičić	University of Zagreb, Faculty of Economics and Business	Trg J.F. Kennedy 6	10000	Zagreb	Croatia	kdumicic@efzg.hr

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
31.	Rija	Erveš	Institute of Mathematics, Physics and Mechanics; University of Maribor, Faculty of Civil Engineering	Jadranska 19; Smetanova 17	1000; 2000	Ljubljana; Maribor	Slovenia	rija.erves@ um.si
32.	Liljana	Ferbar Tratar	University of Ljubljana, Faculty of Economics	Kardeljeva ploščad 17	1000	Ljubljana	Slovenia	liljana.ferbar.tratar@ ef.uni-lj.si
33.	Tamás	Fleiner	Department of Computer Scince and Information Theory, Budapest University of Technology and Economics	Magyar tudósok körútja 2	1117	Budapest	Hungary	fleiner@cs.bme.hu
34.	Helena	Gaspars-Wieloch	Poznan University of Economics	Al. Niepodleglosci 10	61-875	Poznan	Poland	helena.gaspars@ ue.poznan.pl
35.	Mirko	Gradišar	University of Ljubljana, Faculty of Economics	Kardeljeva ploščad 17	1000	Ljubljana	Slovenia	miro.gradisar@ ef.uni-lj.si
36.	Rainer	Graf	University of Applied Sciences Technikum Wien	Höchstädtplatz 6	1200	Vienna	Austria	It10m027@ technikum-wien.at
37.	Petra	Grošelj	University of Ljubljana, Biotechnical Faculty	Jamnikarjeva 101	1000	Ljubljana	Slovenia	petra.groselj@bf.uni-lj.si
38.	Nebojša	Gvozdenović	University of Novi Sad, Faculty of Economics Subotica	Segedinski put 9-11	24000	Subotica	Serbia	nebojsa.gvozdenovic@ gmail.com
39.	Marko	Hell	University of Split, Faculty of Economics	Cvite Fiskovića 5	21000	Split	Croatia	marko.hell@efst.hr
40.	Eloy	Hontoria	GIO Universidad Politécnica de Cartagena	Campus Muralla del Mar	30202	Cartagena	Spain	eloy.hontoria@upct.es

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
41.	Ivan	Horvat	VB Leasing d.o.o.	Horvatova 82	10000	Zagreb	Croatia	ivan.horvat@ vbleasing.hr
42.	Marko	Intihar	University of Maribor, Faculty of Logistics	Mariborska 7	3000	Celje	Slovenia	marko.intihar@fl.uni-mb.si
43.	Marko	Jakšič	University of Ljubljana, Faculty of Economics	Kardeljeva ploščad 17	1000	Ljubljana	Slovenia	marko.jaksic@ ef.uni-lj.si
44.	Jaroslav	Janáček	University of Žilina, Faculty of Management and Informatics	Univerzitná 1	010 26	Žilina	Slovakia	jaroslav.janacek@ fri.uniza.sk
45.	Zied	Jemai	Laboratoire Génie Industriel (LGI), Ecole Centrale Paris	Grande Voie des Vignes	92 295	Châtenay-Malabry	France	zied.jemai@ecp.fr
46.	Dragan	Jukić	J. J. Strossmayer University of Osijek, Department of Mathematics	Trg Ljudevita Gaja 7	31000	Osijek	Croatia	jukicd@mathos.hr
47.	Sandi	Klavžar	Universtity of Ljubljana, FMF; University of Maribor, FNM	Jadranska 19; Koroška 160	1000; 2000	Ljubljana; Maribor	Slovenia	sandi.klavzar@fmf.uni-lj.si
48.	Igor	Klep	University of Maribor, Faculty of Natural Science and Mathematics; Universtity of Ljubljana, Faculty of Mathematics and Physics; The University of Auckland, Department of Mathematics	Koroška 160; Jadranska 19; Private Bag 92019	2000; 1111; 1124	Maribor; Ljubljana; Auckland	Slovenia; New Zealand	igor.klep@ fmf.uni-lj.si igor.klep@ auckland.ac.nz
49.	Michal	Kohány	University of Zilina, Faculty of Management Science and Informatics	Univerzitna 8215/1	01026	Zilina	Slovakia	michal.kohani@ fri.uniza.sk

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
50.	Tadej	Kolmanič	Center odličnosti za biosenzoriko, instrumentacijo in procesno kontrolo	Tovarniška 26	5270	Ajdovščina	Slovenia	tadej.kolmanic@cobik.si
51.	Danijel	Kovačić	Danijel Kovačić s.p., Informacijsko svetovanje	Grič 4	1000	Ljubljana	Slovenia	kovacic.danijel@ gmail.com
52.	Tomaž	Kramberger	University of Maribor, Faculty of Logistics	Mariborska 7	3000	Celje	Slovenia	tomaz.kramberger@ fl.uni-mb.si
53.	Giuseppe	Lancia	Dip. di Matematica e Informatica, University of Udine				Italy	giuseppe.lancia@ uniud.it
54.	Michael	Löffler	Austro Control Corporation				Austria	michael.loeffler@ austrocontrol.at
55.	Grzegorz	Majewski	Institute for Process Management, Faculty of Information Studies in Novo mesto	Ulica talcev 3	8000	Novo mesto	Slovenia	g.majewski@ fis.unm.si
56.	Marjana	Merkač Skok	Fakulteta za poslovne in komercijalne vede	Lava 5	3000	Celje	Slovenia	marjana.merkac@fkpv.si
57.	Nenad	Mirkov	University of Novi Sad, Faculty of Economics Subotica	Segedinski put 9-11	24000	Subotica	Serbia	nenad.mirkov@ gmail.com
58.	Mahdi	Moeini	Laboratoire Génie Industriel (LGI), Ecole Centrale Paris; Braunschweig University of Technology, IBR, Algorithms Group	Grande Voie des Vignes; Mühlenpfordtstr. 23	92 295; 38106	Châtenay-Malabry; Braunschweig	France; Germany	mahdi.moeini@ecp.fr; moeimi@ibr.cs.tu-bs.de
59.	Gerhard	Navratil	Vienna University of Technology, Department for Geodesy and Geoinformation	Gusshausstr. 27-29/E120.2	1040	Vienna	Austria	navratil@ geoinfo.tuwien.ac.at

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
60.	Maciej	Nowak	University of Economics in Katowice, Faculty of Informatics and Communication	1 Maja 50	40-287	Katowicw	Poland	maciej.nowak@ ue.katowice.pl
61.	Kosovka	Ognjenović	Institute of Economic Sciences	12 Zmaj Jovina	11000	Belgrade	Serbia	kosovka.ognjenovic@ ien.bg.ac.rs
62.	Irena	Palić	University of Zagreb, Faculty of Economics and Business	Trg J.F. Kennedy 6	10000	Zagreb	Croatia	ipalic@efzg.hr
63.	Anita	Pavković	University of Zagreb, Faculty of Economics and Business	Trg J.F. Kennedy 6	10000	Zagreb	Croatia	apavkovic@efzg.hr
64.	Polona	Pavlič	University of Maribor; Faculty of Natural Sciences and Mathematics	Koroška cesta 160	2000	Maribor	Slovenia	polona.pavlic@uni-mb.si
65.	Polona	Pavlovčič Prešeren	University of Ljubljana, Faculty of Civil and Geodetic Engineering	Jamova 2	1000	Ljubljana	Slovenia	polona.pavlovcic@ fgg.uni-lj.si
66.	Mirjana	Pejić Bach	University of Zagreb, Faculty Economics and Business, Department of Informatics	Trg J.F. Kennedyja 6	10000	Zagreb	Croatia	mpejic@efzg.hr
67.	Tunjo	Perić	University of Zagreb, Faculty Economics and Business, Department of Informatics	Trg J.F. Kennedyja 6	10000	Zagreb	Croatia	tperic@ efzg.hr
68.	Michel	Petitjean	MTi, INSERM UMR-S 973, University Paris 7	35 rue Hélène Brion	75205	Paris Cedex 13	France	petitjean.chiral@ gmail.com
69.	Sanja	Pfeifer	University of Josip Juraj Strossmayer in Osijek, Faculty of Economics	Gajev trg 7	31000	Osijek	Croatia	pfeifer@efos.hr
70.	Snježana	Pivac	University of Split, Faculty of Economics	Cvita Fiskovića	21000	Split	Croatia	snjezana.pivac@efst.hr

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
71.	Tea	Poklepović	University of Split, Faculty of Economics	Cvite Fiskovića 5	21000	Split	Croatia	tea.poklepovic@efst.hr
72.	Eva	Potpinková	Institute of Mathematics, Faculty of Science, P. J. Šafárik University	Jesenná 5	040 01	Košice	Slovakia	eva.potpinkova@ student.upjs.sk
73.	Janez	Povh	Faculty of Information Studies in Novo mesto	Ulica talcev 3	8000	Novo mesto	Slovenia	janez.povh@ fis.unm.si
74.	Željko	Račić	University of Banja Luka, Faculty of Economics	Majke Jugovića 4	78000	Banja Luka	Bosnia and Herzegovina	zeljko.racic@efbl.org
75.	Lorenzo	Ros-McDonnell	GIO Universidad Politécnica de Cartagena	Campus Muralla del Mar	30202	Cartagena	Spain	lorenzo.ros@upct.es
76.	Darja	Rupnik Poklukar	University of Ljubljana, Faculty of Mechanical Engineering	Aškerčeva ulica 6	1000	Ljubljana	Slovenia	darja.rupnik@fs.uni.lj.si
77.	Evren	Sahin	Laboratoire Génie Industriel (LGI), Ecole Centrale Paris	Grande Voie des Vignes	92 295	Châtenay-Malabry	France	evren.sahin@ecp.fr
78.	Sebastian	Sitarz	Institute of Mathematics, University of Silesia in Katowice	Ul. Bankowa 14	40-007	Katowice	Poland	ssitarz@ ux2.math.us.edu.pl
79.	Andreja	Smole	Cosylab d.d.	Teslova 30	1000	Ljubljana	Slovenia	andreja.smole@ cosylab.com
80.	Renata	Sotirov	Department of Econometrics & Operational Research, Tilburg University	Warandelaan 2 P.O. Box 90153	5000	LE Tilburg	The Netherlands	r.sotirov@uvt.nl
81.	Oskar	Sterle	University of Ljubljana, Faculty of Civil and Geodetic Engineering	Jamova 2	1000	Ljubljana	Slovenia	oskar.sterle@ fgg.uni-lj.si

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
82.	Bojan	Stopar	University of Ljubljana, Faculty of Civil and Geodetic Engineering	Jamova 2	1000	Ljubljana	Slovenia	bojan.stopar@ fgg.uni-lj.si
83.	Tamara	Straživuk	University of Banja Luka, Faculty of Economics	Majke Jugovića 4	78000	Banja Luka	Bosnia and Herzegovina	bytamara@gmail.com
84.	Nataša	Šarlija	University of Josip Juraj Strossmayer in Osijek, Faculty of Economics	Gajev trg 7	31000	Osijek	Croatia	natasa@efos.hr
85.	Simona	Šarotar Žižek	University of Maribor, Faculty of Economics and Business	Razlagova 14	2000	Maribor	Slovenia	simona.sarotar-zizek@uni-mb.si
86.	Slavko	Šimundić	University of Split, Faculty of Law	Domovinskog rata 8	21000	Split	Croatia	slavko.simundic@ pravst.hr
87.	Damjan	Škulj	University of Ljubljana, Faculty of Social Sciences	Kardeljeva ploščad 5	1000	Ljubljana	Slovenia	damjan.skulj@ fdv.uni-lj.si
88.	Sabina	Šmigoc	University of Maribor; Faculty of Natural Sciences and Mathematics	Koroška cesta 160	2000	Maribor	Slovenia	sabina.smigoc@ gmail.com
89.	Petra	Šparl	Institute of Mathematics, Physics and Mechanics; University of Maribor, Faculty of Organizational Sciences	Jadranska 19; Kidričeva cesta 55a	1000; 4000	Ljubljana; Kranj	Slovenia	petra.sparl@ fov.uni-mb.si
90.	Mitja	Štiglic	University of Ljubljana, Faculty of Economics	Kardeljeva ploščad 17	1000	Ljubljana	Slovenia	mitja.stiglic@ ef.uni-lj.si
91.	Andrej	Taranenko	University of Maribor; Faculty of Natural Sciences and Mathematics	Koroška cesta 160	2000	Maribor	Slovenia	andrej.taranenko@ uni-mb.si

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
92.	Simon	Thevenin	Faculty of economics and Social Sciences, HEC-University of Geneva, Uni-Mail	Bd du Pont-d'Arve 40	1211	Geneva 4	Switzerland	simon.thevenin@unige.ch
93.	Luka	Tomat	University of Ljubljana, Faculty of Economics	Kardeljeva ploščad 17	1000	Ljubljana	Slovenia	luka.tomat@ef.uni-lj.si
94.	Katarina	Tomičić-Pupek	University of Zagreb, Faculty of Organization and Informatics	Pavlinska 2	42000	Varaždin	Croatia	ktomicic@foi.hr
95.	Tadeusz	Trzaskalik	University of Economics in Katowice, Faculty of Informatics and Communication	Ul. 1 Maja 50	40-287	Katowice	Poland	tadeusz.trzaskalik@ue.katowice.pl
96.	Ana	Vehovec	DRI, upravljanje investicij d.o.o.	Kotnikova 40	1000	Ljubljana	Slovenia	
97.	Aleksander	Vesel	University of Maribor; Faculty of Natural Sciences and Mathematics	Koroška cesta 160	2000	Maribor	Slovenia	vesel@uni-mb.si
98.	Jelena	Vidović	University of Split, The University Department of Professional Studies	Kopilica 5	21000	Split	Croatia	jvidovic@oss.unist.hr
99.	Tea	Vizinger	University of Maribor, Faculty of Logistics	Mariborska 7	3000	Celje	Slovenia	tea.vizinger@fl.uni-mb.si
100.	Robert	Vodopivec	MEDIFAS			Šempeter pti Gorici	Slovenia	vodopivec.robert@siol.net
101.	Tina	Vuko	University of Split, Faculty of Economics	Cvita Fiskovića	21000	Split	Croatia	tina.vuko@efst.hr
102.	Marino	Widmer	University of Fribourg – DIUF, Decision Support & Operations Research	Bd de Pérolles 90	1700	Fribourg	Switzerland	marino.widmer@unifr.ch

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
103.	Lidija	Zadnik Stirn	University of Ljubljana, Biotechnical Faculty	Jamnikarjeva 101	1000	Ljubljana	Slovenia	lidija.zadnik@bf.uni-lj.si
104.	Marijana	Zekić-Sušac	University of Josip Juraj Strossmayer in Osijek, Faculty of Economics	Gajev trg 7	31000	Osijek	Croatia	marijana@efos.hr
105.	Nicolas	Zufferey	Faculty of economics and Social Sciences, HEC-University of Geneva, Uni-Mail	Bd du Pont-d'Arve 40	1211	Geneva 4	Switzerland	nicolas.zufferey@unige.ch
106.	Janez	Žerovnik	Institute of Mathematics, Physics and Mechanics; University of Ljubljana, FME	Jadranska 19; Aškerčeva 6	1000	Ljubljana	Slovenia	janez.zerovnik@imfm.uni-lj.si; janez.zerovnik@fs.uni-lj.si
107.	Petra	Žigert Pleteršek	Faculty of Chemistry and Chemical Engineering, University of Maribor	Smetanova 17	2000	Maribor	Slovenia	petra.zigert@um.si