Elektrotehniški vestnik 74(4): 195-200, 2007
Electrotechnical Review: Ljubljana, Slovenija

Data Mining Based Decision Support System to Support Association Rules

Rok Rupnik, Matjaž Kukar

University of Ljubljana, Faculty of Computer and Information Science
Tržaška 25, 1000 Ljubljana
E-pošta: rok.rupnik@fri.uni-lj.si

Abstract. Modern organizations use several types of decision support systems to facilitate decision support. In many cases OLAP based tools are used in the business area enabling multiple views on data and through that a deductive approach to data analysis. Data mining extends the possibilities for decision support by discovering patterns and relationships hidden in data and therefore enabling an inductive approach to data analysis. The use of data mining to facilitate decision support enables new approaches to problem solving. The paper introduces our approach to integration of decision support systems with data mining methods. We introduce a data mining based decision support system designed for business users enabling them to use association rule models to facilitate decision support with only a basic level of knowledge of data mining.

Keywords: data mining, decision support system, decision support, association rules

Sistem za podporo odločanju z uporabo metode asociacijskih pravil

Povzetek. Sodobne organizacije uporabljajo različne sisteme za podporo odločanju in s tem podporo odločitvenim procesom. Za podporo odločanju za poslovne namene se večinoma uporabljajo orodja OLAP, ki omogočajo različne poglede na podatke in s tem deduktiven pristop k analiziranju podatkov. Odkrivanje zakonitosti v podatkih razširja možnosti za podporo odločanju prek odkrivanja vzorcev in razmerij, skritih v podatkih, ter s tem induktivni pristop k analiziranju podatkov. Uporaba odkrivanja zakonitosti v podatkih za potrebe podpore odločanju omogoča nove pristope k premagovanju problemov. Prispevek predstavlja sistem za podporo odločanju, ki temelji na metodi asociacijskih pravil in je namenjen poslovnim uporabnikom, ki imajo le osnovno raven znanja o odkrivanju zakonitosti v podatkih.

Keywords: odkrivanje zakonitosti v podatkih, sistem za podporo odločanju, podpora odločanju, asociacijska pravila

1 Introduction

Modern organizations use several types of decision support systems to facilitate decision support. For the purposes of analysis and decision support in the business area in many cases OLAP based decision support systems are used [1]. Performing analysis through OLAP follows a deductive approach of analyzing data [8]. The disadvantage of such an approach is that it depends on coincidence or even luck of choosing the right dimensions at drilling-down to acquire the most valuable information, trends and patterns. We could say that OLAP systems provide analytical tools enabling user-led analysis of the data, where the user has to start the right query in order to get the appropriate answer [9]. Such an approach enables mostly the answers to the questions like: “What is overall revenue for the first quarter grouped by customers?” What about the answers to the questions like: “What are characteristics of our best customers?” Those answers can not be provided by OLAP systems, but by the using data mining. Performing analysis through data mining follows an inductive approach of analyzing data [8].

Data mining is a process of analyzing data in order to discover implicit, but potentially useful information and uncover previously unknown patterns and relationships hidden in data. The use of data mining to facilitate decision support can lead to an improved performance of decision making and can enable the tackling of new types of problems that have not been addressed before. The integration of data mining and decision support can significantly improve current approaches and create new approaches to problem

Received 5 December 2006
Accepted 10 March 2007

solving, by enabling the fusion of knowledge from experts and knowledge extracted from data [9].

The paper introduces decision support system called DMDSS (Data Mining Decision Support System) which is based on data mining. DMDSS enables integration of data mining into decision processes by enabling repeated creation of data mining models. In DMDSS, data mining models are created by data mining experts and exploited by business users.

2 Data Mining

Data mining is a process of analyzing data in order to discover implicit, but potentially useful information and uncover previously unknown patterns and relationships hidden in data. In the last decade, the digital revolution has provided relatively inexpensive and available means to collect and store data. The increase in the data volume causes greater difficulties in extracting useful information for decision support. The traditional manual data analysis has become insufficient, and methods for efficient computer-based analysis indispensable. From this need, a new interdisciplinary field of data mining has been born. Data mining encompasses statistical, pattern recognition, and machine learning tools to support the analysis of data and discovery of principles that lie within the data [11].

In this paper we consider a data mining method called association rules, which belongs to unsupervised learning. In unsupervised learning, the goal is to describe associations and patterns among a set of input measures.

3 Integrating Data Mining and Decision Support

Decision support systems (DSS) are defined as interactive computer-based systems intended to help decision makers to utilize data and models in order to identify problems, solve problems and make decisions. They incorporate both data and models and they designed to assist decision makers in semi-structured and unstructured decision making processes. They provide support for decision making, they do not replace it. The mission of decision support systems is to improve effectiveness, rather than the efficiency of decisions [9].

Several authors discuss integration of data mining into decision support and they all confirm the value of it. Chen argues that the use of data mining helps institutions to make critical decisions faster and with a greater degree of confidence. He believes that the use of data mining lowers the uncertainty in the decision process [2]. Lavrac and Bohanec claim that the integration of data mining and decision support can lead to an improved performance of decision support systems and can enable the tackling of new types of problems that have not been addressed before. They also argue that the integration of data mining and decision support can significantly improve the current approaches and create new approaches to problem solving, by enabling the fusion of knowledge from experts and knowledge extracted from data [9].

4 Introduction of DMDSS

We developed DMDSS on Oracle platform. We did this after Oracle has offered Oracle Data Mining (ODM) option integrated in the database enabling the use of data mining methods on data in Oracle database. Besides ODM there is also Java based API available which enables the development of J2EE applications which use data mining.

Our decision to develop DMDSS was also based on the fact that the traditional use of data mining through data mining software tools does not bring data mining closer to business users because of complexity of data mining tools [4, 5, 6, 7]. Data mining tools are very complex and demand expertise in data mining, i.e. understanding of data mining algorithms and parameters for algorithms. We wanted to develop a decision support system that would enable data mining experts to create data mining models and enable business users to exploit data mining models through easy-to-use GUI. DMDSS was developed for a wireless network operator for the purposes of decision support in the area of analytical CRM (Customer Relationship Management).

4.1 A Process Model for DMDSS

One of the key issues in the design of DMDSS was to determine the data mining process model. The process model for DMDSS is based on the CRISP-DM (CRoss Industry Standard Process for Data Mining) process model. CRISP-DM process model breaks down the data mining activities into the following six phases which all include a variety of tasks [10, 3]: business understanding, data understanding, data preparation, modelling, evaluation and deployment. CRISP-DM process model was adapted to the needs of DMDSS as a two stage model: the preparation stage and the production stage. The division into two stages is based on the following two demands. First, DMDSS should enable repeated creation of data mining models based on an up-to-date data set for every area of analysis. Second, business users should only use it within the deployment phase with only the basic level of understanding of data mining concepts. Area of analysis is a business domain on which business users perform analysis and make decisions.

The preparation stage represents the process model for the use of DMDSS for the purposes of preparation of the area of analysis for the production use (Fig. 1). During the preparation stage, the CRISP-DM phases are performed in multiple iterations with the emphasize on

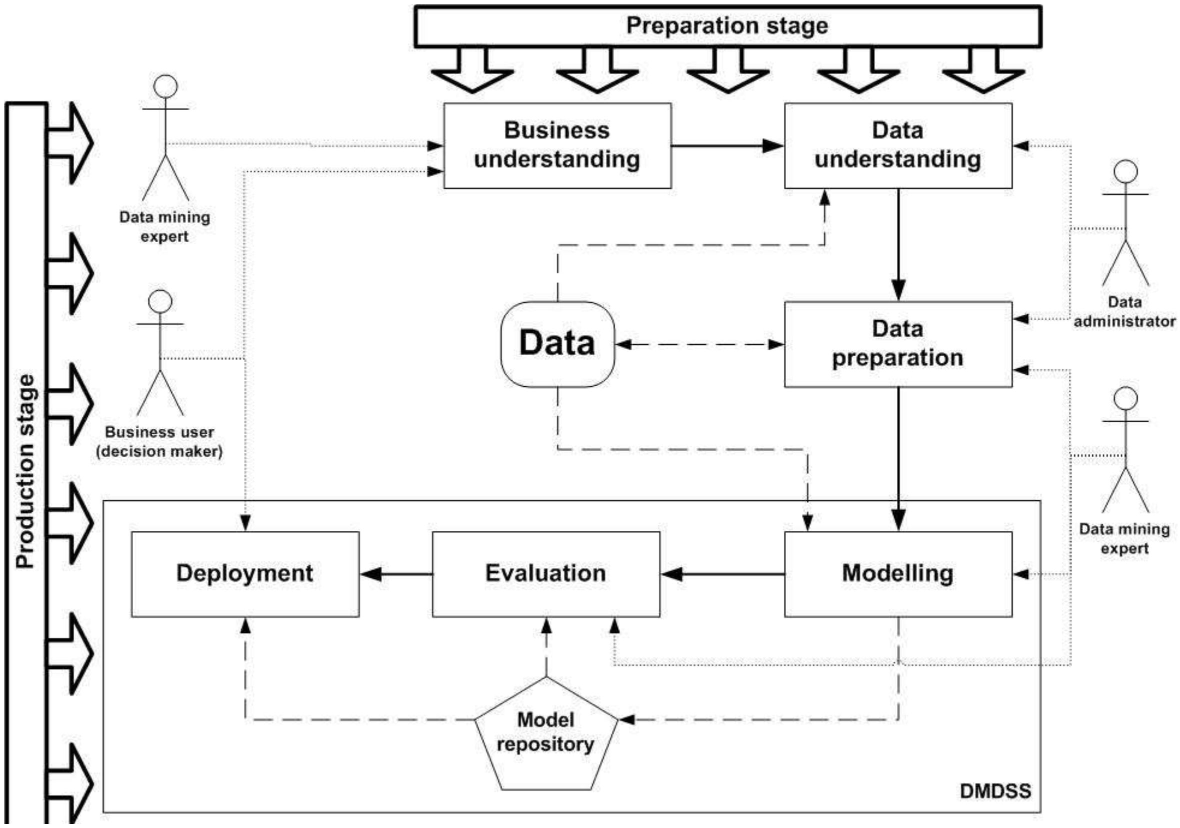

Figure 1: Scheme of the process model for DMDSS

the first five phases starting from business understanding and ending with evaluation. The aim of executing multiple iterations of all CRISP-DM phases for every area of analysis is to achieve step-by-step improvements in any of the phases. In the business understanding phase, slight redefinitions of the objectives can be made, if necessary, according to the results of other phases, especially the results of the evaluation phase. In the data preparation phase the improvements in the procedures which execute recreation of the data set can be achieved. The data set must be recreated automatically every night based on the current state of data warehouse and transactional databases. The problems detected in the data preparation phase can also demand changes in the data understanding phase.

In the modeling and evaluation phase, the model is created and evaluated for several times to allow fine-tuning of data mining algorithms through finding proper values of the algorithm parameters. It is essential to do enough iterations in order to monitor the level of changes in data sets and data mining models acquired and reach the stability of the data preparation phase and parameter values for data mining algorithms. The mission of the preparation stage is to confirm the fulfilling of the objectives of the area of analysis for decision support and to assure the stability of data preparation.

The production stage represents the production use of DMDSS for the area of analysis (Fig. 1). In the production stage the emphasis is on the phases of modeling, evaluation and deployment, which does not mean that other phases are not encompassed in the production stage. Data preparation, for example, is executed automatically based on procedures developed in the preparation stage. Modeling and evaluation are performed by a data mining expert, while the deployment phase is performed by a business user.

4.2 Functionalities of DMDSS

Our analysis of the functional demands for DMDSS was done simultaneously with the design of the data mining process model. Both activities are interrelated, because the process model implicitly defines the functionality of the decision support system. DMDSS supports two roles: the data mining expert and the business user. Each of the roles has access to the forms and their functionalities according to the production stage of the process model. DMDSS enables the use of the following data mining methods: classification, clustering and association rules. In the following part of the paper only the support for association rules will be introduced.

In the modeling phase, the data mining expert can create association rules models using the model creation form. Data mining expert sets three algorithm parameters: minimal support, minimal confidence and maximal rule length. Especially important is the possibility to set the minimal value for support and confidence. The model creation algorithm may produce a large amount of rules; in some cases there might be hundreds of thousands of them.

In the evaluation phase, the data mining expert can view and evaluate the model using the data mining expert model viewing form. This form is very similar to the form for model viewing which is used by the business user in the deployment phase and will be introduced later on. While viewing, the administrator can input comments for the model to business users at model interpretation. In the final step of the evaluation phase, the data mining administrator can change the published status of the model to a value *true* if he estimates that there are changes in the model which would be noteworthy to business users, which can view only models with published status set to *true*. The administrator observes the changes according to the previous model of a particular area of analysis. Due to the fact that in many cases the number of rules can be rather high, the observing of changes is not easy. This problem can be handled through setting the minimal value for confidence and/or support at model creation to a rather high value, e.g. 0.90 or 0.95. There is an enhancement planed in DMDSS to solve this problem which will be introduced later on in the paper.

For the purposes of the deployment phase, business users have access to the form model viewing (Fig. 2). The form shows general information about the model: model name, date of model creation, number of rules and number of comments. The presentation technique

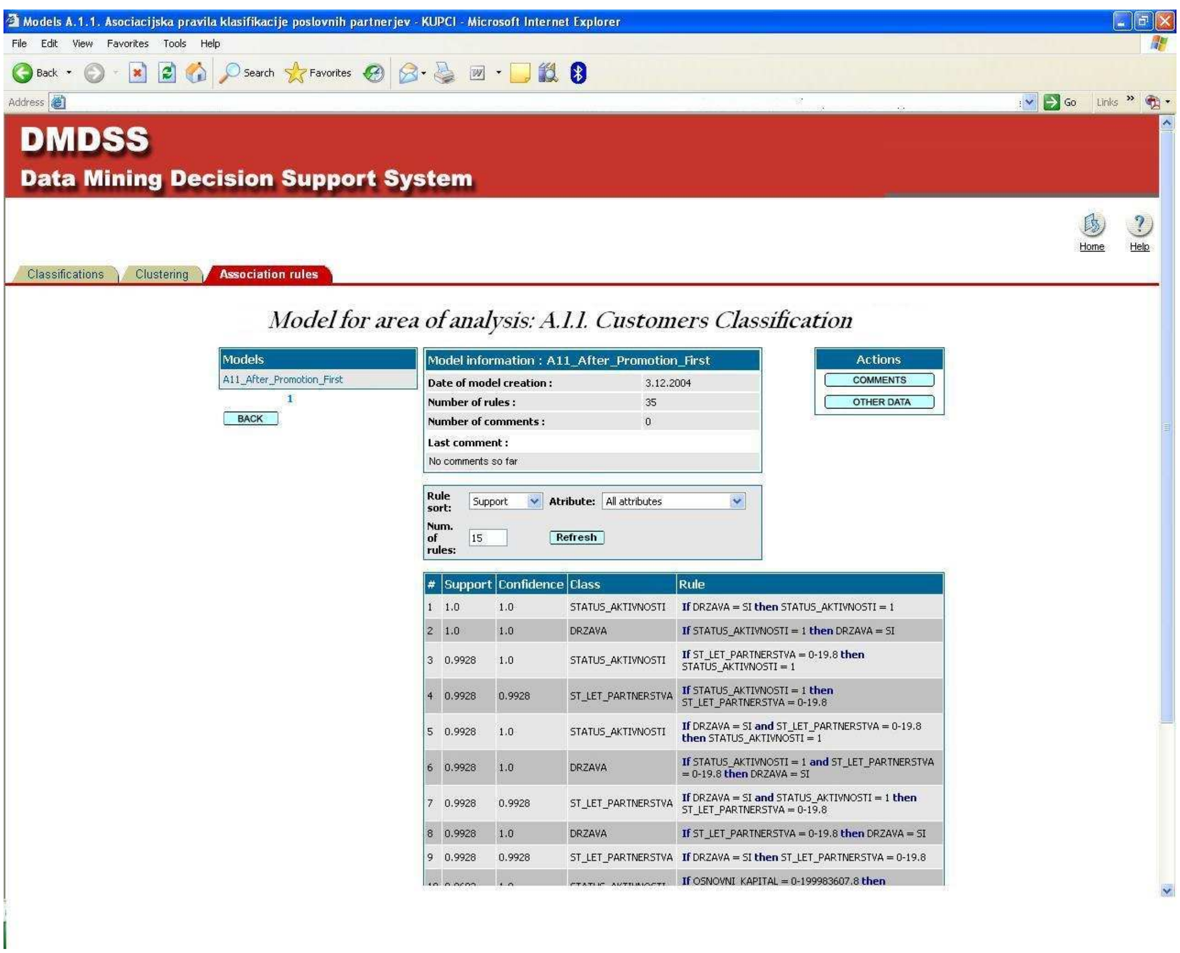

Figure 2: Model viewing form for business users

for association rules is a table showing confidence, support, class and a rule are presented. The class is the attribute on the right-hand side of the rule and serves as a dependent variable for the rule in question. It is important for the user to see the class in a separate column, even though the same information is presented in the rule. This way the user can easily scan through rules and focus on the attribute he is interested in. The form shows by default the rules for all attributes, but the user can filter the rules according to the class and this way focusing on a particular attribute is trivial. The user can sort the rules either by confidence or by support. Due to the fact that the model may have a lot of rules, the user can control the number of rules shown by setting the parameter for that purpose. Through combining of sorting, filtering and controlling the number of rules shown, the user can flexibly form criteria for rules shown according to his wishes. The form also shows the last comment of the model written by the data mining expert and enables access to all previous comments through an opening the form for comments.

The example of the area of analysis which uses association rules method in DMDSS is called "Customers Classification". Through the association rule model and monitoring the relations between attributes business users can get patterns and information to facilitate decision support. Especially interesting are the rules where the class is an attribute, which has the role of the classification attribute in the training set for the model creation using the classification method. This enables business users to acquire additional patterns and rules for a particular customer category.

5 Related Work

Some decision support systems that use data mining have already been developed and introduced in the literature. Polese introduced a decision support system based on data mining [13]. The system was designed to support tactical decisions of a basketball coach during a basketball match through suggesting tactical solutions based on the data of the past games. The decision support system only supports the association rules data mining method and uses the association rule algorithm called *Apriori* algorithm combined with the *Decision query* algorithm. The decision support system enables the coach to submit data about his tactical strategies and data about the game and the rival team. After that the system provides the coach with opinion about the chosen strategies and with suggestions. The system is not designed to support other domains; it only supports the basketball domain.

Bose and Sugumaran introduced the Intelligent Data Miner (IDM) decision support system [1]. IDM is a Web-based application system intended to provide organization-wide decision support capability for business users. Besides data mining it also supports some other function categories to enable decision support: data inquiry and multidimensional analysis through enabling OLAP views on multidimensional data. In the data mining part of IDM it supports the creation of models, manipulation of models and presentation of models in various presentation techniques of, among others, the following data mining methods: association rules, clustering and classifiers (classification). The system also performs data cleansing and data preparation and provides necessary parameters for data mining algorithms. An interesting characteristic of IDM is that it makes a connection to an external data mining software tool which performs data mining model creation. The system enables predefined and ad-hoc data mining model creation. The authors state that the disadvantage of IDM is the fact that non-technical users (business users) need to have a fair amount of understanding of data mining and that the use of data mining and the creation of data mining models still needs to be clearly directed by the user, especially with ad-hoc model creation.

Lee and Park presented the Customized Sampling Decision Support System (CSDSS) which uses data mining [12]. CSDSS is a web-based system that enables the user to select a process sampling method that is most suitable according to his needs at purchasing semiconductor products. The system enables the autonomous generation of the available customized sampling methods and provides the performance information for those methods. CSDSS uses clustering data mining method within the generation of sampling methods. The system is not designed to support other domains; it only supports the domain mentioned.

6 Discussion

DMDSS has now been in production for several months. During the first year of its production there will be supervising and consultancy provided by the development team. The main goal of supervising and consultancy is to assist the data mining administrator in the company. The person responsible for that role has enough knowledge, because he was a member of development team and all the time present at preparation stage for every area of analysis. But, he has not enough experience yet. Supervising will mainly cover support at model evaluation and model interpretation for data mining administrator and business users.

Business users use DMDSS at their daily work. They use patterns and rules identified in models as the new knowledge, which they use for analysis and decision process at their work. It is becoming apparent that they are getting used to DMDSS. According to their words they have already become aware of the advantages of continual use of data mining for analysis purposes. Based on the models acquired they have already prepared some changes in marketing approach and they are planning a special customer group focused campaign utilizing the knowledge acquired in data mining models. The most important achievement after several months of usage is the fact that business users have really started to understand the potentials of data mining. All of a sudden they have got new ideas for new areas of analysis, because they have started to realize how to define areas of analysis to acquire valuable results. A list of new areas of analysis will be made in a couple of months, and after that it will be discussed and evaluated. Selected areas of analysis will then be implemented and introduced to DMDSS according to the process model introduced in the paper.

Besides introducing new areas of analysis, there are other new functionalities planned for implementation. In many cases association rules models consist of a large number of rules. In these cases it is hardly possible for the data mining administrator to check them all in order to evaluate them and detect changes. For a business user it is also difficult to browse through lots of rules and detect valuable new knowledge. For that reason we plan to develop a method which will detect association rules that are different from the rules of the previous model. Business users believe this would significantly increase their efficiency at utilizing association rules.

The experience in the use of DMDSS has also revealed that business users need the possibility to make their own archive of classification rules and to have an option to make their own comments to archived rules in order to record the ideas implied and gained by the rules. These enhancements are planned to be implemented in the future.

7 References

[1] R. Bose, V. Sugumaran, Application of Intelligent Agent Technology for Managerial Data Analysis and Mining, *The DATABASE for Advances in Information Systems*, 30, 1, pp. 77-94.

[2] K.C. Chen, Decision Support System for Tourism Development: System Dynamic Approach, *Journal of Computer Information Systems*, 45, 1, pp. 104-112, 2004

[3] C. Clifton, A. Thuringsigham, Emerging Standards for Data Mining, *Computer Standards & Interfaces*, 23, 3, pp. 187-193, 2001

[4] M. Goebel, L.A. Gruenwald; A Survey on Data Mining Knowledge Discovery Tools, *SIGKDD Explorations*, 1, 1, pp. 20-33, 1999

[5] R. Grossman R, KDD-2003 Workshop on Data Mining Standsrds, Srevices and Platforms, *SIGKDD Explorations*, 5, 2, pp. 197, 2003

[6] J.H. Heindrics, J.S. Lim, Integrating Web-based Data Mining Tools With Business Models for Knowledge Management, *Decision Support Systems*, 35, 1, pp. 103-112, 2003

[7] M. Holsheimer, Data Mining by Business Users: Integrating Data Mining in Business Process, *Proceedings of International Conference on Knowledge Discovery and Data Mining*, pp. 266-291, 1999

[8] JSR-73 Expert Group, Java Specification Request 73: Java Data Mining (JDM), *Java Community Process*, 2004

[9] D. Mladenić, N. Lavrač, M. Bohanec, S. Moyle, *Data Mining and Decision Support: Integration and Collaboratio*, Dordrecht, Kluwer Academic Publishers, 2002

[10] C. Shearer, The CRISP-DM Model: The New Blueprint for Data Mining, *Journal of Data Warehousing*, 5, 4, pp. 13-22, 2000

[11] I.H. Witten, E. Frank, *Data Mining*, New York, Morgan Kaufman, 2000

[12] H.L. Lee, C. Park, Agent and Data Mining Based Decision Support System and its Adaption to a New Customer Centric Electronic Commerce, *Expert Systems With Applications*, 25, 4, pp. 619-635, 2003

[13] G.Polese, M.Troiano, G.Tortora, A Data Mining Based System Supporting Tactical Decisions, *Proceedings of the 14th International Conference on Software Engineering and Knowledge Engineering,* pp. 681-684, 2002

Rok Rupnik received his B.Sc, M.Sc. and Ph.D. degrees in Computer and Information Engineering from the University of Ljubljana in 1994, 1998 and 2002, respectively. Since 1994 he has been assistant-lecturer and senior-lecturer at the Laboratory of Information Systems of the Faculty of Computer and Information Science. His current research interests are data mining, intelligent agents and multi-agent systems, information systems strategic planning, information systems development methodologies and mobile applications. He is a member of IEEE, ACM and PMI.

Matjaž Kukar received his B.Sc., M.Sc. and Ph.D. degrees in Computer and Information Engineering from the University of Ljubljana in 1993, 1996 and 2001, respectively. He is currently Assistant Professor at the Faculty of Computer and Information Science of the same university. His research interests include machine learning, data mining and intelligent data analysis, ROC analysis, cost-sensitive learning, reliability estimation, and latent structure analysis, as well as applications of data mining in medical and business problems. He is a member of SLAIS and ECCAI.