

Oznaka poročila: ARRS-RPROJ-ZP-2013/17



ZAKLJUČNO POROČILO RAZISKOVALNEGA PROJEKTA

A. PODATKI O RAZISKOVALNEM PROJEKTU

1. Osnovni podatki o raziskovalnem projektu

Šifra projekta	J2-2285
Naslov projekta	Podatkovno rudarjenje za integrativno analizo podatkov v sistemski biologiji
Vodja projekta	11130 Sašo Džeroski
Tip projekta	J Temeljni projekt
Obseg raziskovalnih ur	4650
Cenovni razred	B
Trajanje projekta	05.2009 - 04.2012
Nosilna raziskovalna organizacija	106 Institut "Jožef Stefan"
Raziskovalne organizacije - soizvajalke	
Raziskovalno področje po šifrantu ARRS	2 TEHNIKA 2.07 Računalništvo in informatika 2.07.07 Inteligentni sistemi - programska oprema
Družbeno-ekonomski cilj	13.01 Naravoslovne vede - RiR financiran iz drugih virov (ne iz SUF)

2. Raziskovalno področje po šifrantu FOS¹

Šifra	1.02
- Veda	1 Naravoslovne vede
- Področje	1.02 Računalništvo in informatika

B. REZULTATI IN DOSEŽKI RAZISKOVALNEGA PROJEKTA

3. Povzetek raziskovalnega projekta²

SLO

Sistemska biologija je nova veja raziskav v znanostih o življenju, ki poskuša razumeti biološke organizme kot celoto. Preučuje procese v bioloških sistemih na vseh nivojih, od genoma do fenoma. Osredotoča se na integracijo podatkov dobljenih v podatkovno intenzivnih poskusih s področij genomike, transkriptomike, proteomike, metabolomike in fenomike.

Metode za rudarjenje podatkov iščejo vzorce v podatkih. Na področju sistemske biologije že igrajo pomembno vlogo saj so sposobne iz podatkov izluščiti nova znanja. Vseeno pa sistemska biologija pred podatkovno rudarjenje postavlja nove izzive: potreba po obravnavi kompleksnih podatkov, podatkov dobljenih iz heterogenih virov in uporabi domenskega predznanja. Poleg tega potrebujemo metode za odkrivanje znanja, ki podpirajo kompleksne scenarije odkrivanja znanja sestavljene iz več različnih nalog rudarjenja podatkov.

Na omenjene zahteve smo odgovorili z razvojem splošnega okvirja za rudarjenje podatkov, ki vključuje splošne algoritme za rudarjenje podatkov poljubne strukture in kompleksnosti. Okvir vključuje tudi ontologijo rudarjenja podatkov, ki pokriva različne naloge rudarjenja podatkov ter različne tipe podatkov in vzorcev, ter podpira sestavljanje posameznih operacij podatkovnega rudarjenja v proces odkrivanja znanja. Učinkovitost novo razvitega pristopa smo prikazali na več praktičnih problemih s področja medicinske sistemske biologije.

Intenzivno mednarodno sodelovanje, vključno z evropskima projektoma FP6 EETP (European Embryonal Tumor Pipeline) in FP7 PHAGOSYS (Systems Biology of Phagosome Formation and Maturation – Modulation by Intracellular Pathogens) nam je zagotovilo primerne praktične probleme in končne uporabnike ter predstavljalo dodatno motivacijo za razvoj predlaganih metod.

ANG

Systems biology is a new branch of the life sciences, which tries to understand biological organisms as a whole. It studies the processes within a biological system at all levels, from the genome to the phenome. It focuses on the integration of high-throughput data generated by genomics, transcriptomics, proteomics, metabolomics, and phenomics.

Data mining methods are concerned with finding patterns in data: They have the potential to discover knowledge in data and already play a very important role in systems biology. However, systems biology poses a number of challenges for current data mining: the need to handle complex data, data from heterogeneous sources, and domain knowledge. Knowledge discovery methods are also needed that support complex discovery processes involving multiple data mining tasks.

We addressed the above challenges by developing a unified framework for data mining that includes generic data mining algorithms for solving data mining tasks for arbitrarily structured/complex data. It also includes an ontology of data mining, covering a range of data mining tasks, data and patterns types, that facilitates the composition of data mining operations into knowledge discovery processes. We demonstrated the effectiveness of the developed framework on several practical problems from medical systems biology.

Strong European collaborations, including the FP6 project EETP (European Embryonal Tumor Pipeline), FP7 project PHAGOSYS (Systems Biology of Phagosome Formation and Maturation – Modulation by Intracellular Pathogens) provided the practical problems and end users, and motivated the development of the proposed methods.

4. Poročilo o realizaciji predloženega programa dela na raziskovalnem projektu³

Sistemska biologija je nova veja znanosti o življenju, ki biološke organizme obravnava kot celoto. Osredotoča se na integracijo podatkov dobljenih v podatkovno intenzivnih disciplinah kot so genomika in transkriptomika. Rudarjenje podatkov že igra pomembno vlogo na tem področju saj lahko iz podatkov izlušči nova znanja. V tem projektu smo razvili splošen okvir za rudarjenje podatkov, ki vključuje splošne algoritme za rudarjenje podatkov poljubne strukture in kompleksnosti (DS2) ter ontologijo rudarjenja podatkov (delovni sklop DS1), ki omogoča sestavljanje posameznih operacij podatkovnega rudarjenja v proces odkrivanja znanja (DS3). Razvite metode smo ovrednotili in prikazali njihovo uporabnost na več praktičnih problemih s področij sistemske biologije in ekologije (DS4).

Gradnje ontologije področja rudarjenja podatkov smo se lotili zaradi potrebe po poenotenju letega: Tovrstna ontologija nam poda besednjak jezika, ki bo omogočal sestavljanje posameznih operacij podatkovnega rudarjenja v proces odkrivanja znanja. Razvili smo ontologijo, poimenovano OntoDM, (Panov et al. 2009a, Panov et al 2010), ki vsebuje definicije osnovnih entitet oz. konceptov podatkovnega rudarjenja, kot npr. množica podatkov, tip podatkov, naloga podatkovnega rudarjenja, algoritem za podatkovno rudarjenje in njegovi sestavni deli (npr. funkcija razdalje). Omogoča nam tudi definicijo bolj zapletenih entitet, kot npr. scenarij

(zaporedje operacij) podatkovnega rudarjenja, rudarjenje strukturiranih podatkov in rudarjenja podatkov z omejitvami. Za razliko od obstoječih ontologij rudarjenja podatkov je OntoDM globoka ontologija in sledi primerom najboljših praks na področju načrtovanja ontologij.

Razvili smo vrsto metod za rudarjenje kompleksnih podatkov, predvsem metod za napovedovanje ciljnih spremenljivk s strukturiranimi vrednostmi ter metod za rudarjenje podatkov z omejitvami. Razvili smo metode za gradnjo dreves za napovedno razvrščanje, kjer so ciljne spremenljivke časovne vrste (Slavkov et al. 2010, Slavkov in Džeroski 2010) ali pa hierarhije (Schietgat et al. 2010, Vens et al 2010). Pri gradnji tovrstnih dreves lahko upoštevamo tudi različne vrste omejitev tako s stališča napovednega modeliranja kot tudi s stališča razvrščanja (Struyf in Džeroski 2010). Razvili smo tudi metodo za učenje ansamblov pravil za napovedovanje več zveznih spremenljivk hkrati (Aho et al. 2009). Razvili smo metode za rangiranje v kontekstu napovedovanja strukturiranih izhodov, kot npr. več ciljnih spremenljivk, metode za evalvacijo rangiranj in metode za agregacijo rangiranj (Slavkov in Džeroski 2010, Slavkov et al 2010a). Razvili smo tudi metode za učenje eno in večciljnih regresijskih dreves iz obsežnih podatkov in podatkovnih tokov (Ikonovska et al 2011ab), ki jo lahko uporabimo za analizo zelo velikih podatkovnih množic, kot so npr. podatki pridobljeni z različnimi hitropretočnimi tehnikami na področju sistemske biologije. Razvili smo nove metode za gradnjo dreves za napovedno razvrščanje, ki lahko upoštevajo avtokorelacijo v vrednostih zvezne ciljne spremenljivke (Stojanova et al. 2011cd). Metode najbližjih sosedov smo prilagodili za napovedovanje ciljnih spremenljivk s strukturiranimi vrednostmi (Pugelj in Džeroski 2011).

Razvite metode smo uporabili na vrsti problemov iz sistemske biologije oz. biomedicine in ekologije, tudi v okviru dveh mednarodnih projektov: V projektu PHAGOSYS (7. OP EU, zaključen 2012) smo se ukvarjali s sistemsko biologijo bolezni, ki manipulirajo človeški imunski sistem, projekt EETP (6OP EU, zaključen 2009) pa se je ukvarjal s proučevanjem embrionalnih tumorjev. Napovedovali smo funkcije genov (Vens et al 2010) in analizirali podatke iz mikromrež, še zlati časovnih vrst mikromrež ter mikromrež v povezavi s kliničnimi podatki (Slavkov in Džeroski 2010). Metode za rangiranje značilk smo uporabili za odkrivanje biomarkerjev pri embrionalnem tumorju (neuroblastoma) in pri nevrodegenerativni Huntingtonovi bolezni (Lovrečič et al 2010). Metode za razvrščanje časovnih vrst smo uporabili za identifikacijo skupin primerov s podobnim časovnim obnašanjem: S tem pristopom smo obravnavali odziv Schwannovih celic na okužbo z bakterijo *M. leprae* (Slavkov et al 2010b). Z razvitimi metodami smo analizirali različne vrste biomedicinskih slik. Obravnavali smo problem hierarhičnega označevanja medicinskih (rentgenskih) slik (Dimitrovski et al 2010abc, Dimitrovski et al 2011b). Analizirali smo tudi povezavo med lastnostmi posameznih genov in značilkami slik celičnih kultur, kjer so s tehnologijo siRNA bili izklopljeni ti geni (Džeroski in Kocev 2010, Kocev et al 2010b).

Metode razvite v okviru tega projekta smo nenačrtovano uporabili tudi na področju ekološkega modeliranja, in sicer za učenje habitatnih modelov skupnosti diatomej (Kocev et al. 2009a) in napovedovanje kemičnih parametrov kvalitete voda (Naumoski et al. 2009) v Prespanskem jezeru, za ocenjevanje stanja vegetacije (Kocev et al. 2009b) in za analizo časovnega poteka ekonomskih in ekoloških komponent vegetacije (Debeljak et al. 2009). Analizirali smo vpliv invazivnih vrst na strukturo biološke združbe (Everaert et al 2010, Aleksovski in Džeroski 2011, Everaert et al 2011, Keller et al. 2011) kot tudi časovni potek ekonomskih in ekoloških komponent vegetacije v kmetijstvu (Debeljak et al 2010). Iz daljinsko zaznanih podatkov smo ocenjevali tudi stanje slovenskih gozdov (Stojanova et al 2010). Zgradili smo tudi modele za ocenjevanje požarne ogroženosti naravnega okolja (Stojanova et al. 2011a). Na področju kmetijske ekologije smo analizirali časovni potek ekonomskih in ekoloških komponent vegetacije (Debeljak et al 2011a). Analizirali smo tudi vpliv uporabe odpadnih snovi kot gnojil v kmetijstvu (Cortet et al. 2011). Modelirali smo tudi problem koeksistence GSO in klasičnih poljščin in sicer z metodologijo relacijskih odločitvenih dreves (Debeljak et al. 2011b) ter z upoštevanjem prostorske avtokorelacije (Stojanova et al. 2011d).

Omeniti velja še vabljen predavanja neposredno povezana s tematiko projekta. Ta vključujejo predavanje na konferenci ICFA (9th International Conference on Formal Concept Analysis, Džeroski 2011a), predavanje na konferenci SMD (9th Congress of the Slovenian Biochemical Society, 5th Congress of the Slovenian Microbiological Society with International Participation, 3rd Central European Forum for Microbiology, Džeroski 2011b) in predavanje na delavnici MCP (Mining Complex Patterns, Džeroski 2011d). Na temo strojno učenje v sistemski biologiji sta bila tudi dva tutoriala (na zimski šoli FEBS Systems X Advanced Course, Džeroski 2011e, ter na konferenci ICSB, the 12th International Conference on Systems Biology, Džeroski 2011g).

Publikacije objavljene v okviru projekta podajamo v formatu (COBISS.SIID, referenca). Izvirni

znanstveni članki: (22691879, Kocev et al. 2009a), (22991399, Panov et al. 2009a), (23480359, Schietgat et al. 2010), (23488807, Slavkov et al. 2010), (23593255, Stojanova et al 2010), (24218407, Debeljak et al 2010), (24222503, Everaert et al 2010), (23122727, Kocev et al 2010a), (24222247, Lovrečić et al 2010), (24647719, Ikonovska et al 2011a), (24218407, Debeljak et al 2011a), (24559143, Stojanova et al 2011a), (25204775, Stojanova et al 2011b), (25200423, Stojanova et al 2011c), (24674087, Keller et al 2011), (25336615, Cortet et al 2011), (25200679, Pugelj and Džeroski 2011), (25192743, Madjarov et al 2011), (25233703, Dimitrovski et al 2011a), (24848423, Dimitrovski et al 2011b), (24222503, Everaert et al 2011). Objavljeni znanstveni prispevki na konferencah: (23322407, Aleksovski et al. 2009a), (23343399, Aho et al. 2009), (22691367, Naumoski et al. 2009), (22828071, Slavkov et al. 2009), (22847271, Panov et al. 2009b), (23322407, Aleksovski et al. 2009b), (24053287, Kocev et al 2010b), (24053543, Slavkov et al 2010a), (23950375, Dimitrovski et al 2010a), (24514343, Slavkov et al 2010b), (24367143, Dimitrovski et al 2010b), (24646439, Ikonovska et al 2011b), (24693287, Džeroski 2011a), (25110311, Ikonovska et al 2011b), (25335591, Gjorgjioski et al 2011). Objavljeni povzetki znanstvenih prispevkov na konferencah: (23328295, Slavkov in Džeroski 2009), (23021095, Atanasova et al. 2009), (23021351, Debeljak et al. 2009), (22828583, Ženko et al. 2009), (24312615, Džeroski in Kocev 2010), (23984935, Dimitrovski et al 2010c), (24223271, Debeljak et al 2009), (25397799, Schwartz et al 2011), (25547559, Džeroski 2011b), (24437031, Džeroski 2011c), (25549095, Džeroski 2011d), (25547815, Džeroski 2011e), (24797735, Aleksovski and Džeroski 2011), (24796967, Debeljak et al 2011b), (24796711, Stojanova et al 2011d). Samostojni znanstveni sestavki ali poglavja v monografski publikaciji: (24216615, Džeroski 2010a), (24216359, Panov et al 2010), (24216103, Struyf in Džeroski 2010), (24215847, Vens et al 2010), (24215335, Slavkov in Džeroski 2010), (24842535, Debeljak and Džeroski 2011). Predgovori in spremne besede: (24215591, Džeroski et al 2010a), (24052775, Džeroski et al 2010b), (24514087, Džeroski et al 2010c). Predavanja na tujih univerzah: (23237159, Džeroski et al. 2009c), (23235367, Džeroski et al. 2009d), (23235879, Džeroski et al. 2009e), (24368167, Džeroski 2010b), (24368423, Džeroski 2010c), (24367911, Džeroski 2010d), (24368935, Džeroski 2010e), (24369703, Džeroski 2010f), (24368679, Džeroski 2010g), (24438311, Džeroski 2011), (24438311, Džeroski 2011f). Vabljen predavanje na konferenci brez natisa: (24312871, Džeroski 2010h). Prispevek na konferenci brez natisa: (22780455, Škunca et al. 2009), (25548583, Džeroski 2011g). Mentorstvo pri doktorskih disertacijah: (255819008, Kocev 2011). Mentorstvo – drugo: (24645927, Ikonovska 2011). Komentorstvo pri doktorskih disertacijah: (25229351, Dimitrovski 2011), (5489249, Kobler 2011). Uredništvo: (22827559, Džeroski et al. 2009a), (22827815, Džeroski et al. 2009b), (24215079, Džeroski et al 2010d), (24049959, Džeroski et al 2010e), (24513831, Džeroski et al 2010f).

5. Ocena stopnje realizacije programa dela na raziskovalnem projektu in zastavljenih raziskovalnih ciljev⁴

V skladu s predlaganim programom smo zastavljene raziskovalne cilje dosegli v celoti.

6. Utemeljitev morebitnih sprememb programa raziskovalnega projekta oziroma sprememb, povečanja ali zmanjšanja sestave projektne skupine⁵

Med izvajanjem projekta je prišlo do manjših sprememb sestave projektne skupine. V projektini skupini sta tako dodatno sodelovala (z 0 obračunskimi urami) mlada raziskovalca Valentin Gjorgjioski (2010 in 2011) in Elena Ikonovska (2010-2012, kar ob prijavi projekta ni bilo predvideno).

7. Najpomembnejši znanstveni rezultati projektne skupine⁶

Znanstveni dosežek			
1.	COBISS ID	23488807	Vir: COBISS.SI
	Naslov	SLO	Metode za iskanje in razlago skupin genov s podobnimi časovnimi profili izraženosti
		ANG	Methods for finding and explaining groups of genes with similar expression timecourse profiles
			Razvili smo metodo za razvrščanje časovnih vrst (vrednosti zvezne spremenljivke) na osnovi dreves za napovedno razvrščanje. Metodo lahko

	Opis	SLO	uporabimo za identifikacijo skupin primerov s podobnim časovnim obnašanjem, za vsako skupino pa dobimo tudi njen opis. Metodo smo uporabili za identifikacijo skupin genov navadnega kvasa, ki se podobno odzivajo na različne tipe okoljskega stresa, dobljene skupine pa so opisane z oznakami iz ontologije GeneOntology.
		ANG	We have developed a method for predicting time series (values of a continuous variable), based on predictive clustering trees. The method can be used to identify groups of examples with similar temporal profiles and at the same time provides a description for each of the groups. We have used the method to identify groups of yeast genes that respond similarly to various kinds of environmental stress, and to explain the groups in terms of gene annotations with terms from the Gene Ontology.
	Objavljeno v	Royal Society of Chemistry; Molecular bioSystems; 2010; Vol. 6, no. 4; str. 729-740; Impact Factor: 3.825; Srednja vrednost revije / Medium Category Impact Factor: 3.787; WoS: CQ; Avtorji / Authors: Slavkov Ivica, Gjorgjioski Valentin, Struyf Jan, Džeroski Sašo	
	Tipologija	1.01 Izvirni znanstveni članek	
2.	COBISS ID	24216359	Vir: COBISS.SI
	Naslov	SLO	Ontologija podatkovnega rudarjenja
		ANG	A data mining ontology
	Opis	SLO	Razvili smo ontologijo podatkovnega rudarjenja OntoDM. Predstavlja entitete kot so podatki, naloge podatkovnega rudarjenja in algoritme ter generalizacije (na osnovi omenjenih entitet). Z ontologijo lahko opišemo več različnih pristopov podatkovnega rudarjenja, vključno z novjšimi pristopi za rudarjenje strukturiranih podatkov in rudarjenja podatkov z omejitvami. Za razliko od ostalih ontologij podatkovnega rudarjenja je OntoDM globoka ontologija in spoštuje najboljše prakse pri načrtovanju ontologij.
		ANG	We have developed the OntoDM ontology of data mining. It represents entities such as data, data mining tasks and algorithms, and generalizations (resulting from the latter). OntoDM covers much of the diversity in data mining research, including recently developed approaches to mining structured data and constraint-based data mining. In contrast to other ontologies of data mining, OntoDM is a deep ontology and is compliant to best practices in ontology engineering.
Objavljeno v	Springer; Inductive databases and constraint-based data mining; 2010; Str. 27-58; Avtorji / Authors: Panov Panče, Soldatova Larisa N., Džeroski Sašo		
Tipologija	1.16 Samostojni znanstveni sestavek ali poglavje v monografski publikaciji		
3.	COBISS ID	24647719	Vir: COBISS.SI
	Naslov	SLO	Učenje večciljnih dreves iz obsežnih podatkov in podatkovnih tokov
		ANG	Learning multi-target trees from massive or streaming data
	Opis	SLO	Razvili smo metode za učenje eno in večciljnih regresijskih dreves iz obsežnih podatkov in podatkovnih tokov. Kolikor nam je znano je to edina delujoča metoda za napovedovanje strukturiranih podatkov iz obsežnih podatkov ali podatkovnih tokov, ki jo lahko uporabimo za analizo zelo velikih podatkovnih množic, kot so npr. podatki pridobljeni z različnimi hitropretočnimi tehnikami na področju sistemske biologije.
		ANG	We have developed methods for learning trees for single and multi-target regression from massive or streaming data. To our knowledge, no other methods for structured prediction on streaming (or massive) data have been proposed so far. The methods can be used for analyzing very large

		datasets, such as those generated by highthroughput omics techniques in the area of systems biology.
	Objavljeno v	- IKONOMOVSKA, Elena, GAMA, João, DŽEROSKI, Sašo. Incremental multi target model trees for data streams. V: Proceedings of the 26th Annual ACM Symposium on Applied Computing 2011, Taichung, Taiwan, March 21 24, 2011. [S. l.]: ACM, cop. 2011, str. 988993. [COBISS.SIID 24646439] - Kluwer Academic Publishers; Data mining and knowledge discovery; 2011; Vol. 23, no. 1; str. 128-168; Impact Factor: 1.545; Srednja vrednost revije / Medium Category Impact Factor: 1.199; A': 1; WoS: EP, ET; Avtorji / Authors: Ikonomovska Elena, Gama Joao, Džeroski Sašo
	Tipologija	1.01 Izvirni znanstveni članek
4.	COBISS ID	26073895 Vir: COBISS.SI
	Naslov	<i>SLO</i> Avtokorelacija v napovednem razvrščanju <i>ANG</i> Autocorrelation in predictive clustering
	Opis	<i>SLO</i> Razvili smo metodo, ki eksplisitno obravnava prostorsko in omrežno avtokorelacijo v podatkih, ki niso neodvisni in enako porazdeljeni (i.i.d.) ter omogoča večnivojski vpogled v pojav avtokorelacije. Metoda temelji na konceptu dreves za napovedno razvrščanje (PCT) in se uporablja za različne naloge modeliranja, kot sta klasifikacija in regresija, kot tudi za nekatere naloge razvrščanja. To metodo smo uporabili na več realnih problemih prostorske regresije in klasifikacije, in pa na problemih omrežne regresije, ki prihajajo s področij socialnih in prostorskih omrežij. <i>ANG</i> We developed a method that explicitly takes into account spatial and network autocorrelation in data that are not independently and identically distributed (i.i.d.) and provides a multilevel insight into the autocorrelation phenomenon. The method is based on the concept of predictive clustering trees (PCTs) and works for different predictive modeling tasks, including classification and regression, as well as some clustering tasks. We applied this method to several real world problems of spatial regression and classification, as well as problems of network regression coming from the areas of social and spatial networks.
	Objavljeno v	- STOJANOVA, Daniela, CECI, Michelangelo, APPICE, Annalisa, MALERBA, Donato, DŽEROSKI, Sašo. Global and local spatial autocorrelation in predictive clustering trees. Lect. notes comput. sci., 2011, vol. 6926, str. 307322. [COBISS.SIID 25200423] Network regression with predictive clustering trees. - Kluwer; Proceedings of the ECML PKDD 2011, European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, 5-9 September 2011, Athens, Greece; Data mining and knowledge discovery; 2012; Vol. 25, no. 2; str. 378-413; Impact Factor: 1.545; Srednja vrednost revije / Medium Category Impact Factor: 1.199; A': 1; WoS: EP, ET; Avtorji / Authors: Stojanova Daniela, Ceci Michelangelo, Appice Annalisa, Džeroski Sašo
	Tipologija	1.01 Izvirni znanstveni članek
5.	COBISS ID	24848423 Vir: COBISS.SI
	Naslov	<i>SLO</i> Hierarhično označevanje medicinskih slik <i>ANG</i> Hierarchical annotation of medical images
	Opis	<i>SLO</i> Predlagali smo uporabo metode naključnih gozdov, sestavljenih iz dreves za napovedno razvrščanje, za označevanje medicinskih slik z značkami, ki so organizirane v hierarhijo. Poskusi kažejo, da so ansamblji dreves za napovedno razvrščanje bolj učinkoviti kot SVM modeli. Poleg tega smo ugotovili, da nam deskriptorji SIFT ponujajo najboljše razlikovanje med

		slikami, kombinacije različnih deskriptorjev pa izboljšujejo napovedno učinkovitost napovednih modelov. Rezultati dobljeni na obravnavani zbirki slik so boljši od doslej objavljenih rezultatov tako v literaturi kot na tekmovanjih v označevanju slik.
	ANG	We propose the use of random forests and bagging of predictive clustering trees in the domain of medical image annotation with labels organized into a hierarchy. The experiments show that ensembles of predictive clustering trees perform consistently better than SVMs. Second, SIFT descriptors are the most discriminative. Next, combinations of several descriptors improve the predictive performance of the classifiers. Finally, the results of the annotation of the considered image database are the best results reported so far both in the literature and at image annotation competitions.
Objavljeno v		Pergamon; Pattern recognition; 2011; Vol. 44, no. 10/11; str. 2436-2449; Impact Factor: 2.292; Srednja vrednost revije / Medium Category Impact Factor: 1.323; A': 1; WoS: EP, IQ; Avtorji / Authors: Dimitrovski Ivica, Kocev Dragi, Loskovska Suzana, Džeroski Sašo
Tipologija		1.01 Izvirni znanstveni članek

8. Najpomembnejši družbeno-ekonomski rezultati projektne skupine⁷

	Družbeno-ekonomski dosežek	
1.	COBISS ID	23480359 Vir: COBISS.SI
	Naslov	SLO Uporaba ansamblov dreves za hierarhično večznačkovno klasifikacijo za napovedovanje funkcije genov
		ANG Using ensembles of trees for hierarchical multilabel classification for predicting gene function
	Opis	SLO Razvili smo metodo za učenje ansamblov dreves za hierarhično večznačkovno klasifikacijo. Z njo smo napovedovali funkcijo genov pri treh različnih organizmih: <i>S. cerevisiae</i> , <i>A. thaliana</i> in <i>M. musculus</i> . Rezultati so pokazali, da je metoda enako točna kot trenutno najsodobnejše metode za avtomatsko določanje funkcij genom, hkrati pa časovno veliko bolj učinkovita.
		ANG We have developed a method for learning tree ensembles for hierarchical multi-label classification. We have used it for gene function prediction in three different organisms: <i>S. cerevisiae</i> , <i>A. thaliana</i> and <i>M. musculus</i> . The results show that our method is as accurate as state-of-the-art methods for automatic gene function prediction, but has a much lower time complexity.
	Šifra	F.02 Pridobitev novih znanstvenih spoznanj
	Objavljeno v	BioMed Central; BMC bioinformatics; 2010; Vol. 11, no. 2; str. 1-14; Impact Factor: 3.028; Srednja vrednost revije / Medium Category Impact Factor: 1.868; A': 1; WoS: CO, DB, MC; Avtorji / Authors: Schietgat Leander, Vens Celine, Struyf Jan, Blockeel Hendrik, Kocev Dragi, Džeroski Sašo
	Tipologija	1.01 Izvirni znanstveni članek
2.	COBISS ID	24215079 Vir: COBISS.SI
	Naslov	SLO Induktivne baze podatkov in rudarjenje podatkov z omejitvami
		ANG Inductive databases and constraint-based data mining
	Opis	SLO Uredili smo knjigo o induktivnih bazah podatkov in rudarjenju podatkov z omejitvami, ki predstavi omenjeno raziskovalno področje in poda pregled novjših raziskav. Induktivne baze so zelo pomembne za integrativno analizo podatkov na splošno in in še posebej za sistemsko biologijo. Poleg podatkov vsebujejo induktivne baze tudi vzorce, ki nastajajo ob izvedbi

		induktivnih povpraševanj. Napisali smo več poglavij, ki pokrivajo teme kot so ontologije za podatkovno rudarjenje, razvrščanje z omejitvami, napovedovanje funkcije genov in analiza podatkov iz mikromrež.
	ANG	We have edited a book on inductive databases and constraint-based data mining, which introduces this research area and gives an overview of recent research. Inductive databases are of great importance for the integrative analysis of data in general and for systems biology in particular. Besides data, inductive databases contain patterns, which are generated through inductive queries. The book contains several chapters authored by the team of this project, covering topics such as ontologies for data mining, constrained clustering, predicting gene function and analyzing micro-array data.
	Šifra	C.01 Uredništvo tujega/mednarodnega zbornika/knjige
	Objavljeno v	Springer; 2010; XVII, 456 str.; A": 1;A': 1; Avtorji / Authors: Džeroski Sašo, Goethals Bart, Panov Panče
	Tipologija	2.01 Znanstvena monografija
3.	COBISS ID	24222247 Vir: COBISS.SI
	Naslov	SLO Rangiranje značilk za odkrivanje biomarkerjev
		ANG Feature ranking for biomarker discovery
	Opis	SLO Razvili smo več metod za rangiranje značilk in jih uporabili na problemih odkrivanja biomarkerjev. Vključujejo metodo za rangiranje v kontekstu napovedovanja strukturiranih izhodov, kot npr. več ciljnih spremenljivk, metode za evalvacijo rangiranj in metode za agregacijo rangiranj. Omenjene metode smo uporabili za odkrivanje biomarkerjev pri embrionalnem tumorju (neuroblastoma) in pri nevrodegenerativni Huntingtonovi bolezni.
		ANG We have developed several methods for feature ranking and applied them to practical problems of biomarker discovery. These include a method for ranking in the context of predicting structured outputs, such as multiple targets, methods for evaluating rankings, and methods for aggregating rankings. We have applied these methods to discover biomarkers for neuroblastoma, a type of embrional tumors, and the neurodegenerative Huntington's disease.
	Šifra	F.02 Pridobitev novih znanstvenih spoznanj
	Objavljeno v	- I. SLAVKOV, S. DŽEROSKI. Analyzing gene expression data with predictive clustering trees. [COBISSID 24215335] - I. SLAVKOV, B. ŽENKO, S. DŽEROSKI. Evaluation method for feature rankings and their aggregations for biomarker discovery. [COBISSID 24514343] - Cambridge Medical Publications Ltd; Journal of international medical research; 2010; Vol. 38, no. 5; str. 1653-1662; Impact Factor: 1.068;Srednja vrednost revije / Medium Category Impact Factor: 2.875; WoS: QA, TU; Avtorji / Authors: Lovrečić Luca, Slavkov Ivica, Džeroski Sašo, Peterlin Borut
	Tipologija	1.01 Izvirni znanstveni članek
4.	COBISS ID	24513831 Vir: COBISS.SI
	Naslov	SLO Predsedovanje programskima odboroma in organizacija tretje in četrte mednarodne delavnice o strojnem učenju v sistemski biologiji (MLSB-09 in MLSB-10)
		ANG Chairing the program committees and organization of the Third and Fourth International Workshops on Machine Learning in Systems Biology (MLSB-

		09, -10)
Opis	SLO	Organizirali smo tretjo in četrto mednarodno delavnico o strojnem učenju v sistemski biologiji (MLSB-09 in MLSB-10) v Ljubljani in Edinburgu. Sopredsedovali smo programskima odboroma obeh dogodkov. Delavnica je dogodek z visokim ugledom in zelo kvalitetnimi vabljenimi predavanji ter recenziranimi prispevki, ki se ga udeleži več kot šestdeset udeležencev. Prispevki z vsake od delavnic so bili izdani v zborniku. Na delavnicah smo predstavili svoje rezultate dela na tem projektu ter na mednarodnih projektih EETP in PHAGOSYS.
	ANG	We organized the Third and Fourth Workshop on Machine Learning in Systems Biology (MLSB-09, MLSB-10) in Ljubljana and Edinburgh. We also co-chaired the program committees for both events. The workshop is a highly reputed event with high quality invited speakers and reviewed contributions that attracts more than 60 participants. Papers presented at the workshop were published in proceedings. At the workshops, we presented the results of our work in this project and the EU projects EETP and PHAGOSYS.
Šifra	B.02 Predsedovanje programskemu odboru konference	
Objavljeno v	- S. Džeroski, P. Geurts, J. Rousu. Machine learning in systems biology : proc. 3rd Int. Wshp. Helsinki Univ.: Department of Computer Science, 2009. [COBISS-ID 22827815]	
	- S. Džeroski, P. Geurts, J. Rousu. Proc. 3rd Int. Wshp. on Machine Learning in Systems Biology, (Journal of machine learning research, vol. 8, 2010). Cambridge: MIT Press, 2010. [COBISS-ID 24513831]	
	- S. Džeroski, S. Rogers, G. Sanguinetti. Machine learning in system biology : proc. 4th Int. Wshp, 2010. [COBISS-ID 24049959]	
Tipologija	2.31 Zbornik recenziranih znanstvenih prispevkov na mednarodni ali tuji konferenci	
5.	COBISS ID	24218407 Vir: COBISS.SI
Naslov	SLO	Analiza časovnih vrst podatkov o vegetacijskem agroekosistemu
	ANG	Analysis of time series data on agroecosystem vegetation
Opis	SLO	Z uporabo algoritma za k-medoidno razvrščanje in dinamičnega časovnega raztezanja med časovnimi vrstami, smo razvrstili časovne vrste zadržnosti tal z oljno ogrščico. Razvrščanje je oblikovalo pet tipičnih skupin časovnih vrst glede zadržnosti tal, ki se med seboj razlikujejo glede na stopnjo zadržnosti, časovno dinamiko zastiranja in dosežene največje vrednosti zadržnosti. Pokazalo se je, da so oblike časovnih profilov poljščine v veliki meri neodvisne od njenega tipa (jesenska/spomladanska oljčna ogrščica) in od načinov zatiranja plevelov. V nadaljevanju smo izdelali model razvrščanja (generalizirana oblika odločitvenega drevesa) za napovedovanje oblike časovne vrste zastiranja tal s pleveli (časovne vrste) s pomočjo neodvisnih (napovednih) spremenljivk, ki so vključevale obliko časovnega profila zadržnosti tal s poljščino, drugimi opisi značilnosti poljščine in okoljskih spremenljivk. Uporabljen pristop je ugotovil medsebojno odvisnost poljščine s plevelnimi vrstami vegetacije glede zastiranja tal.
	ANG	Using the k-medoids clustering algorithm and the dynamic time warping distance between time series, we clustered the time course profiles of oilseed rape cover crop. The clustering revealed five typical clusters of crop cover profiles that differed in terms of rate of increase, lag phase and maximum value, but were largely independent of the type of crop (winter/spring oil seed rape) and the weed management regime. We then constructed predictive clustering trees (a generalized form of decision trees) that predict the weed cover profile (time series) from independent

	(input) variables that include the crop cover cluster, other crops descriptors and environmental variables. The approach was successful in identifying the inter-dependencies between the weed and crop type of vegetation.
Šifra	F.02 Pridobitev novih znanstvenih spoznanj
Objavljeno v	Elsevier; Proceedigs of the Ecological modelling for enhanced sustainability in management, ISEM 2009, October 6-9, 2009, Québec, Canada; Ecological modelling; 2010; Vol. 222, no. 14; str. 2524-2529; Impact Factor: 1.769; Srednja vrednost revije / Medium Category Impact Factor: 2.52; WoS: GU; Avtorji / Authors: Debeljak Marko, Squire Geoff R., Kocev Dragi, Hawes Cathy, Young Marc W., Džeroski Sašo
Tipologija	1.01 Izvirni znanstveni članek

9. Drugi pomembni rezultati projektne skupine⁸

Projekt je izboljšal vidnost slovenskih znanstvenikov in Slovenije na specifičnih področjih, ki se jih dotika (strojno učenje, bioinformatika in sistemska biologija) kot tudi na širših znanstvenih področjih (informacijske tehnologije in biologija). Tukaj omenimo vabljena tutoriala na zimski šoli FEBS & System X Advanced Lecture Course on Systems Biology: From Molecules to Function, februarja 2011 v Innsbrucku, Avstrija ter na konferenci 12th International Conference on Systems Biology, avgusta 2011 v Heidelbergu, Nemčija. Vsakega od predavanj se je udeležilo več kot 50 poslušalcev, pomembna je pa tudi vidnost v okviru vsakega od teh izjemno odmevnih dogodkov.

Projekt je tudi vzpodbujal prenos znanja na področju sistemske biologije v Sloveniji. Poleg organizacije delavnice MLSB-09 v Ljubljani (ki je omenjena med dosežki zgoraj) je konkreten primer prenosa znanja vabljeno predavanje na skupnem kongresu 9th Congress of the Slovenian Biochemical Society, 5th Congress of the Slovenian Microbiological Society with International Participation, and 3rd Central European Forum for Microbiology, ki je potekal v Mariboru. Šlo je za ugledni mednarodni dogodek z vabljenimi predavatelji in recenziranimi prispevki visoke kvalitete, ki se ga je udeležilo veliko število poslušalcev iz Slovenije.

V času trajanja projekta je pod mentorstvom ali somentorstvom vodje projekta doktorske disertacije zagovarjalo 11 kandidatov. Disertacije so večinoma bile povezane s tematiko projekta. Iz njih je nastala vrsta publikacij, o katerih tukaj poročamo.

10. Pomen raziskovalnih rezultatov projektne skupine⁹

10.1. Pomen za razvoj znanosti¹⁰

SLO

Raziskave, ki so potekale v okviru tega projekta, so pomembne za razvoj več znanstvenih disciplin. Tako so najprej prispevale nova znanja na področju računalništva (informacijskih tehnologij), oz. bolj natančno na področju strojnega učenja in rudarjenja podatkov. Projekt je z razvojem metod za analizo kompleksnih strukturiranih podatkov ter razvojem ontologije podatkovnega rudarjenja za podporo kompleksnim postopkom odkrivanja znanja pomembno premaknil meje trenutno najsodobnejših raziskav.

Naše raziskave so pomembne tudi za znanstvene discipline, ki uporabljajo novo razvite metode. Predvsem tu mislimo na sistemska biologija, kjer je zelo očitna potreba po analizi kompleksnih strukturiranih podatkov, vseeno pa obstaja le malo metod za analizo takih podatkov. Z upoštevanjem teh potreb sistemske biologije naše raziskave pomembno vzpodbujajo njen razvoj.

Problem analize kompleksnih strukturiranih podatkov (iz heterogenih virov in ob upoštevanju domenskega predznanja) pa ni prisoten le v sistemski biologiji, ampak tudi na mnogih drugih znanstvenih področjih. Podatki o okolju lahko na primer zlahka dosežejo podobno kompleksnost kot podatki v sistemski biologiji. Uporabnost razvitih pristopov smo prikazali tudi na več problemih analize okoljskih podatkov. Metode, ki smo jih razvili v okviru tega projekta so tako uporabne tudi na drugih znanstvenih področjih, ki se srečujejo z analizo kompleksnih podatkov.

ANG

The research carried out within this project is relevant for the development of several scientific disciplines. First, it contributes to the field of computer science (information technologies) broadly speaking, and the discipline of machine learning and data mining more specifically. The project has moved well beyond the state of the art in this area, by developing methods for the analysis of complex structured data, as well as developing an ontology of data mining for supporting complex knowledge discovery processes.

Our research is also relevant for the scientific field within which we are applying the developed methods, namely systems biology. Even though the problem of complex structured data analysis is highly relevant, few effective approaches exist for mining such data. By addressing the pressing needs of systems biology, our research greatly facilitates its development.

The problem of analyzing complex structured data (from heterogeneous sources in the presence of domain knowledge) is not only present in systems biology, but also in many other scientific disciplines. Environmental data, for example, can easily reach complexity of the levels encountered in systems biology. We have demonstrated the usefulness of the developed approaches on a number of environmental data analysis problems as well. The methods developed in this project are also relevant to other scientific disciplines that encounter complex data analysis problems.

10.2. Pomen za razvoj Slovenije¹¹

SLO

Menimo, da imajo rezultati naših raziskav direkten vpliv na slovensko gospodarstvo in družbo na področju informacijskih in biotehnologij ter posredno na področju zdravja in trajnostnega razvoja. Bolj podrobno, na področju informacijskih tehnologij predvidevamo, da bi lahko razvite metode za analizo kompleksnih podatkov ob uporabi predznanja v obliki ontologij botrovale nastanku programskega produkta. Tega bi potencialno lahko tržili uporabnikom na več (tudi tehničnih) področjih. Informacijske tehnologije so močno zastopane v slovenski industriji, ki bi lahko rezultate projekta pretvorila v tržni produkt.

Razvite metode uporabljamo na področju sistemske biologije, tj. za integrativno analizo različnih podatkov o celičnih procesih. Genske regulacijske poti odkrite na ta način so lahko uporabljene za razvoj novih terapij za preučevane bolezni, kar je pomembno za farmacevtsko industrijo na sploh ter še posebej za slovensko farmacevtsko industrijo. Bolezni, ki smo jih preučevali, vključujejo embrionalne tumorje ter okužbe s salmonelo in tuberkulozo.

Projekt je izboljšal vidnost slovenskih znanstvenikov in Slovenije na specifičnih področjih, ki se jih dotika (strojno učenje, bioinformatika in sistemska biologija) kot tudi na širših znanstvenih področjih (informacijske tehnologije in biologija). Poleg tega je izboljšal

mednarodno sodelovanje slovenskih znanstvenikov, saj smo pri raziskavah sodelovali (in še naprej sodelujemo) s Katoliško univerzo Leuven v Belgiji, Univerzo v Portu na Portugalskem, Centrom za integrativno sistemsko biologijo na Imperial College v Londonu, Univerzitetnim medicinskim centrom v Leidnu in Inštitutom Max Planck za celično biologijo in genetiko iz Dresdna. Takšno sodelovanje vzpodbuja nastanek konzorcijev ter predlogov mednarodnih in evropskih projektov na obravnavanih področjih.

Projekt je spodbudil prenos znanja na področju sistemske biologije v Sloveniji. Sistemska biologija je porajajoče se raziskovalno področje, ki bo v prihodnosti pridobilo na pomembnosti. Konkreten primer prenosa znanja je bila organizacija delavnic MLSB 2009 (Third Workshop on Machine Learning in Systems Biology) v Ljubljani in MLSB 2010 (Fourth Workshop on Machine Learning in Systems Biology) v Edinburgu na Škotskem. To je ugledna mednarodna delavnica z vabljenimi predavatelji in recenziranimi prispevki visoke kvalitete, ki je vsakič pritegnila več kot 60 udeležencev.

Poleg tega je projekt prispeval k razvoju raziskovalcev na omenjenih področjih tako na doktorskem kot podoktorskem nivoju. Znanstvena vsebina projekta je povezana z vsebinami več predmetov, ki jih poučujemo na Mednarodni podiplomski šoli Jožefa Stefana v Ljubljani in na Univerzi v Novi Gorici. Raziskave so tako prispevale k nadaljnjemu razvoju in izboljšavam teh učnih predmetov.

ANG

We believe the results of our research have a direct impact on Slovenian economy and society in the areas of information and biotechnologies, as well as indirect impact in the areas of health and sustainable development. More specifically, in the area of information technologies it is conceivable that the developed methods for complex data analysis that use background knowledge written in the form of ontologies would give rise to a software product, which could be marketed to a potentially broad customer base in many disciplines (incl. various types of engineering). Slovenian industry has a strong IT sector that would be capable of turning the achieved research results into a commercial product.

The developed methods are also used in the area of systems biology, i.e., for integrative analysis of a variety of data about various processes at the cell level. The knowledge discovered in this fashion can be of use in the development of new therapies for the studied diseases, which would be relevant for the pharmaceutical industry in general and the Slovenian pharmaceutical industry in particular. Diseases that we studied include embryonal tumors, Salmonella infections, and tuberculosis.

The project promoted the visibility of Slovenian researchers and Slovenia in the specific research areas considered (machine learning, bioinformatics, and systems biology) and the corresponding wider scientific areas (information technology and biology). It also increased the international cooperation of Slovenian researchers, since our research was (and still is) performed in cooperation with the Katholieke Universiteit Leuven, Belgium, University of Porto, Portugal, the Center for Integrative Systems Biology at Imperial College, London, the Leiden University Medical Center, and the Max Planck Institute for Cell Biology and Genetics, Dresden. This facilitates the creation of consortia and formulation of international and European project proposals in the areas addressed by the project.

The project also facilitated the transfer of knowledge in the area of systems biology to Slovenia. Systems biology is an emerging research area that will receive increased attention over the coming years. A

concrete example of knowledge transfer was the organization of the MLSB 2009 and 2010 workshops (Third and Fourth Workshop on Machine Learning in Systems Biology) in Ljubljana and Edinburgh. This is a highly reputed workshop with high quality invited speakers and reviewed contributions that attracted more than 60 participants in each edition.

Finally, the project contributes to the development of researchers in its areas of interest, both at the PhD student and at the PostDoc level. The scientific content of the project is related to topics taught within several courses at the Jozef Stefan International Postgraduate School in Ljubljana, and the University of Nova Gorica. Its findings thus contribute to the further development and improvement of these courses.

**11. Samo za aplikativne projekte in podoktorske projekte iz gospodarstva!
Označite, katerega od navedenih ciljev ste si zastavili pri projektu, katere konkretne rezultate ste dosegli in v kakšni meri so doseženi rezultati uporabljeni**

Cilj		
F.01	Pridobitev novih praktičnih znanj, informacij in veščin	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.02	Pridobitev novih znanstvenih spoznanj	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.03	Večja usposobljenost raziskovalno-razvojnega osebja	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.04	Dvig tehnološke ravni	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.05	Sposobnost za začetek novega tehnološkega razvoja	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.06	Razvoj novega izdelka	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>

F.07	Izboljšanje obstoječega izdelka	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.08	Razvoj in izdelava prototipa	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.09	Razvoj novega tehnološkega procesa oz. tehnologije	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.10	Izboljšanje obstoječega tehnološkega procesa oz. tehnologije	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.11	Razvoj nove storitve	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.12	Izboljšanje obstoječe storitve	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.13	Razvoj novih proizvodnih metod in instrumentov oz. proizvodnih procesov	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.14	Izboljšanje obstoječih proizvodnih metod in instrumentov oz. proizvodnih procesov	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.15	Razvoj novega informacijskega sistema/podatkovnih baz	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>

	Uporaba rezultatov	<input type="text"/>
F.16	Izboljšanje obstoječega informacijskega sistema/podatkovnih baz	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.17	Prenos obstoječih tehnologij, znanj, metod in postopkov v prakso	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.18	Posredovanje novih znanj neposrednim uporabnikom (seminarji, forumi, konference)	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.19	Znanje, ki vodi k ustanovitvi novega podjetja ("spin off")	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.20	Ustanovitev novega podjetja ("spin off")	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.21	Razvoj novih zdravstvenih/diagnostičnih metod/postopkov	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.22	Izboljšanje obstoječih zdravstvenih/diagnostičnih metod/postopkov	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.23	Razvoj novih sistemskih, normativnih, programskih in metodoloških rešitev	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.24	Izboljšanje obstoječih sistemskih, normativnih, programskih in metodoloških rešitev	

	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.25	Razvoj novih organizacijskih in upravljavskih rešitev	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.26	Izboljšanje obstoječih organizacijskih in upravljavskih rešitev	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.27	Prispevek k ohranjanju/varovanje naravne in kulturne dediščine	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.28	Priprava/organizacija razstave	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.29	Prispevek k razvoju nacionalne kulturne identitete	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.30	Strokovna ocena stanja	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.31	Razvoj standardov	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.32	Mednarodni patent	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>

F.33	Patent v Sloveniji	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.34	Svetovalna dejavnost	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>
F.35	Drugo	
	Zastavljen cilj	<input type="radio"/> DA <input type="radio"/> NE
	Rezultat	<input type="text"/>
	Uporaba rezultatov	<input type="text"/>

Komentar

12.Samo za aplikativne projekte in podoktorske projekte iz gospodarstva!
 Označite potencialne vplive oziroma učinke vaših rezultatov na navedena področja

	Vpliv	Ni vpliva	Majhen vpliv	Srednji vpliv	Velik vpliv	
G.01	Razvoj visokošolskega izobraževanja					
G.01.01.	Razvoj dodiplomskega izobraževanja	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.01.02.	Razvoj podiplomskega izobraževanja	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.01.03.	Drugo: <input type="text"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02	Gospodarski razvoj					
G.02.01	Razširitev ponudbe novih izdelkov/storitev na trgu	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02.02.	Širitev obstoječih trgov	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02.03.	Znižanje stroškov proizvodnje	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02.04.	Zmanjšanje porabe materialov in energije	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02.05.	Razširitev področja dejavnosti	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02.06.	Večja konkurenčna sposobnost	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02.07.	Večji delež izvoza	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02.08.	Povečanje dobička	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02.09.	Nova delovna mesta	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02.10.	Dvig izobrazbene strukture zaposlenih	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02.11.	Nov investicijski zagon	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.02.12.	Drugo: <input type="text"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.03	Tehnološki razvoj					

G.03.01.	Tehnološka razširitev/posodobitev dejavnosti	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.03.02.	Tehnološko prestrukturiranje dejavnosti	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.03.03.	Uvajanje novih tehnologij	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.03.04.	Drugo:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.04	Družbeni razvoj					
G.04.01	Dvig kvalitete življenja	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.04.02.	Izboljšanje vodenja in upravljanja	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.04.03.	Izboljšanje delovanja administracije in javne uprave	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.04.04.	Razvoj socialnih dejavnosti	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.04.05.	Razvoj civilne družbe	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.04.06.	Drugo:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.05.	Ohranjanje in razvoj nacionalne naravne in kulturne dediščine in identitete	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.06.	Varovanje okolja in trajnostni razvoj	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.07	Razvoj družbene infrastrukture					
G.07.01.	Informacijsko-komunikacijska infrastruktura	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.07.02.	Prometna infrastruktura	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.07.03.	Energetska infrastruktura	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.07.04.	Drugo:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.08.	Varovanje zdravja in razvoj zdravstvenega varstva	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
G.09.	Drugo:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	

Komentar

--

13.Pomen raziskovanja za sofinancerje¹²

	Sofinancer	
1.	Naziv	
	Naslov	
	Vrednost sofinanciranja za celotno obdobje trajanja projekta je znašala:	EUR
	Odstotek od utemeljenih stroškov projekta:	%
	Najpomembnejši rezultati raziskovanja za sofinancerja	Šifra
	1.	
	2.	
3.		
4.		

	5.	
Komentar		
Ocena		

14. Izjemni dosežek v letu 2012¹³

14.1. Izjemni znanstveni dosežek

Razvili smo metodo, ki eksplicitno obravnava avtokorelacijo v podatkih, ki niso neodvisni in enako porazdeljeni. Metoda gradi drevesa za napovedno razvrščanje in se lahko uporablja za različne naloge modeliranja, kot so klasifikacija, regresija in razvrščanje v skupine. Pri gradnji modelov uporablja opisne podatke (lastnosti vozlišča) in strukturo omrežja, medtem ko pri njihovi uporabi upošteva samo opisne podatke. Naučene modele tako lahko uporabimo tudi v primerih, ko je omrežna struktura okrog (novih) testnih primerov neznana ali nezanesljiva. Naučeni modeli se prilagajajo lokalnim lastnostim podatkov ter hkrati zagotavljajo lokalno gladke napovedi. Metodo smo uspešno uporabili na več realnih problemih prostorske regresije in klasifikacije ter na problemih omrežne regresije v socialnih in prostorskih omrežjih.

Objavljeno v: D. Stojanova, M. Ceci, A. Appice, S. Džeroski. Network regression with predictive clustering trees. *Data Mining and Knowledge Discovery*, 25(2):378-413, 2012.

14.2. Izjemni družbeno-ekonomski dosežek

Predlagali smo uporabo ansamblov dreves za napovedno razvrščanje za avtomatsko klasifikacijo kremenastih alg. Iz slik, zajetih pod mikroskopom, najprej izlučimo vrsto značilk. Na osnovi značilk in pravilnih uvrstitev alg v ustrezno taksonomsko kategorijo (za slike iz učne množice), se naučimo modelov za hierarhično klasifikacijo oz. uvrščanje alg.

Predlagano metodologijo smo ovrednotili na tim. ADIAC zbirki slik kremenastih alg. Poskusi kažejo, da so ansambli dreves za napovedno razvrščanje bolj učinkoviti kot konkurenčni SVM modeli. Poleg tega smo ugotovili, da nam deskriptorji SIFT ponujajo najboljše razlikovanje med slikami, kombinacije različnih deskriptorjev pa izboljšujejo napovedno učinkovitost napovednih modelov. Naši rezultati so boljši od vseh doslej objavljenih rezultatov.

Objavljeno v: I. Dimitrovski, D. Kocev, S. Loškova, S. Džeroski. Hierarchical classification of diatom images using ensembles of predictive clustering trees. *Ecological Informatics*, 7:19-29, 2012.

C. IZJAVE

Podpisani izjavljam/o, da:

- so vsi podatki, ki jih navajamo v poročilu, resnični in točni
- se strinjamo z obdelavo podatkov v skladu z zakonodajo o varstvu osebnih podatkov za potrebe ocenjevanja ter obdelavo teh podatkov za evidence ARRS
- so vsi podatki v obrazcu v elektronski obliki identični podatkom v obrazcu v pisni obliki
- so z vsebino zaključnega poročila seznanjeni in se strinjajo vsi soizvajalci projekta

Podpisi:

*zastopnik oz. pooblaščen oseba
raziskovalne organizacije:*

in

vodja raziskovalnega projekta:

Institut "Jožef Stefan"

Sašo Džeroski

ŽIG

Kraj in datum:

Ljubljana	14.3.2013
-----------	-----------

Oznaka prijave: ARRS-RPROJ-ZP-2013/17

¹ Opredelite raziskovalno področje po klasifikaciji FOS 2007 (Fields of Science). Prevajalna tabela med raziskovalnimi področji po klasifikaciji ARRS ter po klasifikaciji FOS 2007 (Fields of Science) s kategorijami WOS (Web of Science) kot podpodročji je dostopna na spletni strani agencije (<http://www.arrs.gov.si/sl/gradivo/sifranti/preslik-vpp-fos-wos.asp>). [Nazaj](#)

² Napišite povzetek raziskovalnega projekta (največ 3.000 znakov v slovenskem in angleškem jeziku) [Nazaj](#)

³ Napišite kratko vsebinsko poročilo, kjer boste predstavili raziskovalno hipotezo in opis raziskovanja. Navedite ključne ugotovitve, znanstvena spoznanja, rezultate in učinke raziskovalnega projekta in njihovo uporabo ter sodelovanje s tujimi partnerji. Največ 12.000 znakov vključno s presledki (približno dve strani, velikost pisave 11). [Nazaj](#)

⁴ Realizacija raziskovalne hipoteze. Največ 3.000 znakov vključno s presledki (približno pol strani, velikost pisave 11) [Nazaj](#)

⁵ V primeru bistvenih odstopanj in sprememb od predvidenega programa raziskovalnega projekta, kot je bil zapisan v predlogu raziskovalnega projekta oziroma v primeru sprememb, povečanja ali zmanjšanja sestave projektne skupine v zadnjem letu izvajanja projekta, napišite obrazložitev. V primeru, da sprememb ni bilo, to navedite. Največ 6.000 znakov vključno s presledki (približno ena stran, velikost pisave 11). [Nazaj](#)

⁶ Navedite znanstvene dosežke, ki so nastali v okviru tega projekta. Raziskovalni dosežek iz obdobja izvajanja projekta (do oddaje zaključnega poročila) vpišete tako, da izpolnite COBISS kodo dosežka – sistem nato sam izpolni naslov objave, naziv, IF in srednjo vrednost revije, naziv FOS področja ter podatek, ali je dosežek uvrščen v A'' ali A'. [Nazaj](#)

⁷ Navedite družbeno-ekonomske dosežke, ki so nastali v okviru tega projekta. Družbeno-ekonomski rezultat iz obdobja izvajanja projekta (do oddaje zaključnega poročila) vpišete tako, da izpolnite COBISS kodo dosežka – sistem nato sam izpolni naslov objave, naziv, IF in srednjo vrednost revije, naziv FOS področja ter podatek, ali je dosežek uvrščen v A'' ali A'.

Družbeno-ekonomski dosežek je po svoji strukturi drugačen kot znanstveni dosežek. Povzetek znanstvenega dosežka je praviloma povzetek bibliografske enote (članka, knjige), v kateri je dosežek objavljen.

Povzetek družbeno-ekonomskega dosežka praviloma ni povzetek bibliografske enote, ki ta dosežek dokumentira, ker je dosežek sklop več rezultatov raziskovanja, ki je lahko dokumentiran v različnih bibliografskih enotah. COBISS ID zato ni enoznačen, izjemoma pa ga lahko tudi ni (npr. prehod mlajših sodelavcev v gospodarstvo na pomembnih raziskovalnih nalogah, ali ustanovitev podjetja kot rezultat projekta ... - v obeh primerih ni COBISS ID). [Nazaj](#)

⁸ Navedite rezultate raziskovalnega projekta iz obdobja izvajanja projekta (do oddaje zaključnega poročila) v primeru, da katerega od rezultatov ni mogoče navesti v točkah 7 in 8 (npr. ker se ga v sistemu COBISS ne vodi). Največ 2.000 znakov, vključno s presledki. [Nazaj](#)

⁹ Pomen raziskovalnih rezultatov za razvoj znanosti in za razvoj Slovenije bo objavljen na spletni strani: <http://sicris.izum.si/> za posamezen projekt, ki je predmet poročanja [Nazaj](#)

¹⁰ Največ 4.000 znakov, vključno s presledki [Nazaj](#)

¹¹ Največ 4.000 znakov, vključno s presledki [Nazaj](#)

¹² Rubrike izpolnite / prepisite skladno z obrazcem "izjava sofinancerja" <http://www.arrs.gov.si/sl/progproj/rproj/gradivo/>, ki ga mora izpolniti sofinancer. Podpisan obrazec "Izjava sofinancerja" pridobi in hrani nosilna raziskovalna organizacija – izvajalka projekta. [Nazaj](#)

¹³ Navedite en izjemni znanstveni dosežek in/ali en izjemni družbeno-ekonomski dosežek raziskovalnega projekta v letu 2012 (največ 1000 znakov, vključno s presledki). Za dosežek pripravite diapozitiv, ki vsebuje sliko ali drugo slikovno gradivo v zvezi z izjemnim dosežkom (velikost pisave najmanj 16, približno pol strani) in opis izjemnega dosežka (velikost pisave 12, približno pol strani). Diapozitiv/-a priložite kot priložitev/-i k temu poročilu. Vzorec diapozitiva je objavljen na spletni strani ARRS <http://www.arrs.gov.si/sl/gradivo/>, predstavitev dosežkov za pretekla leta pa so objavljena na spletni strani <http://www.arrs.gov.si/sl/analize/dosez/>. [Nazaj](#)

Obrazec: ARRS-RPROJ-ZP/2013 v1.00
BC-90-AD-2E-C3-7F-91-25-7D-2F-26-E4-B2-2B-90-B6-0A-DF-1F-3C

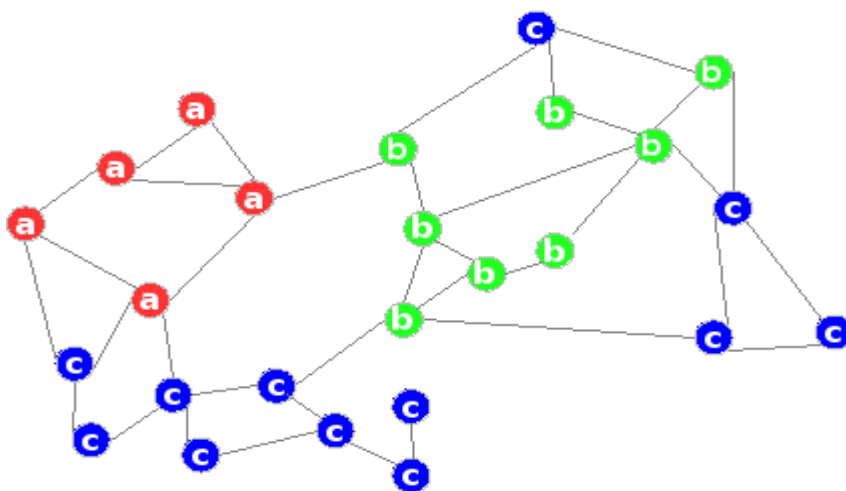
TEHNIKA

Področje: 2.07 – Računalništvo in informatika

Dosežek: Avtokorelacija v napovednem razvrščanju

Vir: D. Stojanova, M. Ceci, A. Appice, S. Džeroski. Network regression with predictive clustering trees. *Data Mining and Knowledge Discovery*, 25(2):378-413, 2012, doi: 10.1007/s10618-012-0278-6.

Avtokorelacija v omrežnih podatkih, kjer imajo povezana vozlišča pogosto enake oznake. Različne oznake vozlišč so predstavljene z različnimi barvami.



Razvili smo metodo, ki eksplicitno obravnava avtokorelacijo v podatkih, ki niso neodvisni in enako porazdeljeni. Metoda gradi drevesa za napovedno razvrščanje in se lahko uporablja za različne naloge modeliranja, kot so klasifikacija, regresija in razvrščanje v skupine.

Pri gradnji modelov uporablja metoda opisne podatke (lastnosti vozlišča) in strukturo omrežja, medtem ko pri njihovi uporabi upošteva samo opisne podatke. Naučene modele tako lahko uporabimo tudi v primerih, ko je omrežna struktura okrog (novih) testnih primerov neznana ali nezanesljiva. Naučeni modeli se prilagajajo lokalnim lastnostim podatkov ter hkrati zagotavljajo lokalno gladke napovedi.

Metodo smo uspešno uporabili na več realnih problemih prostorske regresije in klasifikacije ter na problemih omrežne regresije v socialnih in prostorskih omrežjih.




TEHNIKA

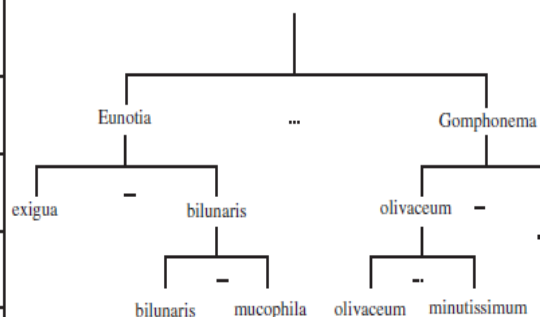
Področje: 2.07 – Računalništvo in informatika

Dosežek : **Avtomska klasifikacija kremenastih alg**

Vir: I. Dimitrovski, D. Kocev, S. Loškovska, S. Džeroski. Hierarchical classification of diatom images using ensembles of predictive clustering trees. *Ecological Informatics*, 7:19-29, 2012, doi: 10.1016/j.ecoinf.2011.09.001

Na osnovi slik, zajetih pod mikroskopom, kremenaste alge avtomatsko uvrščamo v taksonomsko drevo. Za ta namen uporabimo ansamble dreves za napovedno razvrščanje in sicer naključne gozdove dreves za hierarhično klasifikacijo..

image	features/descriptors						taxonomy
	Fourier coefficients						
	48	24	59	66	37	...	olivaceum
	36	25	53	45	15	...	minutissimum
	35	25	56	52	19		exigua
...



Predlagali smo uporabo ansamblov dreves za napovedno razvrščanje za avtomatsko klasifikacijo kremenastih alg. Iz slik, zajetih pod mikroskopom, najprej izluščimo vrsto značilk. Na osnovi značilk in pravih uvrstitev alg v ustrezno taksonomsko kategorijo (za slike iz učne množice), se naučimo modelov za hierarhično klasifikacijo oz. uvrščanje alg.

Predlagano metodologijo smo ovrednotili na tim. ADIAC zbirki slik kremenastih alg. Poskusi kažejo, da so ansambli dreves za napovedno razvrščanje bolj učinkoviti kot konkurenčni SVM modeli. Poleg tega smo ugotovili, da nam deskriptorji SIFT ponujajo najboljše razlikovanje med slikami, kombinacije različnih deskriptorjev pa izboljšujejo napovedno učinkovitost napovednih modelov.

Naši rezultati na omenjeni zbirki so boljši od vseh doslej objavljenih rezultatov in sicer tako po klasifikacijski točnosti kot tudi po številu upoštevanih taksonomskih kategorij.