

Table of Contents

Keynote Talks

Emergent Dynamics of Information Propagation in Large Networks	1
<i>Katia Sycara</i>	
New Applications and Theoretical Foundations of the Dominance-based Rough Set Approach	2
<i>Roman Słowiński</i>	

RSCTC 2010 Discovery Challenge

RSCTC'2010 Discovery Challenge: Mining DNA Microarray Data for Medical Diagnosis and Treatment	4
<i>Marcin Wojnarski, Andrzej Janusz, Hung Son Nguyen, Jan Bazan, ChuanJiang Luo, Ze Chen, Feng Hu, Guoyin Wang, Lihe Guan, Huan Luo, Juan Gao, Yuanxia Shen, Vladimir Nikulin, Tian-Hsiang Huang, Geoffrey J. McLachlan, Matko Bošnjak, Dragan Gamberger</i>	
TunedIT.org: System for Automated Evaluation of Algorithms in Repeat- able Experiments	20
<i>Marcin Wojnarski, Sebastian Stawicki, Piotr Wojnarowski</i>	

Clustering

Consensus Multiobjective Differential Crisp Clustering for Categorical Data Analysis	30
<i>Indrajit Saha, Dariusz Plewczyński, Ujjwal Maulik, Sanghamitra Bandyopadhyay</i>	
Probabilistic Rough Entropy Measures in Image Segmentation	40
<i>Dariusz Matyszko, Jarosław Stepaniuk</i>	
Distance Based Fast Hierarchical Clustering Method for Large Datasets . .	50
<i>Bidyut Kr. Patra, Neminath Hubballi, Santosh Biswas, Sukumar Nandi</i>	
TI-DBSCAN: Clustering with DBSCAN by Means of the Triangle Inequality	60
<i>Marzena Kryszkiewicz, Piotr Lasek</i>	

Multimedia and Telemedicine: Soft Computing Applications

Vehicle Classification Based on Soft Computing Algorithms	70
<i>Piotr Dalka, Andrzej Czyżewski</i>	

Controlling Computer by Lip Gestures Employing Neural Networks	80
<i>Piotr Dalka, Andrzej Czyżewski</i>	
Computer Animation System Based on Rough Sets and Fuzzy Logic	90
<i>Piotr Szczuko</i>	
Adaptive Phoneme Alignment based on Rough Set Theory	100
<i>Konstantinos Avdelidis, Charalampos Dimoulas, George Kalliris, George Papanikolaou</i>	
Monitoring Parkinson’s Disease Patients Employing Biometric Sensors and Rule-Based Data Processing	110
<i>Paweł Żwan, Katarzyna Kaszuba, Bożena Kostek</i>	
Content-Based Scene Detection and Analysis Method for Automatic Classification of TV Sports News	120
<i>Kazimierz Choroś, Piotr Pawlaczyk</i>	
Combined Learning Methods and Mining Complex Data	
Combining Multiple Classification or Regression Models Using Genetic Algorithms	130
<i>Andrzej Janusz</i>	
Argument Based Generalization of MODLEM Rule Induction Algorithm	138
<i>Krystyna Napierała, Jerzy Stefanowski</i>	
Integrating Selective Pre-processing of Imbalanced Data with Ivotes Ensemble	148
<i>Jerzy Błaszczyński, Magdalena Deckert, Jerzy Stefanowski, Szymon Wilk</i>	
Learning from Imbalanced Data in Presence of Noisy and Borderline Examples	158
<i>Krystyna Napierała, Jerzy Stefanowski, Szymon Wilk</i>	
Tracking Recurrent Concepts using Context	168
<i>João Bártolo Gomes, Ernestina Menasalvas, Pedro A. C. Sousa</i>	
Support Feature Machine for DNA Microarray Data	178
<i>Tomasz Maszczyk, Włodzisław Duch</i>	
Is It Important which Rough-Set-Based Classifier Extraction and Voting Criteria are Applied Together?	187
<i>Dominik Ślęzak, Sebastian Widz</i>	
Improving Co-training with Agreement-based Sampling	197
<i>Jin Huang, Jelber Sayyad Shirabad, Stan Matwin, Jiang Su</i>	

Experienced Physicians and Automatic Generation of Decision Rules from Clinical Data	207
<i>William Klement, Szymon Wilk, Martin Michalowski and Ken Farion</i>	

Gene-Pair Representation and Incorporation of GO-based Semantic Similarity into Classification of Gene Expression Data	217
<i>Torsten Schön, Alexey Tsymbal, Martin Huber</i>	

Rough Sets: Logical and Mathematical Foundations

A Fuzzy View on Rough Satisfiability	227
<i>Anna Gomolińska</i>	

Rough Sets in Terms of Discrete Dynamical Systems	237
<i>Marcin Wolski</i>	

A Preference-based Multiple-Source Rough Set Model	247
<i>Md. Aquil Khan, Mohua Banerjee</i>	

Classification of Dynamics in Rough Sets	257
<i>Davide Ciucci</i>	

Relational Granularity for Hypergraphs	267
<i>John G. Stell</i>	

Perceptual Tolerance Intersection	277
<i>Piotr Wasilewski, James F. Peters, Sheela Ramanna</i>	

Rough Approximations: Foundations and Methodologies

Categories of Direlations and Rough Set Approximation Operators	287
<i>Murat Diker</i>	

Approximations and Classifiers	297
<i>Andrzej Skowron, Jarosław Stepaniuk</i>	

A Note on a Formal Approach to Rough Operators	307
<i>Adam Grabowski, Magdalena Jastrzębska</i>	

Communicative Approximations as Rough Sets	317
<i>Mohua Banerjee, Abhinav Pathak, Gopal Krishna, Amitabha Mukerjee</i>	

On the Correctness of Rough-Set Based Approximate Reasoning	327
<i>Patrick Doherty, Andrzej Szalas</i>	

Unit Operations in Approximation Spaces	337
<i>Zbigniew Bonikowski</i>	

Machine Learning: Methodologies and Algorithms

VIII

Weighted Nearest Neighbor Classification via Maximizing Classification Consistency	347
<i>Pengfei Zhu, Qinghua Hu, Yongbin Yang</i>	
Rough Set-Based Incremental Learning Approach to Face Recognition	356
<i>Xuguang Chen, Wojciech Ziarko</i>	
A Comparison of Dynamic and Static Belief Rough Set Classifier	366
<i>Salsabil Trabelsi, Zied Elouedi, Pawan Lingras</i>	
Rule Generation in Lipski's Incomplete Information Databases	376
<i>Hiroshi Sakai, Michinori Nakata, Dominik Ślęzak</i>	
A Fast Randomisation Test for Rule Significance	386
<i>Ivo Düntsch, Günther Gediga</i>	
Ordinal Classification with Monotonicity Constraints by Variable Consistency Bagging	392
<i>Jerzy Błaszczyński, Roman Słowiński, Jerzy Stefanowski</i>	
Learnability in Rough Set Approaches	402
<i>Jerzy Błaszczyński, Roman Słowiński, Marcin Szelaż</i>	
Upper Bounds on Minimum Cardinality of Exact and Approximate Reducts	412
<i>Igor Chikalov, Mikhail Moshkov, Beata Zielosko</i>	
An Extension of Rough Set Approximation to Flow Graph Based Data Analysis	418
<i>Doungnat Chitcharoen, Puntip Pattaraintakorn</i>	
Credibility Coefficients Based on SVM	428
<i>Roman Podraza, Bartosz Janeczek</i>	
On Algorithm for Building of Optimal α -Decision Trees	438
<i>Abdulaziz Alkhalid, Igor Chikalov, Mikhail Moshkov</i>	
Layered Approximation Approach to Knowledge Elicitation in Machine Learning	446
<i>Tuan Trung Nguyen</i>	
Multiagent Systems	
Configuration Management of Mobile Agents Based on SNMP	456
<i>Michał Komorowski</i>	
Adaptive Immunity-based Multiagent Systems (AIBMAS) Inspired by the Idiotypic Network	466
<i>Chung-Ming Ou, C.R. Ou</i>	

Distributed Default Logic for Context-aware Computing in Multi-Agent Systems	476
<i>Dominik Ryżko, Henryk Rybiński</i>	

A Novel Approach to Default Reasoning for MAS	484
<i>Przemysław Więch, Henryk Rybiński</i>	

A Platform for the Evaluation of Automated Argumentation Strategies ...	494
<i>Piotr S. Kośmicki</i>	

Emerging Intelligent Technologies and Net-Centric Applications

Fuzzy Similarity-Based Relative Importance of MPEG-7 Visual Descriptors for Emotional Classification of Images	504
<i>EunJong Park, SungHwan Jeong, JoonWhoan Lee</i>	

The Impact of Recommendation Sources on the Adoption Intention of Microblogging based on Dominance-based Rough Set Approach	514
<i>Yang-Chieh, Chaio-Chen Chang, Chiun-Sin Lin, Gwo-Hshiung Tzeng</i>	

Fault Effects Analysis and Reporting System for Dependability Evaluation	524
<i>Piotr Gawkowski, Monika Anna Kuczyńska, Agnieszka Komorowska</i>	

Solving the Reporting Cells Problem Using a Scatter Search Based Algorithm	534
<i>Sónia M. Almeida-Luz, Miguel A. Vega-Rodríguez, Juan A. Gómez-Pulido, Juan M. Sánchez-Pérez</i>	

Learning Age and Gender using Co-occurrence of Non-Dictionary Words from Stylistic Variations	544
<i>Rajendra Prasath</i>	

Disturbance Measurement Utilization in Easily Reconfigurable Fuzzy Predictive Controllers: Sensor Fault Tolerance and Other Benefits	551
<i>Piotr M. Marusak</i>	

Classification and Decision Support Applications

Biometric-Based Authentication System using Rough Set Theory	560
<i>Hala S. Own, Waheeda Al-Mayyan, Hussein Zedan</i>	

Classification of Facial Photograph Sorting Performance Based on Verbal Descriptions	570
<i>Daryl H. Hepting, Richard Spring, Timothy Maciag, Katherine Arbuthnott, Dominik Słęzak</i>	

Random Musical Bands Playing in Random Forests	580
<i>Miron B. Kursa, Elżbieta Kubera, Witold R. Rudnicki, Alicja A. Wieczorkowska</i>	

An Empirical Comparison of Rule Sets Induced by LERS and Probabilistic Rough Classification	590
<i>Jerzy W. Grzymala-Busse, Shantanu R. Marepally, Yiyu Yao</i>	
DRSA Decision Algorithm Analysis in Stylometric Processing of Literary Texts	600
<i>Urszula Stańczyk</i>	
Blind Music Timbre Source Isolation by Multiresolution Comparison of Spectrum Signatures	610
<i>Xin Zhang, Wenxin Jiang, Zbigniew W. Raś, Rory Lewis</i>	
Rough Sets for Solving Classification Problems in Computational Neuroscience	620
<i>Tomasz G. Smolinski, Astrid A. Prinz</i>	
Towards Approximate SQL – Infobright’s Approach	630
<i>Dominik Ślęzak, Marcin Kowalski</i>	
A Protein Classifier Based on SVM by Using the Voxel Based Descriptor	640
<i>Georgina Mirceva, Andreja Naumoski, Danco Davcev</i>	

Intelligent Methods in Optimization and Control

Explicit Neural Network-Based Nonlinear Predictive Control with Low Computational Complexity	649
<i>Maciej Ławryńczuk</i>	
Solution of the Inverse Heat Conduction Problem by Using the ABC Algorithm	659
<i>Edyta Hetmaniok, Damian Słota, Adam Zielonka</i>	
Application of Fuzzy Wiener Models in Efficient MPC Algorithms	669
<i>Piotr M. Marusak</i>	
Multicriteria Subjective Reputation Management Model	678
<i>Michał Majdan, Włodzimierz Ogryczak</i>	
Application of Fuzzy Preference Based Rough Set Model to Condition Monitoring	688
<i>Xiaomin Zhao, Ming J Zuo, Tejas Patel</i>	
Graph-based Optimization Method for Information Diffusion and Attack Durability in Networks	698
<i>Zbigniew Tarapata, Rafał Kasprzyk</i>	

Granularity and Granular Systems

Paraconsistent and Approximate Semantics for the OWL 2 Web Ontology
 Language 710
Linh Anh Nguyen

Representation of Granularity for Non-Euclidian Relational Data by Jac-
 card Coefficients and Binary Classifications..... 721
Shoji Hirano, Shusaku Tsumoto

Information Systems in Modeling Interactive Computations on Granules .. 730
Andrzej Skowron, Piotr Wasilewski

Distributed Representations to Detect Higher Order Term Correlations in
 Textual Content 740
Pinar Öztürk, Rajendra Prasath, Hans Moen

Author Index 751

Emergent Dynamics of Information Propagation in Large Networks

Katia Sycara

Robotics Institute, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213
`katia@cs.cmu.edu`

Summary

Large scale networked systems that include heterogeneous entities, e.g., humans and computational entities are becoming increasingly prevalent. Prominent applications include the Internet, large scale disaster relief and network centric warfare. In such systems, large heterogeneous coordinating entities exchange uncertain information to obtain situation awareness. Uncertain and possibly conflicting sensor data is shared across a peer-to-peer network. Not every team member will have direct access to sensors and team members will be influenced mostly by their neighbors in the network with whom they communicate directly. In this talk I will present our work on the dynamics and emergent behaviors of a large team sharing beliefs to reach conclusions about the world. Unlike past work, the nodes in the networks we study are autonomous and actively fuse information they receive. Nodes can change their beliefs as they receive additional information over time.

We find empirically that the dynamics of information propagation in such belief sharing systems are characterized by information avalanches of belief changes caused by a single additional sensor reading. The distribution of the size of these avalanches dictates the speed and accuracy with which the team reaches conclusions. A key property of the system is that it exhibits qualitatively different dynamics and system performance over different ranges of system parameters. In one particular range, the system exhibits behavior known as scale-invariant dynamics which we empirically find to correspond to dramatically more accurate conclusions being reached by team members. Due to the fact that the ranges are very sensitive to configuration details, the parameter ranges over which specific system dynamics occur are extremely difficult to predict precisely. I will present results on the emergent belief propagation dynamics in those systems, mathematical characterization of the systems' behavior and distributed algorithms for adapting the network behaviors to steer the whole system to areas of optimized performance.

New Applications and Theoretical Foundations of the Dominance-based Rough Set Approach

Roman Słowiński

Institute of Computing Science, Poznań University of Technology, Poznań, and
Systems Research Institute, Polish Academy of Sciences, 00-441 Warsaw, Poland,
`roman.slowinski@cs.put.poznan.pl`

Summary

Dominance-based Rough Set Approach (DRSA) has been proposed as an extension of the Pawlak's concept of Rough Sets in order to deal with ordinal data [see [2,3]]. Ordinal data are typically encountered in multi-attribute decision problems where a set of objects (also called actions, acts, solutions, etc.) evaluated by a set of attributes (also called criteria, variables, features, etc.) raises one of the following questions: (i) how to assign the objects to some ordered classes (ordinal classification), (ii) how to choose the best subset of objects (optimization), or (iii) how to rank the objects from the best to the worst (ranking). The answer to everyone of these questions involves an aggregation of the multi-attribute evaluation of objects, which takes into account a law relating the evaluation and the classification, or optimization, or ranking decision. This law has to be discovered from the data by inductive learning. In case of decision problems corresponding to some physical phenomena, this law is a model of cause-effect relationships, and in case of a human decision making, this law is a decision maker's preference model. In DRSA, these models have the form of a set of "if..., then..." decision rules. In case of multi-attribute classification the syntax of rules is: "if evaluation of object a is better (or worse) than given values of some attributes, then a belongs to at least (at most) given class", and in case of multi-attribute optimization or ranking: "if object a is preferred to object b in at least (at most) given degrees with respect to some attributes, then a is preferred to b in at least (at most) given degree".

Since its conception, DRSA has been adapted to a large variety of decision problems [10]. Moreover, it has been adapted to handle granular (fuzzy) information [5], and incomplete information [1]. Stochastic version of DRSA has also been characterized in [9].

In this presentation, we will concentrate on two recent applications of DRSA: decision under uncertainty and time preference [6], and interactive robust multi-objective optimization [4,7]. Moreover, we will give account of topological properties of DRSA [8], using the concept of a bitopological space.

References

1. Dembczyński, K., Greco, S., and Słowiński, R.: Rough set approach to multiple criteria classification with imprecise evaluations and assignments, *European Journal of Operational Research*, 198 (2009) 626-636.
2. Greco, S., Matarazzo, B., and Słowiński, R.: The use of rough sets and fuzzy sets in MCDM, in: Gal, T., Stewart, T., and Hanne, T. (eds.), *Advances in Multiple Criteria Decision Making*, Kluwer, Boston, 1999, pp. 14.1-14.59.
3. Greco, S., Matarazzo, B., and Słowiński, R.: Rough sets theory for multicriteria decision analysis, *European Journal of Operational Research*, 129 (2001) 1-47.
4. Greco, S., Matarazzo, B., and Słowiński, R.: Dominance-based Rough Set Approach to Interactive Multiobjective Optimization, chapter 5 in: Branke, J., Deb, K., Miettinen, K., Słowiński, R. (eds.), *Multiobjective Optimization: Interactive and Evolutionary Approaches*, LNCS 5252, State-of-the-Art Surveys, Springer, Berlin, 2008, pp. 121-155.
5. Greco, S., Matarazzo, B., and Słowiński, R.: Granular Computing for Reasoning about Ordered Data: the Dominance-based Rough Set Approach, chapter 15 in: Pedrycz, W., Skowron, A., Kreinovich, V. (eds.), *Handbook of Granular Computing*. John Wiley & Sons, Chichester, 2008, pp. 347-373.
6. Greco, S., Matarazzo, B., and Słowiński, R.: Dominance-based rough set approach to decision under uncertainty and time preference, *Annals of Operations Research*, 176 (2010) 41-75.
7. Greco, S., Matarazzo, B., and Słowiński, R.: **DARWIN**: Dominance-based rough set Approach to handling **R**obust **W**inning solutions in **I**nteractive multiobjective optimization, in: *Proc. 5th Intl Workshop on Preferences and Decisions*, Trento, April 6-8, 2009, pp. 34-41.
8. Greco, S., Matarazzo, B., and Słowiński, R.: Algebra and Topology for Dominance-based Rough Set Approach, in: Raś, Z.W., Tsay, L.-S. (eds.), *Advances in Intelligent Information Systems*, Studies in Computational Intelligence, vol. 265, Springer, Berlin, 2010, pp. 43-78.
9. Kotłowski, W., Dembczyński, K., Greco, S., and Słowiński, R.: Stochastic dominance-based rough set model for ordinal classification, *Information Sciences*, 178 (2008) 4019-4037.
10. Słowiński, R., Greco, S., Matarazzo, B.: Rough Sets in Decision Making, in: R.A. Meyers (ed.), *Encyclopedia of Complexity and Systems Science*, Springer, New York, 2009, pp. 7753-7786.

RSCTC'2010 Discovery Challenge: Mining DNA Microarray Data for Medical Diagnosis and Treatment

Marcin Wojnarski^{1,2}, Andrzej Janusz², Hung Son Nguyen², Jan Bazan³,
ChuanJiang Luo⁴, Ze Chen⁴, Feng Hu⁴, Guoyin Wang⁴, Lihe Guan⁴, Huan
Luo⁴, Juan Gao⁴, Yuanxia Shen⁴, Vladimir Nikulin⁵, Tian-Hsiang Huang^{5,6},
Geoffrey J. McLachlan⁵, Matko Bošnjak⁷, and Dragan Gamberger⁷

¹ TunedIT Solutions,

Zwirki i Wigury 93/3049, 02-089 Warszawa, Poland

`marcin.wojnarski@tunedit.org`

² Faculty of Mathematics, Informatics and Mechanics, University of Warsaw

Banacha 2, 02-097 Warszawa, Poland

`andrzejjanusz@gmail.com`, `son@mimuw.edu.pl`

³ Institute of Mathematics, University of Rzeszow

Rejtana 16A, 35-959 Rzeszow, Poland

`bazan@univ.rzeszow.pl`

⁴ Institute of Computer Science and Technology,

Chongqing University of Posts and Telecommunications,

Chongqing 400065, P.R.China

`{hufeng,wanggy}@cqupt.edu.cn`

⁵ Department of Mathematics, University of Queensland

`v.nikulin@uq.edu.au`, `gjm@maths.uq.edu.au`

⁶ Institute of Information Management, National Cheng Kung University, Taiwan

`huangtx@gmail.com`

⁷ Laboratory for Information Systems, Rudjer Boskovic Institute,

Bijenicka 54, 10000 Zagreb, Croatia,

`matko.bosnjak@irb.hr`, `dragan.gamberger@irb.hr`

Abstract. RSCTC'2010 Discovery Challenge was a special event of Rough Sets and Current Trends in Computing conference. The challenge was organized in the form of an interactive on-line competition, at TunedIT.org platform, in days between Dec 1, 2009 and Feb 28, 2010. The task was related to feature selection in analysis of DNA microarray data and classification of samples for the purpose of medical diagnosis or treatment. Prizes were awarded to the best solutions. This paper describes organization of the competition and the winning solutions.

TunedIT.org: System for Automated Evaluation of Algorithms in Repeatable Experiments

Marcin Wojnarski^{1,2}, Sebastian Stawicki², and Piotr Wojnarowski^{1,2}

¹ TunedIT Solutions

Zwirki i Wigury 93 lok. 3049, 02-089 Warszawa, Poland

² Faculty of Mathematics, Informatics and Mechanics, University of Warsaw
Banacha 2, 02-097 Warszawa, Poland

Abstract. In this paper we present TunedIT system which facilitates evaluation and comparison of machine-learning algorithms. TunedIT is composed of three complementary and interconnected components: TunedTester, Repository and Knowledge Base.

TunedTester is a stand-alone Java application that runs automated tests (experiments) of algorithms. Repository is a database of algorithms, datasets and evaluation procedures used by TunedTester for setting up a test. Knowledge Base is a database of test results. Repository and Knowledge Base are accessible through TunedIT website. TunedIT is open and free for use by any researcher. Every registered user can upload new resources to Repository, run experiments with TunedTester, send results to Knowledge Base and browse all collected results, generated either by himself or by others.

As a special functionality, built upon the framework of automated tests, TunedIT provides a platform for organization of on-line interactive competitions for machine-learning problems. This functionality may be used, for instance, by teachers to launch contests for their students instead of traditional assignment tasks; or by organizers of machine-learning and data-mining conferences to launch competitions for the scientific community, in association with the conference.

Consensus Multiobjective Differential Crisp Clustering for Categorical Data Analysis

Indrajit Saha¹, Dariusz Plewczyński¹, Ujjwal Maulik²,
Sanghamitra Bandyopadhyay³

¹ Interdisciplinary Centre for Mathematical and Computational Modeling (ICM),
University of Warsaw, 02-089 Warsaw, Poland.

Email: (indra,darman)@icm.edu.pl

² Department of Computer Science and Engineering, Jadavpur University,
Kolkata-700032, West Bengal, India.

Email: drumaulik@cse.jdvu.ac.in

³ Machine Intelligence Unit, Indian Statistical Institute,
Kolkata-700108, West Bengal, India.

Email: sanghami@isical.ac.in

Abstract. In this article, an evolutionary crisp clustering technique is described that uses a new consensus multiobjective differential evolution. The algorithm is therefore able to optimize two conflicting cluster validity measures simultaneously and provides resultant Pareto optimal set of non-dominated solutions. Thereafter the problem of choosing the best solution from resultant Pareto optimal set is resolved by creation of consensus clusters using voting procedure. The proposed method is used for analyzing the categorical data where no such natural ordering can be found among the elements in categorical domain. Hence no inherent distance measure, like the Euclidean distance, would work to compute the distance between two categorical objects. Index-coded encoding of the cluster medoids (centres) is used for this purpose. The effectiveness of the proposed technique is provided for artificial and real life categorical data sets. Also statistical significance test has been carried out to establish the statistical significance of the clustering results. Matlab version of the software is available at <http://bio.icm.edu.pl/~darman/CMODECC>.

Keywords: Crisp clustering, differential evolution, multiobjective optimization, Pareto optimal, statistical significance test.

Probabilistic Rough Entropy Measures in Image Segmentation

Dariusz Małyszko and Jarosław Stepaniuk

Department of Computer Science
Białystok University of Technology
Wiejska 45A, 15-351 Białystok, Poland
{d.malyszko,j.stepaniuk}@pb.edu.pl

Abstract. In numerous data clustering problems, the main priority remains a constant demand on development of new improved algorithmic schemes capable of robust and correct data handling. This requirement has been recently boosted by emerging new technologies in data acquisition area. In image processing and image analysis procedures, the image segmentation procedures have the most important impact on the image analysis results.

In data analysis methods, in order to improve understanding and description of data structures, many innovative approaches have been introduced. Data analysis methods always strongly depend upon revealing inherent data structure. In the paper, a new algorithmic **Rough Entropy Framework** - (REF, in short) has been applied in the probabilistic setting. Crisp and Fuzzy RECA measures (**Rough Entropy Clustering Algorithm**) introduced in [?] are extended into probability area. The basic rough entropy notions, the procedure of rough (entropy) measure calculations and examples of probabilistic approximations have been presented and supported by comparison to crisp and fuzzy rough entropy measures. In this way, uncertainty measures have been combined with probabilistic procedures in order to obtain better insight into data internal structure.

Distance Based Fast Hierarchical Clustering Method for Large Datasets

Bidyut Kr. Patra, Neminath Hubballi, Santosh Biswas, and Sukumar Nandi

Department of Computer Science & Engineering,
Indian Institute of Technology Guwahati, Assam 781039, INDIA
{bidyut, neminath, santosh_biswas, sukumar}@iitg.ernet.in

Abstract. Average-link (AL) is a distance based hierarchical clustering method, which is not sensitive to the noisy patterns. However, like all hierarchical clustering methods AL also needs to scan the dataset many times. AL has time and space complexity of $O(n^2)$, where n is the size of the dataset. These prohibit the use of AL for large datasets. In this paper, we have proposed a distance based hierarchical clustering method termed l -AL which speeds up the classical AL method in any metric (vector or non-vector) space. In this scheme, first leaders clustering method is applied to the dataset to derive a set of leaders and subsequently AL clustering is applied to the leaders. To speed-up the leaders clustering method, reduction in distance computations is also proposed in this paper. Experimental results confirm that the l -AL method is considerably faster than the classical AL method yet keeping clustering results at par with the classical AL method.

TI-DBSCAN: Clustering with DBSCAN by Means of the Triangle Inequality

Marzena Kryszkiewicz and Piotr Lasek

Institute of Computer Science, Warsaw University of Technology
Nowowiejska 15/19, 00-665 Warsaw, Poland
{mkr, p.lasek}@ii.pw.edu.pl

Abstract. Grouping data into meaningful clusters is an important data mining task. DBSCAN is recognized as a high quality density-based algorithm for clustering data. It enables both the determination of clusters of any shape and the identification of noise in data. The most time-consuming operation in DBSCAN is the calculation of a neighborhood for each data point. In order to speed up this operation in DBSCAN, the neighborhood calculation is expected to be supported by spatial access methods. DBSCAN, nevertheless, is not efficient in the case of high dimensional data. In this paper, we propose a new efficient TI DBSCAN algorithm and its variant TI-DBSCAN-REF that apply the same clustering methodology as DBSCAN. Unlike DBSCAN, TI-DBSCAN and TI-DBSCAN-REF do not use spatial indices; instead they use the triangle inequality property to quickly reduce the neighborhood search space. The experimental results prove that the new algorithms are up to three orders of magnitude faster than DBSCAN, and efficiently cluster both low and high dimensional data.

Vehicle Classification Based on Soft Computing Algorithms

Piotr Dalka and Andrzej Czyżewski

Gdansk University of Technology, Multimedia Systems Department,
Narutowicza 11/12, 80-233, Gdansk, Poland
{dalken, andcz}@sound.eti.pg.gda.pl

Abstract. Experiments and results regarding vehicle type classification are presented. Three classes of vehicles are recognized: sedans, vans and trucks. The system uses a non-calibrated traffic camera, therefore no direct vehicle dimensions are used. Various vehicle descriptors are tested, including those based on vehicle mask only and those based on vehicle images. The latter ones employ Speeded Up Robust Features (SURF) and gradient images convolved with Gabor filters. Vehicle type is recognized with various classifiers: artificial neural network, K-nearest neighbors algorithm, decision tree and random forest.

Controlling Computer by Lip Gestures Employing Neural Networks

Piotr Dalka and Andrzej Czyżewski

Gdansk University of Technology, Multimedia Systems Department,
Narutowicza 11/12, 80-233, Gdansk, Poland
{dalken, andcz}@sound.eti.pg.gda.pl

Abstract. Results of experiments regarding lip gesture recognition with an artificial neural network are discussed. The neural network module forms the core element of a multimodal human-computer interface called LipMouse. This solution allows a user to work on a computer using lip movements and gestures. A user face is detected in a video stream from a standard web camera using a cascade of boosted classifiers working with Haar-like features. Lip region extraction is based on a lip shape approximation calculated by the means of lip image segmentation using fuzzy clustering. ANN is fed with a feature vector describing lip region appearance. The descriptors used include a luminance histogram, statistical moments and co-occurrence matrices statistical parameters. ANN is able to recognize with a good accuracy three lip gestures: mouth opening, sticking out the tongue and forming puckered lips.

Computer Animation System Based on Rough Sets and Fuzzy Logic

Piotr Szczuko¹

Multimedia Systems Department, Gdansk University of Technology
szczuko@sound.eti.pg.gda.pl,
WWW home page: <http://sound.eti.pg.gda.pl>

Abstract. A fuzzy logic inference system was created, based on the analysis of animated motion features. The objective of the system is to facilitate the creation of high quality animation by analyzing personalized styles contained in numerous animations. Sequences portraying a virtual character acting with a differentiating personalized style (natural or exaggerated) and various levels of fluidity were prepared and subjectively evaluated. Knowledge gathered in subjective evaluation tests was processed utilizing variable precision rough set (VPRS) approach for defining non-ambiguous inverse relation between subjective features of the result animation and objective parameters of the animated motion. Once the mapping is known then the user can define own requirements on animation, and the input motion is processed accordingly to produce the desired result. The paper focuses on employing variable precision rough set methodology for selection of representative parameter values.

Adaptive Phoneme Alignment based on Rough Set Theory

Konstantinos Avdelidis¹, Charalampos Dimoulas¹, George Kalliris², George Papanikolaou¹

¹ Aristotle University of Thessaloniki

School of Engineering, Dept. of Electrical and Computer Engineering
University Campus, 54124 Thessaloniki, Greece

² Aristotle University of Thessaloniki

Dept. of Journalism and Mass Communication, University Campus
54124 Thessaloniki, Greece

avdel@auth.gr, babis@eng.auth.gr, gkal@jour.auth.gr, pap@eng.auth.gr

Abstract. The current work describes a phoneme matching algorithm based on rough set concepts. The objective of this type of algorithms is focused on the localization of the phonemic content of a specific spoken occurrence. According to the proposed algorithm, a number of rough sets containing the multiple expected phonemic instances in a sequence are created, each defined by a set of short term frames of the voice signal. The properties of the corresponding information system are derived from a features set calculated from the speech signal upon initiation. Given the above, an iterative procedure is applied by updating the phoneme instances versus the optimization of the accuracy metric. The main advantage of this algorithm is the absence of a training phase allowing for wider speaker adaptability and independency. The current paper focuses on the feasibility of the task as this work is still in early research stage.

Monitoring Parkinson's Disease Patients Employing Biometric Sensors and Rule-Based Data Processing

Paweł Żwan, Katarzyna Kaszuba and Bożena Kostek

Gdansk University of Technology, Narutowicza 11/12, 80-233 Gdansk, Poland
{zwan, katkasz, bozenka}@sound.eti.pg.gda.pl

Abstract. The paper presents how rule-based processing can be applied to automatically evaluate the motor state of Parkinson's Disease patients. Automatic monitoring of patients by using biometric sensors can provide assessment of the Parkinson's Disease symptoms. All data on PD patients' state are compared to historical data stored in the database and then a rule-based decision is applied to assess the overall illness state. The training procedure based on doctors' questionnaires is presented. These data constitute the input of several rule-based classifiers. It has been proved that the rough-set-based algorithm can be very suitable for automatic assessment of the PD patient's stability/worsening state.

Keywords: Rough sets, Rule-based processing, Parkinson's Disease, decision systems, automatic assessment of the patient's motor state.

Content-Based Scene Detection and Analysis Method for Automatic Classification of TV Sports News

Kazimierz Choroś and Piotr Pawlaczyk

Institute of Informatics, Wrocław University of Technology
Wybrzeże Wyspińskiego 27, 50-370 Wrocław, Poland
kazimierz.choros@pwr.wroc.pl

Abstract. A large amount of digital video data is stored in local or network visual retrieval systems. The new technology advances in multimedia information processing as well as in network transmission have made video data publicly and relatively easy available. Users need the adequate tools to locate their desired video or video segments quickly and efficiently, for example in Internet video collections, TV shows archives, video-on-demand systems, personal video archives offered by many public Internet services, etc. Detection of scenes in TV videos is difficult because the diversity of effects used in video editing puts up a barrier to construct an appropriate model. The framework of automatic recognition and classification of scenes reporting the sport events in a given discipline in TV sports news have been proposed. Experimental results show good performance of the proposed scheme on detecting scenes on a given sport discipline in TV sports news. In the tests a special software called AVI - the Automatic Video Indexer has been used to detect shots and then scenes in tested TV news videos.

Combining Multiple Classification or Regression Models Using Genetic Algorithms

Andrzej Janusz

Faculty of Mathematics, Informatics, and Mechanics, The University of Warsaw,
Banacha 2, 02-097 Warszawa, Poland,
andrzejjanusz@gmail.com

Abstract. Blending is a well-established technique, commonly used to increase performance of predictive models. Its effectiveness has been confirmed in practice as most of the latest international data-mining contest winners were using some kind of a committee of classifiers to produce their final entry. This paper is a technical report presenting a method of using a genetic algorithm to optimize an ensemble of multiple classification or regression models. An implementation of this method in *R*, called Genetic Meta-Blender, was tested during the Australian Data Mining 2009 Analytic Challenge competition and it was awarded with the Grand Champion prize for achieving the best overall result. In the report, the purpose of the challenge is described and details of the winning approach are given. The results of Genetic Meta-Blender are also discussed and compared to several baseline scores.

Argument Based Generalization of MODLEM Rule Induction Algorithm

Krystyna Napierała and Jerzy Stefanowski

Institute of Computing Science, Poznań University of Technology,
ul. Piotrowo 2, 60-965 Poznań, Poland
{krystyna.napierala, jerzy.stefanowski}@cs.put.poznan.pl

Abstract. Argument based learning allows experts to express their domain, local knowledge about the circumstances of making classification decisions for some learning examples. In this paper we have incorporated this idea in rule induction as a generalization of the MODLEM algorithm. To adjust the algorithm to the redefined task, a new measure for evaluating rule conditions and a new classification strategy with rules had to be introduced. Experimental studies showed that using arguments improved classification accuracy and structure of rules. Moreover the proper argumentation improved recognition of the minority class in imbalanced data without essential decreasing recognition of the majority classes.

Integrating Selective Pre-processing of Imbalanced Data with Ivotes Ensemble

Jerzy Błaszczyński, Magdalena Deckert, Jerzy Stefanowski, Szymon Wilk

Institute of Computing Science, Poznań University of Technology,
60-965 Poznań, Poland
{jerzy.blaszczyński, magdalena.deckert, jerzy.stefanowski,
szymon.wilk}@cs.put.poznan.pl

Abstract. In the paper we present a new framework for improving classifiers learned from imbalanced data. This framework integrates the SPIDER method for selective data pre-processing with the Ivotes ensemble. The goal of such integration is to obtain improved balance between the sensitivity and specificity for the minority class in comparison to a single classifier combined with SPIDER, and to keep overall accuracy on a similar level. The Ivotes framework was evaluated in a series of experiments, in which we tested its performance with two types of component classifiers (tree- and rule-based). The results show that Ivotes improves evaluation measures. They demonstrated advantages of the abstaining mechanism (i.e., refraining from predictions by component classifiers) in Ivotes rule ensembles.

Learning from Imbalanced Data in Presence of Noisy and Borderline Examples

Krystyna Napierała, Jerzy Stefanowski and Szymon Wilk

Institute of Computing Science, Poznań University of Technology,
ul. Piotrowo 2, 60-965 Poznań, Poland
{krystyna.napierala, jerzy.stefanowski, szymon.wilk}@cs.put.poznan.pl

Abstract. In this paper we studied re-sampling methods for learning classifiers from imbalanced data. We carried out a series of experiments on artificial data sets to explore the impact of noisy and borderline examples from the minority class on the classifier performance. Results showed that if data was sufficiently disturbed by these factors, then the focused re-sampling methods – NCR and our SPIDER2 – strongly outperformed the oversampling methods. They were also better for real-life data, where PCA visualizations suggested possible existence of noisy examples and large overlapping areas between classes.

Tracking Recurrent Concepts using Context

João Bártolo Gomes¹, Ernestina Menasalvas¹, and Pedro A. C. Sousa²

¹ Facultad de Informatica - Universidad Politecnica Madrid, Spain
joao.bartolo.gomes@alumnos.upm.es
emenasalvas@fi.upm.es

² Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal
pas@fct.unl.pt

Abstract. The problem of recurring concepts in data stream classification is a special case of concept drift where concepts may reappear. Although several methods have been proposed that are able to learn in the presence of concept drift, few consider concept recurrence and integration of context. In this work, we extend existing drift detection methods to deal with this problem by exploiting context information associated with learned decision models in situations where concepts reappear. The preliminary experimental results demonstrate the effectiveness of the proposed approach for data stream classification problems with recurring concepts.

Keywords Data Stream Mining, Concept Drift, Recurring Concepts, Context-awareness, Ubiquitous Knowledge Discovery

Support Feature Machine for DNA Microarray Data

Tomasz Maszczyk and Włodzisław Duch

Department of Informatics, Nicolaus Copernicus University
Grudziądzka 5, 87-100 Toruń, Poland
{tmaszczyk, wduch}@is.umk.pl

Abstract. Support Feature Machines (SFM) define useful features derived from similarity to support vectors (kernel transformations), global projections (linear or perceptron-style) and localized projections. Explicit construction of extended feature spaces enables control over selection of features, complexity control and allows final analysis by any classification method. Additionally projections of high-dimensional data may be used to estimate and display confidence of predictions. This approach has been applied to the DNA microarray data.

Is It Important which Rough-Set-Based Classifier Extraction and Voting Criteria are Applied Together?

Dominik Ślęzak^{1,2}, Sebastian Widz^{3,4}

¹ Institute of Mathematics, University of Warsaw
Banacha 2, 02-097 Warsaw, Poland

² Infobright Inc., Poland
Krzywickiego 34 pok. 219, 02-078 Warsaw, Poland

³ XPLUS SA
ul. Gzégzółki 4, 02-804 Warsaw, Poland

⁴ Polish-Japanese Institute of Information Technology
Koszykowa 86, 02-008 Warsaw, Poland
`slezak@infobright.com,sebastian.widz@xplus.pl`

Abstract. We propose a framework for experimental verification whether mechanisms of voting among rough-set-based classifiers and criteria for extracting those classifiers from data should follow analogous mathematical principles. Moreover, we show that some of types of criteria perform better for high-quality data while the others are useful rather for low-quality data. The framework is based on the principles of approximate attribute reduction and probabilistic extensions of rough-set-based approach to data analysis. The framework is not supposed to produce the best-ever classification results, unless it is extended by some additional parameters known from the literature. Instead, our major goal is to illustrate in a possibly simplistic way that it is worth unifying mathematical background for the stages of learning and applying rough-set-based classifiers.

Improving Co-training with Agreement-based Sampling

Jin Huang¹, Jelber Sayyad Shirabad¹, Stan Matwin^{1,2}, and Jiang Su¹

¹ School of Information Technology and Engineering, University of Ottawa, Canada
{jhuang33, jsayyad, stan, jsu}@site.uottawa.ca

² Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland

Abstract. Co-training is an effective semi-supervised learning method which uses unlabeled instances to improve prediction accuracy. In the co-training process, a random sampling is used to gradually select unlabeled instances to train classifiers. In this paper we explore whether other sampling methods can improve co-training performance. A novel selective sampling method, agreement-based sampling, is proposed. Experimental results show that our new sampling method can improve co-training significantly.

Experienced Physicians and Automatic Generation of Decision Rules from Clinical Data

William Klement¹, Szymon Wilk^{1,3}, Martin Michalowski² and Ken Farion³

¹ MET Research Group, University of Ottawa, Canada.

² Adventium Labs, Minneapolis MN 55401, USA.

³ Institute of Computing Science, Poznan University of Technology, Poland.

⁴ Division of Emergency Medicine, Children's Hospital of Eastern Ontario, Canada.

Abstract. Clinical Decision Support Systems embed data-driven decision models designed to represent clinical acumen of an experienced physician. We argue that eliminating physicians' diagnostic biases from data improves the overall quality of concepts, which we represent as decision rules. Experiments conducted on prospectively collected clinical data show that analyzing this filtered data produces rules with better coverage, certainty and confirmation. Cross-validation testing shows improvement in classification performance.

Gene-Pair Representation and Incorporation of GO-based Semantic Similarity into Classification of Gene Expression Data

Torsten Schön^{1,2}, Alexey Tsymbal², and Martin Huber²

¹ Hochschule Weihenstephan-Triesdorf, Freising, Germany
`info@torsten-schoen.de`

² Corporate Technology Div., Siemens AG, Erlangen, Germany
`alexey.tsymbal@siemens.com`, `martin.huber@siemens.com`

Abstract. To emphasize gene interactions in the classification algorithms, a new representation is proposed, comprising gene-pairs and not single genes. Each pair is represented by L1 difference in the corresponding expression values. The novel representation is evaluated on benchmark datasets and is shown to often increase classification accuracy for genetic datasets. Exploiting the gene-pair representation and the Gene Ontology (GO), the semantic similarity of gene pairs can be incorporated to pre-select pairs with a high similarity value. The GO-based feature selection approach is compared to the plain data driven selection and is shown to often increase classification accuracy.

A Fuzzy View on Rough Satisfiability

Anna Gomolińska

Białystok University, Department of Mathematics,
Akademicka 2, 15267 Białystok, Poland,
anna.gom@math.uwb.edu.pl

Abstract. In the paper, several notions of rough satisfiability of formulas are recalled and discussed from the standpoint of fuzzy set theory. By doing so we aim to better understand what rough satisfiability really means and to look for schemata describing it.

Keywords: fuzzy set, satisfiability of formulas, descriptor language, approximation space.

Rough Sets in Terms of Discrete Dynamical Systems

Marcin Wolski

Department of Logic and Philosophy of Science
Maria Curie-Skłodowska University
marcin.wolski@umcs.lublin.pl

Abstract. In the paper we consider a topological approximation space (U, τ) (induced by a given information system \mathcal{I}) as a discrete dynamical system; that is, we are concerned with a finite approximation space U whose topology τ is induced by a function $f : U \rightarrow U$. Our aim is to characterise these type of approximation spaces by means of orbits which represent the evolution of points of U with respect to the process f . Apart from topological considerations we also provide some algebraic characterisation of orbits. Due to the finiteness condition imposed by \mathcal{I} , any point $a \in U$ is eventually cyclic. In consequence, as we demonstrate, orbits are algebraically close to rough sets, e.g. they induce a Łukasiewicz algebra of order two, where the lower approximation operator may be interpreted as the action of retrieving a cycle from a given orbit and the upper approximation operator may be interpreted as the action of making a given orbit cyclic.

A Preference-based Multiple-Source Rough Set Model

Md. Aquil Khan and Mohua Banerjee

Department of Mathematics and Statistics,
Indian Institute of Technology,
Kanpur 208 016, India
`{mdaquil,mohua}@iitk.ac.in`

Abstract. We propose a generalization of Pawlak's rough set model for the multi-agent situation, where information from an agent can be preferred over that of another agent of the system while deciding membership of objects. Notions of lower/upper approximations are given which depend on the knowledge base of the sources as well as on the position of the sources in the hierarchy giving the preference of sources. Some direct consequences of the definitions are presented.

Classification of Dynamics in Rough Sets

Davide Ciucci

Dipartimento Di Informatica, Sistemistica e Comunicazione
Università di Milano – Bicocca, Viale Sarca 336 – U14, I-20126 Milano (Italia)
`ciucci@disco.unimib.it`

Abstract. A classification of the different dynamics which can arise in rough sets is given, starting from three different standpoints: information tables, approximation spaces and coverings. Dynamics is intended in two broad meanings: evolution in time and originated from different sources. Existing works on this topic are then categorized accordingly.

Relational Granularity for Hypergraphs

John G. Stell

School of Computing, University of Leeds, LS2 9JT, U.K.
j.g.stell@leeds.ac.uk

Abstract. A set of subsets of a set may be seen as granules that allow arbitrary subsets to be approximated in terms of these granules. In the simplest case of rough set theory, the set of granules is required to partition the underlying set, but granulations based on relations more general than equivalence relations are well-known within rough set theory. The operations of dilation and erosion from mathematical morphology, together with their converse forms, can be used to organize different techniques of granular approximation for subsets of a set with respect to an arbitrary relation. The extension of this approach to granulations of sets with structure is examined here for the case of hypergraphs. A novel notion of relation on a hypergraph is presented, and the application of these relations to a theory of granularity for hypergraphs is discussed.

Perceptual Tolerance Intersection

Piotr Wasilewski¹, James F. Peters¹, and Sheela Ramanna²

¹ Computational Intelligence Laboratory,
Department of Electrical & Computer Engineering, Univ. of Manitoba,
75A Chancellor's Circle, Winnipeg, Manitoba R3T 5V6, Canada
`{piotr,jfpeters}@ee.umanitoba.ca`

² Department of Applied Computer Science
University of Winnipeg
Winnipeg, Manitoba R3B 2E9, Canada
`s.ramanna@uwinnipeg.ca`

Abstract. This paper introduces a perceptual tolerance intersection of sets as an example of near set operations. Such operations are motivated by the need to consider similarities between digital images viewed as disjoint sets of points. The proposed approach is in keeping with work by E.C. Zeeman on tolerance spaces and visual perception and J.H. Poincaré on sets of similar sensations used to define representative (aka *tolerance*) spaces such as visual, tactile and motile spaces. Perceptual tolerance intersection of sets is a direct consequence of recent work on near sets and a solution to the problem of how one goes about discovering affinities between digital images. The main contribution of this article is a description-based approach to assessing the resemblances between digital images.

Keywords: description, near sets, perceptual granules, set operations, similarity, tolerance.

Categories of Direlations and Rough Set Approximation Operators

Murat Diker

Hacettepe University, Department of Mathematics,
06800 Beytepe Ankara, Turkey
mdiker@hacettepe.edu.tr

Abstract. In this paper, we define a category **R-APR** whose objects are sets and morphisms are the pairs of rough set approximation operators. We show that **R-APR** is isomorphic to a full subcategory of the category **cdrTex** whose objects are complemented textures and morphisms are complemented direlations. Therefore, **cdrTex** may be regarded as an abstract model for the study of rough set theory. On the other hand, dagger symmetric monoidal categories play a central role in the abstract quantum mechanics. Here, we show that **R-APR** and **cdrTex** are also dagger symmetric monoidal categories.

Approximations and Classifiers

Andrzej Skowron¹ and Jarosław Stepaniuk²

¹ Institute of Mathematics, The University of Warsaw
Banacha 2, 02-097 Warsaw, Poland
skowron@mimuw.edu.pl

² Department of Computer Science, Białystok University of Technology
Wiejska 45A, 15-351 Białystok, Poland,
j.stepaniuk@pb.edu.pl

Abstract. We discuss some important issues for applications that are related to generalizations of the 1994 approximation space definition [?]. In particular, we present examples of rough set based strategies for extension of approximation spaces from samples of objects onto the whole universe of objects. This makes it possible to present methods for inducing approximations of concepts or classifications analogously to the approaches for inducing classifiers known in machine learning or data mining.

Keywords: approximation space, lower approximation, upper approximation, rough sets, extension of approximation space, classifiers

A Note on a Formal Approach to Rough Operators

Adam Grabowski and Magdalena Jastrzębska

Institute of Mathematics
University of Białystok
ul. Akademicka 2
15-267 Białystok, Poland
adam@math.uwb.edu.pl, magjas0@poczta.onet.pl

Abstract. The paper is devoted to the formalization of two elementary but important problems within rough set theory. We mean searching for the minimal requirements of the well-known rough operators – the lower and the upper approximations in an abstract approximation space to retain their natural properties. We also discuss pros and cons of the development of the computer-checked repository for rough set theory based on the comparison of certain rough approximation operators proposed by Anna Gomolińska.

Communicative Approximations as Rough Sets

Mohua Banerjee¹, Abhinav Pathak², Gopal Krishna², and
Amitabha Mukerjee²

¹ Dept of Mathematics and Statistics,
Indian Institute of Technology, Kanpur, India
mohua@iitk.ac.in

² Dept of Computer Science & Engineering,
Indian Institute of Technology, Kanpur, India
amit@cse.iitk.ac.in

Abstract. Communicative approximations, as used in language, are equivalence relations that partition a continuum, as opposed to observational approximations on the continuum. While the latter can be addressed using tolerance interval approximations on interval algebra, new constructs are necessary for considering the former, including the notion of a “rough interval”, which is the indiscernibility region for an event described in language, and “rough points” for quantities and moments. We develop the set of qualitative relations for points and intervals in this “communicative approximation space”, and relate them to existing relations in exact and tolerance-interval formalisms. We also discuss the nature of the resulting algebra.

On the Correctness of Rough-Set Based Approximate Reasoning

Patrick Doherty¹ and Andrzej Szalas^{1,2}

¹ Dept. of Computer and Information Science
Linköping University, SE-581 83 Linköping, Sweden
patdo@ida.liu.se

² Institute of Informatics, Warsaw University
Banacha 2, 02-097 Warsaw, Poland
andsz@mimuw.edu.pl

Abstract. There is a natural generalization of an indiscernibility relation used in rough set theory, where rather than partitioning the universe of discourse into indiscernibility classes, one can consider a covering of the universe by similarity-based neighborhoods with lower and upper approximations of relations defined via the neighborhoods. When taking this step, there is a need to tune approximate reasoning to the desired accuracy. We provide a framework for analyzing self-adaptive knowledge structures. We focus on studying the interaction between inputs and output concepts in approximate reasoning. The problems we address are:

- given similarity relations modeling approximate concepts, what are similarity relations for the output concepts that guarantee correctness of reasoning?
- assuming that output similarity relations lead to concepts which are not accurate enough, how can one tune input similarities?

Unit Operations in Approximation Spaces

Zbigniew Bonikowski

Institute of Mathematics and Informatics
Opole University, Poland
zbonik@math.uni.opole.pl

Abstract. Unit operations are some special functions on sets. The concept of the unit operation originates from researches of U. Wybraniec-Skardowska. The paper is concerned with the general properties of such functions. The isomorphism between binary relations and unit operations is proved. Algebraic structures of families of unit operations corresponding to certain classes of binary relations are considered. Unit operations are useful in Pawlak's Rough Set Theory. It is shown that unit operations are upper approximations in approximation space. We prove, that in the approximation space (U, R) generated by a reflexive relation R the corresponding unit operation is the least definable approximation if and only if the relation R is transitive.

Weighted Nearest Neighbor Classification via Maximizing Classification Consistency

Pengfei Zhu, Qinghua Hu, Yongbin Yang

Harbin Institute of Technology, Harbin 150001, China
huqinghua@hit.edu.cn

Abstract. The nearest neighbor classification is a simple and effective technique for pattern recognition. The performance of this technique is known to be sensitive to the distance function used in classifying a test instance. In this paper, we propose a technique to learn sample weights via maximizing classification consistency. Experimental analysis shows that the distance trained in this way enlarges the classification consistency on several datasets and has a strong ability to tolerate noise. Moreover, the proposed approach has better performance than nearest neighbor classification and several state-of-the-art methods.

Rough Set-Based Incremental Learning Approach to Face Recognition

Xuguang Chen and Wojciech Ziarko

Department of Computer Science University of Regina
Regina, SK, S4S 0A2, Canada

Abstract. The article reports our implementation of a rough set-based incremental learning algorithm involving the application of the hierarchy of probabilistic decision tables to face recognition. The implementation, the related theoretical background such as the basics of the variable precision rough set theory, the algorithm, the classifier structure and experiments with balanced and imbalanced data sets are presented.

A Comparison of Dynamic and Static Belief Rough Set Classifier

Salsabil Trabelsi¹, Zied Elouedi¹, and Pawan Lingras²

¹ Larodec, Institut Supérieur de Gestion de Tunis, Tunisia

² Saint Mary's University Halifax, Canada

Abstract. In this paper, we propose a new approach of classification based on rough sets denoted Dynamic Belief Rough Set Classifier (D-BRSC) which is able to learn decision rules from uncertain data. The uncertainty appears only in decision attributes and is handled by the Transferable Belief Model (TBM), one interpretation of the belief function theory. The feature selection step of the construction procedure of our new technique of classification is based on the calculation of dynamic reduct. The reduction of uncertain and noisy decision table using dynamic approach which extracts more relevant and stable features yields more significant decision rules for the classification of the unseen objects. To prove that, we carry experimentations on real databases using the classification accuracy criterion. We also compare the results of D-BRSC with those obtained from Static Belief Rough Set Classifier (S-BRSC).

Keywords : rough sets, belief function theory, uncertainty, dynamic reduct, classification.

Rule Generation in Lipski's Incomplete Information Databases

Hiroshi Sakai¹, Michinori Nakata², and Dominik Ślęzak^{3,4}

¹ Mathematical Sciences Section, Department of Basic Sciences,
Faculty of Engineering, Kyushu Institute of Technology
Tobata, Kitakyushu 804, Japan
sakai@mns.kyutech.ac.jp

² Faculty of Management and Information Science,
Josai International University
Gumyo, Togane, Chiba 283, Japan
nakatam@ieee.org

³ Institute of Mathematics, University of Warsaw
Banacha 2, 02-097 Warsaw, Poland

⁴ Infobright Inc., Poland
Krzywickiego 34 pok. 219, 02-078 Warsaw, Poland
slezak@infobright.com

Abstract. *Non-deterministic Information Systems (NISs)* are well known as systems for handling information incompleteness in data. In our previous work, we have proposed *NIS-Apriori* algorithm aimed at extraction of decision rules from *NISs*. *NIS-Apriori* employs the minimum and the maximum supports for each descriptor, and it effectively calculates the criterion values for defining rules. In this paper, we focus on *Lipski's Incomplete Information Databases (IIDs)*, which handle non-deterministic information by means of the sets of values and intervals. We clarify how to understand decision rules in *IIDs* and appropriately adapt our *NIS-Apriori* algorithm to generate them. Rule generation in *IIDs* turns out to be more flexible than in *NISs*.

A Fast Randomisation Test for Rule Significance

Ivo Düntsch¹ and Günther Gediga²

¹ Dept of Computer Science, Brock University, St. Catharines, ON, Canada,
`duentsch@brocku.ca`

² Dept of Psychology, Universität Münster, Fliegerstr. 21, D-48149 Münster,
`guenther@gediga.de`

Abstract. Randomisation is a method to test the statistical significance of a symbolic rule; it is, however, very expensive. In this paper we present a sequential randomisation test which in most cases greatly reduces the number of steps needed for a conclusion.

Ordinal Classification with Monotonicity Constraints by Variable Consistency Bagging

Jerzy Błaszczyński¹, Roman Słowiński^{1,2}, Jerzy Stefanowski¹

Institute of Computing Science, Poznań University of Technology¹,
60-965 Poznań, Poland

Systems Research Institute, Polish Academy of Sciences²,
01-447 Warsaw, Poland

{jurek.blaszczyński, roman.slowinski, jerzy.stefanowski}@cs.put.poznan.pl

Abstract. We propose an ensemble method that solves ordinal classification problem with monotonicity constraints. The classification data is structured using the Variable Consistency Dominance-based Rough Set Approach (VC-DRSA). The method employs a variable consistency bagging scheme to produce bootstrap samples that privilege objects (i.e., classification examples) with relatively high values of consistency measures used in VC-DRSA. In result, one obtains an ensemble of rule classifiers learned on bootstrap samples. Due to diversification of bootstrap samples controlled by consistency measures, the ensemble of classifiers gets more accurate, which has been acknowledged by a computational experiment on benchmark data.

Learnability in Rough Set Approaches

Jerzy Błaszczyński¹, Roman Słowiński^{1,2}, Marcin Szeląg¹

¹ Institute of Computing Science, Poznań University of Technology,
60-965 Poznań, Poland

{jblaszczyński, rslowinski, mszelag}@cs.put.poznan.pl

² Institute for Systems Research, Polish Academy of Sciences,
01-447 Warsaw, Poland

Abstract. We consider learning abilities of classifiers learned from data structured by rough set approaches into lower approximations of considered sets of objects. We introduce two measures, λ and δ , that estimate attainable predictive accuracy of rough-set-based classifiers. To check the usefulness of the estimates for various types of classifiers, we perform a computational experiment on fourteen data sets. In the experiment, we use two versions of the rough-set-based rule classifier, called VC-DomLEM, and few other well known classifiers. The results show that both introduced measures are useful for an a priori identification of data sets that are hard to learn by all classifiers.

Upper Bounds on Minimum Cardinality of Exact and Approximate Reducts

Igor Chikalov¹, Mikhail Moshkov¹, and Beata Zielosko²

¹ Mathematical and Computer Sciences & Engineering Division
King Abdullah University of Science and Technology
Thuwal 23955-6900, Saudi Arabia

{igor.chikalov, mikhail.moshkov}@kaust.edu.sa

² Institute of Computer Science, University of Silesia
39, Będzińska St., Sosnowiec, 41-200, Poland
beata.zielosko@us.edu.pl

Abstract. In the paper, we consider the notions of exact and approximate decision reducts for binary decision tables. We present upper bounds on minimum cardinality of exact and approximate reducts depending on the number of rows (objects) in the decision table. We show that the bound for exact reducts is unimprovable in the general case, and the bound for approximate reducts is almost unimprovable in the general case.

Keywords: Exact reduct, approximate reduct, upper bound on minimum cardinality.

An Extension of Rough Set Approximation to Flow Graph Based Data Analysis

Doungrat Chitcharoen and Puntip Pattaraintakorn

School of Mathematics, Faculty of Science,
King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand 10520
s0067111@kmitl.ac.th, kppuntip@kmitl.ac.th

Abstract. This paper concerns some aspects of mathematical flow graph based data analysis. In particular, taking a flow graph view on rough sets' categories and measures leads to a new methodology of inductively reasoning from data. This perspective shows interesting relationships and properties among rough set, flow graphs and inverse flow graphs. A possible car dealer application is outlined and discussed. Evidently, our new categories and measures assist and alleviate some limitations in flow graphs to discover new patterns and explanations.

Credibility Coefficients Based on SVM

Roman Podraza and Bartosz Janeczek

Institute of Computer Science, Warsaw University of Technology
Nowowiejska 15/19, 00-665 Warsaw, Poland
R.Podraza@ii.pw.edu.pl

Abstract. ARES System was a data analysis tool supporting Rough Set theory. It has been expanded to cover other approaches like Emerging Patterns and Support Vector Machine. A special feature of ARES System is ability to identify exceptional objects within information systems by using credibility coefficients. The credibility coefficient is a measure, which attempts to weigh up a degree of typicality of each object in respect to the rest of information system. The paper presents an idea of credibility coefficients based on SVM approach. The new coefficients are compared with the others ones available in the ARES System.

On Algorithm for Building of Optimal α -Decision Trees

Abdulaziz Alkhalid, Igor Chikalov, and Mikhail Moshkov *

Mathematical and Computer Sciences & Engineering Division
King Abdullah University of Science and Technology
Thuwal 23955-6900, Saudi Arabia

{[abdulaziz.alkhalid](mailto:abdulaziz.alkhalid@kaust.edu.sa), [igor.chikalov](mailto:igor.chikalov@kaust.edu.sa), [mikhail.moshkov](mailto:mikhail.moshkov@kaust.edu.sa)}@kaust.edu.sa

Abstract. The paper describes an algorithm that constructs approximate decision trees (α -decision trees), which are optimal relatively to one of the following complexity measures: depth, total path length or number of nodes. The algorithm uses dynamic programming and extends methods described in [?] to constructing approximate decision trees. Adjustable approximation rate allows controlling algorithm complexity. The algorithm is applied to build optimal α -decision trees for two data sets from UCI Machine Learning Repository [?].

* The research has been partially supported by KAUST-Stanford AEA project “Predicting the stability of hydrogen bonds in protein conformations using decision-tree learning methods”.

Layered Approximation Approach to Knowledge Elicitation in Machine Learning

Tuan Trung Nguyen

Polish-Japanese Institute of Information Technology
Koszykowa 86, 02-008 Warsaw, Poland
nttrung@pjwstk.edu.pl

Abstract. Domain knowledge elicitation constitutes a crucial task in designing effective machine learning algorithm, and is often indispensable in problem domains that display a high degree of internal complexity such as knowledge discovery and data mining, the recognition of structured objects, human behavior prediction, or multi-agent cooperation. We show how to facilitate this difficult and sometimes tedious task with a hierarchical concept learning scheme, designed to cope with the inherent vagueness and complexity of knowledge therein used. We present how our approach, based on Rough Mereology and Approximate Reasoning frameworks, correlate to other well established approaches to machine learning.

Configuration Management of Mobile Agents Based on SNMP

Michał Komorowski

Institute of Computer Science,
Warsaw University of Technology,
Nowowiejska 15/19, Warsaw 00-665, Poland
M.Komorowski@ii.pw.edu.pl

Abstract. Mobile agents are small programs that can transport their code, data and execution context from one machine to another and be capable of continuing execution in the new environment. This technology has a lot of advantages and promising applications, unfortunately there is a noticeable absence of deployed solutions. There are few reasons of this situation but one of the most important is the lack of tools that can be used for the configuration management of mobile agents. This process focuses on the monitoring and controlling configuration items and is essential for other processes like incident management or availability management. In this paper a new, flexible and universal solution for the configuration management of mobile agents is proposed. This solution is based on well known and widely used management standard - SNMP (Simple Network Management Protocol).

Adaptive Immunity-based Multiagent Systems (AIBMAS) Inspired by the Idiotypic Network

Chung-Ming Ou¹ and C.R. Ou²

¹ Department of Information Management, Kainan University, Luchu 338, Taiwan
cou077@mail.knu.edu.tw

² Department of Electrical Engineering, Hsiuping Institute of Technology, Taichung
412, Taiwan
crou@mail.hit.edu.tw

Abstract. An adaptive immune-inspired multiagent system (AIBMAS) is proposed. The intelligence behind such system is based on the idiotypic immune network. Tunable activation Threshold (TAT) proposes that agents adapt their activation thresholds. Immune algorithm based on the immune network theory and memory mechanism is derived.

Distributed Default Logic for Context-aware Computing in Multi-Agent Systems

Dominik Ryzko and Henryk Rybiński

Warsaw University of Technology, Institute of Computer Science
Nowowiejska 15/19 00-665 Warsaw, Poland
{d.ryzko, hrb}@ii.pw.edu.pl

Abstract. The paper describes how Distributed Default Logic (DDL) can be used as a formalism for context-aware computing in a Multi-Agent System. It is shown that the original notation does not require any changes. The DDL reasoning engine has been adapted to handle situations like unavailability of sensors. New semantics of Distributed Default Rules in the application to reasoning with context information is also described.

A Novel Approach to Default Reasoning for MAS

Przemysław Więch and Henryk Rybiński

Warsaw University of Technology, Institute of Computer Science
Nowowiejska 15/19 00-665 Warsaw, Poland
{[pwiech](mailto:pwiech@ii.pw.edu.pl),[hrb](mailto:hrb@ii.pw.edu.pl)}@ii.pw.edu.pl

Abstract. In a multi-agent system the sought information can often be found across various knowledge bases, which means that making early assumptions can lead to hasty conclusions. In the paper we present a formalism for distributed default reasoning to be performed by a group of agents that share knowledge in the form of a distributed default theory. The formalism is based on default transformations, which can be used to derive answers to queries in the form of defaults. Such new defaults can then be treated as intermediate results in the reasoning process. It is shown that passing messages containing transformed defaults is more informative than strict statements and enables avoiding early conclusions. Moreover, the extended reasoning features are embedded in the description logic framework.

A Platform for the Evaluation of Automated Argumentation Strategies

Piotr S. Kośmicki

Institute of Computer Science, Faculty of Electronics and Information Technology,
Warsaw University of Technology
Nowowiejska 15/19, 00-665 Warsaw, Poland,
pedrok@wp.pl,
WWW home page: <http://home.elka.pw.edu.pl/~pkosmick/>

Abstract. This paper describes a platform for testing automated argumentation strategies for agents. It is a continuation of the discussion about the Arguing Agents Competition (AAC). The second version of the AAC platform is introduced, including the architecture and the capabilities of the platform, the currently available engine and an automated strategy for the dialogue game.

The argumentation and some of the formalizations for the arguments and dialogues are briefly presented.

Fuzzy Similarity-Based Relative Importance of MPEG-7 Visual Descriptors for Emotional Classification of Images

EunJong Park¹, SungHwan Jeong², and JoonWhoan Lee²

¹ Electronics and Telecommunications Research Institute, Daejeon, South Korea

² Computer Engineering Department, Chonbuk National University, JeonJu, South Korea

for511@etri.re.kr, {chlee,shjeong}@chonbuk.ac.kr

Abstract. Many kinds of attributes are used for various areas of decision making. Sometimes the attributes have complicated vector-types as in MPEG-7 visual descriptors that prevent us from attaching unequal importance to each descriptor for the construction of content- or emotion-based image retrievals. In this paper, fuzzy similarity-based rough approximation is used for determining the relative importance of MPEG-7 visual descriptors for an emotion. In the methods, the relative importance is given to a descriptor itself rather than a component of the vector of a descriptor or a combined descriptor. Also we propose a method for building a classification system based on representative color images. The experimental result shows the proposed classification method is promising for the emotional classification or evaluation of color images.

The Impact of Recommendation Sources on the Adoption Intention of Microblogging based on Dominance-based Rough Set Approach

Yang-Chieh¹, Chaio-Chen Chang², Chiun-Sin Lin³, and Gwo-Hshiung Tzeng^{4,5}

¹ Department of Management Science, National Chiao Tung University
1001 Ta-Hsueh Road, Hsinchu 300, Taiwan

jerry110888@gmail.com

² Department of International Business, National Dong Hwa University
No.1, Sec. 2, Da Hsueh Rd., Shoufeng, Hualien, 97401, Taiwan

aka@mail.ndhu.edu.tw

³ Department of Management Science, National Chiao Tung University
1001 Ta-Hsueh Road, Hsinchu 300, Taiwan

netec7@yahoo.com.tw

⁴ Department of Business Administration, Kainan University
No. 1 Kainan Road, Luchu, Taoyuan 338, Taiwan

ghtzeng@mail.knu.edu.tw

⁵ Institute of Management of Technology, National Chiao Tung University
1001 Ta-Hsueh Road, Hsinchu 300, Taiwan

ghtzeng@cc.nctu.edu.tw

Abstract. Microblogging is a social media tool that allows users to write short text messages to public and private networks. This research focuses specifically on the microblogging on Facebook. The main purposes of this study are to investigate and compare what recommendation sources influence the intention to use microbloggings and to combine gender, daily internet hour usage and past use experience to infer the usage of microbloggings decision rules using a dominance-based rough-set approach (DRSA). Data for this study were collected from 382 users and potential users. The analysis is grounded in the taxonomy of induction-related activities using DRSA to infer the usage of microbloggings decision rules. Finally, the study of the nature of microblogging reflects essential practical and academic value.

Fault Effects Analysis and Reporting System for Dependability Evaluation

Piotr Gawkowski, Monika Anna Kuczyńska, and Agnieszka Komorowska

Institute of Computer Science, Warsaw University of Technology,
Nowowiejska 15/19, Warsaw 00-665, Poland

Abstract. The paper describes the concept and the architecture of the data warehouse and reporting modules dedicated to distributed fault injection testbench. The purpose of this warehouse is to collect the data from fault simulation experiments and support researchers in exploration and analysis of these results. The data model of the warehouse, main ETL processes, multidimensional structure of OLAP cube, and predefined reports are discussed. Practical advantages of the presented system are illustrated with some exemplary analyses of the experimental results collected during dependability evaluation of the chemical reactor control algorithm with software implemented fault injection approach.

Solving the Reporting Cells Problem Using a Scatter Search Based Algorithm

Sónia M. Almeida-Luz¹, Miguel A. Vega-Rodríguez², Juan A. Gómez-Pulido²,
and Juan M. Sánchez-Pérez²

¹ Polytechnic Institute of Leiria, School of Technology and Management
Department of Informatics Engineering, 2400 Leiria, Portugal

`sluz@estg.ipleiria.pt`

² University of Extremadura, Dept. Technologies of Computers and Communications,
Escuela Politécnica

Campus Universitario s/n. 10003 Cáceres. Spain

`{mavega, jangomez, sanperez}@unex.es`

Abstract. The Location Management problem is an important issue of mobility management, which is responsible for determining the network configuration, with the major goal of minimizing the involved costs. One of the most common strategies of location management is the Reporting Cells (RC) scheme, which mainly considers the location update and the paging costs. In this paper we propose a Scatter Search (SS) based approach applied to the Reporting Cells as a cost optimizing solution, with the objective of achieving the best network configuration defining a subset of cells as reporting cells and the others as non-reporting cells. With this work we want to define the most adequate values of the SS parameters, when applied to the RC problem, using twelve test networks that represent 4 distinct groups divided by size. We also want to compare the performance of this SS based approach with a previous study based on Differential Evolution and also with other approaches presented in the literature. The results obtained are very interesting because they outperform those obtained with other approaches exposed in the literature.

Learning Age and Gender using Co-occurrence of Non-Dictionary Words from Stylistic Variations

R.Rajendra Prasath

Department of Computer and Information Science (IDI)
Norwegian University of Science and Technology (NTNU),
Sem Sælands vei 7-9, NO - 7491, Trondheim, Norway,
rajendra@idi.ntnu.no,
<http://www.idi.ntnu.no/~rajendra>;

Abstract. This work attempts to report the stylistic differences in blogging for gender and age group variations using slang word co-occurrences. We have mainly focused on co-occurrence of non dictionary words across bloggers of different gender and age groups. For this analysis, we have focused on the feature *use of slang words* to study the stylistic variations of bloggers across various age groups and gender. We have modeled the co-occurrences of slang words used by bloggers as graph based model where nodes are slang words and edges represent the number of cooccurrences and studied the variations in predicting age groups and gender. We have used demographically tagged blog corpus from ICWSM Spinner dataset for these experiments and used Naive Bayes classifier with 10 fold cross validations. Preliminary results shows that the concurrence of of slang words could be a better choice for predicting age and gender.

Disturbance Measurement Utilization in Easily Reconfigurable Fuzzy Predictive Controllers: Sensor Fault Tolerance and Other Benefits

Piotr M. Marusak

Institute of Control and Computation Engineering, Warsaw University of Technology,
ul. Nowowiejska 15/19, 00-665 Warszawa, Poland,
P.Marusak@ia.pw.edu.pl

Abstract. The easily reconfigurable predictive controllers are supplemented with a mechanism of disturbance measurement utilization. It is done in such a way that the main advantage of the controllers – their simplicity – is maintained. The predictive controllers under consideration are based on fuzzy Takagi–Sugeno (TS) models in which step responses are used as local models. These models are supplemented with the parts describing the influence of disturbances on the outputs of the control plant. Then, the controllers are formulated in such a way that the control signals are easily generated. Efficiency and usefulness of the predictive controllers utilizing disturbance measurement is demonstrated in the example control system of a nonlinear control plant with delay.

Biometric-Based Authentication System using Rough Set Theory

Hala S. Own¹, Waheeda Al-Mayyan², and Hussein Zedan²

¹ Department of Solar and Space Research,
National Research Institute of Astronomy and Geophysics
Cairo, Egypt
halaown@gmail.com

² Software Technology Research Lab, De Montfort University
The Gateway, Leicester LE1 9BH, England, UK
{walmayyan, hzedan}@dmu.ac.uk

Abstract. In this paper we have proposed a biometric-based authentication system based on rough set theory. The system employed signature for authentication purpose. The major functional blocks of the proposed system are presented. Information is extracted as time functions of various dynamic properties of the signatures. We apply our methodology to global features extracted from a 108-users database. Thirty-one features were identified and extracted from each signature. Rough set approach has resulted in a reduced set of nine features that were found to capture the essential characteristics required for signature identification. Low error rates obtained in experiments illustrate the feasibility of using Rough Set as a promising technique for online signature identification systems.

Classification of Facial Photograph Sorting Performance Based on Verbal Descriptions

Daryl H. Hepting¹, Richard Spring¹, Timothy Maciag¹,
Katherine Arbuthnott², and Dominik Ślęzak^{3,4}

¹ Department of Computer Science, University of Regina
3737 Wascana Parkway, Regina, SK, S4S 0A2 Canada

{dhh, spring1r, maciagt}@cs.uregina.ca

² Champion College, University of Regina
3737 Wascana Parkway, Regina, SK, S4S 0A2 Canada
katherine.arbuthnott@uregina.ca

³ Institute of Mathematics, University of Warsaw
Banacha 2, 02-097 Warsaw, Poland

⁴ Infobright Inc.
Krzywickiego 34 pok. 219, 02-078 Warsaw, Poland
slezak@infobright.com

Abstract. Eyewitness identification remains an important element in judicial proceedings. It is very convincing, yet it is not very accurate. To better understand eyewitness identification, we began by examining how people understand similarity. This paper reports on analysis of study that examined how people made similarity judgements amongst a variety of facial photographs: participants were presented with a randomly ordered set of photos, with equal numbers of Caucasian (C) and First Nations (F), which they sorted based on their individual assessment of similarity. The number of piles made by the participants was not restricted. After sorting was complete, each participant was then asked to label each pile with a description of the pile's contents. Following the results of an earlier study, we hypothesize that individuals may be using different strategies to assess similarity between photos. In this analysis, we attempt to use the descriptive pile labels (in particular, related to lips and ears) as a means to uncover differences in strategies for which a classifier can be built, using the rough set attribute reduction methodology. In particular, we aim to identify those pairs of photographs that may be the key for verifying an individual's abilities and strategies when recognizing faces. The paper describes the method for data processing that enabled the comparisons based on labels. Continued success with the same technique as previously reported to filter pairs before performing the rough sets analysis, lends credibility to its use as a general method. The rough set techniques enable the identification of the sets of photograph pairs that are key to the divisions based on various strategies. This may lead to a practical test for people's abilities, as well as to inferring what discriminations people use in face recognition.

Random Musical Bands Playing in Random Forests

Miron B. Kursa¹, Elżbieta Kubera^{2,3}, Witold R. Rudnicki¹, and
Alicja A. Wieczorkowska³

¹ Interdisciplinary Centre for Mathematical and Computational Modelling,
University of Warsaw, Pawińskiego 5A, 02-106 Warsaw, Poland,

{M.Kursa, W.Rudnicki}@icm.edu.pl

² University of Life Sciences in Lublin, Akademicka 13, 20-950 Lublin, Poland

elzbieta.kubera@up.lublin.pl

³ Polish-Japanese Institute of Information Technology,
Koszykowa 86, 02-008 Warsaw, Poland,

alicja@poljap.edu.pl

Abstract. In this paper we investigate the problem of recognizing the full set of instruments playing in a sound mix. Random mixes of 2-5 instruments (out of 14) were created and parameterized to obtain experimental data. Sound samples were taken from 3 audio data sets. For classification purposes, we used a battery of one-instrument sensitive random forest classifiers, and obtained quite good results.

An Empirical Comparison of Rule Sets Induced by LERS and Probabilistic Rough Classification

Jerzy W. Grzymala-Busse^{1,2}, Shantan R. Marepally¹, and Yiyu Yao³

¹ Department of Electrical Engineering and Computer Science University of Kansas,
Lawrence, KS 66045, USA

² Institute of Computer Science, Polish Academy of Sciences,
01-237 Warsaw, Poland
{jerzy, shantan}@ku.edu

³ Department of Computer Science, University of Regina,
Regina, Saskatchewan, Canada S4S 0A2
yyao@cs.uregina.ca

Abstract. In this paper we present results of an experimental comparison (in terms of an error rate) of rule sets induced by the LERS data mining system with rule sets induced using the probabilistic rough classification (PRC). As follows from our experiments, the performance of LERS (possible rules) is significantly better than the best rule sets induced by PRC with any threshold (two-tailed test, 5% significance level). Additionally, the LERS possible rule approach to rule induction is significantly better than the LERS certain rule approach (two-tailed test, 5% significance level).

DRSA Decision Algorithm Analysis in Stylometric Processing of Literary Texts

Urszula Stańczyk

Institute of Informatics, Silesian University of Technology,
Akademicka 16, 44-100 Gliwice, Poland

Abstract. When the indiscernibility relation, fundamental to Classical Rough Set Approach, is substituted with dominance relation, it results in Dominance-Based Rough Set Approach to data analysis. It enables support not only for nominal classification tasks, but also when ordinal properties on attribute values can be observed [?], making DRSA methodology well suited for stylometric processing of texts. Stylometry involves handling quantitative features of texts leading to characterisation of authors to the point of recognition of their individual writing styles. As always, selection of attributes is crucial to classification accuracy, as is the construction of a decision algorithm. When minimal cover gives unsatisfactory results, and all rules on examples algorithm returns very high number of rules, usually constraints are imposed by selection of some reduct and limiting the decision algorithm by including within it only rules with certain support. However, reducts are typically numerous and within them some of conditional attributes are used more often than others, which is also true for conditions specified by decision rules. The paper presents observations how the frequency of usage for features reflects on the performance of decision algorithms resulting from selection of rules with conditional attributes exploited most and least often.

Blind Music Timbre Source Isolation by Multiresolution Comparison of Spectrum Signatures

Xin Zhang¹, Wenxin Jiang², Zbigniew W. Raś^{2,4,5}, and Rory Lewis³

¹ Univ. of North Carolina, Dept. of Math. and Comp. Science,
Pembroke, NC 28372, USA

² Univ. of North Carolina, Dept. of Comp. Science,
Charlotte, NC 28223, USA

³ Univ. of Colorado, Dept. of Comp. Science,
Colorado Springs, CO 80933, USA

⁴ Polish-Japanese Institute of Information Technology,
02-008 Warsaw, Poland

⁵ Polish Academy of Sciences, Institute of Comp. Science,
01-237 Warsaw, Poland

xin.zhang@uncp.edu, {wjiang3, ras}@uncc.edu, rlewis@eas.uccs.edu

Abstract. Automatic indexing of music instruments for multi-timbre sounds is challenging, especially when partials from different sources are overlapping with each other. Temporal features, which have been successfully applied in monophonic sound timbre identification, failed to isolate music instrument in multi-timbre objects, since the detection of the start and end position of each music segment unit is very difficult. Spectral features of MPEG7 and other popular features provide economic computation but contain limited information about timbre. Being compared to the spectral features, spectrum signature features has less information loss; therefore may identify sound sources in multi-timbre music objects with higher accuracy. However, the high dimensionality of spectrum signature feature set requires intensive computing and causes estimation efficiency problem. To overcome these problems, the authors developed a new multi-resolution system with an iterative spectrum band matching device to provide fast and accurate recognition.

Rough Sets for Solving Classification Problems in Computational Neuroscience

Tomasz G. Smolinski and Astrid A. Prinz

Department of Biology, Emory University, Atlanta, GA 30322, USA
{tomasz.smolinski, astrid.prinz}@emory.edu

Abstract. Understanding cellular properties of neurons is central in neuroscience. It is especially important in light of recent discoveries suggesting that similar neural activity can be produced by cells with quite disparate characteristics. Unfortunately, due to experimental constraints, analyzing how the activity of neurons depends on cellular parameters is difficult. Computational modeling of biological neurons allows for exploration of many parameter combinations, without the necessity of a large number of “wet” experiments. However, analysis and interpretation of often very extensive databases of models can be hard. Thus there is a need for efficient algorithms capable of mining such data. This article proposes a rough sets-based approach to the problem of classifying functional and non-functional neuronal models. In addition to presenting a successful application of the theory of rough sets in the field of computational neuroscience, we are hoping to foster with this paper a greater interest among the members of the rough sets community to explore the plethora of important problems in that field.

Towards Approximate SQL – Infobright’s Approach

Dominik Ślęzak and Marcin Kowalski

¹ Institute of Mathematics, University of Warsaw
Banacha 2, 02-097 Warsaw, Poland

² Infobright Inc., Poland
Krzywickiego 34 pok. 219, 02-078 Warsaw, Poland
{slezak,mkowalski}@infobright.com

Abstract. We discuss various ideas how to implement execution of approximate SQL statements within Infobright database engine. We first outline the engine’s architecture, which is designed entirely to work with standard SQL. We then discuss several possible extensions towards approximate querying and point out at some analogies with the principles of the theory of rough sets. Finally, we present the results of experiments obtained at the prototype level, both with respect to the speed of query execution and the accuracy of approximate answers.

Keywords: Analytic Database Engines, Approximate Querying, Rough Sets and Granular Computing, Infobright Community and Enterprise Editions.

A Protein Classifier Based on SVM by Using the Voxel Based Descriptor

Georgina Mirceva, Andreja Naumoski, and Danco Davcev

Faculty of Electrical Engineering and Information Technologies,
Univ. Ss. Cyril and Methodius, Skopje, Macedonia
{georgina, andrejna, etfdav}feit.ukim.edu.mk

Abstract. The tertiary structure of a protein molecule is the main factor which determines its function. All information required for a protein to be folded in its natural structure, is coded in its amino acid sequence. The way this sequence folds in the 3D space can be used for determining its function. With the technology innovations, the number of determined protein structures increases every day, so improving the efficiency of protein structure retrieval and classification methods becomes an important research issue. In this paper, we propose a novel protein classifier. Our classifier considers the conformation of protein structure in the 3D space. Namely, our voxel based protein descriptor is used for representing the protein structures. Then, the Support Vector Machine method (SVM) is used for classifying protein structures. The results show that our classifier achieves 78.83% accuracy, while it is faster than other algorithms with comparable accuracy.

Explicit Neural Network-Based Nonlinear Predictive Control with Low Computational Complexity

Maciej Ławryńczuk

Institute of Control and Computation Engineering, Warsaw University of Technology
ul. Nowowiejska 15/19, 00-665 Warsaw, Poland, tel. +48 22 234-76-73
M.Lawrynczuk@ia.pw.edu.pl

Abstract. This paper describes a nonlinear Model Predictive Control (MPC) algorithm based on neural models. Two neural models are used on-line: from a dynamic model the free trajectory (the influence of the past) is determined, the second neural network approximates the time-varying feedback law. In consequence, the algorithm is characterised by very low computational complexity because the control signal is calculated explicitly, without any on-line optimisation. Moreover, unlike other suboptimal MPC approaches, the necessity of model linearisation and matrix inversion is eliminated. The presented algorithm is compared with linearisation-based MPC and MPC with full nonlinear optimisation in terms of accuracy and computational complexity.

Solution of the Inverse Heat Conduction Problem by Using the ABC Algorithm

Edyta Hetmaniok, Damian Słota, and Adam Zielonka

Institute of Mathematics,
Silesian University of Technology,
Kaszubska 23, 44-100 Gliwice, Poland

Abstract. In this paper, a numerical method of solving the inverse heat conduction problem based on the respectively new tool for combinational optimization, named the Artificial Bee Colony algorithm (ABC), is presented. In the first step, the direct heat conduction problem, associated to the considered inverse heat conduction problem, is solved by using the finite difference method. In the second step, the proper functional, based on the least squares method, is minimized by using the ABC algorithm, giving the solution of the considered problem. An example illustrating the precision and effectiveness of the method is also shown. The proposed approach is original and promising.

Application of Fuzzy Wiener Models in Efficient MPC Algorithms

Piotr M. Marusak

Institute of Control and Computation Engineering, Warsaw University of Technology,
ul. Nowowiejska 15/19, 00-665 Warszawa, Poland,
P.Marusak@ia.pw.edu.pl

Abstract. Efficient Model Predictive Control (MPC) algorithms based on fuzzy Wiener models are proposed in the paper. Thanks to the form of the model the prediction of the control plant output can be easily obtained. It is done in such a way that the MPC algorithm is formulated as a numerically efficient quadratic optimization problem. Moreover, inversion of the static process model, used in other approaches, is avoided. Despite its relative simplicity the algorithm offers practically the same performance as the MPC algorithm in which control signals are generated after solving a nonlinear optimization problem and outperforms the MPC algorithm based on a linear model. The efficacy of the proposed approach is demonstrated in the control system of a nonlinear control plant.

Multicriteria Subjective Reputation Management Model

Michał Majdan^{1,2} and Włodzimierz Ogryczak¹

¹ Institute of Control and Computation Engineering, Warsaw University of Technology, Warsaw, Poland

² National Institute of Telecommunications, Warsaw, Poland

Abstract. The most widely used reputation models assume uniform users' preference structure. In this paper a new reputation management model is presented. It is focused on aggregation of community wide reputation in situation when agents do not share the same preference structure. The reputation is interpreted as vectors of attributes that represent several reputation evaluation criteria. Outcomes of the criteria are transformed by the utility functions and assigned subjective probabilities so that the subjective expected utility values can be obtained. Subjective expected utilities are further aggregated by the weighted ordered weighted average (WOWA) operator. The expressive power of subjective utilities concept along with the WOWA aggregation technique provides the reputation management system with a capability to model various preference structures. It is shown with an illustrative example.

Application of Fuzzy Preference Based Rough Set Model to Condition Monitoring

Xiaomin Zhao, Ming J Zuo, and Tejas Patel

Department of Mechanical Engineering, University of Alberta,
Edmonton, AB, T6G 2G8, Canada

Abstract. Parameters that vary monotonically with fault development are useful in condition monitoring, but not easy to find especially for complex systems. A method using fuzzy preference based rough set model and principle component analysis (PCA) is proposed to generate such an indicator. The fuzzy preference based rough set model is employed to evaluate the monotonic trends of features reflecting machinery conditions. PCA is used to condense the informative features and generate an indicator which can represent the development of machine health condition. The effectiveness of the proposed method is tested for damage level detection of an impeller in a slurry pump.

Graph-based Optimization Method for Information Diffusion and Attack Durability in Networks

Zbigniew Tarapata and Rafał Kasprzyk

Military University of Technology, Cybernetics Faculty,
Gen. S. Kaliskiego Str. 2, 00-908 Warsaw, Poland
{zbigniew.tarapata, rafal.kasprzyk}@wat.edu.pl

Abstract. In this paper we present a graph-based optimization method for information diffusion and attack durability in networks using properties of Complex Networks. We show why and how Complex Networks with Scale Free and Small World features can help optimize the topology of networks or indicate weak or strong elements of the network. We define some efficiency measures of information diffusion and attack durability in networks. Using these measures we formulate multicriteria optimization problem to choose the best network. We show a practical example of using the method based on an analysis of a few social networks.

Paraconsistent and Approximate Semantics for the OWL 2 Web Ontology Language

Linh Anh Nguyen

Institute of Informatics, University of Warsaw
Banacha 2, 02-097 Warsaw, Poland
nguyen@mimuw.edu.pl

Abstract. We introduce a number of paraconsistent semantics, including three-valued and four-valued semantics, for the description logic SROIQ, which is the logical foundation of OWL 2. We then study the relationship between the semantics and paraconsistent reasoning in SROIQ w.r.t. some of them through a translation into the traditional semantics. We also present a formalization of rough concepts in SROIQ.

Representation of Granularity for Non-Euclidian Relational Data by Jaccard Coefficients and Binary Classifications

Shoji Hirano and Shusaku Tsumoto

Department of Medical Informatics, Shimane University, School of Medicine
89-1 Enya-cho, Izumo, Shimane 693-8501, Japan
E-mails: hirano@ieee.org, tsumoto@computer.org

Abstract. In this paper we present a method for representing the granularity for asymmetric, non-Euclidean relational data. It firstly builds a set of binary classifications based on the directional similarity from each object. After that, the strength of discrimination knowledge is quantified as the indiscernibility of objects based on the Jaccard similarity coefficients between the classifications. Fine but weak discrimination knowledge supported by the small number of binary classifications is more likely to be coarsened than those supported by the large number of classifications, and coarsening of discrimination knowledge causes the merging of objects. According to this feature, we represent the hierarchical structure of data granules by a dendrogram generated by applying the complete-linkage hierarchical grouping method to the derived indiscernibility. This enables users to change the coarseness of discrimination knowledge and thus to control the size of granules.

Information Systems in Modeling Interactive Computations on Granules

Andrzej Skowron¹ and Piotr Wasilewski²

¹ Institute of Mathematics

Warsaw University

Banacha 2, 02-097 Warsaw, Poland

² Computational Intelligence Laboratory

Department of Electrical and Computer Engineering

University of Manitoba

75A Chancellor's Circle, MB R3T 5V6 Winnipeg, Canada

skowron@mimuw.edu.pl, piotr@ee.umanitoba.ca

Abstract. In this paper we discuss the importance of information systems in modeling interactive computations performed on (complex) granules and propose a formal approach to interactive computations based on information systems. The basic concepts of information systems and rough sets are interpreted in the framework of interactive computations. We also show that information systems can be used for modeling more advanced forms of interactions such as hierarchical ones. The role of hierarchical interactions is emphasized in modeling interactive computations. Some illustrative examples of interactions used in the hierarchical multimodal classification method as well as in the ACT-R 6.0 system are reported.

Keywords: interactive computing, interactive systems, multi-agent systems, rough sets, granular computing, wisdom technology.

Distributed Representations to Detect Higher Order Term Correlations in Textual Content

Pinar Öztürk, Rajendra Prasath, and Hans Moen

Department of Computer and Information Science (IDI)
Norwegian University of Science and Technology (NTNU),
Sem Sælands Vei 7-9, NO - 7491, Trondheim, Norway,
{pinar,rajendra}@idi.ntnu.no; hnsmoen@gmail.com,

Abstract. Case Based Reasoning(CBR), an artificial intelligence technique, solves new problem by reusing solutions of previously solved similar cases. In conventional CBR, cases are represented in terms of structured attribute-value pairs. Acquisition of cases, either from domain experts or through manually crafting attribute-value pairs from incident reports, constitutes the main reason why CBR systems have not been more common in industries. Manual case generation is a laborious, costlier and time consuming task. Textual CBR (TCBR) is an emerging line that aims to apply CBR techniques on cases represented as textual descriptions. Similarity of cases is based on the similarity between their constituting features. Conventional CBR benefits from employing domain specific knowledge for similarity assessment. Correspondingly, TCBR needs to involve higher-order relationships between features, hence domain specific knowledge. In addition, the term order has also been contended to influence the similarity assessment. This paper presents an account where features and cases are represented using a distributed representation paradigm that captures higher-order relations among features as well as term order information.

Author Index

- Al-Mayyan, Waheeda, 560
Alkhalid, Abdulaziz, 438
Almeida-Luz, Sónia M., 534
Arbuthnott, Katherine, 570
Avdelidis, Konstantinos, 100
- Bandyopadhyay, Sanghamitra, 30
Banerjee, Mohua, 247, 317
Bazan, Jan, 4
Biswas, Santosh, 50
Błaszczyszki, Jerzy, 148, 392, 402
Bonikowski, Zbigniew, 337
Bošnjak, Matko, 4
- Chang, Chaio-Chen, 514
Chen, Xuguang, 356
Chen, Ze, 4
Chikalov, Igor, 412, 438
Chin, Yang-Chieh, 514
Chitcharoen, Doungrat, 418
Choroś, Kazimierz, 120
Ciucci, Davide, 257
Czyżewski, Andrzej, 70, 80
- Dalka, Piotr, 70, 80
Davcev, Danco, 640
Deckert, Magdalena, 148
Diker, Murat, 287
Dimoulas, Charalampos, 100
Doherty, Patrick, 327
Duch, Włodzisław, 178
Düntsch, Ivo, 386
- Elouedi, Zied, 366
- Farion, Ken, 207
- Gamberger, Dragan, 4
Gao, Juan, 4
Gawkowski, Piotr, 524
Gediga, Günther, 386
Gomes, João Bártolo, 168
Gómez-Pulido, Juan A., 534
Gomolińska, Anna, 227
Grabowski, Adam, 307
Grzymala-Busse, Jerzy W., 590
- Guan, Lihe, 4
- Hepting, Daryl H., 570
Hetmaniok, Edyta, 659
Hirano, Shoji, 721
Hu, Feng, 4
Hu, Qinghua, 347
Huang, Jin, 197
Huang, Tian-Hsiang, 4
Hubballi, Neminath, 50
Huber, Martin, 217
- Janeczek, Bartosz, 428
Janusz, Andrzej, 4, 130
Jastrzębska, Magdalena, 307
Jeong, SungHwan, 504
Jiang, Wenxin, 610
- Kalliris, George, 100
Kasprzyk, Rafał, 698
Kaszuba, Katarzyna, 110
Khan, Md. Aquil, 247
Klement, William, 207
Komorowska, Agnieszka, 524
Komorowski, Michał, 456
Kostek, Bożena, 110
Kośmicki, Piotr S., 494
Kowalski, Marcin, 630
Krishna, Gopal, 317
Kryszkiewicz, Marzena, 60
Kubera, Elżbieta, 580
Kuczyńska, Monika Anna, 524
Kursa, Miron B., 580
- Lasek, Piotr, 60
Lee, JoonWhoan, 504
Lewis, Rory, 610
Lin, Chiun-Sin, 514
Lingras, Pawan, 366
Luo, ChuanJiang, 4
Luo, Huan, 4
Ławryńczuk, Maciej, 649
- Maciąg, Timothy, 570
Majdan, Michał, 678
Małyszko, Dariusz, 40

- Marepally, Shantan R., 590
 Marusak, Piotr M., 551, 669
 Maszczyk, Tomasz, 178
 Matwin, Stan, 197
 Maulik, Ujjwal, 30
 McLachlan, Geoffrey J., 4
 Menasalvas, Ernestina, 168
 Michalowski, Martin, 207
 Mirceva, Georgina, 640
 Moen, Hans, 740
 Moshkov, Mikhail, 412, 438
 Mukerjee, Amitabha, 317

 Nakata, Michinori, 376
 Nandi, Sukumar, 50
 Napierała, Krystyna, 138, 158
 Naumoski, Andreja, 640
 Nguyen Hung Son, 4
 Nguyen Linh Anh, 710
 Nguyen Tuan Trung, 446
 Nikulin, Vladimir, 4

 Ogryczak, Włodzimierz, 678
 Ou, C.R., 466
 Ou, Chung-Ming, 466
 Own, Hala S., 560
 Öztürk, Pinar, 740

 Papanikolaou, George, 100
 Park, EunJong, 504
 Patel, Tejas, 688
 Pathak, Abhinav, 317
 Patra, Bidyut Kr., 50
 Pattaraintakorn, Puntip, 418
 Pawlaczyk, Piotr, 120
 Peters, James F., 277
 Plewczyński, Dariusz, 30
 Podraza, Roman, 428
 Prasath, Rajendra, 544, 740
 Prinz, Astrid A., 620

 Ramanna, Sheela, 277
 Raś, Zbigniew W., 610
 Rudnicki, Witold R., 580
 Rybiński, Henryk, 476, 484
 Ryzko, Dominik, 476

 Saha, Indrajit, 30
 Sakai, Hiroshi, 376
 Sánchez-Pérez, Juan M., 534
 Sayyad Shirabad, Jelber, 197

 Shen, Yuanxia, 4
 Schön, Torsten, 217
 Skowron, Andrzej, 297, 730
 Słota, Damian, 659
 Słowiński, Roman, 2, 392, 402
 Smolinski, Tomasz G., 620
 Sousa, Pedro A.C., 168
 Spring, Richard, 570
 Stańczyk, Urszula, 600
 Stawicki, Sebastian, 20
 Stefanowski, Jerzy, 138, 148, 158, 392
 Stell, John G., 267
 Stepaniuk, Jarosław, 40, 297
 Su, Jiang, 197
 Sycara, Katia, 1
 Szałas, Andrzej, 327
 Szczuko, Piotr, 90
 Szeląg, Marcin, 402
 Ślęzak, Dominik, 187, 376, 570, 630

 Tarapata, Zbigniew, 698
 Trabelsi, Salsabil, 366
 Tsumoto, Shusaku, 721
 Tsymbal, Alexey, 217
 Tzeng, Gwo-Hshiang, 514

 Vega-Rodríguez, Miguel A., 534

 Wang, Guoyin, 4
 Wasilewski, Piotr, 277, 730
 Więch, Przemysław, 484
 Widz, Sebastian, 187
 Wieczorkowska, Alicja A., 580
 Wilk, Szymon, 148, 158, 207
 Wojnarowski, Piotr, 20
 Wojnarski, Marcin, 4, 20
 Wolski, Marcin, 237

 Yang, Yongbin, 347
 Yao, Yiyu, 590

 Zedan, Hussein, 560
 Zhang, Xin, 610
 Zhao, Xiaomin, 688
 Zhu, Pengfei, 347
 Ziarko, Wojciech, 356
 Zielonka, Adam, 659
 Zielosko, Beata, 412
 Zuo, Ming J, 688
 Żwan, Paweł, 110