

Jacek Koronacki, Zbigniew W. Raś, Sławomir T. Wierzchoń, and
Janusz Kacprzyk (Eds.)

Advances in Machine Learning II

Studies in Computational Intelligence, Volume 263

Editor-in-Chief

Prof. Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6
01-447 Warsaw
Poland
E-mail: kacprzyk@ibspan.waw.pl

Further volumes of this series can be found on our homepage:
springer.com

Vol. 243. Imre J. Rudas, János Fodor, and Janusz Kacprzyk (Eds.)
Towards Intelligent Engineering and Information Technology, 2009
ISBN 978-3-642-03736-8

Vol. 244. Ngoc Thanh Nguyen, Radosław Piotr Katarzyniak, and Adam Janiak (Eds.)
New Challenges in Computational Collective Intelligence, 2009
ISBN 978-3-642-03957-7

Vol. 245. Oleg Okun and Giorgio Valentini (Eds.)
Applications of Supervised and Unsupervised Ensemble Methods, 2009
ISBN 978-3-642-03998-0

Vol. 246. Thanasis Daradoumis, Santi Caballé, Joan Manuel Marqués, and Fatos Xhafa (Eds.)
Intelligent Collaborative e-Learning Systems and Applications, 2009
ISBN 978-3-642-04000-9

Vol. 247. Monica Bianchini, Marco Maggini, Franco Scarselli, and Lakshmi C. Jain (Eds.)
Innovations in Neural Information Paradigms and Applications, 2009
ISBN 978-3-642-04002-3

Vol. 248. Chee Peng Lim, Lakshmi C. Jain, and Satchidananda Dehuri (Eds.)
Innovations in Swarm Intelligence, 2009
ISBN 978-3-642-04224-9

Vol. 249. Wesam Ashour Barbakh, Ying Wu, and Colin Fyfe
Non-Standard Parameter Adaptation for Exploratory Data Analysis, 2009
ISBN 978-3-642-04004-7

Vol. 250. Raymond Chiong and Sandeep Dhakal (Eds.)
Natural Intelligence for Scheduling, Planning and Packing Problems, 2009
ISBN 978-3-642-04038-2

Vol. 251. Zbigniew W. Ras and William Ribarsky (Eds.)
Advances in Information and Intelligent Systems, 2009
ISBN 978-3-642-04140-2

Vol. 252. Ngoc Thanh Nguyen and Edward Szczerbicki (Eds.)
Intelligent Systems for Knowledge Management, 2009
ISBN 978-3-642-04169-3

Vol. 253. Roger Lee and Naohiro Ishii (Eds.)
Software Engineering Research, Management and Applications 2009, 2009
ISBN 978-3-642-05440-2

Vol. 254. Kyandoghere Kyamakya, Wolfgang A. Halang, Herwig Unger, Jean Chamberlain Chedjou, Nikolai F. Rulkov, and Zhong Li (Eds.)
Recent Advances in Nonlinear Dynamics and Synchronization, 2009
ISBN 978-3-642-04226-3

Vol. 255. Catarina Silva and Bernardete Ribeiro
Inductive Inference for Large Scale Text Classification, 2009
ISBN 978-3-642-04532-5

Vol. 256. Patricia Melin, Janusz Kacprzyk, and Witold Pedrycz (Eds.)
Bio-inspired Hybrid Intelligent Systems for Image Analysis and Pattern Recognition, 2009
ISBN 978-3-642-04515-8

Vol. 257. Oscar Castillo, Witold Pedrycz, and Janusz Kacprzyk (Eds.)
Evolutionary Design of Intelligent Systems in Modeling, Simulation and Control, 2009
ISBN 978-3-642-04513-4

Vol. 258. Leonardo Franco, David A. Elizondo, and José M. Jerez (Eds.)
Constructive Neural Networks, 2009
ISBN 978-3-642-04511-0

Vol. 259. Kasthurirangan Gopalakrishnan, Halil Ceylan, and Nii O. Attoh-Okine (Eds.)
Intelligent and Soft Computing in Infrastructure Systems Engineering, 2009
ISBN 978-3-642-04585-1

Vol. 260. Edward Szczerbicki and Ngoc Thanh Nguyen (Eds.)
Smart Information and Knowledge Management, 2009
ISBN 978-3-642-04583-7

Vol. 261. Nadia Nedjah, Leandro dos Santos Coelho, and Luiza de Macedo de Moura (Eds.)
Multi-Objective Swarm Intelligent Systems, 2009
ISBN 978-3-642-05164-7

Vol. 262. Jacek Koronacki, Zbigniew W. Raś, Sławomir T. Wierchoń, and Janusz Kacprzyk (Eds.)
Advances in Machine Learning I, 2010
ISBN 978-3-642-05176-0

Vol. 263. Jacek Koronacki, Zbigniew W. Raś, Sławomir T. Wierchoń, and Janusz Kacprzyk (Eds.)
Advances in Machine Learning II, 2010
ISBN 978-3-642-05178-4

Jacek Koronacki, Zbigniew W. Raś,
Sławomir T. Wierzchoń, and Janusz Kacprzyk (Eds.)

Advances in Machine Learning II

Dedicated to the Memory of Professor
Ryszard S. Michalski

Prof. Jacek Koronacki
Institute of Computer Science
Polish Academy of Sciences
ul. Ordona 21
01-237 Warsaw
Poland
E-mail: korona@ipipan.waw.pl

Prof. Sławomir T. Wierchoń
Institute of Computer Science
Polish Academy of Sciences
ul. Ordona 21
01-237 Warsaw
Poland
E-mail: stw@ipipan.waw.pl

Prof. Zbigniew W. Raś
Woodward Hall 430C
University of North Carolina
9201 University City Blvd.
Charlotte, N.C. 28223
USA
E-mail: ras@uncc.edu,
ras@pjwstk.edu.pl

Prof. Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6
01-447 Warsaw
Poland
E-mail: kacprzyk@ibspan.waw.pl

ISBN 978-3-642-05178-4

e-ISBN 978-3-642-05179-1

DOI 10.1007/978-3-642-05179-1

Studies in Computational Intelligence

ISSN 1860-949X

Library of Congress Control Number: 2009940119

© 2010 Springer-Verlag Berlin Heidelberg

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Typeset & Cover Design: Scientific Publishing Services Pvt. Ltd., Chennai, India.

Printed in acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Foreword

Professor Richard S. Michalski passed away on September 20, 2007. Once we learned about his untimely death we immediately realized that we would no longer have with us a truly exceptional scholar and researcher who for several decades had been influencing the work of numerous scientists all over the world - not only in his area of expertise, notably machine learning, but also in the broadly understood areas of data analysis, data mining, knowledge discovery and many others. In fact, his influence was even much broader due to his creative vision, integrity, scientific excellence and exceptionally wide intellectual horizons which extended to history, political science and arts.

Professor Michalski's death was a particularly deep loss to the whole Polish scientific community and the Polish Academy of Sciences in particular. After graduation, he began his research career at the Institute of Automatic Control, Polish Academy of Science in Warsaw. In 1970 he left his native country and held various prestigious positions at top US universities. His research gained impetus and he soon established himself as a world authority in his areas of interest – notably, he was widely considered a father of machine learning.

His contacts with the Polish scientific community were very close over all the years; in the last couple of years he was an affiliate scientist at the Institute of Computer Science, Polish Academy of Sciences, Warsaw. This relation culminated some years ago with his election to the rank of Foreign Member of the Polish Academy of Sciences, a distinction granted to only a small number of world-wide best scientists, including numerous Nobel Prize and other prestigious awards winners.

Professor Michalski was one of those active members of the Polish Academy of Sciences who were always interested in solving whatever problems we had, always ready to help us in shaping the research policy of the Academy and discuss with us all difficult issues that are these days unavoidable in any large and prestigious research organization with so many strong links to science worldwide. He was always ready to offer us his deep understanding and scholarly vision of the future of the human scientific endeavor. As President of the Polish Academy of Sciences I sense very personally an enormous loss coming from no longer being able to ask for his opinion and advice.

I wish to congratulate the editors of these scholarly volumes, Professors Jacek Koronacki, Zbigniew Raś, Sławomir T. Wierchoń and Janusz Kacprzyk, for their initiative to pay the tribute to the memory of Professor Michalski. Having known him for many years they realized that the best way to honor his life achievements would be to prepare a collection of high quality papers on topics broadly perceived as Professor Michalski's main interest and to present *in memoriam* volumes of the contributions written by those who had the luck to be his friends or, at least, to meet him on various occasions. I am really impressed that so many prominent authors have accepted the invitation and I thank all of them most deeply.

I believe the memory of Professor Richard S. Michalski should remain with us for ever. Hopefully, these volumes will contribute to reaching this objective in the most appropriate and substantial way.

Professor Michał Kleiber
President
Polish Academy of Sciences

Preface

This is the second volume of a large two-volume editorial project we wish to dedicate to the memory of the late Professor Ryszard S. Michalski who passed away in 2007. He was one of the fathers of machine learning, an exciting and relevant, both from the practical and theoretical points of view, area in modern computer science and information technology. His research career started in the mid-1960s in Poland, in the Institute of Automation, Polish Academy of Sciences in Warsaw, Poland. He left for the USA in 1970, and since then had worked there at various universities, notably, at the University of Illinois at Urbana – Champaign and finally, until his untimely death, at George Mason University. We, the editors, had been lucky to be able to meet and collaborate with Ryszard for years, indeed some of us knew him when he was still in Poland. After he started working in the USA, he was a frequent visitor to Poland, taking part at many conferences until his death. We had also witnessed with a great personal pleasure honors and awards he had received over the years, notably when some years ago he was elected Foreign Member of the Polish Academy of Sciences among some top scientists and scholars from all over the world, including Nobel prize winners.

Professor Michalski's research results influenced very strongly the development of machine learning, data mining, and related areas. Also, he inspired many established and younger scholars and scientists all over the world.

We feel very happy that so many top scientists from all over the world agreed to pay the last tribute to Professor Michalski by writing papers in their areas of research. These papers will constitute the most appropriate tribute to Professor Michalski, a devoted scholar and researcher. Moreover, we believe that they will inspire many newcomers and younger researchers in the area of broadly perceived machine learning, data analysis and data mining.

The papers included in the two volumes, Machine Learning I and Machine Learning II, cover diverse topics, and various aspects of the fields involved. For convenience of the potential readers, we will now briefly summarize the contents of the particular chapters.

Part I, "General Issues" is concerned with some more general issues and problems that are relevant in various areas, notably in machine learning, data mining, knowledge discovery, and their applications in a multitude of domains.

- **Witold Pedrycz** ("Knowledge-Oriented and Distributed Unsupervised Learning for Concept Elicitation") discusses a new direction of unsupervised learning and concept formation in which both domain knowledge and experimental evidence (data) are considered together. This is a reflection of a certain paradigm which could be referred to as knowledge-oriented clustering or knowledge mining (as opposed to data mining). The author presents the main concepts and algorithmic details. The distributed way of forming information granules which is

realized at the level of individual locally available data gives rise to higher order information granules (type-2 fuzzy sets, in particular).

- **Andrzej Skowron and Marcin Szczuka** (“Toward Interactive Computations: A Rough-Granular Approach”) present an overview of Rough Granular Computing (RGC) approach to modeling complex systems and processes. They discuss the granular methodology in conjunction with paradigms originating in rough sets theory, such as approximation spaces. The authors attempt to show the methodology aimed at the construction of complex concepts from raw data in ahierarchical manner. They illustrate, how the inclusion of domain knowledge, relevant ontologies, and interactive consensus finding leads to more poerful granular models for processes and systems.
- **Stan Matwin and Tomasz Szapiro** (“Data Privacy: From Technology to Economics”) deal with an omnipresent and increasingly relevant problem of data privacy. They attempt to relate two different approaches to data privacy: the technological approach, embodied in the current privacy-preserving data mining work, and the economic regulations approach. The authors claim that none of these two approaches alone will be able to address the increasingly important data privacy issues. They advocate a hybrid system, combining both approaches in a complementary manner. A view of privacy is presented in the context of an accepted taxonomy of economic goods, stating the question: if privacy is exchanged and traded, then what kind of good is it? The authors also show that, viewed in the light of an established economic growth theory, the involvement of privacy in the growth process leads to a steady state growth.
- **Phillipa M. Avery and Zbigniew Michalewicz** (“Adapting to Human Gamers using Coevolution”) consider a challenging task of how to mimic a human ability to adapt, and create a computer player that can adapt to its opponent’s strategy. Without this ability to adapt, no matter how good a computer player is, given enough time human players may learn to adapt to the strategy used, and routinely defeat the computer player. However, by having an adaptive strategy for a computer player, the challenge it provides is ongoing. Additionally, a computer player that adapts specifically to an individual human provides a more personal and tailored game play experience. To address specifically this last need, the authors investigate the creation of such a computer player. By creating a computer player that changes its strategy with influence from the human strategy, it is shown that the holy grail of gaming, an individually tailored gaming experience, is indeed possible. A computer player for the game of TEMPO, a zero sum military planning game, is designed. The player was created through a process that reversely engineers the human strategy and uses it to co-evolve the computer player.
- **Mirsad Hadzikadic and Min Sun** (“Wisdom of Crowds in the Prisoner’s Dilemma Context”) provide a new way of making decisions by using the wisdom of crowds (collective wisdom) to handle continuous decision making problems, especially in a complex and rapidly changing world. By simulating the Prisoner’s Dilemma as a complex adaptive system, key criteria that separate a wise crowd from an irrational one are investigated, and different aggregation strategies are suggested based on different environments.

Part II, “Logical and Relational Learning, and Beyond”, is concerned with two very important, well founded, and successful general paradigms for machine learning that are based on logic and relational analyses.

- **Marenglen Biba, Stefano Ferilli, and Floriana Esposito** (“Towards Multistrategic Statistical Relational Learning”) discuss statistical relational learning, a growing field in machine learning that aims at the integration of logic-based learning approaches with probabilistic graphical models. Learning models in statistical relational learning consist in learning the structure (logical clauses in Markov logic networks) and the parameters (weights for each clause in Markov logic networks). Markov logic networks have been successfully applied to problems in relational and uncertain domains. So far the statistical relational learning models have mostly used the expectation-maximization (EM) for learning statistical parameters under missing values. In the paper, two frameworks for integrating abduction in the statistical relational learning models are proposed. The first approach integrates logical abduction with structure and parameter learning of Markov logic networks in a single step. During structure search, clause evaluation is performed by first trying to logically abduce missing values in the data and then by learning optimal pseudo-likelihood parameters using the completed data. The second approach integrates abduction with the structural EM by performing logical abductive inference in the E-step and then by trying to maximize parameters in the M-step.
- **Luc De Raedt** (“About Knowledge and Inference in Logical and Relational Learning”) gives a gentle introduction to the use of knowledge, logic and inference in machine learning which can be regarded as a reinterpretation and revisiting of Ryszard Michalski’s work “A theory and methodology of inductive learning” within the framework of logical and relational learning. At the same time some contemporary issues surrounding the integration of logical and probabilistic representations and types of reasoning are introduced.
- **Marta Fraňová and Yves Kodratoff** (“Two examples of computational creativity: ILP multiple predicate synthesis and the ‘assets’ in theorem proving”) provide a precise illustration of what can be the idea of “computational creativity”, that is, the whole set of the methods by which a computer may simulate creativity. The analysis is restricted to multiple predicate learning in inductive logic programming and to program synthesis from its formal specification. The authors show heuristics the goal of which is to provide the program with some kind of inventiveness. The basic tool for computational creativity is what is called an ‘asset generator’. A detailed description of the authors’ methodology for the generation of assets in program synthesis from its formal specification is given. In a conclusion a result is provided, which is a kind of challenge for the other theorem provers, namely how to ‘invent’ a form of the Ackerman function which is recursive with respect to the second variable instead of the first variable as the usual definitions are. In inductive logic programming multiple predicate synthesis, the assets have been provided by members of the inductive logic programming community, while their methodology tries to make explicit a way to discover these assets when they are needed.

- **Jan Rauch** (“Logical Aspects of the Measures of Interestingness of Association Rules”) discusses the relations of the logical calculi of association rules and of the measures of interestingness of association rules. The logical calculi of association rules, 4ft-quantifiers, and known classes of association rules are first introduced. New 4ft-quantifiers and association rules are defined by the application of suitable thresholds to known measures of interestingness. It is proved that some of the new 4ft-quantifiers are related to known classes of association rules with important properties. It is shown that new interesting classes of association rules can be defined on the basis of other new 4ft-quantifiers, and several results concerning new classes are proved. Some open problems are mentioned.

Part III, “Text and Web Mining”, is concerned with various problems and aspects of data mining and machine learning related to a great challenge we face nowadays that is related to the constantly growing role of the Internet and its related Web services which permeate all aspects of economy and human life. The papers in this part show how tools and techniques from broadly perceived machine learning and data/text mining can help the human being to fully utilize the power of these new services.

- **Katharina Morik and Michael Wurst** (“Clustering the Web 2.0”) present two approaches to clustering in the scenario of Web 2.0 with a special concern of understandability in this new context. In contrast to the Semantic Web type approaches which advocate ontologies as a common semantics for homogeneous user groups, Web 2.0 aims at supporting heterogeneous user groups where users annotate and organize their content without a reference to a common schema so that the semantics is not made explicit. However, it can be extracted by using machine learning and hence the users are provided with new services.
- **Miroslav Kubat, Kanoksri Sarinnapakorn, and Sareewan Dendamrongvit** (“Induction in Multi-Label Text Classification Domains”) describe an original technique for automated classification of text documents. It is assumed, first, that each training or testing example can be labeled with more than two classes at the same time which has serious consequences not only for the induction algorithms but also for how we evaluate the performance of the induced classifier. Second, the examples are usually described by very many attributes which makes induction from hundreds of thousands of training examples prohibitively expensive. Results of numerical experiments on a concrete text database are provided.
- **Boris Mirkin, Susana Nascimento, and Luís Moniz Pereira** (“Cluster-Lift Method for Mapping Research”) present a method for representing research activities within a research organization by doubly generalizing them. The approach is founded on Michalski’s idea of inferential concept interpretation for knowledge transmutation within a knowledge structure taken here to be a concept tree.. To be specific, the authors concentrate on the Computer Sciences area represented by the ACM Computing Classification System (ACM-CCS). Their cluster-lift method involves two generalization steps: one on the level of individual activities (clustering) and the other on the concept structure level (lifting). Clusters are extracted from the data on similarity between ACMCCS topics according to the working in the organization. Lifting leads to conceptual

generalization of the clusters in terms of “head subjects” on the upper levels of ACM-CCS accompanied by their gaps and offshoots. A real-world example of the representation is show.

- **Marzena Kryszkiewicz, Henryk Rybiński, and Katarzyna Cichoń** (“On Concise Representations of Frequent Patterns Admitting Negation”) deal with the discovery of frequent patterns which is one of the most important issues in the data mining field. Though an extensive research has been carried out for discovering positive patterns, very little has been offered for discovering patterns with negation. One of the main difficulties concerning frequent patterns with negation is huge amount of discovered patterns as it exceeds the number of frequent positive patterns by orders of magnitude. The problem can be significantly alleviated by applying concise representations that use generalized disjunctive rules to reason about frequent patterns, both with and without negation. The authors examine three types of generalized disjunction free representations and derive the relationships between them. They also present two variants of algorithms for building such representations. The results obtained on a theoretical basis are verified experimentally.

Part IV, “Classification and Beyond”, deals with many aspects, methods, tools and techniques related to broadly perceived classification which is a key issue in many areas, notably those related to the topics of the present volume.

- **Derek Sleeman, Andy Aiken, Laura Moss, John Kinsella, and Malcolm Sim** [“A system to detect inconsistencies between a domain expert’s different perspectives on (classification) tasks”] discuss the range of knowledge acquisition, including machine learning, approaches used to develop knowledge bases for intelligent systems. Specifically, the paper focuses on developing techniques which enable an expert to detect inconsistencies in two (or more) perspectives which the expert might have on the same (classification) task. Further, the INSIGHT system is developed to provide a tool which supports domain experts exploring, and removing, the inconsistencies in their conceptualization of a task. The authors show a study of intensive care physicians reconciling two perspectives on their patients. The high level task which the physicians had set themselves was to classify, on a 5 point scale (A-E), the hourly reports produced by the Unit’s patient management system. The two perspectives provided to INSIGHT were an annotated set of patient records where the expert had selected the appropriate class to describe that snapshot of the patient, and a set of rules which are able to classify the various time points on the same 5-point scale. Inconsistencies between these two perspectives are displayed as a confusion matrix; moreover INSIGHT then allows the expert to revise both the annotated datasets (correcting data errors, and/or changing the assigned classes) and the actual rule-set. The paper concludes by outlining some of the follow-up studies planned with both INSIGHT and this general approach.
- **Eduardo R. Gomes and Ryszard Kowalczyk** (“The Dynamics of Multiagent Q-learning in Commodity Market Resource Allocation”) consider the commodity market (CM) economic model that offers a promising approach for the distributed resource allocation in large-scale distributed systems. The existing

CM-based mechanisms apply the economic equilibrium concepts, assuming that price-taking entities do not engage in strategic behaviour, and in this paper the above issue is addressed and the dynamics of strategic learning agents in a specific type of CM-based mechanism, called iterative price adjustment, is discussed. The scenario is considered in which agents use utility functions to describe preferences in the allocation and learn demand functions adapted to the market by reinforcement learning. The reward functions used during the learning process are based either on the individual utility of the agents, generating selfish learning agents, or the social welfare of the market, generating altruistic learning agents. The authors' experiments show that the market composed exclusively of selfish learning agents achieve results similar to the results obtained by the market composed of altruistic agents. Such an outcome is significant for a series of other domains where individual and social utility should be maximized but agents are not guaranteed to act cooperatively in order to achieve it or they do not want to reveal private preferences. This outcome is further analyzed, and an analysis of the agents' behaviour from the perspective of the dynamic process generated by the learning algorithm employed by them is also given. For this, a theoretical model of multiagent Q-learning with ε -greedy exploration is developed and applied in a simplified version of the addressed scenario.

- **Christian Borgelt** ("Simple Algorithms for Frequent Item Set Mining") introduces SaM, a split and merge algorithm for frequent item set mining. Its core advantages are its extremely simple data structure and processing scheme, which not only make it quite easy to implement, but also very convenient to execute on external storage, thus rendering it a highly useful method if the transaction database to mine cannot be loaded into main memory. Furthermore, the author's RELim algorithm is shown and different optimization options for both SaM and RELim are discussed. Finally, some numerical experiments comparing SaM and RELim with classical frequent item set mining algorithms (as, e.g., Apriori, Eclat and FP-growth) are given.
- **Michał Dramiński, Marcin Kierczak, Jacek Koronacki, and Jan Komorowski** ("Monte Carlo feature selection and interdependency discovery in supervised classification") consider applications of machine learning techniques in life sciences. Such applications force a paradigm shift in the way these techniques are used because rather than obtaining the best possible classifier, it is of interest which features contribute best to the classification of observations into distinct classes and what are the interdependencies between the features. A method for finding a cut-off between informative and non-informative features is given, followed by the development of a new methodology and an implementation of a procedure for determining interdependencies between informative features. The reliability of the approach rests on a multiple construction of tree classifiers. Essentially, each classifier is trained on a randomly chosen subset of the original data using only a randomly selected fraction of all of the observed features. This approach is conceptually simple yet computationally demanding. The method proposed is validated on a large and difficult task of modelling HIV-1 reverse transcriptase resistance to drugs which is a good example of the aforementioned paradigm shift.

- **Halina Kwaśnicka and Mariusz Paradowski** (“Machine Learning Methods in Automatic Image Annotation”) are concerned with image analysis and more specifically automatic image annotation which grew from such research domains as image recognition and cross-lingual machine translation. Because of an increase in computational, data storage and data transfer capabilities of today's computer technology, an automatic image annotation has become possible. Automatic image annotation methods, which have appeared during last several years, make a large use of many machine learning approaches, with clustering and classification methods as the most frequently applied techniques to annotate images. The chapter consists of three main parts. In the first, some general information concerning annotation methods is presented. In the second part, two original annotation methods are described. The last part presents experimental studies of the proposed methods.

Part V, “Neural Networks and Other Nature Inspired Approaches”, deals with the development and applications of various nature inspired paradigms, approaches and techniques. Notably, diverse aspects related to neural networks, evolutionary computation, artificial immune systems, swarm heuristics, etc. are considered, showing their potentials and applicability.

- **Nikola Kasabov** (“Integrative Probabilistic Evolving Spiking Neural Networks Utilising Quantum Inspired Evolutionary Algorithm: A Computational Framework”) considers integrative evolving connectionist systems (iECOS) that integrate principles from different levels of information processing in the brain, including cognitive, neuronal, genetic and quantum, in their dynamic interaction over time. A new framework of iECOS, called integrative probabilistic evolving spiking neural networks (ipSNN), utilizing a quantum inspired evolutionary optimization algorithm to optimize the probability parameters. Both spikes and input features in ipESNN are represented as quantum bits being in a superposition of two states (1 and 0) defined by a probability density function which allows for the state of an entire ipESNN at any time to be represented probabilistically in a quantum bit register and probabilistically optimized until convergence using quantum gate operators and a fitness function. The proposed ipESNN is a promising framework for both engineering applications and brain data modeling as it offers faster and more efficient feature selection and model optimization in a large dimensional space in addition to revealing new knowledge that is not possible to obtain using other models. As a further development of ipESNN, the neuro-genetic models – ipESNG, are indicated.
- **Boris Kryzhanovsky, Vladimir Kryzhanovsky, and Leonid Litinskii** (“Machine Learning in Vector Models of Neural Networks”) present a review and some extensions of their works related to the theory of vector neural networks. The interconnection matrix is always constructed according to the generalized Hebbian rule which is well-known in machine learning area. The main principles and ideas are emphasized. Analytical calculations are based on the probabilistic approach. The obtained theoretical results are verified via computer simulations.

- **Hongbo Liu, Ajith Abraham, and Benxian Yue** (“Nature Inspired Multi-Swarm Heuristics for Multi-Knowledge Extraction”) present a novel application of two nature inspired population-based computational optimization techniques, namely the Particle Swarm Optimization (PSO) and Genetic Algorithm (GA), for rough set reduction and multi-knowledge extraction. A Multi-Swarm Synergetic Optimization (MSSO) algorithm is presented for rough set reduction and multi-knowledge extraction. In the MSSO approach, different individuals encodes different reducts. The proposed approach discovers the best feature combinations in an efficient way to observe the change of positive region as the particles proceed throughout the search space. An attempt is made to prove that the multi-swarm synergetic optimization algorithm converges with a probability of 1 towards the global optimum. The proposed approach is shown to be very effective for multiple reduct problems and multi-knowledge extraction.
- **Tomasz Maszczyk, Marek Grochowski, and Włodzisław Duch** (“Discovering Data Structures using Meta-learning, Visualization and Constructive Neural Networks”) discuss first several visualization methods which have been used to reveal hidden data structures, thus facilitating discovery of simplest but appropriate data transformations which can then be used to build constructive neural networks. This is an efficient approach to meta-learning, based on the search for simplest models in the space of all data transformations, as shown in the paper. It can be used to solve problems with complex inherent logical structure that are very difficult for traditional machine learning algorithms.
- **Vladimir Golovko, Sergei Bezobrazov, Pavel Kachurka, and Leanid Vaitsekhovich** (“Neural Network and Artificial Immune Systems for Malware and Network Intrusion Detection”) consider neural networks and artificial immune systems as tools applicable to many problems in the area of anomaly detection and recognition. Since the existing solutions use mostly static approaches, which are based on the collection of viruses or intrusion signatures, detection and recognition of new viruses or attacks becomes a major problem. The authors discuss how to overcome this problem by integrating neural networks and artificial immune systems for virus and intrusion detection, as well as combining various kinds of neural networks in a modular neural system for intrusion detection.
- **Alexander O. Tarakanov** (“Immunocomputing for speaker recognition”) proposes an approach to speaker recognition by intelligent signal processing based on mathematical models of immunocomputing. The approach includes both low-level feature extraction and high-level ("intelligent") pattern recognition. The key model is the formal immune network including apoptosis (programmed cell death) and immunization both controlled by cytokines (messenger proteins). Such a formal immune network can be formed from audio signals using a discrete tree transform, singular value decomposition, and the proposed index of inseparability in comparison with the Renyi entropy. An application to the recognition of nine male speakers by their utterances of two Japanese vowels is shown, and the proposed approach outperforms main state of the art approaches of computational intelligence.

We are happy that we initiated and now are able to finalize this exceptional two-volume editorial project the scope and size of which is extraordinary. First, and most importantly, we have gathered astonishingly many eminent researchers and scholars from all parts of the world who have been actively working on a broadly perceived area of machine learning, data analysis, data mining, knowledge discovery, etc. They have contributed with great papers in which a synergistic and inspiring combination of a comprehensive state of the art material, deep technical analysis, novel applications, and much vision and new look on the past, present and future of the fields have been included. Second, the sheer size of this editorial project is exceptional, but the readers will obtain a rare view of what is the best, most original and promising in the areas. This all is certainly the best tribute that our research community can pay to Professor Ryszard Michalski. He had been inspiring all of us for so many years, and many of us owe him so much because his works, and contacts with him, have shaped our careers and maybe even life.

In an email to us, sent upon submitting her chapter to our editorial project, Professor Katharina Morik summarized it all in an unsurpassed way, albeit based on just one personal recollection: *“Ryszard would not want to have stories told and memories shared but just to see scientific work. Hence, I abstained from expressing my feelings [in the chapter submitted]. In 1986, I was just an unknown scientist in Berlin when inviting him to a workshop on knowledge representation and machine learning, admitting that I could not pay any royalty. To him, only the subject mattered. He came, and I learned a lot from him. He was kind of a godfather to machine learning those days, but he was open to the inductive logic programming approach which we were developing then. This gave me confidence that in science you don't need to have connections, be famous, or offer a lot of money: just work hard, discuss, and share the enthusiasm, that's it. Ryszard has never lost this attitude”*.

We wish to thank all the authors for their excellent contributions and an efficient collaboration in this huge and demanding editorial project. Special thanks are also due to Dr. Tom Ditzinger, Senior Editor, and Ms. Heather King, both from Engineering/Applied Sciences of Springer, who have provided much encouragement and support.

Warsaw
August 2009

Jacek Koronacki
Zbigniew W. Raś
Sławomir T. Wierchoń
Janusz Kacprzyk

Table of Contents

Part I: General Issues

Knowledge-Oriented and Distributed Unsupervised Learning for Concept Elicitation	3
<i>Witold Pedrycz</i>	
Toward Interactive Computations: A Rough-Granular Approach	23
<i>Andrzej Skowron and Marcin Szczuka</i>	
Data Privacy: From Technology to Economics	43
<i>Stan Matwin and Tomasz Szapiro</i>	
Adapting to Human Gamers Using Coevolution	75
<i>Phillipa M. Avery and Zbigniew Michalewicz</i>	
Wisdom of Crowds in the Prisoner's Dilemma Context	101
<i>Mirsad Hadzikadic and Min Sun</i>	

Part II: Logical and Relational Learning, and Beyond

Towards Multistrategic Statistical Relational Learning	121
<i>Marenglen Biba, Stefano Ferilli, and Floriana Esposito</i>	
About Knowledge and Inference in Logical and Relational Learning	143
<i>Luc De Raedt</i>	
Two Examples of Computational Creativity: ILP Multiple Predicate Synthesis and the 'Assets' in Theorem Proving	155
<i>Marta Fraňová and Yves Kodratoff</i>	
Logical Aspects of the Measures of Interestingness of Association Rules	175
<i>Jan Rauch</i>	

Part III: Text and Web Mining

Clustering the Web 2.0	207
<i>Katharina Morik and Michael Wurst</i>	
Induction in Multi-Label Text Classification Domains	225
<i>Miroslav Kubat, Kanoksri Sarinnapakorn, and Sareewan Dendamrongvit</i>	

Cluster-Lift Method for Mapping Research Activities over a Concept Tree	245
<i>Boris Mirkin, Susana Nascimento, and Luís Moniz Pereira</i>	

On Concise Representations of Frequent Patterns Admitting Negation	259
<i>Marzena Kryszkiewicz, Henryk Rybiński, and Katarzyna Cichoń</i>	

Part IV: Classification and Beyond

A System to Detect Inconsistencies between a Domain Expert's Different Perspectives on (Classification) Tasks	293
<i>Derek Sleeman, Andy Aiken, Laura Moss, John Kinsella, and Malcolm Sim</i>	

The Dynamics of Multiagent <i>Q</i> -Learning in Commodity Market Resource Allocation	315
<i>Eduardo R. Gomes and Ryszard Kowalczyk</i>	

Simple Algorithms for Frequent Item Set Mining	351
<i>Christian Borgelt</i>	

Monte Carlo Feature Selection and Interdependency Discovery in Supervised Classification	371
<i>Michał Damiński, Marcin Kierczak, Jacek Koronacki, and Jan Komorowski</i>	

Machine Learning Methods in Automatic Image Annotation	387
<i>Halina Kwaśnicka and Mariusz Paradowski</i>	

Part V: Neural Networks and Other Nature Inspired Approaches

Integrative Probabilistic Evolving Spiking Neural Networks Utilising Quantum Inspired Evolutionary Algorithm: A Computational Framework	415
<i>Nikola Kasabov</i>	

Machine Learning in Vector Models of Neural Networks	427
<i>Boris Kryzhanovsky, Vladimir Kryzhanovsky, and Leonid Litinskiĭ</i>	

Nature Inspired Multi-Swarm Heuristics for Multi-Knowledge Extraction	445
<i>Hongbo Liu, Ajith Abraham, and Benxian Yue</i>	

Discovering Data Structures Using Meta-learning, Visualization and Constructive Neural Networks	467
<i>Tomasz Maszczyk, Marek Grochowski, and Włodzisław Duch</i>	

Neural Network and Artificial Immune Systems for Malware and Network Intrusion Detection.....	485
<i>Vladimir Golovko, Sergei Bezobrazov, Pavel Kachurka, and Leonid Vaitsekhovich</i>	
Immunocomputing for Speaker Recognition	515
<i>Alexander O. Tarakanov</i>	
Author Index	531