RSCTC'2000 Conference Committee

Honorary Chairs: Zdzislaw Pawlak, Lotfi A. Zadeh

Conference Chair: Wojciech Ziarko

Program Chairs: Yiyu Yao, Andrzej Skowron (Europe), Shusaku Tsumoto (Asia)

Advisory Board:

Nick Cercone	Jerzy Grzymala-Busse	T.Y. Lin
Toshinori Munakata	Akira Nakamura	Setsuo Ohsuga
Lech Polkowski	Zbigniew Ras	Roman Slowinski
Paul P. Wang	Ning Zhong	

Publicity Chair: Ning Zhong

Local Organization Committee: Mengchi Liu (Chair), David Brindle, Cory Butz

Program Committee

Anna Buczak	Gianpiero Cattaneo	Juan-Carlos Cubero
Andrzej Czyzewski	Didier Dubois	Ivo Duentsch
Maria C. Fernandez	Salvatore Greco	Howard Hamilton
Howard Ho	Xiaohua Hu	Ryszard Janicki
Julia Johnson	Janusz Kacprzyk	Willi Kloesgen
Jan Komorowski	Jacek Koronacki	Bozena Kostek
Wojciech Kowalczyk	Marzena Kryszkiewicz	Pawan Lingras
Chunnian Liu	Jiming Liu	Mengchi Liu
Qing Liu	Tadeusz Luba	Solomon Marcus
Ernestina Menasalvas	Zbigniew Michalewicz	Ryszard Michalski
John Mordeson	Mikhail Moshkov	Hung Son Nguyen
Kanji Ohsima	Ewa Orlowska	Piero Pagliani
Witold Pedrycz	James F. Peters	Zdzislaw Piasta
Henri Prade	Mohamed Quafafou	Vijay Raghavan
Sheela Ramanna	Qiang Shen	Krzysztof Slowinski
Jerzy Stefanowski	Jaroslaw Stepaniuk	Zbigniew Suraj
Roman Swiniarski	Hans-Michael Voigt	Anita Wasilewska
S.K.M. Wong	Xindong Wu	Satoshi Yamane
Lizhu Zhou	Jan Zytkow	

Welcome Message from Conference Chair

Dear Conference Participants:

We are pleased to present the program of RSCTC'2000 and thank you for your research contributions without which the organization of the conference would not be possible. It is our hope that you will enjoy the conference and the spectacular beauty of the Rocky Mountains. We warmly welcome you in Banff and wish you great time in this beautiful place!

Wojciech Ziarko, Conference Chair of RSCTC'2000

RSCTC'2000 Sessions

Monday October 16, 2000	
8:30 - 8:40 (Summit Room) Opening Address	
8:40 - 9:40 (Summit Room) Keynote Speech	
10:00 - 12:00 (Summit Room) Rough Sets and Systems I	10:00 - 12:00 (Assiniboine Room) Fuzzy Sets and Systems
1:30 - 3:30 (Summit Room) Special Session on Granular Computing	1:30 - 3:30 (Assiniboine Room) Non-Classical Logics and Reasoning
4:00 - 6:00 (Summit Room) Rough Sets and Data Mining I	4:00 - 6:00 (Assiniboine Room) Pattern Recognition and Image Processing
7:00PM - 9:00PM (Alpine Meadows) Conference Reception	

Tuesday October 17, 2000	
8:40 - 9:40 (Summit Room) Keynote Speech	
10:00 - 11:00 (Summit Room) Keynote Speech	
11:20 - 12:00 (Summit Room) Plenary Talk	
1:30 - 3:10 (Summit Room) Rough Sets and Systems II	1:30 - 3:10 (Assiniboine Room) Current Trends in Computing I
3:30 - 5:30 (Summit Room) Machine Learning and Data Mining I	3:30 - 5:30 (Assiniboine Room) Current Trends in Computing II
5:30 - 6:20 (Summit Room) Business Meeting	
7:00 - 10:00 Conference Banquet	

Wednesday October 18, 2000	
9:00 - 9:40 (Summit Room) Plenary Talk	
10:00 - 10:40 (Summit Room) Plenary Talk	
11:00 - 11:40 (Summit Room) Plenary Talk	
1:30 - 3:10 (Summit Room) Rough Sets and Data Mining II	1:30 - 3:10 (Assiniboine Room) Current Trends in Computing III
3:30 - 5:30 (Summit Room) Rough Sets and Systems III	3:30 - 5:30 (Assiniboine Room) Current Trends in Computing IV

Thursday October 19, 2000
10:00 - 3:00PM Conference Tour to Lake Louise

Program of RSCTC'2000

Sunday October 15, 2000

15:00 - 20:00 Registration.

Monday October 16, 2000

8:30 - 8:40 (Summit Room) Opening Address
Dr. W. Ziarko

8:40 - 9:40 (Summit Room) Keynote Speech
Session Chair: Dr. N. Zhong

S. Ohsuga, To What Extent Can Computers Aid Human Activity Toward Second Phase Information Technology ?

9:40 - 10:00 Break

10:00 - 12:00 (Summit Room) Parallel Session: Rough Sets and Systems I
Session Chair: A. Skowron

- J.J. Alpigini, J.F. Peters, Dynamic System Visualization With Rough Performance Maps (58)
- P. Apostoli, A. Kanda, Approximation Spaces of Type-Free Sets (66)
- J.G. Bazan, M. Szczuka, RSES and RSESlib - A Collection of Tools for Rough Set Computations (74)
- M. Beynon, An Investigation of β-reduct Selection Within the Variable Precision Rough Sets Model (82)
- M. Borkowski, D. Ślęzak, Application of Discernibility Tables to Calculation of Approximate Frequency Based Reducts (91)
- M. Boussouf, M. Quafafou, Scalable Feature Selection Using Rough Set Theory (99)

10:00 - 12:00 (Assiniboine Room) Parallel Session: Fuzzy Sets and Systems
Session Chair: R. Slowinski

- K. Chakrabarty, On Fuzzy Lattice (200)
- R. Intan, M. Mukaidono, Conditional Probability Relations in Fuzzy Relational Database (213)

- M. Inuiguchi, T. Tanino, A New Class of Necessity Measures and Fuzzy Rough Sets Based on Certainty Qualifications (223)
- J. Saquer, J.S. Deogun, A Fuzzy Approach for Approximating Concepts (231)
- H. Thiele, On Axiomatic Characterizations of Fuzzy Approximation Operators I: The Fuzzy Rough Set Based Case (239)

12:00 - 1:30 Lunch

1:30 - 3:30 (Summit Room) Parallel Session: Special Session on Granular Computing
Session Chair: J.F. Peters

- H.S. Nguyen, A. Skowron, M. Szczuka, Situation Identification by Unmanned Aerial Vehicle (17)
- L. Polkowski, A. Skowron, Rough-Neuro Computing (25)
- A. Skowron, J. Stepaniuk, J.F. Peters, Approximation of Information Granule Sets (33)
- Y. Yamauchi, M. Mukaidono, Probabilistic Inference and Bayesian Theorem on Rough Sets (41)
- Y.Y. Yao, J.P. Zhang, Interpreting Fuzzy Membership Functions in the Theory of Rough Sets (50)
- K.E. Wolff, A Conceptual View of Knowledge Bases in Rough Set Theory (188)

1:30 - 3:30 (Assiniboine Room) Parallel Session: Non-Classical Logics and Reasoning
Session Chair: R. Swiniarski

- T. Bittner, J.G. Stell, Rough Sets in Approximate Spatial Reasoning (407)
- K. Nakamatsu, J.M. Abe, A. Suzuki, Annotated Semantics for Defeasible Deontic Reasoning (432)
- L. Polkowski, Spatial Reasoning via Rough Sets (441)
- P. Sapiecha, H. Selvaraj, M. Pleban, Decomposition of Boolean Relations and Functions in Logic Synthesis and Data Analysis (449)
- S. Tsumoto, Diagnostic Reasoning from the Viewpoint of Rough Sets (457)
- S.K.M. Wong, C.J. Butz, Rough Sets for Uncertainty Reasoning (473)

3:30 - 4:00 Break

4:00 - 6:00 (Summit Room) Parallel Session: Rough Sets and Data Mining I
Session Chair: J. Grzymala-Busse

- S. Greco, B. Matarazzo, R. Slowinski, J. Stefanowski, Variable Consistency Model of Dominance-Based Rough Sets Approach (138)
- S. Greco, B. Matarazzo, R. Slowinski, J. Stefanowski, An Algorithm for Induction of Decision Rules Consistent With the Dominance Principle (266)
- J.W. Guan, D.A. Bell, Data Mining for Monitoring Loose Parts in Nuclear Power Plants (276)
- J. Han, X. Hu, N. Cercone, Supervised Learning: A Generalized Rough Set Approach (284)
- J. Johnson, M. Liu, H. Chen, Unification of Knowledge Discovery and Data Mining Using Rough Sets Approach in a Real-World Application (292)
- A.H. Liang, W. Ziarko, B. Maguire, The Application of a Distance Learning Algorithm in Web-Based Course Delivery (300)

4:00 - 6:00 (Assiniboine Room) Parallel Session: Pattern Recognition and Image Processing
Session Chair: J. Stefanowski

- B. Chakraborty, Feature Subset Selection by Neuro-Rough Hybridization (481)
- S.-W. Jang, K.-J. Kim, H.-Il Choi, Accurate Estimation of Motion Vectors Using Active Block Matching (489)
- M. Last, A. Kandel, O. Maimon, E. Eberbach, Anytime Algorithm for Feature Selection (494)
- B.T. Chun, Y. Bae, T.-Y. Kim, Reconstruction in an Original Video Sequence of the Lost Region Occupied by Caption Areas (502)
- Z.M. Wojcik, Detecting Spots for NASA Space Programs Using Rough Sets (531)

7:00PM - 9:00PM (Alpine Meadows) Conference Reception

Tuesday October 17, 2000

8:40 - 9:40 (Summit Room) Keynote Speech
Session Chair: W. Ziarko

L.A. Zadeh, Toward Perception-Based Probabilistic Reasoning

9:40 - 10:00 Break

10:00 - 11:00 (Summit Room) Keynote Speech
Session Chair: A. Skowron

Z. Pawlak, Rough Sets and Decision Algorithms

11:20 - 12:00 (Summit Room) Plenary Talk
Session Chair: Y.Y. Yao

Jan Zytkow, A Statistics Connection: Rough Sets and Contingency Tables as Tools of Discovery

12:00 - 1:30PM Lunch

1:30 - 3:10 (Summit Room) Parallel Session: Rough Sets and Systems II
Session Chair: L. Polkowski

- A. Chouchoulas, Q. Shen, Rough Set-Based Dimensionality Reduction for Multivariate Adaptive Regression Splines (112)
- S. Greco, B. Matarazzo, R. Slowinski, Rough Set Approach to Decisions Under Risk (128)
- J. Järvinen, Approximations and Rough Sets Based on Tolerances (150)
- H. Midelfart, J. Komorowski, A Rough Set Approach to Inductive Logic Programming (158)
- J. Stefanowski, A. Tsoukiàs, Valued Tolerance and Decision Rules (180)

1:30 - 3:10 (Assiniboine Room) Parallel Session: Current Trends in Computing I
Session Chair: H.S. Nguyen

- R. Janicki, Y. Liu, On Trace Assertion Method of Module Interface Specification With Concurrency (593)
- K. Karimi, J. Johnson, H.J. Hamilton, A Proposal for Including Behavior in the Process of Object Similarity Assessment With Examples from Artificial Life (603)
- P. Levecq, N. Meskens, How Can Help an Inductive Approach in the Resolution of a Problematic α Problem? (608)
- T. Wu, S.-C. Huang, A Timed Petri Net Interpreter: An Alternative Method for Simulation (613)
- C. Zhang, Z. Zhang, O.S. San, An Agent-Based Soft Computing Society (621)

3:10 - 3:30 Break

3:30 - 5:30 (Summit Room) Parallel Session: Machine Learning and Data Mining I
Session Chair: S. Tsumoto

- J.W. Grzymala-Busse, M. Hu, A Comparison of Several Approaches to Missing Attribute Values in Data Mining (340)
- M. Nakata, T. Murai, and G. Resconi, Handling Data With Plural Sources in Databases (366)
- D. Ramamonjisoa, E. Suzuki, I. Hamid, Research Topics Discovery from WWW by Keywords Association Rules (374)
- P. Synak, Temporal Templates and Analysis of Time Related Data (382)
- A. Wojna, Constraint Based Incremental Learning of Classification Rules (390)
- N. Zhong, J. Dong, C. Liu, S. Ohsuga, A Hybrid Model for Rule Discovery in Data (398)

3:30 - 5:30 (Assiniboine Room) Parallel Session: Current Trends in Computing II
Session Chair: M. Moshkov

- D. Ślęzak, J. Wróblewski, Application of Normalized Decision Measures to the New Case Classification (515)
- R. Swiniarski, An Application of Rough Sets and Harr Wavelets to Face Recognition (523)
- Z.M. Wojcik, Solving the Localization-Detection Trade-Off in Shadow Recognition Problems Using Rough Sets (539)
- G. Auda, M. Kamel, A Modular Neural Network for Vague Classification (545)
- Z. Suraj, Some Remarks on Extensions and Restrictions of Information Systems (172)
- J.F. Peters, A. Skowron, L. Han, S. Ramanna, Towards Rough Neural Computing Based on Rough Membership Functions: Theory and Application (572)

5:30 - 6:20 (Summit Room) Business Meeting

7:00 - 10:00 Conference Banquet

Wednesday October 18, 2000

9:00 - 9:40 (Summit Room) Plenary Talk:
Session Chair: T. Tsumoto

J. Komorowski, Predicting Gene Function from Gene Expressions and Background Knowledge A Rough Set Success Story

9:40 - 10:00 Break

10:00 - 10:40 (Summit Room) Plenary Talk:
Session Chair: J.F. Peters

R. Swiniarski, Rough Sets and Their Applications to Feature Reduction and Selection

10:40 - 11:00 Break

11:00 - 11:40 (Summit Room) Plenary Talk:
Session Chair: M. Kryszkiewicz

J. Grzymala-Busse, Learning from Imbalanced Data

12:00 - 1:30 Lunch

1:30 - 3:10 (Summit Room) Parallel Session: Rough Sets and Data Mining II
Session Chair: S. Greco

- J.F. Gálvez, F. Diaz, P. Carrión, A. Garcia, An Application for Knowledge Discovery Based on a Revision of VPRS Model (258)
- M. Moshkov, Classification of Infinite Information Systems (167)
- H.S. Nguyen, On Efficient Construction of Decision Trees from Large Databases (316)
- S. Tsumoto, An Approach to Statistical Extension of Rough Set Rule Induction (324)
- G. Wang, F. Liu, The Inconsistency in Rough Set Based Rule Generation (332)

1:30 - 3:10 (Assiniboine Room) Parallel Session: Current Trends in Computing III
Session Chair: N. Zhong

- I. Chikalov, On Algorithm for Constructing of Decision Trees With Minimal Number of Nodes (107)
- M. Kryszkiewicz, Inducing Theory for the Rule Set (353)
- T. Lin, Y.E. Huang, A Study of Conditional Independence Change in Learning Probabilistic Network (416)
- J. Wierzbicki, Rough Set Approach to CBR (465)
- C. Chiu, J.-T. Hsu, C.-Y. Lin, The Application of Genetic Programming in Milk Yield Prediction for Dairy Cows (559)

3:10 - 3:30 Break

3:30 - 5:30 (Summit Room) Parallel Session: Rough Sets and Systems III
Session Chair: Y. Yao

- T.Y. Lin, H. Cao, Searching Decision Rules in Very Large Databases Using Rough Set Theory (308)
- M.C. Fernández-Baizán, C. Pérez-Llera, J. Feito-García, A. Almeida, LEM3 Algorithm Generalization Based on Stochastic Approximation Space (248)
- A. Breunig, H.R. de Azevedo, E.P. Teixeira, E.B. de Faria, Computer Vision Using Fuzzy Logic for Robot Manipulator - Lab. Implementation (193)

3:30 - 5:30 (Assiniboine Room) Parallel Session: Current Trends in Computing IV
Session Chair: W. Ziarko

- J.-C. Lamirel, J. Ducloy, H. Kammoun, Some Basics of a Self Organizing Map (SOM) Extended Model for Information Discovery in a Digital Library Context (361)
- H. Kiem, D. Phuc, Discovering Motif Based Association Rules in a Set of DNA Sequences (348)
- Q. Liu, λ-level Rough Equality Relation and the Inference of Rough Paramodulation (424)
- L. Tian, The Nature of Crossover Operator in Genetic Algorithms (580)

Thursday October 19, 2000

10:00 - 3:00PM Conference Tour to Lake Louise

If you like to hike, please bring proper foot-ware and warm jackets.
Banff has outdoor Hot Springs and therefore having swimming suit is also recommended.

Abstracts of Keynote Speech and Plenary Talks

S. Ohsuga, To What Extent Can Computers Aid Human Activity Toward Second Phase Information Technology?
High technology expands whole social activity. Each activity is growing large and complex in many areas such as energy, transportation, economy, communication, security, information processing, environment, health care, etc. Needs for design and development of new systems are increasing. But the increased scale and complexity of each task causes troubles such as the delay or cancellation of large software development project, a failure of a new breeder type reactor, and so on. Apart from what have been made as the primary causes of the troubles, the author thinks that there were the limitations of human capability at the basis of the troubles. These limitations are seen in many aspects. As the partner to back up human being computer's role is increasing. But can they aid all aspects of the human act ivity? The answer is No. The human activity is classified into two classes; the first class that can be dealt with by algorithmic method such as business information processing, scientific computation, network applications, and so on and the second class that cannot be dealt with by algorithmic method such as design and development, programming, etc. Currently computers can be used mainly in the first class activities. Growing inequality of computer's aids among activities accelerates the occurrence of troubles. Only possibility for its solution is to make computers more intelligent to aid all aspects of human activity, because human capability has already come to near the limit but computer's capability has the room to be enhanced. In this talk, first the environment of future information technology and second an idea for new information technology to adapt this environmental change are discussed. The approach is to consider the information technology for dealing with large and complex problems. It leads us to an idea to change the ordinary human-led interactive systems to computer-led interactive systems where X-led interactive system is the system in which X has an initiative. It requires a very high level intelligence to computers. To discuss the way to develop such a system is the objective of this talk. Major topics discussed in this talk are; model-based software architecture, modeling scheme, method of externalization and model building, autonomous problem solving, large knowledge base, automatic programming and integration of different information processing methods, and knowledge acquisition. This is an outline of a research work being conducted by the authors group under the sponsorship of The Science and Technology Agency of Japanese Government.

L.A. Zadeh, Toward a Perception-Based Theory of Probabilistic Reasoning
The past two decades have witnessed a dramatic growth in the use of probability-based methods in a wide variety of applications centering on automation of decision-making in an environment of uncertainty and incompleteness of information.

Successes of probability theory have high visibility. But what is not widely recognized is that successes of probability theory mask a fundamental limitation – the inability to operate on what may be called perception-based information. Such information is exemplified by the following. Assume that I look at a box containing balls of various sizes and form the perceptions: (a) there are about twenty balls; (b) most are large; and (c) a few are small. The question is: What is the probability that a ball drawn at random is neither large not small? Probability theory cannot answer this question because there is no mechanism within the theory to represent the meaning of perceptions in a form that lends itself to computation. The same problem arises in the examples:

Usually Robert returns from work at about 6 pm. What is the probability that Robert is home at 6:30 pm? I do not know Michelle's age but my perceptions are: (a) it is very unlikely that Michelle is old; and (b) it is likely that Michelle is not young. What is the probability that Michelle is neither young nor old? X is a normally distributed random variable with small mean and small variance. What is the probability that X is large? Given the data in insurance company database, what is the probability that my car may be stolen? In this case, the answer depends on perception-based information which is not in insurance company database.

In these simple examples – examples drawn from everyday experiences – the general problem is that of estimation of probabilities of imprecisely defined events, given a mixture of measurement-based and perception-based information. The crux of the difficulty is that perception-based information is usually described in a natural language – a language which probability theory cannot understand and hence is not equipped to handle.

To endow probability theory with a capability to operate on perception-based information, it is necessary to generalize it in three ways. To this end, let PT denote standard probability theory of the kind taught in university-level courses. The three modes of generalization are labeled: (a) f-generalization; (b) f.g-generalization: and (c) nl-generalization. More specifically: (a) f-generalization involves fuzzification, that is, progression from crisp sets to fuzzy sets, leading to a generalization of PT which is denoted as PT+. In PT+, probabilities, functions, relations, measures and everything else are allowed to have fuzzy denotations, that is, be a matter of degree. In particular, probabilities described as low, high, not very high, etc. are interpreted as labels of fuzzy subsets of the unit interval or, equivalently, as possibility distributions of their numerical values. (b) f.g-generalization involves fuzzy granulation of variables, functions, relations, etc., leading to a generalization of PT which is denoted as PT++. By fuzzy granulation of a variable, X, what is meant is a partition of the range of X into fuzzy granules, with a granule being a clump of values of X which are drawn together by indistinguishability, similarity, proximity, or functionality. For example, fuzzy granulation of the variable Age partitions its vales into fuzzy granules labeled very young, young, middle-aged, old, very old, etc. Membership functions of such granules are usually assumed to be triangular or trapezoidal. Basically, granulation reflects the bounded ability of the human mind to resolve detail and store information. (c) Nl-generalization involves an addition to PT++ of a capability to represent the meaning of propositions expressed in a natural language, with the understanding that such propositions serve as descriptors of perceptions. Nl-generalization of PT leads to perception-based probability theory denoted as PTp.

An assumption which plays a key role in PTp is that the meaning of a proposition, p, drawn from a natural language may be represented as what is called a generalized constraint on a variable. More specifically, a generalized constraint is represented as X isr R, where X is the constrained variable; R is the constraining relation; and isr, pronounced ezar, is a copula in which r is an indexing variable whose value defines the way in which R constrains X. The principal types of constraints are: equality constraint, in which case isr is abbreviated to =; possibilistic constraint, with r abbreviated to blank; veristic constraint, with r=v; probabilistic constraint, in which case r=p, X is a random variable and R is its probability distribution; random-set constraint, r=rs, in which case X is set-valued random variable and R is its probability distribution; fuzzy-graph constraint, r=fg, in which case X is a function or a relation and R is its fuzzy graph; and usuality constraint, r=u, in which case X is a random variable and R is its usual – rather than expected – value.

The principal constraints are allowed to be modified, qualified, and combined, leading to composite generalized constraints. An example is: usually (X is small) and (X is large) is unlikely. Another example is: if (X is very small) then (Y is not very large) or if (X is large) then (Y is small).

The collection of composite generalized constraints forms what is referred to as the Generalized Constraint Language (GCL). Thus, in PTp, the Generalized Constraint Language serves to represent the meaning of perception-based information. Translation of descriptors of perceptions into GCL is accomplished through the use of what is called the constraint-centered semantics of natural languages (CSNL). Translating descriptors of perceptions into GCL is the first stage of perception-based probabilistic reasoning.

The second stage involves goal-directed propagation of generalized constraints from premises to conclusions. The rules governing generalized constraint propagation coincide with the rules of inference in fuzzy logic. The principal rule of inference is the generalized extension principle. In general, use of this principle reduces computation of desired probabilities to the solution of constrained problems in variational calculus or mathematical programming.

It should be noted that constraint-centered semantics of natural languages serves to translate propositions expressed in a natural language into GCL. What may be called the constraint-centered semantics of GCL, written as CSGCL, serves to represent the meaning of a composite constraint in GCL as a singular constraint X isr R. The reduction of a composite constraint to a singular constraint is accomplished through the use of rules which govern generalized constraint propagation.

Another point of importance is that the Generalized Constraint Language is maximally expressive, since it incorporates all conceivable constraints. A proposition in a natural language, NL, which is translatable into GCL is said to be admissible. The richness of GCL justifies the default assumption that any given proposition in NL is admissible. The subset of admissible propositions in NL constitutes what is referred to as a precisiated natural language, PNL. The concept of PNL opens the door to a significant enlargement of the role of natural languages in information processing, decision and control.

Perception-based theory of probabilistic reasoning suggests new problems and new directions in the development of probability theory. It is inevitable that in coming years there will be a progression from PT to PTp, since PTp enhances the ability of probability theory to deal with realistic problems in which decision-relevant information is a mixture of measurements and perceptions.

Z. Pawlak, Rough Sets and Decision Algorithms
Rough set based data analysis starts from a data table, called an information system. The information system contains data about objects of interest characterized in terms of some attributes. Often we distinguish in the information system condition and decision attributes. Such information system is called a decision table. The decision table describes decisions in terms of conditions that must be satisfied in order to carry out the decision specified in the decision table. With every decision table a set of decision rules, called a decision algorithm can be associated. It is shown that every decision algorithm reveals some well known probabilistic properties, in particular it satisfies the Total Probability Theorem and the Bayes' Theorem. These properties give a new method of drawing conclusions from data, without referring to prior and posterior probabilities, inherently associated with Bayesian reasoning.

Jan Zytkow, A Statistics Connection: Rough Sets and Contingency Tables as Tools of Discovery
Contingency tables represent data in a granular way and are a well-established tool for inductive generalization of knowledge from data. Both rough sets and contingency tables are founded on a similar idea of granular empirical data. Both approaches use the representation of empirical objects by n-tuples or vector of attribute values. We show that the basic concepts of rough sets, such as concept approximation, indiscernibility, and reduct can be used in the framework of contingency tables. We further demonstrate the relevance to rough sets theory of additional probabilistic information available in contingency tables and in particular of statistical tests of significance and predictive strength applied to contingency tables. Tests of both type can help the evaluation mechanisms used in inductive generalization based on rough sets. Granularity of attributes can be improved in feedback with knowledge discovered in data. We demonstrate how various techniques for (1) contingency table refinement, for (2) column and row grouping based on correspondence analysis, and (3) the search for equivalence relations between attributes improve both granularization of attributes and the quality of knowledge. Finally we demonstrate the limitations of knowledge viewed as concept approximation, which is the focus of rough sets. Transcending that focus and reorienting towards the predictive knowledge and towards the related distinction between possible and impossible (or statistically improbable) situations can be very useful in expanding the rough sets approach to more expressive forms of knowledge.

J. Komorowski, Predicting Gene Function from Gene Expressions and Background Knowledge – A Rough Set Success Story

The majority of states in health or disease is most likely controlled by hundreds of genes. Until recently, molecular biomedicine could study one or very few genes in parallel. With the advent of the high throughput micro-array technology it is now possible to observe the levels of activity in literally thousands of genes. At the same time, genome mapping projects such as, for instance, the Human Genome Initiative, which is an international research program for the creation of detailed genetic and physical maps of the human genome, generate enormous quantities of data. For the human genome, there are approximately 100K genes out of which ca 5K genes are known. The major goal of Functional Genomics is an assignment of function to genes. State-of-the-art in Functional Genomics is the use of unsupervised learning methods. Our multidisciplinary team has developed a supervised learning method for inducing predictive rule models for functional classification of gene expressions from microarray hybridization experiments.

In this talk I show how a rough set-based approach implemented in the ROSETTA system helps resolve some of the thorny issues in the functional classification of gene expressions. The predictive and descriptive quality of our rule models is demonstrated on the fibroblast serum response data. Our analysis shows that the rules are capable of representing the complex relationship between gene expressions and function, and that it is possible to put forward high quality hypotheses about the function of unknown genes.

Finally, I analyze why a rough set approach seems to be the right choice for this class of problems.

This is joint research with Torgeir Hvidsten, Astrid Laegreid, Herman Midelfart and Arne Sandvik.

Nota bene. The talk is self-contained with respect to the knowledge of molecular biology.

R.W. Swiniarski, Rough Sets and Their Applications to Feature Reduction and Selection

Reduction of pattern dimensionality through feature extraction and feature reduction/selection belongs to the most fundamentals steps in data preprocessing. Feature selection is often isolated as a separate step in processing sequence. The talk presents some aspects of rough sets methods in context of feature reduction and selection in pattern recognition. The presented rough sets theory description emphasizes a role of rough sets reducts in feature reduction and selection, including dynamic reducts. The overview of methods of feature selection emphasizes the open and closed loop methods and feature selection criteria, including rough sets-based methods. The talk presents an algorithm and an application of rough sets for feature reduction/selection proposed jointly with Principal Component Analysis. Finally, the talk presents numerical results of face recognition experiments using rough sets rule-based and neural network classifiers, with feature selection/reduction technique which is based on the proposed Principal Components Analysis and rough sets methods. This talk also includes a short description of feature extraction from facial images using Singular

Value Decomposition.

J.W. Grzymala-Busse, Learning from Imbalanced Data
Frequently real-life data sets are imbalanced: one class is represented by the majority of cases while the other class is a minority. In medical data the smaller class – as a rule – is more important.. As a result of rule induction from imbalanced data, rule sets are biased: the minority class is not well represented. During classification of unseen cases, almost every case is classified as a member of the majority class. The total accuracy, measured as a ratio of the total number of correctly classified cases to the total number of cases, is good, because almost every case from the majority class is correctly classified. On the other hand, many cases from the smaller class may not be classified correctly. Thus the resulting classification system will be rejected by diagnosticians. A solution, based on the ideas related to the ROC analysis, is presented.