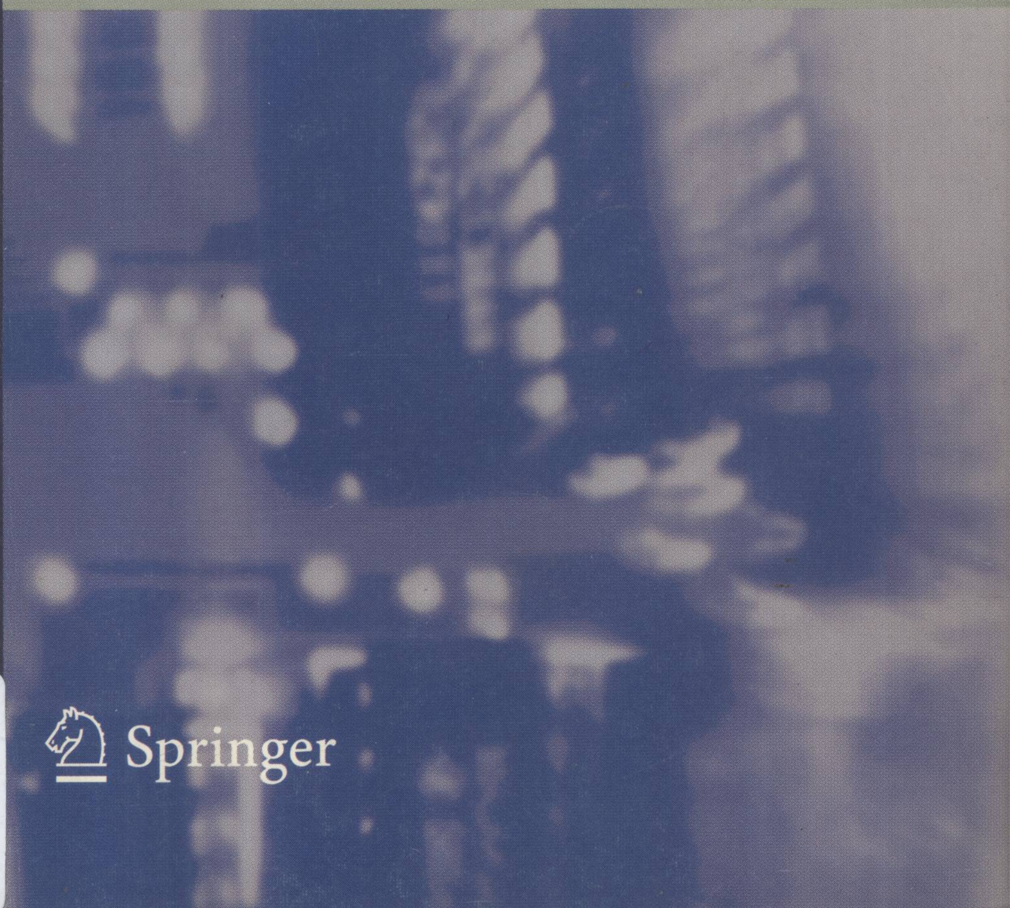
Journal Subline

LNCS 3100

Transactions on
Rough Sets I

James F. Peters · Andrzej Skowron
Editors-in-Chief



Springer

James F. Peters Andrzej Skowron
Jerzy W. Grzymała-Busse Bożena Kostek
Roman W. Świniarski Marcin S. Szczuka (Eds.)

Transactions on Rough Sets I

Springer

Editors-in-Chief

James F. Peters
University of Manitoba, Department of Electrical and Computer Engineering
Manitoba, Winnipeg, Manitoba R3T 5V6 Canada
E-mail: jfpeters@ee.umanitoba.ca

Andrzej Skowron
University of Warsaw, Institute of Mathematics
Banacha 2, 02-097 Warsaw, Poland
E-mail: skowron@mimuw.edu.pl

Volume Editors

Jerzy W. Grzymała-Busse
University of Kansas, Department of Electrical Engineering and Computer Science
3014 Eaton Hall 1520 W. 15th St., #2001 Lawrence, KS 66045-7621, USA
E-mail: Jerzy@ku.edu

Bożena Kostek
Gdansk University of Technology
Faculty of Electronics, Telecommunications and Informatics
Multimedia Systems Department, Narutowicza 11/12, 80-952 Gdansk, Poland
E-mail: bozenka@sound.eti.pg.gda.pl

Roman W. Świniarski
San Diego State University, Department of Computer Science
5500 Campanile Drive, San Diego, CA 92182-7720, USA
E-mail: rswiniar@sciences.sdsu.edu

Marcin S. Szczuka
Warsaw University, Institute of Mathematics
Banacha 2, 02-097 Warsaw, Poland
E-mail: szczuka@mimuw.edu.pl

Library of Congress Control Number: 2004108444

CR Subject Classification (1998): F.4.1, F.1, I.2, H.2.8, I.5.1, I.4

ISSN 0302-9743
ISBN 3-540-22374-6 Springer-Verlag Berlin Heidelberg New York

Springer-Verlag is a part of Springer Science+Business Media

springeronline.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Olgun Computergrafik
Printed on acid-free paper SPIN: 11011873 06/3142 5 4 3 2 1 0

Lecture Notes in Computer Science 3100

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Preface

We would like to present, with great pleasure, the first volume of a new journal, Transactions on Rough Sets. This journal, part of the new journal subline in the Springer-Verlag series Lecture Notes in Computer Science, is devoted to the entire spectrum of rough set related issues, starting from logical and mathematical foundations of rough sets, through all aspects of rough set theory and its applications, data mining, knowledge discovery and intelligent information processing, to relations between rough sets and other approaches to uncertainty, vagueness, and incompleteness, such as fuzzy sets, theory of evidence, etc.

The first, pioneering papers on rough sets, written by the originator of the idea, Professor Zdzisław Pawlak, were published in the early 1980s. We are proud to dedicate this volume to our mentor, Professor Zdzisław Pawlak, who kindly enriched this volume with his contribution on philosophical, logical, and mathematical foundations of rough set theory. In his paper Professor Pawlak shows all over again the underlying ideas of rough set theory as well as its relations with Bayes' theorem, conflict analysis, flow graphs, decision networks, and decision rules.

After an overview and introductory article written by Professor Pawlak, the ten following papers represent and focus on rough set theory-related areas. Some papers provide an extension of rough set theory towards analysis of very large data sets, real data tables, data sets with missing values and rough non-deterministic information. Other theory-based papers deal with variable precision fuzzy-rough sets, consistency measure conflict profiles, and layered learning for concept synthesis. In addition, a paper on generalization of rough sets and rule extraction provides two different interpretations of rough sets. The last paper of this group addresses a partition model of granular computing.

Other topics with a more application-orientated view are covered by the following eight articles of this first volume of Transactions on Rough Sets. They can be categorized into the following groups:

- music processing,
- rough set theory applied to software design models and inductive learning programming,
- environmental engineering models,
- medical data processing,
- pattern recognition and classification.

These papers exemplify analysis and exploration of complex data sets from various domains. They provide useful insight into analyzed problems, showing for example how to compute decision rules from incomplete data. We believe that readers of this volume will better appreciate rough set theory-related trends after reading the case studies.

Many scientists and institutions have contributed to the creation and the success of the rough set community. We are very thankful to everybody within the International Rough Set Society who supported the idea of creating a new LNCS journal subline – the Transactions on Rough Sets. It would not have been possible without Professors Peters' and Skowron's invaluable initiative, thus we are especially grateful to them. We believe that this very first issue will be followed by many others, reporting new developments in the rough set domain. This issue would not have been possible without the great efforts of many anonymously acting reviewers. Here, we would like to express our sincere thanks to all of them.

Finally, we would like to express our gratitude to the LNCS editorial staff of Springer-Verlag, in particular Alfred Hofmann, Ursula Barth and Christine Günther, who supported us in a very professional way.

Throughout preparation of this volume the Editors have been supported by various research programs and funds; Jerzy Grzymała-Busse has been supported by NSF award 9972843, Bożena Kostek has been supported by the grant 4T11D01422 from the Polish Ministry for Scientific Research and Information Technology, Roman Świniarski has received support from the *"Adaptive Data Mining and Knowledge Discovery Methods for Distributed Data"* grant, awarded by Lockheed-Martin, and Marcin Szczuka and Roman Świniarski have been supported by the grant 3T11C00226 from the Polish Ministry for Scientific Research and Information Technology.

April 2004

Jerzy W. Grzymała-Busse
Bożena Kostek
Roman Świniarski
Marcin Szczuka

LNCS Transactions on Rough Sets

This journal subline has as its principal aim the fostering of professional exchanges between scientists and practitioners who are interested in the foundations and applications of rough sets. Topics include foundations and applications of rough sets as well as foundations and applications of hybrid methods combining rough sets with other approaches important for the development of intelligent systems.

The journal includes high-quality research articles accepted for publication on the basis of thorough peer reviews. Dissertations and monographs up to 250 pages that include new research results can also be considered as regular papers. Extended and revised versions of selected papers from conferences can also be included in regular or special issues of the journal.

Lecture Notes in Computer Science

For information about Vols. 1–3025

please contact your bookseller or Springer-Verlag

Vol. 3133: A.D. Pimentel, S. Vassiliadis (Eds.), Computer Systems, Architectures, Modeling, and Simulation. XIV, 562 pages. 2004.

Vol. 3125: D. Kozen (Ed.), Mathematics of Program Construction. X, 401 pages. 2004.

Vol. 3123: A. Belz, R. Evans, P. Piwek (Eds.), Natural Language Generation. X, 219 pages. 2004. (Subseries LNAI).

Vol. 3120: J. Shawe-Taylor, Y. Singer (Eds.), Learning Theory. X, 648 pages. 2004. (Subseries LNAI).

Vol. 3118: K. Miesenberger, J. Klaus, W. Zagler, D. Burger (Eds.), Computer Helping People with Special Needs. XXIII, 1191 pages. 2004.

Vol. 3116: C. Rattray, S. Maharaj, C. Shankland (Eds.), Algebraic Methodology and Software Technology. XI, 569 pages. 2004.

Vol. 3114: R. Alur, D.A. Peled (Eds.), Computer Aided Verification. XII, 536 pages. 2004.

Vol. 3113: J. Karhumäki, H. Maurer, G. Paun, G. Rozenberg (Eds.), Theory Is Forever. X, 283 pages. 2004.

Vol. 3112: H. Williams, L. MacKinnon (Eds.), New Horizons in Information Management. XII, 265 pages. 2004.

Vol. 3111: T. Hagerup, J. Katajainen (Eds.), Algorithm Theory - SWAT 2004. XI, 506 pages. 2004.

Vol. 3109: S.C. Sahinalp, S. Muthukrishnan, U. Dogrusoz (Eds.), Combinatorial Pattern Matching. XII, 486 pages. 2004.

Vol. 3107: J. Bosch, C. Krueger (Eds.), Software Reuse: Methods, Techniques and Tools. XI, 339 pages. 2004.

Vol. 3105: S. Göbel, U. Spierling, A. Hoffmann, I. Iurgel, O. Schneider, J. Dechau, A. Feix (Eds.), Technologies for Interactive Digital Storytelling and Entertainment. XVI, 304 pages. 2004.

Vol. 3104: R. Kralovic, O. Sykora (Eds.), Structural Information and Communication Complexity. X, 303 pages. 2004.

Vol. 3103: K. Deb (Ed.), Genetic and Evolutionary Computation - GECCO 2004. XLIX, 1439 pages. 2004.

Vol. 3102: K. Deb (Ed.), Genetic and Evolutionary Computation - GECCO 2004. L, 1445 pages. 2004.

Vol. 3101: M. Masoodian, S. Jones, B. Rogers (Eds.), Computer Human Interaction. XIV, 694 pages. 2004.

Vol. 3100: J.F. Peters, A. Skowron, J.W. Grzymała-Busse, B. Kostek, R.W. Świniarski, M.S. Szczuka (Eds.), Transactions in Rough Sets I. X, 405 pages. 2004.

Vol. 3099: J. Cortadella, W. Reisig (Eds.), Applications and Theory of Petri Nets 2004. XI, 505 pages. 2004.

Vol. 3098: J. Desel, W. Reisig, G. Rozenberg (Eds.), Lectures on Concurrency and Petri Nets. VIII, 849 pages. 2004.

Vol. 3097: D. Basin, M. Rusinowitch (Eds.), Automated Reasoning. XII, 493 pages. 2004. (Subseries LNAI).

Vol. 3096: G. Melnik, H. Holz (Eds.), Advances in Learning Software Organizations. X, 173 pages. 2004.

Vol. 3094: A. Nürnberger, M. Detyniecki (Eds.), Adaptive Multimedia Retrieval. VIII, 229 pages. 2004.

Vol. 3093: S.K. Katsikas, S. Gritzalis, J. Lopez (Eds.), Public Key Infrastructure. XIII, 380 pages. 2004.

Vol. 3092: J. Eckstein, H. Baumeister (Eds.), Extreme Programming and Agile Processes in Software Engineering. XVI, 358 pages. 2004.

Vol. 3091: V. van Oostrom (Ed.), Rewriting Techniques and Applications. X, 313 pages. 2004.

Vol. 3089: M. Jakobsson, M. Yung, J. Zhou (Eds.), Applied Cryptography and Network Security. XIV, 510 pages. 2004.

Vol. 3086: M. Odersky (Ed.), ECOOP 2004 – Object-Oriented Programming. XIII, 611 pages. 2004.

Vol. 3085: S. Berardi, M. Coppo, F. Damiani (Eds.), Types for Proofs and Programs. X, 409 pages. 2004.

Vol. 3084: A. Persson, J. Stirna (Eds.), Advanced Information Systems Engineering. XIV, 596 pages. 2004.

Vol. 3083: W. Emmerich, A.L. Wolf (Eds.), Component Deployment. X, 249 pages. 2004.

Vol. 3080: J. Desel, B. Pernici, M. Weske (Eds.), Business Process Management. X, 307 pages. 2004.

Vol. 3079: Z. Mammeri, P. Lorenz (Eds.), High Speed Networks and Multimedia Communications. XVIII, 1103 pages. 2004.

Vol. 3078: S. Cotin, D.N. Metaxas (Eds.), Medical Simulation. XVI, 296 pages. 2004.

Vol. 3077: F. Roli, J. Kittler, T. Windeatt (Eds.), Multiple Classifier Systems. XII, 386 pages. 2004.

Vol. 3076: D. Buell (Ed.), Algorithmic Number Theory. XI, 451 pages. 2004.

Vol. 3074: B. Kuijpers, P. Revesz (Eds.), Constraint Databases and Applications. XII, 181 pages. 2004.

Vol. 3073: H. Chen, R. Moore, D.D. Zeng, J. Leavitt (Eds.), Intelligence and Security Informatics. XV, 536 pages. 2004.

Vol. 3072: D. Zhang, A.K. Jain (Eds.), Biometric Authentication. XVII, 800 pages. 2004.

Vol. 3071: A. Omicini, P. Petta, J. Pitt (Eds.), Engineering Societies in the Agents World. XIII, 409 pages. 2004. (Subseries LNAI).

Vol. 3070: L. Rutkowski, J. Siekmann, R. Tadeusiewicz, L.A. Zadeh (Eds.), Artificial Intelligence and Soft Computing - ICAISC 2004. XXV, 1208 pages. 2004. (Subseries LNAI).

Vol. 3068: E. André, L. Dybkj{\}ae r, W. Minker, P. Heisterkamp (Eds.), Affective Dialogue Systems. XII, 324 pages. 2004. (Subseries LNAI).

Vol. 3067: M. Dastani, J. Dix, A. El Fallah-Seghrouchni (Eds.), Programming Multi-Agent Systems. X, 221 pages. 2004. (Subseries LNAI).

Vol. 3066: S. Tsumoto, R. Słowiński, J. Komorowski, J.W. Grzyma{\l}a-Busse (Eds.), Rough Sets and Current Trends in Computing. XX, 853 pages. 2004. (Subseries LNAI).

Vol. 3065: A. Lomuscio, D. Nute (Eds.), Deontic Logic in Computer Science. X, 275 pages. 2004. (Subseries LNAI).

Vol. 3064: D. Bienstock, G. Nemhauser (Eds.), Integer Programming and Combinatorial Optimization. XI, 445 pages. 2004.

Vol. 3063: A. Llamosí, A. Strohmeier (Eds.), Reliable Software Technologies - Ada-Europe 2004. XIII, 333 pages. 2004.

Vol. 3062: J.L. Pfaltz, M. Nagl, B. Böhlen (Eds.), Applications of Graph Transformations with Industrial Relevance. XV, 500 pages. 2004.

Vol. 3061: F.F. Ramas, H. Unger, V. Larios (Eds.), Advanced Distributed Systems. VIII, 285 pages. 2004.

Vol. 3060: A.Y. Tawfik, S.D. Goodwin (Eds.), Advances in Artificial Intelligence. XIII, 582 pages. 2004. (Subseries LNAI).

Vol. 3059: C.C. Ribeiro, S.L. Martins (Eds.), Experimental and Efficient Algorithms. X, 586 pages. 2004.

Vol. 3058: N. Sebe, M.S. Lew, T.S. Huang (Eds.), Computer Vision in Human-Computer Interaction. X, 233 pages. 2004.

Vol. 3057: B. Jayaraman (Ed.), Practical Aspects of Declarative Languages. VIII, 255 pages. 2004.

Vol. 3056: H. Dai, R. Srikant, C. Zhang (Eds.), Advances in Knowledge Discovery and Data Mining. XIX, 713 pages. 2004. (Subseries LNAI).

Vol. 3055: H. Christiansen, M.-S. Hacid, T. Andreasen, H.L. Larsen (Eds.), Flexible Query Answering Systems. X, 500 pages. 2004. (Subseries LNAI).

Vol. 3054: I. Crnkovic, J.A. Stafford, H.W. Schmidt, K. Wallnau (Eds.), Component-Based Software Engineering. XI, 311 pages. 2004.

Vol. 3053: C. Bussler, J. Davies, D. Fensel, R. Studer (Eds.), The Semantic Web: Research and Applications. XIII, 490 pages. 2004.

Vol. 3052: W. Zimmermann, B. Thalheim (Eds.), Abstract State Machines 2004. Advances in Theory and Practice. XII, 235 pages. 2004.

Vol. 3051: R. Berghammer, B. Möller, G. Struth (Eds.), Relational and Kleene-Algebraic Methods in Computer Science. X, 279 pages. 2004.

Vol. 3050: J. Domingo-Ferrer, V. Torra (Eds.), Privacy in Statistical Databases. IX, 367 pages. 2004.

Vol. 3049: M. Bruynooghe, K.-K. Lau (Eds.), Program Development in Computational Logic. VIII, 539 pages. 2004.

Vol. 3047: F. Oquendo, B. Warboys, R. Morrison (Eds.), Software Architecture. X, 279 pages. 2004.

Vol. 3046: A. Laganà, M.L. Gavrilova, V. Kumar, Y. Mun, C.K. Tan, O. Gervasi (Eds.), Computational Science and Its Applications – ICCSA 2004. LIII, 1016 pages. 2004.

Vol. 3045: A. Laganà, M.L. Gavrilova, V. Kumar, Y. Mun, C.K. Tan, O. Gervasi (Eds.), Computational Science and Its Applications – ICCSA 2004. LIII, 1040 pages. 2004.

Vol. 3044: A. Laganà, M.L. Gavrilova, V. Kumar, Y. Mun, C.K. Tan, O. Gervasi (Eds.), Computational Science and Its Applications – ICCSA 2004. LIII, 1140 pages. 2004.

Vol. 3043: A. Laganà, M.L. Gavrilova, V. Kumar, Y. Mun, C.K. Tan, O. Gervasi (Eds.), Computational Science and Its Applications – ICCSA 2004. LIII, 1180 pages. 2004.

Vol. 3042: N. Mitrou, K. Kontovasilis, G.N. Rouskas, I. Iliadis, L. Merakos (Eds.), NETWORKING 2004, Networking Technologies, Services, and Protocols; Performance of Computer and Communication Networks; Mobile and Wireless Communications. XXXIII, 1519 pages. 2004.

Vol. 3040: R. Conejo, M. Urretavizcaya, J.-L. Pérez-de-la-Cruz (Eds.), Current Topics in Artificial Intelligence. XIV, 689 pages. 2004. (Subseries LNAI).

Vol. 3039: M. Bubak, G.D.v. Albada, P.M. Sloot, J.J. Dongarra (Eds.), Computational Science - ICCS 2004. LXVI, 1271 pages. 2004.

Vol. 3038: M. Bubak, G.D.v. Albada, P.M. Sloot, J.J. Dongarra (Eds.), Computational Science - ICCS 2004. LXVI, 1311 pages. 2004.

Vol. 3037: M. Bubak, G.D.v. Albada, P.M. Sloot, J.J. Dongarra (Eds.), Computational Science - ICCS 2004. LXVI, 745 pages. 2004.

Vol. 3036: M. Bubak, G.D.v. Albada, P.M. Sloot, J.J. Dongarra (Eds.), Computational Science - ICCS 2004. LXVI, 713 pages. 2004.

Vol. 3035: M.A. Wimmer (Ed.), Knowledge Management in Electronic Government. XII, 326 pages. 2004. (Subseries LNAI).

Vol. 3034: J. Favela, E. Menasalvas, E. Chávez (Eds.), Advances in Web Intelligence. XIII, 227 pages. 2004. (Subseries LNAI).

Vol. 3033: M. Li, X.-H. Sun, Q. Deng, J. Ni (Eds.), Grid and Cooperative Computing. XXXVIII, 1076 pages. 2004.

Vol. 3032: M. Li, X.-H. Sun, Q. Deng, J. Ni (Eds.), Grid and Cooperative Computing. XXXVII, 1112 pages. 2004.

Vol. 3031: A. Butz, A. Krueger, P. Olivier (Eds.), Smart Graphics. X, 165 pages. 2004.

Vol. 3030: P. Giorgini, B. Henderson-Sellers, M. Winikoff (Eds.), Agent-Oriented Information Systems. XIV, 207 pages. 2004. (Subseries LNAI).

Vol. 3029: B. Orchard, C. Yang, M. Ali (Eds.), Innovations in Applied Artificial Intelligence. XXI, 1272 pages. 2004. (Subseries LNAI).

Vol. 3028: D. Neuenschwander, Probabilistic and Statistical Methods in Cryptology. X, 158 pages. 2004.

Vol. 3027: C. Cachin, J. Camenisch (Eds.), Advances in Cryptology - EUROCRYPT 2004. XI, 628 pages. 2004.

Vol. 3026: C. Ramamoorthy, R. Lee, K.W. Lee (Eds.), Software Engineering Research and Applications. XV, 377 pages. 2004.

Table of Contents

Rough Sets – Introduction

Some Issues on Rough Sets 1
Zdzisław Pawlak

Rough Sets – Theory

Learning Rules from Very Large Databases Using Rough Multisets 59
Chien-Chung Chan

Data with Missing Attribute Values:
Generalization of Indiscernibility Relation and Rule Induction 78
Jerzy W. Grzymala-Busse

Generalizations of Rough Sets and Rule Extraction...................... 96
Masahiro Inuiguchi

Towards Scalable Algorithms for Discovering Rough Set Reducts......... 120
Marzena Kryszkiewicz and Katarzyna Cichoń

Variable Precision Fuzzy Rough Sets 144
Alicja Mieszkowicz-Rolka and Leszek Rolka

Greedy Algorithm of Decision Tree Construction for Real Data Tables 161
Mikhail Ju. Moshkov

Consistency Measures for Conflict Profiles............................ 169
Ngoc Thanh Nguyen and Michal Malowiecki

Layered Learning for Concept Synthesis.............................. 187
Sinh Hoa Nguyen, Jan Bazan, Andrzej Skowron, and Hung Son Nguyen

Basic Algorithms and Tools
for Rough Non-deterministic Information Analysis 209
Hiroshi Sakai and Akimichi Okuma

A Partition Model of Granular Computing 232
Yiyu Yao

Rough Sets – Applications

Musical Phrase Representation and Recognition
by Means of Neural Networks and Rough Sets 254
Andrzej Czyzewski, Marek Szczerba, and Bozena Kostek

Processing of Musical Metadata Employing Pawlak's Flow Graphs 279
Bozena Kostek and Andrzej Czyzewski

Data Decomposition and Decision Rule Joining for Classification of Data with Missing Values 299
Rafał Latkowski and Michał Mikołajczyk

Rough Sets and Relational Learning 321
R.S. Milton, V. Uma Maheswari, and Arul Siromoney

Approximation Space for Software Models 338
James F. Peters and Sheela Ramanna

Application of Rough Sets to Environmental Engineering Models 356
Robert H. Warren, Julia A. Johnson, and Gordon H. Huang

Rough Set Theory and Decision Rules in Data Analysis of Breast Cancer Patients 375
Jerzy Załuski, Renata Szoszkiewicz, Jerzy Krysiński, and Jerzy Stefanowski

Independent Component Analysis, Principal Component Analysis and Rough Sets in Face Recognition 392
Roman W. Świniarski and Andrzej Skowron

Author Index 405

Some Issues on Rough Sets

Zdzisław Pawlak[1,2]

[1] Institute for Theoretical and Applied Informatics
Polish Academy of Sciences
ul. Bałtycka 5, 44-100 Gliwice, Poland
[2] Warsaw School of Information Technology*
ul. Newelska 6, 01-447 Warsaw, Poland
zpw@ii.pw.edu.pl

1 Introduction

The aim of this paper is to give rudiments of rough set theory and present some recent research directions proposed by the author.

Rough set theory is a new mathematical approach to imperfect knowledge.

The problem of imperfect knowledge has been tackled for a long time by philosophers, logicians and mathematicians. Recently it became also a crucial issue for computer scientists, particularly in the area of artificial intelligence. There are many approaches to the problem of how to understand and manipulate imperfect knowledge. The most successful one is, no doubt, the fuzzy set theory proposed by Lotfi Zadeh [1].

Rough set theory proposed by the author in [2] presents still another attempt to this problem. This theory has attracted attention of many researchers and practitioners all over the world, who have contributed essentially to its development and applications. Rough set theory overlaps with many other theories. However we will refrain to discuss these connections here. Despite this, rough set theory may be considered as an independent discipline in its own right.

Rough set theory has found many interesting applications. The rough set approach seems to be of fundamental importance to AI and cognitive sciences, especially in the areas of machine learning, knowledge acquisition, decision analysis, knowledge discovery from databases, expert systems, inductive reasoning and pattern recognition.

The main advantage of rough set theory in data analysis is that it does not need any preliminary or additional information about data – like probability in statistics, or basic probability assignment in Dempster-Shafer theory, grade of membership or the value of possibility in fuzzy set theory.

One can observe the following about the rough set approach:

- introduction of efficient algorithms for finding hidden patterns in data,
- determination of minimal sets of data (data reduction),
- evaluation of the significance of data,
- generation of sets of decision rules from data,

* Former University of Information Technology and Management.

J.F. Peters et al. (Eds.): Transactions on Rough Sets I, LNCS 3100, pp. 1–58, 2004.

- easy-to-understand formulation,
- straightforward interpretation of obtained results,
- suitability of many of its algorithms for parallel processing.

Rough set theory has been extended in many ways (see, e.g., [3–17]) but we will not discuss these issues in this paper.

Basic ideas of rough set theory and its extensions, as well as many interesting applications can be found in books (see, e.g., [18–27, 12, 28–30]), special issues of journals (see, e.g., [31–34, 34–38]), proceedings of international conferences (see, e.g., [39–49]), tutorials (e.g., [50–53]), and on the internet (see, e.g., www.roughsets.org, logic.mimuw.edu.pl,rsds.wsiz.rzeszow.pl).

The paper is organized as follows:

Section 2 (Basic Concepts) contains general formulation of basic ideas of rough set theory together with brief discussion of its place in classical set theory.

Section 3 (Rough Sets and Reasoning from Data) presents the application of rough set concept to reason from data (data mining).

Section 4 (Rough Sets and Bayes' Theorem) gives a new look on Bayes' theorem and shows that Bayes' rule can be used differently to that offered by classical Bayesian reasoning methodology.

Section 5 (Rough Sets and Conflict Analysis) discuses the application of rough set concept to study conflict.

In Section 6 (Data Analysis and Flow Graphs) we show that many problems in data analysis can be boiled down to flow analysis in a flow network.

This paper is a modified version of lectures delivered at the Taragona University seminar on Formal Languages and Rough Sets in August 2003.

2 Rough Sets – Basic Concepts

2.1 Introduction

In this section we give some general remarks on a concept of a set and the place of rough sets in set theory.

The concept of a set is fundamental for the whole mathematics. Modern set theory was formulated by George Cantor [54].

Bertrand Russell discovered that the intuitive notion of a set proposed by Cantor leads to antinomies [55]. Two kinds of remedy for this discontent have been proposed: axiomatization of Cantorian set theory and alternative set theories.

Another issue discussed in connection with the notion of a set or a concept is vagueness (see, e.g., [56–61]). Mathematics requires that all mathematical notions (including set) must be exact (Gottlob Frege [62]). However, philosophers and recently computer scientists have become interested in vague concepts.

In fuzzy set theory vagueness is defined by graduated membership.

Rough set theory expresses vagueness, not by means of membership, but employing a boundary region of a set. If the boundary region of a set is empty it means that the set is crisp, otherwise the set is rough (inexact). Nonempty boundary region of a set means that our knowledge about the set is not sufficient to define the set precisely.

The detailed analysis of sorities paradoxes for vague concepts using rough sets and fuzzy sets is presented in [63].

In this section the relationship between sets, fuzzy sets and rough sets will be outlined and briefly discussed.

2.2 Sets

The notion of a set is not only basic for mathematics but it also plays an important role in natural language. We often speak about sets (collections) of various objects of interest, e.g., collection of books, paintings, people etc. Intuitive meaning of a set according to some dictionaries is the following:

"A number of things of the same kind that belong or are used together."

Webster's Dictionary

"Number of things of the same kind, that belong together because they are similar or complementary to each other."

The Oxford English Dictionary

Thus a set is a collection of things which are somehow related to each other but the nature of this relationship is not specified in these definitions.

In fact these definitions are due to Cantor [54], which reads as follows:

"Unter einer Mannigfaltigkeit oder Menge verstehe ich nämlich allgenein jedes Viele, welches sich als Eines denken lässt, d.h. jeden Inbegriff bestimmter Elemente, welcher durch ein Gesetz zu einem Ganzen verbunden werden kann."

Thus according to Cantor a set is a collection of any objects, which according to some law can be considered as a whole.

All mathematical objects, e.g., relations, functions, numbers, etc., are some kind of sets. In fact set theory is needed in mathematics to provide rigor.

Russell discovered that the Cantorian notion of a set leads to antinomies (contradictions). One of the best known antinomies called the powerset antinomy goes as follows: consider (infinite) set X of all sets. Thus X is the greatest set. Let Y denote the set of all subsets of X. Obviously Y is greater then X, because the number of subsets of a set is always greater the number of its elements. Hence X is not the greatest set as assumed and we arrived at contradiction.

Thus the basic concept of mathematics, the concept of a set, is contradictory. This means that a set cannot be a collection of arbitrary elements as was stipulated by Cantor.

As a remedy for this defect several improvements of set theory have been proposed. For example,

- Axiomatic set theory (Zermello and Fraenkel, 1904).
- Theory of types (Whitehead and Russell, 1910).
- Theory of classes (v. Neumann, 1920).

All these improvements consist in restrictions, put on objects which can form a set. The restrictions are expressed by properly chosen axioms, which say how the set can be build. They are called, in contrast to Cantors' intuitive set theory, axiomatic set theories.

Instead of improvements of Cantors' set theory by its axiomatization, some mathematicians proposed escape from classical set theory by creating completely new idea of a set, which would free the theory from antinomies. Some of them are listed below.

- Mereology (Leśniewski, 1915).
- Alternative set theory (Vopenka, 1970).
- "Penumbral" set theory (Apostoli and Kanada, 1999).

No doubt the most interesting proposal was given by Stanisaw Leśniewski [64], who proposed instead of membership relation between elements and sets, employed in classical set theory, the relation of "being a part". In his set theory, called *mereology*, this relation is a fundamental one.

None of the three mentioned above "new" set theories were accepted by mathematicians, however Leśniewski's mereology attracted some attention of philosophers and recently also computer scientists, (e.g., Lech Polkowski and Andrzej Skowron [6]).

In classical set theory a set is uniquely determined by its elements. In other words, this means that every element must be uniquely classified as belonging to the set or not. In contrast, the notion of a beautiful painting is vague, because we are unable to classify uniquely all paintings into two classes: beautiful and not beautiful. Thus *beauty* is not a precise but a vague concept. That is to say the notion of a set is a *crisp* (precise) one. For example, the set of odd numbers is crisp because every number is either odd or even. In mathematics we have to use crisp notions, otherwise precise reasoning would be impossible. However philosophers for many years were interested also in *vague* (imprecise) notions.

Almost all concepts we are using in natural language are vague. Therefore common sense reasoning based on natural language must be based on vague concepts and not on classical logic. This is why vagueness is important for philosophers and recently also for computer scientists.

Vagueness is usually associated with the boundary region approach (i.e., existence of objects which cannot be uniquely classified to the set or its complement) which was first formulated in 1893 by the father of modern logic Gottlob Frege [62], who wrote:

"Der Begriff muss scharf begrenzt sein. Einem unscharf begrenzten Begriffe würde ein Bezirk entsprechen, der nicht überall eine scharfe Grenzlinie hätte, sondern stellenweise ganz verschwimmend in die Umgebung überginge. Das wäre eigentlich gar kein Bezirk; und so wird ein unscharf definirter Begriff mit Unrecht Begriff genannt. Solche begriffsartige Bildungen kann die Logik nicht als Begriffe anerkennen; es ist unmöglich, von ihnen genaue Gesetze aufzustellen. Das Gesetz des ausgeschlossenen Dritten ist ja eigentlich nur in anderer Form die Forderung, dass der Begriff scharf begrenzt sei. Ein beliebiger Gegenstand x fällt entweder unter der Begriff y, oder er fällt nicht unter ihn: *tertium non datur.*"

Thus according to Frege

"The concept must have a sharp boundary. To the concept without a sharp boundary there would correspond an area that had not a sharp boundary-line all around."

That is, mathematics must use crisp, not vague concepts, otherwise it would be impossible to reason precisely.

Summing up, vagueness is

- Not allowed in mathematics.
- Interesting for philosophy.
- Necessary for computer science.

2.3 Fuzzy Sets

Zadeh proposed completely new, elegant approach to vagueness called *fuzzy set theory* [1]. In his approach an element can belong to a set to a degree $k (0 \leq k \leq 1)$, in contrast to classical set theory where an element must definitely belong or not to a set. For example, in classical set theory language we can state that one is definitely ill or healthy, whereas in fuzzy set theory we can say that someone is ill (or healthy) in 60 percent (i.e., in the degree 0.6). Of course, at once the question arises where we get the value of degree from. This issue raised a lot of discussion, but we will refrain from considering this problem here.

Thus fuzzy membership function can be presented as

$$\mu_X(x) \in < 0, 1 >,$$

where, X is a set and x is an element.

Let us observe that the definition of fuzzy set involves more advanced mathematical concepts, real numbers and functions, whereas in classical set theory the notion of a set is used as a fundamental notion of whole mathematics and is used to derive any other mathematical concepts, e.g., numbers and functions. Consequently fuzzy set theory cannot replace classical set theory, because, in fact, the theory is needed to define fuzzy sets.

Fuzzy membership function has the following properties:

$$\begin{aligned} \mu_{U-X}(x) &= 1 - \mu_X(x) \text{ for any } x \in U, \\ \mu_{X \cup Y}(x) &= max(\mu_X(x), \mu_Y(x)) \text{ for any } x \in U, \\ \mu_{X \cap Y}(x) &= min(\mu_X(x), \mu_Y(x)) \text{ for any } x \in U. \end{aligned} \tag{1}$$

This means that the membership of an element to the union and intersection of sets is uniquely determined by its membership to constituent sets. This is a very nice property and allows very simple operations on fuzzy sets, which is a very important feature both theoretically and practically.

Fuzzy set theory and its applications developed very extensively over recent years and attracted attention of practitioners, logicians and philosophers worldwide.

2.4 Rough Sets

Rough set theory [2, 18] is still another approach to vagueness. Similarly to fuzzy set theory it is not an alternative to classical set theory but it is embedded in it. Rough set theory can be viewed as a specific implementation of Frege's idea of vagueness, i.e., imprecision in this approach is expressed by a boundary region of a set, and not by a partial membership, like in fuzzy set theory.

Rough set concept can be defined quite generally by means of topological operations, *interior* and *closure*, called *approximations*.

Let us describe this problem more precisely. Suppose we are given a set of objects U called the *universe* and an indiscernibility relation $R \subseteq U \times U$, representing our lack of knowledge about elements of U. For the sake of simplicity we assume that R is an equivalence relation. Let X be a subset of U. We want to characterize the set X with respect to R. To this end we will need the basic concepts of rough set theory given below.

- The *lower approximation* of a set X (with respect to R) is the set of all objects, which can be for *certain* classified as X with respect to R (are *certainly* X with respect to R).
- The *upper approximation* of a set X (with respect to R) is the set of all objects which can be *possibly* classified as X with respect to R (are *possibly* X with respect to R).
- The *boundary region* of a set X (with respect to R) is the set of all objects, which can be classified neither as X nor as not-X with respect to R.

Now we are ready to give the definition of rough sets.

- Set X is *crisp* (exact with respect to R), if the boundary region of X is empty.
- Set X is *rough* (inexact with respect to R), if the boundary region of X is nonempty.

Thus a set is *rough* (imprecise) if it has nonempty boundary region; otherwise the set is *crisp* (precise). This is exactly the idea of vagueness proposed by Frege.

The approximations and the boundary region can be defined more precisely. To this end we need some additional notation.

The equivalence class of R determined by element x will be denoted by $R(x)$. The indiscernibility relation in certain sense describes our lack of knowledge about the universe. Equivalence classes of the indiscernibility relation, called *granules* generated by R, represent elementary portion of knowledge we are able to perceive due to R. Thus in view of the indiscernibility relation, in general, we are unable to observe individual objects but we are forced to reason only about the accessible granules of knowledge.

Formal definitions of approximations and the boundary region are as follows:

R-lower approximation of X

$$R_*(X) = \bigcup_{x \in U} \{R(x) : R(x) \subseteq X\}, \tag{2}$$