

Ajith Abraham, Rafael Falcón, and Rafael Bello (Eds.)

Rough Set Theory: A True Landmark in Data Analysis

## Studies in Computational Intelligence, Volume 174

### 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. 151. Tomasz G. Smolinski, Mariofanna G. Milanova  
and Aboul-Ella Hassanien (Eds.)  
*Computational Intelligence in Biomedicine and Bioinformatics*,  
2008  
ISBN 978-3-540-70776-9

Vol. 152. Jarosław Stepaniuk  
*Rough – Granular Computing in Knowledge Discovery and Data  
Mining*, 2008  
ISBN 978-3-540-70800-1

Vol. 153. Carlos Cotta and Jano van Hemert (Eds.)  
*Recent Advances in Evolutionary Computation for  
Combinatorial Optimization*, 2008  
ISBN 978-3-540-70806-3

Vol. 154. Oscar Castillo, Patricia Melin, Janusz Kacprzyk and  
Witold Pedrycz (Eds.)  
*Soft Computing for Hybrid Intelligent Systems*, 2008  
ISBN 978-3-540-70811-7

Vol. 155. Hamid R. Tizhoosh and M. Ventresca (Eds.)  
*Oppositional Concepts in Computational Intelligence*, 2008  
ISBN 978-3-540-70826-1

Vol. 156. Dawn E. Holmes and Lakhmi C. Jain (Eds.)  
*Innovations in Bayesian Networks*, 2008  
ISBN 978-3-540-85065-6

Vol. 157. Ying-ping Chen and Meng-Hiot Lim (Eds.)  
*Linkage in Evolutionary Computation*, 2008  
ISBN 978-3-540-85067-0

Vol. 158. Marina Gavrilova (Ed.)  
*Generalized Voronoi Diagram: A Geometry-Based Approach to  
Computational Intelligence*, 2009  
ISBN 978-3-540-85125-7

Vol. 159. Dimitri Plemenos and Georgios Miaoulis (Eds.)  
*Artificial Intelligence Techniques for Computer Graphics*, 2009  
ISBN 978-3-540-85127-1

Vol. 160. P. Rajasekaran and Vasanth Kalyani David  
*Pattern Recognition using Neural and Functional Networks*,  
2009  
ISBN 978-3-540-85129-5

Vol. 161. Francisco Baptista Pereira and Jorge Tavares (Eds.)  
*Bio-inspired Algorithms for the Vehicle Routing Problem*, 2009  
ISBN 978-3-540-85151-6

Vol. 162. Costin Badica, Giuseppe Mangioni,  
Vincenza Carchiolo and Dumitru Dan Burdescu (Eds.)  
*Intelligent Distributed Computing, Systems and Applications*,  
2008  
ISBN 978-3-540-85256-8

Vol. 163. Pawel Delimata, Mikhail Ju. Moshkov,  
Andrzej Skowron and Zbigniew Suraj  
*Inhibitory Rules in Data Analysis*, 2009  
ISBN 978-3-540-85637-5

Vol. 164. Nadia Nedjah, Luiza de Macedo Mourelle,  
Janusz Kacprzyk, Felipe M.G. França  
and Alberto Ferreira de Souza (Eds.)  
*Intelligent Text Categorization and Clustering*, 2009  
ISBN 978-3-540-85643-6

Vol. 165. Djamel A. Zighed, Shusaku Tsumoto,  
Zbigniew W. Ras and Hakim Hacid (Eds.)  
*Mining Complex Data*, 2009  
ISBN 978-3-540-88066-0

Vol. 166. Constantinos Koutsojannis and Spiros Sirmakessis  
(Eds.)  
*Tools and Applications with Artificial Intelligence*, 2009  
ISBN 978-3-540-88068-4

Vol. 167. Ngoc Thanh Nguyen and Lakhmi C. Jain (Eds.)  
*Intelligent Agents in the Evolution of Web and Applications*, 2009  
ISBN 978-3-540-88070-7

Vol. 168. Andreas Tolk and Lakhmi C. Jain (Eds.)  
*Complex Systems in Knowledge-based Environments: Theory,  
Models and Applications*, 2009  
ISBN 978-3-540-88074-5

Vol. 169. Nadia Nedjah, Luiza de Macedo Mourelle and  
Janusz Kacprzyk (Eds.)  
*Innovative Applications in Data Mining*, 2009  
ISBN 978-3-540-88044-8

Vol. 170. Lakhmi C. Jain and Ngoc Thanh Nguyen (Eds.)  
*Knowledge Processing and Decision Making in Agent-Based  
Systems*, 2009  
ISBN 978-3-540-88048-6

Vol. 171. Chi-Keong Goh, Yew-Soon Ong and Kay Chen Tan  
(Eds.)  
*Multi-Objective Memetic Algorithms*, 2009  
ISBN 978-3-540-88050-9

Vol. 172. I-Hsien Ting and Hui-Ju Wu (Eds.)  
*Web Mining Applications in E-Commerce and E-Services*, 2009  
ISBN 978-3-540-88080-6

Vol. 173. Tobias Grosche  
*Computational Intelligence in Integrated Airline Scheduling*,  
2009  
ISBN 978-3-540-89886-3

Vol. 174. Ajith Abraham, Rafael Falcón and Rafael Bello (Eds.)  
*Rough Set Theory: A True Landmark in Data Analysis*, 2009  
ISBN 978-3-540-89886-3

Ajith Abraham  
Rafael Falcón  
Rafael Bello  
(Eds.)

# Rough Set Theory: A True Landmark in Data Analysis

 Springer

Ajith Abraham  
Centre of Excellence for Quantifiable Quality of  
Service  
Norwegian University of Science and  
Technology  
O.S. Bragstads plass 2E  
N-7491 Trondheim  
Norway  
Email: ajith.abraham@ieee.org

Rafael Bello  
Department of Computer Science  
Central University of Las Villas  
Carretera Camajuaní km 5 1/2  
Santa Clara, Villa Clara 54830  
CUBA  
Email: rbello@uclv.edu.cu

Rafael Falcón  
School of Information Technology and  
Engineering (SITE)  
University of Ottawa  
800 King Edward Ave., Ottawa ON K1N 6N5  
Canada  
Email: rfalcon@uottawa.ca

ISBN 978-3-540-89920-4

e-ISBN 978-3-540-89921-1

DOI 10.1007/978-3-540-89921-1

Studies in Computational Intelligence

ISSN 1860949X

Library of Congress Control Number: 2008941015

© 2009 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

To my loving parents

---

Ajith Abraham

To the Northey family,  
for their continuous support  
and encouragement

---

Rafael Falcón

To Marilyn and my daughters

---

Rafael Bello



---

## Preface

Along the years, rough set theory has earned a well-deserved reputation as a sound methodology for dealing with imperfect knowledge in a simple though fully mathematically supported way. The increasing number of international conferences and prestigious journals addressing the topic are a vivid example of the rapid advancements in the field and the significant outbreak experienced during the last few years. We believe there are at least three major reasons that could possibly shed light on the expanding interest of researchers worldwide about rough sets:

1. Their formulation is deeply rooted on the very nature of ordinary sets, one of the cornerstones of today's mathematical and logical reasoning schemes. This, in turn, facilitates its understanding to a very broad audience.
2. Their role as a complementary and reinforcing paradigm of many other research areas and applications, which has led to a plethora of hybrid models encompassing a still growing number of traditional methodologies, most of them falling under the umbrella of soft computing and computational intelligence. Fuzzy sets, evolutionary algorithms, swarm intelligence, neural networks and decision trees are just a few in a long list of approaches which have largely profited from the incorporation of rough sets into their main research avenues, eventually improving the performance and interpretation of current data analysis tools.
3. The unique capability of rough sets for performing difficult data analysis tasks such as attribute reduction, rule generation, data set characterization in terms of attribute values or the computation of attribute significance and dependency, without leaning upon any preliminary or additional information about data, plus the relative simplicity of the computational implementations. Rough-set-driven data analysis stands today as a true landmark among peer methodologies, enjoying a continued success yet open to future research directions and areas posing severe restrictions and novel computing environments, such as ubiquitous computing.

This edited volume aims at continue stressing the benefits of applying rough sets in many real-life situations while still keeping an eye on topological aspects of the theory as well as strengthening its linkage with other reasoning paradigms.

Part I is comprised of four chapters which intend to outline some theoretical contributions. The original concept of rough set introduced by Z. Pawlak has been extended in many different directions shortly after its very inception. Some of these extensions have to do with loosening the requirement of the underlying indiscernibility relation to be an equivalence relation. Chapter 1 is devoted to present two different extensions of rough sets which rely upon fuzzy proximity and intuitionistic fuzzy proximity relations, respectively. The parallelism of their basic properties with those of the conventional rough set model becomes evident and several real life applications illustrating the practical use of these generalizations are considered along the chapter.

Categories arise in mathematics and appear frequently in computer science where algebraic and logical notions have powerful representations using categorical constructions. Chapter 2 proposes a generalization of the standard rough set operators, which are based on relations involving ordinary sets, to embrace partially-ordered monads from which a theory of rough monads can be derived. This turns rough set theory into a categorical methodology with a broad avenue of further developments both from the theoretical and application standpoints. The chapter illustrates the use of rough monads in management of drug interactions and medical diagnosis.

Chapter 3 is concerned with the interpretation of both rough set and knowledge space theories in light of granular computing. In order to arrive at a unified framework bridging these two methodologies, it becomes necessary to analyze their granular structures and approximations regardless the different types of elementary granules and granular structures each of them utilizes for information representation and processing. The result is a multilevel granular structure with a set-theoretic formulation of granular structures as its building blocks. The framework allows studying rough set analysis and knowledge spaces in a common setting with results being applied interchangeably.

Finally, a study on some topological aspects of rough sets and approximation of classifications is the subject of Chapter 4. The author sheds light on the different types of unions and intersections of rough sets in addition to the formulation of a new concept of rough equivalence which captures the approximate equality of sets at a higher level than the existing notion of rough equality. The approximation of classifications is also revisited and some observations about the properties of rules generated from information systems and their inherent structure are made.

The second part of the book has to do with some applications of rough sets in data mining activities. An interesting direction in rough set research is that of clustering algorithms. Chapter 5 elaborates on how to endow rough clustering approaches with mechanisms of partial supervision, thus far being solely conceived for fuzzy clustering. Partial supervision comes in the form of fuzzy membership grades and allocation of patterns to positive or negative regions



of the sought knowledge structures (clusters). The chapter also provides some insight on the influence of the type and number of knowledge-based hints being furnished to the system.

Rough set theory has been regarded since its very inception as a powerful and feasible methodology for performing data mining and knowledge discovery activities. Prominent results, also embracing the financial and economic realms, can be witnessed in many research studies across literature. Although discriminant analysis continues to be the preferred and most popular approach for making predictions in these areas, the application of rough-set-based data analysis leads, most of the times, to high overall prediction accuracy and interpretable rule bases. Despite this fact, prediction still remains a challenging and difficult goal to achieve. In light of this, Chapter 6 presents a generic scheme for generating decision rules for stock market prediction. Numerical attributes are handled via rough set with Boolean reasoning (RSBR) approach and the rough confusion matrix is used to evaluate the performance of the predicted reducts and classes.

On the other hand, a framework for modeling uncertainty in Web caching scenarios has been crafted and formalized in Chapter 7. Web caching and prefetching have been widely acknowledged as effective schemes to relieve the service bottleneck and minimize the user access latency, both hot problems still encountered in today's web applications. The proposed framework comprises three modules, which exploit rough set capabilities to reduce the rules of log files and, at the same time, enhance the prediction performance of the user behavior. The induction engine module, for example, has an ID3-like learning algorithm based on the minimum entropy principle.

The last four chapters present some rough hybrid models tailored for classification and attribute reduction purposes. Hybridization among computational intelligence techniques has been regarded as both a successful paradigm and a well-settled tendency in decision-making and machine learning fields, thus giving rise to stronger, synergetic models, which can better cope with the huge amounts of information available nowadays and draw precise knowledge in a more interpretable way. Chapter 8 provides an overview of some rough hybrid approaches and a thorough comparative study on their performance versus prominent, conventional classifiers such as support vector machines and decision trees. Hybrid models include neuro-fuzzy decision trees, rough-neuro-fuzzy decision trees and fuzzy-rough classification trees in the context of software defect classification. Conclusions about the types of metric data that need to be collected and the readability of the generated rules are derived.

One of the most appealing properties of rough set theory lies in its ability to spot and remove redundant attributes which are present in the information systems we usually deal with. Given the exponential computational complexity of feature selection, the call for heuristic approaches becomes a must. In particular, evolutionary algorithms offer impressive optimization capabilities due to the multiple handling of potential problem solutions. Chapters 9 and 10 provide an overview of such nature-inspired heuristics in their quest for suitable feature subsets. Settled at the very core of the approaches, some concepts borrowed from

rough set theory have become an integral part of the guiding principle ruling the optimization scheme. Chapter 9 moves from greedy to swarm intelligence approaches, also featuring a novel dynamic evolutionary technique and a search strategy that substantially reduces the time needed to find appropriate solutions (reducts). Chapter 10 emphasizes on particle swarm optimization and its application to functional magnetic resonance imaging.

The main workflow of rough set data analysis (RSDA) is applied in Chapter 11 to draw knowledge from a set of events captured by protection, control and monitoring devices in a substation network. Once again, the reduct computation phase is largely profited from the presence of evolutionary approaches (this time, Genetic Algorithms), which also make possible to dynamically add more events into the system owing to their amazing adaptive capabilities. The ensuing rule base identifies and isolates the most probable faulty section in the network, hence improving the outage response time. The case studies span from simple scenarios to complex distribution substations comprised of various types of relay models and even include time series data.

We want to thank the authors for submitting their contributions to this volume and the prominent experts in the field all over the world who were willing to devote a share of their valuable time in reviewing the articles and providing insightful remarks.

The kind assistance we got during the publication phase from Thomas Ditzinger and Heather King of Springer-Verlag is gratefully acknowledged.

Santa Clara, Cuba  
September 2008

Ajith Abraham  
Rafael Falcón  
Rafael Bello

---

## Contents

---

### Part I: Theoretical Contributions to Rough Set Theory

---

<b>Rough Sets on Fuzzy Approximation Spaces and Intuitionistic Fuzzy Approximation Spaces</b> <i>B.K. Tripathy</i> .....	3
<b>Categorical Innovations for Rough Sets</b> <i>P. Eklund, M.A. Galán, J. Karlsson</i> .....	45
<b>Granular Structures and Approximations in Rough Sets and Knowledge Spaces</b> <i>Yiyu Yao, Duoqian Miao, Feifei Xu</i> .....	71
<b>On Approximation of Classifications, Rough Equalities and Rough Equivalences</b> <i>B.K. Tripathy</i> .....	85

---

### Part II: Rough Set Data Mining Activities

---

<b>Rough Clustering with Partial Supervision</b> <i>Rafael Falcón, Gwanggil Jeon, Rafael Bello, Jechang Jeong</i> .....	137
<b>A Generic Scheme for Generating Prediction Rules Using Rough Sets</b> <i>Hameed Al-Qaheri, Aboul Ella Hassanien, Ajith Abraham</i> .....	163
<b>Rough Web Caching</b> <i>Sarina Sulaiman, Siti Mariyam Shamsuddin, Ajith Abraham</i> .....	187
<b>Software Defect Classification: A Comparative Study of Rough-Neuro-fuzzy Hybrid Approaches with Linear and Non-linear SVMs</b> <i>Rajen Bhatt, Sheela Ramanna, James F. Peters</i> .....	213

---

**Part III: Rough Hybrid Models to Classification and Attribute  
Reduction**

---

<b>Rough Sets and Evolutionary Computation to Solve the Feature Selection Problem</b> <i>Rafael Bello, Yudel Gómez, Yailé Caballero, Ann Nowe, Rafael Falcón . . .</i>	235
<b>Nature Inspired Population-Based Heuristics for Rough Set Reduction</b> <i>Hongbo Liu, Ajith Abraham, Yanheng Li . . . . .</i>	261
<b>Developing a Knowledge-Based System Using Rough Set Theory and Genetic Algorithms for Substation Fault Diagnosis</b> <i>Ching Lai Hor, Peter Crossley, Simon Watson, Dean Millar . . . . .</i>	279
<b>Index . . . . .</b>	321
<b>Author Index . . . . .</b>	323

---

## List of Contributors

**Aboul Ella Hassanien**

Information Tech. Dept., Cairo Univ.  
5 Ahamed Zewal St.  
Orman, Giza, Egypt  
a.hassanien@fci-cu.edu.eg

**Ajith Abraham**

Center for Quantifiable QoS  
Norwegian Univ. of Sci. & Tech.  
O.S. Bragstads plass 2E NO-7491  
Trondheim, Norway  
ajith.abraham@ieee.org

**Ann Nowe**

CoMo Lab, Comp Science Dept.  
Vrije Universiteit Brussel  
Brussels, Belgium  
asnowe@info.vub.ac.be

**Balakrushna Tripathy**

School of Computing Sciences  
VIT University  
Vellore, T.N. 632 014, India  
tripathybk@rediffmail.com

**Ching Lai Hor**

CSM Renewable Energy Group  
University of Exeter in Cornwall  
Tremough Campus, Penryn  
TR10 9EZ, United Kingdom  
c.l.hor@exeter.ac.uk

**Dean Millar**

CSM Renewable Energy Group  
University of Exeter in Cornwall  
Tremough Campus, Penryn  
TR10 9EZ, United Kingdom  
dlmillar@csm.ex.ac.uk

**Duoqian Miao**

Dept. of Comp. Science & Technology  
Tongji University  
Shanghai, China 201804  
miaoduoqian@163.com

**Feifei Xu**

Department of Computer Science  
University of Regina  
Regina, Saskatchewan  
S4S 0A2, Canada  
xufeifei1983@hotmail.com

**Gwanggil Jeon**

Dept. of Electronics & Comp. Eng.  
Hanyang University  
17 Haengdang-dong, Seongdong-gu  
Seoul, Korea  
windcap315@ece.hanyang.ac.kr

**Hameed Al-Qaheri**

Dept. of Quant. Methods & Inf. Syst.  
College of Business Administration  
Kuwait University, Kuwait  
alqaheri@cba.edu.kw

**Hongbo Liu**

School of Comp. Science and Eng.  
Dalian Maritime University  
116026 Dalian, China  
lhb@dlut.edu.cn

**James F. Peters**

Computational Intelligence Lab  
Electrical & Comp. Eng. Department  
University of Manitoba  
Winnipeg, Manitoba  
R3T 5V6 Canada  
jfpeters@ee.umanitoba.ca

**Jechang Jeong**

Dept. of Electronics & Comp. Eng.  
Hanyang University  
17 Haengdang-dong, Seongdong-gu  
Seoul, Korea  
jjeong@ece.hanyang.ac.kr

**Johan Karlsson**

Department of Comp. Architecture  
University of Málaga, Spain  
johan@ac.uma.es

**Maria A. Galán**

Department of Applied Mathematics  
University of Málaga, Spain  
magalan@ctima.uma.es

**Patrik Eklund**

Department of Computing Science  
Umeå University, Sweden  
peklund@cs.umu.se

**Peter Crossley**

Electrical Energy and Power Systems  
University of Manchester  
P.O.Box 88, Manchester  
M60 1QD, United Kingdom  
p.crossley@manchester.ac.uk

**Rafael Bello**

Central Univ. of Las Villas (UCLV)  
Carretera Camajuaní km 5.5  
Santa Clara, Cuba 54830  
rbellop@uclv.edu.cu

**Rafael Falcón**

School of Information Tech. & Eng.  
800 King Edward Ave, U of Ottawa  
Ottawa, ON K1N 6N5, Canada  
rfalcon@uottawa.ca

**Rajen Bhatt**

Samsung India Software Center  
Noida-201305, Uttar Pradesh, India  
rajen.bhatt@gmail.com

**Sarina Sulaiman**

Soft Computing Research Group  
Faculty of Comp. Sci. & Inf. System  
Universiti Teknologi Malaysia  
Johor, Malaysia  
sarina@utm.my

**Sheela Ramanna**

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

**Simon Watson**

Dept. of Electronic & Electrical Eng.  
University of Loughborough  
Holywell Park, Leicestershire  
LE11 3TU, United Kingdom  
s.j.watson@lboro.ac.uk

**Siti Mariyam Shamsuddin**

Soft Computing Research Group  
Faculty of Comp. Sci. & Inf. System  
Universiti Teknologi Malaysia  
Johor, Malaysia  
mariyam@utm.my

**Yailé Caballero**

University of Camagüey  
Carretera Circunvalación Norte km 5  
Camagüey, Cuba 74650  
yaile@inf.reduc.edu.cu

**Yanheng Li**  
School of Comp. Science and Eng.  
Dalian Maritime University  
116026 Dalian, China

**Yiyu Yao**  
Department of Computer Science  
University of Regina

Regina, Saskatchewan  
S4S 0A2, Canada  
yyao@cs.uregina.ca

**Yudel Gómez**  
Central Univ. of Las Villas (UCLV)  
Carretera Camajuaní km 5.5  
Santa Clara, Cuba 54830  
ygomezd@uclv.edu.cu

