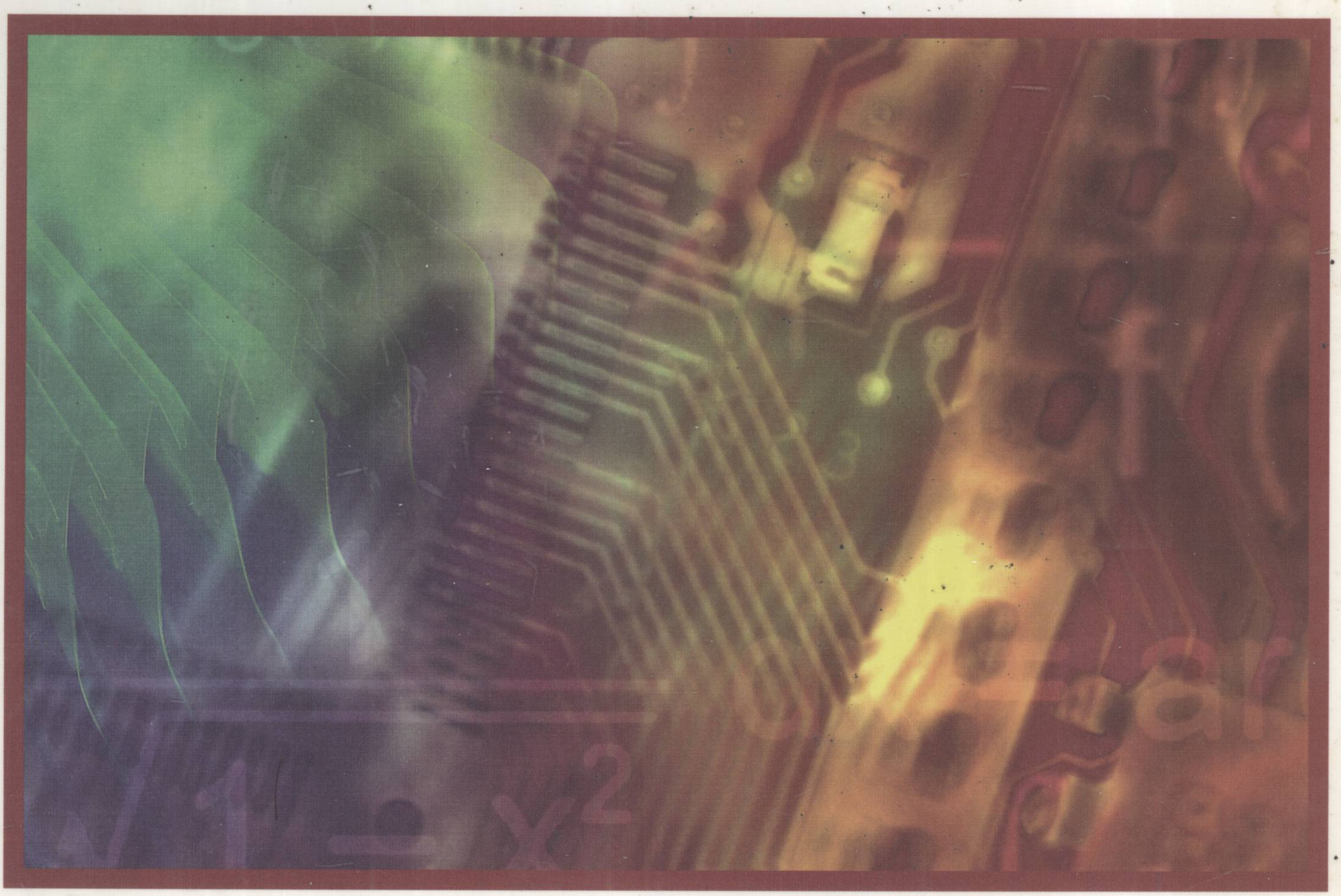
# ROUGH COMPUTING

## *Theories, Technologies, and Applications*



Aboul Ella Hassanien, Zbigniew Suraj,
Dominik Ślęzak, & Pawan Lingras

# Rough Computing: Theories, Technologies, and Applications

Aboul Ella Hassanien
*Kuwait University, Kuwait*

Zbigniew Suraj
*University of Information Technology and Management, Poland*

Dominik Ślęzak
*University of Regina, Canada*

Pawan Lingras
*Saint Mary's University, Canada*



Information Science REFERENCE

INFORMATION SCIENCE REFERENCE
Hershey · New York

| | |
|---|---|
| Acquisitions Editor: | Kristin Klinger |
| Development Editor: | Kristin Roth |
| Senior Managing Editor: | Jennifer Neidig |
| Managing Editor: | Sara Reed |
| Copy Editor: | Angela Thor |
| Typesetter: | Jamie Snavely |
| Cover Design: | Lisa Tosheff |
| Printed at: | Yurchak Printing Inc. |

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue, Suite 200
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-pub.com
Web site: http://www.igi-global.com/reference

and in the United Kingdom by
Information Science Reference (an imprint of IGI Global)
3 Henrietta Street
Covent Garden
London WC2E 8LU
Tel: 44 20 7240 0856
Fax: 44 20 7379 0609
Web site: http://www.eurospanonline.com



Library of Congress Cataloging-in-Publication Data

Rough computing : theories, technologies and applications / Aboul Ella Hassanien ... [et al.], eds.

p. cm.

Summary: "This book offers the most comprehensive coverage of key rough computing research, surveying a full range of topics from granular computing to pansystems theory. With its unique coverage of the defining issues of the field, it provides libraries with a single, authoritative reference to this highly advanced technological topic"--Provided by publisher.

Includes bibliographical references and index.

ISBN-13: 978-1-59904-552-8 (hardcover)

ISBN-13: 978-1-59904-554-2 (ebook)

1. Soft computing. 2. Rough sets. I. Hassanien, Aboul Ella.

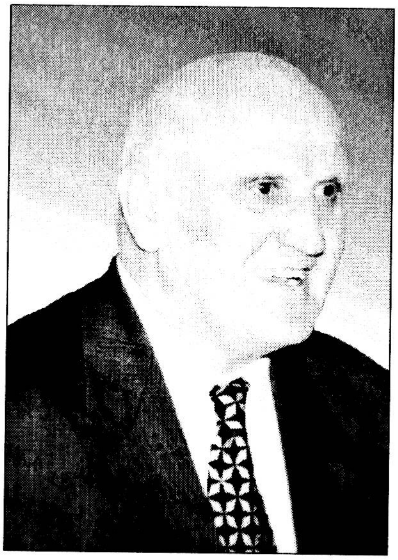
QA76.9.S63R66 2008

006.3--dc22

2007016855

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

# In Memoriam



*This book is dedicated to Professor Zdzisław Pawlak, a father of rough sets, who passed away on April 7, 2006 in Warsaw, Poland.*

# Preface

Since introduction of rough set methodology by Zdzisław Pawlak in the early eighties, we have witnessed its great advances in both theory and applications. There is a growing research interest in foundations of rough sets, with some relationships to other already established methodologies. As a result, rough sets are linked with decision support and intelligent systems, soft and granular computing, data mining and KDD, as well as pattern recognition and machine learning. A wide range of applications of rough sets, alone or combined with other techniques, have been proposed in bioinformatics and medicine, business and finances, environmental and social studies, multimedia data mining and processing, and many others.

The objective of this book is to provide a representative and authoritative source of current research results and methodology in the area of rough set methods and applications. The book consists of 12 chapters organized in three section. Section I (four chapters) presents the foundations of rough sets, including comparison with other methodologies and discussion on future perspectives. This part will be useful for the beginners and students who are interested in carrying out his/her research projects on the rough sets.

Section II (four chapters) continues with current trends of extending, combining, as well as applying rough set techniques. Section III (four chapters) discusses hybrid intelligent systems involving the elements of rough set techniques, as well as illustrates the place for rough sets in real-life applications. These two parts will be useful for the professionals who are interested in catching a new idea or using the book as a reference work.

**Chapter I**, by **Piotr Wasilewski** and **Dominik Ślęzak**, presents the algebraic foundations behind the rough set theory. Algebraic structures arise from various types of data-based knowledge and information. The authors examine those structures with respect to satisfying or violating different versions of the excluded middle principle. Besides some novel results, the chapter may be also treated as a survey on the rough-set-based methodologies of representing and analyzing the information systems.

**Chapter II**, by **Hung Son Nguyen**, presents the approximate Boolean reasoning approach to problem solving. It is based on the general framework for the concept approximation. It combines classical Boolean reasoning with modern methods in machine learning and data mining. The author shows advantages of approximate Boolean reasoning using examples of some representative challenges of KDD, emphasizing ability of balancing between the quality of the designed solution and its computational cost.

**Chapter III**, by **Richard Jensen** and **Qiang Shen**, introduces the fundamental ideas behind rough-set-based approaches and reviews the related feature selection methods. The authors discuss extensions to the traditional rough set approach, including recent feature selection methods based on tolerance rough sets, variable precision rough sets and fuzzy rough sets. The chapter reports also the latest developments in the search methods supporting the rough-set-based feature selection search methods, including hill climbing, genetic algorithms, and ant colony optimization.

**Chapter IV**, by **Yiyu Yao** and **Yaohua Chen**, reviews the existing studies on the comparisons and combinations of rough set analysis and formal concept analysis. Such unified framework provides a better understanding of the current data analysis challenges. The authors also report some new results important for both the theoretical foundations and applications.

**Chapter V**, by **Theresa Beaubouef** and **Frederick E. Petry**, discusses how rough sets can enhance databases by allowing for management of uncertainty. The authors discuss the rough relational database and rough object-oriented database models, as well as their fuzzy and intuitionistic extensions. Benefits of those various methods are discussed, illustrating usefulness and versatility of rough sets for the database extensions.

**Chapter VI**, by **Cory J. Butz** and **Wen Yan**, reviews a recently developed framework for reasoning from data, called the rough set flow graphs (RSFG). The authors examine two methods for conducting inference in RSFG. They further show how the order of variable elimination affects the amount of computation. The culminating result is the incorporation of an algorithm for obtaining a good ordering into the RSFG inference.

**Chapter VII**, by **Annibal Parracho Sant'Anna**, presents a new index of quality of approximation. It measures the mutual information between the relations respectively determined by conditional and decision attributes. It is based on comparison of two graphs, each one representing a set of attributes. Applications in the context of indiscernibility and dominance relations are considered. Combination with the idea of transformation into probabilities of classes being a preferred option is explored as well. The algorithmic procedure to select the most important attributes is outlined.

**Chapter VIII**, by **Zbigniew W. Ras** and **Elzbieta M. Wyrzykowska**, focuses on a novel strategy of construction of action rules, directly from single classification rules instead of the pairs of classification rules. This way, not only the simplicity of the algorithm, but also its reduced time complexity is achieved. The chapter also presents a modified tree-based strategy for constructing action rules, with comparative analysis included.

**Chapter IX**, by **James F. Peters**, **Maciej Borkowski**, **Christopher Henry**, and **Dan Lockery**, discusses the monocular vision system that learns to control the pan and tilt operations of a digital camera that tracks a moving target. The authors consider various forms of the actor critic learning methods. Rough set approximation spaces are applied to handle degrees of overlapping of the behavior patterns. The conventional actor critic methods are successfully extended with the built-in run-and-twiddle rough-set-based control strategy mechanism.

**Chapter X**, by **Tomasz G. Smolinski**, **Astrid A. Prinz**, and **Jacek M. Zurada**, proposes hybridization of rough set theory and multiobjective evolutionary algorithms to perform the task of signal decomposition in the light of the underlying classification problem itself. The authors present results for several variants of implementation of the introduced methodology.

**Chapter XI**, by **Jerzy W. Grzymala-Busse**, **Zdzisław S. Hippe**, **Teresa Mroczek**, **Edward Roj**, and **Boleslaw Skowronski**, presents application of two rough set approaches to mining the data sets related to bed caking during the hop extraction process. The authors use direct rule induction by the MLEM2 algorithm; they also generate belief networks conversed into the rule sets within the BeliefSEEKER system. Statistics for the rule sets are presented. Six rule sets are also ranked by the expert. The overall results show that both approaches are of approximately the same quality.

**Chapter XII**, by **Krzysztof Pancerz** and **Zbigniew Suraj**, constitutes the continuation of research trend binding rough set theory with concurrency. Automatic methods of discovering concurrent system models from data tables are presented. Data tables are created on the basis of observations or specifications of the process behaviors in the modeled systems. The proposed methods are based on rough set theory and colored Petri nets.

Overall, this book provides a critical analysis of many important issues in rough sets and applications, including introductory material and research results, as well as hybrid intelligent systems involving the elements of rough set techniques.

# Acknowledgment

The editors would like to thank all people involved in the development and review process of the book, especially Professor Guoyin Wang, the Chairman of the Rough Set and Soft Computation Society, Chinese Association for Artificial Intelligence, Dean of the College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, and Professor Sheela Ramanna, Dept. of Applied Computer Science, University of Winnipeg. Most of the authors of chapters included in this book also served as referees for chapters written by other authors. Thanks go to all those who provided constructive and comprehensive reviews.

We would also like to thank all the authors for their contributions. Special thanks go to Kristin Roth and Meg Stocking at IGI Global for their great support. Special thanks also go to the rest of the publishing team at IGI Global, whose contributions throughout the whole process from inception of the initial idea to final publication have been invaluable. In particular, to Kristin Roth and Meg Stocking, who continuously prodded via e-mail for keeping the project on schedule, and to Jan Travers, whose enthusiasm motivated me to initially accept his invitation for taking on this project.

# Section I
# Foundations of Rough Sets

# Table of Contents

# Detailed Table of Contents



This chapter presents the algebraic foundations behind the rough-set theory. Algebraic structures arise from various types of data-based knowledge and information. The authors examine those structures with respect to satisfying or violating different versions of the excluded middle principle. Besides some novel results, the chapter may be also treated as a survey on the rough-set-based methodologies of representing and analyzing the information systems.



This chapter presents the approximate Boolean reasoning approach to problem solving. It is based on general framework for the concept approximation. It combines classical Boolean reasoning with modern methods in machine learning and data mining. The author shows advantages of approximate Boolean reasoning using examples of some representative challenges of KDD, emphasizing ability of balancing between the quality of the designed solution and its computational cost.



This chapter introduces the fundamental ideas behind rough-set-based approaches and reviews the related feature selection methods. The authors discuss extensions to the traditional rough set approach, including recent feature selection methods based on tolerance rough sets, variable precision rough sets, and fuzzy-rough sets. The chapter reports also the latest developments in the search methods supporting

# Chapter I
# Foundations of Rough Sets from Vagueness Perspective

**Piotr Wasilewski**
*Warsaw University, Poland*

**Dominik Ślęzak**
*Infobright Inc., Canada*

## ABSTRACT

*We present three types of knowledge that can be specified according to the rough set theory. Then, we present three corresponding types of algebraic structures appearing in the rough set theory. This leads to three following types of vagueness: crispness, classical vagueness, and a new concept of "intermediate" vagueness. We also propose two classifications of information systems and approximation spaces. Based on them, we differentiate between information and knowledge.*

## INTRODUCTION

Handling vagueness was one of the motivations for proposing the rough set theory (Pawlak, 2004, see also Pawlak, 1982, 1991, 2003). In this chapter, we present algebraic foundations of the rough set theory from that perspective. Vagueness is understood according to the *tertium non datur* principle from traditional logic; the contemporary version of this principle is called the law of excluded middle (Frege, 1903; Hempel, 1939; Pawlak, 2004; Russell, 1923). Concepts in the rough set theory are represented by subsets of a universe of discourse. Each concept, $X$, can be represented by two subsets: that consisting of examples of $X$ (i.e., objects, which can be certainly classified as belonging to $X$) called a positive region of $X$, and that consisting of objects that certainly do not belong to $X$, called a negative region of $X$. Positive region of a given concept is represented by its lower approximation. Negative region is represented as a complementation of an upper approximation of the set. A complementation of the union of these two sets is called a boundary. Thus, boundary consists of objects that cannot be certainly classified as belonging to a concept, or as not belonging. A concept is vague if its boundary is nonempty, otherwise it is crisp (Pawlak, 2004). Vague concepts are represented in the rough set theory as rough sets. Crisp concepts satisfy the *tertium non datur* principle, their boundaries are empty, there are no objects of third type or, in other words, the middles between their positive and negative regions are excluded. Rough sets (representing vague concepts) violate this principle. However, classical rough sets, based on equivalence relations, satisfy the weak law of excluded middle. This law states that a pseudocomplement of a set $X$ with a pseudocomplement of a pseudocomplement of a set $X$ is equal to whole space or, in more general version it states that a pseudocomplement of an element $x$ of a pseudocomplemented lattice $L$ with a pseudocomplement of a pseudocomplement of this element $x$ is equal to unit of the lattice $L$. This observation gives a reason for differentiating two types of vagueness: strong vagueness (or simply vagueness) and weak vagueness. Weakly vague concepts satisfy the weak law of excluded middle, while strongly vague concepts violate this law. Concepts of both types violate the law of excluded middle.

We show that an incompleteness of information is a source of these two types of vagueness.

Incomplete information, represented by deterministic incomplete information systems, can be analyzed by means of classical approximation spaces. Such spaces consist of a universe and an indiscernibility relation, which is always equivalence relation. Knowledge granules determined by these spaces are equivalence classes. Rough sets within classical approximation spaces violate the law of excluded middle, but they satisfy the weak law of excluded middle. Thus, such rough sets represent vague concepts in an "intermediate" sense.

Nondeterministic information systems gave a reason for introducing many information relations (Orłowska, 1993, 1997, 1998). Most of these relations are tolerance relations, which are not necessarily transitive, that is are not equivalence relations. Therefore, incomplete information, represented by nondeterministic information systems, can be analyzed by tolerance spaces that consist of a universe and a tolerance relation. Tolerance spaces admit various ways of granulations of the universe. We present a type of knowledge granulations of the tolerance spaces such that if a tolerance relation is not transitive, then rough sets determined by this granulation violate even the weak law of excluded middle. Thus, such rough sets represent vague concepts in the strong sense.

We propose a differentiation between information and knowledge. It follows from foundations of the rough set theory. Starting point is reflected by two suggested classifications of information systems and approximation spaces. We point out that properties of incompleteness and determinacy of information systems are independent and, particularly, that there are complete indeterministic information systems. Information, represented in both complete deterministic and complete indeterministic information systems, enables us to discern between two arbitrary objects from the universe. So, indiscernibility relations determined by information systems of these types are identity relations. Since, according to Pawlak (1982, 1991, 2004), knowledge is based on ability to discern between objects, then complete information systems are bases for constructing complete knowledge, which is represented by identity relations. So, knowledge has more general nature than information. Complete knowledge is necessarily exact; thus, there is no complete inexact knowledge. It follows that properties of completeness and exactness of knowledge are