

---

## 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 pre-fetching 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

Rough Set Theory: A True Landmark in Data Analysis

Abraham, A.; Falcón, R.; Bello, R. (Eds.)

2009, XVI, 324 p., Hardcover

ISBN: 978-3-540-89920-4