Decision-theoretic Rough Sets and Beyond

Zdzisław Pawlak introduced rough sets 30 years ago as a theory for data analysis and classification. Rough set theory has raised rapid attention in academia and for practitioners. Fundamental notions of rough sets are knowledge granulation induced by grouping objects into equivalence classes based on their descriptions, definability of concepts, and approximations of concepts by definable concepts. A concept is presented by a set of instances and is approximated by a pair of definable concepts called the lower and upper approximations, or equivalently, three pair-wise disjoint sets known as the positive, boundary and negative regions. Objects in the positive region surely belong to the set, objects in the negative region surely do not belong to the set, and objects in the boundary region possibly belong to the set. Since its introduction, several attempts have been made to generalize the classical qualitative rough sets and to connect it to the theories.

Decision-theoretic rough sets (DTRS), introduced by Yiyu Yao in 1990, are the results of integrating Bayesian decision theory and rough set theory. DTRS use conditional probability to measure the degree of overlap of two sets. By introducing a pair of thresholds, we can define three quantitative probabilistic regions. By applying Bayesian decision theory, we can systematically compute and interpret the required thresholds. In the past 20 years, DTRS has gained increasing attention from researchers across several fields. Areas of application include data mining, feature selection, information retrieval and web-based support systems.

A special issue on probabilistic rough sets in this journal (International Journal of Approximate Reasoning, Volume 49, Issue 2, 2008) contributed to a surge of research interest on probabilistic approaches to rough sets. As a result, two other new models, namely, game-theoretic rough sets and information-theoretic rough sets, have been developed. The present special issue is a sequel and provides a snap shot of most recent progress on DTRS and beyond. The 12 papers cover many aspects of rough set theory. They will serve as a basis for many further developments such as Bayesian approaches, game-theoretic approaches, three-way decisions, classification and clustering, granular computing and big data analytics.

The paper entitled "**An automatic method to determine the number of clusters using decision-theoretic rough set**" authored by Hong Yu, Zhanguo Liu, and Guoyin Wang and the paper entitled "**An Extension to Rough *c*-Means Clustering Based on Decision-Theoretic Rough Set Model**" authored by Fan Li, Mao Ye and Xudong Chen investigated the use of DTRS model in clustering algorithms. The first paper considers the problem of identifying the number of clusters using the DTRS approach. A new clustering validity evaluation function is proposed that is based on the risk calculated by loss functions and possibilities. The second paper extends the rough c-means clustering algorithm by utilizing the idea of DTRS in the step of data points assignment to clusters. The assignments to a cluster or clusters are based on the expected loss (risk) associated with this action. Experimental results validate the effectiveness of the approach.

The paper entitled "**An axiomatic characterization of probabilistic rough sets**" authored by Tong-Jun Li and Xiao-Ping Yang derived two sets of axioms using the probabilistic rough set approximation operators where each set consists of five axioms. The proposed approach may help in increasing an understandability of probabilistic rough sets and can potentially be used in comprehending other types of probabilistic rough approximations.

The papers entitled "**Analyzing uncertainties of probabilistic rough set regions with game-theoretic rough sets**" authored by Nouman Azam and JingTao Yao and the paper entitled "**On an optimization representation of decision-theoretic rough set model**" Xiuyi Jia, Zhenmin Tang, Wenhe Liao, and Lin Shang addressed a fundamental issue in probabilistic rough sets related to the determination and interpretation of probabilistic thresholds. The two articles provided approaches that can obtain the required threshold parameters directly from the data. The former considers the game-theoretic rough set model to configure the threshold levels by considering the uncertainty involved in the three rough set regions. A repetitive game was formulated where the players cooperatively determine suitable threshold levels. The latter constructs an optimization problem of minimizing the decision costs of the DTRS model. Based on the suggested minimization problem, the thresholds and cost functions can be automatically obtained from the data and minimum cost attribute reduction can be defined.

The paper entitled "**Feature selection with test cost constraint**" authored by Fan Min, Qinghua Hu, and William Zhu considered the selection of important features as constraint satisfaction problem under a test cost constraint. A heuristic algorithm is proposed that is based on backtracking that selects important features by considering the given constraint. The approach may be useful for medium and small sized datasets.

The paper entitled "**Generalized probabilistic approximations of incomplete data**" authored by Jerzy Grzymala-Busse, Patrick Clark and Martin Kuehnhausen introduced two generalizations of probabilistic approximations i.e. global and local approximations. The global approximations are intersections of attribute-value blocks defined by the sets of all objects with the same attribute value. The local approximations are composed of unions of intersections of attribute-value blocks for distinct attributes. The approaches were compared on different datasets.

The paper entitled "**Incorporating logistic regression to decision-theoretic rough sets for classifications**" authored by Dun Liu, Tianrui Li, and Decui Liang introduced a new classification approach based on logistic regression and the DTRS model. The DTRS model was used to obtain three-way decisions that help in reducing the number of misclassifications which is sometimes considered as a major drawback of logistic regression based classification. The suggested approach uses the DTRS model to determine the probabilistic thresholds while logistic regression is used to determine the conditional probabilities used for three-way decision making.

An extension of DTRS model to multi-class was considered in the paper entitled "**Multi-class decision-theoretic rough sets"** authored by Bing Zhou. The presented approach suggested different costs associated with different types of classification errors. A pair of

threshold is determined using the DTRS approach for each class that defines the three probabilistic rough set regions and the implied three-way decisions.

The paper entitled "**Multigranulation decision-theoretic rough sets**" authored by Yuhua Qian, Hu Zhang, Yanli Sang, and Jiye Liang extended the DTRS model from the viewpoint of multigranulation, where the rough sets based lower and upper approximations are induced by granular structures through multiple binary relations. It is argued that multigranulation rough sets may be more useful than single granulation rough sets that include existing rough set models which describe a concept using approximations under a single relation.

The paper entitled "**Qualitative and quantitative combinations of crisp and rough clustering schemes using dominance relations**" authored by Pawan Lingras, Min Chen and Duoqian Miao proposed a framework with strict dominance and indifference relations to capture and axiomatize qualitative information in a crisp and rough clustering scheme. The proposed framework may also used for qualitative combination of clustering schemes based on different attributes and studying its relationship with quantitative combinations. The theoretical development in this study is applied to a synthetic retail dataset to illustrate its applications in data mining.

The paper entitled "**Rule Acquisition and Complexity Reduction in Formal Decision Contexts**" authored by Ming-Wen Shao, Yee Leung, and Wei-Zhi Wu proposed a general approach for rule acquisition and knowledge reduction in formal decision contexts. An attribute reduction approach was proposed for this purpose that deletes non-essential attributes without loss of the maximum rules. The examples suggest the complexity of concept lattice can be reduced by the suggested approach.

This special issue is a result of hard work of both authors and reviewers. The critical and constructive comments from reviewers in several rounds of reviews greatly improved final version all. We thank the Editor-in-Chief, Dr. Thierry Denoeux, of the International Journal of Approximate Reasoning for his support and encouragement.

September 2013

JingTao Yao
University of Regina, Canada

Huaxiong Li,
Nanjing University, China

Georg Peters
Munich University of Applied Sciences, Germany
Australian Catholic University, Australia