Available at
www.ComputerScienceWeb.com
POWERED BY SCIENCE @ DIRECT®

Pattern Recognition Letters 24 (2003) 829–831

Pattern Recognition Letters

www.elsevier.com/locate/patrec

Preface

The theory of rough sets, introduced by Pawlak in 1982, has recently emerged as a major mathematical approach for managing uncertainty that arises from inexact, noisy or incomplete information. It is turning out to be methodologically significant to the domains of artificial intelligence and cognitive sciences, especially in the representation of and reasoning with vague and or imprecise knowledge, data classification analysis, rule generation, machine learning, data mining and knowledge discovery. The theory is also proving to be of substantial importance in many other areas of applications.

The focus of rough set theory is on the ambiguity caused by limited discernibility of objects in the domain of discourse. The idea is to approximate any concept (a crisp subset of the domain) by a pair of exact sets, called the lower and upper approximations, in the granular universe. For the past few years, rough set theory and granular computation has proven to be another soft computing tool which, in various synergistic combinations with fuzzy logic, artificial neural networks and genetic algorithms, provides a stronger framework to achieve tractability, robustness, low cost solution and close resembles with human like decision making. For example, rough-fuzzy integration forms the basis of the computational theory of perceptions (CTP), recently explained by Zadeh, where perceptions are considered to have fuzzy boundaries and granular attribute values. Similarly, rough neural synergistic integration helps in extracting crude domain knowledge in the form of rules for describing different concepts/classes, and then encoding them as network parameters; thereby constituting the initial knowledge base network for efficient learning. A key point of granular computation is that, here computations/operations are performed on granules (clump of similar objects or points), rather than on the individual data points. As a result, the computation time is greatly reduced. At present the results on these investigations, both theory and real life applications, are being available in different journals and conference proceedings mainly in the fields of computer science, information technology, engineering and mathematics.

The objective of this special issue is to present to the pattern recognition community a cross sectional view of the present status of the said research, considering rough set theory both individually and in combination, with applications to different facets of pattern recognition, in particular to data mining tasks. This has nine papers addressing various tools, methodologies and applications. These are authored by experts from different active groups in USA, Canada, Japan, Poland and India, and each of them is reviewed by two to four referees.

The issue starts with the article "Rough Sets Methods in Feature Selection and Recognition" by two pioneers in rough set theory, R.W. Swiniarski and A. Skowron. The significance of the basic constructs of rough set theory, namely reducts and their applications, including dynamic reducts, in feature selection has been emphasized. The effectiveness of the features is demonstrated using a neural classifier for face and mammogram recognition problems. Besides these, the article provides some basic definitions and preliminaries of rough sets, which are useful to understand the remaining contributions.

In the second paper "Tree Structure for Efficient Data Mining Using Rough Sets" by

PII: S0167-8655(02)00195-2

Ananthanarayana, Narasimha Murthy and Subramanian, a novel data structure, called pattern count tree (PC-Tree), for data mining is developed which scans the data base only once to generate the abstractions. It provides a compact and complete representation of the training set. Flexibility offered by rough sets is then exploited to develop a rough PC-Tree which reduces the processing time and memory requirements without affecting the classification accuracy.

The next two articles deal with the symbiotic integration of rough sets with other tools for efficient decision-making. For example, Mitra, Pal and Siddiqi describe in their article "Nonconvex Clustering using Expectation Maximization Algorithm with Rough Set Initialization" an integration of a minimal spanning tree (MST) based graph-theoretic technique and expectation maximization (EM) algorithm with rough sets for non-convex clustering. EM provides statistical modeling of the data and handles the associated uncertainties. Rough set theory is used for generating logical rules to obtain initial approximate mixture model parameters and thus helps in faster convergence and avoidance of the local minima problem of EM; thereby enhancing its performance. While extracting rough set rules, it uses fuzzy granulation of the feature space to model overlapping clusters. MST helps in determining non-convex clusters. Since the algorithm works on Gaussians rather than the original data points, time requirement is very low.

Similarly, a decision support tool based on rough fuzzy computing and case based reasoning is developed by Rao and Sarma in "A Rough Fuzzy Approach for Retrieval of Candidate Components for Software Reuse" for retrieval and selection of reusable software components from a repository. Rough-fuzzy sets are used to model the indiscernibility relation in input patterns as well as the vagueness in the output decisions. While the method reduces the search domain and hence requires less retrieval time, rough fuzzy integration helps in finding suitable components for reuse, particularly when the exact matches are not available.

The significance of discretization of continuous valued attributes, which is a major concern of rough set theoretic computation, is demonstrated in the next two papers. In "Meningitis Data Mining by Cooperatively using GDT-RS and RSBR", Zhong, Dong and Ohsuga have demonstrated an application of generalized distribution table and rough sets (GOT-RS), and rough sets with Boolean reasoning (RSBR) for mining if-then rules in a meningitis data. Here RSBR is used for discretization of real valued attributes as a preprocessing step before GDT-RS, a soft hybrid induction system, is applied for discovery of rules. The quality of the rules, as expected, is found to be affected by the application of RSBR.

While this discusses crisp discretization, the next article by Roy and Pal titled "Fuzzy Discretization of Feature Space for a Rough Set Classifier" introduces a new methodology for generating fuzzy discretization of feature space to be applicable for a rough set theoretic classifier. Fuzzy discretization has three components viz., membership value, group member and affinity value, unlike crisp discretization which is characterized only by the group member. Its role in handling inconsistent decision tables arising from overlapping pattern classes has been demonstrated for both rough set theoretic and multilayer perceptron (MLP) based classifiers.

The remaining three contributions describe different real life applications. In the paper "Increasing Sensitivity of Pattern Birth by Changing Rule Strengths" of Grzymala-Busse, Goodwin and Zhang, an interesting problem of increasing the accuracy of preterm birth by changing the rule strength for the category is studied. The criteria chosen for determining the optimal rule strength include the largest difference between the true-positive and false-positive properties, and the maximum profit. A rule induction system (which learns from examples based on rough sets) is used for classification based on rule strengths.

Classification of storm cells is a challenging problem because it involves complex evolution of storm cells besides high dimensionality and impreciseness/incompleteness in the weather data. A rough set approach is used by Peters, Suraj, Shan, Ramanna, Pedrycz and Pizzi in their article

"Classification of Meteorological Volumetric Radar Data", to detect storm events responsible for the severe summer weather.

In the last article "Automatic Identification of Sound Source Position Employing Neural Networks and Rough Sets" by Czyzewski, the problem of separating an acoustic signal in noisy environment has been addressed. Here a decision system based on rough sets is used for processing parameters representing acoustic signals so that their incoming directions can be identified.

Finally, we take this opportunity to thank Pattern Recognition Letters for giving us an opportunity to act as the Guest Editors for this special issue. We believe, the issue is very timely. We are thankful to all the contributors and reviewers for their co-operation in making this special issue a reality and in bringing it out in time. The support and co-operation from the Department of Science and Technology (DST), India and the Polish State Committee for Scientific Research (KBN), Poland under an INDO-POLISH collaborative project is gratefully acknowledged. Thanks of S.K. Pal are also due to Mr. Indranil Dutta and Mr. Sanjoy K. Das for rendering sincere secretarial assistance.

Guest Editors
S.K. Pal
A. Skowron