# Guest Editorial: Special Issue on Advances in Multiobjective Evolutionary Algorithms for Data Mining

In the present era of automation, a huge amount of data is routinely collected in almost all spheres of life. Manual sifting of this vast repository of data in order to extract meaningful information is no longer feasible. Data mining, the process of discovering novel, interesting, previously unknown and potentially useful patterns from large data sets [1], has evolved as an important area of research in response to this challenge.

The aim of any data-mining technique is to build an efficient predictive or descriptive model of a large amount of data that not only best fits or explains it, but is also able to generalize to new data. Depending on the application, the underlying data set can be simple like numbers and characters, more complicated like trees and graphs, or might even be unstructured like text. Data-mining tasks can broadly be classified into two categories: predictive or supervised and descriptive or unsupervised. The predictive techniques learn a model from the current data in order to make predictions about the behavior of new datasets. On the other hand, the descriptive techniques provide a summary of the available data. Association rule mining, classification, clustering, regression, sequence and link analysis, and dependency modeling are some of the most common data-mining tasks.

Optimization of the model parameters is a requirement for the successful application of any data-mining technique. Often such problems do not lend themselves to closed-form, analytic solutions, and cannot be solved using standard mathematical techniques. Moreover, due to the large size of the input data, the problems become intractable; thus, designing efficient deterministic algorithms is not feasible in most cases. Applications of evolutionary algorithms, with their inherent parallel architecture, have been found to be particularly useful for automatic processing of large quantities of noisy raw data for optimal parameter setting and to discover significant and meaningful information.

Evolutionary algorithms belong to a class of population-based search and optimization techniques that are inspired by the principles of natural evolution. There are different techniques within this broad class of search methods. All the approaches, while being different in the way they sample the search space, essentially encompass an encoding of the set of parameters, have a set of operators to generate new solutions, and a rule for accepting/rejecting the new solutions [2]. Traditionally, evolutionary algorithms have been used to solve problems with an aim to optimize only a single criterion. However, many real-life problems involve multiple conflicting measures of performance, or objectives, which need to be optimized simultaneously. Optimum performance in one objective often implies unacceptably low performance in one or more of the other objectives, creating the need for a compromise to be reached. This facet of multiobjective optimization (MOO) is highly visible in data mining. For example, in association rule mining, a rule may be evaluated in terms of both its support and confidence, a clustering solution may be evaluated in terms of several conflicting measures of cluster validity, while in finance one may wish to form a portfolio exhibiting both high returns and low risk. Such problems have an inherent multiobjective nature, the goal being to simultaneously optimize all the conflicting objectives. By their very nature MOO problems do not admit a single solution that provides optimum values for all the objectives. A set of trade-off solutions, called the Pareto optimal set, is obtained as the result of any MOO technique. Given any Pareto optimal solution, it is not possible to improve it on any one objective without degrading it in at least one other. Due to their population-based nature, evolutionary algorithms have been found to be particularly suitable for solving MOO problems and, consequently, a large number of multiobjective evolutionary algorithms (MOEAs) have been developed in the last few years [3].

This special issue focuses on the novel and advanced designs of MOEAs for data mining and state-of-the-art reviews as well as cutting edge applications for solving problems in diverse domains. A comprehensive survey of MOEAs and their applications to data mining is provided in the two articles by Mukhopadhyay *et al*. The first article provides an introduction to the basic concepts of multiobjective optimization followed by a brief review of the MOEAs most commonly used in data mining. Common data-mining tasks are also discussed, followed by a justification of the use of MOEAs in data mining. The paper then provides an in-depth review of the use of MOEAs in two data-mining tasks, such as feature selection and classification. Applications of MOEAs in clustering, biclustering, association rule mining, and several other data-mining tasks are surveyed in Part II. That paper concludes with some possible paths for future research.

The article by Garcia-Piquer *et al*. is in the broad area of multiobjective evolutionary clustering. Here, the authors conduct an in-depth evaluation of different cluster representation strategies for encoding the solutions as well as the corresponding genetic operators. In particular, they investigate three representations: prototype-based, label-based, and graph-based. Large-scale experiments are carried out on several synthetic data sets and data sets from the UCI Repository using an approach called the clustering algorithm based on

Date of current version January 27, 2014.
Digital Object Identifier 10.1109/TEVC.2013.2292181

a multiobjective strategies platform. The multiobjective clustering techniques are also compared to some conventional clustering algorithms, namely *k*-means, expectation maximization, *x*-means, and self-organizing maps. The primary motivation of this paper is to identify the most robust algorithm that can deal with data sets of varying complexities. The results show that, in general, the performance of multiobjective clustering methods is better than those obtained by the single objective algorithms analyzed, especially for complex problems in which it is difficult to identify clusters according to a single criterion. The prototype-based representation appears to be better and more scalable than the other two forms of representations. The authors note that this may be due to the fact that the length of the individuals is independent of the number of instances. The representation also seems to explore the search space better while being more robust to scattered data sets. Garcia-Piquer *et al*. conclude the article with a mention of possible future research directions.

The problem of mining quantitative association rules that focus both on positive as well as on negative dependencies is tackled in the article by Martín *et al*., which is entitled "A New Multi-objective Evolutionary Algorithm for Mining a Reduced Set of Interesting Positive and Negative Quantitative Association Rules." Incorporating negative dependencies is important since it indicates which items do not appear together. From this perspective, the authors propose a new multiobjective evolutionary algorithm that is based on decomposition, in order to mine a reduced set of positive and negative quantitative association rules with a low computational cost. For this purpose, they suggest an evolutionary learning of intervals of the attributes and a condition selection for each rule, while introducing an external population and a restarting process to store all the non-dominated rules found and to improve the diversity of the rule set obtained. Three objectives are simultaneously optimized; comprehensibility, interestingness, and performance. The resulting rules are interesting, easy to understand, and have a good coverage of the dataset. The effectiveness of the proposed technique is validated over nine real-world datasets, with the number of variables ranging from 4 to 91 and the number of examples ranging from 40 to 22 784. The results indicate that the proposed method can mine rule sets with a good trade-off among the number of rules, support, and coverage. Since the evolved rules have fewer attributes, interpretability is also easier.

An interesting application of multiobjective evolutionary algorithms is presented by Zheng *et al*. in their article "Population Classification in Fire Evacuation: A Multiobjective Particle Swarm Optimization Approach." Here, they have addressed the issue of mining rules for classifying the evacuee population in case of fire. This is important for making evacuation decisions and saving lives. In this regard, they propose a multiobjective particle swarm optimization method that simultaneously optimizes the precision and recall measures of the classification rules. For this, an effective approach has been designed for encoding classification rules and the subsequent use of a comprehensive learning strategy for evolving particles and maintaining diversity of the swarm. Comparative results show that the proposed method performs better than some state-of-the-art methods for classification rule mining for a real-world fire evacuation data set. The paper also reports on a successful application of the proposed technique in a real-world fire evacuation operation that happened recently in China.

Gong *et al*. in "Complex Network Clustering By Multiobjective Discrete Particle Swarm Optimization Based on Decomposition" propose a multiobjective discrete particle swarm optimization algorithm to solve network clustering problems. Based on the network topology, the particle's velocity and position vectors in discrete form are defined. All the arithmetic operators between velocity and position vectors are redefined. Moreover, a problem-specific swarm initialization technique and a turbulence operator are introduced. The performance of the proposed algorithm has been validated on signed as well as on unsigned networks, and its effectiveness has been established through an extensive comparative study with ten state-of-the-art approaches.

The six papers included in this special issue provide a snapshot of the current research trends in multiobjective evolutionary algorithms for data mining. The main issues and challenges in this domain have been highlighted and directions of future research work have also been provided.

## REFERENCES

[1] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "The KDD process for extracting useful knowledge from volumes of data," *Commun. ACM*, vol. 39, no. 11, pp. 27–34, 1996.

[2] J. H. Holland, *Adaptation in Natural and Artificial Systems*. Ann Arbor, MI, USA: Univ. Michigan Press, 1975.

[3] C. A. Coello Coello, G. B. Lamont, and D. A. Van Veldhuizen, *Evolutionary Algorithms for Solving Multi-Objective Problems*. Berlin, Germany: Springer, 2007.

Sanghamitra Bandyopadhyay, *Guest Editor*
Machine Intelligence Unit
Indian Statistical Institute
Kolkata 700108, India
(e-mail: sanghami@isical.ac.in)

Ujjwal Maulik, *Guest Editor*
Department of Computer Science and Engineering
Jadavpur University
Kolkata 700032, India
(e-mail: umaulik@cse.jdvu.ac.in)

Carlos Artemio Coello Coello, *Guest Editor*
CINVESTAV-IPN,
Depto. de Computación,
Colonia San Pedro Zacatenco,
México 07360, Mexico
(e-mail: ccoello@cs.cinvestav.mx)

Witold Pedrycz, *Guest Editor*
Department of Electrical and Computer Engineering
University of Alberta,
Edmonton, AB T6G 2R3, Canada
(e-mail: wpedrycz@ualberta.ca)

**Sanghamitra Bandyopadhyay** (SM'05) received the B.Tech., M.Tech., and Ph.D. degrees in computer science from University of Calcutta, Calcutta, India, Indian Institute of Technology Kharagpur (IITKgp), Kharagpur, India, and Indian Statistical Institute (ISI), Kolkata, India, respectively.

She is currently a Professor at the ISI. She has authored five books and more than 250 technical articles. Her research interests include bioinformatics, soft and evolutionary computation, pattern recognition, and data mining.

Dr. Bandyopadhyay received the Dr. Shanker Dayal Sharma Gold Medal and also the Institute Silver Medal from IITKgp in 1994, the Young Scientist Awards of the Indian National Science Academy, in 2000, the Indian Science Congress Association in 2000, the Young Engineer Award of the Indian National Academy of Engineering in 2002, the Swarnajayanti Fellowship from the Department of Science and Technology in 2007, the prestigious Shanti Swarup Bhatnagar Prize in Engineering Science in 2010, and the Humboldt Fellowship from Germany in 2009. She has been selected as a Senior Associate of ICTP, Italy, in 2013. She is a fellow of the National Academy of Sciences, Allahabad, India, and the Indian National Academy of Engineering.

**Ujjwal Maulik** (SM'05) received the bachelor's degree in both physics and computer science, in 1986 and 1989, respectively, and the master's and Ph.D. degrees in computer science, in 1991 and 1997, respectively.

He is currently a Professor in the Department of Computer Science and Engineering, Jadavpur University, Kolkata, India. He has co-authored seven books and about 250 research publications. His research interests include computational intelligence, bioinformatics, combinatorial optimization, pattern recognition, and data mining. He received the Alexander Von Humboldt Fellowship for Experienced Researchers, Germany. He is a Fellow of the Indian National Academy of Engineering.

**Carlos Artemio Coello Coello** (F'11) received the B.Sc. degree in civil engineering from Universidad Autónoma de Chiapas, Mexico, in 1991 and the M.Sc. and Ph.D. degrees in computer science from Tulane University, New Orleans, LA, USA, in 1993 and 1996, respectively.

He is currently a Professor with Distinction (CINVESTAV-3F Researcher) and Chair of the Computer Science Department of CINVESTAV-IPN, Mexico City, Mexico, USA. He has either authored or co-authored over 355 publications, which currently report over 19 400 citations according to Google Scholar (his h-index is 56). His research interests include evolutionary multiobjective optimization and constraint-handling techniques for evolutionary algorithms.

Dr. Coello Coello received the 2007 National Research Award from the Mexican Academy of Science in the area of exact sciences; the Medal to the Scientific Merit from Mexico City's congress; the Ciudad Capital: Heberto Castillo 2011 Award in Basic Science; the 2012 National Medal of Science and Arts in the area of physical, mathematical and natural sciences; and the 2013 IEEE Kiyo Tomiyasu Award. He is an Associate Editor of IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION and also serves on the editorial board of 14 other international journals. He is also a member of the ACM, Sigma Xi, and the Mexican Academy of Sciences.

**Witold Pedrycz** (F'88) is a Professor and Canada Research Chair (CRC—Computational Intelligence) in the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada. He is also with the Systems Research Institute of the Polish Academy of Sciences, Warsaw, Poland. In 2009, he was elected a Foreign Member of the Polish Academy of Sciences. He has published numerous papers in his areas of research and has authored 14 research monographs covering various aspects of computational intelligence and software engineering. His research interests involve computational intelligence, fuzzy modeling and granular computing, knowledge discovery and data mining, fuzzy control, pattern recognition, knowledge-based neural networks, relational computing, and software engineering.

Dr. Pedrycz is an Editor-in-Chief of *Information Sciences*, and the Editor-in-Chief of the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS PART A. He is also an Associate Editor of IEEE TRANSACTIONS ON FUZZY SYSTEMS and is a member of a number of editorial boards of other international journals. He received the prestigious Norbert Wiener award from the IEEE Systems, Man, and Cybernetics Council in 2007, the IEEE Canada Computer Engineering Medal in 2008, and the Cajastur Prize for Soft Computing from the European Centre for Soft Computing for Pioneering and Multifaceted Contributions to Granular Computing in 2009. He has also been a member of numerous program committees of IEEE conferences in the area of fuzzy sets and neurocomputing.