Intelligent Data Analysis 17 (2013) 365–366
DOI 10.3233/IDA-130583
IOS Press

# Editorial

Dear Colleague

Welcome to volume 17(3) of Intelligent Data Analysis Journal.

This issue of the IDA journal consists of nine articles which represent a variety of topics, all related to the applied and theoretical research in the field of Intelligent Data Analysis.

The first three articles of this issue are about semi-supervised learning methods. Imani and Keyvanpour in the first article of this issue discuss the difficulty in text categorization where a small number of examples are available and what role semi-supervised learning methods can play. They propose use of self-learning as a solution to this class of problems and apply dynamic weighting combined with majority voting approach to analyse un-labeled data. They used their proposed method on real world data sets where their experiments indicate performance improvements in classification. Choi *et al.* in the next article of this issue discuss the difficulty in validation of discovered models when a small number of labeled examples are available from which a subset has to be used for validation. They propose to use ensemble learning and graph sharpening to circumvent this problem. Their experimental results demonstrate the applicability of their method in real world situations where best parameter values are identified. Sugiyama and Yamamoto in the last article of this group propose a new approach for semi-supervised learning that is based on closed set lattices. They present a learning algorithm which performs as multi-class classifier and a label ranker for mixed data that contains both discrete and continuous variables. Their proposed algorithm uses both labeled and unlabeled data to construct a closed set lattice. Their experiments show the competitive performance of their algorithm in classification and ranking compared to other learning algorithms.

Chen *et al.* in the fourth article of this issue discuss the problem of skewed class distribution and costs associated with non-uniform misclassification. They investigate the effects of cost ratio, imbalance ratio and sample size on classification performance using some real-world data sets. They further demonstrate that cost ratio and level of class imbalance have strong effects on prediction performance and mostly recommend near balanced training data sets to be used. Albertini and de Mello discuss the main complexity in analyzing data streams where clustering is applied dynamically to understand and represent data behavior changes. They emphasize the limitations of the above process and propose an on-line and adaptive approach to monitor and react to behavior changes. Their proposed approach that is based on $k$-means is experimentally evaluated and behavior changes are verified by testing the isomorphism of Markov chains over time.

This issue also includes four reviews and applied research articles. Khatoon *et al.* in the sixth article emphasize the importance of mining software source code for identifying useful patterns and provide a survey of tools and techniques which are based on data mining approaches. Their article provides a comparison and evaluation of some current code mining tools and techniques and presents a concise overview of source code mining techniques. Their results show that existing techniques are primarily targeting bug detection and among some of the areas where they identify a need is tools to develop quality software by automatically detecting different bugs. Covões *et al.* in the next article compare a number of clustering algorithms that are suitable to work with constraints. Their studies presented in the article include using 20 data sets where for each of them they experiment over 800 constraints. In order to provide some confidence on the randomness of their results, they provide a number of statistical significance of their results. They also present the robustness and computational complexity of their algorithms in terms of handling noisy constraints and a number of new experimental findings. Mallick *et al.* report on their investigation of the need for incremental mining of sequential patterns and present an analytical study focusing on the characteristics of 20 incremental learning algorithms. Overall, they conclude that the best performance can be achieved in progressive databases. And finally, Majnik and Bosnić in the last article of this issue discuss extensive use of ROC (Receiver Operating Characteristics) curves in machine learning and present a survey of this field where they present a condense report containing important achievements. They present application areas of the ROC analysis in machine learning and describe a number of problems and challenges. They also provide a couple of examples to illustrate some of the existing applications.

In conclusion, with this issue of the IDA journal which is Volume 17(3), we are observing a steady increase in submission of manuscripts to our journal for evaluation and publication. We continue our efforts to select the highest quality papers. In addition, this year's Intelligent Data Analysis Symposium (IDA-2013) will be held in London, UK from October 17–19. The deadline for submission of papers is May 6, 2013. For more information please refer to http://sites.brunel.ac.uk/ida2013. We look forward to receiving your feedback along with more and more quality articles in both applied and theoretical research related to the field of IDA.

With our best wishes,
Dr. A. Famili
Editor-in-Chief