

Editorial

Dear Colleague:

Welcome to volume 15(5) of Intelligent Data Analysis Journal.

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

The first four articles are about all forms of classifications. Borchani *et al.* in the first article argue the complexity of mining concept drifting data streams and propose a new semi-supervised approach for handling these data sets. Their approach is based on monitoring three kinds of drifts which are feature, conditional and dual. The idea which is quite general, is to develop new classifiers when a drift occurs based on these parameters. Experimental studies reported in this paper are based on a naïve Bayes classifier and also logistic regression and it is from both synthetic and real data. In the second article of this issue, Delibasic *et al.* discuss the idea of white box data mining in which the algorithms and their use would be transparent to the users. The authors present a component based algorithm design platform that can be used in several forms such as single or joint influence. Two new heuristics in decision tree algorithm design are part of what is proposed in this paper and their evaluation is based on 80 component based decision tree algorithms on 15 benchmark data sets. Their results showed that one should search for optimal component interplay instead of looking for the optimal among predefined algorithms. Peng and Li in the third article of this issue argue that few attempts exist in subspace clustering on multi-way data and propose a new subspace clustering approach. They show that their proposed algorithm can simultaneously perform subspace identification through various data clustering approaches. Their experimental results on both synthetic and real world data sets demonstrate the effectiveness of their approach. And Pereira *et al.* in the next article propose a new attribute selection strategy that is based on lazy learning and essentially it postpones the identification of relevant attributes until a new instance is submitted for classification. Their strategy relies on the hypothesis that the attribute values of an instance to be classified may contribute to identifying the best attributes for the correct classification of that particular instance. Their experimental results on over 40 data sets show the effectiveness of their approach as part of which they propose a new metric to estimate when a specific data set can benefit from the lazy attribute selection approach.

The next two articles in this issue are on data preprocessing. Montañes *et al.* address a physical system parameter evaluation problem and propose a greedy algorithm that is based on polynomial stepwise regression. This method is to solve the problem of understanding the degree of being polynomial for a set of attributes in physical systems. Their approach is evaluated on some artificial data and some real data sets where they show their method successfully outperforms some learning methods in terms of effectiveness and efficiency. Malik and Unwin in the sixth article of this issue emphasize the importance of data quality and present a framework for automatically detecting unusual or erroneous data values in data sets. The main idea behind their approach is to generate association rules to identify the exceptional cases. Their experimental results show that their proposed framework is able to successfully identify erroneous values in large data sets.

The last three articles of this issue are mostly on applied research. Chen and Wang's article is about forecasting where they propose an approach that is a mix of a hybrid fuzzy linear regression and back

propagation network. In their evaluation some historical data on global CO₂ concentration are used to evaluate the effectiveness of their methodology where the results are quite impressive. Next article by Park *et al.* is about time series forecasting where the authors discuss the limitations of some artificial intelligence methods in reflecting the dynamic characteristics of financial time series. Instead, the authors suggest a new method that constructs the trend forecasting model of financial time series. This method is evaluated using some financial index data and its superiority is demonstrated in the article. In the last article of this issue, Makanju *et al.* address the problem of machine learning detectors for deployment of real life networks and propose a combination of genetic programming and artificial neural nets as part of their machine learning paradigm. Their evaluation results on a number of data sets show that a machine learning based detector can be significantly enhanced or limited by the representation of the training data.

In conclusion, we continue our efforts on organizing and publishing special issues of IDA journal, sometimes 1–3 per year. Currently, there are three special issues under development that are at different stages of completion. The IDA conference that is usually held every two years in Europe, will be held in Porto, Portugal from October 29–31, 2011. For details, important dates and other information please refer to the following web site (<http://www.liaad.up.pt/ida2011/>). 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