Editorial

Dear Colleague

Welcome to volume 17(4) 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 article of this issue by Peres and Olague is on genetic programing applied to object recognition. The authors propose a new approach for learning operators through genetic programming. Their approach that relies on synthesis of mathematical expressions extracts information derived from objects such as images. Their experimental results confirm the validity of their approach using a series of tests where their results are compared with a number of test beds. The second article of this issue by Feng *et al* is about mining high utility itemsets from data streams. Knowing that time and space are quite essential in analyzing this class of data, the authors propose a new data structure for maintaining utility information of transaction itemsets which is based on a sliding window. As a result, searching for patterns in high utility item sets can be done with an efficient scanning of the database. Their experimental results show that their algorithm has better performance in terms of stability and other performance parameters. Al Shaqsi and Wang in the third article of this issue discuss the topic of cluster estimation in unsupervised learning. They propose a new method for estimating the most probable number of clusters. Their approach is based on calculating the length of a constant similarity interval where they use the longest length as the most probable number of clusters. Their experimental results on three synthetic and eight real data sets and its comparison with some well-known methods show that their approach performs significantly better.

Cagliero and Fiori present a framework that entails the discovery of hidden and high-level correlations, in the form of generalized association rules, among the content and contextual features of social media contents. Unlike traditional data mining approaches, their proposed approach takes into account the evolution of extracted patterns across the sequence of the previous mining sessions. Their experiments performed on both real and synthetic data show the effectiveness and the efficiency of their approach. Xie et al. in the fifth article of this issue discuss challenges in multi-category classification problems and propose an approach called one-versus-all twin support vector machine classification for multi-category classification problems. The idea is to use a one-versus-all approach to construct a twin support vector machine classifier. The evaluation of their work reported in the article includes use of synthetic and real data sets where both theoretical and experimental results demonstrate that their approach outperforms traditional classifiers.

Lee and Giraud-Carrier in the sixth article of this issue address the key topic of automatically choosing learning algorithms for data mining applications. They propose a two-step approach in which one first clusters learning algorithms based on behaviour similarity and then redefines meta-learning process as mapping classification tasks to clusters of behaviorally-similar algorithms. Their experiments on a wide variety of classification tasks reported in the paper demonstrate better results than typical selection models. The next article by Rui and Yang propose a smooth support vector learning algorithm suitable for fuzzy rule-based classification systems. The proposed algorithm is capable of generating non-linear separating surfaces using arbitrary kernels where a fuzzy classifier using arbitrary reference functions can

be built from training samples. The performance of their proposed algorithm using different reference function is shown in their experimental results.

Finally, the last two articles of this issue are on applied research. Hatami and Chira discuss the challenging task of feature selection in analyzing microarray gene expression data and propose a random subspace ensemble-based method targeting microarray data in life sciences. Their approach is based on using multiple classifiers that use randomly selected features from a given training data. Experimental results of their study show the use of their approach for analyzing some publicly available gene expression data related to cancer where the proposed method is competitive with related models from the literature. The last article if this issue, by Shin and Park is also about feature identification. The authors discuss the issue in the domain of search engines where they emphasize that the existing relevance assessment approaches applied in web page retrieval sometimes produce unsatisfactory results due to changes in their input. The authors therefore suggest to identify significant features from input data elements with the aim of improving the effectiveness of input retrieval while at the same time reducing computational costs. The author report a set of experiments conducted using real-world data sets where they produced interesting results.

In conclusion, with this issue of the IDA journal which is Volume 17(4), we are glad to report continuous 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. 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