Editorial

Dear Colleague:

Welcome to volume 26(3) of the Intelligent Data Analysis (IDA) Journal.

This issue of the IDA journal is the third issue for our 26th year of publication. It contains twelve articles representing a wide range of topics related to the theoretical and applied research in the field of Intelligent Data Analysis.

The first group of articles are about advanced learning: unsupervised and supervised methods in IDA. Li et al in the first article present an extended clustering algorithm that is based on cluster shape distribution of sample distribution. Their algorithm that automatically determines the number of clusters and classification discrimination boundaries by finding the boundary closures of the clusters from a global perspective of the sample distribution. In their approach, the cluster labels of the boundary are propagated to the entire sample set by a nearest neighbor search. The proposed method is evaluated on multiple benchmark datasets where their results show that the proposed method achieves highly accurate and robust clustering results. Pan et al in the next article of this group present a resistance outlier sampling algorithm for imbalanced data prediction. The authors argue that traditional sampling algorithms are susceptible to outliers and introduce the isolation Forest algorithm to overcome its vulnerability to this problem. They introduce a calculation method for anomaly index which can delete outliers accurately of minority data. Their experimental results from a number of public imbalanced datasets show that the algorithm can effectively improve the accuracy of the minority class and increase the stability. The third article of this issue by Zhao et al argue that learning from multi-class imbalanced data sets is still an open problem and present an ensemble algorithm for multi-class imbalanced data which is an extension of WHMBoost. The idea is based on using random balance based on average size to balance the data distribution. Their results demonstrate that it has obvious advantages compared with state of the art ensemble algorithms and can effectively deal with multi-class imbalanced data sets. Chen et al in the fourth article of this issue present an approach that is called sparse non-negative matrix factorization for uncertain data clustering where the objective is to minimize the total clustering cost. In their approach they not assume that there is a probability distribution for each uncertain data. Thus all possible locations need to be considered to determine the representative to handle large-scale datasets. Their detailed analysis shows the correctness of their method, and provide an effective initialization and peeling strategy to enhance the ability of processing large-scale datasets. In the next article, Neves Oliveira et al present a hierarchical entity-label disambiguation in named entity recognition task that is based on deep learning. The authors argue that it is unfeasible to obtain information from knowledge bases to give a disambiguated meaning between the entity mentions and the actual labels. This information must be extracted directly from the context dependencies. Their experiments conducted on a real police reports dataset show that the proposed approach significantly outperforms baseline methods. In the last article of this group, Tang et al present a multi-instance positive and unlabelled learning with bi-level embedding which arises from real applications. Unlike other methods using only simple single level mapping, the bi-level embedding strategy is designed to customize specific mapping for positive and unlabeled data. In this approach, the weighting measure adopted in positive data can extracts the uncontaminated information of true positive

instances without interference from negative ones. Their experimental results from this research show that their method has better performance than other state-of-the-art methods.

The last group of articles in this issue are about enabling techniques and applied methods in IDA. In the first article of this group Soliman et al propose a transfer learning-based neural learning method that helps to search knowledge graphs to provide probabilistic reasoning between the queries and their results. The problem is formulated as a classification task where graphs are preprocessed to abstract the N-Triples, then encode the abstracted N-triples into a transitional state that is suitable for neural transfer learning. To validate the proposed approach, the authors employ ten-fold cross-validation. Their results have shown that their approach is accurate by acquiring the average accuracy, recall, precision and f-measure. The second article of this group by Naghibi et al introduces an online learning agent system for cost-sensitive automatic topical data acquisition from the Web with minimum bandwidth usage. The proposed method uses online learning topical crawlers to dynamically adapt to the properties of web pages during the crawling process of the target topic, and learn an effective combination of a set of link scoring criteria for that topic. Using standard metrics in empirical evaluation indicates that when non-learning methods show inefficiency, the learning capability of the proposed approach significantly increases the efficiency of topical crawling, and achieves the state of the art results. The next article of this issue Choi et al introduce a deep learning related framework to analyze stock market data using bi-dimensional histogram and autoencoder. Their constructed stock market network represents the latent space of a bi-dimensional histogram, and network analysis is performed to investigate the structural properties of the stock market. The authors demonstrate that the portfolio consisting of stocks corresponding to the peripheral nodes of bi-dimensional histogram network shows better investment performance than other benchmark stock portfolios. In the next article of this group, Guo et al present an integrated model based on feedforward neural network and Taylor expansion for indicator correlation elimination. The idea is based on a generalized n-power correlation and a feedforward neural network to express the relationship between indicators quantitatively and expanded at every sample to eliminate nonlinear relationships. To compare the elimination efficiency, the authors propose the ranking accuracy to measure the distance of the resulting sequence to the benchmark sequence. In the twelfth article of this issue Meghdouri et al argue that compact data models have become relevant due to the massive, ever-increasing generation of data and propose an observers-based data modelling, which is a lightweight algorithm to extract low density data models that are suitable for both static and stream data analysis. The core sets of the proposed approach keep data internal structures while alleviating computational costs of machine learning during evaluation. The authors compare their proposed approach with existing proposals in classification, clustering, and outlier detection where their results show the best trade-off in accuracy, versatility, and speed. And finally, Bostanian et al in the last article of this issue explain that although ensemble learners and deep neural networks are state-of-the-art schemes for classification applications, they suffer from complex structure, need large amount of samples and also require plenty of time to be converged. The authors propose a new orthogonal version of AdaBoost, called ORBoost, in order to desensitize its performance against noisy samples as well as exploiting low number of weak learners. To assess the performance of their proposed ORBoost, the authors apply a large number of repository datasets where the proposed method is compared with existing methods. The achieved results support the significant superiority of ORBoost to the counterparts in terms of accuracy, robustness, number of exploited weak learners and generalization on most of the datasets.

In conclusion, we would like to thank all the authors who have submitted their manuscripts with the results of their excellent applied and theoretical research to be evaluated by our referees and published in the IDA journal. Over the last few years, our submission rate has increased substantially, although

our acceptance rate remains around 12-15%. We are also glad to announce that our impact factor has increased by 32% since last year (from 0.651 to 0.860). 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