Editorial

Dear Colleague:

Welcome to volume 26(2) of the Intelligent Data Analysis (IDA) Journal.

This issue of the IDA journal is the second issue for our 26th year of publication. It contains fourteen 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 in this issue are about advances in data pre-processing and feature selection in IDA. Thi Thu *et al.* in the first article present an anomaly repair-based approach to improve time series forecasting. The authors argue that anomaly patterns cause negative effects on the accuracy of time series forecasting and propose a novel anomaly repair-based approach to improve time series forecasting in the case of anomaly existence. Their experimental results on several time series datasets reveal that their proposed approach improves remarkably the accuracy of many existing time series forecasting methods and achieves better prediction performance when dealing with anomalies in time series forecasting. The second article of this group by Zhu *et al.* is about topic discovery from short reviews based on data enhancement. The authors argue that with the rapid development of social media and mobile internet, short reviews, such as Weibo and Twitter, have exploded online and discovering topics from short reviews is significant for many practical applications. To improve the efficiency of topic discovery, they introduce the concept of data enhancement and strengthen the data in sentences and words in short reviews based on the weight of importance. Their results show that the proposed method outperforms benchmarks in topic discovery and also has better clustering effects. The third article of this issue by Gupta and Chug is about a feature selection strategy for improving software maintainability prediction that is a significant contributor while choosing particular software. The authors apply ImpS algorithm to handle imbalanced data and investigate several feature selection (FS) techniques including symmetrical uncertainty, Random Forest filter, and Correlation-based FS using one open-source, three proprietaries and two commercial datasets. Their results substantiate that FS techniques significantly improve the performance of various prediction models with a very good improvement. Chu *et al.* in the fourth article of this issue present a modal-aware feature learning with an application to multimodal hashing. The authors argue that many retrieval applications can benefit from multiple modalities, for which how to represent multimodal data is the critical component. The authors present a modal-aware operation as a generic building block to capture the non-linear dependencies among the heterogeneous intermediate features, which can learn the underlying correlation structures in other multimodal data. The modal-aware operation consists of a kernel network and an attention network. Their experiments conducted on three public benchmark datasets demonstrate significant improvements in the performance of their method relative to state-of-the-art techniques. Ljubic and others in the last article of this group present an overview discussion on how to augment data with generative adversarial networks. They argue that performance of neural networks greatly depends on quality, size and balance of training dataset. In order to reduce this problem, methods and techniques are borrowed from the traditional machine learning to deal with imbalanced data. The idea is to generate artificial data through neural networks which appears to be a meaningful solution to the imbalance problem. Their experiments show that new samples of minority class could be created and dataset imbalance ratio could be lowered for better results.

The second group of articles are about advanced learning methods in IDA. Zhang *et al*. in the first article of this group present a noise-resilient online learning algorithm with ramp loss for ordinal regression which is widely used in applications, such as credit portfolio management, recommendation systems, and ecology, where the core task is to predict the labels on ordinal scales. The authors propose a noise-resilient online learning algorithm using the ramp loss function, called PA-RAMP, to improve the performance of their method for noisy data streams. Their experiments on real-world data sets demonstrate that the proposed noise-resilient online ordinal regression algorithm is more robust and efficient than state-of-the-art online ordinal regression algorithms. The seventh article of this issue by Du *et al*. is about learning transferable and discriminative features for unsupervised domain adaptation. The authors argue that transferability and discriminability are two key criteria for characterizing the superiority of feature representations to enable successful domain adaptation. The authors introduce a novel method for unsupervised domain adaptation that is meant to optimize these two objectives simultaneously. Their comprehensive experiments on five real-world datasets and their results verify the effectiveness of the proposed method. Liu *et al*. in the eighth article of this issue argue that graph convolution networks (GCNs) have attracted significant attention, have become the most popular method for learning graph representations and present a refined GCN for filtering with concise and expressive embedding. The idea is to match non-semantic ID inputs. Their extensive experiments conducted on three public datasets demonstrates an improved performance over the state-of-the-art baseline and the effectiveness and rationality of each part of their proposed approach. Bahrkazemi and Mohammadi in the next article of this issue present a strategy to estimate the optimal low-rank in incremental SVD-based algorithms for recommender systems. The authors argue that in recommender methods (such as collaborative filtering and content-based) the estimated value of rank for approximating the recommender systems' data matrix is chosen experimentally in the related literature. The authors investigate the role of singular values for estimating a more reliable amount of rank in the mentioned dimensionality reduction technique. Their numerical results illustrate that the suggested strategy improves the accuracy of the recommendations and run times of applied algorithms. Duan *et al*. in the last article of this group present a topic based shortest path set algorithm for influence maximization. Their article focuses on the influence maximization problem in social networks, which aims to find some influence nodes that maximize the spread of information. Their comprehensive set of experiments conducted on large real-world networks, show that their proposal provides more impressive results in the aspects of influence spread and running time.

The last group of articles in this issue are about enabling techniques and applied methods in IDA. González-Pérez and Sánchez-Gutiérrez in the first article of this group discuss an approach for improving the accuracy of multiclass classification in machine learning where they present a case study on cell signalling. The authors argue that it is important to make sense of the data to propose a useful model to solve a problem where domain knowledge includes information not contained in the data. This will help to understand the data to be fed into a machine-learning algorithm and guide us on what features might help our model. This article is about investigating whether the joint use of feature selection techniques would lead to improving the accuracy in multiclass classification in machine learning. The results presented in this article have shown substantial accuracy increase with fewer input features by only using 3 out of the 16 original descriptors. The twelfth's article of this issue by Yousefi and Tucker is about identifying latent variables in dynamic Bayesian networks with bootstrapping where their particular application is type 2 diabetes complication prediction. The authors argue that predicting complications associated with any complex disease is a challenging task given imbalanced and highly correlated disease complications along with unmeasured or latent factors. The authors attempt to deal with complex imbalanced clinical data,

whilst determining the influence of latent variables within causal networks generated from the observation in which they propose appropriate methods for building dynamic Bayesian networks with latent variables. Their obtained results demonstrate an improvement in the prediction performance. Li *et al.* in the next article present an interactive attention capsule network for similar case matching where the idea is to detect which two cases are more similar for a given triplet where the problem plays a significant role in the legal industry and thus has gained much attention. The authors propose a novel model, called Interactive Attention Capsule Network that attempts to simulate the process of judgment by legal experts, which captures fine-grained elements similarity to make an interpretable judgment. Their experiments based on a real-world dataset demonstrate the superiorities and competitiveness of their proposed model. And finally the last article of this issue by Ni *et al.* is about predicting remaining execution time of business process instances via an auto-encoded transition system. The authors argue that most of the traditional remaining-time-prediction approaches only take into account formal process models and cannot handle large-scale event logs in an effective manner. To overcome these weaknesses, the authors propose a remaining-execution-time-prediction approach based on a novel auto-encoded transition system, which can enhance the complementarity of process modelling and deep learning techniques. Their extensive experiments on four real-world datasets show the effectiveness of the proposed approach.

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