

## Editorial

---

Dear Colleague:

Welcome to volume 24(6) of Intelligent Data Analysis (IDA) Journal.

This issue of the IDA journal is the last issue for our 24<sup>th</sup> year of publication. It contains thirteen articles representing a wide range of topics related to the theoretical and applied research in the field of Intelligent Data Analysis.

The first three articles are about advanced data preprocessing in IDA. In the first article Wu *et al.* introduce an entropy difference and Kernel-based oversampling technique for imbalanced data learning. The authors argue that current oversampling techniques simply consider the imbalance of quantity and pay no attention to whether the data distribution is balanced or not. They propose an entropy difference and Kernel-based method which considers the imbalance degree of dataset from distribution by entropy difference and overcomes the limitation of the well-known SMOTE method for nonlinear problems by oversampling in the feature space of support vector machine classifier. Their experimental study evaluates and compares the performance of their method against state-of-the-art algorithms, and then demonstrates that the proposed approach is competitive on multiple benchmark imbalanced datasets. Zhao *et al.* in the next article discuss sentiment analysis via dually-born-again network and sample selection. The authors investigate an alternative technical path to further improve the state-of-the-art performance of text sentiment analysis and introduce a dually-born-again network in which the teacher network and the student network are simultaneously trained through an iterative approach. Their experiments show that the proposed framework enhances the performance of existing networks. Matsuda *et al.* in the third article present a multiclass spectral feature scaling method for dimensionality reduction. The authors argue that irregular features disrupt the desired classification and consider aggressively modifying scales of features in the original space according to the label information to form well-separated clusters in low-dimensional space. Their numerical experiments show that the proposed method outperforms well-established supervised dimensionality reduction methods for toy problems with more samples than features and real-world problems with more features.

The second group of articles in this issue are about advanced learning methods in IDA. The first article in this group by Sopchoke *et al.* is about explainable and unexpectable recommendations using relational learning on multiple domains where the authors combine relational learning with multi-domain to develop a formal framework for a recommendation system. Their proposed idea is based on constructing general rules for recommendations, providing suggested items with clear and understandable explanations, and delivering a broad range of recommendations including novel and unexpected items. Their experiment results show that the proposed algorithm is very promising and although the quality of recommendations provided is moderate, it does produce interesting recommendations not found in the primitive single-domain based system and with simple and understandable explanations. The next article by Suto is a comprehensive study on plant leaf recognition with shallow and deep learning. Apparently the automated leaf classification is a standalone research area inside machine learning and several shallow and deep methods have been proposed to recognize leaf types. This article provides an overview of plant leaf classification literature, where the authors have found an interesting deficiency (lack of hyper-parameter

search) and a key difference between studies (different test sets). This work gives an overall review about the efficiency of shallow and deep methods under different test conditions and it can be a basis to further research. Xie *et al.* in the sixth article of this issue present a hybrid recommendation model that is based on deep learning and stacking integration strategy. The authors argue that in the traditional recommendation algorithms, due to the rapid development of deep learning and internet technology, user-item rating data is becoming increasingly sparse and present a hybrid model based on deep learning and stacking integration strategy. Their fusion-based model learns the more abstract and deeper nonlinear interaction features by deep learning technology and their proposed hybrid recommendation model can significantly improve the accuracy of rating prediction. Ali *et al.* in the next article present a reinforcement learning based metric filtering approach for evolutionary distance metric learning. The authors propose a hybrid system that combines a sequential feature selection performed by reinforcement learning with the evolutionary feature prioritization of evolutionary distance metric learning in a clustering process. The outcome of this approach has been the best accuracy matrix with the least number of elements where the results show a significant decrease in the number of features while maintaining or increasing accuracy.

And finally the third group of articles are about enabling techniques and innovative application in IDA. Nalluri *et al.* in the eight's article present a data-driven disease classification and fine-tuning approach that is based on a multi-objective evolutionary algorithm. The authors have used sequential minimal optimization classifier as the base classifier and three evolutionary algorithms namely cat swarm optimization, invasive weed optimization and eagle search based invasive weed optimization to diagnose particular diseases from available datasets. Their statistical tests substantiate the cogence of the obtained results. Mohanraj *et al.* in the next article present a hybrid deep learning model to predict and target vaccination rates in the less immunized regions of India. The proposed rank-based multi-layer perceptron hybrid deep learning framework uses substantial amount of data and predicts and categorizes the percentage of partly immunized vaccination rates as extreme, low and medium ranges. This predicted finding is cross-verified by deep soft cosine semantic and ranking SVM based model. Their comparative results are evaluated using a number of well-known evaluation measures where the results show that the hybrid deep learning system is clearly superior to any other alternative approach. Meeng *et al.* in the tenth article of this issue discuss uni- and multivariate probability density models for numeric subgroup discovery. The approach is about continuous targets, which are important in many practical applications for which differences are often quantified using z-score and similar measures that compare simple statistics such as the mean and variance of the subset and the data. The proposed approach can be incorporated easily into existing subgroup discovery frameworks, so no new frameworks are developed. Li *et al.* in the next article present an adaptive computing framework for multimedia data streams analysis where the approach is based on efficient online computing/filtering or analysis of multimedia streams. The authors argue that existing methods are based on a greedy strategy which orders the filters according to several factors. The authors propose a dynamic-analytic hierarchy process framework which uses a time-based compositional forecasting approach, which is based on the idea of exponential smoothing to deal with the factors' proportion relationships dynamics. Their experiments on both synthetic and real life multimedia streams demonstrate that their proposed framework provides great adaptability in modeling the factors proportion relationships changing over multimedia stream environment. The twelfth article of this issue by Wang *et al.* is about improving multi-view facial expression recognition through two novel texture-based feature representations. The authors argue that although several automatic computer systems have been proposed to address facial expression recognition problems, the majority of them still fail to cope with some requirements of many practical application scenarios. They explore and investigate one of the most influential and common issues raised in practical application scenarios when

applying automatic facial expression recognition system, head pose variation. And finally, the last article of this issue by Sivanaiah is about content boosted hybrid filtering applied to recommendation systems where the authors argue that content-based and collaborative filtering are the traditional recommendation strategies, each with its own strengths and weaknesses. The authors further emphasize that in content-based filtering techniques pessimistic users get poor recommendations of either uninteresting products or no recommendations at all. This can be alleviated by boosting the content profiles of pessimistic users using the top-n recommendations of collaborative filtering. Their proposed content boosted hybrid filtering system provides a novel list of recommendations even for pessimistic users, with predictive accuracy better than that of a traditional content-based filtering system.

In conclusion, we would like to thank all the authors who have submitted the results of their excellent research to be evaluated by our referees and published in the IDA journal. 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