# Editorial

Dear Colleague:

Welcome to volume 25(4) of Intelligent Data Analysis (IDA) Journal.

This issue of the IDA journal is the fourth issue for our 25th year of publication. This issue 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 are about data preprocessing and data understanding. Ngueilbaye *et al*. in the first article of this group argue that various forms of data analytics use heterogeneous and distributed data where the proliferation of these platforms faces many challenges. The authors present an extended complete formal taxonomy and some algorithms that detect and correct contextual data quality anomalies. Their methods are effective in detecting and correcting more data anomalies than existing taxonomy techniques, and also highlight the demerit of Support Vector Machine. Vahldiek *et al*. in the next article of this group present a data generator for multivariate numerical data with arbitrary correlations and distributions. The authors argue that artificial or simulated data are relevant when real data may not be used for reasons of data protection, or when special distributions or effects should be present in the data to test certain machine learning methods. They present a data generator for multivariate numerical data with arbitrary marginal distributions and arbitrary correlations which is implemented in the open source statistics software R. Additionally, outliers can be integrated and the use of the data generator is demonstrated with a concrete example. Ye *et al*. in the third article of this issue present a human interaction recognition method that is based on parallel multi-feature fusion network. The proposed approach that can improve network performance and reduce the amount of network parameters at the same time. The proposed algorithm is compared with the traditional activity recognition algorithm, and their experiments show that it can accomplish the recognition tasks of six kinds of interactive actions in a much better way with high accuracy. The fourth article of this issue by Jaffar Khan *et al*. is about handling incomplete data classification using imputed feature selected bagging. The idea is to consider the efficiency of classifiers to be more important than building a bunch of classifiers. The authors propose an imputed feature selected bagging method which uses multiple imputation, feature selection and bagging ensemble learning approach to construct a number of base classifiers to classify new incomplete instances without any need for imputation in testing phase. Their experimental results show the proposed method is considerably faster and gives high accuracy for classification when used with incomplete data as compared to other methods.

The second group of articles in this issue are about unsupervised and supervised learning methods. Homayouni and Mansoori in the fifth article of this issue introduce a manifold regularization ensemble clustering with many objectives using unsupervised extreme learning machines. The authors argue that spectral clustering has been an effective clustering method, in the last decades, because it can get an optimal solution without any assumptions on data's structure. In their proposed method, several objective functions are considered simultaneously, as a robust method for constructing the similarity matrix aiming to find robust partitioning that satisfy multiple objectives, handling noisy data, keeping diversity-based goals, and dimension reduction. Their experiments on some real-world datasets besides

three benchmark protein datasets demonstrate the superiority of their method over some state-of-the-art single and ensemble methods. In the next article of this group Tan *et al.* propose a new semi-supervised algorithm combined with mutual-cross imperical competition algorithm that is mostly suitable for motion imagination classification. The proposed algorithm not only reduces the tedious and time-consuming training processes and enhances the adaptability of Brain Computer Interface, but also optimizes the parameters of an SVM method in the entire semi-supervised process. The experiments include selected optimal parameters to a public data set to verify its effectiveness. Merkurjev in the seventh article of this issue introduces a modified auction dynamics technique suitable for semi-supervised multiclass data classification. The proposed approach seems to be an efficient and unconditionally stable semi-supervised graph-based method for multiclass data classification which requires considerably less labelled training data to accurately classify a data set compared to current techniques. Extensive experiments on benchmark datasets are performed and the results are compared to other methods. Krechowicz in the next article of this issue discuss content-aware data distribution over cluster nodes and introduces two custom data items addressing methods for distributed data storage targeting scalable distributed two–layer data store. The basic idea of those methods is to preserve that data items stored on the same cluster node are similar to each other following concepts of data clustering. The proposed methods generate good results efficiently in comparison to traditional clustering techniques like k-means, agglomerative and birch clustering. Yu *et al.* in the last article of this group introduce an instance confidence-based decision tree algorithm for classification. The authors investigate the impact and significance of degree of confidence of instances on the classification performance of decision tree algorithms, taking the classification and regression tree algorithm as an example. Their experimental results show that the algorithm can significantly improve the generalization performance as well as avoiding the over-fitting problem to a certain extent.

The last group of articles in this issue are about enabling methods in IDA. Nannan *et al.* in the first article of this group present an approach for prediction of online time series with concept drift that is based on dynamic intuitionistic fuzzy cognitive map. In order to tackle with the possible hesitancy in the process of modeling, intuitionistic fuzzy set is applied in the construction of dynamic fuzzy concept map, where hesitation degree as a quantitative index explicitly expresses the hesitancy. Their experiments using public data sets verify the effectiveness of the proposed method. Wang and Bai in the next article of this group present an approach for predicting breadth and depth of microblog users' behaviour. They apply machine learning algorithms to a large number of data sets that are is preprocessed, and the extracted features are divided into three types: user features, microblog features and social features. Their experimental results show that the prediction precision of the improved random forest algorithm has less fluctuations, and it is not sensitive to the changes of various features. Da and Yang in the twelfth article of this group present a merged architecture for low-resolution instance segmentation. The domain of instance segmentation has a wide range of applications, including video surveillance, autonomous driving, and behaviour analysis. Their experiments demonstrate that their proposed technique exhibits good pixel-level segmentation performance in terms of both accuracy and computational efficiency for a given low resolution input, and it can be easily implemented in embedded platforms. Zeeshan *et al.* in the next article of this group present a feature-based multi-criteria recommendation system that is based on weighted approach with ranking correlation. The authors argue that recommendation systems are more effective if being used for making a recommendation using multiple criteria of decisions by using the correlations between the features and items content or finding a similar user rating to get targeted results. To evaluate their approach they have tested the proposed algorithm on different datasets having multiple features of information where the results demonstrate that proposed method is efficient in different types of datasets. And finally, Lai and Wu in the last article of this issue present a multimodal emotion recognition with hierarchical memory

networks to address the issues of recognizing utterance level emotions. to capture the interpersonal dependencies, their approach employs global memory networks to integrate the local outputs into global storages. Their experiments on the public datasets show their model outperforms the compared methods with accuracy improvement.

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. Over the last few years, our submission rate has exceeded 600 manuscripts per year, with an acceptance rate of around 12–15%. 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