Intelligent Data Analysis 25 (2021) 1047–1049
DOI 10.3233/IDA-210003
IOS Press

Editorial

Dear Colleague:

Welcome to volume 25(5) of Intelligent Data Analysis (IDA) Journal.

This issue of the IDA journal is the fifth issue for our 25th 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 are about advanced data preprocessing and data understanding. Noering et al. in the first article of this group present an approach for improving discretization based pattern discovery for multivariate time series. The authors argue that load situations in engineering applications are challenging to be defined by hand. In other words, they need to be discovered as recurrent patterns within multivariate time series data of the system under consideration. They propose a series of preprocessing steps for symbolic representation of time series aiming for enhanced multivariate pattern discovery. They also show the performance of their proposed algorithms in a synthetic test data set as well as in a real life example with 100 million time points. Miswan et al. in the next article of this issue present a predictive modelling approach for different preprocessing techniques on machine learning classifiers. Their research is about preprocessing applied by previous hospital readmission that are compared, and the most common approaches highlighted such as missing value imputation, feature selection, data balancing, and feature scaling. Their results indicated that the preprocessing approaches helped improve the model's prediction of hospital readmission. Chen et al. in the third article of this issue present a non-binary hierarchical tree overlapping community detection that is based on multi-dimensional similarity. The authors argue that overlapping communities exist in real networks, where the communities represent hierarchical community structures, such as schools and government departments. The authors further explain that a non-binary tree would allow a vertex to belong to multiple communities to obtain a more realistic overlapping community structure. They propose a non-binary hierarchical tree overlapping community detection that is based on multi-dimensional similarity. Their experiments on real-world networks demonstrate that their proposed algorithm is superior to other representative algorithms in terms of the quality of overlapping community detection. Wan et al. in the last article of this group discuss an attribute interaction aware matrix factorization (MF) method for recommendation tasks. The authors argue that when the original rating matrix is particularly large, the training time complexity of the MF-based model becomes higher and the update cost of the model is also higher. They partition the original rating matrix into different sub-matrices according to the attribute interactions, train each submatrix independently, and merge all the latent vectors to generate the final score. Their experimental results on the two classic public datasets show that their approach can not only effectively improve the accuracy of recommendation, but also make full use of parallel computing technology to improve training efficiency without modifying the matrix factorization model.

The second group of articles in this issue are about learning methods in IDA. Srivastava and Mittal in the first article of this group present an age decay accuracy weighted ensemble method for drifting data stream mining. The authors explain that a data source of a dynamic environment generates data streams in which the underlying data distribution changes very frequently with respect to time and hence results in concept drifts. The authors propose a novel ensemble approach for learning in concept drifting data streams. The proposed method assigns weights to the component classifiers based on its accuracy and its remaining life-time in the ensemble is such a way that ensures maximum accuracy. Their experiments on benchmark artificial drifting data stream generators real datasets and compares its performance with well-known state-of-the-art existing methods which outperform existing methods. Li et al. in the next article of this group present a novel framework for biological network classification via machine learning to predict the label of both small-scale and large-scale biological network data flexibly and efficiently. Their framework firstly presents a simplified graph kernel method to capture the structural information of each graph where obtained informative features are adopted to train different scale biological network data-oriented classifiers to construct the prediction model. Their experiments on five benchmark biological network datasets on graph classification task show that the proposed model outperforms the state-of-the-art methods with a large accuracy rate. He et al. in the next article present a two-stage clustering that is based on cold-start method for active learning. In the first stage, to separate the instances of minority class from that of majority class, a multi-centre clustering is constructed and in the second stage, the initial training instances are selected from each cluster based on an adaptive candidate representative instances. Their experiments demonstrate the effectiveness of the proposed method from the aspects of class coverage, classification performance, and impact on active learning. In the eighth article of this group, Wang et al. present an approach for location prediction for facility placement by incorporating multi-characteristic information. The authors introduce a novel model of multi-characteristic Information based Top-k location prediction that captures the spatio-temporal behaviors of customers based on historical trajectories, exploits the social relevancy from their friend relationships, as well as examines the category competitiveness of specific facilities thoroughly. Their experiments conducted on real-world datasets demonstrate the higher efficiency and effectiveness of the proposed model. Vo and Do in the next article of this group present a novel approach of graph-of-words based mixture model for semantic-enhanced text stream clustering. Different from traditional static text clustering approach, text stream clustering task has specific key challenges related to the rapid change of topics/clusters and high-velocity of coming streaming document batches. They propose a novel approach of graph-of-words based text stream clustering, where they generate clusters from each document batch while sampling clusters/topics can support to overcome the word-independent evaluation challenge. Their extensive experiments on multiple benchmark real-world datasets demonstrate the effectiveness of their proposed model in both accuracy and time-consuming performances.

The last group of articles in this issue are about enabling techniques and applied methods in IDA. Maafiri and Chougdali in the first article of this group present a robust face recognition algorithm that is based on a new kernel-PCA. They propose a new algorithm in order to enhance PCA performance by exploiting the rank-revealing factorization. They also improve the recognition rate of existing PCA methods by developing a nonlinear extension of the approach. Their extensive experiments on well-known standard face databases show that the proposed algorithm is more robust than several variations of conventional PCA, in terms of face recognition accuracy. Chen et al. in the eleventh article of this issue present a personalized trajectory privacy-preserving method that is based on sensitive attribute generalization and location perturbation. The authors argue that since different users have different requirements regarding the anonymity of sensitive information, attackers can use specific background knowledge to analyze published trajectory data and access a user's private information. The results of theoretical analyses and experimental evaluations demonstrate the effectiveness of the proposed method in preserving personalized privacy in published trajectory data. Wang et al. in the next article of this group discuss how a virtual sample based block-diagonal dictionary can be used for learning in face recognition. The authors argue that since the recognition performance may be adversely affected by the noise, they propose a virtual samples based robust block diagonal dictionary learning for face recognition. In the proposed model, both the original and virtual samples are combined to solve the small sample size problem, and both the structure constraint and the low rank constraint are exploited to preserve the intrinsic attributes of the samples. Their extensive experiments are conducted in comparison with many state-of-the-art methods on benchmark face datasets, and experimental results demonstrate the efficacy of the proposed method. Singla et al. in the next article present an ensemble machine learning method applied to spatio-temporal data analysis of plant and ratoon sugarcane. Their proposed method demonstrates that temporal variations in the spectral data and computational efficiency of machine learning methods may be effectively used for the discrimination of types of sugarcane. The performance of their binary classification model is based on the best model in all the possible combination of input images and binary classification on the remotely sensed data can be effectively handled using random forest method. And finally, Saito et al. in the last article of this issue present a method for computation of target-oriented link criticalness centrality in uncertain graph. The authors utilize the bridge detection technique in graph theory and efficiently identify critical links in case the node reachability is taken as the performance measure. They define a set of target nodes and a new measure associated with it, which is the marginal loss of the expected number of nodes in the network that can reach. They apply the proposed method to two real-world networks, one from social network and the other from spatial network, and empirically show that the proposed method has a good scalability with respect to the network size.

In conclusion, we would like to thank all the authors who have submitted their manuscripts with 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 650 manuscripts per year, with an acceptance rate of 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). In addition, there is a special issue of the IDA journal under preparation entitled: "*Cloud/Fog/Edge Computing for Urban Big Data*" that is scheduled for publication for next year. Interested authors can contact: Prof. Jerry Chun-Wei Lin (jerrylin@ieee.org) to submit their manuscript. 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