Editorial

Dear Colleague:
Welcome to volume 26(6) of the Intelligent Data Analysis (IDA) Journal.

This issue of the IDA journal is the last 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 are about state of the art data pre-processing methods in IDA. The first article by Qi and Chen is about a novel density-based outlier detection method that is based on a set of key attributes. The authors argue that although many related technologies have been proposed, most of them are faced with the problem of the neighborhood size of an object which is difficult to determine. To overcome this weakness, the authors propose a novel density-based outlier detection method that is based on the stability of the reverse minimum of the sum of edge set. Their experiments on synthetic and real-world datasets demonstrate that their method is more effective than the existing outlier detection approaches. The second article of this group by Matsue and Sugiyama is about unsupervised feature extraction from multivariate time series for outlier detection. The authors argue that although various feature extraction algorithms have been developed for time series data, it is still challenging to obtain a flat vector representation with incorporating both of time-wise and variable-wise association between multiple time series. The authors introduce an algorithm, that constructs feature vector representation for multiple time series in an unsupervised manner which they examine the effectiveness of the extracted features under the unsupervised outlier detection scenario using synthetic and real-world datasets. They show its superiority compared to well-established baselines. In the last article of this group, Haoran et al. present a collusive anomalies detection method that is based on collaborative Markov random field. The authors propose a novel Markov Random Field-based method, considering node level and community-level behavior features. The proposed method has several advantages that are based on the analysis of the nodes' local structure, the community-level behavioral features among which is that it operates in a completely unsupervised fashion requiring no labeled data, while still incorporating side information, if available. Through their experiments on a user-reviewed dataset, the results show that the proposed approach can significantly outperform state-of-the-art baselines in collusive anomalies detection.

The second group of articles are about advanced learning methods in IDA. Li et al. in the first article of this group, present a density based clustering ensemble approach that is based on selecting internal validity index. The idea is to consistently integrate multiple clustering results to obtain better division as most of ensemble research employs a single algorithm with different parameters for clustering. The innovation of this article consists of setting dynamic thresholds where reconstructed matrices are analyzed by hierarchical clustering to obtain basic clustering results and an internal validity index is designed by the compactness within clusters. The authors show that the clustering effect is Significantly improved. The article reports a series of experiments where the results verify the improvement and effectiveness of the proposed technique. She et al. in the second article of this group present an adaptive fuzzy C-means (FCM) clustering integrated with local outlier factor. The authors argue that the conventional fuzzy C-means is sensitive to the initial cluster centers and outliers, which may cause the centers deviate from

the real centers when the algorithm converges. To improve the performance of FCM, the authors propose a method of initializing the cluster centers that is based on probabilistic suppression. Their experiments on synthetic and real-world datasets demonstrate the clustering performance and anti-noise ability of the proposed method. Hsu and Nguyen in the sixth article of this issue present an approach for discovering suitable number of clusters for fuzzy clustering where the authors argue that the main problem of Fuzzy c-Means (FCM) is deciding on an appropriate number of clusters. The approach is to determine suitable number of clusters without repeated execution based on a singular value decomposition. Based on the percentage of variance, this method can calculate the appropriate number of clusters. The proposed method was applied to several well-known datasets to demonstrate its effectiveness. Yuan et al. in the next article of this issue present a graph structure learning that is based on feature and label consistency. The authors argue that graph neural networks (GNNs) have achieved remarkable success in graph-related tasks by combining node features and graph topology elegantly. The authors design a simple and effective Graph Structure Learning strategy based on feature and label consistency to increase the homophilous level of networks for generalizing any existing GNNs to heterophilous networks. Their empirical results on public networks with homophily or heterophily, and structure attacks show that their methods outperform the state-of-the-art methods in most cases. The eighth article of this issue by Zou et al. is about a credit scoring system that is based on a bagging-cascading boosted decision tree. The authors propose a hybrid ensemble method that combines the advantages of the bagging ensemble strategy and boosting ensemble optimization pattern, which can well balance the trade-off of variance-bias optimization. Their experimental results on a number of datasets show the proposed approach provides a more accurate credit scoring result. In the next article of this issue, Chong et al. introduce a shapley value-based resampling approach for imbalanced data classification. The proposed method removes the noise data according to the Shapley value and undersamples the samples with Shapley values less than zero in the majority class. Their experimental results show that the proposed method can significantly improve the effect of imbalanced data classification. The last article of this group by Shin et al. is entitled a penalized additive neural network regression. The authors introduce an additive neural network model that is constructed by using a linear combination of univariate neural networks, or equivalently functional components. They use a B-spline activation function, which is useful to capture local features of data, for nodes that constitute the model. Their numerical studies show that the fitted functional components adapt to local and sparse structures based on a given dataset.

The last group of articles in this issue are about enabling techniques and innovative case studies in IDA. The first article of this group by Pecar et al. is about evaluation of end-to-end aspect-based sentiment analysis (ABSA) that employs novel benchmark dataset for aspect and opinion review analysis. The authors argue that pipeline approaches do not model correlations between the tasks and address this bottleneck by introducing the first purposely designed and annotated dataset for ABSA. The authors evaluate this dataset on several experiments employing state-of-the-art models and set benchmarks and analyze the strengths as well as weaknesses of the data and their proposed approaches. Zu and Xie in the twelfth article of this issue introduce a keyphrase extraction method that uses deep and wide learning features. The authors argue that traditional statistic-based methods for keyphrase extraction only make use of statistical features of the words and ignore the semantic relationship between words. Their experimental results on two public datasets show that the performance of their proposed model is better than eight common baseline key phrase extraction methods. The thirteenth article of this issue by Wu et al. is about a non-overlapping approximate pattern matching approach which is used to calculate the support of patterns, a key issue in sequential pattern mining. The approach is based on Hamming distance, which cannot be used to measure the local approximation between the subsequence and pattern,

resulting in large deviations in matching results. The authors present numerous experiments to verify the performance of the proposed algorithm. And finally in the last article of this issue Peng et al. present a graph convolutional networks-based robustness optimization approach for scale-free internet of things. The authors argue that the Internet of Things (IoT) devices have limited resources and are vulnerable to attacks, so optimizing their network topology to resist random failures and malicious attacks has become a key issue. The authors propose an intelligent topology robustness optimization model based on a graph convolutional network. Their extensive experimental results demonstrate that the proposed approach can more effectively improve the robustness of scale-free IoT networks against malicious attacks compared to two existing heuristic algorithms.

In conclusion for the last issue of 2022, I would like to mention I founded the IDA journal in 1995–96 where it was launched in July 1996 and the first issue was published in January 1997 (by Elsevier-North Holland and then transferred to the IOS Press in late 1999). Thanks so mush to all the IOS Press staff who have managed the IDA journal so well and to colleagues like yourself, who have who have submitted their manuscripts to be evaluated by our editorial board members and published in the IDA journal. Now, after so many years, I am stepping down as Editor-in-Chief and transferring the responsibility to my colleague, Dr. Jose Maria Pena (from Oxford, UK), whom I have known since 1997. Please join me in welcoming Dr. Pena to the position of the Editor-in-Chief of the IDA Journal. We are also glad to announce that our impact factor has increased by over 50% since last year (from 0.860 to 1.321). 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   **Dr. J.M. Pena**
Founding Editor   **Editor-in-Chief**