

## Editorial

---

Dear Colleague:

Welcome to volume 27(1) of the Intelligent Data Analysis (IDA) Journal.

This issue of the IDA journal is the first issue for our 27<sup>th</sup> 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 state of the art data pre-processing methods in IDA. The first article by Fushimi *et al.* is about outlier detection based on distance minimization. The authors propose a new method of constructing a variable bin width histogram that can accommodate the unbalanced distribution of the samples yet retaining, as a whole, the good aspect of both equal width and equal-area histograms that are being used popularly for data visualization and analysis. The authors also propose a method to annotate the constructed bins if the data for annotation is given for each sample as a set of nominal variables, using z-score with respect to their distribution within each bin. The authors apply the proposed ideas to both real and synthetic datasets, and confirm that both methods work as intended and both can represent the sample distribution with a smaller number of bins than. The last article of this group by Guo and Liu is also about anomaly detection where the authors propose a spatial-temporal trajectory anomaly detection approach that is based on an improved spectral clustering algorithm that is based on the multi-scale threshold and density combined with shared nearest neighbors. Taking Internet usage data on campus as an example, multiple clustering algorithms are used for anomaly detection and four evaluation metrics are applied by the authors of this article to estimate the clustering results. Where abnormal trajectories list is verified to be effective and credible.

The second group of articles are about supervised and unsupervised learning methods in IDA. The first article by Li *et al.* is an approach for large-scale spectral ensemble clustering which is combining multiple base clustering to obtain improved results. The authors propose a large-scale spectral ensemble clustering (LSEC) method to balance efficiency and effectiveness. The authors further combine all the base clustering using a bipartite graph partition-based consensus function to obtain improved consensus clustering results. Their experiments conducted on ten large-scale datasets demonstrate the efficiency and effectiveness of the LSEC method. The second article of this group by Liao *et al.* is a novel application based on hybrid of balanced bagging classification and light gradient boosting machine to predict traffic crash severity to eliminate bias and variance. The proposed model has demonstrated better performance when compared with other models such as Gaussian Naïve Bayes, Support vector machines, and Random Forest. More specifically, the proposed model managed to achieve better performance among all metrics for validation dataset.

The third group of articles are about enabling techniques and innovative case studies in IDA. Peng *et al.* in the first article of this group present an approach for collaborative optimization that is suitable for named entity recognition-based applications. Named entity recognition (NER) as a crucial technology is widely used in many application scenarios, including information extraction, information retrieval, text summarization, and machine translation assisted in AI-based smart communication and networking systems. The authors propose a machine learning-guided model to achieve NER, where the hyper-parameters

of model are automatically adjusted to improve the computational performance. The experimental results demonstrate the satisfactory performance of their proposed model. The seventh article of this issue by Li *et al.* is entitled wide & deep generative adversarial networks for recommendation systems. The approach is capable of extracting both explicit and implicit information of user's preferences. The authors combine Cross-Entropy loss in G with Wasserstein loss in D to get data distribution where the joint loss will be to receive the training information feedback from data distribution. Their empirical results on three public benchmarks show that their proposed approach significantly outperforms state-of-the-art methods. The next article of this issue by Li *et al.* is about inferring student social link from spatiotemporal behavior data via entropy-based analyzing model. The authors argue that social link is an important index to understand master students' mental health and social ability in educational management. Therefore, extracting hidden social strength from students' rich daily life behaviors has also become an attractive research hotspot. Their proposed approach is based on students' multi-source heterogeneous behavioral data to calculate the frequency of co-occurrence under the influence of time intervals. Their experiments show that the proposed method is superior to the traditional methods under many evaluating criteria. In the ninth article of this issue Deng *et al.* present a simple loss function training method for deep neural networks. The approach is based on multiple independent losses scheduling, which allows multiple loss functions to independently participate in the training process according to their performance. Their extensive experiments using various deep architectures on various recognition benchmarks demonstrate the proposed scheme is simple, robust, lightweight, and effective for typical classification tasks. In the last article of this group, Tan *et al.* present a guided node graph convolutional networks for repository recommendation. The authors present an end-to-end framework, namely guided node graph convolutional network, which effectively captures the connections between entities by mining the influence of related nodes. The proposed approach is evaluated on a real-world Github and a music recommendation dataset, and the experimental results show that the method outperforms the recommendation baselines.

The last group of articles in this issue are about advanced learning methods in IDA. The first article by Lv and Dong present an adaptive active learning algorithm with informativeness and representativeness. The authors first present an adaptive active learning framework, in which the weight of informativeness and representativeness criteria can be dynamically updated by the feedback of previous learning processes. Secondly, by formulating the active learning as a Markov decision process, it can adaptively select the suitable sampling strategies according to the reward of the learning process. Different from traditional active learning algorithms, this approach can adaptively select sampling strategies and adjust the weights simultaneously, which helps it more feasible in the application. The second article of this group by Santos *et al.* is entitled: Bayesian estimation of decay parameters in Hawkes processes, a ubiquitous tool for modeling and predicting event times. The authors demonstrate that these estimation difficulties relate to the noisy, non-convex shape of the Hawkes process's log-likelihood as a function of the decay. The authors show that their approach alleviates the decay estimation problem across a range of experiments with synthetic and real-world data. Wu *et al.* in the thirteenth article of this issue present a temporal motif-based attentional graph convolutional network approach for dynamic link prediction. The authors propose a temporal motif matrix construction method to capture higher-order structural and temporal features, then introduce a spatial convolution operation following a temporal motif-attention mechanism to encode these features into node embedding. Their experimental results on various real-world datasets demonstrate that the proposed model is superior to state-of-the-art baselines on the dynamic link prediction task. In the next article of this issue, Zou *et al.* propose an approach for efficient mining of maximal l-reachability co-location patterns from spatial data sets. Co-location patterns are sets of spatial features that are strongly correlated in space. Because the average size of l-reachability

co-location patterns tends to be longer, maximal 1-reachability co-location pattern mining is investigated in this article. The effectiveness and efficiency of their proposed model and algorithms are analyzed by extensive comparison experiments on synthetic and real-world spatial data sets. And finally, in the last article of this issue, Li *et al.* discuss how to find reinforced structural hole spanners in social networks via node embedding. The authors argue that identifying structural hole spanners that benefit from acting as bridges between communities is a core study in social network analysis. The authors propose a node embedding-based method for identifying reinforced structural hole spanners in social networks. Their extensive experimental results show that the performance of hole spanners identified by the proposed method outperforms several existing methods.

In conclusion for the first issue of our volume 27, I would like remind you that as the founding Editor-in-Chief of the IDA journal, I am gradually wrapping up my duties 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***