

## Editorial

---

Dear Colleague:

Welcome to volume 26(5) of the Intelligent Data Analysis (IDA) Journal.

This issue of the IDA journal is the fifth issue for our 26<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 are about state of the art data pre-processing methods in IDA. The first article by Moura *et al.* is about label noise detection at random model with ensemble filters. This research investigates the performance of ensemble noise detection under two different noise models: the noisy at random and the noisy completely at random models, in which the probability of label noise is entirely independent. The authors investigate the effect of class distribution on noise detection performance since it changes the total noise level observed in a given performed experiments. The second article of this issue by Li *et al.* is about identification of informational and probabilistic independence by adaptive thresholding. The authors propose a novel learning technique, called Adaptive Independence Thresholding (AIT), to automatically identify the informational independence and probabilistic independence. Their experimental results on a large number of benchmark datasets from the UCI machine learning repository show that AIT is more effective than other learning techniques. Kim and Kim in the next article of this issue present an improved False Discovery Rate (FDR) method for edge detection that applies an expectation maximization approach. The approach is based on an estimation procedure which is more data sensitive under some conditions. The proposed method is compared favourably with the most popular FDR tool in numerical experiments. In the last article of this group, Boukela *et al.* present an approach for unsupervised contextual anomaly detection and characterization. The authors argue that outlier detection has been widely explored and applied to different real-world problems. However, outlier characterization that consists of finding and understanding the outlying aspects of the anomalous observations is still challenging. The authors present a new approach to simultaneously detecting subspace outliers and characterize them. Their extensive experiments on various datasets show the efficacy of their method and its competitiveness compared to the existing solutions in terms of both detection and characterization accuracy.

The second group of articles are about advanced learning methods in IDA. Sun *et al.* in the first article of this group present a multiple hierarchical clustering ensemble algorithm to recognize clusters that are arbitrarily shaped. The authors argue that most of the existing clustering ensemble algorithms take the linear clustering algorithms as the base clustering. The authors present an Agglomerative Nesting method that is meant to build base clusters for integrating multiple clusterings. The proposed approach can establish the proposed algorithm's superiority in terms of clustering performance by comparing the proposed algorithm's clustering performance to that of existing clustering algorithms on different datasets. In the second article of this group, Zhang et introduce LAMB, which is a novel algorithm for label collaboration-based multi-label learning. The authors argue that multi-label learning is more complex than single-label learning for that the labels tend to be correlated. They also emphasize that traditional multi-label learning algorithms learn independent classifiers for each label and employ ranking

or threshold on the classification results. The authors propose a high-order multi-label learning algorithm that utilizes collaboration between its own prediction and the prediction of other labels. Their extensive experiments on various datasets demonstrate that their proposed algorithm achieves superior performance over existing state-of-the-art algorithms. Ávila Mendes and da Silva in the next article of this issue present an approach for modeling the combined influence of complexity and quality in supervised learning. The authors argue that a significant part of the performance of any classification algorithm depends on the dataset's complexity and quality. The authors apply a structural equation modeling and partial least squares structural equation modelling algorithm to solve the problem of classification quality. Their experimental analysis with 178 datasets showed that the control of complexity improves the classification results more than data quality does. In the eighth article of this issue Wang *et al.* present a novel Bayesian learning approach where they discuss both undirected dependence and directed causality. They argue that for most Bayesian network classifiers the directed edges substantially represent assertions of conditional independence rather than causal relationships although the learned joint probability distributions may fit data well. They apply ensemble learning strategy to build Bayesian causal forest with a set of Bayesian classifiers. Their experiments performed on 32 public datasets achieve outstanding classification performance compared to state-of-the-art single-model methods. The next article by Liu *et al.* is entitled IMLBoost for intelligent diagnosis with imbalanced medical records. The authors introduce a new imbalance lessened boosting algorithm to better classify imbalanced medical records, highlighting the contribution of samples in minor classes as well as hard and boundary samples. Their experimental results on five UCI imbalanced medical datasets have demonstrated the effectiveness of the proposed algorithm compared with other existing classification methods. The tenth article of this issue by Wang *et al.* is about influence maximization (IM) based on network representation learning in social networks. The authors argue that a large number of existing studies mainly focus on the heuristic methods, which generally lead to sub-optimal solutions and suffer time-consuming and inapplicability for large-scale networks where no research has been found that surveyed both of them. The authors attempt to solve the IM problem by introducing a novel influence network embedding approach and a novel influence maximization algorithm. Their experiments on four real-world networks indicate that their proposed method outperforms four state-of-the-art network embedding baselines. In the last article of this group, Espinoza and Norambuena present an approach for evaluating semantic representations for extended association rules. The authors use natural language processing techniques in conjunction with machine learning classifiers where they report the classification performance using the F1 and accuracy metrics. Their experiments indicate that the choice of semantic text representation does not have major effects on the performance of their approach for polarity classification.

The last group of articles in this issue are about enabling techniques and innovative case studies in IDA. He *et al.* in the first article of this group discuss adversarial attacks and fostering enhanced models based on generative adversarial networks. The authors present a creative framework based on a generative adversarial network (GAN) with a series of training algorithms that aims to generate instances of adversarial attacks and utilize them to help establishing a new Intrusion Detection System (IDS) based on a neural network that can replace the old IDS without knowledge of any of its parameters. The proposed framework can generate four types of adversarial attacks with the same number of classifiers with an average detection rate of 98% in coping with both generated adversarial and original attacks. Nabipour *et al.* in the next article of this group present an approach to energy-aware resource management: toward Green NOMA heterogeneous networks. The authors propose a self-optimizing model to enhance network performance and guarantee the users quality of service requirements by considering limited resources and using effective user association, carrier scheduling and handover optimization algorithms.

In order to maximize the network performance. These proposed algorithms use the carrier matching feature and optimal transmission power for problem-solving. Their simulation results prove that, despite the increased computational complexity, effective resource allocation and optimal relations made the proposed approach capable to increase performance indices, such as network throughput, by up to 30%. And finally, Zhao *et al.* in the last article of this issue present a robust intuitionistic fuzzy clustering with bias field estimation for noisy image segmentation. The concept of intuitionistic fuzzy set has been found to be highly useful to handle vagueness in data. Based on intuitionistic fuzzy set theory, intuitionistic fuzzy clustering algorithms are proposed which cannot acquire satisfying performance when applied to segmented images corrupted by noise. In order to solve these problems, a robust intuitionistic fuzzy clustering with bias field estimation (RIFCB) is proposed for noisy image segmentation. Their experimental results on some Berkeley images show that the proposed method achieves satisfactory segmentation results on images corrupted by different kinds of noise in contrast to conventional fuzzy clustering algorithms.

In conclusion, 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). This year, we are completing our volume 26. Thanks so much to all the IOS Press staff and colleagues like yourself, who have who have submitted their manuscripts with the results of their excellent applied and theoretical research 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. It is my pleasure to let you know that Dr. Pena is now the Editor-in-Chief. His e-mail is: jmpena@ida-ij.com. 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  
Founder

Dr. J.M. Pena  
Editor-in-Chief