Intelligent Data Analysis 24 (2020) 741–742
DOI 10.3233/IDA-200003
IOS Press

Editorial

Dear Colleague:
Welcome to volume 24(4) of Intelligent Data Analysis (IDA) Journal.

This issue of the IDA journal is the fourth issue for our 24th year of publication. It contains twelve articles representing a wide range of topics related to the theoretical and applied research in the field of Intelligent Data Analysis.

The first two articles are about advanced data preprocessing and data understanding in IDA. The first article by Boukela *et al*. is about an outlier ensemble for unsupervised anomaly detection where the authors demonstrate how honeypots data can be analyzed in an active defense strategy in communication networks. The authors propose how to improve the unsupervised anomaly detection in honeypots data by varying the data feature subset and the parameterization of the anomaly detection algorithm. Their ensemble method outperforms existing solutions where a detection rate higher than 92% is achieved. Alipour and Ansari in the next article of this group introduce an advanced profile hidden Markov model for malware detection. The author's study identifies "species" of malware families, which are more sophisticated, obfuscated, and structurally diverse and propose a hybrid technique combining aspects of signature detection with machine learning based methods to classify malware families. Based on "consensus sequences", their experimental results show that their proposed approach outperforms other HMM-based techniques even when limited training data is available.

The second group of articles in this issue are about unsupervised and supervised learning in IDA. In the first article of this group da Silva et al argue that Fuzzy C Means (FCM) is strongly affected by the selection of the initial cluster centres where the most common selection method is the trial-and-test random approach. The authors propose two methods to obtain the initial cluster centres which are applied to FCM and its variants. The main advantage of these methods is to provide high quality partitions faster than the original methods as well as other FCM and ckMeans-based algorithms with deterministic selection of cluster centres. Callister *et al*. in the second article of this issue discuss stream clustering using a self-controlled connectivity graph approach and argue that clustering algorithms with sensitive parameters are often not robust to such changes, leading to poor clustering outputs. The authors demonstrate the asymptotic linear dependency of skewness excess against the graph connectivity and propose a novel algorithm, with improved robustness against stream evolution. They evaluate their approach on real-world benchmark data sets against previous stream clustering algorithms, and demonstrate that it provides better clustering performance. Trieu *et al*. in the fifth article of this issue introduce an algorithm for hiding high-sensitive utility itemsets with the objective of minimizing the side effects of the sanitization process. The proposed approach exactly identifies both items and transactions for data modification such that it minimizes the impacts on non-sensitive itermsets, data distortions, and the time to access database. Prakash and Pabitha in the next article of this group introduce a hybrid node classification mechanism for influential node prediction in social networks where the idea is to incorporate probability metrics with regression classifier as part of a support vector Bayesian machine. Their experimental evaluation of the proposed system with the existing support vector machine technique has produced very good results denoting that the true positive influential node classification process from the other existing nodes was higher than SVM. Le *et al*. in the seventh article of this group argue that identifying features to represent

graphs such as social networks and protein graphs is increasingly common and becoming more reliable due to the fact that data has increased not only in quantity but also in complexity. The authors also explain that substantial efforts have been made to convert the graphs to an improved representation, among which is graph embedding. The authors propose an approach to retain more edge information while ensuring the embedding graph is still sufficiently small, compared to the original one. Their experiment results show that their proposed method also increases the accuracy of learning models.

And finally the third group of articles are about enabling techniques and innovative application in IDA. The first article of this group by Fazanaro and Pedrini is a comparative analysis of Bayesian network structure learning applied to crime data. The authors explain that the adoption of concepts from exact sciences is a recent movement, originating it a novel research area, called computational criminology, which employs procedures from applied mathematics, statistics and computer science to provide original or enhanced solutions to problems such as crime prediction. Their comparative analysis investigates the employment of statistical inference by means of Bayesian network for predictive policing, using real data from police department. In the next article of this group Kirtania *et al*. introduce an adaptive k-NN classifier for handling imbalance class of problems and its application to clinical data. Their system can effectively handle data with skewed distributions and varying class densities using the concept of adaptive distance. The results of their experiments demonstrate clear superiority over state-of-the-art algorithms. Dong *et al*. argue that since vast amount of data is generated through social networks, it is essential to detect and trace large events and burst topics in mass social network data based on real-time big data parallel computing. The authors propose a model that uses a negative binomial model to fit the distribution of Weibo topic words. They introduce the concepts of the 'hot degree' and the 'dispersion degree' of a topic with their corresponding computing methods. Their experiments on real data demonstrate that their proposal is effective and efficient in tracking bursting events. The eleventh article by Liu *et al*. is about graph matching where the authors argue that multi-constrained graph pattern matching is an NP-complete problem and the fuzziness of constraint variables may exist in many applications. The authors introduce a multi-fuzzy-constrained graph pattern matching problem in big graph data, and propose an efficient first-k algorithm for solving it. The results of their experiments conducted on three datasets of real social networks illustrate that their proposed algorithm significantly out-performs existing methods in efficiency and the introduction of fuzzy constraints makes the proposed algorithm more efficient and effective. And finally in the last article of this issue Wei *et al*. introduce a community-based algorithm for influence maximization on dynamic social networks where the objective is to find the top k influential seeds which can maximize the influence spread. Their experimental results show that their algorithm obtains a better influence spread than many baseline algorithms as well as an acceptable running time while considering the dynamic social networks.

In conclusion, we would like to thank authors who have submitted the results of their excellent research to be evaluated by our referees and published in the IDA journal. This year, we are also working on a special issue which is from the best papers of CIARP-2019 conference that was held in Havana-Cuba, on October 2019. 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