Intelligent Data Analysis 23 (2019) 499–501
DOI 10.3233/IDA-190002
IOS Press

Editorial

Dear Colleague:

Welcome to volume 23(3) of Intelligent Data Analysis (IDA) Journal.

This issue of the IDA journal, is the third issue for our 23rd year of publication. It contains 12 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 in IDA. Chen *et al.* in the first article of this issue discuss the topic of privacy preservation and explain some of the challenges of published trajectory data anonymization. The authors propose a trajectory privacy preservation method based on 3D-Grid partition in order to reduce information loss in the process of trajectory anonymization. The proposed method considers three scenarios of trajectory distribution and measures trajectory similarity based on time, orientation, spatial locations and other features of trajectory. Their theoretical analysis and experimental results show that, compared to other methods, the proposed algorithm effectively preserves trajectory data privacy and improves the anonymous results of trajectory data in terms of accuracy and availability. Moser and Schramm in the second article argue that complexity of data analysis requirements in use of industrial time series data has led to an urgent need to develop effective methods for extracting structural information from data based on the clustering of system behavior time series. The authors present a review of generic approaches in multivariate dynamic time warping (MDTW) to determine the most promising approaches for use in the automotive domain. They report approaches to be particularly useful for tasks such as the objective assessment of subjective driving perceptions.

The second group of articles in this issue are about unsupervised learning in IDA. Diasse and Li in the first article of this group argue that transfer learning could provide machine learning algorithms with the ability to train a model on a given task, capture the existing relationship in the data and reuse it for another task in the same or similar domain. The authors present a multi-view deep unsupervised transfer leaning via joint auto-encoder coupled with dictionary learning, which is based on two stages. The authors also present an extensive set of experiments performed on two real-world datasets which demonstrate the effectiveness of their approach compared with different state of the art baselines. Zhang *et al.* in the next article of this group argue that in the era of big data, the data provided by multiple sources for the same entity may result in conflicting information or eventually outliers from sources with higher or lower reliability. The authors propose a local linear regression method for addressing the problems caused by outliers. Their proposed method can effectively estimate the source reliability and the truth of the datasets with outliers. Their experiments on two real-world datasets demonstrate that the proposed method yields more accurate results than existing state-of-the-art methods.

The third group of articles are about semantic data analysis. Keith Norambuena and Villegas in the fifth article of this issue introduce a novel technique for association rule extraction in text where the objective is to explore the feasibility of applying the approach in the field of opinion mining and sentiment analysis. The proposed method is based on the construction of association rules, which are extended through a similarity criteria for terms represented in a semantic vector space. Their results show that the proposed method is competitive compared to the baseline provided by Naïve Bayes and Support Vector Machines. Along the same line of research, Tajbakhsh and Bageherzadeh in the next article explain

1088-467X/19/$35.00

that topic modeling encompasses a set of techniques for text clustering and tag recommendation with significant advantages such as unsupervised learning. They also emphasize that the main drawback of topic modeling techniques, specifically LDA, lies on their incapability in clustering short texts in which semantic relation between words is neglected. The authors propose a method of topic modelling named Semantic Knowledge LDA that is based on semantic relations between the words in tweets from Twitter social network based on the co-occurrence of words. The article includes the results of applying their approach to a set of real tweets for testing purposes where they record higher performance in terms of precision, recall and F-Score.

The fourth group of articles in this issue are about recommender systems. Huang in the first article of this group explain that personal information management enables users to manage and classify information via social tagging and the emerging social networks generate new concepts for designing modern recommender systems in personal information management and sharing platforms. The authors propose a recommender mechanism for the personal information management and sharing platforms that incorporates tag-based personalized interest and social network relationships into a modified Bayesian probability model. The performance of the proposed system is evaluated based on the word2vec word embedding model. Li *et al.* in the eighth article of this issue explain that tagging information is the common resource to complement implicit feedbacks to assist collaborative filtering recommendation where existing tag-aware recommendation methods still suffer from the problem of high dimension and sparsity of tagging information. The authors propose a novel tag-aware recommendation framework by incorporating tag mapping scheme into ranking-based collaborative filtering model, to boost ranking-oriented personalized recommendation performance. Their experiments on real-world recommendation datasets show that their proposed recommendation method outperforms competing methods on ranking-oriented recommendation performance.

And finally, the last four articles in this issue are about text analysis in IDA. Asghari *et al.* in the ninth article of this issue define cross language plagiarism as the unacknowledged reuse of text across language pairs where it occurs if a passage of text is translated from source language to target language and no proper citation is provided. The authors argue that although various methods have been developed for detection of cross language plagiarism, less attention has been paid to measure and compare their performance, especially when tackling with different types of paraphrasing through translation. The authors have constructed a bilingual plagiarism detection corpus comprised of seven types of obfuscation. Their results show that word embedding approach outperforms the other approaches with respect to recall when encountering heavily paraphrased passages. Zhu *et al.* in the tenth article of this issue explain that with the emergence of short texts, discovering topics within them has become an important task where Biterm Topic Model (BTM) has been considered as more suitable to discover interesting topics. They emphasize that there are still some challenges that dealing short texts with BTM will always ignore the document-topic semantic information and lack the true intentions of users. The authors propose a joint model based on online algorithms of Latent Dirichlet Allocation (LDA) and BTM, which combines the merits of both models. Their experiments on real world datasets show that their proposed method is better than other baseline methods. Jipeng *et al.* in the next article of this issue discuss Dirichlet Multinomial Mixture (DMM) models and their limitations on clustering short texts. The authors propose a novel model by incorporating word-word correlation into DMM, called WDMM. Their experimental results on real-world datasets demonstrates the substantial superiority of WDMM model over the state-of-the-art methods. And finally Lee *et al.* in the last article of this issue discuss data stream clustering and its challenging issues, such as handling limited memory, dealing with evolving clusters, and detecting noise in the data. The authors propose a hybrid data stream clustering method that combines model-based clustering and density-based clustering. The authors apply their proposed method to several synthetic and real datasets where their experimental results demonstrate that the proposed method works effectively for a data stream that includes noise in the data.

In conclusion, we would like to thank all the 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 will have one special issue which is a collection of the best papers from Fuzzy Sets and Data Mining Conference that was held in Bangkok, Thailand in November 2018. 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