Intelligent Data Analysis 21 (2017) 1–3
DOI 10.3233/IDA-160866
IOS Press

Editorial

Dear Colleague:

Welcome to volume 21(1) of Intelligent Data Analysis (IDA) Journal.

This issue of the IDA journal, the first issue of 2017, represents the beginning of our twenty first year of publication and serving the IDA community. This issue contains twelve articles, all covering a wide range of topics related to the theoretical and applied research in the field of Intelligent Data Analysis.

The first three articles of this issue are about various aspects of word sense and text processing in IDA. In the first article of this issue, Huang *et al.* discuss word sense information and argue that in the process of semantic role labelling, word sense could lead to achieve better result to discover semantic role information. The authors introduce a synergetic theory to semantic analysis and propose a semantic analysis method based on synergetic neural network, which can effectively use semantic and word sense information. Their experiments show that their proposed model can further improve the performance of semantic role labelling, and thus provides an important reference value to future research. Starc and Mladenic in the second article of this issue argue that meaning representation that captures all the concepts in the text may not always be available or may not be sufficiently complete. They also argue that ontologies provide a structured and reasoning-capable way to model the content of a collection of texts. The authors present a novel approach for joint learning of ontology and semantic parser from text. The method is based on semi-automatic induction of a context-free grammar from semantically annotated text and has been evaluated on the first sentences of Wikipedia pages describing people. Makrehchi and Kamel in the last article of this group argue that in the majority of text mining tasks, stopwords are removed according to a standard stopword list and/or using high and low document frequencies. The authors propose a new approach for stopword extraction, based on the notion of backward filter level performance and data sparsity index. Based on the proposed model to evaluate the extracted stopwords, the authors examine high document frequency filtering for stopword reduction and propose a new algorithm for building general and domain-specific stopword lists. Their evaluation experiment show that the proposed approach offers more promising results that guarantee a minimum information loss by filtering out most stopwords.

The next four articles are on fuzzy classification and evolutionary computation. Shahparast and Masoori in the first article of this group discuss the complexity of extracting comprehensive rules from high-dimensional data that is normally a serious challenge for designing fuzzy classifiers. The authors propose a feasible approach for extracting rules from high-dimensional data that works in a bottom-up manner. Their approach that is scalable generates all manageable specific rules and then tries to generalize them. Their evaluation involves scalability of the proposed approach on high-dimensional datasets, where they compare it with some related methods. Along the same line of research, Feng and Guo in the next article emphasize that soft set theory offers a general mathematical tool for dealing with uncertainty and present their investigation in the decision making that is based on interval-valued intuitionistic fuzzy soft sets. The authors present a novel approach to solve the problems of decision making and give an example to illustrate the proposed approach, which demonstrates that it is more flexible and effective. Furthermore, the authors present an adjustable approach to weighted interval-valued intuitionistic fuzzy

1088-467X/17/$35.00 © 2017 – IOS Press and the authors. All rights reserved

soft sets based decision making problems by using distance measures. Lu *et al.* in the sixth article of this issue list maximum likelihood (ML) as a method for estimating the parameters of statistical models and emphasize that it has gained much attention and popularly in a variety of fields. However, ML has limitations among which is its optimization problem. The authors first use the ML method to address the statistical modelling of ML parameter estimation of general linear dynamic systems and then present a novel heuristic particle swarm search algorithm which is an incorporation of a sliding mode controller into a standard particle swarm optimization. Their experiments on simulated data demonstrate that the proposed approach is effective and superior to three existing approaches in estimating the ML parameters of the linear dynamic system. Perez-Alonso *et al.* in the last article of this group argue that association rules (ARs) and approximate dependencies (ADs) are significant fields in data mining and the focus of many research efforts. They also argue that the knowledge, extracted by traditional mining algorithms becomes inexact when new data operations are executed, a common problem in real-world applications. The authors propose two algorithms for incremental maintenance of previously discovered ARs and ADs, inspired by efficient computation of changes. Their experimental results on real education data and repository datasets show that their proposed methods achieve a good performance and can significantly improve traditional mining, incremental mining, and a naive approach.

The last group of articles in this issue are on data understanding and novel applications of IDA methods. Song *et al.* in the first article of this group argue that Symbolic Aggregate Approximation (SAX) has been the de facto standard representation method for knowledge discovery in time series data and emphasize that very little work has been done in empirically investigating the intrinsic properties and statistical mechanics in SAX words. The authors propose a new statistical measurement, called information embedding cost (IEC), to analyze the statistical behaviors of the symbolic dynamics. Their experiments on a number of benchmark datasets demonstrate that SAX can always reduce the complexity while preserving the core information embedded in the original time-series with significant embedding efficiency, as well as its robustness to missing values and noise. Shao *et al.* in the ninth article of this issue discuss the idea of maximal information coefficient (MIC), that has been successfully applied in many fields to discover hidden but valuable two-variable relationships, and emphasize that MIC can only work in the case of two-variable relationships. In this article the authors propose an adaptive algorithm, which can deal with not only relationships of pairs of variables, but also multivariable relationships. From the experiments presented in this article it seems that the algorithm is fast and scalable for high-dimensional relationships, therefore, it is suitable for big data. Homayouni and Mansoori in the next article emphasize that the key point in success of an ensemble algorithm is to build a set of diverse classifiers and propose a novel density-based lazy stacking algorithm, suitable for ensemble learning. The approach takes advantage of both lazy learning, in finding local optimal solutions, and the stacking method, in achieving classifier diversity, to obtain better performance while keeping the complexity intact. The performance of the proposed algorithm is compared against four rival classification methods, using some real-world UCI datasets. Their experimental results confirmed that the proposed algorithm significantly outperforms other methods in terms of classification accuracy. The eleventh article of this issue by Martinez *et al.* is about excitation-emission matrices (EEM's) using complex signals in which the authors attempt to predict the standard biosensor-based measurement utilizing EEMs, pH, turbidity and conductivity, among several other variables. The authors have found that nonparametric techniques offer a better mapping from the original signal measurements and in general, the graphical model obtained can be utilized in analyzing two dimensional data. And finally Karimi-Nejad *et al.* in the last article of this issue emphasize that online social network (OSN) users generate massive amounts of information and argue that detection and analysis of the dense sub-structures of networks, called communities, could facilitate a comprehensive understanding of OSNs. The authors also emphasize that most researchers have tackled this issue by comparing results obtained from community detection algorithms with information on available social grouping as a ground-truth. Their study presents a new scoring function that targets the behavior of nodes in order to validate detected communities where they employ this function as a Cluster Validity Index for evaluating detected communities. Performance of the proposed index is compared with other known functions by ranking in terms of several goodness metrics, on a variety of homogeneous networks.

In conclusion, we would like to let you know that this year, as what we have done since 2014, in addition to our six regular issues, we will also publish a special issue related to three highly relevant conferences (The Fifth ASE International Conference on Big Data, The Social Informatics Conference and Technologies and Applications of Artificial Intelligence) that were held in Taiwan. Details will be provided in the next issue of the IDA journal. 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