# Editorial

Dear Colleague:

Welcome to volume 20(1) of Intelligent Data Analysis (IDA) Journal.

With this issue of the IDA journal, the first issue of 2016, we celebrate the 20th anniversary of our journal. This issue consists of eleven 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 are on various forms of supervised and unsupervised learning. Kotsifakos and Athitsos in the first article of this issue argue that many distance or similarity measures have been proposed for time series similarity search. However, none of these measures guarantee an optimal performance when used for a Nearest Neighbor (NN) classification. The authors study the problem of selecting the most appropriate distance measure, given a pool of time series distance measures. The article contains a framework for solving this problem, by identifying, given the query, the distance measure most likely to produce the correct classification results for that query. Based on their experimental evaluation with 45 datasets, they identify the best-performance of the their methods that provides the best results in terms of classification error rate, which include using the Cross Validation method for selecting the distance measure in each dataset, as well as using a single specific distance measure across all datasets. The next article by Li *et al*. is on cluster boundary detection where the authors propose an algorithm that is suitable for mixed attribute data. The approach is based on defining a new objective function according to the data set which is categorized into three collections, i.e. core, exclusion and shadow. In an iterative optimization process, this algorithm extracts its shadow sets from each cluster to form the boundary of clusters. Their experimental results, on both the synthetic data and real data with mixed attributes, numerical attributes and categorical attributes, show that the proposed approach can effectively detect cluster boundary with higher or similar accuracy to its rival methods. The third article of this issue by Perez-Tellez *et al*. is about clustering for weblogs in which the authors present a methodology to cluster weblog posts according to the topics discussed therein, which they derive through text analysis. The proposed approach is based on using Self-Term Expansion methodology to improve the representation of the data and the generative probabilistic model in order to identify relevant topics discussed in the weblogs. The results of this approach have demonstrated a considerable improvement over the pre-defined baseline and alternative state of the art approaches, achieving an improvement of up to 20% in many cases.

The next group of articles are about supervised vector machines. Chen *et al*. in the first article of this group propose a self-training Support Vector Machine (SVM) algorithm, which only needs a small set of labelled samples. The authors also propose a model selection based self-training and the confidence criterion, which are used for searching the best parameter pair of SVM and selecting the most useful unlabelled data to expand the labelled training data set. To demonstrate the validity of re-extracting features and the robustness of features noise, the authors perform an extensive set of experiments which are included in the paper. Ben Ishak, in the fifth article of this issue, discuss the problem of variable selection for linear and nonlinear regression and perform a comparative study between Support Vector Regression and Random Forests for the purpose of variable importance. Their experiments on simulated

and real-world datasets have been carried where they also present the SVR score for variable ranking and the RF score for nonlinear cases. Among the findings in this article is that the RF models are more efficient for selecting variables especially when used with an external score of importance.

The third group of articles in this issue are on optimization and multi-objective learning. Song *et al.* in the first article of this group argue that traditional multi-objective algorithms such as genetic algorithms are slow and the efficiency of these algorithms is also low. The authors propose a multi-objective binary bat algorithm that is based on Pareto approach and is mostly suitable for association rule mining. This algorithm is independent of minimum support and minimum confidence. Compared with the single-objective optimization, binary bat algorithm and Apriori algorithm, their experimental results on six datasets show that the new algorithm is feasible and highly effective. Abbasi Rad and Hamzeh in the next article of this group argue that ranking solutions of the population in an evolutionary algorithm that solves a many objective optimization problem is a challenging task. They propose a method for approximating the ranking induced by the hypervolume indicator. This method is compared to similar methods and its efficiency and performance is evaluated through some experiments. The method is tested using benchmark test problems from the Walking Fish Group of problems which are scalable both in number of variables and objectives. In the eighth of this issue Fattahi *et al.* introduce a Standard Krill Herd algorithm and propose a Fuzzy Krill Herd (FKH) optimization algorithm which is able to dynamically adjust the participation amount of exploration and exploitation by monitoring the progress of solving the problem in each step. In order to evaluate the proposed FKH algorithm, the authors utilize some standard benchmark functions and also choose an inventory control problem. Their experimental results indicate the superiority of their proposed FKH optimization algorithm in comparison with the standard KH optimization algorithm. Zhan *et al.* in the last article of this group argue that the standard Particle Swarm Optimization (PSO) algorithms can easily get trapped in the local optima and has a slow convergence speed. They propose a new optimization method based on Neighbor Heuristic and Gaussian Cloud Learning that is meant to improve the performance of traditional PSO algorithms. Their proposed algorithm consists of two main steps. First analyzing the relationship among particles in the evolutionary process. Second, a Gaussian cloud learning strategy to enhance population diversity and balance the global and local search abilities. The proposed approach is evaluated using 12 benchmark functions and 6 shifted functions where the results show that it is superior to the recent variants of PSO in terms of convergence speed, solution accuracy, algorithm efficiency and robustness.

The last two articles of this issue are on various forms of networking. Rahmani *et al.* in the tenth article of this issue apply a network approach, previously proposed for discovering the relationships among diseases, to analyze drug data. The authors first construct a Human Drug Network (HDN) for 200 different drugs based on functional and structural information available in the PPI network. They then evaluate their proposed HDN using literature to prove that the proposed HDN is biologically meaningful. In the following step, the authors use the HDN to augment the initial prior knowledge of different drugs. It is concluded that their proposed HDN enables one to generate novel hypotheses (in terms of potential drug target proteins) and produce complementary results comparing to existing methods. Finally in the last article of this issue, Hosseini-Pozveh *et al.* propose a sign-aware cascade (SC) model for modelling the effect of both trust and distrust relationships on activation of nodes with positive or negative opinions towards a product in the signed social networks. Since this is an NP-hard problem, a particle swarm optimization (PSO) method is included in the approach which applies the random keys representation technique to convert the continuous search space of the PSO to the discrete search space of the problem. Their experiments demonstrate the effectiveness of the SC in modelling the real-world cascades.

In conclusion, with this issue of the IDA journal, which is Volume 20(1), we are celebrating the 20th anniversary of our journal. The IOS press office, the publisher of the IDA journal, has several plans for

2016. In addition to our six regular issues that now contain 11–12 articles, our plan starting 2014 has been to publish one special issue per year, which is normally related to a scientific conference for which organizers have submitted an interesting proposal. 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