Intelligent Data Analysis 16 (2012) 165–166
DOI 10.3233/IDA-2012-0516
IOS Press

Editorial

Dear Colleague:

Welcome to volume 16(2) of Intelligent Data Analysis Journal.

This issue of the IDA journal consists of ten articles that represent a variety of topics, all related to the applied and theoretical research in the field of Intelligent Data Analysis.

The first three articles are about feature selection and optimization. In the first article, Chen *et al*. discuss the complexity of feature selection in a high dimensional data and propose a particle swarm optimization (PSO) algorithm that is based on the opposite sign test. This particular test increases population diversity in a typical PSO and avoids local optima trapping. Their approach has been evaluated using a number of UCI data sets where the classification accuracy is shown to be higher than genetic algorithms and sequential search. Chuang *et al*. in the second article also discuss the importance of feature selection and propose a PSO approach that is based on a complementary distribution strategy. Their novel approach introduces new complementary particles that replace the selected particles when the global particle fitness is monitored. Their experimental results show that their approach preserves knowledge of good feature selection combinations and thus can find optimum combination of features. Kaedi and Ghasem-Aghaee in the next article of this issue present a Bayesian Optimization approach to overcome the challenge of using past experiences to solve problems based of a case-based reasoning (CBR) approach. The approach consists of obtaining and storing Bayesian networks in specific case-bases where they are retrieved and applied for future cases. The most important advantages of their method is first it is not problem dependent, and second previous solutions are stored using a probabilistic descriptions which makes the use of stored knowledge more flexible.

The next three articles are about unsupervised and supervised learning. Perez-Suarez *et al*. discuss some challenges in proper clustering of data where objects may belong to more than one cluster and introduce a dynamic clustering system. The proposed system, that has a graph-cover strategy for building a set of clusters that normally overlap, has also the capability of dynamically updating clusters through managing multiple additions and or deletion of objects. Their article includes a number of their experiments with their results. Baumgartner and Serpen in the fifth article of this issue present a new design heuristic for hybrid classifier ensembles in machine learning. Their heuristic includes both global and local learners which allow heterogeneous and homogeneous diversity through the co-existence of global and local learners. The article includes evaluation of their approach for hybrid classification using Weka machine learning with a group of 46 data sets from UCI data collection. Johansson *et al*., in the next article of this group, discuss the limitations of ensemble classifications and present a method for improving predictive performance through generating transparent models. The idea is based on using the oracle coaching approach to predict class labels for some applications like production data in conjunction with the original training set. The results of this work is shown to be robust for a variety of methods. It is also shown by the authors that all kinds of performance gains can be obtained by adjusting the relative weights of training and oracle data.

The article by Soroush *et al*. discusses the two most important issues in customer prediction systems, which are proper feature selection and the application of the most suitable prediction methods. They

introduce a new customer prediction system that is based on a hybrid approach and applies a multiple forward stepwise logistic regression model to achieve its goals. Among their applications, the authors used an insurance company data set to evaluate their system where they demonstrate its better prediction capability comparing to some existing techniques. Loslever *et al*. in the eighth article of this issue present a standard five step model for statistical data analysis. Using two data sets, one simulated and the other a real data set, the authors compare use of two different methods to analyze these data sets based the five step model that they introduce. Abdul-Rahman *et al*. in the ninth article of this issue discuss the vital role of preprocessing in classification and propose a framework for selecting discriminatory features from life science class of data prior to classification. The authors explore using a combination of several multivariate filters and particle swarm optimization with support vector machines for their feature reduction and selection processes. They evaluate their results based on a number of performance measures in which they show higher classification accuracy achieved by the proposed framework. And finally Velasquez in the last article of this issue explore the possibility of analyzing website text contents and propose a methodology to extract the main words in static web sites. The key element in this approach is to identify which pages in a web site can attract user attention when browsing is performed. To validate the approach the author uses web data originated in a complex static website from a Chilean bank where the results consisting of a set of key words are tested on a group of real users and the effectiveness of the approach is shown.

In conclusion, with this second issue of Volume 16 of the IDA journal, we continue our efforts on organizing and publishing one of the best journals in this community. There are two special issues planned for this year. Also, the next IDA conference that is usually organized in Europe, will be held in Helsinki, Finland on October 25–27, 2012. For details please refer to http://ida2012.org/. 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