

## Editorial

---

Dear Colleague:

Welcome to volume 14(6) of Intelligent Data Analysis Journal.

This issue of the IDA journal consists of nine articles, all related to the applied and theoretical research in the field of Intelligent Data Analysis. We have also included some high quality survey articles in this issue.

In the first article of this issue, Jiménez et al., emphasize the importance of data mining of non-linear data structures and in particular provide an interesting survey of some recent work in frequent tree pattern mining. They examine some of the most popular tree mining algorithms and compare them in order to identify and highlight their similarities and differences. Azevedo in the next article of this issue discusses contrast sets which are a way to represent group differences and presents an approach to derive rules that describe contrast sets. This approach is based on Fisher Exact Test to identify significant differences across groups. Detail evaluation of this approach is also presented in the article. Peltonen et al., in the third article of this issue introduce a new learning problem called relevant subtask learning which is a variant of multi-task learning and its objective is to build a classifier for a given task of interest, given too few training examples. The idea is also to investigate if certain amount of supplementary data could enrich the learning process. The authors show how to solve the problem and demonstrate a solution to this with logistic regression classifiers. Two algorithms are given by the authors and evaluated using a number of real data sets.

The next article by Latiri and Smaili describe two methods for mining monolingual and bilingual text corpora. Their first approach is based on association rules and it is applied to some text from French newspapers. Their second approach is based on machine translation and statistical language modeling. The authors report on the integration of both methods into a real statistical translation system with which they have performed several experiments. Their results show that inter-lingual triggers achieve better results than the ones reported in the literature. Tsou et al., in the fifth article of this issue introduce an entropy-based non-linear regression tree algorithm that consists of two steps. In step one, a decision tree is established through selecting the splitting variables and in the second step, splitting points are converted into dummy variables that are fitted into a logistic regression model. Their results include a number of case studies that show the advantages of their approach. Li et al., in the next article discuss the application of simulated annealing in ensemble clustering and its sensitivity to its parameter settings. They also emphasize the effect of cooling functions and demonstrate that the clustering results could be invalid if simulated annealing is used without proper understanding of the behavior of cooling functions. Their research presents how different cooling functions may affect the performance of ensemble clustering and this is done on a number of well known data sets.

In the seventh article, Bilgin and Camurcu discuss development of an efficient clustering framework for analyzing high-dimensional data sets. The idea proposed in their work is based on an extending method where the algorithm is integrated into a relationship clustering approach. Since the overall goal was to improve CPU and memory usage, the approach is evaluated using a number of data sets and how the approach can be extended with larger dimensions in data sets. Ng and Dash in the next article report on their investigation of identifying frequent patterns in a continuous stream of transactions. They report

their comparison of the most popular methods, propose a distance based sampling technique and perform an empirical study of the algorithms using synthetic and bench mark datasets. The outcome of their work is a new algorithm that outperforms the existing algorithms. The final article of this issue by Laleh and Abdollahi Azgomi introduce a hybrid fraud scoring and spike detection technique that works on data streams. The approach that differentiates normal, fraud and anomalous links increases the suspicion of normal links with a dynamic global black list. Their proposed approach has several advantages among which is its ability for anomaly detection without supervision and its ability to foil fraudsters which continuously morph their styles to avoid detection.

In conclusion, this is the last issue of Volume 14 of the IDA journal for which the submission rate has grown very rapidly. That is the main reason that the number of articles included in each issue have increased over the last year. We would like to let the readers know about the IDA conference that will be held in Porto-Portugal from October 29–31, 2011. For details, please refer to the following link (<http://www.liaad.up.pt/ida2011/>). We look forward to receiving your feedback along with more and more quality articles in both applied and theoretical research.

With our best wishes,

Dr. A. Famili  
*Editor-in-Chief*