

Editorial

Dear Colleague:

Welcome to volume 15(2) of Intelligent Data Analysis Journal.

This issue of the IDA journal consists of eight articles, all related to the applied and theoretical research in the field of Intelligent Data Analysis.

In the first article of this issue, Rayward-Smith discusses that aggregated data can take several forms. He emphasizes the need to measure the similarity between such aggregated data for which there is a range of possibilities that are listed in the literature. This article proposes a unifying theory that is not only based on known metrics, it also includes some new measures. The new measures are extensively discussed in the paper and their evaluation is presented. Iglesias *et al.* in the next article of this issue present an approach to automatically create the model of an agent behavior based on the observation and analysis of its atomic behaviors. In their approach, observations of an agent behavior are transformed into a sequence of atomic behaviors. They then analyze this stream of information for primary classifications. Their experimental results show that a system based on their approach can efficiently recognize different behaviors in different domains. Pichara and Soto discuss the importance of anomaly detection in large commercial data sets and propose a new semi-supervised algorithm that actively learns to detect anomalies through interaction with an expert user. Their approach consists of three steps, which are Bayes network identification, subspace clustering and a probabilistic active learning. They evaluate their approach using a number of synthetic and real data sets and demonstrate that under noisy data and anomalies presenting regular patterns, their approach correctly identifies relevant anomalies.

In the fourth article of this issue Ordonez and Zhao present a detailed comparison between constrained association rules and decision trees and identify important differences between both in order to predict multiple target attributes. Their article contains extensive experimental evaluation using real medical data. Based on their experiments, they show that association rules, compared to decision trees tend to produce higher confidence, are better for multiple target attributes, perform better for user defined binning and are easier to interpret. Decherchi *et al.* in the next article, discuss the stationary distribution of the data in data mining tasks and propose an operative criterion to verify the stationary assumptions. They apply their method to theoretical and practical predictions of generalization errors where their experimental results confirm the effectiveness of their proposed approach to detect non-stationarity in large data sets. Similarly in the next article Van-Hulse *et al.* emphasize on understanding the data characteristics and its impacts on learning. They present a comprehensive empirical analysis of learning from imbalanced and noisy data which includes the performance of 11 learning algorithms. Their results show that factors such as data set size, class distribution, and noise level all play a critical role in the learning process.

In the seventh article, Chen *et al.* discuss learning to explain credit rating and propose a methodology that is based on learning vector quantization that is targeted for the development of credit rating models. They apply their methods to a real data set of private companies where their learning algorithm is trained and calibrated in a supervised way. Results presented in this article are in the form of a robust and stable

ranking of company credits. The last article of this issue by Engen *et al.* is on exploring discrepancies in one of the KDD cup applications and its dataset that was used for an intrusion detection problem and has been referenced in the literature extensively. The authors in this article report on an empirical investigation where they demonstrate the impact of several methodological differences in public data sets and report on underlying causes of discrepancies.

In conclusion, in the upcoming volumes of the IDA journal for which the submission rate has grown very rapidly, we are planning to have more special issues. 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, important dates and other information please refer to the following web site (<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