

Editorial

Dear Colleague:

Welcome to volume 9(4) of *Intelligent Data Analysis* – An International Journal!

This issue of IDA journal contains five articles covering various aspects of research contributing to: the enhancements of data analysis processes, document classification and association rule mining. All articles deal with topics in applied and theoretical research in intelligent data analysis. In addition, in their experiments and evaluation procedures, these articles deal with diverse domains that should be of interest to many readers in the IDA community.

Hashemi, in the first article of this issue, introduces a new wrapper approach to deal with the problem of detecting the atypical examples or misclassified objects in large data sets. This research covers both outliers and overlapping samples. The article includes some interesting results that demonstrate the approach is by far faster than some methods such as quadratic wrappers. The results are based on a detailed evaluation using 20 benchmark data sets and 5 classifiers. In the second article of this issue, Khoshgoftaar, Seliya and Gao focus on the issue of data characteristics and its effect on classification. They present an approach to properly detect noise in the data based on a set of Boolean rules generated from the original data set. They evaluate their approach using a software measurement data set that contains some known noisy instances. The effectiveness of their approach is empirically demonstrated in which 100% of known noisy instances are detected. Lee and Calvo in the next article describe a new document classification system that is based on using Naïve Bayes and k-Nearest Neighbour methods and present three methods for its scalability. Their methods are evaluated using two data sets (one small and another one large) and show some very promising results.

The last two articles of this issue are on association rule mining. Scheffer propose an approach to maximize expected accuracy in the production of association rules. The approach is based on a fast algorithm that dynamically prunes redundant rules and sections of hypothesis space that cannot contain better solutions. The evaluation is based on its comparison with the Apriori algorithm. Geller, Zhou and Prathipati, in the last article of this issue suggest using concept hierarchies to improve the results of association rule mining. The method is based on changing the structure of the level of granularity of the data so that the data instances are raised to a higher level. This should improve the problem of data sparseness. To deal with the problem of abstractness of the association rules, they propose formulating the notion of an optimal target level. This apparently reduces the computational efforts.

The 6th International Symposium of Intelligent Data Analysis (IDA-2005) will be held in Madrid from September 8–10. Details of conference information are available at the IDA society home page at: <http://www.ida-society.org> or at the conference home page at: <http://www.ida-2005.org>. Plans are underway to publish a special issue of the IDA journal, in 2006. This special issue would include 5 or 6 extended versions of the best papers presented at this symposium. Over the last 10 years, we have benefited from the participation of researchers and practitioners of the IDA field in this important bi-annual event.

With our best wishes,

Dr. A. Famili
Editor-in-Chief