

## Editorial

---

Dear Colleague:

Welcome to volume 6(6) of the journal *Intelligent Data Analysis*!

With volume 6(6) of IDA, that consists of five articles, we celebrate the completion of our sixth year of continuing success. This issue contains five articles from a broad set of applied research papers related to various topics in data analysis.

In the first article, Garcke and Griebel introduce a new classification method based on regularization techniques. Their method computes a nonlinear classifier but scales only linearly with the number of data points and is most suitable for applications where the feature space is quite high. Their paper discusses the complexity of the algorithm and presents some results that are comparable to existing methods. The second article by Counsell et al. discusses the grouping problem in real world applications and explains a method to decompose large number of objects into mutually exclusive subsets where within-group dependencies are high and between group dependencies are low. Their method is based on an evolutionary algorithm which considers the whole population as a solution. Their method is evaluated on the problem of email user allocation which results in 26% reduction in email messages. The third paper, by Moody et al., introduces and evaluates a data-driven technique for identification of rock mineral compositions. The algorithm, presented in this paper, is based on a two stage process in which components in a given rock are identified first and classification accuracies are improved next. Kim, Street and Menczer, in the fourth article, deal with the problem of feature subset selection. Their work is based on a new clustering technique and use of heuristics to estimate the quality of clusters. The evolutionary algorithm introduced in this paper generates a feature subset and a number of clusters which are evaluated. Their experimental results is based on both real and synthetic data and shows their method can consistently find approximate solutions which can help to identify significant features and an appropriate number of clusters in an unsupervised learning application. The last paper of this issue, by Hruschka and Ebecken, is about prediction of missing values in large sets. The authors introduce a Bayesian algorithm that generates a network that is used to substitute missing values in large data sets. Their approach is based on predicting the most probable instance for a given feature in each object of a database. The evaluation of their algorithm is for classification tasks and includes experiments with or without their data pre-processing approach.

And finally, we are working on a special issue of IDA for our volume 7, in 2003. This will be the results of the efforts by our colleagues who organized a workshop held during PKDD-ECML conference in Helsinki, August 2002. We appreciate all the quality papers submitted by our colleagues working in this field and all the support received.

Best wishes,

*A. Famili*  
Editor-in-Chief