



*Special Track on*

## Data Mining

This special track is devoted to showcasing the latest advances in the field of data mining research. Topics of interest include applications, such as intelligence analysis, genomics, bioinformatics and biometrics, medical and health industry, text, video, and multimedia mining, e-commerce, web, financial data analysis, intrusion detection, remote sensing, earth sciences, and astronomy; modeling algorithms such as hidden Markov models, decision trees, neural networks, or statistical methods and probabilistic methods; case studies in areas of application, or over different algorithms and approaches; feature extraction and selection; post-processing techniques such as visualization, summarization, or trending; preprocessing and data reduction; data engineering or warehousing; or other data mining research which is related to artificial intelligence.

This year the special track received many quality submissions and, of these, accepted eight papers for presentation while three additional contributions were referred to the poster session. The majority of submissions dealt with improving well established algorithms, while a fraction of them took on new applications. Topics of papers presented at the conference reflect current trends in data mining, as follows.

New or improved algorithms presented included a knowledge-based approach to association rule filtering; a method of adapting association rule mining to classification by assessing the statistical significance of candidate rules; a novel approach to instance-based classification which addresses two major problems associated with the k-nearest neighbor method: (i) instead of using a fixed global k, the algorithm uses a flexible number of neighbors contributing a vote and (ii) the value of k depends on the local characteristics of the region in which an example resides; two fitness functions (AUCFitness and BREFFitness) to be used with G-REX (a rule extraction technique that uses Genetic Programming), and a new "brevity" comprehensibility measure, with which G-REX obtains higher accuracy than do decision trees; comprehensive computational experiments comparing boosting and sampling techniques for imbalanced data classification; an ensemble approach to improve minority class classification; and a study of different mixture-based imputation methods for collaborative filtering.

In terms of applications, the track included a paper on the design of a classifier for the diagnosis of retinopathy of prematurity.