Data Mining, Knowledge Discovery, and Information Retrieval

H. Michael Chung  
Department of Information Systems  
Center for Information Systems Technologies  
California State University, Long Beach  
hmchung@csulb.edu  
and  
Fredric C. Gey  
Data Survey Research Center  
University of California, Berkeley  
gey@ucdata.berkeley.edu

The field of data mining is rich in research and application development opportunities: first, current data mining approaches are enhanced through cutting edge research from statistics, machine learning, visualization and database management. Specific opportunities include the development of new and hybrid algorithms, and continuing research into mining of new data formats and structures.

Second, it necessary to make data mining tools easier for domain experts to use. The construction of better and more accurate algorithms is not beneficial unless end users are able to configure and run the algorithms conveniently, understand the results, and use the results for the solution of relevant problems.

Third, a costly aspect of data mining includes the pre-processing tasks of selecting, extracting, cleaning and preparing the data prior to data analysis.

Fourth, performance evaluation of data mining approaches is an issue related to algorithm enhancement as well as to usability.

Fifth, from a business and managerial perspective, data mining research needs to consider managerial decision support and organizational decision-making. Business issues also include the impact of the Internet and e-Commerce, which have resulted in the accumulation of vast amounts of transactional and demographic data.

Sixth, Web mining can investigate Web content, structure, and usage.

Seventh, data mining in a distributed environment is a natural extension of the research considering the decentralized and collaborative nature of the work and data source locations.

The selected papers for this year address the above issues: three papers for the algorithms and data organization, one for feature selection, one for decision support, one each for collaborative data mining and Web mining. Finally, issues related to data mining process standards are presented.