

# Data Mining: Where is it Heading?

(Panel)

Jiawei Han

Database Systems Research Laboratory

School of Computing Science

Simon Fraser University, B.C., Canada V5A 1S6

E-mail: han@cs.sfu.ca

## Abstract

*Data mining is a promising field in which research and development activities are flourishing. It is also a young field with vast, unexplored territories. How can we contribute significantly to this fast expanding, multi-disciplinary field? This panel will bring database researchers together to share different views and insights on the issues in the field.*

Data mining has attracted popular interest recently, due to the high demand for transforming huge amounts of data found in databases and other information repositories into useful knowledge. The rapid progress in this field owes to the joint efforts of researchers and developers in data mining, data warehousing, database systems, knowledge-base systems, statistics, machine learning, information retrieval, data visualization, high performance computing, and a number of other related fields.

An important question facing database researchers is “Where is data mining research and development heading?” This panel is going to address this question.

Based on my observation, there are a number of directions which may require more in-depth research in the years to come.

- Integration with data warehousing and on-line analytical processing (OLAP) technology.

A data warehouse may serve as a valuable source of cleaned and integrated data for on-line analytical processing and data mining. OLAP can be viewed as a simple form of mining for aggregated data characteristics. Integrated with OLAP tools, a data miner may drill along any dimension in a data cube to find interesting patterns at multiple levels of abstraction, which will increase the usefulness of both data mining and data warehousing systems.

- Mining multiple kinds of knowledge.

Besides popularly studied classification and association, there are many other kinds of data mining tasks to

be explored, including characterization, comparison, clustering, predictive modeling, time-related pattern analysis, etc.

- Support of data mining query languages and efficient, interactive, ad-hoc data mining.

Similar to relational languages, high-level data mining languages need to be developed to allow users to specify ad-hoc data mining tasks and patterns to be mined, as well as perform interactive or pattern-directed mining flexibly and efficiently.

- Handling increasingly complex data.

Mining relational and transactional data has been our focus. However, mining knowledge from semi-structured and unstructured data, from hypertext, documents, spatial, and multimedia data, as well as from legacy databases, active databases, and the Internet, are all challenging and highly demanding research issues which need to be dealt with.

- High performance data mining.

Efficient and scalable mining algorithms have been our major research focus, and this trend will be further enhanced by developing parallel, distributed, and incremental data mining algorithms.

- Visualization and data mining.

Visualization of database contents and data mining results may help users comprehend or appreciate mining results and redirect miners to search for promising patterns. Easy-to-use and easy-to-see visualization tools will be an asset for interactive mining.

- Applications of data mining.

Several interesting questions arise, such as: “How can we apply the mined knowledge effectively in business management, decision making, and process control? How can we integrate the mined knowledge into a knowledge-base, an expert system, a decision support system, or even a query optimizer?”