DBMiner: Interactive Mining of Multiple-Level Knowledge in Relational Databases*

Jiawei Han Yongjian Fu Wei Wang Jenny Chiang Osmar R. Zaïane Krzysztof Koperski

Data Mining Research Group, Database Systems Research Laboratory
School of Computing Science, Simon Fraser University, British Columbia, Canada V5A 1S6
E-mail: {han, yongjian, weiw, ychiang, zaiane, koperski}@cs.sfu.ca

Abstract

Based on our years-of-research, a data mining system, DBMiner, has been developed for interactive mining of multiple-level knowledge in large relational databases. The system implements a wide spectrum of data mining functions, including generalization, characterization, association, classification, and prediction. By incorporation of several interesting data mining techniques, including attribute-oriented induction, progressive deepening for mining multiple-level rules, and meta-rule guided knowledge mining, the system provides a user-friendly, interactive data mining environment with good performance.

Introduction

A data mining system, DBMiner, has been developed based on our studies of data mining techniques [1, 3, 2] and our experience in the development of an early system prototype, DBLearn [1]. The system integrates data mining techniques with database technologies and discovers knowledge at multiple concept levels from large relational databases efficiently and effectively.

The system has the following distinct features:

1. It incorporates several interesting data mining techniques, including attribute-oriented induction [1], progressive deepening for mining multiple-level rules [2, 3] and meta-rule guided knowledge mining [3], and implements a wide spectrum of data mining functions including generalization, characterization, association, classification, and prediction.
2. It performs interactive data mining at multiple concept levels on any user-specified set of data in a database using an SQL-like Data Mining Query Language, DMQL, or a graphical user interface. Users may interactively set and adjust various thresholds, control a data mining process, perform *roll-up* or *drill-down* at multiple concept levels, and generate different forms of outputs, including generalized relations, generalized feature tables, multiple forms of generalized rules, visual presentation of rules, charts, curves, etc.
3. Efficient implementation techniques have been explored using different data structures, including generalized relations and multiple-dimensional data cubes, and being integrated with relational database techniques. The data mining process may utilize user- or expert-defined concept hierarchies which can be specified flexibly, adjusted dynamically based on data distribution, and generated automatically for numerical attributes.
4. Both UNIX and PC (Windows/NT) versions of the system adopt a client/server architecture. The latter communicates with various commercial database systems in data mining using the ODBC technology.

The system has been tested on several large relational databases with satisfactory performance. Additional data mining functionalities are being designed and will be added incrementally to the system along with the progress of our research. Detailed information about the DBMiner system is accessible with the Internet address: http://db.cs.sfu.ca/DBMiner.

*Research was supported in part by the grant NSERC-A3723 from the Natural Sciences and Engineering Research Council of Canada, the grant NCE:IRIS/PRECARN-HMI-5 from the Networks of Centres of Excellence of Canada, and grants from the MPR Teltech Ltd., and the Hughes Research Laboratories.

References

[1] J. Han, Y. Cai, and N. Cercone. Data-driven discovery of quantitative rules in relational databases. *IEEE Trans. Knowledge and Data Engineering*, 5:29–40, 1993.

[2] J. Han and Y. Fu. Discovery of multiple-level association rules from large databases. In *Proc. VLDB'95*, pp. 420–431, Zurich, Switzerland, Sept. 1995.

[3] J. Han and Y. Fu. Exploration of the power of attribute-oriented induction in data mining. In U.M. Fayyad, *et al.* (eds.), *Advances in Knowledge Discovery and Data Mining*, pp. 399–421. AAAI/MIT Press, 1996.