

Data Mining: Tool for Knowledge Management

A. Srimurugan*

S. Nattar**

The secret of success is to know something nobody else knows.

Aristotle Onassis

INTRODUCTION

Data mining is the extraction of hidden predictive information from large number of databases, which is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of Decision Support Systems (DSS). Data mining tools can answer business questions that traditionally were too time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations. Data mining techniques can be implemented rapidly on existing software and hardware platforms to enhance the values of existing information resources, and can be integrated with new products and systems as they are brought on-line. This paper deals with data mining in knowledge management.

Author's Affiliation: *Librarian & Head, Department of Dr. T. P. M. Library, Madurai Kamaraj University, Madurai., ** Librarian, Central Library, Sri S. Ramasamy Naidu Memorial College, Sattur, TamilNadu, Cell No.Cell: 9786924840.

Reprint Request: A. Srimurugan*Librarian & Head, Department of Dr. T. P. M. Library, Madurai Kamaraj University, Madurai.

(Received on 02.09.08, accepted on 14.11.2008)

WHAT IS DATA MINING?

The term data mining refers loosely to finding relevant information or "discovering knowledge" from a large volume of data. Like discovery in artificial intelligence data mining attempts to discover statistical rules and patterns automatically from data. However, data mining defers from machine learning in that it deals with large volumes of data, stored primarily on disk

DATA MINING COMPRISES OF THREE BASIC UNITS NAMEDLY DATA, INFORMATION AND KNOWLEDGE

Data

Data are any facts, numbers, or text that can be processed by a computer.

Different kinds of data such as:

1. Operational or Transactional data, such as sales, cost, inventory, payroll, and accounting.
2. Non-operational data, such as industry sales, forecast data, and macroeconomic data.
3. Meta data - data about data itself, such as Bibliographical Databases, logical database design or data dictionary definitions.

Information

The patterns, associations, or relationships among all this data can provide information.

Knowledge

When information is put to use appropriately in right situations, by an individual, knowledge evolves. Information can be converted into

knowledge about historical patterns and future trends.

DATA WAREHOUSE

A Data warehouse is a repository of information gathered from multiple sources, stored under a unified schema, at a single site. Once gathered, the data are stored for a long time, permitting access to historical data. Thus, data warehouses provide the user a single consolidated interface to data, making decision-support queries easier to write. Moreover, by accessing information for decision support from a data warehouse, the decision maker ensures that online transaction-processing systems are not affected by the decision-support workload.

EVOLUTION OF DATA MINING

Data mining are the result of a long process of research and products development. The four steps listed below were revolutionary:

1. Data Collection (1960s) - This technology is characterized by its retrospective nature and static data delivery.
2. Data Access (1980s) - This phase is characterized by its retrospective nature and dynamic data delivery at record level.
3. Data Warehousing & Decision Support (1990s) - The main characteristics are retrospective nature and dynamic data delivery at multiple levels.
4. Data Mining (2000s) - The characteristics include prospective nature and proactive information delivery.

The core components of data mining technology have been under development for decades, in research area such as statistics, artificial intelligence, and machine learning.

THE KNOWLEDGE DISCOVERY PROCESS

The knowledge discovery and data mining

(KDD) process can roughly be separated into four steps.

Data Selection

The target subset of data and the attributes of interest are identified by examining the entire raw dataset.

Data Cleaning

Noise and outliers are removed, field values are transformed to common units and some new fields are created by combining existing fields to facilitate analysis. The data is typically put into a relational format, and several tables might be combined in a denormalization steps.

Data Mining

We apply data mining algorithms to extract interesting patterns.

Evaluation

The patterns are presented to end-users in an understandable form, for example through visualization.

The following figure clearly explains the knowledge discovery processes.

WHAT CAN DATA MINING DO?

Data mining can be primarily used by libraries with strong client focus- use and usage, choice of materials, Budget commitments, communication, and marketing of information and promoting library information services. It enables libraries to determine relationship among "internal" factors such as areas of subjects, product positioning, or staff skills, and "external" factors such as economic indicators, comparative performance of various subjects, and customer demographics. With data mining, a librarian could use point - of - reference and reading records of readers preference to achieve targeted promotions based on an individual's information seeking behaviour pattern history.

HOW DOES DATA MINING WORK?

Data mining provides the link between transaction and analytical systems. Data mining