Chapter 59
From Data to Knowledge:
Data Mining

Tri Kurniawan Wijaya
Sekolah Tinggi Teknik Surabaya, Indonesia

ABSTRACT

This chapter will discuss a very useful technique to get (or to mine) a hidden information or knowledge which is lie in our data namely, data mining, which is a powerful and automatic (or semi-automatic) technique. Not only about the concept and theory, this chapter will also discuss about the application and implementation of data mining. Firstly, the authors will talk about data, information, and knowledge, whether they are different or not. After understand the term, they will discuss about what data mining is and what the importance of it. Second, they describe the process of gaining the hidden knowledge, how it is done, from the beginning until presenting the result. The authors will go through it step by step. In the next section, they will discuss about the several different tasks of data mining. In addition, to get a better understanding, the authors will compare data mining with other terminology which closely related so called data warehouse, and OLAP. For the last, but not the least, as stated before, this chapter will tell us about the real implementation of data mining in several different areas.

INTRODUCTION

Data, Information and Knowledge

Before discussed any further about data mining, it is better for us to describe first what are data, information, and knowledge.

According to Palace (1996), Data are any facts, numbers, or text that can be processed by a computer. Recently, there are huge growing amounts of data in different formats and different databases. This includes:

- **Operational or transactional data:** data that are obtained from sales, cost, inventory, payroll, and accounting.

- **Non-operational data:** data are obtained from forecasting, industry, and macro-economic data.
- **Metadata:** that is data about the data itself, such as logical database design or data dictionary definitions

From data we can get information, which is the patterns, associations, or relationships among all the data. For example, from the analysis of point of sale transaction we can obtain information about which products are sold and when.

Furthermore, information can be converted into knowledge about historical patterns and future trends. For instance, the monthly information from the point of sale transaction can be analyzed to understand the consumer buying behavior. As a consequence, the retailer could determine which promotional effort is worth, or which products are going to be advertised more than the others.

What Data Mining Is

In his books, Berson (2000) stated that data mining is an automated process of detecting relevant patterns in database. In his books, Berson gave an interesting example about how finding a pattern could be useful: suppose that there is a pattern that showed that married males with children tend to drive a sports car compared to married males without children then this pattern would indeed valuable and quite surprising for the marketing manager.

When discovering patterns in data, the process itself should be automatic or (usually) semiautomatic. The patterns discovered must be meaningful such that they lead to some advantage. Moreover, useful patterns allow us to make nontrivial predictions on new data. There are two approaches for the expression of a pattern (Witten & Frank, 2005): as a black box whose innards are effectively incomprehensible and as a transparent box whose construction reveals the structure of the pattern. Both, we are assuming, make good predictions.

The difference between them is on their structure, whether their structure represent the pattern that can be examined and analyzed or not. In other words, we interested in patterns that also explain something about the data itself.

In order to understand data mining better, Palace (1996) gave a good example about data mining, data mining software and analysis. There are retailers that used the data mining capacity of Oracle software to analyze buying patterns of their customers. It is found that when men bought diapers on Thursdays and Saturdays, many of them also buy beer. After some analysis, it is revealed that these men typically did their weekly grocery shopping on Saturdays. On Thursdays, however, they only bought a few items. The retailers concluded that these men purchased the beer to have it available for the upcoming weekend. Then the retailers could use this newly discovered information in various ways to increase revenue. For example, we could sell beer and diapers at full price on Thursday as well as move the beer display next to the diaper display.

Data mining is also about solving problems by analyzing data in databases. For example, we want to figure out about the loyalty of our customer in a highly competitive marketplace. A database of customer, along with customer profiles, could be a good starting point. First we could analyze the behavior of the former customers, distinguishing the loyal ones from the customer who likely to switch to another products. Once the distinguished characteristics are found, we could use these characteristics to identify the present customer. This specific group can be targeted for a special treatment. Using the same techniques, we can also identify customers who might be attracted to another product, or another offers.

However, data mining is not a magic. For many years, experts have manually "mined" databases, looking for statistically significant patterns. Data mining is similar with when the analyst uses statistical and machine learning techniques to predict the patterns. Recently, technologies automate the

Related Content

A New Approach for Supervised Dimensionality Reduction

Yinglei Song, Yongzhong Liand Junfeng Qu (2018). *International Journal of Data Warehousing and Mining (pp. 20-37).*

www.irma-international.org/article/a-new-approach-for-supervised-dimensionality-reduction/215004

Classifying Very High-Dimensional Data with Random Forests Built from Small Subspaces

Baoxun Xu, Joshua Zhexue Huang, Graham Williams, Qiang Wangand Yunming Ye (2012). *International Journal of Data Warehousing and Mining (pp. 44-63).*

www.irma-international.org/article/classifying-very-high-dimensional-data/65573

Word Embedding Techniques for Sentiment Analyzers: A Review

Upendar Rao Rayalaand Karthick Seshadri (2021). *New Opportunities for Sentiment Analysis and Information Processing (pp. 233-252).*

www.irma-international.org/chapter/word-embedding-techniques-for-sentiment-analyzers/286914

Mining for Mutually Exclusive Items in Transaction Databases

George Tzanisand Christos Berberidis (2007). *International Journal of Data Warehousing and Mining (pp. 45-59).*

www.irma-international.org/article/mining-mutually-exclusive-items-transaction/1789

Distributed Privacy Preserving Clustering via Homomorphic Secret Sharing and Its Application to (Vertically) Partitioned Spatio-Temporal Data

Can Brochmann Yildizli, Thomas Pedersen, Yucel Saygin, Erkay Savasand Albert Levi (2011). *International Journal of Data Warehousing and Mining (pp. 46-66).*

www.irma-international.org/article/distributed-privacy-preserving-clustering-via/49640