Chapter III
Data Mining Association Rules for Making Knowledgeable Decisions

A.V. Senthil Kumar
CMS College of Science and Commerce, India

R. S. D.Wahidabanu
Govt. College of Engineering, India

ABSTRACT

This chapter describes two techniques used to explore frequent large itemsets in the database. In the first technique called "closed directed graph approach," the algorithm scans the database once making a count on possible 2-itemsets from which only the 2-itemsets with a minimum support are used to form the closed directed graph which explores possible frequent large itemsets in the database. In the second technique, dynamic hashing algorithm, large 3-itemsets are generated at an earlier stage which reduces the size of the transaction database after trimming and the cost of later iterations will be less. Furthermore the authors hope that these techniques help researchers not only to understand about generating frequent large itemsets, but also assist with the understanding of finding association rules among transactions within relational databases.

INTRODUCTION

Recently, with the advent of the vast growth in applications of computers, large amounts of transaction data are stored in databases. This has occurred in all areas of human endeavors, from the mundane (such as supermarket transaction data, credit card usage records, telephone call details, and government statistics) to the more exotic (such as images of astronomical bodies, molecular databases, and medical records). Little wonder, then that interest has grown in the possibility of tapping these data, of extracting from them information that might be of value to the owner of

the database. The discipline concerned with this task has become known as data mining.

Data mining, *the extraction of hidden predictive information from large databases*, 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. 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.

Most companies already collect and refine massive quantities of data. Data mining techniques can be implemented rapidly on existing software and hardware platforms to enhance the value of existing information resources, and can be integrated with new products and systems as they are brought online. When implemented on high performance client/server or parallel processing computers, data mining tools can analyze massive databases to deliver answers to questions such as, "Which clients are most likely to respond to my next promotional mailing, and why?"

Data mining techniques are the result of a long process of research and product development. This evolution began when business data was first stored on computers, continued with improvements in data access, and more recently, generated technologies that allow users to navigate through their data in real time. Data mining takes this evolutionary process beyond retrospective data access and navigation to prospective and proactive information delivery. Data mining is ready for application in the business community because it is supported by three technologies that are now sufficiently mature:

- Massive data collection
- Powerful multiprocessor computers
- Data mining algorithms

Commercial databases are growing at unprecedented rates. A recent META Group survey of data warehouse projects found that 19% of respondents are beyond the 50 gigabyte level, while 59% expect to be there by second quarter of 1996. In some industries, such as retail, these numbers can be much larger. The accompanying need for improved computational engines can now be met in a cost-effective manner with parallel multiprocessor computer technology. Data mining algorithms embody techniques that have existed for at least 10 years, but have only recently been implemented as mature, reliable, understandable tools that consistently outperform older statistical methods.

Data mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner. The relationships and summaries derived through a data mining exercise are often referred to as models or patterns. Examples include linear equations, rules, clusters, graphs, tree structures, and recurrent patterns in time series. Association rules are among the most popular representations for local patterns in data mining.

BACKGROUND

There are quite a few rules that are available for analyzing data transformation for making intelligent decision. The association rule is by far the most useful method in this respect, which is described next.

Association Rule

An association rule is a simple probabilistic statement about the co-occurrence of certain events in

Emotion-Drive Interpretable Fake News Detection
Xiaoyi Ge, Mingshu Zhang, Xu An Wang, Jia Liuand Bin Wei (2022). *International Journal of Data Warehousing and Mining (pp. 1-17).*
www.irma-international.org/article/emotion-drive-interpretable-fake-news-detection/314585