

Chapter 3

Temporal Association Rule Mining in Large Databases

A.V. Senthil Kumar

Hindusthan College of Arts and Science, Bharathiar University, India

Adnan Alrabea

Al Balqa Applied University, Jordan

Pedamallu Chandra Sekhar

New England Biolabs Inc., USA

ABSTRACT

Over the last couple of years, data mining technology has been successfully employed to various business domains and scientific areas. One of the main unresolved problems that arise during the data mining process is treating data that contains temporal information. A thorough understanding of this concept requires that the data should be viewed as a sequence of events. Temporal sequences exist extensively in different areas that include economics, finance, communication, engineering, medicine, weather forecast and so on. This chapter proposes a technique that is developed to explore frequent temporal itemsets in the database. The basic idea of this technique is to first partition the database into sub-databases in light of either common starting time or common ending time. Then for each partition, the proposed technique is used progressively to accumulate the number of occurrences of each candidate 2-itemsets. A Directed graph is built using the support of these candidate 2-itemsets (combined from all the sub-databases) as a result of generating all candidate temporal k-itemsets in the database. The above technique may help researchers not only to understand about generating frequent large temporal itemsets but also helps in understanding of finding temporal association rules among transactions within relational databases.

INTRODUCTION

In recent years, data mining has attracted more attention in database communities because of its wide applicability. Similarly, the digital data acquisition and storage technology made great process

which has resulted in the growth of huge databases. One can find these huge databases in walks of life, from the mundane (includes supermarket transaction data, credit card usage records, telephone call details, government statistics, etc.) to more exotic (includes images of astronomical bodies, molecular databases, medical records, etc.). There were several advancements occurred in the area of

DOI: 10.4018/978-1-60960-067-9.ch003

generation, collection and storage of data. Some of main contributing factors in these advancements include computerization of businesses, scientific and government transactions, advances in data collection tools ranging from scanned text and image platforms to satellite remote sensing systems and cheap storage space. Furthermore, internet is playing a big role as a global information system that flooded us with a tremendous amount of data and information. These databases play vital role in understanding the real-time systems and helps researchers as knowledge repositories to understand and design strategies for future. This explosion of growth in stored data has generated an urgent need for new and novel techniques and automated tools that can intelligently assist us in transforming the vast amounts of data into useful information and knowledge. The discipline that emerged to retrieve and analyze information from these databases is known as *data mining*. Data mining can be defined as a process of extracting patterns from data. It is becoming an increasingly important tool to transform these data into information. Data mining can also be defined as the analysis of 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 goal of data mining is to discover hidden patterns, unexpected trends or other subtle relationships in the data using a combination of techniques from machine learning, statistics and database technologies. This new discipline today finds application in a wide and diverse range of business, scientific and engineering scenarios. For example, large databases of loan applications are available which record different kinds of personal and financial information about the applicants (along with their repayment histories). These databases can be mined for typical patterns leading to defaults which can help determine whether a future loan application must be accepted or rejected. Several terabytes of remote-sensing image data are gathered from satellites around the globe.

Data mining can help reveal potential locations of some (as yet undetected) natural resources or assist in building early warning systems for ecological disasters like oil slicks etc. Other situations where data mining can be of use include analysis of medical records of hospitals in a town to predict, for example, potential outbreaks of infectious diseases, analysis of customer transactions for market research application etc. Srivatsan Laxmanan and Sastry (2006) listed detailed review on wide variety of application areas for data mining in the recent years.

Data mining can be performed on data represented in quantitative, textual, or multimedia forms. It has a flexibility to use a variety of parameters to examine the data. Those parameters include association (patterns where one event is connected to another event, such as purchasing a pen and purchasing paper), sequence or path analysis (patterns where one event leads to another event, such as the birth of a child and purchasing diapers), classification (identification of new patterns, such as coincidences between duct tape purchases and plastic sheeting purchases), clustering (finding and visually documenting groups of previously unknown facts, such as geographic location and brand preferences), and forecasting (discovering patterns from which one can make reasonable predictions regarding future activities, such as the prediction that people who join an athletic club may take exercise classes). 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. The discovery of association relationship among a huge database has been known to be useful in selective marketing, decision analysis, and business management (Hipp, Guntzer, & Nakhaeizadeh, 2000).

An important research issue that extends from the mining of association rules is the discovery of temporal association patterns in temporal databases due to the wide variety of applications on

16 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/temporal-association-rule-mining-large/46890

Related Content

Deploying Data Warehouses in Grids with Efficiency and Availability

Rogério Luís de Carvalho Costa and Pedro Furtado (2010). *Complex Data Warehousing and Knowledge Discovery for Advanced Retrieval Development: Innovative Methods and Applications* (pp. 208-229).

www.irma-international.org/chapter/deploying-data-warehouses-grids-efficiency/39593

Particle Identification Using Light Scattering: A Global Optimization Problem

M. C. Bartholomew-Biggs, Z. Ulanowski and S. Zakovic (2009). *Intelligent Data Analysis: Developing New Methodologies Through Pattern Discovery and Recovery* (pp. 143-160).

www.irma-international.org/chapter/particle-identification-using-light-scattering/24217

A New Approach to Classification of Imbalanced Classes via Atanassov's Intuitionistic Fuzzy Sets

Eulalia Szmidt and Marta Kukier (2009). *Intelligent Data Analysis: Developing New Methodologies Through Pattern Discovery and Recovery* (pp. 85-101).

www.irma-international.org/chapter/new-approach-classification-imbalanced-classes/24213

A Fuzzy Decision Tree Analysis of Traffic Fatalities in the US

Malcolm J. Beynon (2009). *Intelligent Data Analysis: Developing New Methodologies Through Pattern Discovery and Recovery* (pp. 201-217).

www.irma-international.org/chapter/fuzzy-decision-tree-analysis-traffic/24220

Big Data Adoption: A Comparative Study of the Indian Manufacturing and Services Sectors

Hemlata Gangwar (2020). *Optimizing Data and New Methods for Efficient Knowledge Discovery and Information Resources Management: Emerging Research and Opportunities* (pp. 138-171).

www.irma-international.org/chapter/big-data-adoption/255755