IDEA GROUP PUBLISHING **ITB12004**

701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com

This paper appears in the book, *Business Applications and Computational Intelligence* edited by Kevin E. Voges and Nigel K. L. Pope © 2006, Idea Group Inc.

Chapter XI

Visual Data Mining for Discovering Association Rules

Kesaraporn Techapichetvanich, The University of Western Australia, Australia

Amitava Datta, The University of Western Australia, Australia

Abstract

Both visualization and data mining have become important tools in discovering hidden relationships in large data sets, and in extracting useful knowledge and information from large databases. Even though many algorithms for mining association rules have been researched extensively in the past decade, they do not incorporate users in the association-rule mining process. Most of these algorithms generate a large number of association rules, some of which are not practically interesting. This chapter presents a new technique that integrates visualization into the mining association rule process. Users can apply their knowledge and be involved in finding interesting association rules through interactive visualization, after obtaining visual feedback as the algorithm generates association rules. In addition, the users gain insight and deeper understanding of their data sets, as well as control over mining meaningful association rules.

Introduction

In this chapter, we discuss the role of visualization in data analysis and in mining association rules from large databases. We describe reviews of works on visualization techniques in five categories, as well as problems related to visualization. A new visualization technique, called hierarchical dynamic dimensional visualization (HDDV) (Techapichetvanich, Datta, & Owens, 2004) was designed to overcome some of these problems. We present a brief overview of this technique. In addition, we discuss how visual data mining can be performed using the HDDV technique. Our main aim is to design a visualization process for mining association rules of a certain type called *market basket association rules*. This type of association rule is used for analyzing trends and correlations in shopping patterns of customers in supermarkets or other retail outlets. Finally, we discuss future trends in visualization and its application into other research areas related to business data analysis.

Background

Researchers in many disciplines such as science, statistics, finance, medical research, and mathematics have developed a variety of multidimensional visualization techniques to support visual representation of massive data sets. In the business world, managers need tools that help them understand their key business, in order to make quick and precise decisions, and to improve their management strategies. Visualization plays an important role in enabling users to explore and gain insight into their data, through visual or graphical images, rather than textual forms such as spreadsheet or tables. Visualization helps users to extract important information such as trends, correlations, or relationships between the variables or dimensions.

In recent years, various visualization methodologies have been developed to support interpreting and representing characteristics and relationships of large multidimensional data sets. Some research areas focus only on visualization techniques, while some apply visualization techniques to data mining to gain insight into large amounts of data such as databases and data warehouses and to discover trends, patterns, and relationships. The research areas of visualization can be categorized into five groups.

First, *geometric* techniques (Cleveland, 1993; Inselberg & Dimsdale, 1987; Kandogan, 2001; D. Keim, 1996) such as the Scatterplot Matrix, Parallel Coordinates, and Star Coordinates involve geometric transformation and projection of data. For the Scatterplot Matrix, individual variables are arranged along the diagonal of a matrix and each display panel illustrates relationships or correlations between variables. For the Parallel Coordinates technique, the dimensions are represented by parallel vertical lines, which are perpendicular to and uniformly distributed along a horizontal line, rather than by data points plotted on two orthogonal axes. Each variable or dimension is assigned to each parallel axis and each line across the axes represents a data item. The relationship between closed axes or dimensions is easy to perceive. In Star Coordinates, axes emanating from