Chapter 15

Uncovering Actionable Knowledge in Corporate Data with Qualified Association Rules

Nenad Jukic
Loyola University Chicago, USA

Svetlozar Nestorov
University of Chicago, USA

Miguel Velasco
University of Minnesota, USA

Jami Eddington
Oklahoma State University, USA

ABSTRACT

Association rules mining is one of the most successfully applied data mining methods in today’s business settings (e.g. Amazon or Netflix recommendations to customers). Qualified association rules mining is an extension of the association rules data mining method, that uncovers previously unknown correlations that only manifest themselves under certain circumstances (e.g. on a particular day of the week), with the goal of improving action results, e.g. turning an underperforming campaign (spread too thin over the entire audience) into a highly targeted campaign that delivers results. Such correlations have not been easily reachable using standard data mining tools so far. This paper describes the method for straightforward discovery of qualified association rules and demonstrates the use of qualified association rules mining on an actual corporate data set. The data set is a subset of a corporate data warehouse for Sam’s Club, a division of Wal-Mart Stores, INC. The experiments described in this paper illustrate how qualified association rules supplement standard association rules data mining methods and provide additional information which can be used to better target corporate actions.

DOI: 10.4018/978-1-4666-2650-8.ch015
1. INTRODUCTION

Rapid increase in the magnitude of the available and affordable computing power, storage, and memory has enabled corporations and organizations to sustain, and in many cases accelerate, the trend of storing and maintaining ever-increasing quantities of data. One of the main information management challenges faced by corporations today is how to get valuable and actionable information from the massive amounts of data that they own.

A typical organization maintains and uses a number of operational data sources. These operational data sources include databases and other data repositories, which are used to support the organization’s day-to-day operations. A data warehouse is created within an organization as an additional separate data store whose primary purpose is data analysis for the support of management’s decision-making processes. Often, the same fact can have both operational and analytical purposes. For example, data describing that customer A bought product B in store C can be stored in an operational data store for business-process support purposes, such as inventory monitoring or financial transaction record keeping. That same fact can also be stored in a data warehouse where, combined with vast numbers of similar facts accumulated over a time-period, it is used to analyze important trends, such as sales patterns or customer behavior. A typical data warehouse periodically retrieves selected analytically-useful data from the operational data sources (Jukic, 2006). For a more in depth look see Kimball, Ross, Thorndwaite, Mundy, and Becker (2007) or Inmon (2005).

Unfortunately, many organizations often underutilize their already constructed data warehouses (Glassy, 1998; Gorla, 2003). While some information and facts can be gleaned from the data warehouse directly, much more can remain hidden as implicit patterns and trends. On-line analytical processing (OLAP) tools, which are also known as business intelligence (BI) tools, provide analytical users with a user friendly way of retrieving data from data warehouses. These tools perform their primary reporting function well when the criteria for aggregating and presenting data are specified explicitly and ahead of time. However, it is the discovery of information based on implicit and previously unknown patterns that often yields important insights into the business and its customers, and may lead to unlocking the hidden potential of already collected information. Such discoveries require utilization of data mining methods.

Data mining is defined as a process whose objective is to identify valid, novel, potentially useful, and understandable correlations and patterns in existing data using a broad spectrum of formalisms and techniques (Chung & Gray, 1999; Smyth, Pregibon, & Faloutsos, 2002). Even though mining operational databases containing data related to current day-to-day organizational activities can be of limited use in certain situations, the most appropriate and fertile source of data for meaningful and effective data mining is the corporate data warehouse.

One of the most important and successful data mining methods for finding new patterns and correlations is association-rule mining (also known as market-basket analysis). However, standard association-rule method is limited to mining only selected portions of the data warehouse. Seeking to eliminate this restriction, and by doing so significantly improve the insight and the actionability of the discovered knowledge, we developed the concept of qualified association rules which is capable of answering a variety of questions based on the entire set of data stored in the data warehouse. Formal definitions, algorithms, and other technical details that provided the theoretical foundation for qualified association rules method are described in our previous papers (Jukic & Nestorov, 2003; Jukic & Nestorov, 2006). In this paper we provide a clear illustration of how this method can enable organizations to get more value out the data they already own. We demonstrate the capabilities of