Chapter XI

Toward Integrating Data Warehousing with Data Mining Techniques

Rokia Missaoui
Université du Québec en Outaouais, Canada

Ganaël Jatteau
Université du Québec en Outaouais, Canada

Ameur Boujenoui
University of Ottawa, Canada

Sami Naouali
Université du Québec en Outaouais, Canada

Abstract

In this chapter, we present alternatives for coupling data warehousing and data mining techniques so that they can benefit from each other’s advances for the ultimate objective of efficiently providing a flexible answer to data mining queries addressed either to a bidimensional (relational) or a multidimensional database. In particular, we investigate two techniques: (1) the first one exploits concept lattices for generating frequent closed itemsets, clusters and association rules from multidimensional data, and (2) the second one defines new operators similar in spirit to
online analytical processing (OLAP) techniques to allow “data mining on demand” (i.e., data mining according to user’s needs and perspectives). The implementation of OLAP-like techniques relies on three operations on lattices, namely selection, projection and assembly. A detailed running example serves to illustrate the scope and benefits of the proposed techniques.

Introduction

Data mining (DM) is the process of discovering hidden knowledge (i.e., patterns and associations) from large data sets while data warehousing (DW) aims at integrating and aggregating data from multiple data sources for further analysis (Chaudhuri & Dayal, 1997; Han & Kamber, 2000). The two technologies present some common features such as (1) information/knowledge extraction from very large data sets, (2) support for decision making, (3) use of background knowledge for additional information (knowledge) extraction, and (4) need for a careful and generally time-consuming data preprocessing step.

There are many topics that have attracted researchers in the area of data warehousing: data warehouse design and multidimensional modeling, efficient cube computation, query optimization, discovery-driven exploration of cubes, data mining in cubes, and so on. In order to avoid computing a whole data cube, many studies have focused on iceberg cube calculation (Xin, Han, Li, & Wah, 2003), semantic summarization of cubes (Lakshmanan, Pei, & Zhao, 2002), and approximation of cube computation (Shanmugasundaram, Fayyad, & Bradley, 1999). Recently, there is an increasing interest for applying/adapting data mining techniques and advanced statistical analysis (e.g., cluster analysis, principal component analysis, log-linear modeling) for knowledge discovery (Ben Messaoud, Boussaïd, & Rabaséda, 2004; Lu, Feng, & Han, 2000; Sarawagi, Agrawal, & Megiddo, 1998) and data compression or query approximation purposes in data cubes (Babcock, Chaudhuri, & Das, 2003; Barbara & Wu, 2001).

The objective of this chapter is to propose techniques for reinforcing the collaboration and linkage between DW and DM techniques by using formal concept analysis and concept lattices (Ganter & Wille, 1999) as a sound and theoretical framework for data mining. More precisely, we first present our view of rule mining in data cubes. Then, we adapt the interactive exploratory mechanisms inherent to online analytical processing (OLAP) techniques to the framework of data mining tools and techniques in order to help the user select the appropriate subset of an already existing data mining output. To conduct the first task, we discuss association rule mining in multidimensional data and show how cube clustering using concept lattices and frequent closed itemsets can be exploited for generating meaningful association
Related Content

Data Mining and Homeland Security
[www.igi-global.com/chapter/data-mining-homeland-security/7853?camid=4v1a](www.igi-global.com/chapter/data-mining-homeland-security/7853?camid=4v1a)

Acquiring Semantic Sibling Associations from Web Documents
[www.igi-global.com/chapter/acquiring-semantic-sibling-associations-web/7744?camid=4v1a](www.igi-global.com/chapter/acquiring-semantic-sibling-associations-web/7744?camid=4v1a)

Navigation Rules for Exploring Large Multidimensional Data Cubes
[www.igi-global.com/chapter/navigation-rules-exploring-large-multidimensional/7702?camid=4v1a](www.igi-global.com/chapter/navigation-rules-exploring-large-multidimensional/7702?camid=4v1a)
Content-Based Image Retrieval

www.igi-global.com/chapter/content-based-image-retrieval/10595?camid=4v1a