
IGI PUBLISHING **ITB15218**
701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.igi-global.com

This paper appears in the publication, **Data Mining and Knowledge Discovery Technologies** edited by D. Taniar © 2008, IGI Global

Chapter I

OLEMAR: An Online Environment for Mining Association Rules in Multidimensional Data

Riadh Ben Messaoud, University of Lyon 2, France
Sabine Loudcher Rabaséda, University of Lyon 2, France
Rokia Missaoui, University of Québec, Canada
Omar Boussaid, University of Lyon 2, France

Abstract

Data warehouses and OLAP (online analytical processing) provide tools to explore and navigate through data cubes in order to extract interesting information under different perspectives and levels of granularity. Nevertheless, OLAP techniques do not allow the identification of relationships, groupings, or exceptions that could hold in a data cube. To that end, we propose to enrich OLAP techniques with data mining facilities to benefit from the capabilities they offer. In this chapter, we propose an online environment for mining association rules in data cubes. Our environment called OLEMAR (online environment for mining association rules), is designed to extract associations from multidimensional data. It allows the extraction of inter-dimensional association rules from data cubes according to a sum-based aggregate measure, a more general indicator than aggregate values provided by the traditional COUNT measure. In our approach, OLAP users are able to drive a mining process guided by a meta-rule, which meets their analysis objectives. In

addition, the environment is based on a formalization, which exploits aggregate measures to revisit the definition of the support and the confidence of discovered rules. This formalization also helps evaluate the interestingness of association rules according to two additional quality measures: lift and loevinger. Furthermore, in order to focus on the discovered associations and validate them, we provide a visual representation based on the graphic semiology principles. Such a representation consists in a graphic encoding of frequent patterns and association rules in the same multidimensional space as the one associated with the mined data cube. We have developed our approach as a component in a general online analysis platform called Miningcubes according to an Apriori-like algorithm, which helps extract inter-dimensional association rules directly from materialized multidimensional structures of data. In order to illustrate the effectiveness and the efficiency of our proposal, we analyze a real-life case study about breast cancer data and conduct performance experimentation of the mining process.

Introduction

Data warehousing and OLAP (online analytical processing) technologies have gained a widespread acceptance since the 90's as a support for decision-making. A data warehouse is a collection of subject-oriented, integrated, consolidated, time-varying, and non-volatile data (Kimball, 1996; Inmon, 1996). It is manipulated through OLAP tools, which offer visualization and navigation mechanisms of multidimensional data views commonly called *data cubes*.

A data cube is a multidimensional representation used to view data in a warehouse (Chaudhuri & Dayal, 1997). The data cube contains *facts* or *cells* that have *measures*, which are values based on a set of dimensions where each dimension usually consists of a set of categorical descriptors called *attributes* or *members*. Consider for example a *sales* application where the dimensions of interest may include, *costumer*, *product*, *location*, and *time*. If the measure of interest in this application is the *sales amount*, then an OLAP fact represents the sales measure corresponding to a single member in the considered dimensions. A dimension may be organized into a hierarchy. For instance, the location dimension may form the hierarchy *city* → *state* → *region*. Such dimension hierarchies allow different levels of granularity in the data warehouse. For example, a *region* corresponds to a high level of granularity whereas a *city* corresponds to a lower level. Classical aggregation in OLAP considers the process of summarizing data values by moving from a hierarchical level of a dimension to a higher one. Typically, additive data are suitable for simple computation according to aggregation functions (SUM, AVERAGE, MAX, MIN, and COUNT). For example, according to such a computation, a user may observe the sum of sales of products according to year and region.