Explanation in OLAP Data Cubes

Rahhal Errattahi, Faculty of Science and Technology, Sultan Moulay Slimane University, Beni Mellal, Morocco
Mohammed Fakir, Faculty of Science and Technology, Sultan Moulay Slimane University, Beni Mellal, Morocco
Fatima Zahra Salmam, Faculty of Science and Technology, Sultan Moulay Slimane University, Beni Mellal, Morocco

ABSTRACT

OLAP is an important technology that offers a fast and interactive data navigation, it also provides tools to explore data cubes in order to extract interesting information from a multidimensional data structures. However, the OLAP exploration is done manually, without tools that could automatically extract relevant information from the cube. In addition, OLAP is not capable of explaining relationships that could exist within data. This paper presents a new approach to coupling between data mining and online analytical processing. Its approach provides the explanation in OLAP data cubes by using the association rules between the inter-dimensional predicates. The mining process could be done by one of the two algorithms, Apriori and Fp-Growth, in which aggregate measures to calculate support and confidence are exploited. It also evaluates the interestingness of mined association rules according to the Lift criteria.

Keywords: Apriori, Association Rules, Data Mining, Fp-Growth, Inter-Dimensional Predicates, OLAP

1. INTRODUCTION

Online Analysis Processing (OLAP) has been recognized as an effective solution for helping in the decision-making process. OLAP is a technology for building multidimensional data structure called “cube”, it help the decision makers to explore data in a multidimensional way, and browse the aggregates according to the different hierarchical levels of each dimensions in order to extract relevant information depending on the level of granularity chosen. However, this exploration done manually by the user, without tools to automatically extracts information and explains the relationships and associations between data in the cube. So the user must have an expertise of the domain representing the subject of the cube, to be able to extract—manually—the knowledge encapsulated in the data.

For example: The height laptop sales in some cities can be simply explained by its association with a category of customers relatively young with high educational level. This information couldn’t be directly concluded from the cube exploration, it requires an association rules algorithm which belongs to the data mining algorithms category.

In the recent years, several studies addressed the issue of mining association rules DOI: 10.4018/jitr.2014100105
from data cubes. Kamber et al. (1997), noting that the multidimensional data structure is favorable for mining association rules, Imielinski et al. (1996) and Goil and Chnoudhay (1999) defend the same point of view and considering that the data mining techniques coupled with an OLAP data structure, can enhance and make the BI system more powerful and easy.

In this paper we present a new approach for the explanation of multidimensional data with association rules, the proposed coupling between the on-line analysis and data mining is inspired from the works of Ben Messaoud et al. (2006) except that we will change the structure of multidimensional data in a simple table (bi-dimensional), but at the same time we introduce some changes in the algorithms that we will use for the association rules mining, such that they support measures to calculate the support and confidence.

In this context the user has the ability to guide the process of mining association rules, to supervise and specify the analysis contexts that need explanation and from which the association rules will be mined, by setting the context of extraction.

Our proposition is based on the approach which consists to change the multi-dimensional data to a simple data structure that can be used by the data mining algorithms. Thus, we introduce two algorithms for mining association rules, these algorithms are based on Apriori and Fp-Growth.

This article is organized as follows. In Section 2 we present definitions and the general notation describing the structure of the data cubes. Section 3 presents the important propositions in the problem of mining association rules from a multidimensional data.

The fourth section present the subject of our approach coupling online analysis and data mining in terms of explanation by inter-dimensional association rules, by using algorithms based on Apriori and Fp-Growth, this algorithms employ the measures to calculate the support and confidence.

In Section 5 we present our Platform Cube Explanation implemented in Java for the explanation of multidimensional data in OLAP cubes. Finally, Section 6 concludes this document by presenting a general overview of the work and proposing some future research directions.

2. GENERAL NOTATIONS

Let C be a data cube with \( d (d \neq 0) \) dimensions \( D = \{D^1, ..., D^d\} \), and a non empty set of measures \( M = \{M_1, ..., M_m\} \). \( H_j^i \) is the \( j^{th} \) hierarchical level belong to the dimension \( D^i \). For example in the Figure 1 the dimension store \( D^1 \) contain three hierarchical level: \( H^i_1 \) store country, \( H^i_2 \) store state and \( H^i_3 \) store city.

The set \( A^i_j = \{a^i_1, ..., a^i_n\} \) represent the modalities or the members of the level \( H_j^i \) belong to \( D^i \). In the same example the hierarchical level \( H^i_1 \) store Country contain three members: Canada \( a^i_1 \), Mexico \( a^i_2 \) and USA \( a^i_3 \).

Let \( D' \in D \) a sub-set non empty of \( p \) dimensions \( \{D^1, ..., D^p\} \) from the data cube \( C(p \leq d) \). The p-tuple \( (e^1, ..., e^p) \) defines a sub-cube on \( C \) according to \( D' \) if \( \forall i \in \{1, ..., p\} \), \( e^i \neq \emptyset \) and there exists a unique \( j \leq 0 \) such that \( e^i_1 \in D^i \). The aggregator \( SUM \) of the measure \( M \) in the sub-cube \( (e^1, ..., e^p) \) noted \( SUM_M(e^1, ..., e^p) \) is the \( SUM \) of measure \( M \) of all facts in the sub-cube.

We define a dimensional predicate in \( D^i \), noted \( \hat{a}^i \), as the predicate \( \hat{a}^i \in A^i_j \) which takes a dimension member as a value. We also define \( (\hat{a}^1 \land ... \land \hat{a}^p) \) as an inter-dimensional predicate in \( D' \) if \( \forall i \in \{1, ..., p\} \), \( \hat{a}^i \) is a dimension predicate in \( D^i \).
Digital Divide: A Glance at the Problem in Moldova
www.igi-global.com/chapter/digital-divide-glance-problem-moldova/22833?camid=4v1a

Corporate Collapse and IT Governance within the Australian Airlines Industry
www.igi-global.com/article/corporate-collapse-governance-within-australian/44599?camid=4v1a