

Chapter 16

An Architecture for Query Optimization Using Association Rule Mining

Sikha Bagui

University of West Florida, USA

Mohammad Islam

University of West Florida, USA

Subhash Bagui

University of West Florida, USA

ABSTRACT

This research presents a way to identify attribute-value relationships already existing in a database by using association rule mining to optimize query processing. Once relationships have been determined, these relationships can be used as a basis for creating temporary structures like views to optimize query operations. This paper presents an architecture that shows how table partitions in the form of views, created based on association rules, can be used to optimize queries. The results of this study were statistically significant.

INTRODUCTION

Query optimization has been an active topic of research for the past three decades, but as databases get larger and as more data and new kinds of data get shifted to data warehouses, new approaches have to be found for query optimization. Even though the query optimizer is one of the most complex decision making systems, it lacks

the functional capabilities that are based on the semantic knowledge stored within a database system to optimize queries. This paper presents an architecture that shows how data mining techniques like association rule mining can aid in uncovering functional dependencies, based on the semantic knowledge of the database, which can be used to partition data into views. This view architecture can then be used in the query optimization process.

DOI: 10.4018/978-1-4666-1873-2.ch016

One of the most important concepts in any relational schema is that of functional dependencies. A functional dependency is a property of the semantics or meaning of the attributes. Functional dependencies have been used as a basis for query optimization by many (Simmen et al., 1996; Huhtala, 1998; Gryz et al. 2001, Godfrey, 2001). A functional dependency, denoted by $X \rightarrow Y$, is a constraint between two sets of attributes in a database such that X and Y are subsets of some universal relation, R . This means that the values of the X component functionally determine the values of the Y component.

Association rule mining is a data mining procedure that is used to find patterns for recurring relationships that appear in a data set. Association rules operate at a higher level of granularity than functional dependencies in that association rules find functional dependencies between attribute-values (and not just attributes) in datasets, and also have statistical measures such as support, confidence and lift measures assigned to the attribute-value relationships.

We mostly think of databases, the “regular” data models, as relational or object-oriented, centered on the concept of a set of attribute-attribute functional dependencies with a set of keys and indexed fields. These concepts aid in the query optimization process. But as increasing amounts of data from different areas are being stored, flat-tabular formats are being used more and more rather than the “regular” data models, hence a lot of the “regular” concepts of the data models may not always hold, for example, indexed fields may be lacking. Querying such databases is a challenge and will of course take longer. This is where newer techniques, like the one presented in this paper, are needed for query optimization.

Moreover, in the regular data models, relational and object-oriented, functional dependencies are attribute-attribute relationships, but association rules will help us break down the dependencies to the level of attribute-value ranges, a finer level of granularity. These attribute-value functional

dependencies are what are developed into association rules. So, association rules will determine new relationships within the data, and views can be based on these new relationships to optimize query processing.

It is the goal of this research to find association rules within a database, and use these association rules as a basis to develop views which can be used in query optimization. These association rule based view partitions will have several advantages: (1) a query will not have to go through the complete table/tables to be processed (only the necessary views), hence speeding up query processing; (2) Since the views are based on association rules, rather than arbitrary table partitions, the query process can be simplified -- meaning that equivalent queries and superset queries can be used; (3) Since views are being used, the data will always be up-to-date; (4) they will aid in optimizing query processing even in databases that are not in the “regular” data models; (5) they will be more useful in less changing databases, for example, in data warehouses that are not so fast changing.

The rest of the paper is organized as follows. First we discuss related works in this area. We briefly explain association rule mining. Then we present a formal presentation of association rule based views and the experimental results. We discuss the statistical significance of the results and present the conclusion and plans for future work.

RELATED WORK

Query optimizers have traditionally drawn upon many sources of information about the database to optimize queries. Integrity constraints, though actually designed to maintain the integrity of a database, have been used in various ways in query optimization (Yu & Sun, 1989; Godfrey, 2001). Since the notion of semantic query optimization was introduced by King (1981), rules have been extracted from semantic information by many (Yu & Sun, 1989; Chakravarthy et al., 1990;

22 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/architecture-query-optimization-using-association/67729

Related Content

Examining Consumer Loyalty, Internet of Things (IoT), and Theme Restaurants in the Delhi Region: An Empirical Study

Pooja Ahuja and Pankaj Tiwari (2020). *International Journal of Knowledge-Based Organizations* (pp. 1-12). www.irma-international.org/article/examining-consumer-loyalty-internet-of-things-iot-and-theme-restaurants-in-the-delhi-region/263033

Applying Sense-Making Methodology to Design Knowledge Management Practices

Bonnie Wai-yi Cheuk and Brenda Dervin (2010). *Ubiquitous Developments in Knowledge Management: Integrations and Trends* (pp. 198-214). www.irma-international.org/chapter/applying-sense-making-methodology-design/41864

Enterprise Data Warehouse Governance Best Practices

Nayem Rahman (2016). *International Journal of Knowledge-Based Organizations* (pp. 21-37). www.irma-international.org/article/enterprise-data-warehouse-governance-best-practices/148228

Knowledge Creation, Management, and Dissemination in Impact Communities

Ana Melro and Lídia Oliveira (2018). *Handbook of Research on Knowledge Management for Contemporary Business Environments* (pp. 22-36). www.irma-international.org/chapter/knowledge-creation-management-and-dissemination-in-impact-communities/205067

How E-buses Took off in Latin America to Save Lives, CO2 and Money

Manuel F. Olivera (2022). *Cases on Applying Knowledge Economy Principles for Economic Growth in Developing Nations* (pp. 261-283). www.irma-international.org/chapter/how-e-buses-took-off-in-latin-america-to-save-lives-co2-and-money/296231