SOCQET: Semantic OLAP with Compressed Cube and Summarization

Laks V.S. Lakshmanan† Jian Pei‡ Yan Zhao†
† University of British Columbia, Canada. Email: {laks, yzhao}@cs.ubc.ca
‡ State University of New York at Buffalo, USA. Email: jianpei@cse.buffalo.edu

1. INTRODUCTION

Data warehouses form the essential infrastructure for many data analysis tasks. A core operation in a data warehouse is the construction of a *data cube*, which can be viewed as a multi-level, multi-dimensional database with aggregate data at multiple granularity.

In a data cube, two basic semantic relations among cells are *roll up* and *drill down*. A cell c_1 can be *rolled up* from cell c_2, and c_2 can be *drilled down* from cell c_1, if c_1 generalizes c_2 in some dimensions, that is, in all dimensions where c_1 and c_2 have different values, c_1 has values "all".

A data cube facilitates the online analytical processing (OLAP) substantially. However, there are some inherent problems that the current techniques cannot handle well. First, while the essential rolling up/drilling down semantics are usually kept in a data cube, many kinds of critical semantics are not highlighted. Second, there is no support of semantic navigation of data in data cubes. It is hard for a user to understand the data in a huge cube space with various hierarchies. Third, there is no semantic compression of data cubes. Almost all approaches proposed previously are *syntactic*. In summary, we need compressed cube representations preserving the cube semantics.

In [1, 2], we developed a quotient cube-based data warehousing and OLAP approach. The ideas are summarized as follows. For many applications, the semantics of an aggregate cell c in a data cube can be defined as the set of tuples in the base table that can be rolled up to c. Intuitively, we can "*summarize*" the cells into a "*class*" if they carry identical semantics, and thus reduce the data cube cell lattice into a *quotient lattice*. We showed that the quotient lattice is substantially smaller than the original cube cell lattice.

We illustrated that the semantic summarization approach can solve the problems identified above. First, in a quotient cube, we store not only the roll up and drill down semantics about cells, but also the summarization of semantics of cells and the relation among classes. Second, a user can conduct OLAP operations on semantic classes, which is more effective than those on cells. A user can navigate the data cube by classes. Moreover, she can drill down into classes and investigate the internal structure of a class. Figure 1 illustrates a drill-down *into* class C_3 in a quotient cube. Last, since all cells in a class carry the same semantics, a semantic compression can be achieved accordingly.

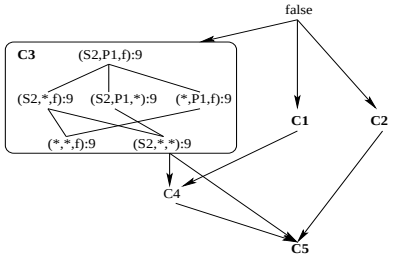

Figure 1: Drill-down into the internal structure of a class.

2. ABOUT THE SOCQET DEMO

The semantic approach brings significant improvement to the effectiveness and efficiency to data warehousing and OLAP. We are developing SOCQET, a systematic approach for effective and efficient semantic summarization for data warehousing and OLAP. Our demo has four major parts.

First, we will present the techniques to materialize quotient cubes using examples. We will analyze why such a materialization method is effective and efficient. We will also illustrate the storage efficiency of the storage techniques using real data sets.

Second, we will demonstrate how various queries can be answered using a materialized quotient cube. Examples and experiments will be used to illustrate the costs of query-answering.

Third, we will present a set of extensive performance studies on the proposed techniques and related methods proposed previously. The experimental results on both real and synthetic data sets will indicate the benefits of the new techniques.

Last, we will showcase a prototype quotient cube-based data warehousing and OLAP system, including a quotient cube engine and an interactive user interface. In particular, we will demonstrate how a quotient cube facilitates the interactive exploration and visualization of a data cube. The audience will be encouraged to play with the demo and experience the exciting tour using semantic navigation services.

3. REFERENCES

[1] L. Lakshmanan, J. Pei, and J. Han. Quotient cube: How to summarize the semantics of a data cube. In *VLDB'02*, Hong Kong, China, Aug. 2002.

[2] L.V.S. Lashmanan, J. Pei, and Y. Zhao. Qc-trees: An efficient summary structure for semantic OLAP. In *SIGMOD'03*, San Diego, CA, June 2003.

SIGMOD 2003, June 9-12, 2003, San Diego, CA.