i^3: Intelligent, Interactive Investigaton of OLAP data cubes

Sunita Sarawagi Gayatri Sathe
School of Information Technology, IIT Bombay, India.
{sunita,gayatri}@it.iitb.ernet.in

The goal of the i^3(eye cube) project is to enhance multidimensional database products with a suite of advanced operators to automate data analysis tasks that are currently handled through manual exploration. Most OLAP products are rather simplistic and rely heavily on the user's intuition to manually drive the discovery process. Such ad hoc user-driven exploration gets tedious and error-prone as data dimensionality and size increases. We first investigated how and why analysts currently explore the data cube and then automated them using advanced operators that can be invoked interactively like existing simple operators.

Our proposed suite of extensions appear in the form of a toolkit attached with a OLAP product. At this demo we will present three such operators: DIFF, RELAX and INFORM with illustrations from real-life datasets.

The DIFF operator

One reason why analysts manually drill down to explore detailed data is to find causes for drops or increases observed at an aggregated level. For automating this task we propose the DIFF [Sar99] operator that can report a summarized difference between two values observed at aggregate levels. For example, a busy executive looking at the annual reports might quickly wish to find the main reasons why sales dropped from the third to the fourth quarter in a region. Instead of digging through heaps of data manually, he could invoke the new DIFF operator which in a single step will do all the digging for him and return the main reasons in a compact form that he can easily assimilate.

The RELAX operator

The RELAX operator is for automatically generalizing the scope of a problem case observed at detailed level. Suppose a local branch manager notices a big drop in sales somewhere in detailed data. Often the next step is to check if this was an isolated case or is this drop part of a bigger problem. For this he rolls up to the next level and views the problem case in the context of combinations of other dimensions using a succession of selection, drill-down and pivot steps. The RELAX operator can be used to automate this search. The output is a set of maximal regions around the problem case with similar changes.

The INFORM operator

Here our goal is to provide an enhanced data exploration environment that is adaptive to a user's prior knowledge of the data [Sar00]. We capture the analysts context by continuously tracking the parts of the cube that a user has visited either through implicit monitoring or by explicit recording. The information in the scattered visited sections of the cube are pieced together to form a model of the user's mental picture of the unvisited parts. Finally, at any time, the user can invoke the INFORM operator that will output the most informative unvisited parts of the cube given his established context. This process of updating the user's context based on visited parts and querying for regions to explore further continues in a loop until the user's mental model perfectly matches the actual cube.

Integrating with OLAP products For each operator we designed fast algorithms that can be invoked *interactively* during cube exploration. They are tightly integrated with existing OLAP systems and ride on the servers' query processing capabilities through dynamically generated queries for all heavy duty processing.

References

[Sar99] S. Sarawagi. Explaining differences in multidimensional aggregates. In *Proc. of the 25th Int'l Conference on Very Large Databases (VLDB)*, 1999.

[Sar00] S. Sarawagi. User adaptive exploration of olap data cubes. Submitted for publication: http://www.it.iitb.ernet.in/~sunita, 2000.