

User-Adaptive Exploration of Multidimensional Data

20062178 Jongwuk Lee

October 31, 2007

Summary

This paper proposes an effective OLAP exploration method based on user's prior knowledge of the data. In particular, this paper presents three scenarios to illustrate the working of the system: (1) The framework recommends informative drill-down paths to the users. Specifically, suppose that users start the process of understanding the data by navigating through subsets of the data cube, *i.e.*, drill-down operator. In such case, the system finds the most surprising drill-down operator which is the most different from the expected values based on users' current knowledge. (2) The framework finds significant information in new data. When inserting new data in the data cube, the system retrieves the helpful content based on users' knowledge of previous data. This is helpful to understand distinct aspects of new data. (3) The framework issues queries based on ad hoc contexts. Among current information, the system finds typical/atypical data in terms of the users' expected value. To address all proposed scenarios, this paper leverages *entropy* as a tool measuring users' expected information. Based on the entropy, this paper introduce *the principle of maximum entropy* to analyze informative contents for current knowledge. This method is very intuitive and helpful to define users' knowledge. The experimental results are also meaningful for effectiveness and efficiency.

Comments

The purpose of this paper is definitely important, but it is not easy to solve this challenge. In such sense, this paper proposes an very meaningful approach using traditional well-formulated method, *i.e.*, *the principle of maximum entropy*. In fact, this method has been applied for finding the most reasonable answer under constrained problem setting. This paper considers this constraints as user's prior knowledge, and finds the most informative contents by comparing entropy between expected value and actual value. Also, the each proposed scenarios is very realistic. In particular, the second scenario finding the most significant changes is very helpful to understand new data by navigating through a few parts of data. I think this framework can be applied for only OLAP analysis but only other environment such as personalized ranked queries.