A survey of query recommendation techniques for datawarehouse exploration

Patrick Marcel*, Elsa Negre**

*Université François Rabelais Tours, Laboratoire d'Informatique, France
patrick.marcel@univ-tours.fr,
** LINA (UMR CNRS 6241) - Nantes University, France
elsa.negre@univ-nantes.fr

Abstract. Lots of data are gathered in datawarehouses that are navigated and explored for analytical purposes. Only recently has the problem of recommending a datawarehouse query to a datawarehouse user attracted attention. In this paper, we propose a simple formal framework for expressing datawarehouse query recommendations. We propose to see the problem of recommending a datawarehouse query for exploration purposes as a function computing a set of queries and associated ratings given a query log, a session, a user profile, a datawarehouse instance, and an expectation function. The rating computed indicates the usefulness of each query for a session. With this viewpoint, we review and categorize the few techniques that, to the best of our knowledge, have been proposed and we illustrate them with a case study.

1 Introduction

Lots of data are gathered and shared in databases that are navigated and explored for analytical purposes. Only recently has the problem of recommending a database query to a database user attracted attention (Chatzopoulou et al., 2009; Khoussainova et al., 2009; Stefanidis et al., 2009). However, in other contexts (like e.g., e-commerce or web search) the problem of computing recommendations is deeply investigated (Adomavicius and Tuzhilin, 2005; Baeza-Yates, 2010).

A typical example of database analysis is a datawarehouse navigated by decision makers using OLAP queries (Sarawagi, 2000). A datawarehouse can be seen as a large database with a particular topology, shared by many analysts who have various interest and viewpoints, explored by sequences of queries. In such a context, query recommendation is particularly relevant since the user is left with the tedious task of navigating a large datacube to find valuable insight.

In this paper, we survey the existing methods for computing datawarehouse query recommendations. We restrict the scope of this survey to methods that, given a user's query over a datawarehouse, use it or transform it into another query, with a supposed added value for the user's exploration. We propose a formal definition of this problem, namely to see the recommendation of datawarehouse queries for exploration purposes as a recommending function