Towards an Intelligent OLAP System Facing Sparse Problems

Rania Koubaa, Miracl Laboratory, Tunisia
Eya Ben Ahmed, Miracl Laboratory, Tunisia
Faiez Gargouri, Miracl Laboratory, Tunisia

ABSTRACT

Exploring intelligent data stored in data warehouses may efficiently assist the knowledge-seeker in his decision process. Such traced information related to performed analysis by decision-makers on data warehouses are stored in OLAP log files. These files contain useful knowledge about the analysts’ preferences. Sometimes, some formulated queries provide no results. Such a dilemma is known as the sparsity problem. In this paper, to overcome this limitation in user-centric data warehouses, the authors focus on a specific class of preferences, namely the conflicting preferences. Indeed, a conflicting preference describes a low frequency preference stored in OLAP log files, so that it is considered as tailored to given analysts. Such preferences are characterized by their rarity. To deal with this issue, the authors introduce a new approach to discover these preferences through mining of rare association rules using a new introduced method for generating the N highest confidence rare association rules. The derived rare preferences will be used to reformulate the launched query avoiding an empty result. The carried out experiments on their built online recruitment data warehouse point out the efficiency of their approach.

Keywords: Conflicting Preference, Data Warehouse, Personalization, Preference, Rare Association Rules

1. INTRODUCTION

Today the Web is emerging exponentially and currently this trend is more than just a hypertext. The Web started to gain essentials of intelligence. In this context, the social web is a collection of social relationships linking people through the web. In fact, the social dimension of Web 2.0 emphasizes the communication between the Internet users with similar interests. Such interaction may be concertized through several online activities (Alfimtsev et al., 2012) such as social network (Cruz-Cunha et al., 2012), education, shopping…etc. Particularly, a social network defines the social organization between actors associated through different relationships ranging from informal contact to familiar relationships. Indeed, distinguished examples of social networks are LinkedIn, Tweeter, Facebook, Viadeo, etc.

DOI: 10.4018/IJWP.2014100103
Several works were devoted to the analysis of interest topics in social network. However, such operation remains effortful and generally does not provide satisfying results due to the large amount of managed information in this context.

Actually, the data warehouse domain delivers numerous solutions for modeling and dealing with huge amount of data. In fact, according to the interest topic, such a social structure may be modeled using multidimensional modeling and the managed information in social network context may be described using diverse dimensions and aggregated through several hierarchies using data warehouse solution.

Indeed, a data warehouse is a set of technologies aimed at enabling the analyst to manage large volumes of data extracted from production information systems in order to assist him in his decision process (Inmon, 2002). The user can interactively explore the multidimensional data by means of On-Line Analytical Processing (OLAP) paradigm (Kimball, 1997) through formulating complex MDX\textsuperscript{1} queries. These launched queries stored in OLAP log files, contain a set of useful knowledge about the analyst preferences. Indeed, a multidimensional preference is closely related to an instance of the data warehouse schema. Based on extracted preferences, several approaches of user-centric OLAP approaches are proposed: (i) OLAP recommender system involves an application predicting the user responses to options in multidimensional context (Jerbi et al, 2009), (Giacometti et al, 2009, 2011), (Khemiri et al., 2012), (Bimonte et al., 2014), (Ben Ahmed et al, 2015) and (Aligon et al., 2015); (ii) OLAP personalization tailors the presentation of the data cube to match the user’s preferences (Ravat et al., 2008), (Garrigós et al., 2009), (Jerbi et al., 2010), (Aligon et al., 2011) and (Golfarelli et al., 2011).

Generally, the personalization is known as a mechanism providing an overall customized, individualized user experience by taking into account the needs, the preferences and the characteristics of the user (Holland et al., 2003).

To the best of our knowledge, few works address the sparsity problem in web-based OLAP analysis when the MDX query generates an empty result. In this context, we distinguish two main trends of preferences: (i) frequent preference related to high frequency multidimensional component and (ii) rare or conflicting preference which has low frequency. In this context, we stress on the rare category of preferences.

In this paper, we focus on the data warehouse as a tool for representation and investigation of multidimensional social networking data. Thus, we introduce a new personalization approach dedicated to discover conflicting preferences from web-based OLAP log files. Such derived patterns will be used in the query reformulation to avoid the sparse results.

To motivate our contribution, we use, throughout this paper, an example of web-based recruitment data warehouse that we built in order to evaluate our contribution. Indeed, the web-based recruitment is the practice of human resources recruitment using electronic technologies. It renovates the recruitment landscape for both employers and job seekers. On the one hand, the companies moved their recruitment process online in order to find speedily relevant applicants. On the other hand, the job seekers may apply swiftly for job offerings. Our designed data warehouse aims to boost the web-based recruiting productivity. Its key purpose is to examine the job offerings and to enhance job opportunities for applicants.

Figure 1 illustrates a part of logical online recruitment warehousing schema. The adopted notations are similar to notations of Golfarelli et al. (1998). Indeed, our schema contains a fact called recruitment measured using the number of vacancies and analyzed according to several dimensions such as the time, the location and the area dimensions. Performing OLAP (Chaudhuri, 1997) in such a data warehouse may help both of companies and job seekers in their recruitment cycle. Recurrently, the launched query may provide an empty result because some user preferences are uncommon. To overcome this problem, we propose to extract the analyst...
The BIZEWEST Portal
www.igi-global.com/chapter/bizewest-portal/17850?camid=4v1a