Closing the Gap between Data Mining and Business Users of Business Intelligence Systems: A Design Science Approach

Ana Azevedo, Instituto Superior de Contabilidade e Administração do Instituto Politécnico do Porto, Porto, Portugal
Manuel Filipe Santos, Departamento de Sistemas de Informação, University of Minho, Guimarães, Portugal

ABSTRACT

Since Lunh first used the term Business Intelligence (BI) in 1958, major transformations happened in the field of information systems and technologies, especially in the area of decision support systems. BI systems are widely used in organizations and their importance is recognized. These systems present themselves as essential parts of a complete knowledge of business and an irreplaceable tool in the support to decision making. The dissemination of data mining (DM) tools is increasing in the BI field, as well as the acknowledgment of the relevance of its usage in enterprise BI systems. BI tools are friendly, iterative, and interactive, allowing business users an easy access. The user can manipulate directly data, having the ability to extract all the value contained into that business data. Problems noted in the use of DM in the field of BI is related to the fact that DM models are complex in order to be directly manipulated by business users, not including BI tools. The nonexistence of BI tools allowing business users the direct manipulation of DM models was identified as the problem. More of these issues, possible solutions and conclusions are presented in this article.

Keywords: Business Intelligence (BI), Business Users, Data Mining (DM), Data Mining Languages, Data Mining Standards, Design Science Research, Inductive Databases, Inductive Data Warehouses, Knowledge Discovery in Databases, Query-by-Example

INTRODUCTION

The research presented in this paper approaches the issue of using Data Mining (DM) languages in the context of Business Intelligence (BI) systems. The aim is to study the viability of developing a DM language oriented to business users and oriented to the BI activities.

Motivation

Business Intelligence (BI) is one emergent area of the Decision Support Systems (DSS) discipline and can be defined as the process that transforms data into information and then into knowledge (Golfarelli, Rizzi, & Cella, 2004). Being rooted in the DSS discipline, BI has suffered a considerable evolution over the last years and is, nowadays, an area of DSS that
attracts a great deal of interest from both the industry and researchers (Arnott & Pervan, 2008; Clark, Jones, & Armstrong, 2007; Davenport, 2010; Hannula & Pirttimäki, 2003; Hoffman, 2009; Negash, 2004; Richardson, Schlegel, & Hostmann, 2009; Richardson, Schlegel, Hostmann, & McMurchy, 2008; Sallam, Hostman, Richardson, & Bitterer, 2010). A BI system is a particular type of system. One of the main aspects is that of user-friendly tools, that makes systems truly available to the final business user.

The term Knowledge Discovery in Databases (KDD) was coined in 1989 to refer to the broad process of finding knowledge in data, and to emphasize the “high-level” application of particular data mining (DM) methods (Fayyad, Piatetski-Shapiro, & Smyth, 1996). The DM phase concerns, mainly, to the means by which patterns are extracted and enumerated from data.

DM is being applied with success in BI and several examples of applications can be found (Linoff, 2008; Turban, Sharda, Aroson, & King, 2008; Vercellis, 2009). Despite that, DM has not yet reached to non-specialized users and thus it is not yet completely integrated with BI. Powerful analytical tools, such as DM, remain too complex and sophisticated for the average consumer of BI systems. McKnight supports that bringing DM to the front line business personnel will increase their potential to attaining BI’s high potential business value (McKnight, 2002). Another fundamental issue that is pointed out by McKnight is the capability of DM tools to be interactive, visual, and understandable, to work directly on the data, and to be used by front line workers for intermediate and lasting business benefits.

Currently, DM systems are functioning as separate isles, and hereby it is considered that only the full integration of the KDD process on BI can conduct to an effective usage of DM in BI (Azevedo & Santos, 2011). Three main reasons can be pointed out for DM to be not completely integrated with BI, each one leading to a specific problem that constraints DM usage in BI. Firstly, the models/patterns obtained from DM are complex and there is the need of an analysis from a DM specialist. This fact can lead to a non-effective adoption of DM in BI, being that DM is not really integrated on most of the implemented BI systems, nowadays. Secondly, the problem with DM is that there is not a user-friendly tool that can be used by decision makers to analyze DM models. Usually, BI systems have user-friendly analytical tools that help decision makers in order to obtain insights on the available data and allow them to take better decisions. Examples of such tools are On-Line Analytical Processing (OLAP) tools, which are widely used (Negash, 2004; Turban, Sharda, Aroson, & King, 2008). There are not equivalent tools for DM that allow business users to obtain insights in DM models. Finally, but extremely important, it has not been given sufficient emphasis to the development of solutions that allow the specification of DM problems through business oriented languages, and that are also oriented for BI activities. With the expansion that has occurred in the application of DM solutions in BI, this is, currently, of increasing importance.

BI systems are, usually, built on top of relational databases and diverse types of languages are involved (Figure 1). As a consequence, DM integration with relational databases is an important issue to consider when studying DM integration with BI. Codd’s relational model for database systems (Codd, 1970, 1982) is long ago adopted in organizations. One of the reasons for the great success of relational databases is related with the existence of a standard language – Structured Query Language (SQL). SQL allows business users to obtain quick answers to ad-hoc business questions, through queries on the data stored in databases. SQL is nowadays included in all the Relational Database Management Systems (RDBMS). SQL serves as the core above which are constructed the various Graphical User Interfaces (GUI) and user friendly languages, such as Query-By-Example (QBE), included in RDBMS (Date, 2004; Elmasri & Navathe, 2007). It is also necessary to define a standard language, which can operate likewise for data mining. Several approaches have been proposed for the definition of data mining languages. In the literature there can be
Business Intelligence 2.0: The eXtensible Markup Language as Strategic Enabler
Rubén A. Mendoza (2010). *International Journal of Business Intelligence Research* (pp. 63-76).
www.igi-global.com/article/business-intelligence-extensible-markup-language/47196?camid=4v1a