Explanatory Business Analytics in OLAP

Emiel Caron, Centre for Science and Technology Studies, Leiden University, Leiden, The Netherlands

Hennie Daniels, Center for Economic Research, Tilburg University, Tilburg, The Netherlands, & Rotterdam School of Management, Erasmus University, Rotterdam, The Netherlands

ABSTRACT

In this paper the authors describe a method to integrate explanatory business analytics in OLAP information systems. This method supports the discovery of exceptional values in OLAP data and the explanation of such values by giving their underlying causes. OLAP applications offer a support tool for business analysts and accountants in analyzing financial data because of the availability of different views and managerial reporting facilities. The purpose of the methods and algorithms presented here, is to extend OLAP applications with more powerful analysis and reporting functions. The authors describe how exceptional values at any level in the data, can be automatically detected by statistical models. Secondly, a generic model for diagnosis of atypical values is realized in the OLAP context. By applying it, a full explanation tree of causes at successive levels can be generated. If the tree is too large, the analyst can use appropriate filtering measures to prune the tree to a manageable size. This methodology has a wide range of applications such as interfirm comparison, analysis of sales data and the analysis of any other data that possess a multi-dimensional hierarchical structure. The method is demonstrated in a case study on financial data.

Keywords: Business Analytics, Exception Reporting, Explanation, OLAP, Variance Analysis

Modern firms can store millions of transaction data in company databases, and consequently the potential of obtaining valuable new business insights from business data has increased enormously. The proliferation of sophisticated software with new analysis tools and the online availability of data will alter the way of working of business and financial analysts. Large amounts of transaction data are nowadays stored in a company data warehouse and multi-dimensional data items like sales (2008, product, region) can be extracted from the data warehouse and organized in so called OLAP cubes for analysis. Typical questions like “Why has sales increased in 2008 compared to 2009” or “Why is performance of our branch office ABC low compared to the average” can be answered by inspection of multidimensional data cubes. In principle the analyst can explore the data by using the standard operators in OLAP like drill-down, roll-up and slice (Han & Kamber, 2005). But as the data sets become large, browsing through the data in search for atypical values may become a complicated and tedious task. Moreover, when it comes to an efficient in depth examination of the
underlying causes, there is still a shortage of tools to intelligently prune a large tree of causes to its essential branches. In this paper we propose several extensions of the OLAP framework for intelligent variance analysis. Remarkable differences of actual versus reference values; like actual versus budget, actual versus historical, etc., are automatically detected by statistical models or normative models. In the next step these differences are explained by generating the most important causes at lower level data. The latter process is guided by several heuristic rules to reduce information overload.

The goal of this paper is to largely automate the current manual managerial discovery process in OLAP systems and to extend these systems with more powerful analysis and reporting functions. This functionality can be provided by extending the conventional OLAP system with an explanation formalism, which supports the work of human decision makers in diagnostic processes, as part of the intelligence phase in the managerial decision-making process (Simon 1960). Here diagnosis is defined as finding the best explanation of unexpected behaviour (i.e., symptoms or exceptions) of a system under study (Verkooijen, 1993). This definition captures two tasks that are central in problem diagnosis, namely exception identification and explanation generation. Mintzberg et al. (1976) describe problem identification as an activity “in which opportunities, problems and crises are recognized and evoke decisional activity” and explanation generation as an activity “in which management seeks to comprehend the evoking stimuli and determine cause-effect relationships”.

In Figure 1, the conceptual architecture of Business Intelligence (BI) framework and its main components are depicted, based on the idea of the enterprise information factory (Inmon 1996). The framework describes how companies conduct and organize BI. In the framework, BI is arranged in components for (1) data production, (2) data assembly, logistics, and storage, and (3) data processing, analysis, and consumption. At the right side of Figure 1, the contributions of our research are depicted in the general BI framework. The OLAP database is extended with novel functionality for the detection of exceptional values, explanation generation, and sensitivity analysis.

The remainder of this paper is organised as follows. In the next section, we summarize the most important OLAP database concepts and notations. In the following section, we show our methodology for explanatory analysis of exceptional values. This section is divided in subsections as follows. First, we show how exceptional values in OLAP databases are defined

---

**Figure 1. Business intelligence framework (based on Inmon (1996)) with extensions for explanatory business analytics for OLAP databases**
Business Information Integration from XML and Relational Databases


[www.igi-global.com/chapter/business-information-integration-xml-relational/4240?camid=4v1a](www.igi-global.com/chapter/business-information-integration-xml-relational/4240?camid=4v1a)