Automated business diagnosis in the OLAP context

Emiel Caron, Hennie Daniels

[1]Erasmus University Rotterdam, ERIM Institute of Advanced Management Studies, PO Box 90153, 3000 DR Rotterdam, The Netherlands, phone +31 010 4082574, e-mail: ecaron@fbk.eur.nl; [2]Tilburg University, CentER for Economic Research, Tilburg, The Netherlands

Abstract. In this paper, we describe an extension of the OLAP (On-Line Analytical Processing) framework with automated causal diagnosis, offering the possibility to automatically generate explanations and diagnostics to support business decision tasks. This functionality can be provided by extending the conventional OLAP system with an explanation formalism, which mimics the work of business decision makers in diagnostic processes. The central goal of this paper is the identification of specific knowledge structures and reasoning methods required to construct computerized explanations from multidimensional data and business mo-dels. The methodology was tested on a case study involving the comparison of financial results of a firm's business units.

Keywords: Datamining, OLAP, Explanation, Business Intelligence.

1. Introduction

Today's OLAP (On-Line Analytical Processing) systems have limited explanation or diagnosis capabilities. The diagnostic process is now carried out manually by analysts, where the analyst explores the multidimensional data to spot exceptions, and navigate the data with operators like drill-down, roll-up, and selection to find the reasons for these exceptions. Such functionality can be provided by extending the conventional OLAP system with an explanation formalism. Here diagnosis is defined as finding the best explanation of unexpected behaviour (or symptoms) of a system under study [10]. This definition assumes that we know which behaviour we may expect from a correctly working system, otherwise we would not be able to determine whether the actual behaviour is what we expect it or not. The expected behaviour in an OLAP environment can be derived from some statistical model or can be domain knowledge from analysts. The objective of this paper is to extend the OLAP system with a diagnostic process. In short, we automate the current user-driven analysis of OLAP data, with an explanation

formalism that finds exceptions, and finds out why these exceptions have emerged.

Our exposition on causal explanation is largely based on Feelders and Daniels' notion of explanations [2, 3], which is essentially based on Humpreys' notion of aleatory explanations [6] and the theory of explaining differences by Hesslow [4]. Causal influences can appear in two forms: contributing and counteracting. Therefore, the following canonical form for causal explanations is proposed in [2, 3]:

$$\langle a, F, r\rangle \text{ because } C^{+}, \text{ despite } C^{-}. \tag{1}$$

where $\langle a, F, r\rangle$ is the event to be explained, C^{+} is non-empty set of contributing causes, and C^{-} a (possibly empty) set of counteracting causes. The explanation itself consists of the causes to which C^{+} jointly refers. C^{-} is not part of the explanation, but gives a clearer notion of how the members of C^{+} actually brought about E. The explanandum is a three-place relation between an object a (e.g. the ABC-company), a property F (e.g. having a low profit) and a reference class r (e.g. other companies in the same branch or industry). The task is not to explain why a has property F, but rather to explain why a has property F *when the members of r do not.* This general formalism for explanation constitutes the basis of the framework for diagnosis in an OLAP context developed in this paper.

To position this paper we mention some related work regarding the explanation of differences in multidimensional data. In [8] Sarawagi presented an operator for data cubes that lets the analyst get summarized reasons for drops or increases observed at an aggregated level. In [9] the authors developed a discovery-driven exploration paradigm that mines the data for exceptions and summarizes the exceptions at appropriate levels in advance. The discovery-driven method is guided by pre-computed indicators of exceptions at various levels of detail in the cube.

The remainder of this paper is organized as follows. Section 2 introduces our notation for the multidimensional model, followed by a description of normative models appropriate for diagnosis. In section 3 the explanation formalism is extended for multidimensional data in order to automatically generate explanations for symptoms derived from multidimensional data. Finally, the complete method is illustrated in a case study on sales data in section 4.

2. Notation, equations and normative models

Many different notations of OLAP concepts are found in the literature. Here we introduce a generic notation that is particular suitable for combining the OLAP concepts of measures, dimensions, and dimension hierarchies. A measure is defined as a function on multiple domains: $y^{i_1 i_2 \ldots i_n} : D_1^{i_1} \times D_2^{i_2} \times \ldots \times D_n^{i_n} \rightarrow \square$. Each