

Theory and support for process frameworks of knowledge discovery and data mining from ERP systems

Elliot Bendoly*

*Decision and Information Analysis, Room 411, 1300 Clifton Road, Goizueta Business School,
Emory University, Atlanta, GA 30322-2701, USA*

Received 8 September 2001; received in revised form 12 February 2002; accepted 17 August 2002

Abstract

Existing theory has framed the process of information extraction and agglomeration, also referred to as the knowledge discovery (KD) process, as a series of strategic search decisions, subject to constraints, with the objective of attaining a sufficient level of domain-specific knowledge for use in strategic planning. Supported by the experiences of firms representative of Client, Developer, and Third-party segments of the data mining (DM) community, this work provides an extension to this basic framework. The implications provided suggest a wealth of untapped opportunities in the area of KD research.

© 2002 Elsevier Science B.V. All rights reserved.

Keywords: Data mining; Knowledge discovery in databases; Process; ERP

1. Introduction

The adoption of enterprise resource planning (ERP) systems over the last fifteen years has been accompanied by an explosion of readily available transactional data. Sales figures, human resource activity, stock-out, and defect occurrences are only a part of the issues accessible to and empowering the modern corporation. Often however, the relevance of this data comes from the information that can be *derived* from examining multiple issues simultaneously and the ability to draw inferences critical to strategic planning [2]. Benefits contribute to the firm's business intelligence and subsequently to the overall competitive advantage of the firm [9,20]. The challenge faced by firms, and the analysts charged with manipulating

this data, rests in their ability to provide acceptable levels of strategically applicable information as the result of allowable effort, in time and money.

One of the fundamental problems of information extraction is that the formats of available data sources are often incompatible, requiring extensive conversion efforts. In an attempt to reduce this difficulty, several ERP systems, such as SAP and Baan, have embedded means by which to organize and archive the transactional data in their application databases. The incorporation of such *data warehousing* schemes has interested both researchers and practitioners. *Knowledge discovery* (KD) describes both the overall process by which information is then extracted/agglomerated and the domain dedicated to research on it [8,27].

There has also recently been a concentrated effort to provide *data mining* (DM) tools able to assist analysts faced with unstructured KD tasks. Distinction between

* Tel.: +1-404-727-7138; fax: 404-727-2053.

E-mail address: elliot_bendoly@bus.emory.edu (E. Bendoly).



Fig. 1. Hierarchical depiction of knowledge discovery and related concepts.

KD and DM concepts remain ambiguous, however, in spite of efforts by various researchers in giving good definitions and examples of their differences [3]. Of the distinguishing features discussed, the most common is the iterative and process-oriented nature of KD, as well as its emphasis on the development of strategic knowledge and domain understanding. Data mining on the other hand is discussed in specific applications and tools for finding rules and relationships among the data.

From a knowledge management standpoint, DM tools allow for the creation of well-defined transferable information [18]. In contrast, KD processes are also characterized by data retrieval, data cleansing, criteria specification, and performance analysis. KD processes agglomerate interim information found by such techniques as data mining in generating understanding and domain knowledge. In an adaptation of the scheme proposed by Haeckel and Nolan in 1993 [10] and inspired by the linkages to decision theory as proposed by Kuhlthau [17], the relationships can be depicted hierarchically as in Fig. 1.

There may be many different goals of a particular KD task, including such objectives as the derivation of dependent relationships, development of forecasts, and classification. The product is an agglomeration of such information, organized in a format that can be applied as knowledge ultimately relevant to a

downstream planning activity. At any one iteration in the process, alternate levels of task information become applicable and thus alternate extraction techniques (DM tools) may prove superior. However, while the majority of recent KD literature has focused on the efficiency of the algorithmic tools, a review of the literature reveals that little attention has been given to the strategic nature and internal dynamics of the discovery process. Since the overall success of a KD process is dependent upon both process efficiency and quality of output, and since certain constraints, such as time, may apply to an overall knowledge discovery task, an analysis of the dynamic nature of these strategies is needed to effect improvements.

2. KD process theory

In 1989, the term knowledge discovery in databases (KDD) was coined to represent the process by which resources are applied towards the transformation of available data into strategic information [5,12]. The task of this transformation can be characterized by a number of issues related to the nature of the data and the features of the outputs desired. The discipline encompasses research including the study and development of data mining (DM) tools.

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات