S

Symbiotic Data Miner

Kuriakose Athappilly
Western Michigan University, USA

Alan Rea
Western Michigan University, USA

INTRODUCTION

Symbiotic data mining is an evolutionary approach to how organizations analyze, interpret, and create new knowledge from large pools of data. Symbiotic data miners are trained business and technical professionals skilled in applying complex data-mining techniques and business intelligence tools to challenges in a dynamic business environment.

BACKGROUND

Most experts agree (Piatetsky-Shapiro, 2000; Thearling, 2007) that data mining began in the 1960s with the advent of computers that could store and process large amounts of data. In the 1980s, data mining became more common and widespread with the distribution of relational databases and SQL. In the 1990s, business saw a boom in data mining as desktop computers and powerful server-class computers became affordable and powerful enough to process large amounts of data in data warehouses (Havenstein, 2007) as well as real-time data via online analytical processing (OLAP). Today we see an increasing use of advanced processing of data with the help of artificial intelligence technology tools such as fuzzy logic, decision trees, neural networks, and genetic algorithms (Brachman, et al., 1996; Gargano & Raggad, 1999). Moreover, current trends are moving organizations to reclassify data mining as business intelligence using such tools as Cognos (Cognos, 2007).

We also see three distinct theoretical approaches to data mining: statistical (classical), Artificial Intelligence (heuristics), and machine learning (blended AI and statistics). The three approaches do not adhere to the historical boundaries applied to data mining; rather, they are embarkation points for data-mining practitioners (Thuraisingham, 1999; Kudyba & Hoptroff, 2001; Kepner & Kim, 2003; Padmanabhan & Tuzhilin, 2003). It is not the intent of this discussion to argue which approach best informs data mining. Instead, we note that many software platforms adhere to one or more methods for solving problems via data-mining tools.

Most organizations agree that sifting through data to create business intelligence, which they can use to gain a competitive edge, is an essential business component (Lee & Siau, 2001; Brown, 2004; MacInnis, 2004; Burns, 2005). Whether it is to gain customers, increase productivity, or improve business processes, data mining can provide valuable information if it is done correctly. In most cases, a triad of business manager, information technology technician, and statistician is needed to even begin the data-mining process. Although this combination can prove useful if a symbiotic relationship is fostered, typically the participants cannot effectively work with one another because they do not speak the same language. The manager is concerned with the business process, the technician with software and hardware performance, and the statistician with analyses of data and interpretations of newfound knowledge. While this may be an overgeneralization, it is not far from the truth (O'Hara, 2007).

What is needed, then, is an individual who can pull all three components together: a symbiotic data miner trained in business, technology, and statistics.

MAIN FOCUS

In this chapter we will discuss how an individual trained not only in business but also in technology and statistics can add value to any data mining and business intelligence effort by assisting an organization to choose the right data-mining techniques and software, as well as interpret the results within an informed business context.

Data Mining in Contemporary Organizations

Data mining is the "semi-automatic discovery of patterns, associations, changes, anomalies, rules, and statistically significant structures and events in data" (Dhond, et al., 2000, p. 480). Analyzed data is many times larger than the sum of its parts. In other words, data mining can find new knowledge from observing relationships among the attributes in the form of predictions, clustering, or associations that many experts might miss. The new knowledge in a continuously changing environment is the most potent weapon for organizations to become and remain competitive (Inmon, 1996; Amato-McCoy, 2006; Corbitt, 2006).

In today's business organizations intelligence is necessary to anticipate economic trends, predict potential revenue streams, and create processes to maximize profits and efficiency. This is especially true for strategic and other mid-level managers (Athappilly, 2003). In the past, many decisions were made using corporate experience and knowledge experts. This is still true today. However, with the increased influx of data—some experts argue that the amount of information in the world doubles every 20 months (Dhond, et al., 2000; Tallon & Scannell, 2007)—many high-level managers now turn to data-mining software in order to more effectively interpret trends and relationships among variables of interest (Deal, 2004).

To support data mining an increasing amount of funds are invested in complex software to glean the data for patterns of information; hardware is purchased that can effectively run the software and distribute the results, and personnel are continually retrained or hired. The personnel include IT technicians, knowledge experts, statisticians, and various business managers. The mix of personnel needed to effectively collect, glean, analyze, interpret, and then apply data-mined knowledge ultimately can lead to one of the biggest data-mining challenges: communicating results to business managers so that they can make informed decisions. Although the managers are the ones who ultimately make the decisions, they do not have the necessary skills, knowledge base, and techniques to assess whether the heuristics, software, and interpreted results accurately inform their decisions. There is ultimately a disjunction between theoretical interpretation and pragmatic application (Athappilly, 2004).

The Challenge for Contemporary Organizations

The challenge is two-fold: 1. A shortcoming of many data-mining tools is the inability of anyone except experts to interpret the results. Business managers must be able to analyze the results of a data-mining operation to "help them gain insights . . . to make critical business decisions" (Apte, et al., 2002, p. 49) and 2. Business managers must rely on IT technicians to apply rules and algorithms, and then rely on statisticians and other experts to develop models and to interpret the results before applying them to a business decision. This process adds at least two layers between the decision and the data. Moreover, there are numerous opportunities for miscommunication and misinterpretation among the team members (O'Hara, 2007).

In order to flatten the layers between the requisite gleaned knowledge and its interpretation and application, a new type of business IT professional is needed to create a symbiotic relationship, which can sustain itself without the triadic team member requirements and the inherent polarities among them.

The Solution for Contemporary Organizations

The solution to the complex data-mining process is symbiotic data miners. The symbiotic data miner is a trained business information system professional with a background in statistics and logic. A symbiotic data miner not only can choose the correct data-mining software packages and approaches but also analyze and glean knowledge from large data warehouses. Combined with today's complex analysis and visualization software, such as Clementine (SPSS, 2007) and Enterprise Miner (SAS, 2007), the symbiotic data miner can create illustrative visual displays of data patterns and apply them to specific business challenges and predictions.

Just as today's business managers use spreadsheets to predict market trends, analyze business profits, or manage strategic planning, the symbiotic data miner can fulfill the same functions on a larger scale using complex data-mining software. Moreover, the miner can also directly apply these results to organizational missions and goals or advise management on how to apply the gleaned knowledge.

Related Content