Symbiotic Data Mining

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 (Anon, 2002; Thearling, 2002) 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 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 (Gargano & Raggad, 1999). Moreover, current trends are moving organizations to reclassify data mining as business intelligence, using such tools as Cognos (2004).

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 (Kudyba & Hoptroff, 2001; Thuraisingham, 1999). 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). 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 even to begin the data-mining process. Although this combination can prove useful if a symbiotic relationship is fostered, typically, the participants cannot work effectively 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.

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 THRUST

In this paper, 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.

In today’s business, organizations intelligence is necessary to anticipate economic trends, predict poten-
Related Content

Complementing the Data Warehouse with Information Filtered from the Web
www.igi-global.com/chapter/complementing-data-warehouse-information-filtered/7869?camid=4v1a

Hyperbolic Space for Interactive Visualization
www.igi-global.com/chapter/hyperbolic-space-interactive-visualization/10663?camid=4v1a

World Wide Web Personalization
www.igi-global.com/chapter/world-wide-web-personalization/10787?camid=4v1a

Aggregate Query Rewriting in Multidimensional Databases
www.igi-global.com/chapter/aggregate-query-rewriting-multidimensional-databases/10560?camid=4v1a