

Data Mining Software

John Wang

Montclair State University, USA

Qiyang Chen

Montclair State University, USA

James Yao

Montclair State University, USA

INTRODUCTION

With increasing amounts of data being generated by businesses and researchers, there is a need for fast, accurate, and robust algorithms for data analysis. Improvements in databases technology, computing performance, and artificial intelligence have contributed to the development of intelligent data analysis. The primary aim of data mining is to discover patterns in the data that lead to better understanding of the data generating process and to make useful predictions (Hand, Mannila, & Smyth, 2001). Most companies now collect and refine massive quantities of data in data warehouses. These companies realize that to succeed in a fast paced world, business users need to be able to get information on demand.

Many organizations now view information as one of their most valuable assets, and data mining software allows a company to make full use of these information assets. Data mining software analyzes relationships and patterns in stored transaction data based on open-ended user queries. Several types of analytical software are available: statistical, machine learning and neural networks, decision trees, naive-Bayes, K-nearest neighbor, rule induction, clustering, rules based, linear and logistical regression time sequence, and so forth (Wang, 2005). There is never enough time to think of all the important questions; that is why the computer should do this itself. It can provide the winning edge in business by exploring the database and it brings back invaluable information.

BACKGROUND

The concept of data mining is relatively new; however, the technology has been around for decades. Companies

have been using computers to sort through data and analyze reports for years. Nevertheless, continuous advancement in technology like improved computer processing power and state of the art statistical software are dramatically increasing the accuracy of analysis, while at the same time lowering overall operating costs. As Zhou (2003, p. 139) observed,

The rapid progress in the digital data acquisition and storage technology has led to the fast growing and tremendous amount of data stored in databases, data warehouses, or other kinds of data repositories such as the World Wide Web. Although valuable information may be hiding behind the data, the overwhelming data volume makes it difficult if not impossible, for human beings to extract them without powerful tools.

Data mining uses advance pattern recognition and mathematical and statistical techniques to find hidden patterns, trends, and correlations within the data sets

Data mining software is the prime result of this ever-improving technological age. It has become one of the most popular diagnostic tools for analyzing data. It gives organizations the ability to examine data from different dimensions, helps them categorize it, and then summarizes the relationships identified in order to provide useful information. Most forms of data mining software typically group data into four types of relationships. They are class, cluster, association, and sequential pattern. A class relationship is formed when stored data are used to locate data in predetermined groups. A cluster is developed when data items are grouped according to logical relationships or consumer preferences. Associations simply links between the items consumers choose and when they chose them. And lastly, sequential pattern is when data are mined to anticipate behavioral patterns and trends.

Data mining technology and software can be used with just about any type of data. It can combine information from many diverse sources to create a detailed data image about each of us—our income, our driving habits, our hobbies, our families, and our political interests (Laudons, 2005). Due to its ability to gather and analyze such diverse data about the preferences of large groups of people, it is the perfect tool for companies, especially retailing and marketing organizations, who have a strong customer focus. It enables these organizations to establish relationships among factors like price, product placement, and competition. They then can use these relationships to determine the impact on sales, customer satisfaction, and profits.

Today, data mining has become an essential part of doing business, and almost all major businesses use some sort of data mining software. Some of the more prevalent software suites are CART by Salford Systems, Clementine by SPSS Inc., and Enterprise Miner by SAS. These programs are the premier in analytical software because they help to make data more predictive, and provide useful insight for decision makers. For instance, Wal-Mart has realized that all their customer data has a great value. They have approximately one hundred million customers entering their stores every week and they are able to collect all the data at the checkout and then map and update it by store, state, and region. Wal-Mart looks at this data to constantly make their operation more efficient; they even use this data when they negotiate with their suppliers. They like to find out what products sell together and then often use one item to attract the customer into the store at a rock bottom price and then have the other items that usually sell with that item at a moderate price that may not be the lowest price in the area. This approach in data mining has worked very well for Wal-Mart stores (Hays, 2004).

MAIN FOCUS

Data mining software is used across many industries. Data mining tools can answer business questions that traditionally were too time-consuming to resolve. Data mining is, in some ways, an extension of statistics, with a few artificial intelligence and machine learning twists thrown in. Like statistics, data mining is not a business solution; it is just a technology. It is designed to help to simplify data mining and eliminate some of

the errors other systems cannot. It also makes it easier to gather information and use it effectively within the company. By combining this with a skilled management team, this program can save money and increase profits in a shorter period of time rather than other database systems.

Executives need to make a sound decision on what type of application fits into the corporate environment and meet project needs. The choice of data mining software is not an easy task. Contrary to common opinion, the best tool suite for one company may not be the most advanced tool; it may not be the one with the most data mining algorithms or the one that gives the greatest accuracy in prediction.

TECHNICAL PERSPECTIVE TO CHOOSE DATA MINING SOFTWARE

The characteristics of the cutting edge data mining software applications can be classified into the following rubric (Giraud-Carrier & Povel, 2003).

Portability

A wide variety of standalone data mining tools are available on the commercial market. Some are general-purpose tools and others are tailored to use in specific industries such as finance or retail. To be truly effective as a tool to improve yield and reliability, data mining software technology needs to be available as part of an integrated yield management software environment. Otherwise, the user will be faced with importing and exporting data from one tool to another depending on where they are in the analysis cycle. To achieve this ease of use, data mining needs to be available as part of a commercially acquired yield management system, or it needs to be integrated in a seamless fashion with an in-house system. The latter case requires both a fully functional programmatic interface as well as an adequate internal staff to set-up and maintain the integration.

Reliability

Investments in high availability technologies, additional backup and restore capabilities, and replication enhancements will enable enterprises to build and deploy highly reliable applications. Moreover, improved

4 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/data-mining-software/16699

Related Content

Integration of E-Learning into Curriculum Delivery at University Level in South Africa

Rabelani Dagadaand Agnes Chigona (2013). *International Journal of Online Pedagogy and Course Design* (pp. 53-65).

www.irma-international.org/article/integration-learning-into-curriculum-delivery/75541

Influence of Class Participation and Coursework on Academic Performance in the Context of Blended Learning

Nazmi Xhomaraand Indrit Baholli (2022). *International Journal of Online Pedagogy and Course Design* (pp. 1-14).

www.irma-international.org/article/influence-of-class-participation-and-coursework-on-academic-performance-in-the-context-of-blended-learning/295951

An Exploratory Study of Incorporating Blogfolios Into a College Writing Course

Jerry Chih-Yuan Sun, Geoffrey Middlebrook, Otto Kheraand Ho-Yuan Chen (2018). *International Journal of Online Pedagogy and Course Design* (pp. 1-15).

www.irma-international.org/article/an-exploratory-study-of-incorporating-blogfolios-into-a-college-writing-course/201112

Didactic Model Aiming at the Use of Digital Information and Communication Technologies in Science Education: Theoretical and Methodological Foundations

Vanessa Arias Gil, Marco Antonio Moreiraand Jesús Ángel Meneses Villagrá (2022). *International Journal of Online Pedagogy and Course Design* (pp. 1-11).

www.irma-international.org/article/didactic-model-aiming-at-the-use-of-digital-information-and-communication-technologies-in-science-education/305729

A Comparative Study on the Curriculum in the Community-Based Chinese Heritage Language (CHL) Schools and Dual Language Immersion (DLI) Program

Shizhan Yuan (2019). *Handbook of Research on Assessment Practices and Pedagogical Models for Immigrant Students* (pp. 139-196).

www.irma-international.org/chapter/a-comparative-study-on-the-curriculum-in-the-community-based-chinese-heritage-language-chl-schools-and-dual-language-immersion-dli-program/229404