Integrated Data Mining and Business Intelligence

S M Monzurur Rahman
United International University, Bangladesh

Md Faisal Kabir
United International University, Bangladesh

Muhammad Mushfiqur Rahman
Samsung R&D Institute Bangladesh Ltd, Dhaka, Bangladesh

INTRODUCTION

The chapter will describe a new integrated approach to data mining by seamlessly combining business intelligence software functions with manual graphical analysis capabilities with automated data mining techniques. The integrated approach provides an analytic system that supports the KDD process in an integrated manner. Furthermore it recognizes the importance of manual analysis as an important part of the KDD process. The data mining process works cooperatively to enhance OLAP analysis and drill down allowing manual and automated approaches to be combined. It will outline both the general Data Mining process and catalogue many of the types of business intelligence tasks those are needed in any typical data mining production environment.

BACKGROUND

Today, the collection and analysis of data is integral to the strategic performance of an organization. When performance does not meet targeted expectations organizations must be in a position to analyse performance data to gain insight on how relevant strategies can be improved. They need to find out why they are having problems, what the cause is and the optimal approach for improvement.

Due to the volumes and complexity of data collected across organizations today manual approaches to analysis are often not effective. The data may be spread across the organisation in a diversity of systems and formats. Analytical systems need to be able to integrate this diversity and provide a comprehensive view of the business. Automated analytic systems are required to help sift through large volumes of data to find interesting patterns that would not be possible manually. Data mining techniques provide, well-proven, techniques to help automate analysis. Data Mining helps transform data into actionable information and provide the insight required to improve strategy (Han, J. & Kamber, M., 2001). What is commonly known as Data Mining is a part of a small process called “Knowledge Discovery in Databases”? According to Piatetsky-Shapiro and Frawley (Piatetsky-Shapiro & Frawley, 1991), “Knowledge Discovery in Databases is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data”.

Knowledge Discovery is a cyclic process. The process starts with the selection of key data from a data warehouse, data mart or the amalgamation of data from other sources (Edelstein H. 1999). When large data bases are available it is often necessary to sample the data to get smaller manageable portions of data. Usually the data requires manual pre-processing or transforming

DOI: 10.4018/978-1-4666-5202-6.ch114

Copyright © 2014, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

to some degree prior to analysis. Pre-processing may involve tasks like cleaning up the data, remove noisy, missing data, or grouping data. Transformation usually involves the creation of new derived attributes. e.g. Trends or Moving Averages but also may involve filtering, ordering, editing, and normalisation. Data visualization is also important at this point to gain a basic understanding of the data.Data mining is the next part of this process, according to Bradley et al., (Bradley, Fayyad, & Mangasarian, 1998): "Data mining is the step in the KDD process concerned with the algorithmic means by which patterns or models (structures) are enumerated from the data under acceptable computational efficiency limitations." Data Mining can be used in a number of ways:

- **Exploration:** The aim is to gain new insight on your business and discover new patterns in your data. We may discover a small segment of customers that are not satisfied and discover their locality is not covered well by service technicians.
- **Confirmation:** You may have a hunch that a pattern exists and need to test for its existence.
- **Prediction:** The aim is to build a model that can predict the likelihood of business events. The model reads in new records and generates a score for that event. For examples, predicting if a customer will churn next month, or respond to marketing campaign, or predict when a system is going to fail.
- **What if Analysis:** Once you have a predictive model it can be used to explore different scenarios by entering hypothetical data.

Data Mining can be used to help solve the following kinds of problems:

- **Classification:** The aim of classification is to determine the membership of objects, into some predefined classes, based on a set of features or measurements. For example, can we predict if a customer will be "profitable" or "unprofitable" when focusing sales and marketing efforts?
- **Estimation:** Estimation is similar to classification but involves continuous valued outcomes. For example, can we predict the revenue return or lifetime value of a prospective customer?
- **Segmentation:** The aim of segmentation is to find natural groupings in your data, called clusters, and then try to describe them. For example we may want to find behavioural market segments that differentiate our customers. Segments can be hierarchical in that a grouping can be broken down into sub groupings.

The outputs of data mining are models or patterns that need to be interpreted in the business context. It is this interpretation that provides insight and knowledge about your business. It is not unusual to discover possible improvements or further questions as a result of interpreting the models. For example, other information or data should be included or excluded, that data needs further pre-processing or transformation. The models may suggest certain hypotheses that can be manually examined using OLAP or graphical analysis under the umbrella of Business Intelligence (BI). These may in turn suggest more improvements to the models. In this way the KDD process repeats and new models or patterns are generated.

The majority of real world applications of data mining have found that automated approaches alone are not enough and that almost always some manual analysis is also required. Today, there are very few data mining tools (One example is CorBusiness™, corvu.rocketsoftware.com) that successfully integrate manual and automated techniques. In most cases separate systems are used for data pre-processing, data mining, manual analysis and final reporting, resulting in a number

18 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/integrated-data-mining-and-business-intelligence/107322

Related Content

The Role of Information and Communication Technology in Competitive Intelligence
Dirk Vriens (2004). *Information and Communications Technology for Competitive Intelligence (pp. 1-33).*
www.irma-international.org/chapter/role-information-communication-technology-competitive/22559

The Performance Mining Method: Extracting Performance Knowledge from Software Operation Data
Stella Pachidiand Marco Spruit (2015). *International Journal of Business Intelligence Research (pp. 11-29).*
www.irma-international.org/article/the-performance-mining-method/132821

Analytical Customer Requirement Analysis Based on Data Mining
Jianxin Jiao, Yiyang Zhangand Martin Helander (2006). *Business Applications and Computational Intelligence (pp. 227-247).*
www.irma-international.org/chapter/analytical-customer-requirement-analysis-based/6027

Information Architectures for Business Continuity
Nijaz Bajgoric (2009). *Continuous Computing Technologies for Enhancing Business Continuity (pp. 60-78).*
www.irma-international.org/chapter/information-architectures-business-continuity/7133

Neighborhood Evaluation in Recommender Systems Using the Realization Based Entropy Approach
Roee Anuar, Yossi Bukchin, Oded Maimonand Lior Rokach (2014). *International Journal of Business Analytics (pp. 34-50).*
www.irma-international.org/article/neighborhood-evaluation-in-recommender-systems-using-the-realization-based-entropy-approach/119496