

Business data mining — a machine learning perspective

Indranil Bose^{a,1}, Radha K. Mahapatra^{b,*}

^a*Department of Decision and Information Sciences, Warrington College of Business Administration, University of Florida, 351 Stuzin Hall, P.O. Box 117169, Gainesville, FL 32611, USA*

^b*Department of Information Systems and Management Sciences, College of Business Administration, University of Texas at Arlington, P.O. Box 19437, Arlington, TX 76019-0437, USA*

Received 29 November 1999; received in revised form 5 February 2001; accepted 12 February 2001

Abstract

The objective of this paper is to inform the information systems (IS) manager and business analyst about the role of machine learning techniques in business data mining. Data mining is a fast growing application area in business. Machine learning techniques are used for data analysis and pattern discovery and thus can play a key role in the development of data mining applications. Understanding the strengths and weaknesses of these techniques in the context of business is useful in selecting an appropriate method for a specific application. The paper, therefore, provides an overview of machine learning techniques and discusses their strengths and weaknesses in the context of mining business data. A survey of data mining applications in business is provided to investigate the use of learning techniques. Rule induction (RI) was found to be most popular, followed by neural networks (NNs) and case-based reasoning (CBR). Most applications were found in financial areas, where prediction of the future was a dominant task category. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Business applications; Data mining; Machine learning

1. Introduction

Data mining, also known as “knowledge discovery in databases” [23], is the process of discovering interesting patterns in databases that are useful in decision making. Data mining is a discipline of growing interest and importance, and an application area that can provide significant competitive advantage to an organization by exploiting the potential of large data warehouses.

The task of finding patterns in business data is not new. Traditionally, it was the responsibility of business analysts, who generally use statistical techniques. The scope of this activity, however, has recently changed. Widespread use of computers and networking technologies has created large electronic databases that store business transactions. Retailers, like Wal-Mart Stores, capture millions of sales transactions through their point-of-sale terminals. Transactions can be analyzed to identify buying patterns of individual customers as well as customer groups, and sales patterns of different stores.

Intense competition is forcing companies to identify innovative ways to capture and enhance market shares while reducing cost. A better appreciation of the buying behavior of customers can enhance the

* Corresponding author. Tel.: +1-817-272-3590; fax: +1-817-272-5801.

E-mail addresses: bosei@ufl.edu (I. Bose), mahapatra@uta.edu (R.K. Mahapatra).

¹ Tel.: +1-352-392-0648; fax: +1-352-392-5438.

effectiveness of target marketing practices. Data warehousing technology has enabled companies to organize and store large volumes of business data in a form that can be analyzed and a maturing of the “artificial intelligence” field has created a set of techniques of “machine learning” that are useful in automating tedious but crucial activities of discovering patterns in databases. These factors have changed the way that business data are analyzed and given rise to data mining, which integrates machine learning, statistical analysis and visualization techniques, with the intuition and knowledge of the business analyst, to discover meaningful and interesting patterns in business data.

Data mining is a complex process involving multiple iterative steps. Fig. 1 gives an overview of this process. The first step is the selection of data for analysis. Normally, historical data is used. The data

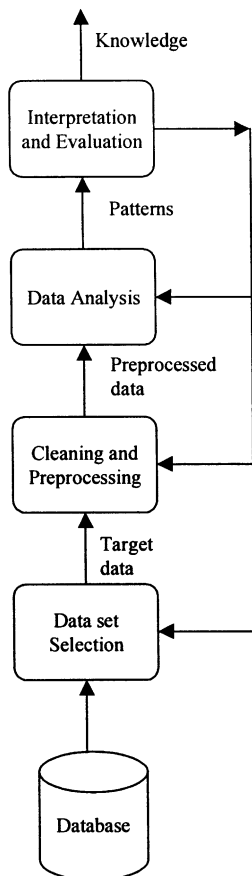


Fig. 1. An overview of the data mining process (adapted from Fig. 1 in [23]).

set may be retrieved from a single source, such as a data warehouse, or may be extracted from several operational databases. The selected data set then undergoes cleaning and preprocessing. Lack of consistency across databases creates serious problem when data is extracted from multiple databases. The cleaning operation removes discrepancies and inconsistencies. Some mining techniques require data to be preprocessed to improve its quality. Examples include transformation of data from one scale to another, identification of predictive attributes in the data set, and reduction of the dimension of the data set through recomposition.

The data set is analyzed next to identify patterns, i.e. models that represent relationships among data. The model is then validated with new data sets to ensure its generalizability. It should be possible to translate the model into actionable business plans that are likely to help the organization achieve its goals. A model or pattern that satisfies these conditions becomes business knowledge. The steps in the mining process are performed iteratively until meaningful business knowledge is extracted.

A number of algorithms have been developed in domains, such as machine learning, statistics, and visualization, to identify patterns in data. Of these, statistical modeling approaches are the oldest. The data set must conform to rigid distribution criteria to employ statistical modeling methods. Pattern discovery algorithms based on machine learning techniques, however, impose fewer restrictions and produce patterns that are easy to understand. They are, therefore, finding wide popularity in data mining applications. Each technique has its own strengths and weaknesses. Understanding these in the context of business data mining is very useful in selecting an appropriate technique for a specific application. The objective of this paper is to inform the information systems (IS) manager and the business analyst about the role of machine learning techniques in business data mining.

2. An overview of machine learning techniques

Machine learning is the study of computational methods to automate the process of knowledge acquisition from examples [34]. This discipline evolved to eliminate the laborious and expensive knowledge

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