Research on Retail Sales Management System Based on Data Mining Technology

Tong Wei[1] and Gong Peng[2]

[1] Anhui Wonder University of Information Engineering – Department of Computer Engineering, Hefei, China
[2] Computer Department of Shandong University at Weihai, China
gongpeng@sdu.edu.cn

Abstract. This paper established a database system for sales, customer, freight, consumption, services, and other information based on the panel sales activity of a building material company through e-commerce. The emphasis of our proposed research in this paper is to apply the decision number C4.5 algorithm of data mining in product sales system. In addition, the original simple query for sales data is replaced with finding the real factors for influencing product sales. The main research work in this paper includes the selection algorithm in data mining, interpretation and evaluation of results.

Keywords: data mining, marketing, Decision tree, Sales Analysis.

1 Introduction

It is well known that Internet of Things is a sort of network connecting information exchange with communication, and intelligently identifying, positioning, tracing, monitoring, and managing articles through Radio Frequency Identification Device (RFID), infrared sensor, global positioning system (GPS), laser scanners and other information sensing devices according to a certain convention protocol. Internet of Things can be divided into three layers, which are sensing, network and application layers. Previously, the technology emphasis is focused on data perception. With the development and application of Internet of Things, however, the application is not limited in query and management of high volumes of data. More importantly, massive fuzzy application data is waiting for analysis to extract the potential useful information from complex and cumbersome data. Therefore, data mining becomes a promising technology which attracts more research interest day by day.

This paper established database of sales, customer, freight, consumption, services, and other information based on the panel sales activity of a building material company through e-commerce. The emphasis is on the following aspects. (1) How can decision number C4.5 algorithm of data mining technology be applied in enterprise e-commerce? (2) How can customer's purchasing behavior be identified; and how can customers shopping patterns and trends be discovered from mass data in order to improve the service quality of enterprise e-commerce, to increase product sales ratio, and to decrease business costs. The main content in this paper is the selection algorithm in data mining, interpretation and evaluation of results.

Q. Zu et al. (Eds.): ICPCA/SWS 2013, LNCS 8351, pp. 586–592, 2014.

2 Database Structure

Data mining is a kind of information or advanced processing process in which some implicit, unknown, and non-trivial information or information with potential application value are identified and extracted from the mass data stored in database or data warehouse.

With respect to the market demand in the boards sales activity of a building material company, an e-commerce sales database structure is designed including time dimension table, customer dimension table, product dimension table, and sales fact table.

Time dimension table (TimeD): tracks the time of products being sold, including the time identifier (TimeID), year (Year), month (Month), day (Day), quarter (Quarter) and other attributes.

Customer dimension table (customer): including customer identifier (customerID), customer name (customer name), customer type (customer type), address (address) and other attributes. Customer type is classified into key account (KA), very important person (VIP), and ordinary customers (client).

Product dimension table (product): including product Number (product ID), product name (product name), specification (spec), unit of measurement (unit), upper limit of inventory (min_sto), lower limit of inventory (max_sto), product subdivision (product class) and other attributes. Among those the production subdivision is further divided into three categories: small class (low_lvl), middle class (middle_lvl), and large class (higher_lvl).

Sales fact table (sales_fact_2013): the sales data of all kinds of decorative products in the major store year (only data of 2010, 2011, and 2012 is calculated).

Among these data, sales amount of product is metric; and product type, customer type, and quarter are the key attributes of sales; while inventory, agent distributer, and promotion means are the import factors which also have an impact on sales of product.

3 Data Mining

Data processing is an important link in the process of implementing data mining. The data preprocessing method introduced in this paper is: extracting relevant training dataset from the product dimension table, customer dimension table, time dimension table, inventory dimension table and sales fact table. That means to classify and disperse all kinds of attributes in the sales database to form a training dataset for data mining. In the training dataset, for instance, sales value is divided into four classes:

(1) large transaction,
(2) medium transaction,
(3) small and medium transaction, and
(4) small transaction.

These four classes are classified according to the following criteria: it is a large transaction when single transaction is greater than or equal to 200 thousands RMB; a medium transaction when single transaction is greater than or equal to 100 thousands and less than 200 thousands RMB; a small and medium transaction when single transaction is greater than or equal to 10 thousands and less than 100 thousands RMB; and