

# The application of data mining technology in e-commerce

Li Yan

Weifang Vocational College, ShanDong WeiFang, 261041, China

**Keywords:** Data mining; Electronic business letter; Application research

**Abstract.** This paper expounds the concept of data mining technology and the architecture of data mining technology. Focuses on the process of data mining. And is the story of the main technical data mining. And then expounds the application of data mining technology in e-commerce activities. Including customer purchasing behavior analysis for mining, analyze customer access to the site, from historical sales data mining trading rule, Positioning customer network gender and path analysis.

## Introduction

E-commerce to promote the development of production and living ways, has penetrated into various industries and fields, plays an important role in stimulating economic growth, promoting transformation. E-commerce service industry is growing up to be a modern service industry a core industry. By the end of 2011, the scale of Chinese Internet users reached 500 million marks, 513 million people. China's online shopping users to scale up to 194 million people, up 20.8% from a year earlier, online shopping usage 37.8%, network shopping user scale will continue to be in the future. The dependence of network shopping users for network shopping is higher. Most of the network shopping users browse the frequency of online shopping website.

## Summary of the data mining technology

**The architecture of data mining .** Data mining is a complex process, so the structure of data mining is more multifarious. Typical data mining system structure is a three-tier structure, as showed in figure 1, from the bottom up, respectively is: data layer, data mining engine, user interface.

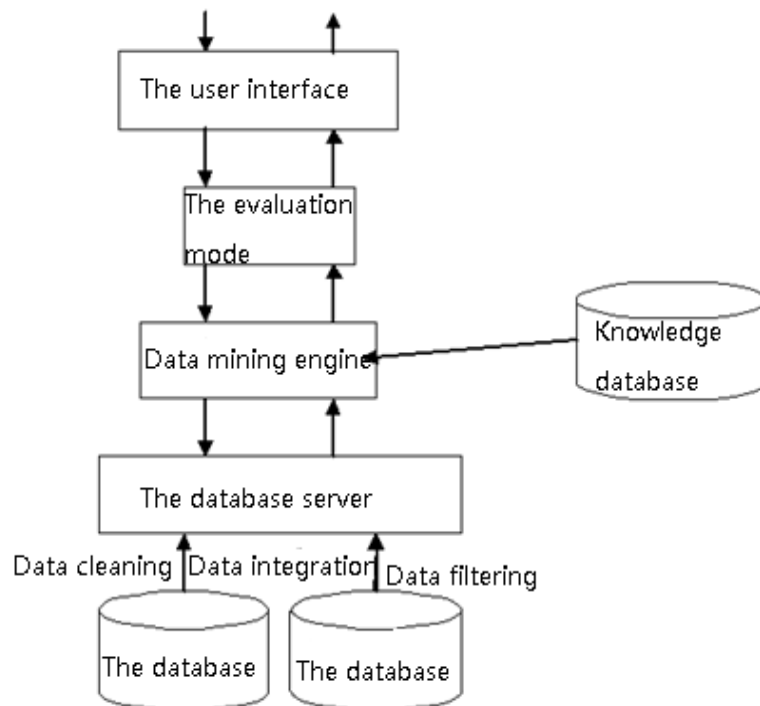


Fig.1 Mining system architecture

1)The data layer. The object of data mining is data. For the purposes of mining, data from various sources, after pretreatment (data cleaning, data integration, data filtering, etc.), will be stored in a unified model of a database or data warehouse. The pattern design of comprehensive consideration, the characteristics of the data mining tasks, the implementation algorithm of diverse factors, such as mining can optimize.

2)Data mining engine. Data mining engine integration within the data mining algorithm. It according to the knowledge base, mining the data in a database or data warehouse have many model, and then to assess these patterns, screening, finally can be meaningful and user interest model.

3)The user interface. The user interface is the human-computer interaction interface. These patterns with intuitive user interface and easy for users to understand the way to the user. In addition, users can also must (sometimes) through a user interface to guide the organizational model of the engine, data model evaluation, mining.

**The process of data mining .** Data mining is a multi-step process, the process from a large database of previously unknown, effective and practical information, and uses this information to make decisions or knowledge. This process is interactive and iterative, many of which processes require the participation of users.

Data mining process mainly includes four steps: problem definition or determine the business object, data preparation, data mining, expression and analysis results. Though we have the numerous steps in sequence, but data mining process is not linear. To obtain good results will be constant and repeat these steps.

**Data preparation.** Determine the target, the appropriate mining is about to start preparing related data for mining. Prepare data is according to the mining requirement gathering data, and establish the database can be found. Data is ready to spend time in the whole process of data mining are one of the most.

**Data mining .** According to the definition of the problem clear mining tasks and goals, first select the appropriate data mining algorithm for mining (such as correlation patterns, sequential patterns, clustering, classification, etc.), and then establish and select proper model and parameters, using the known data for training and testing of the model, the final mining results is obtained using this model. Data mining is a work of "model" of the people in accordance with the design of the process of data processing, analysis and prediction. , of course, an algorithm or tool can't be alone all of the data mining, usually it is hard to start to choose which algorithm is most suitable for the problems facing, so need to set up different model parameters (or algorithm), choose the best.

**Expression and evaluation of mining results .** This phase is subdivided into the expression and result evaluation of two steps. To dig up the results of the represented in the form of a kind of easy to understand, and analyzed and evaluated. Specifically includes eliminating irrelevant and redundant mode, filter the information to be presented to the user. Using visualization techniques to meaningful patterns in the form of visual graphics or logic, said into user understandable language. Successful application of data mining technology can convert the raw data to more concise and easier to understand, can be clearly defined relationship in the form of, also can solve found potential conflict with previous knowledge and use of statistical methods to evaluate model, so as to decide whether to need to be repeated before operation, in order to get the best and most suitable pattern. We also analysis of knowledge can be built into the organizational structure of the business information system, realize the assimilation of knowledge.

### **The main technical data mining**

1)Correlation analysis. Correlation analysis aims to unearth hidden in the relationship between data. Data association is a kind of important that exists in the database can be found. If two or more variables between the values of a certain regularity, is referred to as a link. The commonly used correlation analysis technique is association rules and sequential patterns. Association rule is utilized

to search for in the same event correlation of different items, such as buy different goods in a purchase activity. Sequential patterns for finding events in time correlation between.

2)Classification analysis. The concept of classification is to find a category description, it represents the overall characteristics of this kind of data, namely the connotation of the class description, generally represented by rules or a decision tree model. The connotation of a class description into characteristic description, and the difference between sex description, characterization is a common characteristic of objects in the class description; distinctiveness is the distinction between two or more classes.

3)Cluster analysis. Data in the database can be divided into different classes according to definite rules. In the same category, the distance between the individual is small, and the distance between separate categories of individuals is larger. Clustering analysis and classification analysis: is the difference between a clustering analysis of data in the database before does not contain any category tag, will have a common trend and pattern of data tuple gathered for a class, the class among the tuples similar degree is highest, the biggest difference between the classes. In practice, can accord to customer's data, using cluster analysis market according to the customer's consumption pattern similarity can be divided into several segment of the market, to develop a targeted marketing strategy, improve sales performance.

Data mining technology in e-commerce applications

**Introduction of electronic commerce** . Electronic commerce refers to people or businesses through the Internet network, with digital electronic means business data exchange and to carry out business activities. The current domestic existing online business advertising, electronic clearing, online ordering, online banking, online payment and settlement, and other types of electronic commerce form.

E-commerce has the following advantages:

1)Services without being limited by the time, have brought great convenience to customers. Consumers can never leave home, 24 hours a day to do all kinds of information query, query goods, such as instant shopping activities.

2)Global resources are sharing. On Internet information capacity is tremendous, anyone can benefit from it

3)Broke through the regional limitation, can directly with the city, national, and even clients all over the world

4)Greatly reduce the business operating costs. Businesses no longer need to real shops, but also can direct purchase, reduce storage, accelerate the capital turnover, can save a lot of manpower and material resources and financial resources

5)Businesses can more convenient, more direct, more systematically accept customer feedback, is good for merchants to do after-sales service and market research

6)Because most of the information can be real-time online to send, so can greatly reduce the advertising cost and information cost, and can increase the timeliness

7)The characteristics of the Internet, active, interactive digital telephone, fax, television, newspapers and other traditional media can't replace, incomparable

8)Save shopping space. Because online shopping is non-store direct form, the manufacturer can be directly linked to consumer, and through the information feedback timely adjustment of product supply and demand, buyers feel store is near, manufacturers can avoid blind production caused by the backlog.

### **Mining customer buying behavior**

The analysis of consumer buying behavior is offline sales and online electronic commerce activities and the most basic the most commonly used data mining activities. Classic case of beer and diapers is based on the analysis of the customer's purchase behavior found in a foreign man in the supermarket to buy diapers at the same time also bought beer, thus decided to adjust the placement of the diapers

and beer more reasonable, promote the sales of beer in the supermarket, at the same time, it saves time for beer and diapers, shortens the time of each customer shopping and improve sales performance. This is a classic case of an offline supermarket, also applies in the e-commerce activities. We can from the website of the records in the database of customer purchase records, can be a transaction. It may be a period of time of transaction records, to analyze the customer purchased goods in trading. Find out the connection between the goods and commodities, goods and customer behavior, so as to adjust e-commerce recommendation strategy, purchase suggestions, etc. For the customer to shorten the time of search goods, simplify operation, enhance the user experience. In April 2012, sponsored by the IT168 and ChengTa media database technology conference in 2012 there was a data mining technology, they invited to taboo data mining engineer to share. Tobago's engineers at the meeting to share such a case, they analyzed the annual valentine's day, Chinese valentine's day, Christmas, the three festivals on taboo commodity sales. Found in the three festivals, taboo user basic products such as flowers and chocolates. This is not surprising, is often the case in these festivals are gifts for men and women in love, flowers and chocolates sales also is normal. But after a time and a half months to January, probably they bought flowers and chocolates in these festival customers have made a purchase behavior analysis, the analysis found that this part of the customer after buying flowers or chocolates a weeks to a month between most of the customers are bought another commodity, that is "condoms". Then analyzes the "condom" volume per month, compared the valentine's day, Chinese valentine's day, Christmas holiday in three months, and other "condom" in volume. Finally draw a conclusion that the customer buy flowers and chocolates in important festival after a week to a month of time is very big probability will buy family planning supplies.

According to the analysis conclusion taboo can the customer will recommend customers to buy after flowers or chocolates into the shopping cart to family planning supplies, such as: XXX customers hello, 80% buy flowers and chocolates customers also bought the following goods. Family planning supplies show customers make a list.

Reasonable recommendation strategy, purchase Suggestions may shorten the duration of search goods from the clients, simplify the customer's purchase operation, as well as to enhance the customer purchase experience. Per the consumer's shopping to shorten the time and improve the quantity and each transaction trading amount, make the electronic commerce activity more efficient.

**Analyze customer access to the site** . E-commerce is through the Internet, with digital electronic means business data exchange, and carry out business activities. This activity needs to be based on an e-commerce site, site access behavior reflects the customer's purchase intention, and interested in, and the use of on site function proficiency, such as the site's search, goods classification, etc. Customers access to a site way, a person's online habits, the use of the network frequency, on the degree of cognitive e-commerce (online), etc.

Enemy and know yourself in order to fight in war, e-commerce activities is not war, but the more understanding of your customer base, grasp their habits and hobby to conduct e-commerce activities is absolutely critical, can better to retain customers and develop new customers, to improve the competitiveness of the enterprises is very important. Using data mining technology to analyze customer behavior data on site visit to better understanding to the customer, make appropriate navigation information, within the site classification, site of convenient function, advertising, etc.

The second is to analyze the customer access to the Site, such as direct input to the site address to access the customer may be frequent this site, has the online shopping experience, familiar with even rely on the network. For such customers don't need too much guidance and promotion, because they already a veteran. Such as still have a plenty of through basic search, or any other website link to the site, it can be concluded that these customers are not often on the network activity, even indirectly through other sites they need to access our e-commerce site, for this kind of customer needs to have certain guidance, such as they will prompt our site collection to IE favorites, or guide them how to remember or direct access to our site.

**Gender positioning customer network** . Electronic commerce activities typically require customers to register the Web fill in relevant information and then can be normal e-commerce

activities. Customer personal information must be kept secret not to leak, so the use of such information and analysis needs more considers the customer's information security and privacy. In fact sheet from the customer registration information to judge if the gender or classification of the client is not very accurate, why do you say that? For example, we can do such an analysis, site registration information to be obtained for male customer transaction records was analyzed, and a look at these so-called men have bought what kind of goods from the customers, how many percentage of goods are made. Analysis can be concluded that probably is the male registered users in fact about 50% in women buy goods, this is a common phenomenon, men shopping online in most case, it is not for himself in consumption, more is for women to consume, from the survey of the entire e-commerce industry women consumption consumption subject is greater than the men, if we simple user registration of gender to the classification of the male and female customers in related marketing, advertising, product recommendation and strategy if the effect is not so accurate and clear. It should be combined with customer purchase records for goods of gender analysis, so as to let the consumer analysis in e-commerce activities, gender in the network.

### **Path analysis**

Path analysis can be used to determine the most frequently accessed in a Website path. There is some other information about the path through path analysis can be concluded:

1 ) 70% of the client in access/company/product2, started with/company, after/company/new/company/products/company/product1.

2) 80% of customers visit this site started/company/products.

3) 65% of clients after browsing four or fewer pages left.

The first rule on/company/product2 page has useful information, but because the customer is circuitous of site visit, so useful information is not obvious. The second rule illustrates the customer access to the site in general is not starting from the home page, but since/company/products, if this directory contains some product types of information on a page, will be a good idea. The third rule illustrates the clients resident time on the site. Since consumers browse on the website is generally not more than 4 pages, you can put the important commodity information on these pages. Through path analysis, can improve the structure of pages and website design.

### **Conclusion**

According to the analysis of the data mining technology to help enterprises to carry out. Data Mining, Data Mining, from Data set automatically extract the useful information hidden in the Data of the non trivial process. The information in the form of rules, concepts, rules and patterns, etc., can help decision makers analyze historical data and current data, discover hidden relationships and patterns, then possible to predict the future behavior.

### **References**

- [1] Feng Qian, Linwen Xu. Data mining and its application in marketing. Hangzhou Institute of electronics industry management branch, the 2001-08.
- [2] Qingyi Wang, Enhong Chen. Knowledge discovery problems and application research. Computer science, 1997 vol. 24 No. 5.
- [3] Yi Sun, Zhen Fang. E-commerce. Beijing: Peking University Press, 2010-8-1.
- [4] Shudong Zhou. Introduction to e-commerce. Nanjing: southeast university press, 2011-7-1.
- [5] Jiawei Han, Bo Kan. Data mining concepts and technology (original version 2) mechanical industry publishing house, 2007-3-1.

[6] Xianchun Zou. E-commerce and Web data mining. Chongqing: Southwest normal university, school of computer and information science, 2001-5.