

Web Data Mining and Its Implication in E-commerce

Wu Lihua Deng Tian

NanChang Institute of Science & Technology

Abstract-With the rapid development and popularity of Internet, the development of e-commerce has increasingly drawn the attention of researchers. They hope to make full use of its advantages and obtain more economic benefits in this new business mode. Web has become the foundation for enterprises to carry on e-commerce. Apply the thoughts and methods of data mining to e-commerce, help the e-commerce operators to get the truly valuable knowledge from the vast amount of information and to guide their decisions. Web data mining is combined with e-commerce in this background. How to effectively analyze and use these information, find out the inner link and serve the business activities have become the common concern of e-commerce operators. This paper discusses the origin of data mining, the basic function of data mining, development overview of data mining technology at home and abroad as well as the implication of data mining in e-commerce.

Keywords-Data mining; Web mining; electronic commerce

I. BRIEF INTRODUCTION

With the continuous development of database technology and the widely use of database system in information management, the amount of data stored in database increases sharply. How to discover knowledge from these huge amount of database causes the knowledge discovery and the emergence of data mining field. Data mining (knowledge mining) is a pattern to draw those previously unknown and potentially useful from the large amount of data. Knowledge discovery in database (KDD) and data mining (DM) is a cross research area including statistics, artificial intelligence, pattern recognition, parallel computing, machine learning, database, knowledge acquisition, data visualization, expert system and other technology. The knowledge discovered is used in decision support process, post hoc and other implication.

The extremely complicated web page is dynamic and without structure. These make it difficult for people to find the required data and information on web page rapidly and easily. Web data mining is a process to discover the implied, unknown, with potential application value and extraordinary mode from the massive Web document collection and related data browsed in the website. It's the implication of traditional data mining technology in web environment. Web data mining can be divided web content mining, web structure mining and web usage pattern mining.

E-commerce refers to individuals or enterprises using digital electronic means to exchange business data

and carry on advertising, selling, buying goods or service. Compared with traditional business activities, e-commerce is not subject to regional restriction, cost saving and many other advantages.

This paper mainly discusses how to effectively use several kinds of feasible data mining technologies such as path analysis, association rule analysis, sequential pattern analysis, classification analysis and clustering analysis to discover users' purchase mode and browse mode in e-commerce. It also proposed the methods to realize path analysis and sequential pattern analysis.

II. THE CAUSE OF DATA MINING

Data mining, also called Knowledge discovery in database (KDD), is an advanced treatment process which extracts reliable, innovative, efficient and understandable modes from a large amount of data. KDD was first put forward on the 11th International Joint Artificial Intelligence Academic Conference held in August 1989. Data mining is a information processing method which combines many disciplines and its main functions are as follows:

1. Classification: According to the property and characteristics of the analyzed objects, establish different categories to describe things. For example: divide each article into different categories according to keywords.

2. Clustering: Identify the internal rules of the object being analyzed, and divide the object according to these rules. For example: As for the question such as "What's the customer's favorite promotion way", cluster according to the purchasing habit and divide the customers with similar purchasing habits to one group. Different categories indicate different purchasing habits. Respectively investigate the favorite promotion way of each group of customers.

3. Association rules: Association is a connection when something happens, other things will happen accordingly. For example: People who buy milk everyday are possible to buy bread. How many of them will certainly buy milk and what the proportion is can be described by the supporting rate and credibility of association. Unlike association, sequence is a vertical connection.

4. Prediction: Effective prediction requires to establish a prediction model. The purpose of prediction is to grasp the development rule of the analysis object and anticipate the trend of the future. For example: make judgment to future development of e-commerce.

5. Deviation detection: Describe the minority and extreme special case of the analysis object and reveal the

internal reasons. For example: there are 500 fraud in 1 million transactions in the bank. For the sake of prudent management, the bank needs to find the internal factors of those 500 cases to reduce operation risk. It's important to note that the various functions of data mining don't exist independently but are connected in data mining and play their role.

III. WEB DATA MINING

In recent years, the data mining technology brings a ray of light to solve this problem. Enterprises can

analyze and predict future behavior of customers by the implication on Web of data mining. Enterprise make use of effective customer information by web data mining technology to greatly reduce the operation cost. Web data mining is a process to discover the implied, unknown, with potential application value and extraordinary mode from the massive web document collection and related data browsed in the website.

Web data mining can be divided into Web Content Mining, Web Structure Mining and Web Usage Mining. As shown in Figure 1:

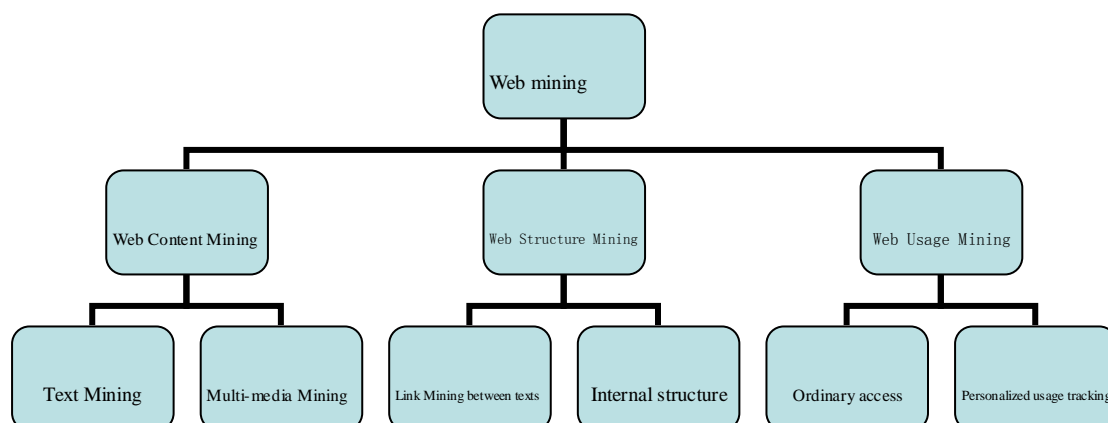


Figure 1. Web Mining Classification Figure

The potential of web data mining lies in applying the existing and latest data mining, analyzing the log, customer information, sales and product related data on Internet server. All the visitors browsing to a site will leave a trail and these information will be stored automatically in the log files of the web server. Web analysis tool generates meaningful information by analyzing and processing the log files on web server. For example: how many people visit the website, where they are from and which sites are the most popular ones and so on. The change in current economic pattern from the traditional entity shop to the electronic trading on the Internet also changes the relationship between the vendors and customers. Now the fluidity of the online customer is big. The main factor they care is the value of goods, not the brand and geographical factors as before. As a result, the next challenge for sellers is to know customers' interest and value orientation as much as possible, so as to ensure the competitiveness in the era of e-commerce.

IV. KNOWLEDGE MODE OF DATA MINING IN E-COMMERCE

The amount of data on web which can be used for data mining analysis is large and of various types. To sum up, there are several types of data which can be used in web data mining technology to produce all kinds of knowledge model.

A. Server data

Customers who visit a site will leave relevant log data on web server and these log data will be stored on

server in the form of files. Generally, they include server logs, error logs, cookie logs, etc.

B. Search data

It's a kind of typical data produced by e-commerce site on server. For example: Online customers may search some products or advertising information, and these search data is linked to server access log of server by cookie or registration information. Currently, there is not a standard search data format.

C. Proxy server data

Web server log only records user's visit to a certain website and proxy server log records user's visit to all the sites. Proxy server equals to providing an intermediary server with cache function between customer's browser and the web server. The cache function reduces the network flow of web server and speeds up web page, at the same time, stores a great deal of customer's visit information in the form of proxy log.

D. Online market data

This kind of data mainly are data related to e-commerce site information stored in traditional relation database, user's purchase information.

V. STRUCTURE MODULES OF PERSONALIZED E-COMMERCE SITE

E-commerce is conducted around three flows, including information flow, capital flow and logistics. E-commerce application is the integration of information flow, capital flow and logistics. Among them, information flow is most important, and it plays a

monitoring role in the entire process. Logistics and capital flow are the guarantee for e-commerce. In the traditional e-commerce solutions and application systems, most of them are taking such a tertiary structure

B/S/S. We designed a B2C e-commerce trading platform frame, mainly for research and develop information flow. See Figure 2.

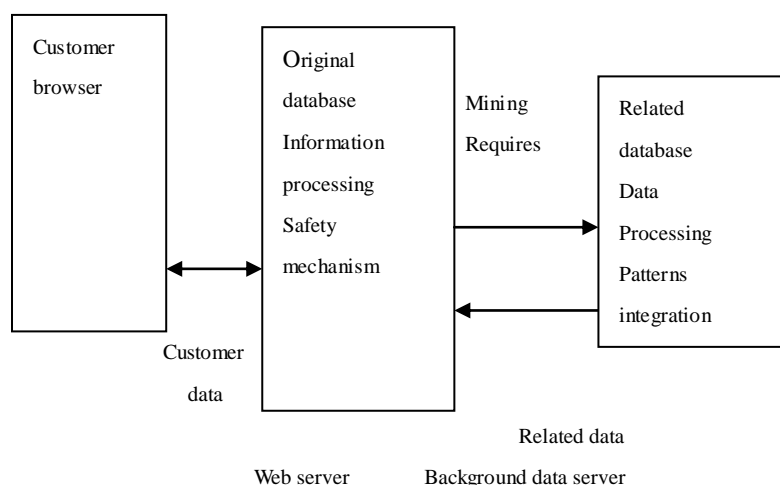


Figure 2. Data Mining Model

The advantages to allocate original database and related database are:

(1)Reduce the burden on the network line. When companies have not given mining task request, the original database system in Web server will operate and maintain automatically, the line is free, and only when a background task request is issued, it will occupy the line.

(2)The information put on web server in the original database system is public and can be provided simultaneously to several companies without influencing on each other.

(3)Data mining process and patterns generated are performed in the background, to ensure the efficiency of the web server.

VI. CONCLUSION

The main job of this paper is to apply web data mining technology to e-commerce , find business intelligence ,improve the structure of e-commerce sites, and provide personalized service for users. Firstly, correlation analysis, sequential pattern analysis, classification analysis and cluster analysis, etc. are used. Pattern analysis is performed for customer behavior

patterns through mining customer information and e-commerce web logs, based on which path analysis and sequential pattern analysis are performed to propose the realization methods; then mining the content of web pages to discover relationships between pages and the products in them, to recommend products online to customers further; followed by a detailed discussion of web log processing technology and frequent path mining algorithm of Web access information mining; Finally, a business website personalized recommendation algorithm based on web mining is proposed with web mining system framework.

REFERENCE

- [1] Mu Z, Xiao H. Computer Aided Process Planning System Design Based on Semi-Generative[J]. International Review on Computers & Software, 2012.
- [2] Witten I H, Frank E, Hall M A. Data Mining: Practical Machine Learning Tools and Techniques[J]. Biomedical Engineering Online, 2011, 5:51(1):95–97.
- [3] Ng R T, Han J. Efficient and Effective Clustering Methods for Spatial Data Mining[J]. Proc of the Vldb Conference, 1994, 88(9):144–155.

- [4] Han J, Kamber M. Data Mining: Concepts and Techniques[J]. Data Mining Concepts Models Methods & Algorithms Second Edition, 2000, 5(4):1 - 18.
- [5] Tan P N, Steinbach M, Kumar V. Introduction to Data Mining, (First Edition)[M]// Addison-Wesley Longman Publishing Co., Inc., 2005.
- [6] Chen M S, Han J, Yu P S. Data mining: an overview from a database perspective[J]. Knowledge & Data Engineering IEEE Transactions on, 1996, 8(6):866-883.
- [7] Witten I H, Frank E. Data mining: practical machine learning tools and techniques with Java implementations[J]. WSEAS TRANSACTIONS on SYSTEMS KAŠPAROVÁ MILOSLAVA, KŘUPKA JIŘÍ, JIRAVA PAVEL ISSN: 1109-2777 1175 Issue 10, Volume 7, 1999, 4(4):76-77.
- [8] Shafer J C, Agrawal R, Mehta M. SPRINT: A Scalable Parallel Classifier for Data Mining[C]// Proceedings of the 22th International Conference on Very Large Data Bases. Morgan Kaufmann Publishers Inc., 1996:544--555.
- [9] Wu X, Kumar V, Quinlan J R, et al. Top 10 algorithms in data mining[J]. Knowledge & Information Systems, 2008, 14(1):1-37.
- [10] Berry M J, Linoff G. Data Mining Techniques: For Marketing, Sales, and Customer Support[J]. 정보과학회지, 1997, 17(1):1 -