Research of Data Mining Based on E-Commerce

Li Yong-hong,
Dept. of Account \Hebei University of Economics and Business
Shijiazhuang, Hebei
liyonghong1966@126.com

Liu Xiao-liang
Dept. of Business \Hebei University of Economics and Business
Shijiazhuang, Hebei
lxl130225@163.com

***Abstract*-this paper describes the Web data mining technology, Web data mining process, data sources and uses, designs a personalized recommendation system based on Web use mining for e-commerce, and discusses in detail the function of each module of the system and implementation techniques. Through the mining of the historical log data of Web server, the system can access to the user browsing patterns to provide personalized service for users.**

Keywords-Web data mining; e-commerce; recommendation system

I. INTRODUCTION

With the increasing popularity of Internet and the rapid development of e-commerce, internet-based business Web site is facing increasing competition. E-commerce sites generate large amounts of data each day, and the data includes consumer-related potential information which is valuable for market analysis and prediction [1]. So how to effectively organize and use such business information and how to understand customer interests and value preferences as much as possible to optimize website design and provide users with personalized service, becomes the issues of the e-commerce to be solved urgently.

Web data mining is the application of data mining technology in the Web environment, is the process to find the implicit, unknown potential value, non-trivial model from a large number of Web documents and the data of users visiting the website. Web data mining can play a role in many areas, and e-commerce provides a wealth of data sources and new research topic for the data mining.

II. WEB DATA MINING

A. Web data mining overview

Web mining is a technology to combine the data mining technology and the Internet, simply said, Web mining is to extract the interesting, potential useful patterns and hidden information from the Web documents and Web activities. On Web there are a mass of data, so how to carry out complex application of these data becomes the research hotspot of current database technology. Data mining is just to find the implicit content with regularity from large amounts of data to resolve data quality problems.

At present, according to the direction of mining, Web data mining is mainly divided into three categories: Web content mining, Web structure mining and Web usage mining, as shown in figure 1.

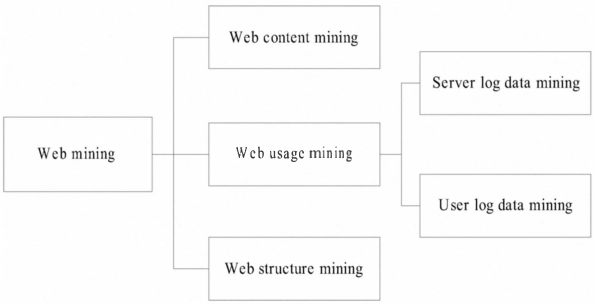

Figure 1. Web mining classification

1) Web content mining

It refers to the process of mining for the Web page content and back-office database, obtaining the useful knowledge from the Web documents, which also can carry out mining for the Web structure and link relationship to obtain useful knowledge from artificial link structure.

2) Web structure mining

Web structure includes the hyperlinks between different pages and the tree structure within the page which can be expressed by HTML, XML, as well as the directory path structures in document URL. Hyperlink structure between the Web pages contains much useful information [2].

3) Web usage mining

The object of Web content mining, Web structure mining is the original data in the network, but the Web usage mining is different from the first two, which faces the second-hand data extracted from the process of user and network interactions. Web usage mining is to find the mode of users to access the Web pages through mining the Web log files and the related data. It also can identify e-commerce potential customers through analyzing and exploring the principles in the Web log records, to strengthen the quality and delivery of end-user Internet information services, and improve the performance of the Web server system.

B. Web data mining application in e-commerce

The emergence of e-commerce has changed the traditional business models, also changed the relationship between the vendors and customers. The expanded customer choice makes them pay more attention to the value of goods, but unlike the previous first consider the brand and geographical factors. Therefore, in terms of the seller to as much as possible understand customer's tastes and values can remain invincible in the competition. Data mining technology can effectively help sellers to understand

978-1-4244-5540-9/10/$26.00 ©2010 IEEE

customer behaviors and improve the efficiency of site, thus having been widely used in the e-commerce design, customer relationship management, Internet marketing, and other fields.

III. E-COMMERCE RECOMMENDATION SYSTEM DESIGN AND IMPLEMENTATION

The task of e-commerce recommendation system is to observe the user's access request online, identify the user session set of each current online user [4], to automatically recommend the pages they may be interested and have not visited according to the user access pattern library established in the pattern mining module.

A. System overall design

The system consists of five modules: data acquisition module, data preprocessing module, offline mining module, online recommendation module and model application module, as shown in figure 2.

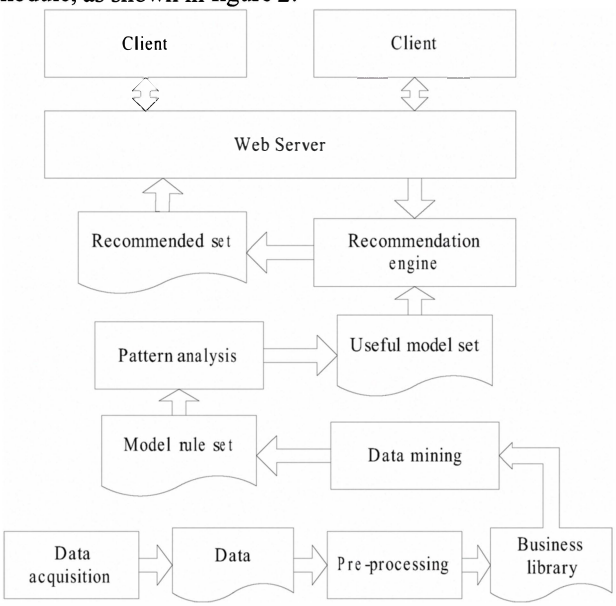

Figure 2. E-commerce recommendation system based on Web mining

The main function of the data acquisition module is to collect the original log data to form data acquisition libraries and prepare for future mining.

Data pre-processing module is mainly to pre-process the data in the data acquisition library to obtain a user transaction database, and a user transaction is the set of a group of meaningful browser pages of a user in a period of time.

Offline mining module is mainly composed by the mining engine, mining algorithm library, pattern discovery sub-module and pattern analysis sub-module. In which the mining engine uses the data mining technology in the mining algorithm library to discover user patterns, through the pattern analysis to analyze and interpret it to obtain meaningful page recommendation set, this system provides a variety of mining techniques such as association rules, service clustering and use clustering.

Online recommendation module sets the recommendation engine in the Web server front-end, which combines the current browse activities of the user and the mined page recommended set together to generate the corresponding recommendation set.

Model application module is to add the page links of the recommended set in the latest user request page, and then through the web server to deliver to the client browser [6], to provide users real-time personalized service; at the same time, the recommended results will be sent to the site management center to adjust the site design, optimize the site structure, thus to improve the site efficiency.

The system applies the data mining ideas and methods to the Web server logs for Web usage mining, to extract the user's access law, and then comes the online access users down to a certain class, according to the law of such user's access to carry out Web page recommendation, thus to provide users with personalized access.

B. System specific module design and implementation

1) Offline module

Offline module mainly consists of data pre-processing, Web mining and other modules, to provide support to the recommended engine, the system structure as shown in figure 3.

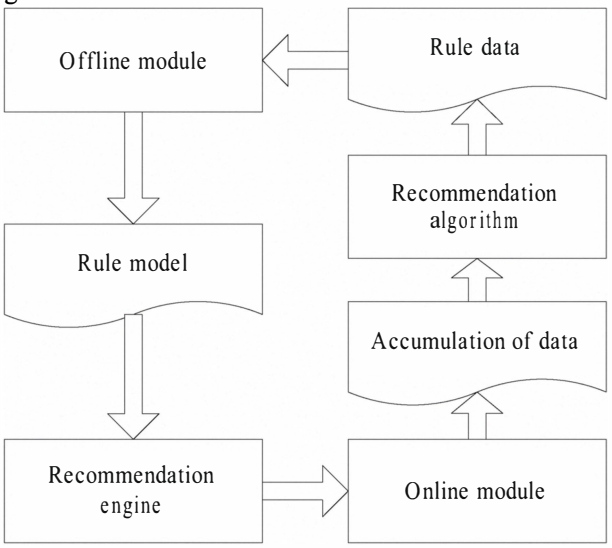

Figure 3. Offline module structure

Data pre-processing and Web mining have relatively large time cost, can not meet the real time requirement of personalized recommendation services, so the two operations must be carried out offline, while the mining results can be directly referred by the online recommendation engine.

Data preprocessing is to complete the work of two parts: delete the data irrelevant to the data source (Web log file), generate the user session documents and transaction database useful for the mining stage and generate site data files for the relevant documents in site. Pattern mining task is to use data

mining algorithms to carry out mining for the result transaction database of data preprocessing, to generate frequent user access page group, that is, user browse mode library [3] . The offline part needs periodical mining to update the user browse mode library.

In general, the offline part of the recommended system mainly aims to registered users, is to clean up the recommended set according to the key information users provided, to display the precise information of user interest in the recommend page set.

2) Online module

The online part mode is finally to provide customers with the best quality of recommendation services when the customer online browsing, so the design of the online part is crucial for the whole recommendation system.

Online part is composed by the pattern analysis tasks and the data source acquisition tasks. In which the pattern analysis tasks use the offline part to generate results and then online recommend the next step operation for users according to the current access behaviors of users, which consists of personalized recommendation engine and Web servers, personalization recommendation engine is to find the user's interest through analyzing the user's current access operation, and generate recommendation results through the recommendation algorithms, thus to achieve personalized services. The acquisition task of the data source is to collect the Web log files on Web server, and this process is automatically completed by configuring the server's corresponding parameters.

Online mode is divided into two situations, one is for registered user login, second is for random non-registered users. Registered users can change any interest, while the recommendation system based on the user's choice will form a recommended set and display the precise recommendation pages, and if the changed interest term does not include the key information of the registered users, then the recommended set will be generated from the original recommended set, so that the recommendation page users obtained will be more accurate.

For non-registered users, the recommended page is closely related to the user interest, while the precise extent largely depends on the user interest item selection, that is, the more the user interest item constraints the more accurate the recommended page [5] , but this has the cost of the user time consumption. The first page of random users includes all the categories and types. And after a series of choices, the recommendation system will finally generate a recommended page, and the more the selected interest items, the more precise the recommended page.

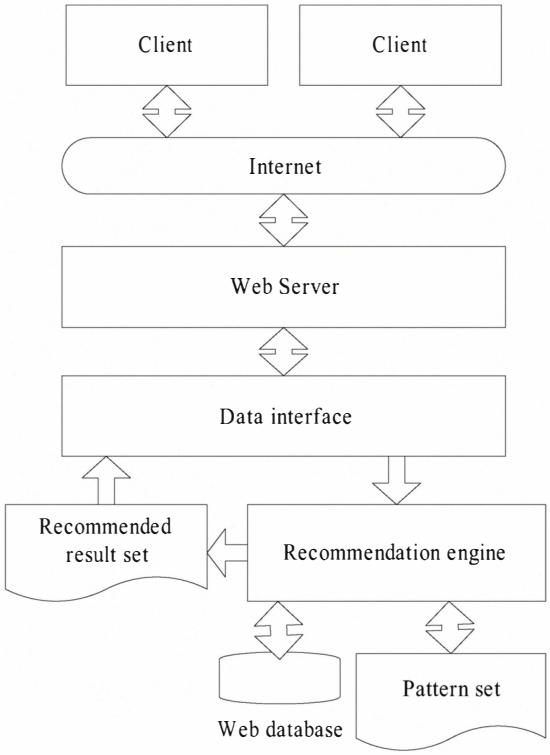

Figure 4. shows the structure diagram of the system online part

From the above figure it can be seen, when users visit the Web site the web page request will be first sent to the Web server to be submitted to the recommendation engine module, and then the latter will use the pattern library the offline module generated and Web database to calculate, generate recommendation result set to be returned to the Web server to return to the client through Http protocol, finally displayed in the browser. Thus the whole recommendation process is achieved. And from this process it can be seen that the entire calculation process is completed by the recommendation engine.

IV. CONCLUSION

At present, Web data mining has gradually become a hot issue in network research, data mining, knowledge discovery, software agent, and other fields. Log mining research for Web site optimization, e-commerce, distant education, information retrieval and other fields has great significance. However, how to complete these technologies and as soon as possible use them in a variety of Internet applications is a new topic we faced.

This paper studies the e-commerce recommendation system models based on Web data mining, designs a personalized recommendation system based on Web use mining for e-commerce, and discusses in detail the function of each module of the system and implementation techniques. The system can access to the user browsing patterns to provide personalized service for users through the mining of the historical log data of Web server.

REFERENCES

[1]. [Mobasher, B. Dai, Improving the Effectiveness of Collaborative Filtering on Anonymous Web Usage Data, ITWP01, 2001.

[2]. Zhen Liu, Minyi Guo. A Proposal of Integrating Data Mining and On-line Analytical Processing in Data Warehouse, In Proceeding of international conference on mechatronics and information technology, 2001:146-51.

[3]. Sarwar B.Karypis, G.Konstan, J., and Riedl, J. Analysis of Recommendation Algorithms for E-Commerce, ACM Conference on Electronic Commerce, 2000.

[4]. N.Colossi, W.Malloy, B. Reinwald Relational extensions For OALP, IBM Systems Journal [J].2002, 41(4):31-35.

[5]. CHEN Yu-ru, HUNG Ming-chuan, Don-lin YANG, Using Data Mining to Construct an Intelligent Web Search System [J], International Journal of Computer Processing of Oriental Languages, 2003.

[6]. S.Foucaud, A.Zanichelli, B.Garilli, M. Scodeggio and P.Franzetti. VIPGI and Elise3D: Reducing VIMOS-IFU Data and Searching for Emission Line Sources in Data Cubes. New Astronomy Reviews, In Press, Corrected Proof, and Available online.19 April 2006.