

Application and Research of E-commerce Recommendation System Based on Web Mining

Xu YAN¹, Bo SUN*

¹ Master of International Business, Guangdong University of Foreign studies, China

* Department of Research, Guangdong University of Foreign Studies, China

sbgz168@gdufs.edu.cn, yanxu0521@yahoo.com

Abstract - As a new branch of data mining, web mining emerge as social requires, it belongs to a new research direction. Web mining is to find potential useful model and hidden information from web documents. It combines the use of web and traditional data mining. The use of web mining in the e-commerce recommendation system can meet the demands of future development of e-commerce. This paper analyzes the background of the use of web mining in e-commerce recommendation system, and establishes a recommendation-system model based on veracity and efficiency. It also describes each module in the model exactly and how they connect with each other, and has a deep research on several important recommendation algorithms; At last it raises the problems and directions of the e-commerce recommendation technology.

Index Terms - e-commerce, web mining, recommendation system, problems and development

1. Introduction of Web Mining

1.1 Definition and characteristics of Web mining

Web mining is also known as Web data mining, is the application of data mining techniques to discover patterns from the Web, it involves Web technology, data mining, computer technology, information science and technology. Web data mining refers to the unknown, potential valuable, models which are found from a large number of Web document collections. Its handled objects include static web pages, Web database, Web structure, user records and other information. Through the data mining, information which cannot be gotten through the text retrieval can be obtained.

Compared with the traditional data mining, web data mining has the following characteristics: 1.The algorithm demands higher efficiency. As in quantity, the amount of Web-based data is much greater than the general warehouse or relational database.2.The requirement of data pre-processing is high. As the information on the Web is always semi-structured or unstructured, it is difficult to form a simple mapping of fixed pattern, which makes the information resources on the Web hardly be supported directly by database systems and the traditional data models, so the data is needed to be pre-treated to provide well-formed source data for mining in the next step.

1.2 Classification of web mining

The traditional data mining and text mining technology have been mature already, the development and accumulation of these techniques make the Web information resource mining become possible. Web data mining is a process using data mining technology to discovery and get information

automatically from Web documents and services, analyzing useful information on the Web, excavating the interesting, useful, unknown information or knowledge of a particular user. The processed information includes Web text, Web pictures, Web video, Web log and other various media information.

The diversity of information on the Web determines the diversity of the task of Web mining. There are lots of classification methods of Web data mining, such as the classification according to Web text language, the attribute of mining site and so on. According to different mining targets, it can be divided into three categories: Web Content Mining, Web Structure Mining and Web Usage Mining. These three types of mining have different emphasis.

1.3 Application of Web mining in e-commerce

In e-commerce, the main task of Web data mining is the mining of customers' access information, so as to obtain the access pattern and browsing behavior of customers, then find valuable market information. Data mining technology in e-commerce mainly includes statistical analysis, association rules, clustering analysis, sequence pattern, path analysis and so on.

The statistical analysis method is the most common method when extracting information from the Web site. The timing analysis tools of many Web browser will submit its statistical analysis report, which helps to improve the performance of the system, enhance the security of the system, and provide decision supports. Association rules is to find the relationship among various web-pages which customers once visited. For instance, through association rules, we can extract the relative relationship between the following pages: 60% users who visit A pages have also visited B page, and over 35% users who visit A pages and B pages have also visited C page. Using this association rules, network product supplier can better organize and edit the structure and content of the site, implementing more effective marketing strategies. Sequential pattern mining is the time series model which excavates transaction sets. Finding a sequence model to predict customers' access pattern, to promote the organization of e-commerce, and to carry out targeted advertising service. Through path analysis, it can improve the connections between web-pages and the interface design of the web structure.

2. Overview of E-commerce Recommendation System

2.1 Definition and composition of recommendation systems for e-commerce

The definition of recommendation systems for e-commerce is that providing product information and suggestion for customers through making use of electronic commerce website, in order to help users decide what to buy, simulating salesman to help customers complete the purchase process.

E-commerce recommendation system is main composed of three parts, they are input module, recommended methods module and output module. The input module is used to accept the information which is inputted by customers, and the most important is the users' rating data about the project. The recommended methods module is the core part of the e-commerce recommendation system, it provides the recommended project to the target customer according to a certain algorithm and user data. What's more, to find the good recommended method is a great part of the personalized recommendation research. The main mission of the output module is to decide the feedback types of the recommendation to users.

According to the project characteristics, there are two types of the recommendation system. One of them takes the web page as the target and adopts the web data mining method and technology to recommend the web page which user are interested in. The other personalized recommendation system takes the commodity as the recommended target in the online shopping environment, and recommend different valuable goods which users are interested in, like video, books, clothing, electrical appliances, etc, so this recommendation system is also called e-commerce personalized recommendation system.

2.2 The role of e-commerce recommendation system

E-commerce recommendation system's biggest advantage is that it can collect interested information to users, and take initiative to provide personalized advice based on preference of users. This recommendation dynamically updates with the users' preference information or the change of products in the system. This not only facilitates users, but also improves the service level of the enterprise.

Generally speaking, the role of e-commerce recommendation system mainly has the following several aspects: on the one hand, it helps users get rid of the limitless network resources and commodities world, which greatly saves their time and commodity procurement cost; on the other hand, the personalized recommendation service effectively improves the loyalty of customers to e-commerce sites. Consequently, the Cross-Selling ability was improved, and e-commerce enterprises obtained more profits and development opportunities.

Practically, compared with the company without e-commerce recommendation system, their sales are increased by 2% - 8%. It is more obvious especially in the movie, CD audio and video, books and daily provisions, whose products are various and at relatively low-price. In this way, application

of recommendation system can greatly improve the enterprise's sales performance.

2.3 Research content of e-commerce recommendation system

Research on the personalized recommendation of e-commerce has four aspects. First of all, the premise of personalized recommendation is to solve the information sources of recommendation system and to express with the right way. The basic recommendation system is the user's personal information and interest information, how to get and effectively handle this information is the key. Secondly, to make the personalized recommendation results accepted and recognized by customers, the core is the accurate and efficient personalized recommendation algorithm. What's more, we must make a comprehensive, objective evaluation of the recommendation system, making it accepted by majority of users. The research goal is the application of recommendation system, recommendation system should provide not only complete personalized shopping environment for users, but also necessary support for customer relationship management and sales decision of enterprises.

3. Algorithm of E-commerce Recommendation System

The core of the recommendation system is the e-commerce recommendation algorithm, its recommendation accuracy and efficiency will directly affect the overall performance of the recommendation system. The typical recommendation algorithms are the recommendation algorithm based on association rule and the collaborative filtering recommendation algorithm. This chapter will discuss these two algorithms.

3.1 The recommendation algorithm based on association rule

Association Rule Mining (ARM) is an important branch of data mining, which is the most typical one in many types. Applying association rules algorithm for the mining of Web access information, we can find the users' access patterns, buying patterns and so on. These patterns can help website designers to better adjust the structure of websites.

3.1.1 Association rule mining algorithm

Association rule mining algorithm is used to reveal the unknown links between data. The design of association rule mining algorithm is ascribed to the following two questions: (1).Find the item sets which the degree of support is greater than or equal to the minimum support, these item sets are called Frequent Item Sets. (2).Using the above (1) to find the frequent item sets, producing the desired planning. Most studies on the association rule follow the above two steps. Because of (2) do not need to read the information from the database, its computation quantity is less, so the emphasis of association rule mining research has been placed on (1).

3.1.2 Application and Realization of association rule mining model

Association rule mining is to count how many people who had bought goods A bought goods B in transaction database. Using association rule mining, we can take the initiative to recommend high correlation products to users, in this way,

sales can be improved, and the customers' loyalty can also be increased. E-commerce recommendation is automatically formed by the recommendation system, which do not need human intervention. The recommendation algorithm based on association rule can be divided into off-line and on-line stages. The off-line stage is time-consuming, it was placed in the offline module; on-line stage provides real-time recommendation services to users according to the users' purchase behavior and the established association rule recommendation model.

3.2 Collaborative filtering recommendation algorithm

3.2.1 Overview of collaborative filtering recommendation algorithm

Collaborative filtering recommendation algorithm is the most successful e-commerce recommendation technology currently, it has been widely applied to many fields. According to the users' registration information and purchasing behavior, collaborative filtering establishes users' behavior model, and then recommends valuable commodity to the users. The achievement of collaborative filtering is generally divided into two steps. At first, getting the users' information, then analyzing the similarity relation between users and predicting users' data.

3.2.2 Collaborative filtering recommendation algorithm based on users

Collaborative filtering recommendation based on users generates recommendation list according to other users' point of view which has high similarity with the target users. The recommended method assumes that if users' evaluations of some projects are similar, and they have similar evaluation of other projects. Firstly it uses statistical techniques to search several neighbor users who have the similar interests with target users, then recommends the highest rated recommendation of neighbor-users to current users.

3.2.3 Nearest neighbor query

Looking for the nearest-neighbor for current users is the core of collaborative filtering recommendation system. For a user X, to produce a neighbor set $N = \{N_1, N_2, \dots, N_t\}$ whose similarity is from small to large by calculation, X does not belong to N, from N_1 to N_t , the distance $d(X, N_k)$ is arranged from small to large. $d(X, N_k)$ is the distance between user X and N_k , can be used to express the similarity between them. There are many ways to measure the similarity, we usually use distance to measure the similarity between users. There are three distance functions as following:

(1) Minkowski distance

Assuming that x, y is the corresponding feature, n is the dimension. The measurement of Minkowski distance of X and Y is as following:

$$d(x, y) = \left[\sum_{i=1}^n |x_i - y_i|^r \right]^{1/r}$$

When r is changed, the measurement formula can evolve into some special distance measurement.

When r=1, Minkowski distance will evolve into the absolute distance:

$$d(x, y) = \sum_{i=1}^n |x_i - y_i|$$

When r=2, the Minkowski distance will evolve into the Euclidean distance:

$$d(x, y) = \left[\sum_{i=1}^n |x_i - y_i|^2 \right]^{1/2}$$

(2) Quadratic distance

The measure form of Quadratic distance is as following:

$$d(x, y) = \left((x - y)^T A (x - y) \right)^{1/2}$$

Among them, A is non negative definite matrix.

When A has different value, the above distance measurement formula will evolve into some special distance measurements. When A is unit matrix, the quadratic distance will evolve into the Euclidean distance;

When A is diagonal matrix, the quadratic distance will evolve into the weighted Euclidean distance:

$$d(x, y) = \left[\sum_{i=1}^n a_n |x_i - y_i|^2 \right]^{1/2}$$

When A is covariance matrix, the quadratic distance will evolve into the Mahalanobis distance.

(3) Cosine distance

The measure form of Cosine distance is as following:

$$d(x, y) = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2}}$$

Among them, the euclidean distance is the most commonly used

4. The Existing Problems and Prospects of E-commerce Recommendation Technology

4.1 Existing problems of e-commerce recommendation technology

The e-commerce recommendation technology has got more and more application and development. However, because of the limitation of the research level and the the complexity of e-commerce websites, there still exists many problems in recommendation systems:

(1) Most current recommendation systems need artificial participation, thus leading to low accuracy in the recommendation systems.

(2) Most recommendation systems require the evaluation of the systems when customers browse the web sites, this requires customers' high coordination to obtain the information.

(3) Most personalized recommendation systems just recommend some different resources to the users, which is similar to ordinary search engine, and it cannot react exactly customers' correct behaviors most times.

(4) Most current e-commerce recommendation systems only apply to single e-commerce website, and it lacks effective recommendation result interpretation model and evaluation system.

4.2 Developing prospects of e-commerce recommendation technology

Combining the existing problems of the e-commerce recommendation system, the future development of e-commerce recommendation technology research mainly embodied in the following aspects:

(1) Research of the combination of recommendation technology. Because all kinds of technologies have some defects, an important issue of the future e-commerce recommendation technology research is whether various recommendation technologies can make complement each other.

(2) Research of the accuracy of recommendation technology. Because of the accuracy of current recommendation technology is not guaranteed and the research is limited in the stage of manual experiment. So it is very important whether we can develop the automatic accuracy verification theoretical model.

(3) In terms of how to get the users' data, the existing recommendation system mainly depends on the users' explicit evaluation, therefore, future research should pay more attention to how to get the users' implicit information automatically.

(4) The system should improve its adaptability. When a user's interest shifts, it can still provide accurate recommendations through a certain way.

(5) The recommendation system should study the solution of the improvement of the algorithm performance, paying more attention to enterprise business intelligence, customer

relationship management and how to improve the market sales of enterprises.

5. Conclusion

At present, an important problem that e-commerce enterprise needs to solve is how to attract more new customers and improve the loyalty of old customers under the virtual e-commerce environment, to help customers find needed or interested items quickly and effectively. However, the e-commerce recommendation system construction of our country is still in the research and exploration stage. Compared with foreign countries, there is still a wide gap.

This research applies web mining technology into the e-commerce system, digs out the trend of network users' interest, and through the collaboration of all modules in the system and the application of various algorithms, gets the accuracy and efficiency of recommendation which can effectively improve the sales ability of e-commerce website and enterprise. It has very important practical significance.

Acknowledgment

This work was supported in part by a grant from Guangdong province science and technology planning project. (2013B040404009)

References

- [1] Li Xingxing. The Research of E-commerce Recommendation System. Computer Engineering and Science.
- [2] Xu Baowen, Zhang Weifeng. Research on Application of Data Mining Technology in Web. Chinese Journal of computers, 2001, Vol.28, No.3:123-127.
- [3] Ji Xiyu. Examples of the Application of Data Mining. Bei Jing: Machinery Industry Press
- [4] Xu Min, Qiu Yuhui. Problems and Countermeasures of E-commerce Recommendation System. Computer Science, 4th 2001.
- [5] Liu Pingfeng, Nie Guihua, Chen Donglin. Review on The Research of E-commerce Recommendation System. Journal of Information, 9th 2007
- [6] Zhang Wei, Liu Yongguo, Peng Jun, Liao Xiaofeng, Wu Zhongfu. 2001. Research on the Development of Data Mining. Computer Science, 28 (7): 79-81.
- [7] Wu Jiansheng. 2001. Research on Web log mining browsing patterns in customers. Journal of Xi'an Jiao Tong University, 35 (6): 621-624.
- [8] Qiu Ming. Electronic Commerce Practical Tutorial. Bei Jing: Higher Education Press, April 2004.
- [9] Zhou Xianchun. Electronic Commerce and Web Data Mining. Computer application, April 2000.
- [10] Chen Jingmin. Data warehouse and data mining technology. Bei Jing: Publishing House of Electronics Industry, 2003. 258-259.
- [11] L. Wu, P. S. Yu, A. Ballman. Speed Tacer: A Web usage mining and analysis tool. IBM System Journal. 37 (1):89-105.1998.