A Web Mining Based Network Personalized Learning System

Hua PANG[1, a], Jian YU[1], Long WANG[2, b]

[1]College of Education Technology, Shenyang Normal University, Shenyang, 110034, China

[2]College of Information, Liaoning University, Shenyang, 110036, China

[a]email:panghua@sohu.com, [b]email:email_wl @163.com

Keywords: Network Learning; Personalized Service; Web Usage Mining; Web Content Mining

Abstract. In order to solve the problems of the traditional network learning system, a network personalized learning system based on web mining is presented in this paper. The web mining technology is simply introduced. The system model is given and the work process of this model is depicted in detail. For improve the intelligence or the system, the personalized information extraction method based on web usage mining and the learning resources recommendation method based on web content mining was proposed. Experiments on the real world dataset show that the system meets the student's demands of personalized learning.

Introduction

With the constant development of the distance education, people pay more and more attention to network education. With the application of the network learning system, the teaching method was enriched and the teaching space was enlarged. The research of network learning system is an important field of the distance education [1].

At present, the network distance education is on the condition of resource sharing, and transplants the traditional classroom education to distance education simply. The traditional teaching mode of network learning system is invariable, and does not consider the characteristics of distance education and the requirements of students. The old systems do not accord with the law that the students are the center of distance education and lack intelligence, so the students could only passive learn the same contents in these systems. The problems can be solved with the application of Web mining in the network learning system.

Web Mining Technology

Web mining [2, 3] is the application of data mining techniques to discover patterns from the World Wide Web. Web mining can be divided into three different types: Web usage mining, Web content mining and Web structure mining.

Web Usage Mining

Web Usage Mining [4, 5] is the application of data mining techniques to discover interesting usage patterns from Web data in order to understand and better serve the needs of Web-based applications. Usage data captures the identity or origin of Web users along with their browsing behavior at a Web site.

Web usage mining itself can be classified further depending on the kind of usage data considered:

(1)Web Server Data: The user logs are collected by the Web server. Typical data includes IP address, page reference and access time.

(2)Application Server Data: Commercial application servers have significant features to enable e-commerce applications to be built on top of them with little effort. A key feature is the ability to track various kinds of business events and log them in application server logs.

(3)Application Level Data: New kinds of events can be defined in an application, and logging

can be turned on for them thus generating histories of these specially defined events. It must be noted, however, that many end applications require a combination of one or more of the techniques applied in the categories above.

Web Structure Mining

Web structure mining [6] is the process of using graph theory to analyze the node and connection structure of a web site. According to the type of web structural data, web structure mining can be divided into two kinds:

(1)Extracting patterns from hyperlinks in the web: a hyperlink is a structural component that connects the web page to a different location.

(2)Mining the document structure: analysis of the tree-like structure of page structures to describe HTML or XML tag usage.

Web Content Mining

Web content mining [7] is the mining, extraction and integration of useful data, information and knowledge from Web page content. The heterogeneity and the lack of structure that permits much of the ever-expanding information sources on the World Wide Web, such as hypertext documents, makes automated discovery, organization, and search and indexing tools of the Internet and the World Wide Web such as Lycos, Alta Vista, WebCrawler, ALIWEB, Met Crawler, and others provide some comfort to users, but they do not generally provide structural information nor categorize, filter, or interpret documents. In recent years these factors have prompted researchers to develop more intelligent tools for information retrieval, such as intelligent web agents, as well as to extend database and data mining techniques to provide a higher level of organization for semi-structured data available on the web. The agent-based approach to web mining involves the development of sophisticated AI systems that can act autonomously or semi-autonomously on behalf of a particular user, to discover and organize web-based information.

Network Personalized Learning System

System Architecture

In traditional network learning system, learning support system and educational resource database exchange data simply. Students can only make autonomous learning through the simple browsing and query function provided by the system, and the learning process lacks intelligence. There are large amount of learning resources in network learning system. Generally, the student doesn’t know if it contains interested contents when he studies new resource. So, he can only check the content or the resource description. If it doesn’t contain the interested content, the checking is useless. In order to reduce the useless checking, the recommendation service [8] should be added to the system. The system forecast the new content or resource whether the student is interested in by analyzing the student personalized information such as his previous learning content, and give the prediction results to the student as a feedback, and on that basis recommend the possible interested content in the teaching system to the student. The system architecture and the relationship among the models are shown in Figure 1.

As shown in Figure 1, when students use the network learning system, some Web logs data are generated. These data include user name, access date, access times, access time, requested URL and so on. Through mining the data, the personalized information of students are obtained. Put the personalized information which as a basis for personalized learning resources recommendation into the student personalized information database. When a student use the system, the service can extract learning information from the learned before based on his personalized information, and train the recommendation model. Through the recommendation model, the service can find the learning resources which the student interested in current subject and recommend them to the student.

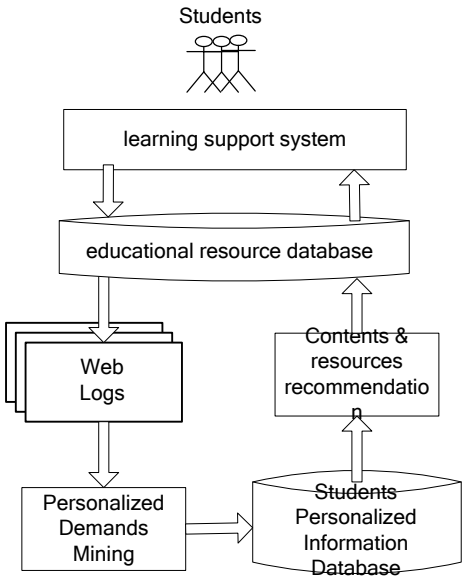

Fig.1 System Architecture

Personalized Information Mining

For a student, what learning resources they are interested in is the most easy and effective way to reflect his personalized demands. From the Web logs, the service can find out the learning resources that the student has browsed and analyze what they are interested in, what they are uninterested in, and then compose the personalized information of the student.

Now, browsing interest measure [9] is widely used to estimate whether the contents the users are interested in. The browsing interest measure P is computed as:

$$P = \frac{C \times T}{B} \tag{1}$$

where C is the browsing times, T is the each browsing time, and B is the number of bytes of the browsing content.

In network learning system, the file types of learning resources are not only text but also image, animation, audio and video. In formula (1), the difference from different file types aren't considered. In addition, the learning resources not only include the resource content, but also include the resource description which is called the brief of resource. According to the characteristics of learning resources, the interest measure of one resource is computed as follow:

$$P = \eta \times \left(\alpha \times \sum_{1}^{n} \frac{T_1}{B_1} + \beta \times \sum_{1}^{m} \frac{T_2}{B_2} \right) \tag{2}$$

where n is the browsing times of the resource content, T_1 is the each content browsing time, B_1 is the number of bytes of the resource content, α is the coefficient of resource content, m is the browsing times or the resource brief, T_2 is the each brief browsing time, B_2 is the number of bytes of the resource brief, β is the coefficient of resource brief, η is the coefficient of resource type.

If $P > P_{\min}$, the learning resource is interested, otherwise the learning resource is uninterested.

Resource Recommendation Method

In this paper, the KNN [10, 11] method is used to recommend the learning resources. Firstly, the all visited resources were divided into two different set: interested resource set and not interested set. For a new resource, the system computes the probability of belonging to interested resources P_1 and the probability of belonging to not interested resources P_2. If $P_1 > P_2$, then the student is interested

in the new resource, otherwise the student is not interested in the new resource.

The detailed algorithm is: the feature representation of the new resource is d, the sample space S has two sorts: interested resources m_1 and not interested resources m_2. The similarity is:

$$Sim(d, s_i) = \frac{\sum_{j=1}^{n} d.t_j \times s_i.t_j}{\sqrt{(\sum_{j=1}^{n} (d.t_j)^2) \times (\sum_{j=1}^{n} (s_i.t_j)^2)}} \tag{3}$$

where n is the amount of features, $s_i \in S$, $d.t_j$ is the weight of the features j in d, $s_i.t_j$ is the weight of the features j in s_i。Composed the neighbor set with k resources which have the bigger similarity, the probability P_1 and P_2 are:

$$P_l = \sum_{S_i \in K} Sim(d, s_i) y(s_i, m_l) - b_l \quad l = 1,2 \tag{4}$$

where $y(s_i, m_l) \in \{0,1\}$, if $s_i \in m_l$, $y(s_i, m_l) = 1$, else $y(s_i, m_l) = 0$; b_l is the threshold of m_l.

Experimental Evaluation

Select the learning logs of 10 students as the experimental data from a network learning system without recommendation service. The learning resources which are placed in “favorite” by students are used as the interested resources, the learning resources which are placed in “recovery” by students are used as the uninterested resources. Firstly test the effectiveness of the personalized information mining, the experimental results are shown in table 1.

Table.1 Personalized Information Mining Results.

student	accuracy rate of interested resources	accuracy rate of uninterested resources	total accuracy rate
1	0.854	0.861	0.858
2	0.863	0.852	0.858
3	0.851	0.863	0.857
4	0.862	0.871	0.867
5	0.891	0.882	0.887
6	0.874	0.865	0.870
7	0.851	0.853	0.852
8	0.860	0.864	0.862
9	0.852	0.871	0.862
10	0.859	0.856	0.858
average	0.862	0.864	0.863

It is shown in table1 that the average accuracy rate can reach 86.3%, and these meet the demands of personalized learning resources recommendation.

Then apply the data set that achieved from personalized information mining to test the resources recommendation method, The experimental results are shown in table 2.

Table.2 Resources Recommendation Results.

student	accuracy	precision	recall
1	0.812	0.822	0.803
2	0.823	0.815	0.812
3	0.832	0.821	0.809
4	0.815	0.831	0.821
5	0.821	0.824	0.811
6	0.831	0.818	0.807
7	0.822	0.833	0.815
8	0.834	0.811	0.813
9	0.811	0.809	0.808
10	0.822	0.820	0.822
average	0.822	0.820	0.812

It is shown in table2 that the accuracy, precision and recall of the method are 82.2%, 82.0% and 81.2%, and these meet the demands of personalized learning resources recommendation.

Conclusion

In this paper, we apply Web mining technology to network learning system, and intelligent recommendation model based on Web content mining and Web usage mining effectively improves the intelligence of the network learning system.. And the system can provide the student personalized demands intelligently and satisfy the personalized network education.

Acknowledgement

This paper is sponsored by the Doctoral Scientific Research Foundation of Liaoning Province (GN: 20141049), the Science and Technology Development Program Funds of Liaoning Province (GN: 2012216007).

Reference

[1]Long Wang, Shaochun Zhong, Dongdai Zhou, Xiaochun Cheng, Jinan Li. A Distance Education System Based on Web Services [J].Journal of Computational Information Systems, 2006, 03:203-213.

[2]J Han, M Kamber. Data Mining: Concepts and Techniques[M]. Beijing: Higher education Press, 2001

[3]Kolari, P., Joshi, A.: Web mining: research and practice . Computing in Science and Engineering, vol.6, pp. 49-53. (2004)

[4]Nasraoui O., Frigui H., Joshi A., and Krishnapuram R.: Mining Web Access Logs Using Relational Competitive Fuzzy Clustering, Proceedings of the Eighth International Fuzzy Systems Association Congress, Hsinchu, Taiwan, August 1999.

[5]Jaideep Srivastava, Robert Cooley, Mukund Deshpande, et al. Web usage mining: discovery and applications of usage patterns from Web data [J]. Appear in SIGKDD Explorations, 2000(2), 01:12-23

[6]Eirinaki, M., Vazirgiannis, M.: Web Mining for Web Personalization, ACM Transactions on Internet Technology, Vol.3, No.1, February 2003.

[7]Srivastava, J., Desikan, P., Kumar, V.: Web Mining: Accomplishments and Future directions. In: National Science Foundation Workshop on Next Generation Data Mining. Baltimore, Maryland(2002).

[8]Xin Jin, Yanzan Zhou, Bamshad Mobasher. A Maximum Entropy Web Recommendation System: Combining Collaborative and Content Features[C]. International Conference on Knowledge Discovery and Data Mining. 2005:612-617

[9]Pierrakos, D., Paliouras, G., Papatheodorou, C., Spyropoulos C. D.: Web usage mining as a tool for personalization: a survey, User modelling and user adapted interaction journal, Vol.13, Issue 4, pp. 311–372.

[10]Sutheera Puntheeranurak, Hidekazu Tsuji. A Multi-clustering Hybrid Recommender System[C]. Proceedings of the 7th IEEE International Conference on Computer and Information Technology. 2007:223-228

[11]K Wang, HQ Liu. Discovering association of structure from semi-structured objects [J]. IEEE Transactions on Knowledge and Data Engineering, 2000(3), 12:353-371.