

The Application of Web Mining in Distance Education Platform

Yuanyuan Liao

Institute of Information Science and Engineering, Xinjiang University,
Shenglistr. 14, 830046, Urumqi, Xinjiang, China
liaoyuan@xju.edu.cn

Abstract—Although the rapid development of the modern distance education, there is still some shortcomings. In order to overcome these shortcomings and use a large number of data in these remote teaching sites effectively, it is necessary for the introduction of data mining technology to the modern remote education system. The paper introduced distance education system based on Web mining, and how data mining of these distance education website.

Keywords—Web mining; Distance Education

Modern distance education is the common trend of the development of education in the world. With the development and application of satellite and cable TV, as well as a variety of electronic communications technology, especially with the advancement of the global computer network and multimedia technology, many universities have established distance education platforms. Because of a large number of students, we are relative lack of educational resources nowadays. For this reason, our country is actively developing modern distance education and integrating the various types of educational resources by advanced information technology and educational technology.

I. THE DISADVANTAGES OF EXISTING DISTANCE EDUCATION PLATFORM

The construction of the remote education platform has set up a solid foundation for college's distance education in our country. Modern Distance Education has made great achievements, but there are also some disadvantages as below:

A. Dynamic Interactive Features

At present, distance education system just a pile of teaching materials, students can only look at teaching content without actually participating in the study actively, which making it difficult to achieve the best learning effect.

B. System with Low Intelligence

First, most systems could not provide suitable intelligent interface according to the specific circumstances of the students, which reduce the learning interest of the students. Second, most systems could not provide appropriate guidance based on the student's actual capacity and learning level, which due to unified teaching.

C. System with Low Level Test and Evaluation Functions

There are only some fixed Web pages with some fixed questions in online test system nowadays. The topics on test paper could not be set up automatically in accordance with

the requirements of the students and the learning situation. Most systems with poor flexibility, could not based on the weaknesses in students' knowledge system. In addition, there are only several levels of difficulties and evaluation criteria.

In addition, a large number of useful teaching information in these remote teaching sites, such as user access logs, registration information, answering questions, learning progress, could not be used effectively. In order to effectively use these data, at the same time, overcome the above shortcomings, it is necessary for the introduction of data mining technology to the modern remote education system. Web mining is used for distance education system can solve the following problems: personality mining system improvements, site modifications, intelligence services, and page recommendation.

II. WEB MINING

Web mining is interested in extracting useful patterns and implicit information from the www resources. Data mining technology is used to discover and extract information which people interested in automatically from Web documents and Web services. Web mining technology is applied to distance education site to find useful information and mode, and analysis the students' learning behavior and learning mode. Essentially, it provides personalized learning environment for distance education students by this way. All in all, Web mining is very useful to distance education.

According to the different objects of data mining, the web data mining is divided into three categories in general: web content mining, web structure mining and web usage mining.

A. Web Content Mining

Web Content Mining technology is the process of founding useful information from the network, data and documentation. Www information resources have become the main body of the network of information resources, at the same time, the network information data is made up of text, image, audio, video, and metadata, so, Web content mining is a multimedia form of Web data mining.

B. Web Structure Mining

Web Structure Mining is mining the Web's the potential link structure mode. This idea stems from citation analysis which is to analyze a page of links, the number of links and objects to create a Web link structure model. The Web structure mining help users find authority site and related topics, so that the user can not only find the useful

knowledge in page, but also discover the knowledge contained in the structure of the page.

C. *Web Usage Mining*

Web Usage Mining has significance for understanding network behavior data. It faced second-hand data is extracted in the process of user and network interaction. These data are mainly left by the users to access the Web in a Web log information, as well as some of the other interactive information, including: network server to access the records, proxy server logging, browser logging, user profiles and registration information, user dialogue or transaction information, user access date. Web Usage Mining is to find and analyze learning information from all kinds of students' log as a basis to provide individualized learning services to students.

III. THE PROCEDURE OF WEB MINING

A. *Pretreatment*

Pretreatment is the most critical aspect of Web mining, and its quality is related to the quality of the mining process and model analysis process. Pretreatment is divided into data cleaning, user identification, session identification, path supplement and event recognition.

1) Data Cleaning

Its purpose is to remove the contents unrelated with data analysis and data mining in log file, such as: removing the records of user failed requests.

2) Subscriber Identity

This is the second step of pretreatment. Because the log file only records the host or the IP address of the proxy server, it is used a cookie technology, or using some heuristic rules to help identify a user.

3) Session Identification

In large Web server logs across a long time period, the user may visit the site many times. Session identification is to split the user's access records into several single sessions. Generally it is used a timeout to identify, if the time exceeds a certain interval between user requests a page, the user starts a new session.

4) Path Supplement

It is to confirm whether some important Web log records are missing. The tasks of the path supplement is to add these missed requests to the user session file, it can also be supplemented path based on the information provided by the reference log network topology.

5) Event ID

Event ID is to define of event for the specific needs of the mining activities, which is related to the knowledge of what the user needs. There are two ways of Events ID, which are reference length and the maximal the forward reference.

B. *Mining Process*

1) Associated Mining Technology

The association mining technology is knowledge discovery technology that a database user access to the knowledge of the sequence association rule. Association

rules for Web data mining is the discovery of the rules of contact between the visited pages in the user's session, these pages do not need to have a direct reference to the relationship. The most common association rule is Apriori algorithm. The association rule will be time-consuming because that Web site has the huge amount of data. A focus for researchers now focuses on the technical reduction of the search space. It can be used a combination of association rules and frequent itemsets mining method to search for related webs that users access a session simultaneously in habits.

2) Cluster Analysis

The cluster analysis is used to combine the user with similar characteristics and data items together. The most common way in Web usage mining areas, are the form of target user clustering and web clustering. Target user clustering is to classify user by the standard of browsing mode. User classification can help us to provide personalized information services for different users. The web clustering can help to establish a group of pages with related content.

3) Classification

Classification rules can tap some of the similarities and differences between Web data, and the new data added to the database items is classified in the appropriate category because of these similarities and differences. The Web mining classification techniques could get the characteristics of the users of a Web page to access the server by a variety of users' personal information found before and common access mode of all kinds of user owned. In addition, User characteristics can be obtained through user registration information on the website. By using classification learning technology, it can be used in the Web mining process to obtain a brief description of the types of users. The common classification methods are naive bayesian method, decision tree, k-nearest neighbor method, support vector machines, neural network method, and so on.

IV. DISTANCE EDUCATION SYSTEM BASED ON WEB MINING

The architecture of the network teaching platform is composed of three parts generally, teaching resource library, learning platform and user. Education Resource Library is a file server to store various types of educational resources; Users are the learners; learning platform is a Web server that presented Web-teaching resources to users, and then users can choose learning resources freely. The web-based distance education system improve the teaching platform based on Web mining. It can provide teaching resources according to the learner's personality characteristics.

The design of the Web-based distance education system is to analysis of Web logs and site files, combined with the personal information of the learners, learning outcomes, learning behavior, and use data mining techniques to derive valuable model and as a basis for improving distance education system to meet the learning needs of different learners. Web-based distance education website is combined with the learner's personal information, learning outcomes, learning behavior by the analysis of Web logs and site files,

and discover useful access patterns and rules by using data mining techniques, and store the valuable patterns after mode analysis in the database of personalized, so, site designers improve the site structure and provide a basis for modifying the teaching content for teachers, and better personalized service and intelligence services for learners to meet the different learning needs of the learners.

V. SUMMARIZES

Web mining technology is used in distance education to analyze multi-dimensional and mining of user learning process, and establish the personalized distance education environment model based on Web mining. It takes full advantage of the accumulated information to help course designers and managers re-design courses and reconstruction of the site. Based on Web mining technology, the design of the curriculum of distance education is more reasonable and more in line with the rules of distance learning and remote teaching, so that students can give full play of their own personality, and the distance education become truly individualized and personality education.

REFERENCES

- [1]. Esposito Dino, Luo Xiaoping. The Discussion JavaScript Library-Dojo, JQuery and Prototype Comparison [J]. Programmer, 2008 (8): 105-107.
- [2]. Jiawei Han, Micheline Kambr. Data Mining Concepts and Techniques, Beijing: Higher Education Press. 2001.5
- [3]. Shubei, Rui-Min Shen, Wang Jiajun. Personalized Distance Learning Model. Computer Engineering and Applications. 2001.9
- [4]. Wang Qijun, Rui Min Shen. Distance Education Model Based on the Web Mining Intelligent, Personalized Computer Engineering. 2000.12.
- [5]. S.Brin and L.Page.The anatomy of a large - scale hypertextual Web searchengine In Senenth International World Wide Web Conference,Brisbane ,Australia,1998
- [6]. Yang Dehua. Personalized Distance Learning System Model [J]. Computer Engineering and Application.2003,39(28):118-121.
- [7]. Rui-Min Shen, Xin Dai, Sun Jian. Based WEB Intelligent Distance Learning Environment. Computer Applications and Software, 2004,21(2):36-38.
- [8]. R. Cooley, B. Mobasher, and J. Srivastava. Data preparation for mining world wide web browsing patterns. Knowledge and Information Systems, 1999.1(1).