

A novel teaching model based on Web Mining and Data Mining

Feng Liu

School of Computer Science and Technology
Nanyang Normal University
Nanyang, China
63699166@qq.com

Jinjiang Liu

School of Computer Science and Technology
Nanyang Normal University
Nanyang, China
nytc@sina.com

Abstract—Aiming at the characteristics of current distance teaching sites with few interactive means and inability to teach students in accordance with their aptitude, a remote teaching model based on Web mining and data mining is proposed. And the realization algorithm is given. This model makes full use of the various information accumulated on the site to help course designers and managers to redesign courses and reconstruct sites, so that the design of courses is more reasonable and in line with teaching regulations. Learners can also give full play to their own personality. The model, compared with the WWW-based one, is more beneficial to distance education, and has broad development and application prospects.

Keywords—Teaching model, Web mining, Data mining, Algorithm

I. INTRODUCTION

With the development of Internet and Web technology, distance teaching technology, one of the hotspots in recent years in network research and application, has overcome the space and time limitations of traditional teaching and made the original scattered teaching resources integrated. As a new supplement to school education, distance teaching system reflects the characteristics of teaching process such as interactivity, fast updating of teaching content, full use of teaching media and so on.

However, the WWW-based distance learning sites have two disadvantages: (1) The existing distance teaching sites are basically static, designers do not know whether the design of the courseware is reasonable and whether it conforms to the teaching rules, Whether the hypertext links between pages conform to the user's visiting habits; (2) Existing distance teaching sites cannot teach students in accordance with their aptitude, and students cannot learn on demand.

Therefore, the distance teaching system is basically just a copy of book teaching to electronic teaching. However, a large amount of useful teaching information has been accumulated on the site, the information such as users' access logs, registration information, Q&A information, test scores, homework and learning progress is not fully utilized, resulting in a waste of resources. In this situation, a new technology is needed to make full use of the information to build an intelligent and personalized distance learning environment.

II. ANOTHER THE DISTANCE TEACHING SYSTEM BASED ON WEB MINING

A. Introduction to Web Mining Development

Web mining is the analysis of the content of documents, the use of available resources, and the relationships between resources to discover effective, novel, potentially valuable, and ultimately understandable patterns and rules. There are two main Web mining technologies currently researched: Web Usage Mining and Web Content Mining.

Web Usage Mining (Web access information mining) is to mine the access records left by the server when the user accesses the Web, that is, to mine the access mode of the user accessing the Web site to find the browsing pattern of the user accessing the site and the frequency of page access, and other information; the objects to be mined are logs on the server including Server Log Data and so on.

The mining methods include path analysis; discovery of association rules and sequence patterns; clustering and classification. Web Content Mining (Web Content Mining) is the mining of Web page content. The content mining can be divided into two categories: mining of text documents (including TEXT, HTML and other formats) and mining of multimedia documents (including media types such as Image, Audio, Video, etc.). Web text mining can summarize, classify, cluster, associate and analyze the content of a large number of document collections on the Web, and use Web documents for trend prediction. (1) Document classification refers to determining a category for each document in the document collection according to a predefined subject category. In this way, users can not only browse the documents conveniently, but also make it easier to find documents by restricting the search scope. (2) The difference between document clustering and classification is that clustering does not have a pre-defined topic category. Its goal is to divide the document collection into several clusters, and the similarity of the document content in the same cluster is required to be as large as possible. The similarity between different clusters is as small as possible. (3) Association analysis refers to finding out the relationship between different words from a collection of documents.

Now Web mining technology is being used in all aspects of WWW, such as e-commerce, E-mail automatic reply,

personalized website design, Web report and so on. A distance teaching model based on Web mining and data mining will be proposed here. It can make full use of the rich information accumulated on the site to better serve distance teaching. Compared with the WWW-based distance teaching model, the main advantages of this model are: (1) Provide the user's browsing mode for the course designer, reconstruct the links between the pages to conform to the user's visiting habits; (2) Provide teachers with information about students' learning of courseware content, and adjust the entire course in a timely manner to make it conform to the teaching rules; (3) Provide different learning content or learning progress according to the students' own conditions to teach students in accordance with their aptitude; (4) According to the students' mastery of the learning courses

degree, offering something deeper or shallower than the course.

B. Distance Teaching Model Based on Web Mining and Data Mining

The distance teaching model based on Web mining mainly uses Web mining technology to find out the patterns and rules that users are interested in by analyzing a large amount of information accumulated on the site, so as to provide curriculum designers and managers with information on improving curriculum design and emphasis. It provides information about the students' learning courseware for teachers, so as to establish an intelligent and personalized distance education environment. The structure of the model is shown in Figure 1.

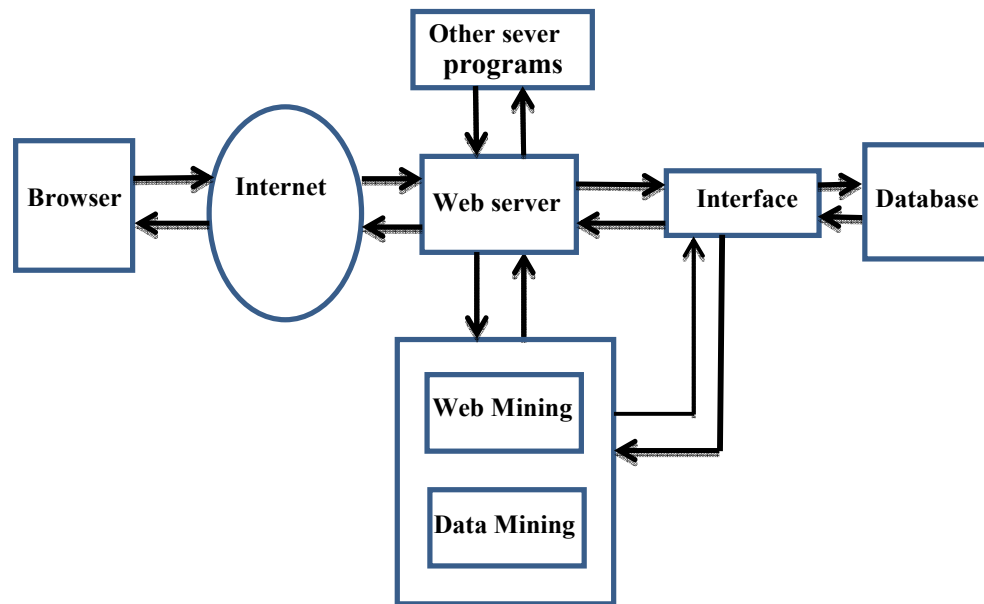


Figure 1. Distance teaching model based on Web mining and data mining

Compared with the distance teaching model based on WWW, a module of Web mining and data mining is added to the distance teaching model based on Web mining. This module is located on the server side and mainly periodically analyzes various information accumulated on the site, such as various server logs, interaction information between students and the site, etc. This module digs out some patterns and rules to redesign courseware for course designers, rebuild sites for administrators, and provide information for teachers to adjust teaching content in a timely manner.

1 Courseware tree

On the server side of the distance learning site, the learning resources of the entire education system are stored, mainly online courses and some other auxiliary files. Logically, these online courses can be regarded as a tree, and a leaf of the tree is a single HTML document, corresponding to a knowledge point or content of a course.

Similarly, all online courses form a tree, which is called CT (Course Tree). In fact, the courseware tree is also a concept tree, so in the mining process, it can be used for concept promotion and multi-level association rule mining. In our model, we pre-extract the courseware tree to form the courseware tree file, which is used for mining and pattern analysis process, so as to effectively carry out knowledge mining.

2 Realization of Distance Teaching Model Based on Web Mining and Data Mining

There are many implementation schemes and technologies for the distance teaching model based on Web mining. On the client side, there is mainly a web browser. There are many solutions on the server side, the most common one is NT Server+IIS+SQL Server. A variety of methods will be integrated in the implementation technology, such as HTML, CGI, VBScript/Jscript, JAVA, ActiveX and Python.

3 Design and Implementation of Web Mining and Data Mining Modules

In the module, the input of Web mining and data mining is mainly composed of four parts: log files of the system, interaction data between users and the site (user's registration information, Q&A information, test scores, homework, etc.), site files (HTML file), courseware tree file. The structure of Web mining and data mining modules is shown in Figure 2.

In the above structure, the method of Web content mining will be used for the processing of the courseware tree file. It can be seen from Figure 2 that the courseware tree file is mainly used in the mining process and pattern analysis process, so as to better carry out concept promotion and multi-layer association rule mining and pattern analysis.

The web mining and data mining modules are implemented as follows:

(1) preprocessing

The preprocessing process here is mainly for the log files of the system, which should be distinguished from the data preprocessing process in the general data mining process.

The preprocessing process is the most critical link in the Web mining process, and the quality of the processing is related to the quality of the subsequent mining process and pattern analysis process. The preprocessing process includes: domain-dependent data cleaning, user

identification, session identification, and path complementation.

(2) mining process

This process mainly uses some data mining algorithms to mine patterns, rules and so on. Some data mining algorithms, such as association rule mining algorithms, sequential pattern mining algorithms, and clustering algorithms, can be used in the process.

(3) pattern analysis

At this stage, we mainly use some methods and tools to analyze the excavated patterns and rules, and find out the patterns and rules that we are interested in. We can use Web roadmaps, OLAP techniques (such as data cubes) and SQL-like language mechanisms to visualize and interpret the rules and patterns mined.

There are many ways to measure the interest in mined rules and patterns. The most commonly used is to use the interest function to measure.

(4) Integration of web mining and data mining modules

Like other server programs, the Web mining and data mining modules run on the server side as an independent application program, analyze the server-side system logs and the interactive database between users and the site at regular intervals, and report to the system administrator, courseware Designers and teachers report various pertinent information.

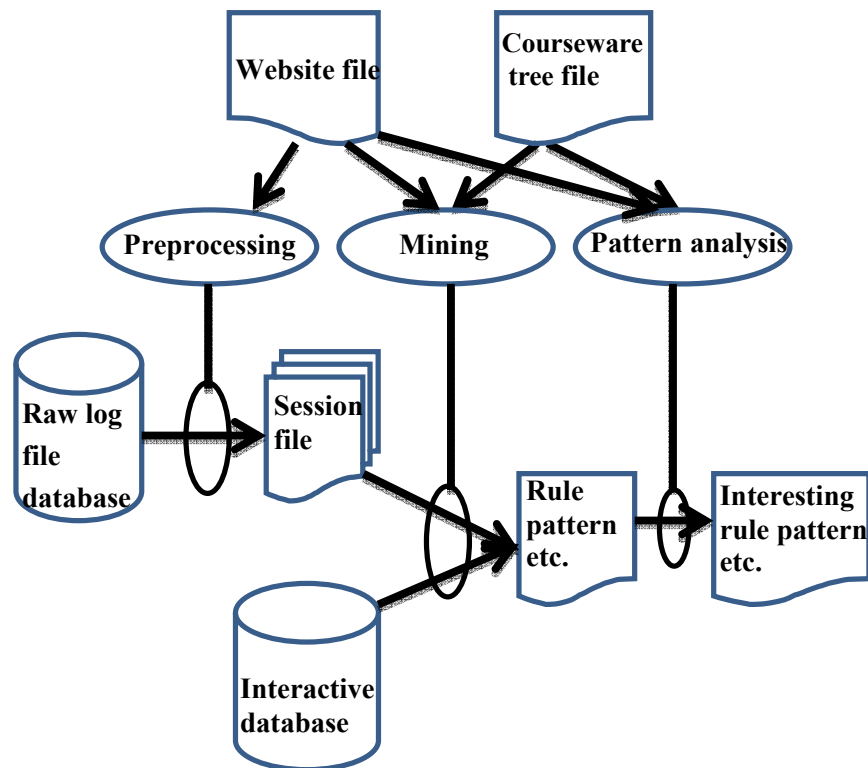


Figure 2 .Web mining and data mining module structure

WWW-based distance learning has become a popular distance learning module today. However, the current distance teaching sites are almost all static, there are few interactive means, and they cannot teach students according to their aptitude. At this time, a lot of useful information has been accumulated on the site, which has not been fully utilized.

In this paper, we propose a distance teaching model based on Web mining and data mining, and give the realization algorithm. In addition to the advantages of the general WWW-based model, this model can also make full use of the various information accumulated on the site to help course designers and managers to redesign courses and reconstruct sites, so that the design of courses is more reasonable and in line with teaching regulations. Learners can also give full play to their own personality.

In conclusion, compared with the WWW-based model, this model is more beneficial to distance education, and has broad development and application prospects.

REFERENCES

- [1] O Meriem and N. Bouyakoub, Teaching and Testing English in the Algerian Educational System[C], Arab World English conference, 2020, vol. 11:444-458.
- [2] T M Silviyanti, S A Gani and D. Junita, The native English teachers' expectations in teaching EFL students[J], *Studies in English Language and Education*, 2021, vol. 8, no. 1:212-226.
- [3] Cui Xuewen. The Application of Mining Algorithm of Association Rulesapriori In Performance Analysis [J]. Hebei North University Journal: Natural Science Edition, 2011, 27(01): 44-47.
- [4] L. Pipanmaekaporn and S. Kamonsantiroj, Mining semantic location history for collaborative poi recommendation in online social networks[J], *OBD*, 2016. vol. 11:31-38.
- [5] J. Han and M. Kamber, Data Mining: Concepts and Techniques[J], Morgan Kaufmann Publishers, 2006,167(2):78-82.