

# Research of Network Teaching Platform based on the Technology of Data Mining

Hongjun Fan<sup>1</sup>

Tourism Management Department, Hainan College of  
Economics and Business  
Hainan College of Economics and Business, HCEB  
Haikou City, Hainan Province, China  
hkfjh2006@163.com

Dong Kang<sup>2 a\*</sup>

Information Management Department, Hainan College  
of Economics and Business  
Hainan College of Economics and Business, HCEB  
Haikou City, Hainan Province, China  
15595621212@163.com

**Abstract**—Basing on the basis of the traditional network teaching platform, this paper has made the intelligent optimization for network teaching platform through the data mining technology. Based on the data analysis, data warehouse has been set up among the database. This new network teaching platform can find the inner link between data resources of the computer system in the data warehouse by the way of applying the correlation analysis, the clustering analysis, the TF – IDF methods, and so on. The new network teaching platform can track user operations intelligently among the system resources, and provide a more humanized service for teaching activities.

**Keywords**—network; teaching; intelligentize; data mining;

## I. INTRODUCTION

We have entered the era of information. On one hand, a large amount of data resources bring convenience to people. On the other hand, they are difficult to use effectively. There are so many same, similar or meaningless related resources because of the lack of corresponding mining analysis means. In the teaching system, how to avoid the repeated and complex information, and how to improve the teaching pertinence, and how to standardize the reasonable distribution of teaching resources have become important key topic field in the teaching system. Because of the generation of data mining technology, the study of the teaching system intelligentize has found the effective means.

## II. QUESTIONS

According to the actual situation of most of the information education in our country at present, the main purpose of the construction of network courses for teachers is not to develop the distance education, but to help teachers and students collect and show resources or activities of teaching and learning by using the curriculum management system (CMS), and to provide information learning environment, and primarily to promote teaching quality [1].

There are so many researches of the curriculum management system (CMS) at home and abroad, but the study of the teaching system intelligentize is a little. In the condition of informatization, as the amount of information increasing, we can just find the basic information through the

regular searching, classifying and tagging in the teaching platform. Usually in the large amount of documents obtained, only a little information is close to target, and has high value, so most of the information is nonsense. The CMS has the inability to support the most effective resources according to the specific circumstances of each student. So when we upload one article, one teaching courseware, or other resources to the system, we cannot accurately know the values, thus it is unable to quantify the benefits. Even more, the system resources can't track the usage of the students. Teachers can't know the internal relationships between students, and unable to effectively provide individuality service for them except the manual way.

## III. THE ANALYSIS OF THE DEMAND AND THE FUNCTION OF THE PLATFORM

The intelligentize teaching platform based on data mining technology will make the systematic analysis for each of the system resources, and find the internal relations between resources automatically rather than the actions of man, and form the valuable information to reduce the workload of human, and to improve the service efficiency of the system. Usually the application of a teaching platform depends on the three main bodies: teachers, students, administrators [2].

In the intelligentize teaching platform, the role of the administrator has been taken place by system. While most of the work will be done by the system, some functions have been distributed between the teachers and the students.

The intelligentize teaching platform can undertake effective service for teachers and students. teachers can open the courses in the system, set curriculum, distribute and correct students' works, divide the students' learning groups, evaluate students' learning on the platform. Students must enter the courses by registration, learn courses, accomplish the homework, search and review resources. Teachers and students can upload their own teaching resources, learning notes, and discuss with each other. In addition to normal resource management, System to complete the job, the intelligentize data analysis has been put forwards into the system, which mainly includes three aspects:

### A. the preprocessing stage

Generally, analyze the text information resources. For the other resources such as video, music and so on should be noted the relevant text information. Similarity comparing

---

This paper is the research of "the resource-sharing course construction for Hotel English basing on the position responsibilities" supported by the Education Department of Hainan Province(2013 scientific research projects for colleges and universities) .Code:Hjg2013-46

will be made when the resources upload into the system, and find the correlation among the resources, and such information will be stored in the corresponding data warehouse for further usage or analysis. The users can gain the assessment and a feedback guidance of resources uploading according to the analysis of the word frequency and transferring integrated data file (TF - IDF), to upload a resource, to show the correlation value of the resources in the system.

#### B. The mining and analyzing stage

Create a model by using data mining technology to discover and summarize the valuable information. When teachers and students are using the system, such as viewing the course information, searching keywords, or asking questions, the model can effectively find the most relevant resources to the current operation to improve the effectiveness of the system.

#### C. The evaluating stage

Users can score for the results of the model mining and analyzing. The scores have been deposited into the system database for further information evaluation in the future. System will make the clustering analysis such as the analysis of the internal relationships of students according to the information of students' operation records, learning process, learning result, ect, and form a result for the reference of administrators and teachers.

### IV. THE TEACHING PLATFORM ARCHITECTURE

A website performance is always influenced by the hardware resources such as CPU, RAM disk; disk I/O, the network bandwidth and so on. On one side, we can improve site performance by the hardware upgrading, on the other side, optimization design of website architecture can achieve the result of half the effort with double results [3]. This system adopts B/S (browser/server) system architecture, with the benefits are: no special software installation in the client, direct usage of the browser, convenience for system upgrade. The system adopts Visual Studio 2005 as the development platform, c # as the programming language, SQL Server 2005 as the database management system.

This system adopts three layer architecture designation of the USL, BLL, DAL. The USL is the user interface presentation layer, namely using the ASPX extension WEB interviewing for satisfying the user's requests and the return of the data; The BLL, say, the business logic layer, with the function of business logic data processing, Which means delivering users' information to the data layer, and returning the results to the browser of the presentation layer; DAL is a data access layer, with the purpose of supplying more rich data services to the USL layer and the BLL layer. Figure 1 shows three layer system architecture based on data mining technology.

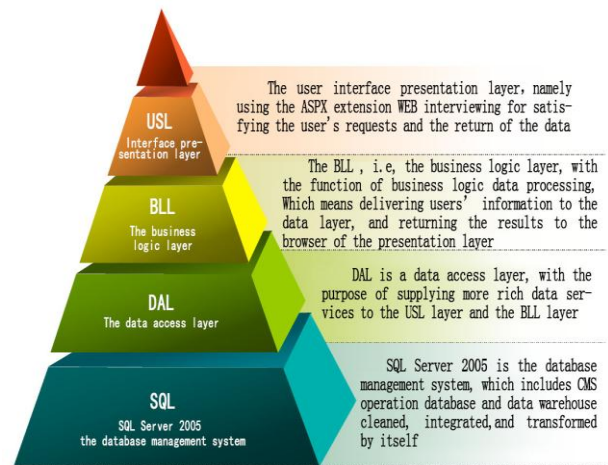


Figure 1. three layer system architecture based on data mining technology

### V. THE APPLICATION OF DATA MINING TECHNOLOGY

#### A. the desination of the data warehouse

According to the explanation in Bill Inmon 's book "Building the Data Warehouse" published in 1991, the Data Warehouse can be shown as follow: it is a Data collection with the characters of theme Oriented (Subject Oriented), Integrated (Integrate), relatively stable (Non - Volatile), and reflecting the historical changes (Time Variant), used for the usage of supporting management decisions. Data in data warehouse contain the basic data, historical data, comprehensive data and metadata[4].

To eliminate the needless, half-baked and inconsistent data in data analysis, the teaching platform has established data warehouse for convenient data mining based on a common database.

Data processing in the data warehouse adopts the way of updating method instead of traditional database query method. The data gained from the teaching system, uploaded by teachers and students will be cleaned, integrated, transformed, and reorganized into a new semantic database for direct query and analysis. This way can not only improve the efficiency of data analysis, but also have no effect the data source in the database of teaching platform.

Data warehouse takes shape when the operation database has been cleaned, integrated, and transformed. Data warehouse can resupply data. Figure 2 shows a three layer data warehouse structure:

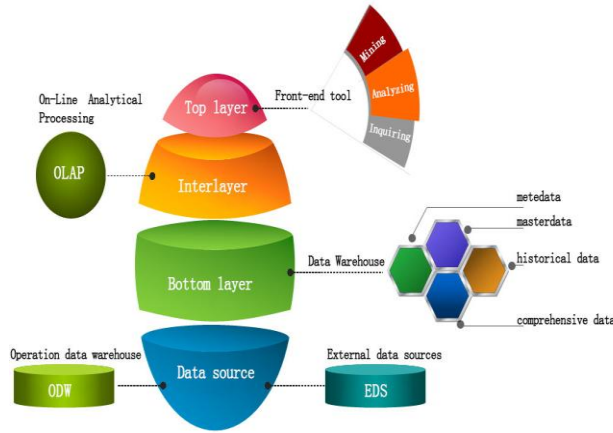


Figure 2. three layer data warehouse structure diagram

### B. The Correlation computing and data preprocessing

Correlation analysis means if there are some correlations between two or more things, one of the things can be predicted by other things. Its purpose is to explore the hidden relationship between data. In the intelligentize teaching system, we often need to preprocess the correlation between materials and students, then to set up a corresponding data warehouse. For example, when we want to analyze the students' learning enthusiasm, we need correlation analysis between the number of log in system for a certain kinds of student and the log in time. Students learning degree can be sensed through the correlation analysis between the completion of the assignment and homework scores. When students use the related materials in the system, we can reach the value correlation between the material and the students' category by the value of material through the number of students skimming and usage, the application of clustering analysis, and the relationship between material value and the classification of the students. All information can be stored in the warehouse in the form of a two-dimensional table.

Here take the students' learning enthusiasm judged by the system as an example. We suppose a student "P" 's learning enthusiasm be IP. Elements of influence for IP are the login number n, the length t, and the systematic operation o. Generally when the times a student logs in the system are more, the time is longer, and systematic operation is positive, the learning enthusiasm "p" is higher, namely, the amount of the  $I_p$  is higher, i.e.,  $I_p = \Delta \times n \times t \times o$ , the role of  $\Delta$  is to set the range into the domain constraints. But in fact, this is only the surface meaning. We can imagine some accident, such as network failure, the unstable client, and so on, causes students to drop after login, or frequent login the browser's cookies functions, which lead to the interference of n. If students leave but no exit system in the process of study, then the amount of t will increase. All of these will enhance the false appearance of  $I_p$ . At the same time, frequently click and meaningless operation of students can cause o repetitive operation amount. To solve these problems, we can detect outliers based on density (partial factors of outliers LOF

technology), and judge what the abnormal login system conditions are for students through the following five steps:

- 1) to calculate the Euclidean distance for one object and other objects
- 2) to sort the Euclidean distance, to calculate the k distance and the k field
- 3) to calculate reachable density of each object
- 4) to calculate the partial outliers divisors of each object
- 5) to rank the divisors of the partial outliers, and put them out.

According to the output, we can take some suspicious data out (i.e., data far away from the outlier), and finally store the data in the data warehouse after transformation. Optimized and improved data can be a very good solution to decrease the impact of the accident brought by the teaching system.

### C. The application of clustering analysis

Usually classification of samples is unknown when we solve practical problems by using data mining technology. Even more we can obtain, it is very difficult to retrieve data. Clustering algorithm can a way to solve this problem [5].

In the teaching system, we often make a division according to certain conditions of students, or certain teaching materials, which helps us to carry out teaching. The data will be analyzed by the data matrix in this system. Variable character P stands for N objects, such as age, gender, course score, and hobbies and so on to describe the student. It should be pointed out that, the interest is not a specific number, so we can make some markers in advance, such as letter "a" represents computer assembly, "b" stands for the web page making, and "c" means animation production... "z" shows the other.

We take clustering analysis to the students by using k-mean value and K center point. First, select K pieces of objects randomly. Each object represents a kind of initial amount. For the rest of the object, the distance of the mean value should be calculated from one to each other to determine which class it belongs to, and then calculate the new mean value of each class. This process is repeated again and again until the classification is stable. Even repetitive computation causes no more changes.

The criterion of clustering algorithm is square error, with its definition being as follows:  $E = \sum_{i=1}^K \sum_{p \in C_i} |p - m_i|^2$ , in which "E" is the sum of all objects' square error, "p" is the object of the calculation,  $m_i$  is the average amount of the class  $C_i$ . We can help teachers to classify students according to the conditions provided by the teachers, such as according to the study scores or interests, and to make the systematic decision analysis of the teaching activities through the clustering calculation.

### D. The application of TF-IDF

TF - IDF (term frequency, inverse document frequency) is a kind of weighted technology commonly used for information retrieval and information mining. TF- IDF is a

statistical method, which is used to evaluate the importance of a word for a set of files or a document in corpus [6].

TF - IDF algorithm is as follows:

$$TF(d,t) = \begin{cases} 0 & freq(d,t) = 0 \\ \frac{1}{1 + \log(1 + \log(freq(d,t)))} & freq(d,t) \neq 0 \end{cases} \quad (1)$$

$$IDF(t) = \log \frac{1 + |d|}{|d_t|} \quad (2)$$

$$TF - IDF(d,t) = TF(d,t) * IDF(t) \quad (3)$$

Freq (d, t) refers to the times of the word “t” appeared in “d” document. In general, the times appeared in one document are too many while a few in other document, which means the differentiation degree of the word is very high, then the function of the word is completely great. We can find out the important words of each article, being the key words through the TF - IDF algorithm. There is no need to note these keywords manually, but is computed by the system. Comparing these keywords, we can calculate the distance between the document and the document, meaning, the similarity. When documents or other materials (need to be noted in the form of text) are uploaded, system can compute intelligently the similarity between it and other data. When the certain resources are used, the system can recommend the similar documents computed beforehand to the user. There is no necessity to find resources, but to make resources automatically show to the user. Document similarity computing is mainly used for related information recommendation in the system, and to reduce the time of searching valuable information in the system to, and to help teachers make the overall decision of information providing. For example, the system similarity computing associated with the records a student searching information, and correlation analysis, show what data the students pay more attention to but the system lacks, and what data there are a

large amount, but the students have no interest, which can help users to optimize system resources.

## VI. CONCLUSION

Intelligentize teaching platform can provide direct query and analysis through the data mining technology with the process of building a data warehouse. The data obtained from the teaching system and uploaded by the teachers and the students in the data warehouse will be filtered, copied, pretreated, integrated, commented, summed up, and reorganized into a semantic database. On the basis of the traditional curriculum management system, the intelligentize teaching platform has been really implemented by the way of data mining technology such as the correlation computing, clustering analysis and TF - IDF. Both teachers and students can get their available information from the system more effectively, and enhance the guidance of making teaching decision and learning.

## REFERENCES

- [1] Li Jiahou, Zhan Yi. Curriculum Management System and its Selection[J]. Modern Educational Technology. 2008 (9) : 64-75
- [2] Kang Dong, Chen Juntao, Zhan Jinmei, Xing Haihua\*, Wu, Shulei. Design of university teacher's files management system[C]. CPS(Conference Publishing Services), Guangzhou. China. 2012.11
- [3] Chen Renzhang, Meng Xiaohua. Large network teaching platform architecture design and implementation[J]. Computer Engineering and Design . 2012, 31(11):2455-2469;
- [4] Chen Buying. The data warehouse and data mining application in vegetable diseases and insect pests forecasting[J]. Journal of Agricultural Mechanization Research. 2013(1):217-219,223;
- [5] Wang Guoxun. DATA MINING MODEL SELECTION BASED ON MULTIPLE CRITERIA DECISION MAKING[D]. Doctoral Dissertation. School of Management and Economics, 2013.9
- [6] Shi Hang. TF - IDF algorithm is used to optimize local news search[J]. Journal of Software Guide. 2011(11):59-60