Design and Application of College Foreign Language Online Teaching System Under the Background of Internet +

Qun Xie and Gengnan Fu(✉)

Changchun University of Chinese Medicine, Changchun 130117, Jilin, China
37055769@qq.com

Abstract. In the new stage of the development of "Internet + Education", online teaching is not only a teaching form, but also an important link in the modern teaching system of colleges and universities. However, there are outstanding problems in the current online teaching process from the construction of digital teaching resources, function and service planning to the design and application of online teaching platform. In this regard, this paper will focus on the analysis of the current situation of college foreign language teaching, and put forward a set of construction scheme of college foreign language online teaching system, aiming at promoting the digital and intelligent transformation and upgrading of college foreign language teaching mode. The whole system is in B/S architecture, which consists of front-end interactive interface, back-end server and data storage server. The function setting of the system will fully cover the teaching needs of the course, and give users such as students and teachers different permission interfaces, which will effectively improve the practicability of the system and strengthen the concurrency control of the system. After simulation test, the system has good adaptability to multimedia teaching resources, and can analyze and mine students' online learning behavior data by means of distributed computing, which makes up for many shortcomings of online teaching and realizes the innovative practice of college foreign language teaching path.

Keywords: Internet + education · College foreign languages · Multimedia technology · Distributed computing · Web application

1 Introduction

In the new era, it is the key to solve many problems in the current teaching practice and the only way to build a high-quality education system for foreign language majors to promote the digital, networked and intelligent transformation of foreign language teaching mode in colleges and universities. Since the "Guiding Opinions on Actively Promoting the" internet plus Action "issued by the State Council, the deep integration of network information technology and education has entered the fast lane of development, and triggered a series of innovations in the teaching field. [1] The emergence of online

D. Kumar et al. (Eds.): IEIT 2023, AHSSEH 10, pp. 1317–1326, 2023.
https://doi.org/10.2991/978-94-6463-230-9_159

teaching not only creates a brand-new form of educational practice, but also can realize the digital construction of teaching resources, optimize teaching programs, increase practical application channels, stimulate students' interest in learning, and improve the teaching effect of foreign language professional courses with the help of the functional advantages of online teaching platform. However, at present, the development and application of online teaching is still in the primary stage, the overall planning and design of the system is incomplete, and the technical system is relatively simple, so that the system itself has obvious equipment limitations, which further affects the presentation and application of multimedia digital teaching resources, weakens the concurrent control of users, and lowers the efficiency of the system. [2] In view of this, this paper holds that colleges and universities should actively grasp the training concept of foreign language compound talents at this stage, adhere to the innovation drive, and build a Web-based online college foreign language teaching system with the help of the practical characteristics of network information technology, multimedia application technology and data analysis and mining technology, so as to lay a solid foundation for the smooth implementation of college foreign language online teaching mode. The system takes students as the main application group, and the function setting covers the whole process of college foreign language education and teaching, which meets the actual needs of many parties and relieves the contradiction among teachers, students and schools in the current teaching process. Moreover, it focuses on the use of distributed computing to strengthen the mining and processing of user learning behavior data, enhance the system function and service dimension, promote the improvement of college foreign language education system, and promote the process of modernization and intelligent construction of college education.

2 Overview of Key Technologies

2.1 Distributed Computing

As a new computing method, the principle of distributed computing is to distribute the computing tasks to a large number of computing function nodes, and after the calculations are completed respectively, the final data conclusion is obtained by unifying the computing results. [3] Distributed computing corresponds to centralized computing, and the main difference lies in the selection and construction of computing function nodes. Numerous computing function nodes can be deployed on a single server, or run on multiple servers through network connection. The outstanding advantages of distributed computing lie in sharing data information and balancing the running load of each computing function node, which greatly saves the processing time of computing tasks and improves computing efficiency. Common distributed computing systems include Apache MapReduce, Apache Spark and MPI. Apache MapReduce represents the idea of batch processing, which can meet the requirements of offline processing, but it takes a long time to run and takes up a lot of resources. The stream processing framework represented by Apache Spark can adapt to real-time tasks well, but there is a great risk of error in data sequence. Because of its own characteristics, MPI can define a series of functions

Table 1. Type of common data mining algorithms

Classify	Data analysis method	Data mining model
Guided data mining	Classification analysis	Decision tree, random forest, neural network
	Predictive analysis	Regression tree, rough set method
	Valuation analysis	SVM, Bayesian method
Unsupervised data mining	Correlation analysis	Pearson Correlation coefficient, Aprior algorithm
	Cluster analysis	K-means clustering algorithm, spectral clustering

to clarify the process and running order of distributed computing, simplify the development difficulty of distributed computing system, maintain the consistency of program scheduling, and further shorten the running time of distributed computing [4].

2.2 Data Mining Technology

As a computer science and technology, data mining technology is a processing method for big data, aiming at extracting potentially useful information and knowledge from a large number of incomplete, noisy, fuzzy and random actual data. [5] The main steps of data mining include: data acquisition, data preprocessing, model establishment and operation, and final result expression. The construction of data mining model is the core link of the whole data mining work, which corresponds to the data analysis method, which not only determines the application direction of data mining results, but also determines the construction of data mining model. Common data analysis methods include classification analysis, prediction analysis, cluster analysis, valuation analysis and correlation analysis, as shown in Table 1 [6].

2.3 Development Process

According to the requirements of the above related application technologies, the configuration and deployment of the development environment of online college foreign language teaching system can be completed. The content of system development is divided into two parts: basic framework construction and functional structure design. The whole system adopts B/S architecture and is divided into three parts: presentation layer, business logic layer and data layer according to the established hierarchical system. [7] In the actual development process, the front-end page takes the React framework as the core, and realizes the complete page function by combining or reusing various components. The development and deployment of the back-end server need to be realized by the Django framework of "request/response" mode, and follow the MVC design pattern, and the front-end interactive interface and the back-end server can be associated and connected under a specific data interface.

The functional structure of the system needs to be designed and developed according to the actual needs of users. Firstly, LDAP authentication mechanism under CAS protocol will be used to support users to complete single sign-on for system user login authentication and authority management. When a user logs in to the system, a Cas-cookie and a one-time ticket will be generated through the CAS server. After receiving the ticket, the Web server will verify the validity of the ticket to the CAS service and return to generate a Service-cookie to complete the authentication and give the user a response. [8] Secondly, the system mainly relies on the React-player component to play multimedia teaching resources, which can read and analyze various URL addresses, so as to improve the system's support for multimedia teaching resources. Finally, the system needs to use data mining algorithm to analyze the characteristics of the complex learning behavior data of student users. The construction of data mining model is supported by sklearn class library in Python language, and distributed operation is realized by MPI protocol, so as to improve the operation efficiency of data mining algorithm and strengthen the calculation ability of the algorithm.

The basic development environment of the whole system will be deployed according to the conventional "LAMP" mode. Among them, Linux CentOS 7.3.1611(64bit) is selected as the bottom operating system, Python is selected as the basic development language environment, version 3.10.2, and Pycharm 2019 is selected as the integrated development tool. Nginx-wsgi-django Web module is selected as the web server, and Mysql 5.7.31 is selected as the database server. After all the software is installed, choose Django directly when creating new items in Pycharm and create an independent virtual environment. At the same time, the basic directory of the project will be set up, mainly including Manage.py, Urls.py, wsgi.py and Settings.py. When the functional modules are designed, all the files are packaged and published on the server, which is convenient for users to log in and visit remotely. Through the introduction of the above key technical theories, the overall environment of system development, the configuration of related software and tools are determined, and the technical feasibility of online college foreign language teaching system is also clarified.

3 Functional implementation

3.1 Student side

A. Online learning

The system will classify the course content according to the foreign language, so that students and users of different majors can choose the corresponding courses independently according to the teaching plan or personal preference. The setting of course content is more flexible and rich than traditional textbooks, which not only fits the reality of life in modern students, but also has clear teaching pertinence. Under this function module, the system uploads various forms of multimedia teaching resources such as video, audio, text and pictures to the database of the system in advance, and formulates the corresponding URL address. When the student user selects and opens the corresponding teaching resources, the React-player component automatically obtains the URL address of the target resource to complete the play or display.

B. Virtual practice

Foreign language courses in colleges and universities need to improve students' comprehensive application ability from listening, speaking, reading, writing and translation. Traditional teaching models often focus on theoretical teaching and ignore the cultivation of practical application ability of foreign languages. In order to solve this problem, the system will use the advantages of computer application to integrate a large number of dynamic graphics, video images and Flash animations, and build a large number of thematic simulation scenes, so that students can complete related tasks in the situation, experience practice and start learning through topic discussion, simulation drills and other forms.

C. Self-test question bank

Under this function module, the platform supports students to complete simulated answering and comprehensive English application ability training online. The item bank subsystem is preset in the system, which contains various types of questions in CET-4, CET-6, TOEFL, IELTS and other professional examinations, so that students can conduct online simulation tests conveniently. The realization of this function mainly depends on the data interface between the front-end interactive page and the back-end function control of the platform.

3.2 Teacher Side

Under the teaching system, the teacher's role orientation is biased towards the organizer and manager of online teaching practice. The corresponding functional authority mostly focuses on the updating and maintenance of teaching resources, virtual practice scenes and self-test question bank. In addition, teachers also focus on using data mining algorithm model to analyze and process the complex learning behavior data of student users, further obtain the real needs of student users, further promote the improvement of personalized teaching system for foreign language majors in colleges and universities, and promote the improvement of teaching effect. In the actual operation process, the data collection objects include CSV log files generated by the actual operation of the system and data records in the system database, as shown in Table 2, which is the main measurement index of student users' online learning behavior.

After data collection, the data is preprocessed, and there are many irrelevant data, duplicate values and missing values in the CSV log file. The data needs to be cleaned, integrated, changed and reduced to keep certain standards and specifications. As shown in Formula 1, it is a normalized calculation formula, where X represents the original data, X_1 and X_2 represent the minimum and maximum values in the original data set respectively, and X' represents the normalized value.

$$X' = \frac{X - X_2}{X_2 - X_1} \tag{1}$$

Data mining mainly focuses on the correlation analysis between learning behavior and learning effect and the clustering analysis of learning behavior. During the simulation test, the system selects more than 1.6 million pieces of data from 300 student users as the

Table 2. Main measures of online learning behavior

Core indicators	Behavioral observation point	Measures
Academic record	Question bank test score	Score, ranking
Learning time	School time	Study start time
	Play time	Study completion time
Learning interaction	Complete the login	Login time
	Complete the homework	Homework completion time and times
	Complete the course	Course completion time and times
	Complete the test	Test completion time and times
	Number of login	Login days
Interpersonal interaction	Participate in the discussion	Number of participants
	Publish a discussion	Published quantity

original data, and uses Pearson correlation coefficient model to analyze and process the data, and obtains Pearson correlation coefficient matrix. Pearson correlation coefficient calculation formula is shown in formula 2, where X represents learning behavior and Y represents the final learning effect. The realization of the algorithm in the system needs to be built in Python language environment with the help of Numpy class library, and some codes are shown below. Finally, the correlation coefficient between students' learning behavior and learning effect is shown in Table 3. After significance test, the significance level of "participating in discussion" is lower than 0.05, which is in a relatively significant state. The significance level of "published discussion" is greater than 0.05, and it is judged that the explanatory power to the learning effect is insufficient, so it is deleted; The significance level of other behaviors is less than 0.01, which has a strong correlation with the learning effect.

$$P = \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \overline{X})^2} \sqrt{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}} \tag{2}$$

```
from math import sqrt
    def multipl(a, b):
    sumofab = 0.0 for i in range(len(a)):
    temp = a[i] * b[i] sumofab + = temp.
    return sumofab.
    def corrcoef(x, y):
    n = len(x).
    sum1 = sum(x) sum2 = sum(y) sumofxy = multipl(x, y).
    sumofx2 = sum([pow(i, 2) for i in x]) sumofy2 = sum([pow(j, 2) for j in y]).
```

num = sumofxy - (float(sum1) * float(sum2) / n).
den = sqrt((sumofx2 - float(sum1 ** 2) / n) * (sumofy2 - float(sum2 ** 2) / n)).
return num / den.
print(corrcoef(x, y)).

In order to further improve the efficiency of students' autonomous learning and improve the learning effect, the cluster analysis of students' learning behavior will continue. K-means algorithm is selected as the cluster analysis model, and the optimal K value is determined according to DBI index. The results of DBI index under different K values are shown in Fig. 1, where the smaller the DBI index, the smaller the intra-cluster clustering and the more reasonable the clustering results, so the final K value is determined as 5 [9].

In Python environment, the platform can automatically call the cluster in Python sklearn class library to build a clustering algorithm. Select the learning behaviors whose Pearson correlation coefficient significance level is less than 0.01, and input the corresponding data values and K values into the K-means clustering analysis model to get the clustering analysis results. Table 4 shows the statistical results of the data after cluster analysis, and the obtained results are standardized to ensure the uniformity and comparability between the data. As shown in Formula 3, it is a standardized calculation

Table 3. Correlation coefficient between learning behavior and learning effect

Behavioral observation point	Correlation coefficient	Behavioral observation point	Correlation coefficient
Question bank test score	0.048	Complete the course	0.384
School time	0.014	Complete the test	0.021
Play time	0.082	Number of login	0.598
Complete the login	0.514	Participate in the discussion	0.221
Complete the homework	0.558	Publish a discussion	0.174

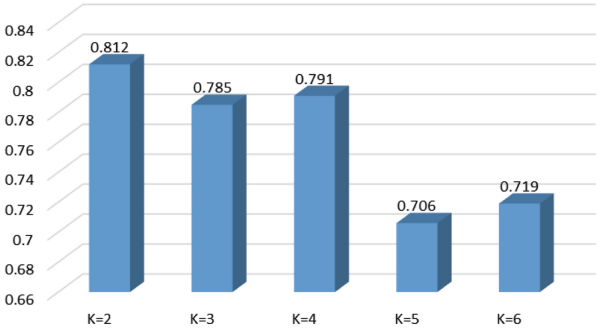

Fig. 1. DBI index results at different K values

Table 4. Cluster analysis results and statistics

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Question bank test score	59.913	56.522	50.514	44.328	22.291
School time	39.964	36.025	32.584	27.501	24.316
Play time	68.221	65.113	62.947	59.506	40.500
Complete the login	471.360	426.539	420.877	297.634	209.101
Complete the homework	19.697	17.075	17.334	14.405	13.225
Complete the course	25.834	23.866	22.488	21.665	18.467
Complete the test	3.687	3.233	4.152	2.334	1.264
Number of login	9.387	7.625	6.339	5.152	0.004
Number of students	30	63	159	33	15
Proportion	10%	21%	53%	11%	5%

formula, where X represents the original data, ω is the mean value and μ is the standard deviation.

$$X\prime = \frac{X - \omega}{\mu} \tag{3}$$

According to the changing rules of the upper limit, lower limit and average value of students' learning effect in each cluster, 300 students are divided into five types, corresponding to five clusters respectively. The analysis results show that the students in cluster 1 perform well as a whole, and the average value of each behavior data is higher than that of other students in the cluster, so they are judged as excellent students. Cluster 2 students' learning effect is second only to excellent students, and their behaviors such as homework completion and course completion are better than excellent students, so they are judged as good students. Cluster 3 students have the highest overall proportion, and the learning effect is also in the middle level, so they are judged as ordinary students. The students in cluster 4 have little difference in course learning, but there are shortcomings in testing and homework, so they are judged as poor students. The overall learning effect of cluster 5 students is not ideal, and they have obvious characteristics in login times, so they are judged as particularly poor students.

In order to improve the operating efficiency of the system for data mining of students' learning behavior, MPI standard will be used to realize the distributed calculation of mining algorithm in Python environment. MPI provides mpi4py binding information transfer interface for Python language, thus allowing any Python program to utilize multiple processors. [10] After MPI 4py is installed, use mpirun -n 3 python3 test.py instruction to run the data mining model, and randomly give each computing node a different rank value. You can directly obtain the rank id through the Get_rank () method. After the computing node program is completed, the results can be summarized to realize the distributed computing of data mining. Taking cluster analysis as an example, the

Table 5. Distributed computing efficiency comparison results

Data volume	Single machine calculation		Distributed computation	
	Time	Accuracy rate	Time	Accuracy rate
4K	288s	84.35%	301s	84.10%
40K	579s	83.94%	610s	83.82%
400K	5399s	83.97%	1326s	83.69%

comparison results of system processing efficiency under single machine environment and distributed computing are shown in Table 5. The results show that there is little difference in accuracy and efficiency between distributed computing and stand-alone computing under a small amount of data, but distributed computing has obvious advantages in computing efficiency for processing a large number of data, which meets the current demand of universities for processing massive student behavior data.

4 Conclusions

In order to improve the effectiveness of foreign language teaching in colleges and universities under the network teaching mode, this paper puts forward the construction scheme of online foreign language teaching system in order to promote the digital and intelligent transformation and upgrading of college foreign language teaching mode. The platform can integrate learning behavior data mining into the teaching management process, highlight the value and significance of students' online learning behavior data, enhance the system function and service dimension, and promote the perfection of college foreign language education system. In the follow-up research, the system should further enrich the construction of teaching resources, optimize the application ability of data analysis module, and provide reference for the informationization and intelligent development of higher education.

References

1. Fu Zhengyun. The Construction of Foreign Language Education Mode in Colleges and Universities Based on Network Informatization[J].Educational Informatization Forum.06 (2021).
2. Liang Xuhong. Challenges and Countermeasures of Foreign Language Teaching in Colleges and Universities in the Network Era[J].New Curriculum Research.03 (2021).
3. Dou Shuai, Li Ziyang, et al. Algorithm Integration Method in Distributed Computing System[J].Computer Systems & Applications.03 (2021).
4. Zhao Baoqi, Li Weidong, et al. Distributed Data Processing System Based on MPI[J].Computer Engineering.06 (2018).
5. Zhang Peng. Data Mining Technology and Practice in the Era of Big Data[J].China New Telecommunications.11 (2021).
6. Wang Lili. The Application of Data Mining Technology in the Background of Big Data[J].Computer & Network.10 (2021).

7. Li Chaoke, Jin Ruixia, et al. Analysis on the Construction of Personalized Network Learning System in Cloud Computing Environment[J].Journal of Fujian Computer.04 (2015).
8. Zhao Kankan, Dai Jianan. Improvement and Implementation of Single Sign-on System Based on CAS Protocol[J].Computer & Network.05 (2020).
9. Yao Ao, Zhang Yu. An Optimized K-Means Clustering Algorithm[J].Industrial Control Computer.11 (2016).
10. Ma Yalei. MPI Based on Python[J].Practical Electronics.12 (2013).