Multimedia self-learning behavior monitoring data mining system based on Web

DENG Chun-Sheng[1], CHENG Hai-Bo[2], LIN Jian-Ming[3], YANG Ju-Rong[4],ZHU Yan-Chen [5*]

{hy20160401@163.com[1],jwcjxjz163@163.com[2],xj20160608@163.com[3],

cl20160629@163.com[4],dddd24616563@163.com[5*]}

(1.School of Computer,Jiangxi University of Traditional Chinese Medicine,Nanchang 330004, China；

2.School of Computer,Jiangxi University of Traditional Chinese Medicine,Nanchang 330004, China

3.School of Computer,Jiangxi University of Traditional Chinese Medicine,Nanchang 330004, China

4.College of Humanities,Jiangxi University of Traditional Chinese Medicine,Nanchang 330004,China；

5.School of Computer,Jiangxi University of Traditional Chinese Medicine,Nanchang 330004, China）

Abstract： In view of the poor clustering effect of information data in the use of the original English multimedia independent learning behavior monitoring data mining system, the original data mining system is optimized according to the Web architecture, and a web-based English multimedia independent learning behavior monitoring data mining system was designed. Install the monitoring sensor, data transmission channel and decoder into the original system hardware framework to complete the system hardware design; According to the requirements of Web architecture design, this paper constructs the English multimedia independent learning behavior monitoring database, adopts the data feature point extraction algorithm, and uses the Web mining technology to achieve the system data mining performance, so as to realize the design of the web-based English multimedia independent learning behavior monitoring data mining system. Through the comparison of data clustering effect, it is verified that the designed system can effectively improve the ability of data clustering and improve the performance of system data mining.

Keywords: Data mining; Autonomous learning; K-means algorithm; Correlation mining;

1 Introduction

With the increasing popularity of Internet services, multimedia data resources on the network are unprecedentedly rich. How to extract useful information from massive multimedia data and deal with the information in a way that can be understood by users, so as to dig out regular potential patterns is a research hotspot with practical significance and application prospect[1].

Monitoring and mining of English Multimedia autonomous learning behavior is a combination of data mining technology and multimedia autonomous learning technology. It is

EAI MOBIMEDIA 2020, August 27-28, Harbin, People's Republic of China

DOI 10.4108/eai.27-8-2020.2294666

an interdisciplinary research field of knowledge discovery, data mining, artificial intelligence, machine learning, database technology and multimedia technology. Data mining of English multimedia autonomous learning is an intelligent multimedia technology, which elevates multimedia processing and management from information access to knowledge acquisition. Multimedia data is unstructured or semi-structured. Each media data has different characteristics and its own way of expressing information. Each media can not only represent information independently, but also jointly represent different characteristics of the same event, and jointly describe the existence, development and results of the event[2]. Therefore, there must be some characteristics, attributes and relations of the information subject in the multimedia data set, or some modes that people can't get intuitively. In previous studies, data mining systems used in this study often had poor data mining effects due to poor clustering performance. Therefore, in this study, Web technology was used to improve the data mining ability of the system.

Web mining the process of extracting patterns and hidden information of interest and potential value from Web documents, media, structures, and user interactions[3]. There are three types of Web mining: Web content mining refers to the summarization, classification, clustering and correlation of Web page text and various media contents; Web structure mining refers to the discovery of knowledge in Web chain structure, which can reflect the flow and distribution of information in Web space, as well as the nature and characteristics of Web elements. Web usage mining refers to the use of access path analysis, classification and clustering and other data mining methods to track the interaction and manipulation between users and the Web, including the mining of server access records and the analysis of access path. This technique is used to optimize the original data mining.

2 System hardware design

Aiming at the problem of poor clustering effect in the original English multimedia self-learning behavior monitoring data mining system, Web mining technology is used to optimize the original system. In order to realize the software module function in this design, the original system hardware is designed. The specific system hardware framework is shown below[4].

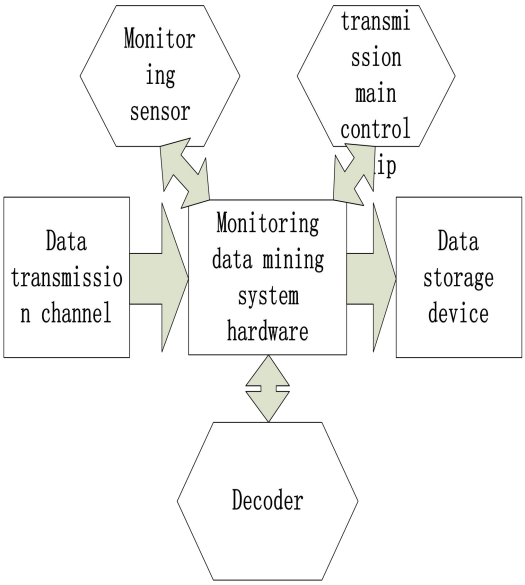


Figure 1 system hardware design framework

The above framework is adopted as the basis and control scheme of hardware framework design to ensure the order and controllability of hardware design and improve the process of system design.

2.1 Design of monitoring sensor

In the process of data mining, data collection is the database of mining technology. In order to ensure the effective data source in this design, a data acquisition device based on sensors is set up. Sensor node is the core of data acquisition, mainly composed of sensor module, processor module, wireless communication module and power module, as shown in figure 2.

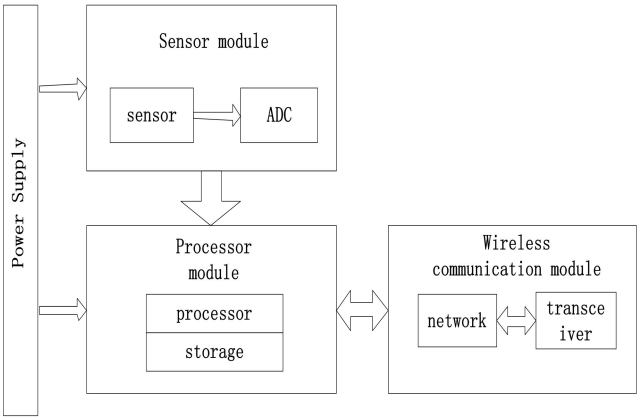


Figure 2 sensor node structure

According to figure 2, in the sensor node, the sensor module is responsible for data collection. The module mainly senses various information through the sensor, and converts the analog parameters into digital parameters, which is convenient for the processor module to process the data[5]. The processor module is mainly used to receive and store data. Finally, the wireless communication module is used to realize the wireless data transmission; the power module provides stable voltage, provides energy for the system, and maintains the normal operation of the system.

According to the actual needs of the system, the CC3200 node master control chip is selected. It is the most widely used and powerful control chip. The internal integration of the two large cores provides user development with a high performance processor running at 80MHz, as well as a wide variety of external devices.

2.2 Design of data transmission channel

The design of data mining system is mainly implemented on digital hardware platform. In this design, FPGA (Field programmable gate array) is selected as the hardware implementation platform of English multimedia independent learning behavior monitoring communication function. FPGA has the advantages of low power consumption, abundant resources and high parallelism, which is very suitable for realizing the functions of the physical layer of the system in this paper. In this design, FPGA core digital processing chip is selected as the main processing chip (as shown in figure 3). The AD/DA chip is also connected through the external FMC (FPGA Mezzanine Card) extension interface to realize the transmitting and receiving at the rf level. In addition, crystal oscillator is required to provide a main reference clock for the internal use of the FPGA. In order to realize the data transmission to the upper computer, the external FPGA connects with the external DDR (Double Data Rate) memory through the ax bus, and then transfers the data to the arm processor.

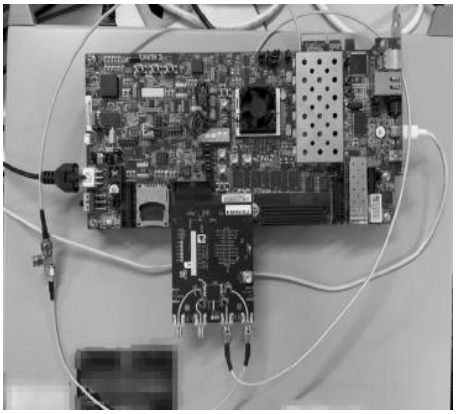


Figure 3 data transmission master control chip

In addition to the above Settings, FMC should also be added to the data transmission device to improve the function of the transmission device[6]. FMC card carrier connector pins are connected with FPGA chip pins that have just configured IO resources through PCB (printed circuit board) design; Connector pins on FMC subboard modules are also connected to IO interfaces by PCB design. Different IO (Input and output) interfaces can be designed on subboard PCB to realize various functions. In this way, the same carrier card can realize different extension functions through different subboard designs, making the chip more flexible to be applied in the scenarios designed by developers. In this project, FMC connector is used to realize the connection between FPGA chip and peripheral AD9361 chip, which can achieve high performance pins of Gb/s and expand FPGA chip to receive AD/DA data. Through the above steps, the construction of English multimedia independent learning behavior data feature extraction module is realized.

2.3 Decoder design

On the basis of the above hardware design, in order to realize remote control of the designed sensor, it is necessary to decode the control signal from the sensor end and the client end, and convert the control signal sent by the software into the actual action level signal[7]. The front end of the decoder is connected to the serial port of the video server through the RS-232/485 converter, and the serial port RS-232 control signal output by the video server is converted into RS-485 signal by the RS-232/485 converter, and then transmitted to the remote decoder. The decoder transcodes the received command signal, obtains the address and action mode of the corresponding cradle head and sensor, controls the action of the corresponding relay, and sends different level control signals to the cradle head and sensor control signal line at the output end of the decoder. The head and sensor are driven by the corresponding control signal level.

Figure 4 decoder design results

The decoder designed above is connected to the data transmission channel and front-end sensor in the paper, and installed into the original hardware framework to complete the hardware design process of the system in the paper. The optimized hardware framework is used as the software development environment to ensure the smooth operation of software modules[8].

3 System software design

In order to realize the corresponding effect of Web mining technology, the software framework structure in the original system is set as the hierarchy suitable for Web technology, as shown below.

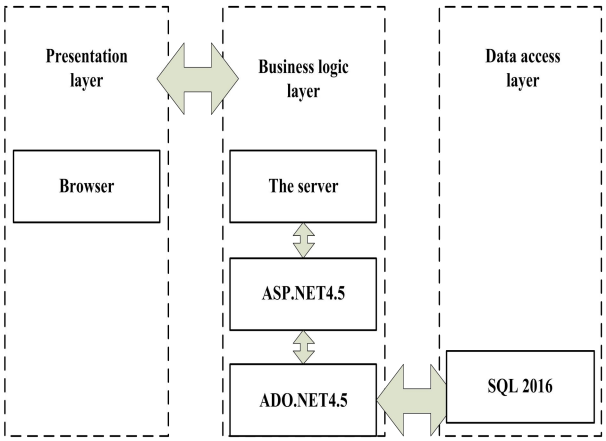


Figure 5 system software hierarchy Settings

In this design, the software results are set as three layers, namely: presentation layer, business logic layer and data access layer. Presentation layer (UI) : this system uses ASP.NET

WEB application development, the client uses IE browser to achieve the user interface; Business logic layer (BLL) : responsible for the processing and data transfer of key businesses. The operation and logical judgment of the database are all processed in this layer. The evaluation process of the system is all processed in this layer. Data access layer (DAL) : responsible for database access, mainly providing data to the business logic layer and manipulating the database based on the parameters passed in. Using ADO.NET data adapters in ASP.NET and SQL server storage process is complete. The structure has the characteristics of convenient system operation, easy maintenance and short information processing cycle[9].

3.1 Build the behavior monitoring database

In view of the problems in the use of the original system, the monitoring data collected by the designed sensor is stored in the form of database, and the stored data is taken as the basis of data mining of the designed system. In order to control the designed database effectively, its internal data table structure is set as follows.

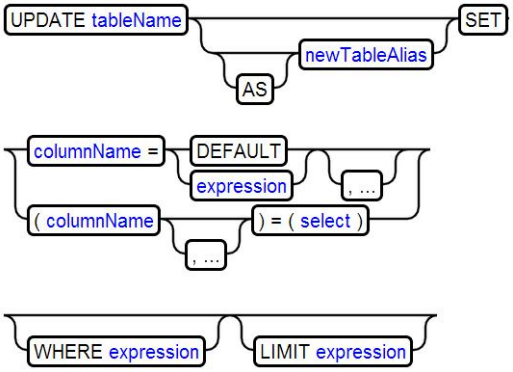


Figure 6 database internal data table connection structure

The above connection results are used as the internal framework of the designed data, and the collected data are edited and stored in the form of data tables. In order to ensure the structural consistency of the information in the database, the following format is set for writing the contents of the data table[10].

Table 1 data tabular Settings

Serial number	field	form
1	data sources	int
2	Sensor number	varchar
3	Data type	varchar
4	Data information content	varchar
5	Data acquisition time	varchar
6	Characteristics of autonomous learning behavior	double

7	Types of autonomous learning	double
8	Self learning operator	double
9	Operator user name	int
10	Operator login time	varchar
11	System notes	double

Use the above data table to organize and store the collected data, and use the above database structure to expand the connection, to ensure the effective data mining part.

3.2 Data feature extraction

Set the internal information of the database designed above in the form of data matrix, set there is a kind of learning behavior information in the database, these variables represent the objects in the database, and reflect the database content in the form of matrix, then:

$$\begin{bmatrix} a_{11} & \cdots & a_{1g} & \cdots & a_{1h} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{i1} & \cdots & a_{ig} & \cdots & a_{ih} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n1} & \cdots & a_{ng} & \cdots & a_{nh} \end{bmatrix} \quad (1)$$

Formula: g and h represent the constant parameter of the matrix, and the interval scale is used to obtain the phase differences of the features in the matrix. In order to improve the clustering accuracy of this clustering method, the scale variable is set as the minimum unit. The design of known variables has an absolute impact on the clustering result. Therefore, the standard metric value is set as a constant value.

Based on the above results, the distance between each object is set as $d(m,n)$, and the characteristic distance between the database contents is obtained by using the minkowski distance function. In the process of using the distance function, the binary variables are represented by 0 and 1. According to the above formula, the degree of dissimilarity between evaluation objects m,n can be obtained. The specific formula is as follows.

$$d(m,n)=\frac{e+f}{j+e+f+k} \quad (2)$$

Formula: j represents the number of variables when the values of the object m,n are all 1; e represents the number of variables when the object value m is 1 and the object value n is 0; f represents the number of variables when the object value m is 0 and the object value n is 1; k represents the number of variables when the values of the object m,n are all 0。

The calculation formula of the internal information heterogeneity of the system database is as follows：

$$d(m,n)=\frac{e+f}{j+e+k} \quad (3)$$

Through the above formula, the feature extraction of data information is completed, and the extracted features are calculated by using the mismatch rate. The specific formula is shown below.

$$d(m,n)=\frac{p-q}{p} \quad (4)$$

Formula: q represents the number of variables with the same state value as m,n ; p represents the total number of variables. Through the above formula, the feature extraction of English multimedia independent learning behavior data is completed, which is used as the auxiliary data support of data mining.

3.3 System data mining based on Web mining technology

The extracted monitoring data feature points are used to classify the data. Combined with the content mining technology in Web mining technology, the data are stored in different data warehouses, and the feature weights are obtained. The data are mined with different feature weights according to the actual needs. Part of the code to implement data mining is shown below.

```
{class Kmdiod()
get_data(self)//receive data
centroids = self.,data[init_index,:]//Initialize cluster center
print('Start iteration');
target = []
stop = false//Initial parameter setting
while(not stop)://Start traversing datastop =true
points = [[can] for in can]
target = []
For sample in self.data;//Calculate the nearest cluster center to the
datatarget.append(level)//
For s in range(self.cen)//
N_distances = sum(distance)//
for(int i=0;i<n;i++){
if(belong[i]){
cout<<end;}//Count all points contained in cluster count all points contained
in cluster}
```

Figure 7 behavior monitoring data mining part of the code

Using the clustering algorithm, all the monitoring data of English multimedia autonomous learning behavior are divided into multiple clusters. Objects in the same cluster have relatively high similarity values, while objects in different clusters have relatively small similarity values. According to this feature, we use the feature weight of the data samples to mine the data, and divide the data into the most similar classes, so as to realize the autonomous learning behavior monitoring of English Multimedia in the process of data mining.

Combining the software module and computing process set above with the system hardware framework designed in this paper, the design of the web-based multimedia English learning behavior monitoring data mining system is realized.

4 Experimental test

Combined with the above hardware design results and software module design, the design of the web-based multimedia English learning behavior monitoring data mining system is completed. In order to ensure the effectiveness of the design, the system testing process was constructed to compare the performance differences between the original system and the designed system.

4.1 System test platform design

In order to understand the performance difference between the designed system and the original system, carry out system test, and ensure that the designed data mining system and the original data mining system can be compared on the same platform, the experimental environment is composed of two parts, namely hardware environment configuration and

software environment configuration. During the experiment, the consistency of all equipment should be ensured to avoid the inaccuracy of the experimental conclusion.

Table 2 system test platform

	software environment	hardware environment
Server side	Ubuntu Linux 14.04 LTS RStudio0.98.1091 Python 3.4	CPU: i7-4790 Memory16GB Hard disk8TB
User side	Windows 7 Chrome	CPU: i7-4790 Memory16GB Hard disk8TB

The above hardware and software were used to construct the experimental environment to ensure the validity of the system test results. The collected English multimedia autonomous learning behavior information database is set as the data sample of this test. The data distribution in the database is as follows.

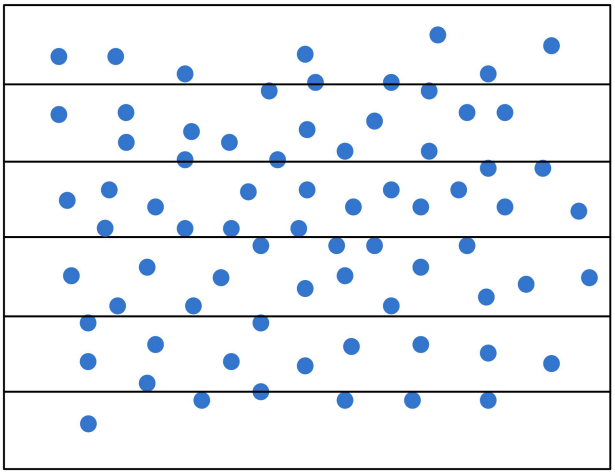
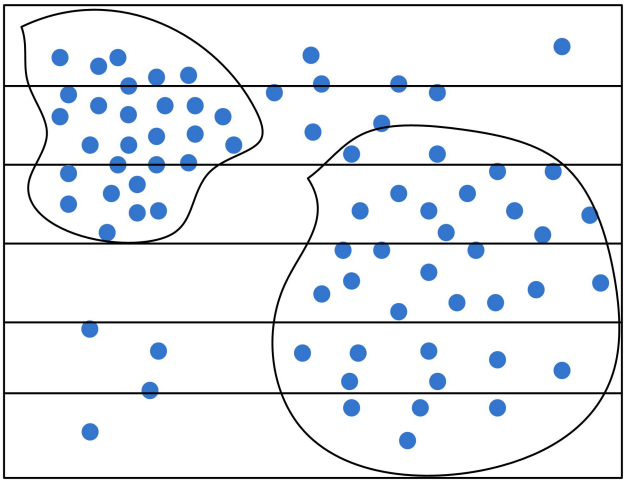


Figure 8 sample distribution of system test data

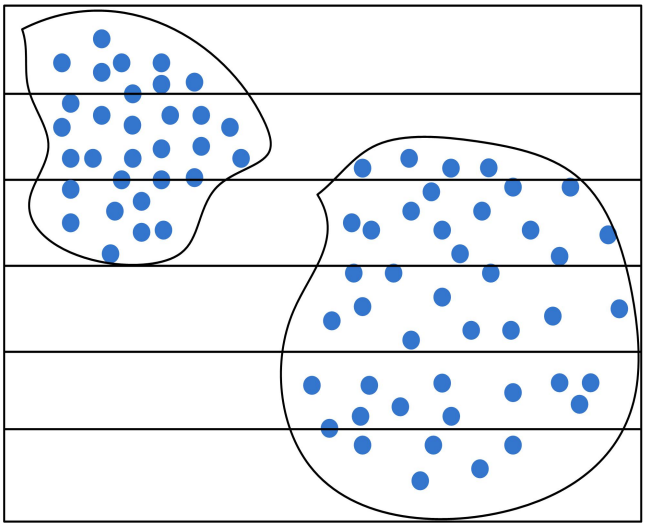
The above database is taken as the data sample of this experiment. The original data mining system and the designed mining system are used to process the test sample database, and the data clustering ability of the two databases is compared.

4.2 Test result analysis

Through the above design, the system test process is completed. The specific system test results are shown below.



(a) clustering results of the original system



(b) the clustering results of the system are designed in this paper

Figure 9 comparison of system test results

The experimental results show that the original data mining system is unable to realize the high-precision clustering convergence for the monitoring information of English multimedia independent learning behavior, and some information points are in discrete state. It can be seen from the above that the clustering effect of the original data mining system is poor. The data mining system designed in this paper has good convergence of clustering results and can complete complete clustering mining within the data set without data dispersion. The data can prove that the original data mining system can not reflect students' learning situation

comprehensively and timely, and the application effect is not ideal. The data mining system designed in this paper can reflect the learning status of all students and is more effective. To sum up, Web mining technology can effectively improve the classification clustering ability of the system, and the system performance of the data mining system designed in this paper is better than the original data mining system.

5 Conclusion

The web-based multimedia self-learning monitoring data mining system can quickly, efficiently and accurately understand the self-learning ability and behavior status of English multimedia in school. For students, this system can stimulate the motivation of independent learning, strengthen the awareness of independent learning, guide and establish the self-confidence of independent learning, develop correct learning habits, and train the ability of self-evaluation process. For teachers, the system can timely feedback the learning situation, promote teachers to continue to learn relevant theoretical knowledge, and actively explore, adjust and improve learning methods in practice, so as to achieve the educational purpose of constantly improving the ability of independent learning and improve the quality of teaching.

In order to further promote the development of this technology, other performance indicators of the system can be tested, so as to continuously improve and improve the system performance, better realize the mining of English multimedia independent learning monitoring data, continuously improve students' English scores, and promote the development of education in China.

6 Fund projects

Science and Technology Research Project of Jiangxi Provincial Department of Education (GJJ160815)

References

[1]Jing,H. Min,Z. Wei,X. et al.: A Study on the Diagnostic Model for the Line Loss in the Low-Voltage Transformer District Based on the Data Mining Technology for the Synchronous Line Loss System. Journal of Chongqing Electric Power College. Vol. 38, no. 06, pp. 48-53（2018）

[2]Fulin, Z. Bo, H. Xuquan, L.: Air conditioning measurement and control system based on ZigBee and data mining technology in the era of big data. Journal of Qingdao Technological University. Vol. 39, no. 06, pp. 81-84+108（2018）

[3] ZuKun, G. WeiShan, Z. ZhiChao, W. et al.: Knowledge Mining System for Petroleum Industry Driven by Web Architecture. Computer Systems & Applications. Vol. 28, no. 11, pp. 132-137（2019）

[4] Ying, Z.: Fast Time Sequential Data Mining Algorithm Based on Grey System Theory. Journal of Xi'an University(Natural Science Edition). Vol. 21, no. 05, pp. 40-43（2018）

[5] Zibo, G. Taili, D. Yong, Z. et al.: Application of Clustering Algorithm for Data Mining in Ship Energy Efficiency. Journal of Wuhan University of Technology(Transportation Science & Engineering). Vol. 43, no. 02, pp. 286-290（2019）

[6] Jianglin, Z. Yachao, Z. Juhua, H. et al.: A load clustering algorithm based on discrete wavelet transform and fuzzy K-modes. Electric Power Automation Equipment. Vol. 39, no. 02, pp. 100-106+122（2019）

[7] Zhejun, X.: Research into a Data Mining Algorithm in Cloud Computing. Bulletin of Science and Technology. Vol. 34, no. 11, pp. 209-215+220（2018）

[8] Xiaoting, Z. Yin, L. Jinglei, T.: Big Data Distribution System Design and Simulation Based on Optimized Clustering Algorithm. Computer Simulation. Vol. 35, no. 12, pp. 204-207（2018）

[9]Lei, X. Longjie, W. Gansan, P. et al.: Uncertain data clustering based on probability distribution. Journal of Shenzhen Institute of Information Technology. Vol. 16, no. 05, pp. 7-14（2018）

[10] Gaozhi, F.: Construction of tourism intelligent management platform based on Web technology. Modern Electronics Technique. Vol. 41, no. 11, pp. 69-72（2018）