Personalized Distance Education System Based on Data Mining

Zhimin Qi
Jilin Normal University, Jilin, China
zhiminqi2123@21cn.com

Abstract—To improve the poor intelligence and personalized service of the learning system in current distance education and training, a personalized learning system model based on data mining technology was proposed. Then, the method of applying the decision tree and BP neural network algorithm to design the system was described in detail. Finally, the core module of the personalized distance education system combined with Web was designed. The personalized function module introduced how the user operates the intelligent Web reasoning in the module according to the user input learning information and gave the most matching learning materials. Finally, the application results showed that the system greatly improved the personalized service module.

Keywords—learning system; data mining technology; BP neural network algorithm

1 Introduction

Modern distance education uses the mature computer network, multimedia technology and information guarantee technology to carry on the transmission of video, audio and multimedia data anytime and anywhere. The educational mode of interaction and visualization is becoming popular. This model is a relatively new type of education. It is a form of education that is aimed at all people in society. This way of education is characterized by fairness, equality and cheap price, which enable the majority of learners to accept and acquire knowledge. Fundamentally, learners will not be constrained by teaching place and teaching time in the past. The most educational resources and the best websites can be owned by learners anywhere in the world at any time.

From traditional teaching mode to modern network teaching, this is a new network teaching mode at first. The starting point of this new teaching mode is to learn and choose based on learners' hobbies and curriculum projects. The overall content, the way of learning, the method of learning, the effective time of learning, the determination of the learning place and the instructor can all choose according to the individual's actual situation. The distance education model is a combination of graphs, sounds, and texts. The personal and personalized interface is more conducive to the actual characteristics of distance learners. It is a collection of knowledge information data-

bases that can stimulate students' interests and hobbies from the actual situation. With such interests and hobbies, more creative learning methods can be developed to create certain favorable conditions, so that learners can quickly and easily learn relevant knowledge. In this education model, resources are shared because of its extensive use of networks. This saves the economic cost and a lot of manpower, material resources, and time, and obtains a greater teaching effect. Students can be taught by famous teachers. Students can also learn about content exchanges and discussions to achieve the goal of brainstorming.

Data mining can discover unknown or valuable data from a large amount of data. It is a scientific research field integrating computer, statistics, simulation, artificial intelligence, and database technology. To build a successful and personalized remote network education platform that meets the requirements, data mining must be used. Each student's browsing methods, notes, and online records are fully utilized. Each student's comprehensive ability data was collected. Teaching experience has been turned into a way that computers can operate.

The emergence of distance education has provided a quick and convenient way for the majority of scholars. Personalized distance education system uses the corresponding mechanism. Based on the student's personality characteristics, a navigation system was created. Multi-functional, multi-channel and multi-format intelligent distance learning environments are designed. It can realize the diversified application of web-based distance education system. The design and implementation of the intelligent learning recommender based on the support vector calculation method has been scientifically studied. Based on the understanding of each student's learning situation and interests, a scheme plan for the learning recommender of the vector calculation method is proposed.

2   State of the art

Modern distance education is mainly based on the Internet. It has gradually been recognized by people. Chawinga pointed out that by the year 2000, many countries in the world had already conducted distance education. 85% of teachers worldwide have their own websites on the Internet. 25% of them have launched online education courses [1]. In Europe and the United States, their online education is at the forefront of the world and has formed a certain scale.

The new development direction of ITS is cooperative teaching mode, cognitive construction of each student model, and multimedia assisted teaching system. The National Science Foundation has invested in these three items for a total of 1.04587 US dollars for scientific research of intelligent system for learning and creation. Memphis University of America Zhang et al. developed a scientific study of the 17-year smart learning system [2]. Through AI cognitive science and complex scientific research, scholar Ghamdi found that this system can teach different subjects. The system can respond appropriately to each student's learning problems, with instructions or hints given by the computer [3].

In recent years, network-based ITS has become a hot topic of scientific research, such as the ELM-ART system developed by Yang et al. [4]. The distance learning system for personalized distance learning has a project developed by Tosho et al. of University of Illinois. From the beginning of 1980, scientific research began and an intelligent instruction system (based on language dialogs) was established. Learners can solve many problems through computer-based dialogue counseling [5].

At present, for the purpose of enhancing the teaching system of personalized distance learning, there is a popular foreign student model based on Bayesian network. According to each student's evidence of the individualized distance learning system in the learning process, reasoning and predicting each student's next operation are performed. At the same time, many foreign scholars such as Gijselaers are currently conducting deeper scientific research on ITS systems. It mainly includes well-established and convenient modules for each student and full-featured teaching modules, friendly collaborative exchanges, and exploration of the mining of learning logs for each student (emotional, intelligent applications in ITS systems, etc.) [6].

Some information systems have been developed in China for many years, and there are rapid developments. This is just starting to be unstable and unsatisfactory. It can be seen that the development of the personalized distance education management system in China is currently unbalanced. For example, the video and education management information system has just begun. Some well-developed regions have followed the situation and are making better use of distance education. The development of the western region is particularly backward, and even temperature and food cannot be solved, let alone distance education.

Various educational institutions at home and abroad attach great importance to the construction of distance education schools. The distance education system will have a period of rapid development in the near future. It is mainly reflected in the following aspects: high efficiency, regionalization, standardization, and intelligence. At present, modern distance education is a form of online education. It includes three levels of work, study, and promotion. The scholar Samigulina thinks that the modern distance education system consists of two parts. This is mainly the hardware system and the software system. The hardware system refers to servers with better performance, basic network top devices, complete courseware, personal operating systems and systems with good performance. Software system mainly refers to the use of computer data in the teaching process, such as learning database, learning data packages, subject information and many other information [7].

Lewis thinks that the hardware system part of the server is a very important part [8]. Because the database is stored in the server, the IBM server is selected, which is currently recognized as the main server in the world. It is characterized by high quality and stable performance. The frequency of use in enterprises and universities is very high. On the network device, CISCO's network router is selected to keep the network communication smoothly.

3 Methodology

3.1 The overall design function of the system

The goal of this system is to design a way for students to learn. At the same time, the student's learning environment is designed based on the teaching of the teacher.
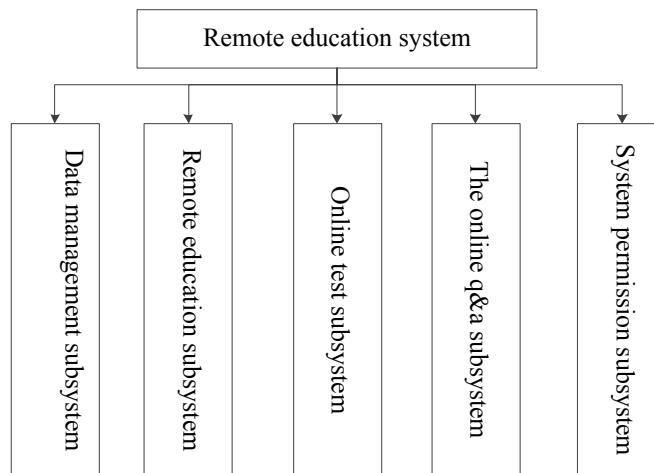

Fig. 1. Overall design frame diagram

Database entity objects are used in distance education systems. The ER diagram of the entity is described. Teachers in distance education occupy a very important part in the system. The attributes of the entity include some fixed attributes including the teacher number, the teacher's name, the type of the teacher's subject, the teacher's contact number, and the teacher's contact address. The attributes of the system user include the attributes such as the user name, user password, user's real name, user's contact phone number, user's affiliation unit, and user contact information. The item bank entity occupies a very important position in the distance education system, which involves the most core part of the system. The item bank entity properties include the item bank number, the item bank contents, the item number of the item bank, and the teacher management number of the item bank. The administrator in the system manages the user's rights, as well as the query of the title data. It mainly includes the administrator number, administrator's name, administrator encrypted password, administrator's unit, contact phone and other attributes. The overall data structure of the asset management information system of the distance education system is formed through the mutual connection and combination of each. The ER diagram of the relational model reflects the interrelationships between internal entities. The ER diagram of this system is shown in Figure 2.
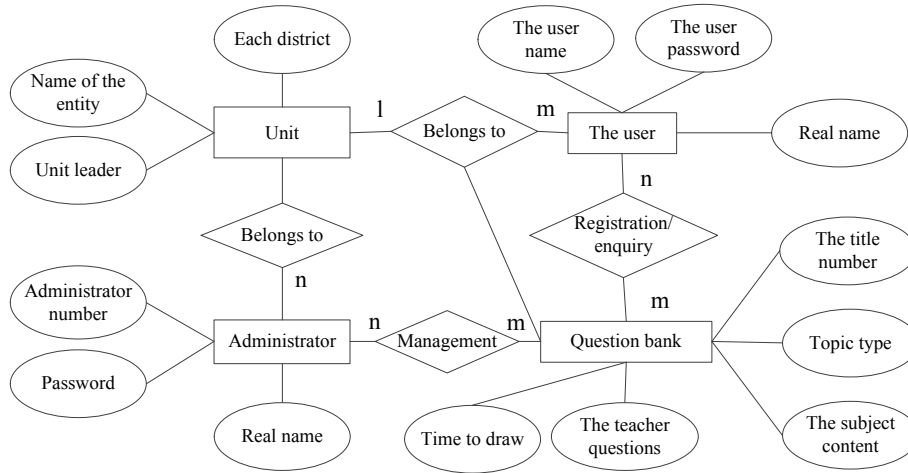
Fig. 2. Education management system data structure ER diagram

3.2 Design and modeling of the system

In this system, the QAM platform module is mainly used as an example. The answering system is mainly divided into three kinds of users: administrators, teachers and students. According to different levels of privilege of users, remote answer questions are mainly divided into three functional modules: administrator, teacher and student modules. Through these three modules, the relevant functions of the system are managed. Through the different permissions, the user manages the module after logging in to the system.

As an important part of the background, the database plays a very important role in the system. The quality of the database directly affects the efficiency of the system and the maintenance of the system. At the same time, design must also take into account the need for good security and scalability.

Taking an administrator's login table data dictionary as an example, the database used for the development of the distance education system at Shaoyang Chaoyang Training School is described as follows:

Table 1. The administrator logs in to the data dictionary

Filename	Administrator login table
Alias	Administrator login information table
Key code operator number	Operator number
Describe	Administrator login basic information
Composition	Operator number + user name + password + permissions
Note	

3.3   Detailed design of the system

Each distance education system has an entry landing subsystem. It is an important standard for measuring the security of system login. There are three levels of users, namely administrators, instructors, and students. The login system flow chart is shown in Figure 3.

Fig. 3. User login flowchart

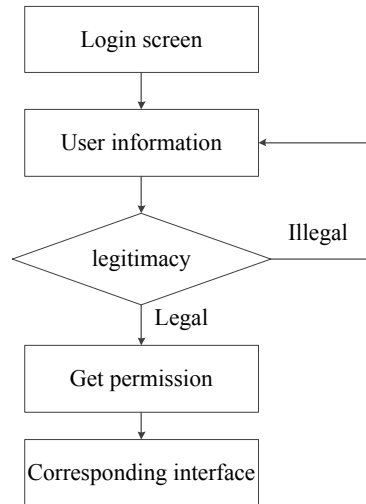According to different user identities, the corresponding interface is opened. Its main code is as follows:

```
<form name="form1" action="">
<tr><td> username:</td> <td><input name="names" type="text"></td></tr>
<tr><td> password:</td> <td><input name="pwds" type="password"></td></tr>
</form>
<%
u=request.form("names")
p=rsquest.form("pwds")
if u="sx0001" and p="sx0001" then Response.Redirect "index.asp"
end if
if u="sx0002" and p="sx0002" then Response.Redirect "index2.asp"
end if
%>
```

The main source of the construction of the examination subsystem is the construction of question banks. Through the construction of the topic, the tutor operates on the questions in the question bank, so that the examination subsystem will rely on the questions in the question bank for examination. The implementation process is as follows. The teacher prepares the questions for the first set of topics for classification. The teacher inputs the relevant information of the questions in the module. After the

questions are entered, user clicks the submit button. It is possible to write the title through the database.

In this system, topics are constantly updated. When the questions in the bank are used for a period of time, the questions are properly deleted and new questions are entered. Some topics are completely denied, and others are subject to modification. The revised topics include the name, attribute and other properties of the title. Repeated questions need to be fully modified.

The second is the design of the test paper subsystem, which includes test paper generation and test paper maintenance. The function of test paper generation is to compose test papers by using manual group questions. The process of implementation is as follows. The system first uses the manual questioning algorithm to automatically organize the test according to the question scores and difficulty factors set by the teacher. The automatic assembly flow chart of the system is shown in Figure 5. Test paper maintenance is mainly to modify, delete, and add operations to the contents of test papers.
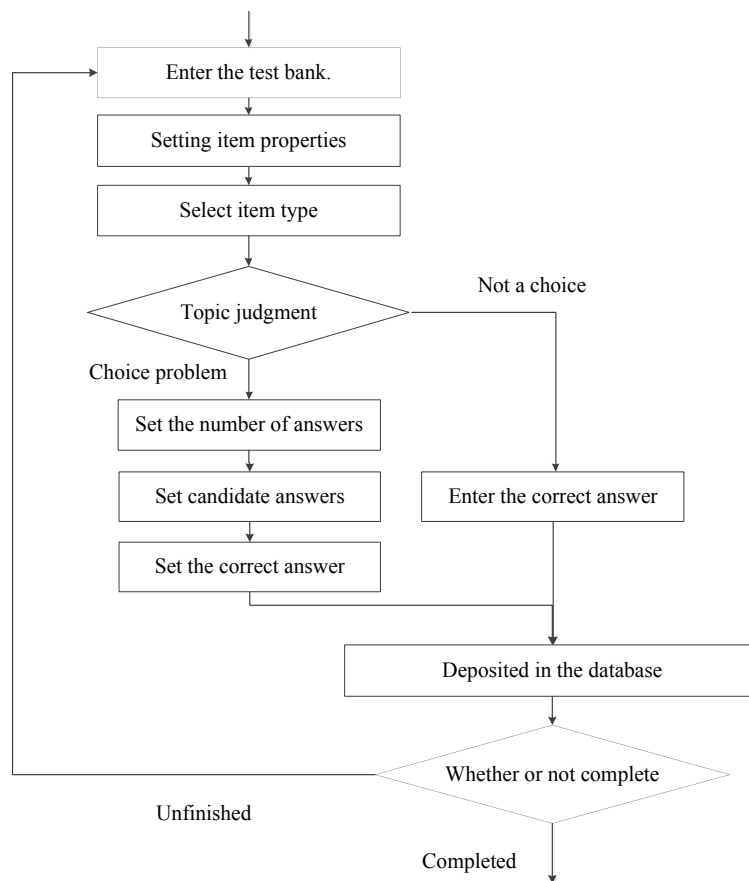
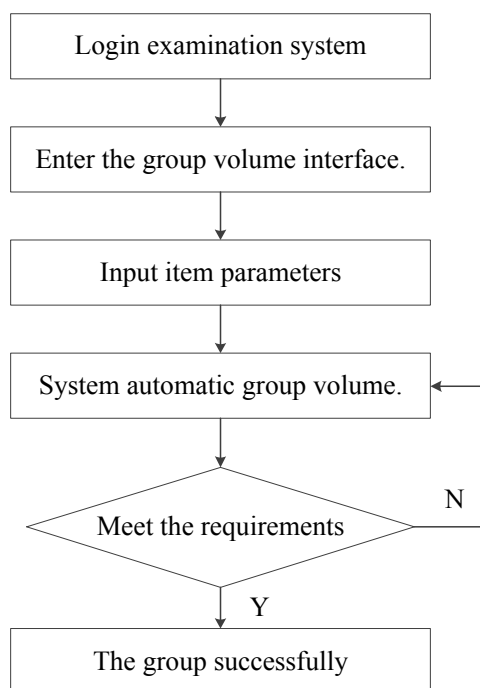**Fig. 4.** All kinds of topics are recorded in the flow chart

Fig. 5. System automatic group volume flow chart

The third is the design of the examination management subsystem. It includes the following aspects. During the examination process, the invigilator handles possible accidents in the examination to ensure that the students are fair and equitable in the process of the exam. The exam was adopted. The question bank group will roll out the questions. During the examination process, if the test student's answers to questions are more than 50%, the teacher can set the student's test status so that he can no longer take the test and re-arrange the test. Through marking, the teacher can understand the students' mastery. There are two methods of marking: one is the automatic judgment of the system, and the other is the teacher's marking. Performance analysis is the last link in the management of the paper. Through this link, students' learning is obtained. This can help teachers understand students' understanding of knowledge. Teachers should provide students with appropriate help for problems.

The fourth is the design of the answering subsystem. The answering system is to answer students' puzzles and difficulties in the learning process. Through this system, students can ask questions to their teachers so that they can better consolidate what they have learned. The system consists of two parts. Students upload questions in the corresponding Q & A section and the teacher answers the questions. When uploading a question, students should clearly state the title, the content of the question, and the confusion in the Q&A section. This will make it easier for tutors to answer questions. The tutor enters the answering system through the corresponding account number and password, and answers the questions of the classmates. The system requires the

teacher to spend no more than 12 hours in the process of answering questions. In this way, the tutor can be asked from a higher level to better answer the questions.

3.4   Multidimensional OLAP of Web log data

The application of On Line Analytical Processing (OLAP) in data warehouse platform is becoming increasingly widespread. OLAP is used for online analysis of data warehouses. It can perform multi-dimensional data analysis of various sizes, which is conducive to effective data mining. Because the core of OLAP is a multi-dimensional data structure, the Web log cube must be established first to perform OLAP analysis of Web logs. According to the actual situation of the TVU remote education website, the following dimensions are set. The domain name dimension attribute values are "COM", "EDU", "GOV", and "NET". The user dimension attribute values are "distance education specialists", "open undergraduate students" and "non-open distance students". The user dimension can be further refined according to the professional. The content dimension is "Training Course", "Teaching Courseware", "Video on Demand", "BBS Forum", "Enrollment Enquiry", "Department Homepage", "VBI Information", "Online Classroom", etc. It can be further subdivided by course.

Log files and user usage of the Web are analyzed online. Some commonly used data are summarized. The user's actions are divided into behaviors. The use of parts or features, the probability of events, the distribution of different types of student users, the access patterns and differences of users in different regions are discussed. The trends in user behavior over time and trends in network traffic over time are analyzed. The application of various components and features in the context, the typical sequence of events, the use and access of different user resources, and the changes in user behavior over time are understood. With the change of the quality of service, that is, the speed of the network, the user's usage patterns change. The distribution of network traffic is also different.

4   Result analysis and discussion

4.1   Application of multidimensional OLAP in Web Log of distance education website

The multi-dimensional OLAP algorithm of Web logs was applied to the Shaoxing Chaoyang training distance education website for experiments. By analyzing, some interesting laws were discovered. Students' learning patterns and learning habits are further mastered and analyzed. The time distribution of learning, the status of popular pages and popular courses are collected. It also provides an evaluation reference for the design of the network and the learning of the course.

The number of visit page requests for the web site visited was analyzed as a function of the month, and it was found that at the end of each semester, the number of visits to the web site before the final exam was the largest. In the final exam, students are required to review the exam and go to the online classroom to find answers and

review exercises. Users generally have a larger number of visits during the day from 12:00 to 13:00, and visits from 17:00 to 21:00 are relatively large. After the analysis, it was learned that many students used the Internet of the unit to access the Internet and had more visits to the site during the lunch break. Some students who do not have access to the Internet at work will return home at night or visit the school's electronic reading room. Teachers visit the site in the evening to answer more questions on the BBS.

It can be seen from the interviews that when the Peking University long-distance education enrollment was introduced, more people visited. Therefore, Peking University is more attractive to students. The Open Academy's information and related content are more frequently visited by students, which shows that students have become more adaptable to find relevant information online. Through multi-dimensional analysis, it has been found that open undergraduate students have more visits to online classes than open majors for a long time. However, the total number of visits is more than the number of open students, and it is known through analysis that the number of college students is far greater than that of open undergraduate students.

In addition, it was found that the teacher's guidance played a positive role in the number of visits. Teacher-led pages and courses have a larger number of visits; otherwise, they are smaller. Through the OLAP of the log file, the student's time on the page and the large amount of page (course) access can be obtained, which is useful information for the course assessment. At the same time, some erroneous request pages on various colleges and departments pages were discovered and adjusted in a timely manner. It plays a very positive role in the further development of distance education and online classroom design.

4.2   Mining and analysis of association rules in Web log

Based on OLAP, the association rule mining algorithm was applied to the Ningbo TV University distance education website for experimentation. The results of the mining association rules are analyzed. It can be found that after visiting the Chaoyang Chaoyang Training School's page, visits on the user's department's home page or visits on the college's home page are often open to question and answer related pages. According to the association rules, it is recommended to put related pages together or create a hyperlink between two pages.

In addition, when users access online classroom, most users will access VOD. Therefore, VOD should be linked on the homepage individually, which is convenient for students and teachers to visit. Users will frequently visit some courses and pages at a certain time, but after a period of time, they will not. Later, it was found that this was related to the teacher's guidance. It can be seen that the guidance of teachers plays a positive role.

Using data mining association rules, users' access behavior can be predicted. It enables the web server to send the web pages that users may browse in advance to the proxy server or user server. This is how to use the association rules to realize the pre-storage of the server. With this method, the slow download problem of distance education website can be solved.

According to this access mode, when the user requests Ui, Uj can be sent to the user's network cache or proxy server in advance. At this point, the user did not request Uj. In this way, when the user next requests Uj, Uj will be displayed immediately. From the user's point of view, the speed of the network is "fastened." If Uj=>Uk, and there is maximum credibility, Uk and Uj can be sent to the user at the same time, and the same rule can be used for analogy.

4.3   Educational decision making and optimizing distance teaching

According to the different management characteristics of each college, information is classified from a higher level, and abstraction and comprehensive implementation are used to establish a data warehouse. Through data warehouses, data mining and OLAP are processed to get valuable information. It provides a scientific data base for each college's decision makers. Teaching management was effectively improved.

The era of popularization of the multimedia network was mainly reflected in teaching, and a new long-distance learning model was realized. The characteristics of this model are as follows. Using multimedia networks, it provides a large amount of learning resources. The main teaching mode is autonomous learning. Compared with traditional learning, the remote network learning environment mainly has the following features. Students can learn as needed. Autonomous learning and ubiquitous learning are the main methods. Learning ability is improved. Students can learn any course section at any time and place. Anyone can start learning and improve their ability to learn and innovate. Research training was improved. However, this learning method also has certain flaws. Teachers can't implement effective management for students' learning and can't grasp the progress of students' learning, students' learning habits and ability. Because knowledge is non-linear, students' learning of hyperlinks is easily disorientated. The burden is increased and learning efficiency is reduced. Fixed interface mode cannot provide students with personalized management services. In addition, it cannot be taught in accordance with their aptitude. The above problems illustrate that this new model is an important factor.

4.4   Function test of distance education system

With the test tool, a variety of normal simulations and loads are performed. All performance indicators of the system are tested. Performance tests typically include pressure and load tests, and pressure and load tests can be tested simultaneously. With load testing, various performance tests can be performed under different workloads. As the load test continues to increase, every indicator of the system changes. With stress testing, the system's bottleneck and performance were determined and tested at the system service level.

Mercury Interactive's LoadRunner test tool and analysis software were used to test the system. From the test results, the response time of each operation increased with the number of users, but none of them exceeded 5 seconds. It was within the acceptable range.

When analyzing different user operating systems, the system resource cup utilization is obtained. It can be seen that when the situation is 2000 users, 3000 users, and 4000 users, the CPU utilization rate is within the normal range, and the system performs well. During the system test, because the occupied memory could not be released, the response time of the system under test was significantly increased after a long period of operation, and the processing capability was significantly reduced. When the user logs in, the system automatically generates a session and occupies part of the memory. The expiration time of the session is set to 2 hours. According to the user's habit analysis, when the user exits by using the method of directly closing the IE window to exit the system, this session is not released and it continues to occupy memory. No exit operation was performed during the test, resulting in a large number of user sessions not being released. After the program is modified, the session is cleared again when users log in again, and after the test is re-executed, this phenomenon disappears.

5   Conclusions

Web mining technology is applied to distance education. The useful information of a large number of users' learning behaviors is tapped and analyzed to help solve problems. A personalized distance education model of online recommendation learning based on web mining is proposed. Web mining technology is applied to distance education, which can make full use of the various information accumulated on the site. Therefore, it can help curriculum designers and managers redesign courses and reconstruct sites. The design of curriculum is more reasonable and more consistent with the laws of distance learning and distance education. Learners can give full play to their personality. Teachers truly teach students in accordance with their aptitude and carry out personalized education.

6   Acknowledgment

7   References

[1] Chawinga, W. D. Paxton, A. (2016). Increasing access to higher education through open and distance learning: empirical findings from mzuzu university, malawi. International Review of Research in Open & Distributed Learning, 17(4):5-8. https://doi.org/10.19173/irrodl.v17i4.2409

[2] Zhang, L. C., Worthington, A. C. (2016). Scale and scope economies of distance education in australian universities. Studies in Higher Education, 42(9): 1785-1799. https://doi.org/10.1080/03075079.2015.1126817

[3] Ghamdi, A. A., Samarji, A., Watt, A. (2016). Essential considerations in distance education in ksa: teacher immediacy in a virtual teaching and learning environment, 6: 17-22.

[4] Yang, Y., Kanhua, Y. U. (2016). Construction of distance education classroom in architecture specialty based on internet of things technology. International Journal of Emerging Technologies in Learning, 11(5): 56-59. https://doi.org/10.3991/ijet.v11i05.5695

[5] Tosho, A., Ariffin A, Sobihatun N. (2016). Conceptual design model of instructional interfaces: courseware for inclusive education system (iid4c) distance learning. International Journal of Distance Education Technologies, 14(4): 68-82. https://doi.org/10.4018/IJDET.2016100106

[6] Gijselaers, H. J. M., Kirschner, P. A., Verboon, P., Groot, R. H. M. D. (2016). Sedentary behavior and not physical activity predicts study progress in distance education. Learning & Individual Differences, 49: 224-229. https://doi.org/10.1016/j.lindif.2016.06.021

[7] Samigulina, G., & Samigulina, Z. (2016). Intelligent system of distance education of engineers, based on modern innovative technologies. Procedia - Social and Behavioral Sciences, 228: 229-236. https://doi.org/10.1016/j.sbspro.2016.07.034

[8] Lewis, C., Contrino, J. (2016). Making the invisible visible: personas and mental models of distance education library users. Journal of Library & Information Services in Distance Learning, 10: 1-15. https://doi.org/10.1080/1533290X.2016.1218813

8 Author

Zhimin Qi is with Jilin Normal University, Jilin, China.