KURAM VE UYGULAMADA EĞİTİM BİLİMLERİ EDUCATIONAL SCIENCES: THEORY & PRACTICE

Received: January 4, 2018
Revision received: July 4, 2018
Accepted: July 6, 2018

www.estp.com.tr

DOI 10.12738/estp.2018.6.163 ♦ December 2018 ♦ 18(6) ♦ 2633-2640

Research Article

Design of and Research on Autonomous Learning System for Distance Education based on Data Mining Technology

Yue Gao[1]
Beihua University

Shuying Zhang[2]
Beihua University

Abstract

Distance education has the time and space flexibility that traditional education can't match, but it also has its own limitations, which mainly lie in the insufficient guidance for students and the lack of effective monitoring of their learning. In order to overcome the inherent defects of distance education, in this paper, the author integrates the various resources that educational institutions can provide, and establishes a complete autonomous learning system for distance education that can provide multiple services. Using data mining technology, the author proposes the method for optimization of autonomous learning programs for distance education, which is described in detail with examples. The research results of this paper can provide some reference for the design of autonomous learning system for distance education.

Keywords

Data Mining • Education • Distance Education • Autonomous Learning

[1]College of Computer Science and Technology, Beihua University, Jilin 132013, China. Email: lunagao@beihua.edu.cn
[2]**Correspondence to:** Shuying Zhang, College of Computer Science and Technology, Beihua University, Jilin 132013, China. Email: jlzhangsy@beihua.edu.cn

Citation: Gao, Y., Zhang, S. Y. (2018). Design of and Research on Autonomous Learning System for Distance Education based on Data Mining Technology. *Educational Sciences: Theory & Practice, 18(6),* 2633-2640. http://dx.doi.org/10.12738/estp.2018.6.163

Compared with traditional education, distance education has many unique advantages (Moore, 1991) – students can participate in learning more flexibly without being limited by time and space. However, distance education also has many deficiencies that have not been unresolved deficiencies (Carswell, Thomas, Petre, Price, & Richards, 2000; Goth, 2009). Firstly, the interactivity of distance education is limited. Communication between teachers and students is insufficient, and difficult problems of students cannot be solved in the first time. To some extent, it reduces the efficiency of learning and the enthusiasm of students (Beldarrain, Y., 2006). Secondly, distance education is highly dependent on students' initiative and enthusiasm. It lacks effective monitoring and restraint in distance education. Compared with full-time students, students engaged in distance education need more subjective initiative. Once students' enthusiasm declines, distance education will not last (Grisolia & de Oliveira, 2016). One way to solve the shortcomings of distance education is to establish a complete autonomous learning system. In this system, teachers and students can communicate well, and students' enthusiasm can be fully mobilized, thereby improving the efficiency of distance education and ensuring the quality of distance education.

In recent years, data mining technology has developed rapidly and penetrated into various fields, making great contributions to the development of different fields. Distance education is actually an autonomous learning in which students have more initiative and greater independence than traditional education. In order to choose an appropriate distance education method, different learning programs must be formulated according to the individual characteristics of students. Data mining technology can be applied to students with different backgrounds, different learning abilities and different levels of knowledge. Therefore, in this paper, the author uses data mining technology for the design of and research on autonomous learning system for distance education.

Basic concepts of data mining and distance education

Data mining

Data mining is a widely used research method, the core of which is to use artificial intelligence, machine learning and other methods to discover hidden rules in big data and to dig out useful information (Romero, Ventura & García, 2008; Baert, Caron, Morge & Routier, 2017). Data mining mainly includes the following processes: 1. understand the research background, obtain prior information in relevant fields, and learn knowledge of relevant disciplines and fields; 2. identify the issues to be studied so that data processing and mining can be more targeted; 3. select the data of interest from the big data, clean the data, and clean up the redundancy and errors in the data; 4. reduce the dimension of data to facilitate subsequent processing; 5. choose the right data processing method. select a method, e.g., clustering analysis, neural network and random forest, to build the model according to the different research purpose; 6. test and evaluate the model; and 7. collate and analyze the results of data mining, and visualize the research results.

Table 1
Comparison of Classification Algorithms

	Decision tree	Neural Networks	Bayesian network	Support Vector Machines
Accuracy	B	C	B	D
Learning speed	C	A	D	A
Classification speed	D	D	D	D
Explanatory ability	D	A	D	A

Data mining methods can be divided into two categories: classification and clustering. Classification methods include neural network, support vector machine, etc.; clustering methods include k-mean clustering, etc. (Ramos, Silva, Rodrigues, Silva & Gomes, 2016; Cicero & Guadalupe, 2013). Data processing effectiveness varies by method. The effectiveness of commonly used data mining methods is shown in Table 1 and 2. "A" indicates the best effectiveness and "D" indicates the worst effectiveness (Liang, 2005; Lavrač, 1999).

Table 2
Comparison of Clustering Algorithms

	K-mean	Cure	Birch	Dbscan
Algorithm efficiency	C	B	A	C
Type of data	Values and symbols	Values	Values	Values

Distance education

Distance education originated from correspondence courses. In recent years, with the development of Internet technology, distance education has been in a vigorous development. The current distance education is mainly a form of education that relies on network technology and new media. Many educational institutions offer a range of online courses to meet the learning needs of different students, which is also a form of distance education.

"Distance" in distance education mainly refers to the geographical distance between teachers and students. Teachers and students no longer communicate face to face, but rely on the media for signal transmission. This transmission may or may not be real-time. Therefore, distance education has certain time and space restrictions on the communication between teachers and students; another characteristic of distance education is that it depends on the organization and planning of educational institutions – the arrangement of courses, the provision of teaching resources and the service support for students by educational institutions are important factors in determining the efficiency of distance education. In addition, distance education is more directed at individual targets of teaching – without partners to learn together, students are prone to loss of enthusiasm for learning and lack of communication.

Establishment of autonomous learning system for distance education

Structure of autonomous learning system for distance education

The structure of autonomous learning system for distance education needs to integrate various material and human resources. The whole system should be able to provide management services, information and consulting services, resource services, learning process services and so on. Among them, management services include teaching management and administration – teaching management is mainly composed of online registration, entrance testing, education management, etc., and administration is mainly the role of administrative department of educational institutions. Information service mainly refers to the indexing of teaching, resources and student information of distance education. Consulting services include both academic consulting services and non-teaching consulting services, which help students understand the structure and specific content of the whole autonomous learning system. Resource services mainly include human resources services and teaching resources services – the former includes teaching staff, tutors and other service personnel, reflecting the human assistance provided by the whole system, and the latter mainly includes various media libraries and test item resources, reflecting the learning resources provided by the whole system. Learning process services cover the whole process of services before, during and after learning. Before the learning process, the system provides entrance guidance and course introduction services. During the learning process, the system provides teaching and mentoring services. After the learning process, the system provides performance test and evaluation services. The structure of the whole autonomous learning is shown in Figure 1.

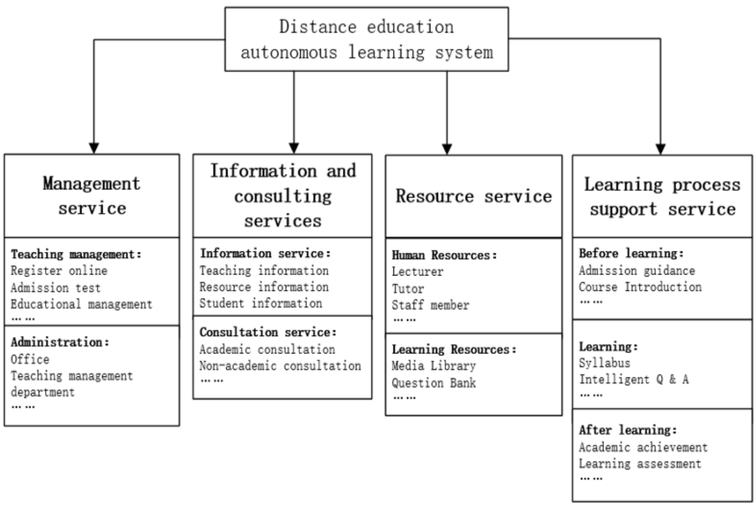

Figure 1. Distance education autonomous learning architecture

Optimization of autonomous learning programs for distance education

Students engaged in distance education have great differences in terms of learning objectives, learning styles, knowledge background and available time. To do well in distance education and achieve good results, we must develop a unique service system for different teaching programs according to the characteristics of each student. Therefore, in the whole autonomous learning system, it is not only necessary to build a complete framework structure, but also to select specific teaching programs and personal services. This process of selecting individual autonomous learning programs based on personal characteristics is referred to herein as optimization of autonomous learning program.

Obviously, the process of optimization of the autonomous learning program is a process of classification. By classifying and analyzing the teaching effectiveness as the target variable, finding the appropriate autonomous learning program and other influencing factors, it is actually a process of classification. The decision tree algorithm is a data mining algorithm that can effectively classify data. This method can be used to judge the feasible risk-making decision by evaluating the risks. It has many advantages: 1. the independent variables related to the target variables can be obtained to help decision makers choose the information they need to pay attention to; 2. large amounts of data can be processed quickly so as to build a simple model; and 3. the information is transparent and easy to understand and analyze. In this paper, the author chooses the decision tree method to optimize the autonomous learning programs. The specific process of program optimization is shown in Figure 2.

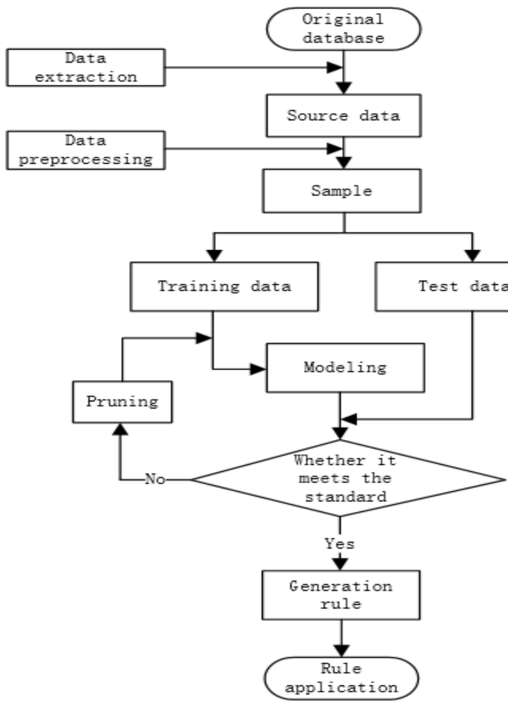

Figure 2. Autonomous learning program optimization process

Example of autonomous learning program optimization

In this paper, the author takes English learning as an example to verify and evaluate the above-mentioned independent learning system. The specific teaching content of the distance learning autonomous learning system includes learning, tutoring and testing. As for learning, it includes courseware learning and directed learning.

Courseware learning is mainly the autonomous learning of students, mainly preview, study and review of courses through reading the contents of courseware. Directed learning mainly refers to students' learning of courses released online by teachers. Tutoring mainly includes real-time and non-real-time tutoring. Real-time tutoring is real-time communication via the platform system. Non-real-time tutoring is conducted by means of students' messages and teachers' answers. The learning programs in this research are divided into four types, namely: courseware learning + real-time tutoring, courseware learning + non-real-time tutoring, directed learning + real-time tutoring, and directed learning + non-real-time tutoring. A total of 400 students participate in the distance education and are randomly and evenly assigned to different teaching programs during the teaching process. In addition, at the end of the courses, the students are tested to show the teaching results.

In this research, the students' age is used to characterize their learning ability, their English proficiency before taking the courses is used to characterize their original knowledge level, and the times of platform login are used to characterize the availability of learning time and the subjective initiative of learning.

The test results of the students after learning are classified – 60 points or more is qualified, and less than 60 points is unqualified. In the model, the author takes the students' scores as the dependent variable, and the students' age, English proficiency before taking the courses, teaching programs and the times of platform login as independent variables.

The decision tree model is implemented by the SPSS Modeler, and the importance ranking of the independent variables in the model is shown in Figure 3. It is found that, firstly, English proficiency before taking the courses has the greatest impact on the test scores after the students taking the courses, which means that the students with higher English proficiency before taking the courses still have higher English proficiency after the distance education; secondly, the more times of platform login, the harder the students are, the better the students' performance; thirdly, teaching program has a great impact on the students' learning effectiveness – when the teaching program of directed learning + real-time tutoring is selected, the students are most likely to pass the test; finally, the age of the students also has a certain impact on the learning effectiveness – when other factors are similar, the older students are less effective than the younger ones.

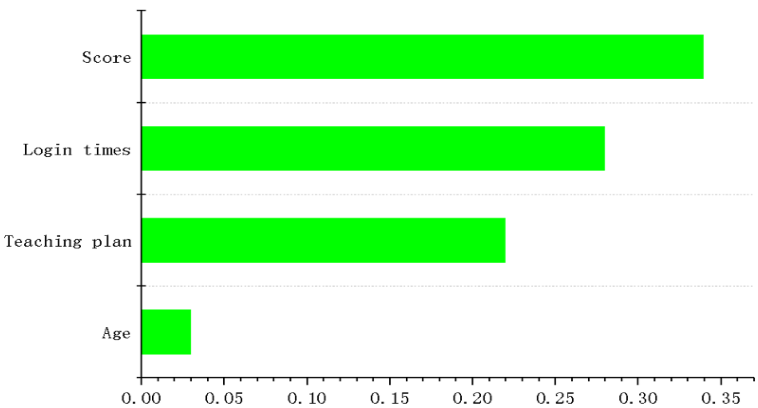

Figure 3. Decision tree model argument ordering

Table 3
Comparison of Clustering Algorithms

	Predicted as qualified	Predicted as failed
Qualified	73.2%	38.1%
Failed	26.8%	61.9%

The accuracy of the model prediction results is shown in Table 3. In the model, 73.2% of qualified students are classified correctly, and 26.8% of them are classified incorrectly; 61.9% of the unqualified students are classified correctly, and 38.1% of them are classified incorrectly. This shows that the model has a higher prediction accuracy for qualified students than unqualified students. Besides, the total accuracy rate of the model is 69.6%, relatively high, which indicates that the three factors of English proficiency before taking the courses, times of platform login and student's age can better predict whether a student can pass the test or not.

Conclusions

(1) In this study, the author establishes a complete autonomous learning system that includes management services, information and consulting services, resource services and learning process services, which are available from educational institutions.

(2) Using data mining technology, the author proposes a method for the optimization of autonomous learning programs for distance education and explains the specific implementation process of the method.

(3) Taking English teaching as an example to practice the method for the optimization of autonomous learning programs, the author finds that the students' original knowledge level, the degree of students' diligence, the type of learning program and their age have different degrees of influence on the their test results. In the optimization of learning programs, instructed learning + real-time tutoring is the best-performing learning program.

References

Baert, Q., Caron, A. C., Morge, M., & Routier, J. C. (2017). Fair task allocation for large data sets analysis. *Revue d'Intelligence Artificielle*, *31*(4), 401-426. http://dx.doi.org/10.3166/RIA.31.401-426

Beldarrain, Y. (2006). Distance education trends: integrating new technologies to foster student interaction and collaboration. *Distance Education, 27*(2), 139-153. http://dx.doi.org/10.1080/01587910600789498

Carswell, L., Thomas, P., Petre, M., Price, B., & Richards, M. (2000). Distance education via the internet: The student experience. *British Journal of Educational Technology*, *31*(1), 29-46.

http://dx.doi.org/10.1111/1467-8535.00133

Cicero, P., & Guadalupe, P. (2013). Pedagogicals recommendations for the use of moodle platform supported by data mining tools. *Applicable Analysis, 92(1)*, 1-14. http://dx.doi.org/10.1080/00036811.2011.601457

Grisolia, A. M., & de Oliveira, A. E. F. (2016). Pedagogical monitoring as a tool to reduce dropout in distance learning in family health. *BMC Medical Education*, *16*(1), 213.http://dx.doi.org/10.1186/s12909-016-0735-9

Goth, G. (2009). Network-enabled compulsory education getting big push. *IEEE Internet Computing, 13(1)*, 7-10. http://dx.doi.org/10.1109/MIC.2009.12

Lavrač, N. (1999). Selected techniques for data mining in medicine. *Artificial Intelligence in Medicine*, *16*(1), 3. http://dx.doi.org/10.1016/S0933-3657(98)00062-1

Liang, M. (2005). Data mining: Concepts, models, methods, and algorithms. *IIE Transactions, 36*(5), 495-496. http://dx.doi.org/10.1109/9780470544341

Moore, M. G. (1991). Distance education: the foundations of effective practice. *Journal of Higher Education, 63*(4), 468. http://dx.doi.org/10.2307/1982124

Ramos, J. L. C., Silva, R. E. D., Rodrigues, R. L., Silva, J. C. S., & Gomes, A. S. (2016). A comparative study between clustering methods in educational data mining. *IEEE Latin America Transactions, 14*(8), 3755-3761. http://dx.doi.org/10.1109/TLA.2016.7786360

Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, *51*(1), 368-384. http://dx.doi.org/10.1016/j.compedu.2007.05.016