



# Intelligent Prediction Method of Online Open Course Passing Rate of Automation Technology

Weiwei Zhang<sup>(✉)</sup> and Jiaofeng Wu

Guangdong Polytechnic of Environmental Protection Engineering, Foshan 528216, China  
wei20220104@163.com

**Abstract.** In the process of analyzing the passing rate of online open courses of automation technology, it is difficult to guarantee the prediction accuracy of the passing rate of courses due to the influence of course factors, learning behavior factors and environmental factors. Based on this, the intelligent prediction method of the passing rate of online open courses of automation technology is optimized and innovated. Firstly, collect the characteristic information of the online open courses of automation technology, then construct the evaluation index of the passing rate of the online open courses of automation technology, and finally optimize the evaluation algorithm to realize the intelligent prediction of the passing rate of the online open courses of automation technology. The experimental results show that the proposed intelligent prediction method of online open course passing rate of automation technology has high practicability and accuracy in the process of practical application, and fully meets the research requirements.

**Keywords:** Automation technology · Online open courses · Course passing rate

## 1 Introduction

With the advancement of information technology in the field of education, the data around the daily teaching operation of colleges and universities and students' learning are collected and stored by various information management platforms. Massive data can help schools understand and master the teaching operation and students' learning. However, at present, most ordinary undergraduate colleges still use these data in the stage of simple data collection and analysis, and do not conduct in-depth mining [1]. How to make the best use of the data stored in the database is a great challenge for most colleges and universities. In recent years, with the rise of the concept of big data, data mining and data analysis technology have gradually penetrated into the field of education, which makes it possible to explore the value of massive data in University databases [2]. By sorting out the research results of educational data mining applications of domestic scholars, it is found that many studies are relatively focused on data mining in the field of network education, while there is less research on the data generated by traditional classroom teaching stored in the database. The network education platform can monitor the whole process of students' course learning, and build a more complex

and scientific evaluation system by setting up a large number of data collection points [3]. Reference [4] introduces e-learning at different levels, describes its characteristics, attributes, advantages and the trend of further development of e-learning, and pays attention to the specific conditions of education policies, institutional types, training, etc. in the region. However, this method has some problems, such as less monitoring points in the learning process and single score structure. These problems bring challenges to data mining.

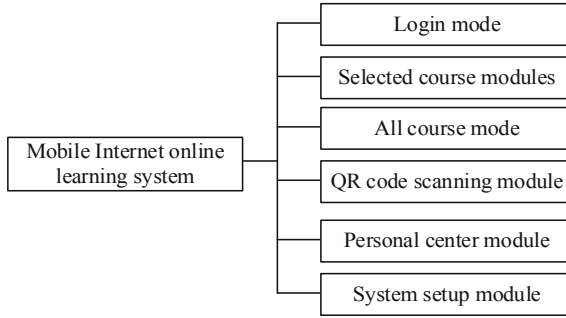
Based on the data mining research of scholars on e-learning, this paper builds a decision tree classification model by adding modeling attributes, decomposing the composition of grades, transforming the meaning of grades, and using the traditional data stored in the database, such as the personal information of undergraduate students majoring in non automated technology online open courses, the scores of automated technology online open courses and grade examinations, Predict the probability of students' online open courses through automation technology, and analyze the main attributes of the model. On this basis, the evaluation algorithm of course passing rate is designed, and based on this, the online open course passing rate evaluation of automation technology is realized. The experimental results show that the prediction performance of this method is good.

## 2 Intelligent Prediction of Online Open Course Passing Rate

### 2.1 Characteristic Information Collection of Online Open Courses of Automation Technology

The purpose of online open course pass rate prediction of automation technology is to calculate unknown system data information according to the known system data information, so as to understand the change trend of the system in advance and provide reference information for stabilizing or optimizing the performance of the system [5]. At present, most of the commonly used prediction methods are based on time series. Mining the change with time and the change law of system data information, such as fuzzy approximation time series prediction method, time series prediction method based on grey theory, neural network time series prediction method, etc. these prediction methods have their own advantages, but they also have common weaknesses, such as the time granularity of prediction, If the time granularity is too low, the algorithm will fall into details and find that the data change is too large to be accurately predicted [6]. According to the functional requirements, the online learning system can be divided into six modules (as shown in Fig. 1): login module, selected courses module, all courses module, QR code scanning module, personal center module and system setting module. The design of each module requires high cohesion and low coupling to maximize the reuse of functions, so as to improve the development efficiency of online teaching system and ensure the scalability of system functions.

According to the influencing factors of online open course learning of automation technology and the characteristics of daily work in Colleges and universities, this paper studies its forecast demand, and uses IBM's SPSS modeler18 0 as a data mining platform. Then the decision tree model based on C5.0 algorithm is selected as the data mining tool. C5.0 algorithm is in C4.5 algorithm, which is suitable for processing large data sets, with



**Fig. 1.** Influencing factors of online open course learning of automation technology

fast running speed and less memory [7]. C5. 0 algorithm divides the samples according to the field that can provide the maximum information gain, recursively divides the sub samples divided for the second time until they can't be divided again, and finally rechecks the division at the bottom, removes the branches with less influence, and obtains the final model. With the help of cloud model research method, this paper puts forward a single normal cloud prediction model, verifies its characteristics of low complexity and quasi prediction in daily work application in the learning process of automation technology online open course, and selects three dimensions: automation technology online open course factors, learning behavior factors and environmental factors to analyze and study whether it affects learners' final passing the course, The selected evaluation indicators are shown in Table 1.

**Table 1.** Evaluation index system affecting course passing rate

Overall indicators	Primary index	Secondary index	Tertiary indicators
Curriculum pass rate evaluation index system	Network course factors	Number of course chapters	0.658
		Number of course videos	0.256
		Number of course references	0.365
		Number of course examination questions	0.423
	Learning behavior factors	Login times	0.925
		Online learning duration	0.898

(continued)

**Table 1.** (continued)

Overall indicators	Primary index	Secondary index	Tertiary indicators
		Homework score	0.785
	Environmental factors	Online hours of teachers in the same period	0.098
		Number of students in the same period	0.523
		Educational background	0.665

**Table 2.** Experimental conditions

Operating system	Windows XP
Application software	Pydurn
Programing language	Pyhon
Frame	Tensorfow
CPU	Pentium MMX200
Graphics card	RTX2010
Memory	64GB
Hard disk	256G

Through the analysis of learners’ learning behavior data of online open courses of automation technology, the courses with the largest number of registered learners are selected, and the correlation between the above data and learners’ final scores is analyzed. The specific formula is as follows:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}} \tag{1}$$

where r is the correlation coefficient of x and y variables. In reality, machine learning tasks often contain many features. These features may have irrelevant or redundant features. Irrelevant features refer to features unrelated to current learning and research, and redundant features refer to those features that can be deduced from other features [8]. Irrelevant or redundant features may lead to model performance degradation, model training time is too long, model accuracy degradation, and dimension disaster. The feature selection method can reduce the feature dimension and optimize the model performance. Therefore, feature selection must be carried out before experimental research, as shown in Fig. 2:

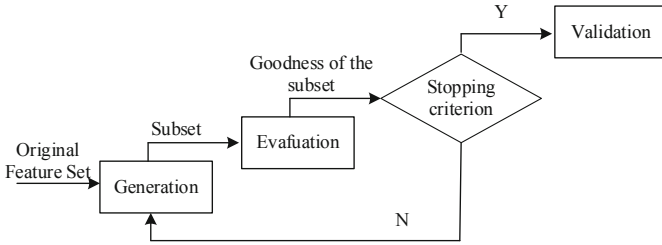


Fig. 2. Feature selection process

In the collection process, it is generally combined with genetic algorithm, particle swarm optimization algorithm, random resampling and so on. Random search will randomly generate a feature subset in the whole feature space as the initial state, give a certain weight to each feature, evaluate the feature with the obtained threshold according to some rules, and the feature higher than the threshold is regarded as an important feature. On the contrary, discard the feature lower than the threshold to ensure that the feature subset found each time is better than the last one. The random search method can effectively avoid the local optimization problem and ensure the good performance of the feature subset.

### 2.2 Evaluation Algorithm of Online Open Course Passing Rate of Automation Technology

From the perspective of online open courses of automation technology, most of them are displayed in the form of linear knowledge points in short video. According to the orderly logical structure of knowledge, find out the link points where people’s thinking affects each other, visualize the link points like a directed network diagram, and show the organic structure diagram of course knowledge. This can not only help teachers

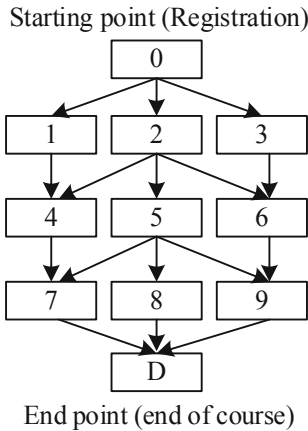


Fig. 3. Course structure network diagram

organize teaching content, but also better guide students to learn curriculum knowledge, provide students with personalized learning navigation, dynamically show the relationship between curriculum knowledge points and the video resources describing the curriculum, and effectively avoid the problem of lost navigation in learning. Therefore, learners can learn each knowledge point in the course through orderly short video, which can be simplified into a directed network diagram, as shown in Fig. 3:

Since the online learning system resources are mostly video based and the amount of uploaded management data is large, the resource management of the server is completed through the web online learning video management platform. The platform is implemented with SSH (struts 2.0 + Spring3.0 + hi hibernate 3.0) multi architecture. The front-end page uses extjs as the interface library and Ajax functional framework to connect the server for data operation and system maintenance. Web online learning video management platform mainly includes three parts: video management module, information notification module and system maintenance module. Its functional structure is shown in Fig. 4.

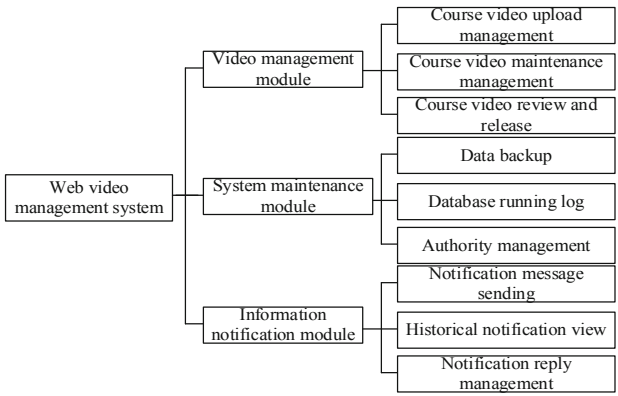


Fig. 4. Function structure of online learning management

The functional structure of server-side online learning system can produce two models: decision tree and rule set. The decision tree is directly generated by dividing samples by the algorithm. Each leaf node represents a specific data subset, and each sample in the data set belongs to only one leaf node. In other words, any given sample can only get one prediction result through the decision tree. The data classification technology in data mining technology is used to predict the results of online open course examination of automation technology, which goes through the steps of data extraction, data processing, decision tree construction, result prediction and decision tree optimization, as shown in Fig. 5:

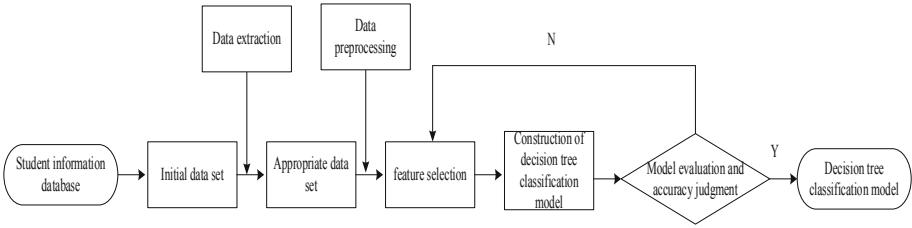


Fig. 5. Construction flow chart of pass rate prediction model

Let  $U$  be a quantitative universe expressed by exact numerical values, and  $C$  be a qualitative concept on  $U$ . if the quantitative value  $x \in U$  and  $x$  are the sub random realization of the qualitative concept  $C$ , the uncertainty  $u(x) \in [0, 1]$  of  $x$  to  $C$  is a random number with stable tendency.

$$u : U \rightarrow [0, 1] \quad \forall x \in U \quad x \rightarrow u(x) \tag{2}$$

For the qualitative concept in the quantitative domain expressed by exact numerical value, if the quantitative value  $x \in U$ , and  $u$  is the sub random realization of the qualitative concept  $U$ , and the uncertainty of  $\mu$ :

$$\mu = u - xe^{-\frac{(x-E_x)^2}{2(Em)^2}} \tag{3}$$

Then the distribution of  $E$  on the universe  $H$  is called a normal cloud. The existing data are used for experimental analysis to build a single normal cloud, which is described as follows:

$$C = \mu(Ex, EU, He) \tag{4}$$

In the prediction model of online education and teaching, it is very important to evaluate online education and teaching. There are various means (indicators) of teaching evaluation, including: course passing rate, achievement quality, student employment rate, student graduation rate and so on. In view of the different problems in the teaching process, we can choose the evaluation index as an important means to evaluate the

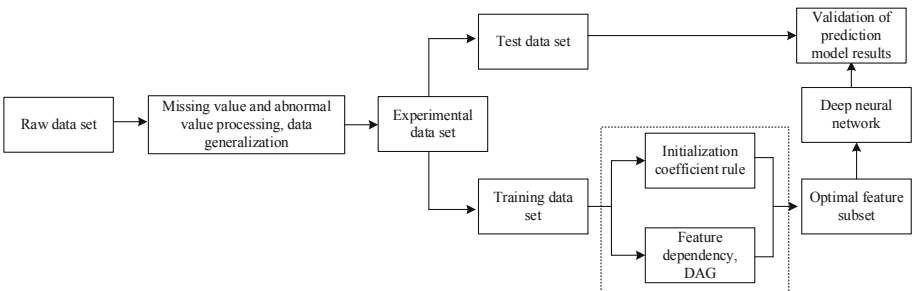


Fig. 6. Online education model

teaching quality, so as to improve the accuracy of education evaluation results and shorten the time consumption. In particular, the prediction model of online education pass rate is shown in Fig. 6:

According to the research, at present, among the many problems existing in the online education and teaching mode, the student passing rate (the student online classroom passing rate is low and the student graduation rate is low) is the main problem. This paper uses the prediction of student passing rate as one of the important methods in online education evaluation. In addition, through the evaluation of online education, teaching managers can also find out the methods to solve the problems existing in current online education as soon as possible. This paper uses fdnn algorithm to build a prediction model, finds out the most important characteristics affecting students' passing rate through experiments, and compares the accuracy of several machine learning algorithms to predict students' passing rate, so as to effectively promote students' learning, And strengthen teaching management to improve teaching quality. This paper is supported by the special fund for basic scientific research business expenses of central universities. In addition, this paper also has reference value for other evaluations of online education.

### 2.3 Implementation of Online Open Course Pass Rate Evaluation of Automation Technology

The automation technology online open course passing rate evaluation data management platform is mainly to collect user information and behavior data, process these data, and recommend relevant user information to DSP in RTB system, so as to realize the automation technology online open course passing rate evaluation, as shown in Fig. 7.

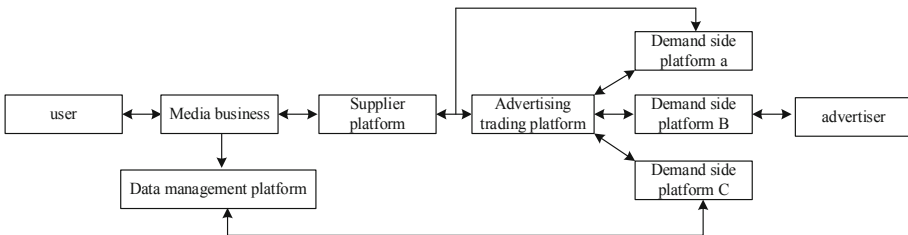
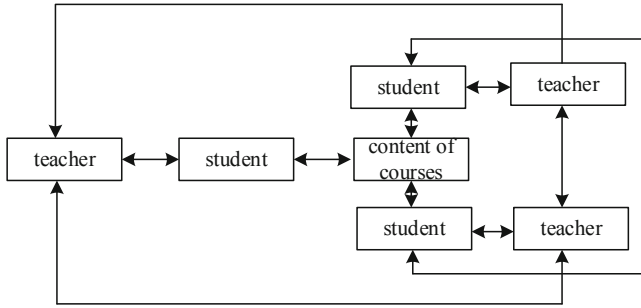


Fig. 7. Schematic diagram of RTB process

In the process of online open course pass rate evaluation of automation technology, the prediction content is the center of the whole prediction structure, that is, the interactive activities are centered on the “interactive content”. Firstly, based on the requirements of the syllabus, teachers should analyze the characteristics of learners “interactive content” and create an interactive situation, some problems and tasks; Second, the surrounding students learn the “content” of the platform in class, collect resources according to the teacher or take the initiative to investigate, solve problems, explore and discuss and communicate with teachers or other students in the process, and finally master the knowledge content, as shown in Fig. 8:

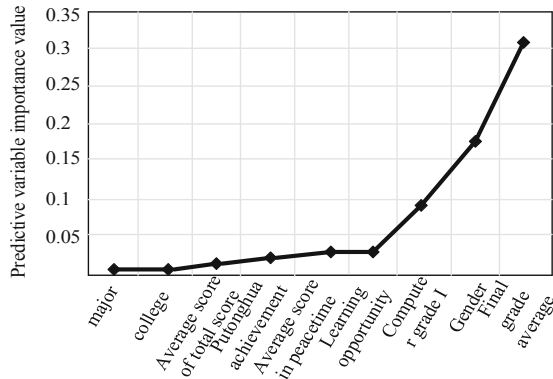


**Fig. 8.** Basic process of course prediction design

In model training, the loss function is usually used to measure the degree of model prediction error. The common loss functions mainly include exponential loss function, square loss function, absolute value loss function and cross entropy loss function. Like other CTR prediction models, the model in this paper also uses cross entropy loss function as the objective function of the model. The prediction platform plays a vital role in prediction activities and has many functional advantages, namely “Visual teaching environment, integrating various forms of media, media materials and intelligent information prediction, multi-directional interactive cooperation, opening teaching resources, multiple information stimulation of teaching, unlimited time and space, personalized autonomous learning, real-time information feedback, network resource sharing and information compatibility with the changes of the whole brain, etc. the interactive platform not only provides technical means and information for student investigation information resources, more importantly, provide a near real information environment for students’ autonomous learning research. Let the interactive content realize various representations, and enable students to obtain knowledge from various sensory stimuli through audio, video, animation and text, so that students can obtain knowledge through hypertext chain communication, virtual experiment and discussion, and finally complete the research task. In this process, the desire prediction platform provides a good environment for students’ autonomous learning with its unique openness, rich, diverse and interactive information resources. The client actively reads the course list information on the server and writes it to the client data cache, updates the client interface, and pushes the course video data to the client player cache, Realize point-to-point online real-time playback. The implementation method of throughput data transmission and communication module is shown in Fig. 9:



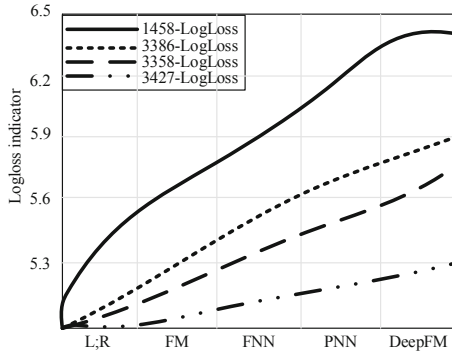
Based on C5.0 The decision tree classification method of 0 algorithm is based on the students' basic information, the scores of online open courses of automation technology and various grade examination scores. In the data mining environment of SPSS modeler, by adding modeling attributes Adjust the model parameters and other steps to finally build the prediction model of probability for the online open course of undergraduate automation technology. When building the decision tree in SPSS modeler environment, the importance of each attribute can be viewed by double clicking the "prediction result" node. The importance value of each attribute is shown in Fig. 10.



**Fig. 10.** Importance value of each predictive variable

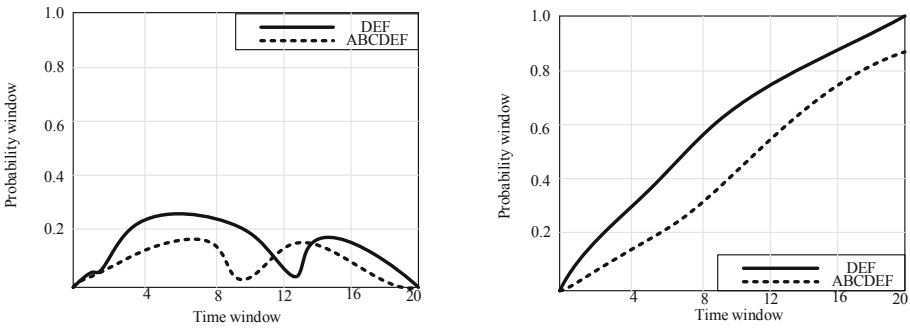
Figure 10 lists the experimental results of the last integrated models on the four verification sets of Ipin you, and compares the best models in each data set. It can be seen intuitively that after combining the model with the model integration framework proposed in Chapter I, the performance has been improved in the test set. Next, the performance of the integrated model will be tested on the test set. For the ensemble2 integrated model, the weight value learned for each single model is reserved for the test set experiment in the next section. The experiment is used to evaluate the results of the selected model combination framework on the test sets in the four data sets of Ipin you. During the experiment, the super parameters of each model are consistent with the super parameters set on the verification set. The test is carried out under different logarithmic losses (LogLoss). The specific results are shown in Fig. 11:

The improvement of the integrated model on the test set is smaller than that on the verification set. After analysis, there may be these reasons: first, the data distribution on the test set and the verification set is different, and the learning of the data between the verification set and the test set by a single model is different, which leads to the different improvement of the prediction performance of the later integrated models; Secondly, in model selection, based on the single model selected in the verification set, it may be the top n single models with the best performance in the test set. The weighted average method is used to calculate the model weight in ensemble2, which is more to fit the data distribution on the verification set. The model combination weight learned from the verification set is still used in model integration in the test set, The performance of the integrated model will not be better.



**Fig. 11.** Comparison of logloss on test set

Use MATLAB’s impulse (path) command for graphical visualization to obtain the frequency domain probability density curve and frequency domain probability distribution curve respectively for intuitive analysis and comparison, as shown in Fig. 12.

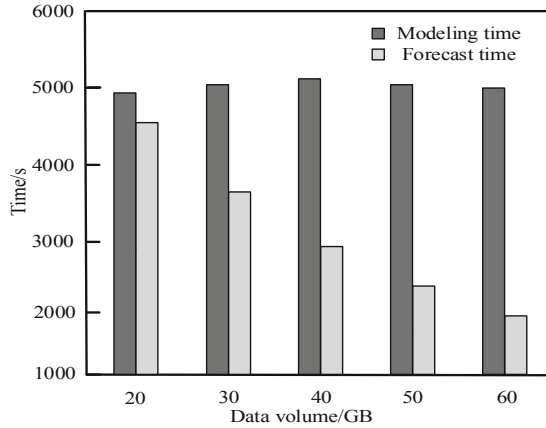


**Fig. 12.** Distribution of course passing rate

In Fig. 12, def represents that the random number of knowledge points is 3, and ABCDEF represents that the random number of knowledge points is 6. Under different constraints, the predicted course passing rate changes. To ensure that the course passing rate is more than 90%, the analysis shows that the course learning time needs more than 17h. The research shows that this method improves the calculation efficiency and can quickly visualize the graphics in the simulation and prediction of the online open course of practical automation technology.

The modeling time and prediction time of the design model are tested under the condition that the traditional data such as personal information, automation technology, online open course scores and grade examination scores are complete. The test results are shown in Fig. 13.

According to the test data of modeling time and prediction time in Fig. 13, when the historical data is complete, the modeling time of the design model is about 5000S, with little fluctuation, indicating that the modeling time is not affected by the amount



**Fig. 13.** Modeling time and prediction time when historical data is complete

of historical data, and the modeling time is short; With the increase of the amount of historical data, the prediction time of the online open course passing rate of automation technology is greatly shortened. The more data, the shorter the prediction time of the online open course passing rate of automation technology. The overall test results show that when the historical data is complete, the modeling time and prediction time of the design model are short.

## 4 Conclusion

Based on the traditional data such as personal information, automation technology, online open course scores and grade examination scores, this paper puts forward an intelligent prediction method for the passing rate of online open courses of automation technology. A decision tree classification model is built to predict the probability of students using automation technology to open courses online. In the research process, the basic attributes of this method are discussed through experiments, so as to better improve the teaching quality of online open courses of automation technology and ensure the teaching effect.

## References

1. Fei, Y.A., Erds, A.: Design for robot assembly: challenges of online education. *Procedia CIRP* **100**(3), 482–487 (2021)
2. Caliskan, S., Kurbanov, R.A., Platonova, R.I., et al.: Lecturers views of online instructors about distance education and adobe connect. *Int. J. Emerg. Technol. Learn. (iJET)* **15**(23), 145 (2020)
3. May, D.: Cross reality spaces in engineering education – online laboratories for supporting international student collaboration in merging realities. *Int. J. Online Eng. (iJOE)* **16**(3), 4 (2020)
4. Trybulska, E.S., Morze, N.: Web-based community-supported online education during the covid-19 pandemic. *Int. J. Web Based Comm.* **17**(1), 1 (2021)

5. Rodriguez, M.E., Guerrero-Roldan, A.E., Baneres, D., et al.: Students' perceptions of and behaviors toward cheating in online education. *Revista Iberoamericana de Tecnologias del Aprendizaje*, 99, 1–1 (2021)
6. Dama, C., Langford, M., Dan, U.: Teachers' agency and online education in times of crisis. *Comput. Hum. Behav.* **121**(3), 106793 (2021)
7. Wong, W.K., Chen, K.P., Lin, J.W.: Real-time data logging and online curve fitting using raspberry pi in physics laboratories. *Int. J. Distance Educ. Technol.* **18**(3), 57–77 (2020)
8. Gao, H.: Engineering training based on virtual simulation technology. *Comput. Simul.* **37**(07), 391–393+408 (2020)