



Research on the predictive model based on the depth of problem-solving discussion in MOOC forum

Jiansheng Li¹ · Linlin Li¹ · Zhixin Zhu¹ · Rustam Shadieva² 

Received: 1 March 2022 / Accepted: 23 February 2023 / Published online: 23 March 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

A discussion forum is an indispensable part of a massive open online course (MOOC) environment as it enables knowledge construction through learner-to-learner interaction such as discussion of solutions to assigned problems among learners. In this paper, a machine prediction model is built based on the data from the MOOC forum and the depth of discussion of solutions to assigned problems on the topic among students was analyzed. The data for this study was obtained from Modern educational technology course through Selenium with Python. The course has been offered to a total of 11,184 students from China seven times since February, 2016. The proposed model includes the formula of the depth of problem-solving discussion in MOOC forum and its prediction probability. The efficiency of the prediction model and the most important factor of the depth of problem-solving discussion in MOOC are explained in the paper. Based on the results, useful suggestions for effective teaching in MOOC forums are provided in the article.

Keywords MOOC · Learning topic · Depth of problem-solving discussion · Predictive Model

1 Introduction

With the spread of COVID-19, the teaching and learning format has changed from face-to-face to online one. For this reason, massive open online courses (MOOCs) have received increased attention (Mays et al., 2021), especially in universities. MOOCs are mostly free higher education courses delivered through the Internet and multimedia tools. Through MOOCs, thousands of students can freely learn and share the courses they are interested in. The three major MOOC platforms are: Coursera,

✉ Rustam Shadieva
rustamsh@gmail.com

¹ Nanjing Normal University, No.122, Ninghai Road, Nanjing 210097, Jiangsu Province, China

² Zhejiang University, No. 866, Yuhangtang Road, Hangzhou 310058, Zhejiang Province, China

Edx and Udacity. In the Chinese-speaking community, Xuetaang is the most popular MOOC platform (Hou, 2016). In general, MOOC platforms include the following three components: course content, community building tools, and platform tools. Of particular interest to this study is the course content component. It includes informational and interactive assets (Fauvel & Yu, 2016).

Despite tremendous success of MOOCs, there are a number of issues associated with MOOCs implementation. The most frequently cited issues are high dropout rate, low participation, and low quality of some MOOCs (AlQaidoom & Shah, 2020). In order to promote MOOCs, these issues need to be identified and addressed as early as possible.

In MOOCs, teachers are separated from students and students are separated from each other. For this reason, their online communication and interaction are very important for courses' success (Alqaidoom & Shah, 2020). Constructivism learning theory holds that knowledge is not directly taught by teachers, but is constructed by students' during in-depth discussion with teachers and/or other students (Moon et al., 2021). Sociological theories state that learning is an interactive process in nature. Therefore, interaction among MOOCs participants is important. For example, efficient interaction not only can promote the acquisition of new knowledge, but also improves the level of interaction (Tokur & Akgün, 2021). Research shows that online discussion is the core of online courses and can arouse students' positive perceptions and attitudes about learning in online mode (Guo et al., 2021).

Scholars often used such research methods as content analysis or questionnaire in their studies on online discussion, and they are usually based on a specific framework, e.g. the cognitive level, knowledge construction (Duruk et al., 2021; Hwang et al., 2018), critical thinking (Warsah et al., 2021) or community of inquiry (Garrison, 1991). Previously, content analysis was often carried out manually, requiring a lot of labor and money. In addition, only part of the data could be extracted for the content analysis. Questionnaire is easier to administer, arranged at the end of the study, and aims to evaluate students' perceptions of their thinking level in online discussions.

Advances in technology, especially data mining, make it possible to automate data processing on MOOCs. Such approach to analyze discussion content of MOOCs has attracted attention of many researchers. Among them, a research area that has gained significant popularity is how artificial intelligence and data mining can contribute to better understanding of the MOOC ecosystem and to improving it.

For example, Almatrafi et al. (2018) built a predictive model based on machine learning algorithms to identify an emergency discussion thread. This model can help teachers effectively navigate a bulletin board system (BBS) and assess posts with some priorities, which is of great help to effectively support teaching and learning and reduce students' dropout rate. Qin (2017) divided the discussion posts into such categories as "confusion," "explanation" and "irrelevant," and trained the support vector machine (SVM) classification model to predict them. Results of the above-mentioned studies demonstrated that applications of data mining were beneficial for MOOC teaching and learning. Therefore, on the basis of the above-mentioned research, this paper intends to establish a mathematical model for automatic prediction of in-depth discussion of learning topics

in MOOC forum. The prediction model of in-depth discussion distinguishes the content of MOOC discussion forum, provides teachers with guidance on the content of discussion forum, and encourages students to think deeply about assigned problems.

More specifically, the research problem in this paper covers the following two aspects:

- (1) To construct the automatic prediction model of the depth of discussion of learning topics;
- (2) To explore the predicted variables of the depth of discussion of learning topics.

2 The influencing factors of the depth of online discussion

Online discussion is a complex learning activity based on online interaction and involves many influencing factors. Mazzolini and Maddison (2007) analyzed discussion posts in 400 course forums and found that teacher participation rate, time of posting (i.e. during or at the end of the forum), and nature of posting (i.e. question, answer, or combination of both) influenced students' participation in online discussion. Based on the question perspective, Zhang and Si (2010) found that the speed of feedback to the question response, the type of question (cognitive-memory questions, convergent questions, divergent questions), and the presentation of question are all important factors in online discussion.

As for the depth of online discussion, Huang (2017) believed that the type of topic (related to / unrelated to teaching content), the number of postings and the size of groups were important variables to influence how deep participants discussed learning content. Similarly, the research by Spatariu et al. (2007) obtained similar results. For example, Spatariu et al. (2007) also thought that the depth of online discussion can be influenced by information type, cooperative learning structure and type (e.g. structured or not, role playing, etc.) and open questions. Besides, teaching assistant, learning support (e.g. designated discussion facilitator, facilitator questioning skills training, etc.) and group structure had impacts on students' critical thinking and their discussion depth. So, the depth of students' online discussion mainly depends on whether teachers provide feedback in time (Nandi et al., 2012) and the teachers' opinions and tendencies towards online discussion activities (i.e. the importance of online discussion) (Gerber et al., 2005). Besides, scholars suggested that the questions related to high cognitive levels also promote in-depth discussion online (Ertmer et al., 2011).

In summary, the influencing factors of the depth of online discussion are active participation of students, timely feedback from teachers, online interactive guidance, open discussion environment, discussion time limit, and challenging discussion tasks. These variables also determine the initiative of learners to participate in online discussion (Choi et al., 2005). Therefore, this study selects research variables according to the above literature.

3 Learning topics and discussion depth in MOOC

Communication and active participation are two crucial determinants in the quality of teaching and learning (Durairaj & Umar, 2015), especially in online learning environments. According to the social development theory (Vygotsky, 1987), the cognitive development of a learner depends on the social environment in which he/she is engaged in knowledge construction process through interaction with peers and the instructor. These interactions facilitate the acquisition, exchange and creation of knowledge together as well as sharing of experiences in the instructor-learner and learner-learner relationships (Tan, 2017).

MOOCs provide students with the opportunity to engage in discussion such as to talk about current coursework, to ask questions about content they are unclear about or to offer assistance to students with lower learning ability. In addition, the lecturer and teaching assistant visit the discussion forum to answer questions from students and provide feedback. Therefore, it is important for students to have active discussions on various learning topics in MOOC forum.

Learning topics in discussion forum may vary, for example, participants can discuss different types of questions of their MOOC, important concepts, participants' responses or feedback. Some of these topics are ephemeral, others are long-lasting. Depth of discussion refers to the level of online discussion such as very thorough discussion of a topic when most of its aspect are considered. Topic discussion is considered as the process of knowledge generation through exchanging of ideas related to a topic. In deep discussion, students need to clearly understand a topic and associated problem, and be able to critically check the logic of arguments (Wang, 2013). The depth of online discussion in this study is defined as students' integration of new knowledge into their original cognitive structure during the MOOC discussion. On this basis, new and existing knowledge is integrated and continuously applied to new learning situations; this process results in a solution to assigned problems and promotion of new knowledge construction. We classified the depth of learning topic discussion in MOOC forum based on Bloom's taxonomy (Anderson & Krathwohl, 2001). Cognitive process dimensions of the taxonomy are remember, understand, apply, analyze, evaluate and create (Anderson & Krathwohl, 2001). So, the present study ranked topic discussion level from low level to high level with such categories as shallow, medium and deep respectively. The topic discussion depth in MOOC forum is shown in Table 1.

Table 1 Topic discussion depth in MOOC forum based on Bloom's taxonomy (Anderson & Krathwohl, 2001)

Category (primary)	Category (secondary)	Cognitive process dimensions
Shallow	Remembering	Remember
	Shallow trigger	Remember
Middle	Induction, integration	Understand
	Comparing, explaining	Understand
	Application	Apply
Deep	In-depth reflection	Evaluate and create
	Analysis, critiquing	Analyze
	Internalized, transfer	Create

Based on the review of above-mentioned literature, this study intends to establish a mathematical model for automatic prediction of the depth of discussion of learning topics in MOOC forum.

4 Related work on learning prediction model

4.1 Feature selection

In order to use the machine learning method, most of the data need to be obtained from realistic scenarios, appropriate features need to be extracted from the data, and then they need to be represented as feature vectors. The original data set generally includes many redundant features or irrelevant features, which increase the dimension of the data and the complexity of the algorithm model. Feature selection is a process that finds the most effective features from the original feature set, reduces the dimension of the feature set and improves the classification accuracy under the premise of ensuring the invariance of similar samples, the distinguishability of different samples and noise robustness (He, 2015). Researchers used various approaches for feature selection. For example, Dash and Liu (1997) proposed a basic framework for feature selection which is composed of four steps: generation, evaluation, stopping criterion and validation. In our study, we propose a model to predict the discussion depth of learning topics in MOOC forums based on this basic framework.

4.2 Logistic regression algorithm

The prediction of topic discussion depth in MOOC forums can be regarded as a classification problem based on machine learning. Logistic regression algorithm is a generalized linear regression analysis method, which can obtain the weight of independent variables, and thus, understands each attribute. The logistic regression algorithm supports two-class and multi-class classification problems. For two-class classification problems, the formula is as follows (Eqs. 1 and 2).

$$P(Y = 1|X) = \frac{\exp(w \cdot x)}{1 + \exp(w \cdot x)} \quad (1)$$

$$P(Y = 0|X) = \frac{1}{1 + \exp(w \cdot x)} \quad (2)$$

$x = (x_1, x_2, \dots, x_d, 1) \in R^{d+1}$ is the feature vector inputted, d is the feature dimension of the sample, $y = \{0, 1\}$ represents the target label outputted, $w = (w_1, w_2, \dots, w_d, b) \in R^{d+1}$ is the weight vector and b is the offset.

The logistic regression algorithm estimates the coefficients of the regression equation by the maximum likelihood estimation, which can be solved by Newton iterative method and gradient descent method. The loss function is shown in Eq. 3.

$$L(w, x, y) = \log(1 + \exp(-yw^T x)) \quad (3)$$

The goal of the logistic regression algorithm is to find the optimal weight vector w according to the sample distribution of the training data. Given a new sample point x , the logistic regression is predicted by the following loss function of Eq. 4.

$$f(z) = \frac{1}{1 + e^{-z}} \quad (4)$$

Among this, $z = w^T x$. The output of the logistic regression algorithm $f(w^T x)$ has the function of probabilistic interpretation. By default, the sample is predicted to be a positive class when $f(w^T x) > 0.5$, and vice versa.

4.3 Model prediction testing

The classification ability of the model to test the data is usually measured by accuracy, precision, recall and F value which are calculated by the constructed confusion matrix, as shown in Table 2.

The accuracy refers to the proportion of the number of samples correctly predicted by the model in the overall sample, which can be calculated by Eq. 5. The precision refers to the proportion of the number of samples that is actually positive and the prediction of which is also positive in samples; it can be calculated by Eq. 6. The recall refers to the proportion of the number of samples predicted to be positive by the model in the actually positive samples, which can be calculated by Eq. 7. The accuracy and the recall are mutually independent and a single prediction criterion cannot reflect the validation of a classification model. However, the F value can make up for this defect which is the weighted harmonic average of the precision and the recall, as shown in Eq. 8.

$$Accuracy = \frac{TP + TN}{C} \quad (5)$$

Table 2 Predicted class matrix

		The predicted class	
		Positive	Negative
The actual category	Positive	TP	FN
	Negative	FP	TN

TP represents the number of samples that is actually positive and the prediction of which is also positive; TN represents the number of samples that is actually negative and the prediction of which is also negative; FP represents the number of samples that is actually negative but is predicted to be positive; FN represents the number of samples that is actually positive but predicted to be negative. So that the positive sample $P = TP + FN$, the negative sample $N = FP + TN$, and the total samples $C = TP + FN + FP + TN$

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$Recall = \frac{TP}{P} \quad (7)$$

$$F1 = \frac{2Recall \times Precision}{Recall + Precision} \quad (8)$$

4.4 Forum posts and prediction model

Posts are presented as discourse in discussion forum. Information exchange and knowledge construction are based on series of participant discussion behaviors. There is a lot of research on teaching and learning-based forum data. For example, Garrison et al. (1999) developed coding framework for learning reflection in online discussion. Babori (2021) focused on exploring learning strategies in MOOCs. Hew and Cheung (2008) compared different types of questions in discussion forum and found that questions about viewpoints were regarded as a more useful technique for encouraging participation than questions for clarification. Bradley et al. (2008) examined how question types influence the quantity and quality of the discussion and determined that a limited focal question type, followed by brainstorming, open focal (an issue with no alternatives) and direct link (such as a quotation) question types were the most influential for increasing word count and the degree of answer completion.

As these data are generated, predictive models are also studied by researchers. Predictive model refers to the quantitative relationship between variables described by mathematical language or formula and is used for prediction of their relationships. It reveals the inner regularity of variables to a certain extent and takes it as the direct basis for calculating a value in prediction. Among them, the regression prediction method is based on the correlation between independent variables and dependent variables to predict relationships among these variables. Therefore, on the basis of the above works, a learning outcome prediction model was established in the present study based on the data such as posts in MOOC discussion forum and their different types.

5 Data preparation

5.1 Data source

Currently, there is no authoritative standard data set available in the research on the mining of posts related to learning topics in MOOC forum. We collected posts on learning topics from a MOOC platform for Chinese universities; it is

a platform for online course learning with a large amount of data. For this reason, the requirement related to a statement about the ethics issues relevant to the research and approvals under which the data was collected and reported is not relevant for the present study.

The discussion area in the platform is generally composed of three sections such as classroom communication area, comprehensive discussion area and question and answer area for teachers, with each having different functions. Students enter a corresponding discussion board to ask questions according to their own needs. Other students and teachers discuss and communicate with each other according to discussion questions and get solutions to assigned problems. Of the 200 courses of this type, Modern Educational Technology course offered by the Fu Gangshan Team of Shaanxi Normal University in China has the highest number of participants. Moreover, this course has been recognized as a “Chinese national top quality” course with rich topics, wide coverage, high learner enthusiasm and good representativeness.

This course has been delivered since February 2016 seven times. Among them, the course delivered third time (third session) was the most complete, including course learning, course management, platform operation and daily communication. So in this study, the data from the third session was used as the research sample. A total of 11,184 participants participated in the third session and 3,680 discussion posts were posted. The content related to learning in the forum mainly consists of three aspects: the first is the question, the second is the knowledge, and the third is the feedback. The question includes type of questions and definition of questions, the knowledge consists of knowledge types and importance of knowledge points, and the feedback is divided into duration of questions, number of questions answered, number of views, number of votes, question–answer feedback, and whether the teacher/teaching assistant participated in the discussion. Among them, the discussion with question solving is mainly based on question–answer feedback. So, these variables are taken as independent variables and dependent variable is the discussion depth.

An application using Selenium with Python was developed in order to capture all posts in discussion forum from the Modern Educational Technology course. Collected metadata included such information as topic title, topic description, topic comment content, comment reply content, topic creation time, last comment reply time, topic reply number, topic browse number, topic vote number, topic initiator ID, and comment reply ID. All information was stored in a data sheet.

5.2 Data preprocessing

In this study, the data preprocessing method included (1) screening against irrelevant or duplicate data and eliminating it, (2) coding variables and (3) data standardization. The specific steps and rules are listed and explained below.

5.2.1 Data screening and deduplication

2,344 threads were obtained and they contained a total of 18,219 posts. Themes of the forum include course learning, course management, platform operation and daily

communication. Among them, the content related to learning is as follows: professional knowledge, test assignments, learning methods, exam review and learning experience. Therefore, only MOOC learning related threads were analyzed in this study.

Next, the threads had to be further cleaned. Anonymous threads were deleted because the variable “question–answer feedback” could not be counted; threads with no reply were also deleted; one of the threads with the same question description and content of reply was reserved; the feature variable “views” was removed because its value was similar as the reply number. Finally, 1,873 threads were reserved, including 8 feature variables.

5.2.2 Variable coding

The variable coding is described as follows:

- Question types: In a mathematics classroom research, Song and Wang (2006) classified classroom questions into: suggestive questions, open questions, guiding questions, understanding questions, evocation questions and judgment questions from the metacognitive and cognitive perspective. However, the judgment question in this question type is only a short answer to what is right or wrong and is inappropriate for the topic characteristics of MOOC discussion. Therefore, this study deleted this category. See Table 3 for detailed description of the classification of questions in this study.
- Knowledge types: (a) the categories of knowledge points (declarative knowledge=0, productive knowledge=1, strategic knowledge=2); and (b) the importance of knowledge points (important point=1, general point=0).
- Question organization: we categorized the organization of questions in MOOC forums. Based on this, questions can be divided into two categories: well-structured question and ill-structured question. Here, well-structured question=1, ill-structured question=2.
- Question duration: it is expressed by the interval between the last post time and the release time of question. Based on the actual situation, we divided the duration of questions into 3 categories: within a day=1, within a week=2, more than a week=3.
- Question–answer feedback: the person who asked the question provided feedback to students’ answers or guided other students to discuss an issue in deep after initiating a discussion. Among them, no feedback=0, one feedback=1, multiple feedbacks=2.
- Teacher participation: teacher or teaching assistant participation=1, no teacher or teaching assistant participation=2.
- Number of replies: the number of answers to each question in MOOC forums. It is a continuous variable.
- Topic discussion depth: based on the cognitive process dimensions of Bloom’s taxonomy, the contents of discussion was analyzed to confirm the discussion depth in MOOC forum. The cognitive level reflected in discussion was divided into 3 levels: shallow=1, middle=2, deep=3. Table 1 demonstrates the classifi-

Table 3 Question types (adopted from Song and Wang (2006))

Types	Value	Description	Example
Arousal	1	Recalling and confirming knowledge	What aspects is the regulation function of instructional prediction reflected in?
Comprehension	2	It requires thinking activities such as simple analysis and calculations and to summarize the content already learned	What is the difference between the prediction of informational instructional design and prediction of traditional instructional design?
Oriented	3	It requires repliers to think actively and point to the goal of problem-solving The aim is to summarize concepts and master rules	How to properly use task-driven informational teaching in the process of information-based teaching?
Open-ended	4	It requires thinking in depth from multiple perspectives	How to achieve the effective class in the information age?
Metacognitive	5	Evoke repliers to adjust his or her own thinking strategy, generate the inquiry of the question and carry out self-monitoring It acts as a guide, guiding repliers to explore the solution to questions	The multimedia environment requires new teaching methods, content organization, process organization and performance prediction. Does it mean we are being asked to improve standards?

cation of depth dimensions and Table 4 includes the description of topic discussion depth in MOOC forums.

To ensure the validity and reliability of the coding, the study involved three experts in the data coding process. The three experts were all with PhD degrees and had long been engaged in teaching and research in the field of educational technology. Their age was as follows: one was 52 and the other two were in their 40 s.

During the variable coding process, the experts worked with posts in 1873 topic categories (except the duration and issues hosted two variables). An independent coding by each expert was carried out and then coding consistency was inspected among coding results from the three experts. To this end, Kappa coefficient was used and the value of 0.791 (the level of consistency between 0.61 and 0.80 is considered as high) was obtained suggesting that high consistency of coding process was reached.

5.2.3 Data normalization

Data normalization was carried out to scale the value of a column of numerical features (assumed to be the i -th column) in the training set to a state with a mean of 0 and a variance of 1. In our study, variable “number of replies” with a large range of data was normalized, as shown in Eq. 9.

$$x_i * = \frac{x_i - \mu}{sd(x)} \quad (9)$$

5.2.4 Variables selection

There are 10 types of the discussion posts: type of questions, definition of questions, knowledge types, importance of knowledge points, duration of questions, number of questions replied, number of views, number of votes, question–answer feedbacks, and whether the teacher/teaching assistant participated in the discussion. The number of views, number of votes and number of replies were merged, and the definition of questions and type of questions were also merged to form 8 variables. The independent variables were: question type, question organization, knowledge type, important knowledge point, question duration, number of replies, teacher participation and question–answer feedback. The dependent variable was topic discussion depth.

5.2.5 Feature variables selection

After data preprocessing, a total of 1482 posts were obtained, each contained nine attributes. The first eight attributes were influencing factors for online discussion depth, i.e. input variables of prediction model, while the last attribute (discussion depth) was the output variable of the model.

Table 4 Topic discussion depth in MOOC forums

Category (primary)	Category (secondary)	Corresponding item	Description
Shallow	Remembering	Recognizing, recalling	Extract existing opinions or factual information from others to simply respond to questions and show your attitude
	Shallow trigger	Interpreting/understanding, classifying and exemplifying	Put forward doubt, description and explanation; briefly explain your point of view and the reason for it
Middle	Induction, integration	Summarizing, inferring	Integrate different resources and others' opinions or infer conclusions from existing knowledge
	Comparing, explaining Application	Comparing, explaining Implementing, executing	Compare different knowledge points and explain the phenomenon Use existing knowledge and tools to solve problems
Deep	In-depth reflection	Organizing, attributing	Evaluate self-awareness and behavior in context
	Analysis, critiquing Internalized, transfer	Checking, judging Planning, producing	Critically check and judge the question or opinions others put forward Form new knowledge construction

Table 5 Variables and key literature

<i>Variable</i>	<i>Name</i>	<i>Key literature for inclusion</i>
X1	question types	Ertmer et al., 2011; Qin, 2017; Song & Wang, 2006
X2	question organization	Spatariu et al., 2007; Zhang & Si, 2010
X3	knowledge types	Huang, 2017; Spatariu et al., 2007
X4	importance of knowledge point	Huang, 2017; Spatariu et al., 2007
X5	question duration	Nandi et al., 2012
X6	number of replies	Huang, 2017; Mazzolini & Maddison, 2007
X7	teacher participation	Mazzolini & Maddison, 2007; Nandi et al., 2012
X8	question–answer feedback	Choi et al., 2005; Mazzolini & Maddison, 2007

The selection of feature variables in this study is mainly based on the above references, see Table 5 for the details of key literature and Table 6 for the variables.

Table 7 shows the result of the correlation analysis of eight variables. There was a positive correlation between discussion depth and types of questions, number of replies, question–answer feedback, teacher participation, duration of questions and types of knowledge points. The correlation coefficients were small, indicating these eight variables were less collinear with each other. Besides, there was a negative correlation between discussion depth and X2, X4, so the two feature variables were removed and the other six variables were retained.

On this basis, the accuracy of the algorithm was tested by a five-fold cross-validation method. The data was equally divided into five parts, four of which were used as training data and the other was used as test data. The average accuracy of the results of the five experiments were used to represent the accuracy of the algorithm selection for two classification models. One model contains eight features and the other contains six features. The accuracy of these two models was calculated using Eq. 5 (mentioned earlier in the paper) and the calculation result is shown in Table 8.

6 Prediction model building

6.1 Optimal parameters

Logistic regression was also a generalized linear regression, and linear classification algorithms can be formulated as:

The objective function $f(w)$ consists of two parts from which one was the regularization term to control the complexity of the model and to prevent the model from overfitting the data, and the other part was the error loss function. In this study, the regularization parameter λ was set to 0.5, 1, 1.5, and each time the parameter selection was done 5 times. The accuracy was used to measure the effectiveness of the model. The optimal combination of parameters obtained by the experiment was {'penalty': 'l2', 'C': 1.5}.

Based on the logistic regression algorithm, six feature variables were selected when the regularization parameter was 1.5. The feature importance ordering was as follows from high to low: X1, X5, X6, X8, X7, X3. See Fig. 1.

Table 6 Description of variables

Variable	Code	Type
Independent variable		
X1 (question types)	Arousal = 1, comprehension = 2, guiding = 3, open-ended = 4, metacognitive = 5	Categorical
X2 (question organization)	Well-structured = 1, ill-structured = 2	Categorical
X3 (knowledge types)	declarative knowledge = 0, productive knowledge = 1, strategic knowledge = 2	Categorical
X4 (importance of knowledge point)	General point = 1, important point = 2	Categorical
X5 (question duration)	within a day = 1, within a week = 2, more than a week = 3	Categorical
X6 (number of replies)	the number of responses to each question	Numerical
X7 (teacher participation)	teacher or teaching assistant participation = 1, no teacher or teaching assistant participation = 2	Categorical
Dependent variable		
X8 (question–answer feedback)	no feedback = 0, a feedback = 1, multiple feedbacks = 2	Categorical
Topic discussion depth	shallow = 1, middle = 2, deep = 3	Categorical

Table 7 The co-linear relation between variables

	X1	X3	X5	X6	X7	X8
Tolerance	0.927	0.962	0.931	0.931	0.962	0.939
VIF	1.078	1.04	1.066	1.075	1.04	1.065

Table 8 The accuracy of the two classification models

Classification model	Accuracy/%
8 features	67.272
6 features	69.438

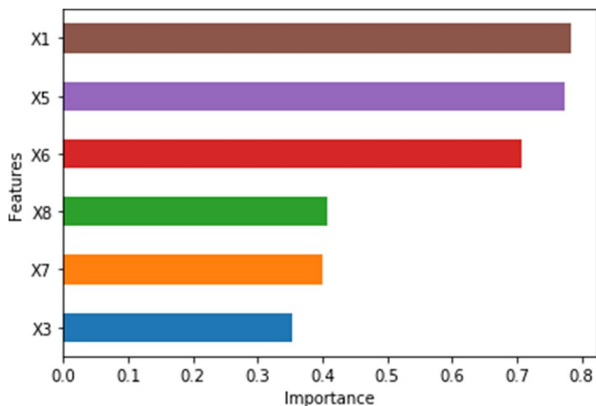
6.2 Regression models

Table 9 is a regression equation constructed with “depth of question discussion” as the dependent variable. This equation consists of 6 models.

In Model 1, the regression coefficient of “question types” was 0.435, $p < 0.001$ (strong). After adding the variable “question duration” to Model 2, the regression coefficient of “question types” decreased to 0.377 but the p value ($p < 0.001$) was still significant. This result shows that the depth of the question discussion is related to the types of question and the duration of the problem.

In Model 3, the regression coefficient of “question types” and “question duration” continued to decrease, but the regression coefficient of “number of replies” factor was 0.311 and the goodness of fit of the model increased to 38.1%. This indicates that the number of replies to questions is a significant influencing factor for the difference in the depth of question discussion. In Model 4, Model 5 and Model 6, three independent variables of “question–answer feedback,” “teacher participation” and “knowledge types” were included respectively. Finally, the regression coefficients of these three factors on the discussion depth of the question were 0.161, 0.151 and 0.072, respectively.

Although the regression coefficients of these three models on the “topic discussion depth” are lower than those of the previous three models, they still have a significant

Fig. 1 Feature importance ordering based on logistic regression

influence ($p < 0.001$), and the goodness of fit of the model increased. Overall, these six factors explained 43.8% of the difference in the depth of discussion.

6.3 Prediction formula and probability

Table 10 decomposes the discussion variable Y into different levels (i.e. deep, middle and shallow) for regression.

From Table 10, we can get the following three formulas:

$$Y_1 = -0.78X_1 - 0.35X_3 - 0.77X_5 - 0.71X_6 - 0.4X_7 - 0.41X_8 \quad (10)$$

$$Y_2 = 0.21X_1 + 0.29X_3 + 0.4X_5 + 0.27X_6 + 0.12X_7 + 0.06X_8 \quad (11)$$

$$Y_3 = 0.87X_1 - 0.08X_3 + 0.57X_5 + 0.84X_6 + 0.53X_7 + 0.4X_8 \quad (12)$$

On the basis of the logistic regression algorithm, the method to predict the probability that a discussion topic reaches the three levels of discussion depth is to select the level with the most probability as the discussion depth. Therefore, the three Eqs. (10), (11) and (12) need to be transformed by the Softmax function. It maps the function values of Y1, Y2 and Y3 to the interval (0,1), and ensure the sum of these mapped values is 1. The transformation results are respectively as follows. See Eqs. (13), (14) and (15).

$$P_{Y1} = \frac{e^{Y1}}{e^{Y1} + e^{Y2} + e^{Y3}} \quad (13)$$

$$P_{Y2} = \frac{e^{Y2}}{e^{Y1} + e^{Y2} + e^{Y3}} \quad (14)$$

$$P_{Y3} = \frac{e^{Y3}}{e^{Y1} + e^{Y2} + e^{Y3}} \quad (15)$$

Table 9 Regression models for dependent variables Y

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
X1	0.435***	0.377***	0.329***	0.309***	0.300***	0.297***
X5		0.324***	0.281***	0.271***	0.359***	0.255***
X6			0.311***	0.302***	0.282***	0.278***
X8				0.168***	0.155***	0.161***
X7					0.162***	0.151***
X3						0.072***
N	1482	1482	1482	1482	1482	1482
R ²	0.189	0.290	0.381	0.408	0.433	0.438

$p < 0.01$ ***; $p < 0.05$ **; $p < 0.1$ *

Table 10 The results of the regression model of 6 feature variables

	X1	X3	X5	X6	X7	X8
Y1	-0.78	-0.35	-0.77	-0.71	-0.4	-0.41
Y2	0.21	0.29	0.4	0.27	0.12	0.06
Y3	0.87	-0.08	0.57	0.84	0.53	0.4
<i>p</i>	0.000	0.000	0.000	0.000	0.000	0.000
<i>R</i> ²	0.663					
<i>value</i>	29.382					

Y1 represents the shallow level, Y2 represents the middle level, Y3 represents the deep level

It is possible to predict the probability that a discussion topic reaches the three levels of discussion depth, thereby evaluating the role of the topic in learning or teaching. The assessment of the discussion topics is also valuable reference in the design of teaching and learning in the MOOC platform.

6.4 Verification of prediction models

Verification of the model in machine learning is mainly measured by precision, recall and *F* value, and the calculation is shown in the above formulas (5), (6), (7), (8). The interpretation of the meaning is shown in Table 2. The test results are shown in Table 11. The accuracy of the feature combination (X1, X5) was 67%; the accuracy of the feature combination (X1, X5, X6) was 65%; the accuracy of the feature combination (X1, X5, X6, X7) was 68%. The results of the models were compared, see Table 9.

Then, the types of questions were divided into arousal, comprehension, oriented, open-ended, and metacognitive. The depth of discussion was divided into shallow, middle and deep for measurement. The results are shown in Table 12. Among them, middle level of discussion is the most effective. Comprehension and oriented are two types of problems that work better. Among them, $X^2 = 910.881, p = 0.000$.

Table 11 Verifying the results of variables in the discussion level (Y1, Y2, Y3)

level	[X1, X5]			[X1, X5, X6]			[X1, X5, X6, X7, X8]			[X1, X3, X5, X6, X7, X8]		
	P	R	F	P	R	F	P	R	F	P	R	F
Shallow	0.64	0.92	0.76	0.63	0.90	0.74	0.67	0.92	0.76	0.65	0.91	0.77
Middle	0.73	0.56	0.63	0.70	0.53	0.60	0.75	0.56	0.65	0.75	0.58	0.64
Deep	0.00	0.00	0.00	0.50	0.03	0.06	0.56	0.15	0.24	0.67	0.18	0.29
Average	0.64	0.67	0.64	0.65	0.65	0.62	0.70	0.68	0.67	0.70	0.70	0.68

P, R and F refer to precision, recall and F value, respectively. See the Eqs. 6, 7 and 8 for reference

Table 12 Verifying the question types and the discussion level (Y1, Y2, Y3)

	Question types				
	Arousal	Comprehension	Oriented	Open-ended	Metacognitive
Shallow	77 (10.8%)	482 (67.5%)	92 (12.9%)	63 (8.8%)	0 (0%)
Middle	124 (19.9%)	56 (9%)	317 (51%)	56 (9%)	69 (11.1%)
Deep	0 (0%)	26 (17.8%)	9 (6.2%)	92 (63%)	19 (13%)
x^2	910.881				
p	0.000				

7 Discussion and conclusion

In this study, we attempted to construct the automatic prediction model of the depth of discussion of learning topics. In addition, we aimed to explore the predicted variables of the depth of discussion of learning topics. Our findings showed that question types, question duration, and number of replies are the three most important variables that affect the prediction model of MOOC deep discussion. Furthermore, we found that the question–answer feedback and knowledge types also affect the prediction model. We discuss these aspects in more details in the following subsections.

7.1 Predictive models and the depth of online discussion

Big data analysis makes possible to automatically evaluate the depth of online discussion. With the help of online course learning data, this study establishes a mathematical model and predicts the depth of discussion questions. In MOOC forums, different question variables are the key to build prediction models. In this study, the important variable in the in-depth discussion is the question types. The prediction model in this study not only can predict the effective questions in the forum for MOOC builders, but also provide effective teaching help for MOOC users. Similarly, Almatrafi et al. (2018) used machine learning algorithms to construct a prediction model that can be used to identify urgent discussion posts that need to be paid attention to and responded to by teachers in MOOC discussion forums. Qin (2017) predicted the categories of discussion posts by training the SVM classification model, thus completing the emotion analysis of the course.

7.2 The question types and the depth of online discussion

The data indicate that question types have strong impacts on shallow, medium, and deep levels of discussion. According to Bloom’s taxonomy, arousal and understanding problems belong to the type of problems with low cognitive level. Therefore, most of online discussion posts were at the shallow level and the meta-cognitive question types could promote the depth of discussion better. This conclusion is consistent with the view of Nancy (2015). Scholars believed that Bloom’s taxonomy

emphasizes the necessity of learning objectives (Nancy, 2015), and problems that require high level cognitive processes will lead to deep learning and efficient problem-solving. Table 12 shows the specific ratio between the question types and the level of discussion in MOOC discussions.

According to Table 12, in MOOC discussion, teachers are suggested to assign students such questions so that students have deep discussion of their question solution or reflection on question -solving process, e.g., open questions with controversial and inconclusive exploratory content. At the same time, teachers can set up reflective discussion topics. For example, discussion of meta-cognitive questions can give online students opportunities for self-reflection and self-assessment. Students on the basis of their analytical thinking, through debate and negotiation, can find questions as useful, and their engagement in discussion can promote their critical thinking development as well as achieve deep online discussion.

7.3 The frequency and duration of discussion and the depth of online discussion

The data showed that the discussion depth score of posts with fewer than seven replies was much lower than that of posts with more than seven replies. Posts with questions lasting more than seven days scored higher for discussion depth than posts with questions lasting less than a day. The reason is that the discussion area of this course has not formed a discussion environment for student to debate and negotiate, and there was a lack of consciousness of exchange and collision between students. The lack of questioning and integration of peer interaction lead to a small number of topic responses and the depth of online discussion was at the shallow level. A post with a short discussion duration couldn't reach the level of in-depth discussion. It may be that students had not enough time to digest and understand what they had learned and had not yet assimilated it into their own knowledge system, so they lacked deep thinking about discussion topics.

In view of this situation, teachers can increase the scope of the discussion forum by encouraging more meaningful posts and discussion such as debating or negotiating around these posts among students. Teachers can take the number of students' replies in the discussion forum as an important indicator of assessment, so that students can attach importance to the discussion forum. On the other hand, teachers can build learning communities where students can communicate with each other, share ideas, brainstorm on discussion issues, and achieve common goals. This not only can ensure students' understanding and mastery of knowledge, but also promote the depth of their discussion.

7.4 The teachers' guidance and the depth of online discussion

The teacher's reply in the discussion forum of this course was mainly about the key and difficult points related to the course content, evaluation feedback, test evaluations and guidance. The leading role of teachers in MOOC affected the continuity and stability of students' participation in learning to some extent (Zhou, 2016). In some posts, the initial discussion of students could not reach the expected effect, the focus of the discussion was not clear, and the topic of the discussion was scattered, which resulted in that most students could only have a superficial understanding of

the discussion, and it was not easy to have further and deeper discussion. At this point, the teacher or teaching assistant can explain the content and key points of a response to a given question, so that students can have an in-depth discussion of a question.

Therefore, the leading role of teachers in forum discussions cannot be ignored. Teachers' positive feedback and guidance play an important role in the in-depth discussion of issues in the forum, and in the process of cultivating students' reflective ability and in-depth cognition. As Wang (2013) pointed out, the teacher guides the students to focus on discussion, uses the threading method to interweave the students' viewpoints, and connects the students' viewpoints to the learning results of the course.

7.5 The questioner's support and the depth of online discussion

The discussion depth score of the discussion post with the questioner responding more than three times was much higher than that of the discussion post with no feedback or less than three feedbacks. Questioners' timely feedback, effective guidance and recognition and praise for other students are important factors to stimulate students' active learning and deep thinking, and important support to promote students' in-depth discussion.

In such case, the course need to pay attention to process prediction. Process prediction is an activity in which students' learning information is explained instantly and dynamically in teaching activities so as to reveal, judge and generate teaching value (Liu et al., 2020). Process prediction is constructed during implementation (Goodman et al., 1996). The questioner can give a timely comment on the content of the reply, such as setting the more insightful reply as "top" and "agree." When there are different opinions in the discussion, the questioner may explain the discussion properly or hold his own position and engage in a debate. Students are also encouraged to collect and sort out the different viewpoints of the respondents, sublimate the discussion, inspire more students to think deeply about problems from different perspectives and ideas, and jointly construct knowledge. This information can be recorded in the student's electronic file, and the student will be awarded a certain number of points, which will be taken into the total score. If the questioner posts only without any feedback, the discussion post will be deemed invalid and will not receive a grade for the discussion section of the course. Therefore, teachers are suggested to give immediate feedback on the posts in the study discussion and make teaching adjustment according to the process prediction.

7.6 The types of knowledge and the depth of online discussion

The types of knowledge have certain influence on the depth of forum discussion. The score of strategic knowledge discussion post is higher than that of declarative knowledge discussion post, which indicates that strategic knowledge can promote in-depth discussion among online learners. This may be because strategic knowledge can guide students to master rules and promote students complete learning.

Therefore, in online discussion, teachers need to provide more strategic knowledge, cultivate students' reflective ability and improve students' creativity.

7.7 Value of this research

In this study, the content analysis is used to quantify, statistics is used to test variables related to MOOC online discussion data, and the machine learning classification algorithm is used to train sample data to build a discussion depth evaluation model, through which unknown discussion data can be classified and predicted. To analyze the MOOC discussion data in an objective, accurate and efficient way, and to study online learning through data model, this kind of data analysis and prediction method is one of potential research directions in online learning context. For example, Almatrafi et al. (2018) built an emergency discussion predictive model and Qin (2017) developed the model on predictive posts. Based on the perspective of deep learning and cognitive development, this study conducted in-depth analysis on the data of MOOC course discussion forum and established a prediction model, providing reference for MOOC course developers and researchers who focus on online learning.

In online learning, in-depth discussion is one of the key points of course teaching. Educators need to focus on the type of questions, duration of questions, and feedbacks to questions in the topic discussion. These are not only important characteristic-indicators for establishing mathematical models (similar indicators are also found in the research by Spatariu et al. (2007), but also important factors affecting the depth of online discussion. Gerber et al. (2005) also reached this conclusion about online discussion. The present research can help identify which problem and knowledge needs attention and guidance during the online teaching process. If the teacher is teaching online, the model can be used to infer whether teaching questions lead to deep learning, etc. Therefore, teachers need to pay more attention to the types of problems and to guide students to think deeply about problems and improve students' cognitive development. This is especially important under the influence of the epidemic when most students learn online.

7.8 Limitations and future research

The classification model trained by the feature set of this study performed well, but did not reach superiority (> 95%). The possible reason is that the selection of prediction features to measure the depth of online discussion, as well as the processing method of online discussion data had some problems. For example, non-linear features need to be converted; the manual data annotation is susceptible to the knowledge background of annotators; the model does not perform well when prediction features are missing.

The logistic regression model is one of machine learning classifications, and decision trees, support vector machine and random forests are common methods of machine learning. In future studies, appropriate machine learning algorithms will be selected to improve machine classification performance according to characteristics of the algorithm, serving for learning and teaching in MOOC.

In addition, this research only collected the data from one course on MOOC platform of Chinese universities. The research scope is too small and the number of research samples is very limited. There are not enough discussion posts available for a course, and the data set was not large enough to train a high-performance model. So, the next step will be to integrate discussion posts from more MOOCs, increase the data set for training, and train models with higher performance. In subsequent studies, models can be constructed by using gradient ascending tree, neural network, Naive Bayes and other methods to further compare their performance, and relevant prediction standards will be established to provide targeted services for learning and teaching in MOOC.

With the emergence and development of big data and analytics techniques, more results will be predicted in future studies based on the data. Machine learning produces models that predict learning outcomes based on data generated in MOOCs. In this study, the learning results of MOOC were closely related to the types of learning questions, with 0.87 types of questions leading to about one in-depth discussion. Therefore, in MOOCs discussion, we firstly need to pay attention to types of questions. The prediction models will assist teaching and learning more efficiently in the future.

Data availability The datasets generated and analyzed during the current study are not publicly available but will be provided by the corresponding author on reasonable request.

Declarations

Conflict of interest None

References

- Almatrafi, O., Johri, A., & Rangwala, H. (2018). Needle in a haystack: Identifying learner posts that require urgent response in MOOC discussion forums. *Computers & Education, 118*, 1–9.
- AlQaidoom, H., & Shah, A. (2020). The role of MOOC in higher education during Coronavirus pandemic: A systematic review. *International Journal of English and Education, 9*(4), 141–151.
- Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
- Babori, A. (2021). Analysis of discussion forums of a programming MOOC. *TEM Journal, 10*(3), 1442–1446. <https://doi.org/10.18421/TEM103-56>
- Bradley, M. E., Thom, L. R., Hayes, J., & Hay, C. (2008). Ask and you will receive: How question type influences quantity and quality of online discussions. *British Journal of Educational Technology, 39*(5), 888–900. <https://doi.org/10.1111/j.1467-8535.2007.00804.x>
- Choi, I., Land, S. M., & Turgeon, A. J. (2005). Scaffolding peer-questioning strategies to facilitate meta-cognition during online small group discussion. *Instructional Science, 33*(5–6), 483–511.
- Dash, M., & Liu, H. (1997). Feature selection for classification. *Intelligent Data Analysis, 1*(3), 131–156. <https://doi.org/10.3233/IDA-1997-1302>
- Durairaj, K., & Umar, I. N. (2015). Analysis of students' listening behavior patterns in an asynchronous discussion forum. *Procedia-Social and Behavioral Sciences, 176*, 27–34. <https://doi.org/10.1016/j.sbspro.2015.01.440>
- Duruk, U., Akgün, A., & Güngörmez, H. G. (2021). Exploring the impact of common knowledge construction model on students' understandings of heat transfer: Exploring the impact of common knowledge. *International Journal of Curriculum and Instruction, 13*(1), 114–136.

- Ertmer, P. A., Sadaf, A., & Ertmer, D. J. (2011). Student-content interactions in online courses: The role of question prompts in facilitating higher-level engagement with course content. *Journal of Computing in Higher Education*, 23(2–3), 157–186.
- Fauvel, S., & Yu, H. (2016). A survey on artificial intelligence and data mining for MOOCs. <http://arxiv.org/abs/1601.06862>. Accessed Dec 2020.
- Garrison, D. R. (1991). Critical thinking and adult education: A conceptual model for developing critical thinking in adult learners. *International Journal of Lifelong Education*, 10(4), 287–303. <https://doi.org/10.1080/0260137910100403>
- Garrison, D. R., Anderson, T., & Archer, W. (1999). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education*, 2(2–3), 87–105. [https://doi.org/10.1016/S1096-7516\(00\)00016-6](https://doi.org/10.1016/S1096-7516(00)00016-6)
- Gerber, S., Scott, L., Clements, D. H., et al. (2005). Instructor influence on reasoned argument in discussion boards. *Educational Technology Research and Development*, 53(2), 25–39.
- Goodman, R. M., Wandersman, A., Chinman, M., Imm, P., Morrissey, E. (1996). An ecological assessment of community-based interventions for prevention and health promotion: Approaches to measuring community coalitions. *American Journal of Community Psychology*, 24, 33–61
- Guo, C., Shea, P., & Chen, X. (2021). Investigation on graduate students' social presence and social knowledge construction in two online discussion settings. *Education and Information Technologies*, 1–19. <https://doi.org/10.1007/s10639-021-10716-8>
- He, M. (2015). *The mobile data security detection technology research based on machine learning*. Unpublished master thesis, Beijing University of Posts and Telecommunications, Beijing, China. (in Chinese)
- Hew, K. F., & Cheung, W. S. (2008). Attracting student participation in asynchronous online discussions: A case study of peer facilitation. *Computers & Education*, 51(3), 1111–1124. <https://doi.org/10.1016/j.compedu.2007.11.002>
- Hou, M. L. (2016). *Data mining and application in MOOC discussion forum[D]*. Unpublished master thesis, Shandong University, Jinan, China. (in Chinese).
- Huang, Q. (2017). *Research on Depth and Influencing Factors of Online Discussion in Teacher Workshops [D]*. Unpublished master thesis, East China Normal University, Shanghai, China. (in Chinese).
- Hwang, W. Y., Li, Y. H., & Shadiev, R. (2018). Exploring effects of discussion on visual attention, learning performance, and perceptions of students learning with STR-support. *Computers & Education*, 116, 225–236.
- Liu, D. F., Sun, D. J., Zhu, Q. Y., Yu, R. R., & Shao, J. T. (2020). Research on mixed teaching model under the guidance of process prediction theory ——A case study of Huainan Normal University. *Journal of Huainan Normal University*, 05, 107–110. (in Chinese).
- Mays, T. J., Oganage, B., Naidu, S., & Perris, K. (2021). Supporting teachers moving online, using a MOOC, during the COVID-19 pandemic. *Journal of Learning for Development*, 8(1), 27–41.
- Mazzolini, M., & Maddison, S. (2007). When to jump in: The role of the instructor in online discussion forums. *Computer & Education*, 49(2), 193–213.
- Moon, A. L., Francom, G. M., & Wold, C. M. (2021). Learning from versus learning with technology: Supporting constructionist reading comprehension learning with iPad applications. *TechTrends*, 65(1), 79–89. <https://doi.org/10.1007/s11528-020-00532-1>
- Nancy, E. A. (2015). Bloom's taxonomy of cognitive learning objectives. *Journal of the Medical Library Association*, 103(3), 152–153. <https://doi.org/10.3163/1536-5050.103.3.010>
- Nandi, D., Hamilton, M., Chang, S., et al. (2012). Evaluating quality in online asynchronous interactions between students and discussion facilitators. *Australasian Journal of Educational Technology*, 28(4), 684–702.
- Qin, C. B. (2017). *A data mining on emotion analysis and knowledge difficulty in Chinese MOOC forum*. Unpublished master thesis, Beijing University of Posts and Telecommunications, Beijing, China. (in Chinese)
- Song, X. P., & Wang, J. H. (2006). Studying learning dynamic and the question of instruction in mathematics classroom. *Journal of Mathematics Education*, 15(3), 13–23. (in Chinese).
- Spatariu, A., Quinn, L. F., & Hartley, K. (2007). A review of research on factors that impact aspects of online discussions quality. *TechTrends*, 51(3), 44–48.
- Tan, K. E. (2017). Using online discussion forums to support learning of paraphrasing. *British Journal of Educational Technology*, 48(6), 1239–1249. <https://doi.org/10.1111/bjet.12491>
- Tokur, F., & Akgün, A. (2021). A new environmental education approach: Environmental emotion-enhanced activities. *Innovative Approaches in Science Education*, 5, 77–94.

- Vygotsky, L. S. (1987). Thinking and speech. In R. W. Rieber & A. S. Carton (Eds.), *The collected works of L.S. Vygotsky, Volume 1: Problems of general psychology* (pp. 39–285). Plenum Press.
- Wang, M. J. (2013). *Research on evaluation of teaching interactive depth in Courses BBS*. Unpublished master thesis, Shaanxi Normal University, Xi'an, China. (in Chinese).
- Warsah, I., Morganna, R., & Uyun, M. (2021). The impact of collaborative learning on learners' critical thinking skills. *International Journal of Instruction, 14*(2), 443–460. <https://doi.org/10.29333/iji.2021.14225a>
- Zhang, M., & Si, Z. (2010). A case study of e-learning from the perspective of problem behavior - taking "Distance Education and E-learning Practice" as an example. *E-Education Research, 01*, 74–78.
- Zhou, L. (2016). The multiple action research on teaching mode based on the concept of MOOC: A case study of modern educational technology. *Journal of Shaoguan University, 37*(11), 126–131. (in Chinese).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.