Research on the Effect of Online Service by Logistic Regression Model and Remote Video Technology

Hanping Zhang*
Computer Department, Wuhan Polytechnic, Wuhan 430074, China

*Corresponding author: hanpingzhang@whpu.edu.cn

Abstract. At present, online teaching is developing rapidly, and with it is how to improve the effectiveness of online teaching. This research constructs learning behavior indicators from the online learning process, and uses logistic regression to analyze the learning data in online teaching, and explores the impact of learners' online learning behavior on academic performance. The correlation between indicators such as the time lag of course registration, the number of login courses, the number of submitted homework tests, the average of the number of saved exercises, and the degree of completion of video viewing and other indicators and scores were discussed. The study found that the submission of homework tests can be used as key indicators for online teaching and learning performance prediction.

Keywords: online teaching; logistic regression; learning behavior; learning effect.

1. Introduction
Since teaching activities, teaching effectiveness has always been an eternal topic for educators, and teachers in the frontline of teaching have been committed to exploring ways and methods to achieve effective teaching effects. Online teaching is a high degree of integration of information technology and education. The teaching environment and teaching methods have undergone tremendous changes, and its teaching effectiveness is bound to give new meanings. At present, online teaching is generally faced with problems such as unsystematic teaching design, overloading of teaching resources, and fragmentation of information. The graduation rate of online academic education is significantly lower than that of traditional education. Large-scale learner registration seems to mean a large-scale dropout rate and failure rate. Improving the learning effect of online teaching learners is a major problem facing online teaching. Through empirical research on the relationship between online learning behavior and learning effect in different environments, it is found that learner's online behavior has an important influence on learning effect. In the existing research on the relationship between online behavior and learning effect, the acquisition of online behavior characteristics is mostly based on certain dimensions of the learner's learning process, such as the learner’s registration time, the degree of teacher-student interaction, the number of exercises, and homework errors. Rate, academic performance, etc. Regarding the research on the relationship between learning behavior and learning effect, due to the complexity of the online learning process, relevant researches have reached different conclusions.
The learning effect of online learners is affected by many factors. In order to explore the influencing factors of learning effect in a comprehensive and in-depth manner, researchers need to mine and analyze a large amount of data reflecting the overall learning process. Starting from the overall learning process of online learners, this research constructs an analysis framework for online learners’ learning behavior. By analyzing the behavior data in one of the online teaching courses "Advanced Mathematics", the logistic regression method is used to analyze the impact of online learning effectiveness.

2. Research on Logistic Regression of Online Learning

2.1. Logistic regression model
Logistic regression model is one of the common methods of data mining and machine learning, and it is a supervised learning method. It analyzes and predicts linear regression of discrete dependent variables based on one or more continuous or discrete independent variables. The dependent variables of logistic regression are usually categorical discrete variables. Binary logistic regression is one of the most commonly used forms, and its dependent variable contains only two categorical values. In online learning analysis, we often encounter some discrete variables that represent the status of research objects, such as whether learners can achieve good learning results after participating in courses, obtain corresponding certificates, and whether they can pass exams and receive corresponding credits. In the online learning environment, the learning behavior of learners is reflected in all aspects of online learning. Binary logistic regression can be used to analyze the impact of learning behavior on learning effects.

2.2. Research on Logistic Regression of Online Learning

2.2.1. Variable selection. Variable selection is the first step of logistic regression. Variables must meet the two conditions of close correlation between the independent variable and the dependent variable and the independence between the independent variables. In order to ensure the reasonable and effective selection of variables, it is necessary to conduct a complete preprocessing and correlation analysis of the variables. When extracting variables from the original data set, the indicators need to be pre-processed, that is, the missing values and outliers of the variables are processed, and the data that does not meet the requirements are eliminated. In order to effectively establish a logistic regression model, it is necessary to carry out correlation analysis on variables, and as far as possible, significantly related independent variables should be selected into the modeling process.

2.2.2. Logistic regression modeling. After determining the variables that enter the logistic regression, the data samples need to be randomly divided into a training set and a validation set according to a certain proportion. The training set is used for training parameters in each experiment, and the validation set is used to verify the prediction accuracy. Logistic regression modeling can be divided into forward introduction method, backward elimination method and stepwise regression method. The three methods have their own advantages and disadvantages.

In the study of logistic regression in online teaching, analyzing the model from the perspective of pedagogy is the key link. Model calculation is an objective process, but its preliminary results may not have a reasonable pedagogical explanation. When there is a difference between the existing conclusions of the model and the actual teaching, it is necessary to repeatedly analyze and verify the data and the modeling process, and even introduce other data analysis and mining methods to assist in the analysis, in order to determine the final conclusion.

2.2.3. Model application effect evaluation. The established logistic regression model can be tested and evaluated through a series of indicators. Common evaluation indicators include correct rate, error rate, sensitivity and special effects, etc., as well as some comprehensive judgment indicators, such as ROC
curve, KS value and Lift value. The most intuitive and effective evaluation index is the prediction accuracy of the model. In addition, the ROC curve uses the area under the curve to characterize the accuracy of the model. The larger the area, the higher the accuracy of the model.

3. Online learning behavior index determination
In order to analyze the relationship between learning behavior and learning effect in online teaching, it is necessary to analyze the learning process of online teaching. In existing research, learning behaviors are mostly based on the basic process of "registration-attending lectures-classroom assessment-homework-discussion-examination-completion-certificate". Jia Jiyou and others analyzed the online behavior data of learners in online teaching on the Coursera platform, and concluded that the results are significantly negatively correlated with the start time of learning, and are related to online time, the number of videos watched, the number of webpages viewed, and the browsing time. It is significantly positively correlated with the six indicators of the number of handouts downloaded, the sum of the usual test scores, and the level of forum participation. Wang Ping conducts research on learners based on the open data of the edX platform, and selects the time to register for the course, the last login time, the number of course interactions, the number of course visits, the number of videos played, the number of learning chapters, and the number of forum posts to explore the learning behaviors of Chinese and foreign MOOCs learners And characteristics, the study found that the learners who obtained the certificate generally browsed more course chapters, but in terms of video viewing, the learners with higher scores did not significantly increase the behavior of video visits.

On the basis of existing research, combined with the online teaching practice in recent years, based on the online learning process, the learner's online learning process is summarized into the stages of pre-school preparation, logging on to the platform, resource learning, exchanges and discussions, and homework assessment, and based on the data The situation has constructed 18 MOOCs learning behavior indicators, and the specific indicators are shown in Table 1.

| Table 1. Online learning behavior indicators |
|----------------------------------------------|
| **Online learning process** | **First-level index** | **Secondary indicators** |
| 1. Preparation and login | Pre-school preparation | Number of visits to the course details page |
| | Attendance | Registration course delay |
| | Number of visits to the course details page | Number of login courses |
| 2. Resource learning | Learning from resources other than videos | Access to resources other than videos |
| | Video resource learning situation | Time to watch the video per login |
| | Video resource learning situation | Watch video completion |
| | Video resources insist on learning | Video viewing density |
| | Video resources insist on learning | Repeated video views |
| | Video resources insist on learning | Video repeat viewing level |
| | Video resources insist on learning | The number of times the video was watched after assignment |
| 3. Forum interaction | Interactive participation | Forum posts |
| | Forum replies | Forum views |
| 4. Homework | Amount of learning tasks completed | Number of submissions |
| | Enthusiasm for task completion | Average number of saved exercises |
| | | Time difference between submission and release |
| | | Assignment submission density |
| | | Time interval for submitting assignments |
4. Model establishment and analysis
This study selects the learning behavior data of a course of "Advanced Mathematics" on the vocational education cloud platform, and uses the logistic regression method to analyze the learners' academic performance. The course starts on September 1, 2020 and ends on January 30, 2021. A total of 256 people participated in the study during the course. There are 45 instructional videos in the course, and the course evaluation adopts chapter assignments, tests and final exams. There are a total of 20 after-school unit assignment tests and a final exam test. In this case, the learner’s course performance is used to characterize the learning effect of the learner, the qualification of the grade is used as the dependent variable, and the 18 indicators of online learning behavior are used as independent variables, and the learning effect of the learner is predicted according to the logistic regression research method in the online learning field.

4.1. Variable determination

4.1.1. Data processing. The 18 preset learning behavior indicators in this study involve multiple types of data such as frequency, time interval, duration, mean value, and ratio. Most of the measurement indicators need to be obtained by calculating the original data of the relevant data sheet through an algorithm. First, compile the algorithm to obtain the index data according to the preset index meaning and the original database table structure, and then compile the SQL function according to the algorithm, calculate the original data, and obtain the data value of the learner sample on the 18 indexes. The indicator data is first processed for missing values and outliers, and missing samples and indicator variables with more missing data are eliminated.

After analysis, it is found that the content of forum posts are all comments and consultations about exams, and has nothing to do with the learning effect of learners. Therefore, the three indicators of forum interaction, the number of forum posts, the number of forum replies, and the number of forum views are excluded. In the case course, the number of visits to other learning resources of the course, the number of repeated views of the video, the degree of repeated viewing of the video, the number of times the corresponding video resource is viewed after the submission of the assignment test, and the time interval of the submission of the assignment test are the missing values and zero values of these indicators. A higher proportion. Therefore, the above-mentioned index variables are eliminated, and the remaining index variables enter the correlation analysis step.

4.1.2. Correlation analysis of learning behavior indicators. Use SPSS Statistics 19 to perform Pearson correlation analysis on the remaining 10 indicators after preprocessing and learner scores. The results show that there is no significant correlation between the number of browsed course details pages and the video viewing density of the two indicators and academic performance; every time you log in to watch the video Although time length and assignment test submission density are significantly correlated with academic performance at the 0.01 level, the correlation coefficients are all <0.2, which is basically irrelevant to academic performance; among the remaining six indicators, the correlation coefficients show that the number of assignments submitted And the mean value of the time difference between the submission and release of the exercises and the mean of the number of saved exercises is significantly correlated at the level of 0.01, and the correlation is greater than 0.6, and the correlation between the number of submissions and the score is greater than 0.9. This research first screened 6 indicators, namely, course registration time lag, number of courses logged in, the mean value of the time difference between the time to submit exercises and the release time, the mean value of the number of exercises saved and the completion of video viewing, and the number of submitted test assignments into the logistic regression modeling process. Based on the collinearity relationship, the course registration time lag, the average value of the time difference between the time of submitting exercises and the release time, the number of times of saving exercises, the number of times of submitting test assignments are eliminated, and finally the two indicators of completion
degree of video viewing and the number of times of logging in to the course are retained to enter the logistic regression modeling process.

4.2. Build a logistic regression model

This study uses the binary logistic regression model in logistic regression analysis to explore the probability of occurrence of learners' qualifications in online teaching. Suppose P is the probability of the learner's learning to pass, and its value range is [0, 1], (1-P) is the probability of unqualified. P/(1-P) for learning

The occurrence ratio of qualified logistic regression is taken as the natural logarithm ln [P/(1-P)].

Assuming that the independent variables are X1, X2,..., Xk, and the dependent variable is P, the logistic linear regression function equation can be expressed as:

\[
\ln\left[\frac{P}{1-P}\right] = B_0 + B_1 X_1 + B_2 X_2 + \cdots + B_k X_k
\]

(1)

Bi (i=0,1,2,…,k) in the formula is the logistic regression coefficient.

According to formula (1), the probability of the learner's qualification is:

\[
P = \frac{\exp (B_0 + B_1 X_1 + B_2 X_2 + \cdots + B_k X_k)}{1 + \exp (B_0 + B_1 X_1 + B_2 X_2 + \cdots + B_k X_k)}
\]

(2)

Formula (2) can be used to calculate the probability of online learners passing.

In this case, the learner’s performance is converted into a qualified (score>=60) and unqualified (score<60) binary dependent variable, and the video viewing completion degree (R) and the number of login courses (L) are used as independent variables. The results of logistic regression are as follows:

The logistic regression linear equation model is:

\[
\ln\left[\frac{P}{1-P}\right] = 0.185L + 10.135R - 7.569
\]

(3)

The equation model of the probability of occurrence of a learner's academic achievement is as follows:

\[
P = \frac{\exp (0.185L + 10.135R - 7.569)}{1 + \exp (0.185L + 10.135R - 7.569)}
\]

(4)

5. Conclusion

Through case studies, on the one hand, we found the correlation between online learning behavior and academic performance, on the other hand, it also verified the practical value of logistic regression in distance education. The study found that a number of learning behavior indicators are significantly related to learning effects, including course registration time lag, the number of courses logged in, the average of the time difference between the submission of exercises and the release time, the average of the number of saved exercises, the completion of video viewing, the number of submissions of test assignments, etc. The 6 index variables that are significantly related to academic performance are distributed in the three dimensions of online learning preparation and login, resource learning, and homework testing, indicating the close relationship between online learners' academic performance and online learning behavior. Among them, the index that has the highest correlation with learners' performance is the number of homework tests submitted. This phenomenon reflects on the one hand the effectiveness of homework in evaluation, and students who put more energy into homework will get better grades; on the other hand, students will get better results. On the one hand, it may be related to the current online learning evaluation mechanism, and the weight of homework is relatively large.

In this study, the method of logistic regression was used to predict the learning effect of online learners, and good prediction results were achieved. The prediction accuracy rate of the prediction model is above 93%, and the area under the ROC curve of the model application is higher than 0.9, indicating that the application of logistic regression method to predict the learning effect based on the
learning behavior has important practical value. The learning behavior can be analyzed through the logistic regression method. And can effectively predict the effect of online learning.

Acknowledgements
Hubei Province Education Planning 2019 Project "Research on the Effectiveness of Online Teaching Model Based on Big Data Analysis" (2019GB149)

References
[1] Sun Honghui. The construction and application of online learning process management big data [J]. China Distance Education, 2016 (11): 17-21.
[2] Jia Jiyou, Miao Jingmin, Wang Qiong. Big data analysis of MOOC learning behaviors and effects—Taking 6 MOOCs of Peking University as an example [J]. Industry and Information Education, 2014, (09): 23-29.
[3] Jiang Zhuoxuan, Zhang Yan, Li Xiaoming. Analysis and prediction of learning behavior based on MOOC data [J]. Computer Research and Development, 2015, (03): 614-628.
[4] Bai Xueimei, Ma Hongliang, Wu Haimei. An Empirical Study on the Effectiveness of Online Courses [J]. Modern Educational Technology, 2017 (2): 17-21.
[5] Jia Jiyou, Miao Jingmin, Wang Qiong. Big data analysis of MOOC learning behaviors and effects—Taking 6 MOOCs of Peking University as an example [J]. Industry and Information Education, 2014, (09): 23-29.
[6] Wang Ping. The new development and application of large-scale online open courses: from cMOOC to xMOOC [J]. Modern Distance Education Research, 2013, (03): 13-19.