Applied Science and Innovative Research
ISSN 2474-4972 (Print) ISSN 2474-4980 (Online)
Vol. 7, No. 1, 2023
www.scholink.org/ojs/index.php/asir

Original Paper

The Influence of Online Learning Behavior on Learning Performance

Huiming Li[1]

[1] University of Science and Technology Liaoning, Anshan 114051, China

Received: February 10, 2023 Accepted: February 21, 2023 Online Published: February 24, 2023
doi:10.22158/asir.v7n1p69 URL: http://doi.org/10.22158/asir.v7n1p69

Abstract

Online education is a significant part of information education. It is an effective way to uncover online learning mechanisms and improve the quality of online teaching by exploiting the behavioral data of online learning platforms for learning performance prediction and analysis. In this paper, we focus on the learner's learning behavior in an online teaching scenario and explore the predictive effectiveness and impact mechanism of each behavioral feature by building a predictive model based on a machine learning algorithm. Experimental results show that three behavioral characteristics, namely the number of visits to course materials, lecture review time, and assignment, intensively influence learning performance. By comparing various machine learning algorithms, it is found that the random forest algorithm has better prediction results.

Keywords

online teaching, learning behavior, learning performance, machine learning

1. Introduction

The development of information technology and Internet technology has promoted the transformation of traditional education to information education. Online teaching methods such as Massive Open Online Courses (MOOC) (Ahmad et al., 2022; Mei et al., 2019), flipped classroom (Chen et al., 2022; Khandakar et al., 2022), personalized learning (Cevikbas, 2022), and blending learning (Ashraf, 2023) have received extensive attention from educators, and have appeared in the classroom of colleges and universities as an auxiliary method of traditional face-to-face teaching. The outbreak of the COVID-19 pandemic in 2020 forced universities to adopt online teaching on a large scale, which led to the rapid expansion and extension of online teaching. Unlike traditional classroom teaching, online teaching is an autonomous learning process with the Internet and information technology as the media and learners as the center. In the post-epidemic era, online teaching is bound to become the main direction of teaching reform in colleges and universities. Learning performance is an important indicator to measure the quality of online learning, so it is of impressive practical significance to carry out learner-centered behavior analysis and effect prediction for building high-quality offline teaching.

Several scholars have used questionnaires to explore the correlation between online learning behavior and learning performance. Mei et al. (2019) analyzed 464 valid questionnaires to explore the relationship between online learning perception and learning performance, and found that learners' recognition of the learning platform has a significant impact on learning performance. Wut et al. (2022) found that students' willingness to participate in online teaching is conducive to improving learning performance through questionnaire survey. The questionnaire analysis method has some drawbacks, such as slow data recovery, and the analysis results are mostly constrained by the subjective awareness of the surveyors. With advances in artificial intelligence, some scholars have attempted to mine online data using AI techniques to discover online learning behaviors that affect learning outcomes and the extent to which these behaviors affect learning outcomes. Ni et al. (2022) used online behavioral data to build a learning performance prediction model by machine learning algorithm, and the results showed that document learning duration was deeply correlated with learning performance. Liu et al. (2022) explored the correlation between students' clicking behavior and academic achievement in the online learning environment, and found that taking weekly clicks as the prediction model parameter was more accurate than monthly clicks. Moreover, clicks on the home page, sub-page, content, and test have a larger impact on the prediction results. Gaftandzhieva et al. (2022) applied statistics and machine learning technology to predict students' final grades according to their online learning behaviors, and pointed out that learning resources (lectures, exercises and source code), activities (homework and lectures) and attendance played a decisive role in the final grades. Dooley et al. (2020) used the data of the learning management system to explore students' learning behaviors in flipped classroom, including learning frequency, learning scope and learning time, analyzed the relationship between these behaviors and academic performance, and found that preview of learning materials in advance is conducive to improving academic performance.

This study takes the Java framework technology course as an example to study the learning behavior of students in online teaching, and explores the correlation between learning behavior and final performance through machine learning algorithms, providing reference suggestions for improving learning performance in the online education environments.

2. Materials and Methods

2.1 Data Source

The research data were derived from a "Java Framework Technology" course offered on the online teaching platform in the fall of the 2022 academic year, which was mainly aimed at third-year students of computer major, with 76 students. The duration of the course is 15 weeks, with one online lecture and one programming exercise per week. Students upload the source code and design report in the form of homework after completing the exercise, and the score is determined by the completion quality and submission time. 7 online behavior features are selected for preliminary analysis, the definition, description and range of variables are shown in Table 1. The final test scores of students were used as the learning performance, and were divided into three categories for classification prediction according to the scores. The score below 70 was considered as the low level and marked as 1; 70 to 84 were classified as medium level, labeled as 2; a score above 85 is considered high and is marked 3.

Table 1. Behavioral Variables of Online Learning

Variable	Description	Range
task	Task point completion percentage, watch live lectures and reach the required length of time, marked as 1 otherwise marked as 0, submitted on time assignment marked as 1 otherwise marked as 0	[0,1]
count	Number of visits to course materials	[0-500]
avg_view	The average duration of watching live lectures	[0-90]
avg_review	The average duration of reviewing lectures	[0-90]
total_view	Sum of avg_view and avg_review	[0-500]
assignment	The score of the assignment	[0-100]
attendance	The score of the attendance	[0-100]

2.2 Data Analysis and Preprocessing

Spearman's rank correlation coefficient is an important indicator to measure the statistical dependence of data. With statistical data X and Y, spearman's rank correlation coefficient can be calculated by the following formula:

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \tag{1}$$

where, r_s is spearman's rank correlation coefficient, with a value range of [-1,1]; $d_i=X_i - Y_i$, and n is the number of samples. In order to explore the relationship between behavior variables and learning performance, we calculate the spearman's rank correlation coefficient between behavior variables and learning performance, and use the heat map to visualize the calculation results. The visualization results are shown in Figure 1. It can be seen that in addition to the negative correlation between attendance and score, the rest are positively correlated with academic performance, among which count, avg_review, total_view, and assignment are weakly correlated with assessment results, while task and avg_view are very weakly correlated. In the online learning environment, students' active participation has a greater relationship with academic performance, and students may passively ensure attendance based on class roll call, higher normal performance, etc., but has a lower learning efficiency, resulting in inconsistency between attendance and test results. Through correlation analysis, we remove attendance and use the remaining variables as parameters of the predictive model.

Boxplots are used to visualize the distribution of behavioral variables and outliers for different scores, and the visualization results are shown in Figure 2. It can be seen that the four behavioral data of avg_view, avg_review, total_view, and assignment have outliers and need to be processed, we used the sample mean to replace these outliers. In addition, we also observe that the mean of count, avg_review, total_view, and assignment increases linearly with the increase of score, which is consistent with the results of correlation analysis.

Finally, in order to eliminate the influence caused by the difference in magnitude between behavioral features, we use the Z-Score method to normalize the datasets. The normalization method is shown in Equation (2), where x_j is the input value of sample j, μ is the sample mean, and σ is the standard deviation of the sample data.

$$z_j = \frac{x_j - \mu}{\sigma} \tag{2}$$

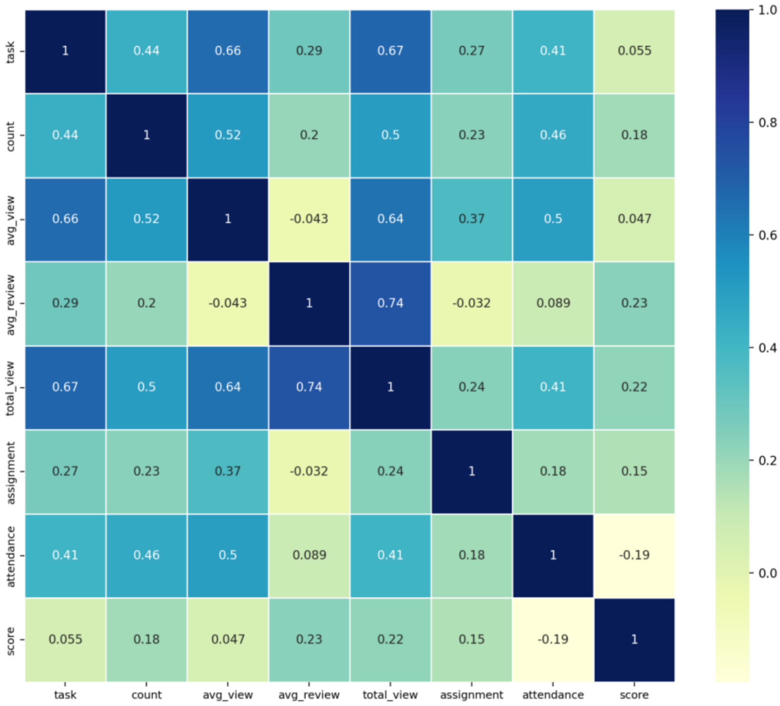

Figure 1. Correlation between Behavioral Features and Learning Performance

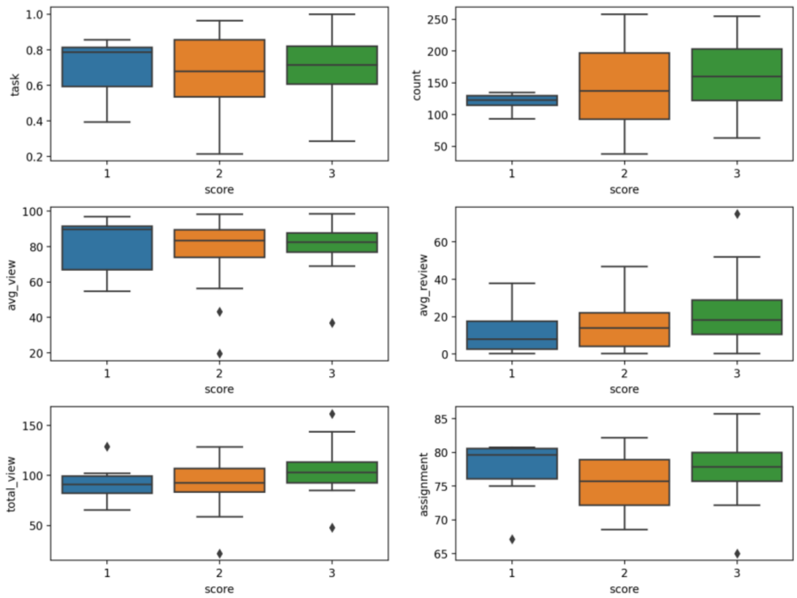

Figure 2. Distribution of Behavioral Features of Different Scores

2.3 Model Constructing

Six machine learning algorithms, including Linear Regression (LR), Bayesian Ridge (BR), Linear SVR (LSVR), Polynomial SVR (PSVR), Decision Tree (DT), and Random Forest (RF), were used to construct learning performance prediction models based on online learning behavior. The performance of the model is evaluated using the metrics Accuracy A, Precision P, Recall R, and F1 score, which are calculated as follows:

$$A = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

$$P = \frac{TP}{TP+FP} \tag{4}$$

$$R = \frac{TP}{TP+FN} \tag{5}$$

$$F_1 = \frac{2\times P\times R}{P+R} \tag{6}$$

where, TP indicates that the instance is positive and the actual prediction is positive; TN indicates that the instance is negative and the actual prediction is negative; FP indicates that the instance is negative and the actual prediction is positive; FN indicates that the instance is positive and the actual prediction is negative. Formula (2-4) is for the binary classification problem. In the multi-classification scenario, when considering a category, the remaining are considered as negative categories. Therefore, the accuracy rate and recall rate of each category can be obtained, and then the average value of the indicator values of multiple categories can be taken.

3. Result

Based on the correlation analysis results combined with the feature distribution of behavioral variables, two sets of feature sets, feature set A (task, count, avg_view, avg_review, total_view and assignment) and feature set B (count, avg_review, total_view and assignment), are used to predict learning performance, and finally the appropriate set of behavioral variables is determined according to the prediction results. The prediction results are shown in Table 2.

Table 2. Prediction Results of Different Algorithms on Feature Set A and Feature Set B

Algorithm	Feature Set A				Feature Set B			
	A	P	R	F1	A	P	R	F1
Linear Regression	0.67	0.71	0.76	0.73	0.68	0.70	0.75	0.72
Bayesian Ridge	0.66	0.68	0.91	0.78	0.65	0.66	0.92	0.77
Linear SVR	0.68	0.70	0.84	0.77	0.66	0.67	0.88	0.76
Poly SVR	0.62	0.67	0.81	0.74	0.66	0.69	0.87	0.76
Decision Tree	0.61	0.65	0.56	0.60	0.64	0.77	0.64	0.70
Random Forest	0.67	0.68	0.62	0.65	0.74	0.78	0.74	0.76

It can be seen from Table 2 that the prediction effect of feature set B is generally better than feature set A, that is, the prediction effect is better with fewer features, indicating the effectiveness of feature selection. Compared with feature set A, the two learning behaviors of task completion and duration of participation in real-time lecture in feature set B are excluded. In the case of online teaching, because it is impossible to effectively supervise students' learning attitudes, although students participate in course lectures and complete the prescribed tasks, the effectiveness of learning is low, resulting in inconsistent behavior and result performance. The accuracy and F1 score of different algorithms on feature set B are shown in Figure 3, and it can be seen that the random forest algorithm outperforms the other algorithms in both metrics.

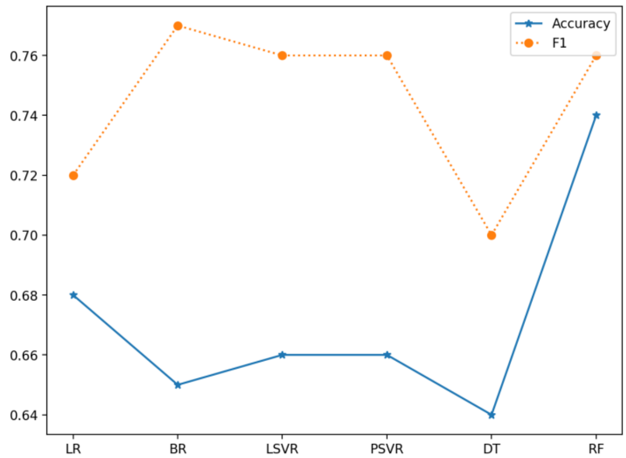

Figure 3. The Accuracy and F1Scores of Different Algorithms on Feature Set B

4. Conclusion

This paper focuses on the problem of learning performance prediction based on learning behavior characteristics in online teaching scenarios. Its core is how to mine the feature set with instructional significance and high predictive power from online data. The effective learning behavior data is extracted based on the basic characteristic information such as learning times, lecture participation and review duration, and homework completion. After data correlation analysis, outlier processing, data normalization, model building, model training, and other processes, the behavior characteristics and algorithms that can effectively evaluate online learning are finally determined. Although this study has achieved good experimental results, the data used for evaluation has problems such as small data volume and few effective features. In the future, we will further explore the generation of learning behavior metrics in order to provide better and more complete suggestions for the construction of the platform and the improvement of the quality of online learning.

References

Ahmad, I., Sharma, S., Singh, R., Gehlot, A., Priyadarshi, N., & Twala, B. (2022). MOOC 5.0: A Roadmap to the Future of Learning. *Sustainability*, *14*(18), 11199.

Mei, H., Wang, J., Zhang, J., & Zhongfeng, X. U. (2019). The influence of MOOC learing percepation on learning performance. *Journal of Beijing University of Technology (social sciences edition)*, *5*, 12.

Chen, J., Li, F., Zhang, X., & Li, B. (2022). An Assistant System for Translation Flipped Classroom. *Applied Sciences*, *13*(1), 327.

Khandakar, A., Chowdhury, M. E., Khalid, M. S., & Zorba, N. (2022). Case Study of Multi-Course Project-Based Learning and Online Assessment in Electrical Engineering Courses during COVID-19 Pandemic. *Sustainability*, *14*(9), 5056.

Cevikbas, M., & Kaiser, G. (2022). Promoting Personalized Learning in Flipped Classrooms: A Systematic Review Study. *Sustainability*, *14*(18), 11393.

Ashraf, M. A., Shabnam, N., Tsegay, S. M., & Huang, G. (2023). Acceptance of Smart Technologies in Blended Learning: Perspectives of Chinese Medical Students. *International Journal of Environmental Research and Public Health*, *20*(3), 2756.

Wut, T. M., Xu, J., Lee, S. W., & Lee, D. (2022). University student readiness and its effect on intention to participate in the flipped classroom setting of hybrid learning. *Education Sciences*, *12*(7), 442.

Ni Q, Xu Y., WEI T., & GAO R. (2022).Influence factor analysis and learning effect model based on online learning behavior data. *Journal of Shanghai Normal University (Natural Sciences)*, (002), 051.

Liu, Y., Fan, S., Xu, S., Sajjanhar, A., Yeom, S., & Wei, Y. (2022). Predicting Student Performance Using Clickstream Data and Machine Learning. *Education Sciences*, *13*(1), 17.

Gaftandzhieva, S., Talukder, A., Gohain, N., Hussain, S., Theodorou, P., Salal, Y. K., & Doneva, R. (2022). Exploring online activities to predict the final grade of student. *Mathematics*, *10*(20), 3758.

Dooley, L., & Makasis, N. (2020). Understanding student behavior in a flipped classroom: Interpreting learning analytics data in the veterinary pre-clinical sciences. *Education Sciences*, *10*(10), 260.