Research on Online Learning Behavior of Higher Vocational Students Based on Data Mining

Fenglin Qu(✉)

Chengdu Agricultural Technology Vocational College, Chengdu 611130, China
396280794@qq.com

Abstract. Exploring the characteristics of online learning behavior of students in higher vocational colleges has important guiding significance for improving the teaching level of higher vocational colleges. This paper collects more than 3000 online learning data of "Fundamentals of digital photography" courses of three majors in Cheng du polytechnic. Firstly, t-test is used to study the relationship between learning behavior and learning results, which selected 8 learning behavior indicators such as video resource learning, classroom performance, brainstorming and teamwork. Then, K-means clustering algorithm is used to analyze the path data of students' online learning behavior. The experiment divides students' learning behavior into four categories, and excavates each category of students' learning behavior habits. Provide optimization suggestions for teachers' teaching and students' learning, and provide reference for improving the quality of teaching (learning).

Keywords: online learning · Learning behavior · T-test · K-means clustering algorithm · Data mining

1 Introduction

The rapid development of information technology promotes the application of new technologies such as artificial intelligence and big data in the field of education, and brings new ideas and methods to the study of education and teaching quality. Classroom learning behavior effectively records various data in the learning process, obtains specific information through the processing and analysis of these data, and uses this information to feed back the teaching process, improve the teaching process and improve the teaching quality. Lan-Mo cloud class realizes the trinity of teachers, learners and curriculum knowledge, and fully records the classroom teaching behavior [4]. Based on the real teaching practice of Lan-Mo cloud, this paper studies based on the analysis of more than 5000 Lan-Mo cloud class teaching data in a higher vocational college.

F. Huang et al. (Eds.): ICAIE 2023, AHCS 15, pp. 30–37, 2024.
https://doi.org/10.2991/978-94-6463-242-2_5

2 Material and Methods

2.1 Research Significance

1) It is helpful for teachers to analyze and diagnose problems and adjust teaching strategies [1]. It can comprehensively evaluate the rationality of teaching objectives, the adaptability of teaching models, and the scientificity of teaching strategies and methods. Through the comprehensive analysis of teachers' speech behavior, students' acceptance, learning status and existing problems, we can find out the causes of students' learning difficulties, help teachers optimize teaching methods and improve the quality of classroom teaching.
2) It helps students improve their learning style and academic performance. Through the research on students' classroom learning behavior, this paper is conducive to promoting teachers' learning guidance to students, helping students improve the standardization of classroom learning behavior, so as to promote the improvement of students' learning methods, promote students' understanding and application of knowledge, and help students improve their grades [3].
3) It is helpful to the realization of students' personalized learning. By analyzing the behavior state of learners, it is helpful for teachers to master students' learning ability, learning style and knowledge mastery, so as to provide personalized learning state support for each student, provide personalized learning state support for each student, promote intelligent education, and promote the application of artificial intelligence in teaching and management.

2.2 Analysis on the Current Situation of Classroom Learning Behavior

Data analysis refers to using appropriate statistical analysis methods to analyze a large number of collected data, integrate and refine the disordered data, and find out the internal value law contained in the target research data object [5]. This process is also the supporting process of the quality management system.

Classroom behavior data analysis includes three stages: resource learning before class, participation and interaction in class and consolidation review after class. Some students will actively participate in these three stages and obtain high completion experience value in these three stages; Some students may only participate in the stage of resource learning before class. The sense of participation in class interaction and consolidation review after class is not very strong, resulting in unsatisfactory final results. By analyzing the user's learning behavior trajectory, we can get the conclusion of behavior preference.

2.3 Data Analysis of Learning Behavior Based on Lan-Mo Cloud Class

1) Statistical analysis of students' learning behavior.

This study is based on the real teaching practice of Lan-Mo cloud class, Obtained the learning data of the course "Fundamentals of digital photography" in a higher vocational college. It includes activity statistics before, during and after class. It contains more than 300 pieces and more than 3000 field data. The statistics of student behavior description are shown in Table 1. The results show that students are active in most teaching activities, but less participate in homework group task activities, which needs further attention and attention from teachers.

2) Use T-test method to test the correlation between learning behaviour and school performance.

a) T-test principle.

T-test uses t-distribution theory to infer the probability of difference. By comparing whether the p value (significance) obtained from the t test of the two groups is greater than 0.05, we can judge whether there is significant difference between the two groups of data. It is mainly used for small sample size (e.g. $n < 30$) and overall standard deviation σ Unknown normal distribution data. If $P < 0.05$, there is a significant difference between the two groups of data. If $P > 0.05$, there is little possibility of significant difference between the two groups of data [7]. Independent sample t-test statistics are:

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$S_1{}^2$ and $S_2{}^2$ is Sample variance, n_1 and n_2 is sample size, X_1 and X_2 is average number of samples.

Table 1. Descriptive statistics of students' learning behavior

Teacher Behavior	Behavioral variables (quantity)	Student Participation Experience Value Situation			
		Minimum Value	Maximum value	Average	Standard deviation
Publish Resources	Video resource learning	1	4	3.53	0.96
	Non video resource learning	271	385	374	34.2
Set Active	Classroom Performance	-15	71	23	20.82
	Brainstorming	18	25	22.2	2.26
	Voting Questionnaire	15	17	16.77	0.57
	Light Live Discussion	0	2	1.75	0.43
	Assignment Group Tasks	778	1258	1046.33	127.9
	test	20	62	41.83	11.71

b) T-test of two groups of achievement samples.

According to the course examination results, the students were divided into two groups with 80 points or more and less. Independent sample t-test was used to analyze whether there were significant differences between the two groups in video resource learning, non video resource learning, classroom performance, brainstorming and other behavioral variables. Test the impact of classroom learning behavior on students' curriculum performance. SPSS 19.0 statistical software was used to process the data. $P < 0.05$ was statistically significant. Count the P value obtained from each t-test into excel table for data statistics. The t-test results of the impact of different behavioral variables on the course performance of the two groups of students are shown in Table 2.

c) T-test result analysis.

The test results show that: a) there are significant differences in the impact of non video resource learning, video resource learning, check-in, brainstorming and group mutual evaluation on students' curriculum performance. b) There is no significant difference in the impact of group task participation on students' curriculum achievement. Group cooperative learning is currently recognized as one of the main learning methods, which helps to cultivate students' team spirit, develop social skills and promote high-level cognition. However, group cooperative learning is also prone to dependence. Students may eat "big pot" in the process of participating in group cooperative learning, which explains to a certain extent why group task participation has no significant impact on students' curriculum performance.

Table 2. Comparative analysis of the impact of different behavior variables on students' curriculum performance

Behavioral variables	**Score range**	**average**	**variance**	**t**	**df**	**p**
Video resource learning	>80	3.67	1	2.04	28	0.63
	<80	3.47	0.96			
Non video resource learning	>80	372.33	1444	2.04	28	0.89
	<80	374.14	1175.83			
Classroom performance	>80	38.89	514.61	2.04	28	0.0037
	<80	15.57	273.05			
Brainstorming	>80	22.66	5	2.04	28	0.47
	<80	22	5.5			
Voting questionnaire	>80	16.77	0.44	2.04	28	0.94
	<80	16.76	0.29			
Light live discussion	>80	1.78	0.19	2.04	28	0.096
	<80	1.23	0.79			
Assignment group tasks	>80	1136.77	7225.94	2.04	28	0.009
	<80	1007.57	16389.46			

3) Cluster analysis.

a) Principle of kmeans clustering analysis algorithm.

The k-means algorithm takes K as the input parameter and divides the set of n objects into K clusters, so that the similarity within the result cluster is high and the similarity between clusters is low. Cluster similarity is a measure of the mean value of objects in the cluster, which can be regarded as the centroid or center of gravity of the cluster.

The processing flow of K-means algorithm is as follows [5]: firstly, K optimized initial objects are selected, and each object represents the initial mean or center of a cluster. For each remaining object, it is assigned to the most similar cluster according to its distance from the mean value of each cluster. Then calculate the new mean of each cluster. This process is repeated until the criterion function converges. Generally, the square error rule is adopted, which is defined as follows:

$$E = \sum_{\mathrm{i}}^{k} \sum_{p \in c_i} |p - m_i|^2$$

E: Sum of squared errors of all objects in the dataset; P: Points in space, Represents a given object; m_i: average value of cluster c_i (both p and m_i are multidimensional); k: Number of clusters;

b) Clustering analysis of learning behavior with kemeans algorithm.

Clustering is an unsupervised pattern recognition technology. Its key is feature extraction and clustering algorithm [3]. In this paper, kmeans algorithm is used to cluster non video resource learning, video resource learning, check-in, brainstorming, group task participation and group mutual evaluation. The statistical results show that the clustering coefficient tends to be flat after the four categories. Based on this, this paper clusters the research objects into four categories, and carries out secondary sorting according to the classification. The sorting results are shown in Table 3.

According to the results of cluster analysis of learning behavior, there are obvious differences in learning effects caused by different learning behaviors. Category I learners whose indicators are higher than the class average are less affected by external influence and have high self-discipline, but they still need to pay attention to knowledge construction. Category 3: some of the indicators are higher than the average value, and some are lower than the average value. Learners are greatly affected by various internal and external factors, and such learners should pay high attention to them. Fully understand their psychological state and learning tendency, so as to create more teaching activities that can stimulate their learning motivation, so as to promote their learning effect. Category 4: all indicators are basically lower than the average value. Relatively speaking, learners with limited learning ability and weak learning initiative should be given more supervision, reminder and encouragement, so as to enhance their self-discipline and promote their integration into online learning. This study also conducted an anonymous survey of the students in the classroom. The results show that no matter what kind of learners, they are more concerned about whether they are paid attention to in the time-space separated teaching environment and whether their network learning behavior is seen or recognized. Compared with face-to-face teaching in the past, learners pay more attention to their position in the network teaching environment.

Table 3. Detailed statistics of cluster analysis results

Category Name	Number of platform visits	Number of task points completed	Number of video views	Video viewing duration	Number of answers to post class questions	Number of participants in discussions	Number of classroom interactions	Assignment score
Category 1: All indicators are higher than the class average	452	84	40	341	19	25	22	90
Category 2: All indicators are equal to the average value	212	82	40	290	19	22	22	85
Category 3: Each indicator has a high or low correlation with the average value	144	82	40	222	19	18	14	85
Category 4: All indicators are basically below the average value	149	80	39	216	18	15	13	60
Average value	210	82	39	261	18	19	17	82

3 Conclusion

3.1 Suggestions on the Improvement of Mixed Teaching Behavior

1) Pre class resource push.

The results show that excessive learning of non video resources will affect the learning effectiveness of three kinds of teaching activities: brainstorming, group task participation and group mutual evaluation; The time for students to browse learning resources is mostly concentrated at 12–13 o'clock and 16–17 o'clock. When teachers upload curriculum resources before class, on the one hand, they should appropriately control the proportion of non video resources; On the other hand, we should try our best to push the most important and up-to-date video resources before the time when students browse resources intensively or within this time period.

2) Classroom activities.

Students' behavior is positive in most teaching activities, but they do not participate enough in the teaching activities that embody the concept of "learner centered" - group tasks and group mutual evaluation [6]; Students' willingness to participate in group cooperation is not strong [8]. To solve this problem, in order to improve students' learning enthusiasm and knowledge construction level, teachers should pay attention to when organizing classroom activities: a) Teachers should refine group task activities to avoid the phenomenon of "idling" in the process of students completing group tasks. Teachers should strengthen the guidance and reward of task division and cooperation in order to enhance students' sense of achievement. b) Increase peer evaluation activities and timely feedback on peer evaluation, so as to attract students to participate in learning activities and promote in-depth learning. c) Strengthen brainstorming activities to stimulate students' innovative thinking and speculative thinking and develop high-level thinking ability.

3) Visually present the learning effect and guide personalized learning.

After class, the teacher can view and analyze the big data report of teaching to summarize and give feedback: a) The use of charts, tables and other visual means to visually present the learning effect of students, so that students can quickly understand their own and class learning, and find out their own shortcomings in learning [7]. b) Teachers should use big data technology to deeply analyze students' learning preferences, learning styles and learning needs, and be aware of each student's advantages and existing problems, so as to accurately guide students' personalized learning. c) Understand the overall learning participation of students, and analyze the specific reasons why some students' "experience value" is lower than the average level of class.

References

1. Sun Zejun, sun Yanan, Chang Xinfeng, etc "Internet plus education" student behavior analysis and improvement [J]. Modernization of education, 2021, Vol. 8 (51), pp. 16–21.
2. Zhang Haihua, Li Nannan Research on online learning behavior path based on big data K-means clustering algorithm [J] Electronic design engineering, 2020, Volume 28 (12), pp. 17–20.
3. Sun Wei1, Wang Jianan 2 Analysis of learning input characteristics of MOOCS learners based on big data [J] China Adult Education, 2020 (21), pp. 53–56.
4. Zhao Wei, Yao Haiying, Research on mixed teaching behavior based on blue ink cloud class——Taking "modern educational technology" course as an example [J] Modern educational technology 2019(5), pp. 46–51.
5. Wang Lan, Research on Higher Vocational Students' learning behavior based on cloud teaching and big data [J] Wireless Internet Technology, 2019(16), pp. 95–97.
6. Chahat Vasal, Harshita Pahwa and Tanu Kukreja.Identifying strengths of children with special needs using multiple intelligence theory[J]. Indian Journal of Health and Well-being 2017, 8(11), pp. 1291–1299.
7. Li Na Analysis of personalized learning mode based on big data [J] Wireless Internet technology 2018(02), pp. 64–65.

8. Wang Shijie 1, Cheng Zhijun 2, sun Haichao 3, etc Analysis of MOOC learning behavior of military cadets based on hierarchical clustering [J] Education and teaching forum, 2018, (2), pp. 107–108.