Online Learning Behavior Analysis and Intervention

1st Wu Yue
Beijing University of Posts and Telecommunications Beijing ,China
Beijing ,China
zhxj20120219@163.com

2nd Wang Jiawen
China University of Mining and Technology-Beijing
Beijing ,China
zxjun@cumtb.edu.cn

3rd Li Lin
China University of Mining and Technology-Beijing
Beijing ,China
zxjun@cumtb.edu.cn

4th Cao Tianyu
China University of Mining and Technology-Beijing
Beijing ,China
zxjun@cumtb.edu.cn

5th Zhao Xuejun
China University of Mining and Technology-Beijing
Beijing ,China
zxjun@cumtb.edu.cn

Abstract—**With the rapid development of big data in recent years, education big data has gradually developed to the application level. Reasonable and full use of existing big data, and promoting the further development of online education has become an important issue in today's education industry. This article divides learners' online learning behaviors into four categories, namely, independent learning behaviors, interactive teaching behaviors, homework behaviors, and resource utilization behaviors. Then based on the four types of learning behaviors of learners, a cluster analysis method is proposed to classify the learning types of different learners. Teachers can adopt different educational methods for different types of learners to teach students in accordance with their aptitude, thereby effectively improving learners. The learning efficiency and learning effect of online learning are improved, and the teaching ecosystem is improved.**

Keywords—Learning behavior analysis, clustering, online classroom

I. INTRODUCTION

In recent years, with the further implementation of education informatization, many institutions and capital have entered the field of online education, which has promoted fair and personalized teaching and services for users. As of December 2020, my country has 342 million online education users, an increase of 109 million users compared to before the epidemic. In the face of the impact of the new crown pneumonia epidemic, the Ministry of Education has initiated the work of "suspending classes without suspension" and opening a national network cloud platform for primary and secondary schools. It is convenient for students to study at home and promotes the further understanding of online education in the whole society.

Online learning refers to the behavior of learners to acquire knowledge on online education platforms through multimedia devices. Compared with offline education, online learning has the advantages of rich learning resources, low education cost, and not being affected by time and space. Learners can draw up learning plans and learning methods according to their own needs, and learning behaviors are more flexible and free. However, this learning method also brings new opportunities and challenges to the education industry. How to enable learners to achieve efficient and comfortable learning in online education has gradually become the focus of researchers.

In today's society, big data is flooding all walks of life, and the online education industry is no exception. There are all kinds of education data. The purpose of this article is to collect and analyze a large number of learners' online learning behavior data to mine the hidden information in the data. The data-based analysis results will be more scientific and accurate, and will have similar online learning behaviors. The learners are classified, and the learning habits of the learners are improved through the intervention of the platform and the teacher, thereby enhancing the learning efficiency and learning initiative of different types of learners.

II. ONLINE LEARNING BEHAVIOR

A. Research on Online Learning Behavior

Observing the previous studies of online learning behavior, we can find that researchers mainly analyze online learning behavior from three aspects. First, use relevant software to track and record learners' online learning behaviors, and analyze learners' online learning behavior patterns and learner models. Relevant learning characteristics and processes are the main purpose. The second is to build a relationship model between online learning effects and online learning behaviors, and to study how to improve learners' online learning effects. The third is to study learners' online learning behavior preferences, learning environment, and learning. Analyze the needs to explore the factors that affect the learning effect.

B. Classification of Online Learning Behaviors

The data about learners on the online education platform is huge, and it is extremely diverse and dynamic. Therefore, before data analysis, we must first filter out the useful information for mining learning behavior, and then conduct further analysis on the data. The learning behavior of learners will be affected by many factors, mainly divided into their own factors and environmental factors. Self-factors are usually expressed in learners' own interest in the course, learning ability, learning efficiency, etc.; while environmental factors are expressed in learners' learning environment, learning equipment, platform education resources, etc.

Online learning itself has certain limitations. For the comprehensiveness of online learning behavior data analysis, useful behavior data is selected for analysis based on the three aspects of learning attitude, learning ability and learning effect. Learning attitude is mainly manifested in the learner's degree of completion and seriousness of learning tasks, and the degree of investment in learning; learning ability is mainly the learner's own concentration, self-control, communication, reading ability and other information implicit in the data; Learning effect is the learning result obtained by learners through online learning, which is mainly reflected by academic performance and homework scores.

Before conducting data mining and analysis of online learning behaviors, based on the above three elements, select statistical attributes from online learning behaviors, and divide online learning behaviors into the following four types:

1) Individual learning behavior refers to the behavior of learners to independently carry out course learning and knowledge digestion. The main behavior measures include course video learning time, course video learning times, self-evaluation times and word count, etc.;

2) Teaching interactive behavior refers to the learning communication behavior of learners in the learning process. The main behavior measures include seeking help from teachers and classmates in online communication, publishing theme posts, browsing theme posts, commenting on theme posts, etc.;

3) Homework behavior refers to the behavior of learners to complete homework and exams after studying the course. The main behavior measures include the number of homework submissions, homework scores, final exam scores, and class test scores;

4) Resource utilization behavior refers to the behavior of learners to complete learning through learning resources published on the platform. The main behavior metrics include resource browsing, uploading resources, downloading resources, and marking resources.

Statistics and analysis of online learning behaviors, the results after quantitative processing are accurate and universal, can help grasp the learning status and learning situation of different learners, and help provide learners with targeted learning guidance and curriculum plans. Realize teaching students in accordance with their aptitude.

III. CLUSTER ANALYSIS OF ONLINE LEARNING BEHAVIOR DATA

A. Cluster Analysis

Input unlabeled data into the algorithm, use the algorithm to find the structure hidden in the data, and then divide the points in the data set into multiple separate point sets (clusters) through this data structure. The algorithm used to find clusters is It is called a clustering algorithm. Through the use of clustering algorithms on online learning behavior data, it is possible to mine and analyze online learning behaviors, aggregate learners with similar learning behaviors, and compare and analyze different learner groups, which is helpful

to discover potential learning patterns. Clustering algorithms commonly used in cluster analysis are K-means algorithm, K-medoids algorithm, Clara algorithm, etc.

B. K-means Algorithm

In order to better mine and analyze learners' online learning behavior patterns, we plan to use the K-means algorithm to cluster learning behavior data and aggregate learners with similar learning behaviors into one category, so as to study the potential learning modes of different types of learners at a macro level.

Before clustering the data, because the gauges used by different variables are not the same, it is first necessary to normalize the variables of different attributes in the online learning behavior. Secondly, it is necessary to preprocess the data, filter the extremely unreasonable learner samples, and then cluster the data according to the established similarity standards. K-means algorithm is an iterative algorithm. It mainly does two things. The first is cluster allocation, and the second is moving cluster centers. Randomly generate K points as cluster centers to divide the data into K categories; then traverse each sample in the data set, calculate the distance from each data to the cluster center, and divide the data samples according to the distance; , Move the cluster center to the mean value of the sample of the same type; repeat these two operations until the cluster center will not move even if the K-means algorithm is iteratively run, and the class of the sample will not change. The function expression of the clustering criterion is:

$$J(c^{(1)}, \ldots, c^{(m)}, \mu_1, \ldots, \mu_k) = \frac{1}{m} \sum_{i=1}^{\infty} |x^{(i)} - \mu_{c^{(i)}}|^2 \qquad (1)$$

In the expression, $c^{(i)}$ represents the current sample, $x^{(i)}$ represents the index or serial number of the cluster, μ_k represents the position of the k-th cluster center, $\mu_{c^{(i)}}$ represents $x^{(i)}$ The cluster center of the cluster to which it belongs.

C. Selection of the Number of Categories K

It is known from the expression of the clustering function that the K-means algorithm is to minimize the square difference between the sample and the mass point as the objective function, and the square of the distance difference between the cluster center and the sample point is usually called the degree of distortion. Therefore, for a cluster, the lower the distortion value, the tighter the samples in the cluster, on the contrary, the higher the distortion value, the looser the structure in the cluster. Generally, the degree of distortion will decrease with the increase of the category. When reaching a certain critical point, the descending speed will tend to be flat, and the obtained curve will be like an elbow. In this case, the K value is the best cluster number. This selection The K value method is called the elbow rule, but in many cases the position of the curved elbow is not clear. This study uses the K-means algorithm to classify different learners. Therefore, while ensuring that data classification is beneficial to online learning behavior research, the K value selection should be close to the optimal number of clusters.

D. Analyze the Learning Types of Learners

In order to better analyze the online learning behavior of different learners, after clustering the data, the learners are divided into the following four types:

1) Active learning type: This type of learner is highly engaged in learning, has a long online learning time, can make better use of online resources, actively communicate with teachers and classmates, and can obtain better academic results through their own efforts. Teaching interactive behavior refers to the learning communication behavior of learners in the learning process. The main behavior measures include seeking help from teachers and classmates in online communication, publishing theme posts, browsing theme posts, commenting on theme posts, etc.;

2) Independent learning type: This type of learners has strong independent learning ability and good completion of courses and homework, but lacks interactive teaching behaviors, hardly expresses opinions on the platform, does not summarize learning experience, and does not make effective use of online learning platforms. Resource utilization behavior refers to the behavior of learners to complete learning through learning resources published on the platform. The main behavior metrics include resource browsing, uploading resources, downloading resources, and marking resources.

3) Ordinary learning type: This type of learner can complete all types of classroom tasks issued by the teacher, and all types of learning tasks also meet the qualified standards, but they have no enthusiasm and interest in the content of learning, learning effects are average, and online learning behaviors are not Obviously outstanding features.

4) Passive learning type: This type of learner has low course participation and low learning input. They only watch part of the courses that are of interest. The completion of courses and homework is low, resource utilization is low, and learning effects are poor.

IV. Analysis Result Application

A. Platform Optimization Strategy

The teaching platform should aim to promote the growth of learners, analyze learning data, improve platform construction, and improve teaching quality. Through the analysis of learners' learning behavior, it is found that some learners cannot make full use of the learning functions of the platform. The platform should design learning guides based on four different types of learners to cultivate good online learning habits for learners. In addition, the interaction between the platform and learners can be increased appropriately. For example, if a daily learning goal is reached or a course is completed, the platform will give learners a certain number of points, and users who reach a certain level of points can get a certain reward. This reward-based teaching can not only attract learners of different ages, but also increase the interest of learning and stimulate learners' enthusiasm.

The online platform has a large amount of learner online learning data. The platform can classify the data according to

courses, and present the data to learners and teachers using visualization techniques such as curve graphs and bar graphs. Because online education and offline education, teachers lack interaction with learners, and they can't understand students' learning situation in time through teaching interaction. If the platform provides detailed visualization data for each student, teachers can use simple data graphs. To understand the learning situation of each student, you can communicate with learners with learning problems in a timely manner and help adjust the status. It can also determine the knowledge mastery of most students based on the data, which is beneficial to the later teaching plan.

Using visual data, learners themselves can also understand their own learning conditions and the learning conditions of most other learners to make self-adjustment, and at the same time enhance the awareness of competition among learners to promote common progress. The platform can also add learner competition functions on this basis, measure the learning status, homework completion status, and learning concentration, compare the learning status between learners, and promote the learning of both parties. In addition, group learning can be increased. Based on online learning behavior analysis, different types of learners are randomly divided into several learning groups. Group members can communicate with each other, and each learn from their strengths to make progress together.

The platform can also implement personalized course recommendation for learners through data collection. It can start with the courses that learners have learned and learners' interests, collect personal data and use recommendation algorithms to make personalized recommendations for users. For a simple example, learners who watched a python course are likely to be interested in machine learning related videos, and machine learning courses can be associated with advanced mathematics courses. On the contrary, learners have a preference for computer courses, so they will study architecture and art. The probability of such courses will be reduced, and the platform can appropriately reduce relevant recommendations. In addition, you can start with the courses that the learners have learned in the past, and calculate whether the learners pay attention to the influence of the course, the strength of the faculty, or the number of students in the course, to analyze the learning preferences of the learners. Through the analysis of learners' learning preferences, it can help the platform increase learners' adhesion to the learning platform, enhance learners' interest in learning and stimulate learning enthusiasm, thereby enhancing learners' learning commitment and enthusiasm.

Online learning platforms have a large amount of learning data. In order to further improve the education platform and optimize the online education environment, it is still necessary to thoroughly analyze the huge online learning data.

B. Teacher Education Strategy

Teachers as educators need to make full use of online learning behavior data to bring better quality education to students. The biggest problem for teachers in online teaching is that they cannot interact with students in real-time teaching, and it is difficult to obtain students' learning from their demeanor. If the teacher does not pay attention to the learning

status of the students and continuously unilaterally outputs knowledge, it is easy to affect the learning efficiency of the students. Therefore, it is very necessary to adjust the teaching activities according to the learning behavior data of the learners.

Teachers can obtain a lot of useful information from the data, such as whether learners' attention is concentrated during a certain period of time. If the data shows that most learners are not concentrating in teaching, then teachers can make improvements in the content and methods of class. Choose to teach in a more simple and interesting way to attract the attention of learners.

In online education, the correction of classwork and off-class homework is highly efficient. Teachers can view the error rate of homework more intuitively, and make timely teaching adjustments based on platform data to help students better absorb knowledge.

In addition, through online learning behavior analysis, different learners can be classified, which helps teachers discover the different learning behaviors of different learning groups, and helps teachers achieve true teaching in accordance with their aptitude. Teachers can intervene learners' online learning behaviors according to different learning habits and learning characteristics of different types of learners, plan different learning plans and learning tasks for learners, help learners improve learning efficiency and learning effects, and promote teaching effects. Promote.

V. CONCLUSION

With the widespread application of big data analysis technology in the education industry, the analysis capabilities of online education data have been continuously enhanced, which facilitates educators and researchers to study the impact of different environmental factors and their own factors on learners' learning effects from many aspects, which is beneficial to Realize student-centered personalized learning and intelligent teaching activities. Through further research,

the online learning system can be optimized, the functions of the online learning system can be improved, and the learning efficiency of learners can be improved based on online learning behaviors, and the learning preferences of learners can be further studied to bring learners a more comfortable learning experience.

Data-based learning behavior analysis is accurate and diverse. The use of K-means clustering algorithm can effectively grasp the learner's online learning behavior status from the macro level, analyze the learning behavior of each learner from the micro level, help teachers improve the curriculum arrangement, help learners improve learning efficiency, and provide the platform at the same time. The construction of the resource bank and the construction of the learning system brought inspiration. Using huge online learning behavior data of learners to classify learners can in a real sense realize teaching students in accordance with their aptitude, and stimulate learners' enthusiasm for learning.

REFERENCES

[1] Lin Dingbao. Analysis of college students' learning behavior based on big data[J]. Journal of Hubei Open Vocational College, 2019, 32(10): 28-29.

[2] Wang Bo. Research on the construction of online learning behavior analysis system for the purpose of good intervention [J]. Electronic World, 2021(02): 98-99+102.

[3] Sun Jinping, Bao Rong, Li Dan. Analysis of student learning behavior in the network teaching environment[J]. Computer Knowledge and Technology, 2021, 17(02): 73-74+88.

[4] Tang Dongdong, Xu Yuanyuan, Meng Yanmei. College English online teaching effect evaluation and online learning behavior analysis [J]. Chinese Journal of Multimedia and Online Teaching (first issue), 2020(01): 61-62.

[5] Yao Jialin, Tan Zhao, Yu Songnan, Ju Yinyin, Liu Sujuan. Study on learning behavior analysis based on K-means clustering algorithm and naive Bayes classifier[J]. Computer Knowledge and Technology, 2019, 15(11): 17-20.

[6] Hu Yanxue, Huai Libo, Cui Rongyi. Online learning behavior analysis method based on interest degree association rules[J]. Journal of Yanbian University (Natural Science Edition), 2019, 45(01): 40-44.