# Analysis of Online Course Learning Data Based on Density Peak Clustering and Research on Teaching Mode

Han Guo, Rui Zhang, and Xiao-Nan Fang(✉)

School of Information Engineering, Shandong Management University, Jinan, China
14438120201031@sdmu.edu.cn, franknan@126.com

**Abstract.** With the development of Zhihuishu platform, online courses have also been constructed and promoted. In this study, taking the web front-end development technology course as an example, this paper uses the Density Peak Clustering method to analyse the online learning data of students. By watching videos and submitting tests, students can be divided into three types: active, medium and passive. The attitudes of medium and passive students towards online courses need to be further improved. It shows that the intervention and guidance of teacher need to be increased in course teaching. Finally, this research explores the "before class - in class - after class" trinity hybrid teaching mode.

**Keywords:** Density Peak Clustering · data analysis · online course · Hybrid teaching mode

## 1 Introduction

Online learning is a direction of education and teaching reform in universities. It refers to providing students with the ability to log in to online websites through the Internet at an irregular time and place, using their own computers, mobile phones and other mobile devices. Compared with traditional online learning, current online learning integrates more interactive elements relying on information technology, so it is generally welcomed by students. It is effectively to combine the advantages of traditional teaching mode with network or digital teaching to form the "Online+offline" hybrid teaching mode. Sun [7] integrated CDIO concept into course teaching in teaching methods and means. To carry out mixed teaching, we should make effective use of network teaching resources and teaching means to improve learning efficiency and stimulate learning interest.

Colleges and universities widely carry out mixed teaching. Especially in the epidemic, online education platforms have accumulated a large amount of education data. Many researchers analyse and predict the learning behaviour of the data [2, 6, 9]. Clustering algorithm is one of the key technologies to deal with big data. It automatically divides data sets according to the similar characteristics of data, and divides objects into different categories according to certain rules and object attributes. Objects of the same category have certain similarities, while objects of different categories have great


Z. Zhan et al. (Eds.): EIMSS 2022, AHCS 7, pp. 457–465, 2023.
https://doi.org/10.2991/978-94-6463-024-4_48

differences. Clustering algorithm is an important unsupervised learning technology in machine learning, which has been widely used in many fields. Xie et al. [8] used k-means algorithm to cluster teaching videos, tests and platform access. Based on the combination of feature selection of genetic algorithm and error correction output code, the student performance level is predicted. The results of the study are helpful to the implementation of personalized teaching for students of different grades and different learning modes. Jovanovic et al. [3] applied cluster models for grouping students based on their cognitive styles in e-learning environment. Clustering students based on cognitive styles and their overall performance should enable better adaption of the learning materials with respect to their learning styles. Luo and Wang [4] performed a time-series clustering algorithm on the 5 test scores of students online and offline, and obtained three distinct student types (Excellent, Moderate and Poor). It is believed that long-term online learning can obtain good and stable test scores.

## 2 Related Works

Li et al. [5] applied K-means, Spectral Clustering and Agglomerative clustering algorithm to analysis the experimental data. The analysis of learning behaviour can effectively realize the division and help to improve the quality of teaching. At present, the commonly used clustering method K-means is only suitable for clustering comparison of convex sample sets. The category of K-means needs to be determined manually, and the iterative selection of family centre points is risky. Density Peak Clustering (DPC) algorithm has the advantages of simple idea and less parameter requirements. It can identify clusters of arbitrary shape. It is suitable for data with irregular spatial distribution. As shown in Fig. 1, the distribution of 3D data is irregular.

DPC is a granular computing model. DPC algorithm is based on two assumptions: One is that the cluster centre is surrounded by nearest neighbour data points with low local density, the other is that any cluster centre is far away from the data points with higher density [1].

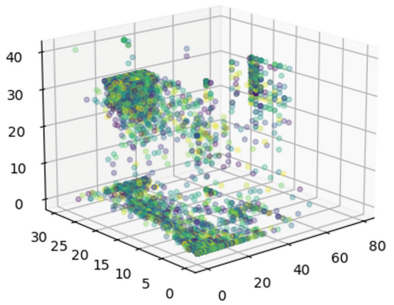


**Fig. 1.** 3D score data distribution map of student.

- The method of DPC first calculates the distance between any two data points;
- It finds the truncation distance $d_c$ for density calculation. It is required that the number of points whose distance around each point is less than $d_c$ accounts for 2% of the total points;
- The method then calculates the local density $\rho_i$ of each data point according to the truncation distance $d_c$;
- It calculates the density distance $\delta_i$ of each data point. Each point finds all points whose density is greater than it, and then finds the distance from the nearest point among these points;
- Then, the method selects the point with large $\rho_i$ and $\delta_i$ as the cluster centre through the threshold.
- Finally, the method allocates the remaining points. Each remaining point is allocated to its nearest neighbour and the cluster where the density of data points is greater than it.

**Fig. 2.** DPC algorithm

For any data point *i*, DPC needs to calculate the local density of two quantities: local density $\rho_i$ and its distance from the nearest point with higher density $\delta_i$. Local density $\rho_i$ is defined as

$$\rho_i = \sum_{j=1}^{n} \chi(d_{ij} - d_c) \tag{1}$$

Where $n$ is the number of data points; $d_{ij}$ is the distance between data points $i$ and $j$; $\chi$ is the indicating function, when $x < 0$, $\chi(x) = 1$, otherwise $\chi(x) = 0$; $d_c$ is the Truncation distance, the density $\rho_i$ is equal to the number of points distributed in the $d_c$ neighbourhood of $i$. $\delta_i$ is measured by calculating the minimum distance between point $i$ and other points with higher density (Fig. 2).

$$\delta_i = \min_{i:\rho_j < \rho_i} d_{ij} \tag{2}$$

## 3 Experimental Results and Analysis

"Web front end development technology" online open course is an online course of Shandong online course alliance established by School of Information Engineering of Shandong Management University. It was officially opened on Zhihuishu platform in the spring and summer of 2019–2020 academic year.

**Table 1.** Course information statistics

| No. | Items of course information | Quantity |
|---|---|---|
| 1 | The number of universities | 29 |
| 2 | The number of disciplines | 57 |
| 3 | The number of students | 4364 |
| 4 | Grade category | 4 |

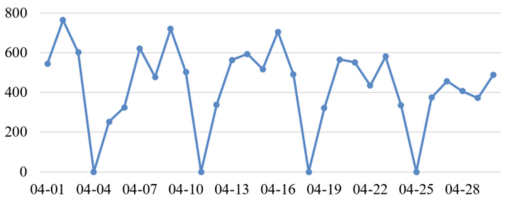


**Fig. 3.** Record of learner participation in the course (part)

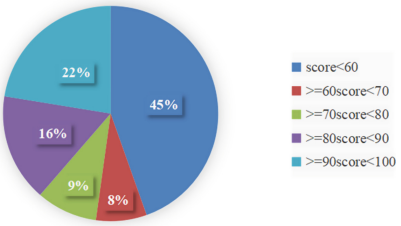


**Fig. 4.** Proportion of students in each section of total score

This paper selects the student learning data of the second semester of the 2019–2020 academic year. There are 4364 students from 29 schools. Table 1 makes statistics on the overview and data records of Web front-end development technology courses currently offered on the Zhihuishu.

The learning time of online course is different from the courses in the fixed classroom, and it is relatively free. Students can arrange their own study time. In this article, the learning situation in one month of the semester is selected for statistics. The results are shown in Fig. 3. It can be seen from the figure that the learning time is regular. However, it shows a downward trend at the end of the month, indicating that students are inactive.

In this research, the final scores of students are counted and divided into five intervals. As shown in Fig. 4, the proportion of people whose scores are less than 60 and greater than or equal to 90 and less than 100 is the highest, indicating that the polarization is very serious.

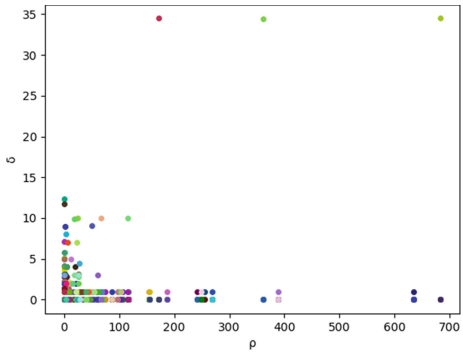


**Fig. 5.** Decision graph of learning progress and chapter test scores

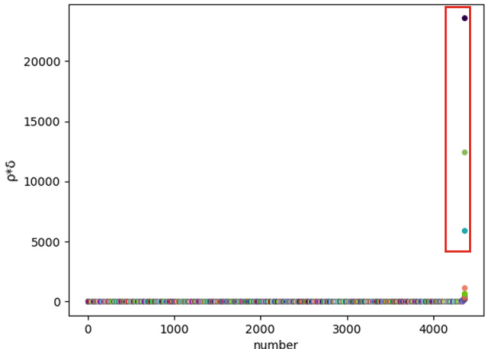


**Fig. 6.** Improved decision graph of learning progress and chapter test scores

Among the many types of learning behaviours, this research chooses watching videos and submitting tests, which are the most extensive and basic behaviours for analysis. Among them, watching videos represents the learning motivation and initiative, which is translated into learning progress scores. Submitting tests indicates the degree to which the learner studies seriously.

Based on these two quantities, DPC divides the data points into three different types through the decision graph named density peak points, normal points and outliers. As shown in Fig. 5, the points are arranged in the order of decreasing density. It is easy to see that three points are very prominent, distributed in the upper right corner of the graph, with very high $\delta$ value and $\rho$ value, indicating that there are no data points with higher density than them in a large neighbourhood. Therefore, these three points are the so-called density peak points, which are suitable for being selected as cluster (category) centres. Although three points have a high $\rho$ value, and $\delta$ is very small. It is distributed in the lower right corner, indicating that there are points with higher density in their immediate neighbours, so they are not peak points and are not suitable to be used as cluster centres.

In order to accurately describe the clustering centre, this research uses the improved decision graph to calculate the product of $\rho$ and $\delta$. These results display in ascending

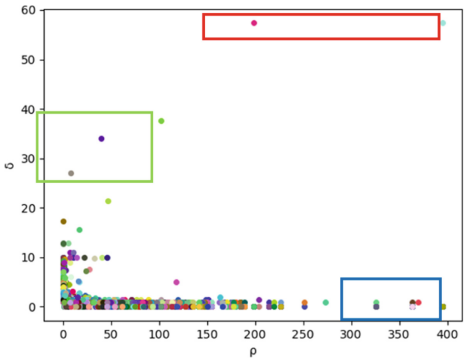


**Fig. 7.** Decision graph of learning progress and chapter test scores

order. As shown in Fig. 6, three points are very prominent and distributed in the upper right corner of the figure, the product of $\delta$ and $\rho$ is very high. According to the results of the experiment, these students can be roughly divided into three categories. The first category is active. It means that the students watch the videos and submit the test well. The second category is passive, and the students fail to complete watching videos and submitting tests. The third type of students are medium, who can well complete the task of watching videos, but fail to complete the task of submitting tests in time, indicating that the attitude of students towards serious learning needs to be further improved. Therefore, the guiding role of teachers in teaching needs to be strengthened.

Considering the learning situation, this paper clusters the scores of four dimensions: learning progress, daily performance, chapter test and final test. The results are shown in Fig. 7. The data points are arranged in the order of decreasing density. It is easy to see that two points are very prominent, distributed in the upper right corner of the graph, with very high $\delta$ Value and $\rho$ value, indicating that there are no data points with higher density than them in a large neighbourhood. Therefore, these two points are the so-called density peak points, which are suitable for being selected as cluster (category) centres. Although six points have a high $\rho$ value, $\delta$ value is very small. It is distributed in the lower right corner, indicating that there are points with higher density in their immediate neighbours, so they are not peak points and are not suitable to be used as cluster centres. Three points in the green box have higher $\delta$, $\rho$ value is very low, distributed in the upper left corner, indicating that They are outliers.

The improved decision graph of learning progress, daily performance, chapter test and final test is shown in Fig. 8. Those two points are very prominent and distributed in the upper right corner of the figure, the product of $\delta$ and $\rho$ is very high. According to the results of the experiment, these students can be roughly divided into two categories: the first category is active, and the other category is passive. This result is consistent with the final score distribution in Fig. 4. The active students performed well in all dimensions, while the passive students performed poorly in all dimensions.

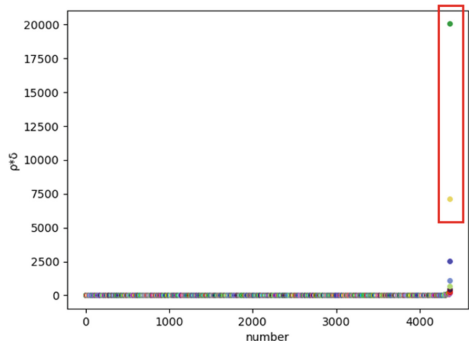


**Fig. 8.** Improved decision graph of learning progress, daily performance, chapter test and final test

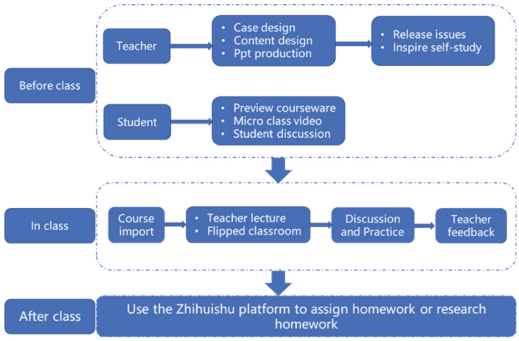


**Fig. 9.** The trinity teaching model of "before class - in class - after class"

## 4 Discussions on Teaching Mode of Network Front-End Development Technology Course

Through data analysis, students do not attach much importance to online teaching, and their attitude towards serious learning needs to be further improved. Therefore, it is necessary to strengthen the guiding role of teachers in teaching. This research discusses the trinity teaching model of "before class- in class- after class".

Before class, teachers make full lesson preparation and teaching design, and can use the Zhihuishu platform to release problems. Students can use the Zhihuishu platform to preview before class, micro class video learning and problem discussion. In the class, the teacher first introduces the course, which can review the knowledge points of the previous class or introduce the teaching content of this class through cases, then the teacher teaches the relevant knowledge points, and then uses the rain class for classroom practice to let the students show the results, and finally the teacher comments and feedback. After class, teachers use the online education platform to assign homework or research homework after class, and finally summarize and feed back in the follow-up courses (Fig. 9).

Based on the OBE concept, realize the mixed teaching of Web front-end development training courses, and improve the depth of students' learning knowledge. Online

resources are the premise of hybrid teaching. Move the traditional classroom teaching forward through the online form of micro video, give students sufficient learning time, and let each student enter the classroom with a better knowledge base as far as possible to fully ensure the quality of classroom teaching. Offline students are organized to consolidate and flexibly apply the basic knowledge learned online through carefully designed classroom teaching activities. Establish a learning feedback mechanism to make teaching activities more targeted and teachers better grasp teaching.

Provide teaching videos for students in each class. Before class, students watch micro class videos through the Zhihuishu platform and preview the courses with other online teaching resources in advance. Each micro class video has strong pertinence, and the playback time of the videos is controlled at about 10–20 min. In class, mixed teaching is carried out with the help of online software to improve class efficiency and help students deepen their understanding of knowledge. The teaching application software can be used for code scanning and check-in, courseware release, online question answering, classroom feedback, online discussion, point answering, online mutual evaluation and other activities.

In the course implementation, the teaching method is combined with teaching method, discussion teaching method, case driven teaching method, flipped classroom teaching method, guiding teaching method, autonomous learning method and so on. Flipped classroom is built under the background of Internet informatization. Students can preview the course content in advance through online resources. The flipped classroom teaching mode breaks through the time and space constraints of the traditional teaching mode and constructs an open teaching space, which can meet the learning needs of different levels.

## 5 Conclusions

This paper uses DPC algorithm to analyse students' online learning data. By analysing the behaviours of watching videos and submitting tests, students are divided into three types: active, medium and passive. This research studies the teaching mode through data analysis, and discusses the trinity teaching mode of "before class - in class - after class". In the process of teaching, the guiding role of teachers in teaching should be strengthened.

**Acknowledgements.** The authors sincerely thank reviewers for their valuable comments. They thank for the learning behaviour data provided in Zhihuishu platform. Finally, they thank teaching research of Shandong Management University (no. YJG2019-13) for support of this project.

## References

1. Chen YW, Shen LL, Zhong CM et al (2020) Survey on density peak clustering algorithm. J Comput Res Dev 57(2):378–394
2. Jiang ZX, Zhang Y, Li XM (2015) Learning behavior analysis and prediction based on MOOC data. J Comput Res Dev 52(3):614–628

3. Jovanovic M, Vukicevic M, Milovanovic M et al (2012) Using data mining on student behavior and cognitive style data for improving e-learning systems: a case study. Int J Comput Intell Syst 5:597–610
4. Luo J, Wang T (2020) Analyzing students' behavior in blended learning environment for programming education. In: WSSE 2020, proceedings of the 2020 the 2nd world symposium on software engineering. Association for Computing Machinery
5. Li Z, You F, Sun D (2020) Research on the evaluation of learning behavior on MOOCs based on cluster analysis. In: 4th international conference on electronic information technology and computer engineering. Association for Computing Machinery
6. Mai TT, Bezbradica M, Crane M (2022) Learning behaviours data in programming education: community analysis and outcome prediction with cleaned data. Future Gener Comput Syst 127:42–55
7. Sun B (2021) Exploration on the construction of "procedural control" course based on concept of CDIO and online-offline hybrid teaching mode. Creat Educ Stud 9(1):1–9
8. Xie ST, Chen Q, Liu KH et al (2021) Learning behavior analysis using clustering and evolutionary error correcting output code algorithms in small private online courses. Sci Program 2:1–11
9. Yang S, Juan L, Lei X et al (2018) Research on data mining of learning behaviours of college students on MOOC platform. MATEC Web Conf 246(3):03025