

Research on Data Mining of Learning Behaviours of College Students on MOOC Platform

Shi Yang¹, Lu Juan²

¹Vocational Education Centre, Naval University of Engineering, PLA, Wuhan, 430033, China

²Vocational Education Centre, Naval Medical University, PLA, Shanghai, 200433, China

Abstract: With the continuous development of computer network and the popularity of internet applications, technology is constantly changing the traditional education model. The rise of the MOOC has set off a worldwide revolution in educational technology, which has been widely welcomed by university teachers and students. On the platforms of MOOC, the learning behaviours of college students have generated massive amounts of relevant data. Teachers can tap learning behaviours, master different types of learning styles to better control the learning steps and urge college students to better participate in all aspects of learning. Based on the MOOC platform, this paper classifies the students into excellent learners, middle learners, poor learners and non-learners by cluster analysis to teach students of different levels in different ways to optimize the MOOC teaching effect.

1 Introduction

With the progress of information education, the traditional teaching methods cannot meet the multi-dimensional learning requirements. The traditional teaching is often through the teacher's words and deeds in the classroom to complete the whole teaching, teaching methods are single, and subject to space and time constraints, college students learning effect cannot be timely feedback to teachers. Online education emerges as the times require in the context of traditional education. It has become the trend of global education. In 2011, the Massachusetts Institute of Technology (MIT) launched the Open Courses Program, setting off a wave of international MOOC (Massive Open Online Courses). MOOC can be so popular, precisely because MOOC as a large-scale, open and emerging network education model, adhering to the common concepts and methods of online education. Domestic educators learn from foreign MOOC platform construction model, combined with specific national conditions, further launched a large-scale online open curriculum research, improve the quality of online teaching, to meet online learning needs.

The big data of college students on the MOOC platform bring new opportunities for emerging online education. Learning analysis technology is also a new hotspot in the field of educational technology in recent years. It is a new technology to interpret and analyse the data generated by college online learning to evaluate and evaluate their learning progress, predict their future performance and discover potential problems in the learning process. Researchers can use these data to find out the internal relationship between learning effect and

college students' learning situation. Learning analysis technology is how to analyse and apply data generated in education. When college students use computers and the internet for online learning, they will produce much data in the learning process, including the basic information submitted by college students when they register, such as personal information, education level, learning preferences, and the learning process information generated by college students learning on the platform. The information is obtained and stored in different data forms through the online learning activities of college students. Some data directly show the progress of college study and test results. Others need to be resolved by means of relevant analysis tools or methods to understand other questions of college students in learning activities. Nowadays, in some advanced online course learning platforms, when college students use the platform, the platform itself will synchronously record college the learning behaviour data. We can analyse the data of learning behaviours of college students on MOOC platform, divide the students into different categories, and give some suggestions to form relevant learning groups.

2 Learning behaviors on MOOC platform of college students

On MOOC platform, learners can carry out video learning, data downloading, discussion area question and answer, which not only includes the process of traditional teaching and learning, but also embodies the advantages of network learning. By collecting and mining learning behaviours data in the network learning environment, the learning process can be better optimized and improved. Learning efficiency. Learning in this paper refers to the process of

acquiring knowledge through the classroom or MOOC, including active and independent learning, but also includes traditional classroom learning. Learning behaviours refers to all kinds of learning activities, such as consulting materials, completing exercises and so on, which are displayed by the participants in the learning process.

In the process of MOOC learning, the effective learning behaviours of college students can be roughly divided into the following categories. Choosing the course to study: Before learning the MOOC course, we need to choose the higher course to study, and the course selection is the premise of learning. Watching instructional video: MOOC platform in the teaching video most directly shows the course of teaching content, is the most willing way of learning for most college students, and therefore is the most important learning behaviours of College students. Download material: Learning material is a supplement to the teaching video. In addition to the course courseware, it is often provided with supplementary information for college students to expand their knowledge according to their own learning situation, so it is also an important way of learning. Forum Exchange: The MOOC platform provides a forum for college students and teachers to exchange learning experience freely in the forum or to seek answers to learning questions. Unit assignment evaluation: Unit assignment evaluation is an effective way to test the learning effect in time. Completing unit assignment evaluation in time can help college students find and fill in the gaps in knowledge. Invalid learning behaviours mainly include: video pause, frequent page refresh, posting unrelated to the curriculum, browsing unrelated to the curriculum, etc. These behaviours will greatly affect the concentration of college students in the learning process, not conducive to the improvement of learning effect. Data mining and analysis of learning behaviours can play a very important role in the process of teaching and learning, and promote learning efficiency. This paper will give a very practical guidance to college students, teachers and the design and development of MOOC platform.

3 Concept and common method of cluster analysis

Clustering analysis is a kind of unsupervised learning, which returns similar objects to the same cluster. It's sort of like automatic classification. Clustering analysis can be applied to almost all objects. The more similar the objects in the cluster, the better the clustering results. Cluster analysis can well reflect the relationship between classes. When two objects belong to the same class, the two objects have the same or similar properties. When two objects belong to two different classes, their properties are different. Objects belonging to the same class have the same characteristics. Cluster analysis can study the properties of objects behind the data, and it is important for us to understand these objects. Cluster analysis can use cluster centres to represent this class well. The nature of it. Clustering centre is generally the average value of each object belonging to this class, which can reflect some data

characteristics of this class. Comparing different kinds of clustering centres, we can find that different clusters have different meanings. Clustering analysis can help us extract important information from data. The meaning of data mining is to extract the information we need from the data information. Cluster analysis reflects the validity of the data information to some extent.

K-Means clustering algorithm based on partition method is the most commonly used clustering algorithm in clustering analysis. Firstly, this kind of clustering algorithm usually needs input parameter K in advance, K represents the number of clustering partitions, and then iteratively updates the data set until the criterion function converges to a certain constant. The workflow of K-means clustering algorithm: Firstly, K initial points are randomly determined as the centre of mass, and then each point in the data set is allocated to a cluster. Specifically, the nearest mass must be found for each point and assigned to the cluster corresponding to the centre of mass; then the mass of each cluster is updated to the average value of all points in the cluster. K-means algorithm is a process of continuous iteration. The final result is to minimize the square of the distance between all samples in the cluster domain and the cluster centre and $J(C)$.

$$J_C = \sum_{j=1}^k \sum_{i=1}^{n_j} \|X_i^{(j)} - Z_j(I)\|^2$$

Among the above formula, I denotes the number of iterations; $Z_j(I)$ denotes the cluster center; $X_i^{(j)}$ represents any data object in class j.

4 Empirical analysis of learning behaviours of college students on MOOC platform

This paper will use clustering algorithm to cluster the data of the learning behaviours of 30 college students on MOOC platform to divide them into excellent learners, middle learners, poor learners and non-learners.

4.1 Step 1: indexes selection

This paper studies college learning behaviours on MOOC platform. Taking Advanced Mathematics as an example, we chose five indexes as basic variables: time of online, time of watching videos, download number of learning materials, posting number in forum and finishing ratio of homework. Their meanings are as follows. Time of online: It reflects the learner's learning time, learning time is too short cannot guarantee the learning effect, learning time is too long also does not mean its high utilization of learning time. Time of watching videos: It is the guarantee for learners to acquire knowledge. The increase of viewing time can obviously improve their learning effect. But if the viewing time of videos is much longer than the length of videos themselves, it indicates that learners have learned in the learning process. Habit related behaviours such as browsing unrelated web pages during video playback. Download number of learning materials: learning materials is an important way to acquire

knowledge and expand knowledge, can supplement teachers in teaching videos did not involve or briefly explain the content, so the download of learning materials to a certain extent indicates that students are not active in acquiring knowledge. The download times of learning materials can also affect the learning effect of students. Posting Number in Forum: It is mainly a place for learners to communicate, answer questions and solve puzzles. Therefore, it is necessary to explain the difficulties encountered in the process of learning in the forum in a timely manner, indicating that the learner is not acquiring knowledge in a round way, actively solving the blind spots

of knowledge, and can solve them in a timely manner. Answer other students' questions while helping themselves consolidate their knowledge. Finishing ratio of homework. It can help learners in a timely manner to check their own knowledge to fill in the gaps, greatly promoting the improvement of learning effect.

4.2 Step 2: data collection

The data of the leaning behaviours of thirty college students on MOOC platform is shown as follows.

Table 1. Data of the selected indexes of 30 students

Index	I ₁	I ₂	I ₃	I ₄	I ₅
Unit	Hour	Hour	Time	Time	1
Student 1	190	124	65	105	0.98
Student 2	155	97	33	15	0.35
Student 3	185	108	68	124	0.99
Student 4	130	93	45	22	0.66
Student 5	170	99	46	12	0.81
Student 6	116	52	15	24	0.52
Student 7	207	125	60	85	1
Student 8	131	96	52	44	0.85
Student 9	142	100	40	12	0.63
Student 10	169	84	39	14	0.71
Student 11	131	98	30	56	0.42
Student 12	160	81	40	38	0.44
Student 13	160	86	39	58	0.56
Student 14	137	91	46	49	0.21
Student 15	132	95	41	36	0.1
Student 16	144	90	49	52	0.94
Student 17	1	0	0	0	0
Student 18	198	107	71	77	0.89
Student 19	160	91	57	15	0.25
Student 20	165	55	55	14	0.36
Student 21	152	94	55	30	0.63
Student 22	160	75	35	24	0.88
Student 23	93	58	55	1	0.61
Student 24	123	88	35	17	0.63
Student 25	142	95	55	5	0.66
Student 26	150	46	55	25	0.65
Student 27	133	80	55	63	0.71
Student 28	1	1	0	0	0
Student 29	129	98	8	36	0.61
Student 30	105	45	25	24	0.35

4.3 Step 3: clustering analysis

We use the system clustering method to do the clustering analysis with the help of the software of SPSS 24.0. We will take 30 students as the "cases" and to select the K-

Means clustering form. We can obtain the clustering results. From the table, we can see that if the samples are divided into four categories. The first class is Student 1, Student 3, Student 7, Student 8 and Student 18; the third

class is Student 6, Student 23, Student 24 and Student 30; the fourth class includes the Student 15 and Student 28; the rest of the students are second class.

Table 5. Clustering result of leaning behaviours of college students of MOOC

Case	Category	Case	Category
Student 1	A	Student 16	B
Student 2	B	Student 17	B
Student 3	A	Student 18	A
Student 4	B	Student 19	B
Student 5	B	Student 20	B
Student 6	C	Student 21	B
Student 7	A	Student 22	B
Student 8	A	Student 23	C
Student 9	B	Student 24	C
Student 10	B	Student 25	B
Student 11	B	Student 26	B
Student 12	B	Student 27	B
Student 13	B	Student 28	D
Student 14	B	Student 29	B
Student 15	D	Student 30	C

4.4 Step 4: conclusion analysis

The students of the first class are excellent learners. They have enough time to study the course, a good learning plan, enough learning time is the basis for good learning results; at the same time, they watch almost all the teaching videos. good completion of homework and actively participate in the discussion area, but Help them to find the missing in time, so we can infer that they are excellent learners and achieve excellent learning results. Secondary learners are tending to complete the viewing of teaching videos, complete the required homework, occasionally participate in the discussion area question and answer, learning links are involved, but because the learning time is not fully guaranteed, it is difficult to have a better learning effect. The third category is poor learners. They can't guarantee to learn at a certain pace. They log on to the system less often. Although they do more homework, they watch less videos than the total length of the course videos. Maybe they don't concentrate on learning after they log on to the system. Face, on the contrary, to do something unrelated to learning, such as browsing unrelated to learning pages, which leads to poor learning results. The fourth category can hardly be called learners. They hardly log on to the MOOC platform and do not conduct effective learning behaviours, so they cannot be called learners. The participation in the discussion area can be divided into two kinds: one is to initiate a topic, the other is to comment on a topic. However, whether as the initiator or commentator of the topic, the content released by college students follows two basic principles, namely, the theme principle of participating content and the topic-related principle of

participating content. The thematic principle of participating in the content is that the contents of university students' participation must contain certain themes. The topic-related principle of participation content refers to that the content of 'participation must be based on topic titles, that is, whether for the description of the depth or breadth of topic content, the description of the content and topic titles must be semantically related. In the hierarchical teaching of MOOC platform, the result evaluation and process evaluation should be combined to ensure the integrity of learning process. The same interest and similar level of college students help each other, which is conducive to the improvement of learning effect.

5 Conclusions

MOOC is an open learning platform and the learning behaviours of college students of MOOC are completely autonomous. There are significant differences in learning behaviours among different learners. We should help teachers to identify the different types through data mining to optimize the teaching effect. The main conclusions are as follows:

- (1) Cluster analysis can be used to mine the learning behaviours of college students on MOOC platform, and K-Means clustering can be selected.
- (2) The empirical study of college students' learning behaviours mainly includes 4 steps, which are indexes selection, data collection, clustering analysis and result analysis.
- (3) The indexes of college students' learning behaviours can be set as time of online, time of watching

videos, download number of learning materials, posting number in forum and finishing ratio of homework.

(4) We can divide the college students on MOOC platform into four categories according to the learning behaviours data and adopt different types of teaching methods for different types of students.

References

1. De Barba P G, Kennedy G E, Ainley M D. The role of students' motivation and participation in predicting performance in a MOOC[J]. *Journal of Computer Assisted Learning*, 2016, 32(3): 218-231.
2. Littlejohn A, Hood N, Milligan C, et al. Learning in MOOCs: Motivations and self-regulated learning in MOOCs[J]. *The Internet and Higher Education*, 2016, 29: 40-48.
3. Drachsler H, Kalz M. The MOOC and learning analytics innovation cycle (MOLAC): a reflective summary of ongoing research and its challenges[J]. *Journal of Computer Assisted Learning*, 2016, 32(3): 281-290.
4. Woodgate A, Macleod H, Scott A M, et al. Differences in online study behaviour between sub-populations of mooc learners[J]. *Educación XXI: revista de la Facultad de Educación*, 2015, 18(2): 147-163.
5. Chang R I, Hung Y H, Lin C F. Survey of learning experiences and influence of learning style preferences on user intentions regarding MOOCs[J]. *British Journal of Educational Technology*, 2015, 46(3): 528-541.
6. Rhoad - Drogalis A, Justice L M, Sawyer B E, et al. Teacher - child relationships and classroom - learning behaviours of children with developmental language disorders[J]. *International journal of language & communication disorders*, 2018, 53(2): 324-338.
7. Guo K, Youssef G J, Dawson A, et al. A psychometric validation study of the Impulsive-Compulsive Behaviours Checklist: a transdiagnostic tool for addictive and compulsive behaviours[J]. *Addictive behaviors*, 2017, 67: 26-33.
8. Fisk C, Dodd A L, Collins A. Response styles, bipolar risk, and mood in students: The Behaviours Checklist[J]. *Psychology and Psychotherapy: Theory, Research and Practice*, 2015, 88(4): 412-426.
9. Cadiz D M, Truxillo D M, O'Neill C. Common risky behaviours checklist: A tool to assist nurse supervisors to assess unsafe practice[J]. *Journal of nursing management*, 2015, 23(6): 794-802.
10. BO W V, Fu M. How is Learning Motivation Shaped Under Different Contexts: An Ethnographic Study in The Changes of Adult Learner's Motivational Beliefs and Behaviours within a Foreign Language Course[J]. *Frontiers in Psychology*, 2018, 9: 1603.
11. Brown K M, Diplock K J, Majowicz S E. The environment in which behaviours are learned: a pilot assessment of high school teaching kitchens as food safety learning environments in Ontario[J]. *Environmental Health Review*, 2016, 59(3): 88-95.
12. Boon A, Vangrieken K, Dochy F. Team creativity versus team learning: transcending conceptual boundaries to inspire future framework building[J]. *Human Resource Development International*, 2016, 19(1): 67-90.