

Behavioral Patterns of Completers in Massive Open Online Courses (MOOCs): The Use of Learning Analytics to Reveal Student Categories

Changsheng Chen¹, Jingyun Long^{2,*}, Junxiao Liu³, Zongjun Wang¹, Minglei Shan¹,
 Yuming Dou¹

¹*Shandong Youth University of Political Science, Jinan 250103, China;*

²*Long Ao School of Jinan, Jinan 250015, China*

³*Library, Shandong Normal University, Jinan 250014, China;*

**Corresponding author*

ABSTRACT

The learning energy of MOOC completers has important reference value for future learners. Existing research focuses on the behavioral performance of dropouts and participants, but ignores the mining of completer behavior patterns. In this paper, Kmeans clustering method, descriptive statistics, one-way analysis of variance, and chi-square test were used to systematically study the behavioral characteristics and academic performance of 1,388 MOOC completers. The results show that there are significant differences in resource and task preferences, effort levels, and so on, and learners can be divided into *hard-working harvesters* and *punch-in participants*; there is no significant difference in demographic characteristics. The article proposes strategies for improving teaching and promoting ordinary learners to improve learner behavior patterns and learning efficiency.

Keywords: *MOOC, Completer, Learning analysis, Kmeans, Behavior pattern*

1. INTRODUCTION

At present, MOOC is regarded as an important basic platform for building a lifelong education system in China with its advantages such as expanding teaching time and space, enhancing teaching attractiveness, stimulating learners' enthusiasm for learning, and expanding the benefits of quality education resources. The Completer is a learner who participates in the course and obtains the final assessment results [1]. Efficient learning can not only exert the potential value of the curriculum, but also stimulate learners' motivation for learning, which has an important impact on strengthening their specific learning styles and continued participation intentions and behaviors [2]. Therefore, digging out behavior patterns that effectively promote learning from the learning experience of MOOC completers is of great significance for adult learners to relieve learning pressure, improve learning experience, and ensure good learning results.

Research shows that there are significant differences in resource preferences, time management, and effort adjustment between the completers and dropouts, and the behavior patterns of completers can provide reference for other learners. At present, the highly open and autonomous online environment formed by MOOC has brought additional pressure on learners (cognitive trek, cognitive overload, insufficient motivation, etc.), making learners have to face different time management and efforts to

adjust and other new challenges [3]. However, many learners, especially adult learners, often have difficulty in meeting the requirements of self-management, resulting in low learning efficiency and high dropouts. This is one of the problems that must be faced in the development of MOOCs. Therefore, how to dig out the effective time management and hard adjustment experience from the learning experience of the MOOC completer is a problem that the course team and platform need to attention. Therefore, this study combines behavioral performance, learning outcomes, and demographic characteristics to analyze the behavioral patterns of MOOC completers, summarize the useful experience of MOOC learning, and provide a reference for improving MOOC teaching and online learning.

2. LITERATURE REVIEW

The analysis of learning behaviors provides an effective means to identify different groups of learners and their unique behavior patterns, which helps researchers understand the process of learner knowledge creation, knowledge sharing and knowledge construction [4]. However, previous studies have paid far more attention to large-scale dropouts than learners who completed the course. In the study of learner behavior patterns, most researchers will use the clustering method. Xiong conducted cluster analysis on the course data of 355

learners and summarized four types of learners in the course, whose grades were significantly stratified. Gifted learners and regular learners who completed the course generally attached great importance to the test link, with few procrastination behaviors, and achieved good results [5]. Tong studied the knowledge sharing behavior of the online knowledge community, explored the relationship between the interaction behavior of the online community and the learning results, and found that highly interactive people are often successful learners [6].

Many researchers use cluster analysis as an effective way to better understand the demographic characteristics of learners, curriculum behaviors and learner needs. De Barba studied the conversation behaviors of 9272 MOOC learners (Session Behaviour), formed learner's conversation pattern classification and conversation activity classification, and explored the results of the two classifications with learner participation, learning achievement, and the relationship between self-regulation skills [7]. Khalil and Ebner studied the learners of Graz University in Austria, and divided the learners into four types: dropouts, outstanding, gamers, and socializers, and found that the behavioral rules of excellent and socializers are of great reference value [8]. Poellhuber comprehensively used clustering algorithms and survival analysis to analyze learner participation behaviors, and found five learner types of viewers, self-evaluators, serious readers, active independent readers, and active social readers. Promoting effect of interactive behavior and self-reflection on learning results [9]. In recent years, some researchers have viewed the online learning process and ecology as a production system. Galbraith believe that the learner's course study time and the number of activities can be considered as an input, and the course results or other academic achievements can be considered as an output [10]. Learning process is a production system. In this system, the technical efficiency mainly comes from the learner's learning technology, including learning strategies, learning methods, and their combination; the scale efficiency mainly comes from the learner's effort, including the participation and persistence of online learning, which requires effort regulation comes into play. Analyze the behavior patterns of the learner groups who have completed the course from multiple dimensions, can deeply understand their behavior patterns and self-management characteristics, and provide valuable learning references for learners, especially inefficient learners. As a result, analysis of behavior patterns,

understanding of learning behaviors and motivations, reconstruction of learning environment, and improvement of online learning experience are of great significance and application value.

3. RESEARCH QUESTIONS

The purpose of this study is to summarize the behavioral patterns and group characteristics of the completers in MOOC, to mine the preferences of different groups on learning resources and tasks, and the time investment of key behaviors. Based on this, this paper raises the following two research questions:

Question 1: what are the characteristics of mooc learners' learning performance and learning results? and what are the categories of completers?

Question 2: what are the characteristics of mooc completers' resource and task preferences and time commitment? And what are the differences between groups?

4. RESEARCH METHODOLOGY

First, the system click stream encoding rules were used to encode the research data, and the validity of the encoding was checked. Some studies believe that video lectures, forums, and assessments are basic learning resources for MOOCs. Analyzing the activities of learners in these components can focus on their behavior during the learning process [11]. Therefore, this article selects the frequency of 5 variables: micro-video, courseware, online quiz, forum participation, and forum posting. The time consuming of micro-video, courseware, online quiz, and forum participation is 4 variables. As an input indicator for online learning. Three variables, process score, final score and forum credit., were selected as output indicators, as shown in Table 1. Then, based on the data preprocessing, Kmeans clustering algorithm was used to cluster the behavior and performance characteristics of the research objects. Based on the analysis results and the existing research conclusions, the behavior types of MOOC learners are defined, and the characteristics of different types of learners are described. Finally, suggestions for teaching and learning are put forward based on qualitative and quantitative research.

Table 1 The encoding scheme of variables

Coded	Variable	Explain
FOR	Frequency of forums	Cumulative times of forums participation
QUI	Frequency of quizzes	Cumulative times of online quizzes
VID	Frequency of browsing micro-videos	Cumulative times of browsing micro-videos
DOC	Frequency of browsing courseware	Cumulative times of browsing text in the chapter
REP	Frequency of forum	Cumulative times of posts and reply sent by learners in the forum

	reply	
FOR_T	Time consuming of forums	Time consuming of the forums(minutes)
QUI_T	Time consuming of quizzes	Time consuming of the quizzes(minutes)
VID_T	Time consuming of micro-videos	Time consuming of the micro-videos (minutes)
DOC_T	Time consuming of coursewares	Time consuming of the coursewares (minutes)
PRO_S	Process score	The score in the process evaluation such as quizzes, assignments, and forums
FIN_S	Final score	The learner's score on the final exam
VOT_S	Forum credit.	Total number of honors received by posting

5. MOOC PLATFORM AND COURSE OVERVIEW

5.1. The MOOC-Platform

The data of this research comes from the platform of China University MOOC (<https://www.icourse163.org>). The platform supports synchronous and asynchronous multiple interactive learning models, providing diversified services for learners who want to improve their academic level. Courses are generally published weekly, with learners participating in the main modules of the study including micro-videos, quizzes, assignments, and forums. Micro-videos are generally developed by course teams or refer to network resources; evaluation consist of quizzes and unit tests, and each unit is set with quiz questions and scores; assignments is an intensive course study, and is generally completed by publishing task requirements online and offline after submitting online, the assignments will be estimated by the peers; the forum includes the Q & A forum and the theme forum. The number of forum postings will be used as the basis for quantitative scoring. In addition, the course also provides information modules such as course introductions and announcements for learners to browse.

5.2. Course Overview and Demographic Analysis

The research object of this article is the learners who take the S course of MOOC in the platform. The course has been running for 7 consecutive sessions, with a total of 32,244 registered students but 1,388 completed course assessments, with a completion rate of 4.3%. The research data is mainly the learner clickstream log records and performance tables recorded by the platform, including three data tables including registration data tables, behavior log tables, and performance summary tables. There are 563795 behavior records and 406.2 per capita. From the perspective of gender structure, 631 were male

(45.5%), 504 were female (36.3%), and 253 were missing in gender (18.2%). In terms of age structure, 27 people (1.9%), 793 people (57.1%) between the ages of 25 and over, 568 people (40.9%) were missing from the age term, and the average age was 42.3 years after the deletion was excluded. Most are working adults. In terms of performance distribution, 748 (53.9%) failed, 310 (22.3%) passed, and 330 (23.8%) excellent.

6. CLUSTERING AND CASE STUDIES

6.1. Clustering Process

Cluster analysis is a data mining method for unsupervised learning. This method can find data points that are naturally combined together and divide the data set into a set of clusters. So far, Kmeans is the most commonly used clustering algorithm in scientific and industrial applications because it is scalable, easy to implement, fast to converge, and adaptable to sparse data [12]. Given that Kmeans algorithm is scalable, easy to implement, fast to converge, and adaptable to sparse data, this study uses Kmeans algorithm to perform cluster analysis of learners' online behaviors to reveal the learning behavior patterns and group characteristics of MOOC learners. The process of Kmeans clustering mainly includes three steps: the first step is to determine the number of clusters K; the second step is to calculate the similarity and classify each sample; the third step is to verify the clustering effect. This research uses R language and its development environment Rstudio to run the clustering algorithm.

First, the hopkins value ($H = 0.052 < 0.5$) was calculated using the *get_clust_tendency* method in the *factoextra* package in R language, indicating that the research indicators are highly polymerizable. Subsequently, the *clusGap* method in the cluster package was used to determine the optimal number of clusters to be 2. Figure 2 reports optimal number of clusters. Then, using the kmeans method, the number of clusters is set to 2, the number of iterations is 1000, and the algorithm is Hartigan-Wong to perform the clustering operation. This

article uses the Silhouette coefficient to verify the rationality of the clustering results (Rousseeuw, 1987). If the score exceeds 0.5, the clustering effect is considered "good"; if the score is between 0.2 and 0.5, the clustering effect is considered "Good"; if the score is less than 0.2, the clustering effect is considered "poor" [13]. The results

show that the silhouette coefficient is 0.32, which indicates that the clustering effect of this paper is good, and it can effectively extract the behavior characteristics and the differences of MOOC completers, and they can be divided into two clusters. Figure 1 reports cluster scatter plot.

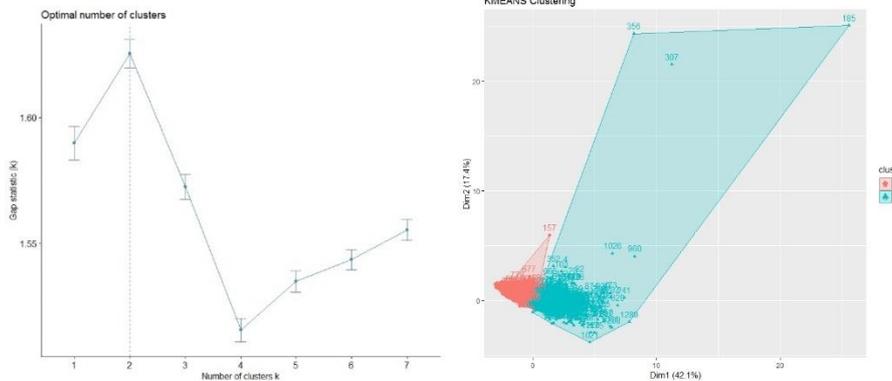


Figure 1 Optimal number of clusters (left) and cluster scatter plot (right)

6.2. Case Studies

To assess the differences between groups of behavioral inputs and outputs, descriptive statistics and one-way analysis of variance were performed based on the types of clustering variables. First, Levene test was performed on the study variables using SPSS20 software. The results showed that the overall variances of the index variables were not equal ($p < 0.05$), which could not meet the

homogeneity of variances. Therefore, in the one-way analysis of variance, the Brown-Forsythe program was selected and Dunnett's T3 algorithm was used for post hoc testing. Use the effect amount (η^2) to estimate the percentage of the variance of the dependent variable explained by the independent variable. When the effect value is [0.01-0.06], it has a small effect. When the effect value is greater than 0.14, it has a large effect [14]. Summary of the clustering results is presented in Table 2.

Table 2 Cluster analysis results and intergroup effect size

	Sample		Cluster1		Cluster2			
	1388 (100%)		593 (42.7%)		795 (57.3%)			
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>F</i>	η^2
FOR	25.62	94.75	47.65	138.61	9.18	26.90	58.28**	0.04
QUI	11.44	8.24	16.98	7.71	7.30	5.84	707.51**	0.34
VID	47.50	41.83	82.16	36.19	21.65	22.67	1456.73**	0.51
DOC	88.17	92.25	160.01	94.07	34.58	39.21	1146.45**	0.45
REP	18.43	19.87	31.55	20.40	8.64	12.44	670.06**	0.33
FOR_T	56.64	257.56	109.35	380.65	17.32	64.99	44.73**	0.03
QUI_T	12.42	25.84	20.26	32.71	6.57	16.98	102.39**	0.07
VID_T	187.65	210.34	312.50	236.47	94.53	123.58	494.50**	0.26
DOC_T	147.38	211.20	277.13	260.36	50.60	73.65	543.64**	0.28
PRO_S	15.81	9.98	23.10	6.65	10.37	8.47	916.43**	0.40
FIN_S	47.31	32.52	70.43	22.19	30.06	27.98	840.73**	0.38
VOT_S	22.01	29.95	41.31	35.10	7.62	12.76	622.73**	0.31
Gender: $\chi^2=3.49, p=0.06$								
	Male	Female	Male	Female	Male	Female		
	631	504	333	217	298	287		

	45.5%	36.3%	56.2%	36.6%	37.5%	36.1%	
Age: Brown-Forsyth: $F=3.77, p=0.53$							
	<25	>=25	<25	>=25	<25	>=25	
	27	793	4	402	23	391	
	1.9%	57.1%	0.7%	67.8%	2.9%	49.2%	

* $p<0.05$, ** $p<0.01$, *** $p<0.001$

There was a significant difference in each variable between the two groups of learners ($p<0.01$), and in addition to the small number of participations in the number of forums (0.04) and the duration of the forums (0.03), the other variables had a medium effect amount (0.07 -0.51), indicating that the clustering variables can well explain the significant differences between the two groups of learners. It is worth noting that although the categorical variable can explain the differences in learners' participation in the forum to a certain extent, the proportion of interpretation is limited, which may require a comprehensive analysis in combination with other input indicators. Based on the above analysis, this article defines the first type of learners as *hard-working harvesters* and the second type of learners as *punch-in participants*.

Case1: *hard-working harvesters*

This type of learners account for 42.7%, and their behavioral investment in various resources and tasks of the course is outstanding. The frequency and time consuming of their behaviors are very high. It can be considered that they are more motivated to learn and are willing to learn through continuous hard work. they have a tendency to "learn well".

Case2: *punch-in participants*

This type of learners accounted for 57.3%, and their behavioral investment in the resources and tasks of the course was doubled, and the frequency and cumulative duration of their behaviors were relatively low. They study lessons with a tendency to try and experience.

7. CLUSTER ANALYSIS

Further analyze the differences in learning demographics, resource preferences, and effort between the two groups of learners in order to select the best behavioral patterns and effective learning patterns. As shown in Table 1, in terms of demographic characteristics, the chi-square test of the gender variables of the two groups of learners failed to reach statistical significance, and the univariate analysis of variance for the age variables of the two groups of learners failed. Dunnett's T3 test for homogeneity of variance. Therefore, it can be considered that there is no difference in gender and age structure between the two groups of learners. Research shows that the frequency of resources and tasks can represent learners' preferences and access to resources,

while the time consuming of behaviors can represent the learners' energy investment and effort tendency.

From the perspective of resource preference and effort, the frequency of hard-working harvesters participating in forums, browsing micro-videos and courseware is significantly higher than that of punch-in participants. Figure 2 reports the behavior frequency, time consuming and relationship of the two types of learners. The multiples of the two are obviously related. Among them, participating in forums, browsing micro-videos and courseware, the difference between the two types of forum posts and other frequencies reached more than 3.65times, and the frequency of participating in the test was 2.33 times. This shows that hard-working harvesters are a group of serious readers who are interested in most learning resources and tasks. At the same time, hard-working harvesters' time investment in the above-mentioned related resources and tasks is also significantly higher than that of punch-in participants, and the proportion of the two types has reached 3.08 ~ 6.31 times, and the multiples of time investment in participating in forums, browsing micro-videos and courseware higher than frequency. It is worth noting that the difference in participating in the forum reached 6.31 times, but the length of the courseware was only 3.08 times, which was lower than the frequency index (3.79 times). This shows that hard-working harvesters will invest more energy in learning, and multi-behavior distribution indicates that they can self-adjust the learning time, but they seem to be more efficient when participating in the test, and do not need to invest too much energy. In addition, in using resources directly related to knowledge, they browsing the most of videos and coursewares. Previous studies have often replaced the test results with test results, which has been criticized by the academic community. In this study, the three indicators of the selection process assessment, final exam and forum praise were used to quantify the learning results, which can better reflect the learner's harvest and learning efficiency in multiple dimensions. As shown in Figure 2, compared with the punch-in participants, the proportion of hard-working harvesters in each performance average reached 2.2 to 5.4 times, of which the forum was praised 5.4 times. This shows that the hard-working harvesters have gained a lot through unremitting investment. On the one hand, they have achieved outstanding results. On the other hand, from the social network, their active participation in discussions has brought a lot of peer recognition to themselves.

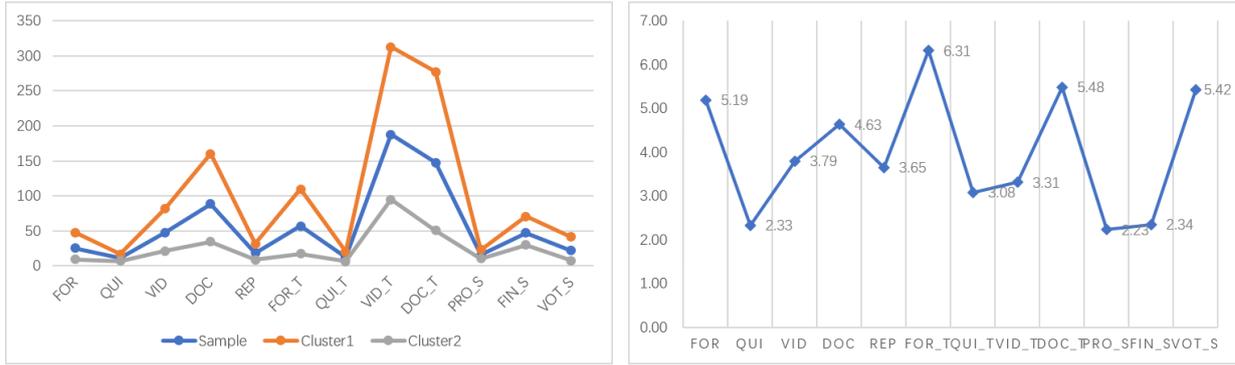


Figure 2 The behavior frequency, time consuming and relationship of the two types of learners

8. DISCUSSION

Previous studies on online courses paid more attention to dropouts and their behavior patterns, but ignored effective investigations of the relationship between the behavior patterns and learning outcomes of learners who completed the course. Studies have shown that learning analytics can stimulate the potential value of online courses, thereby helping learners to make better decisions during the learning process. In order to explore the behavioral patterns and group differences of the learners who completed the MOOCs, this study summarized the learner's behavior patterns through the Kmeans clustering algorithm. The results found that: 1) there were significant differences in the behavioral patterns of MOOC learners. The performance characteristics of the learners can be divided into two groups: hard-working harvesters and punch-in participants group; 2) there are differences between groups in resource and task preferences and time investment of MOOC learners.

In terms of behavior patterns, de Barba et al. confirms that successful learners study lessons more frequently and for longer throughout the course [3]. Hard-working harvesters are representatives of successful learners. They frequently access course resources and tasks during their studies, and each stay for a long time. In contrast, punch-in participants have more learning behaviors in the course, but their time investment in resources and tasks is far less than hard-working harvesters. This low-immersion behavior affects their learning results. This is similar to the serious readers proposed by Poellhuber et al., That is, they will browse more course materials, but the dwell time of each behavior is limited [9]. Therefore, teachers can classify and manage large-scale students in a timely manner during the teaching process according to the group differences of learners in the MOOC, and track and identify their learning status and learning needs in time, in order to learn from the learning methods, the use of cognitive tools, and time investment. Provide learning services from management perspectives to increase their participation.

In terms of resource and task preferences and effort level, the hard-working harvesters' frequency and time

investment in participating in forums, browsing micro-videos and courseware are significantly higher than those who have punched experience, and the multiples have a clear relationship. This is in line with the findings of Broadbent, et al. [15]. To keep learners in an unstructured environment, learners need advanced time management skills. The research in this paper confirms that effective learning broadly learns learning resources and tasks, which is in line with the viewpoint of behaviorist theory. In addition, hard-working harvesters will invest more energy in learning, and often can effectively adjust the learning time and learning performance. They are also very efficient in participating in the quizzes and do not need to invest too much energy, which may be related to fewer repeated submissions. Therefore, teachers should improve the construction of curriculum resources, but also need to strengthen the learner's time management energy saving and efforts to adjust the ability to cultivate, in order to improve learning performance and learning efficiency.

In this study, three indicators, including process assessment, final exam, and forum praise, were selected to quantify the learning results, which can better reflect the learner's harvest and learning efficiency in multiple dimensions. It is worth noting that there is a significant difference in the frequency of praise for the two groups of learners. The hard-working harvesters have gained a lot through unremitting investment. On the one hand, they have outstanding achievements. On the other hand, from the social network, they actively participate in the discussion. Bring a lot of peer recognition to themselves. This is consistent with the research by Liu et al. [16], they believe forum discourse behavior can promote learner thinking expression and knowledge processing, and high-level cognitive behaviors mostly appear in the interactive discourse of active participants. Therefore, the team in MOOC should arrange tutors to lead discussions and answer questions in a timely manner. Not only should the needs of ordinary learners be met, but also open-ended question discussions should be introduced, and the problem-oriented motivate learners to explore their consciousness and critical thinking and higher order thinking skills.

9. CONCLUSION

The main contributions of this research are: 1) the construction of a behavioral pattern classification scheme for learners who completed the course, which can effectively classify learners' input-output variables and demographic variables; The study of the correlation between learning behavior patterns and learning results provides a new idea and method, and provides empirical support for improving learners' behavior patterns and learning effectiveness. The limitations of this study are mainly as follows: First, the research data comes from a single course, so the analysis results may limit the interpretability of relevant situations in other courses. Second, the study lacks measures of learner learning experience, information literacy, and metacognitive levels, which may lead to biased interpretation of results. Finally, the coding mechanism for learning behavior needs to be improved. Therefore, the future research will improve the accuracy of the learning behavior analysis and the timeliness of intervention by enriching the research data and constructing comprehensive indicators.

ACKNOWLEDGEMENT

The work was financially supported by the project of Shandong province higher educational research program(J18RA144), special project of Shandong Youth University of Political Science (XXPY18064, JG201814), research project of Shandong Youth University of Political Science (2015XGY05, SQ2016QN01), undergraduate teaching reform research project of Shandong province undergraduate teaching reform research project (Z2016M102), project of Shandong province higher educational science and technology program (J15LN15).

REFERENCES

- [1] Ramesh, Arti, et al. Uncovering hidden engagement patterns for predicting learner performance in MOOCs. Proceedings of the first ACM conference on Learning@ scale conference. 2014.
- [2] Kolb, David A. *Experiential learning: Experience as the source of learning and development*. FT press, 2014.
- [3] de Barba, Paula G., et al. The importance and meaning of session behaviour in a MOOC. *Computers & Education* 146 (2020): 103772.
- [4] Yang xianmin, wang huaibo, li jihong. Application of lag sequence analysis in the analysis of learning behavior [J]. *China Educational Technology* 2016(2):17-23(in Chinese)
- [5] Xiong lu-ying, guo xing-jun, jiang qi, zheng qin-hua. study on learner academic procrastination based on online behavioral data. *e-Education Research* 12 (2019): 29-35. (in Chinese)
- [6] Tong lili, li ronglu, yan qiang. identification model of opinion leaders in online knowledge communities. *China Educational Technology* 2019 03 (2019): 97-103. (in Chinese)
- [7] de Barba, Paula G., et al. The importance and meaning of session behaviour in a MOOC. *Computers & Education* 146 (2020): 103772.
- [8] Khalil, Mohammad, and Martin Ebner. Clustering patterns of engagement in Massive Open Online Courses (MOOCs): the use of learning analytics to reveal student categories. *Journal of Computing in Higher Education* 29.1 (2017): 114-132.
- [9] Poellhuber, Bruno, Normand Roy, and Ibtihel Bouchoucha. Understanding Participant's Behaviour in Massively Open Online Courses. *International Review of Research in Open and Distributed Learning* 20.1 (2019).
- [10] Galbraith, Craig S., and Gregory B. Merrill. Academic performance and burnout: An efficient frontier analysis of resource use efficiency among employed university students. *Journal of further and Higher Education* 39.2 (2015): 255-277.
- [11] Kahan, Tali, Tal Soffer, and Rafi Nachmias. Types of participant behavior in a massive open online course. *International Review of Research in Open and Distributed Learning: IRRODL* 18.6 (2017): 1-18.
- [12] Kaufman, Leonard, and Peter J. Rousseeuw. *Finding groups in data: an introduction to cluster analysis*. Vol. 344. John Wiley & Sons, 2009.
- [13] Oyelade, O. J., O. O. Oladipupo, and I. C. Obagbuwa. Application of k Means Clustering algorithm for prediction of Students Academic Performance. *arXiv preprint arXiv:1002.2425* (2010).
- [14] Cohen, Mark A. Some new evidence on the seriousness of crime. *Criminology* 26.2 (1988): 343-353.

[15] Broadbent, Jaclyn. Comparing online and blended learner's self-regulated learning strategies and academic performance. *The Internet and Higher Education* 33 (2017): 24-32.

[16] Liu zhi, Yang chongyang, liu sanya et al. differences in cognitive behavior and sequence patterns of SPOC learners. *Open Education Research* 25.02(2019):46-54. (in Chinese)