Research Hotspots and Trend Analysis of MOOC Learner Loyalty Based on Data Mining

Yanrong Huang[a*], Rui Wang[b*]

Corresponding author: hyanrong@whu.edu.cn[a*]; 397673667@qq.com[b*]

College of Economics & Management, Zhejiang University of Water Resources and Electric Power, 583 Xuelin Street, Hangzhou City, Zhejiang Province, China 310018[a];
School of Economics & Management, Jiangxi University of Science and Technology, No. 86 Hongqi Avenue, Ganzhou City, Jiangxi Province, China, 341000[b]

Abstract. Massive open online courses (MOOC) have been widely used in colleges and universities worldwide. During the COVID-19 epidemic, MOOC has become the primary online teaching mode in education system. The phenomenon of "high registration rate" coexisting with "low return rate" and "high dropout rate" on the MOOC platform has attracted scholars' attention to improving MOOC learner loyalty. This paper takes 596 articles of MOOC learner loyalty research from 2013 to 2023 in the Web of Science database as samples, uses the CiteSpace Knowledge graph to conduct literature mining and visual analysis, and proposes the research status, research hotspots, and trends. Research result shows that MOOC learner loyalty research has formed nine clusters, including "empirical investment," "learning experience," "online learner," "self-directed learning," "Kenyan cloud school," "building capacity," "interpretable model," "discovering MOOC learner motivation," and "preference-based group." With the development of computer technology, teaching, and learning methods innovation, interdisciplinary integration research to improve MOOC learner loyalty has become a new trend.

Keywords: MOOC; learner loyalty; trend analysis; data mining

1. Introduction

The rapid development of informatization has promoted the continuous popularization of educational information technology, and the constant transformation of educational methods has also led to significant changes in people's academic concepts [1]. With the deepening of Lifelong learning and the increasing demand for high-quality development of the whole education, MOOC (Massive Online Open Courses) has received extensive attention from the global education community[2]. In 2012, as an essential form of distance education, a large-scale open online course based on the internet, targeting the general public, emerged in a surge. Online learning provides learners with many online resources, breaking the limitations of learning time and space and meeting the diverse learning needs of learners. The number of courses and online learning on China's MOOC platform has increased yearly, especially under the impact of COVID-19 in 2020. The development of large-scale online courses reached a climax. As of 2022, the number of MOOCs in China has reached 61900, with over 370 million registered users [3]. The number of courses and online learning is on the rise.

ICMEIM 2023, September 08-10, Wuhan, People's Republic of China

DOI 10.4108/eai.8-9-2023.2340191

The complete openness and large-scale model recruitment of MOOC enable MOOC objects to low-cost access resources such as course videos, learning materials, course exercises, software tools, etc. [4]. The meticulous course design and MOOC platform's comprehensive learning support service system enable participants to achieve personalized learning according to their requirements[5]. However, user churn among online learners is particularly severe, such as low course completion rates or choosing to give up halfway. The phenomenon of "high registration rate" coexisting with "low return rate" and "high dropout rate" on the MOOC platform has caused a waste of learning resources. Many MOOC learners drop out of school, and only 10% of registered learners can complete the course [6]. Faced with the high turnover rate of MOOC platform learners, many scholars have researched the loyalty of MOOC platform learners by different methods [7, 8], which have become an essential topic for the high-quality development of MOOC platforms [9].

This article intends to use data mining methods to mine and visually analyze relevant literature on MOOC learner loyalty included in the WOS database from 2013 to 2023. It explores the research status, authors, institutions, key issues, research hotspots, and MOOC learner loyalty development trends. It provides policymakers with MOOC education strategy planning, rational allocation of university teaching resources, instructors providing course design, and researchers offering research references.

2. Research Method And Data Resources

2.1 Research Method

CiteSpace is a literature data mining and visualization software developed by Chen Chaomei's team, which integrates multiple methods such as clustering and social network analysis. Use the CiteSpace visualization function to systematically analyze domestic literature and enhance research results' scientific, intuitive, and objective nature. The visual analysis software is developed by taking the sending organization, sending author, and sending trend as the research object, taking the size of nodes and network connectivity in the visual Knowledge graph as the presentation elements, and aiming at exploring the hot topics of the subject frontier, identifying the knowledge association structure, and identifying the content of the field pioneer [10]. The research of scientific knowledge graphs takes cluster analysis and co-occurrence analysis as its research methods, metrology as its theoretical foundation, and involves the cross fields of applied mathematics, graphics, and information science [11].

The implementation of Citespace automatic clustering is based on the spectral clustering algorithm. The spectral clustering algorithm is itself an algorithm based on graph theory, so it has a natural advantage for clustering based on connectivity rather than node attributes like co-reference networks [12]. The idea of spectral clustering is to view the data as nodes in a graph, with less similar points farther apart and more similar points closer together. After forming, the undirected graph is partitioned so that the weight sums between different classes are as low as possible and the weight sums between the same class are as high as possible [13].

The graph G(V, E), V denotes a point ($v_1, v_2, \cdots v_n$) in the dataset, and E denotes the set of edges. For any two points *i* and *j* in the graph, if edges are connected between them, its weight

is $w_{ij} > 0$; if these two points are not connected, then there is $w_{ij} = 0$. Providing that the degree of each node d_i is expressed as the sum of the weights of that node and all other nodes, the diagonal matrix of the matrix can be derived by eliciting the definition of the degree matrix via the definition of:

$$D = \begin{pmatrix} d_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d_n \end{pmatrix} \tag{1}$$

Define the Laplace matrix L=D-W, where D represents the degree matrix and W represents the adjacency matrix. After forming an undirected graph, the goal of clustering is to cut the undirected graph into n unconnected subgraphs, each of which is $A_1, A_2, \cdots A_n$, where $A_1 \cap A_2 \cap \cdots \cap A_n = V$, $A_i \cap A_j = \varnothing$. Define the slice weight between any two subgraphs as:

$$W(A, B) = \sum_{i \in A, j \in B} w_{ij} \tag{2}$$

The tangent weight of the entire undirected graph is:

$$cut(C_1, C_2 \cdots C_k) = \frac{1}{2} \sum_{i=1}^{k} W(C_i, \overline{C_i}) \tag{3}$$

According to $W_{ij} = e^{\frac{-\|x_i - x_j\|^2}{2\alpha^2}}$, calculate the similarity matrix W and assign the diagonal of w to 0, then there is $w(i, j) = 0$. According to the normalized degree matrix D and the normalized Laplace matrix $L = D^{\frac{1}{2}} W D^{\frac{1}{2}}$, calculate the eigenvector of L. Place the vector with the first k eigenvalues maximized into a matrix X in columns, that is, $X = [v_1, v_2 \cdots v_k]$, normalize the matrix X to obtain Y. Perform K-mean clustering on matrix Y based on each row of data points, and the class to which row *i* belongs is the x_i original class.

2.2 Data Resources

The research object of this article is the dataset of research on MOOC learner loyalty in the Web of Science digital library. In the WOS core collection database, 12 search term combinations were created by combining "MOOC" and "massive open online course" with "dropout," "retention," "completion," "attrition," "loyalty," "satisfaction," and the time range was set from 1990 to 2023, and obtained 1167 articles, and duplicate articles and information unrelated to the topic were screened and eliminated. Since there was only one article on the study of MOOC learner loyalty before 2013, 596 articles from January 1, 2013, to June 30, 2023, were ultimately selected as the research sample.

3. Analysis Of Research Status

3.1 The Publication of Research on MOOC Learner Loyalty

The literature publication status in the MOOC learner loyalty is shown in Figure 1. From 2013 to 2022, the annual publication volume was 2 articles, 19,26,34,41,67,87,93,98,92 articles, and 36 articles were published in the first half of 2023. Through analysis, it was found that the number of publications peaked in 2021. The research in the field of intelligent governance started early. Firstly, in the initial stage (2013-2014), the number of published papers increased from 2 to 19, indicating that the loyalty of foreign scholars to MOOC learners began to attract attention during this period; Secondly, during the slow growth stage (2015-2017), the number of articles published on the loyalty of MOOC learners gradually increased, with a total of 101 articles published and an average of 33.6 articles published per year, reflecting the widespread attention paid by the academic community and society to the loyalty of MOOC learners, and the increasing research enthusiasm. Finally, during the rapid development stage (2018-2022), the annual publication volume was 92.5, indicating that this issue has become a research hotspot of widespread concern among scholars. With the increase in the number of registrations for MOOC, the issue of MOOC dropout has become a focus of attention, and research on the loyalty of MOOC learners has also become an important issue in improving the effectiveness of MOOC learning.

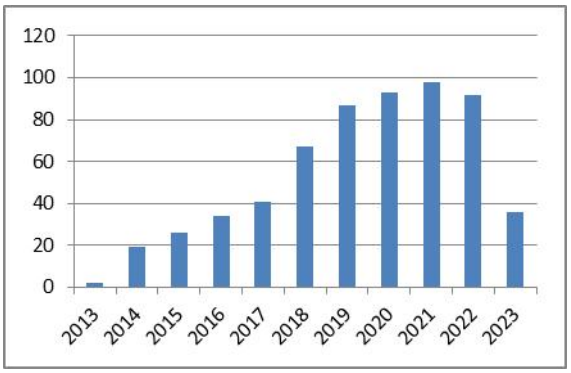

Figure 1. Annual number of documents issued

3.2 Analysis of research authors

Draw a collaborative relationship network graph with the author as the node type. The development of the research field benefits from the contributions of numerous authors, and the research of authors with many publications is representative, as shown in Figure 2. There are 307 nodes in the foreign author cooperation network, with 185 connections between nodes and a network center density of 0.0039, which is relatively sparse. Research has shown that scholars in MOOC learner loyalty have begun to take shape, but a widespread cooperative relationship has not yet been established among scholars. The author's collaboration network graph shows that Yu and Zhonggen have the highest number of publications, followed by Romero Rodriguez, Luis M and Albelbisi, and Nour Awni, who have all published five

articles. A few authors have close communication with each other, and the overall cooperation network shows a dispersed trend.

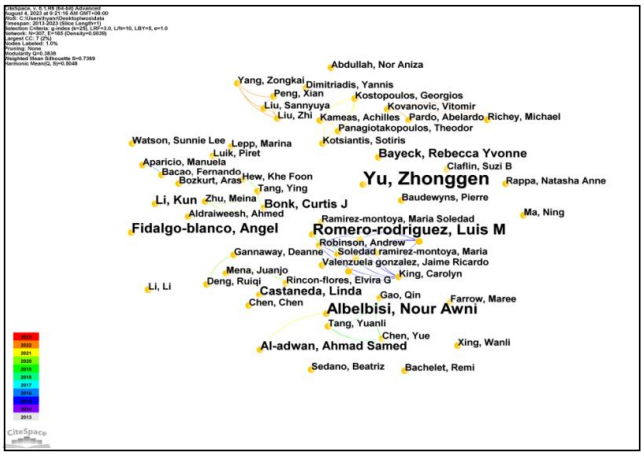

Figure 2. Analysis of research authors

3.3 Analysis of MOOC Learner Loyalty Publishing Institutions

Draw a network graph of cooperative relationships for node types from the literature publishing agency, as shown in Table 1.

Table 1. Distribution of articles publishing institutions

Country	Publication number	Country	Rank
Tecnol Monterrey	15	Mexico	1
University Tasmania	12	Australia	2
Beijing Language & Culture University	10	China	3
Cent China Normal University	9	China	4
Purdue University	7	America	5
Beijing Normal University	7	China	6
University Malaya	7	Malaysia	7
Harvard University	6	America	8
MIT	5	America	9
Duke University	5	America	10

The literature publishing institutions in MOOC learner loyalty research have not yet formed a clear central cooperation network, and only small-scale cooperation networks exist. From the perspective of publishing institutions, the top three are Tecnol Monterey, Univ Tasmania, and Beijing Language&Culture Univ, with a publication volume of 15, 12, and 10 articles, respectively, and there is no close cooperation among the three major publishing institutions. MIT, Univ Granada, and Harvard Univ's cooperative relationship is relatively close. There are 278 nodes in the collaboration network diagram of MOOC learners' loyalty research, with 147

connections between nodes and a network density of 0.0038, indicating sparse network relationships.

4. Research Hotspots And Trend Analysis

4.1 Keyword co-occurrence analysis

Using the CiteSpace word frequency analysis function, analyze the hot research areas related to MOOC learner loyalty literature, as shown in Figure 3. From the perspective of keyword networks, the number of network nodes is 344, with 1760 connections between nodes and a network center density of 0.0298. The connections between networks are relatively close. Analyzing the frequency of keywords "occurrences," "motivation," "mass open online course," "engagement, performance," "quality," etc., have a higher frequency of occurrence. The centrality of keywords such as "design," "mass open online course," "option," and "motivation" in the relevant literature of the WOS database is relatively high, indicating that are essential factors affecting the loyalty of MOOC learners. These high-frequency keywords mean that research on MOOC learner loyalty focuses on the factors that affect it, and keywords such as "technology" and "model" are also highly commented on, indicating the focus of the research.

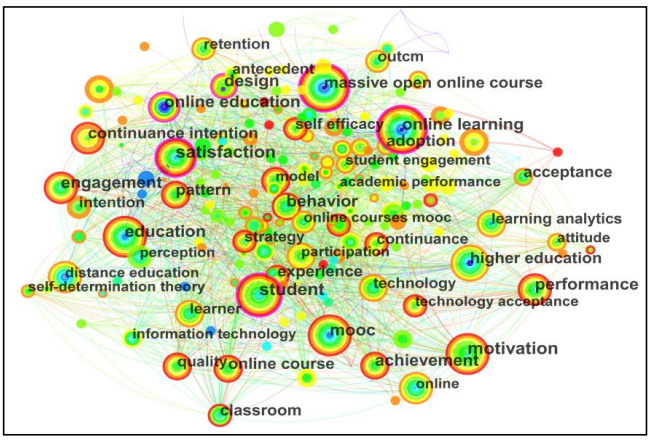

Figure 3. Keyword co-occurrence analysis chart

4.2 Keyword clustering analysis

The clustering module value Q in the keyword clustering graph represents the degree of concentration of internal connections. The larger the Q value, the more internal connections there are, and the better clustering results. It is generally believed that when Q>0.3, the clustering structure is significant. The S value in the clustering graph represents the average contour value of the cluster. It is generally believed that clustering is reasonable when S>0.5 and convincing when S>0.7[14]. The Q value in this study is 0.3838, and the S value is 0.7369, indicating that the clustering results in this study are highly scientific and practical. Nine

major clusters were obtained through cluster analysis, including "#0 empirical investment," "#1 learning experience," "#2 online learner," "#3 self-directed learning," "#4 Kenyan cloud school," "#5 building capacity," "#6 interpretable model," "#7 discovering MOOC learner motivation," and "#8 preference-based group."

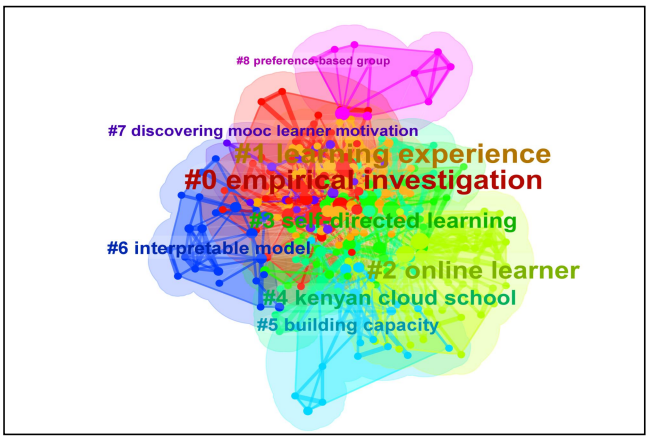

Figure 4. Keyword cluster analysis diagram

Out of nine clusters, 5 clusters conducted research on learner loyalty motivation and behavior, including "# 1 learning experience," "# 2 online learner," "# 3 self-directed learning," "# 7 discovering MOOC learner motivation," "and" # 8 preference based group, as shown in Figure 4. In addition, the use of next-generation information technology tools to research interpretable models is also one of the critical areas in the field. From the literature clustering analysis results, it can be found that studying learners' motivation, analyzing learners' preferences, and studying interpretable models are hot topics in learning MOOC learners' loyalty.

4.3 Research Trend Analysis

By using CiteSpace software to analyze the timeline of keywords, we can further understand the development and evolution process of keyword clustering. We can appreciate the support association strength of different keyword clusters by connecting keywords, as shown in Figure 5. The horizontal line represents the research topic, with time points from 2013 to 2023 from left to right, indicating changes in time. The diamond dots on the horizontal line represent the keywords under clustering. Through analysis, it was found that 'Kenyan cloud school' was the earliest clustering theme that emerged, from lifelong learning, blended learning, design, non-formal learning, and analytical education, to now comparative study computer games. The topic cluster of “discovering MOOC learner motivation” was first proposed in 2018 at the latest, with research focusing on the structural equation, academic performance, continuity, attachment nursing student, intervention, and self-directed learning.

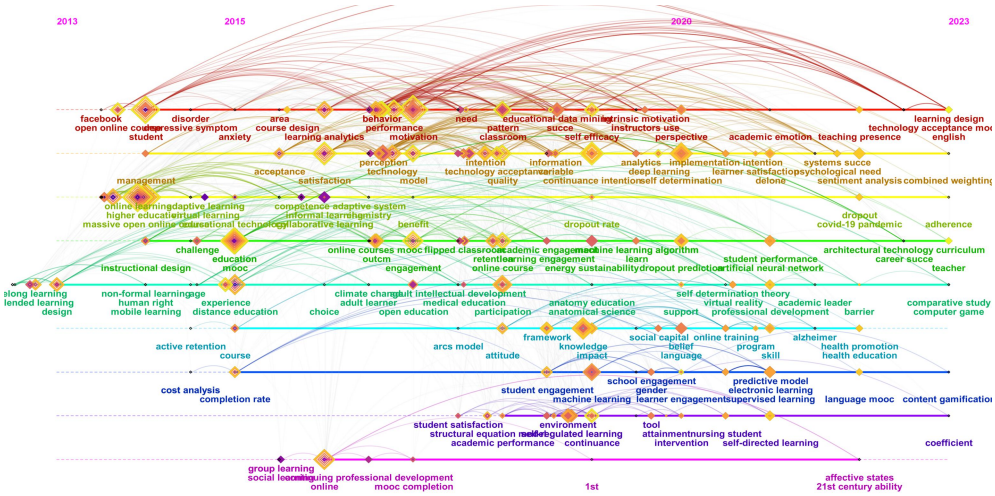

Figure 5. Trend chart of Clustering Time Line

The topic clustering of "interpretable model" first appeared in cost analysis and completion rate in 2015, and the recent focus has been on student engagement, machine learning, predictive model, supervised learning, language MOOC, and content augmentation. The research results on thematic clustering of empirical investment are the most concentrated, from the initial theme words anxiety and learning analytics to

later behavior, motivation, pattern, educational data, and to academic emotion teaching presence, technology acceptance model, indicating significant changes in the research methods of empirical investment. From the evolution of the nine considerable clustering themes, from empirical research and learning experience to explanatory models, research motivation has been found, indicating that the depth and breadth of its research field are gradually enriched.

5. Conclusion

MOOC online courses are the most innovative learning method that has attracted widespread attention in the industry and academia since their emergence, especially during the pandemic, providing high-quality education for students worldwide and helping millions complete their studies online. However, the implementation of MOOCs faces the problem of high dropout rates, and how to improve the loyalty of MOOC learners is an important issue that needs to be urgently addressed. This article conducts data mining and visualization analysis on relevant literature related to MOOC learner research from 2013 to 2023.

From the current research status of MOOC, the study of MOOC learner loyalty has gradually attracted attention from the academic community. The relevant research has undergone initial germination, slow growth, and rapid growth, with an average annual publication volume of 92.5 articles in the past three years, which has become a research hotspot. The publishing institutions and scholars have begun to take shape, but there still needs to be more extensive

cooperative relationships between institutions and scholars, resulting in sparse connections. Cross-regional and interdisciplinary research needs to be strengthened.

From the perspectives of keyword co-occurrence networks, clustering analysis, research hotspots and trends have gradually enriched the depth and breadth of their research fields. Nine major clusters were obtained through cluster analysis, including "empirical investment," "learning experience," "online learner," "self-directed learning," "Kenyan cloud school," "building capacity," "interpretable model," "discovering MOOC learner motivation," and "preference-based group." The literature analysis results show that in addition to the improvement of traditional teaching models, there is also attention to innovation in teaching and learning methods, as well as many issues such as resource construction. With the development of computer technology and teaching theory, the development direction of MOOC research is constantly updating, and interdisciplinary integration is one of the directions for studying interpretable models of learners' motivation and preferences.

Acknowledgments. The research supported by Key Project of Zhejiang Higher Education Association under grant KT2022001; General topic of the first batch of teaching reform research projects in the 14th Five Year Plan of Zhejiang University of Water Resources and Electric Power, under grant JK202205; Zhejiang Provincial Education Science Planning Project under grant 2021SCG328. General Project of Education Science Planning in Jiangxi Province under grant 21YB093.

References

[1] Y. H. Zheng, S. H. Yang, J. Y. Wang. “The current situation, direction fields, and reflection prospects of china's mooc research,” Informationization of Education in China, vol. 29, pp. 26-34, 2023.

[2] Y. R. Huang, R. Wang. “Analysis of influencing factors of mooc learners' loyalty based on online review text mining,” The 2021 IEEE International Conference on Electronic Information Engineering and Computer Science, IEEE, China, pp. 693-698, 2021.

[3] China Youth Daily, The number of MOOCs in China reaches 61900, with over 370 million registered users, 2022.

[4] X. B. Han, W. S. Ge, Q, Zhou, et al.. “A comparative study of mooc platform and typical online teaching platform,” China Educational Technology, vol.8, pp.61-68, 2014.

[5] J. Fan, Y. C. Jiang, Y. Z. Liu, et al.. “Interpretable MOOC recommendation: a multi-attention network for personalized learning behavior analysis,” Internet Research: Electronic Networking Applications and Policy , vol. 2, pp.588-605, 2022.

[6] H. S. Yuan, X. Liu. “Current situation and common issues of mooc practice in chinese universities: A report on MOOC practice in Chinese universities,” Research on Modern Distance Education, vol.4, pp. 3-12, 2014.

[7] G. Ángel, D Carabantes Alarcón. “MOOC: measuring satisfaction, loyalty, success and certification of digital education,” RIED: Revista Iberoamericana de Educación a Distancia, vol. 20, pp. 105-123, 2017.

[8] N. Z. Dina, R. T. Yunardi, A. A. Firdaus. “Utilizing text mining and feature-sentiment-pairs to support data-driven design automation massive open online course,” International Journal of Emerging Technologies in Learning, vol.16, pp.134-151, 2021.

[9] R. Wang, Y. C. Xiao, Y. R. Huang. “Analysis of online course quality elements from the dual perspective of learner attention importance,” Research on Higher Education in Heilongjiang Province, vol.41, pp. 141-148, 2023.
[10] J. X. He, X. T. Zhang, Y. C. Liu. “Research progress and future development trend of intelligent theory of social governance - Knowledge graph analysis based on CiteSpace,” Journal of Chengdu University (Social Sciences Edition), vol.208, pp.1-19, 2023.
[11] L. Chen, Y. Q. Zhang, H. Y. Yu et al.. “Visualization research on the status quo, hotspots and trends of scientific and technological achievements research in universities in China -- based on the analysis of scientific knowledge graph,” Research on Science and Technology Management, vol.17, pp.99-109, 2019.
[12] Y Chen, C.M. Chen, Z. G. Hu, etc. “CiteSpace Practical Guide to Principles and Applications of CiteSpace Analysis,” Published by Science Press, 2023.
[13] L. Song, J. Qi, M. Wu. “Research on Density Sensitive Clustering Algorithm for Non-convex Sets,” Proceedings of the 3rd International Conference on Mechatronics Engineering and Information Technology (ICMEIT 2019), 2019.
[14] J. Li, C. M. Chen, “CiteSpace technology text mining and visualization,” Capital University of Economics and Business Press, 2017.