Analysis of the influencing factors of online classes satisfaction based on text mining Take MOOC platform art education online classes as an example

Qingju Liu[1, *], Yuchen Ding[2], Pengbo Qian[3], Ruoyao Li[4], Jiahao Zhou[1]

[1]School of Art, Qingju Liu, Anhui University, Hefei, Anhui, 230601

[2]School of Business, Anhui University Hefei, Anhui, 230601

[3]School of Management, Anhui University, Hefei, Anhui, 230601

[4]School of law, Anhui University, Hefei, Anhui, 230601

*Corresponding author. Email: liuqingju0317@163.com

Abstract. The normalization of the epidemic and the convenience of online classes have led to the rapid rise of online education courses. The MOOC teaching platform represents the online course field and is popular in various disciplines. Unlike disciplines such as natural science and engineering technology, art education courses require real-time emotional expression and transmission. However, the existing technical framework is difficult to support the interactive needs of art teaching mode, and the learning effect of art students is not ideal, and the user The learning effect of art students is not ideal, and the phenomenon of low user satisfaction and low willingness to continue learning is more serious. Based on the above background, this paper uses Python to crawl the review texts of art online courses and summarizes the factors influencing students' satisfaction with art education online courses through sentiment analysis and LDA model to provide valuable references for the quality of art online education.

Keywords: User satisfaction, art online education, sentiment analysis, LDA model.

1. Introduction

With the normalization and continuation of the new pneumonia epidemic, the "stop teaching, stop learning" model of online classes is rapidly emerging. In the competition with offline institutions, the resource advantages and convenience of online classes are highlighted, and online education courses have achieved from "0" to "1"[1]. MOOC's teaching platform is a typical representative of the online course field. It provides users with systematic teaching courses of various subjects, and has been recognized by many teachers and students for their openness, socialization, and networking [2]. Unlike natural science, engineering technology, medicine, and other disciplines that teach relatively fixed concepts and theorems, art education focuses more on transmitting ideas and feelings [3]. However, the existing technical framework of the MOOC platform is difficult to support the real-time transmission of feelings in art teaching mode. MOOC's teaching platform is a typical representative of the online course field, providing users with systematic teaching courses of various subjects, which many teachers and students recognize for their openness, socialization, and networking. Different from the disciplines that teach relatively fixed concepts and theorems such as natural science, engineering technology and medicine, art education focuses more on the transmission of ideas and feelings, but the existing technical framework of the MOOC platform can hardly support the real-time transmission of feelings in art teaching mode. Therefore, the learning effect of art users may not be as ideal as expected, and there are problems such as poor user satisfaction, low willingness to continue to use, high dropout rate, etc. How to prevent user loss, improve user satisfaction, and reduce resource waste has become an urgent problem for art education, online classes. It is necessary to analyze the factors influencing satisfaction of art education online classes, and provide suggestions for deepening art education reform and expanding the possibilities of teaching.

With the continuous development of technology and its cross-fertilization, students' learning styles have undergone tremendous transformation [4]. As a new teaching mode, online classes have been

attracting much attention and popularity, among which. As an important way for students to perceive the world, art education has gradually become an important development area of China's public education system. Therefore, how to promote the development of art education based on online classes, a new educational model, is important for exploring effective methods and strategies for art perception and enhancing students' artistic intuition [5]. Although the MOOC online platform solves the problems of teachers and resources, allows repeated reproduction of part or all of the teaching content, and enables students to plan their learning time and progress independently [6]. However, some scholars believe that the inability to, on the other hand, MOOC courses all provide learner evaluation functions. However, since the learners of each course are different, even the learner evaluation results of courses on the same subject topic often lack comparability because the learners involved in the evaluation have not participated in other similar courses [7].

Based on the above background, I think studying the satisfaction influencing factors of art education online courses from the learners' perspective is necessary. Therefore, this paper first uses python crawler technology to collect the text of art education online course reviews in MOOC online platforms. Then, it uses text mining techniques such as sentiment analysis and LDA topic model to analyze the review data, study students' sentiment tendency and the distribution of subject words under different sentiment tendencies. Finally, it summarizes and summarizes the satisfaction influencing factors of art education online courses for students, which provides valuable references for improving the quality of art online education and even the whole MOOC teaching.

2. Research Methodology And Process

The path of constructing the model of factors influencing online class satisfaction is shown in Figure 1.

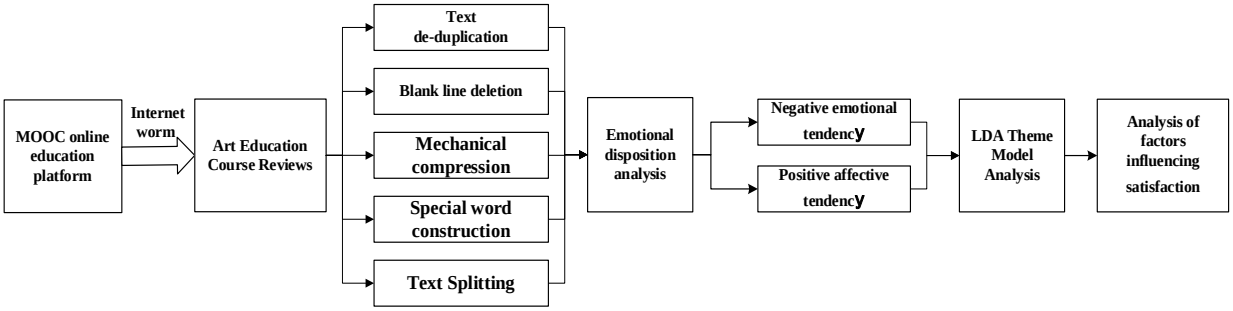

Figure 1 The path of constructing the model of influencing factors of online class satisfaction.

2.1. Data collection and processing

Python was chosen in this paper as theprimary tool for collecting and analyzing art online course review data. The course reviews were retrieved on the MOOC platform using terms related to "art education", and the data were collected from January 1, 2020, to December 31, 2021. We used fiddler to capture network packets, deserialized, parsed fields, and finally accessed the database to store the data in CSV format, and obtained a total of 9728 reviews. The data obtained based on Python contains a large number of spaces, blank lines, and redundant information with low or no value, which can interfere with sentiment analysis. Hence, the noise is reduced by text de-duplication and compression, blank line deletion, synonym, or misspelled word replacement, word separation, and deactivation. Finally, a total of 6494 comment data were obtained, and the data efficiency rate was 66.75%.

2.2. Sentiment Analysis

We use NLP (Natural Language Processing) and data mining algorithms to make sentiment judgments on text language, to grasp the computational research of text opinions and attitudes. In this paper, we call the open-source sentiment analysis document of the Baidu AI platform in python environment to divide the sentiment polarity of the text into three levels: negative, neutral and positive,

which facilitates the training of the LDA topic model for articles with different levels of sentiment tendency later, to study the difference of text topics under different sentiment tendency.

2.3. LDA Topic Model

Latent Dirichlet Allocation (LDA) is a document topic generation model widely used in semantic mining because LDA can reduce the dimensionality of text representation. Let the text of M comments be collected, there are a total of N words, and the number of text topics is K, Φ words distribution of text topics is generated by sampling from the Dirichlet distribution α, and the corresponding topic W words are generated by sampling according to the word distribution. The model continuously repeats the above process until all texts are sampled, and finally obtains the topic distribution and word distribution of each topic for each text.

Since the LDA topic model is unsupervised, the number of topics is an important input parameter of the model. In order to ensure the accuracy and reasonableness of the model construction results, this paper uses perplexity to determine the optimal number of topics for a document. The perplexity is a criterion used to evaluate the model's merit and can be used to adjust the number of topics, which is calculated as follows.

$$Perplexity(D) = \exp\{-\sum_{d=1}^{M} \lg p(w_d) / \sum_{d=1}^{M} N_d\} \quad (1)$$

w_d denotes the word, $p(w_d)$ denotes the probability of the word in the document, N_d denotes the number of documents, and D denotes the set of all words in the document. When using perplexity for evaluation, the more topics there are, the perplexity value will gradually decrease; and the more the number of topics, the more costly the LDA model calculation will be. At the same time, to avoid model overfitting, the selection of the perplexity value and the number of topics should be considered together, and the value with the smallest perplexity and the optimal number of topics should be selected as the optimal number for LDA model training.

3. Sentiment Analysis

This study identifies the processed sample data based on Python calling Baidu AI open-source platform sentiment orientation analysis tool and judges the sentiment orientation of the identified attitude words. The following figure shows that positive emotions accounted for 91.86 %, positive emotions accounted for 91.86 %, and neutral emotions accounted for 1.37 %. Negative emotions account for 6.77 %, and the general public still holds a relatively positive attitude towards quality art education. Positive and negative emotions are then divided into three levels (general, moderate, and high), as shown in Figure 2.

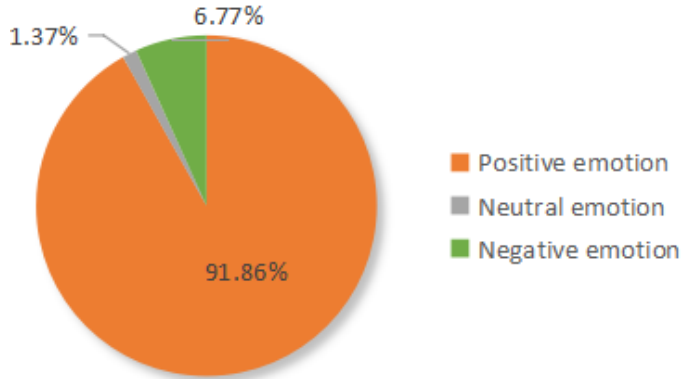

Figure 3 Web text review sentiment distribution view

In Figure 3 and Figure 4, The highest proportion of positive emotions is high, accounting for about 44.73 % of the number of positive emotions. The proportion of general and moderate emotions is close, about 30 % each. The negative emotions generally accounted for the largest proportion, about

72.88 %, while the high proportion is relatively low. The overall emotional tendency of consumers is considerable, and there is a certain degree of negative emotion, but only a few are highly negative.

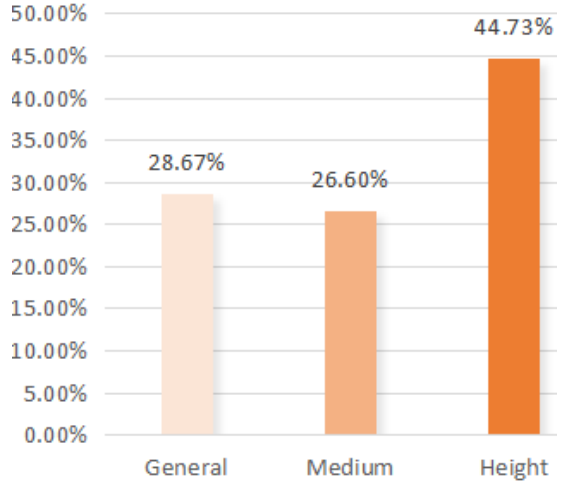

Figure 3 Positive emotion grading statistics

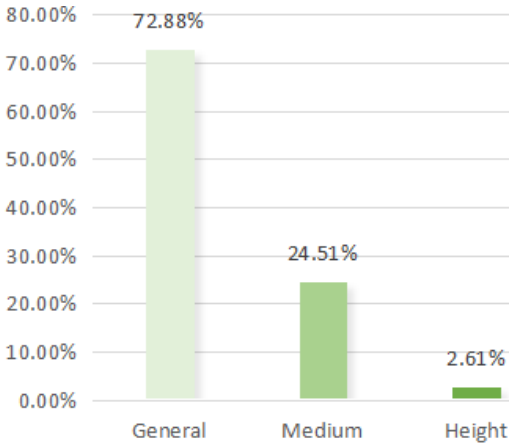

Figure 4 Negative emotion grading statistics

4. Lda Topic Model Analysis

4.1 Analysis of MOOC Users' Comments on Positive Emotional Tendency

By using the gensim package in Python, the LDA topic model is performed on the positive emotion dataset and based on the confusion formula, the confusion values are calculated for different numbers of topics in the interval of 2-10 (interval of 1), and the experimental results are shown in Figure 5, with the horizontal axis showing the number of topics and the vertical axis showing the confusion, from which it can be seen that the confusion fluctuates as the number of topics increases.

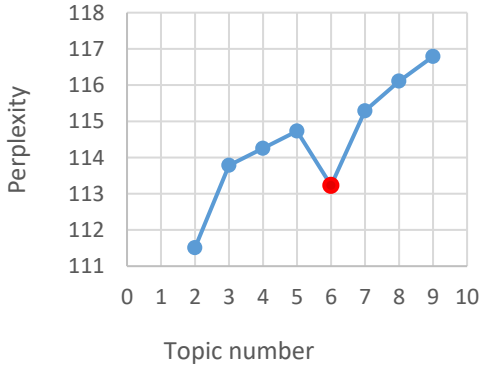

Figure 5 Theme Perplexity of Positive Emotion Sets

As shown in Figure 5, when the number of topics is 6, the inflection point of confusion degree appears, so the optimal number of topic words K is 6. Combined with the optimal number of topic words, the topic words with the highest descriptive value to the text content are screened as the key topic words, and then the text content and its keywords under each topic are summarized, as shown in Table 1.

Table 1. LDA topic model of negative emotion set

Theme 1		Theme 2		Theme 3		Theme 4		Theme 5		Theme 6	
Keywords	Weight	Keywords	Weight	Keywords	Weight	Keywords	Weight	Keywords	Weight	Keywords	Weight
Benefit greatly	0.042	Clear	0.018	Enrichment	0.082	Video	0.004	Free	0.056	Popular	0.028
Great gain	0.036	Vivid	0.015	Structure	0.024	Courseware	0.003	Cost	0.038	Interesting	0.024
Good	0.024	with the heart	0.014	Logical	0.019	Interactive	0.003	Great	0.036	Lucid	0.016
Very good	0.016	Seriously	0.010	Knowledge	0.018	Pictures	0.003	Economical	0.033	Emotional	0.010
More than expected	0.015	College teacher	0.008	Dry Goods	0.012	Titles	0.003	Very good	0.024	Fun	0.009
Enjoy	0.010	with care	0.006	Work	0.010	Beautiful	0.003	Recommend	0.018	Knowledge	0.007
Great	0.010	Very good	0.005	Practical	0.008	Pictures	0.003	Like	0.011	Thinking	0.005
Surprised	0.008	Famous teacher	0.005	Practices	0.008	Assignments	0.002	Not bad	0.006	Vision	0.005
Study	0.008	Not bad	0.004	Promotion	0.007	Exams	0.002	Spend	0.005	Attraction	0.004
Pretty good	0.007	Like	0.003	Useful	0.003	Content	0.002	Quality	0.005	Like	0.003

Table 2. LDA topic model of negative emotion set

Theme 1		Theme 2		Theme 3		Theme 4		Theme 5		Theme 6	
Keywords	Weight	Keywords	Weight	Keywords	Weight	Keywords	Weight	Keywords	Weight	Keywords	Weight
Clip case	0.023	Score	0.029	Answer	0.003	PPT	0.008	Read the manuscript	0.005	Close	0.006
Serious	0.017	Too low	0.016	Filling blanks	0.003	Video	0.006	Reading	0.004	Courses	0.004
Unsmooth	0.008	failing	0.015	Problems	0.003	Too old	0.004	Stiff	0.004	Deadline	0.004
APP	0.006	Certificate	0.012	Not found	0.003	Ugly	0.003	Stumbling	0.003	Earlier	0.003
Platform	0.006	Not shown	0.01	Original sentence	0.003	Perception	0.003	Attraction	0.003	Progress	0.003
Feedback	0.005	Achievements	0.01	Loss of points	0.003	Special effects	0.003	God out	0.003	Not updated	0.003

Softw are	0.0 04	Score	0.0 08	Unhapp y	0.0 03	Cheap	0.0 03	Bored	0.0 02	Time	0.0 03
Feedb ack	0.0 04	Unreaso nable	0.0 07	Eggless	0.0 02	PDF	0.0 02	Sound	0.0 02	No notifica tion	0.0 02
No respo nse	0.0 04	finished brushing	0.0 06	Tesk	0.0 02	Not detaile d	0.0 02	Putonghu a	0.0 02	missed	0.0 02
Waste	0.00 3	Task	0.00 5	No window	0.00 2	Interesti ng	0.00 2	Doing	0.00 2	Too long	0.00 2

In Theme 1, "benefited greatly", "great gain " and "good" are given greater weight, indicating that users are generally more satisfied with the quality of the online courses. "Surprised" and "beyond expectation" reflect that users' expectations are also an important factor affecting user satisfaction. Theme 2 is mainly about users' evaluation of the teachers of online classes, including "clear explanation", "serious", "attentive", "vivid language", etc. "professor", "famous teacher" and "teacher from a university" reflect that the fame and academic background of the online course teachers are factors that affect user satisfaction. Theme 3 is mainly related to the quality of the course, "rich", "structure", "logic" and "dry". Also, "application", "promotion" and "work" indicate that the practical nature of the course is also a factor of interest to users. Theme 4 deals with the format of the lectures, with 'videos', 'courseware', 'illustrations', 'tests', and 'assignments' reflecting users' concerns. "Assignments" reflect users' concern for a variety of learning styles. Theme 5 relates to price, as online classes on MOOC platforms are free to users, and cost greatly affects user satisfaction. Theme 6 mentions "popular", "increase knowledge", "interesting" and "cultivate emotions", indicating that Interestingness, depth, and breadth of course content all influence user choice.

4.2 Analysis of MOOC Users' Comments on Negative Emotional Tendency

As the number of topics increases, the evaluation of negative emotional tendencies fluctuates as shown in Figure 6.

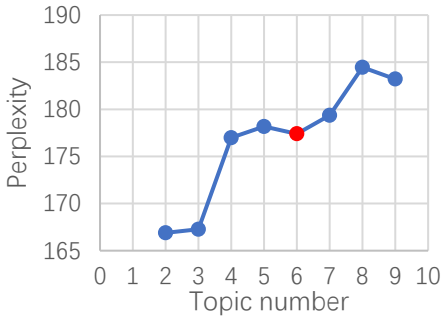

Figure 6 Theme Perplexity of Negative Emotion Sets

When the number of topics is 6, there is an inflection point in the degree of confusion, so the best number of topic words K is 6, with the first 10 words in each topic entry as the topic words. The distribution of subject words is shown in Table 2.

In theme 1, "clip case", "serious", "unsmooth" and "flashback" account for a relatively high proportion, while referring to "APP", "platform" and "software", which may be caused by the network speed and the system problems of the platform itself. Users ' "feedback " on the platform, the platform "no response ", no attention, and follow-up resolution, resulting in users' dissatisfaction. Theme 2 is mainly related to the user's performance and certification, "score ", "too low", "unqualified ", "failing ", "unreasonable " is the users ' main voice, and the users' "certificate " is also because "not shown" can not "prove" performance results, users complain that "finished", "task" performance is not good, the problem may appear in the system or teacher score above. The main content of topic 3 is "answer ", "filling blanks ", "problems " and other related aspects. The user reflects the problem "can not find",

"original sentence " and "lose points ", reflecting the separation of task and teaching content, and even the phenomenon of homework "no window ", teaching methods and platform system need to be improved. In theme 4, "PPT" and "video " occupy a large proportion. The painting style is "too old ", "very ugly ", "special effects " and "cheap ". Experience is poor from the "sense of view ". The content of "PDF " is "not detailed " and lacks "interesting ". Topic 5 repeatedly mentioned the interpreter is "read the manuscript ", "reading ", "stiff ", "stumbling ", and "voice " and "Putonghua " is not good enough, causing users to feel "attractive " is insufficient, easy to "god out " and "bored ". Theme 6 reflects that "courses ", "closing " and "deadline " are "earlier ", progress update speed is slower, "time " is "too long ", and the platform can't notify timely, causing "missed " and other issues.

5. Analysis of results and recommendations

In response to the above analysis of the factors influencing user satisfaction of MOOC art online classes, this paper puts forward the following three suggestions.

(1) In terms of teacher resources, teachers' professional competence and professional quality will directly affect the emotional experience of users. The platform should strictly control the quality of teachers and select teachers with good professionalism and rich teaching experience. In contrast, labels such as "professor" and "famous teacher" will increase the charismatic attributes of teachers. According to the perceived value theory, increasing the number of renowned teachers on the platform will increase users' perceived value and thus meet their expectations.

(2) In terms of course structure, compared with traditional science and social science online courses, art education online courses require more effective emotional links and interactive transmission. Therefore, course developers need to pay attention to establishing communication channels, such as giving feedback on course users' needs in the subsequent course content, to enhance users' sense of participation and dependence.

(3) In terms of course development, problems such as "lagging" and "flashback" lead to negative emotions among users. As there are more real-time demonstrations in online art courses, all aspects of course development need to be closely coordinated. Course recording equipment and playback systems need to ensure stability and fluency to meet users' needs for technical action learning and professional theoretical knowledge learning.

Acknowledgment

*This paper is one of the research results of the 2021 National Student Innovation and Entrepreneurship Training Program "Deha Art--Create a new era of "Internet plus" quality education" (Project No.: 202110357258S)

References

[1] Wang, Juan & Liu, Wei. (2022). A study on the factors influencing the sticky behavior of online course users. China Distance Education(03),61-67+75. doi:10.13541/j.cnki.chinade.2022.03.006.

[2] Mc Auley, M. T., Proctor, C. J., Corfe, B. M., Cuskelly, G. J., & Mooney, K. M. (2013). Nutrition research and the impact of computational systems biology. Journal of Computer Science & Systems Biology, 6(5), 271-285.

[3] Ma, Yun, Zheng, Yanlin & Shi, Huishan. (2017). A study on the current status of subject distribution of MOOC in domestic universities. Modern Distance Education(01),11-19. doi:10.13927/j.cnki.yuan.2017.0002.

[4] Lee, Zeng-Hui. (2017). Research on the construction of a mobile cloud platform for art teaching under the concept of perceptual art education. Journal of Hunan University of Science and Technology (Social Science Edition)(05),175-180. doi:10.13582/j.cnki.1672-7835.2017.05.027.

[5] Wang, Yi, Liu, Zhimin, Liu, Chundi, Jiang, Hongliang & Li, Xinxing. (2014). Design and implementation of a mobile video resource platform for art education. Modern Education Technology (10),68-72.

[6] Wang, Y.F., Qian, S.J., Zhao, J.Y. & Yan, M.X.. (2017). The application of MOOC-based flipped classroom in "literature search and scientific paper writing". Wireless Internet Technology (20),85-87.

[7] Liu, Y., Ji, J.M., Li, N., Zeng, Y., Li, J.X., Zhu, S.Q. & Kang, J.. (2021). Research on the construction of teaching quality evaluation system of MOOC courses from students' perspective: an example of MOOC courses in academic information literacy. Library Journal. (02),95-103. doi:10.13663/j.cnki.lj.2021.02.011.