Educational Administration: Theory and Practice
2024, 30(5), 10186 -10190
ISSN: 2148-2403
https://kuey.net/

Research Article

Construction Of A Sentiment Analysis Model For Chinese MOOC Comments Based On Big Data

XiaoHu*, Helmi Norman[2], Norazah Nordin[3]

[1*,2,3]Faculty of Education, Universiti Kebangsaan Malaysia, Bangi 43600, Selangor, Malaysia , : lilei20233202@163.com, : helmi.norman@ukm.edu.my, drnmn@ukm.edu.my

***Corresponding Author:** XiaoHu
Email: lilei20233202@163.com

Citation: XiaoHu, et al (2024) Construction Of A Sentiment Analysis Model For Chinese MOOC Comments Based On Big Data , *Educational Administration: Theory and Practice*, *30*(5), 10186 -10190
Doi: 10.53555/kuey.v30i5.4722

ARTICLE INFO

ABSTRACT

Chinese MOOCs, as a new educational model developed in the context of information technology, has gained good development in the era of online education. More and more users use the platform to carry out relevant course learning, and leave subjective emotional color comments in the course comment area, covering the evaluation and attitude towards the course. The above data can help MOOC platform managers and learners to obtain some valuable information. Based on this, this paper constructs a big data-based sentiment analysis model for Chinese MOOC Comments, and conducts in-depth mining of the sentiment and potential themes in the comment area, so as to help learners determine the quality of the course, and also to help MOOC platforms to better carry out construction and improvement.

Keywords: big data; Chinese MOOC Comments; sentiment analysis

1 Introduction

At present, social media has become a ubiquitous communication tool for all social classes. It makes a network of communication channels available, connecting us to people, organizations and information networks [1]. With the rapid development of emerging learning technologies, Massive Open Online Courses (MOOCs) have emerged to provide open online learning courses for a large number of students [2]. However, as a new educational model, the construction method of MOOCs is still under exploration, and the quality still varies. Comments are learners' direct feedback on the teaching effect of Chinese MOOC Courses, therefore, mining the characteristics of comments and analyzing the sentiment of comments can help catechism course designers better grasp the direction of course construction, which is of great significance to improve the quality of catechism courses.

Aiming at the characteristics of large number of Chinese MOOC Comments and mixed information, the article proposes a big data-oriented semantic analysis model for Chinese MOOC Comments. Aspect-level sentiment analysis is used to analyze and study the Chinese MOOC course comments, aiming to mine the aspect words and corresponding sentiment tendencies in the comments. In the model, Chinese MOOC course reviews are categorized into three main categories: content-related, sentiment-related and other. Aiming at the differences of different categories, the article proposes semantic features based on word classes for characterizing and classifying the comments.

2 Application of Text Analysis and Mining Based on Online Reviews of Chinese MOOC Comments

2.1 Data Sources

In order to verify the effectiveness of the model, the comment content of the comment area of four resource-sharing courses was collected from the Chinese MOOC Class platform as experimental data, and the relevant basic data are shown in Table 1.

Table 1 Comment data sources and basic information

Course	Total number of comments	Number of content-related comments	Number of emotion-related comments	Number of comments in other categories
Legal Class	490	140	212	146
Computer Class	305	131	102	79
Finance Class	521	125	207	187
Literature Class	77	32	12	35

All the above data were labeled, after training, the two data processors independently categorized the above comment data and discussed the inconsistencies in the categorization results, and finally obtained the consistent results as the manual labeling results of the data. After completing the manual classification and labeling, we used the "Psychological Analysis System for Text in the Media" to count the number of words belonging to each type of feature in each comment. Then, a random forest classification algorithm was used to train classifiers to classify the above course comments. The evaluation of the classification results is mainly based on the common evaluation indexes in text categorization: Accuracy, Precision, Recall and F-value.

2.2 Course Review Classification Results

In order to examine the effectiveness of the classification model within a single course, the group took the review data of "Legal Class" as the experimental data, and conducted classification experiments and statistics. From the civil law comments through systematic sampling, 92 comments were extracted as the test set.

From the "Legal Class" reviews, 92 reviews were sampled as test set and 400 reviews as training set, and then trained and classified by Random Forest Classification Algorithm. The classification results are shown in Table 2. Table 2 Classification results within a single course

Course	Accuracy	Precision	Recall	F-value
Legal Class	84.36%	84.03%	83.97%	83.99%

From the experimental results, it can be seen that the training and prediction of the classification model within a single course can achieve a classification accuracy of 84.36%, with a classification F-value of 83.99%, which is able to obtain a better classification effect. In order to test the cross-course classification effect of the classification model, the classification model trained with the review data of "Legal Class" is applied to three other courses in different fields: Computer Class,Finance Class,Literature Class, and the results are shown in Table 3.

Table 3 Cross-curricular classification results

Course	Accuracy	Precision	Recall	F-value
Computer Class	79.72%	78.72%	75.20%	76.92%
Finance Class	81.34%	82.17%	80.86%	81.51%
Literature Class	80.25%	80.81%	79.21%	80.00%

From the results in Table 3, it can be seen that the classification results of the model can still achieve a high classification accuracy although there is a decrease in the cross-course classification. The classification accuracy of "Finance Class" and "Literature Class" is above 80%, while the classification accuracy of "Computer Class" is 79.72%. The main reason is that the course commentaries of "Computer Class" contain some codes composed of English words, which affects the classification accuracy to a certain extent.

2.3 Text Analysis and Evaluation of Chinese MOOC Class Reviews

2.3.1 Course content evaluation based on word cloud

The content-related category of comments is mainly related to course content[3]. These comments on course content can provide reference for learners' learning, so that learners can understand the hotspots that other learners are concerned about; they can also provide guidance for teachers' teaching, so that teachers can easily understand the content that students are more concerned about and make adjustments to their teaching based on learners' feedback. For content-based comments, content semantics can be mined through word frequency statistics and cloud diagrams.

The word cloud diagram of content-related comments contains a large number of words related to course content, and there will be large differences between different courses. Comment categorization can quickly filter out the comments related to course content, so as to obtain the content that learners discuss more. Vocabulary that appears more frequently in the comments means that this is the content that learners are more concerned about and discuss more, and it may also suggest that learners have some learning difficulties with this content. From the teacher's point of view, the teacher should provide more learning resources for these contents and provide learners with more targeted learning guidance to learners. At the same time, instructional

designers should further optimize the teaching design of the above content to improve the quality of Chinese MOOC courses and better meet the learning needs of learners[4].

2.3.2 Overall evaluation of courses based on affective tendencies

Emotional comments reflect the learners' emotional tendency towards the current course content, which can provide reference for the evaluation of online courses, and at the same time help lecturers further improve the quality of courses. Commonly used sentiment tendency analysis mainly uses the method based on sentiment words, which calculates the sentiment tendency of the whole sentence by weighting positive and negative sentiment words. In this paper, based on the emotion word weighting model, the emotional tendency of a comment is evaluated by the difference in the number of positive and negative emotion words, which is calculated as follows: for all the emotion-related categories of comments in the comments of a course, the number of positive and negative emotion words in each comment is queried from the semantic features. Then the emotional tendency value of the course is calculated by the following formula[5]:

$$E = \frac{\sum_{i=1}^{n} w_{ipos} - \sum_{i=1}^{n} w_{ineg}}{\sum_{i=1}^{n} w_{ipos} + \sum_{i=1}^{n} w_{ineg}}$$

In this formula, E denotes the emotional polarity value of the course comments, n is the number of comments in the emotion-related category in the course, Wipos denotes the number of positive emotion words in the ith comment, and Wineg denotes the number of negative emotion words in the ith comment. the value of E is between 1 and -1: 1 denotes that the emotional polarity is completely positive, and -1 denotes that the emotional polarity is completely negative[6]. The four courses in the dataset were analyzed for sentiment polarity according to the evaluation method described above, and the results obtained are shown in Table 4.

Table 4 Affective polarity analysis of the courses

Course	Total positive words number of emotion	Total number of negative emotion words	Emotion polarity E-value
Legal Class	371	66	0.698
Computer Class	57	14	0.606
Finance Class	176	25	0.751
Literature Class	257	40	0.731

From the results in Table 4, it can be seen that the emotional polarity E-values of the above four courses are all greater than 0.6, indicating that learners' emotional evaluations of these four courses on Love Course Online tend to be positive. Due to the instability and complexity of affective tendency, the existence of small differences in E-value between different courses cannot be used as an absolute standard for course judgment, but it can provide a reference for course evaluation. In addition, teachers should focus on the negative emotions appearing in the comments and carefully screen the reasons for these emotions and make targeted adjustments according to the different reasons[7].

2.3.3 Evaluation of course services based on information feedback

The more noteworthy content in the comments of other categories is the relevant questions and comments about technical support, which can provide reference for technical support personnel, so as to improve the quality of service and enhance the user experience. In other comments, by filtering the questions and sentences and setting relevant keywords, such as "video", "card", "download" and other terms, we can quickly filter out the relevant valuable comments, such as Table 5, which is the list of comments from the course service evaluation. For example, Table 7 shows representative questions and technical comments filtered out from other types of comments in the "Law" course. These comments asking for information and technical support are an important part of the e-learning experience. When the number of comments is large, this information is often overwhelmed and difficult to be noticed by technical support staff. By further mining the categorized comments, it is possible to quickly locate such information and improve the management of Chinese MOOC courses[8].

This kind of information can be quickly localized by further mining the categorized comments, which improves the efficiency of the management and maintenance of the Chinese MOOC courses. Technical support staff of Chinese MOOC courses should carry out targeted investigation of these problems and provide effective technical support to learners to improve their learning experience. At the same time, these comments are also important reference data for the evaluation of the usability and ease of use of the e-learning platform, and have a strong reference value for improving and perfecting the technical architecture of the e-learning platform and the quality of course services.

Table 5 Helpful information comments

Question Comments	Type	1.Why can't you watch the courseware 2.Why can't I open the webpage 3.Why can't I download 4.What if the client doesn't show the video
Keyword Comments	Filter	1. The video has no sound for two days. 2.The video can't be watched because of lagging 3. Audio and video are not synchronized 4.Video clarity is poor, affecting the viewing experience

3 Influencing Factors and Optimization Suggestions of Learners' Emotional Tendencies Based on Sentiment Analysis of Chinese MOOC Comments

3.1 Teacher Influencing Factors and Suggestions

First of all, teachers' verbal expression includes the verbal expression of the content of the lesson, the tone of voice and other aspects. For example, the teacher's speed of speech, intonation, and diction, as well as the way the teacher expresses his/her ideas, all affect the students' emotional tendencies[9-10]. Teachers' excellent language expression can help students understand and digest the lecture content more thoroughly. Therefore, in Chinese MOOC platform, teachers should pay attention to their language expressions and use simple, clear and easy-to-understand language expressions to explain some complex concepts.

In Chinese MOOC platform, teachers should pay attention to choosing appropriate teaching methods in order to better enhance students' emotional experience and promote their learning effects. For example, teachers can use a variety of teaching methods, such as interaction, heuristic, case study, etc., to stimulate students' learning interest and motivation.

3.2 Course content influencing factors and suggestions

Course content includes the quality of courseware production, teaching design and layout of course knowledge points, as well as the degree of difficulty[11-12]. As far as the quality of courseware production is concerned, whether the courseware is well-produced or not has little influence on students, but if there are some wrong contents in the courseware, it will have a great influence on students' emotional tendency. Teaching design and the arrangement of course knowledge points have an important impact on students, when the difficult knowledge points in front of the students will appear to be afraid and negative experience. Therefore, in Chinese MOOC platform, teachers should carefully design and plan the course content according to the learners' needs and interests, reasonably arrange the knowledge points, and improve the learners' emotional engagement and learning effect.

3.3 Influencing factors of learner experience and suggestions

Learner experience includes the sense of acquisition and identity, and is also influenced by multiple factors such as teachers and course content as well as learning motivation[13-14]. In Chinese MOOC platform, the sense of acquisition and identification can have a positive emotional impact on learners. When learners master new knowledge and skills, complete tasks and assignments, and get higher scores in learning, they will have a sense of acquisition and a sense of identity. A sense of acquisition can improve learners' self-confidence and self-efficacy, and enhance learners' motivation and interest in learning, and a sense of identity can improve learners' self-identity and emotional engagement, and promote learners' positive emotional experience in learning. When learners gain a sense of achievement and satisfaction, they feel proud and satisfied, which improves their self-esteem and sense of self-worth, and also can stimulate their self-motivation and desire for inquiry, and push them to make continuous progress and improvement in learning. Therefore, in MOOC platform, teachers should design and arrange a series of challenging and practical tasks and quizzes to stimulate learners' interest and motivation in learning, promote learners' independent inquiry and learning, improve learners' sense of acquisition and satisfaction, and enhance their interest in learning.

3.4 Platform Support Influencing Factors and Suggestions

In the process of learners' Chinese Language Program on the Chinese MOOC platform, the technical support of the platform will also have a certain impact on learners' emotional tendencies. Learners use the platform for learning and communication, and factors such as the platform's user experience and interaction design, technical support, and reliability of information will affect learners' emotions[15]. For example, the issuance of certificates after learning a course and the feedback of answers after taking a test all affect the learners' rating of the course. If the technical support of the platform can respond to learners' questions and feedback in a timely manner and provide effective solutions and help, learners will feel concerned and respected, thus enhancing their positive emotions; on the contrary, if the technical support of the platform cannot respond to learners' questions and feedback in a timely manner, learners will feel neglected and helpless. Therefore, the platform should ensure that it has good user experience and interaction design, and perfect technical support to help learners get a good user experience and emotional experience.

3 Conclusion

Accompanied by the continuous development of information technology, education and teaching are also facing the reform of the Internet, Chinese MOOC platform provides learners with a good platform for online course learning, and also opens a comment section for learners to express their views and opinions. Through the sentiment analysis of the text in the comment area, we can fully understand the learners' emotions and opinions, which is of great significance for both learners and platform management. Based on this, this paper constructs a Chinese MOOC comments sentiment analysis model based on big data, takes the comment data of the law course as the experimental object, analyzes the indicators of course content evaluation, overall course evaluation, and course service evaluation, and puts forward corresponding optimization suggestions from the teachers, course content, individual learners, and the MOOC platform accordingly in order to provide a certain amount of references for efficient implementation of Chinese MOOC platform teaching.

References

1. Helmi Norman,Mohamed Ally,Norazah Nordin.Use of Social Media and Social Network Analysis for Mobile Learning. Mobile and Ubiquitous Learning,. 2018.11:249-259.
2. Norazah Nordin,Helmi Norman,Mohamed Amin Embi,Ahmad Zamri Mansor.Factors for
3. Development of Learning Content and Task for MOOCs in an Asian Context.International Education Studies, 2016,5(9):2-3.
4. Zhong Zhaoman; Huang Xianbo; Xiong Yulong. Sentiment Analysis of Critical Incident Commentary Subjects Based on Interactive Attention [J]. Data Acquisition and Processing, 2023, 38 (05): 1206-1213.
5. Zhang, Lina; Dong, Lulu; Li, Mei; Tan, Juan-sae. Sentiment analysis of online reviews based on machine learning [J]. Journal of Lanzhou College of Arts and Sciences (Natural Science Edition), 2023, 37 (05): 52-57.
6. Sun Chuanyuan; Dong Limin; Li Aiming. A study on the effect of deep learning experience of catechism learners: An analysis based on the content of learners' comments in the course "Psychology and Life" [J]. Adult Education, 2023, 43 (09): 54-60.
7. Liang ZY; Zhu LJ; Chen J; Chang GG. A study on sentiment analysis of online course reviews based on LDA-LSTM model [J]. Modern Information Technology, 2023, 7 (16): 79-83+88.
8. Zhou Yan. A Sentiment Analysis Approach for Online Learning by Incorporating Bi-LSTM and Conditional Random Field [J]. Journal of Xiamen University (Natural Science Edition), 2023, 62 (04): 687-694.
9. Tingting He. A Study on Analyzing Learners' Emotional Tendencies Based on Reviews of Online College English Courses [J]. English Square, 2023, (15): 113-116.
10. Wei, Xiaocong; Yu, Lan. Construction and application of a sentiment recognition corpus for Chinese Chinese MOOC Comments [J]. Journal of Chongqing University of Technology (Natural Science), 2023, 37 (04): 174-181.
11. Xiong Tingyan; Liu Zeping; Jin Meiling; Liao Ying. Text Semantic Mining and Sentiment
12. Analysis Based on Catechism Course Commentary - An Example from the Course of Econometrics Fundamentals and EViews Software Operation [J]. China Journal of Multimedia and Network Teaching (Lenten Edition), 2023, (02): 221-224.
13. Zhu, F.; Tao, Y.. Research on Course Quality Evaluation System of MCT Learners--Taking Tourism Management as an Example [J]. Tourism Overview, 2023, (02): 57-59+63.
14. Wang Nujian; Aizirguli Yusuf; Chen Degang. Sentiment Analysis of MOOC Reviews Based on Lexicon and Heterogeneous Fusion Network [J]. Modern Electronic Technology, 2022, 45 (17): 79-84.
15. Nigram Buysmujiang; Aizilguli Yusuf. Emotional tendency analysis of catechism user comments based on BERT and two-way GRU model [J]. Computer and Modernization, 2021, (04): 20-26.
16. Deng Xiaorui; Zhou Bin; Diao Yajing. Research on the influence of e-commerce Q&A reviews on consumers' purchase intention [J]. China Price, 2023, (11): 82-84+97.
17. Ma R; Li M; Zheng Ziyuan; Kong Xiangrui. Research on Sentiment Analysis and Theme Mining Based on MOOC Course Reviews [J]. Educational Information Technology, 2023, (09): 3-7.