A Method for Online Course Evaluation Based on Continuous Bag-of-Words Model and Semantic Analysis—A Case Study of Statistics

Yongjie Chu[*1, 2], Cengceng Liu[3]

[1]*School of Management, Nanjing University of Posts and Telecommunications, Nanjing, China*
[2]*Institute of High-Quality Development Evaluation, Nanjing University of Posts and Telecommunications, Nanjing, China*
[3]*School of Management Science and Engineering, Nanjing University of Finance and Economics, Nanjing, China*
**chuyongjie@njupt.edu.cn*

Abstract

With the rapid development of the Internet and multimedia, online courses have become one of the main ways for students to learn. Online course reviews are the comments the learners or students published voluntarily based on their real learning experience of a course, which include diverse information. This paper selects the statistics-related courses in the MOOC platform as the research object and obtains online reviews of seven courses, then employs natural language processing techniques to deal with the review data and further develops a method for online course evaluation. First, the continuous bag-of-words model is used to extract the feature words of courses from the online course reviews. Secondly, based on mutual information and semantic similarity analysis, this paper identifies and clusters the learners' preferences for online courses, the preferences include six aspects. Finally, we compute the value and weight of each aspect of preference according to the sentiment propensity and the occurrence frequency of preference words, respectively. Then the overall score of each course is calculated using the above values. Based on the results, we proposed suggestions and enlightenments for the online course development and improvement.

Keywords: *Online course reviews, Course evaluation, Continuous bag-of-words model, Semantic analysis*

1 INTRODUCTION

With the development of the Internet, more and more attention has been paid to online courses. In China, there are many online teaching platforms favored by learners, such as MOOC, and NetEase Cloud classes. With the continuous development and construction of online courses, there are still several problems to be well solved, such as the high dropout rate, low participation, and poor learning effect of students. How to assess the implementation effect of online courses and evaluate the quality of online courses is a key challenge for the development of online courses. At present, there is no unified and authoritative online course quality evaluation system, scholars have made many studies on this issue, and tried to build different online course quality evaluation systems in their research. However, current evaluation methods are mainly based on questionnaires or interviews with the learners and course experts.

The online course learning or experience, just like the currently popular online shopping, more and more learners or students are accustomed to post comments after learning a course to share their experiences and feelings. We name these comments related to online course learning as online course reviews (OCR). OCR refers to the online evaluations of a course written in the text by students or learners on the corresponding course websites after they have learned a course online, which contain personal opinions and emotional information about the courses and the learning experience.

A large amount of review data accumulated on the web page of online courses often truly shows the intuitive learning experience, real feedback, and suggestions to the courses and teachers, as well as the learning effect of learners. These reviews not only reflect the quality of the courses from the perspective of learners, but also affect the potential learners to choose an online course. Therefore, the OCR plays a significant role in the improvement of online courses. In this paper, based on

B. Fox et al. (Eds.): IC-ICAIE 2022, AHCS 9, pp. 765-771, 2023.
https://doi.org/10.2991/978-94-6463-040-4_116

online course reviews, we employ natural language processing technology and data mining methods to design the evaluation process of online courses from the perspective of learners. The framework of this work is illustrated in Figure 1. We employ continuous bag-of-words model to extract the feature words of courses according to the online course reviews captured from the online course platform, then compute the mutual information and the semantic similarity between feature words. The aim of this paper is to recognize learners' preferences and calculate the overall score of each course according the preferences of learners. This study can benefit the teacher to clarify the preferences of learners or students, and to improve the quality of online courses in future course design.

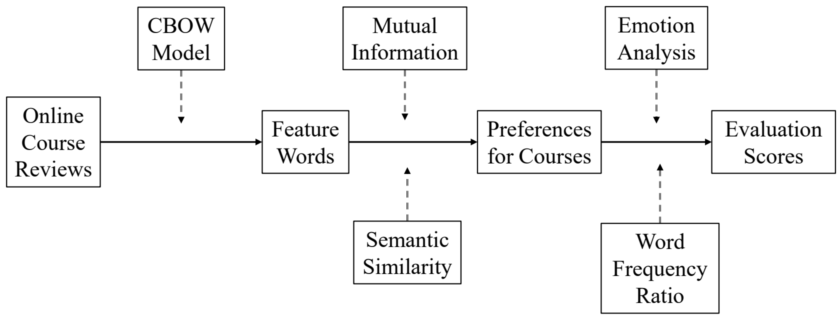

Figure 1 The illustration of OCR-based online course evaluation method

2 LITERATURE REVIEW

The research on the online course evaluation has been carried out in a wide range. McGahan et al. (2015) propose that there is a link between online course quality standards and student learning outcomes, and the course design is essential for online student participation and retention. [6] Jaggars and Xu (2016) analyze the characteristics of online course design, organization and presentation, learning objectives and assessment, and interpersonal interaction. [4] They find that the performance of students is influenced by the course design characteristics, and the score of students is influenced by the interactions. Considering the online learning environment, Gomez et al. (2018) combine the role of teachers, the performance of teachers, and the demographic characteristics of students with the teaching effect, and construct a set of comprehensive online teaching evaluation tools. They point out that the teaching effect is mainly affected by the role and performance of teachers, and the status of students, such as work commitment and family responsibility, has a significant impact on the teaching effect. [3] Barteit et al. (2020) study the quality of online education in the medical field in low-income countries and draw a lot of conclusions through questionnaires and pilot studies. They find that participants are less blinded when they are interviewed and suggest that researchers should take measures to find out more effective and reliable ways to evaluate online courses. [1]

To make an effective evaluation, the criteria must be demonstrated. Many scholars analyze this issue. Martin et al. (2019) Interview the award-winning teachers who design the online courses or teach through the online courses to identify the important elements of course design. [5] They further propose to add or design the elements of online courses by a reverse design method, which can provide diverse opportunities for different learners to interact with each other. Parker et al. (2018) evaluate over 20 online teaching courses from four dimensions, i.e., content, design, interactivity and availability. They conclude that the average score of online courses is 73, while video and network resources only account for 48 and 62, which emphasizes the importance of high interactivity to the online teaching effect. [7] Calderon et al. (2020) make the online course evaluation according to the following three dimensions, they are teaching situation (accuracy and complexity), interactive communication quality, and meta-learning (reflection on tasks and learning process). Recently, some scholars have concentrated on online course evaluation in China. [2] Wang et al. (2018) propose that high registration rates and low completion rates are a major bottleneck in the development of MOOC. In order to improve the quality of MOOC courses and the graduation rate of students, they constructed a set of semantic analysis models (SMA) to analyze the learning status and emotional polarity of students. [8] Xie (2019) points out that the huge amount of commentary data on the MOOC platform is of great significance to the course quality research of distance education. This paper attempts to track the learning log data in the MOOC platform and identify multiple dimensions to be optimized for the teaching. [9] Zhou et al. (2021) apply machine learning and computer vision techniques to investigate the consumer behavior in the online classroom, they explore the effect of video features on the popularity of the online education courses. [10]

The above research has done a lot of work for the course evaluation from various perspectives; however, fewer studies are focusing on the significance of OCR in the online course platforms for the course assessment. In this study, we aim to further improve the course evaluation with the help of OCR data.

3 METHODOLOGIES

3.1 Feature Words Extraction for Online Course Evaluation

Feature word recognition and extraction basically according to word segmentation. In this paper, we use the Jieba lexicon and HIT stop word list to segment the Chinese text content, and then utilize the Word2Vector model to generate the feature word vectors from the online course reviews.

Word2Vector generates word vectors based on the neural network language model. It consists of a three-layer neural network, including an input layer, projection layer and output layer. Word2Vector includes two basic models, one is the continuous bags-of-words (CBOW) model, which predicts the central word through its adjacent word vectors, and the other is a skip-grams model, which uses the central word vector to predict its context. By training the Word2Vector model, a word vector can be obtained from the text of the OCR corpus.

CBOW model is used to train word vectors. Given a training sentence consisting of n words $S = (w_1, w_2, w_3, \cdots, w_n)$, the text sample can be expressed as $(c(w_i), w_i)$, the context of the central word w_i is $context(w_i) = \{w_j | j \in [(i-k, i) \cup (i+1, i+k)]\}$. The word distance from the central word is k. The CBOW model contains the following three layers of networks:

(1) Input layer. Input the context information $context(w_i)$ of a central word w_i, including 2k words. If the dimension of a word vector is m, then the input word vector can be expressed as $\{v(context(w_1)), context(w_2), \cdots, context(w_{2k})\}$.

(2) Projection layer. Calculate the sum of the input 2k word vectors, i.e., $\sum_{i=1}^{2c} v(context(w_i))$.

Then the maximum likelihood estimation is used to estimate the conditional probability of a word $P(w_i | context(w_i))$.

(3) Output layer. Output the probability of a sentence S, which is the product of the probability of *n* words.

$$P(S) = \prod_{i=1}^{n} P(w_i | context(w_i)) \quad (1)$$

3.2 Identification the Course Preferences of learners

In online course reviews, the learners usually comment on some features of a course that they are concerned about, thus these features of the course exactly reflect the learners' preferences for online courses. To further clarify the learners' preferences for the online courses, and to reduce the dimensionality of variables, this paper clusters the preferences into several classes. Considering that the multiple course features the learners concerned often appear together in the reviews, we utilize the word co-occurrence as the basis for clustering and discrimination of course demand preferences. Moreover, feature words with similar semantics usually express similar course features, semantic similarity is also used as the basis for clustering and identifying the preferences for courses.

Word co-occurrence refers to the co-occurrence of certain keywords involved in a text such as the sentences or paragraphs, which implies the semantic association information between keywords. In word co-occurrence models, mutual information (MI) is developed to measure the degree of association between words. The larger the MI value between the two words, the greater the correlation between them. Conversely, a small MI value indicates less interdependence between words. Therefore, this paper chooses the mutual information $MI(w_i, w_j)$ to measure the correlation between the feature word w_i and w_j. $MI(w_i, w_j)$ is defined as follows:

$$MI(w_i, w_j) = log_2 \frac{p(w_i, w_j)}{p(w_i) * p(w_j)} \quad (2)$$

where $p(w_i)$ and $p(w_j)$ denote the occurrence probability of feature word w_i, and w_j, respectively. $p(w_i, w_j)$ is the co-occurrence probability of feature word w_i and w_j.

Besides the mutual information, semantic similarity is also important for word clustering. Semantic similarity refers to the possibility that two words can replace each other in different contexts without changing the syntax and meaning of the text. Compared with MI, it is an intrinsic reflection of the correlation between two words. The higher the semantic similarity between two words, the stronger the correlation between them. In this work, the following similarity function $S_m(w_i, w_j): S \times S \to [0,1]$ is used to measure the semantic similarity of two feature words w_i and w_j in the feature set S.

To identify the correlation between two feature words more elaborately, we combine $MI(w_i, w_j)$ and $Sim(w_i, w_j)$ to calculate the correlation of feature words. Let $R(w_i, w_j)$ be the correlation between the feature word w_i and w_j, then it is computed by

$$R(w_i, w_j) = \alpha MI(w_i, w_j) + \beta Sim(w_i, w_j). \quad (3)$$

Eq. (3) describes the strength of the correlation between the two course feature words. α and β are the balance parameters and α+β=1. To determine the values of these two balance parameters, according to the OCR data, we make couple of experiments to evaluate the parameters. Finally, we set the value of α and β to be 0.4 and 0.6, respectively. The larger the value of $R(w_i, w_j)$, the greater the correlation between the feature word w_i and w_j. As a result, two feature words with a larger value of $R(w_i, w_j)$ can be clustered into the same class of course preferences.

4 EMPIRICAL ANALYSIS

4.1 Data Description

In this paper, we take the courses related to statistics as example and crawl the online course reviews of these courses on the MOOC platform. The basic information of the courses is crawled through crawling technology, including the course name, number of reviews, rating score and course introduction. As only a small part of the courses owns a large number of online course reviews, those courses that contain more than 100 online reviews are selected. Finally, we obtain 2340 reviews of 7 courses which include Statistics, Applied Statistics, Biostatistics and Medical Statistics. Among them, the largest number of reviews of a course is 648, while the smallest number of reviews of a course is 117.

4.2 Results and Analysis

Feature word extraction from the online course reviews consists of three steps. The first step is Chinese word segmentation, part-of-speech labeling and counting the word frequency. We employ a dictionary and stop word list to remove the meaningless function words and auxiliaries in the corpus. Using the precise mode of Jieba, the text is precisely segmented and labeled. Then, we apply the fuzzy matching algorithm to filter out all nouns, and select the nouns with more than 10 occurrences as the candidates of feature words. The second step is feature word pruning. To make the feature words better represent the course characteristics, the candidate set of feature words is further filtered. A single word is pruned by the regular rules, which means that a feature noun with only one Chinese character in the set is removed. Then those words with the same or similar meaning in the feature words are merged to obtain a new list of feature words. The last step is the manual review of the feature words. We manually check the candidates of feature words to determine whether they can be used to represent the course. Finally, we select 33 course feature words and sort them according to the number of occurrences.

Based on the mutual information and semantic similarity, we analyze the correlation among 33 feature words. According to the correlation score $R(w_i, w_j)$, two feature words with higher values are clustered into one class to represent the learners' preferences. the consumer's preference for online courses demand was determined (see Table 1). Table 1 shows the different categories of learners' preferences for online courses, including 6 different categories. Accordingly, the preferences of learners include six dimensions such as Teacher Style, Teaching State, etc.

Table. 1 Evaluation dimensions and corresponding labels for statistical courses

Preferences of Learners	Feature Words
Teacher Style	Careful, Clear-minded, Patient, Knowledgeable, Focused, Interesting
Course Design	Contents, Knowledge points, In-class tests, Exercises, Discussions, Teaching methods
Teaching State	Gentle, Heavy Accent, Speed of speech, Boring
Teaching Implementation	Making the hard easy、Reading the book only, Interactive, Heuristic
Learning Experience	Easy, Understandable, Bad, Hard to understand, Worthwhile, Exercise the thinking, Grateful, Recommended
Course Platform	Quit abruptly, Playback, Smooth video, Subtitle, Speed control

To obtain the scores of each dimension of learners' preferences, we build an effective dictionary based on CNKI emotion dictionary, which contains positive, negative and degree words. Then we discriminate the polarity of an emotional word and assign it a value. If an emotional word is in the dictionary of positive emotional words, then the emotional score is 1. If an emotional word is in the dictionary of negative emotional words, then the emotional score is -1. Otherwise, the emotional score is 0. The negative words and degree adverbs are checked by traversing all of the words. If there are one or more negative words, multiply by $(-1)^m$, where m is the number of negative words. If there are degree adverbs, then it is multiplied by the weight of degree adverbs.

As the result, we map the online course reviews, including the feature words and emotional words, into the learners' preferences, and further compute the score of each preference. A score of 5 indicates the strongest preference, 3 indicates the middle preference, and 1 indicates the weakest preference. Based on the above scoring principles, we find out the preference level of one or more features involved in the online course reviews. We then calculate the score of one dimension in the preference by averaging the total score of the same dimension, and finally determine the score of all online reviews in each dimension of preferences. As a result, the learners' scores on six dimensions of preferences for courses can be obtained. The scores for each dimension of a single course are added up separately, and the score for each dimension of a single course is calculated by dividing the total number of reviews. Moreover, the weight of each dimension of the preference is calculated by the ratio of the frequency of feature words to the frequency of all feature words. Table 2 shows the weight of the six dimensions of preferences. At last, we combine the scores of each dimension of a single course to obtain the overall score of a single course. Table 3 lists the scores of each dimension and the overall scores of courses.

Table 2 Weights for each dimension of course preference

Teacher Style	Teaching State	Teaching Design	Teaching Implementation	Learning Experience	Course Platform
0.15	0.12	0.21	0.26	0.15	0.11

According to Table 2, the learners or students pay more attention to the teaching implementation and teaching design, which implies that compared with other aspects, those preferences related to the learning process are quite important for the online course. The learning experience and the style of the teachers are also important as both weights are big. It demonstrates that the learners are concerned much about the teachers, as well as the feeling in the learning.

Table 3 Scores of each dimension and the overall scores of courses

	Teacher Style	Teaching State	Teaching Design	Teaching Implementation	Learning Experience	Course Platform	Overall Score
Course 1	0.90	0.86	0.94	0.95	0.85	0.90	0.91
Course 2	0.90	0.80	0.91	0.90	0.90	0.83	0.88
Course 3	0.89	0.85	0.93	0.95	0.86	0.94	0.91
Course 4	0.90	0.90	0.86	0.92	0.75	0.75	0.86
Course 5	0.83	0.88	0.89	0.86	0.89	0.87	0.87
Course 6	0.88	0.88	0.86	0.86	0.92	0.86	0.88
Course 7	0.84	0.90	0.83	0.93	0.91	0.71	0.87

Table 3 clearly shows the overall score of a single course and the score of each dimension the learners are concerned about. The evaluation process takes what the learners are interested in into consideration, and reflect the overall quality of a course from the perspective of learners. Based on the above analysis, this paper puts forward the following enlightenments and suggestions on the development of online courses. (1) Carefully mining the online course review information, and effectively identifying the features or preferences of the courses that students are interested in will help find out the shortcomings and advantages of an existing online course. It provides meaningful suggestions for the future development and improvement of online courses. (2) It is suggested that encouraging learners to actively provide information feedback about the online courses to ensure the scale and quality of online review data. As the number of online reviews increases, it helps truly reflect the learners' assessment of the quality of the course and learning experience, thus further scientifically benefiting the analysis of online review data.

5 CONCLUSIONS

This paper crawls the review data of 7 statistics-related online courses on the MOOC platform. With these data, 33 course features of online courses are extracted and 7 different preferences are further obtained through correlation analysis. We utilize natural language processing techniques such as continuous bag-of-words model, and word similarity models such as mutual information and semantic similarity analysis, to extract course features of learners. Finally, we establish a quality evaluation method for online courses based on the seven-dimensional vector of course preferences.

This study provides a new method for the evaluation of online courses, but there are some limitations. Since only one type of courses is selected for this article, the number of samples is very limited. In the future, we will consider designing more detailed indexes for evaluation and develop more reasonable weight functions for these indexes. Besides, courses and reviews on the cross-platforms will benefit the evaluation of online courses.

ACKNOWLEDGEMENTS

This research was partially supported by the Humanities and Social Science Research on Youth Fund Project of the Ministry of Education of China (Grant No. 19YJC870003), the Natural Science Foundation of Jiangsu Province (Grant No. BK20210576), and sponsored by NUPTSF (NY221107).

REFERENCES

[1] Barteit S, Guzek D, Jahn A, et al. Evaluation of e-learning for medical education in low-and middle-income countries: A systematic review[J]. Computers & education, 2020, 145: 103726.

[2] Calderon O, Sood C. Evaluating learning outcomes of an asynchronous online discussion assignment: a post-priori content analysis[J]. Interactive Learning Environments, 2020, 28(1): 3-17.

[3] Gómez-Rey P, Fernández-Navarro F, Barbera E, et al. Understanding student evaluations of teaching in online learning [J]. Assessment & Evaluation in Higher Education, 2018, 43(8): 1272-1285.

[4] Jaggars S, Xu D. How do online course design features influence student performance?[J]. Computers & Education, 2016, 95: 270-284.

[5] Martin F, Ritzhaupt A, Kumar S, et al. Award-winning faculty online teaching practices: Course design, assessment and evaluation, and facilitation[J]. The Internet and Higher Education, 2019, 42: 34-43.

[6] McGahan S, Jackson C, Premer K. Online course quality assurance: Development of a quality checklist [J]. InSight: A Journal of Scholarly Teaching, 2015, 10: 126-140.

[7] Parker R, Boulos L, Visintini S, et al. Environmental scan and evaluation of best practices for online systematic review resources[J]. Journal of the Medical Library Association: JMLA, 2018, 106(2): 208-218.

[8] Wang L, Hu G, Zhou T. Semantic analysis of learners' emotional tendencies on online MOOC education[J]. Sustainability, 2018, 10(6): 1921.

[9] Xie Z. Bridging MOOC education and information sciences: empirical studies[J]. IEEE Access, 2019, 7: 74206-74216.

[10] Zhou M, Chen G H, Ferreira P, et al. Consumer Behavior in the Online Classroom: Using Video Analytics and Machine Learning to Understand the Consumption of Video Courseware[J]. Journal of Marketing Research, 2021, 58(6): 1079-1100.