Received February 5, 2021, accepted February 19, 2021, date of publication February 24, 2021, date of current version March 5, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3062052

Evaluating On-Line Courses via Reviews Mining

CONG QI AND SHUDONG LIU
School of Information and Security Engineering, Zhongnan University of Economics and Law, Wuhan 430073, China

Corresponding author: Shudong Liu (bupt.mymeng@gmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61602518 and Grant 71872180, and in part by the Fundamental Research Funds for the Central Universities, Zhongnan University of Economics and Law under Grant 2722020JCT033.

ABSTRACT Because the participants are not limited by age-, gender-, race-, or geography-related barriers, recently, massive open online courses (MOOC) have witnessed remarkable growth in number of online self-learners, courses providers and online platforms. MOOC learners usually share some learning experiences and release millions of course-related comments in discussion forum. On the one hand, these comments could reflect the learners' attitudes toward the online courses. On the other hand, semantic knowledge hidden in these comments would assist other learners to choose the appropriate courses and help instructors to improve their courses' attraction. Recently, few research works focus on evaluating the courses through reviews mining. Thus, this paper constructs a curriculum evaluation system based on MOOC reviews, which quantifies the curriculum from different topics. Firstly, we employ latent dirichlet allocation (LDA) to mine the reviews generated by students, and obtain a topic-word distribution matrix and a comment-topic distribution matrix which can describe the topics of the course comments. Next, the emotion values of the comments in each topic are calculated by the auto-encoder and Bi-LSTM text classification model. We utilize the emotions and the quantified scores of the courses on different topics to establish a comprehensive curriculum evaluation system. The experimental results show that there are five main indicators abstracted from students' reviews, which are instructor, course content, course assessment, MOOC platform, and hot courses. Moreover, comment texts of each course under different evaluation indicators are objectively and accurately converted into numerical marks, which can provide the students and educators with reliable references.

INDEX TERMS MOOC, curriculum evaluation, review mining, text classification, emotion analysis.

I. INTRODUCTION

With the development of web 2.0 and education information techniques, recently, massive open online courses (MOOC) have witnessed remarkable growth in the number of online self-learners, course providers, and online platforms [1]–[3]. MOOC is large-scale, online, and student-driven, moreover, all educational resources are offered freely and openly for learners to reuse anytime and anywhere. During the past decade, about more than 180 million have registered for at least one MOOC and more than 950 universities worldwide are offering 16.3k courses on different platforms such as Coursera, edX, Udacity, FutureLearn, and Swayam [4]. It is obvious that MOOC plays an increasingly important role in modern education.

The associate editor coordinating the review of this manuscript and approving it for publication was Luigi De Russis.

Another important and attractive characteristic of MOOC is that there is a discussion forum, which facilitates interaction among learners and between learners and instructors to ensure a high-quality learning experience. The instructors use the discussion forum to assign learning tasks, offer learning suggestions, and answer learners' questions. The learners usually use it to report problems, express opinions, share learning tips, and collaboratively complete learning tasks. Millions of comments are posted by learners in the discussion forum. The comments not only reflect the learners' attitudes and learning tips toward the online courses, but also offer an evaluation way of the courses from various aspects. The previous research works on MOOC comments can be divided into two groups [5], [6]: the topic of the comments [7], [8] and the poster role [9]–[11], but few research works focus on evaluating the courses through reviews mining. Thus, this paper attempts to demonstrate how to excavate the hidden evaluation aspects from the reviews and then utilize the related topics to construct a curriculum evaluation system that navigates curriculum advancement and provides guidance for course selection.

The main contributions of this paper include:(1) we propose a short-sentence latent dirichlet allocation (LDA) topic model. In view of comment text features such as short length, sparse content, insufficient word co-occurrence information, and strong context-dependence, we add sentence-level information to LDA topic model. After dividing the comment texts into sentences, and use sentence as a unit to classify topics more accurately by extracting unique topic of the corresponding sentence. (2) we utilize autoencoder and Bi-LSTM to build an emotion classification model for comments. Autoencoder is used for feature extraction and emotion classifier is trained by the recurrent neural network. The comments are quantified into numerical scores according to the ratio of positive emotions.

II. RELATED WORKS

A. CURRICULUM EVALUATION

In the aspect of MOOC curriculum research, the curriculum evaluation system for MOOC has not been scientifically and reasonably established [12], [13], and an appropriate evaluation method has not been found to evaluate the quality of the course accurately [14], [15]. The existing evaluation methods of traditional courses include responsive evaluation model, appearance evaluation model, goal dissociation evaluation model, CIPP evaluation model, and goal evaluation model, etc. One of the primary defects in these methods above is that the quantification-based curriculum evaluation method excludes factors that are related to the quality of curriculum but cannot be quantified, resulting in biased evaluation results and incomplete management mechanism of curriculum evaluation [16]. Drake *et al.* [17] construct a MOOC evaluation system which includes quantitative factors, qualitative description, and curriculum-oriented improvement. Miranda *et al.* [18] propose an evaluation framework based on data mining and vague set methods, which integrates five first-level indicators, including course content, instructional design, interface design, media technology, and curriculum management. Nie *et al.* [15] propose a diagnostic MOOC evaluation approach, which combines the analytic hierarchy process algorithm and learner review mining to integrate expert opinions, standardized rubrics, and learner feedback into the evaluation process. Zhao *et al.* [19] focus on three types of learning quality evaluation and propose a learning quality evaluation approach.

It is vital that appropriate methods are identified and available to determine the impact of these courses, a crucial but under researched element. Aspects such as the effectiveness and quality of learning and impact of knowledge gained are vitally important in determining the strength of MOOCs as a learning tool, but there is no substantial evidence showing that these factors are measured or evaluated [10]. Although there have been recent reviews conducted on MOOCs [15], [21], few researchers have specifically focused on methods used for curriculum evaluation. Cunningham-Nelson *et al.* [22] propose a net method to visualize the sentiment and key themes of qualitative student survey comments, the user data is student evaluation of teaching surveys, but is not MOOC reviews. With the heterogeneity of participants in MOOCs and the low retention rate [23], [24], conducting effective evaluations of MOOCs is critical. To date, little work has been done in this area and it has been highlighted as an area for future research. Despite increasing research about MOOCs, there are limitations in reporting the methods and using valid and reliable measures in the studies [25].

B. MOOC REVIEW MINING

Most previous researches on MOOC review mining have failed to analyze comment texts systematically or consistently due to the difficulty in the evaluation of massive texts. Fortunately, with advancement in the natural language processing, many researchers exploit text mining techniques to systematically study MOOC reviews [26] for emotion detection, sentiment analysis, opinion mining, knowledge extracting, role labeling, etc. First, many people have done lots of research achievement in terms of emotion detection, for example, Liu *et al.* [27] have designed and implemented a framework based on emotion recognition and topic mining to predict the popularity of MOOC courses, and obtain emotional feedback on the content of the course for the teachers and the administrators to improve teaching strategies and enhance the user experience of the platform. Weng *et al.* [28] propose a social-emotional semantic model based on MOOC reviews, which can help teachers determine the reason for the emotion change and generate course adjustment or personalized instruction. To get a comprehensive understanding of the role of emotions in MOOC participants' learning experiences, Xing *et al.* [29] propose a machine learning model to automatically detect the achievement emotions in the forum posts. Peng *et al.* [30] propose a behavior-emotion topic model, which can be used to detect review's semantic content, study significant differences in discourse behaviors, and focused topics between completers and non-completers.

Besides, many researchers have made great progress in sentimental analysis of MOOC reviews. For example, Onan *et al.* [31] propose an efficient sentiment classification framework with a high predictive performance of MOOC reviews, which is based on conventional supervised learning methods, ensemble learning methods, and deep learning methods. Hew *et al.* [32] use a gradient boosting trees and sentiment analysis approach to predict MOOC learner satisfaction.

Moreover, for opinion mining and knowledge extracting, Kastrati *et al.* [33] propose a framework to automatically analyze opinions of students expressed in MOOC reviews, which takes advantage of weekly supervised methods to deal with unlabeled students' reviews. Kastrati *et al.* [34] propose a model for aspect-based opinion mining of students' comments post on online learning platforms, which

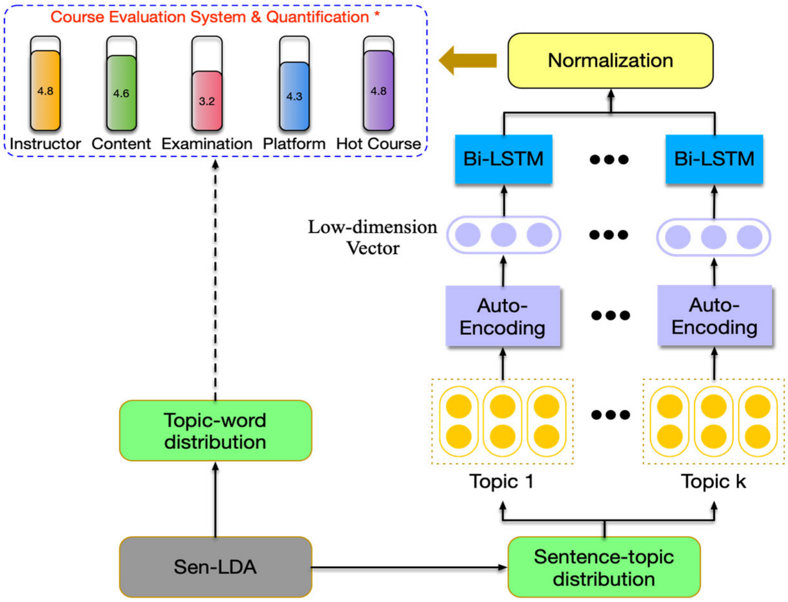

FIGURE 1. **The framework of the proposed approach.**

aims to predict some of the key aspects related to online courses from learners' reviews. Estrada *et al.* [35] build two corpora of expressions into the programming languages domain and use various supervised learning methods to investigate opinion and emotion classification for students' reviews. Misuraca *et al.* [36] use a family technique of opinion mining to compute the sentiment polarity of students' MOOC comments. Koufakou *et al.* [37] use data mining techniques to extract knowledge from student comments, which can help educators and administrators gain insights of student sentiment and views. For the case of MOOC interaction, some researchers have selected some information on MOOC forums, e.g., post quantity, reply time, and teachers' interactive input and made a comparative analysis on the levels, categories, teaching modes, video types, learning support, and evaluation methods of the courses [38], [39]. Shapiro *et al.* [40] show that knowledge, work, convenience, and personal interest are the most prevalent motivations, lack of time is the most common barrier for students in MOOC. Maldonado-Mahauad *et al.* [41] highlight that self-regulated learning is critical for students in MOOCs to succeed.

The difficulty of online MOOC review mining lies in the fact that its key features are very sparse and related to context. This paper will continue and expand the previous work, using the improved LDA topic model to complete the analysis of all the Chinese Universities MOOC course reviews. Furthermore, this paper explores the characteristics of the courses offered by colleges and universities, as well as the topics and concerns in students' comments, and analyzes the obtained topics in detail. In addition, it lays a foundation for the next-step research of text comment quantification. As a popular topic model, LDA [42] has widely used in text mining for data mining, latent data discovery, and finding relationships among data and text documents [43], consequently, we exploit an improved LDA topic model to analyze learners' MOOC comments.

III. OUR PROPOSED MODEL

A. BASIC IDEA

The approach proposed in this paper is shown in Figure 1, and its core processes include: first, collect all the comment text data from the MOOC; second, use short-sentence LDA model to derive the subject information contained in all reviews; then, constructing sentiment classifier for reviews using autoencoder and Bi-LSTM model; then, calculate the emotional values of different subjects in one course comments as the basis for evaluation; lastly, the emotional values of all the comments made by a specific student can also be calculated on different topics, which are matched with the course scores, and courses of higher values are selected for students' reference.

B. SHORT-SENTENCE TOPIC MODEL

The majority of MOOC reviews are less than 50 words long and such short text is likely to create poor context-dependency, insufficient co-occurrence information, and sparse feature matrix, and result in an awful performance of LDA. However, MOOC comments have the advantages of simple and concise content, highly generalized meaning, and usually one sentence (divide comments into sentences at intervals of "," "." "!" "?" ";") covers only one topic, as shown in Table 1.

In order to accurately detect the topics described in the short sentences, this paper improves the LDA model. For MOOC comments, this paper assumes that all the words in a sentence are generated around a topic. Based on this assumption, MOOC comment topic analysis can be done in the unit of sentence. Therefore, in the 3 layers of structure

TABLE 1. MOOC reviews.

Courses	Comments	Topics Contained in Sentences
Python Language Programming - Beijing Institute of Technology	Short sentence 1: The teacher's lecture was very detailed; Short sentence 2: the course is very good and I look forward to the follow-up courses.	Short Sentence 1: Instructor Short Sentence 2: Course Content
"The Essence of C Language Programming" - Harbin Institute of Technology	Short sentence 1: The test is a bit difficult; Short sentence 2: I hope that the programming problems can be done using a computer.	Short Sentence 1: Course Assessment Short Sentence 2: Course Assessment

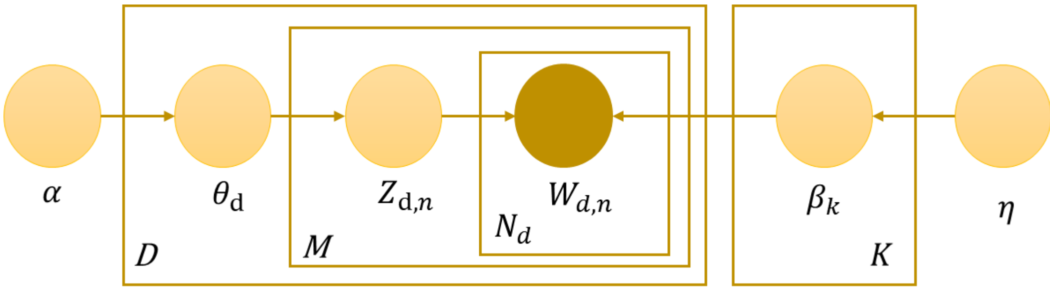

FIGURE 2. The LDA topic model in units of short sentence.

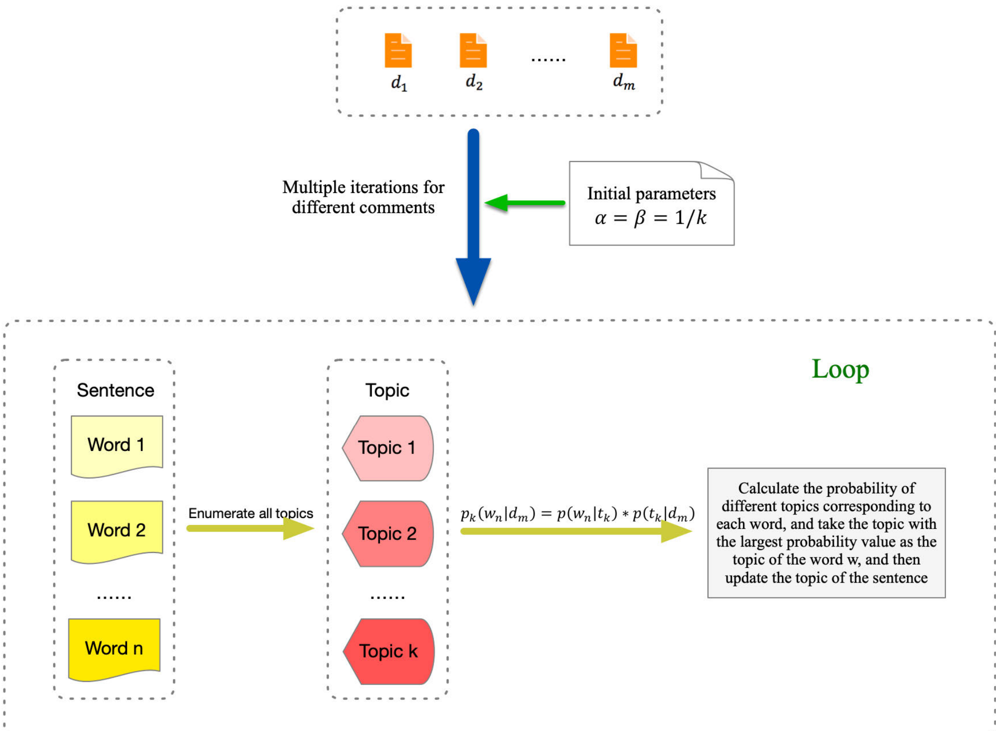

FIGURE 3. Process of MOOC topic model.

of the LDA "text-topic-word" distribution, the new sentence layer is added between the text layer and the topic layer, and the improved model becomes the "short comment sentence-topic-word" model. The graphical representation of this model is shown in Figure 3. The relevant notations are listed in Table 2.

TABLE 2. Notations of LDA model.

Notation	Explanation
D	Quantity of comments
M	Quantity of short sentences
N	Quantity of words
T	Quantity of topics
V	The size of the dictionary
w	Word
z	Topic
ϕ	Topic-Word polynomial distribution
θ	Reviews - Topic polynomial distribution

The improved LDA topic model is still a generative model, and its generating process is:

1) For each topic z, the topic-word distribution matrix ϕ_z is depicted. The polynomial distribution has a Dirichlet distribution of parameter β, and can be represented as:

$$\phi_z \sim Dirichlet(\beta) \tag{1}$$

2) For each comment d:

(1) Describing reviews - topic distribution matrix θ_d is subject to dirichlet distribution of parameter α, and can be represented as:

$$\theta_d \sim Dirichlet(\alpha) \tag{2}$$

(2) For each short sentence in each comment:

Based on the review- topic matrix θ_d, we select a topic z to get the topic of this short sentence; Based on the topic-word matrix ϕ, we extract the unique topic of this short sentence.

Following the above steps, we can utilize the improved LDA model to generate all the text.

As shown as in Figure 3, the set of MOOC participants' comments is defined as $D = d_1, d_2, \ldots d_m (1 \le m \le M)$, where d_m is the preprocessed review record. The short sentence set is represented as $M_{i,s} = \{m_{i,1}, m_{i,2}, \ldots m_{i,s}\}$, where $m_{i,s}$ is the short sentence divided from the comment di. The topic set is defined as $T = t_1, t_2, \ldots t_k (1 \le k \le K)$. In the initial stage of this model, we assign value for priori parameters α and β according to preset parameters (generally $1/K$), and randomly assign value for θ_d and $\beta_k (for\ alld$ and $t)$. Furthermore, we iterate multiple times for sampling as follows:

(1) For the nth word w_n of a certain comment d_m in D, let t_k denote the corresponding topic to w_n, and calculate the probability:

$$p_k (w_n \mid d_m) = p (w_n \mid t_k) * p (t_k \mid d_m) \tag{3}$$

(2) Enumerate the topics in T, and obtain all $p_k (w_n \mid d_m)$ where $k \in [1, K]$. Take the one with the highest probability as the topic of w_n. Based on topic-word distribution, we can obtain the unique topic of the short sentence that includes the word w_n.

(3) In phase (2), if the short sentence $m_{i,s}$ that includes w_n preferred a different topic to the original one, the θ_d and β_k may be affected. Furthermore, it can have an influence on the computation of $p_k (w_n \mid d_m)$.

We consider it an iteration to make We $p(w|d)$ calculations once of all words w in all comments d that belongs to D and reselect their topics. Then, we perform iteration repeatedly until it converges, so as to obtain the topic-word matrix and the sentence-topic matrix. The topic-word matrix indicates the topics of participants' reviews, the hot words of each topic, and the high-frequency concerns of the students. The sentence-topic matrix demonstrates the probability distributions of each sentence under each topic. Here, the short sentences have a 0-1 distribution or an approximate 0-1 distribution. Thus, the implied topics become straightforward after dividing comments into sentences, and the potential topic of each short sentence can stand out.

C. AFFECTIVE COMPUTING OF REVIEWS

Sentiment analysis of MOOC commentary is to transform the sparse vectors of short text into low dimension vectors using the automatic encoding process of deep learning, and the learning process makes the low dimension vectors contain the essential features of text information. With the unnecessary interference in the high dimension vectors removed, the result can be used in sentiment analysis to improve the final classification effect.

The specific process is composed of three steps. We firstly process comment texts in short sentences and build a vector space model. Here, each short text is converted into a vector in space. Second, the high-dimensional sparse vectors are input into the constructed deep noise sparse autoencoder, and the low-dimensional abstract feature vectors are extracted through layer-by-layer abstraction. In this step, the regularization process is also included. Finally, Bi-LSTM is used to obtain the classification results of the short texts. The overall process is shown in Figure 4.

The pre-processing of short text includes cleaning, word segmentation, etc., which obtains the bag of words that make up these short sentences. Each word in the bag of words can be viewed as a measure in a short text feature vector. If the word appears in the short sentence, it is recorded as 1, otherwise it is recorded as 0. Thus, each short text can be represented as a vector in space x, which is shown in formula 4:

$$x = (t_1, t_2 \ldots \ldots t_i \ldots \ldots t_m) \tag{4}$$

where m is the total number of words in the bag of words, $t_i$represents whether the text contains the ith word, and if so, $t_i = 1$, or $t_i = 0$.

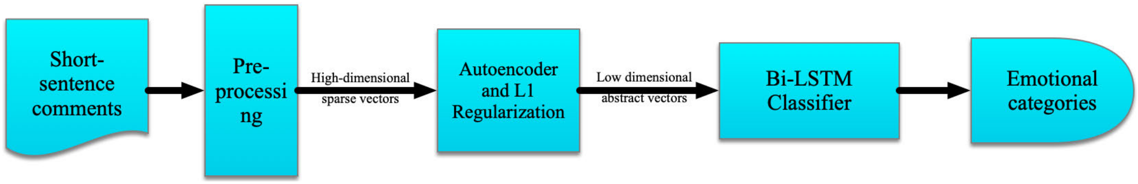

FIGURE 4. Affective computing flowchart.

1) AUTOENCODER

After receiving input vector x, the basic autoencoder first changes it linearly, and then obtains a encoding result y with the process of the activation function. Let the sigmoid function denote the activation function, and the computation is demonstrated in formula 5. The refactored vector z is obtained by decoding y, which is shown in formula 6.

$$y = f_\theta(x) = s(Wx + b) \tag{5}$$

$$z = g_{\grave{\theta}}(y) = s\left(\grave{W}y + \grave{b}\right) \tag{6}$$

The encoding parameter is $\theta = \{W, b\}$, and the decoding parameter is $\grave{\theta} = \left\{\grave{W}, \grave{b}\right\}$. W is a weight matrix of $d' * d$, while W' is the transpose of W, i.e., $W' = W^T$, where b and b' represent bias vectors.

The learning process of the autoencoder is unsupervised, and the target of optimization is to make the refactored vector z restore the input vector x as much as possible, i.e., to minimize the loss caused by the reconstruction, and to obtain the optimal parameters θ^* and $\theta^{*\prime}$. This process can be expressed by formula 7.

The loss function used here is Kullback-Leibler dispersion shown as formula 8.

$$\theta^*, \theta^{*\prime} = \arg\min L(x, z) = \arg\min L\left(x, g_{\grave{\theta}}(f_\theta(x))\right) \tag{7}$$

$$L(x, z) = KL(x \| z) \tag{8}$$

The autoencoder is trained using the classic stochastic gradient descent algorithm. In every iteration, formula 9 is used to update the weight matrix as follows:

$$W \leftarrow W - l \times \frac{\partial L(x, z)}{\partial W} \tag{9}$$

where L is for the learning rate, b and b' are updated in the same way. Coding and decoding of autoencoder can extract the features of text information, and by learning process and error control, the main features of the input text can be shown in the output.

2) L_1 REGULARIZATION

The powerful nonlinear expression ability of the autoencoder often leads to over-fitting of the input data, i.e., the features unique to individual objects are also fully described. The structural differences among short texts are significant, which cannot manifest the public distribution of short text, and the generalization ability of the trained model is inferior. Considering the problems above, in this paper, we constrain the learning ability of autoencoder.

L_1 regularization is a commonly used variable selection method, which is widely applied to model improvement. In this paper, the absolute value function of the autoencoder coefficient is used as a penalty term to compress its quantity of coefficient, and it automatically compresses the coefficient of the smaller absolute value to 0, so as to guarantee the sparseness of the parameters in the algorithm, and to avoid overlearning the non-significant characteristics of short texts. Specifically, we adjust the previous formula 8 to formula 10 and 11, which are calculated as follows:

$$L(x, z) = KL(x \| z) + Lasso(\theta) \tag{10}$$

$$Lasso(\theta) = \lambda \sum_{j=0}^{|\theta|} |\theta_j| \tag{11}$$

where λ is the parameter of the L_1 regularization. The larger the value, the greater the penalty, the sparser the training results will be. The value of λ needs to be adjusted multiple times according to the actual data so as to achieve the balance of fitting ability and generalization ability.

3) Bi-LSTM NETWORK

LSTM is a deep neural network classifier with sequential processing capabilities, but still needs improvement for MOOC review texts, since it cannot encode back-to-front information. For example, "The teacher has an accent, which is terrible, and is not as good as the other teachers". In this sentence, "which is terrible" is a modification of "accent". However, it is difficult for one-way LSTM to capture the information in this way. Therefore, this paper uses a Bi-LSTM to fully capture the context information in order to understand the semantics of the current words comprehensively. The network structure is shown in the Figure 5.

It can be seen that the output H_n of the Bi-LSTM is sliced and merged by the one-way LSTM and the reverse LSTM:

$$H_n = h_n + h'_n \tag{12}$$

In this research, the comments are grouped into different topics in advance, and then the binary classification models are trained under each topic to judge whether a comment is positive or negative.

D. MULTI-DIMENSIONAL EVALUATION OF MOOC

After detecting the topics and emotions of MOOC reviews, we utilize these results to establish a model of curriculum

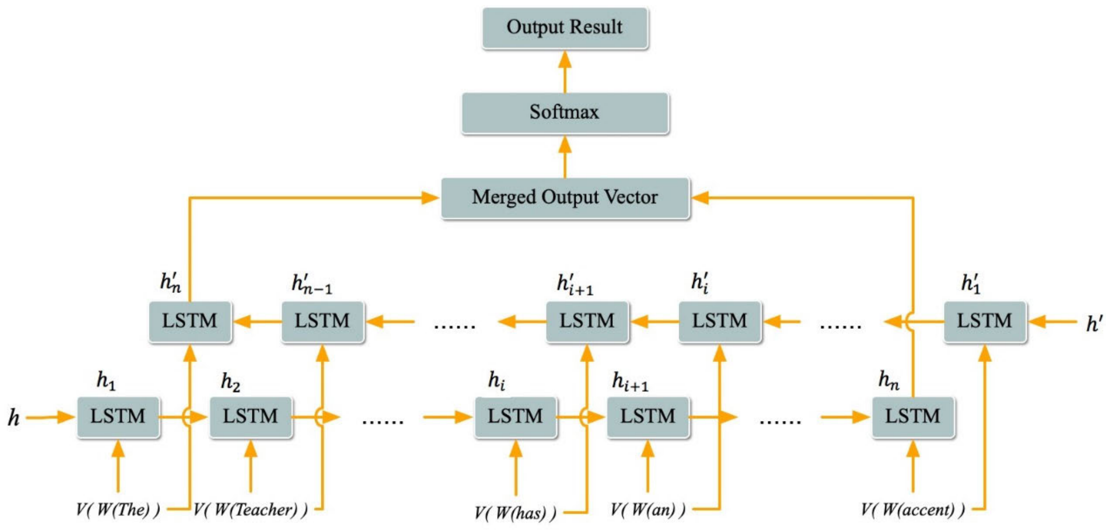

FIGURE 5. Bi-LSTM network structure.

evaluation in the following function:

$$Score_m = \frac{\sum positive_i * \theta_{p,i}}{\sum positive_i * \theta_{p,i} + \sum negative_j * \theta_{n,j}} * \lambda \quad (13)$$

where, $Score_m$ represents the quantitative score of mth topic, $positive_i$ stands the score of the comments of positive polarity and $negative_j$ for the comments of negative polarity, and $\theta_{p,i}$ is defined as the weight of the ith positive comment, and $\theta_{n,j}$ is the weight of the jth negative comment. To simplify the process, we assign both $\theta_{p,i}$ and $\theta_{n,j}$ to be 1. Furthermore, $\lambda$is the score weight which lets $Score_m \in [0, 5]$.

E. COURSE RECOMMENDATION MODEL FOR STUDENTS

According to the emotions scores of the students' comments, we can identify their preferences. If the majority of the comments are positive on a topic, and part students give negative results, it means that this student is more demanding on the previously mentioned topic. Thus, it is necessary to do weighted processing in the inner product computation of students and course comments. The weight coefficient W_i is calculated as follows:

$$W_i = \begin{cases} \log\left(\frac{\sum (S > \bar{S}_i)}{\sum (S < \bar{S}_i) + 1} + 1\right) & S < \bar{S}_i \\ \log\left(\frac{\sum (S < \bar{S}_i)}{\sum (S > \bar{S}_i) + 1} + 1\right) & S > \bar{S}_i \end{cases}$$

where S is the score given by the student to the indicator i, and S_i represents the average score for all participants of indicator i. By using the weight above, we can highlight the concerns of the students who are different from others, and make the matching results more closely tied with the students' interesting index.

We utilize the dot product formula of emotional preference vector and course score vector to calculate the matching degree between participants and courses:

$$\varphi = \frac{\sum_{i=1}^{k} W_i * Score_{s,i} * Score_{c,i}}{\lambda^2} \quad (14)$$

where k is the total number of dimension, W_i is the weight of indicator i, and $Score_{s,i}$ is student scores of indicators in dimension i, and $Score_{c,i}$ is defined as course scores of indicators in dimension i, and λ is the score base. The courses with the highest scores can be recommended as they meet the students' preferences.

IV. EXPERIMENTS AND RESULTS ANALYSIS

This part will introduce the experiment from 5 aspects: data collection, data pre-processing, topic discovery, affective computing, and course recommendation.

A. DATA SOURCE AND DATA PRE-PROCESSING

We use the Beautiful Soup and Selenium Topical Crawler to crawl course information and comment data from the ''iCourse'' Chinese University MOOC website within one year (April 2018 to March 2019). The data includes detailed information about the course: course name, institution, instructor, starting time, course description, number of enrollments, course reviews, and course review information: reviewer username, the time when the review was posted, the order in which the course started, the text of the review and the number of positive reviews, etc.

In the data cleaning stage, structured data are organized by the steps of segmenting words, deleting stop words, and removing low-frequency words, etc. After data cleaning, 584369 short sentences were extracted from the original data of more than 140,000 comments. The changes in the volume of data are shown in Figure 6.

Because of the difficulty to label all the comments, we extract 48505 comments under the five topics after LDA

TABLE 3. Topic- word probability matrix.

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
Probability of words	Probability of words	Probability of words	Probability of words	Probability of words
Professor 0.078	Class 0.047	Think 0.044	Platform 0.025	Knowledge 0.021
Teach 0.056	Method 0.039	Exercise 0.036	Video 0.024	Concept 0.016
Professional 0.049	Novel 0.037	Case 0.018	Courseware 0.020	Thoughts 0.016
Knowledge 0.048	Detailed 0.020	Homework 0.015	Document 0.020	Algorithm 0.013
Mode 0.032	Funny 0.013	Knowledge Point 0.015	Vision 0.020	Programming 0.012
Instruct 0.031	Subject 0.012	Exam 0.015	Clear 0.016	Network 0.012
Theory 0.024	Case 0.010	Easy 0.015	Version 0.013	Music 0.011
Funny 0.023	Start-up 0.010	Arrangement 0.014	Discussion 0.011	Art 0.011
Humor 0.016	Practical 0.010	Finish 0.012	Sound 0.011	Instance 0.011
Thorough 0.012	Boring 0.009	Important 0.012	App 0.009	Artificial Intelligence 0.009

TABLE 4. Sentence -topic probability matrix.

INDEX	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
1	0.8647873	0.0333363	0.0335977	0.0349431	0.0333356
2	0.9330129	0.0167053	0.0167466	0.1670152	0.0168335
……	…….	…….	……	……	……
584369	0.0669501	0.7322530	0.0669958	0.0666669	0.0671341

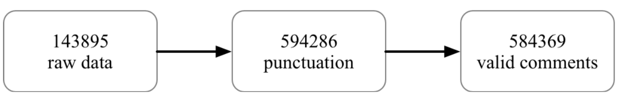

FIGURE 6. Changes in the volume of comments in data cleaning.

and labeled the comments as positive or negative manually. We also noticed that providing emotional values will be better, but limited by the subjectivity of artificial labeling, we just did the polarity classification. We alleviated this problem by using Eq. (13) to quantify the score of each topic.

B. TOPIC ANALYSIS OF COMMENTS

The word segmentation set obtained via the data cleaning phase is input to the short-sentence LDA model. According to the model learning steps, after multiple sampling and repeated testing, the prior parameters α and β are respectively determined to be 0.1 and 0.2. When the quantity of topics K Equals 5, the experiment of topic-word probability matrix in Table 3 achieves the best performance. Here, the corresponding sentence-topic probability matrix is demonstrated in Table 4.

Based on the probability distribution of the words under each topic in Table 3, the comments can be divided into five categories, which are the comments on instructor, course content, course assessment, MOOC platform, and hot courses.

Topic 1: This topic is about the instructors and their lecture-delivering ways. The words "professor", "professional" and "knowledge" indicate the professional and academic level of the teachers. This topic is mostly about the students' recognition and praise of the instructors.

Topic 2: This topic is about course content. Words such as "novel", "detailed", "interesting", "practical", indicate that MOOC courses are different from offline courses in arrangement, and online courses tend to students' stimulate innovative thinking among students. It is also found that some students express inner feelings about the course content, which implies that colleges and universities should offer more innovative and practical courses.

Topic 3: This topic is about course assessment, with focuses on knowledge content, homework completion, and exam schedule, etc. Only those who submit their assignments on time and who have passed the exam are eligible for a certificate of the course, and some courses offer valuable credit certification.

Topic 4: This topic is about MOOC users' experiences, and the users' experiences are essential for teachers to improve their teaching effects and stimulate students' interest in learning. Highlighted such as ''vision'', ''clear'' and ''sound'' in the students' comments reflect that the students are concerned about not only the course content when viewing the videos and previewing the courseware, but also the quality of the course videos and the picture resolution.

Topic 5: The words under this topic are mostly professional words, e.g., ''algorithm'', ''programming'', ''network'', ''music'', ''art'', ''artificial intelligence'', which manifests the students' concerns. It can be seen that most of the MOOC courses are related to computers, music, and art.

According to the topic analysis, the MOOC curriculum evaluation indicators are constructed as instructor, course content, course assessment, MOOC platform, and hot courses. The short sentences are classified into five topics based on probability. Therefore, we can complete the sentiment analysis of the comments on different topics. The result is shown in Table 4.

C. SENTIMENT ANALYSIS OF COMMENTS

1) Bi-LSTM NETWORK LAYER

The Bi-LSTM core code is written in JetBrains PyCharm 2018.2.4 x64, run in Ubuntu 16.04 system and by PC with 32 Intel (R) Xeon (r) CPUS@2.6 GHz / 64 GB RAM. The text classification network is constructed using the Keras, a high-level neural network API that encapsulates the Google open source framework TensorFlow-GPU v1.5.0.

The model training environment is NVIDIA GeForce GTX 1080/8GB VRAM. The quantity of initial hidden layer neurons is assigned as 128. To avoid over-fitting, the dropout rate is set as 0.2, and recurrent dropout is also set as 0.2.

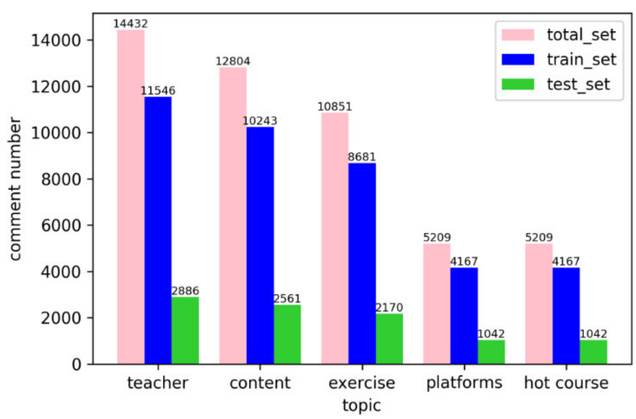

FIGURE 7. Size of comments on each topic.

2) MODEL EVALUATION

We collect comment samples under five topics and label train_set test_set as shown in Figure 7. For comparing

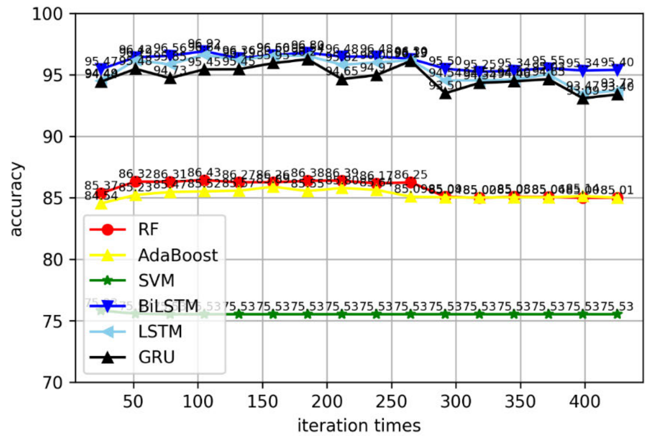

FIGURE 8. Classifier Accuracy with Different Iterations.

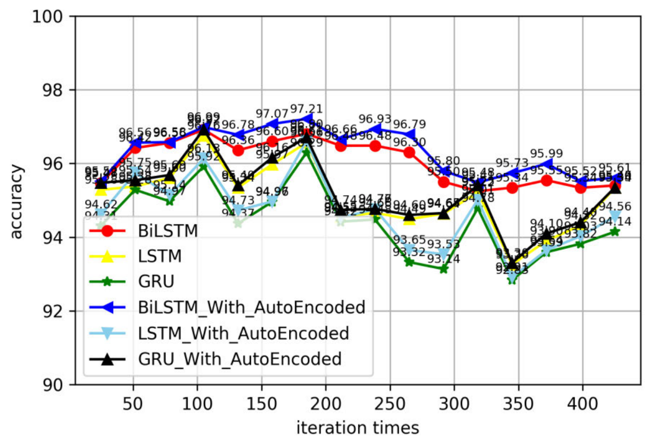

FIGURE 9. Accuracy of different encoding methods with different iterations.

different text classifiers, Random Forest, AdaBoost, SVM, LSTM, and GRU are chosen in the experiment. Furthermore, we choose the 0-1 word vector and the autoencoder word vector for contrast between the two vector representations. Figure 8 shows performances of models with different times of iteration on topic ''Instructor''. The performance of 0-1 encoder and autoencoder with different times of iteration under topic ''Instructor'' is demonstrated in Figure 9.

In the experiment, the Bi-LSTM network used the autoencoder to reach the maximum accuracy of 97.21% at 180 times of iteration. Thus, this model can be used to complete the emotional polarity annotation of the remaining sentences.

D. COMMENT QUANTIFICATION

After completion of the emotional polarity for all sentences, course comments can be quantified under several topics. In the experiment, we quantified the comments on the course ''Python programming'' offered by Peking University, Beijing Institute of Technology, and Harbin Institute of Technology. The results are shown in Table 5 and Figure 10.

Figure 10 indicates that the courses offered by the three universities and the MOOC platform have both received

TABLE 5. Comparison of python course.

Colleges	Instructors	Course content	Course assessment	MOOC Platform	Hot Courses
PKU	4.85	4.72	4.45	4.7	4.89
BIT	4.86	4.75	4.63	4.8	4.6
HIT	3.8	4.07	4.25	4.63	4.5

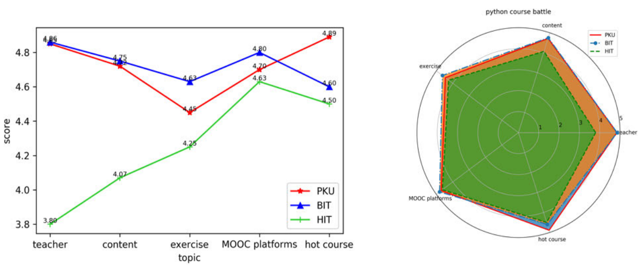

FIGURE 10. Comparison of python course.

positive feedback from participants. According to comments, for the courses offered by Peking University and Beijing Institute of Technology, the teachers' teaching quality is first-class, the communication with the students is effective, and the curriculum arrangement is acceptable. It also shows that the participants have heated discussions on hot courses, which reflects the participants' passion for Python programming. Though popular with students, the courses of Peking University emphasize more to the difficulty of passing the final exam, so the score of the indicator "Course Assessment" is lower than other indicators. However, noted that the courses offered by Harbin Institute of Technology are obviously inferior to those of the other two universities, and the MOOC reviews suggested that the professor is senior in age, still use Python 2.0 is out of date.

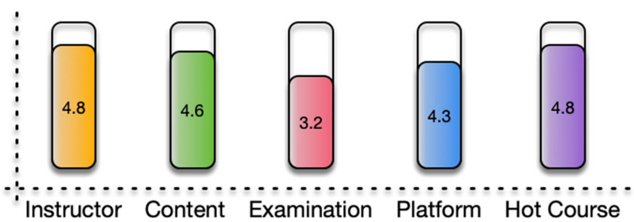

FIGURE 11. The overall scores of five indicators of MOOC.

V. EXTENDED EXPERIMENTS

A. OVERVIEW OF MOOC SENTIMENT ANALYSIS

After the sentimental analysis of MOOC comments, we have done some calculating analysis of the data. Firstly, we detect the comprehensive satisfaction degree of MOOC through the quantification of the five indicators, as shown in Figure 11.

According to the analysis results, students are satisfied with the 'instructor', 'course content' and 'hot course' indicators, but dissatisfied with the 'course assessment', and 'MOOC platform' indicators. In general, MOOCs have received positive reviews from students, reflecting the huge success MOOCs have earned in recent years. Especially, the indicator of instructor gets the highest evaluation, which is closely related to the strict teacher examination and approval system of MOOC. In order to ensure the quality of teaching, teachers should first apply to education institutions before opening courses. The relevant institutions will conduct a detailed review of teachers' professional knowledge, teaching ability, and course content, etc. Then after meeting the requirements the courses can only be opened on the platform. In addition to teachers, hot course indicators are also popular among students. We explore those comment data and find that under such indicators, students often comment on computers, programming, art, and other related contents. The reason is that these courses are more suitable to be offered in the form of network courses. The knowledge points of these courses are many and scattered, and whether or not to take these courses are closely related to students' interests, and students need to learn independently through the learning process. Plus, the learning environment is also not subject to the time and space constraints of campus teaching. Both internal and external reasons contribute to the higher popularity of these courses than others.

B. THE MOST CONCERNED CONTENT

At the same time, it is worth noting that in the overall sentiment analysis, the students are not satisfied with the course assessment and the MOOC platform, and the students give 3.2 and 4.3 scores. We have done a sample analysis of these negative comments to explore what the students complain about most. Some comments fragments are as follows:

"Lack of downloadable written courseware.", "Why can't the slides play?", "The course content is very good, but sometime the video can't open, what a trouble!", "To be honest, actually, take out money also can buy the certificate of completion, pure individual idea.", "The website server is too bad.", etc.

TABLE 6. Some students' comments quantification results.

Students	Instructor	Content	Assessment	MOOC Platform	Hot Course
bwt1997	4.63	4.54	2.39	4.88	3.54
Wings 0 Stars	4.23	4.67	3.39	4.52	4.14

TABLE 7. The matching degree between students and universities.

Students	University	Course Indicators	Matching Degree
bwt1997	PKU	$< 4.85,\ 4.87,\ 4.45,\ 4.70,\ 4.89 >$	**3.755**
	BIT	$< 4.86,\ 4.75,\ 4.63,\ 4.80,\ 4.60 >$	3.736
	HIT	$< 3.80,\ 4.07,\ 4.25,\ 4.63,\ 4.50 >$	3.349
Wings 0 Stars	PKU	$< 4.85,\ 4.87,\ 4.45,\ 4.70,\ 4.89 >$	4.120
	BIT	$< 4.86,\ 4.75,\ 4.63,\ 4.80,\ 4.60 >$	**4.130**
	HIT	$< 3.80,\ 4.07,\ 4.25,\ 4.63,\ 4.50 >$	3.717

From the negative comments, we know that students complained about the course assessment and MOOC platform, including the hardware aspects such as unsmooth video playing, low sound definition, and unstable server. A small number of students also expressed their concern about the surrogate test-taking problems in online courses. These pain points require universities and MOOC platform to take appropriate measures together in their later teaching practice, strengthen the investment of hardware facilities and the supervision of course assessment, so as to bring students a better learning environment.

C. THE CHARACTERISTICS OF HIGH-QUALITY COURSE

For the comment data, we sort the courses according to their emotion scores, and select the top 10 courses, as shown in Figure 12, to explore their characteristics.

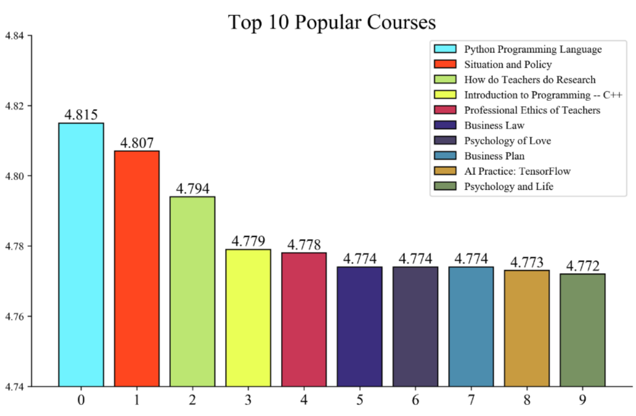

FIGURE 12. The course score ranks top 10 courses.

Based on the above information, it can be seen that MOOC students prefer computer courses, most of which are programming courses, such as Python and C++ programming, etc., and they also pay attention to exchange what they have learned with each other. The following are comments from the Python Programming courses:

— GongXiao Feng 3: "Teacher Song's explanation is lively, interesting and fascinating, which is really quite good. It's very important that the presentation is conceptual and easy for students to understand."

— A bulu: "This course is well organized, starting from the simple to the deep, and is a basic course worth watching for beginners in Python. Only part of the video upload is incomplete, or there is no voice, a little flawed."

It is important to note that in the "Situation and Policy" course holding by Nanjing Normal University courses, only the 4th opening of this course, there are nearly 7000 comments, besides, the students' nicknames are all like NJU-DDDDDDD. From other information, we know that the enrollments occupy a large proportion of the MOOC course evaluation. And the university also has hard rubrics on the comment quantities, so we guess it is a vicious behavior, namely in order to achieve teaching index, teachers command students must choose a course and must comment on that course.

D. STUDENTS-COURSE MATCHING MODEL

According to the course recommendation model in Section E of Part III, we construct the quantitative score distribution of students on the course evaluation system by using the topic-word distribution matrix and the multi-dimensional evaluation model, as shown in Table 6.

Using the formula (14), we calculated the matching degree φ, the results are shown in Table 7.

Based on the matching degree analysis, when using the model in this paper to recommend students "BWT1997" and

"Wings 0 Stars" for Python course learning. For the former, the student is more concerned about the instructors (4.63) and MOOC platform (4.88), and are not interested in course assessment (2.39), then Peking University Python course is suitable for him or her to learn. For the latter, the students are more concerned about the course content (4.67) and MOOC platform (4.52), so the Python course of Beijing Institute of Technology is suitable for him or her to learn.

VI. CONCLUSION

This paper has established evaluation indicators from five aspects, which include instructors, course content, course assessment, MOOC platforms, and hot courses. Based on proposed indicators, we scientifically and reasonably quantify numerical scores of each index, to provide students and educators with an intuitive reference. This research contributes to extracting students' concerns from their reviews so as to assist MOOC managers and universities in improving most of the platform infrastructure and service experience. However, in the comment sentiment analysis, only positive and negative dimensions were considered, with no consideration of neutral emotion or emotional intensity, which is to be covered in future research. Moreover, in future research, a system will be designed that can detect the topic descriptions and emotional polarity changes brought by the new comments in time to realize the dynamically updating the scores of each indicator.

REFERENCES

[1] O. Almatrafi, A. Johri, and H. Rangwala, "Needle in a haystack: Identifying learner posts that require urgent response in MOOC discussion forums," *Comput. Educ.*, vol. 118, pp. 1–9, Mar. 2018.

[2] J. Reich and J. A. Ruipérez-Valiente, "The MOOC pivot," *Science*, vol. 363, no. 6423, pp. 130–131, Jan. 2019.

[3] J. A. Ruiperez-Valiente, M. Jenner, T. Staubitz, X. Li, T. Rohloff, S. Halawa, C. Turro, Y. Cheng, J. Zhang, I. Despujol, and J. Reich, "Macro MOOC learning analytics: Exploring trends across global and regional providers," in *Proc. 10th Int. Conf. Learn. Anal. Knowl. (LAK)*, 2020, pp. 518–523.

[4] D. Shah. (2020). *By the Numbers: MOOCS in 2020*. The Report by Class Central. [Online]. Available: https://www.classcentral.com/report/mooc-stats-2020

[5] G. S. Stump, J. Deboer, J. Whittinghill, and L. Breslow, "Development of a framework to classify MOOC discussion forum posts: Methodology and challenges," in *Proc. NIPS Workshop Data Driven Educ.*, 2013, pp. 1–20.

[6] O. Almatrafi and A. Johri, "Systematic review of discussion forums in massive open online courses (MOOCs)," *IEEE Trans. Learn. Technol.*, vol. 12, no. 3, pp. 413–428, Jul. 2019.

[7] W. Yan, N. Dowell, C. Holman, S. S. Welsh, H. Choi, and C. Brooks, "Exploring learner engagement patterns in teach-outs using topic, sentiment and on-topicness to reflect on pedagogy," in *Proc. 9th Int. Conf. Learn. Anal. Knowl.*, Mar. 2019, pp. 180–184.

[8] H. M. Dai, T. Teo, and N. A. Rappa, "Understanding continuance intention among MOOC participants: The role of habit and MOOC performance," *Comput. Hum. Behav.*, vol. 112, Nov. 2020, Art. no. 106455.

[9] H. Li, M. K. Kim, and Y. Xiong, "Individual learning Vs. Interactive learning: A cognitive diagnostic analysis of MOOC students' learning behaviors," *Amer. J. Distance Educ.*, vol. 34, no. 2, pp. 121–136, Apr. 2020.

[10] X. Lu, X. W. Liu, and W. Zhang, "Diversities of learners' interactions in different MOOC courses: How these diversities affects communication in learning," *Comput. Educ.*, vol. 151, Jul. 2020, Art. no. 103873.

[11] W. Zou, X. Hu, Z. Pan, C. Li, Y. Cai, and M. Liu, "Exploring the relationship between social presence and learners' prestige in MOOC discussion forums using automated content analysis and social network analysis," *Comput. Hum. Behav.*, vol. 115, Feb. 2021, Art. no. 106582.

[12] X. Zhai and L. Lin, "Research into development dilemmas of MOOC in China's higher education and its countermeasures," *Academics*, vol. 190, no. 3, pp. 262–268, 2014.

[13] S. Cooper and M. Sahami, "Reflections on Stanford's MOOCs," *Commun. ACM*, vol. 56, no. 2, pp. 28–30, 2013.

[14] A. Deshpande and V. Chukhlomin, "What makes a good MOOC: A field study of factors impacting student motivation to learn," *Amer. J. Distance Educ.*, vol. 31, no. 4, pp. 275–293, 2017.

[15] Y. Nie, H. Luo, and D. Sun, "Design and validation of a diagnostic MOOC evaluation method combining AHP and text mining algorithms," *Interact. Learn. Environ.*, vol. 28, pp. 1–13, Aug. 2020, doi: 10.1080/10494820.2020.1802298.

[16] T. Levine, "Stability and change in curriculum evaluation," *Stud. Educ. Eval.*, vol. 28, no. 1, pp. 1–33, Jan. 2002.

[17] J. R. Drake, M. T. O'Hara, and E. Seeman, "Five principles for MOOC design: With a case study," *J. Inf. Technol. Educ., Innov. Pract.*, vol. 14, no. 14, pp. 125–143, 2015.

[18] P. Miranda, P. Isaias, and S. Pifano, "Model for the evaluation of MOOC platforms," in *Proc. 8th Int. Conf. Educ., Res. Innov. (ICERI)*, 2015, pp. 1199–1208.

[19] Z. Zhao, Q. Wu, H. Chen, and C. Wan, "Learning quality evaluation of MOOC based on big data analysis," in *Proc. Int. Conf. Smart Comput. Commun. (SmartCom)*, 2016, pp. 277–286.

[20] T. R. Liyanagunawardena, A. A. Adams, and S. A. Williams, "MOOCs: A systematic study of the published literature 2008-2012," *Int. Rev. Res. Open Distrib. Learn.*, vol. 14, no. 3, pp. 1–521, 2013.

[21] N. F. Sa'don, R. A. Alias, and N. Ohshima, "Nascent research trends in MOOCs in higher educational institutions: A systematic literature review," in *Proc. Int. Conf. Web Open Access to Learn. (ICWOAL)*, Nov. 2014, pp. 1–4.

[22] S. Cunningham-Nelson, M. Laundon, and A. Cathcart, "Beyond satisfaction scores: Visualising student comments for whole-of-course evaluation," *Assessment Eval. Higher Educ.*, vol. 45, pp. 1–16, Aug. 2020.

[23] K. F. Hew, "Promoting engagement in online courses: What strategies can we learn from three highly rated MOOCS," *Brit. J. Educ. Technol.*, vol. 47, no. 2, p. 320, 341, 2016.

[24] D. Clow, "MOOCs and the funnel of participation," in *Proc. 3rd Int. Conf. Learn. Anal. Knowl. (LAK)*, 2013, pp. 185–189.

[25] S. Türkay, H. Eidelman, Y. Rosen, D. Seaton, G. Lopez, and J. Whitehill, "Getting to know English language learners in MOOCs: Their motivations, behaviors, and outcomes," in *Proc. 4th ACM Conf. Learn. Scale*, Apr. 2017, pp. 209–212.

[26] D. W. Jordan, "Re-thinking student written comments in course evaluations: Text mining unstructured data for program and institutional assessment," Ph.D. dissertation, Dept. Educ., California State Univ., Long Beach, CA, USA, 2011.

[27] Z. Liu, W. Zhang, J. Sun, H. N. H. Cheng, X. Peng, and S. Liu, "Emotion and associated topic detection for course comments in a MOOC platform," in *Proc. Int. Conf. Educ. Innov. through Technol. (EITT)*, Sep. 2016, pp. 15–19, doi: 10.1109/EITT.2016.11.

[28] J. Weng, W. Gan, G. Ding, Z. Tian, Y. Gao, and J. Qiu, "SESM: Emotional social semantic and time series analysis of learners' comments," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2020, pp. 20–28.

[29] W. Xing, H. Tang, and B. Pei, "Beyond positive and negative emotions: Looking into the role of achievement emotions in discussion forums of MOOCs," *Internet Higher Educ.*, vol. 43, Oct. 2019, Art. no. 100690.

[30] X. Peng and Q. Xu, "Investigating learners' behaviors and discourse content in MOOC course reviews," *Comput. Educ.*, vol. 143, Jan. 2020, Art. no. 103673.

[31] A. Onan, "Sentiment analysis on massive open online course evaluations: A text mining and deep learning approach," *Comput. Appl. Eng. Educ.*, vol. 29, pp. 1–18, May 2020.

[32] K. F. Hew, X. Hu, C. Qiao, and Y. Tang, "What predicts student satisfaction with MOOCs: A gradient boosting trees supervised machine learning and sentiment analysis approach," *Comput. Educ.*, vol. 145, Feb. 2020, Art. no. 103724.

[33] Z. Kastrati, A. S. Imran, and A. Kurti, "Weakly supervised framework for aspect-based sentiment analysis on students' reviews of MOOCs," *IEEE Access*, vol. 8, pp. 106799–106810, 2020.

[34] Z. Kastrati, B. Arifaj, A. Lubishtani, F. Gashi, and E. Nishliu, "Aspect-based opinion mining of students' reviews on online courses," in *Proc. 6th Int. Conf. Comput. Artif. Intell.*, Apr. 2020, pp. 510–514.

[35] M. L. Barrón Estrada, R. Zatarain Cabada, R. Oramas Bustillos, and M. Graff, "Opinion mining and emotion recognition applied to learning environments," *Expert Syst. Appl.*, vol. 150, Jul. 2020, Art. no. 113265.

[36] M. Misuraca, G. Scepi, and M. Spano, "Using opinion mining as an educational analytic: An integrated strategy for the analysis of students' feedback," *Stud. Educ. Eval.*, vol. 68, Mar. 2021, Art. no. 100979.

[37] A. Koufakou, J. Gosselin, and D. Guo, "Using data mining to extract knowledge from student evaluation comments in undergraduate courses," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 3138–3142.

[38] K. F. Hew and W. S. Cheung, "Students' and instructors' use of massive open online courses (MOOCs): Motivations and challenges," *Educ. Res. Rev.*, vol. 12, pp. 45–58, Jun. 2014.

[39] C. M. Stracke, E. Tan, A. M. Texeira, B. Vassiliadis, M. D. C. T. Pinto, A. Kameas, and C. Sgouropoulou, "Gap between MOOC designers' and MOOC learners' perspectives on interaction and experiences in MOOCs: Findings from the global MOOC quality survey," in *Proc. IEEE 18th Int. Conf. Adv. Learn. Technol. (ICALT)*, Jul. 2018, pp. 1–5, doi: 10.1109/ICALT.2018.00007.

[40] H. B. Shapiro, C. H. Lee, N. E. W. Roth, K. Li, M. Çetinkaya-Rundel, and D. A. Canelas, "Understanding the massive open online course (MOOC) student experience: An examination of attitudes, motivations, and barriers," *Comput. Educ.*, vol. 110, pp. 35–50, Jul. 2017.

[41] J. Maldonado-Mahauad, M. Pérez-Sanagustín, R. F. Kizilcec, N. Morales, and J. Munoz-Gama, "Mining theory-based patterns from big data: Identifying self-regulated learning strategies in massive open online courses," *Comput. Hum. Behav.*, vol. 80, pp. 179–196, Mar. 2018.

[42] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.

[43] M. Jelodar, Y. Wang, C. Yuan, X. Feng, X. Jiang, and Y. Li, "Latent Dirichlet allocation (LDA) and topic modeling: Models, applications, a survey," *Multimedia Tools Appl.*, vol. 78, pp. 15169–15211, Nov. 2018.

CONG QI is currently a Master Student with the School of Information and Security Engineering, Zhongnan University of Economics and Law, Wuhan, China. His research interests include user profile knowledge graph and machine learning.

SHUDONG LIU received the Ph.D. degree in computer science and engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 2015. He is currently a Lecturer with the School of Information and Security Engineering, Zhongnan University of Economics and Law, Wuhan, China. His research interests include social networks analysis, recommender systems, and machine learning.