International Journal of Educational Development
ISSN Print: 3104-4263
ISSN Online: 3104-4271
CD ISSN: 3104-428X
https://doi.org/10.63313/IJED.9092

Automatic detection of inconsistencies between MOOC text reviews and ratings based on BERT+RNN model

Guoxia Dong. Tingmei Wang*. Yunfei Xu
Beijing Union University, Beijing 100101, China
Email: yykjttingmei@buu.edu.cn

How to cite this paper: Dong, G., Wang, T., & Xu, Y. (2026). Automatic detection of inconsistencies between MOOC text reviews and ratings based on BERT+RNN model. International Journal of Educational Development, 4(1), 54–69. ISSN Print: 3104-4263, ISSN Online: 3104-4271.
https://doi.org/10.63313/IJED.9092
Published: 2026-07-13

Copyright © 2026 by author(s) and Erytis Publishing Limited.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
CC BY 4.0

Abstract

To address the inconsistency between text reviews and star ratings on Chinese university MOOC platforms, this paper proposes a BERT+RNN-based automatic detection and correction model. The study collected 343519 course reviews and constructed a five-category sentiment labeling system (class 1–class 5) through manual annotation. The performance of three deep learning models, namely, a CNN, an RNN, and a BERT+RNN, was compared. Experimental results show that the BERT+RNN model achieved an accuracy of 93.7% on the test set, outperforming other models. The model was used to automatically correct the ratings of all reviews, with a total of 27794 reviews (8.1% of the total), representing different ratings between the original and model-generated annotations. Subject and course satisfaction rankings were analyzed based on the corrected ratings. This study provides an effective technical approach for improving the reliability of MOOC evaluation data and platform governance, and has strong practical application value.

Keywords

MOOCs; Text Sentiment Analysis; BERT Model; Inconsistency Detection

1. Introduction

With the deep integration of information technology and education, massive open online courses (MOOCs) have become an important part of the global higher education ecosystem[1-3] . Local platforms such as Chinese University MOOC have gathered a large number of high-quality course resources, providing flexible and open lifelong learning access for hundreds of millions of learners[4-5]. In this context, the course evaluation system — especially the textual reviews and five-star ratings provided by learners — serves as the core basis for reflecting course quality, supporting subsequent learners in course selection decisions, and helping instructors and platforms optimize teaching content. Its authenticity and reliability

are particularly critical[6-8].

However, observations of the Chinese University MOOC platform have revealed an inconsistency: a certain contradiction or divergence exists between the sentiment orientation of textual reviews and the quantitative star ratings. For instance, as shown in Table 1, some reviews that describe course advantages in detail and express gratitude are accompanied by extremely low one star ratings; conversely, some reviews containing criticism and complaints are paired with four or five star ratings. This phenomenon may stem from multiple factors: on the one hand, learners may forget to submit a rating, and the system assigns one star by default; on the other hand, some high ratings are accompanied by reviews with no substantive content, which may be hastily completed by learners to meet a minimum word requirement. Such inconsistency severely weakens the credibility and utility of course evaluation data and requires identification and correction.

Table 1. Examples of inconsistency between text reviews and star ratings in MOOC courses.

Course	text reviews	Star rating
Python Programming	I have deeply understood the joy of learning and experienced the sense of achievement from hard work. Thank you to the MOOC platform.	1
Python Programming	Teacher Song taught excellently, especially the classic cases, which enable us to learn by analogy and apply knowledge flexibly..	2
Detailed Explanation and Application Cases of Linear Algebra	It's really excellent.	3
Electronic Technology Application Experiments	It’s an online course, and it’s really tough without experimental equipment.	4
Acupuncture in 100 Days	Why isn’t the grade given yet? It’s already the next academic year.	5

Although existing studies often treat online review texts and star ratings as independent or consistent information sources, research on the detection of inconsistencies between them remains relatively limited. Faced with the massive volume of review data accumulated on MOOC platforms, manual verification one by one is not only time-consuming and labor-intensive but also difficult to guarantee the accuracy of results. Developing a technical model that can automatically and accurately identify inconsistencies between review texts and ratings in the Chinese University MOOC context has significant theoretical significance and practical value.Based on this, this paper constructs an automatic detection model for identifying and correcting inconsistencies between textual reviews and scores. Relying on this model, subsequent statistical analysis is carried out to obtain more accurate and valuable research conclusions.

2. Literature Review

2.1. Development of Text Sentiment Analysis Technology and Its Application in the MOOC Field

Text sentiment analysis is one of the core tasks in the field of Natural Language Processing (NLP). It aims to automatically identify, extract, analyze, and summarize the sentiment orientations, opinions, and emotions contained in text through computational techniques. Early studies mainly relied on sentiment dictionary-based methods and traditional machine learning-based methods [9]. The core of the dictionary-based method is to construct a sentiment lexicon consisting of a large number of sentiment words (e.g., “clear”, “confusing”), degree adverbs (e.g., “very”, “slightly”), and negation words (e.g., “not”, “lack of”). The overall sentiment orientation is determined by calculating the polarity and comprehensive sentiment score of sentiment words in the text via rules. However, such methods struggle with ironic expressions such as “The teacher’s lecture is very ‘thorough’” (used ironically, actually meaning perfunctory). Traditional machine learning-based methods treat sentiment analysis as a text classification problem. Text feature extraction is first performed using techniques such as the bag-of-words model and TF-IDF to generate high-dimensional sparse vectors, followed by training and prediction with classifiers such as Support Vector Machines and Naive Bayes. Liu et al. (2021) [10]applied the BosonNLP sentiment dictionary to conduct sentiment analysis on review texts from three mathematics courses on the MOOC platform, classifying reviews into positive, neutral, and negative categories. This sentiment model is relatively simple and fails to consider contextual semantics.

With the rise of deep learning, sentiment analysis has entered a new stage and solved the problem of semantic capture faced by traditional methods. Recurrent Neural Networks (RNN) and their variants LSTM[11] and GRU [12]can capture contextual dependencies and long-distance semantic correlations in text, with stronger modeling capabilities for transitional sentences. Sun (2023) [13]crawled course evaluation data of 11 courses from the Chinese University MOOC platform and built a Bi-LSTM model for sentiment classification (positive, negative, neutral). However, the model only performed three-class classification and did not subdivide more detailed emotional states. In addition, Convolutional Neural Networks (CNN) adopt multi-scale convolution kernels: for example, 2-gram kernels capture phrases such as “clear course materials”, and 3-gram kernels capture expressions such as “late response to questions”. For short comments such as “Great course!” and “Bad review”, 1-gram convolution kernels are used to improve processing efficiency. Chen et al. (2024)[14] extracted 102,184 reviews from the Class Central platform and trained classifiers including CNN to realize automatic classification of course evaluations into 10 semantic categories (e.g., course quality, instructor, assessment, etc.).

In recent years, pre-trained large language models represented by BERT (Bidirectional Encoder Representations from Transformers) [15,16]have adopted a two stage strategy: general pre-training + domain fine tuning.The BERT base model is first pre trained on general Chinese corpora such as Wikipedia and news data, and

then fine tuned on domain specific reviews, which significantly improves the model's recognition accuracy. Cavalcanti et al. (2023)[17] explored the application of the BERT model in classifying Portuguese teacher feedback texts, including task level, process level, self regulated, and self level feedback.Compared with Cavalcanti et al.'s (2020) [18]study using random forest as the classifier, the Cohen's kappa coefficient of this model increased by 35.71%.Wei and Yu (2023)[19] constructed the first publicly available Chinese MOOC review sentiment corpus based on pre trained language models, which currently only supports binary classification (positive/negative).

2.2. Research on Data Mining and Analysis of MOOC Corpus

Most analyses of MOOC course reviews remain at the level of topic mining. For example, Liu et al. (2017) [20]applied the LDA topic model to analyze course review posts and mined hot topics of concern to learners, finding that they paid particular attention to course content and teaching delivery. Zhang and Duan (2020)[21]extracted hot topics from learner reviews (video production, course content, knowledge volume, teaching style, learning experience) and calculated the sentiment polarity of each topic based on a sentiment dictionary to evaluate course quality. Chen et al. (2025)[22] used course evaluation data from the Class Central platform and automatically identified 10 topic categories (e.g., course quality, instructor, assessment, learning resources) with the BERT model, then analyzed differences in topic distribution across disciplines and rating groups.

2.3. Pain Points in Current Research

At present, most studies on course reviews from the Chinese University MOOC platform focus on topic mining. In sentiment classification tasks, existing research is mostly limited to binary classification (positive/negative) or three-class classification (positive/neutral/negative), and no study has yet conducted fine-grained analysis corresponding to five-star ratings. This paper aims to fill this gap. The core work is to construct an automated method for detecting inconsistencies between review texts and star ratings. Large-scale course review data are crawled from the Chinese University MOOC platform, and automatic scoring models are built based on LSTM, CNN, DPCNN, and BERT neural networks, respectively. The best-performing model is selected to identify and correct reviews where star ratings are inconsistent with textual sentiment.On this basis, the satisfaction levels of different disciplines and courses are further ranked to obtain the top five disciplines and courses. Experimental results show that the proposed model can significantly improve the detection performance of inconsistent reviews.

3. Research Design and Methods

3.1. Research Design and Procedure

Large-scale course review data were crawled from the China University MOOC platform. Invalid data with null values or only punctuation characters were removed; excessively long reviews were truncated to retain a maximum of 256 characters to adapt the input format for subsequent model training.Then A subset of the data was selected as the training set, manually inspected and annotated, after which automatic scoring models based on RNN, CNN, and BERT were constructed respectively. The model with the optimal performance was screened through model evaluation, then applied to the full sample of MOOC corpus to implement automatic star rating. By comparing the automatic ratings with the original ones, reviews with "inconsistency between star ratings and text sentiment" were identified and corrected.Finally, Based on the revised review data, the satisfaction levels of different disciplines and courses are ranked, and the top five disciplines and courses are finally obtained.

3.2. Data Collection

A Python web crawler was developed to collect a course review dataset from the China University MOOC platform, comprising a total of 343,519 entries. The dataset includes six key attributes: review ID, review time, reviewer ID, reviewer nickname, review content, and the corresponding course ID. The reviews employ a 5-star rating system, where scores range from 1 to 5 stars. A higher score indicates a higher level of student satisfaction with the course. The volume and distribution of the data are illustrated in Figure 1.

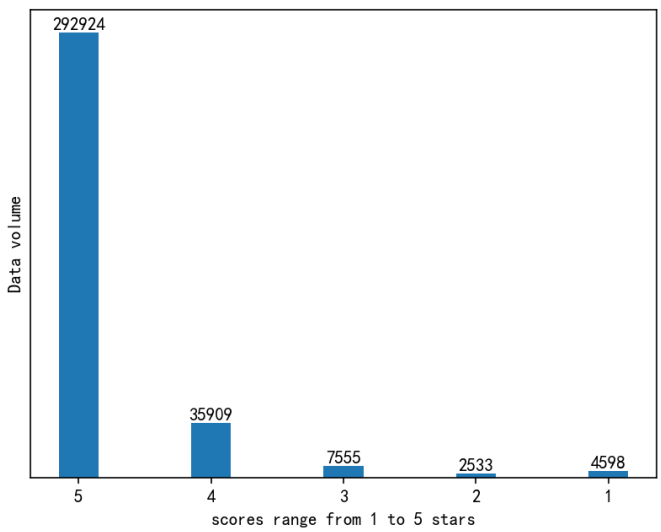


Figure 1. Distribution of MOOC course reviews collected in this study. They will be used for course satisfaction analysis

3.3. Model Selection and Evaluation Metrics

With the rapid development of deep learning and large language model technologies,

this paper selects three representative models: RNN, CNN, and BERT+RNN (to achieve better performance, the BERT model is improved by adding an RNN module). Automatic scoring models for MOOC course reviews are constructed based on these three models respectively, and their prediction performances are compared to determine the most suitable modeling method for the sample data, thereby improving the reliability of prediction results.

For model performance evaluation, two metrics in multi-class classification tasks are used: the confusion matrix and accuracy. The confusion matrix is a standard format for representing accuracy evaluation, expressed as an n×n matrix. Each column represents the predicted class, and the total number of each column indicates the number of data predicted as that class. Each row represents the true class of the data, and the total number of each row indicates the number of data instances belonging to that class. Accuracy is the ratio of the number of correct predictions to the total number of predictions made by the model, which can be calculated from the confusion matrix.

4. Experimental Procedure and Results

4.1. Annotation and Division of Training Samples

Here the annotation of training data and the division of the training set, validation set, and test set are carried out. It can be seen from Figure 1 that the data categories exhibit severe imbalance, while artificial intelligence models require the proportions of various types of data to be as balanced as possible. Approximately 3,000 entries were randomly selected from each category, with a total of 15,000 data entries, for manual inspection and annotation. The annotation rules are as follows:

class1: Completely negative information without any positive evaluation. Example: "Useless, purely unrealistic, disappointing."

class2: Mainly negative information with less intense emotion than class1. Example: "Well, the clarity is poor and the coherence is insufficient."

class3: Neutral stance without emotional attitude, only questions, consultations, suggestions, etc. Example: "Excuse me, teacher, when will the book be published?"

class4: Mainly positive information with less intense emotion than class5. Example:" I have gained an understanding of PPT beautification."

class5: Completely positive information without any negative content. Example: "I have gained a lot through learning. Thank you so much!"

The training samples were verified and re-annotated according to the above rules. Some examples are shown in Table 2.

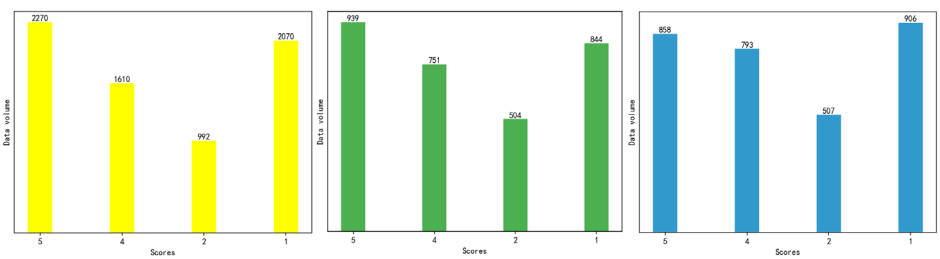
Table 2. Some examples for original scores of training samples & manually annotated scores.

According to the annotation guidelines, the original annotations are not accurate enough.

Examples	Original score	Manually score
To be honest, it is not suitable for beginners.	class1	class2
Can I get in touch with the one who issues the certificate?	class1	class3
It is very beneficial.	class2	class5
Could you stop asking me for evaluations all the time? It's really annoying.	class3	class1
Blah blah blah blah	class4	class1
Great! Great! Great!	class4	class5
If possible, please make the explanations for the after-class exercises simpler and clearer.	class5	class3
This course is pretty good.	class5	class4

Due to the strong subjectivity of scoring, this paper only corrects corpus that originally belongs to class1/2 but was labeled as class4/5 (or vice versa), without considering misclassifications between adjacent categories. In addition, since class3 is uncertain—it may contain both positive and negative evaluations and is highly subjective—we do not process corpus with the true category of class3.

The annotated data were divided into 6,942 training samples, 3,064 test samples, and 3,038 validation samples. The data distribution is shown in Figure 2. The training set was used to train the model and optimize its parameters, such as the weights of the neural network. The validation set was applied to evaluate model performance during training and prevent overfitting. The test set provided an independent dataset for assessing the actual performance of the model, reflecting its behavior in real world applications and was used only after model training and hyperparameter tuning were completed to evaluate the final generalization ability. The distribution trends of the three datasets were kept consistent to avoid additional bias introduced by data splitting and ensure stable accuracy during model training and fitting. In addition, the imbalance between different categories was reduced as much as possible in this study, such that the ratio between any two categories was controlled within 2:1, which alleviated the bias in model prediction.



(a) (b) (c)

Figure 2. Distribution of the training corpus: (a) Distribution of the training set. (b) Distribution of the validation set. (c) Distribution of the test set.

4.2. Experimental Environment and Parameter Fine-tuning

All experiments were conducted on the Google Colab platform with a GPU T4

environment and Linux operating system. The deep learning models were built using the PyTorch 2.0 framework, with Python 3.11 as the programming language. Three models, namely CNN, RNN, and BERT+RNN, were compared. Extensive parameter fine-tuning was performed to achieve the optimal performance for each model, and the corresponding parameter settings are listed in Table 3.

Table 3. Fine-tuning parameters values of the models in the experiment.

Models	CNN	RNN	Bert+RNN
Batch Size	128	128	64
Optimizer	Adam	Adam	Adamax
Epochs	20	20	50
Dropout rate	0.5	0.5	0.8
Learning rate	1e-3	1e-3	1e-3
Convolution kernel size	[2, 3, 4]	——	——
Number of convolution kernels	256	——	——
Pooling window size	3*3	——	——
LSTM hidden layer size	——	128	128
Number of LSTM layers	——	2	1

4.3. Experimental Results

4.3.1. Detection Results of the CNN Model

The accuracy obtained by training the CNN model is shown in Table 4 as a confusion matrix. In this study, only the confusion between class1 & 2 and class4 & 5 was corrected. Based on the confusion matrix statistics, among a total of 3038 data samples, 262 were incorrectly predicted, giving a model accuracy of 91.38%.

Table 4. The confusion matrix and accuracy by the CNN model detected.

Confusion Matrix	Pred_class1	Pred_class2	Pred_class4	Pred_class5
True_class1	752	8	71	13
True_class2	319	85	74	26
True_class4	45	9	193	504
True_class5	22	2	75	840

4.3.2. Detection Results of the RNN Model

The accuracy achieved by the RNN model training is presented as a confusion matrix in Table 5. According to the statistics derived from the confusion matrix, a total of 3038 data samples were evaluated, among which 214 were incorrectly predicted, resulting in a model accuracy of 93.0%.

Table 5. The confusion matrix and accuracy by the RNN model detected.

Confusion Matrix	Pred_class1	Pred_class2	Pred_class4	Pred_class5
True_class1	716	73	51	4
True_class2	256	163	68	17

True_class4	17	27	230	477
True_class5	8	22	82	827

4.3.3. Detection Results of the BERT+RNN Model

The accuracy of the BERT+RNN model is shown in Table 6. According to the statistics, among a total of 3038 data samples, 191 were incorrectly predicted, giving a model accuracy of 93.7%.

Table 6. The confusion matrix and accuracy by the BERT+RNN model detected.

Confusion Matrix	Pred_class1	Pred_class2	Pred_class4	Pred_class5
True_class1	766	46	30	2
True_class2	223	239	32	10
True_class4	17	47	258	429
True_class5	16	37	76	810

By comparing Tables 4–6, it can be seen that the BERT+RNN model achieves the highest accuracy. This model will be used to automatically annotate and correct the collected MOOC reviews.

5. Application Analysis

5.1. Automatic Identification of MOOC Course Reviews Based on the BERT+RNN Model

The collected MOOC corpus was automatically annotated using the BERT+RNN model with the highest accuracy, and compared with the original dataset, as shown in Figure 3. A total of 27794 data samples with inconsistencies between the original annotations and the automatic annotations were corrected, including 1439 samples of class1, 1386 samples of class2, 4940 samples of class4, and 20029 samples of class5, accounting for 8.1% of the total. A sampling inspection was conducted on the corrected dataset, and several typical examples are listed in Table 7. It can be seen that the artificial intelligence model has corrected the obviously wrong original annotations.

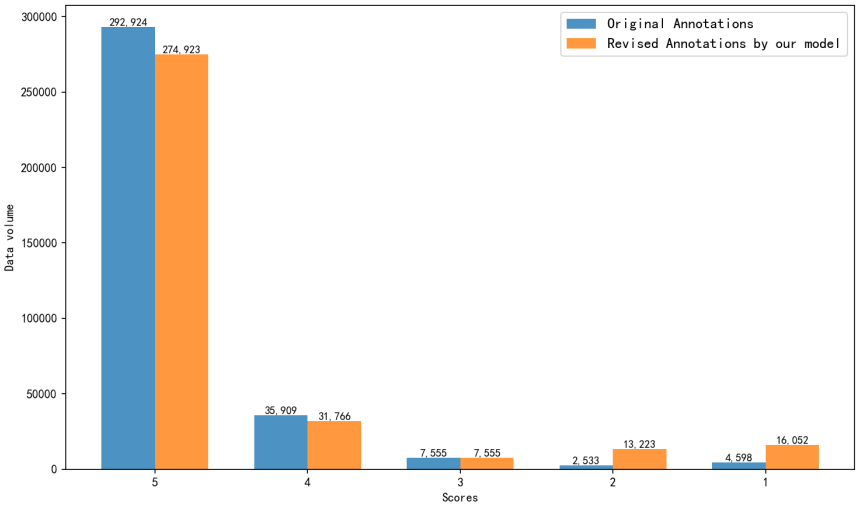


Figure 3. Comparison of distribution between original scores and model automatic scores.

Table 7. Reviews examples are listed for comparision between the original and revised scores by the Model.

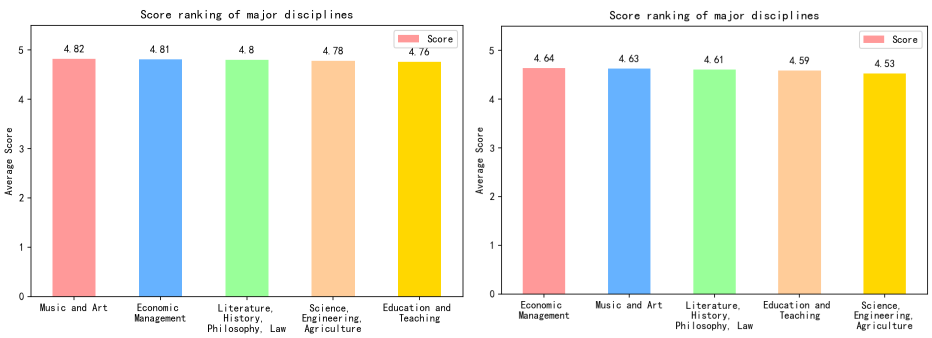
Number	Example	Original Score	Revised Score
1	I have deeply understood the joy of learning and experienced the sense of achievement from hard work. Thank you to the MOOC platform.	1	5
2	The teacher explained very well; it's just that I'm not good enough…	1	5
3	Studying can be addictive, and I feel uncomfortable when I don't study.	1	4
4	Gained some benefits.	1	4
5	The course is excellent!! I hope to achieve good results after completing it.	2	5
6	Teacher Song taught excellently, especially the classic cases, which enable us to draw inferences about other cases from one instance.	2	5
7	Learned a lot, suitable for beginners with zero foundation.	2	4
8	After watching the teacher's videos, I think they are excellent, and I realize I still have a lot to learn.	2	5
9	If we have a relatively objective understanding of ourselves, we can properly and effectively control negative emotions when our mood fluctuates.	2	4
10	The explanation is okay, but I couldn't understand it starting from the chords.	4	2
11	It's an online course, and it's really difficult without experimental equipment.	4	1

12	Insufficient quiz reminders.	4	1
13	It is rather difficult for me as a liberal arts student.	4	1
14	The test time is too short.	4	2
15	Sa da sheng di suo da ao da suo	5	1
16	It would be better if the recording was clearer. Sometimes I can't see the teacher's operations clearly.	5	2
17	Why aren't the grades given yet? It's already the next academic year.	5	2
18	There is no answer.	5	2
19	Can I take the National Computer Rank Examination Level 2 after learning this course?	5	2
20	The course is great, but the volume of some videos fluctuates.	5	2

5.2. Statistical Analysis

5.2.1. Subject Satisfaction Ranking

Based on the MOOC review scores corrected by the model, a statistical analysis of subject satisfaction ranking was conducted. According to the subject category of each course, the average score of each subject was calculated separately. The top five subjects were extracted based on the average scores, and the results are shown in Figure 4. Figure 4 (a) shows the subject ranking under the original scores, and Figure 4 (b) presents the ranking results after model correction. It can be seen that the subject ranking has changed after adjustment by the model.



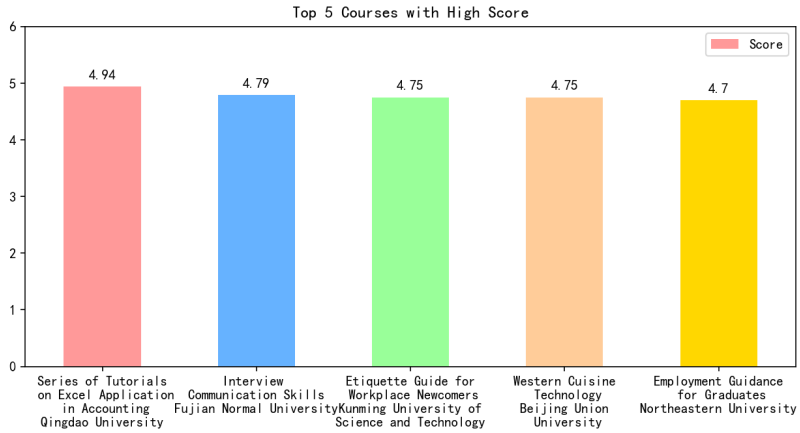
(a) (b)

Figure 4. Subject satisfaction ranking: (a) based on original scores; (b) based on model-corrected scores. The rankings of "Economics and Management" and "Music and Art" were exchanged; the rankings of "Science, Engineering and Agriculture" and "Education and Teaching" were exchanged.

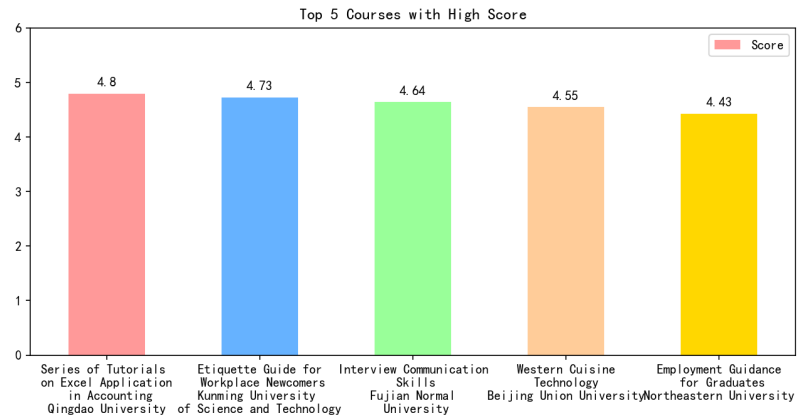
5.2.2. Top 5 Courses in Each Subject by Satisfaction

We further counted the top 5 courses in each subject according to satisfaction. Taking the subject "Economics and Management" in Figure 5 as an example, the course ranking after model correction changed from the original ranking. Interview

Communication Skills from Fujian Normal University and Etiquette Guide for Workplace Newcomers from Kunming University of Science and Technology ranked second and third in the original ranking, respectively. After model correction, their order was exchanged. The course rankings of some other subjects after model correction are shown in Figures 6–9.



(a)



(b)

Figure 5. Course satisfaction ranking of "Economics and Management":(a) based on original scores; (b) based on model-corrected scores.

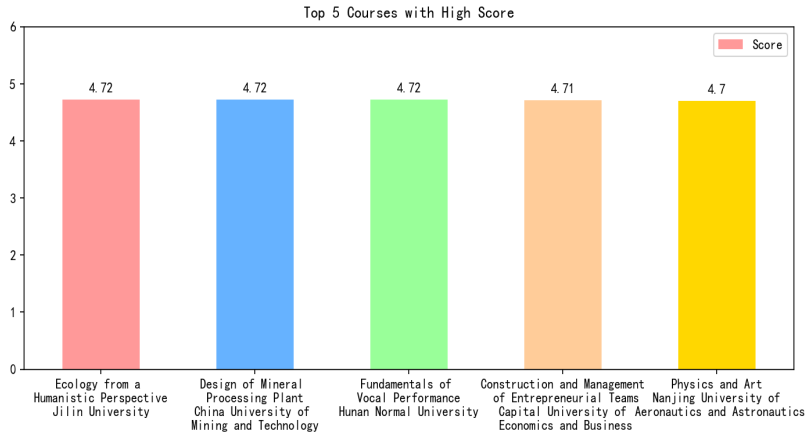


Figure 6. Course satisfaction ranking of "National Excellent Courses", based on model-corrected scores.

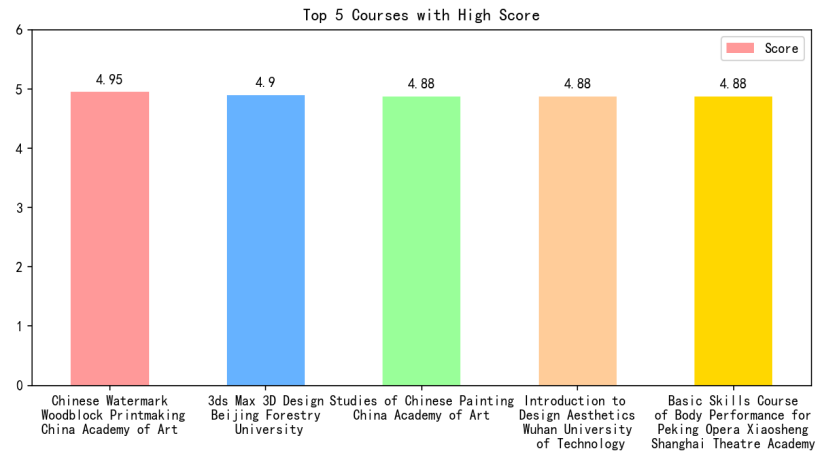


Figure 7. Course satisfaction ranking of "Music and Art", based on model-corrected scores.

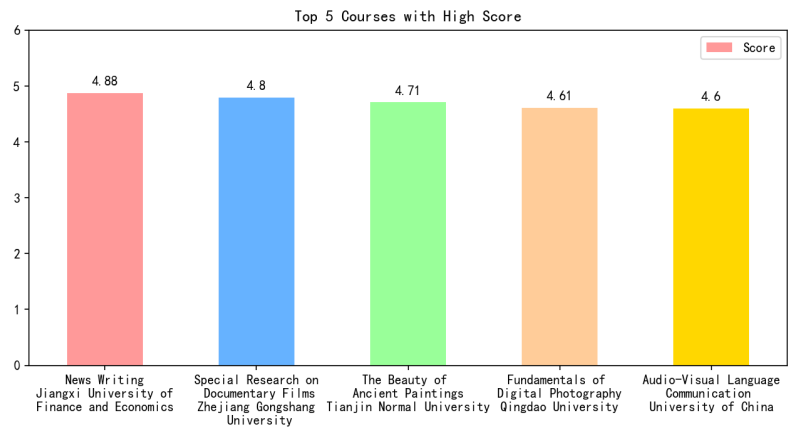


Figure 8. Course satisfaction ranking of "Literature, History, Philosophy and Law", based on model-corrected scores.

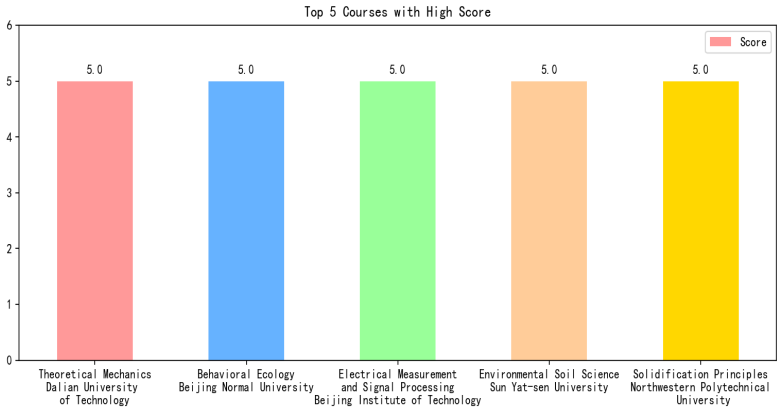


Figure 9. Course satisfaction ranking of "Science, Engineering, and Agriculture", based on model-corrected scores.

6. Discussion

Here we analyzes the detection results and performance of the model. As can be seen from Figure 3, most course ratings are 4 or 5, indicating that students have a relatively high overall satisfaction with MOOC courses. The model corrected the largest number of data for class4 and class5, because these two categories have the largest data volume. However, the correction proportion of class1 and class2 is relatively higher.Some comments with moderate content but low student ratings (Example 4 in Table 5) will be corrected by the model. There are also many comments rated 1 or 2 with relatively positive content (Examples 1, 2, and 3 in Table 5), which may be due to students forgetting to rate, so the system defaults to 1 star. Some comments originally rated 5 stars consist of meaningless sentences (e.g., Example 15), which may be written randomly by students to meet the word count requirement, and the model will automatically correct them to 1 point.Some 5-star comments contain both praise and shortcomings (e.g., Examples 17, 18, 20, and 21). Since the model detects negative evaluations and adopts a stricter scoring criterion, such comments will be labeled as 1–2 stars. Therefore, overall, the model scores comments more strictly, and the overall rating of the corrected corpus will decrease.

7. Conclusions

Course evaluations on the Chinese university MOOC platform serve as an important basis for reflecting course quality and providing course selection references for subsequent learners. However, on this platform, there exists a certain degree of inconsistency between the sentiment orientation expressed in textual comments and the quantitative star ratings.To address this issue, this study collected a total of 343519 course review records and constructed automatic scoring models for reviews based on RNN, CNN, and BERT+RNN respectively. Experimental results show that the BERT+RNN model achieves the best performance. On this basis, the

model is used to automatically score and correct MOOC reviews, and the satisfaction rankings of various disciplines and courses are further counted.The work provides an effective technical tool for purifying the MOOC evaluation environment and improving data quality. The proposed detection model can be integrated into the background management system of the platform as an important part of automatic review and filtering, ensuring the fairness of the evaluation system and promoting the sustainable and high-quality development of the MOOC platform.

It should be noted that this study still has certain limitations. First, the size of the training dataset is insufficient, and there is a problem of class imbalance. Second, the model does not yet cover the prediction of class3. Future research will continue to expand the training dataset, standardize the data annotation process, enhance the fitting performance of pre-trained large language models, and supplement the prediction ability for class3.

Acknowledgements

Guoxia DONG, Tingmei Wang, and Yunfei Xu thank the Youth Special Project of 2023 Beijing Education Science "14th Five-Year Plan" (Grant No. CGCA23129), entitled "Research on the Path of Artificial Intelligence Empowering Classroom Teaching Behavior Analysis," for its financial support.

References

[1] Papadakis S. MOOCs 2012-2022: an overview[J]. Advances in Mobile Learning Educational Research, 2023, 3(1): 682-693.

[2] Moore R L, Blackmon S J. From the learner's perspective: A systematic review of MOOC learner experiences (2008–2021)[J]. Computers & Education, 2022, 190: 104596.

[3] Salama R, Hinton T. Online higher education: Current landscape and future trends[J]. Journal of Further and Higher Education, 2023, 47(7): 913-924.

[4] Huang H, Qi D. Is MOOC really effective? Exploring the outcomes of MOOC adoption and its influencing factors in a higher educational institution in China[J]. Plos one, 2025, 20(2): e0317701.

[5] Cheng J, Yuen A H K, Chiu D K W. Systematic review of MOOC research in mainland China[J]. Library Hi Tech, 2023, 41(5): 1476-1497.

[6] Alturkistani A, Lam C, Foley K, et al. Massive open online course evaluation methods: systematic review[J]. Journal of medical Internet research, 2020, 22(4): e13851.

[7] Mishra M, Dash M K, Sudarsan D, et al. Assessment of trend and current pattern of open educational resources: A bibliometric analysis[J]. The Journal of Academic Librarianship, 2022, 48(3): 102520.

[8] Su P Y, Guo J H, Shao Q G. Construction of the quality evaluation index system of MOOC platforms based on the user perspective[J]. Sustainability, 2021, 13(20): 11163.

[9] Mehta P, Pandya S. A review on sentiment analysis methodologies, practices and applications[J]. International Journal of Scientific and Technology Research, 2020, 9(2): 601-609.

[10] Liu, J., Tan, J. Q., Zhang, J. H., et al. Research on Text Analysis of Emotional Dictionaries for Mathematics Courses Based on MOOC Platforms[J]. Science & Education Guide, 2021(19): 78-80. DOI: 10.16400/j.cnki.kjdk.2021.19.026.

[11] Graves A. Long short-term memory[J]. Supervised sequence labelling with recurrent

neural networks, 2012: 37-45.

[12] Cho K, Van Merriënboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[J]. arXiv preprint arXiv:1406.1078, 2014.

[13] Sun, S. Q. Research on Learners' Emotional Tendency Analysis Based on Course Evaluation in the MOOC Platform[D]. Liaoning Normal University, 2023. DOI: 10.27212/d.cnki.glnsu.2023.000229.

[14] Chen X, Zou D, Cheng G, et al. Deep neural networks for the automatic understanding of the semantic content of online course reviews[J]. Education and Information Technologies, 2024, 29(4): 3953-3991.

[15] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.

[16] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers). 2019: 4171-4186.

[17] Cavalcanti A P, Mello R F, Gašević D, et al. Towards explainable prediction feedback messages using BERT[J]. International Journal of Artificial Intelligence in Education, 2024, 34(3): 1046-1071.

[18] Cavalcanti A P, Mello R F, Miranda P, et al. Utilização de recursos linguísticos para classificação automática de mensagens de feedback[C]//Simpósio Brasileiro de Informática na Educação (SBIE). SBC, 2021: 861-872.

[19] Wei, X. C., Yu, L. Construction and Application of Emotional Recognition Corpus for Chinese MOOC Reviews[J]. Journal of Chongqing University of Technology (Natural Science), 2023, 37(04): 174–181.

[20] Liu, S. N. Y., Peng, X., Liu, Z., et al. Research on Learner Topic Mining for MOOC Course Reviews[J]. Journal of Educational Technology, 2017, 38(10): 30–36. DOI: 10.13811/j.cnki.eer.2017.10.005.

[21] Zhang, X. X., Duan, Y. H. MOOC Quality Evaluation Based on Learners' Online Review Texts: A Case Study of "Chinese University MOOC" Reviews[J]. Modern Educational Technology, 2020, 30(09): 56–63. DOI: CNKI:SUN:XJJS.0.2020-09-009.

[22] Chen X, Zou D, Xie H, et al. Automatic Classification of Online Learner Reviews Via Fine-Tuned BERTs[J]. International Review of Research in Open and Distributed Learning, 2025, 26(1): 57-79.