International Journal of Computing Sciences Research (ISSN print: 2546-0552; ISSN online: 2546-115X)
Vol. 9, pp. 3782-3803
doi: 10.25147/ijcsr.2017.001.1.249
https://stepacademic.net

Short Paper

Sentiment Analysis of Chinese Online Course Reviews Based on Deep Learning

Junchao Dong
Graduate School, University of the East, Philippines
dong.junchao@ue.edu.ph
(corresponding author)

Joan P. Lazaro
Graduate School, University of the East, Philippines
joan.lazaro@ue.edu.ph

Date received: March 13, 2025
Date received in revised form: May 22, 2025
Date accepted: May 22, 2025

Recommended citation:

> Dong, J., & Lazaro, J. P. (2025). Sentiment analysis of Chinese online course reviews based on deep learning. *International Journal of Computing Sciences Research*, 9, 3782-3803. https://doi.org/10.25147/ijcsr.2017.001.1.249

Abstract

Purpose – Sentiment analysis of Chinese online course reviews is one of the key technologies to improve the intelligence level of online learning systems. This study proposes a sentiment analysis model for Chinese online course reviews based on the long short-term memory network (LSTM). It integrates it into the "Mushroom Community" platform independently developed by the researchers to achieve automatic classification of the sentiment tendency of reviews.

Method – Taking 29,985-course reviews from the China MOOC platform as the dataset, four models, namely linear regression, support vector machine, random forest, and LSTM, were constructed and compared. The accuracy, precision, recall, and F1-score were used to evaluate the model performance.

Results – Experimental results show that the LSTM model performs best in this task: accuracy 0.94, precision 0.83, recall 0.86, F1 Score 0.84; and has been successfully deployed in the "Mushroom Community", verifying its feasibility and practicality.

Conclusion – This study confirms the effectiveness of the LSTM model in sentiment analysis of Chinese online course reviews and can be applied to actual online learning systems.

Recommendations – Given LSTM's reliance on large-scale, high-quality annotated data and its "black box" characteristics, subsequent work will explore the introduction of pre-trained language models such as BERT and lightweight integration solutions in small sample scenarios.

Research Implications – The research results provide a reference for Chinese text mining and the development of intelligent online learning systems.

Practical Implications – This study can provide data support and decision-making reference for educational platforms, teaching institutions, and teachers in optimizing course quality, adjusting teaching strategies, and sustainable development of online learning systems.

Keywords – online courses, course reviews, sentiment analysis, deep learning, LSTM

INTRODUCTION

In recent years, the scale of online education users in China has continued to grow and is expected to exceed 380 million by the end of 2025 (Yu et al., 2024). Against the backdrop of the accelerated digital transformation of education, massive amounts of Chinese course review data have become an important basis for evaluating teaching quality and optimizing course design. The number of comments in online learning systems can reach tens of thousands per day, and these data contain learners' emotional feedback on knowledge point explanations and teaching methods (Shaik et al., 2023). Traditional manual analysis methods are inefficient and highly subjective, and it is not easy to meet the needs of real-time feedback. Course review sentiment analysis plays an important role in a variety of educational scenarios, including student satisfaction analysis, personalized recommendation of learning resources, and early warning and timely intervention of dropouts.

Online course review of sentiment literature is one of the core tasks in the field of natural language processing. It aims to identify and classify sentiment tendencies (positive/negative) in course review text data through machine learning algorithms. Sentiment analysis technology can be roughly divided into three stages: rule-driven, machine learning, and deep learning (Onan, 2021, p. 15). In the rule-driven stage, people usually need to build sentiment dictionaries and syntactic rule matching. The classification of sentiment tendencies is achieved through the statistics of polarity words. The methods at this stage mainly have bottlenecks such as poor domain adaptability and the inability to accurately capture implicit emotions. Especially in the Chinese context, when there are more colloquial expressions and ironic expressions, the effect is poor. In order to make up for the shortcomings of the rule-driven stage, researchers introduced a variety of traditional machine learning models such as support vector machines and random forests, combined with feature engineering such as TF-IDF and N-gram to improve the prediction accuracy. However, the feature engineering construction in the machine learning stage is a very large and complex stage, which requires a lot of manpower and material resources. For this reason, deep learning models represented by LSTM have been widely used in a variety of tasks such as text classification and commodity sentiment analysis, and have achieved good results. The LSTM model can achieve a deep understanding of the semantics in the comment text through context encoding and end-to-end training, significantly improving the accuracy of the model and having better generalization capabilities. In addition, the deep learning model can automatically learn abstract features from the original data and directly parse the semantic relationship of the text sequence without human intervention. This avoids the reliance of traditional machine learning methods on bag-of-words models or TF-IDF, greatly reducing the cost of manpower investment, saving time, and improving efficiency.

This study combines the Word2Vec word vector model and the LSTM analysis algorithm to build a deep learning model for Chinese online course review analysis, which can accurately predict the emotional information contained in student reviews. In addition, the researchers integrated it into the online learning system "Mushroom Community" independently developed by the researchers to realize the real-time emotional management function of the online learning system and improve the intelligence of the online learning system.

LITERATURE REVIEW

Jatain et al. (2023) used 105,000 learner review data from Coursera and MSIT platforms to build a two-stage aspect sentiment analysis model. The first stage determines the aspect categories of the reviews, and the second stage determines the emotional tendencies corresponding to these aspects. The experimental results show that course content, teacher quality, technical support, and evaluation methods are the main aspects that affect course popularity. The BERT-based aspect sentiment analysis model has the best effect, with an accuracy of 94.27%. The researchers found that deep learning models such as BERT, EvoMSA, and ELMo have significant advantages in sentiment analysis. The aspect sentiment analysis results can provide guidance and suggestions for course optimization to instructors and course leaders.

Onan (2020) built a variety of classification models including traditional machine learning, ensemble learning, and deep learning to classify and analyze student feedback in teaching evaluation. Based on a dataset of 154,000 teaching evaluation comments, the experimental results show that the deep learning model outperforms the traditional machine learning model in sentiment analysis tasks. The solution combining GloVe word vectors and the RNN model enhanced by the attention mechanism can better handle the complex emotions expressed in student comments and significantly improve the classification accuracy, reaching 98.29%. This study shows that deep learning has great potential in processing massive student evaluation data, and provides new ideas for educational institutions to improve the intelligence level of teaching evaluation systems.

Mamidted and Maulana (2023) studied the comments of 237 college students in a state university in the Philippines on the performance of teachers in online classes. The Orange text mining software was used to extract the positive and negative emotions in the students' comments and analyze the students' feelings about the teachers' online teaching performance. The results showed that 85.53% of the students had a positive attitude towards the performance of teachers in online classes, and only 3.40% of the students expressed negative emotions. This study provides empirical evidence for teacher performance in online education, and teachers play a key role in promoting student engagement and motivation.

Santiago et al. (2022) conducted a sentiment analysis of the online learning experience of students at a state university in the Philippines. The study collected unstructured feedback from 94 students and analyzed the sentiment polarity (positive, neutral, negative) and subjectivity or objectivity of these feedbacks using natural language processing technology and sentiment analysis methods. The results showed that most students had a positive online learning experience and believed that online courses had significant advantages in applying online learning tools, collaborating with classmates to complete academic tasks and the flexibility of time and place. Although students also encountered challenges in network connection and access to learning resources when using online courses, they had good adaptability and self-regulation and were able to overcome these difficulties.

Cahapin et al. (2023) conducted a study on 459 students from Cavite State University in the Philippines, processing 918 open-ended responses (27,316 words) using sentiment analysis technology to explore students' perceptions of limited face-to-face learning after the epidemic. The results showed that students' learning status showed a mixture of positive, negative, and neutral emotions, and opinions on the implementation of limited face-to-face teaching were mainly positive (over 300), mainly reflected in the expectation of face-to-face interaction and improved learning outcomes, but some students were worried about health risks (over 100 negative emotions) and increased transportation costs. The study pointed out that the COVID-19 epidemic poses challenges to students' financial, social, and mental health. Educators need to receive training in health protection teaching methods and optimize course outlines. Schools should dynamically adjust face-to-face policies to balance safety and learning needs. This study provides emotional data support for the optimization of hybrid learning models in the post-epidemic period.

Celestial-Valderama et al. (2021) conducted a study on student feedback in blended learning courses at Jose Rizal University in the Philippines. They explored how to use unstructured feedback to optimize the implementation of blended learning by conducting text mining and sentiment analysis on the responses to open-ended questions in the

Canvas learning management system survey. The study selected courses with low student satisfaction, such as "History Reading in the Philippines" (HIS C101), and preprocessed the responses to three open-ended questions by data cleaning, word segmentation, and stop word removal. The K-means clustering and TF-IDF keyword extraction techniques were used to generate topic clusters, and the polarity scores were calculated through sentiment analysis, as the polarity score of the topic "Teacher Interaction" was 0.985, and the polarity score of the topic "Notification System" was 0.004. The results show that student feedback focuses on teacher preparation, LMS module accessibility, and notification efficiency, with both positive and negative emotions. The study proposes to develop a feedback mining system to match negative sentiment topics, such as "slow network connection", to corresponding departments, such as the Information Technology Office to develop improvement plans, providing a data-driven practical framework for the continuous optimization of blended learning.

Yee et al. (2023) used a Transformer-based pre-trained deep learning model to classify the emotions of forum posts in a massive open online course (MOOC). The results showed that the Transformer-based pre-trained model outperformed the traditional support vector machine (SVM) model in the emotion classification task. The performance of the Transformer-based pre-trained model was further improved when the topic and context information of the post were combined. The study found that there were significant differences in the emotional expression of learners of different ages, genders, and course completion statuses. Older learners posted more posts and their posts often contained richer emotional content. The study believes that by introducing artificial intelligence models, course-related personnel can better understand students' interactive behaviors in MOOCs and develop more targeted teaching strategies.

Koufakou (2024) used deep learning models to classify the sentiment of student course reviews. The study used multiple models such as deep neural network models and state-of-the-art Transformer models such as BERT, RoBERTa, and XLNet to analyze more than 10,000 online course reviews. A detailed comparison of traditional machine learning methods and modern deep learning models in sentiment analysis tasks was conducted. The experimental results showed that deep learning models have advantages in processing large-scale text data, and RoBERTa performed best in the sentiment polarity classification task with an accuracy of 95.5%. At the same time, the study also explored the impact of different hyperparameters on model performance, providing practical guidance for educational institutions on how to use these NLP models to analyze course feedback.

In order to reduce the reliance on manually labeled data and automatically identify the polarity of emotions expressed in student comments, Kastrati et al. (2020) proposed a weakly supervised framework. The framework uses weak supervision techniques to train deep learning models with a small amount of manually labeled data to identify sentiment categories in unannotated comments. The study used two large-scale real-world educational datasets of 105,000 and 5,989 comments collected from Coursera and traditional classroom environments. Experimental results show that the framework performs well in sentiment classification tasks and is significantly better than sentiment analysis techniques that rely on a large amount of manually labeled data. The framework uses weak supervision signals to effectively identify key aspects in MOOC comments and classify the sentiment tendencies of these aspects, thereby improving the accuracy and efficiency of sentiment analysis. The weakly supervised framework reduces the workload of manual annotation, provides teachers with valuable advice on course design and teaching effectiveness, and helps improve the quality of MOOC teaching.

The above literature review shows that in recent years, deep learning technology has shown significant advantages in sentiment analysis research in the field of education. Many studies have verified the excellent performance of various pre-trained models based on LSTM and Transformer architectures in the task of sentiment classification of course reviews (Murithi et al., 2024). However, existing results mainly focus on English contexts and are mostly based on independent analysis systems, lacking deep integration with online learning platforms. In addition, although various pre-trained models based on Transformer architecture have achieved better experimental results than LSTM models in the task of sentiment analysis of course reviews, such models still face challenges in terms of efficiency, domain adaptability, and interpretability, and are not yet fully applicable to the development and use of Mushroom Community. The Word2Vec word vector and LSTM neural network fusion architecture has been widely used in Chinese product reviews due to its advantages in lightweight adaptation to Chinese features, low resource friendliness, and system integration feasibility, and has been successfully integrated and applied in e-commerce platforms. Under the comprehensive consideration of accuracy, efficiency, and practicality, the development of this online learning system also adopts the Word2Vec word vector and LSTM neural network fusion architecture similar to that of e-commerce platforms.

To this end, this study proposes to build a Chinese course review sentiment analysis model that integrates Word2Vec word vectors and LSTM neural networks and integrates the model into the independently developed "Mushroom Community" online learning system to achieve real-time monitoring and intelligent management of review sentiment. This study not only expands the localized application scenarios of educational sentiment analysis but also promotes the evolution of online education platforms toward intelligence and adaptability through system-level sentiment computing modules.

METHODOLOGY

In order to build a Chinese course review sentiment analysis model suitable for the Mushroom Community, the researchers conducted an experimental demonstration as shown in Figure 1.

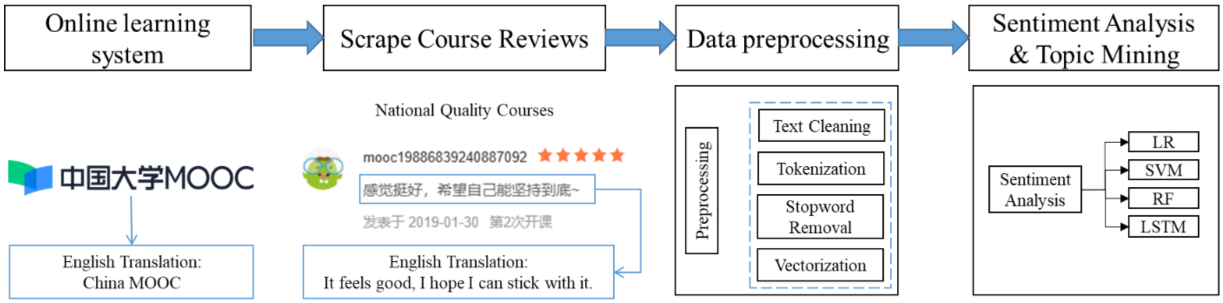

Figure 1. Chinese online course review sentiment analysis process.

The researchers used Python programming language and Selenium library to crawl 107,845 comments from 2,096 national quality courses on Chinese MOOC websites. A combination of automatic and manual methods was used to remove duplicate comments, clean special characters, filter invalid text, standardize language fonts, and filter short text, and a total of 29,985 valid comments were obtained from 2,004 courses. The jieba library was used for text segmentation, and the commonly used Chinese stop word list was applied to remove words with low information content. The sentiment tags of the comments were manually annotated, and the text was vectorized using TF-IDF (for LR, SVM, and RF models) and the Word2Vec word embedding model (for LSTM). Four sentiment analysis models were tested: three machine learning models (LR, SVM, RF) and one deep learning model (LSTM) for binary sentiment classification of course reviews. Accuracy and F1 score were the main evaluation indicators.

Data collection and preprocessing

Data collection

This study selected Chinese university MOOC as the source of experimental data. The Python programming language combined with the Selenium web crawling library was used to collect the course review data of Chinese university MOOCs, as shown in Figure 2. The specific web crawling process is shown in Figure 3.

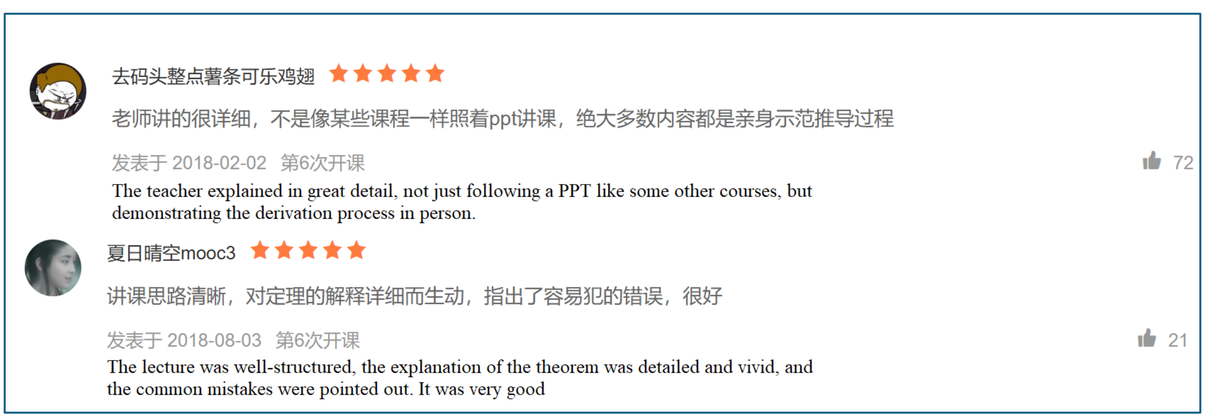

Figure 2. China MOOC online course review page (sample).

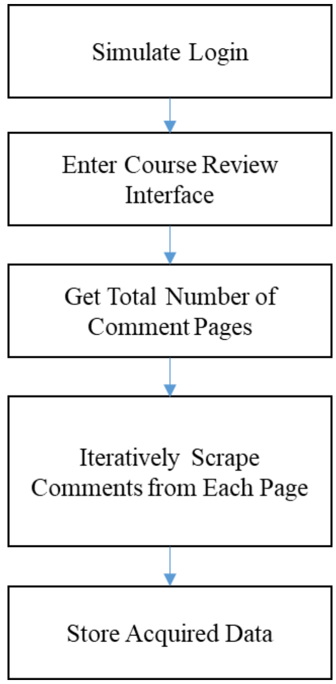

Figure 3. Chinese online course review crawling process.

Step 1: Configure the experimental environment. Use "pip install" from the terminal to download third-party libraries such as Selenium and Pandas. Download the driver "chromedriver" corresponding to the Chrome browser version and place the executable file in the Python project directory.

Step 2: Define collection rules. Use the "find_element" and "find_elements" methods to locate and retrieve page elements that meet the criteria by ID value, class, XPath path, etc. Simulate email login to obtain cookies and write them to the "cookie.txt" file. Use cookies to log in and refresh the page, enter the course comment page in the logged-in state, and obtain the comments on each page. It should be noted that when Selenium opens a page, it operates in the parent frame by default and cannot access nodes in child frames. In this case, use the "switch_to.frame" method to switch to the desired frame.

Step 3: Store the acquired information. If the data volume is very large, you can store it in MySQL or a remote server. If the data volume is small, you can save the data in a local file in the format of CSV, XLSX, JSON, etc. In this study, since the data volume is not large, it is written into an XLSX file.

Data preprocessing

The text data of Chinese online course reviews posted by users is unstructured. In the Internet age, a free, open, rapidly developing, and diversified network culture has emerged. People express their emotions in a variety of ways, and the content of the comments has become complex and diverse. This includes the use of emoticons, Internet buzzwords, text symbols, number combinations, and other special expressions, as well as a large number of invalid symbols and repeated information. A large number of repeated identifiers will affect the frequency of word occurrence, thereby affecting the accuracy of the model. Meaningless text will have a negative impact on data modeling, and noisy data will interfere with experimental results and should not be learned by the model. Therefore, before extracting text features and building a sentiment analysis model, the data needs to be cleaned so that the model can better fit the actual semantic features and enhance its generalization ability. The data preprocessing process is shown in Figure 4.

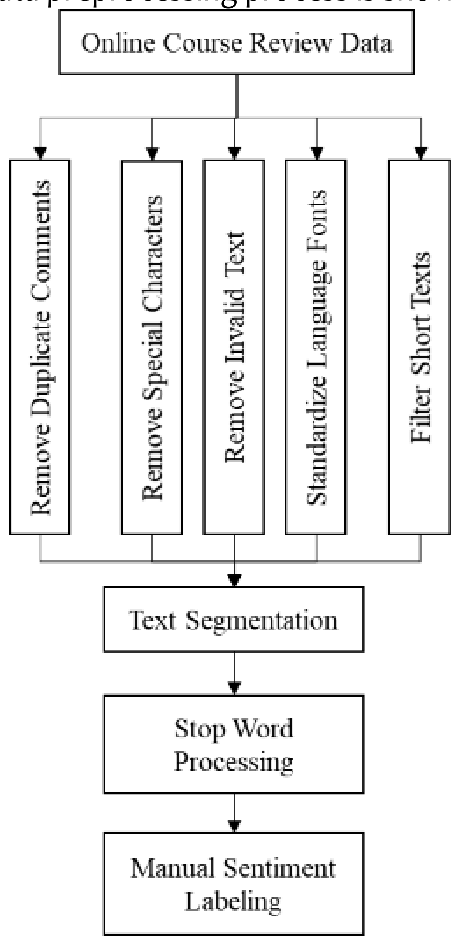

Figure 4. Data preprocessing process.

The researchers used a combination of programmatic automation and manual screening to complete the data cleaning. The specific steps are as follows:

Step 1: Remove duplicate comments. If the same user posts multiple identical comments on the same course, they are considered duplicates, and only the first comment is retained. This task uses the "drop_dupldduplicate()" function in the Pandas library, and 686 duplicate records were removed.

Step 2: Remove special characters. Upon inspection, the researchers found that the comments contained emoticons such as "(_ò _ ó_)", "o(^o^)o", phonetic symbols such as "ㄝ", "ㄡ", and special characters such as "~", "&" that are irrelevant to the content information. Regular expressions are used to remove most unnecessary symbols, and string methods are used to remove specific characters.

Step 3: Remove invalid text. Characters that do not convey any emotional meaning or provide specific evaluations of the topic are considered invalid texts because meaningless statements significantly reduce data quality. The 're' module in Python is used to replace pure numeric comments such as "131300", comments such as "…" that are composed entirely of punctuation marks, and single-character comments such as "hhhh" with empty strings, while uniformly deleting low-quality comments such as "啊啊啊啊啊" and "哈哈哈".

Step 4: Standardize language fonts. Pure English and Japanese comments are deleted, and traditional Chinese characters are converted to simplified Chinese to standardize the format.

Step 5: Filter short texts. Short texts are usually unclear, chaotically structured, and ambiguous, providing little information of analytical value. Therefore, in order to ensure the effectiveness of subsequent data analysis, records with less than or equal to 6 Chinese characters are filtered out.

Step 6: Text segmentation. After cleaning, the annotation data needs to be segmented. Unlike English, where words are separated by spaces, Chinese has a fuzzy boundary between "words" and "phrases" and no formal separators. Words are the smallest, independent, and meaningful language units. The task of text segmentation is to insert spaces or other boundary markers in a continuous sequence of characters to form a sequence of words according to certain rules. Currently, a popular Chinese text segmentation tool is Jieba. Therefore, the researchers used the Jieba package in Python to segment the annotation text.

Step 7: Remove stop words. Stop words are high-frequency words with low information value such as quantifiers, pronouns, conjunctions, etc., which take up storage space and affect text feature extraction. Identifying the emotions contained in the comment text is different from accurately describing the language rules, and it does not require the use of all components in the sentence. Removing stop words does not directly affect the classification results while retaining meaningful words will reduce the amount of calculation and help improve the accuracy of subsequent sentiment analysis. The construction of the stop word list depends on the experimental tasks and goals and needs to be continuously updated during use. Commonly used stopword lists include the Harbin Institute of Technology stopword list, the Baidu stopword list, and the Chinese stopword list. By integrating the contents of these public dictionaries, we manually removed words such as "进步", "容易" and "清楚", which are often associated with user emotions in course reviews. In addition, we added colloquial small words that often appear in the comment section, such as "滴", "鸭" and "哒", to form a custom stop word list for online course reviews. Using this custom stop word set, irrelevant words, punctuation marks, and common terms are removed from the comments. The comparison shows that the preprocessed text data is concise and well-structured, which highlights the key information, facilitates text feature extraction, and improves classification accuracy.

Step 8: Sentiment labeling. In order to obtain the sentiment labels, we selected 15 undergraduate students from Guilin University of Electronic Technology to manually label the sentiment polarity of the original comment text. The labeling rules are as follows: Each group consists of three students who manually label the sentiment polarity of the same comment. If all three agree, the final sentiment label will be assigned accordingly. If there is a disagreement among the three, the majority opinion is used to determine the final sentiment label. The sentiment labels are 0 (negative) and 1 (positive). To ensure the reliability of annotation quality, three annotators were arranged to annotate randomly selected subsets of 100 Chinese course review text samples. The Fleiss' Kappa score was then used to evaluate the consistency between raters. This indicator is often used to measure the consistency level between multiple raters. The evaluation results show that the Fleiss' Kappa score is 0.72. According to the scale standard, this value belongs to the "substantial consistency" range, indicating that the consistency between raters is high.

After the above preprocessing steps, a total of 29,985 valid comments were obtained, forming the online course review dataset required for subsequent model training. The specific text data is shown in Table 1.

Sentiment Analysis Model

Sentiment analysis of Chinese online course reviews belongs to the sentiment analysis task, which is essentially Chinese text analysis in data mining. To find a model suitable for the task of sentiment analysis of Chinese online course reviews, the researchers selected four models that are widely used in text classification and sentiment analysis tasks: logistic regression (LR) (Mantika et al., 2024), random forest (RF) (Khan et al., 2024), support vector machine (SVM) (Arif et al., 2024), and long short-term memory (LSTM) (Alfarhood, 2025).

Experimental environment and model hyperparameters

The experiments were conducted using the Google Colab platform, which provides a high-performance cloud-based environment equipped with NVIDIA GPUs for efficient model training. The following Python libraries were used:

TensorFlow for deep learning model implementation.

Scikit-learn for machine learning algorithms and model evaluation.

Gensim for Word2Vec embedding training.

Pandas and NumPy for data manipulation.

NLTK and Keras for text processing and neural network construction.

Table 1. Chinese online course review text data preprocessing process (example)

Original comment	final comment			
	Data Cleaning	Text Segmentation	Stop Word Removal	Sentiment Label
这门课程讲得非常清楚，收获很多！👍 English translation: This course was very clearly taught and I gained a lot!	这门课程讲得非常清楚收获很多	这门 / 课程 / 讲得 / 非常 / 清楚 / 收获 / 很多	课程 / 讲得 / 清楚 / 收获	1
哈哈哈，这课程真不错！😁 English translation: Hahaha, this course is really good!	这课程真不错	这 / 课程 / 真 / 不错	课程 / 不错	1
课程内容太难了，听不懂，讲解不够清楚😓。 English translation: The course content is too difficult, I can't understand it, and the explanation is not clear enough.	课程内容太难了听不懂讲解不够清楚	课程 / 内容 / 太 / 难 / 了 / 听不懂 / 讲解 / 不够 / 清楚	课程 / 内容 / 难 / 听不懂 / 讲解 / 清楚	0
学习过程中遇到很多问题，效果不如预期，非常失望！。。。。 English translation: I encountered many problems during the learning process and the results were not as expected. I was very disappointed!	学习过程中遇到很多问题效果不如预期非常失望	学习 / 过程 / 中 / 遇到 / 很多 / 问题 / 效果 / 不如 / 预期 / 非常 / 失望	学习 / 过程 / 遇到 / 问题 / 效果 / 不如 / 预期 / 失望	0

In this study, the researchers divided the dataset in a ratio of 8:2, with 80% used for training the model and 20% for testing the model performance. For this purpose, the ‘train_test_split’ function was used, and the ‘random_state’ was set to 42 to ensure the repeatability of the split. In addition, by analyzing the number of positive and negative samples in the training set, the researchers found that there was an obvious sample imbalance in the dataset (positive samples: negative samples were approximately 9:1). To solve the sample imbalance problem in the training set, the researchers used random oversampling to increase the number of minority class samples and balance the sample class distribution. This method generated a more balanced training dataset (positive samples: negative samples were approximately 1:1), which helped improve the performance and classification accuracy of the model on the imbalanced dataset. In this study, in order to achieve sentiment analysis of course reviews, the researchers used different word embedding techniques to construct the input features of the four models. The three machine learning models LR, RF, and SVM used TF-IDF for text vectorization. TF-IDF measures the importance of each word in the document and the entire corpus, generating sparse and high-dimensional feature representations for training traditional machine learning models. For the LSTM model, the researchers used Word2Vec embedding to convert the text into a low-dimensional, dense vector representation while retaining the semantic relationship and contextual information between words. The use of Word2Vec word embedding enables LSTM to more effectively capture the semantic and sequential dependencies in the text, thereby improving its performance in sentiment classification tasks. The specific hyperparameters of the four models are shown in Table 2.

Table 2. Model hyperparameter settings

Model	**Hyperparameter**	**Setting**
LR	Solver	'lbfgs'
	Maximum Iterations	Default
RF	Number of Estimators	100
	Random State	42
	Criterion	'gini'
SVM	Kernel	'linear'
	Regularization Parameter (C)	Default
LSTM	Embedding Dimension	100
	Number of LSTM Units	100
	Dropout Rate	0.2
	Recurrent Dropout Rate	0.2
	Batch Size	64
	Number of Epochs	50
	Optimizer	'adam'
	Loss Function	'binary_crossentropy'

Evaluation metrics

This study divides the sentiment of course reviews into positive (1) and negative (0), which is a binary classification problem in machine learning. Therefore, the researchers used the commonly used evaluation indicators of binary classification models, accuracy, precision, recall, and F1 score, to evaluate the prediction ability of the model (Rainio et al., 2024).

System Integration and Visualization

Save the trained LSTM model as the "lstm_sentiment_model.h5" file. Then use Flask to deploy it as a REST API and integrate it with Spring Boot. In Spring Boot, the sentiment analysis function is implemented by calling the Flask service using "RestTemplate" and creating a controller to handle sentiment analysis requests from the front end.

In order to visualize the sentiment analysis results in the comment management module of the Mushroom Community backend management system, the researchers used the table and chart visualization tool ECharts.js in the backend system to display the results by interacting with the API of the Spring Boot service.

RESULTS

Comparative experiment

In order to verify the effectiveness of the four selected models in the Chinese online course review sentiment analysis task, the researchers conducted comparative experiments under the same conditions. This helps to understand the relative performance of different models and methods. The experimental results are shown in Table 3.

Table 3. Performance of four models in the Chinese online course review sentiment analysis task

Model	Accuracy	Precision	Recall	F1-score
LR	0.94	0.84	0.82	0.83
SVM	0.94	0.83	0.82	0.83
RF	0.94	0.83	0.82	0.82
LSTM	0.94	0.83	0.86	0.84

From the experimental results in Table 3, we can see that the accuracy of the four models all reached 0.94, indicating that they performed consistently in the overall classification task. Despite the high accuracy, there were slight differences between the models when analyzing specific indicators. LR performed stably in the experiment, with an accuracy of 0.84, a recall of 0.82, and an f1 score of 0.83. This shows that logistic regression can balance precision and recall well, but may have limitations when dealing with contextual dependencies of complex texts. The accuracy and recall of SVM are similar to those of LR, at 0.83 and 0.82, respectively, and the f1 score is 0.83. This shows that SVM performs well in high-dimensional feature spaces and is suitable for text classification tasks. RF performs slightly worse than the other models, with both precision and recall of 0.82

and a slight drop in the f1 score to 0.82. This shows that although random forests can capture the importance of features when processing text data, they may be weak in capturing sequential relationships in sentiment analysis. LSTM performs well, especially in terms of recall, which reaches 0.86 and an F1-score of 0.84, demonstrating its advantage in capturing sequential dependencies and contextual information in text. The higher recall rate indicates that LSTM does a better job of identifying actual positive samples and reducing missed detections. The experimental results are consistent with the conclusions obtained by Onan (2020) in English course review texts, which indirectly reflects the strong applicability of the LSTM model in sentiment analysis of course reviews.

Hyperparameter Selection

In order to further evaluate the impact of the two key hyperparameters, Epochs, and learning rate, on the performance of the LSTM model, the researchers plotted the loss curve, accuracy curve, and learning rate curve, as shown in Figure 5.

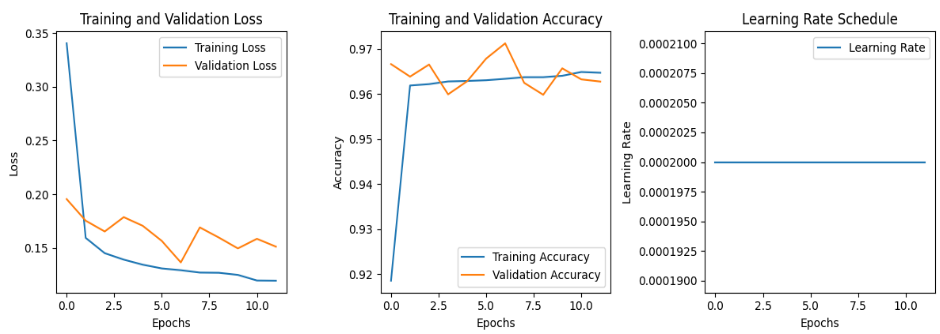

Figure 5. Loss curve, accuracy curve, and learning rate curve.

As shown in Figure 5, both the training loss and validation loss are decreasing, indicating that the model is constantly learning. However, the validation loss fluctuates at certain points, indicating that the model occasionally overfits. Overall, both the training loss curve and the validation loss curve show a consistent downward trend, indicating that there is no obvious overfitting or underfitting during the training process. At the same time, the training and validation accuracies are high and close, which also shows that the model performs well without obvious overfitting or underfitting. There is no obvious change in the learning rate, indicating that the threshold for learning rate adjustment may not have been reached during the training process.

Performance differences across course categories

In order to explore the possible performance differences of the LSTM model in different courses, the researchers divided the collected Chinese course review data into two categories: Humanities and Social Studies (HSS) Courses and Science, Technology, Engineering, and Mathematics (STEM) Courses. Then, the LSTM model was used to conduct experiments on the datasets of the two categories of courses. The experimental results are shown in Table 4.

Table 4. Performance of LSTM model in sentiment analysis tasks for two types of Chinese online course reviews

Model	Accuracy	Precision	Recall	F1-score
HSS	0.91	0.82	0.84	0.83
STEM	0.95	0.88	0.89	0.88

From the experimental results in Table 4, we can see that the performance of the LSTM model in humanities course reviews is better than that of STEM courses, with accuracy, precision, recall, and F1-scores of 4.40%, 7.32%, 5.95%, and 6.02% respectively. This shows that the performance of the LSTM model will be slightly different in different course categories.

Typical misclassification examples and analysis

In order to gain a deeper understanding of the limitations of the model, some representative examples that were misclassified by the model were selected, as shown in Table 5.

Table 5. Performance of LSTM model in sentiment analysis tasks for two types of Chinese online course reviews

Comment text	Actual Sentiment	Model Prediction	Causes of prediction errors
我觉得我听他讲 1000 节课，我会取得非常好的成绩，哈哈哈！ English translation: I think if I listen to his 1,000 classes, I will achieve very good results, hahaha!	Negative	Positive	Taking sarcasm as compliments
我不能说，老师没有尽力。 English translation: I can't say the teacher didn't try his best.	Positive	Negative	Contains complex negation
虽然实验模拟出错了，但是这节课整体还可以。 English translation: Although there were some errors in the experimental simulation, the class was generally okay.	Negative	Positive	Contains professional terms

The researchers manually checked and analyzed the characteristics of the comments that the model misclassified. They found that the model had a higher probability of sending errors when faced with three complex scenarios: the first example comment in Table 5: a comment with irony; the second example comment in Table 5: a comment with multiple negations; and the third example comment in Table 5: a comment with professional terms.

Mushroom Community Sentiment Analysis Web Page

The sentiment analysis page of the Chinese online course comments in the Mushroom Community online learning system is shown in Figure 6. This page intuitively presents the sentiment polarity distribution of user comments in the system through data visualization technology. Administrators can quickly understand the sentiment trend of the entire course through this page, which is a reference for further course optimization.

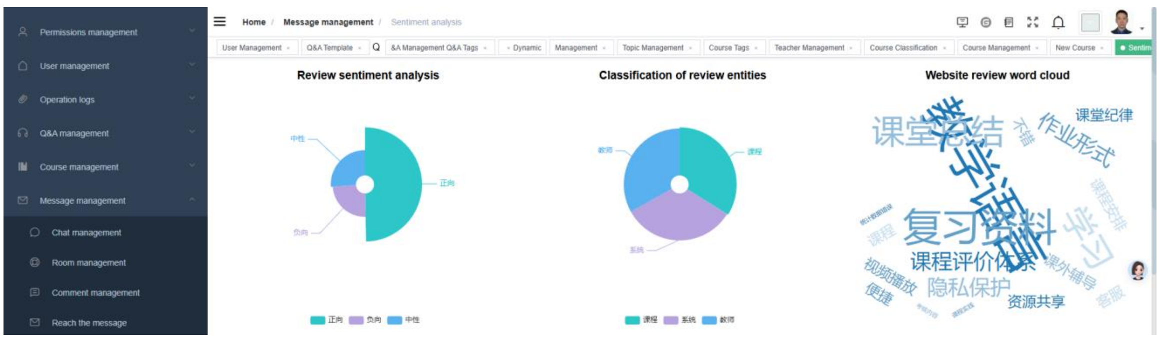

Figure 6. Mushroom Community Chinese course comments sentiment analysis web page.

DISCUSSION

Although all models have the same accuracy in the Chinese online course sentiment analysis task, LSTM has better recall and f1 score, so LSTM has become the preferred model used in the development of the Mushroom Community online learning system. LSTM is more suitable for processing complex language structures and sentiment information in long texts. The traditional LR, SVM, and RF three-week machine learning models perform well in simple classification tasks but are slightly insufficient in terms of context and sequence dependencies.

The results of the loss curve, accuracy curve, and learning rate curve show that the LSTM model trained by the researchers has no obvious overfitting or underfitting phenomenon, and the model has a high accuracy rate and can be used and integrated by the Mushroom Community online learning system.

There are three possible reasons why the LSTM model performs better on humanities course reviews than STEM courses. First, humanities reviews often use richer sentiment vocabulary and subjective expressions, making it easier for the model to capture sentiment tendencies. Second, STEM course reviews contain a lot of professional terms and technical descriptions, which may cause the model to have difficulty understanding sentiment tendencies. Third, in humanities reviews, positive and negative sentiment expressions are more obvious, while STEM course reviews may tend to be more neutral or objective descriptions, increasing the difficulty of classification.

Although the LSTM model performs well in the overall sentiment classification task, it still faces certain challenges when dealing with certain complex language structures. Through manual analysis, researchers found that the comment texts that the LSTM model made mistakes in can be roughly divided into scenarios such as those containing sarcastic and ironic expressions, containing multiple layers of negation, and containing professional terms. In scenarios with sarcasm and irony, these comments use positive words on the surface, but in context, they express negative emotions. The model may be misled by the surface words, resulting in sentiment classification errors. In scenarios with multiple layers of negation, although these comments contain positive words, they usually contain one or more negative words. The model may ignore the role of negative words, resulting in incorrect comment classification. In scenarios with professional terms, the low frequency of rare professional terms in the training corpus, poor embedding quality, and insufficient information also makes the model prone to errors.

After the Mushroom Community online learning system successfully integrates the LSTM model, it can automatically classify and intuitively display the sentiment polarity of the course review articles published by students in the Mushroom Community, providing reliable data support for the optimization of course resources.

CONCLUSIONS AND RECOMMENDATIONS

This study provides a comprehensive technical path and empirical basis for sentiment analysis of comment data in the Mushroom Community online learning system, from data collection and preprocessing, model selection and training, to comparison and adjustment of evaluation indicators. The experimental results show that the combination of the Word2Vec word vector model and LSTM neural network performs better than the traditional three-machine models of logistic regression (LR), support vector machine (SVM), and random forest (RF). It has a stronger ability to capture contextual relationships and semantic continuity in Chinese comment texts. In addition, through sample balancing and data cleaning, deduplication, deletion of invalid text, text segmentation stop words, and other preprocessing operations, the effect of the LSTM model in the sentiment analysis task of Chinese course comments can be effectively improved. The researchers successfully integrated the trained model into the self-developed online learning system Mushroom Community, realized the automated sentiment analysis of Chinese online course comments, and greatly improved the intelligence level of the online learning system. However, the currently constructed LSTM Chinese course review sentiment model still has the following two problems that need to be solved: First, the model is prone to prediction errors in the face of irony, negation, and subtle expressions. Second, the LSTM model is overly dependent on manually labeled data, and the human bias in manual labeling has a greater impact on the model's prediction effect.

To reduce the influence of personal subjective consciousness on the model when manually annotating data, and at the same time improve the prediction quality of the model in various complex scenarios such as irony, negation, and subtle expressions, researchers plan to try to use more advanced pre-trained language models in the future, such as BERTT, XLNet, and RoBERTa. The multi-head attention mechanism in the pre-trained language model makes it better than LSTM when facing Chinese course reviews in complex scenarios. In addition, the pre-trained language model can realize sentiment analysis tasks under small samples, which will greatly reduce the impact of the quality of manually annotated data on the model effect. However, the pre-trained model has a complex structure and requires high hardware resources for training and deployment. We need to explore data distillation technology (Aquino et al., 2025) to reduce the resource consumption of the model before it can be integrated into the Mushroom Community online learning system. In addition, the research also plans to train and integrate more fine-grained sentiment analysis models to achieve more sophisticated sentiment analysis and management, so as to more comprehensively understand learners' needs and problems.

To further improve the intelligent service level of the Mushroom Community online learning system, provide teachers and students with a better online teaching experience, and promote the continuous innovation and development of educational informatization, researchers will continue to optimize the LSTM model and service performance, including continuous tuning and version upgrades of the LSTM model to improve performance and response speed; regular updates of language models and corpora to ensure more accurate sentiment analysis; using the system backend visualization analysis capabilities, regularly collecting user feedback, and continuously optimizing user experience and teaching effects through rapid iteration of model parameters and system functions.

IMPLICATIONS

This study explores the detailed process and steps of applying the LSTM model to the sentiment analysis task of Chinese online course reviews and also studies the method and feasibility of integrating the sentiment analysis model into the online learning system. The research results are of great significance to many fields such as online education platform operators, teachers, and students. The sentiment analysis results of this study can provide important inspiration for online education platforms, educational institutions, and teachers in terms of course quality optimization, teaching strategy adjustment, and sustainable development of online learning systems. In addition, the results of this study also provide useful references for governments and education policymakers in optimizing the allocation of educational resources, online education supervision, and promoting the application of artificial intelligence in the field of education.

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to all those who helped to complete this paper.

FUNDING

The study is self-funded.

DECLARATIONS

Conflict of Interest

The authors declare that they have no competing interests.

Informed Consent

not applicable

Ethics Approval

not applicable

REFERENCES

Alfarhood, S. (2025). Soutcom: Real-time sentiment analysis of Arabic text for football fan satisfaction using a bidirectional LSTM. *Expert Systems, 42(2),* e13641. https://doi.org/10.1111/exsy.13641

Aquino, J. A., Liew, D. J., & Chang, Y. C. (2025). Graph-aware pre-trained language model for political sentiment analysis in Filipino social media. Engineering Applications of Artificial Intelligence, 146, 110317. https://doi.org/10.1016/j.engappai.2025.110317

Arif, M., Hasan, M., Al Shiam, S. A., Ahmed, M. P., Tusher, M. I., Hossan, M. Z., ... & Imam, T. (2024). Predicting Customer Sentiment in Social Media Interactions: Analyzing Amazon Help Twitter Conversations Using Machine Learning. *International Journal of Advanced Science Computing and Engineering, 6(2),* 52-56. https://doi.org/10.62527/ijasce.6.2.211

Cahapin, E. L., Santiago, J. C. S., Malabag, B. A., Reyes, J. L., Legaspi, G. S., & Benedicto, J. (2023). Sentiment analysis of students' perception towards the implementation of the limited in-person learning: A post-pandemic perspective. *International Journal of Computing Sciences Research, 7,* 1664-1684. https://doi.org/10.25147/ijcsr.2017.001.1.126

Celestial-Valderama, A. M., Vinluan, A., & Moraga, S. D. (2021). Mining students' feedback in a general education course: Basis for improving blended learning implementation. *International Journal of Computing Sciences Research, 5,* 568-583. https://doi.org/10.25147/ijcsr.2017.001.1.60

Jatain, D., Singh, V., & Dahiya, N. (2023). An exploration of the causal factors making an online course content popular & engaging. *International Journal of Information Management Data Insights, 3(2),* 100194. https://doi.org/10.1016/j.jjimei.2023.100194

Kastrati, Z., Imran, A. S., & Kurti, A. (2020). Weakly supervised framework for aspect-based sentiment analysis on students' reviews of MOOCs. *IEEE Access, 8,* 106799-106810. https://doi.org/10.1109/ACCESS.2020.3000739

Khan, T. A., Sadiq, R., Shahid, Z., Alam, M. M., & Su'ud, M. B. M. (2024). Sentiment analysis using support vector machine and random forest. *Journal of Informatics and Web Engineering, 3(1),* 67-75. https://doi.org/10.33093/jiwe.2024.3.1.5

Koufakou, A. (2024). Deep learning for opinion mining and topic classification of course reviews. *Education and Information Technologies, 29(3),* 2973-2997. https://doi.org/10.1007/s10639-023-11736-2

Mamidted, A. D., & Maulana, S. S. (2023). The Teaching Performance of the Teachers in Online Classes: A Sentiment Analysis of the Students in a State University in the Philippines. *Randwick International of Education and Linguistics Science Journal, 4(1),* 86-95. https://doi.org/10.47175/rielsj.v4i1.639

Mantika, A. M., Triayudi, A., & Aldisa, R. T. (2024). Sentiment analysis on Twitter using Naïve Bayes and logistic regression for the 2024 presidential election. *SaNa: Journal of Blockchain, NFTs and Metaverse Technology, 2(1),* 44-55. https://doi.org/10.58905/sana.v2i1.267

Murithi, M. K., Oirere, A. M., Ndung'u, R. N., Murithi, M. K., Oirere, A. M., & Ndung'u, R. N. (2024). A Systematic Review of the Sentiment Analysis Models Used in Handling Polarity Shift. *International Journal of Computing Sciences Research, 8,* 2635-2676. https://doi.org/10.25147/ijcsr.2017.001.1.179

Onan, A. (2020). Mining opinions from instructor evaluation reviews: a deep learning approach. *Computer Applications in Engineering Education, 28(1),* 117-138. https://doi.org/10.1002/cae.22179

Onan, A. (2021). Sentiment analysis on massive open online course evaluations: a text mining and deep learning approach. *Computer Applications in Engineering Education, 29(3),* 572-589. https://doi.org/10.1002/cae.22253

Rainio, O., Teuho, J., & Klén, R. (2024). Evaluation metrics and statistical tests for machine learning. *Scientific Reports, 14(1),* 6086. https://doi.org/10.1038/s41598-024-56706-x

Santiago, C. J. S., Centeno, Z. J. R., Ulanday, M. L. P., & Cahapin, E. L. (2022). Sentiment analysis of students' experiences during online learning in a state university in the Philippines. *International Journal of Computing Sciences Research, 7,* 1287-1305. https://doi.org/10.25147/ijcsr.2017.001.1.102

Shaik, T., Tao, X., Dann, C., Xie, H., Li, Y., & Galligan, L. (2023). Sentiment analysis and opinion mining on educational data: A survey. *Natural Language Processing Journal, 2,* 100003. https://doi.org/10.1016/j.nlp.2022.100003

Yee, M., Roy, A., Perdue, M., Cuevas, C., Quigley, K., Bell, A., ... & Miyagawa, S. (2023, September). AI-assisted analysis of content, structure, and sentiment in MOOC discussion forums. *Frontiers in Education, 8,* 1250846. https://doi.org/10.3389/feduc.2023.1250846

Yu, S., Androsov, A., Yan, H., & Chen, Y. (2024). Bridging computer and education sciences: A systematic review of automated emotion recognition in online learning environments. *Computers & Education, 220*, 105111. https://doi.org/10.1016/j.compedu.2024.105111

Author's Biography

Junchao Dong is currently pursuing a Doctorate in Information Technology at the University of the East, Manila, in the Philippines. His research interest is the application of artificial intelligence technology in education.

Dr. Joan P. Lazaro is a full-time professor at the College of Engineering, Computer Engineering Department, and a special lecturer of IT programs at the Graduate School at the University of the East. He is a graduate of Doctor of Information Technology and Master of Engineering Science from the University of the East Manila Graduate School and a Bachelor of Science in Computer Engineering from the University of the East Caloocan. Among the different certifications he earned are the following: Professional Computer Engineer, Fortinet's Network Security Expert Certification – NSE 1 and 2 Network Security Associate, Certified Microsoft Innovative Educator Program, and National Certificate II in Mechatronics Servicing. His research interests include software development, network security, and engineering sciences.