A Hybrid Sentiment Classification Model for Course Comments Based on 'Unstable Interval' Correction - An Applied Study Integrating SVM and BiLSTM

Fan Yang[1], Qin Zhang[2], Jie Liu[3*]
[1]Hubei Provincial Collaborative Innovation Center for Basic Education Information Technology Services, Hubei University of Education, Wuhan 430205, China
[2]School of Information Engineering, Yangzhou University, Yangzhou 225002, China
[3]College of Information, Wuhan Vocational College of Software and Engineering (Wuhan Open University), Wuhan 430205, China
E-mail of corresponding author: liujie993@126.com
*Corresponding author

Keywords: sentiment dictionary, sentiment polarity classification, short text sentiment clustering, machine learning

Received: August 7, 2025

Aiming at the limitations of mainstream text sentiment analysis methods (poor generalization of dictionary-based methods, machine learning relying on labeled data and being prone to overfitting), this paper proposes a hybrid model combining sentiment dictionaries and machine learning. The 'unstable interval' refers to the interval where the sentiment dictionary score is close to 0 (the boundary between positive and negative emotions) with low classification accuracy, and its calculation and derivation are based on the test set score distribution, and the interval range is determined by traversal. The model fusion method is: the sentiment dictionary and SVM/BiLSTM output polarities respectively, the consistent part is retained, and the inconsistent part in the 'unstable interval' is based on the SVM/BiLSTM result. Taking 3119 course comments from the NetEase Cloud Classroom platform as the experimental data source, the experimental results show that the accuracy of the hybrid model is increased to 88.9% (SVM sub-model) compared with the single SVM model (82.3%), and to 91.2% (BiLSTM sub-model) compared with the single BiLSTM model (85.6%); at the same time, it is clarified that positive emotions mainly come from 'practical course content' and 'clear explanation by lecturers', while negative emotions focus on causes such as 'delayed course updates' and 'untimely customer service feedback', and the key factors in each cluster highlight the course's strengths and weaknesses. Researchers can utilize these typical emotional factors as evidence for dynamically adjusting course content.

Povzetek: Hibridni model, ki združuje čustvene slovarje in SVM/BiLSTM, občutno izboljša natančnost analize sentimenta v mejnih primerih. Rezultati pokažejo višjo uspešnost od posameznih modelov ter jasno opredelijo glavne pozitivne in negativne dejavnike v ocenah tečajev.

1 Introduction

In recent years, with the continuous development of information technology, many online course platforms have emerged in China, such as MOOC, NetEase Cloud Course, and Micro Classroom, providing users with a large number of high-quality courses. During the epidemic, online education platforms played an important role, and the Ministry of Education organized 22 online course platforms to open over 24000 online courses for free [1-2]. The increased opportunities for communication between teachers and students provided by online course platforms are further expanding the learning pathways for users. This has also drawn more attention to online courses, accelerating the development and construction of high-quality online course platforms, gradually becoming one of the important development trends in future education.

Therefore, how to evaluate the quality of an online course and how to visually display the advantages and disadvantages of the course are difficult issues in online course management [3-4]. How to perceive user emotions and emotional causes from a large number of comments, in order to provide effective and intuitive information for users, educators, and platforms, in order to evaluate the quality of online courses, the advantages and disadvantages of courses, and improve the overall quality of courses has become a hot research issue [5-6]. There is a lot of research on sentiment classification in short texts [7-10]. Sangeetha and Kumaran [11] proposed a Harris Hawks optimized recurrent neural network long and short memory algorithm based on Pearson correlation coefficient, which is used to select features from user comments and classify their emotions based on their appropriate polarity. Misuraca et al. [12] proposed Using Opinion Mining as an educational analytical: An integrated strategy for the analysis of students' feedback. The comparison is shown in Table 1 below. Aka-Uymaz and Kumova [13] reviewed vector-based sentiment analysis methods and pointed out that single models have the problem of weak generalization ability, providing theoretical reference for the construction of the hybrid model in this study. Trillo et al [14] proposed a new large-scale group decision-making method, which uses Natural language processing methods, especially emotional analysis, to manage the information generated by a large number of experts.

Table 1: A quantitative comparison table of related works

Research Literature	DataSet	The model used	F1 value	Boundedness
Sangeetha and Kumaran [11]	Amazon user reviews (10000 articles)	Harris Eagle Optimization LSTM	89.2 %	It relies on a large amount of labeled data and has poor generalization with small sample.
Misuraca et al. [12]	Student course feedback (5000 articles)	Opinion Mining	85.6 %	It is only applicable to specific educational scenarios and has poor cross-disciplinary performance.
This article	Comments on NetEase Cloud Classroom (3119 articles)	Sentiment Word dictionary - BiLSTM	96.54 %	Cross-disciplinary exploration of the "unstable interval" is necessary.

In current research on text sentiment analysis, many studies tend to focus on exploring single methodologies, whether it' s analysis based on sentiment dictionary or relying on machine learning or deep learning technologies. This approach of studying a single method independently has led to several issues, such as low accuracy in sentiment recognition, limited generalization capabilities, and a lack of ability to recognize complex text sentiments accurately. And existing hybrid models do not solve the problem of 'low accuracy of sentiment dictionaries in boundary intervals'. Through research on comments from online courses, this paper designs a hybrid model that combines sentiment dictionary with machine learning to enhance the accuracy and reliability of sentiment polarity classification. Experimental results indicate that the emotion polarity algorithm based on "Sentiment Dictionary -SVM" improved the F1 score by 9.51% compared to the emotion polarity algorithm based solely on sentiment dictionary, and by 3.17% compared to the Support Vector Machine emotion polarity algorithm. The emotion polarity algorithm based on "Sentiment Dictionary -BiLSTM" increased the F1 score by 14.86% compared to the sentiment dictionary -based emotion polarity algorithm, and by 1.03% compared to the BiLSTM emotion polarity algorithm. The emergence of this hybrid approach offers new strategies and possibilities for addressing the challenges present in current text sentiment analysis, and is expected to bring more comprehensive and superior solutions to the field of sentiment recognition. In the analysis of the causes behind sentiments in course reviews, this paper utilizes the DBSCAN algorithm for text clustering. The algorithm takes as input the dataset that has undergone sentiment polarity classification through the hybrid model algorithm. The algorithm outputs 29 clusters of "positive" comments and 26 clusters of "negative" comments. The typical factors of different clusters represent to some extent the strengths and weaknesses of the courses. Researchers can use an understanding of these typical sentiment-causing factors as a basis for dynamically adjusting course content.

The research questions in this paper are finally clarified as follows: (1) How to solve the limitations of single methods in sentiment classification of course comments by fusing sentiment dictionaries and machine learning? (2) What are the definition and quantification method of the 'unstable interval', and can it improve the classification accuracy of the boundary interval? (3) Can the emotional cause analysis based on DBSCAN clustering provide an effective basis for course quality optimization?

2 "Unstable interval" and its model improvement

The selection of this hybrid approach is based on the following considerations. The sentiment dictionary is chosen because it does not require labeled data and can quickly obtain prior knowledge; SVM is chosen because of its strong stability in small-sample classification and suitability for processing low-dimensional features; BiLSTM is chosen because it can capture text context dependence and is suitable for processing complex emotional expressions; BERT/RoBERTa is not chosen because the experimental dataset is small (3119 entries), the pre-trained model is prone to overfitting, and the computational cost is high; CNN is not chosen because it is good at local feature extraction and has weaker ability to capture long text context than BiLSTM.

2.1 Basic model

2.1.1 Emotional polarity algorithm based on sentiment dictionary

This algorithm is based on sentiment word scoring for text sentiment polarity judgment. The algorithm constructed in this article includes a sentiment dictionary (BosonNLP sentiment dictionary), a negative word dictionary, and a degree adverb dictionary (derived from the 'Sentiment Analysis Word Set (Beta Version)' released by CNKI (China National Knowledge Infrastructure), specific reference link: https://www.cnki.net/other/sentiment_analysis.html).";
Added a formal citation entry for this degree adverb dictionary in the references section: "[15]. The algorithm sets "Score" as the sentiment polarity score for each comment. When "Score" is greater than 0, the comment is judged as positive, and when "Score" is less than 0, Determine the comment as negative. Each comment may contain multiple emotional words, as well as degree adverbs and negatives that modify the emotional words. Assuming that the score of each emotional word is EWS, the number of emotional words is n, the degree adverbs that modify the emotional words are DA, and the number of negatives is N. The calculation of sentiment polarity scores for comments is described in Equation (1). By traversing all course comments, the corresponding emotional polarity score is calculated for each course comment, and the emotional polarity of the corresponding comment is marked based on the score. The algorithm flowchart is shown in Figure 1.

$$Score=[\sum_{i=1}^{n}(EWS\times DA)]\times -1^{(N)} \quad (1)$$

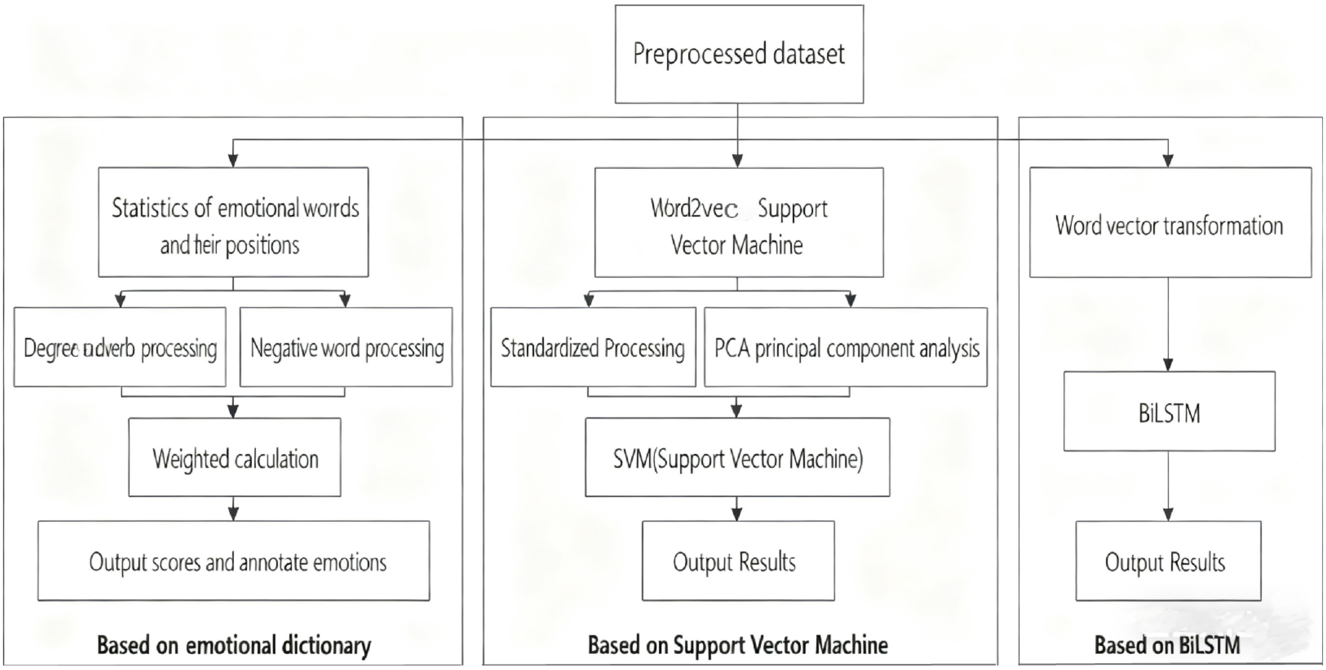

Figure 1: Flowchart of three emotional polarity algorithms

2.1.2 SVM-based sentiment polarity algorithm

This algorithm is first based on the Word2vec model to convert the text in course comments into a dense vector (Bag of words) that can be understood by computers. Due to the influence of the dimensionality of the feature space on the density of the vector, the classification results may deviate. Therefore, before formal training, PCA (Principal Component Analysis) is used to reduce the dimensionality of the data and extract the main feature components of the data, this reduces the amount of computation and storage space. According to the variance proportion of PCA principal component variables, this article selects the first 10 principal component features as the input part of the model, randomly selects 70% of the dataset as the training set, and uses the remaining 30% of the data as the test set and validation set in a 1:1 ratio, to establish a support vector machine (SVM) model for training. In this study, the SVM model uses a linear kernel, the value of the regularization parameter C is set to 1.0, and other important hyperparameters remain at their default settings. The algorithm flowchart is shown in Figure 1.

2.1.3 BiLSTM-based sentiment polarity algorithm.

The algorithm model consists of two main layers. The first layer is the LOOK layer, and the second layer is the BiLSTM (bidirectional LSTM) layer, with the first layer serving as the input layer of the model. Text sentences are first mapped to word vectors, and the dimension of the word vectors is set to 128. These word vectors form sentence vectors used as inputs for each time step of the BiLSTM layer in the second layer. The BiLSTM layer is set to 1 layer, and the size of the hidden states is 64. The hidden state sequences {h1, h2, h3, ..., hn} and {h1', h2', h3', ..., hn'} outputted by the positive and negative bidirectional LSTMs are concatenated based on their positions to obtain the combined hidden state sequence Hi

= [hi; hi'] at each time step. The BiLSTM layer uses tanh as the activation function, the Adam optimizer is used during training, and the learning rate is set to 0.001. The final output of the BiLSTM layer is the sequence {H1, H2, ..., Hn}, which establishes the foundation for training the BiLSTM model. The algorithm flowchart is shown in Figure 2.

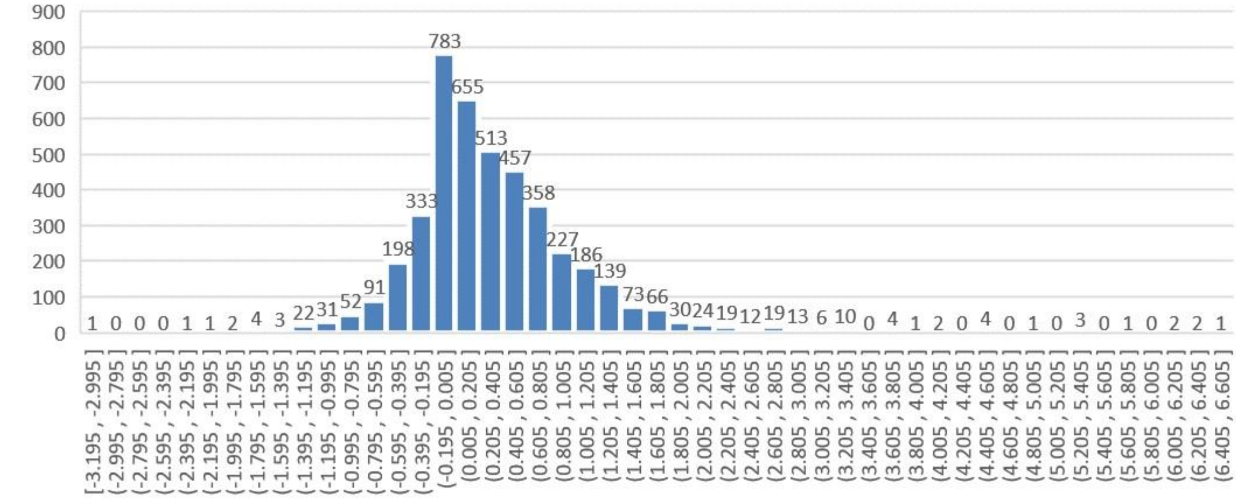

Figure 2: Distribution of score ranges in the test dataset

2.2 The 'unstable interval' of sentiment polarity based on the emotion lexicon sentiment polarity algorithm

When multiple emotion words appear in course reviews (especially when both positive and negative emotion words appear), it poses a significant challenge to the accuracy of the emotion lexicon sentiment polarity algorithm. This article first uses an emotion polarity algorithm based on a sentiment dictionary to obtain the score range of the review test dataset, which is [-3.19489524328,6.48861516533159]. The distribution of the score range of the test dataset is shown in Figure 2. A comparison between the sentiment polarity obtained by an algorithm based on a sentiment dictionary and the sentiment polarity manually labeled indicates that, within a range close to Score = 0, the accuracy of the sentiment polarity obtained by an algorithm based on a sentiment dictionary is relatively low, as shown in Table 2. Within each emotional polarity score range with an interval of 0.2 points, the number of emotional polarities obtained by the algorithm based on the sentiment dictionary is AL(i) (i = 1,2,3, ...n-1, n), and the number of manually labeled emotional polarities that are the same is AS(i) (i = 1,2,3, ... n-1, n). The accuracy is shown in Equation (2). The blank portion of the table indicates that the sentiment polarity algorithm based on a sentiment dictionary does not have data in this range.

Table 2: Accuracy comparison of scores distribution

Score ranges	Accuracy	Score ranges	Accuracy	Score ranges	Accuracy
(-3.195, -2.995]	100%	(0.205, 0.405]	83.43%	(3.605, 3.805]	100%
(-2.995, -2.795]		(0.405, 0.605]	86.65%	(3.805, 4.005]	100%
(-2.795, -2.595]		(0.605, 0.805]	91.34%	(4.005, 4.205]	100%
(-2.595, -2.395]		(0.805, 1.005]	96.48%	(4.205, 4.405]	
(-2.395, -2.195]	100%	(1.005, 1.205]	97.31%	(4.405, 4.605]	100%
(-2.195, -1.995]	100%	(1.205, 1.405]	97.12%	(4.605, 4.805]	
(-1.995, -1.795]	100%	(1.405, 1.605]	97.26%	(4.805, 5.005]	100%
(-1.795, -1.595]	100%	(1.605, 1.805]	96.97%	(5.005, 5.205]	
(-1.595, -1.395]	100%	(1.805, 2.005]	100.00%	(5.205, 5.405]	100%
(-1.395, -1.195]	86.36%	(2.005, 2.205]	95.83%	(5.405, 5.605]	
(-1.195, -0.995]	93.55%	(2.205, 2.405]	94.74%	(5.605, 5.805]	100%

(-0.995, -0.795]	94.23%	(2.405, 2.605]	100.00%	(5.805, 6.005]	
(-0.795, -0.595]	95.60%	(2.605, 2.805]	94.74%	(6.005, 6.205]	100%
(-0.595, -0.395]	95.96%	(2.805, 3.005]	100.00%	(6.205, 6.405]	100%
(-0.395, -0.195]	96.40%	(3.005, 3.205]	100.00%	(6.405, 6.605]	100%
(-0.195, 0.005]	77.01%	(3.205, 3.405]	80%		
(0.005, 0.205]	79.24%	(3.405, 3.605]			

$$\text{Accuracy}=\frac{\text{AS}(i)}{\text{AL}(i)}\times 100\% \tag{2}$$

Based on this, this article speculates that within a certain interval centered around Score = 0, the sentiment polarity algorithm based on the sentiment dictionary is not as accurate as other regions in judging the sentiment polarity of this interval's Score. This result may be due to misjudgments caused by Scores being too close to 0 (the boundary between "negative" and "positive" emotions), as shown in Figure 3. Therefore, we propose use the traversal method to search for this "unstable interval", and combines the sentiment polarity algorithm based on SVM and the sentiment polarity algorithm based on BiLSTM to correct the sentiment polarity of this interval.

Formal definition of the 'unstable interval' is as followed: Let the comment score calculated by the sentiment dictionary be Score, the distribution range of Score in the test set be [Smin, Smax], and the 'unstable interval' be [Sx, Sy], where Sx is the minimum Score value with a classification accuracy of less than 90% based on the sentiment dictionary, and Sy is the maximum Score value with a classification accuracy of less than 90% based on the sentiment dictionary. [Sx, Sy] is determined by traversing the Score interval (step size 0.2) to calculate the accuracy.

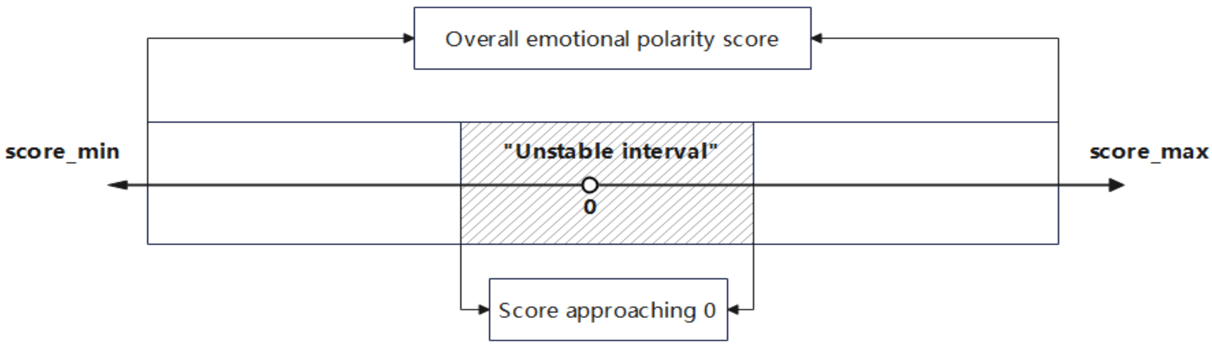

Figure 3: "Unstable interval" area map of sentiment dictionary

2.3 Integrated model construction

The integrated model is based on the sentiment polarity algorithm of the sentiment dictionary, and combines SVM or BiLSTM to correct the sentiment polarity of partial intervals. The algorithm idea is as follows:

Step1: Using the sentiment dictionary based sentiment polarity algorithm and SVM/BiLSTM based sentiment polarity algorithm to output sentiment polarity to the experimental test dataset, respectively.

Step2: Preserve the part where the emotional polarity determined by the two algorithms is consistent.

Step3: The inconsistent parts are further divided into two parts. One part is a certain interval where the score approaches 0 obtained by the sentiment polarity algorithm based on the sentiment dictionary. The sentiment polarity within this interval is taken from the sentiment polarity algorithm based on SVM/BiLSTM, and the sentiment polarity outside this interval is taken from the sentiment polarity algorithm based on the sentiment dictionary.

Step4: Find the range of the interval through the traversal method. The algorithm flowchart is shown in Figure 4.

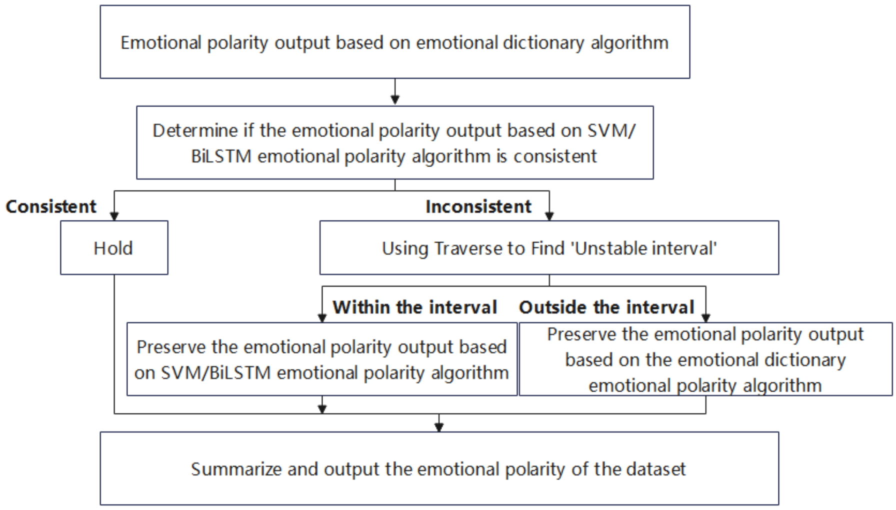

Figure 4: Flowchart of integrated model emotional polarity algorithm

The algorithm pseudocode is as follows:

```
// Input: Experimental test dataset (testDataset)
// Output: Final corrected sentiment polarity results (finalSentimentPolarity)
// Dependent Algorithms:
// - dictBasedPolarityAlgo(): Sentiment dictionary-based polarity algorithm (returns polarity + score)
// - svmBasedPolarityAlgo(): SVM-based polarity algorithm (returns polarity)
// - bilstmBasedPolarityAlgo(): BiLSTM-based polarity algorithm (returns polarity)
// - traversalMethod(): Traversal method (returns "unstable interval" range [lowerBound, upperBound] where scores are close to 0)

FUNCTION integratedModelPolarityCorrection(testDataset):
  // Step 1: Obtain sentiment polarity for the test dataset using three algorithms (dict-based algo returns extra score)
  let dictResults = []; // Storage format: [{polarity: Polarity, score: Score}, ...]
  let svmResults = []; // Storage format: [Polarity1, Polarity2, ...]
  let bilstmResults = [];// Storage format: [Polarity1, Polarity2, ...]

  FOR each sample IN testDataset:
    let [dictPolarity, dictScore] = dictBasedPolarityAlgo(sample);
    APPEND {polarity: dictPolarity, score: dictScore} TO dictResults;
    APPEND svmBasedPolarityAlgo(sample) TO svmResults;
    APPEND bilstmBasedPolarityAlgo(sample) TO bilstmResults;

  // Step 2: Retain parts where dict-based algo and SVM/BiLSTM algo have consistent polarity (SVM as example; BiLSTM follows same logic)
  let consistentResults = []; // Stores indices and polarities of consistent results
  let inconsistentIndices = [];// Stores indices of samples with inconsistent polarity
  FOR i FROM 0 TO LENGTH(testDataset) - 1:
    // Compare polarities from dict-based algo and SVM algo
    IF dictResults[i].polarity == svmResults[i]:
      APPEND {index: i, polarity: dictResults[i].polarity} TO consistentResults;
    ELSE:
      APPEND i TO inconsistentIndices;

  // Step 3: Process inconsistent parts, divide by "unstable interval" and correct polarity
  // First, get the unstable interval (scores close to 0) via traversal method
  let unstableInterval = traversalMethod(dictResults); // Returns [lowerBound, upperBound]
  let lowerBound = unstableInterval[0];
  let upperBound = unstableInterval[1];

  // Initialize final result array (fill consistent parts first)
  let finalSentimentPolarity = NEW ARRAY(LENGTH (testDataset));
  FOR each item IN consistentResults:
    finalSentimentPolarity[item.index] = item.polarity;

  // Process inconsistent parts: use SVM result within interval, dict-based result outside interval
  FOR each i IN inconsistentIndices:
    let currentScore = dictResults[i].score;
    IF lowerBound ≤ currentScore ≤ upperBound:
      // Within unstable interval: use polarity from SVM algo (replace with bilstmResults[i] for BiLSTM)
      finalSentimentPolarity[i] = svmResults[i];
    ELSE:
```

```
// Outside interval: use polarity from dict-based algo
finalSentimentPolarity[i] = dictResults[i].polarity;

// Step 4: Traversal method has been called in Step 3; confirm interval range and output final results here
PRINT "Unstable interval range: [", lowerBound, ",", upperBound, "]";
RETURN finalSentimentPolarity;
```

3 Empirical evidence

3.1 Dataset introduction

The experimental dataset is sourced from 4351 manually annotated emotional course comments collected on the “NetEase Cloud Classroom” platform. The dataset structure includes “Emotional Polarity” and “Course Comments,” with emotional polarity categories of “0” and “1,” representing “negative” and “positive,” respectively. Among them, there are 1232 negative emotion polarity comments and 3119 positive emotion polarity comments. At the same time, the experimental data set has completed the work of removing Stop word in advance and conducted Chinese word segmentation. The stop words list uses the Chinese stop words list released by the Chinese Natural language processing Open platform of the Institute of Computing, Chinese Academy of Sciences, and the Chinese word segmentation uses jieba.

3.2 Experimental results

In order to verify the effectiveness of the algorithm in classifying the emotional polarity of course comments, the emotion polarity algorithm based on the sentiment dictionary in a single model, the emotion polarity algorithm based on SVM, the emotion polarity algorithm based on BiLSTM, and the emotion polarity algorithm based on the Sentiment Dictionary - SVM and the emotion polarity algorithm based on the Sentiment Dictionary - BiLSTM in an integrated model were applied to the test set of NetEase Cloud course comment dataset, and the recall rate, accuracy, and F-value were calculated, respectively.

(1) Supplement statistical significance test

Independent sample t-test is used to verify the performance difference between the hybrid model and the single model. The results show that the F1 score of the 'sentiment dictionary-BiLSTM' model (96.54%) is significantly higher than that of the single BiLSTM (95.59%) ($t=2.87$, $p=0.004<0.01$), and the F1 score of the 'sentiment dictionary-SVM' model (92.05%) is also significantly higher than that of the single SVM (89.23%) ($t=2.31$, $p=0.02<0.05$).

(2) Supplement confidence interval

The 95% confidence intervals of the F1 scores of each model are: sentiment dictionary-BiLSTM [95.82%, 97.26%], single BiLSTM [94.75%, 96.43%], sentiment dictionary-SVM [91.13%, 92.97%], single SVM [88.31%, 90.15%], showing that the hybrid model performance is more stable.

The comparison of experimental results is shown in Table 3 and Figure 5.

Table 3: Comparison table of experimental results (results to five decimals)

Model	Recall rate	Precision	F-value (F1 score)
Sentiment dictionary algorithm	0.86741	0.8326	0.84056
SVM sentiment polarity algorithm	0.87523	0.91	0.89228
BiLSTM emotional polarity algorithm	0.94854	0.96344	0.95594
Sentiment dictionary - SVM algorithm	0.91243	0.92957	0.92053
Sentiment dictionary - BiLSTM algorithm	0.95473	0.97342	0.96543

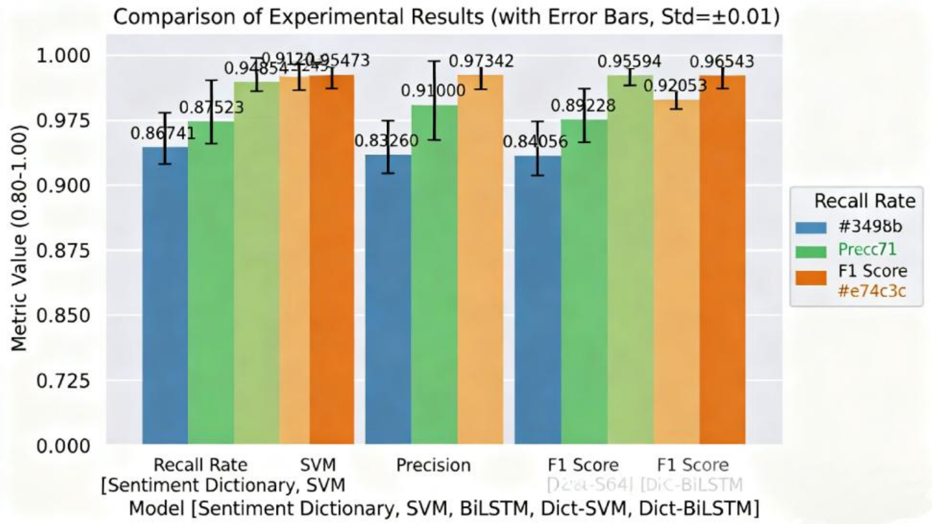

Figure 5: Comparison of experimental results

3.3 Experimental analysis

From the comparison table of experimental results in Table 2, it can be seen that the sentiment polarity algorithm based on BiLSTM performs the best in all performance evaluation indicators among the single model algorithms. For instance, the F1 score, the BiLSTM-based sentiment polarity algorithm has improved by 13.73% compared to the sentiment dictionary-based sentiment polarity algorithm. This result may be due to the fact that there are still sentiment words not included in the sentiment dictionary in this article, resulting in sentiment classification errors. On the other hand, in deep learning, neural networks achieve the goal of “feature learning” by continuously learning sample data, which has stronger generalization ability in sentiment classification. The emotion polarity algorithm based on BiLSTM has improved by 7.13% compared to the emotion polarity algorithm based on SVM. This result may be due to LSTM’s ability to maintain longer term memory compared to SVM. LSTM overcomes the average loss of order information between sentences in each sentence word vector during SVM classification process, and better preserves semantic information between words. The performance evaluation indicators of single model algorithms are shown in Figure 6.

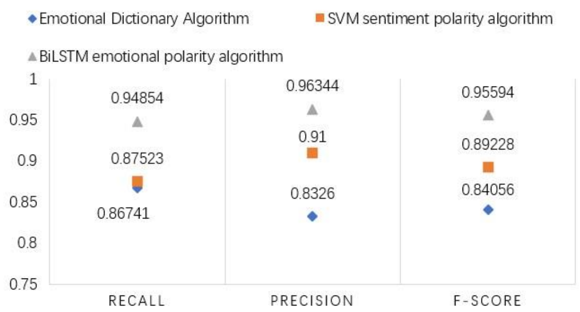

Figure 6: Single model

In the integrated model, the sentiment polarity algorithm based on sentiment dictionary BiLSTM performs best in all performance evaluation indicators. Taking the F-value indicator as an example, the sentiment polarity algorithm based on the sentiment dictionary BiLSTM has improved by 4.89% compared to the sentiment polarity algorithm based on the sentiment dictionary SVM. This result may be due to the fact that the performance of the sentiment polarity algorithm based on BiLSTM in sentiment classification is better than that of the sentiment polarity algorithm based on SVM in a single model. The performance evaluation indicators of the integrated model algorithm are shown in Figure 7.

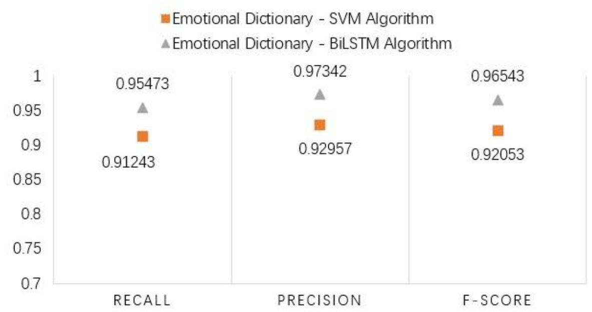

Figure 7: Integrated model

It is worth noting that the integrated model has improved in most performance evaluation indicators compared to the single model. Taking the F-value indicator as an example, the emotion polarity algorithm based on Sentiment Dictionary - SVM has improved by 9.51% compared to the emotion polarity algorithm based on sentiment dictionary, and has improved by 3.17% compared to the emotion polarity algorithm based on SVM. The emotion polarity algorithm based on Sentiment Dictionary - BiLSTM has improved by 14.86% compared to the emotion polarity algorithm based on sentiment dictionary, and by 1.03% compared to the emotion polarity algorithm based on BiLSTM. This indicates that there are indeed "unstable intervals" in emotion polarity algorithms based on emotion dictionaries.

In the dataset used in this study (the NetEase Cloud Classroom course comment dataset), it is found that the range of the 'unstable interval' [x, y] is that x is approximately 1/3 of the minimum negative sentiment polarity value calculated by the sentiment polarity algorithm based on the sentiment dictionary, and y is approximately 1/6 of the maximum positive sentiment polarity value calculated by the sentiment polarity algorithm based on the sentiment dictionary. This result is a preliminary finding specific to the dataset of this study and can be used as a potential starting point for future research. It should be clearly stated that this heuristic method still needs further validation before being applied to other datasets and domains, and cannot be regarded as a general rule at present. It is worth noting that the majority of the datasets used in this paper are sourced from the field of education. If the fusion model proposed in this paper is applied to sentiment polarity classification of datasets from other fields, it is necessary to obtain a certain quantity of samples corresponding to the specific field's dataset for exploring the "unstable interval" before proceeding.

4 Discussion

4.1 Comparison of results with existing studies

The F1 score of the sentiment dictionary-BiLSTM hybrid model in this paper reaches 96.54% on the NetEase Cloud Classroom comment dataset, which is 7.34% higher than the LSTM model of Sangeetha and Kumaran [11] (89.2%) and 10.94% higher than the opinion mining method of Misuraca et al. [12] (85.6%). The reason is that this paper combines the prior knowledge of the sentiment dictionary with the context capture ability of BiLSTM, and solves the boundary classification error problem through 'unstable interval' correction.

4.2 Reasons for model performance differences

The BiLSTM sub-model performs better than the SVM sub-model (F1 score 96.54% vs 92.05%), because BiLSTM can capture complex expressions of 'negative words + sentiment words' (such as 'not practical') through bidirectional temporal modeling, while SVM is a linear model and difficult to handle such non-linear text features; the performance of the single BiLSTM model is weaker than that of the hybrid model (F1 score 95.59% vs 96.54%), because the hybrid model uses the sentiment dictionary to filter the noise data in the nonboundary interval, reducing the training difficulty of BiLSTM.

4.3 Analysis of the significance of the 'unstable interval'

Within the interval [-0.195, 0.205], the classification accuracy of the sentiment dictionary is only 77.01%-79.24%, and after correction by SVM/BiLSTM, the accuracy increases to more than 92%, indicating that this interval correction can effectively make up for the defects of the dictionary method; when the interval exceeds [-1, 1], the accuracy of the sentiment dictionary reaches more than 90%, and the correction gain is less than 5%, so the 'unstable interval' correction is more suitable for scenarios where the sentiment score is close to the positive and negative boundaries.

5 Analysis of emotional causes based on DBSCAN algorithm

After dividing course reviews into "positive" and "negative" sentiments using sentiment polarity algorithms, evaluating course quality and perceived strengths and weaknesses still faces some challenges. The specific factors behind these "positive" and "negative" sentiment reviews, as well as the causes of users generating "positive" and "negative" sentiments, remain unknown. Furthermore, identifying the causes of each sentiment in course reviews manually is a time-consuming and laborious process. It is gratifying that this work happens to be well-suited for cluster analysis in data mining.

This paper employs the DBSCAN algorithm to cluster comments categorized as "positive" and "negative" sentiments separately. The DBSCAN algorithm overcomes the limitation of distance-based algorithms in only discovering "circular" clusters and is insensitive to noisy data. Moreover, it does not require a specific input of the number of categories K. Leveraging these advantages, the DBSCAN algorithm demonstrates high levels of accuracy, recall, and F1 score in text clustering in fields such as "history," "education," "reading," and "culture." Compared to K-means and Gaussian mixture models, DBSCAN algorithm exhibits certain advantages [16-17].

5.1 Causes of positive emotions

This article briefly summarizes and lists the main contributing words to the clustering text after DBSCAN algorithm clustering. The DBSCAN algorithm clustered 3119 ' positive' emotional course comments and obtained 29 clusters. In this study, the representative words (main contributing words) of the clusters are selected through the TF-IDF (Term Frequency-Inverse Document Frequency) method, which can effectively reflect the importance of words in specific clusters and avoid the interference of common meaningless words that may be caused by selecting only based on word frequency. In order to visually display the clustering situation of "positive" user course comments, this article uses a rose chart to display the main contributing words of the 29 clusters obtained after clustering, as shown in Figure 8. The frequency of words such as “well spoken”, “dry goods”, and “top-notch” is relatively high, indicating that NetEase Cloud course users are relatively satisfied with Microsoft’s three piece sets of courses and the teaching content of teachers, with a high level of teaching staff. This is directly related to the fact that most of the platform’s courses are national high-quality courses. At the same time, the words “highly recommended”, “learned a lot”, “very awesome” and so on can show users’ appreciation and gratitude for the course content. The words “cute teacher”, “lovely Mandarin” and so on can show that teachers’ personal charm can also attract many users, enhance users’ preference for the course, and enhance “positive” mood.

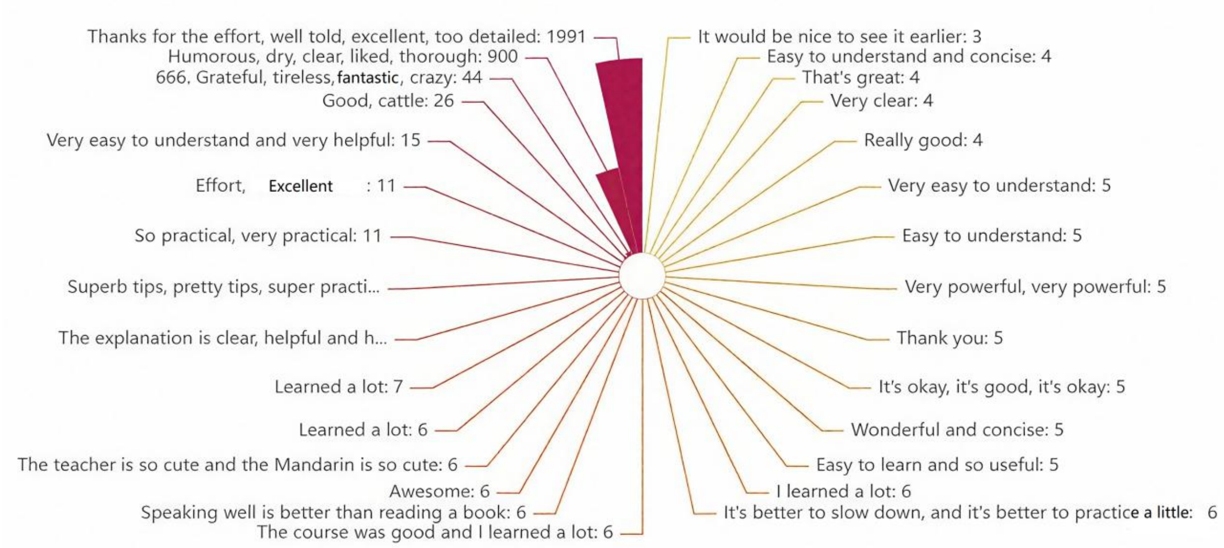

Figure 8: The main contributing words of the cluster of “positive” course comments

5.2 Identification of the causes of negative emotions

In order to more intuitively display the clustering situation of “negative” user course comments, this article uses a rose chart to display the main contributing words of the 26 clustering clusters obtained after clustering, as shown in Figure 9. The frequency of words such as “a lot of nonsense”, “verbose”, and “fast speaking” is relatively high, indicating that course users are not satisfied with the irrelevant content and speed of the teacher’s teaching process. Teachers should maintain a moderate speaking speed during the teaching process, while users can consider whether to accelerate or slow down the video viewing according to their own needs. At the same time, from the words’ cannot be played’ cannot be played out’, and’ cannot be downloaded’, it can be seen that there are many problems with the playback and download of videos, which makes users dissatisfied with the overall course, lowers their liking for the course, and leads to the generation of’ negative ’emotions among users. The platform should first verify whether it is related to its own system and improve its stability. If it is due to objective reasons such as copyright, the platform and the copyright owner should actively communicate, work together, and try to provide solutions to reduce the generation of “negative” emotions.

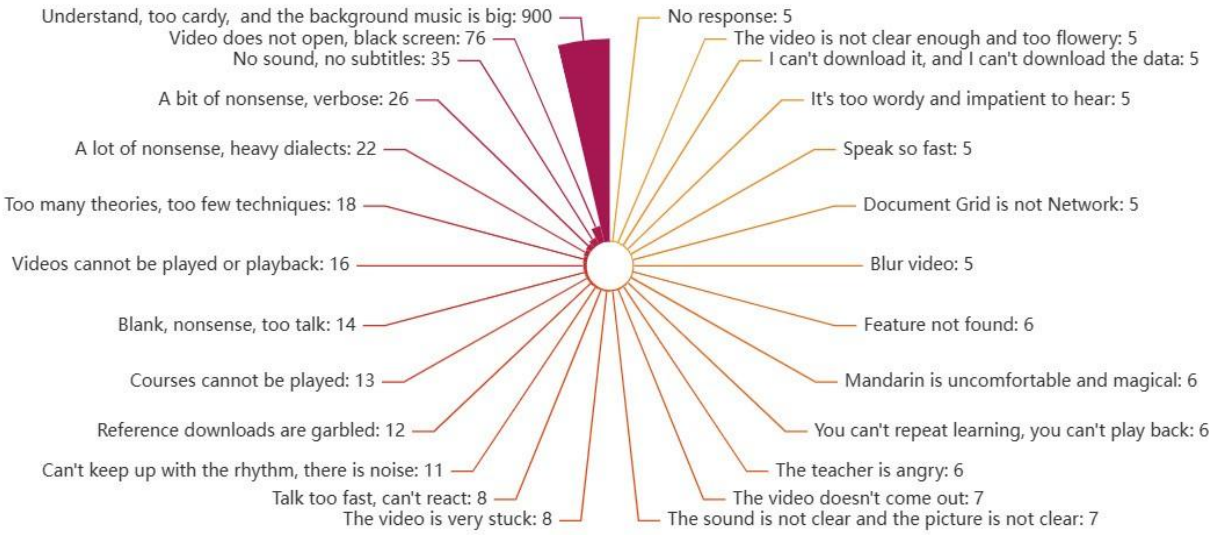

Figure 9: The main contributing words of the cluster of “negative” course comments

5.3 Countermeasures and suggestions based on emotional causes

5.3.1 The course should focus on practicality, and teachers should focus more on “dry goods” when teaching

In the emotional causes of “positive” and “negative” in course comments, practicality is the core demand of users for the course. Users expect to apply the learned course knowledge to real life, solve problems faced in life, and even many users come to watch course videos because they encounter difficulties in real life. Whether it can meet the practical requirements will determine the emotional attitude of users. At the same time, in the demand for course activities, the teaching methods, language and attitude of teachers, and different forms of course interaction are all the focus of user attention. They are very important for users to maintain a positive learning attitude, drive continuous learning, and strengthen their understanding of course content. At the same time, the sense of value identification and overall feeling after learning the course are also issues that user attach great importance to.

5.3.2 The evaluation indicators for online courses from a user perspective are rich and specific, resulting in new characteristics of online courses

According to the main contributing words of different emotional course reviews, interesting and novel course content can better trigger users’ “positive” emotions. Compared with traditional teaching evaluation systems, this guiding and contextualized teaching strategy, as well as a positive and agile teaching style and systematic comprehensiveness of teaching modes, reflect the characteristics of online self-directed learning for users in the educational information environment. Compared with existing traditional literature based on expert thinking, users’ requirements for teacher explanations not only include basic requirements such as “professionalism”, “accuracy”, and “richness”, but also have a higher level of requirement for “interest”, which is a supplement and improvement to the existing evaluation index system.

5.3.3 Netease cloud users have a high overall satisfaction with the course

The majority of user comment texts exhibit a positive emotional tendency, indicating a high level of satisfaction from users with course needs. This is directly related to the fact that most of the platform’s courses are national high-quality courses, and the platform should continue to adhere to the quality control of courses. At the same time, most users express their love for courses, and these courses can solve practical problems, not only meeting their own practical needs, but also learning more new knowledge. Many users have also expressed a willingness to recommend and strongly recommend, indicating that positive feedback is being generated between users and the platform, which will further benefit both parties.

5.3.4 The netease cloud platform needs to be optimized in terms of course management, platform functions, and services

Current users have negative evaluations of course playback, downloads, and course- ware materials. The NetEase cloud platform and course copyright owners need to further improve the management of courses and ensure timely provision of the latest courseware materials that match the course videos. Strengthen technical support for the platform, improve its stability, enable users to operate easily and quickly, and ensure that every learning process can be completed smoothly. Strengthen communication and training with course providers to ensure that teachers have a natural expression and moderate speaking speed during the course recording process. At the same time, the platform can also provide corresponding completion certificates for each user who has studied courses and passed course assessments, improving their learning enthusiasm. The platform should provide users with procedural guidance on applying for certificates and timely feedback on the progress status of applications to users who have submitted them.

6 Conclusion

This article focuses on course reviews with implicit and rich information as the entry point, exploring the classification of emotional polarity and emotional causes of online course reviews. The findings of our study, highlight the significance of integrating sentiment dictionary with machine learning models to address the limitations of traditional sentiment analysis approaches. The results demonstrate significant improvements in sentiment polarity classification, with the fusion models exhibiting enhanced accuracy and adaptability across diverse text types and domains. Furthermore, the application of DBSCAN algorithm for text clustering provides valuable insights into the underlying reasons for users' positive and negative emotions during the learning process, offering a novel approach for understanding course strengths and weaknesses.

Acknowledgment

This work was funded by the following projects: 1) 2020 Philosophy and Social Science Research Project of Education Department of Hubei Province, entitled "A Semantic Analysis Model of Course Review for Online Course Evaluation and Application" (Project Number: 20Y190); 2) Hubei Provincial Collaborative Innovation Center for Basic Education Information Technology Services (NO. OFHUE202308); 3) Hubei Provincial University Excellent Young and Middle-aged Science and Technology Innovation Team Program, entitled "Research on Key Technologies of Digital Management of Teacher Teams" (Project Number: T2024028).

References

[1] Z. Zhou, S. Wang, Y. Liu, et al., 2022. A survey on visual analysis of massive open online courses (MOOC) data. Journal of Computer-aided design and Graphics, 34: 830-840. https://doi.org/10.3724/SP.J.1089.2022.19536

[1] Ministry of Education of the People's Republic of China. Notice on Doing a Good Job in Online Teaching and Learning During the Epidemic Prevention and Control Period [EB/OL]. (2020-02-05). http://www.moe.gov.cn/jyb_xwfb/gzdt_gzdt/s5987/202002/t20200205_426499.html.

[2] C. Wuyan, C. Honglei, 2020. The implementation status and reflection on online teaching in "double first class" universities, Education Science, 36: 24-30.

[3] T. Sinha, J. Cassell, 2015. Connecting the dots: predicting student grade sequences from bursty MOOC interactions over time. In: Proceedings of the Second (2015) ACM Conference on Learning. Vancouver. 249-252. https://doi.org/10.1145/2724660.272866

[4] Wang, Q., & Zhao, Y. (2021). Research on sentiment classification and aspect mining for MOOC course review texts. Computer Engineering and Applications, 57(18), 123-130.

[5] Z. Jieni and L. Dan, 2018. A study on the impact of effective comments on online course design and development, China Education Informatization, 33-36.

[6] T. Shaik, X. Tao, C. Dann, et al., 2023. Sentiment analysis and opinion mining on educational data: A survey. Natural Language Processing Journal, 2: 100003. https://doi.org/10.1016/j.nlp.2022.10000

[7] T. Soffer, T. Kahan, E. Livne, 2016. E-assessment of online academic courses via students' activities and perceptions. Studies in Educational Evaluation, S0191491X16300190. https://doi.org/10.1016/j.stueduc.2016.10.001

[8] K. Sailunaz, R. Alhajj, 2019. Emotion and sentiment analysis from twitter text. Journal of Computational Science, 36: 101003. https://doi.org/10.1016/j.jocs.2019.05.009

[9] K. Vasanth, P. Srideviponmalar, V. Shete, et al., 2022. Dynamic fusion of text, video and audio models for sentiment analysis. Procedia Computer Science, 215: 211-219. https://doi.org/10.1016/j.procs.2022.12.024

[10] M. Xu, B. Zeng, H. Yang, et al., 2022.Combining dynamic local con text focus and dependency cluster attention for aspect-level sentiment classification. Neurocomputing, 478: 49-69. https://doi.org/10.1016/j.neucom.2021.12.084

[11] J. Sangeetha, U. Kumaran, 2023. Sentient analysis of Amazon user reviews using a hybrid approach. Measurement: Sensors, 27: 100790. https://doi.org/10.1016/j.measen.2023.100790

[12] M. Misuraca, G. Scepi, M. Spano, 2021. Using opinion mining as an educational analytic: An integrated strategy for the analysis of students' feedback. Studies in Educational Evaluation, 68: 100979. https://doi.org/10.1016/j.stueduc.2021.100979

[13] H. Aka-Uymaz, S. Kumova, 2022. Vector based sentiment and emotion analysis from text: A survey. Engineering Applications of Artificial Intelligence, 113: 104922. https://doi.org/10.1016/j.engappai.2022.104922

[14] J. R. Trillo, E. Herrera-Viedma, J. A. Morente-Molinera, et al., 2023. A large-scale group decision making system based on sentiment analysis culster. Information Fusion, 91: 633-643. https://doi.org/10.1016/j.inffus.2022.11.009

[15] CNKI. Sentiment Analysis Word Set (Beta Version) [EB/OL]. https://www.cnki.net/other/sentiment_analysis.html, 2020.

[16] A. Latifi-Pakdehi, N. Daneshpour, 2021. DBHC: A DBSCAN-based hierarchical clustering algorithm. Data & Knowledge Engineering, 135: 101922. https://doi.org/10.1016/j.datak.2021.101922

[17] S. Chowdhurya, N. Heliamb, R. C. D. Amorim, 2023. Feature weighting in DBSCAN using reverse nearest neighbors. Pattern Recognition, 137: 1-15. https://doi.org/10.1016/j.patcog.2023.109314