



International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
<https://www.inderscience.com/ijict>

Sentiment classification model for student teaching evaluation based on deep learning technology

Cen Li

DOI: [10.1504/IJICT.2024.10063748](https://doi.org/10.1504/IJICT.2024.10063748)

Article History:

Received:	15 September 2023
Last revised:	27 October 2023
Accepted:	13 November 2023
Published online:	13 June 2024

Sentiment classification model for student teaching evaluation based on deep learning technology

Cen Li

ChangChun Normal University,
ChangChun, 130000, China
Email: m18088682231@163.com

Abstract: In order to improve the effect of emotion classification of students' teaching evaluation, this paper studies the emotion classification of students' teaching evaluation combined with deep learning technology. This paper presents a sentiment analysis method based on the sentiment dictionary of teaching evaluation as a cold start solution when the marking data is insufficient. This method builds a collocation dictionary for teaching evaluation, solves the polar context-related problems of emotion words, introduces parallel relation and similarity calculation, and designs multi-level polarity recognition strategies for emotion words, so as to improve the performance of emotion analysis. The research shows that the classification model of student teaching evaluation based on deep learning proposed in this paper has a good effect on emotion classification, and has a certain effect on promoting interaction between teachers and students.

Keywords: student; teaching evaluation sentiment; classification model; deep learning.

Reference to this paper should be made as follows: Li, C. (2024) 'Sentiment classification model for student teaching evaluation based on deep learning technology', *Int. J. Information and Communication Technology*, Vol. 24, No. 7, pp.1–16.

Biographical notes: Cen Li graduated from Changchun Normal University with a Master's degree in 2017. He is currently working in Changchun Normal University as a Postgraduate Secretary. His research direction is higher education management. He has presided over one scientific research project of the Department of Education and participated in four teaching and research projects at the provincial level.

1 Introduction

Sentiment polarity classification refers to analysing and judging the positive and negative sentiments of texts, that is, positive and negative meanings. There are mainly two methods based on sentiment dictionary and based on machine learning. Generally, due to the richness of semantic expressions, machine learning is more accurate than sentiment dictionaries and has more flexibility. However, machine learning is extremely dependent on corpus, and it is doomed to fail to use classifiers trained in other fields to classify texts for teaching evaluation (Williamson and Eynon, 2020). Therefore, it is often necessary to prepare a large number of artificially labelled corpora for the required

field to retrain the classifier. However, dictionary matching is applicable to a wider range of corpora, but there is currently no universal dictionary that covers all fields of terminology. In general, the method based on machine learning do not consider the role of sentiment lexicon, while the method based on sentiment dictionary ignore the relationship between words (Taylor and Pagliari, 2018). In Kennedy (2018), the author tried to apply the method of combining sentiment dictionary and machine learning to hotel reviews and achieved good results. In Liang et al. (2018), the authors also integrated linguistic rules such as sentiment lexicon, degree adverbs, and negative words with the existing sentence-level long short-term memory (LSTM) sentiment classification model. Moreover, he considered the sentiment distribution between adjacent words and found that the bi-directional long short-term memory (Bi-LSTM) model based on linguistic rules has the best sentiment classification effect. Therefore, according to the characteristics of students' teaching evaluation texts, this paper selects a machine learning classifier with better classification effect combined with a sentiment dictionary to classify the polarity of teaching evaluation texts on the basis of separately processing suggested sentences. At the same time, this paper verifies the effectiveness and feasibility of the method.

The research on sentiment classification phrases is summarised as follows: there are two main ways to extract opinion words at this stage: one is direct extraction based on part of speech, including phrase pattern, sentiment dictionary construction, etc. (Etter et al., 2018); the other is indirect extraction based on opinion word comment object recognition, including extraction method of existing rules, extraction model based on syntactic analysis (Trottier, 2018). The two methods have their own advantages and disadvantages: direct extraction is not easy to miss emotional words, and the accuracy is relatively high, but it is not easy to obtain the comment objects of opinion words and the relationship between them; while indirect extraction can complete the 'comment object – opinion words' in one step. The extraction of opinion pairs, but it is easy to miss opinion words when the comment object is implied, and because of the uncertainty of the type of comment object, the construction of the comment object lexicon is much more complicated than the sentiment dictionary (Assumpção et al., 2018). Because students' evaluation of teachers is multi-faceted, the construction of the thesaurus of comment objects is complex and difficult to be comprehensive. At the same time, the unstructured text of students' evaluation of teaching leads to the more common implicit comment objects, which is prone to the phenomenon of missing opinion words. Therefore, this paper uses part-of-speech-based opinion extraction method (Chatterjee et al., 2015).

The summary of human task research on emotional analysis of teaching evaluation is as follows: according to the different analysis granularity, text sentiment analysis tasks can be divided into: document level, sentence level, word level and attribute level. Since teaching evaluation is generally not very long, the task is divided into fine-grained sentiment analysis and processed as a fine-grained attribute-level sentiment analysis task (Grossi et al., 2018). Attribute-level sentiment analysis can be understood as the process of extracting the evaluation object (attribute) in the text and determining the sentiment tendency for the attribute. Pointed out that the process of extraction and identification of multiple groups is also developing (Pangrazio and Sefton-Green, 2020). In earlier research work, attribute-level sentiment analysis is also known as feature-based opinion mining (feature-based opinion mining). Later work further represents the opinion as a quadruple (g, s, h, t) , where g represents the comment object (target), s represents the sentiment (sentiment), h represents the opinion holder (opinion holder), and t represents

the time. Comment objects usually contain entities and their attributes, so the above four-tuple can be transformed into a five-tuple (e, a, s, h, t) . Attribute-level sentiment analysis is correspondingly defined as the extraction and identification process of the above-mentioned tuples (quadruples or quintuples). Since the teaching evaluation corpus is short, it is treated as an attribute-level corpus, so it is a fine-grained division task (Walsh, 2020).

The research on emotional mining in classroom teaching evaluation texts is summarised as follows: Wiggins and Wilbanks (2019) proposed a sentiment analysis method of short Chinese texts based on teaching evaluation for the purpose of improving the quality of teaching and improving the effectiveness of teaching evaluation from the perspective of supervision experts. Aiming at the characteristics of Chinese short text teaching evaluation data being highly professional and sparse in features, a professional dictionary for teaching evaluation was established, and word vectors were trained with the Word2Vec language model to reduce the dimension of the attributes of the teaching evaluation data. The support vector machine (SVM) algorithm with three cores of the base kernel performs sentiment classification on the Chinese short text data in the teaching evaluation to judge the sentimental tendency of the evaluation. Budiharto and Meiliana (2018) analyses the sentiment tendency of classroom teaching evaluation text information, the purpose is to let teachers realise their own strengths and weaknesses objectively, and at the same time can improve teachers' teaching methods and methods. Berendt et al. (2020) uses a LSTM network method based on multi-prototype word vectors to analyse the sentiment tendency of classroom teaching evaluation texts. The teaching evaluation corpus comes from the educational administration system, and 150,000 teaching evaluations are manually annotated. The accuracy rates of the SVM system and the combined system of LSTM and Word2Vec are also compared. It is a further exploration of the existing teaching evaluation system based on machine learning. Ferguson and Caplan (2010) and others used a comparative analysis method combining data mining and emotion mining to conduct a comparative study on the quality of postgraduate teaching evaluation, showing that the emotional mining data and the teaching evaluation scale data have both correlations and differences. By comparison, we can better understand the nature of the quality of the teaching evaluation questionnaire.

The research on intelligent learning methods for sentiment mining in text is summarised as follows: the step-by-step principle refers to designing test questions based on the differences in students' cognitive and emotional development levels, so that the test results can present a certain gradient, and the levels and differences can be distinguished according to the students' answers. The purpose of this test is to judge the effect of study sheet use according to Bloom's taxonomy of educational goals. It is necessary to determine the cognitive level or emotional level of each student based on the answers of each student. Finally, the overall test results of the students reflect the use of study sheets in the performance of each dimension and level of Bloom's taxonomy of educational goals (Ford et al., 2020). Therefore, the test must ensure the step-by-step nature of the test questions through the following ways: first, design a relatively open answer form and way, so that students' answers can express their thinking process and emotional attitude as clearly as possible, and avoid the answer form restricting students' thinking and emotional performance. Second, when testing whether students' cognitive process has reached the level of inference, students are required to clarify their own thinking process and methods. Students' inference results cannot be used as the only

basis for discrimination, so as to avoid students' 'guessing' behaviour from affecting the accuracy of the test (Wagler and Hanus, 2018).

Machine learning or deep learning methods often rely on a large number of training samples. If there are too few training samples or the distribution of data in different categories is uneven, it may lead to overfitting problems in model training. Therefore, in cases where the sample size is small or the category distribution is uneven, data augmentation methods can be used to improve this problem.

In the field of student evaluation of teaching, this article aims to effectively integrate and analyse the fine-grained emotional analysis results (i.e., evaluation dimensions and corresponding emotional tendencies) of the evaluation text set, which focuses on the feedback of university student evaluation results and comprehensive evaluation of teaching quality, in response to the fact that teachers and their courses often have multiple evaluation texts (i.e., evaluation text sets). This improves the drawbacks of traditional research.

This paper combines the deep learning technology to carry out the research on the sentiment classification of students' teaching evaluation, and constructs the sentiment classification model of students' teaching evaluation, which can effectively improve the teaching effect of colleges and universities.

2 Sentiment classification model based on deep learning

Machine learning or deep learning methods often rely on a large number of training samples. If there are too few training samples or the distribution of data in different categories is uneven, it may lead to overfitting problems in model training. Therefore, in cases where the sample size is small or the category distribution is uneven, data augmentation methods can be used to improve this problem. This article applies deep learning to the classification and evaluation of students' teaching emotions, and improves the teaching effectiveness of intelligent teaching models through algorithmic models

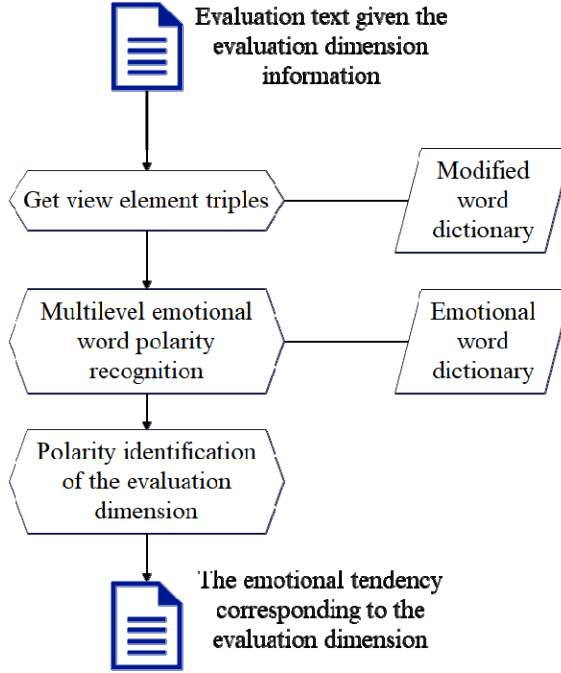
2.1 Sentiment analysis method for teaching evaluation dimension

This paper will use the dictionary-based method and the deep learning-based method to solve the sentiment analysis task oriented to the teaching evaluation dimension.

This paper proposes a teaching evaluation dimension sentiment analysis method based on the teaching evaluation sentiment dictionary. The overall flow chart of Figure 1 is as follows (Makransky and Lilleholt, 2018).

Sentiment words have obvious positive (positive, +1) and derogatory (negative, -1) polar emotional colours. Moreover, sentiment words can usually be divided into two categories: sentiment words that are not polar contexts and sentiment words that are related to polar contexts. Polarity is context-independent emotional words, as the name suggests, the polarity of such emotional words does not change with different contexts (that is, contexts). Generally, we can obtain the polarity of such sentiment words by querying public sentiment dictionary resources (such as WordNet, HowNet, etc.). Therefore, in the research of this paper, a basic dictionary will be constructed based on the public sentiment dictionary resources for the sentiment polarity recognition of students' teaching evaluation texts (Park et al., 2018).

Figure 1 Sentiment analysis process based on teaching evaluation sentiment dictionary (see online version for colours)



Identifying the polarity P_{sent} of sentiment words is the key task of sentiment analysis based on the teaching evaluation sentiment dictionary in this paper. In this paper, parallel sentiment words and similarity calculation are introduced to identify the sentiment polarity of new words and further improve the polarity recognition performance.

2.1.1 Polarity recognition based on juxtaposed emotional words

There are side-by-side dependencies COO in students' teaching evaluation texts, and students sometimes express emotional words side-by-side. At the same time, we think that the emotional tendencies of the juxtaposed emotional words are the same. Therefore, we can use juxtaposed sentiment words to identify the sentiment polarity of new words. The specific method is as follows.

If S is a candidate set of juxtaposed sentiment words w whose sentiment polarity has been identified, we can use the principle of majority polarity to judge the sentiment polarity P_{new} of the new word according to the polarity P_{sent}^i of the sentiment word $sent_i$ in S . The calculation formula is as formula (1). If $score \geq 0$, then the positive polarity co-words are in the majority, so $P_{new} = +1$. Conversely, if $score < 0$, the negative polarity co-words are in the majority, so $P_{new} = -1$.

$$score = \sum_{i \in S} P_{sent_i}. \quad (1)$$

2.1.2 Polarity recognition based on similarity calculation

To further improve the performance of sentiment analysis, polarity recognition based on similarity calculation is supplemented here. For sentiment words whose polarity has not been identified, the cosine similarity value represented by word vector between them and sentiment words with identified polarity is calculated. Moreover, the polarity P_{\max} of the emotional word with the highest similarity value is selected as the polarity P_{new} of the emotional word of unknown polarity, that is, $P_{new} = P_{\max}$ (i Badia et al., 2018).

2.1.3 Multi-level sentiment word polarity recognition strategy

The following introduces the specific content of the multi-level emotional word polarity recognition strategy proposed in this paper:

- Step 1 First, the algorithm uses the teaching evaluation collocation dictionary to perform the first round of matching. The algorithm identifies triples containing positive collocation Pos_Pair and derogatory collocation Neg_Pair, and marks the polarity of their corresponding sentiment words.
- Step 2 The algorithm then uses the basic dictionary to perform a second round of matching on the remaining unrecognised polar emotional words. If the sentiment word sent belongs to Pos_Dic, then $P_{sent} = +1$. If the sentiment word sent belongs to Neg_Dic, then $P_{sent} = -1$.
- Step 3 For the emotional words whose polarity is not identified in Steps 1 and 2, the algorithm uses the polarity recognition method based on parallel emotional words introduced above to determine the emotional polarity of the word.
- Step 4 For sentiment words whose polarity is not identified after the above strategies, the algorithm uses the polarity identification method based on similarity calculation introduced above to determine the sentiment polarity of the word.

The multi-level sentiment word polarity recognition strategy in the previous section can effectively identify the polarity P_{sent} of sentiment words in triples. This section will introduce how to implement the teaching evaluation dimension sentiment analysis of teaching evaluation texts through the sentiment polarity P_{Unit} of the triplet, that is, to obtain the sentimental tendency of praise and depreciation corresponding to each teaching evaluation dimension label in the teaching evaluation text (Quesnel et al., 2018).

1 Three-tuple sentiment polarity calculation

The modifier in the triplet is generally a negative word, which has a turning effect on the sentimental tendency of the opinion. Therefore, the sentiment tendency P_{Unit} of the triple is not only related to the polarity of sentiment words, but also to the modifiers. The formula for calculating P_{Unit} is as follows:

$$P_{Unit} = m \times P_{sent} \quad (2)$$

2 Sentiment analysis of teaching evaluation texts

Each teaching evaluation text contains multiple opinion element triples. After obtaining the polarity of each triplet, it is necessary to combine and deduplicate the triples with the same teaching evaluation dimension label and their polarity, so as to

obtain the sentiment tendency corresponding to each teaching evaluation dimension label in the teaching evaluation text (Pietarinen et al., 2019).

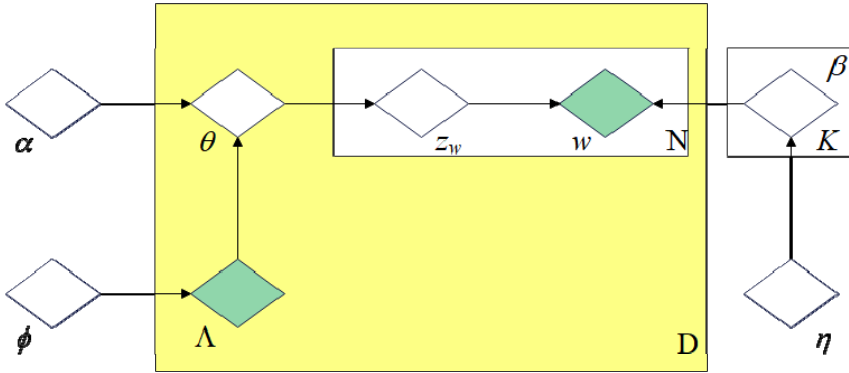
2.2 Sentiment analysis based on DEABi-LSTM

Deep learning has been widely used in the field of sentiment analysis because of its powerful representation ability. This section will study the use of deep learning methods to solve the problem of sentiment analysis under the given teaching evaluation dimension. This paper proposes a teaching evaluation dimension sentiment analysis model based on attention-based Bi-LSTM with dimension embedding (DEABi-LSTM) (Ding et al., 2018).

In the DEABi-LSTM model proposed in this paper, the labelled LDA model is used to generate the teaching evaluation dimension probability distribution. It is used for the vector representation of a given teaching evaluation dimension and the dimension sentiment semantics embedding of the teaching evaluation dimension, which are described in detail next.

The labelled LDA model is a supervised generation model, which is mainly used for multi-label document modelling. Figure 2 shows the structure of labelled LDA.

Figure 2 Structure diagram of labelled LDA (see online version for colours)



Different from the traditional LDA model, labelled LDA can make good use of label information. The labelled LDA model generation process is as follows:

- 1 The algorithm first uses the Bernoulli experiment to generate a set of topic tags $\Lambda^{(d)} = \{l_1, l_2, \dots, l_K\}$ for each document, where K is the number of different tags in the document set, $l_k \in \{0, 1\}$.
- 2 Secondly, the algorithm defines the label vector of document d as $\lambda^{(d)} = \{k \mid \Lambda_k^{(d)} = 1\}$, and obtains the label mapping matrix $L^{(d)}$ of each document.
- 3 Finally, the algorithm uses $L^{(d)}$ to reduce the dimension of the Dirichlet topic prior distribution $\alpha = (\alpha_1, \dots, \alpha_k)^T$ to obtain a vector $\alpha^{(d)}$ [formula (3)], which realises the generation of document topic distribution under the restriction of its own label (Yuan and Ip, 2018).

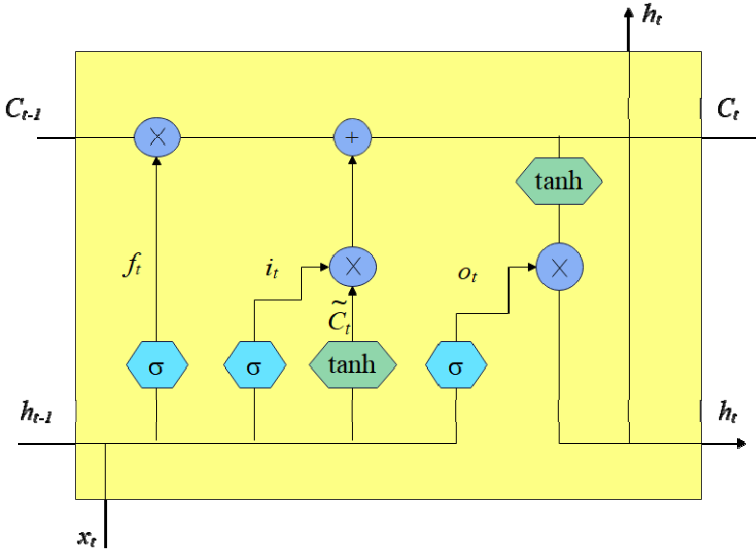
$$\alpha^{(d)} = L^{(d)} * \alpha = (\alpha_{\lambda 1}, \dots, \alpha_{\lambda d}) \quad (3)$$

d_i is the vector representation of the i^{th} teaching evaluation dimension, then the probability mean value d of the teaching evaluation dimension corresponding to the teaching evaluation text can be obtained, as shown in formula (4).

$$d = \frac{1}{N} \sum_{i=1}^N d_i \quad (4)$$

LSTM is an improved model for recurrent neural network (RNN). Figure 3 shows the neuron structure of LSTM, which solves the limitation of long-distance dependence in RNN through its clever structure. Because of its unique advantages, LSTM has attracted extensive attention from scholars, and various LSTM variant models have gradually emerged, which further optimises the performance of the model in applications in various fields.

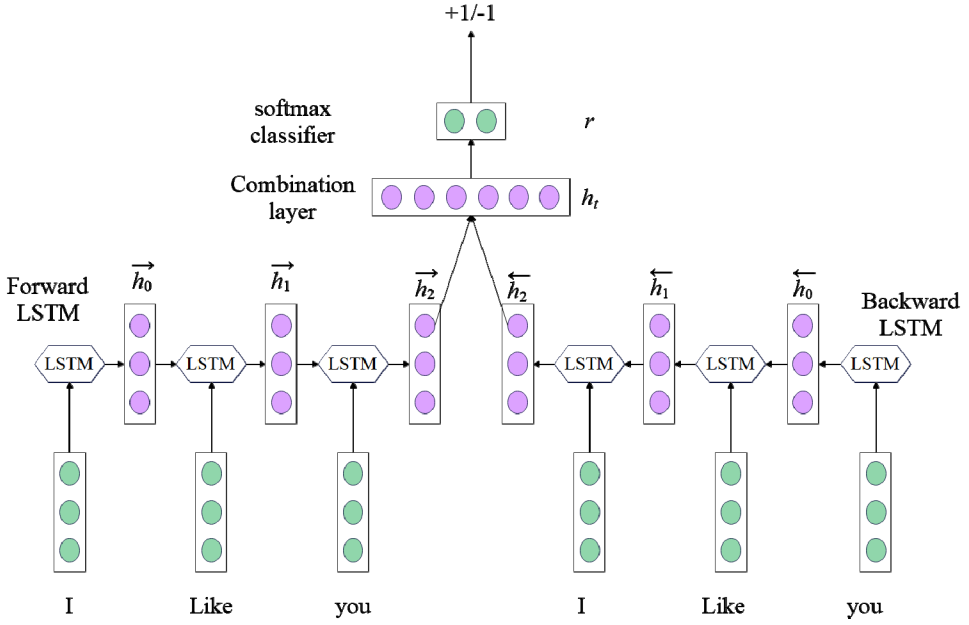
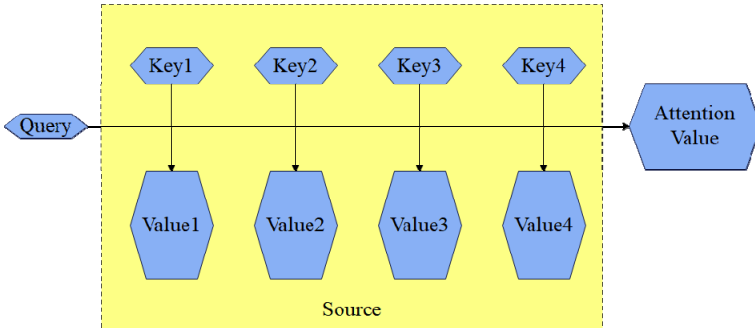
Figure 3 LSTM neuron structure (see online version for colours)



Bidirectional long short-term memory (Bi-LSTM) network is a variant model of LSTM. Figure 4 shows a schematic diagram of the structure of Bi-LSTM when dealing with sequence classification tasks. In the fine-grained sentiment analysis task, more perfect context information modelling and deep semantic feature learning are very important, so this paper adopts the Bi-LSTM model for sentiment analysis.

The attention mechanism is a way of simulating the human visual selection process (attention) for resource allocation.

At present, the attention mechanism has been applied to many types of specific models. The more commonly used attention framework is shown in Figure 4–Figure 5.

Figure 4 Example of classification using Bi-LSTM (see online version for colours)**Figure 5** Attention frame (see online version for colours)

In Figure 5, source consists of a set of elements, query is a target vector, and the attention mechanism can be understood as a mechanism that extracts a specific vector from a given vector set source according to the target vector query for weighted summation. The attention function used to generate attention value is introduced here, and the generation formula of attention value is as formula (5) (Wagler and Hanus, 2018).

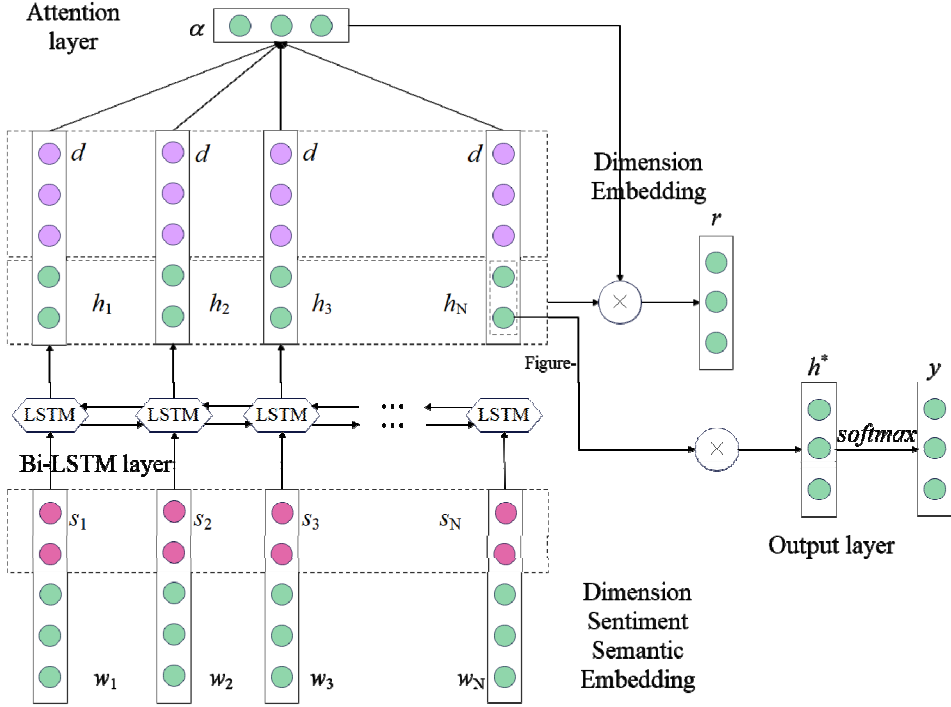
$$AttentionValue = (Query, Source) = \sum_{i=1}^L Similarity(Query, Key_i) * Value_i \quad (5)$$

In the sentiment analysis research on the teaching evaluation dimension in this paper, different teaching evaluation dimensions correspond to different emotional tendencies. Therefore, it is necessary to give more weights to semantic features related to specific teaching evaluation dimensions when building sentiment analysis models. At this time,

introducing the attention mechanism into the model will help improve the model performance.

Figure 6 shows the structure of the DEABi-LSTM model proposed in this paper, and this section will introduce the model construction process in detail.

Figure 6 Structure of the DEABi-LSTM model (see online version for colours)



2.2.1 Input layer

The input to the DEABi-LSTM model consists of two parts. For the teaching evaluation text containing N words, it is first converted into word vector form $\{w_1, \dots, w_N\}$, namely, WordEmbedding, by the pre-trained word2vec model. Secondly, the LabeledLDA model is used to obtain the teaching evaluation dimension sentiment semantic word embedding $\{s_1, \dots, s_N\}$. Finally, the fusion of $\{w_1, \dots, w_N\}$ and $\{s_1, \dots, s_N\}$ is used as the input of the model, which can make full use of the dimension information of teaching evaluation and make the semantics of the input vector of the model more perfect.

2.2.2 Bi-LSTM layer

Bi-LSTM can capture the contextual semantic features of teaching evaluation dimension and enhance long memory ability. \vec{h}_i is the output of the forward LSTM, \overleftarrow{h}_i is the output of the backward LSTM, then the combination of \vec{h}_i and \overleftarrow{h}_i is the output vector h_i of the hidden layer. H is the feature matrix $\{h_1, h_2, \dots, h_N\}$ output by the hidden layer, $H \in R^{2m \times N}$, m is the size of the hidden layer.

2.2.3 Attention layer

In the DEABi-LSTM model, the attention layer makes full use of the output H of the Bi-LSTM hidden layer and the given teaching evaluation dimension probability vector DimensionEmbedding. Moreover, it focuses on key parts of the teaching evaluation text in different teaching evaluation dimensions.

First, it calculates the attention weight α , as in formula (6) and formula (7), where d is the probability distribution of the teaching evaluation dimension.

$$A = \tanh \left(\begin{bmatrix} W_h H \\ W_d d \otimes e_N \end{bmatrix} \right) \quad (6)$$

$$\alpha = \text{softmax}(w^T A) \quad (7)$$

Next, it obtains the text representation r with attention weight under the given teaching evaluation dimension information, as shown in formula (9), where $r \in R^{2m}$.

$$r = H\alpha^T \quad (8)$$

The feature representation h^* of the teaching evaluation text finally obtained after passing through the attention layer, as shown in formula (9), where $h^* \in R^{2m}$.

$$h^* = \tanh(W_r r + W_x h_N). \quad (9)$$

2.2.4 Output layer

The model can map the multi-dimensional teaching evaluation text feature representation h^* to the sentiment tendency category (positive, negative) through the softmax fully connected layer. Finally, the output is the label probability distribution y of the sentiment tendency corresponding to the teaching evaluation text under the given teaching evaluation dimension, as shown in formula (10), where $y \in R^3$.

$$y = \text{softmax}(W_s h^* + b_s) \quad (10)$$

In formula (6) to formula (10) in this section, W_h , W_d , W_r , W_x and W_s are weight matrices, and b_s is bias terms.

For the training of the DEABi-LSTM model, this paper adopts the sigmoid cross-entropy logarithmic loss function as the objective function to minimise the gap between the predicted results and the actual results, as shown in formula (11).

$$J(\theta) = -\frac{1}{2} \sum_{i=0}^2 t_i \log(y_i) + \lambda \|\theta\|_F^2 \quad (11)$$

In formula (11), t_i represents the model's prediction result of sentiment tendency category for the i^{th} teaching evaluation dimension label, and y_i represents the actual result. The L2 regularisation penalty term is added to the formula, λ represents the L2 regularisation coefficient, and θ represents the parameters in the model.

In this paper, Adam algorithm is used as the optimiser, and dropout technology is used in the training process to prevent overfitting. The model initialisation excitation function adopts the tanh function. Finally, the model takes the emotional tendency

category with the highest probability of each given teaching evaluation dimension as the prediction result.

In this paper, the emotional elements of the teaching evaluation dimension (subject words and emotional words), the teaching evaluation dimension and the emotional tendency in the teaching evaluation text are manually annotated. At the same time, in order to ensure the reliability of the annotated data, this paper invites two annotators to annotate the data respectively, and uses the kappa coefficient to test the consistency of the two annotations. The formula for the kappa coefficient test is shown in formula (12).

$$Kappa = \frac{p_a - p_e}{1 - p_e} \quad (12)$$

In formula (12), p_a refers to the agreement rate actually observed by the examiner, and p_e refers to the expected agreement rate in theory.

The confusion matrix shows basic quantitative statistics. In order to better measure the pros and cons of the model, three evaluation indicators, precision, recall, and F1 value, are derived on this basis. The calculation formula is as follows:

$$precision = \frac{TP}{TP + FP} \quad (13)$$

$$recall = \frac{TP}{TP + FN} \quad (14)$$

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \quad (15)$$

The precision rate reflects the proportion of all instances where the model predicted value is 1 correctly. The recall rate reflects the proportion of all instances where the model prediction value is 1 correctly. The F1 value combines the results of precision and recall.

3 Sentiment classification model for student teaching evaluation based on deep learning

When the text contains only emotional words with a certain tendency, the method of emotional dictionary matching is used, and the method of machine learning is used in other cases. This paper combines deep learning algorithms to build a sentiment classification model for student teaching evaluation based on deep learning, as shown in Figure 7.

The sentiment classification framework is shown in Figure 8.

The experimental environment for this article is Win1064bit, and the CPU model is Intel® Core (TM) i56500, with a RAM size of 16 GB. The segmentation tool used in this article is Jieba segmentation, and the segmentation mode is precise. The numerical calculation library used in the classification model is Numpy1.17.4, and the deep learning framework is Tensorflow. For other auxiliary calculations such as recall rate and F1 score, scikit learn is used. The development library used for training Word2vec word vectors in this article is gensim.

Figure 7 Sentiment classification model for student teaching evaluation based on deep learning (see online version for colours)

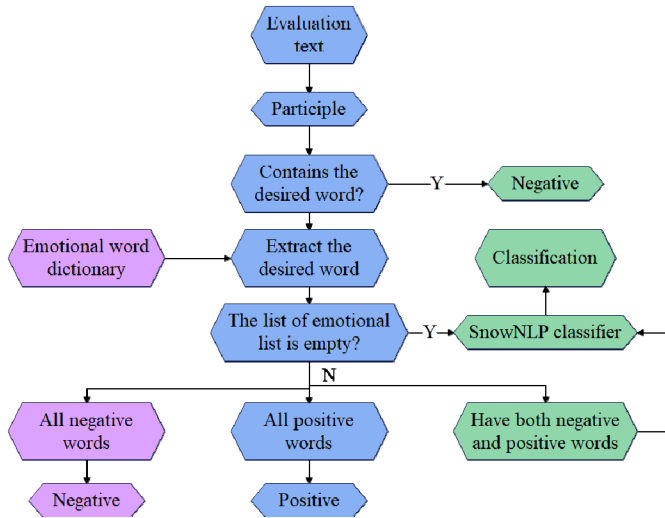
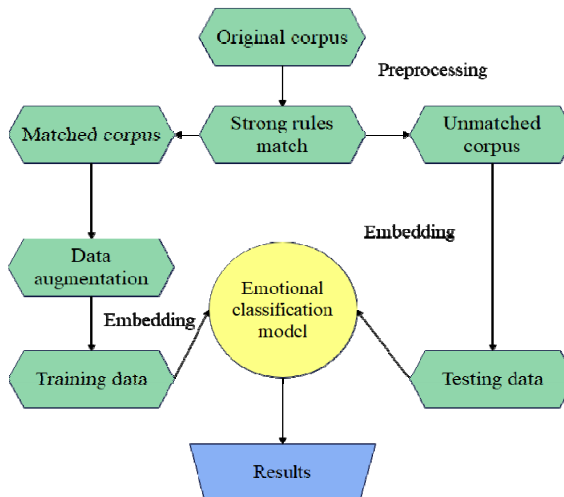


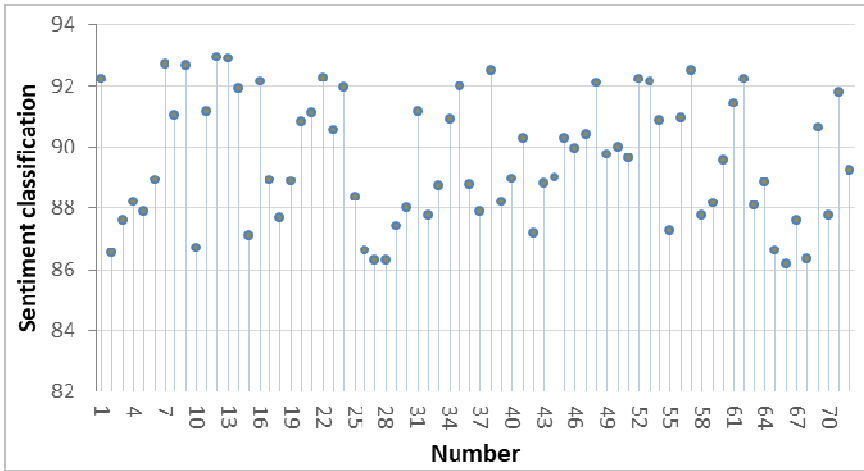
Figure 8 Sentiment classification framework (see online version for colours)



After the above model is constructed, the sentiment classification model for student teaching evaluation based on deep learning proposed in this paper is tested. Moreover, this paper calculates the accuracy of the model for sentiment classification of students' teaching evaluation, and finally obtains the results shown in Figure 9 through multiple sets of experiments.

It can be seen from the above research that the sentiment classification model for student teaching evaluation based on deep learning proposed in this paper has a good effect on sentiment classification, and has a certain effect on promoting the interaction between teachers and students.

Figure 9 Sentiment effect test of sentiment classification model for student teaching evaluation based on deep learning (see online version for colours)



A large number of experts and scholars have conducted analysis and research on feedback on teaching and course comments, and have achieved many research results in teaching evaluation. This study focuses on open teaching comments. Generally speaking, the purpose of feedback comments is to evaluate teachers' teaching level and improve educational quality. In the era of the internet, comments have a new definition under online learning methods, helping students quickly locate the desired target services. Specifically, the newly designed model plays an important role in improving the teaching level of individual teachers, especially in effectively modifying teaching strategies, directly meeting the specific needs of students, and more effectively managing the classroom. The evaluation of closed survey questionnaires currently used by offline domestic schools. If we use an emotion phrase pattern matching model and conduct teaching evaluation through open answer emotion analysis, it can analyse the hidden student emotions in the comments, which is beneficial for its application in the field of education. Emotional phrase pattern matching can extract language patterns from text case studies, which is also a challenging issue in sentiment analysis case studies.

In the future, it is necessary to continuously improve the intelligent algorithms for students to learn emotional analysis, and the improvement directions mainly include:

- The development of custom pre-trained embeddings on education-specific text corpora.
- Leveraging ensemble methods to combine the strengths of various deep learning architectures.
- Incorporating attention mechanisms to highlight key phrases in evaluation texts.
- Exploring transfer learning techniques to adapt pre-trained language models to the education domain.

4 Conclusions

In order to improve the reliability of the teaching evaluation score so that it can better serve the formulation of strong rules, the original teaching evaluation score is modified to some extent. First of all, the teachers in the same teaching class show basically the same characteristics to each student. Second, each student has a self-assessment scale that is stable across all the courses the student takes in the same semester. We can calculate the difference between the original teaching evaluation score and the average of the two in units of teaching class and student self, respectively, and then use this difference to replace the original teaching evaluation score. This paper combines the deep learning technology to carry out the research on the sentiment classification of students' teaching evaluation, and constructs the emotion classification model of students' teaching evaluation. The research shows that the sentiment classification model for student teaching evaluation based on deep learning proposed in this paper has a good effect on sentiment classification, and has a certain effect on promoting the interaction between teachers and students.

For the topic clustering section, the selection of clustering methods and initial sample points has a significant impact on the clustering effect. However, due to space issues, this article did not delve too much into how to select appropriate clustering algorithms and initial cluster centres, which is also worth considering.

References

- Assumpção, T.H., Popescu, I., Jonoski, A. and Solomatine, D.P. (2018) 'Citizen observations contributing to flood modelling: opportunities and challenges', *Hydrology and Earth System Sciences*, Vol. 22, No. 2, pp.1473–1489.
- Berendt, B., Littlejohn, A. and Blakemore, M. (2020) 'AI in education: learner choice and fundamental rights', *Learning, Media and Technology*, Vol. 45, No. 3, pp.312–324.
- Budiharto, W. and Meiliana, M. (2018) 'Prediction and analysis of Indonesia Presidential Election from Twitter using sentiment analysis', *Journal of Big Data*, Vol. 5, No. 1, pp.1–10.
- Chatterjee, S., Kar, A.K. and Mustafa, S.Z. (2021) 'Securing IoT devices in smart cities of India: from ethical and enterprise information system management perspective', *Enterprise Information Systems*, Vol. 15, No. 4, pp.585–615.
- Ding, N., Zhou, W. and Fung, A.Y. (2018) 'Emotional effect of cinematic VR compared with traditional 2D film', *Telematics and Informatics*, Vol. 35, No. 6, pp.1572–1579.
- Etter, M., Colleoni, E., Illia, L., Meggiorin, K. and D'Eugenio, A. (2018) 'Measuring organizational legitimacy in social media: assessing citizens' judgments with sentiment analysis', *Business & Society*, Vol. 57, No. 1, pp.60–97.
- Ferguson, K. and Caplan, A. (2021) 'Love thy neighbour? Allocating vaccines in a world of competing obligations', *Journal of Medical Ethics*, Vol. 47, No. 12, p.e20.
- Ford, E., Oswald, M., Hassan, L., Bozentko, K., Nenadic, G. and Cassell, J. (2020) 'Should free-text data in electronic medical records be shared for research? A citizens' jury study in the UK', *Journal of Medical Ethics*, Vol. 46, No. 6, pp.367–377.
- Grossi, V., Rapisarda, B., Giannotti, F. and Pedreschi, D. (2018) 'Data science at SoBigData: the European research infrastructure for social mining and big data analytics', *International Journal of Data Science and Analytics*, Vol. 6, No. 3, pp.205–216.

- i Badia, S.B., Quintero, L.V., Cameirao, M.S., Chirico, A., Triberti, S., Cipresso, P. and Gaggioli, A. (2018) 'Toward emotionally adaptive virtual reality for mental health applications', *IEEE Journal of Biomedical and Health Informatics*, Vol. 23, No. 5, pp.1877–1887.
- Kennedy, H. (2018) 'Living with data: aligning data studies and data activism through a focus on everyday experiences of datafication', *Krisis: Journal for Contemporary Philosophy*, Vol. 2018, No. 1, pp.18–30.
- Liang, F., Das, V., Kostyuk, N. and Hussain, M.M. (2018) 'Constructing a data-driven society: China's social credit system as a state surveillance infrastructure', *Policy & Internet*, Vol. 10, No. 4, pp.415–453.
- Makransky, G. and Lilleholt, L. (2018) 'A structural equation modeling investigation of the emotional value of immersive virtual reality in education', *Educational Technology Research and Development*, Vol. 66, No. 5, pp.1141–1164.
- Pangrazio, L. and Sefton-Green, J. (2020) 'The social utility of 'data literacy'', *Learning, Media and Technology*, Vol. 45, No. 2, pp.208–220.
- Park, M., Im, H. and Kim, D.Y. (2018) 'Feasibility and user experience of virtual reality fashion stores', *Fashion and Textiles*, Vol. 5, No. 1, pp.1–17.
- Pietarinen, T., Vauras, M., Laakkonen, E., Kinnunen, R. and Volet, S. (2019) 'High school students' perceptions of affect and collaboration during virtual science inquiry learning', *Journal of Computer Assisted Learning*, Vol. 33, No. 4, pp.334–348.
- Quesnel, D., DiPaola, S. and Riecke, B.E. (2018) 'Deep learning for classification of peak emotions within virtual reality systems', *International Series on Information Systems and Management in Creative eMedia (CreMedia)*, Vol. 2, No. 2, pp.6–11.
- Taylor, J. and Pagliari, C. (2018) 'Mining social media data: how are research sponsors and researchers addressing the ethical challenges?', *Research Ethics*, Vol. 14, No. 2, pp.1–39.
- Trottier, D. (2018) 'Scandal mining: political nobodies and remediated visibility', *Media, Culture and Society*, Vol. 40, No. 6, pp.893–908.
- Wagler, A. and Hanus, M.D. (2018) 'Comparing virtual reality tourism to real-life experience: effects of presence and engagement on attitude and enjoyment', *Communication Research Reports*, Vol. 33, No. 4, pp.456–464.
- Walsh, J.P. (2020) 'Social media and moral panics: assessing the effects of technological change on societal reaction', *International Journal of Cultural Studies*, Vol. 23, No. 6, pp.840–859.
- Wiggins, A. and Wilbanks, J. (2019) 'The rise of citizen science in health and biomedical research', *The American Journal of Bioethics*, Vol. 19, No. 8, pp.3–14.
- Williamson, B. and Eynon, R. (2020) 'Historical threads, missing links, and future directions in AI in education', *Learning, Media and Technology*, Vol. 45, No. 3, pp.223–235.
- Yuan, S.N.V. and Ip, H.H.S. (2018) 'Using virtual reality to train emotional and social skills in children with autism spectrum disorder', *London Journal of Primary Care*, Vol. 10, No. 4, pp.110–112.