Online-Journals.org

iJET International Journal of Emerging Technologies in Learning

iJET | eISSN: 1863-0383 | Vol. 18 No. 18 (2023) | OPEN ACCESS

https://doi.org/10.3991/ijet.v18i18.43505

PAPER

Analysis of the Sentiment in the Evaluation Texts of University Students by Means of the Concept of Flexible Management

Changyan Zhu(✉)

College of Marxism,
Yantai Nanshan University,
Longkou, China

zhuchangyan2023@163.com

ABSTRACT

With the development of information technology (IT) and the popularity of the Internet, it is easier to obtain college students' teaching evaluation text data. However, it is still challenging to deeply understand and effectively analyze these texts. Based on the flexible management concept in universities, this study aimed to understand and analyze the teaching evaluation texts of college students more accurately, thereby improving the teaching quality. The implicit features of those texts were recognized first in this study, which aimed to supplement and improve the implicit features neglected by existing sentiment analysis methods. A sentiment analysis method based on Bidirectional Encoder Representations from Transformers (BERT) was adopted to explore the deep semantic information of texts. It was done by using deep learning technology and to improve the accuracy of sentiment analysis. This in turn provides more valuable reference information for university teaching management and further promotes the practical application of the flexible management concept in university teaching management.

KEYWORDS

college students' teaching evaluation texts, sentiment analysis, recognition of implicit features, Bidirectional Encoder Representations from Transformers (BERT), flexible management concept

1 INTRODUCTION

With the rapid development of information technology (IT) and the growing popularity of the Internet in today's society, big data has been widely used in all walks of life, including the education field [1–3]. In this context, it is easier to obtain the teaching evaluation text data of college students, which contains a large amount of teaching evaluation information, helping teachers provide timely feedback on the teaching effects and further optimizing teaching methods [4–7]. However, these

Zhu, C. (2023). Analysis of the Sentiment in the Evaluation Texts of University Students by Means of the Concept of Flexible Management. *International Journal of Emerging Technologies in Learning (iJET)*, 18(18), pp. 261–276. https://doi.org/10.3991/ijet.v18i18.43505

Article submitted 2023-04-27. Revision uploaded 2023-07-24. Final acceptance 2023-07-28.

teaching evaluation texts are diverse and complex, and they contain not only clear evaluation opinions but also a large number of implicit, vague, or emotionally rich descriptions. Therefore, deeply understanding and effectively analyzing the texts is a highly challenging task [8–9]. In view of this situation, it is of extremely important theoretical and practical significance to analyze and utilize the rich and complex text information by applying advanced artificial intelligence (AI) technology to the education field [10–14].

A flexible management concept has been gradually accepted and implemented in the current educational management field, especially in university teaching management. This concept advocates flexible adjustment of teaching methods based on students learning needs and characteristics, aiming to comprehensively improve the teaching quality [15–18]. In this context, accurate understanding and analysis of college students' teaching evaluation texts not only help teachers better adapt to students learning needs to improve teaching quality but also provide strong decision-making support for university education management, thereby effectively improving the efficiency and effect of educational management. However, the urgent problems to be solved are how to efficiently extract valuable information from the massive data of teaching evaluation texts and make accurate sentiment analysis of the information, thereby accurately grasping students' emotional attitudes and needs.

The methods based on dictionaries and machine learning have been mainly adopted in the existing sentiment analysis studies of college students teaching evaluation texts. However, when dealing with large-scale, complex, and unclearly defined text data, these methods often have poor performance and cannot meet the needs of deep and precise analysis [19–21]. At the same time, existing methods often neglect the implicit features of texts, such as implicit emotional expression, irony, and metaphor, which may lead to certain biases in the sentiment analysis results and cannot fully and accurately reflect students' true feelings and needs.

Therefore, this study aimed to explore and solve the above problems. The implicit features of college students' teaching evaluation texts were recognized first, which aimed to supplement and improve the implicit features neglected by existing sentiment analysis methods, thereby understanding students' true emotional attitudes more accurately. A sentiment analysis method based on Bidirectional Encoder Representations from Transformers (BERT) was adopted to explore the deep semantic information of texts. It was done by using deep learning technology, thereby improving the accuracy of sentiment analysis. It is expected that such a study can help understand the teaching evaluation of college students more accurately, provide more valuable reference information for university teaching management, and further promote the practical application of flexible management concepts in university teaching management.

2 IMPLICIT FEATURE RECOGNITION OF COLLEGE STUDENTS' TEACHING EVALUATION TEXTS

Existing implicit feature recognition methods have such obvious shortcomings that the sentiment analysis of college students' teaching evaluation texts under the flexible management concept may not be possible. First, these methods usually infer relationships based on co-occurrence and association rules and mainly rely on the mapping relationships between viewpoint words and feature attributes, which typically require model training using tagged corpora. However, it is not easy to obtain a large number of tagged teaching evaluation corpora for actual text sentiment analysis, which undoubtedly increases the difficulty and complexity of the research. Second, it is difficult to accurately extract implicit information using the mapping relationship method of viewpoint words and feature attributes because the teaching evaluation texts of college students are usually highly complex and diverse and contain rich implicit viewpoints and deep emotional colors. Moreover, these methods often neglect the importance of non-viewpoint words in recognizing implicit features. In the absence of viewpoint words, non-viewpoint words may play a crucial role, providing important contextual information to help understand and recognize implicit features. Figure 1 shows the framework of the implicit feature recognition method.

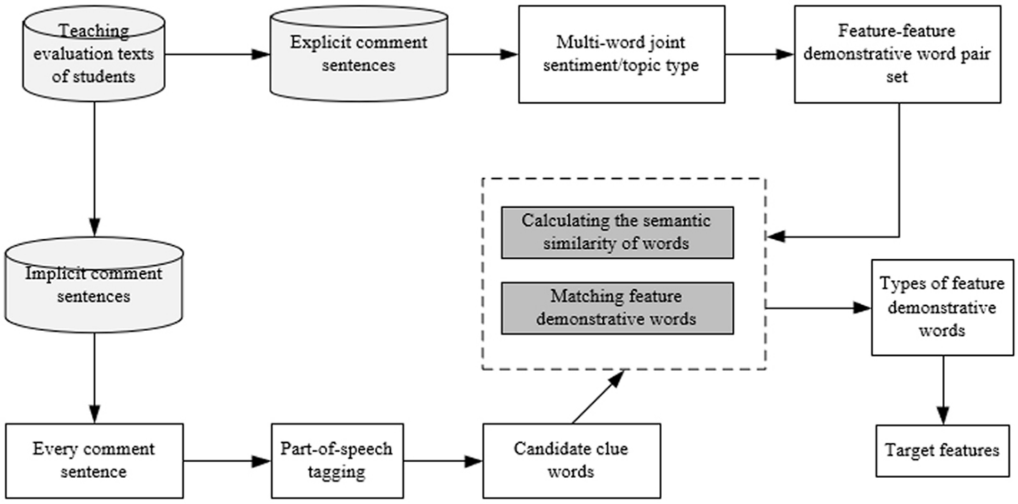

Fig. 1. Framework of implicit feature recognition method

To make sentiment a analysis of college students' teaching evaluation texts under the flexible management concept, this study proposed an implicit feature recognition method based on domain feature demonstrative words. This method mined demonstrative words from feature categories in explicit comment sentences in specific fields using the constructed multi-word joint sentiment/topic model. Even in the absence of clear viewpoint words, the method effectively mined students' implicit evaluations and sentiments towards teaching by recognizing and understanding demonstrative words. Second, the method introduced a word vector model in the recognition process of implicit features, which aimed to measure the semantic correlation of concentrated words between clue words and feature demonstrative words in implicit comment sentences, which enhanced the semantic depth of analysis. Based on the types of clue words, the method recognized implicit features according to different situations, which increased the flexibility and adaptability of the method, enabling it to better cope with the complexity and diversity of students' teaching evaluation texts.

Traditional text sentiment analysis methods often ignore the differences in word types, which play a vital role in actual texts. The expression and interpretation methods of different types of words in texts may be completely different, which has a significant impact on the sentiment analysis results. A multi-word joint sentiment/topic model was constructed, and implicit variables representing word types were introduced that accurately recognized the word types. This further improves the accuracy and meticulousness of sentiment analysis. The model obtained the probability distribution of words of the same type, thereby understanding and analyzing emotional tendencies in the texts more deeply. Especially in the teaching evaluation texts of college students, where various types of words may be used to express their emotions and opinions, the types and probability distribution of these words were crucial for understanding and evaluating their teaching feedback.

The generation process of each document under the multi-word joint sentiment/topic model were given as follows:

Step 1: A type distribution $\vartheta \sim Fue(\gamma)$ and an emotional distribution $\tau_f \sim Fue(\gamma)$ of words were generated;

Step 2: A topic distribution $\varphi_{f,m} \sim Fue(\beta)$ and a universal emotional word distribution $\upsilon_m^H \sim Fue(\alpha_m)$ were generated for various emotional tendencies m of college students' teaching evaluation;

Step 3: A feature word distribution $\upsilon_m^H \sim Fue(\alpha_m)$, a specific emotional word distribution $\upsilon_m^P \sim Fue(\alpha_m)$ and a non-viewpoint feature demonstrative word distribution $\upsilon_m^U \sim Fue(\alpha_m)$ were generated for various emotional tendencies m and topics x;

Step 4: An emotional tag $m \sim MU(\tau_f)$ and a topic $x \sim MU(\varphi_f)$ were selected for various sentences in the teaching evaluation texts of college students;

Step 5: For various words $q_u \in f$ in the teaching evaluation texts, the emotional tags m of their documents were assigned to the words, and a topic $x_u \sim MU(\varphi_{f,m})$, the word type $t_u \sim MU(\vartheta)$ and word $q_u \sim MU\,(\phi_{m,xu}^{tu})$ were selected.

Parameter estimation is an important step in constructing an effective model because it determines the predictive ability and accuracy of the model. As a widely used Markov chain Monte Carlo method, Gibbs sampling effectively estimates the parameters to improve the model's performance. Emotion dictionaries, domain emotion dictionaries, and domain feature dictionaries were introduced as prior knowledge. This not only enhanced the model's accuracy in recognizing word types but also made the model more targeted by introducing specific domain knowledge and thereby reflecting the emotional information in college students' teaching evaluation texts more accurately.

First, a topic and emotional tag were sampled for each word. The topic is the wide-ranging category to which the word belongs, such as "teaching quality," "course content," etc. The emotional tag refers to the emotional tendency expressed by the word, such as "positive," "negative," etc. In this step, the topic and emotional tag were jointly sampled; that is, the sampling considered both the topic and emotional tag of the word, which aimed to obtain the probability distribution of the topic and emotional tag of each word. The sampling condition formula was given as follows:

$$
\begin{aligned}
& o\left(a_u = k, x_u = j \middle| a_{-u}, x_{-u}, q\right) \infty \\
& \frac{V_{fk}^{FA} + \varepsilon_k}{\sum_{k=1}^{A} V_{fk}^{FA} + \varepsilon_k} \frac{V_{fkj}^{FAY} + \beta_j}{\sum_{j=1}^{Y} V_{fkj}^{FAY} + \beta_j} \\
& \frac{\Pi\left(\sum_{q=1}^{Q} V_{kjq}^{AYQ} + \alpha_{kq}\right)}{\Pi\left(\sum_{q=1}^{Q} V_{kjq}^{AYQ} + \alpha_{kq} + l_u\right)} \prod_{q=1}^{Q} \frac{\Pi\left(V_{kjq}^{AYQ} + \alpha_{kq} + l_u\right)}{\Pi\left(V_{klq}^{ayq} + \alpha_{kq}\right)}
\end{aligned}
\tag{1}
$$

Then word types were sampled based on Dirichlet's prior distribution. The word types include emotional words, domain emotional words, and domain feature words. The prior distribution is a kind of probability distribution widely used in Bayesian statistics that represents the uncertainty of a multinomial distribution. The prior distribution was introduced, which obtained the probability distribution of word types, further improving the model's generalization ability and accuracy. The sampling condition formula for the word type *t* of the *u*-th word was given as follows:

$$o\left(t_u = y \middle| t_{-u}, m, x, q\right) \infty \begin{cases} \dfrac{\left(b_{m,y,-u}^{(v)} + \alpha_u\right)\left(b_{m,f,-u}^{(y)} + \gamma_y\right)}{\sum_{u=1}^{C}\left(b_{m,y,-u}^{(v)} + \alpha_u\right)}, y = 0 \\ \dfrac{\left(b_{m,y,-u}^{(v)} + \alpha_u\right)\left(b_{m,f,-u}^{(y)} + \gamma_y\right)}{\sum_{u=1}^{C}\left(b_{m,y,-u}^{(v)} + \alpha_u\right)} \end{cases} \tag{2}$$

It is of great significance to further integrate the feature-demonstrative word pair set into an organized form. Each feature category (D1 = D2 … Dl) corresponded to three types of words, namely, general emotional words, feature emotional words, and non-viewpoint emotional words. For each word type, the words with the highest probability were retained. Then these words were filtered to only retain those with a strong semantic correlation. This screening method was based on a dual consideration of probability distribution and semantic correlation of words, which ensured that the selected words were both statistically significant and semantically relevant, thereby more accurately reflecting students' evaluations and emotions towards teaching. This integration method helped recognize and mine implicit features in students' teaching evaluation texts. The feature demonstrative word pair set was integrated, which recognized what feature categories and related emotional words were the focus of students' attention from a large number of teaching evaluation texts, thereby understanding students' needs and expectations.

During the implicit feature recognition of college students' teaching evaluation texts, the crucial step is to find the demonstrative word that best matches the clue word in the implicit comment sentence. Although the feature demonstrative word set, which was mined using the multi-word joint sentiment/topic model, may overlook words with low frequency but high semantic correlation, a model capturing the semantic correlation between words should be introduced to improve the probability of matching the demonstrative word with the highest correlation with the clue word. The Continuous Bag-of-Words (CBOW) model is a pre-trained word vector model that captures semantic and syntactic correlation between words, making it suitable for calculating the semantic correlation of concentrated words between clue words and feature demonstrative words. Compared to the co-occurrence model based solely on word frequency, the CBOW model better captures words with low frequency but high semantic correlation. This, in turn, improves the accuracy of implicit feature recognition, a critical for understanding the subtle differences and deep needs in students' teaching evaluation texts.

Let $\vec{q} = \left[z_1, z_2, \ldots, z_b\right]$ be the word vectors of college students' teaching evaluation texts in the Rb space, and then the semantic correlation between words was calculated using the following formula:

$$SIM\left(q_1,q_2\right)=\cos\phi=\frac{\overrightarrow{q_1}\cdot\overrightarrow{q_2}}{\left\|\overrightarrow{q_1}\right\|\cdot\left\|\overrightarrow{q_2}\right\|} \tag{3}$$

The implicit feature recognition based on feature demonstrative words includes three aspects, namely: selecting clue words, matching feature demonstrative words, and recognizing implicit features according to different situations based on the types of feature demonstrative words. First, clue words in implicit comment sentences were selected. Clue words mainly include general emotional words, feature emotional words, and non-viewpoint emotional words. These three types of words help understand and analyze users' emotions, which is the key to recognizing implicit features. The scoring formula for calculating candidate feature categories were given as follows:

$$SC\left(D_u\right)=s\times\sigma_u^{GLD}+n\times\pi_u^{GLD} \tag{4}$$

Let σ_u^{GLD} be the correlation between the contextual word CO_k of the clue word and the feature category D_u, then the explicit comment sentences tagged with feature categories in college students' teaching evaluations were represented using weighted logarithmic likelihood probability. Let $B_{COf,Du}$ be the frequency of the contextual word CO_k appearing in the comment sentences of feature category D_k, – be not, CO_{NU} be the number of contextual words in the clue words, and π_u^{GLD} be the probability that the matched feature demonstrative word belongs to the feature category D_u in given cases, then the calculation formula will be follows:

$$\begin{aligned}\sigma_u^{GLD}&=\sum_{k=1}^{CO_{NU}}o\left(CO_k\left|D_u\right.\right)\log\frac{o\left(CO_k\left|D_u\right.\right)}{o\left(CO_k\left|\bar{D}_u\right.\right)}\\&=\frac{B_{CO_k,D_u}}{B_{CO_k,D_u}+B_{\overline{CO}_k,D_u}}\log\frac{B_{CO_k,D_u}\left(B_{CO_k,D_u}+B_{\overline{CO}_k,\bar{D}_u}\right)}{B_{CO_k,\bar{D}_u}\left(B_{CO_k,\bar{D}_u}+B_{\overline{CO}_k,D_u}\right)}\end{aligned} \tag{5}$$

Let WO_{AL} be the number of all words, D_{AL} be the number of all topics, and $s+n=1$ be the weight of σ_u^{GLD} and π_u^{GLD} Based on the Bayesian principle, there were:

$$\pi_u^{GLD}=\frac{\upsilon^z*o\left(z\right)}{o\left(GLD\right)}=\frac{\upsilon^z*WO_{AL}}{Y_{AL}} \tag{6}$$

3 SENTIMENT ANALYSIS OF COLLEGE STUDENTS' TEACHING EVALUATION TEXTS UNDER THE FLEXIBLE MANAGEMENT CONCEPT

The teaching evaluation texts of college students usually belong to short-text data, which often contains a lot of noise, such as wrongly written words, syntax errors, irrelevant information, etc., interfering with the model's learning and affecting its capture of real and effective information. Therefore, this study first needed to clean the data before making sentiment analysis, such as by correcting wrongly written words, deleting useless symbols or tags, and unifying synonyms, thereby improving the data quality. Moreover, stop words, such as "of," "is/are," "in," etc., appeared frequently in the texts but often had little impact on the topics and emotions of the texts.

Therefore, the elimination of these words not only reduced the calculation amount of the model but also enabled the model to focus more on words significantly affecting the sentiment analysis.

This study also used the tokenizer of the BERT-based model for word segmentation. The BERT tokenizer, in addition to breaking down a text into words, preserves certain meaningful sub-words. This approach enhances the model's ability to capture the semantic nuances of words more effectively. In addition, the preprocessed texts were sent into a neural network model for sentiment analysis. The model captured complex semantic information and emotional tendencies at the word level, making it a main tool for sentiment analysis. Under the flexible management concept, the teaching and evaluation texts of college students may contain various complex emotions and viewpoints. The neural network model handled this complexity better.

The BERT model is built on the encoder module of the transformer. Positional encoding needs to be introduced to add positional information to each word because the Transformer model itself does not perceive the positional information of words. The BERT adopts a fixed positional encoding method, which generates positional encoding through the linear transformation of sine and cosine functions, and adds it to the word embedding vector, thereby maintaining the word order information.

In the BERT model, the self-attention mechanism is a crucial part, allowing the model to view information on other positions in the input sequence to better encode words at the current positions. For each word, the self-attention mechanism calculates its relationship or similarity with other words in the sequence and assigns different weights accordingly. In this way, the representation of each word contains contextual information about the entire text. The formula for the calculation of attention mechanism was as follows:

$$Attention(W, J, C) = SM\left(\frac{WJ^{Y}}{\sqrt{fj}}\right)C \tag{7}$$

where, the three matrices W, J, and C in the formula correspond to *Query*, *Key*, and *Value*, respectively. In practical applications, the multi-head attention mechanism is commonly used, which divides the input information into multiple parts, performs the self-attention operation on each part, and then splices the results, allowing the model to capture different types of information in different representation sub-spaces. For example, some heads may focus on capturing grammatical information, while others may focus on semantic information. The following formulas provided the operation process of the multi-head attention mechanism:

$$HE_u = Attention(WQ_u^W, JQ_u^J, CQ_u^C) \tag{8}$$

$$MH(W, J, C) = Concant(HE1, \ldots, HE_g)Q^P \tag{9}$$

The layer normalization operation was performed after the residual connection between the input and output of each self-attention and feed-forward neural network was established. This structure made gradient-back propagation easier, which was helpful for model training. If Z is the input of the residual module and $D(z)$ is the output of the module, then application of the residual module to the transformer results in:

$$Z + Attention(W, J, C) \tag{10}$$

Short-text sentiment analysis often faces the problem of sparse information and ambiguous context because short texts often contain only a small amount of information due to their length constraints, which may make it difficult to accurately reflect their true semantics and emotions. However, as a deep learning model, TextCNN (convolutional neural network for text processing) has excellent performance in dealing with such problems. Figures 2 and 3 show the principles and refined structure of TextCNN.

The input layer is the starting point for TextCNN to process data. In this step, each word is converted into a vector, which is usually completed through pre-trained word embedding. The input received by the input layer is usually a two-dimensional matrix, with the number of rows and columns corresponding to the number of words in texts and the word vector dimension, respectively. Let l be the number of words, and then the two-dimensional matrix of the input text vector is as follows:

$$E' = E^{l \times f} \tag{11}$$

The convolutional layer is the core part of TextCNN responsible for extracting useful features from the input data. It usually contains multiple convolution kernels of different sizes, each of which captures different features in texts. In convolution operations, the convolution kernel slides through the entire input matrix, calculating the dot product of the convolution kernel and the input matrix under the current window to capture local features. Let $g{*}f$ be the size of each convolution kernel, g be the number of words in the vertical direction of the convolution kernel, f be the dimension of the word vector, and $(l - g + 1)$ be the vector obtained by each convolution kernel after the convolution operation.

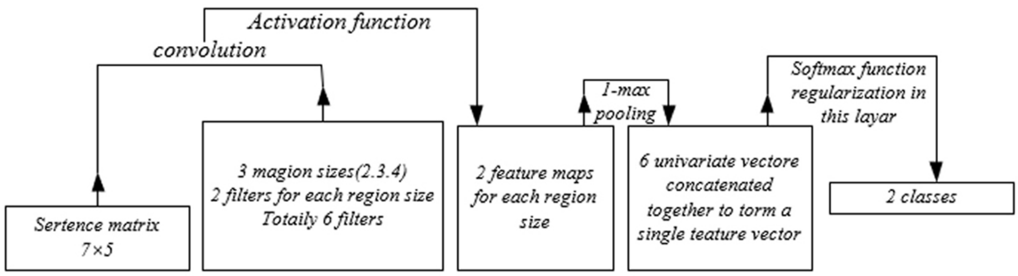

Fig. 2. TextCNN principle

The pooling layer is located after the convolutional layer, and its main functions are dimensionality reduction and abstract features. Global maximum pooling is commonly used in TextCNN, and this operation extracts the most important information (i.e., maximum value) from the feature map output by each convolution kernel, which helps the model capture the most important local features. The fully connected layer is located after the pooling layer, which connects all pooled features together to form a long vector. This vector is used as the input for the softmax layer. The fully-connected layer aims to learn the nonlinear relationships between features, further improving the model's performance.

In this study, the characteristics of BiGRU and TextCNN were combined in the sentiment analysis strategy of college students' teaching evaluation texts. BiGRU considered the contextual information in texts. When making sentiment analysis of texts, BiGRU captured the contextual relationships of words, helping understand their emotional tendencies in specific contexts. TextCNN was used to process text data.

It effectively captured local correlation between adjacent words in sentences through convolution and maximum pooling operations and extracted local n-gram features. The combination of BiGRU and TextCNN simultaneously considered contextual relationships and local correlation, which enabled the model to make a more comprehensive and accurate sentiment analysis of the input texts. After the model extracted and fused all feature information, the output was sent to the Softmax classifier. The classifier determined to which emotional category the input text was most likely to belong. It was based on the calculated probability distribution, which enabled the model to ultimately generate a clear emotional classification result.

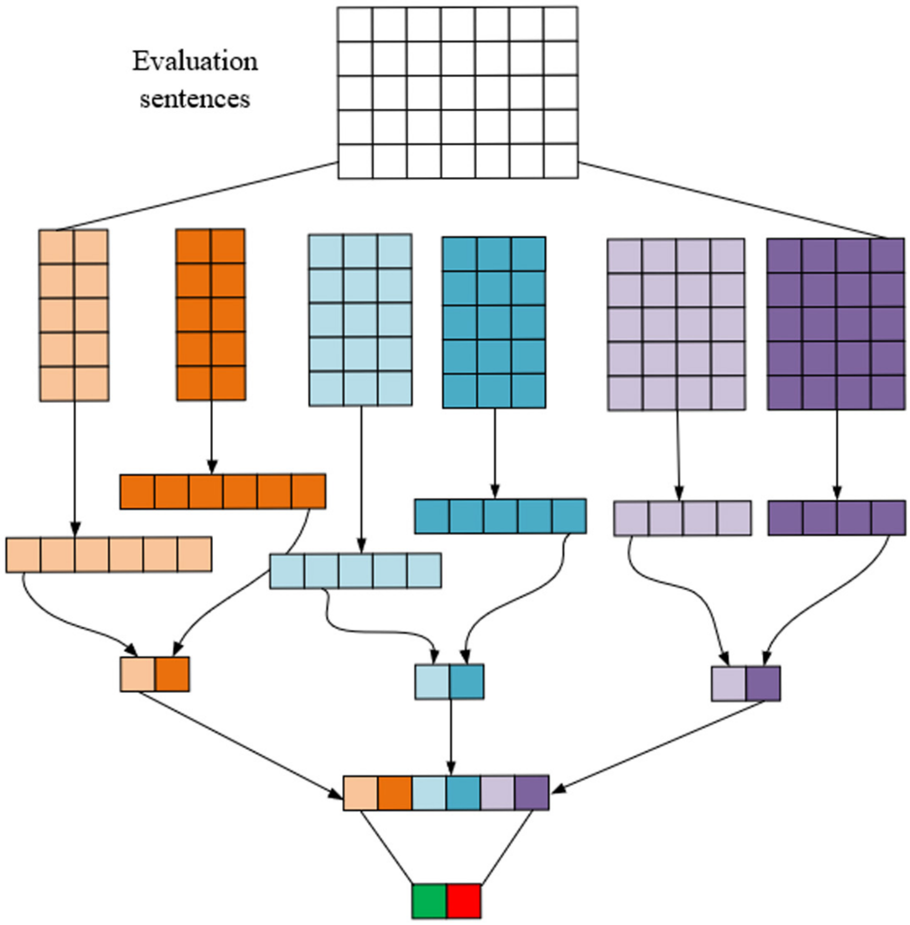

Fig. 3. Refined structure diagram of TextCNN

4 EXPERIMENTAL RESULTS AND ANALYSIS

It can be seen from the data in Figure 4 that the accuracy of the method proposed in this study exceeds or equals that of RNN and pLSA methods in implicit feature recognition of all five categories, which shows that the proposed method has better performance in recognizing the implicit features, such as teaching methods, personal characteristics of teachers, course content, course organization and management, and interaction between teachers and students, when making sentiment analysis of college students' teaching evaluation texts. Specifically, in terms of the two categories of personal characteristics of teachers and interaction between teachers and students, the accuracy of the proposed method increases by 0.05 to 0.06 compared with that of RNN and pLSA, which is a significant improvement. In terms of the other three categories, the accuracy of the proposed method is also slightly higher or equal to that of the RNN and pLSA methods. These results indicate that the method proposed in this study exhibits high accuracy and efficiency in implicit feature recognition, which verifies its effectiveness and superiority.

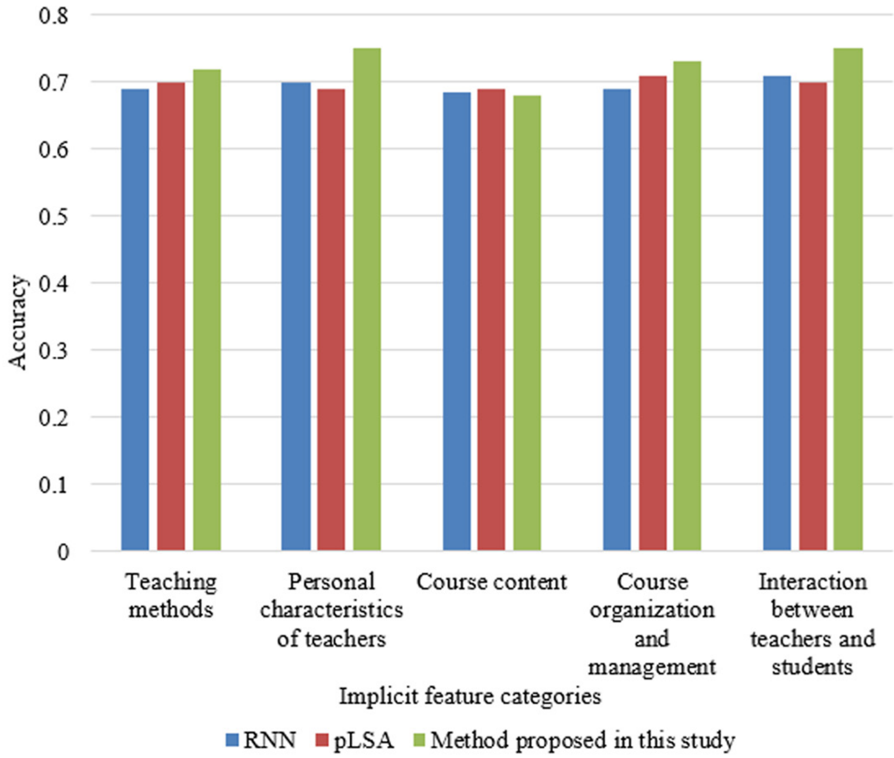

Fig. 4. Accuracy of implicit feature recognition

It can be seen from the data in Figure 5 that the method proposed in this study has good performance in recognizing the recall of implicit features in college students' teaching evaluation texts. Specifically, the recall of the proposed method exceeds that of the RNN and pLSA methods in four categories, namely, teaching methods, personal characteristics of teachers, course organization and management, and interaction between teachers and students. The proposed method lags slightly behind RNN and pLSA only in the category of course content, but the gap is not significant. In terms of teaching methods, the recall of the proposed method is 0.69, which is 0.05 and 0.04 higher than the 0.64 recall of the RNN method and the 0.65 recall of the pLSA method. In terms of the personal characteristics of teachers and the interaction between teachers and students, the recall of the proposed method is also higher than that of the other two methods. The recall of organizations and management also significantly increases. In summary, the method proposed in this study outperforms the RNN and pLSA methods on the whole in terms of recall of implicit feature recognition in college students' teaching evaluation texts, demonstrating the effectiveness and superiority of the proposed method. The lag in the course content category is slight, maybe because it is difficult to accurately recognize the features of the category or because there is still room for optimization in the model parameter setting or training process.

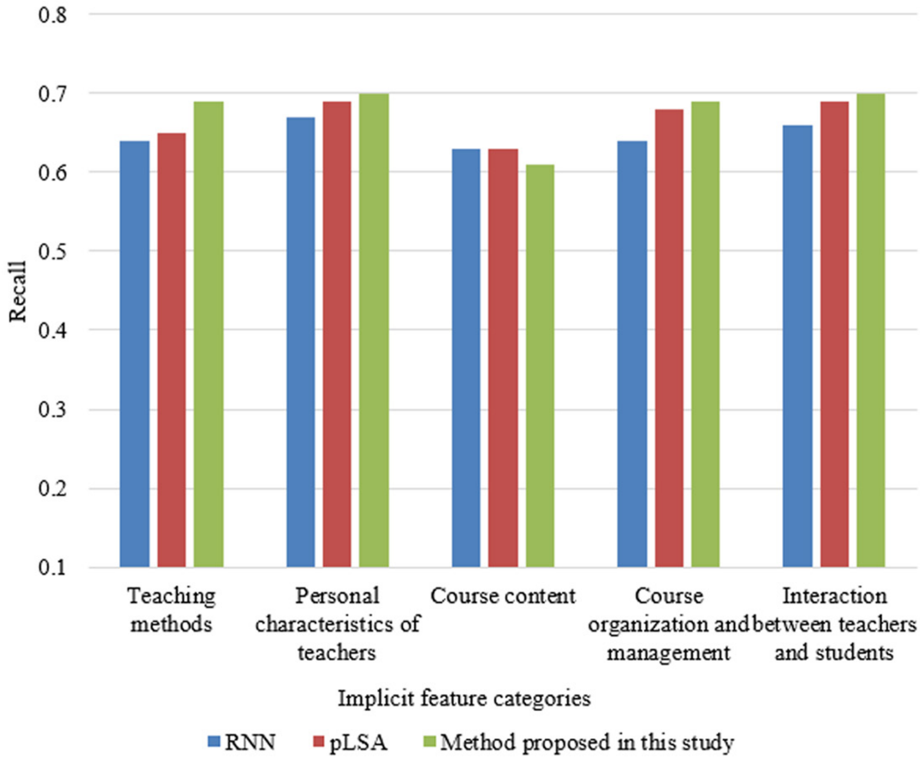

Fig. 5. Recall of implicit feature recognition

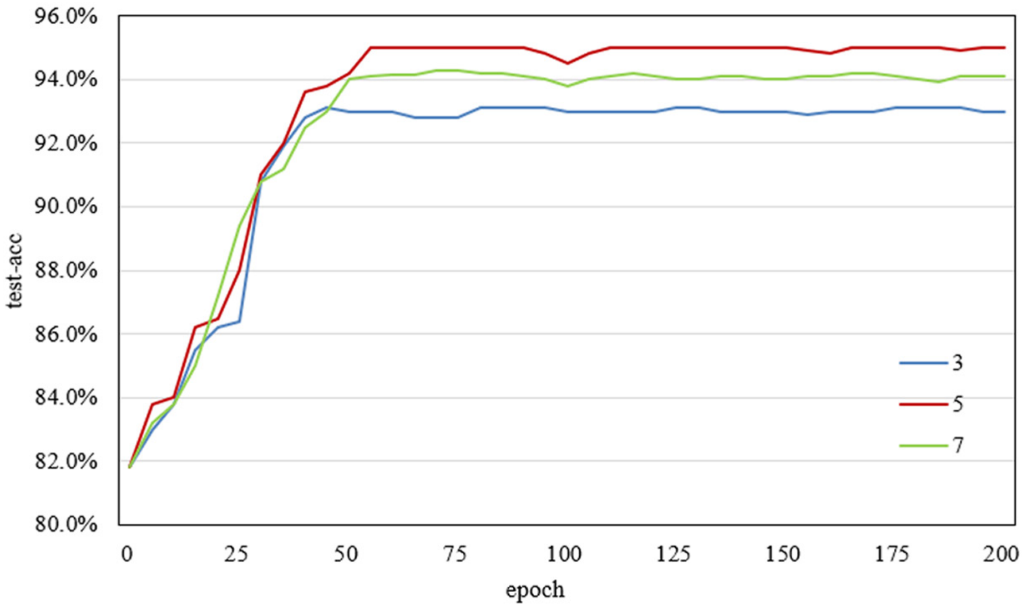

Fig. 6. Accuracy curves under different word vector dimensions

It can be seen from Figure 6 that the model's accuracy increases for three-dimensional, five-dimensional, and seven-dimensional word vectors with the increase in training epoch, indicating that the model's performance constantly increases. However, under the same training epoch, the larger the dimension of the word vector, the higher the accuracy of the model. First, when the training epoch is 200 hours, it can be observed that the accuracy of the three-dimensional, five-dimensional, and seven-dimensional word vectors is 93.0%, 95.0%, and 94.1%, respectively, with the accuracy of the five-dimensional word vector being the highest. Second, although there are no significant differences in the accuracy of three-dimensional, five-dimensional, and seven-dimensional word vectors in the early training stage (e.g., the epoch is 25 or 50 hours), the gap becomes more apparent, and the model with higher word vector dimensions also has higher accuracy as the training continues. This may be because high-dimensional word vectors capture more complex relationships between words. However, it should be noted that though the word vector with higher dimensions has better accuracy, it does not necessarily mean that higher dimensions are better. It can be seen from the data that the accuracy does not significantly increase but slightly decreases when the word vector dimension increases from five to seven, maybe because too large a word vector dimension may make the model too complex, leading to overfitting and affecting the generalization ability of the model. Overall, selecting the appropriate word vector dimension has a significant impact on the performance of the model. In practical applications, it is necessary to comprehensively consider the computational resources, complexity, and performance of the model to select the optimal word vector dimension. The five-dimensional word vector was considered the best choice because it achieved a good balance between accuracy and model complexity in the experiment of this study.

Table 1. Experimental evaluation results of sample set (without stop words)

	Accuracy	Precision	Recall	F1
Naive Bayes	0.8623	0.8525	0.8523	0.8516
Support vector machine	0.6814	0.8564	0.8524	0.8546
Random forest	0.8658	0.8616	0.8635	0.8629
LSTM	0.8685	0.8657	0.8655	0.8635
Method proposed in this study	0.8768	0.8733	0.8743	0.8743

Table 2. Experimental evaluation results of sample set (with stop words)

	Accuracy	Precision	Recall	F1
Naive Bayes	0.8976	0.8934	0.8952	0.8965
Support vector machine	0.9012	0.8945	0.8961	0.8915
Random forest	0.9001	0.8996	0.8992	0.8935
LSTM	0.9021	0.8978	0.8945	0.8942
Method proposed in this study	0.9034	0.8998	0.9121	0.9001

It can be seen from the data in Table 1 that the method proposed in this study exhibits certain advantages in four evaluation indexes, namely, accuracy, precision, recall, and F1-score (F1). First, compared with Naive Bayes, support vector machines, random forest, and Long-Short-Term Memory (LSTM), the accuracy of the proposed method is 0.8768, which is higher to some extent, showing a high accuracy. Second, the precision of the proposed method is 0.8733, which also performs better than the other four methods, showing a higher recognition rate for true positive examples. Third, the recall of the proposed method is 0.8743, which is the highest among the five methods, indicating that the proposed method can find more positive examples among all true positive examples. Finally, the F1-score of the proposed method is also 0.8743, which is the same as the recall and is the highest among the five methods. The F1-score is the harmonic average of precision and recall, which is a good measure when both are important. In summary, the method proposed in this study has the best performance in all four indexes, which proves the effectiveness and superiority of the proposed method in implicit feature recognition.

It can be seen from Table 2 that the method proposed in this study performs well in four evaluation indexes, namely, accuracy, precision, recall, and F1-score. First, the accuracy of the proposed method is 0.9034, which is higher than that of the other four methods, namely, Naive Bayes, support vector machines, random forests, and LSTM, showing the advantages of the proposed method in terms of overall accuracy. Second, the precision of the proposed method is 0.8998. Although the value is slightly lower than that of support vector machines and random forests, the proposed method still has advantages compared with the performance of Naive Bayes and LSTM, which proves the effect of the proposed method in the proportion of true positive samples in predicted positive samples. Third, the recall of the proposed method is 0.9121, which is significantly higher than that of the other four methods, indicating that the proposed method has significant advantages in the proportion of all true positive samples found and demonstrating strong recall. Finally, the F1-score of the proposed method is 0.9001, which considers both precision and recall and is their harmonic average. Although the F1-score of the proposed method is slightly lower than that of Naive Bayes, it is higher than that of support vector machines, random forests, and LSTM, showing the effect of relative balance. In summary, although the method proposed in this study is not optimal in terms of precision and F1-score, it has outstanding performance in accuracy and recall, especially with significant advantages in recall, which proves the effectiveness and superiority of the proposed method in implicit feature recognition.

It can be seen from the data in Figure 7 that the method proposed in this study exhibits certain advantages in four evaluation indexes, namely, accuracy, precision, recall, and F1-score. First, the accuracy of the proposed method shows an upward trend with the increase of evaluation indexes and reaches the highest of 0.782 at 16, which is slightly higher than the linear prediction of 0.78, indicating that the proposed method has good performance in accuracy. Second, although the precision of the proposed method is highly fluctuant throughout the entire process, it shows an overall upward trend and especially reaches the highest of 0.75 at 16, demonstrating high precision. Third, the recall of the proposed method also shows an upward trend and ultimately reaches 0.708 at 16, indicating that the proposed method also performs well in recall. Finally, as the harmonic average of precision and recall, the F1-score of the proposed method shows an upward trend throughout the entire process and reaches 0.71 at 16, which is a quite good result, indicating that the proposed method also has good performance when considering both precision and recall comprehensively. In summary, the method proposed in this study shows superiority in four evaluation indexes, which proves that the proposed method has high accuracy and effectiveness in the sentiment analysis of college students' teaching evaluation texts under the flexible management concept.

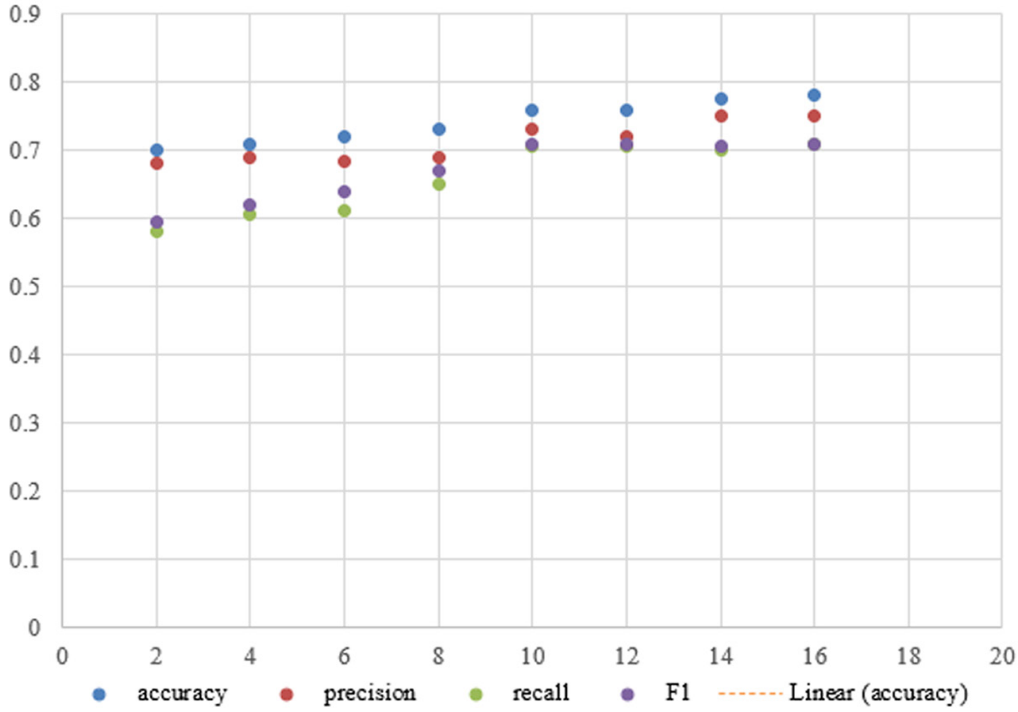

Fig. 7. Comparison of four evaluation indexes of the model

5 CONCLUSION

This study combined the fusion models of BiGRU and TextCNN, which have good performance in sentiment analysis tasks. BiGRU effectively captures the emotional features and contextual semantic correlations of texts. TextCNN, whereas, effectively captures the interrelationships between adjacent words in text sentences, which makes the model demonstrate high accuracy and reliability in sentiment analysis of college students' teaching evaluation texts. The proposed model has better performance than other comparison algorithms in implicit feature recognition. In addition, the proposed model has relatively high accuracy and recall in recognizing the five categories of implicit features in college students' teaching evaluation texts under the flexible management concept, indicating that the proposed model has a good ability to capture implicit features. As the sample size increased, the model's performance improved significantly. When using a larger training sample set, the accuracy and recall of the model increased, indicating that the model effectively learned from more data and improved its performance. Compared with other word vector models, such as Word2Vec, GloVe, and BERT, the model proposed in this study has higher accuracy, which shows its superiority in processing text sentiment analysis tasks.

Overall, the fusion model based on BiGRU and TextCNN proposed in this study has good performance in sentiment analysis of college students' teaching evaluation texts, high accuracy and recall, and a good ability to recognize implicit features, which provides an effective tool for deeply understanding the feedback and emotions of college students towards education and teaching, helping educators manage and improve teaching in a more refined manner.

6 REFERENCES

[1] W. Nuankaew and P. Nuankaew, “Educational engineering for models of academic success in Thai Universities during the COVID-19 pandemic: Learning strategies for lifelong learning,” *International Journal of Engineering Pedagogy*, vol. 11, no. 4, pp. 96–114, 2021. https://doi.org/10.3991/ijep.v11i4.20691
[2] H. Huang and C.T. Hsin, “Environmental literacy education and sustainable development in schools based on teaching effectiveness,” *International Journal of Sustainable Development and Planning*, vol. 18, no. 5, pp. 1639–1648, 2023. https://doi.org/10.18280/ijsdp.180535
[3] H. Wang, “Evaluation and measurement of student satisfaction with online learning under integration of teaching resources,” *International Journal of Emerging Technologies in Learning*, vol. 17, no. 11, pp. 99–113, 2022. https://doi.org/10.3991/ijet.v17i11.32037
[4] T. Vidić, I. Klasnić, and M. Đuranović, “Student evaluation of online teaching quality, their own engagement and success expectancy in the future profession,” *International Journal of Emerging Technologies in Learning*, vol. 17, no. 4, pp. 135–147, 2022. https://doi.org/10.3991/ijet.v17i04.27517
[5] E.Y. Gueddoudj, A. Chikh, and A. Attia, “Os-ETL: A high-efficiency, Open-Scala solution for integrating heterogeneous data in large-scale data warehousing,” *Ingénierie des Systèmes d’Information*, vol. 28, no. 3, pp. 557–565, 2023. https://doi.org/10.18280/isi.280303
[6] S. Anawar, N.F. Othman, S.R. Selamat, Z. Ayop, N. Harum, and F. Abdul Rahim, “Security and privacy challenges of big data adoption: A qualitative study in telecommunication industry,” *International Journal of Interactive Mobile Technologies*, vol. 16, no. 19, pp. 81–97, 2022. https://doi.org/10.3991/ijim.v16i19.32093
[7] R. Raman and P. Nedungadi, “Adoption of web-enabled student evaluation of teaching (WESET),” *International Journal of Emerging Technologies in Learning (IJET)*, vol. 5, no. 24, pp. 191–207, 2020. https://doi.org/10.3991/ijet.v15i24.17159
[8] Z.H. Xu, “Construction and optimization of talent training quality based on data mining,” *Ingénierie des Systèmes d’Information*, vol. 25, no. 4, pp. 419–425, 2020. https://doi.org/10.18280/isi.250403
[9] R. Faizi, “Using sentiment analysis to explore student feedback: A lexical approach,” *International Journal of Emerging Technologies in Learning*, vol. 18, no. 9, pp. 259–267, 2023. https://doi.org/10.3991/ijet.v18i09.38101
[10] J. He, K.Y. Lin, and Y. Dai, “A data-driven innovation model of big data digital learning and its empirical study,” *Information Dynamics and Applications*, vol. 1, no. 1, pp. 35–43, 2022. https://doi.org/10.56578/ida010105
[11] K. Orynbekova, A. Bogdanchikov, S. Cankurt, A. Adamov, and S. Kadyrov, “MapReduce solutions classification by their implementation,” *International Journal of Engineering Pedagogy*, vol. 13, no. 5, pp. 58–71, 2023. https://doi.org/10.3991/ijep.v13i5.38867
[12] K. Wang, “Academic early warning model for students based on big data analysis,” *International Journal of Emerging Technologies in Learning*, vol. 18, no. 12, pp. 16–31, 2023. https://doi.org/10.3991/ijet.v18i12.41087
[13] M.U. Abdullah and A. Alkan, “A comparative approach for facial expression recognition in higher education using hybrid-deep learning from students’ facial images,” *Traitement du Signal*, vol. 39, no. 6, pp. 1929–1941, 2022. https://doi.org/10.18280/ts.390605
[14] N. Sharma and V. Jain, “Evaluation and summarization of student feedback using sentiment analysis,” In *Advanced Machine Learning Technologies and Applications: Proceedings of AMLTA 2020*, vol. 1141, pp. 385–396, 2021. https://doi.org/10.1007/978-981-15-3383-9_35

[15] B.K. Prahani, I.R. Dawana, B. Jatmiko, and T. Amelia, "Research trend of big data in education during the last 10 years," *International Journal of Emerging Technologies in Learning*, vol. 18, no. 10, pp. 39–64, 2023. https://doi.org/10.3991/ijet.v18i10.38453

[16] L. Zi, "Influence of flexible management on students' entrepreneurial competency in private universities: Taking China Nanning university as an example," In *2022 13th International Conference on E-business, Management and Economics*, pp. 664–670, 2022. https://doi.org/10.1145/3556089.3556200

[17] R. Duarte, A.L. de Oliveira Pires, and Â.L. Nobre, "Increasing adult students' learning opportunities with flexible learning pathways: Evidence from a technology and industrial management graduate course," In *2016 2nd International Conference of the Portuguese Society for Engineering Education (CISPEE)*, pp. 1–6, 2016. https://doi.org/10.1109/CISPEE.2016.7777731

[18] I.R. Banday, M. Zaman, S.M.K. Quadri, S.A. Fayaz, and M.A. Butt, "Big data in academia: A proposed framework for improving students performance," *Revue d'Intelligence Artificielle*, vol. 36, no. 4, pp. 589–595, 2022. https://doi.org/10.18280/ria.360411

[19] S. Rusche, J. Weissflog, S. Wenninger, and B. Häckel, "How flexible are energy flexibilities? Developing a flexibility score for revenue and risk analysis in industrial demand-side management," *Applied Energy*, vol. 345, p. 121351, 2023. https://doi.org/10.1016/j.apenergy.2023.121351

[20] F. Ouatik, M. Erritali, F. Ouatik, and M. Jourhmane, "Students' orientation using machine learning and big data," *International Journal of Online and Biomedical Engineering*, vol. 17, no. 1, pp. 111–119, 2021. https://doi.org/10.3991/ijoe.v17i01.18037

[21] T. Wu, C. Wang, Y. Hu, X. Zeng, and M. Song, "Flexible solid-solid phase change materials with high stability for thermal management," *International Journal of Heat and Mass Transfer*, vol. 211, p. 124202, 2023. https://doi.org/10.1016/j.ijheatmasstransfer.2023.124202

7 AUTHOR

Changyan Zhu born in Shandong, China, in 1980. She presently serves as an Associate Professor at Yantai Nanshan University. Her research focuses on college student management and ideological-political education. Zhu has led eight department- and bureau-level projects, contributed to four provincial projects, and participated in a Ministry of Education project. She has published over ten papers and edited two textbooks. (E-mail: zhuchangyan2023@163.com; Orcid: https://orcid.org/0000-0002-8093-1707).