https://doi.org/10.31449/inf.v49i8.9637 Informatica **49** (2025) 355–366 **355**

A Multidimensional-Weighted TextRank and LSTM-Attention Model for Network Public Opinion Sentiment Analysis

Minjie He, Qi Huang*
School of Business, Nantong Institute of Technology, Nantong 226002, China
E-mail: huangqi314159@126.com
*Corresponding author

Keywords: network public opinion, sentiment analysis, TextRank, keyword extraction, deep learning

Received: June 10, 2025

As social media rapidly develops, network public opinion has become an important channel for reflecting social emotions, especially in emergencies and public opinion surges. To improve the accuracy of public opinion sentiment analysis, a network public opinion sentiment analysis model integrating improved TextRank algorithm is proposed. By introducing multidimensional features such as term frequency inverse document frequency, part of speech, and word position, the keyword extraction process is improved, and combined with deep learning, the accuracy of model classification is enhanced. The findings indicated that the accuracy of the proposed model on the test set reached 0.96, and the F1 values on the training and testing sets were 92.6% and 90.9%, respectively, demonstrating the advantages of this method in complex sentiment analysis tasks. In addition, the model proposed by the research performed well in the sentiment classification task of four network public opinion hotspots, with the highest accuracy rates of positive and negative sentiment classification reaching 98% and 96% respectively, a root mean square error as low as 0.176, and a mean absolute percentage error of only 0.081. The results indicate that the model has better fitting and generalization abilities in sentiment classification tasks. This not only provides an efficient technical solution for sentiment analysis of network public opinion, but also lays an important foundation for the intelligent development of social media public opinion monitoring systems.

Povzetek: Model združuje večdimenzionalno utežen TextRank (TF-IDF, besedna vrsta, položaj; G1) z LSTM-pozornostjo za analizo sentimenta javnega mnenja.

1 Introduction

With the widespread use of social media, Network Public Opinion (NPO) has become an indispensable influencing factor in public events, especially in emergency situations where changes in public emotions can quickly spread and form a wide social impact [1]. The Sentiment Analysis (SA) of NPO, as an automated technology, has been widely utilized in fields such as public opinion guidance and sentiment prediction, and has become an important component of public opinion management [2]. SA technology has been broadly utilized in fields such as public opinion monitoring, consumer feedback analysis, and emotion prediction by classifying the emotional tendencies of online texts [3]. However, traditional SA methods often face noise interference and emotional diversity issues when dealing with complex and unstructured social media data. Therefore, how to extract effective emotional features from large-scale and complex network texts to improve the accuracy and robustness of SA has become a research focus in the current field of SA. Xu et al. used text analysis and sentiment calculation to identify fluctuating factors, and combined Granger causality test to screen key variables. Based on the grey prediction model, they constructed an optimized model that integrates public opinion fluctuations, significantly improving prediction accuracy on four types of emergency event data [4]. Xu et al. focused on typical campus public opinion events and used Latent Dirichlet Allocation (LDA) for topic extraction, combined with Sentiment Knowledge Enhanced Pre-training (SKEP) model to complete emotion classification. They revealed the evolution law of public opinion from two dimensions: spatiotemporal and population characteristics, providing theoretical support for campus public opinion governance, but still limited by model accuracy [5]. Qiu et al. used Python to preprocess text data and combined spectral clustering with LDA topic models to mine high-value topics from multiple sources of public opinion. They proposed a method based on spectral clustering algorithm. By means of visual analysis, the core issues were effectively identified, and the evolution of public emotions throughout the process of public opinion dissemination was mapped out [6]. Shackleford et al. proposed a fusion of an improved Valence Aware Dictionary And Sentiment Reasoner (VADER) dictionary with multiple classical machine learning algorithms, and constructed multiple hybrid models. After comparing and evaluating using standard performance indicators, it was found that the combination of VADER dictionary and medium Gaussian support vector machine performed the best, showing significant advantages among the seven comparison schemes [7].

Table 1: Literature summary table.

Authors	Year	Algorithms/Methods used	Key results	Limitations
Xu et al. [4]	2023	Granger causality+Gray prediction model	Improved the accuracy of predicting public opinion on unexpected events	Dependent on accuracy of factor selection and Granger test assumptions
Xu et al. [5]	2024	LDA+SKEP sentiment classification+spatial-temporal analysis	Effectively identified emotional features of campus opinion	Limited by current sentiment classification model accuracy
Qiu et al. [6]	2022	Spectral clustering+LDA+visualization	Identified core topics and emotional shifts in multi-source public opinion	Limited scalability
Shackleford et al. [7]	2023	Improved VADER+Medium Gaussian Support Vector Machine	Achieved best performance in 7 schemes	Generalization to multilingual text not discussed
Guda et al. [9]	2023	TextRank method using FOX stop word list	F1 is 16.59% and 14.22% respectively	Limited robustness across datasets
Lu et al. [10]	2023	SciBERT+TextRank+DPCNN	Optimized citation recommendation system	Dependent on external vocabulary knowledge base
Zhili et al. [11]	2024	SSA-optimized BiLSTM	The model evaluation results are highly consistent with manual scoring	Limited scope of application
Li et al. [12]	2024	GCN+BiLSTM	Significantly improve deep question answering performance	Model structure may increase training cost and data dependency

Recently, the combination of keyword extraction and deep learning methods has gradually become a research hotspot in SA. The TextRank algorithm, an unsupervised learning method based on graph ranking, has obtained notable achievements in tasks such as keyword extraction and text summarization [8]. Guda et al. compared and analyzed the performance of fast automatic keyword extraction algorithm and TextRank algorithm under different stop word lists. The findings denoted that the TextRank method using FOX stop word list had the best performance, with F1 values of 16.59% and 14.22% on text and speech data, respectively [9]. Lu et al. proposed a Scientific Bidirectional Encoder Representation from Transformers (SciBERT) model that integrates vocabulary database knowledge. This method combined TextRank to automatically extract literature topics and used Deep Pyramid Convolutional Neural Networks (DPCNN) to construct a scientific paper semantic representation and citation recommendation system. Findings denoted that the model achieved optimal performance in a single WordNet fusion [10]. In addition, Zhili et al. proposed a deep learning-based method for evaluating semantic similarity of English translation keywords. Firstly, the keywords in the translated text were extracted using the co-occurrence algorithm, and the Sparse Search Algorithm (SSA) was used to adjust the network weights. A Bidirectional Long Short-Term Memory (BiLSTM) neural network model optimized by SSA was constructed. The experimental data showed that the sentence similarity evaluation results obtained by this method were highly consistent with the manual professional rating [11]. Li et al. proposed a hybrid neural network model that integrates Graph Convolutional Network (GCN) and BiLSTM, introducing dual attention and gating mechanisms, and optimizing the joint expression of document and graph structures through contrastive learning. The experimental verification on the HotpotQA dataset showed that this method could effectively improve the performance of deep problem solving [12]. The research methods, core achievements, and existing problems of the literature have been summarized and organized, as shown in Table 1.

Based on Table 1, although research in this field has been progressing steadily, especially in the application of keyword extraction and deep learning models. However, traditional TextRank algorithms and other methods still have certain limitations, especially in terms of improving sentiment classification accuracy and model generalization ability. In view of this, an NPO SA model integrating improved TextRank algorithm is proposed, which enhances the ability to extract sentiment keywords by introducing multidimensional features such as Term Frequency Inverse Document Frequency (TF-IDF), part of speech, and word position for keyword extraction. Unlike previous graph sorting methods that used static weights or single feature initialization, G1 weighting can dynamically adjust the contributions of each feature and enhance the sensitivity of keyword extraction to complex emotional expressions. On this basis, the model utilizes Long Short-Term Memory (LSTM) networks to capture context dependent structures and introduces attention mechanisms to weight and aggregate key information, thereby enhancing the accuracy and robustness of sentiment discrimination. Not only does it form a highly coupled linkage mechanism of "keyword extraction emotion discrimination" in the model structure, but it also demonstrates strong cross topic adaptability and model interpretability through empirical verification in multiple public opinion hot topic tasks. The research aims to bridge the gap between graph sorting methods and deep models, improve the comprehensive performance of NPO SA, and provide a more practical new technological path for social media sentiment recognition in complex contexts.

2 Methods and materials

2.1 Improved textrank keyword extraction algorithm

The traditional TextRank algorithm usually assigns the same initial weight to all candidate word nodes in the keyword extraction process, ignoring the significant differences in semantic structure and text distribution of words, resulting in certain generalization limitations in keyword recognition [13]. To address this issue, the study introduces three semantic related attributes: part of speech, word position, and TF-IDF value, and constructs a multidimensional feature matrix to comprehensively measure the importance of words. TF-IDF is a statistical feature weighting method that evaluates the importance of words in text by calculating term frequency (TF) and document frequency [14]. Among them, TF reflects the frequency of words in the current text, while inverse document frequency (IDF) measures their scarcity in the corpus. The importance of the word is contingent upon the magnitude of the product value. The expressions for TF and IDF are shown in equation (1) [15].

$$\begin{cases} TF(t,d)=\dfrac{f_{t,d}}{\sum_k f_{t,d}} \\ IDF(t,d)=\log(\dfrac{N}{1+n_t}) \end{cases} \tag{1}$$

In equation (1), $f_{t,d}$ refers to the amount of times the word t appears in document d, N represents the total amount of documents in the corpus, and n_t represents the amount of documents containing the word t. The TF-IDF value is the product of TF and IDF, as shown in equation (2).

$$TF\text{-}IDF(t,d,D)=\frac{f_{t,d}}{\sum_k f_{t,d}}\cdot\log(\frac{N}{1+n_t}) \tag{2}$$

In equation (2), D means the collection of all documents in the entire corpus. The importance of keywords is often determined by multiple heterogeneous features, such as word frequency intensity, sentence position, and part of speech category. The impact of these three attributes on the salience of keywords varies in different contexts. Compared with traditional fixed weight allocation or simple arithmetic mean methods, the G1 dynamic weighting algorithm can adaptively adjust weights based on the distribution characteristics of features in the dataset, thereby more accurately characterizing the actual contribution value of each feature in semantic representation. Therefore, the study used the G1 weighting method to weight the differences among the three types of attributes, calculate the comprehensive weight of each word, and use it as the initial score input for graph nodes in the improved TextRank algorithm to enhance the semantic sensitivity of keyword ranking. The G1 weighting method is a subjective objective fusion method for determining weights, which utilizes the degree of difference between adjacent indicators to determine weights and avoid subjective settings. The difference sequence between indicators is calculated as denoted in equation (3) [16].

$$c_j=\sum_{i=1}^{n-1}\left|a_{i+1,j}-a_{i,j}\right| \tag{3}$$

In equation (3), $a_{i,j}$ represents the value of the i th sample on the j th attribute, and n represents the total amount of samples. c_j represents the degree of difference of the j th attribute, which is used to measure the magnitude of its variation in the sample. Then, the relative weight is calculated, as shown in equation (4).

$$\lambda_j=\frac{c_j}{\sum_{k=1}^{3}c_k} \tag{4}$$

In equation (3), λ_j denotes the weight of the j th indicator, and $\sum_{k=1}^{3}c_k$ represents the sum of all attribute differences, used for normalization. 3 represents the total number of attributes, including TF-IDF, part of speech, and word position. After integrating attributes and weights, the initial rating for each word is obtained, as shown in equation (5).

$$\omega=\omega_1\cdot TF\text{-}IDF+\omega_2\cdot loc+\omega_3\cdot pos \tag{5}$$

In equation (5), ω represents the comprehensive weight, while ω_1, ω_1, and ω_3 are the weights of TF-IDF, word position, and part of speech, respectively. loc and loc respectively represent word position features and part of speech features. The comprehensive weight attributes are shown in Figure 1.

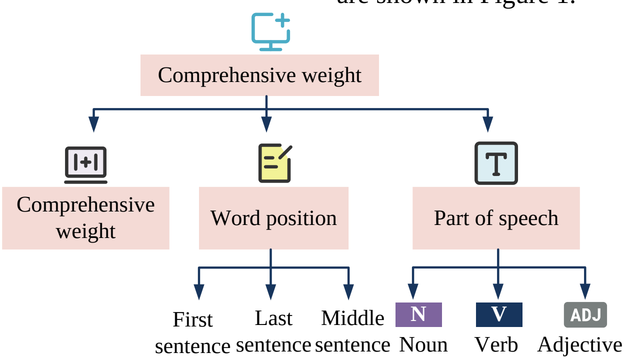

Figure 1: Schematic diagram of comprehensive weight attributes. (Source from: Author's self drawn)

In Figure 1, the comprehensive weights are constructed from three aspects: TF-IDF value, word position, and part of speech. The TF-IDF value corresponds to its weight, and the word position feature weight is divided into the first sentence, last sentence, and middle sentence according to the position in the sentence. The weight of part of speech features includes nouns, verbs, and adjectives. The G1 weighting method is used to determine the comprehensive weights of three attributes, which are used as the initial weights for keyword extraction in the TextRank algorithm. The improved TextRank (I-TextRank) algorithm is obtained, and the expression is denoted in equation (6).

$$S(\omega_i) = (1-\alpha)\cdot\omega + \alpha\cdot\sum_{\omega_j\in In(w_i)}\frac{S(\omega_j)}{|Out(\omega_j)|} \tag{6}$$

In equation (6), $S(\omega_i)$ represents the final weight, α represents the damping coefficient, generally set to 0.85, ω_j is the input node of ω_i, $In(\omega_i)$ stands for the set of all nodes pointing to ω_i, and $Out(\omega_j)$ indicates the set of all output nodes pointing to ω_j. The overall process of the I-TextRank algorithm is denoted in Figure 2.

In Figure 2, the input text is first preprocessed, including TF-IDF value calculation of words, position feature extraction, and part of speech tagging. After completing the three features, the G1 weighting method is used to calculate the comprehensive weights and generate the initial weights for each word. Based on these weights, the algorithm constructs an I-TextRank graph structure and performs iterative calculations to determine word importance through node ranking. After the graph sorting is completed, the algorithm filters candidate words based on a preset threshold, sorts them by score, and outputs the final keyword list.

2.2 NPO sentiment analysis model Integrating I-TextRank and LSTM-attention

The development process of NPO is not only driven by information dissemination mechanisms, but also by the combined effect of public attitudes and media reactions, forming a dynamic chain of "information diffusion-social response-public opinion evolution". The generation of public opinion is not a single dimensional dissemination phenomenon, but a collective construction process of risk perception under multi-party interaction. The social risk evolution of NPO in emergencies is shown in Figure 3 [17].

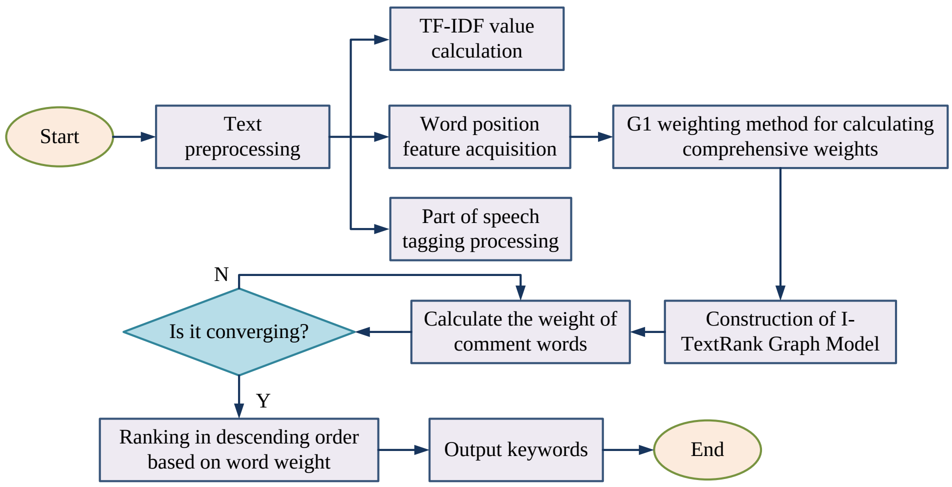

Figure 2: I-TextRank algorithm process. (Source from: Author's self drawn)

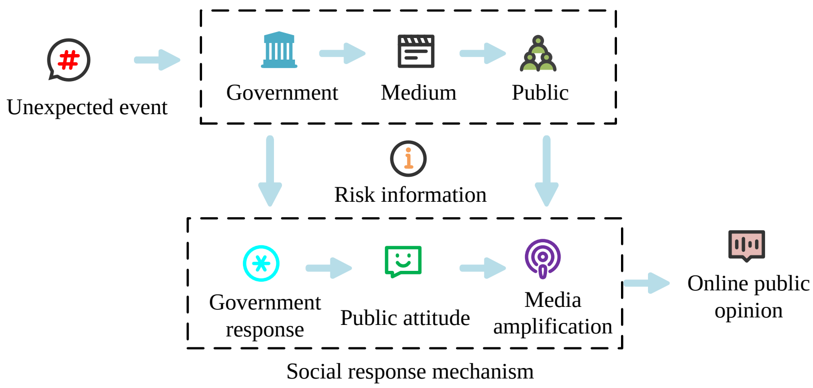

Figure 3: The social risk framework of NPO. (Source from: Author's self drawn)

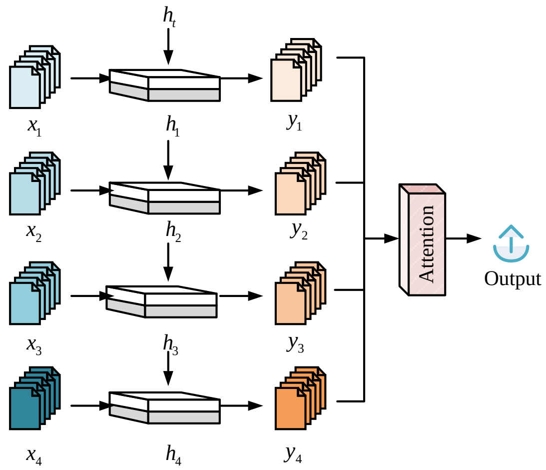

Figure 4: LSTM-Attention structure. (Source from: Author's self drawn)

Figure 3 shows the social amplification process of NPO triggered by emergencies, including three main stages: information dissemination path, amplification mechanism, and social feedback mechanism. After an emergency occurs, relevant information is transmitted to the public through the dissemination chain, with the government, media, and the public forming the initial amplification station, playing a core role as the main body of information diffusion in characterizing risk events. Subsequently, risk information triggers government response and public emotional reactions, and this social feedback process is further amplified by media coverage and public behavior, ultimately forming public opinion fluctuations in cyberspace. SA has become an important tool for understanding and grasping changes in public sentiment in this complex and dynamic public opinion environment. Research extracts keywords based on I-TextRank and constructs a classification model using deep learning techniques for sentiment polarity analysis. Firstly, the LSTM network is employed for the purpose of binary classification, with the objective of discriminating positive and negative emotions. Subsequently, AM is introduced with a view to optimizing the model's ability to capture key emotional information and to improve overall performance. The LSTM-Attention structure is denoted in Figure 4 [18].

In Figure 4, the LSTM-Attention model sequentially inputs sequence data x_1, x_2, x_3, x_4, and performs temporal processing through LSTM units to generate hidden state vectors h_1, h_2, h_3, h_4 and corresponding outputs y_1, y_2, y_3, y_4. h_t represents the hidden state vector at the t-th time step. These outputs are processed through an attention mechanism layer, which calculates the correlation score between each vector and the global context, assigns different attention weights, and then weights y_1, y_2, y_3, and y_4 to obtain the final context aware representation as the model output. LSTM receives the embedded vector sequence and outputs the hidden state sequence as shown in equation (7) [19].

$$h_t = LSTM(e_t, h_{t-1}) \tag{7}$$

In equation (7), e_t represents a low dimensional word vector. To weight each hidden state, the model introduces an AM to calculate the attention score for each time step. The expression for calculating attention score is shown in equation (8).

$$u_t = \tanh(\omega_u h_t + b_u) \tag{8}$$

In equation (8), u_t represents the attention score vector of the t th time step, ω_u represents the trainable weight matrix, and u_t is the bias vector, which increases the expressive power of the model. After normalization, the attention weight of each time step can be normalized to the relative importance of the current hidden state in sentiment classification, as expressed in equation (9).

$$\omega_A = \frac{\exp(u_t^T u_\omega)}{\sum_{k=1}^{T} \exp(u_t^T u_\omega)} \tag{9}$$

In equation (9), ω_A denotes the attention weight, u_ω refers to the trainable context vector, $u_t^T u_\omega$ represents the dot product of the attention score vector and the context vector, T represents the total length of the sequence. It is imperative to normalize all time-step attention scores, thereby ensuring that the sum of the weights is equal to one. To obtain the final weighted hidden state, the attention weights are utilized to weight and sum the hidden states of all time steps, as shown in equation (10).

$$v = \sum_{t=1}^{T} \omega_A h_t \tag{10}$$

In equation (10), v represents sentence sentiment representation that integrates attention information. Finally, the hidden states weighted by the AM are input into the fully connected layer, and the probability distribution of each emotion category is calculated using the softmax function, as denoted in equation (11).

$$\hat{y} = \text{softmax}(\omega_c v = b_c) \tag{11}$$

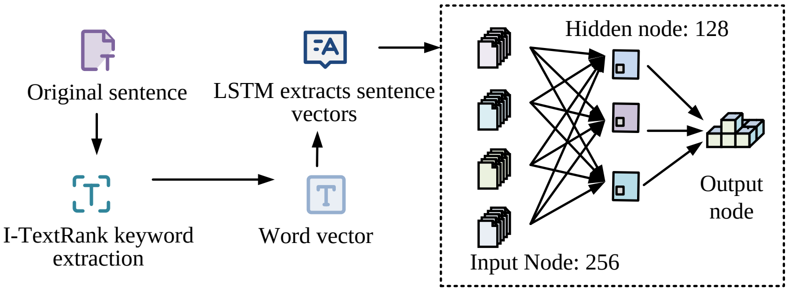

Figure 5: The overall architecture of the I-TextRank-based sentiment analysis framework. (Source from: Author's self drawn)

In equation (11), the probability distribution vector of the emotion category predicted by the $\hat{y}$ model represents the probability that the sentence belongs to each category. ω_c means the weight matrix, and b_c means the bias vector. Finally, the cross-entropy loss function is used as the optimization objective, as expressed in equation (12).

$$L = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \tag{12}$$

In equation (12), L means the total loss value, C means the number of categories, y_i represents the unique heat vector of the true label, and $\hat{y}_i$ means the prediction probability. The process of integrating I-TextRank and LSTM-Attention for NPO SA is shown in Figure 5.

In Figure 5, the emotion classification process mainly includes two core stages, namely sentence feature extraction and deep neural network classification. In the feature extraction stage, the input original sentence is first used to extract keywords through the I-TextRank algorithm. The original sentence and the extracted keywords are jointly input to the word embedding module and converted into a sequence of word vectors. Then, the word vector sequence is input into the LSTM network for sequence modeling, further capturing the contextual semantic relationships in the sentence and generating a complete sentence vector. Finally, the sentence vector is fed into a deep neural network classifier, which consists of a fully connected neural network structure with 256 input nodes and 128 hidden nodes, and outputs a classification result node to determine the emotional category.

3 Results

3.1 I-TextRank performance test

To verify the performance of I-TextRank, the Weibo Sentiment dataset was selected for experimental testing. This dataset was constructed by collecting public opinion data from Sina Weibo, a major Chinese microblogging platform. The data comes from popular topics and search events within two months, covering daily social discussions and emergency public events. The topic selection process involved keyword frequency analysis, real-time hot topic crawling, and manual filtering to ensure relevance and representativeness. In the data preprocessing stage, Jieba word segmentation tool was used for Chinese word segmentation, while removing stop words and noisy characters. The processed text was converted into Word2Vec word vector representation. In the emotional annotation process, the initial sentiment polarity annotation was first performed based on a rule-based sentiment dictionary, and then independently verified manually by three professional annotators to ensure the accuracy and consistency of the annotation results. For annotation cases with differences, the majority voting mechanism was used for final judgment. The final constructed Weibo sentiment dataset contained 5000 annotated samples, with a balanced distribution of positive and negative sentiment categories. The model parameter configuration is shown in Table 2.

Based on the parameter configuration in Table 2, to verify the contribution of each component of the G1 weighting method and model structure to the overall performance, a ablation experiment was designed to compare the performance of four keyword extraction strategies in sentiment classification tasks. The results are shown in Table 3.

Table 2: Hyperparameter settings.

Hyperparameter	Value
Input size	256
Hidden units	128
Output size	2
Batch size	32
Learning rate	0.001
Dropout rate	0.5
Iterations	300
Data set	Weibo Sentiment

Table 3: Results of ablation experiment.

Model variant	Accuracy (%)	F1 value (%)
TextRank (Baseline)	88.2	86.5
TextRank-TF-IDF	90.5	88.3
TextRank-Equal weights	91.2	89.0
I-TextRank	96.3	90.9

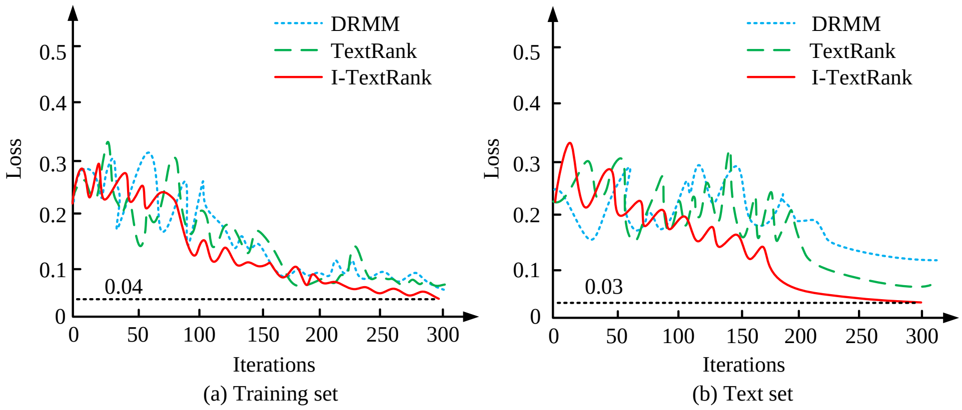

Figure 6: Loss function variation curve. (Source from: Author's self drawn)

From Table 3, there were significant differences in the performance of the four models in sentiment classification tasks. TextRank, as the basic model, had an accuracy of 88.2% and an F1 value of 86.5%, showing the worst performance. This indicates that without introducing any feature weighting mechanism, its keyword ranking results have limited support for sentiment discrimination. After introducing TF-IDF as the unique feature into the initial score, the performance of the TextRank TF-IDF model significantly improved, with an accuracy of 90.5% and an F1 value of 88.3%, verifying the positive role of word frequency information in keyword importance evaluation. On this basis, by further introducing language structure features such as part of speech and word position and assigning equal weights, the model performance was further improved to an accuracy of 91.2% and an F1 value of 89.0%, indicating that multi-feature fusion helps to improve the quality of keyword ranking. The final proposed I-TextRank model adopted the G1 weighting strategy for differentiated fusion of three types of features, achieving the highest accuracy of 96.3% and F1 value of 90.9%, significantly better than other models, fully demonstrating the significant effect of the G1 weighting mechanism in improving the semantic sensitivity of keyword recognition and optimizing sentiment classification performance. In the comparative experiment, with a maximum iteration of 300, the proposed model was compared and tested with traditional TextRank and Deviation Rule Markov Model (DRMM) [20]. The change in loss function is shown in Figure 6.

Figures 6 (a) and 6 (b) respectively show the curves of the loss functions of three algorithms on the dataset as a function of iteration times. In Figure 6 (a), as the number of iterations increased, the I-TextRank decreased the fastest and the curve was relatively stable. After the 200th iteration, it tended to stabilize and eventually dropped to the lowest value of about 0.04, significantly better than the other two models. Although DRMM and TextRank could also achieve a certain degree of loss reduction, their overall decline rate was slower, their fluctuations were greater, and their final convergence level was higher than I-TextRank, indicating poor fitting performance on the Levy function. In Figure 6 (b), I-TextRank also showed significant advantages. Although there were some fluctuations in the initial stage, compared to DRMM and TextRank, its convergence was smoother and faster. The final loss value of I-TextRank decreased to 0.03, while DRMM and TextRank still had significant fluctuations in the later stages of iteration, and the lowest loss value was still higher than I-TextRank, indicating weak generalization ability. The study used the Weibo Sentiment dataset, which was segmented into a training set and a testing set in an 8:2 ratio. The classification accuracy of the three models on the dataset was tested, and the outcomes are denoted in Figure 7.

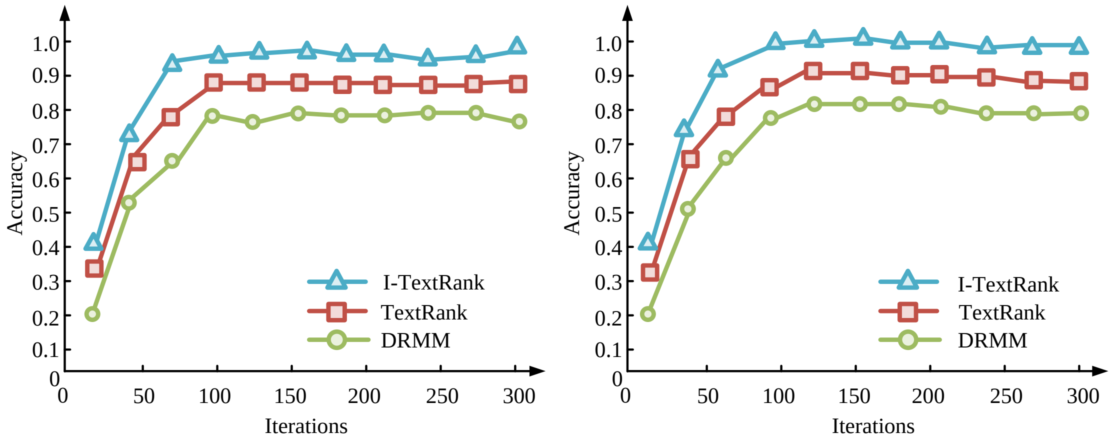

Figure 7: Classification accuracy of three models on datasets. (Source from: Author's self drawn)

Table 4: Multiple indicator test results.

Data set	Model	Precision/%	Recall/%	F1/%
Training dataset	DRMM	77.3	79.1	78.7
	TextRank	86.5	85.5	84.2
	I-TextRank	93.4	91.9	92.6
Test dataset	DRMM	79.8	77.9	78.8
	TextRank	88.1	87.1	86.5
	I-TextRank	91.7	91.1	90.9

Figures 7 (a) and 7 (b) respectively show the trends of the accuracy of the three models on the training and testing sets as a function of the number of iterations. Overall, the I-TextRank model performed better than TextRank and DRMM on both datasets, demonstrating its stronger fitting ability and better generalization performance. In Figure 7 (a), all three models had low accuracy in the initial stage. The I-TextRank quickly increased to 0.75 after the 50th iteration, reached above 0.95 in the 100th iteration, and remained at 0.96 thereafter. The accuracy of the TextRank model remained stable at 0.88, with a slightly slower convergence speed but still acceptable stability. The DRMM model showed the smallest improvement, with an accuracy rate of around 0.79 after the 100th round and slight fluctuations in the later stages, indicating its limited ability to fit the training set. In Figure 7 (b), the accuracy of I-TextRank remained stable at 0.97 after the 100th round, indicating that the model did not exhibit significant overfitting and had strong generalization ability. The accuracy of the TextRank model on the test set was slightly lower than that on the training set, at 0.82, which was almost consistent with the trend of the training set. However, the overall accuracy was low, further verifying its shortcomings in extracting key emotional features. The study conducted another comparison using precision, recall, and F1 value as indicators, and the test findings are denoted in Table 4.

According to Table 4, on the training set, the precision of I-TextRank reached 93.4%, the recall rate was 91.9%, and the F1 value was 92.6%, significantly higher than TextRank and DRMM. This indicated that I-TextRank could better capture emotional key features during the model learning stage, improving the accuracy and stability of classification. On the test set, I-TextRank also performed well, with an F1 value of 90.9%, far higher than TextRank's 86.5% and DRMM's 78.8%. In addition, although TextRank performed better than the training set on the test set, it was still significantly lower than I-TextRank, indicating that I-TextRank not only has strong fitting ability in the training stage, but also has stronger generalization ability and robustness. Overall, I-TextRank outperformed the comparison model in precision, coverage, and overall performance, indicating that the strategy of introducing multidimensional weights and G1 weighting to improve the initial node score can effectively enhance the semantic sensitivity of keyword extraction and sentiment discrimination, and is suitable for NPO SA tasks.

3.2 Application effect of NPO sentiment analysis model integrating I-TextRank

After conducting performance tests on I-TextRank, the study used four different fields of public opinion hotspots, namely AI fraud, college entrance examination reform, short drama money grabbing chaos, and US-China relations. The raw online data for each topic was collected through Sina Weibo, news portals, and forum discussions. The data has undergone cleaning, duplicate data removal, and sentiment annotation. For each hotspot, approximately 2000 samples were compiled and manually labeled as positive or negative emotions through a semi-automatic process.

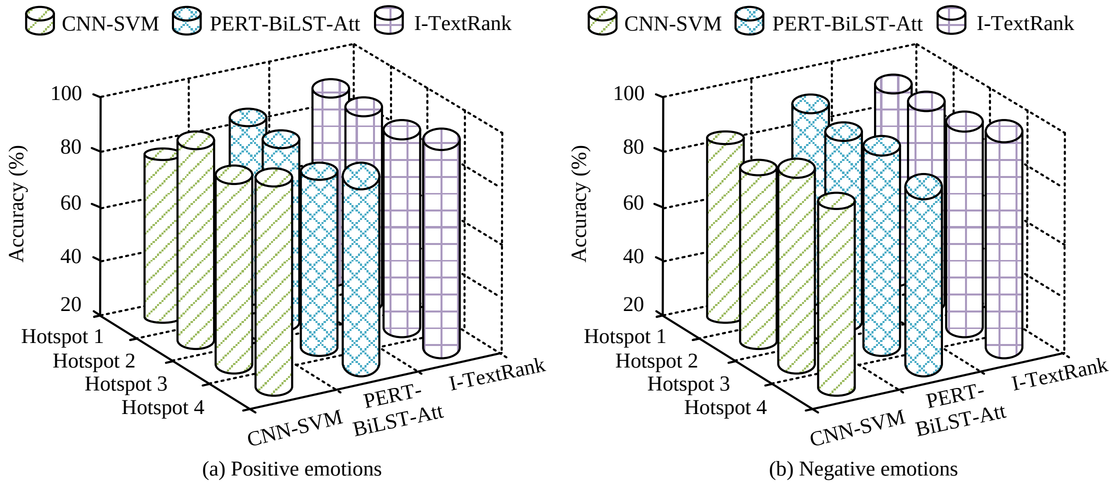

Figure 8: Classification accuracy results under different hotspots. (Source from: Author's self drawn)

Table 5: Classification error results under different hotpots.

Hot topics in public opinion	Model	RMSE	MAPE	R^2
Hotpot 1	CNN-SVM	0.215	0.123	0.892
	PERT-BiLST-Att	0.195	0.105	0.912
	I-TextRank	0.176	0.083	0.932
Hotpot 2	CNN-SVM	0.221	0.135	0.885
	PERT-BiLST-Att	0.205	0.119	0.901
	I-TextRank	0.175	0.079	0.926
Hotpot 3	CNN-SVM	0.238	0.151	0.878
	PERT-BiLST-Att	0.211	0.122	0.909
	I-TextRank	0.192	0.085	0.919
Hotpot 4	CNN-SVM	0.231	0.148	0.874
	PERT-BiLST-Att	0.205	0.113	0.911
	I-TextRank	0.185	0.081	0.921

The emotional category analysis ability of the four models was further validated through network data collection and processing. The NPO SA model based on I-TextRank proposed by the research was compared and analyzed with the mixed Convolutional Neural Network and Support Vector Machine (CNN-SVM) model [21], as well as the SA model that integrates Pretrained Embedding-Bidirectional Long Short-Term Memory-Attention (PERT-BiLST-Att) [22]. AI fraud, college entrance examination reform, short drama money circle chaos, and China-US relations are recorded as hotspot 1~hotspot 4 respectively, and the classification accuracy is shown in Figure 8.

Figures 8 (a) and 8 (b) show the ROC curves of three models on four different public opinion hotspots, respectively. Performance evaluations were conducted on each hotspot, and the classification performance of the models was quantified using AUC. In Figure 8 (a), the I-TextRank model consistently outperformed the other two models in the four public opinion hotspots, especially in the classification of positive emotions, with an accuracy rate of almost 100%. On the four hotspots, the positive emotion classification accuracy of I-TextRank was 98%, 96%, 95%, and 94%, respectively. PERT-BiLST-Att performed relatively stable on these hotspots, with an accuracy rate of around 90% for positive emotion classification. In Figure 8 (b), the accuracy of the I-TextRank model in classifying negative emotions in four public opinion hotspots was 96%, 95%, 93%, and 92%, respectively. The accuracy of PERT-BiLST-Att's negative emotion classification remained above 80%, demonstrating its relative advantage in emotion classification. However, the performance of CNN-SVM was relatively lagging behind, with significantly lower classification accuracy for both positive and negative emotions compared to I-TextRank and PERT-BiLST-Att. Especially in negative emotion classification, its accuracy was relatively low. The study selected Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and Fit Coefficient R^2 as evaluation metrics to compare the error results of different models. The findings are denoted in Table 5.

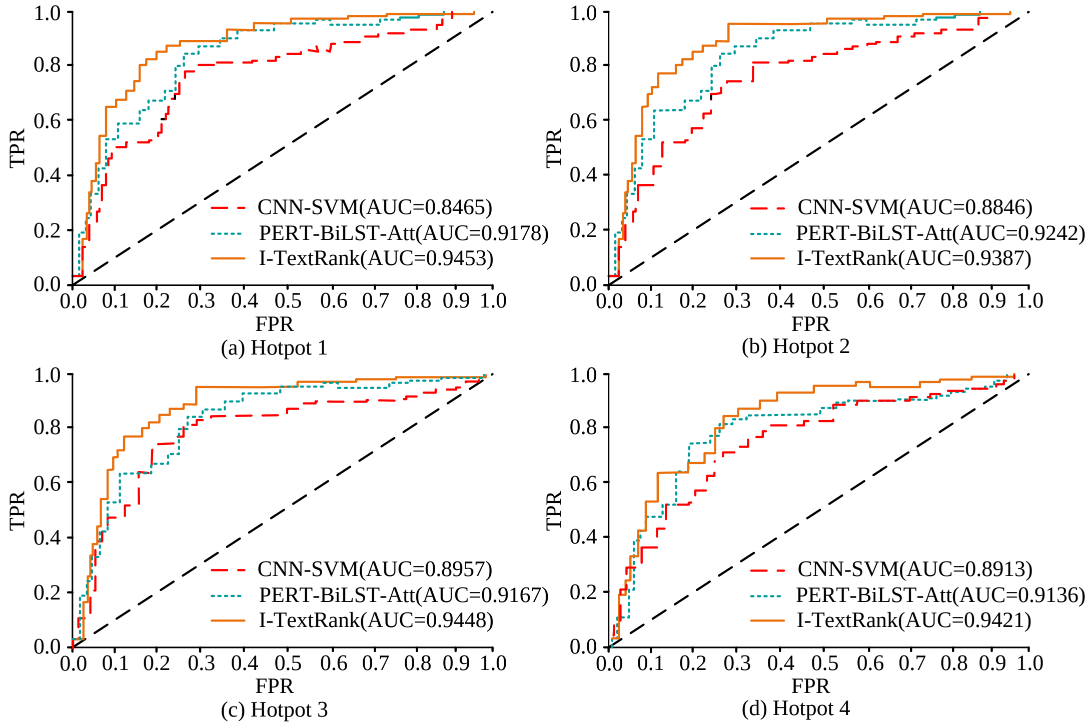

Figure 9: ROC curves of different models under different hotpots. (Source from: Author's self drawn)

Table 6: Cross-validation performance.

Fold	Accuracy (%)	F1 value (%)	AUC
1	96.0	90.7	0.9335
2	95.6	90.2	0.9361
3	95.8	90.4	0.9378
4	96.2	91.0	0.9354
5	95.6	90.2	0.9382
Average value	95.8	90.5	0.9362
Standard deviation	0.24	0.29	0.0017

From Table 5, the I-TextRank model had the best error performance in all four hotspots, consistently showing the lowest RMSE and MAPE, as well as the highest R^2 value, indicating that the model had strong fitting and generalization abilities in sentiment classification tasks. Among them, on hotspot 1, the RMSE of I-TextRank was 0.176, MAPE was 0.083, and R^2 was 0.932, all of which were better than the other two models. PERT-BiLST-Att closely followed, with three indicators of 0.195, 0.105, and 0.912, while CNN-SVM had weaker performance, with with three indicators of 0.215, 0.123, and 0.892. On Hotspot 2, I-TextRank also demonstrated strong performance, with with three indicators of 0.175, 0.079, and 0.926. The performance of PERT-BiLST-Att was relatively stable, with with three indicators of 0.195, 0.105, and 0.912. The three indicators of CNN-SVM were 0.220, 0.119, and 0.885, indicating relatively low performance. On Hotspot 3 and Hotspot 4, I-TextRank maintained the lowest RMSE and MAPE, while R^2 had the highest, at 0.919 and 0.921 respectively, demonstrating its powerful ability in these complex SA tasks. In contrast, CNN-SVM and PERT-BiLST-Att performed poorly. The Area Under ROC Curve (AUC) results obtained from testing on four hot topics are shown in Figure 9.

Figures 9 (a), 9 (b), 9 (c), and 9 (d) show the ROC curves of three models on four different public opinion hotspots. Performance evaluations were conducted on each hotspot, and the classification performance of the models was quantified by Area Under the Curve (AUC). In Figure 9 (a), the I-TextRank model performed the most outstandingly, with an AUC value of 0.9453, far exceeding the other two models, demonstrating its superior performance in handling this public opinion hotspot. The AUC values of PERT-BiLST-Att and CNN-SVM were 0.9178 and 0.8465, respectively, indicating a certain gap compared to I-TextRank. In Figure 9 (b), I-TextRank still performed the best with an AUC of 0.9387. The AUC value of PERT-BiLST-Att was 0.9242, while the performance of CNN-SVM was still low, with an AUC value of 0.8846. The curves of I-TextRank and PERT-BiLST-Att showed a significant difference in the false positive rate range, further demonstrating the excellent performance of I-TextRank in this hotspot. In Figures 9 (c) and 9 (d), I-TextRank consistently demonstrated strong performance, with AUC values of 0.9444 and 0.9421, respectively, consistently at its optimal position. The AUC values of PERT-BiLST-Att were 0.9167 and 0.9136 in hotspot 3 and hotspot 4, respectively, maintaining a relatively stable performance. The AUC value of CNN-

SVM was the lowest, with AUC values of 0.8957 and 0.8913 for hotspot 3 and hotspot 4, respectively, indicating its weaker performance on these hotspots. From this, it can be seen that the I-TextRank curve is almost entirely above the other two curves, indicating that it can better distinguish between positive and negative samples. To avoid overfitting of the model and verify its generalization ability under different data partitions, a five-fold cross validation experiment was conducted on the dataset, and the results are shown in Table 6.

From the results in Table 6, the I-TextRank model performed stably in various performance indicators in the five-fold cross validation, with minimal fluctuations and good generalization ability and robustness. The accuracy fluctuated between 95.6% and 96.2%, with a mean of 95.8% and a standard deviation of only 0.24%, indicating that the model has very little difference in classification performance under different training test partitions. The average F1 value was 90.5%, with a standard deviation of 0.29%, indicating that the model's ability to distinguish positive and negative emotions remains stable. The AUC value remained above 0.9335 in all compromises, with the highest reaching 0.9382 and an average of 0.9362, with a standard deviation of only 0.0017, further demonstrating the model's strong discriminative ability on different subsets. The overall results indicate that the model does not have overfitting issues for a certain data partition, and its performance is not accidentally high, but has stability and universality at the structural level. Therefore, the proposed feature fusion and weighting mechanism is effective and reliable in sentiment classification tasks.

4 Conclusion

An SA model that integrates I-TextRank and LSTM-Attention was proposed to address the limitations of existing SA methods in keyword extraction and sentiment classification accuracy. By combining the advantages of I-TextRank in keyword extraction stage with the contextual modeling ability of LSTM-Attention model, the performance of sentiment feature extraction and classification was effectively enhanced. The performance test results of I-TextRank showed that its accuracy on the test set was 0.96, and its F1 value was as high as 90.9%. From this, I-TextRank outperformed the comparison model in terms of iterative convergence speed, training fitting ability, and testing generalization performance, demonstrating the advantages of this model in NPO SA tasks. When conducting SA on four public opinion hotspots, namely AI fraud, college entrance examination reform, short drama money grabbing chaos, and US-China relations, the accuracy of this model was the best among all tasks. It performed particularly well in the classification of positive and negative emotions, with the highest accuracy of positive and negative emotion classification in AI fraud, at 98% and 96% respectively. In terms of AUC values, this model outperformed the other two models, with the highest AUC value of 0.9448 in the hot topic of short drama money making chaos, demonstrating the strong advantage of this model in handling complex public opinion data. The results demonstrated that the proposed model had significant merits in improving the semantic sensitivity of keyword extraction and sentiment classification, and could effectively enhance the accuracy and stability of public opinion SA tasks. There are also certain limitations in the research. The I-TextRank algorithm relies heavily on the keyword extraction process, and for some texts with subtle or complex emotional expressions, there may still be insufficient accuracy in extraction. Future work could attempt to introduce cross domain transfer mechanisms to enable models to adapt to emotional distribution differences across different themes, contexts, and social platforms, enhancing their cross-scenario robustness. Second, considering extensions to multilingual text processing scenarios, especially for resource-poor languages, model applicability is enhanced through multilingual embedding or cross-language transfer learning. At the same time, multimodal data is further integrated to enhance the model's comprehensive perception ability of emotional signals and improve the recognition effect of complex semantics, ironic metaphors, and other emotional forms.

Funding

The research is supported by Jiangsu Province “14th Five-Year Plan” Business Administration Key Construction Discipline Project (Su Jiaoyanhan [2022] No. 2/Sequence 285), Nantong Institute of Technology Business School Zhongzhi Scientific Research Team Project (NSKT2025-01).

References

[1] Jiahui Wang, Kun Yue, and Liang Duan. Models and techniques for domain relation extraction: A survey. Journal of Data Science and Intelligent Systems, 3(1):16-25, 2023. https://doi.org/10.47852/bonviewJDSIS3202973

[2] Xuegang Chen, Sheng Duan, Shanglin Li, Dong Liu, and Hongbin Fan. A method of network public opinion prediction based on the model of grey forecasting and hybrid fuzzy neural network. Neural Computing and Applications, 35(35):24681-24700, 2023. https://doi.org/10.1007/s00521-023-08205-9

[3] Qingqing Li, Ziming Zeng, Shouqiang Sun, Chen Cheng, and Yingqi Zeng. Constructing a spatiotemporal situational awareness framework to sense the dynamic evolution of online public opinion on social media. The Electronic Library, 41(5):722-749, 2023. https://doi.org/10.1108/EL-05-2023-0134

[4] Liwei Xu, Jiangnan Qiu, and Jie Zhai. Trend prediction model of online public opinion in emergencies based on fluctuation analysis. Natural Hazards, 116(3):3301-3320, 2023. https://doi.org/10.1007/s11069-022-05808-8

[5] Zhengzhi Xu, Zi Ye, Haiyang Ye, Lijia Zhu, Ke Lu, Hong Quan, Jun Wang, Shanchuan Gu, Shangfeng Zhang, and Guodao Zhang. Public opinion evolution law and sentiment analysis of campus online public opinion events. Journal of Advanced Computational

Intelligence and Intelligent Informatics, 28(4):990-1004, 2024. https://doi.org/10.20965/jaciii.2024.p0990
[6] Zeguo Qiu, and Baiyan He. Research on the evolution of public opinion and topic recognition based on multi-source data mining. International Journal of Computer Applications in Technology, 69(3):219-227, 2022. https://doi.org/10.1504/ijcat.2022.127816
[7] Shackleford Matthew Brett, Adeliyi Timothy, and Joseph Seena. A prediction of South African public Twitter opinion using a hybrid sentiment analysis approach. Science and Information Organization, 14(10):156-165, 2023. https://doi.org/10.14569/IJACSA.2023.0141017
[8] Yan Jiang Author, Chunlin Xiang, and Lingtong Li. Keyword acquisition for language composition based on TextRank automatic summarization approach. International Journal of Advanced Computer Science & Applications, 15(4):994-1005, 2024. https://doi.org/10.14569/IJACSA.2024.01504101
[9] Blessed Guda, Bello Kontagora Nuhu, James Agajo, and Ibrahim Aliyu. Performance evaluation of keyword extraction techniques and stop word lists on speech-to-text corpus. The International Arab Journal of Information Technology, 20(1):134-140, 2023. https://doi.org/10.34028/iajit/20/1/14
[10] Yonghe Lu, Meilu Yuan, Jiaxin Liu, and Minghong Chen. Research on semantic representation and citation recommendation of scientific papers with multiple semantics fusion. Scientometrics, 128(2):1367-1393, 2023. https://doi.org/10.1007/s11192-022-04566-5
[11] Zhili Wu, and Qian Zhang. A deep learning-based method for determining semantic similarity of english translation keywords. International Journal of Advanced Computer Science & Applications, 15(5):303-313, 2024. https://doi.org/10.14569/IJACSA.2024.0150531
[12] Jinhong Li, Xuejie Zhang, Jin Wang, and Xiaobing Zhou. Deep question generation model based on dual attention guidance. International Journal of Machine Learning and Cybernetics, 15(11):5427-5437, 2024. https://doi.org/10.1007/s13042-024-02249-6
[13] Yan Jiang, Chunlin Xiang, and Lingtong Li. Keyword acquisition for language composition based on TextRank automatic summarization approach. International Journal of Advanced Computer Science & Applications, 15(4):994-1005, 2024. https://doi.org/10.14569/IJACSA.2024.01504101
[14] Chengzhi Zhang, Lei Zhao, Mengyuan Zhao, and Yingyi Zhang. Enhancing keyphrase extraction from academic articles with their reference information. Scientometrics, 127(2):703-731, 2022. https://doi.org/10.1007/s11192-021-04230-4
[15] Qian Zhou, Hua Dai, Yuanlong Liu, Geng Yang, Xun Yi, and Zheng Hu. A novel semantic-aware search scheme based on BCI-tree index over encrypted cloud data. World Wide Web, 26(5),3055-3079, 2023. https://doi.org/10.1007/s11280-023-01176-w
[16] Xiang Chen, Xing Wang, Hubiao Zhang, Yuheng Xu, You Chen, and Xiaotian Wu. Interval TOPSIS with a novel interval number comprehensive weight for threat evaluation on uncertain information. Journal of Intelligent & Fuzzy Systems, 42(4):4241-4257, 2022. https://doi.org/10.3233/JIFS-210945
[17] Chenyu Wang, Yanjun Ye, Yingqiao Qiu, Chen Li, and Meiqing Du. Evolution and spatiotemporal analysis of earthquake public opinion based on social media data. Earthquake Science, 37(5):387-406, 2024. https://doi.org/10.1016/j.eqs.2024.06.002
[18] Jiahao Wen, and Zhijian Wang. Short-term load forecasting with bidirectional LSTM-attention based on the sparrow search optimisation algorithm. International Journal of Computational Science and Engineering, 26(1):20-27, 2023. https://doi.org/10.1504/ijcse.2023.129154
[19] Haifeng Yang, Juanjuan Hu, Jianghui Cai, Yupeng Wang, Xin Chen, Xujun Zhao, Lili Wang. A new mc-lstm network structure designed for regression prediction of time series. Neural Processing Letters, 55(7):8957-8979, 2023. https://doi.org/10.1007/s11063-023-11187-3
[20] Wei Shi, Guangcong Xue, Xicheng Yin, Shaoyi He, and Hongwei Wang. DRMM: A novel data mining-based emotion transfer detecting method for emotion prediction of social media. Journal of Information Science, 50(3):590-606, 2024. https://doi.org/10.1177/01655515221100728
[21] Jiawen Li, Yuesheng Huang, Yayi Lu, Leijun Wang, Yongqi Ren, and Rongjun Chen. Sentiment analysis using e-commerce review keyword-generated image with a hybrid machine learning-based model. Computers, Materials & Continua, 2024, 80(1):1581-1599, 2024 https://doi.org/10.32604/cmc.2024.052666.
[22] Mingyong Li, Zheng Jiang, Zongwei Zhao, and Longfei Ma. A PERT-BiLSTM-Att model for online public opinion text sentiment analysis. Intelligent Automation & Soft Computing, 37(2):2387-2406, 2023. https://doi.org/10.32604/iasc.2023.037900