IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received January 10, 2021, accepted February 3, 2021, date of publication February 8, 2021, date of current version September 16, 2021.
Digital Object Identifier 10.1109/ACCESS.2021.3058016

Bi-Level Attention Model With Topic Information for Classification

HONGTAO LIU AND QIMIN QIAN
Chongqing University of Posts and Telecommunications, Chongqing 400000, China
Corresponding author: Qimin Qian (s180231887@stu.cqupt.edu.cn)

This work was supported in part by the National Social Science Fund of China under Grant 18BGL266.

ABSTRACT This paper proposes a bi-level attention neural network model (TBAM) that incorporates topic information. This model is suitable for a variety of classification situations and has proven effectiveness in sentiment classification and text classification. Compared with other advanced neural networks that are also used in classification, it has achieved better classification results. The TBAM model has three characteristics: (1) The model reflects the hierarchical structure of the document; (2) The model uses the attention mechanism at the vocabulary and sentence levels respectively, so that it can pay attention to the more important and the less important when constructing the document representation. content. (3) In the vocabulary level, by introducing potential topic information into the semantic representation of the vocabulary level to improve the effect of vocabulary representation, the topic information is regarded as domain information, which makes up for the ignorance of the domain information by the classification model, and increases the model domain adaptability. Through testing on the public data sets CCF-BDCI and THUCNews, the model shows good results on the three indicators of Precision, Recall, and F1.

INDEX TERMS Topic information, hierarchical attention, classification, domain adaptability.

I. INTRODUCTION

Behaviors such as information dissemination and opinion expression are increasingly appearing in various social network media, emerging news platforms, e-commerce and other fields [1]. So every day there will be explosive information flowing into various channels, and how to extract useful information from these information is now the focus of natural language processing research. For example, the application of sentiment classification is conducive to understanding people's views and attitudes on a certain thing or a certain product in time, which is very important for personalized search, public opinion analysis, customer service, etc., [2], [3]. The application of text classification can automatically classify a large amount of information, which is convenient for people to organize and read. It is ubiquitous in applications such as automatic summarization, news classification, and information retrieval [5].

Many researches in the field of natural language processing will be combined with sentiment classification and text classification. For example, the recommended direction will be combined with the sentiment classification to make a more accurate judgment on the user's favorite; Information search direction, text classification is an extremely powerful aid to help push more accurate information. Therefore, improving the accuracy of classification is a difficult problem in the field of natural language processing. The current methods for classification mainly fall into three categories: methods based on language rules, methods based on traditional machine learning, and methods based on deep neural networks. However, with the increase in the volume of text and the diversification of language representation, the advantages of the method based on the deep network model have gradually become apparent. Compared with the previous two types of methods, the method based on the deep network model can better understand the text when dealing with complex and changeable language models, and can capture more comprehensive and deeper text features. The analysis field has achieved very good results, Therefore, this paper chose the network model for classification. However, due to the neglect of topic information, this kind of method still lacks the ability to capture the semantic information of text. Many deep affective analysis models fail to consider the obvious fact that a document has a hierarchical structure (word source sentence, sentence source

The associate editor coordinating the review of this manuscript and approving it for publication was Sunil Karamchandani.

document), and different words and sentences have different importance in the document [5].

Considering the above problems and inspired by Yang *et al.* [5], this research uses a hierarchical neural network that conforms to the characteristics of the document structure, and adds an attention mechanism to apply different weights to different words and sentences. At the same time, the topic model is applied at the lexical level to capture the deep semantic information of the text and increase the generalization ability of the model.

In summary, the main contributions of this research are as follows:

- According to the structural characteristics of the document, a hierarchical neural network with attention mechanism is proposed to obtain the vector representation of the document.
- Propose a method to integrate topic information into the neural network model to expand the semantic representation of the word layer from the topic perspective.
- Through testing on the public data sets CCF-BDCI and THUCNews, the effectiveness of the model in classification is verified from the three indicators of Precision, Recall and F1.

The remainder of this paper is organized as follows. session 2 reviews the research related to existing classification methods and topic models. session 3 describes the TBAM model proposed in this research. session 4 introduces the experimental design in detail and analyzes the experimental results. session 5 summarizes the work of this research.

II. RELATED WORKS

Classification is the cornerstone of many research problems in the field of natural language processing. How to improve the classification performance is the focus of our research. It has been analyzed that the use of neural network is the mainstream method in today's research. Now how to combine event scenarios (fields) to provide better word representation for the model and improve the performance of neural network classification is a problem we need to consider. This research attempts to apply topic models at the vocabulary level to capture the deep semantic information of the text, improve the classification performance of neural networks, and enhance the generalization ability of the model. Therefore, this research is carried out from two aspects, namely the application of neural network model in classification and topic model.

A. NETWORK MODEL FOR CLASSIFICATION

The neural network model can capture the deep features and hidden patterns of the text, so that the model can better understand the text and improve the classification performance.

Wan [6] proposed a deep neural network-based sentiment analysis method for microblog comments by constructing a multi-layer pooling layer and a convolutional layer, the network complexity is increased, data features are effectively extracted, and the accuracy of sentiment analysis is improved. Zhao *et al.* [7] proposed a parallel network of bidirectional LSTM (Bi-LSTM) and convolutional neural network (CNN), and used the concatenation of the output vectors of the two networks as the expression of the final document to improve the accuracy of sentiment classification. Kim [9] proposed the use of convolutional neural network for sentence modeling, which is applied in sentiment classification tasks and achieved good results in multiple data sets. Zhou *et al.* [10] proposed a hierarchical attention-based fuzzy neural network (HABFNN). The model uses convolutional neural (CNN) to model sentences based on word vectors at the word sentence layer; at the sentence document layer, it uses CNN and bidirectional long-term short-term memory (Bi-LSTM) network modeling is based on model documents represented by sentences. Cao *et al.* [11] proposed a decomposing convolutional neural network model based on the attention mechanism, which combines the attention mechanism to extract important feature information and improve and optimize the effect of text sentiment classification. Zhu *et al.* [12] built a new framework through convolutional neural network to significantly improve the classification accuracy. Lu *et al.* [13] proposed a fusion neural network model, which captures different levels of text features through convolutional layers, Bi-GRU layers and Attention layers to improve classification accuracy. Basnet and Timalsina [14] proposed using a recurrent neural network to classify Nepalese news into appropriate categories, and through testing on eight different categories of news data sets, very competitive results were obtained. Zeng *et al.* [15] merged the results of multiple convolution kernels, and then pooled the merged results to construct three CNN models. Experiments show that the modified model is better than the traditional text classification method. Alfian *et al.* [16] proposed a classification system based on real-time data and machine learning to help patients with diabetes do better self-management. the BLE-based sensors and the proposed real-time data processing are sufficiently efficient to monitor the vital signs data of diabetic patients. Furthermore, machine learning–based classification methods were tested on a diabetes dataset and showed that a Multilayer Perceptron can provide early prediction of diabetes given the user's sensor data as input. Ijaz *et al.* [17] proposed a machine learning based classification system (CCPM) for cervical cancer prediction, The CCPM first removes outliers by using outlier detection methods such as density-based spatial clustering of applications with noise (DBSCAN) and isolation forest (I Forest) and by increasing the number of cases in the dataset in a balanced way, for example, through synthetic minority over-sampling technique (SMOTE) and SMOTE with Tomek link (SMOTE Tomek). Finally, it employs random forest (RF) as a classifier. Ijaz *et al.* [18] proposed a machine learning based hybrid classification algorithm (HPM) for early prediction of diabetes and hypertension, the proposed HPM consists of Density-based Spatial Clustering of Applications with Noise (DBSCAN)-based outlier detection to remove the outlier data, Synthetic Minority Over-Sampling Technique (SMOTE) to balance the distribution of class, and Random Forest (RF) to classify the diseases. Three benchmark datasets were utilized to predict and achieved good results. Ma [19] proposed to add a superimposed attention mechanism composed of target level and sentence level attention models to LSTM, which is called perceptual LSTM. It pays special attention to the use of common sense knowledge in deep neural sequence models to improve model emotion classification. Kalchbrenner *et al.* [20] described a dynamic convolutional neural network (DCNN) convolutional architecture and used it for semantic modeling of sentences, in small-scale binary and multi-category emotion prediction, six-way problem classification, and Twitter's sentiment prediction has achieved good results. Pan *et al.* [21] studied the implicit emotion classification model based on deep neural network. A classification model based on DNN, LSTM, Bi-LSTM and CNN is established to judge the trend of users' implicit emotion text. Wang and Zhang [22] proposed a convolutional neural network (CNN) text vector classification model based on a mixture of word vectors and characters. Semberecki and Maciejewski [23] proposed a method for text document classification using deep neural networks with LSTM (long-term and long-term memory) units.

The classification effect of the models proposed in reference [6]–[23] is obviously better than that of the traditional machine learning method. However, these methods ignore the hierarchical structure of the text itself and the topic information it carries.

B. TOPIC MODEL

Topic model is one of the popular methods to learn text representation. The LSA algorithm proposed by Deerwester *et al.* [24] is to reduce the dimension of words in documents by using singular value decomposition (SVD), and map words to potential semantic space. LSA can remove noise and capture synonyms which often appear in the same document, but it can not solve the problem of polysemy. Hofmann [25] proposed probabilistic latent semantic analysis (PLSA). PLSA regards documents and words as random variables and introduces them into potential topic level. PLSA solves the problem of polysemy and multiple near meanings of a word, but over fitting may occur. Blei [26] proposed the Latent Dirichlet Allocation (LDA). LDA model introduced the potential dirichlet distribution of topic and vocabulary, and solved the over fitting problem of PLSA. Its parameters did not increase with the increase of training set, and had good generalization ability. Haidar *et al.* [27] used LDA model to form document based distribution on the topic of each word, and applied it to the recurrent neural network language model. Syamala and Nalini [28] introduced a method to extract the most important information from the views expressed in the input text by combining various machine learning and deep learning techniques with LDA. On the basis of LDA topic model, Niu *et al.* [29] proposed a text classification algorithm by using neural network fitting word topic probability distribution. Neural topic models (NTM) proposed by Miao *et al.* [30] and Srivastava and Sutton [31] are topic models parameterized by deep neural networks. Neural topic model uses the excellent ability of deep neural network as function approximator to learn the complex semantic association between potential topic and word meaning, and achieves good results.

III. TBAM

The structure of the TBAM model proposed in this research is shown in Figure 1. The model consists of four modules: (A) LDA, (B) word level topic information attention mechanism (W-TIAM), (C) sequence level attention mechanism (SLAM), and (D) classifier

At the word level, our model integrates the topic information extracted by LDA to improve the expression effect of the word level, and uses the attention mechanism at the word level and sequence level, so that the model can be more deeply learn the semantic information and domain information of the document context, and improve the classification performance and domain adaptability of the model.

LDA model extracts topic information based on text dictionary. The dictionary of text is represented as $x_{BoW} = R^{|V|}$ and LDA is expressed in the form of word bag, where $|V|$ is the vocabulary. LDA analyzes the potential topics and generates the embedded matrix W (topic information) of the potential topics, and the topic information is passed to the W-TIAM module to be used to expand the word-level representation. The text is initially processed as a sentence sequence and as a sentence unit is processed as a word sequence, expressed as $x_{seq} \in R^T$, where T represents the length of the word sequence. After that, the word sequence is expressed as a word embedding matrix U through the Word2vec tool, and passed into the bidirectional recurrent neural network to obtain the hidden representation matrix H of the vocabulary. W-TIAM obtains the topic-word embedding Q according to the relationship α between the topic information W and the hidden representation of the vocabulary H, then concatenate the word embedding U and the topic-word embedding Q to obtain the word-level semantic representation K, and the sequence-level attention mechanism is applied to the semantic representation of the sentence containing the word-level topic information, thereby achieving the semantic representation S at the semantic level. Finally, based on the semantic representation of the text, the classifier is used to infer the label of the text.

A. LDA

The topic model LDA assumes that there are D documents in the document set, and all documents have K topics and V words (not repeated). After all documents are input, the LDA algorithm will get the probability distribution of each document belonging to these K topics. And the probability distribution of V words under each topic [23], [32].

LDA has strong topic mining ability and is an unsupervised model. It is considered that it does not rely on training samples and has no domain transfer problem, so it has good

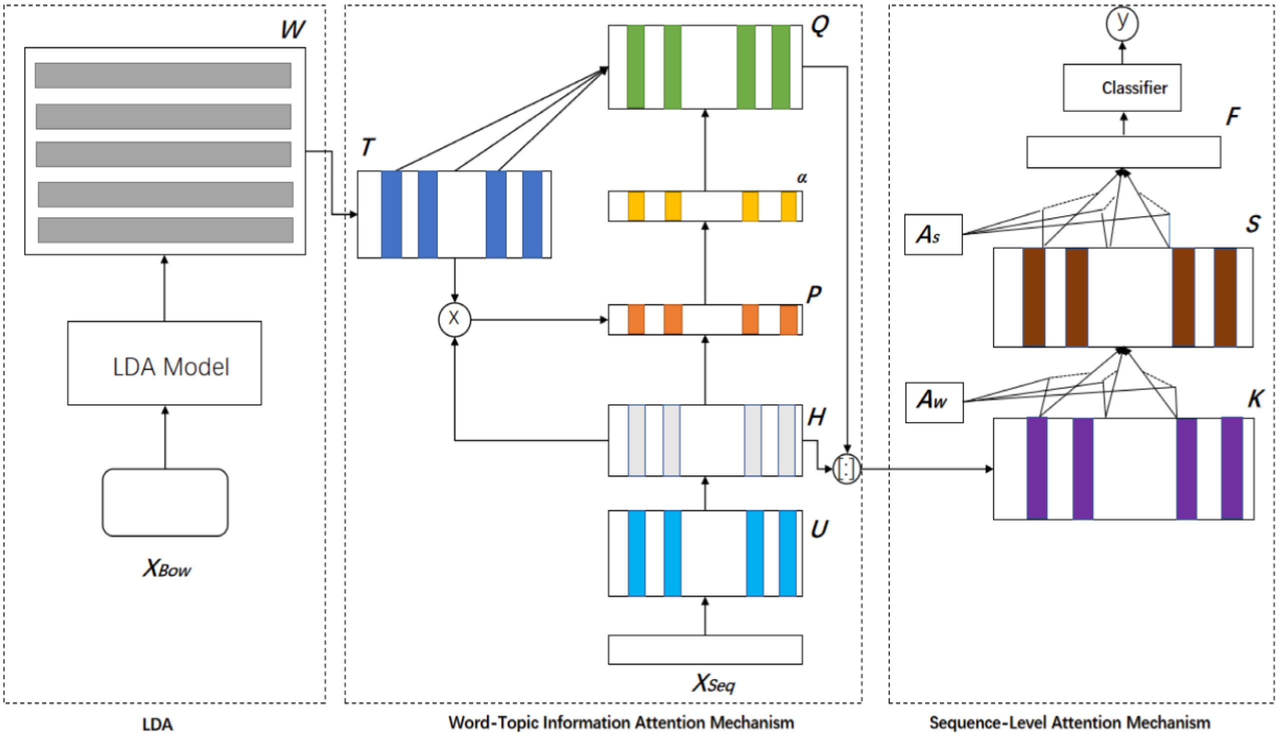

FIGURE 1. TBAM architecture.

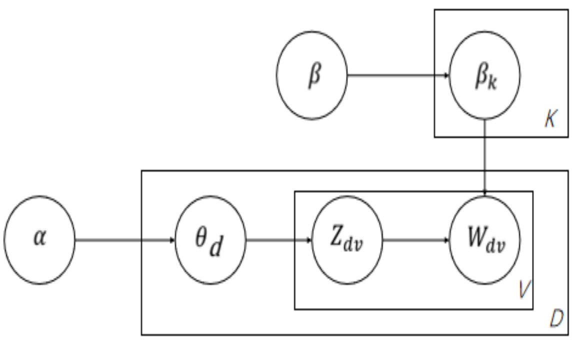

FIGURE 2. LDA.

domain adaptability. LDA model is a complete Bayesian probability graph model, and the inference of parameters needs to infer the posterior distribution of parameters, so Gibbs sampling algorithm is used to estimate model parameters [34]. As shown in Figure 2 is the schematic diagram of LDA probability graph model, which depicts the generation process of the whole long text data set [33]:

1) For each document $d = 1 \ldots D$:

   Sample a document-topic distribution $\theta_d \sim Dir(\alpha)$;

2) For each topic $k = 1 \ldots K$:

   Sample a topic-word distribution $\beta_k \sim Dir(\eta)$;

3) For each word in the text $w = 1 \ldots V$:

   3.1) Sample a topic tag $Z_{dv} \sim Mult(\theta_d)$

   3.2) Sample a word under the topic vocabulary distribution of topic tags $W_{dv} \sim Mult(\beta_{Z_{dv}})$

where $Mult(\bullet)$ is the polynomial distribution, $Dir(\bullet)$ is the Dirichlet distribution, θ_d is the topic distribution of any document, α is the super parameter of the distribution, and is a K dimensional vector, K represents the number of topics. β_k is the lexical distribution of any topic, η is the super parameter of the distribution, and is a V dimensional vector, V represents the number of no repetitive words in all documents in the data set.θ_d and β_k are parameters to be learned by the model. Z_{dv} is a hidden variable, which represents the topic of the nth word in any document in the data set. W_{dv} is an observable variable, $\beta_{Z_{dv}}$ represents the distribution of topic-vocabulary under the topic Z_{dv}.

Then, the topic-word distribution matrix W extracted from LDA topic model is transformed into a low dimensional topic embedding matrix T through the full connection layer, thus obtaining the same dimension as the lexical hiding state vector.

$$t_k = \tanh(W\beta_k) \tag{1}$$

t_k is low dimensional topic embedding, β_k is topic vocabulary distribution, W is parameter matrix, tanh is nonlinear function.

B. WORD LEVEL TOPIC INFORMATION ATTENTION MECHANISM (W-TIAM)

In this study, we propose a word level topic information attention mechanism, which introduces the topic information into the word representation, so as to improve the word level semantic information.

The topic vector matrix W provided by the LDA topic model and the word semantic information representation matrix H are used to calculate the relationship α between the topic and the word, and the topic vector matrix T is reconstructed by α to obtain the word topic embedding Q. After that, we connect the word hidden embedding H and the word topic embedding Q to obtain the word-level semantic representation K. The process is briefly explained as follows:

- Get The Word Semantic Information Representation Matrix H

First, divide the document into a sequence of sentences with punctuation marks such as period and question mark as separators, d $= (s_1 \ldots s_l \ldots s_L)$. and then divide each sentence in the sequence of sentences, remove stop words, etc., and process them into a sequence of words, $s_l = (w_{l1} \ldots w_{lt} \ldots w_{lT})$. second, use Word2vec to express all the words in the sentence as word vectors, and input the word vector of the sentence into the two-way recurrent neural network to obtain the hidden state vector matrix H of the words.

- Reconstruct The Topic Vector Matrix T To Obtain The Word Topic Embedding Q

The relationship between the topic information and the hidden state of the word is calculated as follows:

$$p_{lt,k} = h_{lt} \cdot t_k^T \tag{2}$$

$$\alpha_{lt,k} = softmax\left(p_{lt,k}\right) \tag{3}$$

$p_{lt,k}$ is an intermediate variable used to calculate α. Intuitively, plt,k is a measure of the matching degree between the hidden vector h_{lt} of the word and the topic vectort_k, which reflects the correlation between the word and the topic

to a certain extent. h_{lt} represents the hidden state vector of the vocabulary, t_k represents the low-dimensional topic embedding, $\alpha_{lt,k}$ represents the correlation weight between the t-th vocabulary in the l-th sentence and the k-th topic. After that, W-TIAM constructs the word topic embedding vector q_{lt}.

$$q_{lt} = \sum_{k=0}^{k} \alpha_{lt,k} \cdot t_k \tag{4}$$

word topic embedding vector q_{lt} and the word hidden state vector h_{lt} are spliced into k_{lt}, which represents the topic feature vector of the t-th vocabulary of the l-th sentence.

$$k_{lt} = [h_{lt} : q_{lt}] \tag{5}$$

C. SEQUENCE LEVEL ATTENTION MECHANISM (SLAM)

The sequence-level attention mechanism further integrates the semantics of the text into the sequence dimension, thereby constructing the expression F of the text. The process is briefly explained as follows:

1) INTEGRATED SENTENCE REPRESENTATION S

Not all words contribute equally to the meaning of a sentence. Therefore, we introduce attention mechanism to integrate word level semantic representation into sentence representation S, which is represented as follows:

$$s_l = \sum_t \beta_{lt} \cdot k_{lt} \tag{6}$$

$$\beta_{lt} = \frac{\exp\left(e_{lt}^T A_w\right)}{\sum_{\mathrm{t}} \exp\left(e_{lt}^T A_w\right)} \tag{7}$$

$$e_{lt} = \tanh\left(W_w k_{lt} + b_w\right) \tag{8}$$

k_{lt} represents a lexical vector representation containing topic information by fusing the hidden state vector and the topic feature vector. e_{lt} is a hidden representation obtained by k_{lt} through a single layer perceptron. β_{lt} is a measure of the importance of the word in the whole sentence. s_l is the vector representation of the sentence. $W_{w,}, b_w, A_w$ are the model parameter to be learned.

- Integrated Text Representation F

Similarly, each sentence has different contributions to the document. The attention mechanism is used to integrate the sentence sequence representation into the text representation F, which is represented as follows:

$$F = \sum_l \varphi_l \cdot s_l \tag{9}$$

$$\varphi_l = \frac{\exp\left(e_l^T A_s\right)}{\sum_{\mathrm{t}} \exp\left(e_l^T A_s\right)} \tag{10}$$

$$e_l = \tanh\left(W_s s_l + b_s\right) \tag{11}$$

e_l is the hidden representation of s_l through a single layer perceptron; φ_l is used to measure the importance weight of the sentence in the whole document; F is the vector representation of the document; W_s, b_s, A_s are the model parameter to be learned.

D. CLASSIFIER

The classifier infers the label y based on the above-mentioned semantic vector F. The classifier is composed of two fully connected multi-layer perceptron layers, and then uses the *softmax* (or *sigmoid*) function to map the result of the analysis into a probability form [5].

E. LOSS FUNCTION

Regarding model training, the model training in this article uses an end-to-end back propagation training method, and the objective function uses a cross-entropy cost function. Y is the known classification category, $\hat{y}$ is the category predicted by the model, and our goal is to minimize the cross entropy of the known classification category and the predicted classification category.

$$H\left(\hat{y}, Y\right) = -\sum\nolimits_{\mathrm{i}} \left(Ylog\left(\hat{y}\right) + (1 - Y)log\left(1 - \hat{y}\right)\right) \tag{12}$$

F. EVALUATION CRITERIA

This paper selects Precision, Recall and F-value as the evaluation indicators for sentiment/text analysis. Precision refers to the probability that the sample is correctly classified in

TABLE 1. Dataset partition.

Corpus	Classes	Instance	Train	Test
CCF-BDCI	3	5000	4400	600
THUCNews	6	30000	27000	3000

the classification. Specifically, in this text analysis, it refers to the proportion of the number of correctly classified texts to the total number of comments, usually represented by P, and the calculation method is as follows:

$$P = \frac{Number\ of\ correctly\ classified\ texts}{Total\ text} \tag{13}$$

Recall rate refers to the proportion of a certain type of sample in the classification. Specifically, in this text sentiment analysis, it refers to the proportion of the number of correctly classified sentimentally positive texts to the number of all sentimentally positive texts. Usually expressed by R, the calculation method is as follows:

$$R = \frac{Number\ of\ correctly\ classified\ positive\ texts}{The\ total\ number\ of\ positive\ texts} \tag{14}$$

Precision and Recall are generally negatively correlated in text classification. Pursuing higher accuracy often means sacrificing part of the Recall. Therefore, the F value is introduced to comprehensively consider the accuracy and Recall rates. The calculation method of F value is as follows:

$$F = \frac{2 * P * R}{P + R} \tag{15}$$

IV. EXPERIMENTAL RESULTS

A. DATASET

In order to verify the proposed model, this study conducted experiments on the CCF-BDCI and THUCNews data sets, both of which are Chinese data sets.

- **CCF-BDCI** created by CCF, there are 7345 news reports in total, sentiment is marked into three categories, positive, neutral, and negative. Considering the characteristics of the model, 5000 pieces of data were filtered out.
- **THUCNews** is filtered according to the historical data of the Sina News RRS subscription channel in the past years. This research has selected six categories of sports, entertainment, home, real estate, education, and fashion, with a total of 30,000 data.

B. EVPERIMENTAL DEVICE

In order to verify the effectiveness of the TBAM model, we use RCNN [35], Bi-GRU, AttBiLSTM [36], HAN [5], BERT [37] as baselines, all of which are effective methods for classification problems. RCNN, Bi-GRU, AttBiLSTM, and BERT are all advanced neural network models for classification. The inspiration for this study is from the HAN model of Yang *et al.*, but our model is different from HAN. Based on the HAN model, we introduce potential topic information into the semantic representation at the lexical level to improve the effect of lexical representation, and the topic information is regarded as domain information, which makes up for the ignorance of domain information by the classification model and increases the domain adaptability of the model. Experiments show that our model's effect on text classification and sentiment classification is better than that of the baseline model.

First of all, perform some filtering on the text data in the data set. From the two data sets, the number of sentences is not less than three and the sentence length is not less than ten words. Secondly, to preprocess the filtered text, we use jieba to mark each text, and delete some special marks and remove stop words. Then construct the Word2vec training corpus on the basis of the two data sets, and train the word vector through the gensim tool to obtain the word vector matrix U. Finally, 600 articles were randomly selected from the processed CCF-BDCI to construct a test set, and the rest were used as a training set. Randomly select 3000 articles from the processed THUCNews data set as the test set, and the rest as the training set.

C. RESULTS AND DISCUSSION

•CCF-BDCI

On the CCF-BDCI data set, the experimental results of CCF-BDCI and the baseline model are shown in Table 2

1) **Considering the hierarchical results of the text, the use of attention mechanism at the word level and sentence level plays an important role in text sentiment analysis.** The experimental results show that the Recall and F1-score of the TBAM model that uses the attention mechanism at both the word level and the sequence level are higher than those of the AttBiLSTM model used only at the sequence level, and are also higher than other baseline models. The Precision of the TBAM model and the AttBiLSTM model that introduce the attention mechanism is higher than that of the Bi-GRU with a similar structure, and the RCNN and BERT, which are also advanced neural network models. And BERT has a multi-head self-attention mechanism, Precision, Recall and F1-score are higher than Bi-GRU and RCNN. It can be inferred from this that it is effective to use the attention mechanism at the word level and sentence sequence level.
2) **TBAM extends word-level semantic representation through topic models and improves the classification performance of the model.** The most important difference between the HAN model and the TBAM model is the introduction of the theme. After the introduction of the theme, the Precision, Recall, and F1-score of the TBAM model are all higher than HAN. The results

TABLE 2. Experiment results.

Models	CCF-BDCI			THUCNews		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
RCNN	0.652	0.544	0.592	0.950	0.952	0.952
Bi-GRU	0.679	0.608	0.640	0.951	0.608	0.951
AttBiLSTM	0.703	0.589	0.639	0.936	0.930	0.933
HAN	0.676	0.562	0.610	0.940	0.939	0.940
BRET	0.680	0.670	0.675	0.953	0.952	0.952
TBAM	0.703	0.614	**0.694**	0.952	0.953	**0.953**

TABLE 3. Hyper parameter K.

K	CCF-BDCI	THUCNews
25	0.566	0.940
30	**0.654**	**0.953**
35	0.564	0.933

show that the use of topics can expand the semantic representation of the word level and bring better results to the model

3) **Through this experiment, it proved the effectiveness of TBAM in the field of sentiment analysis.** It can be seen intuitively that the F1-score of TBAM is higher than other baseline models.

•THUCNews

On the THUCNews data set, the experimental results of THUCNews and the baseline model are shown in Table 2.

It can be seen intuitively that the F1-score of TBAM is higher than other baseline models.

The experimental results of TBAM are better than all other baselines, which shows that our model is effective in text analysis tasks. The other observations in Table 3 are consistent with Table 2, and therefore further support the previous discussion. According to the experimental results in Table 2 and Table 3, the following conclusions can be drawn:

1) Pay attention to the original hierarchical structure of the article. Using the attention mechanism at the word level and sequence level respectively will improve the model's effect.
2) The topic model extends the semantic representation at the word level, introduces potential topic information into the semantic representation at the vocabulary level to improve the effect of vocabulary representation, treats topic information as domain information, makes up for the classification model's neglect of domain information, and increases The domain adaptability of the model, so whether it is sentiment analysis on the news data set or classification judgment on the classification data set, the model can better complete the classification task.

In order to prove the effectiveness of the model, we conducted experiments on two data sets (CCF-BDCI and THUCNews), both of which are well-known public data sets., The results run on it will be representative. At the same time, the experiments on these two data sets can prove the following two points according to their experimental results: 1. The correctness of the model's idea of adding topic information, 2. Note that the information brought by the sentence layer to the article can better improve the classifier Effect. However, our model also has shortcomings. The addition of topics makes the model unable to process too short sentences. Therefore, it also has high requirements on the data set. Before conducting the experiment, we have done the emotion data set currently used. Strict screening, other data sets will be excluded because of the length of the text, so we can only experiment on these two data sets at present. Now, we are revising and improving the model's shortcomings, and we will conduct experiments on more well-known sentiment classification data sets in the future.

D. HYPER PARAMETER K

Further considering the impact of the number of topics on the model, the following experiments were carried out, and the F1-score index was verified on the data set for different number of topics:

It can be seen that the number of topics still has a great impact on the performance of the model, so try to determine the optimal number of topics in the data set before doing the experiment. In LDA, there is no good way to determine the number of topics. We use more attempts. The number of topics in the data set is from 10, 20, 25, 30, and 35 steps, and the parameters are manually adjusted according to the training results. Will be determined eventually.

E. TIME COMPLEXITY AND SPACE COMPLEXITY

The amount of computation and access (the amount of storage space occupied in the process of operation) of the model are also the key factors that affect the performance of the model. Too much computation will lead to a lot of time-consuming model training and prediction, which can not quickly verify the idea and improve the model, and also can not achieve fast classification prediction. The spatial complexity determines the number of parameters of the model. Due to the limitation of dimension disaster, the more parameters of the model, the more data the training model needs. The data used in the experiment for training is often limited, and the training data can not reach the level required by the model, which often leads to the problem of over fitting. In the proposed model, epoch size is set to 200, batch size is set to 200, the number of training iterations is 20000, the training time is 6 hours, and the model size is 244mb.

V. CONCLUSION

This paper proposes a hierarchical attention neural network model (TBAM) that incorporates topic information. The TBAM model reflects the hierarchical structure of the document, and uses the attention mechanism at the vocabulary and sentence levels to make it possible to construct the document representation separately. Pay attention to the more important and the less important content. In the vocabulary level, the potential topic information is introduced into the semantic representation at the vocabulary level to improve the effect of vocabulary representation. The topic information is regarded as domain information, which makes up for the ignorance of domain information by the classification model and increases the domain adaptability of the model. Through testing on the public data sets CCF-BDCI and THUCNews, the effectiveness of the model in classification is verified from the three indicators of Precision, Recall and F1. However, the shortcoming of this model is that the classification ability of short text data is not outstanding. The theme of short text data set is relatively scattered, and there is not much useful information, and feature extraction is difficult. The next research will explore whether it can be combined with other short text data topics. The combination of extraction methods can make up for this deficiency.

REFERENCES

[1] W. Liu, G. Cao, and J. Yin, "Bi-level attention model for sentiment analysis of short texts," *IEEE Access*, vol. 7, pp. 119813–119822, 2019, doi: 10.1109/ACCESS.2019.2936457.

[2] M. Soleymani, D. Garcia, B. Jou, B. Schuller, S.-F. Chang, and M. Pantic, "A survey of multimodal sentiment analysis," *Image Vis. Comput.*, vol. 65, pp. 3–14, Sep. 2017, doi: 10.1016/j.imavis.2017.08.003.

[3] H. Xie, X. Li, T. Wang, R. Y. K. Lau, T.-L. Wong, L. Chen, F. L. Wang, and Q. Li, "Incorporating sentiment into tag-based user profiles and resource profiles for personalized search in folksonomy," *Inf. Process. Manage.*, vol. 52, no. 1, pp. 61–72, Jan. 2016, doi: 10.1016/j.ipm.2015.03.001.

[4] Z. Zheng, "Research on sentiment analysis algorithm based on topic features and deep learning," China Univ. Technol. Taipei Campus, Taipei, Taiwan, Tech. Rep., 2018.

[5] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, San Diego, CA, USA, 2016, pp. 1480–1489, doi: 10.18653/v1/N16-1174.

[6] F. Wan, "Sentiment analysis of Weibo comments based on deep neural network," in *Proc. Int. Conf. Commun., Inf. Syst. Comput. Eng. (CISCE)*, Haikou, China, Jul. 2019, pp. 626–630, doi: 10.1109/CISCE.2019.00146.

[7] X. Zhao, C. Wang, Z. Yang, Y. Zhang, and X. Yuan, "Online news emotion prediction with bidirectional LSTM," in *Web-Age Information Management*, vol. 9659, B. Cui, N. Zhang, J. Xu, X. Lian, and D. Liu, Eds. Cham, Switzerland: Springer, 2016, pp. 238–250.

[8] X. Li, Y. Rao, H. Xie, X. Liu, T.-L. Wong, and F. L. Wang, "Social emotion classification based on noise-aware training," *Data Knowl. Eng.*, vol. 123, Sep. 2019, Art. no. 101605, doi: 10.1016/j.datak.2017.07.008.

[9] Y. Kim, "Convolutional neural networks for sentence classification," Sep. 2014, *arXiv:1408.5882*. Accessed: Nov. 26, 2020. [Online]. Available: http://arxiv.org/abs/1408.5882

[10] G. Zhou, L. Lv, X. Qiao, and L. Jin, "Hierarchical attention-based fuzzy neural network for subject classification of power customer service work orders," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE)*, New Orleans, LA, USA, Jun. 2019, pp. 1–6, doi: 10.1109/FUZZ-IEEE.2019.8858852.

[11] D. Cao, C. District, Y. Huang, and H. Li, "Text sentiment classification based on attention mechanism and decomposition convolutional neural network model," in *Proc. IEEE Int. Conf. Adv. Elect. Eng. Comput. Appl.*, Aug. 2020, pp. 234–238.

[12] X. Zhu, J. Huang, Z. Zhou, and Y. Han, "Chinese article classification oriented to social network based on convolutional neural networks," in *Proc. IEEE 1st Int. Conf. Data Sci. Cyberspace*, Jun. 2016, pp. 33–36.

[13] H. Lu, X. Liu, Y. Yin, and Z. Chen, "A patent text classification model based on multivariate neural network fusion," in *Proc. 6th Int. Conf. Soft Comput. Mach. Intell. (ISCMI)*, Johannesburg, South Africa, Nov. 2019, pp. 61–65, doi: 10.1109/ISCMI47871.2019.9004335.

[14] A. Basnet and A. K. Timalsina, "Improving nepali news recommendation using classification based on LSTM recurrent neural networks," in *Proc. IEEE 3rd Int. Conf. Comput., Commun. Secur. (ICCCS)*, Kathmandu, Nepal, Oct. 2018, pp. 138–142, doi: 10.1109/CCCS.2018.8586815.

[15] S. Zeng, Y. Ma, X. Zhang, and X. Du, "Term-based pooling in convolutional neural networks for text classification," *China Commun.*, vol. 17, no. 4, pp. 109–124, 2020.

[16] G. Alfian, M. Syafrudin, M. Ijaz, M. Syaekhoni, N. Fitriyani, and J. Rhee, "A personalized healthcare monitoring system for diabetic patients by utilizing BLE-based sensors and real-time data processing," *Sensors*, vol. 18, no. 7, p. 2183, Jul. 2018, doi: 10.3390/s18072183.

[17] M. F. Ijaz, M. Attique, and Y. Son, "Data-driven cervical cancer prediction model with outlier detection and over-sampling methods," *Sensors*, vol. 20, no. 10, p. 2809, May 2020, doi: 10.3390/s20102809.

[18] M. Ijaz, G. Alfian, M. Syafrudin, and J. Rhee, "Hybrid prediction model for type 2 diabetes and hypertension using DBSCAN-based outlier detection, synthetic minority over sampling technique (SMOTE), and random forest," *Appl. Sci.*, vol. 8, no. 8, p. 1325, Aug. 2018, doi: 10.3390/app8081325.

[19] Y. Ma, H. Peng, T. Khan, E. Cambria, and A. Hussain, "Sentic LSTM: A hybrid network for targeted aspect-based sentiment analysis," *Cogn. Comput.*, vol. 10, pp. 639–650, Mar. 2018.

[20] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," Apr. 2014, *arXiv:1404.2188*. Accessed: Nov. 26, 2020. [Online]. Available: http://arxiv.org/abs/1404.2188
[21] D. Pan, J. Yuan, L. Li, and D. Sheng, "Deep neural network-based classification model for sentiment analysis," in *Proc. 6th Int. Conf. Behav., Econ. Socio-Cultural Comput. (BESC)*, Beijing, China, Oct. 2019, pp. 1–4, doi: 10.1109/BESC48373.2019.8963171.
[22] L. Wang and B. Zhang, "Fault text classification based on convolutional neural network," in *Proc. IEEE 7th Int. Conf. Ind. Eng. Appl. (ICIEA)*, Bangkok, Thailand, Apr. 2020, pp. 937–941, doi: 10.1109/ICIEA49774.2020.9101960.
[23] P. Semberecki and H. Maciejewski, "Deep learning methods for subject text classification of articles," in *Proc. Federated Conf. Comput. Sci. Inf. Syst.*, Sep. 2017, pp. 357–360, doi: 10.15439/2017F414.
[24] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, "Indexing by latent semantic analysis," *J. Amer. Soc. Inf. Sci.*, vol. 41, no. 6, pp. 391–407, 1990.
[25] T. Hofmann, "Unsupervised learning by probabilistic latent semantic analysis," *Mach. Learn.*, vol. 42, pp. 177–196, Jan. 2001.
[26] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.
[27] B. G. Rosenberg, J. K. Balogh, H. Gerjuoy, J. Bond, and R. Mc Devitt, "The DL scale: The measurement of clinical status of a psychiatric ward," *J. Clin. Psychol.*, vol. 18, no. 3, pp. 290–294, 1962.
[28] M. Syamala and N. J. Nalini, "LDA and deep learning: A combined approach for feature extraction and sentiment analysis," in *Proc. 10th Int. Conf. Comput., Commun. Netw. Technol.*, vol. 5, 2019, pp. 1–5.
[29] S. Niu, X. Chai, and Q. Li, "A text classification algorithm based on neural network and LDA," *J. Comput. Eng.*, pp. 208–214, 2019.
[30] Y. Miao, E. Grefenstette, and P. Blunsom, "Discovering discrete latent topics with neural variational inference," May 2017, *arXiv:1706.00359*. Accessed: Nov. 26, 2020. [Online]. Available: http://arxiv.org/abs/1706.00359
[31] A. Srivastava and C. Sutton, "Autoencoding variational inference for topic models," Mar. 2017, *arXiv:1703.01488*. Accessed: Nov. 26, 2020. [Online]. Available: http://arxiv.org/abs/1703.01488
[32] F. Li, M. Huang, and X. Zhu, "Sentiment analysis with global topics and local dependency," in *Proc. AAAI Conf. Artif. Intell.*, 2010, pp. 1–7.
[33] S. Qian, T. Zhang, and C. Xu, "Boosted multi-modal supervised latent Dirichlet allocation for social event classification," in *Proc. 22nd Int. Conf. Pattern Recognit.*, Stockholm, Sweden, Aug. 2014, pp. 1999–2004, doi: 10.1109/ICPR.2014.349.
[34] H. M. Wallach, D. Mimno, and A. McCallum, "Rethinking LDA: Why priors matter," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 1973–7981.
[35] S. Lai, L. Xu, K. Liu, and J. Zhao, "Recurrent convolutional neural networks for text classification," in *Proc. AAAI Conf. Artif. Intell.*, 2015, pp. 1–7.
[36] P. Zhou, W. Shi, J. Tian, Z. Qi, B. Li, H. Hao, and B. Xu, "Attention-based bidirectional long short-term memory networks for relation classification," in *Proc. 54th Annu. Meeting Assoc. Comput. Linguistics*, vol. 2. Berlin, Germany, 2016, pp. 207–212, doi: 10.18653/v1/P16-2034.
[37] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," May 2018, *arXiv:1810.04805*. Accessed: Nov. 26, 2020. [Online]. Available: http://arxiv.org/abs/1810.04805

HONGTAO LIU received the master's degree in computer application technology from Chongqing Normal University, China, in 2001, and the Ph.D. degree in artificial intelligence from Southwest University, Chongqing, China, in 2004. He is currently an Associate Professor with the School of Computer Science, Chongqing University of Posts and Telecommunications. His research interests include natural language processing, social networking, and swarm intelligence.

QIMIN QIAN received the master's degree from the Chongqing University of Posts and Telecommunications, in 2021. Her research interest includes natural language processing.