Research Article

Short-Text Classification Detector: A Bert-Based Mental Approach

Yongjun Hu ,[1] Jia Ding,[2] Zixin Dou ,[1,3] and Huiyou Chang[2]

[1]*School of Management, Guangzhou University, Guangzhou 510006, China*
[2]*School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510006, China*
[3]*Research Center of e-commerce, Guangzhou University, Guangzhou 510006, China*

Correspondence should be addressed to Zixin Dou; 13826064897@163.com

With the continuous development of the Internet, social media based on short text has become popular. However, the sparsity and shortness of essays will restrict the accuracy of text classification. Therefore, based on the Bert model, we capture the mental feature of reviewers and apply them for short text classification to improve its classification accuracy. Specifically, we construct a model text at the language level and fine tune the model to better embed mental features. To verify the accuracy of this method, we compare a variety of machine learning methods, such as support vector machine, convolution neural networks, and recurrent neural networks. The results show the following: (1) Through feature comparison, it is found that mental features can significantly improve the accuracy of short text classification. (2) Combining mental features and text as input vectors can provide more classification accuracy than separating them as two independent vectors. (3) Through model comparison, it can be found that Bert model can integrate mental features and short text. Bert can better capture mental features to improve the accuracy of classification results. This will help to promote the development of short text classification.

1. Introduction

With the proliferation of online text information, text classification plays a vital role in obtaining information resources [1]. As an efficient and well-known natural language processing technology, text classification can identify the content of a given document and find the relationship between document features and document categories. It is widely used in various fields, such as event detection [2,3], media analysis [4, 5], viewpoint mining [6, 7], and predicting product revenue [8,9]. Although text classification has always been a well-known problem, a suitable solution for short text classification has not been found. Especially, with the rapid growth of the digital media scale, a complex environment will affect the results of text content retrieval and analysis. This makes short text classification a challenging task. Therefore, to promote content analysis of online text information, a reliable text classification tool is needed [10].

Recently, a large number of scholars have studied text classification. Traditional classification algorithm models

include *K*-nearest neighbor (KNN) [11], naive Bayes (NB) [12], and support vector machine (SVM) [13]. These models have good classification results and have been widely used. They extract the features of text documents and then use one or more classifiers to predict multiple related tags [14, 15]. However, such methods are time-consuming and require the extensive domain knowledge of experts [10]. At present, with the development of deep learning, the traditional classification methods are gradually integrated and replaced by deep learning classification algorithms. As deep learning can learn representation from data without complex feature engineering, it has become a hot research topic in this field [16, 17]. To obtain a better classification effect, many researchers use a convolutional neural network (CNN) [18] and recurrent neural network (RNN) [19] to extract and calculate text features. In particular, the bidirectional encoder representations from transformers (Bert) [20] developed by Google. Different from the previous network architecture, Bert is based on attention mechanism and transformer coding structure. However, previous studies

have not used the mental feature of the speaker, that is, mental features, for short text classification.

The current research challenges are as follows. Although the emergence of machine learning will improve the effect of text classification, the sparsity and shortness of text will limit the accuracy of text classification. At the same time, with the growth of smart phones, short text has been integrated into daily life. Therefore, a method is needed to quickly identify the publisher's intention, to improve the accuracy of classification. In addition, when it comes to cross languages, the traditional method uses the same corpus to train the classifier, but this method cannot be extended to multilingual environment [21]. To address this knowledge gap, in this research work, we focus on text classification according to users' mental features, and use cross-lingual data sets for experiments. Given this, this paper proposes a method, which effectively integrates mental features with text content on a linguistic level. Specially, we have designed two methods to integrate text context and all features on a linguistic level. Then, we fine-tune the model by evaluating the significance of each feature. The mental features can reflect the speaker's behavior and improve the accuracy of the method. Meanwhile, to verify the accuracy of this method, we compare a variety of machine-learning methods. The results show that our method has significant advantages in short text classification tasks.

The main contributions of this paper are as follows:

–A Bert-based mental approach is proposed for the classification of short text content. The proposed method combines the user's mental feature with the short text content. It can help better identify the user's intention contained in the short text, that is, false comments or text topic detection.

–Compared with other existing machine-learning research, the proposed method effectively integrates mental feature vector and short text vector. It improves the accuracy of text classification and achieves good accuracy on cross-corpus data sets.

The rest of this paper is shown as follows. Section 2 summarizes the literature reviews. In Section 3, we introduce the methodology. Section 4 is an experiment, including data set, research settings, evaluation metrics, and experimental results. Section 5 discusses the key findings, theoretical implications, and practical implications. Finally, the conclusion is presented.

2. Literature Review

2.1. Mental Theory. The concept of mental theory was first proposed by Craik [22]. The basic assumption of the mental theory is that any form of communication is based on the situational way people talk about [23, 24]. It represents how people imagine and understand the situation in the world. Although it may contain factual information, it is not only able to identify facts, it can also be used to make judgments and inferences, to affect people's behavior. When people use signs to represent objects, it is often quite abstract. Therefore, mental models need to be used to explain existing concepts by adding more information [25], which will have a potential impact on people. Mental theory has been widely used in many fields, such as management education [26] and management decision-making [27, 28].

In the field of text, the mental theory has been proved to improve text comprehension [29–31]. Specifically, mental theory reflects the different levels of representation formed by the speaker in the process of text writing. Representation refers to an abstract propositional representation between the thought contained in the text and its linguistic information. This is also a cognitive representation of reality, which is related to the speaker's cognition, perception, and behavior in various situations [25, 32]. It is worth noting that it represents the content of the text (the events, objects, and processes described in the text), rather than the characteristics of the text itself [33].

At the same time, deep learning can be used as a powerful tool to expand mental theory. For example, to improve teaching efficiency, Tawatsuji et al. [34] and Matsui et al. [35] extracted the relationship between students' mental state and mental information through deep learning, supplemented by teachers' speech behavior. To ensure driving safety, Darwish et al. [36] analyzed the driver's psychology through in-depth learning and judge how the driver perceives the environment. Dutta et al. [37] believed that machine-learning algorithm is used as a classification tool of mental state, which can improve the accuracy of classification.

Therefore, this paper combines mental theory with deep learning to apply the classification of short texts. By discovering the relationship between text and mental features, more accurate classification results will be obtained.

2.2. Text Classification Method. Recently, most classification methods are mainly based on machine learning. In terms of KNN, Moldagulova et al. [11] and Trstenjak et al. [38] use the KNN algorithm to classify documents. The results show that this algorithm has good classification performance. In terms of Xgboost, Wang et al. [39] took the Xgboost algorithm and granularity parameters as input characteristics to predict sample categories. Li and Zhang [40] proposed a classification prediction model based on the Xgboost algorithm. In terms of NB, Zhu et al. [12] used the NB algorithm for text classification. Jiang et al. [41] proposed an improved NB technology for text classification performance. This method solves the problem of unsatisfactory results caused by the uneven distribution of training data. Bilal et al. [42] used the NB algorithm for periodical literature classification. The results show that the accuracy of this method is high enough. In addition, some scholars also studied SVM, which is a method used to predict and define how to classify data sets [43]. It can classify text data into predefined classes [44]. In terms of SVM, Luo [13] applied the SVM model to the classification of English texts and documents. The results show that this method has good performance. According to Vijayarani et al. [45], the accuracy of SVM is slightly higher than that of NB. However, traditional target classification focuses on feature engineering to maximize the use of classifiers [46–48], such as SVM. Such methods are time-consuming and require the extensive domain knowledge of experts.

With the development of machine learning, a large number of NN models have emerged in natural language processing tasks [49,50]. As these methods can learn representation from data without complex feature engineering, NN has become a hot research topic in this field. Mainstream NN include RNN [19,51], gated neural network (GNN) [52,53], CNN [50, 54, 55], and long short-term memory (LSTM) [56]. The most popular neural network architectures are CNN and RNN. CNN has a good performance on features extraction by convolution kernel, which improves the accuracy of feature descriptors. RNN is widely used to capture flexible context information. However, Kandhro et al. [57] found that the performance of the LSTM method has more advantages than CNN and RNN. LSTM solves the vanishing gradient problem and has long-term dependence, which can retain the characteristics of previous learning [56]. Tang et al. [58] found that LSTM can effectively capture the information of sentences. Lee et al. [59] mine tourists' destinations and preferences through text classification and spatial clustering based on LSTM. The results show that this method has good results. Subsequently, Google proposed Bert [20]. This model has made a breakthrough in the text field and achieved the most advanced results. A large number of scholars widely used Bert, which is enough to prove its great advantages in feature extraction [60,61].

However, previous studies have not used human mental-related features for short text classification. We assume that in a specific linguistic pattern, mental features can provide more information for short text classification. Therefore, this paper studies the Bert method for short text classification, and fine-tune it. Specifically, to obtain more accurate classification results, we combine mental features with text content according to specific short text patterns.

3. Methodology

3.1. Theory.
The sparsity and shortness of short text may seriously destroy the representation of short text. An important solution is to enrich the short text representation by involving the cognitive aspects of the text, such as mental features. Generally, the short text content of users can be enriched from external mental and internal mental features (as shown in Figure 1).

Figure 1 shows the conceptual framework proposed in this study, which is divided into two main stages: mental model processing and classification model training. First, we propose two methods to embody mental features, namely, history information and Maslow's need. Our text data set contains cross-lingual corpus. The approach proposed in this paper can effectively integrate mental feature vector and short text vector, classify short text with higher accuracy, and help readers understand the intention contained in the text. Next, the application of nonconcentric zhite is introduced in methods 1 and 2, respectively.

3.1.1. Method 1.
The historical information of users contain their behavior laws. There will be great differences in the historical information of different users. Based on this, this paper introduces the user history feature into the short text classification model as an external mental feature. Its concept is as shown in Figure 2. By enriching the user's intention expressed in a short text, accurate classification is carried out.

3.1.2. Method 2.
This paper takes Maslow's need as the internal mental feature. Maslow's needs include five levels of cognitive needs, including physiological needs, security needs, emotional needs, respect needs, and self-realization needs. It can be expected that different levels of needs play an important role in everyone's character formation. Therefore, it is also important to understand these basic requirements in short text classification. If the mental feature is constructed according to the different needs of users, the short text can be enriched, and the readers can better understand the content of the text. Its concept is as shown in Figure 3.

3.2. Model Structure.
The Bert is realized by constructing mental features and combining domain-related knowledge. Its structure is shown in Figure 4. The model integrates text and features into a corpus instead of inputting them with different vector matrices. The data input rules can be described as follows:

(i) [Text] indicates the text content

(ii) [F] represents a Feature item

(iii) [CLS] indicates that the corpus is used for the classification model

(iv) [SEP] represents a clause symbol, which is used to disconnect two sentences in the input corpus

The Bert mainly consists of the following three parts:

Input layer: the feature and text data are used to establish the input sequence for this model. Then, the final input representation is obtained by summing the position embedding, word embedding, and segmentation embedding of each tag sequence.

Encoder layer: it consists of 12 transform blocks, which input the marked sequence and output the representation of the sequence.

Output layer: it consists of a simple softmax classifier at the top of the encoder layer, which is used to calculate the conditional probability distribution on the predefined classification label.

3.2.1. Input Layer.
The difference between the two methods is how we input meta-data into the Bert:

(i) Pair Method (PM): this method inputs the feature text into the model as a sentence independent of the claim. That is, at the token level, the claim is separated from the feature text by the special token '[SEP]'.
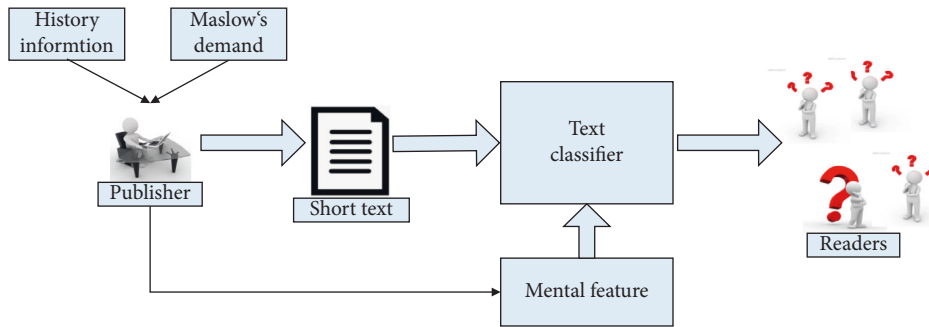
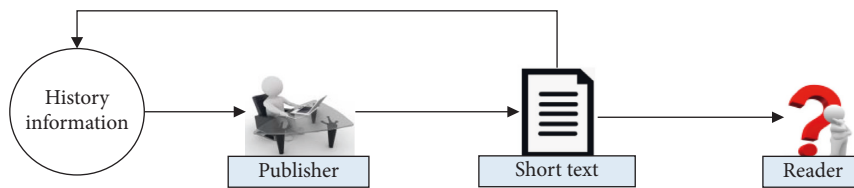FIGURE 1: The concept of text classification based on mental feature.

FIGURE 2: The concept of text classification based on history information.
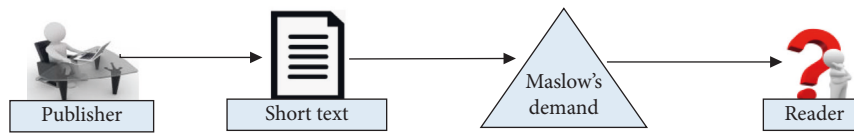

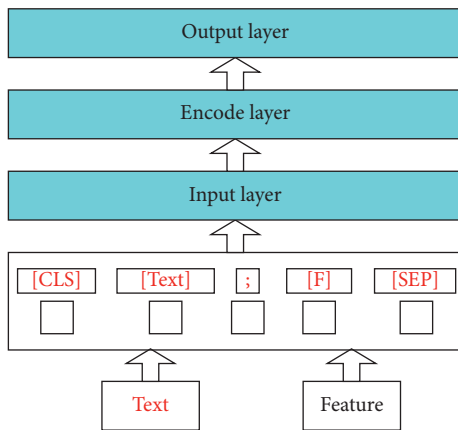FIGURE 3: The concept of text classification based on Maslow's need.

FIGURE 4: Basic structure of the Bert.

(ii) News Text Method (NTM): this method inputs the news text into the model as a single sentence. Noteworthy, the News Text is composed of Claim and Feature Text. They are separated by '; '. Only one '[SEP]' token is added at the end of the entire token sequence.

The input representation of each token e is obtained by adding its token embedding (W), segment embedding (S), and position embedding (P). For visualization of this structure, see Figure 5. The embedding features of these three words are as follows:

(1) Token embedding: this embedding is a vector to convert each word into a fixed dimension. The input text will be tokenized when it is sent into this embedding. In addition, two special tokens, [CLS] and [SEP], will be inserted at the beginning and the end of the tokenization result. They are regarded as the following classification tasks and the effect of dividing sentences on services, respectively.

(2) Position embedding: this embedding refers to encoding the position information of words into feature vectors. It is a crucial link to introduce the word position relationship into the model.

(3) Segment embedding: there are only two vector representations of this embedding. The former vector assigns 0 to each token in the first sentence, and the latter vector assigns 1 to each token in the second sentence. If the input has only one sentence, its segment embedding is all 0.

3.2.2. Encoder Layer. The model architecture is composed of 12 layers of transformers. Its basic structure is shown in
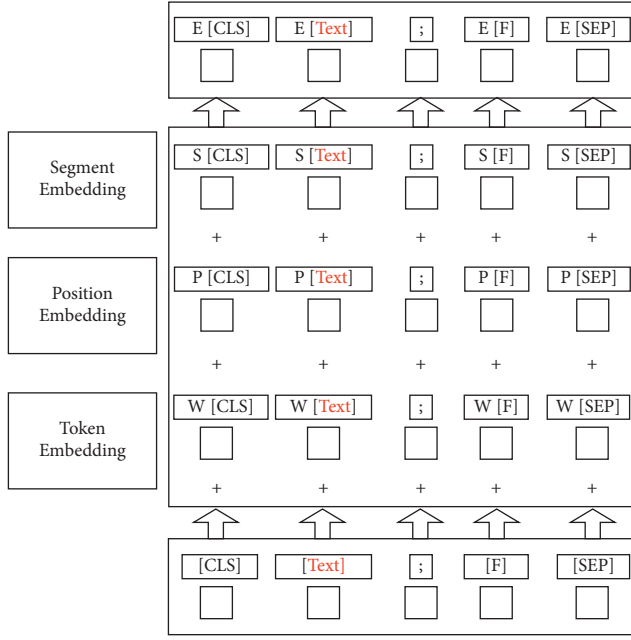
FIGURE 5: Construction of input sequence representations for Bert.

Figure 6. Each transformer is composed of a self-attention module, add&norm module, feed-forward module, and add&norm module:

(1) Self-attention module: this module is to find the correlation between words. Each self-attention mechanism first converts the input data into three vectors through three parameter matrices Q, K, and V. Where, Q is the query vector parameter matrix, K is the key vector parameter matrix, and V is the value vector parameter matrix. The converted vector dimension is smaller than the input dimension. Then, the machine calculates the self-attention vector of the Q. The main process is to divide the Q and the K by the square root of the K. This will reduce the vector size to a certain extent, which is conducive to keeping the gradient stable during backpropagation. After that a softmax operation is performed on all normalized dot products. Its purpose is to normalize, which can strengthen the influence of relevant time step data and weaken the influence of irrelevant time steps. Finally, multiply the aforementioned results by V. The calculation of self-attention is shown in the following formula:

$$Z = \text{Attention}(Q, K, V)$$
$$= \text{softmax}\left(\frac{QK^T}{\sqrt{d_t}}\right)V, \quad (1)$$

where Z is the output vector of the attention module.

(2) add&norm module: in this module, the Z vector is input into LayerNorm, which normalizes Z. The purpose of this is not to let the Z vector to fall in the saturation region of the activation

function. Therefore, the normalized N vector is obtained.

(3) Feed-forward module: as the calculation in the self-attention module is linear, to improve the nonlinear fitting ability of the model, a feed-forward network needs to be connected behind it. The network consists of two linear mapping parts. The first part is linear mapping and nonlinear activation relu function. The second part is a linear mapping. The formula is as follows:

$$F(Z) = \text{relu}(0, ZW_1 + b_1)W_2 + b_2, \quad (2)$$

where F is the output vector of the feed-forward neural network, W_1 and W_2 are the weight matrix, b_1 and b_2 are the bias.

Then the output of the feed-forward network is normalized by the add&norm module.

The aforementioned steps are a transformer. After 12 times, the output of the 12th transformer is a hidden state vector, that is, the T vector.

3.2.3. Output Layer. The output layer is a simple softmax classifier at the top of the model. This model only uses the final hidden state vector $T[\text{CLS}]$ as the aggregate representation of the sequence, that is, the T vector output through the transformer. The final classification result is obtained according to the following formula:

$$P = \text{softmax}\left(T[\text{CLS}]V^T\right) = \frac{\exp\left(P\left(y_i | T[\text{CLS}], \theta\right)\right)}{\sum_{i=1}^{c} \exp\left(P\left(y_i | T[\text{CLS}], \theta\right)\right)}, \quad (3)$$

where $V \in R^{c*h}$ is the trainable task-specific parameter matrix, and c is the number of labels. h is the dimension with a default value. It is worth noting that the category y_i with the probability of occurrence is the category to which the T[CLS]. Therefore, the final distribution function will output a C-dimensional vector. Each dimension represents the probability that $T[\text{CLS}]$ belongs to y_i. When the C-dimensional vector elements are normalized, the sum of them is 1.

4. Experiment

4.1. Data Set

4.1.1. For English Fake News Detection. The data set we experimented with is based on the LIAR data set, which was published by Wang [62]. It consists of a large number of claims, namely text content and related features. These features include subject, speaker, job, state, party, history, and context. Where Claim is the text content (Text). History indicates the speaker's statistics on the historical behavior of news speakers lying. From the perspective of psychology, this behavior can fully describe an individual's psychology. Therefore, this paper takes this index as a mental feature. For truthfulness, it is labeled as true, mostly true, half-true,

| T [CLS] | T [Text] | ; | T [F] | T [SEP] |

......

Add & Normalize

| Feed Forward | Feed Forward | Feed Forward | Feed Forward | Feed Forward |

| N [CLS] | N [Text] | ; | N [F] | N [SEP] |

LayerNorm

| Z [CLS] | Z [Text] | ; | Z [F] | Z [SEP] |

Add & Normalize

Self-Attention

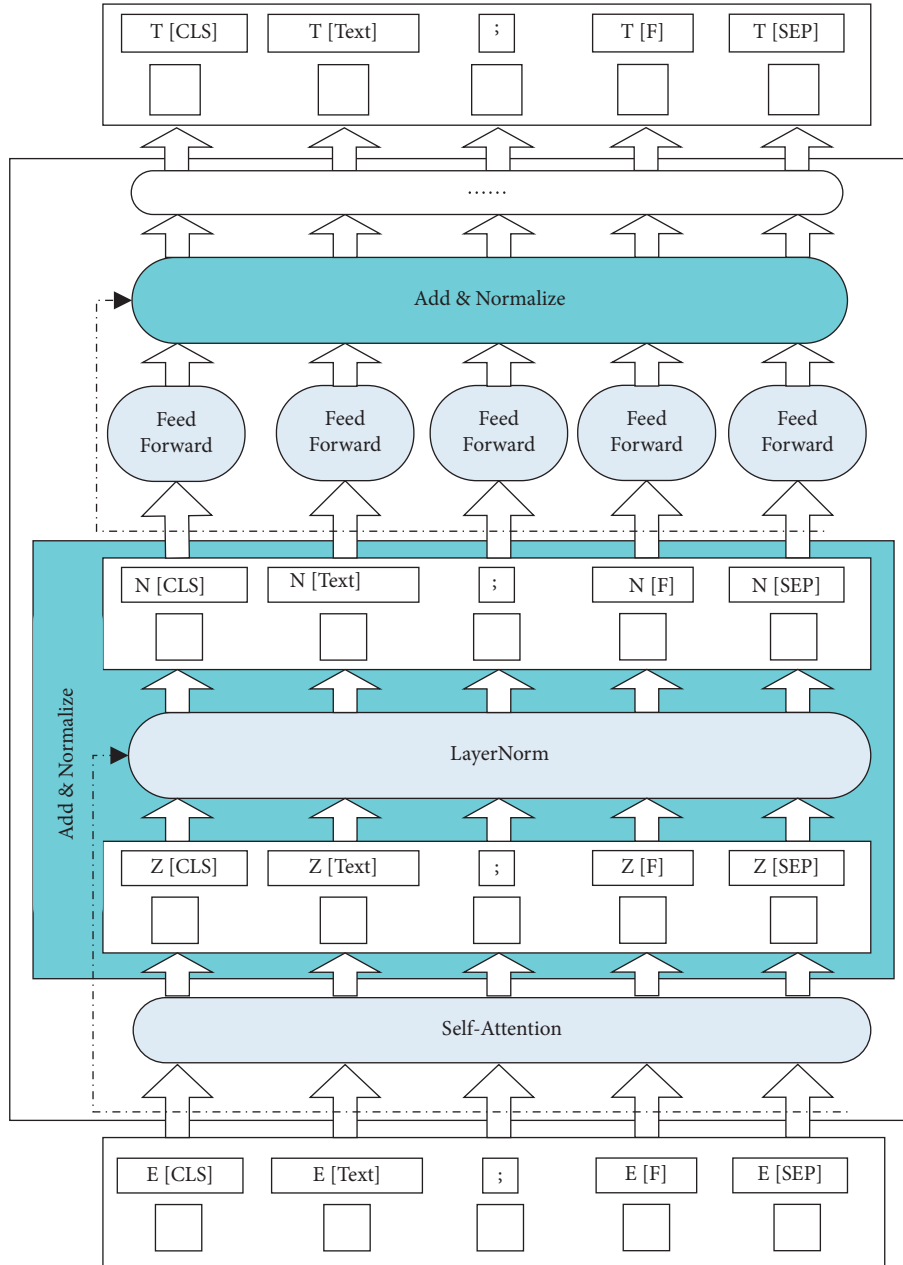| E [CLS] | E [Text] | ; | E [F] | E [SEP] |

FIGURE 6: The basic structure of the encoder.

barely-true, false, and pants-on-fire by journalists. To avoid these insignificant symbols affecting the results, we replaced some specific punctuation marks.

4.1.2. For Chinese Topic Detection. The data set we experimented with is based on social messaging data. The data set is a calculation of a great amount of text content and the Maslow's need features. We regard this feature as a mental feature because it can effectively reflect the mental state of the text publisher. For topic, it is labeled as meaningless, work/study, family, affection, leisure, and a blessing. To avoid these insignificant symbols affecting the results, we replaced some specific punctuation marks.

4.2. Experiment Settings. To explore the effectiveness of text classification, we first fine-tune the Bert method. Then, we evaluate the three methods, including all the features. Moreover, we use different feature combinations to evaluate the significance of mental features. Specifically, for model training, we set the learning rate to 2e–5, the batch size to 8, and the training time to 3.0.

4.3. Evaluation Metric. For evaluation metric, we use the following equation:

$$\text{accuracy} = \frac{(\text{TP} + \text{TN})}{(P + N)}, \tag{4}$$

where TP is for true positive, P is for total positive, FN is for false negative, and N is for total negative. By calculating these values, we can get the accuracy of the results.

4.4. Experimental Results

4.4.1. Fine-Tuning Analysis for Bert. We first compare the two input methods with all features included of Bert, namely PM and NTM. From Table 1, the accuracy of NTM results is 0.476 in English Fake News Detection (EFND) and 0.960 in Chinese Topic Detection (CTD), respectively. Both of them are significantly better than PM. The main reason is that PM inputs text and features as separate sequences. Although the model can learn the representation of each sentence sequence through fine-tuning, due to the segmentation of text and features, it will not be able to associate any feature information with the text, effectively. In other words, as text and features are input as different sentences, some information may be lost, which is the main reason for poor performance. We only need to input the text and features as a whole sentence. It is not necessary to split it with [SEP].

4.4.2. Feature Selection and Analysis. To verify that mental features contribute most to short text classification, we experimented with NTM composed of text and single feature items in the Bert method.

From Table 2, it can see that the model performs worst when it is fine-tuned only with text in EFND. Meanwhile, different features make different contributions to fake news detection. The mental feature has the greatest improvement, and their accuracy is more than 0.4 in EFND. This shows that mental features can effectively improve the accuracy of fake news detection.

In Table 3, we also found that the accuracy of text content is much lower than that of mental features in CTD. It verifies the contribution of mental features to topic detection.

4.4.3. Comparative Text Classification Method. In this section, we first compare our method with existing methods, using text and mental feature data. From Tables 4 and 5, the performance of Bert is significantly better than other models. It improves the accuracy by 0.2. This confirms the effectiveness of Bert. It also shows that the self-attention mechanism has a better ability to capture sequence semantics. Based on ensuring the task awareness of the model, it can directly learn the relationship between the target text sequence and the corresponding classification label, which simplifies the training.

In addition, when only plain text is used, our method is also used to compare existing models. In Tables 4 and 5, whether EFND or CTD, the results show that the performance of models using the mental feature is significantly better than those using plain text data. The mental feature approximately improves the absolute value of accuracy by 0.2.

In summary, the experimental results of EFND and CTD are consistent, indicating that mental features play a key role in short text classification, whether false news detection or

TABLE 1: The result of Bert in different methods.

	EFND	CTD
PM	0.265	0.754
NTM	0.460	0.960

TABLE 2: The results with different features in EFND.

EFND	Accuracy	EFND	Accuracy
Only text	0.273	Text + State	0.293
Text + Subject	0.281	Text + Party	0.288
Text + Speaker	0.284	Text + Context	0.288
Text + Job	0.285	Text + Mental	0.459

TABLE 3: The results with different features in CTD.

CTD	Accuracy	CTD	Accuracy
Text + Mental	0.960	Only text	0.780

TABLE 4: The result of different methods in EFND.

	Accuracy	
Model	Text + Mental	Only text
Bert	0.460	0.273
Bays	0.243	0.236
SVM	0.259	0.256
Xgboost	0.250	0.217
KNN	0.222	0.206
CNN	0.165	0.155
LSTM	0.166	0.155
GRU	0.164	0.1437
BP	0.164	0.159
RNN	0.162	0.158

TABLE 5: The result of different methods in CTD.

	Accuracy	
Model	Text + Mental	Only text
Bert	0.960	0.780
Bays	0.582	0.564
SVM	0.578	0.564
Xgboost	0.577	0.575
KNN	0.564	0.397
CNN	0.550	0.497
LSTM	0.529	0.525
GRU	0.552	0.527
BP	0.533	0.527
RNN	0.512	0.496

topic detection. In especial, our method, regardless of whether it has mental features or not, the result is the best. Therefore, Bert not only has higher accuracy but also has universality. In other words, it can be applied to cross-linguistic and multidomain.

5. Discussion

5.1. Key Findings. The purpose of this study is to explore the role of mental features in short text classification. In general, our findings show that there are two factors in the mental feature of information publishers. The first kind of factor is an external mental feature, that is, historical information. The data show that the classification result of the model is more accurate under the joint action of short text and historical information. This shows that the authenticity of the short text of the information published is affected by its subject. For those who publish false reviews more often in history, the probability of new short texts being false will increase. By combining the effective features extracted from short text, we can better understand the authenticity of publishers.

The second type of factor is the internal mental feature, that is, Maslow's need. The results show that this feature is related to the accuracy of short text classification. This feature reflects the publisher's deep-seated aspects, that is, psychological changes. This shows that the content of short texts can be enriched by depicting the psychological changes of information publishers. People will have different psychological changes when publishing different texts. By taking psychological changes as auxiliary features, they can map with text features. It can improve the accuracy of text topic classification.

In summary, short text cannot completely reflect the meaning that users want to express. Through mental features, users' intentions expressed in short texts can be enriched.

5.2. Theoretical Implications. We apply mental features to short text classification. Specifically, we prove through experiments that the improvement of short text accuracy is achieved through the integration of mental and text features. In addition, through experiments, we find the form in which these two features are input into Bert to achieve more effective fusion.

Our results are better than traditional methods. To improve the accuracy of classification results, learning mental features through the Bert model and recognizing their relationship with short texts. This provides an innovative perspective and enriches the literature of short text classification.

5.3. Practical Implications. False review identification is very important. With the vigorous development of the Internet economy, the credibility of reviews is of great significance to consumers. False reviews imitate the tone of real reviews, which makes it difficult to distinguish between true and false. Its content distorts the facts and misleads consumers, which has a great negative impact on the interests of consumers and the platform. From an academic point of view, there is still a huge research space for the exploration of this direction, which is very worthy of in-depth excavation by interested researchers. By mining more consumers' mental features and depicting users' mental portraits, the users' credibility can obtain. Combining it with the text content, it is expected to effectively identify the authenticity of reviews.

Text topic classification method is the key technical basis of network public opinion analysis. Internet public opinion refers to people's opinions or remarks with certain influence and tendency on the Internet with the help of Internet media. With the rapid growth of communication technology and intelligent devices, there has been a huge surge in data traffic. Different applications, users, and devices generate large amounts of data every second [63]. Once the wrong or extreme public opinion is spread, with its influence in the network world, it will often cause huge public opinion pressure and even uncontrollable consequences. Therefore, it is necessary to control the dynamics of public opinion. Through the in-depth mining of users' mental features, the performance features of network public opinion can be reflected. It is hoped that it can provide a reference for effectively understanding the evolutionary process of network public opinion.

6. Conclusion

Due to the sparsity of the short text, it is difficult for the machine to understand its content. We find two mental feature that reflect the information publisher. By integrating this feature with the text, the accuracy of short text topic classification can be improved. In this context, we propose a method which effectively integrates mental features with text content on a linguistic level. Specially, we have designed two input methods to integrate text context and feature on a linguistic level. Namely, the pair method and new text method. Also, to verify the accuracy of Bert method, we compare a variety of machine-learning methods, such as SVM, CNN, and RNN. The results show that (1) through feature comparison, it is found that mental features can significantly improve the accuracy of short text classification. (2) Combining mental features and text as input vectors can provide more classification accuracy than separating them as two independent vectors. Namely, the new text method is better than the pair method. (3) Through model comparison, it can be found that the Bert model can integrate mental features and short text. Bert can better capture mental features to improve the accuracy of classification results.

There are still some limitations in this paper. Our results demonstrate the effectiveness of mental features in false review detection and topic classification. However, in different situations, the interference factors are different. Therefore, it will be interesting to test our method for this study in other contexts. We encourage other researchers to take mental features as a meaningful framework and integrate different features to improve the accuracy of short text classification. In addition, this paper classifies short text on offline data sets. Therefore, in the future, we expect to automatically record users' daily information data to facilitate real-time analysis of users' mental status. In this way, short text can be detected in real-time.

Data Availability

For English fake news detection (EFND), the data set we experimented with is based on the LIAR data set (http://www. cs.ucsb.edu/~william/data/liar_dataset.zip); For Chinese topic

detection (CTD), the data set we experimented with is based on social messaging data. We welcome interested partners to contact the first author for data (hyjsdu96@126.com).

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Authors' Contributions

Conceptualization was performed by Yongjun Hu and Huiyou Chang. Methodology was contributed by Jia Ding, Zixin Dou, and Yongjun Hu. Software was implemented by Jia Ding and Zixin Dou. Validation was done by Yongjun Hu and Huiyou Chang. Formal analysis was performed by Zixin Dou. Investigation was contributed by Yongjun Hu. Resources were provided by Yongjun Hu and Zixin Dou. Data curation was done by Yongjun Hu. The original draft was prepared by Jia Ding and Zixin Dou. Review and editing were done by Zixin Dou and Yongjun Hu. Supervision was performed by Yongjun Hu and Huiyou Chang. Funding acquisition was contributed by Yongjun Hu. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

References

[1] WT. Li, SB. Gao, H. Zhou, ZH. Huang, and W. Li, "The automatic text classification method based on BERT and feature union," in *Proceedings of the IEEE 25th International Conference on Parallel and Distributed Systems (ICPADS)*, Tianjin, China, December 2019.

[2] S. Takeshi, O. Makoto, and M. Yutaka, "Earthquake shakes twitter users: real-time event detection by social sensors," in *Proceedings of the 19th International Conference on World Wide Web*, pp. 851–860, ACM, Raleigh North, CA, USA, April 2010.

[3] V. Sarah, LH. Amanda, S. Kate, and P. Leysia, "Microblogging during two natural hazards events: what twitter may contribute to situational awareness," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1079–1088, ACM, Atlanta, GA, USA, 2010.

[4] DL. Altheide, "Qualitative media analysis," *Qualitative Research Methods*, vol. 38, 1996.

[5] MU. Khan, AR. Javed, M. Ihsan, and T. Usman, "A novel category detection of social media reviews in the restaurant industry," *Multimedia Systems*, 2020.

[6] JJ. Bernard, Z. Mimi, S. Kate, and C. Abdur, "Twitter power: tweets as electronic word of mouth," *Journal of the American Society for Information Science and Technology*, vol. 60, no. 11, pp. 2169–2188, 2009.

[7] A. Tumasjan, TO. Sprenger, PG. Sandner, and IM. Welpe, "Predicting elections with twitter: what 140 characters reveal about political sentiment," in *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media*, pp. 178–185, Washington, DC, USA, May 2010.

[8] LJ. Huang, ZX. Dou, YJ. Hu, and RY. Huang, "Textual analysis for online reviews: a polymerization topic sentiment model," *IEEE Access*, vol. 99, p. 1, 2019.

[9] L. Huang, Z. Dou, Y. Hu, and R. Huang, "Online sales prediction: an analysis with dependency scor-topic sentiment model," *IEEE Access*, vol. 7, pp. 79791–79797, 2019.

[10] ZJ. Gao, A. Feng, XY. Song, and X. Wu, "Target-dependent sentiment classification with bert," *IEEE Access*, vol. 8, 2019.

[11] A. Moldagulova and RB. Sulaiman, "Document classification based on KNN algorithm by term vector space reduction," in *Proceedings of the 18th international conference on control, automation and systems (ICCAS)*, pp. 387–391, PyeongChang, Repubic of Korea, October 2018.

[12] J. Zhu, H. Wang, and X. Zhang, "Discrimination-based feature selection for multinomial naïve bayes text classification," *Computer Processing of Oriental Languages. Beyond the Orient: The Research Challenges Ahead*, vol. 4285, pp. 149–156, 2006.

[13] X. Luo, "Efficient English text classification using selected Machine Learning Techniques," *Alexandria Engineering Journal*, vol. 60, no. 3, pp. 3401–3409, 2021.

[14] G. Nadia and M. Andrew, "Collective multi-label classification," in *Proceedings of the International Conference on Information and Knowledge Management*, pp. 195–200, Bremen, Germany, October 2005.

[15] I. Katakis, G. Tsoumakas, and I. Vlahavas, "Multilabel text classification for automated tag suggestion," *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, vol. 32, no. 5, pp. 75–83, 2008.

[16] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," in *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pp. 655–665, Baltimore, Maryland, June 2014.

[17] K Shuang, H Guo, ZX Zhang, J Loo, and S Su, "A word-building method based on neural network for text classification," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 31, no. 2, pp. 455–474, 2019.

[18] K. Yoon, "Convolutional neural networks for sentence classification," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 1746–1751, Doha, Qatar, October 2014.

[19] M. Tomas, K. Martin, B. Lukas, C. Jan, and K. Sanjeev, "Recurrent neural network based language model," in *Proceedings of the 11th Annual Conference of the International Speech Communication Association*, pp. 1045–1048, Makuhari, Chiba, Japan, September 2010.

[20] J. Devlin, MW. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 4171–4186, Minneapolis, Minnesota, June 2019.

[21] W. Zehra, A. R. Javed, Z. Jalil, U. K. Habib, and R. G. Thippa, "Cross corpus multi-lingual speech emotion recognition using ensemble learning," *Complex and Intelligent Systems*, vol. 7, pp. 1845–1854, 2021.

[22] KW. Craik, *The Nature of Explanation*, Cambridge University Press, Cambridge UK, 1943.

[23] TA. Dijk, *Society and Discourse: How Social Contexts Influence Text and Talk*, Cambridge University Press, Cambridge, UK, 2009.

[24] TA. Dijk, *Discourse and Knowledge: A Sociocognitive Approach*, Cambridge University Press, Cambridge,. UK, 2014.

[25] PN. Johnson-Laird, *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*, Harvard University Press, Cambridge, MA, USA, 1983.

[26] R. Hogan and R. Warrenfeltz, "Educating the modern manager," *The Academy of Management Learning and Education*, vol. 2, no. 1, pp. 74–84, 2003.

[27] MS. Gary and R. Wood, "Mental models, decision rules, and performance heterogeneity," *Strategic Management Journal*, vol. 32, no. 6, pp. 569–594, 2011.

[28] LM. Palmunen, T. Lainema, and E. Pelto, "Towards a manager's mental model: conceptual change through business simulation," *International Journal of Management in Education*, vol. 19, no. 2, 2021.

[29] RM. Best, MP. Rowe, Y. Ozuru, and DS. McNamara, "Deep-level comprehension of science texts: the role of the reader and the text," *Topics in Language Disorders*, vol. 25, pp. 65–83, 2005.

[30] J. Dunlosky and KA. Rawson, "Why does rereading improve metacomprehension accuracy? Evaluating the levels-of-disruption hypothesis for the rereading effect," *Discourse Processes*, vol. 40, pp. 37–55, 2005.

[31] LG. Hathorn and KA. Rawson, "The roles of embedded monitoring requests and questions in improving mental models of computer-based scientific text," *Computers & Education*, vol. 59, no. 3, pp. 1021–1031, 2012.

[32] M. Magzan, "Mental models for leadership effectiveness: building future different than the past," *Journal of Engineering Management and Competitiveness*, vol. 2, no. 2, pp. 57–63, 2012.

[33] M. Arthur, "GlenbergMarion MeyerKaren LindemMental models contribute to foregrounding during text comprehension," *Journal of Memory and Language*, vol. 26, no. 1, pp. 69–83, 1987.

[34] Y. Tawatsuji, T. Uno, SY. Fang, and T. Matsui, "Real-time estimation of learners' mental states from learners' physiological information using deep learning," in *Proceedings of the 26th International Conference on Computers in Education (ICCE 2018)*, pp. 107–109, Metro Manila, Philippines, November 2018.

[35] T. Matsui, Y. Tawatsuji, S. Fang, and T. Uno, "Conceptualization of IMS that estimates learners' mental states from learners' physiological information using deep neural network algorithm," in *Proceedings of the International Conference on Intelligent Tutoring Systems*, pp. 63–71, Kingston, Jamaica, June 2019.

[36] A. Darwish and HJ. Steinhauer, "Learning individual driver's mental models using POMDPs and BToM," in *Proceedings of the 6th International Digital Human Modeling Symposium*, pp. 51–60, Skövde, Sweden, September 2020.

[37] S. Dutta and A. Nandy, "An extensive analysis on deep neural architecture for classification of subject-independent cognitive states," in *Proceedings of the 7th Acm Ikdd Cods And 25th Comad*, pp. 180–184, Hyderabad India, January 2020.

[38] B. Trstenjak, S. Mikac, and D. Donko, "KNN with TF-IDF based framework for text categorization," *Proc Eng*, vol. 69, pp. 1356–1364, 2014.

[39] FF. Wang, Y. Jia, ZJ. Liu, and M. Kong, YF. Wu, Study on offshore seabed sediment classification based on particle size parameters using XGBoost algorithm," *Computers & Geosciences*, vol. 149, no. 4, Article ID 104713, 2021.

[40] SL. Li and XJ. Zhang, "Research on orthopedic auxiliary classification and prediction model based on XGBoost algorithm," *Neural Computing and Applications*, vol. 32, no. 59, pp. 1971–1979, 2020.

[41] Y. Jiang, H. Lin, X. Wang, and D. Lu, "A Technique for improving the performance of Naive Bayes text classification," in *Proceedings of the Lecture notes in computer science*, pp. 196–203, Tokyo, Japan, February 2011.

[42] M. Bilal, H. Israr, M. Shahid, and A. Khan, "Sentiment classification of roman-Urdu opinions using naïve bayesian, decision tree and KNN classification techniques," *Journal of King Saud University - Computer and Information Sciences*, vol. 28, no. 3, pp. 330–344, 2016.

[43] Y. Ke and M. Hagiwara, "An English neural network that learns texts, finds hidden knowledge, and answers questions," *Journal of Artificial Intelligence and Soft Computing Research*, vol. 7, no. 4, pp. 229–242, 2017.

[44] B. Ghaddar and J. Naoum-Sawaya, "High dimensional data classification and feature selection using support vector machines," *European Journal of Operational Research*, vol. 265, no. 3, pp. 993–1004, 2018.

[45] S. Vijayarani, MJ. Ilamathi, and M. Nithya, "Preprocessing techniques for text mining-an overview," *International Journal of Computer Science and Communication Networks*, vol. 5, no. 1, pp. 7–16, 2015.

[46] L. Jiang, M. Yu, M. Zhou, XH. Liu, and TJ. Zhao, "Target-dependent Twitter sentiment classification," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp. 151–160, Portland, OR, USA, June 2011.

[47] S. Kiritchenko, X. Zhu, C. Cherry, and S. Mohammad, "NRC-Canada- 2014: detecting aspects and sentiment in customer reviews," in *Proceedings of the 8th International Workshop on Semantic Evaluation*, pp. 437–442, Dublin, Ireland, August 2014.

[48] J. Wagner, P. Arora, S. Cortes et al., "Dcu: aspect-based polarity classification for semeval task 4," in *Proceedings of the 8th International Workshop on Semantic Evaluation*, pp. 223–229, Dublin, Ireland, August 2014.

[49] P. Chen, ZQ. Sun, LD. Bing, and W. Yang, "Recurrent attention network on memory for aspect sentiment analysis," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Process*, pp. 452–461, Copenhagen, Denmark, September 2017.

[50] B. Huang and K. Carley, "Parameterized convolutional neural networks for aspect level sentiment classification," in *Proceedings of the Conf. Empirical Methods Natural Lang. Process*, pp. 1091–1096, Brussels, Belgium, October 2018.

[51] N. Kalchbrenner and P. Blunsom, "Recurrent continuous translation models," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 1700–1709, Washington, DC, USA, October 2013.

[52] W. Xue and T. Li, "Aspect based sentiment analysis with gated convolutional networks," in *Proceedings of the Meeting of the Association for Computational Linguistics*, Melbourne, Australia, July 2018.

[53] M. Zhang, Y. Zhang, and D. Vo, "Gated neural networks for targeted sentiment analysis," in *Proceedings of the 13th AAAI Conference on Artificial Intelligence*, pp. 3087–3093, Phoenix Arizona, February 2016.

[54] YL. Shen, XD. He, JF. Gao, L. Deng, and G. Mesnil, "Learning semantic representations using convolutional neural networks for web search," in *Proceedings of the 23rd international conference on World wide web companion*, pp. 373-374, Seoul, Republic of Korea, April 2014.

[55] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A Convolutional neural network for modelling sentences," in *Proceedings of the 52th Annual Meeting of the Association for Computational Linguistics*, Baltimore, Maryland, June 2014.

[56] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[57] IA. Kandhro, SZ. Jumani, K. Kamlash, A. Hafeez, and F. Ali, "Roman Urdu headline news text classification using RNN, LSTM and CNN," *Advances in Data Science and Adaptive Analysis*, vol. 12, no. 2, 2020.

[58] DY. Tang, B. Qin, XC. Feng, and T. Liu, "Effective LSTMs for targetdependent sentiment classification," *Computer Science*, vol. 7, p. 3, 2015.

[59] H. Lee and Y. Kang, "Mining tourists' destinations and preferences through LSTM-based text classification and spatial clustering using Flickr data," *Spatial Information Research*, vol. 29, no. 7, pp. 825–839, 2021.

[60] JY. Khan, TI. Khondaker, S. Afroz, G. Uddin, and A. Iqbal, "A benchmark study of machine learning models for online fake news detection," *Machine Learning with Applications*, vol. 4, pp. 1–12, 2021.

[61] TT. Yang, F. Li, DH. Ji, and XH. Liang, "Fine-grained depression analysis based on Chinese micro-blog reviews," *Information Processing & Management*, vol. 58, no. 6, Article ID 102681, 2021.

[62] WY. Wang, "Liar, liar pants on fire: a new benchmark dataset for fake news detection," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, Vancouver, Canada, January 2017.

[63] A. Abbasi, AR. Javed, C. Chakraborty, J. Nebhen, W. Zehra, and Z. Jalil, "ElStream: an ensemble learning approach for concept drift detection in dynamic social big data stream learning," *IEEE Access*, vol. 9, pp. 66408–66419, 2021.