IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 5 March 2023, accepted 30 March 2023, date of publication 10 April 2023, date of current version 13 April 2023.
Digital Object Identifier 10.1109/ACCESS.2023.3265712

RESEARCH ARTICLE

Six-Granularity Based Chinese Short Text Classification

XINJIE SUN[1,2], **ZHIFANG LIU**[1], **AND XINGYING HUO**[1]
[1]Institute of Computer Science, Liupanshui Normal University, Liupanshui 553004, China
[2]Guizhou Xinjie Qianxun Software Service Company Ltd., Liupanshui 553004, China

Corresponding author: Xinjie Sun (sxj123lps@163.com)

This work was supported in part by the National Science Research Project of Department of Education in Guizhou Province under Grant KY[2020]112, in part by the Liupanshui Normal University High Level Talent Research Launch Fund under Grant LPSSYKYJJ201708, in part by the Liupanshui Science and Technology Bureau Fund Project 52020-2018-04-15, in part by the Liupanshui Normal University Major Comprehensive Reform Pilot Project LPSSYzyzhggsd202003, in part by the Science and Technology Foundation of Guizhou Province under Grant ZK[2022]528, and in part by the Youth Science and Technology Talent Growth Project of Department of Education in Guizhou Province under Grant KY[2022]054.

ABSTRACT Short text classification is an important task in Natural Language Processing (NLP). The classification result for Chinese short text is always not ideal due to the sparsity problem of them. Most of the previous classification models for Chinese short text are based on word or character, considering that Chinese radical can also represent the meaning individually, so word, character and radical are all used to build a Chinese short text classification model in this paper, which solves the data sparsity problem of short text. In addition, in the process of segmenting sentences into words, considering that jieba will cause the loss of key information and ngram will generate noise words, both jieba and ngram are used to construct a six-granularity (i.e. word-jieba, word-jieba-radical, word-ngram, word-ngram-radical, character and character-radical) based Chinese short text classification (SGCSTC) model. Additionally, different weights are assigned to the six granularities and are automatically updated in the process of back-propagation using cross-entropy loss due to the different influence of them on the classification results. The classification Accuracy, Precision, Recall and F1 of SGCSTC in THUCNews-S dataset are 93.36%, 94.47%, 94.15% and 94.31% respectively, and that in CNT dataset are 92.67%, 92.38%, 93.15% and 92.76% respectively, and multiple comparative experiment results on THUCNews-S and CNT datasets show that SGCSTC outperforms the state-of-the-art text classification models.

INDEX TERMS Chinese short text, six-granularity, weights, data sparsity.

I. INTRODUCTION

With the rise of short messages, online communities and micro-blogs, the network short text has begun to take shape and its development trend has been irreversible. Especially for the short text information represented by instant messages, online chat records, mobile phone short messages, microblog interaction, blog comments, news comments, etc. In order to better understand these short texts, short text classification has become an important research field.

As one of the most important tasks in Natural Language Processing (NLP), short text classification is widely used in sentiment analysis [1], spam filtering [2], commodity evaluation [3], etc. Traditional short text classification methods use Bag-Of-Words (BOW) [4], Ngram [5], or Frequency-Inverse Document Frequency of Terms (TF-IDF) [6] as features, and use Support Vector Machine [7], Naive Bayes [8], and other machine learning models as classifiers. Ahmed et al. [9] proposed a hybrid model based on simulated annealing (SA) and generalized normal distribution optimizer (GNDO). Nurcahyawati et al. [10] proposed a comment classification model based on SVM. Parlak et al. [11] proposed a text classification feature filtering method based on machine learning. Zhao et al. [12] proposed a new bio-inspired optimization algorithm to solve optimization problems.

The traditional methods have been able to perform most of the classification tasks accurately, however, since they rely heavily on feature engineering and the classifica-

The associate editor coordinating the review of this manuscript and approving it for publication was Pasquale de Meo.

tion results are largely depends on the selection of features [13]. In contrast, deep learning model can achieve excellent text classification results without complex feature engineering. Therefore, deep learning models such as Convolutional Neural Network (CNN) [14], Recurrent Neural Networks (RNN) [15], and Attention mechanisms [16] have been widely used in short text classification in recent years. Zhao et al. [17] proposed the Inversed Attention Orthogonal Projection Module, which used inversed attention to iteratively reverse the attention map on text features. Magalhães et al. [18] proposed a text classification model based on CNN, LSTM and Graph Neural Networks (GNN). Huang et al. [19] proposed a text classification model that combined the improved self-attention mechanism with Skip-grate recurrent unit network.

Although there have been some good short text classification models, most of them are for English [16], [20], [21], and there are few excellent Chinese short text classification models. With the rapid development of Chinese social networks, it is urgent to develop short text classification models suitable for Chinese short texts. Among the existing Chinese short text classification models, most of them are designed based on word or character, which is difficult to solve the data sparsity problem of Chinese short texts. Deng et al. [22] proposed a Chinese long text classification model based on BiLSTM, CNN and Attention. Zhai et al. [23] proposed a Chinese news classification model based on multi-scale CNN and LSTM. Hao et al. [24] proposed a mutual-attention convolutional neural networks based Chinese short text classification model. Since radical is an important part of Chinese that can express certain information, and the character with the same radical can often express similar meanings. Therefore, the addition of radical can help solve the data sparsity problem of Chinese short text.

Based on the above reasons, a Chinese short text classification model based on the characteristics of word, character and radical is constructed. In addition, considering that the most common methods for segmenting Chinese sentences into words are jieba and ngram, while jieba tends to lose some important information and ngram tends to produce noisy words, both jieba and ngram are used to segment Chinese sentences. Finally, SGCSTC, a Chinese short text classification model based on the characteristics of six granularities (i.e. word-jieba, word-jieba-radical, word-ngram, word-ngram-radical, character, and character-radical) is constructed. Additionally, considering the different contributions of different granularities to text classification, different weights are set for each of the six granularities, and the cross-entropy loss function are used to continuously optimize them in the process of back-propagation. Finally, a deep learning model based on BIGRU, Attention and CNN is constructed to fully extract the deep hidden features of Chinese short text. Multi group comparison experiment results on two publicly available datasets show that SGCSTC outperforms the state-of-the-art text classification models.

The main contributions of this paper are as follows:

- As far as we know, this paper is the first research work to construct the classification model of Chinese short text by using six gradualities: word-jieba, word-jieba-radical, word-ngram, word-ngram-radical, character and character-radical, which can well solve the sparsity problem of Chinese short text.
- Considering that different gradualities may have different influence on the Chinese short text classification results, SGCSTC assigns different weights to the six gradualities, and automatically optimizes them by using cross-entropy loss in the process of back-propagation.
- A classification model based on BIGRU, Attention and CNN is constructed to fully extract the deep hidden features of Chinese short text. The forward, backward and bidirectional hidden vectors of BIGRU are used for short text classification at the same time, which can further solve the sparsity problem of Chinese short text.
- Extensive experimental results on 2 publicly available datasets show that SGCSTC outperforms the state-of-the-art text classification models.

II. RELATED WORK

A. FEATURE ENGINEERING METHODS

The feature engineering method first extracts the n-gram, TF-IDF or other features of the text, and then classifies them through machine learning models.

Ahmed et al. [9] proposed a binary simulated normal distribution optimizer (BSNDO), which is a hybrid version of simulated annealing (SA) and generalized normal distribution optimizer (GNDO). BSNDO uses SA as local search to achieve higher classification accuracy. Multiple comparison experiments on multiple datasets demonstrated the effectiveness of BSNDO as a feature selection method. In order to solve the highly nonlinear optimization problem, Yuan et al. [25] proposed a learning and chaos k-optimal gravity search strategy (EOCS) based on elite opposition for the Grey Wolf Optimizer (GWO) algorithm. In the EOCS based grey wolf optimizer (EOCSGWO) algorithm, the elite opposition-based learning strategy (EOBLS) is proposed to take full advantage of better-performing particles for optimization in the next generations. Zhao et al. [12] proposed an artificial hummingbird algorithm (AHA) to solve optimization problems. AHA simulated the special flight skills and intelligent foraging strategies of hummingbirds in nature, and axial flight, diagonal flight and omnidirectional flight techniques used in foraging strategy were modeled. Yuan et al. [26] proposed a new swarm intelligence optimization algorithm called alpine skiing optimization (ASO). In ASO, physical fitness and sprint were two important factors for skiers to win the competition. Skiers revealed the behavior of winning the competition according to static sliding and dynamic sliding. The statistical results confirmed that ASO could provide competitive results compared with other

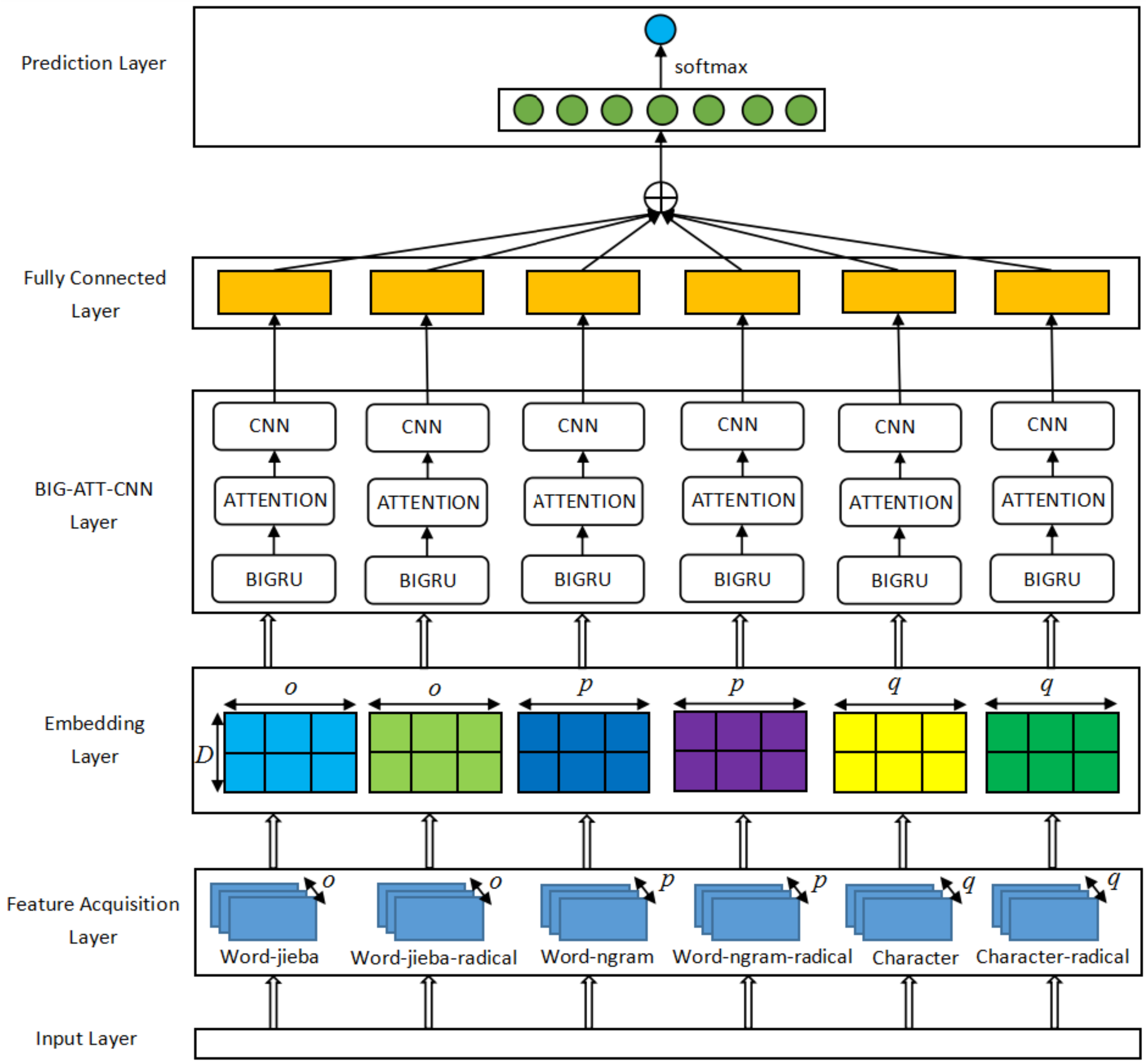

FIGURE 1. The structure of SGCSTC.

most advanced optimization algorithms. Luo [27] implemented a SVM model for classifying English texts and documents. Their experimental results on a dataset containing 1033 documents showed that SVM had excellent classification performance, and Rocchio achieved the optimal classification results when the dataset was small. Chia et al. [28] explored the method of using machine learning algorithms to classify satirical texts on Twitter, and obtained ideal classification results. Thirumoorthy et al. [29] constructed a feature selection method based on term frequency distribution measurement, and used Naive Bayes and SVM for text classification. The classification results on two common datasets proved that their feature selection method had better classification accuracy. Aljedani et al. [30] proposed a machine learning algorithm for multi label text classification based on Arabic, and analyzed the impact of feature selection methods and feature set dimensions on the classification results. The experimental results showed that the proposed method had superior classification accuracy and lower classification cost. Janani et al. [31] proposed a text classification method based on feature selection optimization algorithm and machine learning model. The algorithm had good performance in feature optimization and text classification.

B. DEEP LEARNING METHODS

Although machine learning algorithms often have good classification performance, however, feature selection is often unsatisfactory, and complex feature selection requires a lot of tedious work, which makes the classification effect of text classification model based on feature engineering less satisfactory. As the deep learning model does not require tedious feature engineering, more and more researchers use the deep learning model for text classification.

Mohammed et al. [32] put forward an integrated learning model integrating the baseline deep learning model for text classification, and the classification results of the proposed method on the six benchmark datasets were better than other baseline integrated learning methods. Akhter et al. [33] studied the effect of four baseline depth learning models in Urdu text classification, and discussed the impact of different text preprocessing schemes on the classification results. The experimental results showed that CNN achieved the best classification results. Ibrahim et al. [34] proposed a hybrid multi label classification method based on general deep learning. The model used CNN to extract the most discriminating features and BiLSTM to obtain context information, and achieved significant recall and F1 improvement on the public medical dataset. Liang et al. [35] proposed a text classification model based on spatial view attention CNN, which used a set of innovative and well-designed multi view representation learning, combined heterogeneous attention mechanisms and CNN based operations to automatically extract and weight multi granularity and fine granularity representations. The proposed model achieved the optimal classification results on five common datasets. Al-Garadi et al. [36] experimented with the most advanced language model based on bidirectional converter, which used tweet level representation to support transfer learning for text classification. The experimental results on the public dataset proved that the transformer based model was more stable and required less annotation data.

C. CHINESE TEXT CLASSIFICATION

Although the deep learning model has made great achievements in text classification, most of them are designed based on English, which cannot be directly applied to Chinese because of the differences between Chinese and English structures. Therefore, it is necessary to develop Chinese based text classification models with the wide application of Chinese text in Internet system.

Deng et al. [22] proposed a Chinese long text classification model based on Attention mechanism, BILSTM and CNN, and achieved the best classification results on two Chinese long text classification datasets. Cai et al. [37] proposed an entity-level knowledge-enhanced pre-trained language model, which leverages several distinct self-supervised tasks for Chinese medical text mining. Liang et al. [38] proposed an integrated model of Chinese text classification based on improved CNN and RNN, and the classification error rates on SINA_P and THUC_P news datasets decreased by 8.36% and 8.28% respectively.

Unlike long texts, Chinese short texts often have the problem of data sparsity, which is a problem that researchers need to focus on. Wang et al. [39] proposed a Chinese short text classification model based on CNN and semantic expansion, which improved the similarity measurement method to improve the coverage of word vector table during short text preprocessing, and added attention mechanism to find related words in short text. Zhang et al. [40] proposed a Chinese short text emotion classification model based on ELECTRA, self-attention mechanism and BiLSTM, and significantly improved the classification accuracy. Feng et al. [41] proposed a multi-channel CNN model based on multi-head attention mechanism to solve the data sparsity problem of Chinese short texts by mining short text features from multiple perspectives.

Almost all the existing Chinese short text classification models are based on word or character, since single-mode feature extraction is difficult to solve the data sparsity problem of short texts, therefore, their classification results are not that ideal. To solve the above problems, six-granularity (i.e. word-jieba, word-jieba-radical, word-ngram, word-ngram-radical, character and character-radical) based Chinese short text classification model is constructed.

III. METHODOLOGY

SGCSTC is constructed for Chinese short text classification, and the structure of it is shown in Figure 1. As can be seen from Figure 1, SGCSTC consists of input layer, feature acquisition layer, embedding layer, BIG-ATT-CNN layer, fully connected layer and prediction layer. Feature acquisition layer extracts the Chinese short text of input layer into six granularity features: word-jieba, word-jieba-radical, word-ngram, word-ngram-radical, character and character-radical. Embedding layer converts the features of six granularities into vectors, respectively. BIG-ATT-CNN layer is a deep learning model based on BIGRU, Attention and CNN. Fully connected layer and prediction layer are used to get the prediction results of Chinese short text.

A. FEATURE ACQUISITION LAYER

Feature acquisition layer is used to convert the Chinese short text of the input layer into six granularity features. The specific process is shown in Figure 2. Given a Chinese short text T containing q characters, first cut T into q characters $L^c = \{c^1, c^2, \ldots, c^q\}$. Then, jieba and ngram are used to split T into two granularity features, i.e. word-jieba feature L^{wj} and word-ngram feature L^{wn}. Assume that L^{wj} contains o words and L^{wn} contains p words, i.e. $L^{wj} = \{wj^1, wj^2, \ldots, wj^o\}$, $L^{wn} = \{wn^1, wn^2, \ldots, wn^p\}$. For stop-words will be discarded in the process of using jieba, and the word obtained by word segmentation contains at least one character, therefore, $o < q$. Assume that the sliding window size of ngram is $n\,(2 \leqslant n \leqslant q)$, then $p = q - n + 1$ and $p < q$. Since a word and character correspond to a radical, after getting the radical of word-jieba, word-ngram and character, respectively, the features $L^{wjr} = \{wjr^1, wjr^2, \ldots, wjr^o\}$, $L^{wnr} = \{wnr^1, wnr^2, \ldots, wnr^p\}$ and $L^{cr} = \{cr^1, cr^2, \ldots, cr^q\}$ of word-jieba-radical, word-ngram-radical and character-radical are respectively obtained.

B. EMBEDDING LAYER

Embedding layer aims to convert the features of the six gradualities obtained by the feature acquisition layer into embedding vectors. Since a word and character correspond to a radical, therefore, $|L^{wj}| = |L^{wjr}| = o$, $|L^{wn}| = |L^{wnr}| = p$, $|L^c| = |L^{cr}| = q$. For the convenience of calculation,

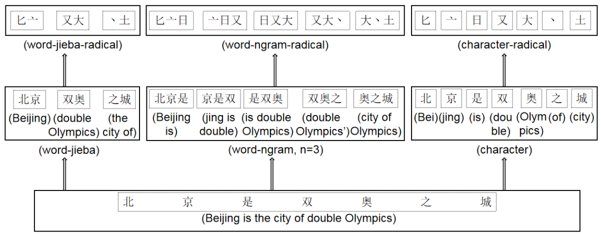

FIGURE 2. The structure of SGCSTC.

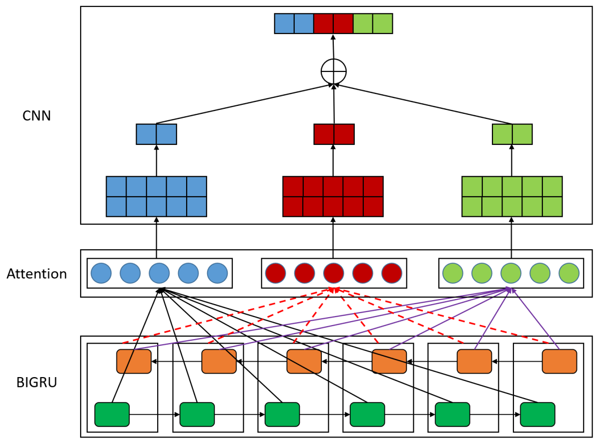

FIGURE 3. The structure of the BIG-ATT-CNN layer.

assuming that each word, character and radical is transformed into a vector with a length of D, the transformed embedding vectors of the six gradualities of word-Jieba, word-Jieba-radical, word-ngram, word-ngram-radical, character and character-radical are $V^{wj} = \{wj_v^1, wj_v^2, \ldots, wj_v^o\}$, $V^{wjr} = \{wjr_v^1, wjr_v^2, \ldots, wjr_v^o\}$, $V^{wn} = \{wn_v^1, wn_v^2, \ldots, wn_v^p\}$, $V^{wnr} = \{wnr_v^1, wnr_v^2, \ldots, wnr_v^p\}$, $V^c = \{c_v^1, c_v^2, \ldots, c_v^q\}$, $V^{cr} = \{cr_v^1, cr_v^2, \ldots, cr_v^q\}$, respectively, and $V^{wj} \in R^{o\times D}$, $V^{wjr} \in R^{o\times D}$, $V^{wn} \in R^{p\times D}$, $V^{wnr} \in R^{p\times D}$, $V^c \in R^{q\times D}$, $V^{cr} \in R^{q\times D}$.

C. BIG-ATT-CNN LAYER

BIG-ATT-CNN layer consists of BIGRU, Attention and CNN, and the structure is shown in Figure 3. Gate Recurrent Unit (GRU) is a variant of RNN that can be used to solve the gradient disappearance problem during long-term memory and back-propagation [45]. Since GRU can only perform forward propagation and tends to ignore the important information that appears after the current time node, therefore, BIGRU is proposed to use both forward and backward informations to obtain the output of the current time node. Specifically, given the embedding vector $V = \{v_1, v_2, \ldots, v_l\}$ of length l, The forward and backward hidden vectors of BIGRU are calculated as follows:

$$\overrightarrow{h_i} = \overrightarrow{GRU(v_i)} \quad (i = 1, 2, \ldots, l) \tag{1}$$

$$\overleftarrow{h_i} = \overleftarrow{GRU(v_i)} \quad (i = l, l-1, \ldots, 1) \tag{2}$$

where $\overrightarrow{h_i}$ and $\overleftarrow{h_i}$ are the forward and backword hidden vectors respectively, and v_i is the output of the $i - th$ step. $\overrightarrow{h_i}$ and $\overleftarrow{h_i}$ combined to obtain bidirectional long dependence h_i:

$$h_i = \left[\overrightarrow{h_i}, \overleftarrow{h_i}\right] \tag{3}$$

$\overrightarrow{h_i}$, $\overleftarrow{h_i}$ and h_i are input into three parallel Attention networks respectively, so that the forward, backward and bidirectional hidden vectors can be used for short text classification at the same time, so as to effectively solve the sparsity problem of short text. Specifically, first, the hidden representations $\overrightarrow{u_i}$, $\overleftarrow{u_i}$ and u_i of $\overrightarrow{h_i}$, $\overleftarrow{h_i}$ and h_i are obtained respectively:

$$\overrightarrow{u_i} = tanh(W_f \overrightarrow{h_i} + b_f) \tag{4}$$

$$\overleftarrow{u_i} = tanh(W_b \overleftarrow{h_i} + b_b) \tag{5}$$

$$u_i = tanh(W_d h_i + b_d) \tag{6}$$

where W_f, W_b and W_d are weights and b_f, b_b and b_d are biases, $tanh(.)$ is activation function. Then, the importance of $\overrightarrow{u_i}$, $\overleftarrow{u_i}$ and u_i are calculated according to their similarity with $\overrightarrow{d_i}$, $\overleftarrow{d_i}$ and d_i. $\overrightarrow{d_i}$, $\overleftarrow{d_i}$ and d_i are word level context vectors which are randomly initialized and jointly learned during the training process. Then, softmax function is used for normalization to obtain normalized weight:

$$\overrightarrow{a_i} = \frac{exp(\overrightarrow{u_i}\,\overrightarrow{d_i})}{\sum_{i=1}^{l} exp(\overrightarrow{u_i}\,\overrightarrow{d_i})} \tag{7}$$

$$\overleftarrow{a_i} = \frac{exp(\overleftarrow{u_i}\,\overleftarrow{d_i})}{\sum_{i=1}^{l} exp(\overleftarrow{u_i}\,\overleftarrow{d_i})} \tag{8}$$

$$a_i = \frac{exp(u_i d_i)}{\sum_{i=1}^{l} exp(u_i d_i)} \tag{9}$$

where $\overrightarrow{a_i}$, $\overleftarrow{a_i}$ and a_i are forward, backward and bidirectional normalized weight respectively. Finally, forward, backward and bidirectional context representations $\overrightarrow{c_i}$, $\overleftarrow{c_i}$ and c_i are calculated through the weighted summation of $\overrightarrow{a_i}$, $\overleftarrow{a_i}$ and a_i, respectively:

$$\overrightarrow{c_i} = \sum_{i=1}^{l} \overrightarrow{a_i}\,\overrightarrow{h_i} \tag{10}$$

$$\overleftarrow{c_i} = \sum_{i=1}^{l} \overleftarrow{a_i}\,\overleftarrow{h_i} \tag{11}$$

$$c_i = \sum_{i=1}^{l} a_i h_i \tag{12}$$

Input $\overrightarrow{c_i}$, $\overleftarrow{c_i}$ and c_i into three parallel convolution layers in CNN respectively to obtain forward, backward and bidirectional convolution results:

$$\overrightarrow{t_i} = relu(W_c^f \overrightarrow{c_i} + b_c^f) \tag{13}$$

$$\overleftarrow{t_i} = relu(W_c^b \overleftarrow{c_i} + b_c^b) \tag{14}$$

$$t_i = relu(W_c^d c_i + b_c^d) \tag{15}$$

where W_c^f, W_c^b and W_c^d are weights and b_c^f, b_c^b and b_c^d are biases, $relu(.)$ is activation function. Input $\overrightarrow{t_i}$, $\overleftarrow{t_i}$ and t_i into three parallel max-pooling layers to obtain $\overrightarrow{k_i}$, $\overleftarrow{k_i}$ and k_i:

$$\overrightarrow{k_i} = max\left\{\overrightarrow{t_i}\right\} \tag{16}$$

$$\overleftarrow{k_i} = max\left\{\overleftarrow{t_i}\right\} \tag{17}$$

$$k_i = max\left\{t_i\right\} \tag{18}$$

Concatenating $\overrightarrow{k_i}$, $\overleftarrow{k_i}$ and k_i together to obtain the output vector $f_i = \overrightarrow{k_i} \oplus \overleftarrow{k_i} \oplus k_i$ of BIG-ATT-CNN layer, where $\oplus$ represents the row connection of matrix.

D. FULLY CONNECTED LAYER

Suppose that the output vectors of the six granularity embedding vectors (i.e. V^{wj}, V^{wjr}, V^{wn}, V^{wnr}, V^c, V^{cr}) after passing through the BIG-ATT-CNN layer are F^{wj}, F^{wjr}, F^{wn}, F^{wnr}, F^c and F^{cr} respectively, and input them into the fully connected layer to obtain the output results:

$$X^{wj} = tanh(F^{wj}W^{wj}) \tag{19}$$

$$X^{wjr} = tanh(F^{wjr}W^{wjr}) \tag{20}$$

$$X^{wn} = tanh(F^{wn}W^{wn}) \tag{21}$$

$$X^{wnr} = tanh(F^{wnr}W^{wnr}) \tag{22}$$

$$X^{c} = tanh(F^{c}W^{c}) \tag{23}$$

$$X^{cr} = tanh(F^{cr}W^{cr}) \tag{24}$$

where W^{wj}, W^{wjr}, W^{wn}, W^{wnr}, W^c and W^{cr} are weights, $tanh(.)$ is activation function.

E. PREDICTION LAYER

Concatenating X^{wj}, X^{wjr}, X^{wn}, X^{wnr}, X^c and X^{cr} to get $X = X^{wj} \oplus X^{wjr} \oplus X^{wn} \oplus X^{wnr} \oplus X^c \oplus X^{cr}$, and the softmax function is used to get the classification result $class$:

$$class = argmax(softmax(X)) \tag{25}$$

F. LOSS FUNCTION

One-hot coding is used to represent the category label, and the cross-entropy loss function is used to calculate the loss in the process of back-propagation:

$$\begin{aligned} loss = &-\alpha loss(word-jieba) \\ &-\varepsilon loss(word-jieba-radical) \\ &-\beta loss(word-ngram) \\ &-\zeta loss(word-ngram-radical) \\ &-\gamma loss(character) \\ &-\delta loss(character-radical) \end{aligned} \tag{26}$$

$$loss(word-jieba) = \sum_{i=1}^{n}\sum_{e=1}^{c} y_{i,e} ln x_{i,e}^{word-jieba} \tag{27}$$

$$\begin{aligned} &loss(word-jieba-radical) \\ &= \sum_{i=1}^{n}\sum_{e=1}^{c} y_{i,e} ln x_{i,e}^{word-jieba-radical} \end{aligned} \tag{28}$$

$$loss(word-ngram) = \sum_{i=1}^{n}\sum_{e=1}^{c} y_{i,e} ln x_{i,e}^{word-ngram} \tag{29}$$

$$\begin{aligned} &loss(word-ngram-radical) \\ &= \sum_{i=1}^{n}\sum_{e=1}^{c} y_{i,e} ln x_{i,e}^{word-ngram-radical} \end{aligned} \tag{30}$$

$$loss(character) = \sum_{i=1}^{n}\sum_{e=1}^{c} y_{i,e} ln x_{i,e}^{character} \tag{31}$$

$$\begin{aligned} &loss(character-radical) \\ &= \sum_{i=1}^{n}\sum_{e=1}^{c} y_{i,e} ln x_{i,e}^{character-radical} \end{aligned} \tag{32}$$

where $loss$ is the overall loss value, $loss(word-jieba)$, $loss(word-jieba-radical)$, $loss(word-ngram)$, $loss(word-ngram-radical)$, $loss(character)$ and $loss(character-radical)$ are the loss values of the 6 granularities respectively. $\alpha, \varepsilon, \beta, \zeta, \gamma, \delta \in [0, 1]$ and $\alpha + \varepsilon + \beta + \zeta + \gamma + \delta = 1$. $\alpha, \varepsilon, \beta, \zeta, \gamma, \delta$ are randomly initialized and continuously optimized in the process of back-propagation. c is the total number of categories. $y_{i,e}$ is the tag value of the $i-th$ sample belonging to class e, and when the $i-th$ sample belongs to class e, $y_{i,e} = 1$, otherwise $y_{i,e} = 0$. $x_{i,e}^{word-jieba}$, $x_{i,e}^{word-jieba-radical}$, $x_{i,e}^{word-ngram}$, $x_{i,e}^{word-ngram-radical}$, $x_{i,e}^{character}$ and $x_{i,e}^{character-radical}$ are the prediction probability that the $i-th$ sample of the six gradualities belongs to class e respectively.

IV. EXPERIMENTS

A. DATASET

THUCNews: THUCNews[1] is a dataset generated by filtering the historical data of Sina News from 2005 to 2011, which contains 836,075 news texts in 14 categories. The titles of 10000 texts of each category are selected to build the Chinese short text dataset THUCNews-S.

CNT: CNT is provided by Zhou et al. [46], which contains 63938 Chinese news headlines in 32 categories. After removing the headlines with non-Chinese characters, there are 43,598 Chinese short texts left.

The two datasets are divided into training set, test set and validate set according to the ratio of 6:2:2. Table 1 shows the statistics of the two datasets.

B. EXPERIMENTAL ENVIRONMENT

1) PARAMETERS OF THE SIX DIMENSIONAL FEATURES

Xinhua Dictionary[2] is used to get the radical of Chinese word and character. Pre-trained Word2Vec is used to obtain the

[1] http://thuctc.thunlp.org/
[2] http://tool.httpcn.com/zi/

TABLE 1. Statistics of THUCNews-S and CNT datasets. (Ave is the average length, Max is the maximum length, and Min is the minimum length.)

Dataset	Type	Count	Ave	Max	Min	Class
THUCNews-S	Train	84,000	12.6	37	6	14
	Test	28,000	12.8	41	5	14
	Validate	28,000	12.5	35	7	14
CNT	Train	26,158	16.3	53	8	32
	Test	8,720	16.1	52	6	32
	Validate	8,720	16.4	57	7	32

embedding vectors. The dimension of the embedding vector is 200, and each sentence is converted into a vector with a length of 100.

2) PARAMETERS OF BIG-ATT-CNN LAYER

The number of neurons in the BIGRU layer is set to 256 and a dropout of 0.5 is used. 256 convolutional kernels of size 3 with a pooling template of size 3 and step size 3 are used for the first CNN layer, and 128 convolutional kernels of size 3 with a pooling template of size 3 and step size 3 are used for the second CNN layer, and a dropout of 0.5 is used after each convolutional layer. The learning rate is set to 0.0001, the batch size is set to 128, and the epoch is set to 200.

C. BASELINE METHODS

DCNN [47]: The model uses dynamic K-Max pooling and global pooling operation on linear sequence, so it can process input sentences with different lengths, and lead to a feature graph on the sentence, which can explicitly capture short-term and long-term relationships.

GRU [45]: The model is a variant of recurrent neural networks that can be used to solve the gradient disappearance/explosion problem during long-term memory and backpropagation.

ABCDM [48]: The model extracts temporal features using two parallel BiLSTMs and BiGRU, respectively, and outputs them to the Attention layer. Then, local features are extracted using the convolutional layer, and global and average pooling operations are used to reduce the sampling of its feature maps, respectively.

AC-BiLSTM [42]: The model firstly uses convolutional layers to extract local features, secondly uses BILSTM to obtain forward and backward contextual representations, and finally, the forward and backward results are fed into the Attention mechanism separately to obtain different Attention levels.

NeuralCRF [43]: The model is the first to propose a combination of learning Chinese character embedding using radical, by which and combined with a specific deep learning model, Chinese text can be accurately classified.

RAFG [44]: The model uses words, words-radical, character and character-radical to build a four-granularity based model, and classifies Chinese text by building a classification model based on BILSTM and Attention.

TABLE 2. Comparison of the results obtained on the THUCNews-S dataset.

Methods	Accuracy	Precision	Recall	F1
DCNN	82.53%	81.66%	81.92%	81.79%
GRU	78.53%	79.14%	76.68%	77.89%
ABCDM	84.75%	82.92%	81.87%	82.39%
AC-BiLSTM	85.52%	86.77%	87.03%	86.90%
NeuralCRF	86.69%	87.73%	86.55%	87.14%
RAFG	87.29%	88.36%	85.77%	87.05%
Xlnet	91.19%	92.23%	90.88%	91.55%
AlBert	91.73%	92.91%	92.08%	92.49%
SGCSTC	93.36%	94.47%	94.15%	94.31%

TABLE 3. Comparison of the results obtained on the CNT dataset.

Methods	Accuracy	Precision	Recall	F1
DCNN	78.58%	80.02%	77.96%	78.98%
GRU	75.66%	76.93%	77.01%	76.97%
ABCDM	80.14%	79.96%	83.45%	81.67%
AC-BiLSTM	85.06%	85.22%	84.12%	84.67%
NeuralCRF	85.59%	84.34%	85.11%	84.72%
RAFG	86.84%	85.99%	87.31%	86.64%
Xlnet	89.94%	88.79%	91.03%	89.9%
AlBert	90.31%	89.62%	90.79%	90.2%
SGCSTC	92.67%	92.38%	93.15%	92.76%

TABLE 4. Comparison of the results on THUCNews-S dataset after replacing the BIG-ATT-CNN in SGCSTC with other deep learning models.

Methods	Accuracy	Precision	Recall	F1
SGCSTC+DCNN	88.86%	87.71%	87.44%	87.57%
SGCSTC+GRU	86.93%	86.64%	85.31%	85.97%
SGCSTC+ABCDM	90.11%	91.35%	90.83%	91.09%
SGCSTC+AC-BiLSTM	92.69%	93.80%	92.04%	92.91%

Xlnet [49]: The model enables learning bidirectional contexts by maximizing the expected likelihood over all permutations of the factorization order based on Bert.

AlBert [50]: The model uses two parameter-reduction techniques to lower memory consumption and increase the training speed of BERT.

D. EVALUATION INDEX

Accuracy, Precision, Recall and F1 are chosen to evaluate SGCSTC and baseline methods, which are defined as follows:

$$Accuracy = \frac{TP + TN}{P + N}$$
$$Precision = \frac{TP}{TP + FP}$$
$$Recall = \frac{TP}{TP + FN}$$
$$F1 = \frac{2 * Precision * Recall}{Precision + Recall}$$

where TP is the number of correctly predicted forward samples, TN is the number of correctly predicted reverse samples, FP is the number of incorrectly predicted forward samples, and FN is the number of incorrectly predicted reverse samples. In order to make the model prediction more accurate, the final prediction results of the model are obtained using 5-fold cross validation, respectively.

TABLE 5. Comparison of the results on CNT dataset after replacing the BIG-ATT-CNN in SGCSTC with other deep learning models.

Methods	Accuracy	Precision	Recall	F1
SGCSTC+DCNN	87.59%	85.82%	86.40%	86.11%
SGCSTC+GRU	84.25%	83.99%	85.06%	84.52%
SGCSTC+ABCDM	88.32%	89.68%	89.17%	89.42%
SGCSTC+AC-BiLSTM	91.75%	92.09%	92.23%	92.16%

E. EXPERIMENTAL RESULTS

SGCSTC and baseline methods are used to perform the classification tasks on THUCNews-S and CNT datasets, and the classification results are shown in Tables 2 and 3, respectively. It can be seen from Tables 2 and 3 that the model achieves the optimal classification results on both datasets. Specifically, the Accuracy, Precision, Recall and F1 of SGCSTC on THUCNews-S dataset are 93.36%, 94.47%, 94.15% and 94.31% respectively, which are at least 1.63%, 1.56%, 2.07% and 1.82% higher than those of the baseline methods. The Accuracy, Precision, Recall and F1 of SGCSTC on CNT dataset are 92.67%, 92.38%, 93.15% and 92.76% respectively, which are at least 2.36%, 2.76%, 2.12% and 2.56% higher than those of the baseline methods.

F. MODEL ANALYSIS

1) ANALYSIS OF THE DEEP LEARNING MODEL

It can be seen from the above experimental results that the classification results of the baseline methods using multiple deep learning models (i.e. ABCDM and AC-BiLSTM) are higher than those using one deep learning model (i.e. NeuralCRF, RAFG and SGCSTC). BIG-ATT-CNN is constructed in this paper to solve the data sparsity problem of Chinese short text by integrating the forward, backward and bi-directional temporal features of the embedding vector. The following experiment are performed to verify the classification effect of BIG-ATT-CNN. The other layers of SGCSTC are kept unchanged, and the BIG-ATT-CNN layer are replaced with DCNN, GRU, ABCDM and AC-BiLSTM respectively, and the classification results of the replaced models on THUCNews-S and CNT are shown in Tables 4 and 5, respectively.

From Tables 4 and 5, it can be seen that the text classification results are much better than those using DCNN, GRU, ABCDM and AC-BiLSTM alone when the BIG-ATT-CNN of SGCSTC is replaced by them, whcih shows that the proposed classification method combining word, character and radical can effectively improve the classification results of Chinese short text. In addition, combining the results in Tables 2-5, it can be seen that the text classification results of SGCSTC are still optimal in both datasets, which indicates that the BIG-ATT-CNN model used in SGCSTC can well solve the data sparsity problem of Chinese short texts.

2) ANALYSIS OF THE CLASSIFICATION EFFECT OF RADICAL

It can be seen from the above experimental results that the use of radical can effectively increase the classification results, and three granularities of radical are used in this paper, i.e. word-jieba-radical(R_1), word-ngram-radical(R_2) and character-radical(R_3). The following experiments are performed to analyze the effects of these three granularities of radical on the classification results of Chinese short text. The Chinese text classification is performed on THUCNews-S and CNT using only 0, 1 or 2 of R_1, R_2, R_3 respectively, while all other layers of SGCSTC are kept unchanged. The results are shown in Tables 6 and 7, respectively.

TABLE 6. Comparison of the results of combinations of different granularity of radicals in SGCSTC on the THUCNews-S dataset.

Methods	Accuracy	Precision	Recall	F1
SGCSTC(no radical)	88.67%	87.92%	88.31%	88.11%
SGCSTC(R_1)	91.03%	91.54%	90.89%	91.21%
SGCSTC(R_2)	90.18%	90.28%	90.74%	90.51%
SGCSTC(R_3)	91.20%	90.96%	91.08%	91.02%
SGCSTC(R_1+R_2)	92.38%	92.85%	93.03%	92.94%
SGCSTC(R_1+R_3)	93.05%	93.77%	93.16%	93.46%
SGCSTC(R_2+R_3)	92.84%	92.67%	92.93%	92.80%

TABLE 7. Comparison of the results of combinations of different granularity of radicals in SGCSTC on the CNT dataset.

Methods	Accuracy	Precision	Recall	F1
SGCSTC(no radical)	85.86%	86.69%	86.74%	86.71%
SGCSTC(R_1)	89.36%	90.05%	89.52%	89.78%
SGCSTC(R_2)	88.78%	87.90%	88.44%	88.17%
SGCSTC(R_3)	90.11%	89.57%	89.18%	89.37%
SGCSTC(R_1+R_2)	91.66%	91.31%	91.08%	91.19%
SGCSTC(R_1+R_3)	92.04%	91.98%	92.36%	92.17%
SGCSTC(R_2+R_3)	91.75%	91.19%	91.36%	91.27%

It can be seen from Tables 6 and 7 that the classification results are the worst when there is no radical included. When only one granularity of radical is included, the classification results are appropriately improved, and the classification results with R_1 and R_3 are better than that with R_2. When two granularities of radical are included, the classification results are better than that of only one granularity of radical, but they are still inferior to that of three granularities of radical (see Tables 2 and 3). This indicates that the idea of simultaneously based on three granularities of radical proposed in this paper can further solve the data sparsity problem and improve the classification results of Chinese short texts.

3) CLASSIFICATION RESULTS WITH DIFFERENT LENGTHS

The shorter the text length is, the more obvious the data sparsity problem is, and the worse the classification results are. On the contrary, the better the classification results are. The following experiments are performed to verify the classification results of SGCSTC for texts of different lengths. Texts with lengths of 10-30 in THUCNews-S and texts with lengths of 10-50 in CNT are selected according to the length statistics in Table 1, and the classification results of SGCSTC at different lengths are shown in Figures 4 and 5, respectively. From Figures 4 and 5, it can be seen that the Accuracy, Precision, Recall and F1 of SGCSTC on both datasets gradually increase with the increase of the length, and tend to stabilize and slightly decrease with the further increase of the length.

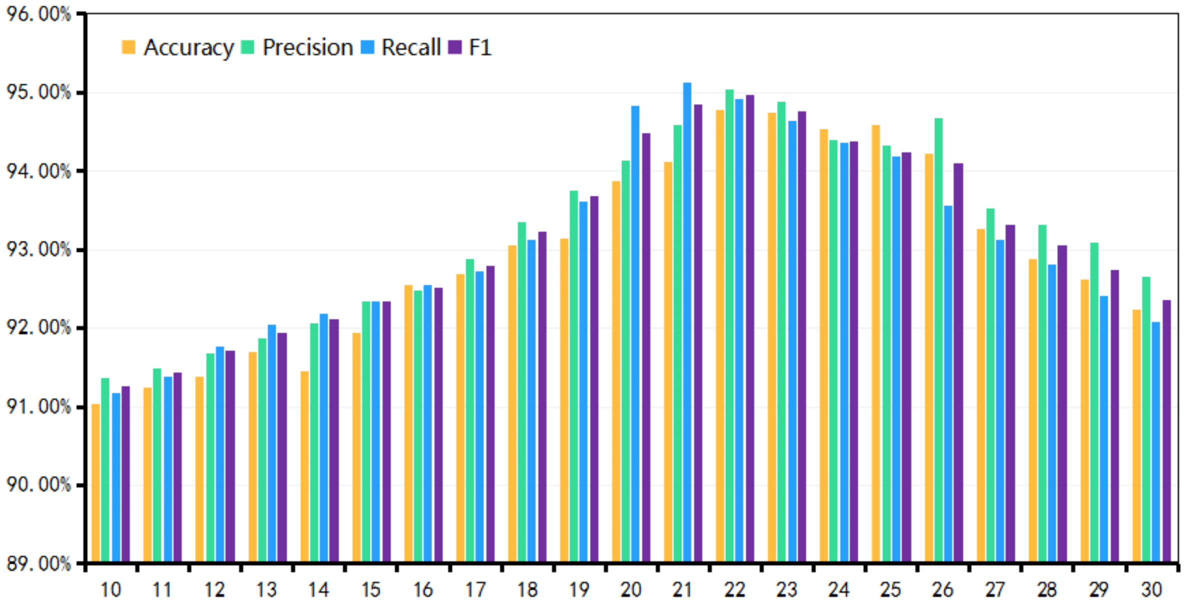

FIGURE 4. Results of evaluation indexes under THUCNews-S dataset with different lengths.

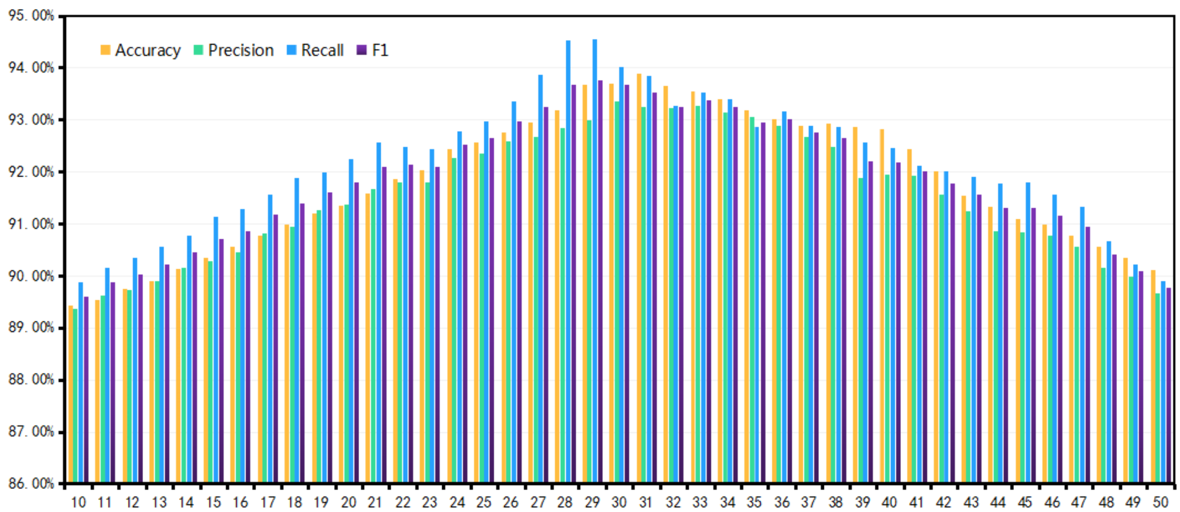

FIGURE 5. Results of evaluation indexes under CNT dataset with different lengths.

We analyze the reason why the classification results decrease with the further increase of the length, mainly because the number of corresponding text datasets decreases when the length is large, and the classification results decrease due to the lack of sufficient training datasets.

4) CLASSIFICATION RESULTS OF DIFFERENT CATEGORIES

SGCSTC achieves good classification results on the THUCNews-S and CNT datasets. Since both datasets are multi-label classified datasets, the following experiments are performed to verify the classification results of SGCSTC for different categories of Chinese short text. SGCSTC and the baseline methods are selected to analyze the classification accuracy on 14 categories in THUCNews-S and 32 categories in CNT. The experimental results are shown in Figures 6 and 7, respectively. From Figures 6 and 7, it can be seen that SGCSTC achieves optimal classification accuracy for most of the categories except for very few categories (design, star and news) in CNT. We analyze why SGCSTC does not achieve the optimal classification accuracy in some categories in CNT, mainly because the content in these categories is relatively new and there are many new web terms, therefore, the radical-based approach cannot help improve the text classification accuracy. From the results in Figures 6 and 7, it can be seen that the classification accuracy of SGCSTC on each category is much better than that of the method without

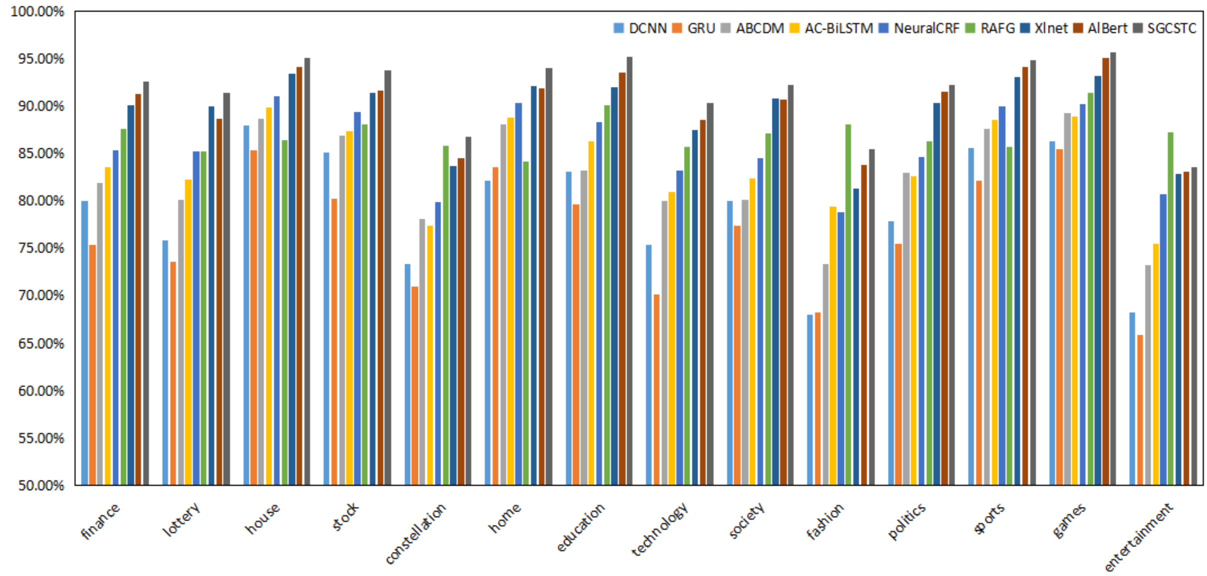

FIGURE 6. Comparison of the classification accuracy of the model in each category of THUCNews-S dataset.

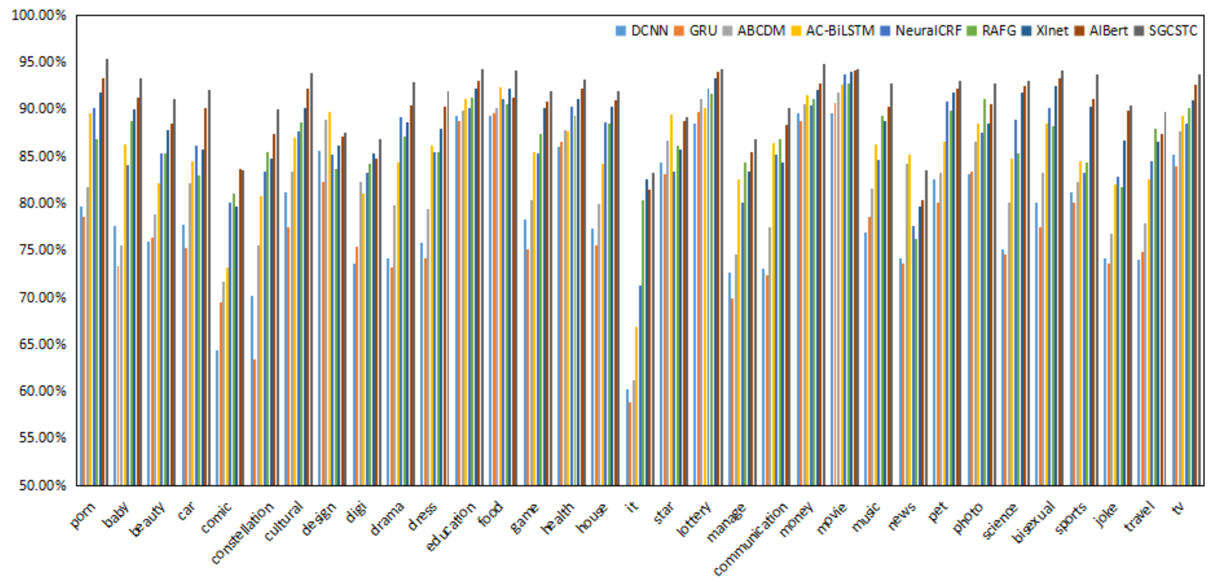

FIGURE 7. Comparison of the classification accuracy of the model in each category of CNT dataset.

using radical, which further illustrates the superiority of the model based on six granularities proposed in this paper.

V. CONCLUSION

SGCSTC is constructed in this paper for the Chinese short text classification. In order to avoid the impact of losing important information or generating noisy data caused by separate Chinese word segmentation methods, both jieba and ngram are used to segment Chinese short texts into words. Considering that Chinese radical can also express rich semantic information, a Chinese short text classification model based on six gradualities is constructed, which can well solve the data sparsity problem of Chinese short text. In order to reflect the different degree of influence of different granularity features on the classification results of Chinese short texts, different weights are assigned to the six granularities, which are updated automatically in the process of back-propagation using cross-entropy loss. In addition, a classification model based on BIGRU, Attention and CNN is constructed, and the forward, backward and bidirectional hidden vectors of BIGRU are used for short text classification at the same time, which can further solve the sparsity problem of Chinese short text. The classification Accuracy, Precision, Recall and F1 of SGCSTC in THUCNews-S dataset are 93.36%, 94.47%, 94.15% and 94.31% respectively, and that in CNT dataset are 92.67%, 92.38%, 93.15% and 92.76% respectively, and multiple comparative experiment results on THUCNews-S and CNT datasets show that SGCSTC outperforms the state-of-the-art text classification models.

However, the embedding vectors in this paper are obtained based on the pre-trained word2vec, so the robustness of the model is poor. When the text contains some of the latest network terms, the model classification results are reduced. In addition, SGCSTC cannot process Chinese text with special symbols, but many Chinese short text, especially comment text, contain many special symbols. SGCSTC's classification results for this kind of Chinese short text are not ideal.

Therefore, our future work will both focus on building a Chinese embedding dictionary containing enough of the latest network terms to improve the robustness of the model and building a classification model that can handle Chinese short text with special symbols well.

REFERENCES

[1] M. Birjali, M. Kasri, and A. Beni-Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," *Knowl.-Based Syst.*, vol. 226, Aug. 2021, Art. no. 107134.

[2] P. Garg and N. Girdhar, "A systematic review on spam filtering techniques based on natural language processing framework," in *Proc. 11th Int. Conf. Cloud Comput., Data Sci. Eng.*, Jan. 2021, pp. 30–35.

[3] Z. Wu, S. Shen, H. Zhou, H. Li, C. Lu, and D. Zou, "An effective approach for the protection of user commodity viewing privacy in e-commerce website," *Knowl.-Based Syst.*, vol. 220, May 2021, Art. no. 106952.

[4] L. Galke and A. Scherp, "Bag-of-words vs. graph vs. sequence in text classification: Questioning the necessity of text-graphs and the surprising strength of a wide MLP," in *Proc. 60th Annu. Meeting Assoc. Comput. Linguistics*, 2022, pp. 4038–4051.

[5] T. Hasan and A. Matin, "Extract sentiment from customer reviews: A better approach of TF-IDF and BOW-based text classification using N-Gram technique," in *Proc. Int. Joint Conf. Adv. Comput. Intell.*, 2021, pp. 231–244.

[6] H. Toosi, M. A. Ghaaderi, and Z. Shokrani, "Comparative study of academic research on project management in Iran and the world with text mining approach and TF–IDF method," *Eng., Construct. Architectural Manag.*, vol. 29, pp. 1553–1583, Apr. 2021.

[7] J. S. Manoharan, "Capsule network algorithm for performance optimization of text classification," *J. Soft Comput. Paradigm*, vol. 3, no. 1, pp. 1–9, Apr. 2021.

[8] S. Ruan, B. Chen, K. Song, and H. Li, "Weighted naive Bayes text classification algorithm based on improved distance correlation coefficient," *Neural Comput. Appl.*, vol. 34, pp. 2729–2738, Feb. 2022.

[9] S. Ahmed, K. H. Sheikh, S. Mirjalili, and R. Sarkar, "Binary simulated normal distribution optimizer for feature selection: Theory and application in COVID-19 datasets," *Exp. Syst. Appl.*, vol. 200, Aug. 2022, Art. no. 116834.

[10] V. Nurcahyawati and Z. Mustaffa, "Improving sentiment reviews classification performance using support vector machine-fuzzy matching algorithm," *Bull. Electr. Eng. Informat.*, vol. 12, no. 3, pp. 1817–1824, Jun. 2023.

[11] B. Parlak and A. K. Uysal, "A novel filter feature selection method for text classification: Extensive feature selector," *J. Inf. Sci.*, vol. 49, no. 1, pp. 59–78, Feb. 2023.

[12] W. Zhao, L. Wang, and S. Mirjalili, "Artificial hummingbird algorithm: A new bio-inspired optimizer with its engineering applications," *Comput. Methods Appl. Mech. Eng.*, vol. 388, Jan. 2022, Art. no. 114194.

[13] M. M. Mironczuk and J. Protasiewicz, "A recent overview of the state-of-the-art elements of text classification," *Exp. Syst. Appl.*, vol. 106, pp. 36–54, Sep. 2018.

[14] Y. Zhou, J. Li, J. Chi, W. Tang, and Y. Zheng, "Set-CNN: A text convolutional neural network based on semantic extension for short text classification," *Knowl.-Based Syst.*, vol. 257, Dec. 2022, Art. no. 109948.

[15] S. Bodapati, H. Bandarupally, R. N. Shaw, and A. Ghosh, "Comparison and analysis of RNN-LSTMs and CNNs for social reviews classification," in *Advances in Applications of Data-Driven Computing*. 2021, pp. 49–59.

[16] T. Yang, L. Hu, C. Shi, H. Ji, X. Li, and L. Nie, "HGAT: Heterogeneous graph attention networks for semi-supervised short text classification," *ACM Trans. Inf. Syst.*, vol. 39, no. 3, pp. 1–29, Jul. 2021.

[17] H. Zhao, C. Zhang, and A. Wang, "An enhanced text classification model by the inverted attention orthogonal projection module," *Connection Sci.*, vol. 35, no. 1, pp. 1–18, Dec. 2023.

[18] D. Magalhães, R. H. R. Lima, and A. Pozo, "Creating deep neural networks for text classification tasks using grammar genetic programming," *Appl. Soft Comput.*, vol. 135, Mar. 2023, Art. no. 110009.

[19] Y. Huang, X. Dai, J. Yu, and Z. Huang, "SA-SGRU: Combining improved self-attention and skip-GRU for text classification," *Appl. Sci.*, vol. 13, no. 3, p. 1296, Jan. 2023.

[20] B. Skrlj, M. Martinc, J. Kralj, N. Lavrac, and S. Pollak, "tax2vec: Constructing interpretable features from taxonomies for short text classification," *Comput. Speech Lang.*, vol. 65, Jan. 2021, Art. no. 101104.

[21] S. A. Sulaimani and A. Starkey, "Short text classification using contextual analysis," *IEEE Access*, vol. 9, pp. 149619–149629, 2021.

[22] J. Deng, L. Cheng, and Z. Wang, "Attention-based BiLSTM fused CNN with gating mechanism model for Chinese long text classification," *Comput. Speech Lang.*, vol. 68, Jul. 2021, Art. no. 101182.

[23] Z. Zhai, X. Zhang, F. Fang, and L. Yao, "Text classification of Chinese news based on multi-scale CNN and LSTM hybrid model," *Multimedia Tools Appl.*, pp. 1–14, Feb. 2023.

[24] M. Hao, B. Xu, J. Y. Liang, B. W. Zhang, and X. C. Yin, "Chinese short text classification with mutual-attention convolutional neural networks," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 19, no. 5, pp. 1–13, 2020.

[25] Y. Yuan, X. Mu, X. Shao, J. Ren, Y. Zhao, and Z. Wang, "Optimization of an auto drum fashioned brake using the elite opposition-based learning and chaotic K-best gravitational search strategy based grey wolf optimizer algorithm," *Appl. Soft Comput.*, vol. 123, Jul. 2022, Art. no. 108947.

[26] Y. Yuan, J. Ren, S. Wang, Z. Wang, X. Mu, and W. Zhao, "Alpine skiing optimization: A new bio-inspired optimization algorithm," *Adv. Eng. Softw.*, vol. 170, Aug. 2022, Art. no. 103158.

[27] X. Luo, "Efficient English text classification using selected machine learning techniques," *Alexandria Eng. J.*, vol. 60, no. 3, pp. 3401–3409, Jun. 2021.

[28] Z. L. Chia, M. Ptaszynski, F. Masui, G. Leliwa, and M. Wroczynski, "Machine learning and feature engineering-based study into sarcasm and irony classification with application to cyberbullying detection," *Inf. Process. Manag.*, vol. 58, Jul. 2021, Art. no. 102600.

[29] K. Thirumoorthy and K. Muneeswaran, "Feature selection for text classification using machine learning approaches," *Nat. Acad. Sci. Lett.*, vol. 45, no. 1, pp. 51–56, Feb. 2022.

[30] N. Aljedani, R. Alotaibi, and M. Taileb, "HMATC: Hierarchical multi-label Arabic text classification model using machine learning," *Egyptian Informat. J.*, vol. 22, no. 3, pp. 225–237, Sep. 2021.

[31] R. Janani and S. Vijayarani, "Automatic text classification using machine learning and optimization algorithms," *Soft Comput.*, vol. 25, no. 2, pp. 1129–1145, Jan. 2021.

[32] A. Mohammed and R. Kora, "An effective ensemble deep learning framework for text classification," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, no. 10, pp. 8825–8837, Nov. 2022.

[33] M. P. Akhter, Z. Jiangbin, I. R. Naqvi, M. Abdelmajeed, and M. Fayyaz, "Exploring deep learning approaches for Urdu text classification in product manufacturing," *Enterprise Inf. Syst.*, vol. 16, no. 2, pp. 223–248, Feb. 2022.

[34] M. A. Ibrahim, M. U. G. Khan, F. Mehmood, M. N. Asim, and W. Mahmood, "GHS-NET a generic hybridized shallow neural network for multi-label biomedical text classification," *J. Biomed. Informat.*, vol. 116, Apr. 2021, Art. no. 103699.

[35] Y. J. Liang, H. Li, B. Guo, Z. Yu, X. Zheng, S. Samtani, and D. D. Zeng, "Fusion of heterogeneous attention mechanisms in multi-view convolutional neural network for text classification," *Inf. Sci.*, vol. 548, pp. 295–312, Feb. 2021.

[36] M. A. Al-Garadi, Y. C. Yang, H. Cai, Y. Ruan, K. O'Connor, G. H. Graciela, J. Perrone, and A. Sarker, "Text classification models for the automatic detection of nonmedical prescription medication use from social media," *BMC Med. Inform. Decis. Making*, vol. 21, no. 1, pp. 1–13, 2021.

[37] Z. Cai, T. Zhang, C. Wang, and X. He, "EMBERT: A pre-trained language model for Chinese medical text mining," in *Proc. Asia–Pacific Web (APWeb) Web-Age Inf. Manag. (WAIM) Joint Int. Conf. Web Big Data*, 2021, pp. 242–257.

[38] J. Liang, Z. Zeng, S. Yu, and Z. Hong, "An improved Chinese text classification ensemble model," in *Proc. 5th Int. Conf. Digit. Signal Process.*, Feb. 2021, pp. 332–336.

[39] H. Wang, K. Tian, Z. Wu, and L. Wang, "A short text classification method based on convolutional neural network and semantic extension," *Int. J. Comput. Intell. Syst.*, vol. 14, no. 1, pp. 367–375, 2021.

[40] S. Zhang, H. Yu, and G. Zhu, "An emotional classification method of Chinese short comment text based on ELECTRA," *Connection Sci.*, vol. 34, no. 1, pp. 254–273, Dec. 2022.

[41] Y. Feng and Y. Cheng, "Short text sentiment analysis based on multi-channel CNN with multi-head attention mechanism," *IEEE Access*, vol. 9, pp. 19854–19863, 2021.

[42] G. Liu and J. Guo, "Bidirectional LSTM with attention mechanism and convolutional layer for text classification," *Neurocomputing*, vol. 337, pp. 325–338, Jan. 2019.

[43] Y. Sun, L. Lin, N. Yang, Z. Ji, and X. Wang, "Radical-enhanced Chinese character embedding," in *Proc. Int. Conf. Neural Inf. Process.*, 2014, pp. 279–286.

[44] H. Tao, S. Tong, H. Zhao, T. Xu, B. Jin, and Q. Liu, "A radical-aware attention-based model for Chinese text classification," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 5125–5132.

[45] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014, *arXiv:1412.3555*.

[46] Y. Zhou, B. Xu, J. Xu, L. Yang, C. Li, and B. Xu, "Compositional recurrent neural networks for Chinese short text classification," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell. (WI)*, Oct. 2016, pp. 137–144.

[47] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," 2014, *arXiv:1404.2188*.

[48] M. E. Basiri, S. Nemati, M. Abdar, E. Cambria, and U. R. Acharya, "ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis," *Future Gener. Comput. Syst.*, vol. 115, pp. 279–294, Feb. 2021.

[49] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, and Q. V. Le, "XLNet: Generalized autoregressive pretraining for language understanding," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–11.

[50] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "ALBERT: A lite BERT for self-supervised learning of language representations," 2019, *arXiv:1909.11942*.

XINJIE SUN received the M.E. degree from the School of Computer Science and Engineering, Northeastern University, Shenyang, China, in 2013. He is currently a Teacher with Liupanshui Normal University. His research interests include next generation internet and artificial intelligence.

ZHIFANG LIU received the M.E. degree from the North China University of Science and Technology, Tangshan, China, in 1999. He is currently a Teacher with Liupanshui Normal University. His research interests include next generation internet and deep learning.

XINGYING HUO received the Ph.D. degree in electrical science and technology from Beijing Jiaotong University, Beijing, China, in 2018. From 2016 to 2017, she was a Visiting Student with Houston University, TX, USA. In 2018, she joined the Department of Computer Science, Faculty of Engineering, Liupanshui Normal University, where she became an Associate Professor, in 2019. Her current research interests include leaky-wave structures and radio-wave propagation.