A Hybrid Classification Method via Character Embedding in Chinese Short Text With Few Words

YI ZHU\textsuperscript{1,2,3}, YUN LI\textsuperscript{1}, YONGZHENG YUE\textsuperscript{2,3}, JIPENG QIANG\textsuperscript{1}, AND YUNHAO YUAN\textsuperscript{1}

\textsuperscript{1}School of Information Engineering, Yangzhou University, Jiangsu 225009, China
\textsuperscript{2}Key Laboratory of Knowledge Engineering with Big Data (Hefei University of Technology), Ministry of Education, Hefei 230009, China
\textsuperscript{3}School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230009, China

Corresponding author: Yun Li (liyun@yzu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61906060 and Grant 61703362, in part by the National Key Research and Development Program of China under Grant 2016YFB1000901, and in part by the scientific and technological research in Anhui Province under Grant 1704a902029.

ABSTRACT

Last decades have witnessed the significance development of research in short text classification. However, most existing methods only focus on the text which contained dozens of words like Twitter or MicroBlog, but not take the short text with few words like news headline or invoice name into consideration. Meanwhile, contemporary short text classification methods either to expand feature of short text with external corpus or to learn the feature representation from all the texts, which have not take the difference between words of short text into full consideration. Notably, the classification of short text with few words are usually determined by a few specific key words contrary to documents classification or traditional short text classification. To address these problems, this paper propose a hybrid classification method of Attention mechanism and Feature selection via Character embedding in Chinese short text with few words, called AFC. More specifically, firstly, the character embedding is computed to represent Chinese short texts with few words, which takes full advantage of short text information without external corpus. Secondly, attention-based LSTM is introduced in our method to project the data into feature representation space with weighting, which make the keywords in classification have more subtle value. Furthermore, the semantic similarity between content and class label information is calculated for feature selection, which reduces the possible negative influence of some redundant information on classification. Experiments on real-world datasets demonstrate the effectiveness of our method compared to other competing methods.

INDEX TERMS

Short text with few words, character embedding, attention mechanism, feature selection.

I. INTRODUCTION

With the rapid development of web services, more and more Chinese short texts with few words are generated on the Internet, such as microblog, news title and search snippets. The main difference between Chinese short texts with few words and traditional Chinese short text is the extremely short length. The Chinese short texts with few words are considered to just contains few Chinese characters, for instance the length of one news title is usually required to less than 20 characters. Recently, Chinese short texts with few words, e.g. news title and invoice name are posted at unprecedented rates, and they are usually an overview of all the content. In addition, the extreme short length, feature sparsity, high ambiguity and other inherent properties of these texts pose huge challenges to text classification. The urgent demand to process Chinese short texts with few words has attracted vast amount of attention and research [1]–[4].

However, most existing short text classification methods rarely notice the Chinese short text with few words. Considering the uniqueness of these texts, it is hard to apply existing classification methods to these texts directly due to the following problems. First, the extreme short text length incurs the lack of information for meaningful analysis and effective classification. Second, there is no explicit word delimitation in Chinese texts which is different from the
blank space between two words in English, and Chinese word segmentation may deteriorate the performance of classification. In addition, due to the existence of interfering words in Chinese short text with few words, it is more difficult to identify keywords for correct classification. Thus, it is a challenge in the tackling of very short Chinese text classification due to the efficiency and effectiveness.

To handle the short text classification, existing methods can roughly be divided into two categories: auxiliary information based methods and representation learning based methods. Auxiliary information based methods utilize the external knowledge base like Freebase, Probase and DBpedia to expand the features of short text. For example, Wang et al. [5] mapped short document to Wikipedia concepts for short text classification. Li et al. [6] used Probase to expand the feature space, and more semantic contexts of terms are introduced to make up of the data sparsity and disambiguate terms. Representation learning based methods seek to learn better feature representations of short text for improving the performance of classification, and many researches and methods have been proposed along this line, such as Naive Bayes(NB) [7] and Attention Mechanism [8]. For example, Zhou et al. [9] proposed the Compositional Recurrent Neural Network (RNN) for Chinese short text classification, which is a hybrid model of character-level and word-level future representations based on long short term memory (LSTM). Ma et al. [10] proposed to learn distributional representations with Gaussian process approach by assuming that short text is a specific sample of one distribution in a Bayesian framework, and the short texts classification is converted to select the most probable Gaussian distribution. Meanwhile, Yu et al. [11] proposed an open source library toolkit for short text analysis and classification which supports effective text pre-processing and fast training/prediction procedures based on representation learning, the toolkit is widely applied as an easy-to-use and extensible tool.

Among the above approaches, auxiliary information based methods can improve the classification performance with the help of external knowledge base [12], but most of the methods in this category are restricted in conditions and require plenty of time, they suffer from the problem of information unavailable and high communication overhead. It is hence hard to be applied in short text classification efficiently. On the other hand, while the representation learning based short text classification methods have achieved tremendous success, two problems prevent the further development of these methods in Chinese short text with few words. The first problem is the identification of key words. The Chinese text with few words have fewer keywords even just one because of the extreme short length, and the category of text usually determined by this key word. The second problem is the feature representation of Chinese short text with few words, the Chinese nonstandard representation, Chinese word segmentation and existence of distractors may result in the ambiguous meaning and misclassify.

For the first problem, the attention mechanism and feature selection are adopted to identify key words. The attention-based LSTM is utilized to weight each word for focusing on the word close to the sentence meaning. The query vectors and key-value vectors are mapped to an output vector for assigning the weight to each word, the vital information is assigned to a larger weight and the features which may not helpful for classification is assigned to a small weight. In addition, the feature selection is leveraged to reserve the notional words like nouns, verbs and adjectives that hold much information than other functional words, which reduces the possible negative influence of some redundant information. In this way, the key words in short texts with few words can be recognized for classification.

For the second problem, Chinese character embedding and feature selection are conducted for representation learning. Character embedding avoid the errors that generated by word segmentation and make full use of Chinese character information which is pretty meaningful. Meanwhile, the semantic similarity between texts and class label information is calculated for feature selection, then the results are sorted and the last few characters are considered to interfering words and removed. As the feature selection is the process to extract feature subsets from the original datasets for reducing the algorithm time complexity, it can effectively identify keywords and improve classification accuracy [13].

In summary, we integrate the attention mechanism and the feature selection via character embedding for the classification of Chinese short text with few words (short for AFC). The Chinese character embedding vectors are calculated as text representation firstly, then the attention mechanism is applied to assign each word a different weight, for which one is useful for classification would be larger, vice is small. Meanwhile, the semantic similarity between content and class label information are computed, then the all words are sorted and last words in list are deleted for feature selection. After that, the weighted vectors are aligned based on feature selection. The main contributions of this paper are summarized as follows:

- The Chinese character embedding is utilized for representation learning, which overcomes the deficit of word segmentation and takes full advantage of significant information of character.
- The attention mechanism is utilized to weight the characters of text to make a further weakening of noise, which enhance the key words influence in classification.
- The feature selection is leveraged to reduce the impact of irrelevant information and align sentence vectors for text classification.
- The comprehensive experiments over three real data sets show that our method outperforms state-of-the-art models and evaluate the effectiveness of our method.

The remainder of this paper is organized as follows. Related work is introduced in Section II and details of the proposed AFC method are provided in Section III. Experimental results and analysis on three real world datasets are
presented in Section IV, followed by our main conclusions in Section V.

II. RELATED WORK

Short text classification has played a crucial role in many applications of Natural Language Processing (NLP) for providing proper service of document management. It aims to process the text which length is very short, typically no longer than 100 characters, such as blog content, online reviews, news title and so on. Due to the sparseness, noises and non-standardability of short text, traditional text classification methods usually fail to achieve satisfied performance. In recent years, short text classification has attained much importance, traditional text classification methods with feature representation optimization, to take representation learning based methods, feature selection based methods and attention based methods.

A. REPRESENTATION LEARNING BASED METHODS

The feature representation learning of text is the fundamental problem of short text classification, which is the main intuition behind the deep learning methods in essence [14]. Mikolov et al. proposed Word2vec language model [15], which is a distributed representation of vocabulary based on neural networks. Peters et al. proposed Embeddings from Language Models(ELMo) [16] that is a new type of deep contextualized word representation, this model addresses both the challenges of different characteristics for word use and different words meaning across linguistic contexts. The vectors are derived from a bidirectional LSTM and trained with a coupled language model objective on a large scale corpus. Devlin et al. proposed a language representation model called Bidirectional Encoder Representation from Transformers(BERT) [17], which utilized the masked language model and next sentence prediction to pre-train a deep bidirectional transformer and text-pair representations. Although afore mentioned model can obtain better representation vectors of words, it is difficult to achieve satisfied performance on small data scale and be applied in Chinese short text with few words directly.

Different from English text, Chinese text is made up of character sequences rather than word sequences, and word is not a natural concept without argument. Therefore, recently, there have already been some effort on devoting character embedding to Chinese NLP task. For example, Zhao et al. [18] initially investigate the possibility of exploiting character dependencies for Chinese, and character-level dependency is shown as a good alternative to word boundary representation. Zhang et al. [19] developed a character-based parsing model that can produce character-level constituent trees for Chinese NLP, and demonstrated the importance and effectiveness of character-level information in Chinese parsing. Sun et al. [20] proposed a method to utilize radical for learning Chinese character embedding, and applied it on Chinese character similarity judgement and Chinese word segmentation. Li et al. [21] presented a character-level neural dependency parser together with character-level dependency treebank for Chinese, it showed that Chinese character embedding played an important role for NLP task performance. However, most of previous character embedding methods combined word-level and character-level features together for NLP task, it is a challenge in the classification of Chinese short text with few words by character embedding only.

B. ATTENTION BASED METHODS

While attention mechanism based methods have achieved tremendous success in many fields like natural language processing [22], there have been more and more researchers strive to develop attention based methods to short text classification [23], [24]. Li et al. proposed [25] a two-level attention networks to identify the sentiment of short text, both local and long-distance dependent features are captured simultaneously by the attention mechanism, and then the attention-based features are utilized to captures more relevant features. Hu et al. proposed [26] a heterogeneous graph neural network for semi-supervised short text classification. In this method, a dual-level attention mechanism which include node-level and type-level attention is utilized to learn the importance of neighbour node and different types to a current node. Meanwhile, the self-attention mechanism has been successfully applied in many NLP tasks, Wang et al. [27] proposed a method which densely connect convolutional neural network(CNN) with multi-scale feature attention, which can produce variable n-gram features to adaptively select multi-scale features for short text classification.

In recent years, due to the special properties of Chinese text, many efforts have been made to solve the problem of Chinese short text classification. For example, Zhou et al. proposed [28] a hybrid attention networks for Chinese short text classification, the word-level and character-level features are utilized for capturing class-related attentive representation to improve classification performance. Lu et al. proposed [29] a multi-representation mixed model with attention and ensemble learning for Chinese news headline classification. However, few of methods focus on the classification problem of Chinese short text with few words like news headline which average length is usually less than 20 characters. The extreme short length, no explicit word delimitation and the identification of keywords are still the huge challenges for classification.

C. FEATURE SELECTION BASED METHODS

The objective of feature selection is to prepare more clean and understandable data for building more simple and comprehensible model to improve the performance of data mining and machine learning [30]. Along this line, many works extend the basic feature selection with external knowledge or just optimize the feature representation in NLP tasks. In the methods with feature representation optimization, to take the external knowledge base like Probase into consideration, Liu et al. proposed [13] a feature selection method based
on part-of-speech and HowNet for Micro-blog Mining, the words with larger amount of information by different part-of-speech are selected based on HowNet, and HowNet is a common knowledge base which describes the concepts represented by Chinese and English and reveals the relationship between concepts and attributes of concepts. Méndez et al. proposed [31] a feature selection method based on semantic ontology cluster to build feature vectors, the words are grouped into topics for training different Machine Learning tasks.

In the short text classification, feature selection has been proven to be an effective and efficient method to process data. Meng et al. proposed [32] a feature selection method to solve the sparseness problem of short text classification, it considers the number of the short texts that shared some words with the same importance, and the intersections among different short texts’ feature vectors are increased. Tommasel et al. proposed [33] an online short text feature selection method for social media, which focuses on discovering implicit relations amongst new posts, already known ones and their corresponding authors to identify groups of socially related posts in the situation of short text stream. Liu et al. applied [34] four different feature selection algorithms to solve the multi-class sentiment classification problem, the experimental results show that feature selection is an effective method to improve classification accuracies and more features do not necessarily lead to better result.

III. PROPOSED METHOD

The whole framework of our proposed AFC is illustrated as FIGURE 1. Our method contains three components: Chinese character embedding, attention mechanism and feature selection. The motivation and details of our proposed method is described as follows.

A. MOTIVATION

The Chinese short text with few words refers to Chinese text with much more short length compared to the traditional normal text or even short text. For instance, the length of Chinese news title is usually required to less than 20 characters (e.g. Intel: no worries about talent backup), and the average length of Chinese invoice name is also about 20 characters. Meanwhile, these Chinese short texts with few words are usually an overview of all the content, such as the title is the summarization of the news and the invoice name is the generalization of the bills. However, in contrast to short texts classification, Chinese short texts with few words attracted far less attention in recent decades.

The extreme short length of Chinese texts with few words lead to the difficult of feature representation learning and key words identification. More specifically, firstly, different from English and other western languages, Chinese sentence is made up of character sequences rather than word sequences, and there is no delimiter between Chinese words. Due to Chinese character embedding has proven to be an effective representation method and the extreme short length of Chinese texts with few words, we propose to learn Chinese character embedding alone, which is the difference between our proposed method with the method in [9], [28]. Secondly, the extreme short length and the few Chinese words in our task means the importance of keywords identification, and the texts classification usually determined by even one keyword. The attention mechanism contained in our proposed method is to assign much more weights to the words that have closer semantic relations to the sentence meaning.

Finally, the feature selection method is applied to remove meaningless and disturbing features in Chinese texts with few words, which can improve the quality of feature vectors and help to identify keywords. Due to the reservation of notional words and semantic similarity calculation, the feature selection method reduces the possible negative influence of some interfering words and redundant information, which is the main difference between our proposed method with the method in [9], [28].
B. CHARACTER EMBEDDING REPRESENTATION
The Chinese texts are different from English for the meaning of each character in words, and there is no explicit word delimitation in Chinese sentences. The traditional method to learn feature representations is to create Chinese word segmentation, and the result of segmentation is treated as the input of learning model [35]. Although the errors caused by word segmentation may be detrimental to the performance, the word embedding has been proven to be effective in Chinese NLP tasks like Named Entity Recognition (NER) [36]. On the other hand, character embedding representations have been used in NLP tasks such as word relatedness computation [37], Part-Of-Speech (POS) tagging [38] and short text classification [28]. In addition, Yin et al. [39] demonstrated that characters contain rich information and can be effectively applied in word similarity computation and analogical reasoning, the method showed that character features are capable of indicating semantic meanings of words and significantly improve the performance. Therefore, these character embedding methods can achieve sound performance in Chinese NLP tasks [40], [41].

However, the feature information in Chinese short text with few words like news headline and invoice name is much less compared to document or even short text, and the effect of word embedding for indicating semantic meanings and reducing sparsity is limited. In addition, since the Chinese texts are written as continuous sequences of characters, the word segmentation may incur errors as afore mentioned. Therefore, we introduce the character embedding based on word2vec tool [15] for feature representation learning. The aim of the word2vec tool is to map each Chinese character in texts into a vector. Due to the extreme length of Chinese short texts with few words, we set the value of window parameter as 1 in word2vec tool, which refers that we just consider one character left and right.

Suppose the Chinese short text is denoted as \( D = \{ (d_i, y) \mid y \in Y \}_{i=1}^{l} \), where \( l \) is the number of characters in text and \( Y \) is the set of text labels. Each character can be represented as a vector, which denoted as \( d_i = [R^M] \), where \( R \) is the short text space and \( M \) denotes the attribute dimensionality. For an input Chinese short text with few words, the character embedding vectors are encoded firstly by the model of word2vec, and these vectors are then utilized as the input of the attention mechanism.

C. ATTENTION MECHANISM
On the basis of character embedding representation, we assign different weights to different embedding vectors for the observation that not all words contribute to classification equally. The aim of the attention mechanism is to focus on the words which have closer semantic relations to the meaning of short texts. In our proposed method, the attention-based LSTM is utilized for mapping vectors into feature space with weighting, which make the keywords in classification have more subtle value.

Suppose the output of LSTM is denoted as \( H = \{ h_i \}_{i=1}^{l} \). Similar to character embedding representation, each \( h_i \) can be represented as \( h_i = [R^N] \), where \( N \) is the attribute dimensionality. The new representation \( z_a \) can be computed as the sum of weighted vectors \( h_i \), which is shown as (1):

\[
z_a = \sum_{i=1}^{l} \alpha_i h_i
\]

where \( \alpha_i \) denotes the attention weight, and it can be computed as (2):

\[
\alpha_i = g(W_1 f(W_2 h_i + b))
\]

where \( f \) and \( g \) are the activation functions of network, \( W_1 \) and \( W_2 \) are the weight matrices and \( b \) denotes the bias vector of model.

D. SHORT TEXT FEATURE SELECTION
Feature selection is a process of removing meaningless and disturbing features from the original feature set to improve the performance, it greatly increases the classification accuracy of text by extracting the most representative feature subsets [33]. Due to that not all words have a positive effect on classification especially in the Chinese short texts with few words, the reservation of notional words and semantic similarity calculation are adopted to improve the quality of feature vectors and help to identify keywords. Different from the traditional feature selection methods which applied in text process, we propose to select features in Chinese short text with few words as the following steps:

Firstly, we have observation that the notional words like nouns, verbs and adjectives hold much information than other functional words, and the latter words only play the role for making the whole sentence fluent and complete [13]. Therefore, we propose to reserve only nouns, verbs and adjectives in each text, where part-of-speech tagging is implemented with Jieba Python Library.

Secondly, since the existence of interfering words and word ambiguities produce difficulties in the comprehension of languages and result in the unsatisfying performance, we apply word embedding based on word2vec to calculate the semantic similarity between each word in short text and all class label information, then the results are sorted and the last words are removed. More specifically, we will add a removed word when the length of sentence increased 5, and no words will be removed when the sentence length is less than 5. The process of feature selection can be shown as FIGURE 2:

For the representations of hybrid method, we combine the outputs of the attention mechanism with feature selection and get the sentence vectors \( z \). The output vectors can be computed as (3):

\[
z = \sum_{i=1}^{l} FS(\alpha_i h_i)
\]
where FS denotes the process of select characters based on the result of feature selection.

IV. EXPERIMENTS

In this section, we conduct extensive experiments to evaluate the effectiveness of our proposed method. Three real Chinese short texts datasets have been used in our experiments, where two of them are Chinese news titles datasets, namely THUCNews dataset and Toutiao dataset, and the rest one is Chinese invoice dataset.

In the following, we firstly introduce the details of datasets. Secondly the benchmark methods and experiment settings are introduced in detail. Then the classification results with observations of our proposed AFC and other competing methods are given. Finally, the properties of AFC are analyzed with certain dataset.

A. DATASETS

THUCNews Data Set

Thucnews is a dataset which generated by collecting historical data of Sina News from 2005 to 2011 [42]. In this paper, we only pick the news title of 200,000 articles, and 10 categories are selected to operate the experiment including finance, realty, stocks, education, science, society, politics, sports, game and entertainment. More specifically, 5000 news title are chosen from each category for training and 500 news for testing and validating respectively.

Toutiao Data Set

Toutiao data set is crawled from JinriToutiao, which is the well known news portal website in China. In the experiment, this dataset contains 32 categories including car, education, movie, news, science and so on. More specifically, there are 43,156 news title in training set, 15,986 in testing dataset and 4796 in validating dataset respectively.

Invoice Data Set

It is a real invoice dataset that comes from the support program of scientific and technological research in Anhui Province (No.1704a0902029), and we have uploaded this dataset to Github. There is a total of 4,200 categories, 6.5 million manually labeled data and 10 million unlabeled data, and the data format after pre-processing is shown in TABLE 1. In the experiment, 10 classes are chosen, and 4000 invoice name are selected from each class for training and 1000 for testing.

Finally, the statistics of all three real-world data sets are summarized in Table 2.

B. COMPARED METHODS

We compare our proposed AFC with the following baseline methods:

- Bigrams + LR/SVM, which are the baselines for text classification. In the experiment, we conduct these algorithms with bigrams as proposed in [43], Logistic Regression (LR) and SVM are adopted as classifiers respectively.
- LibShortText [11], which is an open source library tool for for large-scale short text classification. It has achieved sound performance in short text analysis.
- TextGrocery [44], which is a simple and efficient short-text classification tool based on LibLinear, the Jieba Tool is embedded as default tokenizer to support Chinese tokenize.
- Character enhanced word embedding model (CWE) [37], which takes characters as basic unit and learn embeddings according to the internal structures of words. It proposed a multiple prototype character embeddings and an effective word selection method.
- Compositional Recurrent Neural Networks (C-RNN) [9]. It is a hybrid model of character-level and word-level features based on RNN with LSTM or Bidirectional LSTM, which tends to construct the missing semantic information caused by word segmentation. In the experiment, we implemented this method with LSTM.
- Hybrid Attention Networks (HANs) [28]. It is a hybrid model which combines the word- and character-level selective attentions. The method applies RNN and CNN to extract the semantic features of texts, and it captures

![FIGURE 2. Short Text Feature selection model.](image-url)
class-related attentive representation from word- and character-level features.

C. EXPERIMENT SETTINGS
The hyper-parameters of the neural networks are described as follows. In the experiments, we choose the hyper-parameter settings that followed the previous studies [9], [28], [37]. In addition, the size of LSTM hidden layer is set as 100 and the vector size of the character embedding as 100. The method of stochastic gradient descent is utilized to train all models with learning rate which is 0.01 and momentum of 0.9. Moreover, we train word embedding and character embedding using word2vec tool with the window parameter as 1,4 and accuracy is adopted in the experiments as the metric.

D. EXPERIMENTAL RESULTS
Table 3 shows the experimental results on three data sets. We have the following observations from experimental results:

- Neural network based methods (e.g., CWE and C-RNN) deliver a relatively good result compared to traditional Bag-of-Word methods (e.g., Bigrams based baselines and LibShortText) in most cases, which reveals the ability of neural network to capture effective semantic representations of Chinese short texts.
- We compare our proposed AFC to the CWE, and these two methods are all based on Chinese character embedding. We find that AFC performs better than the CWE in all datasets. We believe that the feature selection and LSTM based attention mechanism used in AFC can assign more subtle value to the keywords in classification.
- We also compare the AFC to the C-RNN and HANs. The experimental results show that AFC outperforms in Toutiao and Invoice datasets, and obtains a result that is competitive in the THUCNews dataset. We believe that the reason is the feature selection method used in AFC can help identify keywords and remove useless data. This result also demonstrate the effectiveness of our proposed method.
- Overall, in all datasets expect THUCNews, our proposed AFC performs best compared to the state-of-the-art methods. In the THUCNews dataset, AFC has a competitive result with the best baseline method. The results prove the effectiveness of the proposed method.

E. ANALYSIS OF PROPERTIES IN AFC
To learn the better feature representations for short text classification, we introduce the Chinese character embedding into our proposed method. In experiments, we use word2vec approach to train character feature vectors. We apply our method on Invoice dataset with word embedding and character embedding under different data size, and the classification results are shown in Table 4. We can see that the proposed AFC based on character embedding outperforms the one with word embedding significantly, and achieves a relatively good result even with less data. It is conducive to learn powerful features on stream data in the future work.

In our method, feature selection is introduced to reduce the possible negative influence of some redundant information and sentence vector alignment. To analyze the influence of feature selection in our proposed method, we conduct additional experiments with and without feature selection. The vectors in AFC without feature selection is learnt by sentence embedding [45]. The results are reported as Table 5. We can observe that AFC with feature selection has a more desirable performance than sentence embedding, which proves the importance of feature selection in our designed method.

V. CONCLUSION
In this paper, we propose a hybrid method of feature selection and attention network for Chinese short text classification with few words, called AFC. It improves previous methods with a novel framework by incorporating Chinese character embedding, attention mechanism and feature selection to obtain better feature representations for short text classification. In our proposed method, character embedding representations are computed, LSTM-based attention network is
utilized weight each word for identify keywords and feature selection is adopted to reduce the possible negative influence of some redundant information. The hybrid model can obtain better feature representations from training data in Chinese short texts with few words. Extensive experiments conducted on three real-world datasets show the proposed method out-performs competing methods in the effectiveness.

At the same time, experimental results show that our proposed method can achieve better results on Chinese short texts with few words, but when the Chinese texts consist of several to dozens of words, the performance may be not satisfied. In future work, we will try to improve the classification accuracy on Chinese short text.

REFERENCES

[1] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas, “Short text classification in Twitter to improve information filtering,” in Proc. 33rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. (SIGIR), Geneva, Switzerland, Jul. 2010, pp. 841–842.

[2] P. Li, L. He, H. Wang, X. Hu, Y. Zhang, L. Li, and X. Wu, “Learning from short text streams with topic drifts,” IEEE Trans. Cybern., vol. 48, no. 9, pp. 2697–2711, Sep. 2018.

[3] X. Li, Y. Wang, A. Zhang, C. Li, J. Chi, and J. Ouyang, “Filtering out the noise in short text topic modeling,” Inf. Sci., vol. 456, pp. 83–96, Aug. 2018.

[4] 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., vol. 33, Jul. 2019, pp. 5125–5132.

[5] X. Wang, R. Chen, Y. Jia, and B. Zhou, “Short text classification using wikipedia concept based document representation,” in Proc. Int. Conf. Int. Technol. Appl., Nov. 2013, pp. 471–474.

[6] P. Li, L. He, X. Hu, Y. Zhang, L. Li, and X. Wu, “Concept based short text stream classification with topic drifting detection,” in Proc. IEEE 16th Int. Conf. Data Mining (ICDM), Dec. 2016, pp. 1009–1014.

[7] Q. Yuan, G. Cong, and N. M. Thalmann, “Enhancing naive bayes with various smoothing methods for short text classification,” in Proc. 21st Int. Conf. Companion World Wide Web (WWW Companion). New York, NY, USA: ACM, 2012, pp. 645–646.

[8] L. Yang, Q. Ai, J. Guo, and W. B. Croft, “ANMM: Ranking short answer texts with attention-based neural matching model,” in Proc. 25th ACM Int. Conf. Inf. Knowl. Manage. New York, NY, USA: ACM, Oct. 2016, pp. 287–296.

[9] Y. Zhou, G. Xu, X. Li, and B. Wu, “Short text feature selection technique for social media,” in Proc. Proc. 27th Int. Joint Conf. Artif. Intell. 2018, pp. 4468–4474.

[10] Y. Zhou, J. Xu, J. Cao, B. Xu, C. Li, and B. Xu, “Hybrid attention networks for Chinese short text classification,” Computación y Sistemas, vol. 21, no. 4, pp. 759–769, 2017.

[11] Z. Lu, W. Liu, Y. Zhou, X. Hu, and B. Wang, “An effective approach for Chinese news headline classification based on multi-representation mixed model with attention and ensemble learning,” in Proc. Nat. CCF Conf. Natural Lang. Process. Chin. Comput., 2017, pp. 339–350.

[12] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, and H. Liu, “Feature selection: A data perspective,” ACM Comput. Surv., vol. 50, no. 6, pp. 94–139, 2018.

[13] J. R. Méndez, T. R. Cotos-Yáñez, and D. Ruano-Ortós, “A new semantic-based feature selection method for spam filtering,” Appl. Soft Comput., vol. 76, pp. 89–104, Mar. 2019.

[14] M. Wang, L. Lin, and F. Wang, “Improving short text classification through better feature space selection,” in Proc. 9th Int. Conf. Comput. Intell. Secur., Dec. 2013, pp. 120–124.

[15] A. Tommasel and D. Godoy, “A social-aware online short-text feature selection technique for social media,” Inf. Fusion, vol. 40, pp. 1–18, Mar. 2018.

[16] Y. Liu, J.-W. Bi, and Z.-P. Fan, “Multi-class sentiment classification: The experimental comparisons of feature selection and machine learning algorithms,” Expert Syst. Appl., vol. 80, pp. 323–339, Sep. 2017.

[17] Q. Zhang, J. Kang, J. Qian, and X. Huang, “Continuous word embeddings for detecting local text reuses at the semantic level,” in Proc. 37th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. (SIGIR). New York, NY, USA: ACM, 2014, pp. 797–806.

[18] N. Peng and M. Dredze, “Named entity recognition for chinese social media with jointly trained embeddings,” in Proc. Conf. Empirical Methods Natural Lang. Process., 2015, pp. 548–554.

[19] X. Chen, L. Xu, Z. Liu, M. Sun, and H. Luan, “Joint learning of character and word embeddings,” in Proc. 24th Int. Conf. Joint Artif. Intell., 2015, pp. 1236–1242.

[20] Y. Shao, C. Hardmeier, J. Tiedemann, and J. Nivre, “Character-based joint segmentation and POS tagging for Chinese using bidirectional RNN-CRF,” in arXiv:1704.01314. [Online]. Available: http://arxiv.org/abs/1704.01314

[21] R. Yin, Q. Wang, P. Li, R. Li, and B. Wang, “Multi-granularity Chinese word embedding,” in Proc. Conf. Empirical Methods Natural Lang. Process., 2016, pp. 981–986.

[22] Y. Lu, Y. Zhang, and D. Ji, “Multi-prototype Chinese character embedding,” in Proc. 10th Int. Conf. Lang. Resour. Eval. (LREC), 2016, pp. 855–859.
Y. Zhu et al.: Hybrid Classification Method via Character Embedding in Chinese Short Text With Few Words

[41] A. Chen, F. Peng, R. Shan, and G. Sun, “Chinese named entity recognition with conditional probabilistic models,” in Proc. 5th SIGHAN Workshop Chin. Lang. Process., 2006, pp. 173–176.

[42] D. Huang and J. Wang, “An approach on chinese microblog entity linking combining Baidu encyclopedia and word2vec,” Procedia Comput. Sci., vol. 111, pp. 37–45, Jan. 2017.

[43] S. Wang and C. D. Manning, “Baselines and bigrams: Simple, good sentiment and topic classification,” in Proc. 50th Annu. Meeting Assoc. Comput. Linguistics, vol. 2. 2012, pp. 90–94.

[44] V. Niculae, K. Sun, X. Chen, Y. Cheng, X. Du, E. Durmus, A. Katiyar, and C. Cardie, “Cornell belief and sentiment system at TAC 2016,” in Proc. TAC. 2016, pp. 1–2.

[45] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality,” in Proc. Adv. Neural Inf. Process. Syst., 2013, pp. 3111–3119.

YI ZHU received the B.S. degree from Anhui University, the M.S. degree from the University of Science and Technology of China, and the Ph.D. degree from the Hefei University of Technology. He is currently an Assistant Professor with the School of Information Engineering, Yangzhou University, China. His research interests are in data mining and knowledge engineering.

YUN LI received the M.Eng. degree in computer science and technology from the Hefei University of Technology, China, in 1991, and the Ph.D. degree in control theory and control engineering from Shanghai University, China, in 2005. He is currently a Professor with the School of Information Engineering, Yangzhou University, China. He has published more than 70 scientific articles. His research interests include data mining and cloud computing.

YONGZHEN YUE received the B.S. degree from the Anhui University of Chinese Medicine. He is currently pursuing the master’s degree with the Key Laboratory of Knowledge Engineering with Big Data (Hefei University of Technology) and the School of Computer Science and Information Engineering at the Hefei University of Technology, Hefei, China. His research interest is in natural language processing.

JIPENG QIANG received the B.S. degree in computer science from Hefei University, China, in 2010, and the M.S. and Ph.D. degrees in computer science from the Hefei University of Technology, China, in 2013 and 2016, respectively. He was a Visiting Ph.D. Student with the University of Massachusetts Boston, from October 2014 to June 2016. He is currently an Assistant Professor with the School of Information Engineering, Yangzhou University, China. His research interests include data mining and natural language processing.

YUNHAO YUAN received the Ph.D. degree in pattern recognition and intelligence system from the Nanjing University of Science and Technology (NUST), China, in 2013. He is currently an Associate Professor with the School of Information Engineering, Yangzhou University. He is the author or coauthor of more than 60 scientific articles. His research interests include pattern recognition, machine learning, multimedia search, and information fusion. He received two National Scholarships from the Ministry of Education, China, and an Outstanding Ph.D. Thesis Award from NUST.