Journal of Computers Vol. 33 No. 2, April 2022, pp. 115-126
doi: 10.53106/199115992022043302010

Named Entity Recognition Model Based on TextCNN-BiLSTM-CRF with Chinese Text Classification

Ji-Ru Zhang, Bo He*

School of Computer Science and Engineering, ChongQing University of Technology, Chongqing, 400054, China
jrzhangjr@163.com, hebo@cqut.edu.cn

Received 30 June 2021; Revised 1 October 2021; Accepted 1 November 2021

Abstracts. The traditional Chinese Named Entity Recognition (NER) method is difficult to define the entity category of the word according to the specific language environment, and the category is ambiguous, so it is difficult to accurately identify the entity. Named entity recognition based on deep learning can find entity categories in text, so it has received widespread attention. On the basis of the neural network model, this paper proposes a model based on TextCNN-BiLSTM-CRF and text classification (TextCNN-BiLSTM-TC-CRF) for Chinese NER. First, the TextCNN model is used to extract the word vector information of the text data; secondly, bidirectional LSTM is used the model extracts the contextual features of the text; then the neural network model is used to automatically extract the word features and the global features of the text for text classification; finally, the text sequence labeling and entity recognition are performed. Experiments verify that on a large-scale Chinese NER data set, the entity recognition model proposed in this paper has better evaluation indicators than other algorithms, with an F1 score of 98.7%.

Keywords: Chinese named entity recognition, text classification, TextCNN-BiLSTM, CRF

1 Introduction

With the continuous development of information technology and the increasing complexity of network language, accurate labeling of entity categories in different language environments has become the focus of attention. How to accurately label entity categories and eliminate ambiguity from massive text data is a hot topic in Chinese named entity recognition.

Chinese named entity recognition refers to finding entities in a paragraph of text and marking them out. Generally, entity labeling is called sequence labeling. Entity generally includes location, person, organization, time, etc. In some special fields, there will be corresponding entity names, such as crime types and ordinance names in legal documents. These entity information are helpful in semantic understanding and information decision-making. Traditional Chinese named entity recognition methods include rule-based methods, dictionary-based methods, statistics-based methods, and deep learning-based methods.

In order to solve the problem of entity labeling in a specific language environment, Ruokolainen et al. [1] proposes a method of using manual organization rules. This method is expensive, and the quality of the system mainly depends on the experience of experts. Moreover, the portability of the system is poor, and every rule change requires a lot of manual modification and even reorganization. In order to solve the problem of poor portability quickly, efficiently and cost-effectively, Jin et al. [2] proposes the use of machine learning methods, such as language model method, Hidden Markov Model, maximum entropy model, decision tree method, etc. The system of this type of method is portable the performance is strong, but the accuracy rate is not high. The subsequent methods based on deep learning have good adaptability and high accuracy, are currently the most used methods.

Although neural networks have good results in the field of named entity recognition, there are still many problems in entity recognition in specific language environments. Compared with traditional named entity recognition, there are mainly the following difficulties:

(1) Internet culture is developing rapidly, with a large number of Internet terms, and the frequency of change is very high. Unregistered words or phrases constantly appear as new entities, for example, new Internet terms, new abbreviations, etc. Therefore, the recognition rate of the method based on word segmentation is low.

(2) In different language scenarios, there is a problem of fuzzy classification of entity categories, and the boundaries between different types of entities are not clear. The same sequence of Chinese characters may have different entity types in different contexts, or it may be an entity in some states, but not an entity in some states. For instance:

* Corresponding Author

Person's name: “潮汐” refers to a person's name under certain conditions, and is a landscape under certain conditions;

Place name: “河南” is the name of a province under certain conditions, and is a general term under certain conditions;

Organization name: “大世界” refers to the name of the organization under certain conditions, and is just a phrase under certain conditions.

To solve the above problems, this paper proposes a Chinese NER model based on TextCNN-BiLSTM-CRF and text classification (TextCNN-BiLSTM-TC-CRF). The model first uses TextCNN to extract features for each word; secondly, BiLSTM is used to extract the context of the specific language environment of the text data; then the vector features of the words and the context features of the language environment are combined and passed to the TextCNN-BiLSTM model, the classification of the text language environment is carried out to achieve the purpose of semantic disambiguation; finally, the classification results of the language environment and the vector representation of the text are combined and passed into the CRF model for training to obtain the sequence labeling of the text, and the category entities are disambiguated to improve the accuracy of the Chinese NER model.

The main contributions of our work can be summarized as follows:

1. We present a novel neural network based on TextCNN-BiLSTM-CRF and text classification (TextCNN-BiLSTM-TC-CRF) for Chinese named entity recognition and entity disambiguation.

2. The combination of character-level neural network, Bidirectional LSTM network and Chinese NER which not only improves accuracy but also reduces time consumption. Results show that the combination of the character-level neural network, Bidirectional LSTM network and Chinese NER is effective.

3. We compare with other different models on the same data set. Experimental results indicate that our model can improve the accuracy of entity recognition and achieve the purpose of entity disambiguation. And our model has also achieved good results in other categories.

2 Related Work

Our work is inspired by two lines of research: deep learning for NER and text classification for entity disambiguation.

Collobert et al. [3] proposed a neural network architecture for the NER task without relying on any task-specific engineering or a lot of prior knowledge. Since then, researchers have paid more attention to using deep neural networks to deal with NER. Lee [4] used long short-term memory (LSTM) network and conditional random fields (CRF), which uses output-label dependencies with transition features and a CRF-like sequence-level objective function. Tang et al. [5] used a bidirectional LSTM network to model both past and future input features, and a CRF layer to recognize Chinese judicial named entity. Due to the ambiguous word boundaries and complex composition [6] Chinese NER task is more challenging compared with English NER. Wu et al. [7] used a CNN-LSTM-CRF neural architecture to capture both local and long-distance contexts for Chinese NER, which effectively improve the performance of Chinese named entity recognition, especially when training data is insufficient. Liu et al. [8] used Bi-directional Quasi-Recurrent Neural Networks to replace BiLSTM, which focused on model training efficiency. Zhao et al. [9] integrated the traditional bi-directional long-short-term memory structure and self-attention mechanism with dilated convolutional neural network to capture context information. He et al. [10] proposed a named entity recognition method that combines knowledge graph embedding with a self-attention mechanism, which the model can get additional auxiliary information.

Entity disambiguation is a fundamental task in natural language processing and computational linguistics. Sun et al. [11] developed various neural network architectures and explored a simple yet effective way that enables to collect millions of training examples from Wikipedia without using any manual annotation. Zuheros et al. [12] proposed a neural network architecture grounded in the use of long short-term memory recurrent neural network for encoding the context of a target geographical entity. Adjali et al. [13] used entity semantic similarity, context similarity, and mention probability for entity disambiguation. Francis-Landau et al. [14] used convolutional neural networks to capture semantic correspondence between a mention's context and a proposed target entity, these convolutional networks operate at multiple granularities to exploit various kinds of topic information, and their rich parameterization gives them the capacity to learn which n-grams characterize different topics. Hu et al. [15] used the multilayer perceptron to extract interaction features of missing data and observational data. Geng et al. [16] integrated convolutional and recurrent neural networks to disambiguate entities and extract relations together. It can acquire rich semantics and utilizes the full advantage of the associated information between entities and relations don't need external features. Chen et al. [17] integrated a BERT-based entity similarity score into the local context model of a state-of-the-art model to better capture latent entity type information.

The advantages and disadvantages of the above methods of entity recognition and entity disambiguation are summarized in Table 1 and Table 2.

Table 1. The advantages and disadvantages of entity recognition

Method	Advantages	Disadvantages
Named entity recognition based on CRF	Get the best label sequence in the global sense	Need a lot of manual annotation data
Named entity recognition based on deep learning	Discover hidden features	Convergence rates are uneven, requiring experience and skills
Named entity recognition based on BERT	Learn the contextual information of words well	Training time is too long, pre-training cost is too high

Table 2. Advantages and disadvantages of entity disambiguation

Method	Advantages	Disadvantages
Entity disambiguation based on neural network	No entity candidate set	Unclear distinction between entities
Entity disambiguation based on entity link	With target library, disambiguation is more accurate	Need a lot of labeled data, which is labor intensive

Inspired by these studies, this paper proposes a network model combining TextCNN-BiLSTM and text classification for the tasks of Chinese named entity recognition and entity disambiguation. The TextCNN network model can be used to obtain word vectors with character information, and the bidirectional long-short-term memory network model can obtain semantic sentence vectors. The second-layer TC model network takes the word vectors and sentence vectors obtained in the first layer as input to screen related topics. Remove irrelevant subject information to achieve the purpose of disambiguation, and at the same time obtain the global optimal annotation sequence through the conditional random field network, and effectively improve the accuracy of Chinese named entity recognition through the purpose of entity disambiguation.

3 Model Description

3.1 Overview

For Chinese NER, this paper proposes a named entity recognition model based on TextCNN-BiLSTM-CRF and text classification, as shown in Fig. 1. The model first convolves and pools each word through TextCNN to extract the word-level features of the word; secondly, the feature is passed to the BiLSTM neural network, and the BiLSTM neural network is used to extract specific language environment information features; then the word-level feature and the contextual text information feature are combined for text classification processing to obtain the category in the specific language environment; finally, the entity tagging of each word is performed using CRF, and the entity information in the sentence sequence is marked to obtain the optimal tag sequence.

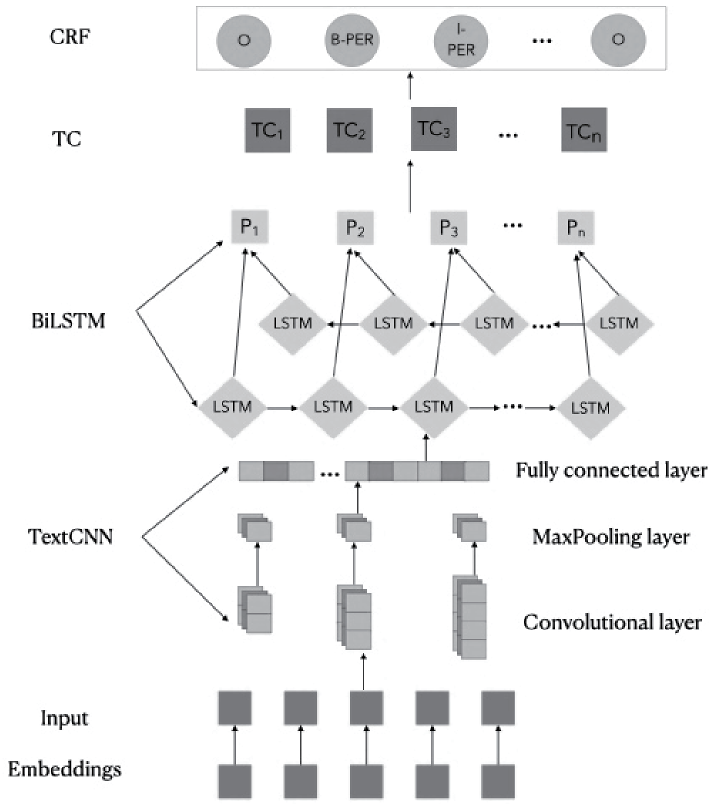


Fig. 1. Chinese NER model based on TextCNN-BiLSTM-TC-CRF

3.2 TextCNN

According to Kim's [18] research, this paper found that TextCNN can be effectively applied in text data classification. When TextCNN extracts key information in a text sequence, it uses multiple kernel convolutions of different sizes for extraction, which can better obtain the local correlation in the sequence.

The TextCNN structure used in this paper is shown in Fig. 2, w1hich is mainly composed of a word vector table, a convolutional layer and a pooling layer. First, the one-hot method is used to map the text data into a vector representation; secondly, based on the longest word, placeholders are filled at both ends of the word to make all word vector matrices the same size; then, fill the latter word vector matrix is passed into the convolutional layer to extract local features; finally, through the pooling layer, the dimensionality of the features is reduced to extract word-level features.

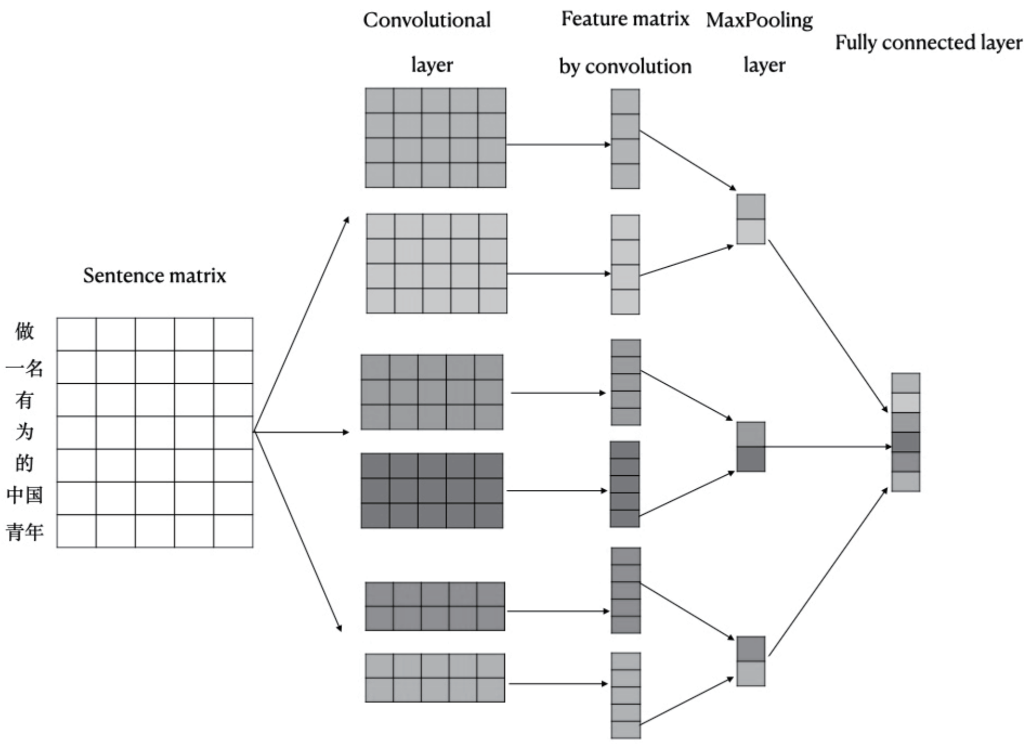


Fig. 2. TextCNN structure diagram

3.3 BiLSTM

In the task of text classification, text information is not only related to information at subsequent moments but also related to information at past moments. Therefore, not only past input features but also future input features are needed in model training. Therefore, this paper uses a bidirectional LSTM network as a training model. In the feature extraction process, the past features are extracted through the forward state, and the future features are extracted through the backward state. By using the bidirectional LSTM model, the performance of the LSTM model is greatly improved [19]. The structure diagram of bidirectional LSTM for text classification is shown in Fig. 3. Input the sequence of word vectors to the BiLSTM layer, forward LSTM to get the forward hidden layer sequence of the word vector $\overrightarrow{h_t} = \{\overrightarrow{h_1}, \overrightarrow{h_2}, \ldots, \overrightarrow{h_t}\}$. The backward LSTM obtains the backward hidden layer sequence of the word vector $\overrightarrow{h_t} = \{\overrightarrow{h_1}, \overrightarrow{h_2}, \ldots, \overrightarrow{h_t}\}$, splicing in order of position to get the final hidden layer sequence $h_c = [\overrightarrow{h_t}, \overrightarrow{h_t}]$, input the last hidden layer sequence h_c to the next layer. Use the softmax function to classify text, and the category with the highest probability is the category of the text.

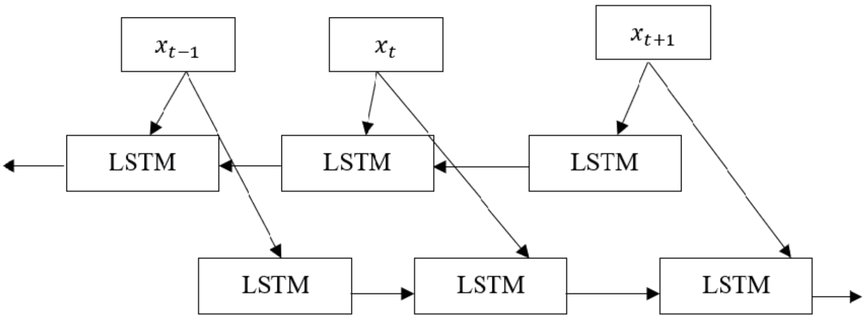


Fig. 3. BiLSTM structure diagram

3.4 TC

Text classification is the process of automatically categorizing texts according to a certain classification system or rules [20]. The purpose of text classification is to sort and classify text resources, while solving the problem of text information overload [21]. The TextCNN-BiLSTM-TC-CRF model in this paper can add specific language environment information to the sequence annotation model through text classification. The specific language environment refers to the context information of the corpus for entity recognition. Text classification is performed according to the context information, that is, the language environment of the corpus is limited. This can improve the accuracy of entity recognition and can also perform semantic disambiguation, identify the entity category.

3.5 CRF

In this paper, entity recognition is transformed into a sequence labeling problem. BiLSTM is very powerful in sequence modeling and can obtain long-term contextual information [22]. However, the tags of entity recognition tasks are not independent, but have strong dependencies, especially word-based entities, such as I-ORG after the B-ORG label, but not I-PER. Therefore, in this paper, we will access CRF behind TextCNN-BiLSTM-TC, because the CRF method can automatically set some legal constraints between tags, such as: the tag of the first word in a sentence can only be "B" or "O", not "I"; the label "B-label I-label I-label" and "label" in the statement should be the same named entity label [23].

The input sequence of the model in this paper is $X = (x_1, x_2, ..., x_n)$, where x_n is the input vector of the nth word, the corresponding label sequence is $Y =(y_1, y_2, ..., y_n)$. The score for each label is

$$Pi = W_s h^{(t)} + b_s \,, \tag{1}$$

where $h^{(t)}$ is the upper layer of the input data $X^{(t)}$ at time t; W_s and b_s are linear mapping functions. On this basis, CRF defines a label transfer function, and the score from the input sequence to the label sequence can be expressed as

$$s(x,y) = \sum_{i=1}^{n} \left(W_{y_{i-1},y_i} + P_{i,y_i}\right) \quad , \tag{2}$$

where, W is the transfer matrix $W_{i,j}$, denotes a transfer fraction; $P_{i,yi}$ represents the fraction of the tag y_i character.

The maximum likelihood estimation is used for the training of the CRF algorithm. The maximum likelihood function is shown in Equation (3), and the probability calculation formula is shown in Equation (4), which represents the probability corresponding to the original sequence to the predicted sequence.

$$L = \sum_{i=1}^{n} log\left(P(y_i|x_i)\right) + \frac{\lambda}{2}|\theta|^2 \;, \tag{3}$$

$$P(y|x) = \frac{e^{s(x,y)}}{\sum_{y \subseteq Y_x} e^{s(x,y)}} \;. \tag{4}$$

4 Experiments

4.1 Training Process

This paper implements a Chinese NER model based on TextCNN-BiLSTM-TC-CRF on the Tensorflow platform. Table 3 shows the training process of the model.

Table 3. TextCNN-BiLSTM-TC-CRF model training process

Algorithm 1. Model Training

Input: $X = (x_1, x_2, ..., x_n)$
Output: $Y = \{y_1, y_2, ..., y_n\}$
for each epoch *do*:
for each batch *do*:
1. Corpus preprocessing
2. TextCNN model to extract word-level feature information
3. The BiLSTM model extracts the contextual information of the text
4. TextCNN- BiLSTM-TC model, the combination of word vector features and text context information realizes the classification of specific language environment categories
5. CRF is passed forward and backward to calculate the optimal probability of the tag sequence
6. TextCNN-BiLSTM-TC-CRF model, entity category disambiguation, calculate the optimal sequence annotation

end *for*
end *for*

It can be seen from Table 3 that for each epoch, the model will be processed in batches. Each epoch processes a batch of data, and the size of each batch of data is determined by batch_size. The model training process first preprocesses the corpus; then performs word-level information feature extraction through TextCNN, then uses BiLSTM to extract the location and context information of the text; then classifies the text according to the word vector and specific language environment information; finally uses the CRF model the information is extracted forward and backward, and the optimal sequence label is calculated.

4.2 Experimental Setup

Dataset

The corpus used in the experiment includes the data from the Today's headlines and the news tag corpus of "People's Daily". Among them, there are 382,688 pieces of text classification data, of which 100,000 pieces are used for training and 180,000 pieces are used for testing. There are 19,484 pieces of entity recognition corpus data. As shown in Table 4, it is a statistical table of corpus data.

Table 4. Corpus statistics

Dataset	Training	Testing
Sentence	11690	7794
Labels	91885	76962
PER	20000	27790
LOC	37186	25438
ORG	34699	23734

Evaluation Standard

This paper uses precision (Pression, P), recall (Recall, R) and F1-score (F1-score, F) as evaluation indicators. The calculation formula is as follows:

$$P = TP/\ TP + FP\ , \quad \textbf{(5)}$$

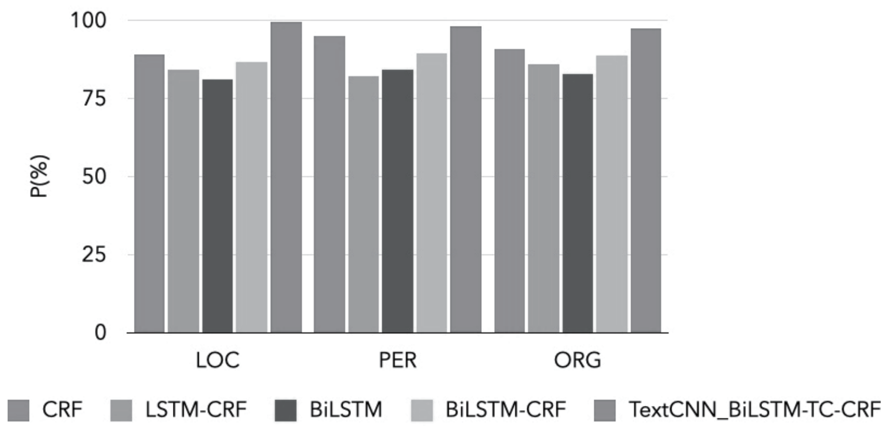
$$R = TP/\ TP + FN\ , \quad \textbf{(6)}$$

$$P = 2 \times P \times R/\ P + R\ . \quad \textbf{(7)}$$

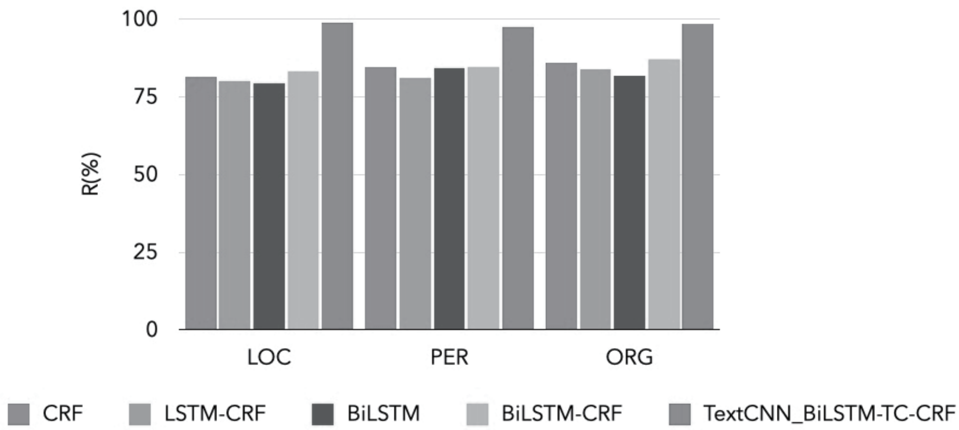
Where P represents the proportion of correctly predicted positive classes to the total number of all predicted positive classes; R represents the proportion of correctly predicted positive classes to the total number of true positive classes in the sample; F is the harmonic average of P and R, which means both the combined effect of.

4.3 Overall System Results

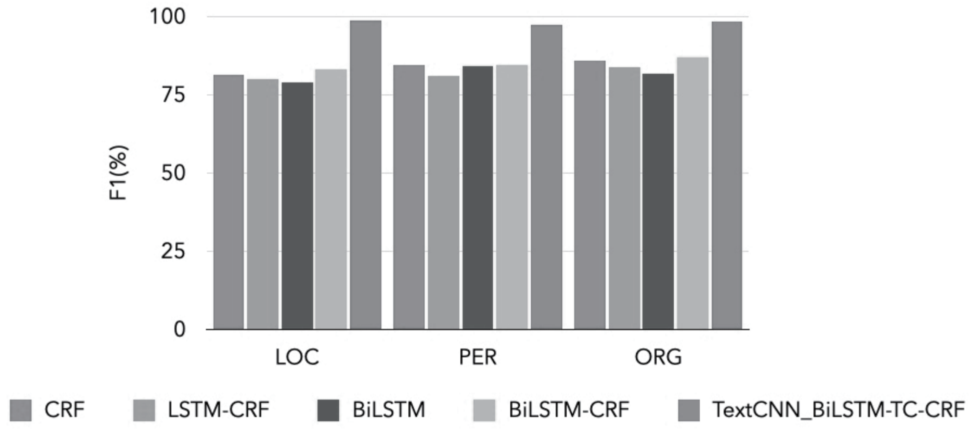
In order to verify the effect of the proposed model in Chinese NER, comparative experiments were carried out on different algorithm models. The comparison models in the experiment include: CRF [24], LSTM-CRF [25], BiLSTM [26], BiLSTM-CRF [27] and TextCNN-BiLSTM-TC-CRF proposed in this paper. All experimental comparisons are trained and tested on the same public data set. Fig. 4 shows the comparison results of the three types of entities on different models.



(a) Precision



(b) Recall



(c) F1-score

Fig. 4. Experimental results in different models, evaluated by precision, recall and F1-score (%)

It can be seen from Fig. 4 that the precision of the CRF model is significantly higher than that of the LSTM and BiLSTM models, because CRF can obtain the global optimal sequence, and LSTM and BiLSTM can only calculate the ratio to obtain the label. Especially in the PER category, the precision of the CRF model is particularly high. This is because the Chinese PER has obvious characteristics, and most of the names are two or three characters. Although CRF can constrain some tags, it cannot consider long-term text information, and has obvious disadvantages in labeling longer text information. LSTM can effectively make up for this disadvantage. It can be seen from Fig. 4 that although the LSTM-CRF model is not necessarily more accurate than the CRF model, BiLSTM has been improved for the first time on this basis. This is because on the basis of LSTM, BiLSTM can obtain information in both directions. The BiLSTM-CRF model combines the advantages of CRF and BiLSTM models. It can not only obtain contextual information but also use the constraints of the CRF model to obtain the optimal sequence labeling. Nevertheless, the model ignores the word information, especially in Chinese named entity recognition, the word information is particularly important in entity recognition. In this paper, the TextCNN model is used to extract the characteristics of the word information, and then the BiLSTM model is used to obtain the information characteristics of the context, and then the data layering of the TC layer is carried out to filter out unnecessary information, which improves the accuracy to a large extent, and then uses the CRF model Select the global optimal label. Compared with the BiLSTM-CRF model, the F1-score of the TextCNN-BiLSTM-TC-CRF model is increased by 3.2%, indicating that the vector information extracted by TextCNN can represent the feature information to a certain extent, and BiLSTM can obtain the language environment information of the text in both directions. The addition of TC makes the judgment of the language environment and the disambiguation of the entity category reached the highest scores compared with other models, which are 97.3%, 98% and 98.7%, respectively. It also further verifies that the text classification model is added to extract text information. The language environment can improve the recognition performance of entity recognition.

Table 5. Comparison of different named entity recognition methods (%)

Models	P	R	F1
CNN	86.52	55.07	67.06
CRF	79.5	81.5	81.5
LSTM-CRF	84.2	84.3	80.2
BiLSTM	81.08	79.21	84.3
BiLSTM-CRF	86.82	83.21	83.21
BiLSTM-CNN-CRF	88.1	83.2	85.6
Attention-BiLSTM-CNN-CRF	89.3	88.8	89.5
Yang et al. [28]	92.04	91.31	91.67
Dong et al. [29]	91.28	90.62	90.95
BERT-BiLSTM-CRF	91.69	92.71	96.27
TextCNN-BiLSTM-TC-CRF (ours)	*97.3*	*98*	*98.7*

By analyzing the experimental data in Table 5, it is found that compared with traditional named entity recognition methods such as CRF, LSTM-CRF, BiLSTM-CRF, the experimental results of our method are better than the former method. The method after adding the LSTM neural network model to the CRF is better than the CRF method, because LSTM solves the problem of length dependence and can deeply mine the feature information hidden in the text. It also shows that the deep learning method is superior in the effect of named entity recognition based on traditional statistical methods. The difference between the BiLSTM-CRF and the LSTM-CRF is not particularly obvious. It may be because the parameter settings are not optimal or are affected by other minor factors. The experimental effect after adding CNN is significantly improved compared with the previous experimental results. This is because the high computing power of CNN can calculate the character representation vector, which is conducive to the expression of sentence context information. With the addition of the Attention mechanism, the model's tagging prediction capabilities are further strengthened. Different words in a sentence contribute to the context in different degrees. The addition of local features of the text in the feature extraction process makes up for the traditional method that only focuses on global feature extraction. The defects can effectively extract more contextual features. Yang et al. [28] gave the combination of CNN, Bi_LSTM, and CRF-based character to extract stroke embedding and n-gram features for Chinese NER. Dong et al. [29] introduce radical features into LSTM-CRF. The experimental results had been improved to a certain extent. We use BERT pre-training for Chinese named entity recognition, and the result has been greatly improved, but the training time is too long. In this paper, TextCNN is used to reduce the training time while ensuring that sufficient context information is obtained. After that, the TC layer is added to filter text data, and then entity recognition is performed. The experimental results are significantly improved. These show that proposed techniques are very useful for Chinese entity disambiguation. Additionally, we achieved F1-score of 98.7%.

4.4 Chinese NER Instance

Apply the model parameters obtained from the above training to other specific language environments. The field scalability experiment results are shown in the Fig. 5.

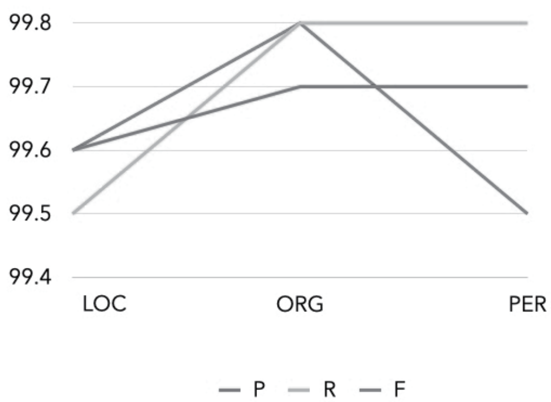


Fig. 5. Field scalability experiment results (%)

It can be seen from the figure that the model parameters are migrated to other language environments. This paper uses travel news data, which also shows better performance. The recall, precision and F1-score are all over 99%, which fully explains the model has good domain scalability, and effectively solves the problem of low accuracy of Chinese NER. The proposed model in this paper can more accurately discover Chinese entities, help users make decisions and solve related problems.

5 Conclusions and Future Work

Aiming at Chinese NER, this paper proposes a Chinese NER model based on TextCNN-BiLSTM-CRF and text classification. The model first obtains word-level vector representation through TextCNN and BiLSTM obtains the specific context information of the text, and then vectorizes it. The representation and the contextual information of the text are combined for text classification processing, and finally the entities in the specific language environment are marked through the CRF model to obtain the optimal mark sequence. This paper uses a labeled text data set to conduct a comparative experiment of Chinese entity recognition methods. The experimental results show that the model proposed in this paper has higher precision and F1-score than other models. Therefore, the TextCNN-BiLSTM-TC-CRF model extracts word-level features and rich language environment features. The addition of word-level features enables the model to effectively recognize newly emerging word categories. A large number of rich language environment features effectively compensate for the entity ambiguity in Chinese entity recognition, thereby improving the performance of the model. And the model proposed in this paper also has a better domain expansion type, which has a better performance on different domain data sets.

Since this paper only considers the contextual features of the entity and does not consider more local features, it is planned to introduce the attention mechanism into the model of this article in the future work, and the use of the attention mechanism to extract local features is the next step direction.

Acknowledgement

This research is supported by the Humanities and Social Sciences of Ministry Education of China (No. 19XJA910001) and the postgraduate innovation fund project of Chongqing University of Technology (No. clgycx 20203114).

References

[1] T. Ruokolainen, P. Kauppinen, M. Silfverberg, K. Lindén, A Finnish News Corpus for Named Entity Recognition, Lang Resources & Evaluation 54(2020) 247–272.
[2] C. Jin, Z. Shi, W. Li, Y. G, Bidirectional LSTM-CRF Attention-based Model for Chinese Word Segmentation. <https://arxiv.org/abs/2105.09681>, 2021 (accessed 21.10.08).
[3] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, P. Kuksa, Natural Language Processing (almost) from Scratch, Journal of machine learning research 12(2011) 2493-2537.
[4] C. Lee, LSTM-CRF Models for Named Entity Recognition, IEICE transactions on information and systems 100(4)(2017) 882-887.
[5] P. Tang, P. Yang, Y. Shi, F. Lin, Y. Wang, Recognizing Chinese Judicial Named Entity Using BiLSTM-CRF, in: Proc. The Third International Conference on Physics, Mathematics and Statistics, 2020.
[6] T. Gui, Y. Zou, Q. Zhang, M. Peng, J. Fu, Z. Wei, X. Huang, A Lexicon-based Graph Neural Network for Chinese NER, in: Proc. 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, 2019.
[7] F. Wu, J. Liu, C. Wu, Y. Huang, X. Xie, Neural Chinese Named Entity Recognition Via CNN-LSTM-CRF and Joint Training with Word Segmentation, in: Proc. The World Wide Web Conference, 2019.
[8] K. Liu, Q. Yu, S. Zhong, Chinese Named Entity Recognition Based on Bi-directional Quasi-Recurrent Neural Networks Improved with BERT: New Method to Solve Chinese NER, in: Proc. The 5th International Conference on Innovation in Artificial Intelligence, 2021.
[9] D. Zhao, J. Cao, D. Huang, J. Meng, P. Zhang, Dual Neural Network Fusion Model for Chinese Named Entity Recognition, International Journal of Computational Intelligence Systems 14(1)(2021) 471-481.
[10] S. He, D. Sun, Z. Wang, Named Entity Recognition for Chinese Marine Text with Knowledge-based Self-attention, Multimedia Tools and Applications (2021) 1-15.
[11] Y. Sun, Z. Ji, L. Lin, D. Tang, X. Wang, Entity Disambiguation with Decomposable Neural Networks, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 7(5)(2017) e1215.
[12] C. Zuheros, S. Tabik, A. Valdivia, E. Martínez-Cámara, F. Herrera, Deep Recurrent Neural Network for Geographical Entities Disambiguation on Social Media Data, Knowledge-Based Systems 173(2019) 117-127.
[13] O. Adjali, R. Besançon, O. Ferret, H. Le Borgne, B. Grau, Multimodal Entity Linking for Tweets, in: Proc. of the Advances in Information Retrieval-42nd European Conference on {IR} Research, 2020.
[14] M. Francis-Landau, G. Durrett, D. Klein, Capturing Semantic Similarity for Entity Linking with Convolutional Neural Networks, in: Proc. of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016.
[15] X. Hu, Y. Han, Z. Geng, A Novel Matrix Completion Model Based on the Multi-layer Perceptron Integrating Kernel Regularization, IEEE Access 9(2021) 67042-67050.
[16] Z. Geng, Y. Zhang, Y. Han, Joint Entity and Relation Extraction Model Based on Rich Semantics, Neurocomputing 429(2021) 132-140.
[17] S. Chen, J. Wang, F. Jiang, C. Lin, Improving Entity Linking by Modeling Latent Entity Type Information, in: Proc. of

the AAAI Conference on Artificial Intelligence, 2020.
[18] Y. Kim. Convolutional Neural Networks for Sentence Classification, in: Proc. 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014.
[19] Y.-M. Huang, Y. Jiang, T. Hasan, Q. Jiang, C. Li, A topic BiLSTM Model for Sentiment Classification, in: Proc. 2nd International Conference on Innovation in Artificial Intelligence, 2018.
[20] J.-X. Yang, L. Bai, Y.-M. Guo, A Survey of Text Classification Models, in: Proc. 2020 2nd International Conference on Robotics, Intelligent Control and Artificial Intelligence, 2020.
[21] M. Iqbal, M.-M. Abid, M.-N. Khalid, A. Manzoor, Review of Feature Selection Methods for Text Classification, International Journal of Advanced Computer Research 10(49)(2020) 2277-7970.
[22] B. Lin, F. Xu, Z. Luo, K. Zhu, Multi-channel Bilstm-CRF Model for Emerging Named Entity Recognition in Social Media, in: Proc. 3rd Workshop on Noisy User-generated Text, 2017.
[23] J. Zhang, Y. Wang, X. Yang, F. Wang, Entity Recognition of Chinese Medical Literature Based on BiLSTM-CRF and Fusion Features, in: Proc. 2020 3rd International Conference on Big Data Technologies, 2020.
[24] J. Lafferty, A. McCallum, F.-C.-N. Pereira, Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data, in: Proc. 18th International Conference on Machine Learning, 2001.
[25] C. Lee, LSTM-CRF Models for Named Entity Recognition, IEICE transactions on information and systems 100(4)(2017) 882-887.
[26] L. Luo, Z. Yang, P. Yang, Y. Zhang, L. Wang, H. Lin, J. Wang, An Attention-based BiLSTM-CRF Approach to Document-level Chemical Named Entity Recognition, Bioinformatics 34(8)(2018) 1381-1388.
[27] Z. Huang, W. Xu, K. Yu, Bidirectional LSTM-CRF Models for Sequence Tagging. <https://arxiv.org/abs/1508.01991>, 2015 (accessed 20.07.12).
[28] F. Yang, J. Zhang, G. Liu, J. Zhou, C. Zhou, H. Sun, Five-stroke Based CNN-BiRNN-CRF Network for Chinese Named Entity Recognition, in: Proc. CCF International Conference on Natural Language Processing and Chinese Computing, 2018.
[29] C. Dong, J. Zhang, C. Zong, M. Hattori, H. Di, Character-based LSTM-CRF with Radical-level Features for Chinese Named Entity Recognition, Natural Language Understanding and Intelligent Applications 10102(2016) 239-250.