

## Research Article

# Chinese Named Entity Recognition Based on Character-Word Vector Fusion

Na Ye , Xin Qin, Lili Dong, Xiang Zhang, and Kangkang Sun

*School of Information and Control Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China*

Correspondence should be addressed to Na Ye; [yananaye2003@126.com](mailto:yananaye2003@126.com)

Received 6 April 2020; Revised 28 May 2020; Accepted 13 June 2020; Published 4 July 2020

Academic Editor: Fuhong Lin

Copyright © 2020 Na Ye et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Due to the lack of explicit markers in Chinese text to define the boundaries of words, it is often more difficult to identify named entities in Chinese than in English. At present, the pretreatment of the character or word vector models is adopted in the training of the Chinese named entity recognition model. Aimed at the problems that taking character vector as an input of the neural network cannot use the words' semantic meanings and give up the words' explicit boundary information, and taking the word vector as an input of the neural network relies on the accuracy of the segmentation algorithms, a Chinese named entity recognition model based on character word vector fusion CWVF-BiLSTM-CRF (Character Word Vector Fusion-Bidirectional Long-Short Term Memory Networks-Conditional Random Field) is proposed in this paper. First, the Word2Vec is used to obtain the corresponding dictionaries of character-character vector and word-word vector. Second, the character-word vector is integrated as the input unit of the BiLSTM (Bidirectional Long-Short Term Memory) network, and then, the problem of an unreasonable tag sequence is solved using the CRF (conditional random field). By using the presented model, the dependence on the accuracy of the word segmentation algorithm is reduced, and the words' semantic characteristics are effectively applied. The experimental results show that the model based on character-word vector fusion improves the recognition effect of the Chinese named entity.

## 1. Introduction

In a broad sense, the purpose of named entity recognition (NER) is to recognize the named entity in the text and classify it into the corresponding entity types. Usually, entity types include people's names, place names, names of organizations, and dates [1, 2]. As a basic task in Natural Language Processing (NLP), NER is widely used in knowledge map construction, information extraction, machine translation, public opinion monitoring, etc. [3]. In the early time, the rule-based methods which mainly applied the rule templates constructed by the linguistic experts are mostly adopted. These methods have problems such as high artificial cost and poor portability. So now, the research is primarily focused on the probability statistics-based and deep learning-based methods [4].

The NER methods based on probability statistics and deep learning both need to represent the text as a vector. However, there are no explicit marks in the Chinese text to define the boundary among words, so two types of prepro-

cessing methods including the character vector-based methods and the word vector-based methods are generally used to transform the text to the vector. In the character vector-based pretraining, the character dictionary is obtained by learning the context features of the character. The advantage is that the dictionary has a small dimension and there is no hidden noise data in it. The disadvantage is that the boundary information of the word is cut off and the semantic features hidden in the word are lost. In the pretraining of the word vector, the word segmentation is carried out first and then the corresponding word dictionary is generated. The problem of this kind of method is that the word segmentation error may cause the named entity to be wrongly segmented, and the dimension of the dictionary is big. The advantage is that it retains the word boundary information and semantic features.

To solve the problems in Chinese named entity recognition based on character vector or word vector, Chen et al. [5] improved the generation of the word vector. Through

the joint training of the character and word vector, the information of a single Chinese character that makes up a word is introduced, which improved the quality of word vector generation, but this method is carried out on the premise of correct word segmentation, and it cannot solve the problem that a single Chinese character has different semantics in different words. Ma and Hovy [6] used CNN (Convolutional Neural Network) to extract features at the character level and then took the pre-trained word vector as the input of the bidirectional recurrent neural network. Finally, the conditional random field was used to establish the dependence on the output tag. This method achieved good results in two foreign language evaluation tasks. Lample et al. [7] used BiLSTM to extract character-level features, which were fused with the word vectors in dictionaries to form the final input vector, and the BiLSTM and the CRF model were combined to do the named entity recognition, which has achieved good results in English, German, Spanish, and other testing corpus. Both the methods proposed by Ma and Hovy and Lample et al. leveraged the word vector to do named entity recognition in foreign language corpus, during which the accuracy of word segmentation needed not to be considered, but the accuracy of word segmentation in Chinese corpus cannot be avoided. Zhang and Yang [8] adopted the bidirectional recurrent network structure of the word grid for named entity recognition. Compared with the named entity recognition models based on the character vector and the word vector, the model took the use of significant word information without word segmentation errors. However, when the model was used for named entity recognition, it needed to dynamically match the input character sequence to the corresponding words in the dictionary, which undoubtedly increased the training time and complexity of the model.

In order to solve the above problems, a preprocessing method of the character vector is proposed in this paper, which not only ensures that the word segmentation results are not affected but also merges the characteristics of the word, and experiments are carried out based on the model of the BiLSTM network combined with the CRF.

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

- (1) A Chinese NER model combining the character-word vector fusion, BiLSTM, and the CRF is proposed
- (2) The character-word vector fusion is key to the Chinese named entity recognition, and we propose a way to process the vector by fusing the character vector and the word vector which the character is contained

The rest of the paper is organized as follows. Section 2 gives a detailed description of the proposed model. Section 3 presents extensive experiments to verify the effectiveness of our proposal, and Section 4 summarizes this work.

## 2. CWVF-BiLSTM-CRF Model

The overall structure of the Chinese named entity recognition model CWVF-BiLSTM-CRF constructed in this paper is shown in Figure 1.

The model is divided into three layers, that is, embedding layer with character word vector fusion, BiLSTM layer, and CRF layer. The transformation of the input to the vector is carried on in the embedding layer. The characters and words in the annotated data set are replaced by pretrained character-word vector, and the character vector and the word vector are added as the representation of the character to form the final input vector. When the BiLSTM layer receives the current input vector, it extracts the context features of the current input and then integrates the output of the forward LSTM (Long-Short Term Memory) network and the reverse LSTM network as the input of the CRF layer. The CRF layer calculates the output at the current moment based on the output of the previous moment and finally predicts the label of a single character.

*2.1. Embedding Layer.* Firstly, the Word2Vec model is used to train the distributed character vector and word vector for the corpus, and the corresponding dictionary of character-character vector and word-word vector was obtained. The corresponding character vector and word vector of the training data annotated by BIO can be read in the dictionary and then are fused as the input of the model. In order to solve the problem that different characters have different meanings in different words, this paper expands the length of the corpus after word segmentation to the same as the length of that after the character segmentation. Figure 2 shows a concrete example of character-word vector fusion. By using the word placeholder processing, the length of the corpus after word segmentation is transformed as the same as the length after character segmentation. And in the training process, the corresponding vectors of the words and the characters are obtained through the dictionary, and the final fused vector which is formed by adding the word and character vectors is used as the input vector of the model. The words' placeholders are used to ensure that the same character has different vectors in different words, for example, "Ying-de Wang" and "Ying Guo" represent the name of a person and England, respectively. Due to the "Ying" in the two words has different semantics, using the character vector as the representation will ignore the semantic information of the word that the character is in, while using the word vector as the representation will not only make the dimension of the vector larger but also cause a potential error due to the word segmentation. Therefore, the character-word vector fusion is proposed to solve the above problems. In order to facilitate the fusion of the word vectors and the character vectors during the training, the words' information is added in the character-word vectors. The character-word vector after the fusion is expressed as

$$e = e_{\text{character}} + e_{\text{word which the character is in}} \quad (1)$$

*2.2. BiLSTM Layer.* LSTM [9] is a variant of the recurrent neural network (RNN) [10]. It is capable of selectively "remembering" the previous features while retaining the ability to process time series, so as to solve the problem that the common neural network cannot process the information of the long-time before moments [11]. BiLSTM [12] is

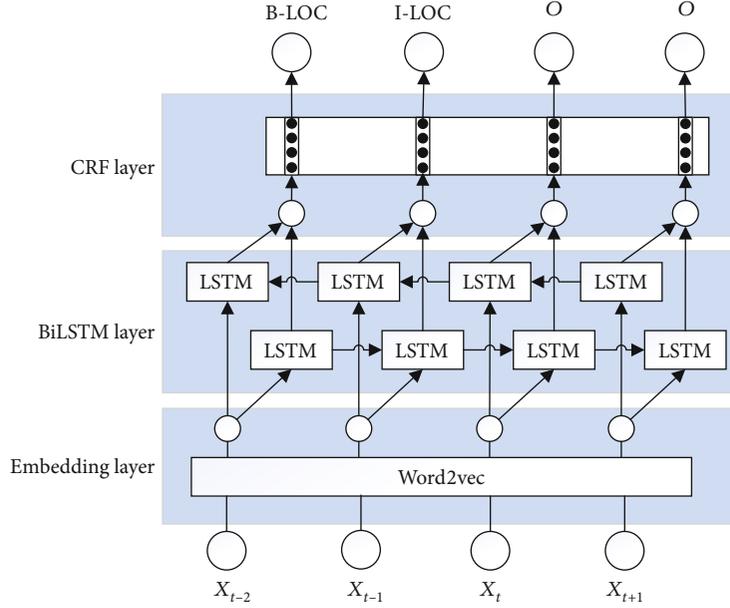


FIGURE 1: CWVF-BiLSTM-CRF entity recognition model. The model includes the embedding layer, BiLSTM layer, and CRF layer.

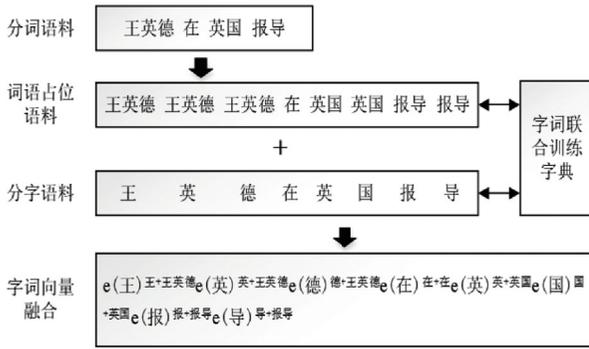


FIGURE 2: Character word vector fusion. The words' information is added in the character-word vectors.

composed of two LSTM networks in different time directions. By extracting the context features of the input unit, the exact meaning of the input unit can be obtained. The BiLSTM network consists of a two-layer LSTM network in the direction of the positive sequence and the reverse sequence. The input sequence is simultaneously input into the two-layer LSTM network at a certain moment. The whole time series is as follows: the input sequence in the positive sequence direction is  $X_1, X_2, X_3, \dots, X_t$ . The input direction in the reverse time sequence is  $X_t, X_{t-1}, X_{t-2}, \dots, X_1$ . During the learning of the context characteristics in the positive and reverse directions, the LSTM networks in the two directions do not share the state. The output of the two LSTM networks in different directions will be concatenated, the dimension transformation will be carried out in the linear layer and the normalization will be done by the Softmax, and then, the final output is obtained.

**2.2.1. Forward Propagation.** The forward propagation of the BiLSTM model is based on the general deep neural network,

which is the process in which the training data is calculated by the input layer, hidden layer, output layer, and the weights between each layer. In the CWVF-BiLSTM-CRF model shown in Figure 1, suppose that the input unit is  $X = \{\}$ , the forward hiding unit of the BiLSTM is  $H^0 = \{\}$ , the reverse hiding unit of the BiLSTM is  $H^1 = \{\}$ , and the output unit is  $O = \{\}$ . The weight between the input unit and the hidden unit is set as  $W_0$ , and the weight between the hidden unit and the output unit is set as  $W_1$ . The specific steps of the forward propagation process are as follows.

- (1) Firstly, character segmentation and word segmentation processing are carried out on the training data, and then, the corpus data after the segmentation are pieced together and labeled with BIO. The distributed vector training is carried out based on the character-word fused vector on the Chinese corpus of Wikipedia. On the premise of taking the character vector as the input unit, in order to ensure that the character can retain the semantic information of the word it belongs to, the word information is integrated into the character vector. The fused vector is taken as the input  $X_t$  of the model
- (2) The output  $H^0_t$  of the forward hidden unit at the current moment is obtained after the linear transformation and nonlinear transformation of the input  $X_t$  at the current moment and the output  $H^0_{t-1}$  at the previous moment received by the forward hidden unit of BiLSTM. The backward hiding unit of the BiLSTM is similar to the forward hiding unit. The output  $H^1_t$  of the backward hiding unit at the current moment is obtained after the transformation of the input  $X_t$  at the current moment and the output  $H^1_{t-1}$  at the previous moment. The output  $O_t^0$  of the BiLSTM model at the current time is generated by the Softmax

activation function after splicing the output vectors of the forward and backward hidden units by location at the current time

- (3) The output  $O_t^0$  of the BiLSTM layer only represents the tag with the highest probability in the probability distribution of all tags at the current time, but the dependence between the output at the previous moment and the current moment in the BiLSTM layer is not taken into account. After using the CRF layer,  $O_t^0$  expanded according to the time series can be seen as a general form of the linear chain conditional random field. The final tag prediction sequence  $O_t^1$  can be obtained from the given random variable sequence  $O_t^0$ , and the conditional random field problem can be solved by using the Viterbi algorithm

**2.2.2. Backward Propagation.** The training process of the neural network can be divided into two stages. In the first stage, the predicted value of the model can be obtained through forward propagation, and in the second stage, the backward propagation is carried out. The idea of the backward propagation is to calculate the error between the predicted value and the real value of the neural network using the loss function. Then, the error will be transferred in the reverse direction layer by layer. The gradient descent method is used to update the model parameters layer by layer, and finally, the model reaches the convergence state. The backward propagation process of this model is as follows.

The loss function is used to calculate the error between the predicted tag value  $O_t^1$  and the real tag value of the model, namely, the CRF layer error, and the CRF layer transition matrix parameters are updated according to the gradient descent method.

The backward hidden layer error of the BiLSTM layer is calculated, and the parameters of the backward hidden layer are updated by the gradient descent method. The forward hidden layer error of the BiLSTM layer is calculated, and the forward hidden layer parameters are also updated by the gradient descent method.

**2.3. CRF Layer.** The probability of each character corresponding to the tags is obtained by the output of the BiLSTM layer through the Softmax function, but the output of the Softmax function at each moment is independent of each other, without considering the sequential nature of the tags. After applying the CRF layer, the model combines the probability calculated by the Softmax function of the tag corresponding to each word with the mutual transfer probability between labels, instead of just the combination of the words corresponding to the maximum probability of labels at each moment, which makes up for the deficiency of the BiLSTM model.

### 3. Experimental Analysis

**3.1. The Experimental Scheme.** According to the task of named entity recognition, two sets of experiments are designed. The first experiment is the tuning experiment of the model parameters, which aims to find the optimal

parameters. The second experiment is a comparison with the model based on word vector training, which aims to verify the effectiveness of the character-word fusion vector on named entity recognition. The detailed design of the two groups of experiments is as follows.

**3.1.1. Experiment 1: Model Parameter Tuning Experiment.** When the same algorithm is adopted, the structure of the neural network has a great influence on the accuracy of the model. In order to search for the optimal structure of the named entity recognition model, this experiment performs a tuning experiment on common parameters that affect the performance of the model. The main parameters considered in the experiment are batch\_size, optimizer, hidden layer nodes, and learning rate. According to the training experience of other papers, the initial parameters of the model are set as follows: Batch\_size: 128, optimizer: SGD (Stochastic Gradient Descent), number of nodes in the hidden layer: 200, and learning rate: 0.005. Considering that the training corpus is too large and in order to avoid overfitting in the training process, the dropout is set to 0.5 because at this value, the number of the network structure generated randomly is the largest and the effect is the best.

**3.1.2. Experiment 2: Comparative Experiment with the Reference Model.** Through the parameter tuning in experiment 1, the optimal hyperparameters of this model are selected for training. In addition, in order to verify the effectiveness of character-word vector fusion, the comparison experiment is carried out with the reference model which adopts word vector on the premise that the data and the model are the same, and the  $F_1$  value is selected as the evaluation criteria in the comparison experiment.

**3.2. Preprocessing of the Dataset.** In this paper, the annotated corpus data of the People's Daily in 1998 is adopted to train the Chinese named entity recognition model. In this corpus, the part of speech tagging of each word is carried out using 26 basic speech markers, among which the annotation for 4 types of nouns is very important for the identification of the named entities, that is, person name *nr*, place name *ns*, organization name *nt*, and other proper nouns *nz*.

**3.2.1. Pretreatment Steps.** The specific steps to process the corpus data of People's Daily are as follows.

- (1) Only the labels of the human name, the place name, and the organization name in the original corpus are retained, these three labels are annotated with single Chinese characters in the format of BIO annotation, and the format of each line of data after annotation is *character-the corresponding label*
- (2) The original corpus is processed by word segmentation, and the words are copied and inserted so as to expand the corpus to the same length as the corpus in step (1)
- (3) The processed corpus in step (1) and step (2) is pieced together to form the final annotated corpus. The format of each line of the corpus is *character-the corresponding word-the corresponding label*

19980101-03-005-004/m	叶利钦/nr 是/v 在/p 接受/v 莫斯科/ns 新闻/n 媒体/n	叶 叶利钦 B-PER	新 新闻 0	发 发表 0
19980101-03-005-005/m	叶利钦/nr 指出/v , /w 1997年/t 俄罗斯/ns 政治/n	利 叶利钦 I-PER	闻 新闻 0	表 发表 0
19980101-03-005-006/m	在/p 谈/v 到/v 外交/n 形势/n 时/Ng 叶利钦/nr 说/v	钦 叶利钦 I-PER	媒 媒体 0	上 上述 0
19980101-03-005-007/m	叶利钦/nr 指出/v , /w 虽然/c 1997年/t 俄罗斯/ns	是 是 0	体 媒体 0	述 上述 0
19980101-03-005-008/m	新华社/nt 波恩/ns 12月/t 30日/t 电/n (/w 记者/n	在 在 0	的 的 0	看 看法 0
19980101-03-005-009/m	金克尔/nr 说/v , /w 1997年/t 是/v 欧盟/j 各/r 成	接 接受 0	联 联合 0	法 看法 0
19980101-03-005-010/m	在/p 谈/v 到/v 欧洲/ns 的/u 政治/n 和/c 外交/n 问题	受 接受 0	合 联合 0	的 的 0
19980101-03-005-011/m	他/r 说/v : /w “ /w 《/w 阿姆斯特丹/ns 条约/n 》/w	莫 莫斯科 B-LOC	采 采访 0	。 。 0</td
19980101-03-005-012/m	新华社/nt 基辅/ns 12月/t 30日/t 电/n 乌克兰/ns 外	斯 莫斯科 I-LOC	访 采访 0	他 他 0
		科 莫斯科 I-LOC	时 时 0	在 在 0

(a)

(b)

FIGURE 3: Contrast graph before and after data labeling. (a) The data before labeling. (b) The data after the BIO labeling.

TABLE 1: Relationship between predicted and actual values.

Predication Actual	Positive	Negative
True	TP	TN
False	FP	FN

(4) Finally, the corpus data in step (3) is divided into a training set, verification set, and test set according to the ratio of 3:1:1

Figure 3 shows part of the corpus before and after labeling by BIO of the People’s Daily. The BIO label in the figure (b) is to mark the character at the beginning of each line, and the words between the character and the BIO label are added to facilitate the fusion of the character-word vector during the training.

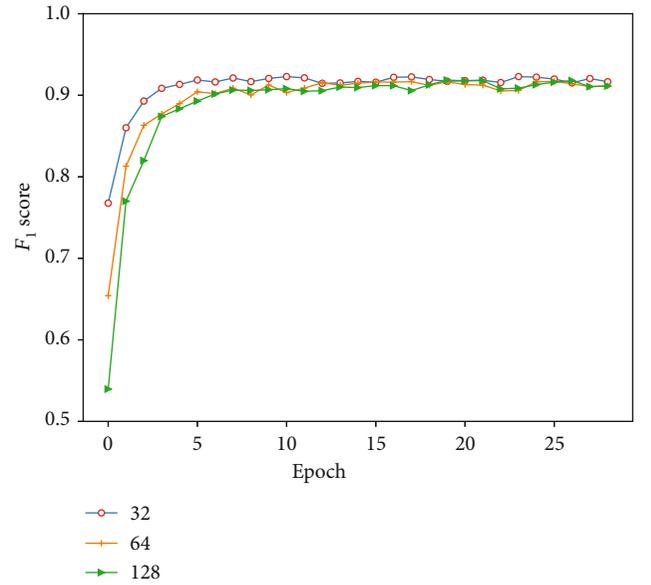
**3.2.2. Character-Word Vector Fusion.** In the data preprocessing stage, the original data is relabeled as the format of *character-the corresponding word-the corresponding label*, and the training of the character vector and the word vector is carried out to generate dictionaries of the character-character vector and the word-word vector.

The training method of the Chinese named entity recognition model based on character-word vector fusion is basically the same as that based on character vector and word vector in general. The only difference is that the input vector of the training model is the vector which combines the character vector and the word vector. In essence, the training in this paper is carried out based on the character vector, but the character vector fuses the characteristics of the word which it is in. In this way, the influence of the word segmentation result need not to be considered. In addition, the character vector integrates the relevant features of the corresponding word, which provides a guarantee for solving the problem that it is difficult for a single Chinese character to embody the semantic information of the word it is in.

### 3.3. Experimental Results and Analysis

**3.3.1. The Evaluation Indexes.** In order to verify the proposed model, the precision, recall, and  $F_1$  score are selected as evaluation indexes.

Firstly, the relationship between the predicted value and the actual value is shown in Table 1, where  $T$  represents the

FIGURE 4: Batch\_size contrast. The  $F_1$  value varies with the value of the batch\_size.

correct prediction,  $F$  represents the wrong prediction,  $P$  represents the positive example, and  $N$  represents the negative example. Then, TP is positive examples correctly predicted, TN is negative examples correctly predicted, FP is examples wrongly predicted as positive, and FN is examples wrongly predicted as negative.

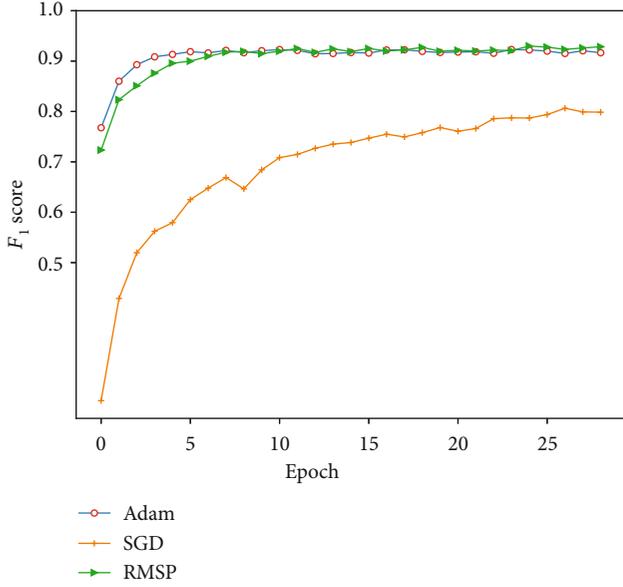
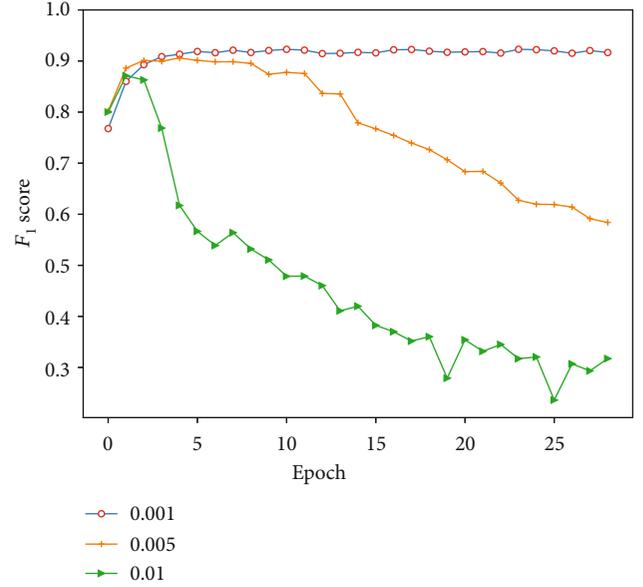
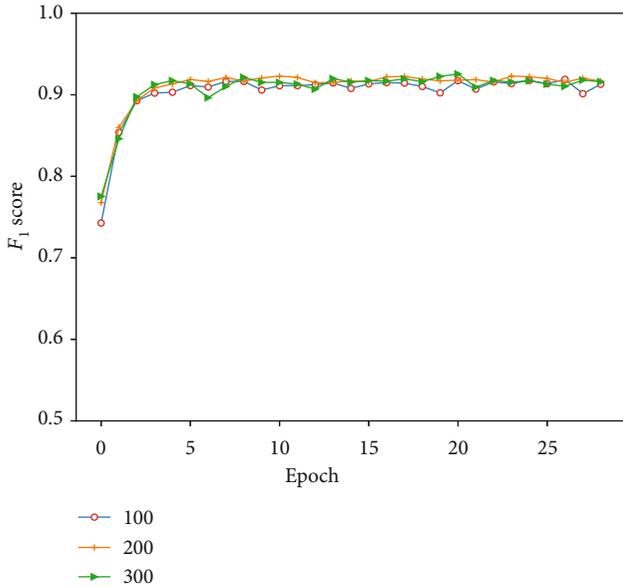
The precision is the ratio between the positive examples correctly predicted and all the examples predicted as positive, and its calculation formula is shown as

$$\text{precision} = \frac{TP}{TP + FP} \times 100\%. \quad (2)$$

The recall is the ratio between the positive examples correctly predicted and all the actual positive examples, and the calculation formula is shown as

$$\text{recall} = \frac{TP}{TP + FN} \times 100\%. \quad (3)$$

The relationship between precision and recall is inversely proportional so that they cannot reflect the overall situation.

FIGURE 5: Optimizer. The  $F_1$  value varies with the type of optimizer.FIGURE 7: Learning rate. The  $F_1$  value varies with the learning rate.FIGURE 6: Number of hidden layer nodes. The  $F_1$  value varies with the number of hidden layer nodes.

Therefore, the  $F_1$  value is comprehensively calculated on the basis of the precision and the recall, and its calculation formula is shown as

$$F_1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \times 100\%. \quad (4)$$

### 3.3.2. Results and Analysis

(1) *Experiment 1: Model Parameter Tuning Experiment.* According to the design of experiment 1, parameters that affect the performance of the model are tuned in this paper. Figures 4–7, respectively, show the variation trend of the  $F_1$

TABLE 2: Named entity recognition experiment results.

Type	Precision	Recall	$F_1$
Name	93.75%	93.35%	93.55%
Place	93.22%	92.07%	92.64%
Institution	91.02%	85.34%	88.09%
All label	92.73%	91.77%	92.25%

TABLE 3: Comparison experiment with a character vector model.

Model type	Vector model	$F_1$
BiLSTM-CRF	Character vector	91.54%
BiLSTM-CRF	Character-word vector fusion	92.25%

value when the batch\_size value, optimizer type, number of the hidden layer node, and learning rate are tuned. During the experiment, the parameters of the previous round are selected for the tuning of the parameters of this round. According to the changing trend of the  $F_1$  value, the batch\_size value, optimizer, number of the hidden layer node, and learning rate are finally determined as 32, Adam, 200, and 0.001, respectively.

(2) *Experiment 2: Compared with the Model Which is Based on Word Vector Training.* By setting the above parameters, the final precision, recall, and  $F_1$  value of the model are 92.73%, 91.77%, and 92.25%, respectively. Table 2 shows the detailed results for different types of label.

The conclusion that the model based on the character vector is better than that based on the word vector has been drawn from the literatures [13, 14]. In order to verify the validity of the character-word vector fusion, we compared the proposed model with the BiLSTM-CRF model based on the character vector, which also trained on the People’s Daily

corpus, and on the premise of using the same parameters; the  $F_1$  value of the proposed model is 92.25%, which is 0.71% higher than that of the compared model. The experimental results are shown in Table 3.

#### 4. Conclusions

In view of the problem that the traditional word vector representation cannot represent semantic features sufficiently, the CWVF-BiLSTM-CRF model based on character-word vector fusion is proposed to carry out the Chinese named entity recognition in this paper. The model takes the character vector integrated with the word vector information as the input unit to supplement the semantic features and obtains the label sequence according to the context through the BiLSTM layer. Finally, the output of BiLSTM is taken as the input of the CRF layer, which makes good use of the predictive label information of the output of the BiLSTM layer. The experimental results proved that compared with the character vector-based BiLSTM-CRF method, the proposed character-word vector fusion is effective for the Chinese named entity recognition task and the value of  $F_1$  reaches 92.25%. In future work, the accuracy of the word segmentation can be further considered and improved in order to add more accurate information into the character-word vector.

#### Data Availability

The data of the People's Daily in 1998 used to support the findings of this study has been deposited in the Key Laboratory of Computational Linguistics (Peking University) Ministry of Education, China, and the URL is <https://klcl.pku.edu.cn/gxzy/231686.htm>.

#### Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

#### Acknowledgments

This work is supported in part by the Key Science and Technology Program of Xi'an (Grant No. Z20180253) and the Xi'an Science and Technology Innovation Leading Project (Grant No. Z20180260).

#### References

- [1] S. D. Chen and X. Y. Ouyang, "Overview of named entity recognition technology," *Radio Communication Technology*, vol. 46, no. 3, pp. 251–260, 2020.
- [2] S. C. Ding, Z. Fang, and N. Wang, "Business domain named entity identification based on BiLSTM-CRF," *Modern Intelligence*, vol. 40, no. 3, pp. 103–110, 2020.
- [3] N. Li, H. M. Guan, P. Yang, and W. Y. Dong, "Chinese named entity recognition method based on BERT-IDCNN-CRF," *Journal of Shan Dong University (Science Edition)*, vol. 55, no. 1, pp. 102–109, 2020.
- [4] L. Liu and D. B. Wang, "A review on named entity recognition," *Journal of the China Society for Scientific and Technical Information*, vol. 37, no. 3, pp. 329–340, 2018.
- [5] X. X. Chen, L. Xu, Z. Y. Liu, M. S. Sun, and H. B. Luan, "Joint learning of character and word embeddings," in *Proceedings of the International Conference on Artificial Intelligence*, pp. 1236–1242, Austin Texas, USA, January 2015.
- [6] X. Ma and E. Hovy, "End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1064–1074, Berlin, Germany, August 2016.
- [7] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, and C. Dyer, "Neural architectures for named entity recognition," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 260–270, San Diego, California, June 2016.
- [8] Y. Zhang and J. Yang, "Chinese NER using lattice LSTM," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1554–1564, Melbourne, Australia, July 2018.
- [9] S. Sachin, A. Tripathi, N. Mahajan, S. Aggarwal, and P. Nagrath, "Sentiment analysis using gated recurrent neural networks," *SN Computer Science*, vol. 1, no. 2, article 74, 2020.
- [10] K. D. Onal, Y. Zhang, I. S. Altinogvde et al., "Neural information retrieval: at the end of the early years," *Information Retrieval Journal*, vol. 21, no. 2-3, pp. 111–182, 2018.
- [11] L. Liu and D. B. Wang, "Survey of the named entity recognition," *Journal of the China Society for Scientific and Technical Information*, vol. 3, pp. 329–340, 2018.
- [12] X. H. Wu, L. Chen, T. T. Wei, and T. T. Fan, "Sentiment analysis of Chinese short text based on self-attention and BiLSTM," *Journal of Chinese Information Processing*, vol. 33, no. 6, pp. 100–107, 2019.
- [13] Z. Liu, C. Zhu, and T. Zhao, "Chinese named entity recognition with a sequence labeling approach: based on characters, or based on words?," in *Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence*, pp. 634–640, Springer, 2010.
- [14] Y. Liu, Y. Zhang, W. Che, T. Liu, and F. Wu, "Domain adaptation for CRF-based Chinese word segmentation using free annotations," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 864–874, Doha, Qatar, October 2014.