Chinese Named Entity Recognition Based on Dynamically Adjusting Feature Weights

Qing Lv, Limin Zheng(✉), and Miao Wang

College of Information and Electrical Engineering, China Agriculture University, Beijing, China
{S20193081363,zhenglimin,S20193081360}@cau.edu.cn

Abstract. Named entity recognition is a basic task in NLP, and it is an important basic tool for many NLP tasks such as information extraction, parsing, question answering system and machine translation. The extraction of sequence features of datasets directly affects the recognition effect of named entities, and only the accumulation of local sequence features cannot capture the long distance dependencies. The extraction of global sequence features improves this problem, but loses some local features. Long entities are nested within short entities and have different entity attributes from short entities, resulting in identification errors. To solve these problems, a Chinese named entity recognition algorithm based on Bert +FL-LGWF+CRF is proposed. In this method, the text is encoded into a word vector matrix by Bert as the input to FL-LGWF (Entity Level-Local And Global Weighted Fusion). FL-LGWF utilizes CNN (Convolutional Neural) to extract the local sequence features of the text vector, and use BISTM (Bidirectional Long Short-Term Memory) to extract contextual global sequence features, and perform dynamic weight fusion on the extracted sequence features. Then the score matrix of the tag is obtained according to the entity attribute level. Finally, the global optimal tag sequence is obtained through the CRF layer. Experimental results show that the proposed Bert +FL-LGWF+CRF model has higher F1 value on both public data sets and self-created data sets.

Keywords: Named entity recognition · Dynamic weight fusion · Entity level local · CNN · BILSTM

Supported by the National Key Research and Development Program of China (2017YFC1601803) and Beijing innovation team project of modern agricultural industrial technology system(BAIC02-2020).

H. Gao and X. Wang (Eds.): CollaborateCom 2021, LNICST 406, pp. 3–19, 2021.
https://doi.org/10.1007/978-3-030-92635-9_1

1 Introduction

Named entities are the entities with specific meaning or strong reference in the text, usually including the name of the person, place name, organization name and so on. Named Entity Recognition (NER) refers to the extraction of Named entities from unstructured input text [1]. Named entity recognition has been widely applied in various fields, such as knowledge graph construction [2], knowledge base construction [3,4], network search [5], machine translation [6], automatic question and answer, etc. [7].

Initially, named entity recognition was a rule-based and dictionary-based approach, which adopted specific rule templates or special dictionaries built by linguists manually according to the characteristics of data sets, and used matching methods to process the text to realize named entity recognition. On this basis, Rau et al. [8] proposed for the first time to combine manual writing rules with heuristic ideas and realize automatic extraction of named entities of company name type from text. However, it consumes a lot of manpower and is not easy to expand in other data sets or types, so there are very obvious limitations. In the machine learning approach, named entity recognition is treated as a sequential labeling problem. Yoshua [9] et al. proposed the Hidden Markov Model (HMM) to directly Model transition probability and performance probability, and to calculate co-occurrence probability, so as to carry out sequence labeling. Sutton et al. [10] proposed a named entity recognition method based on Conditional Random Field (CRF), and manually annotated the feature template of the entity, and annotated the sequence. CRF can use internal and contextual characteristic information to label a location. The above methods not only consume time and effort, but also have a certain error in manual annotation because feature selection relies on the prior experience of human, which leads to a low accuracy. In recent years, Deep Neural Network has been widely concerned in the field of natural language processing, and named entity recognition based on Deep Neural Network (DNN) [11] has become a research hotspot. This technology mainly utilizes the strong computing power of neural network to automatically extract the context features at the sentence level and realize entity recognition. Collobert et al. [12] proposed the named entity recognition method based on neural network for the first time. In this method, each word has a fixed size window, but the effective information between long words is ignored. Lample [13] proposed a named entity recognition method based on Long Short-Term Memory (LSTM), which effectively solved the problem of Long time dependence effect. However, these methods do not take into account the sequence of words in sentences. Huang [14] proposed a Bidirectional Long Short-Term Memory (BILSTM) that can better capture Bidirectional semantic dependence by combining forward and backward LSTM. However, because the training is fixed word embedding, there is a problem that the polysemy cannot be expressed. Devlin et al. [15] proposed the Bidirectional Encoder Representation from Transformers (Bert), which represents the context through pre-training and fine-tunes to solve the problem of inadequate context representation in NER. In order to make full use of the advantages of each model, scholars put forward a series of combination models. Dong [16] proposed a combination model of LSTM+CRF to extract sequence features by LSTM, and input the extracted sequence features into the CRF layer to add constraints to the predictive labels. Li [17] proposed Bert +BILSTM+CRF model, in which Bert was used for text encoding, and BILSTM was used for sequence feature extraction. CRF layer added some constraints for the final predicted tags to ensure that the predicted tags were legitimate. Bert +BILSTM+CRF model has excellent performance in accuracy and recall rate, and is one of the mainstream models for named entity recognition at present. However, BILSTM can not take into account local features when extracting global features of sequences. For the existence of entity nesting problem, it will also produce errors in the identification of named entities.

In summary, existing models cannot extract sequence features locally or globally according to sequence features, and the problem of entity nesting in named entity recognition task is not taken into account. For the above problems. In this paper, a hybrid sequence feature extraction and fusion entity attribute priority classification method are proposed to judge entity tags. The contributions of this paper are as follows:

(1) A weighted fusion method of CNN and BILSTM is proposed. CNN is used to extract the local features of the sequences. BILSTM extracts global sequence features related to text context. The contribution of local features to each word in the sequence was extracted. The contribution of global features to each word in the sequence was extracted. Multiply local and global sequence features with corresponding extracted contribution degrees to get the weights of local and global features to the words in the sequence, and add the weights.

(2) An EL is proposed based on the entity nesting problem, and the priority level of the attribute containing the nested entity is set to be higher than that of the nested entity attribute.

(3) The named entity recognition based on Bert+EL-LGWF+CRF is proposed. Bert is used to encode the text as a sequence composed of word vector matrix; CNN and BILSTM are used to extract sequence features. The generated sequence features fuse the attribute level value of the entity; The CRF layer calculates the optimal solution of the whole sequence for global optimization, and obtains the optimal prediction result of the whole sequence. The BERT pre-training model directly calls the model trained by Y [17].

2 Model Approach

2.1 BERT

Bert (Bidirectional Encoder Representation from Transformers) addresses the polysemy problem by dynamically encoding each word in text into a low-dimensional representation of a real-valued vector. This enhances the semantic representation of words. The network architecture of BERT is shown in Fig. 1. Bert is composed of multiple layers of Transformer (TRM in Fig. 1) [18]. The input encoding vector of Bert is the unit sum of three embedded features, and the output is the trained word encoding matrix $T_1, T_2..., T_n$ Fig. 1). The Transform coding unit is the most important part of BERT, and its core is to model a piece of text based on self-attention mechanism, as shown in Fig. 2. Input X1 and X2 of the Transform coding unit, add relative position coding information (Positinal Encoding in Fig. 2), and code through self-attention layer (self-attention in Fig. 2) and feedforward neural network (FNN in Fig. 2). Residual network and layer normalization (Add & Normal in Fig. 2) are added to solve the deep learning degradation problem Fig. 2.

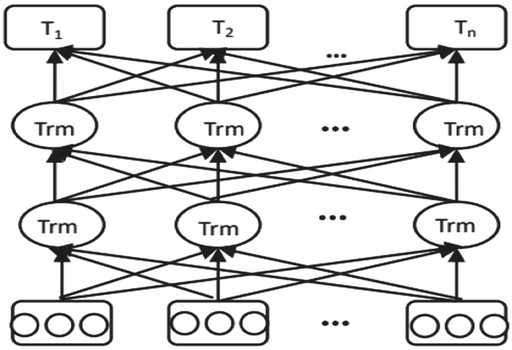

Fig. 1. The network architecture of BERT.

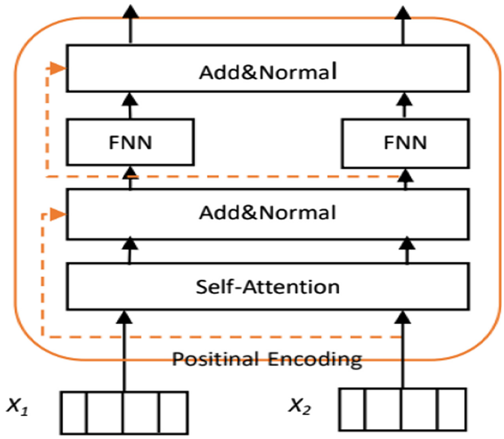

Fig. 2. Transform encoding unit.

2.2 CNN

CNN [19] is a neural network capable of representation learning and parallel computing. Convolutional neural networks can process multi-dimensional data, as shown in Fig. 3. The convolution layer is composed of convolution kernel (Conv) and activation function (ReLU). The convolution layer performs feature extraction on input data. After feature extraction in the convolutional layer, the output features will be transferred to the pooling layer for feature selection and information filtering. Full-connection layer (FC) is located in the last part of Convolutional Neural Network hidden layer. Full-connection layer is the nonlinear combination of extracted features to get the output Fig. 3).

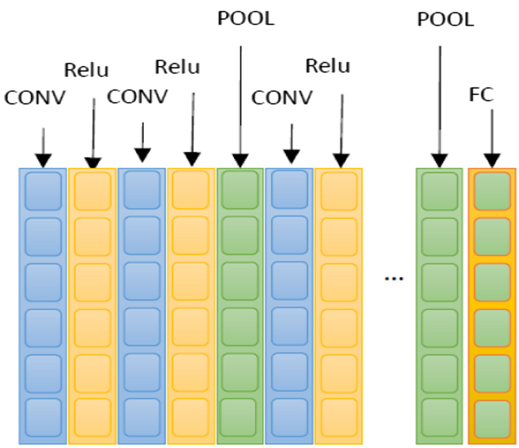

Fig. 3. Convolutional neural network diagram.

2.3 BILSTM

BILSTM is a bidirectional cyclic neural network, which can process long sequences of sentences and extract the global features of sentences [20]. The structure of BILSTM is shown in Fig. 4. BILSTM consists of a forward LSTM and a backward LSTM, and combines a set of sequences $X_1, X_2, ..., X_n$ as input, returns a set of sequences $H_1, H_2, ..., H_n$, these sequences contain information about each step of the input sequence; BILSTM can deal with the remote dependence problem, and takes into account the information before and after the current moment. BILSTM comprehensively considers all the information of the whole sequence Fig. 4).

2.4 CRF

CRF uses the dependency information between adjacent tags for sentence-level labeling, calculates the optimal solution of the overall sequence by adding the

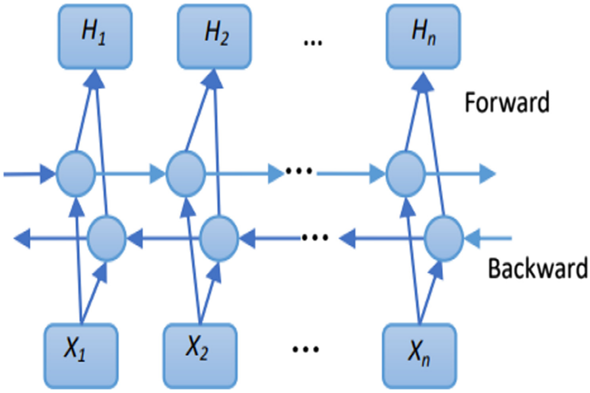

Fig. 4. The structure of BILSTM.

transfer score matrix of tags for global optimization, and obtains the optimal prediction result of the overall sequence [21].

3 BERT+EL-LGWF+CRF

The design idea of the model is to use Bert to encode text to solve the problem of polysemy. By making full use of the ability of CNN to extract local sequence features and the ability of BILSTM to extract global sequence features, the weight value is used as the contribution of extracting sequence features from the same data. In the case of ambiguity in identifying entity attributes, the rank of entity attributes is taken as the condition of priority judgment. CRF was optimized globally to get the final result. A named entity recognition model based on Bert+EL-LGWF+CRF was constructed, and an optimal sequence feature was finally obtained through continuous training and adjustment of the model.

3.1 Weighted Fusion According to CNN and BILSTM

CNN extracts the local features of the sequence, inputs the word vector matrix encoded by BERT into the CNN network, and uses the convolutional neural network to generate the local features of fixed size and independent from the input. This is shown in Fig. 5.

Let $(Pad, Pad, x1, x2, ..., xn, Pad, Pad)$ be the input of the CNN network, where $(x1, x2, ..., Xn)$ represents the input sentence, Pad (element 0) represents the data filling part, xi represents the vector coding of the ith word in the sentence, n represents the number of words in the sentence sequence. After passing through the CNN neural network, n outputs are obtained, $y1, y2..., yn$ is shown in formula (1).

$$y_i = \left([f\theta]_{[w]_t,i}\right) \tag{1}$$

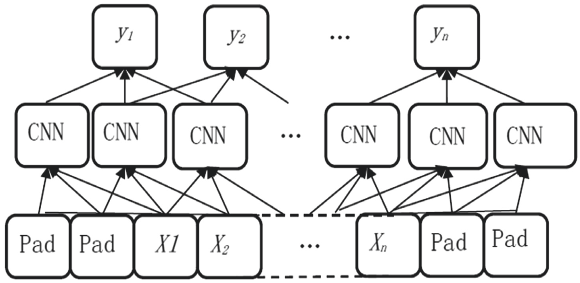

Fig. 5. CNN extracts local features of data.

where, f_θ is the fractional matrix with parameter θ, I is the sequence number of words in the sequence, [w] stands for the tag set, $[[\mathrm{w}]_4 t$ stands for the tag number of 0 t in the tag set, and t is the last tag number in the tag set.

CNN extract local features affect the accuracy of the named entity recognition, such as "poison eggs" entity attributes as involved food name, CNN first to identify the word "poison", and then identify the word "chicken", finally identify the word "egg", before and after such dependencies will be missing, causing the attribute recognition as independent entities, "poison" Identify the "egg" attribute as the name of the food in question.

BILSTM extracts the global characteristics of the data. The character coding sequence is input into the BILSTM layer, and the global feature extraction is performed to obtain the representation fraction matrix of the sequence. This is shown in Fig. 6. The model consists of input layer W, hidden layer H and output layer Y. W0, W1,... , Wn represents the features of input matrix, the hidden layer H is composed of the combination of forward LSTM and backward LSTM, and the output layer Y obtains the score matrix of each label in the word vector Fig. 6). Although BILSTM can ensure the long distance dependency, the accuracy of BILSTM is lower than that of CNN in the identification of independent entity attributes.

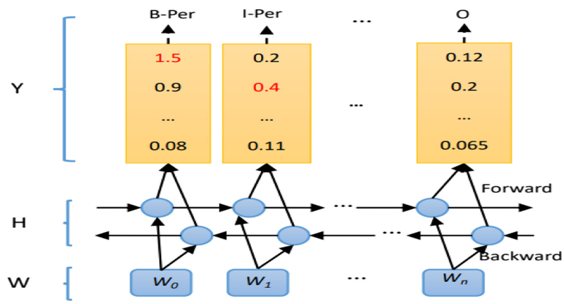

Fig. 6. BILSTM extracts global characteristics of the data.

Only the accumulation of local sequence features can not better capture the long-distance dependence. The extraction of global sequence features improves this problem, but some local features are lost. In order to distinguish different features, hybrid features are extracted from the perspective of weighted fusion and weight vector is introduced, The weight connected to the sequence feature is regarded as the contribution to the final sequence feature. Feature weighted fusion obtains mixed features by multiplying vectors with their own weights and adding corresponding items. The weights of global features and local features on the sequence tags belonging to the same word vector are extracted respectively for weighted fusion, as shown in Fig. 7.

CNN and BILSTM output each tag in the word vector score matrix (y1, y2... , yn), denoted by y, and the representation fraction matrix of the sequence output by the CNN model is denoted by y-CNN; The representation fraction matrix of the sequence output by the BILSTM model is denoted as Y-BILSTM; Y-CNN and Y-BILSTM pass through the network of full connection layer, and the calculation process of full connection layer is shown in the following formula.

$$F_{CNN} = g\left(W_{CNN} y_{CNN} + b\right) \tag{2}$$

$$F_{BILSTM} = g\left(W_{BILSTM} y_{BILSTM} + b\right) \tag{3}$$

where, g(·) is the activation function, b is the offset quantity, W-CNN and W-BILSTM are the weights of the connection in extracting sequence features. The weight changes dynamically with the training.

$$K_{CNN} = Dense_{unit=1}\left(y_{CNN}\right) \tag{4}$$

$$K_{BILSTM} = Dense_{unit=1}\left(y_{BILSTM}\right) \tag{5}$$

Y obtains the weight vector matrix (K1, K2, ... , Kn), the weight vector matrix is referred to as K. In which, $Dense_(unit = 1)$ is a fully connected function whose weight matrix is 1 dimension.

Pass K through Softmax layer, the α weight vector matrix $(\alpha 1, \alpha 2, ..., \alpha n)$, and the value of α i is shown in Formula (6).

$$\alpha_i = e^{K_i} / \sum_{j=1}^{n} e^{K_j} \tag{6}$$

where, $0 < \alpha_i < 1, \quad \sum_i \alpha_i = 1$; The score matrices output by CNN and BILSTM are weighted and fused with their respective weight value vector matrices, as shown in Eq. (7)

$$L = \left(y_{CNN} \otimes \alpha_{CNN}\right) \oplus \left(y_{BILSTM} \otimes \alpha_{\bar{B}ILSTM}\right) \tag{7}$$

where L represents the final output score matrix of the weighted fusion of CNN and BILSTM, $\alpha - CNN$ represents the vector matrix of weight value extracted from Y-CNN, $\alpha_B ILSTM$ represents the vector matrix of weight value extracted from Y-BILSTM, $\otimes$ represents matrix multiplication, and $\oplus$ represents matrix

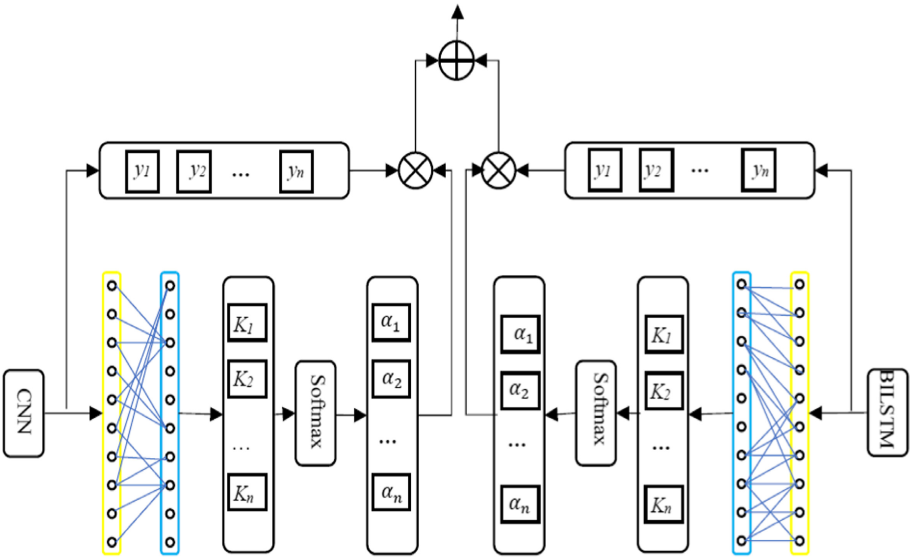

Fig. 7. Schematic diagram of feature extraction for weighted fusion.

addition Fig. 7). The advantage of mixed features is that the local sequence features extracted by CNN and the global sequence features extracted by BILSTM can be taken as the proportion of features extracted by word vectors in the same sequence according to the weight value. By adopting the weighted fusion method, the information contained in mixed features can be made more comprehensive through the training of the network.

3.2 EL

The process of named entity recognition is that the model scores the category of entity attributes to which each word in the input text belongs. The attribute tag with a higher score is the entity category to which the word belongs. To solve the problem of entity nesting, the priority of entity properties is set.

$$EL = EL_1 + EL_2 + \cdots + EL_n \tag{8}$$

$$S = L \oplus EL \tag{9}$$

As shown in Eqs. (8) and (9). Where, ELi represents the priority ranking matrix of the ith entity attribute, and I is the number of entity attributes defined in the named entity recognition task. L is the final output score matrix of the weighted fusion of CNN and BILSTM; EL is the priority grade matrix composed of all entity attribute grade matrices; $\oplus$ represents matrix addition; + represents matrix stitching operation.

In the case that the priority level of entity attribute is not set, the score value of the recognized word under each type of entity attribute tag is obtained through sequence feature extraction, and the entity attribute tag of the word is obtained through CRF global optimization and decoding. If there is a nesting problem between entity attributes, the model will produce an error in extracting the score value of the entity attribute to which the word belongs, as shown in Fig. 9 in the left half of changshu yong hui supermarket entity recognition, changshu yong hui supermarket entity attributes for an organization, the model of the entity recognition, changshu in place under the label of score may be superior to scored under the label of the organization, Causes the entity attribute of Changshu to be identified as a place.

Therefore, prioritize the entity attributes with nested problems. For the entity attribute of Changshu Yonghui Supermarket in the right part of Fig. 8, the priority of the entity attribute as organization should be higher than that of the entity attribute as location. The priority of the entity attribute as organization is defined as 1, and the priority of the entity attribute as location is defined as 0. After the score value of the recognized word under each type of entity attribute tag is obtained through sequence feature extraction, the obtained score value is added to the priority score of the entity attribute to obtain a new score value, so that Changshu's score under the organization tag is better than that under the location tag. After global optimization and decoding by CRF, the entity attribute tag of the word is obtained. Fig. 8).

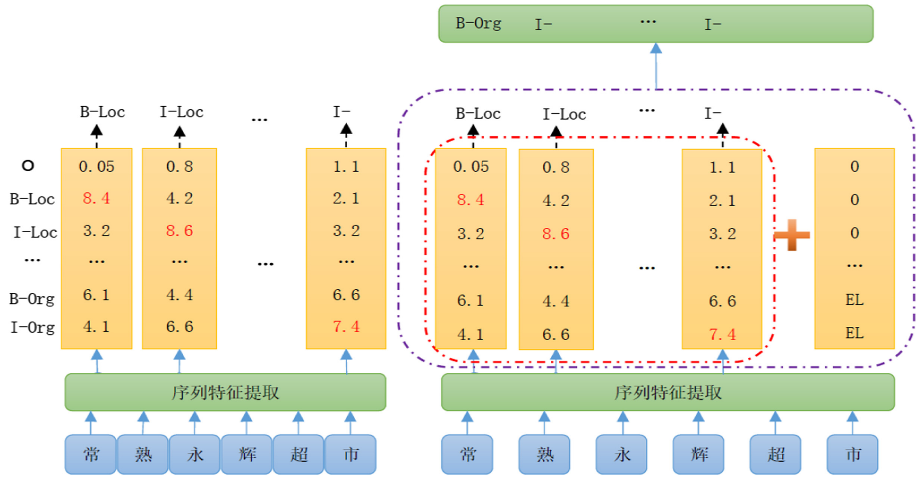

Fig. 8. Fusion El diagram.

Therefore, the proposed Chinese named entity model is shown in Fig. 9.

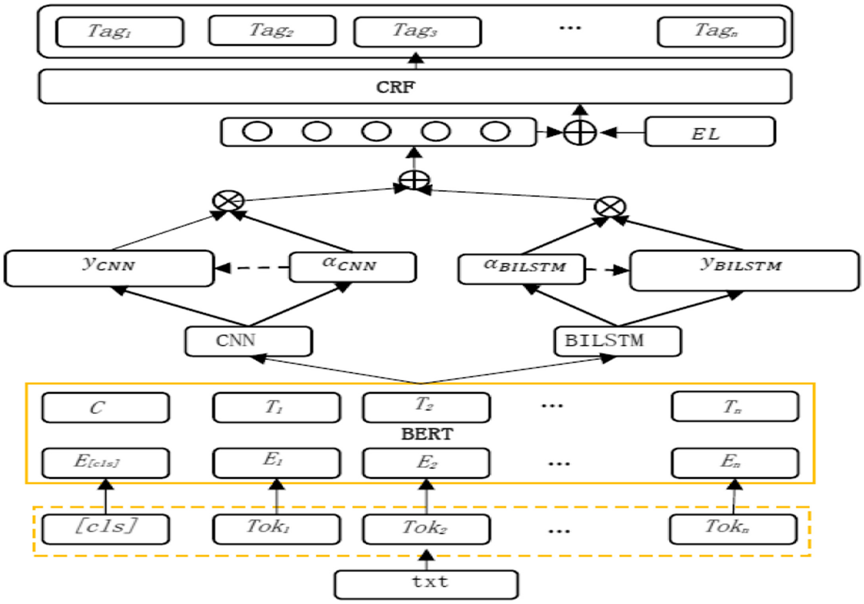

Fig. 9. A Chinese named entity recognition model based on BERT+EL-LGWF+CRF.

As shown in Fig. 9, the text is first partitioned to get TOK1, TOK2, ... ,Tokn; The obtained word segmentation sequence extracts text features through Bert, obtains word granularity vector matrix, and obtains the vector representation of each word in the sentence T1, T2,... Tn; Bert outputs the word vector encoding matrix as CNN and BILSTM inputs. The generated representations of CNN and BILSTM are dynamically weighted and fused, and the score value of the category label obtained are added to the priority score of the entity attribute to obtain the final score value. Finally, the global optimal sentence level tag sequence is obtained by learning the dependency between tags through CRF layer.

3.3 Loss Function

The maximum likelihood method is used for training, and the corresponding loss function is shown as the formula.

$$-\log p(y \mid x) = -score(x, y) + logZ(x) \tag{10}$$

where

$$score(x, y) = \sum_i \alpha p_{B(x_i y_i)} + (1-\alpha) p_{P(x_i y_i)} + EL\,(y_i) + Trans\,(y_{i-1}, y_i) \tag{11}$$

α represents the weight of BILSTM obtained by model training, $p_{B(x_i,y_i)}$ represents the probability score of i-th position labeled y_i in the case of feature extraction by BILSTM$(1-\alpha)$ represents the weight of CNN obtained by model training, $p_P(x_i, y_i)$ represents the probability score of i-th position labeled y_i in the case of feature extraction by CNN.$EL(y_i)$ represents the characteristic weight of the label y_i. Trans$(y_(i-1), y_i)$ is given by TransitonScore [16].

where

$$
\begin{aligned}
&\log Z(x) = \log \sum_y \exp(score(x,y)) \\
&= \log \sum_{y_1} \sum_{y_2} \ldots \sum_{y_n} \exp\left(p\left(y_1\right)\right. \\
&+ EL\left(y_1\right) + T\left(\text{Start}, y_1\right) + p\left(y_2\right) \\
&+ EL\left(y_2\right) + T\left(y_1, y_2\right) + \cdots \\
&\left. + p\left(y_n\right) + EL\left(y_n\right) + T\left(y_n, End\right)\right)
\end{aligned} \tag{12}
$$

p represents EmissionScore, n represents the length of the input sentence. T represents TransitonScore?

4 Experiment and Analysis

4.1 Dataset

People's Daily, MSRA, Golden-Horse data sets are adopted to self-labeled food additive related food safety event news report data set, referred to as food data set, as the experimental data set. The size of the data set is shown in Table 1. The validity of Chinese named entity recognition is verified for the proposed model.

In this paper, 10,680 news reports about food safety incidents related to food additives in the past five years are retrieved from the Food Partner website, and the entity attributes are screened out. These entity attributes are divided into 10 categories, and the corpus is marked with Yedda. The data set of news reports on food safety events related to food additives, referred to as food data set, is obtained. The entity attributes and labels of the food data set as well as the number of details are shown in Table 2.

The entity attribute was marked with BIO-style labeling strategy. For example, the name of the Food concerned is labeled as "FOOD", and a nine-season shrimp (I-food) of the Food concerned appeared in the data set, that is, the shrimp (I-food) is labeled as "B-food" and the additive is seaweed gelatin. The words that do not belong to any one of the 10 categories are marked as O. Examples of BIO labeling are shown in Table 3.

4.2 Evaluation Indices

Set of evaluation index for P (accuracy), R (a) recall rate, F1 value of P to correctly identify the number of entities of identify the percentage of the number

Table 1. Experimental dataset.

Dataset type	Data	Dataset size
Raw dataset	MSRA	42452
Raw dataset	People's Daily	20864
Raw dataset	Golden-horse	11731
Labeled datasets	Food datasets	625518 words

Table 2. Entity attributes and labels of food datasets.

Entity attribute	Label	Number of entities
Time	Time	4260
Location	Loc	5292
Person	Per	1164
Name of the food involved	Food	9042
Production enterprise	Org	5244
Additive	Add	4458
Trace elements, heavy metals	Elements	1044
Hazards, first aid measures	Influences	1326
Involved in the file	Document	648
Reason	Reason	2016

Table 3. An example BIO tag.

九	节	虾	注	有	海	藻	明	胶	。
B-Food	I-Food	O	O	O	B-Add	I-Add	I-Add	I-Add	O

of entities, R for the correct identification number of named entities standards results in the percentage of the number of entities, the weighted harmonic mean formula one is P and R is the influence of balance P and R composite indicator, The definitions are shown in Eqs. (10), (11) and (12).

$$P = \frac{TP + TN}{TP + FN + FP + TN} \times 100\% \tag{13}$$

$$R = \frac{TP}{TP + FN} \times 100\% \tag{14}$$

$$F1 = \frac{2 \times P \times R}{P + R} \times 100\% \tag{15}$$

Where, TP – Predicts positive cases to be true Fn – Predicts the positive example to be false FP –Predicts the counterexample to be true TN – Predicts the counterexample to be false.

4.3 Experimental Results and Analysis

Bert+BILSTM+CRF and Bert+CNN+CRF were selected as comparison models. Where: (1) BILSTM extracts global features of sequences, but some local features are lost due to the consideration of long distance dependencies; (2) CNN extracts the local features of the sequence and splices them together for sequence tag recognition. However, only the accumulation of local sequence features cannot capture the long-distance dependencies well. (3) This model gives consideration to both global and local features, and the theoretical effect is better than that of the comparative model.

All model parameters remain consistent. Experiments are carried out on the open data set and the food data set respectively, and the obtained experimental results are shown in Table 4. By analyzing Table 4, it can be seen that the proposed weighted fusion model obtained optimal results in P, R and F1 of all data sets on the common data sets. Compared with Bert + Bilstm +CRF, Bert +CNN+CRF and other single mode structures, the weighted fusion model significantly improves P, R and F1 of the three data sets. Compared with the weighted fusion model, the weighted fusion model has significantly improved P, R and F1 on the three data sets, indicating that the mixed feature extraction model has better performance in performance. In MSRA data set, the weighted fusion model improves the F1 value by 0.88% compared with Bert+CNN+CRF, which is the minimum value in the table. In the People's Daily data set, the weighted fusion model is 1.78% higher than that of Bert+CNN+CRF in the F1 value, which is the maximum value in the table. In the food dataset, the F1 value of the weighted fusion model reaches 82.69, which is 4.46% higher than that of the single-mode extraction method Bert+CNN+CRF with the lowest F1 value. This is because the entity attribute of the food data set is more non-standard than that of the public data set, and the entity attribute has different lengths and obvious gaps. By comparison, it is shown that the weighted fusion of local features extracted by CNN and global features extracted by BILSTM constitutes a mixed feature extraction model, which has a great improvement in the performance of the named entity recognition task. It is fully proved that the mixed feature extraction model can improve the accuracy of named entity recognition compared with the single-mode.

Table 4. Experimental results of each method on datasets.

	MSRA			People's Daily			Golden-horse			Dataset about food		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Weighted fusion model	96.77	96.92	96.84	95.85	95.74	95.79	92.98	92.23	93.12	83.19	82.19	82.69
BERT+BILSTM+CRF	95.96	95.96	95.96	95.04	93.99	94.51	92.12	92.52	92.32	82.39	80.41	81.39
BERT+CNN+CRF	96.50	96.53	96.52	94.03	93.99	94.01	91.86	92.89	92.37	78.68	77.78	78.23

There is no problem of entity nesting in the public data set, so the food data set labeled by oneself is used to determine the effectiveness of entity attribute level. To reduce named entity identification errors due to entity nesting problems, a priority level is defined for each entity attribute. 10% of the attributes of each entity are randomly selected for observation, in which there is no nested relationship between entity attributes such as a person, time, place, name of related food, additive, trace element, and heavy metal, so the priority level of these entity attributes is classified as 0. In the production enterprises with physical attributes, there are entities with nested locations with physical attributes, such as "Zhangjiagang Oshang Supermarket", "Suzhou Food and Drug Administration", etc., so the priority level of the production enterprises with physical attributes is classified as EL. Entities with nested entity attribute heavy metals and additives, such as "excessive lead", "excessive Fempni", etc., are caused by entity attribute events. Therefore, the priority level of entity attribute events is classified as EL. For the harm caused by physical attributes, the first aid measures have entities nested with trace elements of physical attributes, such as "affecting the absorption of calcium in the human body", etc. Therefore, the priority level of the harm caused by physical attributes is classified as EL. The entity attribute refers to the entity whose documents have nested the entity attribute manufacturer, place, and time, such as "Regulations on the Protection of Consumer Rights and Interests of Liaoning Province", "Notification of Food Safety Sampling Inspection of Shenzhen Market Supervision Administration in 2020", etc. Since the priority level of the entity attribute manufacturer is EL, the priority level of the entity attributes involved in the file is 2EL. The priority levels of each entity's attributes are shown in Table 5.

Table 5. Attribute priority of each entity

Entity attribute	Priority
Time	0
Location	0
Person	0
Name of the food involved	0
Production enterprise	EL
Additive	0
Trace elements, heavy metals	0
Hazards, first aid measures	EL
Involved in the file	2EL
Reason	EL

On the basis of the experiment of weighted fusion model (using CNN) in Table 4, EL is defined as different values respectively, and the experimental data obtained are shown in Table 6. From the analysis of Table 6, it can be seen that the performance of the model is constantly changing with different EL. When EL is between 0 and 0.05, F1 value fluctuates little, the difference is less than

Table 6. Experimental results of different El values.

EL	P	R	F1
0.025	83.48	82.19	82.83
0.05	83.90	81.79	82.83
0.1	85.29	83.11	84.18
0.2	84.34	83.20	83.77
0.4	85.82	84.27	85.05
0.8	86.14	84.54	85.27
1.3	85.95	83.20	84.55
1.6	84.86	82.82	83.83

1%, indicating that the level weight assigned to entity attributes is too small to affect the effect of named entity recognition. When EL is 0.8, F1 value reaches a maximum of 85.27 which is 2.58% higher than that when EL is 0. When EL is 1.3, F1 value is 84.55, which is lower than the peak value of 84.55. When EL is 1.6, F1 value drops to 83.83. The experimental results show that defining EL values for entity attributes can improve the performance of entity recognition in named entity recognition tasks with nested entities.

5 Summary

This paper proposes a sequential labeling method based on Bert+EL-LGWF-CRF model. This method mainly solves two problems in Chinese named entity recognition. (1) The recognition accuracy of irregular entity attributes by a single model is low. (2) Long entity attributes nested short entity attributes of different categories.

The method is evaluated on the Chinese named entity recognition task. The Experimental results show that the hybrid feature extraction model with weighted fusion of CNN and BILSTM is superior to the model with single CNN or BILSTM feature extraction in accuracy, recall value and F1. Especially in the small sample data set, the convergence speed is fast, and it is easier to learn the characteristics of the data set. On the basis of mixed feature extraction, the priority level of entity attributes is increased according to the nested features of the identified entity attributes, which can effectively solve the nested problem between Chinese entities and has excellent performance.

Use a large number of data sets to train the model. The trained model can extract entities from related text without any tags, which not only saves a lot of manpower and time, but also the accuracy and speed of manpower when the amount of text is too large. Incomparable, it is of great significance for the statistics and analysis of information needed in large-scale texts.

References

1. Sun, P., Yang, X., Zhao, X., et al.: An overview of named entity recognition. In: 2018 International Conference on Asian Language Processing (IALP). IEEE (2019)
2. Xie, R., Liu, Z., Jia, J., et al.: Representation learning of knowledge graphs with entity descriptions. In: Thirtieth AAAI Conference on Artificial Intelligence (2016)
3. Riedel, S., Yao, L., McCallum, A., et al.: Relation extraction with matrix factorization and universal schemas. In: Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies 2013, pp. 74–84 (2013)
4. Shen, W., Wang, J., Luo, P., et al.: Linden: linking named entities with knowledge base via semantic knowledge. In: Proceedings of the 21st International Conference on World Wide Web, pp. 449–458 (2012)
5. Zhu, J., Uren, V., Motta, E.: ESpotter: adaptive named entity recognition for web browsing. In: Althoff, K.-D., Dengel, A., Bergmann, R., Nick, M., Roth-Berghofer, T. (eds.) WM 2005. LNCS (LNAI), vol. 3782, pp. 518–529. Springer, Heidelberg (2005). https://doi.org/10.1007/11590019_59
6. Babych, B., Hartley, A.: Improving machine translation quality with automatic named entity recognition. In: Proceedings of the 7th International EAMT Workshop on MT and Other Language Technology Tools, Improving MT Through Other Language Technology Tools, Resource and Tools for Building MT at EACL 2003 (2003)
7. Bordes, A., Usunier, N., Chopra, S., et al.: Large-scale simple question answering with memory networks. arXiv preprint arXiv:15060.02075 (2015)
8. Rau, L.F.: Extracting company names from text. In: IEEE Conference on Artificial Intelligence Application. IEEE (1991)
9. Bengio, Y., Frasconi, P.: An input output HMM architecture. Adv. Neural Inf. Process. Syst. **7**(4), 427–434 (1995)
10. Sutton, C.: Dynamic conditional random fields: factorized probabilistic models for labeling and segmenting sequence data. In: Proceedings of the 21st International Conference on Machine Learning 2004 (2007)
11. Bishop, C.M.: Neural Networks or Pattern Recognition (2005)
12. Collobert, R.: Natural language processing from scratch (2011)
13. Lample, G., Ballesteros, M., Subramanian, S., et al.: Neural architectures for named entity recognition (2016)
14. Huang, Z., Xu, W., Yu, K.: Bidirectional LSTM-CRF Models for Sequence Tagging (2015)
15. Devlin, J., Chang, M.W., Lee, K., et al.: BERT: pre-training of deep bidirectional transformers for language understanding (2018)
16. Dong, C., Zhang, J., Zong, C., et al.: Character-based LSTM-CRF with radical-level features for Chinese named entity recognition (2016)
17. Cui, Y., Che, W., Liu, T., et al.: Pre-training with whole word masking for Chinese BERT (2019)
18. Vaswani, A., Shazeer, N., Parmar, N., et al.: Attention is all you need. arXiv (2017)
19. Gu, J., Wang, Z., Kuen, J., et al.: Recent advances in convolutional neural networks. Pattern Recogn. (2015)
20. Ye, J., Zou, B., Hong, Y., Shen, L., Zhu, Q., Zhou, Q.: Negation and speculation scope detection in Chinese. J. Comput. Res. Dev. **56**(7), 1506–1516 (2019)
21. Pinto, D., McCallum, A., Wei, X., et al.: Table extraction using conditional random fields. In: 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 235–242 (2003)