Enhancing named entity recognition from military news with bert

Hao Qin 1, Yongli Wang1*

1School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China)
*yongliwang@njust.edu.cn

Abstract. Information extraction from news pages is often used in various fields. Named entity recognition is a key step in information extraction. News data is difficult to process because of its unstructured characteristics. So the author proposes a method to perceive the relationship between words and sentences. The method is integrated with BERT (Bidirectional Encoder Representations from Transformers)-BiLSTM (Bi-directional LongShortTerm Memory)-CRF (Conditional Random Field), referred to as the Bertbc model. The model was tested on the People's Daily dataset and text data from military news. The results show that this method improves the recognition accuracy, recall rate, f value and recognition effect.

1. Introduction

With the rapid development of network information, it is very meaningful to extract key information from news web pages. Information extraction can transform unstructured text information into structured text information. Entity extraction is the key step of information extraction. By extracting key entity words from military news, you can quickly grasp important news information, screen important military news, and monitor military public opinion.

Named entity recognition from military news is to extract important entities in the field such as people name, geographical location, organization name and so on. With the development of computing power, algorithm and corpus [1], deep learning will become the mainstream method of named entity recognition. Hammerton et al[2] used one-way LSTM for named entity recognition; Haoyuan Shan et al[3] used CRF model to learn text features and identified military named entities; Santos et al[4] proposed the use of character CNN to enhance the CNN-CRF model; Strubell et al[5] proposed the use of a hollow convolutional network (IDCNN-CRF) for named entity recognition, extracting sequence information and speeding up the training: Huang et al[6] used the BiLSTM-CRF model to deal with NLP sequence labeling problems; Yanbin Zheng et al[7] designed a two-way GRU-CRF model and selected appropriate optimization algorithms and hyperparameters for China. Xuefeng Wang et al[8] used the word2vec-bilstm-crf deep learning model to realize the identification of military texts.

At present, most of the word vector representations of named entity recognition are context-independent. The authors propose a named entity recognition method that can perceive multiple granularity semantic relationships. This method integrates BERT[9], BiLSTM-CRF entity recognition model, referred to as BertBC model. This model can generate word vector, text vector and position vector, and realize context association with word vector. It not only avoids the complexity of manual extraction, but also improves the accuracy of news named entity identification.
2. Named entity recognition model
For the named entity recognition from military news, the deep learning model based on BertBC model is adopted, and the structure is shown in Figure 1. The model consists of the BERT model layer, the BiLSTM neural network layer and the CRF layer. First, the BERT model is used for text pre-training. The BERT model will represent the input and can perceive the multi-granularity semantic relationship. The output vector of the BERT model is feed to the BiLSTM neural network layer, the context features are extracted and the feature vector is output. The CRF layer labels the named entities according to the feature vector and outputs the corresponding tags. Entity annotations use BIO labeling rules[10].

2.1. BERT model layer
BERT model layer[11] is the first layer of BertBC model, the word vector representation layer of context, and the core layer of BertBC model. The word vector is going from word embedding to contextual word embedding. The BERT model can perceive multi-granularity semantic relations and further increase the generalization ability of the word vector model. It can describe character-level, word-level, sentence-level and even inter-sentence relationship characteristics. The BERT model uses transformer to encode to achieve context-dependent word vector representation. The BERT model can simultaneously use the information in both directions. Specific steps are as follows:

**Step 1** Split and enter the text sequence in sentence units.

**Step 2** The BERT model will generate word vectors, text vectors, and position vectors. Word vector is a vector representation which integrates semantic information. Text vector characterizes the global semantic information of text. Position vectors are used to distinguish words in different positions.

**Step 3** The BERT model takes the sum of the word vector, text vector and position vector as the model input.

**Step 4** It is used for the sequence labeling task, and the pre-training task of the model is performed, that is, the semantics of the word vector is enhanced by the Transformer of the BERT model.

**Step 5** Enter the results into the BiLSTM neural network layer of the model.

2.2. BiLSTM neural network layer
The BiLSTM neural network layer is the second layer of the BertBC model, which combines the forward LSTM with the backward LSTM. Schuster et al[12] proposed a bidirectional cyclic neural network, and then Graves et al[13] proposed the BiLSTM model. This layer not only learns about the connections between texts, but also solves the problem of entity remote dependencies.

The LSTM model is composed of input word $X_t$ from time t, Cell state $C_t$, Temporary cell status $\tilde{C}_t$, Hidden layer state $h_t$, Forgotten door $f_t$, Memory gate $i_t$ and Output gate $O_t$. The LSTM is calculated by forgetting information and remembering new information from the cell state. Useful information is passed on at subsequent moments, while useless information is discarded and hidden state

![Figure 1. Military domain named entity recognition model structure](image-url)
\( h_t \) is output at each time step. Forgetting, memory and output are controlled by the forgetting gate \( f_t \), the memory gate \( i_t \) and the output gate \( O_t \) calculated by the hidden layer state \( h_{t-1} \) and the current input \( X_t \).

\[
\begin{align*}
    f_t &= \sigma(w_f \cdot [h_{t-1}, x_t] + b_f) \quad (1) \\
    i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2) \\
    \tilde{C}_t &= \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3) \\
    C_t &= f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (4) \\
    O_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5) \\
    h_t &= O_t \cdot \tanh(C_t) \quad (6)
\end{align*}
\]

Finally, you can get the same hidden layer state sequence as the length of the sentence \( \{h_0, h_1, \ldots, h_{n-1}\} \). The forward output gate \( LSTM_L \) inputs three words in turn to get three vectors \( \{h_{L0}, h_{L1}, h_{L2}\} \). The backward output gate \( LSTM_R \) inputs three words in turn to get three vectors \( \{h_{R0}, h_{R1}, h_{R2}\} \). Finally, the forward and backward vectors are spliced to get \( \{h_{L0}, h_{L1}, h_{L2}, h_{R0}, h_{R1}, h_{R2}\} \), namely \( \{h_{L0}, h_1, h_2\} \). The BiLSTM model diagram is shown in Figure 2.

![Figure 2. BiLSTM model diagram](image)

2.3. CRF layer

The CRF layer is the third layer of the BertBC model, which is used to learn the annotation features. To ensure that the predicted tag is legal, it adds some constraints to the last predicted tag \(^{[14]}\). Constraints can be automatically learned during data training. Let \( X = (X_1, X_2, \ldots, X_n) \) and \( Y = (Y_1, Y_2, \ldots, Y_m) \) are joint random variables. If an undirected graph \( G = (V, E) \) from a random variable \( Y \) represents Markov random field, then the conditional probability distribution \( P(Y|X) \) is called conditional random field. \( X \) is called an input variable and an observation sequence. \( Y \) is called an output sequence, a marker sequence, and a state sequence. The whole sequence is determined partly by the BiLSTM output matrix and partly by the CRF transfer matrix. The optimal scoring sequence of text can be obtained by maximizing logarithmic likelihood function during model training.

3. Identification step

3.1. Named entity annotation

An entity annotation is a label for each element in a text sequence. We annotate the entities using the bio annotation method. "B" is used to represent the first word of the named entity. "I" is used to indicate the internal and ending words of the named entity. "O" indicates a character that is not a named entity. The "B" and "I" labels are followed by "LOC", "ORG", and "PER". "LOC" stands for place name. "ORG" represents the name of the organization. "PER" stands for name. As shown in Table 1 below.

| Entity type         | Entity begins | Inside and end of the entity |
|---------------------|---------------|------------------------------|
| Place name          | B-LOC         | I-LOC                        |
| People name         | B-PER         | I-PER                        |
| Institution name    | B-ORG         | I-ORG                        |
3.2. Named entity recognition pre-training
In order to improve the pre-training speed, the author also used the Google BERT Chinese pre-training model(chinese_L-12_H-768_A-12). The pre-training steps are as follows:

**Step 1** Use the BERT model to represent our input. The last layer of the BERT is the last dimension of all the transformer results, which is a three-dimensional vector dimension.

**Step 2** Use the results of the BERT pre-training as input to the BiLSTM neural network.

**Step 3** feed the output of the BiLSTM neural network to the CRF for decoding.

3.3. News Named Entity Recognition Algorithm
After training with the BertBC model, a model file will be formed to be called by the named entity recognition algorithm. After the news named entity recognition algorithm, the input news text will become the corresponding entity result. The algorithm description is shown in Table 2.

| Table 2. News named entity recognition algorithm |
|-----------------------------------------------|
| **News named entity recognition algorithm**   |
| Input: news text                             |
| 1. Load BertBC pre-trained model as output, Google model (chinese_L-12_H-768_A-12) |
| 2. result_h←id                               |
| 3. for row ∈ size                            |
| 4. if id==0 or label in range                |
| 5. return result_h                           |
| 6. compare(tokens, tags)                     |
| 7. for each label ∈ labels                   |
| 8. index ← label                             |
| 9. for char,tag in (string, tags)            |
| 10. if char=BIO append char                  |
| 11. return result: Output: Entity (ORG, PER, LOC) |

First, the algorithm loads the model and converts the text id into the real sequence result. The model is then used to determine whether the id is consistent with the entity of the training model. Finally, add the entity sequence and return the result.

4. Analysis of results

4.1. Experimental data
The Chinese data was pre-trained using the January 1998 People's Daily data set as a training data set. The data set partitioning is adopted as shown in Table 3. And the author uses a baidu military news to verify the effect of named entity recognition.

| Table 3. Training, verification, test set division |
|-----------------------------------------------|
| Types   | Training data | Test Data | verify the data |
|---------|---------------|-----------|----------------|
| Proportion/%  | 75%           | 15%       | 10%            |

4.2. Experimental setup and environment
The experiment used the pre-trained language model BERT-Base provided by Google.

BERT parameter setting: The maximum sequence length is 128. The train_batch_size is 16. The learning_rate is 5e-5. The dropout_rate is 0.5 [15].

BiLSTM parameter setting: The word vector dimension is 100. The BiLSTM hidden layer is 1 layer. The number of neurons in the forward and reverse LSTM is 128. The learning rate is 0.001. The batch size batch_size is 128. The number of iterations epoch is 40 [16].

The experimental environment is shown in Table 4. The experiment uses the accuracy rate P, the recall rate R and the F value to evaluate the recognition effect. The F value reflects the overall test results. The calculation formula for the evaluation index is as follows:
4.3. Analysis of experimental results

In this paper, the CRF model, the word2vec-BiLSTM-CRF model and the BertBC model are selected for comparative experiments. The word2Vec is used as the representation of sentences in the first two groups of models. The third group of experiments adopted the BertBC model of this paper, and the results were shown in Table 5.

| Model             | Accuracy | Recall rate | F1    |
|-------------------|----------|-------------|-------|
| CRF               | 82.39%   | 81.77%      | 82.08%|
| WORD2VEC-BILSTM-CRF | 89.58%   | 87.32%      | 88.43%|
| BertBC            | 91.11%   | 92.36%      | 91.73%|

The experimental results show that the CRF model has a good accuracy and can complete the target task. The effect of the word2vec-BiLSTM-CRF model is significantly improved compared to the CRF model. This shows that BiLSTM network can extract the context information of a sentence, so as to carry out deep modeling of a sentence. BertBC works better than word2vec-bilstm-crf. This shows that BERT can use the transformer as an encoder to achieve context-sensitive. It can have better parallelism, deeper layers, and enhance the semantics of words through self-attention to achieve better results.

The BertBC model test used in this paper uses the accuracy P, the recall rate R and the F value to verify the results. The test data is shown in Table 6.

| Entity type | P   | R   | F    |
|-------------|-----|-----|------|
| overall     | 91.11% | 92.36% | 91.73% |
| LOC         | 91.90% | 93.71% | 92.80% |
| ORG         | 85.02% | 87.03% | 86.01% |
| PER         | 97.06% | 96.15% | 96.61% |

It can be seen from the test results that this model has a high accuracy for name entities. The accuracy of the organization name is still slightly worse. This may be due to the limited number of institutional data and the large changes in the organization’s name.

4.4. Experimental effect analysis

For the part of military news, we can see that the model extraction effect is shown in Table 7.
Table 7. BertBC model military news extraction experiment results

| news content | LOC | ORG           | PER     |
|--------------|-----|---------------|---------|
| Yonhap News Agency reported that the first long-term (deputy ministerial level) of the National Security Office in Cheong Wa Dae, South Korea, said on the 22nd that the government decided not to renew the Korea-Japan Military Intelligence Protection Agreement. Jin Yougen said that the Japanese government will be notified through diplomatic channels within the time limit for renewal according to the agreement. | Korea | Yonhap, Cheong Wa Dae, National Security Office | Jin Yougen |

It can be seen from the table that the model can effectively identify place names, institution names, and person names. It achieves the purpose of identifying the named entity of news text.

5. Conclusion

Traditional military public opinion analysis has the problem of huge news information. For this reason the author puts forward the enhancement named entity recognition from the military news. And the author USES relevant data to verify the model. The results show that the accuracy, recall rate, f value and recognition effect of the method are improved.

ACKNOWLEDGMENT

This article has been awarded by the Fundamental Research Fund for the Central Universities (30918015103, 30918012204), Nanjing Science and Technology Development Plan Project (201805036), and "13th Five-Year" equipment field fund (61403120501), China Academy of Engineering Consulting Research Project(2019-ZD-1-02-02).

References

[1] Zhang Fan, Wang Min. Medical Named Entity Recognition Based on Deep Learning[J]. Computing Technology and Automation, 2017, 36 (1): 124-125.
[2] Hammerton J . Named Entity Recognition with Long Short-Term Memory[C]// Conference on Natural Language Learning at Hlt-naacl. Association for Computational Linguistics, 2003.
[3] Shan Heyuan, Zhang Haisu, Wu Zhaolin. A Military Named Entity Recognition Method Based on CRFs under Small Granularity Strategy[J]. Journal of Armored Force Engineering Institute, 2017, 31 (1): 87-88.
[4] Santos C N D , Guimarães, Victor. Boosting Named Entity Recognition with Neural Character Embeddings[J]. Computer Science, 2015.
[5] Strubell E, Verga P, Belanger D, et al. Fast and accurate entity recognition with iterated dilated convolutions[J]. arXiv preprint arXiv:1702.02098, 2017.
[6] Huang Z, Xu W, Yu K. Bidirectional LSTM-CRF models for sequence tagging. Computer Science. 2015-8[2018-8].https://arxiv.org/abs/1508.01991v1
[7] Zheng Yanbin, Xia Zhichao, Guo Zhi, Huang Yongzhong, Liu Wenfen. The Named Entity Recognition of the News Texts of the Ten ASEAN Countries[J]. Science Technology and Engineering, 2018, 18(35):162-168.
[8] Wang Xuefeng, Yang Ruopeng, Zhu Wei. A Military Named Entity Recognition Method Based on Deep Learning[J]. Journal of Armored Force Engineering Institute, 2018, 32(04):94-98.
[9] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.
[10] Zhang Guohua. Research on Automatic Labeling of Chinese Framework Semantic Roles[D]. Shanxi University, 2008.
[11] Rei M. Semi-supervised multitask learning for sequence labeling[J]. arXiv preprint arXiv:1704.07156, 2017.
[12] Schuster M, Paliwal K K. Bidirectional recurrent neural networks[J]. 1997, 45(11):2673-2681.
[13] Graves A , Jürgen Schmidhuber. Framewise phoneme classification with bidirectional LSTM and
other neural network architectures[J]. Neural Netw, 2005, 18(5):602-610.

[14]Bale T L, Vale W W. CRF and CRF receptors: role in stress responsivity and other behaviors.[J]. Annual Review of Pharmacology & Toxicology, 2004, 44(44):525.

[15]Yang Piao, Dong Wenyong. Chinese Named Entity Recognition Method Based on BERT Embedded [J/OL]. Computer Engineering: 1-7[2019-07-24].https://doi.org/10.19678/j.issn.1000-3428.0054272.

[16]Wu Hui,Lü Li,Yu Bihui.Chinese Named Entity Recognition Based on Migration Learning and BiLSTM-CRF[J].Mini-micro Systems,2019,40(06):1142-1147.