A Nested Named Entity Recognition Model Based on Multi-agent Communication Mechanism (Student Abstract)

Canguang Li[1, 2], **Guohua Wang**[1, 2], **Jin Cao**[1, 2], **Yi Cai**[1, 2*]
[1] School of Software Engineering, South China University of Technology, Guangzhou, China
[2] Key Laboratory of Big Data and Intelligent Robot (South China University of Technology), Ministry of Education
{secanguangli, sejincao}@mail.scut.edu.cn, {ghwang, ycai}@scut.edu.cn

Abstract

Traditional sequence tagging methods for named entity recognition (NER) face challenges when handling nested entities, where an entity is nested in another. Most previous methods for nested NER ignore the effect of entity boundary information or type information. Considering that entity boundary information and type information can be utilized to improve the performance of boundary detection, we propose a nested NER model with a multi-agent communication module. The type tagger and boundary tagger in the multi-agent communication module iteratively utilize the information from each other, which improves the boundary detection and the final performance of nested NER. Empirical experiments conducted on two nested NER datasets show the effectiveness of our model.

Introduction

NER is a task to extract named entities from texts and classify them into pre-defined types such as *protein*, *DNA* and *RNA*. NER is usually considered as a sequence tagging task where each word is tagged with a single label, such as *B-DNA* and *E-protein*. The single label is consisted of a boundary label (*B*, *E*) and a type label (*DNA*, *protein*). However, traditional sequence tagging methods can not handle nested entities, where an entity is nested in another. Considering the phrase "IL - 6 gene" from GENIA dataset, "IL - 6" is a *protein* and "IL - 6 gene" is a *DNA*. Words "IL - 6" shared by two entities have two kinds of labels. Only tagging these words with a single label can not extract all entities.

In recent studies, the two-stage methods for nested NER are fairly popular. Sohrab and Miwa (2018) propose an exhaustive model which firstly enumerates possible entity regions and then classifies entity regions. Zheng et al. (2019) and Tan et al. (2020) propose similar models which firstly detect entity boundaries and then classify candidate entities derived from entity boundaries. These models work to a certain extent. However, model of Sohrab and Miwa (2018) ignores that entity boundary information helps reduce the scope for entity type recognition and extract many non-entity regions. Models of Zheng et al. (2019) and Tan et al. (2020) ignore that entity type information can help detect entity boundaries and the performance of boundary detection limits the performance of nested NER. For example, "gene" is often the boundary (*E*) of *DNA* or the inside (*I*) of *protein* in GENIA dataset. Without the type information, it may be hard to determine the boundary label of "gene".

In this paper, we propose a multi-agent communication module (Wang et al. 2019) to utilize boundary information and type information to improve boundary detection. It contains a type tagger and a boundary tagger. The type tagger can utilize the output of the boundary tagger for type tagging and the boundary tagger can utilize the output of the type tagger for boundary tagging. They communicate and collaborate on boundary detection iteratively. The final output of the boundary tagger is used to derive candidate entities for entity classification. Compared with models of Zheng et al. (2019) and Tan et al. (2020) which only use context feature of words to detect entity boundaries, our model effectively utilizes the boundary information and type information for boundary detection through the multi-agent communication module, which improves the performance of boundary detection and so that improves the performance of nested NER.

Our Model

Feature Representation Module We employ two independent bidirectional LSTM (Bi-LSTM) to respectively get the boundary feature and type feature. For word x_i, the boundary feature is represented as the concatenation of the forward and backward hidden states $\mathbf{h}_i^b = [\overrightarrow{\mathbf{h}_i^b}; \overleftarrow{\mathbf{h}_i^b}]$. The type feature is obtained in a similar way and is represented as $\mathbf{h}_i^t = [\overrightarrow{\mathbf{h}_i^t}; \overleftarrow{\mathbf{h}_i^t}]$.

Initial Boundary tagger We employ this module to obtain the initial boundary labels without type information at the beginning. For word x_i, we feed its boundary feature $\mathbf{h}_i^b$ into a multi-label linear layer with a sigmoid function to predict boundary labels. We adopt binary cross entropy function to compute loss L^b of the predicted and true class distributions.

Multi-agent Communication Module This module is used to improve boundary detection based on initial boundary labels. It contains a multi-label type tagger and a multi-label boundary tagger. **For type tagger**, we use an attention mechanism to match and aggregate boundary information:

$$\mathbf{g}_i^b = \sum_{k=1}^{n} \alpha_k \mathbf{h}_k^t, with\ \alpha_k = \frac{exp(\mathbf{h}_i^b \odot \mathbf{m}_i^b \mathbf{h}_k^t)}{\sum_{j=1}^{n} exp(\mathbf{h}_i^b \odot \mathbf{m}_i^b \mathbf{h}_j^t)} \quad (1)$$

---

[*] Corresponding author: Yi Cai, ycai@scut.edu.cn

Model	GENIA			GermEval 2014		
	P	R	F1	P	R	F1
Exhaustive[1]	73.3	68.3	70.7	75.0	60.8	67.2
Layered[1]	76.1	66.8	71.1	72.9	61.5	66.7
BA[1]	75.9	73.6	74.7	74.5	69.1	71.7
Seq2Seq	78.5	72.2	75.2	**83.7**	75.4	79.4
BENSC	**79.0**	66.4	72.2	75.8	68.2	71.8
Our model	78.2	**74.8**	**76.5**	80.2	**82.9**	**81.6**

Table 1: Overall results on two nested NER datasets.

where $\mathbf{m}_i^b$ is a mask vector whose value is 1 or 0 if i-th word has or has no positive predicted boundary labels, $\odot$ denotes element-wise multiplication. Then we concatenate the boundary information $\mathbf{g}_i^b$ and type feature $\mathbf{h}_i^t$ and feed it into the type tagger to predict type labels. **For boundary tagger**, we adopt the similar attention mechanism to match and aggregate type information:

$$\mathbf{g}_i^t = \sum_{k=1}^{n} \beta_k \mathbf{h}_k^b, with\ \beta_k = \frac{exp(\mathbf{h}_i^t \odot \mathbf{m}_i^t \mathbf{h}_k^b)}{\sum_{j=1}^{n} exp(\mathbf{h}_i^t \odot \mathbf{m}_i^t \mathbf{h}_j^b)} \quad (2)$$

Then we feed the concatenation of the type information $\mathbf{g}_i^t$ and boundary feature $\mathbf{h}_i^b$ into the boundary tagger to predict boundary labels. In this way, the type tagger and boundary tagger then communicate and collaborate with each other iteratively until the given iteration number l. At each iteration $r \in \{1, 2, ..., l\}$, we adopt binary cross entropy function to compute the type loss L_r^t and boundary loss L_r^b.

Candidate Entity Classifier We employ a Bi-LSTM to encode words in the entity region and concatenate the last forward and backward output to represent the entity region $\mathbf{h}^e = [\overrightarrow{\mathbf{h}}_i^e; \overleftarrow{\mathbf{h}}_j^e]$. Then we feed $\mathbf{h}^e$ into a linear classifier with a softmax function to predict the entity type. We adopt cross entropy function to compute the classification loss L^e.

Optimized Objective The optimized objective is the linear sum of losses L^b, L_r^t, L_r^b and L^e.

Experiments

Experimental Settings

Datasets We conduct empirical experiments on two nested NER datasets: **GENIA** and **GermEval 2014**. We use the preprocessed version released by Zheng et al. (2019).

Baselines We denote baselines as Exhaustive (Sohrab and Miwa 2018), Layered (Ju, Miwa, and Ananiadou 2018), BA (Zheng et al. 2019), Seq2Seq (Straková, Straka, and Hajic 2019) and BENSC (Tan et al. 2020). Experiments of Seq2Seq and BENSC are reproduced on the datasets we use.

Experimental Results

Table 1 shows the overall results of nested NER. Our model outperforms other models in terms of Recall (R) and F1. It illustrates that our model can extract more correct entities. BENSC and Seq2Seq achieve higher Precision (P) but lower Recall because they extract fewer entities and miss some correct entities. Similar phenomenon occurs on boundary detection. To illustrate the effectiveness of our multi-agent communication module to improve boundary detection, we compare with BA and BENSC which only use the context feature of words to detect entity boundaries. Table 2 shows that our model outperforms BA and BENSC on Recall and F1 because our model effectively utilizes boundary and type information through the multi-agent communication module.

[1]The results are reported by Zheng et al. (2019)

Model	P	R	F1
BA[1]	79.7	76.9	78.3
BENSC	**83.1**	69.8	75.9
Our model	81.8	**78.3**	**80.0**

Table 2: Boundary detection on GENIA dataset.

Conclusions

This paper proposes a nested NER model based on multi-agent communication mechanism. Considering that the boundary information can help recognize entity type and type information can help recognize entity boundary, we propose a multi-agent communication module to effectively utilize these two kinds of information. In this way, we improve the performance of boundary detection and so that improve the performance of nested NER. The results of experiments show that our model outperforms existing models.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities, SCUT (No.D2182480), the National Key Research and Development Program of China, the Science and Technology Programs of Guangzhou (No.201802010027, 201902010046), National Natural Science Foundation of China (62076100).

References

Ju, M.; Miwa, M.; and Ananiadou, S. 2018. A Neural Layered Model for Nested Named Entity Recognition. In *NAACL-HLT*, 1446–1459.

Sohrab, M. G.; and Miwa, M. 2018. Deep Exhaustive Model for Nested Named Entity Recognition. In *EMNLP*, 2843–2849.

Straková, J.; Straka, M.; and Hajic, J. 2019. Neural Architectures for Nested NER through Linearization. In *ACL*, 5326–5331.

Tan, C.; Qiu, W.; Chen, M.; Wang, R.; and Huang, F. 2020. Boundary Enhanced Neural Span Classification for Nested Named Entity Recognition. In *AAAI*, 9016–9023.

Wang, S.; Wei, Z.; Fan, Z.; Liu, Y.; and Huang, X. 2019. A Multi-Agent Communication Framework for Question-Worthy Phrase Extraction and Question Generation. In *AAAI*, 7168–7175.

Zheng, C.; Cai, Y.; Xu, J.; Leung, H.-f.; and Xu, G. 2019. A Boundary-aware Neural Model for Nested Named Entity Recognition. In *EMNLP-IJCNLP*, 357–366.