Incorporating Bidirection-Interactive Information and Semantic Features for Relational Facts Extraction (Student Abstract)

Yang Yu, [1,2] **Guohua Wang,** [1,2] **Haopeng Ren,** [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
{yangyu_scut, renhp_scut}@foxmail.com, {ghwang, ycai}@scut.edu.cn

Abstract

The interaction between named entity recognition and relation classification is quite essential for the extraction of relational triplets. However, most of jointly extraction works only consider unidirectional interaction between the two subtasks. They even neglect the interactive information totally. In order to tackle these problems, we propose a novel unified joint extraction model which considers bidirection-interactive information between the two subtasks. Our model consists of two modules. The first module utilizes Bi-LSTM and GCN to capture the sequential and the structure-semantic features of a sentence, The second module utilizes two layers to capture bidirection-interactive information between the two subtasks and generates relational triplets respectively. The experimental results show that our proposed model outperforms the state-of-the-art models on two public datasets.

Introduction

Joint extraction of entities and relations simultaneously from unstructured text is crucial to build large structural knowledge. It is widely applied in many NLP tasks, including the construction of knowledge base, text mining and question answering. A pair of entities and their corresponding semantic relation are often described as a relational triplet, such as <*Trump*, *president of*, *United States*>.

Joint extraction of entities and relations contains two subtasks that are named entity recognition and relation classification. Although most recent works can extract overlapping relational facts efficiently and overcome the problem of error propagation, most of them neglect the bidirectional interaction between these two subtasks. That is to say, the interaction between the two subtasks is not fully modeled. However, the bidirection-interactive information is quite essential for the extraction of relational triplets. For example, the sentence "Chicago is located in the United States" contains an entity pair <*Chicago*, *United States*>. If we have learnt that the relation of the entity pair <*Chicago*, *United States*> is "country", then the label of the entity pair should not be "S-PERSON" and "B-PERSON, I-PERSON". If we have learnt that the label of the entity pair are "S-LOCATION" and "B-LOCATION, I-LOCATION", then the relation of the entity pair should not be "president". Therefore, we can conclude that relation extraction can facilitate named entity recognition. At the same time, named entity recognition is also helpful for relation extraction.

To address the problems above, we propose a unified joint extraction model, which considers bidirection-interactive information between named entity recognition and relation classification. More specially, our model consists of two main modules. One is a lower sub-network: *sentence encoding module*. This module utilizes Bi-LSTM (Hochreiter and Schmidhuber 1997) and GCN (Kipf and Welling 2017) to capture the sequence-semantic features and the structure-semantic features of a sentence. The other one is the upper sub-network: *triplets decoding module*. This module consists of interactive layer and decode layer. Interactive layer captures the bidirection-interactive information between named entity recognition and relation classification. Decode layer generates the relational triplets.

Methodology

In this paper, we propose a novel Seq2Seq neural model, which considers bidirection-interactive information between named entity recognition and relation classification. Our model consists of two main modules, as detailed below.

Sentence Encoding Module

To consider both the sequence-semantic features and the structure-semantic features of the sentence. We first apply bidirectional RNN to extract sequential features, then we use GCN to further extract structure-semantic features.

Bi-LSTM For the word w_t of the input sequence at time t, we utilize Bi-LSTM to get its forward and backward hidden states. We calculate the average of the two hidden states to represent the output of the word at time step t.

GCN We use a dependency parser to create a dependency tree for the input sentence. Then, we apply GCN to extract structure-semantic features through convolving the features of neighboring nodes of the dependency tree.

$$o_t^{l+1} = ReLU(\sum_{v \in N(t)} (W^l o_v^l + b^l)) \tag{1}$$

where t is the target node and $N(t)$ represents the neigh-

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

borhood node of t, including t itself. o_v^l denotes the hidden feature of node v at layer l. W^l and b^l are learnable weights.

Triplets Decoding Module

This module consists of two layers. We introduce the interactive layer to capture the bidirection-interactive information between named entity recognition and relation classification. We add the decode layer to generate triplets.

Interactive Layer Interactive layer consists of several decoders. Each decoder consists of three time steps and generates the interactive hidden states $h_t^{D_1}$ of one triplet. At time step one, it decodes the interactive hidden states of the relation. At time step two and time step three, it decodes the interactive hidden states of the two entities respectively.

$$h_t^{D_1} = BiLSTM(W^{D_1} c_t, h_{t-1}^{D_1}) \tag{2}$$

where c_t is the context attention vector calculated by encoding outputs O^E and $h_{t-1}^{D_1}$. W^{D_1} is learnable weight matrix.

Decode Layer Decode layer is similiar to interactive layer, each decoder consists of three time steps and decodes a relational triplet. The hidden state $h_t^{D_2}$ at time step t of decode layer is calculated by the word vector extracted at time step $t-1$ and previous hidden state $h_{t-1}^{D_2}$. After that, we use $h_t^{D_2}$ to calculate the logits which are used to acquire the relational triplets prediction.

$$q_i^{D_2} = f([h_t^{D_2}, o_i^E] W_1^{D_2}) W_2^{D_2} \tag{3}$$

where $q_i^{D_2}$ is the confident value of i-th word. o_i^E is the encoding output of i-th word. $W_1^{D_2}$ and $W_2^{D_2}$ are both learnable weight matrices. f is activation function.

Experiments

Datasets

(1) **NYT dataset**. It consists of 1.18M sentences sampled from 294k 1987-2007 New York Times news articles. There are totally 24 valid relations.
(2) **WebNLG dataset**. It is originally created for Natural Language Generation (NLG) task. This dataset contains 246 valid relations.

Baseline Methods

We compare our models with several state-of-the-art baseline models, including NTS (Zheng et al. 2017), OneDecoder (Zeng et al. 2018), MultiDecoder (Zeng et al. 2018), GraphRel$_1p$ (Fu, Li, and Ma 2019), GraphRel$_2p$ (Fu, Li, and Ma 2019), CopyMTL-One (Zeng, Zhang, and Liu 2020), and CopyMTL-Mul (Zeng, Zhang, and Liu 2020).

Experimental Results

We use the standard micro Precision, Recall and F1 score to evaluate the experimental results. Triplets are regarded as correct when their relations and entities are both correct. Table 1 shows the precision, recall, and F1 score of baseline models and our model on two public datasets. We can see that our model significantly outperforms other baseline models on both datasets. Considering the interactive information is not fully captured in baseline models, the results well justify the effectiveness of incorporating bidirection-interactive information and semantic features in our model.

Model	NYT			WebNLG		
	P	R	F1	P	R	F1
NTS	62.4	31.7	42.0	52.5	19.3	28.3
OneDecoder	59.4	53.1	56.0	32.2	28.9	30.5
MultiDecoder	61.0	56.6	58.7	37.7	36.4	37.1
GraphRel$_1p$	62.9	57.3	60.0	42.3	39.2	40.7
GraphRel$_2p$	63.9	60.0	61.9	44.7	41.1	42.9
CopyMTL-One	72.7	**69.2**	70.9	57.8	60.1	58.9
CopyMTL-Mul	75.7	68.7	72.0	58.0	54.9	56.4
Our model	**78.2**	69.1	**73.3**	**59.2**	**61.3**	**60.2**

Table 1: Overall results on NYT and WebNLG datasets.

Conclusion

In this paper, we propose a unified model for joint extraction of entities and overlapping relations. Our model firstly utilizes two layers to extract the sequence-semantic features and the structure-semantic features of a sentence. Then, it employs an interactive layer to consider bidirectional interaction between named entity recognition and relation extraction. Finally, it uses decode layer to generate relational triplets. Experimental results show that our method can extract relational facts effectively and achieve the start-of-the-art results on two public datasets.

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

Fu, T.-J.; Li, P.-H.; and Ma, W.-Y. 2019. GraphRel: Modeling text as relational graphs for joint entity and relation extraction. In *ACL*, 1409–1418.

Hochreiter, S.; and Schmidhuber, J. 1997. Long short-term memory. *Neural computation* 9(8): 1735–1780.

Kipf, T. N.; and Welling, M. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *ICLR*.

Klambauer, G.; Unterthiner, T.; Mayr, A.; and Hochreiter, S. 2017. Self-Normalizing Neural Networks. In *NIPS*, 971–980.

Zeng, D.; Zhang, H.; and Liu, Q. 2020. CopyMTL: Copy Mechanism for Joint Extraction of Entities and Relations with Multi-Task Learning. In *AAAI*, 9507–9514.

Zeng, X.; Zeng, D.; He, S.; Liu, K.; and Zhao, J. 2018. Extracting relational facts by an end-to-end neural model with copy mechanism. In *ACL*, 506–514.

Zheng, S.; Wang, F.; Bao, H.; Hao, Y.; Zhou, P.; and Xu, B. 2017. Joint Extraction of Entities and Relations Based on a Novel Tagging Scheme. In *ACL*, 1227–1236.