SCIENCE CHINA
Information Sciences

• LETTER •

June 2020, Vol. 63 169103:1–169103:3
https://doi.org/10.1007/s11432-018-9721-7

Fine-grained relation extraction with focal multi-task learning

Xinsong ZHANG[1], Tianyi LIU[1], Weijia JIA[2,1*] & Pengshuai LI[1]

[1]*School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai* 200240*, China;*
[2]*State Key Lab of IoT for Smart City, University of Macau, Macau* 999078*, China*

Received 11 November 2018/Revised 26 November 2018/Accepted 19 December 2018/Published online 15 April 2020

Citation Zhang X S, Liu T Y, Jia W J, et al. Fine-grained relation extraction with focal multi-task learning. Sci China Inf Sci, 2020, 63(6): 169103, https://doi.org/10.1007/s11432-018-9721-7

Dear editor,

Relation extraction aims to identify relation facts for pairs of entities in raw texts to construct triplets such as [Arthur Lee, place_born, Memphis]. To automatically extract relation facts, the distant supervision strategy [1] has been proposed, which assumes that, if there exists a relation between two entities in a known knowledge base, all the sentences that mention these two entities will likely express the same relation. Recently, neural networks have been widely applied to distant supervised relation extraction and have achieved good performances by precisely extracting semantic features [2]. However, most current studies have not paid sufficient attention to distinguish fine-grained relations. Relations, such as company/shareholder, company/advisors, and company/founders, contain similar relation features. All three relations can be easily recognized from others such as location/contain; however, it is difficult for relation extractors to distinguish them from each other. Meanwhile, previous studies have not focused on relations that are difficult to distinguish. Even though remarkable relations can be easily extracted, similar relations are difficult to be recognized precisely to improve the fine-grained relation extraction.

This study proposes a novel approach for the extraction of fine-grained relations. Herein, type-based attention and a new training algorithm, namely focal multi-task learning, were proposed for extracting fine-grained relation features. In the proposed fine-grained relation extractor, the bidirectional gated recurrent unit (BGRU) network serves as the fundamental neural model for the sentence encoder. Next, a type-based attention approach was built to represent the relation features as a weighted sum of the sentence embeddings in a bag according to the entity-type information. To overcome the limitations associated with similar relations, the proposed model was trained with focal multi-task learning. Entity-type predictions were proposed as auxiliary tasks and were trained together with the relation extraction task in parallel by sharing all the hidden representation layers. Finally, multi-task learning was integrated with the focal loss to focus on similar relations which are difficult to distinguish. The proposed method was evaluated on a widely used benchmark. Extensive experiments demonstrated that the proposed approach is effective and outperforms the state-of-the-art approaches.

The proposed approach. In the distant supervised relation extraction paradigm, all sentences labeled by a relation triple constitute a bag. The relation triple is described as [head, relation, tail], where head and tail are both entities. Suppose that there are N bags $\{B_1, B_2, \ldots, B_N\}$ in the training set and that the i-th bag contains q_i sentences $B_i = \{b_1^i, \ldots, b_{q_i}^i\}$ $(i = 1, \ldots, N)$. Our innovative solution for fine-grained relation extraction in distant supervision comprises three key mod-

* Corresponding author (email: jia-wj@cs.sjtu.edu.cn)

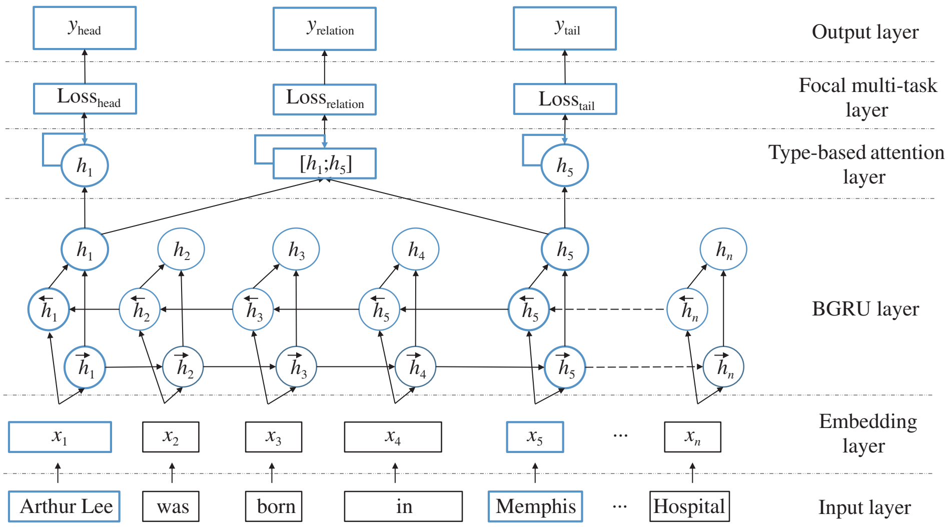

Figure 1 (Color online) The architecture of our fine-grained relation extractor illustrating the procedure for handling one sentence and predicting the relation between [Arthur Lee] and [Memphis].

ules, as shown in Figure 1; these modules are sentence encoder, type-based attention, and focal multi-task learning.

• Sentence encoder. Given a sentence b^* and two target entities, BGRU is used to construct a distributed representation of the relation features for the sentence. The input representations of the proposed model are pre-trained word embeddings and position embeddings. We used BGRU, including both forward and backward subnetworks, to capture the global sequence information. The hidden state of each word in the sentence is $h_t = \text{BGRU}(x_t)$, as shown in Figure 1. The final relation representation of the sentence is concatenated with the hidden states h^{head} and h^{tail}.

• Typed-based attention. Given a bag of sentences B^* and two target entities, the representation of the bag was computed by summing up the weighted sentence representations, which are given by the sentence encoder. The weights are learned using a type-based attention mechanism. Attention mechanism has been proven to effectively alleviate the wrong label problem in the distant supervised relation extraction [3]. However, previous studies have generated attention weights by aligning sentences to external relation features, which have a low capability of distinguishing the fine-grained relations owing to their similar features. In the proposed model, we generate attention weights with entity-type features, which are learned from entity-type predictions. The entity-type information is helpful to recognize the fine-grained relations. We propose the following formulas to compute the type-based attention weights:

$$w_i = W_\alpha^{\mathrm{T}}(\tanh[h_i^{\text{head}}; h_i^{\text{tail}}]), \tag{1}$$

$$\alpha_i = \frac{\exp(w_i)}{\sum_{j=1}^{q}\exp(w_j)}, \tag{2}$$

$$h^{(\text{relation})} = \sum_{i=1}^{q} \alpha_i [h_i^{\text{head}}; h_i^{\text{tail}}]. \tag{3}$$

Finally, we feed the relation representation $h^{(\text{relation})}$ of the bag B^* into the focal multi-task layer and the output layer for the estimated probability of each relation.

• Focal multi-task learning. After computing the representation of a sentence bag, we applied the focal multi-task learning to extract the fine-grained relation features. We trained the proposed model using multi-task learning by designing three relevant tasks in parallel, which are the head-type prediction, relation extraction, and tail-type prediction, as shown in Figure 1. The estimated probability distributions are computed for all three tasks using the following equation:

$$\begin{aligned} &\hat{p}^{(m)} = \text{softmax}(W^{(m)}h^{(m)} + b^{(m)}),\\ &m \in \{\text{head}, \text{tail}, \text{relation}\}. \end{aligned} \tag{4}$$

To further focus on the similar relations that are difficult to distinguish, we assigned more weights for training sentences with ambiguous relations in the loss function by integrating a focal loss function [4] with the proposed multi-task learning. We also focused on ambiguous types for the entity-type prediction tasks with the focal loss function. Therefore, a novel joint cost function is proposed as the following equation, which is a linear combination of the focal cost functions for all tasks:

$$\phi = \sum_{i\in m} \lambda_i J(\hat{p}^i, y^i, \theta^i),$$

$$m = \{\mathrm{head}, \mathrm{tail}, \mathrm{relation}\}, \quad (5)$$

where λ_i is the weight for each respective task whose sum should be 1 and θ^i represents the parameters of the task i. The focal loss function for each task is the negative log-likelihood of the true class labels:

$$J(\hat{p}, y, \theta) = -\frac{1}{z}\sum_{j=1}^{z} \kappa(1-\hat{p}_j)^{\gamma} y_j \log(\hat{p}_j) + \beta\|\theta\|^2, \quad (6)$$

where $y \in \mathcal{R}^z$ is a one-hot vector representing the ground truth, $\hat{p} \in \mathcal{R}^z$ is the estimated probability for each class, and z is the number of class labels. β and θ represent the $L2$ regularization strength and all the training variables, respectively. κ and γ are parameters of the focal loss function. In the $J(\hat{p}, y, \theta)$ equation, a relation with higher confidence will have a smaller influence on the final loss. Herein, the joint cost function ϕ not only adjusts the weights of similar relations but also balances the importance of the different tasks. The ambiguous task is assigned bigger weights. Finally, we compute the average loss of all the training sentences.

Our results. We conducted experiments on a widely used benchmark for distant supervised relation extraction, i.e., NYT-10 [5]. We evaluated the proposed method through a classic hold-out evaluation. This evaluates our models by comparing the relation facts discovered from the test articles with those presented in Freebase. Both the aggregate precision/recall (PR) curves and the precision at top N predictions (P@N) are reported in our experiments. PCNN+ATT [3] is presented as the strongest baseline.

• Overall performance. The proposed method achieves a better PR curve than that of the baseline. The PR curve area of PCNN+ATT is 0.35, whereas the proposed method increases that to 0.39, which indicates a significant improvement of 11.4%. To further evaluate the performance of the proposed method, we proposed P@100, P@200, P@300 and their mean as the quantitative indicators to compare the proposed method with the baseline. We increased these four indicators by 7.8%, 7.4%, 6.6%, and 7.3%, respectively. The results indicate that the proposed method outperforms the baseline significantly.

• Effect of focal multi-task learning for fine-grained relations. To further demonstrate the effect of focal multi-task learning, we compared the proposed method to a fundamental BGRU network. The proposed method achieves better performance than the BGRU model and increases the PR curve area by 15.2%. Additionally, we evaluated the proposed focal multi-task learning using top predictions of fine-grained relation sentences that contain similar relation features. A significant increase of 28.0% at P@100 for fine-grained relation sentences indicates that focal multi-task learning is beneficial for distinguishing similar relations.

Conclusion. Herein, we exploited a novel approach for fine-grained relation extraction with a knowledge base by integrating a BGRU network with focal multi-task learning and type-based attention. The focal multi-task learning helps distinguish similar relations precisely, and the type-based attention sufficiently utilizes the entity-type features to accurately recognize fine-grained relations. Our experiments show that the proposed approach achieves a significant improvement over the strong baseline.

Acknowledgements This work was supported by National Basic Research Program of China (973 Project) (Grant No. 2015CB352401), National Natural Science Foundation of China (Grant Nos. 61532013, 61872239). FDCT/0007/2018/A1, DCT-MoST Joint-Project (Grant No. 025/2015/AMJ), University of Macau Grants (Grant Nos. MYRG2018-00237-RTO, CPG2018-00032-FST, SRG 2018-00111-FST).

References

1 Mintz M, Bills S, Snow R, et al. Distant supervision for relation extraction without labeled data. In: Proceedings of Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, Singapore, 2009. 1003–1011

2 Zeng D J, Liu K, Chen Y B, et al. Distant supervision for relation extraction via piecewise convolutional neural networks. In: Proceedings of Conference on Empirical Methods in Natural Language Processing, Lisbon, 2015. 1753–1762

3 Lin Y K, Shen S Q, Liu Z Y, et al. Neural relation extraction with selective attention over instances. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Berlin, 2016. 2124–2133

4 Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection. In: Proceedings of IEEE International Conference on Computer Vision, Venice, 2017. 2999–3007

5 Riedel S, Yao L M, McCallum A. Modeling relations and their mentions without labeled text. In: Proceedings of European Conference on Machine Learning and Knowledge Discovery in Databases, Barcelona, 2010. 148–163