

Entity Type Enhanced Neural Model for Distantly Supervised Relation Extraction (Student Abstract)

Long Bai, Xiaolong Jin, Chuanzhi Zhuang, Xueqi Cheng

CAS Key Lab of Network Data Science and Technology, Institute of Computing Technology,
Chinese Academy of Sciences (CAS)

School of Computer and Control Engineering, University of Chinese Academy of Sciences, Beijing
{bailong18b, jinxiaolong, zhuangcz, cxq}@ict.ac.cn

Abstract

Distantly Supervised Relation Extraction (DSRE) has been widely studied, since it can automatically extract relations from very large corpora. However, existing DSRE methods only use little semantic information about entities, such as the information of entity type. Thus, in this paper, we propose a method for integrating entity type information into a neural network based DSRE model. It also adopts two attention mechanisms, namely, sentence attention and type attention. The former selects the representative sentences for a sentence bag, while the latter selects appropriate type information for entities. Experimental comparison with existing methods on a benchmark dataset demonstrates its merits.

Introduction

Relation extraction is a typical type of information extraction, which is essential for a number of natural language processing related applications, such as knowledge base population and question answering. Its goal is to extract the semantic relationship between two given entities. A common approach of relation extraction is supervised learning, which needs a large labeled training set. However, labeling a large training set is labor-intensive and time consuming.

Distantly Supervised Relation Extraction (DSRE) is an extensively used method aiming at saving time and labor in relation extraction. This method automatically labels training data by aligning facts in an existing knowledge base to an unlabeled corpus. For this reason, DSRE methods suffer from the problem of wrong labeling. Multi-instance learning methods are thus used to reduce noises caused by wrong labeling. These methods collect all sentences that contain a pair of entities into a sentence bag, and then predict the relation on this bag instead of each sentence.

Recently developed neural network based methods (Zeng et al. 2015; Lin et al. 2016; Ji et al. 2017) follow the above multi-instance learning framework. In these methods, each sentence in a sentence bag is first encoded into a sentence vector. Then the bag representation (i.e., a vector representing the whole bag of sentences) is produced through summing up the weighted sentence vectors, where the weight

of each sentence is calculated via an attention mechanism. Finally, the relation type of the sentence bag is predicted according to its bag representation.

Feature based relation extraction methods view the information of entity type as a necessary type of features. However, few studies on neural network based models have taken such information about entities into consideration. To fill this gap, this paper applies entity type constraints to a neural network based DSRE model, where an attention mechanism is adopted to select appropriate type information for entities. Experiments on a widely-used dataset validates the superiority of our model, as compared with those existing ones.

The Proposed Model

Our model is a significant extension to the one in (Ji et al. 2017). We employ the same sentence encoder, i.e., PCNN. However, we make two important modifications. First, we improve the original sentence attention mechanism, where attention weights on sentences are fixed even when predicting different relations. Second, we replace the entity description information with fine-grained entity type. In addition, we employ another attention mechanism, called type attention, to select the appropriate types for different entities. In what follow, we describe these two modifications in detail.

Sentence Attention. Let B be the currently processing sentence bag. Firstly, following (Zeng et al. 2015; Lin et al. 2016; Ji et al. 2017), for each sentence $s_i \in B$, we encode it via PCNN as $\mathbf{s}_i = \text{PCNN}(s_i)$, where $\mathbf{s}_i \in \mathbb{R}^{3n_c}$. Here, n_c is the number of convolution kernels of size $(n_w + 2n_p) \times l$, where n_w is the word embedding size, n_p is the position embedding size and l is the window size.

Next, we derive the bag representation $\mathbf{b} \in \mathbb{R}^{3n_c}$, where an adaptive mechanism is adopted to adjust the attention weight α_i of each sentence s_i . Specifically, $\mathbf{b} = \sum_i \alpha_i \mathbf{s}_i$, where $\alpha_i = \frac{\exp(e_{i,j})}{\sum_k \exp(e_{k,j})}$ and $e_{k,j} = \mathbf{s}_k^T \mathbf{r}_j^{(s)}$. Here, $e_{k,j}$ can be viewed as the relatedness between the sentence s_i and the specified relation r_j ; $\mathbf{r}_j^{(s)} \in \mathbb{R}^{3n_c}$ represents the attention vector used for sentence attention, and $\mathbf{r}_j^{(s)} = R_s^T \mathbf{r}_j + \mathbf{b}_s$, where R_s is the transformation matrix and \mathbf{b}_s is the bias; $\mathbf{r}_j \in \mathbb{R}^{n_r}$ represents the embedding of the relation r_j at hand; n_r is the size of the relation vector.

Finally, the probability of bag B representing the relation r_j , which will be used later, can be calculated by $p(r_j|B) = \frac{\exp(o_j)}{\sum_k \exp(o_k)}$, where $\mathbf{o} = W^T \mathbf{b} + \mathbf{b}_o$, W is the transformation matrix and \mathbf{b}_o is the bias.

Type Attention. As aforesaid, fine-grained entity type provides rich semantic information about entities, which may enhance the performance of relation extraction. This motivates us to integrate type information to the word embedding of entities. As one entity may correspond to multiple types, we represent the type information of an entity as the attention sum of its all types. Specifically, we have $\mathbf{t}_e = \sum_i \beta_i \mathbf{t}_{e,i}$, where $\beta_i = \frac{\exp(e_{t,i})}{\sum_k \exp(e_{t,k})}$ and $e_{t,k} = \mathbf{t}_{e,k}^T \mathbf{r}_j^{(t)}$. Here, $\mathbf{r}_j^{(t)} \in \mathbb{R}^{n_w}$ represents the attention vector used for type attention, and $\mathbf{r}_j^{(t)} = R_t^T \mathbf{r}_j + \mathbf{b}_t$; R_t is the transformation matrix and \mathbf{b}_t is the bias; $\mathbf{t}_{e,k} \in \mathbb{R}^{n_w}$ represents the embedding of the i -th fine-grained type of entity e .

Training Objective. The objective function of training is the sum of the bag prediction loss function and the entity type constraint function, $L = L_B + \lambda L_C$, where λ is a weight factor. In this paper, we use cross-entropy loss as the bag prediction loss $L_B = -\sum_{i=1}^n \log p(r_i|B_i)$, where B_i represents the i -th sentence bag in the training set, which corresponds to the relation label r_i and the entity pair $(e_{i,1}, e_{i,2})$. The entity type constraint is defined as the sum of two constraints of the corresponding entities, $L_C = \sum_{i=1}^n (L_t(e_{i,1}) + L_t(e_{i,2}))$. The type constraint $L_t(e)$ is applied by minimizing the distance between the word embedding and type representation of the entity, namely, $L_t(e) = \|\mathbf{w}_e - \mathbf{t}_e\|_2$, where $\mathbf{w}_e \in \mathbb{R}^{n_w}$ is the word embedding of the entity e . This constraint integrates type semantics into entity representation, which brings more information when predicting relations.

Evaluation

The New-York-Times dataset is widely used in DSRE. We use the version presented in (Lin et al. 2016), which links entity mentions to the entities in Freebase. We map entity types in Freebase to fine-grained entity types in (Ling and Weld 2012). Relation and type embeddings are randomly initialized. Following the previous studies, the precision-recall curve is employed to evaluate the proposed model. Specifically, we take top-2000 predictions, discard NA ones, and finally draw the precision-recall curve. In addition, we conduct only held-out evaluation.

We choose 5 models as baselines for comparison, including 3 feature-based ones (i.e., Mintz (Mintz et al. 2009), MultiR (Hoffmann et al. 2011) and MIML-RE (Surdeanu et al. 2012)) and 2 neural network based ones implemented on our own (i.e., PCNN+ATT (Lin et al. 2016) and PCNN+ATT+D (Ji et al. 2017)). For PCNN+ATT+D, only freebase description information is used, but we reach consistent results as in (Ji et al. 2017). In the experiment, we select the relation embedding size n_r from $\{20, \underline{30}, 50\}$, weight factor λ from $\{0, \underline{0.1}, 1\}$. The best configurations are underlined. For other parameters, we select the best parameter reported in (Lin et al. 2016). Specifically, we set word embedding size n_w to 50, position embedding size n_p

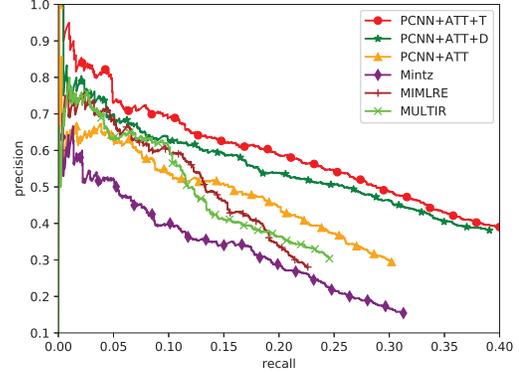


Figure 1: Experimental results in terms of precision-recall curves.

to 5, window size l to 3, convolution kernel number n_c to 230, batch size to 160, and dropout rate to 0.5.

As shown in Figure 1, the curve corresponding to our model, denoted as PCNN+ATT+T, is consistently above the ones of all baseline models. That means our model achieves the best performance in terms of precision and recall. This well justifies the effectiveness of entity type and the modified attention mechanisms. Moreover, PCNN+ATT+T consistently outperforms PCNN+ATT+D, which suggests that, as compared with entity descriptions, entity types are more precise information and thus enable the model to catch more accurate semantics about the entities.

Acknowledgement

This work is supported by National Key R&D Program of China under grant 2016YFB1000902, NSFC under grants 61772501, 61572473, 61572469 and 91646120, and the GF Innovative Research Program.

References

- Hoffmann, R.; Zhang, C.; Ling, X.; Zettlemoyer, L. S.; and Weld, D. S. 2011. Knowledge-based weak supervision for information extraction of overlapping relations. In *ACL-11*, 541–550.
- Ji, G.; Liu, K.; He, S.; and Zhao, J. 2017. Distant supervision for relation extraction with sentence-level attention and entity descriptions. In *AAAI-17*, 3060–3066.
- Lin, Y.; Shen, S.; Liu, Z.; Luan, H.; and Sun, M. 2016. Neural relation extraction with selective attention over instances. In *ACL-16*.
- Ling, X., and Weld, D. S. 2012. Fine-grained entity recognition. In *AAAI-12*.
- Mintz, M.; Bills, S.; Snow, R.; and Jurafsky, D. 2009. Distant supervision for relation extraction without labeled data. In *ACL-09*, 1003–1011.
- Surdeanu, M.; Tibshirani, J.; Nallapati, R.; and Manning, C. D. 2012. Multi-instance multi-label learning for relation extraction. In *EMNLP-CoNLL-12*, 455–465.
- Zeng, D.; Liu, K.; Chen, Y.; and Zhao, J. 2015. Distant supervision for relation extraction via piecewise convolutional neural networks. In *EMNLP-15*, 1753–1762.