Title
Progressive Graph Learning over Pruned Dependency Trees For Relation Extraction

Permalink
https://escholarship.org/uc/item/6r1428g4

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Authors
Wang, Chuang
Ren, Qianqian
Li, Shiyang
et al.

Publication Date
2023

Peer reviewed

Progressive Graph Learning over Pruned Dependency Trees For Relation Extraction

Chuang Wang, Qianqian Ren*, Shiyang Li, Hui Xu*
{renqianqian, xuhui}@hlju.edu.cn
Department of Computer Science and Technology, Heilongjiang University
Harbin, 150080, China

Abstract

Dependency tree is efficient for relation extraction model to exploit relations between words. Recent approaches have achieved promising performance while still suffering from inherent limitations, such as the computation efficiency and flexible pruning strategies. In this paper, we propose a novel relation extraction framework called Progressive Graph Learning over pruned dependency tree (PGLNet). PGLNet constructs a set of graphs by progressively adapting to input sentence. Specially, we implement the model to construct progressive weighted adjacency matrices by learning the relations among graph nodes with multi-head self-attention mechanism. Then, the model takes the learned weights as reference to prune dependency tree in order to preserve useful relevant substructures for the relation extraction while removing irrelevant words. Next, progressive convolution module is designed to encode the relations of entities and followed by relations classification. We evaluate our proposed model using public real-world datasets, experimental results demonstrate that the proposed model achieves state-of-the-art performance with consistency in all datasets. We conclude that the ability of PGLNet to progressively adapt to input data and enable the model with robustness.

Keywords: Machine learning; Neural Networks; Relation Extraction

Introduction

Relation extraction has long been studied as a key technical capability of natural language processing. The issue aims to recognize relations among entities in the text. Relation extraction is of great importance to many real-world applications, such as predicting medicine effectiveness, knowledge graph construction, and question answering(Zhang, Guo, & Lu, 2019). Yet, the redundancy and complexity of relation prediction make the problem especially challenging.

To extract both entities and relations, early works usually adopt pipeline methods. These methods first recognize entities and then predict the relation among entities. However, pipeline methods may not make full use of the correlations between operations of entity recognition and relation extraction. Recent breakthroughs in neural network based techniques enable promising results in feature learning for relation extraction. In particular, neural network based models have shown the ability to improve the scope and accuracy of relation extraction. Recurrent neural network (RNN) or its variants long short-term memory (LSTM) have naturally obtained popularity of for their ability to extract features of sentences(Xu, Mou, Li, et al., 2015). However, the gradient vanishing problem brought by RNNS has made them difficult to obtain satisfactory results in processing long sequence. To overcome such limitations, convolution neural network (CNN) and Self-attention mechanism have been utilized in relation extraction problem(D. Zeng, Liu, Lai, Zhou, & Zhao, 2014). However, CNN could only work in 2D space and can not model the inherent structure of the sentence.

In recent years, Graph Neural Networks(GNNs) have been achieving increasing attentions to extract relations between words for relation extraction. Especially, applying GNN over dependency tree and exploiting long-range relations have obtained promising results(Zhang et al., 2019; Zhang, Qi, & Manning, 2018). Although GNNs and dependency tree based relation extraction has improved the relation extraction performance efficiently. However, there are still some problems that need to further explore:

- Existing improvements in relation extraction performance mainly rely on directly parsing tree structure, which is at the price of poor parallelization and high computation cost. Designing a computation efficient solution for entity and relation extraction is meaningful.
- Existing rule-based pruning methods might eliminate partial important information in the full tree. Designing an efficient pruning strategies to remain important information while ignoring irrelevant information from the dependency trees is a key solution.

To address the above problems, we propose a Progressive Graph Learning based Framework over pruned dependency trees for relation extraction (PGLNet) in this paper. The progressive graph learning module in the proposed model first constructs a progressive weighed graph to learn the relations between words with multi-head attention mechanism, which is referenced to prune the dependency tree. Then, a graph convolution is operated on dependency tree to encode relations.

The contributions of this work are summarized as follows.

1. We propose a novel relation extraction model, Progressive Graph Learning Network(PGLNet). By constructing a progressive weighted graph that can adapt to sentence, the model can efficiently capture the relations of words. In-

*Corresponding author.

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*.

tegrated with dependency tree structure, the model is designed to encode relations.

2. A progressive graph learning module is implemented to model relations of words. Using multi-head self-attention mechanism, the model can learn the relations of words in the sentence regardless of their distance, which not only brings a global receptive fields but also exploits the features in a flexible manner.

3. We conduct extensive experiments on public real-world datasets to validate the performance of our proposed model. The experimental results show that our model can achieve competitive performances compared to the existing baselines.

The rest of the paper is organized as follows. Section II presents related work. Section III elaborates on the proposed model. Section IV presents an extensive experiment evaluation. Section V concludes the paper.

Related Work

As the popularization of deep learning technique, researchers have started to use many large-scale training sets to train the constructed relation extraction models, and classify relations for practical application scenarios. However, in real-world applications, relation extraction faces many challenges, for example, large-scale dataset is very expensive to obtain(Zehe, Viswanathan, Cai, & Knoll, 2016), the long-tail phenomenon of relationships is serious(Guo, 2010) and many factual knowledge is hidden in more complex contexts(Banerjee & Tsioutsiouliklis, 2018). Tang et al.(Tang et al., 2017) proposed a distant supervised relation extraction method. The method automatically generates relational training data by correlating textual entities in the corpus with known triples in an external knowledge base, however, multiple relationships may be contained between pairs of the same entity, which is the noisy data resulting from remote supervision. Zeng et al.(D. Zeng, Kang, Chen, & Zhao, 2015) proposed the Multi-Instance Multi-Label model to solve the problem of noise by introducing the bag mechanism. Lin et al.(Lin, Shen, Liu, Luan, & Sun, 2016) proposed a sentence-level attention approach based on this, assigning weights to each packet to minimise the noise from remote supervision, and Feng et al.(Feng, Huang, Li, Yang, & Zhu, 2018) also achieved better noise reduction by using reinforcement learning to remove noisy data. Han et al.(Han et al., 2018) present the Fewrel large-scale few-times learning dataset, which contains a total of 100 relations with 700 samples each. Snell et al.(Snell, Swersky, & Zemel, 2017) propose a model for learning semantic metrics. Finn et al.(Finn, Abbeel, & Levine, 2017) proposed a Model-agnostic model for learning how to quickly initialise the parameters of a new task. Gao et al.(Gao, Han, Liu, & Sun, 2019) proposed a prototype network with a hybrid attention mechanism, which contains both sentence-level attention and feature-level attention.

Since much of the factual knowledge is hidden in more complex contexts, Yao et al.(Yao et al., 2019) proposed DocRED, a large-scale document-level relation extraction dataset. Based on this, Huang et al.(Huang et al., 2021) designed heuristic rules to select a collection of paths containing information from the entire document by analysing three document relationship extraction benchmark datasets, and these path sets were further combined with BiLSTM to obtain good performance on the benchmark dataset. Zhao et al.(Zhao, Zeng, Xu, & Dai, 2022) proposed ATLOP, a document-level relation extraction model using adaptive thresholding and local context pooling.

Dependency tree is useful for relation extraction model to exploit long-range relations between words(Culotta & Sorensen, 2004). Zhang et al.(Zhang et al., 2018) proposed a dependency-based model that captures non-local syntactic relation that cannot be adequately represented from the surface alone. Due to the richness and diversity of linguistic expressions, a text often contains a large amount of complex information such as modifiers and inflections, and direct input to the model for feature extraction adds a large amount of invalid computation and brings a lot of invalid data to the results. To allow further performance improvements, Xu et al.(Xu, Mou, Ge, Chen, & Zhi, 2015) proposed to apply neural networks on the shortest dependency paths between entities in the book. Zhang et al.(Zhang et al., 2018) proposed the application of graph convolutional networks (GCNs) to prune dependency structure trees with promising results.

Methodology

In this section, we first give the problem definition of relation extraction task, then we describe each component of the proposed PGLNet model. Figure 1 illustrates the overall architecture of PGLNet.

Problem Statement

Definition 1 *(Relation Graph). Given a sentence* $X = \{x_1, x_2, \cdots, x_n\}$*, n is the number of tokens. A relation graph is represented* $G = (V, E, R)$*, where* $V = v_{i}{}_{i=1}^{|V|}$ *denotes the set of vertices (i.e tokens) and* $E = \{e_{ij}\}$ *is the set of edges.* $A \in R^{N \times N}$ *represents the adjacency matrix of the graph. For* $v_i, v_j \in V$*, if* $(v_i, v_j) \in E$*, then* A_{ij} *is 1 otherwise it is 0.*

Let $S = \{s_1, s_2, \cdots, s_m\}$ and $O = \{o_1, s_2, \cdots, o_t\}$ denote the subject entity and object entity respectively. Our goal is to obtain the relation between two entities $r \in R$, where R is the predefined relation set.

PGL Encoder

Recent research has shown BERT has the advantages of trainability and inference speed compared with other training techniques. Thus, we employ a pre-trained BERT (Devlin, Chang, Lee, & Toutanova, 2018) to encode the input sentence. In the model, the sentence is fed into the encoder to obtain the semantic encoding.

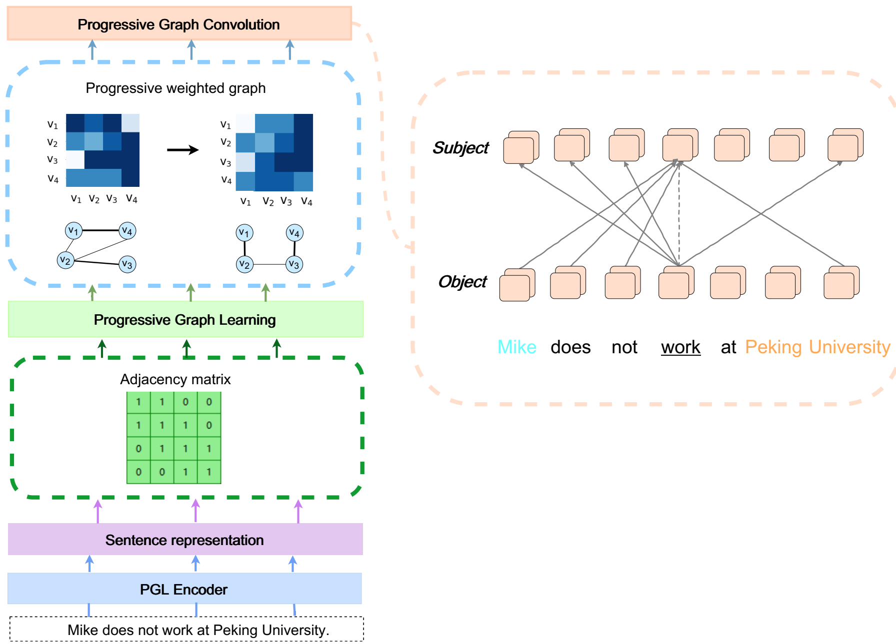

Figure 1: The architecture of PGLNet. The left side illustrates the components, and the right side shows the progressive convolution operation for the word "work". The blue box illustrates the progressive weighted graph. The element in the matrix represents the relation strength between two nodes. The darker the color of the element is, the greater the correlation strength is. In the example, v_1 shows the greatest relation with v_4.

Given a sentence $X = \{x_1, x_2, \cdots, x_n\}$, the encoder outputs the hidden state as following:

$$H = \mathrm{Bert}(x_i) \tag{1}$$

where $H = \{h_1, h_2, \cdots, h_n\}$ is the hidden state vector.

Progressive Graph Learning

Definition 2 *(Progressive Weighted Graph(pw-graph)). Given a sentence X and the corresponding relation graph G, the relations between different nodes under different weighted mechanism are varied. It is intuitive to represent the relations of nodes progressively with different weighted mechanisms. Progressive weighted graph(pw-graph) $\{G^L\}$ is denoted as a a set of graphs, where $G^L = (V, A^L)$. Where A^L is the progressive adjacency matrices set. For $A^{(t)}(t = 1, 2 \cdots, N) \in A^N$, $A^{(t)}_{ij}$ represents the relation strength between v_i and v_j under the t weighted mechanism, and its value is between 0 and 1.*

The blue box of figure 1 describes the idea of progressive weighted graph. Given the nods set $\{v_1, v_2, v_3, v_4\}$, the element on the $i-$th row and $j-$th column denotes the correlation between node v_i and v_j. The darker the color of the grid is, the greater the correlation strength is.

The goal is to impose a higher weight between nodes with greater correlations. In our model, we use multi-head attention mechanism to model the correlations of nodes regardless of their distance. The multi-head attention in our model first maps the queries, keys and values into three feature spaces, namely the query Q, the key K, and the value V. The matrix of these features are formulated as:

$$\begin{aligned} Q &= HW_i^Q \\ K &= HW_i^K \\ V &= HW_i^V \end{aligned} \tag{2}$$

where W_i^Q, W_i^K and W_i^V are learnable linear mapping matrices applied on Q, K and V. Next, we adopt a scaled dot product function to compute the matrix of the self-attention (*single-head attention*) as:

$$MultiHead(Q,K,V) = Concat(A^{(1)}, ..., A^{(i)}, ..., A^{(N)}) \tag{4}$$

$$A^{(i)} = Attention(QW_i^Q, KW_i^K, VW_i^V), \tag{5}$$

where N is the number of attention heads. Multi-head self-attention efficiently learns the correlations of nodes in the sentence regardless of their distance, which not only brings a global receptive fields but also exploits the features in a flexible manner.

Progressive Graph Convolution Module

As the graph convolution network(GCN) has been proven to be an efficient graph structure based model to learn correlations between the node and its neighbors, we thus employ GCN to lean relations between nodes over the graph. Following (Cetoli, Bragaglia, O'Harney, & Sloan, n.d.), GCN is defined as follows:

$$h_v^{k+1} = ReLU\Big(\sum_{u \in \mathcal{N}(v)} (AW^k h_u^k + b^k)\Big) \tag{6}$$

where $\{u, v\} \in V$ are nodes in the graph G. $\mathcal{N}$ is the set of neighbors of node v excludes itself. h_u^k is the embedding of node u at the $k-$ layer. W^k is a weight matrix and b^k is a bias.

Based on formula (6), our progressive convolution module is formulated as:

$$h_v^{k+1} = ReLU\Big(\sum_{u \in \mathcal{N}(v)} (A^{(t)} W^k h_u^k + b_t^k)\Big) \tag{7}$$

where $t \in \{1, 2, \cdots, N\}$, $A^{(t)}$ is the progressive weighted matrix for pw-graph $G^{(t)}$ and b_t^k is the bias for the $t-$th adjacency matrix of graph $G^{(t)}$.

Conducting L-lay progressive convolution operation on word vectors, the hidden representations of each node obtained, which are impacted by its neighbor no more than L edges in the dependency tree. Following by (Zhang et al., 2018), the representation of a sentence is represented as:

$$h_{\text{sent}} = f(\mathbf{h}^{(L)}) = f(\text{PGCN}(\mathbf{h}^{(0)})) \tag{8}$$

where $\mathbf{h}^{(0)}$ is the input of progressive convolution (PGCN) module, f is the max pooling function. In the same way, we obtain the representation of a subject and object(Zhang et al., 2018):

$$h_s = f(\mathbf{h}_{s_1:s_2}^{(L)}) \tag{9}$$

$$h_o = f(\mathbf{h}_{o_1:o_2}^{(L)}) \tag{10}$$

where s_1 and s_2 are two spans in the sentence for a subject entity, respectively, o_1 and o_2 are two spans in the sentence for an object entity.

After obtaining the representation of the sentence, a feedforward neural networks (FFNN)(Zhong, Wang, & Miao, 2019; Zhang et al., 2018) is utilized to obtain the final representation:

$$h_{\text{final}} = \text{FFNN}([h_{\text{sent}}; h_s; h_o]) \tag{11}$$

The final model takes the h_{final} input logistic regression classification among them and makes a relational classification prediction.

Experiments

In this section, we present the experimental evaluation of PGLNet and the existing competing baselines over public datasets. The core target of PGLNet is to implement sentence-level relations extraction and cross-sentence *n*-ary relation extraction task. We report and evaluate the experimental results to verify the effectiveness of our model.

Dataset and Parameters Setting

In the experiments, we evaluate the performance of the proposed model for two kinds relation extraction tasks, namely sentence-level relation extraction and cross-sentence *n*-ary relation extraction. The datasets for each relation extraction task is as follows:

Sentence-level relation extraction For fair and comprehensive comparison, we follow (Zhang et al., 2018) to evaluate our proposed model on two public datasets.

- TACRED(Zhang, Zhong, Chen, Angeli, & Manning, 2017): The dataset includes over 106,000 mention pairs, and it contains 41 relation types and a *no_relation* type.
- SemEval 2010 Task 8(Hendrickx et al., 2010): The SemEval dataset is a popular public dataset. The examples for training and testing are 8000 and 2717 respectively. In particular, it contains 19 relation types.

N-ary relation extraction For cross-sentence *n*-ary relation extraction, the model uses the dataset of Peng et al.(Song, Zhang, Wang, & Gildea, 2018), which contains a total of over 12,000 ternary relationship instances and binary relationship instances extracted from the PubMed corpus. Most of these instances contain multiple sentences, and all instances are grouped into five labels, including "distance or nonresponse", "sensitivity", " response", "resistance" and "none".

During the training process, we tune the hyper parameters to obtain the best performance. In particular, the heads number of attention $N \in \{3, 4, 5, 6\}$, and the number of encoder layers $L \in \{1, 2, 3\}$. For sentence-level relation extraction, $N = 3$ and $L = 2$, while for cross-sentence *n*-ary relation extraction, $N = 2$ and $L = 1$.

Comparison Methods and Metrics The baseline approaches for comparison are as follows.

- Feature-Based(Quirk & Poon, 2017): It refers to relation extraction method based on the shortest dependency path between all entity pairs.
- G-LSTM(Tai, Socher, & Manning, 2015): It is a LSTM based relation extraction method with graph structure.
- T-GCN(X. Zeng et al., 2019): It is a pruned tree-based graph neural network relation extraction.
- AGGCN(Zhang et al., 2019):It is an attention mechanism based graph convolutional networks for relation extraction.

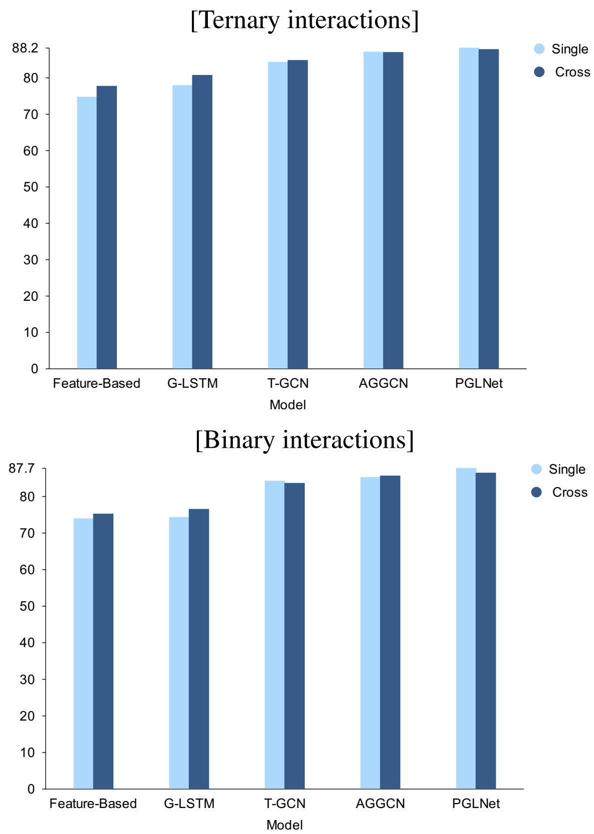

Figure 2: Cross-sentence *n*-ary relation extraction results.

- LR(Zhang et al., 2017): It is a dependency-based logistic regression classifier relation extraction model.
- PA-LSTM (Zhang et al., 2017): It is a position-aware attention based relation extraction model.
- C-GCN(Zhang et al., 2018): It is a contextualized pruned tree graph neural network for relation extraction.

All comparison approaches are evaluated by the following metrics: precision, recall and $F1$. For these metrics, higher values are better.

Results

Figure 2 reports the comparison of performance among PGLNet and baseline approaches on two datasets for cross-sentence *n*-ary relation extraction. As expected, feature based method and LSTM based method work worst. The performance of graph neural network based models is better. It implies the efficiency of graph neural network for relation extraction tasks.

Figure 3 reports the sentence level relation extraction performance of all models over TACRED dataset. We observe that our model improves upon other LSTM and GCN based models in precision, recall and $F1$. The scores of our model are 78.4, 71.3 and 74.7 respectively. Compared with baselines, where the best precision, recall and $F1$ are 69.9, 64.5 and 66.4, respectively, PGLNet produces about 12%, 10% and 16% performance improvement. Our method also improves the recall obviously. It implies that our model can obtain stable performance.

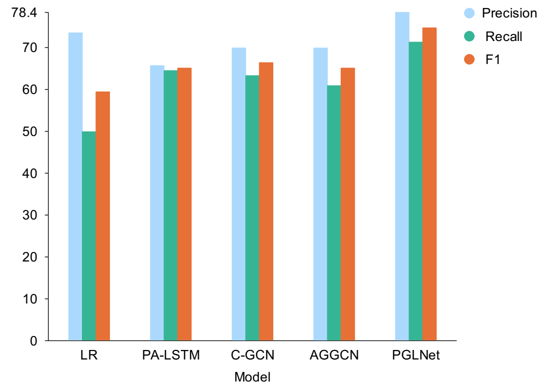

Figure 3: Sentence level relation extraction on TACRED.

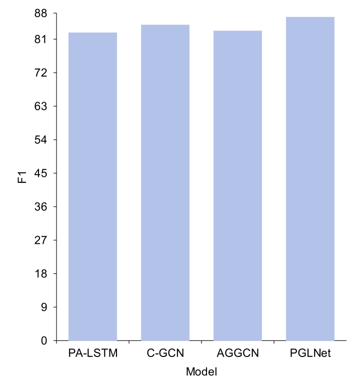

Figure 4: Sentence level relation extraction on SemEval.

Figure 4 shows the relation extraction performance of all models over SemEval dataset. From the figure, we observe that our model outperforms all other LSTM and GCN based models in all metrics. Specifically, our model achieves 87.9 $F1$ score. Compared with the baselines, where the best $F1$ is 84.7 by C-GCN model. PGLNet produces about 2.4% performance improvement.

Acknowledgements

This work is supported by National Training Program of Innovation and Entrepreneurship for Undergraduates under Grant No.202210212186.

Conclusion

In this paper, we propose a relation extraction model, Progressive Graph Learning over pruned dependency trees(PGLNet). The proposed model progressively captures the correlations among nodes in the generated graph structure of sentence. Different from the traditional method of pruning dependency tree, we construct progressive weighted graph for the dependency tree and use multi-head attention mechanism to weight the correlations among nodes in the graph. The experimental results show that the model with a progressive weighted graph can consistently obtain excellent performance in the three datasets used in this study. In the future work, we will work implementing the progressive graph learning framework to solve the the problem of overlapping triples in the same sentence.

References

Banerjee, S., & Tsioutsiouliklis, K. (2018). Relation extraction using multi-encoder lstm network on a distant supervised dataset. In *Ieee international conference on semantic computing* (p. 235-238).

Cetoli, A., Bragaglia, S., O'Harney, A. D., & Sloan, M. (n.d.). Graph convolutional networks for named entity recognition. *Context Scout*.

Culotta, A., & Sorensen, J. (2004). Dependency tree kernels for relation extraction.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding.

Feng, J., Huang, M., Li, Z., Yang, Y., & Zhu, X. (2018). *Reinforcement learning for relation classification from noisy data.*

Finn, C., Abbeel, P., & Levine, S. (2017). Model-agnostic meta-learning for fast adaptation of deep networks.

Gao, T., Han, X., Liu, Z., & Sun, M. (2019). Hybrid attention-based prototypical networks for noisy few-shot relation classification. In (p. 6407-6414).

Guo, J. (2010). The phenomenon of hidden rules in the relationship of central and local governments based on the opportune authorization system. *Zhejiang Social Sciences*.

Han, X., Zhu, H., Yu, P., Wang, Z., Yao, Y., Liu, Z., & Sun, M. (2018). *Fewrel: A large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation.*

Hendrickx, I., Su, N. K., Kozareva, Z., Nakov, P., Séaghdha, D., Padó, S., … Szpakowicz, S. (2010). Semeval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals. In *Meeting of the association for computational linguistics.*

Huang, Q., Zhu, S., Feng, Y., Ye, Y., Lai, Y., & Zhao, D. (2021). Three sentences are all you need: Local path enhanced document relation extraction.

Lin, Y., Shen, S., Liu, Z., Luan, H., & Sun, M. (2016). Neural relation extraction with selective attention over instances. In *Proceedings of the 54th annual meeting of the association for computational linguistics (volume 1: Long papers).*

Quirk, C., & Poon, H. (2017). Distant supervision for relation extraction beyond the sentence boundary. In *Conference of the european chapter of the association for computational linguistics.*

Snell, J., Swersky, K., & Zemel, R. S. (2017). Prototypical networks for few-shot learning.

Song, L., Zhang, Y., Wang, Z., & Gildea, D. (2018). N-ary relation extraction using graph state lstm.

Tai, K. S., Socher, R., & Manning, C. D. (2015). Improved semantic representations from tree-structured long short-term memory networks. *Computer Science*, *5*(1), : 36.

Tang, S., Zhang, J., Zhang, N., Wu, F., Xiao, J., & Zhuang, Y. (2017). [acm press the 40th international acm sigir conference - shinjuku, tokyo, japan (2017.08.07-2017.08.11)] proceedings of the 40th international acm sigir conference on research and development in information retrieval - sigir Ï7 - encore. , 1113-1116.

Xu, Y., Mou, L., Ge, L., Chen, Y., & Zhi, J. (2015). *Classifying relations via long short term memory networks along shortest dependency paths.*

Xu, Y., Mou, L., Li, G., Chen, Y., Peng, H., & Jin, Z. (2015). Classifying relations via long short term memory networks along shortest dependency paths. In *Proceedings of the 2015 conference on empirical methods in natural language processing, EMNLP 2015, lisbon, portugal, september 17-21, 2015* (pp. 1785–1794).

Yao, Y., Ye, D., Li, P., Han, X., Lin, Y., Liu, Z., … Sun, M. (2019). Docred: A large-scale document-level relation extraction dataset.

Zehe, D., Viswanathan, V., Cai, W., & Knoll, A. (2016). Online data extraction for large-scale agent-based simulations. In *the 2016 annual acm conference.*

Zeng, D., Kang, L., Chen, Y., & Zhao, J. (2015). Distant supervision for relation extraction via piecewise convolutional neural networks. In *Conference on empirical methods in natural language processing.*

Zeng, D., Liu, K., Lai, S., Zhou, G., & Zhao, J. (2014). Relation classification via convolutional deep neural network. In J. Hajic & J. Tsujii (Eds.), *COLING 2014, 25th international conference on computational linguistics, proceedings of the conference: Technical papers, august 23-29, 2014, dublin, ireland* (pp. 2335–2344).

Zeng, X., He, S., Zeng, D., Liu, K., Liu, S., & Zhao, J. (2019). Learning the extraction order of multiple relational facts in a sentence with reinforcement learning. In *Empirical methods in natural language processing.*

Zhang, Y., Guo, Z., & Lu, W. (2019). Attention guided graph convolutional networks for relation extraction.

Zhang, Y., Qi, P., & Manning, C. D. (2018). Graph convolution over pruned dependency trees improves relation extraction.

Zhang, Y., Zhong, V., Chen, D., Angeli, G., & Manning, C. D. (2017). Position-aware attention and supervised data improve slot filling. In *Proceedings of the 2017 conference on empirical methods in natural language processing.*

Zhao, C., Zeng, D., Xu, L., & Dai, J. (2022). Document-level relation extraction with context guided mention integration and inter-pair reasoning.

Zhong, P., Wang, D., & Miao, C. (2019). *Knowledge-enriched transformer for emotion detection in textual conversations.*