An Overview of Entity Relation Extraction

Yixuan Xie¹, a
¹Zhuhai No.1 High School, Zhuhai, Guangdong 519000, China;
a180109101337@bitzh.edu.cn

Abstract. Plenty of digital text is generated and shared in nowadays life, which constitutes many unstructured text resources. An automatically information extraction method is highly in demand as a large amount of useful information can be stored and output through it. In this paper, several basic concepts about entity relation extraction are introduced, including word embedding, positional embedding and convolutional neural networks. Several types of relation extraction are also discussed: supervised relation extraction, relation extraction using distant supervision and relation extraction using few-shot approach. In addition, the existing challenges and problems are discussed, like overlapped triples, wrong labeled data and the difficulties that few-shot learning approaches are now facing.

1. Introduction
Relation extraction is a sub-task of natural language processing (NLP) which aims to discover relations between entity pairs Head and Tail given unstructured text data.

Methods based on chronological sequence are divided into two main kinds: method based on features and method based on deep learning. Method based on features requires human to establish some particular features according to the sentences, for example, dependency identifications and semantic meanings. And then one needs to apply these pre-defined features to the conventional machine learning methods such as support vector machines (SVM) to identify the relation between entity pairs. However, this method is both time-consuming and laborious which needs human to establish features manually. With the development of deep learning, methods based on deep learning have been widely used in the relation extracting. This kind of methods can automatically extract the features in the sentences and thus boost the efficiency of relation extraction.

2. Relation Extraction Approach
Sentences-level relation extraction uses labeled data to train models, so as to predict new relations in the given sentences. The major issue faced by such approach is the lack of available training data. Table 1 shows the total number of relations and sentences provided in common relation extraction datasets according to OpenNRE framework [1].

| Dataset            | Relations | Sentences |
|--------------------|-----------|-----------|
| SemEval-2010 Task 8| 9         | 6647      |
| TACRED             | 42        | 21,478    |
| Wiki80             | 80        | 56,000    |
Bag-level relation extraction uses knowledge bag to enhance the efficiency of relation extraction. To build distance supervision datasets (more details in 4.2.), entities are aligned with sentences contained Head and Tail in original texts. Base on this, sentences with same entity pairs are matched together, which then are put into an entity-pair bag. The drawbacks are that the datasets are noisy and instances are uneven distributed.

Document-level relation extraction is built and developed from long distance extraction. This approach can extract relations that can only be deduced by understanding several sentences. This can be vital for some particular domains but many improvements are still needed to enhance the efficiency while maintaining the accuracy at the same time.

3. Basic concepts

3.1. Word embedding

This approach aims to capture the syntactic and semantic information of the word by representing each word as a vector in a low-dimensional array. The process is shown on Table 2.

| Words  | X-axis | Y-axis |
|--------|--------|--------|
| Sam    | 242    | 27     |
| sends  | 9      | 16     |
| an     | 18     | 68     |
| massage| 187    | 42     |
| to     | 29     | 15     |
| Jack   | 88     | 37     |

3.2. Positional embedding

Based on the idea that the closer the word and entity are, the more chance they have to contain useful information, encoding the distance of each word from entities is needed.

3.3. Convolutional neural networks

To make the data more compact to better present the features, the so-called convolution is used to process the data, which not only can reduce the size of data but also extract the features. Imaging there is a 3x3 cube scanner, the procession of convolution is to conduct each element in the scanner with a setting equation:

$$M = X \ast w + b$$  \hspace{1cm} (1)

After this, we calculate the sum of the proceed elements to find M. The process is showing on Figure 1.
4. Types of relation extraction

4.1. Supervised relation extraction
Supervised relation extraction is to extract sentence-level relations. Training by labeled relation data, this approach can identify the relations between entity pairs based on the specific relations they are trained.

4.1.1. Some conventional models
In relation extraction, depicting a global trait is extremely vital. Zeng et al. [2] utilized convolutional neural networks that can work out an over-all representative of the whole context by combining with local features.

Words with the shortest dependency path (SDP) are more informational, in general. Xu et al. [3] proposed an LSTM model that takes advantage of the SDP between entities. They also divided the path into both forward and reverse direction to calculate the reverse SDP.

Because approaches that first match entities rather than identify the relation tend to make more errors, recent model usually extracts entities together with their relations. Wei et al. [4] introduced a hierarchical tagging scheme, that is, first to find the subjects, and then relations with their related subject, which can be an empty set. This method can maximize the number of extracted relation triples.

4.1.2. Pre-trained models
Building a new model usually needs plenty of time, to avoid this, we can use a model that had been used in some similar tasks, the model be used is the so-call pre-trained model.

Bidirectional encoder representation from transformers (BERT), an unsupervised transformer, is one of the commonly used pre-trained language models. Many researchers have tried to adjust and improve it, Table 3 demonstrates some of the works done.

| Tables 3. State F1 scores of relation extraction approaches using Semeval 2010 task-8 as input dataset. |
|-----------------------------------------------|
| **Semeval 2010**                          |
| **CNN based**                             |
| Zeng et al. [2]                            | 82.7 |
| Nguyen and Grishman[5]                     | 82.8 |
| Santos et al. [6]                          | 84.1 |
| Wang et al. [7]                            | 88.0 |
| **RNN based**                              |
| Zhou et al. [8]                            | 84.0 |
| Cai et al. [9]                             | 86.3 |
| **BERT based**                             |
| Wei et al. [4]                             | 87.5 |
| Soares et al. [11]                         | 89.2 |
| Wu and He et al. [12]                      | 89.25|
| Zhao et al. [13]                           | 90.2 |

4.2. Relation extraction using distance supervision
Relation extraction using distance supervision is first introduced by Mintz et al. [10], which is an approach that can provide plenty of labeled data. The model is first given some pre-trained data that contains some triples, sentences that have same head and tail will be deduced to have the same relation and labeled with the assumed relation. Then, these labeled sentences would be bundled together that constitutes a bag with same labeled sentences. However, this approach is imperfect because same head and tail do not guarantee the same relation, which will produce many errors.
4.2.1. Methods to address the issue of incorrect labelling

Distance supervision provides plenty of labeled data and reduces the manual work efficiently, however, it also creates many wrong labeled sentences that affect the accuracy. There are many methods and models designed to solve this problem. Here in this paper, we will list three of them.

Zeng et al. [2] introduced the first neural network model for multi-instances learning with distant supervision, which made use of a piece-wise convolutional neural network to gain relational features. The model assumes that each given sentence contains at least one entity pair that is informative, and it only considers the most expressive sentence in training and predicting. Although it can reduce the wrong labeled problems, it ignores plenty of data which may also be informative. Ji et al. [18] provided a method that can help extract the most appropriate relations, which is useful dealing with ambiguous relations. It builds entity descriptors to contain background information which operate on weighting the sentences. Noticing that relations are in semantic correlation with each other, instead of single individual, Han et al. [17] utilized a hierarchical attention on each bag of information to better extract appropriate relations.

There are many other models and approaches that are designed to solve the problem of wrong labeled sentences. Some of them are presented in Table 4 which also gives a score for each model.

| Tables 4. State mean precision scores of distantly supervised relation classification methods on NYT dataset, ‘Held-out’ and ‘Manual’ indicate that the scores are from held-out and manual evaluation respectively. |
|---|
| **NYT** |
| **Held-out** | Jiang et al. [14] | 72.0 |
| | Lin et al. [15] | 72.2 |
| | Han et al. [16] | 71.0 |
| | Han et al. [17] | 81.6 |
| **Manual** | Zeng et al. [19] | 78.3 |
| | Ji et al. [18] | 81.3 |
| | Wang et al. [20] | 86.9 |

4.3. Relation Extraction Using Few-shot Approach

Few-shot approach learning method aims to use precise and powerful algorithms to compensate for the lack of training data. For instance, when people learning how to distinguish cat and dog, they don’t need hundreds of thousands of data to practice, instead, a few photos will be sufficed.

5. Challenges

There are mainly three types of challenges in relation extraction: overlapping triples, noisy in distance supervision and few-shot learning.

Overlapping triples exist when giving a sentence like: Jackie Chen was born in Hongkong, one of the cities of China. We can find that there are three relations: between Jackie Chen and Hongkong; between Hongkong and China; and between Jackie Chen and China, which are birthplace, municipality and birthplace, respectively. With this in mind, because most relation extraction models assume that each entity only have one relation, a large amount of useful information could be lost in this case.

Also, distance supervision can bring us many labeled data, however these data can be very noisy. Errors are made underlying the assumption that if two entities participate in a relation, any sentence that contains those two entities might express that relation. However, this is not always the case in terms of semantic relations between entity pairs.

Lastly, few-shot learning requires a precise and powerful algorithm. This is especially challenging since the data size is small and the data are noisy. For instance, the model might take the noisy as a part of features and thus might work fine on training data but collapses when facing real life problems.
6. Conclusion
In this paper, we summarized some basic concepts of neutral relation extraction including word embedding, positional embedding and convolutional neutral network. In addition, we introduced three types of neutral relation extraction: supervised relation extraction, relation extraction using distance supervision, and relation extraction using few-shot approach. Existing challenges and problems were also discussed, the major ones are overlapping triples, noisy labeled data, and the lack of training data used to build algorithms for few-shot method.

References
[1] Han, X., Gao, T., Yao, Y., Ye, D., Liu, Z., and Sun, M. (2019). Opennre: An open and extensible toolkit for neural relation extraction. arXiv preprint arXiv:1909.13078.
[2] Zeng, D., Liu, K., Lai, S., Zhou, G., Zhao, J., et al. (2014). Relation classification via convolutional deep neural network.
[3] Xu, Y., Mou, L., Li, G., Chen, Y., Peng, H., and 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, pages 1785–1794.
[4] Wei, Z., Su, J., Wang, Y., Tian, Y., and Chang, Y. (2019). A novel hierarchical binary tagging framework for joint extraction of entities and relations. arXiv preprint arXiv:1909.03227.
[5] Nguyen, T. H. and Grishman, R. (2015). Relation extraction: Perspective from convolutional neural networks. In Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing, pages 39–48, Denver, Colorado. Association for Computational Linguistics.
[6] Santos, C. N. d., Xiang, B., and Zhou, B. (2015). Classifying relations by ranking with convolutional neural networks. arXiv preprint arXiv:1504.06580.
[7] Wang, L., Cao, Z., De Melo, G., and Liu, Z. (2016). Relation classification via multi-level attention cnns. In Proceedings of the 54th annual meeting of the Association for Computational Linguistics (volume 1: long papers), pages 1298–1307.
[8] Zhou, P., Shi, W., Tian, J., Qi, Z., Li, B., Hao, H., and Xu, B. (2016). Attention-based bidirectional long short-term memory networks for relation classification. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pages 207–212.
[9] Cai, R., Zhang, X., and Wang, H. (2016). Bidirectional recurrent convolutional neural network for relation classification. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 756–765.
[10] Mintz, M., Bills, S., Snow, R., and Jurafsky, D. (2009). Distant supervision for relation extraction without labeled data. In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2, pages 1003–1011. Association for Computational Linguistics.
[11] Soares, L. B., Fitzgerald, N., Ling, J., and Kwiatkowski, T. (2019). Matching the blanks: Distributional similarity for relation learning. arXiv preprint arXiv:1906.03158.
[12] Wu, S. and He, Y. (2019). Enriching pre-trained language model with entity information for relation classification. arXiv preprint arXiv:1905.08284.
[13] Zhao, Y., Wan, H., Gao, J., and Lin, Y. (2019). Improving relation classification by entity pair graph. In Asian Conference on Machine Learning, pages 1156–1171.
[14] Jiang, X., Wang, Q., Li, P., and Wang, B. (2016). Relation extraction with multi-instance multi-label convolutional neural networks. In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pages 1471–1480.
[15] Lin, Y., Shen, S., Liu, Z., Luan, H., and 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), pages 2124–2133.
[16] Han, X., Liu, Z., and Sun, M. (2018a). Neural knowledge acquisition via mutual attention between knowledge graph and text. In Thirty-Second AAAI Conference on Artificial Intelligence.
[17] Han, X., Yu, P., Liu, Z., Sun, M., and Li, P. (2018b). Hierarchical relation extraction with coarse-to-fine grained attention. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 2236–2245.

[18] Ji, G., Liu, K., He, S., and Zhao, J. (2017). Distant supervision for relation extraction with sentence-level attention and entity descriptions. In Thirty-First AAAI Conference on Artificial Intelligence.

[19] Zeng, D., Liu, K., Chen, Y., and Zhao, J. (2015). Distant supervision for relation extraction via piecewise convolutional neural networks. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 1753–1762.

[20] Wang, G., Zhang, W., Wang, R., Zhou, Y., Chen, X., Zhang, W., Zhu, H., and Chen, H. (2018a). Label-free distant supervision for relation extraction via knowledge graph embedding. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 2246–2255.