A Survey Deep Learning Based Relation Extraction

Xiaxia Zhang, Yugang Dai*, Tao Jiang
Key Laboratory of China's Ethnic Languages and Information Technology of Ministry of Education, Northwest Minzu University, Lanzhou, Gansu 730000, China

* Corresponding author
e-mail: 49666594@qq.com

Abstract. From human understanding of natural language to machine understanding of natural language, NLP has become the mainstream technology in the world. Extracting useful information from massive information, namely information extraction (IE), such as relation extraction (RE), is one of the important semantic processing tasks. With the explosion of web texts and the emergence of new relations, the margin of human knowledge has increased dramatically, and unstructured text processing has become a problem that needs to be overcome, so people have been working on RE for many years. From the earliest pattern matching to the current popular neural network, RE has made significant progress. After reviewing the development history of relation extraction, this paper analyzes and discusses the development of existing neural network models and pre-trained language models from static technology to dynamic technology and reinforcement learning; finally, combined with the latest development of NLP research technology, the future research direction and trend of entity relation extraction are prospected.

1. Introduction

In recent years, the wave of artificial intelligence caused by deep learning has swept the world. Under the double support of massive data resources brought by the popularization of the Internet and rapidly improved computing power resources, deep learning has deeply influenced all directions of natural language processing and greatly promoted the development of NLP.

Today, the limitations of deep learning are slowly becoming widely recognized. For natural language processing, to achieve a precise and in-depth semantic understanding, it is impossible to solve the essential problems simply by relying on data annotation and computing power input. The massive text data information contains a lot of valuable knowledge, to obtain these useful information, involves a key research problem of artificial intelligence - knowledge acquisition.

At present, these structured knowledge has been widely used in search engine, question answering system and other natural language processing applications. But compared with the rapidly increasing amount of knowledge in the real world, the coverage of knowledge maps is still inadequate. Because of the large scale of knowledge and the high cost of manual tagging, the new knowledge can hardly be completed by human tagging alone. In order to add more world knowledge to the knowledge map as timely and accurately as possible, researchers strive to explore an efficient and automatic way to acquire world knowledge, namely relation extraction technology.
Relation extraction is a classic task, which has been continuously studied in the past 20 years. Feature engineering, nuclear method and graph model have been widely used in it, and some phased achievements have been achieved. However, many problems such as data scale, learning ability, complex context and open relation in practical application reflect the challenges that NLP continues to face. With the advent of the era of deep learning, neural network model brings a new breakthrough for entity relation extraction.

2. Relation extraction framework based on deep learning
The relational extraction framework based on deep learning mainly includes three parts: input preprocessing, data representation and network model learning, as shown in Figure 1. The data preprocessing part takes the whole sentence or the specific range of information in the sentence as the input of the neural network, and uses the natural language processing tool to represent the characteristics of the data. In the data representation part, the input data in the previous step is represented by a low-dimensional vector. In the network model learning part, the network model is designed according to the previous input feature information, and the expression of sentences or sentence packages is obtained. Then, the model is learned through training data.

Figure 1 framework of relation extraction method based on deep learning

2.1. A subsection Relation extraction model based on neural network
There are already many types of neural networks for natural language text sequences, such as Recurrent neural network (RNN, LSTM), convolutional neural networks (CNN) and Transformer, etc. These models can be used for relation extraction through appropriate modifications. Initially, work\cite{1} first proposed the use of CNN to encode sentence semantics for relation classification, which is significantly improved over non-neural network methods. In addition to the simple CNN, a series of improved models are also proposed\cite{2,3,4}. However, the CNN method limits the model's ability to handle remote relations. The information fused through CNN network is often local, so it is difficult to deal with the long-distance relation. Therefore, these methods are limited to a large extent and cannot reach a good level of application. Work\cite{5,6} applies RNN and LSTM to relation extraction. Cai\cite{7} proposes a relation classification model combining cyclic neural network and convolutional neural network on the shortest dependent path, and introduces the bidirectional cyclic convolutional neural network to learn the information on the shortest dependent path from both positive and negative directions. Based on the convolutional neural network, Wang\cite{8} designed the input attention layer and the relational attention layer, and proposed the BiAtt-pooling-CNN model. Zeng\cite{9} first combined deep learning models with multi-example learning to remotely monitor relation extraction. In response to the "noise problem", Lin\cite{10} proposed to use the attention mechanism to assign different weights to the sentences in the sentence package to reduce the noise, so that the information can be more fully used. Zhang\cite{11} first used bidirectional circulation neural network for relation extraction and proposed a BRNN model. Due to the internal structure of RNN itself, it will cause gradient explosion. In order to
solve this problem, Hochreiter[12] adopted gate mechanism and proposed long-term short-term memory neural network. Xu[13] used LSTM network to capture sentence information for relation extraction for the first time. To solve the problem that one-way LSTM network cannot fully extract context information, Zhang[14] proposed BiLSTM model to extract the bidirectional implicit state output of sentences. On the basis of Zhang[11], Zhou[6] proposed Att-BiLSTM model in combination with attention mechanism. Considering from the perspective of entity pair, Qiu[15] proposed an attention mechanism based on entity pair and reduced the computational complexity of two-way LSTM network by constructing a two-way GRU network.

In addition, in order to better extract the features in sentences, various neural networks and other machine learning combinations are modeled for relation extraction. Work[16] proposed to use recursive neural network to model the syntax analysis tree of sentences, and attempted to consider the lexical and syntactic features of sentences while extracting semantic features. This idea has also been further explored by many subsequent works. In order to make better use of the information in the dependency tree, Liu[7] constructed the network model DepNN based on the enhanced dependency subtree. Zhang[18] used path-centric pruning for dependency analytic trees, and then combined with neural network method to obtain sentence vectors to classify relation.

The deep learning model is mainly composed of neural network, tree structure and various mechanisms, as well as the combination of different linguistic knowledge and the use of different structural characteristics of neural network to jointly design the network model to achieve the ideal effect.

2.2. Pre-trained language model

The early pre-training techniques were static, including NNLM, Word2Vec, GloVe and FastText. NNLM, proposed by Bengio[19] in 2003, is a classical model in the early stage of using neural network to implement language model. In 2013, Word2Vec drew lessons from NNLM. Mikolov proposed to use language model to obtain word vectors, which is a simplification and improvement of NNLM, mainly reflected in model structure and training skills. Subsequently, GloVe and FastText were proposed and this static pre-training technique gradually became the most commonly used text representation technique.

The static pre-training technology promotes the rapid development of natural language processing, but the static word vector technology cannot handle the polysemy problem well. Therefore, the pretraining language model provides a series of dynamic pretraining technical schemes. In 2018, ELMo proposed a context-specific text representation that effectively handles polysemy and migrates it downstream for specific tasks. Subsequently, the GPT pre-training language model replaces the two-way LSTM in the ELMo model with Transformer, which uses the decoder structure of Transformer. It is the powerful characterization ability of Transformer that lays a solid foundation for the final model performance. Although the GPT model has achieved good results, it is still a unidirectional language model in essence, and its modeling ability of semantic information is limited. Therefore, a bidirectional thought was provided for later scholars, and a bidirectional pre-training language model BERT model based on Transformer emerged. This was followed by a series of improvements to the BERT model and the release of XLNet by Google. The Emergence of the BERT model has swept across many typical tasks in the field of natural language processing. Some recent approaches have also made relation extraction a downstream task of PLMs. These methods have had some success[20,21].

2.3. New model for relation extraction

In recent years, with the rise of reinforcement learning method, entity relation extraction is given a new way of thinking. Some scholars try to combine the reinforcement learning method with the deep learning method to jointly extract entities and relations. Feng[22] proposed the model of extracting entities and relations based on the joint learning method based on enhanced learning and deep learning in 2017. By designing a reward function, the extracted entity information is transferred to the relation...
extraction and feedback is obtained, so as to extract both entities and relations simultaneously. A remote supervised entity relation extraction method for DEEP reinforcement learning was proposed by Qiu[23]. Based on the mislabeled data, the data were divided into positive and negative sample data by deep reinforcement learning method, and the information of positive example data and negative data was fully utilized.

Generative adversarial network (GAN) is a new method in entity relation extraction, which is widely used in image and visual field at the beginning. GAN are first introduced into text classification tasks. Then Wu[24] introduced the GAN into weakly supervised entity relation extraction in 2017. Qin[25] added the idea of confrontation to the model to classify the relation of the implicit discourse. Qin[26] introduced GAN into remote supervised relation extraction in 2018 to filter out wrong labels and finally achieved the effect of noise reduction.

3. Summary and Prospect
Various neural networks and pre-training techniques are emerging, and many different networks are designed to extract relations. Under the framework of relation extraction based on deep learning, how to combine various technologies for relation extraction is an important direction of future research. The existing relational extraction methods already contain most possible inputs and features. How to directly construct the end-to-end extraction model while introducing domain knowledge and no longer use all kinds of natural language processing tools to reduce the transmission of errors, it is still worth exploring. In the relation extraction task, we often face the problem of small data set relation categories and a small number of samples. The knowledge or patterns learned from different tasks are transferred to the relation extraction task, as well as the migration of features and shared parameters in the same task. The combination of learning and transfer learning has become a problem that we need to study further. Combined with the existing information knowledge, extract more coarse-grained relationships such as cross-sentences and cross-paragraphs, which have strong practical research value. The relation extraction from entity pair to multiple entities needs further study. How to balance the advantages and disadvantages of relation extraction between open domain and specific domain is also worth studying.

Deep learning has become the main technology of relation extraction. All research is for practical applications, so it is necessary to continue to explore and research in the specific application process. This article summarizes some relation extraction methods based on deep learning, and summarizes the related work difficulties and challenges. And on this basis, the future research direction is analyzed and prospected.

Acknowledgments
The work was supported by Ministry of Education Humanities and Social Science Research Project (No.18YJCZH065), the Program of Natural Science Program (Yxm2020104), and National Key Research and Development Plan (No.2017YFB1002103).

References
[1] ChunYang Liu, WenBo Sun, WenHan Chao, Wanxiang Che. Convolution Neural Network for Relation Extraction. The 9th International Conference on Advanced Data Mining and Applications (ADMA 2013).
[2] Thien Huu Nguyen and Ralph Grishman. 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.
[3] Lin Y, Shen S, Liu Z, et al. Neural Relation Extraction with Selective Attention over Instances[C]. meeting of the association for computational linguistics, 2016: 2124-2133.
[4] Jiang X, Wang Q, Li P, et al. Relation Extraction with Multi-instance Multi-label Convolutional Neural Networks.[C]. international conference on computational linguistics, 2016: 1471-1480.
[5] Zhang D, Wang D. Relation Classification via Recurrent Neural Network[J]. arXiv:
Computation and Language, 2015.

[6] Zhou P, Shi W, Tian J, et al. Attention-Based Bidirectional Long Short-Term Memory Networks for Relation Classification[C]. meeting of the association for computational linguistics, 2016: 207-212.

[7] Cai R, Zhang X, Wang H, et al. Bidirectional Recurrent Convolutional Neural Network for Relation Classification[C]. meeting of the association for computational linguistics, 2016: 756-765.

[8] Wang L, Cao Z, De Melo G, et al. Relation Classification via Multi-Level Attention CNNs[C]. meeting of the association for computational linguistics, 2016: 1298-1307.

[9] Zeng D, Liu K, Chen Y, et al. Distant Supervision for Relation Extraction via Piecewise Convolutional Neural Networks[C]. empirical methods in natural language processing, 2015: 1753-1762.

[10] Lin Y, Shen S, Liu Z, et al. Neural Relation Extraction with Selective Attention over Instances[C]. meeting of the association for computational linguistics, 2016: 2124-2133.

[11] Zhang D, Wang D. Relation Classification via Recurrent Neural Network[J]. arXiv: Computation and Language, 2015.

[12] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.

[13] Xu Y, Mou L, Li G, et al. Classifying relations via long short term memory networks along shortest dependency paths[C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015: 1785-1794.

[14] Zhang S, Zheng D, Hu X, et al. Bidirectional long short-term memory networks for relation classification[C]//Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation. 2015: 73-78.

[15] Qin P, Xu W, Guo J. Designing an adaptive attention mechanism for relation classification[C]//2017 International Joint Conference on Neural Networks (IJCNN). IEEE, 2017.

[16] Socher R, Huval B, Manning C D, et al. Semantic Compositionality through Recursive Matrix-Vector Spaces[C]. empirical methods in natural language processing, 2012: 1201-1211.

[17] Liu Y, Wei F, Li S, et al. A Dependency-Based Neural Network for Relation Classification[C]. international joint conference on natural language processing, 2015: 285-290.

[18] Zhang Y, Qi P, Manning C D, et al. Graph Convolution over Pruned Dependency Trees Improves Relation Extraction[C]. empirical methods in natural language processing, 2018: 2205-2215.

[19] Bengio Y, Ducharme R, Vincent P, et al. A neural probabilistic language model[J]. Journal of Machine Learning Research, 2003, 3(6): 1137-1155.

[20] Soares L B, Fitzgerald N, Ling J, et al. Matching the Blanks: Distributional Similarity for Relation Learning[C]. meeting of the association for computational linguistics, 2019: 2895-2905.

[21] Wei Z, Su J, Wang Y, et al. A Novel Hierarchical Binary Tagging Framework for Joint Extraction of Entities and Relations[J]. arXiv: Computation and Language, 2019.

[22] Feng Y, Zhang H, Hao W, et al. Joint Extraction of Entities and Relations Using Reinforcement Learning and Deep Learning.[J]. Computational Intelligence and Neuroscience, 2017, 2017(2017): 7643065-7643065.

[23] Qin P, Xu W, Wang W Y. Robust Distant Supervision Relation Extraction via Deep Reinforcement Learning[J]. 2018.

[24] Wu Y, Bamman D, Russell S, et al. Adversarial Training for Relation Extraction[C]. empirical methods in natural language processing, 2017: 1778-1783.

[25] Qin L, Zhang Z, Zhao H, et al. Adversarial Connective-exploiting Networks for Implicit Discourse Relation Classification[C]. meeting of the association for computational linguistics, 2017: 1006-1017.
[26] Qin P, Xu W, Wang W Y, et al. DSGAN: Generative Adversarial Training for Distant Supervision Relation Extraction.[J]. arXiv: Computation and Language, 2018.