A Review and Outlook for Relation Extraction

Yan Yu
Capital Normal University, Beijing, China
2191002038@cnu.edu.cn

Haolin Sun
Capital Normal University, Beijing, China
sunhaolin05@163.com

Jie Liu
Capital Normal University, Beijing, China
liujxxxy@126.com

ABSTRACT

Relation extraction (RE) aims to identify and determine the specific relation between entity pairs from natural language texts. As a key technology of Natural Language Processing (NLP), RE has broad application prospects in the fields of information retrieval, knowledge graphs, and automatic question answering systems. From pattern matching to neural network, we have made a detailed review of supervised relation extraction methods. In this paper, we focus on the two challenges of current RE: Few-shot learning and dealing with more complex context. We also make a comparative analysis of the existing methods and summarize the technical difficulties. Finally, we look forward to the development of RE.

CCS CONCEPTS

• **Computing methodologies**; • **Artificial intelligence**; • **Natural language processing**; • **Information extraction.**;

KEYWORDS

Relation extraction, Pattern matching, Machine learning, Neural network

ACM Reference Format:
Yan Yu, Haolin Sun, and Jie Liu. 2021. A Review and Outlook for Relation Extraction. In *The 5th International Conference on Computer Science and Application Engineering (CSAE 2021), October 19–21, 2021, Sanya, China.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3487075.3487103

1 INTRODUCTION

With the proliferation of Web texts, the Internet contains a huge amount of information, which implies a large number of relation facts. Relation extraction (RE) obtains relation facts by extracting the semantic relation between two or more entities from the text. The relation between entities is formally expressed as a relation triple <E1, R, E2>, where E1 and E2 refer to the entity type, and R refers to the relation type. For example, in the sentence *"The current president of the United States is Biden"*, it is recognized that the relation between the entities *"the United States"* and *"Biden"* is *"the president of"*, expressed as *(the United States, the president of, Biden).*

RE converts the unstructured information in the text into structured information and stores it in the knowledge base, which provides certain support and helps for subsequent intelligent retrieval and semantic analysis [1] [2]. Researchers use RE technology to extract relations in a unified format from unstructured natural language texts, which facilitates the processing of massive data. RE correlates the extracted entities and the relation between the entity pairs and promotes the automatic construction of the knowledge graph [3] and etc.. Therefore, RE not only has theoretical significance but also has very broad application prospects.

From early pattern matching [4] to machine learning [5], RE has received great attention. With the development of deep learning, neural models are widely used in RE [6, 7]. The neural models effectively improved the defects of traditional annotation tools and achieved good results. These methods bridge the gap between unstructured text and structured knowledge and show their effectiveness on several public benchmarks.

With the advancement of machine learning and deep learning technology [8], the existing RE methods have achieved remarkable results, but most of the methods are only suitable for simple scenarios, and training data also requires a lot of manual annotations. The real world is much more complicated than this simple setting: many long-tail relations cannot provide enough training examples, different requirements for RE models [9] and a large number of relation facts are expressed in multiple sentences. Therefore, in order to build an effective and stable RE system, RE requires further research towards more complex scenarios.

In this paper, we review the development of supervised RE methods (Section 2) and the latest RE methods for more complex scenarios (Section 3). We hope that all these contents could encourage the community to make further exploration and breakthrough towards better RE.

2 BACKGROUND AND EXISTING WORK

Relation extraction is to extract the semantic relation between two or more entities from the text. Relation extraction is closely related to entity extraction. The RE system framework [10] is shown in Figure 1. Named entity recognition refers to the identification of entities with specific meanings in the text, which mainly includes names of persons, organizations, etc. Relation trigger word recognition is to classify the words that trigger the entity relation, identify whether it is a trigger word or a non-trigger word, and determine whether the extracted relation is a positive or negative type. For example: *"The current president of the United States is Biden"*, preprocess the sentence, and then identify the entities *"the United States"* and *"Biden"*. *"President"* as a relation trigger indicates that there may be a certain relation between these two entities. Finally, through the RE model, it is concluded that there is a relation of *"location"* between the two entities.

CSAE 2021, October 19–21, 2021, Sanya, China

ACM ISBN 978-1-4503-8985-3/21/10…$15.00
https://doi.org/10.1145/3487075.3487103

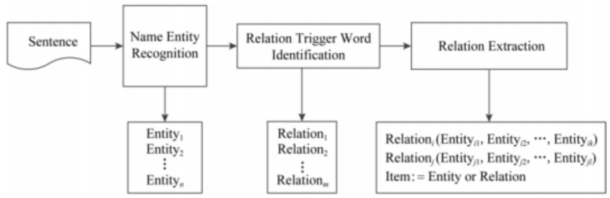

Figure 1: The RE System Framework.

In this section, we introduce pattern matching, statistics relation extraction, and the supervised neural relation extraction models and briefly shows the connections and differences between the various methods.

2.1 Pattern Extraction Models

Early relation extraction methods are mostly based on pattern matching. This type of methods are based on linguistic knowledge, combined with the characteristics of the corpus, and a template is manually written by domain experts to match entities with specific relations from the text. Soderland [11] uses sentence analysis tools to identify syntactic elements in the text, and then automatically constructs pattern rules from these elements. In order to improve the accuracy of pattern matching and higher coverage, Carlson [12] built a system that continuously extracts information from the network to fill the growing structured knowledge base. In order to simplify the pattern writing work of human experts and achieve rapid generalization of new relation types, Shun Zheng [13] used reinforcement learning to extract potential patterns from the NRE model, and proposed a neural pattern diagnostic framework for automatically summarizing and improving high-quality relation patterns from noisy data.

Although the existing pattern matching method reduces the difficulty of human experts, and manual annotation, it alleviates the inevitable label noise problem in distant supervision. However, the possible errors in the automatically constructed model still make the recall rate of the RE system based on pattern matching generally low.

2.2 Statistics Relation Extraction (SRE)

With pattern matching, statistical methods increase coverage and reduce human workload and difficulty. Therefore, Statistical Relation Extraction (SRE) has been extensively studied. Typical SRE methods are divided into two categories: feature-based methods and kernel-based methods.

The core idea of the feature-based method is to extract the features of the relevant sentence from the example sentence, such as grammatical information and lexical information [14], construct feature vectors, and then use the method of calculating the similarity of feature vectors to train the RE model. Therefore, it is very important to extract high-quality features. Kambhatla [15] proposed a method that uses the maximum entropy model to combine different vocabulary, syntactic and semantic features to construct a training model. Zhou et al. [16] used Kambhatla's method to integrate basic grammatical block information, semi-automatically collect features, use Support Vector Machine (SVM) for relation classification, and F1 achieves 55.5%. To further improve the accuracy of RE, Jiang [17] comprehensively considered the complexity of the technology and the features of different dimensions, divided the features into different sub-spaces, and discussed three different representations and features of different complexity.

The model training speed of the feature-based method is faster, but the artificially designed features need to rely on expert knowledge, which affects the applicability of the model. Cristianini [18] proposed a kernel-based relation extraction method. The kernel-based method make full use of the features of the context at a distance, and use the kernel function to calculate the distance between sentences in a high-dimensional space as their similarity, avoiding the direct calculation of feature vectors.

Zelenko [19] is the first to define the kernel function and its algorithm in the text shallow analysis description, and combine the kernel function with SVM and voting perceptron to extract the relation. Compared with the feature-based methods, it proves that the kernel-based method can be developed to mine feature sets for RE. Zhao [20] is the first to distinguish multi-category features and input them into the support vector classifier of the composite kernel function to extract the semantic relations of entities. Other scholars have also proposed methods such as the shortest path dependency tree kernel function method [21] and the combined kernel function [22]. The comparison between feature-based and kernel-based is shown in Table 1

2.3 Neural Relation Extraction Models

Neural relation extraction models introduce neural networks to automatically extract semantic features from text. Compared with SRE models, neural relation extraction methods can effectively capture textual information and generalize to wider range of data. The existing supervised neural relation extraction methods mainly include two categories: pipeline method and joint extraction method. The pipeline method treats entity recognition and relation extraction as two separate processes.The joint extraction method combines entity extraction and relation extraction and optimizes together in a unified model.

2.3.1 Pipeline-based relation extraction. The early pipeline method mainly used Convolutional Neural Networks (CNN) [23] that effectively model local textual patterns,and Recurrent Neural Networks (RNN) [24] that can better handle long sequential data. Santos [25] proposed the CR-CNN model. This method divides the vector of each word into two parts, the word vector and the position vector, and obtains the vector representation of the entire sentence after convolution. The model obtained an F1 value of 84.1% on the SemEval-2010 Task 8 dataset, which is better than the best non-deep learning methods at the time. Wang [26] introduced the attention mechanism to CNN and proposed Multi-level Attention CNN. The introduction of the attention mechanism gives words that reflect relations more weight, and the F1 value reaches 88%. Zhou et al. [27] calculated the weight of each state vector through the attention mechanism based on the word vector generated by BiLSTM, and calculated the sum of the weights of all state vectors, to obtain the sentence vector, which effectively improved the result of relation classification.

Table 1: Comparison of Relation Extraction Methods Based on Surprised Machine Learning

Method	Feature Space	Representation	Key Factor	Speed
Feature-based	Context, Syntactic Tree, et al.	Explicit	Feature Vector	Faster
Kernel-based	Tree kernels, Convolution Kernels, et al.	Implied	Kernel	Slower

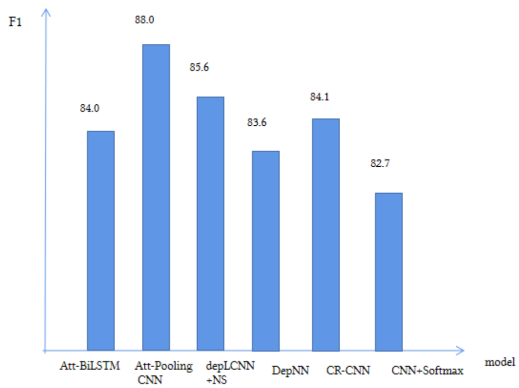

Figure 2: F1 Comparison Based on the Pipeline Model on the SemEval-2010 Task 8 dataset (%).

In addition, many RE models have been proposed, such as depLCNN+NS [28], DepNN [29], CNN+Softmax [30]. Figure 2 is a comparison of the results of some representative pipeline-based relation extraction models on the SemEval-2010 Task 8 dataset.

2.3.2 Joint relation extraction. The pipeline-based extraction method separates entity extraction and relation extraction. Joint extraction is to combine them. There are two methods of joint extraction: parameter sharing method and sequence labeling method.

Parameter sharing methods model entities and relations respectively, and then named entity recognition and relation extraction are mutually dependent on the shared parameters generated during the training process through the shared coding layer, and finally, the best global parameters are obtained through training. Miwa M et al. [31] uses BiLSTM and tree-LSTM structures, and the embedding layer and sequence layer are shared by entity recognition and relation classification tasks while extracting entities and its relation. The joint model obtained an F1 value of 84.4%. After using WordNet as external knowledge, the model obtained a value of 85.6%. Miwa M ignores the long dependency relation between tags when predicting entities. Suncong Zheng [32] uses the LSTM decoder to solve the long dependency problem of tags: after predicting the entity, the entity pair is input into the relation classification module to identify the relation between them. Because the entity pairs that do not have relation are also inputted into the relation classification module, it causes information redundancy and increases the error rate.

Suncong Zheng [33] proposed a novel labeling mechanism to solve the problem of information redundancy, transforming the relation extraction into a sequence labeling problem, extract entities and relations at the same time through sequence labeling. It reduces the influence of invalid entities on the model, and improves the recall and accuracy of RE to 72.4% and 43.7%, respectively. However, this method cannot solve the problem of relation overlap. To solve this problem, Bekoulis et al. [34] regard named entity recognition and relation extraction as a multi-head selection problem, which can represent multiple relations between entities. S Wang [35] proposed a new graph-based joint learning model, which converts the joint extraction task into a directed graph problem, using a conversion-based analysis framework to solve. This method not only can effectively solve the overlapping relation, but also uses the loss function of paranoid weight to strengthen the association between related entities. The F1 value reaches 50.9%. Different from extracting discrete relation labels by two entities, Zhepei Wei et al. [36] proposed the CasRel model increases the F1 value to 89.6%.

In addition, with the emergence of per-trained language models represented by BERT [37], many tasks in the field of NLP have achieved great breakthroughs [38] [39].

3 FUTURE DIRECTIONS AND CHALLENGES

Even though RE has made great progress, it still faces many challenges, such as few-shot learning, and relation extraction for more complex contexts.

3.1 Learning on Long-tailed Data

Large-scale datasets play a key role in training accurate and efficient models. However, in the real world, large-scale datasets often exhibit extreme long-tailed distributions, that is, there are only a small number of samples for multiple relation types in the datasets. As shown in Figure 3, nearly 70% of the data in the NYT data set has a long tail distribution [40] [41]. Under this distribution, the deep neural network was found to perform poorly on the tail class [42].

In recent years, Few-shot Learning (FSL) [43] [44] has been proposed for long-tail distribution. The model can use the generalized knowledge learned from the past data, combined with a small number of training samples of new types of data, to achieve rapid migration learning, and has a certain ability to draw inferences. In order to advance this field, Han [45] first established a large-scale few-shot relation extraction dataset (FewRel).

In order to solve the unbalanced sample distribution, Wang [46] uses the idea of transfer learning to transfer the knowledge obtained from the head class to the tail class. Meta-learning refers to mastering a basic model by learning a series of task sets, rather than individual data. When a new task comes, it can complete the training process with the smallest data. Larochelle [47] proposed a meta-learner model based on LSTM, which learns an algorithm for training another learner's neural network in a few-sample state by learning specific parameter updates and regular initialization of the learner network. Mishra et al. [48] proposed a simple and general

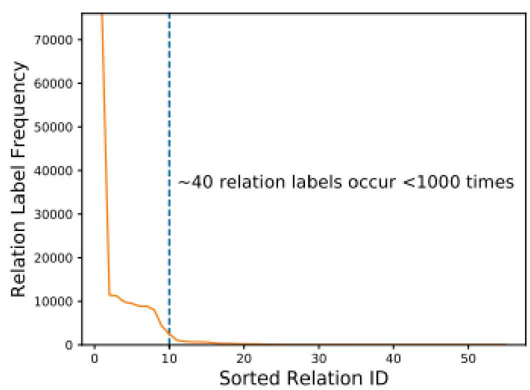

Figure 3: Label Frequency Distribution of Classes without NA in NYT Dataset.

meta-learner architecture that uses a novel combination of temporal convolution and soft attention. The former collects information from past experience, while the latter is used to determine specific information. In the most extensive meta-learning experiment so far, this structure can achieve the most advanced performance.

FSL has made great progress. However, there are still many undiscussed challenges important to its application.

- Existing FSL methods usually use prior knowledge from a single mode (such as image, text or video), and different modes may contain different structures, so considering the application of multi-modal information in the design of FSL method is a very promising direction.
- Few-shot domain adaptation studies how few-shot models are transferred across domains. Some people think that in practical applications, the test domain usually lacks annotations and may be very different from the training domain. Therefore, it is very important to evaluate the transferability of the cross-region few-shot model.

3.2 More Complex Contexts

Relation extraction (RE) can be divided into sentence-level relation extraction [49] and document-level relation extraction [50] according to the form of the text. The sentence-level RE is for a sentence. The document-level RE is an article containing multiple sentences. Figure 4 is a case of document-level relation extraction. In order to infer the relation between Yulia Tymoshenko and the Ukrainian (country of nationality), we need to connect evidence in the document and perform step-by-step reasoning, we can infer that Yulia Tymoshenko is also Ukrainian.

At present, there are mainly two types for document-level relation extraction: sequence-based modeling and graph-based model.

- Sequence-based modeling. Kumutha [51] and Quirk [52] rely on text features extracted from various syntactic structures to connect sentences in documents. The disadvantage of this sequence modeling method is the lack of rich associated information in the document.

Figure 4: Example of Document-Level Relation Extraction.

- Graph-based model. This method focuses on how to build a better document graph to retain more semantic information and better spread information on the graph. Zeng [53] construct an entity graph between sentences, which can use multi-hop paths between entities to infer the correct relation. In order to better use graphs for relational reasoning, Christooulou [54] uses various inter-sentence and intra-sentence dependencies to construct a graph convolutional neural network-based relation extraction model between sentences.

In addition, some researchers have created datasets for document-level relation extraction. At present,the three main datasets used for document-level relation extraction:Chemical Disease Relation(CDR) [55], Gene Disease Association(GDA) [56], Document-Level Relation Extraction Dataset (DocRED) [50].

4 CONCLUSIONS

In this paper, we review the development of relation extraction in a comprehensive and detailed manner, and introduce the three major challenges faced by the current RE system: nested relation extraction, learning on long-tailed data and relation extraction for more complex contexts. Through this paper, we hope to show the progress and problems of existing RE research and encourage more efforts in this area.

ACKNOWLEDGMENTS

This work is supported by National Science and technology innovation 2030 major projects (2020AAA0109700), National Natural Science Foundation of China (62076167), and Beijing Municipal Education Commission-Beijing Natural Fund Joint Funding Project (KZ201910028039).

REFERENCES

[1] Guha R, McCool R, Miller E(2003). Semantic search. Proceedings of the 12th international conference on World Wide Web.ACM:700-709.
[2] DONG X, Gabrilovich E, Heitz G, *et al*(2014). Knowledge vault: A Web-Scale Approach to Probabilistic Knowledge Fusion. Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 601-610.
[3] Singhal Amit (2012). Introducingthe Knowledge Graph: Things, Not Strings. Official Blog(of Google).
[4] Kambhatla N(2004).Combining lexical, Syntactic, and semantic features with maximum entropy models for extracting relations. Proceedings of the ACL 2004 on Interactive poster and demonstration sessions. Association for Computational Linguistics, 22.

[5] Mooney R J, Bunescu R C(2006). Subsequence Kernels for Relation Extraction. Advances in Neural Information Processing Systems, 171-178.
[6] Xu Y, Jia R, Mou L, *et al*(2016). Improved Relation Classification by Deep Recurrent Neural Networks with Data Augmentation[C]// COLING. arXiv.
[7] Liu C Y, Sun W B, Chao W H, *et al*(2013). Convolution Neural Network for Relation Extraction[J].
[8] Javed AR, *et al*(2021). Automated cognitive health assessment in smart homes using machine learning. Sustainable Cities and Society. Feb 1, :102572.
[9] Sharma G, *et al*(2019). A framework for automatic categorization of social data into medical domains. IEEE Transactions on Computational Social Systems. Dec 4, 7(1):129-40.
[10] Yangarber R(1998). NYU: Description of the Proteus/PET System as Used for MUC-7 ST[J]. University of Stuttgart.
[11] Soderland S, Fisher D, Aseltine J, *et al*(1995). CRYSTAL: Inducing a conceptual dictionary. Morgan Kaufmann Publishers Inc.
[12] Carlson A, Betteridge J, Kisiel B, *et al*(2010). Toward an Architecture for Never-Ending Language Learning[C]// Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2010, Atlanta, Georgia, USA, July 11-15. AAAI Press, 2010.
[13] Zheng S, Xu Han, Lin Y, *et al*(2019). DIAG-NRE: A Neural Pattern Diagnosis Framework for Distantly Supervised Neural Relation Extraction[C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics.
[14] Marnee M, Manning C D(2008). Stanford typed dependencies manual.
[15] Kambhatla N(2004). Combining lexical, syntactic, and semantic features with maximum entropy models for information extraction[C]// Annual Meeting of Association of Computational Linguistics.
[16] Zhou G, Su J, Zhang J, *et al*(2005). Exploring Various Knowledge in Relation Extraction[C]// ACL 2005, 43rd Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference, 25-30 June 2005, University of Michigan, USA.
[17] Jing Jiang and ChengXiang Zhai(2007). A systematic exploration of the feature space for relation extraction. In Proceedings of NAACL.
[18] Cristianini N, Shawe-Taylor J(2000). An Introduction to Support Vector Machines and Other Kernel-based Learning Methods: Preface[J].
[19] Zelenko D, Aone C, Richardella A(2003). Kernel Methods for Relation Extraction[J]. Journal of Machine Learning Research, 3(3):1083-1106.
[20] Zhao S, Grishman R(2005). Extracting relations with integrated information using kernel methods[C]// ACL 2005, 43rd Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference, 25-30 June 2005, University of Michigan, USA.
[21] Bunescu R C, Mooney R J(2005). A shortest path dependency kernel for relation extraction[C]// HLT/EMNLP 2005, Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference, 6-8 October 2005, Vancouver, British Columbia, Canada. Association for Computational Linguistics.
[22] Lecun Y A, Bengio Y O(1998). Convolutional networks for images, speech, and time series. MIT Press.
[23] Elman J L(1990). Finding Structure in Time[J]. Cognitive Science, 14(2):179-211.
[24] Santos C, Bing X, Zhou B(2015). Classifying Relations by Ranking with Convolutional Neural Networks[J]. Computer Science, 86(86):132-137.
[25] Wang L, Cao Z, de Melo G, *et al*(2016).Relation Classification via Multi-Level Attention CNNs.Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 1:1298-1307.
[26] Peng Z, Wei S, Tian J, *et al*(2016). Attention-Based Bidirectional Long Short-Term Memory Networks for Relation Classification[C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers).
[27] Xu K, Fe Ng Y, Huang S, *et al*(2015). Semantic Relation Classification via Convolutional Neural Networks with Simple Negative Sampling[J]. Computer ence, 71(7):941-9.
[28] Yang L, Wei F, Li S, *et al*(2015). A Dependency-Based Neural Network for Relation Classification[J]. Computer ence.
[29] Zeng D, Liu K, Lai S, *et al*(2014). Relation classification via convolutional deep neural network.
[30] Miwa M, Bansal M(2016). End-to-End Relation Extraction using LSTMs on Sequences and Tree Structures[C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers).
[31] Zheng S, Hao Y, Lu D, *et al*(2017). Joint Entity and Relation Extraction Based on A Hybrid Neural Network[J]. Neurocomputing, 257(000):1-8.
[32] Zheng S, F Wang, Bao H, *et al*(2017). Joint Extraction of Entities and Relations Based on a Novel Tagging Scheme[J].
[33] Giannis B, Johannes D, Thomas D, *et al.* Joint entity recognition and relation extraction as a multi-head selection problem[J]. Expert Systems with Application, 2018, 114(DEC.):34-45.
[34] Wang S, Yue Z, Che W, *et al*(2018). Joint Extraction of Entities and Relations Based on a Novel Graph Scheme[C]// Twenty-Seventh International Joint Conference on Artificial Intelligence {IJCAI-18.
[35] Wei Z, Su J, Wa Ng Y, *et al*(2020). A Novel Cascade Binary Tagging Framework for Relational Triple Extraction[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics.
[36] Devlin J, Chang M W, Lee K, *et al*(2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[J].
[37] Wu S,Y He(2019). Enriching Pre-trained Language Model with Entity Information for Relation Classification[J]. ACM.
[38] Yi Zhao, Huaiyu Wan, Jianwei Gao, Youfang Lin(2019). Improving Relation Classification by Entity Pair Graph. ACML: 1156-1171.
[39] Shen D, Zhang J, Zhou G, *et al*(2003). Effective adaptation of a hidden markov model-based named entity recognizer for biomedical domain. Association for Computational Linguistics.
[40] Riedel S, Yao L, Mccallum A K(2010). Modeling relations and their mentions without labeled text[C]// Machine Learning and Knowledge Discovery in Databases, European Conference, ECML PKDD 2010, Barcelona, Spain, September 20-24, 2010, Proceedings, Part III. Springer-Verlag.
[41] Kai L, Chen D, Y Li, *et al*(2018). Cooperative Denoising for Distantly Supervised Relation Extraction[C]// The 27th International Conference on Computational Linguistics (COLING 2018).
[42] Mateusz Buda, Atsuto Maki, and Maciej A Mazurowski(2018). A systematic study of the class imbalance problem in convolutional neural networks. Neural Networks, 106:249–259.
[43] Li F F, Member, IEEE, *et al*(2006). One-Shot Learning of Object Categories. IEEE Computer Society.
[44] Fink M(2004). Object Classification from a Single Example Utilizing Class Relevance Metrics[C]// Advances in Neural Information Processing Systems 17 [Neural Information Processing Systems, NIPS 2004, December 13-18, 2004, Vancouver, British Columbia, Canada]. DBLP.
[45] Han X, H Zhu, Yu P, *et al*(2018). FewRel: A Large-Scale Supervised Few-Shot Relation Classification Dataset with State-of-the-Art Evaluation[J].
[46] Wang Y, Ramanan D, Hebert M H(2017). Learning to model the tail. Curran Associates Inc.
[47] Sachin Ravi and Hugo Larochelle(2017). Optimization as a model for few-shot learning. In Proceedings of ICLR.
[48] Nikhil Mishra, Mostafa Rohaninejad, Xi Chen, and Pieter Abbeel(2018). A simple neural attentive metalearner. In Proceedings of ICLR.
[49] Hendrickx I, Kim S N, Kozareva Z, *et al*(2019). SemEval-2010 Task 8: Multi-Way Classification of Semantic Relations Between Pairs of Nominals.
[50] Yao Y, Ye D, Li P, *et al*(2019). DocRED: A Large-Scale Document-Level Relation Extraction Dataset[C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics.
[51] Kumutha Swampillai and Mark Stevenson(2011). Extracting relations within and across sentences. In Proceedings of RANLP, pages 25-32.
[52] Quirk C, Poon H(2017). Distant Supervision for Relation Extraction beyond the Sentence Boundary[C]// Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers.
[53] Wenyuan Zeng, Yankai Lin, Zhiyuan Liu, and Maosong Sun(2017). Incorporating relation paths in neural relation extraction. In Proceedings of EMNLP, pages 1768–1777.
[54] Sahu S K, Christopoulou F, Miwa M, *et al*(2019). Inter-sentence Relation Extraction with Document-level Graph Convolutional Neural Network[J].
[55] Jiao L, Sun Y, Johnson R J, *et al*(2016). BioCreative V CDR task corpus: a resource for chemical disease relation extraction[J]. Database The Journal of Biological Databases and Curation, 2016:baw068.
[56] Y Wu, Luo R, Leung H, *et al*(2019). RENET: A Deep Learning Approach for Extracting Gene-Disease Associations from Literature[J]. Department of Computer Science, The University of Hong Kong, Pokfulam, Hong Kong.