A Survey on Relation Extraction

Meiji Cui[1(✉)], Li Li[1], Zhihong Wang[2], and Mingyu You[1]

[1] Tongji University, Shanghai 201804, China
cui_mj@163.com
[2] Shanghai Hi Knowledge Information Technology Ltd., Shanghai 200433, China

Abstract. Relation extraction, as an important part of information extraction, can be used for many applications such as question-answering and knowledge base population. To thoroughly comprehend relation extraction, the paper reviews it mainly concentrating on its mainstream methods. Besides, open information extraction (OIE), as a different relation extraction paradigm, is introduced as well. Also, we exploit the challenges and directions for relation extraction. We hope the paper will give the overview of relation extraction and help guide the path ahead.

Keywords: Relation extraction · Distant supervision · Neural network · OpenIE
End-to-end extraction

1 Introduction

Relation extraction (RE) is the task of detecting and classifying predefined relationships between entities identified in text. The key objective of the RE is to extract tuples of the form relation <entity1, entity2>. Many efforts have been invested in RE, and there are many varied existing approaches to this problem. At present, the mainstream approaches of RE can be classified as follows: rule-based methods and statistic-based methods, while statistic-based methods include unsupervised approaches, semi-supervised approaches, supervised approaches, distant supervision, neural network. We will discuss these approaches in detail in subsequent sections.

The rest of the paper is organized as follows: Sect. 2 introduces some datasets which are used in the task of RE. The existing mainstream approaches for RE are summarized in Sect. 3. Open information extraction is demonstrated in Sect. 4. Challenges and possible directions are depicted in Sect. 5 and some conclusions are drawn in Sect. 6.

2 Datasets for Relation Extraction

The previous works on RE mainly employed supervised training datasets. These datasets are intensively human annotated. ACE 2004 dataset, ACE 2005 dataset and SemEval-2010 Task 8 are the main supervised datasets for RE.

Since the human-annotated datasets for relation extraction are time-consuming and generally small, Mintz et al. proposed a distant supervision approach for automatically

J. Li et al. (Eds.): CCKS 2017, CCIS 784, pp. 50–58, 2017.
https://doi.org/10.1007/978-981-10-7359-5_6

generating large amounts of training data [1]. Normally, Wikipedia, NY Times corpus [2] and KBP shared tasks [3] are used as evaluation datasets for relation extraction.

3 Mainstream Methods

Since RE plays a vital role in robust knowledge extraction from unstructured texts and serves as an intermediate step in a variety of natural language processing applications, many efforts have been invested in it. Some mainstream RE approaches will be depicted as follows.

3.1 Rule-Based Approaches

The most previous works on RE are based on rules. Rule-based approaches need to predefine rules that describe the structure of the entity mentions. These methods require that the rules builder have a deep understanding of the background and characteristics of the field. Hence, the obvious drawbacks are the huge demand of human participation and poor portability.

3.2 Statistic-Based Approaches

Nowadays, statistic-based approaches can be simply classified into five categories: unsupervised, semi-supervised, supervised, distant supervision and neural network. The details of such approaches will be demonstrated.

Unsupervised Approaches. Unsupervised methods extract strings of words between entities in large amounts of text, and clusters and simplifies these word strings to produce relation-strings [4]. Without demand of annotated data, unsupervised methods can use very large amounts of data as well as extract very large amount of relations. However, since there is no standard form of relations, the output resulting may not be easy to map to relations which is necessary for a particular knowledge base.

Semi-supervised Approaches. The main idea of semi-supervised methods is bootstrapping, thereby the methods are also known as bootstrapping methods. This was first introduced in DIPRE [5], and then extended in Snowball, KnowltAll and TextRunner. Bootstrapping methods typically suffer from semantic drift and poor precision.

Supervised Approaches. Supervised approaches are the most commonly used methods for relation extraction and yield relatively high performance. In the supervised paradigm, relation extraction is treated as a classification task. Supervised method can be simply divided into two types: feature-based methods and kernel-based methods.

In feature-based methods, classification clues (e.g., sequences, parse trees) need to be exploited to convert into feature vectors [6]. The key of feature-based approaches is how to select a suitable feature set when converting structured representations into feature vectors. However, choice of features is usually guided by intuition and experiments.

Kernel-based methods provide a natural alternative to exploit rich representations of input classification clues, such as syntactic parse trees. Kernel defines similarity between objects (e.g. strings, word sequences and parse trees) implicitly in a higher dimensional space. Reference [7] proposed the shortest-path dependency kernels. Lately, a new feature-enriched tree kernel for relation extraction was proposed, by annotating each tree node with a set of discriminate features to refine the syntactic tree representation [8].

However, supervised methods rely on the availability of extensive training data. Additionally, because the relations are labeled on a particular corpus, the resulting classifiers are of poor portability toward other domains. What's more, some NLP tools used in supervised methods are prone to produce some errors, harmful to relation extraction.

Distant Supervision. Distant supervision automatically generates training examples and learns features through aligning free text with KBs such as Freebase, a large semantic database. Thus the method does not need any human intervention, and can extract vast numbers of features from a large amount of data. Table 1 represents the main development process of distant supervision.

Table 1. Distant supervision based methods for RE

Model/System	Main points	Addressed problems
Logistic classifier [1]	DS for RE SISL	Manual annotation
At-Least-One Model [2]	Relaxing SISL to MISL	Predicting relations
MultiR [10]	MIML two approximations	Overlapping relations
MIML-RE [3]	Graphical model	Overlapping relations

The method was first introduced in the context of biological KBs [9]. With the availability of large scale of KBs such as Freebase, Reference [1] extended distant supervision method to any texts. They made the assumption that if two entities participate in a relation, any sentence that contains those two entities might express that relation (SISL). However, the assumption is too strong, treating the relation extraction as a single-instance single-label task. Hence, Reference [2] relaxed the assumption which is called expressed-at-least-once assumption holding with more certainty (e.g. multi instances, MISL). Reference [10] further relaxed the consumption by constructing a graphical model, treating the relation extraction as a multi-instance multi-label problem (e.g. MIML). Reference [3] used a multi-label classifier to denote the latent relation types of all the mentions involving that pair and a set of binary classifiers to decide if the relation holds for the given entity pair.

The above methods ignored the key point that knowledge bases used as distant supervision were highly incomplete. Therefore, they proposed a semi-supervised MIML algorithm modeling the bag-level label noise where entity-pair level labels were either positive or unlabeled. Reference [11] also extended the model to resolve the incompleteness of knowledge bases. However, these methods only exploit one specific kind of indirect supervision knowledge, but ignore many other kinds of useful supervision knowledge. Thus, some distant supervision methods by using additional knowledge to eliminate wrongly labeled instances were proposed [12].

Neural Network. All of the above methods' performance strongly depends on the quality of the extracted features, derived from the existing natural language processing (NLP) tools. As we all know, the errors are inevitably produced during the processing. Hence, how to extract features by reducing the use of existing NLP tools becomes the important research point. The resurgence of neural network (NN) provides the new insight to such problem. Neural network was first applied to relation classification by Reference [13]. Since then, many neural networks, such as RNN, CNN (convolution neural network), LSTM (long short term memory), are exploited to relation extraction.

Zeng et al. employed a CNN-based framework for relation extraction, which was the first attempt [14]. They employed a CNN to extract both lexical and sentence level features, and then concatenated to form the final feature vector to predict the relation between entity pair. Reference [15] used multiple window sizes and pre-trained word embeddings to improve the performance of CNN-based model. To deal with the impact of artificial classes, Reference [16] also adopted a CNN-based model, using a new pair-wise ranking loss function to perform relation classification by ranking (CR-CNN). Xu et al. acquired relation representations from shortest dependency path through a CNN [17]. Reference [18] further combined distant supervision and piecewise CNN (PCNN). Additionally, Jiang et al. considered different sentences contained the entity pair by performing cross-sentence max-pooling to select features from different sentences of interest after extracting each sentence of interest based on a CNN [19]. Reference [20] took attention mechanism to relation extraction by considering more sentences of interest in order to make full use of supervision information. Attention mechanism aimed

Table 2. CNN-based methods for relation extraction

Classifier	Additional features	Main points
MVRNN [13]	Syntactic parsing tree, POS, NER, WordNet	Matrix-vector representation
CNN [14]	PF, WordNet, words around nominals	PF
CR-CNN [16]	PF	Ranking approach
depLCNN + NS [17]	Dependency parsing, WordNet, words around nominals	DSP + Negative sampling strategy
PCNN [18]	PF	Piecewise max pooling
MIMLCNN [19]	PF	Cross-sentence max pooling, multi-label relation modeling
ACNN/APCNN [20]	PF	Attention mechanism
Att-Pooling-CNN [21]	PF	Attention both on entity level and pooling level
Path + Max [22]	PF	Inference chains built on relation path

to automatically increase the weights of invalid instances while reducing the weights of those noisy instances. A CNN with two levels of attention was adopted in Reference [21]. To address the problem that many sentences containing only one target entity while rich in information for RE, Reference [22] originally incorporated relation path into RE. Table 2 shows the CNN-based methods for RE. Since word embeddings are all used in neural network, they don't be considered as additional features.

However, CNN-based method can't capture temporal features, crucial for the performance of RE while the distance between target entities is long. Another neural network called recurrent neural network (RNN) was used in RE for that purpose. Reference [23] proved that RNN particularly performed better in learning relations within long context, attributing to the bi-directional network. The long short term memory (LSTM) architecture extended by RNN was adopted by Reference [24] to model the complete sentence. Xu et al. combined shortest dependency path (SDP) with LSTM units, while LSTM networks were multichannel corresponding to different types of information along the SDP, allowing heterogeneous sources integrated [25]. Similar to the CNN model, attention mechanism is also used in LSTM, which was proposed by Reference [26]. Table 3 represents the methods based on RNN for relation extraction.

Table 3. RNN-based methods for relation extraction

Classifier	Additional features	Main points
RNN [23]	Position indicator	Simple RNN model
BLSTM [24]	PF, POS, NER, WNSYN, DEP, RELATIVE-DEP	BLSTM model
SDP + LSTM [25]	Dependency parsing, POS, GR, WordNet	Combination of SDP and LSTM
Att + BLSTM [26]	PI	Attention mechanism

All above methods treated entity and relation extraction as a pipeline of two separated tasks, i.e., named entity recognition (NER) and relation classification (RC). However, the two sub-tasks are much relevant such that joint model can avoid cascading of errors. Therefore, end-to-end relation extraction method is proposed, which infers to modeling entities and relations in a single model. Reference [27] firstly used a single model to jointly predict entities and relations by an incremental beam-search algorithm in conjunction with structured perceptron. Reference [28] used a history-based structured learning approach to jointly extract entities and relations. Reference [29] utilized bi-directional LSTM-RNNs on both word sequence level and dependency tree substructure information level, while stacking tree-structured LSTM-RNNs on sequential LSTM-RNNs, allowing the two LSTM-RNNs sharing parameters in a model. Reference [30] further proposed a hybrid neural network, while the first layer is a bi-directional LSTM unit, shared by both NER module and RC module. Then the second layer includes a LSTM unit to recognize entities, which can capture tag dependencies and a CNN layer to extract relations.

4 Open Information Extraction (OIE)

Compared with the above methods, open information extraction (OIE), originally introduced by Reference [31] called TextRunner, aims to obtain tuples with highly scalable extraction in portable across domain by identifying a variety of relation phrases and their arguments in arbitrary sentences. Reference [32] presented a novel OIE system called WOE which means Wikipedia-based open extractor. The training data they used were produced by heuristically matching Wikipedia infobox attribute values with corresponding sentences. Additionally, they also proved that dependency parse features performed better than shallow linguistic features. Furthermore, Reference [33] proposed the second generation of OIE to deal with the problem of incoherent extractions and uninformative extractions, by introducing REVERB which implements a general model of verb-based relation phrases expressed as two simple constraints, e.g. syntactic constraint and lexical constraint. Also, they proposed R2A2 adding an argument learning component, ARGLEARNER, which performed significantly better than the previous work. Reference [34] introduced a model called OLLIE, e.g. open language learning for information extraction. Reference [35] proposed an OIE system called ClausIE, aiming at solving the problems of relations expressed via appositions, possessives and participial modiers.

5 Challenges and Directions

Relation extraction, as a key component of information extraction, is still a challenging research field due to the complexity of human language.

For distant supervision, KBs are still far away from the total completed ones. On the one hand, we can complete the KBs as possible, which can be used as powerful distant supervision for relation extraction. On the other hand, some effective inference chains can be considered to extract more relations while no completed entity mentions in KBs.

As we can see, neural network is widely used in relation extraction and has achieved comparable results. We can make progress is to make full use of neural network with little manual features or features extracted by NLP toolkits. Hence, we should fully extract features automatically from all possible instances of interest with neural network. The above mentioned attention mechanism is the promising method for relation extraction together with neural network. We can exploit more levels attention to deal with the multi-instances multi-relations problem.

Regarding that entities extraction and relations extraction influence each other. Joint extraction can be seen as the promising method. We can use more effective or efficient methods, such as method used in Reference [36] for NER, to improve the performance.

6 Conclusion

In this paper, we review the relation extraction in a system. We include the mainstream relation extraction methods, rule-based methods and statistic-based methods respectively. OpenIE, as a different relation extraction paradigm, is introduced as well.

Particularly, we exploit the challenges and directions for relation extraction based on distant supervision and neural network. We can see, relation extraction, as an important part for question-answering, knowledge base population and so on, is a promising and meaningful research field as it is still faced with many challenges.

Acknowledgment. This research is supported by the National Natural Science Foundation of China under Grant No. 51475334, the Science and Technology Project of Shanghai under Grant No. 16dz1206102 and the Fundamental Research Funds for the Central Universities under Grant No.22120170077.

References

1. Mintz, M., Bills, S., Snow, R., et al.: 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, vol. 2, pp. 1003–1011 (2009)
2. Riedel, S., Yao, L., McCallum, A.: Modeling relations and their mentions without labeled text. In: Balcázar, J.L., Bonchi, F., Gionis, A., Sebag, M. (eds.) ECML PKDD 2010. LNCS (LNAI), vol. 6323, pp. 148–163. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15939-8_10
3. Surdeanu, M., Tibshirani, J., Nallapati, R., et al.: Multi-instance multi-label learning for relation extraction. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 455–465 (2012)
4. Yan, Y., Okazaki, N., Matsuo, Y., et al.: Unsupervised relation extraction by mining Wikipedia texts using information from the web. 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, vol. 2, pp. 1021–1029 (2009)
5. Brin, S.: Extracting patterns and relations from the world wide web. In: Atzeni, P., Mendelzon, A., Mecca, G. (eds.) WebDB 1998. LNCS, vol. 1590, pp. 172–183. Springer, Heidelberg (1999). https://doi.org/10.1007/10704656_11
6. Kambhatla N.: Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations. In: Proceedings of the ACL 2004 on Interactive Poster and Demonstration Sessions (2004)
7. Bunescu R.C., Mooney R.J.: A shortest path dependency kernel for relation extraction. In: Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, pp. 724–731 (2005)
8. Sun, L., Han, X.: A feature-enriched tree kernel for relation extraction (2014)
9. Craven, M., Kumlien, J.: Constructing biological knowledge bases by extracting information from text sources. In: Proceedings of ISMB, pp. 77–86 (1999)
10. Hoffmann, R., Zhang, C., Ling, X., et al.: Knowledge-based weak supervision for information extraction of overlapping relations. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, vol. 1, pp. 541–550 (2011)
11. Ritter, A., Zettlemoyer, L., Etzioni, O.: Modeling missing data in distant supervision for information extraction. Trans. Assoc. Comput. Linguist. **1**, 367–378 (2013)
12. Han, X., Sun, L.: Global distant supervision for relation extraction. In: Proceedings of Thirtieth AAAI Conference on Artificial Intelligence (2016)

13. Socher, R., Huval, B., Manning, C.D., et al.: Semantic compositionality through recursive matrix-vector spaces. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 1201–1211 (2012)
14. Zeng, D., Liu, K., Lai, S., et al.: Relation classification via convolutional deep neural network. In: Proceedings of COLING, pp. 2335–2344 (2014)
15. Nguyen, T.H., Grishman, R.: Relation extraction: perspective from convolutional neural networks. In: Proceedings of VS@ HLT-NAACL, pp. 39–48 (2015)
16. Santos, C.N., Xiang, B., Zhou, B.: Classifying relations by ranking with convolutional neural networks. arXiv preprint arXiv:1504.06580 (2015)
17. Xu, K., Feng, Y., Huang, S., et al.: Semantic relation classification via convolutional neural networks with simple negative sampling. arXiv preprint arXiv:1506.07650 (2015)
18. Zeng, D., Liu, K., Chen, Y., et al.: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp. 1753–1762 (2015)
19. Jiang, X., Wang, Q., Li, P., et al.: Relation extraction with multi-instance multi-label convolutional neural networks. In: COLING, pp. 1471–1480 (2016)
20. Lin, Y., Shen, S., Liu, Z., et al.: Neural relation extraction with selective attention over instances. In: Proceedings of ACL, vol. 1 (2016)
21. Wang, L., Cao, Z., de Melo, G., et al.: Relation classification via multi-level attention CNNs. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, vol. 1, pp. 1298–1307 (2016)
22. Zeng, W., Lin, Y., Liu, Z., et al.: Incorporating relation paths in neural relation extraction. arXiv preprint arXiv:1609.07479 (2016)
23. Zhang, D., Wang, D.: Relation classification via recurrent neural network. arXiv preprint arXiv:1508.01006 (2015)
24. Zhang, S., Zheng, D., Hu, X., et al.: Bidirectional long short-term memory networks for relation classification. In: 29th Pacific Asia Conference on Language, Information and Computation, pp. 73–78 (2015)
25. Xu, Y., Mou, L., Li, G., et al.: 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, pp. 1785–1794 (2015)
26. Zhou, P., Shi, W., Tian, J., et al.: Attention-based bidirectional long short-term memory networks for relation classification. In: Proceeding of the 54th Annual Meeting of the Association for Computational Linguistics, pp. 207–212 (2016)
27. Li, Q., Ji, H.: Incremental joint extraction of entity mentions and relations. In: Proceedings of ACL, vol. 1, pp. 402–412 (2014)
28. Miwa, M., Sasaki, Y.: Modeling joint entity and relation extraction with table representation. In: Proceedings of EMNLP, pp. 1858–1869 (2014)
29. Miwa, M., Bansal, M.: End-to-end relation extraction using LSTMs on sequences and tree structures. arXiv preprint arXiv:1601.00770 (2016)
30. Zheng, S., Hao, Y., Lu, D., et al.: Joint entity and relation extraction based on a hybrid neural network. Neurocomputing **257**(27), 59–66 (2017)
31. Yates, A., Cafarella, M., Banko, M., et al.: TextRunner: open information extraction on the web. In: Proceedings of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations, pp. 25–26 (2007)

32. Wu, F., Weld, D.S.: Open information extraction using Wikipedia. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pp. 118–127 (2010)
33. Etzioni, O., Fader, A., Christensen, J., et al.: Open information extraction: the second generation. IJCAI **11**, 3–10 (2011)
34. Schmitz, M., Bart, R., Soderland, S., et al.: Open language learning for information extraction. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 523–534 (2012)
35. Del Corro, L., Gemulla, R.: ClausIE: clause-based open information extraction. In: Proceedings of the 22nd International Conference on World Wide Web, pp. 355–366 (2013)
36. Ma, X., Hovy, E.: End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF. arXiv preprint arXiv:1603.01354 (2016)