Turkish Journal of Electrical Engineering and Computer Sciences

Volume 29 | Number 2 Article 35

1-1-2021

Neural relation extraction: a review

MEHMET AYDAR

ÖZGE BOZAL

FURKAN ÖZBAY

Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Review Article

Turk J Elec Eng & Comp Sci
(2021) 29: 1029 – 1043
© TÜBİTAK
doi:10.3906/elk-2005-119

Neural relation extraction: a review

Mehmet AYDAR[1,2], **Özge BOZAL**[1,3,*], **Furkan ÖZBAY**[1,4]
[1]AI Enablement Department, Huawei Turkey Research and Development Center, İstanbul, Turkey
[2]Research and Development Center, Ford Otosan, İstanbul, Turkey
[3]Department of Computational Science and Engineering, Boğaziçi University, İstanbul, Turkey
[4]Department of Computer Engineering, Faculty of Electrical and Electronics Engineering, Yıldız Technical University, İstanbul, Turkey

Received: 17.05.2020 • **Accepted/Published Online:** 19.11.2020 • **Final Version:** 30.03.2021

Abstract: Neural relation extraction discovers semantic relations between entities from unstructured text using deep learning methods. In this study, we make a clear categorization of the existing relation extraction methods in terms of data expressiveness and data supervision, and present a comprehensive and comparative review. We describe the evaluation methodologies and the datasets used for model assessment. We explicitly state the common challenges in relation extraction task and point out the potential of the pretrained models to solve them. Accordingly, we investigate additional research directions and improvement ideas in this field.

Key words: Neural relation extraction, deep learning, pretrained model, distant supervision

1. Introduction

Never-ending information generation and sharing on the Web provides us with abundant data, most of which constitute the unstructured text sources. To better make sense of and draw associations among those data, we, human beings, use relational facts among the subjects (entities) in the text. For a more comprehensive understanding of specific domains such as bioinformatics, finance, social networking etc., we need computers to process those information.

Representing the information delivered by the text in machine-readable format is essential. One way to do that is to represent entities and their relations in so called triples, which indicate unambiguous facts about entities. A triple (h, r, t) implies that entity h has relation r with another entity t. Knowledge graphs (KG) such as FreeBase [1] and DBpedia [2] are examples of such representations. They are directed and labeled graph structured data which aim to express such explicit semantics and relations of entities in triple form.

Relation extraction from text is a subtask of natural language processing (NLP) which aims to discover relation r between entity pairs h and t given unstructured text data. Earlier work on relation extraction from text heavily relies on kernel-based and feature-based methods [3]. However, recent research studies make use of data-driven deep learning methods, eliminating conventional NLP approaches for relation extraction. [4] explained how the conventional deep learning methods are integrated into relation extraction. The authors of [5] reviewed relation extraction literature focusing on distant supervision. To the best of our knowledge, there is no survey paper in this research field which explained how the state-of-the-art pretrained models are utilized in relation extraction models. In this work, we followed a similar taxonomy to the prior works, such

*Correspondence: ozge.bozal@huawei.com

that we discussed various relation extraction studies in terms of the adopted data supervision. We provided a comprehensive and comparative review in this research field focusing on the challenges and we also proposed possible remedies to solve these common difficulties. Additionally, we made a clear categorization of the relation extraction studies regarding their assumption on data expressiveness, which we believe that prior works are missing.

Section 2 explains various approaches on data expressiveness for relation extraction. In Section 3 neural relation extraction methods are classified and explained in terms of data supervision. Section 4 describes existing challenges in this field of research. In Section 5, benchmark datasets and evaluation metrics are presented. We discuss possible future research directions and improvement ideas in Section 6 and we conclude our survey in Section 7.

2. Relation extraction approaches

In this section, we categorize neural relation extraction methods regarding their assumptions on expressiveness of training instances about the relations.

2.1. Sentence-level relation extraction

In this approach, sentence-based annotated training data is used. Annotation contains sentence-triple alignment information, such that sentences in the training set are labeled with the triples. Once trained, the model's objective is to predict new relations given new entity pairs. However, insufficient amount of training data is a major drawback as labeled data is not always available in real life scenarios. Table 1 shows total number of relations and sentences provided in common relation extraction datasets according to OpenNRE framework [6].

Table 1. Number of relations and sentences provided in common relation extraction datasets.

Dataset	#relations	#sentences
SemEval-2010 Task 8	9	6647
TACRED	42	21,784
Wiki80	80	56,000

2.2. Bag-level relation extraction

Since labeling data in deep learning requires a lot of manual effort, external knowledge bases are used to enhance weakly labeled training set. Knowledge graphs contain information regarding the relations between the entities in the form of triples ($head$, $relation$, $tail$). For creating distant supervision datasets such as NYT, entity pairs in a triple are aligned with the sentences that contain *head* and *tail* entities in the natural text. In this approach, the sentences matched by an entity pair constitute a bag. For this reason these datasets are noisy. Besides that, they are imbalanced, that is, the instances are not evenly distributed across relations.

There are different selection methods to weigh the expressiveness of a bag's instances. One might choose a maximum, average or attention selector which admits the most relevant instance, all instances or weighted average of all instances, respectively [7–9]. More details regarding this approach is given in section 3.2.

2.3. Document-level relation extraction

Sentence-level approach lacks in grasping entity pair relations across a document [10, 11], that is to say, it ignores relations which can be deduced only by understanding several sentences within a document. This can especially be vital for some domains such as drug-side effect relations in pharmaceutical documents [12]. The authors of [11] are the first to address this problem in distantly supervised setups and proposed a document-level graph representation to extract more relations. DocRED [10] provides a benchmark dataset for document-level relation extraction which contains relations that can only be extracted from multiple sentences. As of today, performance of document-level relation extraction methods fall behind the human performance when it comes to cross-sentence reasoning, therefore this approach needs more attention.

3. Types of relation extraction

3.1. Supervised relation extraction

In supervised neural relation extraction from text the sentence-level relation extraction approach is adopted, which requires training data with relational tags. Many of the studies rely on classifying entity pairs according to the particular relations they are assigned to. We list the results of existing methods in Table 2.

Table 2. States F1 scores of supervised relation classification approaches using Semeval 2010 Task-8 as input dataset.

		SemEval 2010
CNN-based	Zeng et al. [14]	82.7%
	Nguyen and Grishman [30]	82.8%
	Santos et al [15]	84.1%
	Wang et al. [31]	88.0%
RNN-based	Zhou et al. [20]	84.0%
	Cai et al. [32]	86.3%
BERT-based	Wei et al. [21]	87.5%
	Soares et al. [26]	89.2%
	Wu and He [25]	89.25%
	Zhao et al. [27]	90.2%

3.1.1. Conventional neural models for relation extraction

Recent research studies focus on extracting relational features with neural networks instead of manual work [13–15]. The authors of [13] proposed a recurrent deep neural network model which admits a compositional vector representation of words and phrases on a parse tree. Each expression is represented by both a vector and a matrix, the former encodes semantic information of an expression and the latter encodes how much it influences the meaning of syntactically neighboring expressions.

In relation classification, drawing global features of relations within a sentence is a crucial task. Accordingly, the authors of [14] utilized convolutional neural networks which can combine local features to obtain a globally representative one. To decrease effects of undesirable artificial classes of relations in prediction, a convolutional deep learning model was introduced in [15], that admits a pairwise ranking loss function and achieve better results than the former model.

TACRED, introduced in [16], is a relation extraction dataset collected based on yearly TAC KBP evaluations. In this work, proposed LSTM sequence model coupled with entity position-aware attention mechanism trained on TACRED outperformed the TAC KBP 2015 slot filling system.

In the study [17], it is claimed that the RNN-based relation extraction models excel the CNN-based models, for the reason that CNNs can capture only the local features, whereas RNNs are capable of learning the long-distance dependency between entities.

An LSTM model proposed in [18] takes advantage of the shortest dependency path (SDP) between entities. They claim that the words along SDP are more informative. Dependency trees are directed graphs, therefore, there is the need of differentiating whether the first entity is related to the second entity or the relation implies the reverse direction. For this purpose, the SPD is divided into two sub-path, each is directed from the entity towards the ancestor node.

Unidirectional LSTM models lack in expressing the complete sequential information of the sentences. The authors of [19] used bidirectional LSTM model (BLSTM) to better represent the sentences.

Meaningful information can be located anywhere in the sentence. Instead of using features from lexical sources such as dependency parsers and named entity recognizers, [20] incorporated attention mechanism to BLSTM network to capture more informative parts of the sentences.

Pipeline approaches which first find the entities than match them with the appropriate relations are prone to error-propagation, namely, the errors in the first part cannot be alleviated in the relation classification part. Recent models study the extraction of entities and their relations, simultaneously. The authors of [21] introduced a hierarchical tagging scheme that maximizes the likelihood of input data and the relational triples. Given a sentence, firstly it finds the subjects, then for each relation r, it tags the appropriate objects, which can also be an empty set. This way, multiple triples can be extracted.

3.1.2. Pretrained language models for relation extraction

Transfer learning is a type of deep learning method used for transferring the ability of a model to another similar or related model. These models are called pretrained models and they save plenty of time and computational power for complex tasks. For NLP tasks there are several widely used pretrained models such as BERT [22], Transformer-XL [23] and OpenAI's GPT-2 [24].

Commonly preferred pretrained language model in relation extraction studies, BERT, is an unsupervised transformer which is trained to predict the next sentence given a sequence of sentences, and also for masked language model. BERT's model captures the contextual information of a word in a given sentence, along with the semantic relation of a sentence to the neighboring sentences in building the whole text. The authors of [25] adjust the pretrained BERT model to handle both sentence and its entities and they achieved better results on SemEval-2010 task 8 dataset than other conventional deep learning methods. The study of [26] aimed to build task agnostic, efficient relation representations from natural text using BERT. The authors achieved better results than previous models on SemEval-2010 task 8 and other models trained on TACRED. [27] achieved the best result on SemEval-2010 task 8. They extract graph topological features on top of BERT embeddings. On the other hand, the study of [21], which also utilized BERT, achieved best results on the distantly supervised NYT dataset.

The pretrained language models overcome the polysemy disambiguation problem which old methodologies suffer from, i.e. they can distinguish between words having multiple lexical meaning, in different context.

In standard transformer structure [28], the attention score $A^{abs}_{i,j}$ between the query i and the key vector j is formulated as

$$A^{abs}_{i,j} = E^T_{x_i} W^T_q W_k E_{x_j} + E^T_{x_i} W^T_q W_k U_j + U^T_i W^T_q W_k E_{x_j} + U^T_i W^T_q W_k U_j,$$

where E_x is the embedding matrix for x, U stands for the absolute positional embedding matrix, W_q and W_k are the weight matrices for the query and the key, respectively.

The contributions of transformer-XL to this attention formulation is two-fold: The first one is the absolute positional embeddings are abandoned and only the relative distances R_{i-j} of the query and the key vectors are attended. The second is that the query is replaced with a trainable parameter u and v to eliminate the bias towards query position for content-based and location-based key vectors, respectively.

$$A_{i,j}^{rel} = E_{x_i}^T W_q^T W_{k,E} E_{x_j} + E_{x_i}^T W_q^T W_{k,R} R_{i-j} + u^T W_{k,E} E_{x_j} + v^T W_{k,R} R_{i-j}$$

Transformer-based language models trained on large and diverse training data such as GPT-2 can handle various NLP tasks. GPT-2 model, takes advantage of byte pair encoding representation which combines the benefits of word-level language models with the generality of byte-level encoding. Using this representation allows evaluating GPT-2 on any dataset regardless of tokenization, vocabulary size and pre-processing. Recently introduced GPT-3 [29] is the scaled up version of GPT-2 in the sense of the hyperparameters and the diverse training data. It proves that high task-agnostic, few-shot performance can be achieved by scaling up the language models.

3.2. Relation extraction using distant supervision

Distant supervision aligns triples in a related knowledge graph with the sentences in input text, in order to automatically generate training data. It assumes the responsibility to determine to what degree a sentence expresses a relation. In other words, sentences are labeled with appropriate relations, and thus an error-prone training set consisting of possibly wrong labeled instances is generated.

The authors of [33] are the first to use this technique. The assumption was that given a triple from the knowledge graph, all sentences that contain the head and tail entities of the triple express the corresponding relation. Unfortunately, this does not reflect the reality. Consider a triple (Bill_gates, Founder, Microsoft) from knowledge base and two sentences below:
“Bill Gates is the co-founder of Microsoft.” and
“The greatest mistake of Bill Gates cost Microsoft $400 billion.”
The first sentence expresses Founder relation, whereas the latter does not. Therefore, the training set including the second sentence is said to be noisy. Subsequent studies in distant supervision assume the same trivial idea of triples in the sentence alignment, however, they differ in feature encoders along with their approach to sentence labeling with appropriate relations and solving the wrong labeling problem caused by distant supervision. Reader can find the results of existing distantly supervised methods in Table 3.

3.2.1. Sentence-triple alignment

Four different frameworks exist for labeling sentences with appropriate relations, which are single-instance single-label (SISL), multiinstance single-label (MISL), single-instance multilabel (SIML) and multiinstance multilabel (MIML) learning. Regarding distant supervision, an instance refers to a sentence in the natural text and a label refers to a relation captured by the knowledge base. Single-instance models assume that a particular relation is derived from only one sentence, while multiinstance approach admits more than one sentence to represent a relation. Multiinstance learning is the bag-level distant supervision approach as explained in subsection 2.2. In single-label methods a particular sentence is relevant to only one relation, whereas in multilabel approach a

Table 3. States mean precision scores of distantly supervised relation classification methods on NYT dataset. ”Held-out” and ”manual” indicates that the scores are from held-out and manual evaluation, respectively.

		NYT
Held-out	Jiang et al. [46]	72.0%
	Lin et al. [9]	72.2%
	Han et al. [39]	71.0%
	Han et al. [38]	81.6%
Manual	Zeng et al. [36]	78.3%
	Ji et al. [37]	81.3%
	Wang et al. [40]	86.9%

sentence can express more than one relation. In this sense, MIML learning framework is more realistic, however, it is necessary to employ efficient ranking and denoising strategies.

3.2.2. Solving wrong-labeling problem with deep learning methods

Distant supervision takes on the annotation burden, however it is obliged to wrong labeling problem. Multiinstance learning [34] aims to relieve problems caused by ambiguously-labeled training data. To denoise training instances of distant supervision, multiinstance learning has become the remedy in relation extraction studies [7, 8, 35, 36]. The work of [7] tried to remedy wrong labeling with an undirected graphical model. [8] focused on multiinstance learning with a probabilistic graphical model. Entity pairs in a corpus do not necessarily imply only one relation. In this direction, the authors of [35] introduced a graphical model with latent variables, which can jointly model the entities and relations in multiinstance multilabel learning fashion.

First neural network model for multiinstance learning with distant supervision was proposed by [36]. The method draws relational features with piece-wise convolutional neural network. The assumption is that, given a relation type, at least one of the input sentences that contain a specific entity pair, is informative, and it considers only the most expressive sentence in the training and the prediction. Obviously, this method neglects a large amount of data which might also be informative on that relation.

In the work of [9], each sentence was ranked using the attention mechanism based on how well it represented a specific relation. Therefore, it suppressed the noisy ones rooted from distant supervision. To better extract the most appropriate relations, especially in ambiguous cases, [37] formulated entity descriptors to include background information which operate on the instances weighted by sentence-level attention.

Relations are not individual tags, on the contrary, they are in semantic correlation with each other. To incorporate the rich information covered by relational correlations, [38] applied a hierarchical attention on each bag of instances.

Another approach is accounting for the information covered by knowledge graphs. The authors of [39] introduced a joint representation learning model for knowledge graph and text, the mutual guidance of which is fed back to the model under an attention mechanism to highlight the significant features of both. To benefit more from the knowledge graphs, [40] proposed a novel distant supervision approach which refuse the hard labels imposed by regular distant supervision methods, rather, they trained the relation classifier directly from KGs with soft labels.

Recent research papers [41–43] confirmed that including high quality human annotation brings significant improvement in relation extraction by alleviating the noise. [43] suggested a reinforcement learning based pattern extraction method to ease pattern-writing work for human experts. The pattern-instance pairs are subject to human annotation to be used in fusing the different labeling methods such as distant supervision and relational patterns.

Based on the complementarity and consistency properties of different languages, [44] combined monolingual and crosslingual attention to take advantage of both language-specific features and the patterns that bear resemblance across languages. They aggregated sentence encodings with weighted attentions to further use in relation prediction. Sequel to this work, the study of [45] investigated the effect of incorporating adversarial training in relation extraction. To alleviate possible incompetency in finding consistent patterns across different languages, this work defined a discriminator which can determine the language of each instance.

3.3. Relation extraction using few-shot approach

Few-shot learning is a learning method, in which in contrast to regular deep learning methods the amount of available training data is small. The assumption is that reliable algorithms can compete the models trained with abundant data. We list some studies related to few-shot learning for relation extraction in Table 4. For the purpose of experimenting few-shot learning algorithms for relation extraction, the authors of [47] provided the "FewRel" dataset. Prototypical networks [48], which admit prototypes rather than classes, were used in few-shot learning scenarios for relation extraction[49]. The model proposed by [26] outperforms human accuracy on few-shot relation matching. [50] introduced an aggregation network model and a matching mechanism which is multilevel.

Table 4. Accuracy scores of few-shot relation classification methods with the best performing configuration on FewRel dataset.

	Best configuration	FewRel
Soares et al. [26]	5 way 1 shot	88.9%
Gao et al. [49]	5 way 10 shot	92.06%
Ye and Ling [50]	5 way 5 shot	92.66%

4. Challenges of relation extraction

Challenges in neural relation extraction regarding available data and existing contextual and structural approaches are presented in this section.

4.1. Overlapping triples

An entity (SingleEntityOverlap) or even an entity pair (EntityPairOverlap) may imply more than one relation in a sentence. Most studies identify entities before the relation classification which assumes that each entity pair is assigned to a single relation (see subsection 3.1). The authors of [51] proposed an end-to-end model which considers relation extraction to be a triple generation problem and applies a copy mechanism to cope with overlapping triples. Another approach proposed by [52] admits a hierarchy of high-level relation indicator detection to mine the relations in a sentence and low-level entity mention extraction to match these relation to the corresponding entities. GraphRel, introduced by [53] is a graph convolutional network based neural model that jointly learns entities and relations. It excels the former methods in solving the overlapping triples problem by incorporating the regional and sequential dependency features of words. Unlike the aforementioned methods,

[21] offered a new formulation for learning relational triples that first identifies subjects, then the relations with a BERT-based subject tagger module, and finally identifies the objects with a relation-specific object module.

4.2. Noise in distant supervision

Relation extraction needs large amount of annotated data. To handle this problem, recent studies incorporate distant supervision which brings its own drawbacks. Distant supervision faces the problem of wrong labeled sentences troubling the training due to the excessive amount of noise. Related studies try to remedy this problem by sentence-level attention [9], hierarchical attention [38], multilingual knowledge extraction [44], joint extraction with knowledge graphs [39] or introducing human annotation to relation extraction [41–43]. Detailed information on these methods is given in subsubsection 3.2.2.

4.3. Few-shot instances

Few-shot based modelling is especially challenging for NLP tasks, since text data is noisy and human annotators tend to be mistaken in language-specific tasks [49]. [47] investigated few-shot learning for relation extraction and provide a dataset for this specific task. [54] improved the former dataset by addressing domain adaptation issues and "none-of-the-above" case which adds extra class to the model. Prototypical networks which assume classification models built on prototypes rather than class labels enable the classifier to identify new classes when only few instances are present for each of those [48, 49].

Class prototypes c_k are typically formulated as the mean of the embeddings of individual instances belonging to the set S_k under an embedding function $f_\varnothing$:

$$c_k = \frac{1}{|S_k|} \sum_{(x_i,y_i)\in S_k} f_\varnothing(X_i).$$

5. Datasets and evaluation

5.1. Datasets

SemEval 2010 Task-8 Dataset [55] contains 2717 sentences, which do not overlap with the 8000 training instances from the version that was released on March 5, 2010 and the instances from SemEval 2007 Task-4. The dataset has 9 distinct relation types.

NYT Dataset (NYT10) [7], was created by aligning relations in Freebase with the sentences in the New York Times Annotated Corpus. Training and test set is generated by splitting the dataset by specific years. Numerous previous work used NYT dataset for relation extraction tasks, however they leverage the dataset as their option.

FewRel [54] is a supervised dataset for relation classification methods utilizing few-shot learning approach. A large set of sentences was first assigned to relations via distant supervision, next, annotated by human experts for denoising. The dataset contains 100 relations, each of which has 700 instances.

Wiki80 is created based on FewRel dataset for few-shot relation extraction tasks, however it is not recognized as a benchmark dataset. It consists of 56,000 samples of 80 distinct relations. The samples were gathered from Wikidata and Wikipedia.

TACRED is the crowd-annotated TAC Relation Extraction Dataset developed by The Stanford NLP Group [16]. TACRED contains 106,264 samples and 41 relation types with "no_relation" label to indicate that there is no relation between entities.

ACE-2005 Multilingual Training Corpus was created for English, Chinese and Arabic languages [56] for the 2005 Automatic Content Extraction (ACE) technology evaluation. The datasets consist of various types of annotated data for entities, relations and events.

WebNLG [57] is another dataset generated for NLP methods. [51] adapted this dataset for relation extraction task. The processed dataset contains 246 relation types, 5019 training, 703 test and 500 validation instances.

5.2. Evaluation

For supervised relation classification tasks the standard precision, recall and F-measure are used for evaluation. Authors usually provide the precision-recall curves for their classification results. In distant supervision, the labels of the text aligned with a knowledge base are not always true labels. For this reason two different evaluation methods are adopted for evaluation of distantly supervised relation classification models: held-out and manual evaluation. In held-out evaluation, the relational facts coming from the knowledge base are considered to be true for the test set, whereas newly predicted relations are treated as false. Since this assumption does not express the reality, some work (see Table 4) conduct manual evaluation which requires human effort. For distantly supervised models, there is no certainty on negative classes, meaning the relation type assigned to a pair of entity by the model might be in fact true, although it does not exist in the knowledge base. Therefore, recall is not a meaningful evaluation metric for distantly supervised models. Instead, P@N (precision when N sentences used for evaluation) and mean precision (mean of all P@N results) metric are preferred. In few-shot learning, there are configurations in the form of m way n shot, m representing the number of relations (classes), and n representing labeled instance number per relation, in that case, sentences. The accuracy performance of the models are evaluated with different data configuration (m, n).

6. Discussion

Neural relation extraction heavily makes use of the research on both deep learning and the semantic web. In this section, we discuss possible research directions regarding relation extraction.

6.1. Question generation and question answering

Neural question generation from text is an emerging research field [58–62]. Question-answering on the knowledge graph is also a well-studied research topic [63]. The joint use of question generation from text and question answering on knowledge bases can discover missing relations between entities. Each appropriate question generated from sentences can be asked to the knowledge graph using question-answering methods that work on the knowledge base. If the system gets a response, it can be added to the natural text. Furthermore, a new triple can be generated based on those question and the corresponding responses. As a result, the training data provide more insight in deep learning methods, as both the natural text and the knowledge graph is enhanced using question generation and question answering methods.

6.2. Possible solutions to improve results of attention mechanism

In relation extraction, attention mechanism usually works best with sentences ranked according to how well they match a specific relation. A similarity metric is involved in matching the relation with the sentence. In this regard, various similarity methods can be explored. Based on the results of the relation extraction, importance weights can be refined, until the algorithm comes up with the optimum results.

Another possible approach is to run event detection algorithms on each of the sentences, paragraphs or documents, and make use of the events in sentence encoding and attention mechanism. Once the main event in a sentence is determined, event-specific triples can be given higher ranking when it comes to alignment of triples to the sentences. Event detection from a knowledge graph is also an emerging research area [64, 65].

6.3. Multilingual bitext mining in machine translation

Machine translation models require a comprehensive training corpus which comprise aligned sentences in different languages. Consequently, sentence alignment is significant in machine translation.

Google's Universal Sentence Encoder (USE) [66] embeds sentences into vectors by maintaining the context of the whole sentence, with a pretrained model available for public use. An extension to USE which supports multilingual sentence encoding has also been published [67]. Facebook also published a similar multilingual sentence encoding study called Laser [68]. These studies make ”bitext mining” possible, which captures similarity scores of sentences even if they are in different languages and matches sentences having close similarity scores.

Relational triples can also be converted to sentences using natural language generation (NLG). Several studies exist for generating natural text from triples [69–73]. Once triples are converted to natural text, the problem reduces to bitext mining, where the pretrained models such as USE and Laser can be utilized. It is also possible to align knowledge bases and natural text in different languages using NLG and a multilingual sentence encoding tool. Since the most of the general-purpose knowledge bases have more content in English, aligning those with their counterparts in other languages can lead to tremendous improvement in distant supervision.

6.4. Paraphrasing

Several paraphrased version of each sentence encoded into a vector space along with the original sentence can help to capture latent words for distant supervision tasks. Thus, higher similarity score can be achieved with related triples from knowledge bases. One possible drawback of this approach might be that while producing abundant data, it simultaneously increases the number of false positives. The remedy can be extending the model with reinforcement learning or adversarial learning.

6.5. Possible solutions for document-level relation extraction

As stated in section 2.3, current methods on document-level relation extraction give poor results comparing to human performance. Enhancing the knowledge base and natural text can be relevant in this scenario, as it assists at finding hidden relation using neural relation prediction methods on the knowledge graph. Besides, external ontologies can be used to enhance the natural text, as ontologies include vocabularies and rule-sets. Locality-sensitive hashing (LSH) methods [74, 75] can also be adopted to quickly determine which ontology aligns well with the input sentence, paragraph or document.

6.6. Integration with few-shot relation extraction

Modern methods on neural relation extraction filter out instances that do not have a sufficient amount of training data. As stated in section 3.3, the few-shot relation extraction is suitable when small amount of training samples is available. In real scenarios, eliminating instances might not be desired. As a result, a joint method of neural relation extraction using distant supervision with a few-shot relation extraction algorithm might be more convenient for real-life scenarios.

7. Conclusion

In this survey, we categorized and summarized neural relation extraction methods regarding the approaches to data expressiveness and data supervision. In addition, we presented datasets for model quality assessment and evaluation metrics. We also explained common challenges and discussed possible remedies to them. To acquire abundant training instances, the latest studies made use of distant supervision. However, it brings noise to data which greatly affects the training of relation extraction models. In addition, there are no explicit negative samples, since the data itself have wrong annotations due to ill-alignment of the unstructured text and the knowledge graph. For that reason, instead of sentence-level approaches in supervised relation extraction, multiinstance approaches are developed for relation extraction with distant supervision. Instead of treating entity recognition and relation extraction separately as in pipeline approaches, later studies adopt end-to-end approaches jointly extracting the entities and relations, which can cope with the problems associated with overlapping triples and long-tail relations. Supervised approaches are not to be abandoned. Also, few-shot learning for relation extraction is a research area that has still room for improvement. Pretrained language models have broken fresh ground in natural language processing. Indeed, incorporating the pretrained language models in supervised relation extraction have made significant improvements in comparison to using conventional deep learning methods. Those language models such as BERT has multilanguage support. Also there are community driven projects to extend pretrained models in other languages. For instance, BERTurk [1] extends Bert for Turkish language. It would be interesting to apply these models for neural relation extraction. As a future work, we are planning to study relation extraction from text for Turkish language.

References

[1] Bollacker K, Evans C, Paritosh P, Sturge T, Taylor J. Freebase: a collaboratively created graph database for structuring human knowledge. In: Proceedings of the 2008 ACM SIGMOD international conference on Management of data (SIGMOD '08); Vancouver, Canada; 2008. pp. 1247-1250.

[2] Auer S, Bizer C, Kobilarov G, Lehmann J, Cyganiak R et al. Dbpedia: a nucleus for a web of open data. In: The Semantic Web; Berlin, Heidelberg; 2007. pp. 722-735.

[3] Pawar S, Palshikar GK, Bhattacharyya P. Relation extraction: a survey. arXiv 2017; 1712.05191.

[4] Kumar S. A survey of deep learning methods for relation extraction. arXiv 2017; 1705.03645.

[5] Smirnova A, Cudré-Mauroux P. Relation extraction using distant supervision: a survey. ACM Computing Surveys (CSUR) 2018; 51 (5): 1-35. doi: 10.1145/3241741

[6] Han X, Gao T, Yao Y, Ye D, Liu Z et al. OpenNRE: an open and extensible toolkit for neural relation extraction. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): System Demonstrations; Hong Kong, China; 2019. pp. 169-174.

[7] Riedel S, Yao L, McCallum A. Modeling relations and their mentions without labeled text. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases; Berlin, Heidelberg; 2010. pp. 148-163.

[8] Hoffmann R, Zhang C, Ling X, Zettlemoyer L, Weld DS. Knowledge-based weak supervision for information extraction of overlapping relations. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies; Portland, Oregon, USA; 2011. pp. 541–550.

[1] Schweter S (2020). Berturk-bert models for Turkish [online]. Website https://github.com/stefan-it/turkish-bert [accessed 29 August 2020].

[9] Lin Y, Shen S, Liu Z, Luan H, Sun M. 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); Berlin, Germany; 2016. pp. 2124–2133.

[10] Yao Y, Ye D, Li P, Han X, Lin Y et al. DocRED: A large-scale document-level relation extraction dataset. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics; Florence, Italy; 2019. pp. 764-777.

[11] Quirk C, Poon H. Distant supervision for relation extraction beyond the sentence boundary. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers; Valencia, Spain; 2017. pp. 1171-1182.

[12] Verga P, Strubell E, McCallum A. Simultaneously self-attending to all mentions for full- abstract biological relation extraction. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers); New Orleans, LA, USA; 2018. pp. 872–884.

[13] Socher R, Huval B, Manning CD, Ng AY. 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; Jeju Island, Korea; 2012. pp. 1201-1211.

[14] Zeng D, Liu K, Lai S, Zhou G, Zhao J. Relation classification via convolutional deep neural network. In: Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers; Dublin, Ireland; 2014. pp. 2335-2344.

[15] Santos CN, Xiang B, Zhou B. Classifying relations by ranking with convolutional neural networks. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers); Beijing, China; 2015. pp. 626-634.

[16] Zhang Y, Zhong V, Chen D, Angeli G, Manning CD. Position-aware attention and supervised data improve slot filling. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing; Copenhagen, Denmark; 2017. pp. 35-45.

[17] Zhang D, Wang D. Relation classification via recurrent neural network. arXiv 2015; 1508.01006.

[18] Xu Y, Mou L, Li G, Chen Y, Peng H 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; Lisbon, Portugal; 2015. pp. 1785-1794.

[19] Zhang S, Zheng D, Hu X, Yang M. Bidirectional long short-term memory networks for relation classification. In: Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation; Shanghai, China; 2015. pp. 73–78.

[20] Zhou P, Shi W, Tian J, Qi Z, Li B et al. 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); Berlin, Germany; 2016. pp. 207–212.

[21] Wei Z, Su J, Wang Y, Tian Y, Chang Y. A novel hierarchical binary tagging framework for joint extraction of entities and relations. arXiv 2019; 1909.03227.

[22] Devlin J, Chang MW, Lee K, Toutanova K. Bert: pre-training of deep bidirectional transformers for language understanding. arXiv 2018; 1810.04805.

[23] Dai Z, Yang Z, Yang Y, Carbonell JG, Le Q et al. Transformer-XL: attentive language models beyond a fixed-length context. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics; Florence, Italy; 2019. pp. 2978-2988.

[24] Radford A, Wu J, Child R, Luan D, Amodei D et al. Language models are unsupervised multitask learners. San Francisco, CA, USA: OpenAI Blog, 2019.

[25] Wu S, He Y. Enriching pre-trained language model with entity information for relation classification. In: Proceedings of the 28th ACM International Conference on Information and Knowledge Management; Beijing, China; 2019. pp. 2361-2364.

[26] Soares LB, FitzGerald N, Ling J, Kwiatkowski T. Matching the blanks: distributional similarity for relation learning. arXiv 2019; 1906.03158.

[27] Zhao Y, Wan H, Gao J, Lin Y. Improving relation classification by entity pair graph. In: Asian Conference on Machine Learning; Nagoya, Japan; 2019. pp. 1156-1171.

[28] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L et al. Attention is all you need. In: Advances in neural information processing systems; Long Beach, CA, USA; 2017. pp. 5998-6008.

[29] Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J et al. Language models are few-shot learners. arXiv 2020; 2005.14165.

[30] Nguyen TH, Grishman R. Relation Extraction: Perspective from Convolutional Neural Networks. In: 1st Workshop on Vector Space Modeling for Natural Language Processing; Denver, CO, USA; 2015. pp. 39-48.

[31] Wang L, Cao Z, De Melo G, Liu Z. Relation classification via multi-level attention cnns. In: 54th annual meeting of the Association for Computational Linguistics (Volume 1: Long Papers); Berlin, Germany; 2016. pp. 1298-1307.

[32] Cai R, Zhang X, Wang H. Bidirectional recurrent convolutional neural network for relation classification. In: 54th annual meeting of the Association for Computational Linguistics (Volume 1: Long Papers); Berlin, Germany; 2016. pp. 756-765.

[33] Mintz M, Bills S, Snow R, Jurafsky D. Distant supervision for relation extraction without labeled data. In: 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; Suntec, Singapore; 2009. pp. 1003-1011.

[34] Dietterich TG, Lathrop RH, Lozano-Pérez T. Solving the multiple instance problem with axis-parallel rectangles. Artificial Intelligence 1997; 89 (1-2): 31-71. doi: 10.1016/S0004-3702(96)00034-3

[35] Surdeanu M, Tibshirani J, Nallapati R, Manning CD. Multi-instance multi-label learning for relation extraction. In: Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning; Jeju Island, Korea; 2012. pp. 455-465.

[36] Zeng D, Liu K, Chen Y, Zhao J. Distant supervision for relation extraction via piecewise convolutional neural networks. In: Empirical Methods in Natural Language Processing; Lisbon, Portugal; 2015. pp. 1753-1762.

[37] Ji G, Liu K, He S, Zhao J. Distant supervision for relation extraction with sentence-level attention and entity descriptions. In: AAAI Conference on Artificial Intelligence; San Francisco, CA, USA; 2017. pp. 3060-3066.

[38] Han X, Yu P, Liu Z, Sun M, Li P. Hierarchical relation extraction with coarse-to-fine grained attention. In: Empirical Methods in Natural Language Processing; Brussels, Belgium; 2018. pp. 2236-2245.

[39] Han X, Liu Z, Sun M. Neural knowledge acquisition via mutual attention between knowledge graph and text. In: AAAI Conference on Artificial Intelligence; New Orleans, LA, USA; 2018. pp. 4832-4839.

[40] Wang G, Zhang W, Wang R, Zhou Y, Chen X et al. Label-free distant supervision for relation extraction via knowledge graph embedding. In: Empirical Methods in Natural Language Processing; Brussels, Belgium; 2018. pp. 2246-2255.

[41] Liu A, Soderland S, Bragg J, Lin CH, Ling X et al. Effective crowd annotation for relation extraction. In: North American Chapter of the Association for Computational Linguistics: Human Language Technologies; San Diego, CA, USA; 2016. pp. 897-906.

[42] Ratner AJ, De Sa CM, Wu S, Selsam D, Ré C et al. Data programming: creating large training sets, quickly. In: Advances in Neural Information Processing Systems; Barcelona, Spain; 2016. pp. 3567-3575.

[43] Zheng S, Han X, Lin Y, Yu P, Chen L et al. DIAG-NRE: a neural pattern diagnosis framework for distantly supervised neural relation extraction. In: 57th Annual Meeting of the Association for Computational Linguistics; Florence, Italy; 2019. pp. 1419-1429.

[44] Lin Y, Liu Z, Sun M. Neural relation extraction with multi-lingual attention. In: 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers); Vancouver, Canada; 2017. pp. 34-43.

[45] Wang X, Han X, Lin Y, Liu Z, Sun M. Adversarial multi-lingual neural relation extraction. In: 27th International Conference on Computational Linguistics; Santa Fe, NM, USA; 2018. pp. 1156-1166.

[46] Jiang X, Wang Q, Li P, Wang B. Relation extraction with multi-instance multi-label convolutional neural networks. In: COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers; Osaka, Japan; 2016. pp. 1471-1480.

[47] Han X, Zhu H, Yu P, Wang Z, Yao Y et al. FewRel: a large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation. In: Empirical Methods in Natural Language Processing; Brussels, Belgium; 2018. pp. 4803-4809.

[48] Snell J, Swersky K, Zemel R. Prototypical networks for few-shot learning. In: Advances in Neural Information Processing Systems; Long Beach, CA, USA; 2017. pp. 4077-4087.

[49] Gao T, Han X, Liu Z, Sun M. Hybrid attention-based prototypical networks for noisy few-shot relation classification. In: AAAI Conference on Artificial Intelligence; Honolulu, HI, USA; 2019. pp. 6407-6414.

[50] Ye ZX, Ling ZH. Multi-level matching and aggregation network for few-shot relation classification. arXiv 2019; 1906.06678.

[51] Zeng X, Zeng D, He S, Liu K, Zhao J et al. Extracting relational facts by an end-to-end neural model with copy mechanism. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics; Melbourne, Australia; 2018. pp. 506-514.

[52] Takanobu R, Zhang T, Liu J, Huang M. A hierarchical framework for relation extraction with reinforcement learning. In: Proceedings of the AAAI Conference on Artificial Intelligence; Honolulu, Hawaii, USA; 2019. pp. 7072-7079.

[53] Fu TJ, Li PH, Ma WY. Graphrel: modeling text as relational graphs for joint entity and relation extraction. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics; Florence, Italy; 2019. pp. 1409-1418.

[54] Gao T, Han X, Zhu H, Liu Z, Li P et al. FewRel 2.0: towards more challenging few-shot relation classification. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing; Hong Kong, China; 2019. pp. 6249-6254.

[55] Hendrickx I, Kim SN, Kozareva Z, Nakov P, Séaghdha DÓ et al. SemEval-2010 Task 8: multi-way classification of semantic relations between pairs of nominals. In: Proceedings of the 5th International Workshop on Semantic Evaluation 2010; Boulder, CO, USA; 2010. pp. 94-99.

[56] Walker C, Strassel S, Medero J, Maeda K. ACE 2005 multilingual training corpus. Linguistic Data Consortium 2006; 57: 45. doi: 10.35111/mwxc-vh88

[57] Gardent C, Shimorina A, Narayan S, Perez-Beltrachini L. The WebNLG challenge: generating text from RDF data. In: Proceedings of the 10th International Conference on Natural Language Generation; Santiago de Compostela, Spain; 2017. pp. 124-133.

[58] Zhou Q, Yang N, Wei F, Tan C, Bao H et al. Neural question generation from text: a preliminary study. In: National CCF Conference on Natural Language Processing and Chinese Computing; Dalian, China; 2017. pp. 662-671.

[59] Xingdi Y, Tong W, Caglar G, Alessandro S, Philip B et al. Machine comprehension by text-to-text neural question generation. In: Proceedings of the 2nd Workshop on Representation Learning for NLP; Vancouver, Canada; 2017. pp. 15-25.

[60] Kupiec JM. Method for extracting from a text corpus answers to questions stated in natural language by using linguistic analysis and hypothesis generation. May 21 1996. US Patent 5,519,608.

[61] Du X, Shao J, Cardie C. Learning to ask: neural question generation for reading comprehension. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics; Vancouver, Canada; 2017. pp. 1342-1352.

[62] Wolfe JH. Automatic question generation from text - an aid to independent study. In: Proceedings of the ACM SIGCSE-SIGCUE Technical Symposium on Computer Science and Education; New York, NY, USA; 1976. pp. 104–112.

[63] Pranav R, Jian Z, Konstantin L, Percy L. SQuAD: 100,000+ questions for machine comprehension of text. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing; Austin, Texas, USA; 2016. pp. 2383-2392.

[64] Xiaozhi W, Xu H, Zhiyuan L, Maosong S, Peng L. Adversarial training for weakly supervised event detection. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; Minneapolis, Minnesota, USA; 2019. pp. 998-1008.

[65] Xiaozhi W, Ziqi W, Xu H, Zhiyuan L, Juanzi L et al. HMEAE: hierarchical modular event argument extraction. In: HMEAE: Hierarchical Modular Event Argument Extraction; Hong Kong, China; 2019. pp. 5777-5783.

[66] Daniel C, Yinfei Y, Sheng-yi K, Nan H, Nicole L et al. Universal sentence encoder for English. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations; Brussels, Belgium; 2018. pp. 169-174.

[67] Yinfei Y, Daniel C, Amin A, Mandy G, Jax L et al. Multilingual universal sentence encoder for semantic retrieval. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations; Online; 2020. pp. 87-94.

[68] Artetxe M, Schwenk H. Massively Multilingual Sentence Embeddings for Zero-Shot Cross-Lingual Transfer and Beyond. Transactions of the Association for Computational Linguistics 2019; 7: 597-610. doi: 10.1162/tacl_a_00288

[69] Bouayad-Agha N, Casamayor G, Wanner L. Natural language generation in the context of the semantic web. Semantic Web 2014; 5 (6): 5493-513. doi: 10.3233/SW-130125

[70] Duma D, Klein E. Generating natural language from linked data: unsupervised template extraction. In: Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013)–Long Papers; Potsdam, Germany; 2013. pp. 83-94.

[71] Sun X, Mellish C. Domain independent sentence generation from RDF representations for the Semantic Web. In: Combined Workshop on Language-Enabled Educational Technology and Development and Evaluation of Robust Spoken Dialogue Systems, European Conference on AI; Riva del Garda, Italy; 2006.

[72] Wen C, Minghui Z, Rongwen Z, Narges N. KB-NLG: from knowledge base to natural language generation. In: Proceedings of the 2019 Workshop on Widening NLP; Florence, Italy; 2019. pp. 80-82.

[73] Zhu Y, Wan J, Zhou Z, Chen L, Qiu L et al. Triple-to-text: converting RDF triples into high-quality natural languages via optimizing an inverse KL divergence. In: SIGIR '19: The 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval; Paris, France; 2019. pp. 455-464.

[74] Datar M, Immorlica N, Indyk P, Mirrokni VS. Locality-sensitive hashing scheme based on p-stable distributions. In: Proceedings of the Twentieth Annual Symposium on Computational Geometry; Brooklyn, NY, USA; 2004. pp. 253-262.

[75] Aydar M, Ayvaz S. An improved method of locality-sensitive hashing for scalable instance matching. Knowledge and Information Systems 2019; 58 (2): 275-294. doi: 10.1007/s10115-018-1199-5