Mention-centered Graph Neural Network for Document-level Relation Extraction

Jiaxin Pan[1], Min Peng[1], and Yiyan Zhang[2]

[1] Computer School, Wuhan University pjx_1997@whu.edu.cn
[2] National University of Singapore e0261914@u.nus.edu

Abstract. Document-level relation extraction aims to discover relations between entities across a whole document. How to build the dependency of entities from different sentences in a document remains to be a great challenge. Current approaches either leverage syntactic trees to construct document-level graphs or aggregate inference information from different sentences. In this paper, we build cross-sentence dependencies by inferring compositional relations between inter-sentence mentions. Adopting aggressive linking strategy, intermediate relations are reasoned on the document-level graphs by mention convolution. We further notice the generalization problem of NA instances, which is caused by incomplete annotation and worsened by fully-connected mention pairs. An improved ranking loss is proposed to attend this problem. Experiments show the connections between different mentions are crucial to document-level relation extraction, which enables the model to extract more meaningful higher-level compositional relations.

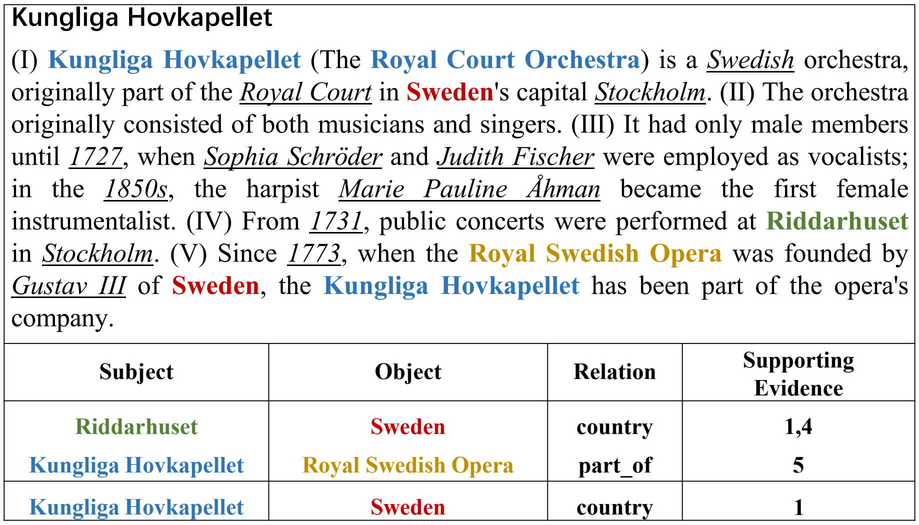
Keywords: Document-level Relation Extraction · Graph Neural Network · Ranking Loss.

1 Introduction

Relation Extraction(RE) is the task of predicting relations between named entities in plain text. It is an important task for many downstream applications such as knowledge graph construction [1] and question answering [2]. Most existing approaches [3–6] focus on sentence-level RE, which discover relational facts from a single sentence. However, as is shown in Figure 1, in real world scenario, many relational facts lie in different sentences in a document. The task of identifying such relations is called document-level RE. To accelerate the development of document-level RE task, [7] published the first and so far unique dataset for large-scale document-level relation extraction, DocRED[3], constructed from Wikipedia.

Current baseline[8] on document-level relation extraction calculates attention scores between entity pairs and sentences to aggregate information through the whole document. Other efforts [9, 10] link the dependency trees of adjacent

[3] https://github.com/thunlp/DocRED



Kungliga Hovkapellet

(I) **Kungliga Hovkapellet** (The **Royal Court Orchestra**) is a *Swedish* orchestra, originally part of the *Royal Court* in **Sweden**'s capital *Stockholm*. (II) The orchestra originally consisted of both musicians and singers. (III) It had only male members until *1727*, when *Sophia Schröder* and *Judith Fischer* were employed as vocalists; in the *1850s*, the harpist *Marie Pauline Åhman* became the first female instrumentalist. (IV) From *1731*, public concerts were performed at **Riddarhuset** in *Stockholm*. (V) Since *1773*, when the **Royal Swedish Opera** was founded by *Gustav III* of **Sweden**, the **Kungliga Hovkapellet** has been part of the opera's company.

Subject	Object	Relation	Supporting Evidence
Riddarhuset	Sweden	country	1,4
Kungliga Hovkapellet	Royal Swedish Opera	part_of	5
Kungliga Hovkapellet	Sweden	country	1

Fig. 1. An example document from DocRED. The first 2 relational facts from annotations are presented with named entity mentions involved in various colors. Other named entity mentions are underlined. The last instance is not included in the original annotation. Supporting evidences of each relation are provided as well.

sentences to capture interactions among inter-sentence entities. However, we figure out the relations between inter-sentence entities can be inferred directly from the intermediate relations between their coreference mentions. Taking Fig 1 as an example, to infer the "country" relation between "Riddarhuset" and "Sweden", first we need to discover the relation "in" between "Riddarhuset" and "Stockholm" in sentence (IV), the relation "capital" between "Sweden" and "Stockholm" in sentence(I) and finally make decisions through these intermediate relations. Unlike prior methods which explore their relations roundaboutly by dependency trees[11] or aggregate sentence representation, we construct document-level graphs which connect inter-sentence mentions directly as cross-sentence dependencies and perform relational reasoning on the combination of mention nodes' representation. By transferring useful information step by step through connected mentions, compound relations can be spread to entity pairs' representations incrementally. We argue that document-level graph enables the model capture long-distance relational facts transcending adjacent limits and alleviates the noises of unrelated text.

We also notice the relational facts annotated by document-level dataset can hardly be complete due to the large number of potential entity pairs. For example, in DocRED, every document has an average of 26.2 entities, requiring 660.24 cross-sentence annotations. To alleviate the laborious work, human annotators are provided with recommendations from RE models and distant supervision based on entity linking. As a result, many NA instances in DocRED indeed have relations but they are not recommended by RE models or entity linking.

Fig 1 shows an example where (Kungliga Hovkapellet, country, Sweden) is not included in original annotations. This means forcing the prediction score of mislabelling NA instances to 0 by traditional BCE loss may hurt the model's ability to generalize relational representations. The proposed fully-connected mention-centered document-level graph exacerbates the situation as it includes countless direct connections between mislabelling NA entity pairs and subordinate none NA mention pairs. To mitigate the problem, we design a new training objective which changes the classification task to ranking task, allowing the model to reach a balance between capturing the distribution of original annotations and preserving genuine relational representations.

Our contributions can be summarised as follows:

- We propose a novel mention-centered model for document-level RE using fully-connected mention pairs to capture cross-sentence dependencies. By exchanging information between mentions pairs iteratively, entity pairs are capable of discovering more accurate inter-sentence relations. Our proposed model is independent of syntactic dependency tools and can achieve state-of-the-art performance on DocRED. We demonstrate the connections between mentions are the core component of inter-sentence relation extraction.
- We show detailed analysis about the incomplete annotation problem in DocRED, which interferes the generalization of NA instances. To relieve the negative impact of aggressive linking strategy on this problem, an improved version of ranking loss is proposed. Qualitative comparison between ranking loss and BCE loss further reveals the significance of the proposed training objective on our model.

2 Model

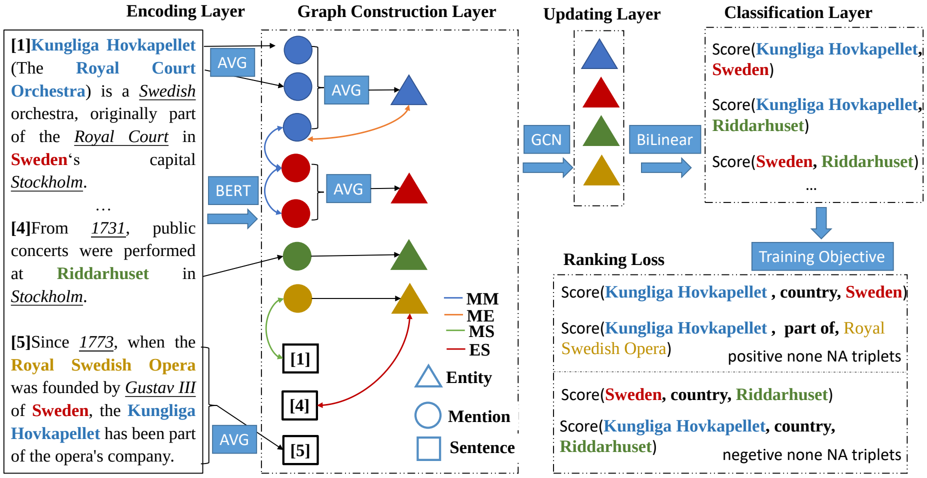
The overall architecture of proposed model is illustrated in Fig 2. The proposed model consists of four layers: Encoding Layer, Graph Construction Layer, Updating Layer and Classification Layer.

2.1 Encoding Layer

Let $[x_1, x_2, ..., x_n]$ denote the document input, we use BERT to encode the document. A Linear layer is used to compress BERT embedding into low-dimensional space for afterwards processing. Entity type embeddings and coreference embeddings used in [7] are concatenated afterwards.

$$[h_i] = [W_b(\mathrm{BERT}(x_i)); e(i); c(i)] \tag{1}$$

where W_b is a trainable weight of linear map, e and c are embedding matrix for entity type and coreference type respectively. $[;]$ denotes concatenation.



[1]Kungliga Hovkapellet (The Royal Court Orchestra) is a Swedish orchestra, originally part of the Royal Court in Sweden‘s capital Stockholm.

…

[4]From 1731, public concerts were performed at Riddarhuset in Stockholm.

[5]Since 1773, when the Royal Swedish Opera was founded by Gustav III of Sweden, the Kungliga Hovkapellet has been part of the opera's company.

Score(Kungliga Hovkapellet, Sweden)

Score(Kungliga Hovkapellet, Riddarhuset)

Score(Sweden, Riddarhuset)

...

Score(Kungliga Hovkapellet , country, Sweden)

Score(Kungliga Hovkapellet , part of, Royal Swedish Opera)

positive none NA triplets

Score(Sweden, country, Riddarhuset)

Score(Kungliga Hovkapellet, country, Riddarhuset)

negetive none NA triplets

Fig. 2. Overall architecture of the proposed model. The model first encodes the whole document by BERT. Then a document-level graph is constructed and fed into the updating layer to exchange information between different nodes. Finally, the classification layer calculates the probabilities of all entity pairs. Some node connections and self-loops are not shown for brevity.

2.2 Graph Construction Layer

Node Construction We form three types of nodes in the graph: mention nodes(M) n_m, entity nodes(E) n_e, and sentence nodes(S) n_s. A mention node n_m ranging from the s-th word to the t-th word is represented as the average of hidden state from s to t. The representation of an entity node n_e is computed as the average of the mention representations associated with the entity. Finally, a sentence node n_s is represented as the average of the word representations in the sentence. In order to distinguish different node types in the graph, we concatenate a node type t embedding to each node representation. The final node representations are then estimated as $n_m = [avg_{w_i \in m}(w_i); t_m]$, $n_e = [avg_{w_i \in e}(w_i); t_e]$, $n_s = [avg_{m_i \in s}(w_i); t_s]$

Edge Construction Accumulating compositional entity relations through representations of relevant mention pairs is the fundamental idea of our model. For this reason, we construct document-level graphs using the following 4 types of edges.

Mention-Mention(MM): To detect the implicit relations between mention pairs, we create mention-mention edges according to their relative distance. Unlike previous methods [13, 14] which only connect mentions within a sentence or coreference mentions within an entity, our model creates mention-mention edges in an aggressive strategy by connecting every mention pairs in one document, as previous illustration about (Riddarhuset, country, Sweden) in Fig 1 shows intermediate mentions can reside in different sentences or entities. In this way, document-level dependencies will be established through the chains of mentions which scattered in different sentences rather than sentence connection[13] or the roots of parse trees between neighboring sentences[14]. We generate the edge representation between two mentions starting from i-th, j-th word as:

$$A_{M_i M_j} = \sigma\left(w_{\mathrm{m}} D(d_{ij})\right) \tag{2}$$

where w_{m} are trainable model parameters and sigmoid activation function σ is used, d_{ij} are the relative distances of the mentions, D is distance embedding matrix. In this way, A_{ij} will be assigned to a real value between 0 and 1 according to their relative distance.

Mention-Entity(ME): To enable entities collect information gathered by subordinate mentions, we add ME edge. The edge between mention i and entity j is represented by:

$$A_{M_i E_j} = \begin{cases} 1, \, i \in j \\ 0, \, i \notin j \end{cases} \tag{3}$$

Mention-Sentence(MS): If a mention belongs to a sentence, we intuitively think the sentence's representation encode the mention's relation information. We set the edge representation of mention i and sentence j as:

$$A_{M_i S_j} = \begin{cases} 1, \, i \in j \\ 0, \, i \notin j \end{cases} \tag{4}$$

Entity-Sentence(ES): In the experiment we find that connecting entity nodes with their residing sentence nodes will improve the performance marginally. So, we set ES edge represented by:

$$A_{E_i S_j} = \begin{cases} 1, \, i \in j \\ 0, \, i \notin j \end{cases} \tag{5}$$

2.3 Updating Layer

To update the representations of entity pairs, we apply GCN operation on the constructed document-level graph. Vanilla GCN is designed for node classification task and weighs the importance of original nodes and neighbouring nodes equally. However, our interest is accumulating supplementary information to mention/entity nodes without losing their local expressive power. Besides, since $A_{ij} \leq 1$ in our model, the node representations will become smaller as the layers deepen especially when the node connects faraway mentions. To handle this problem, we integrate residual connections to original GCN operations:

$$h_i^{(k)} = ReLU\left(\sum_{j=1}^{n} c_i A_{ij}\left(W^{(k)} h_j^{(k-1)} + b^{(k)}\right) + h_j^{(k-1)}\right) \tag{6}$$

where $h_j^{(k)}$ is the embedding of node j at the k^{th} layer, $b^{(k)}$ is a bias term, $W^{(k)}$ is a weight matrix. $c_i = 1/\sum_{j=1}^{n} A_{ij}$ is a normalization constant.

As pointed out in [15], vanilla GCN operation makes the features of connected nodes similar. For one thing, we adopt this characteristic to synthesis higher-order information in the document-level graph. For another, by adding residual connections, we maintain original node representations rich of context information.

2.4 Classification Layer

Following [7], we compute the probability of the given entity pair$(e_i,\ e_j)$ by sigmoid function.

$$P(r|e_i, e_j) = sigmoid(N_{e_i}^T W_t N_{e_j} + b_t) \tag{7}$$

where W_t, b_t are relation type dependent trainable weights and bias.

2.5 Training Objective

We divide the triplets produced by our model into 3 types: NA triplets, positive none NA triplets(the final outputs) and negative none NA triplets. In DocRED task setting, the output scores of none NA triplets are arranged in descending order. Triplets whose scores are greater than a threshold will be selected to the testing procedure and rest of candidates will be omitted as negative none NA class . Apparently, as long as the model ranks higher scores to annotations rather than NA or negative none NA triplets, the results will be credible. Therefore, based on [16], we propose an improved version of ranking loss to simulate this procedure.

During a training batch D, $y^+ \in D$ denotes the positive none NA triplets and $y^{na} \in D$ denotes the negative none NA triplets. Let $s_\theta(y^+)$ and $s_\theta(y^{na})$ be respective scores for triplets y^+ and y^{na} generated by the network with parameter set θ. The new loss function is as follows:

$$L = \log\left(1 + \exp\left(m^+ - s_\theta(y^+)\right)\right) + \log\left(1 + \exp\left(m^- + s_\theta(y^{na})\right)\right) \tag{8}$$

where m^+ and m^- are margins which help to measure the errors between predicted scores and labels. Training by minimizing this loss function will restrict the scores of negative none NA triplets smaller than those of positive none NA triplets. Scores of NA triplets are neglected because they will not be tested. In this way, we circumvent to make hard decisions of mislabelling NA triplets.

3 Experiment

In this section, we will introduce the DocRED dataset and model settings of our experiments.

3.1 DocRED Dataset

We use the DocRED dataset to evaluate the proposed method. DocRED contains 3,053 /1,000 /1,000 documents for training, development and test respectively, with 132,375 entities and 96 relation types. Manual analysis shows about 40.7% of relational facts can only be extracted from multiple sentences and 61.1% relational instances require a variety of reasoning.

3.2 Baselines

- CNN/LSTM/BiLSTM/Context-Aware[7]: The CNN/LSTM/BiLSTM based models encode the whole document word by word with CNN/LSTM/BiLSTM as encoder. Context-Aware model[17] considers other relations in the context when predicting the target relation.
- EoG[13]/GCNN[14]: EoG connects sentence nodes in the document as inter-sentence dependencies and aggregates information through attention mechanism. GCNN constructs inter-sentence interactions by linking the roots of adjacent sentences' parse trees and co-reference mentions within an entity, and then updates information through GCN.
- BERT/BERT-2step[18]: [18] replaces the encoder with BERT[12]. BERT-2step further predicts if the relation exists between entity pairs before decides the accurate relation type of entity pairs.
- HIN[8]: HIN uses a hierarchical inference method to aggregate the inference information from entity, sentence, document levels respectively by attention mechanism and translation constraint.
- LSR[11]: LSR treats the document-level graph structure as a latent variable and induces it through structured attention of shortest dependency path.

3.3 Model Settings

We use "BERT-Base, Uncased" version as BERT encoder in our experiments. The learning rate of BERT layer is 10^{-5} while the learning rate of GCN layer is 10^{-3}. The embedding size of BERT model is 768. The layer number of GCN is set to be 2. We set m^{+} to -1 and m^{-} to -2. Other settings are the same as [7]. "+wiki" means we use relation data from Wikidata [4] to facilitate the learning of ranking loss.

4 Result

4.1 Model Performance

We use the same evaluation metrics as [7] and evaluations on the test set are reported from CodaLab.

[4] https://www.wikidata.org

Table 1. Performance of different RE models on DocRED(%). In Ignore F1 setting, relational facts appeared in training set are discarded during evaluation.

Model	**Dev**		**Test**	
	Ignore F1	**F1**	**Ignore F1**	**F1**
CNN[7]	37.99	43.45	36.44	42.33
LSTM[7]	44.41	50.66	43.60	50.12
BiLSTM[7]	45.12	50.95	44.73	51.06
Context-Aware[7]	44.84	51.10	43.93	50.64
EoG[13]	45.94	52.15	49.48	51.82
GCNN[14]	46.00	51.32	49.79	51.52
BiLSTM-LSR[11]	48.82	55.17	52.15	54.18
BiLSTM-MCN	51.89	54.00	51.24	53.54
BiLSTM-MCN+wiki	54.33	56.31	52.95	54.83
BERT[18]	-	54.16	-	53.20
BERT-2step[18]	-	54.42	-	53.92
HIN[8]	54.29	56.31	53.70	55.60
BERT-LSR[11]	52.43	59.00	56.97	59.05
BERT-MCN	56.11	57.76	56.00	58.26
- *ranking loss*	55.32	57.00	55.20	57.50
BERT-MCN+wiki	**57.33**	**60.20**	**57.00**	**59.40**

Table 1 shows the results of different models under supervised settings. From the table, we have the following observations:(1) MCN is substantially beneficial to F_1 improvement, which indicates the information flow carried by the proposed document-level graph can enhance the dependencies between entities. (2) BERT encoder and ranking loss contributes 2% F_1 and 0.76% F_1 improvement respectively. (3) By linking mentions with Wikidata, our model benefits from its accurate relation annotations. Overall, our BERT-MCN model outperforms sequence-based models, attention-based models and parse tree-based models.

4.2 Model Analysis

In this subsection, we demonstrate the effectiveness of each component using the development set of DocRED.

Aggressive Mention Linking Matters Compared with GCNN which limits the exchange of information to neighbouring sentences or inner-entity mentions, our model is able to exchange information between heterogeneous mention pairs regardless of distance. To investigate the usefulness of aggressive mention linking strategy, we restrict MM edge in 2.2.2 to only link mentions belonging to neighboring sentences(the same sentence include) or the same entities. As shown in Table 2, the GCNN-similar implementation of BERT-MCN, BERT-ADJ, achieves 55.30% F1 score in development set, which indicates the inference of cross-sentence mention pairs is the key of document-level relation extraction. Implementation of BiLSTM-ADJ achieves 52.00% F1 score in development set.

Table 2. Edge ablations on dev set (%). We remove the ranking loss of BERT-MCN for fair comparison

Model	**F1**
BERT-MCN	57.00
- *MM*	54.00
- *ME*	54.72
- *MS*	55.76
- *ES*	56.36
GCNN	51.52
BiLSTM-ADJ	52.00
BERT-ADJ	55.30

Additionally, we remove each type of edge in the constructed graph one by one to examine their effectiveness. As shown in Table 2, removal of MM and ME edges significantly degrades the model's performance while MS and ES edges do not significantly affect the performance. The result provides another evidence that the edges about mentions are the critical points of document-level relation extraction.

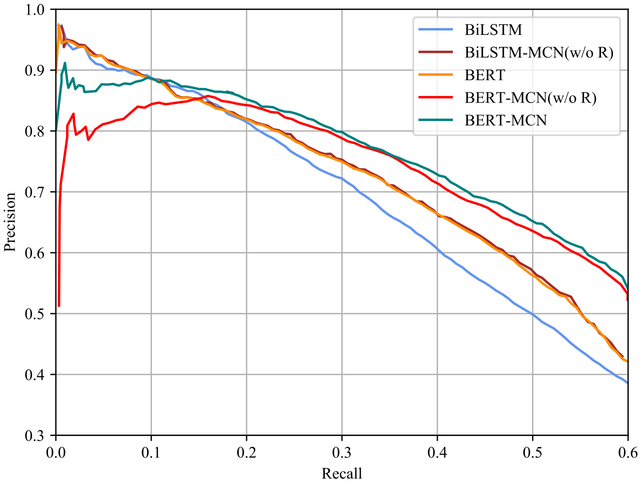


Fig. 3. Aggregate precision/recall curves of different models on DocRED

Ranking Loss and NA Class We present PR curves of BiLSTM, BiLSTM-MCN(w/o ranking loss), BERT , BERT-MCN(w/o ranking loss), BERT-MCN models in Fig 3. As shown in the figure, a peculiar sharp decline occurs in BERT-MCN(w/o R) model in the low recall area, which sharply hurts the performance but does not happen in other models. In order to find out what happens in this sharp decline, we analyse the incorrect samples from top 10% recall area. Surprisingly, we discover most samples are indeed correct or partially alluded by the text but not included in the annotations. During the annotation process of DocRED, because of the large number of potential entity pairs in the DocRED, [7] first generate triplet candidates from RE models and distant supervision based on entity linking, then ask human annotators to label these candidates. This process inevitably ignores some instances which traditional models are not good at, randomizing the distribution of DocRED. By replacing BCE loss with our proposed ranking loss, The PR curve of BERT-MCN is much smoother by circumventing predicting polluted NA instances in the training set. This problem also indicates the performance of our proposed model could be underestimated. We further relax restrictions if the entity pairs predict the highest none NA relations which are negative but listed in Wikidata. As Table 1 shows, our model gets additional 1.2% F_1 improvement.

Intra- and Inter-sentence Performance It is obvious that more supporting evidences indicate the model should consider more information from other sentences. According to the number of supporting evidences of gold relations in dev set, we divide them into $d = 0/1/2 - 3$ (0/1/2 or 3 supporting evidences) and $d \geq 4$ (more than 4 supporting evidences), and then analyse the recall on relational facts in Fig 4[5]. Apparently, MCN greatly boost the accuracy especially when the supporting evidences are large, which illustrates the MCN's ability to synthesize the information between multiple sentences.

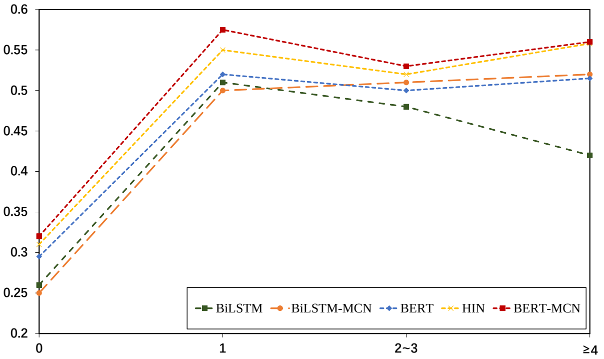


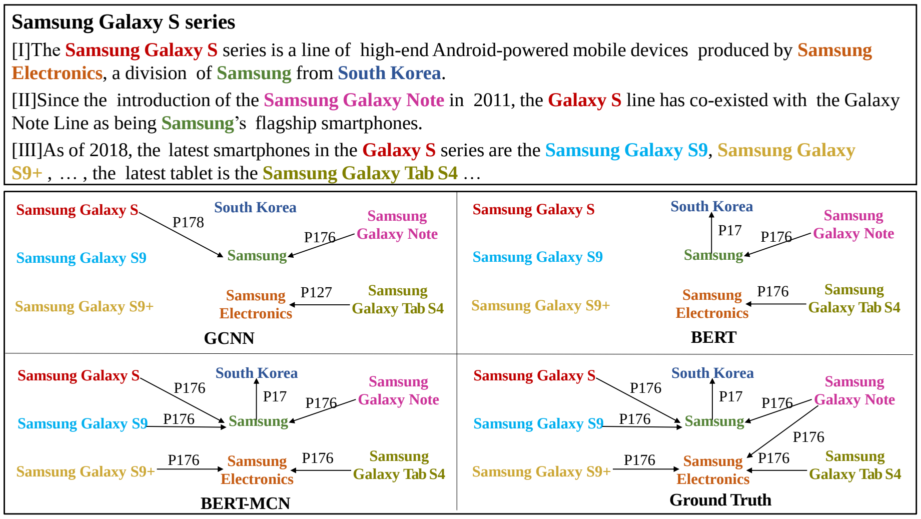
Fig. 4. Models' recall of relational facts with different number of supporting evidences.

[5] We omit BERT-LSR for the lack of source code

4.3 Case Study

Figure 5 presents some relational facts predicted by our BERT-MCN model and two baselines.

Commonsense reasoning: “Samsung” and “South Korea” has a relation “country” which can be identified by the word “from” in sentence (I). GCNN model(as well as BiLSTM model) fails to recognize this pattern possibly because “country” is more often related with another preposition “in”. However, models with BERT layer successfully predicts this relation. Additionally, BERT model successfully predicts 2 of 3 “manufacturer” relational facts connected with “Samsung Electronics”. After consulting wikipedia, we find the text “produced by Samsung Electronics” appears in both pages of correct mentions while the incorrect “Samsung Galaxy S9+” does not appear in the original text. We argue BERT may encode commonsense knowledge during pre-training process while BiLSTM encoder relies more on pattern recognition.



Samsung Galaxy S series

[I]The Samsung Galaxy S series is a line of high-end Android-powered mobile devices produced by Samsung Electronics, a division of Samsung from South Korea.

[II]Since the introduction of the Samsung Galaxy Note in 2011, the Galaxy S line has co-existed with the Galaxy Note Line as being Samsung’s flagship smartphones.

[III]As of 2018, the latest smartphones in the Galaxy S series are the Samsung Galaxy S9, Samsung Galaxy S9+ , … , the latest tablet is the Samsung Galaxy Tab S4 …

Fig. 5. Case study of “Samsung Galaxy S series” of DocRED. We visualize the predicted relational facts by different models in the lower column. “P17” refers to “country”. “P127” refers to “owned by”. “P176” refers to “manufacturer” and “P178” refers to “developer”. Some relations and words are omitted for brevity.

Logical reasoning: BERT-MCN model successfully predicts every “manufacturer” relational facts connected with “Samsung” while others do not. To identify this relation, the model first needs to identity “Samsung Electronics” produces “Samsung Galaxy S” and “Samsung Electronics” belongs to “Samsung” in sentence (I), then identify these products such as “Samsung Galaxy S9” belong to “Galaxy S” from sentence (III). We argue our MCN structure collects compositional information from various intermediate mention pairs, so that it can discover complicated higher-order inter-sentence relationships.

5 Related Work

Traditional approaches in RE focus on sentence-level relation extraction, using CNN [3] or RNN [19] to encode sentences as relation representation. Later works add attention mechanism[5, 20], reinforcement learning [21, 22], generative adversarial network [23] and capsule network [24] to deal with distant supervision[25] setting. The existing approaches which deal with document-level relation extraction mainly focus on medical relations. [26] propose the first approach for applying distant supervision to cross-sentence relation extraction. They add an edge between the dependency roots of adjacent sentences and extract features from the graph to copy with inter-sentence relations in GDKD. Later [9, 27] extend this method by introducing Graph LSTMs or Dependency-Based RNN to encode the whole graph. But they only consider up to 3 consecutive sentences. Recently, [28] propose to classify relation from entity-level pairs rather than mention-level pairs. [13] construct similar document-level graph, but they limit mention interactions within one sentence and construct relation-dependent edge representations by attention scores between entity nodes. [8] leverages translation constrict to capture relation representation between entities and aggregates information across the document by attention scores between entities and sentences.

Recent GCN methods used in RE most utilize the dependency tree to construct graph. [29] use the GCN layer to encode the dependency path for relation extraction and achieves state-of-art performance on TACRED dataset. [30] extend this method by introducing attention mechanism to extract a more precise dependency graph. [31] first perform GCN operation on dependency tree and then performs a 2nd-phase prediction based on relation-weighted graph to consider interaction between named entities and relations. [14] construct document-level graph by linking adjacent roots of parse trees and coreference mentions. [11] dynamically learn a document-level graph through structured attention of shortest dependency trees nodes and Matrix-Tree Theorem.

6 Conclusion and Future Work

In this paper, we establish cross-sentence dependencies through document-level fully-connected mention pairs. Moreover, considering our model is sensitive to widespread mislabeled NA instances of document-level relation extraction dataset, we propose an improved version of ranking loss to generalize relational representations for mislabeled NA instances. Experimental results show our model achieves comparable results with previous methods which rely on parse trees or attention mechanism. Our future work aims to design more subtle ways to exchange information between node representations of document-level graphs.

References

1. Yi, L., Luheng, H., Mari, O., Hannaneh, H.: Multi-task identification of entities, relations, and coreference for scientific knowledge graph construction. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 3219–3232, Brussels, Belgium. Association for Computational Linguistics(2018)
2. Mo, Y., Wenpeng, Y., Kazi, S.H., Cicerodos, S., Bing, X., Bowen, Z.: Improved neural relation detection for knowledge base question answering. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 571–581, Vancouver, Canada. Association for Computational Linguistics(2017)
3. Daojian, Z., Kang, L., Siwei, L., Guangyou, Z., Jun, Z.: Relation classification via convolutional deep neural network. In: Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, pages 2335–2344, Dublin, Ireland. Dublin City University and Association for Computational Linguistics(2014)
4. Peng, Z., Wei, S., Jun, T., Zhenyu, Q., Bingchen, L., Hongwei, H., Bo, X.: Attention-based bidirectional long short-term memory networks for relation classification. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pages 207–212, Berlin, Germany. Association for Computational Linguistics(2016)
5. Yankai, L., Shiqi, S., Zhiyuan, L., Huanbo, L., Maosong, S..: Neural relation extraction with selective attention over instances. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 2124–2133, Berlin, Germany. Association for Computational Linguistics(2016)
6. Van-Thuy, P., Joan, S., Masashi, S., Yuji, M.: Ranking-based automatic seed selection and noise reduction for weakly supervised relation extraction. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pages89–95, Melbourne, Australia. Association for Computational Linguistics(2018)
7. Yuan, Y., Deming, Y., Peng, L., Xu, H., Yankai, L., Zhenghao, L., Zhiyuan, L., Lixin, H., Jie, Z., Maosong, S.: DocRED: A large-scale document-level relation extraction dataset. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 764–777, Florence, Italy. Association for Computational Linguistics(2019)
8. Hengzhu, T., Yanan, C., Zhenyu, Z., Jiangxia, C., Fang, F., Shi, W., Pengfei, Y.: Hin: Hierarchical inference network for document-level relation extraction. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining, pages 197–209. Springer(2020)
9. Nanyun, P., Hoifung, P., Chris, Q., Kristina, T., Wen-tau, Y.: Cross-sentence n-ary relation extraction with graph LSTMs. Transactions of the Association for Computational Linguistics, 5:101–115(2017)
10. Wei, Z., Hongfei, L., Zhiheng, L., Xiaoxia, L., Zhengguang, L., Bo, X., Yijia, Z., Zhihao, Y., Jian, W.: An effective neural model extracting document level chemical-induced disease relations from biomedical literature. Journal of biomedical informatics, 83:1–9(2018)
11. Guoshun, N., Zhijiang, G., Ivan, S., and Wei, L.: Reasoning with latent structure refinement for document-level relation extraction. In: Proceedings of the 58th An-

nual Meeting of the Association for Computational Linguistics, pages 1546–1557, Online. Association for Computational Linguistics(2020)

12. Jacob, D, Ming-Wei, C., Kenton, L., Kristina, T.: BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics(2019)
13. Fenia, C., Makoto, M., Sophia, A.: Connecting the dots: Document-level neural relation extraction with edge-oriented graphs. 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), pages 4927–4938, Hong Kong, China. Association for Computational Linguistics(2019)
14. Sunil, K.S., Fenia, C., Makoto, M., Sophia, A.: Inter-sentence relation extraction with document-level graph convolutional neural network. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 4309–4316, Florence, Italy. Association for Computational Linguistics(2019)
15. Qimai, L., Zhichao, H., Xiao-Ming, W.: Deeper insights into graph convolutional networks for semi-supervised learning. In: Thirty-Second AAAI Conference on Artificial Intelligence(2018)
16. Cicero, d.S., Bing, X., Bowen, Z.: 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), pages 626–634, Beijing, China. Association for Computational Linguistics(2015)
17. Daniil, S., Iryna, G.: Context-aware representations for knowledge base relation extraction. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 1784–1789, Copenhagen, Denmark. Association for Computational Linguistics(2017)
18. Hong, W., Christfried, F., Rob, S., Nilesh, M., William, W.: Fine-tune bert for docred with two-step process. arXiv preprint arXiv:1909.11898(2019)
19. Javid, E., Dejing, D.: Chain based RNN for relation classification. In: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 1244–1249, Denver, Colorado. Association for Computational Linguistics(2015)
20. Daojian, Z., Kang, L., Yubo, C., Jun, Z.: Distant supervision for relation extraction via piecewise convolutional neural networks. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 1753–1762, Lisbon, Portugal. Association for Computational Linguistics(2015)
21. Jun, F., Minlie, H., Li, Z., Yang, Y., Xiaoyan, Z.: Reinforcement learning for relation classification from noisy data. In: Thirty-Second AAAI Conference on Artificial Intelligence(2018)
22. Xiangrong, Z., Shizhu, H., Kang, L., Jun, Z.: Large scaled relation extraction with reinforcement learning. In Thirty-Second AAAI Conference on Artificial Intelligence(2018)
23. Pengda, Q., Weiran, X., William, Y.W.: DSGAN: Generative adversarial training for distant supervision relation extraction. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 496–505, Melbourne, Australia. Association for Computational Linguistics(2018)

24. Xinsong, Z., Pengshuai, L., Weijia, J., Hai, Z.: Multi-labeled relation extraction with attentive capsule network. In: Proceedings of the AAAI Conference on Artificial Intelligence, volume 33, pages 7484–7491(2019)
25. Sebastian, R., Limin, Y., Andrew, M.: Modeling relations and their mentions without labeled text. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pages 148–163. Springer(2010)
26. Chris, Q., Hoifung, P.: 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, pages 1171–1182, Valencia, Spain. Association for Computational Linguistics(2017)
27. Pankaj, G., Subburam, R., Hinrich, S.,and Thomas, R.: Neural relation extraction within and across sentence boundaries. In: Proceedings of the AAAI Conference on Artificial Intelligence, volume 33, pages 6513–6520(2019)
28. Robin, J., Cliff, W., Hoifung, P.: Document-level n-ary relation extraction with multi-scale representation learning. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 3693–3704, Minneapolis, Minnesota. Association for Computational Linguistics(2019)
29. Yuhao, Z., Peng, Q., Christopher, D.M.: Graph convolution over pruned dependency trees improves relation extraction. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 2205–2215, Brussels, Belgium. Association for Computational Linguistics(2018)
30. Zhijiang, G., Yan, Z., Wei, L.: Attention guided graph convolutional networks for relation extraction. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 241–251, Florence, Italy. Association for Computational Linguistics(2019)
31. Tsu-Jui, F., Peng-Hsuan, L., Wei-Yun, M.: 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, pages 1409–1418, Florence, Italy. Association for Computational Linguistics(2019)