HIN: Hierarchical Inference Network for Document-Level Relation Extraction

Hengzhu Tang, Yanan Cao, Zhenyu Zhang, Jiangxia Cao, Fang Fang, Shi Wang, and Pengfei Yin

Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China
Institute of Computing Technology, Chinese Academy of Sciences, China

PAKDD 2020, Singapore, May 2020

OUTLINE

- Background
- Proposed Approach
- Experiments
- Conclusion

OUTLINE

- Background
- Proposed Approach
- Experiments
- Conclusion

Document-Level Relation Extraction

- **Relation extraction (RE)** aims to detect the semantic relations between entities in plain text, which plays an important role in knowledge base population and natural language understanding.

- For **document-level RE**, the input is a document with annotated entities, as well as multiple occurrences of each entity, i.e., entity mentions, the goal is to identify all the related entity pairs in the document.

Main Challenges

- Multiple entities are mentioned in the document and exhibit complex interactions.
- It requires reading and reasoning over multiple sentences to recognize the relations between target entities.

Input:
[1] "Nisei" is the ninth episode of the third season of the American science fiction television series The X-Files. [2] It premiered on the Fox network on November 24, 1995. [3] It was directed by David Nutter, and written by ***Chris Carter***, Frank Spotnitz and Howard Gordon. [4] "Nisei" featured guest appearances by Steven Williams, Raymond J. Barry and Stephen McHattie ... [8] The show centers on FBI special agents ***Fox Mulder*** (David Duchovny) and Dana Scully (Gillian Anderson) who work on cases linked to the paranormal, called X-Files ...

Subject: *Chris Carter*
Object: *Fox Mulder*
Relation: *creator* **Supporting Sentences:** 1, 3, 8

Fig. 1. An example from DocRED. Each document in DocRED is annotated with named entity mentions, coreference information, relations, and supporting sentences.

Motivation

- Most previous document-level RE methods used only entity-level information and this is not adequate. We argue that the document-level RE model requires taking advantage of **multi-granularity inference information**: entity level, sentence level and document level.

- This paper mainly solves two problems:

(1) How to obtain the inference information with different granularity;

(2) How to aggregate these different granularity inference information and make the final prediction;

OUTLINE

- Background
- Proposed Approach
- Experiments
- Conclusion

Proposed Approach

- The RE model is given a relation candidate (E_a, E_b, D) and expected to output the relations between E_a and E_b, where E_a and E_b are entities in the document D.

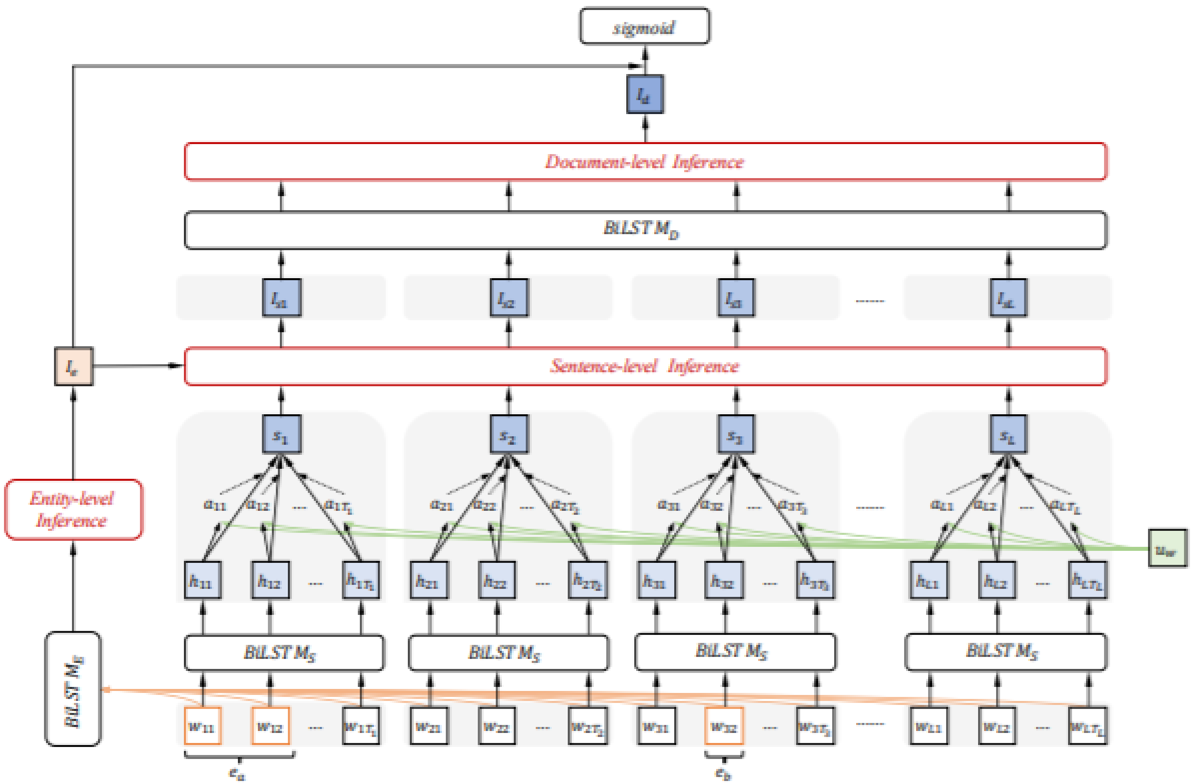

Fig. 2. The overall architecture of the Hierarchical Inference Network (HIN)

Input Layer

- We concatenate the following three embeddings as the representation of each word
  - **Word Embeddings**, which are used to capture the meaningful semantic information of words;
  - **Entity Type Embeddings**, which are obtained by mapping the entity type (e.g., PER, LOC, ORG) into a vector;
  - **Coreference Embeddings**, we assign entity mentions corresponding to the same entity with the same entity id, then entity ids are embedded into vectors.

Entity-Level Inference Module

- Compute the entity representation

$$\mathbf{h}_i = \mathrm{BiLSTM}_E(\mathbf{w}_i), i \in [1, n].$$

$$\mathbf{e}_l = avg_{w_i \in e_l}(\mathbf{h}_i), \quad \mathbf{E}_a = avg_{e_l \in E_a}(\mathbf{e}_l)$$

Average pooling is used to compute the representation of each entity

- Obtain the entity-level inference information

$$\mathbf{E}_a^k = \mathbf{W}_k^{(1)}(ReLU(\mathbf{W}_k^{(0)}\mathbf{E}_a))$$

$$\mathbf{I}_e^k = Concat\left(\mathbf{E}_a^k \mathbf{R}^k \mathbf{E}_b^k; \mathbf{E}_b^k - \mathbf{E}_a^k; \mathbf{E}_a^k; \mathbf{E}_b^k\right)$$

$$\mathbf{I}_e = G_e\left(\left[\mathbf{I}_e^1; \ldots; \mathbf{I}_e^K; \mathbf{M}(d_{ba}) - \mathbf{M}(d_{ab})\right]\right)$$

Where $\boldsymbol{E_a^k}$ and $\boldsymbol{I_e^k}$ are entity representation and entity-level inference representation obtained in the k-th representation subspace, respectively. $\boldsymbol{I_e}$ is the final entity-level inference representation.

Hierarchical Document-Level Inference Module

❖ Sentence-Level Inference

To represent words in their context, each sentence is fed into a BiLSTM encoder:

$$\mathbf{h}_{jt} = \mathrm{BiLSTM}_S\left(\mathbf{w}_{jt}\right), t \in [1, T_j].$$

Since different words in a sentence are differentially informative, the self-attention mechanism is employed to form a sentence vector:

$$\alpha_{jt} = \mathbf{u}_w^\top \tanh\left(\mathbf{W}_w \mathbf{h}_{jt} + \mathbf{b}_w\right)$$
$$a_{jt} = \frac{\exp\left(\alpha_{jt}\right)}{\sum_t \exp\left(\alpha_{jt}\right)}$$
$$\mathbf{S}_j = \sum_t a_{jt} \mathbf{h}_{jt}$$

where is the vector representation of the j-th sentence.

A semantic matching method is adopted to obtain the sentence-level inference information from the j-th sentence:

$$\mathbf{I}_{sj} = G_s\left(\left[\mathbf{S}_j; \mathbf{I}_e; \mathbf{S}_j - \mathbf{I}_e; \mathbf{S}_j \circ \mathbf{I}_e\right]\right).$$

Hierarchical Document-Level Inference Module

❖ Document-Level Inference

Self-attention mechanism is again used to distinguish crucial sentence-level inference information for overall document-level inference representation:

$$\mathbf{c}_{sj} = \mathrm{BiLSTM}_D\left(\mathbf{I}_{sj}\right), j \in [1, L]$$

$$\alpha_j = \mathbf{u}_s^\top \tanh\left(\mathbf{W}_s \mathbf{c}_{sj} + \mathbf{b}_s\right)$$

$$a_j = \frac{\exp\left(\alpha_j\right)}{\sum_j \exp\left(\alpha_j\right)}$$

$$\mathbf{I}_d = \sum_t a_j \mathbf{c}_{sj}$$

where I_d is the document-level inference information.

Prediction Layer

❖ Prediction

we concatenate entity-level inference representation I_e and document-level inference representation I_d together to make the final prediction.

$$P\left(r|E_a, E_b\right) = sigmoid\left(\mathbf{W}_r \begin{bmatrix} \mathbf{I}_e \\ \mathbf{I}_d \end{bmatrix} + \mathbf{b}_r\right).$$

❖ Loss Function

The binary cross entropy (BCE) loss is employed as training loss:

$$Loss = -\sum_{r=1}^{l} y_r \log\left(p_r\right) + \left(1 - y_r\right) \log\left(1 - p_r\right).$$

OUTLINE

- Background
- Proposed Approach
- Experiments
- Conclusion

Overall Performance

Table 1. Performance of different models on DocRED (%).

Model	Dev		Test	
	Ign F1	F1	Ign F1	F1
CNN-RE [15]	41.58	43.45	40.33	42.26
LSTM-RE [15]	48.44	50.68	47.71	50.07
BiLSTM-RE [15]	48.87	50.94	48.78	51.06
Context-Aware [12]	48.94	51.09	48.40	50.70
HIN-GloVe	**51.06**	**52.95**	**51.15**	**53.30**
BERT-RE [13]	-	54.16	-	53.20
BERT-Two-Step [13]	-	54.42	-	53.92
HIN-BERT	**54.29**	**56.31**	**53.70**	**55.60**

❖ DocRED is the largest human-annotated document-level RE dataset constructed from Wikidata and Wikipedia, and more details about DocRED can be found in [15].

[15] Yao et al. DocRED: A Large-Scale Document-Level Relation Extraction Dataset. ACL. 2019.

Analysis by the number of supporting sentences

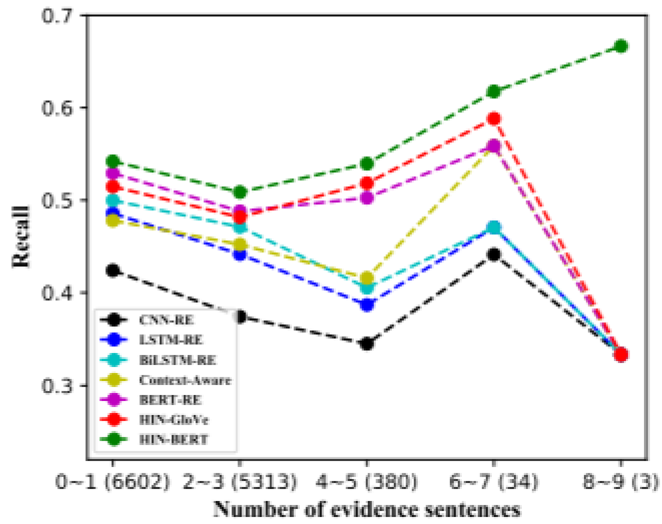

Fig. 3. Recall of models on relational facts with different number of supporting sentences. Numbers in parentheses represent the number of relational facts with different number of supporting sentences in dev set.

- Our model always performs better than other baselines, especially when the number of supporting sentences increases gradually.

Case Study

Table 3. The results predicted by BERT-RE and HIN-BERT. The reasoning type of each example is different and the first row for each example is the input document. The ***head***, ***tail***, ***relation*** and supporting sentences are colored accordingly.

Logical reasoning	[0] The Galaxy S series is a line of Samsung Electronics, a division of ***Samsung*** [2] Galaxy S line has ... being ***Samsung*** 's flagship smartphones. [4] the latest smartphones in Galaxy S series are the ***Samsung Galaxy S9*** ...
Relation	**Lable**: ***manufacturer*** **BERT-RE**: ***None*** **HIN-BERT**: ***manufacturer***
Coreference reasoning	[0] ***Robert Kingsbury Huntington***, was a naval aircrewman and member of Torpedo Squadron 8. [2] ... ***Huntington*** was shot down during the Battle of Midway ... [3] He was born in ***Los Angeles*** , California ...
Relation	**Lable**: ***birth place*** **BERT-RE**: ***death place*** **HIN-BERT**: ***birth place***
Common-sense reasoning	[0] IBM Research – Brazil is one of twelve research laboratories comprising IBM Research , its first in ***South America***. [1] It was established in June 2010 , with locations in ***São Paulo*** and Rio de Janeiro ...
Relation	**Lable**: ***continent*** **BERT-RE**: ***country*** **HIN-BERT**: ***country***

- Our model can extract triples that need multiple types of reasoning, but it doesn't perform well in common-sense reasoning scenarios. We think the problem can be solved by adding external knowledge and we leave it as our future work.

OUTLINE

- Background
- Proposed Approach
- Experiments
- Conclusion

Conclusion

- We proposed a Hierarchical Inference Network (HIN) for document-level RE, which can aggregate multi-granularity inference information for relation prediction.

- Experimental results show that our method achieves state-of-the-art performance on the largest human-annotated DocRED dataset.

- In the future, we plan to incorporate external knowledge to further improve the proposed model.

Thanks!

Questions and Advices?