Supplementary Material of Global-to-Local Neural Networks for Document-Level Relation Extraction

Difeng Wang[†]   **Wei Hu**[†, ‡, *]   **Ermei Cao**[†]   **Weijian Sun**[§]

[†] State Key Laboratory for Novel Software Technology, Nanjing University, China
[‡] National Institute of Healthcare Data Science, Nanjing University, China
[§] Huawei Technologies Co., Ltd.

{dfwang,emcao}.nju@gmail.com, whu@nju.edu.cn, sunweijian@huawei.com

A Notations

To help understanding, Table 1 summarizes the key notations used in this paper.

Symbols	Descriptions
$\mathcal{D}, k$	a document, the document length
$w, \mathbf{h}$	a word, the hidden states of a word
$\mathbf{H}$	the output of BERT
m, e, s	a mention, an entity, a sentence
$\mathbf{n}_m, \mathbf{t}_m$	a mention's node rep. & type rep.
$\mathbf{n}_e, \mathbf{t}_e$	an entity's node rep. & type rep.
$\mathbf{n}_s, \mathbf{t}_s$	a sentence's node rep. & type rep.
L	the number of R-GCN layers
$x, \mathcal{X}$	an edge type, the set of edge types
$\mathcal{N}$	a node's neighbors linked by an edge
$\mathcal{Q}, \mathcal{K}, \mathcal{V}$	queries, keys, values of multi-head attn.
z	the number of attention heads
$\mathcal{M}, \mathcal{S}$	the sets of mention & sentence nodes
$\mathbf{e}^{\text{glo}}, \mathbf{e}^{\text{loc}}$	an entity global rep. & local rep.
$\delta, \boldsymbol{\Delta}$	a distance, the distance rep. matrix
$\hat{\mathbf{e}}$	an entity final rep.
$r, \mathcal{R}$	a relation, the set of relations
$\mathbf{o}_r$	a target relation rep.
$\mathbf{o}_c$	a context relation rep.
θ	the attn. weight for a relation rep.
$\mathbf{y}$	the probability distribution of a relation
y^*	the true label of a relation

Table 1: Notations in the paper.

B Dataset Availability

The CDR dataset (Li et al., 2016) is available at https://biocreative.bioinformatics.udel.edu/media/store/files/2016/CDR_Data.zip. The DocRED dataset (Yao et al., 2019) is available at https://github.com/thunlp/DocRED. Note that, the gold standard of the test set of DocRED is unknown, and only F1 scores can be obtained

*Corresponding author

via an online interface at https://competitions.codalab.org/competitions/20717.

C Experimental Setup

In this section, we provide more details of our experiments. We implemented GLRE with PyTorch 1.5 and trained it on a server with an Intel Xeon Gold 5117 CPU, 120 GB memory, two NVIDIA Tesla V100 GPU cards and Ubuntu 18.04.

Analogous to Christopoulou et al. (2019), we pre-processed the CDR dataset, including sentence splitting, word tokenization and hypernym filtering.

When using the training set only, we trained a model on the training set, searched the best epoch in terms of F1 scores on the development set, and tested on the test set. Under the "train + dev" setting, we first trained on the training set and evaluated on the development set, in order to find the best epoch. Then, we re-ran on the union of the training and development sets until the best epoch and evaluated on the test set. For both cases, we employed dropout and layer normalization (Ba et al., 2016) to prevent model overfitting.

Hyperparameters	Values
Batch size	8
Learning rate	0.0005
Gradient clipping	10
Early stop patience	15
Regularization	10^{-4}
Dropout ratio	0.2 or 0.5
Dimension of words	768
Dimension of nodes	256
Dimension of node types	20
Number of R-GCN layers	CDR = 3, DocRED = 2
Number of attention heads	CDR = 4, DocRED = 2
Dimension of distance	20
Final dimension of entities	$532 (= 256 \times 2 + 20)$
Dimension of relations	$1064 (= 532 + 532)$

Table 2: Hyperparameters in the experiments.

The parameters of GLRE were initialized with a Gaussian distribution (mean = 0 and SD = 1.0) using a fixed initialization seed. We trained GLRE by Adam optimizer (Kingma and Ba, 2015) with mini-batches. The hidden size of BERT was set to 768. A transformation layer was used to project the BERT output into a low-dimensional space of size 256. All hyperparameter values used in the experiments are shown in Table 2.

D Case Study on DocRED

In this section, we show a few examples from the DocRED development set in Table 3, as a supplement to the case study in Section 4.5.

(1) Logical reasoning is needed in Case 1. In order to identify the relational fact, one needs to use two mentions of *"Conrad Johnson"* in S1 and S2, respectively. Specifically, one first identifies the fact that *"Conrad Johnson"* was born in *"Texas"* from S2, and then infers the fact that *"Conrad Oberon Johnson"* (coreference) is an *"American"* educator from S1.

(2) Coreference reasoning is needed in Case 2. In order to recognize the relation between *"The Hungry Ghosts"* and *"Michael Imperioli"*, one has to infer that *"He"* refers to *"Michael Imperioli"* in S5.

(3) Prior knowledge is needed in Case 3. Through some external prior knowledge, one can know that *"North America"* is a continent and *"California"* is a state, which are the valuable information to help judge their relation.

We also compare GLRE against the model without entity local representations and the model without context relation representations.

(4) In order to predict the relation between *"Kunar"* and *"Afghanistan"*, GLRE attends more to *"Afghanistan"* in [S3] by entity local representations, and correctly identifies the relation. However, GLRE without entity local representations outputs "N/A".

(4) To predict the relation between *"Breaking Dawn"* and *"Stephenie Meyer"*, GLRE relies on the context relation *"author"* between *"Eclipse"* and *"Stephenie Meyer"*, which is easier to be predicted. In contrast, GLRE without context relation representations imprecisely predicts it as *"creator"* (for general work).

[S1] **Conrad Oberon Johnson** (November 15, 1915–February 3, 2008) was an **American** music educator, long associated with the city of Houston, who was inducted into the Texas Bandmasters Hall of Fame in 2000. [S2] Born in Victoria, **Texas**, **Conrad Johnson** was nine when his family moved to Houston. ...

| **Case 1** | Label: country | GLRE: country | Wang et al.: N/A |

[S1] **Michael Imperioli** (born March 26, 1966) is an American actor, writer and director best known for ... [S4] He was starring as Detective Louis Fitch in the ABC police drama Detroit 1-8-7 ... [S5] **He** wrote and directed his first feature film, **The Hungry Ghosts**, in 2008. ...

| **Case 2** | Label: director | GLRE: director | Wang et al.: cast |

[S1] The Pleistocene coyote (Canis latrans orcutti), also known as the Ice Age coyote, is an extinct subspecies of coyote that lived in western **North America** during the Late Pleistocene era. [S2] Most remains of the subspecies were found in southern **California**, though at least one was discovered in Idaho. ...

| **Case 3** | Label: continent | GLRE: continent | Wang et al.: country |

[S1] Operation Unified Resolve is an air and ground operation to flush out and trap al - Qaeda fighters hiding in the eastern **Afghanistan** provinces. [S2] Launched on 23 June 2003, Operation Unified Resolve is a joint operation between Pakistan, United States, and **Afghanistan**. [S3] Over 500 troops, mostly from the U.S. 82nd Airborne Division, began hunting the Taliban and al - Qaeda fighters in the provinces of Nangarhar and **Kunar** on **Afghanistan**'s eastern border. ...

| **Case 4** | Label: country | GLRE: country | w/o local rep.: N/A |

[S1] **Eclipse** is the third novel in the Twilight Saga by **Stephenie Meyer** ... [S3] **Eclipse** is preceded by New Moon and followed by **Breaking Dawn**. ... [S6] Eclipse was the fourth bestselling book of 2008, only behind Twilight, New Moon, and Breaking Dawn. ...

| **Case 5** | Label: author | GLRE: author | w/o context rel.: creator |

Table 3: Case study on the DocRED development set. **Target entities** and **related entities** are colored.

References

Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E. Hinton. 2016. Layer normalization. *arXiv:1607.06450*.

Fenia Christopoulou, Makoto Miwa, and Sophia Ananiadou. 2019. Connecting the dots: Document-level neural relation extraction with edge-oriented graphs. In *EMNLP-IJCNLP*, pages 4925–4936, Hong Kong. ACL.

Diederik P. Kingma and Jimmy Lei Ba. 2015. Adam: A method for stochastic optimization. In *ICLR*, San Diego, CA, USA. OpenReview.net.

Jiao Li, Yueping Sun, Robin J Johnson, Daniela Sciaky, Chih-Hsuan Wei, Robert Leaman, Allan Peter Davis, Carolyn J Mattingly, Thomas C Wiegers, and Zhiyong Lu. 2016. BioCreative V CDR task corpus: A resource for chemical disease relation extraction. *Database*, page baw068.

Hong Wang, Christfried Focke, Rob Sylvester, Nilesh Mishra, and William Wang. 2019. Fine-tune Bert for DocRED with two-step process. *arXiv:1909.11898*.

Yuan Yao, Deming Ye, Peng Li, Xu Han, Yankai Lin, Zhenghao Liu, Zhiyuan Liu, Lixin Huang, Jie Zhou, and Maosong Sun. 2019. DocRED: A large-scale document-level relation extraction dataset. In *ACL*, pages 764–777, Florence, Italy. ACL.