Large-scale Exploration of Neural Relation Classification Architectures
The Appendices

A Pre- and Post-processing Rules

We applied two pre-processing rules for removing negative instance in **DDI** corpus as used in Zhou et al., 2018.

- Rule 1: Instances with two target drugs referring to the same drug are removed.
- Rule 2: Instances with two target drugs being in coordinate position should be removed.

For **CDR** corpus, in the post-processing step, if there is no CID relations can be identified in an abstract, the following heuristic rules are applied to find the most likely relations (Gu et al., 2017):

- Rule 1: All chemicals in the title are associated with all diseases in the entire abstract.
- Rule 2: When there is no chemical in the title, the most frequently mentioned chemical in the abstract is associated with all diseases in the entire abstract.

For **ScienceIE** corpus, we applied the following rules for post-processing. In which, rules marked by $(*)$ are which used in (Lee et al., 2017) - the system which has the best state-of-the-art result on this corpus; others are our extended linguistic rules:

Rules for recognizing *Synonym-of* relations:

- $E1$ $(E2[/w+])$ and $E2$ is 90% uppercase $(*)$
- $E1$ *or* $E2$

Rules for recognizing *Hynonym-of* relations:

- $E1\ is\ a[n][/w+]E2$
- $E1,\ a[n][/w+]E2$
- $E1\ includ[/w+]\ E2,\ E3$
- $E1\ such\ as\ E2,\ E3$
- $E1$ $([i.e.]\ E2,\ E3)$
- $E1(E2,\ /w+)$ and $E2$ is under 80% uppercase $(*)$

Rules for *None* (no relation):

- $E1/E2$ $(*)$
- $(E1)$ $E2$ $(*)$
- $E1\ and\ E2$
- $E1,\ E2$

B Model's Hyper Parameters

We implement the neural networks using the Tensor Flow library[1] and generate the dependency tree using spaCy[2]. Batch-padding is applied to pad the length of all tokens to be equal to the maximum length in each batch. The mini batch training size is set to 128.

In the experiment, we kept 10% of training data as the validation set to fine-tune the model as follows. All LSTMs and CNN employ the $RMSProp$ optimizer with the learning rate and the momentum value being 0.0005 and 0.9 respectively. They both use the $Glorot\ random\ uniform$-based initializer. The $tanh$ activation function is applied to the output of all LSTM units. The embeddings have various numbers of dimensions: 100 dims of FT, 50 dims of WN, 50 dims of $Char$, 25 dims of POS, 50 dims of DEP, 100 dims of $Dtyp$ and 100 dims of $Ddir$. The CNN filter's region sizes are $1-2-3$, each has 128, 64 and 32 filters respectively. Two hidden layers with 128 units are used after the convolution layer and before the $softmax$. The priority weight α of two directed $softmax$ classifiers is set as 0.55.

C Examples of Errors

Table 1 shows some examples of our system errors on test set. There are two types of errors: (i) FP indicated a wrong predicted relation; (ii) FN indicated a missing relation. Note that our comments for the cause of errors are empirical, based on the heuristic survey on the system outputs.

[1] *TensorFlow is an Open Source Software Library for Machine Intelligence: https://www.tensorflow.org*

[2] spaCy: Industrial-Strength Natural Language Processing in Python: https://spacy.io

Table 1: Example of errors

#	Entity pair	Golden	Predict	Corpus	ID	Type of error		Cause of errors
						FP	FN	
01	dasatinib, paclitaxel	Ef(T24,T25)	–	DDI	ML:21813412		✓	Parser error
02	coarse curvilinear mesh, meshes	HO(T9,T20)	–	ScienceIE	S0021999113006955		✓	
03	face preprocessing, eye location	HO(T31,T30)	None	ScienceIE	S2212671612000431		✓	SDP - Missing information
04	alcoholphentermine, hydrochloride	Int(T1,T2)	Ef(T1,T2)	DDI	DB:Phentermine	✓	✓	
05	lovastatin, hyperlipidemia	None	CID	CDR	1615846	✓		SDP- Redundant information
06	report, requirements	None	MT(e1,e2)	SemEval	9610	✓		
07	epinephrine, toxicity	None	CID	CDR	24091473	✓		Missing negation
08	antacid, oxybutynin	None	Ef(T14,T19)	DDI	DB:Oxybutynin	✓		
09	anthology, songs	MC(e2,e1)	MC(e1,e2)	SemEval	9681	✓	✓	Relation's directionality
10	ataxia-telangiectasia, OMA	Pr(T4,T2)	Pr(T2,T4)	Phenebank	PMC3751478	✓	✓	
11	Pseudoachondroplasia, growth retardation	Pr(T1,T13)	–	Phenebank	PMC3119180		✓	Cross-sentence relations
12	hydrochlorothiazide, dizziness	CID	–	CDR	3833372		✓	
13	acid, eyes	ED(e1,e2)	CC(e1,e2)	SemEval	8051	✓	✓	Other causes of errors
14	downturn, people	None	PP(e2,e1)	SemEval	8644	✓		
15	ribavirin, anemia	None	CID	CDR	15482540	✓		Imperfect annotation
16	Support vector machine, classification method	None	HO(T6,T27)	ScienceIE	S221267161400105X	✓		

–: Cannot generate the SDP. CID: Chemical-induced Disease. Ef: Effect. Pr: Promotes. HO: Hyponym-of.
MC: Member-Collection. ED: Entity-Destination. CC: Content-Container. PP: Product-Producer. MT: Message-Topic.

References

Jinghang Gu, Fuqing Sun, Longhua Qian, and Guodong Zhou. 2017. Chemical-induced disease relation extraction via convolutional neural network. *Database (Oxford)*, 2017:bax024.

Ji Young Lee, Franck Dernoncourt, and Peter Szolovits. 2017. Mit at semeval-2017 task 10: Relation extraction with convolutional neural networks. In *Proceedings of the 11th International Workshop on Semantic Evaluations (SemEval-2017)*, pages 978–984.

Deyu Zhou, Lei Miao, and Yulan He. 2018. Position-aware deep multi-task learning for drugdrug interaction extraction. *Artificial intelligence in medicine*, In Press.