Inter-sentence Relation Extraction with Graph Convolutional Neural Network

Sunil Kumar Sahu, Fenia Christopoulou, Makoto Miwa, Sophia Ananiadou

ACL 2019 | Florence, Italy

Relation Extraction

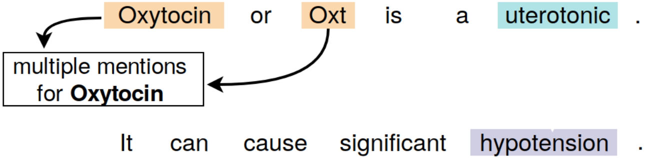
The task of **identifying interactions between named entities**

Relation Extraction

The task of **identifying interactions between named entities**

- Entity-based Relation Extraction

  mention: **unique** named **entities**
  concept: multiple entity **mentions**
  (aliases) mapped to the **same concept**



Relation Extraction

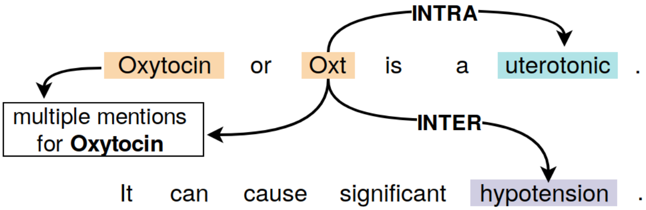
The task of **identifying interactions between named entities**

- Entity-based Relation Extraction

  mention: **unique** named **entities**
  concept: multiple entity **mentions** (aliases) mapped to the **same concept**

- Context-based Relation Extraction

  Intra-sentence: Entities in the **same sentence**
  Inter-sentence: Entities in **different sentences**

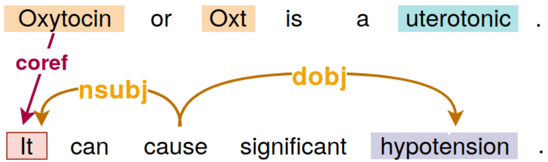


Motivation

- **Local** dependencies: **within** sentences
  - Dependency parsing [Culotta and Sorensen, 2004; Liu et al., 2015]
  - Adequate for intra-sentence relations

- **Non-local** dependencies: **across** sentences
  - Coreference [Ma et al., 2016]
  - Discourse dependencies
  - Required for inter-sentence relations

- Relations depend on both **local** and **non-local dependencies**



Proposed Approach

IDEA: Utilise local and non-local dependencies in combination

Proposed Approach

IDEA: Utilise local and non-local dependencies in combination

- Represent a snippet as a graph
- Words = nodes
  Edges = local, non-local dependencies
- Incorporate GCNN for graph encoding
- *Multi-instance Learning* for concept-level relation extraction

Proposed Approach

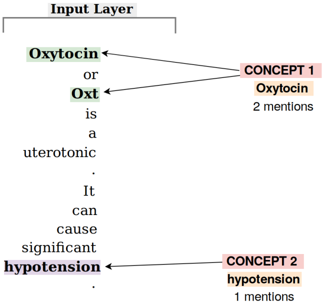
IDEA: Utilise local and non-local dependencies in combination

- Represent a snippet as a graph
- Words = nodes
  Edges = local, non-local dependencies
- Incorporate GCNN for graph encoding
- *Multi-instance Learning* for concept-level relation extraction

Task Definition

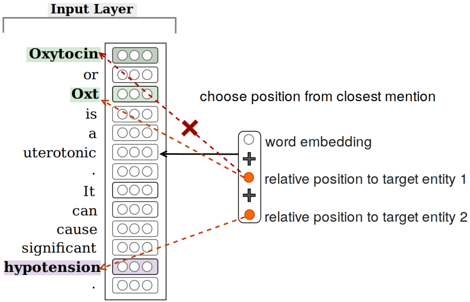
- Inter-sentence, concept-level relation extraction
- *Input*: entity concepts (c_1, c_2)
  entity mentions for each concept $(c_1^{m_1}, \ldots, c_1^{m_i}), (c_2^{m_1}, \ldots, c_2^{m_j})$
  textual snippet t
- *Output*: relation r between two concepts (c_1, r, c_2) in snippet t

Proposed Approach: Architecture



Input: Marked **named entity concepts** and their **mentions**

Proposed Approach: Architecture

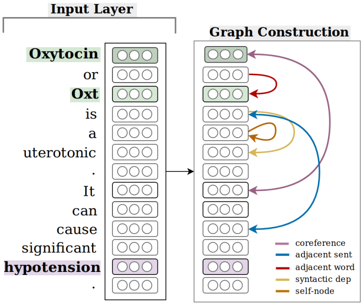


Input: Marked **named entity concepts** and their **mentions**

Node representations are built as the concatenation of

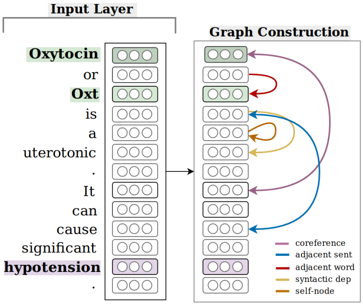
- **Word** representations
- **Relative Position** representations from closest target mention

Proposed Approach: Architecture



- Graph construction $\rightarrow$ **map** entire **document** to a **graph**
- Words $=$ nodes
  Edges $=$ semantic, syntactic, sequential dependencies
- Combination of **intra**- and **inter**-sentence information

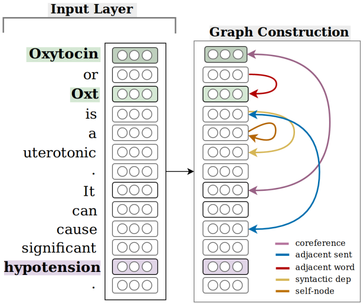
Proposed Approach: Architecture

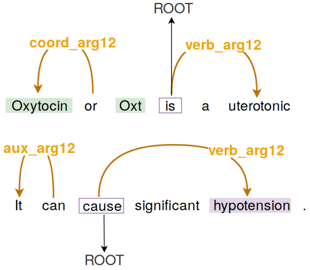


Local dependencies:

Non-local dependencies:

Proposed Approach: Architecture



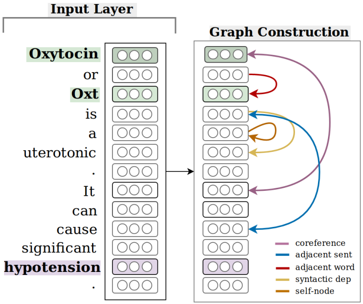


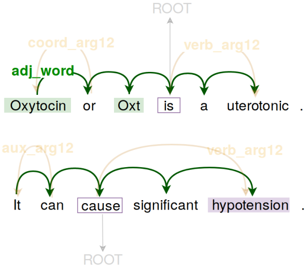
Local dependencies:

- **Syntactic dependency** → clues for intra-sentence relations

Non-local dependencies:

Proposed Approach: Architecture



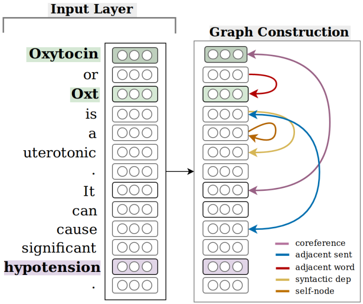


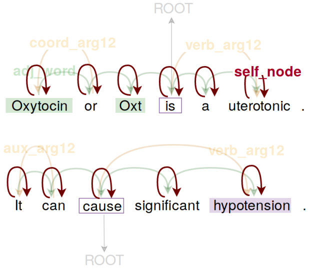
Local dependencies:

- **Syntactic dependency** → clues for intra-sentence relations
- **Adjacent word** → sequential information

Non-local dependencies:

Proposed Approach: Architecture



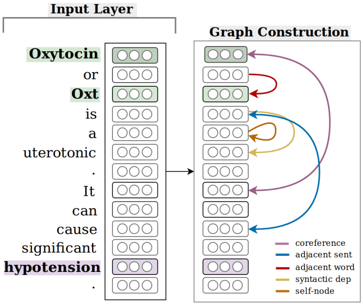


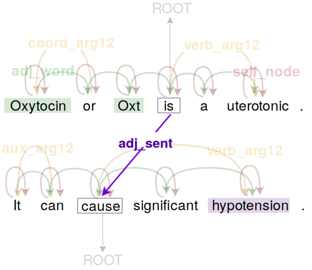
Local dependencies:

- **Syntactic dependency** → clues for intra-sentence relations
- **Adjacent word** → sequential information
- **Self-node** → node semantic information

Non-local dependencies:

Proposed Approach: Architecture





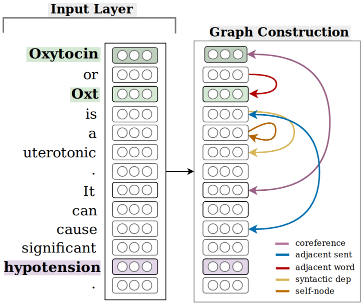
Local dependencies:

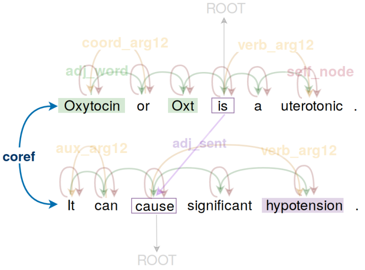
- **Syntactic dependency** → clues for intra-sentence relations
- **Adjacent word** → sequential information
- **Self-node** → node semantic information

Non-local dependencies:

- **Adjacent sentence** → discourse dependencies [Quirk and Poon, 2017]

Proposed Approach: Architecture



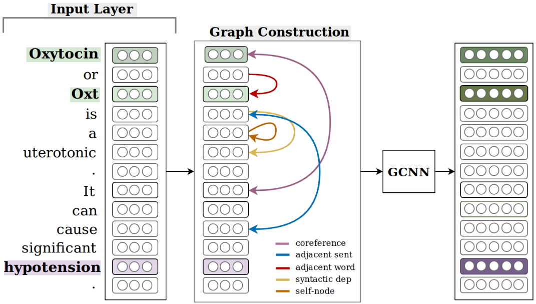


Local dependencies:

- **Syntactic dependency** → clues for intra-sentence relations
- **Adjacent word** → sequential information
- **Self-node** → node semantic information

Non-local dependencies:

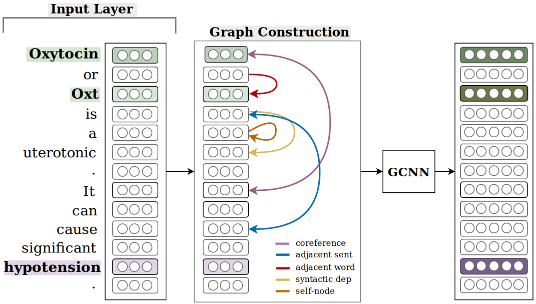
- **Adjacent sentence** → discourse dependencies [Quirk and Poon, 2017]
- **Coreference** → helpful for both intra- and inter- sentence relations



GCNN [Marcheggiani and Titov, 2017]:

$$\mathbf{x}_i^{k+1} = f\left(\sum_{u \in \nu(i)} \left(\mathbf{W}_{l(i,u)}^k \quad \mathbf{x}_u^k \; + \; \mathbf{b}_{l(i,u)}^k \right)\right),$$

Proposed Approach: Architecture



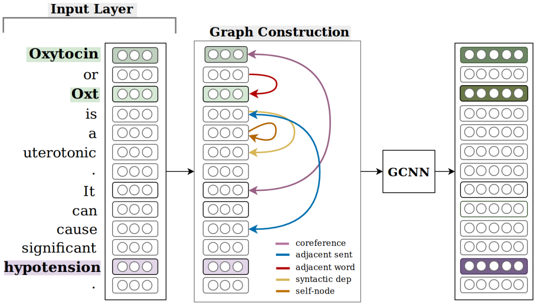
GCNN [Marcheggiani and Titov, 2017]:

$$\mathbf{x}_i^{k+1} = f\left(\sum_{u \in \nu(i)} \left(\mathbf{W}^k_{l(i,u)} \mathbf{x}^k_u + \mathbf{b}^k_{l(i,u)}\right)\right),$$

k-stacked GCNN blocks

$l(\cdot)$ labelled edge type

$v(i)$ neighboring nodes



GCNN [Marcheggiani and Titov, 2017]:

$$\mathbf{x}_i^{k+1} = f\left(\sum_{u \in \nu(i)} \left(\mathbf{W}^k_{l(i,u)}\,\mathbf{x}^k_u + \mathbf{b}^k_{l(i,u)}\right)\right),$$

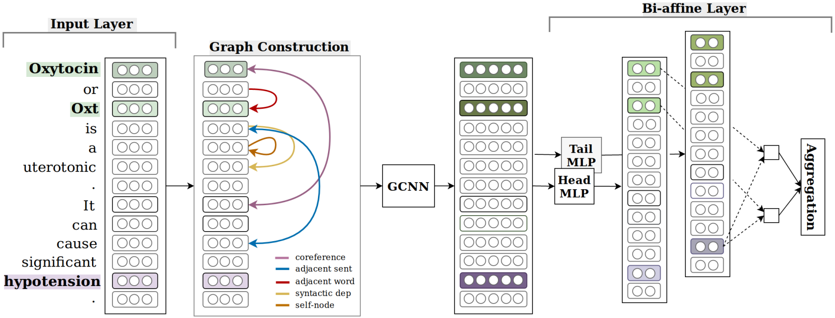
k-stacked GCNN blocks

$l(\cdot)$ labelled edge type

$v(i)$ neighboring nodes

tune number of parameters keeping

top-N most **frequent** types & merging **rare** types

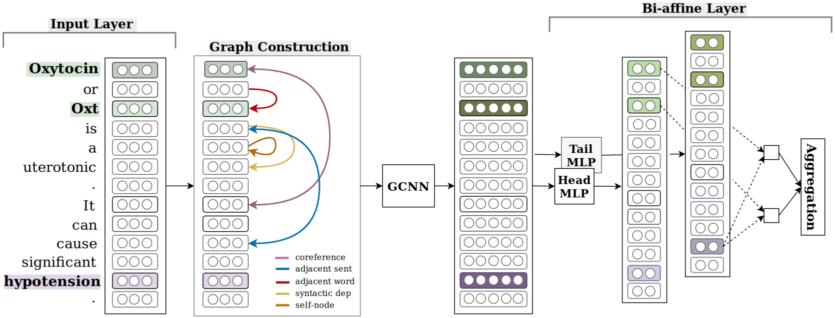


MIL Relation Classification [Verga et al., 2018]:

$$\mathbf{x}_i^{head} = \mathbf{W}_{head}^{(1)} \left(\text{ReLU} \left(\mathbf{W}_{head}^{(0)} \; \mathbf{x}_i^K \right) \right),$$

$$\mathbf{x}_i^{tail} = \mathbf{W}_{tail}^{(1)} \left(\text{ReLU} \left(\mathbf{W}_{tail}^{(0)} \; \mathbf{x}_i^K \right) \right)$$

Proposed Approach: Architecture



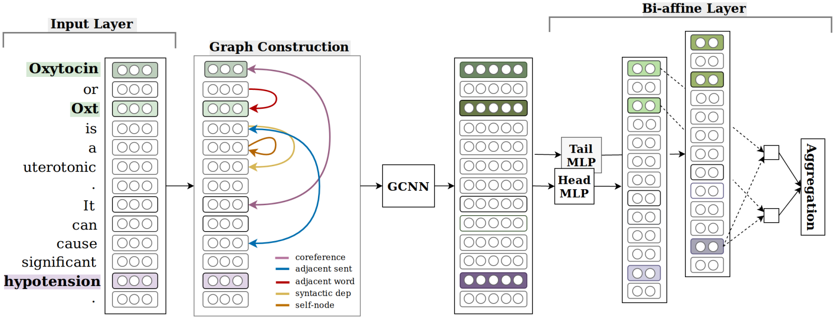
MIL Relation Classification [Verga et al., 2018]:

$$\mathbf{x}_i^{head} = \mathbf{W}_{head}^{(1)} \left(\text{ReLU} \left(\mathbf{W}_{head}^{(0)} \mathbf{x}_i^K \right) \right),$$

$$\mathbf{x}_i^{tail} = \mathbf{W}_{tail}^{(1)} \left(\text{ReLU} \left(\mathbf{W}_{tail}^{(0)} \mathbf{x}_i^K \right) \right)$$

2-layer FFNN
for each argument (head/tail)

Proposed Approach: Architecture



MIL Relation Classification [Verga et al., 2018]:

$$scores(e^{head}, e^{tail}) = \log \sum_{i \in E^{head},\, j \in E^{tail}} \exp\left(\left(\mathbf{x}_i^{head}\ \mathbf{R}\right) \mathbf{x}_j^{tail}\right),$$

bi-affine pairwise scoring

aggregate mention pairs (**x**) → concept pairs (**e**)

Datasets

1. CDR [Wei et al., 2015]
   - Chemical-Disease Relations (binary)
   - Abstract-level annotations
   - **manually annotated**

2. CHR [**new** dataset]
   → limited abstract, concept-level datasets
   - CHemical Reactions (binary)
   - Abstract-level annotations
   - **distantly supervised**

⋆ Abstracts: **PubMed**
⋆ Chemicals: **THALIA** [Soto et al., 2018]
⋆ Relations: **BioChem4j** [Swainston et al., 2017]

Data	**Item**	**Train**	**Dev.**	**Test**
CDR	# Articles	500	500	500
	# Positive pairs	1,038	1,012	1,066
	# Negative pairs	4,198	4,069	4,119
CHR	# Articles	7,298	1,182	3,614
	# Positive pairs	19,643	3,185	9,578
	# Negative pairs	69,843	11,466	33,339

Table 1: Statistics of the CDR and CHR datasets.

Experimental Settings

Tools

- Enju **dependency parser**: predicate-argument structures [Miyao and Tsujii, 2008]
- Stanford CoreNLP for **coreference** [Manning et al., 2014]
- GENIA Sentence Splitter for **sentence splitting**
- GENIA Tagger for **tokenization** [Tsuruoka et al., 2005]

Experimental Settings

Tools

- Enju **dependency parser**: predicate-argument structures [Miyao and Tsujii, 2008]
- Stanford CoreNLP for **coreference** [Manning et al., 2014]
- GENIA Sentence Splitter for **sentence splitting**
- GENIA Tagger for **tokenization** [Tsuruoka et al., 2005]

Baselines (no dependecies)

- CNN-RE: Re-implementation [Kim, 2014]
- RNN-RE: Re-implementation [Sahu and Anand, 2018]

Experimental Settings

Tools

- Enju **dependency parser**: predicate-argument structures [Miyao and Tsujii, 2008]
- Stanford CoreNLP for **coreference** [Manning et al., 2014]
- GENIA Sentence Splitter for **sentence splitting**
- GENIA Tagger for **tokenization** [Tsuruoka et al., 2005]

Baselines (no dependecies)

- CNN-RE: Re-implementation [Kim, 2014]
- RNN-RE: Re-implementation [Sahu and Anand, 2018]

✓ Adapt CNN-RE and RNN-RE to use bi-affine pairwise scoring

Results

Data	Model	P (%)	R (%)	F1 (%)
CDR	Xu et al. [2016] (SVM)	59.6	44.0	50.7
	Zhou et al. [2016] (SVM + LSTM + Kernel)	64.8	49.2	56.0
	Gu et al. [2017] (CNN + ME)	60.9	59.5	60.2
	Li et al. [2018] (RPCNN)	55.1	63.6	59.1
	Verga et al. [2018] (Transformer)	49.9	63.8	55.5
	CNN-RE	51.5	65.7	57.7
	RNN-RE	52.6	62.9	57.3
	GCNN	52.8	66.0	58.6

Table 2: CDR test set in comparison with the state-of-the-art.

- 3rd best compared to systems without additional enhancements
  (joint NER training [Verga et al., 2018], post-processing [Gu et al., 2017])
- -1.6% [Gu et al., 2017] $\rightarrow$ separate intra & inter extraction, feature-based inter-sentence model
- -0.5% [Li et al., 2018] $\rightarrow$ mention-pairs treated separately, usage of entity indicators

Results

Data	Model	P (%)	R (%)	F1 (%)
CDR	Xu et al. [2016] (SVM)	59.6	44.0	50.7
	Zhou et al. [2016] (SVM + LSTM + Kernel)	64.8	49.2	56.0
	Gu et al. [2017] (CNN + ME)	60.9	59.5	60.2
	Li et al. [2018] (RPCNN)	55.1	63.6	59.1
	Verga et al. [2018] (Transformer)	49.9	63.8	55.5
	CNN-RE	51.5	65.7	57.7
	RNN-RE	52.6	62.9	57.3
	GCNN	52.8	66.0	58.6
CHR	CNN-RE	81.2	87.3	84.1
	RNN-RE	83.0	90.1	86.4
	GCNN	84.7	90.5	**87.5**

Table 3: CDR and CHR test set in comparison with the state-of-the-art.

- Outperforms other encoders (CNN, RNN) on both datasets

Analysis: Number of edge types

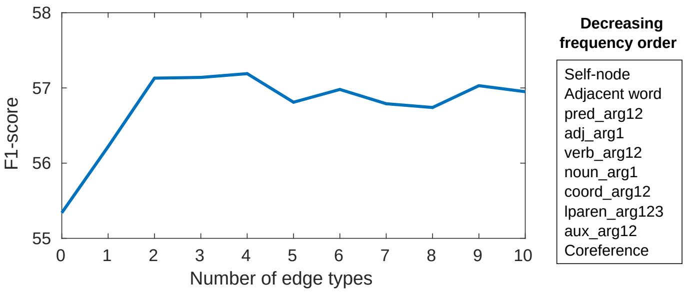


Figure 1: CDR dev. set: top-*N* most frequent edge types (rest considered as a single "rare" type).

Analysis: Number of edge types

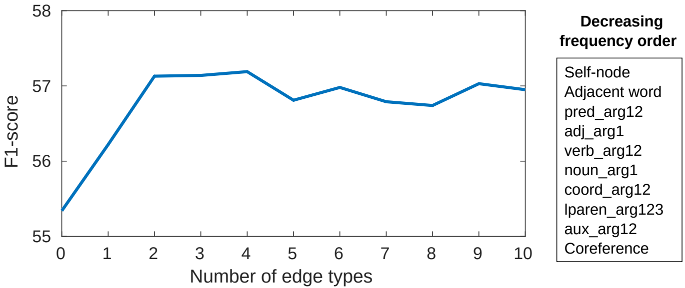


Figure 1: CDR dev. set: top-*N* most frequent edge types (rest considered as a single "rare" type).

- Keep all edges, adjust # edge types (equal to # parameters)
- **Top-4** different edge types → performs best
- Keeping **predicates** and **adjective types** → most important

Analysis: Ablation

Model	Overall	Intra	Inter
GCNN (best)	**57.19**	63.43	**36.90**
– Adjacent word	55.75	62.53	35.61
– Syntactic dependency	56.12	62.89	34.75
– Coreference	56.44	63.27	35.65
– Self-node	56.85	63.84	33.20
– Adjacent sentence	57.00	**63.99**	35.20

Table 4: Ablation analysis on the CDR development set. F1-score (%), for intra- (Intra) and inter-sentence (Inter) pairs.

- **Intra** pairs influenced more by **local dependencies** (syntax, adjacent word)
- **Inter** pairs identification supported by **all edges**

Conclusions and Future Work

Conclusions

- Proposed a **GCNN** model for **inter-sentence relation extraction**
- Creation of a **Chemical**-driven **distantly supervised corpus**
  - Motivation: limited number of abstract, concept-level datasets
- Effectiveness of **local** and **non-local** dependencies on **inter-sentence pairs**

Conclusions and Future Work

Conclusions

- Proposed a **GCNN** model for **inter-sentence relation extraction**
- Creation of a **Chemical**-driven **distantly supervised corpus**
    - Motivation: limited number of abstract, concept-level datasets
- Effectiveness of **local** and **non-local** dependencies on **inter-sentence pairs**

Future Work

- Joint NER traning
- Sub-word embeddings [Sennrich et al., 2016]
- Application to other domains

Thank you!

Questions?

efstathia.christopoulou@manchester.ac.uk
https://twitter.com/NactemNlp

Corpus

Data pre-processing:

- GENIA sentence splitter
- GENIA tagger
- Merge common Knowledge Base IDs into the same concept
- Remove self-relations (between a concept and itself)
- CDR: Hypernym filtering
- CHR: Extraction of both directions for each instance

References

Aron Culotta and Jeffrey Sorensen. Dependency tree kernels for relation extraction. In *Proceedings of Annual Meeting on Association for Computational Linguistics*, pages 423–430. Association for Computational Linguistics, 2004.

Jinghang Gu, Fuqing Sun, Longhua Qian, and Guodong Zhou. Chemical-induced disease relation extraction via convolutional neural network. *Database*, 2017:1–12, 2017.

Yoon Kim. Convolutional neural networks for sentence classification. In *Proceedings of Conference on Empirical Methods in Natural Language Processing*, pages 1746–1751. Association for Computational Linguistics, 2014.

Haodi Li, Ming Yang, Qingcai Chen, Buzhou Tang, Xiaolong Wang, and Jun Yan. Chemical-induced disease extraction via recurrent piecewise convolutional neural networks. *BMC Medical Informatics and Decision Making*, 18(2):60, Jul 2018. ISSN 1472-6947.

Yang Liu, Furu Wei, Sujian Li, Heng Ji, Ming Zhou, and Houfeng Wang. A dependency-based neural network for relation classification. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pages 285–290. Association for Computational Linguistics, 2015.

Xuezhe Ma, Zhengzhong Liu, and Eduard Hovy. Unsupervised ranking model for entity coreference resolution. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1012–1018. Association for Computational Linguistics, 2016.

Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60. Association for Computational Linguistics, 2014.

Diego Marcheggiani and Ivan Titov. Encoding sentences with graph convolutional networks for semantic role labeling. In *Proceedings of Conference on Empirical Methods in Natural Language Processing*, pages 1506–1515. Association for Computational Linguistics, 2017.

Yusuke Miyao and Jun'ichi Tsujii. Feature forest models for probabilistic HPSG parsing. *Computational Linguistics*, 34(1):35–80, 2008.

Chris Quirk and Hoifung Poon. Distant supervision for relation extraction beyond the sentence boundary. In *Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics*, pages 1171–1182. Association for Computational Linguistics, April 2017.

Sunil Kumar Sahu and Ashish Anand. Drug-drug interaction extraction from biomedical texts using long short-term memory network. *Journal of Biomedical Informatics*, 86:15 – 24, 2018.

Rico Sennrich, Barry Haddow, and Alexandra Birch. Neural machine translation of rare words with subword units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1715–1725, Berlin, Germany, August 2016. Association for Computational Linguistics. doi: 10.18653/v1/P16-1162. URL https://www.aclweb.org/anthology/P16-1162.

Axel J Soto, Piotr Przybyła, and Sophia Ananiadou. Thalia: Semantic search engine for biomedical abstracts. *Bioinformatics*, 2018.

Neil Swainston, Riza Batista-Navarro, Pablo Carbonell, Paul D Dobson, Mark Dunstan, Adrian J Jervis, Maria Vinaixa, Alan R Williams, Sophia Ananiadou, Jean-Loup Faulon, et al. biochem4j: Integrated and extensible biochemical knowledge through graph databases. *PloS one*, 12(7):e0179130, 2017.

Yoshimasa Tsuruoka, Yuka Tateishi, Jin-Dong Kim, Tomoko Ohta, John McNaught, Sophia Ananiadou, and Jun'ichi Tsujii. Developing a robust part-of-speech tagger for biomedical text. In *Panhellenic Conference on Informatics*, pages 382–392. Springer, 2005.

Patrick Verga, Emma Strubell, and Andrew McCallum. Simultaneously self-attending to all mentions for full-abstract biological relation extraction. In *Proceedings of Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 872–884. Association for Computational Linguistics, 2018.

Chih-Hsuan Wei, Yifan Peng, Robert Leaman, Allan Peter Davis, Carolyn J Mattingly, Jiao Li, Thomas C Wiegers, and Zhiyong Lu. Overview of the BioCreative V chemical disease relation (CDR) task. In *Proceedings of the fifth BioCreative challenge evaluation workshop*, pages 154–166, 2015.

Jun Xu, Yonghui Wu, Yaoyun Zhang, Jingqi Wang, Hee-Jin Lee, and Hua Xu. CD-REST: a system for extracting chemical-induced disease relation in literature. *Database*, 2016:1–10, 2016.

Huiwei Zhou, Huijie Deng, Long Chen, Yunlong Yang, Chen Jia, and Degen Huang. Exploiting syntactic and semantics information for chemical–disease relation extraction. *Database*, 2016:1–12, 2016.