Global Textual Relation Embedding for Relational Understanding

Zhiyu Chen[1], **Hanwen Zha**[1], **Honglei Liu**[1], **Wenhu Chen**[1], **Xifeng Yan**[1], and **Yu Su**[2]
[1]University of California, Santa Barbara, CA, USA
[2]The Ohio State University, OH, USA
{zhiyuchen, hwzha, honglei, wenhuchen, xyan}@cs.ucsb.edu, su.809@osu.edu

Abstract

Pre-trained embeddings such as word embeddings and sentence embeddings are fundamental tools facilitating a wide range of downstream NLP tasks. In this work, we investigate how to learn a general-purpose embedding of textual relations, defined as the shortest dependency path between entities. Textual relation embedding provides a level of knowledge between word/phrase level and sentence level, and we show that it can facilitate downstream tasks requiring relational understanding of the text. To learn such an embedding, we create the largest distant supervision dataset by linking the entire English ClueWeb09 corpus to Freebase. We use global co-occurrence statistics between textual and knowledge base relations as the supervision signal to train the embedding. Evaluation on two relational understanding tasks demonstrates the usefulness of the learned textual relation embedding. The data and code can be found at https://github.com/czyssrs/GloREPlus

1 Introduction

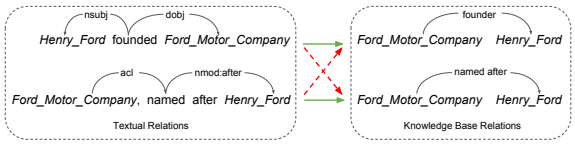
Pre-trained embeddings such as word embeddings (Mikolov et al., 2013; Pennington et al., 2014; Peters et al., 2018; Devlin et al., 2018) and sentence embeddings (Le and Mikolov, 2014; Kiros et al., 2015) have become fundamental NLP tools. Learned with large-scale (e.g., up to 800 billion tokens (Pennington et al., 2014)) open-domain corpora, such embeddings serve as a good prior for a wide range of downstream tasks by endowing task-specific models with general lexical, syntactic, and semantic knowledge.

Inspecting the spectrum of granularity, a representation between lexical (and phrasal) level and sentence level is missing. Many tasks require relational understanding of the entities mentioned in the text, e.g., relation extraction and knowledge base completion. Textual relation (Bunescu and Mooney, 2005), defined as the shortest path between two entities in the dependency parse tree of a sentence, has been widely shown to be the main bearer of relational information in text and proved effective in relation extraction tasks (Xu et al., 2015; Su et al., 2018). If we can learn a *general-purpose embedding for textual relations*, it may facilitate many downstream relational understanding tasks by providing general relational knowledge.

Similar to language modeling for learning general-purpose word embeddings, distant supervision (Mintz et al., 2009) is a promising way to acquire supervision, at no cost, for training general-purpose embedding of textual relations. Recently Su et al. (2018) propose to leverage global co-occurrence statistics of textual and KB relations to learn embeddings of textual relations, and show that it can effectively combat the wrong labeling problem of distant supervision (see Figure 1 for example). While their method, named GloRE, achieves the state-of-the-art performance on the popular New York Times (NYT) dataset (Riedel et al., 2010), the scope of their study is limited to relation extraction with small-scale in-domain training data.

In this work, we take the GloRE approach further and apply it to large-scale, domain-independent data labeled with distant supervision, with the goal of learning general-purpose textual relation embeddings. Specifically, we create the largest ever distant supervision dataset by linking the entire English ClueWeb09 corpus (half a billion of web documents) to the latest version of Freebase (Bollacker et al., 2008), which contains 45 million entities and 3 billion relational facts. After filtering, we get a dataset with over 5 million unique textual relations and around 9 million co-occurring textual and KB relation pairs. We then train textual relation embedding on the collected

Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 1324–1330
Florence, Italy, July 28 - August 2, 2019.



	$\xleftarrow{dobj}$ *founded* $\xrightarrow{nsubj}$	$\xrightarrow{acl}$ *named* $\xrightarrow{nmod:after}$
founder	2468.0	24.0
named_after	305.0	347.0
...	...	...

Figure 1: *Left:* The wrong labeling problem of distant supervision. The Ford Motor Company is both founded by and named after Henry Ford. The KB relation *founder* and *named_after* are thus both mapped to all of the sentences containing the entity pair, resulting in many wrong labels (red dashed arrows). *Right:* Global co-occurrence statistics from our distant supervision dataset, which clearly distinguishes the two textual relations.

dataset in a way similar to (Su et al., 2018), but using Transformer (Vaswani et al., 2017) instead of vanilla RNN as the encoder for better training efficiency.

To demonstrate the usefulness of the learned textual relation embedding, we experiment on two relational understanding tasks, relation extraction and knowledge base completion. For relation extraction, we use the embedding to augment PCNN+ATT (Lin et al., 2016) and improve the precision for top 1000 predictions from 83.9% to 89.8%. For knowledge base completion, we replace the neural network in (Toutanova et al., 2015) with our pre-trained embedding followed by a simple projection layer, and gain improvements on both MRR and HITS@10 measures. Our major contributions are summarized as following:

- We propose the novel task of learning general-purpose embedding of textual relations, which has the potential to facilitate a wide range of relational understanding tasks.
- To learn such an embedding, we create the largest distant supervision dataset by linking the entire English ClueWeb09 corpus to Freebase. The dataset is publicly available[1].
- Based on the global co-occurrence statistics of textual and KB relations, we learn a textual relation embedding on the collected dataset and demonstrate its usefulness on relational understanding tasks.

2 Related Work

Distant supervision methods (Mintz et al., 2009) for relation extraction have been studied by a number of works (Riedel et al., 2010; Hoffmann et al., 2011; Surdeanu et al., 2012; Zeng et al., 2015; Lin et al., 2016; Ji et al., 2017; Wu et al., 2017). (Su et al., 2018) use global co-occurrence statistics of textual and KB relations to effectively combat the wrong labeling problem. But the global statistics in their work is limited to NYT dataset, capturing domain-specific distributions.

Another line of research that relates to ours is the universal schema (Riedel et al., 2013) for relation extraction, KB completion, as well as its extensions (Toutanova et al., 2015; Verga et al., 2016). Wrong labeling problem still exists since their embedding is learned based on individual relation facts. In contrast, we use the global co-occurrence statistics as explicit supervision signal.

3 Textual Relation Embedding

In this section, we describe how to collect large-scale data via distant supervision (§3.1) and train the textual relation embedding (§3.2).

3.1 Global Co-Occurrence Statistics from Distant Supervision

To construct a large-scale distant supervision dataset, we first get the English ClueWeb09 corpus (Callan et al., 2009), which contains 500 million web documents. We employ the FACC1 dataset (Gabrilovich et al., 2013) to map ClueWeb09 to Freebase. We identify over 5 billion entity mentions in ClueWeb09 and link them to Freebase entities. From the linked documents, we extract 155 million sentences containing at least two entity mentions. We then use the Stanford Parser (Chen and Manning, 2014) with universal dependencies to extract textual relations (shortest dependency paths) between each pair of entity mentions[2], leading to 788 million relational triples (subject, textual relation, object), of which 451 million are unique.

Following (Su et al., 2018), we then collect the global co-occurrence statistics of textual and KB relations. More specifically, for a relational triple (e_1, t, e_2) with textual relation t, if (e_1, r, e_2) with

[1] https://github.com/czyssrs/GloREPlus

[2] To be more precise, only shortest dependency paths without any other entity on the path are extracted.

KB relation r exists in the KB, then we count it as a co-occurrence of t and r. We count the total number of co-occurrences of each pair of textual and KB relation across the entire corpus. We then normalize the global co-occurrence statistics such that each textual relation has a valid probability distribution over all the KB relations, which presumably captures the semantics of the textual relation. In the end, a bipartite relation graph is constructed, with one node set being the textual relations, the other node set being the KB relations, and the weighted edges representing the normalized global co-occurrence statistics.

Filtering. When aligning the text corpus with the KB, we apply a number of filters to ensure data quality and training efficiency: (1) We only use the KB relations in Freebase Commons, 70 domains that are manually verified to be of release quality. (2) Only textual relations with the number of tokens (including both lexical tokens and dependency relations) less than or equal to 10 are kept. (3) Only non-symmetric textual relations are kept, because symmetric ones are typically from conjunctions like "and" or "or", which are less of interest. (4) Only textual relations with at least two occurrences are kept. After filtering, we end up with a relation graph with *5,559,176* unique textual relations, *1,925* knowledge base (KB) relations, and *8,825,731* edges with non-zero weight. It is worth noting that these filters are very conservative, and we can easily increase the scale of data by relaxing some of the filters.

3.2 Embedding Training

Considering both effectiveness and efficiency, we employ the Transformer encoder (Vaswani et al., 2017) to learn the textual relation embedding. It has been shown to excel at learning general-purpose representations (Devlin et al., 2018).

The embedded textual relation token sequence is fed as input. For example, for the textual relation $\xleftarrow{dobj}$ $founded$ $\xrightarrow{nsubj}$, the input is the embedded sequence of $\{< -dobj >, founded, < nsubj >\}$. We project the output of the encoder to a vector z as the result embedding. Given a textual relation t_i and its embedding z_i, denote $\{r_1, r_2, ..., r_n\}$ as all KB relations, and $\tilde{p}(r_j|t_i)$ as the global co-occurrence distribution, the weight of the edge between textual relation t_i and KB relation r_j in the relation graph. The training objective is to minimize the cross-entropy loss:

$$L = -\sum_{i,j} \tilde{p}(r_j|t_i) log(p(r_j|t_i)), \quad (1)$$

Where

$$p(r_j|t_i) = (softmax(Wz_i + b))_j. \quad (2)$$

W and b are trainable parameters.

We use the filtered relation graph in §3.1 as our training data. To guarantee that the model generalizes to unseen textual relations, we take 5% of the training data as validation set. Word embeddings are initialized with the GloVe (Pennington et al., 2014) vectors[3]. Dependency relation embeddings are initialized randomly. For the Transformer model, we use 6 layers and 6 attention heads for each layer. We use the Adam optimizer (Kingma and Ba, 2015) with parameter settings suggested by the original Transformer paper (Vaswani et al., 2017). We train a maximum number of 200 epochs and take the checkpoint with minimum validation loss for the result.

We also compare with using vanilla RNN in GloRE (Su et al., 2018). Denote the embedding trained with Tranformer as **GloRE++**, standing for both new data and different model, and with RNN as **GloRE+**, standing for new data. We observe that, in the early stage of training, the validation loss of RNN decreases faster than Transformer. However, it starts to overfit soon.

4 Experiments

In this section, we evaluate the usefulness of the learned textual relation embedding on two popular relational understanding tasks, relation extraction and knowledge base completion. *We do not fine-tune the embedding*, and only use in-domain data to train a single feedforward layer to project the embedding to the target relations of the domain. We compare this with models that are specifically designed for those tasks and trained using in-domain data. If we can achieve comparable or better results, it demonstrates that the general-purpose embedding captures useful information for downstream tasks.

4.1 Relation Extraction

We experiment on the popular New York Times (NYT) relation extraction dataset (Riedel et al., 2010). Following GloRE (Su et al., 2018), we aim at augmenting existing relation extractors with the textual relation embeddings. We first average the

[3] https://nlp.stanford.edu/projects/glove/

Precision@N	100	300	500	700	900	1000
PCNN+ATT	97.0	93.7	92.8	89.1	85.2	83.9
PCNN+ATT+GloRE	97.0	97.3	94.6	93.3	90.1	89.3
PCNN+ATT+GloRE+	**98.0**	**98.7**	**96.6**	93.1	89.9	88.8
PCNN+ATT+GloRE++	**98.0**	97.3	96.0	**93.6**	**91.0**	**89.8**

Table 1: Relation extraction manual evaluation results: Precision of top 1000 predictions.

textual relation embeddings of all contextual sentences of an entity pair, and project the average embedding to the target KB relations. We then construct an ensemble model by a weighted combination of predictions from the base model and the textual relation embedding.

Same as (Su et al., 2018), we use PCNN+ATT (Lin et al., 2016) as our base model. GloRE++ improves its best F_1-score from 42.7% to 45.2%, slightly outperforming the previous state-of-the-art (GloRE, 44.7%). As shown in previous work (Su et al., 2018), on NYT dataset, due to a significant amount of false negatives, the PR curve on the held-out set may not be an accurate measure of performance. Therefore, we mainly employ manual evaluation. We invite graduate students to check top 1000 predictions of each method. They are present with the entity pair, the prediction, and all the contextual sentences of the entity pair. Each prediction is examined by two students until reaching an agreement after discussion. Besides, the students are not aware of the source of the predictions. Table 1 shows the manual evaluation results. Both GloRE+ and GloRE++ get improvements over GloRE. GloRE++ obtains the best results for top 700, 900 and 1000 predictions.

4.2 Knowledge Base Completion

We experiment on another relational understanding task, knowledge base (KB) completion, on the popular FB15k-237 dataset (Toutanova et al., 2015). The goal is to predict missing relation facts based on a set of known entities, KB relations, and textual mentions. (Toutanova et al., 2015) use a convolutional neural network (CNN) to model textual relations. We replace their CNN with our pre-trained embedding followed by one simple feed-forward projection layer.

As in (Toutanova et al., 2015), we use the best performing DISTMULT and E+DISTMULT as the base models. DISTMULT (Yang et al., 2015) learns latent vectors for the entities and each relation type, while model E (Riedel et al., 2013) learns two latent vectors for each relation type, associated with its subject and object entities respectively. E+DISTMULT is a combination model that ensembles the predictions from individual models, and is trained jointly. We conduct experiments using only KB relations (*KB only*), using their CNN to model textual relations (*Conv*), and using our embedding to model textual relations (*Emb*).

The models are tested on predicting the object entities of a set of KB triples disjoint from the training set, given the subject entity and the relation type. Table 2 shows the performances of all models measured by mean reciprocal rank (MRR) of the correct entity, and HITS@10 (the percentage of test instances for which the correct entity is ranked within the top 10 predictions). We also show the performances on the two subsets of the test set, with and without textual mentions. The pre-trained embedding achieves comparable or better results to the CNN model trained with in-domain data.

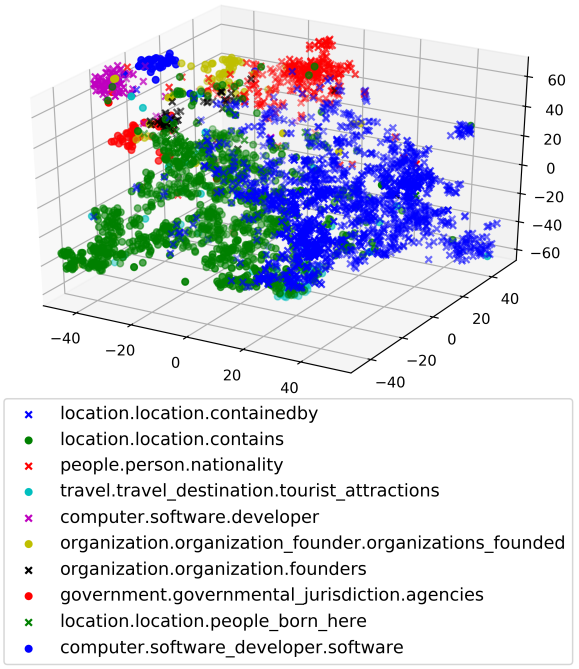


Figure 2: t-SNE visualization of our textual relation embeddings on ClueWeb validation data

5 Analysis

t-SNE visualization To measure the intrinsic property of the learned textual relation embedding,

[4]The result of our implementation is slightly different from the original paper. We have communicated with the authors and agreed on the plausibility of the result.

Model	Overall		With mentions		Without mentions	
	MRR	HITS@10	MRR	HITS@10	MRR	HITS@10
DISTMULT (KB only)	35.8	51.8	27.3	39.5	39	56.3
Conv-DISTMULT	36.5	52.5	28.5	41.4	39.4	56.5
Emb-DISTMULT (GloRE+)	36.4	52.6	**28.8**	**41.8**	39.3	56.7
Emb-DISTMULT (GloRE++)	**36.6**	**53.0**	28.0	40.8	**39.8**	**57.1**
E+DISTMULT (KB only)	37.8	53.5	29.5	43	40.9	57.3
Conv-E+Conv-DISTMULT	38.7	**54.4**	**30.0**	**43.8**	41.9	58.2
Emb-E+Emb-DISTMULT (GloRE+)	38.8	54.2	**30.0**	43.3	42.0	58.2
Emb-E+Emb-DISTMULT (GloRE++)	**38.9**	**54.4**	**30.0**	43.5	**42.1**	**58.3**

Table 2: Results of KB completion on FB15k-237 dataset[4], measured by MRR and HITS@10 (Both scaled by 100).

Subject and object	Francis Clark Howell, Kansas City		
KB relation	people.person.place_of_birth		
Textual relation in NYT train set	$\xleftarrow{nsubjpass}$ *born* $\xrightarrow{nmod:on}$ *nov.* $\xrightarrow{nmod:in}$		
Corresponding sentence in NYT train set	...Francis Clark Howell was born on nov. 27, 1925, in Kansas City, ...		
Top-5 nearest neighbors in ClueWeb train set	Textual relation	Cosine similarity	A corresponding sentence in ClueWeb raw data
	$\xleftarrow{nsubjpass}$ *born* $\xrightarrow{nmod:in}$ 1295 $\xrightarrow{nmod:in}$	0.61	...According to the Lonely Planet Guide to Venice, St. Roch was born in 1295 in Montpellier, France, and at the age of 20 began wandering...
	$\xleftarrow{nsubjpass}$ *born* $\xrightarrow{nmod:in}$ 1222 $\xrightarrow{nmod:in}$	0.61	...Isabel BIGOD was born in 1222 in Thetford Abbey, Norfolk, England...
	$\xleftarrow{nsubjpass}$ *born* $\xrightarrow{dobj}$ *Lannerback* $\xrightarrow{nmod:in}$	0.60	...Yngwie (pronounced "ING-vay") Malmsteen was born Lars Johann Yngwie Lannerback in Stockholm, Sweden, in 1963, ...
	$\xleftarrow{nsubjpass}$ *born* $\xrightarrow{nmod:in}$ *Leigha* $\xrightarrow{appos}$ *Muzaffargarh* $\xrightarrow{nmod:in}$	0.57	...Satya Paul - Indian Designer Satya Paul was born in Leigha, Muzaffargarh in Pakistan, and came to India during the partition times...
	$\xleftarrow{nsubjpass}$ *born* $\xrightarrow{nmod:on}$ *raised* $\xrightarrow{nmod:in}$	0.55	...Governor Gilmore was born on October 6, 1949 and raised in Richmond, Virginia...

Table 3: Case study: Textual relation embedding model can well generalize to unseen textual relations via capturing common shared sub-structures.

we apply t-SNE visualization (Maaten and Hinton, 2008) on the learned embedding of ClueWeb validation set.

We filter out infrequent textual relations and assign labels to the textual relations when they co-occur more than half of the times with a KB relation. The visualization result of GloRE++ embedding associating with the top-10 frequent KB relations is shown in Figure 2. As we can see, similar textual relations are grouped together while dissimilar ones are separated. This implies that the embedding model can well generate textual relation representation for unseen textual relations, and can potentially serve as relational features to help tasks in unsupervised setting.

Case Study To show that the embedding model generalizes to unseen textual relations via capturing crucial textual sub-patterns, we randomly pick some textual relations in NYT train set but not in ClueWeb train set, and compare with its top-5 nearest neighbors in ClueWeb train set, based on the similarity of the learned embedding. A case study is shown in Table 3. We can see that the KB relation *place_of_birth* often collocates with a preposition *in* indicating the object fits into a location type, and some key words like *born*. Together, the sub-structure *born in* serves as a strong indicator for *place_of_birth* relation. There is almost always some redundant information in the textual relations, for example in the textual relation $\xleftarrow{nsubjpass}$ *born* $\xrightarrow{nmod:on}$ *nov.* $\xrightarrow{nmod:in}$, the sub-structure $\xrightarrow{nmod:on}$ *nov.* does not carry crucial information indicating the target relation. A good textual relation embedding model should be capable of learning to attend to the crucial semantic patterns.

Acknowledgment

The authors would like to thank the anonymous reviewers for their thoughtful comments. This research was sponsored in part by the Army Research Laboratory under cooperative agreements W911NF09-2-0053 and NSF IIS 1528175. The views and conclusions contained herein are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notice herein.

References

Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the ACM SIGMOD International conference on Management of data*, pages 1247–1250. ACM.

Razvan C Bunescu and Raymond J Mooney. 2005. A shortest path dependency kernel for relation extraction. In *Proceedings of Conference on Empirical Methods in Natural Language Processing*, pages 724–731. Association for Computational Linguistics.

Jamie Callan, Mark Hoy, Changkuk Yoo, and Le Zhao. 2009. Clueweb09 data set.

Danqi Chen and Christopher D Manning. 2014. A fast and accurate dependency parser using neural networks. In *Proceedings of Conference on Empirical Methods in Natural Language Processing*, pages 740–750. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Evgeniy Gabrilovich, Michael Ringgaard, and Amarnag Subramanya. 2013. FACC1: Freebase annotation of ClueWeb corpora, version 1 (release date 2013-06-26, format version 1, correction level 0). http://lemurproject.org/clueweb09/.

Raphael Hoffmann, Congle Zhang, Xiao Ling, Luke Zettlemoyer, and Daniel S Weld. 2011. Knowledge-based weak supervision for information extraction of overlapping relations. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 541–550. Association for Computational Linguistics.

Guoliang Ji, Kang Liu, Shizhu He, Jun Zhao, et al. 2017. Distant supervision for relation extraction with sentence-level attention and entity descriptions. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Diederik Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *Proceedings of the International Conference on Learning Representations*.

Ryan Kiros, Yukun Zhu, Ruslan R Salakhutdinov, Richard Zemel, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. 2015. Skip-thought vectors. In *Advances in neural information processing systems*, pages 3294–3302.

Quoc Le and Tomas Mikolov. 2014. Distributed representations of sentences and documents. In *Proceedings of the International Conference on Machine Learning*, pages 1188–1196.

Yankai Lin, Shiqi Shen, Zhiyuan Liu, Huanbo Luan, and Maosong Sun. 2016. Neural relation extraction with selective attention over instances. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 2124–2133. Association for Computational Linguistics.

Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Proceedings of the Annual Conference on Neural Information Processing Systems*, pages 3111–3119.

Mike Mintz, Steven Bills, Rion Snow, and Dan Jurafsky. 2009. Distant supervision for relation extraction without labeled data. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 1003–1011. Association for Computational Linguistics.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of Conference on Empirical Methods in Natural Language Processing*, pages 1532–1543.

Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. In *Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics*.

Sebastian Riedel, Limin Yao, and Andrew McCallum. 2010. Modeling relations and their mentions without labeled text. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 148–163. Springer.

Sebastian Riedel, Limin Yao, Andrew McCallum, and Benjamin M Marlin. 2013. Relation extraction with matrix factorization and universal schemas. In *Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics*.

Yu Su, Honglei Liu, Semih Yavuz, Izzeddin Gur, Huan Sun, and Xifeng Yan. 2018. Global relation embedding for relation extraction. In *Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics*.

Mihai Surdeanu, Julie Tibshirani, Ramesh Nallapati, and Christopher D Manning. 2012. Multi-instance multi-label learning for relation extraction. In *Proceedings of Conference on Empirical Methods in Natural Language Processing*, pages 455–465. Association for Computational Linguistics.

Kristina Toutanova, Danqi Chen, Patrick Pantel, Hoifung Poon, Pallavi Choudhury, and Michael Gamon. 2015. Representing text for joint embedding of text and knowledge bases. In *Proceedings of Conference on Empirical Methods in Natural Language Processing*.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Proceedings of the Annual Conference on Neural Information Processing Systems*, pages 6000–6010.

Patrick Verga, David Belanger, Emma Strubell, Benjamin Roth, and Andrew McCallum. 2016. Multilingual relation extraction using compositional universal schema. In *Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics*.

Yi Wu, David Bamman, and Stuart Russell. 2017. Adversarial training for relation extraction. In *Proceedings of Conference on Empirical Methods in Natural Language Processing*.

Yan Xu, Lili Mou, Ge Li, Yunchuan Chen, Hao Peng, and Zhi Jin. 2015. Classifying relations via long short term memory networks along shortest dependency paths. In *Proceedings of Conference on Empirical Methods in Natural Language Processing*, pages 1785–1794. Association for Computational Linguistics.

Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2015. Embedding entities and relations for learning and inference in knowledge bases. In *Proceedings of the International Conference on Learning Representations*.

Daojian Zeng, Kang Liu, Yubo Chen, and Jun Zhao. 2015. Distant supervision for relation extraction via piecewise convolutional neural networks. In *Proceedings of Conference on Empirical Methods in Natural Language Processing*, pages 1753–1762. Association for Computational Linguistics.