Improving Discriminative Learning for Zero-Shot Relation Extraction

Van-Hien Tran[1][*]**, Hiroki Ouchi**[1]**, Taro Watanabe**[1]**, Yuji Matsumoto**[2]

[1]Nara Institute of Science and Technology, Japan

`{tran.van_hien.ts1,hiroki.ouchi,taro}@is.naist.jp`

[2]RIKEN Center for Advanced Intelligence Project (AIP), Japan

`yuji.matsumoto@riken.jp`

Abstract

Zero-shot relation extraction (ZSRE) aims to predict target relations that cannot be observed during training. While most previous studies have focused on fully supervised relation extraction and achieved considerably high performance, less effort has been made towards ZSRE. This study proposes a new model incorporating discriminative embedding learning for both sentences and semantic relations. In addition, a self-adaptive comparator network is used to judge whether the relationship between a sentence and a relation is consistent. Experimental results on two benchmark datasets showed that the proposed method significantly outperforms the state-of-the-art methods.

1 Introduction

Relation extraction is a fundamental task in Natural Language Processing (NLP) that predicts the semantic relation between two entities in a given sentence. It has attracted considerable research effort as it plays a vital role in many NLP applications such as Information Extraction (Tran et al., 2021a,b) and Question Answering (Xu et al., 2016).

Most recent studies (Tran et al., 2019; Tian et al., 2021) treated this task in a fully supervised manner and achieved excellent performance. However, the supervised models cannot extract relations that are not predefined or observed during training, while many new relations always exist in real-world scenarios. Thus, it is worth enabling models to predict *new relations* that have never been seen before. Such a task is considered as zero-shot learning (Xian et al., 2019), where a key to achieving high performance is how to generalize a model to unseen classes by using a limited number of seen classes.

However, there are only a few existing studies on zero-shot relation extraction (ZSRE). Levy et al. (2017) tackled this task by using reading comprehension models with predefined question templates.

Obamuyide and Vlachos (2018) simply reduced ZSRE to a text entailment task, utilizing existing textual entailment models. Recently, Chen and Li (2021) presented ZS-BERT, which projects sentences and relations into a shared space and uses the nearest neighbor search to predict unseen relations.

The previous studies overlooked the importance of learning discriminative embeddings. In essence, the discriminative learning helps models to better distinguish relations, especially on similar relations. Our study focuses on this aspect and demonstrates its significance for improving ZSRE. Specifically, we propose a new model that incorporates discriminative embedding learning (Han et al., 2021) for both sentences and semantic relations, which is inspired by contrastive learning (Chen et al., 2020) commonly used in computer vision. In addition, instead of using distance metrics to predict unseen relations as done by Chen and Li (2021), we use a self-adaptive comparator network to judge whether the relationship between a sentence and a relation is consistent. This verification process helps the model to learn more discriminative embeddings. Experimental results on two datasets showed that our method significantly outperforms the existing methods for ZSRE.

2 Related Work

To date, ZSRE has been under-investigated so far. Levy et al. (2017) formulated ZSRE as a question-answering task. They first manually created 10 question templates for each relation type and then trained a reading comprehension model. Because it requires the effort of hand-crafted labeling, this method can be unfeasible and impractical to define templates of new-coming unseen relations. Obamuyide and Vlachos (2018) converted ZSRE to a textual entailment task, in which the input sentence containing two entities is considered as the premise *P*, whereas the relation description containing the same entity pair is regarded as the hypothesis *H*.

---

[*]Corresponding author.

They then used existing textual entailment models (Rocktäschel et al., 2016; Chen et al., 2017) as their base models, although these models may not be entirely relevant for ZSRE. The closest to our work is research by Chen and Li (2021). First, they proposed the ZS-BERT model, which learns two functions to project sentences and relation descriptions into a shared embedding space. Then, they used the nearest neighbor search to predict unseen predictions; however, it is prone to suffer the hubness problem (Radovanovic et al., 2010). Unlike the previous studies, our work emphasizes the necessity of discriminative embedding learning that may play a vital role in solving the ZSRE.

3 Proposed Model

3.1 Task Definition

Let $\mathcal{Y}_\mathcal{S}$ and $\mathcal{Y}_\mathcal{U}$ denote the sets of *seen* and *unseen* relation labels, respectively. They are disjoint, *i.e.*, $\mathcal{Y}_\mathcal{S} \cap \mathcal{Y}_\mathcal{U} = \emptyset$. Given a training set with n_S samples, the ith sample consists of the input sentence X_i, the entities e_{i1} and e_{i2}, and the description D_i of the corresponding *seen* relation label $y_s^i \in \mathcal{Y}_\mathcal{S}$, hereby denoted as $\left\{ S_i = \left(X_i, e_{i1}, e_{i2}, D_i, y_s^i \right) \right\}_{i=1}^{n_S}$. Using the training set, we train a relation model $\mathcal{M}$, *i.e.*, $\mathcal{M}\left(S_i \right) \to y_s^i \in \mathcal{Y}_\mathcal{S}$. In the test stage, given a testing sentence S' consisting of two entities and the descriptions of all *unseen* relation labels in $\mathcal{Y}_\mathcal{U}$, $\mathcal{M}$ predicts the *unseen* relation $y_u^j \in \mathcal{Y}_\mathcal{U}$ for S'.

3.2 Framework

Sentence Encoder. From the input sentence, we add four entity marker tokens ([E1], [/E1], [E2], and [/E2]) to annotate two entities, e_{i1} and e_{i2}. Then, we tokenize and input them through a pretrained BERT encoder (Devlin et al., 2019). Finally, we obtain the vector representing the relation between the two entities by concatenating the two vectors of the start tokens ([E1] and [E2]).

Relation Encoder. Most relations are well defined, and their descriptions are available from open resources such as Wikidata (Chen and Li, 2021). For each relation, e.g., "*founded by*", we input its description to the pre-trained Sentence-BERT encoder (Reimers and Gurevych, 2019) and obtain the representation vector by using the mean pooling operation on the outputs.

Overview of the Model. On the basis of the two modules above, we present our full model in Figure 1. Given a training mini-batch of N sentences,

we feed them into the **Sentence Encoder** and a subsequent nonlinear projector to obtain N final sentence embeddings. Simultaneously, we acquire K different relations from the N sentences. The K corresponding descriptions of the K relations are then fed into the **Relation Encoder** and a subsequent nonlinear projector to acquire the final relation embeddings. To obtain more discriminative embeddings, we introduce the learning constraints described in detail later. Finally, we concatenate pairs from the two spaces and use a network $\boldsymbol{F}$ to judge whether the relationship between a sentence and a relation is consistent.

3.3 Model Training

To boost the learning of discriminative embeddings for sentences and relations, we consider three main goals in training: (1) discriminative sentence embeddings, (2) discriminative relation embeddings, and (3) an effective comparator network $\boldsymbol{F}$.

Discriminative Sentence Embeddings. In Figure 1, given a mini-batch of N sentences, we obtain N corresponding sentence embeddings: [$\mathbf{s}_1$, $\mathbf{s}_2$, ..., $\mathbf{s}_N$]. To learn the discriminative features, we first use a softmax multi-class relation classifier to predict the seen relation for each sentence:

$$\mathcal{L}_{\text{Softmax}} = -\frac{1}{N} \sum_i^N y_s^i \log(\hat{y_s}^i), \qquad (1)$$

where $y_s^i \in \mathcal{Y}_\mathcal{S}$ is the ground-truth seen relation label of the ith sentence and $\hat{y_s}^i$ is the predicted probability of y_s^i. However, such a softmax loss only encourages the separability of the inter-class features. Meanwhile, discriminative power characterizes features in both the separable inter-class differences and the compact intra-class variations (Wen et al., 2016). Thus, we use another loss to ensure the intra-class compactness. First, the similarity distance between two sentences is given by

$$d\left(\mathbf{s}_i, \mathbf{s}_j \right) = 1/(1 + \exp(\frac{\mathbf{s}_i}{\|\mathbf{s}_i\|} \cdot \frac{\mathbf{s}_j}{\|\mathbf{s}_j\|})). \qquad (2)$$

Clearly, this value should be small for any sentence pair of the same relation (*positive* pair) and large for a *negative* pair. We then apply such distance constraints on all T unordered sentence pairs, where
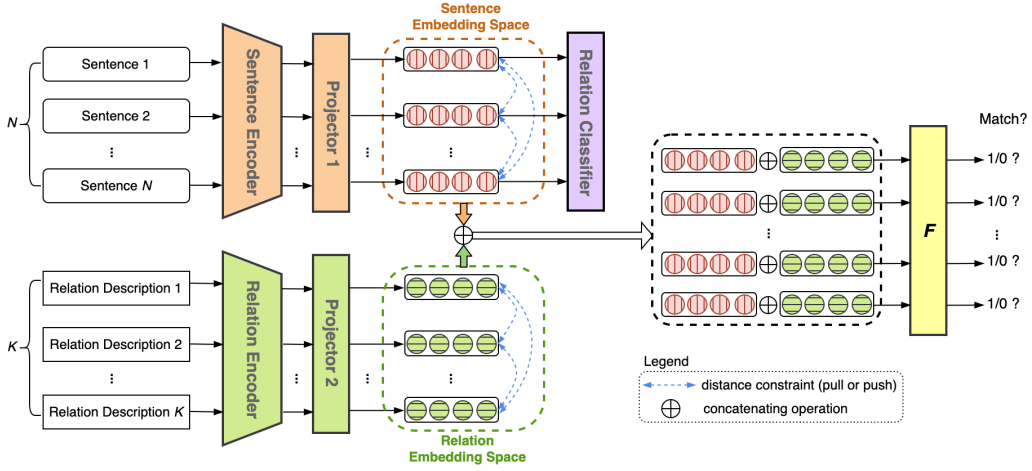
Figure 1: Overview of our proposed model with an input training mini-batch of size N.

$T = N(N-1)/2$, and formulate the loss as

$$\mathcal{L}_{\text{S2S}} = -\frac{1}{T} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left(\mathbb{I}_{ij} \log d(\mathbf{s}_i, \mathbf{s}_j) \right. \tag{3}$$
$$\left. + (1 - \mathbb{I}_{ij}) \log(1 - d(\mathbf{s}_i, \mathbf{s}_j)) \right),$$

where $\mathbb{I}_{ij} = 1$ if the pair $(\mathbf{s}_i, \mathbf{s}_j)$ is *positive* or 0 otherwise. $\mathcal{L}_{\text{S2S}}$ not only ensures the intra-relation compactness but also encourages the inter-relation separability further. Finally, the final loss of learning discriminative sentence embeddings in the sentence embedding space is defined as follows:

$$\mathcal{L}_{sent} = \mathcal{L}_{Softmax} + \gamma \cdot \mathcal{L}_{\text{S2S}}, \tag{4}$$

where γ is a hyperparameter. With this joint supervision, it is expected that not only the inter-class sentence embedding differences are enlarged, but also the intra-class sentence embedding variations are reduced. Thus, the discriminative power of the learned sentence embeddings will be enhanced.

Discriminative Relation Embeddings. In Figure 1, we obtain K corresponding relation embeddings: $[\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_K]$ for K different relations in the relation embedding space. From the K relations, we have a total of Q pairs ($Q = K(K-1)/2$), where each pair includes two different unordered relations. Thus, we maximize distance for each of these pairs and define the loss of learning discriminative relation embeddings by

$$\mathcal{L}_{rel} = -\frac{1}{Q} \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \log(1 - d(\mathbf{r}_i, \mathbf{r}_j)), \tag{5}$$

where $d(\mathbf{r}_i, \mathbf{r}_j)$ is the similarity distance between two relations using Equation 2.

Comparator Network. After obtaining the discriminative embeddings of sentences and relations, we use a comparator network $\boldsymbol{F}$ to judge how well a sentence is consistent with a specific relation. This validation information will guide our model to learn more discriminative embeddings. In Figure 1, we concatenate sentences and relations as pairs and feed them into $\boldsymbol{F}$. To enhance the training efficiency, we control the rate of positive and negative pairs. Specifically, we keep all positive pairs but randomly keep only a part of negative pairs (e.g., positive:negative *rate* is 1:3). The $\boldsymbol{F}$ is a two-layer nonlinear neural network that outputs a scalar similarity score in the range of (0,1). Finally, the loss of training $\boldsymbol{F}$ is defined as

$$\mathcal{L}_F = -\frac{\sum_{i=1}^{\mathcal{N}_{pos}} \log v_i + \sum_{j=1}^{\mathcal{N}_{neg}} \log(1 - v_j)}{\mathcal{N}_{pos} + \mathcal{N}_{neg}}, \tag{6}$$

where v_i and v_j are the similarity scores of the ith positive pair and j^{th} negative pair, respectively; $\mathcal{N}_{pos}$ and $\mathcal{N}_{neg}$ are the number of positive pairs and negative pairs for training.

Total Loss. Based on the three aforementioned losses, the full loss function for training our model is as follows:

$$\mathcal{L} = \mathcal{L}_F + \alpha \mathcal{L}_{sent} + \beta \mathcal{L}_{rel}, \tag{7}$$

where α and β are hyperparameters that control the importance of $\mathcal{L}_{sent}$ and $\mathcal{L}_{rel}$, respectively.

3.4 Zero-Shot Relation Prediction

In the testing stage, we conduct zero-shot relation prediction by comparing the similarity score of a

given sentence with all the unseen semantic relation representations. We classify the sentence $\mathbf{s}_i$ to the unseen relation that has the largest similarity score among relations, which can be formulated as

$$P_{zsre}\left(\mathbf{s}_i\right) = \max_j \left\{v_{ij}\right\}_{j=1}^{|\mathcal{Y}_{\mathcal{U}}|}. \qquad (8)$$

4 Experiments

4.1 Dataset

Following the previous work (Chen and Li, 2021), we evaluate our model on two benchmark datasets: **Wiki-ZSL** and **FewRel** (Han et al., 2018). FewRel is a human-annotated balanced dataset consisting of 80 relation types, each of which has 700 instances. Wiki-ZSL is a subset of Wiki-KB dataset (Sorokin and Gurevych, 2017), which filters out both the "none" relation and relations that appear fewer than 300 times. The statistics of Wiki-KB, Wiki-ZSL, and FewRel are shown in Table 1. Note that descriptions of the relations in the above datasets are available and accessible from the open source Wikidata[1].

	#instances	#relations	avg. len.
Wiki-KB	1,518,444	354	23.82
Wiki-ZSL	94,383	113	24.85
FewRel	56,000	80	24.95

Table 1: The statistics of the datasets.

4.2 Experimental Settings

Following Chen and Li (2021), we randomly selected m relations as unseen ones ($m = |\mathcal{Y}_{\mathcal{U}}|$) for the testing set and split the entire dataset into the training and testing datasets accordingly. This guarantees that the m relations in the testing dataset do not appear in the training dataset. We used macro precision (P), macro recall (R), and macro F1-score ($F1$) as the evaluation metrics.

We implemented the neural networks using the PyTorch library[2]. The $tanh$ function is used with each nonlinear projector in our model. The comparator network $\boldsymbol{F}$ is a two-layer nonlinear neural network in which the hidden layer is equipped with the $tanh$ function, and the output layer size is outfitted with the $sigmoid$ function. The dropout

---

[1] https://www.wikidata.org/wiki/Wikidata:Main_Page
[2] PyTorch is an open-source software library for machine intelligence: https://pytorch.org/

	Wiki-ZSL			FewRel		
$m = 5$	P	R	F1	P	R	F1
ESIM*	48.58	47.74	48.16	56.27	58.44	57.33
CIM*	49.63	48.81	49.22	58.05	61.92	59.92
ZS-BERT*	71.54	72.39	71.96	76.96	78.86	77.90
ZS-BERT†	74.32	71.72	72.97	80.96	78.00	79.44
Ours	**87.48**	**77.50**	**82.19**	**87.11**	**86.29**	**86.69**
$m = 10$	P	R	F1	P	R	F1
ESIM*	44.12	45.46	44.78	42.89	44.17	43.52
CIM*	46.54	47.90	45.57	47.39	49.11	48.23
ZS-BERT*	60.51	60.98	60.74	56.92	57.59	57.25
ZS-BERT†	64.53	58.30	61.23	60.13	55.63	57.80
Ours	**71.59**	**64.69**	**67.94**	**64.41**	**62.61**	**63.50**
$m = 15$	P	R	F1	P	R	F1
ESIM*	27.31	29.62	28.42	29.15	31.59	30.32
CIM*	29.17	30.58	29.86	31.83	33.06	32.43
ZS-BERT*	34.12	34.38	34.25	35.54	38.19	36.82
ZS-BERT†	35.42	33.47	34.42	39.09	36.70	37.84
Ours	**38.37**	**36.05**	**37.17**	**43.96**	**39.11**	**41.36**

Table 2: Results with different m values in percentage. * indicates the results reported by Chen and Li (2021); † marks the results we reproduced using the official source code of Chen and Li (2021).

technique was applied at a rate of 0.3 on the hidden layer and embeddings of sentences and relations in the two embedding spaces. We used Adam (Kingma and Ba, 2015) as the optimizer, in which the initial learning rate was $5e - 6$; the batch size was 16 on FewRel and 32 on Wiki-ZSL; and $\alpha = 0.7$, $\beta = 0.3$, and $\gamma = 0.5$.

4.3 Results and Analysis

Main Result. The experimental results obtained by varying m unseen relations are shown in Table 2. It can be observed that our model steadily outperforms the competing methods on the test datasets when considering different values of m. In addition, the improvement in our model is smaller when m is larger. An increase in m leads to a rise in the possible choices for prediction, thereby making it more difficult to predict the correct unseen relation.

Obamuyide and Vlachos (2018) simply used two basic text entailment models (ESIM and CIM) that may not be entirely relevant for ZSRE. Besides, they ignored the importance of discriminative feature learning for sentences and relations. Chen and Li (2021) also overlooked the necessity of learning discriminative embeddings. In addition, the nearest neighbor search method in ZS-BERT is prone to cause the hubness problem (Radovanovic et al., 2010). Thus, our model was designed to overcome the existing limitations. Compared with ZS-BERT, our model significantly improved its performance when $m = 5$, by 9.22 and 7.25 $F1$-score on Wiki-

$m = 5$	F1	
	Wiki-ZSL	FewRel
Ours	**82.19**	**86.69**
Ours ($\alpha = 0$)	74.42	81.05
Ours ($\beta = 0$)	78.92	84.27
Ours ($\gamma = 0$)	77.13	82.95

Table 3: Ablation study.

ZSL and FewRel, respectively.

Impact of Discriminative Learning. To gain more insight into the improvement in our model, we analyzed the importance of learning discriminative features in both the sentence and relation spaces. In Table 3, we consider three special cases of Equation 7: (1) $\alpha = 0$ means no $\mathcal{L}_{sent}$; (2) $\beta = 0$ means no $\mathcal{L}_{rel}$; and (3) $\gamma = 0$ means no $\mathcal{L}_{S2S}$, which is a part of $\mathcal{L}_{sent}$. Clearly, all three losses are important for training our model to obtain the best performance. In particular, $\mathcal{L}_{sent}$ for learning discriminative sentence features is more important than $\mathcal{L}_{rel}$ for learning discriminative relation embeddings, as the performance decreases significantly after removing it. In addition, $\mathcal{L}_{S2S}$ plays a vital role in $\mathcal{L}_{sent}$ since it mainly ensures the intra-relation compactness property of discriminative sentence embeddings.

Feature Space Visualization. We visualized the testing sentence embeddings produced by ZS-BERT and our model in a case of $m = 5$ on the FewRel[3] dataset using t-SNE (Maaten and Hinton, 2008). Figure 2 shows that the embeddings generated by our model express not only a larger inter-relation separability but also a better intra-relation compactness, compared with the embeddings by ZS-BERT. Besides, we focus on two relations: "country" and "location". According to their descriptions (country[4] and location[5]), we can see that they are somewhat similar but different in some details. Specifically, an ordered entity pair (**e1**, **e2**) in a sentence expresses the relation "country" if and only if **e2** must be a country and **e2** has sovereignty over **e1**. Meanwhile, if the entity pair (**e1**, **e2**) does not hold the relation "country", it may appear the relation "location" when **e2** is a place that **e1** happens or exists. Thus, the two similar re-
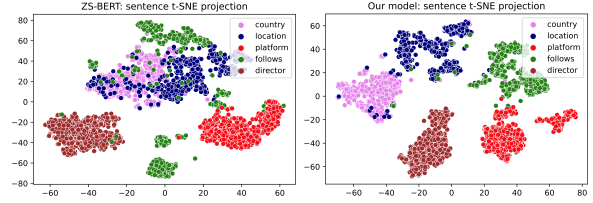

Figure 2: Visualization of the sentence embeddings by ZS-BERT and our model when $m = 5$ on the FewRel.

lations make it difficult for ZS-BERT to distinguish them. Meanwhile, our model can discriminate between them. These observations again prove the necessity of learning discriminative features for ZSRE.

5 Conclusion

In this work, we present a new model to solve the ZSRE task. Our model aims to enhance the discriminative embedding learning for both sentences and relations. It boosts inter-relation separability and intra-relation compactness of sentence embeddings and maximizes distances between different relation embeddings. In addition, a comparator network is used to validate the consistency between a sentence and a relation. Experimental results on two benchmark datasets demonstrated the superiority of the proposed model for ZSRE.

References

Chih-Yao Chen and Cheng-Te Li. 2021. ZS-BERT: Towards zero-shot relation extraction with attribute representation learning. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3470–3479, Online. Association for Computational Linguistics.

Qian Chen, Xiaodan Zhu, Zhen-Hua Ling, Si Wei, Hui Jiang, and Diana Inkpen. 2017. Enhanced LSTM for natural language inference. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1657–1668, Vancouver, Canada. Association for Computational Linguistics.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A simple framework for contrastive learning of visual representations. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 1597–1607. PMLR.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of

---

[3]The FewRel dataset is annotated by crowdworkers, thereby ensuring high quality.

[4]https://www.wikidata.org/wiki/Property:P17

[5]https://www.wikidata.org/wiki/Property:P27

deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Jiale Han, Bo Cheng, and Guoshun Nan. 2021. Learning discriminative and unbiased representations for few-shot relation extraction. In *Proceedings of the 30th ACM International Conference on Information amp; Knowledge Management*, CIKM '21, page 638–648, New York, NY, USA. Association for Computing Machinery.

Xu Han, Hao Zhu, Pengfei Yu, Ziyun Wang, Yuan Yao, Zhiyuan Liu, and Maosong Sun. 2018. FewRel: A large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4803–4809, Brussels, Belgium. Association for Computational Linguistics.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Omer Levy, Minjoon Seo, Eunsol Choi, and Luke Zettlemoyer. 2017. Zero-shot relation extraction via reading comprehension. In *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, pages 333–342, Vancouver, Canada. Association for Computational Linguistics.

Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(86):2579–2605.

Abiola Obamuyide and Andreas Vlachos. 2018. Zero-shot relation classification as textual entailment. In *Proceedings of the First Workshop on Fact Extraction and VERification (FEVER)*, pages 72–78, Brussels, Belgium. Association for Computational Linguistics.

Milos Radovanovic, Alexandros Nanopoulos, and Mirjana Ivanovic. 2010. Hubs in space: Popular nearest neighbors in high-dimensional data. *Journal of Machine Learning Research*, 11:2487–2531.

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, Hong Kong, China. Association for Computational Linguistics.

Tim Rocktäschel, Edward Grefenstette, Karl Moritz Hermann, Tomás Kociský, and Phil Blunsom. 2016. Reasoning about entailment with neural attention. In

4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings.

Daniil Sorokin and Iryna Gurevych. 2017. Context-aware representations for knowledge base relation extraction. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1784–1789, Copenhagen, Denmark. Association for Computational Linguistics.

Yuanhe Tian, Guimin Chen, Yan Song, and Xiang Wan. 2021. Dependency-driven relation extraction with attentive graph convolutional networks. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 4458–4471, Online. Association for Computational Linguistics.

Van-Hien Tran, Van-Thuy Phi, Akihiko Kato, Hiroyuki Shindo, Taro Watanabe, and Yuji Matsumoto. 2021a. Improved decomposition strategy for joint entity and relation extraction. *Journal of Natural Language Processing*, 28(4):965–994.

Van-Hien Tran, Van-Thuy Phi, Hiroyuki Shindo, and Yuji Matsumoto. 2019. Relation classification using segment-level attention-based CNN and dependency-based RNN. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2793–2798, Minneapolis, Minnesota. Association for Computational Linguistics.

Vu Tran, Van-Hien Tran, Phuong Nguyen, Chau Nguyen, Ken Satoh, Yuji Matsumoto, and Minh Nguyen. 2021b. CovRelex: A COVID-19 retrieval system with relation extraction. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, pages 24–31, Online. Association for Computational Linguistics.

Yandong Wen, Kaipeng Zhang, Zhifeng Li, and Yu Qiao. 2016. A discriminative feature learning approach for deep face recognition. In *Computer Vision – ECCV 2016*, pages 499–515. Springer International Publishing.

Yongqin Xian, Christoph H. Lampert, Bernt Schiele, and Zeynep Akata. 2019. Zero-shot learning—a comprehensive evaluation of the good, the bad and the ugly. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(9):2251–2265.

Kun Xu, Siva Reddy, Yansong Feng, Songfang Huang, and Dongyan Zhao. 2016. Question answering on Freebase via relation extraction and textual evidence. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2326–2336, Berlin, Germany. Association for Computational Linguistics.