



# A Relational Instance-Based Clustering Method with Contrastive Learning for Open Relation Extraction

Xiaoge Li<sup>1</sup>(✉) , Dayuan Guo<sup>2</sup> , and Tiantian Wang<sup>1</sup> 

<sup>1</sup> School of Computer Science and Technology, Xi'an University of Posts and Telecommunications, Xi'an 710121, China

lixg@xupt.edu.cn

<sup>2</sup> Shaanxi Key Laboratory of Network Data Analysis and Intelligent Processing, Institute, Xi'an 710121, China

**Abstract.** Unsupervised text representations significantly narrow the gap with supervised pretraining, and relation clustering has gradually become an important method of open relational extraction (OpenRE). However, different relational categories generally overlap in the high-dimensional representation space, and distance-based clustering is difficult to separate different categories. In this work, we propose a relational instance-based clustering method with contrastive learning (RICL) - a framework to leverage similarity distribution information and contrastive method to promote better aggregation and relational representation. Specifically, to enable the model to better represent relation instances with word-level features, we construct an augmented dataset using only standard dropout as noise and iteratively optimize the vector representation of relation instances by fully using self-supervised signals. Experiments on real-world datasets show that RICL can achieve excellent performance compared with previous state-of-the-art methods.

**Keywords:** Relation Extraction · Unsupervised Clustering · Contrastive learning

## 1 Introduction

Relation extraction is an important basic work for building large-scale knowledge bases such as semantic networks and knowledge graphs [1–3]. However, conventional relation extraction methods such as semi-supervision and distant supervision are generally used to deal with pre-defined relations and cannot well identify emerging relations in the real world.

Against this background, OpenRE has been widely studied for its ability to mine novel relation from massive text data. At present, OpenRE is mainly based on unsupervised methods, which can be divided into two categories. The first group is pattern extraction models [4–6], which usually uses sentence analysis tools, combined with linguistics and professional domain knowledge, to construct artificial rules based on

lexical, syntactic and semantic features. When performing relation extraction tasks, different relation types are obtained by matching rules with the preprocessed text. However, with the expansion of the relational model set, the complexity of the system is greatly increased, and it is difficult to be widely used in the open field. The second group is to discover various relation types through unsupervised methods [7–9]. This work optimizes the representation of relations to improve the accuracy of unsupervised clustering while overcoming the instability of unsupervised training. Recently, some RE methods work begin to study better utilization of hand-crafted features, which only use named entities to induce relation types [10]. The hierarchy information in relation types is further exploited for better novel relation extraction [11].

However, much research has shown that complex linguistic information requires high-dimensional embeddings so that the meaning of the text becomes clear [12]. This complex information may contain local syntactic [13] and semantic structures [14]. Therefore, the position and relative distance in the high-dimensional vector space is not completely consistent with the relational semantic similarity. Especially before model training starts, even with deep neural networks, different classes may still overlap in high-dimensional space [15].

We propose a relational instance-based clustering method with contrastive learning in this work. In order to make the model better mine the information of the relation instance itself to produce better clustering results, the nonlinear mapping is optimized by using the difference information of the constructed relation instance’s comparative dataset and the distribution information of the original instance dataset. High-dimensional relational instance features of complex information are transformed into relation-oriented low-dimensional feature representations. Specifically, we pull together instances representing the same relationship while pushing apart those from different ones by jointly optimizing distribution loss and contrastive loss so that the learned representation is cluster-friendly. In addition, the proposed method obtains supervision from the data itself and its corresponding augmented dataset and iteratively learns better feature representations for relation classification tasks to improve the quality of supervision, which in turn improves cluster purity and separates distances between different clusters.

Overall, our work has the following contributions: (1) we propose a self-supervised framework which can fine-tune pretrained MLMs into capable universal relational encoders and extensively learn to cluster relational data; (2) we show how to use contrastive learning to learn and improve representations of relation instances in a self-supervised manner.

## 2 Related Work

Self-supervised learning has recently achieved excellent results on multiple tasks in the image and text domains, and many studies have been further developed thanks to its effectiveness in feature representation work. The quality of learned representations is assured by a theoretical framework based on contrast learning [16], which learns self-features from unlabeled data and formalize the concept of semantic similarity through latent classes to improve the performance of classification tasks. Hu et al. [9] propose adaptive clustering algorithms and uses pseudo-labels of relations as self-supervised

signals to optimize their semantic representations. Recently, there has been an increasing interest in contrast learning using individual raw sentences based on PLMs [15, 17, 18].

Meanwhile, inspired by research related to contrast learning in computer vision [19, 20], we utilize “multi-view” contrastive learning for relation extraction. Previous work mainly uses sentences as the smallest unit of text input, builds enhanced datasets by randomly masking characters or replacing words, and uses semantic similarity as the goal of the measurement model. In contrast, our work takes entity word pairs as the minimum granularity of semantic representation, abstracts various types of relations, and obtains their vector representations with the help of the idea of clustering. It not only maintains the advantages of unsupervised learning, which can deal with deal with undefined relation types, but also exerts the advantages of supervised learning, which has a strong guiding ability for relational feature learning.

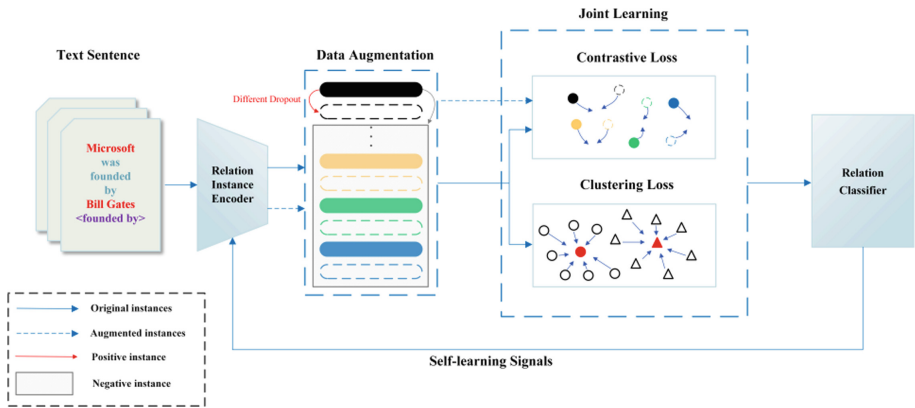


Fig. 1. Overall architecture of RICL

### 3 Methodology

In this work, we propose a simple and effective approach to relation clustering, which exploits relation instance distribution information in unlabeled data and semantic information from pretrained models, enabling the model to optimize the representation of relations.

In order to alleviate the overlap of different relation clusters in the representation space, we improve the clustering of unlabeled data by contrastive learning to promote better separation. The proposed method is shown in Fig. 1.

We build a “multi-view” of the training corpus, gradually optimize the representation of relation instances in a joint learning manner and aggregate to generate pseudo-labels, and fine-tune the pre-trained language model through the classification. As shown in Figure 1., we mainly iteratively perform the following steps:

(1) First, we use the pretrained BERT as the encoder of relational instances  $\{h_i\}_{i=1, \dots, N}$ ; each relational instance  $h_i$  is composed of an entity pair vector as the output

vector. However, high-dimensional representations of  $h$  contain too much information (structural features, semantic information, etc.), and the direct use of high-dimensional vectors for clustering cannot align well with the relationships corresponding to instances.

(2) In order to better reflect the semantic similarity between each other through the distance between the relation representation spaces, we transform the high-dimensional representations of relation instances  $h_i$  into low-dimensional representations  $h'_i$  through non-linear mapping  $g$ . However, the quality of pseudo-labels produced by direct clustering is not high, which is not conducive to downstream classification tasks.

(3) In order to reduce the negative impact of pseudo-label errors, we apply different dropouts under the same pre-training model to construct a positive set and other data under the same batch as a negative set. During the training process, aiming at the aggregation of clusters of similar relational instances and the separation of different instances, the representation of relation instances is optimized to improve the quality of pseudo-labels produced by clustering. Pseudo-labels serve as prior knowledge of the dataset and are finally used for supervised relation classification. The above steps are executed iteratively until the clustering result tends to be stable.

### 3.1 Relational Instance Encoder

The relational instance encoder is to extract the semantic relation representations between two arbitrary given entities in a sentence. We utilize a large pretrained language model to efficiently encode entity pairs and their contextual information.

For sentence  $S = [s_1, \dots, s_n]$ , we introduce two pairs of special identifiers  $[E1\setminus], [\setminus E1], [E2], [\setminus E2]$  to mark entities and inject them to  $S = [s_1, \dots, [E1\setminus], s_i, \dots, s_k, [\setminus E1], \dots, [E2\setminus], s_m, \dots, s_j, [\setminus E2], \dots, s_n]$ . We adopt BERT [21] as our encoder  $I(\bullet)$  due to its strong performance and wide application in extracting semantic information. Formally:

$$v_1, \dots, v_n = I(s_1, \dots, s_n) \quad (1)$$

$$h = [v_{[E1\setminus]}, v_{[E2\setminus]}] \quad (2)$$

where  $v_i$  is a word vector generated by BERT, we use the outputs concatenated by  $v_{[E1\setminus]}$  and  $v_{[E2\setminus]}$  as the representation of the relational instance. This method of relational representation has been widely used in previous RE methods [9, 22, 23].

### 3.2 Instance-Relational Contrastive Learning

We use the distribution information of relation instances and their own feature information to build a joint model to achieve deep clustering. As shown in **Fig. 1**, our joint learning model is composed of two components,  $f(\bullet)$  and  $g(\bullet)$ , using clustering loss and contrastive loss, respectively. We describe the specific structure of the model in Sect. 4.

**Dropout Noise as Data Augmentation.** We use different dropouts to obtain different vector representations of the same text. Specifically, for each batch  $B = \{h_i\}_{i=1}^M$ , we

generate a new vector representation for each relation instance in  $B$  and then get an augmented batch  $B^a = \{h_i, \tilde{h}_i\}_{i=1}^M$ . The positive pair  $h_i, \tilde{h}_i$  takes exactly the same relational instance, and their embeddings only differ in dropout masks, while treating the other  $2M - 2$  instances as negative instances  $N$  of this positive pair. Here the dropout rate  $p$  is 0.1.

Given a batch of data  $B^a$ ,  $\tau$  denotes a temperature parameter. We leverage the standard InfoNCE loss [24] to aggregate the positive pairs together and separate the negative pairs in the embedding space:

$$L_a = - \sum_{i=1}^M \log \frac{\exp(\cos(g(h_i), g(\tilde{h}_i))) / \tau}{\sum_{h_j \in N} \exp(\cos(g(h_i), g(h_j)))} \quad (3)$$

### 3.3 Clustering

Different from contrastive learning, clustering focuses on the similarity between different instances, encodes abstract semantic information into representations of relation instances, and finally aggregates instances of the same relation.

The known dataset consists of  $K$  relation classes. The centroid representation of each class denoted as  $u_k, k \in \{1, \dots, K\}$ . We compute the probability of assigning  $h_i$  to the  $k^{\text{th}}$  cluster by student's t-distribution [25]:

$$q_{ik} = \frac{(1 + \|h_i - u_k\|_2^2 / \alpha)^{-\frac{\alpha+1}{2}}}{\sum_{k'=1}^K (1 + \|h_i - u_{k'}\|_2^2 / \alpha)^{-\frac{\alpha+1}{2}}} \quad (4)$$

Here  $\alpha$  denotes the degree of freedom of the student's t-distribution and  $q_{ik}$  can be regarded as the probability of the cluster assignment. In general, we follow Maaten and Hinton [25] by setting  $\alpha = 1$ .

A linear layer  $f(\bullet)$  is used to fit the centroid of each relation cluster and then iteratively improve it by the auxiliary distribution proposed by Xie et al. [26] Concretely, defining  $p_{ik}$  as the auxiliary probability:

$$p_{ik} = \frac{q_{ik}^2 / f_k}{\sum_{k'} q_{ik'}^2 / f_{k'}} \quad (5)$$

where  $f_k = \sum_{i=1}^M q_{ik}, k = 1, \dots, K$  is the cluster frequency within a batch, the purpose of this is to encourage learning from high-confidence cluster assignments while improving low-confidence tasks against biases caused by imbalanced clusters, resulting in better clustering performance.

We optimize the KL divergence loss between the cluster assignment probability and the target distribution:

$$L_b = KL(P||Q) = \sum_i \sum_k p_{ik} \log \frac{p_{ik}}{q_{ik}} \quad (6)$$

In conclusion, our overall objective is,

$$L = (1 - \varepsilon)L_a + \varepsilon L_b \quad (7)$$

$\varepsilon$  balances between the clustering loss and the contrastive loss of RICL is set to 0.65. Note that  $L_b$  is only optimized on the initial data, and the parameters for  $f(\bullet)$  and  $g(\bullet)$  will be updated-parameters in the  $l(\bullet)$  are not improved when minimizing  $L$ .

Finally, we obtain  $\{h'_i\}_{i=1}^M$  using the optimized  $g(\bullet)$  and  $f(\bullet)$ , and then generate pseudo-labels  $y'$  by k-means algorithm:

$$y' = Kmeans(h') \quad (8)$$

### 3.4 Relation Classification

Based on the pseudo-labels  $y'$  generated by clustering, we can use supervised learning to train the classifier and refine relational instance  $h$  to encode more relational semantic information:

$$l_n = \mu_\tau(l_\theta(S)) \quad (9)$$

$$L_C = \min_{\theta, \tau} \frac{1}{M} \sum_{n=1}^M loss(l_n, one\_hot(y'_n)) \quad (10)$$

where  $\mu_\tau$  denotes the relation classification module parameterized by  $\tau$  and  $I_n$  is a probability distribution over  $K$  pseudo-labels for the original data. In order to find the best-performing parameters  $\theta$  for Relational Instance Encoder and  $\tau$  for Relation Classification, we optimize the above classification loss.

## 4 Experimental Setup

We first introduce publicly available datasets for training and evaluation. Then we briefly introduce the baseline models used for comparison. Finally, we elaborate on the hyperparameter configuration and implementation details of RICL.

### 4.1 Datasets

We conduct experiments and comparisons on three open-domain datasets.

**FewRel.** Few-Shot Relation Classification Dataset is derived from Wikipedia and annotated by humans [27]. FewRel contains 80 types of relations, each with 700 instances. Following the paper [7], we use all instances of 64 relations as training set, and the test set of FewRel, which randomly selects 16 relations with 1600 instances.

**T-REx SPO and T-REx DS.** They come from the T-Rex dataset [28], which is generated by aligning Wikipedia corpus with Wiki-data. At first, we need to preprocess each sentence in the dataset. If there are multiple entity pairs in a sentence, the sentence will be retained for the same number of times according to the number of occurrences of different entity pairs. And then, we built two datasets, T-REx SPO and T-REx DS, according to Hu et al. [9]. In both datasets, 80% of sentences will be used for model training, and the remaining 20% were set aside for validation, the rest for testing.

## 4.2 Baseline and Evaluation Metrics

We use standard unsupervised evaluation metrics for comparisons with the other six baseline algorithms. For all models, we assume the number of target relation classes is known in advance, but no human annotations are available to extract relations from the open-domain data. We set the number of clusters to the number of ground-truth classes and evaluate performance with B3, V-measure, and ARI [8, 9, 29]. To evaluate the effectiveness of our method, we select the following SOTA OpenRE models for comparison.

**VAE** [30] consists of a classifier that predicts relations and a factorization model which reconstructs arguments. The model is jointly optimized by reconstructing entities from pairing entities and predicted relations.

**UIE** [8] trains a discriminative relation extraction model by introducing a skewness loss and a distribution distance loss to make the model confidently predict each relation and encourage the average prediction of all relations.

**SelfORE** [9] uses an adaptive clustering algorithm to obtain relation sets based on a large pretrained language model and then uses the pseudo-labels of relations as self-supervised signals to optimize their semantic representations.

**EL\_ORE** [29] conduct Element Intervention, which intervenes on the context and entities respectively to obtain the underlying causal effects of them, to address the spurious correlations from entities and context to the relation type.

**RW-HAC** [31] reconstructs word embeddings and uses single feature reduction to alleviate the feature sparsity problem for relation extraction through clustering.

**Etype +** [10] consists of two regularization methods and a link predictor and uses only named entity types to induce relation types.

## 4.3 Implementation Details

Follow the settings used in previous work [8, 9, 29, 30], at T-REx SPO and T-REx DS datasets, RICL are trained with 10 relation classes. Although it is lower than the number of real relationships in the dataset, it still reveals important insights due to the very imbalanced distribution of relationships on the 10 relation classes of data used for training and testing.

For Relational Instance Encoder, we use the default tokenizer in BERT to preprocess all datasets and set the max length of a sentence as 128. We use the BERT-base-uncased model to initialize parameters for  $l(\bullet)$  and use BertAdam to optimize the loss.

For Instance-relational Contrastive Learning, we use an MLP  $g(\bullet)$  with fully connected layers with the following dimensions  $\mathbb{R}^d$ -512-512-256. We randomly initialize weights following Xie et al. [26]. For Clustering, we use a linear layer  $f(\bullet)$  of size  $256 \times K$  with  $K$  indicating the number of clusters, and initialize the cluster centers by the Kmean algorithm.

For Relation Classification, we use a fully connected layer as  $\mu_\tau$  and set the dropout rate to 10%, the learning rate to  $5e - 5$ , and the warm-up rate to 0.1. In the process of fine-tuning BERT, we freeze its first 8 layers. All experiments are conducted using an NVIDIA GeForce RTX 3090 with 24GB memory.

## 5 Results and Analysis

In this section, we present the experimental results of RICL on three open-domain datasets, and verify the rationality of the framework through ablation experiments. Finally, we prove its effectiveness by combining data characteristics and visual analysis.

**Table 1.** Main results on three relation extraction datasets.

Dataset	Model	B3			V-measure			ARI
		F1	Prec.	Rec.	F1	Hom.	Comp.	
T-Rex SPO	VAE [30]	24.8	20.6	31.3	23.6	19.1	30.6	12.6
	UIE-BERT [8]	38.1	30.7	50.3	39.1	37.6	40.8	23.5
	SelfORE [9]	41.0	39.4	42.8	41.1	40.3	42.5	33.7
	EI_ORE [29]	45.0	46.7	43.4	45.3	45.4	45.2	36.6
	Our	44.6	42.9	44.4	47.2	46.2	48.2	37.1
T-Rex DS	VAE [30]	9.0	6.4	15.5	5.7	4.5	7.9	1.9
	UIE-BERT [8]	22.4	17.6	30.8	31.2	26.3	38.3	12.3
	SelfORE [9]	32.9	29.7	36.8	32.4	30.1	35.1	20.1
	EI_ORE [29]	42.9	40.2	45.9	47.3	46.9	47.8	25.0
	Our	43.3	41.3	46.6	47.1	47.3	48.6	28.2
FewRel	VAE [30]	36.5	30.9	44.6	47.3	44.8	50.0	29.1
	RW-HAC [31]	33.7	25.6	49.2	43.3	39.1	48.5	25.0
	EType + [10]	31.9	23.8	48.5	40.8	36.4	46.3	24.9
	SelfORE [9]	51.2	50.8	51.6	58.3	57.9	58.8	34.7
	Our	53.9	50.9	57.4	65.3	63.2	67.6	47.3

### 5.1 Main Results

**Table 1** reports model performances on T-Rex SPO, T-Rex DS, and FewRel dataset, which shows that the proposed method achieves state-of-the-art results on the OpenRE task. Benefiting from the rich information in the pre-trained model, RICL exploits the relation distribution in unlabeled data and optimizes the relation representation through the method of contrastive learning so as to achieve a better clustering effect, thus greatly surpassing previous cluster-based baselines.

### 5.2 Ablation Study

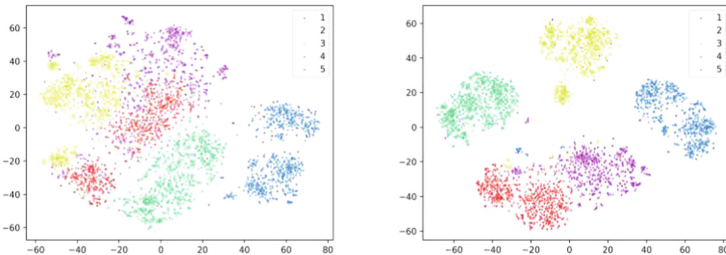
In order to study the effect of each algorithm in the proposed framework, we conduct ablation experiments on two datasets, respectively, and the results are presented in **Table 2**.



**Table 2.** Ablation results on T-Rex SPO and FewRel

Method	T-Rex SPO			FewRel		
	B3-F1	V-F1	ARI	B3-F1	V-F1	ARI
w/o contrastive learning	41.8	43.1	22.9	51.0	57.8	34.1
w/o clustering	39.9	39.8	19.6	47.5	52.0	34.9
w/o classification	42.9	46.6	32.0	51.9	59.2	45.5
RICL	44.6	47.2	37.1	53.9	65.3	47.3

The results show that the model performance is degraded if  $L_a$  is removed, indicating that Instance-relational Contrastive Learning can produce superior relation embeddings from either unlabeled data. It is worth noting that Clustering has an important role in RICL. It prevents the excessive separation of the same relation instance in the space, avoids the collapse of the relation semantic space. At the same time, it provides guidance for downstream relation classification and optimizes the representation of relation instances. In addition, joint optimizing on both the Clustering and the Contrastive Learning is also very important. While alleviating the overlap of different relation classes in the representation space, different instances under the same class are aggregated.

**Fig. 2.** Visualization of feature embeddings on FewRel-5

### 5.3 Visualization and Analysis

To further explore the performance of RICL and the rationality of its design, we randomly select 5 types of data in the FewRel dataset and visualize the embedded features from BERT-base-uncased (left) and RICL (right) with t-SNE in **Fig. 2**. It is convenient for us to observe the changes in class distribution.

In the initial distribution, we observe that classes 2, 3, 4 have high purity, but these classes are not highly clustered and have slight overlap at the boundaries. The relation instances of class 1 and 5 are heavily overlapped in space. Through the analysis of relationship classes and their instances, class 1 describes the “located in” relation between the airport and the place it belongs to, and class 5 describes the “located in” relation between the regional locality and the city or country. These two classes are affected by

factors such as relational semantics and entity types [10], and some relation instances are spatially closely distributed.

From a global perspective, RICL achieves better separation of each class in space, solves the problem of blurred boundaries, ensures the overall consistency, and explores the possibility of further subdividing categories under the same class. While classes 2, 3, 4 are aggregated, they are separated from different class as much as possible in space to ensure semantic consistency. When dealing with class 1 and class 5 overlapping problems, RICL locally aggregates discretely distributed class 5 instances and separates them from class 1 while guaranteeing relational consistency, thereby improving class purity as much as possible.

## 6 Conclusions

In this paper, we propose a novel self-supervised learning framework for open-domain relation extraction, namely RICL. It aims to enable the neural network to obtain better relation-oriented representation encoding and how to better handle relational instances in the open domain in a self-supervised manner. We utilize instance distribution information and contrastive learning to promote better aggregation and relational representation, improving clustering accuracy and reducing error propagation, thus benefiting downstream classification tasks. Moreover, we iteratively improve the robustness of the neural encoder by using pseudo-labels as self-supervised signals for relation classification. Our experiments show that RICL can perform more efficient and accurate relation extraction on open-domain corpora than previous methods, and can construct a representation space more suitable for semantic tasks.

## References

1. Xiong, C., Power, R., Callan, J.: Explicit semantic ranking for academic search via knowledge graph embedding. In: Proceedings of WWW, pp. 1271–1279 (2017)
2. Wang, Z., Zhang, J., Feng, J., Chen, Z.: Knowledge graph embedding by translating on hyperplanes. In: Proceedings of AAAI, pp. 1112–1119 (2014)
3. Dong, L., Wei, F. R., Zhou, M., Xu, K.: Question answering over freebase with multicolumn convolutional neural networks. In: Proceedings of ACL-IJCNLP, pp. 260–269 (2015)
4. Anthony, F., Stephen, S., Oren, E.: Identifying relations for open information extraction. In: Proceedings of EMNLP, pp. 1535–1545 (2011)
5. Jiang, M., Shang, J., Taylor, C., Ren, X., Lance, M., Timothy, P., Han, J.: Metapad: Meta pattern discovery from massive text corpora. In: Proceedings of KDD, pp. 877–886 (2017)
6. Zheng, S., et al.: DIAG-NRE: A neural pattern diagnosis framework for distantly supervised neural relation extraction. In: Proceedings of ACL, pp. 1419–1429. (2019)
7. Wu, R., et al.: Open relation extraction: Relational knowledge transfer from supervised data to unsupervised data. In: Proceedings of EMNLP-IJCNLP, pp. 219–228 (2019)
8. Étienne, S., Vincent, G., Benjamin, P.: Unsupervised information extraction: Regularizing discriminative approaches with relation distribution losses. In: Proceedings of ACL, pp. 1378–1387 (2019)
9. Hu, X., Wen, L., Xu, Y., Zhang, C., Philip Y.: SelfORE: self-supervised relational feature learning for open relation extraction. In: Proceedings of EMNLP, pp 3673–3682 (2020)

10. Tran, T., Le, P., Ananiadou, S.: Revisiting unsupervised relation extraction. In: Proceedings of ACL, pp. 7498–7505 (2020)
11. Zhang, K., et al.: Open Hierarchical Relation Extraction. In: Proceedings of ACL, pp. 5682–5693 (2021)
12. Choudhary, R., Doboli, S., Minai, A.: A Comparative Study of Methods for Visualizable Semantic Embedding of Small Text Corpora. In: Proceedings of IJCNN, pp. 1–8 (2021)
13. Hewitt, J., Manning, C.: A structural probe for finding syntax in word representations. In: Proceedings of NAACL, pp. 4129–4138 (2019)
14. Richie, R., White, B., Bhatia, S., Hout, M.C.: The spatial arrangement method of measuring similarity can capture high-dimensional semantic structures. *Behav. Res. Methods* **52**(5), 1906–1928 (2020). <https://doi.org/10.3758/s13428-020-01362-y>
15. Zhang, D., Nan, F., Wei, X., et al.: Supporting clustering with contrastive learning. In: Proceedings of ACL, pp. 5419–5430 (2021)
16. Arora, S., Khandeparkar, H., Khodak, M., Plevrakis, O., Saunshi, N.: A Theoretical Analysis of Contrastive Unsupervised Representation Learning. arXiv preprint arXiv: 1902.09229 (2019)
17. Liu, F., Vulić, I., Korhonen, A., et al.: Fast, effective, and self-supervised: Transforming masked language models into universal lexical and sentence encoders. In: Proceedings of EMNLP, pp. 1442–1459 (2021)
18. Gao, T., Yao, X., Chen, D.: SimCSE: Simple contrastive learning of sentence embeddings. In: Proceedings of ACL, pp. 6894–6910 (2021)
19. Chen, T., Zhai, X., Ritter, M., Lucic, M., Houlsby, N.: Self-supervised gans via auxiliary rotation loss. In: Proceedings of CVPR, pp. 12154–12163 (2019)
20. Chen, X., Fan, H., Girshick, R., He, K.: Improved Baselines with Momentum Contrastive Learning. arXiv preprint arXiv: 2003.04297 (2020)
21. Devlin, J., Chang, M., Lee, K., Toutanova, K.: BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of NAACL, pp. 4171–4186 (2019)
22. Zhao, J., Gui, T., Zhang, Q., et al.: A Relation-Oriented Clustering Method for Open Relation Extraction. In Proceedings of ACL, pp. 9707–9718 (2021)
23. Wang, Y., Sun, C., Wu, Y., Zhou, H., Li, L., Yan, J.: ENPAR: enhancing entity and entity pair representations for joint entity relation extraction. In Proceedings of EMNLP, pp. 2877–2887 (2021)
24. Oord, A., Li, Y., Vinyals, O.: Representation learning with contrastive pre-dictive coding. arXiv preprint arXiv: 1807.03748 (2018)
25. Maaten, L., Hinton, G.: Visualizing data using t-sne. *J. Mach. Learn. Res.* **9**(86), 2579–2605 (2008)
26. Xie, J., Girshick, R., Farhadi, A.: Unsupervised deep embedding for clustering analysis. In: ICML, pp. 478–487 (2016)
27. Han, X., et al.: Fewrel: a large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation. In: Proceedings of EMNLP, pp. 4803–4809 (2018)
28. Elshahar H., et al.: T-rex: a large scale alignment of natural language with knowledge base triples. In: Proceedings of LREC, pp. 3448–3452 (2018)
29. Liu, F., Yan, L., Lin, H., et al.: Element intervention for open relation extraction. In: Proceedings of ACL, pp. 4683–4693 (2021)
30. Marcheggiani, D., Titov, I.: Discretestate variational autoencoders for joint discovery and factorization of relations. In: Proceedings of TAACL, pp. 231–244 (2016)
31. Elshahar, H., Demidova, E., Gottschalk, S., Gravier, C., Laforest, F.: Unsupervised Open Relation Extraction. In: Blomqvist, E., Hose, K., Paulheim, H., Ławrynowicz, A., Ciravegna, F., Hartig, O. (eds.) *ESWC 2017. LNCS*, vol. 10577, pp. 12–16. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-70407-4\\_3](https://doi.org/10.1007/978-3-319-70407-4_3)