

Learning Discriminative Representations for Open Relation Extraction with Instance Ranking and Label Calibration

Shusen Wang*, Bin Duan*, Yanan Wu*, Yajing Xu
Pattern Recognition & Intelligent System Laboratory,
Beijing University of Posts and Telecommunications, Beijing, China
{shusenw, bobtuan, yanan.wu, xyj}@bupt.edu.cn

Abstract

Open relation extraction is the task to extract relational facts without pre-defined relation types from open-domain corpora. However, since there are some hard or semi-hard instances sharing similar context and entity information but belonging to different underlying relation, current OpenRE methods always cluster them into the same relation type. In this paper, we propose a novel method based on Instance Ranking and Label Calibration strategies (IRLC) to learn discriminative representations for open relation extraction. Due to lacking the original instance label, we provide three surrogate strategies to generate the positive, hard negative, and semi-hard negative instances for the original instance. Instance ranking aims to refine the relational feature space by pushing the hard and semi-hard negative instances apart from the original instance with different margins and pulling the original instance and its positive instance together. To refine the cluster probability distributions of these instances, we introduce a label calibration strategy to model the constraint relationship between instances. Experimental results on two public datasets demonstrate that our proposed method can significantly outperform the previous state-of-the-art methods¹

1 Introduction

Open relation extraction (OpenRE) has been proposed to extract new relational facts where the types of target relations are not pre-defined. Previous methods can be classified into two types: open information extraction (OpenIE) and unsupervised relation discovery. For OpenIE (Yates et al., 2007; Etzioni et al., 2008; Fader et al., 2011), the relations are directly represented by relation phrases

*The first three authors contribute equally. Yajing Xu is the corresponding author.

¹Our code and implementation details are publicly available at <https://github.com/ShusenWang/NAACL2022-IRLC>

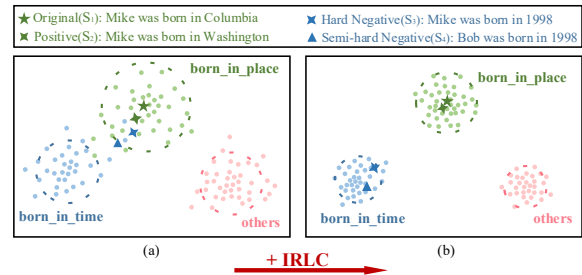


Figure 1: (a) The distribution of baselines in the relational feature space, where the similar context and entities make the hard negative S_3 and semi-hard negative S_4 instance clustered into a wrong relation type. (b) The distribution refined by our method, where these negative instances are separated from the original instance and correctly predicted.

extracted in the sentence. However, the generalization capabilities of these methods are limited since they severely rely on surface-form relations and a relation can be expressed by many surface forms.

Recently, much attention has been focused on unsupervised relation discovery, which is commonly formulated as a clustering task to learn effective relation representations and cluster them (Yao et al., 2011; Marcheggiani and Titov, 2016; Simon et al., 2019). Hu et al. (2020) leverage BERT to extract relational feature and propose a self-supervised framework to learn relation representations from pseudo labels. Because current methods are unstable and easily collapsed (Simon et al., 2019), Liu et al. (2021) solve above-mentioned problems from a causal view and propose element intervention to alleviate the spurious correlations in OpenRE models. However, there are still some hard or semi-hard samples wrongly predicted in the representation space due to the spurious correlations from entities and context to the relation type.

As shown in Figure 1(a), there are two types of negative instances for the relation type *BORN_IN_PLACE*: **Hard negative** and **Semi-hard negative**. For Semi-hard negative instances like S_4 , OpenRE models will assign S_1 and S_4 into

the same relation type *BORN_IN_PLACE* since S_1 and S_4 share similar context information. This problem can be even more severe if the representation space exists some hard negative instances like S_3 , because S_3 possesses a similar context "was born in "and similar entity "Jon" to S_1 . An intuitive way to solve this problem is to refine the relational feature space, as shown in Figure 1(b). Besides, all instances should follow the same relative relationship in the label semantic space which means the original and positive instances have a more similar cluster probability distribution than the hard and semi-hard negative instances. Therefore, it is important to model the constraint relationship between these instances in the label semantic space.

In this paper, we propose a novel method based on **Instance Ranking** and **Label Calibration** strategies (**IRLC**) to better identify the hard and semi-hard negative instances by learning discriminative representations in relational feature and label semantic space simultaneously. However, due to lacking of the instance label, we cannot directly obtain the positive, hard negative and semi-hard negative instances of the original instance. To solve this, we use three data augmentation strategies to generate the positive, hard negative and semi-hard negative instances for the original instance. To refine the relational feature space, we introduce instance ranking to make the original instance close to its positive instance and away from its hard and semi-hard negative instances. To correct the cluster assignment probabilities of hard and semi-hard negative instances, and keep the probability distributions of the original and positive instances aligned, in the label semantic space, Label Calibration strategy is designed to model two constraint relationships between the original and hard negative instance, and between the hard and semi-hard negative instance.

To summarize, the major contributions of our work are as follows: (1) We propose a novel method based on instance ranking and label calibration to learn discriminative representations in relational feature and label semantic space simultaneously. (2) We introduce three surrogate strategies to generate the positive, hard negative and semi-hard negative instances under unsupervised manner. (3) Experimental results show that our proposed method significantly outperforms the previous state-of-the-art models with the improvements of average performance of 11.1% and 11.8%, on two datasets respectively.

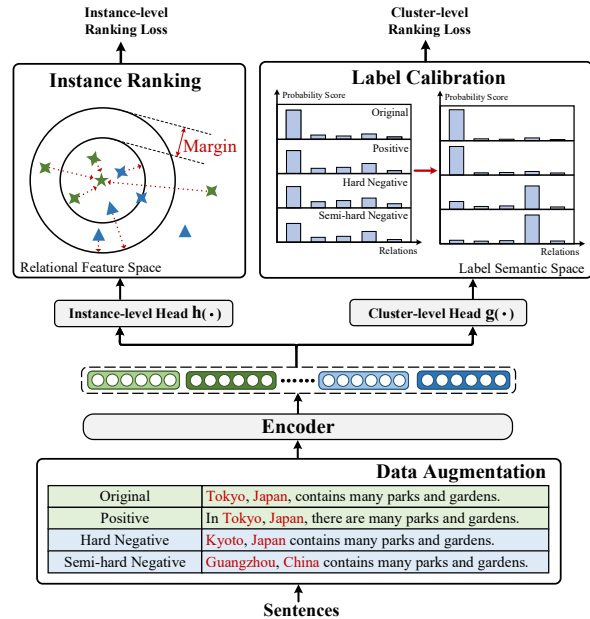


Figure 2: Framework of our proposed method.

2 Method

In this work, we propose a novel method to learn relation representations in feature and semantic space simultaneously. As shown in Figure 2, our method mainly consists of three components: data augmentation, instance ranking, and label calibration modules. We will introduce these module details in the following subsections.

2.1 Data Augmentation

Since there are no pre-defined relation types, it is difficult to directly obtain the positive, hard negative, and semi-hard negative instances of the original instance. To solve this problem, we introduce three surrogate data augmentation strategies to generate above-mentioned instances for the original instance. Specifically, for an original relation instance X_i , we use the following strategies:

Back Translation for Positive: To keep the relation type consistent with the original instance and introduce minimal semantic impact, we use back translation to generate the high-quality positive instance by first translating the original instance to another language and then back to English.

Entity Replacing for Hard Negative: We choose T5 (Raffel et al., 2019) to generate the most similar word to head or tail entity, and then replace the head or tail entity with its augmented word to obtain the hard negative instance, which possesses the similar entity and context to original instance.

Entity Swap for Semi-Hard Negative: To con-

struct a semi-hard negative instance for the original instance, we follow the setting of Entity Swap (Cao and Wang, 2021), which swaps the target entities with other randomly selected entities of the same entity type in the original instance.

2.2 Instance Ranking

After instance construction, we obtain a group of augmented instances of the original instance. Instance Ranking aims to refine the relational feature space. Specifically, given a group of instances $(X_i, X_i^p, X_i^{hn}, X_i^{sn})$, where $X_i^p, X_i^{hn}, X_i^{sn}$ are positive, hard negative, and semi-hard negative instances respectively. We first encode them to obtain their relation representations $(r_i, r_i^p, r_i^{hn}, r_i^{sn})$, and then map these representations into the relational feature space with an instance-level head \mathbf{h} to obtain a group of relational feature $(t_i, t_i^p, t_i^{hn}, t_i^{sn})$. Then we can obtain the instance-level ranking loss:

$$\begin{aligned} \mathcal{L}_i^{IR} = & \max(0, D(t_i, t_i^p) - D(t_i, t_i^{hn}) + m_H) \\ & + \max(0, D(t_i, t_i^{hn}) - D(t_i, t_i^{sn}) + m_S) \end{aligned} \quad (1)$$

where $D(x, y)$ is the euclidean distance between x and y , m_H and m_S are two margins for instance-level ranking loss. Optimized by the objective \mathcal{L}_i^{IR} , model can make the original relation instance closer to its positive instance and away from its correspondingly hard and semi-hard negative instances with different margins.

2.3 Label Calibration

In addition to refining the feature space, we introduce Label Calibration to model the constraint relationship between instances to correct the cluster assignment probabilities of hard and semi-hard negative instances and keep the probability distributions of the original and positive instance aligned in the label semantic space. With a group of relation representations $(r_i, r_i^p, r_i^{hn}, r_i^{sn})$ encoded from their corresponding instances, we first generate the group of cluster representations $(z_i, z_i^p, z_i^{hn}, z_i^{sn})$ by mapping them into the label semantic space with a cluster-level head \mathbf{g} , and then obtain the cluster-level ranking loss:

$$\begin{aligned} \mathcal{L}_i^{LC} = & \max(0, D(z_i, z_i^p) - D(z_i, z_i^{hn})) \\ & + \max(0, D(z_i, z_i^{hn}) - D(z_i, z_i^{sn}) + m_L) \end{aligned} \quad (2)$$

where $D(x, y)$ is the KL distance between x and y to measure the difference between the cluster assignment probabilities of the instances, m_L is the

margin for cluster-level ranking loss. The first term is to model the constraint relationship between the original and hard negative instance, and the second term is to the constraint relationship between the hard and semi-hard negative instance. The final loss function is as follows:

$$\mathcal{L} = -\frac{1}{n} \sum_{i=1}^n (\mathcal{L}_i^{IR} + \mathcal{L}_i^{LC}) \quad (3)$$

3 Experiments

3.1 Datasets

To assess the performance of our method, we conduct experiments on **T-REx SPO** and **T-REx DS**, which both come from T-REx² (Elsahar et al., 2018) but differ in whether having surface-form relations or not. Following the setup of Liu et al. (2021), we use 80% of instances for model training and 20% for validation on both two datasets.

3.2 Baselines

For comparison, we consider the following baselines:

- **rel-LDA** A generative method proposed by Yao et al. (2011), which treats unsupervised relation discovery as a topic model. In our experiment, we choose the full rel-LDA to compare with our method.
- **March** A method (Marcheggiani and Titov, 2016) based on self-supervised signal from entity link predictor to learn a VAE-based model.
- **UIE** A method proposed by Simon et al. (2019) to solve instability and use two regularization to train a discriminative model for OpenRE. In our experiments, we compare our method with two versions of UIE, which only differ in the relation encoding model, i.e., PCNN and BERT.
- **SelfORE** A self-supervised framework propose by Hu et al. (2020), which learn contextual relation representations from pseudo labels.
- **Element Intervention** A method proposed by Liu et al. (2021), which formulates OpenRE by using a structural causal model.

²<https://hadyelsahar.github.io/t-rex/>

Dataset	Model	B^3			V-measure			ARI	Avg.
		F1	Prec.	Rec.	F1	Homo.	Comp.		
T-REx SPO	rel-LDA-full (Yao et al., 2011)	18.5	14.3	26.1	19.4	16.1	24.5	8.6	15.5
	March (Marcheggiani and Titov, 2016)	24.8	20.6	31.3	23.6	19.1	30.6	12.6	20.3
	UIE-PCNN (Simon et al., 2019)	36.3	28.4	50.3	41.4	33.7	53.6	21.3	33.0
	UIE-BERT (Simon et al., 2019)	38.1	30.7	50.3	39.1	37.6	40.8	23.5	33.6
	SelfORE (Hu et al., 2020)	41.0	39.4	42.8	41.4	40.3	42.5	33.7	38.7
	Element Intervention (Liu et al., 2021)	45.0	46.7	43.4	45.3	45.4	45.2	36.6	42.3
	IRLC	57.4	77.1	45.7	60.4	71.1	52.5	42.3	53.4
	IRLC w/o instance ranking	53.8	68.5	44.3	51.0	58.9	44.9	36.6	47.1
IRLC w/o label calibration	50.9	65.4	41.7	46.5	54.0	40.8	40.5	46.0	
T-REx DS	rel-LDA-full (Yao et al., 2011)	12.7	8.3	26.6	17.0	13.3	23.5	3.4	11.0
	March (Marcheggiani and Titov, 2016)	9.0	6.4	15.5	5.7	4.5	7.9	1.9	5.5
	UIE-PCNN (Simon et al., 2019)	19.7	14.0	33.4	26.6	20.8	36.8	9.4	18.6
	UIE-BERT (Simon et al., 2019)	22.4	17.6	30.8	31.2	26.3	38.3	12.3	22.0
	SelfORE (Hu et al., 2020)	32.9	29.7	36.8	32.4	30.1	35.1	20.1	28.5
	Element Intervention (Liu et al., 2021)	42.9	40.2	45.9	47.3	46.9	47.8	25.0	38.4
	IRLC (ours)	58.5	77.1	47.2	47.0	58.1	39.4	45.0	50.2
	IRLC w/o instance ranking	46.5	76.4	33.4	42.1	57.0	33.4	28.8	39.1
IRLC w/o label calibration	46.8	73.6	34.2	38.9	52.3	31.0	28.3	38.0	

Table 1: Experimental results(%) produced by the baseline models and the proposed model IRLC on T-REx SPO and T-REx DS in terms of B^3 , V-measure, ARI.

3.3 Evaluation Metrics

As the previous work (Simon et al., 2019; Hu et al., 2020; Liu et al., 2021), we adopt B^3 (Bagga and Baldwin, 1998), V-measure (Rosenberg and Hirschberg, 2007), and Adjusted Rand Index (ARI) (Hubert and Arabie, 1985) to evaluate different methods. Considering that any of the three metrics can measure the clustering performance from different angles, we take the average of B^3 F1, V-measure F1 and ARI for comprehensive evaluation.

3.4 Implementation Details

For fair comparison, all model are trained and evaluated on 10 relation types, same as (Simon et al., 2019; Hu et al., 2020; Liu et al., 2021). We implement our model in PyTorch³ (Paszke et al., 2017) with transformers package⁴ (Wolf et al., 2020). We adopt bert-base-cased as backbone to generate contextual relation representations. The output size of the instance-level head in instance ranking is 128, while the size is set to 10 in the cluster-level head for label calibration, same as the number of relation types. We use Adam (Kingma and Ba, 2014) as optimizer with a learning rate of 1e-5 for backbone and a learning rate of 1e-3 for two heads. The max length of input sentence is 96 and the batch size is 32. All experiments are conducted by using a GeForce RTX 3090Ti with 24 GB memory.

³<https://pytorch.org/>

⁴<https://github.com/huggingface/transformers>

3.5 Main Results

We summarize the performances of the baselines and our method in Table 1. From the experimental results, we can see that our method IRLC significantly outperforms baselines by a large margin and achieves new state-of-the-art results on both two datasets. For T-REx SPO, compared with the previous SOTA model, IRLC improves the average performance by 11.1%, B^3 F1-score by 12.4%, V-measure F1-score by 15.1%, and ARI by 5.7%. The results confirm IRLC can learn discriminative representations to help model extract novel relations. For T-REx DS, our method IRLC outperforms the SOTA model with an average performance gain of 11.8%, proving the effectiveness of IRLC for OpenRE.

3.6 Ablation Study

To study the effect of instance ranking and label calibration in the proposed method, we conduct ablation experiments on two datasets and report the results in Table 1. We find that the performance of IRLC will severely degrade without instance ranking or label calibration. It proves both two strategies proposed in our method are important and effective, and combining these two strategies can achieve a noticeable performance gain. More specifically, we can observe that instance ranking or label calibration is effective enough to outperform previous SOTA models with an average performance gain of at least 3.7% in T-REx

Model		Min	Max	Mean	Median
T-REx SPO	SelfORE	0.38	0.84	0.63	0.67
	IRLC	0.16	0.37	0.27	0.28
T-REx DS	SelfORE	0.49	0.90	0.69	0.62
	IRLC	0.25	0.54	0.41	0.40

Table 2: The intra-class variance statistics between SelfORE and our proposed method.

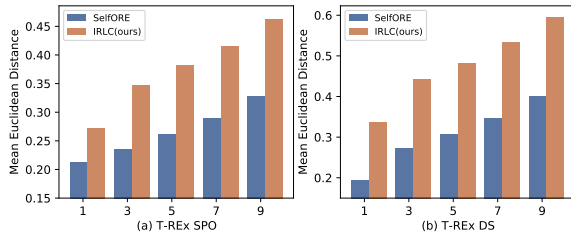


Figure 3: Inter-class distance statistics with different number of the nearest class centers.

SPO dataset, showing the effectiveness of these two strategies.

3.7 Qualitative Analysis

In this section, we first analyse the representation distribution of novel relations on two datasets from two perspectives, intra-class and inter-class, to study how our method refines the representation space. And then we visualize the representations of novel relations to show the effectiveness of our method.

IRLC leads to smaller intra-class distance. Table 2 shows the intra-class variance statistics. Specifically, we use intra-class variance to indicate the intra-class distance of relation type. Each cluster intra-class variance is obtained by calculating the average variances of all normalized relation representations corresponding to the same relation type, and we report the min/max/mean/median variance values on all relation types. From the results, we can see that the intra-class variance are much smaller than compared method in four aspects. It confirms IRLC can make the relation representations from same relation type closer.

IRLC leads to larger inter-class distance. Figure 3 shows the inter-class distance statistics. The X-axis is the number of the nearest class centers. We obtain the euclidean distances between each class center and its nearest class centers with different number, and then average these distances of all relation types as the inter-class distance. From the results, we can observe that IRLC significant increases the inter-class distance with different num-

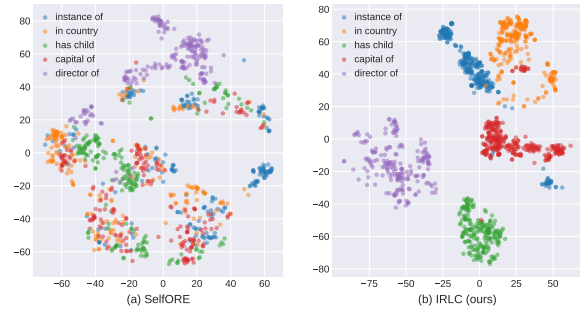


Figure 4: Visualization of the relation representations on T-REx SPO.

ber of the nearest class centers, especially in T-REx SPO. In summary, IRLC can obtain a better relation representation space with smaller intra-class distance and larger inter-class distance for OpenRE.

Visualization of Relation Representations. To intuitively show how our method helps to refine the relation representation space, we visualize the representations of novel relations by using t-SNE (Van der Maaten and Hinton, 2008) to reduce the dimension to 2. We randomly choose 5 relations and sample 200 instances in each relation. As shown in Figure 4(a), the relation representation space of compared model is chaotic and somewhat dense. However, the relation representations from different types are mostly separated in our proposed method, as shown in Figure 4(b).

4 Conclusion

In this paper, we propose a novel method based on instance ranking and label calibration (IRLC) to learn discriminative representations for better identifying the hard and semi-hard negative instances, in the relational feature and label semantic space simultaneously. Due to lacking the label of each instance, we introduce three surrogate strategies to generate the augmented views for the original instance. And then instance ranking is used to refine the relational feature space, and label calibration is designed to model the constraint relationship between instances. Experiments and analysis confirm the effectiveness of IRLC for OpenRE.

Acknowledgements

We thank all anonymous reviewers for their valuable suggestions. This work was supported by the National Natural Science Foundation of China (NSFC No.62076031).

References

- Amit Bagga and Breck Baldwin. 1998. [Entity-based cross-document coreferencing using the vector space model](#). In *36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 1*, pages 79–85, Montreal, Quebec, Canada. Association for Computational Linguistics.
- Shuyang Cao and Lu Wang. 2021. Cliff: Contrastive learning for improving faithfulness and factuality in abstractive summarization. *arXiv preprint arXiv:2109.09209*.
- Hady Elsahar, Pavlos Vougiouklis, Arslan Remaci, Christophe Gravier, Jonathon Hare, Frederique Laforest, and Elena Simperl. 2018. [T-REx: A large scale alignment of natural language with knowledge base triples](#). In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).
- Oren Etzioni, Michele Banko, Stephen Soderland, and Daniel S Weld. 2008. Open information extraction from the web. *Communications of the ACM*, 51(12):68–74.
- Anthony Fader, Stephen Soderland, and Oren Etzioni. 2011. [Identifying relations for open information extraction](#). In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 1535–1545, Edinburgh, Scotland, UK. Association for Computational Linguistics.
- Xuming Hu, Lijie Wen, Yusong Xu, Chenwei Zhang, and Philip Yu. 2020. [SelfORE: Self-supervised relational feature learning for open relation extraction](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3673–3682, Online. Association for Computational Linguistics.
- Lawrence Hubert and Phipps Arabie. 1985. Comparing partitions. *Journal of classification*, 2(1):193–218.
- Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Fangchao Liu, Lingyong Yan, Hongyu Lin, Xianpei Han, and Le Sun. 2021. [Element intervention for open relation extraction](#). 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 4683–4693, Online. Association for Computational Linguistics.
- Diego Marcheggiani and Ivan Titov. 2016. [Discrete-state variational autoencoders for joint discovery and factorization of relations](#). *Transactions of the Association for Computational Linguistics*, 4:231–244.
- Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. 2017. Automatic differentiation in pytorch. In *NIPS-W*.
- Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2019. Exploring the limits of transfer learning with a unified text-to-text transformer. *arXiv preprint arXiv:1910.10683*.
- Andrew Rosenberg and Julia Hirschberg. 2007. [V-measure: A conditional entropy-based external cluster evaluation measure](#). In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 410–420, Prague, Czech Republic. Association for Computational Linguistics.
- Étienne Simon, Vincent Guigue, and Benjamin Piwowarski. 2019. [Unsupervised information extraction: Regularizing discriminative approaches with relation distribution losses](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1378–1387, Florence, Italy. Association for Computational Linguistics.
- Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-sne. *Journal of machine learning research*, 9(11).
- Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. [Transformers: State-of-the-art natural language processing](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.
- Limin Yao, Aria Haghighi, Sebastian Riedel, and Andrew McCallum. 2011. [Structured relation discovery using generative models](#). In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 1456–1466, Edinburgh, Scotland, UK. Association for Computational Linguistics.
- Alexander Yates, Michele Banko, Matthew Broadhead, Michael Cafarella, Oren Etzioni, and Stephen Soderland. 2007. [TextRunner: Open information extraction on the web](#). In *Proceedings of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT)*, pages 25–26, Rochester, New York, USA. Association for Computational Linguistics.