ONTOLOGY-GUIDED AND TEXT-ENHANCED REPRESENTATION FOR KNOWLEDGE GRAPH ZERO-SHOT RELATIONAL LEARNING

Ran Song[1], **Shizhu He**[2], **Suncong Zheng**[3],
Shengxiang Gao[1], **Kang Liu**[2], **Jun Zhao**[2], **Zhengtao Yu**[1*]
[1]Faculty of Information Engineering and Automation,
Yunnan Key Laboratory of Artificial Intelligence,
Kunming University of Science and Technology, Kunming, China
[2]National Laboratory of Pattern Recognition,
Institute of Automation Chinese Academy of Sciences, Beijing, China
[3]Tencent AI Lab
song_ransr@163.com,congzheng@tencent.com,ztyu@hotmail.com
gaoshengxiang.yn@foxmail.com,{shizhu.he,kliu,jzhao}@nlpr.ia.ac.cn

ABSTRACT

Knowledge graph embedding (KGE) have been proposed and utilized to knowledge graph completion (KGC), but most KGE methods struggle in unseen relations. Previous studies focus on complete zero-shot relational learning by incorporating text-features and proximity relations, which are difficult to accurately represent the complete semantic of relations. To overcome the above-mentioned issues in zero-shot relation learning, we propose an ontology-guided and text-enhanced representation, which could improve the effect of current KGE for unseen relations. In fact, each KG contain ontology and text descriptions that describe the meta-information of knowledge. To combine text-embedding space and graph-embedding space, we design TR-GCN to obtain the meta-representation of relations based on the ontology structure and their textual descriptions. It will be used directly to guide previous KGE methods such as TransE and RotatE on zero-shot relation learning. The experimental results on multiple public datasets demonstrate that the proposed ontology-guided and text-enhanced representation can enrich KGs embedding, and significantly improves the KGC performance on unseen relations.

1 INTRODUCTION

KGs (Knowledge Graphs) such as Freebase Bollacker et al. (2008), DBpedia Lehmann et al. (2015) and YAGO Mahdisoltani et al. (2014) contain large amounts of entities, relations and triples, but the incompleteness of those KGs is an urgent issue for its wide utilization. Recently, embedding based methods Nickel et al. (2011); Bordes et al. (2013); Sun et al. (2019) have been proposed and widely used in knowledge graph completion (KGC), which attempts to embed a KG into a low-dimensional continuous space. And the numerical representations (e.g., vectors) of entities and relations can be used to reason potential facts.

However, embedding-based methods are struggle in completing facts involving unseen entities and relations. The main reason is that current embedding based methods merely relied on the inter-linked triples of entities and relations in the factual graph Zhang et al. (2021), therefore these method cannot learn any effective representations for unseen entities and relations. For example, as illustrated in Figure 1, the unseen relation "*academicAdvisor*", which does not appear in the

*Corresponding author

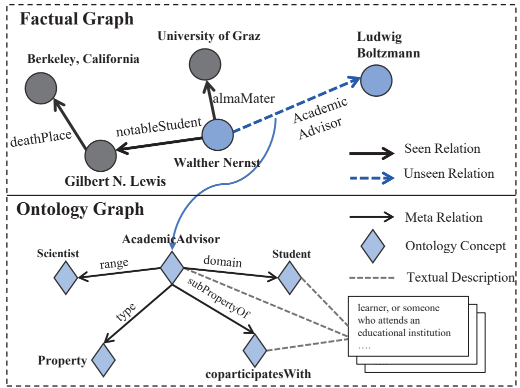

Figure 1: Knowledge graph consists of relatively fixed ontology (the lower part) and frequently evolving facts (the upper part). The knowledge completion involving unseen relations, which struggle in previous embedding methods, could be alleviated by utilized their structure information and textual descriptions in the ontology.

observed facts, is represented as a random initial tensor (e.g., vector) and cannot be predicted at all. In fact, with the expansion of knowledge, new entities and relations are constantly emerging from the real world and should be encoded (added and updated) in the knowledge graph. Therefore, how to model unseen entities and relations, and conduct zero-shot learning is an important task of KGC.

To overcome the above mentioned challenge, we propose to use the ontology (as shown in Figure 1) to learn the semantic representation of unseen relations. In fact, as a universal knowledge description framework, KG represents ontology and facts in a unified form (inter-linked triplets) Hogan et al. (2020). A KG usually contains a relatively fixed ontology and frequently updated facts, the former describes concepts and properties that are meta-information of entities and relations, and the latter encodes the evolving entities and their relations. For unseen relations, the ontology express their relationships with other concepts and properties through meta-relations such as "*domain*", "*range*" and "*subPropertyOf*". And we can obtain accurate representations of unseen relation based on the structure of ontology graph. As shown in Figure 1, although the unseen relation "*academicAdvisor*" was not observed, its semantic representation can still be obtained by other concepts and properties such as "*(academicAdvisor, domain, Student)*", "*(academicAdvisor, range, Scientist)*" and "*(academicAdvisor, subPropertyOf, coparticipatesWith)*". The ontology constructs a semantic association between unseen relations and seen information (relation and entity). In addition, the textual descriptions will also enrich representation of unseen relations. Therefore, we can obtain accurate and rich representation of unseen relations through learning the semantic transform from their related ontology elements (such as concepts, properties, etc.) and textual descriptions.

In this paper, we propose an ontology-guided learning method for zero-shot relation learning, which could represent the unseen relations by fusing ontology graph-embedding spaces and description text-embedding spaces. Inspire by RGCN Schlichtkrull et al. (2018), we design Text-Relation Graph Convolution Network (TR-GCN) to obtain the meta-representation of relations based on the ontology graph structure of them inter-lined by meta-relations and the textual descriptions of ontology node. The representation of relations (especially for the unseen relations) can be supplemented by the meta-representation of the TR-GCN outputs. We adapt TransE, RotatE and other mainstream embedding-based methods as base models, and the proposed method can be used in combination with most of KGE methods.

In short, our main contributions are as follows:

- We propose a zero-shot relation learning to obtain the rich and accurate semantic representations for unseen relations, which model ontology graph and textual description in a unified graph neural network. The ontology graph establish connection between the unseen relations and other related ontology elements. The textual description makes rich for semantic representation of unseen relations.

- We design Text-Relation Graph Convolution Network (TR-GCN) to obtain the meta-information of relation through adding text information during message passing of GNN.
- We implement our method with some mainstream methods such as TransE and RotatE. And the experimental results show that our method is effective for representation unseen relations. It also significantly improves the KGC performance on the zero-shot relational learning.

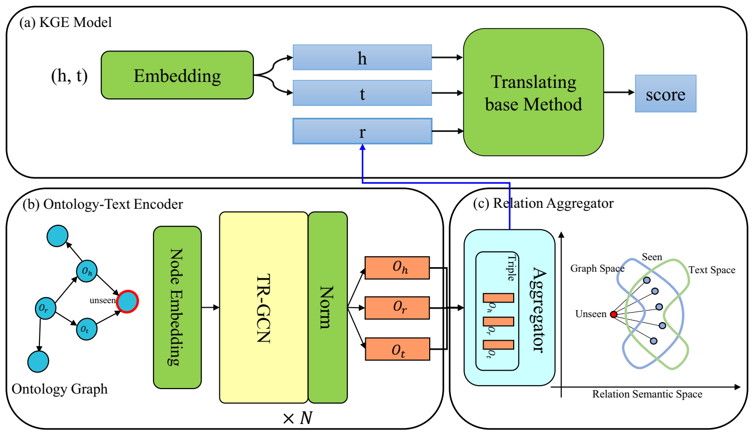

Figure 2: The proposed method contains three core parts. (a) The traditional embedding-based method, which does not contain a proper relation embedding for unseen relations. therefore we represent unseen relations by Ontology-Text Encoder (b) and Relation Aggregator (c). (b) Ontology-Text Encoder, which includes node embedding layer and TR-GCN layer. All node (including concepts and properties) representations can be trained directly and represented by its neighbors. (c) The Relation Aggregator, three different type nodes of ontology are used to obtain the relation representations.

2 Related Work

2.1 Knowledge Graph Embedding

Recently, massive work focused on embedding-based method for knowledge graph completion Zhang et al. (2021). The key issue of knowledge graph embedding is to learn low dimensional distributed embedding of entities and relations Ji et al. (2021); Bordes et al. (2013); Sun et al. (2019).This paper proposes and integrates the ontology into knowledge graph embedding, and the proposed method can be used in combination with most of KGE methods.

2.2 Zero-shot Learning for KGC

In the area of knowledge graph completion, more work focus on zero-shot entity learning which devote to deal with unseen entity. While few works consider zero-shot relation learning and model unseen relations. The limited works take text-embedding spaces as semantic spaces of relation to represent unseen relations Qin et al. (2020); Geng et al. (2021).And Zhang et al. (2020) design a classifier-based method, which select an appropriate seen relation to replace the unseen relation. Our work focuses on unseen relation in knowledge graph completion, propose a method that incorporate ontology graph and textual description to leaning the unseen relation representations.

3 Method

In this section, we introduce our proposed method as shown in Figure 2. Our method improves effects of previous KGE models for unseen relations by making rich and accuracy to their representation. KGC aims at scoring a triple (h_k, r_k, t_k) from KG (Factual Graph, contains inter-linked

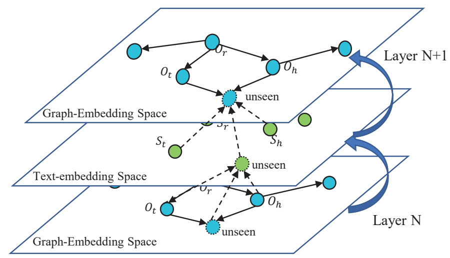

Figure 3: Text-Relational Graph Convolutional Network (TR-GCN) add text-embedding information in message passing round.

factual triples) $\mathcal{G}_f = (\mathcal{R}_f, \mathcal{E}_f)$, where $r_k \in \mathcal{R}_f$ is relation and $h_k, t_k \in \mathcal{E}_f$ are entities. Ontology also describe as a graph (Ontology Graph , contains mete-relations among concepts and properties) $\mathcal{G}_o = (\mathcal{R}_o, \mathcal{E}_o)$, which uses meta-relations to associate between ontology nodes (concepts and properties). Each factual triple corresponds to ontology nodes (h_o, r_o, t_o), where $h_o, r_o, t_o \in \mathcal{E}_o$ are the ontology nodes corresponding the head and tail entities and relation. For example, "*LeBron James*" and "*Kevin Durant*" in factual graph are mapped to "*BasketballPlayer*" in ontology graph. And the relation $\mathcal{R}_f$ and entities $\mathcal{E}_f$ of factual graph all find own meta information in ontology.

3.1 Text-Relational Graph Convolutional Network

Ontology is the skeleton of KGs, which provide meta-data descriptions for guiding the knowledge graph construction and completion. In addition, the descriptions of relation contain semantic information which can also be used to represent relations.

Inspired by RGCN Schlichtkrull et al. (2018), we design Text-Relational Graph Convolutional Network (TR-GCN) model to combine ontology graph and textual descriptions, as shown in Figure 3. We add text embedding of ontology nodes during message passing of GNN. First, we embed all nodes into node embedding and text embedding as follow:

$$\mathbf{h}_o = h_o \mathbf{W}_o^E \tag{1}$$

$$\mathbf{h}_s = \sum_{i \in N_w} x_i \mathbf{W}_s^E \tag{2}$$

where h_o is the concept in the ontology, which map the head entity in factual graph, x_i is a word of node and $\mathbf{W}_s^E$ is word embedding come from pretrained model (GloVe), $\mathbf{W}_o^E$ is node embedding and $\mathbf{h}_o^0$ is initialization of TR-GCN. Cause t_o and r_o have same representation process, we only display the h_0.

The TR-GCN has high effective at aggregating and encoding features from structured neighborhoods and semantic information, and gain high performance on ontology graph. In this paper, the following TR-GCN is used to obtain the node representation of ontology graph and textual description:

$$\begin{aligned}\mathbf{h}_{o,i}^{(l+1)} = ReLU(&\sum_{r \in \mathcal{R}} \sum_{j \in \mathcal{N}_i^r} \frac{1}{C_{i,r}} \mathbf{W}_r^{(l)} \mathbf{h}_{o,j}^{(l)} \\ &+ \mathbf{W}_0^{(l)} \mathbf{h}_{o,i}^{(l)} + \mathbf{W}_S \mathbf{h}_{s,i})\end{aligned} \tag{3}$$

$$\mathbf{h}_{o,i}^{(l+1)} = Norm\_Layer(\mathbf{h}_{o,i}^{(l+1)}) \tag{4}$$

where $\mathbf{h}_{o,i}^{(l)} \in \mathbb{R}^d$ is hidden state of ontology node $h_{o,i}$ in the l-th layer of TR-GCN, and d is dimension of layer's representation. $\mathcal{N}_i^r$ denotes the set of neighbor indices of node i under meta-relation $r \in \mathcal{R}_o$. $\mathbf{W}_r^{(l)}$ is relation parameters of meta-relation r which weight for node i neighboring node in l-th layer. $\mathbf{W}_0^{(l)}$ is self-loop weight for encoding self-node features. $c_{i,r}$ is a normalization constant that can either be learned or chosen in advance. $\mathbf{W}_S$ is semantic weight, and we fixed text-embedding representation in all layer. ReLU is activate function. We also use layer normalization to speed the training.

3.2 RELATION AGGREGATING

The core issue of this paper is how to obtain the effective representation of the relation r in factual triple (h_k, r_k, t_k), especially for the triple involving unseen relations. For each factual triple (h_k, r_k, t_k), we first obtain their mapped ontology nodes (h_o, r_o, t_o). h_o and t_o are the concepts of two entities h_k and t_k, and r_o is the relation h_k in ontology. It should be noted that h_o, r_o and t_o are all node in the ontology graph. We select the features $(\mathbf{h}_o^L, \mathbf{r}_o^L, \mathbf{t}_o^L)$ corresponding to ontology nodes (h_o, r_o, t_o) from outputs of TR-GCN $\mathbf{H}_o^L \in \mathbb{R}^{d*N}$, where N is number of ontology nodes, L is the number of ontology encoder layer. Based on the node representations, we design aggregating strategies with ontology graph-embedding space and text-embedding space.

In order to make full use of all aspects of information, we leverage three ontology nodes representations to represent relations of KG:

$$\mathbf{r}_a = [\mathbf{h}_o^L : \mathbf{r}_o^L : \mathbf{t}_o^L] \cdot \mathbf{W}^A \tag{5}$$

where $\mathbf{W}^A \in \mathbb{R}^{3d*d}$ is transformation matrix.

The above aggregator leverage graph-embedding and text-embedding semantic space. Intuitively, a node representation is computed by neighboring nodes, it ensures that the representation of the nodes are all in the same space. With the enrichment of text-embedding and graph-embedding space, our method can deal with that the relation that disappear in training can be represented, this can be described as:

$$\mathbf{r}_a = \sum_{n \in \mathcal{N}} g(\mathbf{h}) \tag{6}$$

where $g(*)$ is represent the aggregating function, $\mathcal{N}$ is neighboring nodes for missing node.

Following previous KGE models, we train our model with the margin-based ranking loss, and use a negative sampling loss function for effectively optimizing ranking loss :

$$\begin{aligned}\mathcal{L} = &- log\sigma(\gamma - f(h_k\mathbf{W}^E, \mathbf{r}_a, t_k\mathbf{W}^E)) \\ &- \sum_{i=1}^{n} \frac{1}{k} log\sigma(f(h_i'\mathbf{W}^E, \mathbf{r}_a, t_i'\mathbf{W}^E) - \gamma)\end{aligned} \tag{7}$$

where $\mathbf{W}^E$ indicates entity embedding. The relation utilized to obtain the relation representation of triple (h_k, r_k, t_k)/(h_k', r_k, t_k'), γ is a fixed margin, σ is the sigmoid function, and (h_i', r_k, t_i') is the corresponding negative triple. The loss function can sample multiple negative triples for each positive triple at once.

4 EXPERIMENTS

We conduct extensive experiments with KGC task on several public dataset, and mainly evaluate the performance of proposed framework on zero-shot relational learning.

4.1 DATASET

We select datasets from four public knowledge graphs, DBpedia, NELL, YAGO, and Wikidata, to evaluate the effectiveness of KGC on zero-shot relational learning. The current benchmark datasets contain only factual graph and not ontology graph. Therefore, we extract ontology from their origin websites[1][2][3]. Generally, we collect series ontology: **DBpedia** have human-created high-quality ontology, which has 26,586 nodes and 8 meta-relations. The ontology of **NELL** has 1,494 nodes, 6,907 triples and 14 meta-relations (e.g. *antisymmetric*, *mutexpredicates*). And ontology of **YAGO** has 654 nodes, 2,452 triples and 28 meta-relations (e.g. *causes*, *synonym*). It should be noted that

[1] `https://www.dbpedia.org/resources/ontology/`

[2] `http://resources.mpi-inf.mpg.de/yago-naga/yago3.1/yagoSchema.tsv.7z`

[3] `http://rtw.ml.cmu.edu/resources/results/08m/NELL.08m.1115.ontology.csv.gz`

Model	NELL-ZS		Wiki-ZS		DB100K-ZS			
	UNSEEN		UNSEEN ♣		UNSEEN		SEEN	
	MRR	Hits@10	MRR	Hits@10	MRR	Hits@10	MRR	Hits@10
DistMult♣	23.50	32.60	18.90	23.60	5.53	10.44	10.13	22.97
TransE♣	9.70	20.30	5.30	11.90	2.64	9.18	17.55	43.34
$ZSGAN_{KG}$(DistMult)♣	25.30	37.10	20.80	29.40	-	-	-	-
$ZSGAN_{KG}$(TransE) ♣	24.00	37.60	18.50	26.10	-	-	-	-
OntoZSL(DistMult) ♣	25.60	38.50	21.10	28.90	-	-	-	-
OntoZSL(TransE) ♣	25.00	39.90	18.40	26.50	-	-	-	-
TransE+GRL	-	-	-	-	4.42	12.72	16.72	40.48
DistMult+S	8.08	12.53	8.87	13.57	6.72	10.91	10.18	23.97
TransE+S	6.41	16.34	22.09	31.45	14.75	30.36	15.31	40.06
TransE+GA	28.79	45.63	26.32	36.55	23.12	40.43	**27.96**	**52.09**
TransE+GTA	**31.80**	**50.61**	**26.77**	**37.07**	**26.58**	**44.65**	27.79	51.80

Table 1: Zero-shot relation learning results on NELL-ZS, Wiki-ZS and DB100K-ZS. Seen Relation is that relation of triples exist in training. The results of ♣ consider candidate sets, which are constructed by using the entity type constraint Qin et al. (2020) Toutanova et al. (2015). The other results match whole entity set. **Bold** numbers denote the best results for dataset.

Wikidata has no publish ontology by official, we collect 20,899 triples including 8,907 nodes and 604 meta-relations (e.g. *instance of* (P31), *see also* (P1659)) as their ontology from the released dump data [4].

4.2 Implementation Details

In our experiments, we adopts the following embedding-based methods because of their efficiency and effectiveness on link predictions: DistMult, Complex, TransE and RotatE. Our codes are based on Sun et al. (2019) and adopt the PyTorch Paszke et al. (2017) framework. For TR-GCN, we used the implementation in the deep graph library (DGL). The initial word embedding are come from GloVe Pennington et al. (2014). The entity embedding size is set to 100 for all embedding-based methods. The TR-GCN hidden size is set to 100, the number of layer is set to 2 and use self-loop for each node. We selected the hyperparameters corresponding to learning rate and batch size from $\{0.0001, 0.0005, 0.001\}$ and $\{128, 256, 512, 1024\}$. And we use Adam to optimize all the parameters.

4.3 Zero-Shot (Unseen) Relation Learning

The unseen relations denote that relation of the triples in the test set but do not appear in the training set. Previous embedding-based models are transductive inference methods , and cannot deal with those relations because they can only learn the representation of the relations that appears in the training set. Table 1 show the result on NELL-ZS, WiKi-ZS and DB100K-ZS, which the testing set of NELL-ZS and WiKi-ZS are all unseen relations Qin et al. (2020).

We select some baseline models to compare. DistMult+S and TransE+S are to replace relation by text embedding of relation. The GRL Zhang et al. (2020) is the classifier-based methods and hard to solve massive unseen relation and only unseen relation in test dataset. ZSGAN Qin et al. (2020) and OntoZSL Geng et al. (2021) always generate represent for relation, therefore it is hard to keep traditional embedding-based method performance in the seen dataset, and them do not work in DB100K-ZS which match whole entity set. DistMult and TransE have same tendency, we show the better one. In general, our method base on ontology can alleviate this issue through learning common node information between seen and unseen relation.

5 Conclusion

In this paper, we propose an ontology-guided learning method for KGE, which use TR-GCN to get ontology graph representation and replace relation representation of embedding-based method by

[4] `https://www.wikidata.org/wiki/Wikidata:Database_download`

ontology graph. Our method shows significant effects on unseen relations and improve performance of previous KGE methods such as TransE, RotatE. Experimental results also demonstrate that our method significantly outperforms existing state-of-the-art method on unseen relation learning.

6 Acknowledgements

The work is supported by the National Natural Science Foundation of China (U21B2027, 61972186, 61732005), the Program For Applied Basic Research of Yunnan Province in Key Areas (202002AD080001, 202103AA080015).

References

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

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In *Neural Information Processing Systems (NIPS)*, pp. 1–9, 2013.

Yuxia Geng, Jiaoyan Chen, Zhuo Chen, Jeff Z Pan, Zhiquan Ye, Zonggang Yuan, Yantao Jia, and Huajun Chen. Ontozsl: Ontology-enhanced zero-shot learning. In *Proceedings of the Web Conference 2021*, pp. 3325–3336, 2021.

Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard de Melo, Claudio Gutierrez, José Emilio Labra Gayo, Sabrina Kirrane, Sebastian Neumaier, Axel Polleres, et al. Knowledge graphs. *arXiv preprint arXiv:2003.02320*, 2020.

Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and S Yu Philip. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick Van Kleef, Sören Auer, et al. Dbpedia–a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic web*, 6(2):167–195, 2015.

Farzaneh Mahdisoltani, Joanna Biega, and Fabian Suchanek. Yago3: A knowledge base from multilingual wikipedias. In *7th biennial conference on innovative data systems research*. CIDR Conference, 2014.

Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. A three-way model for collective learning on multi-relational data. In *Proceedings of the 28th International Conference on International Conference on Machine Learning*, pp. 809–816, 2011.

Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017.

Jeffrey Pennington, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532–1543, 2014.

Pengda Qin, Xin Wang, Wenhu Chen, Chunyun Zhang, Weiran Xu, and William Yang Wang. Generative adversarial zero-shot relational learning for knowledge graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 8673–8680, 2020.

Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In *European semantic web conference*, pp. 593–607. Springer, 2018.

Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. Rotate: Knowledge graph embedding by relational rotation in complex space. *arXiv preprint arXiv:1902.10197*, 2019.

Kristina Toutanova, Danqi Chen, Patrick Pantel, Hoifung Poon, Pallavi Choudhury, and Michael Gamon. Representing text for joint embedding of text and knowledge bases. In *Proceedings of the 2015 conference on empirical methods in natural language processing*, pp. 1499–1509, 2015.

Jing Zhang, Bo Chen, Lingxi Zhang, Xirui Ke, and Haipeng Ding. Neural, symbolic and neural-symbolic reasoning on knowledge graphs. *AI Open*, 2:14–35, 2021.

Yao Zhang, Xu Zhang, Jun Wang, Hongru Liang, Wenqiang Lei, Zhe Sun, Adam Jatowt, and Zhenglu Yang. Generalized relation learning with semantic correlation awareness for link prediction. *CoRR*, abs/2012.11957, 2020. URL https://arxiv.org/abs/2012.11957.