Relational Graph Neural Network with Hierarchical Attention for Knowledge Graph Completion

Zhao Zhang[1], Fuzhen Zhuang[1], Hengshu Zhu[2], Zhiping Shi[3], Hui Xiong[2], Qing He[1]

[1]Institute of Computing Technology, CAS; [2]Baidu Inc.; [3]Capital Normal University.

Introduction

Knowledge graphs (KGs) are comprised of knowledge triples in the form of (h, r, t), where h and t correspond to the head and tail entities and r denotes the relation between them, e.g. (Beijing, capitalOf, China). Knowledge graph completion aims to fill the missing values into incomplete triples.

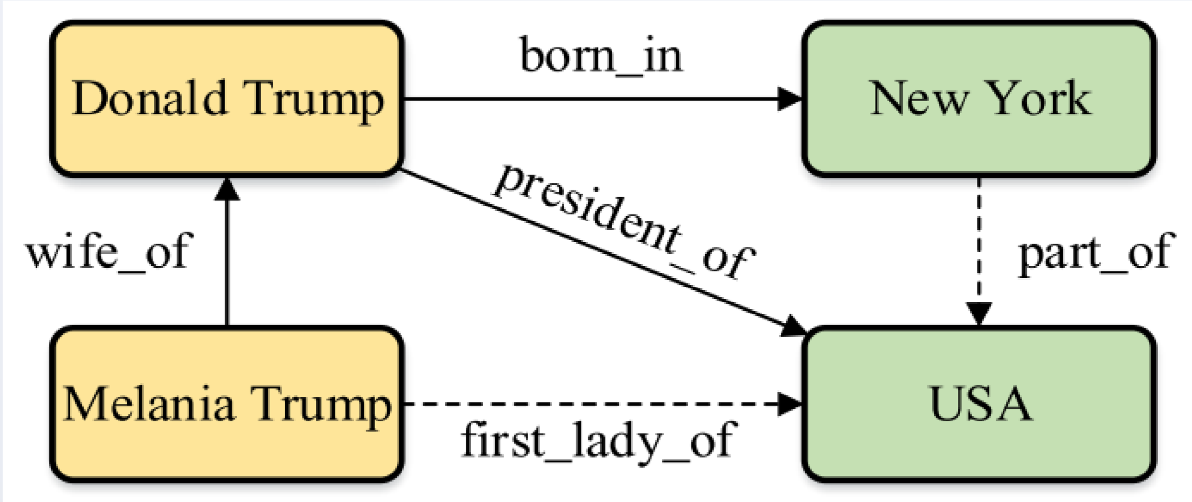


Figure 1: Subgraph of a KG containing existing triples (solid lines) and the inferred ones (dashed lines).

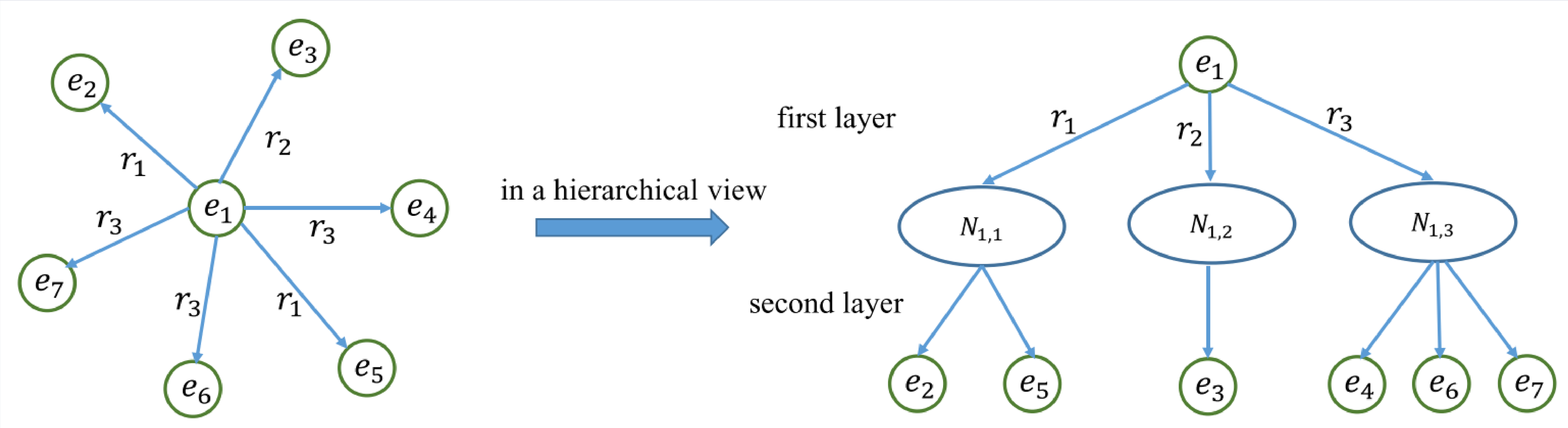
Motivation

1. Most existing models treat the triples in KGs independently, and fail to pay attention to the local neighborhood information of an entity.
2. Graph neural network (GNN) enables each node to gather information from its neighborhood.

Question: Can we leverage the local neighborhood of an entity for the KG completion task using GNNs?

Observation

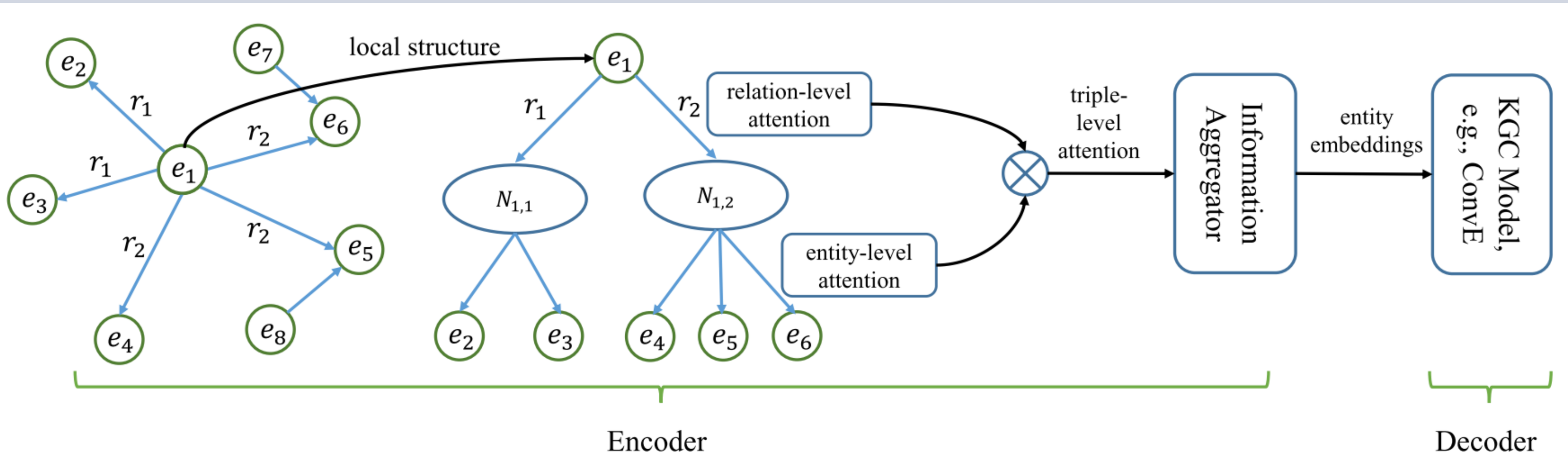
We find the neighborhood of an entity can be viewed as a hierarchical structure.



Question: Can we take advantage of the local neighborhood information of an entity with the hierarchical structure?

Methodology

1. We design an encoder-decoder framework.
2. The encoder learns entity representations in a GNN manner with the hierarchical attention mechanism.
3. The decoder is an existing model, ConvE.



Hierarchical attention mechanism:

1. First layer: relation-level attention, which is inspired by the fact that the weights of different relations differ greatly in indicating an entity.

$$\mathbf{a}_{h,r} = \mathbf{W}_1 \left[\mathbf{h} \middle\| \mathbf{v}_r\right],$$

$$\alpha_{h,r} = \mathrm{softmax}_r(\mathbf{a}_{h,r}) = \frac{\exp(\sigma(\mathbf{p}\cdot\mathbf{a}_{h,r}))}{\sum_{r'\in\mathcal{N}_h}\exp(\sigma(\mathbf{p}\cdot\mathbf{a}_{h,r'}))},$$

2. Second layer: entity-level attention, which enables our model to highlight the importance of different neighboring entities under the same relation.

$$\mathbf{b}_{h,r,t} = \mathbf{W}_2 \left[\mathbf{a}_{h,r} \middle\| \mathbf{t}\right],$$

$$\beta_{r,t} = \mathrm{softmax}_t(\mathbf{b}_{h,r,t}) = \frac{\exp(\sigma(\mathbf{q}\cdot\mathbf{b}_{h,r,t}))}{\sum_{t'\in\mathcal{N}_{h,r}}\exp(\sigma(\mathbf{q}\cdot\mathbf{b}_{h,r,t'}))},$$

Finally, the two-level attention scores are further combined into a triple-level attention score.

$$\mu_{h,r,t} = \alpha_{h,r}\cdot\beta_{r,t},$$

Experiment

Dataset: FB15k, WN18, FB15k-237 and WN18RR

Dataset	$\lvert\mathcal{E}\rvert$	$\lvert\mathcal{R}\rvert$	#triples in Train/Valid/Test
FB15k	14,951	1,345	483,142 / 50,000 / 59,071
WN18	40,943	18	141,442 / 5,000 / 5,000
FB15k-237	14,541	237	272,115 / 17,535 / 20,466
WN18RR	40,943	11	86,835 / 3,034 / 3,134

Results

	FB15k					WN18				
			Hits@N					Hits@N		
	MR	MRR	@1	@3	@10	MR	MRR	@1	@3	@10
TransE (Bordes et al. 2013)¶	-	0.463	0.297	0.578	0.749	-	0.495	0.113	0.888	0.943
DistMult (Yang et al. 2015)¶	42	0.798	-	-	0.893	665	0.797	-	-	0.946
ComplEx (Trouillon et al. 2016)¶	-	0.692	0.599	0.759	0.840	-	0.941	0.936	0.945	0.947
RotatE (Sun et al. 2019)¶	40	0.797	0.746	0.830	0.884	**309**	0.949	0.944	**0.952**	0.959
ConvE (Dettmers et al. 2018)¶	51	0.657	0.558	0.723	0.831	374	0.943	0.935	0.946	0.956
R-GCN (Schlichtkrull et al. 2018)	-	0.696	0.601	0.760	0.842	-	0.819	0.697	0.929	**0.964**
RGHAT (Ours)	**37**	**0.812**	**0.760**	**0.843**	**0.898**	342	**0.954**	**0.949**	0.951	**0.964**

	FB15K-237					WN18RR				
			Hits@N					Hits@N		
	MR	MRR	@1	@3	@10	MR	MRR	@1	@3	@10
TransE (Bordes et al. 2013)¶	357	0.294	-	-	0.465	3384	0.226	-	-	0.501
DistMult (Yang et al. 2015)¶	254	0.241	0.155	0.263	0.419	5110	0.43	0.39	0.44	0.49
ComplEx (Trouillon et al. 2016)¶	339	0.247	0.158	0.275	0.428	5261	0.44	0.41	0.46	0.51
RotatE (Sun et al. 2019)¶	**177**	0.338	0.241	0.375	0.533	3340	0.476	**0.428**	0.492	0.571
ConvE (Dettmers et al. 2018)¶	244	0.325	0.237	0.356	0.501	4187	0.43	0.40	0.44	0.52
ConvKB (Nguyen et al. 2018)§	216	0.289	0.198	0.324	0.471	**1295**	0.265	0.058	0.445	0.558
R-GCN (Schlichtkrull et al. 2018)§	600	0.164	0.10	0.181	0.30	6700	0.123	0.08	0.137	0.207
Nathani's (Nathani et al. 2019)§	210	0.518	0.46	0.54	0.626	1940	0.44	0.361	0.483	0.581
A2N (Bansal et al. 2019)	-	0.317	0.232	0.348	0.486	-	0.45	0.42	0.46	0.51
RGHAT (Ours)	196	**0.522**	**0.462**	**0.546**	**0.631**	1896	**0.483**	0.425	**0.499**	**0.588**

Conclusion

In this paper, we proposed a novel neighborhood-aware model RGHAT for the KGC task. RGHAT is equipped with a hierarchical attention mechanism, which can effectively aggregate the local neighborhood information of each entity. Particularly, the hierarchical attention mechanism provides a fine-grained learning process for the proposed model, which increased the interpretability of RGHAT. Moreover, further analysis showed the results of RGHAT were more consistent with human intuition compared to other neighborhood-aware models. Finally, extensive experiments on popular benchmarks clearly validated the superiority of RGHAT against various sate-of-the-art baselines.

www.PosterPresentations.com