From Discrimination to Generation: Knowledge Graph Completion with Generative Transformer

Xin Xie[1,2], Ningyu Zhang[1,2*], Zhoubo Li[1,2], Shuming Deng[1,2],
Hui Chen[3], Feiyu Xiong[3], Mosha Chen[3] and Huajun Chen[1,2*]

[1] Zhejiang University & AZFT Joint Lab for Knowledge Engine,
[2] Hangzhou Innovation Center, Zhejiang University
[3]Alibaba Group

Introduction

1. We convert link prediction to sequence to sequence generation and propose GenKGC, which can reduce the inference time while maintaining the performance.
2. We propose relation-guided demonstration and entity-aware hierarchical decoding, which can better represent entities and relations and reduce the time complexity of generation.
3. We report results on two datasets and release a new largescale KG dataset, OpenBG500, for research purposes.

WHY Generation?

Most previous KG completion methods embed entities and relations into a vector space and obtain the predicted triples by leveraging pre-defined scoring functions to those vectors as discrimination paradigm.

1. **For training and inference**: a discrimination strategy has to costly scoring of all possible triples in inference and suffers from the instability of negative sampling.

2. **For data storage efficiency**: a discrimination strategy allocates a large memory footprint for the large-scale real-world knowledge graph.

Model

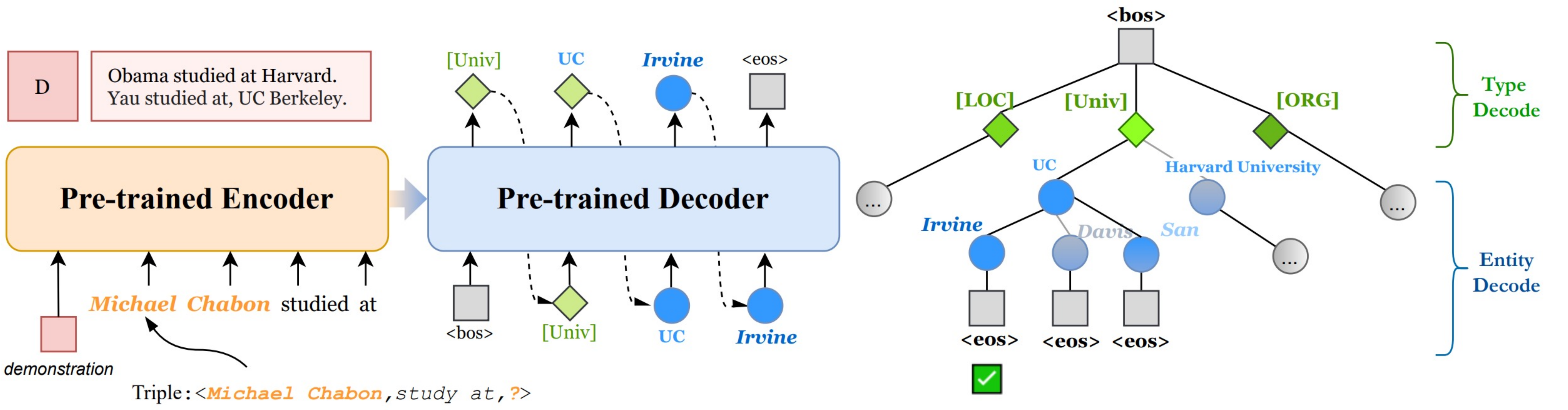

Figure 1: Architecture of GenKGC. We augment the input text of entity and relation with demonstrations, and introduce entity-aware hierarchical decoding for fast inference.

Experiments

Main result in WN18RR, FB15k-237 and OpenBG500.

	WN18RR			FB15k-237			OpenBG500		
Method	Hits@1	Hits@3	Hits@10	Hits@1	Hits@3	Hits@10	Hits@1	Hits@3	Hits@10
Graph embedding approach									
TransE [2] ⋄	0.043	0.441	0.532	0.198	0.376	0.441	0.207	0.340	0.513
DistMult [13] ⋄	0.412	0.470	0.504	0.199	0.301	0.446	0.049	0.088	0.216
ComplEx [11] ⋄	0.409	0.469	0.530	0.194	0.297	0.450	0.053	0.120	0.266
RotatE [9]	0.428	0.492	0.571	0.241	0.375	0.533	-	-	-
TuckER [1]	0.443	0.482	0.526	0.226	0.394	0.544	-	-	-
ATTH [4]	0.443	0.499	0.486	0.252	0.384	0.549	-	-	-
Textual encoding approach									
KG-BERT [14]	0.041	0.302	0.524	-	-	0.420	0.023	0.049	0.241
StAR [12]	0.243	0.491	0.709	0.205	0.322	0.482	-	-	-
GenKGC	0.287	0.403	0.535	0.192	0.355	0.439	0.203	0.280	0.351

Inference and training efficiency comparison.

For One Triple	Method	Complexity	Time under RTX 3090
Training	TransE	$O(k+1)$	0.08ms
	KG-BERT	$O(\|d\|^2 \times (k+1))$	72ms
	GenKGC	$O(\|d\|^2)$	2.35ms
Inference	TransE	$O(\|\mathcal{E}\|)$	0.02s
	KG-BERT	$O(\|d\|^2 \times \|\mathcal{E}\|)$	10100s
	GenKGC	$O(\|d\|^2 \times \|d\|^k)$	0.96s

A query and first five entities with their probability predicted by GenKGC w/o entity-aware decoding, and its reranking with GenKGC.

Query:(?,student,Michael Chabon)		
Rank	GenKGC w/o hierarchical decoding	Probability
1	University of California	
2	**University of California, Irvine**	
3	University of California, San Francisco	
4	University of California, Davis	
5	University of California, Santa Cruz	
Rank	GenKGC	Probability
1	**University of California, Irvine**	
2	University of California, San Francisco	
3	University of California, Davis	
4	University of California, Santa Cruz	
5	University of Calgary	

Autoregressive loss function with type constrained decoding

$$p_\theta(y \mid x) = \prod_{i=1}^{|c|} p_\theta(z_i \mid z_{<i}, x) \prod_{i=|c|+1}^{N} p_\theta(y_i \mid y_{<i}, x),$$