ZeroKBC: A Comprehensive Benchmark for Zero-Shot Knowledge Base Completion

Pei Chen[1], Wenlin Yao[2], Hongming Zhang[2], Xiaoman Pan[2], Dian Yu[2], Dong Yu[2] and Jianshu Chen[2]

[1]Texas A&M University
[2]Tencent AI Lab

Motivation

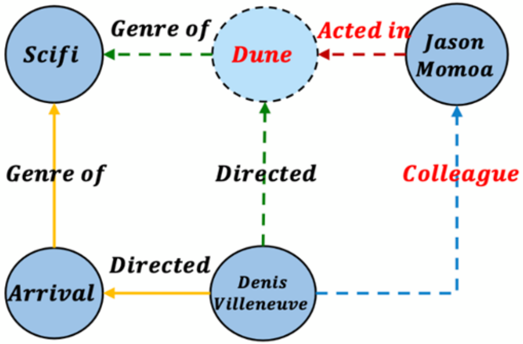
- Knowledge base completion (KBC): predict the missing links in knowledge graphs.
- Challenges:
  - Previous KBC tasks and approaches mainly focus on the setting where all test entities and relations have appeared in the training set.
  - KG is growing constantly - new (unseen) entities and relations emerge every day
  - Limited research on the Zero-shot KBC settings

- What are all possible scenarios of zero-shot KBC in real-world application?
- Can we develop a comprehensive benchmark to cover all scenarios  matically evaluate different KBC models?
- What is the most important aspect to solve zero-shot KBC?

Three Zero-shot Scenarios

- Zero-Shot Entity KBC (ZeroE)
  - When we see a new entity, we need to predict all the possible links between this new entity and existing entities
  - We may know some context for it (some connections between this new entity and the existing KG) or we need the textual descriptions of it in the absence of connections
- Zero-Shot Relation KBC (ZeroR):
  - When we see a new relation, predict if this relation exists in any pair of the entities
  - We will always have the textual descriptions of new relations to distinguish them
- Zero-Shot Both KBC (ZeroB):
  - Predict the link for both a newly added relation and a newly added entity
  - Textual description is necessary for the relations as in ZeroR and entity descriptions are also required in the absence of context

Different Scenarios of Zero-shot KBC



Entity Description:
Denis Villeneuve: *a Canadian film director and screenwriter …*
Relation Description:
Genre of: *category of creative works based on stylistic, thematic or technical criteria …*

Task Definition		Head + Relation -> ?Tail Tail + Relation -> ?Head				
Scenarios	Settings	Input head/tail Entity			Input Relation	
		Seen	Description	Context	Seen	Description
KBC + Descriptions	1	Y	Y	Y	Y	Y
	2	Y	Y	Y	Y	N
	3	Y	N	Y	Y	Y
Standard KBC	4	Y	N	Y	Y	N
Zero-shot Entity	5	N	Y	Y/N	Y	Y
	6	N	Y	Y/N	Y	N
	7	N	N	Y	Y	Y
	8	N	N	Y	Y	N
Zero-shot Relation	9	Y	Y	Y	N	Y
	10	Y	N	Y	N	Y
Zero-shot Entity/Relation	11	N	Y	Y/N	N	Y
	12	N	N	Y	N	Y
N/A		All other Cases				

Zero-shot Scenario Coverage Comparison

Related Work	Settings												Knowledge Sources		
	1	2	3	4	5	6	7	8	9	10	11	12	Linguistic	World	Commonsense
TransE (Bordes et al., 2013)				✓									✓	✓	
(Neelakantan et al., 2015)										✓				✓	
(Xie et al., 2016)		✓				✓								✓	
(Hamaguchi et al., 2017)								✓					✓	✓	
RotatE (Sun et al., 2019)				✓									✓	✓	
(Albooyeh et al., 2020)								✓					✓	✓	
(Baek et al., 2020)								✓					✓	✓	
(Malaviya et al., 2020)		✓													✓
(Qin et al., 2020)										✓			✓	✓	
StAR (Wang et al., 2021)	✓												✓	✓	
Our Work	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

- Each previous work only covers a small portion of KBC settings, while our work has more exhaustive coverage.
- We found that there are 6 important KBC settings that were missing in previous study: standard setting #3 and zero-shot settings #5, #7, #9, #11 and #12.

Statistics of our ZeroKBC

Scenarios	Statistics	WN18RR	FB15K237	ATOMIC
Standard	# All Triples	93,003	310,116	389,437
	# Ent.	40,943	14,541	48,645
	# Rel.	11	237	9
	# Avg. Wrd. Ent.	14.24	7.34	3.72
	# Avg. Wrd. Rel.	2.55	19.24	4.44
ZeroE	# Seen Ent.	32,270	11,579	38,903
	# Un. Dev Ent.	2,848	1,395	4,768
	# Un. Test Ent.	2,848	1,396	4,691
ZeroR	# Seen Rel.	6	157	5
	# Un. Dev Rel.	2	20	2
	# Un. Test Rel.	3	60	2
ZeroB	# Seen Ent.	30,015	11,252	32,919
	# Un. Dev Ent.	54	449	2,577
	# Un. Test Ent.	225	962	2,977
	# Seen Rel.	6	154	5
	# Un. Dev Rel.	2	19	2
	# Un. Test Rel.	3	59	2

- ZeroKBC includes three types of knowledge resources:
  - Linguistic knowledge (WN18RR)
  - World knowledge (FB15K-237)
  - Commonsense knowledge (ATOMIC)
- Three zero scenarios
  - ZeroE: use the out-of-sample entities as the unseen entities for evaluation
  - ZeroR: set aside a part of relations for evaluation only
  - ZeroB: take the intersection of the previous two zero-shot scenarios

Baselines

- TransE:
  - Embeds entities and relations into low-dimension vectors and interprets relations r as translations operating on the head h and tail t entities
  - In its vanilla settings, the initial embeddings of entities and relations are random
  - In our adaptation, we use BERT to encode the textual descriptions of entities and relations for initialization when the descriptions are available
  - Unseen entities or relations will use the BERT embeddings for prediction.
- RotatE:
  - Embeds entities and relations into complex space vectors and interprets relations r as rotations operating in the space
  - Similar to TransE, we use BERT to initialize the real part of the entities and relations

Baselines

- StAR:
  - Adopts BERT as its textual encoder and uses both the translation property of a triple and their textual semantics that can be more generalizable for unseen entities and relations.

Experimental Results

Scenarios	Settings	Models	WN18RR					FB15K237					ATOMIC				
			MR	MRR	Hits @10	Hits @3	Hits @1	MR	MRR	Hits @10	Hits @3	Hits @1	MR	MRR	Hits @10	Hits @3	Hits @1
Standard	1	TransE	2368	20.8	50.8	36.4	1.7	193	31.5	50.3	35.0	22.0	2442	20.2	29.6	22.2	15.2
		RotatE	2814	**42.6**	54.0	**46.1**	**36.1**	193	**32.2**	**51.5**	**35.8**	**22.7**	2318	**24.1**	**32.6**	**25.2**	**19.8**
		StAR	**68**	34.6	**62.3**	41.9	20.3	**127**	27.4	45.8	30.0	18.4	**1566**	7.8	12.5	7.5	5.2
ZeroE	5 w/ context	TransE	3598	7.6	12.1	7.9	5.1	787	11.8	19.9	12.4	7.7	6170	0.8	1.5	0.7	0.3
		RotatE	4064	7.6	11.6	7.8	5.4	2412	2.6	4.9	2.5	1.1	11872	0.2	0.4	0.2	0.1
		StAR	**487**	**36.3**	**58.0**	**41.0**	**25.5**	**236**	**22.5**	**38.7**	**24.2**	**14.5**	**1399**	**5.7**	**10.5**	**5.3**	**3.0**
ZeroR	9	TransE	19584	0.3	0.4	0.2	0.1	6922	2.7	4.3	2.8	1.7	7799	0.5	0.8	0.2	0.1
		RotatE	21876	0.2	0.4	0.2	0.1	6645	1.6	2.3	1.6	1.2	18562	0.1	0.1	0.0	0.0
		StAR	**2810**	**9.7**	**24.6**	**11.1**	**2.4**	**1836**	**11.8**	**21.2**	**12.1**	**7.0**	**2787**	**2.6**	**4.3**	**2.3**	**1.4**
ZeroB	11 w/ context	TransE	**4830**	4.0	9.0	3.2	2.1	3581	3.8	6.9	4.2	1.9	15032	2.2	3.8	**2.0**	0.1
		RotatE	7116	4.3	8.9	4.1	2.6	3855	1.5	3.8	1.1	0.4	10405	1.8	1.9	1.6	**1.6**
		StAR	5317	**14.1**	**29.6**	**15.4**	**7.7**	**1561**	**10.4**	**19.2**	**10.5**	**5.8**	**2102**	**2.3**	**4.1**	**2.0**	1.0

- Compared with the standard KBC setting, the performance of canonical KBC systems (i.e., TransE, RotatE, and StAR) drops dramatically under zero-shot entity or relation setting.
- ZeroB is the most challenging setting since both entities and relations in test data are unseen during training

Scenarios	Settings	Models	WN18RR					FB15K237					ATOMIC				
			MR	MRR	Hits @10	Hits @3	Hits @1	MR	MRR	Hits @10	Hits @3	Hits @1	MR	MRR	Hits @10	Hits @3	Hits @1
Standard	1	TransE	2368	20.8	50.8	36.4	1.7	193	31.5	50.3	35.0	22.0	2442	20.2	29.6	22.2	15.2
		RotatE	2814	**42.6**	54.0	**46.1**	**36.1**	193	**32.2**	**51.5**	**35.8**	**22.7**	2318	**24.1**	**32.6**	**25.2**	**19.8**
		StAR	**68**	34.6	**62.3**	41.9	20.3	**127**	27.4	45.8	30.0	18.4	**1566**	7.8	12.5	7.5	5.2
ZeroE	5 w/ context	TransE	3598	7.6	12.1	7.9	5.1	787	11.8	19.9	12.4	7.7	6170	0.8	1.5	0.7	0.3
		RotatE	4064	7.6	11.6	7.8	5.4	2412	2.6	4.9	2.5	1.1	11872	0.2	0.4	0.2	0.1
		StAR	**487**	**36.3**	**58.0**	**41.0**	**25.5**	**236**	**22.5**	**38.7**	**24.2**	**14.5**	**1399**	**5.7**	**10.5**	**5.3**	**3.0**
ZeroR	9	TransE	19584	0.3	0.4	0.2	0.1	6922	2.7	4.3	2.8	1.7	7799	0.5	0.8	0.2	0.1
		RotatE	21876	0.2	0.4	0.2	0.1	6645	1.6	2.3	1.6	1.2	18562	0.1	0.1	0.0	0.0
		StAR	**2810**	**9.7**	**24.6**	**11.1**	**2.4**	**1836**	**11.8**	**21.2**	**12.1**	**7.0**	**2787**	**2.6**	**4.3**	**2.3**	**1.4**
ZeroB	11 w/ context	TransE	**4830**	4.0	9.0	3.2	2.1	3581	3.8	6.9	4.2	1.9	15032	2.2	3.8	**2.0**	0.1
		RotatE	7116	4.3	8.9	4.1	2.6	3855	1.5	3.8	1.1	0.4	10405	1.8	1.9	1.6	**1.6**
		StAR	5317	**14.1**	**29.6**	**15.4**	**7.7**	**1561**	**10.4**	**19.2**	**10.5**	**5.8**	**2102**	**2.3**	**4.1**	**2.0**	1.0

- StAR can encode the semantics of text descriptions more efficiently, which significantly outperforms TransE and RotatE across knowledge graph genres and zero-shot settings
- It indicates the importance of considering textual semantics in zero-shot KBC
  - Design an efficient approach that can take descriptive information into consideration

Scenarios	Settings	Models	WN18RR					FB15K237					ATOMIC				
			MR	MRR	Hits @10	Hits @3	Hits @1	MR	MRR	Hits @10	Hits @3	Hits @1	MR	MRR	Hits @10	Hits @3	Hits @1
Standard	1	TransE	2368	20.8	50.8	36.4	1.7	193	31.5	50.3	35.0	22.0	2442	20.2	29.6	22.2	15.2
		RotatE	2814	**42.6**	54.0	**46.1**	**36.1**	193	**32.2**	**51.5**	**35.8**	**22.7**	2318	**24.1**	**32.6**	**25.2**	**19.8**
		StAR	**68**	34.6	**62.3**	41.9	20.3	**127**	27.4	45.8	30.0	18.4	**1566**	7.8	12.5	7.5	5.2
ZeroE	5 w/ context	TransE	3598	7.6	12.1	7.9	5.1	787	11.8	19.9	12.4	7.7	6170	0.8	1.5	0.7	0.3
		RotatE	4064	7.6	11.6	7.8	5.4	2412	2.6	4.9	2.5	1.1	11872	0.2	0.4	0.2	0.1
		StAR	**487**	**36.3**	**58.0**	**41.0**	**25.5**	**236**	**22.5**	**38.7**	**24.2**	**14.5**	**1399**	**5.7**	**10.5**	**5.3**	**3.0**
ZeroR	9	TransE	19584	0.3	0.4	0.2	0.1	6922	2.7	4.3	2.8	1.7	7799	0.5	0.8	0.2	0.1
		RotatE	21876	0.2	0.4	0.2	0.1	6645	1.6	2.3	1.6	1.2	18562	0.1	0.1	0.0	0.0
		StAR	**2810**	**9.7**	**24.6**	**11.1**	**2.4**	**1836**	**11.8**	**21.2**	**12.1**	**7.0**	**2787**	**2.6**	**4.3**	**2.3**	**1.4**
ZeroB	11 w/ context	TransE	**4830**	4.0	9.0	3.2	2.1	3581	3.8	6.9	4.2	1.9	15032	2.2	3.8	**2.0**	0.1
		RotatE	7116	4.3	8.9	4.1	2.6	3855	1.5	3.8	1.1	0.4	10405	1.8	1.9	1.6	**1.6**
		StAR	5317	**14.1**	**29.6**	**15.4**	**7.7**	**1561**	**10.4**	**19.2**	**10.5**	**5.8**	**2102**	**2.3**	**4.1**	**2.0**	1.0

- Graph properties of knowledge bases can affect models' generalization capabilities.
  - Compared with WN18RR and FB15K237, ATOMIC has a much sparser graph structure
  - The models that only leverage the graph structures suffer from the largest performance drop from the standard setting to zero-shot ones
- The model (i.e., StAR) could generalize better to unseen entities and relations with text descriptions

Conclusion

- We systematically examines different zero-shot KBC scenarios and develops a comprehensive benchmark - **ZeroKBC**, which covers all scenarios with three types of knowledge sources.
- Experimental results show that canonical and SOTA KBC systems suffer great performance degradation on this challenging yet practical benchmark
- We further present several important observations and reveal the importance of designing more efficient methods to **integrate semantics** of entities and relations into KG learning

Code and data: https://github.com/brickee/ZeroKBC