Open-World Taxonomy and Knowledge Graph Co-Learning

Presented at KDD'22-OKN

Jiaying Lu (jiaying.lu@emory.edu) [1] Carl Yang (j.carlyang@emory.edu) [1]

[1]Emory University, Atlanta, GA, USA

Aug 15, 2022

Table of Contents

Challenges for Closed-World Knowledge Base

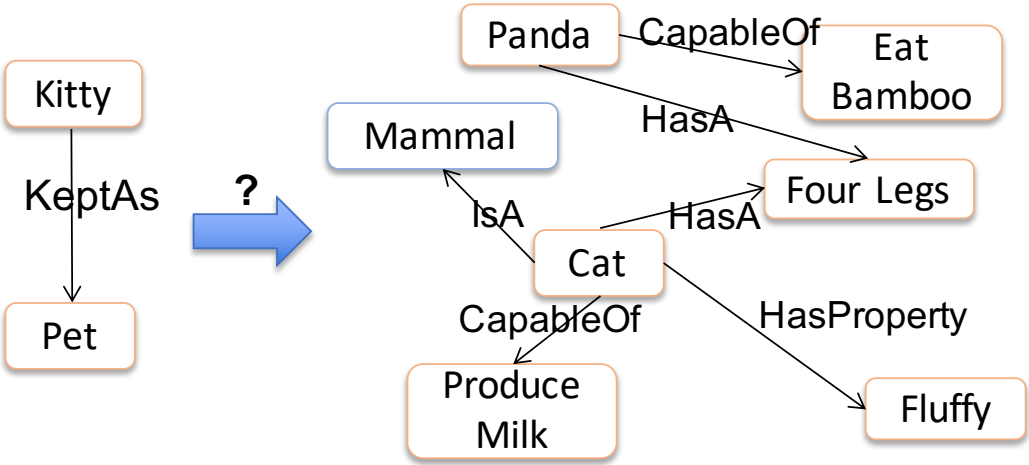


Figure 1: Closed-World KB

Taxonomy Is the Silver Bullet

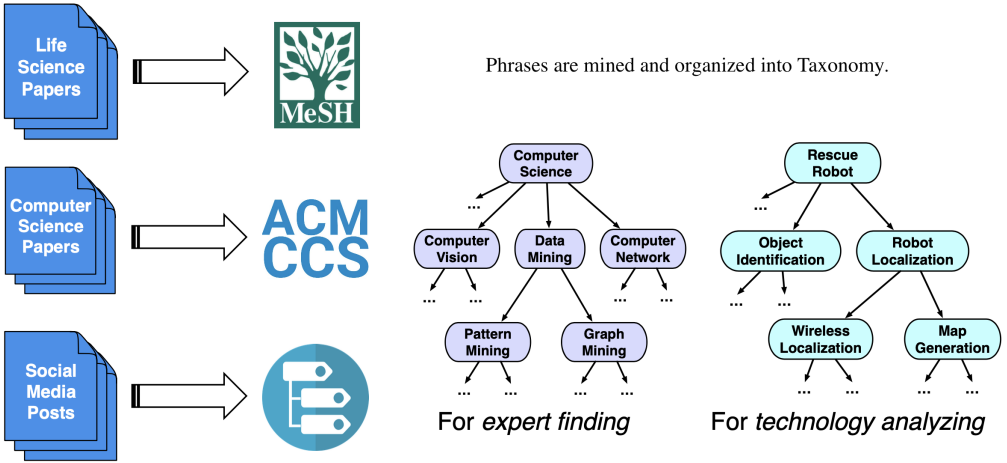


Figure 2: Toy example for automated taxonomy discovery. (Image credit: Shen et al. [4])

Integration of Auto-constructed Taxonomy and Open Knowledge Graph

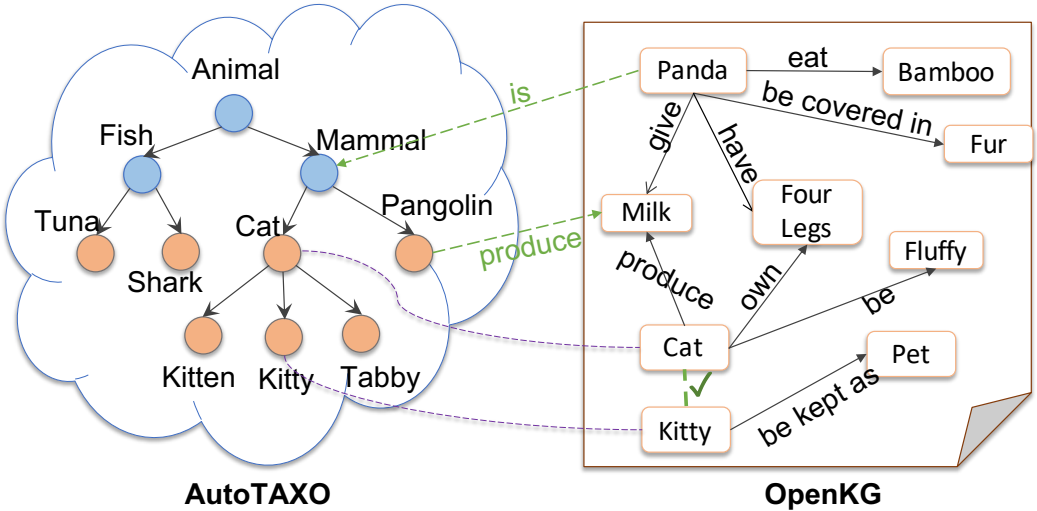
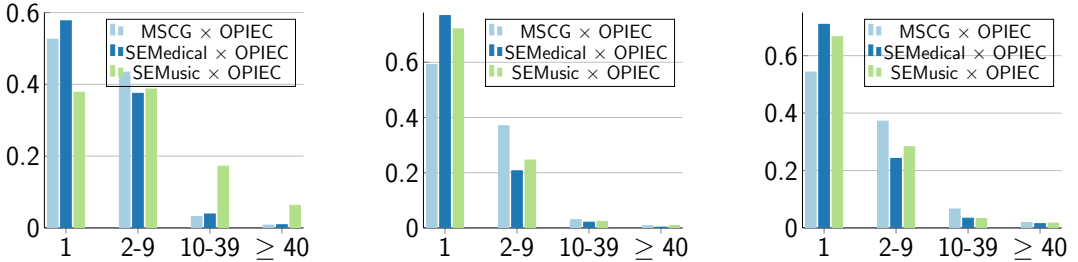


Figure 3: Open-World TaxoKG

TaxoKG-Bench: A New Benchmark with Six Datasets for TaxoKG

TaxoKG-Bench[1] is created from:

- ◇ Three AutoTAXOs: MS Concept Graph [5], SemEval'18 Task9 2A:Medical and 2B:Music [1];
- ◇ Two OpenKGs: ReVerb [2] and OPIEC [3].



(a) Concept on * × OPIEC (b) Entity on * × OPIEC (c) Relation on * × OPIEC

Figure 5: Concept, entity and relation histograms on three aligned TaxoKGs.

[1]We release TaxoKG-Bench: https://figshare.com/s/ca54dd1ca5f08a203017

HakeGCN: A Novel Method for Effective TaxoKG Completion

We propose HakeGCN with the following novel components:

- ◇ The **polar coordinates-based GCN encoder**;
- ◇ The **taxonomy-based sampling strategy**;
- ◇ The **GCN-oriented phased bounded decoder**.

Pilot Study on Combing TAXO and KG

Table 1: OpenKG completion results on OKG v.s. TaxoKG.

Method	Data	R-MRR	R-H@ 10, 30, 50
HAKE	OPIEC (OKG Only)	.350	**.454**, **.517**, **.545**
HAKE	SEMedical x OPIEC (TaxoKG)	**.352**	.450, .509, .544
CompGCN	OPIEC (OKG only)	.006	.012, **.030**, **.049**
CompGCN	SEMedical x OPIEC (TaxoKG)	**.009**	**.013**, .023, .034
HakeGCN	OPIEC (OKG only)	.375	.478, .555, .607
HakeGCN	SEMedical x OPIEC (TaxoKG)	**.412**	**.508**, **.600**, **.652**

Table 2: Concept assignment results on AutoTaxo v.s. TaxoKG.

Method	Data	C-MAP	C-P@1, 3, 10
HAKE	SEMedical (TAXO only)	.186	.344, **.355**, .177
HAKE	SEMedical x OPIEC (TaxoKG)	**.262**	**.371**, .309, **.256**
CompGCN	SEMedical (TAXO only)	**.075**	**.284**, **.117**, **.109**
CompGCN	SEMedical x OPIEC (TaxoKG)	.041	.060, .044, .032
HakeGCN	SEMedical (TAXO only)	.105	.093, .093, .123
HakeGCN	SEMedical x OPIEC (TaxoKG)	**.271**	**.377**, **.366**, **.251**

Performance Comparison on TaxoKG Completion Task

Table 3: Experiment results on TaxoKG completion in the medical domain.

	SEMedical × ReVerb				SEMedical × OPIEC			
	C-MAP	C-P@1, 3, 10	R-MRR	R-H@10, 30, 50	C-MAP	C-P@1, 3, 10	C-MRR	R-H@10, 30, 50
TransE	.036	.104, .083, .050	.002	.002, .009, .012	.025	.045, .061, .030	.005	.007, .019, .030
HAKE	.203	.307, **.286**, **.216**	.170	.343, .430, .459	.262	.371, .309, **.256**	.352	.450, .509, .544
DistMult	.065	.188, .069, .033	.023	.070, .135, .187	.022	.159, .068, .032	.032	.061, .158, .218
CompGCN	.119	.191, .184, .150	.003	.005, .012, .017	.041	.060, .044, .032	.009	.013, .023, .034
LtCaG	.186	.245, .247, .172	.004	.005, .006, .008	.126	.166, .157, .122	.013	.021, .041, .051
HakeGCN	**.233**	**.331**, .278, .204	**.275**	**.424**, **.545**, **.603**	**.271**	**.377**, **.366**, .251	**.412**	**.508**, **.600**, **.652**

In-Depth Analysis

Table 4: KG neighbors used in taxonomy concept prediction.

Concept	KG Neighbors
technique	(make from, recycled material, -) ✓ (architecture, be a thing of, -) ✓ (-, be apply, biology) ✓ (-, mean of, expression) ✗

Table 5: Taxonomy neighbors used in KG relation prediction.

Relation	Taxonomy Neighbors
be marry to	control ✗, family name ✓, guest ?
die from	illness ✓, disease ✓, disorder ✓
listen to	work of art ?, musical work ✓, piece of music ✓

Conclusion

- ◇ the first paper to study the joint modeling of existing KGs and taxonomies.
- ◇ the new TaxoKG completion task with a new benchmark.
- ◇ the novel HakeGCN model.
- ◇ comprehensive experimental studies.

References

[1] Jose Camacho-Collados, Claudio Delli Bovi, Luis Espinosa Anke, Sergio Oramas, Tommaso Pasini, Enrico Santus, Vered Shwartz, Roberto Navigli, and Horacio Saggion. "SemEval-2018 Task 9: Hypernym Discovery". In: *Proceedings of The 12th International Workshop on Semantic Evaluation*. 2018.

[2] Anthony Fader, Stephen Soderland, and Oren Etzioni. "Identifying relations for open information extraction". In: *EMNLP*. 2011.

[3] Kiril Gashteovski, Sebastian Wanner, Sven Hertling, Samuel Broscheit, and Rainer Gemulla. "OPIEC: An Open Information Extraction Corpus". In: *AKBC*. 2018.

[4] Jiaming Shen, Xiaotao Gu, Yu Meng, and Jiawei Han. "Automated Taxonomy Discovery and Exploration". In: *ICDM Tutorial*. 2021.

[5] Wentao Wu, Hongsong Li, Haixun Wang, and Kenny Q Zhu. "Probase: A probabilistic taxonomy for text understanding". In: *SIGMOD*. 2012.