Type-augmented Relation Prediction in Knowledge Graphs

Zijun Cui (cuiz3@rpi.edu) — Rensselaer Polytechnic Institute
Pavan Kapanipathi (kapanipa@us.ibm.com) — IBM reserach
Kartik Talamadupula (krtalamad@us.ibm.com) — IBM research
Tian Gao (tgao@us.ibm.com) — IBM research
Qiang Ji (jiqr@rpi.edu) — Rensselaer Polytechnic Institute

Motivation

- Leverage prior type information to improve relation prediction performance
- Relation Prediction in Knowledge Graphs:
  - (Helen Mirren, ?, Chiswick)
- Prior Knowledge: type information of entities/relations
  - Helen Mirren is a *person/award_winner/actor/*
  - (*Person*, place_of_birth, *Location*)

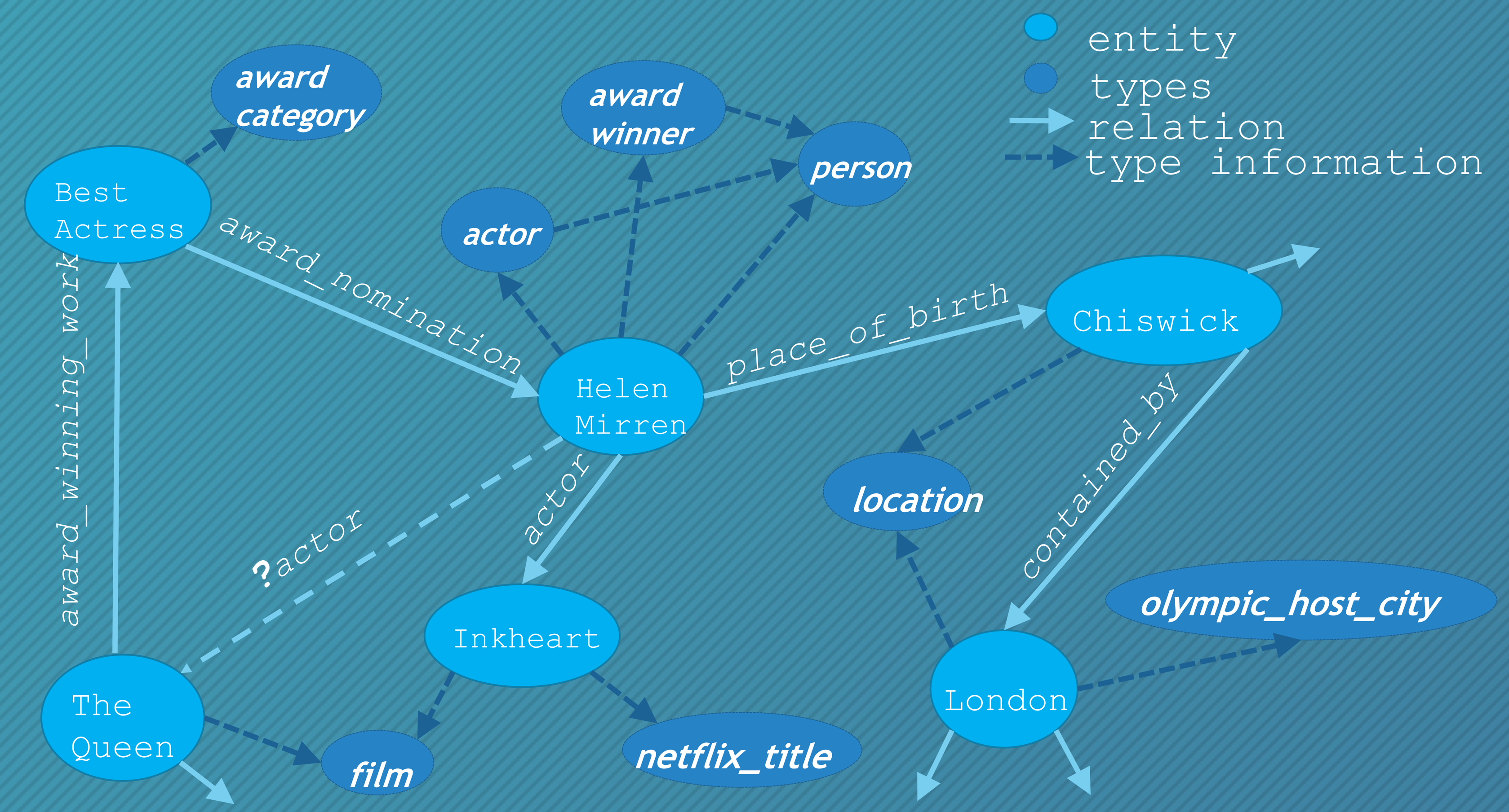

Figure: An example in a knowledge graph(KG)

Overview

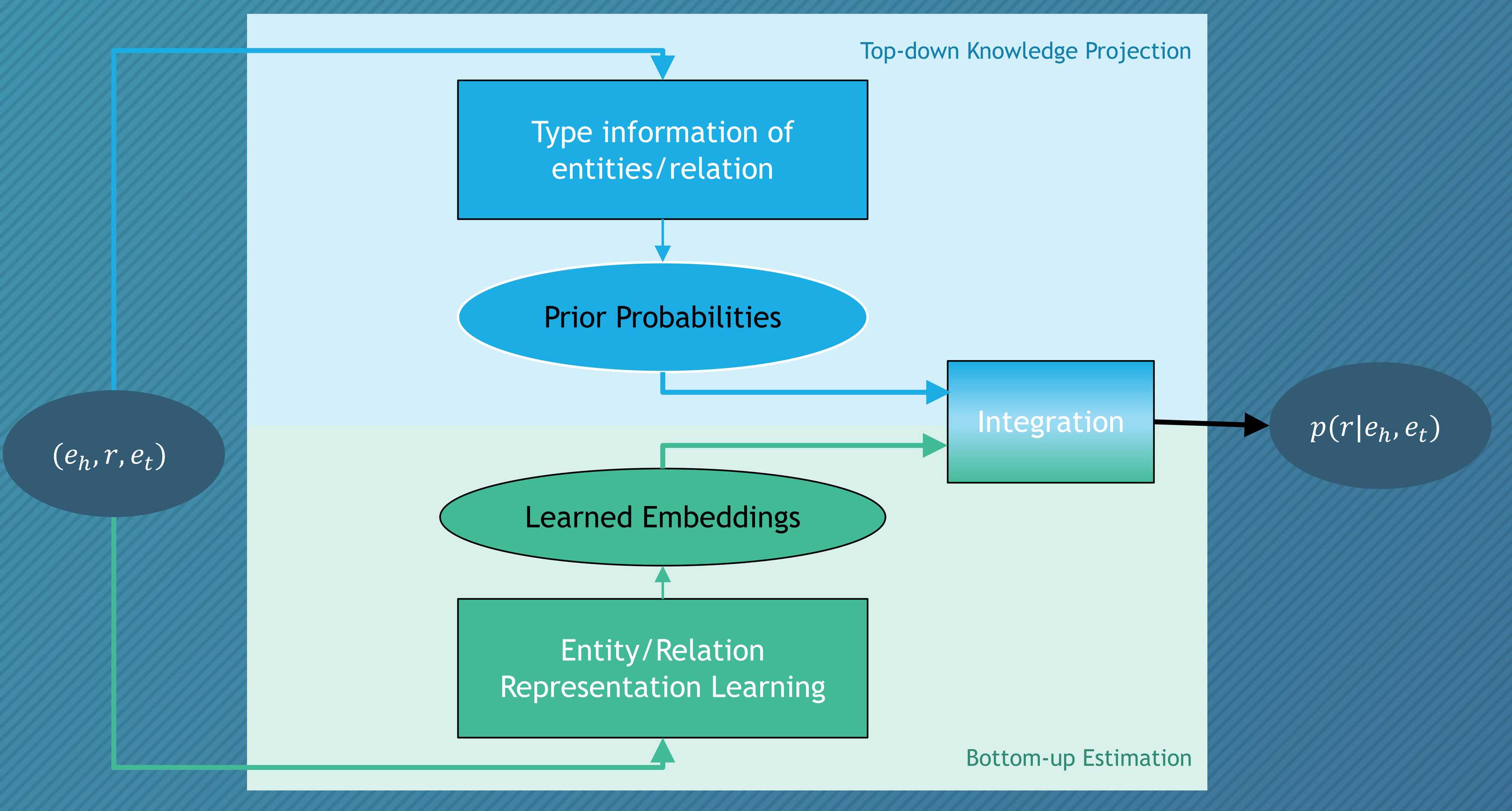

Figure: Overview of the proposed TaRP model

Type Information Encoding

- We encode the type information as prior probabilities by considering hierarchical structures among types
- Type sets usually have an underlying hierarchy, such as the structure among types {***actor, award_winner, person***} :
  - H_1 = */person/actor*
  - H_2 = */person/award_winner*
  - H_3 = */person*

Hierarchy-based type weights

- We define hierarchy-based type weights to assign different weights to types based on their locations in the hierarchy
- We hypothesis that types of more specific semantic meaning are more helpful, and higher weights are automatically assigned to these types
- For example, given three hierarchies H_1, H_2 and H_3 , we have type weights:
  - $w_e(person) = \min\{0.27, 0.27, 1\} = 0.27$
  - $w_e(actor) = 0.73$
  - $w_e(award\_winner) = 0.73$

Type-based prior probability

- Given a triple $(e_h, r, e_t) \in \mathcal{G}$, we define two similarity score $s(e_h, r)$ and $s(e_t, r)$ based on the correlation between type sets
- The prior probability $p(r|\mathcal{T}(e_h, e_t, \mathcal{R}))$ is then defined as

$$p(r|\mathcal{T}(e_h, e_t, \mathcal{R})) = \frac{s(e_h, r)s(e_t, r)}{\sum_{r' \in \mathcal{R}} s(e_h, r')s(e_t, r')}$$

where $\mathcal{T}(e_h, e_t, \mathcal{R})$ is the type information for entity pair (e_h, e_t) and the relation set $\mathcal{R}$

- The higher the correlation between type sets, the higher the prior probability of the relation

Embedding-based Models

- Embedding-based models learn representations of relations and entities by minimizing the distance $f_r(\boldsymbol{e_h}, \boldsymbol{e_t})$ in a continuous embedding space
- Given the learned embeddings, we compute the likelihood by taking the exponential

$$p(e_h, e_t|r) = \exp\left(f_r(\boldsymbol{e_h}, \boldsymbol{e_t})\right)$$

- The lower the distance, the lower the likelihood

Type Information Integration

- Type Information Integration is performed based on probabilities
- For each pair of entities (e_h, e_t), the posterior probability is

$$p(r|e_h, e_t, \mathcal{T}(e_h, e_t, \mathcal{R})) \propto p(e_h, e_t|r)p(r|\mathcal{T}(e_h, e_t, \mathcal{R}))$$

Experiments

- Evaluation of the TaRP model

Baseline 1: embedding-based model trained on observed triples
Baseline 2: embedding-based model trained on (observed triples + type triples)

Models		FB15K			YAGO26K-906			DB111K-174		
		MR	Hits@1	Hits@10	MR	Hits@1	Hits@10	MR	Hits@1	Hits@10
Embedding-based model	TransE	3.64	76.50	92.30	1.12	90.70	99.92	4.76	66.60	86.70
	RotatE	2.38	80.20	97.80	1.10	92.84	99.90	4.53	65.90	93.80
	QuatE	4.01	82.20	94.90	1.33	91.65	98.96	8.56	58.60	88.90
Embedding-based model (trained with type triples)	TransE(w/Type)	3.32	79.37	91.56	1.12	90.70	99.93	4.16	67.64	91.91
	RotatE(w/Type)	3.67	73.63	96.44	**1.08**	**93.31**	99.93	3.47	70.08	96.42
	QuatE(w/Type)	3.98	80.82	92.97	1.32	91.98	99.09	7.63	60.49	89.14
TaRP	TaRP-T	1.84	88.90	99.00	1.10	90.80	**99.98**	1.61	74.80	99.40
	TaRP-R	**1.16**	**92.91**	**99.84**	**1.08**	92.84	**99.98**	**1.52**	76.50	**99.50**
	TaRP-Q	1.64	91.60	99.50	1.14	92.93	99.79	1.56	**76.60**	99.40

- Compare to SoTAs

FB15K	MR	Hits@1	Hits@10
DKRL(CNN)+TransE (Xie et al. 2016)	2.03*	--	90.8*
TKRL(RHE) (Xie, Liu and Sun 2016)	1.73*	92.8*	--
SSP(Std.) (Xiao et al. 2017)	1.22*	--	89.2*
SSP(Joint) (Xiao et al. 2017)	1.47*	--	90.9*
TransT (Ma et al. 2017)	1.19*	--	94.1*
TaRP-R	**1.16**	**92.9**	**99.8**

FB15K-237	MR	Hits@1	Hits@10
HAKE(Zhang et al. 2020)	1.85	92.85	99.13
TaRP-R	**1.19**	**94.25**	**99.79**

YAGO26K-906	MR	Hits@1	Hits@10
JOIE(Hao et al. 2019)	1.47	90.1	97.1
TaRP-R	**1.08**	**92.8**	**99.9**

DB111K-174	MR	Hits@1	Hits@10
JOIE(Hao et al. 2019)	2.22	71.8	89.6
TaRP-R	**1.52**	**76.5**	**99.5**

Conclusions

- We achieve significantly better performance by leveraging type information compared to SoTAs on four benchmark datasets
- Our proposed approach is effective in integrating type information
- In the paper, we also show that our method is more data efficient. Through cross-dataset evaluation, we show that type information extracted from a specific dataset can generalize well to different datasets