THE UNIVERSITY of EDINBURGH

INFR11215 Knowledge Graphs

Combined KG Reasoning

Jeff Z. Pan

http://knowledge-representation.org/j.z.pan/

1

THE UNIVERSITY of EDINBURGH

Lecture Outline

- Motivation
- Discussions on Schema-aware KG Embedding

2

Motivations

- Pros of KG embeddings:
  - Help to address the incompleteness
  - Useful in many downstream applications like link prediction, similarity search and question answering
- Cons of KG embeddings:
  - typically only consider relation assertions
  - do not consider schema
    - Why is it a problem?
    - If it is a problem, how to address it?

Lecture Outline

- Motivation
- Discussions on Schema-aware KG embedding

Brief Summary of KGE for LP

- Link prediction: can be cast as a learning to rank problem

$$\phi_{spo} \triangleq \phi(\langle s,p,o \rangle; \Theta)$$

- KGE for link prediction: 2-layer neural network architecture
  - Encoding layer
  - Scoring layer

$$\begin{aligned} \phi(\langle s,p,o \rangle; \Theta) &\triangleq \phi_p^{\theta}(\mathbf{e}_s, \mathbf{e}_o) \\ \mathbf{e}_s, \mathbf{e}_o &\triangleq \psi(s), \psi(o) \end{aligned}$$

- Loss function require both positive and negative samples

Why Ignoring Schama is a Problem?

- Closed world assumption
- Limited expressiveness of KGE models
- Inconsistency

Closed World Assumption

- Key question: how to pick samples
  - Positive samples are easy
  - Negative samples are a lot trickier, as all input triples are positive
- Closed World Assumption (CWA) is used to pick negative samples:
  - given a KG G, its schema S and a new triple (s,p,o): if (s,p,o) ∈ G, then (s,p,o) is correct; otherwise, (s,p,o) is incorrect
  - Procedure: given any (s,p,o), replace s with s' s.t. (s',p,o) ∉ G

Example: Closed World Assumption

Example 1: if the KG contains the triple (John, likes, Ice_Cream), under the CWA, the negative sample (John, likes, Pizza) would be unequivocally considered as false, suggesting that John does not like pizza.

Example 2: Given a positive triple *(English_Americans, population_place, New_England)* in DB15K dataset, the CWA negative sampling strategy replaces the tail entity with random entities, such as:
(English_Americans, population place, Hawaii),
(English_Americans, population place, Arizona),
(English_Americans, population place, New York metropolitan area),
(English_Americans, population place, Vietnam),
(English_Americans, population place, Uruguay),
(English_Americans, population place, Seattle metropolitan area),
(English_Americans, population place, Chicago metropolitan area),
(English_Americans, population place, Guatemala).
However, there are a few false negative triples.

Alternatives to Closed World Assumption

- Schema-aware Closed World Assumption (SCWA): given a KG G, its schema S and a new triple (s,p,o): if (s,p,o) ∈ Closure(G ∪ S), then (s,p,o) is correct; otherwise, (s,p,o) is incorrect (Wang et al. 2023)

- Open World Assumption: given a KG G, its schema S and a new triple (s,p,o): if S ∪ G ∪ (s,p,o) $\not\models \bot$, then (s,p,o) is correct; otherwise, (s,p,o) is incorrect

Limited expressiveness

- Fully expressiveness not enough
  - Given T^+ positive and T^- negative sample sets
  - $\forall(s,p,o) \in T^+,\ \phi^{\theta}_p(\mathbf{e}_s, \mathbf{e}_o) \le \lambda_p$
  - $\forall(s,p,o) \in T^-,\ \phi^{\theta}_p(\mathbf{e}_s, \mathbf{e}_o) > \lambda_p$

Model	Symmetry	Antisymmetry	Inversion	Composition
SE	✗	✗	✗	✗
TransE	✗	✓	✓	✓
TransX	✓	✓	✗	✗
DistMult	✓	✗	✗	✗
ComplEx	✓	✓	✓	✗
RotatE	✓	✓	✓	✓

From Sun et al. - RotatE: Knowledge Graph Embeddings by Relational Rotation in Complex Space

THE UNIVERSITY *of* EDINBURGH

Inconsistency

Explanation for: dbo:governingBody Range dbo:Organisation

1)	dbo:Broadcaster **DisjointWith** dbo:Location	In NO other justifications
2)	dbpedia:Lower_Hutt dbo:broadcastArea dbpedia:Wellington	In NO other justifications
3)	dbo:broadcastArea **Domain** dbo:Broadcaster	In NO other justifications
4)	dbpedia:Lower_Hutt **Type** dbo:Location	In NO other justifications

Explanation for: dbo:governingBody Range dbo:Organisation

1)	dbo:SportsTeam **SubClassOf** schema:Organization	In NO other justifications
2)	'dbpedia:Keith_Andrews_(footballer)' dbo:team dbpedia:Republic_of_Ireland	In NO other justifications
3)	dbo:Location **DisjointWith** schema:Organization	In NO other justifications
4)	dbo:team **Range** dbo:SportsTeam	In NO other justifications
5)	dbpedia:Republic_of_Ireland **Type** dbo:Location	In NO other justifications

Note: the two relation assertions have been removed in latest DBpedia release, but still exist in DB15K dataset.

THE UNIVERSITY *of* EDINBURGH

How to Address the Expressiveness Issue

- Incorporate schema information in loss function good approaches expected to be applicable to different KGE methods
- Combine symbolic reasoning with KGE Use OWA to check if the triples learned from KGE methods are consistent with the schema
  - Bonus: consistent triples can be combined with existing triples and schema to infer further triples

Incorporate Schema Information in Loss Function

- Example schema axioms:
  - Relation equivalence: p ≡ q (p ⊑ q, q ⊑ p)

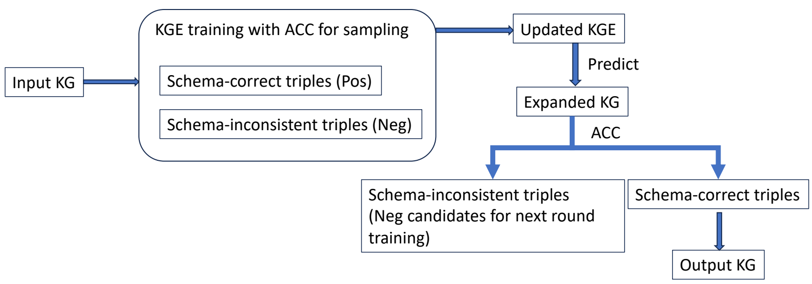
$$\phi(\langle s,p,o\rangle;\Theta) = \phi(\langle s,q,o\rangle;\Theta) \quad \forall s,o \in \mathcal{E}.$$

  - Inverse relations: p ≡ q^- (p ⊑ q^-, q^- ⊑ p)

$$\phi(\langle s,p,o\rangle;\Theta) = \phi(\langle o,q,s\rangle;\Theta) \quad \forall s,o \in \mathcal{E}.$$

- Scalable implementation of such revised scoring functions often demands KGE dependent revisions
- Optional reading: https://link.springer.com/chapter/10.1007/978-3-319-71249-9_40

Combine Symbolic Reasoning with KGE



ACC: Approximated Consistency Checking

- **Schema-correct**: consistent with the schema of the Knowledge Graph and satisfying the constraints, such as domain and range.
- **Schema-unknown**: they are consistent with the schema, but not yet satisfying the constraints, due to lack of some type information for their heads or tails, i.e., neither schema-correct nor schema-incorrect
- **Schema-inconsistent**: not consistent with the schema.

Optional reading: https://knowledge-representation.org/j.z.pan/pub/SICKLE2023.pdf

THE UNIVERSITY *of* EDINBURGH

Lecture Outline

- Motivation: KGE lack significant expressiveness
- Introduction: Limitations of classic KGE methods
- Focus: Solutions of these limitations
- Practical
  - Combine symbolic reasoning with KGE
  - Schema aware zero shot learning
- Next time we introduce tractable Description Logic EL