Learning Rules from Incomplete KGs Using Embeddings
Vin Thinh Ho¹, Daria Stepanova¹, Mohamed Gad-Eirab¹, Evgeny Kharlamov², Gerhard Weikum¹
hlthinh@mpi-inf.mpg.de, dstepano@mpi-inf.mpg.de, gadelrab@mpi-inf.mpg.de, evgeny.kharlamov@cs.ox.ac.uk, weikum@mpi-inf.mpg.de
¹Max Planck Institute for Informatics, Saarbrücken, Germany
²University of Oxford, Oxford, United Kingdom

1. Motivation and Contributions

Knowledge graphs: huge collections of positive unary and binary facts treated under Open World Assumption (e.g. isMarriedTo(clara, dave), researcher(dave))

Automatically constructed, thus incomplete ⇒ KG completion task

- Hard to interpret
- Allow for reasoning
- Not extensible
- Local patterns

Rule-based approach

- Interpretable

Embedding-based approach

- No reasoning
- Extensible (e.g., text)
- Global patterns

Our approach: rule-based with embeddings support

Challenges:

- Structurally different output
- Large embedding size
- Large rule search space

Contributions:

- Framework for rule learning with external sources
- Hybrid embedding based rule measure
- Experiments on real world KGs

2. Our Proposal: Rule Learning with External Sources

- Problem statement:
  
  Given: \( \mathcal{G} = (G, f) \)
  
  - Knowledge graph \( G \)
  
  - Probability function \( f: \) truthfulness of \( G \)'s missing facts

  Find: Ordered set of rules, which
  
  - Describe \( G \) well and predict highly probable facts based on \( f \)

- Our solution:

  Hybrid rule quality function to prune search space of rules \( r: \)
  
  \[
  \mu(r, \mathcal{P}) = (1 - \lambda) \times \mu_1(r, \mathcal{G}) + \lambda \times \mu_2(G, \mathcal{P})
  \]

  - Descriptive quality \( \mu_1 \) of rule \( r \) over \( G \):
    
    \[
    \mu_1 (r, \mathcal{G}) \rightarrow \alpha \in [0, 1]
    \]

  - Predictive quality \( \mu_2 \) of \( r \): truthfulness of predictions \( \mathcal{G} \), made by \( r \) on \( G \):
    
    \[
    \mu_2 (G, \mathcal{P}) \rightarrow \alpha \in [0, 1]
    \]

  - Weighting factor \( \lambda \in [0, 1] \) to control the distribution of \( \mu_1 \) and \( \mu_2 \)

  Realization of \( f \) and \( \mu_2 \) relying on embeddings:

  \[
  f(\text{fact}) = 0.5 \times (1/\text{subject_rank} + 1/\text{object_rank})
  \]

  \[
  \mu_2(G, \mathcal{P}) = \frac{\sum_{\text{source} \in G} f(\text{fact})}{|G|}
  \]

3. General Architecture

4. Rule Refinement

Extended AMIE [Galárraga, et al, VLDB 2015] (additions are in blue):

- Refinement operators: add
  
  - dangling atom
  
  - instantiated atom
  
  - closing atom
  
  - negated instantiated atom
  
  - negated closing atom

- Rule filtering:
  
  - language bias
  
  - support
  
  - head coverage
  
  - confidence
  
  - embedding-based measure (\( \mu \))

  - exception confidence:
    
    \[
    e \text{-} \text{conf}(r, \mathcal{G}) = \text{conf}(r', \mathcal{G})
    \]

    where \( r' \) = body(r) → body(r), not head(r)

5. Experiments

- Approximation of complete KG: original

- Available KG: random 80% of original KG, preserving the distribution of facts over predicates.

- Embedding models:

  - TransE, HoloE, SSP (with text)

- Examples of mined rules:

  - \( r_1: \) nationality(\( X, Y \)) ← graduated_from(X, Z), in_country(Z, Y), not research_uni(Z)

  - \( r_2: \) scriptwriter_of(X, Y) ← preceded_by(X, Z), scriptwriter_of(Y, Z), not iv_series(Z)