Rule Learning from Knowledge Graphs Guided by Embedding Models

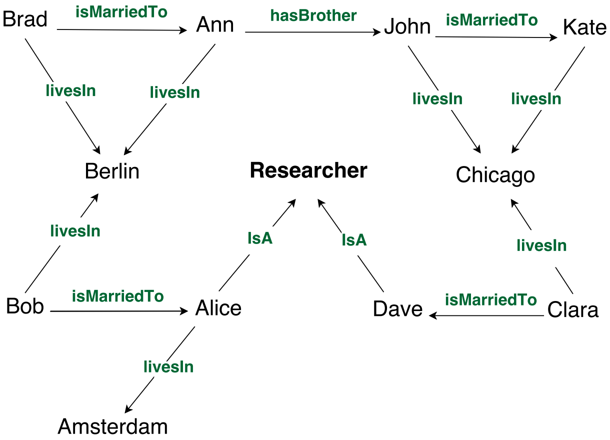
Vinh Thinh Ho[1], Daria Stepanova[1], Mohamed Gad-Elrab[1], Evgeny Kharlamov[2], Gerhard Weikum[1]

[1]Max Planck Institute for Informatics, Saarbrücken, Germany

[2]University of Oxford, Oxford, United Kingdom

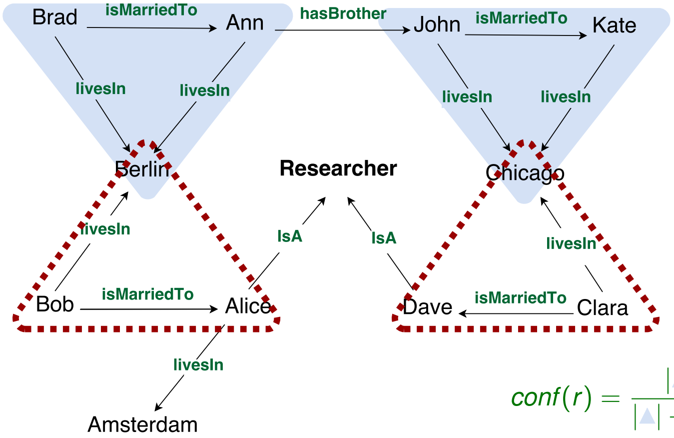
ISWC 2018

Rule Learning from KGs



Rule Learning from KGs

Confidence, e.g., WARMER [Goethals and den Bussche, 2002]
CWA: whatever is missing is false

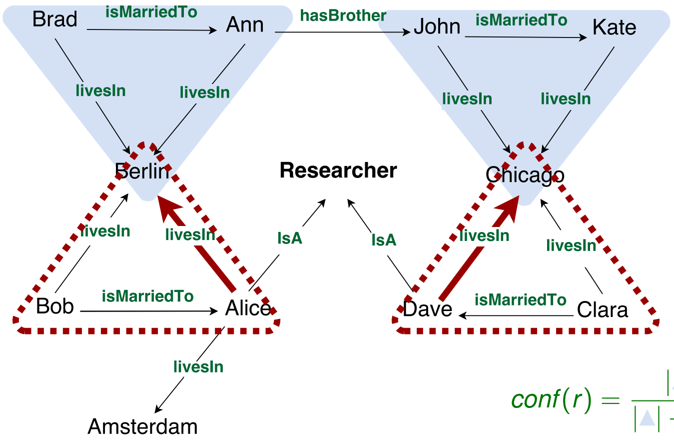


$$conf(r) = \frac{|\blacktriangle|}{|\blacktriangle| + |\triangle|} = \frac{2}{4}$$

$$r: \ livesIn(X, Y) \leftarrow isMarriedTo(Z, X), livesIn(Z, Y)$$

Rule Learning from KGs

Confidence, e.g., WARMER [Goethals and den Bussche, 2002]
CWA: whatever is missing is false



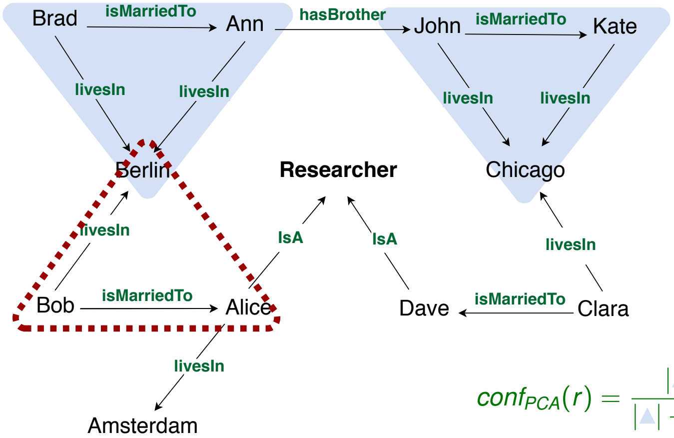
$$conf(r) = \frac{|\blacktriangle|}{|\blacktriangle| + |\triangle|} = \frac{2}{4}$$

$r: \; livesIn(X, Y) \leftarrow isMarriedTo(Z, X), livesIn(Z, Y)$

Rule Learning from KGs

PCA confidence AMIE [Galárraga *et al.*, 2015]

PCA: Since Alice has a living place already, all others are incorrect.

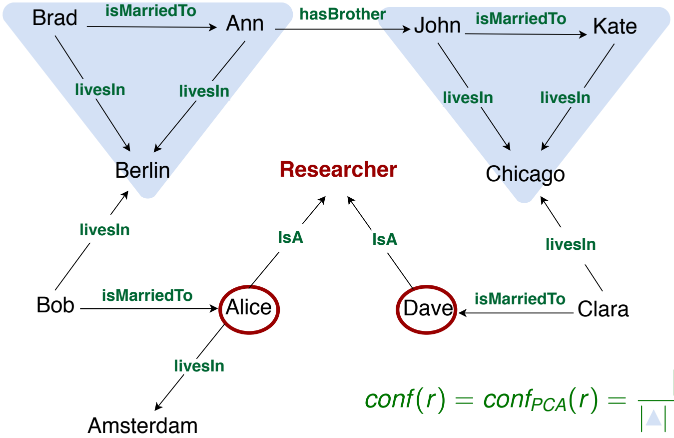


$$conf_{PCA}(r) = \frac{|\blacktriangle|}{|\blacktriangle| + |\triangle|} = \frac{2}{3}$$

$$r: \ livesIn(X, Y) \leftarrow isMarriedTo(Z, X), livesIn(Z, Y)$$

Rule Learning from KGs

Exception-enriched rules: [ISWC 2016, ILP 2016]

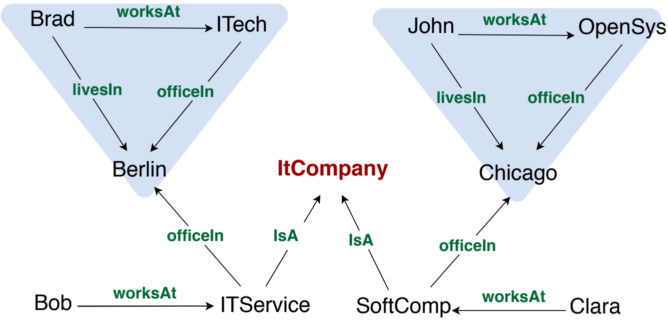


$$conf(r) = conf_{PCA}(r) = \frac{|\blacktriangle|}{|\blacktriangle| + |\triangle|} = 1$$

$r: \ livesIn(X, Y) \leftarrow isMarriedTo(Z, X), livesIn(Z, Y), not\ isA(X, researcher)$

Absurd Rules due to Data Incompleteness

Problem: rules learned from highly incomplete KGs might be absurd..

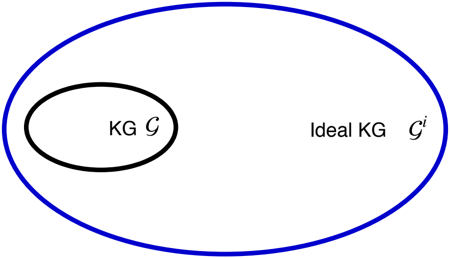


$$conf(r) = conf_{PCA}(r) = 1$$

$$livesIn(X, Y) \leftarrow worksAt(X, Z), officeIn(Z, Y), not\ isA(Z, itCompany)$$

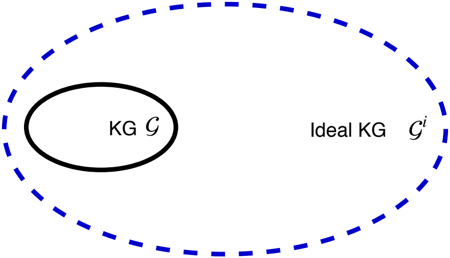
Ideal KG

$\mu(r, \mathcal{G}^i)$: measure quality of the rule r on $\mathcal{G}^i$



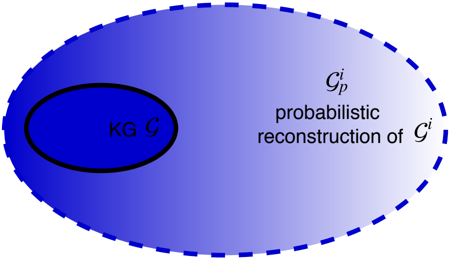
Ideal KG

$\mu(r, \mathcal{G}^i)$: measure quality of the rule r on $\mathcal{G}^i$, but $\mathcal{G}^i$ is unknown



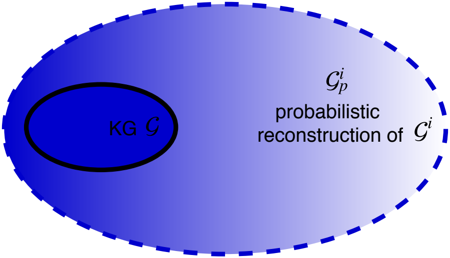
Probabilistic Reconstruction of Ideal KG

$\mu(r, \mathcal{G}_p^i)$: measure quality of r on $\mathcal{G}_p^i$



Hybrid Rule Measure

$$\mu(r, \mathcal{G}_p^i) = (1-\lambda) \times \mu_1(r, \mathcal{G}) + \lambda \times \mu_2(r, \mathcal{G}_p^i)$$



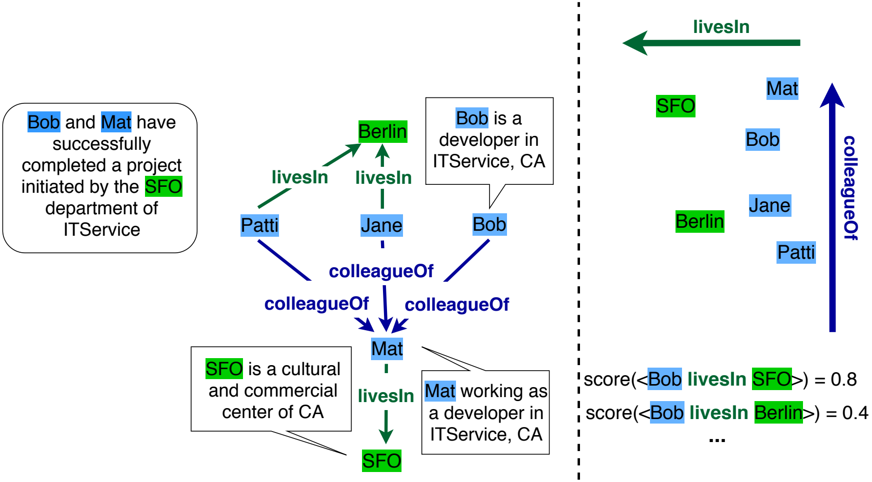
Hybrid Rule Measure

$$\mu(r, \mathcal{G}_p^i) = (1 - \lambda) \times \mu_1(r, \mathcal{G}) + \lambda \times \mu_2(r, \mathcal{G}_p^i)$$

- **$\lambda \in [0..1]$: weighting factor**
- **μ_1 : descriptive quality** of rule r over the available KG $\mathcal{G}$
  - confidence
  - PCA confidence
- **μ_2 : predictive quality** of r relying on $\mathcal{G}_p^i$ (probabilistic reconstruction of the ideal KG $\mathcal{G}^i$)

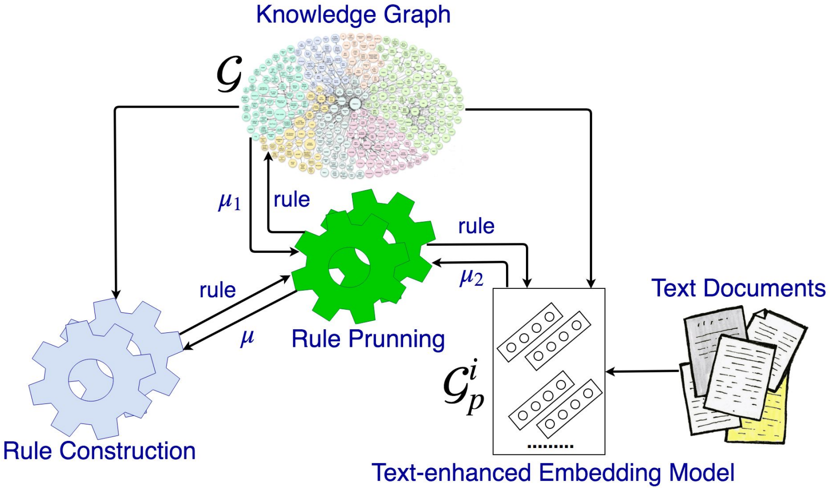
KG Embeddings

- Popular approach to KG completion, which proved to be effective
- Relies on translation of entities and relations into vector spaces



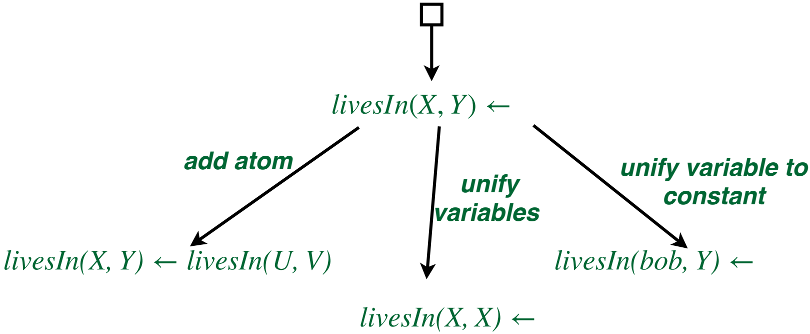
TransE [Bordes *et al.*, 2013], SSP [Xiao *et al.*, 2017], TEKE [Wang and Li, 2016]

Our Approach



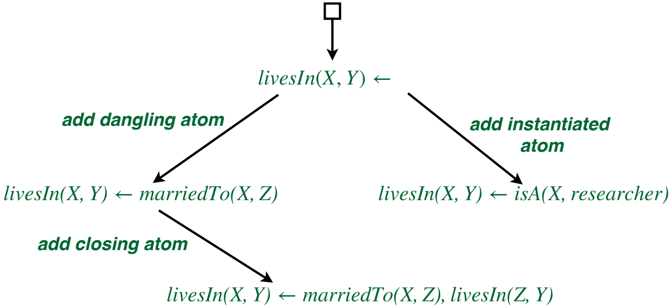
Rule Construction

- Clause exploration from general to specific
  - all first-order clauses: [Shapiro, 1991]



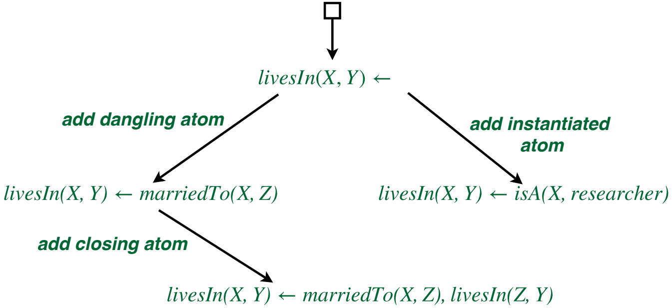
Rule Construction

- Clause exploration from general to specific
  - closed rules: AMIE [Galárraga *et al.*, 2015]
    $livesIn(X, Y) \leftarrow marriedTo(X, Z), livesIn(Z, Y)$



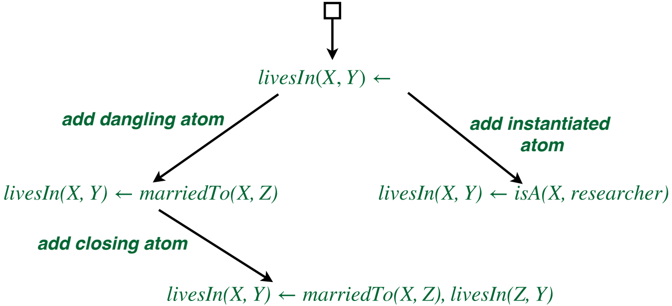
Rule Construction

- Clause exploration from general to specific
  - closed rules: AMIE [Galárraga *et al.*, 2015]
    $livesIn(X, Y) \leftarrow marriedTo(X, Z), livesIn(Z, Y)$



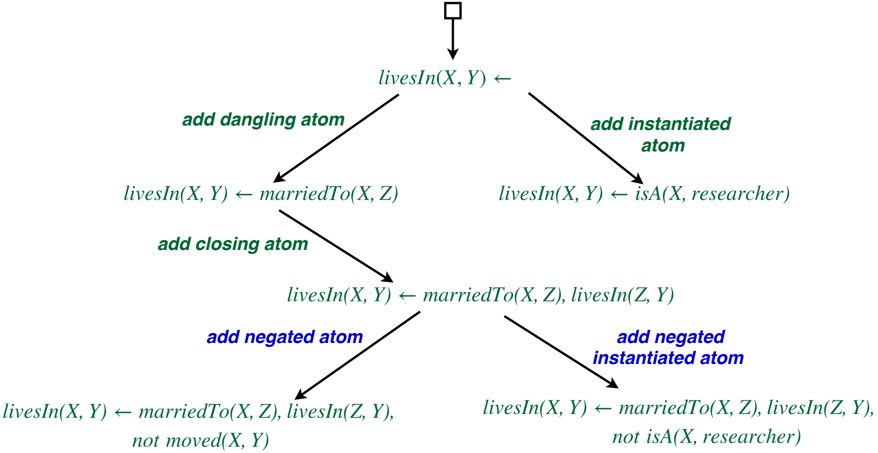
Rule Construction

- Clause exploration from general to specific
  - closed rules: AMIE [Galárraga *et al.*, 2015]
    $livesIn(X, Y) \leftarrow marriedTo(X, Z), livesIn(Z, Y)$



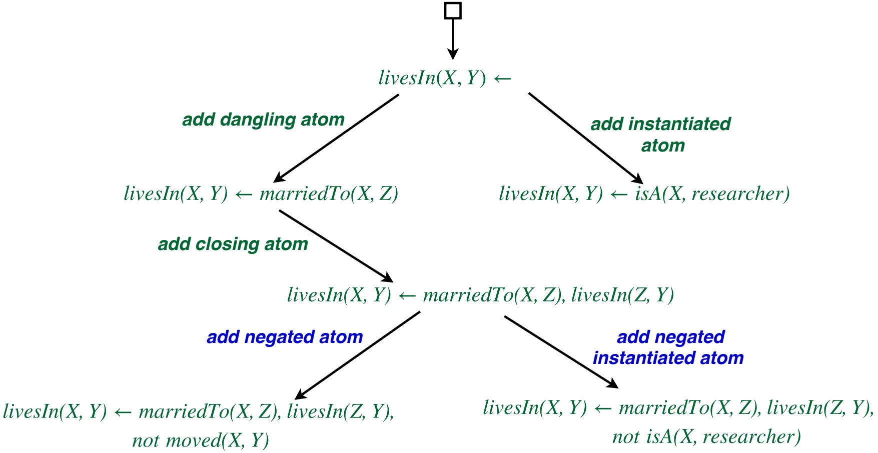
Rule Construction

- Clause exploration from general to specific
  - This work: closed and safe rules with negation
    $livesIn(X, Y) \leftarrow marriedTo(X, Z), livesIn(Z, Y), not\ isA(X, researcher)$

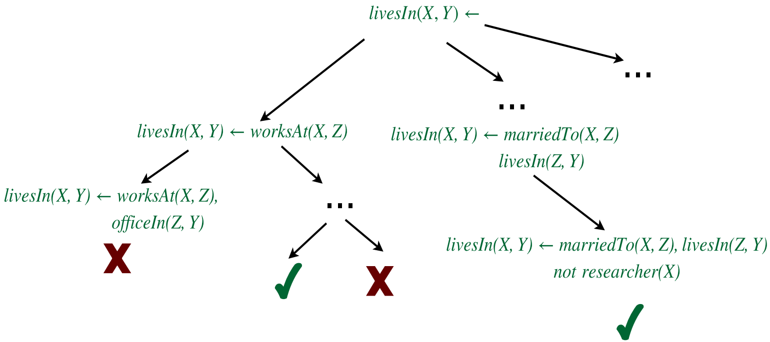


Rule Construction

- Clause exploration from general to specific
  - This work: closed and safe rules with negation
    $livesIn(X, Y) \leftarrow marriedTo(X, Z), livesIn(Z, Y), not\ isA(X, researcher)$



Rule Prunning



Prune rule search space relying on

- novel hybrid embedding-based rule measure

Embedding-based Rule Quality

- Estimate average quality of predictions made by a given rule r

$$\mu_2(r, \mathcal{G}_p^i) = \frac{1}{|predictions(r, \mathcal{G})|} \sum_{fact \in predictions(r, \mathcal{G})} \mathcal{G}_p^i(fact)$$

  - Rely on truthfulness of **predictions made by** r based on the probabilistic reconstruction $\mathcal{G}_p^i$ of $\mathcal{G}^i$

Embedding-based Rule Quality

- Estimate average quality of predictions made by a given rule r

$$\mu_2(r, \mathcal{G}_p^i) = \frac{1}{|predictions(r, \mathcal{G})|} \sum_{fact \in predictions(r,\mathcal{G})} \mathcal{G}_p^i(fact)$$

  - Rely on truthfulness of **predictions made by** r based on the probabilistic reconstruction $\mathcal{G}_p^i$ of $\mathcal{G}^i$

Example:

$livesIn(X, Y) \leftarrow marriedTo(X, Z), livesIn(Z, Y)$

- Rule predictions: $livesIn(mat, monterey)$,$livesIn(dave, chicago)$

$$\mu_2(r, \mathcal{G}_p^i) = \frac{\mathcal{G}_p^i(<\text{mat } \textbf{livesIn} \text{ monterey}>) + \mathcal{G}_p^i(<\text{dave } \textbf{livesIn} \text{ chicago}>)}{2}$$

Embedding-based Rule Quality

- Estimate average quality of predictions made by a given rule r

$$\mu_2(r, \mathcal{G}_p^i) = \frac{1}{|predictions(r, \mathcal{G})|} \sum_{fact \in predictions(r, \mathcal{G})} \mathcal{G}_p^i(fact)$$

  - Rely on truthfulness of **predictions made by** r based on the probabilistic reconstruction $\mathcal{G}_p^i$ of $\mathcal{G}^i$

Example:

$livesIn(X, Y) \leftarrow marriedTo(X, Z), livesIn(Z, Y), not\ isA(X, surfer)$

- Rule predictions: $\cancel{livesIn(mat, monterey)}$, $livesIn(dave, chicago)$

$$\mu_2(r, \mathcal{G}_p^i) = \frac{\mathcal{G}_p^i(<\text{dave } \textbf{livesIn} \text{ chicago}>)}{1}$$

- $\mu_2(r, \mathcal{G}_p^i)$ goes down for noisy exceptions

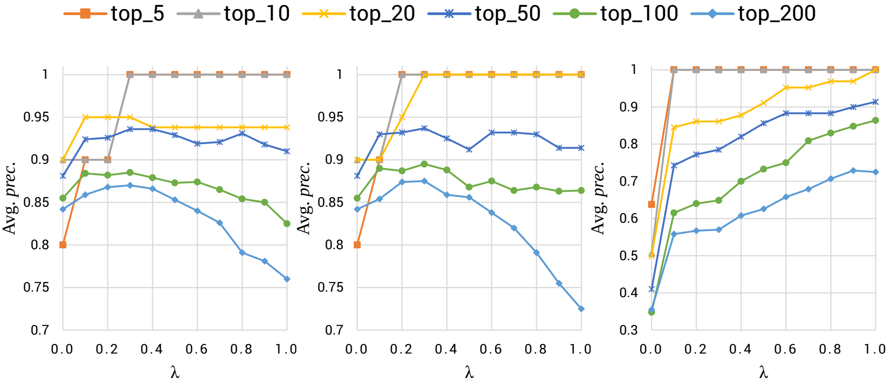
Evaluation Setup

- Datasets:
  - FB15K: 592K facts, 15K entities and 1345 relations
  - Wiki44K: 250K facts, 44K entities and 100 relations
- Training graph $\mathcal{G}$: remove 20% from the available KG
- Embedding models $\mathcal{G}_p^i$:
  - TransE [Bordes *et al.*, 2013], HolE [Nickel *et al.*, 2016]
  - With text: SSP [Xiao *et al.*, 2017]
- Goals:
  - Evaluate effectiveness of our hybrid rule measure

$$\mu(r, \mathcal{G}_p^i) = (1 - \lambda) \times \mu_1(r, \mathcal{G}) + \lambda \times \mu_2(r, \mathcal{G}_p^i)$$

  - Compare against state-of-the-art rule learning systems

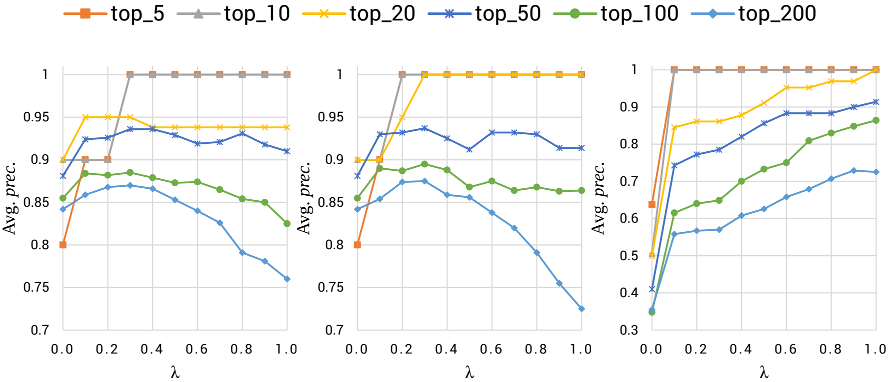
Evaluation of Hybrid Rule Measure



(a) Conf-HolE (b) Conf-SSP (c) PCA-SSP

Precision of *top-k* rules ranked using variations of μ on FB15K.

Evaluation of Hybrid Rule Measure



(a) Conf-HolE (b) Conf-SSP (c) PCA-SSP

Precision of *top-k* rules ranked using variations of μ on FB15K.

- Positive impact of embeddings in all cases for $\lambda = 0.3$
- **Note:** in (c) comparison to AMIE [Galárraga *et al.*, 2015] ($\lambda = 0$)

Example Rules

Examples of rules learned from Wikidata

Script writers stay the same throughout a sequel, but not for TV series

r_1 : $scriptwriterOf(X, Y) \leftarrow precededBy(Y, Z), scriptwriterOf(X, Z),$ **not** $isA(Z, tvSeries)$

Nobles are typically married to nobles, but not in the case of Chinese dynasties

r_2 : $nobleFamily(X, Y) \leftarrow spouse(X, Z), nobleFamily(Z, Y),$ **not** $isA(Y, chineseDynasty)$

Related Work

Inductive logic programming
[Muggleton *et al*, 1990]

Data mining
[Mannila, 1996]

Learning from interpretations
[De Raedt *et al*, 2014]

Neural-based rule learning
[Yang *et al*, 2017, Evans *et al, 2018*]

Relational association rule learning
[Goethals *et al*, 2002, Galárraga *et al*, 2015, ISWC 2016, ILP 2017]

Interactive pattern mining
[Goethals *et al*, 2011, Dzyuba *et al*, 2017]

Exploiting cardinality info
[ISWC 2017]

Rule Learning guided by embedding models

✓

Conclusion

- Summary:
  - Framework for learning rules from KGs with external sources
  - Hybrid embedding-based rule quality measure
  - Experimental evaluation on real-world KGs
  - Approach is orthogonal to a concrete embedding used
- Outlook:
  - Other rule types, e.g., with existentials in the head or constraints
  - Plug-in portfolio of embeddings
  - Mimic framework of exact learning [Angluin, 1987] by establishing complex queries to embeddings

References I

Dana Angluin.
Queries and concept learning.
Machine Learning, 2(4):319–342, 1987.

Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana Yakhnenko.
Translating Embeddings for Modeling Multi-relational Data.
In *Proceedings of NIPS*, pages 2787–2795, 2013.

Luis Galárraga, Christina Teflioudi, Katja Hose, and Fabian M. Suchanek.
Fast rule mining in ontological knowledge bases with AMIE+.
VLDB J., 24(6):707–730, 2015.

Bart Goethals and Jan Van den Bussche.
Relational association rules: Getting warmer.
In *PDD*, 2002.

Maximilian Nickel, Lorenzo Rosasco, and Tomaso A. Poggio.
Holographic embeddings of knowledge graphs.
In *AAAI*, 2016.

Ehud Y. Shapiro.
Inductive inference of theories from facts.
In *Computational Logic - Essays in Honor of Alan Robinson*, pages 199–254, 1991.

Zhigang Wang and Juan-Zi Li.
Text-enhanced representation learning for knowledge graph.
In *IJCAI*, 2016.

Han Xiao, Minlie Huang, Lian Meng, and Xiaoyan Zhu.
SSP: semantic space projection for knowledge graph embedding with text descriptions.
In *AAAI*, 2017.