Learning Rules from Incomplete KGs using Embeddings[*]

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

[1]Max Planck Institute for Informatics, Saarland Informatics Campus, Germany
[2]University of Oxford, Oxford, United Kingdom

Abstract. Rules over a Knowledge Graph (KG) capture interpretable patterns in data and various methods for rule learning have been proposed. Since KGs are inherently incomplete, rules can be used to deduce missing facts. Statistical measures for learned rules such as confidence reflect rule quality well when the KG is reasonably complete; however, these measures might be misleading otherwise. So it is difficult to learn high-quality rules from the KG alone, and scalability dictates that only a small set of candidate rules is generated. Therefore, the ranking and pruning of candidate rules is a major problem. To address this issue, we propose a rule learning method that utilizes probabilistic representations of missing facts. In particular, we iteratively extend rules induced from a KG by relying on feedback from a precomputed embedding model over the KG and external information sources including text corpora. Experiments on real-world KGs demonstrate the effectiveness of our novel approach both with respect to the quality of the learned rules and fact predictions that they produce.

Motivation. Rules are widely used to represent relationships and dependencies among data items in datasets and to capture the underlying patterns in data [1,16]. Applications of rules include health-care [24], telecommunications [9], manufacturing [2,10,11,12,7], and commerce [17,6]. In order to facilitate rule construction, a variety of rule learning methods have been developed, see e.g. [8] for an overview. Moreover, various statistical measures such as confidence, actionability, and unexpectedness to evaluate the *quality* of the learned rules have been proposed.

Rule learning has recently been adapted for the setting of Knowledge Graphs (KGs) [4,22] where data is represented as a graph of entities interconnected via relations and labeled with classes, or more formally as a set of grounded binary and unary atoms typically referred to as facts. Examples of large-scale KGs include Wikidata [20], Yago [19], and Google's KG. Since many KGs are constructed from semi-structured knowledge, such as Wikipedia, or harvested from the Web with a combination of statistical and linguistic methods, they are inherently incomplete [15,4].

Rules over KGs are of the form $head \leftarrow body$, where $head$ is a binary atom and $body$ is a conjunction of, possibly negated, binary or unary atoms. When rules are automatically learned, statistical measures like support and confidence are used to assess the quality of rules. Most notably, the confidence of a rule is the fraction of facts predicted by the rule that are indeed true in the KG. However, this is a meaningful measure for rule quality only when the KG is reasonably complete. For rules learned from largely incomplete KGs, confidence and other measures may be misleading, as they do not reflect the patterns in the

---

[*] This poster is accompanying our conference paper [5].

missing facts. For example, a KG that knows only (or mostly) male CEOs would yield a heavily biased rule $gender(X, male) \leftarrow isCEO(X, Y), isCompany(Y)$, which does not extend to the entirety of valid facts beyond the KG. Therefore, it is crucial that rules can be ranked by a meaningful quality measure, which accounts for KG incompleteness.

Example. Consider a KG about people's jobs, residence and spouses as well as office locations and headquarters of companies. Suppose a rule learning method has computed the following two rules:

$$r_1 : livesIn(X, Y) \leftarrow worksFor(X, Z), hasOfficeIn(Z, Y) \tag{1}$$
$$r_2 : livesIn(Y, Z) \leftarrow marriedTo(X, Y), livesIn(X, Z) \tag{2}$$

The rule r_1 is quite noisy, as companies have offices in many cities, but employees live and work in only one of them, while the rule r_2 clearly is of higher quality. However, depending on how the KG is populated with instances, the rule r_1 could nevertheless score higher than r_2 in terms of confidence measures. For example, the KG may contain only a specific subset of company offices and only people who work for specific companies. If we knew the complete KG, then the rule r_2 should presumably be ranked higher than r_1.

Suppose we had a perfect oracle for the true and complete KG. Then we could learn even more sophisticated rules such as

$$r_3 : livesIn(X, Y) \leftarrow worksFor(X, Z), hasHeadquarterIn(Z, Y), not\ locatedIn(Y, USA).$$

This rule would capture that most people work in the same city as their employers' headquarters, with the USA being an exception (assuming that people there are used to long commutes). This is an example of a rule that contains a negated atom in the rule body (so it is no longer a Horn rule) and has a partially grounded atom with a variable and a constant as its arguments.

Problem. The problem of KG incompleteness has been tackled by methods that (learn to) predict missing facts for KGs (or actually missing relational edges between existing entities). A prominent class of approaches is statistics-based and includes tensor factorization, e.g., [14] and neural-embedding-based models, e.g. [3,13]. Intuitively, these approaches turn a KG, possibly augmented with external sources such as text [23,25,18], into a probabilistic representation of its entities and relations, known as *embeddings*, and then predict the likelihood of missing facts by reasoning over the embeddings [21].

These kinds of embeddings can complement the given KG and are a potential asset in overcoming the limitations that arise from incomplete KGs. Consider the following gedankenexperiment: we compute embeddings from the KG and external text sources, that can then be used to predict the complete KG that comprises all valid facts. This would seemingly be the perfect starting point for learning rules, without the bias and quality problems of the incomplete KG. However, this scenario is way oversimplified. The embeddings-based fact predictions would themselves be very noisy, yielding also many spurious facts. Moreover, the computation of all fact predictions and the induction of all possible rules would come with a big scalability challenge: in practice, we need to restrict ourselves to computing merely small subsets of likely fact predictions and promising rule candidates.

Our Approach. In this work we propose an approach for rule learning guided by external sources that allows to learn high-quality rules from incomplete KGs. In particular, our method extends rule learning by exploiting probabilistic representations of missing facts computed by embedding models of KGs and possibly other external information sources.

More formally, let $\mathcal{G}$ be a KG, then a *probabilistic KG $\mathcal{P}$* is a pair $\mathcal{P} = (\mathcal{G}, f)$ where f is a probability function over the facts (or KGs edges), where we assume $f(a) = 1$ for each fact $a \in \mathcal{G}$, which is already known to be true. Our proposal is to learn rules that not only describe the available graph $\mathcal{G}$ well, but also predict highly probable facts based on the function f. The key questions here are how to define the quality of a given rule r based on $\mathcal{P}$ and how to exploit it during rule learning for pruning out not promising rules. We define a quality measure μ for rules over probabilistic KGs as a function $\mu : (r, \mathcal{P}) \mapsto \alpha$, where $\alpha \in [0, 1]$. To measure the quality μ of r over $\mathcal{P}$ we propose:

- to measure the quality μ_1 of r over $\mathcal{G}$, where $\mu_1 : (r, \mathcal{G}) \mapsto \alpha \in [0, 1]$,
- to measure the quality μ_2 of $\mathcal{G}_r$ (i.e., $\mathcal{G}$ extended with edges derived using r) by relying on $\mathcal{P}_r = (\mathcal{G}_r, f)$, where $\mu_2 : (\mathcal{G}', (\mathcal{G}, f)) \mapsto \alpha \in [0, 1]$ for $\mathcal{G}' \supseteq \mathcal{G}$ is the quality of extensions $\mathcal{G}'$ of $\mathcal{G}$ over the signature of $\mathcal{G}$ given f, and
- to combine the result as the weighted sum.

That is, we define our hybrid rule quality function $\mu(r, \mathcal{P})$ as follows:

$$\mu(r, \mathcal{P}) = (1 - \lambda) \times \mu_1(r, \mathcal{G}) + \lambda \times \mu_2(\mathcal{G}_r, \mathcal{P}) \tag{3}$$

In this formula μ_1 can be any classical quality measure of rules over complete graphs. Intuitively, $\mu_2(\mathcal{G}_r, \mathcal{P})$ is the quality of $\mathcal{G}_r$ wrt f that allows us to capture the information about facts missing in $\mathcal{G}$ that are relevant for r. The weighting factor λ, we call it *embedding weight*, allows one to choose whether to rely more on the classical measure μ_1 or on the measure μ_2 of the quality of the extension $\mathcal{G}_r$.

We propose to realize this approach by iteratively constructing rules over a KG and by collecting feedback from a precomputed embedding model, through specific queries issued to the model for assessing the quality of (partially constructed) rule candidates. This way, the rule induction loop is interleaved with the guidance from the embeddings, and we avoid scalability problems. Our machinery is also more expressive than many prior works on rule learning from KGs, by allowing non-monotonic rules with negated atoms as well as partially grounded atoms. Within this framework, we devise confidence measures that capture rule quality better than previous techniques and thus improve the ranking of rules.

Contribution. We propose a rule learning approach guided by external sources, and show how to learn high-quality rules by utilizing feedback from embedding models. We implement our approach and present extensive experiments on real-world KGs, demonstrating the effectiveness of our approach with respect to both the quality of the mined rules and predictions that they produce. Our code and data are made available to the research community at https://github.com/hovinhthinh/RuLES.

References

1. R. Agrawal, T. Imieliński, and A. Swami. Mining association rules between sets of items in large databases. *SIGMOD Rec.*, 22(2):207–216, 1993.
2. E. Auriol, M. Manago, and J. Guiot-Dorel. Cassiopée: Fehlerdiagnose von CFM 56-3 triebwerken für boing 737 flugzeuge. *KI*, 10(1):47–53, 1996.

3. A. Bordes, N. Usunier, A. García-Durán, J. Weston, and O. Yakhnenko. Translating Embeddings for Modeling Multi-relational Data. In *NIPS*, pages 2787–2795, 2013.
4. L. Galarraga, C. Teflioudi, K. Hose, and F. M. Suchanek. Fast rule mining in ontological knowledge bases with AMIE+. *VLDB J.*, 24(6):707–730, 2015.
5. V. T. Ho, D. Stepanova, M. Gad-Elrab, E. Kharlamov, and G. Weikum. Rule learning from knowledge graphs guided by embedding models. In *ISWC*, 2018.
6. S. Im, Z. W. Ras, and H. Wasyluk. Action rule discovery from incomplete data. *Knowl. Inf. Syst.*, 25(1):21–33, 2010.
7. E. Kharlamov, O. Savkovic, G. Xiao, R. Peñaloza, G. Mehdi, M. Roshchin, and I. Horrocks. Semantic Rules for Machine Diagnostics: Execution and Management. In *CIKM*, pages 2131–2134, 2017.
8. S. Kotsiantis and D. Kanellopoulos. Association rules mining: A recent overview. *GESTS Int. Trans. on CS and Eng.*, 32(1):71–82, 2006.
9. H. Mannila, H. Toivonen, and A. I. Verkamo. Discovering frequent episodes in sequences. In *KDD-95*, pages 210–215, 1995.
10. G. Mehdi, E. Kharlamov, O. Savkovic, G. Xiao, E. G. Kalayci, S. Brandt, I. Horrocks, M. Roshchin, and T. A. Runkler. Semantic Rule-Based Equipment Diagnostic. In *ISWC*, pages 225–240, 2017.
11. G. Mehdi, E. Kharlamov, O. Savkovic, G. Xiao, E. G. Kalayci, S. Brandt, I. Horrocks, M. Roshchin, and T. A. Runkler. Semantic Rule-Based Equipment Diagnostics. In *ISWC*, pages 314–333, 2017.
12. G. Mehdi, E. Kharlamov, O. Savkovic, G. Xiao, E. G. Kalayci, S. Brandt, I. Horrocks, M. Roshchin, and T. A. Runkler. SemDia: Semantic Rule-Based Equipment Diagnostics Tool. In *CIKM*, pages 2507–2510, 2017.
13. M. Nickel, L. Rosasco, and T. A. Poggio. Holographic embeddings of knowledge graphs. In *AAAI*, pages 1955–1961, 2016.
14. M. Nickel, V. Tresp, and H.-P. Kriegel. A three-way model for collective learning on multi-relational data. In *ICML*, pages 809–816, 2011.
15. H. Paulheim. Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic Web*, 8(3):489–508, 2017.
16. G. Piatetsky-Shapiro. Discovery, analysis, and presentation of strong rules. In *KDD*, pages 229–248. 1991.
17. Z. W. Ras and A. Wieczorkowska. Action-rules: How to increase profit of a company. In *PKDD*, pages 587–592, 2000.
18. M. Ringsquandl, E. Kharlamov, D. Stepanova, M. Hildebrandt, S. Lamparter, R. Lepratti, I. Horrocks, and P. Kröger. Event-Enhanced Learning for KG Completion. In *ESWC*, pages 541–559, 2018.
19. F. M. Suchanek, G. Kasneci, and G. Weikum. Yago: A Core of Semantic Knowledge. In *WWW*, pages 697–706, 2007.
20. D. Vrandecic and M. Krötzsch. Wikidata: a free collaborative knowledgebase. *CACM*, 57(10):78–85, 2014.
21. Q. Wang, Z. Mao, B. Wang, and L. Guo. Knowledge graph embedding: A survey of approaches and applications. *IEEE Trans. Knowl. Data Eng.*, 29(12):2724–2743, 2017.
22. Z. Wang and J. Li. RDF2Rules: Learning rules from RDF knowledge bases by mining frequent predicate cycles. *CoRR*, abs/1512.07734, 2015.
23. Z. Wang and J. Li. Text-enhanced representation learning for knowledge graph. In *IJCAI*, pages 1293–1299, 2016.
24. J. Wojtusiak. Rule learning in healthcare and health services research. In *Machine Learning in Healthcare Informatics*, pages 131–145. 2014.
25. H. Xiao, M. Huang, L. Meng, and X. Zhu. SSP: semantic space projection for knowledge graph embedding with text descriptions. In *AAAI*, pages 3104–3110, 2017.