Negative Sampling and Rule Mining for Explainable Link Prediction in Knowledge Graphs

Kamrul Islam [1], Sabeur Aridhi [1], Malika Smail-Tabbone [1] Details

1 CAPSID - Computational Algorithms for Protein Structures and Interactions
Inria Nancy - Grand Est, LORIA - AIS - Department of Complex Systems, Artificial Intelligence & Robotics

Abstract : Several KG embedding methods were proposed to learn low dimensional vector representations of entities and relations of a KG. Such representations facilitate the link prediction task, in the service of inference and KG completion. In this context, it is important to achieve both an efficient KG embedding and explainable predictions. During learning of efficient embeddings, sampling negative triples was highlighted as an important step as KGs only have observed positive triples. We propose an efficient simple negative sampling (SNS) method based on the assumption that the entities which are closer in the embedding space to the corrupted entity are able to provide high-quality negative triples. As for explainability, it actually constitutes a thriving research question especially when it comes to analyse KGs with their rich semantics rooted in description logics. Hence, we propose in this paper a new rule mining method on the basis of learned embeddings. We extensively evaluate our proposals through several experiments. We evaluate our SNS sampling method plugged to several KG embedding models through link prediction task performances on well-known datasets. Experimental results show that the SNS improves the prediction performance of KG embedding models, and outperforms the existing sampling methods. To assess the performance of our rule mining method with and without SNS, we mine and evaluate rules on three popular datasets. The extracted rules are evaluated as knowledge nuggets extracted from the KG and also as support for explainable link prediction. The overall results are good and open the way to many improvements and new perspectives.

Keywords : knowledge graph embedding | link prediction | negative sampling | rule mining | explainability

Document type : Journal articles

Domain : Chemical Sciences