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

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. Among the existing efficient embedding, sampling negative instances is highlighted as an important step as KGs only have observed positive triples. We propose an efficient sampling negative instance (SN) method based on the assumption that the entities which are close in the embedding space to the correct entity are also close in the embedding space to the candidate entity. As far as we know, it actually contradicts a common research question especially when it comes to complex KGs with their rich semantics rooted in description logics. Hence, we propose in this paper a new rule mining method based on the basis of learned embeddings. We extensively evaluate our proposals through several experiments. We evaluate our SN sampling method plugged to several KG embedding models through link prediction task performances on well-known datasets. Experimental results show that the SNP improves the prediction performance of KG embedding models, and supports the existing sampling methods. To testify the performance of our rule mining method with and without SN, 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, SN mining, explainable