Knowledge Graph Representation with Embeddings

1 Introduction

Knowledge graphs (KGs) serve as structured repositories of real-world facts in the form of triples comprising of entities and relations. KGs such as *Yago* and *Wikidata* (Vrandečić and Krötzsch 2014) have been applied to a number of applications including question answering, rule mining and web search. Knowledge graph embeddings have recently emerged as a popular technique for representation learning, where entities and relations are represented by low-dimensional dense vectors that can capture the interactions within the knowledge graph and then used for predicting missing links. Several popular KG embedding models have been successfully used for the task of link prediction (Wang et al. 2017). Such models have gained considerable attention and are being exploited for various other semantic tasks.

2 Semantic Limitations

As the basic premise of KG embeddings is centered around the semantic relationships between various entities, there is a widespread notion that embeddings must be able to capture the semantics and features of KG entities and relations very well. Embeddings have been, therefore, used for many similarity-based tasks including entity similarity (Sun et al. 2020) and relation similarity (Kalo, Ehler, and Balke 2019), as well as conceptual clustering (Gad-Elrab et al. 2020; Wang et al. 2019).

While the results look promising, none of the previous works have performed a detailed analysis of the benefits of the embeddings across different datasets as well as across different entities within a single dataset. In some cases, a measurement of the consistency and scalability of the proposed embedding-based approach for different real-world datasets is largely lacking. The oversight of the limitations of KG embeddings and emphasis on the success for the simpler cases might prove misleading to research community.

In our recent work (Jain et al. 2021a), we aim to address this issue by performing a critical study of the characteristics of the latent vectors obtained from several KG embedding models and quantitatively measuring their ability for semantic representation and learning. The key insight from our detailed analysis is that while embedding models used for representation learning of KGs are assumed to encapsulate the semantics for entities and relations, in reality their semantic soundness is severely restricted and highly dependent on the datasets on which they are trained. These findings indicate that a thorough inspection of the advantages and weaknesses of KG embeddings is necessary when employing them for semantic tasks.

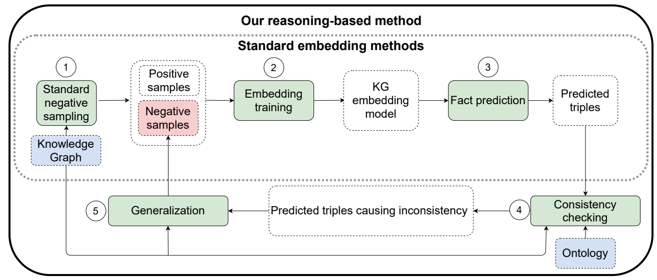

Figure 1: Standard embedding pipeline (grey dotted frame) and our reasoning-based method (black frame) in a nutshell

3 Reasoning-based Embeddings

In spite of their state-of-the-art performance on link prediction task, embedding models can lead to undesirable and nonsensical predictions (Wiharja et al. 2020) due to the limitations in the semantic representations for entities and relations of the KGs. This calls for more goal-oriented approaches in which ontological reasoning is used to verify and improve the actual predictions made by embedding models. To address this issue we proposed to improve the accuracy of embeddings using ontological reasoning in our work (Jain et al. 2021b).

The paper proposes an iterative method that dynamically identifies inconsistent predictions produced by a given embedding model via symbolic reasoning and feeds them as negative samples for retraining this model. The method starts with any available negative sampling procedure, e.g., (Kotnis and Nastase 2017; Zhang et al. 2019) and trains the embedding model as usual. Then, among predictions made by the model, those that cause inconsistency when being added to the KG are selected as negative samples for the next iteration of our method. In order to address the scalability problem that arises when integrating ontological reasoning into the training process, we proposed an advanced technique to generalize the inconsistent predictions to other semantically similar negative samples during retraining.

Our *ReasonKGE* approach can support any embedding model, and with the increasing number of iterations it yields better embeddings that make less inconsistent predictions and achieve higher prediction accuracy w.r.t. standard metrics. Experimental results demonstrated the improvements in accuracy of facts produced by our method compared to the state-of-the-art (details in the paper), indicating that reasoning-based training of embedding models can effectively improve their semantic representation.

References

Gad-Elrab, M. H.; Stepanova, D.; Tran, T.-K.; Adel, H.; and Weikum, G. 2020. ExCut: Explainable Embedding-Based Clustering over Knowledge Graphs. In *Int. Semantic Web Conf.*, 218–237.

Jain, N.; Kalo, J.-C.; Balke, W.-T.; and Krestel, R. 2021a. Do embeddings actually capture knowledge graph semantics? In *Extended Semantic Web Conference*, 143–159. Springer.

Jain, N.; Tran, T.-K.; Gad-Elrab, M.; and Stepanova, D. 2021b. Improving knowledge graph embeddings with ontological reasoning. In *International Semantic Web Conference*. Springer.

Kalo, J.-C.; Ehler, P.; and Balke, W.-T. 2019. Knowledge graph consolidation by unifying synonymous relationships. In *Int. Semantic Web Conf.*, 276–292.

Kotnis, B., and Nastase, V. 2017. Analysis of the impact of negative sampling on link prediction in knowledge graphs. *CoRR* abs/1708.06816.

Sun, Z.; Zhang, Q.; Hu, W.; Wang, C.; Chen, M.; Akrami, F.; and Li, C. 2020. A benchmarking study of embedding-based entity alignment for knowledge graphs. *Proc. VLDB Endow.* 13(12):2326–2340.

Vrandečić, D., and Krötzsch, M. 2014. Wikidata: A free collaborative knowledge base. *Communications of the ACM* 57(10):78–85.

Wang, Q.; Mao, Z.; Wang, B.; and Guo, L. 2017. Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering* 29(12):2724–2743.

Wang, C.; Pan, S.; Hu, R.; Long, G.; Jiang, J.; and Zhang, C. 2019. Attributed graph clustering: A deep attentional embedding approach.

Wiharja, K.; Pan, J. Z.; Kollingbaum, M. J.; and Deng, Y. 2020. Schema aware iterative knowledge graph completion. *J. Web Semant.* 65:100616.

Zhang, Y.; Yao, Q.; Shao, Y.; and Chen, L. 2019. Nscaching: Simple and efficient negative sampling for knowledge graph embedding. In *ICDE*, 614–625.