Do Embeddings Actually Capture Knowledge Graph Semantics?

Nitisha Jain[1] , **Jan-Christoph Kalo**[2] , **Wolf-Tilo Balke**[2] , **Ralf Krestel**[1]
[1]Hasso-Plattner-Institut, University of Potsdam, Germany
[2]Technische Universität Braunschweig, Germany
{nitisha.jain,ralf.krestel}@hpi.de, {kalo,balke}@ifis.cs.tu-bs.de

This extended abstract describes the main contributions of our work presented at the research track of the Extended Semantic Web Conference 2021 as a negative results paper (Jain et al. 2021).

1 Introduction

Knowledge graphs (KGs) serve as structured repositories of real-world facts in the form of triples comprising of entities and relations e.g. *(head entity, relation, tail entity)*. KGs such as *Yago* and *Freebase* 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 or triple completion in knowledge graphs (Wang et al. 2017) to address the issue of incompleteness in real-world KGs.

Due to their state-of-the-art performance, KG embedding models have gained considerable attention and are being exploited for various other semantic tasks. 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). Moreover, several previous works have attempted to leverage KG embeddings for performing reasoning with rules (Yang et al. 2015; Zhang et al. 2019).

While the results look promising, none of these 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. Our work aims 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.

2 Experiments and Results

KG embeddings are trained to capture the structural information of the underlying dataset. Ideally, if latent embeddings were able to embody all the latent features of entities, then entities with similar features would be similar in the vector space as well. That is, entities belonging to a particular type or class would result in similar vectors (Wang et al. 2019) and it should be possible to identify the entities belonging to a particular type from the KG embeddings. Therefore, in this work, we focus on verifying whether the entities can be categorized or assigned to their respective types from their corresponding latent vector representations. We perform a systematic investigation with two distinct sets of classification and clustering experiments for the entity embeddings in the vector space for two popular benchmark datasets - *Yago3-10* and *FB15K-237*. We used five different embedding techniques from the LibKGE library (Broscheit et al. 2020) that are widely popular: *TransE*, *RESCAL*, *ComplEx*, *DistMult* and *ConvE*.

Fig. 1 shows the weighted F1 measures for the Yago3-10 dataset across all the embedding models (color coded) as well the different classifiers (pattern coded) for entities belonging to classes at different levels of granularity in the Yago ontology (refer to (Jain et al. 2021) for details). These results indicate that the classification performance drops significantly for fine-grained classes, indicating that the semantic capability of embeddings is limited and heavily dependent on the dataset characteristics. While entities belonging to a small set of high-level, easy classes are relatively well-represented, the same does not hold true for most of the entities corresponding to other important classes in the dataset.

To compare the performance of the embeddings with a non-embeddings baseline, we leveraged a heuristics based technique *SDType* that simply uses the links between the entities to infer their type (Paulheim and Bizer 2013). It was seen that *SDType* was able to achieve quite competitive results as compared to several prominent embedding models. This provides strong evidence for the shortcomings of embeddings for representing fine-grained classes for which even simple statistical approach can already give compara-

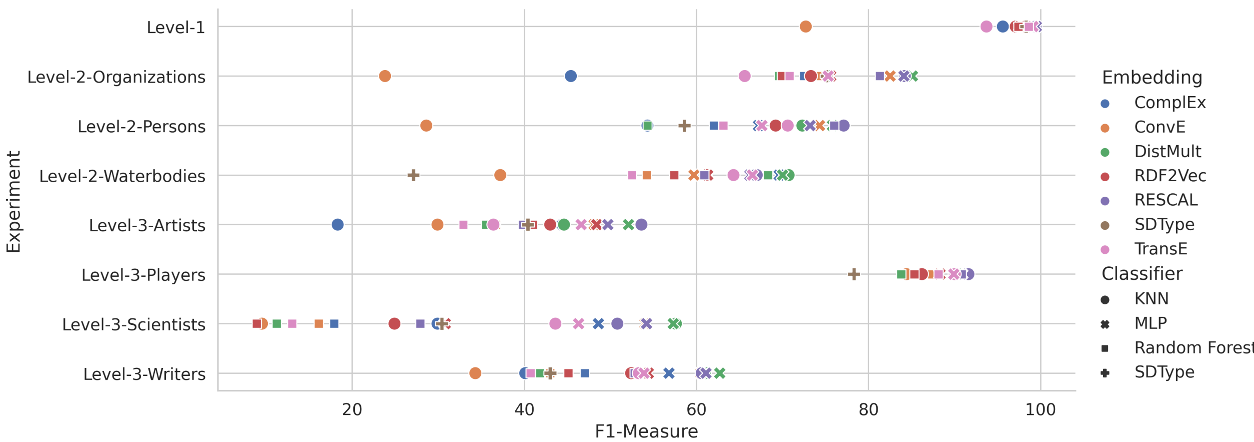

Figure 1: F1 measure for Yago3-10 classification experiments (best viewed in color).

ble results by deriving the semantics directly from the KG.

Furthermore, this work explores and identifies the factors that determine a good semantic representation of the entities and relations for any given KG, as well as the reasons for the shortcomings of current embedding models. An analysis of the relations associated with the different classes that were used in our experiments for the Yago3-10 dataset showed the presence of overlapping relations between entities belonging to different semantic types in real-world KGs (refer to Section 5 of (Jain et al. 2021)). This lack of unique relations makes it harder for embedding models to derive type-specific features about the entities, thus limiting their capability to learn similar entities or identify any common traits for all entities belonging to the same class. These observations raise strong doubts about the applicability of KG embeddings for various semantic and learning tasks.

3 Insight and Relevance to KR

The key insight from our detailed analysis in this work 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. While the research community is focused on developing novel architectures for training the KG embedding models, a careful eye on the generalizability of these models in terms of their semantic representation also deserves more attention. We believe the results from this work would serve as a precautionary tale and help the KR community become cognizant of the realistic semantic benefits of knowledge graph embeddings, such that they can make prudent decisions when applying these embeddings to new problem statements in representation learning and reasoning. We have made our datasets and code publicly available[1] to facilitate further research in this direction.

References

Broscheit, S.; Ruffinelli, D.; Kochsiek, A.; Betz, P.; and Gemulla, R. 2020. LibKGE - A knowledge graph embedding library for reproducible research. In *Proc. of the 2020 Conf. on Empirical Methods in Natural Language Processing: System Demonstrations*, 165–174.

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. 2021. Do embeddings actually capture knowledge graph semantics? In *Extended Semantic Web Conference*, 143–159. 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.

Paulheim, H., and Bizer, C. 2013. Type inference on noisy rdf data. In *Int. Semantic Web Conf.*, 510–525.

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.

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.

Yang, B.; Yih, S. W.-t.; He, X.; Gao, J.; and Deng, L. 2015. Embedding Entities and Relations for Learning and Inference in Knowledge Bases. In *Proc. of the Int. Conf. on Learning Representations (ICLR) 2015*.

Zhang, W.; Paudel, B.; Wang, L.; Chen, J.; Zhu, H.; Zhang, W.; Bernstein, A.; and Chen, H. 2019. Iteratively Learning Embeddings and Rules for Knowledge Graph Reasoning. In *Proc. of the 2019 World Wide Web Conf.*, WWW '19, 2366–2377.

[1]https://github.com/nitishajain/KGESemanticAnalysis