Extended Abstract: Re-evaluating Knowledge Graph Embedding Models Performance on Domain Specific Datasets

Victor Ciupec[0000−0002−1025−9791] and Peter Bloem[0000−0002−0189−5817]

Vrije Universiteit, Amsterdam, The Netherlands
v.ciupec@student.vu.nl, p.bloem@vu.nl

Abstract. Knowledge graph embedding models (KGEs) have mostly been evaluated and compared on generic benchmark datasets. In this paper we research if training on domain specific datasets instead has any performance impact. We conducted an hyperparameter search experiment on five KGE models and found that the models perform generally better on domain specific datasets, although the relative performance and hyperparameter impact are in line with previous studies.

Keywords: knowledge graph embedding · domain specific dataset · hyperparameters · link prediction.

Introduction and Motivation. Knowledge graph embedding (KGE) models have become popular solutions for the link prediction problem in knowledge graphs. KGE models learn algebraic representations of the entities and relations in a knowledge graph and use a scoring function to predict and rank new triples, thus separating correct from incorrect triples. Various KGE models vary not only in their embeddings and scoring function, but also in the choice of hyperparameters, most notably loss function and training strategy. This paper researches the question on whether KGE models perform differently on domain specific datasets compared to generic ones like Freebase and Wordnet and specifically which of the studied models perform best and which hyperparameters impact performance the most. We investigate how certain properties of domain specific datasets such as ontological structure and redundancy in expressing facts influence the performance and the selection of hyperparameters. Our experiments follow the methodology in [3], with which we compare the performance metrics.

Experimental Setup. We trained and evaluated the KGE models in this study using two domain specific datasets AIFB [1] and MUTAG [2], both bound by strong ontological information, with very detailed schemas that contain full hierarchies of classes and sub-classes using the RDF model. We compare performance of five of the most popular KGE models: RESCAL, DistMult, TransE, ComplEx and ConvE. Our experiment uses quasi-random search across a large discrete hyperparameter space, followed by a Bayesian optimization for fine tuning numerical ones. The best model is selected using the entity ranking protocol metrics MRR and Hits@10 after training the five best configurations for each architecture.

		MRR	Hits@10
AIFB	RESCAL	42.1	56.9
	TransE	46.01	59.8
	DistMult	**49.2**	**60.1**
	ComplEx	48.7	60.0
	ConvE	47.2	58.7
MUTAG	RESCAL	35.63	46.65
	TransE	26.82	47.39
	DistMult	**48.07**	**60.32**
	ComplEx	38.68	50.49
	ConvE	31.63	47.39

Table 1. Performance on test data of the best performing models.

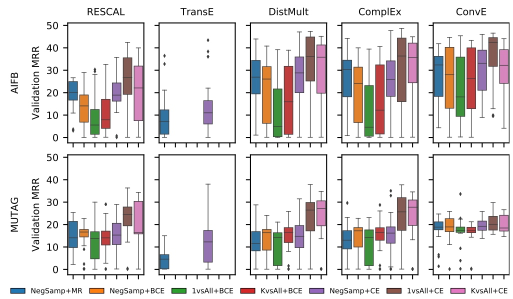

Fig. 1. Distribution of filtered MRR (%) on validation data over the train type/ loss function combinations during quasi-random hyperparameter search.

Results. In our study, DistMult has outperformed the other models on both domain specific datasets, followed by ComplEx. The relative performance gaps between the models is vastly reduced compared to the original publications, attributed perhaps to consistent training methodology. We also noted that the MRR and Hits@10 scores are remarkably higher for all models on AIFB than on MUTAG, justified by the existence of symmetrical relations in AIFB and by the KGE models' known ability to predict inverse relations.

We observed that the choice of loss function has by far the highest impact on the performance, followed by the training strategy. With some exceptions, cross entropy loss and 1vsAll training strategy performed best across the board. Furthermore, the higher MRR variance across the domain on AIFB suggests that models are more sensitive to hyperparameter change on AIFB than on MUTAG.

Compared with the results obtained by [3] on FB15K-237 and WNRR, most models performed notably better in our experiment, which can be explained by domain specific biases in our datasets. Our study[1] showed that domain specific datasets contribute to better KGE performance mostly due to the ontological structure and their intrinsic redundancy in expressing facts through the triples.

References

1. Bloehdorn, S., Sure, Y.: Kernel methods for mining instance data in ontologies. In: Aberer, K., Choi, K.S., Noy, N., Allemang, D., Lee, K.I., Nixon, L., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., Schreiber, G., Cudré-Mauroux, P. (eds.) The Semantic Web. pp. 58–71. Springer Berlin Heidelberg (2007)
2. Bühmann, L., Lehmann, J., Westphal, P., Bin, S.: Dl-learner structured machine learning on semantic web data. In: Companion Proceedings of the The Web Conference 2018. p. 467471. International WWW Conferences Steering Committee (2018)
3. Ruffinelli, D., Broscheit, S., Gemulla, R.: You CAN teach an old dog new tricks! on training knowledge graph embeddings. In: International Conference on Learning Representations (2020)

[1] https://github.com/Vixci/bachelor/blob/master/thesis.pdf