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

Victor Ciupec, Peter Bloem

What is the performance impact of training KGE models on *domain specific datasets*? Which *hyperparameters* make the most impact on performance?

Experiment

- We trained 5 KGE models (RESCAL, DistMult, TransE, ComplEx and ConvE) on two domain specific datasets AIFB and MUTAG
- Used libKGE for hyperparameter search:
  - quasi-random search across large discrete hyperparam space: training strategy (1vsAll, kvsAll, neg-sampling), loss function (cross-entropy, binary cross entropy, margin ranking),
  - followed by Bayesian optimization for numerical ones
- Best model is selected using entity ranking protocol metrics MRR and Hits@10

.

Results

- DistMult outperformed the other models on both AIFB and MUTAG (Table 1)
- Training type and and loss function have the highest impact on the performance
- Choice of loss function has by far the highest impact on performance, followed by the training strategy
- The best models had the 1vsAll training strategy and cross-entropy (CE) loss function (except TransE where neg-sampling + CE) (Figure 1)
- Overall all models performed better on domain specific datasets compared to generic ones in previous studies with same setup
- We concluded that ontological structure and intrinsic redundancy in expressing facts and domain specific datasets contribute to better KGE performance

		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.

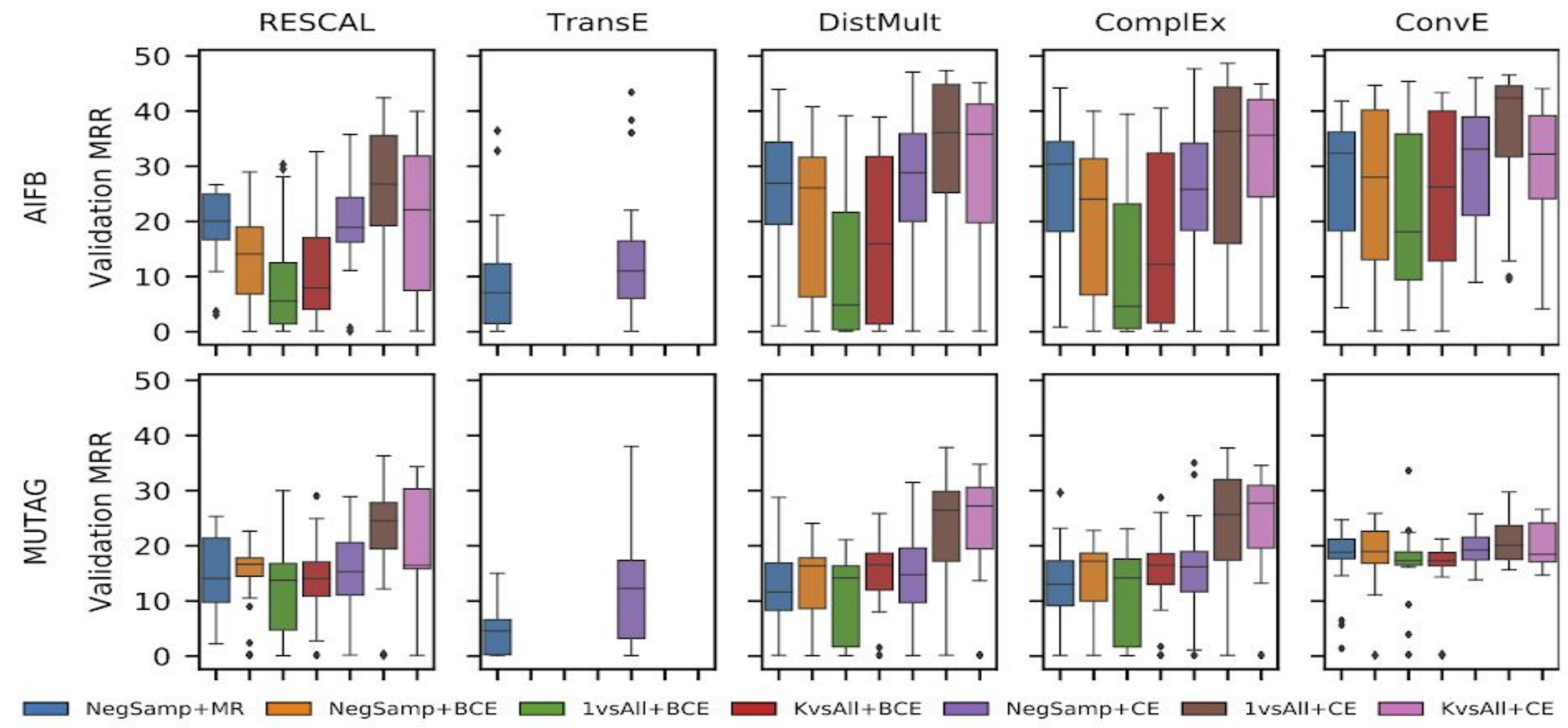

Figure 1: Distribution of filtered MRR (%) on validation data over the train type/ loss function combinations during quasi-random hyperparameter search.

http://bit.ly/vixcibsc