Towards Understanding the Impact of Schema Conceptualization on Knowledge Graph Embeddings

Brandon Dave, Dr. Cogan Shimizu
dave.15@wright.edu, cogan.shimizu@wright.edu
Computer Science & Engineering
KASTLE Lab, Joshi 390

Introduction

- Knowledge graph expresses data as relationships between nodes
- Ontology focuses on the modeling of KGs, as seen in the schema diagrams in Figure 1 and 2
- KG Embeddings (KGE) is the research field in utilizing machine learning to auto-fill and predict nodes and relationships

Is there an impact to KGE with the schema's design when utilizing a shallow versus rich approach?

Embedding Models

- Trans-E: prediction of entity-to-entity relationships
- Trans-R: prediction of an entity to a relationship
- DistMult: prediction of pair-wise interactions and semantic reasoning
- ComplEx: an improvement from DistMult utilizing a Hermitian Dot Product with complex values

Evaluation Metrics

- Hit@k: the rate of correct entities appearing within the top *k* entries
- Mean Rank (MR): average ranked position across all predictions
- Mean Reciprocal Rank (MRR): overall average of how close an instance of an entity and a relationship is to another node

Schema Concepts

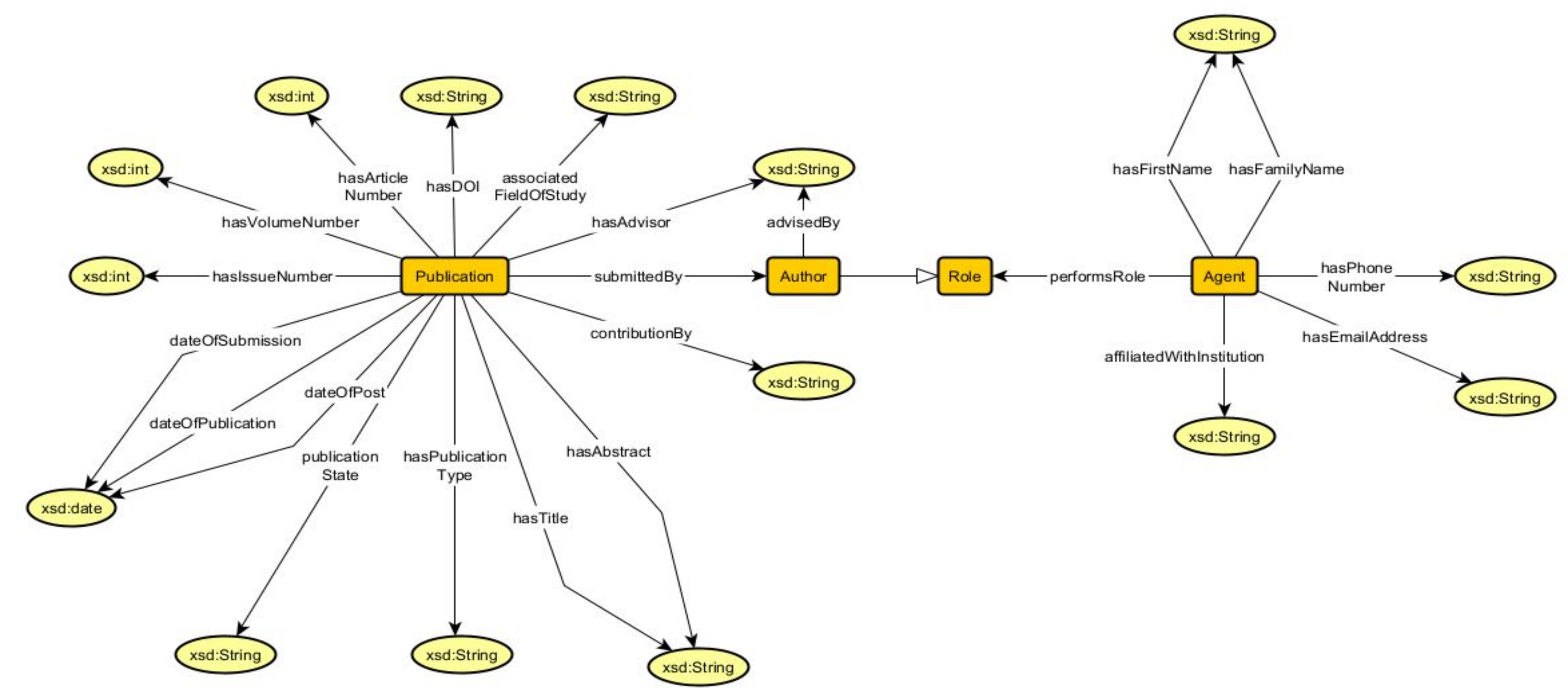

Figure 1: Shallow Design

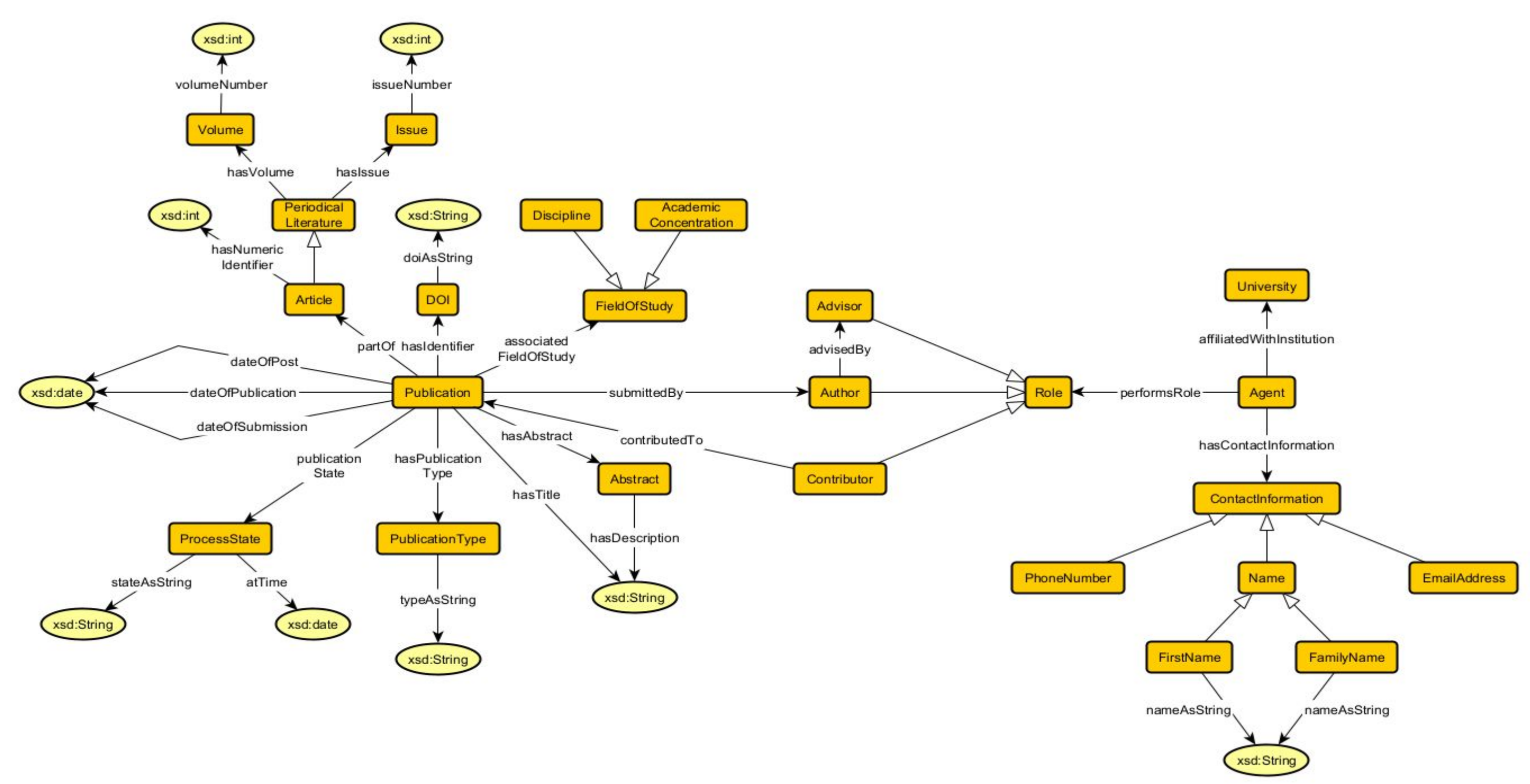

Figure 2: Rich Design

Preliminary Results

ALL	MR	MRR	H@1	H@3	H@10
TransE_NotClean					
Shallow_All	**17.671754**	**0.510615**	**0.411413**	**0.548074**	**0.711626**
Rich_All	37.021571	0.478205	0.391775	0.507560	0.651298
Delta	-19.349817	0.032411	0.019638	0.040514	0.060327
TransE_Clean					
Shallow_All	**16.004682**	**0.510416**	**0.409428**	**0.549654**	**0.712017**
Rich_All	41.615039	0.482275	0.395826	0.511988	0.654758
Delta	-25.610357	0.028141	0.013601	0.037665	0.057258
TransR					
Shallow_All	**413.633237**	0.092006	0.075774	0.095220	0.113197
Rich_All	434.098484	**0.112367**	**0.096677**	**0.115643**	**0.134727**
Delta	-20.465247	-0.020361	-0.020903	-0.020423	-0.021530
DistMult					
Shallow_All	**165.963736**	**0.340006**	**0.289562**	**0.353055**	**0.433333**
Rich_All	265.549778	0.261939	0.228910	0.267384	0.320704
Delta	-99.586042	0.078067	0.060652	0.085671	0.112630
ComplEx					
Shallow_All	**171.694639**	**0.348654**	**0.300776**	**0.360821**	**0.437884**
Rich_All	263.238816	0.268759	0.235774	0.274572	0.326594
Delta	-91.544177	0.079895	0.065001	0.086249	0.111290

SPLIT	MR	MRR	H@1	H@3	H@10
TransE_NotClean					
Shallow_Split	**138.438027**	0.293748	0.249561	0.301931	0.364737
Rich_Split	176.536839	**0.320323**	**0.274590**	**0.330523**	**0.398368**
Delta	-38.098812	-0.026574	-0.025030	-0.028592	-0.033631
TransE_Clean					
Shallow_Split	**136.215533**	0.297559	0.251864	0.306459	0.372850
Rich_Split	174.791412	**0.326508**	**0.279918**	**0.338311**	**0.405980**
Delta	-38.575879	-0.028950	-0.028054	-0.031852	-0.033130
TransR					
Shallow_Split	**427.716955**	**0.094434**	**0.078577**	**0.096875**	**0.115589**
Rich_Split	434.585313	0.083940	0.070556	0.085262	0.100761
Delta	-6.868358	0.010494	0.008021	0.011613	0.014828
DistMult					
Shallow_Split	**292.566215**	**0.133801**	**0.096202**	**0.137969**	**0.200195**
Rich_Split	375.044530	0.087382	0.065789	0.086065	0.119559
Delta	-82.478315	0.046419	0.030413	0.051904	0.080636
ComplEx					
Shallow_Split	**308.120902**	**0.130338**	**0.095243**	**0.133567**	**0.191433**
Rich_Split	387.443978	0.088012	0.064216	0.087500	0.126698
Delta	-79.323076	0.042327	0.031027	0.046067	0.064735

Discussions

- KGE has no significant impact for clean versus dirty data
- Trained models evaluated with seen data outperforms standard split training
- Discernible differences in MR Metrics where the Shallow design has a better performance on KGE

Future Work

- Reproduction of research on other datasets
- Investigation of causal mechanisms of differing results
- Evaluation of KGE model weights

References

1. Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., Yakhnenko, O.: Translating embeddings for modeling multi-relational data. In: Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2. p.2787–2795. NIPS'13, Curran Associates Inc., Red Hook, NY, USA (2013)
2. Lin, Y., Liu, Z., Sun, M., Liu, Y., Zhu, X.: Learning entity and relation embeddings for knowledge graph completion. Proceedings of the AAAI Conference on Artificial Intelligence 29(1) (Feb 2015). https://doi.org/10.1609/aaai.v29i1.9491, https://ojs.aaai.org/index.php/AAAI/article/view/9491
3. Shimizu, C., Hammar, K., Hitzler, P.: Modular ontology modeling. Semantic Web 14, 459–489 (2023). https://doi.org/10.3233/SW-222886, https://doi.org/10.3233/SW-222886, 3
4. Trouillon, T., Welbl, J., Riedel, S., ´Eric Gaussier, Bouchard, G.: Complex embeddings for simple link prediction (2016)
5. Yang, B., tau Yih, W., He, X., Gao, J., Deng, L.: Embedding entities and relations for learning and inference in knowledge bases (2015)
6. Zheng, D., Song, X., Ma, C., Tan, Z., Ye, Z., Dong, J., Xiong, H., Zhang, Z., Karypis, G.: Dgl-ke: Training knowledge graph embeddings at scale. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. p. 739–748. SIGIR '20, Association for Computing Machinery, New York, NY, USA (2020)

Acknowledgement. This work was funded by the National Science Foundation under Grant 2333532; Proto-OKN Theme 3: An Education Gateway for the Proto-OKN. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We also thank Andrew Eells for providing a list of supporting publications in the field of deep learning research.