Embedding Knowledge Graphs as Languages with BERT

Anonymous EurNLP submission

Abstract

Knowledge Graphs (KGs) are graph-structured knowledge bases, where facts are represented in the form of relationships (edges) between entities (nodes). One core issue with KGs is incompleteness - missing facts. Link prediction seeks to solve this problem by embedding KG nodes and edges to predict missing links. A crucial problem with traditional link prediction is that one cannot explicitly include context into the prediction. In industry we often want to predict what someone is likely to do (e.g. eat) in a particular moment. A triplet prediction such as *(Person1, Eating, ?)* relies on the notion that *Person1* must be fully represented by a single embedding. However, if we instead have *(Person1, Where, Office, Time, Midday, Eating, ?)* then we can use this potentially unseen subgraph to make a prediction conditioned on the context. In this paper, we solve this context problem by translating KGs into languages and embed them using a BERT-based architecture. We show this method is competitive with state-of-the-art (SOTA) methods, but also contains a number of vital advantages over previous techniques such as contextualised node/edge embeddings, prediction over unseen subgraphs, natural-language-like querying, and explainability via attention. This work is motivated by previous works which combine language modelling (word2vec, Mikolov et al. 2013) with unbiased (Perozzi et al., 2014) and biased (Grover and Leskovec, 2016; Abu-El-Haija et al., 2017) walks over graphs. Current SOTA language models have improved upon word2vec significantly by using deep bidirectional architectures which produce contextualised embeddings (e.g. BERT, Devlin et al. 2018). We hypothesize that applying these architectures would increase the accuracy of embedding KGs. We embed a benchmark KG FB15k-237 (Toutanova and Chen 2015) and our [COMPANY] KG using walks to form 'sentences'. We use these sentences to train a BERT language model (without pretrained weights) to embed the nodes/edges in the graph. We combine the walks with a sentence manipulation stage (the Translator). Translators are task-specific, allowing them to be tuned to the particular KG and application. These task-specific differences vary the walk-bias (e.g. up-weight specific edges) and the structure of the sentence. We present results from our hyperparameter study on the maximum walk length and model size in Table 1, alongside scores of SOTA models. This demonstrates that our method is competitive with SOTA KG embedding models. We present examples where the extra functionality of contextualised predictions over unseen subgraphs, natural language-like querying, and increased explainability has been vital in an industry setting. We discuss enhancements of the method that could improve the performance beyond SOTA.

Model	L_{Max}	**Hits@1**	**@3**	**@10**
BERT-Base	1	0.06	0.12	0.22
BERT-Base	2	0.12	0.17	0.28
BERT-Base	5	0.14	0.20	0.3
BERT-Base	10	0.19	0.27	0.43
BERT-Base	20	0.22	0.31	0.45
BERT-Large	20	0.24	0.34	0.48
DistMult	—	0.16	0.26	0.42
R-GCN	—	0.16	0.28	0.42
ConvE	—	0.24	0.36	0.50
ComplEx-N3	—	0.27	0.40	0.57

Table 1: Hits@K scores for multiple models and hyperparameter selections. L_{Max} is the maximum length of the walk. References: DistMult (Yang et al., 2014), R-GCN (Schlichtkrull et al., 2017), ConvE (Dettmers et al., 2017), ComplEx-N3 (Lacroix et al., 2018).

References

Sami Abu-El-Haija, Bryan Perozzi, Rami Al-Rfou, and Alex Alemi. 2017. Watch your step: Learning graph embeddings through attention. *CoRR*, abs/1710.09599.

Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. 2017. Convolutional 2d knowledge graph embeddings. *CoRR*, abs/1707.01476.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805.

Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. *CoRR*, abs/1607.00653.

Timothe Lacroix, Nicolas Usunier, and Guillaume Obozinski. 2018. Canonical tensor decomposition for knowledge base completion.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 3111–3119. Curran Associates, Inc.

Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. *CoRR*, abs/1403.6652.

Michael Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. 2017. Modeling relational data with graph convolutional networks.

Kristina Toutanova and Danqi Chen. 2015. Observed versus latent features for knowledge base and text inference. In *Proceedings of the 3rd Workshop on Continuous Vector Space Models and their Compositionality*, pages 57–66, Beijing, China. Association for Computational Linguistics.

Bishan Yang, Wen tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2014. Embedding entities and relations for learning and inference in knowledge bases.