Technical Report for OGB Link Property Prediction: ogbl-wikikg2

Liang Yao[a], Jiazhen Peng[a], Qiang Liu[a], Hongyun Cai[a], Shengong Ji[a], Feng He[a], Xu Cheng[a]

[a] *Tencent Inc., Shenzhen, China*

Abstract

This technical report introduces our solution for the OGB Link Property Prediction challenge ogbl-wikikg2.

1. Introduction of the Open Graph Benchmark Challenge

The Open Graph Benchmark (OGB) is a collection of realistic, large-scale, and diverse benchmark datasets for machine learning on graphs [1]. The ogbl-wikikg2 dataset is a Knowledge Graph (KG) extracted from the Wikidata knowledge base. It contains a set of triplet edges (head, relation, tail), capturing different types of relations between entities, e.g., (Canada, citizen, Hinton). ogbl-wikikg2 contains 2,500,604 entities and 535 relation types. The task is to predict new triplet edges given the training edges. The evaluation metric follows the standard filtered metric widely used in KG. Specifically, each test triplet edge is corrupted by replacing its head or tail with randomly-sampled 1,000 negative entities (500 for head and 500 for tail), ensuring the resulting triplets do not appear in KG. The goal of the task is to rank the true head (or tail) entities higher than the negative entities, which is measured by the Mean Reciprocal Rank (MRR) metric.

2. Method

Existing methods for the challenge only use information in the provided KG, but external data such as entity descriptions, entity types, and visual information are ignored [2]. In this report, we utilize entity descriptions of Wikidata to enhance the KG representation. Specifically, we downloaded and parsed the JSON dump of all Wikibase entries for Wikidata generated on

Table 1: The MRR score on the test and validation sets. The higher, the better.

Method	Test MRR	Validation MRR
StarGraph + TripleRE	0.7201 ± 0.0011	0.7288 ± 0.0008
StarGraph + TripleRE + Text	**0.7305 ± 0.0010**	**0.7442 ± 0.0006**
InterHT+	0.7293 ± 0.0018	0.7391 ± 0.0023

November 30, 2015 from https://archive.org/details/wikidata-json-20151130. For each entity, we embed the description text with Sentence-BERT [3] (the best performing all-mpnet-base-v2 model), and then initialize the entity embeddings in existing methods with our text embeddings. We call our method X + Text, where X is the base method. In this report, we use StarGraph + TripleRE [4] as the base method.

Our method achieves better results compared with the state-of-the-art model InterHT+ [5] and StarGraph + TripleRE [4], as shown in Table 1. To the best of our knowledge, our method ranks No.1 on the leaderboard for ogbl-wikikg2 until now.

References

[1] W. Hu, M. Fey, M. Zitnik, Y. Dong, H. Ren, B. Liu, M. Catasta, J. Leskovec, Open graph benchmark: Datasets for machine learning on graphs, arXiv preprint arXiv:2005.00687 (2020).

[2] S. Ji, S. Pan, E. Cambria, P. Marttinen, S. Y. Philip, A survey on knowledge graphs: Representation, acquisition, and applications, IEEE Transactions on Neural Networks and Learning Systems 33 (2) (2021) 494–514.

[3] N. Reimers, I. Gurevych, Sentence-bert: Sentence embeddings using siamese bert-networks, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2019.
URL `https://arxiv.org/abs/1908.10084`

[4] H. Li, X. Gao, Y. Deng, Stargraph: A coarse-to-fine representation method for large-scale knowledge graph (2022). arXiv:2205.14209.

[5] B. Wang, Q. Meng, Z. Wang, D. Wu, W. Che, S. Wang, Z. Chen, C. Liu, Interht: Knowledge graph embeddings by interaction between head and tail entities, arXiv preprint arXiv:2202.04897 (2022).