Reranking Triples by Leveraging Text Descriptions for Link Prediction

Bin Zhang[1], Ximin Sun[1⋆], Mingda Wang[2], Bin Zheng[2], Bo Sun[2], and Zhenfeng Han[3]

[1] STATE GRID ELECTRONIC COMMERCE CO.,LTD. / STATE GRID FINANCIAL TECHNOLOGY GROUP
[2] STATE GRID ECOMMERCE TECHNOLOGY CO.,LTD.
[3] College of Intelligence and Computing, Tianjin University, Tianjin, China
{zhangbin, wangmingda, zhengbin,}@sgec.sgcc.com.cn
sunsemon@126.com
nsauska@163.com
zhenfenghan@tju.edu.cn
⋆ Corresponding Author

Abstract. Link prediction task intends to complete Knowledge Graphs (KGs) which are always far from complete. Textual descriptions of entities in KG provide additional information that may not be explicitly represented in the structured part of the KG. Current methods aim to learn the representation of KG and predict missing links by utilizing structured and textual information. In this poster, we propose a novel rerank method that introduces the natural language inference task to leverage textual information of entities in a different way. The experiment demonstrates that our rerank method improves the quality of link prediction.

Keywords: Link prediction · Knowledge graph · Natural language inference.

1 Introduction

Various Knowledge Graphs (KGs) such as Freebase and ConceptNet have been published to share linked data and have been crucial for many tasks. However, according to the Open World Assumption, KGs are never complete. Due to this fact, different KG representation learning (RL) models map KGs to a low dimensional vector space and predict missing facts. TransE [1] regards the relationships as translating operations between two entities on the same vector space. TransH [2] models relationships as translation on hyperplanes and entities are projected to the hyperplanes which allow entities to play different roles.

However, these translation-based RL methods only utilize the structural information of KG and ignore the rich information contained in entity descriptions. Fig. 1 presents an example of a tirple with entity descriptions sampled from Freebase. Therefore, some methods such as DKRL [3] leverage the textual descriptions of entities to enhance the representations of entities, which improves the quality of link prediction.

Motivated by [4], we reduce triple classification to NLI and propose a novel rerank method to improve the quality of link prediction by differently utilizing the textual descriptions of entities. Specifically, we train translation-based RL methods and use them to generate data used to train the NLI model. We sort the triples depend on the scores of translation-based RL methods. Then we use a linear combination of scores calculated by two types of model to rerank the triples.

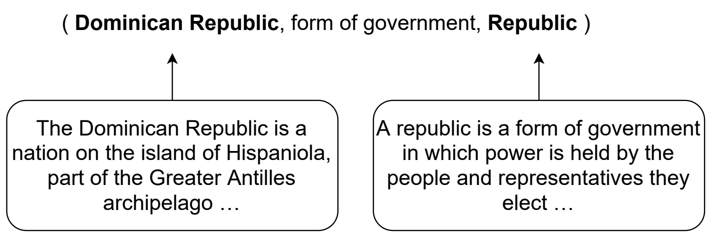

Fig. 1. Example of entity descriptions.

2 Approach

2.1 Translation-based RL Model

We only introduce the TransE model because these translation-based RL models are similar. Given entity set E, relationship set R and triple set S in which each triple (h, r, t) consist of two entities h, t and a relationship r, the task of the model is to learn embeddings of entities and relationships. TransE[1] regards the relationship r as translation from h to t. The score function is

$$d(\mathbf{h} + \mathbf{r}, \mathbf{t}) = \|\mathbf{h} + \mathbf{r} - \mathbf{t}\|_2^2 \tag{1}$$

TransE uses margin-based ranking loss:

$$L_{RL} = \sum_{(h,r,t)\in S} \sum_{(h',r,t')\in S'} \left[\gamma + d(\mathbf{h} + \mathbf{r}, \mathbf{t}) - d\left(\mathbf{h}' + \mathbf{r}, \mathbf{t}'\right)\right]_+ \tag{2}$$

where γ is a margin hyperparameter, $[n]_+$ represents the positive part of a number n, and the negative triple set S' consists of corrupted triples which replace the head or tail entity with a random entity.

2.2 Natural Language Inference

Given a dataset $\mathcal{D} = \{(s_1, s_2)_i, y_i\}_{i=1}^{N}$, Natural Language Inference (NLI) intents to learn a function $f_{\text{NLI}}(s_1, s_2) \rightarrow \{E, N, C\}$ which predict the relationship of input pair (s_1, s_2). The input (s_1, s_2) are two natural language sentences and denote the premise and hypothesis respectively. The label y is one of three classes {entailment, neutral, contradiction} which represent entailment, natural and contradiction relationships between premise and hypothesis.

2.3 Reducing Triple Classification to NLI

To leverage text description of entities through NLI, we need to generate sequence pair and relevant label from triples S. We transform three classes {entailment, neutral, contradiction} to two classes {entailment, contradiction} for the consistency of datasets. We construct NLI dataset $\mathcal{D} = \{(s_1, s_2)_i, y_i\}_{i=1}^{N}$ from triple datasets S and S'. For each triple $(h, r, t) \in S$, we construct $\{(s^1, s^2), y\}$ by mapping two entities (h, t) to relevant textual descriptions (h_{text}, t_{text}), where $s^1 = [h_{text}; t_{text}]$, $s^2 = [h; r; t]$ and $y = 1$. If triple $(h, r, t) \in S^{'}$, $y = 0$. In order to ensure the quality of generated dataset, we construct the negative triple dataset $S^{'}$ by transforming the first top negative sample of link prediciton by TransE. We use Bert [5] as our NLI model, which use cross entropy loss:

$$L_{NLI} = \frac{1}{N} \sum_{i} - [y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i)] \tag{3}$$

where p_i is the classification probability of sequence pair with Bert.

2.4 Rerank Method

First, given a triple (h, r, t), we use all entites to replace head or tail entity and calculate scores of all triples called $score_{TransE}$. Secondly, we sort triples depended on scores and get the top 10 triples. Then, we get their $score_{Bert}$ which is the classification probability computed by Bert. New scores are then computed as:

$$score^{rerank} = score_{TransE} - \lambda * score_{Bert} \tag{4}$$

and the 10 triples are reranked according to $score^{rerank}$.

3 Experiments

We verify our rerank method on two datasets, namely, FB15k and FB15k-237. To confirm that all entities in datasets have descriptions, we follow the process in DKRL [3] to remove some entities and relevant triples in both datasets. The goal of our rerank method is to improve the usability of link prediction, so we only rerank the top 10 triples. And we also follow the evaluation setting named "Filter" in TransE [1] and use Hits@1, Hits@2, Hits@3 to evaluate our method.

By Table 1, we show that our rerank method achieves higher scores of all metrics compared to original TransE and TransH models on two datasets. Besides, we even achieve 43% higher score of Hits@1 on FB15k, which demonstrates the usability of link prediction by our model. We can conclude the effectiveness of our rerank method.

Table 1. Link predcition results on two datasets.

Dataset	FB15k			FB15k-237		
Metric	Hits@1	Hits@2	Hits@3	Hits@1	Hits@2	Hits@3
TransE	0.369	0.539	0.598	0.153	0.234	0.284
TransH	0.347	0.576	0.640	0.143	0.231	0.281
TransE(rerank)	**0.458**	**0.552**	**0.605**	**0.181**	**0.243**	**0.287**
TransH(rerank)	**0.498**	**0.597**	**0.649**	**0.185**	**0.250**	**0.294**

4 Conclusion

Using text descriptions of entities has been proved to be an valid way to improve the quality of link prediction. In this poster, we propose a rerank method for link prediction which is a different way to leverage the text information of entities. Experiments demonstrate that our approach is useful and promising. In future work, we are interested in extending our rerank method for more KG Embedding models.

References

1. Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, Oksana Yakhnenko.: Translating Embeddings for Modeling Multi-relational Data. In: 27th Annual Conference on Neural Information Processing Systems 2013, pp. 2787–2795. Lake Tahoe, Nevada, United States (2013)
2. Zhen Wang, Jianwen Zhang, Jianlin Feng, Zheng Chen.: Knowledge Graph Embedding by Translating on Hyperplanes. In: 58th AAAI Conference on Artificial Intelligence, pp. 1112–1119. AAAI Press, Québec City, Québec, Canada (2014)
3. Ruobing Xie, Zhiyuan Liu, Jia Jia, Huanbo Luan, Maosong Sun.: Representation Learning of Knowledge Graphs with Entity Descriptions. In: 4th IEEE International Conference on Multimedia Big Data, pp. 1–5. IEEE, Xi'an, China (2018)
4. Sean Welleck, Jason Weston, Arthur Szlam, Kyunghyun Cho.: Dialogue Natural Language Inference. In: 57th Conference of the Association for Computational Linguistics, pp. 3731–3741. Association for Computational Linguistics, Florence, Italy (2019)

5. Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: 14th Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 4171–4186. Association for Computational Linguistics, Minneapolis, MN, USA (2019)