Hierarchy-Based Link Prediction in Knowledge Graphs

Manling Li
Institute of
ComputingTechnology,CAS
Beijing, China 100190
limanlingcs@gmail.com

Yantao Jia
Institute of
ComputingTechnology,CAS
Beijing, China 100190
jiayantao@ict.ac.cn

Yuanzhuo Wang
Institute of
ComputingTechnology,CAS
Beijing, China 100190
wangyuanzhuo@ict.ac.cn

Jingyuan Li
Institute of
ComputingTechnology,CAS
Beijing, China 100190
lijingyuan@ict.ac.cn

Xueqi Cheng
Institute of
ComputingTechnology,CAS
Beijing, China 100190
cxq@ict.ac.cn

ABSTRACT

Link prediction over a knowledge graph aims to predict the missing entity h or t for a triple(h, r, t). Existing knowledge graph embedding based predictive methods represent entities and relations in knowledge graphs as elements of a vector space, and employ the structural information for link prediction. However, knowledge graphs contain many hierarchical relations, which existing methods have pay little attention to. In this paper, we propose a hierarchy-constrained locally adaptive knowledge graph embedding based link prediction method, called hTransA, by integrating hierarchical structures into the predictive work. Experiments over two benchmark data sets demonstrate the superiority of hTransA.

Keywords

Link prediction; hierarchy; knowledge graph embedding

1. INTRODUCTION

Link prediction over a knowledge graph aims to predict the missing head entity h or tail entity t for a triple (h, r, t) in the knowledge graph. Namely, it predicts t given (h, r) or predict h given (r, t). In recent years, many knowledge graph embedding based predictive methods have been conducted for link prediction, which embed a knowledge graph into a continuous vector space and then a score function $f_r(h, t)$ for each triple (h, r, t) is learned according to the structures of the graph. Finally, a list of candidate entities is returned in the decreasing order of their scores. For example, TransA [3], the state-of-the-art predictive method, utilizes the structural information of knowledge graphs by computing the local distance of entities and the proximity of relations. However, these methods focus on the general structure of the graphs, and there are many particular structures which can be employed to promote the performance of link prediction. One typical structure is the hierarchical structure, which is a structure where entities are organized in a tree, and their relations are hierarchical relations [1]. For instance, "Barack Obama" and his two children "Sasha Obama" and "Malia Obama" compose a tree with "Barack" as the root node. Specifically, relations in a knowledge graph can be classified into two categories, hierarchical and non-hierarchical. Hierarchical relations between entities, such as "child" between "Barack Obama" and "Sasha Obama", have directions from the head entity to the tail entity. Non-hierarchical relations between entities, such as "colleague" between "Barack Obama" and "Hillary Clinton", mean that one can interchange the head entity with the tail entity, and the triple obtained is still correct. One way to distinguish hierarchical relations in a knowledge graph is to study the mapping properties of relations. Namely, we can find the *1-to-N* (e.g., father-to-children) or *N-to-1* (e.g., children-to-father) relations. Hierarchical structures are extremely common in knowledge graphs due to the ubiquitousness of hierarchical relations. For instance, WN18, a subset of the knowledge graph WordNet, has about 50% hierarchical relations. Furthermore, it is intuitive that considering the hierarchical structure can promote the performance of link prediction. For example, knowing that "Sasha" is the child of "Barack" and "child" is a hierarchical relation helps to predict the child relationship between "Barack" and "Malia".

Motivated by this intuition, we propose a hierarchy-constrained locally adaptive knowledge graph embedding based link prediction method, called hTransA, by considering the hierarchical structures. Experiments over two benchmark data sets validate the effectiveness of the proposed method.

2. THE PREDICTIVE METHOD HTRANSA

The idea of hTransA is to define a hierarchy-based entity-specific margin M_{ent} by classifying the relations of h (or t) into hierarchical and non-hierarchical ones, where M_{ent} is used to separate positive entities from negative ones for the given entity h (or t). Then the embedding of entities and relations to a vector space $\mathbb{R}^d$ for each triple (h, r, t) is learned by minimizing a loss function concerning M_{ent}. Specifically, for a given entity h and its related relation r, the set of positive entities, denoted by P_r, contains entities which have relation with h of type r, and the set of negative entities, denoted by N_r, contains entities which have relations with h of type other than r. And R_h is the set of all relations related to h and $|R_h|$ is the cardinality of R_h. More formally,

$$M_{ent} = \frac{\sum_{r \in R_h} m_r}{|R_h|},$$

where m_r is the margin between positive and negative enti-

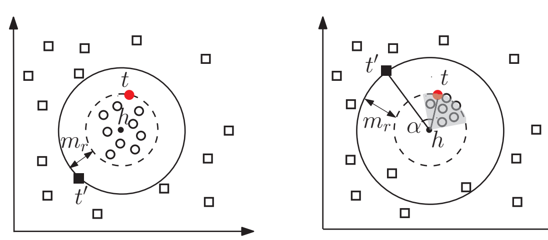
WWW'16 Companion, April 11–15, 2016, Montréal, Québec, Canada.
ACM 978-1-4503-4144-8/16/04.
http://dx.doi.org/10.1145/2872518.2889387.

ties for the given h and $r \in R_h$, which is defined according to whether r is hierarchical or not. More formally, let H be the set of hierarchical relations in a knowledge graph, and the boldface characters denote the embedding vectors of entities and relations. For instance, $\mathbf{h}$ is the embedding vector of the entity h. Then for all $t \in P_r$ and $t' \in N_r$, we define

$$m_r = \begin{cases} \min_{t,t'} \sigma(||\mathbf{h} - \mathbf{t}'|| - ||\mathbf{h} - \mathbf{t}||), & r \notin H \\ \min_{t,t'} \sigma(||\mathbf{h} - \mathbf{t}'|| - ||\mathbf{h} - \mathbf{t}||) + \lambda\phi(\alpha), & r \in H \end{cases}$$

where λ is a regularization parameter with $0 \leq \lambda \leq 1$, α is the angle between the two vectors $\mathbf{h}-\mathbf{t}$ and $\mathbf{h}-\mathbf{t}'$, and $\phi(\alpha)$ is a penalty function which is monotonically increasing with respect to α. And $\sigma(x)$ returns the absolute value of x.

The geometric meaning of m_r is illustrated in Figure 1. If r is non-hierarchical, the value $\sigma(||\mathbf{h}-\mathbf{t}'||-||\mathbf{h}-\mathbf{t}||)$ obtains the minimum when it takes the farthest positive entity t and the nearest negative entity t' with respect to h, such that $||\mathbf{h} - \mathbf{t}'||$ is small. In this case, m_r is the distance between the two concentric spheres, shown in Figure 1(a). If r is hierarchical, it is shown in [1] that the positive entities should lie close to each other since they are siblings with h as the common father. In other words, the positive entities can be enclosed in a circular sector (shaded area in Figure 1(b)), where the value $\sigma(||\mathbf{h} - \mathbf{t}'|| - ||\mathbf{h} - \mathbf{t}||) + \lambda\phi(\alpha)$ obtains the minimum when it takes the farthest positive entity t and the nearest negative entity t' with respect to h, such that both $||\mathbf{h} - \mathbf{t}'||$ and α are small. In this case, m_r is the distance between the two concentric spheres, shown in Figure 1(b). Furthermore, the setting of m_r, when r is hierarchical, is similar to the soft margin defined in SVM. The introduction of $\phi(\alpha)$ intends to penalize negative entity t' in which the angle between the two vectors $\mathbf{h} - \mathbf{t}$ and $\mathbf{h} - \mathbf{t}'$ is large.



(a) (b)

Figure 1: The illustration of m_r based on whether r is hierarchical, where circles stand for positive entities and rectangles represent negative ones in $\mathbb{R}^d$.

In order to predict t given (h, r) or predict h given (r, t), we follow the instruction in TransA [3] to adaptively choose the optimal margin M_{opt} such that $M_{opt} = \mu M_{ent} + (1-\mu)M_{rel}$, where $0 \leq \mu \leq 1$ and $M_{rel} = \min_{r_i \in R_h}(||r_i|| - ||r||)$ with $||r_i|| \geq ||r||$ is the relation-specific margin. Then we learn the representations of entities and relations by minimizing the loss function $L = \sum_{(h,r,t)\in\Delta} \sum_{(h',r,t')\in\Delta'} \max(0, f_r(h,t) + M_{opt} - f_r(h', t'))$ and rank candidate entities in terms of $f_r(h,t) = ||\mathbf{h}+\mathbf{r}-\mathbf{t}||$,where $||\cdot||$ is the L_1-norm or L_2-norm.

3. EXPERIMENTS

The experiments were carried out on two public knowledge graphs, WN18 used in [1] and FAMILY used in [2]. WN18 is a subset of the knowledge graph WordNet, which has 18 types of relations and 40,943 entities. FAMILY is an artificial hierarchical knowledge graph where entities are organized in a tree, and the number of relation types and entities are 7 and 721, respectively. Following [1], the relations are classified into *1-to-1*, *1-to-N*, *N-to-1*, *N-to-N* and the proportion of the four classes are 25.5%, 17.4%, 30.9%, 26.2% for WN18, and 0.3%, 32.0%, 19.0%, 48.7% for FAMILY. We also filter out the corrupted triples which are correct ones for evaluation, denoted as "filter" and "raw" otherwise.

The baseline methods include classical embedding methods, such as TransE [1], TransA [3], and other methods shown in Table 1. Since WN18 is also used by our baselines, we compare our results with them reported in [3]. All parameters are determined on the validation set. The penalty function adopts two monotonically increasing function $\phi(\alpha) = -log(cos\alpha)$ and $\phi(\alpha) = 1 - cos\alpha$. The optimal settings are: $\eta = 0.001$, $d = 100$, $B = 1440$, $\mu = 0.5$, $\lambda = 0.2$ for $\phi(\alpha) = -log(cos\alpha)$ and $\lambda = 0.5$ for $\phi(\alpha) = 1 - cos\alpha$, as well as taking L_1 as dissimilarity.

Table 1: Evaluation results on link prediction.

Data sets	WN18		FAMILY	
Metric	Mean Rank		Mean Rank	
	Raw	Filter	Raw	Filter
Unstructured	315	304	374	357
SE	1,011	985	362	351
SME(linear)	545	533	26	9
SME(bilinear)	526	509	29	12
TransE	263	251	30	10
TransA	165	153	23	8
hTransA($-log(cos\alpha)$)	129	117	17	6
hTransA($1 - cos\alpha$)	138	128	17	6

It can be seen from Table 1 that on both data sets, hTransA obtains the lowest mean rank, and decreases the mean rank of the state-of-the-art method, TransA, by 20% $\sim$ 30%. It is unsurprising since that hTransA employs the hierarchical structures to promote the performance of link prediction.

4. CONCLUSIONS

In this paper, we propose hTransA for link prediction in knowledge graphs, which adaptively chooses the entity-specific margin by modeling the hierarchical structures. Experiments demonstrate the effectiveness of hTransA.

5. ACKNOWLEDGMENTS

This work is supported by National Grand Fundamental Research 973 Program of China (No. 2013CB329602, 2014CB340401),National Natural Science Foundation of China (No.61572469, 61402442, 61572473, 61303244, 61402022, 61572467), Beijing nova program (No. Z121101002512063), and Beijing Natural Science Foundation (No. 4154086).

6. REFERENCES

[1] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko. Translating embeddings for modeling multi-relational data. In *NIPS*, pages 2787–2795, 2013.

[2] A. Garcıa-Durán, A. Bordes, and N. Usunier. Effective blending of two and threeway interactions for modeling multi-relational data. *Machine Learning and Knowledge Discovery in Databases*, pages 434–449, 2014.

[3] Y. Jia, Y. Wang, H. Lin, X. Jin, and X. Cheng. Locally adaptive translation for knowledge graph embedding. In *AAAI*, 2016.