[redacted]

tures 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 ObamaŽ and his two children •Sasha ObamaŽ and •Malia ObamaŽ compose a tree with •BarackŽ

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 R_h, which is de“ned 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 P_r and t N_r, we de“ne

$$m_r = \begin{cases} \min_{t,t'} (\|\mathbf{h} - \mathbf{t}\| - \|\mathbf{h} - \mathbf{t}\|), & r \,/\, H \\ \min_{t,t'} (\|\mathbf{h} - \mathbf{t}\| - \|\mathbf{h} - \mathbf{t}\|) + \ (\quad), & r \quad H \end{cases}$$

where is a regularization parameter with 0 1, is the angle between the two vectors $\mathbf{h} - \mathbf{t}$ and $\mathbf{h} - \mathbf{t}$, and ( ) is a penalty function which is monotonically increasing with respect to . And (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 ($\|\mathbf{h} - \mathbf{t}\| - \|\mathbf{h} - \mathbf{t}\|$) 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 ($\|\mathbf{h} - \mathbf{t}\| - \|\mathbf{h} - \mathbf{t}\|$) + ( ) obtains the minimum when it takes the farthest positive entity t and the nearest negative entity t with respect to h, such that both $\|\mathbf{h} - \mathbf{t}\|$ and 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 de“ned in SVM. The introduction of ( ) 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.

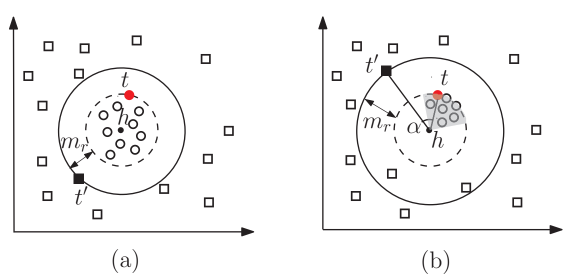

Figure 1: The illustration of m_r based on whether r is hierarchical, where circles stand for positive entities and rectangles represent negative ones in $\mathbb{R}^d$.

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 µ 1 and $M_{rel} = \min_{r_i \ R_h} (\|r_i\| - \|r\|)$ with $\|r_i\|$ $\|r\|$ is the relation-speci“c margin. Then we learn the representations of entities and relations by minimizing the loss function L = $_{(h,r,t)\,\Delta}$ $_{(h',r,t')\,\Delta'}$ 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 arti“cial 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 classi“ed 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 “lter out the corrupted triples which are correct ones for evaluation, denoted as •“lterŽ and •rawŽ 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 ( ) = −log(cos) and ( ) = 1 − cos . The optimal settings are: = 0.001, d = 100, B = 1440, µ = 0.5, = 0.2 for ( ) = −log (cos) and = 0.5 for ( ) = 1 − cos , 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))	129	117	17	6
hTransA(1 − cos)	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% 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-speci“c margin by modeling the hierarchical structures. Experiments demonstrate the e ectiveness 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. Garcša-Durán, A. Bordes, and N. Usunier. E ective 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.