Hyperbolic Hierarchy-Aware Knowledge Graph Embedding for Link Prediction

Zhe Pan and Peng Wang*

School of Computer Science and Engineering, Southeast University, Nanjing, China
{zpan, pwang}@seu.edu.cn

Abstract

Knowledge graph embedding (KGE) using low-dimensional representations to predict missing information is widely applied in knowledge completion. Existing embedding methods are mostly built on Euclidean space, which are difficult to handle hierarchical structures. Hyperbolic embedding methods have shown the promise of high fidelity and concise representation for hierarchical data. However, the logical patterns in knowledge graphs are not considered well in these methods. To address this problem, we propose a novel KGE model with extended Poincaré Ball and polar coordinate system to capture hierarchical structures. We use the tangent space and exponential transformation to initialize and map the corresponding vectors to the Poincaré Ball in hyperbolic space. To solve the boundary conditions, the boundary is stretched and zoomed by expanding the modulus length in the Poincaré Ball. We optimize our model using polar coordinate and changing operators in the extended Poincaré Ball. Experiments achieve new state-of-the-art results on part of link prediction tasks, which demonstrates the effectiveness of our method.

1 Introduction

Knowledge graphs (KGs) (Dong et al., 2014) , consisting of *(subject, relation, object)* triples, are essential for question answering, information extraction and recommendation systems. In real-world, KGs are usually incomplete (Lin et al., 2015), so predicting missing links in KGs via knowledge graph embedding (KGE) into vector spaces becomes more and more important. Hierarchical structures are common in KGs and used to manage the relations and concepts, typically forming hierarchical data. However, existing KGE methods often encounter challenges when dealing with hierarchical structures, because it is particularly difficult
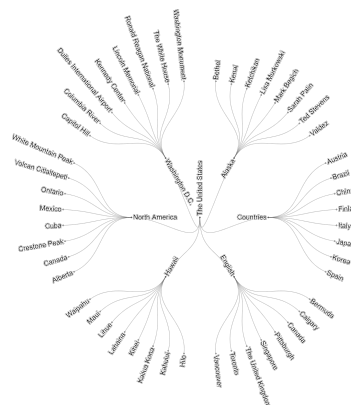
---

*Corresponding author.

Figure 1: An example of long-tail distribution and Poincaré Ball in the NELL.

for models which are built on Euclidean space to preserve hierarchical structures (Nickel and Kiela, 2017).

Recent works proposed hyperbolic representation learning represented by Poincaré Ball (Cannon et al., 1997). Figure 1 shows part of knowledge in the NELL KG (Mitchell et al., 2018), in which the entities show a long-tailed distribution with distance from the *The United States*. Namely, the hierarchical relationships between entities can be approximated as a tree structure, while the number of entities in each layer increases exponentially with depth of tree increasing. Such a knowledge structure can be well represented with the Poincaré Ball (Ungar, 2001), which is a type of hyperbolic space suitable for embedding the hierarchical structures and entities in KGs. Even most hyperbolic KGE models choose Poincaré Ball model to embed the structures, they still suffer from the problems of restricted capacity and floating-point precision when majority of points are embedded near by the boundary of Poincaré Ball due to long-tail distribution.

To tackle these challenges, this paper proposes a novel hyperbolic knowledge embedding method named HBE (**H**yperbolic extended Poincaré **B**all

2941

Embedding), which employs a extended Poincaré Ball for KG embedding and captures hierarchical structures with polar coordinate system in a hyperbolic transformation (Chami et al., 2019). First, HBE uses tangent space to initialize the entity and relation vectors in conventional Euclidean space. Then, it projects embedded entities into polar coordinate to gain hierarchical information and takes embeddings into radius and angle parts to absorb hyperbolic information. The boundary is stretched and zoomed via expanding the modulus length in the Poincaré Ball. At the same time, the addition rule in the extended Poincaré Ball are changed accordingly, so that the model can be optimized in the extended Poincaré Ball. The overall training process adopts the idea of TransE, that is, through the transformation of positive and negative examples, the measurement distance between the result of the head entity and the relationship and the tail entity is as small as possible.

In summary, the main contributions of this paper are three-fold: (1) We propose a novel hyperbolic knowledge embedding method, HBE, to apply extended Poincaré Ball for KGE and captures hierarchical structures. (2) In order to enable the model to be optimized in extended Poincaré Ball, we fine-tune the operator and fix model in polar coordinate system to embed entities and relations. (3) Experiments show that HBE outperforms state-of-the-art methods on link prediction tasks at a moderate dimension.

2 Hyperbolic Geometry

The hyperbolic space is one of the three kinds of isotropic spaces, which includes Euclidean (flat), spherical (positively curved) and hyperbolic (negatively curved) spaces (Cannon et al., 1997). Compared with the Euclidean and spherical spaces, the amount of space covered by a hyperbolic geometry increases exponentially rather than polynomially (Buser, 1992). This property allows us to capture KG structures with hyperbolic space and suits those forming hierarchies. For the hyperbolic geometry, there are several important isometric models including the hyperbolic model, Klein disk model and Poincaré Ball model. This paper chooses the extended Poincaré ball model due to its feasibility for gradient optimization (Abramowicz et al., 2002) and its infinite boundary. We hereby introduce some basic operations of hyperbolic geometry and Poincaré Ball, and then address the way to ex-

tend Poincaré Ball and modify some operators in extended one.

Specifically, a d-dimensional Poincaré Ball with a negative curvature $-c$ ($c > 0$) is defined by the manifold $(\mathcal{B}^d, g_{\boldsymbol{x}})$ (Ungar, 2001). The geodesic distance (or hyperbolic distance) $d(\boldsymbol{u}, \boldsymbol{v})$ between vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in the Poincaré Ball with $c = 1$ is given by (Ungar, 2001):

$$d(\boldsymbol{u}, \boldsymbol{v}) = \cosh^{-1}\left(1 + 2\frac{\|\boldsymbol{u} - \boldsymbol{v}\|^2}{(1 - \|\boldsymbol{u}\|^2)(1 - \|\boldsymbol{v}\|^2)}\right) \tag{1}$$

When the points move from the origin towards the ball boundary, their geodesic distance increases exponentially, offering a larger capacity of space for embedding concepts and entities (Ungar, 2001).

The vector translation in the Poincaré Ball is defined by the Möbius addition (Ungar, 2001):

$$\boldsymbol{x} \oplus_c \boldsymbol{y} = \frac{\left(1 + 2c\boldsymbol{x} \cdot \boldsymbol{y} + c\|\boldsymbol{y}\|^2\right)\boldsymbol{x} + \left(1 - d\|\boldsymbol{x}\|^2\right)\boldsymbol{y}}{1 + 2c\boldsymbol{x} \cdot \boldsymbol{y} + c^2\|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2} \tag{2}$$

where $\boldsymbol{x}, \boldsymbol{y}$ are hyperbolic vectors and c is the curvature of hyperbolic space.

Previous work defines the matrix-vector multiplication between Poincaré Balls using the exponential and logarithmic maps (Ungar, 2001). The hyperbolic vectors are first projected into the tangent space at 0 using the logarithmic map ($log_{\mathbf{0}}$) then multiplied the transformation matrix like what in the Euclidean space, and finally projected back on the manifold with the exponential map ($exp_{\mathbf{0}}$) (Nickel and Kiela, 2017). Specifically, the two projections on vector $v \in \mathcal{B}$ are defined as follows:

$$\exp_{\mathbf{0}}^c(\mathbf{v}) = \tanh(\sqrt{c}\|\mathbf{v}\|)\frac{\mathbf{v}}{\sqrt{c}\|\mathbf{v}\|} \tag{3}$$

$$\log_{\mathbf{0}}^c(\mathbf{y}) = \tanh^{-1}(\sqrt{c}\|\mathbf{y}\|)\frac{\mathbf{y}}{\sqrt{c}\|\mathbf{y}\|} \tag{4}$$

Through such projections, we can apply any Euclidean counterpart operations on hyperbolic vectors. The transformation can be done using the Möbius version of matrix-vector multiplication:

$$\mathbf{M} \otimes_c \boldsymbol{x} = \exp_{\mathbf{0}}^c\left(\mathbf{M}\log_{\mathbf{0}}^c(\boldsymbol{x})\right) \tag{5}$$

Möbius scalar multiplication can be obtained in the same way as:

$$r \otimes_c \boldsymbol{x} = \exp_{\mathbf{0}}^c\left(r\log_{\mathbf{0}}^c(\boldsymbol{x})\right)$$
$$= \left(\frac{1}{\sqrt{c}}\right)\tanh\left(r\tanh^{-1}(\sqrt{c}\|\boldsymbol{x}\|)\right)\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|} \tag{6}$$

3 Methodology

Poincaré Ball model can suit hierarchical structure of KGs, in which entities can form different hierarchies under different relations. For example, in WordNet (Fellbaum, 1998) the $chair$ is a parent node to different chair types (e.g. $folding\_chair$, $armchair$) under the relation $hypernym$. And both chair and their types are parent nodes to parts of a typical chair (e.g. $backrest$, leg) under the relation has_part. The parent node $chair$ would be embedded closer to the origin and node $backrest$ would be farther to origin.

An ideal embedding model should capture all hierarchies simultaneously. Take bilinear models (Wang et al., 2014) as an example, they can measure similarity between the subject entity embedding and an object entity embedding using the Euclidean inner product. However, a clear correspondence to the Euclidean inner product does not exist in hyperbolic space. The Euclidean inner product can be expressed as a function of Euclidean distance and norms. Noting this, Poincaré GloVe (Tifrea et al., 2019) absorbed the squared norms of the embeddings into the biases by replacing the Euclidean with the Poincaré distance $d_{\mathcal{B}}(\boldsymbol{u}, \boldsymbol{v})$ to obtain the hyperbolic version of GloVe.

Meanwhile, the capacity of Poincaré Ball model is restricted by floating-point precision when majority of points locate near by the boundary due to long-tail distribution. To tackle this problem, we utilize an extended Poincaré Ball to expand the border into infinity and adjust some operators to align with Euclidean geometry, which redefines the distance $d_{\mathcal{B}}(\boldsymbol{u}, \boldsymbol{v})$.

3.1 Extended Poincaré Ball

In Poincaré Ball, it is obvious that the whole space is symmetric along the center but the apparent Euclidean distance from the origin to any point is not equal to the hyperbolic distance (Chami et al., 2019). In order to make the apparent distance consistent with the actual hyperbolic distance, we establish a new model to ensure the distance from any point to the center is just equal to which in hyperbolic space, which is called the extended Poincaré Ball. Suppose that the polar coordinates of any point in the original coordinate system (Poincaré disk) is (r, θ), and that in the new space is $(2tanh^{-1}r, \theta)$ (Buser, 1992). As Figure 2 shown, circles in extended Poincaré Disk (in 2-dimension) are twisted.
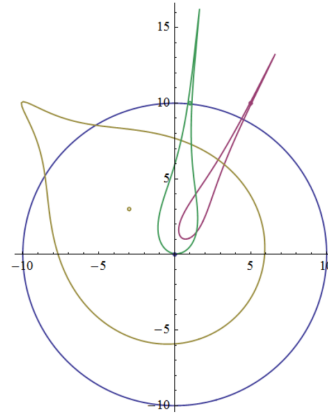

Figure 2: Circle in extended Poincaré Disk

In this way, the radius of ball space is infinite. Therefore, points near the boundary are extremely compressed in Poincaré Ball while there is no such problem in the extended one. Meanwhile, it can be proved that Hyperbolic Cosine Theorem still holds for the operators in extended Poincaré Ball: $cosh(c) = cosh(a)cosh(b) - sinh(a)sinh(b)\cos\gamma$ (a, b, c stand for the geodesic distance of triangle and γ stands for the angle between a, b). Extended Poincaré Ball and Poincaré Ball share the same distance form when calculated by cosine theorem as well (see appendix A.1 for detailed information).

Furthermore, inspired by the Hyperbolic Cosine Theorem, in which the hyperbolic distance can be composed of modulus and angle, we use polar coordinates to embed entity and relation into extended Poincaré Ball. The score function can be formed as two parts – polar radius and polar angle.

$$d_{\mathcal{B}} = \alpha d_{r\mathcal{B}} + \beta d_{\theta\mathcal{B}} \tag{7}$$

where α and β are the weights to be learned. The radius part plays an essential role in levels of entities in extended Poincaré Ball. And the angle aims to distinguish entities in the same level. The whole function shares the similar way as works in entity typing proposed by Federico (Lopez et al., 2019) which does not satisfy Cauchy inequality.

So we can formulate the polar radius with Möbius addition and multiplication as follows:

$$d_{\boldsymbol{r}\mathcal{B}} = \|2\tanh^{\text{-}1}((R \otimes_c \boldsymbol{h}) \oplus_c -(r \oplus_c \boldsymbol{t}))\|^2 \tag{8}$$

where $\boldsymbol{h}, \boldsymbol{r}, \boldsymbol{t}$ stand for hyperbolic embeddings of head entity, relation and tail entity. And R stands for relation matrix in hyperbolic space inspired by MuRP (Balazevic et al., 2019). Due to the property

of extended Poincaré Ball, one can classify the embedding levels of different entities by Euclidean Norm.

A point in polar coordinates system in high dimension can be formulated as:

$$x_{\mathcal{B}} \begin{cases} x_1 & = r\cos\theta_1 \\ x_2 & = r\sin\theta_1\cos\theta_2 \\ \dots \\ x_{d-1} & = r\sin\theta_1\dots\sin\theta_{d-2}\cos\theta_{d-1} \\ x_d & = r\sin\theta_1\dots\sin\theta_{d-2}\sin\theta_{d-1} \end{cases} \tag{9}$$

In consider of convergence and efficiency, we can simplify and formulate the polar angle as: $\Delta\theta = \pi - |\pi - |\theta - \theta'||$. Consequently, we simplify Equation 9 in TransE form (Bordes et al., 2013) as:

$$d_{\theta\mathcal{B}} = \|(\theta_h + \theta_r - \theta_t)\,mod\,2\pi\| \tag{10}$$

From another perspective, angle parts can be replaced with radius parts by Cosine Theorem in hyperbolic space. However, to better capture complex relation such as symmetry, anti-symmetry, inversion and composition, it is necessary to utilize extra angle part for downstream tasks like link predictions. On the other hand, the introduction of angle part can simulate the rotation in RotatE (Sun et al., 2019). Theoretically, any algebraic system hold the fundamental properties of congruence can be used as angle part in HBE when embedding complex relations. Take angles as an example, suppose that a relation $\theta_r \in [0, 2\pi)$ is close to π, then a symmetric relation can be formed as $(\theta_h + \theta_r + \theta_r)\,mod\,2\pi = \theta_h\,mod\,2\pi$ with arbitrary θ_h and $\neq$ for asymmetric relations.

3.2 Optimization

Since the Poincaré Ball has a Riemannian manifold structure, we optimize radius parameters with stochastic Riemannian optimization methods such as RSGD or RSVRG (Bonnabel, 2013). Let ∇E denote the Euclidean gradient of $L(P)$. Using RSGD, the Riemannian gradient can be computed as $\nabla_R = \frac{(1-\|P_t\|^2)^2}{4}\nabla_E$. In summary, the full update for a single embedding is calculated by:

$$\boldsymbol{P}_{t+1} = \boldsymbol{P}_t - \eta_t\frac{\left(1 - \|\boldsymbol{P}_t\|^2\right)^2}{4}\nabla_E \tag{11}$$

where η denotes the learning rate.

According to the isometric projection of Poincaré Ball, the angle part can be optimized by Euclidean optimization methods such as SGD or Adam (Kingma and Ba, 2015).

To train the model, we use the negative sampling loss functions with self-adversarial training (Sun et al., 2019).

$$s = -\log\left(\sigma\left(\lambda - d_{\mathcal{B}}(h, r, t)\right)\right)$$
$$-\sum_{i=1}^{n} p\left(h_i', r, t_i'\right)\log\left(\sigma\left(d_{\mathcal{B}}\left(h_i', r, t_i'\right) - \lambda\right)\right) \tag{12}$$

where λ is margin. For negative samples,

$$p\left(h_j', r, t_j' \mid \{(h_k, r_k, t_k)\}\right) = \frac{e^{yf\left(h_j', t_j\right)}}{\sum_{k=1}^{size} e^{yf\left(h_k', t_k'\right)}} \tag{13}$$

where p is the probability distribution of sampling negative triples, and α is the temperature of sampling.

4 Experiments

To evaluate our approach, we choose the widely used KG datasets WN18RR (Dettmers et al., 2018) and FB15K-237 (Bordes et al., 2013). WN18RR is a subset of WordNet, a hierarchical collection of relations between words, created from WN18 by removing the inverse of many relations from validation and test sets to make the dataset more challenging, containing 40,943 entities and 11 relations. FB15k-237 is a subset of Freebase, a collection of real world facts, created from FB15k in the same way as WN18RR. FB15k-237 contains 14,541 entities and 237 relations. The statistics of datasets are shown in Table 1. Noteworthily, the lower the metric $\mathcal{E}_G$ is, the more tree-like the KG is ($\mathcal{E}_G$ is the mean of the estimated curvatures of the sampled triangles. See (Chami et al., 2020) for more details).

We evaluate HBE on the task of KG link prediction, which is critical for practical applications. We use the scoring function to rank the correct tail or head entity against all possible entities for link prediction tasks in KGs. The evaluation metrics are: (1) mean reciprocal rank (MRR), which measures the mean of inverse ranks assigned to correct entities; and (2) hits at K (H@K, $K \in 1, 3, 10$), which measures the proportion of correct triples among the top-K predicted triples.

Table 2 and table 3 shows the performance of HBE and previous models. Our model outperforms more complex models such as DistMult, ConvE on all datasets, and beats the model RotatE (Sun et al.,

2944

Dataset	Ent.	Train	Valid.	Test	$\mathcal{E}_G$	Ave. deg.
WN18RR	40,493	86,835	3,034	3,134	-2.54	2.19
FB15k-237	14,541	272,115	17,535	20,466	-0.65	18.71

Table 1: Statistics of WN18RR and FB15k-237.

	WN18RR			
	MRR	H@1	H@3	H@10
TransE	.226	-	-	.501
DisMult	.437	.397	.444	.490
ConvE	.436	.401	.448	.527
HyperKG	.411	-	-	.500
MuRP	.481	.440	.495	.566
RotatE	.476	.428	.492	**.571**
HBE	**.488**	**.448**	**.502**	.570

Table 2: WN18RR results for link predictions.

	FB15k-237			
	MRR	H@1	H@3	H@10
TransE	.294	-	-	.465
DisMult	.241	.155	.263	.419
ConvE	.325	.237	.356	.501
HyperKG	.280	-	-	.452
MuRP	.323	.235	.353	.501
RotatE	**.338**	**.241**	**.375**	.533
HBE	.336	.239	.372	**.534**

Table 3: WN18RR results for link predictions.

	WN18RR			
	MRR	H@1	H@3	H@10
HBE	.488	.448	.502	.570
HBE-polar	.160	.013	.262	.448
HBE-dis	.377	.304	.417	.515
HBE-raw	.479	.429	.494	.564

Table 4: WN18RR ablation results for link predictions.

which may be related to the design of complex number and rotation. Figure 3(b) shows that the points in the Poincaré ball are obviously more sensitive to hierarchy, which is to say there are sparse in the middle and dense near by boundaries. Meanwhile, it is specific that most of the points are concentrated near the boundary, which makes the model prone to the problems of poor convergence and shortage of floating-point precision. HBE in Figure 3(c) utilizes extend boundary and releases the problem of dense distribution.

Finally, according to Table 4, HBE-polar is the polar version in Euclidean space. HBE-dis shows the result of radius part only, HBE-raw is the model in Poincaré Ball without extension of boundary. Detail information about weights of polar coordinate, analysis on hierarchical embeddings, and the relation case study is addressed in A.2, A.3, and A.4, respectively.

5 Conclusion

We introduce a novel translational method for embedding hierarchical KGs in extended Poincaré ball of hyperbolic geometry. Our model learns hierarchy-specific parameters with polar coordinate by Möbius multiplication and Möbius addition. We show that HBE outperforms on the link prediction task on some hierarchical KG dataset.

2019) on WN18RR datasets, which demonstrates the promising potential of hyperbolic space and polar coordinate. Nevertheless, the results in FB15k-237 shows that HBE has similar preformance with RotatE . The reason may lies in special structure of FB15k as many points with low level of hierarchy and great degree in KGs, which confuse the radius part of HBE. Because low-level points with low level trends to be embedded near by border in raw Poincaré ball while great-degree points not.

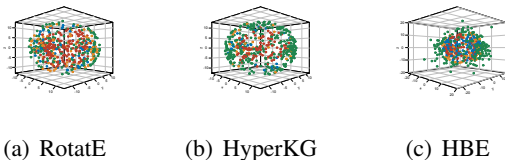(a) RotatE   (b) HyperKG   (c) HBE

Figure 3: Visualizations of models.

As shown in the Figure 3(a), the hierarchy in RotatE is not distinguished, and the overall distribution is more uniform after dimension reduction,

References

M. A. Abramowicz, I. Bengtsson, V. Karas, and et al. 2002. Poincaré ball embeddings of the optical geometry. *Classical and Quantum Gravity*, 19(15):3963.

Ivana Balazevic, Carl Allen, and Timothy M. Hospedales. 2019. Multi-relational poincare graph embeddings. In *Proceedings of 33rd Conference on Neural Information Processing Systems*, pages 4463–4473.

Silvere Bonnabel. 2013. Stochastic gradient descent on riemannian manifolds. *IEEE Transactions on Automatic Control*, 58(9):2217–2229.

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, and et al. 2013. Translating embeddings for modeling multi-relational data. In *Proceedings of 27th Conference on Neural Information Processing Systems*, pages 2787–2795.

Peter Buser. 1992. *Geometry and Spectra of Compact Riemann Surfaces*. Springer Science and Business Media, MA Birkhäuser Boston.

James W Cannon, William J Floyd, Richard Kenyon, and et al. 1997. Hyperbolic geometry. *Flavors of geometry*, 31:59–115.

Ines Chami, Adva Wolf, Da-Cheng Juan, Frederic Sala, Sujith Ravi, and Christopher Christopher Re. 2020. Low-dimensional hyperbolic knowledge graph embeddings. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL*, pages 6901–6914.

Ines Chami, Zhitao Ying, Christopher Re, and Jure Leskovec. 2019. Hyperbolic graph convolutional neural networks. In *Proceedings of 33rd Conference on Neural Information Processing Systems*, pages 4869–4880.

Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and et al. 2018. Convolutional 2d knowledge graph embeddings. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*, pages 1811–1818.

Xin Dong, Evgeniy Gabrilovich, Geremy Heitz, and et al. 2014. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 601–610.

Christiane Fellbaum. 1998. Wordnet: An electronic lexical database. *Language*, 76(3):706.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representations, ICLR*.

Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence*, pages 2181–2187.

Federico Lopez, Benjamin Heinzerling, and Michael Strube. 2019. Fine-grained entity typing in hyperbolic space. In *Proceedings of the 4th Workshop on Representation Learning for NLP*, pages 169–180.

Tom Mitchell, William Cohen, Estevam Hruschka, and et al. 2018. Never-ending learning. *Communications of the ACM*, 61(5):103–115.

Maximilian Nickel and Douwe Kiela. 2017. Poincaré embeddings for learning hierarchical representations. In *Proceedings of 31st Conference on Neural Information Processing Systems*, pages 6341–6350.

Zhiqing Sun, ZhiHong Deng, JianYun Nie, and et al. 2019. Rotate: Knowledge graph embedding by relational rotation in complex space. In *Proceedings of 7th International Conference on Learning Representations*.

Alexandru Tifrea, Gary Bécigneul, and Octavian-Eugen Ganea. 2019. Poincaré glove: Hyperbolic word embeddings. In *Proceedings of the 7th International Conference on Learning Representations*.

Abraham A Ungar. 2001. Hyperbolic trigonometry and its application in the poincare ball model of hyperbolic geometry. *Computers and Mathematics with Applications*, 41(1):135–147.

Zhen Wang, Jianwen Zhang, Jianlin Feng, and et al. 2014. Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence*, pages 1112–1119.

A Appendix

A.1 Hyperbolic Distance in Extended Poincaré Ball

Theorem A.1. *Given $P, Q : (r, \theta), \quad (r', \theta')$ are points in the polar coordinate system of extended Poincaré Ball, then the hyperbolic distance is x:*

$$\cosh x = \cosh(r)\cosh(r') - \sinh(r)\sinh(r')\cos(\Delta\theta) \tag{14}$$

where $\Delta\theta = \pi - |\pi - |\theta - \theta'||$.

Proof: Suppose a hyperbolic triangle with P, Q, O (O is the origin) which forms a 2-dimension plane from Poincaré Ball into Poincaré Disk. According to the distance formulation in Poincaré Disk:

$$x = \log\frac{1+s}{1-s}, s = \frac{|w-w'|}{1-w\bar{w}'} \tag{15}$$

Suppose the polar coordinate of points P and Q in the raw Poincaré disk:

$$(R, \theta), (R', \theta') \tag{16}$$

Then:

$$s = \sqrt{\frac{R^2 + R'^2 - 2RR'\cos\Delta\theta}{1 + R^2R^2 - 2RR'\cos\Delta\theta}} \quad (17)$$

And,

$$\cosh(x) = \frac{\exp(x) + \exp(-x)}{2}$$
$$= \frac{1}{2}\left(\frac{1+s}{1-s} + \frac{1-s}{1+s}\right) = \frac{1+s^2}{1-s^2} \quad (18)$$

Take s into Equation 18:

$$\cosh(x) = \frac{1 + R^2 + R'^2 + R^2R'^2 - 4RR'\cos\theta}{(1-R^2)(1-R'^2)} \quad (19)$$

Take this transformation

$$r = \tanh(R/2), r' = \tanh(R'/2) \quad (20)$$

into Extended Poincaré Disk (Ball):

$$\cosh(x) = \cosh(r)\cosh(r') - \sinh(r)\sinh(r')\cos(\Delta\theta)$$
$$= \frac{1+R^2}{1-R^2} \cdot \frac{1+R^2}{1-R'^2} - 4\frac{RR'\cos\Delta\theta}{(1-R^2)(1-R^2)} \quad (21)$$

Theorem is proved.

A.2 Weights of Polar Coordinate

To analyze influence of curvature and the weight of radius and angle part, we collect ratio and curvature with different dimensions.

Figure 4(a) shows the weight ratio of radius and angle in different dimension. And x-axis stands for dimension and y-axis is the ratio of $\frac{\beta}{\alpha+\beta}$, which stands for weight of radius part and dimension. With dimension grows by, the share of radius part descends rapidly and the curvature of extended Poincare Ball tends to be 0 (more flat and close to Euclidean Space) in Figure 4(b). And the performance of KEEN in high dimensions will close to RotatE.

A.3 Analysis on Hierarchical Embeddings

For certain relations, we sample some triplets with hierarchical relation $\_hypernym$ from WN18RR and show their head and tail entity embeddings' radius parts of distribution.

As we expected in Figure 5(a), the tail level is higher than head level. And relation $/film/film/genre$ in FB15k-237 has a similar situation which can be leveled and distinguished by radius part in Figure 5(b). To compare with, we
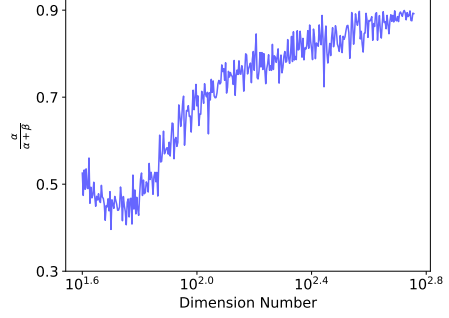
(a) ratio.
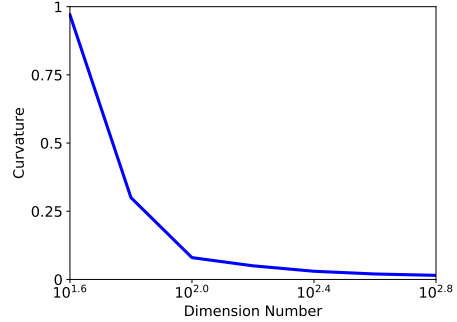
(b) curvature.

Figure 4: Ratio and Curvature. The ratio $\frac{\beta}{\alpha+\beta}$ stands for weight of angle part in sum of total weights. With dimension increases, angle part accounts for more part of total weights.

(a) hypernym.   (b) genre.

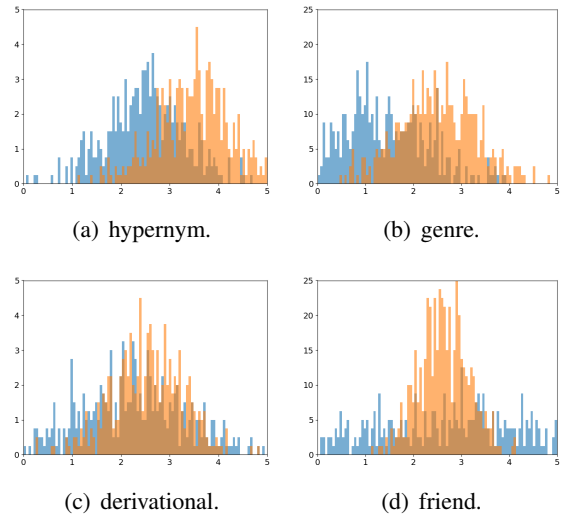(c) derivational.   (d) friend.

Figure 5: Distribution of Entities. For a certain relation, we sample triplets and gather the radius part distribution of head and tail entities. And blue histogram stands for tail while orange one for head.

choose relation $\_derivationally\_related\_form$ which is symmetric as an example in Figure 5(c). And radius part distribution of head and tail entities with $\_derivationally\_related\_form$

2947

relations	WN18RR				
	$\mathcal{E}_G$	MRR	H@1	H@3	H@10
also_see	-2.09	.627	.589	.643	696
has_part	-1.43	.192	.108	.230	363
hypernym	-2.46	.190	.135	.209	.306
verb_group	-0.50	.974	.974	.974	974
synset_domain	-0.69	.370	.294	.412	509
member_meronym	-2.90	.287	.204	.320	.437

Table 5: Link prediction results for some relations on WN18RR.

is much more irregular than hierarchical ones in $\_hypernym$. In other words, angle parts are in a leading position when representing this kinds of symmetric relations. Furthermore, this happens on relation $/celebrities/friendship/friend$ in FB15k-237.

Furthermore, we can calculate the KL divergence of several head-tail distribution mentioned above for further analysis. On the other hand, it is apparent that $\_hypernym$ has more hierarchical structure than $/film/film/genre$ when comparing 0.084 with 0.078 in radius part, which can be inferred in Table 5 by $\mathcal{E}_G$ in Appendix A.4.

A.4   Relation Case Study

In order to further analyze the performance on different relations, we conducted the relation case study. Table 5 illustrates the link prediction performances of 6 relations, in which a relation is supposed to be more hierarchical with lower $\mathcal{E}_G$.

There are obvious different performances between semantic hierarchical relations (such as $hypernym$ or $haspart$) and semantic non-hierarchical relations (such as $verb\_group$ or $also\_see$). As is mentioned above, $\mathcal{E}_G$ is used for measuring tree-like level of a KG. And relation $also_see$ with smaller $\mathcal{E}_G$ and semantic symmetric property could be well embedded by HBE. It is worth noting that results of $verb\_group$ are surprisingly good and it may not be so reliable due to its small amount in test set, which needs further analysis. From another standpoint, the hierarchical level of a KG may not be well defined by $\mathcal{E}_G$.