Hyperbolic Recommender Systems

Tran Dang Quang Vinh[1], Yi Tay[1], Shuai Zhang[2], Gao Cong[1], Xiao-Li Li[3]

[1]Nanyang Technological University, Singapore
[2]University of New South Wales, Australia
[3]Institute for Infocomm Research(I2R), A*STAR, Singapore
trandang001@e.ntu.edu.sg, ytay017@e.ntu.edu.sg, shuai.zhang@student.unsw.edu.au,
gaocong@ntu.edu.sg, xlli@i2r.a-star.edu.sg

Abstract

Many well-established recommender systems are based on representation learning in Euclidean space. In these models, matching functions such as the Euclidean distance or inner product are typically used for computing similarity scores between user and item embeddings. This paper investigates the notion of learning user and item representations in Hyperbolic space. In this paper, we argue that Hyperbolic space is more suitable for learning user-item embeddings in the recommendation domain. Unlike Euclidean spaces, Hyperbolic spaces are intrinsically equipped to handle hierarchical structure, encouraged by its property of exponentially increasing distances away from origin. We propose HyperBPR (Hyperbolic Bayesian Personalized Ranking), a conceptually simple but highly effective model for the task at hand. Our proposed HyperBPR not only outperforms their Euclidean counterparts, but also achieves state-of-the-art performance on multiple benchmark datasets, demonstrating the effectiveness of personalized recommendation in Hyperbolic space.

Introduction

The pervasive impact that recommender systems have on the web is evident. This widespread ubiquity is understandable, given the growth of data in recent years whereby users are commonly plagued with over-choice. After all, interaction data (clicks, purchases, etc.) lives at the heart of many web applications such as content streaming sites, e-commerce and so on. To this end, recommender systems serve as not only a great mitigation strategy, but also create an overall better user experience on the web. This paper is concerned with the task of personalized (or collaborative) ranking, in which a ranked list of prospective candidate items is served to each user.

Learning representations of user and item pairs forms the crux of the personalized ranking problem. Across the literature, a diverse plethora of machine learning models have been proposed (Rendle et al. 2009; Rendle 2010; Mnih and Salakhutdinov 2008; He et al. 2017). A variety of matching functions have been traditionally adopted, such as the inner product (Bayesian Personalized Ranking) (Rendle et al. 2009), Euclidean distance (Collaborative Metric

Learning (Hsieh et al. 2017) and/or neural networks (He et al. 2017). Notably, a common denominator is that all of these models operate in Euclidean space which may be sub-optimal for interaction data.

This paper investigates the notion of learning user-item representations in Hyperbolic space $\mathbb{H}$ in which the distance increases exponentially relative to the origin. Hyperbolic representation learning have recently demonstrated great promise across a diverse range of applications such as learning entity hierarchies (Nickel and Kiela 2017) and/or natural language processing. (Tay, Tuan, and Hui 2018; Dhingra et al. 2018). In a similar vein, we hypothesize that a non-conformal space, provides a more suitable inductive bias for interaction data that is commonplace in recommender systems. Intuitively, Hyperbolic spaces induce a tree-structured (hierarchical) embedding space, which is inherently more suitable for modeling hierarchical structure. We show that a conceptually simple Hyperbolic adaptation of the popular Bayesian Personalized Ranking (BPR) algorithm is capable of not only achieving very competitive results, but also outperforms more complex neural models on multiple personalized ranking benchmarks.

It is intuitive that hierarchical structures exists as one of the predominant flavors in recommender systems. Naturally, items generally exhibit hierarchical structure (i.e., movies, products tend to follow a product hierarchy). Similarly, implicit user interactions may also inhibit hierarchical qualities due to intrinsic power-law nature of the problem domain. The notion of exploiting hierarchical structure has been established in many existing works in the literature (Wang et al. 2015; Zhao et al. 2017; Wang et al. 2018). However, this work is the first work to explore a hierarchical inductive bias for training machine learning models for recommender systems. Our experiments show that our proposed model, trained with this inductive bias, leads to considerable improvements in ranking performance of the model.

The usage of Hyperbolic distance qualifies our model as a metric learning approach, albeit in Hyperbolic space as opposed to Euclidean space. Metric learning models such as the Collaborative Metric Learning (Hsieh et al. 2017) have reasonably demonstrated empirical success. However, it has been argued to introduce instability according to (Tay, Anh Tuan, and Hui 2018) due to its inability to fit a large number of interactions with a fixed set of parameters. To this

end, we argue that the Hyperbolic space can be interpreted to be seemingly larger than Euclidean spaces in the sense that the norm (distance from the origin) captures some information. Due to the increasing distance from the origin, this causes the embedding space to have a greater extent of representation capability as opposed to Euclidean spaces. This reinforces the key intuition of modeling user-item pairs in Hyperbolic space, while maintaining the simplicity and effectiveness of the CML model.

Our Contributions   All in all, the key contributions of this work are summarized as follows:

- We investigate the notion of training recommender systems in Hyperbolic space as opposed to Euclidean space. We propose *Hyperbolic Bayesian Personalized Ranking* (HyperBPR), a strong competitive model for one-class collaborative filtering (i.e., personalized ranking). To the best of our knowledge, this is the first work that explores the use of Hyperbolic space for the recommender systems domain.

- We conduct extensive experiments on eight benchmark datasets. Our proposed HyperBPR demonstrates the effectiveness of the Hyperbolic space, outperforming not only it's Euclidean counterparts but also a suite of competitive baselines. Notably, HyperBPR outperforms the state-of-the-art neural collaborative filtering (NCF) and collaborative metric learning (CML) models on all benchmarks. We achieve a reasonable performance gain over competitors, pulling ahead by up to 5.7% performance in terms of standard ranking metrics.

- We conduct extensive qualitative and visualization experiments, delving into the inner workings of our proposed HyperBPR.

Related Work

Across the rich history of recommender systems research, a myriad of machine learning models have been proposed (Rendle 2010; Mnih and Salakhutdinov 2008; Rendle et al. 2009; He et al. 2016; Koren 2008; He et al. 2017; Hsieh et al. 2017). Traditionally, many works are mainly focused on factorizing the interaction matrix, i.e, Matrix Factorization (Mnih and Salakhutdinov 2008; Koren, Bell, and Volinsky 2009), learning latent factors for user and items based on their preferences. Naturally, the formulation of matrix factorization is equivalent combining the user-item embeddings using the inner product (He et al. 2017). To this end,(Hsieh et al. 2017) argued that this formulation lack expressiveness due to its violation of the triangle inequality. As a result, the authors proposed Collaborative Metric Learning (CML), a strong recommendation baseline based on Euclidean distance. Notably, many recent works have moved into neural models (He et al. 2017; Zhang et al. 2018), in which stacked nonlinear transformations have been used to approximate the interaction function.

Our work is concerned with recommendation with implicit feedback (i.e., clicks, likes of binary nature). In this task, the Bayesian Personalized Ranking (BPR) model (Rendle et al. 2009) remains a strong competitive baseline. BPR has seen widespread success across a myriad of domains and applications (Dave et al. 2018b; Zhang et al. 2016; He and McAuley 2016b; Dave et al. 2018a). Our work trains the BPR model in Hyperbolic space, by incorporating the Hyperbolic distance as the similarity function between user and item.

Our work is inspired by recent advances in Hyperbolic representation learning (Nickel and Kiela 2017; Cho et al. 2018; Nickel and Kiela 2018; Ganea, Bécigneul, and Hofmann 2018; Sala et al. 2018; Davidson et al. 2018). For instance, (Tay, Tuan, and Hui 2018) proposed training a question answering system in Hyperbolic space. (Dhingra et al. 2018) proposed learning word embeddings using a Hyperbolic neural network. (Gülçehre et al. 2018) proposed an Hyperbolic variation of self-attention and the transformer network, and applies it to tasks such as visual question answering and neural machine translation. While the advantages of Hyperbolic space seems eminent in the wide variety of application domains, there is no work that investigates this embedding space within the context of recommender systems and implicit interaction data. This constitutes the key novelty of our work. A detailed primer on Hyperbolic spaces is given in the technical exposition of the paper.

Hyperbolic Recommender Systems

This section outlines the overall architecture of our proposed model. The key motivation behind our architecture is to embed the two user-item pairs into the hyperbolic space and then maximize the margin between the scores of the positive user-item pair and the negative user-item pair through pairwise learning. Figure 2 depicts the overall model architecture.

Input Encoding

Our proposed model takes a user (denoted as $\mathbf{u}_i$), a positive (observed) item (denoted as $\mathbf{v}_j$) and a negative (unobserved) item (denoted as $\mathbf{v}_k$) as an input. Each user and item are represented as one-hot vectors which maps onto a dense low-dimensional vector by indexing onto an user/item embedding matrix. Our model then leverages Bayesian Personalized Ranking (BPR) to optimize the pairwise ranking between the positive and negative item.

Hyperbolic Geometry & Poincaré Embeddings

The hyperbolic space is uniquely defined as a complete and simply connected Riemannian manifold with constant negative curvature (Krioukov et al. 2010) as visualized in Figure 1[1]. In fact, there are only three types of the Riemannian manifolds of constant curvature, which are Euclidean geometry (constant vanishing sectional curvature), spherical geometry (constant positive sectional curvature) and hyperbolic geometry (constant negative sectional curvature). Some properties of the three geometries can be found at Table 1. In this paper, we pay attention to the Euclidean

---

[1]Images were taken at `https://en.wikipedia.org/wiki/Hyperbolic_geometry`

Property	Euclidean	Spherical	Hyperbolic
Curvature K	0	>0	<0
A line	no finite length; unbounded	finite length; unbounded	finite length
Two distinct lines	not enclose a finite area	enclose a finite area	not enclose a finite area
Parallel lines	1	0	∞
Sum of triangle angles	π	$>\pi$	$<\pi$
Circle length	$2\pi r$	$2\pi\sin(\epsilon r)$	$2\pi\sinh(\epsilon r)$
Disk area	πr^2	$2\pi(1-\cos(\epsilon r))$	$2\pi(\cosh(\epsilon r)-1)$

Table 1: Some properties of Euclidean, spherical and hyperbolic geometry; in which r is the radius and $\epsilon = \sqrt{|K|}$.

(a) M. C. Eschers Circle Limit III, 1959

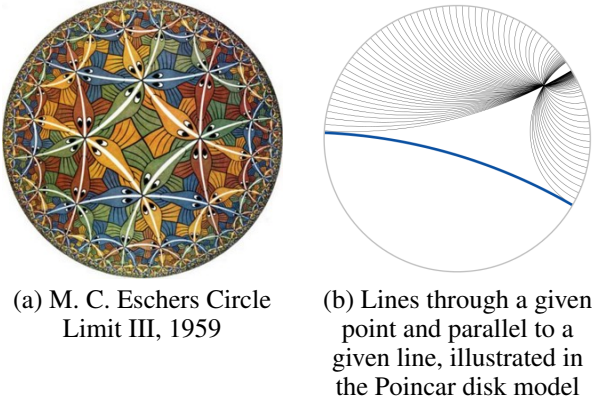(b) Lines through a given point and parallel to a given line, illustrated in the Poincar disk model

Figure 1: Visualizations of Hyperbolic space.

spaces and hyperbolic spaces due to the key difference in their space expansion. Indeed, hyperbolic spaces expand faster (exponentially) than Euclidean spaces (polynomially). Specifically, for instance, in the two-dimensional hyperbolic space $\mathbb{H}_\epsilon^2$ of constant curvature $K = -\epsilon^2 < 0$, $\epsilon > 0$ with the hyperbolic radius of r, we have:

$$L(r) = 2\pi\sinh(\epsilon r), \tag{1}$$
$$A(r) = 2\pi(\cosh(\epsilon r) - 1), \tag{2}$$

in which $L(r)$ is the length of the circle and $A(r)$ is the area of the disk. Hence, both Eqn. (1) and (2) illustrate the exponentially growing/expansion of the hyperbolic space $\mathbb{H}_\epsilon^2$ with respect to the radius r.

Although hyperbolic space cannot be isometrically embedded into Euclidean space, there exists multiple models of hyperbolic geometry that can be formulated as a subset of Euclidean space and are very insightful to work with depends on different tasks. Amongst these models, we prefer the Poincaré ball model as proposed by (Nickel and Kiela 2017) due to its conformality (i.e., angles are preserved between hyperbolic and Euclidean space) and convenient parameterization.

The Poincaré ball model is the Riemannian manifold $\mathcal{P}^n = (\mathcal{B}^n, g_p)$, in which $\mathcal{B}^n = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| < 1\}$ is the *open* n-dimensional unit ball that equipped with the metric as:

$$g_p(\mathbf{x}) = \left(\frac{2}{1 - \|\mathbf{x}\|^2}\right)^2 g_e, \tag{3}$$

where $\mathbf{x} \in \mathcal{B}^n$; and g_e is the Euclidean metric tensor with components $\mathbf{I}_n$ of $\mathbb{R}^n$.

The distance between two points on $\mathcal{P}$ is given by:

$$d_p(\mathbf{x}, \mathbf{y}) = \cosh^{-1}\left(1 + 2\frac{\|\mathbf{x} - \mathbf{y}\|^2}{(1 - \|\mathbf{x}\|^2)(1 - \|\mathbf{y}\|^2)}\right). \tag{4}$$

We adopt the hyperbolic distance function to model the relationships between users and items. Specifically, the hyperbolic distance $d_p(\mathbf{u}, \mathbf{v})$ between user u and item v is calculated based on Eqn. (4). On a side note, it is worth mentioning that $d_p(\mathbf{u}, \mathbf{v})$ helps to discover the latent hierarchies automatically as the distance within the Poincaré ball changes smoothly with respect to the norm of u and v. Notably, the distance between points grow exponentially as the norm of the vectors approaches 1. Geometrically, if we place the root node of a tree at the origin of $\mathcal{B}^n$, the children nodes thus spread out exponentially with their distance to the root towards the boundary of the ball due to the above mentioned property.

Learning Hyperbolic Representations of User-Item Pairs

Inspired by (Gülçehre et al. 2018), the hyperbolic distance is then passed into an extra layer called hyperbolic matching layer for matching pairs of users and items. Given a user u and an item v that are both lying in $\mathcal{B}^n$, we take:

$$\alpha(\mathbf{u}, \mathbf{v}) = f(d_p(\mathbf{u}, \mathbf{v})), \tag{5}$$

where $f(\cdot)$ is simply preferred as a linear function $f(x) = \beta x + c$ with $\beta \in \mathbb{R}$ and $c \in \mathbb{R}$ are scalar parameters and learned along with the network.

Optimization and Learning

This section illustrates the optimization and learning process of HyperBPR.

BPR Triplet Loss. HyperBPR leverages BPR pairwise learning to minimize the pairwise ranking loss between the positive and negative items. The objective function is defined as follows:

$$\arg\min_\Theta \sum_{(i,j,k)\in\mathcal{D}} -\ln\sigma\{\alpha(\mathbf{u}_i, \mathbf{v}_j) - \alpha(\mathbf{u}_i, \mathbf{v}_k)\} + \lambda(\|\Theta\|^2), \tag{6}$$
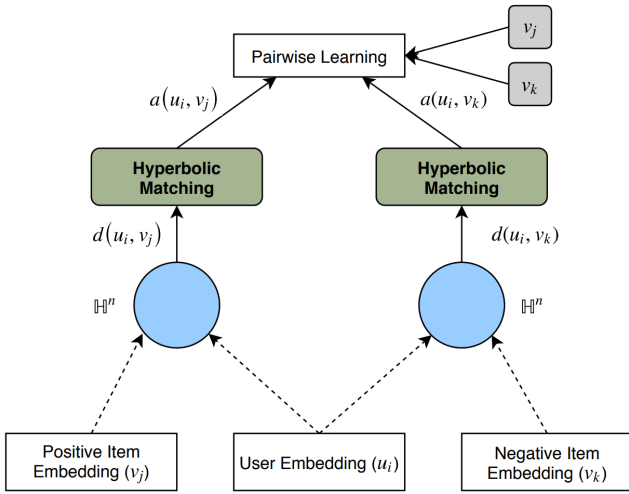
Figure 2: Illustration of our proposed HyperBRP architecture.

where (i, j, k) is the triplet that belongs to the set $\mathcal{D}$ that contains all pairs of positive and negative items for each user; σ is the logistic sigmoid function; Θ represents the model parameters; and λ is the regularization parameter.

Gradient Conversion. The parameters of our model are learned by using RSGD (Bonnabel 2013). As similar to (Nickel and Kiela 2017), the parameter updates have the form:

$$\boldsymbol{\theta}_{t+1} = \mathfrak{R}_{\theta_t}(-\eta_t \nabla_R \mathcal{L}(\boldsymbol{\theta}_t)), \qquad (7)$$

where $\mathfrak{R}_{\theta_t}$ denotes a retraction onto $\mathcal{B}$ at $\boldsymbol{\theta}$; η_t is the learning rate at time t; and $\nabla_R \mathcal{L}(\boldsymbol{\theta}_t)$ is the Riemannian gradient with respect to $\boldsymbol{\theta}$.

The Riemannian gradient ∇_R is then calculated from the Euclidean gradient by rescaling ∇_E with the inverse of the Poincaré ball metric tensor:

$$\nabla_R = \frac{(1 - \|\boldsymbol{\theta}_t\|^2)^2}{4} \nabla_E. \qquad (8)$$

The details of gradient conversion can be referred to (Nickel and Kiela 2017; Tay, Tuan, and Hui 2018).

Experiments

Experimental Setup

In this section, we introduce the overall experimental setup.

Datasets For our experimental evaluation, we adopt eight datasets from Amazon datasets (He and McAuley 2016a). The selection is based on promoting diversity based on dataset size and domain, in which we ensure the inclusion of both large/small datasets across various domains. The datasets can be obtained at `http://jmcauley.ucsd.edu/data/amazon/` with their domain names truncated

Dataset	Interactions	# Users	# Items	% Density
Clothing	235,906	7,917	171,760	1.74
Sports	113,119	3,740	54,744	5.53
Cell Phones	32,885	1,141	18,797	15.33
Toys & Games	111,301	3,143	61,733	5.74
Tools & Home	64,182	2,047	35,793	8.76
Automotive	34,167	1,211	26,096	10.81
Patio/Lawn	10,702	374	7,293	39.24
Musical	16,501	471	12,206	28.70

Table 2: Statistics of all datasets used in our experimental evaluation

in the interest of space. The statistics of the datasets are reported in Table 2.

Evaluation Setup and Metrics We experiment on the collaborative ranking (or one-class collaborative filtering) setup. We adopt Hit Ratio (HR@10) and nDCG@10 (normalized discounted cumulative gain) evaluation metrics, which are well-established ranking metrics for the task at hand. Following (He et al. 2017; Tay, Anh Tuan, and Hui 2018), we randomly select n negative samples which the user have not interacted with and rank the ground truth amongst these negative samples. We set $n = 100$ since we empirically found this to be sufficient for probing differences in relative performance amongst compared baselines. For all datasets, the last item the user has interacted with is withheld as the test set while the penultimate serves as the validation set. During training, we report the test scores of the model based on the best validation scores.

Compared Baselines In our experiments, we compare with five well-established and competitive baselines.

- **Bayesian Personalized Ranking (BPR)** (Rendle et al. 2009) is a strong collaborative filtering (CF) baseline that takes three inputs include users, positive items, and negative items. The triplet objective is to rank positive item higher than negative item for that user.

- **Multi-layered Perceptron (MLP)** is a feedforward neural network that applies multiple layers of nonlinearities to capture the relationship between users and items. Following (He et al. 2017), we use a three layered MLP with a pyramid structure.

- **Matrix Factorization (MF)** is the standard baseline for recommender systems. It models the user-item representation using the inner product.

- **Neural Collaborative Filtering (NCF)** (He et al. 2017) is the state-of-the-art method for collaborative filtering. The key idea of NCF is to fuse the last hidden representation of MF and MLP together into a joint model.

- **Collaborative Metric Learning (CML)** (Hsieh et al. 2017) is a strong metric learning baseline that learns user-item similarity using the Euclidean distance. CML can be considered a key ablative baseline in our experiments, signifying the difference between Hyperbolic and Euclidean metric spaces.
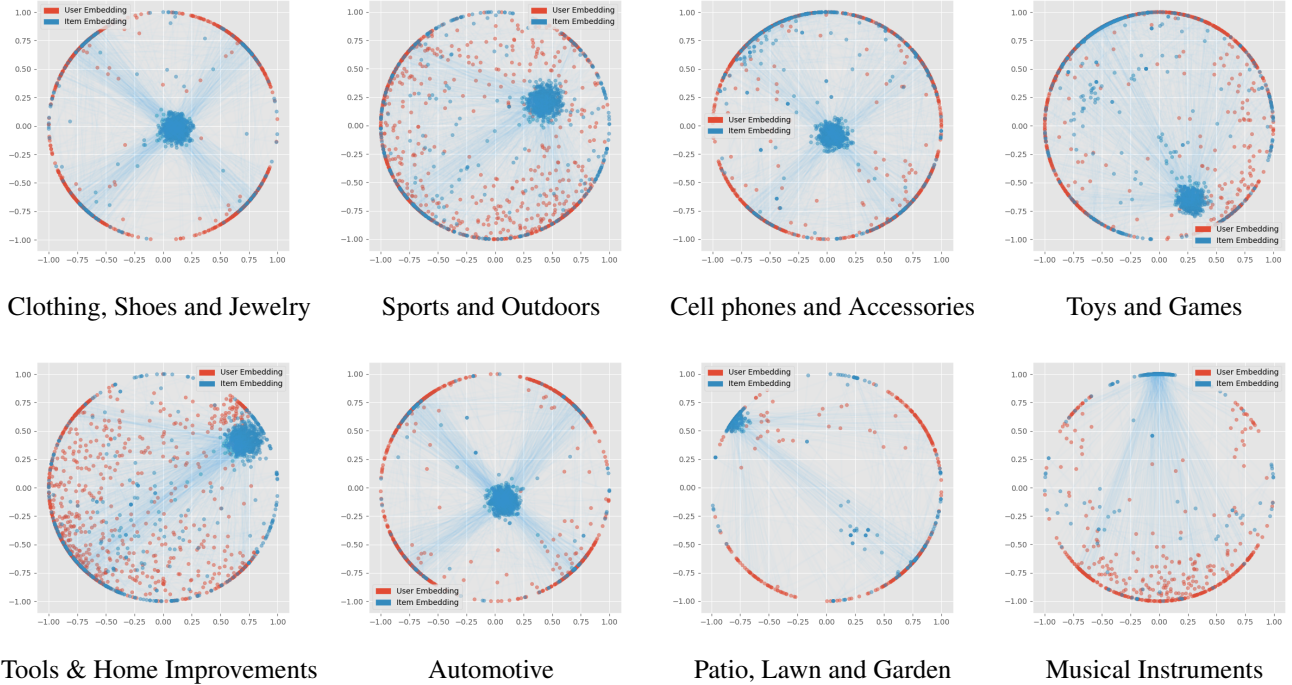
Figure 3: Two-dimensional hyperbolic embedding of 8 Amazon datasets in the Poincaré disk. The images illustrate the embedding of user and item pairs after the convergence.

Implementation Details We implement all models in Tensorflow. All models are trained using Adam (Kingma and Ba 2014) with a learning rate is tuned amongst $\{0.001, 0.005, 0.01\}$. The embedding size d of all models is tuned amongst $\{50, 100, 150\}$ and selectively set to 100. The number of batch B is tuned amongst $\{10, 100, 1000\}$. For models that optimize the hinge loss, the margin λ is tuned amongst $\{0.1, 0.2, 0.5\}$. The NCF and MLP models are implemented following the configuration and architecture in (He et al. 2017); however, the pretrained MF and MLP are not applied to NCF for a fair comparison. All the embeddings and parameters are randomly initialized using the Gaussian distribution with mean of 0 and standard deviation of 0.01. For most datasets and baselines, we empirically set the hyperparameters with the learning rate of 10^{-3}, the number of batches is 10, the embedding size of 100 and the margin is set to 0.1.

Experimental Results

This section experimentally presents our results on all datasets. For all obtained results, the best result is in boldface whereas the second best is underlined. As reported in Table 3, our proposed model significantly outperforms all the baselines on both HR@10 and nDCG@10 metrics across all datasets.

Pertaining to the baselines, CML outperforms other baselines in most of the datasets. We observe that the performance of MF and CML is extremely competitive, i.e. both MF and CML consistently achieve good results across the datasets. The performance gain of CML on the datasets is approximately 1%-2%. Notably, the performance of MF is much better than CML on Patio dataset. One possible reason is that for the small datasets with high density (e.g., Patio with density of 39.24%), a simple model such as MF should be considered as a priority choice. In addition, the performance of NCF is often only comparable to vanilla MLP and MF in most cases. The explanation is because of using the dual embedding spaces (since NCF combines MLP and MF), this kind of usage could possibly lead to the overfitting if the dataset is not large enough (Tay, Anh Tuan, and Hui 2018).

Remarkably, our proposed model HyperBPR significantly outperforms the best baseline method. The percentage improvements in term of nDCG on eight datasets (in the same order as reported in Table 3) are +3.39%, +2.50%, +2.83%, +5.54%, +2.00%, +3.76%, +5.72% and +2.45% respectively. We also observe similar high performance gains on the hit ratio (HR@10). Note that the Amazon datasets follow power-law distribution due to its rich and detailed category hierarchy (McAuley et al. 2015). Therefore, it enables us to achieve very competitive results of our proposed HyperBPR in the hyperbolic space over other strong Euclidean baselines. Informally, since trees require an exponential space for branching in which only hyperbolic geometry has this characteristic, trees prefer to be embedded in the hyperbolic space instead of Euclidean space. In other words, trees can be considered as *discrete hyperbolic spaces* (Krioukov et al. 2010). Our experimental evidence shows the remarkable
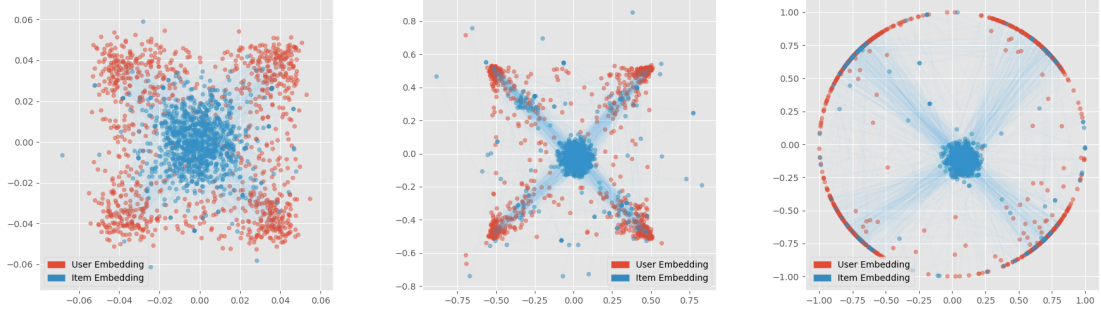
(a). Intermediate embedding of HyperBPR after 10 epochs, 100 epochs and embedding after convergence.
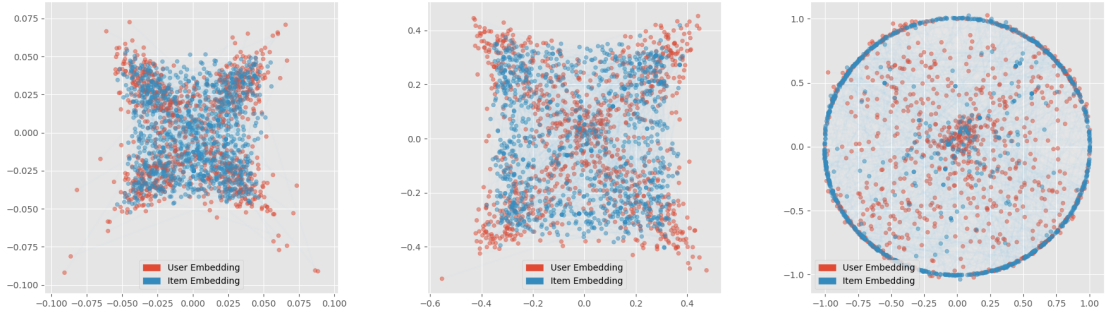
(b). Intermediate embedding of CML after 10 epochs, 100 epochs and embedding after convergence.

Figure 4: Comparison between two-dimensional Poincaré embedding and Euclidean embedding on Automotive dataset. The images illustrate the intermediate embedding of HyperBPR and CML after 10 epochs, 100 epochs and the embedding after convergence.

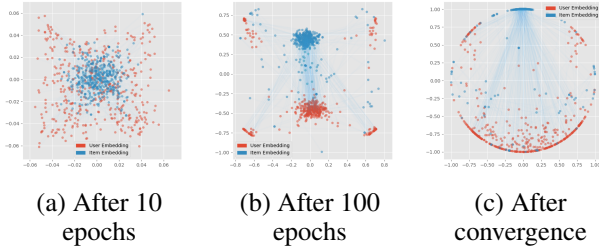(a) After 10 epochs   (b) After 100 epochs   (c) After convergence

Figure 5: Transformation of the user/item embeddings on Musical dataset with respect to the number of epochs.

recommendation results of our proposed HyperBPR model on the variety of datasets and the advantage of hyperbolic space over Euclidean space in handling hierarchical data structure.

Qualitative Analysis

This section investigates the qualitative analysis of our proposed model to understand the behavior of the embeddings in hyperbolic space.

Hyperbolic convergence Figure 3 represents the two-dimensional hyperbolic embedding on the test set of 8 Amazon datasets after the convergence. We observe that item embeddings form a *sphere* over the user embeddings. Moreover, since we conduct the analysis on the **test** set, the visualization of the user/item embeddings in Figure 3 demonstrates the ability of HyperBPR to self-organize and automatically detect the hierarchical structure in the user/item embeddings, as similar to (Tay, Anh Tuan, and Hui 2018).

On a side note, we observe that smaller datasets with high density such as Patio and Musical tends to force the user-item pair embeddings to the boundary of the ball at the convergence. We take the Musical dataset as an example to visualize the transformation of the embeddings. Figure 5 illustrates the user/item embeddings transformation with respect to the number of epochs. It is apparent that after the first 100 epochs, the user and item embeddings are likely to converge and form linking pairs between the embeddings. At the convergence, the embeddings are then pushed toward the boundary, which also give a sign of no hierarchical structure in the dataset.

Convergence comparison Figure 4 illustrates the comparison between two-dimensional Poincaré embedding (Hy-

	BPR		MLP		MF		NCF		CML		HyperBPR	
	HR	nDCG	HR	nDCG	HR	nDCG	HR	nDCG	HR	nDCG	HR	nDCG
Clothing	0.039	0.024	0.058	0.035	0.051	0.032	0.059	0.035	<u>0.066</u>	<u>0.040</u>	**0.120**	**0.074**
Sports	0.149	0.100	0.120	0.071	0.148	0.100	0.118	0.076	<u>0.159</u>	<u>0.107</u>	**0.193**	**0.132**
Cell Phones	0.186	0.128	0.147	0.092	0.200	<u>0.130</u>	0.157	0.101	<u>0.203</u>	0.127	**0.243**	**0.158**
Toys & Games	0.274	0.209	0.255	0.178	0.288	<u>0.216</u>	0.236	0.167	<u>0.292</u>	0.212	**0.360**	**0.272**
Tools & Home	0.139	0.095	0.134	0.087	0.161	<u>0.115</u>	0.146	0.086	<u>0.167</u>	0.112	**0.198**	**0.135**
Automotive	0.034	0.023	0.047	0.030	0.048	<u>0.031</u>	0.048	0.030	<u>0.059</u>	<u>0.037</u>	**0.121**	**0.074**
Patio	0.175	0.116	0.164	0.102	<u>0.208</u>	<u>0.126</u>	0.151	0.092	0.156	0.099	**0.290**	**0.183**
Musical	0.055	0.037	0.037	0.018	0.055	0.033	0.050	0.023	<u>0.059</u>	<u>0.043</u>	**0.116**	**0.068**

Table 3: Experimental results on 8 Amazon datasets. Our proposed HyperBPR achieves very competitive results, outperforming strong Euclidean baselines such as CML and BPR.
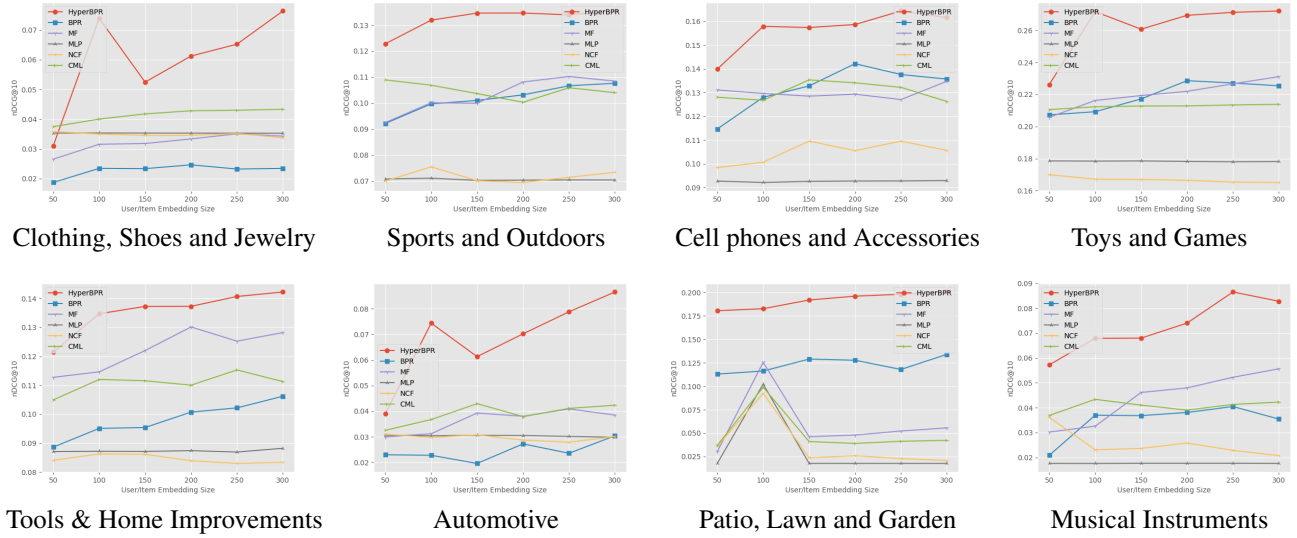

Figure 6: Effects of the embedding size on 8 Amazon datasets.

perBPR) and Euclidean embedding (CML) on the Automotive dataset. For the CML, we decide to clip the norm, i.e. the norm of the embeddings is constrained to 1, for an analogous comparison.

At first glance, we notice the difference between the two types of embedding by observing the distribution of user and item embeddings in the spaces regarding the number of epochs. While HyperBPR has the item embeddings gradually assemble as the number of epochs increases, the item embeddings of CML have the opposite movement. The reason is because the learned metric of CML pulls the positive items closer while simultaneously pushing the negative items further apart; thus, the item embeddings are then pushed toward the boundary. In addition, the convergence of CML shows no hint of hierarchy which is a deficiency compare to HyperBPR.

Effect of Embedding Size

In this section, we study the effect of the embedding size on the performance of our proposed model and the baselines. Figure 6 represents the effect of the embedding size d for

$d \in \{50, 100, 150, 200, 250, 300\}$ on 8 Amazon datasets in term of nDCG@10. In general, we observe that HyperBPR always significantly outperforms the baselines regardless of the embedding size. While NCF maintains a stable performance throughout embedding size, the performance of other baselines seem to slightly fluctuate. Additionally, we notice that HyperBPR has its nDCG@10 only slightly decreases at $d = 150$ but then still maintains superb performance as the embedding size d increases.

Conclusion

In this paper, we introduce a new effective and competent recommendation model called HyperBPR. To the best of our knowledge, HyperBPR is the first model to explore the hyperbolic space in recommender system. Through extensive experiments on 8 datasets, we are able to demonstrate the effectiveness of HyperBPR over other baselines in Euclidean space, even state-of-the-art models such as CML or NCF. The promising results of HyperBPR may inspire other future works to explore hyperbolic space in solving recommendation problems.

References

[Bonnabel 2013] Bonnabel, S. 2013. Stochastic gradient descent on riemannian manifolds. *IEEE Trans. Automat. Contr.* 58(9):2217–2229.

[Cho et al. 2018] Cho, H.; Demeo, B.; Peng, J.; and Berger, B. 2018. Large-margin classification in hyperbolic space. *CoRR* abs/1806.00437.

[Dave et al. 2018a] Dave, V. S.; Zhang, B.; Al Hasan, M.; AlJadda, K.; and Korayem, M. 2018a. A combined representation learning approach for better job and skill recommendation.

[Dave et al. 2018b] Dave, V. S.; Zhang, B.; Chen, P.-Y.; and Hasan, M. A. 2018b. Neural-brane: Neural bayesian personalized ranking for attributed network embedding. *arXiv preprint arXiv:1804.08774*.

[Davidson et al. 2018] Davidson, T. R.; Falorsi, L.; Cao, N. D.; Kipf, T.; and Tomczak, J. M. 2018. Hyperspherical variational auto-encoders. *CoRR* abs/1804.00891.

[Dhingra et al. 2018] Dhingra, B.; Shallue, C. J.; Norouzi, M.; Dai, A. M.; and Dahl, G. E. 2018. Embedding text in hyperbolic spaces. *arXiv preprint arXiv:1806.04313*.

[Ganea, Bécigneul, and Hofmann 2018] Ganea, O.; Bécigneul, G.; and Hofmann, T. 2018. Hyperbolic entailment cones for learning hierarchical embeddings. In *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, 1632–1641.

[Gülçehre et al. 2018] Gülçehre, Ç.; Denil, M.; Malinowski, M.; Razavi, A.; Pascanu, R.; Hermann, K. M.; Battaglia, P.; Bapst, V.; Raposo, D.; Santoro, A.; and de Freitas, N. 2018. Hyperbolic attention networks. *CoRR* abs/1805.09786.

[He and McAuley 2016a] He, R., and McAuley, J. 2016a. Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In *Proceedings of the 25th International Conference on World Wide Web, WWW 2016, Montreal, Canada, April 11 - 15, 2016*, 507–517.

[He and McAuley 2016b] He, R., and McAuley, J. 2016b. Vbpr: Visual bayesian personalized ranking from implicit feedback.

[He et al. 2016] He, X.; Zhang, H.; Kan, M.-Y.; and Chua, T.-S. 2016. Fast matrix factorization for online recommendation with implicit feedback. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, 549–558. ACM.

[He et al. 2017] He, X.; Liao, L.; Zhang, H.; Nie, L.; Hu, X.; and Chua, T. 2017. Neural collaborative filtering. In *Proceedings of the 26th International Conference on World Wide Web, WWW 2017, Perth, Australia, April 3-7, 2017*, 173–182.

[Hsieh et al. 2017] Hsieh, C.; Yang, L.; Cui, Y.; Lin, T.; Belongie, S. J.; and Estrin, D. 2017. Collaborative metric learning. In *Proceedings of the 26th International Conference on World Wide Web, WWW 2017, Perth, Australia, April 3-7, 2017*, 193–201.

[Kingma and Ba 2014] Kingma, D. P., and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

[Koren, Bell, and Volinsky 2009] Koren, Y.; Bell, R.; and Volinsky, C. 2009. Matrix factorization techniques for recommender systems. *Computer* (8):30–37.

[Koren 2008] Koren, Y. 2008. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 426–434. ACM.

[Krioukov et al. 2010] Krioukov, D. V.; Papadopoulos, F.; Kitsak, M.; Vahdat, A.; and Boguñá, M. 2010. Hyperbolic geometry of complex networks. *CoRR* abs/1006.5169.

[McAuley et al. 2015] McAuley, J. J.; Targett, C.; Shi, Q.; and van den Hengel, A. 2015. Image-based recommendations on styles and substitutes. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, Santiago, Chile, August 9-13, 2015*, 43–52.

[Mnih and Salakhutdinov 2008] Mnih, A., and Salakhutdinov, R. R. 2008. Probabilistic matrix factorization. In *Advances in neural information processing systems*, 1257–1264.

[Nickel and Kiela 2017] Nickel, M., and Kiela, D. 2017. Poincaré embeddings for learning hierarchical representations. In Guyon, I.; Luxburg, U. V.; Bengio, S.; Wallach, H.; Fergus, R.; Vishwanathan, S.; and Garnett, R., eds., *Advances in Neural Information Processing Systems 30*. 6338–6347.

[Nickel and Kiela 2018] Nickel, M., and Kiela, D. 2018. Learning continuous hierarchies in the lorentz model of hyperbolic geometry. In *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, 3776–3785.

[Rendle et al. 2009] Rendle, S.; Freudenthaler, C.; Gantner, Z.; and Schmidt-Thieme, L. 2009. BPR: bayesian personalized ranking from implicit feedback. In *UAI 2009, Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, Montreal, QC, Canada, June 18-21, 2009*, 452–461.

[Rendle 2010] Rendle, S. 2010. Factorization machines. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, 995–1000. IEEE.

[Sala et al. 2018] Sala, F.; Sa, C. D.; Gu, A.; and Ré, C. 2018. Representation tradeoffs for hyperbolic embeddings. In *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, 4457–4466.

[Tay, Anh Tuan, and Hui 2018] Tay, Y.; Anh Tuan, L.; and Hui, S. C. 2018. Latent relational metric learning via memory-based attention for collaborative ranking. In *Proceedings of the 2018 World Wide Web Conference*, WWW '18, 729–739. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee.

[Tay, Tuan, and Hui 2018] Tay, Y.; Tuan, L. A.; and Hui, S. C. 2018. Hyperbolic representation learning for fast and efficient neural question answering. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, WSDM '18.

[Wang et al. 2015] Wang, S.; Tang, J.; Wang, Y.; and Liu, H. 2015. Exploring implicit hierarchical structures for recommender systems. In *Proceedings of the 24th International Conference on Artificial Intelligence*, 1813–1819. AAAI Press.

[Wang et al. 2018] Wang, S.; Tang, J.; Wang, Y.; and Liu, H. 2018. Exploring hierarchical structures for recommender systems. volume 30, 1022–1035. IEEE.

[Zhang et al. 2016] Zhang, B.; Choudhury, S.; Hasan, M. A.; Ning, X.; Agarwal, K.; Purohit, S.; and Cabrera, P. P. 2016. Trust from the past: Bayesian personalized ranking based link prediction in knowledge graphs. *arXiv preprint arXiv:1601.03778*.

[Zhang et al. 2018] Zhang, S.; Yao, L.; Sun, A.; Wang, S.; Long, G.; and Dong, M. 2018. Neurec: On nonlinear transformation for personalized ranking. *arXiv preprint arXiv:1805.03002*.

[Zhao et al. 2017] Zhao, P.; Xu, X.; Liu, Y.; Zhou, Z.; Zheng, K.; Sheng, V. S.; and Xiong, H. 2017. Exploiting hierarchical struc-

tures for poi recommendation. In *2017 IEEE International Conference on Data Mining (ICDM)*, 655–664. IEEE.