Hyperbolic Personalized Tag Recommendation

Weibin Zhao[1], Aoran Zhang[2], Lin Shang[1], Yonghong Yu[2], Li Zhang[3], Can Wang[4], Jiajun Chen[1], and Hongzhi Yin[5]

[1] State Key Lab for Novel Software Technology, Nanjing University, Nanjing, China
`njzhaowb@gmail.com, {shanglin, chenjj}@nju.edu.cn`

[2] Tongda college, Nanjing University of Posts and Telecommunications, Yangzhou, China
`njuptzhangaoran@gmail.com, yuyh@njupt.edu.cn`

[3] Department of Computer Science, Royal Holloway, University of London, Surrey, UK
`li.zhang@rhul.ac.uk`

[4] School of Information and Communication Technology, Griffith University, Brisbane, Australia
`can.wang@griffith.edu.au`

[5] School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, Australia
`db.hongzhi@gmail.com`

Abstract. Personalized Tag Recommendation(PTR) aims to automatically generate a list of tags for users to annotate web resources, the so-called items, according to users' tagging preferences. The main challenge of PTR is to learn representations of involved entities (i.e., users, items, and tags) from interaction data without loss of structural properties in original data. To this end, various PTR models have been developed to conduct representation learning by embedding historical tagging information into low-dimensional Euclidean space. Although such methods are effective to some extent, their ability to model hierarchy, which lies in the core of tagging information structures, is restricted by Euclidean space's polynomial expansion property. Since hyperbolic space has recently shown its competitive capability to learn hierarchical data with lower distortion than Euclidean space, we propose a novel PTR model that operates on hyperbolic space, namely HPTR. HPTR learns the representations of entities by modeling their interactive relationships in hyperbolic space and utilizes hyperbolic distance to measure semantic relevance between entities. Specially, we adopt tangent space optimization to update model parameters. Extensive experiments on real-world datasets have shown the superiority of HPTR over state-of-the-art baselines.

Keywords: Hyperbolic spaces · Personalized · Tag recommendation · Embedding.

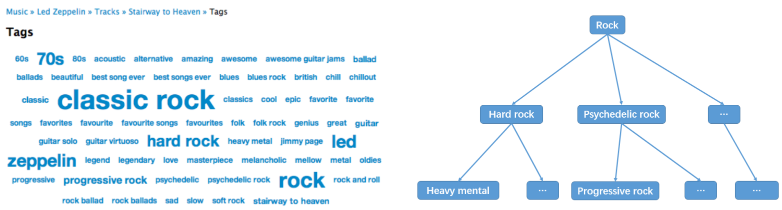
1 INTRODUCTION

Tagging systems have become essential in many web applications, such as Last.FM, Flickr, and YouTube. In the process of tagging, users are allowed to freely add metadata to the songs, videos, products, and other web resources(called items) in the form of keywords, the so-called tags. Besides annotating items, tags are beneficial to the systems and users for efficiently organizing, searching, and sharing the related items. With the increasing availability of tags in various domains, tag recommendation has become a popular service to help users acquire their desired tags more conveniently.

As a subtask of top-N ranking recommendation, the tag recommendation aims to assist users' tagging process by automatically suggesting a ranked list of tags. According to whether users' personalized preferences are considered, the tag recommendation tasks can be divided into the non-personalized and the personalized. The non-personalized tag recommendation (NPTR) [20, 40, 38, 47] aims to generate a list of candidate tags ranked merely by their semantic relevance to the target item, this kind of recommendation will suggest the same tag list to all involved users. On the other hand, the personalized tag recommendation (PTR) [14, 15, 37, 39] takes user's tagging preferences into account, its ultimate goal is to suggest tags that are relevant to both target item and target user, such goal make PTR more complex than NPTR, and as a result, PTR will recommend different tag lists to different users for the same target item. Due to users' diverse intentions and interests, PTR is more meaningful and practical in real scenarios of tag recommendation. Moreover, as indicated in [35], the PTR could outperform the theoretical upper bound of any NPTR.

Serving as the source data for PTR tasks, the users' historical tagging information implies a complex structure that involves three kinds of entities (i.e., users, items, and tags) and multiple interactive relationships, so it is a challenge for PTR to accurately learn latent representations of entities with preservation of their real semantic relevance in such data. To this end, various learning methods have been proposed to boost the performance of the PTR model. The core of tagging information is the ternary interaction, i.e., user-item-tag, which can be naturally represented by a three-order tensor. Thus tensor factorization techniques are widely adopted [3, 9, 35, 37, 39, 49] to boost the performances of PTR models. Although tensor factorization-based PTR models are effective to a certain extent, all of them are conducted in Euclidean spaces. That is, they learn latent representations of entities in a low-dimensional embedding space and adopt matching functions that only cover the scope of Euclidean space, such as inner product, Euclidean distance, and neural networks, to compute the semantic relevance between embeddings [42, 43] because Euclidean spaces are the natural generalization of our intuition.

On the other hand, recent research [8, 11, 25, 29, 30] has shown that Euclidean embedding have distortion for many real-world data, which follows power-law distribution or exhibits scale-free. It means that the actual logical patterns and semantic relevance can not be well preserved when embedding such data in Euclidean spaces. As mentioned in [29, 30], the data with power-law distribution tend to have tree-like structures. For a tree, the number of its nodes grows exponentially with the tree depth, but the volume of Euclidean spaces grows polynomially with distance from the origin point. In terms of tagging information, the interactive relationships it contains have been found to follow power-law distribution [12, 23, 31, 34], e.g., a small portion of tags (hot tags) are frequently used to annotate a certain kind of items while massive tags are seldom adopted. As power-law distributions and tree-like structures can be explained by assuming an underlying hierarchy in the data [29, 30], it is therefore crucial for PTR to leverage this insight to develop an optimal learning model with lower embedding distortions [8]. Moreover, through our observation, this phenomenon does exist: For real users, some prefer to use tags with general semantics, while others prefer tags with descriptive semantics. Such personal preference divides users into different hierarchies with respect to the semantics. As shown in Fig. 1(a), in Last.FM, a popular track like Led Zeppelin's "Stairway to Heaven" has dozens of unique tags applied hundreds of times. According to our knowledge, some latent semantic hierarchies (Fig.1(b)) may exist among these annotated tags.



(a) (b)

Fig. 1. Hierarchies implied in tags

Notably, hyperbolic space has shown promise in modeling hierarchical data [5, 8, 29, 30] in recent years. Hyperbolic space is a kind of non-Euclidean space with constant negative curvature. If we embed a disk into a two-dimensional hyperbolic space with curvature $K = -1$, its corresponding circumference $(2\pi \sinh r)$ and area $(2\pi(\cosh r - 1))$ both grow exponentially with the radius r, as opposed to the two-dimensional Euclidean space where the corresponding circumference $(2\pi r)$ and area $\left(\pi r^2\right)$ grows linearly and quadratically respectively. Thus hyperbolic space can be viewed as the continuous version of a tree, and it is well-suited for embedding hierarchical data with lower dimensions than the Euclidean space.

Motivated by the above merits of hyperbolic space, we develop a novel PTR model with hyperbolic embedding, namely, HPTR. Our idea is to learn representations of entities by embedding historical tagging information into the Poincaré ball, which is an isometric hyperbolic space model and feasible to perform a gradient-descent step. As the PTR is an implicit feedback recommendation task, we build an objective function based on Bayesian Personalized Ranking (BPR) optimization criterion. Besides, there exist multiple relations between entities, which more or less affect the performance of PTR, so the main difficulty of HPTR is how to discover users' tagging preferences by dealing with these relations in hyperbolic space. Unlike the traditional item recommendation, in PTR scenarios, the users' preferences for items make no sense for learning and predicting. Thus we take two relations into consideration: $user-tag$ and $item-tag$, which have been proven effective in many PTR models [9, 37, 44, 49]. Finally, we can utilize the hyperbolic distance between involved entities to reflect one user's tagging preference for a target item.

The main contributions of our work are summarized as follows:

- We bridge the gap between PTR and hyperbolic geometry by discovering common structural properties between tagging information and hyperbolic space. With the expectation for achieving better performance in the PTR task, we propose a novel PTR model conducted in hyperbolic space, namely HPTR, which learns the representations of entities on Poincaré ball and measures users' tagging preferences by hyperbolic distance. To the best of our knowledge, this is the first work to integrate PTR with hyperbolic space.
- We conduct extensive experiments on three real-world datasets to verify the efficiency of the proposed model, and experimental results show that our HPTR can outperform the state-of-the-art PTR model, especially with lower embedding dimensions.

2 RELATED WORK

Our work is related to the following research directions.

2.1 Personalized Tag Recommendation Methods

With the popularization of tagging systems in various web applications, personalized tag recommendation (PTR) is becoming more attractive in the field of recommender systems. Considering the core of users' historical tagging information is the ternary interaction between entities, i.e., user, item, and tag, which can be represented by a three-order tensor naturally, so most early studies utilized tensor factorization techniques, especially the tucker decomposition (TD) to learn representations of involved entities in PTR tasks [3, 35, 39].

Due to the model equation of TD resulting in a cubic runtime in the factorization dimension, the computation cost of TD makes it infeasible for large-scale PTR tasks. Rendle et al. [37] proposed the pairwise interaction tensor factorization (PITF) model to tackle this problem, which explicitly models the pairwise interactions between entities and results in linear runtime. PITF has been extensively studied for its outperformance, and more learning methods derived from it [9, 16] have been proposed to fit new problem scenarios. Moreover, to take advantage of the end-to-end learning ability of deep neural networks(DNN), several learning frameworks based on DNN [27, 28, 49] are developed to improve the performance of traditional PTR models further.

Note that all the above models are conducted in Euclidean spaces. As we mentioned before, their capabilities of learning the representations of hierarchical data are restricted by the polynomial expansion property of Euclidean space.

2.2 Hyperbolic Embedding

In the field of representation learning, hyperbolic spaces have started to get attention from the studies on how to discover suitable embedding spaces to model complex networks [21]. As scale-free and strong clustering are typical properties of complex networks, and such properties can be traced back to hierarchical structure within them, hyperbolic spaces have become a better choice for their capability of modeling hierarchical structures. Since then, there has been an increasing interest in utilizing hyperbolic embedding to learn representations of data with explicit or implicit hierarchies, representative research or applications of hyperbolic embedding include but are not limited to: Natural Language Processing [8, 25, 29], Knowledge Graph embedding [1, 5, 19], Heterogeneous Information Network embedding [46], Neural Networks based representation learning [6, 11, 24] and Computer Vision tasks [17, 32]. In the domain of recommender systems, the original work exploring the use of hyperbolic space for the recommender systems is [42]. Subsequently, a number of models enhanced by hyperbolic embedding [4, 10, 22, 26, 43] have been proposed in order to get better performance in traditional recommendation tasks or cope with new tasks, via a series of corresponding experiments, these models have demonstrated the superiority over their Euclidean counterparts and state-of-the-art baselines. Nevertheless, the existing studies of hyperbolic recommender systems have not covered the scope of PTR.

3 PRELIMINARIES

3.1 Problem Description

Unlike the item recommendation systems containing two types of entities, i.e., users and items, PTR consists of three types of entities: the set of users U, the set of items I, and the set of tags T. The historical tagging information between users, items, and tags is represented as $S \subseteq U \times I \times T$. A ternary $(u, i, t) \in S$ indicates that the user u has annotated the item i with the tag t. From the ternary relation set S, personalized tag recommendation methods usually deduce a three-order tensor $Y \in \mathbb{R}^{|U|\times|I|\times|T|}$, whose element $y_{u,i,t}$ is defined as follows:

$$y_{u,i,t} = \begin{cases} 1, & (u, i, t) \in S \\ 0, & otherwise, \end{cases} \tag{1}$$

where $y_{u,i,t} = 1$ indicates a positive instance, and the remaining data are the mixture of negative instances and missing values. In addition, the tagging information for the user-item pair (u, i) is defined as $\mathbf{y}_{u,i} = \{y_{u,i,t} | y_{u,i,t}, t \in T\}$.

PTR aims to recommend a ranked list of tags to a certain user for annotating a certain item. Usually, a score function $\widehat{Y} : U \times I \times T \longrightarrow \mathbb{R}$ is employed to measure users' preferences on tags for their target items. The entry $\widehat{y}_{u,i,t}$ of $\widehat{Y}$ indicates the degree to which a user u prefers to annotate the item i with the tag t. After predicting the score $\widehat{y}_{u,i,t}$ for all candidate tag t given a $user - item$ pair (u, i), the personalized tag recommender system returns a ranked list of Top-N tags in terms of the obtained scores. Formally, the ranked list of Top-N tags given to the $user - item$ pair (u, i) is defined as follows:

$$Top(u, i, N) = \underset{t \in T}{\overset{N}{argmax}}\ \widehat{y}_{u,i,t}, \tag{2}$$

where N denotes the number of recommended tags.

3.2 Hyperbolic Embedding

Hyperbolic space is a smooth Riemannian manifold with constant negative curvature, and five isometric models can describe it [33], which are the Lorentz (hyperboloid) model, the Poincaré ball model, the Poincaré half space model, the Klein model, and the hemishpere model. Our work chooses the Poincaré ball to describe the embedding space, for it is relatively suitable for modeling a tree.

Let $\mathcal{B}^d = \left\{ \boldsymbol{x} \in \mathbb{R}^d \mid \|\boldsymbol{x}\| < 1 \right\}$ be the an open d-dimensional unit ball, where $\| \cdot \|$ denotes the Euclidean norm. The Poincaré ball can be defined by the Riemannian manifold $\left(\mathcal{B}^d, g_{\boldsymbol{x}}^B\right)$, in which $g_{\boldsymbol{x}}^B$ is the Riemannian metric tensor given as:

$$g_{\boldsymbol{x}}^B = \left(\frac{2}{1 - \|\boldsymbol{x}\|^2} \right)^2 g^E \tag{3}$$

where $x \in \mathcal{B}^d$ and $g^{\mathbb{E}} = \mathbf{I}$ denotes the Euclidean metric tensor. Furthermore, the distance between points $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{B}^d$ is given as:

$$d_B(\boldsymbol{x}, \boldsymbol{y}) = \operatorname{arcosh} \left(1 + 2 \frac{\|\boldsymbol{x} - \boldsymbol{y}\|^2}{(1 - \|\boldsymbol{x}\|^2)(1 - \|\boldsymbol{y}\|^2)} \right) \tag{4}$$

It is worth noting that the Poincaré ball model is conformal: the angles of embedded vectors are equal to their angles in the Euclidean space, thus making Poincaré ball suitable for the gradient-based learning method.

4 HPTR MODEL

Due to the core of tagging information is the ternary interaction among entities, i.e., $user - item - tag$, our HPTR is committed to learn three embedding matrices: $\mathbf{U} \in \mathbb{R}^{|U| \times d}, \mathbf{I} \in \mathbb{R}^{|I| \times d}, \mathbf{T} \in \mathbb{R}^{|T| \times d}$ (d is the embedding dimension). In addition, there exist multiple relations between entities, which more or less affect the performance of PTR, so the main difficulty of HPTR is how to discover users' tagging preferences by dealing with these relations in hyperbolic space. Inspired by the work [35], which claims that the relation between user and item is meaningless for modeling users' tagging preference. So we take two factors into consideration: $user - tag$ relation and $item - tag$ relation.

Furthermore, as indicated in section 3, the final output of PTR is a list of Top-N tags, which ranked by the predicting score $\widehat{y}_{u,i,t}$ for all candidate tag t with respect to a certain pair (u, i), so we merely need to calculate the semantic relevance between (u, i) and t in embedding space, instead of u and t, i and t, respectively. Hence we add embedding of *user* and *item* together, i.e, $u + i$ to represent each (u, i) pair. Meanwhile, in the embedding space, if a user u prefers to annotate the item i with the tag t, the distance d_B between point $u + i$ and point t on $\mathcal{B}^d$ should be relatively shorter, and vice versa. Consequently, our HPTR can measure users' tagging preferences by using the score function $\widehat{y}_{u,i,t}$ defined as:

$$\widehat{y}_{u,i,t} = p\left(d_B\left(\left(\mathbf{U_u} + \mathbf{I_i}\right), \quad \mathbf{T_t}\right)\right) \tag{5}$$

where $\mathbf{U}_u = \mathbf{U}.onehot(u), \quad \mathbf{I}_i = \mathbf{I}.onehot(i), \quad \mathbf{T}_t = \mathbf{T}.onehot(t)$ are embeddings of a given triple (u, i, t); $.onehot()$ indicates the operation of lookup in embedding table according to one-hot id encoding; $p(\cdot)$ is the transformation function for converting hyperbolic distances d_B to users' tagging preference, here we take it as $p(x) = \beta x + c$ with $\beta \in \mathbb{R}$ and $c \in \mathbb{R}$ similar to [42].

4.1 Objective Function

In this work, we agree with the assumption in [37]: When we observe a certain pair (u, i) in tagging information S, we believe that the user u should prefer tag t over tag t' iff the triple (u, i, t) can be observed from historical tagging information and (u, i, t') can not be observed. Based on this assumption, the training set D_S (i.e., the set of quadruple (u, i, t, t')) with the pairwise constraint is defined as:

$$D_S = \{(u, i, t, t') \mid (u, i, t) \in S \wedge (u, i, t') \notin S\} \tag{6}$$

The objective of model training is to maximize the margin between the scores $\widehat{y}_{u,i,t}$ of the positive triple (u, i, t) and negative triple (u, i, t'), so we adopt the Bayesian Personalized Ranking (BPR) optimization criterion [36] to learn model parameters $\Theta = \{\mathbf{U}, \mathbf{I}, \mathbf{T}, \beta, \mathbf{c}\}$, and build the objective function of HPTR as:

$$\mathcal{L}^{HPTR} = \min_{\Theta} \sum_{(u,i,t,t') \in D_S} - \ln \sigma\left(\widehat{y}_{u,i,t,t'}\right) + \lambda_\Theta \|\Theta\|_F^2 \tag{7}$$

4.2 Optimization

As the Poincaré ball is a Riemannian manifold with constant negative curvature, the parameters lies in the ball should be updated by Riemannian gradient, so the Riemannian stochastic gradient descent(RSGD) [2] has been applied to optimize most of Poincaré embedding based models [10, 29, 42, 46]. In terms of HPTR, the model parameters consist of embedded parameters(i.e., $\{\mathbf{U}, \mathbf{I}, \mathbf{T}\} \in \mathcal{B}^d$) and non-embedded parameters(i.e., $\{\beta, \mathbf{c}\} \notin \mathcal{B}^d$), therefore, we update the two types of parameters together via tangent space optimization [5, 6] to avoid using two corresponding optimizers.

$$\boldsymbol{\theta}_{t+1} = \operatorname{proj}\left(\boldsymbol{\theta}_t - \eta \frac{\left(1 - \|\boldsymbol{\theta}_t\|^2\right)^2}{4} \nabla_E\right) \tag{8}$$

We recall that a d-dimensional hyperbolic space is a Riemannian manifold $\mathcal{M}$ with a constant negative curvature $-c(c > 0)$, the tangent space $\mathcal{T}_{\boldsymbol{x}}\mathcal{M}$ at point $\boldsymbol{x}$ on $\mathcal{M}$ is a d-dimensional flat space that best approximates $\mathcal{M}$ around $\boldsymbol{x}$, and the elements $\mathbf{v}$ of $\mathcal{T}_{\mathrm{x}}\mathcal{M}$ are referred to as tangent vectors. In our work, We define all of parameters in the tangent space of the Poincaré ball so that we can learn them via powerful Euclidean optimizers(e.g., Adam). In particular, for the calculation of $\mathbf{U_u} + \mathbf{I_i}$ in Equation 5, we also do it in tangent space beforehand.

When it comes to calculate the hyperbolic distance d_B, we use the exponential map $\exp_{\boldsymbol{x}}^c(\mathbf{v})$ to recover the corresponding parameters (map $\mathbf{v}$ of tangent space back to $\mathcal{B}^d$) as following:

$$\exp_{\boldsymbol{x}}^c(\mathbf{v}) = \boldsymbol{x} \oplus_c \left(\tanh\left(\sqrt{c}\frac{\lambda_{\boldsymbol{x}}^c\|\mathbf{v}\|}{2}\right)\frac{\mathbf{v}}{\sqrt{c}\|\mathbf{v}\|}\right) \tag{9}$$

Where $\oplus^c$ denotes the $M\ddot{o}bius$ addition operator [11] that provides an analogue to Euclidean addition for hyperbolic space.

5 EXPERIMENTS AND ANALYSIS

In this section, we conduct several groups of experiments on two real-world datasets to compare the performance of HPTR with other state-of-the-art PTR models.

5.1 Datasets and Evaluation Metrics

In our experiments, we choose two public available datasets [6], i.e., LastFM and ML10M, to evaluate the performance of all compared methods. Similar to [35, 37], we preprocess each dataset to obtain their corresponding p-core, which is the largest subset where each user, item, and tag has to occur at least p times. In our experiments, every datasets is 5-core or 10-core. The general statistics of datasets are summarized in Table 1.

To evaluate the recommendation performance of all compared methods, we adopt the $leave-one-out$ evaluation protocol, which has been widely used in related studies. Specifically, for each pair (u, i), we select the last triple (u, i, t) according to the tagging time and remove it from S to S_{test}. The remaining observed $user-item-tag$ triples are the training set $S_{train} = S - S_{test}$. Similar to the item recommendation problem, the PTR provides a top- N highest ranked

[6] https://grouplens.org/datasets/hetrec-2011/

Table 1. Description of datasets.

Dataset	Users	Items	Tags	Tag assignments	Density
LastFM-core5	1348	6927	2132	162047	8.13989E-06
LastFM-core10	966	3870	1024	133945	3.49896E-05
ML10M-core5	990	3247	2566	61688	7.47871E-06
ML10M-core10	469	1524	1017	37414	5.14701E-05

list of tags for a pair (u, i). We employ two typical ranking metrics to measure the performance of all compared methods, i.e., Precision@N and Recall@N. For both metrics, we set $N = 3, 5, 10$.

5.2 Experiment Settings

We choose the following traditional tag recommendation algorithms as baselines:

- PITF: PITF [37] explicitly models the pairwise interactions among users, items and tags by inner product, it is a strong competitor in the field of personalized tag recommendation.
- NLTF [9] is a non-linear tensor factorization model, which enhances PITF by exploiting the Gaussian radial basis function to capture the nonlinear interaction relations among users, items and tags.
- ABNT: ABNT [49] utilizes the multi-layer perception to model the nonlinearities of the interactions among users, items and tags.

We empirically set the parameters of compared models according to their corresponding literature in order to recover their optimal performance: the dimension of embedding d is set to 64. In addition, for the ABNT model, the number of hidden layers is set to 2. For HPTR, we set curvature $-c = -1$, and map tangent space $\mathcal{T}_{\boldsymbol{x}}\mathcal{M}$ at origin point $\boldsymbol{x} = 0$ on the Poincaré ball, the dimension of embedding d is tuned amongst $\{8, 16, 32, 64\}$. We choose Adam [18] as the optimizer for all involved models.

5.3 Performance Comparison

Tables 2-5 present the tag recommendation quality of all compared method on the selected four datasets.

From the inspection of Tables 2-5, we have observed the following experimental results:

- PITF is superior to NTLF and ABNT with respect to all evaluation metrics, which indicates that, for the adoption of Euclidean matching functions, the traditional inner product might be a better choice for the PTR model to measure the semantic relevance between entities.

Table 2. Recommendation Quality Comparisons on LastFM-core5.

Model	PITF	NLTF	ABNT	HPTR-8	HPTR-16	HPTR-32	HPTR-64
Precision@3	0.2127	0.1949	0.1563	0.1944	0.2425	0.2634	**0.2813**
Precision@5	0.1789	0.1678	0.1353	0.1590	0.1925	0.2112	**0.2229**
Precision@10	0.1274	0.1191	0.1018	0.1095	0.1275	0.1413	**0.1424**
Recall@3	0.2571	0.2275	0.1569	0.2477	0.3127	0.3432	**0.3600**
Recall@5	0.3479	0.3239	0.2194	0.3201	0.3857	0.4061	**0.4382**
Recall@10	0.4814	0.4523	0.3298	0.4065	0.4722	0.5129	**0.5191**

Table 3. Recommendation Quality Comparisons on LastFM-core10.

Model	PITF	NLTF	ABNT	HPTR-8	HPTR-16	HPTR-32	HPTR-64
Precision@3	0.2513	0.2443	0.1641	0.2343	0.2797	0.3094	**0.3162**
Precision@5	0.2088	0.2062	0.1367	0.1875	0.2153	0.2431	**0.2555**
Precision@10	0.1458	0.1249	0.0941	0.1261	0.1486	0.1631	**0.1675**
Recall@3	0.3204	0.2845	0.1579	0.2961	0.3552	0.3894	**0.4001**
Recall@5	0.4158	0.4017	0.2190	0.3740	0.4346	0.4762	**0.4914**
Recall@10	0.5654	0.5541	0.3034	0.4704	0.5262	0.5607	**0.5696**

Table 4. Recommendation Quality Comparisons on ML10M-core5.

Model	PITF	NLTF	ABNT	HPTR-8	HPTR-16	HPTR-32	HPTR-64
Precision@3	0.1398	0.1323	0.0822	0.0693	0.1218	0.1610	**0.1711**
Precision@5	0.1021	0.0972	0.0628	0.0523	0.0881	0.1116	**0.1206**
Precision@10	0.0641	0.0596	0.0400	0.0359	0.0547	0.0661	**0.0707**
Recall@3	0.3208	0.2974	0.2089	0.1574	0.2723	0.3480	**0.3717**
Recall@5	0.3910	0.3560	0.2538	0.1949	0.3206	0.3938	**0.4200**
Recall@10	0.4623	0.4270	0.3039	0.2622	0.3840	0.4507	**0.4766**

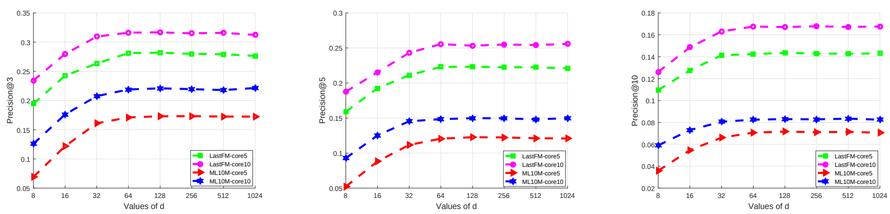
Table 5. Recommendation Quality Comparisons on ML10M-core10.

Model	PITF	NLTF	ABNT	HPTR-8	HPTR-16	HPTR-32	HPTR-64
Precision@3	0.1699	0.1436	0.0896	0.1259	0.1761	0.2075	**0.2189**
Precision@5	0.1173	0.1143	0.0759	0.0928	0.1251	0.1454	**0.1484**
Precision@10	0.0744	0.0714	0.0501	0.0590	0.0730	0.0806	**0.0825**
Recall@3	0.3770	0.3388	0.2210	0.3051	0.4078	0.4742	**0.4970**
Recall@5	0.4523	0.4334	0.3015	0.3605	0.4743	0.5395	**0.5486**
Recall@10	0.5205	0.5341	0.3858	0.4406	0.5358	0.5877	**0.5960**

- Compared against the most competitive PITF with respect to the same embedding dimension $d = 64$, HPTR improves the Precision@3 of PITF by 32.2%, 25.8%, 22.3%, and 28.9% on Lastfm-core5, Lastfm-core10, ML10M-core5, and ML10M-core10, respectively. For Recall@3, the improvements of HPTR over PITF are 40.0%, 24.9%, 15.9%, and 31.8% on the above four datasets, respectively. It implies that Hyperbolic space can provide a more suitable inductive-bias for modeling interactive relationships in tagging information.
- For each compared method, its recommendation performance is better on the 10-core datasets than that on the corresponding 5-core datasets. This observation indicates that HPTR may result in better recommendation performance on datasets with higher density.
- The HPTR has achieved the best recommendation performance over al1 evaluation metrics. Notably, its performance with the lower hyperbolic embedding dimension: on LastFM-core5, LastFM-core10, and ML10M-core10 with $d = 16$, and on ML10M-core5 with $d = 32$, HPTR has met or exceeded the performance of other models with higher Euclidean embedding dimension ($d = 64$), which confirms the prominent advantage of hyperbolic space in the representation capacity.

5.4 Parameters sensitivity analysis

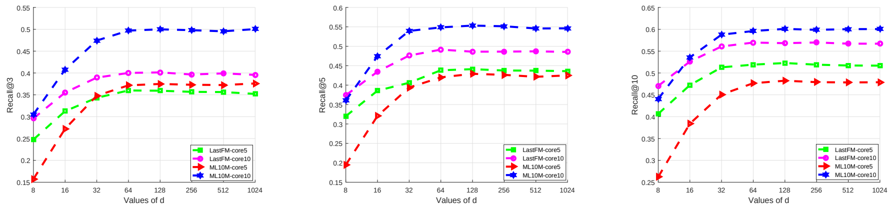
In our proposed HPTR, the dimension of embeddings d is the most important parameter since it controls the capacity of the whole model, so we conduct additional experiments to study the sensitivity of d to the performance of PTR by tuning it within $\{8, 16, 32, 64, 128, 256, 512, 1024\}$. We also take Precision@N and Recall@N, and set $N = 3, 5, 10$, to give an insight of impact on performance with respect to parameter d.



(a) Precision@3 (b) Precision@5 (c) Precision@10

Fig. 2. Impact of d on Precision@N

From Fig 2 and Fig 3, we can observe that the curves of Precision@N and Recall@N show similar changing trends on four datasets. In the beginning, the values of Precision@N and Recall@N both increase stably with the growth of d, when d exceeds 64, most of Precision@$N$and Recall@N are no longer in an



(a) Recall@3 (b) Recall@5 (c) Recall@10

Fig. 3. Impact of d on Recall@N

uptrend, which indicates that merely increasing the dimension is not conducive to sustained improvement of recommendation. One possible reason is that HPTR will obtain sufficient learning ability when d reaches a certain threshold, and after that, the higher dimension of embeddings will lead to the over-fitting problem.

6 CONCLUSION

In this paper, we take the initiative to conduct PTR task in hyperbolic space. Being aware of the common structural properties of hyperbolic space and tagging information, we propose the HPTR model to pursue better recommendation performances. HPTR is committed to learning optimal representation of users, items, and tags in the hyperbolic space. By embedding training data into Poincaré ball and adopting hyperbolic distance as the matching function, our HPTR is expected to exhibit better performance in the recommendation for its capability of capturing the hierarchical relationship within the training data. We conduct extensive experiments to verify the validity of HPTR, and the experimental results have shown its superiority over state-of-the-art baselines. Furthermore, HPTR with a lower hyperbolic embedding dimension can outperform baselines with higher Euclidean embeddings in the experiments.

It should be noted that, in our work, we have not made the most of side information within the tagging information and have overlooked the graph structure among entities. Recently, there appears some work [45, 50, 48, 7, 41, 13] that has been made to improve the performance of recommendation models by inducing side information and graph learning methods. This will motivate us to consider whether we can utilize such information and methods to further boost the performance of PTR. Furthermore, since hyperbolic space can be described by more than one isometric model, our future work will mainly focus on exploring the performance of the other isometric models in PTR tasks.

Acknowledgements The authors would like to acknowledge the support for this work from the Natural Science Foundation of the Higher Education Institutions of Jiangsu Province (17KJB520028), Tongda College of Nanjing University of Posts and Telecommunications (XK203XZ21001), and Future Network Scientific Research Fund Project(FNSRFP-2021-YB-54)

References

1. Balažević, I., Allen, C., Hospedales, T.: Multi-relational poincaré graph embeddings. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems. pp. 4463–4473 (2019)
2. Bonnabel, S.: Stochastic gradient descent on riemannian manifolds. IEEE Transactions on Automatic Control **58**(9), 2217–2229 (2013)
3. Cai, Y., Zhang, M., Luo, D., Ding, C., Chakravarthy, S.: Low-order tensor decompositions for social tagging recommendation. In: Proceedings of the fourth ACM international conference on Web search and data mining. pp. 695–704 (2011)
4. Chamberlain, B.P., Hardwick, S.R., Wardrope, D.R., Dzogang, F., Daolio, F., Vargas, S.: Scalable hyperbolic recommender systems. arXiv preprint arXiv:1902.08648 (2019)
5. Chami, I., Wolf, A., Juan, D.C., Sala, F., Ravi, S., Ré, C.: Low-dimensional hyperbolic knowledge graph embeddings. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 6901–6914 (2020)
6. Chami, I., Ying, R., Re, C., Leskovec, J.: Hyperbolic graph convolutional neural networks. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems. pp. 4868–4879 (2019)
7. Chen, L., Wu, L., Hong, R., Zhang, K., Wang, M.: Revisiting graph based collaborative filtering: A linear residual graph convolutional network approach. In: Proceedings of the AAAI conference on artificial intelligence. vol. 34, pp. 27–34 (2020)
8. Dhingra, B., Shallue, C., Norouzi, M., Dai, A., Dahl, G.: Embedding text in hyperbolic spaces. In: Proceedings of the Twelfth Workshop on Graph-Based Methods for Natural Language Processing (TextGraphs-12). pp. 59–69 (2018)
9. Fang, X., Pan, R., Cao, G., He, X., Dai, W.: Personalized tag recommendation through nonlinear tensor factorization using gaussian kernel. In: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence. pp. 439–445 (2015)
10. Feng, S., Tran, L.V., Cong, G., Chen, L., Li, J., Li, F.: Hme: A hyperbolic metric embedding approach for next-poi recommendation. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 1429–1438 (2020)
11. Ganea, O.E., Bécigneul, G., Hofmann, T.: Hyperbolic neural networks. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems. pp. 5350–5360 (2018)
12. Halpin, H., Robu, V., Shepherd, H.: The complex dynamics of collaborative tagging. In: Proceedings of the 16th international conference on World Wide Web. pp. 211–220 (2007)
13. Han, J., Zheng, L., Xu, Y., Zhang, B., Zhuang, F., Philip, S.Y., Zuo, W.: Adaptive deep modeling of users and items using side information for recommendation. IEEE transactions on neural networks and learning systems **31**(3), 737–748 (2019)
14. Hotho, A., Jäschke, R., Schmitz, C., Stumme, G.: Information retrieval in folksonomies: Search and ranking. In: European semantic web conference. pp. 411–426. Springer (2006)

15. Jäschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L., Stumme, G.: Tag recommendations in folksonomies. In: European conference on principles of data mining and knowledge discovery. pp. 506–514. Springer (2007)
16. Jiang, F., Yu, Y., Zhao, W., Zhang, L., Jiang, J., Wang, Q., Chen, X., Huang, G.: Personalized tag recommendation via adversarial learning. In: Developments of Artificial Intelligence Technologies in Computation and Robotics: Proceedings of the 14th International FLINS Conference (FLINS 2020). pp. 923–930. World Scientific (2020)
17. Khrulkov, V., Mirvakhabova, L., Ustinova, E., Oseledets, I., Lempitsky, V.: Hyperbolic image embeddings. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6418–6428 (2020)
18. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
19. Kolyvakis, P., Kalousis, A., Kiritsis, D.: Hyperkg: Hyperbolic knowledge graph embeddings for knowledge base completion. arXiv preprint arXiv:1908.04895 (2019)
20. Krestel, R., Fankhauser, P., Nejdl, W.: Latent dirichlet allocation for tag recommendation. In: Proceedings of the third ACM conference on Recommender systems. pp. 61–68 (2009)
21. Krioukov, D., Papadopoulos, F., Kitsak, M., Vahdat, A., Boguná, M.: Hyperbolic geometry of complex networks. Physical Review E **82**(3), 036106 (2010)
22. Li, A., Yang, B., Chen, H., Xu, G.: Hyperbolic neural collaborative recommender. arXiv preprint arXiv:2104.07414 (2021)
23. Li, X., Guo, L., Zhao, Y.E.: Tag-based social interest discovery. In: Proceedings of the 17th international conference on World Wide Web. pp. 675–684 (2008)
24. Liu, Q., Nickel, M., Kiela, D.: Hyperbolic graph neural networks. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems. pp. 8230–8241 (2019)
25. López, F., Heinzerling, B., Strube, M.: Fine-grained entity typing in hyperbolic space. In: Proceedings of the 4th Workshop on Representation Learning for NLP (RepL4NLP-2019). pp. 169–180 (2019)
26. Ma, C., Ma, L., Zhang, Y., Wu, H., Liu, X., Coates, M.: Knowledge-enhanced top-k recommendation in poincaré ball. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 35, pp. 4285–4293 (2021)
27. Nguyen, H.T., Wistuba, M., Grabocka, J., Drumond, L.R., Schmidt-Thieme, L.: Personalized deep learning for tag recommendation. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining. pp. 186–197. Springer (2017)
28. Nguyen, H.T., Wistuba, M., Schmidt-Thieme, L.: Personalized tag recommendation for images using deep transfer learning. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. pp. 705–720. Springer (2017)
29. Nickel, M., Kiela, D.: Poincaré embeddings for learning hierarchical representations. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. pp. 6341–6350 (2017)
30. Nickel, M., Kiela, D.: Learning continuous hierarchies in the lorentz model of hyperbolic geometry. In: International Conference on Machine Learning. pp. 3779–3788. PMLR (2018)
31. Pan, Y., Huo, Y., Tang, J., Zeng, Y., Chen, B.: Exploiting relational tag expansion for dynamic user profile in a tag-aware ranking recommender system. Information Sciences **545**, 448–464 (2021)

32. Peng, W., Shi, J., Xia, Z., Zhao, G.: Mix dimension in poincaré geometry for 3d skeleton-based action recognition. In: Proceedings of the 28th ACM International Conference on Multimedia. pp. 1432–1440 (2020)
33. Peng, W., Varanka, T., Mostafa, A., Shi, H., Zhao, G.: Hyperbolic deep neural networks: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence (2021)
34. Rader, E., Wash, R.: Influences on tag choices in del. icio. us. In: Proceedings of the 2008 ACM conference on Computer supported cooperative work. pp. 239–248 (2008)
35. Rendle, S., Balby Marinho, L., Nanopoulos, A., Schmidt-Thieme, L.: Learning optimal ranking with tensor factorization for tag recommendation. In: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 727–736 (2009)
36. Rendle, S., Freudenthaler, C., Gantner, Z., Schmidt-Thieme, L.: Bpr: Bayesian personalized ranking from implicit feedback. In: Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence. pp. 452–461 (2009)
37. Rendle, S., Schmidt-Thieme, L.: Pairwise interaction tensor factorization for personalized tag recommendation. In: Proceedings of the third ACM international conference on Web search and data mining. pp. 81–90 (2010)
38. Sun, B., Zhu, Y., Xiao, Y., Xiao, R., Wei, Y.: Automatic question tagging with deep neural networks. IEEE Transactions on Learning Technologies **12**(1), 29–43 (2018)
39. Symeonidis, P., Nanopoulos, A., Manolopoulos, Y.: Tag recommendations based on tensor dimensionality reduction. In: Proceedings of the 2008 ACM conference on Recommender systems. pp. 43–50 (2008)
40. Tang, S., Yao, Y., Zhang, S., Xu, F., Gu, T., Tong, H., Yan, X., Lu, J.: An integral tag recommendation model for textual content. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33, pp. 5109–5116 (2019)
41. Vasile, F., Smirnova, E., Conneau, A.: Meta-prod2vec: Product embeddings using side-information for recommendation. In: Proceedings of the 10th ACM Conference on Recommender Systems. pp. 225–232 (2016)
42. Vinh, T.D.Q., Tay, Y., Zhang, S., Cong, G., Li, X.L.: Hyperbolic recommender systems. arXiv preprint arXiv:1809.01703 (2018)
43. Vinh Tran, L., Tay, Y., Zhang, S., Cong, G., Li, X.: Hyperml: A boosting metric learning approach in hyperbolic space for recommender systems. In: Proceedings of the 13th International Conference on Web Search and Data Mining. pp. 609–617 (2020)
44. Wang, K., Jin, Y., Wang, H., Peng, H., Wang, X.: Personalized time-aware tag recommendation. In: Thirty-Second AAAI Conference on Artificial Intelligence (2018)
45. Wang, Q., Yin, H., Wang, H., Nguyen, Q.V.H., Huang, Z., Cui, L.: Enhancing collaborative filtering with generative augmentation. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 548–556 (2019)
46. Wang, X., Zhang, Y., Shi, C.: Hyperbolic heterogeneous information network embedding. In: Proceedings of the AAAI conference on artificial intelligence. vol. 33, pp. 5337–5344 (2019)
47. Wu, Y., Yao, Y., Xu, F., Tong, H., Lu, J.: Tag2word: Using tags to generate words for content based tag recommendation. In: Proceedings of the 25th ACM international on conference on information and knowledge management. pp. 2287–2292 (2016)

48. Ying, R., He, R., Chen, K., Eksombatchai, P., Hamilton, W.L., Leskovec, J.: Graph convolutional neural networks for web-scale recommender systems. In: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 974–983 (2018)
49. Yuan, J., Jin, Y., Liu, W., Wang, X.: Attention-based neural tag recommendation. In: International Conference on Database Systems for Advanced Applications. pp. 350–365. Springer (2019)
50. Zhang, S., Yin, H., Chen, T., Hung, Q.V.N., Huang, Z., Cui, L.: Gcn-based user representation learning for unifying robust recommendation and fraudster detection. In: Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval. pp. 689–698 (2020)