

# 双曲推荐系统

Tran Dang Quang Vinh<sup>1</sup>, Yi Tay<sup>1</sup>, 张帅<sup>2</sup>, 高聪<sup>1</sup>, 李小丽<sup>3</sup>

<sup>1</sup>新加坡南洋理工大学

<sup>2</sup>澳大利亚新南威尔士大学

<sup>3</sup>新加坡A \* STAR信息通信研究所 (I2R) 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

## 摘要

许多成熟的推荐系统都是基于欧几里德空间中的表示学习。在这些模型中,诸如欧几里德距离或内积的匹配函数通常用于计算用户和项目嵌入之间的相似性得分。本文研究了在双曲空间中学习用户和项目表示的概念。在本文中,我们认为双曲空间更适合于在推荐域中学习用户项嵌入。与欧几里德空间不同,双曲空间本身具有处理层次结构的能力,受到其距离原点指数增加的属性的鼓励。我们提出HyperBPR (双曲贝叶斯个性化排名),这是一个概念上简单但非常有效的模型,可用于手头的任务。我们提出的HyperBPR不仅优于其欧几里德同行,而且在多个基准数据集上实现了最先进的性能,展示了双曲空间中个性化推荐的有效性。

## 介绍

推荐系统对网络的普遍影响是显而易见的。鉴于近年来用户普遍受到过度选择的数据增长,这种普遍存在的普遍存在是可以理解的。毕竟,交互数据(点击,购买等)是许多网络应用程序的核心,例如内容流媒体网站,电子商务等。为此,推荐系统不仅可以作为一种很好的缓解策略,还可以在Web上创建更好的用户体验。本文涉及个性化(或协作)排名的任务,其中向每个用户提供预期候选项的排序列表。

学习用户和项目对的表示形成了个性化排名问题的关键。在整个文献中,已经提出了各种各样的机器学习模型(Rendle等, 2009; Rendle 2010年; Mnih和Salakhutdinov 2008; 他等人。 2017年)。传统上采用了各种匹配功能,例如内部产品(贝叶斯个性化排名)(Rendle等。 2009年),欧几里德距离(Collaborative Metric

学习(Hsieh等人2017)和/或神经网络(他等。 2017年)。值得注意的是,一个共同点是所有这些模型都在欧几里德空间中运行,这对于交互数据来说可能是次优的。

本文研究了在双曲空间H中学习用户项表示的概念,其中距离相对于原点呈指数增长。最近,双曲线表示学习在学习实体层次结构等各种应用中表现出了巨大的希望(Nickel和Kielbaso 2017)和/或自然语言处理(Tay, Tuan和Hui 2018; Dhingra等。 2018)。类似地,我们假设非共形空间为在推荐系统中常见的交互数据提供更合适的归纳偏差。直观地,双曲空间诱导树结构(分层)嵌入空间,其本质上更适合于建模分层结构。我们表明,流行的贝叶斯个性化排序(BPR)算法的概念上简单的双曲线自适应不仅能够实现非常有竞争力的结果,而且还能够在多个个性化排名基准上优于更复杂的神经模型。

直观的是,分层结构作为推荐系统中的主要风格之一而存在。自然地,项目通常呈现分层结构(即,电影,产品倾向于遵循产品层级)。类似地,由于问题域的固有幂律性质,隐式用户交互也可能抑制分层质量。在文献的许多现有著作中已经建立了利用层次结构的概念(王等人。 2015年; 赵等人。 2017年; 王等人。 2018)。然而,这项工作是探索用于推荐系统的训练机器学习模型的分层归纳偏差的第一项工作。我们的实验表明,我们提出的模型,通过这种归纳偏差训练,导致模型的排名性能的显著改善。

双曲线距离的使用使我们的模型成为度量学习方法,尽管在双曲空间中与欧几里德空间相对。度量学习模型,例如Collaborative Metric Learning(Hsieh等人2017)合理地证明了经验上的成功。然而,有人认为它会引入不稳定性(泰,

Anh Tuan和Hui 2018)由于无法适应大型

与一组固定参数的交互次数。对此

版权所有© 2018, 人工智能促进协会

情报(www.aaii.org)。版权所有。

最后，我们认为双曲空间可以被解释为看似比欧几里德空间更大，因为标准（距离原点的距离）捕获了一些信息。由于与原点的距离增加，这导致嵌入空间具有更大程度的表示能力，而不是欧几里德空间。这加强了在双曲空间中建模用户 - 项目对的关键直觉，同时保持了CML模型的简单性和有效性。

我们的贡献总而言之，这项工作的主要贡献概括如下：

- 我们研究了双曲空间中训练推荐系统的概念，而不是欧几里德空间。我们提出了双曲贝叶斯个性化排名（HyperBPR），这是一种用于一类协同过滤（即个性化排名）的强大竞争模型。据我们所知，这是第一个探索将双曲空间用于推荐系统领域的工作。
- 我们对八个基准数据集进行了大量实验。我们提出的HyperBPR证明了双曲线空间的有效性，不仅优于欧几里德空间，而且优于一系列竞争基线。值得注意的是，HyperBPR在所有基准测试中都优于最先进的神经协同过滤（NCF）和协作度量学习（CML）模型。与竞争对手相比，我们实现了合理的性能提升，在标准排名指标方面提升了5.7%。
- 我们进行了广泛的定性和可视化实验，深入研究了我們提出的HyperBPR的内部工作原理。

## 相关工作

在推荐系统研究的丰富历史中，已经提出了无数的机器学习模型（Rendle 2010; Mnih和Salakhutdinov 2008; Rendle等人。2009年；他等人。2016年韩国2008年；他等人。2017年；Hsieh等人。2017年）。传统上，许多作品主要集中在分解交互矩阵，即矩阵分解（Mnih和Salakhutdinov 2008; Koren, 贝尔, 和Volinsky 2009），根据用户的喜好学习用户和项目的潜在因素。当然，矩阵分解的公式相当于使用内积组合用户项嵌入（他等人，2017年）。为此，（谢谢等。2017年）他认为，由于违反了三元不平等，这一公式缺乏表现力。因此，作者提出了协作度量学习（CML），这是一种基于欧几里德距离的强推荐基线。值得注意的是，许多最近的作品已经转变为神经模型（他等人，2017年；张等人。2018），其中堆叠的非线性变换

已经使用tensions来近似交互函数。

我们的工作涉及具有隐式反馈的推荐（即点击，喜欢二元性质）。在这

任务，贝叶斯个性化排名（BPR）模型（Rendle等。2009）仍然是一个强大的竞争基线。BPR在众多领域和应用程序中取得了广泛的成功（Dave等人，2018b；张等人。2016年。他和McAuley 2016b；戴夫等人。2018A）。我们的工作通过将双曲线距离作为用户和项目之间的相似度函数来训练双曲空间中的BPR模型。

我们的工作受到双曲线表示学习的最新进展的启发（Nickel和Kiela 2017；Cho等人。2018；Nickel和Kiela 2018；Ganea, Be'cigneul和Hof-曼2018年；Sala等人。2018；戴维森等人。2018）。例如，（Tay, Tuan和Hui 2018）建议在双曲空间中训练一个问答系统。（Dhingra等。2018）使用双曲线神经网络提出学习单词嵌入。（Guel'ehre等人，2018）提出了自我关注和变压器网络的双曲线变化，并将其应用于视觉问答和神经机器翻译等任务。虽然双曲面空间的优势在各种应用领域中显得突出，但是在推荐系统和隐式交互数据的上下文中没有研究这种嵌入空间的工作。这构成了我们工作的关键新颖性。关于双曲空间的详细入门在本文的技术论述中给出。

## 双曲推荐系统

本节概述了我们提出的模型的总体架构。我们架构背后的关键动机是将两个用户 - 项目对嵌入双曲线空间，然后通过成对学习最大化正用户 - 项目对和负用户 - 项目对的得分之间的差距。数字 2 描绘了整体模型架构。

## 输入编码

我们提出的模型将用户（表示为 $u_i$ ），正（观察）项（表示为 $v_j$ ）和负（未观察到）项（表示为 $v_k$ ）作为输入。每个用户和项目被表示为单热向量，其通过索引到用户/项目嵌入矩阵上映射到密集的低维向量上。然后，我们的模型利用贝叶斯个性化排名（BPR）来优化正面和负面项目之间的成对排名。

## 双曲几何和Poincare' 嵌入

双曲空间被唯一地定义为具有恒定负曲率的完整且简单连接的黎曼流形（Krioukov et al. 2010）如图所示<sup>1</sup>。事实上，只有三种类型的具有恒定曲率的黎曼流形，它们是欧几里德几何（恒定消失截面曲率），球面几何形状（恒定正截面曲率）和双曲线几何形状（恒定负截面曲率）。表格中可以找到三种几何形状的一些属性<sup>1</sup>。在本文中，我们关注欧几里德

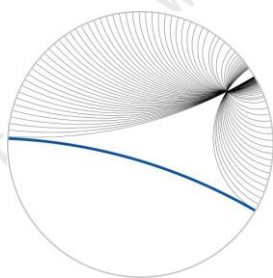
<sup>1</sup>图像拍摄于 [http://产生维基/Hyperbolic\\_geometry](http://产生维基/Hyperbolic_geometry)

属性	欧几里德	球形	夸张的
曲率K.	0	>0	<0
一条线	无限长；无界	有限长度；无界	有限长度
两条截然不同的线	不包围有限区域	包围一个有限的区域	不包围有限区域
平行线	1	0	$\infty$
三角形的角度之和	$\pi$	$>\pi$	$<\pi$
圆长	$2\pi r$	$2\pi \sin(r)$	$2\pi \sinh(r)$
磁盘区域	$\pi R^2$	$2\pi(1 - \cos(r))$	$2\pi(\cosh(r) - 1)$

表1：欧几里德，球面和双曲线几何的一些性质；其中r是半径， $E = |K|$ 。



(a) MC Eschers  
Circle Limit  
III, 1959



(b) 穿过给定点并平行于给定线的线，如Poincar磁盘模型中所示

图1：双曲空间的可视化。

由于空间扩展的关键差异，空间和双曲空间。实际上，双曲空间比欧几里德空间（多项式）更快（指数地）扩展。具体地，例如，在具有双曲半径r的恒定曲率 $K = E^2 < 0$ ,  $E > 0$ 的二维双曲空间 $H^2$ 中，我们具有：

$$\begin{aligned} L(r) &= 2\pi \sinh(Er), \\ A(r) &= 2\pi (\cosh(Er) - 1), \end{aligned} \quad (1)$$

2) 其中 $L(r)$ 是圆的长度， $A(r)$ 是圆盘的面积。因此，Eqn. (1)和(2)说明了双曲空间的指数增长/扩展 $H^2$ 相对于半径r。

尽管双曲空间不能等距地嵌入到欧几里德空间中，但是存在多个双曲几何模型，这些模型可以被表述为欧几里德空间的子集，并且非常有洞察力地依赖于不同的任务。在这些模型中，我们更喜欢由Poincaré提出的球模型（Nickel和Kiela 2017）由于其保形性（即在双曲线和欧几里德空间之间保持角度）和方便的参数化。

庞加莱球模型是黎曼流形

$P^n = (B^n, g_P)$ ，其中 $B^n = \{x \in \mathbb{R}^n : \|x\| < 1\}$ 是开放的n维单位球，其配备公制为：

$$g_P(x) = \frac{4}{(1 - \|x\|^2)^2} dx^2, \quad (3)$$

其中 $x \in B^n$ ；并且 $g_P$ 是具有 $\mathbb{R}^n$ 的组分 $I_n$ 的欧几里德度量。

P上两点之间的距离由下式给出：

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

我们采用双曲线距离函数来模拟用户和项目之间的关系。具体地，基于Eqn计算用户u和项目v之间的双曲线距离 $d_P(u, v)$ 。(4)。另外值得一提的是， $d_P(u, v)$ 有助于在庞加莱球内的距离发生变化时自动发现潜在的等级

关于你和v的规范是顺利的。值得注意的是当矢量的范数接近1时，点之间的距离呈指数增长。几何上，如果我们将树的根节点放置在“原点”，则子节点因此以它们到根的距离指数地扩展到边界。球由于上述属性。

## 学习用户项对的双曲表示

灵感来自（Gulcehre等人，2018年），然后将双曲线距离传递到称为双曲线匹配层的额外层，以匹配用户和项目对。给用户u

和一个项目v都在 $B^n$ ，我们采取：

$$\alpha(u, v) = f(d_P(u, v)), \quad (5)$$

5) 其中 $f(\cdot)$ 仅作为线性函数 $f(x) = \beta x + c$ 而优选 $\beta$ 和 $c$ 是标量参数和与网络一起学习。

## 优化和学习

本节说明了HyperBPR的优化和学习过程。

BPR三联体损失。HyperBPR利用BPR成对学习来最小化两者之间的成对排名损失。目标函数是

罚款如下：

$$\begin{aligned} \mathcal{L} = & \sum_{(i, j, k) \in D} \left[ \ln \sigma(\alpha(u_i, v_j) - \alpha(u_i, v_k)) + \lambda \right. \\ & \left. + \sum_{(l, j, k) \in D} \left[ \ln \sigma(\alpha(u_l, v_j) - \alpha(u_l, v_k)) + \lambda \right] \right] \end{aligned} \quad (6)$$

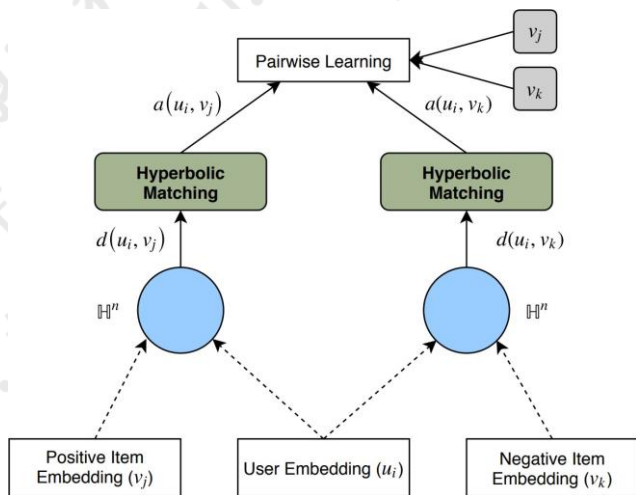


图2：我们提出的HyperBRP架构的图示。

其中  $(i, j, k)$  是属于集合的三元组，其中包含每个用户的所有正面和负面项目对； $\sigma$ 是逻辑sigmoid函数； $\Theta$ 代表模型参数； $\lambda$ 是正则化参数。

**渐变转换。** 使用RSGD学习我们模型的参数 (Bonnabel 2013)。与之类似 (Nickel和Kielia 2017)，参数更新具有以下形式：

$$\theta_{t+1} = R_{\text{tht}}(-\eta_t \nabla_{\theta} L(\theta_t)),$$

(7) 其中  $R_{\text{tht}}$  表示收缩到  $B_R(\theta_t)$  关于  $\theta$ 。  $\eta_t$  是时刻  $t$  的学习率；和  $\nabla_{\theta} L(\theta_t)$  是黎曼梯度

然后通过重新缩放  $E$  与庞加莱球尺度张量的倒数，从欧几里得梯度计算黎曼梯度  $R$ ：

$$(1 - \theta^2)^2$$

数据集	互动	#用户	#Items	%密度
服装	235,906	7,917	171,760	1.74
体育	113,119	3,740	54,744	5.53
手机	32,885	1,141	18,797	15.33
玩具和游戏	111,301	3,143	61,733	5.74
工具和家庭	64,182	2,047	35,793	8.76
汽车	34,167	1,211	26,096	10.81
庭院/草坪	10,702	374	7,293	39.24
音乐	16,501	471	12,206	28.70

$$\nabla_R = \frac{1}{4} \nabla_E. \quad (8)$$

可以参考梯度转换的细节 (Nickel和Kielia 2017; Tay, Tuan和Hui 2018)。

## 实验

### 实验装置

在本节中，我们将介绍整体实验设置。

**数据集** 对于我们的实验评估，我们采用来自亚马逊数据集的八个数据集 (他和McAuley 2016a)。选择基于促进基于数据集大小和域的多样性，其中我们确保在各个域中包含大/小数据集。数据集可以在。获得

表2：我们的实验评估中使用的所有数据集的统计数据

为了太空利益。表中报告了数据集的统计数据 2.

**评估设置和指标** 我们在协作排名 (或一类协同过滤) 设置上进行实验。我们采用命中率 (HR @ 10) 和 nDCG @ 10 (标准化折扣累积增益) 评估指标，这些指标是针对手头任务的完善的排名指标。以下 (他等人, 2017年; Tay, Anh Tuan和Hui 2018), 我们随机选择用户未与之交互的  $n$  个阴性样本，并在这些阴性样本中对基本事实进行排序。我们设置  $n = 100$ ，因为我们凭经验发现这足以探测比较基线之间相对性能的差异。对于所有数据集，用户与之交互的最后一项被保留为测试集，而倒数第二项用作验证集。在培训期间，我们会根据最佳验证分数报告模型的测试分数。

**比较基线** 在我们的实验中，我们将与五个成熟且具有竞争力的基线进行比较。

- **贝叶斯个性化排名 (BPR)** (Rendle等人, 2009) 是一个强大的协作过滤 (CF) 基线，它包含三个输入，包括用户，正项和负项。三元组目标是将正项目排名高于该用户的负项目。
- **多层感知器 (MLP)** 是一个前馈神经元

应用多层非线性来捕获用户和项目之间关系的ral网络。Fol-

HTTP://jmcauley.ucsd. EDU /数据/亚马逊/ 他们的域名被截断

降脂（他等人，2017年），我们使用具有金字塔结构的三层MLP。

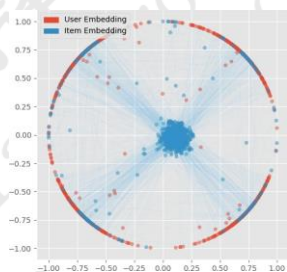
矩阵分解（MF）是推荐系统的标准基线。它使用内积来模拟用户项目表示。

神经协同过滤（NCF）（他等人，2017年）是最先进的协同过滤方法。NCF的关键思想是将MF和MLP的最后隐藏表示融合在一起成为一个联合模型。

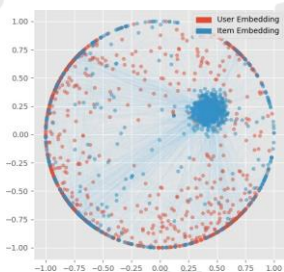
协作度量学习（CML）（Hsieh等人，2017）是一个强大的度量学习基线，使用欧几里德距离学习用户项目相似性。在我们的实验中，CML可以被认为是关键的一个消融基线，它标识了双曲线和欧几里德度量空间之间的差异。

- 
- 
-

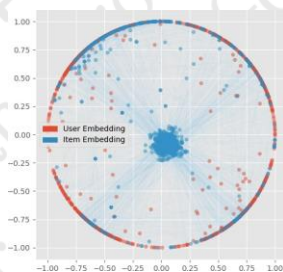




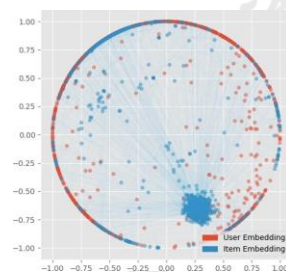
服装, 鞋帽和珠宝



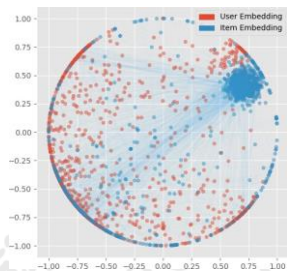
运动和户外活动



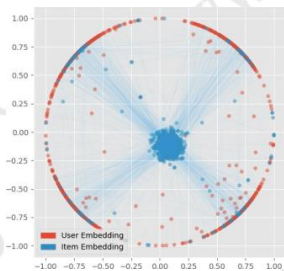
手机和配件



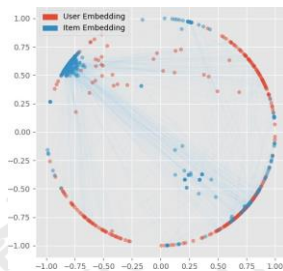
玩具和游戏



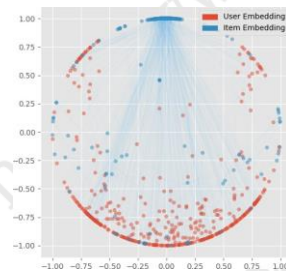
工具和家庭改进



汽车



天井, 草坪和花园



乐器

图3: Poincare磁盘中8个亚马逊数据集的二维双曲线嵌入。 图像说明了收敛后用户和项目对的嵌入。

实现细节我们在Tensorflow中实现所有模型。所有模型都使用Adam训练(金玛和巴2014)学习率在0.001, 0.005, 0.01之间调整。所有型号的嵌入尺寸d在50, 100, 150之间调整, 并有选择地设置为100。批次B的数量在10, 100, 1000之间调整。对于优化铰链损耗的模型, 边距 $\lambda$ 在0.1, 0.2, 0.5之间调整。NCF和MLP模型是在配置和架构之后实现的(他等人, 2017年); 但是, 为了公平比较, 预训练的MF和MLP不适用于NCF。使用高斯分布随机初始化所有嵌入和参数, 均值为0, 标准偏差为

0.01。对于大多数数据集和基线, 我们根据经验设置超参数, 学习率为 $10^{-3}$ , 批次数为10, 嵌入大小为100, 保证金设定为0.1。

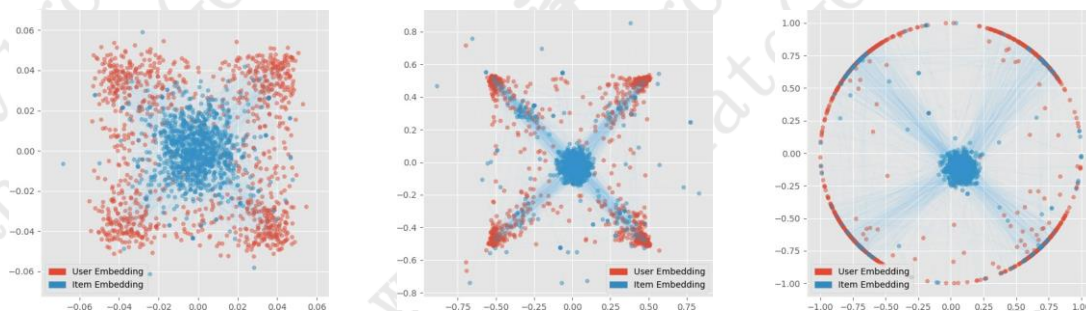
## 实验结果

本节实验性地介绍了我们在所有数据集上的结果。对于所有获得的结果, 最好的结果是粗体, 而第二个最好的结果是下划线。如表中所述3, 我们提出的模型明显优于所有数据集上HR @ 10和nDCG @ 10指标的所有基线。

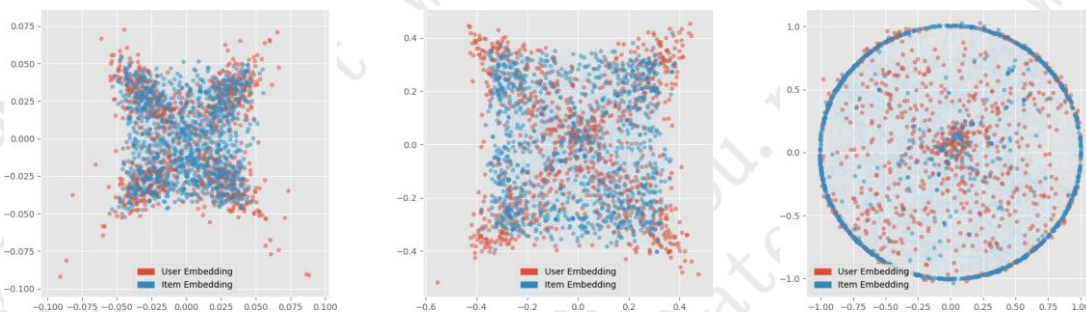
与基线有关, CML在大多数数据集中优于其他基线。我们观察到MF和CML的性能极具竞争力, 即MF和CML在整个过程中始终如一地取得了良好的效果

数据集。CML对数据集的性能增益约为1%-2%。值得注意的是, MF的性能比Patio数据集上的CML要好得多。一个可能的原因是对于具有高密度的小数据集(例如, 密度为39.24%的Patio), 诸如MF的简单模型应被视为优先选择。此外, 在大多数情况下, NCF的性能通常仅与香草MLP和MF相当。解释是因为使用了双嵌入空间(因为NCF结合了MLP和MF), 如果数据集不够大, 这种用法可能会导致过度拟合(Tay, Anh Tuan和Hui 2018)。

值得注意的是, 我们提出的HyperBPR模型明显优于最佳基线方法。nDCG在八个数据集上的百分比改进(与表中报告顺序相同) 3) 是+ 3.39%, + 2.50%, + 2.83%, + 5.54%, + 2.00%, + 3.76%, + 5.72%和+ 2.45% respectively。我们还观察到命中率(HR @ 10)的类似高性能增益。请注意, 由于其丰富详细的类别层次结构, 亚马逊数据集遵循幂律分布(McAuley等人, 2015年)。因此, 它使我们能够在双曲线空间中实现我们提出的HyperBPR与其他强欧几里德基线相比具有非常有竞争力的结果。非正式地, 由于树需要一个指数空间用于分支, 其中只有双曲线几何具有这种特性, 树更喜欢嵌入双曲空间而不是欧几里德空间。换句话说, 树可以被认为是离散的双曲空间(Krioukov et al. 2010)。我们的实验证据显示了非凡的

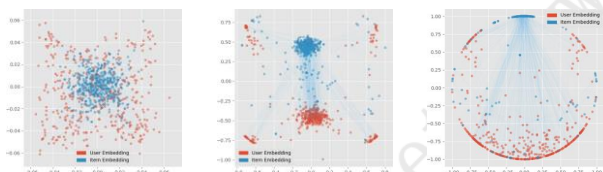


(a). HyperBPR的中间嵌入在10个时期之后，100个时期和收敛后的嵌入。



(b). 在10个时期，100个时期和收敛后的嵌入之后中间嵌入CML。

图4：二维庞加莱嵌入和欧几里德嵌入在汽车数据集上的比较。 图像说明了在10个时期，100个时期和收敛后的嵌入之后HyperBPR和CML的中间嵌入。



(a) 10点以后  
时代之后

(b) 100后  
时代

(c) 在  
收敛

图5：音乐数据集上的用户/项目嵌入相对于时期数量的转换。

我们提出的HyperBPR模型对各种数据集的推荐结果以及欧几里德空间上双曲空间处理分层数据结构的优势。

## 定性分析

本节研究我们提出的模型的定性分析，以了解嵌入双曲空间的行为。

**双曲线收敛图** 3 表示在收敛后的8个亚马逊数据集的测试集上的二维双曲线嵌入。 我们观察到项目嵌入在用户嵌入上形成一个球体。 此外，由于我们对测试集进行了分析，因此可视化用户/项目嵌入图 3 演示了HyperBPR自我组织和自动检测用户/项目嵌入中的层次结构的能力，类似于 (Tay, Anh Tuan和Hui 2018)。

另外，我们观察到较小的数据集较高

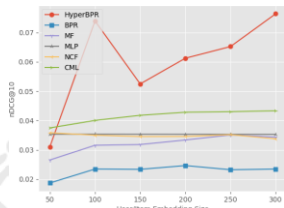
诸如Patio和Musical之类的密度倾向于迫使用户 - 项目对嵌入到收敛的球的边界。 我们以Musical数据集为例，可视化嵌入的转换。 数字 5 图示了关于时期数量的用户/项目嵌入变换。 显然，在前100个时期之后，用户和项目嵌入可能会聚并在嵌入之间形成链接对。 在收敛时，嵌入被推向边界，这也在数据集中给出了没有分层结构的标志。

**收敛比较图** 4 说明二维庞加莱嵌入之间的比较 (Hy-

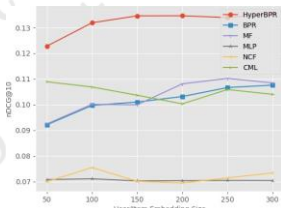


	业务流程再造					多层感知器 (MLP) 慢性粒细胞白血病		今天		NCF		HyperBPR	
	人力资源 源		nDCG		人力资源	nDCG	人力资源	nDCG	人力资源	nDCG	人力资源	nDCG	人力资
服装	0.039	0.024	0.058	0.035	0.051	0.032	0.059	0.035	<u>0.066</u>	<u>0.040</u>	<b>0.120</b>	<b>0.074</b>	
体育	0.149	0.100	0.120	0.071	0.148	0.100	0.118	0.076	<u>0.159</u>	<u>0.107</u>	<b>0.193</b>	<b>0.132</b>	
手机玩具和游	0.186	0.128	0.147	0.092	0.200	<u>0.130</u>	0.157	0.101	<u>0.203</u>	0.127	<b>0.243</b>	<b>0.158</b>	
戏工具和家庭	0.274	0.209	0.255	0.178	0.288	<u>0.216</u>	0.236	0.167	<u>0.292</u>	0.212	<b>0.360</b>	<b>0.272</b>	
汽车露台	0.139	0.095	0.134	0.087	0.161	<u>0.115</u>	0.146	0.086	<u>0.167</u>	0.112	<b>0.198</b>	<b>0.135</b>	
音乐	0.034	0.023	0.047	0.030	0.048	0.031	0.048	0.030	<u>0.059</u>	<u>0.037</u>	<b>0.121</b>	<b>0.074</b>	
	0.175	0.116	0.164	0.102	<u>0.208</u>	<u>0.126</u>	0.151	0.092	0.156	0.099	<b>0.290</b>	<b>0.183</b>	
	0.055	0.037	0.037	0.018	0.055	0.033	0.050	0.023	<u>0.059</u>	<u>0.043</u>	<b>0.116</b>	<b>0.068</b>	

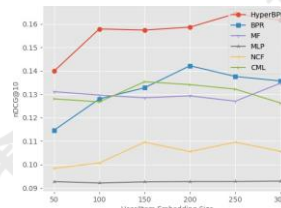
表3: 8个亚马逊数据集的实验结果。我们提出的HyperBPR取得了非常有竞争力的结果, 优于强大的欧几里德基线, 如CML和BPR。



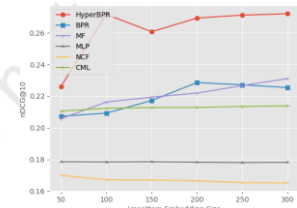
服装, 鞋帽和珠宝



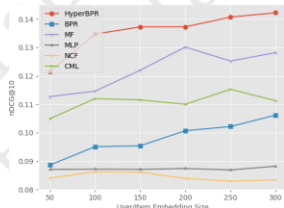
运动和户外活动



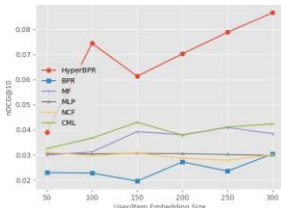
手机和配件



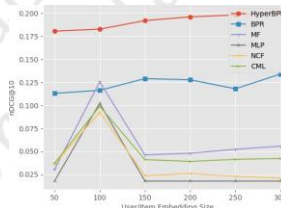
玩具和游戏



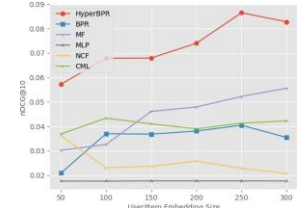
工具和家庭改进



汽车



天井, 草坪和花园



乐器图6: 嵌入大

小对8个亚马逊数据集的影响。

perBPR) 和Euclidean嵌入 (CML) 在Automotive数据集上。对于CML, 我们决定剪切范数, 即嵌入的范数被约束为1, 以进行类似的比较。

乍一看, 我们通过观察用户和项目嵌入在有关时期数量的空间中的分布, 注意到两种嵌入类型之间的差异。虽然随着时期数量的增加, HyperBPR项目嵌入逐渐聚集, 但CML的项目嵌入具有相反的运动。原因是因为CML的学习度量将正项拉近, 同时将负项进一步分开; 因此, 项目嵌入被推向边界。此外, CML的收敛性没有显示层次结构, 这与HyperBPR相比是一个缺陷。

## 嵌入尺寸的影响

在本节中, 我们将研究嵌入大小对我们提出的模型和基线性能的影响。数字 6 表示嵌入大小d的效果

$\in \{50, 100, 150, 200, 250, 300\}$  在 dCG @ 10 的 8 个亚马逊数据集上 d 通常, 我们观察到无论嵌入大小如何, HyperBPR总是明显优于基线。虽然NCF在整个嵌入尺寸中保持稳定的性能, 但其他基线的性能似乎略有波动。此外, 我们注意到HyperBPR的nDCG @ 10仅在d = 150时略有下降, 但随着嵌入大小d的增加, 仍然保持了极好的性能。

## 结论

在本文中, 我们介绍了一种新的有效且称职的推荐模型, 称为HyperBPR。据我们所知, HyperBPR是第一个探索推荐系统中双曲线空间的模型。通过对8个数据集的大量实验, 我们能够证明HyperBPR相对于欧几里德空间中其他基线的有效性, 甚至是最先进的模型, 如CML或NCF。HyperBPR的有希望的结果可能激发其他未来的工作, 以探索解决推荐问题的双曲空间。



## 参考

- [Bonnabel 2013] Bonnabel, S. 2013. 黎曼流形上的随机梯度下降。IEEE Trans. 自动售货机。对照。58 (9): 2217-2229。
- [Cho等人。2018] Cho, H.; Demeo, B.; 彭, J.; 和 Berger, B. 2018. 双曲空间中的大边分类。CoRR abs / 1806.00437。
- [戴夫等人。2018a] 戴夫, VS; 张, B.; Al Hasan, M.; AlJadda, K.; 和Korayem, M. 2018a. 结合代表性学习方法, 以获得更好的工作和技能推荐。
- [戴夫等人。2018b] 戴夫, VS; 张, B.; 陈, P.-Y.; 和哈桑, MA 2018b. Neural-brane: 归因于网络嵌入的神经贝叶斯个性化排名。arXiv preprint arXiv: 1804.08774。
- [戴维森等人。2018年] Davidson, TR; Falorsi, L.; 曹, ND; Kipf, T.; 和Tomczak, JM 2018. 超球面变分自动编码器。CoRR abs / 1804.00891。
- [Dhingra等。2018] Dhingra, B.; Shallue, CJ; Norouzi, M.; 戴上午; 和Dahl, GE 2018. 在双曲空间中嵌入文本。arXiv preprint arXiv: 1806.04313。
- [Ganea, Be'cigneul和Hofmann 2018] Ganea, O.; B'cigneul, G.; 和Hofmann, T. 2018. 用于学习分层嵌入的双曲线蕴涵。在第35届国际机器学习会议论文集, ICML 2018, Stockholmsmassan, Stockholm, Sweden, July 10-15, 2018, 1632-1641。
- [Gu'lc ehre等。2018] Gu'lc ehre, C.; Denil, M.; Malinowski, M.; Razavi, A.; Pascanu, R.; 赫尔曼, KM; Battaglia, P.; Bapst, V.; Raposo, D.; Santoro, A.; 和de Freitas, N. 2018. 双曲线注意网络。CoRR abs / 1805.09786。
- [他和McAuley 2016a] He, R. 和McAuley, J. 2016a. 起伏: 用一流的协同过滤模拟时尚潮流的视觉演变。在第25届万维网国际会议论文集, 2016年WWW, 加拿大蒙特利尔, 2016年4月11日至15日, 507-517。
- [他和McAuley 2016b] He, R. 和McAuley, J. 2016b. Vbpr: 隐式反馈的视觉贝叶斯个性化排名。
- [他等。2016] 他, X.; 张, H.; Kan, M.-Y.; 和蔡, T.-S. 2016. 具有隐式反馈的在线推荐的快速矩阵分解。在第39届国际ACM SIGIR信息检索研究与发展会议论文集中, 549-558。ACM。
- [他等。2017] 他, X.; 廖, L.; 张, H.; Nie, L.; 胡, X. 和Chua, T. 2017. 神经协同过滤。在第26届万维网国际会议论文集, 2017年WWW, 澳大利亚珀斯, 2017年4月3-7日, 173-182。
- [Hsieh等人。2017] Hsieh, C.; 杨, L.; 崔, Y.; 林, T.; 和Belongie, S.J.; 和Estrin, D. 2017年. 协作度量学习。在第26届万维网国际会议论文集, 2017年WWW, 澳大利亚珀斯, 2017年4月3日至7日, 193-201。
- [Kingma and Ba 2014] Kingma, DP和Ba, J. 2014. Adam: 一种随机优化方法。arXiv preprint arXiv: 1412.6980。
- [Koren, Bell和Volinsky, 2009] Koren, Y.; Bell, R.; 和沃林斯基, C. 2009. 推荐系统的矩阵分解技术。计算机 (8): 30-37。
- [Koren 2008] Koren, Y. 2008. 分解符合邻域: 多方面的协作过滤模型。在第14届ACM SIGKDD知识发现和数据挖掘国际会议论文集中, 426-434。ACM。
- [Krioukov等。2010] Krioukov, DV; Papadopoulos, F.; Kitsak, M.; Vahdat, A.; 和Bogun', M. 2010. 复杂网络的双曲几何。CoRR abs / 1006.5169。
- [McAuley等人。2015年] McAuley, JJ; Targett, C.; 施, Q.; 和van den Hengel, A. 2015. 基于图像的风格和替代品建议。在第38届国际ACM SI-GIR信息检索研究与发展会议论文集, 智利圣地亚哥, 2015年8月9日至13日, 第43-52页。
- [Mnih and Salakhutdinov 2008] Mnih, A. 和Salakhutdinov, R. R. 2008. 概率矩阵分解。在神经信息处理系统的进展中, 1257-1264。
- [Nickel 和 Kiela 2017] Nickel, M. 和 Kiela, D. 2017. Poincare' 嵌入式用于学习分层表示。在盖恩, 我。Luxburg, UV; Bengio, S.; Wallach, H.; Fergus, R.; Vishwanathan, S.; 和Garnett, R., eds., Advances in Neural Information Processing Systems 30. 6338-6347。
- [Nickel和Kiela 2018] Nickel, M. 和Kiela, D. 2018. 在双曲线几何的洛伦兹模型中学习连续的层次结构。在第35届国际机器学习会议论文集, ICML 2018, Stockholmsmassan, Stockholm, Sweden, July 10-15, 2018, 3776-3785。
- [Rendle等人。2009年] Rendle, S.; Freudenthaler, C.; Gantner, Z.; 和Schmidt-Thieme, L. 2009. BPR: 来自隐式反馈的贝叶斯个性化排名。在UAI 2009, 第20届人工智能不确定性会议论文集, 蒙特利尔, QC, 加拿大, 2009年6月18日至21日, 452-461。
- [Rendle 2010] Rendle, S. 2010. 分解机器。在数据挖掘 (ICDM), 2010年IEEE第10届国际会议上, 995-1000。IEEE。
- [Sala et al. 2018年] Sala, F.; Sa, CD; 顾, A.; 和Re', C. 2018. 双曲线嵌入的表示权衡。在第35届国际机器学习会议论文集, ICML 2018, Stockholmsmassan, 瑞典斯德哥尔摩, 2018年7月10日至15日, 4457-4466。
- [Tay, Anh Tuan和Hui 2018] Tay, Y.; Anh Tuan, L.; 和慧, S. C. 2018. 通过基于记忆的注意力进行协作排序的潜在关系度量学习。在2018年万维网会议论文集中, WWW' 18, 729-739。瑞士日内瓦共和国: 国际万维网会议指导委员会。
- [Tay, Tuan和Hui 2018] Tay, Y.; Tuan, LA; 和Hui, SC 2018. 用于快速有效的神经问题回答的双曲线表示学习。在第十一届ACM国际网络搜索和数据挖掘会议论文集中, WSDM '18。
- [王等人。2015] Wang, S.; 唐, J.; 王, Y. 和Liu, H. 2015. 探索推荐系统的隐式分层结构。在第24届国际人工智能会议论文集, 1813-1819。AAAI出版社。
- [王等人。2018] Wang, S.; 唐, J.; 王, Y. 和Liu, H. 2018. 探索推荐系统的层次结构。第30卷, 1022-1035。IEEE。
- [张等人。2016] 张, B.; Choudhury, S.; Hasan, MA; 宁, X. Agarwal, K.; Purohit, S.; 和Cabrera, PP 2016. 过去的信任: 知识图中基于贝叶斯个性化排名的链接预测。arXiv preprint arXiv: 1601.03778。
- [张等人。2018年] 张, S. 姚, L.; 太阳, A.; 王, S. 长, G.; 和Dong, M. 2018. Neurec: 关于个性化排名的非线性转换。arXiv preprint arXiv: 1805.03002。
- [赵等人。2017] Zhao, P.; 徐, X. 刘, Y. 周, Z.; 郑, K. 盛, VS; 和Xiong, H. 2017. 利用等级结构 -

poi推荐。 2017年IEEE国际数据挖掘大会 (ICDM) , 655-664。  
IEEE。