

双曲空间互学习捆绑推荐模型 HyperMLBR

刘婕^a, 李琳^{b*}, 柯豪乐^b, 谢伟平^a

(武汉理工大学 a. 土木工程与建筑学院; b. 计算机与人工智能学院, 武汉 430070)

摘要 在许多工程与商务应用中,项目与商品捆绑是一种广泛采用的推荐策略.但是现有方法对捆绑推荐中交互数据的特征认识存在不足,而且欧氏空间的互学习方法难以在双曲空间中充分发挥视图之间互补优势,因此提出了一种双曲空间互学习的捆绑推荐模型(HyperMLBR).该模型设计了图掩码模块和互学习模块,利用掩码自编码器和在线知识蒸馏技术,促使两个视图在双曲空间中进行有效的协作建模.最后 HyperMLBR 模型与其他模型在 Youshu 和 NetEase 数据集上进行性能分析、消融实验和互学习效果分析,实验结果表明:HyperMLBR 模型具有有效性和动机的正确性.

关键词 捆绑推荐;交互数据;双曲空间;互学习

中图分类号 TP391 文献标志码 A 文章编号 1672-4321(2025)05-0654-10

doi:10.20056/j.cnki.ZNMDZK.20250511

HyperMLBR: Hyperbolic space mutual learning bundle recommendation model

LIU Jie^a, LI Lin^{b*}, KE Haole^b, XIE Weiping^a

(Wuhan University of Technology, a. School of Civil Engineering and Architecture; b. School of Computer Science and Artificial Intelligence, Wuhan 430070, China)

Abstract Project and commodity bundling is a widely used recommendation strategy in many engineering and business applications. However, the existing methods have insufficient understanding of the characteristics of interaction data in bundle recommendation, and the Euclidean mutual learning methods fail to fully exploit the complementary advantages between views in hyperbolic space. Therefore, a hyperbolic space mutual learning bundle recommendation model (HyperMLBR) is proposed. The model designs a graph masking module and a mutual learning module, utilizing masked autoencoder and online knowledge distillation technique to promote effective collaborative modeling of two views in hyperbolic space. HyperMLBR model and other models were analyzed in the Youshu and NetEase data sets for performance analysis, ablation experiment and mutual learning effect analysis. The experimental results demonstrate that the HyperMLBR model is effective and its motivation is correct.

Keywords bundle recommendation; interactive data; hyperbolic space; mutual learning

近年来,捆绑推荐受到了越来越多的关注,捆绑推荐通过学习用户的历史交互,直接向用户推荐一个包含相似项目的捆绑包.捆绑推荐的目标是通过学习用户的历史交互信息,准确预测用户与捆绑包互动的概率.大多数现有的方法都是从用户-捆绑包交互图(即捆绑包视图)和用户-项目-捆绑包交互

图(即项目视图)中学习用户和捆绑包嵌入^[1-4].模型从这两个视图学到的知识是不同的,且模型在这两个视图上的学习趋势也不同.因此,两个视图对用户偏好的不同学习趋势将导致不同的推荐结果.那么如何使两个视图可以学习各自推荐结果中包含的知识将是进一步提高模型推荐性能的关键.然

收稿日期 2025-01-14

* 通信作者 李琳(1977-),女,教授,博士,研究方向:信息检索与推荐系统、数据挖掘与知识管理、多模态数据机器学习,
E-mail: cathylin@whut.edu.cn

基金项目 国家自然科学基金资助项目(62276196)

而,这种重要的协作关系在现有的工作^[1-3]中没有得到很好的利用,对两个视图的松散建模难以保证知识传递的效果^[5].最近的工作 MIDGN^[6] (Multi-view Intent Disentangle Graph Networks) 和 CL²BRec^[7] (Cross-level relational graph Contrastive Learning for Bundle Recommendation) 意识到了该问题,并利用对比学习框架改进从不同视图学到的表征.然而,将两个视图中具有不同树状程度的交互图均匀嵌入到欧几里得空间的方式存在着严重的失真问题^[8-9].嵌入本身的不准确会在视图协作建模的过程中引入更多负面的噪声,影响互学习的效果.本文认为对于用户与项目之间交互关系的利用程度还可以进一步提高,项目视图更注重捆绑包与项目之间相似性. Attlist^[2] (Hierarchical Self-attentive Recommendation Model) 和 GRAM-SMOT^[10] (GRaph Attention Mechanism and SubModular OpTimization) 引入了注意力机制去建模同一个捆绑包中不同项目的重要性.这种方式是利用了捆绑包和项目之间的附属关系,没有充分考虑到用户-项目交互图中不同数据对捆绑包预测贡献程度的差异.

针对上述问题,本文提出了一种基于双曲空间的互学习捆绑推荐模型(Hyperbolic Space Mutual Learning Bundle Recommendation Model, HyperMLBR). HyperMLBR 主要有以下两个创新点,一是针对没有充分考虑用户-项目交互图中不同数据对捆绑包预测贡献程度具有差异的问题,本文将用户和项目的交互划分为两个不同的集合,分别对应重要交互和次要交互.利用图掩码模块对重要交互集合中的边进行不同概率的遮盖和恢复,从而加强模型对重要交互的学习力度,帮助项目视图得到更好的实体表征.二是针对视图之间学习倾向不同的问题,本文采用在线蒸馏的方式,在学习到相对准确的实体嵌入之后,通过互学习框架使两个视图不断提取软标签中包含的互补知识.而且,为了更好地结合双曲空间和相互学习的优势,本文还提出了基于双曲距离的互学习损失函数,以便两个视图更好地传递知识.

1 相关研究

1.1 捆绑推荐方法

比起项目推荐,捆绑推荐有着交互类型复杂,数据更加稀疏等特点,技术难度更高.现有研究大

致可以分为两类:基于深度学习的方法^[1-2]和基于图的方法^[3,6-7].典型的传统深度学习方法就是依靠深度神经网络,通过共享底层参数层或者利用注意力机制的方式去学习实体的表征.CHEN等^[1]提出了一个名为 DAM (Deep Attentive Multi-Task Model) 的模型,采用了分解注意力网络来聚合捆绑包中的项目,以此表示捆绑包的特征.HE等^[2]提出了一种自注意力聚合层来捕获用户和捆绑包特征的一致性,并以此为权重构建用户和捆绑包的实体表征.HE等^[11]通过一致性感知门控网络,并结合两个基于注意力的用户偏好模型,共同捕获用户对捆绑包的偏好.

但是,这些方法难以利用推荐系统数据中潜在的结构特征,对捆绑推荐中复杂的实体关系建模较为松散.后来,图神经网络开始逐渐在推荐领域被使用^[12-14],图神经网络对于图数据强大的建模能力,使得模型可以更好地从用户的交互图中学习到用户的偏好,极大地提高了推荐的准确性.BGCN^[3] (Bundle Graph Convolutional Network) 是图神经网络在捆绑推荐领域的经典应用.BGCN 将整个交互图划分为两个不同的级别:捆绑包级别和项目级别.然后分别在两个级别内进行图卷积操作以捕获用户的不同偏好.DENG等^[15]提出一个 BundleNet (Bundle Recommendation Network) 框架,该框架基于历史数据构建的用户-项目-捆绑包三方交互图,并将捆绑推荐问题形式化为三方图中的连接预测问题.VIJAIKUMAR等^[10]提出了基于图注意力机制的模型 GRAM-SMOT 和一个基于度量学习的损失函数.这种损失函数可以将与用户和捆绑包交互的项目拉近,并将不相关的项目相对于给定的边距推得更远.ZHAO等^[6]提出了一种名为多视图意图分离图神经网络模型 MIDGN,该模型能够以更细的粒度精确、全面地捕捉用户意图和项目关联的多样性.WANG等^[7]提出了一种跨级关系图对比学习的捆绑推荐模型 CL²BRec.该方法将用户-项目和用户-捆绑包的交互图视为局部交互关系视图,将捆绑包-项目图视为局部依赖关系视图,将完整的用户-捆绑包-项目视为全局结构视图,最终构建了一种跨级别捆绑推荐模型.

1.2 非欧式空间的应用

在人工智能的诸多领域,数据常会在欧氏空间中进行表示,并且各领域都在欧氏空间的基础上取得显著成就,例如自然语言处理领域的 Transformer^[16]、图像分类领域的 ResNet^[17]、推荐系统

领域的神经协同过滤^[18]等.但是,传统的基于向量内积公式的相似度计算方式在某些情况下存在违反三角不等式的问题^[19],而且欧氏空间在嵌入某些类型的数据(例如树状数据、无标度数据或球形数据)时会遭受较大的失真^[20].在这种情况下,非欧氏空间可以为相应数据提供更好的学习基础.非欧氏空间根据曲率的不同可以划分为双曲空间(曲率小于0)和椭圆空间(曲率大于0)^[21].两个空间有着各自的几何特性和优势.

近年来,学者们发现推荐系统中的用户交互数据大多呈现出长尾效应,其分布呈现为一条长尾向右下方延伸的曲线,即幂律分布.KE等^[5]通过计算Youshu和NetEase两个数据集的双曲度(Gromov's δ -hyperbolicity),验证了捆绑推荐中的交互图具有幂律分布特性.这种现象在推荐系统中常称之为长尾效应,即小部分项目受到了大部分的关注和喜欢,而大部分项目仅受到少数用户的关注.然而,欧氏空间难以在保持低失真的同时嵌入具有上述特征的数据^[22-23].双曲空间和双曲嵌入为解决这个问题提供了一个有希望的方向.双曲空间有着指数级的空间增长速度,恰好于幂律分布的数据相互匹配.同时,其强大的分层数据建模能力也适用于树状结构的数据.许多学者开始尝试在双曲空间中解决传统的项目推荐任务.HGCF^[24](Hyperbolic Graph Collaborative Filtering)利用双曲空间的优势开发了一种新颖的项目推荐模型,可以在双曲空间中有效地学习用户和项目的嵌入.LGCF^[25](Lorentz Graph Collaborative Filtering)提出了一种模型,其中模型的所有操作都在双曲空间中完成,以减少空间映射操作带来的负面影响.然而将双曲空间应用于捆绑推荐的工作却很少.HyperMBR^[5](Hyperbolic Mutual Learning model for Bundle Recommendation)将双曲图卷积与互学习相结合用于捆绑推荐,在双曲空间中准确地学习了用户嵌入,但是捆绑包视图和项目视图仍相互隔离且独立训练,对于项目视图中交互图的不同边也没有进行区分.

2 任务描述

本文分别定义 $U=\{u_1, u_2, \dots, u_M\}$, $B=\{b_1, b_2, \dots, b_N\}$ 且 $I=\{i_1, i_2, \dots, i_J\}$ 为用户、捆绑包和单个项目的集合,其中 M 、 N 和 J 分别表示三个实体集的元素数量.根据用户的历史交互记录和捆绑包构成信息,定义

了两个交互矩阵 $X_{M \times N} = \{x_{ub} | u \in U, b \in B\}$ (用户-捆绑包), $Y_{M \times J} = \{y_{ui} | u \in U, i \in I\}$ (用户-项目),以及捆绑包和项目之间的附属关系矩阵 $Z_{N \times J} = \{z_{bi} | b \in B, i \in I\}$,其中 $x_{ub} = 1$ 或 $y_{ui} = 1$ 表示用户 u 与捆绑包 b 或项目 i 发生过交互(点击、评论、收藏等).如果 $z_{bi} = 1$ 表示捆绑包 b 包含项目 i .根据上述定义,捆绑推荐任务可以被描述为如下形式:

输入:给定用户与项目交互矩阵 X ,用户与捆绑包交互矩阵 Y 以及捆绑包与单个项目的附属关系矩阵 Z .

输出:用户 u 最有可能发生交互的 K 个捆绑包组成的推荐列表.

3 HyperMLBR 模型设计

3.1 HyperMLBR 模型整体框架

HyperMLBR模型包含两个不同的视图以及三种不同的损失函数,即边际排名损失函数、互学习损失函数和图掩码预测损失函数,整体框架如图1所示.两个视图中编码模型采用的是HyperMBR^[5]模型,确保两个视图本身学习到的嵌入更加准确,以减少知识传递过程中产生的误导.边际损失函数作为有监督的损失函数,每个视图都会受到它的指导,尽可能往正确的方向优化,不会随意受到另一个视图的影响.通过监督学习,两个视图会逐渐对测试集中的每一个用例给出自己预测的标签.同时,图掩码预测损失函数则会指导项目视图的编码器更加注重有效交互的学习.

由于两个视图的输入和嵌入初始化都不同,最终学习到的表征也不同,因此它们对同一个用例给出的预测值也会有所不同.根据另一个视图的预测找出每一个测试用例其他最有可能的结果,可以增加自身结果的后验熵,从而帮助模型收敛到更鲁棒的结果.因此本文进一步引入了互学习损失函数,两个视图视作为有着共同目标的同级模型,在向另一个视图传递知识的同时,也在接受另一个视图的知识.最后,三个损失函数以不同的权重组成最终的损失函数,共同优化模型.

3.2 边际排名模块设计

本文采用边际排名损失^[5]来优化模型参数.该损失函数分为两个部分,首先需要利用负样本集 N 来估计正样本 j 的近似排名.其中负样本集 N 是在训练之前随机采样生成的.近似排名估计的具体公式

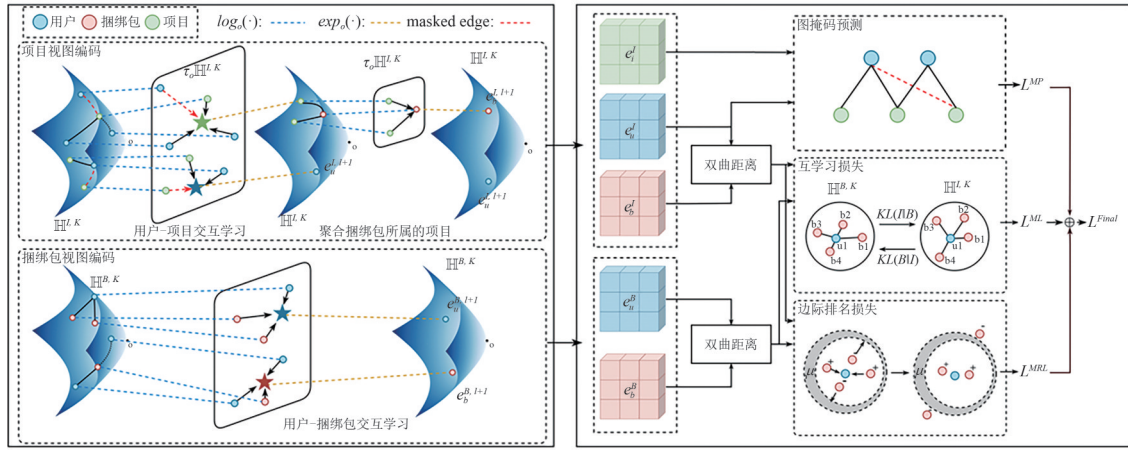


图 1 HyperMLBR 整体框架图

Fig. 1 HyperMLBR overall framework

表示如下:

$$\text{rank}(i, j) \approx r_{i,j} = \sum_{k \in N} |\mu + d_{\mathbb{H}}^2(u_i, b_j) - d_{\mathbb{H}}^2(u_i, b_k)|_+, \quad (1)$$

其中 $d_{\mathbb{H}}(u, b_k)$ 是用户 i 和捆绑包 k 之间的双曲距离, $|\cdot|_+ = \max(0, \cdot)$ 表示取非负部分, 即近似排名的最小值就是 0. $\mu \in \mathbb{R}^+$ 是安全边际参数, 表示需要将负样本推离的最小距离. 最后, 根据每个样本的近似排名, 边缘排名损失的计算定义如下:

$$L_{i,j}^{\text{MRL}} = \log(1 + r_{i,j}). \quad (2)$$

3.3 图掩码模块设计

项目视图更注重捆绑包与项目之间相似性, 项目是目标用户与可能喜欢的捆绑包之间的桥梁. 基于此, 用户-项目交互图中的边可以根据用户交互过的项目是否被用户交互过的捆绑包所包含分为两类. 这两类交互可被定义如下:

$$\mathcal{E}_1 = \{(u, i) \mid Y_{ui} = 1, XZ_{ui} > 0\}, \quad (3)$$

$$\mathcal{E}_2 = \{(u, i) \mid Y_{ui} = 1, XZ_{ui} = 0\}, \quad (4)$$

其中 \mathcal{E}_1 表示用户-项目交互图中更重要的边, 这些边为捆绑包的预测做出了更多的贡献; \mathcal{E}_2 表示用户-项目交互图中次重要的边. 受计算机视觉领域中掩码自编码器的启发^[26], 本文在部分被屏蔽的用户-项目图上进行信息传播, 然后利用图掩码预测损失函数指导编码器重建显著缺失的信息, 并进一步推测潜在的有效交互.

首先, 最简单遮掩边的方式是固定遮掩概率, 对 \mathcal{E}_1 中的每一条边进行等概率的遮盖. 然而, Youshu 和 NetEase 数据集存在明显的长尾效应, 每个实体的交互密度变化非常明显. 所以合理的方式是根据边两端节点的度来计算该节点的交互密度, 并根据交互密度设定该条边的遮掩概率. 对于 \mathcal{E}_1 中的每一条边 (u, i) 计算对应两端实体的交互密度:

$$\text{density}(u, i) = \frac{\text{deg}(u) + \text{deg}(i)}{2}, \quad (5)$$

其中 $\text{deg}(u)$ 表示用户 u 在用户-项目交互图中的度. 为了缓解长尾效应带来的度分布差距过大的问题, 本文设定 $s_{ui} = \text{density}(u, i)$. 基于此, 设定每条边的遮掩概率如下:

$$p_{ui} = p_l + \frac{s_{ui} - s_{\min}}{s_{\max} - s_{\min}} (p_u - p_l), \quad (6)$$

其中, p_l 和 p_u 是超参数, 分别是遮掩概率的下限和上限. 式 6 保证了一条边两端的实体交互密度越高, 这条边被遮掩的概率就越高. 定义被遮掩边的集合为 $\mathcal{E}_{\text{mask}}$, 经过双曲图卷积模块的信息传播之后, 可以得到用户和项目位于双曲空间的实体表征 e_u^I, e_i^I . 可以利用这两个实体表征去预测被遮掩的边, 并构建损失函数, 损失函数如下:

$$L^{MP} = \sum_{(u, i) \in \mathcal{E}_{\text{mask}}} \log(1 + \sum_{k \in N} |\mu + d_{\mathbb{H}}^2(u, i) - d_{\mathbb{H}}^2(u, k)|_+), \quad (7)$$

其中 $d_{\mathbb{H}}(u, i)$ 表示 u 和 i 嵌入之间的双曲距离.

3.4 互学习模块设计

捆绑推荐中两个视图有着截然不同的学习倾向, 而且存在着结果互补的可能. HyperMBR^[5] 获得了相对准确的实体嵌入, 但是两个视图之间的学习过程仍然互相隔离, 无法从另一个视图中学习到有益的知识. 利用跨视图的互学习方法去促进两个视图之间的相互增强, 同时考虑到传统的基于向量内积的优化方式不利于嵌入在双曲空间的优化, 本文提出的模型并没有采用常规的基于概率的软标签作为知识蒸馏学习的目标, 而是使用了一个更适合双曲空间的知识蒸馏损失.

在获得用户和捆绑包的嵌入 (例如 $e_u^B, e_b^B, e_u^I, e_b^I$)

之后,针对两个双曲空间 \mathbb{H}^a 和 \mathbb{H}^b ,使用式 8^[5]计算嵌入之间的双曲距离.

$$d_{\mathbb{H}}^K(x, y) = \sqrt{K} \operatorname{arcosh}(-\langle x, y \rangle_{\mathbb{H}} / K), \quad (8)$$

其中 $d_{\mathbb{H}}^K(x, y)$ 表示的是 x 和 y 之间的双曲距离,而 $\langle \cdot, \cdot \rangle_{\mathbb{H}}$ 是双曲面中的向量内积操作, $\langle x, y \rangle_{\mathbb{H}} = -x_0 y_0 + x_1 y_1 + \dots + x_d y_d$.

然后,使用 softmax 将双曲距离转换为用于知识蒸馏的对象.具体描述如下:

$$p^I(u^I, b^I) = \operatorname{softmax}(d_{\mathbb{H}}^K(e_u^I, e_b^I) / \tau), \quad (9)$$

$$p^B(u^B, b^B) = \operatorname{softmax}(d_{\mathbb{H}}^K(e_u^B, e_b^B) / \tau),$$

其中 τ 是一个称为温度的超参数.后验预测(软标签) p^I 和 p^B 揭示了实体之间的隐藏关系,并显示了两个视图学习倾向的差异,这些差异在训练数据中并未明确包含.通过将彼此的软标签视为自身的附加训练经验,两个视图可以学习到更全面的知识.然后,使用 KL 散度来促进两个视图之间的相互学习. KL 距离可以表述如下:

$$\operatorname{KL}(p^I \| p^B) = \sum_j^C P^I(x_j) \log \frac{P^I(x_j)}{P^B(x_j)}, \quad (10)$$

其中 C 是为每个用户随机采样的捆绑包的数量.最终的互学习损失函数如下:

$$L^{\text{ML}} = \alpha \operatorname{KL}(p^I \| p^B) + (1 - \alpha) \operatorname{KL}(p^B \| p^I), \quad (11)$$

其中 α 是一个超参数,用于控制两个视图之间的相互学习程度. α 越高,项目视图向捆绑包视图传递知识的力度就越大.

最后,将三个损失函数组合在一起作为最终的损失函数.公式如下:

$$L^{\text{Final}} = L^{\text{MRL}} + \lambda L^{\text{ML}} + \beta L^{\text{MP}}, \quad (12)$$

其中 λ 和 β 分别是用于加权互学习损失和图掩码预测损失的超参数.

4 实验与结果分析

4.1 实验设置及评价指标

本文所使用的数据集 Youshu 和 NetEase, NetEase 数据集是由 CAO 等^[4]爬取自网易云音乐平台.与 NetEase 数据集相似, Youshu 是 CHEN 等^[1]从优书网爬取的真实数据.这两个数据集也是捆绑推荐领域使用最为广泛的公共数据集^[1,4,6,10].训练集、验证集和测试集的划分与官方设置保持一致,划分比例为 7:1:2,划分基准是用户与捆绑包的交互个数.数据的划分情况如表 1 所示.所有的实验都是在

GPU 上进行运算, GPU 型号为 NVIDIA GTX 3090, 显存 24GB, CUDA 版本为 10.2, 服务器的内存容量为 256GB, Python 版本为 3.7.7, PyTorch 版本为 1.9.1.

表 1 实验数据集划分

Tab. 1 Division of experimental dataset

| | NetEase | Youshu |
|-----|---------|--------|
| 训练集 | 217864 | 34416 |
| 验证集 | 27857 | 5189 |
| 测试集 | 56555 | 9746 |

在实验中,对于每个基线模型,本文遵循引用参考文献中提供的最佳参数.优化器选用 Adam, 训练时并采用 4096 的批量大小.对于使用 BPR loss 的模型,负样本数量设置为 1. HyperMBR 模型学习率为 $1e-3$, 训练批量大小为 4096, 捆绑包视图和项目视图中双曲空间的曲率初始值分别设为 2.0 和 1.0. 对于边际排名损失,将边际 μ 设置为 1.5, 负样本数量设置为 256. HyperMLBR 模型学习率设置为 $1e-3$, 分别调整边际超参数 μ 、图掩码采样概率下限 p_l 和上限 p_u 、互学习系数 α 、互学习损失权重 λ 、图掩码损失权重 β 和蒸馏温度 τ , 调整范围为 $\{0.5, 1, 1.5, 2\}$ 、 $\{0, 0.2, 0.4, 0.6, 0.8, 1.0\}$ 、 $\{0.1, 0.3, 0.5, 0.7, 0.9\}$ 、 $\{0.05, 0.1, 0.5, 1, 2\}$ 、 $\{0, 0.005, 0.01, 0.1, 0.5, 1, 2\}$ 和 $\{0.8, 0.9, 1.0, 1.1, 1.2\}$. 训练时的批量大小为 4096, 嵌入的维度设置为实验的最优参数 64, 优化器采用 Adam. \mathbb{H}^a 和 \mathbb{H}^b 的曲率选择在范围 $\{1, 2, 3, 4\}$ 内. 值得一提的是两个双曲空间的曲率不一定相同.

在推荐系统领域,模型最终会向用户推荐 K 个用户交互概率最高的对象,也称为 TopK 推荐.捆绑推荐领域普遍采用的评价指标是 Recall@K 和 NDCG@K^[1,3,6,10]. 本文所使用的基线模型 (DAM、Attlist、BGCN、GRAM-SMOT、MIDGN 和 CL²BRec) 在其原论文中对性能指标的计算方式略有不同. DAM, GRAM-SMOT 用的是留一法 (leave-one-out), 用户的一个交互记录会被随机删除用作模型测试. 在测试阶段,这两个模型会对测试集中的每个正样本随机挑选 99 个用户从未交互过的捆绑包作为负样本,以此来计算 Recall@K 以及 MAP@K. 这种方式无需对所有捆绑包计算推荐分数,能够减少时间消耗,加快训练进程. 而 Attlist、BGCN、MIDGN 和 CL²BRec 则是对全部的捆绑包候选集计算 Recall@K 和 NDCG@K, 训练耗时较长,但对模型性能要求更加严格. 为了保证评价指标的严格性和不同模型性能的可比性,本文统一采用后一种方式进行指标的计算.

指标 Recall@K 是指召回命中率, 衡量的是模型的召回能力, 即模型推荐的 K 个捆绑包中有多少是用户真实交互过. 计算公式如下:

$$\text{Recall@K} = \frac{\text{Hit_Num@K}}{K}. \quad (13)$$

NDCG@K 则在 Recall@K 的基础上, 进一步考察模型对正负样本的排序能力, 希望用户喜欢的捆绑包能够在排序结果中相对靠前. 最好的情况就是模型能把所有正样本都排在负样本前面, 此时 NDCG@K 的分数最高, 为 1. 计算公式如下:

$$\text{NDCG@K} = \frac{\text{DCG@K}}{\text{IDCG@K}}, \quad (14)$$

DCG@K 的计算方式如下:

$$\text{DCG@K} = \sum_{i=1}^K \frac{\text{rel}_i}{\log_2(i+1)}, \quad (15)$$

其中, $\text{rel}_i \in [0, 1]$ 表示根据模型推测的概率排序后的位置 i 的捆绑包是否命中, 1 为命中. IDCG@K 则是指最理想的排序情况下的 DCG@K 的分数, 即所有正样本都排在负样本前面的情况.

4.2 总体性能和结果分析

本文将 HyperMLBR 与基线模型、HyperMBR 进行比较, 表 2 和表 3 分别展示了模型在 Youshu 和 NetEase 上 Recall@K 和 NDCG@K 指标的表现情况. 从表中可以看到 HyperMLBR 的性能与其他模型相比得到了进一步的提升.

表 2 在 Youshu 数据集上的实验结果

Tab. 2 Experimental results on Youshu dataset

| 模型 | Recall@K | | | | NDCG@K | | | |
|---------------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| | K=5 | K=10 | K=20 | K=40 | K=5 | K=10 | K=20 | K=40 |
| DAM ^[1] | 0.1098 | 0.1460 | 0.2082 | 0.2890 | 0.0873 | 0.0996 | 0.1198 | 0.1418 |
| AttList ^[2] | 0.0414 | 0.0613 | 0.0869 | 0.1245 | 0.0370 | 0.0424 | 0.0499 | 0.0599 |
| BGCN ^[3] | 0.0889 | 0.1515 | 0.2336 | 0.3415 | 0.0848 | 0.1032 | 0.1268 | 0.1559 |
| GRAM-SMOT ^[10] | 0.1129 | 0.1673 | 0.2408 | 0.3451 | 0.1037 | 0.1184 | 0.1392 | 0.1675 |
| MIDGN ^[6] | 0.1210 | 0.1728 | 0.2519 | 0.3466 | 0.1076 | 0.1210 | 0.1440 | 0.1696 |
| CL2BRec ^[7] | 0.1173 | 0.1810 | 0.2649 | 0.3615 | 0.1106 | 0.1290 | 0.1531 | 0.1798 |
| HyperMBR ^[5] | 0.1275 | 0.1875 | 0.2774 | 0.3762 | 0.1149 | 0.1343 | 0.1589 | 0.1857 |
| HyperMLBR (Ours) | 0.1329 | 0.2024 | 0.2826 | 0.3816 | 0.1233 | 0.1439 | 0.1664 | 0.1940 |
| % Improv. | 4.24% | 7.95% | 1.87% | 1.44% | 7.31% | 7.15% | 4.72% | 4.47% |

表 3 在 NetEase 数据集上的实验结果

Tab. 3 Experimental results on NetEase dataset

| 模型 | Recall@K | | | | NDCG@K | | | |
|---------------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| | K=5 | K=10 | K=20 | K=40 | K=5 | K=10 | K=20 | K=40 |
| DAM ^[1] | 0.0099 | 0.0163 | 0.0411 | 0.0690 | 0.0077 | 0.0102 | 0.0210 | 0.0281 |
| AttList ^[2] | 0.0194 | 0.0325 | 0.0548 | 0.0898 | 0.0163 | 0.0211 | 0.0281 | 0.0373 |
| BGCN ^[3] | 0.0211 | 0.0361 | 0.0581 | 0.0934 | 0.0187 | 0.0239 | 0.0308 | 0.0401 |
| GRAM-SMOT ^[10] | 0.0204 | 0.0346 | 0.0595 | 0.0972 | 0.0174 | 0.0225 | 0.0303 | 0.0402 |
| MIDGN ^[6] | 0.0225 | 0.0374 | 0.0607 | 0.0968 | 0.0192 | 0.0245 | 0.0319 | 0.0415 |
| CL2BRec ^[7] | 0.0233 | 0.0400 | 0.0670 | 0.1064 | 0.0208 | 0.0267 | 0.0351 | 0.0455 |
| HyperMBR ^[5] | 0.0239 | 0.0426 | 0.0712 | 0.1126 | 0.0213 | 0.0277 | 0.0381 | 0.0490 |
| HyperMLBR (Ours) | 0.0287 | 0.0470 | 0.0774 | 0.1194 | 0.0260 | 0.0321 | 0.0407 | 0.0518 |
| % Improv. | 20.08% | 10.33% | 8.71% | 6.04% | 22.07% | 15.88% | 6.82% | 5.71% |

从实验结果可见, HyperMLBR 在两个数据集上比 HyperMBR 的 Recall@K 指标分别提高 1.44%~7.95% 和 6.04%~20.08%; 比 HyperMBR 的 NDCG@K 指标分别提高 4.47%~7.31% 和 5.71%~22.07%. HyperMLBR 模型性能的提升可以归因于以下三个原因:

(1) 图掩码模块加强了项目视图的学习效果. 图掩码模块加强了项目视图中较为重要边的学习

力度, 并通过边遮掩重构的方式重建了显著缺失的信息, 并进一步推测潜在的有效交互, 提升了项目视图的学习效果, 最终帮助模型提升了整体性能.

(2) 互学习模块起到了视图之间相互增强的效果. 两个视图之间的相互学习成功帮助两个视图学到了更全面的实体表征, 最终提升了模型的性能.

(3) 双曲空间缓解了嵌入的失真问题. 模型充分考虑了捆绑推荐中交互图的树状结构, 并利用双

曲空间对于这类数据强大的嵌入能力,使得模型能够学到更准确的实体表示,从而让视图之间传递更准确的知识,减少了误导.

4.3 消融实验

为了探究和分析 HyperMLBR 模型中图掩码模块以及跨视图的互学习模块对整体性能以及模型其他模块的影响,本文在 Youshu 和 NetEase 数据集上设计了 4 个消融实验模型,分别如下:

(1)w/o IV (Item View) 和 w/o BV (Bundle View): 这两个变体在训练阶段保持一致,仍然是两个视图同步训练,但是在最终预测阶段,w/o IV 只保留捆绑包视图的预测结果,w/o BV 只保留项目视图的预测结果.这两个变体实验是为了验证加入图掩码模块、互学习模块之后,两个视图是否如期望的一样能够得到增强.

(2)w/o Hyp (without Hyperbolic): 这个变体将模型中的编码器改为传统的基于欧式空间的图卷积模型,探究互学习模块在失去了相对准确嵌入的基础后,是否还能起到相互增强的目的.

(3)w/o MP (without mask predict): 这个变体去掉了图掩码模块,探究不对项目视图中的交互数据进行有区别的学习对模型性能的影响.

(4)w/o ML (without Mutual Learning): 这个变体实验在 HyperMLBR 的基础上移除了互学习模块,探究跨视图的相互学习对模型性能的影响.

实验结果如表 4 所示,可得出如下结论:

表 4 在 Youshu 和 NetEase 数据集上的消融实验结果

Tab. 4 The ablation experiment results on Youshu and NetEase datasets

| 模型 | Youshu | | NetEase | |
|-----------|---------------|---------------|---------------|---------------|
| | Recall@5 | NDCG@5 | Recall@5 | NDCG@5 |
| HyperMLBR | 0.1329 | 0.1233 | 0.0287 | 0.0260 |
| w/o IV | 0.0649 | 0.0627 | 0.0154 | 0.0145 |
| w/o BV | 0.1135 | 0.0978 | 0.0152 | 0.0121 |
| w/o Hyp | 0.1176 | 0.1109 | 0.0238 | 0.0212 |
| w/o ML | 0.1284 | 0.1182 | 0.0249 | 0.0219 |
| w/o MP | 0.1321 | 0.1227 | 0.0280 | 0.0251 |

(1)图掩码模块是有效的.为了评估图掩码模块是否对模型性能有贡献,本文设计了变体模型 HyperMLBR w/o MP,该变体模型在 HyperMLBR 中移除了图掩码模块.这个变体模型在两个数据集上都出现了不同程度性能的下降,这就证明项目视图确实更加注重项目和捆绑包之间的相似度,更偏向于向用户推荐那些包含用户历史交互过的项目的

捆绑包.

(2)跨视图的互学习模块是有效的.为了评估互学习模块是否对模型性能有贡献,本文设计了一个模型的变体,使其在式 12 中删除互学习损失.本文将该变体模型称为 HyperMLBR w/o ML.该变体模型在 NetEase 数据集上的 Recall@5 和 NDCG@5 指标上分别下降 13.24% 和 15.77%,在 Youshu 数据集上分别下降 3.39% 和 4.14%.这证明了促使两个视图相互学习的重要性.同时,w/o IV 和 w/o BV 两个变体模型的指标出现了不同程度的下降,这说明相互学习让视图之间传递了自身学习到的知识,但视图本身仍然保留了自身的优势,无法被完全取代.而且在增添了互学习的损失函数之后,两个视图的表现都比原先更好,这说明互学习模块确实使得两个视图实现相互增强的目的.

(3)准确的嵌入减少了互学习过程中的误导.w/o Hyp 模型中,让模型在欧式空间中执行对应操作,并确保其他条件与双曲空间完全一致.该变体模型在 NetEase 数据集上的 Recall@5 和 NDCG@5 指标上分别下降 17.07% 和 18.46%,在 Youshu 数据集上的 Recall@5 和 NDCG@5 分别下降 11.51% 和 10.06%.变体模型在两个数据集上的性能甚至略微不如 HyperMLBR.这说明准确的实体嵌入,是互学习模块的基础.如果模型的底层嵌入本身就带有较大的误差,视图本身的预测结果就会带有较大的错误,这个错误会在相互学习的过程中传递给对方,进一步放大了错误,最终导致模型性能下降.

4.4 互学习效果分析

为了进一步探索互学习模块的作用,本文进行了相关实验来研究以下两个问题:(1)两个视图在训练过程中是否从彼此中提取了知识?(2)两个视图是否加强了彼此?

为了证明知识确实在两个视图之间被蒸馏,即两个视图会达成一定程度的公式,输出更相似的 logits,本文可视化了训练期间 KL 距离的变化曲线.训练时会给予两个视图一个热身过程,以便它们有时间学到一定水平的基础知识,而不是过早地将未成熟的知识传递给彼此.因此,从第 10 个 epoch 开始计算 KL 距离的变化,结果如图 2 所示.互学习模块的加入显著减小了两个视图 logits 之间的 KL 距离,这说明随着训练的进行两个视图确实从彼此中学到了知识,并逐渐在推荐结果上达成了一定程度的共识.而且可以看到 KL 距离一直处于大于 0 的状态,这说明尽管两个视图互相作为对方的教师模型,传递知识,

两者还是存在自身的特点, 保持着与对方的差异.

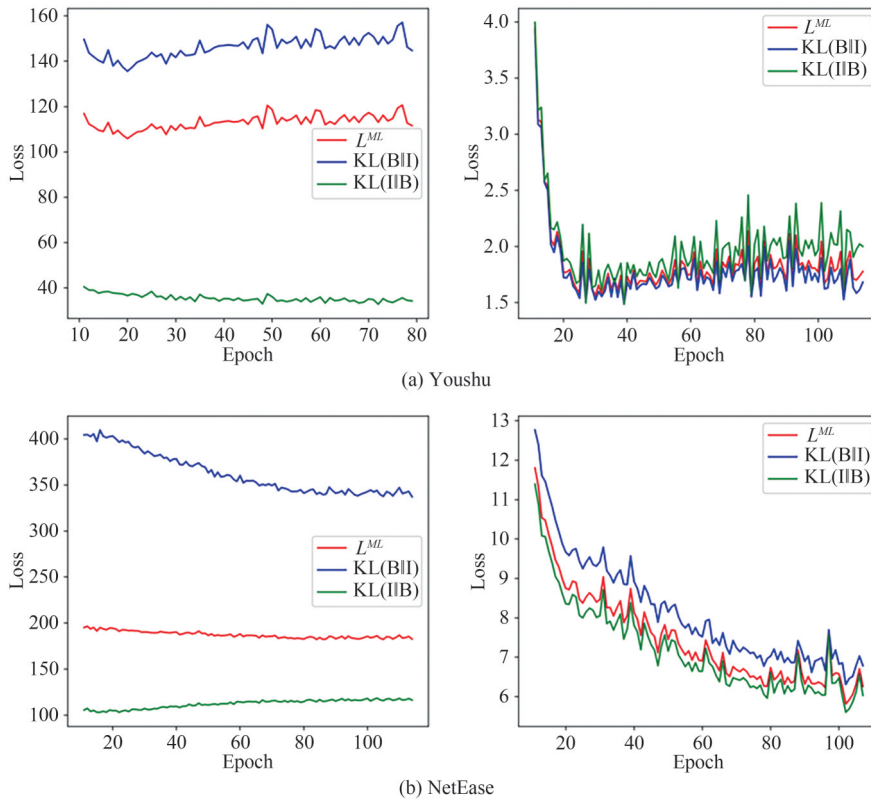


图2 两个数据集在训练过程中的KL距离变化

Fig. 2 KL distance changes of the two datasets during training

为了验证两个视图是否加强了彼此, 在图 3 中展示了基于单视图和双视图计算的 Recall@20 指标. 捆绑包视图预测使用 $d_{\text{diff}}(e_u^B, e_b^B)$, 而项目视图预测使用 $d_{\text{diff}}(e_u^I, e_b^I)$. 从实验结果中得出以下结论: 1) 互学

习模块显著提高了所有三种预测类型的推荐性能. 2) 单一视图的性能远远不及双视图, 说明在两个视图学到的知识之间存在强烈的互补性, 并证明使用互学习模块来鼓励它们相互加强的动机是合理的.

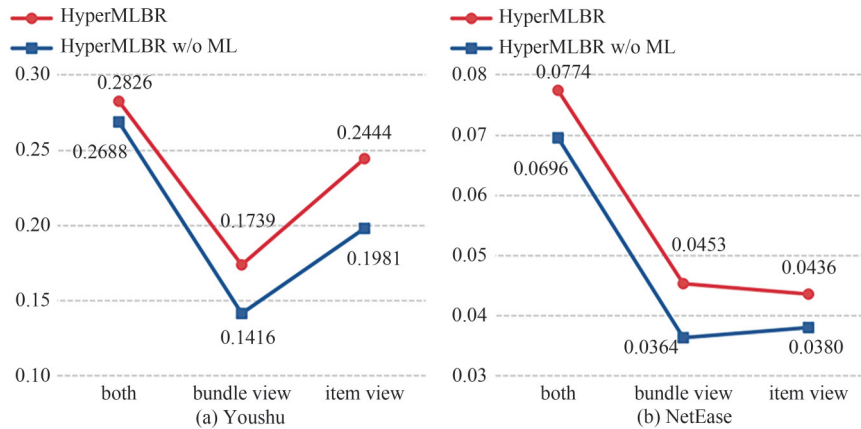


图3 三种视图下 Recall@20 指标的比较

Fig. 3 Comparison of the Recall@20 indicator in three views

5 结语

经过对捆绑推荐的分析, 发现捆绑推荐中捆绑包视图和项目视图有着各自的学习倾向, 并且这种

不同的学习倾向会使得最终推荐结果产生变化, 而且用户-项目视图中不同边具有不同的重要性. 两个视图的独立训练以及最终推荐结果的简单相加, 未能有效利用两个视图各自的优势, 而且对于不同

边的重要性建模也不够理想.针对这两个问题,本文提出了一种双曲空间互学习的捆绑推荐模型 HyperMLBR,该模型通过图掩码模块加强了项目视图中嵌入的学习,然后利用互学习损失函数促使两个视图向对方学习,在保持自身优势的同时,学习对方的长处,最终达到互相增强的目的.最终的实验结果和效果分析不仅表明了两个模块的有效性,也表明了互学习模块确实如先前分析的一样产生了正向效果,而且说明了一个准确的嵌入是互学习模块的重要基础.

参 考 文 献

- [1] CHEN L, LIU Y, HE X, et al. Matching user with item set: Collaborative bundle recommendation with deep attention network [C]//Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. Macao: International Joint Conferences on Artificial Intelligence Organization, 2019: 2095-2101.
- [2] HE Y, WANG J, NIU W, et al. A hierarchical self-attentive model for recommending user-generated item lists [C]//Proceedings of the 28th ACM International Conference on Information and Knowledge Management. Beijing: ACM, 2019: 1481-1490.
- [3] CHANG J, GAO C, HE X, et al. Bundle recommendation with graph convolutional networks [C]//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. Virtual Event: ACM, 2020: 1673-1676.
- [4] CAO D, NIE L, HE X, et al. Embedding factorization models for jointly recommending items and user generated lists [C]//Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. Tokyo: ACM, 2017: 585-594.
- [5] KE H, LI L, WANG P, et al. Hyperbolic mutual learning for Bundle recommendation [C]// Database Systems for Advanced Applications. Cham: Springer Nature Switzerland, 2023: 417-433.
- [6] ZHAO S, WEI W, ZOU D, et al. Multi-view intent disentangle graph networks for bundle recommendation [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(4): 4379-4387.
- [7] WANG N, SUN J, LI J. Cross-level relational graph contrastive learning for bundle recommendation [C]// 2023 IEEE International Conference on Web Services. Chicago: IEEE, 2023: 112-117.
- [8] CHEN W, FANG W, HU G, et al. On the hyperbolicity of small-world and treelike random graphs [J]. Internet Mathematics, 2013, 9(4): 434-491.
- [9] RAVASZ E, BARABÁSI A L. Hierarchical organization in complex networks [J]. Physical Review E, Statistical, Nonlinear, and Soft Matter Physics, 2003, 67 (2) : 026112.
- [10] VIJAIKUMAR M, SHEVADE S, MURTY M N. GRAM-SMOT: Top-N personalized bundle recommendation via graph attention mechanism and submodular optimization [C]// Machine Learning and Knowledge Discovery in Databases. Cham: Springer International Publishing, 2021: 297-313.
- [11] HE Y, ZHANG Y, LIU W, et al. Consistency-aware recommendation for user-generated item list continuation [C]//Proceedings of the 13th International Conference on Web Search and Data Mining. Houston: ACM, 2020: 250-258.
- [12] WANG X, HE X, WANG M, et al. Neural graph collaborative filtering [C]//Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. Paris: ACM, 2019: 165-174.
- [13] YING R, HE R, CHEN K, et al. Graph convolutional neural networks for web-scale recommender systems [C]// Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London: ACM, 2018: 974-983.
- [14] 葛尧, 陈松灿. 面向推荐系统的图卷积网络 [J]. 软件学报, 2020, 31(4): 1101-1112.
- [15] DENG Q, WANG K, ZHAO M, et al. Personalized bundle recommendation in online games [C]//Proceedings of the 29th ACM International Conference on Information & Knowledge Management. Virtual Event: ACM, 2020: 2381-2388.
- [16] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//Proceedings of the Neural Information Processing Systems. Long Beach: NIPS Foundation, 2017, 30: 5998-6008.
- [17] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 770-778.
- [18] HE X, LIAO L, ZHANG H, et al. Neural collaborative filtering [C]//Proceedings of the 26th International Conference on World Wide Web. Perth: International World Wide Web Conferences Steering Committee, 2017: 173-182.
- [19] HSIEH C K, YANG L, CUI Y, et al. Collaborative metric learning [C]//Proceedings of the 26th International

- Conference on World Wide Web. Perth; International World Wide Web Conferences Steering Committee, 2017: 193-201.
- [20] NICKEL M, KIELA D. Poincaré embeddings for learning hierarchical representations [EB/OL]. 2017: 1705.08039. <https://arxiv.org/abs/1705.08039v2>.
- [21] LI H, HESTENES D, ROCKWOOD A. A universal model for conformal geometries of Euclidean, spherical and double-hyperbolic spaces [M] Berlin: Springer Berlin Heidelberg, 2001.
- [22] CHAMI I, YING R, RÉ C, et al. Hyperbolic graph convolutional neural networks [J]. Advances in Neural Information Processing Systems, 2019, 32: 4869-4880.
- [23] ZHANG S, CHEN H, MING X, et al. Where are we in embedding spaces? [C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. Virtual Event: ACM, 2021: 2223-2231.
- [24] SUN J, CHENG Z, ZUBERI S, et al. Hgcf: Hyperbolic graph convolution networks for collaborative filtering [C]//Proceedings of the Web Conference 2021. Ljubljana: SIGWEB 2021: 593-601.
- [25] WANG L, HU F, WU S, et al. Fully hyperbolic graph convolution network for recommendation [C]//Proceedings of the 30th ACM International Conference on Information & Knowledge Management. Virtual Event: ACM, 2021: 3483-3487.
- [26] HE K, CHEN X, XIE S, et al. Masked autoencoders are scalable Vision learners [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 16000-16009.

(责编&校对 雷建云)