

百度2024：HCTS——双曲神经网络在跨域推荐系统中的迁移学习效能优化



SmartMindAI

专注搜索、广告、推荐、大模型和人工智能最新技术，欢迎关注我

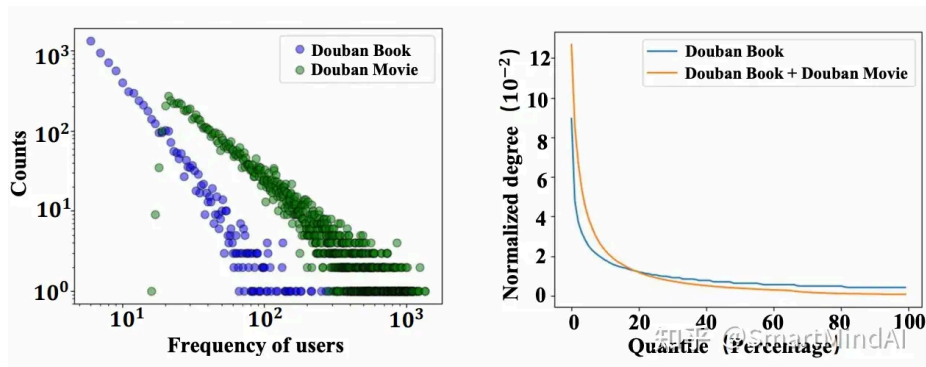
已关注

10 人赞同了该文章

Introducton

推荐系统⁺被广泛应用于电子商务、视频流媒体平台和智能手机应用市场等领域，以解决信息过载问题。推荐系统利用用户的交互数据来预测其偏好。然而，推荐系统面临冷启动问题和数据稀疏性问题。冷启动问题指的是系统遇到新用户或没有足够历史数据的物品时无法给出准确的推荐。数据稀疏性问题是由于用户与物品之间的有限交互导致的。

为了解决推荐场景中的挑战，一种方法是通过集成其他数据源或情景的数据来增强模型性能。跨域推荐（CDR）是一种通过知识迁移策略，利用源域数据来提升目标域模型性能的方法。尽管在CDR领域取得了进展，长尾问题仍然存在。在CDR任务中，不同域数据的合并可能加剧长尾分布。传统神经网络模型⁺在处理长尾分布数据时表现不佳，因为它们将物品和用户嵌入欧几里得⁺空间进行编码。相比之下，双曲流形是一种非欧几何空间，适合表示树状或层次结构。双曲流形的指数扩展特性使其能够更好地表示具有指数级增长的项目。



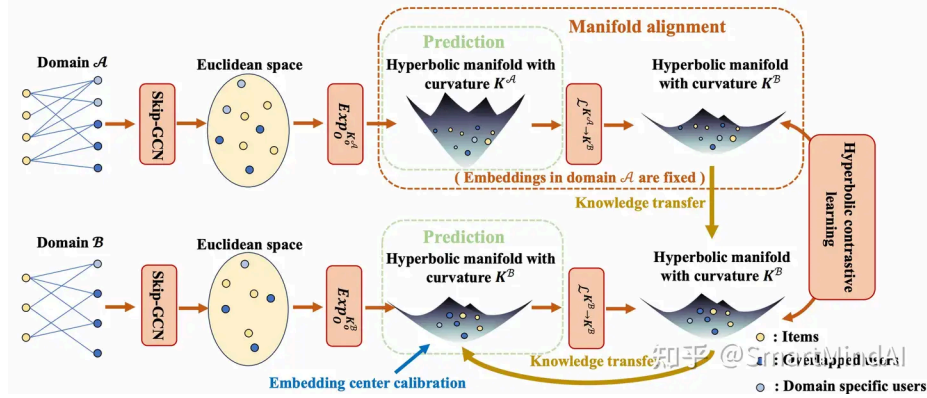
近年来，双曲神经网络在处理长尾分布数据方面取得了巨大进展。为了解决CDR任务，本文开发了基于双曲神经网络的模型。然而，源域⁺和目标域之间的差异带来了两个挑战：1. 如何在双曲空间⁺中捕捉不同域之间的一致性；2. 如何在双曲流形上实现知识转移。

针对上述挑战，我们提出了一种名为HCTS的创新框架用于CDR任务。首先，我们使用两个GNN模块在两个域的节点上执行邻域传播，然后将用户和项目嵌入到曲率适应的双曲流形上。为了解决双曲流形间的知识传递问题，我们引入了基于双曲对比学习的知识传递模块，包括流形对齐、双曲对比学习和嵌入中心校正。这些方法能够有效地捕捉域的特异性并传递知识。

Methodology

Framework Overview

损失、校正损失和双曲边缘排名损失来训练模型。



Embedding Layer

所有 $u^S \in \mathcal{U}^S$, $i^S \in \mathcal{I}^S$, $u^T \in \mathcal{U}^T$ 和 $i^T \in \mathcal{I}^T$ 均独立输入到嵌入层中, 从这些输入我们得到源用户 $u^{\mathcal{E},S}$, 源项目 $i^{\mathcal{E},S}$, 目标用户 $u^{\mathcal{E},T}$ 和目标项目 $i^{\mathcal{E},T}$ 。我们将用户和项目嵌入到欧几里得空间中, 分别标记为 \mathcal{E} 的上标。重叠的用户在源域和目标域中都有嵌入, 而非重叠的用户只有一个嵌入。简化后的表示为 $u^{\mathcal{E}}$ 和 $i^{\mathcal{E}}$ 。

Single Domain GNN Aggregator

在推荐系统领域, 现有的GNN模型在传递节点之间的消息时往往会遇到问题。推荐任务中的用户和项目特征通常是高度稀疏的, 使用特征转换和激活函数⁺可能无法改善结果。此外, 堆叠多个GNN层可能会导致梯度消失或过度平滑的问题。为了解决这些问题, 我们使用skip-GCN进行邻域聚合。

$$\begin{aligned} u^{\mathcal{E},(l+1)} &= u^{\mathcal{E},(l)} + \sum_{i \in N_u} \frac{1}{|N_u|} i^{\mathcal{E},(l)}, \\ i^{\mathcal{E},(l+1)} &= i^{\mathcal{E},(l)} + \sum_{u \in N_i} \frac{1}{|N_i|} u^{\mathcal{E},(l)}, \end{aligned}$$

其中的上标 (l) 表示 skip-GCN 的层, skip-GCN 的输出是

$$\tilde{u}^{\mathcal{E}} = \sum_{k=1}^L u^{\mathcal{E},(k)}, \quad \tilde{i}^{\mathcal{E}} = \sum_{k=1}^L i^{\mathcal{E},(k)}$$

对于源域和目标域, 我们分别执行邻居聚合, 每个用户和项目最终的嵌入可以表示为源域的 $\tilde{u}^{\mathcal{E},S}$, $\tilde{i}^{\mathcal{E},S}$

以及目标域的 $\tilde{u}^{\mathcal{E},T}$, $\tilde{i}^{\mathcal{E},T}$

Hyperbolic Manifold Projection

在CDR任务中, 目标域的交互图通常稀疏且受冷启动问题影响, 而源域的交互图更密集。因此, 在下游预测任务中, 源域和目标域的双曲流形需要不同的最佳曲率。为此, 我们为每个域引入可训练的曲率, 以便将交互图嵌入到具有最优曲率的双曲流形中。然后, 我们使用指数映射函数将这些嵌入映射⁺到双曲流形中。

$$\begin{aligned} u^{\mathcal{H}^{K_S},S} &= \text{Exp}_o^{K_S}(0, \tilde{u}^{\mathcal{E},S}), & i^{\mathcal{H}^{K_S},S} &= \text{Exp}_o^{K_S}(0, \tilde{i}^{\mathcal{E},S}), \\ u^{\mathcal{H}^{K_T},T} &= \text{Exp}_o^{K_T}(0, \tilde{u}^{\mathcal{E},T}), & i^{\mathcal{H}^{K_T},T} &= \text{Exp}_o^{K_T}(0, \tilde{i}^{\mathcal{E},T}), \end{aligned}$$

其中 K_S 和 K_T 是可训练参数。

Knowledge Transfer via Hyperbolic Contrastive Learning

流形对齐是该框架的关键部分。由于源域和目标域的嵌入被投影到具有不同曲率的双曲流形上，现有的双曲对比学习方法无法直接应用。这是因为双曲对比学习基于双曲距离的计算，而在不同双曲流形上计算距离是不可行的。双曲距离函数只在对应双曲流形的切空间中有效，具有线性共轭、非负性和对称性等性质。

$\langle \cdot, \cdot \rangle_{\mathcal{M}}$ 在 \mathbb{R}^{n+1} 上，它只是一个伪内积，但它是限制在 $\mathcal{H}^{n,K}$ 的切空间上的内积，即

$\langle \cdot, \cdot \rangle_{\mathcal{M}} : T_x \mathcal{H}^{n,K} \times T_x \mathcal{H}^{n,K} \rightarrow \mathbb{R}$ 对于 $\mathcal{H}^{n,K}$ 中的所有 x ， $\langle \cdot, \cdot \rangle_{\mathcal{M}}$ 是一个明确定义的内积。然后， $\|u\|_{\mathcal{M}} = \sqrt{\langle u, u \rangle_{\mathcal{M}}}$ 是一个明确定义的范数。

为了在不同曲率的双曲流形上进行双曲对比学习，我们需要将嵌入投影到具有相同曲率的双曲流形上。根据不同曲率的双曲流形的特性，我们定义了从曲率为 K_A 的流形到曲率为 K_B 的流形的线性变换。这个线性变换保持了流形的切空间在北极点的一致性。

$$\mathcal{L}^{K_A \rightarrow K_B}(x^{\mathcal{H}^{K_A}}) = \text{Exp}_o^{K_B} \left(\text{ReLU} \left(W \text{Log}_o^{K_A}(x^{\mathcal{H}^{K_A}}) \right) \right)$$

为了确保

$$\text{ReLU} \left(W \text{Log}_o^{K_A} x^{\mathcal{H}^{K_A}} \right)$$

保持在北极点的切空间中，我们将 W 设置为一个可训练矩阵，其第一行为零。

双曲对比学习

基于上述流形对齐函数，我们可以使用对比学习技术在不同曲率的双曲流形之间传递知识。为了从域 \mathcal{A} 传递知识到域 \mathcal{B} ，其中 \mathcal{A} 和 \mathcal{B} 可以是 \mathcal{S} 或 \mathcal{T} ，我们首先使用线性转换将具有曲率 K_A 的 \mathcal{A} 域中的嵌入映射到具有曲率 K_B 的双曲流形。

$$e^{\mathcal{H}^{K_B}, \mathcal{A}} = \mathcal{L}^{K_A \rightarrow K_B}(e^{\mathcal{H}^{K_A}, \mathcal{A}}),$$

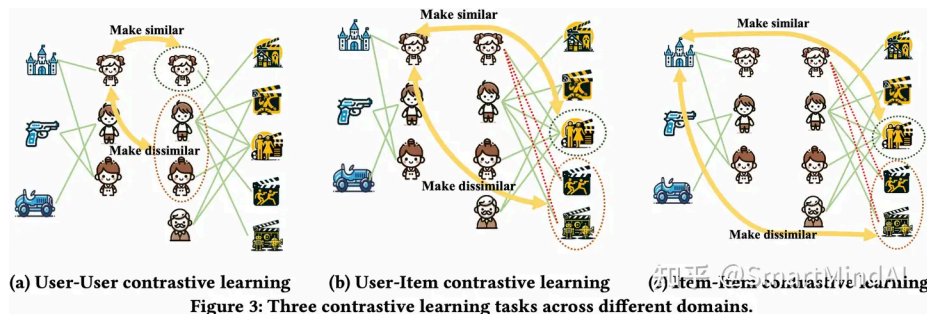
其中 $e^{\mathcal{H}^{K_B}, \mathcal{A}}$

在具有曲率 K_B 的双曲流形上表示域 \mathcal{A} 的嵌入，而 e 可以是 u 或 i 。在将知识从 \mathcal{A} 域传递到 \mathcal{B} 域的过程中，我们保持嵌入 $e^{\mathcal{H}^{K_B}, \mathcal{A}}$

在训练过程中，这些嵌入保持不变。为了保持对称性，我们还对 \mathcal{B} 域的嵌入进行了线性转换。

$$e^{\mathcal{H}^{K_B}, \mathcal{B}} = \mathcal{L}^{K_B \rightarrow K_B}(e^{\mathcal{H}^{K_B}, \mathcal{B}}).$$

在具有曲率 K_B 的双曲流形上，我们定义了三种对比学习策略：用户-用户对比学习、用户-项目对比学习和项目-项目对比学习。这些策略用于在重叠用户之间传递知识。尽管用户在两个不同的域中的行为可能会有所不同，但是同一用户在两个域中的行为之间的相关性自然高于不同用户之间行为的相关性。因此，我们使用对比学习策略来增强知识传递。



$$L_{u-u}^{K_B} = - \sum_{i \in \mathcal{U}^o} \log \frac{\exp(\text{sim}(u_i^{\mathcal{H}^{K_B}, \mathcal{A}}, u_i^{\mathcal{H}^{K_B}, \mathcal{B}})/\tau)}{\sum_{j \in \mathcal{U}^o \setminus \{i\}} \exp(\text{sim}(u_i^{\mathcal{H}^{K_B}, \mathcal{A}}, u_j^{\mathcal{H}^{K_B}, \mathcal{B}})/\tau)},$$

$$\text{sim}(x, y) = -d_{\mathcal{M}}^K(x, y),$$

在两个相关的数据集中，一个用户在一个域中的行为与另一个域中的行为高度相关。我们定义了用户-项目对比学习，通过双曲对比学习方法，将一个域中用户的嵌入与他们在另一个域中互动的项目紧密对齐。在用户-项目对比学习中，我们选择了一个正面项目和几个负例样本来进行学习。

$$L_{u-i}^{K_B} = - \sum_{i \in \mathcal{U}^O} \log \frac{\exp(\text{sim}(\mathbf{u}_i^{\mathcal{H}^{K_B}, B}, \mathbf{i}_{pos}^{\mathcal{H}^{K_B}, A})/\tau)}{\sum_{j \in N_{neg}} \exp(\text{sim}(\mathbf{u}_i^{\mathcal{H}^{K_B}, B}, \mathbf{i}_j^{\mathcal{H}^{K_B}, A})/\tau)},$$

其中 \mathbf{i}_{pos} 表示在另一域中与用户 i 互动的项。 N_{neg} 表示在另一域中未与用户 i 互动的一组项目。为了在不同域中相同用户互动的项之间传递知识，我们定义了项目-项目对比学习如下：

$$L_{i-i}^{K_B} = - \sum_{i \in \mathcal{U}^O} \log \frac{\exp(\text{sim}(\mathbf{i}_{pos}^{\mathcal{H}^{K_B}, B}, \mathbf{i}_{pos}^{\mathcal{H}^{K_B}, A})/\tau)}{\sum_{j \in N_{neg}} \exp(\text{sim}(\mathbf{i}_{pos}^{\mathcal{H}^{K_B}, B}, \mathbf{i}_j^{\mathcal{H}^{K_B}, A})/\tau)}.$$

基于上述讨论的对比策略，我们首先考虑通过目标流形 K_T 上的对比学习，从源域向目标域 K_T 转移知识，并获得三个损失函数： $L_{u-u}^{K_T}$ 、 $L_{u-i}^{K_T}$ 和 $L_{i-i}^{K_T}$ 。此外，由于源域中嵌入的质量决定了向目标域转移的知识质量，我们还考虑了从目标域向源域转移知识，并获得了 $L_{u-u}^{K_S}$ 、 $L_{u-i}^{K_S}$ 和 $L_{i-i}^{K_S}$ 。最后，对比知识转移任务的整体优化目标为：

$$L_{cts} = L_{u-u}^{K_T} + L_{u-i}^{K_T} + L_{i-i}^{K_T} + L_{u-u}^{K_S} + L_{u-i}^{K_S} + L_{i-i}^{K_S}$$

嵌入中心校准。

对比学习使得相似节点的嵌入更接近，不相似节点的嵌入更远，增强了表示的辨别性。然而，这可能导致目标域中嵌入的中心偏离双曲流形的北极点，导致双曲表示失真，降低了其在建模数据层次结构时的有效性。为了纠正这个问题，我们使用校准函数来调整嵌入的中心偏差。

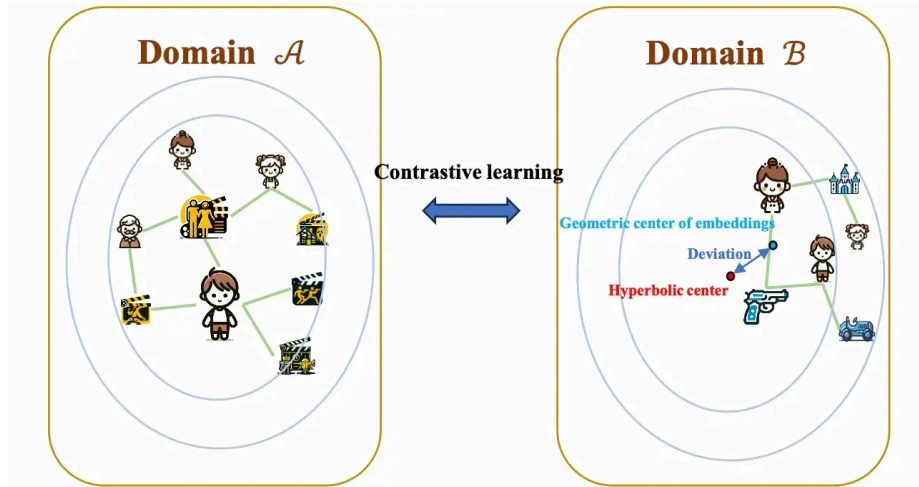


Figure 4: The deviation of embeddings from the north pole.

具体来说，我们将双曲嵌入映射回切空间：

$$\mathbf{e}^{\mathcal{E}} = \text{Log}_o^K(\mathbf{e}^{\mathcal{H}^K})$$

其中 \mathbf{e} 是用于预测的向量，可以是用户或项目。然后我们计算它们的几何中心。

$$\mathbf{e}_c^{\mathcal{E}} = \frac{1}{|\mathcal{V}|} \sum_{i \in \mathcal{V}} \mathbf{e}_i^{\mathcal{E}}$$

定义校准损失函数 L_{clib} ，用于确保嵌入的几何中心接近北极点。该损失函数使用欧几里得距离 $d^{\mathcal{E}}$ 计算。

Hyperbolic Margin Ranking Loss

研究中，我们采用双曲边际排序损失作为预测损失。在源域，预测损失为：

$$L_S(u^{K_S, \mathcal{S}}, i_{pos}^{K_S, \mathcal{S}}, i_{neg}^{K_S, \mathcal{S}}) = \max \left(d_{\mathcal{M}}(u^{K_S, \mathcal{S}}, i_{pos}^{K_S, \mathcal{S}})^2 - d_{\mathcal{M}}(u^{K_S, \mathcal{S}}, i_{neg}^{K_S, \mathcal{S}})^2 + m, 0 \right),$$

其中 m 是一个非负超参数 $u^{K_S, \mathcal{S}}$ 表示源域中用户在双曲流形上的嵌入。 K_S 是通过训练以最佳拟合源域的数据得到的。

$i_{pos}^{K_S, \mathcal{S}}$ 是该用户在这个双曲流形上的正样本的嵌入，而 $i_{neg}^{K_S, \mathcal{S}}$ 则是该用户在这个相同双曲流形上的一个负样本的嵌入。对于目标域，我们以相同方式获取预测损失 L_T 。

Multi-task Optimization

我们对整个网络的训练进行了多任务优化。损失函数定义如下：

$$L = L_S + L_T + \lambda_{cts} L_{cts} + \lambda_{clib} L_{clib},$$

其中 λ_{cts} 和 λ_{clib} 是从0到1的超参数。

Experimental Setup

Datasets and Evaluation Protocols.

我们通过多个实际世界的数据集评估了我们提出的模型和基线模型，从Amazon数据集选择了四个子集，从Douban数据集选择了三个子集。我们通过HR@10和NDCG@10对所有这些模型进行评估。

Overall Performance Comparison (RQ1)

在表中，前三个模型是单域方法，只在目标域数据上训练。后面的模型是跨域方法，它们在源域和目标域数据上都进行训练。观察结果如下：

Models	Amazon						Douban					
	Book→Movie		Book→Music		Movie→Toy		Book→Music		Movie→Book		Movie→Music	
	N@10	H@10	N@10	H@10	N@10	H@10	N@10	H@10	N@10	H@10	N@10	H@10
HGCF	0.0347	0.0942	0.0488	0.1163	0.0253	0.0568	0.0444	0.1722	0.0481	0.1990	0.0448	0.1722
LightGCN	0.0230	0.0653	0.0449	0.1066	0.0299	0.0591	0.0409	0.1643	0.0295	0.1260	0.0409	0.1643
GCF	0.0234	0.0668	0.0456	0.1076	0.0278	0.0544	0.0414	0.1626	0.0443	0.1718	0.0416	0.1661
BiTGCF	0.0271	0.0832	0.0460	0.1143	0.0303	0.0615	0.0403	0.1608	0.0430	0.1779	0.0451	0.1775
CoNet	0.0131	0.0395	0.0102	0.0281	0.0095	0.0249	0.0212	0.0949	0.0373	0.1422	0.0216	0.0975
DTCDR	0.0132	0.0413	0.0231	0.0630	0.0150	0.0333	0.0207	0.0923	0.0406	0.1615	0.0276	0.1169
CMF	0.0241	0.0714	0.0421	0.1017	0.0289	0.0604	0.0384	0.1591	0.0428	0.1633	0.0430	0.1696
DeepAPF	0.0225	0.0649	0.0347	0.0930	0.0282	0.0568	0.0331	0.1371	0.0371	0.1555	0.0340	0.1371
CLFM	0.0157	0.0484	0.0253	0.0698	0.0161	0.0385	0.0215	0.1116	0.0345	0.1482	0.0258	0.1134
EMCDR	0.0202	0.0573	0.0148	0.0453	0.0249	0.0568	0.0290	0.0833	0.0425	0.1045	0.0303	0.0833
CCDR	0.0171	0.0557	0.0118	0.0397	0.0289	0.0605	0.0194	0.0448	0.0253	0.1125	0.0242	0.1142
ART-CAT	0.0236	0.0718	0.0334	0.1008	0.0285	0.0586	0.0308	0.1616	0.0462	0.1784	0.0458	0.1792
HCTS (ours)	0.0361*	0.0969*	0.0512*	0.1279*	0.0328*	0.0645*	0.0474*	0.1898*	0.0455*	0.2061*	0.0471*	0.1915*
Improvement	+4.03%	+2.86%	+4.91%	+9.97%	+8.25%	+4.87%	+6.76%	+10.22%	+1.03%	+2.76%	+3.49%	+2.95%

1. HCTS在性能上明显优于HGCF，因为它不仅捕捉了数据的层次结构，还有效地从源域转移知识来辅助目标域的学习。
2. 在所有基线中，HCTS的表现都优于其他基线，但受源域和目标域之间相关性的影响。在某些数据集上，改进适中，而在其他数据集（如Amazon和Douban数据集中的Book-Music任务）上，改进显著。

原文《Hyperbolic Knowledge Transfer in Cross-Domain Recommendation System》

发布于 2024-07-22 10:50 · IP 属地北京