

基于度量学习的多空间推荐系统

檀彦超 郑小林 魏翔宇 阳 及

(浙江大学计算机科学与技术学院 杭州 310007)

(埃默里大学计算机科学与技术学院 亚特兰大 30322 美国)

摘 要 隐式反馈具有数据获取成本小、形式广泛的特点,因此在现代推荐系统中被广泛使用.由于用户的隐式反馈通常是稀疏,不平衡,且含义不明确的.因此,想要准确学习用户和物品之间的复杂交互具有挑战性.传统的基于矩阵分解的推荐方法只能建模用户-物品之间的相似性.同时,矩阵分解使用点积运算作为相似度评估方式,而点积运算不满足三角不等式,即不能将用户-物品相似性传递到用户-用户以及物品-物品的相似性建模.因此,矩阵分解不足以在隐式反馈中充分建模用户和物品的关系.尽管现在有基于隐式反馈使用欧式距离来度量用户-物品相似度的度量学习方法,使得对应的推荐方法能够满足三角不等式.但是,现有的度量方法通常会将每个用户或者物品表示为度量空间中的单个点,进而在单个空间内通过用户-物品之间的距离来表征用户-物品之间的相似性.由于在不同的环境下,用户对于同一种类型的物品的偏好也可能存在差异.基于单个空间的用户、物品嵌入向量有可能无法满足用户具有的多种偏好和物品具有的多种属性,进而限制了推荐系统的性能.为了充分刻画用户和物品,我们尝试从多个侧面对于用户和物品进行表示,并提出了一个基于多空间的度量学习(MML)框架.通过设计整合多个空间相似性的度量方式,我们将用户和物品投影到多个空间中进行细粒度的表示.另外,我们设计了一种经过校准的优化策略,包括经过校准的最大间隔损失函数和经过校准的采样方法.在保持多空间度量学习表示能力的同时,确保框架的有效性.最后,模型通过训练好的用户、物品向量,对于稀疏的用户-物品交互矩阵进行填补.在动态更新空间权重的同时,可以赋予模型新的训练视角,最终实现端到端的训练.通过四个真实世界推荐数据集上进行的大量实验表明,MML可以在 Recall 和 nDCG 衡量指标上将目前最优的对比算法提高 40%以上.

关键词 度量学习;多侧面;隐式反馈;推荐系统;矩阵填补

中图法分类号 TP18

DOI号 10.11897/SP.J.1016.2022.00001

Multi-Space Recommender Systems via Metric Learning

TAN Yan-Chao ZHENG Xiao-Lin WEI Xiang-Yu Carl Yang

(College of Computer Science, Zhejiang University, Hangzhou 310007)

(College of Computer Science, Emory University, Atlanta 30322 USA)

Abstract Compared with explicit feedback, implicit feedback is more abundant and easier to obtain, which is widely explored by modern recommender systems. However, the implicit feedback is often sparse, imbalanced, and has ambiguous meaning, which poses great challenges to the learning of complex interactions among users and items. Based on the interaction matrix, one of the most popular methods is matrix factorization, which has been widely studied and applied to model user preferences and item properties. However, the performances of these methods are restricted due to the implicit feedback, since most of them can only treat the missing entries as negative feedback. Moreover, the matrix factorization methods cannot satisfy the triangle inequality. In other words, it can only model the similarity between users and items, which cannot be transferred to model the similarity of user-user and item-item relationship. To overcome this limita-

tion, the recommendation models based on metric learning aim to capture some relationships beyond user-item interactions from implicit feedback, which have been popular and achieved great performance in many fields such as computer vision. However, the existing methods based on metric learning for recommendation represent users and items as single points in the metric space, and then the preference is modeled by the distance between the user and the item. In this case, the model ignores the fact that users can have multiple preferences, and items can have multiple properties, which cannot be modeled by single vectors and limit the model's performance in the recommendation. To fully capture and exploit the multiple facets of user preferences and item properties, we propose a novel framework of Multi-space Metric Learning (MML) for the recommendation. Firstly, by designing a cross-space similarity measurement, we project users and items into multiple spaces for fine-grained representation. The different spaces aim to capture different facets of users and items, which alleviate the mix-up problem of both single user's and item's embedding. Secondly, based on the increasing dimension of the user and item embedding, it is important to alleviate the learning cost of the model. Considering the effectiveness and efficiency of the multi-space modeling, we design a calibrated optimization strategy, including a calibrated margin loss function and a calibrated sampling method. Finally, to fully integrate the learned user-item, user-user, and item-item relationships for further mining the multi-facet of user and item, we design an iterative loop mechanism for the End-to-end training. By imputing the missingness of the user-item interaction matrix, we can update the multiple spaces projection of users and items, which can provide a new perspective for the proposed MML model to mine the relationship between users and items. Extensive experiments on four real-world recommendation datasets show that MML can achieve up to 40% improvements over the state-of-the-art baselines regarding both Recall and nDCG metrics.

Keywords metric learning; multi-space; implicit feedback; recommender systems; rating imputation

1 引言

随着 Web 上各种应用程序以及网络信息的数量快速增长,推荐系统已经成为帮助用户减轻信息过载问题的基础.在推荐系统中,信息通常包含显式反馈和隐式反馈.显示反馈包括正反馈和负反馈,而隐式反馈指的是仅收集交互为“1”的反馈.隐式反馈中,所有负反馈与缺失值都统一反馈为空(隐式反馈的定义如图 1 左侧中所示).传统的推荐方法主要基于矩阵分解^[1-5],这些方法基于显式反馈取得了不错的效果,但是在隐式反馈的基础上还有存在挑战.尽管与显式反馈(例如:1-5 星评级)相比,隐式反馈(例如购买记录和浏览历史记录)在当今的现实世界系统中具有更为丰富和更容易获取的特性.但是,隐式反馈中的正反馈通常极为稀疏且正负反馈的数量不平衡.同时,负反馈与缺失值混合在一起,也容易让推荐系统在区分用户偏好时产

生歧义.例如,用户 i 对物品 j 的标注是空既有可能由于用户 i 不喜欢物品 j 而导致的未给出评分,也有可能由于用户 i 没有见到过物品 j 而导致的未给出评分.

为了对没有交互过的物品进行排序,传统矩阵分解使用点积拟合用户-物品之间的相似度,并以此学习用户、物品对应的嵌入向量.但是,点积不是度量学习,因此无法传达关键的不等式属性和处理缩放比例不同的输入变量.我们在图 1 的示例中说明了点积的缺点,其中隐空间维度为 d 等于 2.假设我们需要对于蓝框中的用户 u_2, u_3 和三部电影 v_1, v_2, v_3 建模,我们可以看到 u_2 和 u_3 相似(即都与 v_1 和 v_2 有交互).在这样的情况下,用户 u_2 (未与 v_3 有交互)也可能对电影 v_3 感兴趣.然而,基于点积学习,可能会出现以下结果: $u_2 = (0, 1), u_3 = (1, 0), v_1 = (1, 1), v_2 = (1, 1), v_3 = (1, -1)$, 因为 $u_2^T v_1 = 1, u_3^T v_1 = 1, u_2^T v_2 = 1, u_3^T v_2 = 1, u_3^T v_3 = 1$. 由于

u_2 和 v_3 之间的点积为 -1 , 因此 v_3 不会被推荐给 u_2 . 这样的结论就和传统的协同过滤有冲突. 但是, 如果我们使用度量学习, 则会得到与基于点积计算相似度相反的结果. 通过使用不等式属性将相似的用户-物品对拉近(如图 1 最右所示), u_2 和 v_3 之间的距离将缩小为 0, 在这样的情况下, v_3 将会被推荐给 u_2 .

由于度量学习既能捕获用户-物品之间的交互关系, 又能捕获复杂的、细粒度的用户-用户和物品-物品关系, 能够更加全面地建模用户和物品^[6]. 因此, 学术研究和工业应用中都提出了许多基于度量学习的算法来探索隐式反馈以进行推荐^[6-9]. 尽管基于度量学习的推荐系统在性能上已经有所提高, 但目前的方法都是在单个空间内投影用户和物品, 并基于在单个空间表征所有的用

户-物品交互关系. 这容易忽略由用户的多个偏好和物品的多种属性带来的多种用户-物品关系, 具体可以体现在由不同的用户物品空间分布带来的偏好冲突问题. 如图 2 左侧所示, 在单空间内, 基于图 1 继续训练, 能得到单空间内训练好的 u_1, u_2, u_3 和 v_1, v_2, v_3, v_4, v_5 的局部最优分布. 此时, 可以观测到 u_3 和 v_4 的距离和 u_3 和 v_5 一致, 与实际情况中 u_3 和 v_4 有交互而 u_3 和 v_5 无交互互相冲突, 从而间接导致用户建模不完善的情况. 如果用户和物品能投影至多个空间, 那么同一个用户或物品将会拥有多种不同的空间分布, 因此可以建模出不同的用户-用户以及物品-物品分布. 根据图 2 右侧基于多空间计算相似度的分布可得, 用户 u_3 和 v_4 以及 u_3 和 v_5 的对应物品距离可以同时被满足.

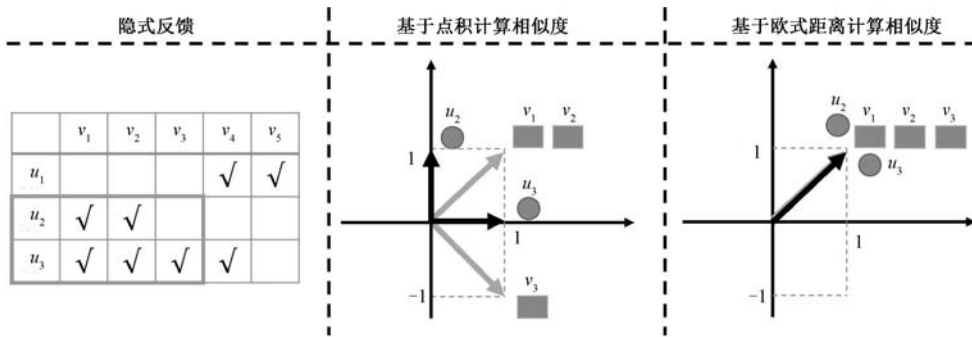


图 1 隐式反馈下基于点积进行学习和基于欧式距离进行学习的对比

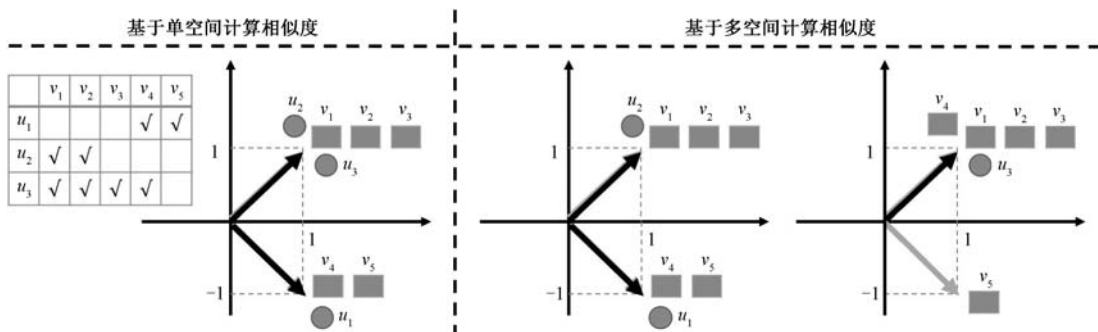


图 2 隐式反馈下单空间投影和多空间投影的对比

此外, 传统的度量学习在正样本和负样本之间采用全局固定的最大间隔, 该间隔没有把用户和物品的多样性考虑在内. 在许多现实世界的推荐系统中, 用户可以具有多个偏好, 物品也可以具有多种属性(统称为多侧面). 例如, 以电影作为物品举例, 一部电影可以属于浪漫爱情电影类别又属于喜剧类别, 推荐系统可以根据这两个物品属性推荐给不同的用户. 考虑到用户 u_1 可能被电影的浪漫情节吸引, 而用户 u_2 可能被有趣的演员表演吸引. 虽然最

终用户 u_1, u_2 均点击了该电影, 但是原因却可能不同. 因此, 我们认为, 从多个侧面全方位建模用户和物品在推荐系统中非常重要, 但这也带来了以下两个不可避免的技术挑战: (1) 如何基于隐式反馈, 从多个侧面合理地表示用户潜在的多偏好以及物品潜在的多属性; (2) 多侧面的建模将会导致表示向量的维度上升, 在这样的背景下, 如何设计有效的优化策略, 从而支持推荐系统从多个侧面学习用户、物品的表示以获取准确的推荐结果.

为了解决上述两个挑战,我们提出了一个基于多空间的度量学习(MML)框架.如图2右侧所示,通过将用户和物品投影到多个空间中,使得一个空间对应于一个侧面.基于多空间的框架,我们得以从多个侧面使用度量学习进行推荐.由于用户和物品可能在多个侧面观测下均有对应含义,因此,为了支持用户/物品可以被分入多个空间,本框架利用非负矩阵分解(NMF)来进行软聚类,并以此获得用户和物品对于多个空间的初始权重.在此过程中,无需借助其他辅助信息.而在初始化之后,本框架则支持NMF获取经模型训练填补后的矩阵,动态更新用户和物品的空间权重.基于学习到的多空间权重,MML可以通过提出的多空间投影方法来解决在图2中无法满足用户 u_3 同时喜欢物品 v_3 和物品 v_4 的挑战,即不同空间中,基于不同侧面表示出用户对于物品的偏好.具体如图2右侧所示,两个投影空间中物品 v_3 和物品 v_4 之间的距离不同,因此可以同时满足多个用户的多个偏好.为了确保多空间投影的结果共同作用于推荐效果,我们设计了一种单调的跨空间相似性度量,从而在统一多空间建模结果的同时,支持对用户、物品的细粒度度量.此外,针对多空间建模带来的参数增加等训练问题,我们设计了一种经过校准的优化策略,以确保进行有效的模型训练.具体而言,提出的MML可以于多空间对用户和物品进行细粒度建模,并且权衡训练的开销和效率,分别对样本的损失函数和正负样本的采样过程进行校准.我们在四个真实世界的推荐数据集上进行了广泛的实验,验证了我们提出的MML框架的有效性.

本文的主要贡献概述如下:

- 我们研究了基于隐式反馈的推荐系统面临的挑战,并发现现有方法在针对用户的多偏好和物品的多属性进行建模和优化时的低效率.
- 我们提出了一种新颖的多空间度量学习(MML)框架,该框架包含跨空间相似性度量和经过校准的优化策略,可以有效地基于多空间投影提升推荐性能.
- 综合实验结果清楚地证明了MML与各种最新推荐对比方法相比的优越性.

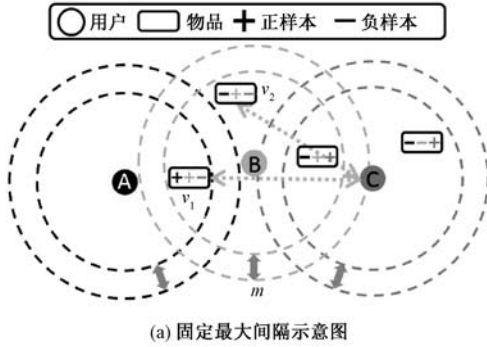
2 相关工作

相较于显式反馈,隐式反馈通常更为丰富且存在较少偏差^[10].基于隐式反馈的传统推荐方法一般依赖统一性假设^[11-12],即认为所有未标记数据均为

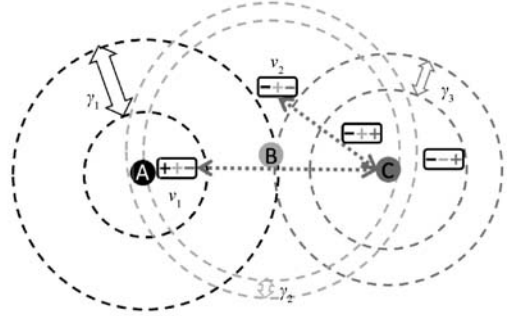
负样本,并分配统一权重以降低数据缺失对于建模的影响.然而,在这种情况下,造成数据缺失的原因既可能是用户没有被曝光过对应的物品,也有可能是因为用户看到了对应的物品,但是用户对于该物品没有兴趣.这两种不同原因造成的数据缺失对于推荐的影响是相反的.如果是用户没有被曝光过的物品,那么推荐系统后续可能需要推荐该商品,但是如果是由于用户不感兴趣,那么推荐系统应该尽可能回避推荐该物品.因此,这种统一将缺失数据的物品都当作负样本并降低数据权重的假设通常是不合理的,并且限制了推荐系统的准确性.目前,矩阵分解已成为推荐的流行方法,该方法使用内积运算,对用户-物品的相似性进行建模^[12-13].基于矩阵分解,有两种经典的top-N推荐模型:加权正则化矩阵分解^[1]和贝叶斯个性化排名^[5].两者都可以推广到解决个性化物品的排序问题.但是,矩阵分解假定用户和物品之间存在线性关系,这将限制推荐系统的推荐性能.因为实际场景中,用户和物品之间的交互通常比线性关系复杂得多^[14];与此同时,矩阵分解使用的内积在数学上不能满足三角不等式,这样的不稳定性容易限制模型捕获细粒度的向量表征,即用户-物品的相似性无法传递到用户-用户和物品-物品相似度的计算^[6].

最近,基于度量学习的推荐算法引起了广泛的研究关注^[6,15-18].这类方法能够在用户和物品之间寻求适当的距离函数来替代点积,并确保选择的度量方式满足三角不等式.这在很大程度上解决了矩阵分解不能度量用户-用户以及物品-物品之间相似度的局限性.基于欧式距离,Hsieh等人首先提出了一种称为协同度量学习(CML)的方法^[6],该方法致力于学习一种基于深度学习的联合度量空间,通过“推”的方式调整(用户,正样本,负样本)之间的相对距离,在建模用户-物品的相似度的同时,还可以对用户-用户和物品-物品之间的相似性进行建模.尽管CML取得了成功,但模型仍有一些局限,许多后续研究都将其作为基准,进行对应的改进.例如,为了避免将可能的推荐候选集“推”到过远的地方,Zhang等人提出了一种基于度量的分解方法^[9].与用“推”操作来调整用户-物品之间距离的CML相比,该模型改为通过“拉”操作来调整对应距离.由于CML没有考虑到用户-物品之间的关系是一对多的映射关系,使得一个用户无法同时“推”到多个与其交互的物品附近.受到翻译机制在知识图谱等模型上取得成功的启发^[19],Park等人考虑使用多个

一对一映射替代 CML 模型的一对一映射^[7]. 通过利用用户和物品的邻域信息构造用户的翻译向量^[7], 使得模型在将负样本“推”离用户的同时, 而且还可以将用户“拉”近所有与其交互的物品. 由于



(a) 固定最大间隔示意图



(b) 经过校准的最大间隔示意图

图3 固定最大间隔和经过校准后最大间隔之间的对比. v_1 的“++”表示用户 A 和 B 与 v_1 交互, 而用户 C 没有与 v_1 交互. v_2 的“--”表示用户 B 与 v_2 交互而用户 A 和 C 没有.

尽管以上基于单个空间度量学习方法在推荐中已经取得了不错的推荐效果, 但是它们在表示具有多侧面性质的用户和物品上存在局限. 最近, 在图嵌入等相关领域的研究表明, 使用多个嵌入空间建模具有必要性和有效性^[21-24]. 值得注意的是, 本文提出的多侧面学习与传统意义上的多视角学习之间存在差异. 多视角学习主要是致力于利用额外信息, 例如文本等词向量, 从数据的不同层面对用户进行分析, 从而缓和单一推荐视图导致的用户画像建模粗放等问题^[23-24]. 但这样的多视角学习对于数据的要求较高, 需要获取不同模态、不同来源、不同背景等形式的数据进行建模. Tan 等人^[25]提出在多个空间中进行度量学习, 但是该模型不适用于有空间有重叠的建模场景. 而本文提出的多侧面模型是在不需要额外数据的背景下, 针对基础的用户-物品交互矩阵进行有重叠的多空间建模.

当前推荐系统中基于度量学习的优化策略, 大多采用均匀采样和固定最大间隔的损失函数^[7,9]. 相比于广泛使用困难样本挖掘来改善度量学习中正负样本不均衡的相关领域(如计算机视觉领域)^[26-27], 这样的优化方式没有充分关注难以分类的困难样本. 受他们的启发, Tran 等人提出了一种基于物品的两阶段负采样方法, 用以替代均匀采样^[28]. 通过二次筛选充分挖掘含信息量的三元组, 从而提升训练效果. 但是, 在推荐系统的建模中, 用户和物品对于建模的有效应该是同等重要的^[29]. 因此, 仅挖掘信息量大的物品的采样方法是不充分的. 此外, 针对不同用户共享相同最大间隔的不合理性,

Tikhonov 等人证明了 CML 是一个不稳定的代数系统^[20], 即 CML 几何评分上存在限制, Tay 等人基于记忆网络和注意力机制改进了潜在的用户-物品交互关系, 以此减轻 CML 潜在的几何问题^[8].

Li 等人设计了一种自适应间隔, 旨在动态调整用户和物品之间的间隔^[30]. 但是, 引入额外参数进行联合学习自适应间隔通常是困难且不稳定的.

3 问题与挑战

3.1 问题定义

基于隐式反馈矩阵 $\mathbf{X} \in \mathbb{R}^{M \times N}$ 的推荐系统中, M 表示用户数量, N 表示物品数量. 当 $X_{uv} = 1$ 时, 我们认为对应的用户-物品对 (u, v) 是正样本, 当 $X_{uv} = 0$ 时, 则认为对应的 (u, v) 信息丢失. 为了在多个度量空间中表示用户和物品的嵌入向量, 我们令每个用户 u_i 的嵌入向量属于集合 $\mathbf{U}_i \in \mathbb{R}^{K \times D}$, 同样的, 物品 v_j 的嵌入向量属于 $\mathbf{U}_j \in \mathbb{R}^{K \times D}$, 其中 K 代表度量空间的数量, D 是嵌入向量的维数. 具体来说, 在第 k 个度量空间中, 用户和物品的嵌入向量可以表示为 $\mathbf{U}_i^k \in \mathbb{R}^{D \times 1}$ 和 $\mathbf{V}_j^k \in \mathbb{R}^{D \times 1}$.

3.2 挑战与动机

对于深度学习模型, 大间隔最近邻居 (LMNN) 是一种广泛采用的成对损失函数^[15], 它使用的三元组损失包含“推”操作和“拉”操作. 以 CML 为例^[6], 度量学习模型广泛使用“推”和“拉”操作来调整正负样本之间的距离, 具体公式如下:

$$\mathcal{L}_m(d) = \sum_{(u, v_p) \in I} \sum_{(u, v_q) \notin I} [m + d(u, v_p)^2 - d(u, v_q)^2]_+ \quad (1)$$

其中, $d(u, v_p)$ 是用户 u 与正样本 v_p 之间的距离, 而 $d(u, v_q)$ 是 u 与负样本 v_q 之间的距离. I 表示

用户交互物品的集合. 三元组损失的主要思想是将负样本对 (u, v_q) “推”离正样本对 (u, v_p) , 而且这个距离至少要间隔 m . 如图 3(a) 所示, 固定间隔意味着不同偏好的用户将共享相同的用户-物品间隔 m .

尽管基于三元组损失的 CML 模型在度量学习中取得了不错的成效, 但该模型仍然存在以下局限性:

首先, 现实生活中用户和物品都可以通过多个侧面进行刻画, 即用户可以针对物品的多种属性拥有多种偏好. 但 CML 无法实现将用户多种偏好下的物品同时“拉”到用户附近^[6]. 同时, 模型仅将用户和物品投影到空间中的单个点, 导致用户不同侧面的偏好可能彼此发生冲突, 进而影响整体偏好的建模.

其次, 全局固定的最大间隔将会限制用户偏好的多样性和传递性. 例如, 在图 3(a) 中, 基于用户 A 和 B 对物品 v_1 和 v_2 的偏好, 固定间隔 m 会使得用户 C 对物品 v_1 和 v_2 的偏好难以区分.

最后, 度量学习中备选的三元组的数量与用户和物品的数量呈立方项相关^[31]. 为了避免遍历的开销, 传统的度量学习模型采用随机采样的方式对负样本进行选取, 这种方式采集到的三元组信息含量具有随机性, 并容易让模型陷入局部最优^[32]. 此外, 在正负样本不均的数据集中, 单纯的随机采样策略可能导致选择的三元组都是一些信息含量少的样本, 即带来损失梯度可忽略的三元组, 这样的三元组对于训练没有任何帮助, 仅仅是增加了对应模型收敛的时间.

为了解决上述挑战, 我们设计了一个基于隐式反馈的推荐框架, 包括以下部分:

- 我们将用户和物品投影至多个空间来缓解第一个局限. 考虑到单个空间不方便表示用户和物品的多个侧面, 因此我们利用多个空间分别表示用户和物品的多个侧面, 并设计了跨空间相似性度量, 通过一维的隐式权重将多个空间的结果融合.

- 我们通过校准用户物品的最大间隔以解决第二个局限. 如图 3(b) 所示, 每个用户都将拥有一个能体现其个性化特征的最大间隔. 在这种情况下, 用户 C 的最大间隔介于用户 A 和 B 之间. 于是, 用户 C 和物品 v_2 之间的距离可以小于其与 v_1 的距离, 由此得以清晰地反映出用户 C 对 v_2 的偏好大于 v_1 .

- 我们提出一种经过校准的采样方法来替代随机采样. 通过校准采样, 模型更加关注具有多个侧面的用户和物品, 这很大程度上缓解了第三个局限.

4 模型

在本节中, 我们将详细介绍我们设计的多空间度量学习 (MML) 框架. 具体而言, 我们首先定义跨空间相似性度量, 该度量允许模型基于隐式反馈, 使用一维隐式交互向量, 将用户和物品投影到多个空间中. 然后, 本文提出的校准优化策略包括经过校准的最大间隔损失函数和经过校准的采样方法. MML 的总体框架如图 4 所示.

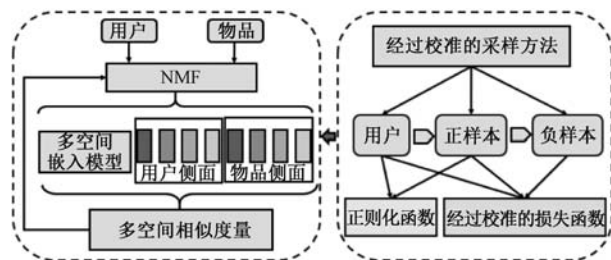


图 4 本文提出的 MML 模型的框架
(其中斜体字指代模型设计部分)

4.1 跨空间相似性度量

我们观察到, 用户的不同行为来自于他们对于自己不同偏好的表达. 传统的推荐算法通常会将这些不同的偏好带来的多个侧面合并到单个向量空间中进行表示^[22]. 如图 2 中所示, 用户可能同时喜欢多个类别的电影, 而电影本身可能属于多个类别. 在单个度量空间中, 不同类别的电影不太可能互相接近, 并且用户可能无法做到同时接近不同类别的电影.

受语言建模中单词具有歧义性并且需要将歧义考虑进嵌入向量的启发^[33], 我们将用户和物品投影到多个度量空间中, 从而对他们的不同侧面进行建模. 这样, 用户 u_i 的嵌入向量 $\{U_i^k\}_{k=1}^K$ 可以同时拉进代表不同侧面偏好的不同物品. 同理可得, 物品 v_j 的嵌入向量 $\{V_j^k\}_{k=1}^K$ 也支持物品被不同偏好的用户喜欢. 尽管现在用户和物品的表示已经投影到了多个空间, 但是现有的相似性度量依然是基于单个空间的. 为了整合多个空间的结果来度量用户和物品之间的相似性, 我们提出了一个单调的跨空间相似性度量方式. 该度量通过一个一维用户-物品交互矩阵, 反映用户和物品对于多个空间的不同权重. 单调性确保了用户和物品之间的距离越小, 它们之间的相似性越高. 用户 u_i 和物品 v_j 在 K 个空间累积相似度的定义如下:

$$s(u_i, v_j) = - \sum_{k=1}^K p(k | (u_i, v_j)) \| \mathbf{U}_i^k - \mathbf{V}_j^k \|^2 \quad (2)$$

其中, $\mathbf{U}_i^k \in \mathbb{R}^{D \times 1}$ 和 $\mathbf{V}_j^k \in \mathbb{R}^{D \times 1}$ 分别是用户和物品等待学习的嵌入向量. $p(k | (u_i, v_j))$ 表示用户 u_i 和物品 v_j 对于第 k 个空间的联合权重, 具体计算方式我们将在后续的章节中介绍.

4.2 多空间权重

在多空间的假设下, 一个用户由 K 个空间组成, 每一个空间对应一个有针对性的侧面. 具体来说, 每一个用户要么完全属于一个空间, 要么或多或少地落入几个空间中, 即每个用户可以表示为 K 个空间的线性组合. 同理可以得到物品的精确建模. 受文档分类的启发^[31,34-35], 此时的线性组合对应的具体含义是用户基于每个空间属性的叠加混合, 因此所有的线性组合都应取非负值. 此外, 用户/物品对应的空间并不完全彼此独立也是很普遍的, 在这种情况下, 捕获用户/物品每个属性的空间轴也不一定是正交的. 基于以上讨论, 我们提出非负矩阵分解(NMF)来得到用户/物品的潜在空间, 以及对应每个空间的权重大小. 用户/物品对应的空间权重具体定义如下所示:

$$\min_{P \geq 0, Q \geq 0} \| \mathbf{X} - \mathbf{P}\mathbf{Q}^T \|_F^2 + \tau (\| \mathbf{P} \|_F^2 + \| \mathbf{Q} \|_F^2), \quad \| \mathbf{P} \| \leq 1, \| \mathbf{Q} \| \leq 1 \quad (3)$$

其中 $\mathbf{P} \in \mathbb{R}^{M \times K}$, $\mathbf{Q} \in \mathbb{R}^{N \times K}$, $\| \cdot \|_F^2$ 表示平方和, τ 是 L-2 正则化的权重. 其中 $P_{u_i, k}$ 是第 k 个空间和用户 u_i 之间的原始用户空间权重, 而 $Q_{v_j, k}$ 是第 k 个空间和物品 v_j 之间的原始物品空间权重. 在实验中, 我们设置 $\tau = 0.05$. 根据对应矩阵含义, 我们需要对分解得到的 \mathbf{P} 、 \mathbf{Q} 进行归一化, 确保 \mathbf{P} 、 \mathbf{Q} 中的值在 $[0, 1]$ 之间:

$$p(k | u_i) = \frac{P_{u_i, k}}{\sum_{c=1, \dots, K} P_{u_i, c}} \quad (4)$$

$$p(k | v_j) = \frac{Q_{v_j, k}}{\sum_{c=1, \dots, K} Q_{v_j, c}} \quad (5)$$

由于空间权重 $p(\cdot | u_i)$ 和 $p(\cdot | v_j)$ 描述的是将原始的用户 u_i 和物品 v_j 的分布经过归一化后得到的归一化空间权重分布. 受空间权重的可叠加性以及协同过滤思想的启发, 我们将用户-物品联合权重定义为用户和物品的平均空间权重:

$$p(k | (u_i, v_j)) = \frac{p(k | u_i) + p(k | v_j)}{2} \quad (6)$$

该公式用于衡量每对(用户, 物品)被投影到每个空间的概率, 即每对用户-物品面向每个空间的权重.

此外, 为了缓解交互矩阵的稀疏问题, 我们考虑迭代式填补矩阵中的空缺值. 具体来说, 我们先利用观测到的用户-物品交互矩阵 \mathbf{X} 得到初始的空间权重矩阵, 此时 $\mathbf{X}_{ij} \in \{0, 1\}$. 此后, 模型可以将公式(2)得到的预测值进行归一化, 并动态更新交互矩阵. 此时 $\mathbf{X}_{ij} \in [0, 1]$, 填补后的矩阵信息将越来越丰富. 基于动态变化的交互矩阵 \mathbf{X} , 模型可以继续通过非负矩阵分解得到用户空间权重 \mathbf{P} 和物品空间权重 \mathbf{Q} .

4.3 经过校准的优化策略

为了有效、准确地学习用户和商品的多侧面特征, 提高推荐效果, 我们提出了一种经过校准的优化策略, 其中包括经过校准的最大间隔损失函数和经过校准的采样方法.

4.3.1 经过校准的最大间隔损失函数

传统的度量学习方法通常将最大间隔设置为三元组损失的超参数, 如图 3(a)所示. 但是, 这样的固定间隔不能反映出用户的不同偏好, 进而会影响训练的效果. 由于现有的损失函数不能适应用户的多种偏好和本文提出的多空间度量框架, 因此我们设计了经过校准的间隔损失函数.

由于最大间隔具有相应的物理含义, 我们基于接受度的概念设计了经过校准的最大间隔, 从而改善多空间用户的投影效果. 接受度是指用户接受新事物的程度. 通过调整正样本和负样本之间的最大间隔, 接受度可以建模用户在接纳新事物上的区别. 另外, 为了改进多空间建模下交互矩阵稀疏的问题, 我们为(用户, 物品)对的损失设计了对应的置信度. 置信度是通过考虑物品被曝光给用户的概率, 进而给予采样出的三元组不同的权重.

接受度: 用户接受度 γ_u 表示用户对新事物的态度. 对于经常接纳新事物的用户来说, 正负样本之间的距离应该较小, 对应的具体含义是用户较容易点击未交互过的物品; 而对于不常接纳新事物的用户来说, 正负样本之间的距离则比较大. 由于 m 定义了度量空间中正样本和负样本之间的最大间隔, 因此, 用户接受度 γ_i 和最大间隔之间呈负相关. 为了让模型从隐式反馈中计算用户 u_i 的接受度 γ_i , 并自适应地设置用户对应的个性化最大间隔, 我们定义

$$\gamma_i = 1 - \frac{\sum_{v_j \in \mathcal{V}_i} | \mathcal{U}_j |}{M}, \gamma_i \in [0, 1] \quad (7)$$

其中 \mathcal{U}_i 表示与用户 u_i 交互的物品集, 而 \mathcal{U}_j 表示与物品 v_j 交互的用户集, M 表示用户数. 公式(7)背后的想法是在用户-物品的二分图上, 利用两跳邻居(例如, 用户-物品-用户)来表示用户的接受度. 即用户的两跳邻居越多, 说明用户的偏好就越多, 也因此反应出用户对于新事物的接纳度越高. 当用户总数为 N 时, 其邻居用户数量的最大值为 $N-1$, 因此 γ_i 的取值在 $[0, 1]$ 区间内.

置信度: 置信度 α 表示当前采样的三元组对应损失的置信度. 由于物品被曝光给用户的概率不同, 在不同空间中区分标签为“?”的样本是源于用户的不喜欢还是用户未被曝光对应物品是很必要的. 基于相似用户行为的传递性, 如果有很多点击过当前正样本的用户也点击过当前的负样本, 则很大概率上, 用户未与当前缺失标签的样本 v_q 有交互是由于用户的不喜欢, 因此该三元组的置信度是高的, 与之对应的损失权重也应该更大. 而正负样本对应的用户交集小, 则代表当前用户很可能未被曝光过 v_q , 在这样的情况下, 应通过调小置信度来降低“推”远 v_q 的力度. 基于元组置信度和正负样本交集呈正相关的考虑, 模型从交互数量的交集中计算元组的置信度 α , 所以我们针对用户 u_i 定义

$$\alpha_i = \frac{|\mathcal{U}_p \cap \mathcal{U}_q|}{|\mathcal{U}_p \cup \mathcal{U}_q|} \quad (8)$$

其中 \mathcal{U}_p 表示与当前正样本 v_p 交互过的用户集合, \mathcal{U}_q 表示与当前负样本 v_q 交互过的用户集合.

基于多空间用户物品度量框架(公式(2)), 我们设计了一个经过校准的最大间隔损失函数:

$$\mathcal{L}_{push}(s) = \sum_{(u, v_p) \in I} \sum_{(u, v_q) \notin I} \alpha_i [\gamma_i - s(u_i, v_p) + s(u_i, v_q)]_+ \quad (9)$$

其中 I 表示交互过的用户物品集合. 因此, (u_i, v_p) 是正样本对, 而 (u_i, v_q) 是负样本对. $s(\cdot, \cdot)$ 在公式(2)中定义, γ_i 在公式(7)中定义, α_i 在公式(8)中定义.

考虑到用户和物品已经在多个度量空间中表示, 所以可以缓解 CML 模型中无法将用户喜欢的物品“拉”近用户的限制(3.2节). 因此, 我们可以如下添加一个简单的逐点正则化函数, 其目的是通过在不同空间中拉近用户和物品来突出用户的喜好, 同时也可以和基于三元组的“推”操作相平衡:

$$\mathcal{L}_{pull}(s) = \sum_{(u, v_p) \in I} -s(u_i, v_p) \quad (10)$$

最后, 为避免过度拟合的问题, 我们将欧几里得球面的约束放宽到在约束点在球内部即可^[9], 并获得最终的目标函数:

$$\min_{u_*, v_*} \mathcal{L}_{push}(s) + \lambda_{pull} \mathcal{L}_{pull}(s), \\ ||u_*||^2 \leq 1 \text{ and } ||v_*||^2 \leq 1 \quad (11)$$

4.3.2 经过校准的采样方法

正如我们在 3.2 节中讨论的那样, 由于遍历所有的三元组开销很高, 因此在度量学习中广泛通过随机采样来选择样本. 但是, 与分类等其他任务相比, 基于隐式反馈的推荐更容易出现样本含义不明确以及正负样本不平衡的情况, 进而容易导致随机采样下样本的信息量少, 损失梯度小. 在多空间的度量框架下, 梯度是由 K 个空间加权组成, 在这样的情况下, 回传到对应空间的梯度更小, 容易导致训练困难. 以计算机视觉中分类任务为例, 该任务正样本和负样本的标签是明确的, 并且可以使用公式(1)直接训练模型. 然而, 在基于隐式反馈的推荐系统中, 负样本包含用户不喜欢和未被曝光的物品. 尽管两个样本都标记为缺失样本, 但候选集应该由未被曝光的物品组成, 而不是由已被曝光但用户不感兴趣的物品组成. 在这种情况下, 如果所有负样本的权重相同并通过公式(1)优化目标, 则未被曝光的物品可能会和用户不感兴趣的物品一起被用户“推”开, 进而影响推荐的性能. 此外, 由于流行度更高的用户和物品对应着更多的交互. 例如, 一部流行的电影就意味着有许多人观看. 考虑到该电影能够满足不同用户的不同偏好, 则对应着该电影具有丰富的属性/侧面. 相比于小众的用户和物品, 流行度高的用户/物品不仅对应更多的侧面, 也更容易影响每个空间内的分布. 因此, 将用户和物品的流行度考虑进建模十分必要.

考虑到带权采样和基于流行度建模的必要性, 本文采用基于流行度的采样方法^[28-29, 36-37]进行带权采样. 但是, 目前用于推荐系统的流行度采样大多基于物品的流行度, 而未考虑到用户的流行度也会影响训练. 为了有效率地利用用户的多偏好性质, 我们设计了经过校准的采样方法.

相较于只关注流行物品的采样, 本文提出的经过校准的采样方法也关注具有丰富偏好的用户. 为了确保在多个空间内充分挖掘偏好丰富的用户, 我们表示第 k 个空间中用户 u_i 被采样到的概率如下:

$$Pr(u_i) = \frac{p(k | u_i) f(u_i)}{\sum_{n=1}^N p(k | u_n) f(u_n)} \quad (12)$$

其中 $f(u_i)$ 是用户 u_i 与物品的交互频率, 即 u_i 交互过的物品数量. N 是用户数, $p(k | u_i)$ 是用户 u_i 对于当前空间 k 的权重. 通过调整用户在不同空间内被采样概率, 从而实现在不同空间中探索不同的

用户-物品关系的目标。

根据被采样的用户,我们将继续按照经过校准的采样方法分别对正样本和负样本进行采样.进行物品采样时,我们有

$$Pr(v_j) = \frac{p(k | v_j)f(v_j)}{\sum_{m=1}^M p(k | v_m)f(v_m)} \quad (13)$$

值得注意的是,传统的基于单个空间利用流行度采样物品的方式,很容易造成用户和流行物品在度量空间中的混杂^[38].因此,模型往往需要进一步过滤和限制^[28].但是,基于本文提出的多空间度量框架,模型可以在不同的空间中有效地表示用户和物品的不同侧面.在这种情况下,经过校准的采样方法能够利用用户和物品对于不同空间的权重,基于侧面挖掘当前空间下的流行用户和物品.实现将聚集在单个空间中的热门用户和物品根据侧面分散到多个空间的目标.因此,多空间投影和经过校准的采样组合相比于单空间内采样,能显著提高性能^[28].

5 实验

在本节中,我们通过关注以下四个研究问题来评估我们提出的 MML 框架:

- 问题 1:与各种最新的推荐方法相比,MML 的效果如何?
- 问题 2:MML 中模型的不同部分对模型的影响是什么?(消融研究)
- 问题 3:不同的超参数如何影响最终推荐效果以及如何选择最佳参数?
- 问题 4:MML 如何利用用户和物品的多个侧面进行建模?(可视化)

5.1 实验准备

5.1.1 数据集

为了全面验证 MML 的有效性,我们使用来自不同领域的四个真实世界数据集,它们具有不同的大小和数据密度,包括 Ciao^①,Delicious₁,MovieLens^② 和 BookX₁.这些数据集已在之前的文献中被广泛采用^[6-7,9],其中基本的统计信息如表 1 所示.

表 1 数据集的基本信息

Dataset	# User	# Item	# Interaction	Density
Ciao	6.7K	11K	147K	0.19%
Delicious	1.1K	1.2K	7.7K	0.61%
MovieLens	6K	4K	1M	4.20%
BookX	19K	40K	605K	0.08%

5.1.2 评估协议

考虑到留一法^[7]涉及到随机采样,可能会造成评估误差^[39],本文采用传统的划分进行评估.具体来说,我们参考^[6],将评分数据集以 60%/20%/20%分为训练/验证/测试集.由于我们的推荐模型本质上是为了学习用户对于所有物品的偏好排序,因此,我们综合信息检索和推荐系统中的指标来对提出的模型性能进行评估.其中包括:召回率(Recall@50, Recall@100)^[6]、归一化折损累计增益(nDCG@50, nDCG@100)^[40].具体而言,Recall@X 指标是用户真实点击的物品中有多少物品被推荐了,而 nDCG@X 则是对顺序敏感的指标.两个指标都是越大越好.

5.1.3 对比算法

本文开展对比实验的目的有两个:(1)验证使用度量学习的推荐算法是否能够有效进行用户-物品之间以外的用户-用户、物品-物品的关系传递,缓解数据稀疏给推荐性能带来的影响;(2)相比同样使用度量学习的算法,验证本文提出的 MML 是否能够通过融合不同侧面的深层特征进一步提高推荐系统的性能.为了达到上述两个实验目的,我们将 MML 与 8 个目前较为先进的模型进行对比,具体罗列如下:

- BPR^[5]:贝叶斯个性化排名(BPR)模型是 Top-N 推荐的流行方法.
- NMF^[13]:非负矩阵分解(NMF)是一种经典的矩阵分解模型,可从交互数据中学习潜在的用户和物品向量.本文将 NMF 用于获取用户和物品对应的空间权重.
- NeuMF^[40]:NeuMF 是将神经网络应用于协作过滤的框架.它在其框架中结合了多个全连接层和矩阵分解.
- CML^[6]:协同度量学习(CML)是第一个使用深度度量学习解决推荐系统协同过滤问题的模型,也是许多基于度量学习的算法基础.
- MetricF^[9]:MetricF 是一种基于度量学习的推荐系统,其通过将用户偏好转换为距离来对用户-物品之间的交互进行建模.该算法使用的度量方式是欧式距离.
- TransCF^[7]:TransCF 借用知识图谱中的翻译思想来改进 CML,并通过学习用户与物品之间的多对关系向量来计算距离度量.

① <https://github.com/pcy1302/TransCF/tree/master/data>

② <https://grouplens.org/datasets/movielens/>

- LRML^[8]: LRML 使用额外的存储模块进行潜在关系度量学习,并为每个用户-物品交互引入潜在关系。

- SML^[30]: SML 算法设计了具有自适应最大间隔的度量损失函数,在度量学习中分别引入了对称的用户中心和物品中心,最终完成推荐。

5.1.4 实验细节

我们基于 Pytorch^① 实现本文提出 MML 框架。比较算法的实现来自开源项目或作者本身(BPR/MetricF/ CML^②, NMF^③, NeuMF^④, TransCF^⑤, LRML^⑥ 和 SML^⑦)。在具体训练中,由于使用 SGD 和 Adam 优化算法对于模型的训练差别较小,而 SGD 相比 Adam 计算复杂度低,迭代速度较快,因此我们使用 SGD 优化器优化 MML,并通过网格搜索调整验证集上的所有超参数。其中,空间个数 K 的设置范围是[1, 6],学习率的范围是{0.0005, 0.001, 0.005, 0.01, 0.1}, λ_{pull} 的范围是{0.001, 0.01, 0.1, 1},不同数据集中的嵌入向量的维度范围是{32, 64, 128, 256, 512, 1024}。对比算法的超参是按照原始论文描述并经过交叉验证调整以实现最佳性能。

表 2 整体推荐结果对比(最佳表现已加粗表示,最优的对比算法用下划线标出)

Dataset	Metric	BPR	NMF	NeuMF	CML	MetricF	TransCF	LRML	SML	MML	提升
Ciao	Recall@50	0.0769	0.0772	0.0763	0.0904	0.0855	0.1102	0.1006	0.1218	0.1557	27.83%
	Recall @100	0.1059	0.1031	0.1036	0.1219	0.1140	<u>0.1412</u>	0.1231	0.1269	0.2018	42.92%
	nDCG@50	0.0275	0.0284	0.0267	0.0424	0.0332	0.0445	0.0409	<u>0.0539</u>	0.0744	38.03%
	nDCG@100	0.0325	0.0342	0.0318	0.0498	0.0409	0.0530	0.0462	<u>0.0598</u>	0.0855	42.97%
Delicious	Recall @50	0.0541	0.0849	0.0646	0.0649	0.1088	<u>0.1742</u>	0.1732	0.1563	0.1841	5.68%
	Recall @100	0.1172	0.1698	0.1452	0.1202	0.1819	<u>0.2384</u>	0.2347	0.2133	0.2554	7.13%
	nDCG@50	0.0153	0.0325	0.0230	0.0185	0.0678	<u>0.0779</u>	0.0741	0.0726	0.0835	7.19%
	nDCG@100	0.0262	0.0480	0.0327	0.0287	0.0752	<u>0.0893</u>	0.0851	0.0863	0.0968	8.40%
Movielens	Recall @50	0.3301	0.3086	0.2882	0.3265	0.3198	0.3276	<u>0.3501</u>	0.3225	0.3659	4.51%
	Recall @100	0.4640	0.4437	0.4277	0.4634	0.4607	0.4619	<u>0.4849</u>	0.4597	0.4965	2.39%
	nDCG@50	0.2609	0.2433	0.2358	0.2749	0.2543	0.2575	<u>0.2865</u>	0.2611	0.3080	7.50%
	nDCG@100	0.3385	0.3011	0.2978	<u>0.3415</u>	0.3131	0.3164	0.3263	0.3178	0.3680	7.75%
BookX	Recall @50	0.0746	0.0539	0.0662	0.1037	0.0741	0.1225	0.1109	<u>0.1232</u>	0.1423	15.50%
	Recall @100	0.1201	0.1356	0.1192	0.1436	0.1162	0.1504	0.1492	<u>0.1516</u>	0.1952	29.78%
	nDCG@50	0.0212	0.0183	0.0207	0.0268	0.0226	0.0284	0.0274	<u>0.0310</u>	0.0379	22.26%
	nDCG@100	0.0262	0.0223	0.0235	0.0324	0.0285	0.0357	0.0345	<u>0.0369</u>	0.0446	20.87%

- 在 Ciao 和 BookX 上,基于度量学习的六个模型(CML, MetricF, TransCF, LRML, SML 和本文提出的 MML)优于基于矩阵分解的其他对比算法(BPR, NMF 和 NeuMF)。这主要是由于度量学习方法能捕获用户-物品关系之外,还可以很好地捕获用户-用户和物品-物品关系。较于其他基于单空间度量学习的算法,本文提出的 MML 可以从多个侧面建模,能有效建模用户和物品的细粒度特征。具体来看, MML 相较于其他度量学习算法有明显的提升,例

5.2 总体效果比较(问题 1)

表 2 显示了本文提出的 MML 和其他算法的指标对比。通过分析实验结果,我们有如下结论:

- 总体上看, MML 在所有评估指标上以及在所有数据集上均超过其他八个对比算法。这个结果回答了问题 1,表明 MML 具有有效的协同排名能力。此外,许多对比算法的性能相比于 MML 不是那么稳定,例如,不同数据集上的最佳的对比算法分散在不同模型。在 Delicious 数据集上, TransCF 是最佳对比算法,而在 BookX 数据集上, SML 是最佳对比算法。这是由于不同数据集的稀疏程度以及用户、物品数量有差异所致。与最佳对比算法相比, MML 模型在 Ciao 和 BookX 数据集上的性能提升范围从 15.50%(在 BookX 上通过 Recall@50 得到)到 42.97%(在 Ciao 上通过 nDCG@100 得到)。在, Delicious 和 MovieLens 中,相较于模型最佳对比算法 LRML, MML 性能提升较小,范围从 2.39%(在 Movielens 上通过 Recall@100 得到)到 8.40%(在 Delicious 上通过 nDCG@100 得到)。

如,在 Ciao 上相较于 SML 至少有 42.97%的提升(通过 nDCG@100 得到), BookX 上相较于 SML 至少有 29.78%的提升(通过 nDCG@100 得到)。

- 注意到 MML 无法在 Delicious 和 MoiveLens

① <https://pytorch.org/>

② <https://github.com/cheungdaven/DeepRec>

③ <https://github.com/ninghaohello/Polysemous-Network-Embedding/>

④ https://github.com/hexiangnan/neural_collaborative_filtering

⑤ <https://github.com/pcyl302/TransCF>

⑥ https://github.com/vanztyay/WWW2018_LRML

⑦ <https://github.com/MingmingLie/SML>

上获得与其他数据集一个数量级的提升,对应的分析如下.由于 Delicious 的数据集较小,且 Delicious 数据的数据密度相较于 Ciao 和 BookX 而言较大,这样的情况下,用户-物品的交互中对应的空值较少,因此数据挖掘以及点击预测的难度较小.而针对 MovieLens 而言,首先,MovieLens 的数据密度较大,同时,在电影推荐中,电影受欢迎程度在推荐结果中起着重要作用^[6],该影响程度甚至超过了基于用户-物品交互的建模.也正因为电影自身的特性,现有的度量学习算法在这种类型的数据集上的表现和基于矩阵分解的对比算法较为一致.而本文提出的 MML 仍然超过其他对比算法,因为 MML 通过校准采样方法,已经将用户和物品的受欢迎程度考虑在内.

• 表 3 显示了不同嵌入尺寸 d 下,TransCF, SML 和本文提出的 MML 的结果(Ciao 数据集). TransCF 和 SML 的嵌入向量对应的总维度数量等于 d ,而 MML 的嵌入向量对应的总维度数等于 $d * k$.我们已经观察到,在维度尺寸过拟合之前,单空间模型和多空间模型的性能都会随着 d 的增加而略有上升.但是,与单空间模型和 MML 之间的差距相比,嵌入尺寸的增加(例如 TransCF 上的 4.22%和 SML 上的 4.55%)带来的改进是微不足道的(在相同的总尺寸下,MML 在 nDCG @50 上优于 TransCF 高达 67.19%,在 Recall@100 上优于 SML 高达 59.02%),这支持了我们的观点,即正确使用多个嵌入空间比单纯增加单个空间的尺寸更有效.值得注意的是,当单个空间的尺寸增加到 1024 时,TransCF 和 SML 都有过度拟合的倾向,而 MML 仍可以进一步提高性能.

表 3 Ciao 上维度结果对比(单个算法最佳表现已加粗)

	Recall @50	Recall @100	nDCG @50	nDCG @100	d	k
TransCF	0.1077	0.1382	0.0427	0.0515	128	1
	0.1093	0.1401	0.0441	0.0525	256	1
	0.1102	0.1412	0.0445	0.0530	512	1
	0.1089	0.1402	0.0435	0.0486	1024	1
SML	0.1167	0.1250	0.0526	0.0572	128	1
	0.1205	0.1258	0.0532	0.0579	256	1
	0.1218	0.1269	0.0539	0.0598	512	1
	0.1214	0.1262	0.0530	0.0583	1024	1
MML	0.1320	0.1768	0.0634	0.0721	32	4
	0.1366	0.1803	0.0655	0.0736	64	4
	0.1406	0.1842	0.0683	0.0797	128	4
	0.1557	0.2018	0.0744	0.0855	256	4

5.3 模型消融研究(问题 2)

为了更好地理解 MML 中不同部分对于模型总

体性能提升的影响,我们设计了 MML 的五种变体,进行了全面的消融研究.

首先,我们将 CML 设置为 base 模型,以此为基础逐步添加 MML 中不同的设计组件,得到对应的变体如下所示:

- base+M:加入本文提出的多空间投影和跨空间相似度(M)的 CML;
- base+C:加入本文设计的经过校准的优化策略(C)的 CML;
- base+M+C:加入本文设计的组件 M 和组件 C 的 CML;
- base+M+C+P:base+M+C 模型加上正则化器 \mathcal{L}_{pull} ;
- base+M+C+P+N:base+M+C+P 模型加上迭代填补原始矩阵(N),动态更新用户和物品的空间权重分配,最终得到本文提出的 MML 完整模型.

表 4 显示了不同模型部分对模型性能的影响.基于表 4,我们有以下观察结果:

- 整体上看,5 种 MML 变体在所有指标上的表现都优于基准模型 CML.增加 MML 中组件导致的性能增益在不同数据集上遵循相同趋势.可以看到,单独加入多空间投影来改善模型的建模粒度,对应的推荐效果提升不明显.单独使用经过校准的优化策略,则可以给模型带来较为显著的提升.如果我们将这两个部分一起使用,推荐效果会优于单独使用多空间投影(M)和校准的优化策略(C).最后,正则化器 \mathcal{L}_{pull} 和动态更新空间权重还可以基于 M 和 C 的部分,继续提升模型性能,达到最终的 MML 效果.

- 值得注意的是,多空间投影(M)能够增加用户和物品的表示能力,但是其需要配合经过校准的优化策略(C),才能对模型有显著提升.例如,在 Ciao 上,仅通过应用多空间投影,base+M 只能将 base 的 Recall@50 略微提高,对应数值从 0.0904 提升至 0.1013.但是,通过进一步应用经过校准的优化策略,base+M+C 则可以实现 Recall@50 = 0.1502 的明显性能提升.base+M 的性能表明,传统的优化策略对多空间框架提不显著,即组件 M 不能通过均匀的随机采样提高性能.只有通过有效的训练,获得用户和物品的细粒度表示,模型才能充分体现出多空间投影的优势.因此,经过校准的优化策略(C)通过校准最大间隔损失函数,捕获用户的多个偏好,并通过校准采样方法从多个侧面考虑用户和商品的受欢迎程度,最终带来了性能提升.以

Ciao 为例, base + C 可以将 base 的 Recall@10 从 0.0904 显著提高到 0.1267.

• 因为正则化器 \mathcal{L}_{pull} 可以主动增强用户和他们喜欢的物品之间的交互, 并通过将正样本拉进目标用户来提高多空间框架的性能, 该部分也可以直接提升模型性能. 例如, base + M + C + P 性能在 Ciao 的 Recall@10 上可以达到 0.1535, 优于 base

+M+C 的 0.1502. 而矩阵填补的方式, 则可以将探索到的用户-物品关系通过矩阵填补, 动态更新空间权重. 与初始的 0-1 交互矩阵相比, 更新后的数值矩阵信息量更多, 能尝试将用户不感兴趣与未见过这两种情况加以区分, 使得模型效果再度提升. 例如, 在 Ciao 的 Recall@50 上将性能从 0.1535 提升至 0.1557.

表 4 不同模型部分对模型性能的影响

数据集	指标	CML	CML+M	base+C	base+M+C	base+M+C+P	base+M+C+P+N
Ciao	Recall@50	0.0904	0.1013	0.1267	0.1502	0.1535	0.1557
	Recall@100	0.1219	0.1334	0.1674	0.1884	0.1972	0.2018
	nDCG@50	0.0424	0.0489	0.0662	0.0723	0.0731	0.0744
	nDCG@100	0.0498	0.5752	0.0712	0.0810	0.0833	0.0855
Delicious	Recall@50	0.0649	0.0751	0.1581	0.1678	0.1807	0.1841
	Recall@100	0.1202	0.1413	0.2010	0.2305	0.2461	0.2554
	nDCG@50	0.0185	0.0253	0.0546	0.0712	0.0803	0.0835
	nDCG@100	0.0287	0.0362	0.0609	0.0803	0.0912	0.0968
MovieLens	Recall@50	0.3265	0.3302	0.3447	0.3513	0.3599	0.3659
	Recall@100	0.4634	0.4716	0.4810	0.4891	0.4900	0.4965
	nDCG@50	0.2749	0.2789	0.2819	0.2903	0.3020	0.3080
	nDCG@100	0.3415	0.3463	0.3534	0.3598	0.3602	0.3680
BookX	Recall@50	0.1037	0.1066	0.1242	0.1321	0.1402	0.1423
	Recall@100	0.1436	0.1475	0.1701	0.1841	0.1921	0.1952
	nDCG@50	0.0268	0.0283	0.0291	0.0298	0.0310	0.0379
	nDCG@100	0.0324	0.0347	0.0361	0.0379	0.0413	0.0446

5.4 超参数研究(问题 3)

本文提出的 MML 引入了两个超参数, 分别是空间数量 K 和正则项 \mathcal{L}_{pull} 对应的权重 λ_{pull} . 在这里, 我们展示了这两个超参数如何影响性能并阐明了如何设置它们.

首先, 对于超参数 K , 不同数据集的对应的最优空间数量的 K 有所不同. 如图 5 所示, 在 Ciao 数据集中和 Delicious 数据集中, 最佳的 K 是 4; 在 Moivelens 数据集中和 BookX 数据集中, 最佳 K 分别是 2 和 5. 可以看到, 无论 K 取值多少, MML 的性能均优于最佳对比算法. 此外, 我们观察到, 随着 K 数量的增加, 推荐性能在达到峰值后逐渐降低. 除了具有先验知识的数据集外, 通常, 我们可以根据数据集的大小, 将空间数量设置为 3 或 4.

其次, 基于不同数据集的最佳空间 K , 我们显示了在 Ciao, Delicious, Moivelens 和 BookX 四个数据集上, 通过 λ_{pull} 变化带来的推荐性能的变化. 由于正则项 \mathcal{L}_{pull} 可以主动地使空间中的正样本靠近用户, 而权重 λ_{pull} 则是用来控制 \mathcal{L}_{pull} 在模型中的强度. 如果 \mathcal{L}_{pull} 太小, 则会削弱将用户与其喜欢物品之间的交互; 而当 \mathcal{L}_{pull} 的值太大时, 则很容易导致模型过拟合. 图 6 中显示了以 nDCG@10 和 nDCG@20 为指标下的对比结果. 我们发现 Ciao, Delicious, MovieLens 和 BookX 数据集上最佳的 λ_{pull} 值分别约为 0.001、1.0、0.1 和 0.1. 因此, 我们可以得出 MML 模型对 λ_{pull} 值不敏感的结论. 在 $[0, 1]$ 的范围内, 我们可以通过验证集调参获得最佳的 λ_{pull} 参数.

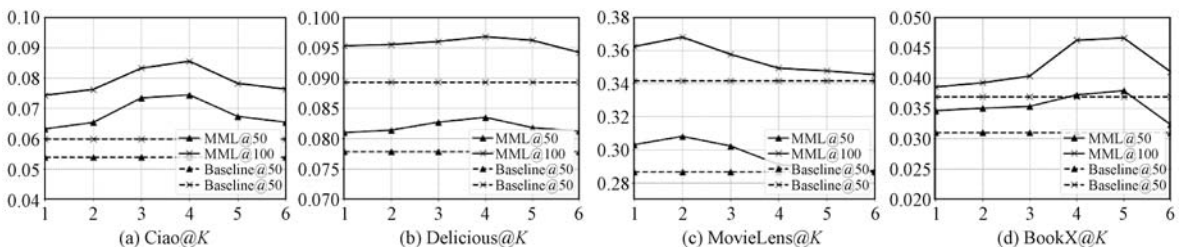


图 5 不同空间数 K 对于模型性能的影响(评价指标为 $nDCG$)

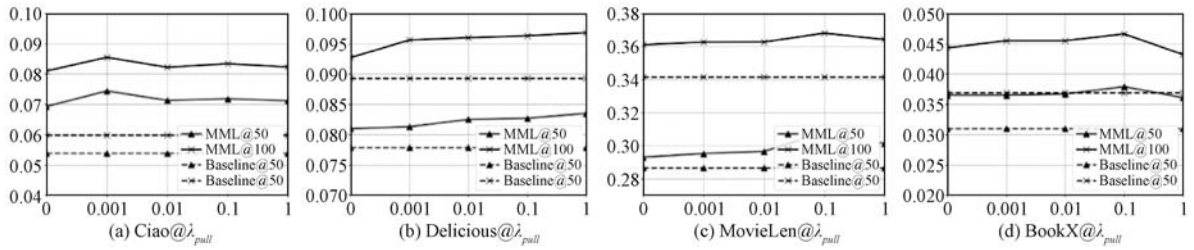


图6 正则化项权重 λ_{pull} 对于模型性能的影响(评价指标为 nDCG)

5.5 多空间可视化(问题4)

为了更加清晰地对比单个空间投影和多空间投影的区别,我们在 Ciao 数据集上可视化了基于 CML 和本文提出的 MML 学习得到的部分嵌入向量,如图 7 所示.其中,CML 和 MML 使用相同的图例来表示不同的物品的所属分类.

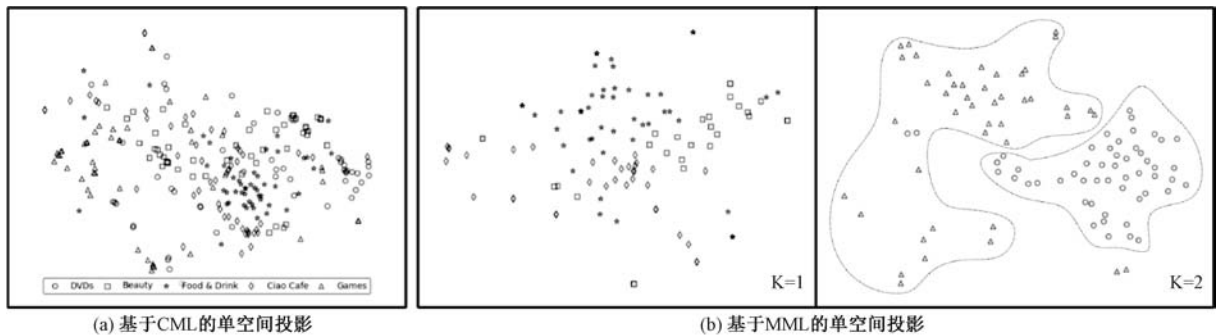


图7 单空间投影和多空间投影对比

尽管用户和物品会出现在多个空间内,但是其在每个空间内展示的侧面是不相同的,具体可以体现为不同的用户-物品、用户-用户和物品-物品关系.为了显示多空间投影框架可以展示更加细粒度的用户-物品建模,我们基于所提出的 MML 推荐框架,利用用户点击物品的类别信息进行一些有趣的推测.以空间数为 2 的划分为例(图 7(b)),我们可以看到,投影到不同空间内的类别是不一样的,而这些不同的类别关系则能够帮助我们对于用户偏好进行更加细致地划分.在 $K=1$ 的空间中,我们可以看到“Beauty”(美妆)、“Ciao Coffee”(咖啡)和“Food & Drink”(美食),而在 $K=2$ 的空间中,占比最高的类别则是“DVDs”和“Games”(游戏).由此推测,位于 $K=1$ 空间中的用户可能是一个偏向于户外出行、通过美食咖啡放松的用户,而 $K=2$ 空间中的用户则是喜欢居家休闲、通过游戏观影放松的用户.通过将用户和物品投影到不同的空间中,MML 可以有效地学习用户和物品的细粒度特征,从而可以充分利用多种用户-物品的空间位置,获得一些

在 CML 中,每一个物品都对应着空间中的一个点.而在 MML 中,每个物品可能对应多个空间中的多个点.从图 7 可以看出,CML 在单个空间中物品类别的分布规律较为散乱,相比之下,MML 的嵌入向量根据物品在不同空间中的对应类别进行聚类,聚集效果优于 CML.

无法直接从原始数据中得到的多侧面建模.

6 总结

在本文中,我们提出了一种从基于多空间的度量学习(MML)框架用于基于隐式反馈的推荐.我们首先使用 NMF 在没有任何辅助信息的情况下,初始化多个度量空间,并获得用户和物品对应的多空间权重.然后,我们设计了针对用户和物品的跨空间相似性度量.此方法可以有效地将用户和物品的多个侧面进行整合,以进行更细粒度的建模.此外,为了保证所提出的多空间框架的有效训练,我们提出了经过校准的优化策略,其中包括经过校准的最大间隔损失函数和经过校准的采样方法.该优化不仅能够将用户和物品的多侧面性质考虑在内,而且能够提升了优化的效率和效果.最后,大量实验表明,提出的 MML 优于最新的推荐算法,最多能将现有的对比算法提高 40%.

参 考 文 献

- [1] Deng A., Zhu Y., Shi B. A collaborative filtering recommendation algorithm based on item rating prediction. *Journal of Software*, 2003, 14(9):1621-1628(in Chinese)
邓爱林, 朱扬勇, 施伯乐. 基于项目评分预测的协同过滤推荐算法. *软件学报*, 2003, 14(9):1621-1628
- [2] Li L., Liu J., Meng X., et al. Recommendation models by exploiting rating matrix and review text. *Chinese Journal of Computers*, 2018, 41(7):1559-1573(in Chinese)
李琳, 刘锦行, 孟祥福等. 融合评分矩阵与评论文本的商品推荐模型. *计算机学报*, 2018, 41(7):1559-1573
- [3] Feng X., Zeng Y.. Joint deep modeling of rating matrix and reviews for recommendation. *Chinese Journal of Computers*, 2020, 43(5):884-900(in Chinese)
冯兴杰, 曾云泽. 基于评分矩阵与评论文本的深度推荐模型. *计算机学报*, 2020, 43(5):884-900
- [4] Hu, Y., Koren, Y., Volinsky, C. Collaborative filtering for implicit feedback datasets//*International Conference on Data Mining*. Pisa, Italy, 2008; 263-272
- [5] Rendle, S., Freudenthaler, C., Gantner, Z., et al. (2012). BPR: Bayesian personalized ranking from implicit feedback. *arXiv preprint arXiv:1205.2618*
- [6] Hsieh, C. K., Yang, L., Cui, Y., et al. Collaborative metric learning//*Proceedings of the 26th International Conference on World Wide Web*. Perth, Australia, 2017; 193-201
- [7] Park, C., Kim, D., Xie, X., et al. Collaborative translational metric learning//*IEEE International Conference on Data Mining*. Singapore, 2018; 367-376
- [8] Tay, Y., Anh Tuan, L., Hui, S. C. Latent relational metric learning via memory-based attention for collaborative ranking//*Proceedings of the World Wide Web Conference*. Lyon, France, 2018; 729-739
- [9] Zhang, S., Yao, L., Tay, Y., et al. Metric factorization: Recommendation beyond matrix factorization. *arXiv preprint arXiv:1802.04606*
- [10] Joachims, T., Granka, L., Pan, B., et al. Accurately interpreting clickthrough data as implicit feedback. *ACM SIGIR Forum*. 2017;4-11
- [11] Pan, R., Zhou, Y., Cao, B., et al. One-class collaborative filtering//*International Conference on Data Mining*. Pisa, Italy, 2008; 502-511
- [12] Paatero, P., Tapper, U. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 1994, 5(2):111-126
- [13] Lee, D. D., Seung, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature*, 1999, 401(6755): 788-791
- [14] Tran, T., Liu, X., Lee, K., et al. Signed distance-based deep memory recommender//*Proceedings of the World Wide Web Conference*. San Francisco, USA, 2019:1841-1852
- [15] Weinberger, K. Q., Saul, L. K.. Distance metric learning for large margin nearest neighbor classification. *Journal of Machine Learning Research*, 2009; 10(2)
- [16] Yang, L., Jin, R. Distance metric learning: A comprehensive survey. *Michigan State University*, 2006, 2(2): 4
- [17] Zadeh, P., Hosseini, R., Sra, S. Geometric mean metric learning//*International Conference on Machine Learning*. New York, USA, 2016; 2464-2471
- [18] Song K, Nie F, Han J, et al. Parameter free large margin nearest neighbor for distance metric learning//*Proceedings of the AAAI Conference on Artificial Intelligence*. San Francisco, USA, 2017; 31(1)
- [19] Wang, Z., Zhang, J., Feng, J, et al. Knowledge graph embedding by translating on hyperplanes//*Proceedings of the AAAI Conference on Artificial Intelligence*. Québec, Canada, 2014; 1112-1119
- [20] Tikhonov, A. N., Arsenin, V. Y. *Solutions of Ill-Posed Problems*. New York, USA, 1977; 1-30
- [21] Chen, H., Yin, H., Wang, W., et al. PME: projected metric embedding on heterogeneous networks for link prediction//*Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. London, UK, 2018; 1177-1186
- [22] Liu, N., Tan, Q., Li, Y., et al. Is a single vector enough? Exploring node polysemy for network embedding//*Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. Anchorage, USA, 2019; 932-940
- [23] Zhang Y, Zhu X, Xu C, Dong S. Hybrid recommendation approach based on deep sentiment analysis of user reviews and multi-view collaborative fusion. *Chinese Journal of Computers*, 2019, 42(6): 1316-1333(in Chinese)
张宜浩, 朱小飞, 徐传运, 董世都. 基于用户评论的深度情感分析和多视图协同融合的混合推荐方法. *计算机学报*, 2019, 42(6): 1316-1333
- [24] Zhang L, Zhao Y, Zhu Z. Advances in semantically shared subspace learning for cross-media data. *Chinese Journal of Computers*, 2017, 40(6): 1394-1421(in Chinese)
张磊, 赵耀, 朱振峰. 跨媒体语义共享子空间学习研究进展. *计算机学报*, 2017, 40(6): 1394-1421
- [25] Tan Yanchao, Carl Yang, Wei Xiangyu, Ma Yun, Zheng Xiaolin. Multi-facet recommender networks with spherical optimization. *arXiv preprint arXiv:2103.14866*
- [26] Shrivastava, A., Gupta, A., Girshick, R. Training region-based object detectors with online hard example mining//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA, 2016; 761-769
- [27] Wang, X., Han, X., Huang, W., et al. Multi-similarity loss with general pair weighting for deep metric learning//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Miami, USA, 2009; 5022-5030
- [28] Tran, V. A., Hennequin, R., Royo-Letelier, J., et al. Im-

- proving collaborative metric learning with efficient negative sampling//Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. Paris, France, 2019; 1201-1204
- [29] Yin, H., Chen, H., Sun, X., et al. SPTF: a scalable probabilistic tensor factorization model for semantic-aware behavior prediction//IEEE International Conference on Data Mining. New Orleans, USA, 2017; 585-594
- [30] Li, M., Zhang, S., Zhu, F., et al. Symmetric metric learning with adaptive margin for recommendation//Proceedings of the AAAI Conference on Artificial Intelligence. New York, USA, 2020; 4634-4641
- [31] Wang, C., Zhang, X., Lan, X. How to train triplet networks with 100k identities? //Proceedings of the IEEE International Conference on Computer Vision Workshops. Venice, Italy, 2017; 1907-1915
- [32] Wu, C. Y., Manmatha, R., Smola, et al. Sampling matters in deep embedding learning//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy, 2017; 2840-2848
- [33] Huang, E. H., Socher, R., Manning, C. D., et al. Improving word representations via global context and multiple word prototypes//Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics. Jeju Island, Korea, 2012; 873-882
- [34] Wang, Y. X., Zhang, Y. J. Nonnegative matrix factorization: A comprehensive review. IEEE Transactions on Knowledge and Data Engineering, 2012, 25(6): 1336-1353
- [35] Xu, W., Liu, X., Gong, Y. Document clustering based on non-negative matrix factorization//Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval. Toronto, Canada, 2003; 267-273
- [36] Barkan, O., Koenigstein, N. Item2vec: neural item embedding for collaborative filtering//IEEE 26th International Workshop on Machine Learning for Signal Processing. Salerno, Italy, 2016; 1-6
- [37] Caselles-Dupr e, H., Lesaint, F., Royo-Letelier, J. Word2vec applied to recommendation: Hyperparameters matter//Proceedings of the 12th ACM Conference on Recommender Systems. New York; USA, 2018; 352-356
- [38] Steck, H. Item popularity and recommendation accuracy//Proceedings of the Fifth ACM Conference on Recommender systems. Chicago, USA, 2011; 125-132
- [39] Krichene, W., Rendle, S. On Sampled Metrics for Item Recommendation//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Virtual Event, USA, 2020; 1748-1757
- [40] He, X., Liao, L., Zhang, H., et al. Neural collaborative filtering//Proceedings of the 26th International Conference on World Wide Web. Perth, Australia, 2017; 173-182



TAN Yan-Chao, Ph. D. Her research interests include recommender system, machine learning.

ZHENG Xiao-Lin, Ph. D, professor. His research interests include smart business, financial technology, big data analysis.

WEI Xiang-Yu, master. His research interests include recommender system, machine learning.

Carl Yang, Ph. D, assistant professor. His research interests include graph data mining.

Background

Implicit feedback is widely explored by modern recommender systems. Since the feedback is often sparse and imbalanced, it poses great challenges to the learning of complex interactions among users and items. Recently, many algorithms based on metric learning have been proposed to explore the implicit feedback for recommendation. They capture complex user-user and item-item relationships, which represent the finer-grained preference information beyond the user-item interactions. Despite their improved performance, by using a single space to project all users and items, they ignore the possible multiple preferences of users and multiple properties of items, which is of great importance in recommender systems to fully model user portrait.

In this paper, we study the challenge of recommendation with implicit feedback and spot the inefficacy of existing methods in modeling and optimizing multi-facet user preferences and item aspects. Inspired by metric learning methods, we extend them to the multi-space recommendation scenario by projecting users and items into multiple spaces, and proposed a novel framework of Multi-space Metric Learning (MML) for the recommendation. We first leverage non-negative matrix factorization (NMF) to obtain the initial multiple spaces without any auxiliary information. Based on multi-space projection, we design a cross-space similarity measurement to project users and items with finer granularity. Moreover, a calibrated optimization strategy is designed to guar-

antee efficient and effective model training. Specifically, it ensures that the loss function and sampling process are calibrated towards the learning of the multi-space user and item projections. Finally, an iterative loop mechanism is designed to fully integrate the learned relationships for further mining the multi-facet of user and item. The experimental results on four real-world recommendation datasets validate our proposed MML

framework and outperform the-state-of-art recommendation algorithms with up to 40% relative improvements.

This work described in this study is supported in part by the National Key R&D Program “Big Data Credit Investigation and Intelligent Evaluation Technology” No. 2018YFB1403001 and National Key R&D Program of China (No. 2019YFB1404901).