

DOI:10.13364/j.issn.1672-6510.20220161

一种基于用户信息映射的跨领域推荐算法

苑树强, 史艳翠

(天津科技大学人工智能学院, 天津 300457)

摘要: 冷启动问题一直是推荐系统中的一大难点,而跨领域推荐可以通过迁移其他领域的信息缓解这个问题.本文提出一种基于用户信息映射的跨领域推荐算法,它可以在具有少量标记数据的情况下有效地学习跨域关系,并利用标签提高用户和物品编码的准确性.首先学习用户和物品的潜在向量,并改进降维方法将其处理成低维稠密向量;然后生成用户种群,利用偏好不同的用户种群训练映射函数,通过映射函数映射用户信息提高推荐效果.通过实验探究各个参数对实验结果的影响,验证了本文模型对重叠用户数量的依赖更低,相较于基准实验,有更好的推荐准确性,同时表明本文方法有更好的推荐效果.

关键词: 冷启动; 跨领域推荐; 推荐算法

中图分类号: TP391

文献标志码: A

文章编号: 1672-6510(2023)03-0068-07

A Cross-Domain Recommendation Algorithm Based on User Information Mapping

YUAN Shuqiang, SHI Yancui

(College of Artificial Intelligence, Tianjin University of Science & Technology, Tianjin 300457, China)

Abstract: The cold start problem has always been a difficulty in recommendation systems, and cross-domain recommendation can alleviate this problem by migrating information from other domains. In this article, we propose a cross-domain recommendation model based on user information mapping, which can effectively learn cross-domain relationships with a small amount of labeled data, and improve the accuracy of user and item coding by using tags. In this model, firstly, the potential vectors of users and objects are learned and processed into low-dimensional dense vectors by improving the dimensionality reduction method. Then, the user population is generated and the mapping function is trained by the user population with different preferences. User information is mapped by mapping function to improve recommendation effect. Through experiments to explore the impact of each parameter on the experimental results, it is verified that the proposed model in this article has a lower dependence on the number of overlapping users, and has better recommendation accuracy compared with the benchmark experiment. At the same time, it shows that the method in this article has better recommendation effect.

Key words: cold start; cross-domain recommendation; recommendation algorithm

随着网络技术的不断发展,人们的生活越来越丰富,越来越多的用户展现出丰富的兴趣,并在各个领域都有所涉猎.研究^[1]表明不同平台的用户偏好和项目特征存在相似性和相关性,因此推荐系统开始利用用户在多个领域的交互信息提高推荐质量,即跨领域推荐^[2].跨领域推荐是通过将一个领域的知识转移到

另一个领域而缓解数据的稀疏性问题.尤其是当用户进入一个新的领域时,也就是冷启动问题,可以利用用户以前已有交互领域中的信息为其推荐.这两个问题也是推荐系统中长期存在的两个问题.

在跨领域推荐中,最关键的问题是如何建立两个领域之间的桥梁^[3].Li等^[4]提出一种密码本迁移学习

收稿日期: 2022-06-24; 修回日期: 2022-09-09

基金项目: 天津市教委理工类基本科研业务费项目(2018KJ105); 天津市企业科技特派员项目(20YDTPJC00560)

作者简介: 苑树强(1990—),男,山东德州人,硕士研究生;通信作者: 史艳翠,副教授, syc@tust.edu.cn

算法,认为跨领域推荐不一定需要迁移具体的数据,而是将用户的评分模式从密集的源域迁移至稀疏的目标域,以此丰富目标域的评分矩阵.基于共享知识模型的跨领域推荐算法 SKP^[5],将用户属性信息加入推荐过程以解决正迁移不足的问题,引入其他可用的辅助信息增加可用数据量,缓解数据稀疏带来的影响.以上方法属于传统的跨领域推荐,没有对每个域自身的特点进行域适应,也没有进行特征映射.这可能会忽视不同域之间不同特征的影响,甚至会引入负迁移.

在不同的任务中,领域间的联系有所不同,因此对于不同的情况,例如用户、物品、文本属性、评分、标签、类别等都可能作为跨领域推荐的共享信息.Zhang 等^[6]通过捕获不同标记之间的语义关系,识别域间相似的用户和物品,通过用户和物品的域内和域间相关性约束矩阵分解;但域间用户和物品相似度的准确性对推荐结果的影响很大.Guo 等^[7]在源域和目标域各自利用用户配置文件生成用户标签矩阵,并利用矩阵分解找到与源域标签相似的目标域标签,将源域的用户配置文件转换为目标域的用户配置文件,使跨域推荐转为单域推荐;但稀疏的用户-标签矩阵导致标签相似性计算困难.

对于跨领域推荐,可迁移特征的质量也是影响推荐效果的重要因素.近年来,随着深度学习的发展,研究人员开始利用深度学习的方法改善跨域推荐的效果^[8].Man 等^[9]提出一种利用多层感知器捕获跨域的非线性映射函数的框架,为学习各域实体的领域特征提供了很高的灵活性.Zhu 等^[10]认为不同用户的偏好不同,所以其映射函数也应该不同.通过用户特征嵌入反馈的元网络生成个性化的映射函数,以实现每个用户的个性化偏好传递.上述算法利用深度学习的方法映射用户特征,但训练依靠域间的重叠用户,而现实情况下域间重叠用户较少.Kang 等^[11]提出使用欧氏距离反映用户对物品的偏好,并使用半监督的方法训练映射函数,但度量空间中用户-物品偏好表示的准确性明显影响推荐结果的准确性.

大部分情况下,物品、标签等信息很难出现重叠;但随着网络的发展,信息越来越丰富,用户可能会同时多个领域存在交互信息.所以,针对源域和目标域存在重叠用户的情况,本文提出一种基于用户信息映射的跨域推荐算法.虽然不同领域中标签没有重叠,但是不同领域中可能会存在含义相似的标签,即标签对用户和物品嵌入模型的影响应该是跨领

域的.因此,首先通过用户和标签的交互信息得到用户和物品特征,并使用降维方法处理成低维特征,再将用户的源域特征映射到目标域.此外,用户映射之后应保持自己的偏好类别,所以通过聚类方法生成用户种群,并利用不同种群的偏好信息和重叠用户训练映射函数.综上所述,本文通过映射用户在源域中的信息,为用户推荐目标域的物品,本质上属于基于内容的推荐系统.

本文的主要工作如下:(1)提出一种新的用户、物品信息表示方式,利用直接标签和间接标签表示用户和物品,更好地体现用户偏好和物品特性;(2)改进降维方式,利用用户物品的交互信息降维,并考虑标签信息特征的重要性;(3)对于映射函数优化时标记样本过少的问题,利用种群信息标记非重叠用户,并优化映射函数.

1 算法描述

1.1 问题定义

跨领域推荐中存在一种情况:在一个系统中,存在一个信息丰富的领域 A 和一个信息较稀疏的领域 B ,当 A 中的用户进入 B 领域时,由于该用户在 B 领域中没有信息,因此会遇到冷启动问题,例如电商的图书领域和 CD 光盘领域,此时需要迁移 A 中的信息为用户改善推荐效果.在本文中,将信息丰富的领域称为源域 S ,信息稀疏的领域称为目标域 T .源域和目标域的用户集合为 U^S, U^T ,物品集合为 V^S, V^T ,标签集合为 B^S, B^T ,评分集合为 R^S, R^T . m 为用户数量, n 为物品数量.图书领域和光盘领域在现实情况下一般不会存在重叠的物品和标签,但可能会存在一些重叠用户,使用 $U^O = U^S \cap U^T$ 表示重叠用户集合.

根据以上描述,给出问题的定义:假设存在两个领域,源域 S 和目标域 T . $U^S \cap U^T \neq \emptyset$, $V^S \cap V^T = \emptyset$, $B^S \cap B^T = \emptyset$,目标是为源域中的非重叠用户推荐目标域物品.

1.2 用户和物品的向量化表示

要将源域的用户特征映射到目标域,首先要得到用户的向量化表示.大部分方法(如矩阵分解^[12])的用户向量都依靠用户-物品的交互矩阵得到.但在目标域中用户-物品的交互矩阵一般较为稀疏,影响生成用户向量的准确性.并且,根据上述方法生成的用户向量得到的用户种群难以确认种群中用户偏好相似性,缺乏可解释性.标签作为用户自己理解和添加

的信息,一定程度上反映了用户的兴趣,可以更好地反映用户的偏好信息. 因此,本文主要依据用户和物品交互过的标签信息进行推荐.

由于用户和物品的向量化表示形式相同,下面用源域用户集合 U^S 为例,介绍源域用户向量生成的过程. 对于 U^S 来说,每个用户的唯一标识是离散的,互相没有关系,可以使用 one-hot 编码表示^[13],向量维度为 U^S 中用户的数量. 对于标签而言,用户通常会交互多个标签,将其表示成 multi-hot 的形式. 用户在打标签时是对某一物品标注,不仅是用户自己标注的标签可以作为标签信息,用户标注过的物品的其他标签也可以作为用户的信息表示.

如图 1 所示,对于用户 1 而言,除了标签 1 外,标签 2、3 也可以作为标签信息. 将标签 1 称为用户 1 的直接标签,标签 2、3 称为用户 1 的间接标签.

同理,对于物品 1 而言,标签 1、2、3 为直接标签,标签 4 为间接标签. 用户和物品的标签信息由直接标签和间接标签组成.

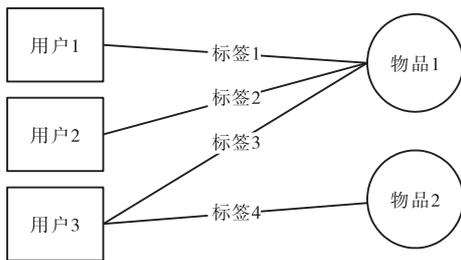


图 1 标签关系表示

Fig. 1 Representation of label relation

以用户为例,标签信息表示为 b ,包括直接标签和间接标签两部分. 直接标签表示成 multi-hot 向量形式,用符号 \bar{b} 表示,间接标签为交互过的物品的直接标签的向量集合,使用符号 b' 表示. $b' = \bar{b}_1 + \dots + \bar{b}_z$, \bar{b}_z 表示该用户交互的第 z 个物品直接标签表示.

$$b = \bar{b} + \alpha b' \quad (1)$$

其中 α 为超参数,表示间接标签在用户表示中的影响力. 以用户 1 为例, $\bar{b} = [1 \ 0 \ 0 \ 0]$, $b' = [0 \ 1 \ 1 \ 0]$, 最终的标签信息为 $b = [1 \ \alpha \ \alpha \ 0]$. 用户信息的整体表示如图 2 所示.



图 2 用户信息的整体表示

Fig. 2 Overall representation of user information

同理,物品向量也由 one-hot 编码和标签信息表

示,标签包括物品本身的标签以及与该标签交互的用户的其他标签.

本文源域和目标域中用户、物品表示向量的集合分别使用 u^S 、 v^S 、 u^T 、 v^T 表示.

1.3 特征向量降维

在现实情况下,如在亚马逊数据集中,用户数量和物品数量都非常庞大,所以 1.2 节的用户向量和物品向量都是一个非常稀疏的高维向量,难以直接使用. 因此,本文提出一种降维方式,将高维向量降维. 采用深度学习嵌入方法,借鉴双塔模型的思想^[14],通过深度学习神经网络(deep neural networks, DNN)处理稀疏向量. 但如果直接将该向量输入 DNN 中,会导致网络参数过多. 为了解决这个问题,将输入向量分成两个部分,如图 3 所示. 下面以源域中的用户 i 和物品 j 为例,将上一节得到的用户向量或物品向量 u_i^S 和 v_j^S 作为输入 ($u_i^S \in U^S, v_j^S \in V^S$),并将 one-hot 编码部分和标签编码部分分别处理输入隐藏层中:

$$\begin{cases} l_{u_i^S} = [u_i^S[\text{start} : m] \cdot W_u, u_i^S[m+1 : \text{end}] \cdot W_1] \\ l_{v_j^S} = [v_j^S[\text{start} : n] \cdot W_v, v_j^S[n+1 : \text{end}] \cdot W_1] \end{cases} \quad (2)$$

经过处理得到的 $l_{u_i^S}$ 和 $l_{v_j^S}$ 作为图 3 中 l_u 和 l_v 层的输入,其中, $W_u \in R^{m \times p}$, $W_v \in R^{n \times p}$,用于处理 one-hot 编码部分, $W_1 \in R^{e \times q}$, e 表示标签数量,用于处理标签信息部分. p 和 q 的比值代表了两部分的重要性比较,比如直观上标签信息部分的信息量更大,因此 q 的值可能更大一些. 隐藏层与双塔模型类似,使用 tanh 作为输出层和隐藏层的激活函数. 最终输出层的两个输出 l_{out} 分别作为用户向量 u_i^S 和物品向量 v_j^S 降维后的向量,用 yu_i^S 和 yv_j^S 表示.

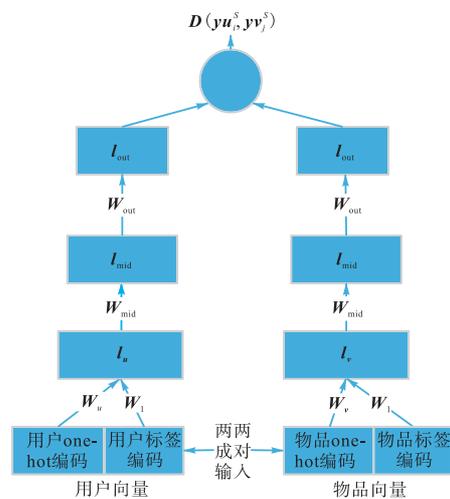


图 3 深度学习降维方法

Fig. 3 Deep learning dimension reduction method

对于降维后的向量 $\mathbf{y}u_i^S$ 和 $\mathbf{y}v_j^S$, 通过计算它们之间的欧氏距离, 表示用户和物品的交互信息, 使用户与交互过的物品的距离更近. 损失函数为

$$\min \sum_{i=1}^m \sum_{j=1}^n \|(\mathbf{y}u_i^S - \mathbf{y}v_j^S) \circ \mathbf{W}_1\|^2 \quad (3)$$

式中: \mathbf{W}_1 为二元指示矩阵, 表示用户物品是否有交互, \circ 表示公式中只考虑交互过的用户和物品的信息, 因为未交互过的用户和物品距离不应该更近. 将源域和目标域中用户、物品表示向量集合 $\mathbf{u}^S, \mathbf{v}^S, \mathbf{u}^T, \mathbf{v}^T$ 分别降维得到对应的降维向量集合 $\mathbf{y}u^S, \mathbf{y}v^S, \mathbf{y}u^T, \mathbf{y}v^T$.

1.4 种群映射函数

1.3 节中得到了用户和物品的低维向量表示, 将源域和目标域的用户 $\mathbf{y}u^S$ 通过 k 均值聚类 (k -means) 算法分别生成 k 个用户种群, 其中用户之间的距离使用欧氏距离表示.

本模型生成映射函数将源域的用户向量映射到目标域中. 该函数反映了源域和目标域之间的关系. 将重叠用户在源域的向量映射到目标域中, 并使它和目标域中交互的物品距离更近. 通过重叠用户在目标域中的交互信息优化映射函数

$$L_r = \min \sum_{\substack{i \in U^a \\ j \in V^T}} \|f(\mathbf{y}u_i^S, \theta) - \mathbf{y}v_j^T\|^2 \quad (4)$$

式中: \mathbf{W}_2 为二元指示矩阵, 表示重叠用户在目标域的交互信息; $f(\theta)$ 表示映射函数, θ 为全局参数.

由于现实情况中一般只存在少量的重叠用户, 利用重叠用户作为标记数据的有监督的训练方法难以很好地训练模型, 所以本文利用用户种群关系标记非重叠用户. 如图 4 所示, 对于非重叠用户来说, 映射后的位置应该靠近同类的重叠用户.

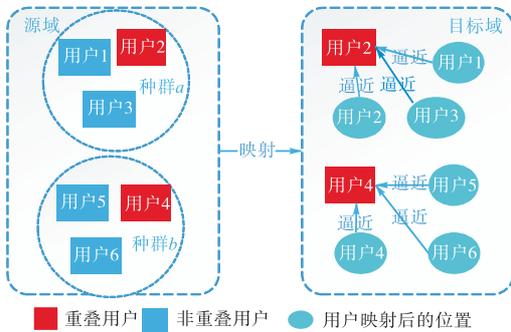


图 4 非重叠用户映射

Fig. 4 Non-overlapping user mappings

因此, 重叠用户映射后在目标域的距离逼近原本的位置, 而非重叠用户则逼近同类重叠用户. 基于以

上思想, 给出用户关系的损失函数为

$$L_u = \min \left(\sum_{i \in U^a} \|f(\mathbf{y}u_i^S, \theta) - \mathbf{y}u_i^T\|^2 + \lambda \sum_{a \in k} d_a \right) \quad (5)$$

其中: λ 代表训练时种群信息的权重, d_a 代表种群 a 内非重叠用户映射后与该类重叠用户的距离, k 代表用户种群个数, d_a 表示为

$$d_a = \sum_{i \in U^a} \sum_{j \in U^a} \|f(\mathbf{y}u_j^S, \theta) - \mathbf{y}u_i^T\|^2 \quad (6)$$

用户映射函数的完全损失函数可以表示为

$$L = L_r + L_u \quad (7)$$

为了获得映射函数, 本文采用随机梯度下降法训练参数 θ . 通过训练集中用户向量反向传播训练映射函数参数, 直到模型收敛.

1.5 跨领域推荐

当给一个源域用户推荐目标域物品时, 这个用户在目标域没有任何交互信息, 相当于冷启动用户. 推荐与映射用户最接近的 N 个目标域物品, 如用户 i 和物品 j 的距离 D 使用欧氏距离表示为

$$D = \|f(\mathbf{y}u_i^S, \theta) - \mathbf{y}v_j^T\|^2 \quad (8)$$

通过计算冷启动用户 i 和目标域物品的距离, 将物品按欧氏距离由低到高排序, 得出目标用户 i 的推荐列表, 记为 $R_i = \{v_1, v_2, \dots, v_n\}$, 最终选取前 N 个物品为目标用户 i 进行推荐.

1.6 总体程序

输入: 源域和目标域的用户物品向量集合 $\mathbf{u}^S, \mathbf{v}^S, \mathbf{u}^T, \mathbf{v}^T$.

输出: 推荐物品序列.

(1) 特征降维: 利用用户物品的交互信息训练深度学习模型, 将 $\mathbf{u}^S, \mathbf{v}^S$ 和 $\mathbf{u}^T, \mathbf{v}^T$ 中元素两两成对输入模型降维, 得到降维后的向量表示 $\mathbf{y}u^S, \mathbf{y}v^S, \mathbf{y}u^T, \mathbf{y}v^T$.

(2) 源域用户聚类: 将集合 $\mathbf{y}u^S$ 通过 k -means 聚类算法生成 k 个用户种群.

(3) 种群个性化映射: 利用源域用户 $\mathbf{y}u^S$ 中种群信息和重叠用户交互信息优化映射函数.

(4) 生成推荐列表: 将源域用户 $\mathbf{y}u^S$ 映射到目标域, 计算用户和目标域物品的欧氏距离, 生成推荐列表.

2 实验

实验环境: 16 GB 内存, Windows 10 操作系统,

RTX 2060 显卡, Python 3.6, 机器学习框架 TensorFlow 2.5.0.

2.1 数据集及其处理

实验部分使用的数据集是亚马逊数据集, 该数据集包含了 42 个不同的物品域, 选取两个类别: 图书和光盘. 其中图书域用户数量为 603 668, 光盘域用户数量为 75 258; 重叠用户数量为 16 738.

将图书域视为源域, 光盘域视为目标域. 数据集内的 reviewerID 字段作为用户 ID; asin 字段作为物品 ID; summary 字段作为标签. 对原始数据集中的数据进行清洗, 删除交互少、孤立的数据. 对于非重叠用户, 在源域数据中去除了交互物品少于 150 的用户、交互用户少于 150 的物品和出现次数少于 30 的标签. 在目标域中去除了交互数据少于 30 的用户和物品以及出现次数少于 4 的标签. 对于重叠用户而言, 去除了交互数据少于 10 的用户. 数据集统计信息见表 1.

表 1 数据集统计信息
Tab. 1 Data set statistics

数据集	用户数量	项目数量	交互信息	稀疏度/%	重叠用户数量
图书	5384	6699	53 055	99.86	1367
光盘	5291	6595	24 715	99.93	

2.2 实验设置

对于目标域, 选取 30% 的重叠用户, 删除其所有的交互信息, 将其作为冷启动用户(测试用户)计算推荐准确性. 在实验中, 为了解决对所有物品排名的耗时问题, 在用户没有交互过的物品(负样本)中随机抽取 999 个物品, 并评估交互过的物品(正样本)的排名是否高于未交互物品(负样本). 为了探究重叠用户数量对实验结果的影响, 在模型训练时, 对训练集中的重叠用户进行限制, 分别选取重叠用户的 10%、50% 和 100% 对模型进行训练. 为了消除实验结果的偶然性, 每组实验分别进行 10 次并取平均值作为最终结果.

2.3 评价指标

使用精确率(precision)和平均精度(mean average precision, mAP)衡量和评价推荐结果. 精确率是广泛用于推荐系统中的度量标准, 精确率针对预测结果而言, 表示预测为正的样本中有多少是真正的正样本. mAP 是一种位置敏感的评价指标, 反映每个用户的平均精确率的平均值; 推荐列表中相关物品越靠前, mAP 就越高.

$$P = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |R(u)|} \tag{9}$$

$$P_{av,j} = \frac{\sum_{j \in |R(u) \cap T(u)|} N_1}{N_2} \tag{10}$$

$$P_{avm} = \frac{1}{|U^o|} \sum_{i \in U^o} P_{av,i} \tag{11}$$

其中: P 表示精确率; $R(u)$ 表示推荐列表; $T(u)$ 表示用户实际交互的物品; $P_{av,i}$ 表示单个用户平均精确率, 针对的是一个用户的排序结果; N_1 和 N_2 分别表示 j 在命中列表和推荐列表中的排名; P_{avm} 表示平均精度.

2.4 实验过程与结果分析

为了研究本算法的推荐效果, 对模型中的几个参数分别单独测试它们对推荐效果的影响, 包括聚类 k 值、降维的维度、间接标签的权重 α 、 q 和 p 的比值、种群信息权重 λ , 并与几组跨领域推荐的模型对比, 分析模型的性能.

2.4.1 聚类 k 值

聚类 k 值对推荐结果的影响如图 5 所示, 即用户聚类数目 k 值对冷启动用户 top 10 推荐效果的影响. 聚类 k 值表示源域用户和目标域用户的种群个数, 代表着用户偏好的分类, 聚类效果也影响着用户偏好映射的准确性. 推荐结果在 k 值大于 8 时持续下降, 在 k 值为 6 时取得最高值. 分析原因可能是当 k 值过大或过小时, 源域和目标域的分类难以做到很好的对应, 导致映射效果降低, 因此在后续实验中将 k 值设为 6.

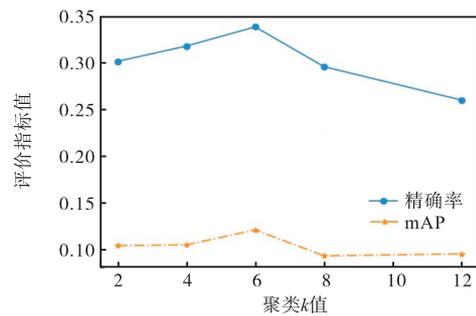


图 5 聚类 k 值对推荐结果的影响
Fig. 5 Effect of clustering k value on the recommendation result

2.4.2 降维维度

在实验过程中, 选择合适的降维维度非常重要, 降维维度对推荐结果的影响如图 6 所示.

为了研究降维维度对实验结果的影响, 将降维维度逐渐增大, 分别考虑降维到 2、4、6、8 时精确率和

mAP 的取值(图 6). 随着维度的提高,推荐结果准确性下降. 在降维维度为 2 时,模型的性能表现最好.

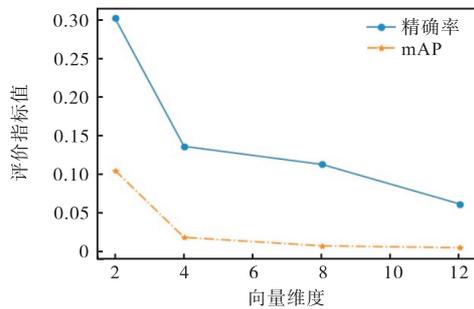


图 6 降维维度对推荐结果的影响

Fig. 6 Effect of dimension reduction length on recommendation results

2.4.3 间接标签权重

间接标签权重 α 代表着对间接标签信息在用户向量、物品向量中比重的大小,间接标签权重对推荐结果的影响如图 7 所示.

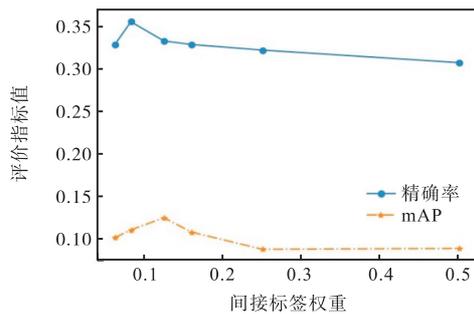


图 7 间接标签权重对推荐结果的影响

Fig. 7 Effect of indirect label weight on recommendation results

由图 7 可以看出,随着间接标签权重的提高,推荐精确率和 mAP 逐渐升高,达到最高点后开始下降. 这主要是因为只考虑直接标签难以反映用户的全部偏好,适量加入间接标签可以提高用户偏好的描述准确性,而间接标签为其他用户打的标签,有可能带有与本用户不同的偏好信息,过高的权重可能会放大间接标签中的无关因素.

2.4.4 降维中 q 和 p 的比值

q 和 p 的比值反映降维过程中标签信息权重. q 和 p 的比值对推荐结果的影响如图 8 所示.

由图 8 可以看出,标签信息权重为 4.5 时效果最好,主要是因为标签信息代表用户的偏好和物品的类别,适当提高标签信息的权重有利于得到更好的推荐结果,而过高的标签信息权重可能会忽略用户项目的标识信息.

2.4.5 种群信息权重

种群信息权重 λ 代表在训练映射函数时,非重叠用户信息的影响力大小. 种群信息权重对推荐结果的影响如图 9 所示.

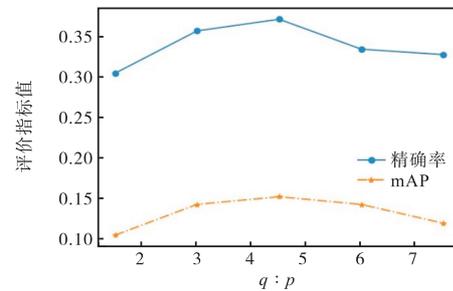


图 8 q 和 p 的比值对推荐结果的影响

Fig. 8 Effect of the ratio of q and p on the recommendation results

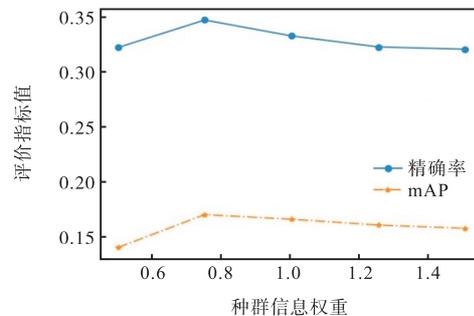


图 9 种群信息权重对推荐结果的影响

Fig. 9 Effect of population information weight on recommendation results

由于重叠用户数量较少,适当提高非重叠用户信息的权重,可以改善推荐效果. 由图 9 可知,在 λ 为 0.75 时效果最好;但非重叠用户在目标域不存在用户交互信息,所以过高的权重可能会影响映射的准确性.

2.4.6 消融实验

为研究间接标签和种群信息的有效性,在选取重叠用户 100% 时,分别在缺失间接标签和种群信息的条件下进行消融实验,结果见表 2.

表 2 消融实验结果

Tab. 2 Ablation experimental results

实验类型	P@10	M@10	P@20	M@20
缺失间接标签	0.236 5	0.145 6	0.263 1	0.194 8
缺失种群信息	0.244 2	0.102 2	0.241 9	0.160 3
完整模型结果	0.279 4	0.165 5	0.379 0	0.201 1

P@10、P@20 分别代表精确率在 top 10 和 top 20 上的表现, M@10、M@20 分别代表 mAP 在 top 10 和 top 20 上的表现. 由表 2 可以看出,加入间接标签信

息明显提高了模型的推荐精确率,因为用户与物品的直接交互较少,直接标签不能很好地代表用户和物品的信息.而加入种群信息标记非重叠用户,使映射函数的训练不仅仅依靠重叠用户,提高了映射的准确性,进而得到更好的推荐效果.

2.4.7 基线设置

为了评估算法性能,选取以下几组不同的方法和本文提出的模型对比,以验证本文模型的有效性.

EMCDR^[9]: 一个用于处理跨域推荐的嵌入和映射框架,通过潜在空间中的映射函数将源域和目标域中的用户和物品投影到两个不同的潜在空间中.

SSCDR^[11]: 一种基于半监督方法的冷启动用户CDR 框架,利用非重叠用户的数据学习映射函数.

PTUPCDR^[10]: 一个 CDR 用户偏好个性化传输框架,学习一个由用户特征嵌入组成的元网络生成个性化的桥接函数,以实现用户偏好的个性化传递.

表 3 列出了各模型在 top 10 和 top 20 上的详细数据. EMCDR 使用重叠用户训练映射函数,受重叠用户数量影响较大;本文模型利用种群信息解决非重叠用户在目标域中信息较少的问题,对重叠用户数据的依赖性降低. 对于同样利用非重叠用户数据和利用用户交互信息训练模型的 SSCDR 和 PTUPCDR, 本文模型也表现出更好的推荐结果.

实验表明,本文模型可以更好地应对真实情况下重叠用户过少的问题,可以有效地为具有跨域信息的冷启动用户推荐物品.

表 3 不同模型的评价指标对比

Tab. 3 Comparison of evaluation indexes of different models

模型	10% 重叠用户				50% 重叠用户				100% 重叠用户			
	P@10	M@10	P@20	M@20	P@10	M@10	P@20	M@20	P@10	M@10	P@20	M@20
EMCDR	0.217 8	0.086 2	0.318 1	0.092 1	0.235 1	0.096	0.339 0	0.105 9	0.241 3	0.101 5	0.344 5	0.105 2
SSCDR	0.236 2	0.096 8	0.334 1	0.102 7	0.253 6	0.106 4	0.355 0	0.113	0.256 2	0.111 5	0.355 9	0.114 3
PTUPCDR	0.248 2	0.081 2	0.236 3	0.117 5	0.247 5	0.080 6	0.244 2	0.126 2	0.247 2	0.081 3	0.247 6	0.096 8
本文模型	0.258 2	0.369 2	0.332 2	0.169 2	0.292 5	0.141 8	0.375 1	0.185 2	0.279 4	0.165 5	0.379	0.201 1

3 结 语

本文通过将源域中的用户偏好映射到目标域,提高用户在目标域中的推荐效果,适用于为其他领域存在交互信息的冷启动用户进行推荐,如向电商平台中已购买过某领域(如图书)物品的用户推荐目标域(如衣服)的物品.利用标签信息获得更准确的用户偏好,经过深度学习降维处理后将其映射到目标域,并通过种群信息利用非重叠用户训练映射函数,提高映射准确性.通过大量的实验,探究各个参数对实验结果的影响,验证了本文模型对重叠用户数量的依赖更低,相较于基准实验,有更好的推荐准确性.

参考文献:

[1] 刘真,田靖玉,苑宝鑫,等. 基于知识聚合和迁移的跨领域推荐算法[J]. 电子学报, 2020, 48(10) : 1928-1932.

[2] 陈雷慧,匡俊,陈辉,等. 跨领域推荐技术综述[J]. 华东师范大学(自然科学版), 2017(5) : 101-116.

[3] 陶鸿,吴国栋,孙成,等. 跨领域推荐研究进展[J]. 长春师范大学学报, 2019, 38(12) : 44-54.

[4] LI B, QIANG Y, XUE X. Can movies and books collaborate? Cross-domain collaborative filtering for spar-

city reduction[C]//Proceedings of the 21st International Joint Conference on Artificial Intelligence. San Francisco : Morgan Kaufmann Publishers Inc. , 2009 : 2052-2057.

[5] 李林峰,刘真,魏港明,等. 基于共享知识模型的跨领域推荐算法[J]. 电子学报, 2018, 16(8) : 1947-1953.

[6] ZHANG Q, HAO P, LU J, et al. Cross-domain recommendation with semantic correlation in tagging systems [C]//IEEE. 2019 International Joint Conference on Neural Networks (IJCNN). New York : IEEE, 2019 : 8852049.

[7] GUO Y, CHEN X. A framework for cross-domain recommendation in folksonomies[J]. Journal of automation and control engineering, 2013, 1(4) : 326-331.

[8] 任豪,刘柏嵩,孙金杨. 面向知识迁移的跨领域推荐算法研究进展[J]. 计算机科学与探索, 2020, 14(11) : 1813-1827.

[9] MAN T, SHEN H, JIN X, et al. Cross-domain recommendation: an embedding and mapping approach[C]// Proceedings of the 26th International Joint Conference on Artificial Intelligence. Menlo Park : AAAI Press, 2017 : 2464-2470.

[10] ZHU Y, TANG Z, LIU Y, et al. Personalized transfer of user preferences for cross-domain recommendation (下转第 80 页)

- [10] ZHAO Y, QI H S, CHENG D Z. Input-state incidence matrix of Boolean control networks and its applications[J]. *Systems & control letters*, 2010, 59(12) : 767–774.
- [11] FORNASINI E, VALCHER M E. Observability, reconstructibility and state observers of Boolean control networks[J]. *IEEE Transactions on automatic control*, 2012, 58(6) : 1390–1401.
- [12] LASCHOV D, MARGALIOT M. A maximum principle for single-input Boolean control networks[J]. *IEEE Transactions on automatic control*, 2011, 56(4) : 913–917.
- [13] ZHANG K Z, ZHANG L J. Observability of Boolean control networks: a unified approach based on finite automata[J]. *IEEE Transactions on automatic control*, 2016, 61(9) : 2733–2738.
- [14] ZHANG K Z. Synthesis for observability of logical control networks[J]. *Automatica*, 2022, 144: 110481.
- [15] CHENG D Z, QI H S, ZHAO Y. An introduction to semi-tensor product of matrices and its applications[M]. Singapore: World Scientific, 2012.
- [16] CHENG D, QI H, LI Z. Analysis and control of Boolean networks: a semi-tensor product approach[J]. *Acta automatica sinica*, 2011, 37(5) : 529–540.

责任编辑: 周建军

(上接第 21 页)

- 2020, 98: 150–161.
- [19] JIANG L, LI S, WANG N, et al. Preparation of dextran casein phosphopeptide conjugates, evaluation of its calcium binding capacity and digestion in vitro[J]. *Food chemistry*, 2021, 352: 129332.
- [20] NARA M, MORII H, TANOKURA M. Coordination to divalent cations by calcium-binding proteins studied by FTIR spectroscopy[J]. *BBA-Biomembranes*, 2013, 1828(10) : 2319–2327.
- [21] 丁媛媛, 王莉, 张新霞, 等. 麦胚多肽-钙螯合物制备工艺优化及其结构表征[J]. *食品科学*, 2017, 38(10) : 215–221.

责任编辑: 郎婧

(上接第 74 页)

- [EB/OL]. [2022-06-01]. <https://arxiv.org/abs/2110.11154v1>.
- [11] KANG S K, HWANG J, LEE D, et al. Semi-supervised learning for cross-domain recommendation to cold-start users[C]//The 28th ACM International Conference on Information and Knowledge Management. New York: Association for Computing Machinery, 2019: 1563–1572.
- [12] KOREN Y, BELL R, VOLINSKY C. Matrix factorization techniques for recommender systems[J]. *Computer*, 2009, 42(8) : 30–37.
- [13] RENDLE S. Factorization machines[C]//IEEE. 2010 IEEE International Conference on Data Mining. New York: IEEE, 2010: 5694074.
- [14] HUANG P S, HE X, GAO J, et al. Learning deep structured semantic models for web search using clickthrough data[C]//ACM. 22nd ACM International Conference on Information and Knowledge Management. New York: Association for Computing Machinery, 2013: 2333–2338.

责任编辑: 郎婧