

# 一种新的用于跨领域推荐的迁移学习模型

王 俊<sup>1)</sup> 李石君<sup>1)</sup> 杨 莎<sup>1),2)</sup> 金 红<sup>1)</sup> 余 伟<sup>1)</sup>

<sup>1)</sup>(软件工程国家重点实验室 武汉大学计算机学院 武汉 430072)

<sup>2)</sup>(汉口学院计算机科学与技术学院 武汉 430212)

**摘 要** 协同过滤是一种简单常用的推荐方法,但是当目标数据非常稀疏时,其性能会严重退化,借助与目标数据跨域关联的辅助数据进行跨领域推荐是解决此问题的一种有效途径.已有的跨领域推荐模型大多假设不同领域完全共享一个评分模式,忽略了领域特有评分模式,可能导致推荐性能退化.此外,许多模型基于单一桥梁迁移跨领域信息,正迁移不足.特别是在考虑领域特有被评分模式的前提下,据作者所知目前还没有模型利用项目的共享被评分模式进行跨领域推荐.因此,该文提出一种新的三元桥迁移学习模型,用于跨领域推荐.首先通过评分矩阵的集合分解提取用户的潜在因子和共享评分模式,以及项目的潜在因子和共享被评分模式,在此过程中考虑了领域特有模式,并对潜在因子施加相似性约束;然后利用潜在因子中的聚类信息构造邻接图;最后通过用户端和项目端的基于共享模式、潜在因子和邻接图的三元桥迁移学习联合预测缺失评分.在三个公开的真实数据集上进行的大量实验表明,该模型的推荐精度优于一些目前最先进的推荐模型.

**关键词** 迁移学习;推荐;协同过滤;跨领域;稀疏;矩阵分解

**中图法分类号** TP311 **DOI号** 10.11897/SP.J.1016.2017.02367

## A New Transfer Learning Model for Cross-Domain Recommendation

WANG Jun<sup>1)</sup> LI Shi-Jun<sup>1)</sup> YANG Sha<sup>1),2)</sup> JIN Hong<sup>1)</sup> YU Wei<sup>1)</sup>

<sup>1)</sup>(State Key Laboratory of Software Engineering, School of Computer Science, Wuhan University, Wuhan 430072)

<sup>2)</sup>(School of Computer Science and Technology, Hankou University, Wuhan 430212)

**Abstract** Collaborative filtering is a recommendation method to predict the missing ratings from target users to target items based on the historical rating data of users similar to target users or items similar to target items. Collaborative filtering is simple and usual, but its performance will degrade seriously when the target data is very sparse. Cross-domain recommendation with the help of auxiliary data associated across domains with target data is an effective way to solve this problem. Most existing cross-domain recommendation models suppose that different domains share a rating pattern completely, which ignores domain-specific rating patterns and may result in degradations of the recommendation performance. In addition, many models transfer cross-domain information based on a single bridge, the positive transfer of which is insufficient. In particular, as far as we know, no model has utilized the shared rated pattern of items for cross-domain recommendation under the premise of considering domain-specific rated patterns. Existing cross-domain recommendation models ignore that the shared rating patterns may have different focuses on users and items, and always only utilize the rating pattern of users. In fact, the rating pattern

收稿日期:2016-12-13;在线出版日期:2017-03-28. 本课题得到国家自然科学基金项目(61272109)、中央高校基本科研业务费专项资金项目(2042014KF0057)、湖北省自然科学基金项目(2014CFB289)资助. 王 俊,男,1986年生,博士研究生,主要研究方向为迁移学习、推荐系统、数据挖掘. E-mail: wjwj@whu.edu.cn. 李石君(通信作者),男,1964年生,博士,教授,中国计算机学会(CCF)会员,主要研究领域为大数据、互联网搜索与挖掘、数据库技术、移动数据挖掘与时空一致性研究. E-mail: shjli@whu.edu.cn. 杨 莎,女,1980年生,博士研究生,讲师,主要研究方向为电子商务数据挖掘与分析、电子商务服务质量评估、互联网经济行为数据挖掘. 金 红,女,1984年生,博士,主要研究方向为机器学习、复杂网络、社区发现. 余 伟,男,1987年生,博士,讲师,主要研究方向为数据抽取、数据融合、数据质量评估.

of users focuses on the interest pattern of users, however, the rated pattern of items focuses on the popularity pattern of items. Therefore, we propose a novel triple-bridge transfer (TRBT) learning model for cross-domain recommendation. Firstly we extract latent factor and shared rating pattern of users as well as latent factor and shared rated pattern of items by collective factorizations on rating matrices, while considering domain-specific patterns and imposing similarity constraints on latent factors; then we construct adjacency graphs utilizing clustering information contained in latent factors; finally we predict the missing ratings jointly by user-side and item-side triple-bridge transfer learning based on shared pattern, latent factors and adjacency graphs. During the triple-bridge transfer, the interactive information between users and items is transferred based on the shared patterns, the feature information of users and items is transferred based on the latent factors, and the adjacency information of users and items is transferred based on the adjacency graphs. We highlight the different focuses of the shared pattern on users and items and utilize the shared rated patterns of items for the first time. Our proposed TRBT model not only can transfer useful cross-domain information of three different types, but also can choose transfer bridges and adjust transfer weights of different bridges flexibly according to the specific situation. Although TRBT model may seem a bit complex, the core algorithms of it can be solved by iterations with a small number and converge with a fast speed, so it can be applied to cross-domain recommendation under the environment of massive data in theory. We adopt both mean absolute error (MAE) and root mean square error (RMSE), which are two popular evaluation metrics for collaborative filtering recommendation models at present, to evaluate the predicting accuracy for missing ratings of different recommendation models. Extensive experiments on three public real world datasets demonstrate that the recommendation accuracy of TRBT model outperforms several state-of-the-art recommendation models.

**Keywords** transfer learning; recommendation; collaborative filtering; cross-domain; sparse; matrix factorization

## 1 引 言

协同过滤是一种广泛应用于推荐系统的推荐方法,它基于用户对项目的历史评分数据,利用与目标用户相似的用户或与目标项目相似的项目预测目标用户对目标项目的缺失评分<sup>[1-2]</sup>.协同过滤不需要用户或项目的内容信息,简单实用,但它依赖历史评分数据,历史评分数据的稀疏性问题成为它的主要瓶颈<sup>[3]</sup>.

在现实中,当目标领域中的评分数据很稀疏时,经常存在许多与其跨域关联的评分数据.例如,在一个出售电影票、图书、唱片等商品的电子商务网站中,当用户对图书的评分很少时,发现存在许多用户对电影和唱片的评分.事实上,用户如果喜欢某种风格的电影或音乐,可能也会喜欢同种风格的图书,因

而可利用用户对电影和唱片的评分数据为用户推荐图书.这属于跨领域推荐问题,目的是利用与目标领域数据相关联的辅助领域数据来辅助目标领域中的推荐.通过跨领域推荐,可有效解决目标数据的稀疏性问题.迁移学习是一种流行的机器学习方法,可用于学习辅助数据中的知识并将其迁移到目标数据进行跨领域推荐<sup>[4-6]</sup>.然而,迁移学习能将辅助数据中的有用信息迁移到目标数据辅助推荐,产生正迁移;也可能同时迁移一些干扰的信息误导推荐,产生负迁移.许多已有的用于跨领域推荐的迁移学习模型还存在一些不足:

(1) 基于评分模式进行迁移学习,但假设关联的领域完全共享一个评分模式,忽略了领域间的评分模式差异,可能导致严重的负迁移<sup>[7]</sup>.

(2) 忽略了共享的评分模式对用户和项目的不同侧重,只利用侧重于用户兴趣模式的用户评分模

式,没有利用侧重于项目受欢迎模式的项目被评分模式。

(3)仅基于潜在因子或评分模式等单一桥梁进行迁移学习,可能无法产生足够的正迁移。

为了解决上述问题,我们提出一种新的用于跨领域协同过滤推荐的三元桥迁移学习模型,在有效减少负迁移的同时显著增加了正迁移。我们的主要贡献如下:

(1)基于相似性约束,通过评分矩阵的集合分解从辅助领域中提取用户和项目的潜在因子,以及与目标领域共享的用户评分模式和项目被评分模式,并考虑领域特有模式,从而有效减少了负迁移。

(2)首次凸显了共享评分模式的不同侧重,并基于共享的用户评分模式和项目被评分模式,分别迁移用户的兴趣模式信息和项目的受欢迎模式信息。

(3)提取用户和项目的潜在因子中的聚类信息,分别构造用户和项目的邻接图。基于共享的用户评分模式、潜在因子和邻接图进行用户端的迁移学习,基于共享的项目被评分模式、潜在因子和邻接图进行项目端的迁移学习,并联合预测目标评分矩阵中的缺失评分,从而显著增加了正迁移。

本文在第2节介绍论文的相关研究工作;在第3节形式化定义论文的研究问题;在第4节简要介绍两种基于矩阵分解的基础推荐模型;在第5节详细介绍我们新提出的一种用于跨领域推荐的迁移学习模型;在第6节从理论上分析该新模型的优势,并将其与目前一些最先进的模型进行对比分析;在第7节通过大量实验,验证该新模型的推荐精度优于对比模型;在第8节总结论文并展望我们未来的研究方向。

## 2 相关工作

当目标数据非常稀疏时,如何利用与目标数据跨域关联的较密集的辅助数据进行跨领域推荐<sup>[8]</sup>成为近年来的研究热点。Mirbakhsh等人<sup>[9]</sup>利用跨领域信息改善冷启动用户的推荐。Tan等人<sup>[10]</sup>有效融合媒体描述、用户生成的文本以及评分等多类型媒体信息,通过跨领域的用户兴趣迁移解决推荐系统的稀疏性问题。跨领域信息可能辅助推荐,也可能误导推荐,其中如何选择和处理辅助领域数据是一个关键问题。Yi等人<sup>[11]</sup>研究了跨领域协同过滤中的辅助领域选择问题,发现更低的KL散度相对更高的用户重叠度更有意义。Li等人<sup>[12]</sup>提出了一种用于匹

配不同领域中重叠的用户或项目的新模型,并利用匹配的映射关系进行跨领域推荐。Zhao等人<sup>[13]</sup>通过主动学习构造实体一致性映射,从而促进不同推荐系统之间的信息迁移。

识别不同领域中数据间的映射关系需要花费较大成本,因此许多跨领域推荐模型直接利用领域之间共享的潜在结构作为信息迁移的桥梁。Shi等人<sup>[14]</sup>基于用户生成的标签提取项目在不同领域中的共同特征,通过标签迁移学习进行跨领域协同过滤推荐。Pan等人<sup>[15]</sup>提出一种坐标系迁移模型,提取用户和项目的潜在因子,并基于它们进行跨领域的特征信息迁移。Li等人<sup>[16]</sup>提出了一种经典的密码本迁移模型,从辅助数据中迁移聚类层次的评分模式到目标数据,称为CBT(CodeBook Transfer)模型。然而,跨领域的信息迁移对目标领域的推荐不一定是有帮助的。例如,上述CBT模型假设所有领域完全共享一个评分模式,忽略了领域间评分模式的差异,可能导致严重的负迁移,使推荐性能退化。Gao等人<sup>[7]</sup>提出了一种聚类层次的潜在因子模型,不仅提取了领域间共享的用户评分模式,还考虑了各领域特有的用户评分模式,缓解了负迁移,但是该模型没有从项目的角度研究项目在不同用户领域中被评分模式的差异。而且据我们所知,目前还没有模型在考虑领域特有被评分模式的前提下,利用项目在不同用户领域中的共享被评分模式进行跨领域推荐。另外需要指出的是,如果只基于标签、潜在因子或评分模式等单一桥梁进行信息迁移,正迁移可能不足。

Gu等人<sup>[17]</sup>基于用户或项目的特征信息和近邻信息构造信息图,并通过图正则化加权非负矩阵三元分解来利用用户或项目的关联信息,称为GMTF(Graph Regularized Weighted Nonnegative Matrix Tri-Factorization)模型。然而,GMTF模型基于单一领域进行协同过滤,且需要密集的数据来构造信息图,当目标数据非常稀疏时其推荐性能将严重下降。为了迁移更多有用信息到稀疏的目标数据,Shi等人<sup>[18]</sup>基于GMTF模型从辅助数据中提取潜在因子并构造相似度图,然后同时利用潜在因子和相似度图进行迁移学习,称为TWBT(Twin Bridge Transfer)模型,但是它忽略了评分模式信息,且要求辅助数据中的用户或项目与目标数据完全相同。

不同于上述模型,我们提出一种新的跨领域推荐模型。在考虑了领域特有模式的基础上,我们首次凸显了共享模式的不同侧重,并通过用户端和项目端的基于共享模式、潜在因子和邻接图的三元桥迁

移学习联合预测缺失评分. 我们提出的新模型在有效减少负迁移的同时, 显著增加了正迁移, 从而得到更高的推荐精度; 而且, 因为充分利用了共享的用户评分模式信息以及项目被评分模式信息, 该新模型在用户或项目的重叠度较低时仍然有效.

### 3 问题定义

假设  $\mathbf{R} \in \mathbf{R}^{M \times N}$  为数据比较稀疏的目标评分矩阵, 表示  $M$  个用户对  $N$  个项目的评分.  $\mathbf{Z}$  为  $\mathbf{R}$  的指标矩阵, 与  $\mathbf{R}$  维数相同, 用来标明  $\mathbf{R}$  中的评分是否缺失. 若  $\mathbf{R}_{ij}$  为已观察评分, 则  $\mathbf{Z}_{ij} = 1$ ; 否则  $\mathbf{Z}_{ij} = 0$ . 假设  $\mathbf{R}_1$  和  $\mathbf{R}_2$  为两个数据比较密集的辅助评分矩阵, 其中  $\mathbf{R}_1 \in \mathbf{R}^{M \times N_1}$ , 表示  $M$  个用户对  $N_1$  个项目的评分;  $\mathbf{R}_2 \in \mathbf{R}^{M_2 \times N}$ , 表示  $M_2$  个用户对  $N$  个项目的评分.  $\mathbf{R}_1$  和  $\mathbf{R}_2$  对应的指标矩阵分别为  $\mathbf{Z}_1$  和  $\mathbf{Z}_2$ ,  $\mathbf{Z}_1$  和  $\mathbf{Z}_2$  的定义与  $\mathbf{Z}$  同理.  $\mathbf{R}_1$  与  $\mathbf{R}$  含有相似的用户和领域关联的项目,  $\mathbf{R}_2$  与  $\mathbf{R}$  含有相似的项目和领域关联的用户. 请注意, 鉴于我们研究的跨领域信息迁移问题是基于矩阵分解而展开, 为了便于进行矩阵的集合分解, 这里我们假设  $\mathbf{R}_1$  与  $\mathbf{R}$  含有维数相同的用户,  $\mathbf{R}_2$  与  $\mathbf{R}$  含有维数相同的项目.

为了解决  $\mathbf{R}$  的数据稀疏性问题, 我们研究如何利用  $\mathbf{R}_1$  和  $\mathbf{R}_2$  辅助预测  $\mathbf{R}$  中的缺失评分, 从而提高  $\mathbf{R}$  中缺失评分的预测精度.

### 4 基础模型简介

#### 4.1 GMTF 模型<sup>[17]</sup>

GMTF 模型首先构造用户图和项目图, 然后对目标评分矩阵  $\mathbf{R}$  进行图正则化非负矩阵三元分解, 并基于分解后得到的  $\mathbf{UBV}^T$  预测  $\mathbf{R}$  中的缺失评分, 其目标最优化问题为

$$\min_{\mathbf{U}, \mathbf{V}, \mathbf{B} \geq 0} O = \|\mathbf{Z} \odot (\mathbf{R} - \mathbf{UBV}^T)\|_F^2 + \gamma_U G_U + \gamma_V G_V \quad (1)$$

这里  $O$  为目标最优化函数,  $\|\cdot\|_F$  表示矩阵的  $F$  范数,  $\odot$  表示 element-wise 积; 矩阵  $\mathbf{U}$  和  $\mathbf{V}$  分别表示用户和项目的潜在特征因子, 矩阵  $\mathbf{B}$  表示用户对项目的评分模式;  $G_U$  和  $G_V$  分别表示用户图和项目图的正则化信息, 其中  $G_U = \text{tr}(\mathbf{U}^T \mathbf{L}_U \mathbf{U})$ ,  $G_V = \text{tr}(\mathbf{V}^T \mathbf{L}_V \mathbf{V})$  ( $\text{tr}$  表示矩阵的迹),  $\mathbf{L}_U$  和  $\mathbf{L}_V$  分别表示用户图和项目图的拉普拉斯矩阵;  $\gamma_U$  和  $\gamma_V$  为正则化参数.

#### 4.2 TWBT 模型<sup>[18]</sup>

TWBT 模型首先利用 GMTF 模型分别从  $\mathbf{R}_1$  和  $\mathbf{R}_2$  中提取用户潜在因子  $\mathbf{U}_0$  和项目潜在因子  $\mathbf{V}_0$ , 并构造用户相似度图  $P_U$  和项目相似度图  $P_V$ , 其中  $P_U$

和  $P_V$  的邻接矩阵分别为  $\mathbf{Q}_U$  和  $\mathbf{Q}_V$ ; 然后基于  $\mathbf{U}_0, \mathbf{V}_0, P_U$  和  $P_V$  进行迁移学习并预测  $\mathbf{R}$  中的缺失评分. TWBT 模型的目标最优化问题为

$$\min_{\mathbf{U}, \mathbf{V}, \mathbf{B} \geq 0} O = \|\mathbf{Z} \odot (\mathbf{R} - \mathbf{UBV}^T)\|_F^2 + \lambda_U \|\mathbf{U} - \mathbf{U}_0\|_F^2 + \lambda_V \|\mathbf{V} - \mathbf{V}_0\|_F^2 + \gamma_U G_U + \gamma_V G_V \quad (2)$$

这里  $O, \mathbf{U}, \mathbf{B}, \mathbf{V}$  的含义同 GMTF 模型;  $\lambda_U, \lambda_V, \gamma_U, \gamma_V$  为正则化参数;  $G_U$  和  $G_V$  分别表示  $P_U$  和  $P_V$  的正则化信息, 其中  $G_U = \text{tr}(\mathbf{U}^T \mathbf{L}_U \mathbf{U})$ ,  $G_V = \text{tr}(\mathbf{V}^T \mathbf{L}_V \mathbf{V})$ ,  $\mathbf{L}_U$  和  $\mathbf{L}_V$  为图拉普拉斯矩阵, 分别基于  $\mathbf{Q}_U$  和  $\mathbf{Q}_V$  生成. 若  $u_{j_*}^0 \in N_p(u_{i_*}^0)$  或  $u_{i_*}^0 \in N_p(u_{j_*}^0)$ ,  $(\mathbf{Q}_U)_{ij} = 1$ ; 否则  $(\mathbf{Q}_U)_{ij} = 0$ . 这里  $u_{j_*}^0$  表示  $\mathbf{U}_0$  的第  $j$  行,  $N_p(u_{i_*}^0)$  表示  $u_{i_*}^0$  的  $p$  最近邻集, 另外  $\mathbf{Q}_V$  的定义与  $\mathbf{Q}_U$  同理.

### 5 TRBT 模型

受到上述模型的启发, 我们提出一种新的三元桥迁移学习模型, 通过潜在因子、邻接图、共享模式合成的三元桥, 迁移辅助评分矩阵  $\mathbf{R}_1$  和  $\mathbf{R}_2$  中的有用信息到目标评分矩阵  $\mathbf{R}$ , 从而辅助预测  $\mathbf{R}$  中的缺失评分, 称为 TRBT (Triple-Bridge Transfer) 模型, 其基本流程如图 1 所示.

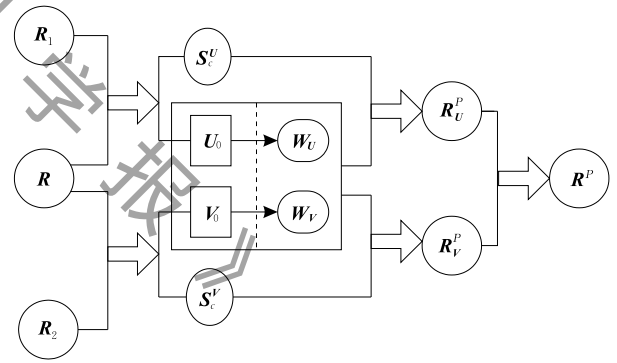


图 1 TRBT 模型的基本流程

首先从  $\mathbf{R}_1$  中提取与  $\mathbf{R}$  相似的用户潜在因子  $\mathbf{U}_0$  以及与  $\mathbf{R}$  共享的用户评分模式  $S^U$ , 并在  $\mathbf{R}_2$  中提取与  $\mathbf{R}$  相似的项目潜在因子  $\mathbf{V}_0$  以及与  $\mathbf{R}$  共享的项目被评分模式  $S^V$ . 然后基于  $\mathbf{U}_0$  和  $\mathbf{V}_0$  分别构造用户邻接图  $\mathbf{W}_U$  和项目邻接图  $\mathbf{W}_V$ . 随后基于  $S^U, \mathbf{U}_0, \mathbf{V}_0, \mathbf{W}_U, \mathbf{W}_V$  进行用户端的三元桥迁移学习, 得到用户端的  $\mathbf{R}$  的预测评分矩阵  $\mathbf{R}_U^P$ ; 基于  $S^V, \mathbf{U}_0, \mathbf{V}_0, \mathbf{W}_U, \mathbf{W}_V$  进行项目端的三元桥迁移学习, 得到项目端的  $\mathbf{R}$  的预测评分矩阵  $\mathbf{R}_V^P$ . 最后基于  $\mathbf{R}_U^P$  和  $\mathbf{R}_V^P$  得到  $\mathbf{R}$  的最终预测评分矩阵  $\mathbf{R}^P$ .

#### 5.1 潜在因子和共享模式的提取

在非负性及正交化约束条件下, 并基于相似度正则化, 我们对评分矩阵  $\mathbf{R}$  和  $\mathbf{R}_1$  进行集合分解, 从

$\mathbf{R}_1$  中提取与  $\mathbf{R}$  相似的用户潜在因子  $\mathbf{U}_0$  以及与  $\mathbf{R}$  共享的用户评分模式  $S_c^u$ ; 对评分矩阵  $\mathbf{R}$  和  $\mathbf{R}_1$  进行集合分解, 从  $\mathbf{R}_2$  中提取与  $\mathbf{R}$  相似的项目潜在因子  $\mathbf{V}_0$  以及与  $\mathbf{R}$  共享的项目被评分模式  $S_c^v$ .

### 5.1.1 用户潜在因子和共享的用户评分模式

因为辅助评分矩阵  $\mathbf{R}_1$  与目标评分矩阵  $\mathbf{R}$  含有相似的用户, 以及属于不同领域但相关联的项目, 我们提取  $\mathbf{R}_1$  中与  $\mathbf{R}$  相似的用户潜在因子  $\mathbf{U}_0$ , 以及与  $\mathbf{R}$  共享的用户评分模式  $S_c^u$ , 作为  $\mathbf{R}$  和  $\mathbf{R}_1$  之间信息迁移的桥梁.

我们对  $\mathbf{R}$  和  $\mathbf{R}_1$  进行正交化非负矩阵集合分解<sup>[7,19]</sup>, 并求解目标最优化问题:

$$\min_{\mathbf{U}, \mathbf{U}_1, S_0^u, S_1^u, S_c^u, \mathbf{V}, \mathbf{V}_1} f_1 = \|\mathbf{R} - \mathbf{U}[\mathbf{S}_0^u, S_c^u] \mathbf{V}^T\|_F^2 + \|\mathbf{R}_1 - \mathbf{U}_1[\mathbf{S}_0^u, S_1^u] \mathbf{V}_1^T\|_F^2 + \alpha_u \|\mathbf{U} - \mathbf{U}_1\|_F^2 \quad (3)$$

其中  $\mathbf{U}^T \mathbf{U} = \mathbf{I}$ ,  $\mathbf{U}_1^T \mathbf{U}_1 = \mathbf{I}$ ,  $\mathbf{V}_1^T \mathbf{V}_1 = \mathbf{I}$ ,  $\mathbf{V}^T \mathbf{V} = \mathbf{I}$  为正交化约束;  $\mathbf{R}$ ,  $\mathbf{R}_1$ ,  $\mathbf{Z}$  和  $\mathbf{Z}_1$  已在问题定义中说明;  $S_0^u$  表示  $\mathbf{R}$  和  $\mathbf{R}_1$  共享的用户评分模式,  $S_c^u$  和  $S_1^u$  分别表示  $\mathbf{R}$  和  $\mathbf{R}_1$  特有的用户评分模式, 在用户评分模式矩阵中, 每行对应一个用户聚类, 每列对应一个项目聚类, 每个元素表示用户聚类对项目聚类的评分模式;  $\mathbf{U}$  和  $\mathbf{U}_1$  分别表示  $\mathbf{R}$  和  $\mathbf{R}_1$  的用户潜在因子, 其中每行对应一个用户, 每列对应一个用户聚类;  $\mathbf{V}$  和  $\mathbf{V}_1$  分别表示  $\mathbf{R}$  和  $\mathbf{R}_1$  的项目潜在因子, 其中每行对应一个项目, 每列对应一个项目聚类;  $\alpha_u$  为正则化参数, 用来约束  $\mathbf{U}$  和  $\mathbf{U}_1$  的相似度.

除了非负性约束, 我们还对潜在因子矩阵  $\mathbf{U}$ ,  $\mathbf{U}_1$ ,  $\mathbf{V}$  和  $\mathbf{V}_1$  等施加了正交化约束, 但请注意, 这里只是基于正交化的思想对它们进行约束, 不代表它们都为正交矩阵. 在非负性以及正交化约束条件下,  $\mathbf{U}$ ,  $\mathbf{U}_1$ ,  $\mathbf{V}$  和  $\mathbf{V}_1$  等潜在因子矩阵中的每行都只有一个非零元, 且该非零元为正数. 若  $U_{ij} > 0$ , 则表示  $\mathbf{U}$  中第  $i$  行对应的用户属于第  $j$  列对应的用户聚类; 若  $V_{ij} > 0$ , 则表示  $\mathbf{V}$  中第  $i$  行对应的项目属于第  $j$  列对应的项目聚类. 在  $\mathbf{U}_1$  和  $\mathbf{V}_1$  中, 非零元的含义与此类同.

下面我们使用交替最小二乘法求解方程式 (3) 描述的目标最优化问题.

假设  $S_0^u \in \mathbf{R}^{K \times D}$ ,  $S_c^u \in \mathbf{R}^{K \times (L-D)}$ ,  $S_1^u \in \mathbf{R}^{K \times (L_1-D)}$ ,  $\mathbf{U} \in \mathbf{R}^{M \times K}$ ,  $\mathbf{U}_1 \in \mathbf{R}^{M \times K}$ .

令  $\mathbf{V} = [\mathbf{V}_{00}^T, \mathbf{V}_{01}^T] \in \mathbf{R}^{N \times L}$ ,  $\mathbf{V}_1 = [\mathbf{V}_{10}^T, \mathbf{V}_{11}^T] \in \mathbf{R}^{N_1 \times L_1}$ , 其中  $\mathbf{V}_{00}^T = \mathbf{V}(:, 1:D)$ ,  $\mathbf{V}_{01}^T = \mathbf{V}(:, (D+1):L)$ ,  $\mathbf{V}_{10}^T = \mathbf{V}_1(:, 1:D)$ ,  $\mathbf{V}_{11}^T = \mathbf{V}_1(:, (D+1):L_1)$ .

这里我们以  $S_c^u$  的更新规则推导为例, 固定其它矩阵变量, 方程式 (3) 被改写为

$$\min_{S_c^u} f_1(S_c^u) = \|\mathbf{R} - \mathbf{U} S_0^u \mathbf{V}_{00} - \mathbf{U} S_c^u \mathbf{V}_{01}\|_{\odot \mathbf{Z}}^2 + \|\mathbf{R}_1 - \mathbf{U}_1 S_0^u \mathbf{V}_{10} - \mathbf{U}_1 S_1^u \mathbf{V}_{11}\|_{\odot \mathbf{Z}_1}^2 \quad (4)$$

对  $S_c^u$  求偏导:

$$\frac{\partial f_1(S_c^u)}{\partial S_c^u} = 2(\mathbf{U}^T([\mathbf{U} S_0^u \mathbf{V}_{00}] \odot \mathbf{Z}) \mathbf{V}_{01}^T - \mathbf{U}^T(\mathbf{R} \odot \mathbf{Z}) \mathbf{V}_{01}^T) + 2\mathbf{U}^T([\mathbf{U} S_c^u \mathbf{V}_{01}] \odot \mathbf{Z}) \mathbf{V}_{01}^T.$$

使用保证  $S_c^u$  非负性的 KKT 互补条件, 得到  $[\mathbf{U}^T([\mathbf{U} S_0^u \mathbf{V}_{00}] \odot \mathbf{Z}) \mathbf{V}_{01}^T + \mathbf{U}^T([\mathbf{U} S_c^u \mathbf{V}_{01}] \odot \mathbf{Z}) \mathbf{V}_{01}^T - \mathbf{U}^T(\mathbf{R} \odot \mathbf{Z}) \mathbf{V}_{01}^T] \odot S_c^u = 0$ .

从而得到  $S_c^u$  的更新规则:

$$S_c^u \leftarrow S_c^u \sqrt{\frac{\mathbf{U}^T(\mathbf{R} \odot \mathbf{Z}) \mathbf{V}_{01}^T}{\mathbf{U}^T([\mathbf{U} S_0^u \mathbf{V}_{00}] \odot \mathbf{Z}) \mathbf{V}_{01}^T + \mathbf{U}^T([\mathbf{U} S_c^u \mathbf{V}_{01}] \odot \mathbf{Z}) \mathbf{V}_{01}^T}} \quad (5)$$

请注意, 方程式 (5) 表示其左右两边矩阵中一一对应(下标相同)的元素之间的更新规则, 为了简化表示, 文中我们统一省去元素下标.

类似地, 我们可得到其他矩阵变量的更新规则:

$$S_1^u \leftarrow S_1^u \sqrt{\frac{\mathbf{U}_1^T(\mathbf{R}_1 \odot \mathbf{Z}_1) \mathbf{V}_{11}^T}{\mathbf{U}_1^T([\mathbf{U}_1 S_0^u \mathbf{V}_{10}] \odot \mathbf{Z}_1) \mathbf{V}_{11}^T + \mathbf{U}_1^T([\mathbf{U}_1 S_1^u \mathbf{V}_{11}] \odot \mathbf{Z}_1) \mathbf{V}_{11}^T}} \quad (6)$$

$$\mathbf{U} \leftarrow \mathbf{U} \sqrt{\frac{(\mathbf{R} \odot \mathbf{Z}) \mathbf{V} [S_0^u, S_c^u]^T + \alpha_u \mathbf{U}}{([\mathbf{U} [S_0^u, S_c^u] \mathbf{V}^T \odot \mathbf{Z}] \odot \mathbf{V} [S_0^u, S_c^u]^T + \alpha_u \mathbf{U})}} \quad (7)$$

$$\mathbf{U}_1 \leftarrow \mathbf{U}_1 \sqrt{\frac{(\mathbf{R}_1 \odot \mathbf{Z}_1) \mathbf{V}_1 [S_0^u, S_1^u]^T + \alpha_u \mathbf{U}_1}{([\mathbf{U}_1 [S_0^u, S_1^u] \mathbf{V}_1^T \odot \mathbf{Z}_1] \odot \mathbf{V}_1 [S_0^u, S_1^u]^T + \alpha_u \mathbf{U}_1)}} \quad (8)$$

$$\mathbf{V} \leftarrow \mathbf{V} \sqrt{\frac{[S_0^u, S_c^u]^T \mathbf{U}^T (\mathbf{R} \odot \mathbf{Z})}{[S_0^u, S_c^u]^T \mathbf{U}^T ([\mathbf{U} [S_0^u, S_c^u] \mathbf{V}^T \odot \mathbf{Z}] \odot \mathbf{V} [S_0^u, S_c^u]^T)}} \quad (9)$$

$$\mathbf{V}_1 \leftarrow \mathbf{V}_1 \sqrt{\frac{[S_0^u, S_1^u]^T \mathbf{U}_1^T (\mathbf{R}_1 \odot \mathbf{Z}_1)}{[S_0^u, S_1^u]^T \mathbf{U}_1^T ([\mathbf{U}_1 [S_0^u, S_1^u] \mathbf{V}_1^T \odot \mathbf{Z}_1] \odot \mathbf{V}_1 [S_0^u, S_1^u]^T)}} \quad (10)$$

$$S_0^u \leftarrow S_0^u \sqrt{\frac{\mathbf{U}^T(\mathbf{R} \odot \mathbf{Z}) \mathbf{V}_{00}^T + \mathbf{U}_1^T(\mathbf{R}_1 \odot \mathbf{Z}_1) \mathbf{V}_{10}^T}{E + F}} \quad (11)$$

对于更新规则 (11), 其中的  $E$  和  $F$  分别为

$$E = \mathbf{U}^T([\mathbf{U} S_0^u \mathbf{V}_{00}] \odot \mathbf{Z}) \mathbf{V}_{00}^T + \mathbf{U}^T([\mathbf{U} S_c^u \mathbf{V}_{01}] \odot \mathbf{Z}) \mathbf{V}_{00}^T \quad (12)$$

$$F = \mathbf{U}_1^T([\mathbf{U}_1 S_0^u \mathbf{V}_{10}] \odot \mathbf{Z}_1) \mathbf{V}_{10}^T + \mathbf{U}_1^T([\mathbf{U}_1 S_1^u \mathbf{V}_{11}] \odot \mathbf{Z}_1) \mathbf{V}_{10}^T \quad (13)$$

参考文献 [7] 可证明, 基于更新规则 5~11 迭代更新矩阵变量, 方程式 (3) 中的目标函数  $f_1$  将单调下降, 最终收敛. 通过足够的迭代, 我们得到  $f_1$  收敛时  $\mathbf{U}_1$  和  $S_0^u$  的值, 分别记为  $\mathbf{U}_0$  和  $S_c^u$ .

### 5.1.2 项目潜在因子和共享的项目被评分模式

评分模式对用户和项目可有不同的侧重, 若侧

重于用户对不同项目评分的模式,称为用户评分模式;若侧重于项目被不同用户评分的模式,称为项目被评分模式.相应地,基于共享用户评分模式的迁移学习侧重于相似用户在不同项目领域的兴趣模式迁移;基于共享项目被评分模式的迁移学习侧重于相似项目在不同用户领域的受欢迎模式迁移.为了增加正迁移,我们希望同时利用共享的用户评分模式和项目被评分模式进行跨领域信息迁移.

因为辅助评分矩阵  $\mathbf{R}_2$  与目标评分矩阵  $\mathbf{R}$  含有相似的项目,以及属于不同领域但相关联的用户,我们提取  $\mathbf{R}_2$  中与  $\mathbf{R}$  相似的项目潜在因子  $\mathbf{V}_0$ ,以及与  $\mathbf{R}$  共享的项目被评分模式  $\mathbf{S}_c^V$ ,作为  $\mathbf{R}$  和  $\mathbf{R}_2$  之间信息迁移的桥梁.

我们对  $\mathbf{R}$  和  $\mathbf{R}_2$  进行正交化非负矩阵集合分解,并求解目标最优化问题:

$$\min_{\mathbf{U}, \mathbf{U}_2, \mathbf{S}_c^U, \mathbf{S}_c^V, \mathbf{S}_1^U, \mathbf{S}_1^V, \mathbf{V}, \mathbf{V}_2 \geq 0} f_2 = \|\mathbf{R} - \mathbf{U}[\mathbf{S}_c^U, \mathbf{S}_c^V]\mathbf{V}^T\|_F^2 + \|\mathbf{R}_2 - \mathbf{U}_2[\mathbf{S}_c^U, \mathbf{S}_c^V]\mathbf{V}_2^T\|_F^2 + \alpha_v \|\mathbf{V} - \mathbf{V}_2\|_F^2 \quad (14)$$

其中  $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ ,  $\mathbf{U}_2^T\mathbf{U}_2 = \mathbf{I}$ ,  $\mathbf{V}_2^T\mathbf{V}_2 = \mathbf{I}$ ,  $\mathbf{V}^T\mathbf{V} = \mathbf{I}$ ;  $\mathbf{S}_c^U$  表示  $\mathbf{R}$  和  $\mathbf{R}_2$  共享的用户评分模式;  $\mathbf{S}_c^V$  和  $\mathbf{S}_1^U$  分别表示  $\mathbf{R}$  和  $\mathbf{R}_2$  特有的用户评分模式;  $\mathbf{U}$  和  $\mathbf{U}_2$  分别为  $\mathbf{R}$  和  $\mathbf{R}_2$  的用户潜在因子,  $\mathbf{V}$  和  $\mathbf{V}_2$  分别为  $\mathbf{R}$  和  $\mathbf{R}_2$  的项目潜在因子;  $\alpha_v$  为正则化参数,用来约束  $\mathbf{V}$  和  $\mathbf{V}_2$  的相似度;与方程式(3)的求解过程类似,我们通过求解方程式(14),得到目标函数  $f_2$  收敛时  $\mathbf{V}_2$  和  $\mathbf{S}_c^U$  的值,分别记为  $\mathbf{V}_0$  和  $\mathbf{S}_c^U$ .

## 5.2 邻接图构造

当目标评分矩阵  $\mathbf{R}$  中的数据非常稀疏时,兴趣相同的用户很少对同一项目评分,特征相似的项目也很少被同一用户评分,这时很难利用  $\mathbf{R}$  中用户或项目的邻接信息.因为与  $\mathbf{R}$  相关的两个辅助评分矩阵  $\mathbf{R}_1$  和  $\mathbf{R}_2$  含有比  $\mathbf{R}$  更密集的数据,潜在因子  $\mathbf{U}_0$  和  $\mathbf{V}_0$  中的数据又比  $\mathbf{R}_1$  和  $\mathbf{R}_2$  更密集,所以我们基于  $\mathbf{U}_0$  和  $\mathbf{V}_0$ ,分别构造用户和项目的邻接图,然后基于邻接图将  $\mathbf{R}_1$  中的用户邻接信息和  $\mathbf{R}_2$  中的项目邻接信息迁移到  $\mathbf{R}$  中.

我们在评分矩阵的集合分解中,对潜在因子  $\mathbf{U}_0$  和  $\mathbf{V}_0$  的非正交约束保证了它们每行都只有一个正元素,且该行其它元素都为零.  $\mathbf{U}_0$  和  $\mathbf{V}_0$  中每行的正元素已分别包含了用户和项目的聚类信息,通过余弦相似度计算,我们即可提取这些聚类信息,得到基于聚类的用户相似度和项目相似度.我们用邻接矩阵表示邻接图,假设用户邻接图和项目邻接图的邻接矩阵分别为  $\mathbf{W}_U$  和  $\mathbf{W}_V$ ,它们的元素定义分别如下:

$$(\mathbf{W}_U)_{ij} = \cos(u_{i_*}^0, u_{j_*}^0) = \frac{u_{i_*}^0 u_{j_*}^{0T}}{\|u_{i_*}^0\|_2 \cdot \|u_{j_*}^0\|_2} \quad (15)$$

$$(\mathbf{W}_V)_{ij} = \cos(v_{i_*}^0, v_{j_*}^0) = \frac{v_{i_*}^0 v_{j_*}^{0T}}{\|v_{i_*}^0\|_2 \cdot \|v_{j_*}^0\|_2} \quad (16)$$

这里  $u_{i_*}^0$  和  $v_{i_*}^0$  分别表示  $\mathbf{U}_0$  和  $\mathbf{V}_0$  的第  $i$  行,  $u_{j_*}^0$  和  $v_{j_*}^0$  分别表示  $\mathbf{U}_0$  和  $\mathbf{V}_0$  的第  $j$  行,  $\|\cdot\|_2$  表示向量的 L2 范数.在构造的邻接图中,如果用户或项目在同一个聚类中,则邻接权重为 1,否则邻接权重为 0.

## 5.3 基于三元桥迁移学习的评分预测

共享模式  $\mathbf{S}_c^U$  和  $\mathbf{S}_c^V$ ,潜在因子  $\mathbf{U}_0$  和  $\mathbf{V}_0$ ,以及邻接图  $\mathbf{W}_U$  和  $\mathbf{W}_V$  可合成一座信息迁移的三元桥.其中,基于  $\mathbf{S}_c^U$  可迁移用户的兴趣模式信息,基于  $\mathbf{S}_c^V$  可迁移项目的受欢迎模式信息;基于  $\mathbf{U}_0$  和  $\mathbf{V}_0$  分别可迁移用户和项目的特征信息;基于  $\mathbf{W}_U$  和  $\mathbf{W}_V$  分别可迁移用户邻接信息和项目邻接信息.我们基于三元桥迁移学习,将  $\mathbf{R}_1$  和  $\mathbf{R}_2$  中的有用信息迁移到  $\mathbf{R}$  中.

基于  $\mathbf{S}_c^U$  和  $\mathbf{S}_c^V$  迁移的信息是用户与项目之间聚类层次的模式信息,比较稳定.与特征信息和邻接信息相比,它对用户或项目的重叠度较不敏感.考虑到  $\mathbf{S}_c^U$  和  $\mathbf{S}_c^V$  在矩阵三元分解中的表示方式不同,我们分别进行用户端和项目端的三元桥迁移学习,并联合预测  $\mathbf{R}$  中的缺失评分.

### 5.3.1 用户端的三元桥迁移学习

$\mathbf{R}_1$  与  $\mathbf{R}$  含有相似的用户和领域关联的项目,它们各自特有一部分用户评分模式,也共享一部分用户评分模式.我们基于已获得的共享用户评分模式  $\mathbf{S}_c^U$ ,潜在因子  $\mathbf{U}_0$  和  $\mathbf{V}_0$ ,邻接图  $\mathbf{W}_U$  和  $\mathbf{W}_V$  进行三元桥迁移学习,目标最优化问题如下:

$$\min_{\mathbf{U}, \mathbf{V}, \mathbf{B}, \mathbf{B}' \geq 0} O_1 = \|\mathbf{Z} \odot (\mathbf{R} - \mathbf{U}[\mathbf{B}, \mathbf{B}']\mathbf{V}^T)\|_F^2 + \lambda_U \|\mathbf{U} - \mathbf{U}_0\|_F^2 + \lambda_V \|\mathbf{V} - \mathbf{V}_0\|_F^2 + \beta_U \|\mathbf{B} - \mathbf{S}_c^U\|_F^2 + \gamma_U G_U + \gamma_V G_V \quad (17)$$

这里  $G_U = \text{tr}(\mathbf{U}^T \mathbf{L}_U \mathbf{U})$ ,  $G_V = \text{tr}(\mathbf{V}^T \mathbf{L}_V \mathbf{V})$  ( $\text{tr}$  表示矩阵的迹),  $\mathbf{L}_U$  和  $\mathbf{L}_V$  分别表示  $\mathbf{W}_U$  和  $\mathbf{W}_V$  的拉普拉斯矩阵,  $\mathbf{L}_U = \mathbf{D}_U - \mathbf{W}_U$ ,  $\mathbf{L}_V = \mathbf{D}_V - \mathbf{W}_V$ ,其中  $\mathbf{D}_U$  和  $\mathbf{D}_V$  为对角矩阵,分别基于  $\mathbf{W}_U$  和  $\mathbf{W}_V$  的行元素之和生成.  $\mathbf{B}$  表示  $\mathbf{R}$  和  $\mathbf{R}_1$  共享的用户评分模式,  $\mathbf{B}'$  表示  $\mathbf{R}$  相对  $\mathbf{R}_1$  特有的用户评分模式.  $\mathbf{U}$  和  $\mathbf{V}$  分别表示  $\mathbf{R}$  的用户潜在因子和项目潜在因子.  $\gamma_U, \gamma_V, \lambda_U, \lambda_V$  和  $\beta_U$  为正则化参数,用来描述置信度以及防止过拟合.

下面使用交替梯度下降法求解方程式(17)描述的目标最优化问题.

假设  $\mathbf{S}_c^U \in \mathbf{R}^{K_c \times D_c}$ ,  $\mathbf{U}_0 \in \mathbf{R}^{M \times K_c}$ ,  $\mathbf{V}_0 \in \mathbf{R}^{L_c \times N}$ , 则  $\mathbf{U} \in \mathbf{R}^{M \times K_c}$ ,  $\mathbf{V} \in \mathbf{R}^{L_c \times N}$ ,  $\mathbf{B} \in \mathbf{R}^{K_c \times D_c}$ ,  $\mathbf{B}' \in \mathbf{R}^{K_c \times (L_c - D_c)}$ .

以学习  $\mathbf{U}$  为例,  $O_1$  对  $\mathbf{U}$  的偏导数为

$$\begin{aligned} \frac{\partial O_1}{\partial \mathbf{U}} = & -2\mathbf{Z} \odot \mathbf{R} \mathbf{V} [\mathbf{B}, \mathbf{B}']^T + \\ & 2\mathbf{Z} \odot (\mathbf{U} [\mathbf{B}, \mathbf{B}'] \mathbf{V}^T) \mathbf{V} [\mathbf{B}, \mathbf{B}']^T + \\ & 2\lambda_U (\mathbf{U} - \mathbf{U}_0) + 2\gamma_U \mathbf{L}_U \mathbf{U}. \end{aligned}$$

使用保证  $\mathbf{U}$  非负的 KKT 互补条件, 得到  $[-\mathbf{Z} \odot \mathbf{R} \mathbf{V} [\mathbf{B}, \mathbf{B}']^T - \lambda_U \mathbf{U}_0 + \mathbf{Z} \odot (\mathbf{U} [\mathbf{B}, \mathbf{B}'] \mathbf{V}^T) \cdot \mathbf{V} [\mathbf{B}, \mathbf{B}']^T + \lambda_U \mathbf{U} + \gamma_U \mathbf{L}_U \mathbf{U}] \odot \mathbf{U} = 0$ .

因  $\mathbf{L}_U$  的元素值可能有正负两种情形, 我们令  $\mathbf{L}_U = \mathbf{L}_U^+ - \mathbf{L}_U^-$ , 其中  $\mathbf{L}_U^+ = \frac{1}{2}(|\mathbf{L}_U| + \mathbf{L}_U)$ ,  $\mathbf{L}_U^- = \frac{1}{2}(|\mathbf{L}_U| - \mathbf{L}_U)$ ,  $\mathbf{L}_U^+$  和  $\mathbf{L}_U^-$  都为正值矩阵.

从而得到  $\mathbf{U}$  的更新规则:

$$\mathbf{U} \leftarrow \mathbf{U} \odot \sqrt{\frac{\mathbf{Z} \odot \mathbf{R} \mathbf{V} [\mathbf{B}, \mathbf{B}']^T + \lambda_U \mathbf{U}_0 + \gamma_U \mathbf{L}_U^- \mathbf{U}}{\mathbf{Z} \odot (\mathbf{U} [\mathbf{B}, \mathbf{B}'] \mathbf{V}^T) \mathbf{V} [\mathbf{B}, \mathbf{B}']^T + \lambda_U \mathbf{U} + \gamma_U \mathbf{L}_U^+ \mathbf{U}}} \quad (18)$$

类似地, 我们可得到  $\mathbf{V}$ ,  $\mathbf{B}$  和  $\mathbf{B}'$  的更新规则:

$$\mathbf{V} \leftarrow \mathbf{V} \odot \sqrt{\frac{(\mathbf{Z} \odot \mathbf{R})^T \mathbf{U} [\mathbf{B}, \mathbf{B}'] + \lambda_V \mathbf{V}_0 + \gamma_V \mathbf{L}_V \mathbf{V}}{(\mathbf{Z} \odot (\mathbf{U} [\mathbf{B}, \mathbf{B}'] \mathbf{V}^T))^T \mathbf{U} [\mathbf{B}, \mathbf{B}'] + \lambda_V \mathbf{V} + \gamma_V \mathbf{L}_V^+ \mathbf{V}}} \quad (19)$$

$$\mathbf{B} \leftarrow \mathbf{B} \odot \sqrt{\frac{\mathbf{U}^T (\mathbf{Z} \odot \mathbf{R}) \mathbf{V}_c^T + \beta_U \mathbf{S}_c^U}{\mathbf{U}^T (\mathbf{Z} \odot [\mathbf{U} \mathbf{B}' \mathbf{V}_d]) \mathbf{V}_c^T + \mathbf{U}^T (\mathbf{Z} \odot [\mathbf{U} \mathbf{B} \mathbf{V}_c]) \mathbf{V}_c^T + \beta_U \mathbf{B}}} \quad (20)$$

$$\mathbf{B}' \leftarrow \mathbf{B}' \odot$$

$$\sqrt{\frac{\mathbf{U}^T (\mathbf{Z} \odot \mathbf{R}) \mathbf{V}_d^T}{\mathbf{U}^T (\mathbf{Z} \odot [\mathbf{U} \mathbf{B} \mathbf{V}_c]) \mathbf{V}_d^T + \mathbf{U}^T (\mathbf{Z} \odot [\mathbf{U} \mathbf{B}' \mathbf{V}_d]) \mathbf{V}_d^T}} \quad (21)$$

请注意, 在推导  $\mathbf{B}$  和  $\mathbf{B}'$  的更新规则的过程中, 我们令  $\mathbf{V} = [\mathbf{V}_c^T, \mathbf{V}_d^T]$ , 其中  $\mathbf{V}_c^T = \mathbf{V}(:, 1:D_c)$ ,  $\mathbf{V}_d^T = \mathbf{V}(:, (D_c+1):L_c)$ .

参考文献[17]可证明, 基于更新规则 18~21 依次迭代更新  $\mathbf{U}$ ,  $\mathbf{V}$ ,  $\mathbf{B}$  和  $\mathbf{B}'$ , 方程式(17)中的目标函数  $O_1$  将单调下降, 最终收敛. 当  $O_1$  收敛时, 我们将  $\mathbf{U}$ ,  $\mathbf{V}$ ,  $\mathbf{B}$  和  $\mathbf{B}'$  的值分别记为  $U_m$ ,  $V_m$ ,  $B_m$  和  $B'_m$ , 令  $\mathbf{R}_U^p = U_m [B_m, B'_m] \mathbf{V}_m^T$ , 为用户端的预测评分矩阵.

### 5.3.2 项目端的三元桥迁移学习

$\mathbf{R}_2$  和  $\mathbf{R}$  含有相似的项目和领域关联的用户, 它们各自特有一部分项目被评分模式, 也共享一部分项目被评分模式. 我们基于已获得的共享项目被评分模式  $\mathbf{S}_c^V$ , 潜在因子  $\mathbf{U}_0$  和  $\mathbf{V}_0$ . 邻接图  $\mathbf{W}_U$  和  $\mathbf{W}_V$  进行三元桥迁移学习, 目标最优化问题如下:

$$\begin{aligned} \min_{\mathbf{U}, \mathbf{V}, \mathbf{I}, \mathbf{I}' \geq 0} O_2 = & \|\mathbf{Z} \odot (\mathbf{R} - \mathbf{U} [\mathbf{I}', \mathbf{I}] \mathbf{V}^T)\|_F^2 + \\ & \lambda'_U \|\mathbf{U} - \mathbf{U}_0\|_F + \lambda'_V \|\mathbf{V} - \mathbf{V}_0\|_F + \\ & \beta_V \|\mathbf{I} - \mathbf{S}_c^V\|_F^2 + \gamma'_U G_U + \gamma'_V G_V \quad (22) \end{aligned}$$

这里矩阵  $\mathbf{I}$  表示  $\mathbf{R}$  和  $\mathbf{R}_2$  共享的项目被评分模式, 矩阵  $\mathbf{I}'$  表示  $\mathbf{R}$  相对  $\mathbf{R}_2$  特有的项目被评分模式.  $G_U$ ,  $G_V$ ,  $\mathbf{U}$  和  $\mathbf{V}$  的含义同方程式(17);  $\gamma'_U$ ,  $\gamma'_V$ ,  $\lambda'_U$ ,  $\lambda'_V$  和  $\beta_V$  为正则化参数, 用来描述置信度以及防止过拟合.

与方程式(17)的求解过程类似, 我们求解方程式(22), 得到目标函数  $O_2$  收敛时矩阵变量  $\mathbf{U}$ ,  $\mathbf{V}$ ,  $\mathbf{I}$  和  $\mathbf{I}'$  的值, 分别记为  $U_n$ ,  $V_n$ ,  $I_n$  和  $I'_n$ , 令  $\mathbf{R}_V^p = U_n [I'_n, I_n] \mathbf{V}_n^T$ , 为项目端的预测评分矩阵.

### 5.3.3 缺失评分预测

令  $\mathbf{R}^p$  表示最终得到的  $\mathbf{R}$  的预测评分矩阵, 则

$$\mathbf{R}^p = \omega \mathbf{R}_U^p + (1-\omega) \mathbf{R}_V^p, \quad \omega \in [0, 1] \quad (23)$$

这里  $\omega$  为比例系数, 可依据具体的领域关联情况, 通过调节迁移学习的比例  $\omega$  对  $\mathbf{R}^p$  进行调优.

## 6 模型分析

为了充分挖掘和利用模式信息, TRBT 模型分别提取领域共享的用户评分模式和项目被评分模式, 其中用户评分模式侧重于用户的兴趣模式, 项目被评分模式侧重于项目的受欢迎模式; 然后基于这两种侧重不同的共享模式进行迁移学习. TRBT 模型在提取共享模式的过程中还考虑了领域特有模式, 有效减少了负迁移. 而且, TRBT 模型基于三元桥, 不仅能迁移多种类型的有用信息, 还可根据具体情况灵活选择迁移方式以及调节迁移权重.

TRBT 模型的核心求解算法能通过有限的迭代快速收敛, 不妨假设文中 5.1 节和 5.3 节求解各变量的迭代更新次数统一为  $t$ . 在 TRBT 模型中, 时间成本主要分为 3 个步骤, 具体如下:

(1) 潜在因子提取的时间复杂度为

$$O(((M+M_1)K_c + K_c L_c + (N+N_1)L_c)t);$$

(2) 邻接图构造的时间复杂度为  $O(K_c M^2 + L_c N^2)$ ;

(3) 通过三元桥迁移学习进行评分预测的时间复杂度为  $O((K_c M + K_c L_c + L_c N)t)$ .

综上可得, TRBT 模型总的时间复杂度为  $O(((M+M_1)K_c + K_c L_c + (N+N_1)L_c)t + K_c M^2 + L_c N^2)$ .  $K_c$  和  $L_c$  分别为用户潜在因子矩阵和项目潜在因子矩阵的列数, TRBT 模型通过矩阵分解对目标评分矩阵  $\mathbf{R}$  进行降维, 使得  $K_c$  和  $L_c$  分别相比  $M$  和  $N$  要大幅减小, 即有  $K_c \ll M$ ,  $L_c \ll N$ . 邻接图构造步骤中需要计算余弦相似度的向量是维数为  $K_c$  或  $L_c$ , 且存在许多零元素的向量. 另外, 潜在因子提取和三元桥迁移学习两个步骤中各变量的迭代更新可

以并行处理,因此理论上 TRBT 模型可扩展到大规模数据的跨领域推荐。

下面我们对 TRBT 模型与一些相关的先进模型进行理论对比分析。

### 6.1 TRBT 模型相对 GMTF 模型和 CBT<sup>[16]</sup> 模型

GMTF 模型基于单一领域进行协同过滤推荐,当目标数据非常稀疏时它的推荐性能将严重下降;CBT 模型和 TRBT 模型通过跨领域推荐解决目标数据的稀疏性问题, CBT 模型基于评分模式进行迁移学习,但假设不同领域完全共享一个评分模式,忽略了领域间的评分模式差异,可能导致严重的负迁移;并且忽略了评分模式对用户和项目的不同侧重,没有充分挖掘和利用评分模式信息, TRBT 模型在考虑领域特有模式的同时,分别提取共享的用户评分模式和项目被评分模式,并基于三元桥进行迁移学习,相对 CBT 模型,能从其它领域迁移更多有用信息。

### 6.2 TRBT 模型相对 TWBT 模型

我们主要从潜在因子提取,邻接图构造以及评分预测中的信息迁移 3 个方面来对 TRBT 模型和 TWBT 模型进行对比分析,具体分析如下:

(1) TWBT 模型直接对辅助评分矩阵进行三元分解并提取潜在因子;TRBT 模型对目标评分矩阵和辅助评分矩阵进行集合分解并提取潜在因子和共享模式,不仅通过考虑领域特有模式减少负迁移,还对潜在因子进行相似性约束,相对 TWBT 模型,TRBT 模型从辅助评分矩阵中提取的潜在因子与目标评分矩阵的潜在因子更相似且更准确,因而更适合用来构造邻接图以及作为跨领域信息迁移的桥梁。

(2) TWBT 模型基于  $P$  最近邻方法将用户或项目之间的相似度指定为 0 或 1,并构造相似度图,对于任意两个用户, TWBT 模型认为:如果其中某个用户在另外一个用户的  $P$  最近邻集合中,那么这两个用户相似,相似度记为 1;否则这两个用户不相似,相似度记为 0。同理, TWBT 模型定义项目之间的相似度,但  $P$  的大小难以确定,  $P$  过大会导致用户或项目之间的相似关系不准确,  $P$  过小会导致用户或项目之间的相似关系被遗漏,而且 TRBT 模型对所有的用户或项目用一个统一的  $P$  值来界定相似度,没有充分考虑到用户或项目之间距离分布的不均衡。另外计算用户或项目的  $P$  最近邻集合的过程比较费时, TWBT 模型在构造邻接图上的时间复杂度为  $O(K_c M^3 + L_c N^3)$ 。而 TRBT 模型在集合分

解的过程中已经对用户或项目进行了聚类,同一个聚类中的用户或项目可被视为是相似的,聚类过程是无监督的, TRBT 模型无需通过设置  $P$  值并基于  $P$  最近邻来界定相似关系,其中相似的用户或项目自动被聚成不同大小的类组,而且用户或项目的聚类信息已包含在提取的潜在因子中,基于余弦相似度计算便可提取这些聚类信息并构造邻接图, TRBT 模型在构造邻接图上的时间复杂度为  $O(K_c M^2 + L_c N^2)$ ,优于 TWBT 模型,因此相比 TWBT 模型, TRBT 模型更方便且更有效地构造邻接图。

(3) TRBT 模型比 TWBT 模型多了一种信息迁移桥梁:共享的用户评分模式或项目被评分模式,特别是创新使用了共享的项目被评分模式进行跨领域的信息迁移,因而能迁移更多有用信息,而且,用户评分模式和项目被评分模式是聚类层次的信息,比较稳定,当属于不同领域的用户或项目之间的重叠度较低时, TRBT 模型可通过提高共享的用户评分模式或项目被评分模式的迁移权重进行有效的跨领域推荐,不必同 TWBT 模型一样要求不同领域的用户或项目完全相同。

综合上述理论分析可得,相对 TWBT 模型, TRBT 模型能从辅助评分矩阵提取到与目标评分矩阵更相似且更准确的潜在因子;能更方便且更有效地构造邻接图;新增了共享模式,特别是共享的项目被评分模式作为跨领域信息迁移的桥梁,不仅能迁移更多有用信息,而且能有效适用于用户或项目不完全重叠的情况。

## 7 实 验

### 7.1 数据集

我们基于下面 3 个公用的真实数据集进行实验评估:

(1) MovieLens10M<sup>①</sup>. 包括 10 000 054 个评分,由 71 567 个用户对 10 681 部电影进行评分,评分以半个星级为增量,最高为 5 星级。

(2) Amazon<sup>②</sup>. 包含 548 552 个不同类别产品的元数据和评价信息,其中包括 393 561 本图书和 19 828 部电影,评分为从 1 到 5 范围内的整数。

(3) Book-Crossing<sup>③</sup>. 包含 1 149 780 个评分,由 278 858 个用户对 271 379 本图书进行评分,评分为

① [www.grouplens.org/node/73/](http://www.grouplens.org/node/73/)

② <http://snap.stanford.edu/data/amazon-meta.html>

③ <http://www2.informatik.uni-freiburg.de/chiegler/BX/>



从 1 到 10 范围内的整数。

## 7.2 对比模型

在模型分析部分,我们已经从理论上对 TRBT 模型与 CBT,GMTF,TWBT 等模型进行了对比分析,阐述了 TRBT 的优势,下面我们继续对 TRBT 模型与 CBT,GMTF,TWBT 等模型进行实验对比。

(1) CBT 模型<sup>[16]</sup>. 一种先进的跨领域协同过滤模型,假设领域间完全共享一个评分模式,并利用聚类层次的密码本描述共享的评分模式,然后基于密码本进行跨领域的信息迁移。

(2) GMTF 模型<sup>[17]</sup>. 一种基于单一领域的图正则化协同过滤模型,在文中 4.1 节已介绍。

(3) TWBT 模型<sup>[18]</sup>. 一种基于潜在因子和相似度图进行迁移学习的跨领域协同过滤模型,在文中 4.2 节已介绍。

(4) TRBT 模型. 我们提出的新模型。

## 7.3 评价指标

平均绝对误差 MAE (Mean Absolute Error) 和均方根误差 RMSE (Root Mean Square Error) 是两种常用的评估协同过滤推荐性能的指标,其中 RMSE 对大误差比较敏感,而 MAE 对小误差的积累比较敏感。CBT, GMTF, TWBT 等模型均采用 MAE 指标或 RMSE 指标来评估评分预测的精度,因此我们在实验中采用它们来评估各推荐模型的推荐精度,MAE 和 RMSE 的定义分别为

$$MAE = \frac{\sum_{i \in T_E} |r_i - r_i^*|}{|T_E|} \quad (24)$$

$$RMSE = \sqrt{\frac{\sum_{i \in T_E} |r_i - r_i^*|^2}{|T_E|}} \quad (25)$$

这里  $|T_E|$  表示测试评分的数量,  $r_i$  表示真实评分,  $r_i^*$  表示通过协同过滤推荐得到的  $r_i$  的预测值。更小的 MAE 或 RMSE 代表更高的推荐精度。

## 7.4 参数设置及实验结果

假设  $C$  为任意评分矩阵,令  $l(C)$  表示  $C$  的密集度,即  $C$  中已观察评分在  $C$  中所有评分的占比,  $l(C)$  越小代表  $C$  的数据越稀疏。令  $C(i_1:i_2, j_1:j_2)$  表示  $C$  的  $i_1$  到  $i_2$  行和  $j_1$  到  $j_2$  列组成的子矩阵,  $C(i_1:i_2, :)$  表示  $C$  的  $i_1$  到  $i_2$  行组成的子矩阵,  $C(:, j_1:j_2)$  表示  $C$  的  $j_1$  到  $j_2$  列组成的子矩阵。

我们在 Amazon 数据集中选择 2000 个用户对 2000 部电影和 2000 本图书的评分数据,每个用户对电影和图书都至少有 25 个评分,其评分矩阵记为  $R^A$ ,其中  $R^A(:, 1:2000)$  描述用户对电影的评分,密

集度为 2.03%;  $R^A(:, 2001:4000)$  描述用户对图书的评分,密集度为 1.75%。在 MovieLens10M 数据集中选择 2000 个用户对 2000 部电影的评分数据,其评分矩阵记为  $R^M$ ,  $l(R^M) = 2.92%$ ,  $R^M$  与  $R^A$  的电影重叠度为 13.6%。在 Book-Crossing 数据集中选择 2000 个用户对 2000 本图书的评分数据,其评分矩阵记为  $R^B$ ,  $l(R^B) = 3.17%$ ,  $R^B$  与  $R^A$  的图书重叠度为 31.2%。为了有效迁移共享模式信息,我们对  $R^A$ ,  $R^B$  和  $R^M$  中的评分统一进行归一化,使评分尺度保持一致。

我们基于  $R^A$  构造目标评分矩阵  $R$ ,并随机选取  $R$  中 80% 的评分用于训练,记为  $R_T^0$ ,剩下 20% 的评分用于测试,记为  $R_E^0$ 。在训练矩阵  $R_T^0$  中随机选取不同密集度的数据用于训练,记为  $R_T$ ,其中每个训练用户的已观察评分数分别为 5, 10, 15 和 20,对应的密集度  $l(R_T)$  分别为 0.25%, 0.5%, 0.75%, 1%。对于不同  $l(R_T)$  下的训练,我们每次随机选择已观察评分,训练 10 次并呈现平均结果,其中各变量的迭代更新次数  $t$  统一取 1000。

鉴于邻接图是基于潜在因子构造的,为了方便实验对比,我们将它们的正则化参数设为一致,并基于用户端和项目端,分别设置各个参数,令  $\lambda_u = \gamma_u = \lambda'_u = \gamma'_u = \alpha_u$ ,  $\lambda_v = \gamma_v = \lambda'_v = \gamma'_v = \alpha_v$ 。假设  $\theta_1$  表示  $R$  和  $R_1$  的用户重叠度,  $\theta_2$  表示  $R$  和  $R_2$  的项目重叠度。当  $\theta_1$  和  $\theta_2$  不同时为 0 时,令  $\omega = \frac{\theta_1}{\theta_1 + \theta_2}$ ; 当  $\theta_1$  和  $\theta_2$  同时为 0 时,令  $\omega = \frac{1}{2}$ 。当重叠度较低时,我们提高共享模式的迁移权重,令  $\beta_u = \frac{\lambda_u}{\theta_1 + 0.01}$ ,  $\beta_v = \frac{\lambda_v}{\theta_2 + 0.01}$ 。各模型的正则化参数的变化最后归结为

对  $\lambda_u$  和  $\lambda_v$  的设置,我们基于集合  $\{0.01, 0.1, 1, 10, 100\}$  生成的网格搜索  $\lambda_u$  和  $\lambda_v$  的值,最后选取最好的结果进行比较。在各模型中,  $K_c$  和  $L_c$  分别为分解  $R$  后得到的用户聚类 and 项目聚类的大小。在一定范围内,  $K_c$  或  $L_c$  越大, MAE 和 RMSE 越小,但在我们的实证研究中,当它们增加到一定大小,各模型的 MAE 以及 RMSE 不再有明显变化,而计算成本会提高。我们在保证各模型得到临界最优值的前提下,权衡计算成本,统一设置:  $K_c = 50$ ,  $L_c = 85$ 。另外在 TRBT 模型中,共享用户评分模式  $S_u^v$  的项目聚类数  $D_c \in (0, 85]$ , 共享项目被评分模式  $S_v^u$  的用户聚类数  $H_c \in (0, 50]$ 。下面我们评估各模型在不同的用户重叠度  $\theta_1$ , 项目重叠度  $\theta_2$  和训练矩阵密集度  $l(R_T)$  下

的推荐精度。

我们首先基于  $R^A$  中的电影评分数据构造目标评分矩阵  $R$ , 进行跨领域的电影推荐。

令  $R = R^A(1:1000, 1:2000)$ ,  $R_2 = R^A(1001:2000, 1:2000)$ ,  $R_1 = R^A(1:1000, 2001:4000)$ 。

此时  $\theta_1 = \theta_2 = 1$ 。在 TRBT 模型中, 取  $D_c = 49$ ,  $H_c = 31$ 。不同模型在不同  $l(R_T)$  下的 MAE 如表 1 所示, RMSE 如表 2 所示。

表 1 电影推荐的 MAE( $\theta_1 = \theta_2 = 1$ )

$l(R_T)/\%$	CBT	GMTF	TWBT	TRBT
0.25	0.9024	1.0296	0.8520	<b>0.7435</b>
0.5	0.8578	0.9543	0.8016	<b>0.7051</b>
0.75	0.8141	0.9095	0.7625	<b>0.6813</b>
1	0.7917	0.8774	0.7386	<b>0.6682</b>

表 2 电影推荐的 RMSE( $\theta_1 = \theta_2 = 1$ )

$l(R_T)/\%$	CBT	GMTF	TWBT	TRBT
0.25	1.1876	1.3352	1.1356	<b>1.0107</b>
0.5	1.0921	1.1564	1.0463	<b>0.9543</b>
0.75	1.0240	1.0652	0.9735	<b>0.8967</b>
1	0.9469	0.9874	0.9153	<b>0.8454</b>

令  $R = R^A(1:1600, 1:2000)$ ,  $R_2 = R^M(1:1600, :)$ ,  $R_1 = R^A(401:2000, 2001:4000)$ , 此时  $\theta_1 = 0.75$ ,  $\theta_2 = 0.136$ , 取  $D_c = 43$ ,  $H_c = 20$ , 不同模型在不同  $l(R_T)$  下的 MAE 如表 3 所示, RMSE 如表 4 所示。

表 3 电影推荐的 MAE( $\theta_1 = 0.75, \theta_2 = 0.136$ )

$l(R_T)/\%$	CBT	GMTF	TWBT	TRBT
0.25	0.9786	1.0137	0.9658	<b>0.8143</b>
0.50	0.9115	0.9468	0.9236	<b>0.7829</b>
0.75	0.8753	0.9024	0.8815	<b>0.7514</b>
1	0.8569	0.8832	0.8687	<b>0.7296</b>

表 4 电影推荐的 RMSE( $\theta_1 = 0.75, \theta_2 = 0.136$ )

$l(R_T)/\%$	CBT	GMTF	TWBT	TRBT
0.25	1.2032	1.3241	1.1835	<b>1.0621</b>
0.5	1.1146	1.1602	1.0987	<b>1.0137</b>
0.75	1.0298	1.0539	1.0493	<b>0.9536</b>
1	0.9635	0.9806	0.9781	<b>0.8942</b>

令  $R = R^A(1:1250, 1:2000)$ ,  $R_2 = R^M(1:1250, :)$ ,  $R_1 = R^A(751:2000, 2001:4000)$ , 此时  $\theta_1 = 0.4$ ,  $\theta_2 = 0.136$ , 取  $D_c = 34$ ,  $H_c = 18$ , 不同模型在不同  $l(R_T)$  下的 MAE 如表 5 所示, RMSE 如表 6 所示。

表 5 电影推荐的 MAE( $\theta_1 = 0.4, \theta_2 = 0.136$ )

$l(R_T)/\%$	CBT	GMTF	TWBT	TRBT
0.25	0.9930	1.0273	1.0157	<b>0.8319</b>
0.5	0.9348	0.9552	0.9691	<b>0.7916</b>
0.75	0.9006	0.9044	0.9315	<b>0.7713</b>
1	0.8921	0.8836	0.9172	<b>0.7358</b>

表 6 电影推荐的 RMSE( $\theta_1 = 0.4, \theta_2 = 0.136$ )

$l(R_T)/\%$	CBT	GMTF	TWBT	TRBT
0.25	1.2673	1.3296	1.2608	<b>1.1054</b>
0.5	1.1209	1.1537	1.1325	<b>1.0316</b>
0.75	1.0455	1.0604	1.0728	<b>0.9846</b>
1	0.9763	0.9731	1.0141	<b>0.9379</b>

我们再基于  $R^A$  中的图书评分数据构造目标评分矩阵  $R$ , 进行跨领域的图书推荐。

令  $R = R^A(1:1000, 2001:4000)$ ,  $R_2 = R^A(1001:2000, 2001:4000)$ ,  $R_1 = R^A(1:1000, 1:2000)$ , 此时  $\theta_1 = \theta_2 = 1$ , 取  $D_c = 51$ ,  $H_c = 34$ , 不同模型在不同  $l(R_T)$  下的 MAE 如表 7 所示, RMSE 如表 8 所示。

表 7 图书推荐的 MAE( $\theta_1 = \theta_2 = 1$ )

$l(R_T)/\%$	CBT	GMTF	TWBT	TRBT
0.25	0.8743	1.0472	0.8205	<b>0.7026</b>
0.5	0.8201	0.9638	0.7813	<b>0.6764</b>
0.75	0.7997	0.9224	0.7582	<b>0.6621</b>
1	0.7531	0.8873	0.7204	<b>0.6359</b>

表 8 图书推荐的 RMSE( $\theta_1 = \theta_2 = 1$ )

$l(R_T)/\%$	CBT	GMTF	TWBT	TRBT
0.25	1.1355	1.2879	1.0622	<b>0.9418</b>
0.5	1.0274	1.1187	0.9654	<b>0.8762</b>
0.75	0.9843	1.0276	0.9256	<b>0.8512</b>
1	0.8916	0.9557	0.8743	<b>0.8129</b>

令  $R = R^A(1:1200, 2001:4000)$ ,  $R_2 = R^B(1:1200, :)$ ,  $R_1 = R^A(801:2000, 1:2000)$ , 此时  $\theta_1 = \frac{1}{3}$ ,  $\theta_2 = 0.312$ , 取  $D_c = 33$ ,  $H_c = 26$ , 不同模型在不同  $l(R_T)$  下的 MAE 如表 9 所示, RMSE 如表 10 所示。

表 9 图书推荐的 MAE( $\theta_1 = \frac{1}{3}, \theta_2 = 0.312$ )

$l(R_T)/\%$	CBT	GMTF	TWBT	TRBT
0.25	0.9671	1.0395	0.9862	<b>0.8204</b>
0.5	0.9006	0.9607	0.9315	<b>0.7812</b>
0.75	0.8842	0.9231	0.9006	<b>0.7627</b>
1	0.8635	0.8862	0.8891	<b>0.7283</b>

表 10 图书推荐的 RMSE( $\theta_1 = \frac{1}{3}, \theta_2 = 0.312$ )

$l(R_T)/\%$	CBT	GMTF	TWBT	TRBT
0.25	1.1902	1.2803	1.1867	<b>1.0385</b>
0.5	1.0571	1.1354	1.0942	<b>0.9633</b>
0.75	0.9788	1.0249	1.0064	<b>0.9012</b>
1	0.9237	0.9502	0.9441	<b>0.8565</b>

令  $R = R^A(1:1000, 2001:2500)$ ,  $R_1 = R^A(1001:2000, 1:500)$ ,  $R_2$  取自  $R^B$ , 这里  $R_2$  与  $R$  维数相同, 但与  $R$  中的用户以及项目均不重叠。此时  $\theta_1 = \theta_2 = 0$ , 取  $D_c = 15$ ,  $H_c = 10$ , 不同模型在不同  $l(R_T)$  下的 MAE 如表 11 所示, RMSE 如表 12 所示。

表 11 图书推荐的  $MAE(\theta_1 = \theta_2 = 0)$ 

$l(R_T)/\%$	CBT	GMTF	TWBT	TRBT
0.25	0.9734	1.0641	1.0453	<b>0.9375</b>
0.5	0.9387	0.9925	0.9976	<b>0.9089</b>
0.75	0.9189	0.9618	0.9764	<b>0.8906</b>
1	0.9023	0.9352	0.9531	<b>0.8737</b>

表 12 图书推荐的  $RMSE(\theta_1 = \theta_2 = 0)$ 

$l(R_T)/\%$	CBT	GMTF	TWBT	TRBT
0.25	1.2495	1.2961	1.2756	<b>1.1953</b>
0.5	1.1239	1.1454	1.1656	<b>1.0882</b>
0.75	1.0489	1.0528	1.0945	<b>1.0246</b>
1	1.0193	1.0247	1.0631	<b>1.0034</b>

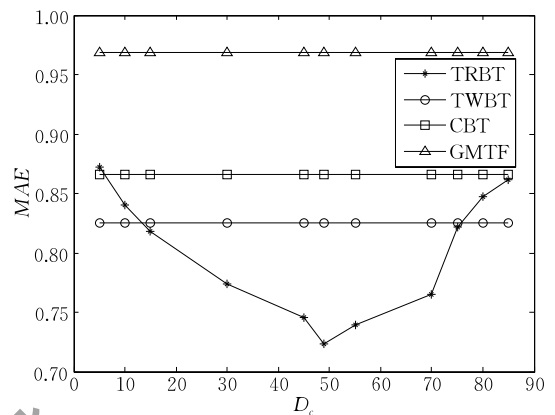
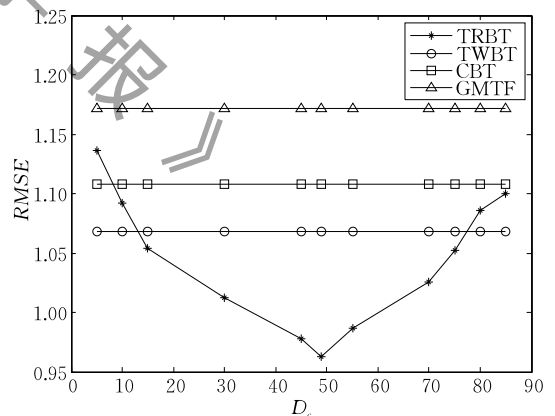
表 1~表 12 中的结果显示:不管是基于  $MAE$  指标,还是  $RMSE$  指标,与 CBT、GMTF、TWBT 等模型相比,TRBT 模型的推荐精度在各种情况下都明显较高;另外当数据越稀疏时,除了个别情况,TRBT 模型与其它模型的精度差距有扩大的趋势,且这些个别情况的偏差很小. 该结果表明,相对 CBT、GMTF、TWBT 等模型,TRBT 模型能更有效地解决协同过滤中目标数据的稀疏性问题,且在数据更稀疏的情况下其优势通常更为显著.

下面我们对表 1~表 12 中的实验结果进行更深入的分析,从中可得,基于单一领域的 GMTF 模型对目标数据的密集度最敏感,其  $MAE$  和  $RMSE$  几乎总是大于其它模型,该结果表明,利用辅助数据可改善目标数据的稀疏性问题. TWBT 模型对用户或项目的重叠度最敏感,当重叠度降低时,其  $MAE$  和  $RMSE$  的增大最显著,甚至出现大于 GMTF 模型的情况;TRBT 模型和 CBT 模型对重叠度较不敏感,不同于 TWBT 模型,它们利用了比较稳定的共享模式进行迁移学习. 当用户或项目完全重叠时,TRBT 模型的  $MAE$  和  $RMSE$  比 CBT 模型更小;当用户或项目的重叠度较低时,CBT 模型的  $MAE$  和  $RMSE$  比 TWBT 模型更小;TRBT 模型根据实际情况调节三元桥的具体迁移权重,在各种情况下其  $MAE$  和  $RMSE$  总是最小. 特别是当用户或项目完全非重叠时,如表 11~表 12 所示,TRBT 模型仍能得到明显优于其它模型的  $MAE$  和  $RMSE$  结果. 然而,此时只有在训练矩阵的密集度为 0.25% 的情况下,TRBT 模型的  $MAE$  和  $RMSE$  小于 GMTF 模型;在其它 3 种训练矩阵密集度情形下,TRBT 模型的  $MAE$  和  $RMSE$  均大于 GMTF 模型. 该结果表明,当用户或项目完全重叠时,共享的特征信息与邻接信息相对共享模式信息能产生更多的正迁移;但是当用户或项目的重叠度较低时,共享的特征信息与邻接信息难以准确提取,容易产生负迁移;依

据具体的重叠度调节不同类别信息的迁移权重,可得到更小的  $MAE$  和  $RMSE$ .

接下来我们进一步探讨  $D_c$  和  $H_c$  的变化对 TRBT 模型推荐精度的影响. 我们使用表 1 对应的实验数据集,其中  $\theta_1 = \theta_2 = 1$ ,先单独利用  $R_1$  预测  $R$  中的缺失评分,令  $\lambda_v = \omega = 1, \lambda_v' = 0$ ,探讨  $D_c$  的变化对推荐精度的影响;再单独利用  $R_2$  预测  $R$  中的缺失评分,令  $\lambda_v' = \omega = 0, \lambda_v = 1$ ,探讨  $H_c$  的变化对推荐精度的影响.  $l(R_T)$  均设置为 0.5,其它参数值可基于当前的参数设置得出.

随着  $D_c$  的变化,TRBT 模型的  $MAE$  和  $RMSE$  分别如图 2 和图 3 所示,我们同时呈现其它模型在相同参数设置下的  $MAE$  和  $RMSE$ .

图 2  $MAE$  随  $D_c$  的变化 ( $\theta_1 = \theta_2 = 1$ )图 3  $RMSE$  随  $D_c$  的变化 ( $\theta_1 = \theta_2 = 1$ )

随着  $H_c$  的变化,TRBT 模型的  $MAE$  和  $RMSE$  分别如图 4 和图 5 所示,我们同时呈现其它模型在相同参数设置下的  $MAE$  和  $RMSE$ .

从图 2~图 5 中可以看出,除了  $D_c$  和  $H_c$  过大或过小的情况,TRBT 模型的  $MAE$  和  $RMSE$  比其它模型更小,而且通过调节  $D_c$  或  $H_c$  的大小,可得到最小的  $MAE$  和  $RMSE$ . 若假设不同领域完全共享一个评分模式,即  $D_c$  和  $H_c$  均取最大值,将出现 TRBT 模型的  $MAE$  和  $RMSE$  大于 TWBT 模型的情况,

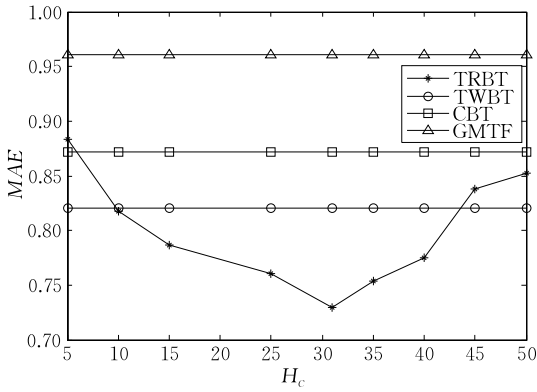


图4 MAE随 $H_c$ 的变化( $\theta_1 = \theta_2 = 1$ )

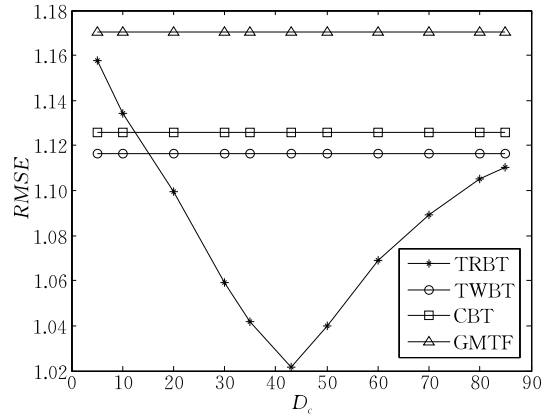


图7 RMSE随 $D_c$ 的变化( $\theta_1 = 0.75, \theta_2 = 0.136$ )

随着 $H_c$ 的变化,各模型的MAE和RMSE分别如图8和图9所示.

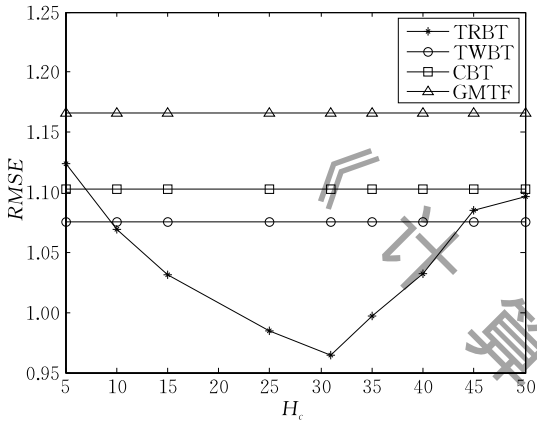


图5 RMSE随 $H_c$ 的变化( $\theta_1 = \theta_2 = 1$ )

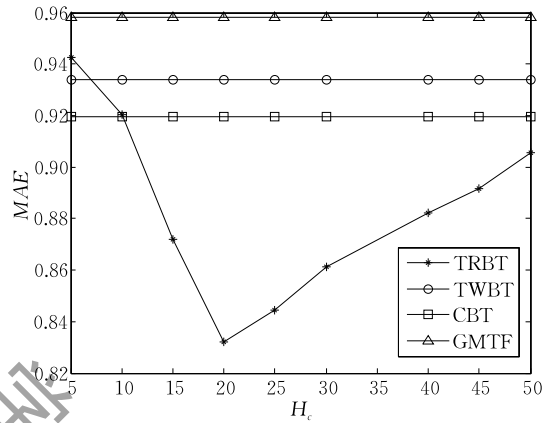


图8 MAE随 $H_c$ 的变化( $\theta_1 = 0.75, \theta_2 = 0.136$ )

导致推荐性能退化. 另外, $D_c$ 和 $H_c$ 不能过小,否则不仅无法产生足够的正迁移,而且使构造的三元桥不准确,从而产生负迁移. 此实验中, $R$ 与 $R_1$ 含有相同的用户, $R$ 与 $R_2$ 含有相同的项目,结果表明,有必要在考虑领域特有模式的基础上,选择合适的 $D_c$ 和 $H_c$ .

下面我们使用表3对应的实验数据集,其中 $\theta_1 = 0.75, \theta_2 = 0.136$ ,在参数设置与图2~图5对应的实验相同的情形下,探讨 $D_c$ 和 $H_c$ 的变化对TRBT模型推荐精度的影响.

随着 $D_c$ 的变化,各模型的MAE和RMSE分别如图6和图7所示.

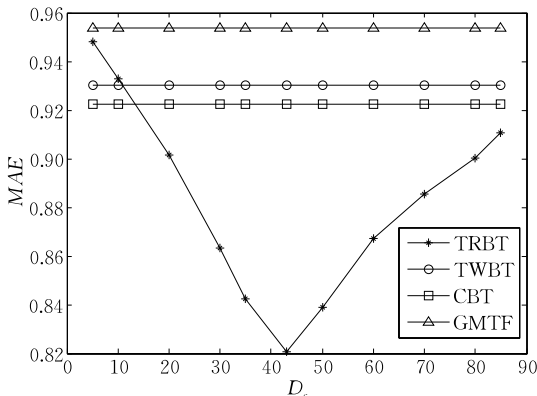


图6 MAE随 $D_c$ 的变化( $\theta_1 = 0.75, \theta_2 = 0.136$ )

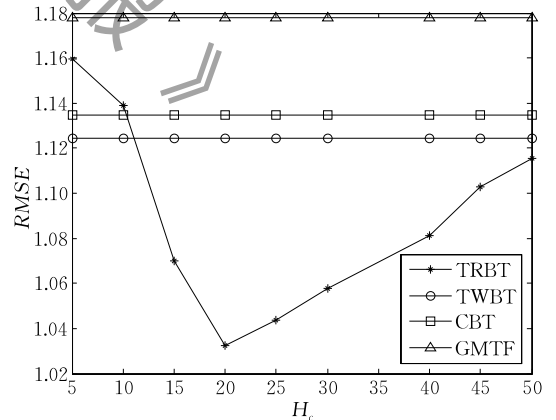


图9 RMSE随 $H_c$ 的变化( $\theta_1 = 0.75, \theta_2 = 0.136$ )

图6~图9的结果与图2~图5类似,不同的是,图6~图9中GMTF模型与其它模型之间的MAE差距以及RMSE差距相对图2~图5中更大. 所以不难理解,当 $R$ 与 $R_1$ 的用户,或 $R$ 与 $R_2$ 的项目不完全同时,更有必要考虑各领域特有的用户评分模式以及项目被评分模式.

综上,实验结果与文中的模型分析部分相吻合,

验证了 TRBT 模型的优越性.

## 8 结论与展望

我们提出一种新的用于跨领域推荐的迁移学习模型 TRBT, 可有效解决目标数据的稀疏性问题. 在考虑领域特有模式的前提下, TRBT 模型提取辅助领域中与目标领域共享的用户评分模式和项目被评分模式, 以及与目标领域相似的用户潜在因子和项目潜在因子; 并利用潜在因子构造邻接图; 然后基于共享模式、潜在因子和邻接图合成的三元桥迁移有用信息. 其中, 基于共享模式迁移用户与项目的互动信息; 基于潜在因子迁移用户及项目的特征信息; 基于邻接图迁移用户及项目的邻接信息. TRBT 模型在有效减少负迁移的同时显著增加了正迁移. 实验表明, TRBT 模型的推荐精度优于一些目前最先进的模型.

未来我们希望充分利用时空信息进行跨领域的移动推荐, 并基于 MapReduce 框架实现大数据环境下的跨领域推荐.

**致 谢** 审稿专家们在百忙之中对本文给出了非常宝贵的修改意见,《计算机学报》编辑部的老师们付出了辛勤劳动,在此表示衷心感谢!

### 参 考 文 献

- [1] Qi Q, Chen Z, Liu J, et al. Using inferred tag ratings to improve user based collaborative filtering//Proceedings of the 27th Annual ACM Symposium on Applied Computing, Trento, Italy, 2012: 2008-2013
- [2] Sarwar B, Karypis G, Konstan J, et al. Item-based collaborative filtering recommendation algorithms//Proceedings of the 10th International Conference on World Wide Web, Hong Kong, China, 2001: 285-295
- [3] Su X, Khoshgoftaar T M. A survey of collaborative filtering techniques. *Advances in Artificial Intelligence*, 2009, 2009(4): 1-19
- [4] Zhang Bo, Shi Zhong-Zhi, Zhao Xiao-Fei, et al. A transfer learning based on canonical correlation analysis across different domains. *Chinese Journal of Computers*, 2015, 38(7): 1326-1336(in Chinese)  
(张博, 史忠植, 赵晓非等. 一种基于跨领域典型相关性分析的迁移学习方法. *计算机学报*, 2015, 38(7): 1326-1336)
- [5] Li B. Cross-domain collaborative filtering; A brief survey//Proceedings of the 23rd IEEE International Conference on Tools with Artificial Intelligence, Boca Raton, USA, 2011: 1085-1086
- [6] Pan W. A survey of transfer learning for collaborative recommendation with auxiliary data. *Neurocomputing*, 2016, 177: 447-453
- [7] Gao S, Luo H, Chen D, et al. Cross-domain recommendation via cluster-level latent factor model//Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Prague, Czech Republic, 2013: 161-176
- [8] Cantador I, Fernández-Tobías I, Berkovsky S, et al. Cross-domain recommender systems//Ricci F, Rokach L, Shapira B eds. *Recommender Systems Handbook*, 2nd Edition. New York, America; Springer, 2015: 919-959
- [9] Mirbakhsh N, Ling C X. Improving top-N recommendation for cold-start users via cross-domain information. *ACM Transactions on Knowledge Discovery from Data*, 2015, 9(4): 33(1-19)
- [10] Tan S, Bu J, Qin X, et al. Cross domain recommendation based on multi-type media fusion. *Neurocomputing*, 2014, 127: 124-134
- [11] Yi C, Shang M S, Zhang Q M. Auxiliary domain selection in cross-domain collaborative filtering. *Applied Mathematics & Information Sciences*, 2015, 9(3): 1375-1381
- [12] Li C Y, Lin S D. Matching users and items across domains to improve the recommendation quality//Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, USA, 2014: 801-810
- [13] Zhao L, Pan S J, Xiang E W, et al. Active transfer learning for cross-system recommendation//Proceedings of the 27th AAAI Conference on Artificial Intelligence, Washington, USA, 2013: 1205-1211
- [14] Shi Y, Larson M, Hanjalic A. Tags as bridges between domains: Improving recommendation with tag-induced cross-domain collaborative filtering//Proceedings of the 19th International Conference on User Modeling, Adaption, and Personalization, Girona, Spain, 2011: 305-316
- [15] Pan W, Xiang E W, Liu N N, et al. Transfer learning in collaborative filtering for sparsity reduction//Proceedings of the 24th AAAI Conference on Artificial Intelligence, Atlanta, USA, 2010: 230-235
- [16] Li B, Yang Q, Xue X. Can movies and books collaborate? cross-domain collaborative filtering for sparsity reduction//Proceedings of the 21st International Joint Conference on Artificial Intelligence, Pasadena, USA, 2009: 2052-2057
- [17] Gu Q, Zhou J, Ding C H Q. Collaborative filtering: Weighted nonnegative matrix factorization incorporating user and item graphs//Proceedings of the 10th SIAM International Conference on Data Mining, Columbus, USA, 2010: 199-210
- [18] Shi J, Long M, Liu Q, et al. Twin bridge transfer learning for sparse collaborative filtering//Proceedings of 17th Pacific-Asia Conference, Advances in Knowledge Discovery and Data Mining, Gold Coast, Australia, 2013: 496-507
- [19] Shen Guo-Wei, Yang Wu, Wang Wei, et al. Large-scale heterogeneous data co-clustering based on nonnegative matrix

factorization. *Journal of Computer Research and Development*, 2016, 53(2): 459-466(in Chinese)

(中国伟, 杨武, 王巍等. 基于非负矩阵分解的大规模异构数据联合聚类. *计算机研究与发展*, 2016, 53(2): 459-466)



**WANG Jun**, born in 1986, Ph. D. candidate. His research interests include transfer learning, recommendation systems and data mining.

**LI Shi-Jun**, born in 1964, Ph. D., professor. His research interests include big data, web search and mining, database technology, mobile data mining and research on spatial and temporal consistency.

**YANG Sha**, born in 1980, Ph. D. candidate, lecturer. Her research interests include mining and analysis of the electronic commerce data, service quality evaluation of electronic business and data mining of the web economy behavior.

**JIN Hong**, born in 1984, Ph. D. Her research interests include machine learning, complex network and community detection.

**YU Wei**, born in 1987, Ph. D., lecturer. His research interests include data extraction, data fusion and data quality evaluation.

## Background

In this paper, we research cross-domain collaborative filtering recommendation, which is a hot and challenging problem in recommendation systems for the past few years. Collaborative filtering is a simple and usual recommendation method, but its performance will degrade seriously when the target data is very sparse. Cross-domain recommendation with the help of auxiliary data associated across domains with target data is an effective way to solve this problem. Transfer learning is a popular method which can be used for cross-domain recommendation. However, there are still some shortages in existing transfer learning models for cross-domain recommendation:

(1) They conduct transfer learning based on the rating pattern, and usually assume that relevant domains share a rating pattern completely, which ignores the differences of rating patterns across domains and may lead to serious negative transfer.

(2) They ignore that the shared rating pattern may have different focuses on users and items. And they always only utilize users' rating pattern which focuses on the interest pattern of users, but have not utilized items' rated pattern which focuses on the popularity pattern of items.

(3) They conduct transfer learning only based on a single bridge such as latent factors or rating pattern, which may not generate sufficient positive transfer.

In order to solve aforementioned problems, we propose a new triple-bridge transfer learning model for cross-domain collaborative filtering recommendation. Our proposed model reduces negative transfer effectively, as well as increases

positive transfer significantly. Our main contributions are as follows:

(1) Based on the similarity constraints between latent factors, we extract latent factor and shared rating pattern of users as well as latent factor and shared rated pattern of items from auxiliary domains via collective factorizations on rating matrices. By considering domain-specific patterns, our model reduces the negative transfer efficiently.

(2) We consider the different focuses of the shared pattern on users and items for the first time. Based on the shared rating pattern of users and the shared rated patterns of items, we transfer the interest pattern of users and the popularity pattern of items separately.

(3) We extract the cluster information from latent factors of users and items separately to construct adjacency graphs of users and items. The user-side triple-bridge transfer learning is conducted based on the shared rating pattern of users, latent factors and adjacency graphs. And the item-side triple-bridge transfer learning is conducted based on the shared rated pattern of items, latent factors and adjacency graphs. We predict the missing ratings of objective rating matrix jointly by the user-side and item-side triple-bridge transfer learning, which increases positive transfer significantly.

The work is supported by the National Natural Science Foundation of China (61272109), the Fundamental Research Funds for the Central Universities (2042014KF0057), and the Natural Science Foundation of Hubei Province of China (2014CFB289).