

# 基于用户类别偏好相似度和联合矩阵分解的推荐算法

何海洋<sup>1,2</sup> 王勇<sup>1,2,3</sup> 蔡国永<sup>1,3</sup>

(1. 桂林电子科技大学计算机与信息安全学院, 桂林, 541004; 2. 桂林电子科技大学广西密码学与信息安全重点实验室, 桂林, 541004; 3. 桂林电子科技大学广西可信软件重点实验室, 桂林, 541004)

**摘要:** 利用上下文信息来提高推荐准确率并增强用户体验是当前推荐领域研究热点之一, 然而现有的上下文感知推荐算法依然面临数据稀疏性问题的挑战。为了进一步缓解数据稀疏性问题, 本文提出一种基于用户类别偏好相似度及联合矩阵分解的推荐算法(Joint matrix factorization with user category preference, JMF-UCP), 它结合用户评分数据及用户类别偏好进行物品推荐, 以解决用户评分数据稀疏时评分预测准确率较低的问题。算法的时间复杂度随着数据量的增加呈线性增长, 因此适用于大规模数据。通过在真实数据集 Movielens 上的实验结果表明, 本文提出的方法在 RMSE 评价指标上优于现有代表性的算法, 验证了本文所提出的推荐算法的有效性。

**关键词:** 推荐系统; 联合矩阵分解; 用户类别偏好; 评分预测

**中图分类号:** TP391.4 **文献标志码:** A

## Recommendation Algorithm Based on User Category Preference Similarity and Joint Matrix Factorization

He Haiyang<sup>1,2</sup>, Wang Yong<sup>1,2,3</sup>, Cai Guoyong<sup>1,3</sup>

(1. School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin, 541004, China; 2. Guanxi Key Laboratory of Cryptography and Information Security, Guilin University of Electronic Technology, Guilin, 541004, China; 3. Guangxi Key Laboratory of Trusted Software, Guilin University of Electronic Technology, Guilin, 541004, China)

**Abstract:** Using context information to improve the accuracy of recommendation systems and enhance user experience is one of the hottest topics in the domain of recommend systems. However the issue of data sparse still challenges the existing context-aware recommender system. To better alleviate the data sparse problem, this paper proposes a rating prediction method, i. e., joint matrix factorization with user category preference(JMF-UCP). Based on the joint matrix factorization, the method addresses the data sparse problem by combining user's rating information and user category preference to predict the rating score with higher accuracy. The time complexity of the proposed method linearly increases with the number of amount of dataset and is scalable to very large datasets. Experimental results on real world rating dataset MovieLens demonstrate that the proposed method can achieve better accuracy.

**Key words:** recommender system; joint matrix factorization; user category preference; rating prediction

## 引言

互联网的出现和普及给用户带来了大量的信息,网络信息的获取变得非常简单。但随着互联网技术的飞速发展,网络信息呈现指数级增长的态势,使得用户不容易从纷繁的信息中找到适合自己的内容。如何向用户准确地推荐他们想要的信息成为亟待解决的问题。推荐系统作为信息过滤中的一种重要技术,主要是通过分析用户的行为记录(包括:购买行为、收藏行为、浏览记录等)对用户的兴趣偏好进行建模,向用户推荐他们需要的产品和信息,并生成个性化推荐以满足用户个性化的需求。根据美国科技博客 VentureBeat 的官网调查,Amazon 中 35% 的商品及应用的销售额来自于它的推荐系统。由于推荐系统巨大的商业价值,受到了学术界和工业界的广泛关注。然而,随着电商网站上用户数目和项目数量的日益增加,在整个项目空间上用户和项目的评分数据变得极端稀疏,这就产生了数据稀疏性问题,导致推荐系统准确率偏低。

对于系统中大部分不活跃的用户,他们的评分数据通常比较少,因此从用户项目评分关系中很难刻画用户之间的偏好。针对评分数据稀疏性问题,学者们从不同的方面进行了研究,其中最为有效的方法就是融入额外的社交媒体上下文信息来缓解评分数据稀疏性问题<sup>[1]</sup>,但大部分都是直接利用上下文信息进行信息的融合,从而产生了辅助信息过多、推荐算法的时间复杂度较高的问题。本文首先从用户类别的角度构建用户类别偏好相似度,从而有效地缓解当前推荐系统所面临的冷启动、数据稀疏性等问题。另外,本文提出推荐算法的时间复杂度随着观测数据的增加呈线性增长,可应用于大规模的数据集。

## 1 相关工作

传统的协同过滤推荐技术主要包括基于模型的协同过滤推荐算法和基于记忆的协同过滤推荐算法。其主要思想是根据用户的历史行为数据来分析用户的偏好,并假设用户对某一项目的喜好通常与其有着相似兴趣的用户相一致<sup>[2]</sup>。基于记忆的协同过滤算法通过在用户项目评分矩阵中找相似的最近邻居,并根据推算出的最近邻居的评分对待评估项目进行评分<sup>[3, 4]</sup>。基于模型的协同过滤推荐算法事先根据已知的用户项目评分矩阵建立一个有效的模型,然后根据该模型来预测未知评分。基于模型的协同过滤推荐算法主要包括分类算法<sup>[5]</sup>、聚类算法<sup>[6]</sup>和矩阵分解等。

矩阵分解技术因其评分预测精度高、可扩展性及可适用于稀疏的数据集而广泛应用于推荐系统<sup>[7-11]</sup>。矩阵分解模型通过分解用户项目评分矩阵将推荐系统中的用户和项目映射到一个低秩的联合潜在特征空间,然后通过用户和项目相应的潜在特征预测用户对项目的未知评分。近年来,为了缓解数据稀疏性问题,进一步提高推荐系统的准确率,一些基于基本矩阵分解的扩展方法被提出。文献[8]提出一种将电影情感特征相似度矩阵和用户电影评分矩阵进行联合分解的电影推荐算法。文献[9]提出一种将社会关系作为正则化项添加到基本的用户项目矩阵分解中的社会化推荐算法。文献[10]提出一种将用户地标偏好矩阵和基于类别的地标偏好相似度进行联合分解的个性化地标推荐算法。文献[11]中利用共享潜在特征,将不同的社会上下文信息作为正则化项融入低秩矩阵分解过程。其中通过融合用户信任关联矩阵提出了 SoRec 模型、融合用户标签关联矩阵提出了 SoRecUser 模型、融合项目标签关联矩阵提出了 SoRecItem 模型。本文将用户类别偏好相似度矩阵和用户项目评分矩阵进行联合分解,有效地缓解了数据集稀疏性问题,从而达到提高评分预测准确率并增强用户体验的目的。

## 2 问题的定义

假设一个推荐系统中有  $n$  个物品和  $m$  个用户,则令  $U = \{u_1, u_2, \dots, u_n\}$  表示用户集合,  $I = \{i_1, i_2, \dots, i_m\}$  表示项目集合。在推荐系统中,对项目进行分类,可以更好地帮助用户找到自己感兴趣的项目。例如电影评论网站 MovieLens,根据电影的分类,为不同类型的电影打上标签(如:喜剧、爱情等)。令

$C = \{c_1, c_2, \dots, c_p\}$  表示类别集合。其中一个用户可以评论多个项目,一个项目可以分属不同的类别。在实际的应用中,用户对物品是否喜爱倾向于受到与自己兴趣相同的用户影响,并且用户间的兴趣相似度越大,用户受到影响的可能性也越大。由于用户项目评分矩阵的稀疏性,本文从用户类别的角度来构建用户之间的相似度矩阵,同时结合用户-项目评分矩阵进行联合矩阵分解,提出一种基于用户类别偏好相似度和联合矩阵分解的推荐算法(Joint matrix factorization with user category preference, JMF-UCP),以缓解数据集稀疏性问题,从而达到提高评分预测准确率的目的。

### 3 JMF-UCP 模型框架

图 1 给出了推荐算法的详细流程,基于用户类别偏好相似度和联合矩阵分解的推荐算法 JMF-UCP 主要包含以下 3 个部分:

(1) 对用户类别偏好相似度进行建模。根据用户项目(User-Item)评分矩阵及项目类别(Item-Category)关系矩阵建立用户类别(User-Category)关联度,然后计算用户类别偏好相似度。

(2) 求解隐含特征向量。该算法将用户评分矩阵和用户类别偏好相似度矩阵进行联合分解形成最优化目标函数,使用梯度下降算法学习得到用户和项目的特征向量。

(3) 评分预测。根据求解的特征向量,求解用户的未知评分完成推荐。

#### 3.1 用户类别偏好相似度

为了度量用户类别偏好相似度,本文采用文献[12]中的方法。假设某分类系统为某类项目预先定义了  $p$  个分类标签,用户  $u$  对项目  $i$  打过评分且项目  $i$  分属  $n$  个分类,则项目  $i$  对应的每个分类标签将获得用户  $u$  的关注度为  $1/t$ 。由此用户  $u$  对类别  $c$  的关注度公式为

$$a_{uc} = \frac{\sum_{i \in D_k(u)} \frac{\text{sgn}(u, i, c)}{N(i)}}{k} \quad (1)$$

式中:如果项目  $i$  属于类别  $c$ ,则  $\text{sgn}(u, i, c) = 1$ ,否则为 0;  $a_{uc}$  为用户  $u$  对类别  $c$  的偏好值;  $D_k(u)$  为用户  $u$  评论过的项目集合;  $k$  为集合  $D_k(u)$  中的元素个数。由此可建立用户  $u$  的 User-Category 偏好向量为

$$\mathbf{A}_u = (a_{u1}, a_{u2}, \dots, a_{up}) \quad (2)$$

本文采用余弦相似度来度量用户类别偏好相似度<sup>[13]</sup>,用户  $i$  和用户  $j$  之间的类别偏好相似度为

$$S_{ij}^{(\text{ucp})} = \frac{\sum_{o=1}^p a_{io} a_{jo}}{\sqrt{\sum_{o=1}^p a_{io}^2} \sqrt{\sum_{o=1}^p a_{jo}^2}} \quad (3)$$

#### 3.2 融入用户类别偏好相似度的矩阵分解

基于矩阵分解模型的推荐方法将用户的评分项目矩阵 User-Item 用 1 个稀疏的二维矩阵表示,然后将稀疏且高维的评分矩阵分解为 2 个低秩稠密矩阵,通过重构的低维矩阵预测用户对项目的评分<sup>[7]</sup>。其需要的优化函数为

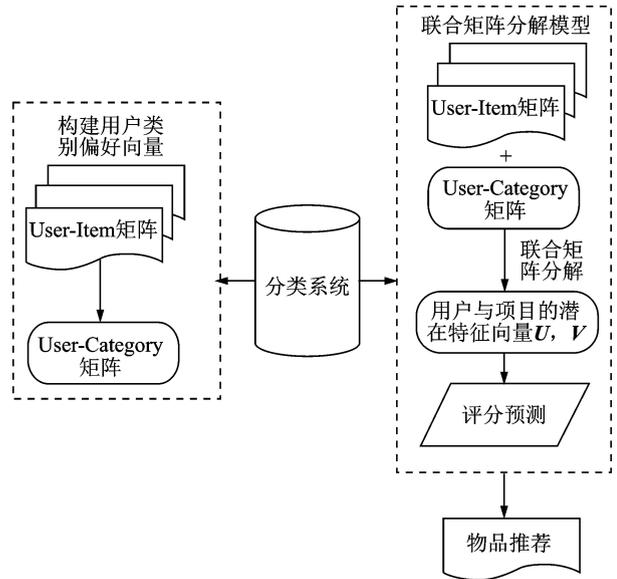


图 1 算法流程图

Fig. 1 Flow diagram of algorithm

$$\min_{\mathbf{U}, \mathbf{V}} L_1(\mathbf{R}, \mathbf{U}, \mathbf{V}) = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n I_{ij} (\mathbf{R}_{ij} - \mathbf{U}_i^T \mathbf{V}_j)^2 + \frac{\lambda_U}{2} \|\mathbf{U}\|^2 + \frac{\lambda_V}{2} \|\mathbf{V}\|^2 \quad (4)$$

式中:  $m, n$  分别表示用户和项目的数量;  $\|\cdot\|_F^2$  表示 Frobenius 范数;  $\mathbf{R}_{m \times n}$  表示评分矩阵;  $\mathbf{U}_{m \times d}$  表示用户的偏好特征矩阵;  $\mathbf{V}_{n \times d}$  表示项目的特征矩阵;  $I$  表示一个指示函数, 如果  $\mathbf{R}_{ij}$  有值, 则  $I_{ij}$  为 1, 如果  $\mathbf{R}_{ij}$  缺失, 则  $I_{ij}$  为 0;  $\|\mathbf{U}\|^2$  和  $\|\mathbf{V}\|^2$  是为防止过拟合的正则化项;  $\lambda_U$  和  $\lambda_V$  为正则化系数。

文献[10]中提出的 SoRecUser 模型通过共享用户潜在特征方法同时分解用户项目评分矩阵和用户分类标签关联矩阵, 以缓解项目评分矩阵数据稀疏性问题, 提高推荐准确率。其优化函数为

$$\min_{\mathbf{U}, \mathbf{V}} L_2(\mathbf{R}, \mathbf{S}, \mathbf{U}, \mathbf{V}) = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n I_{ij} (\mathbf{R}_{ij} - \mathbf{U}_i^T \mathbf{V}_j)^2 + \frac{\partial}{2} \sum_{i=1}^m \sum_{k=1}^m J_{ik} (\mathbf{F}_{ik} - \mathbf{U}_i^T \mathbf{T}_k)^2 + \frac{\lambda_U}{2} \|\mathbf{U}\|^2 + \frac{\lambda_V}{2} \|\mathbf{V}\|^2 + \frac{\lambda_T}{2} \|\mathbf{T}\|^2 \quad (5)$$

式中:  $\mathbf{F}$  为用户与分类标签的关联矩阵;  $J$  为一个指示函数;  $\mathbf{T}$  为标签潜在特征矩阵。

为了进一步缓解数据稀疏性问题, 本文首先通过用户项目的评分关系与项目类别的关联关系构建用户类别偏好相似度矩阵  $\mathbf{S}^{(\text{ucp})}$ , 然后通过联合矩阵分解同时分解用户项目评分矩阵和用户类别偏好相似度矩阵  $\mathbf{S}^{(\text{ucp})}$ , 其优化函数为

$$\min_{\mathbf{U}, \mathbf{V}} L_3(\mathbf{R}, \mathbf{S}, \mathbf{U}, \mathbf{V}) = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n I_{ij} (\mathbf{R}_{ij} - \mathbf{U}_i^T \mathbf{V}_j)^2 + \frac{\partial}{2} \sum_{i=1}^m \sum_{k=1}^m J_{ik} (\mathbf{S}_{ik}^{(\text{ucp})} - \mathbf{U}_i^T \mathbf{U}_k)^2 + \frac{\lambda_U}{2} \|\mathbf{U}\|^2 + \frac{\lambda_V}{2} \|\mathbf{V}\|^2 \quad (6)$$

式中:  $J$  为一个指示函数, 如果  $\mathbf{S}_{ik}^{(\text{ucp})}$  有值, 则  $J_{ik}$  为 1, 如果  $\mathbf{S}_{ik}^{(\text{ucp})}$  缺失, 则  $J_{ik}$  为 0;  $\partial$  为权衡系数, 用来控制用户类别偏好相似度对评分预测的影响。目标函数主要包含两个部分, 即在将用户评分矩阵分解成低维的用户潜在特征矩阵和项目潜在特征矩阵时, 同时分解用户类别偏好相似度。该联合矩阵分解模型可以有效地缓解数据稀疏性问题, 对于评分较少的用户, 很难从用户评分矩阵中捕捉其潜在的特征, 在此模型中可以通过用户类别偏好相似度来学习此类用户的潜在特征。

为了得到满足式(6)中的两个低秩矩阵  $\mathbf{U}$  和  $\mathbf{V}$ , 这里使用梯度下降搜索目标函数  $L_3$  的局部最小值。为此, 式(6)分别对  $\mathbf{U}, \mathbf{V}$  进行求导

$$\frac{\partial L_3}{\partial \mathbf{U}_i} = \sum_{j=1}^n I_{ij} (\mathbf{U}_i^T \mathbf{V}_j - \mathbf{R}_{ij}) \mathbf{V}_j + \partial \sum_{k=1}^m J_{ik} (\mathbf{U}_i^T \mathbf{U}_k - \mathbf{S}_{ik}^{(\text{ucp})}) \mathbf{U}_k + \lambda_U \mathbf{V}_j \quad (7)$$

$$\frac{\partial L_3}{\partial \mathbf{V}_j} = \sum_{i=1}^n I_{ij} (\mathbf{U}_i^T \mathbf{V}_j - \mathbf{R}_{ij}) \mathbf{U}_i + \lambda_V \mathbf{U}_i \quad (8)$$

### 3.3 JMF-UCP 算法

基于用户类别偏好相似度和联合矩阵分解的推荐方法。算法 1 给出 JMF-UCP 推荐算法的算法描述。

#### 算法 1 JMF-UCP 推荐算法

输入: 用户的评分矩阵  $\mathbf{R}$

用户类别偏好相似度矩阵  $\mathbf{S}^{(\text{ucp})}$

权衡系数  $\partial$

正则化系数  $\lambda_U, \lambda_V$

最大迭代次数  $\text{maxIter}$

停止迭代条件  $\epsilon$

输出: 完整的用户评分矩阵  $\hat{\mathbf{R}}$

// 变量初始化

$\mathbf{U}_{m \times d} \leftarrow (\text{normal}(0, 1)), \mathbf{V}_{n \times d} (\text{normal}(0, 1))$

迭代次数  $t=0$ ;

学习速率  $r=1$ ;

收敛指示器  $f=0$

//利用梯度下降算法求解  $\mathbf{U}, \mathbf{V}$

while( $t < \text{maxIter}$  and  $f=0$ )

do{

    根据式(7,8)计算  $\frac{\partial L_3}{\partial \mathbf{U}^{(t)}}$ ,  $\frac{\partial L_3}{\partial \mathbf{V}^{(t)}}$

    while( $L_3(\mathbf{R}, \mathbf{S}, \mathbf{U}^{(t)} - r \frac{\partial L_3}{\partial \mathbf{U}^{(t)}}, \mathbf{V}^{(t)} - r \frac{\partial L_3}{\partial \mathbf{V}^{(t)}}) \geq L_3^{(t)}$ )

do{

$r = r/2$ ;

}

$\mathbf{U}^{(t+1)} = \mathbf{U}^{(t)} - r \frac{\partial L_3}{\partial \mathbf{U}^{(t)}}$ ,  $\mathbf{V}^{(t+1)} \leftarrow \mathbf{V}^{(t)} - r \frac{\partial L_3}{\partial \mathbf{V}^{(t)}}$

    根据式(6)计算  $L_3^{(t+1)}$

    if( $L_3^{(t)} - L_3^{(t+1)} \leq \epsilon$ )

do{

$f = 1$

}

$t = t + 1$ ;

}

//输出预测评分  $\hat{\mathbf{R}}$

$\hat{\mathbf{R}} = \mathbf{U}\mathbf{V}^T$

### 3.4 JMF-UCP 算法时间复杂度分析

由于现有推荐系统的应用场景中,项目的类别数目往往远小于系统中用户个数和项目个数,且用户类别偏好可事先存放在内存中,单独来计算。因此 JMF-UCP 算法的计算开销主要来自于式(6)中的目标函数  $L_3$  和梯度下降变量的迭代更新。因此目标函数  $L_3$  的时间复杂度为  $O(n_R l + n_S l)$ , 其中  $n_R l, n_S l$  分别表示矩阵  $\mathbf{R}, \mathbf{S}^{(\text{ucp})}$  中的非零元素个数。同理,可以推导出式(7,8)的时间复杂度。因此,每迭代一次总的时间复杂度是  $O(n_R l + n_S l)$ 。综合上述分析,算法的时间复杂度随着矩阵  $\mathbf{R}, \mathbf{S}^{(\text{ucp})}$  的非零元素个数的增加呈线性增长,因此本文算法可以应用于大规模的数据集。

## 4 实验结果及评价

为了验证 JMF-UCP 算法在评分预测方面的有效性,实验将基于推荐领域 MovieLens 1 MB 真实数据集 (<http://grouplens.org/datasets/movieLens/>)。该数据集包含 2000 年 6 400 个独立匿名用户对 3 900 部电影作的 1 000 209 次评分,评分的取值为 1~5 之间的离散值,标签的种类数共有 18 种,电影都被打上不同的分类标签,每个电影对应一个或多个分类标签。

### 4.1 评分预测准确性评估

为了验证本文中提出算法在评分预测中的准确性,采用了评分预测常用的评估方法均方根误差(Root mean squared error, RMSE)<sup>[8, 11]</sup>, 公式如下

$$\text{RMSE} = \sqrt{\frac{\sum_{u,i} (R_{ui} - \hat{R}_{ui})^2}{|R_t|}} \quad (9)$$

式中:  $R_{ui}$  表示用户  $u$  对项目  $i$  的真实评分;  $\hat{R}_{ui}$  表示预测评分;  $|R_t|$  表示测试集中的评分个数。

可以看出,均方根误差 RMSE 越低,评分预测的准确率越高,推荐系统的性能越好。

为了评估本文提出的推荐算法的性能,通过实验把 JMF-UCP 与另外一些推荐方法进行了比较:(1)Random 方法,该方法为目标用户随机产生邻居的方法;(2)UserAvg 方法,该方法根据每个用户的历史平均评分对未知的评分进行预测;(3)协同过滤方法<sup>[14]</sup>(CF),该方法是目前使用最为广泛的基于内存的推荐方法;(4)非负矩阵方法<sup>[15]</sup>(NMF),该方法是基本的矩阵分解,其基本形式如式(4)所示,其中正则项参数  $\lambda_U$  及  $\lambda_V$  的取值与 JMF-UCP 方法相同;(5)JMF-UP 方法,该方法首先根据用户项目评分矩阵构建用户间的相似度矩阵  $S$ ,然后利用联合矩阵分解融合基本的用户偏好  $S$  来完成评分预测,其形式如式(6)所示,将其中的用户类别偏好相似度  $S^{(ucp)}$  替换成  $S$ ,其中所需要的参数取值与 JMF-UCP 方法中相同;(6)SoRecUser 方法,该方法利用共享潜在特征,将用户和分类标签的关联关系融入评分矩阵的低秩矩阵分解过程<sup>[10]</sup>。

试验中评分数据集被分为两个部分:随机抽取 80% 的评分数据作为训练集,剩下 20% 的评分数据作为测试集。为了获取真实可信的实验结果,相同条件下实验重复运行 10 次,最终的实验结果是这 10 次的平均值。试验中正则化参数  $\lambda_U$  及  $\lambda_V$  取值为 0.001,潜在因子个数  $d=10$ 。图 2 给出了不同推荐方法的实验结果对比。

从图 2 的实验结果可以得知,矩阵分解方法 NMF 要优于传统的协同过滤推荐方法,而采用基本用户偏好的联合矩阵分解的 JMF-UP 方法由于其存在较严重的数据稀疏性问题,相比较基本矩阵分解方法提高效果不明显,而融合了上下文信息的 SoRecUser 模型有明显的提高。本文提出的推荐算法在 RMSE 的评价指标上为最优,相比其他方法具有更高的评分预测准确率,更好地缓解了数据集稀疏性问题。

#### 4.2 参数 $\vartheta$ 对 RMSE 的影响

在 3.1 小节的实验中,JMF-UCP 中的参数  $\lambda_U$ ,  $\lambda_V$  和  $d$  使用了普遍可以接受的经验值,但是参数  $\vartheta$  控制着用户类别偏好在推荐系统中的重要性, $\vartheta$  取值越大则用户类别偏好对推荐系统影响也越大。因此,本文针对参数  $\vartheta$  进行实验,着重研究参数  $\vartheta$  对 JMF-UCP 模型的性能影响,通过调整参数  $\vartheta$  的值,观察 JMF-UCP 模型的性能。实验结果如图 3 所示,参数  $\vartheta$  取不同值的情况下,JMF-UCP 评分预测的 RMSE 值有不同的变化。从图 3 中看出用户类别偏好的控制参数  $\vartheta$  大于 0.1 或者小于 0.1 时,实验结果的均方根误差值都会上升。因此,3.1 小节中实验参数  $\vartheta$  设置为 0.1 是合理的。同时上述实验结果也说明,适当的考虑用户对类别的偏好可以进一步提高推荐系统的性能。

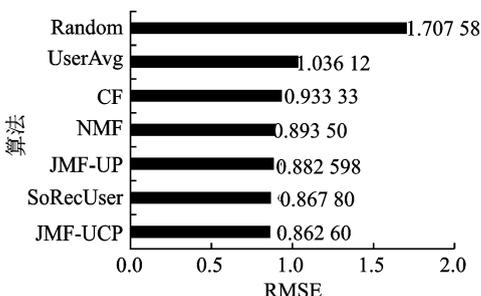


图 2 不同算法对比图

Fig. 2 Comparison of RMSE by different algorithms

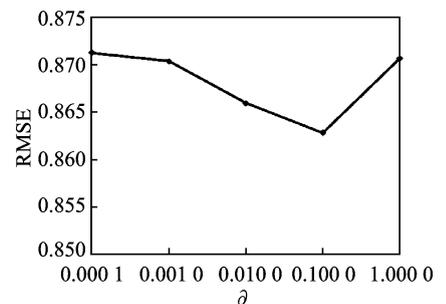


图 3 参数  $\vartheta$  对 RMSE 的影响

Fig. 3 Impact of parameter  $\vartheta$  on RMSE

## 5 结束语

针对推荐系统实际应用中数据稀疏性问题,本文提出一种基于用户类别偏好相似度和联合矩阵分解的推荐算法。首先通过推荐系统中的项目分类功能以及用户项目评分矩阵,构建用户类别偏好相似度,然后通过联合矩阵分解同时分解用户项目评分矩阵和用户类别偏好相似度,构建最优化目标函数,

并利用梯度下降的算法求解目标函数,从而获取用户未知评分完成推荐。在真实数据集上的实验评估证明了本文提出的推荐算法具有较好的准确率,有效地缓解了数据稀疏性问题。同时本文所提出算法的时间复杂度随着观察数据的增加呈现线性增长,因此可应用于大规模数据。然而,本文模型也存在一定的局限性,当推荐系统中的数据所涉及时间跨度比较大时,用户类别偏好的建模准确性也会下降,从而影响算法的准确率。通过对分类系统进行及时的跟踪,及时更新用户类别偏好有助于保持评分预测的准确率。另外未来可考虑结合用户间的朋友关系与信任关系更加贴切地对用户间的相似度进行建模,进一步提高推荐的准确率。

### 参考文献:

- [1] 高新波, 沈钧戈. 基于社会媒体的旅游数据挖掘与分析[J]. 数据采集与处理, 2016, 31(1):18-27.  
Gao Xinbo, Shen Junge. Social media based travel data mining and analysis[J]. Journal of Data Acquisition & Processing, 2016, 31(1):18-27.
- [2] 郭磊, 马军, 陈竹敏. 一种信任关系强度敏感的社会化推荐算法[J]. 计算机研究与发展, 2015, 50(9): 1805-1813.  
Guo Lei, Ma Jun, Chen Zhumin. Trust strength aware social recommendation method[J]. Journal of Computer Research and Development, 2015, 50(9): 1805-1813.
- [3] Witten I, Milne D. An effective, low-cost measure of semantic relatedness obtained from Wikipedia links[C]//Proceeding of AAAI Workshop on Wikipedia and Artificial Intelligence: An Evolving Synergy. Chicago, USA: AAAI Press, 2008: 25-30.
- [4] Shi Y, Larson M, Hanjalic A. Exploiting user similarity based on rated-item pools for improved user-based collaborative filtering[C]//Proceedings of the third ACM conference on Recommender Systems. [S.l.]:ACM, 2009: 125-132.
- [5] Park D H, Kim H K, Choi I Y, et al. A literature review and classification of recommender systems research[J]. Expert Systems with Applications, 2012, 39(11): 10059-10072.
- [6] Tsai C F, Hung C. Cluster ensembles in collaborative filtering recommendation[J]. Applied Soft Computing, 2012, 12(4): 1417-1425.
- [7] Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems[J]. Computer, 2009 (8): 30-37.
- [8] Shi Y, Larson M, Hanjalic A. Mining mood-specific movie similarity with matrix factorization for context-aware recommendation[C]//The Workshop on Context-Aware Movie Recommendation. [S.l.]:ACM, 2010:34-40.
- [9] Ma H, Zhou D, Liu C, et al. Recommender systems with social regularization[C]//ACM International Conference on Web Search and Data Mining. [S.l.]: ACM, 2011:287-296.
- [10] Shi Y, Serdyukov P, Hanjalic A, et al. Personalized landmark recommendation based on Geotags from photo sharing sites [C]// International Conference on Weblogs & Social Media. Barcelona, Catalonia, Spain; DBLP, 2011:622-625.
- [11] Ma H, Zhou T C, Lyu M R, et al. Improving recommender systems by incorporating social contextual information[J]. ACM Transactions on Information Systems (TOIS), 2011, 29(2): 9.
- [12] Cai G, Lv R, Wu H, et al. An improved collaborative method for recommendation and rating prediction[C]//IEEE International Conference on Data Mining Workshop. [S.l.]: IEEE, 2014:781-788.
- [13] Linden G, Smith B, York J. Amazon. com recommendations: Item-to-item collaborative filtering[J]. IEEE Internet Computing, 2010, 7(1):76-80.
- [14] Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions[J]. Knowledge and Data Engineering, IEEE Transactions on, 2005, 17(6): 734-749.
- [15] Zhang S, Wang W, Ford J, et al. Learning from incomplete ratings using non-negative matrix factorization [C]// Proceedings of the Sixth SIAM International Conference on Data Mining. Bethesda, MD, USA:[s. n.], 2006, 6: 548-552.

### 作者简介:



**何海洋**(1990-),男,硕士研究生,研究方向:社交媒体挖掘、社交媒体信任评估, E-mail: haiyang0902@gmail.com。



**王勇**(1977-),男,副研究员,硕士生导师,中国计算机学会高级会员,研究方向:大数据环境下的数据安全、信息论以及电子商务安全。



**蔡国永**(1971-),男,博士,教授,中国计算机学会高级会员、IEEE 会员,研究方向:社交媒体信任评估、谣言检测与传播管控、群体情感或观点挖掘以及信息推荐。

(编辑:张黄群)

