•信息工程•

DOI:10.15961/j.jsuese.201800528

基于用户兴趣评分填充的改进混合推荐方法

李　征[1,2],段　垒[1]

(1.河南大学 计算机与信息工程学院,河南 开封 475004;2.三峡大学 湖北省水电工程智能视觉监测重点实验室,湖北 宜昌 443002)

摘　要:针对传统协同过滤推荐方法中的用户项目评分数据稀疏和推荐准确度不高的问题,提出了一种基于用户兴趣评分填充的改进混合推荐方法。首先,分析用户对项目类型的偏好,计算用户兴趣评分并进行矩阵填充;然后,考虑用户主观评分差异化及项目自身质量的影响,对传统皮尔逊相关系数进行改进,并基于已填充评分矩阵进行用户相似性及项目相似性计算;在此基础上分别基于用户和项目两个方面进行评分预测,并将两者的预测评分进行加权求和,进而进行混合推荐;最后,以Movielens100k为数据集进行实验,先分析了用户兴趣评分矩阵的填充效果,再将文中方法和传统协同过滤混合推荐方法以及文献中提出方法进行了对比分析。实验结果表明:提出的评分矩阵填充方法能有效缓解数据稀疏的影响,填充效果优于传统评分矩阵填充方法;提出的改进混合推荐方法(IHRIRF)比传统的混合协同过滤推荐方法HCFR及WPCC方法具有更好地推荐效果。

关键词:协同过滤;数据稀疏;评分差异化;混合推荐;皮尔逊相关系数

中图分类号:TP311　　**文献标志码**:A　　**文章编号**:2096-3246(2019)01-0189-08

Improved Hybrid Recommendation Approach Based on User Interest Ratings Filling

LI Zheng[1,2], *DUAN Lei*[1]

(1.School of Computer and Info. Eng., Henan Univ., Kaifeng 475004, China; 2.Key Lab. of Intelligent Vision Monitoring for Hydropower Project of Hubei Province, Three Gorges Univ., Yichang 443002, China)

Abstract: Aiming at the problems of data sparseness of user item ratings and low recommendation accuracy in traditional collaborative filtering recommendation methods, an improved hybrid recommendation approach based on user interest ratings filling was proposed. Firstly, the users' preference to the item types was analyzed, and the user interest ratings were calculated. Afterwards, the operation of matrix filling was performed. Then the impact of users' subjective ratings differentiation and the item's own quality were considered, and the traditional Pearson correlation coefficient was improved. Based on the filled ratings matrix, users' similarity and items' similarity were computed, to predict ratings from the perspective of users and items respectively. Moreover, the weighted sum of two predicted ratings was calculated further to perform the hybrid recommendation. Finally, experiments were carried out on the Movielens100k dataset. The filling effect of user interest ratings matrix was analyzed firstly, and then the proposed approach, traditional collaborative filtering recommendation methods, and previous methods in the literature were compared and analyzed. The results show that the proposed matrix filling method can effectively alleviate the effect of data sparseness, and the filling effect is better than traditional ratings matrix filling methods. Furthermore, our improved hybrid recommendation approach (IHRIRF) has better recommendations than traditional collaborative filtering recommendation method HCRF as well as WPCC method.

Key words: collaborative filtering; data sparseness; rating differentiation; hybrid recommendation; Pearson correlation coefficient

信息过载问题使得用户选择个性化需求的服务变得越来越困难。推荐系统作为一种能够解决信息过载的有效手段受到越来越多的关注。推荐系统能够根据用户的历史服务使用记录分析用户的偏好,从而挖掘用户的潜在需求。

已有的推荐技术主要包括协同过滤推荐、基于

收稿日期:2018 – 05 – 16

基金项目:国家重点基础研究发展计划资助项目(2014CB340404);国家自然科学基金资助项目(61402150);中国博士后科学基金资助项目(2016M592286);河南省科技研发专项资助(182102410063);三峡大学水电工程智能视觉监测湖北省重点实验室开放基金资助项目(2016KLA04);河南大学科研基金资助项目(2013YBZR015)

作者简介:李　征(1984—),女,副教授,博士. 研究方向:Web服务发现与推荐;软件工程. E-mail:lizheng@henu.edu.cn

网络出版时间:2018 – 12 – 21 09 : 28 : 00　　**网络出版地址**:http://kns.cnki.net/kcms/detail/51.1773.TB.20181219.1145.001.html

内容的推荐以及混合推荐等等，已实际应用于大型电子商务系统中[1]。有研究表明：协同过滤方法相比其他推荐方法具有更好的准确性和扩展性[2–4]。协同过滤方法可分为基于记忆和基于模型的方法，本研究侧重基于记忆的协同过滤。基于记忆的协同过滤一般又分为基于用户和基于项目的方法。基于用户的协同过滤就是找到一组和目标用户相似性较高的邻居用户，根据邻居用户进行项目评分预测[5]。类似地，基于项目的协同过滤就是找到目标项目的邻居项目，根据邻居项目进行项目评分预测。但是，在很多系统中，用户使用过的项目所占比例非常小，评分信息极其稀疏，而数据稀疏问题容易导致推荐结果不准确[6]。

目前，相关研究者针对协同过滤推荐已开展了大量研究。Liu等[7]提出一种考虑到用户共同使用项目数的改进皮尔逊相关系数计算方法WPCC，相比传统协同过滤推荐方法准确率有了一定的提高。王海燕等[8]根据服务的推荐属性特征改进了传统的皮尔逊相关系数计算方法，并结合信任关系找到可信的邻居用户，提出一种能抵抗恶意攻击的推荐方法。Guo等[9]考虑社会信任信息，通过合并用户的可信赖邻居的评分补充和表示用户的偏好，找到具有类似偏好的其他用户，进而提出一种合并信任的推荐方法。

Cai等[10]结合认知心理学借鉴客体典型性的思想，根据用户群中的用户典型程度找到用户的邻居，提出一种结合典型的协同过滤方法。Yu等[11]考虑不同用户间的共同知识，使用基于角色的方法向同一上下文组中的其他成员推荐服务，提出一种高效的角色挖掘推荐方法。范波等[12]打破单一的评分相似度限制，对用户相似性的计算从不同项目类型进行考虑，提出一种多角度计算相似性的推荐方法。

上述方法在一定程度上提高了推荐准确率，但数据稀疏问题并没有得到有效解决。为此，王详德等[13]选择非精确拉格朗日乘子法弥补数据缺失项，对填充矩阵使用基于模型的奇异值分解方法进行推荐。潘涛涛等[14]利用置信系数区分填充值与真实值间的可信度，同时考虑物品可信度与可预测性，提出矩阵填充和物品可预测性的方法。韩亚楠等[15]进行矩阵填充时考虑用户对项目的偏好度，同时将用户兴趣可变性的特点融合到传统协同过滤推荐中。袁卫华等[16]提出一种非负约束下的低阶矩阵填充模型LR-NMF，并在此基础上提出一种基于群组的协同推荐方法。

这些方法虽然能缓解数据稀疏并提高一定的推荐准确率，但仍存在以下两个问题：

1）原始评分矩阵的缺失项往往利用用户或项目的评分均值等进行填充，填充数据的可靠性程度有待考量；

2）使用传统的皮尔逊公式进行项目或用户的相似性计算时，忽略了用户主观评分的差异化以及项目自身质量所带来的影响。

针对上述问题，首先，分析用户对项目类型的兴趣（偏好），进而对原始评分矩阵中的缺失项进行填充；然后，考虑用户主观评分差异化以及项目自身质量的影响，提出了一种改进的基于用户和项目的混合协同过滤推荐方法，有效缓解了数据稀疏问题，从而进一步提高了推荐准确率。

1 融合用户兴趣评分的矩阵填充方法

目前，大多数学者仅通过用户或项目评分均值对评分矩阵中的缺失数据进行填充。这种同等对待的填充方法缺乏可靠性，在解决数据稀疏问题的同时可能造成准确率的下降。针对传统矩阵填充的不足，作者提出一种结合用户兴趣评分的矩阵填充方法，该方法充分考虑了用户对项目类型的兴趣偏好，使填充的评分具有更好的推荐效果。

在进行矩阵填充时，充分考虑项目所包含的属性类别信息，同时结合用户项目评分，进一步找到用户对项目所包含属性类型的偏好。这样，对评分缺失项进行填充时，考虑用户对该项目包含类型的偏好信息，使得填充的数据更具有真实性。比如，推荐系统中常用的Movielens数据集，不仅有用户项目评分，而且还包含用户个人信息及电影类型等信息，如表1和2所示。

表 1 用户项目评分信息

Tab. 1 User item ratings information

用户	I_1	I_2	I_3	I_4
U_1	3	5	0	0
U_2	5	2	3	4
U_3	0	5	0	4
U_4	2	0	0	4

表 2 项目所属类型信息

Tab. 2 Item type information

项目	动作	科幻	喜剧	爱情
I_1	0	1	1	0
I_2	1	0	1	1
I_3	0	1	1	0
I_4	1	1	1	0

表1中，数字为用户对项目的评分，范围在“1～5”之间。其中，“0”分表示用户没有使用过该项

目。表2中，“1”表示项目属于该类型，“0”表示项目不属于该类型。

通过分析用户项目评分及项目所属类型信息，能够得到用户对具体项目类型的兴趣评分，计算方法如式（1）所示。然后，根据用户对具体项目类型的兴趣评分信息对缺失项进行填充。

$$a_u = \frac{\sum_{i \in I_{u,a}} R_{u,i}}{|I_{u,a}|} \tag{1}$$

式中，a_u为用户u对项目类型a的评分，$I_{u,a}$为用户u已评分且包含类型a的项目集，$|I_{u,a}|$为该集合中的元素个数，$R_{u,i}$为用户u对项目i的评分。

以表1、2中的数据为例，通过式（1）计算用户U_2对喜剧片类型的兴趣评分。首先，找到用户U_2已评分且包含喜剧片类型的项目集$\{I_1,I_2,I_3,I_4\}$；然后，计算该项目集中评分均值作为喜剧类型的评分，即（5+2+3+4）/4=3.5。类似地，得到表1中每个用户的项目类型兴趣评分，如表3所示。

表 3　用户项目类型兴趣评分信息

Tab. 3　User item type interest ratings information

用户	动作	科幻	喜剧	爱情
U_1	5.0	3.0	4.0	5.0
U_2	3.0	4.0	3.5	2.0
U_3	4.5	4.0	4.5	5.0
U_4	4.0	3.0	3.0	0

具体对一个用户评分缺失项进行填充时，分为以下3种情况：

1）在已评分项目集中存在与预填充项目所有类型完全一致的项目。在这种情况下，为了排除项目的特定类型组合对用户偏好的影响，计算所有已评分项目集中与预填充项目所有类型完全一致的项目的评分均值，并进行填充，如式（2）所示：

$$r_{u,i} = \frac{\sum_{j \in I_{u,i}} R_{u,j}}{|I_{u,i}|} \tag{2}$$

式中，$r_{u,i}$为用户u对项目i的填充评分，$I_{u,i}$为用户u已评分的与预填充项目i类型完全一致的项目集合，$|I_{u,i}|$为该集合中的元素个数，$R_{u,j}$为用户u对项目j的评分。

2）在已评分项目集中不存在与预填充项目所有类型完全一致的项目，但该预填充项目所包含的全部类型均包含在已评分项目集的所有项目类型中。在这种情况下，根据由式（1）得到的用户对项目类型的兴趣评分，计算预填充项目包含类型的评分均值，并进行填充，如式（3）所示：

$$r_{u,i} = \frac{\sum_{a \in P_i} a_u}{|P_i|} \tag{3}$$

式中，$r_{u,i}$为用户u对项目i的填充评分，P_i为项目i包含的类型集合，$|P_i|$为该集合中的元素个数，a_u为用户u对项目类型a的兴趣评分。

3）在预填充项目所包含全部类型中，存在已评分项目集中不包含的项目类型。在这种情况下，预填充项目包含用户从未评分的类型，为了尽可能避免矩阵填充导致的预测准确度降低问题，不对该项目进行填充。

经过上述步骤，能对用户原始项目评分矩阵缺失项进行极大的填充，从而缓解数据稀疏问题，并在实验部分对本文填充方法和传统填充方法进行了对比，验证了方法的填充效果。

2　考虑用户评分差异化及项目自身质量的相似性计算方法

皮尔逊相关系数被广泛应用于协同过滤算法的相似性计算中[17]，即被用于计算两个项目或者两个用户的相似性，如式（4）所示：

$$sim(i,j) = \frac{\sum_{u \in U_{i,j}} (r_{u,i} - \bar{r}_i) * (r_{u,j} - \bar{r}_j)}{\sqrt{\sum_{u \in U_{i,j}} (r_{u,i} - \bar{r}_i)^2 \sum_{u \in U_{i,j}} (r_{u,j} - \bar{r}_j)^2}} \tag{4}$$

式中，$U_{i,j}$为同时使用项目i、j的用户集合，$r_{u,i}$为用户u对项目i的评分，$\bar{r}_i$为评分矩阵中项目i的平均分。

但传统的皮尔逊相关系数计算公式往往忽略了用户主观评分差异化和项目自身质量对相似性结果产生的影响，导致推荐准确率下降。作者考虑到传统的皮尔逊相关系数计算相似度时存在的不足，从项目和用户两个角度对传统皮尔逊相关系数计算公式进行改进。计算项目相似性时，考虑用户评分的差异化对项目相似性计算时的贡献度，对整体评分差异较小的用户进行惩罚，以降低对项目相似性的影响；计算用户相似性时，考虑项目自身质量的差异化，对项目自身质量较高或较低的项目进行惩罚，以降低对用户相似性的影响。

2.1　考虑用户评分差异化的项目相似性计算

由皮尔逊相关系数计算式（4）可知，在计算两个项目的相似性时，需要找到对这两个项目均有评分的用户集，根据用户集中每个用户对两个项目评分的差别计算两个项目的相似性。然而，用户对项目进行评分时，往往存在差异较小的情况。如果一个用户对所有项目都评低分或者对所有项目都评高分。即

该用户整体评分差异化较小，那么，该用户任意两个项目的评分很可能一样，必然导致项目相似性偏大。但是，这种情况下导致项目相似性偏大的原因是该用户整体评分差异性较小，并非因为两个项目就很相似。此时，认为该用户对项目相似性计算的贡献度偏小。如果一个用户对所有项目评分高低均匀时，即该用户整体评分差异化较大，则该用户对项目相似性计算的贡献度也较大。

采用用户所有评分的离散性表示该用户整体评分的差异化情况，即该用户对项目相似性计算时的贡献度，并将其作为一个惩罚因子加入到传统的皮尔逊相关系数计算公式中，如式（5）所示：

$$sim(i,j)=\frac{\sum_{u\in U_{i,j}}(r_{u,i}-\bar{r}_i)*(r_{u,j}-\bar{r}_j)*X_{\mathrm{norm},u}}{\sqrt{\sum_{u\in U_{i,j}}(r_{u,i}-\bar{r}_i)^2\sum_{u\in U_{i,j}}(r_{u,j}-\bar{r}_j)^2}} \tag{5}$$

式中，$X_{\mathrm{norm},u}$为惩罚因子，即用户u对项目相似性计算时的贡献度，是用户u评分标准差的归一化结果，由式（6）计算得到，取值范围为[0～1]。

$$X_{\mathrm{norm},u}=\frac{X_u-X_{\min}}{X_{\max}-X_{\min}} \tag{6}$$

式中：X_u为用户u评分的标准差，即离散性，由式（7）计算得到；$X_{\max}$为所有用户评分标准差的最大值，$X_{\min}$为所有用户评分标准差的最小值。

$$X_u=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-\mu)^2} \tag{7}$$

式中，N为用户u参与评分的所有项目个数，y_i为用户对第i个项目的评分，μ为该用户对其参与评分项目的平均评分。

式（5）与传统的皮尔逊相关系数计算式（4）相比，加入了惩罚因子$X_{\mathrm{norm},u}$，即计算项目相似性时，考虑了每个用户由于评分差异化所产生的贡献度不同，从而使得到的项目相似性更加准确。

2.2 考虑项目自身质量的用户相似性计算

类似地，计算两个用户相似性时，需要找到两个用户共同使用的项目集，根据两个用户对项目集中每个项目评分的差别计算相似性。由于项目自身质量的差异化，同样会影响用户相似性的计算结果。

例如：一个项目的性价比较低，多数用户对该项目的评分都是1分；或者一个项目的性价比非常高，多数用户对该项目的评分都是5分。这种情况下，任意两个用户对该项目的评分很可能也是1分或者5分，产生这种结果的原因可能是该服务的自身质量很高或者很低，并非因为两个用户的偏好相同。如果利用传统的皮尔逊相关系数计算相似性时就会出现结果偏高的情况，但实际上并不能说明两个用户的相似性就很高。针对上述情况，认为该项目对两个用户相似性计算的贡献度偏小。

为解决上述问题，作者将对项目所有评分的离散性作为惩罚因子加入到传统的皮尔逊相关系数公式中，如式（8）所示。

$$sim(u,v)=\frac{\sum_{i\in I_{u,v}}(r_{u,i}-\bar{r}_u)*(r_{v,i}-\bar{r}_v)*X_{\mathrm{norm},i}}{\sqrt{\sum_{i\in I_{u,v}}(r_{u,i}-\bar{r}_u)^2\sum_{i\in I_{u,v}}(r_{v,i}-\bar{r}_v)^2}} \tag{8}$$

式中：$I_{u,v}$为用户u,v共同评分的项目集；$r_{u,i}$为用户u对项目i的评分；$\bar{r}_u$为用户u的平均评分；$X_{\mathrm{norm},i}$为惩罚因子，是项目i评分标准差的归一化结果，取值范围为[0～1]。

式（8）相对于传统的皮尔逊计算式（4），加入了惩罚因子$X_{\mathrm{norm},i}$，即考虑了每个项目的自身质量对用户相似性计算结果的影响，从而使得到的用户相似性更加准确。

3 基于用户兴趣评分填充的改进混合推荐方法

通过改进的相似度计算方法能找到更加合适的邻居用户，进而根据邻居用户对项目进行预测评分。基于用户的协同过滤评分预测如式（9）所示：

$$P_u(r_{u,i})=\bar{r}_u+\frac{\sum_{(v\in N_u)\wedge(r^1_{v,i}=0)}Sim(u,v)*(r_{v,i}-\bar{r}_v)}{\sum_{(v\in N_u)\wedge(r^1_{v,i}=0)}Sim(u,v)} \tag{9}$$

式中，$P_u(r_{u,i})$为用户u的预测评分，$\bar{r}_u$为用户u参与评分的所有项目的平均评分，N_u为用户u的邻居用户集合。

类似地，基于项目的协同过滤评分预测如式（10）所示：

$$P_i(r_{u,i})=\bar{r}_i+\frac{\sum_{(j\in N_i)\wedge(r^1_{uj}=0)}Sim(i,j)*(r_{u,j}-\bar{r}_j)}{\sum_{(j\in N_i)\wedge(r^1_{uj}=0)}Sim(i,j)} \tag{10}$$

式中，$P_i(r_{u,i})$为项目i的预测评分，$\bar{r}_i$为项目i的平均评分，N_i为项目i的邻居项目集合。

最后，使用系数λ（$0\le\lambda\le1$）将基于用户和项目的方法进行结合[18]，如式（11）所示：

$$P(r_{u,i})=\lambda\cdot P_u(r_{u,i})+(1-\lambda)\cdot P_i(r_{u,i}) \tag{11}$$

由式（11）计算出候选项目预测评分，在此基础

上选择评分较高的项目进行推荐。

根据以上工作, 作者提出一种基于用户兴趣评分填充的改进混合推荐方法(improved hybrid recommendation approach based on user interest rating fill, IHRIRF)。首先, 根据算法1得到用户对项目类型的兴趣评分, 并对用户原始评分矩阵进行填充; 然后, 通过算法2利用改进的皮尔逊相关系数分别计算用户以及项目间相似性, 进而将基于用户的和基于项目的方法结合进行推荐。

算法 1　融合用户兴趣评分的矩阵填充方法

输入:

1. 评分矩阵$\boldsymbol{R}=\{[(u_1,i_1),\cdots,(u_1,i_m)],\cdots,[(u_n,i_1),\cdots,(u_n,i_m)]\}$;

2. 项目类型矩阵$\boldsymbol{A}=\{[(i_1,a_1),\cdots,(i_1,a_q)],\cdots,[(i_m,a_1),\cdots,(i_m,a_q)]\}$。

输出:

已填充矩阵$\boldsymbol{R}_{\text{filled}}$。

```
begin
1.  for a to a_q ∈ A do
2.      for i to i_m ∈ R do
3.          if i 包含 a,r_{u,i}>0   then //项目i包含类型a且有评分
4.              sum+=r_{u,i}; count++;
5.      end for
6.      a_u=sum/count;
7.  end for
8.  for i to i_m do
9.      for j to i_m do
10.         if i_type==j_type   then //项目i,j类型完全一致
11.             sum+=r_{u,j}; count++;
12.     end for
13.     if r_{u,i}==0   then
14.         if count!=0   then
15.             r_{u,i}=sum/count;
16.         else
17.             for a to a_q   do
18.                 if i 包含 a   then //项目i包含类型a
19.                     sum+=a_u; count++;
20.             end for
21.             r_{u,i}=sum/count;
22.         R_filled←r_{u,i};
23.  end for
24.  return R_filled;
end begin
```

算法1中, 语句1～7行通过式(1)计算用户对具体项目类型的兴趣评分; 语句8～15行针对文中第1节所述的评分矩阵填充的第1种情况计算缺失项的评分; 语句16～21行针对第2种情况计算缺失项的评分; 语句22～24行根据计算的评分对原始评分矩阵进行填充, 并返回填充后的评分矩阵。

算法 2　基于用户兴趣评分填充的改进混合推荐方法

输入:

$\boldsymbol{R}_{\text{filled}}=\{[(u_1,i_1),\cdots,(u_1,i_m)],\cdots,[(u_n,i_1),\cdots,(u_n,i_m)]\}$。

输出:

Top–k_i: 评分最高的前k个项目。

```
begin
1.  for u to u_n ∈ R_filled   do
2.      for v to u_n ∈ R_filled   do
3.          for i to i_m ∈ R_filled   do
4.              X_max←X_i; X_min←X_i;
5.          end for
6.          X_{norm,i}=(X_i–X_min)/(X_max–X_min); //项目i评分标准差归一化
7.          sim(u,v)_pro←r_{u,i},r_{v,i},X_{norm,i}; //计算用户间相似性
8.          S_u←sim(u,v)_pro;
9.          rank(S_u);
10.     end for
11. end for
12. for i to i_m ∈ R_filled   do
13.     for j to i_m ∈ R_filled   do
14.         for u to u_n ∈ R_filled   do
15.             X_max←X_u; X_min←X_u;
16.         end for
17.         X_{norm,u}=(X_u–X_min)/(X_max–X_min);
18.         sim(i,j)_pro←r_{u,i},r_{u,j},X_{norm,u}; //计算项目间相似性
19.         S_i←sim(i,j)_pro;
20.         rank(S_i);
21.     end for
22. end for
23. for v_neighbor to top–k ∈ S_u   do
24.     P_u(r_{u,i})←sim(u,v)_pro,r_{v,i}; //计算基于用户的预测评分
25. end for
26. for j_neighbor to top–k ∈ S_i   do
27.     P_i(r_{u,i})←sim(i,j)_pro,r_{u,j}; //计算基于项目的预测评分
28. end for
29. P(r_{u,i})=λ*P_u(r_{u,i})+(1 – λ)*P_i(r_{u,i});
30. rankedP(r_{u,i})←rank(P(r_{u,i}));
```

31. Top–k_i ←chooseTop$k$(ranked$P(r_{u,i})$, k);
32. return Top–k_i。

end begin

算法2中，语句1～11行通过式（8）计算用户间相似性并得到用户相似性矩阵；语句12～22行通过式（5）计算项目间相似性并得到项目相似性矩阵；语句23～28行分别通过式（9）和（10）计算基于用户的和基于项目的预测评分；语句29行通过式（11）将基于用户的和基于项目的预测评分进行加权求和；语句30～32行对求和得到的预测评分进行排序，并将评分较高的前k个项目推荐给用户。

4 实验与分析

将本文方法与传统混合协同过滤方法及WP-CC方法[7]进行比较，并分析相关结果。

4.1 实验准备

所有实验和算法均通过Java实现，开发环境为Eclipse。所有实验运行在一台具有Intel Core i5-4590 CPU，4 GB内存，操作系统为Windows 7的PC上。

实验采用的是Movielens100k数据集[19]。该数据集是基于协同过滤的推荐系统中常用的数据集，本文提出的方法也是基于协同过滤的，所以选取它为实验数据集，以方便进行实验对比分析。Movielens100k数据集包括943个用户对1 682个项目的10万条评分，数据集中还包括用户个人信息、项目相关属性信息等。

4.2 评估指标

为了评估推荐算法的准确度，选择推荐系统中常用的平均绝对误差（mean absolute error，*MAE*）和归一化折损累积增益（normalized discounted cumulative gain，*NDCG*）作为评价标准[20]。其中，*MAE*反映预测值和真实值间的差距，*NDCG*反映对多个候选项目评分预测排序情况的优劣，计算方法如式（12）～（14）。

$$MAE=\frac{\sum_{i=1}^{n}|P_i-t_i|}{n} \tag{12}$$

式中，p_i为对候选项目i的预测评分，t_i为候选项目的实际评分，n为候选项目个数。式（12）表明，*MAE*值越小，预测结果越准确。

$$NDCG_n(u)=\frac{DCG_n(u)}{\max DCG_n(u)} \tag{13}$$

式中，$NDCG_n(u)$为用户u候选项目预测评分排序的$DCG_n(u)$值与实际情况下的max $DCG_n(u)$值之比，$DCG_n(u)$值由式(14)得到：

$$DCG_n(u)=r(u,i_1)+\sum_{k=2}^{n}\frac{r(u,i_k)}{\mathrm{lb}(k)} \tag{14}$$

式中，$r(u,i_k)$为所有候选项目根据预测评分从高到低排序中的第k个项目的实际评分。可以看出，*NDCG*值越大，表示对候选项目的预测评分越准确。

4.3 结果分析

首先，通过实验验证作者提出的矩阵填充方法的有效性；然后，通过实验确定式（11）中系数λ的最佳取值；通过*MAE*和*NDCG*两项指标对相关方法的推荐结果进行对比分析。

选取用户评分均值（URA）、项目评分均值（IRA）以及本文融合用户兴趣评分（UIR）的3种矩阵填充方法进行实验对比。实验时，从用户所有已评分项目中随机挑选项目并将其认定为评分缺失项，再通过3种方法进行填充，根据填充结果计算3种方法在不同用户规模下的*MAE*值，如表4所示。

表 4 填充数据*MAE*值对比

Tab. 4 Comparison of *MAE* values of filled data

	URA	IRA	UIR
U_{50}	1.046	1.039	0.948
U_{100}	1.001	1.038	0.860
U_{300}	0.961	0.956	0.856
U_{500}	0.959	0.977	0.882
U_{all}	0.994	1.001	0.897

从表4可以看出，使用作者提出的UIR方法填充数据的*MAE*值在不同用户规模下都比URA和IRA方法小，说明用户评分矩阵填充时考虑了用户对具体项目类型的兴趣能够达到更好的推荐效果。

分析参数λ的取值对本文IHRIRF算法的影响。将邻居用户数分别设置为10、20、30、40、50时，计算本文算法的*MAE*值，结果如图1所示。

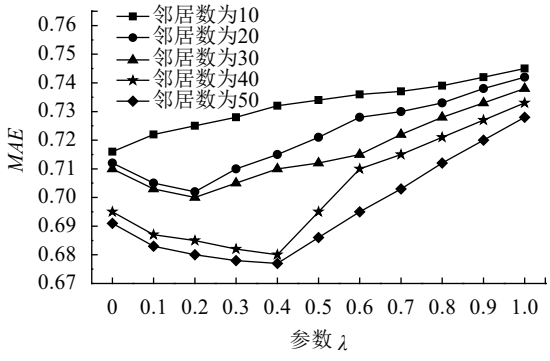


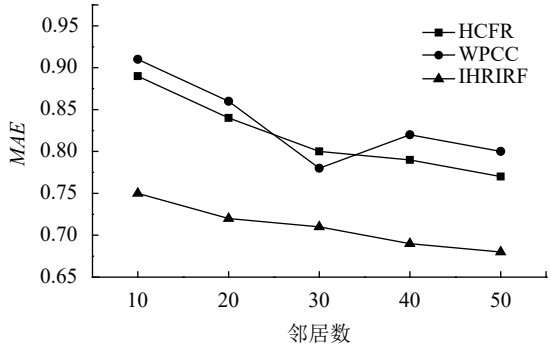
图 1 参数λ对IHRIRF算法的影响

Fig. 1 Impact of parameter λ on the IHRIRF algorithm

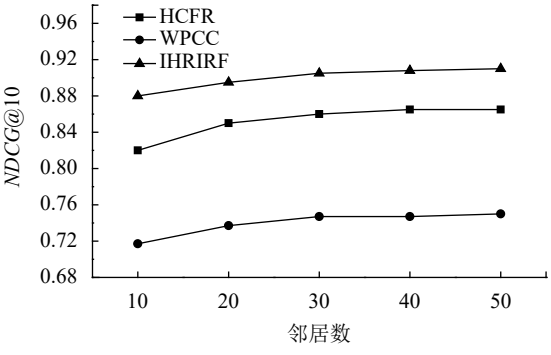
图1表明，在不同规模的邻居数下，本文IHRIRF算法的*MAE*值随着参数λ的增大出现不同的变化趋势。其中，邻居规模为20和30时，λ值为0.2时

*MAE*值最小；邻居规模为40和50时，λ值为0.4时*MAE*值最小。综合考虑两种情况，λ值取0.3时，本文算法能够接近最优值。因此，将参数λ的值设置为0.3，进行如下对比实验。

选取传统的混合协同过滤推荐方法HCFR（hybrid collaborative filtering recommendation）、考虑用户共同使用项目数的改进皮尔逊相关系数计算方法WPCC[7]和本文的IHRIRF方法进行比较，结果如图2所示。



(a) *MAE* 值对比



(b) *NDCG* 值对比

图 2　实验结果对比

Fig. 2　Comparison of experimental results

图2中，以邻居用户数为横坐标，分别以*MAE*和*NDCG*值为纵坐标，观察3种方法的*MAE*值和*NDCG*值随着邻居用户数增加而变化的情况，具体分析如下。

从图2(a)可以看出：3种方法的*MAE*值都随着邻居数的增加呈现降低的趋势。其中，HCFR和WPCC方法的MAE值基本相差不大，本文IHRIRF方法的MAE值最小。表明考虑评分差异化对计算相似度的影响能够达到更好的推荐效果，验证了本文方法的有效性。

从图2(b)可以看出：3种方法的*NDCG*值随着邻居数的增加波动很小，可见根据候选项目预测的评分排序结果没有因为邻居数的增加而有很大的变化。对比3种方法发现，本文方法得到的*NDCG*值比另外两种大，推荐效果更好，进一步验证了本文方法的有效性。

5　结　论

提出了一种基于用户兴趣评分填充的改进混合推荐方法。首先，通过分析用户对项目类型的偏好，计算用户兴趣评分并进行矩阵填充，缓解了数据稀疏的问题；然后，在进行相似性计算时考虑了用户主观评分差异化以及项目自身质量的影响，提出了一种改进皮尔逊相关系数的混合推荐方法。将本文IHRIRF方法和传统的混合协同过滤推荐方法HCFR及WPCC方法[7]进行了对比，结果表明本文的IHRIRF方法能够产生更好的推荐效果。

在今后的工作中，将尝试使用其他更大规模的数据集，以验证方法的通用性；同时，尝试将基于记忆的方法和基于模型的方法相结合，以进一步提高推荐的准确度。

参考文献：

[1] Ja Dongyan,Zhang Fuzhi.Collaborative filtering recommendation system based on double neighborhood selection strategy[J].Journal of Computer Research and Development,2013,50(5):1076–1084.[贾冬艳,张付志.基于双重邻居选取策略的协同过滤推荐系统[J].计算机研究与发展,2013,50(5):1076–1084.]

[2] 向小东,邱梓咸.基于slope one算法改进评分矩阵填充的协同过滤算法研究[J/OL].2019,36(5)[2018-03-09].http://www.arocmag.com/article/02-2019-05-031.html.

[3] Koren Y,Bell R.Advances in collaborative filtering[M].New York:Springer,2015:145–186.

[4] Huang Xianying,Xiong Liyuan,Li Qindong.Personalized news recommendation technology based on improved collaborative filtering algorithm[J].Journal of Sichuan University (Natural Science Edtion),2018,55(1):49–55.[黄贤英,熊李媛,李沁东.基于改进协同过滤算法的个性化新闻推荐技术[J].四川大学学报(自然科学版),2018,55(1):49–55.]

[5] Liu Hanqing,Zhu Min,Su Yabo,et al.A collaborative prediction model for user interest shift feature[J].Journal of Sichuan University (Natural Science Edtion),2016,53(3):548–554.[刘汉清,朱敏,苏亚博,等.一种考虑用户兴趣转移特征的协同预测模型[J].四川大学学报(自然科学版),2016,53(3):548–554.]

[6] Zang Xuefeng,Liu Tianqi,Sun Xiaoxin,et al.Collaborative filtering algorithm based on Bhattacharyya coefficient and item correlation[J].Computer Science,2017,44(12):52–57.[臧雪峰,刘天琦,孙小新,等.一种基于Bhattacharyya系数和项目相关性的协同过滤算法[J].计算机科学,2017,44(12):52–57.]

[7] Liu Haifeng,Hu Zheng,Mian A,et al.A new user similarity model to improve the accuracy of collaborative filtering[J]. Knowledge-Based Systems,2014,56:156–166.

[8] Wang Haiyan,Yang Wenbin,Wang Suichang,et al.A method of service recommendation based on trusted alliance[J]. Chinese Journal of Computers,2014,37(2):301–311.[王海艳,杨文彬,王随昌,等.基于可信联盟的服务推荐方法[J].计算机学报,2014,37(2):301–311.]

[9] Guo Guibing,Zhang Jie,Thalmann D.Merging trust in collaborative filtering to alleviate data sparsity and cold start[J]. Knowledge-Based Systems,2014,57:57–68.

[10] Tang Jie,Cai Yi,Li Qing,et al.Towards typicality-based collaborative filtering recommendation[J].IEEE Transactions on Knowledge and Data Engineering,2014,26(3):766–779.

[11] Yu Zhiwei,Wong K R,Chi Chihung.Efficient role mining for context-aware service recommendation using a high-performance cluster[J].IEEE Transactions on Services Computing,2017,10(6):914–926.

[12] Fan Bo,Cheng Jiujun.Collaborative filtering recommendation algorithm based on user's multi-similarity[J].Computer Science,2012,39(1):23–26.[范波,程久军.用户间多相似度协同过滤推荐算法[J].计算机科学,2012,39(1):23–26.]

[13] Wang Xiangde,Lei Yuxia,Yan Yushu.Research on singular value decomposition collaborative filtering algorithm based on matrix completion[J].Information Technology and Cyber Security,2017,36(19):55–61.[王祥德,雷玉霞,闫昱姝.基于矩阵填充的SVD协同过滤算法研究[J].信息技术与网络安全,2017,36(19):55–61.]

[14] Pan Taotao,Wen Feng,Liu Qinrang.Collaborative filtering recommendation algorithm based on rating matrix filling and item predictability[J].Acta Automatica Sinica,2017,43(9):1597–1606.[潘涛涛,文锋,刘勤让.基于矩阵填充和物品可预测性的协同过滤算法[J].自动化学报,2017,43(9):1597–1606.]

[15] Han Yanan,Cao Han,Liu Liangliang.Collaborative filtering recommendation algorithm based on scoring matrix filling and use interest[J].Compute Engineering,2016,42(1):36–40.[韩亚楠,曹菡,刘亮亮.基于评分矩阵填充与用户兴趣的协同过滤推荐算法[J].计算机工程,2016,42(1):36–40.]

[16] Yuan Weihua,Wang Hong,Du Xianghua.Collaborative filtering algorithm integrating non-negative matrix completion and subgroups partitioning[J].Journal of Chinese Computer Systems,2017,38(12):2645–2651.[袁卫华,王红,杜向华.结合非负矩阵填充及子集划分的协同推荐算法[J].小型微型计算机系统,2017,38(12):2645–2651.]

[17] Fang Chen,Zhang Hengwei,Zhang Ming,et al.Trust expansion and listwise learning-to-rank based service recommendation method[J].Journal on Communications,2018,39(1):147–158.[方晨,张恒巍,张铭,等.基于信任扩展和列表级排序学习的服务推荐方法[J].通信学报,2018,39(1):147–158.]

[18] Cheng Yanping,Wang Sai.A hybrid collaborative filtering algorithm based on user-item[J].Computer Technology and Development,2014,24(12):88–95.[陈彦萍,王赛.基于用户–项目的混合协同过滤算法[J].计算机技术与发展,2014,24(12):88–95.]

[19] Fernando O,Antonio H,Jesus B,et al.Recommending items to group of users using matrix factorization based collaborative filtering[J].Information Sciences,2016,345:313–324.

[20] Polatidis N,Georgiadis C K.A dynamic multi-level collaborative filtering method for improved recommendations[J]. Computer Standards & Interfaces,2017,51:14–21.

（编辑　张凌之）

引用格式：Li Zheng,Duan Lei.Improved hybrid recommendation approach based on user interest ratings filling[J].Advanced Engineering Sciences,2019,51(1):189–196.[李征,段垒.基于用户兴趣评分填充的改进混合推荐方法[J].工程科学与技术,2019,51(1):189–196.]