

文章编号: 1000-5862(2016)05-0481-06

一种基于混合相似度的用户多兴趣推荐算法

滕少华¹ 麦嘉俊¹ 张 巍¹ 赵淦森²

(1. 广东工业大学计算机学院 广东 广州 510006; 2. 华南师范大学计算机学院 广东 广州 510631)

摘要: 针对传统协同过滤推荐数据稀疏会影响推荐质量, 以及项目最近邻居集的计算忽略用户多兴趣及提高推荐的准确度问题, 该文采用混合模型改进了相似性度量计算. 综合 Pearson 相关系数与修正余弦相似性, 提出了一种基于混合相似度的用户多兴趣推荐算法. 实验表明: 该推荐方法的相似度计算更高效, 不仅提高推荐准确率, 而且使用户有更好的推荐体验.

关键词: 用户多兴趣; 推荐算法; 协同过滤; 混合相似度

中图分类号: TP 311 **文献标志码:** A **DOI:** 10.16357/j.cnki.issn1000-5862.2016.05.07

0 引言

个性化服务的出现是互联网快速发展的一个重要标志之一. 个性化推荐服务应用到各领域为人们在生活上都带来便利, 并深受人们认同. 个性化推荐是一种商家或企业通过收集用户行为或购买记录, 使用推荐算法计算得出满足用户需求的推荐结果, 最后以推送等方式反馈给用户的主动服务. 国外, 亚马逊电子商务网站最早使用推荐系统来向网站顾客主动推送推荐商品^[1]. Netflix 视频公司举行的推荐算法改进大赛、潘多拉音乐电台对客户进行精确主动推荐, 这些都极大推动了推荐系统研究领域的发展. 如今国内, 豆瓣网不仅仅为用户提供电影方面推荐, 还对音乐、图书也进行满足用户需求的推荐. 优酷网、土豆网、爱奇艺等影视类网站的个性化推荐技术也是越趋成熟. 近年新浪微博和微信的快速发展与广泛使用, 精准广告推荐、好友推荐、文章推荐和各种服务的推荐将成为个性化推荐研究的新方向.

协同过滤系统是在各行各业中应用日益广泛^[2]. Ringo、Tapestry、GroupLens 以及 Video Recommender 是协同过滤推荐系统广泛应用和推广的典型代表^[3]. 协同过滤推荐技术的发展, 能为商家带来更大的商业价值.

协同过滤推荐虽然应用广泛, 在不同领域给商家带来商业价值并给用户都带来便利, 但其仍然存在着许多不足之处. 如实际上用户的兴趣爱好各不相同, 但传统协同过滤推荐只是按照用户同一喜好来推荐, 忽略了用户兴趣多样性. 另外传统的协同过滤推荐方法会带了冷启动问题. 然而, 目前的大多数推荐算法只采用近邻模型来确定最近邻居项目, 并未考虑用户兴趣多样性的问题, 另外简单的相似度方法未能评分矩阵极端稀疏等问题, 使得推荐质量大幅降低. 因此, 考虑用户实际的推荐体验, 提高推荐准确率, 合理改善推荐效果成为一个重要的热点问题.

1 相关工作

协同过滤(collaborative filtering)采用相似近邻模型. 根据目标用户相似近邻来计算预测目标用户对测试物品的打分, 然后进行相应推荐. 基于近邻的推荐算法与基于模型的推荐算法是目前协同过滤推荐算法的两大类^[4-7].

实际上, 已有大量学者对合理的推荐方法以及可行的推荐机制做出研究. 刘东宁等^[8]提出一种角色协同方法, 用 E-CARGO 模型建模最终使推荐机制人性化. 马小军等^[9]提出一种基于内容的机器学习

收稿日期: 2016-01-12

基金项目: 国家自然科学基金(61402118), 广东省科技计划项目(2012B091000173, 2013B010401034, 2013B090200017, 2013B010401029), 广东省教育厅项目(ZYGX008), 广东省重点实验室开放基金(15zk0132)和广州市科技计划(2012J5100054, 2013J4500028, 2013J4100004, 201508010067)资助项目.

作者简介: 滕少华(1962-), 男, 江西南昌人, 教授, 博士, 主要从事协同计算、数据挖掘、网络安全、大数据和复杂网络下的协同建模的研究.

习推荐方法,并在相似度计算中使用区分度,从而提高推荐准确率.李国等^[10]提出一种非线性组合的协同过滤推荐算法,增加时间属性、非线性组合用户特征和项目属性.高灵渲等^[11]提出了改进聚类模式的推荐算法,建立模型来预测项目的所属类别,提高推荐的效率.庄永龙等^[12]提出了基于项目特征建模,引入权重因子来对相似性进行改进,该算法虽然提高了预测准确率,但在权重计算时只用 Pearson 相关系数度量,没有考虑共同评分项目的数量对相似性的影响.吴月萍等^[13]提出了新相似度计算方法,将特征相似度和协同相似度各占 50% 来组合,没有考虑项目相似性问题.

针对上述分析,本文提出了混合改进相似度计算的用户多兴趣推荐算法设计与分析.第 1 步线性综合 Pearson 相关系数和修正余弦相似度,用新的相似性度量方法求解目标项目的最近邻项目集,第 2 步在目标项目的最近邻项目集上计算用户最近邻居来进行协同过滤预测推荐,最终形成推荐.目的是考虑用户的实际推荐体验,改善推荐效果,同时提高推荐准确度.

2 用户多兴趣推荐算法建模

2.1 传统的协同过滤推荐算法

传统推荐算法基本思想:建立相似近邻模型,是根据对同一项目或产品具有相似兴趣的用户来找用户最近邻,再对项目进行预测和产生推荐列表.首先输入用户和项目数据,计算用户相似度,找出与目标用户的相似邻居,最后对目标项目进行预测并向用户产生推荐列表.这种算法主要分 3 个步骤:评分矩阵(数据输入)、形成邻居和产生推荐列表.

(i) 评分矩阵(数据输入).用户对项目的评分数据记录在一个评分矩阵上,如表 1 所示.该矩阵主要记录了用户集合与项目集合,以及用户对不同项目的打分情况.

表 1 用户-项目评分矩阵

User	Item					
	I_1	I_2	...	I_j	...	I_n
U_1	R_{11}	R_{12}	...	R_{1j}	...	R_{1n}
U_2	R_{21}	R_{22}	...	R_{2j}	...	R_{2n}
\vdots	\vdots	\vdots		\vdots		\vdots
U_i	R_{i1}	R_{i2}	...	R_{ij}	...	R_{in}
\vdots	\vdots	\vdots		\vdots		\vdots
U_m	R_{m1}	R_{m2}	...	R_{mj}	...	R_{mn}

(ii) 形成邻居.根据步骤(i)评分数据,使用相似度公式计算两用户之间的相似度,从计算结果寻找目标用户最近邻居.这步骤中的相似度公式选取,会间接影响最终推荐的质量.相关相似性度量 PC(Pearson Correlation)、余弦相似性度量 CC(Cosine Correlation)和修正的余弦相关性度量 ACC(Adjusted Cosine Correlation)这 3 种是常用的计算相似性度量的方法.

(iii) 产生推荐列表.根据步骤(ii)中形成的最近邻居集,计算用户对待测项目的预测评分,最后根据对项目预测分值的排序,可以得出目标用户的项目 Top-N 推荐集.设用户 u 和用户对应的项目集 R 对任意项目 t 的预测值计算为

$$prediction_{u,t} = \bar{R}_u + \sum_{i=1}^N (corr_{u,i}) \times (rating_{i,t} - \bar{R}_i) / \sum_{i=1}^N (corr_{u,i}), \quad (1)$$

其中 i 是最近邻居集中的用户, $corr_{u,i}$ 是用户 u 和用户 i 之间的相似度, $rating_{i,t}$ 是用户 i 对项目 t 的评分值, \bar{R}_u 是用户 u 对所有项目的平均分; \bar{R}_i 是用户 i 对所有项目的平均分.

Top-N 推荐集.将计算出的预测值 $prediction_{u,t}$ 按由大到小进行排序,选取前 N 个项目推荐给目标用户.

2.2 混合改进相似度的用户多兴趣推荐过程

目前的大多数推荐算法在计算项目相似性时,仅采用传统相似度量方式计算,忽略用户对不同项目有不同兴趣爱好的表现,从而导致相似性计算不够精确.同一用户对不同的项目表现出来的兴趣和评分数据是不一样的,所以在预测不同的项目时,需要用到的最近邻居用户也应该不一样.这说明了待测项目和与该待测项目相对应的最近邻居用户有关.为了提高预测的准确率,应该找出目标用户在待测项目上有一定相似性的最近邻居用户来进行计算.此外,只采用一种相似度去计算相似度,解决不了评分稀疏性问题,所以需要根据用户对项目评分情况来对相似性度量方法进行优化.本文提出的方法重点在于首先在用户-项目评分矩阵上,通过混合改进相似度来找出相似的项目最近邻居,根据项目最近邻居来计算其他用户与目标用户的相似性,得出用户最近邻居来预测目标用户对待评分项目对象的预测评分并产生推荐列表,最后按评分降序排列,选取前 N 个推荐对象推送给用户.该推荐过程如图 1 所示.

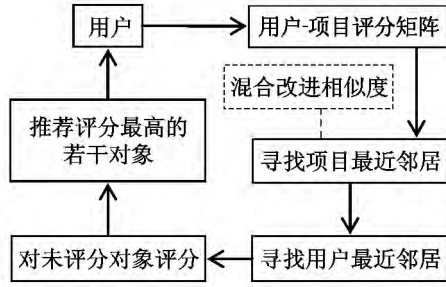


图1 混合改进相似度的用户多兴趣推荐过程

2.3 混合改进的相似性度量计算

2.3.1 常见相似度计算方法 目前,有很多相似度的计算方法,从简单的余弦相似度到复杂的 K 最近邻等算法。本文介绍最常用的3种: Pearson 相关系数、余弦相似度和修正余弦相似度^[14]。

Pearson 相关相似度 PC(Pearson Correlation): 主要反映2个用户之间线性关系,用 Pearson 相关系数计算得出。普通的余弦相似度 CC(Cosine Correlation) 主要通过计算用户在 n 维项目的向量的余弦夹角来评估用户 a 和用户 b 的相似度。修正余弦相似度 ACC(Adjusted Cosine Correlation) 能较好地确保用户邻居的相似性,普通余弦相似度 CC 没有考虑用户评分尺度问题。例如在评分范围1~5的电影数据集中,用户 a 认为3分就是满足其兴趣爱好的电影,而用户 b 则会认为5分电影才是他喜欢的。而 ACC 则是通过计算时减去用户的总体平均分来解决这个问题。

2.3.2 混合模型相似度计算 在项目邻居的相似度计算中,由于用户-项目评分矩阵稀疏性大,如果只简单采用了相关性系数来度量相关性,会大大降低推荐的准确率。为了提高推荐质量,本文提出混合模型相似度计算,采用权重参数来综合 Pearson 相关系数和修正余弦相似度,通过实验计算阈值来调整2种相似度的比重,并将此混合模型的相似度作为新的组合相似度计算方法。Pearson 相关系数是反映两项目的线性关系,不需要考虑用户行为,能客观计算出项目相似性。修正余弦相似度则考虑了用户主观评分尺度问题。因此,为了提高推荐精度,需要结合上述2种相似度来进行计算。

在混合模型相似度计算中,定义一个权重参数 λ 。权重参数 λ 用来调整属 Pearson 相关系数和修正余弦相似性。权重参数 λ 的计算公式为

$$\lambda = \begin{cases} UI/\beta, & \text{当 } UI < \beta \text{ 时,} \\ 1, & \text{当 } UI \geq \beta \text{ 时,} \end{cases} \quad (2)$$

其中 UI 表示两项目共同评分的用户个数,参数 β 为阈值。计算共同评分用户个数大于 β 时,使 $\lambda = 1$,此时表示只用修正余弦相似性来作为计算度量。

在计算项目相似度时,阈值 β 的取值会直接影响混合相似性度量的准确度,从而影响推荐效果。因此,在计算项目预测准确率的实验中,会以5为步长来计算,且需要通过多次实验来确定合适的一个阈值 β 。

混合 Pearson 相关系数 $sim_1(i, j)$ 与修正的余弦相似性 $sim_2(i, j)$ 使用权重参数 λ 用来动态调整两相似度的比重,最终得出混合相似性度量方法的计算公式为

$$sim(i, j) = (1 - \lambda) sim_1(i, j) + \lambda sim_2(i, j) = \begin{cases} (1 - UI/\beta) sim_1(i, j) + (UI/\beta) sim_2(i, j) & \text{当 } UI < \beta \text{ 时} \\ sim_2(i, j) & \text{当 } UI \geq \beta \text{ 时} \end{cases} \quad (3)$$

$$sim_1(i, j) = \sum_{c \in U_{ij}} (R_{ci} - \bar{R}_i) (R_{cj} - \bar{R}_j) / \left(\sqrt{\sum_{c \in U_{ij}} (R_{ci} - \bar{R}_i)^2} \sqrt{\sum_{c \in U_{ij}} (R_{cj} - \bar{R}_j)^2} \right), \quad (4)$$

其中 \bar{R}_i 和 \bar{R}_j 分别表示项目 i 、项目 j 的平均分。 R_{ci} 和 R_{cj} 分别表示用户 c 对项目 i 与项目 j 的评分。 U_{ij} 表示对项目 i 与项目 j 共同评分的用户集合。

$$sim_2(i, j) = \sum_{c \in U_{ij}} (R_{ci} - \bar{R}_c) (R_{cj} - \bar{R}_c) / \left(\sqrt{\sum_{c \in U_{ij}} (R_{ci} - \bar{R}_c)^2} \sqrt{\sum_{c \in U_{ij}} (R_{cj} - \bar{R}_c)^2} \right), \quad (5)$$

其中 \bar{R}_c 表示用户 c 对所有项目的平均分。

2.4 混合相似度的用户多兴趣推荐算法

首先,根据上述优化后的混合相似性度量通过计算两项目的相似性,找出相似的项目最近邻居,然后根据项目最近邻居来计算用户的相似性,寻找相似的用户最近邻居,其次根据用户最近邻居集在待评分项目上的评分来计算对待评分项目预测评分并产生推荐列表,最后按评分降序排列出前 N 个推荐项目对象,推荐给目标用户。其具体步骤如下。

(i) 计算找出最近邻居项目集合。对所有项目 $i \in I$ 采用优化的混合相似性度量方法来计算项目 i 的最近邻居项目,记作 SI_i 。此步骤相似度 $sim(i, j)$ 的计算方法使用上文提出的优化混合相似度方法来计算。按 $sim(i, j)$ 降序排列找出最近邻居项目集 SI_i 。

(ii) 计算找出最近邻居用户集合。 $usim(a, b)$ 表示计算用户 a 和用户 b 在项目集 SI_i 上的相似性, $usim(a, b) =$

$$\frac{\sum_{k \in CSI_i(a, b)} (R_{ak} - \bar{R}_a) (R_{bk} - \bar{R}_b)}{\sqrt{\sum_{k \in CSI_i(a, b)} (R_{ak} - \bar{R}_a)^2} \sqrt{\sum_{k \in CSI_i(a, b)} (R_{bk} - \bar{R}_b)^2}}, \quad (6)$$

其中 $CSI_i(a, b)$ 表示用户 a 和用户 b 在项目集 SI_i 中共同评分的项目子集。 \bar{R}_a 表示用户 a 对所有项目的

平均分; \bar{R}_b 表示用户 b 对所有项目的平均分.

将 $usim(a, b)$ 降序排列, 找出用户 a 在项目 i 上的最近邻居集 $Userset_{a, b}$.

(iii) 计算项目预测评分. 根据邻居用户在项目 i 上的评分, 预测用户 a 对项目 i 的评分 $P_{a, i}$,

$$P_{a, i} = \bar{R}_a + k \sum_{b \in Userset_{a, b}} usim(a, b) (R_{b, i} - \bar{R}_b), \quad (7)$$

其中 $\frac{1}{k} = \sum_{b \in Userset_{a, b}} usim(a, b)$.

(iv) 按由大到小对不同项目的预测评分进行排序, 选取前 N 个项目推荐给用户 a . 基于此步骤, 设计的混合相似度用户多兴趣推荐算法伪代码如下:

Input: User-Item 评分矩阵; 目标用户 a 和待测项目 i ;

Output: 推荐集;

(i) 计算项目最近邻居:

foreach j in Item 集合

用优化混合相似度方法来计算项目 i 与项目 j 之间的相似度 $sim(i, j)$;

if 当前相似项目的数量 $< SI_i$ 设定的大小

将 j 加入 SI_i

else if $sim(i, j) > SI_i$ 中的最小值

将 j 加入 SI_i

Sort(SI_i); // 降序排序;

(ii) 计算用户最近邻居:

foreach b in User 集合

if 目标用户 a 与 b 在 SI_i 上有共同评分项目

foreach k in $CSI_i(a, b)$

用修正余弦相似度 ACC 计算用户 a 和用户 b 的相似度 $usim(a, b)$;

if 当前相似用户的数量 $< Userset_{a, b}$ 设定的大小

将 b 加入 $Userset_{a, b}$

else if $usim(a, b) > Userset_{a, b}$ 中的最小值

将 b 加入 $Userset_{a, b}$

Sort($Userset_{a, b}$); // 降序排序;

(iii) 计算项目预测评分:

foreach 待测项目 i in 用户 a 的未评分项目集合
 $sumRate = 0$;

$sumSim = 0$;

$RateSet = \{ \}$;

foreach 用户 b in $Userset_{a, b}$

$sumRate += usim(a, b) * (R_{b, i} - \bar{R}_b)$

$sumSim += usim(a, b)$

$P_{a, i} = \bar{R}_a + sumRate / sumSim$;

将 $P_{a, i}$ 加入 $RateSet$;

(iv) 给出推荐集

Sort($RateSet$); // 降序排序

从中选取前 N 个项目推荐给用户 a .

2.5 推荐质量评价

(i) 推荐算法的精确度计算通常采用平均绝对偏差 MAE(mean absolute error). 文中采用其作为度量标准(值越小表示精度越高).

实际的用户评分集合用 q_N 表示, 而预测的用户评分集合用 p_N 表示. 平均绝对偏差 MAE 的计算公式为^[15]

$$MAE = \sum_{i=1}^N |p_i - q_i| / N, \quad (8)$$

其中 N 为测试集中评分记录个数.

(ii) 准确率(Precision). 准确率定义为预测准确的项目总数与预测的所有项目总数之比. 计算公式为

$$Precision = \sum_{u \in U} |R(u) \cap T(u)| / \sum_{u \in U} |R(u)|, \quad (9)$$

其中 $R(u)$ 为推荐系统对用户 u 在测试集中推荐的项目集合, $T(u)$ 为测试集中用户 u 喜欢的项目集合.

(iii) 召回率(Recall) 计算公式为

$$Recall = \sum_{u \in U} |R(u) \cap T(u)| / \sum_{u \in U} |T(u)|. \quad (10)$$

3 实验与结果分析

3.1 数据集

本文采用 MovieLens 网站提供的数据集来进行实验分析. 该数据集包含 1 682 部电影, 943 位用户, 记录了全部用户对电影 100 000 条评分. 记录评分分值是 1 ~ 5 之间的整数. 本文的实验数据集是从其原始数据集中抽取 471 100 条电影评分记录, 包含 350 位用户分别对 1 346 部电影的评分.

3.2 实验结果与分析

3.2.1 比较与分析 3 种相似性度量方法 为了提高推荐的精度, 选择合适的相似性度量方法来进行线性组合. 本实验需要比较分析这 3 种相似性度量方法——相关相似性 PC、余弦相似性 CC 和修正的余弦相似性 ACC. 最近邻个数 K 依次从 5 逐步递增至 50, 步长为 5. 实验比较结果如图 2 所示.

通过图 2 比较实验可知, 在最近邻个数 K 依次递增的情况下, 修正的余弦相似性度量方法的 MAE 值比其它 2 种相似性度量方法要低, 故表明其推荐质量更高. 因此选取修正的余弦相似性作为本文提出的混合相似度计算的一种相似性度量来研究.

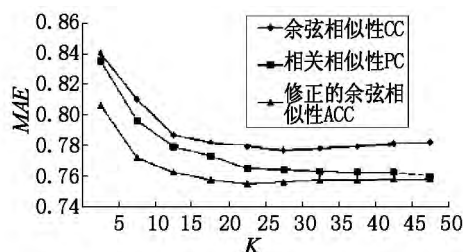
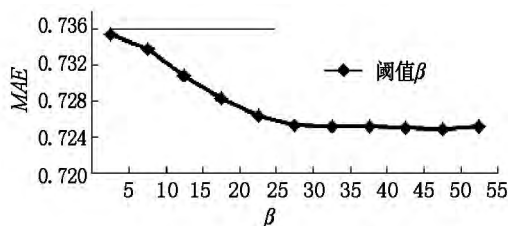


图2 PC、CC、ACC 的比较

3.2.2 阈值 β 的确定 权重参数 λ 需要通过阈值 β 来确定.在计算项目之间的相似度时 β 的值会直接影响推荐的最终得出的项目相似度.由于不同 β 值会间接影响了推荐的质量,因此,需要进行多次实验来为阈值 β 确定一个合适的值,实验如图3所示.

图3 阈值 β 的确定

通过图3实验结果可知 β 的值介于30~55时,MAE值趋向稳定.而当 $\beta=50$ 时,MAE值最小,说明阈值 β 的值定为50能使混合模型推荐算法有较高的推荐准确率.

3.2.3 比较与分析推荐算法 将本文提出的推荐算法,分别和传统协同过滤算法、用户多兴趣混合模型推荐算法进行MAE值比较.实验结果如图4所示.

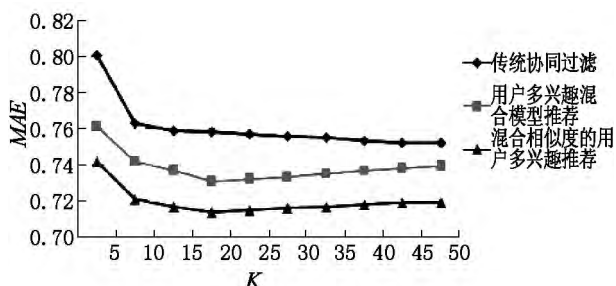


图4 混合相似度用户推荐算法的比较

从图4可以得出,本文提出的混合改进相似度用户多兴趣推荐算法,在最近邻个数 K 的依次递增情况下,其MAE值比其他2种对比算法低.故实验说明本文提出的推荐算法,推荐精度有明显的提高.准确率的实验如图5所示,召回率的实验如图6所示.

从图5和图6可以看出,本文提出的混合相似度的用户多兴趣推荐算法相比于其他2种推荐算法,有更高的准确率和召回率.

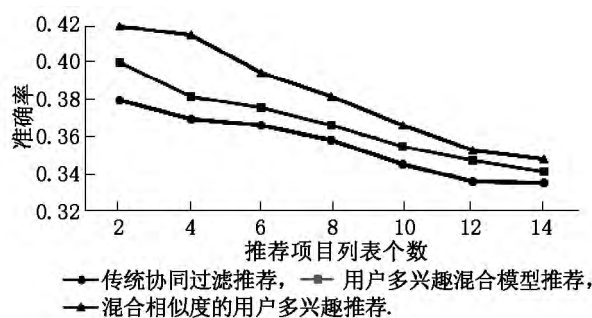


图5 准确率比较

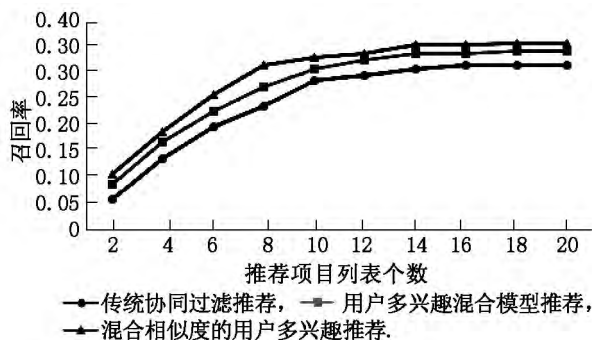


图6 召回率比较

4 结束语

在协同过滤基础上分析研究了PC、CC、ACC这3种相似性度量方法,本文提出了混合相似度计算的用户多兴趣推荐算法,不仅使相似度计算更高效,并且提高了推荐准确度,为用户提供更好的推荐体验.该算法对相似度进行混合改进,使用权重参数综合Pearson相关系数和修正余弦相似度,用这一新的相似度计算方法来求解目标项目的相似项目集,提高相似度计算精度.通过MovieLens公开数据集实验表明,该算法相比于其它2种推荐算法有着更好的推荐效果.

接下来的工作将对冷启动问题进行研究.综合考虑新项目的搜索热度和项目类别的新颖性等指标,通过数据挖掘对其进行更合理有效的推荐.

5 参考文献

- [1] Sarwar B, Karypis G, Konstan et al. Item-based collaborative filtering recommendation algorithms [C]. New York: ACM, 2001: 285-295.
- [2] 刘建国,周涛,汪秉宏.个性化推荐系统的研究进展[J].自然科学进展,2009,19(1):1-15.
- [3] 黄创光,印鉴,汪静,等.不确定近邻的协同过滤推荐算法[J].计算机学报,2010,33(8):1369-1377.

- [4] Liu Qi ,Chen Enhong ,Xiong Hui ,et al. Enhancing collaborative filtering by user interests expansion via personalized ranking [J]. IEEE Trans on Systems ,Man and Cybernetics-B 2012 ,42(1) : 218-233.
- [5] 邓爱林 ,朱扬勇 ,施伯乐. 基于项目评分预测的协同过滤推荐算法 [J]. 软件学报 2003 ,4(9) : 1621-1628.
- [6] Zhao Zhidan ,Shang Mingsheng. User-based collaborative-filtering recommendation algorithms on hadoop [EB/OL]. [2015-12-17]. https://www.researchgate.net/publication/221306166_User-Based_Collaborative-Filtering_Recommendation_Algorithms_on_Hadoop.
- [7] 孙光福 ,吴乐 ,刘淇 ,等. 基于时序行为的协同过滤推荐算法 [J]. 软件学报 2013 ,24(11) : 2721-2733.
- [8] 刘冬宁 ,刘艳 ,滕少华 ,等. 基于角色协同的在线社交网络好友推荐机制 [J]. 广西大学学报: 自然科学版 , 2014 ,39(6) : 1316-1323.
- [9] 马小军 ,赵伟. 改进相似度的分布式个性化推荐 [J]. 计算机工程与应用 2014 ,50(4) : 126-131.
- [10] 李国 ,张智斌 ,刘芳先 ,等. 非线性组合的协同过滤推荐算法 [J]. 计算机应用 2011(11) : 3063-3067.
- [11] 高灵渲 ,张巍 ,霍颖翔 ,等. 改进的聚类模式过滤推荐算法 [J]. 江西师范大学学报: 自然科学版 2012 ,36(1) : 106-110.
- [12] 庄永龙. 基于项目特征模型的协同过滤推荐算法 [J]. 计算机应用与软件 2009 ,26(5) : 244-246.
- [13] 吴月萍 ,郑建国. 协同过滤推荐算法 [J]. 计算机工程与设计 2011 ,32(9) : 3019-3021.
- [14] Teng Shaohua ,Li Junlei ,Li Rigui ,et al. The calculation of similarity and its application in data mining [J]. Lecture Notes in Computer Science 2014 ,8351: 563-574.
- [15] 许智宏 ,王宝莹. 基于项目综合相似度的协同过滤算法 [J]. 计算机应用研究 2014 ,31(2) : 398-400.

User Multi-Faced Interests Recommendation Algorithm Based on Hybrid Similarity

TENG Shaohua¹ ,MAI Jiajun¹ ,ZHANG Wei¹ ,ZHAO Gansen²

(1. School of Computer Science and Technology ,Guangdong University of Technology ,Guangdong Guangzhou 510006 ,China;

2. School of Computer ,South China Normal University ,Guangdong Guangzhou 510631 ,China)

Abstract: The traditional collaborative filtering recommendation's sparse data will affect the quality ,and it fails to take into account the user multi-faced interests to determine the projects nearest neighbor set. Coupling with the traditional similarity measure method without considering user's behavior ,leads to lower quality of the recommendation. In order to improve the recommendation accuracy ,the hybrid model improved similarity measure calculated by Pearson correlation linear combination of adjusted cosine correlation has been used ,and then an user multi-faced interests recommendation algorithm of hybrid similarity computing is proposed in the paper. The experimental results show that the similarity calculation of recommendation method is more efficient ,improve the accuracy of recommendation ,and make the better recommendation of user experience.

Key words: user multi-faced interests; recommendation algorithm; collaborative filtering; hybrid similarity computing

(责任编辑: 冉小晓)