

基于混合模型推荐算法的优化

李鹏飞 吴为民

(北京交通大学计算机与信息技术学院计算机工程系 北京 100044)

摘 要 现代电子商务系统用户和物品数目的日益增加使得 User-Item 矩阵变得越来越稀疏,再加上目前相似性度量方法均存在一定弊端,致使推荐系统的推荐质量降低了。针对传统混合模型推荐算法做了优化,其相似性度量方法由物品属性相似性和改进的修正余弦相似性线性组合而成,权重因子自动生成,考虑了用户评分尺度及用户活跃度对物品相似性的影响。为解决冷启动问题,使用用户基本信息获得用户间的相似度,各属性权重因子由 SVDFeature 计算得到。实验结果表明,该算法有效地提升了推荐系统的推荐质量,同时还有效解决了用户冷启动与物品冷启动问题。

关键词 协同过滤,相似度,混合模型,权重因子,冷启动

中图法分类号 TP181 文献标识码 A

Optimized Implementation of Hybrid Recommendation Algorithm

LI Peng-fei WU Wei-min

(Department of Computer Engineering, School of Computer, Beijing Jiaotong University, Beijing 100044, China)

Abstract The ever-increasing number of users and items of modern electronic commercial system has made the user-item matrix to become more and more sparse. This situation, in combination with somewhat inappropriate similarity calculation methods currently used, makes the recommendation quality of recommender system to gradually reduce. For this, we presented an optimized recommender algorithm which is based on a hybrid model. In our algorithm, the similarity function is a linear combination of the item property similarity and a modified correlation cosine similarity. The weighting factor, which is generated automatically, is related to the number of users who rated both items. The modification to the correlation cosine similarity measure considers both the rating tendency and the activity from users. To deal with the cold start problem, we also acquired user similarity through user property information with weighting factors computed by SVDFeature. The experimental results demonstrate that our algorithm effectively improves the recommendation quality and alleviates cold starting problem resulting from both users and items.

Keywords Collaborative filtering, Similarity, Hybrid recommendation, Weighting factor, Cold start

1 引言

随着互联网信息的爆炸性增长,电子商务网站中用户和物品数目急剧增加,如何在短时间内找到用户感兴趣的物品已成为人们关注的焦点,因此个性化推荐系统在电子商务、基于 social 的社会化站点、作为互联网经济支柱的在线广告和新颖的在线应用推荐等领域都发挥着越来越重要的作用。

目前,已经存在许多较为成熟的个性化推荐技术,如基于人口统计学的过滤、内容过滤、协同过滤^[1,2]等。然而,这些推荐算法均存在一定弊端,基于人口统计学的过滤单纯地使用用户的基本信息对用户进行分类,过于粗糙;内容过滤忽略了人对物品的态度,并有用户冷启动问题;协同过滤是应用最为广泛的推荐技术之一。然而不管是基于用户的协同过滤(User-Based CF)^[3],还是基于物品的协同过滤(Item-Based CF)^[4],它们依然存在数据稀疏性问题和冷启动等问题,推荐质量仍然较低。

为提高推荐质量,之前有大量学者做了相关研究。Dempster^[5]和 Thiesson 等人^[6]提出了聚类技术,即将具有相似兴趣爱好的用户分配到相同的簇中,聚类产生之后,根据簇中其他用户对物品的评价预测目标用户对该物品的评价,但此方法对数据集中的噪声数据较为敏感,且没有解决冷启动问题。Sarwar 等人^[7]提出了基于奇异值分解(SVD)的办法来降低项目空间的维度,但维度降低的同时导致了信息的损失,效果难以保证。Peng 等人^[8]采用了内容过滤与 Item-Based 协同过滤相结合的混合模型推荐算法,将物品属性相似性与物品评分相似性进行线性组合,但其平衡因子是静态的,且其只考虑了相同属性的个数,而忽略了其他属性的影响。Wu 等人^[9]在此基础上虽将平衡因子变为动态的,但平衡因子所代表的意义并不大,不能准确突出物品属性相似性与评分相似性的轻重,且其没有解决用户冷启动问题。

本文在传统的混合模型推荐算法基础上,对修正余弦相似性进行了改进,再与物品属性相似性动态结合,提出了一种

到稿日期:2013-05-20 返修日期:2013-07-26

李鹏飞(1988—),男,硕士生,主要研究方向为数据挖掘、信息检索、推荐系统,E-mail:11120454@bjtu.edu.cn;吴为民(1966—),男,副教授,主要研究方向为信息检索、形式验证。

新的相似性度量方法,同时解决了冷启动问题。本文第2节详细介绍本文对传统混合模型推荐算法的优化;第3节主要描述在 MovieLens 数据集上的实验结果并进行分析;最后总结全文,介绍下一步工作。

2 基于混合模型推荐算法的优化

本节介绍了对传统 Item-Based 协同过滤算法中修正余弦相似性的改进办法、物品属性相似性的计算方法、权重因子的动态调整、用户冷启动问题的解决方案以及产生推荐的过程。

2.1 Item-Based 协同过滤中修正余弦相似性的改进

Item-Based 协同过滤算法中,3种传统的 item 相似性度量方法^[10]分别为余弦相似性、相关相似性和修正余弦相似性,实验结果表明修正余弦相似性在 MovieLens 数据集上能够获得最佳效果,故本次实验在修正余弦相似性的基础上,通过加入主观评分归一化^[11]以及用户活跃度对物品相似性的影响对其进行改进。

2.1.1 主观评分归一化

不同用户的评分习惯不尽相同,有些用户较为宽松,评分较高,而有些用户相对严格,评分略低,故不同的评分尺度对物品相似度的计算影响较大,修正余弦相似性虽然考虑了此问题,但仍然略显粗糙,不够精确。本文引入了主观评分归一化算法^[11],通过算法变换,将各用户的平均评分统一,把所有评分转换到同一价值坐标系,得到新的 User-Item 评分矩阵,消除了这种因用户评分尺度不一致而造成的误差。

具体算法过程如下:

算法1 主观评分归一化

输入:用户-物品评分矩阵 $A(m,n)$

输出:经归一化算法变换后的用户-物品评分矩阵 $A'(m,n)$

Step1 根据输入的评分矩阵,计算每个用户 u 的平均评分 \bar{R}_u ,计算公式如下:

$$\bar{R}_u = \frac{\sum_{i=1}^n R_{ui}}{n} \quad (1)$$

其中, R_{ui} 表示用户 u 对物品 i 的评分, n 为用户 u 评价物品的个数。

Step2 重新设定用户评分区间,将之定为:

$$\left[\frac{\sum_{u=1}^m \min(R_{u1}, R_{u2}, \dots, R_{un})}{m}, \frac{\sum_{u=1}^m \max(R_{u1}, R_{u2}, \dots, R_{un})}{m} \right] \quad (2)$$

此区间的平均值为 k ,

$$k = \frac{\sum_{u=1}^m \min(R_{u1}, R_{u2}, \dots, R_{un}) + \sum_{u=1}^m \max(R_{u1}, R_{u2}, \dots, R_{un})}{2m} \quad (3)$$

Step3 重新计算用户 u 对物品 i 的评分 R_{ui}' ,得到新的用户-物品评分矩阵 $A'(m,n)$:

$$R_{ui}' = \frac{k * R_{ui}}{\bar{R}_u} \quad (4)$$

其中, $\frac{k}{\bar{R}_u}$ 称为归一化系数。

由式(4)可知,经上述算法变换后的用户 u 的平均评分 \bar{R}_u' 为:

$$\bar{R}_u' = \frac{\sum_{i=1}^n R_{ui}'}{n} = \frac{\sum_{i=1}^n \frac{k * R_{ui}}{\bar{R}_u}}{n} = \frac{k}{\bar{R}_u} * \left[\frac{\sum_{i=1}^n R_{ui}}{n} \right] = \frac{k}{\bar{R}_u} * \bar{R}_u = k \quad (5)$$

其中, n 为用户 u 评价物品的个数。

通过此变换可使 $\bar{R}_1' = \bar{R}_2' = \dots = \bar{R}_m' = k$,用户的价值坐标系因此得到了统一,使物品的得分既能体现该物品在每位用户心目中的位置,又能在相同的价值坐标系下进行计算,有效降低了用户评分尺度对物品相似度计算产生的影响。

2.1.2 用户活跃度对物品相似性的影响

在协同过滤中,两物品产生相似度是因为它们共同出现在很多用户的评分列表中,每个用户的评分列表都对物品相似度产生贡献,活跃的用户对很多物品都有评分,而不活跃的用户可能只对少数物品有所评价。JohnS. Breese 在文献^[12]中提出一种称为 Item CF-IUF 的算法,IUF (Inverse User Frequency)表示用户活跃度对数的倒数,据此,我们认为活跃用户对物品相似度的贡献应该小于不活跃的用户,用 θ_u 表示用户 u 的活跃度对数的倒数,则 θ_u 的计算公式为:

$$\theta_u = \frac{1}{\log(1 + |N(u)|)} \quad (6)$$

其中, $N(u)$ 表示用户 u 评分的物品的集合, $|N(u)|$ 表示用户 u 评分的物品的数量,可以间接反映出用户的活跃度。

2.1.3 修正余弦相似性的改进

传统的修正余弦相似性计算公式如下所示:

$$\text{sim}(i, j) = \frac{\sum_{u \in U} (R_{ui} - \bar{R}_u)(R_{uj} - \bar{R}_u)}{\sqrt{\sum_{u \in U} (R_{ui} - \bar{R}_u)^2} \times \sqrt{\sum_{u \in U} (R_{uj} - \bar{R}_u)^2}} \quad (7)$$

其中, U 表示对物品 i 和 j 共同评分的用户集合。

根据式(4)与式(6)可知,加入主观评分归一化与用户活跃度对物品相似度的影响之后得到的改进的修正余弦相似性计算方法为:

$$\text{sim}_1(i, j) = \frac{\sum_{u \in U} \theta_u * (R_{ui}' - \bar{R}_u') * (R_{uj}' - \bar{R}_u')}{\sqrt{\sum_{u \in U} (R_{ui}' - \bar{R}_u')^2} \times \sqrt{\sum_{u \in U} (R_{uj}' - \bar{R}_u')^2}} = \frac{\sum_{u \in U} \left[\frac{1}{\log(1 + |N(u)|)} \right] * (R_{ui}' - \bar{R}_u') * (R_{uj}' - \bar{R}_u')}{\sqrt{\sum_{u \in U} (R_{ui}' - \bar{R}_u')^2} \times \sqrt{\sum_{u \in U} (R_{uj}' - \bar{R}_u')^2}} \quad (8)$$

在后面的实验结果中可以看到,改进后的相似度计算方法较普通的修正余弦相似性方法能够获得更高的精确度,推荐质量得到了提升。

2.2 物品属性相似性的计算

根据内容过滤可知,两物品的相似性由其自身的属性信息决定,通过分析物品属性信息,我们可以得到物品-属性矩阵 A 。假设数据集中 Item 个数为 n ,每个 Item 的属性个数为 k ,则 A 的表示如表1所列。

表1 物品-属性矩阵 A

	Attr ₁	...	Attr _i	...	Attr _k
Item ₁	0	...	1	...	0
⋮	⋮		⋮		⋮
Item _i	1	...	0	...	1
⋮	⋮		⋮		⋮
Item _n	0	...	1	...	0

其中,1表示物品拥有此属性,0则表示不拥有此属性。

为了计算物品 i 和物品 j 的物品属性相似性,我们给出了如下公式:

$$\text{sim}_2(i, j) = \frac{|C_i \cap C_j|}{|C_i \cup C_j|} + \omega \quad (9)$$

其中, C_i 和 C_j 分别表示物品 i 和 j 的属性集合, ω 是为了防止大多数物品之间物品属性相似性为0,在本实验中,其值定

为 0.001。

2.3 混合模型相似性度量方法的优化

为了提高推荐质量并且能够解决物品冷启动问题,我们将物品属性相似性与改进的修正余弦相似性进行线性组合,得到最终的相似性度量方法。属性相似性可以认为是物品间的客观相似性,与用户行为无关,而改进的修正余弦相似性则是主观相似性,根据文献[9]可知,协同过滤的推荐效果一般要优于内容过滤,因此,在设定线性组合系数时,我们主要考虑改进的修正余弦相似性的精确程度。

修正余弦相似性最主要的缺点是由于评分数据的极端稀疏性,对两物品 i 和 j 共同评分的用户集合可能很小,而两物品只有在较多的用户评分比较相似时,才认为它们之间的相似度确实较高,如果只有一两个用户对物品 i 和 j 共同评分,即使算出的相似度较高,其可信度也是较低的,改进的修正余弦相似性依旧存在这个问题。

由此可知,权重因子 λ 与两物品共同评分的用户个数息息相关,当共同评分用户个数较多时,改进的修正余弦相似性所占的比重就较大,反之则不然。权重因子的计算公式如下所示:

$$\lambda = \begin{cases} \frac{|N_i \cap N_j|}{a}, & \text{当 } |N_i \cap N_j| < a \text{ 时} \\ 1, & \text{当 } |N_i \cap N_j| \geq a \text{ 时} \end{cases} \quad (10)$$

其中, $|N_i \cap N_j|$ 表示同时对物品 i 和物品 j 评分的用户个数,参数 a 为阈值,当共同评分用户个数超过 a 时,直接令 $\lambda=1$,阈值 a 的取值范围为: $[\min(|N_i \cap N_j|), \max(|N_i \cap N_j|)]$,其值需要经过多次实验来获得。

我们使用权重因子 λ 将物品属性相似性以及改进的修正余弦相似性进行线性组合,物品 i 和物品 j 的属性相似性用 $\text{sim}_1(i, j)$ 表示,改进的修正余弦相似性用 $\text{sim}_2(i, j)$ 表示,则混合模型相似性度量方法的计算公式为:

$$\text{sim}(i, j) = (1 - \lambda) * \text{sim}_1(i, j) + \lambda * \text{sim}_2(i, j) =$$

$$\begin{cases} \left(\frac{a - |N_i \cap N_j|}{a} \right) * \text{sim}_1(i, j) + \left(\frac{|N_i \cap N_j|}{a} \right) * \text{sim}_2(i, j), & \text{当 } |N_i \cap N_j| < a \text{ 时} \\ \text{sim}_2(i, j), & \text{当 } |N_i \cap N_j| \geq a \text{ 时} \end{cases} \quad (11)$$

实验结果表明,通过此方法得到的相似性度量方法,不仅解决了物品冷启动问题,而且使得计算物品间的相似性更加精确,有效提高了推荐质量。

2.4 用户冷启动问题的解决

如今,各大电子商务网站每天新注册的用户成百上千,这些新用户并未对任何物品产生过评价,单纯使用上述方法无法进行预测,当出现冷启动用户时,本文利用用户的特征属性计算用户的最近邻居集合,并根据最近邻居集合中其他用户对此物品的评分,来预测目标用户对其的评分。

假设数据集中用户的特征属性个数为 n ,则用户 u 的特征属性集合为 $C_u = \{c_{u1}, c_{u2}, \dots, c_{un}\}$,其中 c_{ui} 表示用户 u 的第 i 个特征属性,每个特征属性都有自己的值,对于分类属性,例如:性别分为男和女,我们令男=0,女=1;对于数值属性,例如:年龄可按 10 年一个时间段来划分,令 0~10 岁值为 0,11~20 岁值为 1,以此类推,其他特征属性也按此方法来赋值。如果用户 u 和用户 v 的第 i 个特征属性值相同,即 $c_{ui} = c_{vi}$,则我们令 $|c_{ui} \cap c_{vi}| = 1$,否则 $|c_{ui} \cap c_{vi}| = 0$ 。则用户 u 与

用户 v 的相似度计算公式为:

$$\text{sim}(u, v) = \alpha * |c_{u1} \cap c_{v1}| + \beta * |c_{u2} \cap c_{v2}| + \dots + \delta * |c_{un} \cap c_{vn}| \quad (12)$$

其中 α, β 与 δ 为权重因子,且有 $\alpha + \beta + \dots + \delta = 1$ 。我们使用 SVDFeature^[13-15] 工具,通过逐轮流添加特征的方式,计算出各个特征对预测结果的影响程度,并据此计算出各权重因子所占的比例。

2.5 产生推荐

通过本文提出的相似性度量方法,根据式(11),可离线获得物品间的相似度 $\text{sim}(i, j)$ 及最近邻居集合,根据式(12),可离线获得用户间的相似度 $\text{sim}(u, v)$ 及最近邻居集合。产生推荐的过程如下所示:

Step1 判断目标用户是否为冷启动用户,若是,跳转到 Step2,否则,跳转到 Step3。

Step2 冷启动用户 u 对物品 i 的评分预测公式为:

$$p_{ui} = \frac{\sum_{v \in N_u} \text{sim}(u, v) * R_{vi}}{\sum_{v \in N_u} |\text{sim}(u, v)|} \quad (13)$$

其中, N_u 为用户 u 的最近邻居集合。

计算完成后结束。

Step3 判断目标 Item 是否为冷启动物品,若是,跳转到 Step4,否则跳转到 Step5。

Step4 用户 u 对冷启动物品 i 的评分预测公式为:

$$p_{ui} = \frac{\sum_{j \in M_i} \text{sim}(i, j) * R_{uj}}{\sum_{j \in M_i} |\text{sim}(i, j)|} \quad (14)$$

其中, M_i 表示物品 i 的 K 最近邻居集合。

计算完成后结束。

Step5 用户 u 对物品 i 的评分预测公式为:

$$P_{ui} = \bar{R}_i + \frac{\sum_{j \in M_i} \text{sim}(i, j) * (R_{uj} - \bar{R}_j)}{\sum_{j \in M_i} |\text{sim}(i, j)|} \quad (15)$$

其中, \bar{R}_i 和 \bar{R}_j 分别代表物品 i 和物品 j 上所有用户打分的平均值。

计算完成后结束。

3 实验结果及分析

3.1 数据集

本文使用 MovieLens 站点提供的数据集 (<http://movielens.umn.edu>),该站点是一个基于 Web 的研究型推荐系统,用于接收用户对电影的评分并提供相应的电影推荐列表。数据集中包含了 943 个用户对 1682 部电影的 100000 条评分记录,每个用户至少对 20 部以上的电影进行了评分,评分范围为 1~5,数值越高表示用户对电影的偏爱程度越高。我们将数据集按照 80% 和 20% 的比例划分成训练集和测试集。我们选择的电影数据集的稀疏等级^[10]为 $1 - \frac{100000}{943 \times 1682} = 0.9369$ 。

每部电影都有其所属的类别,我们将这 19 种类别作为电影的属性,用于获得电影的属性相似度,数据集中还提供了用户的性别、年龄、职业、邮编等特征属性信息。另外,本实验在训练集中分别删除了 10 个物品和 10 个用户的所有用户评分数据,将删除后的数据集作为新的测试集,分别测试本算法解决物品冷启动和用户冷启动的能力。

3.2 度量标准

推荐系统中常用的评估推荐系统推荐质量的度量标准主要有平均绝对偏差 MAE^[7] 和均方根误差 RMSE。本文采用平均绝对偏差 MAE 作为度量标准,它主要通过计算预测的用户评分与实际的用户评分之间的偏差度量预测的准确性,MAE 越小,推荐质量越高。

假设测试集中实际的用户评分集合为 $\{q_1, q_2, \dots, q_N\}$, 预测的用户评分集合表示为 $\{p_1, p_2, \dots, p_N\}$, 则平均绝对偏差 MAE 的计算公式为:

$$MAE = \frac{\sum_{i=1}^N |p_i - q_i|}{N} \quad (16)$$

其中, N 为测试集中评分记录的个数。

3.3 实验结果

3.3.1 3种传统相似性度量方法的比较

本实验对余弦相似性、相关相似性和修正的余弦相似性的计算精确度进行比较,选取具有最佳推荐质量的相似性度量方法进行改进。最近邻居集合的个数 K 从 5 递增到 50, 每次增加 5, 实验结果如图 1 所示。

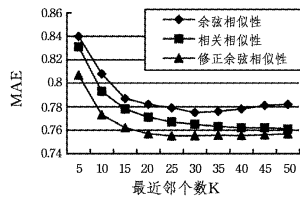


图 1 3种传统相似性度量方法的比较

由图 1 可见,修正的余弦相似性度量方法在推荐质量上确实要优于其它两种相似性计算方法,故而,本实验选取修正的余弦相似性度量方法作为研究基础。

3.3.2 修正余弦相似性度量方法改进后的效果

为证实改进的修正余弦相似性较改进前性能确实有所提升,本实验将二者的推荐质量进行了对比,具体如图 2 所示。

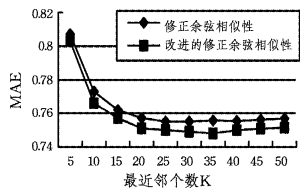


图 2 改进前后修正余弦相似性的比较

由图 2 可见,改进后的相似性度量方法在评分预测精度上确实有所提升,有效提高了推荐系统的推荐质量。

3.3.3 阈值 a 的获取

由式(10)和式(11)可知,阈值 a 的变化直接影响到物品间相似度的精确程度,间接决定了推荐系统的推荐质量,因此,我们需要为 a 设定一个合适的值,如图 3 所示。

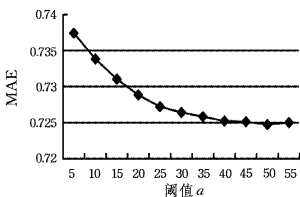


图 3 阈值 a 的选取

通过图 3 可知,当 $a = 50$ 时,混合推荐算法能够取得最佳的推荐效果,因此,将阈值 a 的值定为 50。

3.3.4 用户相似度权重因子的测定

根据 MovieLens 数据集中提供的用户特征属性信息,提取各个特征,并将其组织成 SVDFeature 所需的输入格式,本实验采用 SVDFeature 中的随机排序输入方式,采用的损失函数公式为:

$$Loss = (r - r')^2 + regularization \quad (17)$$

其中, r 为数据集中原始评分, r' 为预测评分, $regularization$ 为正则化参数。

将提取出的各个特征属性轮流加入到 SVDFeature 的 Group 特征集中,通过比较最终预测结果判定各特征属性的权重。结果如图 4 所示。

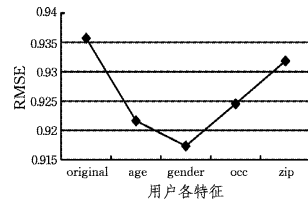


图 4 SVDFeature 对用户特征参数的测定

从图 4 可以看出,性别特征对预测结果的影响最大,年龄次之,邮编影响最小,根据上述影响的比例,分别将性别权重因子 α 定为 0.4,年龄权重因子 β 定为 0.3,职业权重因子 γ 定为 0.2,邮编权重因子 δ 定为 0.1。将这些权重因子代入式(12),便得到用户相似度的计算公式。

3.3.5 算法解决冷启动问题的能力

经本文优化后的混合模型推荐算法既可解决物品冷启动问题,又能解决用户冷启动问题,我们分别在训练集中提取 10 个物品的所有用户评分记录和 10 个用户的所有用户评分记录,并分别将它们当作新的测试集,以验证本算法解决冷启动问题的能力。传统混合模型推荐算法中冷启动用户对物品评分和用户对冷启动物品的评分均设定为 3 分。测试结果如图 5 所示。

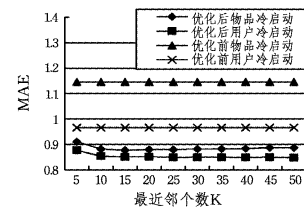


图 5 解决冷启动问题的能力

由此可知,本文优化后的混合模型推荐算法确实能够有效地解决物品冷启动和用户冷启动问题。

3.3.6 优化前后混合模型推荐算法的比较

为了验证本文提出算法的有效性,我们分别将其与内容过滤和改进的 Item-Based 协同过滤以及传统的混合模型推荐算法进行比较,如图 6 所示。

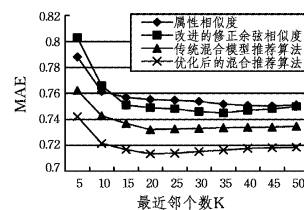


图 6 优化后混合模型推荐算法的效果

(下转第 98 页)

shers, 1994

- [2] Daniel A, Juan F, Fernando Lopez-Mesas. Nonsmooth Analysis and Hamilton-Jacobi Equations on Riemannian Manifolds [J]. Journal of Functional Analysis, 2005, 220(2): 304-361
- [3] LedyaeV Y S, Zhu Q J. Techniques for Nonsmooth Analysis on Smooth Manifolds; Local Problems [J]. Lecture Notes in Control and Information Sciences, 2004, 301: 283-297
- [4] LedyaeV Y S, Zhu Q J. Techniques for Nonsmooth Analysis on Smooth Manifolds; Deformations and Flows [J]. Lecture Notes in Control and Information Sciences, 2004, 301: 299-311
- [5] Daniel A, Juan F. Proximal Calculus on Riemannian Manifolds [J]. Mediterranean Journal of Mathematics, 2005, 2(4): 437-450
- [6] LedyaeV Y S, Zhu Q J. Techniques for nonsmooth analysis on smooth manifolds i; local problems [J]. Lecture Notes in Control and Information Sciences, 2004, 301: 283-297
- [7] LedyaeV Y S, Zhu Q J. Techniques for nonsmooth analysis on smooth manifolds ii; deformations and flows [J]. Lecture Notes in Control and Information Sciences, 2004, 301: 299-311
- [8] Azagra D, Ferrera J, Lopez-Mesas F. Nonsmooth analysis and Hamilton-Jacobi equations on smooth manifolds [J]. Journal of Functional Analysis, 2005, 220: 304-361

- [9] Azagra D, Ferrera J. Proximal calculus on Riemannian Manifolds [J]. Mediterranean Journal of Mathematics, 2005, 2: 437-450
- [10] 刘三阳, 朱石焕, 肖刚. 黎曼流形上 Fritz John 必要最优性条件 [J]. 辽宁师范大学学报: 自然科学版, 2007, 30(3): 268-272
- [11] 肖刚, 刘三阳, 尹小艳. 微分流形上的最优化算法 [J]. 西安电子科技大学学报: 自然科学版, 2007, 34(3): 472-475
- [12] 肖刚, 刘三阳. 黎曼流形上非可微多目标规划的必要最优性条件 [J]. 吉林大学学报: 理学版, 2008, 46: 209-213
- [13] 肖刚. 几类向量变分不等式的研究 [D]. 西安: 西安电子科技大学, 2008
- [14] 李志启. 黎曼流形上 Aronson-Benilan 非线性方程的局部梯度估计 [D]. 上海: 华东师范大学, 2012
- [15] 宋如意. 基于黎曼流形上的优化方法求解稀疏 PCA [D]. 上海: 复旦大学, 2012
- [16] 张晶. 黎曼流形上拉普拉斯算子的特征值估计 [D]. 新乡: 河南师范大学, 2012
- [17] 高恩芝, 王士同. 黎曼流形的距离均方差最小降维改进算法 [J]. 计算机工程与应用, 2013, 49(2): 198-202
- [18] Lang S. Analysis II [M]. Addison-Wesley
- [19] 付巍巍. 黎曼流形上非光滑优化最优性条件的研究 [D]. 阜新: 辽宁工程技术大学, 2009

(上接第 71 页)

从图 6 可以得出, 本文提出的优化混合模型推荐算法相对于单纯的内容过滤和改进的 Item-Based 协同过滤以及传统的混合模型推荐算法在推荐质量上确实有明显的提升。

结束语 本文对传统混合模型的推荐算法进行了优化, 增加了用户评分归一化及用户活跃度对物品相似性的影响, 新的相似性度量方法使得计算得到的目标物品的最近邻居更加准确。实验结果表明, 上述方法显著提高了推荐系统的推荐质量, 并且可以有效解决冷启动问题。实验不足之处是在解决冷启动问题上, 计算精度还有待提高, 需要进一步的深入研究。

参 考 文 献

- [1] Breese J, Heckerman D, Kadie C. Empirical analysis of predictive algorithms for collaborative filtering [C] // Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence (UAI-98). 1998: 43-52
- [2] Herlocker L J, Konstan A J, Riedl T J. Empirical analysis of design choices in neighborhood-based collaborative filtering algorithms [J]. Information Retrieval, 2002, 5(4): 287-310
- [3] Zeng C, Xing C X, Zhou L Z. A personalized search algorithm by using content-based filtering [J]. Journal of Software, 2003, 14(5): 999-1004
- [4] Sarwar B, Karypis G, Konstan J, et al. Item-Based Collaborative Filtering Recommendation Algorithms [A] // Proceedings of the 10th International World Wide Web Conference [C]. Paris: IEEE Computer Society Press, 2001: 285-295
- [5] Dempster A, Laird N, Rubin D. Maximum likelihood from incomplete data via the EM algorithm [J]. Journal of the Royal Statistical Society, 1977, B39: 1-38
- [6] Thiesson B, Meek C, Chickering D, et al. Learning mixture of DAG models [R]. Technical Report, MSR-TR-97-30. Redmond;

- Microsoft Research, 1997
- [7] Sarwar B M, Karypis G, Konstan J A, et al. Application of dimensionality reduction in recommender system—A case study [C] // ACM WebKDD 2000 Workshop. 2000
- [8] Peng Yu, Cheng Xiao-ping. Item-based collaborative filtering algorithm using attribute similarity [J]. Computer Engineering and Applications, 2007, 43(14): 144-147
- [9] Wu Yue-ping, Zheng Jian-guo. Collaborative Filtering Recommendation Algorithm on improved similarity measure method [J]. Computer Applications and Software, 2011, 28(10)
- [10] Deng Ai-lin, Zhu Yang-yong, Shi Bai-le. A Collaborative Filtering Recommendation Algorithm Based on Item Rating Prediction [J]. Journal of Software, 2003, 14(9): 1621-1628
- [11] Cheng Shu, Gui Lin, Ji Hang. Subjective Rating Normalization Algorithm and Error Analysis [J]. Journal of Higher Correspondence Education (Natural Sciences), 2007, 21(5)
- [12] Breese J S, Heckerman D, Kadie C. Empirical Analysis of Predictive Algorithms for Collaborative Filtering [R]. Technical Report, MSR-TR-98-12. May 1998
- [13] Chen Tian-qi, Zheng Zhao, Lu Qiu-xia, et al. Informative Ensemble of Multi-Resolution Dynamic Factorization Models [C] // KDD-Cup Workshop. 2011
- [14] Chen Tian-qi, Tang Lin-peng, Liu Qin, et al. Combining Factorization Model and Additive Forest for Collaborative Followee Recommendation [C] // KDD-Cup Workshop. 2012
- [15] Chen Tian-qi, Zheng Zhao, Yong Yu, et al. Svdfeature: User Reference Manual [R]. Technical Report, APEX-TR-2011-09-17. Apex Data & Knowledge Management Lab, Shanghai Jiaotong University, 2011
- [16] Li Xue, Zuo Wan-li, He Feng-ling, et al. An Improved Item-Based Collaborative Filtering Recommendation Algorithm [J]. Journal of Computer Research and Development, 2009, 46(Suppl.): 394-399