

文章编号: 1001-9081(2013)03-0834-04

doi: 10.3724/SP.J.1087.2013.00834

基于填充和相似性信任因子的协同过滤推荐算法

郝立燕*, 王 靖

(华侨大学 计算机科学与技术学院, 福建 厦门 361000)

(*通信作者电子邮箱 249664306@qq.com)

摘要:为了提高推荐系统在数据稀疏情况下的推荐质量,提出一种改进的协同过滤算法。该方法使用一种数据挖掘算法对稀疏评分矩阵进行填充;在完整的填充矩阵上计算用户相似性,并引入相似性信任因子;最终做出推荐预测。典型数据集上的对比实验结果表明,即使在评分数据极为稀疏的情况下,该算法仍能取得较好的结果。

关键词:推荐; 协同过滤; 稀疏性; 矩阵填充; 信任因子

中图分类号: TP391 文献标志码:A

Collaborative filtering recommendation algorithm based on filling and similarity confidence factor

HAO Liyan*, WANG Jing

(College of Computer Science and Technology, Huaqiao University, Xiamen Fujian 361000, China)

Abstract: In order to improve the recommendation quality of recommendation system when the data are sparse, an improved collaborative filtering algorithm was proposed. Using a data mining algorithm, the sparse rating matrix was filled firstly. Afterwards user-similarities and their confidence factors were calculated using the complete filling matrix. Ultimately, the recommendation forecast was made. Comparative experiments on typical dataset show that the algorithm is able to achieve better results even with extremely sparse data.

Key words: recommend; collaborative filtering; sparsity; matrix completion; confidence factor

0 引言

协同过滤推荐技术是解决信息过载问题的有效手段,其优势在于推荐的个性化、不受被推荐项目具体内容的限制并且具有较高的推荐可靠性^[1]。然而,协同过滤推荐技术和其他类型的推荐技术在实际应用中普遍面临数据稀疏性的问题^[2]。以电子商务推荐系统为例,在这些系统中用户购买商品的总量通常只会占网站商品总量的1%左右^[3]。用户只对极少数项目进行评分,用户项目评分矩阵极为稀疏。这给协同过滤推荐算法中用户相似性的计算带来了很大困难,并影响最终的推荐质量。许多研究者将聚类的思想应用于协同过滤算法中以缓解数据稀疏性问题。吴湖等^[4]提出一种两阶段评分预测方法。首先对原始矩阵中的评分模式进行用户和项目两个维度的联合聚类,然后在这些类别的内部通过加权非负矩阵分解方法进行未知评分预测。Tsai等^[5]提出在协同过滤推荐算法中使用融合聚类(cluster ensemble)技术。然而,这些使用聚类技术的算法一定程度上缓解了数据稀疏性问题,但是也存在对处于聚类边缘的用户推荐精度比较低的问题。也有一些研究者将一些降维技术应用于推荐算法。Sarwar等^[6]最先提出采用奇异值分解(Singular Value Decomposition, SVD)技术对评分矩阵进行降维处理,寻找用户之间的隐含关系。赵宏霞等^[7]为解决协同过滤推荐算法中数据稀疏性的问题,采用因子分析的方法对数据降维,并使用回归分析方法预测待评估值,既减少了数据量又较大限度保留了信息。这些采用降维技术的算法也取得一定成果,但

不可避免地会牺牲一部分有用信息。在数据本身就极为稀疏情况下,这一问题是不容忽视的。综上所述,面对数据稀疏性问题,如何充分有效地利用已有数据信息,提高算法推荐质量仍然是值得研究的课题。

为了提高数据稀疏情况下的推荐质量,本文在传统协同过滤算法的基础上提出基于填充和相似性信任因子的协同过滤推荐算法。

1 相关工作和定义

1.1 协同过滤推荐算法的一般过程及相关定义

在推荐系统中应用协同过滤的思想,需要各用户对各项目的兴趣度信息。这一信息通常以用户项目评分矩阵的形式表示。一个包含m个用户,n个项目的评分矩阵可以表示为:

$$\mathbf{R} = \begin{bmatrix} r_{1,1} & \cdots & r_{1,x} & \cdots & r_{1,n} \\ \vdots & & \vdots & & \vdots \\ r_{p,1} & \cdots & r_{p,x} & \cdots & r_{p,n} \\ \vdots & & \vdots & & \vdots \\ r_{m,1} & \cdots & r_{m,x} & \cdots & r_{m,n} \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_p \\ \vdots \\ u_m \end{bmatrix}$$
$$i_1 \quad \cdots \quad i_x \quad \cdots \quad i_n$$

矩阵 \mathbf{R} 中每一行向量表示一个特定用户的评分集合,每一列向量表示一个特定项目的被评分集合,每一个元素 $r_{u,i} \in \mathbf{R}$ 表示用户 u 对项目 i 的评分值。在获取用户项目评分矩阵后,一个典型的协同过滤推荐过程表述如下:

1) 利用评分矩阵,计算用户 $u, v \in U$ (U 表示评分矩阵中所有用户的集合)之间的相似性。相似性公式的选择至关重要。本文采用实践中表现较好的Pearson相关系数公式:

收稿日期: 2012-09-14; 修回日期: 2012-10-12。

基金项目: 国家自然科学基金资助项目(10901062); 福建省高等学校杰出青年科研人才培育计划项目(11FJ01)。

作者简介: 郝立燕(1988-), 男, 山东临沂人, 硕士研究生, 主要研究方向: 数据挖掘、推荐系统; 王靖(1981-), 男, 福建泉州人, 副教授, 博士, 主要研究方向: 模式识别、推荐系统。

$$sim_{uv} = \frac{\sum_{i \in I_{uv}} (r_{u,i} - \bar{r}_{u,I_{uv}})(r_{v,i} - \bar{r}_{v,I_{uv}})}{\sqrt{\sum_{i \in I_{uv}} (r_{u,i} - \bar{r}_{u,I_{uv}})^2} \sqrt{\sum_{i \in I_{uv}} (r_{v,i} - \bar{r}_{v,I_{uv}})^2}} \quad (1)$$

其中:累加范围 I_{uv} 表示用户 u 与用户 v 共同评价过的项目的集合, $\bar{r}_{u,I_{uv}}$ 是指用户 u 在项目集合 I_{uv} 上的所有评分的均值。

2) 确定目标用户 u_o 的邻居用户集合 U_{u_o} , 通常采用 k 近邻方式, 即用户 $u \in U$ 按照与目标用户的相似性值 sim_{uu_o} 从大到小排序, 前 k 个用户构成邻居用户集合 U_{u_o} 。

3) 计算目标用户 u_o 对他的未评分项目 $i \in C_I I_{u_o}$ ($C_I I_{u_o}$ 表示集合 I_{u_o} 在集合 I 中的补集) 的预测评分值 $r_{u_o,i}$:

$$r_{u_o,i} = \bar{r}_{u_o,I_{u_o}} + \frac{\sum_{u \in U_{u_o}} (r_{u,i} - \bar{r}_{u,I_u}) \cdot sim_{uu_o}}{\sum_{u \in U_{u_o}} |sim_{uu_o}|} \quad (2)$$

1.2 奇异值分解技术

奇异值分解是一种矩阵分解技术, 它深刻揭示了矩阵的内部特征^[8]。奇异值分解可以将 $m \times n$ 阶矩阵 R 分解为三个矩阵 U, Σ, V 的乘积, 即

$$R = U \times \Sigma \times V^T \quad (3)$$

其中: U, V 分别是 $m \times m$ 和 $n \times n$ 阶正交矩阵; $\Sigma = [\Sigma_1 \ 0 \ 0 \ 0]$ 且

Σ_1 是对角矩阵, 对角线上的元素为矩阵 R 的奇异值。利用 SVD 可以获得矩阵 R 的最佳低秩逼近矩阵 R_r , 即

$$R_r = \underset{\text{rank}(M)=r}{\operatorname{argmin}} \|R - M\|_F \quad (4)$$

记 $\Sigma_r = \text{diag}(\sigma_1, \dots, \sigma_r)$ 为 R 的前 r 大奇异值构成的对角矩阵; U_r, V_r 分别为 $\sigma_1, \dots, \sigma_r$ 相应的左奇异向量和右奇异向量构成的矩阵, 则

$$R_r = U_r \times \Sigma_r \times V_r^T \quad (5)$$

SVD 在协同过滤算法中有着重要的应用。

2 本文算法

2.1 算法设计思想

在用户项目稀疏评分矩阵上, 直接计算用户相似性面临诸多问题。例如, 相似性公式(1)利用的是用户之间共同评分过的项目对应的评分。大多数用户之间的共同评分项目极少甚至没有, 对应的用户相似性值只受少数几个共同项目评分的影响, 计算结果充满了偶然性^[9]。另外, 很多没有共同评分项目的用户之间, 会存在潜在的相似性^[10]。直接计算用户相似性, 也无法发现那些潜在的相似性信息。对此, 充分挖掘用户项目特征, 首先对稀疏评分矩阵进行填充。在填充后的完整矩阵上计算用户相似性, 可以避免结果的偶然性, 并发现潜在的用户相似性。为了缓解推荐系统中普遍存在的数据稀疏性问题, 提出基于填充和相似性信任因子的协同过滤推荐算法。算法流程如图 1 所示。接下来, 本文将详细讨论该算法的具体过程及原理。

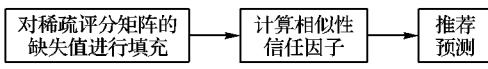


图 1 本文算法流程

2.1.1 矩阵填充及相似性计算

对稀疏评分矩阵的缺失值进行填充是为了更有效地利用已有评分信息计算用户相似性。充分挖掘到原始矩阵中的评分特征信息, 再加以有效利用, 是矩阵填充的关键。研究者已经提出了许多种解决稀疏矩阵缺失值填充问题的方法, 其中最基本的方法是利用原始矩阵中用户项目的评分均值填充对应的缺失值^[11]。这种填充方法增强了相似性计算的稳定性, 但是也在一定程度上削弱了稀疏矩阵的数据特征。

Sarwar 等^[6]首次把奇异值分解技术应用到协同过滤算法中, 使用 SVD 技术挖掘用户与项目之间的潜在关系预测用户对特定项目的喜好。这在一定程度上缓解了数据稀疏性问题。在这之后, SVD 技术以各种方式在协同过滤算法中得到广泛运用。孙小华等^[10]综合基于 SVD 技术的协同过滤算法与基于 k 近邻的协同过滤算法的优势提出了 Pear_After_SVD 算法。该算法首先使用 SVD 技术对原稀疏评分矩阵进行填充, 然后利用填充后的完整矩阵计算用户之间的相似性, 最后采用 k 近邻方式选择邻居用户并最终作推荐预测。使用 SVD 技术的填充(例如 Pear_After_SVD)通常只是在平均值填充的基础上增加一部分通过 SVD 提取的矩阵特征信息。然而, 只通过一次 SVD 提取到的矩阵特征信息是极为有限的。

以上两种填充方式都在一定程度上改善了稀疏评分数据情况下的推荐质量, 但是评分矩阵的数据特征并没能被充分利用。为了有效挖掘稀疏矩阵数据特征, 本文提出使用 SOFT-IMPUTE 算法^[11]对协同过滤算法中用到的稀疏评分矩阵进行填充。对于稀疏的评分矩阵 R , SOFT-IMPUTE 算法通过求解最优化问题(式(6))获取 R 的近似填充矩阵 R^* 。

$$\min_{R^*} \frac{1}{2} \|W \odot (R - R^*)\|_F^2 + \lambda \|R^*\|_* \quad (6)$$

其中: \odot 表示两个矩阵在对应位置元素上进行乘积运算; $\|R\|_*$ 表示矩阵 R 的核范数, 即矩阵 R 所有奇异值的和; W 为与评分矩阵 R 同规模的 0-1 矩阵: 当矩阵 R 中元素 $r_{i,j}$ 为缺失值时, 矩阵 W 中元素 $w_{i,j} = 0$; 否则 $w_{i,j} = 1$ 。SOFT-IMPUTE 算法采用迭代的 SVD 方法求解模型(式(6)), 从理论上保证了迭代的收敛性。与传统的 SVD 方法相比, 模型(式(6))通过多次 SVD 提取稀疏矩阵特征信息, 可以更为充分利用评分矩阵的数据特征。另外, 该模型引入了正则项 $\lambda \|R^*\|_*$, 很大程度缓解了数据稀疏性带来的过拟合问题。因此本文采用 SOFT-IMPUTE 算法对协同过滤算法中用到的稀疏评分矩阵进行初始填充。

采用 SOFT-IMPUTE 算法对用户项目稀疏评分矩阵 R 进行填充, 填充结果记为 R^* 。填充矩阵中的每一个元素 $r_{u,i}^* \in R^*$, 表示用户 u 对项目 i 在填充矩阵 R^* 上的评分值。在填充矩阵上, 计算用户 $u, v \in U$ 之间的相似性:

$$sim_{uv}^* = \frac{\sum_{i \in I} (r_{u,i}^* - \bar{r}_{u,I}^*)(r_{v,i}^* - \bar{r}_{v,I}^*)}{\sqrt{\sum_{i \in I} (r_{u,i}^* - \bar{r}_{u,I}^*)^2} \sqrt{\sum_{i \in I} (r_{v,i}^* - \bar{r}_{v,I}^*)^2}} \quad (7)$$

2.1.2 相似性信任因子及用户评分预测

在填充矩阵 R^* 上, 利用式(7)可以计算得到任意用户 $u, v \in U$ 之间的相似性。根据相似性值得大小, 可以确定邻居用户范围。继而可以利用式(2)计算目标用户 u_o 对项目 $i \in C_I I_{u_o}$ 的预测评分值, 最终做出推荐预测。然而, 直接将式(7)计算得到的用户相似性应用于式(2)并不合理。原评分矩阵数据极为稀疏, 数据填充存在误差, 这都导致式(7)得到的用户相似性未必能反映用户之间的关系。在进行用户的评分预测时, 应该考虑到式(7)得到的相似性具有不同的可靠性, 本文称它为相似性信任因子。相似性信任因子高, 在评分预测式(2)中该相似性值具有较高的权重; 反之, 具有较低的权重。

在填充矩阵上计算得到不同用户之间的相似性值, 对应不同的信任度因子。本文认为该相似性信任因子受两个因素的影响: 数据稀疏性和数据填充误差。使用 SOFT-IMPUTE 算法对原稀疏评分矩阵进行处理, 可以得到对应的填充矩阵。在这个过程中, 原稀疏评分矩阵中评分数越稠密的用户, 被挖掘到的信息越多, 填充矩阵中包含的相应用户特征信息也

越多,该用户对应的填充数据越可靠;反之,用户对应的填充数据越不可靠。本文用 $\underline{\text{num}}(u)$ 表示任意用户 $u \in U$ 在稀疏评分矩阵 R 中的评分数目, $\overline{\text{num}}(U)$ 表示稀疏矩阵 R 中所有用户 $u \in U$ 的平均评分数目。那么,原稀疏评分矩阵中任意用户 $u \in U$ 的评分数据稠密程度可以表示为:

$$d_u = \begin{cases} 0, & \underline{\text{num}}(u) = 0 \\ \exp\left(\frac{\underline{\text{num}}(u)}{\overline{\text{num}}(U)} - 1\right), & 0 < \underline{\text{num}}(u) < \overline{\text{num}}(U) \\ 1, & \underline{\text{num}}(u) \geq \overline{\text{num}}(U) \end{cases} \quad (8)$$

另一方面,数据填充误差也需要被考虑。原稀疏矩阵已经有部分数据,填充矩阵在对应位置上也有填充值。最好情况下,稀疏矩阵的已有数据与填充矩阵中的对应数据具有相同值。然而,以这样的目标进行填充,会出现过拟合的问题。因此,稀疏矩阵的已有数据与填充矩阵中的对应数据必然存在差别,即填充误差。任意用户 $u \in U$ 的填充误差可以表示为:

$$s_u = \frac{1}{\text{sim}_{uu}} = \frac{\sqrt{\sum_{i \in I_u} (r_{u,i} - \bar{r}_{u,I_u})^2} \sqrt{\sum_{i \in I_{u_p}} (r_{u,i}^* - \bar{r}_{u,I_u}^*)^2}}{\sum_{i \in I_u} (r_{u,i} - \bar{r}_{u,I_u}) (r_{u,i}^* - \bar{r}_{u,I_u}^*)} \quad (9)$$

结合以上分析,本文构造填充矩阵 R^* 中任意用户 $u \in U$ 的用户信任因子:

$$f_u = \left| \frac{d_u}{s_u} \right| \quad (10)$$

显然,用户信任因子 f_u 满足以下几个特点:1) 取值范围在 $0 \sim 1$;2) 与用户评分稠密度成正比例关系;3) 与数据填充误差成反比例关系。用户相似性表达了两个用户的评分特征关系。使用填充矩阵 R^* 中数据,计算任意用户 $u, v \in U$ 之间的相似性,该相似性对应的相似性信任因子表示为:

$$w_{uv} = f_u \cdot f_v \quad (11)$$

最后,将相似性信任因子引入式(2),以期望获得更准确的评分预测。目标用户 u_o 对项目 $i \in C_I I_{u_o}$ 的评分预测为:

$$\hat{r}_{u_o, i} = \bar{r}_{u_o, I_{u_o}} + \frac{\sum_{u \in U_{u_o}} (r_{u,i} - \bar{r}_{u,I_u}) \cdot \text{sim}_{uw_o}^* \cdot w_{uw_o}}{\sum_{u \in U_{u_o}} |\text{sim}_{uw_o}^* \cdot w_{uw_o}|} \quad (12)$$

2.2 本文算法的实现

基于以上分析,本文提出的基于填充和相似性信任因子的协同过滤算法。

算法 基于填充和相似性信任因子的协同过滤算法。

输入 稀疏评分数据集 R , 目标用户 u_o , 邻居用户数 k 。

输出 目标用户 u_o 对项目 $i \in C_I I_{u_o}$ 的评分预测值。

Step1 利用 SOFT-IMPUTE 算法对稀疏评分矩阵 R 进行填充,填充结果记为 R^* ;

Step2 在填充矩阵 R^* 上,利用式(7)计算目标用户 u_o 与任意用户 $u \in U$ 的相似性 sim_{uw_o} ;

Step3 用户 $u \in U$ 按照相似性值 sim_{uw_o} 从大到小排序,前 k 个用户构成邻居用户集合 U_{u_o} ;

Step4 利用式(8)~(11)计算相似性值 sim_{uw_o} ($u \in U_{u_o}$) 对应的相似性信任因子 w_{uw_o} ;

Step5 利用式(12)计算目标用户 u_o 对他的未评分项目 $i \in C_I I_{u_o}$ 的评分值 $\hat{r}_{u_o, i}$ 并输出。

3 实验

3.1 数据集和度量标准

本文使用 MovieLens 站点 (<http://movielens.umn.edu>) 提

供的测试数据集。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。从该站点下载 1997 年 9 月 19 日至 1998 年 5 月 22 日的数据集,包括 943 个用户对 1682 个项目的 10 万条投票记录,用户评分数据集的稀疏等级为 $1 - 100\ 000 / (943 \times 1682) = 93.7\%$ 。针对这一数据集,有多种划分训练集和测试集的方式。该站点提供了一种划分好的数据集,训练集和测试集分别占整个数据集的 80% 和 20%。本文实验也将首先使用这种划分方式下的数据验证本文算法的可行性;然后按照不同的比例重新划分子集,进一步验证算法可行性。

推荐系统推荐质量的度量标准主要有统计精度度量方法和决策支持精度度量方法^[12]。实验采用统计精度度量方法中广泛使用的平均绝对偏差 (Mean Absolute Error, MAE) 作为度量标准,即通过计算预测的用户评分与实际用户评分之间的偏差来衡量预测的准确性。MAE 值越小说明算法推荐的精度越高。假设预测的用户评分集合为 $\{p_1, p_2, \dots, p_n\}$, 对应的实际评分集合为 $\{q_1, q_2, \dots, q_n\}$, 那么预测结果的 MAE 定义为:

$$MAE = \frac{\sum_{i=1}^n |p_i - q_i|}{n} \quad (13)$$

3.2 实验结果和性能比较分析

为了验证本文所使用的填充算法 (SOFT-IMPUTE) 对传统协同过滤算法的改善作用,首先将以下几种算法进行对比测试:

- 1) Pearson-CF: 传统基于 Pearson 的协同过滤算法;
- 2) CF-mean^[9]: 使用评分均值进行矩阵填充,在填充矩阵上计算用户相似性;
- 3) CF-SVD^[10]: 使用奇异值分解技术进行矩阵填充,在填充矩阵上计算用户相似性;
- 4) CF-SOFT: 利用 SOFT-IMPUTE 算法进行矩阵填充,在填充矩阵上计算用户相似性。

本实验中,算法中使用到的参数为通过多次测试取它们的最优值。SOFT-IMPUTE 算法中设置衰减因子 $\lambda = 10$, 阈值 $\varepsilon = 10^{-6}$ 。CF-SVD 算法中设置奇异值数目为 38。图 2 列出了不同邻居数目下几种算法的 MAE 值,从图中可看出,CF-SOFT 算法在所有的邻居数目下,都取得了最小的 MAE。由此可见,SOFT-IMPUTE 算法在这几种填充算法中,最适用于用户的相似性计算,从而得到最高的推荐精度。

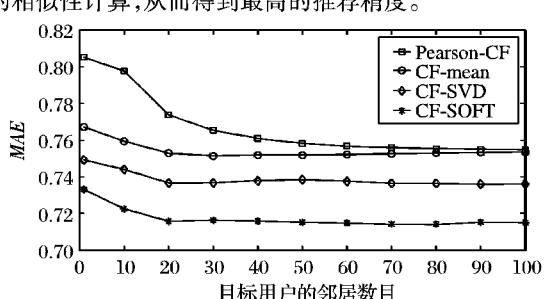


图 2 不同邻居数目下各算法的 MAE 比较

进一步地,为了验证本文提出的用户相似性权重因子对算法推荐精度的改善作用,比较直接 SOFT-IMPUTE 填充后的协同过滤算法 (CF-SOFT) 和在此基础上进一步引入相似性信任因子后的协同过滤算法 (WCF-SOFT)。图 2 列出了不同邻居数目下这两种算法的 MAE 值,从图中可看出,在不同邻居数目下,引入本文提出的相似性信任因子都能减小 MAE,从

而提高推荐精度。

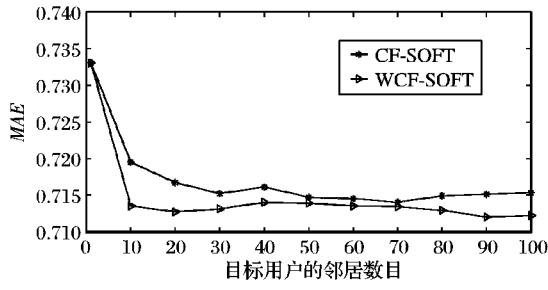


图3 相似性信任因子的影响

最后,为了验证算法在不同稀疏度评分数据下的性能,将MovieLens数据集中训练集和测试集分别按照4:6、5:5、6:4和7:3的比例随机划分。在不同稀疏度的评分数据下对WCF-SOFT、CF-SVD、CF-mean和Pearson-CF进行对比实验(邻居数目均采用20个),实验结果如表1所示。实验结果表明,同CF-SVD、CF-mean、Pearson-CF算法相比,本文算法在各种稀疏度的数据下都取得最好的结果。即使在评分数据最稀疏的情况下,本文算法得到的结果也接近于CF-SVD算法在最多评分数据下的结果。由此可见,本文算法在缓解用户项目评分稀疏性问题上有较好的效果。

表1 不同数据稀疏度下各种算法的MAE比较

算法	4:6	5:5	6:4	7:3
WCF-SOFT	0.7280	0.7198	0.7131	0.7081
CF-SVD	0.7526	0.7426	0.7304	0.7222
CF-mean	0.7591	0.7526	0.7428	0.7392
Pearson-CF	0.7739	0.7696	0.7642	0.7642

4 结语

本文在分析传统协同过滤推荐算法及其稀疏性问题的基础上,提出基于填充和相似性信任因子的协同过滤推荐算法。采用SOFT-IMPUTE算法进行填充,对数据特征挖掘更加充分,填充效果较好,改善了推荐质量。相似性信任因子的引入使基于填充的相似性值得到更合理使用,进一步改善推荐质量。在不同稀疏度数据集下,该算法都能取得较好的推荐结果,缓解推荐系统普遍面临的数据稀疏性问题。需要指明的是,该算法的执行需要存储完整的填充矩阵,对存储空间有一

定的要求。算法使用SOFT-IMPUTE对稀疏矩阵进行填充,每次迭代需要一次奇异值分解,计算量较大。因此,该算法适合中小规模的推荐系统,而不适合大规模的推荐系统。在改善推荐质量的同时,提高算法的可扩展性是有意义的下一步研究方向。

参考文献:

- [1] YANG J M, LI K F. An inference-based collaborative filtering approach[C]// Proceedings of the 3rd IEEE International Symposium on Dependable, Autonomic and Secure Computing. Piscataway, NJ: IEEE Press, 2007: 84–94.
- [2] PARK S T, PENNOCK D M. Applying collaborative filtering techniques to movie search for better ranking and browsing[C]// Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2007: 550–559.
- [3] CASTRO-SCHEZ J, MIGUEL R. A highly adaptive recommender system based on fuzzy logic for B2C ecommerce portals[J]. Expert Systems with Applications, 2011(38): 2441–2454.
- [4] 吴湖,王永吉.两阶段联合聚类协同过滤算法[J].软件学报,2011,21(5):1042–1054.
- [5] TSAI C F, HUNG C. Cluster ensembles in collaborative filtering recommendation[J]. Applied Soft Computing, 2012, 12(4): 1417–1425.
- [6] SARWAR B, KARYPIS G, KONSTAN J. Item-based collaborative filtering recommendation algorithms[C]// Proceedings of the 10th International World Wide Web Conference. New York: ACM Press, 2001: 285–295.
- [7] 赵宏霞,王新海,杨皎平.基于用户和项目因子分析的混合协同推荐算法[J].计算机应用,2011,31(5):1382–1386.
- [8] KALMAN D. A singularly valuable decomposition: the SVD of a matrix[J]. The College Mathematics Journal, 1996, 27(1): 2–23.
- [9] LANDERMAN L R, PIEPER K C. An empirical evaluation of the predictive mean matching method for imputing missing values[J]. Sociological Methods & Research, 1997, 26(1): 3–33.
- [10] 孙小华,陈洪,孔繁胜.在协同过滤中结合奇异值分解与最近邻方法[J].计算机应用研究,2006,23(9):206–208.
- [11] MAZUMDER R, HASTIE T, TIBSHIRANI R. Spectral regularization algorithms for learning large incomplete matrices[J]. The Journal of Machine Learning Research, 2010, 11: 2287–2322.
- [12] 陈健,印鉴.基于影响集的协作过滤推荐算法[J].软件学报,2007,18(7):1685–1694.

(上接第673页)

最大类间方差的二维阈值分割方法中。实验结果表明,相对于现有的几种经典的二维阈值分割方法,本文方法的分割效果更好。图像分割质量评价指标的评测结果也证实了本文方法的有效性。

参考文献:

- [1] 张新明,孙印杰,郑延斌.二维直方图准分的Otsu图像分割及其快速实现[J].电子学报,2011,39(8):1779–1784.
- [2] OTSU N. A threshold selection method from gray-level histogram [J]. IEEE Transactions on Systems, Man, Cybernetics, 1979, 9(1): 62–66.
- [3] 刘建庄,栗文青.灰度图像的二维Otsu自动阈值分割法[J].自动化学报,1993,19(1):101–105.
- [4] GONG J, LI L Y, CHEN W N. Fast recursive algorithm for two-dimensional thresholding[J]. Pattern Recognition, 1998, 31(3): 295–300.
- [5] 汪海洋,潘德炉,夏德深.二维Otsu自适应阈值选取法的快速实现[J].自动化学报,2007,33(9):968–971.
- [6] 郝颖明,朱枫.二维Otsu自适应阈值的快速算法[J].中国图象

图形学报,2005,10(4):484–488.

- [7] 吴一全,张金矿.平均灰度级-梯度二维直方图最大相关熵阈值分割[J].小型微型计算机系统,2009,30(8):1675–1679.
- [8] 范九伦,赵凤.灰度图像的二维Otsu曲线阈值分割法[J].电子学报,2007,35(4):751–755.
- [9] 吴一全,潘喆,吴文怡.二维直方图区域斜分阈值分割及快速递推算法[J].通信学报,2008,29(4):77–83.
- [10] 吴一全,张金矿.二维直方图θ-划分最大平均离差阈值分割算法[J].自动化学报,2010,36(5):634–643.
- [11] FERWERDA J A. Element of early vision for computer graphics[J]. IEEE Computer Graphics and Applications, 2001, 21(5): 22–33.
- [12] 侯志强,韩崇昭,郑林,等.一种基于视觉模型的边缘检测阈值选取策略.光电子工程,2004,31(2):59–62.
- [13] YASNOFF W, MUI J, BACUS J. Error measures for scene segmentation[J]. Pattern Recognition, 1977, 9(4): 217–231.
- [14] LEVINE M D, NAZIF A M. Dynamic measurement of computer generated image segmentations[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1985, 7(2): 155–164.