融合隐性特征的群体推荐方法研究

刘 毅 钟 忺 李 琳
(武汉理工大学计算机科学与技术学院 武汉 430070)

摘 要 作为目前最成功的主流推荐方法,奇异值分解算法(SVD)将已知的海量数据建模并通过矩阵分解降维处理来得到有效信息;非负矩阵分解(NMF)则通过分解出非负矩阵元素来解释特征意义。这两种较为成功的方法均通过对显性反馈信息进行基于矩阵分解的处理得到用户的喜好信息来进行群体推荐。然而,仅凭用户的显性反馈信息有时无法准确反映用户的真实喜好。为解决上述问题,提出了一种针对这两种模型的改进方法,将隐性特征和基于隐性特征的群体权重计算方法融合进经典的矩阵分解算法,其中隐性特征可以完善用户的喜好信息,基于隐性特征的群体权重计算方法则根据群体的特点给予用户相应的权重,使得推荐的准确率得到提升。对该方法在KDD Cup 2012 Track1中的腾讯微博数据集上进行测试,实验结果表明在该数据集上融合方法的平均绝对偏差(MAE)和准确率要优于SVD算法与NMF算法,推荐的性能有较明显的提升。

关键词 群体推荐,隐性特征,群体权重,平均绝对偏差
中图法分类号 TP319.41 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.03.048

Research on Method of Group Recommendation for Fusion of Hidden Features

LIU Yi ZHONG Xian LI Lin
(Department of Computer Science and Technology, Wuhan University of Technology, Wuhan 430070, China)

Abstract As the most successful mainstream recommendation method, singular value decomposition (SVD) algorithm builds the model from known huge data and uses the matrix decomposition dimension reduction to get effective information, and non negative matrix factorization (NMF) uses the decomposition of nonnegative matrix elements to explain the meanings of characteristics. These two kinds of successful methods are based on matrix decomposition of explicit feedback information, and obtain the user's preference information. However, they cannot accurately reflect the true preferences of the users only according to user's explicit feedback. To solve the problem, this paper put forward an improved method for the two models, integrating the hidden features and weight calculation method based on hidden features into the classical matrix decomposition algorithm, the hidden features can perfect the information of user's preferences. Weight calculation method based on hidden features can judge the group characteristics and give the appropriate weight to users, which improve the recommendation accuracy. The method was tested on the Tencent micro blog data set in KDD Cup 2012 Track1. The results show that from the experimental standard of the MAE and the precision, the fusion method is better than SVD and NMF on this data set, and significantly improves the recommendation performance.

Keywords Group recommendation, Hidden features, Weights of the group, MAE

1 引言

互联网的蓬勃发展使社会高度信息化,在为人们带来前所未有的方便的同时也产生了"数据爆炸"——海量信息导致人们需要浪费很多时间来寻找自己需要的资源。推荐系统根据用户的爱好,可以通过人工智能和数据挖掘等方法向用户提供个性化的服务、商品以及信息的推荐。目前,推荐系统除了能提供生活与娱乐方面的推荐服务外,还可以提供更加多样化的推荐,如科技论文、专利技术推荐[1-3]。推荐系统已经成为数学、计算机科学、物理学等学科研究的热点,其成果将促进社会各方面的发展。

目前已有的主流推荐系统分为以下几类:协同过滤推荐系统、基于内容的推荐系统、混合推荐系统。虽然不同的推荐系统采用了不同的方法向用户提供推荐服务,但几乎所有的个性化推荐算法都是通过分析已经获得的用户显性信息,进而为用户提供个性化的商品或信息服务。协同过滤推荐系统

到稿日期:2016-01-26 返修日期:2016-05-17 本文受国家社会科学基金(15BGL048),武汉市创新团队项目(201307020402005),湖北省自然科学基金:城市隧道监控大数据分析预测及视频语义检索方法研究(ZRY2015001126)资助。
刘 毅(1991—),男,硕士生,主要研究方向为智能技术、大数据挖掘;钟 忺(1985—),男,博士,副教授,主要研究方向为视频语义检索、大数据分析,E-mail:345128484@qq.com;李 琳(1978—),女,博士,教授,博士生导师,主要研究方向为推荐系统、信息检索等。

由于处理较为容易，分析数据时不需要了解数据所对应的领域知识，是应用最成功的推荐系统。协同过滤算法分为两类：基于内存的协同过滤算法，准确性在密集的评分矩阵中能达到预期要求，然而对于稀疏的评分矩阵其准确性很差；基于模型的协同过滤算法，能通过特征提取和模型建立解决矩阵稀疏的问题，其中最有代表性的是基于矩阵分解的推荐算法[4-5]，如奇异值分解、非负矩阵分解等。奇异值分解(SVD)算法最早出现于文献[6]中的协同过滤推荐并展现出了优秀的性能，如今 SVD 算法是最为流行的基于矩阵分解降低数据维度的算法[6-10]；非负矩阵分解(NMF)通过分解出非负的矩阵元素来解释特征意义，也是常用的矩阵分解方法。

群体推荐系统是个人推荐系统的延伸，旨在为一个群体提供项目推荐。由于群体成员的兴趣爱好具有差异性、多样性和复杂性，因此不能简单地按照个人推荐系统的算法来为群体提供项目推荐。此外，群体推荐涉及群体决策，而基于群体决策的本质，决策的结果可能会因为群体成员在群体中所拥有的地位与发言权不同而不同，简而言之，在构建群体推荐系统时必须考虑到群体中的每个用户可能具有不同权重的实际情况，而这将使得群体推荐的过程更为复杂。本文在常规的 SVD 方法中引入含蓄的用户反馈信息——隐性特征，并研究针对不同情况的群体的权重计算方法，以提高推荐的准确性。

2 群体权重模型设计

用户所组成的网络中最基本的单位是群体，而用户之间存在熟悉和不熟悉两种关系，用户间的交互作用既有直接的也有间接的。一般情况下，用户网络中单独的一个用户节点会与其他 $n(n\geqslant 1)$ 个用户相连，共同组成用户网络图谱。在此，定义熟悉程度为群体成员互相之间对彼此兴趣爱好的了解程度。根据群体成员间的熟悉程度，可以将群体划分为陌生关系群体和熟悉关系群体。

2.1 陌生关系群体

如果构成某群体的成员之间感兴趣的对象不同并且无交互行为，这种具有多元性、临时性、复杂性的群体即称为陌生关系群体。鉴于此群体的特点，可将其视为若干个“孤岛”所组成的“群体岛”，“孤岛”之间没有任何联系。通过兴趣图谱和打分数据可以分析出陌生关系群体中群体用户的权重。

2.2 熟悉关系群体

如果构成某群体的成员之间存在交互作用且兴趣不尽相同，则称这种群体为熟悉关系群体，如图 1 所示。这类群体最为普遍。为了得到熟悉关系群体中群体成员的权重，需要分析用户之间的社交图谱折射出的用户之间的熟悉程度，但通过分析发现用户之间对于相互间熟悉程度的认知是不同的，如图 1 中，A,B,C,D,E 共同构成一个熟悉群体用户，一个用户用一个单独节点表示，两个用户之间的熟悉关系用箭头表示，箭头上的数字代表熟悉程度，数字 5 代表的熟悉程度最高，1 最低。例如：$A->C$ 代表用户 A 认为自己对用户 C 的熟悉程度是 4。在熟悉关系群体中没有孤立的节点，任意两个节点都是可以通过其他节点互相抵达的。

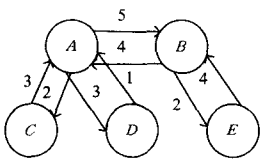

图 1 熟悉关系群体结构

3 权重计算

3.1 陌生关系群体用户权重计算

在陌生关系群体中，由于处于同一群体中的用户没有交互行为，使得该群体具有随机性和临时性。考虑到陌生关系群体中的用户均为各自孤立的节点，计算此类群体用户的权重的过程较熟悉关系群体更为复杂，主要考虑从隐性特征分析数据，如用户在网上的隐性特征以及权威程度。其中，权威程度指群体成员在群体社交中的影响力大小[7,10]，例如：一个对某领域非常了解或者很有威望的专家往往拥有较高的权威程度，这是因为他所给出的评价信息具有更高的指导价值，并且有许多人会赞成并附和。

隐性特征的一般指标有：月上线次数、打分次数、发表评论次数、系统中的页面访问总量、回复内容匹配度等；衡量权威程度的指标则包括：所在系统内部等级、关注/点赞人数、回复被采纳次数等。具体指标结构如图 2 所示。

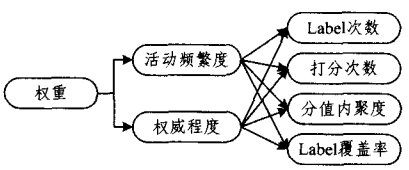

图 2 陌生关系群体的权重指标构造

Label 次数：一段时间内用户产生的标签行为次数。

打分次数：一段时间内用户打分行为发生的次数。

分值内聚度：所评分内容的平均评分和群体用户给出分值的误差平方和。假设整个群体的总量是 a，一起打分过的 $item$ 数量是 m，用 $\overline{p_1},\overline{p_2},\cdots,\overline{p_m}$ 表示通过平均聚合策略计算所得的陌生关系群体对 $item_1,item_2,\cdots,item_m$ 所给出的平均分值。对于单个用户，p_{xy} 表示第 x 个用户对第 $y(1\leqslant y\leqslant m)$ 个项目的评分值，该分值内聚度公式为：

$$\sum_{1\leqslant x\leqslant a}^{1\leqslant y\leqslant m}(p_{xy}-\overline{p_m})^2 \tag{1}$$

Label 覆盖率：群体中的每一位用户在网络上都有属于各自的爱好范围，能够查询到用户感兴趣的对象；在 Label 数据中，以用户自身为中心，可以构建用户的爱好图谱，如图 3 所示。

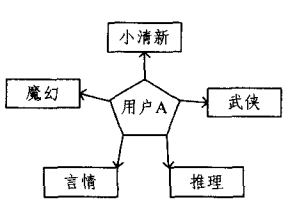

图 3 陌生关系群体中某用户的爱好图谱

依照此方法可以构建陌生群体中所有用户 Label 数据组成的爱好图谱，并根据用户爱好图谱中的相交爱好点的数目

得出相应用户的权重。如上所述，即使是互相陌生的用户也可以通过相同的爱好点联系起来。假设某个用户所有爱好点的连接数为 A，该群体总的爱好点数目为 B，那么该用户的 *Label* 覆盖率为 A/B。

构造陌生关系群体中的指标判断矩阵：

$$\begin{bmatrix} 1 & i & j & k \\ 1/i & 1 & e & g \\ 1/i & 1/e & 1 & h \\ 1/k & 1/g & 1/h & 1 \end{bmatrix}$$

通过 AHP 层次分析法可以得出 Label 次数、打分次数、分值内聚度和 Label 覆盖率的重要程度，其权重分别为 w_a，w_b，w_c 和 w_d。设用户 A 在一段时间 t 内得出的相应指标数值为：Label 次数 X，打分次数 Y，分值内聚度 Z，Label 覆盖率 M。使用离差标准化的方法[11]对数据进行归一化处理，将结果值映射到[0,1]之间，转换函数为：

$$x^* = \frac{x - min}{max - min} \tag{2}$$

其中，变量 min 代表样本数据最小值，max 则代表样本数据最大值。经过归一化处理得到 Label 次数 x、打分次数 y、分值内聚度 z 以及 Label 覆盖率 m。

$$W_A = x * w_a + y * w_b + z * w_c + m * w_d \tag{3}$$

通过相同的方法求出群体内其余用户的权威程度，则该陌生关系群体的权威程度 W 为每个用户权威程度的总和；每一个用户的相应权重计算公式则为：

$$(x * w_a + y * w_b + z * w_c + m * w_d)/W \tag{4}$$

3.2　熟悉关系群体用户权重计算

某熟悉关系群体如图 4 所示。

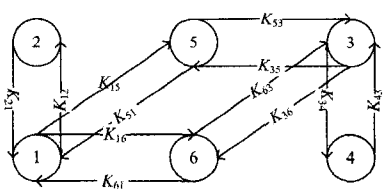

图 4　熟悉关系群体

图中的节点 1－节点 5 分别代表 5 个用户，箭头代表用户之间的熟悉关系，而箭头上的数据代表着熟悉程度。例如：K_{21} 数值为 5，则代表用户 2 对用户 1 的熟悉程度为 5。在对图 1 中的结构进行分解之后，可以计算出每个用户的权重，算法思想如下：假设要求用户 A 的权重 W_A，需要求出用户 A 对其他用户的熟悉程度 R_A。首先分析用户 A 对某一用户 B 的熟悉程度，若用户 A 与用户 B 之间存在直接路径，则用户 A 对用户 B 的熟悉程度为 K_{AB}；若用户 A 与用户 B 之间不存在直接路径，即 A 与 B 中间存在 n 个节点，那么用户 A 对用户 B 的熟悉程度为 $\max\{\sum_{1\leqslant i\leqslant n} K_i\}$。算法步骤如下：

Step1　在熟悉关系群体中遍历用户 A 和用户 B 之间的所有路径，得到路径集合 R。

Step2　若集合中存在用户 A 到用户 B 的直接路径（即两者中间无其他结点），则用户 A 对用户 B 的熟悉程度为 $R_{AB} = K_{AB}$。

Step3　若集合中不存在用户 A 到用户 B 的直接路径，但有其他 n 条包含中间节点的间接路径，遍历每条间接路径，转到 Step1。

Step4　K_i 代表第 i 条路径所计算出的熟悉程度，统计所有路径 $\sum_{1\leqslant i\leqslant n} K_i$，计算出 A 与 B 之间的熟悉程度 $R_{AB} = \max\{\sum_{1\leqslant i\leqslant n} K_i\}$。

Step5　m 代表其他用户数量，按照 Step1－Step4 的方法计算出 A 对其他用户的熟悉程度总和 R_A，$R_A = \sum_{1\leqslant j\leqslant m} R_{AJ}$。

Step6　假设群体中用户数为 x，先计算群体中所有用户的熟悉程度总和 $R = \sum_{1\leqslant i\leqslant x} R_i$，则用户 A 在群体中的权重计算公式为：

$$W_A = R_A/R \tag{5}$$

例如：在图 4 中，若要计算用户 1 的权重，则

用户 2 到用户 1：K_{21}

用户 3 到用户 1：有路径 Road(351)和 Road(361)，且

Road(351)：$\min\{K_{35}, K_{51}\}$

Road(361)：$\min\{K_{36}, K_{61}\}$

最终结果为 $\max\{\min\{K_{35}, K_{51}\}, \min\{K_{36}, K_{61}\}\}$

用户 4 到用户 1：有路径 Road(4361)和 Road(4351)，且

Road(4361)：$\min\{K_{43}, K_{36}, K_{61}\}$

Road(4351)：$\min\{K_{43}, K_{35}, K_{51}\}$

最终结果为 $\max\{\min\{K_{43}, K_{36}, K_{61}\}, \min\{K_{43}, K_{35}, K_{51}\}\}$

用户 5 到用户 1：优先选择 K_{51}

用户 6 到用户 1：优先选择 K_{61}

R_1 代表用户 1 和其他用户(2,3,4,5,6)的熟悉程度之和，且

$$R_1 = K_{21} + \max\{\min\{K_{35}, K_{51}\}, \min\{K_{36}, K_{61}\}\} + \max\{\min\{K_{43}, K_{36}, K_{61}\}, \min\{K_{43}, K_{35}, K_{51}\}\} + K_{51} + K_{61}$$

同理可求出 R_2，R_3，R_4，R_5，R_6，得 $R = R_1 + R_2 + R_3 + R_4 + R_5 + R_6$，则用户 1 的权重 $W_1 = R_1/R$。

4　群体推荐项目的确定与排序

4.1　划分用户簇

通过 k-means 聚类方法可以较好地把整个用户群体分为 n 个用户簇，簇与簇之间的相似度低，而每个簇的内部用户的相似度高。划分好之后，将同一个簇中所有用户成员的权重相加，得到该簇总的权重：

$$W = \sum_{1\leqslant i\leqslant m} w_i, m \text{ 为该簇用户数} \tag{6}$$

一个用户簇的权重大小代表着该簇对整个推荐结果的影响力。一般来说，用户的权重越大，其在群体中的地位越高，在推荐列表里的比例也越高。

4.2　确定群体推荐项目

4.2.1　评分数据推荐列表

假设某簇内总共有 s 位用户，簇内用户一起评价过的项目为 j，$j \in Item$，则第 $i(1<i<s)$ 位用户的评分向量 V_i 可以表示为 $V_i = (v_{i1}, v_{i2}, \cdots, v_{ij})$。由于第 3 节已计算出簇内各用户的权重，该用户的权重 W_i 已知，进而通过对其聚合可以得到整个簇的评分向量 V_{total}，计算公式为：

$$V_{total} = \sum_{i=1}^{s} V_i \times W_j/s \tag{7}$$

将 V_{total} 视为一个虚拟用户，通过比较 V_{total} 与其他用户的 Pearson 相关相似度，选出相似度最高的 top8 作为 V_{total} 的邻居集合 V_{ne}，在 V_{ne} 中某一用户 u 给项目 i 的评分可以用 $rec_{u,i}$ 表示，虚拟用户 V_{total} 与另一待比较相似度的用户 $u1$ 之间的相似度值则用 $S_{V_{total},u1}$ 表示，那么 V_{total} 对某一项目 j 的爱好程度的计算公式为：

$$Pear_{V_{total},j}=\frac{\sum_{neN} S_{V_{total},N}(rec_{N,j}-\overline{rec_N})}{\sum_{neN}|S_{V_{total},N}|}+\overline{rec_{V_{total}}} \tag{8}$$

计算出 V_{total} 对每个项目的预测评分后，将项目按分值降序排列，选择排名前 8 位的项目组合成集合 R_{score} 推荐列表并推荐给 V_{total}。由于 V_{total} 是虚拟用户，其代表一个相似用户簇，因此该推荐列表也等同于推荐给相似簇内的所有用户。

4.2.2 Label 信息推荐列表

从各个相似用户簇里抽取 Label 信息以构建每个簇的初始 Label 信息集合，用 L_{all} 表示；然后逐个统计各初始 Label 信息在 L_{all} 中的出现频数并按降序排列，选取 top8 构成集合 L_8，并将它作为该簇所有用户爱好的 Label 信息代表；最后，用 L_8' 来代表各 Label 信息集合中的所有频率的列向量，用 Label 的频数作为对应 Label 的权重 $W=\log L_8'$。

在得到对应 Label 的权重之后，以余弦相似度的计算方法来衡量 Label 之间的相似程度，以此推断出项目间的相似度。V_i 是推荐集合中第 i 个项目的空间向量，则该项目和 L_8 所代表的簇的相似度为[11-13]：

$$Sim=\frac{\vec{V_i}*\vec{L_8}}{|\vec{V_i}|*|\vec{L_8}|} \tag{9}$$

将计算出的值按降序排列，选取前 8 位作为推荐项目集合，用 R_{label} 表示。

取评分数据推荐列表和 Label 信息推荐列表的交集作为最终推荐项目集合，用 R_{final} 表示，则：

$$R_{final}=R_{label}\cup R_{score} \tag{10}$$

整个算法的步骤如下：

Step1 从各个相似簇中抽取 Label 信息构成爱好矩阵。

Step2 以 Label 的出现频数作为评判标准计算 Label 的权重。

Step3 通过余弦相似度计算各 Label 之间的相似度并按降序排列，从而得到 Label 信息的推荐列表。

Step4 根据权重计算算法计算出用户权重，聚合得到相似用户簇的评分向量。

Step5 通过 Pearson 相似度计算出各用户对应的相似度，按降序排列得到评分推荐列表。

Step6 取 Label 信息的推荐列表和评分推荐列表的交集作为最终推荐列表。

5 检验标准与实验结果

5.1 检验标准

5.1.1 平均绝对偏差(MAE)

$$MAE=\frac{\sum_{u,i\in T}|r_{ui}-\hat{r}_{ui}|}{|T|} \tag{11}$$

MAE 被用来检测真实评价值和推荐值之间的偏差，u 代表评估用户数，T 代表项目数[14-16]。MAE 的值越小，表示预测值与真实值越接近，也说明算法的精确度越高。另外，平方根误差(RMSE)也可用于检验推荐算法的精确度，计算公式为：

$$RMSE=\frac{\sqrt{\sum_{u,i\in T}(r_{ui}-\hat{r}_{ui})^2}}{|T|} \tag{12}$$

5.1.2 准确率(Precision)

该指标用于检测算法推荐的准确度，用推荐给用户的群体用户项目中被用户所喜爱的项目数目 $T(u)$ 与推荐项目总数 $R(u)$ 的比例来衡量，计算公式[17-20]为：

$$Precision=\frac{\sum_{u\in U}|R(u)\cap T(u)|}{\sum_{u\in U}|R(u)|} \tag{13}$$

5.1.3 召回率(Recall)

召回率是评价推荐系统的标准之一，具体计算公式[21-23]为：

$$Recall=\text{提取出的命中项目数}\ /\ \text{样本项目总数} \tag{14}$$

5.2 数据预处理和参数设置

本实验使用 KDD Cup 2012 Track1 中的腾讯微博数据集，由于其数据量较大，首先需要进行 SVD 矩阵分解以达到降维除噪的目的，然后利用 k-means 聚类方法把整个群体分为若干个簇，使簇与簇之间的相似度较低，而每个簇的内部用户相似度较高。群体规模参数设置为 θ，θ 的值可以设为 4，8，12，16，20，24。通过调查收集信息，整合 28 名计算机专业的硕士生对于打分次数、Label 次数、分值内聚度、Label 覆盖率这 4 个指标的分析数据，并通过 AHP 层次分析法计算出这 4 个指标的对应权重分别是：0.223，0.285，0.364，0.128。最后通过 k-means 方法，把整个群体分解成成员数可能不尽相同的 N 个用户簇，每个用户簇的所有用户的权重之和即为该簇的权重。

5.3 实验结果与分析

本实验结果主要反映群体规模和用户簇的数值分别对整个推荐结果产生的各种影响。

5.3.1 用户簇数量固定，群体规模不断扩大

如图 5 所示，当用户簇的数量设为 6 时，SVD 群体推荐算法和 NMF 群体推荐算法所计算的 MAE 值会随着群体规模的增大呈现出增加趋势，其中 SVD 群体推荐算法递增速率逐步减小。分析可知，当用户的群体规模扩大时，由于相应的用户簇数量不变，每个簇内的用户数增多，导致每个簇中信息的复杂程度和多元性提高，从而使得 MAE 值的递增速率降低；而引进了 Label 信息的群体推荐算法由于信息复杂程度导致 MAE 值在刚开始会呈现下降趋势，但是随着规模的不断增大，MAE 值会随着群体成员复杂程度和多元性的增加而递增。

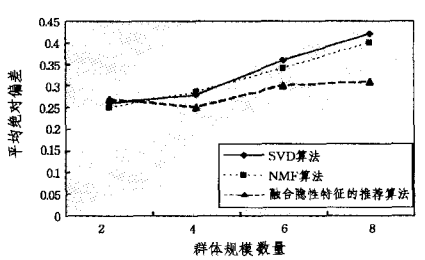

图 5 群体规模扩大时 MAE 值的变化

如图 6 所示，SVD 群体推荐算法和 NMF 群体推荐算法的准确率会一直持续下降，而考虑了 Label 信息的算法的准确率则是先上升后下降。

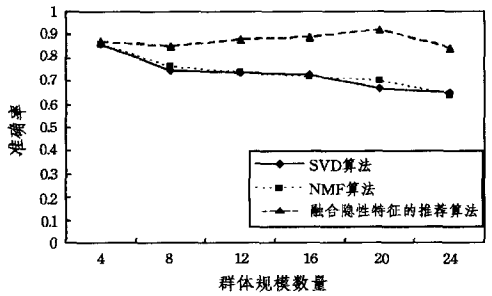

图 6　群体规模扩大时准确率的变化

如图 7 所示，3 种类型的算法的召回率均会随着群体规模扩大而下降，究其原因是用户爱好的多元性和种类复杂性随着规模扩大会不断增加。

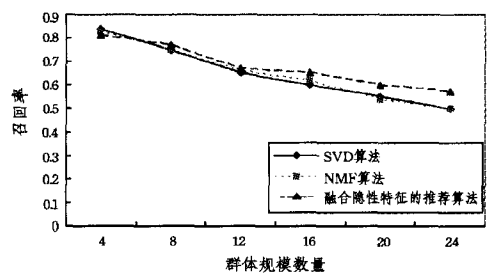

图 7　群体规模扩大时召回率的变化

5.3.2　*群体规模固定，用户簇数量变化*

如图 8、图 9 所示，随着用户簇数的增加，考虑了 Label 信息的算法的 MAE 值明显比 SVD 群体推荐算法和 NMF 群体推荐算法更为稳定，这是因为 Label 信息包含了更为全面的用户爱好信息，因此也使准确率得到了提高。

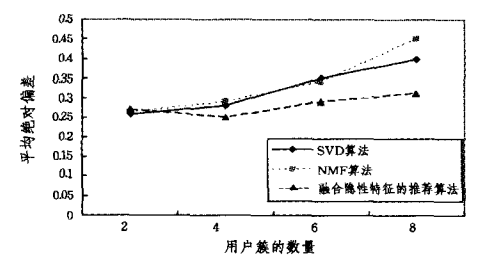

图 8　用户簇数量增加时 MAE 值的变化

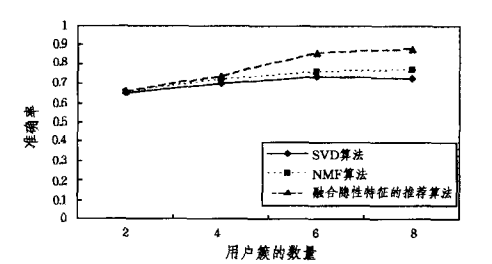

图 9　用户簇数量增加时准确率的变化

如图 10 所示，3 种算法的召回率随着用户簇数量的增多均呈现递减趋势，但引入 Label 信息的算法的递减速率较为缓慢。

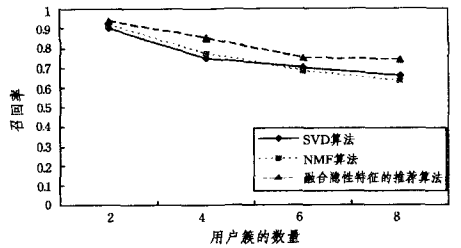

图 10　用户簇数量增加时召回率的变化

结束语　本文针对主流群体推荐算法的典型——SVD 群体推荐算法和 NMF 群体推荐算法在群体规模快速增长、用户簇逐渐增多的情况下推荐准确性下降、性能不稳定等问题，提出了一种基于隐性特征信息建立的用户权重计算方法。在上述实验数据上进行的实验表明，在群体规模不断扩大、用户簇数量增加时，基于隐性特征的群体推荐算法的平均绝对偏差值比 SVD 群体推荐算法和 NMF 群体推荐算法的更为稳定，准确率更高，而召回率差别则不明显。本文的创新点在于以隐性特征为主要参考标准综合用户反馈信息，并在此基础上建立针对不同群体情况的用户权重计算方法以提高推荐的准确率和稳定性。如何筛选最优隐性特征，如何结合显性特征建立自适应性的权重计算方法，还有待进一步的探索和研究。

参 考 文 献

[1] LIU Q, CHEN E H, XIONG H, et al. Enhancing collaborative filtering by user interest expansion[J]. IEEE Transactions on Systems, Man, and Cybernetics-part B: Cybernetics, 2012, 42(1): 218-233.

[2] KONSTAN J A, RIEDL J T. Recommender systems: from algorithms to user experience[J]. User Modeling and User-Adapted Interaction, 2012, 22(1): 101-123.

[3] CACHEDA F, CARNEIRO V, FERNANDEZ D, et al. Comparison of collaborative filtering algorithms: limitations of current techniques and proposals for scalable, high-performance recommender systems[J]. ACM Transactions on the Web, 2011, 5(2): 1-33.

[4] ARMSTRONG J R, WADE J. Development of Systems Engineering Expertise[J]. Procedia Computer Science, 2015, 44(2): 689-698.

[5] ZHENG X, CHEN C C, HUNG J L, et al. A Hybrid Trust-based Recommender System for Online Communities of Practice [J]. IEEE Transactions on Learning Technologies, 2015, 8(4): 345-356.

[6] COLOMBO-MENDOZA L O, VALENCIA-GARCÍA R, RODRÍGUEZ-GONZÁLEZ A, et al. RecomMetz: A context-aware knowledge-based mobile recommender system for movie showtimes[J]. Expert Systems with Applications, 2015, 42(3): 1202-1222.

[7] ZHU Y X, LV L Y. A review of the evaluation index of the recommendation system[J]. Journal of the University of Electronic Science and technology of University of science and techno-

logy,2012,41(2):163-175. (in Chinese)
朱郁筱,吕林媛. 推荐系统评价指标综述[J]. 电子科技大学学报,2012,41(2):163-175.

[8] TEJEDA-LORENTE Á, PORCEL C, PEIS E, et al. A quality based recommender system to disseminate information in a university digital library[J]. Information Sciences, 2014, 261(5): 52-69.

[9] ANBARASU V, LINDA X, MAHALAKSHMI S. An Efficient Recommender System based on Collaborative Filtering[J]. Asian Journal of Applied Sciences, 2015, 3(1).

[10] MIU P. Research on Weibo user interest model based on information push technology [D]. Wuhan: Wuhan University of Technology, 2012. (in Chinese)
谬平. 基于微博用户兴趣模型的信息推送技术的研究[D]. 武汉:武汉理工大学,2012.

[11] MONTUSCHI P, LAMBERTI F, GATTESCHI V, et al. A Semantic Recommender System for Adaptive Learning[J]. It Professional, 2015, 17(5): 50-58.

[12] MEHDI M, BOUGUILA N, BENTAHAR J. Probabilistic approach for QoS-aware recommender system for trustworthy web service selection[J]. Applied Intelligence, 2014, 41(2): 503-524.

[13] YU Z W, LI L, WONG H S, et al. Probabilistic cluster structure ensemble[J]. Information Sciences, 2014, 267(5): 16-34.

[14] SIHME A, LEHOCINEM B, MINIAIH A. Batch Adsorption of Phenol From Industrial Waste Water Using Cereal By-Products As A New Adsrbent[J]. Energy Procedia, 2012, 18(4): 1135-1144.

[15] CAO Y M. Research and implementation of personalized news recommendation algorithm based on collaborative filtering [D]. Beijing: Beijing University of Posts and Telecommunications 2013. (in Chinese)
曹一鸣. 基于协同过滤的个性化新闻推荐算法的研究与实现[D]. 北京:北京邮电大学 2013.

[16] ALPHY A, PRABAKRAN S. A Dynamic Recommender System for Improved Web Usage Mining and CRM Using Swarm Intelligence[J]. Scientific World Journal, 2015, 2015: 1-16.

[17] XIA P Y. Collaborative filtering algorithm in personalized recommendation technology [D]. Qingdao: Ocean University of China, 2011. (in Chinese)
夏培勇. 个性化推荐技术中的协同过滤算法研究[D]. 青岛:中国海洋大学,2011.

[18] RAHMAN R M, SIDDIQUEE M R, HAIDER N. Movie Recommendation System Based on Fuzzy Inference System and Adaptive Neuro Fuzzy Inference System[J]. International Journal of Fuzzy System Applications, 2015, 4(4): 31-69.

[19] CHEN L, CHEN G, WANG F. Recommender systems based on user reviews: the state of the art[J]. User Modeling and User-Adapted Interaction, 2015, 25(2): 99-154.

[20] WANG L C, MENG X W, ZHANG Y J. Chinese Journal of context aware recommendation system[J]. Software, 2012(1): 1-20. (in Chinese)
王立才,孟祥武,张玉洁. 上下文感知推荐系统[J]. 软件学报,2012(1):1-20.

[21] HUANG Q. Personalized recommendation algorithm of network book resources [D]. Chengdu: Southwest Jiaotong University, 2014. (in Chinese)
黄琼,网络图书资源个性化推荐算法研究[D]. 成都:西南交通大学,2014.

[22] DING B Z. Collaborative filtering algorithm based on citation information[D]. Jilin: Jilin University, 2014. (in Chinese)
丁彬钊. 基于引文信息的协同过滤算法研究[D]. 吉林:吉林大学,2014.

[23] CHEN N Y. Design and implementation of recommendation system based on personalized recommendation engine[D]. Guangzhou: South China University of Technology, 2012. (in Chinese)
陈诺言. 基于个性化推荐引擎组合的推荐系统的设计与实现[D]. 广州:华南理工大学,2012.

(上接第 219 页)

[12] HERZIG D M, MIKA P, BLANCOR, et al. Federated entity search using on-the-fly consolidation[M]// The Semantic Web—ISWC 2013. 2013: 167-183.

[13] DONG X, NAUMANN F. Data fusion-resolving data confiicts for integration[C]// VLDB. 2009.

[14] ZHANG Y X, LI Q Z, PENG Z H. 2-Stage Data Conflict Resolution Based on Markov Logic Networks[J]. Chinese Journal of Computers, 2012, 35(1): 101-111. (in Chinese)
张永新,李庆忠,彭朝晖. 基于 Markov 逻辑网的两阶段数据冲突解决方法[J]. 计算机学报,2012,35(1):101-111.

[15] GUO S, DONG X, SRIVASTAVA D, et al. Record linkage with uniqueness constraints and erroneous values[J]. PVLDB, 2010, 3(1): 417-428.

[16] LIU X, DONG X L, OOI B C, et al. Online data fusion[J]. PVLDB, 2011, 4(12): 932-943.

[17] FAN W F, GEERTS F, TANG N, et al. Inferring data currency and consistency for conflict resolution[C]// ICDE. 2013: 470-481.

[18] LI M H, LI J Z, GAO H. Evaluation of Data Currency[J]. Chinese Journal of Computers, 2012, 35(1): 2348-2360. (in Chinese)
李默涵,李建中,高宏. 数据时效性判定问题的求解算法[J]. 计算机学报,2012,35(11):2348-2360.

[19] FAN W F, GEERTS F, et al. Determining the currency of data [C]// PODS. 2011.