

# 考虑负相关性信息的协同过滤算法研究\*

郭强, 周继平, 郭迎迎, 胡兆龙, 刘建国

(上海理工大学复杂系统科学研究中心, 上海 200093)

**摘要:** 为了研究 Pearson 负相关性信息对协同过滤算法的影响, 提出了一种考虑负相关性信息的协同过滤算法。该算法选取正相关用户作为最近邻居, 负相关用户作为最远邻居, 使用参数调节最近邻居和最远邻居在推荐过程中的作用。MovieLens 数据集上的对比实验表明, 负相关性不仅可以提高推荐结果的准确性, 而且可以增加推荐列表的多样性; 进一步分析发现, 负相关性还可以大幅度提高不活跃用户的推荐准确性。该工作表明, 负相关性有助于解决推荐系统中准确性、多样性两难的问题和冷启动问题。

**关键词:** 协同过滤; 负相关性; 最远邻居集

**中图分类号:** TP311; TP301.6

**文献标志码:** A

**文章编号:** 1001-3695(2013)12-3543-03

doi:10.3969/j.issn.1001-3695.2013.12.006

## Collaborative filtering algorithm by considering negative correlations

GUO Qiang, ZHOU Ji-ping, GUO Ying-ying, HU Zhao-long, LIU Jian-guo

(Research Center of Complex System Science, University of Shanghai for Science & Technology, Shanghai 200093, China)

**Abstract:** In order to study the effect of the negative correlation of Pearson to collaborative filtering algorithm, this paper presented an improved collaborative filtering algorithm. Firstly this algorithm selected the positive and negative correlation users as the nearest neighbor set and furthest neighbor set respectively, and then made use of a tunable parameter to adjust the effect of the nearest and furthest neighbor set on recommendation. The experiment results on MovieLens dataset show that negative correlations can not only significantly improve the accuracy of recommendations, but also increase the diversity of recommendation lists. It has found that collaborative filtering algorithm by considering the negative correlations can greatly improve the recommendation accuracy of users with small degrees. This work suggests that the negative correlations help solve the dilemma of accuracy and diversity and cold start problem of the recommender systems.

**Key words:** collaborative filtering; negative correlation; furthest neighbors set

个性化推荐算法利用用户的历史选择信息预测其喜好, 成为解决信息过载的有效手段之一。目前已经提出了多种推荐算法, 如协同过滤算法<sup>[1]</sup>、基于内容的推荐算法<sup>[2]</sup>、混合推荐算法<sup>[2]</sup>和基于网络结构的推荐算法<sup>[3]</sup>等。其中协同过滤是推荐系统中运用最广泛, 也是最成功的一种个性化推荐算法。在基于协同过滤的推荐系统中, K-近邻模型是使用最广泛的经典模型<sup>[4]</sup>, 其主要原理是利用评分相似度构造 K 个最近邻居, 然后根据最近邻对用户进行推荐。传统协同过滤算法常常利用 Pearson 相关系数<sup>[2]</sup>计算相似度, 根据相似度大小选择 K-近邻, 往往只考虑 Pearson 相关系数的正相关性, 没有考虑其负相关性。某些文献虽然考虑到了负相关性, 但是都没能表示出负相关性的真实含义, 比如认为负相关性对推荐结果影响不大, 进而将其视为无效值直接舍弃掉<sup>[5]</sup>或使用绝对值来代替<sup>[1]</sup>或将其投影成正相关性<sup>[6]</sup>。因此为了研究负相关性对协同过滤算法的影响, 本文提出了一种考虑负相关性信息的协同过滤 (collaborative filtering by considering negative correlation, CBCF) 算法, 该算法在选择 K-近邻时, 同时考虑用户的正负相关邻居, 然后同时利用正负相关邻居产生推荐。

## 1 相关工作

已有的对协同过滤算法的改进主要集中在数据稀疏性问题和相似性度量问题。针对稀疏性问题, 研究者提出了很多解决方法, 其中包括矩阵填充<sup>[7]</sup>和矩阵降维<sup>[8]</sup>。关于相似性度量的已有研究包括向相似性<sup>[9]</sup>和项目间相似性<sup>[10]</sup>等。

本文主要从邻居选择的角度对协同过滤算法进行改进。邻居选择的研究主要包括以下几个方面: 罗辛等人<sup>[11]</sup>提出相似度支持度的概念并将其作为一种新的 K 近邻度量; 黄创光等人<sup>[12]</sup>提出不确定近邻的协同过滤算法, 自适应地选择预测目标的近邻对象作为推荐群; 张尧等人<sup>[13]</sup>综合考虑邻居选择权重, 提出了基于用户分类的邻居选择方法; Zeng 等人<sup>[14]</sup>研究了不相似用户对传统协同过滤算法的影响。当用户评分数据相当稀疏或者分布不均衡时, Pearson 相似度值可能为负值, 现有文献往往使用其绝对值代替<sup>[1]</sup>或直接舍弃掉<sup>[5]</sup>。针对负相关性, Wang 等人<sup>[6]</sup>将负相似度映射到区间 [0, 1] 上, 正相似度映射到区间 (0.5, 1] 上。不论是使用绝对值还是映射到指定

**收稿日期:** 2013-04-22; **修回日期:** 2013-05-28 **基金项目:** 国家自然科学基金资助项目 (91024026, 71071098, 71171136); 上海市科研创新基金资助项目 (11ZZ135, 11YZ110); 国家教育部科学技术研究重点资助项目 (211057); 上海市系统科学一流学科建设项目 (XTKX2012); 上海市青年科技启明星计划资助项目 (A) 类 (11QA1404500)。

**作者简介:** 郭强 (1975-), 女, 辽宁大连人, 副教授, 硕导, 主要研究方向为数据挖掘 (qiang.guo@usst.edu.cn); 周继平 (1988-), 男, 湖北孝感人, 硕士研究生, 主要研究方向为推荐算法; 郭迎迎 (1988-), 女, 河南安阳人, 硕士研究生, 主要研究方向为推荐算法; 胡兆龙 (1988-), 男, 江西南昌人, 硕士研究生, 主要研究方向为社会网络分析; 刘建国 (1979-), 男, 山西临汾人, 教授, 博导, 博士, 主要研究方向为网络科学、个性化推荐。

区间,均没有考虑到 Pearson 负相似度信息。

综上所述,关于邻居选择的研究,较少关注到 Pearson 相关系数的负相关性<sup>[1,5,6]</sup>,如直接舍弃掉<sup>[5]</sup>、使用绝对值代替<sup>[1]</sup>或者投影成正相似度<sup>[6]</sup>。为此本文提出了考虑 Pearson 负相关性的改进协同过滤算法,从负相关性的真实含义出发,将负相关用户集成到协同过滤算法中,与正相关用户一起产生推荐。

## 2 传统的协同过滤算法

协同过滤算法一般分为三步:

a) 数据表述。给定用户集  $U$  和项目集  $I$ , 则用户对于项目的兴趣可以表示为一个  $m \times n$  的矩阵  $R$ 。在该矩阵中,  $m$  代表用户数, 每一个行向量表示特定用户的评分集合;  $n$  代表项目数, 每一个列向量表示特定项目的被评分集合, 每一元素  $r_{ui} \in R$  表示用户  $u$  对项目  $i$  的评分。

b) 近邻选择。按照相似度从大到小为当前用户或项目选择最近邻集合。

c) 推荐产生。利用最近邻居评分的加权平均值预测目标用户未评分项目的评分, 根据预测评分值可以得到用户对任意项目的兴趣度及其推荐集。

### 2.1 计算相似度

传统的相似性度量方法主要有余弦相似性 (cosine)<sup>[3]</sup>、修正的余弦相似性 (adjusted cosine)<sup>[3]</sup> 和相关相似性 (correlation)<sup>[4]</sup>, 其中 Breese 等人发现相关相似性优于余弦相似性<sup>[4]</sup>, 本文在此仅介绍相关相似性。用户  $u_i$  和  $u_j$  的相似性可以通过计算 Pearson 相关系数得到。假设  $I_{ij}$  表示用户  $u_i$  和  $u_j$  共同评分的项目集合, 那么用户  $u_i$  和  $u_j$  的相似度为

$$\text{sim}(u_i, u_j) = \frac{\sum_{c \in I_{ij}} (R_{ic} - \bar{R}_i)(R_{jc} - \bar{R}_j)}{\sqrt{\sum_{c \in I_{ij}} (R_{ic} - \bar{R}_i)^2} \sqrt{\sum_{c \in I_{ij}} (R_{jc} - \bar{R}_j)^2}} \quad (1)$$

其中:  $R_{ic}$  和  $R_{jc}$  分别表示  $u_i$  和用户  $u_j$  对项目  $c$  的评分;  $\bar{R}_i$  和  $\bar{R}_j$  分别表示用户  $u_i$  和  $u_j$  对已评分项目的平均评分。

### 2.2 选择近邻

邻居选择通常采用 K-近邻法, 即选择与当前用户相似度最大的前  $k$  个用户作为邻居。对一个活动用户  $a$ , 要产生一个依相似度由大到小排列的邻居集合  $U = \{u_1, u_2, \dots, u_k\}, a \notin U$ 。

### 2.3 产生推荐

利用最近邻居预测目标用户对未评分项目的评分, 从而形成 top-N 推荐。目标用户  $u_i$  对项目  $i$  的评分  $P_{u_i,i}$  预测为

$$P_{u_i,i} = \bar{R}_{u_i} + \frac{\sum_{v \in N_{u_i}} \text{sim}(u_i, v) \times (R_{v,i} - \bar{R}_v)}{\sum_{v \in N_{u_i}} (|\text{sim}(u_i, v)|)} \quad (2)$$

其中:  $N_{u_i}$  为用户  $u_i$  的最近邻居集合;  $\text{sim}(u_i, v)$  为用户  $u_i$  和用户  $v$  的相似度;  $\bar{R}_{u_i}$  为用户  $u_i$  的平均评分;  $\bar{R}_v$  为用户  $v$  的平均评分;  $R_{v,i}$  为用户  $v$  对项目  $i$  的评分。

## 3 考虑负相关性信息的协同过滤算法

### 3.1 传统协同过滤算法的不足

基于 Pearson 的最近邻协同过滤算法在选取邻居时, 忽视了 Pearson 相关系数的两面性, 只考虑 Pearson 相关系数的正相关性, 而没有考虑其负相关性。下面用一个实例来更好地说明

该问题。

图1表示用户  $a, b, c$  对项目  $d, e, f, g$  的评分。采用 Pearson 相似度计算图1中用户  $a, b$  和  $c$  之间的两两相似度。其中用户  $a$  和其他用户间的相似度分别为:  $\text{sim}(a, b) = 0.8, \text{sim}(a, c) = -0.94$ 。由式(1)可知, Pearson 相关系数  $r \in [-1, 1], r > 0$  表明用户在共同评分项目上的评分是正相关的。如图1中,  $\text{sim}(a, b) = 0.8, a$  和  $b$  在共同评分项目上的评分正相关, 即一方对某个项目给高分 (低分) 时, 另一方也倾向于给高分 (低分), 因此可以利用用户  $b$  对项目  $f$  的评分来预测用户  $a$  对项目  $f$  的评分。同理,  $r < 0$  表示用户在共同评分项目上的评分是负相关的。如图1中,  $\text{sim}(a, c) = -0.94$ , 用户  $a$  和  $c$  在共同评分项目上的评分负相关, 即一方对某个项目给高分 (低分) 时, 另一方倾向于给低分 (高分), 因此可以利用用户  $c$  对项目  $f$  的评分来预测用户  $a$  对项目  $f$  的评分。

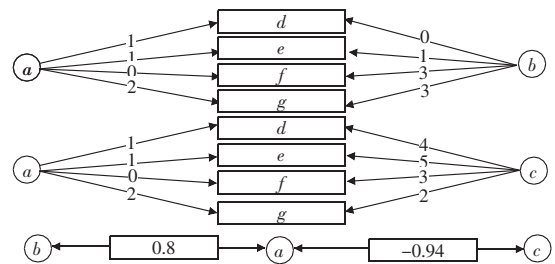


图1 用户评分示意图

由上述分析可知, 基于 Pearson 的最近邻协同过滤算法未充分利用负相关性信息, 因此本文对传统协同过滤算法在邻居选择和评分预测两个步骤上进行改进。综合考虑 Pearson 相关系数的正负两面性, 将负相关用户集成到协同过滤算法中, 提出了考虑负相关性信息的协同过滤算法。

### 3.2 算法介绍

#### 3.2.1 邻居选取

a) 最远邻居 (furthest neighbor)。与当前用户的 Pearson 相似度为负值的用户集合, 即负相关用户集合。

$$\text{FN}(u_i) = \{u \mid \text{sim}(u_i, u_j) < 0\} \quad i \neq j \quad (3)$$

其中:  $\text{FN}(u_i)$  为  $u_i$  的最远邻居集合;  $\text{sim}(u_i, u_j)$  为用户  $u_i$  和  $u_j$  的 Pearson 相似度值。

b) K-最远邻居集合 (K-furthest neighbor)。从最远邻居集中选择前  $K$  个绝对值最大的用户所组成的集合。

$$\text{KFN}(u_i) = \{u_1, u_2, \dots, u_k\} \quad u_i \notin \text{KFN}(u_i) \quad (4)$$

其中: 用户  $u_j (1 \leq j \leq k)$  按照与  $u_i$  的相似度绝对值由大到小排列。

#### 3.2.2 评分预测

本文将最远邻居集成到协同过滤算法中, 故采用如下公式进行评分预测。

$$P_{u_i,i} = \bar{R}_{u_i} + (1 - \alpha) \frac{\sum_{v \in N_{u_i}} \text{sim}(u_i, v) \times (R_{v,i} - \bar{R}_v)}{\sum_{v \in N_{u_i}} (|\text{sim}(u_i, v)|)} + \alpha \frac{\sum_{w \in \text{FN}_{u_i}} \text{sim}(u_i, w) \times (R_{w,i} - \bar{R}_w)}{\sum_{w \in \text{FN}_{u_i}} (|\text{sim}(u_i, w)|)} \quad (5)$$

其中:  $\text{FN}_{u_i}$  为用户  $u_i$  的最远邻居;  $\alpha$  为阈值, 用于调节最近邻居和最远邻居的作用。当  $\alpha = 0$  时, 推荐完全按照最近邻居进行, 退化成传统协同过滤算法; 当  $\alpha = 1$  时, 推荐完全根据最远邻居进行。  $\alpha$  取值范围为  $\alpha \in [0, 1]$ 。

## 4 实验结果分析

### 4.1 数据集

本文使用 MovieLens 数据集来测试改进后的算法,该数据集是由 GroupLens 研究产品组 (<http://www.grouplens.org>) 提供的一个著名电影评分数据集,包含 943 个用户对 1682 个电影的 10 万条打分记录,且每个用户至少对 20 部电影进行过评分,用户会对自己看过的电影按照 5 分制打分,1 分表示最不喜欢,5 分表示最喜欢。本实验从 10 万条记录中随机选取 80% 作为训练集,剩下的 20% 作为测试集。

### 4.2 评价标准

1) 平均绝对误差<sup>[15]</sup> 本文使用平均绝对误差 (mean absolute error, MAE) 来衡量推荐算法的准确性。MAE 通过计算预测的用户评分与实际用户评分之间的偏差来度量预测的准确性。预测的用户评分集合为  $\{p_1, p_2, \dots, p_n\}$ , 相应的实际用户评分集合为  $\{q_1, q_2, \dots, q_n\}$ , 则 MAE 通常定义为

$$MAE = \frac{\sum_{i=1}^N |p_i - q_i|}{N} \quad (6)$$

显然, MAE 越小, 表明评分预测的偏差越小, 算法准确性越好。

2) 平均 Hamming 距离<sup>[15]</sup> 本文使用平均 Hamming 距离 (average Hamming distance) 度量推荐列表的多样性。用户  $u_i$  和  $u_j$  的推荐列表的多样性被定义为

$$H_{ij} = 1 - Q_{ij}/L \quad (7)$$

其中:  $L$  为推荐列表的长度;  $Q_{ij}$  为用户  $u_i$  和  $u_j$  的推荐列表中相同项目的个数。推荐列表的多样性定义为  $H_{ij}$  的平均值  $\langle H \rangle$ , 可见  $\langle H \rangle$  越大, 推荐列表的多样性越好。

### 4.3 实验结果及分析

#### 4.3.1 Pearson 相似度值分布

本文采用式(1)计算训练集中用户间的两两相似度, 相似度值的分布如图2所示。负相似度值在协同过滤算法中未得到充分利用, 往往被直接舍弃, 或使用其绝对值或投影参加到推荐过程中。如果用户相似度矩阵中负相似度值所占的比例较少, 那么毫无疑问可以直接舍弃掉, 然而从图2中可以看到, 用户相似度矩阵中有 35% 的相似度值小于 0, 为负值, 数量还是相当多的, 因此不能随意舍弃掉这些为数不少的负相似度值。

#### 4.3.2 参数 $\alpha$ 估计

考虑负相关性信息的协调过滤算法有邻居个数  $k$  和用于调节最近邻居和最远邻居作用的阈值  $\alpha$  两个参数, 本实验通过设置不同的  $k$  值和  $\alpha$  值, 估计到一个较优的参数值  $\alpha$ , 然后应用到本文其他实验当中。本实验计算不同邻居个数  $k$  和阈值  $\alpha$  下的 MAE, 邻居个数  $k$  从 10 ~ 50, 间隔为 10, 阈值  $\alpha$  从 0 ~ 1, 间隔为 0.1, 实验结果如图2所示。

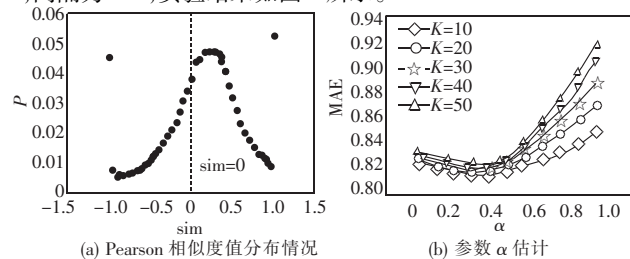


图2 相似度值分布图

由图2可知, 在不同的邻居个数  $k$  下,  $\alpha = 0.3$  时, MAE 都具有最小值, 由此可知在该数据集下  $\alpha = 0.3$  时 CBCF 算法的准确性最好, 因此在后续实验中选择参数  $\alpha$  的值为 0.3。

#### 4.3.3 准确性比较

为了检验本文提出算法的准确性, 本文在同等数据集的基础上变换邻居个数, 比较 CBCF 算法和基于 Pearson 相似度的传统协同过滤算法 (standard collaborative filtering, SCF), 计算 MAE, 邻居个数  $k$  从 10 ~ 30, 间隔为 4。由图3(a)可知, 在各种实验条件下, 与 SCF 算法相比, 本文提出的 CNCF 均具有较小的 MAE。由此可知, 本文提出的考虑负相关性信息的协同过滤算法能够明显提高推荐准确度, 这表明考虑负相关性可以提高算法的准确性。

#### 4.3.4 多样性比较

为了检验本文提出算法的多样性, 将 CBCF 和 SCF 算法进行比较, 以平均 Hamming 距离作为评价指标, 邻居个数从 10 到 70, 间隔为 10, 推荐列表长度  $L = 50$ 。由图3(b)可知, 当邻居个数  $k$  大于 40 时, CBCF 算法具有较大的平均 Hamming 距离。随着邻居个数的增加, CBCF 算法的多样性呈上升趋势, 而 SCF 算法却呈下降趋势, CBCF 算法的提升幅度越来越大, 这表明考虑负相关性可以提高算法的多样性。

给用户推荐流行产品, 可以提高准确性, 但是会让用户的视野变得狭窄; 而为了提高推荐的多样性, 给用户推荐一个冷门产品或者打分很低的产品, 很容易引起用户的反感。准确性和多样性之间存在竞争关系, 通常只能牺牲精确性来提高多样性或者牺牲多样性来提高精确性<sup>[16]</sup>。本文发现, 与传统协同过滤算法相比, 考虑负相关性信息的协同过滤算法将负相关用户集成到推荐过程中, 可以同时提高推荐的准确性和多样性。

#### 4.3.5 负相关性信息推荐准确性的影响

用户的度表示该用户选择过多少产品<sup>[3]</sup>。为了验证负相关性对不同用户的影响, 分别计算训练集中度最大的前 100 个用户和度最小的前 100 个用户的 MAE。由图3(c)(d)可知, 与活跃用户相比, 考虑负相关性信息的协同过滤算法的准确性对不活跃用户的提升幅度更大, 比如当  $k = 16$  时, 不活跃用户的提升比例可以达到 3.01%, 而活跃用户的提升比例仅为 0.31%, 这表明负相关性可以提高协同过滤算法对不活跃用户的预测准确性。

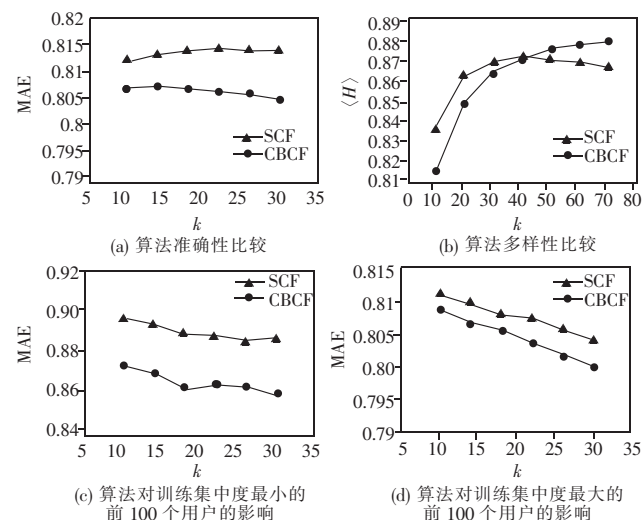


图3 在 SCF 和 CBCF 两种算法中, MAE 及 Hamming 距离随邻居个数  $K$  的变化情况

- nary Computation, 2011, 1(1): 32-49.
- [3] DEB K, PRATAP A, AGARWAL S, *et al.* A fast and elitist multi-objective genetic algorithm: NSGA-II [J]. *IEEE Trans on Evolutionary Computation*, 2002, 6(2): 182-197.
  - [4] ZITZLER E, LAUMANN S, THIELE L. SPEA2: improving the strength Pareto evolutionary algorithm [C]//Proc of Evolutionary Methods for Design, Optimization and Control with Applications to Industrial Problems. Berlin: Springer-Verlag, 2002: 95-100.
  - [5] CORNE D W, KNOWLES J D, OATES M J. The Pareto envelope-based selection algorithm for multi-objective optimization [C]//Proc of Parallel Problem Solving from Nature-PPSN VI. 2000: 839-848.
  - [6] KNOWLES J D, CORNE D W. The Pareto archived evolutionary strategy: a new baseline algorithm for Pareto multiobjective optimization [C]//Proc of Congress on Evolutionary Computation. Piscataway, NJ: IEEE Press, 1999: 98-105.
  - [7] NAIN P K S, DEB K. A multi-objective search and optimization procedure with successive approximate models, KanGAL Report 2004012 [R]. [S. l.]: Indian Institute of Technology Kanpur, 2004.
  - [8] JIN Yao-chu, SENDOFF B. A systems approach to evolutionary multi-objective structural optimization and beyond [J]. *IEEE Computational Intelligence Magazine*, 2009, 4(3): 62-76.
  - [9] SHI L, RASHEED K. A survey of fitness approximation methods applied in evolutionary algorithms [C]//Proc of Computational Intelligence in Expensive Optimization Problems. 2010: 3-28.
  - [10] JIN Yao-chu. Surrogate-assisted evolutionary computation: recent advances and future challenges [J]. *Swarm and Evolutionary Computation*, 2011, 1(2): 61-70.
  - [11] LIM D, JIN Yao-chu, ONG Y S, *et al.* Generalizing surrogate-assisted evolutionary computation [J]. *IEEE Trans on Evolutionary Computation*, 2010, 14(3): 329-355.
  - [12] LIM D, ONG Y S, JIN Yao-chu, *et al.* A study on meta-modeling techniques, ensembles, and multi-surrogates in evolutionary computation [C]//Proc of the 9th Annual Conference on Genetic and Evolutionary Computation. London: ACM Press, 2007: 1288-1295.
  - [13] GUO Guan-qi, LI Wu, YANG Bo, *et al.* Predicting Pareto dominance in multi-objective optimization using pattern recognition [C]//Proc of the 2nd International Conference on Intelligent System Design and Engineering Application. 2012: 456-459.
  - [14] SCHAFFER J D. Multiple objective optimization with vector evaluated genetic algorithms [C]//Proc of the 1st International Conference Genetic Algorithms. New Jersey: Lawrence Erlbaum, 1987: 93-100.
  - [15] JENSEN M T. Reducing the run-time complexity of multiobjective EAs: the NSGA-II and other algorithm [J]. *IEEE Trans on Evolutionary Computation*, 2003, 7(5): 503-515.
  - [16] 郑金华, 史忠植, 谢勇. 基于聚类的快速多目标遗传算法 [J]. *计算研究与发展*, 2004, 41(7): 1081-1087.
  - [17] SERGIOS T, KONSTANTINOS K. Pattern recognition [M]. 4th ed. Beijing: Machine Press, 2010.
  - [18] POLONI C. Hybrid GA for multi-objective aerodynamic shape optimization [C]//Proc of Genetic Algorithms in Engineering and Computer Science. New York: Wiley, 1997: 397-414.
  - [19] VIENNET R. Multi-criteria optimization using a genetic algorithm for determining a Pareto set [J]. *International Journal of Systems Science*, 1996, 27(2): 255-260.
  - [20] ZITZLER E, DEB K, THIELE L. Comparison of multi-objective evolutionary algorithm: empirical results [J]. *Evolutionary Computation*, 2000, 8(2): 173-195.
  - [21] FONSECA C M, FLEMING P J. Multi-objective optimization and multiple constraint handling with evolutionary algorithms—Pareto II: application example [J]. *IEEE Trans on Systems, Man, and Cybernetics: Part A: Systems and Humans*, 1998, 28: 26-37.
  - [22] BOSMAN P A N, THIERENS D. The balance between proximity and diversity in multi-objective evolutionary algorithms [J]. *IEEE Trans on Evolutionary Computation*, 2003, 7(2): 174-188.

(上接第3545页)新用户由于没有打分信息或者有很少的打分信息(即不活跃用户),协同过滤无法为其产生推荐,这就是一直困扰推荐系统领域的冷启动问题<sup>[16]</sup>。然而本文发现与传统协同过滤算法相比,考虑负相关性信息的协同过滤算法由于将负相关用户集成到推荐过程中,能大幅度提高对不活跃用户的推荐准确性,因此负相关性有助于解决推荐系统中的冷启动问题。

## 5 结束语

本文深入研究了负相关性对协同过滤算法的影响,提出了考虑负相关性信息的协同过滤算法,该算法分别选取正相关用户和负相关用户作为最近和最远邻居,然后使用参数调节最近邻居和最远邻居在推荐过程中的作用,最后根据预测出的评分产生推荐列表。MovieLens 实验表明负相关性可以同时提高推荐准确性和推荐列表多样性,说明负相关性有助于解决准确性和多样性两难问题。进一步分析发现,与活跃用户相比,负相关性能大幅度提高不活跃用户的推荐准确度。与传统协同过滤算法相比,本文算法对不活跃用户的 MAE 提高比例可以达到 3.01%,这表明负相关性有助于解决冷启动问题。

### 参考文献:

- [1] GOLDBERG D, NICHOLS D, OKI B M, *et al.* Using collaborative filtering to weave an information tapestry [J]. *Communications of the ACM*, 1992, 35(12): 61-70.
- [2] ADOMAVICIUS G, TUZHILIN A. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions [J]. *IEEE Trans on Knowledge and Data Engineering*, 2005, 17(6): 734-749.
- [3] 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.
- [4] BREESE J S, HECKERMAN D, KADIE C. Empirical analysis of predictive algorithms for collaborative filtering [C]//Proc of the 14th Conference on Uncertainty in Artificial Intelligence. San Francisco: Morgan Kaufmann Publishers Inc, 1998: 43-52.
- [5] 朱丽中, 徐秀娟, 刘宇. 基于项目和信任的协同过滤推荐算法 [J]. *计算机工程*, 2013, 39(1): 58-63.
- [6] WANG Jing, YIN Jian. Enhancing accuracy of user-based collaborative filtering recommendation algorithm in social network [C]//Proc of the 3rd International Conference on System Science, Engineering Design and Manufacturing Informatization. 2012: 142-145.
- [7] 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, 14(9): 1621-1628.
- [8] 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.
- [9] 石珂瑞, 刘建国, 郭强, 等. 有向相似性对协同过滤推荐系统的影响研究 [J]. *复杂系统与复杂性科学*, 2012, 9(3): 46-49.
- [10] 邹永贵, 望靖, 刘兆宏, 等. 基于项目之间相似性的兴趣点推荐方法 [J]. *计算机应用研究*, 2012, 29(1): 116-118.
- [11] 罗辛, 欧阳元新, 熊璋, 等. 通过相似度支持度优化基于 K 近邻的协同过滤算法 [J]. *计算机学报*, 2010, 33(8): 1437-1445.
- [12] 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.
- [13] 张尧, 冯玉强. 协同过滤推荐中基于用户分类的邻居选择方法 [J]. *计算机应用研究*, 2012, 29(11): 4216-4219.
- [14] ZENG Wei, SHANG Ming-sheng, ZHANG Qian-ming, *et al.* Can dissimilar users contribute to accuracy and diversity of personalized recommendation [J]. *International Journal of Modern Physics C*, 2010, 21(10): 1217-1227.
- [15] 刘建国, 周涛, 郭强, 等. 个性化推荐系统评价方法综述 [J]. *复杂系统与复杂性科学*, 2009, 6(3): 1-10.
- [16] 周涛. 个性化推荐技术的十大挑战 [J]. *程序员*, 2012(6): 107-111.