

一种改进的协同过滤的商品推荐方法

薛亮^{1,2}, 徐慧², 冯尊磊³, 贾俊铖⁴

1. 苏州城市学院 计算机工程系, 江苏 苏州 215104;
2. 苏州大学文正学院 计算机工程系, 江苏 苏州 215104;
3. 浙江大学 计算机科学与技术学院, 浙江 杭州 310027;
4. 苏州大学 计算机科学与技术学院, 江苏 苏州 215006)

摘要:传统的电子商务平台无法实现对用户进行个性化优质商品的推荐,大量的商品信息无法被充分应用在帮助客户选择商品上。针对上述问题,提出了改进的协同过滤算法(new Pearson collaborative filtering, NP-CF)为用户进行个性化的电子商品推荐。该算法弱化了活跃用户对商品相似度的贡献程度并且利用标准差的计算降低电子商品本身质量对相似度的影响,将两者计算获得的系数与皮尔逊关系系数相结合,从而计算出更加准确的用户相似度,再利用相似度值计算商品的推荐值并且通过加权评价公式对该值进行加权处理。最后在真实的数据集 MovieLens 和人工数据集 Mobile-Data 上对该算法进行实验测试,且与传统的基于用户信息的协同过滤算法(user collaborative filtering, User-CF)进行比对,该算法(NP-CF)整体上优化了推荐结果并且提高了推荐的准确率。

关键词:电子商品;推荐;协同过滤;皮尔逊相关系数;相似度

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2022)07-0201-07

doi:10.3969/j.issn.1673-629X.2022.07.035

An Improved Co-filtered Goods Recommendation Method

XUE Liang^{1,2}, XU Hui², FENG Zun-Lei³, JIA Jun-cheng⁴

1. Computer Engineering Department, Suzhou City University, Suzhou 215104, China;
2. Computer Engineering Department, Wenzheng College of Soochow University, Suzhou 215104, China;
3. School of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China;
4. School of Computer Science and Technology, Soochow University, Suzhou 215006, China)

Abstract: Traditional e-commerce platforms are unable to realize personalized high-quality goods recommendations for users, causing users to face a large amount of goods information, but the utilization rate of information is gradually decreasing. In response to the above problems, we propose an improved collaborative filtering algorithm (new Pearson collaborative filtering, NP-CF) to provide users with personalized electronic goods recommendations. The algorithm weakens the contribution of active users to goods similarity and uses the calculation of standard deviation to reduce the impact of the quality of electronic goods on similarity. The calculated coefficients of the two with the Pearson relationship number is combined to calculate a more accurate user similarity, then the similarity value is used to calculate the recommended value of the goods and weight the value through a weighted evaluation formula. Finally, the proposed algorithm is tested on the real data set called MovieLens and the artificial data set named Mobile-Data and compared with the traditional user collaborative filtering based on user information (user-CF). The proposed NP-CF optimizes the recommendation results as a whole and improves the accuracy.

Key words: electronic goods; recommendation; collaborative filtering; Pearson correlation coefficient; similarity

0 引言

随着全球科技的飞速发展,并且在当今社会疫情的背景下,网络购物已经成为人们日常购物的重要方式。根据中国报告大厅网络购物行业数据统计,网络

在线的零售交易额正在以每年多于40%的速度增长,在国内电子购物发展非常迅猛的背景下,在线商品交易的用户数量已经高达8.6亿^[1]。但根据商务部的相关反映,网络购物中的一些违法犯罪行为突出,例如:

收稿日期:2021-07-16

修回日期:2021-11-17

基金项目:国家自然科学基金青年科学基金(62003218);教育部高等教育司产学研合作协同育人项目(201902295009)

作者简介:薛亮(1991-),男,硕士,讲师,CCF会员(92165M),研究方向为数据挖掘。

侵犯消费者的合法权益、制售假冒伪劣商品等,已经成为侵权假冒的关键手段,为消费者推荐优质的商品也就变得非常必要。针对上述商品信息过载的问题,为了方便、便捷、有效地为用户提供所需的商品,推荐算法逐渐成为研究热点。通过分析用户行为以及物品内容特征,经过一系列的数据处理,帮助用户获得需要的信息。推荐算法在网络购物产品中得到了较为广泛的应用,但是商家为了提高销售量使得推荐的商品鱼龙混杂,导致消费者难以分辨出自己想要的商品。因此,想要为用户进行个性化推荐,就必须要对推荐算法有较为深入的研究。

目前,大众使用的较为主流的推荐算法是协同过滤算法,但传统的协同过滤算法常常伴有数据稀疏性以及冷启动的问题。因此,国内外许多专家提出了众多改进算法。Choi Y 等人^[2]对用户的情感进行分类研究,发现可以通过该方式影响电子商务推荐的结果,他们认为可以通过计算机相关技术、机器学习等方法,将用户的情感分类,再根据分类之后所得到的结果对用户进行个性化推荐。该系统采用了 Naive Bayes(朴素贝叶斯)、SVM 和决策树等方法,通过大量数据集的实验分析可知决策树具有更好的分类效果。杨茂保等人^[3]通过分析用户的评论从而发现用户潜在的偏好,系统主要采用协同过滤算法分析用户在评论中的情感因素,预测用户之后的产品需求。

根据上述的研究内容,该文的电子产品推荐采用改进的协同过滤算法实现,通过降低活跃用户对相似度的影响并且将电子产品本身属性的质量通过标准差计算获得其离散性并将该离散值融入进相似度计算的公式,对电子商品的推荐值进行加权处理从而优化推荐的预测效果。最后,在真实的数据集 Movie-Lens 上对算法的性能进行测试,并且利用人工数据集 Mobile-Data 在设计实现的电子产品推荐系统上进行推荐性能实验测试。

1 传统的协同过滤算法

协同过滤推荐算法(collaborative filtering, CF)^[4]是迄今为止较为热门的推荐技术。协同过滤推荐算法的目标是根据用户之前的喜好和其他志同道合的用户意见,为该用户或者新用户推荐新项目或者预测他们将对哪些项目感兴趣。总的来说,该算法可以实现对某些特定项目的推荐与预测功能。传统的协同过滤

算法主要分为两类:一是基于用户的相似度计算进行相关推荐,二是基于物品的相似度计算进行相关推荐。基于用户的协同过滤推荐算法(user collaborative filtering, User-CF)是通过计算用户之间的相似度得到与目标用户志趣相投的用户,并根据相似度较高的用户评分,将评分较高的项目信息推荐给目标用户。基于物品的协同过滤推荐算法(item collaborative filtering, Item-CF)是通过计算物品之间的相似度,举例来说,如果用户喜欢物品 a,就会向用户推荐与物品 a 具有较高相似性的其他物品信息^[1]。该文主要采用基于用户的协同过滤推荐算法,该算法的主要流程如图 1 所示。

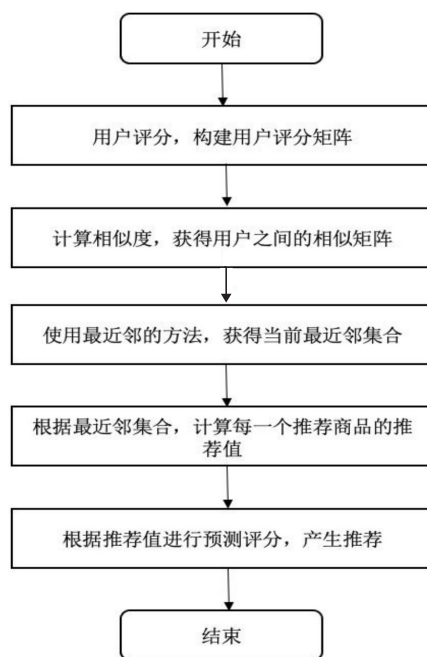


图 1 传统的 User-CF 算法流程

1.1 构建用户评分矩阵

用户评分矩阵数据模型的构建,主要是根据用户对物品的评分表,构建出相应的评分矩阵。针对一个系统具有多个用户与物品信息,记用户数目为 M ,物品数目为 N ,从而根据用户对物品的评分构建出 $M \times N$ 的用户-物品的评分矩阵, $R = M \times N$,故 R_{ij} 表示用户 i 对物品 j 的评分数据。User 表示用户用 $User_i$ 表征,Item 表示物品用 $Item_j$ 表征。为了更加直观地表现数据模型的构建过程,假设用户数 $M = 3$,物品数 $N = 4$,则构建出一个 $R = 3 \times 4$ 阶的用户-评分矩阵,见表 1。其中数据越低表示评分越低,若数值为 0 表示没有进行相关的评分。

表 1 用户评分

User/Item	Item ₁	Item ₂	Item ₃	Item ₄
User ₁	$R_{11} = 4$	$R_{12} = 3$	$R_{13} = 0$	$R_{14} = 0$

续表 1

User/Item	Item ₁	Item ₂	Item ₃	Item ₄
User ₂	$R_{21} = 4$	$R_{22} = 2$	$R_{23} = 4$	$R_{24} = 0$
User ₃	$R_{31} = 1$	$R_{32} = 1$	$R_{33} = 2$	$R_{34} = 1$

根据上述表格分析:假设要为 User₁ 推荐物品,就先要找到与该用户有相同兴趣的用户,通过观察上表发现 User₂ 和 User₁ 均对 Item₁ 具有相同的评分,然而 User₃ 对该物品的评分较低为 1 分,因此选取 User₂ 为最近邻,可见 User₂ 对 Item₃ 的评分为 4 分并且 User₁ 未对该物品做出任何评分,故可以将 Item₃ 推荐给 User₁,预测 User₁ 对 Item₃ 的评分也是 4 分。这里将用户对电子商品的评分作为输入,设计出评分表。该表包含多个用户的评分组合。设计评分矩阵如图 2 所示。

	Item ₁	Item ₂	Item ₃	Item ₄
User ₁	4	3	0	0
User ₂	4	2	4	0
User ₃	1	1	2	1

图 2 物品的评分矩阵

1.2 传统的相似度计算

推荐系统的设计还离不开计算模型中的相似度,在构建完数据模型得到用户对电子商品的评分矩阵后,根据该矩阵计算数据模型内用户之间的相似度,继而得到相似度矩阵。常见的计算用户之间相似度的方法主要有:余弦相似度(cos)^[4]、杰卡德系数(Jaccard)^[5]以及皮尔逊相关系数(Pearson)^[6]等。具体的计算公式如下:

(1) 基于余弦的相似度(cos):

$$\cos(\vec{a}, \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \times \|\vec{b}\|} \quad (1)$$

(2) 杰卡德系数(Jaccard):

$$\text{Jaccard}(u, v) = \frac{|I_u \cap I_v|}{|I_u \cup I_v|} \quad (2)$$

(3) 皮尔逊相关系数(Pearson):

$$\text{Pearson}(u, v) = \frac{\sum_{i \in C_{uv}} (R_{ui} - \bar{R}_u) \times (R_{vi} - \bar{R}_v)}{\sqrt{\sum_{i \in C_u} (R_{ui} - \bar{R}_u)^2} \times \sqrt{\sum_{i \in C_v} (R_{vi} - \bar{R}_v)^2}} \quad (3)$$

值得指出的是,基于上述传统的相似度计算公式,在用户评分矩阵极其稀疏的情况下很难通过直接计算得到真实的最近邻集合。例如,用户 A 的评分数据为(1,2,1),用户 B 与用户 C 的评分分别为(5,4,5)、

(4,5,4)。通过观察分析可以知道用户 B 与用户 C 更为相似,但是通过皮尔逊相关系数计算得到用户 A 与用户 C 最为相似,因此可见无法获得真正的最近邻。

1.3 评分预测推荐结果

传统的基于用户评分信息的协同过滤算法^[7]在获得用户的相似度矩阵后,通过 K 近邻的方法获取与目标用户最为相似的 K 个用户集合,再通过加权评价公式进行项目推荐的预测评分。

常用的预测评分的公式如下:

$$P_{u,i} = \frac{\sum_{\text{all simill item. N}} S(u, v) * R_{v,i}}{\sum_{\text{all simill item. N}} |S(u, v)|} \quad (4)$$

该公式是推荐算法评分预测常用的公式, u, v 表示两个不同的用户, i 表示物品, $R_{v,i}$ 表示用户 v 对物品 i 的评分数据, $S(u, v)$ 是用户 u 与用户 v 之间的相似度值。通过计算用户对与 i 相似的商品给出的评分总和,以计算对用户 u 商品的评分预测,其本质上是对基于皮尔逊相关系数计算后所得的相似度以及最近邻评分的加权平均数,然后将所计算得到的加权平均数进行排序,从而进行评分预测。

2 改进的基于用户的协同过滤算法

综上所述,传统的基于用户评分信息的协同过滤算法(user collaborative filtering, User-CF),利用一般的相似度计算公式时会受到活跃用户大量的评分信息以及商品本身质量属性的影响。首先,为了消除上述相关因素对相似度计算的影响,该文将活跃用户对商品相似度的贡献程度降低并且通过计算标准差降低电子商品本身质量对相似度的影响,将两者计算获得的系数与皮尔逊相关系数(Pearson)结合。其次,采取对推荐项目的推荐值进行加权处理,从而得到优化后的结果。

2.1 改进的相似度计算

在推荐系统中,最为主要的工作就是计算用户之间的相似度。因此,考虑到活跃用户以及商品自身质量对相似度计算结果的影响,该文对相似度计算的改进工作主要如下:

(1) 活跃用户的惩罚。

活跃用户对商品的评分行为过多,由于该类浏览过多的商品对商品的甄别性较差,因此,非活跃用户的评分比活跃用户的评分更具有价值。该文参考 John

S. Breese^[8] 论文中所提出的 IUF (inverse user frequency), 用于降低活跃用户对商品评分的贡献程度, 具体的实现公式如下:

$$IUF = \frac{1}{\log(1 + |N(u)|)} \quad (5)$$

该公式对活跃用户进行了软性的惩罚, 其中 $N(u)$ 代表当前训练的用户数。

(2) 降低电子商品本身质量的影响。

该文利用标准差的计算方式表示电子商品本身的质量因素对用户评分数值的影响。评分的标准差的值越高, 表明电子商品本身的质量因素对用户评分的影响比重就越大, 反之, 则表明影响的程度越小。其离散性降低商品本身质量^[9]对用户评分的计算公式如下:

$$D_i = \frac{\sigma_i}{\gamma_i} = \frac{\sqrt{\frac{1}{N-1} \cdot \sum_{m=1}^N I_m - \gamma_i^2}}{\gamma_i} \quad (6)$$

其中, N 表示各用户对电子商品的评分个数, I_m 表示用户对某电子商品的第 m 个评分数值, γ_i 表示用户对电子商品 i 评分的均值。

该文采用皮尔逊相关系数计算用户之间的相似度, 为提高该公式的精确度, 将式(5)与式(6)融入公式(3), 从而降低活跃用户对相似度值的贡献程度以及电子商品本身质量对其的影响, 可得修正后的公式如下:

$$\text{simNP}(u, v) = \frac{\sum_{i \in C_{uv}} (R_{ui} - \bar{R}_u) \times (R_{vi} - \bar{R}_v) \times IUF \times D_i}{\sqrt{\sum_{i \in C_{uv}} (R_{ui} - \bar{R}_u)^2} \times \sqrt{\sum_{i \in C_{uv}} (R_{vi} - \bar{R}_v)^2}} \quad (7)$$

其中, C_{uv} 表示用户 u 与用户 v 共同感兴趣评分的物品集合, R_{ui} 表示用户 u 对物品 i 的评分信息, $\bar{R}_u$ 表示用户 u 对所有他感兴趣商品评分的均值。

2.2 选择最近邻

最近邻的搜索 (nearest neighbor search, NNS)^[9-14] 主要是在相似度计算的基础上, 通过获得的相似度矩阵, 找到最近邻集合。在 K 近邻实现的过程中, 整个数据分为训练数据和测试的样本数据, 在分类问题中, 使用以下步骤实现 K 近邻的算法:

(1) 确定 K 的数值, 该 K 值是特征空间中训练模型的数量;

(2) 根据之前相似度计算获得各个用户之间的数据点的距离;

(3) 在训练模型数据中搜索与最接近 K 值的各个数据点, 其中这些数据点必须在训练模型的观察点内;

(4) 计算 K 值与各个数据点之间的距离, 从而获得最近邻集合。

举例说明根据图 3 的特征空间图, 可以获得最近邻集合:

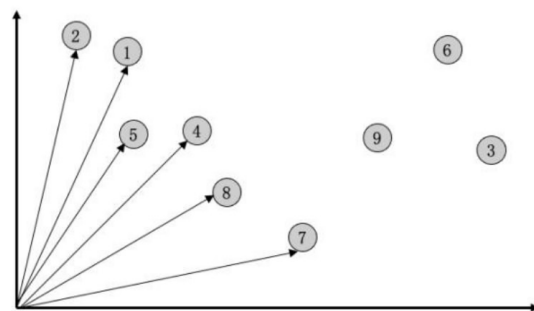


图 3 训练模型特征空间图

从该特征空间图中可以找出 5 的五个最近邻, 即为 $NNS(5) = [4, 1, 8, 2, 7]$ 。获得最近邻集合为后续进行评分预测奠定了较好的基础, 在该集合内排序越靠近前面说明与目标项目的相关程度越大。

2.3 算法整体流程

通过改进的皮尔逊相关系数的公式计算获得用户的相似度矩阵, 利用 K 近邻的方式选择 K 个最为相似的用户, 对此 K 个用户的相似度值进行推荐值的计算公式如下:

$$VSim(u, v) = \text{SimNP}(u, v) * K * R_{vi} \quad (8)$$

然后通过如下公式进行评分预测:

$$NP_{u,i} = \frac{\sum_{\text{all small item. N}} VSim(u, v) * (R_{vi} - \bar{R}_v)}{\sum_{\text{all small item. N}} |VSim(u, v)|} + \bar{R}_u \quad (9)$$

其中, $VSim(u, v)$ 是用户 u 与用户 v 加权后的两者相似度的值。对比公式(4), 此公式引入用户 u 与用户 v 对已有电子商品评分的均值, 优化了原先的加权评价公式。最后, 使用该公式进行对用户的预测推荐。本算法的具体流程如下:

算法: 改进的协同过滤算法。

Input: User-Item 即用户与项目的评分矩阵。

Output: 预测推荐的项目列表。

Steps:

(1) 初始化: 根据读入的用户、评分和项目的数据文件, 通过预处理构建 User-Item 的评分矩阵;

(2) 实现求均值的函数 average_rating, 计算用户评分的均值 $\bar{R}_u$ 和 $\bar{R}_v$;

(3) For $u = 1$ in User

(4) 通过公式(5)计算对活跃用户的惩罚系数 IUF;

(5) 根据公式(6)计算商品本身质量的离散系数 D_i ;

(6) 将步骤 4、5 计算获得的 IUF 与 D_i 运用于公式(7), 从而利用改进后的皮尔逊相关系数公式计算用户之间的相似性 $\text{SimNP}(u, v)$;

(7) End

(8) For $n = 1$ to N do

(9)遍历当前用户的相似度利用选择 K 近邻的方法,选择 K 个相似度值较高的用户形成最近邻集合;

(10)End

(11)For $n = 1$ to K do

(12)遍历(该过程在 for 循环内实现)最近邻集合,通过公式(8)计算当前项目的推荐值;

(13)利用公式(9)进行加权评分预测;

(14)最后根据预测评分的结果进行遍历,从而得到最终推荐的项目列表。

(15)End

3 实验

3.1 数据集

MovieLens100K 是电影评分数据集^[15],由 943 个用户对 1 682 部电影的 100 000 个电影的评分组成,评分范围为 1 分至 5 分。数据文件中主要分为用户编号 (userId)、电影编号 (itemId)、电影评分 (rating) 以及时间戳 (timestamp),并且该文件中的数据是随机排序的。在实验过程中,将 MovieLens 数据集以 7 : 3 的比例随机分为训练集与测试集。

人工数据集 Mobile-Data,是利用 Intel Core i5 的 CPU,8 GB 运行内存进行爬虫收集数据,该数据集主要收集汇总了手机信息。将该人工数据集 Mobile-Data 用于测试电子商品系统的推荐功能。

3.2 算法评价标准

算法评测指标是计算准确率 (Precision)、召回率 (Recall) 以及覆盖率 (Cov),将该算法 (记为 new Pearson collaborative filtering, NP-CF) 与传统的 User-CF 算法进行比较。算法评测的方法如下:

(1)准确率。

准确率是用于衡量推荐算法对整个数据集中训练集最终做出多少正确的预测。计算公式如下:

$$\text{Precision} = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{|R(u)|} \quad (10)$$

(2)召回率。

召回率是衡量从测试的数据集中为用户推荐的信息获得反馈的数据占用户实际点击感兴趣的数据的比例。计算公式如下:

$$\text{Recall} = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{T(u)} \quad (11)$$

公式(10)与(11)中, $R(u)$ 表示推荐算法对用户 u 所推荐的项目, $T(u)$ 表示用户 u 在所训练的数据集上面喜欢的项目集合。

(3)覆盖率。

覆盖率^[16]表明预测推荐的项目对目标用户是否有效,该值越大则说明此算法预测推荐的项目其分布就越均匀。计算公式如下:

$$\text{Cov} = \frac{\sum_{u \in U} N_u}{n} \quad (12)$$

其中, N_u 表示预测评分为用户 u 所推荐的所有项目的数目, n 表示整个数据集中所有项目的数目。

3.3 实验结果与分析

文中算法 (NP-CF) 与传统的基于用户的协同过滤算法 (User-CF) 进行比较。其一,采用上述的 MovieLens 数据集比较两者的准确率、召回率。其二,设计实现电子商品推荐系统,将人工数据集 Mobile-Data 存入数据库作为数据来源,利用上述两种推荐算法实验测试推荐结果。

3.3.1 MovieLens 数据集实验

采用文中算法 (NP-CF) 与传统的基于用户信息的协同过滤算法 (User-CF) 在 MovieLens 数据集并且通过不同的最近邻个数计算相应的准确率、召回率。这里采用随机的方式将数据集的 70% 作为训练集,遍历训练集的数据形成用于计算用户相似度的倒排表,其次实现相似度的计算并且通过加权评价公式形成预测推荐的列表,最后根据预测推荐的项目计算准确率、召回率,比较二者的数据。

实验结果见表 2 和表 3。

表 2 准确率

最近邻个数 N 算法	10	20	30	40	50	60	65
NP-CF	0.246 2	0.267 8	0.281 1	0.286 7	0.291 8	0.291 3	0.291 5
User-CF	0.085 0	0.091 7	0.100 0	0.106 7	0.111 6	0.116 1	0.119 3

表 3 召回率

最近邻个数 N 算法	10	20	30	40	50	60	65
NP-CF	0.077 5	0.084 3	0.088 5	0.090 3	0.091 9	0.091 7	0.091 8
User-CF	0.026 8	0.028 9	0.031 5	0.033 6	0.035 1	0.036 6	0.037 6

从表 2 可以看出,文中算法(NP-CF)相比传统的基于用户信息的协同过滤算法(User-CF)准确率有了明显的提高。准确率的比对结果如图 4 所示。从图中可以看出最近邻个数 N 对两种算法的准确率有一定的影响。当 N 达到 30 之后,NP-CF 的准确率逐渐趋于平稳,在 $N < 50$ 时,该算法的准确率随着最近邻个数 N 的增加而提高。当 $N > 50$ 时,准确率开始下降,但是降低的程度较低。综上所述,在最近邻个数变化的情况下,文中算法(NP-CF)的准确率明显优于传统的基于用户信息的协同过滤算法(User-CF)。

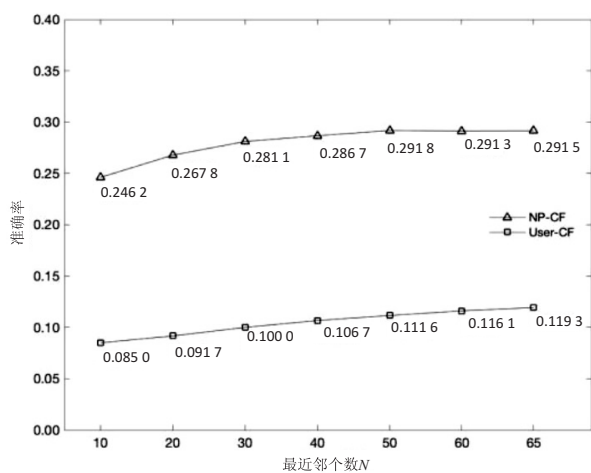


图 4 两种算法准确率比较结果

根据表 3 对两种算法召回率的统计,随着最近邻数值的增加,两种算法的召回率都在提高。

比较两种算法召回率的数值,可见文中算法(NP-CF)召回率提高的更显著。

从图 5 可以看出,通过文中算法(NP-CF)计算获得的召回率,随着最近邻个数的增加,当最近邻个数 $N > 40$ 时,召回率的数值逐渐趋于稳定在 0.09 左右,当最近邻个数 $N > 50$ 时,召回率的数值开始降低。由此得出,文中算法(NP-CF)的召回率具备优势。

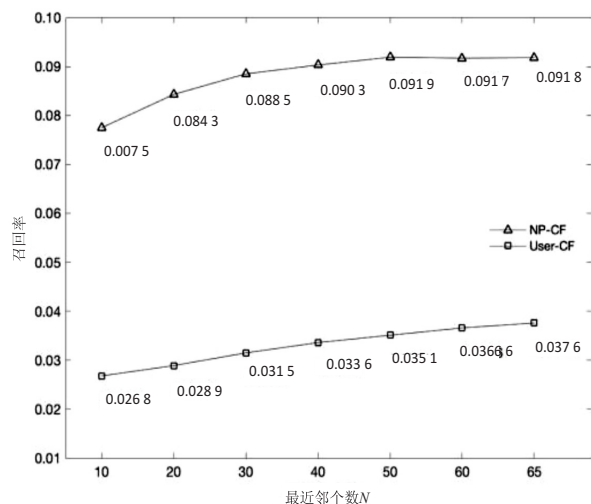


图 5 两种算法召回率比较结果图

根据上述两种评价方法的比较,最终设计实现的电子推荐系统,最近邻的个数选择 30,根据上述的测试分析可见当最近邻个数为 $N = 20$ 时,文中算法的准确率达到 28.11%,而传统的基于用户的协同过滤算法的准确率只有 10.00%,可见预测推荐的结果更为准确。 $N = 30$ 的准确率数值与最近邻个数 $N > 30$ 的准确率值都比较接近。

3.3.2 人工数据集

采用 eclipse 开发工具设计实现电子商品推荐系统,将人工数据集 Mobile-Data 存入数据库作为数据源,利用文中算法(NP-CF)和传统的基于用户信息的协同过滤算法(User-CF),分别设计实现系统的推荐功能并且利用人工数据集 Mobile-data 进行推荐功能的实例测试。在系统通过多个用户对不同的手机进行浏览与评分,测试系统最终对目标用户进行推荐,观察其结果发现能够形成较好的推荐。测试数据包含 4 个测试用户 A、B、C、D,用户喜欢商品包括 4 个品牌,10 个型号手机产品。

采用文中算法(NP-CF)以及传统的基于用户信息的协同过滤算法(User-CF)经过该数据集实际测试结果见表 4。其中文中算法(NP-CF)最终形成了 5 个推荐结果,选用传统的基于用户信息的协同过滤算法(User-CF)形成了 3 个推荐结果。

表 4 系统测试结果

算法	推荐结果
NP-CF	1. IPHONE11PRO
	2. HUAWEIP40
	3. ONEPLUS9PRO
	4. IPHONE12
	5. HUAWEIMATE30EPRO
User-CF	1. HUAWEIP40
	2. ONEPLUS9PRO
	3. IPHONE12

根据皮尔逊相关系数相关技术的公式,用户 A 与用户 B 的相似度值应该与用户 A 与用户 D 的相似度值相等,但是传统的基于用户的系统过滤算法未将用户 D 感兴趣的电子产品 IPHONE11PRO 以及 HUAWEIMATE30EPRO 推荐给目标用户 A,造成该推荐结果的原因是只利用皮尔逊相关系数计算获得最近邻,常常无法获得准确的最近邻集合,而文中算法(NP-CF)对皮尔逊相关系数进行了优化改进后获得的推荐结果相较而言更加具有优势。值得指出的是,该方法推荐结果与最终推荐次序无关,通过计算目标用户的相似度,最终按照商品的 ID 顺序将推荐的商品添加进入推荐列表。

4 结束语

传统的基于用户的协同过滤算法由于受活跃用户以及商品自身质量因素的影响,得到的预测推荐效果较差。对此,该文将活跃用户的惩罚系数与商品自身质量对用户评分的惩罚因子融入皮尔逊相关系数的计算公式。经过 MovieLens 数据集测试的实验表明,该算法(NP-CF)能够有效地提高准确率(Precision)、召回率(Recall),并且在使用人工数据集 Mobile-Data 作为电子商品推荐系统测试的数据源时,该算法在仿真测试中能够得到较好的推荐效果。

参考文献:

- [1] 王敏. 基于协同过滤的电子商务个性化推荐系统设计和实现[D]. 广州:广州大学,2019.
- [2] CHOI Y, LEE H. Data properties and the performance of sentiment classification for electronic commerce applications[J]. Information Systems Frontiers, 2017(3):1-20.
- [3] 杨茂保. 电子商务中基于客户偏好和情感评估的个性化推荐算法[J]. 智能计算机与应用, 2018, 8(3):118-121.
- [4] HERLOCKER J L, KONSTAN J A, TERVEEN L G, et al. Evaluating collaborative filtering recommender systems[J]. ACM Transactions on Information Systems, 2004, 22(1):5-53.
- [5] PATRA B K, LAUNONEN R, OLLIKAINEN V, et al. A new similarity measure using Bhattacharyya coefficient for collaborative filtering in sparse data[J]. Knowledge-Based Systems, 2015, 82:163-177.
- [6] SARWAR B. Item-based collaborative filtering recommendation algorithms[C]//Proc. the 10th international world wide web conference (WWW10). Hong Kong: [s. n.], 2001.
- [7] 潘丽芳, 张大龙, 李慧. 基于用户的协同过滤(UserCF)新闻推荐算法研究[J]. 山西师范大学学报:自然科学版, 2018, 32(4):26-30.
- [8] BREESE J S. Empirical analysis of predictive algorithms for collaborative filtering[C]//Proc. 14th conference on uncertainty in artificial intelligence. Madison, WI: Morgan Kaufmann Publisher, 1998.
- [9] 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.
- [10] 张洪波, 王佳蕾, 张丽娟, 等. 基于信任网络的协同过滤推荐方法[J]. 计算机科学, 2018, 45(8):146-150.
- [11] 薛美琦. 基于协同过滤技术的物品选购推荐系统设计[D]. 成都:电子科技大学, 2019.
- [12] 戎静月. 改进的协同过滤推荐算法研究[D]. 保定:河北大学, 2020.
- [13] 王建芳, 韩鹏飞, 苗艳玲, 等. 一种基于用户兴趣联合相似度的协同过滤算法[J]. 河南理工大学学报:自然科学版, 2019, 38(5):118-123.
- [14] LI X, ZHANG J, UNIVERSITY L N. Study on the privacy protection of mobile recommendation system[J]. Journal of Library and Information Sciences in Agriculture, 201830(11):127-130.
- [15] 王娟, 熊巍. 基于矩阵分解的最近邻推荐系统及其应用[J]. 统计与决策, 2019, 35(6):17-20.
- [16] 王岩, 张杰, 许合利. 结合用户兴趣和改进的协同过滤推荐算法[J]. 小型微型计算机系统, 2020, 41(8):1665-1669.