

文章编号:1007-5429(2015)06-0034-08

协同过滤推荐算法的改进

——帕累托最优视角

焦媛媛,熊剑芳,沈志锋

(南开大学 商学院,天津 300071)

摘要: 在已有的基于项目协同过滤推荐技术基础上,应用帕累托最优理论对该算法进行改进。首先基于调整余弦相似度构件项目相似矩阵,对待评价项目的 K 近邻选择过程进行优化,以事先过滤掉一些代表性不高的邻近项目;其次在推荐评价过程中,为了适应改进算法,对准确率和召回率计算也进行了相应的优化。应用 movielens 1M 数据检验算法优化效果。发现在项目推荐的准确度、覆盖率、召回率和算法可扩展性等方面都有显著提高。

关键词: 协同过滤; 推荐系统; 帕累托最优; 基于项目推荐

中图分类号: F270

文献标识码: A

Improvement of the Collaborative Filtering Recommendation Algorithm: Perspective of Pareto Optimality

JIAO Yuan-yuan, XIONG Jian-fang, SHEN Zhi-feng

(School of business, Nankai University, Tianjin 300071, China)

Abstract: Based on the existing item collaborative filtering recommendation technology, Pareto optimality theory is adopted to improve the algorithm. The similarity matrix of items based on adjusted cosine similarity is built. The K-neighbor selection process of the being evaluated items is optimized, which is used to filter out some less representative adjacent items in advance. In the process of recommendation evaluation, in order to adapt to the improved algorithm, the accuracy and recall rate calculation was also optimized. 1 m movielens data was used to inspect the algorithm's results. We found that there is some significantly improvement in all the accuracy of item recommendation, coverage, recall rate and extensibility of algorithm etc.

Key words: collaborative filtering; recommendation system; pareto optimality; item based recommendation

1 引言

自 20 世纪 90 年代中期以来,源于互联网信息技术的快速发展和应用,推荐系统研究逐渐受到重视,并迅速发展成为一个独立热门研究领域^[1],推荐

系统 RS(Recommender Systems)是为用户解决网络信息过载的一类技术。其主要任务是针对众多类型的项目或信息,为用户提供一系列个性化的建议(推荐)^[1]。推荐系统的工作方式可能会有巨大的不同,这主要取决于被推荐的项目类型,和可获得的关

收稿日期:2015-04-28; 修回日期:2015-07-28

基金项目:中央高校基本科研业务费专项资金资助项目(NKZXA1101)

作者简介:焦媛媛(1970-),女,天津人,教授,博士,主要研究方向为项目管理,电子商务、企业间信息系统, E-mail:yyjiao2007@126.com。

于用户偏好的信息^[2]。因此,RS必须要能够应用不同的机制,从而为每个用户推荐对他们最有用的项目。

按照推荐技术和使用的信息内容不同,推荐技术可分为四类^[1,12]:协同过滤推荐^[2]、基于内容的推荐^[7,8]、基于知识的推荐^[9,10]和混合过滤推荐^[13]。

协同过滤推荐最早应用于各类社区网站,用于电影、音乐等的推荐,随着网络技术和信息技术的发展,电子商务网站也逐渐开发推荐系统,为用户主动推荐他们潜在需要,或感兴趣的项目,协同过滤推荐技术被首先应用于电商推荐,随后也逐渐包含了其他三类推荐技术^[6],基于内容和基于知识的推荐在搜索引擎和门户网站得到广泛应用^[7],同时基于知识的推荐也广泛应用于科研院所图书和情报系统,混合过滤推荐主要应用于一些复杂检索和信息需求。另外,一些社交网络应用主要基于协同过滤和混合推荐技术为用户扩展朋友圈,挖掘用户娱乐喜好和需求等。同时协同过滤方法也被应用于搜索引擎、门户网站和图书情报系统等几乎所有需要信息推荐和检索的领域^[3]。

协同过滤技术是发展最早,并依然是应用最广泛的推荐技术^[2,11],因为互联网经济就是用户为王经济,因此不只是电商和社交服务企业,其他几乎任何互联网应用服务企业都要抢用户,抢流量。用户数量越庞大,该互联网应用价值就越大,因此,首先,良好的协同过滤推荐设计对互联网企业发展是至关重要的。

协同过滤推荐是指利用已有用户群过去的行为或评论信息预测当前用户最可能喜欢哪些东西,或对哪些东西感兴趣^[2]。按照不同的实现方式,协同过滤可分为基于内存的算法和基于模型的算法。

基于内存的协同过滤算法一般只进行在线推荐,即所有的计算过程都在内存中进行,它的基本思想是通过一定的相似度评价标准找出用户或者项目的邻集合,并利用邻居集合的标注信息进行预测。由于所有的计算都是在线完成的,因而可扩展性是该算法的主要挑战。传统的基于用户的最近邻推荐技术是基于内存的;基于模型的方法一般分为两个步骤,即离线模型训练和在线推荐预测,寻找有效的模型训练方法对基于模型的协同过滤算法至关重要,在模型训练好以后,可实现实时计算推荐。基于项目的过滤和某些降维技术都属于基于模型的方法。

由于基于模型的协同过滤推荐方法的相对优越

性,本研究主要基于调整余弦相似度选择邻近项目,并基于帕累托最优理论思想,优化基于项目的协同过滤推荐模型算法,并在推荐质量评价方面,也做了相应的优化调整。

本文后面的结构安排如下:第二部分,论述本研究的理论基础和基于项目的协同过滤推荐算法优化思路。第三部分,详细论述优化推荐算法的计算实现过程。第四部分,实验设计,评价改进算法的推荐质量,并与传统的推荐方法对比分析。第五部分,研究总结并提出未来研究方向。

2 新方法理论基础与算法优化方案

2.1 基于项目的近邻推荐

协同过滤推荐面临的一个主要问题就是评分数据库稀疏性^[21],因为对于大多数用户来讲,他们所评价过的项目占数据库中所有项目的比例通常是非常低的。当计算两两用户之间的相似性时,必须考虑两个用户共同评过的项目。而实际情况是,可能每个用户都评价了上百个项目,却只有两三个共同评价的项目。特别是在大型电子商务网站,这种弊端尤其明显。即基于用户的近邻推荐可扩展性能差^[12]。而基于项目的最近邻推荐主要应用项目间相似度,这种方式很容易实现离线预计算,从而可以大大减少在线运算内存和时间开销,并且相对于用户相似度而言,项目相似度更稳定,这种预处理计算不会过于影响预测准确度^[12]。

基于项目的近邻推荐主要目标是从活跃用户已做出评价的项目中,寻找出与其未评分项目相似的项目,而这种相似性衡量主要基于不同项目之间的被评分相似性。这一协同过滤过程可以总结为下面三个步骤^[13-15]:

(1) 首先构建一个项目相似度矩阵,并通过相应算法得出与候选项目 p 最相似的 k 个项目集合(k —邻近)。这一步骤对推荐质量具有最重要影响,本文正是努力提供一种新方法,为待评价项目寻找更合适的邻近集。

(2) 通过相应算法,计算 u 对这些邻近项目的评分加权总和,作为用户 u 对项目 p 的预测评分。

(3) 寻找最合适的 N 个项目推荐给活跃用户(由于它们的高评分和新颖性)。

然而,如果完全计算当前用户购物篮所有项目(用户所购买过的所有项目)的相似性,计算量也是非常庞大的。尽管相对于基于用户的协同过滤算法,基于项目的协同过滤算法计算复杂度更低,但考

考虑到内存要求, N 个项目的相似度矩阵理论上会有 N^2 项,所以我们有必要进一步降低复杂度。Sarwar等通过对每个项目只记录有限近邻的方法,以降低计算复杂度^[12],然而这种方法会增加无法预测某个特定项目的风险。同样的问题还存在于Linden等仅考虑那些与其他项目同时评分最少的项目^[22]。另外,Salter等提出的基于条件概率的推荐还同时考虑了不同项目的被购买频率对相似性计算的影响,从而进一步提高了推荐准确性^[3]。然而却反而增加了计算复杂度。

也有研究者试图同时降低计算复杂度和提升推荐准确率,如J. Bobadilla等设计的方法使那些最重要的用户(对很多项目掌握更多知识的人)被选为邻近的可能性增加了^[5]。同样是该作者通过构造一个指数结合均方误差,以使相似用户之间具有更多共同关注项目^[23],然而这两个研究中的推荐质量测量指标验证却并没有完全得到改进。A. Hernando等使用一个权重乘以相似性指标^[24],这个权重是活跃用户所评价项目和其他用户所评价项目数量的比例,这一指标思想被本研究吸收应用,我们在推荐质量评价中构建了项目奇异性指标——候选项目与活跃用户评论的所有项目数之比。

2.2 帕累托最优理论及其应用

帕累托最优理论来源于经济学,是指资源分配最具效率的一种理想状态。这一理论思想也被广泛应用于多目标优化问题中^[16,17]。在这类问题中,(在最简化假设条件下)我们把所有的可行解集称为 x ,在其他条件不变差的条件下,如果没有其它可行解使目标函数值更优,那么我们说一个解决方法 x 达到帕累托最优、帕累托效率、帕累托支配。正式来讲,我们面临一个多目标最优化问题,必须从所有可行解 $x=(x_1, \dots, x_2)^T \in X$ 中找到一个最优解,以使多目标函数 $f(x)=(f_1(x), \dots, f_m(x))^T$ 达到最优解。

如果对于所有的 $x \in X$ 都有 $f_i(x) \leq f_i(x')$, $i=1, 2, \dots, m$,并且至少存在一个 i ,使 $f_i(x) < f_i(x')$,那么我们说一个解 $x' \in X$ 是支配解(帕累托最优解),支配的概念在多目标智能优化算法问题中得到广泛应用^[16,17]。例如P. Hansen等基于帕累托拖效率优化思想,对多目标禁忌搜索算法进行优化,大大减少了搜索路径数,使算法快速收敛^[16]。K. Deb等专门研究了非支配快速提升分类遗传算法的优化问题^[17],也是基于帕累托优化思

想对算法搜索路径和计算效率进行优化。还有J. Horn等提出的非线性多目标遗传算法^[18],N. Srinivas等研究的动态优化遗传算法^[27],E. L. Ulungu等研究的马尔可夫链优化问题^[19]等等,都应用了占优/支配的思想使它们的解趋向于帕累托最优。这些问题都属于多目标智能优化的NP难问题^[15],而基于项目的近邻推荐也属于NP难问题,因为计算候选项目与邻近项目之间的相似度,必须考虑众多的评分属性,众多的邻近项目。这类问题很难通过明确的计算路径快速解决,只能通过一定的算法思想趋近最优解。

2.3 算法优化思路方案

在本研究中,帕累托最优思想被应用于决定哪些项目最可能使用户感兴趣^[20],在进行项目相似度计算之前,对项目之间的相似度进行预评估,以减少后面的相似矩阵计算量,但同时又不会影响预测精确度。

应用 k 最邻近算法,很容易为待评项目找到很多并不太适合的邻近(一些和待评价项目相关性非常差的项目)。我们的假设是,如果应用帕累托最优思想,通过在 k 邻近选择过程中剔除代表性小的邻近项目,并保留最具代表性的邻近项目。这样既能减少项目间相似矩阵计算量,还能显著提升预测精确度。从而大大提高协同过滤计算效率和推荐精确度、可扩展性等。

图1为本研究算法流程的图示说明。第1~3步都是离线模型训练过程。首先,系统会自动构建数据库内所有项目的相似度矩阵;第2,3两步是一个动态循环过程, $\#R$ 为活跃用户已评价过的项目数, $(m-\#R)$ 即为活跃用户未评价过的项目数;第2步,推荐系统模型将分别离线计算每个 i 与购物篮中所有项目的相似度,基于帕累托最优思想选出 i 的 K_i 个最优邻近项目;第3步,基于活跃用户对 K_i 个项目的评分,预测项目 i 的评分;第4步为在线评分排序和推荐算法过程,系统将根据活跃用户的在线浏览行为,推荐相应的项目类别。

显然,本文提出的方法并没有对传统方法做出修改,它只是识别一些最合适的初始邻近项目,确保传统方法能从购物篮中计算出最具代表性的邻近。把代表性低的一些候选项目在预处理中剔除。本文提出方法的一个优势在于,它能很好地嵌入到任何其他传统过滤方法中,如融合进相似性测量改进,聚类改进等。

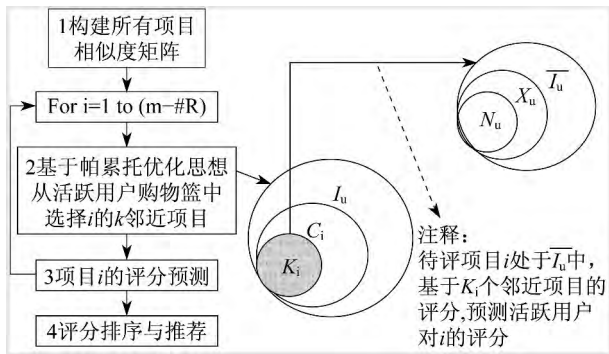


图1 改进的基于项目协同过滤推荐算法流程图示
(所用数学符号在第3节都有定义)

3 优化算法设计项目

3.1 基本变量定义

我们定义应用协同过滤推荐系统的数据库内包含 l 个用户, m 个项目, 对每个项目的评分区间为 $\{\min, \dots, \max\}$, 评分缺失值用“ \cdot ”表示, 并包含下列集合定义。

表1 相关集合定义

集合名	集合定义
用户集	$U = \{u \in \mathbb{N} / 1 \leq u \leq l\}$
项目集	$I = \{i \in \mathbb{N} / 1 \leq i \leq m\}$
可能的评分集	$V = \{v \in \mathbb{N} / \min \leq v \leq \max\} \cup \{\cdot\}$
项目 i 获得的评分集	$R_i = \{(u, v) / u \in U, v \in V\}$
项目 i 获得用户 u 的评分 v 为 $r_{u,i} = v$	
用户 u 的购物篮为 I_u, \bar{I}_u 为其补集	
用户 u 评价过的项目的平均评分为 \bar{r}_u	
项目 i 所收到的所有用户评分均值为 \bar{r}_i	
有效数据集	C
有效数据数	$\#C = \#\{x \in C / x \neq \cdot\}$
u 评分个数	$\#R_u = \#\{i \in I / r_{u,i} \neq \cdot\}$

3.2 选择候选邻近(占优邻近)

我们决定所有与待评估项目相似的项目, 或者更正式地讲, 为占优邻近, 设 $U_i = \{u \in U / r_{u,i} \neq \cdot\}$ 为评价过项目 i 的用户集。 $d(r_{u,x}, r_{u,y})$ 为同一用户对不同项目 x 和 y 的评分差别。

$$d(r_{u,x}, r_{u,y}) = \begin{cases} |r_{u,i} - r_{u,x}|, & r_{u,x} \neq \cdot \\ \infty, & r_{u,x} = \cdot \end{cases}$$

我们说, 对于项目 i , 项目 x 更优于项目 y , 当满足下列条件时:

$$x >_i y \Leftrightarrow \forall u \in U_i: d(r_{u,i}, r_{u,x}) \leq d(r_{u,i}, r_{u,y}) \\ \wedge \exists j \in U_i / d(r_{j,i}, r_{j,x}) < d(r_{j,i}, r_{j,y}) \quad (1)$$

因此, 相对于最优项目, 次优项目并没有与待评

项目之间表现出更高的相似性, 只是表现出低度的相似性。也就是说, 次优项目并不能比最优项目为待评项目提供更多相似性信息, 因此可以剔除掉。

我们定义 C_i 为项目 i 的候选邻近(占优邻近项目)。它必须满足以下条件: 其中 D_i 是非占优项目集

$$C_i \subseteq I, i \notin C_i, C_i = I - (D_i \cup \{i\}), \\ \forall y \in D_i, \exists x \in C_i / x >_i y \quad (2)$$

3.3 寻找帕累托最优 k 邻近

我们主要通过两步来寻找待评项目的 k 邻近。

(1) 计算待评项目和占优邻近项目集 C_i 中所有项目之间的相似性。

(2) 找出与待评项目相似性最高的 k 个项目(对每个不同 i 循环这两步)。

该方法的优点在于, 计算过程只在 C_i 和待评项目之间进行, 而并不是象传统方法那样, 计算待评项目与用户购物篮中所有项目 I_u 之间的相似性。

已有学者研究证实^[17], 对于基于项目的推荐技术, 余弦相似度方法比 Pearson 相关度量表现更好。而由于基本的余弦方法不会考虑用户评分平均值之间的差异。改进的余弦方法能很好地解决这个问题, 做法是在评分值中减去平均值。相应地, 改进余弦方法的取值在 -1 到 $+1$ 之间, 就像 Pearson 方法一样, 本文正是采用调整余弦相似度计算项目之间的相似度。

3.3.1 设 $A_{x,y} = \{i \in I / r_{x,i} \neq \cdot \wedge r_{y,i} \neq \cdot\}$ 为同时被活跃用户 u 评价过的项目集, 设 x 为待评项目, 则:

$$\text{sim}(x, y) = \frac{\sum_{u \in U} (r_{u,x} - \bar{r}_x) \cdot (r_{u,y} - \bar{r}_y)}{\sqrt{\sum_{u \in U} (r_{u,x} - \bar{r}_x)^2 \cdot \sum_{u \in U} (r_{u,y} - \bar{r}_y)^2}} \\ \Leftrightarrow y \in C_x \quad (3)$$

$$\text{sim}(x, y) = \cdot \Leftrightarrow y \notin C_x \quad (4)$$

我们可以看出, 邻近只从占优邻近集中选出。

3.3.2 设 K_i 为待评估项目 i 的 k 个邻近, k 邻近必须满足下面的条件:

$$K_i \subseteq C_i \wedge \#K_i \leq k \wedge i \notin K_i \\ \forall x \in K_i, \forall y \in (C_i - K_i): \text{sim}(x, i) \geq \text{sim}(y, i) \quad (5)$$

选择最优邻近数量 k 要考虑多方面因素。包括数据库的大小、稀疏性和数据性质、用户的奇异性和推荐的项目类型等。因此, 需要人为确定一些 k 值, 正如 Schafer 等^[25]所述: “计算一个项目的特征是非

常昂贵的,需要与其他所有的项目相对比”。在本文中,我们为每个实验测试了若干邻近 k 值的效果,以发现最优邻近值。

候选邻近 C_i 应该能够覆盖待评项目的所有被评分信息特征。因此,待评项目的候选邻近规模与该项目所有评分数和评分奇异性有关。首先,如果待评项目收到更多的用户评论,则需要更多的候选邻近以覆盖评分的全部特征;其次,如果待评项目所有的被评论毫无规律(非常奇异),则会非常难以寻找到合适的邻近(需要的邻近数将会更多)。反之,如果待评项目的所有评论很普通,特点很强,则只需少量的邻近项目就能代表待评项目的特征。

3.4 项目推荐

为活跃用户做出推荐的计算过程可以分为两步:

(1) 基于待评价项目 i 的 k 个邻近,预测待评项目的评分(对每个不同 i 循环此过程);

(2) 寻找出预测评分最高的 N 个项目,推荐给活跃用户。

3.4.1 计算活跃用户对所有 k -邻近项目的加权评分总和作为用户 u 对待评价项目 i 的评分预测^[26]。形式上为

$$\begin{aligned} pred(u, i) &= \bar{r}_i + \\ &\frac{\sum_{p \in ratedItems(u)} sim(i, p) * (r_{u,i} - \bar{r}_u)}{\sum_{p \in ratedItems(u)} sim(i, p)} \\ &\Leftrightarrow ratedItems(u) \neq \emptyset \end{aligned} \quad (6)$$

3.4.2 为完成第二步(推荐),我们定义 X_u 为可能推荐给活跃用户 u 的项目集, N_u 为最可能被推荐的 top- N 个项目集,它必须满足下面的关系式:

$$\begin{aligned} X_u &\subset I \wedge \forall i \in X_u, r_{u,i} = \cdot, p_{u,i} \neq \cdot \\ N_u &\subseteq X_u, \forall x \in N_u, \forall y \in (X_u - Z_u): p_{u,x} \geq p_{u,y} \end{aligned} \quad (7)$$

4 实验设计

4.1 测量指标

为验证本文提出改进方案的效果,我们应用下面的推荐精确度和推荐质量测量指标:平均绝对误差,覆盖率,准确率和召回率^[7, 27, 28]。

使用平均绝对误差(Mean Absolute Error) MAE 测量预测值聚合的平均误差。

设 $B_u = \{u \in U | r_{u,i} \neq \cdot \wedge p_{u,i} \neq \cdot\}$ 为项目 i 收到的所有评分集。

$$mae_u = \frac{1}{\#B_u} \sum_{i \in B_u} |r_{u,i} - p_{u,i}| \quad (8)$$

定义整个系统的 MAE 为

$$mae = \frac{1}{\#U} \sum_{u \in U} mae_u \quad (9)$$

应用覆盖率测量一个项目的 k 个相似项目的推荐能力,我们定义项目的覆盖率为

$$coverage_u = 100 \times \frac{\#\{i \in I | r_{u,i} = \cdot \wedge p_{u,i} \neq \cdot\}}{\#\{i \in I | r_{u,i} = \cdot\}} \quad (10)$$

系统覆盖率为

$$coverage = \frac{1}{\#U} \sum_{u \in U} coverage_u \quad (11)$$

应用准确率和召回率测量推荐的质量。准确率是指推荐的相关项目数量占总推荐项目数量的比例;召回率是指推荐的相关项目数量占总相关项目数量的比例。我们根据系数 θ 判断一个项目是否相关,为计算准确率和召回率,我们重新定义 J. Serrano 和 N. Srinivas 等^[29, 30]中的准则为

$$\begin{aligned} X'_u &\subset I \wedge \forall i \in X'_u, r_{u,i} \neq \cdot, p_{u,i} \neq \cdot \\ Z'_u &\subseteq X'_u, \#Z'_u \leq N, \forall y \in (X'_u - Z'_u): p_{u,x} \geq p_{u,y} \end{aligned} \quad (12)$$

定义一个用户的准确率为

$$precision_u = \frac{\#\{i \in Z'_u | r_{u,i} \geq \theta\}}{\#Z'_u} \quad (13)$$

系统准确率为

$$precision = \frac{1}{\#U} \sum_{u \in U} precision_u \quad (14)$$

定义一个用户的召回率为

$$recall_u = \frac{\#\{i \in Z'_u | r_{u,i} \geq \theta\}}{\#\{i \in I | r_{u,i} \geq \theta\}} \quad (15)$$

系统召回率为

$$recall = \frac{1}{\#U} \sum_{u \in U} recall_u \quad (16)$$

下一节的实验都将测量系统指标值。

4.2 实验数据和参数设置

本文提出的方法应用于传统协同过滤方法之前的预处理过程。并应用调整余弦相似度,在实验中,我们通过上节提出的四个推荐质量测量指标检验该组合方法对推荐效果的影响,并与传统推荐系统中的皮尔逊相关性方法得出的推荐质量进行对比分析。

在实验中,我们主要计算经过前期近邻帕累托优化预处理的调整余弦相似度^[31](Pareto Dominance Adjust Cosine)、皮尔逊相关系数^[17]

(Pearson Correlation) 和调整余弦相似度^[32] (Adjust Cosine), 基于这三种相似性, 计算推荐结果, 并分别对这三种推荐结果测量 Mae, Coverage, Precision, Recall 水平, 进行对比分析。在图 2 中, 我们依次用 PDAC, PEA, ACO 代表基于三种方法得出的推荐质量测量值。

本实验应用 Movielens 1M 数据 (<http://grouplens.org/datasets/movielens/>) 做测试。它包含 6040 个用户对将近 3900 部电影的 1000209 条匿名评分信息。我们应用 20% 的评分数据作为测试数据, 对实验进行交叉验证, 相关实验参数如表 1 所示。

表 1 实验参数

	预测		推荐		交叉验证	
	k	K 值个数	N	k	θ	用户 (%) 项目 (%)
Movielens	50, ..., 600	56	{1, ..., 20}	100	5	20 20

4.3 实验结果分析

图 2 表示应用本文的改进方法对 Movielens 中电影推荐质量改进的测量指标曲线对比图。图 A, B, C, D 分别为基于三种方法推荐结果得出的 Mae, Coverage, Precision, Recall 对比图。

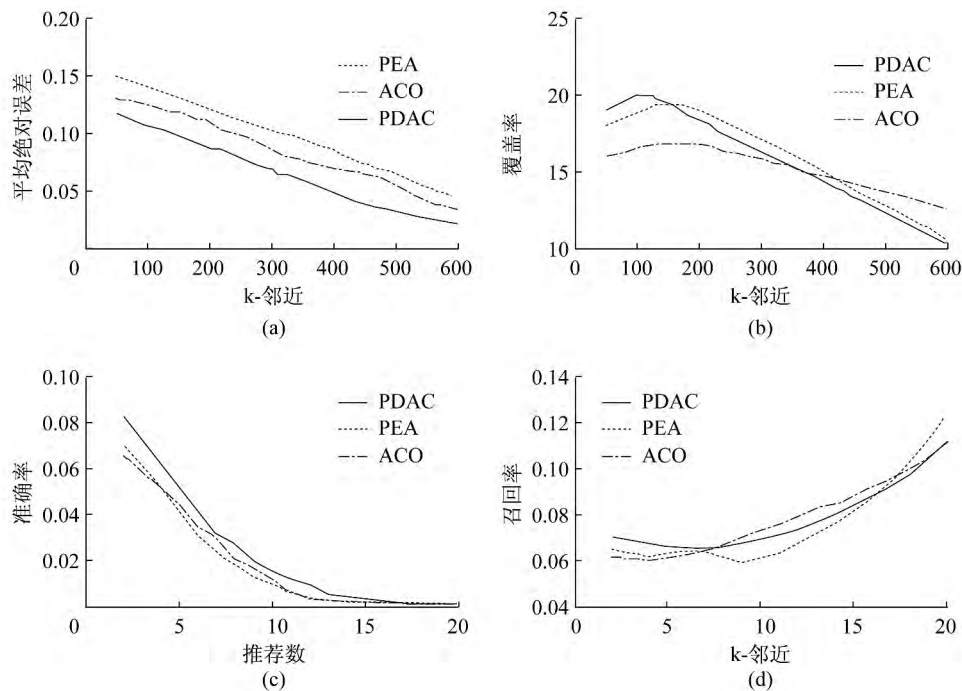


图 2 新方法与传统方法的推荐质量测量指标对比图。A: MAE, B: Coverage, C: precision, D: Recall

图 2(a)表明对 MAE 的改进, 基于帕累托改进的调整余弦相似性方法对 MAE 有很大提升, 无论 k —邻近规模处于什么水平, 都小于其他两种方法得出的值。当 k 值越来越大时, 三条曲线靠的越来越近, 这是因为, 随着 k 值的增大, 三种方法的邻近项目都趋向于相似, 从而误差也趋于相似。

图 2(b)表明, 相对于余弦相似度 (ACO) 方法得出的测量值, PDAC 方法在覆盖率方面有显著提升, 在 k —邻近处于低值时, 这一提升更明显, 甚至比 PEA 方法得到的覆盖率更高。但是在非常高的 k 值处, 覆盖率却处于最低处, 体现出在高邻近规模时, 该方法会过滤掉一些特别奇异的评分项目。

图 2(c)和 2(d)分别是对准确率和召回率的改进, 在这两个指标测量中, 我们使用 $K=100, \theta=5$ 。

本文提出的方法对这两个指标都有改进。随着项目推荐数的增长, 准确率一直都比另外两种方法高。召回率在项目推荐数少时, 明显高于另外两种方法, 体现出该方法推荐的最相关项目最符合用户的需求, 但随着项目推荐数的增多, 召回率并没有显著改进, 因为, 随着推荐数的增多, 过滤掉的相关项目就更少了, 和前面三个图的原因是一样的。

总结四个测量指标值对比图, 我们可以看出, 基于帕累托最优的改进方法在平均绝对误差、覆盖率、准确率和召回率等各方面都有改进, 特别是在 k —邻近值和推荐数处于低值的阶段, 改进更明显, 同时也表现出, 在 k —邻近和推荐数高值处, 指标值趋于一致, 或是比另外两种方法表现稍差。这些现象都在我们的预期之中, 因为随着邻近的增多, 被过滤掉

的相似项目就越来越少了,与传统的基于项目协同过滤方法所用数据就越来越相似了;同样的道理,随着推荐项目数增多,被推荐的项目种类也趋于相似了,差异主要表现在,推荐项目排序不一样,基于帕累托优化的项目协同过滤方法优先把用户最期望的项目优先推荐,排在前面。

5 结论与展望

传统基于项目的协同过滤方法并不能为待评价项目选择足够多的有代表性相似项目,从而不能保证后验预测评分的准确性。本文提出的方法主要应用于前期项目相似性计算的邻近筛选过程,保留下来的邻近项目能完全反应待评项目的特征,预期这一改进能减少计算量,并提高了推荐质量。

从本文的实际数据库实验可以看出,PDAC 方法无论在评分预测上还是项目推荐质量上,都得到了显著提升。这就表明一些相关性较差的项目确实对相似性计算和推荐质量有严重影响,然而传统方法却并不能把这些次相关项目排除掉。从本文的实验结果可以看出,在 k -邻近低值处和推荐项目数量少时,这一算法改进所得效果更显著,这也是非常重要的,因为用户往往只关心推荐最靠前项目。

这一方法也适用于其他协同过滤推荐系统,包括基于用户和基于知识的推荐,因为它并没有改变已有算法的结构,只是在传统算法前加了一个预过滤过程,本研究预过滤的是邻近项目,将来研究也可以基于帕累托最优思想预过滤邻近用户、邻近相关知识等等。另外,今后研究也可以着重应用项目奇异性指标,或用户的奇异性指标——活跃用户所评价的项目和其他用户所评价项目数量的比例,用于提高推荐的准确率或推荐可信度。

参考文献:

- [1] Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions[J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(6): 734-749.
- [2] Campos L M, Fernández-Luna J M, Huete J F, et al. Combining content-based and collaborative recommendations: a hybrid approach based on Bayesian networks[J]. International Journal of Approximate Reasoning, 2010(53): 785-799.
- [3] Mehrbakhsh N, Karamollah B, Othman I, et al. Collaborative filtering recommender systems[J]. Res. J. Appl. Sci., Eng. Technol, 2013(5): 4168-4182.
- [4] Bobadilla J, Ortega F, Hernando A. A collaborative filtering similarity measure based on singularities [J]. Information Processing and Management, 2011, 48(2): 204-217.
- [5] Bobadilla J, Serradilla F, Bernal J. A new collaborative filtering metric that improves the behavior of recommender systems [J]. Knowledge-Based Systems, 2010(23): 520-528.
- [6] Bobadilla J, Hernando A, Ortega F, et al. A framework for collaborative filtering recommender systems [J]. Expert Systems with Applications, 2011, 38(12): 14609-14623.
- [7] Antonopoulos N, Salter J. Cinema screen recommender agent: combining collaborative and content-based filtering[J]. IEEE Intelligent Systems, 2006: 35-41.
- [8] Serrano J, Viedma E H, Olivas J A, et al. A Google wave-based fuzzy recommender system to disseminate information in university digital libraries 2. 0, Information Sciences [J]. Artificial Intelligence Review, 2011, 181(8): 1503-1516.
- [9] Michael J P. A Framework for Collaborative, Content-Based and Demographic Filtering[J]. Artificial Intelligence Review, 1999, 12(13): 393-408.
- [10] Burke. Knowledge-based recommender systems [J]. Encyclopedia of library and information Science, 2000, 69(32): 180-200.
- [11] Antonio R A, Manuel Luque, Tomás García-Saiz. Recommender system in collaborative learning environment using an influence diagram[J]. Expert Systems with Applications, 2013 (40): 7193-7202.
- [12] Sarwar B, Karypis G, Konstan J, et al. Item-Based Collaborative Filtering Recommendation Algorithms[J]. Hong Kong, 2001(5): 1-5.
- [13] Al-Shamri M Y, Bharadwaj K K. Fuzzy-genetic approach to recommender systems based on a novel hybrid user model[J]. Expert Systems with Applications, 2008, 35(3): 1386-1399.
- [14] Barragán A B, Costa E, Burguillo J C, et al. A hybrid content-based and item-based collaborative filtering approach to recommend TV programs enhanced with singular value decomposition [J]. Information Sciences 2010, 180 (22): 4290-4311.
- [15] Bobadilla J, Ortega F, Hernando A, et al. Improving collaborative filtering recommender system results and performance using genetic algorithms [J]. Knowledge-Based Systems, 2011, 24(8): 1310-1316.
- [16] Hansen P. Tabu search for multi-objective optimizations: MOTS [C]. University of Cape Town, South Africa, Proceedings of the 13th International Conference on Multiple Criteria Decision Making, 1999.
- [17] Deb K, Agrawal S, Pratap A, et al. A fast elitist non-dominated sorting genetic algorithm for multi-objective optimisation: NSGA-II[C]. PPSN VI, Paris, France, Proceedings of the 6th International Conference on Parallel Problem Solving from Nature, 2000: 849-858.
- [18] Horn J, Nafpliotis N, Goldberg D E. A niched Pareto genetic algorithm for multiobjective optimization [J]. IEEE World Congress on Computational Intelligence, 1994: 82-87.
- [19] Ulungu E L, Teghem J, Fortemps P H, et al. MOSA method, a

- tool for solving multiobjective combinatorial optimization problems[J]. Journal of Multi-criteria Decision Analysis, 1999 (8): 221-236.
- [20] Trabelsi W, Wilson N, Bridge D, et al. Comparing approaches to preference dominance for conversational recommenders[C]. ICTAI'10, Arras, France, 22th IEEE International Conferences on Tools with Artificial Intelligence, 2010: 113-120.
- [21] Pappelis M, Plexousakis D, Kutsuras T. Alleviating the sparsity problem of collaborative filtering using trust inferences [J]. Lectures Notes on Computer Science, 2005, 3477: 224-239.
- [22] Linden, Smith B, York J. Amazon.com recommendations: item-to-item collaborative Filtering [J]. Internet Computing IEEE 2003, 7(1): 76-80.
- [23] Bobadilla J, Hernando A, Ortega F, et al. Collaborative filtering based on significances[J]. Information Sciences, 2012, 185(1): 1-17.
- [24] Hernando A, Bobadilla J, Ortega F, et al. Incorporating reliability measurements into the predictions of a recommender system[J]. Information Sciences, 2013(218): 1-16.
- [25] Schafer J B, Frankowski D, Herlocker J, et al. Collaborative filtering recommender system [J]. LNCS, 2007 (4321): 291-324.
- [26] Lee S K, Cho Y H, Kim S H. Collaborative filtering with ordinal scale-based implicit ratings for mobile music recommendation[J]. Information Sciences, 2010, 180(11): 2142-2155.
- [27] Deshpande M, Karypis G. Item-Based Top-N Recommendation Algorithms[J]. ACM Transactions on Information Systems, January 2004, 22(1): 143-177.
- [28] Hernández F, Gaudioso E. Evaluation of recommender systems: a new approach [J]. Expert Systems with Applications, 2008, 35(3): 790-804.
- [29] Serrano J, Viedma E H, Olivas J A, et al. A Google wave-based fuzzy recommender system to disseminate information in university digital libraries 2.0[J]. Information Sciences, 2011, 181(8): 1503-1516.
- [30] Srinivas N, Deb K. Multi-objective optimization using non-dominated sorting in genetic algorithms [J]. Evolutionary Computation, 1994.
- [31] Aher S B, Lobo L M R J. Combination of machine learning algorithms for recommendation of course in E-Learning System based on historical data[J]. Knowledge-Based Systems, 2013 (51): 1-14.
- [32] Herlocker J L, Konstan J A, Riedl J T, et al. Evaluating collaborative filtering recommender system [J]. ACM Transactions of Information Systems, 2004, 22(1): 5-53.

(上接第 33 页)

功能与花费成本的比值(根据价值工程理论 $V = \frac{F}{C}$

得到),其新增价值为所获价值与花费成本的差值($V_{i\text{新增}} = V_i - C_i$);②建设项目价值链某一环节的利益相关者所花费的成本直接影响到下一环节利益相关者的价值增值;③建设项目各利益相关者的新增价值都大于零($V_i - C_i > 0$),是项目价值管理成功的表现,是项目各利益相关者决定参与项目的根源所在。

参考文献:

- [1] 戚安邦等. 项目管理学[M]. 北京: 科学出版社, 2012: 110.
- [2] Anuradha A, Sekhar V. Systemic approach to project management: a stakeholders perspective for sustainability. India Conference[C]. Hyderabad, 2011: 1-4.
- [3] Kate D. Different stakeholder groups and their perceptions of project success[J]. International Journal of Project Management, 2014, 32(2): 189-201.
- [4] Harrell B L. Effective owner-contractor relationships [J]. Power Engineering International, 2010, 18(7): 42-46.
- [5] Moree W. The influence of different contract types on the effectiveness of owner-contractor relationship in construction projects[D]. Delft: Delft University of Technology, 2013.
- [6] Mohamed M G, Aminah R F. Fuzzy preference relations consensus approach to reduce conflicts on shared responsibilities in the owner managing contractor delivery system[J]. American Society of Civil Engineers, 2011, 137(8): 609-618.
- [7] Rashmi A, Timothy J K. Managing stakeholders for project management success: an emergent model of stakeholders[J]. Journal of General Management, 2010, 35(3): 67-80.
- [8] 戚安邦, 高山. 代建制企业为核心的项目全团队价值管理模型与方法[J]. 科学学与科学技术管理, 2010(6): 194-199.
- [9] Claus B, Daniel J, Alexander K. Behavior of internal stakeholders in project portfolio management and its impact on success [J]. International Journal of Project Management, 2013, 31(3): 830-846.
- [10] Hanh T, David G C. The likelihood of subcontractor payment: downstream progression via the owner and contractor[J]. Journal of Financial Management of Property and Construction, 2012, 17(2): 135-152.
- [11] 许婷. 工程项目采购供应链中的竞合博弈[J]. 中国管理科学, 2009, 17(1): 83-88.
- [12] Carolyn B, Akintola A, Eamon F. An analysis of success factors and benefits of partnering in construction [J]. International Journal of Project Management, 2000, 18(6): 423-434.
- [13] 盛峰, 戚安邦, 王进同. 政府投资项目利益相关者博弈与合作伙伴关系管理模式研究[J]. 生产力研究, 2008, (7): 86-88.
- [14] Huňka F, Žáček J, Meliš Z et al. REA value chain and supply chain[J]. Faculty of Economics & Administration, 2011, (16): 68-77.