

面向C2M个性化定制的协同过滤推荐算法研究

曾宪均, 左清, 肖金根

江西理工大学, 江西 赣州

收稿日期: 2022年2月24日; 录用日期: 2022年3月31日; 发布日期: 2022年4月6日

摘要

协同过滤算法是现今电商模式下最有效和最广泛的推荐方法, 但是传统算法仅考虑用户兴趣, 忽略项目信息的影响。特别在C2M (用户直连制造) 模式下, 借助于工业定制平台, 项目的有效信息量倍增, 项目信息对于用户的个性化定制有重大的影响。针对这个问题, 本文构建用户满意度模型来预测项目的信息对用户的影响, 结合传统协同过滤算法, 综合考虑用户兴趣和满意度来产生推荐结果。通过实验结果表明, 改进协同过滤算法的具有可行性, 并提高了推荐的有效性。

关键词

协同过滤, C2M, 用户满意度, 模糊时间窗

Research on Collaborative Filtering Recommendation Algorithm for C2M Personalized Customization

Xianjun Zeng, Qing Zuo, Jingen Xiao

Jiangxi University of Science and Technology, Ganzhou Jiangxi

Received: Feb. 24th, 2022; accepted: Mar. 31st, 2022; published: Apr. 6th, 2022

Abstract

Collaborative filtering algorithm is widely used in e-commerce mode. Traditional algorithms only take into account the interest of users, which ignore the impact of item information. Especially in the C2M (Customer-to-Manufacturer) mode, with the help of the industrial customization platform, the item information is increased rapidly and makes a significant impact on the user's personalized customization. To solve the problem, this text builds a customer satisfaction model to

predict the impact of item information on users, and combines traditional collaborative filtering algorithm to generate recommendation results that take into consideration the interest and satisfaction of users. Experimental results show that the improved collaborative filtering algorithm is feasible and improves the effectiveness of the recommendations.

Keywords

Collaborative filtering, C2M, Customer Satisfaction, Fuzzy Time Windows

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着电子商务和互联网技术的快速发展,当前的商业模式急速演变成新的模式,而用户的需求也从最初对产品功能(价格、质量、性能)等的需求,变成对消费体验(在线互动,信息共享)等的需求,更高等级是对服务生态系统(产品生态、产品生命周期服务)等的需求[1]。由此 C2M (用户直连制造)商业模式被提出。C2M 模式需要满足的客户个性化定制的需求,工厂客户提供订单定制的平台,客户能够全程参与到工厂的产品设计和订单配置的过程中。由于产品属性的多样性和组合性,可以供客户定制的产品订单能够达到成千上万的数量级,因此通过个性化智能推荐算法,能为客户的订单配置提供建议,减少客户定制的难度和时间。

目前个性化推荐算法基本上有以下几种:基于协同过滤的推荐技术、基于内容的推荐技术、基于关联规则的推荐技术等[2]。而协同过滤技术是当前推荐系统最广泛应用的和最有效的。协同过滤算法是根据用户历史评分数据来预测目标用户对项目评分。大量的学者研究协同过滤推荐技术。于洪提出了遗忘曲线模型的协同过滤算法[3],刘超慧等提出惩罚因子和时间权重的改进[4]。吴彦文等结合时间效应和信任度算法提高了算法推荐的速度和准确率[5]。Liu C 等提出了应有于大规模定制的协同过滤推荐方法[6]。上述学者研究了协同过滤推荐算法的改进方法,说明了时间参数能提高推荐结果的准确性。但是大多数改进算法仅考虑用户历史行为参数,针对的是用户的兴趣,而没有考虑项目的信息参数,如用户 A 与 B 的兴趣非常相似,根据协同过滤方法, A 喜欢的项目 x, B 也极有可能喜欢,但是如果考虑项目 x 的时间参数,比如电影的上映年份,因电影的年份久远会造成电影分辨率低,历史文化差异大,甚至不能获取观影资源等问题,即使推荐项目 x 给用户 B,也是无效推荐。即失效项目会影响用户的兴趣,降低推荐的成功率。

不同于电商的产品, C2M 模式推荐的是工厂的定制产品,具有更多的属性,并且 C2M 模式在实时定制平台的能够提供给用户更多的定制信息,比如存货数量,生产时间,发货时间等。这些信息虽然不会影响产品的品质,但是极大地影响用户下单的兴趣,如果一个订单的项目的属性超过用户的接受范围,那么用户将不考虑这个产品。为了描述项目本身信息对用户的影响,通过用户满意程度来表示用户对项目信息的接受程度。在量化用户的满意程度方面,余建军[7]和陈萍[8]等学者提出了模糊时间窗隶属函数来决定用户的满意度。曹霞[9]和袁杰[10]等学者通过服务质量评价来决定用户的满意度。

本文为了解决 C2M 模式项目信息利用的问题,提出基于用户满意度的协同过滤算法。利用用户对项目本身的时间信息的交互记录,通过模糊时间窗函数构建用户的满意度模型,计算当前用户对项目的时

间信息的价值大小。再结合协同过滤算法，同时考虑用户的兴趣和项目信息属性，得到更具有价值的推荐结果。

2. 算法研究与设计

2.1. 基于用户的协同过滤算法

协同过滤技术分为基于用户的协同过滤算法和基于项目的协同过滤算法[11]。两种的算法的主要区别是根据现有的评分数据来找出相似用户还是相似项目，Sarwar [12]比较了两种方法的性能和实用性。吕锋[13]和傅鹤岗[14]的研究说明了基于项目的协同过滤算法更适用实时性和个性化产品推荐的系统，由于C2M模式性化项目数量大且评分数据稀疏，不适用于基于用户的协同过滤算法。本文采用基于项目的过滤算法进行研究。其步骤如下

1) 收集用户数据，生成用户 - 评分矩阵

定制平台通过收集用户对产品的历史互动数据，转换成用户对产品的评分数据，用户 u 对项目 i 评分记为 R_{ui} ，通过量化处理转换为 $m \times n$ 的用户 - 评分矩阵，评分矩阵表示为 R ：

$$R = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1n} \\ R_{21} & R_{22} & \cdots & R_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ R_{m1} & R_{m2} & \cdots & R_{mn} \end{bmatrix} \quad (1)$$

2) 相似度计算

相识度计算目前常用余弦相似度法、皮尔逊相似度和修改余弦相似度计算方法，皮尔逊不适用于稀疏度较高的数据集[15]，而本文重点是关注用户满意度对推荐的影响，故本文选择余弦相似度计算出项目 i 和 j 之间的相识度，记为 $sim(i, j)$ ：

$$sim(i, j) = \frac{\sum_{i=1}^n R_{ui} * R_{uj}}{\sqrt{\sum_{i=1}^n R_{ui}^2} \sqrt{\sum_{i=1}^n R_{uj}^2}} \quad (2)$$

其中 R_{ui} 、 R_{uj} 表示用户 u 对项目 i 和 j 的评价。

3) 最近邻居集合

采用 Top-K 方法选出最近邻居集合，即通过设定最近邻居数 K ，取相似度最大的前 K 个项目作为最近邻居集 N ，本文选取 $K = 30$ 作为参数取值。

4) 预测评分值计算

通过最近邻居集对项目 i 的评分和相识度可以预测目标用户 u 对项目 i 的评分 P_{ui} ，计算公式如下：

$$P_{ui} = \bar{R}_i + \frac{\sum_{v \in N_u} sim(i, j) * (R_{vj} - \bar{R}_j)}{\sum_{v \in N_u} |sim(i, j)|} \quad (3)$$

其中， N_u 表示最近邻居中对项目 i 评价过的用户， \bar{R}_i 是目标项目的平均评分，考虑到不同用户评分的差异性， \bar{R}_j 取的是目标项目和最近邻居共同的项目的平均评分。

2.2. 用户满意度模型

用户满意度是用户对生产时间的要求及对服务体验的评价，用户的满意度是用户对服务能力评价的硬指标，对用户的购买需求有很大的影响[16]。所以本文研究通过项目的时间信息来生成用户满意度模

型。而为了量化用户满意度，本文通过模糊时间窗隶属函数处理生产时间，如下公式所示：

$$S_t = \begin{cases} 1, & e_i \leq t_i \leq l_i \\ \frac{L_i - t_i}{L_i - l_i}, & l_i \leq t_i \leq L_i \\ 0, & t_i \geq L_i \end{cases} \quad (4)$$

其中 t_i 表示产品实际生产时间， S_t 表示用户的满意度，实际生产时间的有效范围为 $[e_i, L_i]$ ，用户理想的生产时间范围为 $[e_i, l_i]$ ，实际生产时间在此之内表示用户能够完全接受，而随着时间的增加，用户满意度降低直至 0。如图 1(a) 所示。实际上人对时间的感知过程并不是线性的，而是指数衰减，参考艾宾浩斯遗忘曲线[17]，可以对满意度函数调整，如下公式所示。图形如如图 1(b) 所示。

$$S_t = \begin{cases} 1, & e_i \leq t_i \leq l_i \\ e^{-\alpha(t_i - l_i)}, & l_i \leq t_i \end{cases} \quad (5)$$

其中 α 为衰减常数， $\alpha > 0$ 。

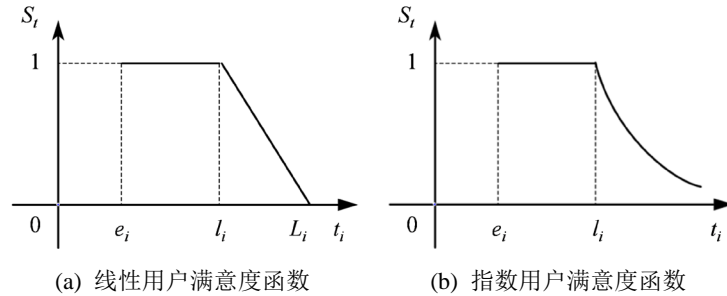


Figure 1. Relationship between project time and user satisfaction
图 1. 项目时间与用户满意度关系

2.3. 协同过滤推荐算法改进策略

基于以上模型，本文改进相似度的计算方法，根据用户的评分偏好和用户的满意度综合评价，对项目之间的相似度用加权的方式重新计算有效值，公式表示如下

$$SS(i, j) = \lambda * Sim(i, j) + (1 - \lambda) * S_t \quad (6)$$

其中的 λ 是融合参数， λ 越大表示系统推荐的方案关注用户的兴趣，相反说明系统用户的满意度影响更大。 λ 参数的取值根据应用环境有所变化，根据王真[18]的物流服务质量评价体系的研究，时效性对用户满意度影响约为 20%，故这里的取值为 0.8。通过优化的相似度可以得到优化的最近邻居集合 NS。

为了进一步对推荐结果进行优化，防止推荐无效结果，结合用户满意度得出预测评分值公式为：

$$PS_{ui} = S_t * P_{ui} \quad (7)$$

3. 实验

本文通过实验进行验证提出的方法可行性。实验环境为：CPU：i5-3470 双核 3.2 GHz，6 G 内存，Windows10 操作系统，Python3.8 开发语言，Pycharm11.0 集成环境。

3.1. 数据集

本文将采用 MovieLens 的 ml-1m 数据集进行实验。该数据集是 GroupLens 工作组收集电影评分数据

集, 包含 6040 个用户在 2000 年对 3952 个电影的 1,000,209 个评分, 评分值在 1~5 的整数, 分值越高, 表示电影越偏爱。数据包括 3 张数据表, 其内容如下表 1 所示。

Table 1. Structure description of data table

表 1. 数据表的表结构描述

USERS		MOVIES		RATINGS	
字段名称	字段类型	字段名称	字段类型	字段名称	字段类型
UserID	用户编号	MovieID	电影编号	UserID	用户编号
Gender	用户性别	Title	电影名称	MovieID	电影编号
Age	用户年龄	Genres	电影种类	Rating	评分值
Occupation	用户职业			Timestamp	评分时间

3.2. 实验结果与分析

1) 用户满意度模型验证

用户满意度关注的是项目时间信息的影响, 通过对 MovieLens 数据进行处理, 从而得到用户对电影上映年份之间的交互记录, 统计用户对同一时间上映的评论数量, 可以得到按电影年份的评论数量, 为了方便显示, 将年份分成了 8 个时间段, 并求出时间段内的评论数量的平均值。如图 2 所示。

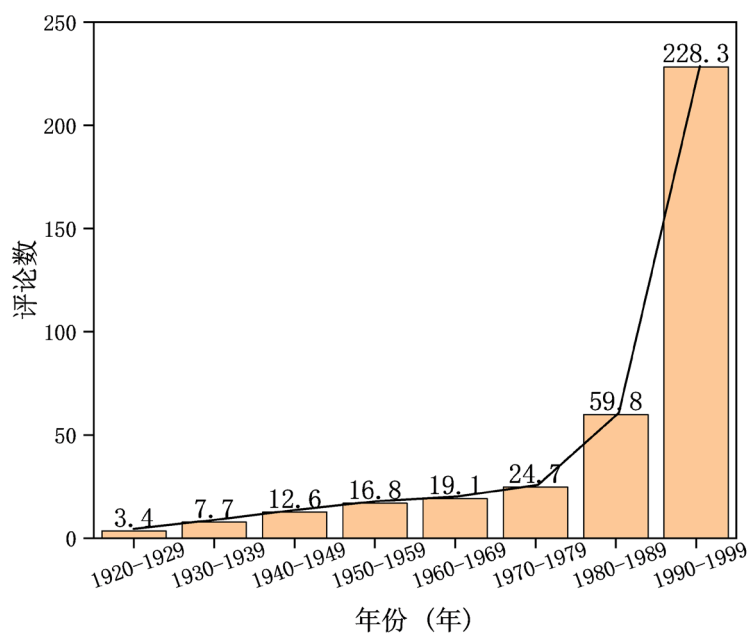


Figure 2. The average number of comments in different time periods

图 2. 不同时间段的评论数的平均值

可以从图 2 明显看出电影上映年份跟用户的评论数的关系, 电影上映年份距离当前时间越近, 评论数越多, 也就表示用户对项目的信息越满意, 并且整个趋势呈指数增长。与预期相符合。

2) 满意度衰减常数 α 的影响

为了确定 α 对算法的影响, 使用 MAE 评估推荐的准确率, MAE 值越小, 准确的越高, MAE 计算公式如下所示:

$$\text{MAE} = \frac{\sum_{i=1}^n |p_i - q_i|}{n} \quad (8)$$

其中 p_i 表示测试集评分, q_i 表示为实际评分, n 表示评分个数。

实验 α 按步长 0.1 进行实验, 实验结果如图 3 所示。

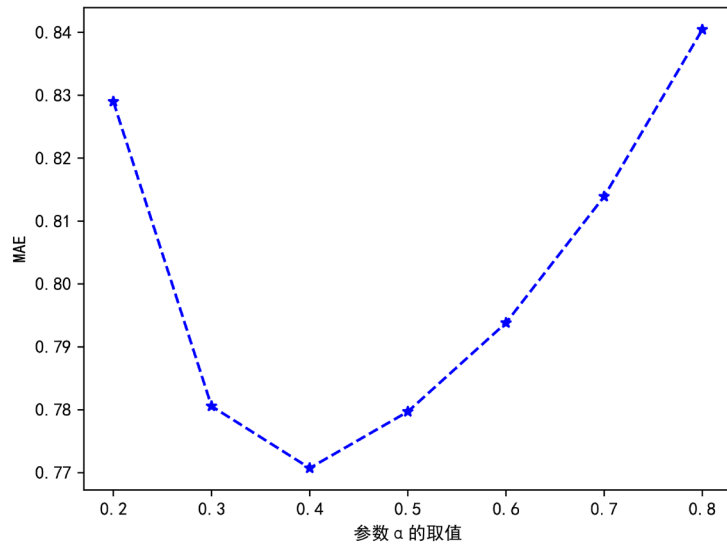


Figure 3. Relationship between α and MAE

图 3. α 与 MAE 的关系

根据实验结果, α 取 0.4 时 MAE 值最小。故实验选取 0.4 作为参数取值。

3) 比较不同测试集比率和 MAE 的关系

通过改变测试集比率比较传统协同过滤算法和本文改进的算法的推荐精度, 比较结果如图 4 所示。

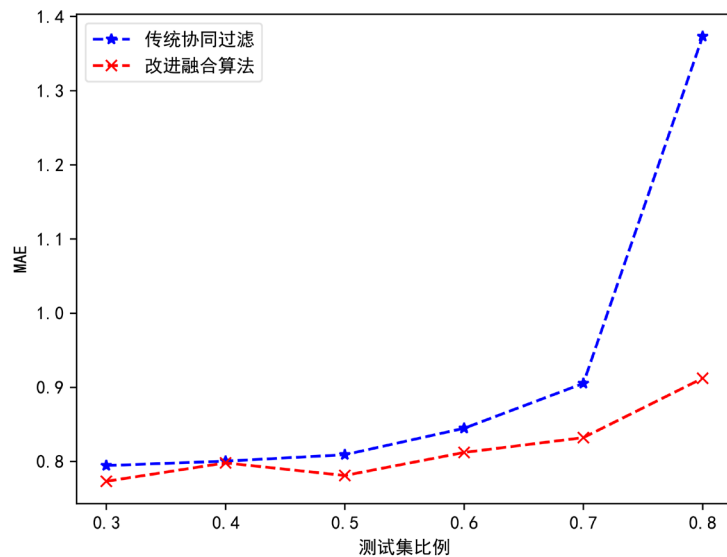


Figure 4. MAE comparison of recommended algorithms

图 4. 推荐算法的 MAE 比较

实验表明随着测试集比例的增加, MAE 呈现上升趋势, 这是由于训练集比例降低造成数据稀疏, 这也是协同过滤算法存在的缺点。而本文提出的改进算法, 能够通过项目信息得到满意度的权重, 不存在数据稀疏性的问题, 推荐的精度也比传统算法精度更高, 同时改进算法也能在一定程度上缓解传统推荐算法的冷启动和数据稀疏的问题, 在推荐结果上, 项目的价值也更符合用户的偏好, 更具有实际意义。

4. 结论

本文面向 C2M 模式, 考虑项目信息对用户产生的价值, 并通过用户满意度模型量化项目信息的价值大小, 并且通过实验验证了模型的准确性, 并结合传统协同过滤算法提出一种基于用户满意度的协同过滤推荐算法, 同时考虑了用户的历史评分偏好和用户对项目信息的满意度, 通过实验也证明了改进算法改进的算法能够推荐更加准确的结果, 产生更具有实际意义的项目, 提高了推荐的效率。

参考文献

- [1] Zhang, X.Y., Ming, X.G., Liu, Z.W., *et al.* (2019) A Framework and Implementation of Customer Platform-Connection Manufactory to Service (CPMS) Model in Product Service System. *Journal of Cleaner Production*, **230**, 798-819. <https://doi.org/10.1016/j.jclepro.2019.04.382>
- [2] 姚婷. 基于协同过滤算法的个性化推荐研究[D]: [硕士学位论文]. 北京: 北京理工大学, 2015.
- [3] 于洪, 李转运. 基于遗忘曲线的协同过滤推荐算法[J]. 南京大学学报(自然科学版), 2010, 46(5): 520-527.
- [4] 刘超慧, 韩传福, 陈天成, 孔先进. 融合惩罚因子和时间权重的协同过滤推荐算法[J]. 信息技术与网络安全, 2020, 39(5): 17-21. <https://doi.org/10.19358/j.issn.2096-5133.2020.05.004>
- [5] 吴彦文, 王洁, 王飞. 混合粒子群遗传算法的协同过滤推荐模型[J]. 小型微型计算机系统, 2017, 38(3): 527-530.
- [6] Liu, C. and Wu, Q.L. (2014) A Research on Mass Customization Personalized Recommendation Based on Collaborative Filtering. *Industrial Engineering Journal*, **17**, 24-28.
- [7] 余建军, 程文琪, 吴永忠. 考虑顾客满意度的生鲜外卖路径规划[J]. 工业工程与管理, 2021, 26(4): 158-167. <https://doi.org/10.19495/j.cnki.1007-5429.2021.04.020>
- [8] 陈萍, 李航. 基于时间满意度的 O2O 外卖配送路径优化问题研究[J]. 中国管理科学, 2016, 24(S1): 170-176.
- [9] 曹霞, 李玲. 基于 SERVQUAL 模型的网购快递服务质量评价体系研究[J]. 铁道运输与经济, 2015, 37(8): 93-98.
- [10] 袁杰. B2C 电子商务快递物流服务质量评价体系的构建研究[J]. 中国商论, 2015(26): 86-88.
- [11] 常昊, 杨盛泉. 基于协同过滤决策树的商品推荐算法的研究[J]. 价值工程, 2020, 39(9): 127-129.
- [12] Sarwar, B., Karypis, G., Konstan, J. and Riedl, J. (2001) Item-Based Collaborative Filtering Recommendation Algorithms. *Proceedings of the 10th International World Wide Web Conference (WWW10)*, Hong Kong, 1-5 May 2001, 285-295.
- [13] 吕锋, 冯壮壮, 张杨航. 面向用户的个性化产品服务系统协同过滤推介新方法[J/OL]. 吉林大学学报(工学版), 1-9. <https://doi.org/10.13229/j.cnki.jdxbgxb20210964>, 2022-02-23.
- [14] 傅鹤岗, 李冉. 基于用户实时反馈的协同过滤算法[J]. 计算机应用, 2011, 31(7): 1744-1747.
- [15] 王永贵, 李昕. 融合狼群算法和模糊聚类的混合推荐算法[J/OL]. 计算机工程与应用, 1-14. <http://kns.cnki.net/kcms/detail/11.2127.TP.20210419.1322.013.html>, 2022-02-23.
- [16] 李雅琴. 融合标签和时间信息的协同过滤推荐[D]: [硕士学位论文]. 兰州: 兰州大学, 2019.
- [17] 何云, 邹正兴. 基于客户时间满意度的集成物流多目标优化模型研究[J]. 物流技术, 2014, 33(7): 118-120+131.
- [18] 王真. B2C 电子商务快递物流服务质量评价体系构建研究[D]: [硕士学位论文]. 成都: 成都理工大学, 2013.