



天津科技大学学报

Journal of Tianjin University of Science & Technology

ISSN 1672-6510, CN 12-1355/N

《天津科技大学学报》网络首发论文

题目：融合用户人格与项目吸引力的群组推荐算法
作者：杨浩，史艳翠，刘凌云
DOI：10.13364/j.issn.1672-6510.20240163
收稿日期：2024-08-08
网络首发日期：2025-04-30
引用格式：杨浩，史艳翠，刘凌云. 融合用户人格与项目吸引力的群组推荐算法[J/OL]. 天津科技大学学报. <https://doi.org/10.13364/j.issn.1672-6510.20240163>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。



融合用户人格与项目吸引力的群组推荐算法

杨 浩, 史艳翠, 刘凌云
(天津科技大学人工智能学院, 天津 300457)

摘要: 传统的群组推荐算法在偏好建模时, 忽略了用户人格特征以及项目自身属性对用户偏好的影响。针对这一问题, 提出了一种融合用户人格与项目吸引力的群组推荐算法。首先, 利用基于评分与人格的混合相似度代替评分相似度进行群组划分; 其次, 综合用户信任度、专业度以及人格因子计算决策权重, 并在此基础上实现偏好融合; 最后, 根据每个项目的历史评分数据计算项目吸引力, 将其与群组偏好加权结合, 以此获取最终推荐列表。在公开数据集上的实验结果表明, 与基线算法的最优结果相比, 本文算法在评价指标 NDCG@5、NDCG@10 和 NDCG@20 分别提升了 4.04%、1.46% 和 1.21%, 验证了该算法的有效性。

关键词: 群组推荐; 偏好融合; 用户人格; 吸引力

中图分类号: TP391

文章编号: 1672-6510 (0000)00-0000-00

Group Recommendation Algorithm Incorporating User Personality and Item Attractiveness

YANG Hao, SHI Yancui, LIU Lingyun

(College of Artificial Intelligence, Tianjin University of Science & Technology, Tianjin 300457, China)

Abstract: Traditional group recommendation algorithms often overly prioritize the preferences of in-group users in preference modeling, neglecting the impact of user personality traits and item attributes. To address this issue, a group recommendation algorithm incorporating user personality and item attractiveness is proposed. Firstly, a hybrid similarity based on ratings and personality was utilized for group division, replacing the conventional ratings similarity. Secondly, decision weights were computed by integrating user trustworthiness, professionalism, and personality factors, achieving the aggregation of preferences. Lastly, item attractiveness was determined by the historical ratings data of each item, and was weighted and combined with group preferences to derive the final recommendation list. Experimental results on publicly available datasets demonstrated that the proposed algorithm improves 4.04%, 1.46% and 1.21% in the evaluation metrics NDCG@5, NDCG@10 and NDCG@20, respectively, compared with the optimal results of the baseline algorithm. The effectiveness of the algorithm is verified.

Key words: group recommendation; preference aggregation; user personality; attractiveness

互联网技术的快速发展, 极大地加快了信息的传播速度, 也促使网络数据发生爆炸式增长。这使用户在使用网络服务时, 容易受“信息过载”问题的困扰^[1]。推荐系统作为解决信息过载问题的重要技术之一, 得到了广泛的关注。目前, 推荐系统已被应用在各个领

域, 如电子商务、社交平台、娱乐软件等。

推荐系统通过分析用户的历史行为记录, 获取用户偏好, 从而向用户推荐产品或服务^[2]。传统的推荐系统大部分是针对单个用户设计, 主要用于满足个体的偏好。然而, 在现实生活中, 人们有时候会组团活

动,比如看电影、同事聚餐、结伴旅游等。在这些场景下,针对单个用户设计的推荐系统就难以适用。所以,近年来群组推荐成为推荐领域热门研究方向之一^[3]。群组推荐是一个复杂的过程。不同于个人推荐系统只需考虑单个用户的偏好,群组推荐需要考虑组内所有用户的偏好以及偏好间的差异。因此,需要设计合理的策略进行偏好融合,并使根据融合后偏好获得的推荐结果满足群组内尽可能多的用户,这也是群组推荐的主要任务。

心理学研究表明,用户的人格会影响其偏好和社交行为。在推荐领域,偏好的不同以及用户间的交互会影响最终的推荐结果。目前已有许多学者将用户人格特征引入推荐算法中,取得了一定成果。Nguyen等^[4]将用户的人格特征融入到推荐的生成过程,提高了用户的满意度。Ning等^[5]提出了一种好友推荐系统Personet,该系统利用大五人格模型(FFM)增强混合过滤朋友选择过程,所获推荐效果优于传统的好友推荐系统。大五人格模型是心理学中完整量化的人格信息模型,其将用户人格分为5个维度:开放性、尽责性、外向性、宜人性和神经质,从心理学角度出发,能够更为合理地帮助推荐系统进行推荐^[6]。Dhelim等^[7]提出了一个基于用户兴趣挖掘和元路径发现的个性化产品推荐算法,该算法利用用户的大五人格特征增强兴趣挖掘过程,并进行个性感知产品过滤,结果显示该算法提高了推荐结果的精度。上述研究证明了将用户人格应用于推荐系统的合理性和可行性。目前,群组推荐领域的研究对用户人格特征考虑较少,这可能影响群组推荐结果的满意度以及准确率。此外,现有的群组推荐算法在获取推荐结果时,更加关注群组偏好的影响,对项目属性利用不充分。为了弥补上述不足,本文提出融合用户人格和项目吸引力的群组推荐算法(group recommendation algorithm incorporating user personality and item attractiveness, PAGR),主要贡献如下:

(1)改进用户相似度计算方法,通过结合用户评分相似度与用户人格相似度得出用户的混合相似度。其中,在计算人格相似度时引入人格权重更新用户初始人格得分,再利用更新后的人格得分计算人格相似度。

(2)提出一种偏好融合策略。通过比较单个用户与群组整体人格得分上的差异,衡量群组内用户的人格因子,然后综合人格因子、信任度以及专业度,得出用户在群组内的决策权重,并在此基础上实现偏好融合。

(3)提出一种项目吸引力计算方法,将项目获得评分的时间划为若干评分周期,通过分析不同周期内项目热度及好评率变化计算项目吸引力,并且将其与群组偏好结合,以此获取推荐结果。

1 相关工作

1.1 群组推荐系统

1.1.1 群组划分

群组划分是实现群组推荐的基础,也是群组推荐系统的重要一步。群组划分是否合理,会影响用户的满意度以及推荐结果的准确率。根据不同划分标准,群组划分可大致分为3种方式(表1)。

表1 群组划分方式

Tab. 1 Group division method

方式	划分标准	优点
基于相似度	距离(相似度)	准确率高,对原始数据要求低
基于内容	标签、主题等	可解释性高
基于规则	关联规则、信任关系等	提高结果准确率与可信度

虽然3种划分方式各有优势,但是基于内容和规则的划分方式对数据质量要求较高,实现难度大。前者有严格的数据类别要求,后者对用户评分数据挖掘不充分^[8]。基于相似度的群组划分更易实现,且具有较高的准确率。基于相似度的群组划分方式包括模糊聚类、谱聚类、k-means聚类等。本文主要研究的是用户对单一维度项目评分下的群组推荐情况,因此文中一个用户只属于一个群组。由于模糊聚类允许用户同时属于多个群组,需要计算和多个群组的隶属度,增加了计算负担。谱聚类需要进行特征分割,复杂度较高,而k-means聚类容易实现,且效率高。因此,本文选用k-means算法进行群组划分。

1.1.2 偏好融合

偏好融合是群组推荐的研究重点。由于组内用户的偏好不可能完全相同,因此在对群组用户推荐之前,必须进行偏好融合。偏好融合可以发生在推荐过程的不同时期,先融合再推荐的方法称为模型融合,先推荐再融合的方法称为推荐结果融合^[9]。两种方法的具体流程如图1所示,其中推荐结果融合又可细分为推荐列表融合和预测评分融合^[10]。

两种偏好融合方法各有优劣,但相较于模型融合,基于推荐结果融合的方法效率更高。目前,对偏好融合策略的研究主要集中在对组内交互关系的利

用上。Wang 等^[11]基于用户属性和社会关系融合得到混合信任矩阵, 以此筛选出可信用户, 实现偏好融合。该方法考虑了用户间的信任关系, 有较高的准确性。Zhang 等^[12]设计了一种用于群体用户间粒度交互的多头注意机制, 其利用群组用户的交互数据进行建模, 充分考虑了用户与用户之间、用户与项目间的交互关系, 推荐效率显著提升。Xu 等^[13]提出了一种旅游群组推荐模型。该模型利用用户的信任关系和社会影响力, 设计一个综合群内和群间预测得分的共识模型, 提升了推荐性能。以上研究利用组内用户的交互信息改进推荐模型, 虽然均取得了一定效果, 但是都忽略了用户人格特征可能产生的影响。为了充分挖掘用户人格因素对用户交互及推荐结果的影响, 本文在偏好融合阶段, 引入用户人格因子计算决策权重, 然后在此基础上, 对组内用户进行偏好融合。

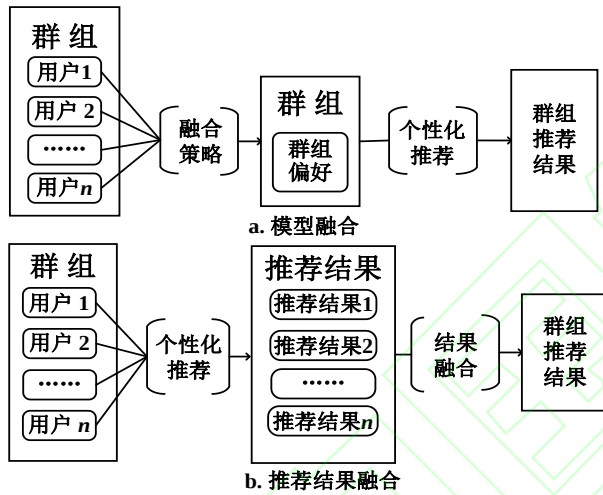


图1 群组偏好融合方法

Fig. 1 Group preference aggregation methods

1.2 用户人格特征

人格指个人所具有的独特而稳定的思维方式和行为风格, 它描述了人类在情感和人际关系上的差异^[14]。此外, 人格也会影响人的兴趣和社交表现。例如: 为了照顾其他人的情绪而放弃自己的想法、内向而不敢发表意见等。研究^[15]表明, 人格相似的人往往有相似的兴趣和偏好。因此, 人格因素也被广泛应用于推荐系统中。本文利用大五人格模型(FFM)衡量用户的人格特征。FFM 模型目前公认的最全面、应用最广泛的人格模型, 它将人格定义为五个因素: 开放性(O)、尽责性(C)、外向性(E)、宜人性(A)和神经质(N)^[16], 每种人格进行 1~7 的打分量化, 从而获取个人的人格得分。

2 相关定义

使用 u 和 v 代表用户和项目, 用户集和项目集分别表示为 $U=\{u_1, \dots, u_i, \dots, u_{|U|}\}$ 和 $V=\{v_1, \dots, v_n, \dots, v_{|V|}\}$; 群组集表示为 $G=\{g_1, \dots, g_a, \dots, g_{|G|}\}$; $r_{i,n}$ 表示用户 u_i 对项目 v_n 的评分, $P_{i,F}$ 为用户 u_i 的 F 型人格得分 ($F \in \{O, C, E, A, N\}$); 群组 g_a 的用户初始评分矩阵和用户初始人格得分矩阵分别表示为 $R_a^{s \times t}$ 和 $H_a^{s \times 5}$, 其中 s 表示群组 g_a 的用户数量, t 表示被评分项目的数量。

定义 1: 假设 $P_{i,F}$ 为用户 u_i 的 F 型人格得分, 该得分大小在 1~7 之间 (即 $P_{i,F} \in [1, 7]$)。将人格得分划分为 3 个区域, 用 Y_i 表示用户 u_i 的人格得分 $P_{i,F}$ 具体所在区域, 则 Y_i 可表示为

$$Y_i = \begin{cases} L(\text{低分区}), P_{i,F} \leq 3 \\ M(\text{中等区}), 3 < P_{i,F} < 5 \\ H(\text{高分区}), P_{i,F} \geq 5 \end{cases} \quad (1)$$

定义 2: 假设 D_F 表示所筛选偏好相似用户的 F 型人格得分的聚集度, 聚集度代表用户的 F 型人格得分分布的聚集程度, 若用户的 F 型人格得分分布越集中于一个区域, 则 D_F 越大, 用户的 F 型人格也就越相似。此处 D_F 的计算公式参考了文献[17]中的波动率计算公式, 为

$$D_F = \frac{\frac{1}{3} \sum_{T \in \{L, M, H\}} |d_T - d_{\text{avg}}|}{d_{\text{avg}}} \quad (2)$$

其中: T 代表人格得分所属区域 ($T \in \{L, M, H\}$), d_T 代表所筛选用户的 F 型人格得分分布在区域 T 的人数; d_{avg} 代表 F 型人格得分分布在 3 个区域人数的平均值。

假设表 2 为根据 3.2.2 节中筛选出的偏好相似用户的人格得分表。

表2 人格得分表

Tab. 2 Table of Personality Scores

	u_1	u_2	u_3	u_4	u_5	u_6	u_7	u_8	u_9	u_{10}
O	5	7	4	5.5	5.5	6	4.5	5.5	5	6
A	2	4	3	5.5	5.5	3	4	4.5	5	6.5
N	3	6	4.5	4	3.5	4	6	6	5	2
C	2.5	5.5	2	4.5	4.5	3.5	5	3	5	2.5
E	6.5	4	2.5	4	2.5	1.5	4	4	3	3

根据定义 1 统计表 2 中用户的人格分数分布情况分别为: O 型人格 {L:0 人, M:2 人, H:8 人}; A 型人格 {L:3 人, M:3 人, H:4 人}; N 型人格 {L:2 人, M:4 人, H:4

人}; C 型人格 {L:4 人, M:3 人, H:3 人}; E 型人格 {L:5 人, M:4 人, H:1 人}。然后依据公式(2)分别计算 5 种人格的聚集度, 其中 D_O 的计算过程(保留 2 位小数)表示为: $D_O = \frac{1}{3} \times (|0 - \frac{10}{3}| + |2 - \frac{10}{3}| + |8 - \frac{10}{3}|) / \frac{10}{3} = 0.93$, 其他同理, 得到最终结果为: $[D_O, D_A, D_N, D_C, D_E] = [0.93, 0.13, 0.27, 0.13, 0.47]$ 。

定义 3: $P_F^{g_a}$ 为群组 g_a 的 F 型人格得分, 考虑到用户人格得分存在差异, 为了使群组整体人格得分尽可能与组内大部分用户人格得分一致, 此处统计组内用户各型人格在 3 种区域分布情况, 然后根据不同分布情况分别计算 $P_F^{g_a}$:

(1) 群组内用户 F 型人格得分分布在任意一个区域占比大于或等于 0.7 时, $P_F^{g_a}$ 的计算公式为

$$P_F^{g_a} = \begin{cases} l \cdot \text{avg}(\{P_F^L\}) + m \cdot P_{mF}^M + h \cdot P_{mF}^H, l \geq 0.7 \\ l \cdot P_{mF}^L + m \cdot \text{avg}(\{P_F^M\}) + h \cdot P_{mF}^H, m \geq 0.7 \\ l \cdot P_{mF}^L + m \cdot P_{mF}^M + h \cdot \text{avg}(\{P_F^H\}), h \geq 0.7 \end{cases} \quad (3)$$

其中: l 、 m 和 h 分别表示群组 g_a 中用户 F 型人格分布在 L 区、M 区和 H 区的人数占比; $\{P_F^L\}$ 表示人格分布在 L 区组内用户的 F 型人格得分集合; $\text{avg}(\{P_F^L\})$ 表示 $\{P_F^L\}$ 的均值; P_{mF}^L 表示 L 区用户 F 型人格得分重复次数最多的分数, 若同一区域含多个重复出现次数最多的得分, 则选择更靠近中间区域得分的分数。例如: 在 L 区若重复出现次数最多得分为 1 和 3, 则 $P_{mF}^L = 3$, 其它同理。

(2) 排除(1)中分布情况, 若此时, 出现分布在任意一个区域人数占比不超过 0.1, 则 $P_F^{g_a}$ 的计算公式为

$$P_F^{g_a} = \begin{cases} l \cdot P_{mF}^L + (m+h) \cdot \text{avg}(\{P_F^{M \cup H}\}), l \leq 0.1 \\ m \cdot P_{mF}^M + (l+h) \cdot \text{avg}(\{P_F^{L \cup H}\}), m \leq 0.1 \\ h \cdot P_{mF}^H + (l+m) \cdot \text{avg}(\{P_F^{L \cup M}\}), h \leq 0.1 \end{cases} \quad (4)$$

其中, $\{P_F^{M \cup H}\}$ 表示人格分布在 M 区和 H 组内用户的 F 型人格得分集合。

(3) 若群组内用户 F 型人格得分分布未出现以上两种情况, 则 $P_F^{g_a}$ 的计算公式为

$$P_F^{g_a} = \text{avg}(\{P_F^{g_a}\}) \quad (5)$$

其中, $\{P_F^{g_a}\}$ 表示群组 g_a 中所有用户的 F 型人格得分集合。

假设表 3 为使用 k-means 算法划分后得到的群组 g_1 的用户评分数据以及人格得分数据表(评分为 0 表示用户未对该项目进行评分)。

表 3 群组 g_1 的评分数据和人格得分数据

Tab. 3 Rating data and personality score for group g_1

	评分数据										5 种人格得分				
	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}	O	A	N	C	E
u_1	5	0	4	4	3.5	0	3	5	2	3	5	2	3	2.5	6.5
u_2	0	4	0	3	5	5	0	2	0	1	7	4	6	5.5	4
u_3	5	1	2	0	0	2	4	5	0	3.5	4	3	4.5	2	2.5
u_4	0	0	0	5	0	3	3	2	4	4	5.5	5.5	4	4.5	4
u_5	0	2	0	0	4	0	5	0	5	0	5.5	5.5	3.5	4.5	2.5

以群组 g_a 为例, 组内用户的各型人格得分分布在 3 个区域的占比情况为: O 型人格 {L:0, M:0.2, H:0.8}; A 型人格 {L:0.4, M:0.2, H:0.4}; N 型人格 {L:0.2, M:0.6, H:0.2}; C 型人格 {L:0.4, M:0.4, H:0.2}; E 型人格 {L:0.4, M:0.4, H:0.2}。然后根据上述占比情况计算群组 g_1 的整体人格得分, 其中 $P_O^{g_1}$ 的计算过程(O 型人格 H 区占比为 0.8, 符合公式(4)中情况)表示为

$$P_O^{g_1} = 0 + 0.2 \times 4 + 0.8 \times (7 + 5 + 5.5 + 5.5 / 4) = 5.4 \quad (6)$$

其他同理, 得到最终结果为: $[P_O^{g_1}, P_A^{g_1}, P_N^{g_1}, P_C^{g_1}, P_E^{g_1}] = [5.4, 4, 4.2, 3.8, 3.9]$ 。

3 算法设计

3.1 算法整体框架

本文算法主要由群组划分、群组偏好建模以及群组推荐 3 个部分组成, 算法框架如图 2 所示。

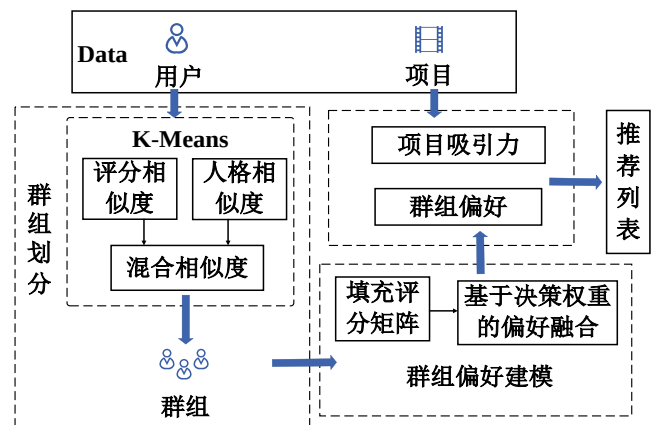


图 2 本文算法框架图

Fig. 2 The algorithm frame diagram of this paper

(1) 群组划分: 分别计算用户间的评分相似度以及人格相似度, 将二者结合得到用户间的混合相似度, 并在此基础上使用 k-means 算法实现群组划分。

(2) 偏好建模: 首先获取每个群组的初始评分矩

阵, 通过矩阵分解 (matrix factorization, MF) 方法获得预测评分, 并利用预测评分填充初始评分矩阵; 其次, 分别计算组内用户的信任度、专业度以及人格因子以获取用户的决策权重, 并根据该决策权重进行偏好融合, 获得群组偏好。

(3) 群组推荐: 首先计算各项目的吸引力, 将其与群组偏好加权结合获得各个项目的最终得分; 然后将最终得分排序, 生成推荐列表。

3.2 基于混合相似度的群组划分

群组推荐系统中, 推荐结果不仅受限于偏好融合策略, 同时还受限于群组内用户的相似度^[1]。组内用户相似度越高, 用户的偏好就越接近, 群组推荐结果的满意度也就越高。此外, 由于用户人格会对用户偏好产生影响, 因此本文在进行群组划分时同样考虑用户人格相似度。通过用户评分相似度以及人格相似度得出混合相似度, 然后以混合相似度为度量, 采用 K-Means 聚类算法实现群组划分。

3.2.1 评分相似度

首先, 利用 Pearson 相关系数计算用户 u_i 与 u_j 的评分相似度 $simr_{i,j}$, 为

$$simr_{i,j} = \frac{\sum_{v_n \in V_{i \cap j}} (r_{i,n} - \bar{r}_i)(r_{j,n} - \bar{r}_j)}{\sqrt{\sum_{v_n \in V_{i \cap j}} (r_{i,n} - \bar{r}_i)^2} \sqrt{\sum_{v_n \in V_{i \cap j}} (r_{j,n} - \bar{r}_j)^2}} \quad (7)$$

其中, $V_{i \cap j}$ 表示用户 u_i 与 u_j 的共同评分项目集; $\bar{r}_i$ 和 $\bar{r}_j$ 分别表示用户 u_i 与 u_j 的评分均值。

由公式(7)计算得到群组 g_1 中用户的评分相似度矩阵 $RS_1^{5 \times 5}$ 表示如下:

$$RS_1^{5 \times 5} = \begin{bmatrix} 1 & -0.0286 & 0.5020 & -0.4231 & -0.7559 \\ -0.0286 & 1 & -0.7303 & -0.0756 & 1 \\ 0.5020 & -0.7303 & 1 & -0.4899 & 1 \\ -0.4231 & -0.0756 & -0.4899 & 1 & 0 \\ -0.7559 & 1 & 1 & 0 & 1 \end{bmatrix} \quad (8)$$

3.2.2 人格相似度

使用 FFM 模型刻画群组用户的人格特征, 因此, 对于用户人格相似度的计算是基于该模型下的 5 种人格得分, 所选数据集中包含这些信息。文献[18]提出了一种用户人格相似度计算方法, 公式为

$$simp_{i,j} = \frac{\sum_F (P_{i,F} - \bar{P}_i)(P_{j,F} - \bar{P}_j)}{\sqrt{\sum_F (P_{i,F} - \bar{P}_i)^2} \sqrt{\sum_F (P_{j,F} - \bar{P}_j)^2}} \quad (9)$$

其中: $simp_{i,j}$ 表示用户 u_i 与 u_j 的人格相似度, $\bar{P}_i$ 和 $\bar{P}_j$

分别表示用户 u_i 与 u_j 的 5 种人格得分的均值。

该方法利用用户的 5 种人格得分代替用户对项目的评分, 然后利用 Pearson 相关系数公式进行计算。然而, 该方法未考虑 5 种人格因素对用户偏好的不同影响。文献[19]中描述了 5 种人格对用户偏好存在不同影响, 所以本文在计算人格相似度之前, 首先计算各型人格权重。

由于人格相似的用户往往具有相似的偏好, 因此通过分析数据集中那些偏好相似用户的各型人格的相似程度(以人格聚集度衡量)量化人格权重, 若这些偏好相似用户的某种人格相似程度越高, 则这种人格对用户偏好影响也就越大。计算人格权重的步骤如图 3 所示。

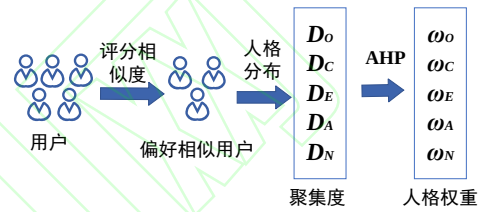


图 3 计算人格权重的步骤

Fig. 3 Steps for calculating personality weight.

首先, 计算用户的偏好(评分)相似度, 并设置相似度阈值 θ ; 其次, 筛选出相似度高于阈值 θ 的用户人格得分数据, 并统计他们各型人格得分的分布情况; 然后, 利用式(2)计算所统计用户的人格聚集度 D_F ; 最后, 利用层次分析法(AHP)^[20]求出各型人格权重, 为

$$[w_F, w_C, w_E, w_A, w_N] = \text{AHP}(D_O, D_C, D_E, D_A, D_N) \quad (10)$$

对于偏好相似的用户, 聚集度 D_F 越大, 用户 F 型人格得分分布就越集中, 此时用户的 F 型人格得分就越相似。即: 相似偏好的用户, D_F 越大, F 型人格越相似, 其对学生偏好影响也就越大。

在前文定义 2 中, 已经求出了所举例的偏好相似用户的各型人格的聚集度大小: $[D_O, D_A, D_N, D_C, D_E] = [0.93, 0.13, 0.27, 0.13, 0.47]$ 。接下来将求得的聚集度分别代表 5 种人格对用户偏好影响的相对重要性, 以此建立判别矩阵(以 Santy 的 1-9 标度方法给出, 具体参照层次分析法步骤), 判别矩阵为

$$\begin{bmatrix} & O & A & N & C & E \\ O & 1 & 7 & 6 & 7 & 5 \\ A & 1/7 & 1 & 1/3 & 1 & 1/5 \\ N & 1/6 & 3 & 1 & 3 & 1/3 \\ C & 1/7 & 1 & 1/3 & 1 & 1/5 \\ E & 1/5 & 5 & 3 & 5 & 1 \end{bmatrix} \quad (11)$$

将其输入到 AHP 算法中, 得出各型人格权重 ω_F 为: $\{\omega_O: 0.5601, \omega_A: 0.0514, \omega_N: 0.1134, \omega_C: 0.0514, \omega_E: 0.2239\}$ 。对群组 g_1 中用户的人格得分数据进行更新, 更新后的人格得分为

$$\hat{H}_1^{5 \times 5} = \begin{bmatrix} 2.8005 & 0.1028 & 0.3402 & 0.1285 & 1.4554 \\ 3.9207 & 0.2056 & 0.6804 & 0.2827 & 0.8956 \\ 2.2404 & 0.1542 & 0.5103 & 0.1028 & 0.5598 \\ 3.0806 & 0.2878 & 0.4536 & 0.2313 & 0.8956 \\ 3.0806 & 0.2878 & 0.3969 & 0.2313 & 0.5598 \end{bmatrix} \quad (12)$$

其中, $\hat{H}_a^{s \times 5}$ 为更新后的人格得分矩阵。根据更新后的人格得分, 利用式(9)计算用户的人格相似度, 求出群组 g_1 中用户的人格相似度矩阵 $PS_1^{5 \times 5}$ 表示为

$$PS_1^{5 \times 5} = \begin{bmatrix} 1 & 0.3742 & 0.5710 & 0.9174 & 0.13389 \\ 0.3742 & 1 & 0.1057 & 0.7109 & -0.5703 \\ 0.5710 & 0.1057 & 1 & 0.5052 & -0.3187 \\ 0.9174 & 0.7109 & 0.5052 & 1 & -0.1649 \\ 0.1339 & -0.5703 & -0.3187 & -0.1649 & 1 \end{bmatrix} \quad (13)$$

3.2.3 混合相似度

根据得到的评分相似度 $simr_{i,j}$ 和人格相似度 $simp_{i,j}$, 将两者结合得到用户 u_i 与 u_j 间的混合相似度 $SIM_{i,j}$, 计算公式为

$$SIM_{i,j} = \lambda simr_{i,j} + (1 - \lambda) simp_{i,j} \quad (14)$$

其中, λ 是评分相似度和人格相似度的权重调整参数, 此处 λ 的计算参考文献[17], $\lambda = 0.8 \times |V_{i \cap j}| / (|V_{i \cap j}| + 5)$, 其中, $|V_{i \cap j}|$ 表示用户 u_i 与 u_j 的共评分数量。

得出混合相似度后, 以该混合相似度为度量, 通过 k-means 算法对用户进行多次聚类, 实现群组划分。

群组 g_1 中用户的混合相似度矩阵 $MS_1^{5 \times 5}$ 根据公式(14)求出, 表示为

$$MS_1^{5 \times 5} = \begin{bmatrix} 1 & 0.2310 & 0.5434 & 0.3812 & -0.1331 \\ 0.2310 & 1 & -0.1915 & 0.4313 & -0.2114 \\ 0.5434 & -0.1915 & 1 & 0.1514 & -0.0173 \\ 0.3812 & 0.4313 & 0.1514 & 1 & -0.1272 \\ -0.1331 & -0.2114 & -0.0173 & -0.1272 & 1 \end{bmatrix} \quad (15)$$

3.3 群组偏好建模

本文算法通过群组内用户的历史评分数据获取群组偏好。主要步骤为: 首先构建各群组的用户初始评分矩阵, 然后利用矩阵分解获得预测评分矩阵, 根据获得预测评分填充初始评分矩阵中的空白值; 其次, 通过计算群组用户信任度、专业度以及人格因子衡量组内用户决策权重; 最后, 根据用户决策权重进行偏好融合, 获取群组偏好。

3.3.1 填充空白评分

群组划分完成后, 根据群组 g_a 用户的历史评分数据构建初始评分矩阵 $R_a^{s \times t}$ 。然后通过 MF 将该初始

评分矩阵分解为低维的用户特征矩阵 $P_a^{s \times k}$ 和项目特征矩阵 $Q_a^{k \times t}$, 其中 k 表示维度 ($k < s, t$)。将两个低维特征矩阵作点积运算进行拟合, 并不断迭代更新使结果逼近 $R_a^{s \times t}$ 。

$$R_a^{s \times t} \approx P_a^{s \times k} \cdot Q_a^{k \times t} = \hat{R}_a^{s \times t} \quad (16)$$

其中, $\hat{R}_a^{s \times t}$ 为群组 g_a 预测评分矩阵, 该矩阵内元素为用户的预测评分。

此处假定矩阵分解维度 k 为 2, 对群组 g_1 的用户初始评分矩阵 $R_1^{5 \times 10}$ 进行矩阵分解, 得到其预测评分矩阵, 表示为:

$$R_1^{5 \times 10} = \begin{bmatrix} 5 & 0 & 4 & 4 & 3.5 & 0 & 3 & 5 & 2 & 3 \\ 0 & 4 & 0 & 3 & 5 & 5 & 0 & 2 & 0 & 1 \\ 5 & 1 & 2 & 0 & 0 & 2 & 4 & 5 & 0 & 3.5 \\ 0 & 0 & 0 & 5 & 0 & 3 & 3 & 2 & 4 & 4 \\ 0 & 2 & 0 & 0 & 4 & 0 & 5 & 0 & 5 & 0 \end{bmatrix} \approx \begin{bmatrix} 3.2203 & 0.5241 \\ 0 & 1.5158 \\ 3.1791 & 0 \\ 1.0378 & 1.1194 \\ 0 & 1.3225 \end{bmatrix} \times \begin{bmatrix} 1.4845 & 0 \\ 0 & 1.5622 \\ 0.8925 & 0 \\ 0.5878 & 1.8955 \\ 0.1869 & 2.5422 \\ 0.1931 & 1.8634 \\ 0.9747 & 1.5725 \\ 1.4947 & 0.6515 \\ 0.2183 & 2.0660 \\ 1.0485 & 0.8213 \end{bmatrix}^T \quad (17)$$

$$= \begin{bmatrix} 4.7805 & 0 & 4.7194 & 1.5406 & 0 & 0 \\ 0.8187 & 2.3679 & 0 & 1.7488 & 2.0659 & 0 \\ 2.8745 & 0 & 2.8377 & 0.9264 & 0 & 0 \\ 2.8862 & 2.8731 & 1.8687 & 2.7319 & 2.5067 & 0 \\ 1.9339 & 3.8533 & 0.5940 & 3.0397 & 3.3619 & 0 \\ 1.5984 & 2.8244 & 0.6139 & 2.2863 & 2.4643 & 0 \\ 3.9631 & 2.3835 & 3.0989 & 2.7719 & 2.0795 & 0 \\ 5.1547 & 0.9876 & 4.7517 & 2.2805 & 0.8616 & 0 \\ 1.7857 & 3.1315 & 0.6940 & 2.5393 & 2.7322 & 0 \\ 3.8068 & 1.2449 & 3.3332 & 2.0075 & 1.0862 & 0 \end{bmatrix} = \hat{R}_1^{5 \times 10}$$

通过获得的预测评分对 $R_a^{s \times t}$ 中的空白值进行填充, 获得群组 g_1 的用户偏好矩阵 $FR_1^{5 \times 10}$, 表示为

$$FR_1^{5 \times 10} = \begin{bmatrix} 5 & 0.8187 & 4 & 4 & 3.5 & 1.5984 & 3 & 5 & 2 & 3 \\ 0 & 4 & 0 & 3 & 5 & 2 & 2.3835 & 2 & 3.1315 & 1 \\ 5 & 1 & 2 & 1.8687 & 0.5940 & 2 & 4 & 5 & 0.6940 & 3.5 \\ 1.5406 & 1.7488 & 0.9264 & 5 & 3.0397 & 3 & 3 & 2 & 4 & 4 \\ 0 & 2 & 0 & 2.5067 & 4 & 2.4643 & 5 & 0.8616 & 5 & 1.0862 \end{bmatrix} \quad (18)$$

3.3.2 用户决策权重计算

群组进行决策时, 组内每位用户都可以提出意见, 但是每位用户的决策权重大小不一。这是因为用户在参与群组决策时易受到自己信任用户的影响, 同时群组内专家用户的意见也被其他用户重视, 这就导致了用户决策权重的动态变化。信任度高和更专业的用户具有更大的决策权重。除此之外, 还同样考虑了用户人格对其在群组内决策权重的影响。综合以上因素, 本文在衡量用户决策权重时首先计算用户的信任度和专业度, 然后引入用户人格因子对二者加权, 以此获得群组内各用户的决策权重。

(1) 信任度

信任度代表某位用户对群组内其他用户的信任程度, 信任度越高的用户在群组中影响力越大, 其决策权重也就越大。本文通过用户间的共评分数量占比和相似度衡量用户间的信任度, 具体如公式(19)所示。

$$Tru_{i,j} = \frac{2 \times SIM_{i,j} \times \frac{|V_{i \cap j}|}{|V_{i \cup j}|}}{SIM_{i,j} + \frac{|V_{i \cap j}|}{|V_{i \cup j}|}} \quad (19)$$

其中: $Tru_{i,j}$ 表示用户 u_i 与 u_j 间的信任度, $|V_{i \cap j}|$ 表示用户 u_i 与 u_j 的总评分数量, $SIM_{i,j} = 1 + SIM_{i,j}$ (相似度的取值区间为 $[-1, 1]$, 此处是为了避免 $SIM_{i,j}$ 取负值时对信任度产生影响)。

根据公式(19)计算群组 g_1 中用户 u_1 与 u_2 的信任度, 表示为: $Tru_{1,2} = 2 \times 1.2310 \times \frac{4}{14} / (1.2310 + \frac{4}{14}) = 0.4638$, 其他同理, 得到群组 g_1 内用户的信任度矩阵 $T_1^{5 \times 5}$ 为

$$T_1^{5 \times 5} = \begin{bmatrix} 1 & 0.4638 & 0.5483 & 0.5675 & 0.3881 \\ 0.4638 & 1 & 0.4457 & 0.5407 & 0.3191 \\ 0.5483 & 0.4457 & 1 & 0.4856 & 0.3069 \\ 0.5675 & 0.5407 & 0.4856 & 1 & 0.3254 \\ 0.3881 & 0.3191 & 0.3069 & 0.3254 & 1 \end{bmatrix} \quad (20)$$

(2) 专业度

除信任度外, 用户专业度同样会影响用户决策权重的大小。用户进行决策时, 会更加关注组内专家用户的意见。本文以用户评分次数占比以及评分价值衡量用户专业度。评分价值是为了防止用户恶意刷分的影响, 如果某位用户评分数据不存在差异或差异很小, 则其评分价值就低, 此处选取用户评分的方差代表该用户评分价值。由于多数情况下, 用户的评分占比数较小, 因此为了减小误差, 将其作为指数函数的幂引入公式。

$$Pro_i = \sigma_i^2 \times e^{\frac{|v_i|}{|V_{g_a}|}} \quad (21)$$

其中, Pro_i 表示用户 u_i 的专业度; $|v_i|$ 为用户 u_i 的评分次数, $|V_{g_a}|$ 为群组 g_a 中所有用户评分数量; σ_i^2 为用户 u_i 历史评分的方差。

根据公式(21)计算 u_1 在群组 g_1 内的专业度 Pro_1 表示为: $Pro_1 = 0.9336 \times e^{\frac{8}{31}} = 0.9336 \times 1.2944 = 1.2085$ 。同理群组 g_1 内其他用户的专业度为: $[Pro_2, Pro_3, Pro_4, Pro_5] = [2.6968, 2.6729, 1.1124, 1.7066]$ 。

(3) 人格影响因子

人格特征在用户进行决策时也会产生影响, 人格差异越大的用户其自身偏好也会存在差异。本文考虑比较用户人格得分与群组整体的人格得分, 以此衡量用户人格因子。若二者差异越大, 则群组偏好与该用户偏好差异也越大。具体如式(22)所示, 当用户 u_i 人格得分与群组 g_a 人格得分差异越小, 用户 u_i 的人格因子越接近 1, 反之越接近于 0。

$$Per_i = e^{-\frac{1}{5} \sum_F \sqrt{(\hat{P}_{i,F} - \hat{P}_{F^{g_a}})^2}} \quad (22)$$

其中, Per_i 表示用户 u_i 人格因子, $\hat{P}_{i,F}$ 和 $\hat{P}_{F^{g_a}}$ 分别为用户 u_i 和群组 g_a 根据人格权重更新后的 F 型人格得分。

根据公式(22)计算用户 u_1 的人格因子 Per_1 表示为 $Per_1 = e^{-\frac{1}{5} \times (0.2240 + 0.1028 + 0.1361 + 0.0668 + 0.5822)} = 0.8006$ 。同理其他用户的人格因子为 $[Per_2, Per_3, Per_4, Per_5] = [0.7850, 0.7748, 0.9581, 0.8937]$ 。

(4) 决策权重

综合考虑用户的平均信任度、专业度以及人格因子, 获得初始决策权重计算公式为

$$Dw_i = \frac{Per_i \times (\overline{Tru_i} + \overline{Pro_i})}{2} \quad (23)$$

其中: Dw_i 表示用户 u_i 的初始决策权重; $\overline{Tru_i}$ 表示用户 u_i 的组内平均信任度。 Dw_i 为用户 u_i 的最终决策权重, $Dw_i = Dw_i / \sum_{u_i \in g_a} Dw_i$ (使群组 g_a 内用户决策权重之和为 1)。

根据式(23)计算用户 u_1 在群组 g_1 内的初始决策权重, 表示为 $Dw_1 = 0.8006 \times (1.2085 + 0.4919) / 2 = 0.6807$, 同理其他用户的初始决策权重为: $[Dw_2, Dw_3, Dw_4, Dw_5] = [1.2322, 1.2086, 0.7627, 0.9122]$ 。则群组 g_1 内所有用户最终决策为: $[Dw_1, Dw_2, Dw_3, Dw_4, Dw_5] = [0.1419, 0.2569, 0.2520, 0.1590, 0.1902]$ 。

3.3.3 偏好融合

根据填充后的评分矩阵获取用户偏好(评分), 然后将群组内每个用户决策权重与各自偏好相乘并累加获得群组偏好, 如公式(24)所示。

$$r_{g_a,n} = \sum_{u_i \in g_a} r_{i,n} \times Dw_i \quad (24)$$

其中 $r_{g_a,n}$ 为群组 g_a 对项目 v_n 的偏好。

根据式(24)求出群组 g_1 对项目 v_1 的偏好为

$$r_{g_1,1} = 0.1419 \times 5 + 0.2520 \times 5 + 0.1590 \times 1.5406 = 2.2145 \quad (25)$$

3.4 项目吸引力

群组决策时, 除群组本身偏好外, 项目吸引力也会影响推荐结果, 吸引力大的项目往往更受用户关注。本文以项目热度占比和好评率衡量项目吸引力。首先将项目初次获得评分时间到最后评分时间分为 x 个评分周期(每部项目得第一个评分周期为其首次出现评分的周期), 考虑到项目对用户的吸引程度会随时间而衰减, 同时引入时间衰减函数表示项目吸引力随时间的改变。表 4 是根据每部项目在各个周期内被评价次数占比以及好评率定义的项目热度等级以及项目好评等级表。

表 4 项目热度以及好评率等级的划分

Tab. 4 The level of popularity and positive ratings of the

	items					
数量占比区间	0	(0,0.2)	[0.2,0.4)	[0.4,0.6)	[0.6,0.8)	[0.8,1]
项目热度等级	0	1	2	3	4	5
项目好评等级	0	1	2	3	4	5

根据划分的项目热度等级以及项目好评等级计算项目吸引力, 为

$$Atr_n = \frac{x_r}{x} \sum_{z \in x} f(z) \cdot \frac{popular_n^z \times positive_n^z}{5} \quad (26)$$

其中: Atr_n 表示项目 v_n 的吸引力; $popular_n^z$ 和 $positive_n^z$ 分别表示第 z 个周期项目 v_n 的热度等级以及好评等级(本文将评分超过 3.5 的评价定义为好评), 当二者中有一项为 0 时, 则在该周期内, 项目 v_n 的吸引力为 0; x_r 表示项目 v_n 获得评分的周期数(若 x_r 越大, 则表示项目具有长期吸引力, Atr_n 也就越大); $f(z)$ 为时间衰减函数, $f(z) = 1/(1 + e^{-z})$ 。

假设表 5 为数据集中用户对项目评分的记录。

表 5 项目评分表

Tab. 5 Item rating table

项目 ID	评分数据
v_1	评分 5 5 3 2 3
	时间戳 2015-1 2016-4 2016-9 2020-8 2021-1
v_2	评分 4 3 3 4 1
	时间戳 2015-2 2016-4 2018-5 2018-7 2020-2
v_3	评分 1 3 2 2 1
	时间戳 2017-9 2018-3 2018-4 2020-5 2020-12

按照上文项目吸引力计算方法, 首先对评分时间划分周期, 此处假定一个周期为 2 年, 获得 3 个评分周期, 分别为: [2015-1, 2017-1], (2017-1, 2019-1], (2019-1, 2021-1]。由于每个项目得第一个评分周期为其首次出现评分的周期。所以对 v_1 和 v_2 来说其第一个周期为[2015-1, 2017-1], 而对 v_3 来说, 其首次出现评分的周期是(2017-1, 2019-1], 因此其第一个周期为(2017-1, 2019-1]。

以项目 v_1 为例, 其在第一个周期内获得 3 条评分记录, 评分次数占比为 0.6, 好评率占比约为 0.67, 因此其热度等级以及好评等级均为 4; 在第二个周期没有评分记录, 在第三个周期热度等级和好评等级分别为 3 和 0。根据公式(26)计算项目 v_1 的吸引力表示为:

$$Atr_1 = \frac{2}{3} \times \left(\frac{1}{1+e^2} \times \frac{4 \times 4}{5} + 0 + \frac{1}{1+e^0} \times \frac{3 \times 0}{5} \right) = 0.2543 \quad (27)$$

3.5 获取群组推荐结果

根据群组偏好和项目吸引力获取项目最终得分, 为

$$\hat{r}_{g_a, n} = \varphi_1 r_{g_a, n} + \varphi_2 Atr_n \quad (28)$$

其中 $\hat{r}_{g_a, n}$ 表示群组 g_a 对项目 v_n 最终偏好, φ_1 和 φ_2 为群组偏好与项目吸引力的权重因子 ($\varphi_1 + \varphi_2 = 1$)。根据项目最终得分进行排序, 选取前 N 个评分最高的项目进行推荐。

此处假设 φ_1 取 0.8 时群组推荐结果最佳, 则最终偏好为: $\hat{r}_{g_1, 1} = 0.8 \times 2.2145 + 0.2 \times 0.2543 = 1.8225$ 。

3.6 算法描述

本文群组推荐算法的描述如下:

算法 1 群组划分算法

输入 用户评分矩阵 $R^{s \times t}$ 和用户人格得分矩阵 $H^{s \times 5}$ 。

输出 k 个群组。

步骤 1. 分别计算评分相似度 $simr$ 和人格相似度 $simp$;

步骤 2. $SIM = \lambda simr + (1 - \lambda) simp$; // 计算混合相似度。

步骤 3. 肘部法则获取最佳 k 值;

步骤 4. **K-Means**(SIM, k); // 使用 K-Means 算法聚类。

步骤 5. **Return** k groups.

算法 2 群组推荐算法

输入 群组 g_a 的评分矩阵 $R_a^{s \times t}$ 和人格得分矩阵 $H_a^{s \times 5}$, 项目集 V 。

输出 Top-N 推荐列表。

步骤 1. $R_a^{s \times t} \approx P_a^{s \times k} \cdot Q_a^{k \times t} = \hat{R}_a^{s \times t}$; // 矩阵分解填充空白评分

步骤 2. **FOR** 用户 u_i **IN** 群组 g_a

步骤 3. 分别计算用户组内平均信任度 $\overline{Tru_i}$ 、用户专业度 Pro_i 以及用户人格因子 Per_i ;

步骤 4. $Dw_i = Per_i \times (\overline{Tru_i} + Pro_i) / 2$; // 初始决策权重

步骤 5. $Dw_i = Dw_i / \sum_{u_i \in g_a} Dw_i$; // 用户最终决策权重

步骤 6. **END FOR**;

步骤 7. **FOR** 项目 v_n **IN** 项目集 V

步骤 8. $Atr_n = \frac{x_r}{x} \sum_{z \in x} f(z) \cdot \frac{popular_n^z \times positive_n^z}{5}$; // 吸引力

力

步骤 9. **END FOR**;

步骤 10. $\hat{r}_{g_a, n} = \varphi_1 r_{g_a, n} + \varphi_2 Atr_n$; // 群组偏好

步骤 11. **SORT by** $\hat{r}_{g_a, n}$;

步骤 12. **Return** Top-N list.

4 实验与结果分析

4.1 实验数据集

为了评估算法的推荐性能, 选取 3 个公开数据集进行实验, 并采取随机划分的方式按 8 : 2 划分训练集和测试集。数据集的具体信息见表 6。

表 6 数据集

Tab. 6 The statistic of datasets

数据集	用户数	项目数	评分数	稀疏度/%
Personality ^[4]	1820	35196	1028751	98.39
ML-latest-small	600	9000	100000	98.15
PER_dataset ^[14]	279	745	5278	97.46

其中 Personality 和 ML-latest-small (<https://group.lens.org/datasets/movielens/latest/>) 均由明尼苏达大学 GroupLens 团队收集, 包含用户对电影的评分信息。此外 Personality 数据集还包含用户的人格得分信息。PER_dataset 为文献[14]收集, 包含用户对歌曲的评分信息以及用户的人格信息。本文利用 Personality 和 PER_dataset 数据集验证引入用户人格特征后的推荐效果, 利用 Personality 和 ML-latest-small 数据验证引入项目吸引力后的推荐效果。

4.2 评价指标

选取应用广泛的平均绝对误差(MAE)以及归一化折损累计增益(NDCG)作为算法性能的评价指标。

(1)MAE 用于衡量模型的预测值与实际值之间的误差大小, 它的值越小表示预测评分和实际评分的误差越小, 算法的效果就越好。公式为

$$MAE = \frac{\sum_{n=1}^c (\hat{r}_n - r_n)}{c} \quad (29)$$

其中, $\hat{r}_n$ 和 r_n 分别表示项目 v_n 的预测得分与实际得

分, c 表示推荐列表的项目数量。

(2)NDCG 常作为群组推荐算法的评价指标, 综合考虑项目在推荐列表中的位置和项目的实际评分衡量推荐列表的质量, 其取值在 0~1 之间, NDCG@N 越大, 说明推荐列表越准确。公式为

$$DCG@N = r_{rel,1} + \sum_{y=2}^N \frac{r_{rel,y}}{\log_2 y} \quad (30)$$

$$NDCG@N = \frac{DCG@N}{IDCG@N} \quad (31)$$

其中: DCG@N 表示用户推荐列表中前 N 个项目满意程度, $r_{rel,y}$ 表示用户对推荐列表中第 y 个项目的实际评分。IDCG@N 是理想情况下的 DCG@N 的值。

4.3 实验与分析

4.3.1 各参数和权重的选取

实验中各参数的值会对结果产生影响, 如图 4 所示, 其中图 4 中的(a)、(b)和(c)、(d)分别为在数据集 Personality 和 ML-latest-small 上 ϕ_1 对三个 NDCG 值和 MAE 值的影响情况, 由图(a)和(b)可知, 在 Personality 数据集上, 当 ϕ_1 取 0.6 时, 算法的三个 NDCG 值均大于 ϕ_1 取 0 和 1 时的结果, 并且此时算法的 MAE 值达到最低; 同理, 由图(c)和(d)可知, 在数据集 ML-latest-small 上, ϕ_1 取 0.7 时, 推荐效果最佳。图 4 中(e)为相似度阈值 θ 的选取对于算法 MAE 的影响, 在 Personality 和 PER_dataset 数据集上实验发现, 当 θ 取 0.9 时, 用户 MAE 值最小且开始趋于稳定, 因此实验时 θ 取 0.9; 图 4 中(f)为矩阵分解维度对于算法 MAE 的影响, 在 3 个数据集中分别选取误差最小时的维度进行后续实验; 图 4 中(g)为评分周期对于算法 MAE 值的影响, 选取结果最优时的周期时间进行后续实验。

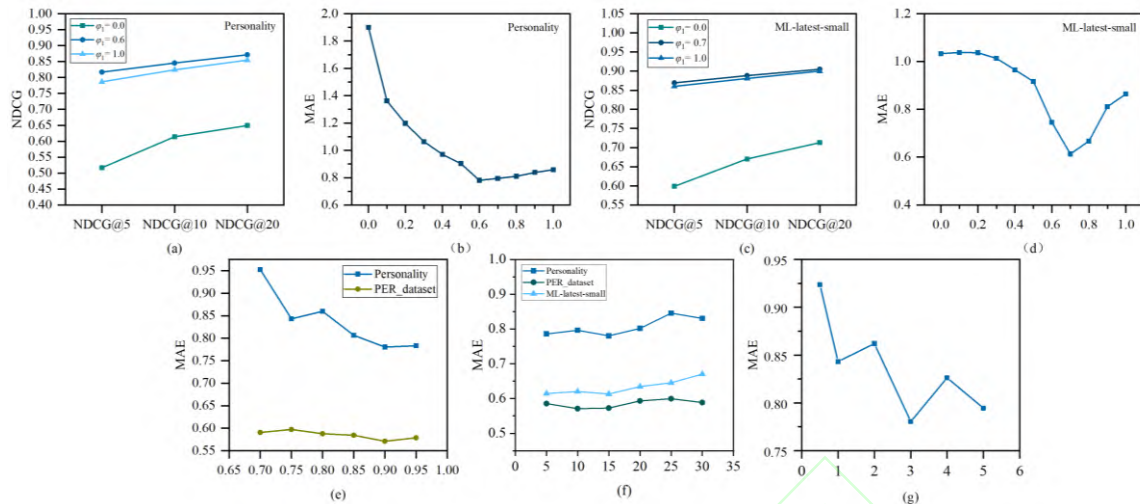


图4 参数取值对推荐结果的影响

Fig. 4 Influence of the value of parameter on the recommendation results

最终实验时各个参数和权重的选择结果见表7。

表7 各参数和权重的选取

Tab. 7 Selection of parameters and weights

数据集	分解维度 k	阈值 θ	评分周期	权重因子	
				ϕ_1	ϕ_2
Personality	15	0.9	3 年	0.6	0.4
PER_dataset	10	0.9	N/A	N/A	N/A
ML-latest-small	15	N/A	3 年	0.7	0.3

4.3.2 不同偏好融合策略的比较

在 Personality 和 PER_dataset 两个包含人格得分的数据集上探究了融入人格因子的偏好融合策略的效果,并将其与其他 3 种传统偏好融合策略进行对比,结果见表8。

表8 不同偏好融合策略的比较

Tab. 8 Comparison of different preference fusion strategies

数据集	融合策略	NDCG@N			MAE
		5	10	20	
Personality	均值策略	0.8005	0.8321	0.8674	0.8429
	最开心策略	0.6343	0.6975	0.7533	1.5020
	最小痛苦策略	0.6520	0.7129	0.7347	1.0487
	本文策略	0.8168	0.8454	0.8712	0.7805
PER_dataset	均值策略	0.6311	0.6626	0.7011	0.5658
	最开心策略	0.5765	0.6117	0.6501	1.2080
	最小痛苦策略	0.5314	0.5420	0.5568	0.5618
	本文策略	0.6335	0.6654	0.7031	0.5710

由表8可知,与传统偏好融合策略所相比,本文偏好融合策略在 Personality 和 PER_dataset 两个数据集上取得了更优的结果。在 Personality 数据集中,本文策略在所选评价指标上的结果均优于传统偏好融

合策略。其中,在 NDCG@5、NDCG@10 和 NDCG@20 上对比传统策略的最优结果分别提升了 2.0%、1.6%和 0.4%,在 MAE 上对比最优结果降低了 7.4%。在 PER_dataset 数据集中,本文策略的 MAE 值虽然略大于最小痛苦策略,但是在 NDCG@5、NDCG@10 和 NDCG@20 相较于传统策略的最优结果均有一定程度提升。这是因为本文根据用户决策权重进行偏好融合,一定程度上提高用户对推荐结果的满意度,同时降低了预测评分的误差。

4.3.3 与其他算法的对比

为了评价本文算法(PAGR)的综合性能,将本文算法与以下4个基线算法进行对比。

UL^[21]:一种考虑偏差的群组推荐改进聚合方法,结合项目评分的偏差增强群组偏好聚合过程。

PwAvg^[22]:一种人格感知聚合技术,使用五个基本人格特征确定群体中每个成员的影响程度,以此实现偏好融合并进行推荐。

IBGR^[23]:基于影响力的群组推荐算法,结合相似性和信任度计算成员间的相互影响,同时更新用户偏好,并进行偏好融合,以此获取推荐结果。

ATLGR^[24]:基于聚合 Top-k 列表的新颖的群组推荐方法,根据项目的相关性得分和排名位置衡量群体偏好。

DREAGR^[25]:一种基于依赖关系的增强型注意力群推荐算法,该算法引入了项目之间的依赖关系作为边信息,以增强用户/组-项目交互,同时提出了一种路径感知注意力嵌入方法用于建模用户对不同类型路径的偏好。

FUA^[26]:一种基于用户行为的群推荐算法,该算

法设计了一个滑动时间窗, 考察子群划分时群内每个成员的近期活动, 将群划分为活跃子群和非活跃子群。

图 5 为本文算法和基线算法在 3 个 NDCG@N 值和 MAE 值上结果的对比。

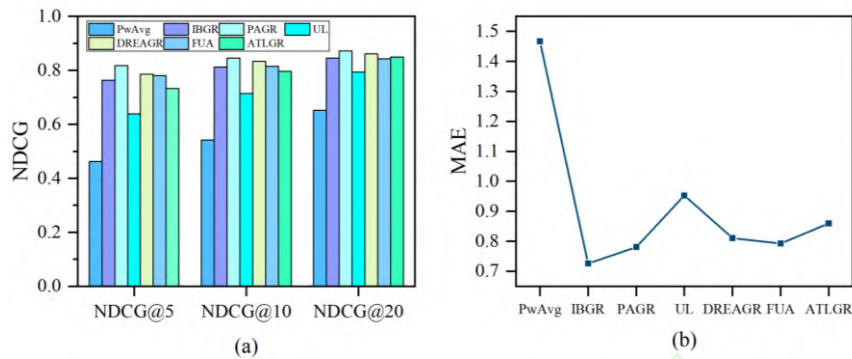


图 5 PAGR 与基线算法的三个 NDCG@N 值和 MAE 值对比

Fig. 5 Comparison of three NDCG@N and MAE for the PAGR with the baseline algorithmic

由图 5(a)可知, 本文算法在 N 取 5、10 和 20 时的 NDCG@N 值均优于基线算法, 根据实验结果计算, 对比基线算法中的最优结果, 本文算法在 NDCG@5、NDCG@10 和 NDCG@20 上分别提升了 4.04%、1.46%和 1.21%; 由图 5(b)可知, 本文算法的 MAE 值低于基线算法 PwAvg、UL、DREAGR、FUA 和 ATLGR。与基线 IBGR 相比, 本文算法的 MAE 值略高, 原因之一是 IBGR 算法设置的群组规模较小, 使得其进行偏好融合时的误差更小, 因此其 MAE 值低于本文算法。

为探究与 IBGR 相同的群组规模下, 二者间 MAE 的差异性, 此处选取群组划分后所有群组中群组规模接近 5, 15, 30, 50 的群组进行推荐, 所获结果与基线算法 IBGR 进行对比, 实验结果如图 6 所示。

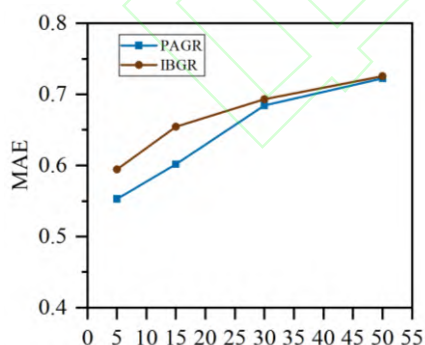


图 6 不同群组规模下 PAGR 与 IBGR 的 MAE 值对比

Fig. 6 Comparison of MAE values of PAGR and IBGR in different group sizes

由图 6 可知, 本文算法与 IBGR 算法 MAE 值均随群组规模增大而增加, 且二者间的差异在逐渐减小。相同群组规模下, 本文算法获得 MAE 值更优。

4.3.4 消融实验

为了验证在本文算法 (PAGR) 中各组件的有效性, 此处定义去除人格因子的 PAGR 算法为 PAGR-P, 去除项目吸引力的 PAGR 算法名为 PAGR-A, 将二者与完整的 PAGR 算法进行对比实验, 其中, 在 Personality 和 ML-latest-small 两个数据集上评估去除项目吸引力后算法的推荐效果, 在 Personality 和 PER_dataset 评估去除人格因子后算法的推荐效果。结果见表 9。

由表 9 可知, PAGR 的性能要优于 PAGR-P 和 PAGR-A, 引入用户人格特征和项目吸引力后, 算法在所选的评价指标上的结果均得到优化, 这验证了本文中各组件的有效性。此外, 本文在 Personality 数据集中同时验证了两种组件对于推荐结果的影响, 实验结果表明, 引入人格特征使得算法的 NDCG@5、NDCG@10 和 NDCG@20 分别提升了 1.69%、1.44%和 0.59%, 同时算法的 MAE 值降低了 8.60%。而引入项目吸引力使得算法的 NDCG@5、NDCG@10 和 NDCG@20 分别提升了 3.97%、2.60%和 1.97%, 算法的 MAE 值降低了 9.08%。项目吸引力组件带来性能提升要稍优于人格特征组件。因此在推荐过程中, 不仅要考虑用户个人因素, 也要关注项目本身对于用户的影响。

表 9 消融实验的结果

Tab. 9 The results of the ablation experiment

数据集	模型	NDCG@N			MAE
		5	10	20	
Personality	PAGR	0.8168	0.8454	0.8712	0.7805
	PAGR-P	0.8032	0.8334	0.8661	0.8539
	PAGR-A	0.7865	0.8240	0.8544	0.8584

PER_dataset	PAGR	0.6335	0.6654	0.7031	0.5710
	PAGR-P	0.6278	0.6576	0.6935	0.5728
ML-latest-small	PAGR	0.8691	0.8882	0.9050	0.6132
	PAGR-A	0.8602	0.8808	0.9000	0.8641

5 结 语

本文研究了用户人格特征和项目吸引力对群组推荐结果的影响,提出了一种融合用户人格和项目吸引力的群组推荐算法。利用基于用户人格和评分的混合相似度进行群组划分,同时在偏好融合时综合考虑了用户的专业度、信任度以及人格因子的影响;并结合群组偏好和项目吸引力两种因素获取最终推荐列表。通过实验验证了所提算法的有效性。接下来的工作是对根据用户历史行为学习用户人格特征进行深入研究,并挖掘用户和群组的更多的特征,继续优化群组推荐的结果。

参考文献:

- [1] 贾俊杰,姚叶旺,陈旺虎. 基于非负矩阵分解的群组推荐算法[J]. 计算机工程与科学, 2022, 44(5): 933-943.
- [2] JIA J, YAO Y, LEI Z, et al. Dynamic group recommendation algorithm based on member activity level[J]. Scientific programming, 2021, 2021: 1969118.
- [3] 张闯,王巍,杜雨珩,等. 融合上下文信息与自注意力机制的群组推荐算法[J]. 计算机应用研究, 2023, 40(2): 498-503.
- [4] NGUYEN T T, MAXWELL H F, TERVEEN L, et al. User personality and user satisfaction with recommender systems[J]. Information systems frontiers, 2018, 20: 1173-1189.
- [5] NING H, DHELMIM S, AUNG N. PersoNet: Friend recommendation system based on big-five personality traits and hybrid filtering[J]. IEEE Trans on computational social systems, 2019, 6(3): 394-402.
- [6] 杨梓涵. 基于人格特征的个性化学习内容推荐研究[D]. 长春: 吉林大学, 2023.
- [7] DHELMIM S, NING H, AUNG N, et al. Personality-aware product recommendation system based on user interests mining and metapath discovery[J]. IEEE Trans on computational social systems, 2020, 8(1): 86-98.
- [8] 李珊,姚叶慧,厉浩,等. 基于 ISA 联合聚类的组推荐算法研究[J]. 数据分析与知识发现, 2019, 3(8): 77-87.
- [9] 成静. 组推荐的偏好融合策略研究[D]. 桂林: 桂林电子科技大学, 2021.
- [10] 宋玉龙, 马文明, 刘彤彤. 融合用户信任度的概率矩阵分解群组推荐算法[J]. 计算机工程, 2022, 48(1): 105-111.
- [11] WANG H, CHEN D, ZHANG J. Group recommendation based on hybrid trust metric[J]. Automatika, 2020, 61(4): 694-703.
- [12] ZHANG J, CHENG J, ZHOU Y, et al. GRUIFI: a group recommendation model covering user importance and feature interaction[J]. Journal of internet technology, 2021, 22(5): 1143-1155.
- [13] XU Z, ZHENG X, ZHANG H, et al. A novel travel group recommendation model based on user trust and social influence[J]. Mobile information systems, 2021, 2021: 1-10.
- [14] KLEĆ M, WIECZORKOWSKA A, SZKLANNY K, et al. Beyond the big five personality traits for music recommendation systems[J]. Eurasip journal on audio, speech, and music processing, 2023, 2023(1): 4.
- [15] SUN J, GENG J, CHENG X, et al. Leveraging personality information to improve community recommendation in e-learning platforms[J]. British journal of educational technology, 2020, 51(5): 1711-1733.
- [16] DE BIASIO A, MERYLIN M, LUCA O, et al. On the problem of recommendation for sensitive users and influential items: simultaneously maintaining interest and diversity[J]. Knowledge-based systems, 2023, 275: 110699.
- [17] 史艳翠, 孟祥武, 张玉洁, 等. 一种上下文移动用户偏好自适应学习方法[J]. 软件学报, 2012, 23(10): 2533-2549.
- [18] HU R, PU P. Enhancing collaborative filtering systems with personality information[C]//ACM. Proceeding of the 5th ACM Conference on Recommender Systems. New York: ACM Digital Library, 2011: 197-204.
- [19] TKALCIC M, KUNAVAR M, TASIC J, et al. Personality based user similarity measure for a collaborative recommender system [C]// Proceeding of the 5th Workshop on Emotion in Human-Computer Interaction-Real world challenges. 2009: 30-37.
- [20] 刘畅. 个性化电影推荐算法研究[D]. 沈阳: 沈阳理工大学, 2023.
- [21] SEO Y D, KIM Y G, LEE E, et al. An enhanced aggregation method considering deviations for a group recommendation[J]. Expert systems with applications. 2018, 93: 299-312.
- [22] YALCIN E, BILGE A. A personality-based aggregation

- technique for group recommendation[J]. Eskişehir technical university journal of science and technology: applied sciences and engineering, 2020, 21(4): 486-498.
- [23] NOZARI R B, KOOHI H. A novel group recommender system based on members' influence and leader impact[J]. Knowledge-based systems, 2020, 205: 106296.
- [24] ABDRABBAH S B. A novel recommendation approach for groups based on aggregating top-k lists[J]. Procedia computer science, 2023, 225: 3067-3076.
- [25] ZHANG J, WANG S, CHEN Z, et al. Dependency relationships-enhanced attentive group recommendation in HINs[EB/OL]. [2024-12-01]. <https://doi.org/10.48550/arXiv.2311.11239>.
- [26] JIA J, WANG F, WANG H, LIU S. A fairness group recommendation algorithm based on user activity[J]. International journal of computational intelligence systems. 2024, 5;17(1):204.

