

融合信任因子的多样化电影推荐算法^①



王雨晨

(北京邮电大学 计算机学院 (国家示范性软件学院), 北京 100876)

通讯作者: 王雨晨, E-mail: wyuchenc@126.com

摘要: 传统的协同过滤算法过于依赖用户之间的评分, 容易出现冷启动和数据稀疏性问题, 同时推荐结果单一, 针对以上问题, 本文提出了一种融合信任因子的多样化电影推荐算法. 首先对用户相似度计算方法进行改进, 引入用户间信任度关系和属性特征信息. 接着使用聚类方法把具有相同兴趣的用户划分在同一社群. 最后在评分时综合考虑用户活跃度对电影的推荐度, 引入惩罚因子, 从而为目标用户提供个性化、多样化的电影推荐. 实验结果表明, 本文提出的算法在推荐精度和多样性指标上均有所提高, 有较好的推荐效果.

关键词: 协同过滤; 信任因子; 聚类; 用户活跃度; 多样化

引用格式: 王雨晨. 融合信任因子的多样化电影推荐算法. 计算机系统应用, 2021, 30(12): 187-193. <http://www.c-s-a.org.cn/1003-3254/8290.html>

Diversified Movie Recommendation Algorithm Based on Trust Factor

WANG Yu-Chen

(School of Computer Science (National Pilot Software Engineering School), Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: Traditional collaborative filtering algorithm relies much on ratings among users, which is prone to cold start and data sparsity. In addition, the recommendation results are single. To solve the above problems, this study proposes a diversified movie recommendation algorithm based on trust factor. Firstly, the calculation method of user similarity is improved, and the trust relationship and attribute characteristic information between users are introduced. Next, clustering is conducted to divide users with the same interest into the same community. Finally, user activity, as the movie recommendation degree, is taken into consideration comprehensively in the rating. The penalty factor is introduced, so as to facilitate personalized and diversified movie recommendations for target users. Experimental results show that the proposed algorithm can improve the recommendation accuracy and diversity, achieving a good recommendation effect.

Key words: collaborative filtering; trust factor; clustering; user activity; diversity

大数据时代下, 互联网为我们提供方便的同时, 也给我们带来了困扰, 网络中的信息呈爆炸式增长, 从海量的信息中挖掘出对我们需要的信息非常重要^[1]. 随着网络上电影数量的增加, 用户不得不花费大量精力来挑选自己感兴趣的电影. 此时, 个性化推荐系统的出现在一定程度上解决了信息过载问题^[2], 它可以通过分析用户的历史行为特征和有关的数据信息, 为用户推荐

可能感兴趣的项目, 起到了信息筛选的作用.

现有的个性化电影推荐算法主要有协同过滤算法、基于内容的推荐算法和混合推荐算法等, 其中, 传统的协同过滤算法 (collaborative filtering) 应用最为广泛^[3]. 它具有算法复杂性较低、对原始数据的数据内容要求较低的优势. 该算法通过分析用户对于电影的历史评分, 进而找到相似用户, 并根据相似用户的电影评

① 收稿时间: 2021-03-09; 修改时间: 2021-04-09; 采用时间: 2021-05-14

价记录为目标用户推荐电影。

但是协同过滤算法在实际应用中也存在很多缺陷,比如冷启动、数据稀疏性问题。基于此,国内外学者做了大量实验研究。Lokhande等^[4]提出了一种融合层次聚类和主成分分析的混合推荐算法模型,降低了用户-评分矩阵的稀疏性,提高了传统协同过滤算法的推荐精度。杨秀梅等^[5]根据已有用户对不同新闻的浏览历史记录提取其上下文信息,解决了新闻推荐系统中的新用户冷启动问题,提高了用户满意度。但是以上研究基本上是依据用户的评分数据进行分析的,并没有考虑用户间的信任度关系,对于电影推荐系统中的用户来说,其信任朋友的推荐的影片可能会更值得信赖,推荐效果更能使用户感兴趣。基于此,俞美华^[6]提出一种融合用户兴趣度的电影推荐算法,通过改进用户相似度的计算方法,提供个性化推荐。Sasmita等^[7]提出了用户相似性和加权信任传播相结合的信任矩阵度量方法,通过使用反向传播、深度神经网络来进行电影推荐。但是上述算法没有充分挖掘用户之间的隐式信任关系,没有建立相对完整的信任机制,会出现信任关系不完整的问题。

另一方面,目前电影推荐算法大多数都是基于用户之间的相似度或者物品之间的相似度,主要关注推荐的准确率,推荐热门物品给用户,而忽略了非热门物品的推荐机会,从整体推荐结果上来看,推荐种类多样性较低、用户容易产生审美疲劳。针对用户的个性化需求,为用户推荐他们感兴趣但难以发现的商品,即将长尾商品准确地推荐给需要它的用户是个亟待解决的问题^[8-10]。Zhou等^[11]使用传导的方法将热门项目的权重传递给长尾项目,提高长尾项目的推荐度,提高了系统的推荐多样性,但是系统准确率较低。Adamopoulos等^[12]针对物品的概率分布计算相似度,提高了系统的多样性和准确度,但是作者只分析了显性评分。因此,在保证准确率的前提下提升推荐算法的多样性具有一定的研究意义。

综上所述,本文对传统的协同过滤算法进行改进,针对于推荐准确性和多样性,提出了一种融合信任因子的多样化电影推荐算法。本算法首先挖掘用户的属性特征和用户之间的信任度关系,将其和传统的相似度计算方法结合,保证了每位用户推荐的可靠度和可信度。然后使用改进类簇中心确认方法的K-means聚类方法,将拥有相同兴趣爱好的用户聚集在同一个社

群中。最后在用户推荐阶段使用重排序方法,通过加入惩罚因子来计算电影推荐度,提高了电影推荐算法的整体多样性,同时能够在一定程度上解决冷启动和数据稀疏性问题。

1 融合信任因子的相似度算法

信任关系是指用户对于邻居用户的推荐可靠性的认可程度,认可度越高,推荐越符合用户兴趣^[13]。而信任关系受到诸多因素的影响,比如用户自身属性、社会信任度关系等,这些因素通常被统称为信任因子^[14]。本文将信任因子与传统相似度计算方法相融合,旨在充分挖掘用户信息特征。

1.1 社会信任度关系

1.1.1 传统直接信任度模型

传统的信任关系模型大多数是通过简单计算用户间的交互次数来建立的^[15,16]。文献^[17]提出了一种计算用户直接信任度的方法,通过最初信任度与用户间成功交互次数的乘积计算得到用户直接信任度,如式(1)所示。其中 $Init(u, v)$ 为初始信任度,计算方法如式(2)所示, $I_u \cap I_v$ 为用户 u 和用户 v 交互的次数, d 为设置的阈值,表示两个用户成功建立信任关系最少的交互次数,文献实验证明, d 取值为80效果最佳。 $count_t$ 表示两用户交互的总次数, $count_s$ 表示用户交互成功的次数,此处交互成功的标准为两用户对于同一项目的评分小于2分。

$$T(u, v) = Init(u, v) \frac{count_s}{count_t} \quad (1)$$

$$Init(u, v) = \frac{\min(I_u \cap I_v, d)}{d} \quad (2)$$

1.1.2 改进的直接信任度模型

传统的信任关系模型能够有效地表示用户的信任关系,但是式(1)中成功交互的计算方法忽略了用户之间对于非感兴趣电影的信任度。如果两个用户对于某部电影的评分都为1.5分,根据传统信任度计算方法,由于两个用户都看过同一电影并且对于电影的评分相同,系统会认定两用户间有较高的相似度。但是这个分数只能说明两个用户对于这部电影不喜爱,并不能代表用户间喜爱的项目相似。因此,针对此问题,本文在传统信任度模型中引入感兴趣因子,定义如下:

$$InterestT(u, v) = \frac{|I_{uv} > D|}{count_t} \quad (3)$$

其中, $|I_{uv} > D|$ 表示用户 u 和用户 v 共同评分的电影分数大于 D 的数量, $count_i$ 表示两个用户交互的总次数. 如果用户 u 和用户 v 对于同一部电影的评分都大于阈值 D , 则说明两用户的兴趣比较相似. 基于电影的评分分值区间为 1-5, 因此设定阈值 D 为 3.5.

综合得到用户直接信任度模型, 定义如下:

$$Trust_1(u, v) = T(u, v) \times InterestT(u, v) \quad (4)$$

1.1.3 间接信任度模型

由于用户-评分矩阵本身比较稀疏, 可能会出现用户 u 和用户 v 没有直接信任关系, 但是他们有共同朋友 k , 根据信任关系的传递性特征, 用户能够信任自己的直接朋友, 也会在一定程度上信任朋友的朋友. 故本文引入间接信任度, 用 $ITrust(u, v)$ 表示. 但是随着信任链的长度增加, 信任度会逐渐减弱. 在保证时间复杂度和保留有效用户的前提下, 根据文献 [18] 对于间接信任的关系研究, 本文只考虑目标用户的两级信任关系.

假设用户 v 是用户 u 的二级信任用户, 两用户之间至少存在一条路径, 每条路径表示为 $Path(u_i, v_i) = (u_i, k, v_i)$, 每条路径的信任度计算方法如下:

$$ITrust_i(u_i, v_i) = Trust_1(u_i, k) \times Trust_1(k, v_i) \quad (5)$$

用户的间接信任模型通过计算路径信任度的均值得到, 如式 (6), 其中 n 表示从用户 u 到用户 v 的路径数目之和.

$$ITrust(u, v) = \frac{\sum_{i \in n} ITrust_i(u_i, v_i)}{n} \quad (6)$$

得到用户间接信任度矩阵后, 将其填充到直接信任度矩阵 $Trust_1$ 中, 构建社会信任度关系矩阵 $Trust(u, v)$.

1.2 用户属性信息

充分挖掘用户的属性信息是获得用户特征的重要方法, 通过数据挖掘, 我们可以对用户进行更全面而深入的了解, 从而能够更加准确地预测、并推荐适合的电影给用户. 较早尝试将属性信息与推荐算法融合的典型代表有 Kazienko 等^[19] 开发的电商推荐系统, 他通过挖掘用户注册的信息进行推荐, 对于新用户来说, 利用属性特征的推荐算法的点击率达到了 89%, 相比于随机推荐提高了 62%. 属性信息从某种程度上反映了用户的喜好度, 因此本文对用户的年龄、性别、职业信息进行挖掘.

本文参考文献 [20] 中的计算方法, 对于用户年龄, 将其划分为 7 个年龄段, 分别用数值 1-7 表示 (年龄段

为: 小于 18、18~24、25~34、35~44、45~49、50~55、大于 56). 对于用户性别, 用 1 表示男 (M), 0 表示女 (F). 对于用户职业, 将其分为七大类, 分别用数值 1-7 表示 (包含企事业单位负责人、专业技术人员、服务业人员、文艺从事人员、教育行业、家政以及其他 7 类). 组合的属性信息相似度如式 (7), 其中 $S(u, v)$ 代表性别相似度, $A(u, v)$ 代表年龄相似度, $O(u, v)$ 代表职业相似度.

$$sim_f(u, v) = 0.6S(u, v) + 0.1A(u, v) + 0.3O(u, v) \quad (7)$$

1.3 融合信任因子模型

融合信任因子模型综合考虑了信任关系的潜在因素和用户个体差异, 主要包含社会信任度关系和属性特征相似度. 用户 u 和用户 v 的信任因子模型如式 (8), 其中, α 为权重因子, $\alpha \in [0, 1]$.

$$sim_t(u, v) = \alpha Trust(u, v) + (1 - \alpha) sim_f(u, v) \quad (8)$$

1.4 改进的相似度计算

用户之间的相似度可以使用余弦相似度来计算, 但是不同用户评分可能会有不同的评分标准, 有的用户习惯性打高分, 而有的用户习惯性打低分, 考虑到不同用户的评分偏好问题, 本文使用修正的余弦相似度作为传统相似度计算方法. 其中 \bar{r}_u, \bar{r}_v 表示用户 u, v 对于所有评分电影的均值, r_{uv} 表示两个用户共同的评分集合, $r_{u,i}, r_{v,i}$ 分别表示用户 u, v 对于电影 i 的评分, 修正的余弦相似度计算如下:

$$sim_c(u, v) = \frac{\sum_{i \in r_{uv}} (r_{u,i} - \bar{r}_u) \times (r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in r_u} (r_{u,i} - \bar{r}_u)^2} \times \sqrt{\sum_{i \in r_v} (r_{v,i} - \bar{r}_v)^2}} \quad (9)$$

结合本节所述, 将信任因子和修正的余弦相似度融合得到用户综合相似度 $sim(u, v)$, β 为调整综合相似度的系数, $\beta \in (0, 1)$, 用户相似度计算如下:

$$sim(u, v) = \beta sim_t(u, v) + (1 - \beta) sim_c(u, v) \quad (10)$$

2 重排序多样化推荐

2.1 聚类

K-means 算法是基于划分的无监督聚类算法^[21,22], 用途广泛且收敛速度快. 通过 K-means 聚类算法对用户进行划分, 能够将用户分为 k 个用户社群, 社群内用户的相似度较高, 社群间用户的相似度较低. 但是 K-means 算法容易受到初始聚类中心的影响^[23], 对初始聚类中心具有较高的敏感性, 而且合适的初始聚类中

心的选择能够减少聚类迭代的次数, 加快收敛速度, 因此, 选择合适的初始聚类中心至关重要. 本文对 K-means 进行一些改进, 在选取初始类簇中心时保证其各中心之间的距离尽可能大, 然后进行聚类并不断更新类簇中心, 用户-评分矩阵聚类的步骤如算法 1.

算法 1. 用户-评分矩阵聚类算法

输入: 用户-评分矩阵 $R_{m \times n}$, 聚类的类簇个数 k
 输出: k 个用户社群 $C = \{C_1, C_2, \dots, C_k\}$

- 1) 随意选取一个用户作为初始聚类的中心 C_1 .
- 2) 首先计算每个用户与当前存在的类簇中心之间最小相似度, 用 $D(x)$ 表示, 则每个用户被选为下一个聚类中心的概率为 $D(x)^2 / \sum D(x)^2$, 按照轮盘概率法选择下一个聚类中心.
- 3) 重复 2), 直到选出 k 个聚类中心 $C' = \{C_1', C_2', \dots, C_k'\}$.
- 4) 对于矩阵中每个用户 u , 计算它到 k 个聚类中心的相似度并将其划分到最近的聚类中心所对应的簇中.
- 5) 针对每个类簇, 重新计算它的聚类中心 C_i .
- 6) 重复步骤 4) 和步骤 5), 直到聚类中心的位置不再变化.

2.2 多样化推荐

用户的推荐列表是根据相似用户的电影评分和相似度计算得到的预测评分来推荐的. 以得到用户 u 的推荐列表为例, 首先需要计算目标用户 u 与所有聚类中心的距离, 找到距离最近的类簇. 然后计算用户 u 与该类簇中所有用户的相似度, 选取最近的 k 个邻居, 记作 U_u , 则用户 u 对于电影 m 的预测评分计算如下:

$$R_{u,i} = \bar{R}_u + \frac{\sum_{v \in U_u} sim(u,v) \times (R_{v,i} - \bar{R}_v)}{\sum_{v \in U_u} sim(u,v)} \quad (11)$$

其中, \bar{R}_u, \bar{R}_v 分别表示用户 u, v 的电影评分均分, $R_{v,i}$ 表示用户 v 对电影 i 的评分.

但是在计算预测评分时, 推荐算法的多样性容易受到用户活跃度的影响^[24], 如果用户比较活跃, 那么他的感兴趣范围更广, 更容易和其他用户有共同的评价电影, 但是这不能表示活跃用户的兴趣特征, 推荐的电影参考价值不大, 因此本文使用重排序技术, 通过引入惩罚因子, 使非活跃用户的推荐权重增大, 活跃用户的推荐权重减小. 重排序技术不会改变电影推荐的候选集, 只是在排序过程中改变电影的推荐概率, 把一些新物品移到推荐列表前面. 本文引入惩罚因子修正电影的推荐度, 改进后的电影预测评分计算如下:

$$R_{u,i} = \bar{R}_u + \frac{\sum_{v \in U_u} sim(u,v) \times (R_{v,i} - \bar{R}_v)}{\ln(1 + |N(v)|) \times \sum_{v \in U_u} sim(u,v)} \quad (12)$$

其中, $|N(v)|$ 表示邻居 v 的评分电影数. 用户越活跃, 其评价的电影数量就越多, 那么惩罚因子就越大, 该用户的推荐度参考性就相对更低一些.

2.3 推荐算法模型

结合上文所述, 融合信任因子的多样化电影推荐算法整体的流程如图 1 所示.

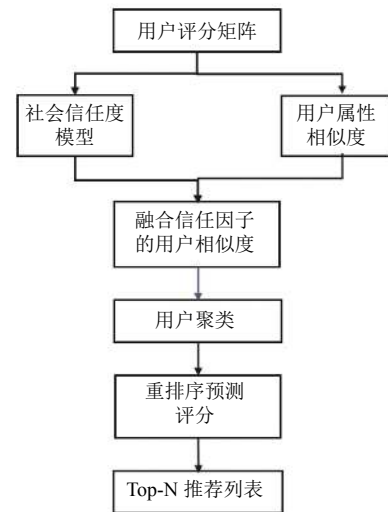


图 1 推荐算法模型

3 实验结果与分析

3.1 实验数据集及评价指标

本实验使用美国明尼苏达大学 GroupLens 实验组创建的开源电影数据集 Movielens-100K. 此数据集包含了 943 名用户对 1682 部电影的 100 000 条评分记录, 每位用户至少评分 20 部电影, 评分值区间为 1~5, 评分越高表明用户对于电影的喜爱度越高. 同时数据集中还包含用户信息表、职业列表等, 能够对用户数据进行挖掘. 为避免偶然性, 实验采用五折交叉验证法, 将数据均分为 5 份, 其中 1 份作为测试数据集, 另外 4 份作为训练数据集. 重复进行实验 5 次, 5 次实验结果的平均值为实验最终结果.

实验软件环境为 Windows10 x64、PyCharm x64, 硬件环境为 Intel(R) Core(TM) I5-10210U CPU @ 1.60 GHz 8 核.

本文采用平均绝对误差 MAE (Mean Absolute Error)^[25], 整体多样性^[26] 作为算法度量标准. MAE 能够衡量预测评分和真实评分的差距, 其值越小, 误差越小, 推荐准确度就越高. 其定义为:

$$MAE = \frac{\sum_{i \in test} |q_i - p_i|}{n} \quad (13)$$

其中, p_i 表示目标用户 u 对于电影 i 的预测评分, q_i 表示目标用户 u 对电影 i 的真实评分. n 表示推荐的电影数量.

多样性指标^[25] 显示了推荐列表中电影的多种类、非相似性, 能够覆盖到用户不同的兴趣和爱好, 推荐系统中用户 u 的推荐列表的多样性定义如下:

$$iDiversity(R(u)) = 1 - \frac{2 \times \sum_{i, j \in R(u), i \neq j} sim(i, j)}{|R(u)|(|R(u)| - 1)} \quad (14)$$

其中, $|R(u)|$ 表示用户 u 的推荐列表数目, $sim(i, j)$ 表示推荐列表中的电影 i 和 j 之间的相似度.

用户的个体推荐结果多样性越高, 系统整体的多样性越高, $|U|$ 表示用户个数, 系统整体多样性定义如下:

$$Diversity = \frac{1}{|U|} \sum_{u \in U} iDiversity(R(u)) \quad (15)$$

3.2 实验结果分析与对比

实验 1. 聚类个数 k , 相似度计算方法的选择.

为了达到更好的推荐效果, 使用聚类算法时要确定一个合适的类簇数目 k . 此处先采用传统的协同过滤方法进行实验, 相似度计算方法使用余弦相似度和修正的余弦相似度方法进行对比实验, 设置类簇的取值范围为 $[5, 50]$, 增量为 5, 然后计算 MAE 值, 实验结果如图 2.

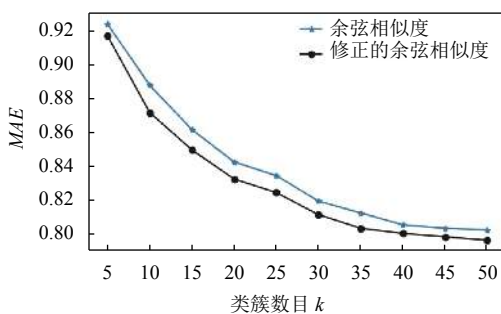


图 2 聚类个数对应的 MAE 值

根据实验可得出结论, 修正后的余弦相似度计算方法比余弦相似度方法得到的结果整体误差要小一些, 随着类簇数量的增多, 两种方法的 MAE 值减小的速度逐渐变慢趋于收敛, 故本文采用修正的余弦相似度方法作为相似度的计算方法.

在基于修正的余弦相似度方法中, 当类簇数目达

到 35 之后, MAE 值趋于平稳, 考虑到算法的计算效率, 因此选取类簇数目 k 为 35 进行实验.

实验 2. 信任因子模型权重.

将用户的社会信任度和属性特征相似度相融合得到信任因子模型 $sim_t(u, v)$, 设置权重系数 $\alpha \in (0, 1)$, 增量为 0.1, 计算不同权重下的 MAE 值, 结果如图 3.

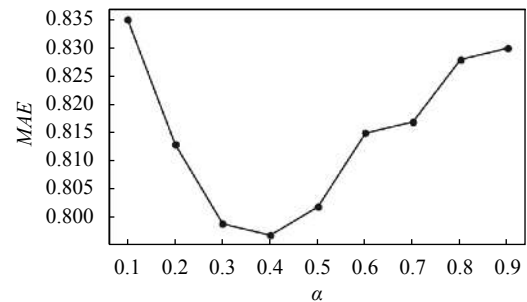


图 3 权重系数对应的 MAE 值

根据实验可知, 当 α 为 0.4 时, MAE 值较小, 因此取 0.4 为权重系数.

实验 3. 用户相似度参数选择.

将用户的信任因子模型和修正的余弦相似度相融合得到用户相似度 $sim(u, v)$, 设置参数 $\beta \in (0, 1)$, 增量为 0.1, 计算不同参数下的 MAE 值, 结果如图 4.

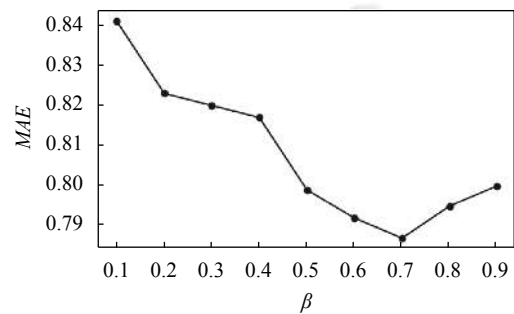


图 4 相似度参数对应的 MAE 值

根据实验可知, 当 β 为 0.7 时, MAE 值较小, 因此取 0.7 为权重系数.

实验 4. 本文算法与其他算法性能对比.

本实验选取基于皮尔逊相似性的协同过滤算法 (PB-UCF), 文献 [17] 提出的融入用户喜好度到信任关系中的协同过滤算法 (WTUBCF), 文献 [27] 提出的 topN 推荐算法 Expectation approach, 文献 [28] 提出的 DNMF 算法和本文提出的算法做对比, 5 种算法的最近邻居用户区间取值为 $[5, 50]$, 增量为 5, 比较 MAE 和多样性指标.

从图5中可以看出,本文提出的算法在最近邻居数达到35能够取得较好的推荐效果.4种算法随着最近邻居数目的增多MAE值逐渐下降,本文算法的MAE值明显低于其他4种算法,说明本文提出的算法能够提高算法推荐的准确度.

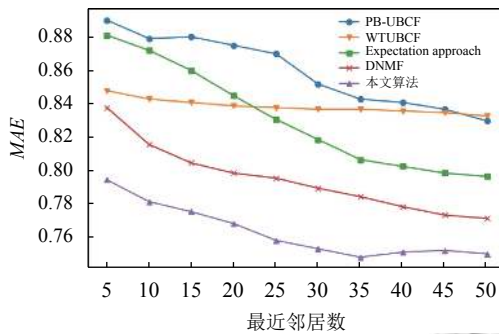


图5 不同算法的MAE值对比

从图6中可以看出,随着推荐物品的数量增多,系统的多样性逐步增加. PB-UCF和WTUBCF两种算法由于没有对多样性问题进行改进,所以算法多样性相对较低.和Expectation approach相比,随着最近邻居数的增加,本文提出的算法在多样性指标中相对较高些.在最近邻居数小于40时, DNMF算法的多样性较其他算法高一些,在邻居数大于40后,本文提出的算法的多样性相对较高,可见随着推荐电影的数目的增多,本文提出的算法能够更好地分析用户属性特征并为其提供个性化的推荐.因此本文提出的算法在保证准确度的情况下,能够在一定程度上提升电影推荐系统的多样性.

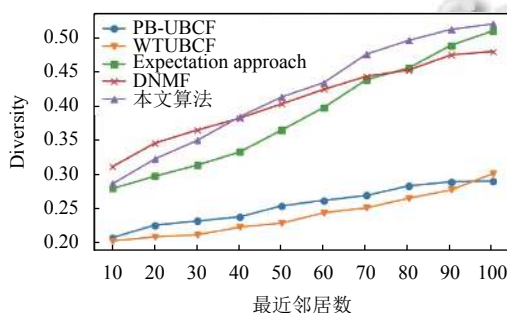


图6 不同算法的多样性对比

4 结论与展望

本文提出的融合信任因子的多样化电影推荐算法,通过分析社会性信任度关系和用户特征信息,充分挖

掘了用户的特征信息.另外通过引入惩罚因子,结合用户活跃度实现了电影多样化推荐.通过实验分析,证明了本文提出的算法有效,相对于之前的电影推荐算法来说,算法平均绝对误差更小,同时推荐算法的多样性有了相对地提升.另外,本文提出的算法具有一定的通用性,除了可以用于电影推荐系统之外,还能应用于电商推荐系统、新闻推荐系统等.

但是对于实际推荐系统,隐式反馈对于推荐算法也起着至关重要的作用.在接下来的工作中将会对隐式反馈进行研究,使推荐算法更迎合用户的偏好.

参考文献

- Edmunds A, Morris A. The problem of information overload in business organisations: A review of the literature. *International Journal of Information Management*, 2000, 20(1): 17–28. [doi: 10.1016/S0268-4012(99)00051-1]
- Verbert K, Manouselis N, Ochoa X, et al. Context-aware recommender systems for learning: A survey and future challenges. *IEEE Transactions on Learning Technologies*, 2012, 5(4): 318–335. [doi: 10.1109/TLT.2012.11]
- Ben Schafer J, Konstan JA, Riedl J. E-commerce recommendation applications. *Data Mining and Knowledge Discovery*, 2001, 5(1–2): 115–153. [doi: 10.1023/A:1009804230409]
- Lokhande A, Jain P. Hybrid collaborative filtering model using hierarchical clustering and PCA. *Proceedings of Recent Advances in Interdisciplinary Trends in Engineering & Applications (RAITEA)*. Indore, 2019. [doi: 10.2139/ssrn.3365525]
- 杨秀梅, 孙咏, 王美吉, 等. 新闻推荐系统中用户冷启动问题的研究. *小型微型计算机系统*, 2016, 37(3): 479–482. [doi: 10.3969/j.issn.1000-1220.2016.03.016]
- 俞美华. 融合用户兴趣度与项目相关度的电影推荐算法研究. *电脑知识与技术*, 2017, 13(8): 22–26. [doi: 10.14004/j.cnki.ckt.2017.0811]
- Choudhury SS, Mohanty SN, Jagadev AK. Multimodal trust based recommender system with machine learning approaches for movie recommendation. *International Journal of Information Technology*, 2021, 13(2): 475–482. [doi: 10.1007/s41870-020-00553-2]
- Adomavicius G, Kwon Y. Improving aggregate recommendation diversity using ranking-based techniques. *IEEE Transactions on Knowledge and Data Engineering*, 2012, 24(5): 896–911. [doi: 10.1109/TKDE.2011.15]
- Zhou T, Kuscsik Z, Liu JG, et al. Solving the apparent

- diversity-accuracy dilemma of recommender systems. Proceedings of the National Academy of Sciences of the United States of America, 2010, 107(10): 4511–4515. [doi: [10.1073/pnas.1000488107](https://doi.org/10.1073/pnas.1000488107)]
- 10 邓明通, 刘学军, 李斌. 基于用户偏好和动态兴趣的多样性推荐方法. 小型微型计算机系统, 2018, 39(9): 2029–2034. [doi: [10.3969/j.issn.1000-1220.2018.09.025](https://doi.org/10.3969/j.issn.1000-1220.2018.09.025)]
- 11 Zhou T, Su RQ, Liu RR, *et al.* Accurate and diverse recommendations via eliminating redundant correlations. New Journal of Physics, 2009, 11(12): 123008. [doi: [10.1088/1367-2630/11/12/123008](https://doi.org/10.1088/1367-2630/11/12/123008)]
- 12 Adamopoulos P, Tuzhilin A. On over-specialization and concentration bias of recommendations: Probabilistic neighborhood selection in collaborative filtering systems. Proceedings of the 8th ACM Conference on Recommender Systems. Foster City: ACM, 2014. 153–160. [doi: [10.1145/2645710.2645752](https://doi.org/10.1145/2645710.2645752)]
- 13 郝立燕, 王靖. 基于填充和相似性信任因子的协同过滤推荐算法. 计算机应用, 2013, 33(3): 834–837. [doi: [10.3724/SP.J.1087.2013.00834](https://doi.org/10.3724/SP.J.1087.2013.00834)]
- 14 Moradi P, Ahmadian S, Akhlaghian F. An effective trust-based recommendation method using a novel graph clustering algorithm. Physica A: Statistical Mechanics and its Applications, 2015, 436: 462–481. [doi: [10.1016/j.physa.2015.05.008](https://doi.org/10.1016/j.physa.2015.05.008)]
- 15 Yang ZH, Xu L, Cai ZM, *et al.* Re-scale AdaBoost for attack detection in collaborative filtering recommender systems. Knowledge-Based Systems, 2016, 100: 74–88. [doi: [10.1016/j.knsys.2016.02.008](https://doi.org/10.1016/j.knsys.2016.02.008)]
- 16 Gan MX, Sun L, Jiang R. Trinity: Walking on a user-object-tag heterogeneous network for personalised recommendations. Journal of Computer Science and Technology, 2016, 31(3): 577–594. [doi: [10.1007/s11390-016-1648-0](https://doi.org/10.1007/s11390-016-1648-0)]
- 17 刘智捷, 徐小良, 王宇翔. 基于融合信任关系的协同过滤推荐算法. 杭州电子科技大学学报(自然科学版), 2018, 38(3): 44–48. [doi: [10.13954/j.cnki.hdu.2018.03.009](https://doi.org/10.13954/j.cnki.hdu.2018.03.009)]
- 18 赵海燕, 侯景德, 陈庆奎. 结合时间权重与信任关系的协同过滤推荐算法. 计算机应用研究, 2015, 32(12): 3565–3568. [doi: [10.3969/j.issn.1001-3695.2015.12.008](https://doi.org/10.3969/j.issn.1001-3695.2015.12.008)]
- 19 Kazienko P, Kołodziejcki P. WindOwls-adaptive system for the integration of recommendation methods in E-commerce. Proceedings of the 3rd International Atlantic Web Intelligence Conference. Lodz: Springer, 2005. 218–224.
- 20 韩立锋, 陈莉. 融合用户属性与项目流行度的用户冷启动推荐模型. 计算机科学, 2021, 48(2): 114–120. [doi: [10.11896/jsjx.200900152](https://doi.org/10.11896/jsjx.200900152)]
- 21 Feldman D, Schmidt M, Sohler C. Turning big data into tiny data: Constant-size coresets for k -Means, PCA and projective clustering. SIAM Journal on Computing, 2020, 49(3): 601–57.
- 22 Huang SD, Ren YZ, Xu ZL. Robust multi-view data clustering with multi-view capped-norm K -means. Neurocomputing, 2018, 311: 197–208. [doi: [10.1016/j.neucom.2018.05.072](https://doi.org/10.1016/j.neucom.2018.05.072)]
- 23 Rezaei K, Rezaei H. HFSSMOOK-means: An improved K -means algorithm using hesitant fuzzy sets and multi-objective optimization. Arabian Journal for Science and Engineering, 2020, 45(8): 6241–6257. [doi: [10.1007/s13369-020-04620-5](https://doi.org/10.1007/s13369-020-04620-5)]
- 24 师扬波. 面向多样性的推荐算法研究 [硕士学位论文]. 成都: 电子科技大学, 2018.
- 25 朱郁薇, 吕琳媛. 推荐系统评价指标综述. 电子科技大学学报, 2012, 41(2): 163–175. [doi: [10.3969/j.issn.1001-0548.2012.02.001](https://doi.org/10.3969/j.issn.1001-0548.2012.02.001)]
- 26 Ziegler CN, McNee SM, Konstan JA, *et al.* Improving recommendation lists through topic diversification. Proceedings of the 14th International Conference on World Wide Web. Lyon: ACM, 2005. 22–32.
- 27 刘慧婷, 岳可诚. 可提高多样性的基于推荐期望的 top-N 推荐方法. 计算机科学, 2014, 41(7): 270–274. [doi: [10.11896/j.issn.1002-137X.2014.07.056](https://doi.org/10.11896/j.issn.1002-137X.2014.07.056)]
- 28 Zhao P, Peng FR, Cui ZH, *et al.* A new matrix factorization model for novel and diverse recommendation. Journal of Shanxi University (Natural Science Edition), 2020, 43(4): 746–755. [doi: [10.13451/j.sxu.ns.2020061](https://doi.org/10.13451/j.sxu.ns.2020061)]