

项目多属性模糊联合的多样性视频推荐算法

张艳红¹ 张春光² 周湘贞³ 王怡鸥⁴

(广东技术师范大学天河学院计算机科学与工程学院 广州 510540)¹

(北京科技大学计算机与通信工程学院 北京 100083)² (北京航空航天大学计算机学院 北京 100191)³

(北京市科学技术情报研究所 北京 100044)⁴

摘要 针对视频协同过滤推荐算法多样性较低的问题,提出了一种基于多属性联合的多样性视频协同过滤推荐算法。根据用户与推荐系统的互动历史记录,判断用户是否满意系统的推荐项目,如果某个用户过去观看同一个主题的视频节目,并且不关心视频的作者,那么认为该用户对视频作者表现出较高的多样性,对视频节目主题表现出的多样性较低。采用信息熵与用户配置信息长度两个指标来评估项目各个属性的多样性,根据两个指标的组合将用户对每个项目属性的多样性分为 4 个象限,并且对用户多样性进行模糊化处理,以获得用户多样性对于 4 个象限的隶属度。在第一个阶段预测未评分项目的评分;在第二个阶段将所有项目重新排序,以提高推荐列表的多样性。最终,基于公开的 Movielens 1M 数据集进行了对比实验,实验结果证明本算法可实现接近 top-N 算法的准确率性能,同时具有一定的多样性增强效果。在推荐准确率与多样性平衡的应用场景下,设置合适的参数能够在损失较少推荐准确率的前提下,显著提高个体多样性、总体多样性与新颖性。

关键词 电子商务,视频推荐系统,多样性增强,协同过滤推荐算法,重新排序算法,长尾分布

中图分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2019.08.012

Diverse Video Recommender Algorithm Based on Multi-property Fuzzy Aggregate of Items

ZHANG Yan-hong¹ ZHANG Chun-guang² ZHOU Xiang-zhen³ WANG Yi-ou⁴

(School of Computer Science and Engineering, Tianhe College of Guangdong Polytechnic Normal University, Guangzhou 510540, China)¹

(School of Computer & Communication Engineering, University of Science Technology, Beijing 100083, China)²

(School of Computer Science and Engineering, Beihang University, Beijing, 100191, China)³

(Beijing Institute of Science and Technology Information, Beijing 100044, China)⁴

Abstract In order to improve the diversity of the collaborative filtering recommender system of videos, this paper proposed a diverse videos collaborative filtering recommender algorithm based on multi-property aggregate. According to the history of interaction between users and recommendation system, users are judged whether they are satisfied with the recommendation items of the system. If a user watches the videos on the same topic produced by different video authors, it indicates that this user shows high diversity to the video authors, and low diversity to the video subjects. Information entropy and user profile length are used to evaluate the diversity of each item's attributes. According to the combination of the two indicators, the user's diversity of each item's attributes is divided into four quadrants, and the user's diversity is fuzzified to obtain the membership degree of user's diversity to the four quadrants. In the first phase, it predicts the rates of unrated items. In the second phase, it re-ranks all items, which improves the diversity of recommendation list. At last, experimental results based on the public Movielens 1M dataset show that, the proposed algorithm can realize the similar accuracy with top-N algorithm, at the same time, it enhances the diversity effectively. In the application scenario of balancing recommendation accuracy and diversity, setting appreciate parameters can improve the individual diversity, total diversity and freshness significantly with acceptable recommendation accuracy reduction.

Keywords Electronic commerce, Video recommender system, Diversity enhancement, Collaborative filtering recommender algorithm, Re-ranking algorithm, Long tail distribution

到稿日期:2019-03-04 返修日期:2019-05-16 本文受国家自然科学基金面上项目(61672077),广东省教育厅教育科学规划教育信息技术研究项目(14JXN060),广东省教育厅项目(2017SZ03)资助。

张艳红(1978—),女,硕士,讲师,主要研究方向为大数据分析、推荐算法,E-mail:33582210@qq.com(通信作者);张春光(1976—),男,博士,讲师,主要研究方向为人工智能、推荐算法、计算机网络;周湘贞(1976—),女,博士,副教授,主要研究方向为图像处理、推荐算法;王怡鸥(1990—),女,博士,助理研究员,主要研究方向为情报研究、智能信息处理。

1 引言

随着互联网内容的持续丰富,各类资源网站包括视频资源网站与 IPTV 领域中的自动推荐系统能够有效地帮助用户高效地获得需要的资源。借助推荐系统获取喜爱的视频资源或电视节目,已经成为各大视频资源网站、智能电视网站的必备模块^[1]。目前大多数的视频与电视节目均将推荐系统的推荐准确率作为评估推荐效果的指标^[2],然而,近期许多研究发现,音乐应用、视频网站、电影应用等领域均存在长尾分布的现象^[3-5]。长尾理论认为提高推荐列表的多样性对项目的销售与用户满意度均有利^[6],因此,提高视频类资源推荐系统的多样性成为新的研究热点^[7]。

许多研究人员对音乐推荐、电子商务、电影推荐等应用设计了提高多样性的推荐方案,其中推荐重新排序方案^[8-9]是较为重要的一种,该方案可有效地增强推荐的多样性,同时对推荐准确率的影响极小。目前 top-N 算法是主流的推荐算法,该算法的推荐准确率较高,重新排序算法^[8-9]则为 top-N 算法引入了排名阈值,能够有效地提高 top-N 算法的多样性。此外,文献^[10]认为重新排名算法并未从全局角度选择推荐列表,并提出了基于推荐期望的协同过滤算法,该算法获得了更高的总体多样性和更为均衡的推荐组合。文献^[11]定义了项目类别贡献函数以改进预测评分公式,提高了与目标项目类别不完全相同的项目得分,该算法增强了推荐系统的个体多样性,具有更高的推荐质量。上述算法均针对一般的应用提出了一般性的推荐算法,而视频网站与电视节目具有许多特殊的特征,因此,本文设计了增强多样性的视频推荐算法。

2 研究背景

目前,主流的大型电影评分网站、音乐评分网站、视频网站等均将推荐准确率作为唯一指标。图 1(a)给出豆瓣电影网站为“喜剧之王”推荐的 5 部电影,5 部电影均为“周星驰”主演的喜剧电影,该推荐系统的总体多样性、个体多样性与项目类别覆盖率均较低。图 1(b)是另一种推荐系统对于“喜剧之王”的推荐结果,可以看出提高推荐列表的多样性十分重要,能够有效地提高用户的满意度,这也符合长尾理论的特点。

为了合理地提高推荐系统的推荐多样性,须考虑以下情况:用户对于某些维度具有多样性,而对某些维度则不期望具有多样性,例如:某些用户对电影主题不具有多样性,但对电影类型具有多样性。因此,本文思路是通过分析用户与推荐系统的交互记录,提取用户期望具有多样性的维度(项目属性)。最终,通过考虑用户对不同项目维度的多样性偏好(是否期望多样性),本文设计了自适应的多属性多样性算法。本文基于公开的电影数据库进行了实验,结果表明,所提算法在保持较高推荐准确率的前提下,有效地提高了推荐的多样性、新颖性与类别覆盖率。



(a) 豆瓣电影网站的推荐结果



(b) 对于“喜剧之王”的推荐结果(电影类型保持为喜剧电影)

图 1 不同推荐系统对“喜剧之王”的推荐列表

Fig. 1 Recommendation lists of different recommender systems for “King of Comedy”

3 自适应的多样性计算

重新排序算法能够有效地提高推荐列表的多样性,并且不影响推荐程序的总体计算复杂度^[8-9],然而重新排序算法并未从全局角度选择推荐列表。本文对重新排序程序进行了改进,设计了提高推荐系统多样性的用户-项目属性模型,对普通协同过滤算法的推荐列表进行了自适应多属性的配置,以期解决项目多属性导致的组合优化问题,在视频推荐系统的准确率与多样性之间实现了较好的权衡。

图 2 给出了包含重新排序处理的协同过滤推荐系统的框架,该框架主要包含以下几个阶段。

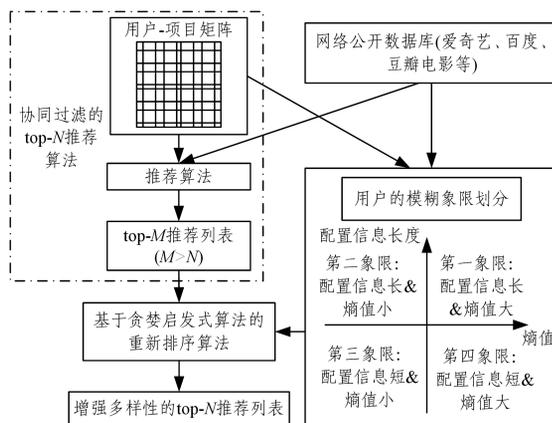


图 2 包含重新排序的协同过滤推荐系统框架

Fig. 2 Framework of collaborative filtering recommender system including e-ranking process

(1)输入数据:1)根据每个用户的评分记录建立用户-项目矩阵;2)描述项目所属类别的结构体,这些数据可以从视频网站或者电影评分网站提取,如优酷网、爱奇艺网、豆瓣电影、百度视频等。

(2)用户建模:推荐系统基于输入的信息为每个用户计算该用户对不同推荐的偏好模型,本算法在该阶段计算用户的模糊化二维坐标系象限。

(3)计算推荐列表:推荐算法采用用户-项目矩阵与项目所属类别的信息计算推荐的项目列表。

(4)重新排序:基于用户的坐标系象限模糊隶属度结果,系统将之前生成的推荐列表重新排序。

(5)输出:系统从重新排序的推荐列表中选择 top-N 的项目。

该模型依赖信息熵,要求每个用户对项目进行足够多的评分,故无法应用于冷启动用户,因此结合用户的个人配置信息^[12]来提高冷启动用户的推荐效果。

本节首先定义了用户-项目属性的坐标系象限概念,并且将用户象限模糊化处理,然后,对多样性推荐算法 MMR, xQUAD 进行修改,加入本文的多样性计算方法与自适应的多属性机制,并且修改了 MMR 与 xQUAD 的目标函数,以提高用户对不同项目属性的多样性判断效果。

3.1 用户的二维坐标系象限划分

为了度量用户对于项目属性偏好的多样性,使用香农熵度量每个用户 u 对于属性 $A \in A'$ 的偏好程度,通过每个属性 $A \in A'$ 的熵值来评估用户 u 的多样性。用户 u 的香农熵值可定义为:

$$H_A'(u) = - \sum_{k=1}^{|\text{dom}(A)|} p_k \cdot \log p_k \quad (1)$$

假设所有的项目均属于用户配置信息的范围,其中 p_k 是属性 A 中第 k 个值的相对频率。

本模型将用户对项目的多样性分为坐标系的 4 个象限,计算用户各个熵值分布与用户配置信息长度分布的中位数,将两个中位数作为用户象限分类的判别阈值。计算用户对于每个项目属性的分类结果,例如:如果 u 的熵值 $H'(u)$ 小于所有用户熵值的中位值,并且 u 的用户配置信息较少,那么用户 u 对于“类别”属性的多样性结果属于坐标系的第三象限。同一个用户对不同属性的结果可能属于不同的象限,图 3 给出了用户的坐标系划分策略。

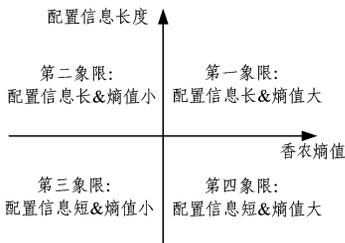


图 3 用户的坐标系划分策略

Fig. 3 Coordinate system division strategy for users

本文分类的理论基础是用户与推荐系统的互动历史记录,判断用户是否满意系统的推荐项目,如果某个用户过去观看同一个主题的视频节目,并且不关心视频的作者,那么认为该用户对视频作者表现出较高的多样性,而对视频节目主题表现出较低的多样性。

给定一个属性 A ,若 A 的熵值较高,则说明用户倾向于选择不同于属性 A 的项目,熵值较低则说明用户倾向于选择类似于属性 A 的项目。因为每个用户与推荐系统的互动记录不同,所以互动记录多的用户,其熵值的效果较好,而对于互动记录少的用户,本文将用户配置信息长度作为用户分类的主要判别依据。

3.2 模糊化二维坐标系象限划分

对于不同的属性,用户所属象限不同,因此,重新给出式

(1)。给定一个用户 u 与一个项目的属性集合 A' ,假设函数 $q_u: A' \rightarrow \{1, 2, 3, 4\}$ 表示用户对项目各个属性的偏好所属的象限,此外引入象限的权重变量 $\omega_k \in [0, 1]$,其中 $k \in \{1, 2, 3, 4\}$ 。为了提高用户象限划分的准确率,设计了用户的模糊化象限划分机制,允许一个用户同时属于多个象限。本文通过定义函数来计算用户对每个象限的隶属度,将熵值与用户配置信息长度进行归一化处理,然后将熵值与用户配置信息长度分别作为坐标系的 x 轴与 y 轴,将 x 值与 y 值作为 4 个双变量高斯函数 (f_1, f_2, f_3, f_4) ,其中 $f_1 \sim N'((0, 0), \sigma^2)$, $f_2 \sim N'((1, 0), \sigma^2)$, $f_3 \sim N'((0, 1), \sigma^2)$, $f_4 \sim N'((1, 1), \sigma^2)$ 为输入变量。模糊化处理将对权重进行加权求和计算:

$$\omega_{q_u(A)} = \sum_{k=1}^4 \omega_k \cdot f_k(x, y) \quad (2)$$

其中, x 为熵值, y 表示用户 u 的配置信息长度, ω_k 为属性 A 的绝对象限权重。因 4 个象限的 σ^2 值相等,故选择坐标 $(1/2, 1/2)$ 作为 σ^2 值,即函数 f 的最大值除以 $4(\sigma^2 = 0.1803)$ 。

3.3 基于贪婪算法的多样性推荐

推荐系统分为两个阶段:第一个阶段预测未评分项目的评分,第二个阶段是将所有项目重新排序,提高推荐列表的多样性则主要在第二个阶段中进行。许多研究已证明搜索多样化推荐列表属于 NP-hard 问题^[13],文献^[14]证明采用贪婪启发式算法能够有效地搜索多样化推荐列表。

搜索多样化推荐列表的贪婪启发式算法如算法 1 所示。

算法 1 搜索多样化推荐列表的贪婪启发式算法

输入: top-N 推荐列表 R, N

输出: 重新排序的推荐列表 S'

1. $S' = \{\}$; // 初始化推荐列表
2. WHILE ($S' < N$) {
3. $i^* = \operatorname{argmax}_{i \in R} f_{\text{obj}}(i, S', u)$;
4. $S' = S' \cup i^*$;
5. $R = R / i^*$;
6. }
7. RETURN S' .

提高推荐列表多样性的方法可分为隐式与显式两种。隐式方法的目标是增加推荐列表中项目之间的平均距离;显式方法的目标是提高用户感兴趣项目类型的覆盖率。

3.3.1 改进的 MMR 算法

MMR 算法^[15]属于隐式的多样性增强算法,通过增加推荐列表中项目之间的平均距离来提高了 top-N 推荐算法的多样性。MMR 的目标函数定义为:

$$f_{\text{obj}}(i, S', u) = \lambda \cdot r^*(u, i) + (1 - \lambda) \cdot \operatorname{avg}_{j \in S} (1 - \operatorname{sim}(i, j)) \quad (3)$$

其中, r^* 是估计评分的函数, sim 是两个项目的相似性度量, λ 为平衡准确率与多样性的权重变量。

因为本文主要求解多属性问题,所以必须考虑属性之间的相似性,对于每个属性 $A \in A'$, $\operatorname{sim}_A(i, j)$ 表示项目 i 与项目 j 关于属性 A 的相似性。将式(3)中项目 i 与项目 j 之间的总体相似性转化为关于用户所属象限的形式,定义为:

$$\operatorname{sim}(i, j) = \left(\sum_{A \in A'} \omega_{q_u(A)} \cdot \operatorname{sim}_A(i, j) \right) / (m * |A'|) \quad (4)$$

其中, $m = \max\{\omega_k | k = 1, 2, 3, 4\}$, $\operatorname{sim}_A(i, j)$ 表示 i 与 j 之间关

于属性 A 的相似性度量。式(4)中用户关于象限的权重值对相似性度量 $sim(i, j)$ 具有影响,因此对式(3)的目标函数也存在影响。

如果一个用户对于某个属性的偏好属于第二象限或第四象限,那么应为 ω_2 与 ω_4 分配一个足够大的值使得原相似性指标具有一个较高的值,因此,如果 S' 中的项目 i 与已有项目极为相似,则会导致 $f_{obj}(i, S', u)$ 的值降低。为了提高推荐列表的多样性,应该减少 S' 中相似度高的项目,换句话说,用户偏好与列表 S' 差异大的项目。反之,为第一象限和三象限分配一个低权重值,则原相似性值明显降低,从而导致对应的 $f_{obj}(i, S', u)$ 的值升高,这说明用户偏好与列表 S' 相似的项目。

3.3.2 自适应 xQUAD

xQUAD 算法^[16]属于显式的多样性增强算法,该算法的思路是减少与已推荐项目中特征相似的项目,增加与已推荐项目中特征不相似的项目。xQUAD 算法的目标函数为:

$$f_{obj}(i, S', u) = \lambda \cdot r * (u, i) + (1 - \lambda) \cdot div(i, S', u) \quad (5)$$

其中, $div(i, S', u)$ 定义为:

$$div(i, S', u) = \sum_f p(i|f) \cdot p(f|u) \cdot \prod_{j \in S'} (1 - p(j|f)) \quad (6)$$

其中, $p(i|f)$ 表示项目 i 包含特征 f 的决策函数。该决策函数的逻辑为:如果项目 i 包含 f ,那么结果为 1,否则结果为 0。 $p(f|u)$ 表示用户 u 对于特征 f 的偏好,即特征 f 在 u 评分项目中的相对频率。

本文为 xQUAD 算法引入了自适应多属性多样性机制。假设 A' 表示项目的属性集合, $A \in A'$ 表示其中一个属性, $f \in dom(A)$ 表示 A 的一个特征。则可将式(6)的 div 函数改写为:

$$div(i, S', u) = \frac{\sum_{f \in dom(A)} p(i|f) \cdot p(f|u) \cdot (1 - \text{avg}_{j \in S'} p(j|f))}{\sum_{f \in dom(A)} p(f|u)} \quad (7)$$

MMR 算法包含一个相似性函数,因此可以直接为 MMR 算法增加象限的权重,xQUAD 算法则使用式(7)通过对各个属性特征之间的多样性进行评估,计算出所有属性的多样性。为 xQUAD 引入权重量,将式(7)改写为加权求和形式,即将 $div(i, S', u)$ 改写为:

$$div(i, S', u) = \sum_{A \in A'} \omega_{q_A(A)} \cdot \frac{\sum_{f \in dom(A)} p(i|f) \cdot p(f|u) \cdot (1 - \text{avg}_{j \in S'} p(j|f))}{\sum_{f \in dom(A)} p(f|u)} \quad (8)$$

4 实验方法与环境

4.1 实验数据集

为了测试本文提出的多样性视频推荐算法的效果,采用公开的 MovieLens 1M 数据集¹⁾进行实验。MovieLens 1M 数据集包含 6040 个用户对于 3952 个电影共计 1000000 个评分。

原数据集包含了电影类型与发布年份等信息,文献[17]为 MovieLens 1M 数据集增加了边际信息,如演员、导演等信息,文献[17]的数据集包含 6040 个用户对于 3883 个项目共计 998963 个评分,并按 6:4 的比例将数据集分为训练数据集和测试数据集。表 1 列出了实验数据集的参数信息。MovieLens 1M 数据集中年份、演员与导演等参数跨度过大,因此,首先对数据集进行预处理,使用 K-means 聚类算法将电影分类,将十年作为电影发行时间的最小粒度。表 2 列出了 MovieLens 1M 数据集预处理后的参数设置。

表 1 MovieLens 1M 数据集的参数

Table 1 Parameters of MovieLens 1M dataset

电影数据集属性	数值
用户数量	6040
项目数量	3883
评分数量	998963
数据稀疏度/%	95.7
平均每个项目的用户数量	275.57
平均每个用户的项目数量	165.39

表 2 MovieLens 1M 数据集的预处理参数

Table 2 Preprocessing parameters of MovieLens 1M dataset

电影属性	数量	分类数量
电影类型	19	—
电影出版跨越的年数	10	—
演员	14736	20
导演	6194	20

4.2 性能评估指标

为了评估视频推荐算法的综合推荐性能,采用了多个性能指标,这些指标可归纳为准确性指标和非准确性指标。非准确性指标包含覆盖率、多样性与新颖性等。仿真实验中将 top-N 的参数 N 设为 10。

4.2.1 推荐准确率

使用精度与召回率作为推荐准确率的性能指标。精度表示 top-N 推荐列表中相关项目的比例。假设 $rel(u, i)$ 为一个布尔函数,代表项目 i 与用户 u 的相关性,结果为 1 表示相关,结果为 0 表示不相关,则精度的定义为:

$$\text{精度} = (\sum_{i=1}^N rel(u, i)) / N \quad (9)$$

召回率表示测试集中包含 top-N 列表中相关项目的比例。假设 $test(u)$ 表示测试集中用户 u 的相关项目集合,则召回率的定义为:

$$\text{召回率} = (\sum_{i=1}^N rel(u, i)) / |test(u)| \quad (10)$$

精度与召回率均无法评估推荐列表中项目位置的准确率,因此采用文献[18]的 nDCG 指标来评估推荐列表项目位置的准确率。nDCG 的定义为:

$$nDCG = \frac{1}{iDCG} \cdot \sum_{i=1}^N \frac{2^{rel(u, i)} - 1}{\log_2(1 + i)} \quad (11)$$

其中, $iDCG$ 表示归一化因子。

4.2.2 个体多样性

采用文献[19]的个体多样性评估方案,假设 R 表示一个推

¹⁾ <https://grouplens.org/datasets/movielens/1m/>

荐集合, $|R|$ 表示 R 的元素数量, ILD(个体多样性)可定义为:

$$ILD(R) = \frac{1}{|R| \cdot (|R| - 1)} \sum_{i \in R} \sum_{j \in R, j \neq i} div(i, j) \quad (12)$$

本文使用基于内容的 ILD, 即将 $div(i, j)$ 作为 $sim(i, j) = avg_{A \in A} sim_A$ 的补充, 其中相似性表示与属性 A 的相似性, 使用 Jaccard 指标计算 sim_A 。

4.2.3 总体多样性

总体多样性是评估电子商务与用户满意度的一个重要指标, 通过提高项目类别的覆盖率可有效地提高电子商务的销售额与用户的满意度^[20]。本文采用项目类别覆盖率与 Gini 系数作为总体多样性的性能指标。Gini 系数可分析 top-N 推荐列表的集中程度, 因此 Gini 系数值越小, 表示分布的质量越低, 反之则越高。

$$覆盖率 = \frac{|\bigcup_{u \in U} top-N(u)|}{|I|} \quad (13)$$

$$Gini \text{ 系数} = 2 * \sum_{i \in I} \left(\frac{|I| + 1 - rank(i)}{|I| + 1} \right) \cdot \left(\frac{rec(i)}{|U|} \right) \quad (14)$$

式(14)中, $rec(i)$ 表示被推荐项目 i 的用户数量, $|U|$ 表示用户总数量, 而 $rank(i)$ 表示 i 在列表中的位置。此外, 覆盖率是 Gini 系数的一个补充性能指标, 覆盖率评估了推荐系统所覆盖的项目类别的数量。只有覆盖率与 Gini 系数均有所提高, 才能说明算法的总体多样性得到了提升。

4.2.4 推荐的新颖性

使用两个基于种群的新颖性指标: 期望的流行度补偿 (EPC) 与推荐的长尾项目比例, 将训练集中评分最低的 80% 的项目作为长尾项目:

$$EPC = \frac{\sum_{i \in R} (1 - pop(i))}{|R|} \quad (15)$$

$$长尾 = \frac{\sum_{i \in Long-tail} rec(i)}{total} \quad (16)$$

其中, $pop(i)$ 表示对项目 i 评分的用户数量(归一化值)。

5 实验结果

本文提出了提高视频推荐系统推荐多样性的算法, 因为

该算法是对 MMR 算法和 xQUAD 算法的改进, 所以将该算法与 MMR 算法、xQUAD 算法进行比较, 而由于 MMR 算法和 xQUAD 算法均是对原 top-N 算法的多样性进行了增强, 因此加入原 top-N 算法进行比较。此外, MMR_ND^[21] 与 ADE^[22-24] 是两个性能较好的多样性增强推荐算法, 因此也将这两个算法作为比较算法。MMR_ND 是对 MMR 算法的改进, 提高了 MMR 算法的新颖性指标, 本文算法也对 MMR 进行了改进以提高多样性。ADE^[22] 则是一种多样性增强的 top-N 推荐算法, 该算法与本文算法的目标一致, 均为提高 top-N 推荐算法的多样性。

通过大量实验发现, 权重 $\{\omega_1, \omega_2, \omega_3, \omega_4\} = \{0.0, 0.0, 0.1, 0.9\}$, $\lambda = 0.6$ 时, 本文推荐算法的推荐准确率最高; 当权重 $\{\omega_1, \omega_2, \omega_3, \omega_4\} = \{0.0, 0.1, 0.3, 0.6\}$, $\lambda = 0.1$ 时, 本文算法的平均性能最好。表 3 列出了几种算法的各个性能指标的平均值(每组实验独立运行 10 次, 将 10 组性能指标的平均值作为最终的结果)。表 3 中本文算法 1 和本文算法 2 表示本文算法在不同参数下的两个版本。本文算法 1 是将本文算法的参数设为 $\{\omega_1, \omega_2, \omega_3, \omega_4\} = \{0.0, 0.0, 0.1, 0.9\}$, $\lambda = 0.6$ 。该算法的推荐精度略高于原 top-N 算法, 召回率略低于原 top-N 算法。可以看出, 本文算法通过调节 4 个象限的权重值与 λ 参数可以完全等价于原 top-N 算法的效果。

表 3 中本文算法 2 是将本文算法的参数设为 $\{\omega_1, \omega_2, \omega_3, \omega_4\} = \{0.0, 0.1, 0.3, 0.6\}$, $\lambda = 0.1$ 。该参数设置下, 本文算法获得了最佳的平均性能, 虽然期推荐准确率低于原 top-N 算法, 但是 ILD、覆盖率、Gini 系数、EPC、长尾均明显优于 top-N 算法。此外, 将本文算法 2 与 MMR 和 xQUAD 两个算法进行比较, 本文算法的推荐准确率与 xQUAD 接近, 并且明显高于 MMR 算法, 而其多样性指标明显优于 xQUAD 算法。虽然本文算法的多样性指标略高于 MMR 算法, 但其推荐准确率明显优于 MMR 算法。因此总体来看, 本文算法有 效地提高了 MMR 和 xQUAD 两个算法的综合推荐性能。

表 3 本文算法与其他推荐算法的平均性能结果

	精度	召回率	nDCG	ILD	覆盖率	Gini 系数	EPC	长尾
原 top-N 算法	0.1488	0.0692	0.1634	0.3551	0.3786	0.2421	0.5736	0.0708
MMR	0.1377	0.0569	0.1509	0.4203	0.3832	0.2487	0.5824	0.0812
xQUAD	0.1417	0.0616	0.1554	0.4109	0.3853	0.2498	0.5844	0.0817
本文算法 1	0.1490	0.0689	0.1637	0.3685	0.3797	0.2438	0.5778	0.0741
本文算法 2	0.1415	0.0610	0.1551	0.4151	0.4072	0.2514	0.7789	0.0824
MMR_ND	0.1402	0.0597	0.1524	0.4117	0.3877	0.2502	0.6654	0.0816
ADE	0.1421	0.0621	0.1533	0.4124	0.3901	0.2512	0.6712	0.0819

MMR_ND 算法是对 MMR 算法的改进, 可看出 MMR_ND 算法对 MMR 算法的推荐准确率、个体多样性、总体多样性、新颖性均实现了增强。与 MMR_ND 算法相比, 本文算法 2 的推荐准确率略优于 MMR_ND 算法, 而个体多样性、总体多样性与新颖性均明显优于 MMR_ND 算法, 其中本文算法的流行度补偿指标 (EPC) 具有较大的优势。ADE 算法的准确率较高, 略高于本文算法 2, 但是其多样性、新颖性均明显低于本文算法 2。

结束语 本文多样性视频推荐算法的理论基础是通过用

户与推荐系统的互动历史记录, 判断用户是否满意系统的推荐项目。如果某个用户过去观看同一个主题的视频资源, 并且不关心视频的作者, 那么认为该用户对视频作者表现出较高的多样性, 对视频节目主题表现出较低的多样性。基于该思想, 本文设计了基于信息熵与用户配置信息的多样性度量方案, 并且将用户对于项目的多属性多样性进行了模糊化处理, 以期提高总体多样性的度量准确性。最终, 基于公开的 Movielens 1M 大型数据集进行了对比实验, 实验结果证明了本文算法对于多样性的增强效果, 并且保证了较小的准确率性能衰减。

参 考 文 献

- [1] ZHUO Y, YOU J L, WANG J L, et al. Measurement and Recommendation System Oriented to Online Video Service in Sea Service[J]. Computer Engineering, 2018, 44(4): 28-34, 40. (in Chinese)
卓煜, 尤佳莉, 王劲林, 等. 海服务中面向在线视频服务的测量与推荐系统[J]. 计算机工程, 2018, 44(4): 28-34, 40.
- [2] WANG X, NIE X, WANG X, et al. A new recommender system framework for TV video[C]// International Conference on Information Science & Technology. IEEE, 2016: 147-152.
- [3] SHI M Z, WU G D, ZHANG Q, et al. Research on the Long Tail Distribution Recommendation of the Multi-topic and RBM[J]. Journal of Chinese Computer Systems, 2018, 39(2): 304-309. (in Chinese)
史明哲, 吴国栋, 张倩, 等. 多主题受限玻尔兹曼机的长尾分布推荐研究[J]. 小型微型计算机系统, 2018, 39(2): 304-309.
- [4] DÍEZ J, MARTÍNEZ-REGO D, ALONSO-BETANZOS A, et al. Optimizing novelty and diversity in recommendations[J]. Progress in Artificial Intelligence, 2018, 1(3): 1-9.
- [5] BENHAMOU F. Fair use and fair competition for digitized cultural goods; the case of eBooks[J]. Journal of Cultural Economics, 2015, 39(2): 123-131.
- [6] SHEUGH L, ALIZADEH S H. A novel 2D-Graph clustering method based on trust and similarity measures to enhance accuracy and coverage in recommender systems[J]. Information Sciences, 2018, 432(1): 210-230.
- [7] WANG S. A recommendation algorithm based on aggregate diversity enhancement[J]. Computer Engineering and Science, 2016, 38(1): 183-187. (in Chinese)
王森. 一种基于整体多样性增强的推荐算法[J]. 计算机工程与科学, 2016, 38(1): 183-187.
- [8] REN C, PING Z, HUA Z. A new Collaborative Filtering technique to improve recommendation diversity[C]// IEEE International Conference on Computer & Communications, 2017.
- [9] DI NOIA T, OSTUNI V C, ROSATI J, et al. An analysis of users' propensity toward diversity in recommendations[C]// Proceedings of the 8th ACM Conference on Recommender systems. ACM, 2014: 285-288.
- [10] HE M, XIAO R, LIU W S, et al. Collaborative Filtering Recommendation Algorithm Combing Category Information and User Interests[J]. Computer Science, 2017, 44(8): 236-241. (in Chinese)
何明, 肖润, 刘伟世, 等. 融合类别信息和用户兴趣度的协同过滤推荐算法[J]. 计算机科学, 2017, 44(8): 236-241.
- [11] WEN J H, YUAN P L, ZENG J, et al. Research on Collaborative Filtering Recommendation Algorithm Based on Topic of Tags[J]. Computer Engineering, 2017, 43(1): 247-252. (in Chinese)
文俊浩, 袁培雷, 曾骏, 等. 基于标签主题的协同过滤推荐算法研究[J]. 计算机工程, 2017, 43(1): 247-252.
- [12] BRAUNHOFER M, ELAHI M, RICCI F. Alleviating the new user problem in collaborative filtering by exploiting personality information[J]. User Modeling and User-Adapted Interaction, 2016, 26(2/3): 221-255.
- [13] REN C, PING Z, HUA Z. A new Collaborative Filtering technique to improve recommendation diversity[C]// IEEE International Conference on Computer & Communications, 2017.
- [14] GOGNA A, MAJUMDAR A. Balancing accuracy and diversity in recommendations using matrix completion framework[J]. Knowledge-Based Systems, 2017, 125(1): 83-95.
- [15] CARBONELL J, GOLDSTEIN J. The use of MMR, diversity-based reranking for reordering documents and producing summaries[C]// International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 1998: 335-336.
- [16] ULLAH M Z, SHAJALAL M, CHY A N, et al. Query Subtopic Mining Exploiting Word Embedding for Search Result Diversification[C]// Asia Information Retrieval Symposium. Cham: Springer, 2016: 308-314.
- [17] RAN J, LEJEUNE M A. Risk-budgeting multi-portfolio optimization with portfolio and marginal risk constraints[J]. Annals of Operations Research, 2018, 262(2): 1-32.
- [18] OSTUNI V C, NOIA T D, SCIASCIO E D, et al. Top-N recommendations from implicit feedback leveraging linked open data[C]// ACM Conference on Recommender Systems. ACM, 2013: 85-92.
- [19] CLARKE C L A, KOLLA M, CORMACK G V, et al. Novelty and diversity in information retrieval evaluation[C]// International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2008: 659-666.
- [20] NIKOLAKOPOULOS A N, KALANTZIS V, GALLOPOULOS E, et al. Factored Proximity Models for Top-N Recommendations[C]// IEEE International Conference on Big Knowledge, 2017.
- [21] CASTELLS P. Improving sales diversity by recommending users to items[C]// ACM Conference on Recommender Systems. ACM, 2014: 145-152.
- [22] ZHANG Z, FAN X Y, GUO Y T, et al. Dynamic Summarization Update Method Based on Topic Signature[J]. Computer Engineering, 2018, 44(6): 169-175. (in Chinese)
张祯, 樊兴悦, 郭禹田, 等. 基于 Topic Signature 的动态文摘更新方法[J]. 计算机工程, 2018, 44(6): 169-175.
- [23] YU X S, SUN S. Research on Personalized Recommendation Model Based on Network Users' Information Behavior[J]. Journal of Chongqing University of Technology (Natural Science), 2013, 27(1): 47-50. (in Chinese)
余肖生, 孙珊. 基于网络用户信息行为的个性化推荐模型[J]. 重庆理工大学学报(自然科学), 2013, 27(1): 47-50.
- [24] WANG Y, WAN X Y, TAO Y Z, et al. Collaborative filtering recommendation algorithm based on K-medoids item clustering[J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2017, 29(4): 521-526. (in Chinese)
王永, 万潇逸, 陶娅芝, 等. 基于 K-medoids 项目聚类的协同过滤推荐算法[J]. 重庆邮电大学学报(自然科学版), 2017, 29(4): 521-526.