

doi: 10.7690/bgzdh.2024.03.005

航空航天科技科普视频资源个性化推荐方法

张晓帆, 张传秋

(中国航天科技国际交流中心, 北京 100048)

摘要: 针对当前航空航天科技科普视频个性化推荐准确率较低的问题, 提出一种航空航天科技科普视频资源个性化推荐方法。构建二值函数结合用户科普视频行为数据, 实现用户航空航天科技科普兴趣挖掘; 利用 Ffmpeg 提取视频帧, 使用关键帧算法, 划分航空航天科技科普视频资源类别; 应用协同过滤算法与优化矩阵分解模型, 设计科普视频资源个性化推荐算法。构建实验环节, 实验结果表明: 此方法的视频类型划分准确性较高, 且可对用户兴趣进行高精度分析, 提升科技科普视频资源个性化推荐准确率。

关键词: 航空航天科技科普视频; 用户兴趣; 个性化推荐; 个性化服务; 聚类挖掘; 关键帧提取
中图分类号: TP391.1; X738 **文献标志码:** A

Personalized Recommendation Method of Aerospace Science and Technology Popular Video Resources

Zhang Xiaofan, Zhang Chuanqiu

(China Aerospace Science and Technology International Exchange Center, Beijing 100048, China)

Abstract: Aiming at the problem of low accuracy of personalized recommendation of popular science video in aerospace science and technology, a personalized recommendation method of popular science video resources in aerospace science and technology is proposed. This paper constructs a binary function combined with the user behavior data of popular science video to realize the user's interest mining of popular science video of aerospace science and technology. It uses Ffmpeg to extract video frames, and uses the key frame algorithm to divide the categories of popular science video resources of aerospace science and technology. It applies the collaborative filtering algorithm and the optimization matrix decomposition model to design the personalized recommendation algorithm of popular science video resources. The experimental results show that this method has high accuracy of video classification, and can analyze the user's interest with high precision, and improve the accuracy of personalized recommendation of science and technology popular science video resources.

Keywords: popular science video of aerospace science and technology; user interest; personalized recommendation; personalized service; cluster mining; key frame extraction

0 引言

随着互联网时代的到来, 人们的网络环境得到了改善。互联网成为人们生活中不可或缺的工具。网络中的数据具有双向性, 用户既是网络数据的浏览者也是数据的生产者^[1-2]。随着数据量的增加, 大量的无用数据出现在用户面前, 严重影响了用户体验, 降低了互联网的便利性。面对海量信息, 如何为用户提供个性化服务, 成为互联网技术中需要研究的热点问题。同时, 在航空航天业发展的推动下, 人们通过互联网观看航空航天科技视频, 大部分的网络用户希望得到更为人性化的个性化推荐, 避免花费大量的检索时间^[3]。

将现行的个性化推荐方法应用到航空航天科技科普视频资源中, 成为解决视频观看用户的重要推荐; 但此部分方法应用后, 存在推荐结果准确性较

差的问题, 影响了用户满意度^[4-5]。针对此问题, 笔者提出一种新型航空航天科技科普视频资源个性化推荐方法, 希望通过个性化推荐方法优化的方式, 提升科普视频资源推荐准确性, 为视频观看用户提供更为便利、更加人性化的服务。

1 用户航空航天科技科普兴趣挖掘

将用户以往的航空航天科技科普视频浏览内容与次数作为基础, 在此基础上对用户的科普兴趣进行捕获与挖掘, 具体操作流程如图 1 所示。

为可以独立反映用户的航天科普兴趣, 在图 1 内容的基础上构建二值函数^[6-7], 则有:

$$f_q(a_i) = \begin{cases} 1, & q = \alpha_1, \alpha_2, \alpha_3 \\ 0, & \text{else} \end{cases} \quad (1)$$

式中: α_1 为保存; α_2 为收藏; α_3 为下载。如 $f_q(a_i)=1$,

收稿日期: 2023-11-16; 修回日期: 2023-12-20

作者简介: 张晓帆(1990—), 女, 河北人, 硕士。

说明用户对此类型的航天科普视频有兴趣； $f_q(a_i)=0$ ，说明用户对此类型的航天科普视频没有兴趣。由于此函数在应用过程中存在一定的不确定性，对其进行优化。考虑到视频自身的内容，将视频资源的时间长度设定为 G_I ，当此取值越大时，用户对其的兴趣度越低。使用 N 表示用户点击科普视频的次数，此参数与用户对此类科普视频兴趣度成正相关关系，当 N 取值越大时，说明用户对此类型视频的兴趣越高。使用 T_I 表示用户的科普视频观看时间，用户对类型科普视频的兴趣度与其呈正相关关系，但并不是 T_I 值越大，用户对其越感兴趣。在对 T_I 进行评定时，需要设定 $\max T_I$ 对其进行约束。根据上述分析结果，用户对某一科普视频的兴趣度可通过下述公式计算：

$$f_q = \delta \frac{N}{G_I * \max T_I} * r^{-|T_I - \max|} \quad (2)$$

式中： δ 为计算过程中的调节因子； r 为兴趣系数。综合上述分析，用户对某一特定航天科普视频类型的兴趣度可表示为：

$$f_q = \min\{1, f_q(a_i) + f_q\} \quad (3)$$

根据上述公式，当用户对某一特定航天科普视频类型的兴趣度 $> f_q$ 时，可将其作为科普视频推荐的决策因素。

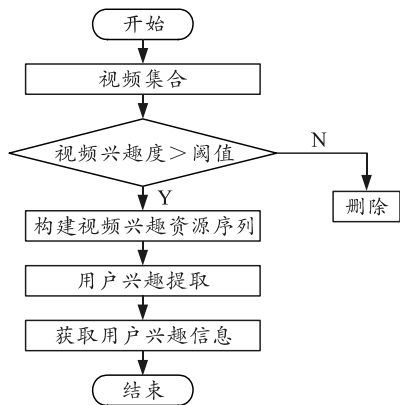


图 1 用户航空航天科技科普兴趣捕获流程

2 航空航天科技科普视频资源类别划分

对于航空航天科技科普视频仅通过基础类别对其科普类型进行划分具有一定的不合理性，使用基础技术划分完成后视频进行个性化推荐存在相应的误差，笔者使用关键帧技术对科普视频类型进行重新划分。

针对某一航天科普视频，利用 Ffmpeg^[8-9]对视频流进行解码获取视频帧，假设共获取到 M 帧，计算每一帧的 RGB 直方图，并根据此计算结果确定

任意两帧 I, j 在 R 通道中的颜色相似度，则有：

$$D_{i,j}^R = \frac{\sum_{k=1}^K (R_i^k - PR_i)(R_j^k - PR_j)}{\sqrt{\sum_{k=1}^K (R_i^k - PR_i)^2 (R_j^k - PR_j)^2}} \quad (4)$$

式中：

$$\left. \begin{aligned} PR_i &= \sum_{k=1}^k R_i^k / k \\ PR_j &= \sum_{k=1}^k R_j^k / k \end{aligned} \right\}, \text{ 则有:}$$

$$D_{i,j} = \max(D_{i,j}^R, D_{i,j}^G, D_{i,j}^B) \quad (5)$$

考虑到 RGB 各通道极大值函数^[10-11]，在计算过程中可使用平均值函数或是其他函数计算颜色相似度。假设给定帧数为 $m(m=M/10)$ ，此时相似度最低的 m 帧为关键帧，则有：

$$(r_1, r_2, \dots, r_n) = \arg \min(\sum_{1 \leq i \leq j \leq n} W_{r_i, r_j}^2)^{1/2} \quad (6)$$

使用上述公式获取视频关键帧。由于此分析过程会对视频信息造成影响，为此设定损失函数，对此过程进行约束。具体损失函数设定如下：

$$L(u_i, t_i) = \sum_i L(u_i, u_i') / N + \chi \sum_i Lu_i'(t_i, t_i') / N \quad (7)$$

式中： u_i 为当前帧为关键帧的概率； t_i 为视频内容对应的航天科普知识。根据此公式，得到损失计算公式：

$$L(b_i) = \begin{cases} b^2/2, & |b| \leq 1 \\ |b| - 0.5, & \text{otherwise} \end{cases} \quad (8)$$

式中 b 为关键帧公式计算值。对上述内容进行整理，使用其对航空航天科技科普视频进行类别划分，为视频推荐提供视频基础。

3 科普视频资源个性化推荐算法设计

在上文设定内容的基础上，构建科普视频资源个性化推荐算法。对比多种方法后，选择协同过滤算法^[11-12]作为基础，实现视频个性化推荐。根据获取到的用户兴趣分析结果，对用户的科普兴趣度及相关排序，则有：

$$v(h, i) = \sum_{c \in n} q_c y_i \quad (9)$$

式中： q_c 为用户对航空航天知识的兴趣度； y_i 为目标用户对第 i 类科普视频的兴趣度。

将此公式与相似度算法进行融合，得到视频推荐大类别计算公式：

$$x_{ij} = |S(i) \cap S(j)| / \sqrt{|S(i)| |S(j)|} \quad (10)$$

式中： $S(i)$ 为用户感兴趣度最高值的科普视频类型； $S(j)$ 为与 $S(i)$ 相似度较高的科普视频类型。为保证推荐结果的精准度，使用优化矩阵分解模型^[13-15]对式(10)进行修正，而后进行求解。在此环节完成后，修正另一个函数，进行二次求解。循环交替运算，直至得到最优解。当目标用户与科普视频较多时，将优化矩阵分解为 2 个小矩阵，2 个小矩阵的乘积无限接近原始矩阵，具体公式设定如下：

$$Z_{M*N} = P_{M*L} * Q^T_{N*L} \quad (11)$$

式中：

$$Z_{M*N} = \begin{bmatrix} z_{11} & \cdots & z_{1n} \\ \vdots & & \vdots \\ z_{m1} & \cdots & z_{mn} \end{bmatrix} \quad (12)$$

此公式表示用户对科普视频的兴趣度矩阵。对此矩阵进行分析，则有：

$$P_{M*L} = \begin{bmatrix} p_{11} & \cdots & p_{1k} \\ \vdots & & \vdots \\ p_{m1} & \cdots & p_{m'n} \end{bmatrix} \quad (13)$$

此公式表示包含用户观看科普视频兴趣的隐性特征。

$$Q_{M*L} = \begin{bmatrix} q_{11} & \cdots & q_{1k} \\ \vdots & & \vdots \\ q_{m1} & \cdots & q_{m'n} \end{bmatrix} \quad (14)$$

此公式表示科普视频包含隐形特征的视频量。对上述公式进行整理，在此公式中增加最小损失函数，提升推荐精度。为避免出现过拟合的问题，使用正则化函数 γ 对其进行优化，则有：

$$\text{loss}(P, Q) = \sum_{i,j} (z_{ij} - p_i q_j^T) + \alpha (\sum_i \|p_i\|^2 + \sum_i \|q_i\|^2) \quad (15)$$

对此公式进行迭代算法，在每次迭代计算中只迭代一个参数，根据此设定进行交替执行，直至结果收敛完成，输出此结果。根据此结果选择合适的科普视频，推荐给目标用户。对上文中设定进行整理，至此，航空航天科技科普视频资源个性化推荐方法设计完成。

4 实验论证分析

笔者提出一种新型航空航天科技科普视频资源个性化推荐方法，为确定此方法具有实用性，构建实验环节对此方法的应用效果进行评定。

4.1 实验准备

此次研究中使用的航空航天科技科普视频均为互联网下载的航空航天科普视频，涵盖 5 类航空知

识内容，共计 5 000 个科普视频作为视频集。实验中使用的视频情况如表 1 所示。

表 1 实验视频概况

实验组	类别数	视频个数	各类视频最多 视频个数	各类视频最少 视频个数
X_1	5	500	20	100
X_2	5	1 500	120	200
X_3	5	1 200	150	300
X_4	5	800	40	80
X_5	5	1 000	95	150

将上述数据作为此次实验的视频数据基础，随机打乱后导入实验平台数据库中。随机选择 10 名用户作为实验对象，对此 10 位用户在近 1 个月内的航空航天科普视频观看浏览记录进行整理，得到每个用户最为喜爱的科普视频类型。在此次实验中，将 5 类科普视频类型分别设定为 A_1 、 A_2 、 A_3 、 A_4 、 A_5 ，根据此分类对用户科普视频的喜好进行分析，具体分析结果如表 2 所示。

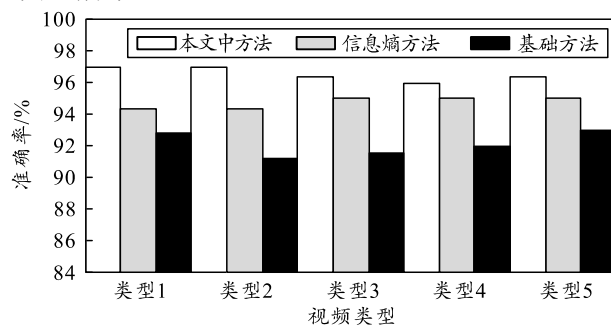
表 2 用户兴趣信息汇总

用户序号	行为数据条数	视频个数	兴趣类型
1	154 810	1 205	A_1 、 A_3
2	215 421	1 200	A_3 、 A_4
3	167 881	952	A_2 、 A_3
4	89 927	3 157	A_2 、 A_5
5	145 266	5 152	A_1 、 A_5

分析用户行为数据量以及观看的视频个数后，得到不同用户的科普兴趣类型，为便于后续的实验分析，笔者仅对用户最感兴趣的视频类型与第 2 感兴趣的类型进行汇总，为实验分析过程提供对照组。

4.2 视频类型划分准确性实验

将实验平台中的视频数据集作为此次实验的数据来源，原有的视频类型划分多按照文本内容进行区分，具有较大的差异性，与真实的视频内容之间有区别；因此，在此次实验中，使用本文中方法、基础方法以及信息熵推荐方法对视频类型进行划分，确定不同方法的视频类型划分准确率。笔者设计了高分辨率和低分辨率 2 组视频，具体实验结果如图 2 所示。



(a) 高分辨率视频

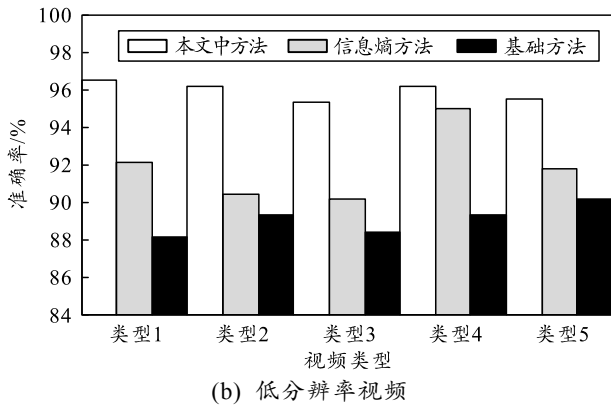


图 2 视频类型划分准确性实验结果

对上述实验结果进行分析可以看出，本文中方法与其他 2 种方法均对原始视频类型展开了二次划分，但划分结果存在较大的差异。本文中方法可对不同类型的视频进行高准确率划分，每一组视频类型划分准确性均可达到 97% 以上。与本文中方法相比，其他 2 种方法的视频类型划分准确性相对较差。基础方法的视频类型划分准确性仅能达到 92%，信息熵推荐方法的视频类型划分准确性达到 95%，相较于基础方法得到了部分提升，但使用效果不及本文中方法。同时，3 种方法在低分辨率视频类型划分过程中，应用差异明显提升，基础方法与信息熵方法的划分准确性大幅度下降。综上所述，本文中方法的视频类别划分能力相对较高。

4.3 个性化推荐准确性实验

将实验准备中预先获取到的用户兴趣作为对照数据，使用实验中选定的 3 种方法对用户进行个性化推荐。此实验共设有 2 个环节，分别为用户首要视频推荐准确率以及用户次要视频推荐准确率。为进一步实现此实验结果量化分析，通过对比每个类别视频推荐数量与该类型视频数量比值的方式，获取实验结果，具体实验结果如图 3 和 4 所示。

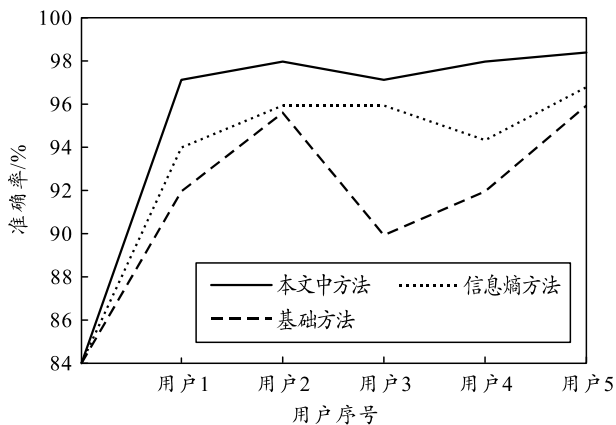


图 3 用户首要视频推荐准确率

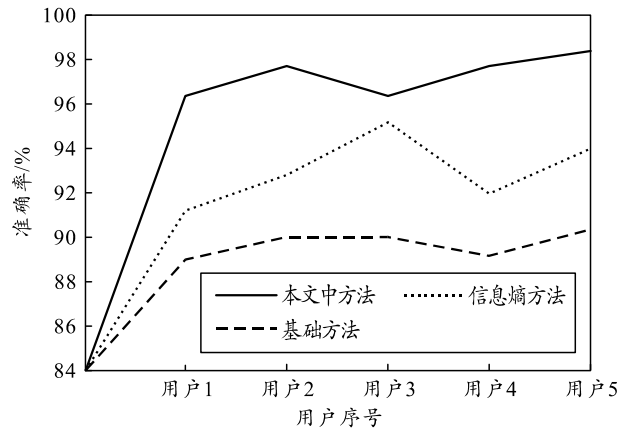


图 4 用户次要视频推荐准确率

从上述实验结果可以看出，在对用户的首要兴趣类型视频进行推荐时，本文中方法的推荐准确率与其他 2 种方法的准确率差异较大。本文中方法可针对用户对科普视频的兴趣度，为其推荐较为喜爱的视频，以此达到用户的满意度。与本文中方法相比，其他 2 种方法对用户的兴趣度分析能力相对较差，无法准确地为用户提供其最为感兴趣的科普视频。

分析上述实验结果可以看出，在对用户第 2 感兴趣的视频进行推荐时，本文中方法的推荐准确率明显高于其他 2 种推荐方法。本文中方法通过用户兴趣分析，准确地确定了用户的喜好类型并对其进行了精准的个性化推荐。基础方法与信息熵推荐方法，在进行个性化推荐时未对用户的个人喜好进行细化分析，对用户推荐时容易出现异常与误差。综合上述实验结果可以确定，本文中方法的个性化推荐能力较强。

5 结束语

此次研究中将用户兴趣作为核心，提出一种新型的航空航天科技科普视频资源个性化推荐方法，经实验证实此方法具有较高的应用价值。同时，此方法还通过对科普视频类型进行了全方位的分析，为提升视频推荐准确率提供基础。由于本文中优化内容相对较多，设定了多个计算环节，为避免计算异常导致此方法的应用效果出现问题，在日后的研究中还需对计算环节进行更新。

参考文献：

- [1] 付鹏强, 苗宇航, 王义文, 等. 航空领域机器人自动钻孔研究进展及关键技术综述[J]. 智能系统学报, 2022, 17(5): 874-885.
- [2] 穆晓霞, 董星辉, 柴旭清, 等. 融合 LDA 主题模型和支

- 持向量机的商品个性化推荐方法[J]. 郑州大学学报(理学版), 2022, 54(3): 34-39.
- [3] 周嘉凡, 杜岳峰, 宋宝燕, 等. 基于元路径注意力机制的MOOC视频推荐方法[J]. 计算机应用, 2022, 42(6): 1808-1813.
- [4] 邵景钰, 董桷朋, 郑泽敏. 高维稀疏个性化推荐方法研究(英文)[J]. 中国科学技术大学学报, 2022, 52(3): 34-45.
- [5] 刘冬邻. 高校科研资源的个性化融合推荐[J]. 重庆大学学报, 2022, 45(7): 122-130.
- [6] 刘芳, 田枫, 李欣, 等. 融入学习者模型在线学习资源协同过滤推荐方法[J]. 智能系统学报, 2021, 16(6): 1117-1125.
- [7] 杨金庆, 程秀峰, 周玮琰. 基于情境感知的资源推荐研究综述与实践进展[J]. 现代情报, 2020, 40(2): 153-159, 167.
- [8] 张翔宇, 张强, 吕明琪. 基于GPS轨迹挖掘的兴趣地点个性化推荐方法[J]. 高技术通讯, 2021, 31(1): 75-83.
- [9] 毛文山, 赵红莉, 孙凤娇, 等. 基于Item2Vec负采样优化的专题地图产品个性化推荐方法研究[J]. 地球信息科学学报, 2020, 22(11): 2128-2139.
- [10] 马晓悦, 马昊. 考虑标签情绪信息的图书资源个性化推荐方法研究[J]. 情报理论与实践, 2020, 43(9): 115-124.
- [11] 孙传明, 周炎, 涂燕. 基于混合协同过滤的个性化推荐方法研究[J]. 华中师范大学学报(自然科学版), 2020, 54(6): 956-962.
- [12] 王嘉琦, 顾晓梅, 王永祥. 混合学习情景下英语视听资源的个性化协同推荐研究[J]. 外语电化教学, 2020(3): 54-60.
- [13] 密成形技术的研究进展[J]. 南京航空航天大学学报, 2020, 52(1): 1-11.
- [14] 郭雪梅. 基于社会化标签的用户标注行为和时间因素的个性化推荐方法研究[J]. 情报科学, 2020, 38(2): 68-74.
- [15] 王娜, 何晓明, 刘志强, 等. 一种基于用户播放行为序列的个性化视频推荐策略[J]. 计算机学报, 2020, 43(1): 123-135.
- *****
- (上接第20页)
- [8] 陶飞, 刘蔚然, 张萌, 等. 数字孪生五维模型及十大领域应用[J]. 计算机集成制造系统, 2019, 25(1): 1-18.
- [9] 张玉良, 张佳朋, 刘检华, 等. 面向航天器在轨装配的数字孪生技术[J]. 导航与控制, 2018, 17(3): 75-82.
- [10] 张旭辉, 张超, 王妙云, 等. 数字孪生驱动的悬臂式掘进机虚拟操控技术[J]. 计算机集成制造系统, 2021, 27(6): 1617-1628.
- [11] 张旭辉, 张雨萌, 王岩, 等. 融合数字孪生与混合现实技术的机电设备辅助维修方法[J]. 计算机集成制造系统, 2021, 27(8): 2187-2195.
- [12] 陈根. 数字孪生[M]. 北京: 电子工业出版社, 2020: 3.
- [13] 陶飞, 马昕, 胡天亮, 等. 数字孪生标准体系[J]. 计算机集成制造系统, 2019, 25(10): 2405-2418.
- [14] 王巍, 刘永生, 廖军, 等. 数字孪生关键技术及体系架构[J]. 邮电设计技术, 2021(8): 10-14.
- [15] 董雷霆, 周轩, 赵福斌, 等. 飞机结构数字孪生关键建模仿真技术[J]. 航空学报, 2021, 42(3): 113-141.
- [16] 陶飞, 刘蔚然, 刘检华, 等. 数字孪生及其应用探索[J]. 计算机集成制造系统, 2018, 24(1): 1-18.
- [17] 朱春明, 何仁平, 周来. 基于数字孪生的总装车间质量智能管控决策应用技术[J]. 兵工自动化, 2022, 41(6): 24-30.
- [18] 苗田, 张旭, 熊辉, 等. 数字孪生技术在产品生命周期中的应用与展望[J]. 计算机集成制造系统, 2019, 25(6): 1546-1558.
- [19] 刘青, 刘滨, 王冠, 等. 数字孪生的模型、问题与进展研究[J]. 河北科技大学学报, 2019, 40(1): 68-78.
- [20] 孙俊义, 刘治红, 郭延涛. 基于数字孪生的关键工装寿命预测综述[J]. 兵工自动化, 2022, 41(6): 7-13.
- [21] 吴健平. 基于AHP-SWOT法的我国BIM应用现状与发展对策研究[D]. 福州: 福建工程学院, 2018.