

Talking about the Challenges Faced by Personalized TV Product Recommendation System

Xingwei Zhang, Chao Chen

Chengdu University of Information Technology, Chengdu Sichuan
Email: zxw-y@qq.com

Received: Jul. 1st, 2019; accepted: Jul. 15th, 2019; published: Jul. 22nd, 2019

Abstract

The triple play enables broadcast and television operators to interact with a large number of home users in real time, making full-scale, personalized product marketing and paid services a reality. As one of the means to solve the information overload problem in the current era of big data, the personalized recommendation system is how to select the results of the user's interest preference and the correct rate in many TV products, so that the recommendation results can be better interpreted sex, novelty and diversity requirements. This paper discusses the key technologies in the personalized TV product recommendation system, discusses the current challenges, and helps to better complete the design and implementation of the recommendation system.

Keywords

Recommendation System, User Portrait, TV Program Label, Data Mining

浅谈个性化电视产品推荐系统面临的挑战

张兴伟, 陈 超

成都信息工程大学, 四川 成都
Email: zxw-y@qq.com

收稿日期: 2019年7月1日; 录用日期: 2019年7月15日; 发布日期: 2019年7月22日

摘 要

三网融合使得广播电视运营商可以与众多的家庭用户实现信息的实时交互, 这使得全方位、个性化的产

品营销和有偿服务成为现实。个性化推荐系统作为解决当前大数据时代下信息过载问题的手段之一, 如何在众多电视产品中挑选出既符合用户兴趣偏好又在保证正确率的条件下, 使推荐结果能够更好的诠释发现性、新颖性和多样性要求。本文对个性化电视产品推荐系统中关键技术进行浅谈, 对目前遇到的挑战进行讨论, 从而有助于更好的完成推荐系统的设计实现。

关键词

推荐系统, 用户画像, 电视节目标签, 数据挖掘

Copyright © 2019 by author(s) and Hans Publishers Inc.

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

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



Open Access

1. 引言

推荐系统作为一种信息过滤技术, 用于提供个性化推荐服务, 帮助用户在“信息超载”的互联网中找到感兴趣的物品, 提高用户选择决策的质量[1]。个性化推荐系统作为解决当前大数据时代性信息过载问题的手段之一[2], 其广泛应用于电子商务产品推荐、新闻推荐、音乐视频推荐等领域。三网融合要求实现网络层面上能够相互联通[3], 在业务发展方面能够相互渗透, 从而确保资源共享程度达到最大化[4]。随着 Internet 相关技术的发展与应用程序开发的快速高效化, 加速了互联网、电信网、广播电视网三网融合的进度, 为传统电视广播媒体带来了新的机遇与挑战。广播电视运营商可以与众多的家庭用户实现信息的实时交互, 这使得全方位、个性化的产品营销和有偿服务成为现实。

1995年3月, Marko Balabanovic 等人在美国人工智能协会首次提出了个性化推荐系统 LIRA。1997年 Resnick 在发表的文章中指出, 人们在生活中常常需要对自己不了解的事务做出判断, 在这种情况下只能依据他人的意见来进行决策, 而推荐系统就是对这种行为的模拟, 通过使用相关算法对他人的用户行为信息进行分析处理, 并将结果提供给需要推荐的用户[5]。2000年清华大学的路海明等提出了基于多 Agent 混合智能实现个性化推荐的研究[6], 作为个性化推荐系统的模型, 标志着我国开始了个性化推荐系统的研究。为应对不同使用场景, 目前各大互联网公司都有自己的个性化推荐系统, 从算法到框架的更新使得个性化推荐系统越来越智能。

通俗来说, 个性化推荐系统为缺乏足够个人经验和能力的使用者提供服务, 因为这些用户不能评估潜在的大量的可供选择的物品[7]。与搜索引擎不同的是, 推荐系统并不需要相应用户提供自己明确的需求。作为搜索引擎的互补模型, 推荐系统需要对使用者的历史行为进行分析与挖掘, 同时完成对用户兴趣的预测建模。以帮助使用者在目的不明确的时候, 发现自身可能比较有兴趣的内容。随着中国广播电视网络公司取得了 5G 商用牌照, 加速了三网融合基础下的个性化电视节目推荐系统的实现。对每个观看用户有针对性的进行电视产品推荐, 从而帮助用户在海量信息中找到适合自己的信息。推荐系统充分利用和挖掘用户的行为数据, 从而帮助用户精准推荐自身比较有兴趣的新产品。通过个性化电视产品推荐系统, 可以减少用户在面对众多节目的选择困难以及为了选择喜欢的节目而不断切换频道造成的时间消耗, 帮助用户在海量电视产品节目中快速的寻找到喜欢的电视产品[8]。个性化推荐系统与搜索引擎的相互补充, 最大限度的使系统比用户更懂用户, 随着深度学习、知识图谱、用户体验建模和视觉体验优化等新技术的发展, 个性化推荐系统不仅增加用户的体验好感度, 而且在算法实现过程中更具有高效性和实时性, 为场景带来更多的点击率和更低的疲劳度。

2. 相关技术

2.1. 实现原理

个性化的推荐系统需要尽可能多完成对数据的采集, 大体来说系统包含使用者数据、待推荐产品数据和相关推荐算法三个基本内容。如图 1 原理框图所示, 通过机顶盒已有数据和网络爬虫相关技术, 实现对数据的采集。用户的收视数据模块用于完成对用户基本资料和收视行为数据的清洗、分析与存储。用户基本资料是在用户申请表中所填写的包括用户年龄、用户性别和用户的职业资料等静态资料数据。收视行为信息, 如观看时长、评分、浏览、收藏、购买等, 能更精确反映单个用户的兴趣偏好, 从而有助于构造用户画像。电视产品数据包含如名称、导演、演员、语种、地区、标清高清、内容描述等多维度数据, 使用数据挖掘、自然语言处理等技术, 提取电视产品的关键信息、物品间的关系和特征, 从而构建合适的标签体系结构使能够更好的达到推荐的效果。针对不同的使用场景与需求, 选择合适的推荐算法能够将用户和商品的关联最大化。在电视产品推荐系统中不仅需要考虑构建用户画像表征的是一个家庭的所有特征, 还需要在兼顾算法准确率时考虑算法的运行时长和不同界面下推荐不同内容等问题。不同的推荐算法各有优劣[9], 用户并不能看到算法的运行过程, 一切都将通过用户界面进行反馈。所以一个好的、科学的、合理的用户界面, 可以提高使用者的体验度, 以此来吸引用户的使用[10]。简单来说, 一个好的电视产品推荐系统需要将效率与体验充分考虑进去。

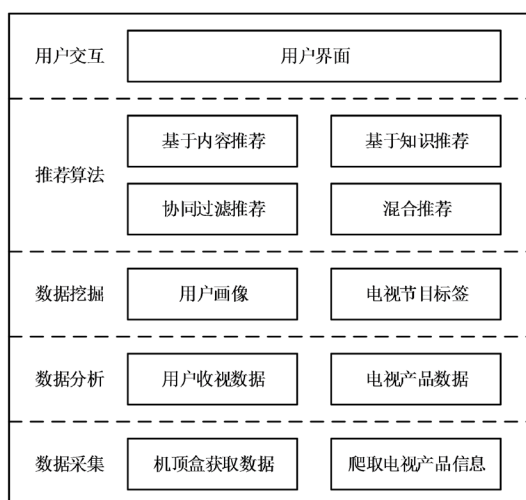


Figure 1. Principle block diagram

图 1. 原理框图

2.2. 数据处理

2.2.1. 所需数据

推荐系统所需数据主要分为两类, 一类是和用户有关的数据信息, 用于构建用户画像; 另一类与电视相关产品信息有关, 用于构建电视产品标签体系库。从机顶盒获取到的信息主要有用户行为信息(收视、点播、回看等)、电视产品信息和用户的消费信息。消费信息包含用户入网时间、用户状态、状态改变时间、套餐等信息。电视相关产品信息包括正题名、创建日期、导演、演员、出品年代、内容描述、连续剧分类、分类名称等。

2.2.2. 用户画像

用户画像表征用户数据的特定业务需求对使用者进行形式化表述, 是通过数据分析和挖掘手段, 将

用户信息进行标签化的处理。用户画像应用于使用者使用过程中的所有生命周期内, 从刚开始对潜在用户挖掘, 到对新用户的引流, 再到对老用户的培育和对流失用户的回流, 用户画像完成对各用户各阶段的不同特征进行深入分析[11]。由于获取的数据是以机顶盒号为单元的, 则所构建的用户画像包含一个家庭内所有成员的信息。构建用户画像的主要流程为:

1) 明确问题导向, 从数据集的规模和数据能够解决什么问题进行考虑。

2) 数据预处理, 主要包括对数据的清洗、集成、变化和规约[12]。数据清洗用于删除原始数据集中的无关、重复数据, 筛选掉与主题无关的数据, 完成对缺失值和异常值处理。数据集成是将多个数据源合并存在一起的过程。数据变换是将数据进行规约化处理, 将数据转化成“适当的、合适的”形式, 以适用于挖掘任务及算法需求。数据规约是为了降低无效、错误数据对建模的影响, 提高建模的准确性, 降低算法的时间复杂度。

3) 特征工程, 最大限度地从原始数据中提取特征, 减少特征数量或者仅选择最佳特征, 供算法和模型使用。

4) 数据建模, 对用户行为构建模型从而形成用户画像。

可以把构建用户画像的过程理解为数据挖掘的过程, 构建一个好的用户画像, 能够判断出该机顶盒用户家庭中的成员分布(家里是否有老人、是否有小孩、孩子或老人的性别和大致的年龄段等)、成员中的活跃用户和每个成员的观看兴趣等信息。针对不能很好的区分家庭成员内每个人的收视偏好, 可以通过与时间段、用户观看习惯相结合进行模糊逻辑推理。

2.2.3. 电视产品标签

建立电视产品标签体系库, 对电视产品所属标签进行系统的、统一的归类。电视产品信息中所涉及到的分类问题是一个多标签问题, 需要给每一个电视节目添加一系列的特征标签。这些特征标签是通过对电视产品信息进行分析和挖掘后得到的。如何建立合适的标签体系, 在现有应用中如爱奇艺、PPTV 等是通过设置一、二级标签的形式进行划分的, 这样的划分方式并不能很好的解决如一部电影/电视剧中同时包含喜剧、爱情或者喜剧、爱情、都市的问题。还需要考虑到的一点就是标签自丰富的问题, 从机顶盒获取到的电视产品信息可能存在数据缺失, 电视产品信息自身的数据可能维度不够等问题, 通常使用爬虫技术和自然语言处理中分词词频统计、同义词转换等技术来解决这类问题。

2.3. 推荐算法

推荐的主要过程是将用户个人信息的特征和物品对象的特征相对应。推荐算法是个性化推荐系统的核心[13], 推荐算法通俗的讲就是利用用户的一些行为, 通过数学算法, 推测出用户可能喜欢的东西, 其很大程度上决定了推荐系统性能的优劣。推荐系统由于数据类型的不同, 常用的推荐算法有基于内容的推荐, 基于协同过滤的推荐, 基于知识的推荐和混合推荐四大类。

2.3.1. 基于内容的推荐

基于内容的推荐算法(Content-Based Recommendations, CB)基于推荐物品或内容找到产品或者内容的相关性, 然后基于用户过去的偏好记录向用户推荐相似的产品。其工作原理为: 数据挖掘技术用于从产品和用户信息中提取产品特征和用户兴趣偏好特征, 从而为用户推荐在特征上与其过去兴趣偏好相似的物品[14]。

2.3.2. 基于协同过滤的推荐

协同过滤推荐(Collaborative Filtering recommendation, CF)与传统的基于内容过滤直接分析内容进行推荐不同, 协同过滤分析用户兴趣, 在用户群中找到指定用户的相似(兴趣)用户, 综合这些相似用户对某

一信息的评价,形成系统对该指定用户对此信息的喜好程度预测。协同过滤推荐算法不依赖于物品特征,仅依赖用户的历史行为数据。基于协同过滤推荐算法可分为:基于用户的协同过滤算法 UserCF 和基于物品的协同过滤算法 ItemCF。

2.3.3. 基于知识的推荐

基于知识的推荐(Knowledge-based Recommendation, KB)由于不需要评分数据,高度重视知识源,因此不存在冷启动问题。主要包含基于约束推荐和基于实例推荐两种方法,这两种方法都需要先收集用户需求,在找不到推荐方案的情况下,自动修复与需求的不一致性,并给出推荐的解释。其区别在于推荐方案是如何被计算出来的。

2.3.4. 混合推荐

在电视推荐系统领域常用混合推荐系统进行推荐,这是因为单一算法并不能很好解决电视产品推荐中所遇到例如系统冷启动问题和新用户的问题。如表 1 常用推荐算法比较所示,对基于内容推荐、协同过滤推荐和基于知识推荐三种推荐方法进行了算法优缺点的分析。

Table 1. Comparison of recommendation algorithms

表 1. 推荐算法常用比较

推荐算法	优点	缺点
基于内容推荐	推荐结果直接,容易解释; 不需要相关领域的知识。	不能解决新用户的问题; 复杂属性不好处理; 要有足够数据构造分类器。
协同过滤推荐	不需要相关领域的知识。 随着时间推移,性能提高; 推荐个性化、自动化程度高; 能处理复杂的非结构化对象。	数据稀疏下效果不佳; 新用户问题; 推荐效果取决于历史数据; 系统开始时推荐质量差。
基于知识推荐	能够将用户需求映射到产品上; 能考虑到非产品属性	知识难获得; 推荐是静态的

混合推荐简单来说就是整合多种推荐算法以获得更好的推荐结果,只有综合运用各种方法的优势、扬长避短、具体问题具体分析,才能形成一个效果强大的推荐系统。混合推荐系统的设计结构主要分为三大类,分别是:整体式混合设计、并行式混合设计和流水线式混合设计。根据设计结构大体上分为推荐算法的直接混合和推荐结果的混合。推荐算法直接混合,例如将基于内容的推荐算法和基于协同过滤的算法混合,在使用协同过滤算法时,如果两个用户共同评分项很少甚至没有,则无法计算其相似度,此时便可以利用用户已评分物品的内容特征或者使用用户的基本属性(性别、年龄、职业等)来计算他们的相似度。推荐结果混合,即单独使用不同的推荐算法,然后对各自的推荐结果进行整合,常用的整合方式有预测评分的线性组合、投票机制等。

3. 面临的挑战

3.1. 用户隐私保护

数据信息作为一种资源,具有普遍性、共享性、增值性、可处理性和多效用性,而数据安全的基本目标就是要确保数据的 3 个安全属性:机密性、完整性和可用性[15]。对推荐系统而言,需要更多的显式

或隐式数据信息进行挖掘, 随之而来的便是个人数据的隐私保护问题。挖掘后的数据更容易被关注, 这是因为这些数据可能会比用户自身更懂自己。实际推荐系统中不仅需要考虑推荐的准确性、新颖性, 同样也要关注对用户隐私的保护, 实现用户数据的安全性。

现阶段, 大数据安全隐私保护技术依然存在诸多漏洞和不足, 关于法律方面的责任划分也没有落实到位[16]。个人信息收集得越多, 安全隐患就越大, 如何保证用户的敏感隐私数据不被泄露出去, 需要在数据采集及预处理阶段完成数据的脱敏工作。数据脱敏又称数据去隐私化或数据变形, 在很大程度上能够解决敏感数据在非可信环境中的安全问题[17]。数据脱敏与数据加密不同, 它是一个不可逆的过程。对数据的全生命周期进行用户隐私保护将成为大数据时代下不可避免的问题。

3.2. 用户兴趣预测建模

人的兴趣会随着年龄和经历进行变化。长期建立起的兴趣爱好由个体的倾向性引起[18], 相对稳定、不会轻易改变, 而短期兴趣爱好是由于当前环境下某些特定条件和刺激而引起产生的[19], 随着外部的情境而不断发生着变化。建立用户画像过程中需要兼顾用户的短期兴趣和长期兴趣, 通过兴趣预测建模对用户长期与短期兴趣进行预测, 从而有助于提高推荐的精准度。

3.3. 系统冷启动问题

系统冷启动问题是针对出现新用户、新电视产品时系统该怎么样进行推荐的问题。对于新用户可以将新用户与历史用户基本信息进行 K 近邻聚类分析形成相似用户簇进行相似用户分类, 同时, 结合非个性化推荐方法(如根据热度推荐、排行榜推荐等)。当新用户拥有一定数量的用户收视行为后选用混合推荐算法继续推荐。对于新出现的电视节目可以通过已有标签与现有标签库进行比对找到最近邻的方式和通过宣传海报与节目内容相结合的方式推荐。系统冷启动问题有助于增强个性化电视产品推荐系统的产品多样性和系统的鲁棒性[20], 帮助解决新用户与新产品出现时所遇到的问题。

3.4. 用户界面的问题

开机首页作为用户打开电视后的第一印象, 所推荐的商品极大地据定了用户接下来的行为, 对用户流量的承接与分发、提升用户收视体验起到至关重要的作用, 成为提升电视用户体验的关键环节之一。由于电视机屏幕和遥控器按键有限的多种问题, 选择合适的用户界面进行展示也至关重要, 推荐系统不仅需要好的算法也需要好的界面将其展示出来。

其主要需要考虑个性化首页的展示风格和电视产品页面(综艺、新闻、体育、财经、电视剧、电影等)推荐列表的展现方式。好的用户界面可以使单一的软件变得更加有个性[21], 可以增加用户体验感。目前电视产品的首页推荐更多的是在相关性推荐的单一数据目标上进行优化, 而随着深度学习、知识图谱等新技术的出现, 推荐系统不仅仅考虑推荐结果的相关性, 还需要在推荐结果的发现性、多样性等方面上做更深度的优化。用户界面作为人机交互的视觉传达和信息传递, 设计适合电视产品推荐系统的用户界面具有很重要的意义。

4. 结束语

随着互联网的发展, 人们逐渐从信息匮乏的时代走向信息过载的大数据时代。大数据时代下, 信息消费者和信息生产者都遇到了很多的挑战, 如何在海量数据中找到有用信息, 怎么样让信息消费者获取到有用信息, 这些挑战加速了推荐系统的发展。电视产品作为大众的一种休闲娱乐的方式之一, 如何快速精准的筛选出用户的兴趣偏好和相应的电视产品, 从而实现系统比自己更懂自己的效果。

参考文献

- [1] 刘春霞, 陆建波, 等. 一种电影推荐模型:解决冷启动问题[J]. 计算机与现代化, 2018(11): 83-87.
- [2] 高凤丽, 孙连山. 个性化推荐系统概述[J]. 技术与市场, 2015, 22(2): 78-79.
- [3] 文臣黎. 三网融合形势下广播电视网络技术发展分析[J]. 广播电视信息, 2019(2): 103-105.
- [4] 徐玉峰. 三网融合背景下广电网络技术问题分析[J]. 数字技术与应用, 2017(2): 241-244.
- [5] 代丽, 樊粤湘. 个性化推荐系统综述[J]. 计算机时代, 2019(6): 9-11+15.
- [6] 路海明, 卢增祥, 李衍达. 基于多 Agent 混合智能实现个性化网络信息推荐[J]. 计算机科学, 2000(7): 32-34.
- [7] (美)弗朗西斯科·里奇, 等. 推荐系统: 技术、评估及高效算法[M]. 李艳民, 等, 译. 北京: 机械工业出版社, 2015.
- [8] 刘炜楠, 杨泽灿, 等. 电视节目推荐系统[J]. 电脑与电信, 2018(7): 30-32.
- [9] 刘辉, 郭梦梦, 等. 个性化推荐系统综述[J]. 常州大学学报(自然科学版), 2017, 29(3): 51-59.
- [10] 黄丽. 基于用户体验的 APP 界面设计探讨[J]. 科技风, 2019(7): 63.
- [11] 王宪朋. 基于视频大数据的用户画像构建[J]. 电视技术, 2017, 41(6): 20-23.
- [12] 张良均, 王路, 等. python 数据分析与挖掘实战[M]. 北京: 机械工业出版社, 2017.
- [13] 袁璐, 沈浩. 基于深度学习的电影海报推荐系统[J]. 现代电影技术, 2018(5): 42-45.
- [14] 赵培. 面向家庭用户的电视节目动态推荐方法研究[D]. [硕士学位论文]. 合肥: 合肥工业大学, 2017.
- [15] 马立川, 裴庆祺, 等. 万物互联背景下的边缘计算安全需求与挑战[J/OL]. 中兴通讯技术, 1-7. <http://kns.cnki.net/kcms/detail/34.1228.TN.20190612.0847.002.html>, 2019-06-14.
- [16] 郑袁平, 贺嘉, 陈珍文, 李雁. 关于大数据安全与隐私保护的研究[J]. 数字通信世界, 2019(3): 166-241.
- [17] 周倩伊, 王亚民, 等. 基于互联网大数据的脱敏分析技术研究[J]. 数据分析与知识发现, 2018, 2(2): 58-63.
- [18] 宗胜. 基于情境兴趣和个人兴趣的用户偏好模型研究与设计[D]. [硕士学位论文]. 上海: 上海交通大学, 2007.
- [19] 王洪伟, 邹莉. 考虑长期与短期兴趣因素的用户偏好建模[J]. 同济大学学报(自然科学版), 2013, 41(6): 953-960.
- [20] 郝予实, 范玉顺. 服务系统中冷启动服务协作关系挖掘与预测[J]. 清华大学学报(自然科学版), 1-8.
- [21] 曾原. UI 设计中的视觉表现分析[J]. 计算机产品与流通, 2019(4): 235.

知网检索的两种方式:

1. 打开知网首页: <http://cnki.net/>, 点击页面中“外文资源总库 CNKI SCHOLAR”, 跳转至: <http://scholar.cnki.net/new>, 搜索框内直接输入文章标题, 即可查询;
或点击“高级检索”, 下拉列表框选择: [ISSN], 输入期刊 ISSN: 2161-8801, 即可查询。
2. 通过知网首页 <http://cnki.net/>顶部“旧版入口”进入知网旧版: <http://www.cnki.net/old/>, 左侧选择“国际文献总库”进入, 搜索框直接输入文章标题, 即可查询。

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: csa@hanspub.org