

用户画像与协同过滤的知识付费平台推荐方法

唐 佳

湖南信息学院 湖南长沙 410000

摘要: 在知识经济时代, 国内知识付费平台的规模、用户数和交易额持续快速增长, 伴随着网民付费习惯的养成、在线交易技术的完善和数字版权保护力度的加强, 知识付费在线上开始高速发展。所以如何把精准的个性化推荐内容提供给知识付费平台的用户, 提升用户感知平台的服务品质, 就变得至关重要了。尽管传统的协同过滤算法具有简单、有效等优点, 但评分矩阵在用户数量和项目数量增加、单纯依靠评分数据进行计算、降低推荐质量、难以满足用户个性化推荐需求的情况下, 在不考虑用户特征和偏好影响的前提下, 存在着严重的稀疏问题, 导致用户相似程度的计算准确度难以保证, 推荐质量也难以得到满足。

关键词: 用户画像; 协同过滤; 知识付费; 推荐方法

User portrait and collaborative filtering of knowledge payment platform recommendation method

Jia Tang

Hunan University of Information Technology, Changsha, Hunan 410000

Abstract: In the era of knowledge economy, the scale of domestic knowledge payment platform, the number of users and transaction volume continue to grow rapidly, with the development of Internet users' payment habits, the improvement of online transaction technology and the strengthening of digital copyright protection, knowledge payment online began to develop rapidly. Therefore, how to provide accurate personalized recommendation content to the users of the knowledge payment platform and improve the service quality of the users' perception of the platform has become crucial. Although the traditional collaborative filtering algorithm has advantages such as simplicity and effectiveness, the scoring matrix has serious sparse problems when the number of users and items increases, the calculation solely depends on the scoring data, the recommendation quality decreases, and it is difficult to meet the user's personalized recommendation needs without considering the influence of user characteristics and preferences. As a result, it is difficult to ensure the accuracy of calculation of user similarity degree, and it is difficult to meet the quality of recommendation.

Keywords: User portrait; Collaborative filtering; Pay for knowledge; Recommended method

一、知识付费及知识付费平台

1. 知识付费定义

“知识付费”是在邹伯涵等人认为“内容付费”与“知识分享”的交集为“知识付费”的基础上逐渐兴起的一个概念, 主要体现为发布者提供付费问答、付费课程、技能付费分享等服务模式, 面向平台上的其他用户。一般而言, 内容付费包含知识付费, 但同时也包含了影视、游戏、音乐等内容的付费, 因此, 内容付费是包含知识付费的一种付费方式。知识付费范畴中包含了在线教育, 因为知识付费涵盖的是课程付费, 而在线教育的表现形式大部分是付费课程, 但在线教育并不只属于知识付费, 毕竟线教育的兴起要比知识付费来得更早, 因此二者是一种包含但又各具特色的关系。

2. 知识付费平台类型

知识付费平台是传统知识付费行业的网络平台化, 或者是个人、团队等内容发布者创造、分享知识并获得一定收益的媒介, 是在互联网普及、人们已经养成在线付费习惯之后产生的。对于如何划分知识付费平台类型, 研究者们并未达成统一的意见, 或者将其划分为三种类型: 垂直化平台、社交化平台和大众化平台, 以平台受众为对象。

或者把在线教育类, 专业网站类, 社交媒体类, 从电子商务的角度来划分。或以数字出版知识服务的视角切入, 分为知识资源库、知识获取终端设备、社交媒体知识分享、开放知识众编、知识付费订阅等多个版块, 通过数字出版知识服务的视角切入高速发展的知识付费丰富了知识付费平台的种类, 用户选择使用最适合自己的知识付费平台类型, 可以根据自己当前的需求变化随时进行选择。总的来讲依据现有知识付费平台类型的划分, 可以将知识付费平台划分为知识电商类、直播讲座类、第三方付费工具类、社交问答类、付费文档类、内容打赏类以及付费课程类, 本文的研究主体知乎 live 则属于直播讲座类的知识付费平台。

3. 用户付费现状分析

2020年春节期间, 由于新型冠状病毒疫情防控, 我国线下实体行业遭受巨大冲击, 而互联网尤其是在线知识付费行业不仅经受住了艰难考验还爆发了许多的商机。调查发现, 这一时间段国内有63.1%的用户购买过知识付费产品, 而知识付费平台原有用户的购买率更是高达88.8%。随着社会竞争压力的增大, 用户购买知识付费产品的目的大多为专业学习、自身发展、提高经济收入等。随着社会

的发展、职场竞争压力的增强,我国消费者愿意支付一定金额获取优质的知识付费产品,并且每月付费金额普遍较高。据2020年调查数据显示,只有17.1%的用户认为所购买的知识付费产品完全符合期望,因此如何让每个用户在较短的时间找到最适合自己的知识付费产品,提高知识付费平台的个性化推荐质量显得尤其重要,亦是本文的研究重点。

二、用户画像分析

用户画像的用途是十分广泛的,如何利用庞大的用户数据完成用户画像的构建,从而帮助企业实现商业价值。这一章节主要介绍用户画像的概念、构建流程、影响因素以及数据来源,为构建知识付费平台用户画像模型奠定基础。

1. 用户画像构建流程

建立平台的用户画像模型要从实际应用需求出发,以解决平台面临的业务问题为根本目的,根据用户的信息数据进行设计。

第一阶段:数据收集与处理。搜集数据的渠道是多样的,对于用户的基础信息主要是利用注册表信息、平台私信调查等方式进行采集;对于用户的其他信息主要是利用数据挖掘手段搜集系统日志的形式进行获取。通常获取的初始数据都存在脏数据,比如数据缺失、冗余、异常等问题,为了避免这些问题数据对用户画像建模产生影响,需要对初始数据进行筛选、集成、归约等数据预处理的操作。

第二阶段:用户标签生成。标签生成阶段主要是基于第一阶段数据的基础上利用机器学习算法进行特征提取,这一过程往往需要借助于关键词提取技术,通过分析上下文之间的关联选出可以代表文本内容的关键词,从而获得可以准确描述用户的特征标签。通常使用人工提取或机器学习的方式进行特征提取,对于图片、视频和语音等内容来说,采用前者的方式会比较精准,但对于文本信息来说,采用后者进行处理的速度和效率则远远高于前者,因此选择合适的处理方式能够提升标签抽取的准确度和可靠度。

第三阶段:用户画像建模。这一阶段主要是根据不同的业务需求,使用不同的机器学习算法对处理过的数据进行建模,并随着用户行为的变化不断更新用户标签,使用户画像更加丰富具体、更有实效性。还可以通过数据可视化的方式将用户画像的标签特征进行展示,让运营者可以清晰的了解用户类别。

2. 用户画像影响因素

用户画像早期被应用到产品设计、市场营销等方面,通过问卷调查等手段获取用户的需求,绘制产品目标用户的画像,使设计的产品与用户的需求相符,从而使企业实现商业化目标。大数据时代,相关用户画像的研究发生了新的变化,研究人员开始利用数据挖掘、数据分析等手段,通过平台上大量的用户数据来分析用户的基本特征、行为习惯、情感偏好等信息,为每个用户提炼出了高匹配度的特征标签,从而建立更为全面的用户画像。同时,用户画像的应用范围和场景也在不断扩大,从市场营销、社交媒体延伸到公共卫生、情报服务等范围,应用场景也覆盖了个性化推荐、异常行为检测等。目前若要构建精准的用户画像模型,主要存在以下几个影响因素:

(1) 数据基础。用户数据的获取如注册信息、个性信息、行为数据、社交互动、交易频率等,越丰富的数据描绘的用户画像越精准,但往往需要的搜集成本是越高的。

(2) 信息隐私限制。在获取用户数据信息时存在隐私限制问题,目前各大平台都鼓励甚至奖赏用户填写个人信息,但是为避免隐私暴露可能带来的隐患,多数用户选择不填写或是编造个人信息,为建立精准用户画像提高了难度。

(3) 业务需求。通过对搜集到的数据进行深层次的分析,将用户需求和具体业务相结合,充分考虑实际应用的场景,这一步骤是实施用户画像能否成功的关键性步骤,但实际业务的需求往往是很复杂的,通常需要分析大量的数据才能解决一个问题。

3. 用户画像数据来源

构建用户画像的目的是将平台上的用户信息全貌尽量展示出来,因此平台上只要是与用户相关的数据均为用户画像的数据来源,这些数据根据属性的不同可以分为静态数据和动态数据。

静态数据。往往指的是用户比较固定的,通常不会轻易发生改变的数据,比如性别、身高、学历、地区等用户基本信息。这种数据通常为结构化数据,主要依靠显式方式进行收集,比如利用注册表单要求用户填写与个人相关的信息。通过这种显式收集方式得到的数据,系统只需对其进行筛选、标准化等方面的整理性操作即可,无须对其建立数据模型或者进行下一步的预测等工作。

(2) 动态数据。通常指会经常发生改变的用户数据,比如搜索记录、浏览内容、位置定位信息、发表个人评论等用户行为信息。这种数据里隐藏着用户的偏好信息,通常为非结构化或半结构化的数据,主要依靠隐式方式进行收集并且需要全力进行挖掘。通常系统会在不被用户察觉的情况下,利用用户的操作行为采集相关数据。采用隐式方式收集的数据在准确性方面往往较差,但通过对平台上大量的用户行为数据进行建模,可在一定程度上进行弥补。

三、基于用户画像与协同过滤的推荐方法

用户画像是根据用户信息和行为分析出的用户特征描述。协同过滤是一种推荐技术,通过分析用户或物品间的相似性来进行推荐。结合用户画像与协同过滤,可以实现更精准的个性化推荐,提高推荐的准确性和用户满意度

1. 基于协同过滤推荐的方法

① 用户协同过滤

用户协同过滤通过计算用户之间的相似度,找出与目标用户相似的其他用户,然后推荐这些相似用户喜欢的物品给目标用户。相似度计算可以采用多种方法,如皮尔逊相关系数、余弦相似度等。

② 物品协同过滤

物品协同过滤通过计算物品之间的相似度,为用户推荐与其喜欢的物品相似的其他物品。物品相似度的计算同样可以采用多种方法,如余弦相似度、Jaccard 相似度等。

2. 基于用户画像推荐的方法

① 根据用户兴趣推荐

通过分析用户在平台上的浏览、点击、收藏、购买等行为数据,挖掘用户的兴趣点,为用户推荐具有相关性的

内容。例如,若用户在知识付费平台上频繁浏览和购买关于摄影的课程,推荐系统可以为该用户推荐更多与摄影相关的课程。

②根据用户行为推荐

用户在知识付费平台上的行为数据(如登录频率、停留时长、互动情况等)可以帮助推荐系统更好地理解用户的需求。例如,若某用户喜欢在晚上学习,推荐系统可以在该时段为用户推荐新课程。此外,根据用户的行为数据,推荐系统还可以为用户推荐短时高效的课程、适合碎片化学习的课程等。

③根据用户消费能力推荐

用户的消费能力是影响推荐策略的另一个重要因素。通过分析用户的消费记录和购买力,推荐系统可以为用户推荐价格合适的课程,从而提高用户的购买意愿。例如,针对购买力较高的用户,推荐系统可以为其推荐高价值、专业度较高的课程;而针对购买力较低的用户,推荐系统可以为其推荐性价比较高、适合入门级用户的课程。

3. 结合用户画像与协同过滤的推荐优化

①多维度融合推荐

结合用户画像的多维度信息(如兴趣、行为、消费能力等)与协同过滤的相似度计算,实现对用户个性化需求的深度挖掘。通过多维度融合推荐,可以提高推荐的准确性和覆盖面。

②动态调整推荐权重

根据用户行为的变化和反馈,动态调整用户画像与协同过滤在推荐中的权重,以实现推荐结果的实时优化。例如,当发现用户对某一类别的兴趣明显增加时,可以提高该类别在推荐中的权重。

③防止过度个性化和信息茧房问题

在进行推荐时,需要关注过度个性化和信息茧房问题。过度个性化可能导致用户仅接触到狭窄的兴趣范围内的信息,限制了用户的视野。为解决这一问题,推荐系统可以适当引入热门、新颖或跨领域的内容,以保持用户对新知识和新领域的探索欲望。

四、结论

综上所述,融合用户画像与协同过滤算法的知识付费平台个性化推荐模式,不仅可以弥补协同过滤算法忽略用户偏好的缺陷,提高平台推荐质量,还可以全面刻画平台用户画像,帮助运营者清晰了解平台用户定位,促进平台可持续的高质量发展。

参考文献:

[1] 仇阿根,张用川,罗宁,郑莹莹,陆文.结合用户特征的政务服务协同过滤推荐方法[J].集成技术,2023,12(01):42-55.

[2] 凌坤,姜久雷,李盛庆.基于改进用户画像的协同过滤推荐算法[J].计算机仿真,2022,39(12):533-541.

[3] 谢少辉,段旭磊,张仰森,侯振瑜,陈琳,范国梁.基于用户画像的军事信息推荐方法[J].指挥信息系统与技术,2022,13(03):72-77+84.

[4] 李悦.基于用户画像的图书馆图书精准推荐研究[D].燕山大学,2022.

[5] 宫丽霞.基于用户画像与改进协同过滤的混合推荐算法研究[D].内蒙古大学,2022.

[6] 李凯.面向信息茧房的用户画像多样化推荐模型研究[D].燕山大学,2022.

课题名称:【基金课题】“基于用户画像与协同过滤的知识付费平台推荐方法研究”(编号:21C1361)