

基于用户画像的信息后组织与个性化推荐 ——以图书馆为例

章元元

(华中师范大学第一附属中学初中部 湖北 武汉 430000)

[摘要] (目的) 大数据环境下, 图书馆面临着服务转型, 容易出现用户难以准确描述自身需求, 盲目推荐易与用户期望产生偏差, 图书馆的服务需要由被动转为主动等问题。用户画像正是精准描述用户的一种策略。将其应用于图书馆, 可以更好地满足用户多样的个性化需求, 提升图书馆开展各项服务的精准性。(内容) 图书馆大多数应用的信息组织方式是信息先组, 这种方式不利于准确反映用户的需求和偏好。可以将信息后组与用户画像的策略联系起来, 应用在图书馆中。在对用户准确画像的基础上, 将检索结果与用户画像相匹配, 再次筛选, 提供给读者更为精准的信息。(结果) 用画画像当前还是面临着一定的困境, 其中数据安全、数据标准、激励机制、共识机制是亟待解决的问题。在用户画像的解决上, 区块链技术的引入是一种较有效的途径。基于区块链可以建立一种全新的且能量化的信任方式, 改变传统信用机制的弊端。

[关键词] 用户画像; 信息后组织; 个性化推荐; 图书馆

【DOI】 10.12252/j.issn.2096-627X.2021.09.995

一、用户画像在图书馆中的应用必要性

(一) 大数据环境下用户难以准确描述自身需求

“互联网+”、大数据存储技术、云计算技术和无线通信技术的发展, 衍生出海量的信息数据内容、人们每天都主动被动地接受各种碎片化的信息, 信息的冗余更加严重。在这样的背景下, 人们面临更多信息的选择, 也被被动地被各种信息所影响, 更容易受到碎片化信息的左右, 容易产生群体的从众行为。

(二) 盲目推荐易与用户期望产生偏差

图书馆的资源数据量呈爆炸式增长, 浩瀚如烟的数字资源使得用户易产生信息迷航。在用户偏好未知的情况下, 无法对用户的行为进行准确把握, 此时很难向用户推荐能够产生吸引力的信息。读者个人具有自身特点和特殊的性格特征, 这种个性化是无法避免的, 如果盲目地向用户推荐信息容易与用户的期望产生偏差, 无差别的信息推送也很难满足这种个性化的需求, 难以吸引读者的参与。

二、基于用户画像的信息后组织

信息后组是相对于信息先组织来说的。需要在对检索结果利用事先建立的用户需求模型进行筛选并将筛选后的结果进行二次组织整理, 最后将经过处理的结果返回给用户。是在用户一次检索结果之上的再次组织。

图书馆大多数应用的信息组织方式是信息先组, 将数据库中的信息进行组织之后, 用户再进行检索。这种方式不利于准确反映用户的需求和偏好。可以将信息后组与用户画像的策略联系起来, 应用在图书馆中。在对用户准确画像的基础上, 将检索结果与用户画像相匹配, 再次筛选, 提供给读者更为精准的信息。

(一) 用户画像的数据抓取范围

用户画像建立在广泛的数据分析之上, 数据抓取的越全面, 画像就会越准确。图书馆的读者与营销用户的特点还是有

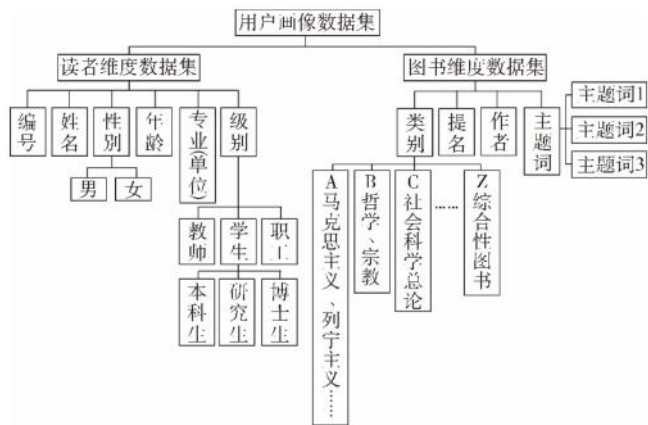
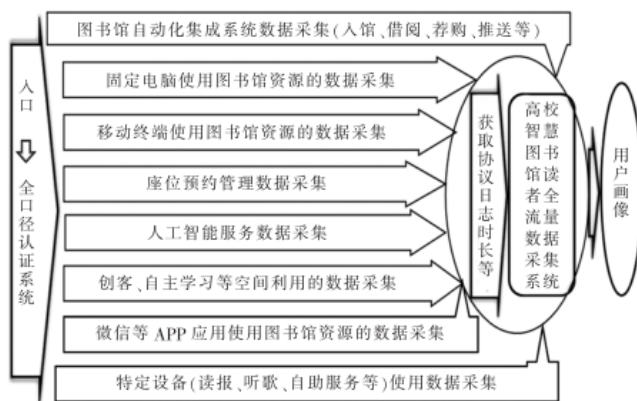
一定的区别的, 因此, 图书馆对读者的数据抓取应该结合图书馆读者的自身特点。

图书馆的读者的自身属性, 例如年龄、性别、专业, 阅读时的情绪状态、兴趣爱好等都会影响读者的阅读选择。笔者认为读者的画像数据集主要包含两个方面, 读者维度数据集、图书维度数据集。读者维度中, 需要收集读者的属性信息, 例如姓名、性别、年龄、专业、编号等。读者行为偏好信息, 例如用户的浏览时间、收藏、评论、分享等行为。用户情景信息, 例如用户的所处的地点、情绪状态等。图书维度的数据集主要包含读者经常浏览的图书类别、提名、作者、主题词等。

(二) 用户画像的数据抓取渠道

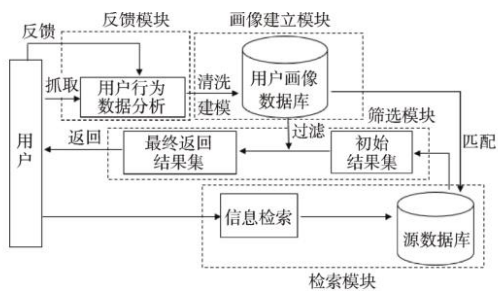
图书馆的现有读者相关数据获取渠道比较有限, 但以上所涉及到的数据范围比较广泛, 需要图书馆扩大数据的抓取渠道。

图书馆在数据抓取渠道上主要分为三种: 一是通过用户注册、登记信息采集用户特征数据, 如姓名、性别、院系、专业、研究方向、兴趣爱好等; 二是通过图书馆视频监控系统、管理信息系统、网络监控器、传感器、阅读终端设备、用户可穿戴智能设备、自助预约系统等采集数据; 三是第三方服务商协议共享的用户查阅的相关数据, 如社会活动数据、用户偏好、移动终端记录等。



(三) 基于用户画像的信息后组织系统

在读者数据全面搜集的基础之上图书馆可以建立基于用户画像的信息后组织系统, 信息后组织系统的实现需建立在用户需求模型的基础上, 而目前随着大数据分析技术的日渐成熟, 用户的行为习惯和需求特征可被更精确地描述, 进而形成用户画像, 作为对用户需求的表征被用于信息后组织系统中。信息后组织系统主要包括用户画像建立、信息检索、过滤筛选和信息反馈4个模块。



三、基于用户画像的个性化推荐

个性化推荐在市场营销中引用较多，可以将此概念引入图书馆服务。图书馆在转型过程中，需要更多地思考如何将服务更加主动化，而个性化推荐就是在充分解读读者需求的基础上，更加主动地向读者提供信息，保持与读者的联系。

基于用户画像，图书馆的个性化可以分为两个方面：推荐方式的个性化与推荐内容的个性化。图书馆向用户进行个性化推荐时应当同时涉及两个方面，将方式与内容结合起来，打造个性化的、符合读者个人特点和偏好的信息推荐。

(一) 推荐方式的个性化

当下图书馆不是只能依靠物理实体的储存与查阅信息的机构，人们对阅读途径的选择更加多样化。图书馆在用画图像的数据搜集之前，就应当把阅读的途径和方式考虑在内，作为对用户推荐方式个性化的基础。

用户画像在数据采集时，通过图书馆视频监控系统、管理信息系统、网络监控器、传感器、阅读终端设备、用户可穿戴智能设备、自助预约系统等采集数据，另外第三方服务商协议共享的用户查阅的相关数据，如社会活动数据、用户偏好、移动终端记录等这些渠道都可以作为推荐方式个性化的数据基础。读者的阅读方式主要可以分为纸质化阅读和数字化阅读，除此之外，无障碍的听书阅读也正在发展中。读者在纸质化阅读和数字化阅读之间会有一定的偏好选择。通过图书馆内的智慧化管理系统可以采集到读者进馆阅读的频率，图书馆的移动终端应用的服务平台也可以收集读者的相关信息，将同意读者的信息相关联，从而刻画出用户的用户画像，将用户的使用偏好进行标记。

对于偏好纸质化阅读的读者，图书馆可以在馆内举办相关的活动，结合读者的兴趣内容，将图书馆中的数据库与读者用户画像进行匹配，给用户推荐相匹配的信息。对于偏好数字化阅读的读者来说，图书馆可以监测用户经常使用的数字化阅读渠道，建设微信、微博、门户网站、阅读设备等多种阅读平台。

(二) 推荐内容的个性化

推荐内容的个性化涉及到一定的信息组织与筛选技术。信息筛选的实现主要以满足用户的个性化需求为目标，利用用户画像对初始结果集进行过滤，提高最终结果集与真实需求的拟合程度。可以采用基于内容的过滤技术来完成对初始结果集的筛选，即根据初始结果集中信息内容与用户画像之间的相似程度来进行信息过滤。将用户画像中用于表征用户需求偏好的标签以向量形式表示 $U_i=(U_{i1}, U_{i2}, U_{i3}, \dots, U_{in})$ ，再将初始结果集中的信息经过主题分析后，形成表征内容的特征向量 $P_i=(P_{i1}, P_{i2}, P_{i3}, \dots, P_{in})$ ，接着采用向量空间余弦值的方法进行两个向量的相似度计算，最后根据预先设立的条件，将超过一定余弦值的向量所对应的信息作为筛选出的信息，集成最终结果集返回给用户。

四、用户画像面临的困境与解决途径

(一) 用画图像面临的困境

1. 数据安全问题

用户画像所采集的数据越全面，画像标签就越准确。但另一方面，用户的个人隐私也就更可能面临威胁。这是二者相矛盾的地方。图书馆对读者数据采集的过程中，数据库有遭到黑客攻击而造成数据泄露或篡改的风险。另外，这些数据涉及到的用户的知识产权问题。有些数据涉及到的用户知识产权一

旦不能被有效保护，也容易引起知识产权的相关问题。笔者认为，以上图书馆的数据采集过程都应当以告知读者并征得读者的前提下进行的。

2. 数据收集的标准难以统一

图书馆大数据涉及利益相关者的多个层面和维度，但在收集数据的过程中，由于技术和应用领域的限制，往往是采用单方面的标准收集那些便于收集、易于收集以及想收集的数据，这相当于对移动图书馆用户画像数据进行了人工的降维，因此限制了画像数据的全面收集。

3. 激励机制不健全

数据收集方为了提高画像数据的价值密度，投入大但成效不甚明显，难以在短期内收回成本，进而降低了其利用大数据进行价值创造的效率与积极性，这样就在逻辑上陷入了画像数据获取的困境。因此，如何通过先进技术和有效措施来保证用户的数据安全以及数据所有权的清晰归属，成为突破画像数据获取困境的重点与难点所在。

(二) 区块链技术的引入

从区块链的特点来看，区块链具有分布式记账、非对称加密和授权技术、智能合约和共识机制等核心特征。去中心化特征可以很好地处理交易过程中的交易流程繁杂、交易成本高、信息不对称与相互信任问题，且区块链系统中每个区块都包含了整个系统内的全部信息，可通过交叉验证来保证数据的真实性与有效性。此外，每一数据块中都保存了加密过的交易备用数据，某一区块的数据丢失不会导致整个系统工作瘫痪，因此还兼有安全性高、数据库可靠的特征。因此，笔者认为，将区块链技术引入移动图书馆用户画像领域，以区块链技术为底层支撑技术，基于区块链理念可以解决目前图书馆用户画像服务面临的困境。

五、结论

将用户画像的概念和技术与图书馆相结合，能够将图书馆的信息较为精准地推荐给读者。基于用户画像，图书馆需要在告知用户的前提下全面采集读者的数据。并应用信息后组织，对用户的一次检索结果再进行组织，并建立用户画像的信息后组织系统。在此基础之上，将推荐方式与推荐内容的个性化结合起来，充分体现读者的个人特征，实现精准推送。

参考文献

[1]肖海清,朱会华.基于参与式用户画像的高校图书馆精准阅读推广模式构建[J].图书馆工作与研究,2020(06):122-128.

[2]董宁.面向群组用户画像的图书馆精准服务体系构建[J].图书馆论坛,2020(06).

[3]刘颖.大数据技术的图书馆用户画像模型建构策略[J].现代电子技术,2020(09):55-57+61.

[4]戴莹.图书馆用户画像服务系统的构建研究[J].四川图书馆学报,2020(02):44-48.

[5]张宇,阮雪灵.大数据环境下移动用户画像的构建方法研究[J].中国信息化,2020(04):65-68.

[6]尹婷婷,曾宪玉.用户画像视角下数字图书馆个性化信息服务建模与分析[J].新世纪图书馆,2020(04):57-61.

[7]康存辉.基于用户画像的高校智慧图书馆服务空间再造探索[J].图书馆工作与研究,2020(04):79-83.

[8]杨群,曾真,莫再峰.面向用户画像的图书馆知识服务体系构建分析[J].图书馆,2020(03):81-87.

[9]杨彦荣,张莹.图书馆用户画像设计与应用研究[J].数字通信世界,2020(02):34+57.

[10]刘海鸥,姚苏梅,黄文娜,苏妍嫒.移动图书馆用户画像大数据应用的困境与对策——基于区块链理念[J].图书馆学研究,2019(23):26-33.

[11]王娜,陈倩玉.基于用户画像的信息后组织系统的构建研究[J].情报理论与实践,2020(04):145-151.

作者简介:
章元元(1992—),女,安徽省庐江县人,硕士研究生,初级,华中师范大学第一附属中学初中部。