

基于用户画像的高校数字图书馆精准服务模式研究

张志清^{1,2}, 聂志璇¹

(1. 武汉科技大学恒大管理学院, 武汉 430081;
2. 武汉科技大学服务科学与工程研究中心, 武汉 430081)

摘要: 网络技术的发展促进了图书馆服务模式从知识型服务向智能型服务的转变。用户画像是真实用户的虚拟代表, 能够更准确地描述用户特征。研究以用户画像为核心, 构建了高校数字图书馆精准服务的框架模型, 包括数据获取、画像生成、智能推荐等。将用户画像技术作为智能图书馆精准服务的关键技术, 能够有效提高图书馆的服务质量及满意度, 可为高校数字图书馆服务模式的转变提供思路和方法。

关键词: 人工智能; 用户画像; 图书馆精准服务; 智能推荐

中图分类号: TP391; G250.7 文献标识码: A 文章编号: 1674-2850(2019)03-0355-08

Research on precise service mode of university digital library based on user portrait

ZHANG Zhiqing^{1,2}, NIE Zhixuan¹

(1. *Evergrande School of Management, Wuhan University of Science and Technology, Wuhan 430081, China;*
2. *Center for Service Science and Engineering, Wuhan University of Science and Technology, Wuhan 430081, China*)

Abstract: The development of network technology promotes the transformation of library service mode from knowledge service to intelligent service. User portrait is a virtual representative of real users and can describe user's characteristics more accurately. Taking user portrait as the core, the framework model of precise service of digital library in colleges and universities is constructed, including data acquisition, portrait generation, intelligent recommendation and so on. Taking the user portrait technology as the key technology of the precise service of the intelligent library can effectively improve the service quality and satisfaction of the library, and provide ideas and methods for the transformation of the service mode of the digital library in colleges and universities.

Key words: artificial intelligence; user portrait; library precision service; intelligent recommendation

0 引言

图书馆是一个搜集、整理、收藏图书资料以供人们阅读和参考的场所, 是一个资源共享的地方。随着网络技术的发展、数字资源的产生和大数据的出现, 信息资源具有数量大、种类多和速度快的特点。因此, 国内外大型数据库的资源急剧增加, 大数据时代的高校图书馆成为了信息的海洋。虽然在表面上大量信息对读者来说是有益的, 但当读者面对庞大的数据库信息时, 往往不能有效地快速检索到需要的信息, 特别是对大学生和老师来说, 这个问题尤其值得关注。因此, 需求的多样性和层次性应运而生,

作者简介: 张志清 (1969—), 男, 教授, 主要研究方向: 信息管理与信息系统、商务智能、数据技术与精准服务、运营管理 (物流与供应链管理、业务流程管理)、项目管理等. E-mail: zhangzhiqing@wust.edu.cn

这就要求高校图书馆提供准确的服务，以满足教师和学生的需要。随着互联网技术的飞速发展，数据挖掘、人工智能等技术也得到了飞速发展，其海量的客户信息数据也越来越受到人们的关注，这些数据反映了图书馆用户日益个性化的需求特征，为图书馆精准服务提供了可能。

1 国内外研究现状

用户画像，又称用户角色^[1]，是对用户进行标识并利用画像技术来确定用户的相关特征：他们的基本状况如何、喜欢哪种产品、喜欢做什么事情，他们的社会交际关系是什么样的，这甚至可以反映出用户的性格。在图书馆环境中，通过对用户早期特征的大量研究，可以将用户画像理解为与用户信息需求相关的关键指标，并对不同用户的典型特征进行建模。准确把握用户信息需求的各个方面的特点，实现读者与资源的准确匹配。

用户画像的概念最早由交互设计之父 ALAN COOPER 提出，其是用户信息的标签化，是真实用户的虚拟代表，是建立在一系列数据之上的目标用户模型^[2]。TRAVIS 认为完整的用户画像需要满足“PERSONA”条件，即基本性（primary）、移情性（empathy）、真实性（realistic）、独特性（singular）、目标性（objectives）、数量（number）和应用性（applicable）^[3]。

用户画像的应用领域十分广泛，在不同的应用领域内，用户画像的构建思路不尽相同，亓丛等^[4]将典型的用户画像思路归结为四大类画像视角：基于虚构的视角（fiction-based perspective）、基于目标导向的视角（goal-directed perspective）、基于角色的视角（role-based perspective）和基于参与的视角（engaging-based perspective）。针对用户画像的构建过程，不同学者拥有不同的看法。AMATO 等^[5]认为，用户画像的构建过程分为两大步：一是明确描述哪些是与用户有关的信息；二是如何有效地描述上述信息。HENCZEL^[6]将用户画像的开发过程分为六步：一是确定需要了解用户的哪些信息；二是识别查找已有数据并判断需补充的数据；三是展开调查；四是分析用户数据特征并生成用户画像；五是评估测试并制定应对变化的方案；六是完善和维护用户画像。在图书馆应用领域，刘速^[7]关注了图书馆用户的访问方式、访问数据库种类的次数、检索的有效关键词、关注的领域等属性，以此进行数据的统计分析，从而提升图书馆的服务质量。

用户画像的构建离不开技术和算法的支持。聚类分析、文本挖掘、关联分析、机器学习等方法得到了广泛的应用。BILLSUS 等^[8]针对用户短期偏好采用最近邻分类算法，针对长期偏好采用朴素贝叶斯分类法。NASRAOUI 等^[9]提出可以利用 Web 站点的日志数据发现用户行为模式，采用 K -均值聚类算法划分用户集群并以此构建集群的用户画像。

2 基于用户画像的高校数字图书馆精准服务的构建

2.1 精准服务模型简介

高校图书馆作为高等院校中较为特殊的服务提供机构，其主要职能是为高校师生提供书籍、文献等知识型数据服务以供高校学生及教职工在日常科研、教学、学习中使用，保证其工作的正常进行。然而，目前高校图书馆在服务整合程度与个性化服务的内容提供方面较为薄弱，尤其是在个性化服务内容方面，高校图书馆针对全校用户普遍提供无差异性的服务，很少甚至不去考虑不同用户之间对信息资源的需求和对数据库资源使用之间的差异。因此，为在高校图书馆个性化服务方面取得突破，构建智能图书馆，并需要对图书馆智能推荐模型进行建立和加以实现。

使用用户画像技术实现图书馆智能推荐是可行的方式之一。基于用户画像的图书馆用户智能推荐应

分为四个主要阶段：相关数据获取、用户画像生成、基于用户画像的智能推荐和评估与修正，如图 1 所示。

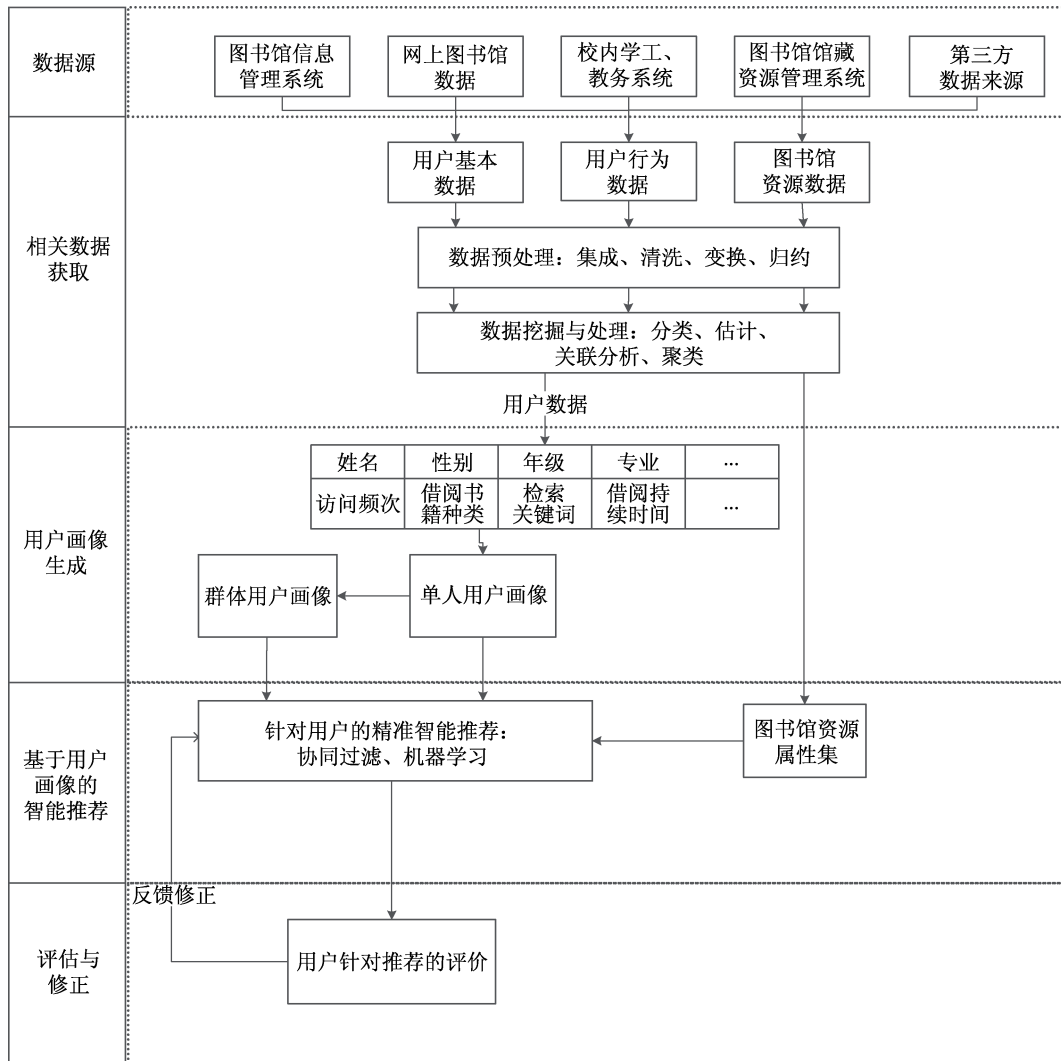


图 1 图书馆精准服务构建流程图

Fig. 1 Flowchart of library precision service construction

实现精准服务，其前提是用户在某些方面拥有较强的差异性，会对产品、服务拥有各自不同的需求，采用构建用户画像的方式对用户进行区分是可行的且能够较为准确地对用户进行分类、预测其需求、提供精准服务、准确构建用户画像，并以此作为推荐服务的基础之一，因此一定数量的数据是必不可少的。需获取的相关数据类型如表 1 所示。

2.2 图书馆用户画像模型的设计

2.2.1 用户画像的属性维度

高校图书馆用户画像模型的构建设计需要根据实际需要，同时还要考虑到对应高校图书馆相关数据的存储种类。通过对已有文献^[10]的研究及与多位图书馆老师的交谈，对本校图书馆信息管理系统的相关数据结构有了一定的了解，结合实际设计需要及相关流程规范，对高校图书馆用户画像模型进行如下设计。

该用户画像模型主要分为三大部分标签维度：用户的基本属性维度、用户的校园属性维度、用户的

表1 用户画像需获取的数据

Tab. 1 Data required for user portraits

一级数据类型	二级数据类型	获取途径	使用方式
基本属性数据	姓名	学工处登记信息	统计分析维度
	性别	一卡通便携携带信息	协同推荐
	年龄（出生年月）	图书馆账户注册信息	
	所在区域		
	用户职业	图书馆账户注册信息	
校园属性数据	用户学历		
	专业课程信息	教务信息	短期兴趣
	研究课题方向		
	院系	学工信息	统计分析维度
	专业	图书馆账户信息	长期兴趣
行为兴趣数据	年级		
	图书馆出入记录	图书馆门禁记录	聚类分析
	书籍借阅记录	借还处记录	短期兴趣
	电子图书馆访问记录	网上用户日志	
	电子资源检索、收藏、评价记录	文本挖掘	
图书馆数据	收藏/下载文献主题信息		
	图书馆实体书信息	图书馆馆藏数据库	推荐内容
	图书馆电子资源数据库信息		
	图书馆相关业务数据	图书馆业务记录	
其他数据	用户日常行为数据	第三方接入记录设备	

行为兴趣维度。用户的基本属性即用户的人口统计学属性，在面向高校图书馆用户的用户画像中，应包含用户性别、用户年龄（出生年月）、用户所在区域、用户职业、用户学历。用户的校园属性即用户因其在高校内作为教职工或学生生活而产生的相关属性，其应该包括用户学号或用户工号（作为图书馆用户的唯一标识）、用户所在专业、用户所在院系、用户所在年级（针对学生用户）、用户的职称（针对教职工用户）、用户主修专业课程。用户的行为兴趣属性指的是用户所感兴趣的图书馆馆藏资源及用户与图书馆相关的行为，其中应包括用户线下访问图书馆频次、用户线上访问图书馆频次、用户线上访问图书馆方式（终端类型）、用户借阅行为频次、用户借阅行为持续时间、用户所感兴趣的图书馆资源特征。其中，图书馆资源属性是基于图书馆文献资源及电子资源的相关属性进行提取的特征，应包括：资源名称、资源作者、发表（出版）年份、出版厂家（发表杂志）、所属领域、被何种数据库所收录。其对应关系如图2、图3所示。

2.2.2 用户画像的标签设计

针对用户画像的三个主要维度，按照其标签维度的稳定性将其进行分类，可以分为稳定标签和非稳定标签。其中，稳定标签指在一定时间内不进行变动（如性别、学号、工号）或拥有规律性变化（如年龄）的标签；非稳定标签指由用户行为数据分析转换而来的标签，其标签值随用户的不同行为有较大变动。因此，使用一个二元组来表示一个用户的用户画像信息，即：用户= \langle 稳定标签，非稳定标签 \rangle 。

稳定标签包含不易变或规律性变化的属性，是一个九维向量，稳定标签向量为：稳定标签= \langle 学号，

性别, 年龄, 所在区域, 职业, 学历, 专业, 院系, 年级>。用以保存用户的静态属性信息。

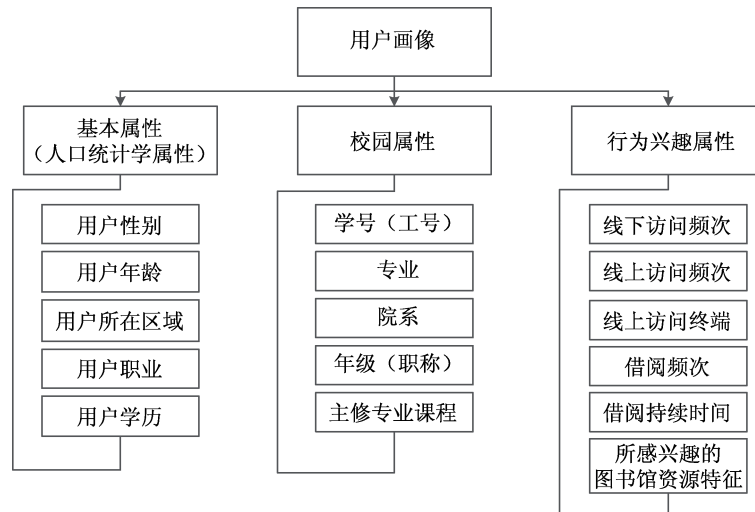


图 2 用户画像包含属性

Fig. 2 Attributes contained in user portraits

非稳定标签是因用户行为时时产生变动的属性, 针对非稳定标签, 采取一个二元组进行表示, 即非稳定标签= \langle 主题, 标签 \rangle , 其中主题= \langle 主题 1, 主题 2, 主题 3, $\dots\rangle$, 标签= \langle 标签 1, 标签 2, 标签 3, $\dots\rangle$, 用以针对诸如用户访问频次、访问终端、查阅书籍及资源等随用户行为动态改变的标签属性。而针对每个标签, 其又为一个二元组, 即: 标签 1= \langle tag, 权重 \rangle , 其中 tag 与权重为多维向量, 用以针对不同需要, 如针对主题 1 “用户访问频次” 所对应的标签 1, 其 tag 向量为二维向量 \langle 线上访问, 线下访问 \rangle , 其权重向量为 $\langle 0.4, 0.6 \rangle$; 针对主题 2 “线上访问终端” 所对应的标签 2, 其对应 tag 向量为四维向量 \langle 移动手机, 移动平板, 个人笔记本电脑, 个人台式机 \rangle , 对应权重向量为 $\langle 0.1, 0.2, 0.5, 0.2 \rangle$ 。“查阅书籍及资源”主题标签根据图书馆资源属性表及用户行为兴趣属性中借阅频次、借阅持续时间进行划分, 其向量维度较多, 因其标签构建思想与前例一致, 不再多加赘述。标签结构如图 4 所示。

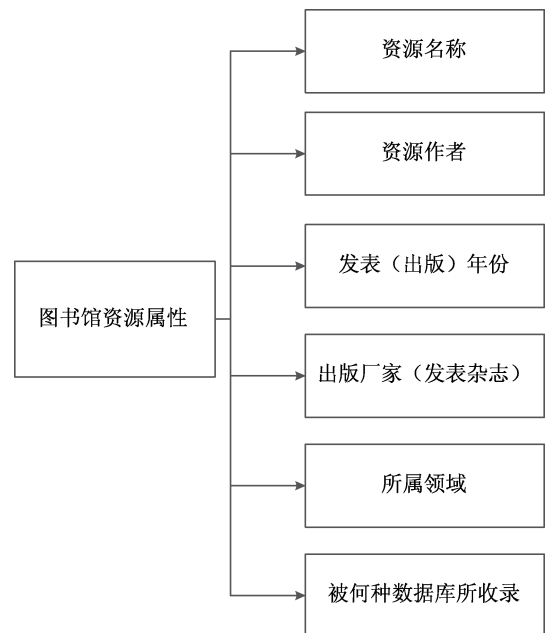


图 3 图书馆资源属性

Fig. 3 Library resource attributes

图书馆用户画像的生成即是对用户画像标签的生成, 具体来说就是稳定标签与非稳定标签的生成。针对稳定标签, 其生成过程十分简单, 仅为已收集相关数据的迁移复制。通过已收集用户的人口统计学属性及校园属性即可构建完成。针对非稳定标签, 其生成过程分为两个主要步骤: 一是主题与 tag 的确定; 二是对应权重的计算。

不同于稳定标签, 非稳定标签的主题与 tag 并非完全固定, 可以通过预设可设想项目来确定一定的主题与 tag。在本文所述面向高校图书馆用户的用户画像中, 针对“用户访问频次”主题, 可预见到的有诸如线下访问实体图书馆、线上访问数字图书馆两个 tag, 可以对此进行预设。而对于部分不可预见的 tag 或 tag 的变动, 如线上图书馆数据库资源的变动 (购买或放弃购买), 使得 tag 维度增多或减少, 可以通

过聚类的方法进行分析，通过对图书馆所有与主题种类相对应的用户行为信息进行聚类，获得其中聚集明显的簇，即为应当设立的 tag 维度。

权重指的是某一个因素相对于整体所占的比重。针对某一个确定的 tag，其维度设定为 n 维向量，即此确定的 tag 中有 n 种可能的用户行为。以 W_i 表示第 i 种用户行为的权重， f_i 表示第 i 种用户行为的频次，其权重 W_i 可表示为 $W_i = \frac{f_i}{\sum f_i}$ 。通过此方法生成权重，权重 W_i 为一个介于 $0 \sim 1$ 之间的小数，其越接近于 0 ，表示用户对于此种用户行为的偏好程度越低；其越接近于 1 ，表明用户对于此种行为的偏好程度越高。与此同时，针对图书、文献资源内容的权重可直接提取并作为读者用户对于各类图书、文献资源特征的偏好矩阵，用以进行后续的智能推荐。

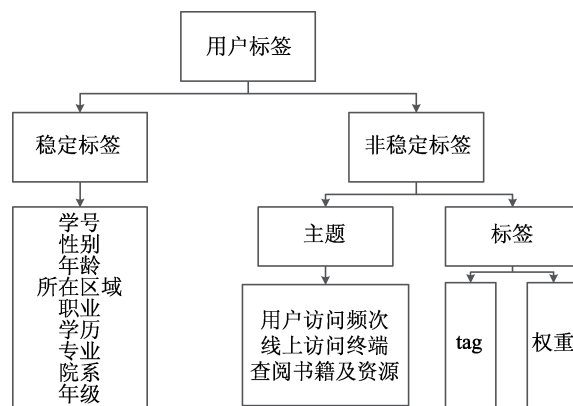


图4 用户标签结构

Fig. 4 User label structure

2.2.3 用户画像的更新设计

用户的行为习惯、兴趣爱好并非一成不变，用户画像在使用过程中也需要进行更新，以此来保证用户画像对用户描述的准确性。针对稳定与非稳定标签，用户画像标签的更新时间无需完全统一。

稳定标签具有较好的稳定性或可预知的规律性变化，其更新周期可依据其规律进行设定，如针对年龄的更新一年一次即可保证数据的准确性。而诸如院系、专业等不易变更但可能会于某些特殊时间段发生的属性改变，可根据实际情况（学校关于专业分流、转换时间安排的相关规定）在时间点与更新周期上进行设置。

针对非稳定标签，不同用户标签的具体内容来自于读者用户的各类用户行为数据。若要使其具有较高的准确性，数据的实时更新无疑是最好的选择，然而受限于客观资源及成本等相关影响因素，使用这种方法进行用户画像标签的更新是不现实的。在生产与库存管理领域中，针对如何在成本最低的前提下保证库存的充足供应已有广泛的讨论与解决方案，其中针对库存管理订货点法的 (t, R, S) 策略是值得借鉴的策略。其基本思想是，库存补给中固定的检查周期为 t ，固定订货水平（即预设数据积累量）为 R ，最大库存量为 S ，每经过一个检查周期 t ，判断当前的存货数量（即用户行为数据积累量） Q ，若 $Q > R$ ，则不订货，否则订货。

将此方法的思想迁移到用户画像更新上，关注检查周期 (t) 、用户行为数据积累量 (Q) 及预设数据积累量 (R) ，进行如下判断：经过周期 t ，对此时用户行为数据（自上次更新起）的积累量 (Q) 进行检查，若 $Q < R$ ，则不对用户画像进行更新；若 $Q \geq R$ ，则对用户画像进行更新。这样可以较好地平衡用户行为数据的积累与画像更新之间的关系。具体思路如图 5 所示，在一个检查周期内，数据积累量随时间为非递减函数，图 5 中的假设为均匀增加，实际非均匀情况在本问题中并无影响。

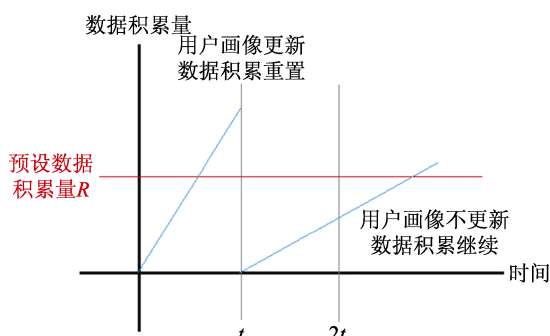


图5 用户画像更新示意图

Fig. 5 Schematic diagram of user portrait update

2.2.4 群体用户画像

除针对某个特定用户的个人用户画像之外，针对由多

个具有相似特征的用户组成的用户群体，如同学院、同专业的用户，其在用户行为上具有较强而普遍的相似性。因此，可以将这些具有相似特征的所有用户看作同一整体，对其进行群体用户画像的构建。

2.3 基于用户画像的智能推荐

用户画像的建模与实现最终是为了实现针对用户的精准服务，因此，使用用户画像并最终实现针对于图书馆用户的精准推荐服务的推荐算法选择便显得尤为重要。智能推荐算法种类繁多，适用的情况更是复杂多样，算法的选择更应结合实际应用的需要。

协同过滤 (collaborative filtering) 最早应用于电子邮件领域，用于对数量大的电子邮件进行筛选分类，并在发展中先后应用在新闻筛选及电子商务推荐系统中。在算法运行过程中，不需要用户提供明确的需求，系统便可以根据用户过去的行为进行预测，向用户推荐其可能感兴趣的内容信息。其算法的核心工作流程包括三个步骤^[1]：收集用户偏好，用户可以以多种方式向系统提供其偏好需求，并不局限于打分、投票、书写评论等直接方式，包括浏览、收藏等行为在内的隐性方式也可以作为偏好表达的一种；寻找最近的邻居，为目标用户寻找最近的邻居；预测评分，根据寻找的邻居集合，计算目标用户可能会基于各个目标项目的评分，在此预测评分的基础之上，向目标用户提供推荐。具体流程如图 6 所示。

基于此，使用协同过滤的推荐方法是可行的方法之一。同时，由于用户（教职工与学生）和项目（书籍与文献资源）均有相似集合的划分，不论是以用户为基础的协同过滤算法（user-based collaborative filtering algorithms），还是以项目为基础的协同过滤算法（item-based collaborative filtering algorithms），均可根据实际需要做出选择。

3 结论

随着图书馆由知识服务逐渐转化为智慧服务，大数据技术在各行业领域内的应用不断扩展，用户对于图书馆提供服务的精准性要求越来越高。本文先对目前图书馆精准服务所面临的新要求进行了论述，提出使用用户画像技术来实现在高校图书馆领域针对图书馆用户实施精准推荐是可行的。随后分析高校图书馆实现精准服务的需求与目的，构建高校图书馆精准服务模型，包括相关数据获取、用户画像生成、基于用户画像的智能推荐和评估与修正四个阶段。并对高校图书馆用户画像进行设计，包含人口统计学属性、校园属性、行为兴趣属性三个维度，并依照其属性稳定性分为稳定标签与非稳定标签，同时结合武汉科技大学图书馆相关数据对其相关非稳定标签进行生成。随后参考库存订货管理的思想提出用户画像更新的周期设计，最后对推荐系统的相关算法进行选择，以提升大数据时代图书馆服务保障力，为读者提供满意的个性化智慧服务，为高校数字图书馆服务模式的转变提供思路和方法。

[参考文献] (References)

- [1] 柳益君, 何胜, 熊太纯. 大数据挖掘视角下的图书馆智慧服务——模型、技术和服务[J]. 现代情报, 2017, 37(11): 81-86.
LIU Y J, HE S, XIONG T C. The smart service of library from the perspective of big data mining——model, technology and

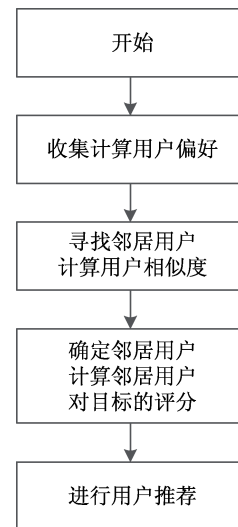


图 6 协同过滤流程图

Fig. 6 Flowchart of collaborative filtering

- service[J]. *Modern Information*, 2017, 37(11): 81-86. (in Chinese)
- [2] 曾鸿, 吴苏倪. 基于微博的大数据用户画像与精准营销[J]. *现代经济信息*, 2016 (16): 306-308.
ZENG H, WU S N. User image and precision marketing on account of big data in Weibo[J]. *Modern Economic Information*, 2016(16): 306-308. (in Chinese)
- [3] 陈慧香, 邵波. 国外图书馆领域用户画像的研究现状及启示[J]. *图书馆学研究*, 2017 (20): 16-20.
CHEN H X, SHAO B. Research status and enlightenment of user portraits in foreign libraries[J]. *Research on Library Science*, 2017(20): 16-20. (in Chinese)
- [4] 亓丛, 吴俊. 用户画像概念溯源与应用场景研究[J]. *重庆交通大学学报(社会科学版)*, 2017, 17 (5): 82-87.
QI C, WU J. Research on the conceptual origin and application of persona[J]. *Journal of Chongqing Jiaotong University (Social Science Edition)*, 2017, 17(5): 82-87. (in Chinese)
- [5] AMATO G, STRACCIA U. User profile modeling and applications to digital libraries[C]//*Proceedings of the Third European Conference on Research and Advanced Technology for Digital Libraries*. London: Springer-Verlag, 1999: 184-197.
- [6] HENCZEL S. Creating user profiles to improve information quality[J]. *Online*, 2004, 28(3): 30-33.
- [7] 刘速. 浅议数字图书馆知识发现系统中的用户画像——以天津图书馆为例[J]. *图书馆理论与实践*, 2017(6): 103-106.
LIU S. The persona in digital library knowledge discovery system——taking Tianjin library as an example[J]. *Library Theory and Practice*, 2017(6): 103-106. (in Chinese)
- [8] BILLSUS D, PAZZANI M J. A hybrid user model for news story classification[C]//*Proceedings of the Seventh International Conference on User Modeling*. New York: Springer-Verlag, 1999: 99-108.
- [9] NASRAOUI O, SOLIMAN M, SAKA E, et al. A web usage mining framework for mining evolving user profiles in dynamic web sites[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2008, 20(2): 202-215.
- [10] 白丽荣. 图书馆电子资源的采购策略[J]. *图书馆学刊*, 2013 (7): 43-45.
BAI L R. Procurement strategy of library electronic resources[J]. *Journal of Library Science*, 2013(7): 43-45. (in Chinese)
- [11] 翁小兰, 王志坚. 协同过滤推荐算法研究进展[J]. *计算机工程与应用*, 2018, 54 (1): 25-31.
WENG X L, WANG Z J. Research progress of collaborative filtering recommendation algorithm[J]. *Computer Engineering and Applications*, 2018, 54(1): 25-31. (in Chinese)

(责任编辑: 李曦)