

基于用户群体画像分析的慕课平台 知识服务策略研究^{*}

刁雪桦^{1,2} 朱学芳^{1,2}

(1. 南京大学信息管理学院, 南京 210023; 2. 南京大学江苏省数据工程与知识服务重点实验室, 南京 210023)

摘要: 慕课网站作为重要的在线学习平台, 存放大量知识服务资源, 也是知识服务落地的重要场景, 属于知识服务体系的一部分。研究当前慕课用户的群体结构与特征既有利于提高平台服务质量, 也有助于丰富知识服务渠道, 从而更好地满足知识服务用户多样化的需求。聚焦慕课场景的知识服务, 从用户视角出发, 以中国大学MOOC平台的用户为样本, 应用K-means++聚类算法进行用户群体画像与分析, 把慕课用户划分为沉默型用户、积极型用户、需求型用户、潜力型用户4种类型。基于画像结果, 提出开展慕课场景知识服务的优化策略, 以期提升慕课服务的用户满意度, 完善相应的智慧知识服务体系建设。

关键词: 用户画像; 慕课; 知识服务; 聚类

中图分类号: G312 **DOI:** 10.3772/j.issn.1673-2286.2023.12.002

引文格式: 刁雪桦, 朱学芳. 基于用户群体画像分析的慕课平台知识服务策略研究[J]. 数字图书馆论坛, 2023 (12): 11-20.

慕课 (Massive Open Online Courses, MOOC), 即大型开放式在线课程, 该概念于2008年被首次提出, 2012年被称为“慕课元年”^[1], 此后慕课服务在国内外迅速发展。慕课的发展推动了高等教育模式的变革, 反映了社会对于教育模式与环境智慧化的期望^[2]。

在国际方面, 美国依托其众多知名高校建设的国际慕课平台引起了全球广泛关注, 如edX、Coursera、Udacity等知名网站; 欧洲则凭借欧洲一体化相关的政策优势, 于慕课发展初期就开启了多国合作, 共建慕课平台, 采用资源共享的发展模式。黄红兵等^[3]认为目前国际慕课平台已形成较为成熟的运行和服务机制。目前, 我国的高校信息化教育服务以结合慕课和小规模在线课程 (Small Private Online Courses, SPOC) 的混合模式为主^[4]。教育部2022年上线“国家高等教育智

慧教育平台”, 同时通过“iCourse”和“xuetangX”两个中国高校在线教学国际平台推动我国慕课的国际化交流^[5]。

尽管从课程数量和平台建设方面看, 我国的慕课资源建设已经取得一定成效, 但其资源利用率仍有很大的提升空间, 其服务模式仍然存在发展与创新空间。慕课网站作为重要的在线学习平台, 不仅汇集了大量知识服务资源, 也是知识服务落地的场景之一, 属于知识服务体系的一部分。智慧知识服务体系始终以用户的需求为导向^[6], 因此本文从慕课用户的视角展开, 应用近年来在智慧信息服务领域较热门的用户画像方法, 以中国大学MOOC平台的用户为研究样本, 分析慕课用户的构成、特征, 以及他们的需求, 探讨慕课平台如何更好地契合不同群体的特点, 以期提升服务质量与服务效果, 完善相应的智慧知识服务体系建设。

收稿日期: 2023-10-07

^{*}本研究得到国家社会科学基金项目“5G环境下中国智慧知识服务体系构建研究”(编号: 22BTQ017)资助。

1 相关研究

1.1 基于慕课场景的知识服务研究

在慕课服务不断发展、信息技术持续革新的情况下,学者们针对慕课平台设计和课程资源建设情况展开了研究,以期提升慕课平台的智慧化程度,完善慕课资源建设。李芳慧等^[7]认为“互联网+”时代的技术发展拓展了知识服务平台的服务渠道和运营模式,并归纳总结了果壳网MOOC学院等优质知识服务平台的运营策略。曹雅宁等^[8]基于劝导技术理论和劝导系统设计模型,探讨当前我国主流慕课平台的劝导设计特征和效果,针对慕课平台的交互系统提出了优化建议。Stefan^[9]比较与分析了多个慕课平台现有的信息素养课程内容和教学设计,揭示基于慕课平台的信息素养教学的现状。邵杨芳^[10]通过对慕课用户评论的文本分析,探究了知识服务用户对健康教育类在线课程的需求和使用评价。

慕课服务领域内引发学者广泛关注的一个问题是超高的中断学习率^[11],因此部分学者围绕慕课服务效果的影响因素和提升用户留存率的方法展开了研究。虞鑫等^[12]从创新扩散视角出发,对慕课用户的活跃度和满意度的影响因素进行了研究。和文斌等^[13]基于复杂理论,结合技术接受模型和社会认知理论,对慕课用户开展问卷调查,验证了使用者自我效能感、感知有用性、感知易用性、社群影响、游戏化设计以及内容质量对用户留存意愿的影响。此外,部分学者基于对慕课平台用户和课程的分析,构建能够对中断学习行为进行预警的算法模型,从而帮助平台提前对有中断学习倾向的用户进行干预,例如:陈辉等^[14]提出基于长短时记忆机和多头注意力机制的用户学习情况预测融合框架;魏玲等^[15]基于RFM模型和数据分组处理网络,提出慕课用户流失预测模型。

慕课服务作为知识服务体系的一部分,与智慧图书馆建设联系紧密,部分学者探讨了慕课平台和图书馆在知识服务过程中的角色和合作模式。早在2014年,秦鸿^[16]就指出图书馆可以深度参与慕课服务。赵莉娜^[17]总结了慕课资源与用户群体的特征,指出高校图书馆可以深入慕课场景,充分利用自身资源,在教育与科研两个方向开展嵌入式信息服务,完善知识服务功能。李佳玉^[18]结合我国高校图书馆基于慕课的知识服务实践现状,提出图书馆应该提供个性化、知识化与合作式

的深度嵌入服务,构建全方位满足用户需求的嵌入式知识服务模型。聂慧^[19]以高校图书馆的阅读推广服务为基点,界定了基于慕课平台的嵌入式阅读推广服务的内容。

1.2 知识服务用户画像研究

用户画像能够提炼用户的特征、发现用户的需求,从而实现对用户的个性化服务。对相似用户进行聚类 and 群体画像,可以发现服务用户的一些隐藏特征,实现用户分群,制定有针对性的服务策略。在知识服务领域的相关研究中,许多学者已经引入用户画像的思想或方法。徐海玲等^[20]基于概念格的方法,构建了高校图书馆的本科生、研究生、教工3类群体的兴趣画像,挖掘并分析了不同群体的行为规律。武慧娟等^[21]基于用户的阅读习惯和付费行为,应用聚类分析方法对数字阅读用户进行了群体画像,探究了用户支付意愿的影响因素。王大阜等^[22]基于读者的基本属性、行为标签、兴趣标签构建了高校图书馆用户画像,采用了基于内容过滤和协同过滤的混合推荐算法向用户进行图书推荐。李宇佳等^[23]提出了在新媒体环境下构建学术用户动态画像的方法,从而实现新媒体信息与资源的智能推荐和智慧服务。Shahbazi等^[24]应用了自然语言处理技术和语义分析方法,根据用户的个人信息、偏好、搜索历史构建用户画像,向在线学习用户和学者推荐匹配的内容和提供课程选择建议。杨凡^[25]以澳洲八校集团成员图书馆为调研样本,提出我国高校图书馆可以利用慕课平台的用户数据构造用户个人画像和用户群体标签,以向用户提供精准、智慧的嵌入式知识服务。

2 研究对象与研究思路

戎军涛等^[26]把知识中介定义为知识生产者和知识获取者的连接桥梁,知识中介活动就是知识中介应用科学手段对知识传递、流通过程进行控制和管理的过程。得到、喜马拉雅、中国大学MOOC、丁香医生等互联网知识服务平台通过文献、音频、视频、咨询等丰富多样的形式完成知识中介活动,为用户提供知识服务,成为知识服务体系的重要组成部分。其中,慕课平台汇聚来自高等教育学校和专业知识机构的课程资源,通过视频课程、课堂讨论、课后作业等形式满足用户的知识需求,使用户可以在任何时间、任何地点获得以往在

线下授课中才能获取的知识服务。慕课平台能够提供广泛、稳定、可靠的知识服务,在知识服务场景中占据重要地位,但其资源利用率和服务效果还有很大提升空间。慕课平台的服务对象是具有知识需求的用户,属于知识服务体系的用户群体之一。本文以慕课平台及其用户为主要研究对象,探讨慕课平台提升知识服务效果和满意度的策略,以期完善知识服务体系建设。

根据张晓林^[27]的“日心说”,知识服务要始终以用户需求为导向。因此,要探索基于慕课场景的知识服务改进策略,首先要对用户有充分的认识,把握慕课场景中的知识服务用户需求。以往的许多研究也强调了慕课用户,常用的用户研究方法包括问卷调查、访谈等。问卷调查和访谈等方法更重视对用户需求的深度分析,但样本量有限,可能受受访者的主观因素干扰;且其研究对象往往是较常使用慕课服务的用户,而慕课平台整体用户的构成和低活跃度用户的需求较难被关注。因此应用用户画像方法,根据慕课用户的客观行为数据进行用户分群,提升慕课平台对用户群体的认识,根据用户特点提出有针对性的服务策略,帮助慕课平台完善其知识服务模式。中国大学MOOC平台是由网易与高等教育出版社合作推出的在线教育平台,是目前国内成熟、主流的慕课平台之一。其提供丰富课程,总选课人数最多^[28],具备代表性。为此,以中国大学MOOC平台用户为研究对象,进行用户群体画像分析,其结果可以认作我国慕课平台用户的特征,并基于此讨论慕课平台的服务改进路径,提升慕课场景的知识服务质量,从而完善相应的智慧知识服务体系建设。

3 中国大学MOOC平台用户群体画像分析

3.1 用户数据采集

为了保障选取的用户样本具备代表性和全面性,首先从中国大学MOOC平台的国家精品、认证学习、计算机、外语、理学工学农学、考研、期末突击、四六级、专升本、求职就业、经济管理、文史哲法、心理学、教育教学、音乐与艺术、有声课堂,共16个板块的热门课程的评论中爬取16 514位平台用户的个人空间链接。然后在分析平台用户的个人空间公开信息后,将用户信息分为三大类:基本数据、社交数据、学习数据。最后使用Python语言编写爬虫程序,获取每位样本用户的信息

(见表1)。

表1 采集的平台用户信息

数据属性	数据名称	数据详情
基本数据	用户标签	用户职业和具体从事领域
	用户简介	用户自定义的个性签名或介绍
社交数据	关注人数	用户在平台内关注的用户数量
	粉丝人数	用户在平台内的粉丝数量
	获赞数量	用户发表主题或回复获得的总点赞数
学习数据	发表内容数	发表主题和回复的总数量
	学习时长	用户累计学习时长(单位为小时)
	已选课程数量	用户选修过的课程总数
	已选课程名称	用户选修的所有课程名称
	优秀证书数量	用户获得了优秀证书的课程总数
	优秀证书名称	用户获得了优秀证书的课程名称
	合格证书数量	用户获得了合格证书的课程总数
	合格证书名称	用户获得了合格证书的课程名称

用户基本数据包括用户标签和用户简介,均为用户自定义的描述性数据。其中用户标签一项包含了用户的职业属性。中国大学MOOC平台向用户提供了学生、老师、在职、其他共计4个可选项:“老师”通常是职业为高校教师的用户;职业标签为“其他”的用户则包括了已退休、自由职业等情况。部分用户会提供更详细的信息,例如所在学校院系、从事行业等。用户简介一项则更为个性化,一定程度上反映了用户的自我展示意愿,不过该项空值率较高(85.1%)。

用户社交数据包括关注人数、粉丝人数、获赞数量。事实上,中国大学MOOC平台的社交属性与功能较弱,不能给用户额外的社交体验,例如私信、发布动态提醒等功能。不过该类数据能在一定程度上反映用户是否关心平台的任课教师、共同学习者,是否有与其他用户产生联系、交流的意愿,因此能在一定程度上反映平台用户黏性,并且能够为平台建设用户社区、开发含社交属性的服务提供参考。

学习数据是平台的核心数据,也是信息量最大的数据,包括用户的发表内容数、学习时长、已选课程数量/名称、优秀证书数量/名称以及合格证书数量/名称。经过对部分样本的分析可以看出,在课程中发表讨论内容通常是课程要求的一部分。相较于反映用户的社交倾向,上述数据更能够反映用户参与课程的积极性,因此将其归类为学习数据。学习时长和已选课程能够直观地反映用户使用平台进行在线学习的频率和需求程度。只有当学生全程参与课程,并达到老师设置的考核标准,才能获取证书,因此已获得证书的信息能够反映用户在某门课程中的学习积极性、投入程度

和学习质量。

3.2 数据预处理

通过对平台用户数据进行聚类分析,可以构建用户群体画像,进而分析慕课用户的特征和需求。在进行聚类之前需要对采集的用户数据进行预处理,从而便于应用聚类算法计算,预处理的过程如下。

首先对数据进行清洗,删除采集到的异常样本,以便使聚类结果更加稳定。以下情况均视为异常样本:①由爬取过程中网页的不稳定性导致的大部分属性为空值的样本;②学习时长为0但显示已获得证书的样本;③已选课程数量为0但学习时长非0的样本,推测这是因为学生学习了已撤销课程。排除异常样本后,剩余总样本数为16 190个。

然后进行聚类特征选择,挑选更能反映用户差异、学习特点、使用习惯的特征进行分析。在基本属性方面,直接删除空值率过高的用户简介项。用户标签划分为用户职业、用户所属细分领域两部分,其中:细分领域部分空值率为72.3%,也予以删除;用户职业予以保留,但其分类方式不便转换为数值型变量,且分布比例不均衡,因此不参与聚类分析,但会用于对画像结果的分析。社交属性方面的3项数据均予以保留。此外,考虑到研究主要对用户的群体画像进行分析,探究慕课用户的群体特征,暂不关注用户个人的学习兴趣领域特征,因此在学习属性中排除用户已选课程名称、优秀/合格证书名称等数据,仅保留数值型数据:发表内容数、学习时长、已选课程数量、优秀证书数量、合格证书数量。由于用户获取证书数量普遍较少,在所有样本中合格、优秀证书数量两项数据的区分度不大,把两项相加合并为用户获得证书数量。

由于中国大学MOOC平台并不在用户的个人空间公开展示每位用户每门课程的完成情况,构造两个新指标,用于反映用户的学习完成情况和积极性,分别为用户参与课程的平均学习时长和证书获得率:平均学习时长为学习时长与已选课程数量之比;证书获得率为获得证书数量与已选课程数量之比。最终获得9项数据用于聚类分析。

3.3 用户聚类过程

本次聚类分析实验中使用的9项数据均为数值型特

征,但各个特征的分布范围差异较大,存在数量级的差异,直接进行聚类分析会导致部分数值较大的特征权重过大,影响聚类效果,因此在聚类前要对所有特征数据进行归一化处理。由于样本中的用户数据分布较为集中,使用最大最小标准化方法进行归一化,把每个特征下的数据转换到[0, 1],接下来就可以进行聚类分析。

聚类是一种无监督学习算法,目的是将一组数据点划分为簇,使簇内相似度最大、簇间相似度最小。使用的聚类分析方法为划分式聚类算法K-means++^[29],该算法是对经典的K-means聚类算法的优化,需要提前确定聚类的簇数 k ,其具体流程如下。

(1) 初始化聚类中心点,随机选取一个样本 u 作为聚类中心。

①当聚类中心数小于 k 时,计算每个样本与当前已有的每个聚类中心之间的距离,记其中最短的为每个样本的 $D(u)$;

②选择 $D(u)$ 最大的一个样本作为新的中心,重复步骤①直到选出 k 个中心。

(2) 计算每个样本到每个聚类中心点的距离,将样本 u 划分到距离其最近的中心点对应的簇上。

(3) 所有样本分类完毕后,重新计算每个簇的质心,将其作为新的聚类中心点。

(4) 重复步骤(2)和(3),直至聚类中心点不再移动。

其中,采用欧氏距离计算样本之间的距离。上述聚类过程使用Python语言编程实现,为了确定最佳聚类簇数 k ,使用了肘部法则和轮廓系数^[30]结合的方法。

误差平方和(Sum of the Squared Errors, SSE)是一个常用的聚类效果评价指标,其值为每个样本点与自己所属簇中心的距离的平方和,该指标可以反映每个簇的样本聚合程度。随着 k 值的增大,SSE会不断下降,但当 k 的取值超过最合适值时,下降速度会变缓。因此肘部法则认为,做出SSE随 k 值的变化图后,曲线拐点处的 k 值即为最佳取值。

肘部法则的局限性在于仅考虑了聚类簇内的聚合度,为了确定 k 的最佳取值还需要综合考虑簇间不相似度和簇内相似度的轮廓系数。对于样本 u ,计算其与同簇其他样本的平均距离 a ,计算其与最近的相邻簇内所有样本的平均距离 b ,则其轮廓系数如式(1)所示。

$$s_u = \frac{b-a}{f_{\max}(a, b)} \quad (1)$$

式中: $f_{\max}$ 表示取最大值函数。所有样本的 s_u 的平均

值则为本次聚类的轮廓系数,取值范围为 $[-1, 1]$ 。轮廓系数越大说明簇间分离程度越高、簇内聚合程度越高,即聚类效果越好。

设 k 值为2~15,分别进行聚类分析,计算每次聚类的 SSE 值和轮廓系数,结果如图1、图2所示。综合两个图可以看出,当 k 取值为4时, SSE 值下降趋势明显减缓,此时轮廓系数也维持在较高的值0.785,因此最终选择 $k=4$ 。聚类完成后,4类用户的簇中心特征如表2所示。

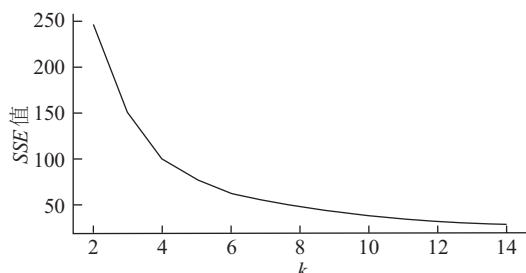


图1 聚类分析的肘部法则图

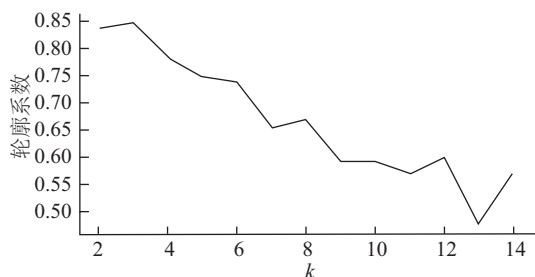


图2 聚类分析的轮廓系数图

表2 聚类后各个簇中心的特征

特征项目	簇编号			
	0	1	2	3
关注人数	0.740	2.393	25.678	1.094
粉丝人数	0.105	2.085	4.296	0.627
发表内容数	20.310	171.891	292.727	108.883
获赞数量	7.298	113.891	165.655	54.682
学习时长	79.068	105.253	611.793	97.737
已选课程数量	28.974	13.748	633.551	15.867
获得证书数量	0.278	10.783	14.393	5.001
平均学习时长	6.108	10.258	0.924	7.452
证书获得率/%	0.893	87.967	2.171	33.723

注:簇中心值为该簇所有样本在各个特征内数据的平均值,该数值能够反映各类用户的平均表现,但并不指向其中任一实际存在的用户。

3.4 画像结果分析

根据聚类结果,平台用户可以被划分为4个类型,根据其9个特征的表现,按照簇编号分别命名为沉默型

用户、积极型用户、需求型用户、潜力型用户,并且统计了所有样本中各个类型用户的占比。

(1) 沉默型用户。沉默型用户对应表2中簇编号为0的用户,在所有样本中占比最大,达到89.07%,这说明平台内存在大量的沉默用户。该类用户虽然有一定的已选课程数量、学习时长,并且平均学习时长中等,但关注人数和粉丝人数都很少,发表内容数很少,极少获得证书。由此可见他们虽然有使用慕课资源的需求,但使用习惯以观看视频学习为主,参与课程讨论的积极性低,社交欲望较低,不太关心是否获得证书这类学习成就。虽然他们并非平台的核心用户,但该类用户的庞大占比能够暴露出当前慕课平台存在知识资源传播完成率不高的问题,也说明当前慕课平台内存在大量有一定知识需求的冷却用户,这部分用户具有相当强的发展潜力。

(2) 积极型用户。积极型用户对应表2中簇编号为1的用户,在所有样本中仅占2.67%。该类用户虽然相对其他用户已选课程数量较少,但平均学习时长最长、证书获得率最高,簇中心点的证书获得率为87.967%。可见他们是平台中参与课程学习积极性最高的一类用户,学习目的明确,有良好的自主学习能力和自我驱动力。此外,这类用户的关注人数和粉丝人数基本持平,可见他们有一定的社交倾向,也具备对其他用户的社交吸引力。综上所述,这类用户是慕课平台中资源利用率最高,并且具备潜在社区号召力的核心优质用户。

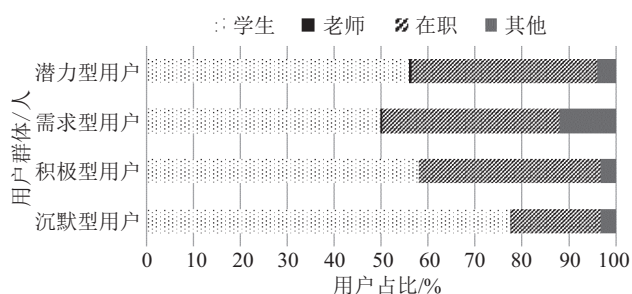
(3) 需求型用户。需求型用户对应表2中簇编号为2的用户,在所有样本中占比最少,仅占1.65%。该类用户的关注人数、粉丝人数、发表内容数、已选课程数量、学习时长、获得证书数量最多,可见他们是平台内使用时间最长、学习需求最大的用户。但相对于他们选修的课程数来说,这类用户的讨论参与度较低,证书获得率较低,仅为2.171%,而且他们的平均学习时长在4类用户中最短,可见他们存在已选课但未开始学习的情况,中断学习情况较多,学习行动力一般。综上所述,这类用户是慕课服务的长期用户,对慕课学习资源有很高的需求,但他们的学习积极性仅体现在小部分课程中。这类用户对慕课资源的需求较强,是慕课平台的核心长期用户,但他们的资源利用率并不高。

(4) 潜力型用户。潜力型用户对应表2中簇编号为3的用户,在所有样本中占6.61%,是除沉默型用户外数量最多的用户。该类用户的关注人数与粉丝人数较少、已选课程数量较少、学习时长中等、平均学习时长较

长、证书获得率中等。可见, 尽管他们目前对平台资源的需求程度不高, 社交欲望一般, 但其学习参与度仅次于积极型用户, 优于沉默型用户和需求型用户, 具备一定的学习积极性。综上所述, 这类用户使用平台的学习目的较明确, 学习习惯良好, 是平台中最有潜力发展为核心用户的一批用户。

3.5 4类用户群体比较

各个类型用户的职业分布如图3所示, 平台内的用户存在分布不均衡的情况。4类用户占比差异大, 用户以学生群体为主, 其他群体用户量较少。其中: 沉默型用户中的学生用户占比明显多于另外3类用户, 说明总体上学生用户的参与积极性和学习主动性略低于其他用



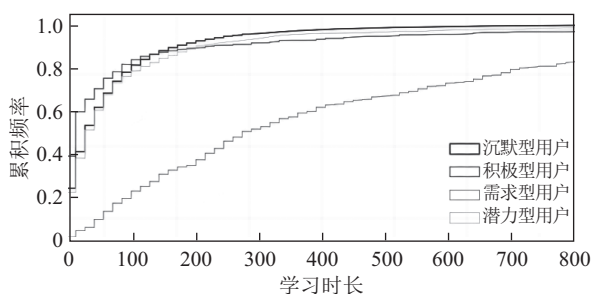
注: 柱形内数字为相应用户数量(人)。

图3 4类用户职业构成

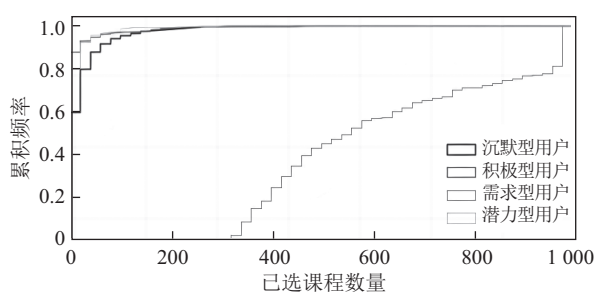
户群体, 可能相当一部分用户的使用目的仅是完成学校任务; 需求型用户中约有50%为学生用户, 该比例低于另外3类用户, 这也能说明需求型用户通常资历较深, 有终身学习的倾向。

为了更直观地比较4类用户在不同特征下的表现以及突出他们的差异, 选取了学习时长、已选课程数量、平均学习时长、证书获得率4个特征, 采用直方图的形式做出所有用户在各个特征下的累积频率图。对于每一项特征, 在所有样本的最小值和最大值区间内均匀划分出多个小区间, 分别统计各类用户样本在每个小区间内的出现频率, 再和小于该区间的所有区间的样本频率进行累加, 获得该区间的累积频率值。以横坐标表示各个特征对应的数值, 纵坐标表示用户在该特征的每个数值区间对应的累积频率, 结果如图4所示。该阶梯状图线可以反映每类用户在单个特征中的分布情况, 图线增长快的地方说明该区间用户分布集中, 增长平缓的地方说明该区间用户分布少。由于部分类型用户在已选课程数量、学习时长两个特征中的分布差异不明显, 又做出4类用户已选课程数量与学习时长特征散点图, 如图5所示。

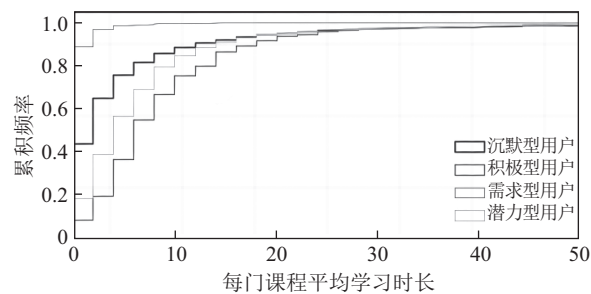
综合图4、图5可见: 沉默型用户、积极型用户、潜在型用户3类群体中均有80%以上的用户学习时长在200小时以内, 需求型用户的学习时长分布则较均衡, 没有出现集中在某段的情况; 学习时长在100小时内的沉默型用户和潜在型用户分布情况非常接近, 而在学习时



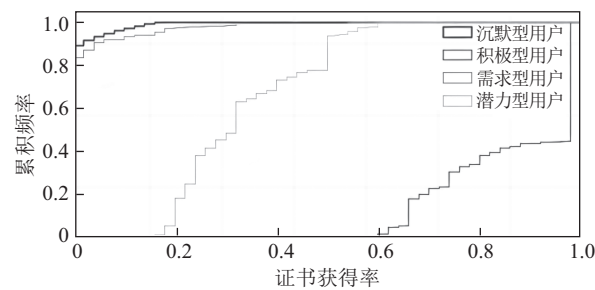
(a) 用户学习时长



(b) 用户已选课程数量



(c) 用户平均学习时长



(d) 用户证书获得率

图4 4类用户部分特征的累积频率图

长超过100小时的用户中, 尽管沉默型用户的规模更庞大, 但其平均表现不如潜力型用户和积极型用户。在已选课程数量方面, 需求型用户表现最突出, 所有用户的选课数量均在300门以上; 积极型用户、潜力型用户的分布情况非常相似, 沉默型用户的平均选课数量较前两类用户更多。

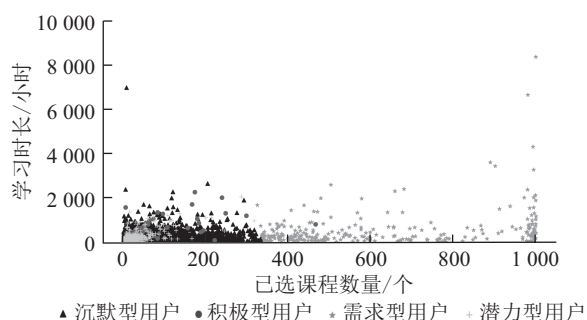


图5 4类用户已选课程数量与学习时长特征

平均学习时长方面的分布情况则略有不同, 4类用户区分较明显, 积极型用户表现最佳, 其后依次为潜力型用户、沉默型用户、需求型用户。在证书获得率方面, 沉默型用户和需求型用户分布接近; 而积极型用户表现最好, 证书获得率在60%以上, 并且有超过50%的用户证书获得率接近100%, 说明这类用户在使用慕课服务的过程中有很强的自我驱动力和学习积极性; 潜力型用户的证书获得率则居中, 大部分为20%~60%, 这类用户的学习积极性尚有提升空间。

综上所述, 沉默型用户代表的是大体量的待激活用户, 优质的慕课平台建设和服务有助于这类用户向潜力型用户或积极型用户发展; 但如果慕课平台只一味鼓励用户选课, 而不维持用户学习积极性的话, 可能会使这3类用户转化为需求型用户, 从而弱化知识服务的实际效果, 这不利于慕课平台发挥作用。

4 基于慕课场景的知识服务策略

根据前文对中国大学MOOC平台的用户群体画像, 可以看出当前国内慕课平台有大量低活跃度的一般用户, 课程的学习完成度低, 大量资源未被充分利用, 用户黏性低。要完善当前慕课平台建设, 提升服务智慧化程度和服务质量, 平台需要根据用户画像开展更个性化、更精准的服务。

针对大量待激活的沉默型用户, 慕课平台需要提升新用户和正在使用用户的留存率, 满足他们对知识资

源和知识服务的需求; 针对已流失的用户, 慕课平台需要采取有效策略促进用户回流。积极型用户是慕课平台的核心用户, 慕课平台需要促进其他用户向该类用户转化, 从而提升核心优质用户的占比, 这不仅能提升平台服务的成效, 还能促进平台良性运转。同时, 平台也要关注积极型用户的使用动向和满意度, 保障优质用户的留存率。针对需求型用户, 慕课平台需要提升资源推荐的准确率, 提供一些提升学习积极性的帮助, 降低中断学习的概率。对于潜力型用户, 慕课平台应当充分发掘他们的学习需求和兴趣领域, 依托智慧化的知识服务, 提供匹配他们需求的资源, 并且帮助他们进一步提升学习的参与度, 培养他们的长期使用习惯, 促进他们成长为积极型用户。

如果慕课平台能够向用户提供更科学、更体系化、更个性化、更智慧化的知识服务, 就能更好地满足用户的需求, 提升用户活跃度和积极性, 促进一般用户向优质的核心用户转化。基于慕课用户群体的特征、慕课服务的环节与流程, 可以从以下几个方面推进当前慕课场景的知识服务模式改进。

4.1 加快资源开发与更新, 促进知识服务体系化

在慕课资源建设环节, 平台需要紧跟用户需求, 进行资源开发与更新, 保障所提供知识的前沿性, 促进知识服务体系化。慕课平台中依然还存在现有资源不足以满足用户需求的情况, 这也是慕课平台出现大量沉默型用户的原因之一。

慕课平台上分布着很多主题相似, 但由于制作机构不同, 在难易程度、知识侧重点等方面存在差异的课程。此外, 各个学科、专业内的课程分布存在不均衡的现象, 例如: 高等数学这类基础课程繁杂, 但一些冷门专业提供的可选课程则较少。完善的知识服务的基础是体系化的慕课资源, 课程资源应尽可能地覆盖各个学科领域, 既要有适合初学者的通识课程, 也要有适合研究人员等专业用户的专业课程。为此, 作为知识服务中介的慕课平台需要根据用户需求加强与知识生产机构的交流和合作, 协调课程生产过程中的资源分配。一方面, 引导课程生产机构及时更新课程, 保障知识资源的时效性; 另一方面, 鼓励打造差异化的课程, 使平台的信息资源在知识侧重点和难度上形成差异, 以满足具有不同需求和知识水平的用户, 同时也能减少资源的浪费。此外,

慕课平台可以梳理现有知识资源的分布情况,构建慕课资源的知识图谱,为平台知识资源建设提供参考,从而有助于形成完整的知识体系,帮助用户系统地获得知识。

知识流动加速的时代,知识日新月异。只有不断更新知识服务的资源,形成完整的资源体系,才能更好地满足用户的需求,减少由资源落后和缺失导致的用户流失情况。

4.2 力求资源推荐智慧化,不断提升用户满意度

在慕课资源分配环节,慕课平台的目标是推进资源推荐智慧化,不断提升用户满意度。部分沉默型用户的学习参与度不高,这可能是由于课程资源并不完全匹配用户需求,用户深度完成学习的欲望较低。需求型用户虽然有一定的学习行动力,但也有大量课程未完成学习,这可能是由于在长期使用的过程中,用户关注领域不断变化。

个性化推荐已经在慕课平台中得到应用,但由于知识资源的特殊性,推荐的智慧化程度还有提升空间。用户注册时填写的用户标签、兴趣领域等静态信息反映的用户兴趣较为有限,并且可能会随着用户的成长而失效,却往往不能得到及时更新。因此智慧资源推荐系统需要敏锐捕捉用户兴趣的变化^[31],关注用户的兴趣漂移情况,给用户推荐最符合当下需求的资源。此外,智慧推荐系统中,对知识资源的画像也同样重要^[32]。智慧推荐系统不仅要及时、精准地发现用户兴趣,还应充分考虑用户的知识水平。在高校中,学习的过程往往是一个由浅到深、由通识到专业的过程。如果学生选修了与其当前已有知识深度不匹配的课程,其学习会变得更加困难,学习效果大打折扣^[33],学生更容易中断学习。慕课平台在进行资源推荐时,可以结合用户的个人信息、已完成的课程、未完成的课程,判断用户对某一领域知识的认知深度,推荐难度与深度适宜的课程。对于课程深度和用户不匹配的情况,可以在用户选修前给予提示,并给用户推荐相应的先修慕课资源和其他知识服务平台的电子信息服务等。

慕课平台应该引导用户选择最符合自身需求和知识水平的课程,减少欠精准的推荐对用户的干扰,从而提升用户对于推荐服务的满意度,解决由课程深度和方向不匹配导致的用户参与度不高、选择课程闲置等问题。

4.3 提供跟踪学习指导,建设用户交流社区

在慕课服务开展环节,慕课平台可以为用户提供全程的学习辅助与指导,建设用户交流社区。国内外的慕课服务中均存在用户学习中断率高的问题。沉默型用户极低的证书获得率、需求型用户较低的平均学习时长也反映了用户中断学习的情况广泛存在。此外,用户群体画像充分说明了部分用户在使用慕课平台时有一定的社交倾向,但当前慕课平台的社交功能并不完善,用户只能在某一课程下的评论区进行讨论,无法进行更深入的互动。

对于在线学习这一场景,用户往往缺乏监督和学习动力,所以比线下学习更容易放弃。针对这一现状,慕课平台可以从线下学习的场景中获得启发,完善平台功能,改善用户学习体验。线下学习过程中,学生在遇到困难时可以寻求参考教材、教师、同学等的帮助;慕课学习过程中,用户获取学习指导和辅助的渠道相对较为单一,通过评论区提问并不能及时获得回复,用户进行讨论和知识信息共享的空间有限。基于此,平台可以完善用户的交流社区建设,给学习同一门课程、兴趣相似的用户提供深度交流学习的渠道,通过社区让积极型用户和需求型用户带动其他类型用户参与讨论与学习,营造良好的学习氛围,引导用户形成共同学习的圈子,互相监督,从而提升用户学习积极性和用户黏性。此外,平台可以整合各方学习资源,鼓励课程制作团队推荐匹配的参考教材等学习辅助资源,与高校图书馆、公共图书馆展开合作,把对应资源的获取渠道接入慕课平台,帮助用户在遇到困难时直达学习资源,辅助用户理解和学习,从而形成一站式服务。

全流程的智慧知识服务通过向用户提供完善的学习辅助与指导功能,以及鼓励用户进行学习交流和互相监督,降低用户学习中断率和流失率,保障用户学习效果。

4.4 激发现有用户潜力,发展壮大知识服务用户群体

从当前慕课平台的用户构成特征来看,慕课平台需要激发现有用户潜力,同时发展壮大用户群体。当前慕课用户以学生群体为主,但也有部分职场人表现出了对慕课资源的需求,学生群体之外的用户表现出了更强的学习积极性和黏性,需求型用户已经把慕课平台作为长

期学习、终身学习的渠道。虽然慕课这一概念始于高校开放课程,但慕课课程资源在不断丰富,服务的对象正在不断拓展。

一方面,慕课平台可以从多个方向完善平台资源建设与服务功能,激发现有用户潜力,促进用户向核心用户转化,提升他们对服务的满意度,从而驱动用户主动宣传慕课服务,引入有知识服务需求的其他用户。另一方面,慕课平台可以积极主动参与知识服务体系建设,通过与图书馆等其他知识服务机构合作,跨越平台限制向高校图书馆等场景的用户推荐合适的资源,扩展知识服务用户的资源选择渠道,这也有利于发展慕课场景的新用户。此外,针对高校老师、职场人士等专业用户,平台可以深入调研他们的需求,提供有针对性的资源与服务,从而拓展慕课平台用户,这也有利于把学生用户培养为具有长期学习与终身学习习惯的需求型用户。

基于慕课场景的知识服务在满足现有用户群体的需求的基础上,可以提升资源多样性,不断拓宽知识服务对象范围,推进知识服务满足专业化与大众化需求。

5 结语

本文聚焦慕课场景的知识服务,以中国大学MOOC平台的用户为样本,进行用户群体画像,把慕课用户划分为沉默型用户、积极型用户、需求型用户、潜力型用户4种类型。基于群体画像分析了慕课场景中用户对知识服务的需求,提出了慕课平台开展知识服务的改进策略,从而更好地满足平台用户的需求,促进用户向核心用户转化,维持用户活跃度,降低核心用户流失率,以有效利用慕课知识资源,使智慧知识服务的理念融入慕课场景,完善相应的知识服务体系建设。

慕课场景下知识服务的发展目标是提升各个环节的智慧化程度,提高资源利用率和服务质量,不断提升用户满意度。更进一步地,从促进知识服务体系发展的目标来看,完善、高质量的慕课服务可以丰富用户获取知识服务的渠道,从而发挥在知识服务体系中的作用,促进知识服务体系运转。知识服务体系建设离不开所有知识服务提供者、知识服务中介的共同参与和努力。慕课平台是知识服务的前端场景,掌握了大量用户在学习时产生的数据,这些数据有重要的潜在价值,借此可以挖掘用户深层次、隐性的需求。如果未来的知识服务体系可以整合各个场景的用户信息,促进信息共享,让其中各个环节和场景协同发挥作用,发挥各平台的

信息优势,就可以进一步提升各类知识服务资源的利用率,最大限度满足用户知识需求。

参考文献

- [1] JORDAN K. Initial trends in enrolment and completion of massive open online courses[J]. The International Review of Research in Open and Distributed Learning, 2014, 15 (1): 133-160.
- [2] 杨文建. MOOC模式下高校图书馆服务拓展[J]. 新世纪图书馆, 2015 (4): 44-48.
- [3] 黄红兵, 董艳, 夏侯勤, 等. 国际慕课研究的计量分析及其对我国的启示: 基于2012—2022年1775篇国际慕课研究文献的分析[J]. 现代教育技术, 2023, 33 (3): 100-108.
- [4] 于岩, 朱鹏威. “互联网+”环境下高校“MOOC+SPOC”信息资源管理模式建设研究[J]. 情报科学, 2022, 40 (5): 59-64.
- [5] 新华社. “国家高等教育智慧教育平台”上线, 我国上线慕课数量超过5万门[EB/OL]. [2023-11-03]. <https://m.gmw.cn/baijia/2022-03/29/1302871227.html>.
- [6] 马费成, 周利琴. 面向智慧健康的知识管理与服务[J]. 中国图书馆学报, 2018, 44 (5): 4-19.
- [7] 李芳慧, 王玲. “互联网+”时代下的知识服务平台运营模式探究[J]. 图书馆学研究, 2018 (12): 63-67.
- [8] 曹雅宁, 柯青, 王笑语, 等. MOOC平台设计如何引导用户深度参与学习: 基于劝导技术理论的分析[J]. 中国远程教育, 2023, 43 (11): 39-49.
- [9] STEFAN D. Content and instructional design of MOOCs on information literacy: a comprehensive analysis of 11 xMOOCs[J]. Information and Learning Sciences, 2019, 120 (3/4): 173-189.
- [10] 邵杨芳. 健康教育类在线课程的用户需求及评价挖掘分析[J]. 中国大学教学, 2023 (S1): 100-113.
- [11] NARAYANASAMY S K, ELÇI A. An effective prediction model for online course dropout rate[J]. International Journal of Distance Education Technologies, 2020, 18 (4): 94-110.
- [12] 虞鑫, 陆洪磊. 学习者采用MOOC的影响因素研究: 基于清华大学“学堂在线”的用户调查[J]. 现代远程教育, 2014 (5): 3-8.
- [13] 和文斌, 董永权, 滕希, 等. 基于复杂理论的MOOC用户留存意愿组态分析[J]. 数字教育, 2023, 9 (4): 29-36.
- [14] 陈辉, 白骏, 殷传涛, 等. 基于学习行为的MOOC用户持续学习预测框架[J]. 北京航空航天大学学报, 2023, 49 (1): 74-82.
- [15] 魏玲, 郭新悦. 基于改进RFM与GMDH算法的MOOC用户流失

- 预测[J]. 中国远程教育, 2020 (9): 39-43, 61, 76.
- [16] 秦鸿. MOOCs的兴起及图书馆的角色[J]. 中国图书馆学报, 2014, 40 (2): 19-26.
- [17] 赵莉娜. 慕课背景下高校图书馆嵌入式信息服务研究[J]. 图书馆研究, 2017 (10): 46-49.
- [18] 李佳玉. 基于慕课的高校图书馆嵌入式知识服务研究[J]. 图书馆工作与研究, 2021 (4): 11-15, 64.
- [19] 聂慧. 基于慕课平台的高校图书馆嵌入式阅读推广服务研究[J]. 图书馆工作与研究, 2021 (11): 110-118.
- [20] 徐海玲, 张海涛, 张泉慧, 等. 基于概念格的高校图书馆群体用户兴趣画像研究[J]. 情报科学, 2019, 37 (9): 153-158, 176.
- [21] 武慧娟, 赵天慧, 孙鸿飞, 等. 基于支付意愿的数字阅读用户画像聚类研究[J]. 情报科学, 2022, 40 (5): 118-127, 136.
- [22] 王大阜, 邓志文, 贾志勇, 等. 基于用户画像的高校图书馆个性化图书推荐研究[J]. 河南师范大学学报(自然科学版), 2022, 50 (3): 95-103.
- [23] 李宇佳, 王益成. 基于用户动态画像的学术新媒体信息精准推荐模型研究[J]. 情报科学, 2022, 40 (1): 88-93, 101.
- [24] SHAHBAZI Z, BYUN Y C. Agent-based recommendation in E-learning environment using knowledge discovery and machine learning approaches[J]. Mathematics, 2022, 10 (7): 1192.
- [25] 杨凡. 澳洲八校集团成员图书馆基于慕课的嵌入式知识服务实践与启示[J]. 图书馆工作与研究, 2022 (4): 51-56.
- [26] 戎军涛, 索传军, 祝小静. 知识中介活动的内涵、功能与运行机制分析[J]. 图书情报工作, 2023, 67 (17): 3-14.
- [27] 张晓林. 颠覆数字图书馆的大趋势[J]. 中国图书馆学报, 2011, 37 (5): 4-12.
- [28] 刘安然. 中美高校慕课比较研究: 以“中国大学MOOC”与Coursera为例[J]. 高教探索, 2021 (9): 88-94.
- [29] ARTHUR D, VASSILVITSKII S. K-means++: the advantages of careful seeding[R]. New York: ACM Press, 2006.
- [30] ROUSSEEUW P J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis[J]. Journal of Computational and Applied Mathematics, 1987, 20: 53-65.
- [31] ZHOU J L, JIANG G R, DU W, et al. Profiling temporal learning interests with time-aware transformers and knowledge graph for online course recommendation[J]. Electronic Commerce Research, 2023, 23 (4): 2357-2377.
- [32] 卢思佳, 王凤姣. 基于画像分析的高校图书馆精准知识服务研究[J]. 大学图书馆学报, 2021, 39 (5): 55-62, 86.
- [33] RAWAT B, DWIVEDI S K. Discovering learners' characteristics through cluster analysis for recommendation of courses in E-learning environment[J]. International Journal of Information and Communication Technology Education, 2019, 15 (1): 42-66.

作者简介

刁雪桦, 女, 硕士研究生, 研究方向: 信息服务、知识服务。

朱学芳, 男, 博士, 教授, 通信作者, 研究方向: 数字信息资源管理及服务、多媒体信息处理、信息可视化、数字人文、模式识别与人工智能, E-mail: xfzhu@nju.edu.cn。

Knowledge Service Strategy of MOOC Platforms Based on User Group Portrait Analysis

DIAO XueHua^{1,2} ZHU XueFang^{1,2}

(1. School of Information Management, Nanjing University, Nanjing 210023, P. R. China;

2. Jiangsu Key Laboratory of Data Engineering and Knowledge Service, Nanjing University, Nanjing 210023, P. R. China)

Abstract: As an important kind of online learning platform, MOOC website stores a large number of knowledge service resources, which is also an important scene for the implementation of knowledge services and a part of the knowledge service system. Studying the group structure and characteristics of current MOOC users is not only conducive to improving the quality of platform service, but also to enriching the knowledge service channels, so as to better meet the diversified needs of knowledge service users. This paper focuses on the knowledge service of MOOC. From the perspective of users, this paper takes users of Chinese University MOOC platform as samples, applies K-means++ clustering algorithm to conduct user group portrait and analysis, and finally divides MOOC users into four types: silent users, active users, demand-oriented users, and potential users. Based on the portrait results, this paper proposes the optimization strategies of knowledge service based on the MOOC scene, in order to improve the user satisfaction of MOOC and improve the corresponding construction of intelligent knowledge service system.

Keywords: User Portrait; MOOC; Knowledge Service; Clustering

(责任编辑: 王玮)