

科研信息服务平台用户活跃度测算研究^{*}

——以国家科技图书文献中心为例

陈阳 白海燕

(中国科学技术信息研究所, 北京 100038)

摘要: 在科研信息服务背景下, 探究如何从海量业务数据中对用户活跃度进行准确测算与分析, 为实时评价活动效果、调整用户政策, 提供基于数据计算和科学分析的决策依据。以国家科技图书文献中心 (NSTL) 为例开展实证分析, 采用衰减算法计算用户行为量, 通过熵权法细分每个行为的贡献量, 对一段时间内NSTL不同类型用户活跃度进行测试与分析, 提出适用于科研信息服务平台的应用场景, 实现用户活跃度算法工程化应用。

关键词: 科研信息服务; 信息服务平台; 用户活跃度; 衰减算法; 熵权法

中图分类号: G251 **DOI:** 10.3772/j.issn.1673-2286.2023.03.004

引文格式: 陈阳, 白海燕. 科研信息服务平台用户活跃度测算研究——以国家科技图书文献中心为例[J]. 数字图书馆论坛, 2023 (3): 28-35.

数字化科研信息服务平台的发展对管理者的服务模式与理念提出了新的要求, 以用户为导向的信息化服务成为主流^[1]。许多数字图书馆逐渐从“以资源为中心”向“以用户为中心”转变^[2], 将用户视为赖以生存和发展的根本, 利用用户积累内容、增长社会关系和增加访问流量^[3], 为其带来源源不断的活力。然而并不是所有的用户都能为平台带来价值: 一些用户会处于“潜水”状态, 他们并不经常使用平台, 对平台的贡献价值较小; 而活跃用户会产生检索、浏览、下载等行为, 对平台产生了至关重要的作用。由此可见, 对于用户活跃度的分析与研究, 可以在一定程度上评价科研信息服务平台的服务效果, 帮助决策者了解平台对用户的吸引力是否充足、推出的活动是否增加了用户活跃行为、哪些政策可以激发用户购买行为等。通过分析解决这些问题, 为实时评价活动效果、监测平台用户行为状况、改善服务质量提供数据支持, 从而形成用户满意、科学管理的双赢局面。

近年来, 用户活跃度研究已经形成较为完善的体系, 但主要聚焦于社交媒体。一般网站将用户发布原创信息、转发消息、分享信息以及关注好友、评论、点赞等作为评价用户活跃度的指标^[4]。用户活跃度越高, 越能衡量平台运行的整体效果, 给平台带来提升的空间。与此同时, 决策者通过了解用户行为的变化趋势, 进一步探索和挖掘平台用户使用情况, 推出有针对性的运行策略, 增强用户的体验感。对于图书馆等信息服务平台来说, 根据读者证使用频率划分读者活跃度, 通过计算图书馆流通信息和读者信息之间的增益度得出对活跃读者的评价规则, 可以更有针对性地分析图书馆工作的潜在价值规律^[5]。

本文通过借鉴现有用户活跃度研究方法, 提出科研信息服务平台用户活跃度测算算法, 以国家科技图书文献中心 (National Science and Technology Library, NSTL) 为例进行实证分析与测试, 设计工程化应用场景, 为实时评价平台活动效果、改善服务策

收稿日期: 2022-11-24

^{*} 本研究得到中国科学技术信息研究所创新研究基金青年项目“NSTL用户活跃行为的测度与评价” (编号: QN2022-15) 资助。

略、加强业务数据管理提供决策支持。

1 研究背景

1.1 用户活跃度概念

用户活跃度并没有统一的学术概念,一般根据不同的场景赋予不同的含义。例如:手机银行的用户活跃度是指用户使用手机银行的一项功能或服务的频繁程度,使用的频率越高,就表明用户活跃度越高,对业务贡献程度越大^[6];微博用户活跃度指的是用户每天使用微博的各种行为发生的频次之和,如用户发布博文、转发、私信等^[7]。此外,也有平台将用户活跃度定义为用户在网站上的各种行为以及属性。以微博用户为例:用户行为指的是发布博文、转发博文、评论博文、关注其他用户和被其他用户关注;用户属性指的是原创博文数、转发博文数、评论博文数、粉丝数和关注数^[8]。

可以看出,现有用户活跃度的定义一般与用户在使用平台时的各种行为有关。因此,本文针对用户行为指标的维度进行了调研与梳理。在传统虚拟社群方面,2001年Nonnecke等^[9]将用户行为划分为发表行为和潜水行为:发表行为是一种外向型行为,用户会主动发帖并做出内容贡献;潜水行为是一种内向型行为,用户从不发帖,只有浏览行为。但是,用二分法定义用户行为存在一定的局限性,不能准确地刻画用户行为。后来,随着研究的不断完善与深入,Heinonen^[10]通过了解用户行为的发生动机,将用户社交媒体活动分为消费、参与以及贡献3种类型,以此定义用户活跃行为包括信息处理活动、娱乐活动以及社会联系活动。在社交网络服务方面,Pagani等^[11]根据不同的网络用户参与程度,将使用行为分为5个维度,分别为创造行为(creator)、评论行为(critic)、收集行为(collector)、参与行为(joiner)、旁观行为(spectator)。陈爱辉等^[12]基于社交媒体的内容与关系属性,将社交网络服务(Social Network Service, SNS)用户的活跃行为划分为4个维度,分别是内容创造行为、内容传播行为、个人关系贡献行为和群体关系贡献行为,以便帮助运营商更好地理解并促进用户的使用行为。由此可见,应用场景不同,用户活跃度的定义也不同,且都与用户的网络使用行为有关。因此,把握用户在科研信息服务平台中的各种使用行为是明确平台用户活跃度的前提。

1.2 用户活跃度测算

一般学者会依据用户在平台中各种行为的频次测算用户活跃度,然而周静等^[13]在对SNS网站活跃度的研究中,将用户活跃度的计算公式设置为“(登录频率+停留时间)/登录周期”,以此探究用户在社交网站的使用行为如何影响活跃度,为提高SNS网站活跃度提供了理论依据。刘文沛等^[14]提出度量软件生态系统的用户活跃度测评方法,通过设计层级结构指标体系(包括应用吸引力、用户忠诚度以及用户流动性等),监控社会网络服务平台的运营情况并进行趋势分析。在微博用户行为活跃度研究方面,陈俊宇等^[7]以数据流的形式系统地分析了微博用户的行为数据,并利用衰减窗口计数方法来计算用户活跃度,且实验结果具有一定的稳定性。高迎等^[8]选取用户粉丝数、关注数以及博文数这3个属性,利用K-means聚类算法将用户分为4类,分别是极度活跃用户、活跃用户、中度活跃用户和普通活跃用户,为提升用户活跃度提供可借鉴的依据。

综上所述,目前没有对用户活跃度的统一学术定义,缺乏分类标准及测算方法规范。同时,本文研究的是科研信息服务平台(与一般的社交媒体有所不同),在用户活跃度研究方面缺乏可借鉴的经验。因此,本文结合用户活跃度相关研究成果和科研信息服务平台的特点提出科研信息服务平台用户活跃度测算方法,为科研管理人员深入挖掘用户活跃度提供借鉴经验。

2 用户活跃度测算

2.1 指标选取

将用户活跃度定义为在一段时间内使用过科研信息服务平台的用户发生的各种行为量的加权和。在使用科研信息服务平台时,登录用户和非登录用户存在权限差异:当用户登录时,拥有使用平台各项功能的权利,平台提供基于用户行为的个性化服务;当用户未登录时,只拥有平台部分功能的使用权利(如检索、浏览、在线下载等),不能使用下单、分享、收藏等功能。因此,结合现有用户行为日志,对于登录用户活跃度的测算,选取检索量、访问量、浏览量、下载量、订单量、订阅量、分享量、收藏量8个指标,其中访问量是指用户访问全部页面的累计量,浏览量是指用户浏览文献页面的累计量,且访问量大于浏览量;对于非登录用户活跃

度的测算,选取检索量、访问量、浏览量、下载量4个指标。

2.2 测算方法

2.2.1 衰减算法

对于一个用户来说,在一段时间内的活跃行为量是其在社交网络中发生的各种行为的频次和,但曾经有过频繁的使用行为不代表用户当前依然活跃,简单地计算行为频次之和不能准确地反映用户在这一段时间内的活跃行为情况。用户在一段时间内的活跃行为量应该是随着时间衰减的,求得的是衰减和。因此,使用衰减算法求得用户活跃行为量,使得计算结果更贴近用户真实的活跃程度。

衰减算法的基本思想是赋予数据流从开始时刻到当前时刻所有元素不同的权值。离当前时刻最近的元素被赋予较大的权值,随着时间的不断推移,权值按照略小于1的比值衰减,从而逐渐消除历史活跃行为数据对用户行为量的影响^[15]。将一段时间内用户发生的使用行为作为一个数据流,采用衰减算法进行计算,可得到每个用户在一段时间内的使用行为量。

将用户行为指标个数定义为 n , $n=8$,表示登录用户行为指标数量; $n=4$,表示非登录用户行为指标数量。定义从选定时间开始($t=1$)到选定时间结束,用户 m 的活跃行为矩阵如公式(1)所示。

$$A_m(m) = \begin{bmatrix} A_1(m) \\ A_2(m) \\ \vdots \\ A_t(m) \end{bmatrix} = \begin{bmatrix} x_{11}(m) & x_{12}(m) & \cdots & x_{1n}(m) \\ x_{21}(m) & x_{22}(m) & \cdots & x_{2n}(m) \\ \vdots & \vdots & & \vdots \\ x_{t1}(m) & x_{t2}(m) & \cdots & x_{tn}(m) \end{bmatrix} \quad (1)$$

可以看出,用户 m 的行为是一个关于时间 t 的矩阵流。因此,定义用户 m 的矩阵流衰减窗口^[17]如公式(2)所示。

$$B_t(m) = \sum_{k=1}^t A_k(m) (1-c)^{k-1} \quad (2)$$

式中: $B_t(m)$ 为用户 m 在时间 t 内的行为量; c 为衰减常数。从式(2)可以看出,矩阵流呈指数型衰减,出现越晚的元素权值越大,出现越早的元素权值越小。

依次使用上述方法,得到一段时间内用户 m 的行为加权矩阵 Y ,具体见公式(3)所示。

$$Y = \begin{bmatrix} B_t(1) \\ B_t(2) \\ \vdots \\ B_t(m) \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mn} \end{bmatrix} \quad (3)$$

式中: y_{mn} 为在时间 t 内,用户 m 每个指标的行为加权。

2.2.2 熵权法

熵原本是热力学概念,最先由申农引入信息论,现已在工程技术、社会经济等领域得到广泛的应用^[16]。在具体的使用过程中,可以利用熵值来判断某个指标的离散程度。指标的信息熵值越小,离散程度越大,提供的信息量就越多,该指标在综合评价中的影响也就越大。该方法既考虑了数据本身具有的信息又避免了主观因素的干扰,使得对评价指标权重的确定更加客观与科学。因此,在得到每个用户行为总量的基础上,利用信息熵来计算各指标的权重,从而得到较为客观合理的评价结果,为多指标综合评价提供依据。

利用熵权法确定指标权重,求出登录用户、非登录用户以及总体活跃度情况,主要有5个步骤。

第一,设参与评价的对象集,也就是用户数为 $M=(M_1, M_2, \cdots, M_m)$,指标集为 $N=(N_1, N_2, \cdots, N_n)$,评价对象 M_i 对指标 N_j 的值记为 y_{ij} ($i=1, 2, \cdots, m; j=1, 2, \cdots, n$)。对该样本数据可直接使用用户 m 的行为加权矩阵 Y ,具体见公式(4)所示。

$$Y = \begin{bmatrix} & N_1 & N_2 & \cdots & N_n \\ M_1 & y_{11} & y_{12} & \cdots & y_{1n} \\ M_2 & y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ M_m & y_{m1} & y_{m2} & \cdots & y_{mn} \end{bmatrix} \quad (4)$$

第二,对矩阵 Y 进行标准化处理,得到标准化矩阵 $V=(v_{ij})_{m \times n}$, v_{ij} 计算方式具体见公式(5)所示。

$$v_{ij} = \frac{y_{ij} - \min(y_j)}{\max(y_j) - \min(y_j)} \quad (5)$$

第三,根据信息熵的定义^[17]计算熵值。定义第 j 个指标下第 i 个评价对象的特征比重为 p_{ij} ,则 $p_{ij} = v_{ij} / \sum_{i=1}^m v_{ij}$ 。由此可得第 j 个指标的指标熵值 e_j ,具体见公式(6)所示。

$$e_j = -\frac{1}{\ln(m)} \sum_{i=1}^m p_{ij} \ln(p_{ij}) \quad (6)$$

当 $p_{ij}=0$ 或 $p_{ij}=1$ 时,认为 $p_{ij} \ln(p_{ij})=0$ 。

第四, 确定各指标的熵值 ω_j , 其表达为公式(7)。

$$\omega_j = \frac{1 - e_j}{\sum_{j=1}^n (1 - e_j)} \quad (7)$$

第五, 求得第 i 个用户的活跃度, 其表达为公式(8)。

$$f_i = \sum_{j=1}^n \omega_j v_{ij} \quad (8)$$

3 实证分析

3.1 实验数据

经过调研多家科研信息服务平台, 决定选取NSTL作为研究对象。NSTL是科技部联合财政部等六部门, 经国务院批准, 于2000年6月12日成立的一家基于网络环境的科技文献信息资源服务机构^[18]。经过20余年的发展, 目前NSTL拥有理、工、农、医各学科领域的科技文献资源, 面向全国提供公益普惠的科技文献信息服务, 吸引了大批科研工作者, 用户类型繁多。NSTL用户可分为个人用户(以个人方式进行实名注册的用户)、公益集团用户(公益性、非营利性团体机构的管理用户)、企业集团用户(公司、企业等商业机构的管理用户)。

从NSTL用户日志获取实验数据, 在Python 3.8环境下, 对2020年1月—2022年8月的用户数据进行测试, 包含登录用户数据约800万条、非登录用户数据约3亿条。使用Elasticsearch作为日志存储方式, 日志由8个索引构成, 分别对应检索量、访问量、浏览量、下载量、订单量、订阅量、分享量、收藏量8个指标。

3.2 实验分析

采用上述用户活跃度指标, 对登录用户、非登录用户以及整体用户活跃度进行分析与评价。由于一些指标存在稀疏性, 考虑将下载量和订单量合并为“下单量”, 将订阅量、分享量和收藏量合并为“其他”。利用衰减算法求得一段时间内用户的行为量, 再对数据进行标准化处理, 采用熵权法对各项指标赋权, 求得每天/月/年维度下的登录用户、非登录用户以及整体用户活跃度情况。

由于篇幅有限, 仅展示2022年1—8月NSTL登录用

户以及非登录用户活跃度, 具体包括用户活跃度值、同比与环比情况, 如图1所示。

可以看出, 2022年1—8月登录用户活跃度在5.5以下, 同比增长率几乎均为负值, 环比增长率正值稀少, 说明在2022年1—8月登录用户不是非常活跃。相比而言, 非登录用户活跃度基本在20~40, 明显高于登录用户活跃度, 同比增长率超过50%, 说明非登录用户使用NSTL非常频繁, 使用行为保持继续增长的态势。

根据上述登录用户以及非登录用户活跃度情况, 可得到2022年1—8月用户整体活跃度情况, 如图2所示。

可以看出, 整体活跃度在20以上, 且在8月达到峰值, 同比增长率较高, 但环比增长率较低。在活跃用户数方面, 2022年1—2月活跃用户数较少, 原因可能是部分用户处于放假状态, 使用NSTL的人数不多。3月的活跃用户数骤增, 在随后的几个月活跃用户数又出现下降趋势。综合图1和图2可以看出, 提升用户活跃度的关键在于积极促进登录用户对NSTL的使用行为频次, 开展丰富的活动来提升平台对用户的吸引力。

运用统计学方法, 根据Spearman相关性对整体活跃度和活跃用户数两个变量进行相关性分析, 选取2020年1月—2022年8月共32个月的数据, 结果如表1所示。

可以看出, 整体活跃度与活跃用户数的Spearman相关系数为0.395, 在0.05水平下显著性值为0.025, 小于0.05, 说明两个变量在0.05置信水平下存在正相关关系, 即整体活跃度与活跃用户数有关。

在实际的应用过程中, 社交媒体平台会利用日活跃用户数、月活跃用户数来评估用户活跃度, 了解目前市场状况。本研究提出的用户活跃度测算算法改进现有算法, 可在一定程度上帮助决策者了解用户活跃情况, 对科研信息服务平台进行实时监控与评估。

4 应用场景

4.1 活动监测与评价

在数据规模日益庞大的当下, 许多科研信息服务平台都有对用户使用效果进行评价的需求, 尤其是在活动上新前后, 需要监测与评价活动对用户的吸引力,

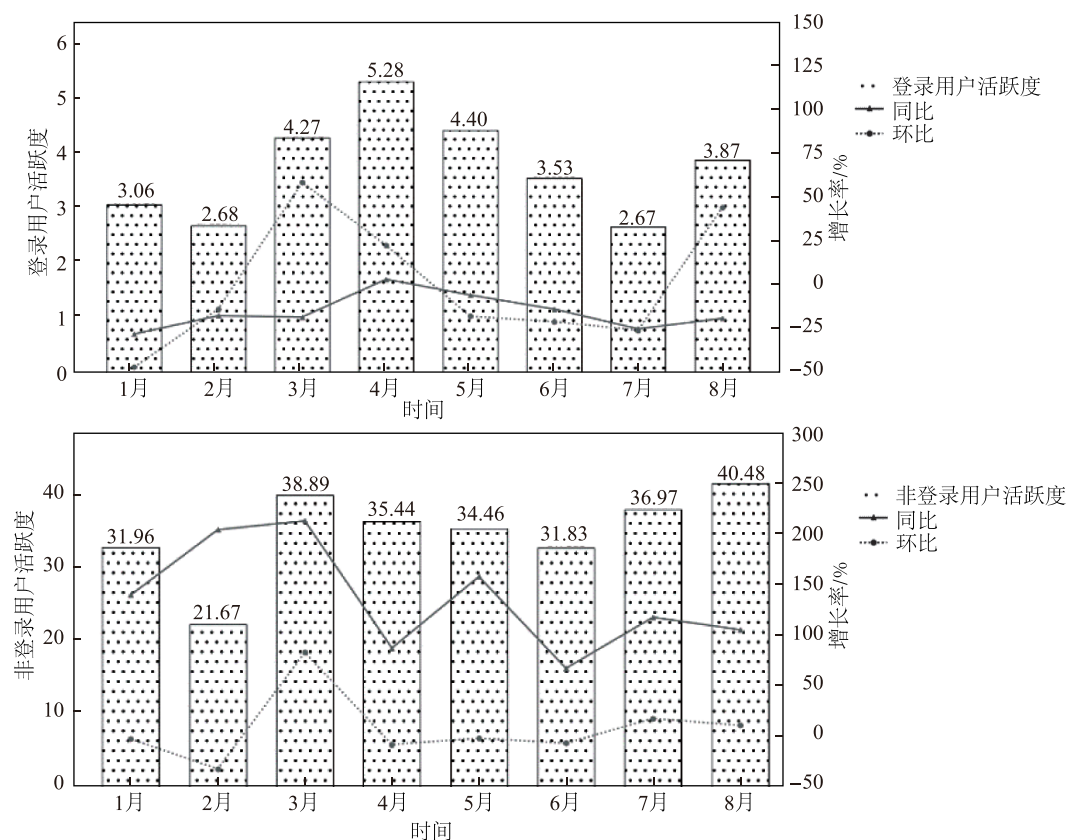


图1 登录用户以及非登录用户活跃度情况

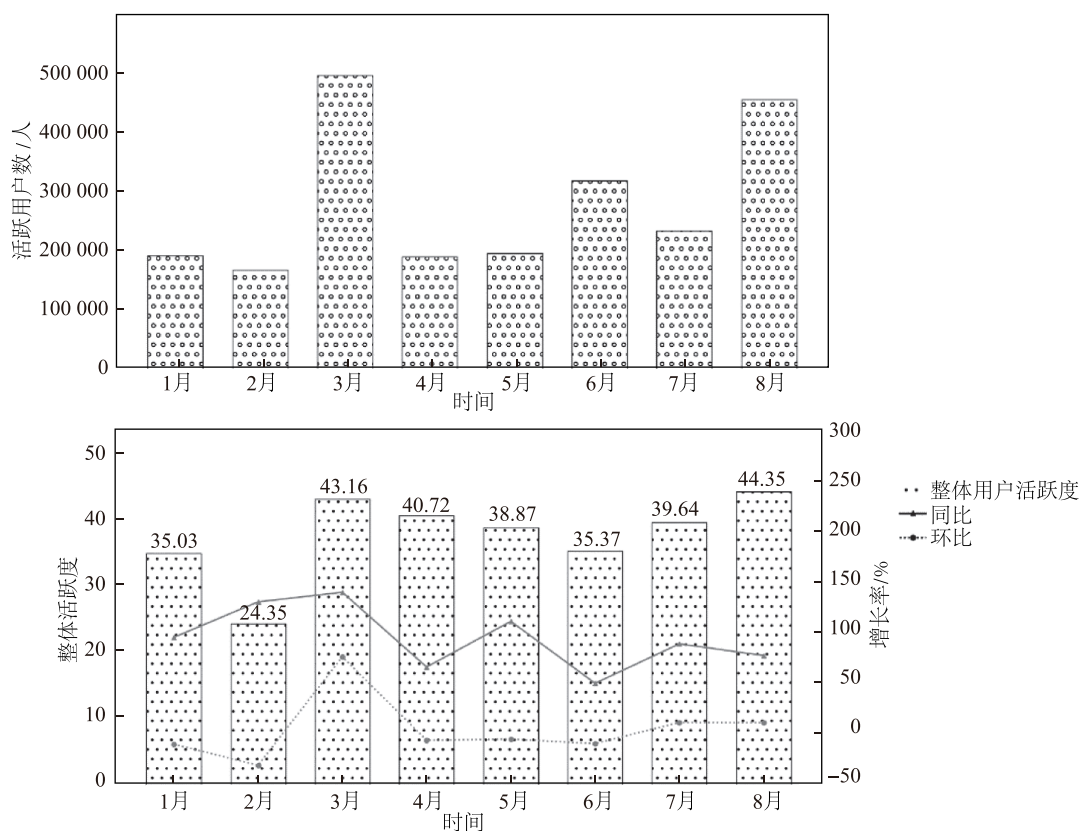


图2 活跃用户数和整体活跃度情况

表1 相关性分析结果

变量	统计量	变量	
		整体活跃度	活跃用户数
整体活跃度	Spearman相关系数	1	0.395 [*]
	双尾检验的显著性值		0.025
	个案数	32	32
活跃用户数	Spearman相关系数	0.395 [*]	1
	双尾检验的显著性值	0.025	
	个案数	32	32

注：“*”表示相关性显著。

以此增加平台的活跃度。科研信息服务平台用户活跃度测算算法在活动期间可以对活动效果形成每周或每月的评估报告,便于决策者及时调整和优化活动内容。活动结束后,可以对此次活动做出完整的效果评估,为今后的类似活动提供指导与事实依据。

在活动效果评估方面,提出的算法可统计活动页面每周或每月的用户行为量、活跃用户数以及用户活跃度,展示同比、环比情况。同时,对于用户活跃度不高的周或月,可查看其活跃行为指标贡献度。若指标贡献度低(如浏览量过少),说明推出的活动可能没有被更多的用户看到,需要加大活动推广力度来吸引用户。此外,还可对每份评估报告进行综合打分(0~100分),划分4个等级,85~100分为非常活跃,75~84分为比较活跃,60~74分为一般活跃,低于60分为欠活跃。最终,可以得到每周或每月的用户活跃度评价报告。

提出的算法还可应用于科研信息服务平台数据中台。通过设置用户活跃度筛选条件(包括时间维度、用户类型、星级、注册站点等),科研管理人员可随时查看每个筛选维度下的用户活跃度情况,具体展示效果如图3所示。



图3 数据中台展示图

同时,根据对用户活跃度的分析,可以得到一段时间内用户活跃度排名情况。通过点击用户身份信息(ID),科研管理人员可查看该用户的基本信息,如性别、年龄、地域、用户类型等;查看用户行为统计结果及趋势图,如检索量、浏览量、订购量、访问量等;得到用户行为排行情况,如文献订购量、文献下载量、文献浏览量等排行;开展用户兴趣分析,如主题词(该用户近期关注的主题词)、同兴趣用户(根据该用户的主题词匹配有相似主题词的其他用户)、同兴趣用户主题词(与该用户有共同兴趣的用户所关注的主题词)。

4.2 基于用户活跃度的用户分群

通过数据挖掘与分析这一手段,参照登录用户活

跃度情况对用户群体分类,不仅能够反映出用户的需求,还能帮助科研管理人员细分用户,精准定位目标群体,推出有针对性的活动策略。通过关注一段时间内的活跃用户,一方面,可以获取这些用户的浏览量、检索量、下载量等一系列行为情况;另一方面,由于每个活跃用户起到的作用不同,影响力也有所差别,有必要识别这些用户中的细分群体来分清不同用户的作用。对于活跃用户与僵尸用户这些差异性大的用户群体,可以进行筛选识别,探究其行为指标背后的规律,发现科研信息服务平台用户行为新的增长点。

聚类是一种无监督分类方法,是指将没有分类标签的数据集分成若干个簇的过程^[19]。其根本目标是使属于同一类的样本尽量相似,使属于不同类的样本体现出明显的差别。目前,聚类算法大致可划分为层次化

聚类算法、划分式聚类算法、基于密度和网格的聚类算法。首先,可利用聚类算法将用户进行分群,根据提出的用户活跃度测算算法,计算出每个簇群用户活跃度情况,将其分为高活跃用户、活跃用户、低活跃度用户、僵尸用户等。其次,为每个用户添加标签,对不同用户进行个性化精准营销,同时也为后续用户画像相关研究提供参考。

利用用户活跃度将用户分群,能够为科研信息服务平台服务提供改进的依据。对于活跃度较高的用户,通过分析其使用行为,了解该用户感兴趣的领域,在其使用科研信息服务平台时进行相关资源与服务的推荐,增强用户的使用黏性;对于活跃度较低的用户,可定期通过短信或邮件方式推送优惠活动,促进其登录或下单行为,使其继续使用该平台。同时,设定更新时间间隔,对用户分群情况进行动态更新,及时调整用户标签。

5 结语

借助现有用户活跃度研究,以NSTL用户日志数据为基础开展实证分析,提出科研信息服务平台用户活跃度测算方法。首先考虑用户曾经有过频繁的使用行为不代表当前同样活跃,利用衰减算法求得一段时间内所有用户的行为量,然后借助熵权法求得每个行为指标的权值,得出一段时间内登录用户活跃度、非登录用户活跃度以及整体活跃度情况。加强对用户行为数据的挖掘与分析,有利于对上新活动实时监测,便于决策者调整活动方案,提升平台整体活跃度。同时,根据每个用户活跃度情况,借助聚类这一方法可有效地将用户分群,有针对性地对不同用户采取不同策略。

本文仍存在一定的局限性,需要从以下3个方面进一步优化与完善:指标选取较为单一,后续考虑用户停留时长、访问频次、点击情况等指标,以系统地分析用户活跃情况;熵权法无法考虑各指标之间的横向影响,权重对样本数据的依赖性较大,后续在算法研究层面,考虑主客观相结合的综合赋值法,充分参考专家知识和经验,得到优化的权值;进一步拓展用户活跃度测算的应用范围。

参考文献

- [1] 常春,邹萌萌.科技资源与用户需求之间的生态学关系研究[J].情报杂志,2020,39(11):139-142,150.
- [2] 李春旺.数字图书馆以用户为中心的特征属性[J].图书情报工作,2003,47(3):72-75.
- [3] CHEN R. Member use of social networking sites: an empirical examination[J]. Decision Support Systems, 2013, 54(3): 1219-1227.
- [4] BOYD D M, ELLISON N B. Social network sites: definition, history, and scholarship[J]. Journal of Computer-Mediated Communication, 2007, 13(1): 210-230.
- [5] 张金镒.基于数据挖掘的图书馆活跃读者研究[J].现代图书情报技术,2008(7):96-99.
- [6] 张昱,谢怀军.手机银行用户活跃度与粘度的影响因素与业务发展研究[J].财经界,2012(4):168-170.
- [7] 陈俊宇,王濛,周刚.微博用户行为活跃度的衰减窗口分析方法[J].中原工学院学报,2013,24(4):59-64,75.
- [8] 高迎,闫绍山,侯小培.基于K-means聚类的微博用户活跃度研究[J].管理观察,2018,38(10):88-89.
- [9] NONNECKE B, PREECE J. Why lurkers lurk[EB/OL]. [2022-11-09]. <http://www.cis.uoguelph.ca/~nonnecke/research/whylurk.pdf>.
- [10] HEINONEN K. Consumer activity in social media: managerial approaches to consumers' social media behavior[J]. Journal of Consumer Behaviour, 2011, 10(6):356-364.
- [11] PAGANI M, MIRABELLO A. The influence of personal and social-interactive engagement in social TV web sites[J]. International Journal of Electronic Commerce, 2011, 16(2):41-68.
- [12] 陈爱辉,鲁耀斌.SNS用户活跃行为研究:集成承诺、社会支持、沉没成本和社会影响理论的观点[J].南开管理评论,2014,17(3):30-39.
- [13] 周静,李季.从人口统计特征和生活方式探讨SNS社交网站用户的使用行为及其活跃度[J].科技与管理,2011,13(2):72-77.
- [14] 刘文沛,崔强,杨达.在线软件生态系统用户活跃度的经验研究[J].计算机工程与设计,2013,34(1):185-189.
- [15] COHEN E, STRAUSS M J. Maintaining time-decaying stream aggregates[J]. Journal of Algorithms, 2006, 59(1):19-36.
- [16] 王昆,宋海洲.三种客观权重赋权法的比较分析[J].技术经济与管理研究,2003(6):48-49.
- [17] 田振清,周越.信息熵基本性质的研究[J].内蒙古师范大学学报(自然科学汉文版),2002,31(4):347-350,360.

- [18] 国家科技图书文献中心. 关于我们[EB/OL]. [2023-03-01]. https://www.nstl.gov.cn/Portal/zzjg_jgjj.html. and applications [M]. Boca Raton: Chapman and Hall/CRC, 2013.
- [19] AGGARWAL C C, REDDY C K. Data clustering: algorithms
-

作者简介

陈阳, 女, 1997年生, 硕士, 助理工程师, 研究方向: 应用统计、数据挖掘, E-mail: chen@istic.ac.cn。
白海燕, 女, 1973年生, 硕士, 研究馆员, 研究方向: 信息组织、数字图书馆。

Research on User Activity Measurement of Scientific Research Information Service Platform:
Taking National Science and Technology Library as an Example

CHEN Yang BAI HaiYan
(Institute of Scientific and Technical Information of China, Beijing 100038, P. R. China)

Abstract: Under the background of scientific research service platform, this paper explores how to accurately measure and analyze user activity from massive business data, so as to provide decision-making basis based on data calculation and scientific analysis for real-time evaluation of activity effect and adjustment of user policy. We take the National Science and Technology Library (NSTL) as an example to carry out empirical analysis. The attenuation window is used to calculate the amount of user behavior, and then the entropy weight method is used to subdivide the contribution of each behavior. The activity of different types of users in NSTL is tested and analyzed over a period of time, and the application scenario applicable to the scientific research service platform is proposed to realize the engineering application of user activity algorithm.

Keywords: Scientific Research Information Service; Information Service Platform; User Activity; Attenuation Algorithm; Entropy Weight Method

(责任编辑: 王玮)