

# 自媒体平台中的用户亚群分类分析方法研究

梁豪<sup>1</sup>, 王超<sup>1,2\*</sup>, 李波<sup>3</sup>

1. 江西财经大学 统计与数据科学学院, 江西 南昌 330000
2. 江西出版传媒集团 博士后管理办公室, 江西 南昌 330000
3. 华中师范大学 数学与统计学院, 湖北 武汉 410000

DOI:10.61369/ASDS.2025090013

**摘 要 :** 本文围绕自媒体平台用户亚群分类分析方法展开研究。在自媒体蓬勃发展的当下, 深入了解用户亚群特征对平台运营和内容精准推送至关重要。研究采用多种方法, 综合了平台用户的行为数据、交互数据和内容偏好数据。通过聚类分析等算法对海量用户数据进行处理, 识别出不同的用户亚群。对各亚群在活跃度、关注领域、内容创作类型等方面进行了分析。结果显示, 自媒体平台用户可被划分为多个有显著特征差异的亚群, 例如高互动创作型亚群、低活跃浏览型亚群等。本研究为自媒体平台优化用户体验、制定个性化营销策略提供了理论依据和实践指导, 未来可进一步研究亚群间的动态转化机制。

**关 键 词 :** 自媒体平台; 用户亚群; 分类分析方法

## A Study on the Classification and Analysis of User Subgroups on Social Media Platforms

Liang Hao<sup>1</sup>, Wang Chao<sup>1,2\*</sup>, Li Bo<sup>3</sup>

- 1.School of Statistics and Data Science, Jiangxi University of Finance and Economics, Nanchang, Jiangxi 330000
- 2.Postdoctoral Management Office, Jiangxi Publishing Group, Nanchang, Jiangxi 330000
- 3.School of Mathematics and Statistics, Central China Normal University, Wuhan, Hubei 410000

**Abstract :** This paper focuses on the research of classification and analysis methods for user subgroups on social media platforms. With the rapid development of social media, gaining a deeper understanding of subgroup characteristics is crucial for platform operations and precise content delivery. The study adopts multiple approaches by integrating user behavior data, interaction data, and content preference data. Through clustering analysis and related algorithms, massive user data are processed to identify distinct user subgroups. Further analysis is conducted on subgroup differences in activity levels, areas of interest, and content creation types. The results show that users on social media platforms can be divided into multiple subgroups with significant distinguishing features, such as highly interactive content creators and low-activity content consumers. This research provides both theoretical foundations and practical guidance for optimizing user experience and developing personalized marketing strategies on social media platforms. Future work may further explore the dynamic transformation mechanisms among subgroups.

**Keywords :** social media platform; user subgroup; classification and analysis method

## 引言

在当今数字化时代, 自媒体平台如雨后春笋般涌现并蓬勃发展, 它们不仅改变了信息传播的方式, 也重塑了人们的社交互动模式。自媒体平台汇聚了来自不同背景、拥有不同兴趣爱好和行为习惯的海量用户, 这些用户在平台上的活动呈现出高度的多样性和复杂性。为了更好地理解自媒体平台用户的行为特征, 实现精准的内容推荐、有效的用户运营以及优化平台服务, 对用户进行亚群分类分析至关重要。通过深入研究用户亚群的分类方法, 能够挖掘出隐藏在庞大用户数据背后的有价值信息, 帮助平台和内容创作者更好地满足用户需求, 进而提升用户体验和平台价值。本文旨在探讨自媒体平台中的用户亚群分类分析方法, 并呈现由此得出的重要结论。

基金项目: 江西省自然科学基金(S20259977)。

作者简介:

梁豪(2001-), 男, 江西财经大学在读硕士;

李波(1977-), 男, 华中师范大学教授。

通讯作者: 王超(1988-), 男, 江西财经大学讲师, 江西省出版传媒集团博士后; 研究方向: 统计学习方法的应用研究。

## 一、问题概述

### （一）研究背景

近年来，社交媒体的快速发展为人们获取事件、新闻等信息带来了巨大的便利，成为了人们从事网络活动的重要工具。社交媒体账户作为用户在社交媒体上获取信息、发表言论和沟通交友的重要媒介，包含了大量有价值信息，通过对社交媒体账户进行分析可以有效地将具有共性特征的用户识别为一个集合。有效的用户分类方法不仅可以快速地帮助用户发现感兴趣的信息，同时还能实现可靠的社交媒体管理以及个性化的内容和用户推荐。国内有关社交媒体用户分类的研究开始阶段主要集中于挖掘用户行为信息来识别用户特征、划分用户类型。王华<sup>[1]</sup>等利用用户兴趣度结合页面停留时间和点击次数对海量数据进行约简，减小计算规模后结合编网法直接对 Web 用户聚类，提高了效率；赵文兵<sup>[2]</sup>等使用计量学方法，基于用户特性分析使用 Pajek 软件进行可视化分析，将和讯财经微博用户分为 8 类，其中信息搜寻者和网络游民约占总量的 90%；薛云霞<sup>[3]</sup>等提取微博用户的命名特征，使用贝叶斯融合的方法将微博用户分为个人和非个人用户；贺超波<sup>[4]</sup>等设计了一种基于随机游走模型的多标签分类方法 MLCMRW，通过学习用户初始化标签、迭代推理获得用户稳定标签分布来对用户进行分类；蒋翠清<sup>[5]</sup>等基于产品论坛中的用户生成内容，构建人口统计学、文体、情感、行为和关键词特征集合，使用 Stacking 分类算法将用户划分为具有潜在购买意愿用户和无购买意愿用户；凌海峰<sup>[6]</sup>等将蚁群算法与 K-means 算法结合，优化用户会话的聚类效果，在 Web 导航推荐中表现更好；李纲<sup>[7]</sup>等将社交媒体用户分为个体民众、媒体、政府和组织四类，提取人口统计学、命名和自我描述特征，基于支持向量机分类模型，发现自我描述特征对分类的贡献最大；吴海涛<sup>[8]</sup>等结合信息内容（LDA 主题模型提取主题分布）和拓扑关系（相同兴趣用户倾向于有共同粉丝）对社会媒体用户兴趣分类，综合方法的性能优于单一方法；He S<sup>[9]</sup>等利用多尺度熵方法对用户的关注者、发布行为、Twitter 文本内容等进行分析以识别 Twitter 用户行为，将用户划分为个人、媒体、企业、机器人和其他类型用户等；陈洲<sup>[10]</sup>等针对 K-Means 算法初始中心点选取和孤立点问题，提出 ICKM 算法，以密度参数最大的对象为第一中心点，结合 KCR 算法寻找后续中心点；高大菊<sup>[11]</sup>利用 K-均值或 DBSCAN 等聚类算法挖掘用户行为模式，结合情感分析识别用户对话题的情感倾向，关联聚类与情感结果探索群体行为与情感的关系；刘玲<sup>[12]</sup>等从用户基本信息、使用行为及感受角度设计问卷，通过因子分析、聚类分析将抖音用户分为五类，经矩阵分析确定点赞关注高端用户为首选目标用户，其次是发布评论转发高端和点赞关注低端用户，并针对首选目标用户提出营销策略。在自媒体算法方面，谢敏<sup>[13]</sup>对抖音算法与爆款短视频打造进行了研究，分析了抖音的算法推荐机制，包括流量池、叠加推荐、热度加权等方面；总结了抖音爆款短视频的六大类型及强烈反差、娱乐性强等特点；还提出了从账号标识、视频封面、视频制作等方面打造爆款短视频的策略；赵辰玮<sup>[14]</sup>等聚焦抖音短视频平台的算法推荐模式，指出抖音主

要采用基于用户信息的基本协同过滤、基于“去中心化”的精准推送、基于“流量池”的叠加推荐三种算法方式；那晓婷<sup>[15]</sup>从数字劳工理论视角，分析智能算法下抖音用户面临的算法驱动权力分化、内容取向商品化、劳动边界消解与隐蔽剥削、媒介依赖与算法焦虑等困境，指出算法通过培养使用习惯、引导免费创作、利用用户数据营销三种方式促使用户付出数字劳动，并从政府、平台、用户三方面提出突围建议，强调用户需提升信息获取、评价、传播能力。

### （二）问题描述

抖音作为主流短视频平台，用户浏览行为数据中蕴含着用户偏好、内容传播规律等关键信息，对平台优化推荐机制、提升用户体验及指导创作者内容创作具有重要价值。本文的研究基于网络公开的 2019 年抖音用户浏览行为数据集（包含 1737312 条记录），涵盖用户 ID、作者 ID、作品信息、互动行为（如点赞、完整观看）等 13 个指标，旨在通过数据分析与挖掘，揭示用户、作者及作品的特征规律。抖音一个视频从创作者处诞生到用户浏览所携带的数据如表 1。该数据集指标多，内容广，针对这类数据，如何有效地进行分析揭示用户、作者及作品的特征规律。

表 1：数据集的说明

字段名	释义	字段名	释义	字段名	释义
uid	用户 id	user_city	用户城市	item_id	作品 id
author_id	作者 id	item_city	作者城市	channel	作品频道
finish	是否看完	like	是否点赞	music_id	音乐 id
duration_time	作品时长	real_time	具体发布时间	H_date	时，天（发布）

首先，构建用户、作者、作品的特征指标体系，通过统计与可视化分析，明确各主体的行为分布规律（如浏览量、点赞量的集中性、视频时长对完整观看率的影响等）。

其次，利用聚类算法对用户和作者进行分类，识别不同群体的核心特征（如高互动用户、头部作者等），为精准运营提供依据。

最后，结合分析结果，提炼对平台推荐机制优化、创作者内容创作及用户留存提升有价值的结论。

## 二、模型构建及相应算法

### （一）指标构建

首先，是对特征指标的构建，每一条数据都是由用户主动发起的，与创作者视频进行交互的行为记录，包括什么人（用户）、看了谁的（作者）、什么视频（作品、发布时间、音乐、时长）、什么渠道、是否看完、是否点赞、音乐和时间等。



图 1：用户、作者、作品指标构建

### （二）聚类算法

聚类算法是一种无监督学习方法，它通过分析数据集中样

本的特征相似性，将具有相似特征的样本自动划分到同一类别（簇），而不同类别的样本特征差异较大。其核心目标是“物以类聚”，无需预先知道数据的类别标签，适用于探索数据内在结构、发现潜在模式等场景，在用户分群、市场细分、图像识别等领域应用广泛。

聚类算法的核心思想是基于样本间的“相似性”或“距离”进行分组：相似性高（距离近）的样本被归为同一簇，相似性低（距离远）的样本被分到不同簇。常见聚类算法及其特点如表2。

表2：聚类算法分类及其特点

	算法类别	典型算法	核心原理	优点	缺点	适用场景
划分式聚类	K-Means		以k个中心为基准，按距离分配样本并迭代更新中心至收敛	高效、适用于大规模数据，易实现	需预设k值，对初始中心和异常值敏感，不适用非凸簇	大规模、凸形簇数据（如用户分群）
层次聚类	凝聚式层次聚类		从单个样本为簇，逐步合并相似簇，形成层次关系	无需预设k值，可展示簇间层次	计算复杂，不适用于大规模数据	小数据、需层次关系场景（如物种分类）
密度聚类	DBSCAN		基于样本密度，通过核心点连通性成簇，自动识别噪声	无需预设k值，可识别不规则形状簇和噪声	对参数敏感，高维数据效果差	含噪声、不规则簇数据（如异常检测）
模型式聚类	高斯混合模型（GMM）		假设数据来自多个高斯分布，通过概率分配样本	支持软聚类，适用于重叠簇	计算复杂，仅适用于近似高斯分布数据	高斯分布数据（如图像分割）

本文要对大规模数据进行分析，所以采用K-Means算法。K-Means算法高效，为聚类问题提供了一种解决方案：把数据集中的n个样本点划分到k个集群中，使每一个点都属于离它最近的均值对应的集群中，使每个集群中的点存在高关联度，使每个集群之间的对象存在低关联度。算法开始之前，首先确定将数据集划分为几个集群，而划分集群个数我们则是依据聚类效果进行评估指标：

1. 内部评估指标（无需真实标签）

（1）轮廓系数（Silhouette Coefficient）：衡量样本与自身簇的相似度（a）和与最近簇的相似度（b），取值范围[-1,1]，越接近1说明聚类效果越好。文档中用户聚类时参考了轮廓系数。

（2）误差平方和（SSE）：所有样本到所在簇中心的距离平方和，SSE越小说明簇内样本越集中，但随k值增大单调递减，需结合业务选择合适k值。

2. 外部评估指标（需真实标签）

（1）纯度（Purity）：每个簇中最多类别的样本数占总样本数的比例，值越高越好。

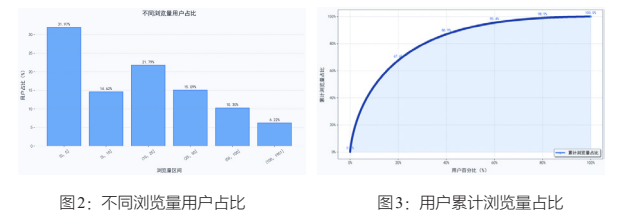
（2）兰德指数（Rand Index）：衡量聚类结果与真实标签的一致性，取值范围[0,1]，1表示完全一致。

三、求解与结果分析

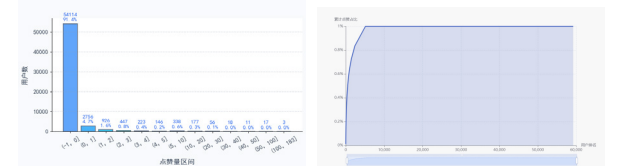
（一）可视化分析

通过对浏览行为数据的认识和分析，在构建并提取了用户、作者、作品的特征之后，进行简单的可视化分析。

1. 用户可视化分析



从图2可以得出大部分用户（约73.47%，即31.97%+21.79%+15.09%+4.62%）的浏览量集中在0到50之间，属于低到中等浏览量用户。只有一小部分用户（约6.23%）有非常高的浏览量（100以上），这些用户可能是抖音的重度使用者。图3是按照用户浏览量从大到小进行排序后，绘制的用户累积浏览量占比曲线。通过这种曲线可以分析抖音用户浏览量的集中度。图中表明曲线在前期快速上升，说明少数用户贡献了大部分浏览量。



通过对图4和图5分析来看，大部分用户没有点赞行为，而少部分用户（5.8%）贡献了大量的点赞，这反映了点赞行为在用户中的不均衡分布。进一步分析，可以知道5.8%的用户点了90%的赞，94.2%的用户只贡献了10%的赞。初步推断，点赞与否可能与用户的性格相关。



图6展示了用户完整观看视频的时长区间分布，区间（0，5）的用户占比最高，为44.43%。这表明有近一半的用户完整观看视频的时长在5秒以内。区间（50，284）的用户占比最低，为3.98%。这说明在抖音上面，较短的视频用户更加容易看完，更有用户留存。初步推断原因可能是抖音用户的观看耐心较低，或者视频内容在开头几秒内没有吸引住用户。只有较少的用户（约3.98%）能够完整观看较长时间（50秒以上）的视频。图7是用户



观看作品的平均完整时长分布,大部分用户(约50.35%)平均完整观看视频的时长在1到10秒之间,再次说明抖音用户的观看耐心较低。只有很少的用户(约3.69%)平均完整观看视频的时长在15秒以上。

这些数据表明抖音用户在观看视频时,倾向于快速浏览,大部分用户在视频开始的几秒内就做出了是否继续观看的决定。这对于内容创作者和平台运营者来说,意味着需要在视频的开头几秒内迅速抓住用户的注意力,以提高用户的完整观看率。

## 2. 创作者可视化分析

创作者可视化分析采用和用户可视化一样的分析方法。

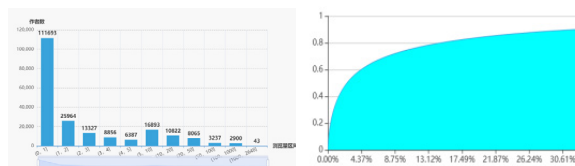


图8: 不同浏览量创作者对比

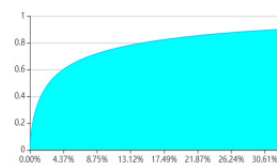


图9: 创作者累计浏览量

根据图8和图9可以看到,在浏览量方面,创作者和用户基本保持一致,依旧是少部分人贡献出了大量的浏览量。不同的是抖音创作者较抖音用户在浏览量上更为均衡,即抖音创作者基本都保持了一定的浏览量,这也符合常理,抖音创作者需要浏览更多的视频来判断当下的浏览流量来源,才能更好的创作出流量较高的视频。

从图10和图11来看,在点赞量方面,抖音创作者和抖音用户保持一致,少部分人贡献大量的点赞。

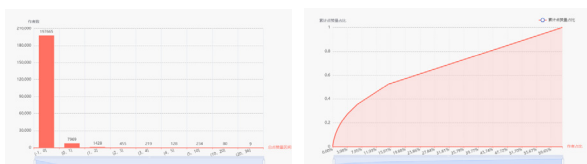


图10: 不同点赞量创作者对比

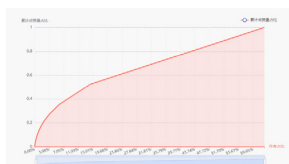


图11: 创作者累计点赞量

## 3. 作品可视化分析



图12: 各日单日作品发布量

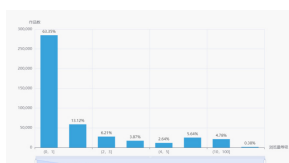


图13: 作品浏览量分布

从图12中可以得出:在2019年9月21日,每日作品发布量非常低,接近0。

从9月21日开始,每日作品发布量逐渐上升,但上升幅度较小。在2019年10月18日左右,每日作品发布量达到了一个峰值,接近35000。10月18日之后,每日作品发布量开始下降,但仍然保持在较高水平。这可能说明在10月18日左右发生了某个事件或者活动,导致用户大量发布作品。这个事件可能是平台举办的活动、热门话题或者其他能够激励用户创作和发布的因素。之后发布量的下降可能是活动结束或者热度减退的结果。

从图13可以得出:

(0,1)区间:占比最大,为63.36%。这表明大部分作品的

浏览量在0到1之间,可能是因为这些作品刚刚发布或者没有得到足够的曝光。

(1,2)区间:占比为13.12%。这些作品的浏览量略高于(0,1)区间,但仍然属于较低浏览量范围。

(2,4)区间:占比为10.07%。这个区间的作品浏览量适中。

(4,1600)区间:占比为13.45%。这个区间的作品浏览量相对较高,虽然占比不是很大,但这些作品可能是较为热门的。

大部分作品(63.36%)的浏览量非常低,这可能反映了抖音平台上内容竞争激烈,只有少数作品能够获得较高的浏览量。有一定比例(13.45%)的作品能够突破到较高的浏览量区间(4以上),这些作品可能具有较高的吸引力或者得到了平台的推荐。对于内容创作者来说,要想获得更高的浏览量,需要在内容质量、创意和吸引力上下功夫,以进入更高的流量池。对于抖音平台,平台也需要进一步优化推荐算法,帮助更多优质内容获得曝光机会,从而提高用户参与度和内容的整体质量。

## (二) 聚类分析

在对用户、创作者、作品进行简单的描述性统计分析与可视化展示后,下一步进行聚类分析。对于抖音平台本身而言,如何对用户进行分类,或者分级,然后差异化的提供服务,是一个非常重要的方向,对于商务合作和广告投放者而言,如何对创作者进行分类,如何选择合作的创作者,也是有一定价值导向。

这里首先要处理两个问题:

一是数据可视化分析中,可以看到有一部分用户使用程度平台低,浏览少,不点赞,对这样的用户进行聚类分析是无效多余的,增加筛选认为至少观看过一个完整短视频且有一定浏览量的用户才具有分析意义。

二是在对创作者的考量上,聚类的结果是服务于商务合作和广告投放,此时核心是浏览量而大部分的创作者总浏览量非常小,这些创作者是无需考虑的,故进行筛选。

两个目标同时达到最优的即所谓的最优决策<sup>[1]</sup>,我们追求的只能是,在确定的订购量下生产商的净收益最大。这就是说在不同的约束条件下,只要建模合理,答案可以是多种。

### 1. 用户特征聚类

通过对用户聚类得到的两个指标:sse(Sum of Squared Errors,误差平方和)和sc(Silhouette Coefficient,轮廓系数)与聚类数量的关系如下:

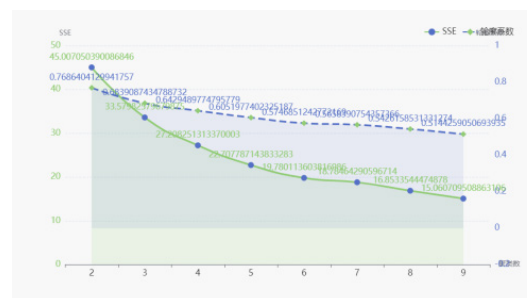


图14: 用户聚类效果

综合考虑选择 $k=4$ 为用户聚类的数量,聚类效果如下:

表3：聚类结果

	浏览量	点赞量	观看作者数	观看作品数	观看作品平均时长	观看配乐数	完整观看数	去过的城市数	观看作品城市数	人数
1	65.559893	0.651947	63.616853	65.559040	11.175167	58.035413	27.440747	1.274133	47.094827	9366
2	381.603365	3.012019	353.144231	381.598558	10.969891	276.408654	92.346154	1.324519	140.896635	416
3	162.612409	1.259288	154.847185	162.610494	11.078860	131.901187	54.075450	1.293757	88.666794	2610
4	16.791653	0.190634	16.512585	16.791518	11.315000	15.934307	8.124806	1.170288	14.572070	29648

2. 作者特征聚类

根据图 15，综合考虑选择为作者聚类的数量，聚类效果如下：

表4：聚类结果

	总浏览量	总点赞量	总观完量	总作品数	作品平均时长	使用配乐数量	发布作品日数	创作活跃度(日)	去过的城市数	人数
1	11.973167	0.115981	5.019598	3.636747	10.783111	3.198138	3.636625	11.301008	1.114046	57356
2	376.060858	3.428769	165.146611	20.035961	11.126590	13.811895	20.034578	24.224066	1.289073	723
3	1092.821053	8.347368	461.315789	31.126316	11.165662	19.957895	31.126316	27.926316	1.326316	95
4	117.075717	1.148983	51.548395	12.432737	10.993801	9.434942	12.431267	21.253124	1.236217	4079

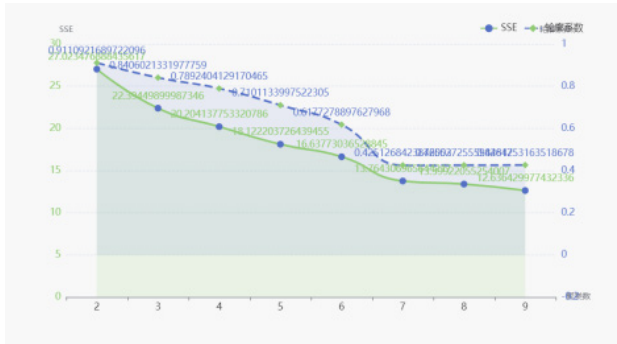


图 15：作者聚类效果

四、结论与建议

上文已经对用户特征和作者特征成功聚类，从上述表3和表4可以得出：

（一）对用户聚类的4个类别，通过分析数据可以发现每个组的特点以及相应建议如下

1. 组1：高浏览量、高观看创作者数、高观看作品数、高观看配乐数、高完整观看数、高观看作品城市数，但点赞量相对较低，这组用户可能是内容消费大户，但不太倾向于点赞。

建议：平台可以设计一些激励机制，例如点赞抽奖、点赞排行榜等，鼓励这类用户多点赞。同时，可以优化点赞操作的便捷性，减少操作步骤，让用户更容易点赞。

2. 组2：高点赞量、高观看作者数、高观看作品数、高观看配乐数、高完整观看数、高观看作品城市数，这组用户比较活跃且喜欢点赞。

建议：将这类用户视为核心用户，给予他们更多的特权和福利，如专属徽章、优先体验新功能等，保持他们的高活跃度，并鼓励他们带动其他用户。

3. 组3：各项指标适中，没有特别突出的高或低，属于比较均衡的用户组。

建议：分析他们的浏览偏好，为他们精准推荐内容，进一步提升他们在平台上的参与度。可以通过个性化推荐系统，推送他们可能感兴趣的作品和作者，促使他们在某一指标上有所提升。

4. 组4：各项指标普遍较低，是活跃度最低的用户组，可能是新用户或者对平台兴趣较低的用户。

建议：开展新用户引导活动，帮助他们熟悉平台功能，了解平台的价值。发送个性化的欢迎邮件或推送消息，介绍平台上的热门内容和热门作者，吸引他们的兴趣。同时，可以对这类用户进行定期回访，了解他们的使用体验和遇到的问题，针对性地解决，提高他们的活跃度。

（二）对作者聚类的4个类别，通过分析数据可以发现每个组的特点以及相应建议如下如下：

1. 组1：作者数量较多，总浏览量、总点赞量和总观完量相对较低，但发布作品日数和创作活跃度较高，作品平均时长适中，可能是新手作者或者是内容质量一般但比较勤奋的作者。

建议：平台可以为这类作者提供内容创作指导，帮助他们提升作品质量，从而提高浏览量和点赞量。同时，可组织一些新作者推广活动，增加他们的曝光机会。

2. 组2：总浏览量、总点赞量、总观完量较高，创作活跃度最高，使用配乐数量和发布作品日数也较高，去过的城市数最多，可能是比较活跃且受欢迎的作者。

建议：将这类作者作为平台的重点扶持对象，给予更多资源倾斜，如推荐位、合作机会等，进一步扩大他们的影响力，并鼓励他们分享创作经验，带动其他作者成长。

3. 组3：各项指标大多处于较高水平，尤其是总浏览量、总点赞量、总观完量和创作活跃度，可能是平台上的头部作者，创作能力强且作品受欢迎。

建议：为这类头部作者提供更高级的创作服务和商业合作机会，帮助他们实现内容变现，同时激励他们持续创作优质内容，维持高人气。

4. 组 4：各项指标大多处于中等偏低水平，人数较多，可能是平台上的普通作者，有一定的创作能力但尚未获得高人气。

建议：针对这类普通作者，平台可以开展培训课程和创作激励计划，帮助他们提升创作能力。同时，优化平台的推荐算法，使他们的作品有更多机会展示给目标受众，逐步积累人气。

# 参考文献

[1] 王华, 王治和, 王平. Web 用户聚类研究 [J]. 甘肃联合大学学报, 2010, 24(1): 79-82

[2] 赵文兵, 朱庆华, 吴克文, 等. 微博客用户特性及动机分析——以和讯财经微博为例 [J]. 现代图书情报技术, 2011, (02): 69-75.

[3] 薛云霞, 李寿山, 阮进. 微博中个人与非个人用户分类方法研究 [J]. 山西大学学报 (自然科学版), 2015(2). DOI: 10.13451/j.cnki.shanxi.univ(nat.sci).2015.02.002.

[4] 贺超波, 杨镇雄, 洪少文, 等. 应用随机游走的社交网络用户分类方法 [J]. 计算机科学, 2015(2). DOI: 10.11896/j.issn.1002-137X.2015.2.042.

[5] 蒋翠清, 宋凯伦, 丁勇, 等. 基于用户生成内容的潜在客户识别方法 [J]. 数据分析与知识发现, 2018(3). DOI: 10.11925/infotech.2096-3467.2017.0849.

[6] 凌海峰, 刘业政, 杨善林. 基于蚁群算法与 K-means 算法相结合的 Web 用户聚类 [J]. 情报学报, 2009, 28(1): 105-108

[7] 李纲, 周华阳, 毛进, 陈思菁. 基于机器学习的社交媒体用户分类研究 [J]. 数据分析与知识发现, 2019, 3(8): 1-9

[8] 吴海涛, 应时. 基于信息内容和拓扑关系的社会媒体用户兴趣分类 [J]. 计算机科学, 2015, 42(4): 185-189

[9] He S, Wang H, Jiang Z H. Identifying user behavior on Twitter based on multi-scale entropy [J]. IEEE, 2014. DOI: 10.1109/SPAC.2014.6982720.

[10] 陈洲, 陆南. 基于 K-Means 算法的 Web 日志用户聚类研究 [J]. 计算机与数字工程, 2020, 48(3): 643-647

[11] 高大菊. 基于计算机大数据分析的社交媒体用户行为挖掘与情感分析研究 [J]. 信息记录材料, 2024, 25 (10): 118-120.

[12] 刘玲, 刘茂芸, 黎丽霞. 对不同抖音用户行为的分析研究 [J]. 科技创新导报, 2021(13). DOI: 10.16660/j.cnki.1674-098X.2103-5640-9275.

[13] 谢敏. 抖音算法与爆款短视频打造研究 [J]. 传媒, 2022(24). DOI: 10.3969/j.issn.1009-9263.2022.24.018.

[14] 赵辰玮, 刘韬, 都海虹. 算法视域下抖音短视频平台视频推荐模式研究 [J]. 出版广角, 2019(18). DOI: 10.16491/j.cnki.cn45-1216/g.2.2019.18.023.

[15] 那晓婷. 智能算法下抖音用户的困境及突围路径探索 [J]. 新闻研究导刊, 2022(10). DOI: 10.3969/j.issn.1674-8883.2022.10.002.