









布可能呈现从低等级到高等级逐步减少的分布状况。

分析: 利用所采集的3个主流直播平台用户交互数据, 建立用户经验等级分布模型. 如图8所示, 用户等级主要集中在低等级部分, 且用户数量随着等级的增加而逐渐减少. 大部分用户主要集中于15级以下, 这也符合用户使用在线直播平台的娱乐目的: 趋向于直接观看直播和参与实时评论, 而较少赠送高额礼物的特点. 所赠送的礼物总价值越多等级上升越快, 经验等级为15级以上的用户虽然所占比例不多但却是平台直播间的忠实粉丝用户群体, 直接影响主播的收益高低.

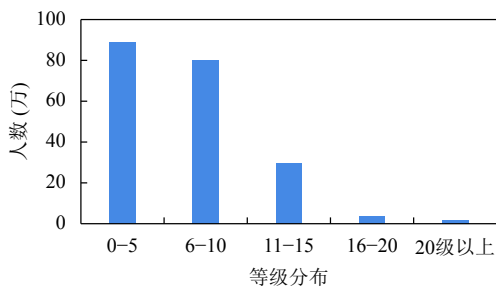


图8 用户等级分布情况

## 2.2 用户行为驱动的用户理解

网络直播作为网络时代的一种新风尚, 对人们的生活侵入程度越来越强. 用户群体往往对直播关注度高, 收看时间长, 弹幕发送频繁, 并通过赠送大量虚拟礼物满足互动消费需求, 自主性强. 因此, 用户行为相比较用户属性而言具有较高的灵活性与不确定性. 对于用户行为, 本文综合大量数据, 寻求其普遍规律, 做出如下假设:

(1) H2a, 用户发送弹幕的频率随着时间的增长而逐步降低.

(2) H2b, 用户发送弹幕的长度更加短小并呈现集群分布的特点.

依据: 不同于微博文本, 用户所发表的弹幕评论通常小于32个字, 结构更加自由、形式更加多样、内容更加丰富, 具有更强的实时性. “在网上没人知道你是一条狗”, 这句话生动地反映了网络的虚拟性和匿名性. 匿名机制的存在, 使得用户可以肆无忌惮地发表自己的看法, 参与直播的实时评论, 因此弹幕更新相当频繁. 另一方面, 弹幕在表达个人的观点的同时利用直播屏幕上的滑动效果分享给其他用户, 这也限制弹幕的有效时长. 弹幕这一讲究快捷的特点, 无疑加速了短文本

的产生. 根据中国互联网络信息中心公布的第38次全国互联网发展统计报告显示, 我国网民以10~39岁的年轻群体为主. 因此, 在内容语义上, 弹幕更贴近网络化用语以及口语, 多使用特殊表达形式, 如数字或者表情的组合, 这就满足了用户群体实时弹幕互动的需求.

分析: 为了探究用户发送弹幕频率的分布情况, 我们对间隔时间在10分钟以内 (以每分钟为间隔单位) 的弹幕数量进行了统计. 如图9所示, 说明用户发送弹幕频率高, 在3分钟内弹幕频率均高达百万人次级别, 并且随间隔时间增长而逐渐减少, H2a得证.

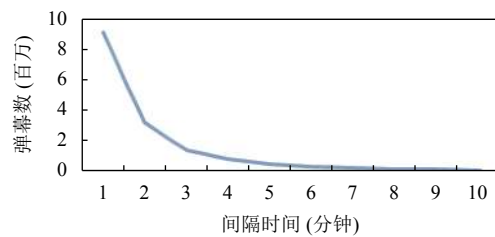


图9 用户发送弹幕频率情况

同理, 为探究弹幕长度特征, 我们统计了弹幕长度分布情况. 如图10所示, 弹幕长度以短文本为主, 1~5字左右的弹幕最为常见, 且随着文本长度的增长弹幕数量出现了明显的减少, 总体呈现出弹幕集群分布于短小处的特征, 这与H2b假设相吻合.

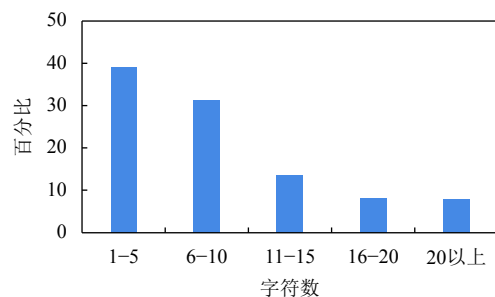


图10 弹幕长度分布情况

弹幕总体呈现出发送频率高, 简洁多样, 即时性强的特征, 一定程度上也切合了当前弹幕文化碎片化和互动性强的特点.

(3) H2c, 用户活跃度与用户历史发送弹幕量、打赏金额以及互动天数正相关.

依据: 首先对活跃用户进行定义: 在直播时间段中, 参与在线网络直播互动的用户定义为活跃用户.

作为直播平台的生命线, 用户活跃度是用户黏性

的反应,直接决定着一个直播平台的商业价值.用户活跃度的具体量化指标往往因应用场景而异,可以根据用户来访、互动情况、核心功能使用频率等综合确定.例如:吴慧等<sup>[18]</sup>从用户背景、社交关系、发表内容质量及社交行为4个方面来设计微博用户活跃度指标.张效尉等<sup>[19]</sup>从某社交网站用户群数据中,选取在一段时间内社交网络用户群中用户平均发布消息的数量作为衡量社交网络用户群用户活跃度指标.王锦坤等<sup>[20]</sup>在设计基于协同过滤的推荐算法时,提出用户活跃度与用户浏览项目数量成正比的模型.本文选择用户发送弹幕量、用户打赏金额以及用户互动时长(天数)作为直播平台用户的活跃度指标,主要出于这样的考虑:弹幕是直播用户对主播人与主播内容的评论,发送弹幕是直播用户观看的体验与参与度的表现;打赏金额是直播观众对主播服务质量的认可与肯定,是深度参与的重要体现;互动时间越长说明该直播用户对此直播平台提供的服务越认可,也是用户黏性的体现.

为此,本文假设用户历史发送弹幕量、打赏金额以及互动天数影响用户活跃度.

分析:为了提取更多的用户行为特征,我们统计了连续两周时间内用户在线天数、发送弹幕数量、赠送礼物总额、最大发送弹幕数量、最大赠送礼物价值以及这段时间内最大连续在线天数.由于用户活跃度存在个体差异性,故本文假设这一组数据能够用来表征不同的用户活跃度.为了验证用户活跃度和这些变量之间的相关关系,我们对这一组数据进行了主成分分析.主成分分析法能将较多的数据变量通过删除变量间的重叠部分得到较少的综合变量,从而将用户活跃度更加准确直观地表示出来.从表2可以看出,共有5个主成分.由于第一主成分相比其它主成分和这些变量之间具有更突出的相关关系,故抽取出第一主成分F1作为用户活跃度的表征.根据主成分分析结果,可以得到在线天数和用户活跃度、弹幕数量、礼物总量、最大弹幕量、最大礼物量、最大连续在线天数的相关系数分别为0.724、0.762、0.218、0.838、0.212和0.792.因此,对于用户活跃度满足 $F1=0.724 \times T1+0.762 \times T2+0.218 \times T3+0.838 \times T4+0.212 \times T5+0.792 \times T6$ , $T1-T6$ 分别代表在线天数、弹幕数量、礼物总量、最大弹幕量、最大礼物量、最大连续在线天数.在线天数越长,发送弹幕量越大,赠送礼物越多,F1的值越大,用户活跃度越高,这个结果与H2c假设吻合.

## 2.3 用户属性与用户行为相融合的用户理解

从H1a以及H1b可以看出用户在线观看直播的时间以及地域方面有明显的时空分布特征,从H2a以及H2b可以看出弹幕长度以及频率特征,从H2c可以得到用户平台活跃度相关因素.在线直播主要给用户提供了即时性互动(即弹幕礼物)的全新观看直播体验.由于在线直播集中在深夜时段以及具有消费性的特点,用户人群具有明显的时空分布特征.弹幕礼物作为在线直播互动体验的媒介,承载着用户行为的产出.而用户平台活跃度作为用户互动的产物和在线直播平台衡量一个客户价值的重要标准,自然也是刻画用户人群的重要因素.因此,本文认为用户模型的构建可从用户属性和用户行为两个方面着手.

表2 成分相关性分析表

成分	1	2	3	4	5
在线天数	0.724	-0.053	-0.631	0.273	0.019
弹幕数量	0.762	-0.209	0.589	0.008	0.168
礼物总量	0.218	0.972	0.075	0.002	-0.002
最大弹幕量	0.838	-0.202	0.475	0.028	-0.176
最大礼物量	0.212	0.974	0.079	0.001	0.003
最大连续在线天数	0.792	-0.066	-0.535	-0.288	0.006

## 3 结论与展望

用户属性和用户行为共同驱动用户理解,构成网络在线直播背景下的用户模型.

用户属性方面,观众观看网络直播的时间选择上具有周期性变化规律,并且得出这与主播的开播时间与日常作息密切相关的结论.在沿海发达城市的用户相对集中,特别是人口多,经济发达同时拥有众多外来人口的城市,在观看直播的观众中占有较大比重.根据观看直播平台分布情况来观察,如今观众选择的直播平台具有“网页端为主移动端为辅”的主要特点.

而在用户行为方面,用户在直播平台的互动方式居多,其中以发送弹幕为主.用户所发送的弹幕内容具有“短文本,高频率”的特点.弹幕内容网络化,接近口语,贴切生活同时也容易产生误解与冲突.根据用户等级可以判断出赠送礼物的比重较少,并且用户等级从低到高呈减少的分布状态.而这两种互动方式与在线天数共同支持用户活跃度的理论推断.

为此,本文针对具体的活跃度给出定量的活跃度计算指标,提取影响用户活跃度的主要特征利用PCA技术根据影响力最大化方向提取主成分作为用户

活跃值的参考。

本次的实验依旧有存在部分不足。首先,在抓取数据方面,由于所使用的工具的局限性,只抓取到监测期间有进行互动的用户资料信息,而这一点就忽略了在监测期只进行观看没有参与互动的用户;其次,用户信息只停留在进行互动时的状态,互动后获得收益状态没有跟踪到;再者,选取的数据源较为局限,可能会出现数据偏差的问题。

### 参考文献

- 1 Geerts D, Vaishnavi I, Mekuria R, *et al.* Are we in sync? Synchronization requirements for watching online video together. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY, USA. 2011. 311.
- 2 Liu LL, Suh A, Wagner C. Investigating communal interactive video viewing experiences online. In: Kurosu M, ed. Human-Computer Interaction. Cham: Springer, 2016.
- 3 李君贤. 网络直播中弹幕语言暴力机制的形成与消解. 西部学刊, 2016, (10): 59–60.
- 4 He M, Ge Y, Chen EH, *et al.* Exploring the emerging type of comment for online videos: Danmu. ACM Transactions on the Web, 2018, 12(1): 1–33. [doi: [10.1145/3180440](https://doi.org/10.1145/3180440)]
- 5 Chen XP, Chen JY, Ma L, *et al.* Fine-grained video attractiveness prediction using multimodal deep learning on a large real-world dataset. International World Wide Web Conferences Steering Committee. Geneva, Switzerland. 2018. 671–678.
- 6 Chung CT, Hsiung HK, Wei CK, *et al.* Towards personalized video summarization using synchronized comments and Probabilistic Latent Semantic Analysis. 2014 IEEE 3rd Global Conference on Consumer Electronics. Tokyo, Japan. 2014. 414–415.
- 7 Shen Y, Chan H, Hung I. Let the comments fly: The effects of flying commentary presentation on consumer judgment. International Journal of Systems Science, 2014, 12(12): 1469–1475.
- 8 Schultes P, Dorner V, Lehner F. Leave a comment! An in-depth analysis of user comments on YouTube. 11th International Conference on Wirtschaftsinformatik. Leipzig, Germany. 2013. 659–673.
- 9 Hu M, Zhang ML, Luo NA. Understanding participation on video sharing communities: The role of self-construal and community interactivity. Computers in Human Behavior, 2016, 62: 105–115. [doi: [10.1016/j.chb.2016.03.077](https://doi.org/10.1016/j.chb.2016.03.077)]
- 10 邓扬, 张晨曦, 李江峰. 基于弹幕情感分析的视频片段推荐模型. 计算机应用, 2017, 37(4): 1065–1070, 1134. [doi: [10.11772/j.issn.1001-9081.2017.04.1065](https://doi.org/10.11772/j.issn.1001-9081.2017.04.1065)]
- 11 郑颺颺, 徐健, 肖卓. 情感分析及可视化方法在网络视频弹幕数据分析中的应用. 现代图书情报技术, 2015, 31(11): 82–90.
- 12 Wu B, Zhong EH, Tan B, *et al.* Crowdsourced time-sync video tagging using temporal and personalized topic modeling. Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA. 2014. 721–730.
- 13 Xu LL, Zhang C. Bridging video content and comments: Synchronized video description with temporal summarization of crowdsourced time-sync comments. AAAI Publications, Thirty-First AAAI Conference on Artificial Intelligence. 2017. 1611–1617.
- 14 肖云鹏. 在线社会网络用户行为模型与应用算法研究[博士学位论文]. 北京: 北京邮电大学, 2013.
- 15 樊鹏翼, 王晖, 姜志宏, 等. 微博网络测量研究. 计算机研究与发展, 2012, 49(4): 691–699.
- 16 Thelwall M, Sud P, Vis F. Commenting on youtube videos: From guatemalan rock to el big bang. Journal of the American Society for Information Science and Technology, 2012, 63(3): 616–629. [doi: [10.1002/asi.21679](https://doi.org/10.1002/asi.21679)]
- 17 中国互联网络信息中心. 中国互联网络发展状况统计报告. 2018.
- 18 吴慧, 张绍武, 林鸿飞. 微博社交网络的用户影响力评价方法. 中文信息学报, 2017, 31(4): 184–190. [doi: [10.3969/j.issn.1003-0077.2017.04.026](https://doi.org/10.3969/j.issn.1003-0077.2017.04.026)]
- 19 张效尉, 余云霞, 王伟. 社交网络群中用户活跃度分析与预测. 西南师范大学学报(自然科学版), 2018, 43(12): 115–121.
- 20 王锦坤, 姜元春, 孙见山, 等. 考虑用户活跃度和项目流行度的基于项目最近邻的协同过滤算法. 计算机科学, 2016, 43(12): 158–162. [doi: [10.11896/j.issn.1002-137X.2016.12.028](https://doi.org/10.11896/j.issn.1002-137X.2016.12.028)]