

厦门大学新媒体丛书

主 编：林升栋 副主编：苏俊斌

中国微博活跃用户 研究报告



视频



文字



图片



音乐

林升栋 ▸ 编著

- 个体的微博使用习惯与**消费者行为中的AIO**
- **局外观察**与微博内容分析解读方式
- **2600**多位微博用户样本 重量级的微博用户**250**位



厦门大学出版社
XIAMEN UNIVERSITY PRESS

国家一级出版社
全国百佳图书出版单位

厦门大学新媒体丛书

中国微博活跃用户 研究报告



视频



文字



图片



音乐

林升栋 编著



厦门大学出版社
XIAMEN UNIVERSITY PRESS

国家一级出版社
全国百佳图书出版单位

图书在版编目(CIP)数据

中国微博活跃用户研究报告/林升栋编著. —厦门:厦门大学出版社,2014.4
(厦门大学新媒体丛书)

ISBN 978-7-5615-5003-8

I. ①中… II. ①林… III. ①互联网络-传播媒介-研究报告-中国
IV. ①G206.2

中国版本图书馆 CIP 数据核字(2014)第 069064 号

厦门大学出版社出版发行

(地址:厦门市软件园二期望海路 39 号 邮编:361008)

<http://www.xmupress.com>

xmup@xmupress.com

厦门集大印刷厂印刷

2014 年 4 月第 1 版 2014 年 4 月第 1 次印刷

开本:720×970 1/16 印张:9.75 插页:1

字数:165 千字 印数:1~2 000 册

定价:30.00 元

本书如有印装质量问题请直接寄承印厂调换

自序

2012年年底,“第二届新浪高校微博协会周年峰会暨中国高校社团与新媒体结合高峰论坛&教育盛典”,好长的一个会议名称,邀请我在北京做了一个简短的报告——《新媒体时代的人才需求——数据分析师》。期间我毫不掩饰地宣称,数据分析师将是21世纪最贵的人才。在超市、银行业、航空业、酒店业、通信业等传统行业之后,电子商务、社交媒体都表现出对数据分析人员的强烈需求。但目前国内项目数据分析领域的考核太注重对分析人员的技术要求,忽略了行为科学的重要性——不仅数据分析结果的解读需要借助行为科学,数据如何分析也需要行为科学的指引,决策时间有限的情况下,更不能东一锄头西一钁头地“乱刨”数据。看到与人相关的一连串单独的数据后,“还原”出一个“鲜活人”,这种能力恰是最难且最缺乏的。

在大数据时代,中小样本的研究为很多团队和专家所忽略,本书研究的微博活跃样本却是随机抽取的中小样本。我的一位同事,质疑这一研究跟传统的抽样研究并无不同,或者在他眼里,



大数据时代的研究似乎必须以“大而全”来彰显意义与价值。微博上也有人吐槽,业界有些部门负责人在大数据侦测指标还未确立的情况下,要求自己的团队不准做小样本研究。然而,北京大学的刘德寰教授却坚持认为大数据挖掘有风险:计算速度;大数据得出的假规律;断裂数据;缺失数据等。相当一部分数据分析师的现有工作是通过中小数据进行数据的解读与建模工作。在某些阶段,小样本研究通过解剖麻雀式的分析,可能挖掘出许多有价值的信息,为大数据分析提供方向和基础。

新浪博客上有个叫 fangmonk 的博主,写了一篇博文《和 90 后 mm 聊夜店生态圈》(http://blog.sina.com.cn/s/blog_537809c30101e0bc.html),话题从淘宝假豪车钥匙开始,博主说某些屌丝在淘宝上买假名车钥匙,冒充高富帅去夜店泡 mm。这位 90 后的 mm 说博主胡编,假钥匙的重量、材质无法达到以假乱真的程度。而且,即使假钥匙的仿真度很高,仍可通过戴的名表、点酒的数量(高富帅一点都是一排酒,没钱可撑不了这个场面)、银行卡的颜色和号码等来判断。当博主感叹夜店的学问也这么多时,这位 90 后 mm 回答“每天我要花很多时间了解品牌的潮流,要了解时尚服饰、汽车、手表、金融,要懂得时间管理,投入产出分析,要懂一些心理学……”,雷到我了。不管是大数据,还是小样本,我们难道不都在追求数据背后的意义与应用价值么?

从数据中挖掘出金子的故事很多,最广为流传的,莫过于沃尔玛的啤酒和尿片了。沃尔玛的分析人员发现在下午的某个时段里,消费者常常一起购买啤酒和尿片,他们通过观察和走访了解到,在美国有孩子的家庭中,太太经常嘱咐丈夫下班后要为孩子买尿片,丈夫们在买尿片时又顺手带回自己钟爱的啤酒,啤酒和尿片两个风牛马不相及的物品就这样走在一起。沃尔玛的工作人员因此决定将啤酒和尿片摆在一起,以方便顾客购买。2004 年天气预报飓风伊万将袭击美国佛罗里达州,沃尔玛早已在飓风途经地区的超市大量配送甜脆饼,他们根据以往的销售数据发现飓风后人们急需甜脆饼。前几个月,互联网上流传的另一个故事,美国一名父亲投诉

一家超市,称他 17 岁的女儿收到婴儿尿片和童车的优惠券。之后这个愤怒的父亲发现女儿的确怀孕了,据说超市的促销源于对他女儿 facebook 日志的分析。

有人说,在大数据时代,数据之间的相关关系足够说明问题。我个人也赞同此说,但我仍坚持在相关的基础上应尽力探求,至少去推测因果关系。举个例子来说,一个小孩在长高,一株小树也在长高,二者之间有相关关系,但没有因果关系。数据分析人员将很多人口统计变量和行为变量放在一起求相关,就好比在河边放了很多根鱼杆,当你放 100 条鱼杆的时候,总有几支鱼杆会钓到鱼,这完全可能是偶然的。当人们知其然,而且也知其所以然时,不仅决策会更有底气,应运而生的方案也更贴近人性需求。运用小样本数据来探索因果关系是比较经济可行的。

我曾经给某地区移动授课,指导他们对一个小样本客户群进行解剖麻雀式的分析。每个移动用户的数据都包含着丰富的内容:人口统计方面的资料、通话时间、通话时长、主叫/被叫、彩信使用、短信数量、漫游、移动上网、投诉等。这些数据,除了可以分析用户的通话行为模式外,甚至可以用于推测用户的性格与职业。举例来说,如果该用户被叫时聊得比较长,主叫时比较短,通常一分钟就结束了,我们可以推断该用户比较节俭;如果该用户在每年 9 月 10 号都收到大量短信,我们可以推测其职业是教师;如果一个人在节假日收到的短信条数很多且远多于他发出的短信,说明该用户有较高的社会地位。随着国内大量模拟电视转换成数字电视,也产生大量收视数据流。这些数据流不仅可以用于确定每个家庭喜欢收看的节目,也可以根据用户的收视行为来推送一些广告。有些用户可能晚上 10 点关了电视睡觉,但在凌晨两三点重新开机,说明这些用户睡眠不佳,或许可以在法律允许的情况下推送安眠药片的广告信息。

2013 年 4 月《财经国家周刊》上的一篇文章声称,每个月 facebook 上分享的内容达 30 亿条, twitter 每天产生 12TB 的信息。在中国,数亿用户在新浪微博上制造了大量数据,峰值数字是每分钟 73 万条数据,这些数据

近 80%是由个人用户产生的。社交媒体用户喜欢在网上分享自己的喜好,晒幸福,分享自己的购物行为,甚至像郭美美那样炫富。及时捕捉这些数据,就意味着你将更了解用户需要,因而能提供更好的用户/顾客体验。这些数据庞大、繁多、复杂,尽可深入挖掘,进行深刻洞察,很多人迷信技术,认为技术可以解决一切问题,我个人对此存疑,尽管我相信算法模型力量强大,但社交媒体上的数据大量是非结构性的,包括视频、语音和图片等,即使是单纯的文字,读起来依然仁者见仁、智者见智。

社交媒体日复一日累积个人的轨迹,记录并展现一个个真实的自我。很多时候,庞大数据就像一个麻乱的大线团,只有找出线头,拎出一个个的网络人格,才能清晰地再现出一个个鲜活的自我,我相信,小样本数据的深度解读有助于这一“再现”。基于这一信念,我们对中国微博活跃用户在日志上表现出来的生活形态进行深入定性的解读,结合肉眼判断与结构化数据两种方法,尝试描摹微博活跃用户不同细分人群的特征。

本书分三个层次(篇)展开。

第一篇是“微博活跃用户群体研究”,通过抽样的方式,从新浪微博上抓取 2400 多个活跃用户,结合微博风云的数据以及 200 多个用户日志的定性解读,对中国微博活跃用户进行分群及生活形态描摹。这是我指导宋玉蓉、梁玉麒、赵成栋、高尚、翁路易五位研究生共同完成的大型研究报告,其中第三节“基于个人影响力的分群描摹”主要由梁玉麒同学撰写,第四节“基于使用动机的分群描摹”主要由宋玉蓉同学撰写,第五节“基于发布微博时间的分群描摹”主要由赵成栋同学撰写,第六节“基于网络身份识别的分群描摹”主要由高尚同学撰写,第七节“人口统计差异与极端值个体用户研究”主要由翁路易、高尚同学撰写。

第二篇是“奢侈品微博活跃粉丝群研究”,在 LV、GUCCI 等六大奢侈品品牌官微中随机抓取约 200 个活跃用户,采用同样的方式进行解读,该部分文章由我指导的研究生李光媛的硕士毕业论文修改而成。

第三篇是“杜蕾斯官方微博活跃粉丝群研究”,通过非随机抓取的方式

获得 200 多个活跃用户,该部分文章的基础是由我指导的本科生欧阳张蓓(组长)、吴蓉蓉、夏英英、高山晶、邵舒、刘雅梦、李新宇、李睿和李昕霖的一份课堂期末作业。

三个层次,从大的群体,到某一种特殊消费品的粉丝群体,这个群体拥有强大的购买力,再到某一个企业官微的粉丝群体分析与管理,层层展现小样本分析的魅力。本书锁定“活跃用户”,是因为在社交媒体上如果不“发声”或者不“上来”,我们就无法入手研究。每种媒体都拥有特定的受众群,只有活跃用户,才有较高的传播价值与商业价值。

与各种机构发布的报告不同,我们更注重意义的捕捉与商业应用的思考,而不仅仅是描述数据统计结果。举例来说,微博上经常更换头像的用户,可能很关注其公众形象;经常@别人,该用户可能期待得到别人的回应,被经常@的人可能跟该用户有密切的联系;被@较多的人,可能是一个圈子中有影响力的人物;经常转发中奖促销信息的用户,比较相信小概率事件,现实生活中可能也会购买体育彩票;提供较详细个人信息的用户,隐私保护的观念不强;经常发布自拍照的人,可能比较自恋;关注对象中明星比率较高的人,易受名人代言的影响;关注特别多而粉丝很少的用户,很可能是僵尸粉;关注很少而粉丝众多的,通常是明星;主要采用移动终端来发布微博的用户,其生活形态与采用 PC 终端发布的用户有较大差别;等等。我们设定了远大的目标,尽管因为时间、财力、人力各方面的局限,现有的研究成果还有诸多局限。我将这两年来积累汇总出版,抛砖引玉,旨在引起更多的关注与探讨。

在这里,我还要特别声明,无论何种研究和商业行为,都要注重保护用户的隐私。我在学校里讲授“顾客关系管理”(Customer Relationship Management)课程,这门课很依赖于客户信息的搜集。很多好东西传到中国后都走了样,小区物业管理的人员将住户信息卖给外面的二手房公司,然后每天我们都会收到很多骚扰电话,不厌其烦地询问是否要出租或出售。甚者,顾客的信息在未经顾客本人许可的情况下被数度转卖。更致命

的是,获取数据的公司根本不会用脑袋去思考数据中蕴藏的意义,他们宁可聘来一些业务员将所有的手机号都拨通一遍,重复着那些毫无创意的说辞,期待着万分之一的顾客会上钩。这种做法已经相当普遍,严重扰乱顾客的生活,引起普遍的反感。在大数据时代,我们更应当尊重顾客,慎用数据,悄悄地去“猜”顾客的心,给顾客惊喜,不动声色地改善顾客体验。

林升栋

2013 年 8 月 10 日

目 录

第一篇 微博活跃用户群体研究/1

- 第一节 研究背景/2
- 第二节 抽样及分析方法/6
- 第三节 基于个人影响力的分群描摹/10
- 第四节 基于使用动机的分群描摹/32
- 第五节 基于微博发布时间的分群描摹/45
- 第六节 基于网络身份识别的分群描摹/59
- 第七节 人口统计差异与极端值个体用户研究/78

第二篇 奢侈品微博活跃粉丝群研究/85

- 第一节 研究背景/86
- 第二节 抽样及分析方法/92
- 第三节 整群特征和分群描摹/98
- 第四节 超级活跃粉丝的生活形态/113
- 第五节 研究局限与未来展望/126

第三篇 杜蕾斯官方微博活跃粉丝群研究/129

- 第一节 企业背景与微博简介/130
- 第二节 抽样说明/132
- 第三节 基于微博发布时间的分群描摹/132



第四节	基于微博发布平台的分群描摹/136
第五节	杜蕾斯微博粉丝主要地域分布/137
第六节	粉丝星座分析/139
第七节	意见领袖特征分析/142

后记	/146
----	------

第一节 研究背景

第二节 抽样及分析方法

第三节 基于个人影响力的分群描摹

第四节 基于使用动机的分群描摹

第五节 基于微博发布时间的分群描摹

第六节 基于网络身份识别的分群描摹

第七节 人口统计差异与极端值个体用户研究

第一节 研究背景

Web3.0 来临,脸书(facebook)、推特(twitter)、博客、视频、照片分享及 LBS^① 类社交网站等自媒体(we media)平台^②和社交媒体放大了受众的声量并促进社会交往,使受众与媒介的互动发生质变。近年来,社交媒体用户爆炸性增长,社交媒体逐渐成为欧美学者研究的热点。在中国,网民 40% 的上网时间花在社交媒体上,登录主流社交媒体已经成为许多网民日常生活不可或缺的一部分。了解中国的社交媒体用户极为重要:中国社交媒体市场规模超过美国,全球最大;中国的社交媒体用户极为活跃,社交意愿强烈,80% 以上的用户拥有多个社交媒体账户,日本只有 39% 的用户拥有多个账户;中国大多数用户只在本土社交媒体上注册,形成相对独立的庞大用户群。^③ 相较于网民整体,社交媒体用户更年轻,受过更好的教育且收入更高。无论就用户数量而言还是就质量而言,社交媒体用户群均有重要的研究价值。

中国目前主流社交媒体中,微博发展最为成熟,使用频率高,覆盖率高于其他类型社交媒体。据 CNNIC 的报告,截止 2012 年,中国微博用户数已超过 3 亿,19 岁及以上的微博用户占中国互联网用户的 88.81%,且近七成用户仅拥有一个微博账号,用户黏着度高,微博成为网民重要的信息获取渠道。^④ 从用户增长速度来看,微博用户的增长率在 2011 年上半年达到高峰,2012 年 6 月底,其在网民中的渗透率已过半(50.9%),用户规

① LBS=location-based services,基于位置的服务。

② Shayne Bowman & Chris Willis, *We media: How audiences are shaping the future of news and information*, The Media Center at The American Press Institute, 2003.

③ Cindy Chiu, Chris Ip & Ari Silverman, *Understanding social media in China*, McKinsey Quarterly, 2012(2): 78~81.

④ 中国互联网数据中心(DCCI). 2012 中国微博蓝皮书. 2013.

模已进入平稳增长期,手机端微博用户的增长幅度仍然强劲。^① 2012年,AC尼尔森通过在线消费者调查^②年龄大于10岁的中国主流社交媒体用户,发现:微博触达97%的主流社交媒体使用者;占总体17%的重量级在线消费者,他们的网购花费占总网购额约一半,他们中有近八成人同时使用微博和SNS;新浪微博在各级城市的覆盖率超过腾讯微博、人人网,绝对优势超过开心网、网易微博、搜狐微博及豆瓣。同年,Hitwise发布全年用户上网总时间统计,新浪微博位列第三,成为用户登陆时长最长的社交网络平台。^③ 在2011年年底,新浪微博用户日均使用时长达1小时。^④ 按总浏览时间计算,新浪微博占微博市场87%的份额。^⑤ 新浪微博在社交媒体中(甚至网络平台中)位于核心地位,这毋庸置疑。本书选择新浪微博为社交媒体代表,采用样本抽样的方式,对其活跃用户群体进行深入研究。

对微博用户群体的研究,多将之视为整体,比较粗糙;或独立研究某一细分群体,比较零散。对一个新媒体用户群进行深入的研究,首先要将其进行系统分群,“物以类聚,人以群分”,对各细分用户群的轮廓描述与洞察,不仅有助于社交媒体的传播管理,还为广告商的精准营销奠定了理论基础和方法基础。目前,尽管业界有研究者采用网络调查的方式,或者采用个别的微博指标,对微博用户进行分群,但大多数研究仍沿用传统媒体用户研究的思路来细分新媒体用户。以往的用户研究常通过问卷调查的方式,借助用户“显性特征”(即人口统计特征,如年龄、性别等),辅以“隐性特征”(如人格、价值观等)来描述细分群体的轮廓。这种调查的可靠性有

① 中国互联网络信息中心(CNNIC). 中国互联网络发展状况统计报告. 2012.

② 尼尔森. 中国社交媒体受访用户研究报告. 2012.

③ 199IT中文互联网数据资讯中心. 2012年新浪微博数据 - 数据信息图. 2013-02-20. <http://www.199it.com/archives/95537.html>.

④ Techweb. 新浪微博称用户日均使用时间达1小时. 2011-12-19. <http://www.techweb.com.cn/data/2011-12-19/1132469.shtml>.

⑤ I美股. 未来资产发布《新浪微博研究报告》. 2011-01-28. <http://news.imeigu.com/a/1296195686757.html>.

赖于受访者的主观估计、记忆与判断。而在微博等社交媒体上,用户的“隐性特征”更为重要,“显性特征”反而沦为辅变量。微博用户愿意在网上“晒”幸福、生活及各种观点,这为我们深入了解用户的价值观与生活方式提供了角度。要在社交媒体研究上取得成功,就要准确解读用户的“隐性特征”。借用网络技术,可以精确获取用户使用行为的数据流,但对其的质化解读还有待努力。本文结合微博风云工具对每个微博用户进行量化分析,花费大量人力对抽样用户的微博日志进行主观式解读与讨论,从而全面深入地了解微博用户各类族群的生活形态。

社交媒体吸引了广告商的广泛关注,正成为他们的新战场。全球最大的广告主宝洁,其首席执行官 McDonald 对华尔街的分析师说,宝洁想要精准的广告预算,从这些年看,Facebook 与谷歌的广告效果要比传统媒体的好,但宝洁现在的广告费却大多用在传统媒体上。^① 美国口碑营销协会将社交媒体理解为“网络口碑活动”,将其看作实现口碑影响的引擎或工具集。^② 社交媒体用户规模庞大,广告商偏爱那些收入较高且不排除广告的用户群。麦肯锡一项针对 5 700 名中国互联网用户的调研发现,不同的社交媒体确实吸引不同类型的用户:喜欢新浪微博的消费者大多收入较高,其月收入在 8 000 元(约 1 300 美元)以上,多数人居住在一线城市。^③ 《2012 年新浪微博用户发展报告》(有效样本量达 10.1 万)显示,新浪微博国内用户量,广东、北京、江苏排名前三,三地均为中国经济最发达的地区,且总体不排除广告的比例超过 76%。^④ 对广告商而言,对新浪微博用户群的研究

① 新浪财经. 宝洁发现 Facebook 可免费打广告 将裁员 1600 人. 2012-01-31. <http://finance.sina.com.cn/stock/usstock/c/20120131/093711279233.shtml>.

② Word of Mouth Marketing Association. WOMMA Guidebook. 2013. <http://www.womma.org/resources/Online-resources>.

③ Cindy Chiu, Davis Lin & Ari Silverman. China's social-media boom. McKinsey & Company, 2012.

④ 新浪公司、北京大学市场与媒介研究中心、第一象限公司. 2012 年新浪微博用户发展报告. 2012.

也有重要的意义。

然而,社交媒体上分布着大量的非活跃用户和“僵尸粉”,对广告商的运营来说,这些都是无效的群体。《华尔街日报》报道,香港大学2012年进行了一项研究,抽取约30 000名用户组成的随机样本,57%的研究对象在调查期间未发过微博,这说明,其若不是“非活跃用户”,就是“僵尸账户”(即营销机构为真实账户操纵关注者数量而创建的账户)。该研究还表明,只有约3 000万用户一周内会发一篇原创帖。据新浪微博官方统计,截止2011年,日活跃用户有4 620万人,换言之,用户中“非活粉”或“僵尸粉”的比例不低。研究这两类粉丝的价值不大,也不适合于本书所用的日志分析法,因此本书只研究新浪微博的活跃粉丝。法国研究公司Semiocast将Twitter(与新浪微博功能相近)的活跃用户定义为:在为期三个月的时间内发过信息、关注过另外一位用户或改变过头像的用户。^①使用日志研究方法,要求对象用户每天都写大量日志,这才有分析和研究的价值,所以本书所谓的活跃粉丝,标准更高。

目前,国内对微博用户进行分群研究的主要是业界人士。麦肯锡针对中国社交媒体用户使用动机和行为分析,认为四类人群较有营销价值:热衷社交型(约占15%,以经营友谊人脉为主)、积极转发型(占15%,原创率极低,但积极转发如笑话等,通常粉丝极多)、安静阅读型(占14%,一般浏览不发表意见)、发表意见型(约占14%,热衷发布个人看法,往往观点强烈,被转发量巨大)。^②新浪与CIC合作的《微博白皮书》用标签将用户兴趣分为三类——娱乐八卦(年轻男女)、情感生活(成年女性)及务实信息(年长男女,尤其男性)。^③该研究发现,不同人口统计特征的人群在新浪微博上的行为方式不尽相同:30岁以下用户是最主要的参与者与跟随者,

① Mozur & Paul. *How Many People Really Use Sina Weibo*. Wall Street Journal, 2012.

② Cindy Chiu, Davis Lin & Ari Silverman. *China's social-media boom*. McKinsey & Company, 2012.

③ CIC、新浪. *CIC·新浪合作微博白皮书—微博引领的中国社会化商业发展与变革*. 2012.

以转发、评论、收发私信为主;31~45岁用户以搜索感兴趣的人或话题为主;46岁以上用户则属于围观者,较多通过增加对他人的关注获取资讯。如前文所述,这些分群在研究方法上仍沿用传统的问卷调查,侧重从性别、年龄等人口统计特征来划分,对用户隐性特征的挖掘仍然较粗浅。

也有学界、业界人士从某个或某几个群体在微博上的日志分析入手研究。群邑智库联合CIC对比“独生代”中的80后、90后在新浪微博中的行为,发现两类群体在消费形态的差异与消费态度上的相似:前者注重个性与品质、享受生活、超前消费;后者属主动型消费者,颇具消费潜力和影响力;两者均喜爱购物,了解各行业的领先品牌。^①他们对用户微博内容进行分析,由于人力所限,分析较浅,且结果局限于两类群体。盛宇以学术类用户这一特定领域为例,完善微博用户的特征指标体系,指出,现有研究集中关注用户显性指标而忽略更稳定的隐性指标。^②黄京华、赵晓初、钟活灵对90后大学生自陈问卷进行数据分析,采用探索性因子分析方法建构不同指向的网络认知指标(情感、社交及工具指向),据此应用聚类分析将90后大学生细分为四种不同类型的网络使用人群。^③综上所述,目前不管是业界还是学界,尚未有人花费大量人力对微博用户的隐性特征深入挖掘,将之与显性特征联结起来,对新浪微博的活跃用户群整体进行系统的分群研究。

第二节 抽样及分析方法

本篇通过对“新浪微博的活跃用户”分层抽样获得2 452个样本(纳入

① CIC、群邑智库. 中国80后90后网论观察白皮书, 2011.

② 盛宇. 微博特定领域用户外在特征研究——以新浪微博学术类用户为例. 情报杂志, 2012(12).

③ 黄京华、赵晓初、钟活灵. 九零后大学生的网络认知定势对其微博应用的影响. 现代传播, 2012(4).