

社会关注度的新闻大数据分析^{*}

——社会、时间与空间的维度

郑 路 吴文兵 葛天任 张 翀

摘要:随着手机移动新闻客户端的爆发式增长和个体用户大量数据的产生,利用大数据挖掘和分析方法对网络移动新媒体进行受众分析尤其是社会属性分析成为可能,这将在很大程度上推动移动传播媒介时代的受众分析研究。本文基于搜狐新闻客户端2014年5月至10月的用户行为抽样数据,探索性地从社会、时间和空间三个维度对移动新闻客户端用户的阅读行为社会关注特征和影响因素进行了分析。研究发现:性别、阶层群体、社会地理空间等因素在新闻关注主题的影响方面存在显著差异;移动新闻客户端用户在不同时间节点所关注的新闻类别也具有明显的差异化特征;“北上广深”等超大城市的用户群体在特定新闻话题上的关注度体现了鲜明的社会性特征。

关键词: 大数据 移动新闻 受众分析 社会关注度

DOI:10.16235/j.cnki.33-1005/c.2018.03.010

一、导 言

2014年,全球移动用户达到36亿,智能手机采用量为26亿,同年12月底,我国手机用户达到5.57亿,我国网民使用手机上网的数量首次超过个人电脑终端数量,成为第一大上网终端。^①移动终端数量的增加,颠覆了传统的新闻媒体传播方式,全面影响人们的日常生活行为。随着近年来网络软硬件基础设施的提高,网民规模的扩大,移动新闻传播规模出现了井喷之势。移动新闻增长幅度在此前3年内(2014年以前)达到66%,移动新闻用户渗透率达到41.4%。^②在2016

年第一季度,人们从新闻客户端获取新闻的方式占64.1%,^③移动新闻客户端已经成为人们获取信息的重要途径,这已经在很大程度上改变了人们的阅读行为。对这一问题的关注和研究,不仅成为新闻传播领域的热点议题,也是社会学、情报学、图书馆学等学科关注的热点。“媒介即讯息”,传播学大师麦克卢汉的经典名言对于互联网大数据时代的最大启示,或许就在于提醒研究者们从网络移动媒介的受众分析中寻找用户使用行为的传播学规律。因为,研究用户的阅读行为特征与偏好,可以让我们更加清晰地了解社会群体在信息获取方面的规律——社会关注度的分布与影响机制,进而促进相关领域的

* 本项目获得搜狐新闻客户端、清华大学社会科学学院和数据研究院合作计划、清华大学自主科研计划(项目号20151080448)和国家社科基金(16ZDA085)资助。

① 李力、丁宁:《国内外移动社交类应用用户信息行为研究进展》,《图书情报工作》2015年第10期。

② 陆瑶:《移动新闻高频化对受众的影响研究——基于访谈的探索分析》,《东南传播》2017年第8期。

③ 2016年一季度移动新闻客户端市场报告:<http://www.askci.com/news/hlw/20160603/17160924854.shtml>。

应用性技术创新。

对社会关注度分布和影响机制的研究,主要集中在个体性格因素、社会因素、环境因素以及系统因素四个方面。^① 具体而言,相关研究从个体用户的性格心理学特征、^②社会习惯、媒介特质和易感性方面讨论了影响社会关注度的多重因素。^③ 尽管一些学者尝试探索性地研究大数据时代的受众分析,^④抑或移动新闻传播的受众分析,^⑤聚焦社会性因素也就是受众群体分析的相关研究仍然缺乏实证经验分析,尤其是在互联网的数据爆炸式增长时代。一方面,相关研究仍然延续了麦奎尔基于电视传播的受众分析理论。当然,麦奎尔理论的高度抽象性为我们今天思考网络媒介的受众行为提供了基础。他的思想中深深带有社会学的分析路径,他认为,在日益复杂的媒体环境下没有单一的对受众进行划分的标准,群体性特征、传播者与使用者之间的张力、社会性因素和时间、空间因素都对受众行为产生了深刻影响。^⑥ 但另一方面,由于传统的调查方法不能获取充分数据,至今研究网络媒介用户使用行为的实证研究仍不充分。大数据方法的普及使获取数据变得更加容易,也为本文提供了研究这一问题的工具和方法。

回顾上述文献,本文尝试采取大数据分析方法对网络受众的新闻关注度进行社会因素分析。本文的具体研究问题为:移动新闻客户端的用户阅读行为规律是什么?不同社会属性的用户阅读行为是否不同?用户的阅读行为在时间和空间上是否有差别?

根据受众分析理论,我们通过数据挖掘的方法采集到了新闻用户的阅读行为数据,从社会、时间和空间的维度对新闻用户的阅读行为进行了大数据分析。通过利用手机号码对搜狐新闻客户端用户进行随机抽样,最终获取个体手机用户18万以及这些用户在2014年5月至10月期间的阅读行为数据。此外,为了研究不同用户对不同属性新闻的关注,我们还根据新闻热点的分布,选取了此段时间24个新闻热点主题,分析不同用户关注特征。

通过分析,本文得出以下结论:首先,从社会维度来看,女性用户更关注娱乐、养生和健康新闻,而男性偏好政治、体育和社会问题新闻;“低档手机用户”更关

注极端社会事件,“中档手机用户”对民生与经济新闻关注较多,而“高档手机用户”偏好娱乐新闻。其次,从时间维度看,一天有三个关注的峰值,分别是早上8时,中午12时和晚上22时,关注峰值在休息日和工作日有细微差别,休息日的早上阅读峰值占比大于工作日早上,而晚上却相反。我们也发现,人们在不同时间段内倾向于关注不同主题的新闻,有白天话题,中午话题与夜晚话题的区分,在夜晚,用户对新闻关注主题也更加广泛。最后,从空间维度来看,新闻关注在空间上呈现较大差异,省会城市和特大城市新闻阅读消费量。北上广深四特大城市新闻阅读居首位,深圳作为我国的创新之都,更关注高科技的发展;北京是关注政治新闻的重地;上海地区偏经济话题,如汽油价格、公务员涨薪。

本文的研究贡献在于:1. 对网络移动媒体的受众行为进行社会学分析;2. 开创性地利用新闻客户端大数据,从社会、时间和空间的不同维度对移动新闻客户端用户的阅读行为进行了较为全面和深入的分析。相比较传统的依靠问卷方法对回溯信息的收集和分析,这种大数据的分析方式具有更高的精度和真实性,而且能够通过把握社会民众的关注点更深入地把握社会脉动。本文将从四个部分展开分析,第一部分为引言,第二部分介绍研究方法,第三部分从社会、时间和空间三个维度对新闻客户端用户行为进行分析,第四部分为结论。

二、研究方法和数据来源

本文的数据来源于搜狐新闻移动客户端2014年5月至10月的流量数据,我们根据用户手机号进行随机抽样,最终抽取18万用户。我们针对这些用户的阅读行为,进行数据挖掘。由于用户阅读新闻的种类形式多样,不易统计。因此,我们还根据搜狐新闻客户端的新闻热度排行,在每个新闻频道中选定新闻热点主题进行分析,这样做的目的是为了分析用户对不同类新闻的偏好。样本及具体的新闻主题见表1,表2:

① 李力、丁宁:《国内外移动社交类应用用户信息行为研究进展》,《图书情报工作》2015年第10期。

② 邓胜利、杨丽娜:《用户个性特征对信息行为影响的研究进展》,《情报理论与实践》2014年第5期。

③ 唐晓波等:《社会化媒体使用行为影响因素实证分析》,《同济大学学报(自然科学版)》2015年第3期。

④ 聂磊:《新媒体环境下大数据驱动的受众分析与传播策略》,《新闻大学》2014年第2期。

⑤ 陆瑶:《移动新闻高频化对受众的影响研究——基于访谈的探索分析》,《东南传播》2017年第8期。

⑥ McQuail, D. 1997, *Audience Analysis*, London: Sage Publication, 转引自刘燕南:《麦奎尔学术背景探源——评丹尼斯·麦奎尔的〈受众分析〉》,《国际新闻界》2013年第1期。

表1 样本月份分布

月份	新闻阅读消费	占比(%)
5月	8218273	16.13
6月	8217361	16.13
7月	7322143	14.37
8月	7354827	14.44
9月	10710253	21.02
10月	9120045	17.90

表2 新闻频道及新闻主题分布

新闻频道	频数	占比(%)	新闻主题	频数	占比(%)
汽车	1655720	3.25	牌照;汽油价格	5911	2.67
时政	4979421	9.77	反腐;公务员涨薪	55067	24.91
教育	1496534	2.94	教师性侵;高考制度改革	6220	2.81
娱乐	11926630	23.41	来自星星的你;明星吸毒	29612	13.4
财经	1484209	2.91	互联网理财产品;国企高管降薪	14242	6.44
游戏	859928	1.69	游戏展会	1593	0.72
科技	1372328	2.69	大数据;诺贝尔科学奖	2871	1.30
军事	3512999	6.9	钓鱼岛;ISIS	24704	11.18
社会	7747531	15.21	扫黄;强拆	27693	12.53
体育	2899151	5.69	李娜;假球黑哨	5409	2.45
房产	601112	1.18	房价;房产税	18292	8.28
文化	1934752	3.8	文艺座谈会	4545	2.06
生活	10472587	20.56	雾霾;养生保健	24886	11.26

三、新闻客户端用户行为分析

(一)新闻阅读的社会维度:社会阶层与性别差异分析

1. 低档用户关注社会事件,中档用户关注经济话题,高档用户关注娱乐信息

在数据后台中寻找社会阶层的变量并不容易。但是,我们发现,当用户利用手机阅读新闻时,该用户的手机型号会被记录,这意味着我们可以根据用户手机价格判断用户的经济条件,我们将样本中的手机类型与2014年11月份手机市场价格进行匹配。用手机价格去识别用户的社会经济地位。最终,我们根据手机价格分布,将手机用户分成三档,即(1)“高档手机用户”:手机价格在3800元以上,(2)“中档手机用户”:手机价格在1000-3800元之间;(3)“低档手机用户”:手机价格在1000元以下。

首先,我们对不同手机档次用户的新闻阅读消费进行统计。我们发现“低档手机用户”和“中档手机用户”新闻阅读占比均为38%左右,高档手机用户点击占

比则较小,为24%,这和客户端内不同档次手机用户的比例相关。

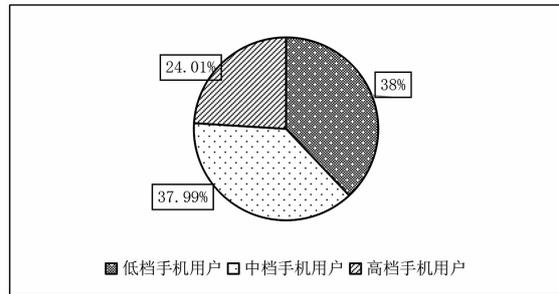


图1 不同档次手机新闻阅读消费占比

其次,我们将高,中,低档手机用户对不同话题的点击率分别用其在全部话题中的点击率标准化后,绘制出如图2至图4折线图。

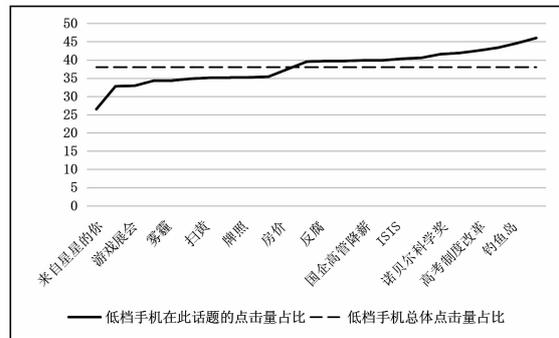


图2 低档手机用户关注话题

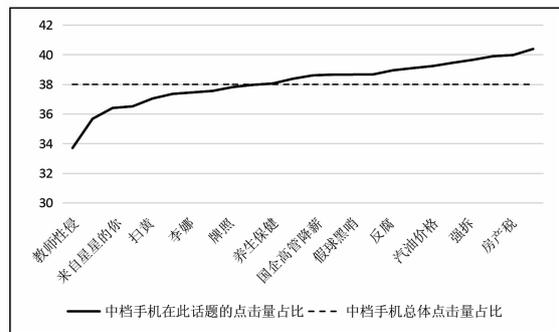


图3 中档手机用户关注话题

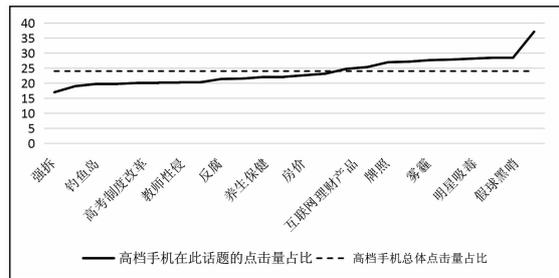


图4 高档手机用户关注话题

如图 2 至图 4 所示,虚线线条代表该档次手机用户的总体点击占比,而实线代表特定新闻主题的点击占比。当实线比例线条超出虚线比例线条时,超过部分的新闻主题为该档次手机用户较为关注的主题。

由图 2 可知,低档手机用户相对关注民生话题(关注度前四话题依次为教师性侵、钓鱼岛、强拆、高考制度改革)。关注度位列前四的四个话题与其他话题相比有着较大的差别,这也说明低档手机用户对这些问题较为关注。一般来说,低档手机用户大部分收入水平较低,其通常关心与自身生活质量、安全等切身利益的社会事件。由于社会负面事件的受害者大部分是弱势群体(教师性侵、强拆等),此类负面话题与其所处的社会位置息息相关。我们发现 2014 年公开报道的教师性侵事件多发生在普通学校,并且以小学和初中生为主(表 3)。

表 3 2014 年教师性侵事件

时间	地点	涉及人物	新闻来源
4 月 15 日	广东省韶关市曲江区	小学体育老师	搜狐新闻
4 月 15 日	广东省茂名市电白县	中学老师	新华网
4 月 21 日	贵州毕节	小学老师	四川新闻网
5 月 2 日	宁夏灵武市	乡村教师	中国网
5 月 18 日	安徽潜山县天明小学	小学校长	新浪网
5 月 27 日	眉县	小学老师	四川新闻网
6 月 16 日	广东惠州市龙门县龙华镇	小学老师	四川新闻网
7 月 19 日	云南省龙场镇	乡村小学教师	梅州日报
9 月 5 日	广东省茂名信宜市大成镇	小学老师	新华网
9 月 7 日	河南平顶山新华区焦店镇	幼儿园老师	四川新闻网
9 月 9 日	吉林省公主岭市育文中学	中学老师	四川新闻网
10 月 13 日	四川美术学院	大学教授	温州新闻
10 月 15 日	厦门大学吴春明	大学教授	南海网

同理,强拆事件发生的地方通常为农村或城乡结合部,这些地方低档手机用户比例远高于中高档手机用户的比例。而高考是弱势群体向上社会流动的最重要途径,社会弱势群体对此非常关注。

由图 3 可知,“中档手机用户”对民生话题较为关注(互联网理财、房产税、房价、油价等)。比较而言,“中档手机用户”关注的民生话题更多集中于经济层面而非社会问题,可能是因为他们相对收入较高、文化水

平较高,具有一定的财富管理理念,因此会更加关注相关话题。“中档手机用户”更加关注互联网理财、房产税与房价。由于“中档手机用户”所对应的用户具有一定财富意识,他们愿意尝试通过互联网进行理财,因此也更关注相关信息,从中学习或找寻理财机会。而相比起西方中产阶级,“中档手机用户”所代表的中国中层阶级也遭受着更大的压力,如房价、油价等,这迫使他们更多地关注现实问题而非享受、兴趣类话题。

由图 4 可知,对于“高档手机用户”而言,“来自星星的你”这一话题超过了平均水平。相比其他用户,“高档手机用户”的财富水平较高、生活环境较好,各种社会问题、投资理财等话题对其生活并没有太大影响,因此这些话题的点击情况较为平均。由于其生活质量较高,其关注点为科技、文化、娱乐等与其个人兴趣更加相关的话题。恰逢统计时韩剧《来自星星的你》在国内掀起一股热潮,在“高档手机用户”中,对《来自星星的你》这一话题的关注远高于其他话题。

2. 男性用户阅读总量大,女性用户平均点击次数多

尽管女性总体上不如男性关注新闻和公共议题,但是我们发现,女性一旦成为新闻读者,她们点击新闻的平均频次比男性读者更高。在“搜狐新闻”的全部点击量中,女性点击量约占 31%,而男性的点击量占到 69%(表 4)。但是,如果根据平均点击频次,以女性读者的平均点击指数^①100 为参照,我们算出男性的平均点击指数约为 76,那么我们可以说,女性的平均点击频次高出男性读者三分之一左右。这说明已经有相当比例的中国大陆女性在关注公共议题。

表 4 分性别点击量与平均点击频数

性别	总体点击量比例(%)	平均点击指数
女	30.81	100
男	69.19	75.74

当我们以用户的平均点击频次为因变量,性别为自变量,控制用户所使用手机的档次^②、用户所处的省份、看新闻的日期和时刻等变量进行多元回归时,女性的平均点击频次比男性的平均点击频次仍然高约三分之一。同时,使用的手机越贵,平均点击频次也越高。

① 平均点击指数根据平均点击次数计算得出,以女性的平均点击频次为参照,相应得出男性的平均点击指数为 75.74。

② 根据手机价格把手机分为三档,低档手机价格为 1000 元以下,中档手机价格为 1000 元至 3800 元,高档手机价格为 3800 元以上。

表5 回归表格

	因变量为平均点击次数		
	1	2	3
男性	-10.751*** (0.422)	-9.402*** (0.422)	-9.399*** (0.422)
中档手机	No	4.906*** (0.407)	4.907*** (0.407)
高档手机	No	19.548*** (0.507)	19.541*** (0.507)
东部	No	1.286** (0.480)	1.284** (0.480)
西部	No	1.282* (0.539)	1.281* (0.539)
月中	No	No	0.184 (0.460)
月末	No	No	-0.691 (0.451)
上班时间	No	No	-0.194 (0.376)
常数项	44.319*** (0.365)	36.726*** (0.565)	37.039*** (0.669)
观测数	136250	136250	136250
R-square	0.005	0.016	0.016

注:(1)模型为线性回归模型;(2)括号中为标准差;(3)显著性水平:*为5%,**为1%,***为0.1%;(4)参照组分别为女性、低档手机、中部地区、月初、下班时间。

根据同样的方法,通过对平均点击频次标准化为平均点击指数,以“高档女性用户”平均点击频次为参照组,假设其平均点击指数为100,那么高档男性用户的平均点击指数为83,低档女性用户的平均点击指数为65,低档男性用户的平均点击指数为49。因此,从关注新闻角度看,高档女性用户>高档男性用户>低档女性用户>低档男性用户,具体见表6。

表6 平均点击频数差异

类别	平均点击指数
高档手机女性用户	100
高档手机男性用户	83.39
低档手机女性用户	65.46
低档手机男性用户	48.85

本文对“高档手机女性用户”的定义为使用高档手机的女性,该人群的特点是生活富足,教育程度较高,有闲暇时间;与此形成对应的是使用高档手机的男性群体——“高档手机男性用户”,同样作为社会精英阶层,其在关注新闻方面比“高档手机女性用户”要低。对“低档手机女性用户”而言,在新闻关注度方面,排在精英男性之后,但是却高于另一个使用低档手机的男性群体——“低档手机男性用户”。

3. 女性关注主题倾向于娱乐生活,男性关注偏政

治、体育和社会

图5为女性在各类话题的阅读占比,31%水平线是女性在总点击量所占的比例。与水平线相交的右上方的曲线表明女性用户在明星吸毒、养生保健、来自星星的你和雾霾等新闻主题的点击量都超过了平均占比。相对男性而言,女性对娱乐、生活类新闻更为关注。女性对涉及明星吸毒的新闻尤为关注,在该话题的点击量占比位居24个热门话题之首。女性在总的点击量中所占的比例为31%,但在这一话题上的点击量占比则达到了42%以上。

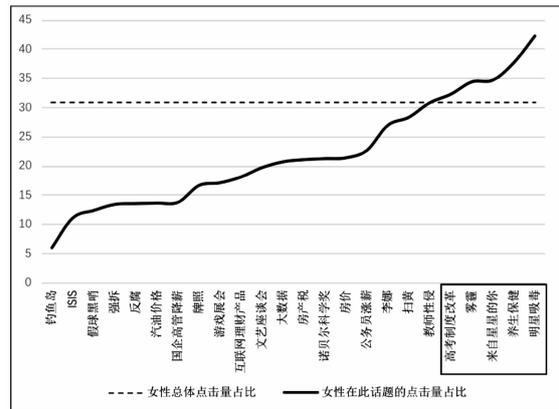


图5 新闻阅读的性别差异

表7 男女用户新闻阅读比例

读者比例	男性(%)	女性(%)
总体	69.19	30.81
明星吸毒	57.8	42.2
养生保健	62.35	37.65
来自星星的你	65.34	34.66
雾霾	65.65	34.35
钓鱼岛	94.05	5.95
ISIS	88.94	11.06
假球黑哨	87.67	12.33
强拆	86.65	13.35
反腐	86.51	13.49

而从男性的角度来看,男性阅读量占总体比例69.19%,男性在钓鱼岛(94.05%),ISIS(88.94%),强拆(86.65%)和反腐(86.51%)等政治类新闻以及假球黑哨等体育类新闻阅读比例占据最高。这说明,男性用户倾向于关注政治、体育类话题。

(二)新闻阅读的时间维度分析

移动新闻阅读呈现出时间上的差异。从图6(左)中可以明显看出,一天之中,新闻阅读有三个峰值。第一个为早上8时,此时人们刚起床或者正在上班途中;第二个峰值在中午的12时,此时人们处于午休阶段;第三个峰值为晚上22时。在这三个峰值中,阅读量最

大的为 22 时,其次为 12 时,再次为早上 8 时。移动新闻阅读的时间规律和人们的作息时相吻合,这说明,人们利用移动端阅读新闻主要是在休息期间。

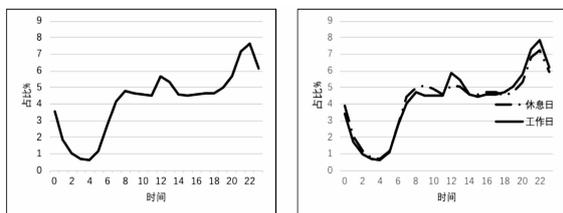


图 6 新闻阅读时间的分布差异

注:(1)图中纵轴为该时段的新闻阅读比全天的阅读量;(2)左图为新闻阅读的整体时间变化图;右图为休息日和工作日新闻阅读的时间变化图

而就休息日与工作日对比而言,从单日阅读量的时间趋势来看,二者的阅读趋势一致。但是休息日与工作日仍存在差异,从总体阅读量来看(表格未报告),工作日的阅读量大于节假日的阅读量;其次,节假日的阅读占比的第一个峰值点占比比工作日高,这意味着和工作日的早上相比,人们更喜欢在休息日的早上阅读新闻。最后,工作日中午和晚上的阅读峰值占比大于此时间段节假日的阅读峰值占比。

(三)新闻消费的空间维度:

1. 新闻消费空间差异大,省会城市新闻消费多,北上广深为新闻阅读消费中心。

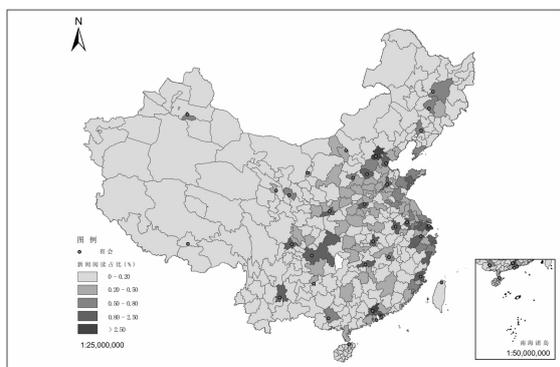


图 7 新闻阅读的城市分布

如图 7 所示,新闻阅读在地域分布有较大差异。首先,省会城市作为区域的政治,经济中心,亦是该省的新闻信息传播中心。其次,从区域上看,北上广深占据了新闻阅读的前四位。北京作为首都,新闻阅读总量上占据首位,新闻阅读消费占全国的 5.03%。其次为上海,消费占全国的 3.8%,广州和深圳分别占全国的 2.9%和 2.66%。北上广深成为新闻阅读消费的中

心,原因可能和各个城市的人口、经济发展和教育水平相关。

表 8 新闻阅读的城市差异

城市	新闻阅读	百分比(%)
北京市	2547097	5.03
上海市	1921001	3.80
广州市	1469618	2.90
深圳市	1347421	2.66
武汉市	1166068	2.30
成都市	1165305	2.30
南京市	1001983	1.98
西安市	908082	1.79
苏州市	892880	1.76
重庆市	880060	1.74

2. 新闻阅读消费的特大城市分析

(1)北京用户规模大,深圳关注度最高

北上广深的用户数量(UV^①)比例为 8 : 6 : 5 : 5,北京用户数量显著高于其他城市。在全部新闻点击量中,北京用户的新闻点击量占四个城市总点击量 35%,亦遥遥领先于其他三个城市。

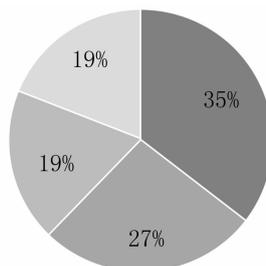


图 8 北上广深新闻阅读消费

考虑到北上广深人口规模各异,北京上海超过 2000 万,而深圳广州仅有 1000 多万人口,为剔除人口因素影响,本文用新闻阅读数(PV^②)除以各市人口数量来衡量各个城市“平均关注度”。为了更好地刻画用户行为,进一步以用户数(UV)除以该市人口作为“市场渗透度”,点击量除以用户数作为“使用深度”,从新的角度来衡量各市对新闻的关注程度。最终,以“平均关注度”、“市场渗透度”和“使用深度”三项指标加总并进行排序,得出各地“综合新闻关注度排名”。

表 9 各市综合新闻关注度排名

	北京	上海	广州	深圳
女	1	3	3	2
男	1	4	3	2
总	1	4	3	2

① UV 是指不同的、通过互联网访问、浏览这个网页的自然人。

② PV(page view)即页面浏览量,通常是衡量一个网络新闻频道或网站甚至一条网络新闻的主要指标。

深圳虽然用户规模和点击量最少,但由于其人口基数较低,市场渗透度和平均关注度更高,因而综合新闻关注度反而仅排在北京之后。上海虽然用户使用频度较高,但由于人口规模最大,市场渗透度最低,综合关注度排名靠后。从点击率最高的24个新闻话题关注度排名情况看,北京关注度最高,上海则关注度最低。

(2) 男性用户是主力,中高端用户占主导

数据显示,男性用户更愿意通过新闻客户端获取新闻资讯。在全部抽样客户中,北上广深四地男性用户占到全体用户的75%,且四地男性用户比例均超过70%,远远高于女性用户所占比例。

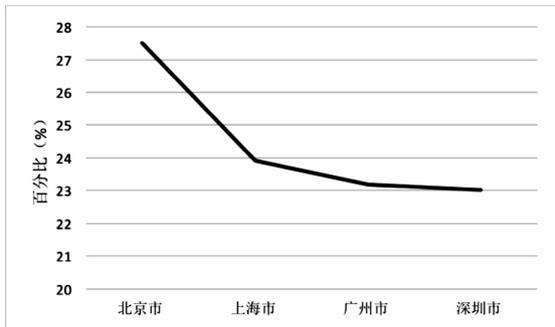


图9 各市女性用户比例

其中,女性用户占比最高的为北京,但占比也仅达到27%。这可能与不同性别的阅读习惯差异相关,尤其是移动新闻客户端的主要用户以70、80、90后为主,这部分群体中的男性用户,普遍对周边事务和时政要闻更加关注。最后,男性用户对新科技接受度较高,更容易采取移动客户端的方式获取新闻资讯。

随着收入增长和智能手机普及,中高端手机已经成为手机市场主导产品。根据中国互联网消费调研中心数据,2014年上半年,1000元以下手机产品,消费者关注度仅为22%,消费者消费偏好主要倾向与1000元

以上中高端产品。从图中可以看出,北京和上海的高档用户和中档用户都远多于广州深圳,

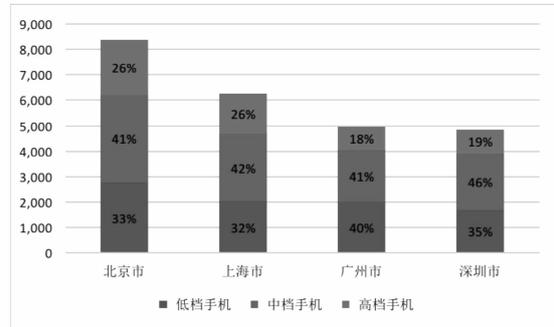


图10 各市手机档次分布比例

定义“话题渗透度”为某一话题该条件下阅读人数与该条件下总人数之比,以表10中为例,北京阅读教师性侵新闻的低档用户占到北京全体低档用户的5.58%。从表11中可以看出,在选择搜狐手机客户端阅读新闻的用户在各大话题中,高档手机用户渗透度最高,其次为中档手机用户,排名最后的是低档手机用户。也就是说通常情况下,高档手机用户阅读某一新闻的用户比例总是高于中档,中档又高于低档。但在各大话题中,“教师性侵”和“钓鱼岛”并不完全符合这一规律。钓鱼岛事件有更多的低档用户关注,可能与社会底层对社会现状不满情绪而采取网络发泄的渠道有关。

表10 特殊议题渗透度

	教师性侵			钓鱼岛		
	低	中	高	低	中	高
北	5.58%	5.49%	6.71%	21.36%	23.02%	22.71%
上	6.39%	4.59%	6.01%	25.58%	23.80%	27.01%
广	5.42%	5.25%	5.69%	22.50%	22.06%	23.63%
深	5.44%	5.13%	7.36%	22.34%	21.65%	26.79%

表11 各市各档手机用户渗透度排名

		李娜	假球黑哨	牌照	汽油价格	房价	房产税	来自星星的你	文艺座谈会	养生保健	雾霾	公务员涨薪	反腐	高考制度改革	教师性侵	明星吸毒	国企高管降薪	互联网理财产品	游戏展会	大数据	诺贝尔科学奖	钓鱼岛	I S I S	扫黄
		低档	北	3	3	3	3	3	3	3	2	3	3	3	3	3	2	3	2	3	3	3	3	3
	上	3	3	3	3	3	3	3	2	3	3	3	3	3	1	3	2	3	3	3	3	2	3	3
	广	3	3	3	3	3	3	3	3	3	3	3	3	3	2	3	3	3	3	3	3	2	3	3
	深	3	3	3	3	3	3	3	3	3	3	3	3	3	2	3	3	3	3	3	3	2	2	2

中档	北	2	2	2	2	2	2	2	3	2	2	1	2	2	3	2	1	2	2	2	1	2	2	
	上	2	2	2	2	2	2	2	3	2	2	2	2	2	3	2	3	2	2	2	2	3	2	2
	广	2	2	2	1	2	2	2	2	2	2	2	2	1	3	2	2	2	2	2	3	2	2	
	深	2	2	2	2	2	2	2	2	2	2	2	2	2	3	2	2	2	2	2	3	3	1	
高档	北	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	3	1	1	1	1	2	1	1
	上	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1
	广	1	1	1	2	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1
	深	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3

(3)新闻点击三高峰,生活作息各不同

我们对北上广深用户的浏览时点与点击量进行分别统计,并以该时点内新闻点击量除以全天新闻点击总数,得出新闻阅读时间分布折线。

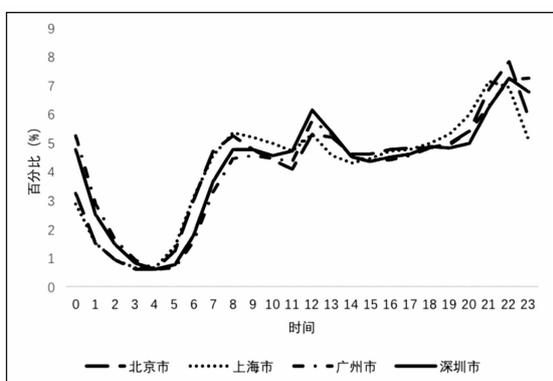


图 11 点击时间分布

从小时分布曲线可以看出,四个城市新闻点击出现三个高峰期,这与全国新闻阅读的时间规律一致,分别为早上 7—8 点,中午 12 点和晚 22 点。从总体上看,用户在夜间使用新闻客户端的活跃度高于白天,休息时间活跃度高于工作时间,尤其是借助移动客户端优势,利用早上通勤和午间休息的碎片化时间获取新闻资讯。

有趣的是,通过大数据我们可以看出广深两地用户熬夜更多,北上两地用户作息更加健康。从小时分布曲线看,深夜期间广深用户点击比例持续高于北上两地,在 22 点以后北上两地用户点击比例快速下降,而同期广深两地用户点击比例下降相对缓慢,一直到凌晨 12 点,用户点击比例水平仍与早高峰期间基本持平,显示广深用户熬夜更多,而北上用户入睡更早。早上情况则相反,北上用户自 4 点起点击比例开始回升,5 点后回升速度加快,而广深用户则到 5 点后才开始回升,且用户点击比例持续低于北上两地,显示北上用户更爱早起,这或许是由于 5—10 月处于夏秋季节,北上两地纬度高,日出较早导致人们相对早起。

(4)京谈政治沪务实,深挖数据穗看球

从图 12 中可以看出,各个城市各话题的关注度变化相似,反腐都是最受欢迎的话题。下面本文将分析各个话题中不同城市渗透率的差异,从而划分不同城市的特点。表 12 列出了各个城市渗透度排名第一的话题。以“李娜”为例,在这个话题中,北京的渗透度——也就是用户点击比例最高,因而李娜归于北京这一列。

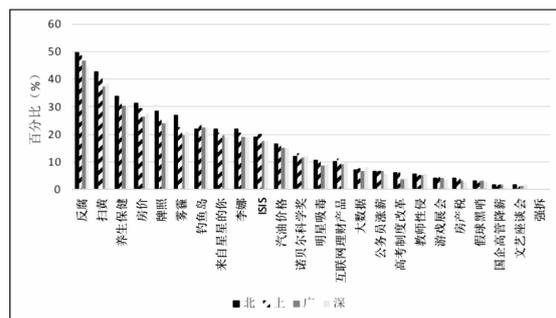


图 12 北上广深新闻主题渗透率

表 12 各市渗透率排第一的话题

北京	上海	广州	深圳
李娜	汽油价格	假球黑哨	大数据
牌照	公务员涨薪		
房价	高考制度改革		
房产税	互联网理财产品		
来自星星的你	游戏展会		
文艺座谈会	诺贝尔科学奖		
养生保健			
雾霾			
反腐			
教师性侵			
明星吸毒			
国企高管降薪			
扫黄			
钓鱼岛			

相对于其他三座城市的性质,北京对于雾霾格外关注。2014 年北京雾霾天气严重,PM2.5 指数比天气预报更受关注,外出活动不仅受到雨雪的影响,还必须

考虑雾霾严重与否。北京马拉松比赛在雾霾中开跑，APEC期间政府用超常规手段治理出的“APEC蓝”成为热词等都体现出“雾霾”二字在北京市民生活中的重要程度。而上海污染程度较轻，广州、深圳重污染天数为0。同样在北京备受关注的还有反腐、扫黄和明星吸毒，这与北京政治中心的地位和人气正火的柯震东吸毒被拘留，微博大V薛蛮子、国民女婿黄海波嫖娼被捕等事件都发生在北京有关；有趣的是，在其他城市对钓鱼岛问题热情高涨的时候，北京却成了最沉默的一家，这也许与北京作为政治中心政治成熟度较高，爱“新闻”胜于“争论”。最后，李娜今年在北京宣布退役，可能是北京最关注“李娜”话题的原因之一。

表 13 四大一线城市空气质量对比^①(单位:天)

	北京	上海	广州	深圳
一级良	41	94	81	122
二级良	13	174	179	199
三级轻度污染	84	56	90	37
四级中度污染	47	21	15	5
五级重度污染	45	17	0	0
六级严重污染	13	1	0	0
重污染合计	58	18	0	0

上海是一座走在前面的城市，上海对高考制度改革的话题关注度在四城市中最高。2014年9月，国务院印发《关于深化考试招生制度改革的实施意见》，勾勒出新一轮高考改革方案的轮廓。随后，上海、浙江两地率先试水高考综合改革：高考不分文理科，英语“一年两次考试，取最好成绩计入高考”。试点改革关乎上海多数家庭的切身利益，对话题的关注度高于其他三个城市。China joy连续10年在上海举办，且上海的动漫产业在全国具有领先优势，平均每部动画影片的票房成绩均能占到全国票房总量的近五分之一，^②此外，上海对于游戏展会的关注度也远高于其他三个城市。

在广州，假球黑哨成为热门话题，可能与广州恒大队在11月2日傍晚第四次捧起冠军奖杯，中央巡视组批中国足球假球黑哨横行，球员靠送礼进国家队等事件有关。巨大的卫冕让广州人更关注足球新闻。

深圳拥有中国最多的A股上市民营科技企业，中

国最大的互联网企业腾讯、最大的通讯制造商华为与中兴、最大的新能源汽车比亚迪等集聚于此，腾讯、华傲数据等公司均被认为已在大数据领域占位。

从上面各个城市的分析可以看出，人们关注新闻，很大程度上是在看自己的“身边事”。记录着城市变化的新闻把这座城的目光聚焦，在零散的新闻库里划分出清晰的城市印象：政治的北京，务实的上海，足球的广州和科技的深圳。

四、研究结论

由于数据获得的局限性，以往针对移动端新闻用户阅读行为的研究相对缺乏，而且准确度较低。本研究通过大数据分析的方式，克服了数据获得的局限性，提供了行为分析的准确性。我们采取手机号码随机抽样的方法，随机抽取搜狐新闻18万用户，挖掘这些用户5月至10月的阅读行为数据，并在各个新闻大类中，选出24个新闻主题，从社会、时间和空间的角度对其进行分析。我们发现了如下结论：

第一，从社会维度看，使用不同档次手机的用户所关注的新闻主题有较大差别，低档手机用户更关注极端社会事件，中档手机用户偏重民生与经济，高档手机用户则笃爱娱乐新闻。此外，新闻关注上在性别亦存在差异，女性用户更关注与娱乐、养生、健康等有关的新闻，而男性对政治、体育和社会问题新闻关注更多；

第二，从时间维度看，新闻关注时间差异较大，一天之中有三个新闻关注的峰值点，分别是早上8时，中午12时和晚上22时。

第三，从空间维度来看，首先，新闻关注在城市间存在较大差异，省会城市一般也是区域的新闻消费中心；其次，通过对“北上广深”四大特大城市分析，发现不同城市有自己的特点，如深圳作为经济的后起之秀，从新闻关注的时点看，在深圳工作的人更喜欢熬夜；而对于北京来说，是关注政治新闻的重地；上海地区偏经济话题，如汽油价格、公务员涨薪。

〔作者郑路，清华大学社会学系副教授；吴文兵，香港科技大学社会科学部研究生；葛天任，同济大学政治与国际关系学院助理教授，上海 200092；张翀，中国农业银行总行职员。〕

责任编辑：陈 拓

^① 资料来源：PM2.5监测网。

^② 《2005—2019年上海动漫产业投资分析及前景预测报告》。