

商务分析整合实践第五组期末报告

2018 年 12 月 20 日

指导老师：郭凡、廖闯

快手用户行为预测——取消关注

宋辰菲

清华大学经济管理学院

2018211650

songchf.18@sem.tsinghua.edu.cn

肖念瑶

清华大学经济管理学院

2018211660

xny18@mails.tsinghua.edu.cn

闫嘉文

清华大学经济管理学院

2018211671

yanjw.18@sem.tsinghua.edu.cn



目录

执行摘要	3
一、研究背景	4
1. 短视频定义	4
2. 短视频行业概览	4
二、研究目标与选题意义	5
三、数据介绍	6
1. 样本汇总数据	6
2. 社交网络数据	7
3. 视频消费数据	7
4. 视频生产数据	7
四、研究思路设计	8
1. 研究变量选择	8
2. 数据整合	10
五、取关机制分析	10
1. APP 运行机制.....	10
2. 关注强度.....	11
3. 作品质量.....	13
4. 用户互动.....	15
5. 案例分析.....	15
六、预测模型.....	18
1. 基准模型.....	18
2. 机器学习模型准确率比较.....	19
3. 集成模型.....	21
七、Spark 平台全量数据分析	22
八、总结与工作展望	23

执行摘要

2018 年秋季商务分析整合实践课中，小组成员通过在快手公司 6 周的综合实践，尝试将课程知识运用于实际业界生产环境，并在业界导师指导下，进行用户取消关注行为的探索性研究和分析、培养业界处理问题思维、获得大数据处理实际经验。

快手 App 作为短视频平台，对接了愿意展现美好生活的“视频生产者”和希望看到他人生活状态的“视频消费者”。在快速发展的同时，也在思考如何站在用户的角度理解其行为，为用户提供更好的消费体验，建设良好的短视频社区。用户在社交平台中最重要、最优先的行为之一就是“关注”和“取消关注”。在 2018 年 9 月 1 日当天，有近 80 万用户在快手上发生了超过 220 万次关注行为，并有超过 10% 的用户在 30 天内取消关注。

对于取消关注的识别和预测，可以帮助平台了解用户行为；识别和划分合理取关和不合理取关，允许并鼓励合理的取关行为，减少并避免不合理的取关行为发生，可以引导用户建立长期稳定的社会网络连接。小组成员将通过分析用户行为数据了解取消关注的机制，并建模预测取消关注事件的发生。

项目开展前期，我们主要完成了数据理解与机制探索的工作。我们尝试探究用户基本特征、用户行为、用户社交网络这三张表背后的数据逻辑和内在联系，进行基本变量的统计分析，分别从用户画像特征和行为特征出发，探索用户取消关注的原因，对取关行为细分定义，并且分析了两种典型案例。

随着项目深入，我们建立模型预测用户的取关行为。通过从数据中抽取均衡样本，Logistic 基准回归预测准确率达到 70.3%。我们还对比使用了 SVM、决策树、神经网络等 10 种不同机器学习模型，对取关行为进行预测，并对比分析了不同模型之间的特征重要性，还使用集成模型来提高模型稳健性和准确率。最后，我们还尝试了在大数据 Spark 平台上用全量数据进行建模，并基于研究结果做出总结汇报，对快手 APP 提出部分建议。

期末报告后续分为以下几个章节展开：（1）研究背景；（2）研究目标与选题意义；（3）数据介绍；（4）研究思路设计；（5）取关机制分析；（6）预测模型；（7）Spark 平台全量数据分析；（8）总结与工作展望

一、研究背景

1. 短视频定义

短视频起源于美国。2011 年 4 月 Viddy 正式发布了移动短视频社交应用，随后 Twitter 推出 Vine，Instagram 开始拓展短视频业务的传统图片分享应用。除此之外，还有 Keek，Mixbit 等许多类似应用出现。在中国，短视频的发展是从 2005 年《一个馒头引发的血案》这一 20 分钟的微电影开始的，微电影的出现促使草根网民加入创作热潮，并进一步推动视频的碎片化和平民化，从 20 分钟逐渐过渡到火爆的 8 秒钟微视。现在，一般将短视频定义为一种视频长度以秒计数、主要依托于移动智能终端实现快速拍摄和美化编辑、可在社交媒体平台上实时分享和无缝对接的一种集文字、语音、视频等多种元素为一体的新型视频形式¹。目前市场上流行的短视频 APP，主要有以下三大特点：

1. 视频长度短，传播速度快；
2. 生产流程简单化，制作门槛低；
3. 具备广泛参与性，社交媒体属性强；

在泛娱乐化和碎片娱乐化的当下，互联网用户对于娱乐的要求愈来愈高，草根用户的心声也逐渐被发掘，相应的互联网产品应运而生。直播 APP 比如花椒、斗鱼，短视频 APP 比如抖音、火山等，针对不同层次的用户设计了令人眼花缭乱的产品。虽然 2016 年直播类 APP 风头强劲，但是 2017 年后，手机短视频类 APP 占了上风。

2. 短视频行业概览

市场数据：

2016-2017 年，是短视频爆发的一年，中国短视频 APP 用户在一年内翻倍，2017 年 12 月已经达到 4 亿 1 千万，年同比增长 116.5%（见图 1）。

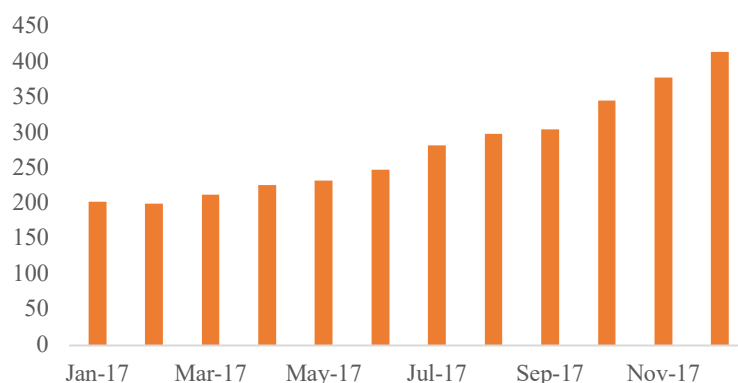


图 1. 2017 年短视频 APP 用户数量(millions)

Source: QuestMobile TRUTH, Dec. 2017

¹ 来自中国产业研究院报告

与成倍增长的用户相匹配的，是更加疯狂增长的短视频 APP 使用时长。从 2016 年的 8.8 千万分钟增长到 2017 年的 3.7 亿分钟，几乎是增长了三倍之多²。与之对应的，是碎片化娱乐在用户间的盛行。根据 QuestMobile TRUTH 的数据显示，人们平均花在每一则视频上的浏览时间减少，从 2016 年 1 月份平均每则视频浏览 4.3 分钟，下降到 2017 年 1 月份的平均浏览时间 3.9 分钟。

产业生态：

短视频的生态圈较为复杂，从文化类监管部门（诸如广电、文化部等）、视频内容提供者、平台渠道方以及相关的硬件支持方，不仅涵盖了专业领域的技术人才和网络名人，更包含了大量草根网民，形成了特有的 UGC（user-generated content）文化。其产业生态如下图所示：

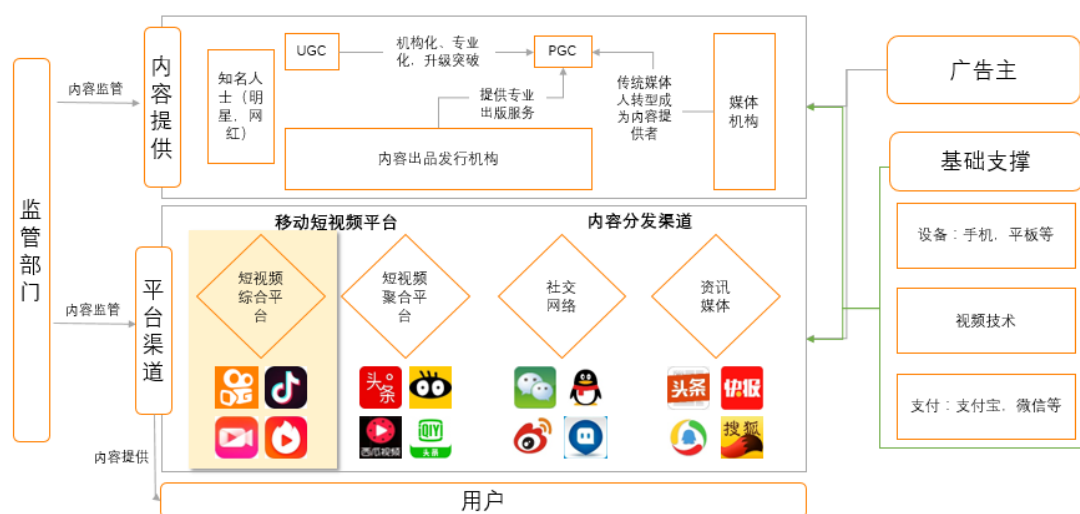


图 2 短视频产业生态圈

本文所研究的对象——快手 APP，属于此生态圈中的移动端视频综合平台，快手旨在用自己的平台，为各类用户提供一个分享、制作上传短视频，并进行社交的机会。

二、研究目标与选题意义

该课题的主要研究目标为：通过分析用户行为数据了解取消关注的机制，理解用户取消关注背后的逻辑，并建模预测取消关注事件的发生，从而优化产品设计，并根据不同的取关原因采取相应措施。

取消关注行为并不是非黑即白的。事实上，一部分取消关注是因为软件限制、自然遗忘等因素导致的不合理取关行为；另一部分则是由于质量低劣、兴趣转移等因素导致的较为合理的取关行为。合理的取关行为是一种自然的社交网络迭代过程，有助于引导短视频社区的良好生态建设；而不合理的取关行为不仅影响了用户体验，也

² 来自 QuestMobile TRUTH, Dec. 2017

破坏了社交网络的正常迭代。因此，我们有必要细分研究取关行为的机制，思考不同渠道下取关行为的合理性，鼓励并引导合理取关行为；减少或避免不合理取关行为的发生，从而更好地引导社区生态建设。我们也将从 APP 机制、关注强度、作品质量、互动行为等四个方面，分别思考取关行为的合理性。

对于快手公司而言，该研究项目在以下三个方面均有积极意义。首先，有助于优化快手 APP 本身。通过取消关注这一行为背后机制的研究，有助于进一步加深对用户偏好的理解，优化快手的内容设计、内容推荐系统。例如，用户可能会取关一些喜欢以极高频率发视频的用户，对其“刷屏”行为无法容忍，那么在掌握这一规律后，快手的推荐页可以适当设置引导，避免“刷屏”行为的出现。其次，有助于帮助快手加强用户社区的建设。在用户的关注行为中，有一部分是出于互粉、骗粉的目的，通过识别由于该行为产生的取消关注，有助于建设高质量的用户社区。第三，有助于为内容生产者提供有效引导。快手平台上不乏网红、大 V 的存在，扶植这些知名用户可以吸引更多流量，提高曝光率。通过研究取关行为，也可以让快手引导这些内容生产者按照用户偏好生产投其所好的内容。

面对庞大的用户群体，竞争激烈的市场，快手想要脱颖而出必须从细节着手，提高用户社区的质量。通过对于取消关注的识别、预测，帮助平台了解用户行为，建立稳定、高质量的用户社区，为用户带来更好的体验，提高用户粘性的同时实现用户增长，从而构建优质的快手用户生态圈。

三、数据介绍

本案例中涉及汇总表、视频生产信息表、视频消费信息表和社交网络四张数据表。我们以下对四张数据表详细描述。

1.样本汇总数据

样本数据记录了在 2018.9.1 发生关注行为的用户在 30 天后（2018.9.30）是否取消关注了被关注用户。其中，是否取消关注是最终的因变量，如果用户取消关注，则记录该用户的取关时间和取关路径。样本数据集中包含了 776,643 个快手用户的 2213939 条关注记录，其中取关记录为 222,623 条，取关率约为 10%。变量说明如下表 1 所示。

表 1：样本数据变量

变量	含义	形式	变量类型
<i>from_uid</i>	发起关注用户 ID	aNoobpTEvBi4Ax1Xyk/gKa==	字符串
<i>to_uid</i>	被关注用户 ID	sEMVrjridxoS3ds6rjCQeQ==	字符串
<i>follow_time</i>	关注时间戳(毫秒)	1535776508000	时间戳
<i>follow_stacj_array</i>	关注栈(关注路径)	[16,42,45]	列表
<i>is_unfollow</i>	是否取消关注	0	0-1 变量
<i>unfollow_time</i>	取关时间	NULL	时间戳
<i>unfollow_page_ref</i>	取关路径	NULL	列表

2.社交网络数据

社交网络数据记录了样本数据中涉及的用户在快手上的所有关注关系，通过该表可以构建样本用户间的社交网络，包括关注者的关注对象、关注时间和关注路径，数据量高达 3.7 亿条。变量说明如下表 2 所示。

表 2：社交网络数据变量

变量	含义	形式	变量类型
<i>From_uid</i>	发起关注用户 ID	aNoobpTEvBi4Ax1Xyk/gKa==	字符串
<i>To_uid</i>	被关注用户 ID	sEMVrjridxoS3ds6rjCQeQ==	字符串
<i>Follow_time</i>	关注时间戳(毫秒)	1535776508000	时间戳
<i>Follow_stack_array</i>	关注栈(关注路径)	[16,42,45]	列表

3.视频消费数据

用户关注后在关注页的视频消费数据包括在 2018.9.1 发生关注的用户在 2018.9.30 之前对被关注用户的视频消费情况。该数据中，每条记录代表被关注者的一次真实展示(real show)，具体的视频消费变量包括关注者是否点击、播放时长、是否点赞、评论次数在内的行为记录。变量说明如下表 3 所示。

表 3：视频消费数据变量

变量	含义	形式	变量类型
<i>from_uid</i>	发起关注用户 ID	fxBJR7Zfu6NXe2PHtJsgHg==	字符串
<i>to_uid</i>	被关注用户 ID	0fQwAEoxBEZLkEnEdJczhA==	字符串
<i>sever_time</i>	视频展示时间	1533746528000	时间戳
<i>photo_id</i>	视频 ID	8164265919	整型
<i>is_click</i>	是否播放	0	0-1 变量
<i>playing_times</i>	播放时长(毫秒)	NULL	整型
<i>is_like</i>	是否点赞	0	0-1 变量
<i>comment_times</i>	评论次数	0	整型

4.视频生产数据

用户关注后的视频生产数据包括了用户的基本特征以及 2018.9.1-2018.9.30 时间窗口内的快手视频生产情况，变量说明如表 4 所示。

表 4：视频生产数据变量

变量	含义	形式	变量类型
<i>user_id</i>	用户 ID	0fQwAEoxBEZLkEnEdJczhA=	字符串
<i>active_data_list_30d</i>	过去 30 天活跃日期列表	["20180908","20180915"]	列表
<i>register_time</i>	注册时间	2011/10/1 10:02	日期
<i>fre_province</i>	过去 30 天常驻省份	北京	字符型
<i>fre_city</i>	过去 30 天常驻城市	北京	字符型
<i>fans_cnt</i>	粉丝数	1798402	整型
<i>follow_cnt</i>	关注数	385	整型
<i>friend_cnt</i>	好友数	189	整型
<i>photo_id</i>	过去 30 天内上传作品 ID	NULL	整型
<i>live_stream_id</i>	过去 30 天内打开直播 ID	3450804815	整型
<i>photo_or_live_create_time</i>	作品或直播创建时间	2018/9/15 18:56	日期
<i>photo_or_live_province_name</i>	作品或直播创建时所在省份	北京	字符型

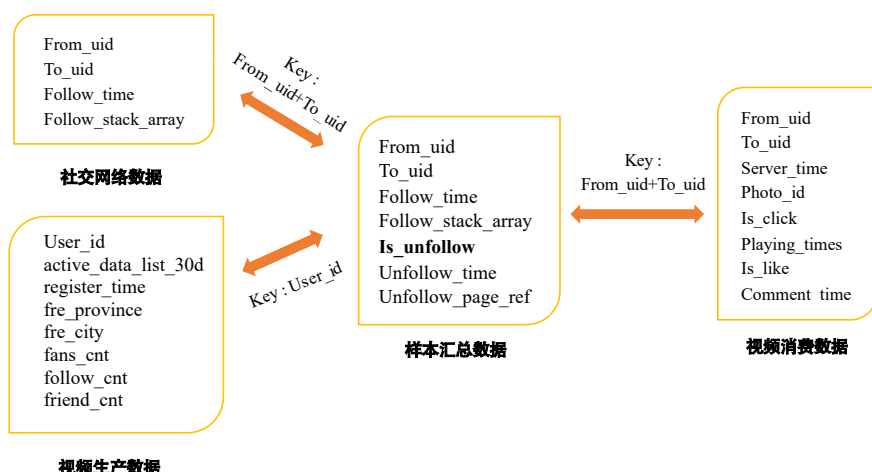


图3 数据表内部逻辑关系示意图

四、研究思路设计

正确将取关行为划分为合理取关行为和不合理取关行为需要通过诸多维度的特征进行考量。为进一步研究用户取关行为的合理性，我们从以下四个方面进行思考，并通过计算构建得到 41 个相关指标：（1）与快手 APP 特征相关的用户属性；（2）视频消费者的关注强度；（3）视频生产者的内容整体质量；（4）互关人群互动行为。详细的用户结构特征变量参考表 5。

1. 研究变量选择

（1）与快手 APP 特征相关的用户属性（14 个）

与快手 APP 特征相关的用户属性也可以被理解成为用户的基本信息属性，这些属性往往与 APP 设置相关。值得注意的是，每个用户的结构属性特征可以被分为固定特征和可变特征。固定特征为该用户在创立账户时的固有属性，这样的变量不会随时间变化而改变，例如用户的注册时间、注册地点等；可变特征则随着用户在平台中的行为发生变化，例如关注数、粉丝数、活跃地点等。我们将试图从关注者 f (*follower*) 属性和被关注者 u (*user*) 属性两方面进行衡量。

（2）视频消费者的关注强度（7 个）

关注强度从网络联系强弱角度出发，有助于我们更加深入了解用户取关行为的逻辑和作用渠道。例如，双向关注（互粉）代表了双方共同的行为特征和趋势，因此理论上比单向关注结构更加稳定，关系也更不容易破裂（取关）。且双方的共同好友数、粉丝数也影响着双方社交关系的距离，从而间接影响了用户的取关行为。

（3）视频生产者的内容整体质量（14 个）

整体内容质量则重点衡量了短视频、直播生产者的内容质量。由于我们并没有具体的内容信息，我们这里通过相关的点赞、评论等信息间接衡量作品质量。具体而言，行为数据可以被分为单一用户在区间内的行为数据和用户之间在区间内的交互数据。单一用户的行为数据包括但不限于一个用户活跃天数、直播次数、收获点击数、评论数等。

(4) 互关人群互动行为（6个）

用户互动行为属性的数据则更加细化地考虑了用户与用户每日在平台上发生的交互行为。行为数据覆盖时间和范围为关注者 f 和被关注者 u 从 2018 年 9 月 1 日至 9 月 30 日期间所有的互动行为数据。关注者和被关注者之间的交互数据包括但不限于展示次数、点击次数、播放次数等。

表 5：用户特征变量与行为变量

	关注者 f / 被关注者 u	衡量内容	变量名称
用户特征变量			
用户和 App 相关	f	粉丝数量	$fans_f$
	f	关注数量	$followers_f$
	f	相互关注数量	$friends_f$
	f	粉丝关注比	$fans_f/followers_f$
	f	好友粉丝比	$friends_f/fans_f$
	f	注册时间	$register_f$
	f	活跃地点	$location_f$
	u	粉丝数量	$fans_u$
	u	关注数量	$followers_u$
	u	相互关注数量	$friends_u$
	u	粉丝关注比	$fans_u/followers_u$
	u	好友粉丝比	$friends_u/fans_u$
	u	注册时间	$register_u$
	u	活跃地点	$location_u$
用户关注强度	$f-u$	关注渠道	$channel$
	$f-u$	关注渠道深度	$depth$
	$f-u$	双向关注	$reciprocal$
	$f-u$	关注次序	n_th
	$f-u$	共同关注	$common_follows$
	$f-u$	共同粉丝	$common_fans$
$f-u$	共同好友	$commen_friends$	
用户行为变量			
内容整体质量	f	总活跃天数	$active_f$
	f	总直播次数	$live_f$
	f	总上传短视频次数	$photo_f$
	f	总收获点击数	$click_f$
	f	总收获播放数	$play_f$
	f	总收获点赞数	$like_f$
	f	总收获评论数量	$comment_f$
	u	总活跃天数	$active_u$
	u	总直播次数	$live_u$
	u	总上传短视频次数	$photo_u$
	u	总收获点击数	$click_u$
	u	总收获播放数	$play_u$
	u	总收获点赞数	$like_u$
	u	总收获评论数量	$comment_u$
用户互动变量	$f-u$	展示次数	$show$
	$f-u$	点击数	$click$
	$f-u$	点击展示比	$click/show$
	$f-u$	播放次数	$play$
	$f-u$	点赞数量	$like$
	$f-u$	评论数量	$comment$

2. 数据整合

我们整个数据集包含 9 月 1 日约 80 万人的近 220 次关注行为。在 30 天之内，所有被关注者产生了 2500 万次视频生产行为，所有的关注者产生了 6100 万次视频消费行为，而社交群体背后的社交网络规模高达 3.7 亿条记录。因此，代码执行效率成为我们更加关注的重点。

在行业老师和同学的帮助指导下，我们在快手的集群服务器上，通过创建 18 个临时表，消耗超过 200GB 内存空间，编译了超过 260 行 SQL 语句，创建了我们需要的“宽”数据表（见表 6）。我们后续的研究发现均通过该表分组和统计得到。

表 6：“宽”数据表实例

From_uid	To_uid	Follow From	Fan_from	Reci Follow	Follow Stack	Click To	Clicks	[41 个]
A	X	43	120	1	直播	10	1	...
A	Y	43	120	0	搜索	17	0	...
B	X	21	68	0	同城	10	0	...
C	Y	52	10	1	直播	17	2	...
C	Z	52	10	0	搜索	5	1	...
[220 万]

五、取关机制分析

1. APP 运行机制

(1) 关注人数

快手对于可关注人数设置了 1000 人的上限，根据行业经验，在用户关注人数即将接近上限时，会触发其整理已关注用户的列表，从而可能引发取关行为，为此我们对用户的关注人数进行如下分析。可以从图 4 中看出，在用户关注人数较少的阶段

（100 人以下），其后续取关的比例达到 18%；而当关注人数逐渐增加时，取关率逐渐降低；而当关注人数达到 800 人，取关比例又呈现上升趋势；当关注人数超过 900 人时，取关率达到了 16%。

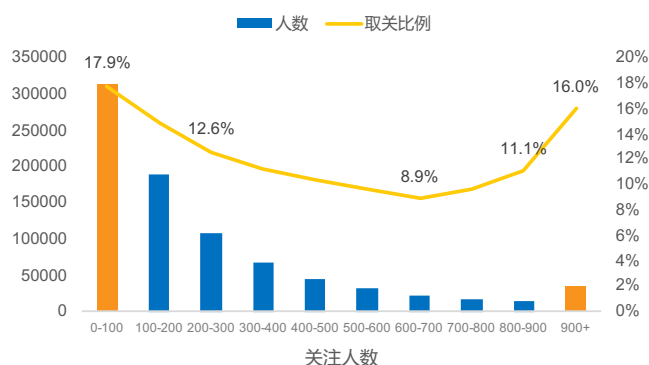


图 4：关注人数对取关率的影响

从用户的使用习惯来分析，当关注人数较少（低于 100 人）时，用户处于刚接触快手的探索阶段，会经常浏览所关注用户的视频，更新自己的关注列表也较为细心频繁；而随着关注人数的增加，新奇感和对被关注用户的关注度逐渐下降，用户不再经常检查自己的关注列表，也无暇顾及取关自己不喜欢的用户；当关注人数超过 800

人次时，由于将达到可关注人数上限，又促使用户不得不取关自己关注列表中不感兴趣的用户，从而可以关注新用户。

（2）点击率

用户关注了某一被关注者后，会收到被关注者发布的短视频、直播等一系列作品。为此我们定义点击率为用户的视频点击量除以展示的总视频数，并进一步探究点击率对取关行为的影响。通过图 5 可以看出，点击率小于 10% 的记录数有近 120 万条，在总记录数中的占比超过了 50%，且这一分组的取关率最低，可能的原因是用户完全没有看到被关注用户的视频；而当点击率大于 90% 时，取关率达到了 21%，我们分析这表明视频作者存在“刷屏”现象，或用户可能在初期浏览视频时由于不感兴趣就取关了视频作者。

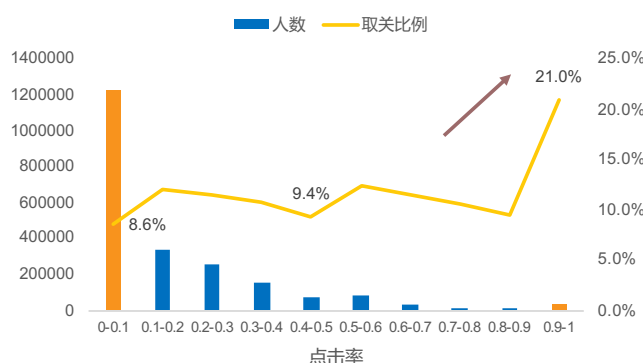


图 5：点击率对取关率的影响

2. 关注强度

（1）关注路径

在 220 万条数据集中，有约 18 万条数据关注路径缺失。在去除缺失数据之后，经统计，快手用户的关注路径有 64 种，我们选取了主要的 8 个关注路径进行分析。通过图 6 可以看出，通过推荐关注的用户取关率最低，而通过作品评论页、直播及私信进行关注的取关率均明显超过平均取关率，达到 14% 以上。我们分析认为以上三种途径更容易发生“骗粉”现象，从而导致用户更容易取消关注。

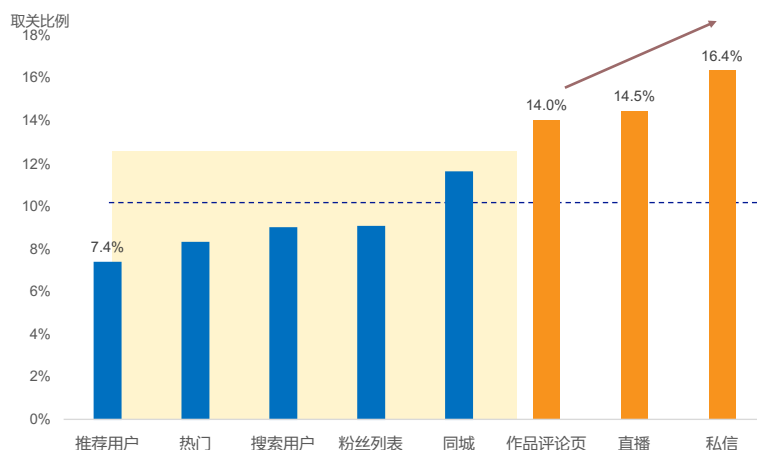


图 6：关注路径对取关率的影响

（2）双向关注

双向关系往往更为牢固，为此我们对用户相关关注的情况进行分析。我们发现单向关注用户的继续关注率为 89.4%，而双向关注的用户继续关注的比例会增加 6 个

百分点，可达 95%（见图 7）。因此我们认为，单向关注关系更容易破裂，相对地，如果用户间相互关注，他们的关注关系更为稳定。

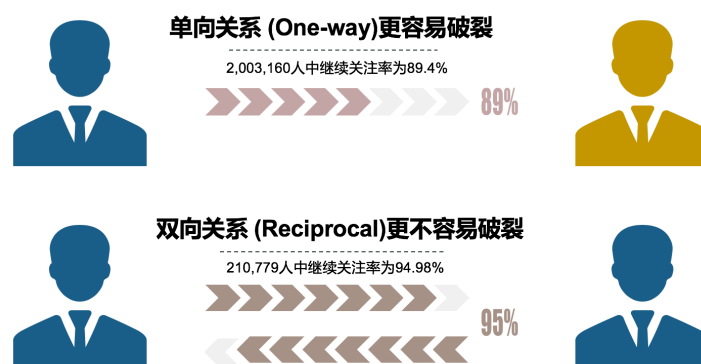


图 7：双向关注对取关率的影响

（3）平等关系

在研究用户间的关系时，我们通过比较用户的粉丝数量比，定义了一个用户关系是否平等的指标。如果用户双方的粉丝数量比在 0.5-2 之间时，认为用户的关系相对平等；当该比例小于 0.5 或大于 2 时，认为用户间是非平等关系。通过分析可以发现，在平等关系的数据中，继续关注率为 86%；而若用户间的关系不平等，继续关注率达到 90%。因此我们认为，平等关系中的用户更容易取关对方，而非平等的关系更为稳定。我们分析产生这一现象的原因是，在非平等关系中，如果粉丝数量比小于 1，这意味着被关注用户更受欢迎，关注用户也就更倾向于对其作品进行消费；而若粉丝数量比大于 1，关注用户的作品质量可能很高，那么他在关注用户时或许会更为挑剔谨慎，从而发生关注后也更不易取关。（见图 8）

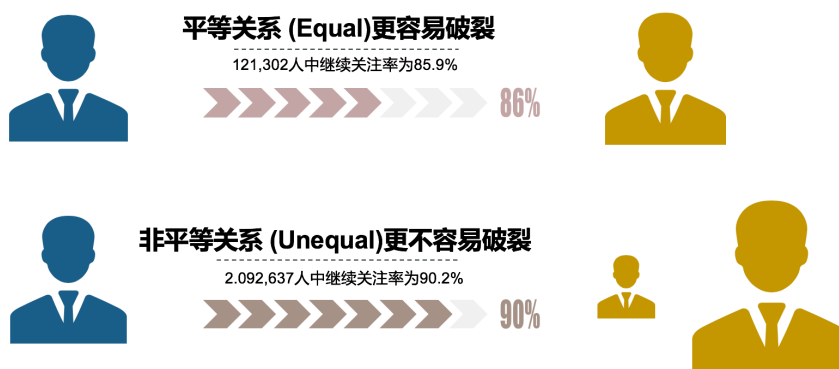


图 8：双向关注对取关率的影响

（4）共同关注

在分析用户双方粉丝数之后，我们还研究了双方好友数，并定义了共同好友数的概念，即关注用户和被关注用户的关注列表中相同用户的数量。由图 9 可以看出，当用户双方没有共同关注用户时，用户的取关率为 13.5%，而若存在两个以上共同好友时，取关率在平均取关率 10% 以下。我们分析认为，当用户双方存在共同关注时，用户可能为熟人关注或者兴趣相同，因此更不容易取关对方。

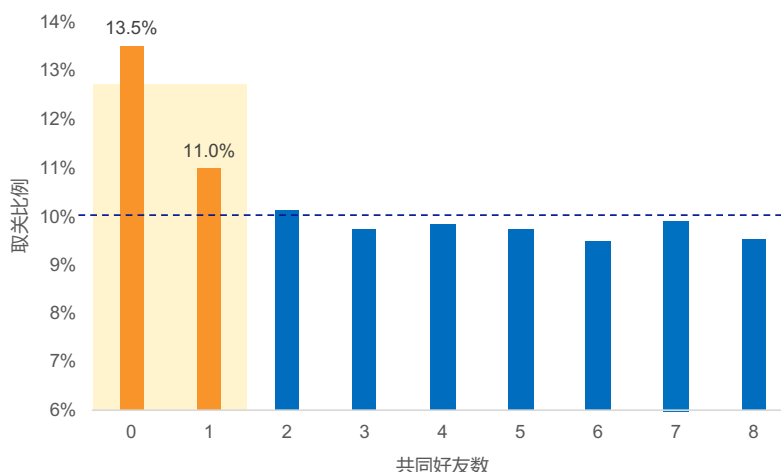


图 9：共同关注人数对取关率的影响

3. 作品质量

一般来说，关注某一个用户是为了进行作品消费，所以作品质量与是否取关息息相关。根据已有数据，并结合小组对于快手 APP 的使用经验，我们认为一个用户的作品质量与其作品所收获的点击数（click_to）、点赞数（like_to）、评论数（comment_to）、播放时长（playing_time_to）等指标有关。但是这些指标的绝对值会受到用户作品数量的影响，一般作品数量越多，其收获的点击数、点赞数、评论数和总观看时长就会越高。所以我们将这些指标除以这个用户所有作品的总数（live_and_photo_to）得到平均点击数、平均点赞数、平均评论数和平均播放时长，本小节，将分别从这四个角度来阐述其对取关率的影响。

（1）平均点击数

可以从图 10 看出的是，当作品没有被点击（即没有被任何人点开看过）或平均点击数较少时，取关率低于平均水平 10%。此处可以理解为用户在浏览时直接跳过了这些内容，没有特别注意，也就没有想到取关该作者。当平均点击数达到 0.5 以上，其取关率随着平均点击数的上升而下降，也就是说，作品的封面足够吸引人或作品质量越好，作品会被更多人阅读和喜欢，人们取关的概率就会越小。

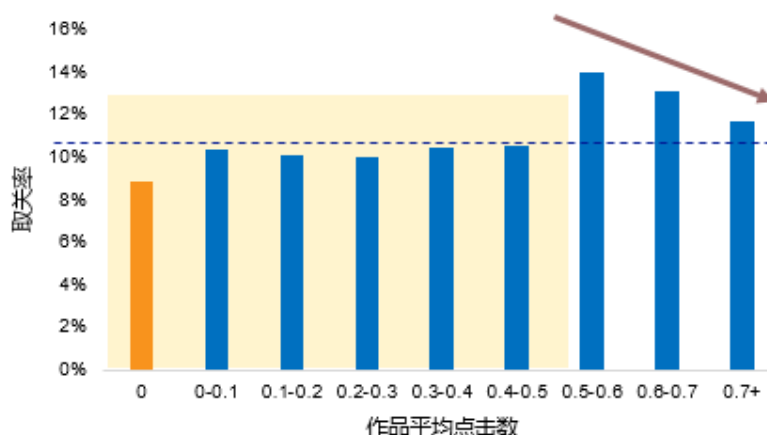


图 10：作品平均点击数对取关率的影响

（2）平均点赞数

从图 11 可以看出，当作者作品的平均点赞数小于 0.6 时，作者被取关的概率没有明显差异。而当平均点赞数超过一定程度（0.6）之后，取关率下降到了 8% 左右。

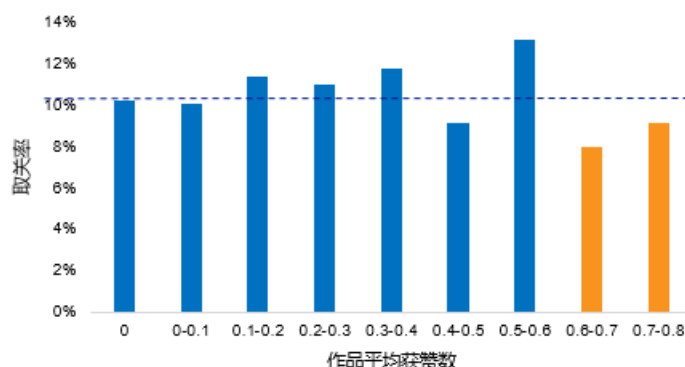


图 11：作品平均点赞数对取关率的影响

（3）平均评论数

关注者对被关注者作品的评论数与取关率的统计结果如下图 12 所示，如果关注者从来未评论过被关注者的作品，取关率也基本与 10% 持平，但当评论数增大时，取关率处于下降趋势。可能的原因是用户在作品下的留言得到作者回复，增加了用户对被关注者的互动程度。且互动得越频繁，心理感觉越亲密，于是就不会取消关注。

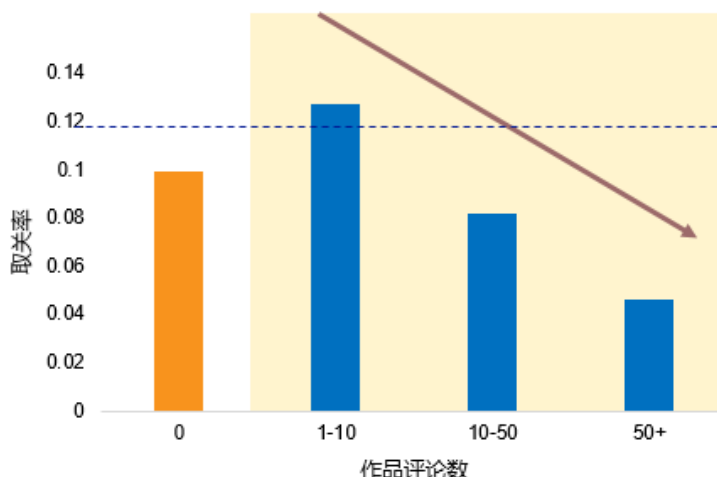


图 12：作品评论数对取关率的影响

（4）平均播放时长

平均播放时长对取关率的影响分析如下图 13 所示，当作品未被播放，即未引起关注者注意时，用户可能没有想到取消关注。但当作品播放时间较短时，取关率高达 12%。小组认为，极短的播放时间可能意味着用户点击观看了视频，发现视频不符合自己的期望很快退出，这样用户取关的可能性比较高。但是随着作品平均被观看时间越长，说明用户对作者的视频越兴趣，就更不容易取关作者。

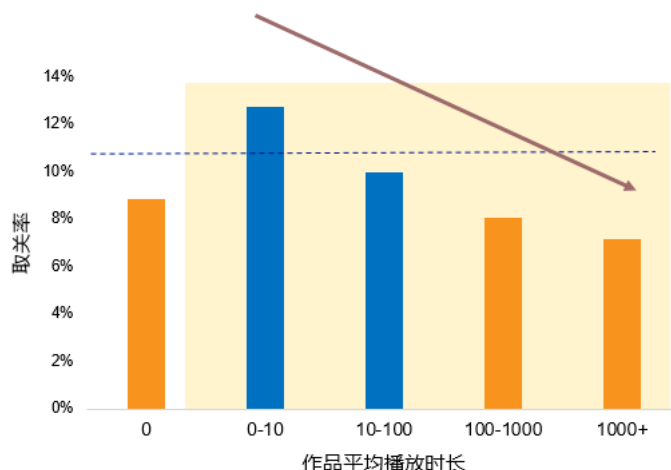


图 13: 作品平均被观看时长（秒）对取关率的影响

通过以上四个指标的研究，我们可以得出的基本结论是，如果作品没有被看到，那么作者就很难引起用户的注意，也就很少会触发取关行为。但当作品被看到时，其内容越吸引人、越有话题性，用户就越感兴趣，取关的行为也就会减少。

4. 用户互动

(1) 关注者对被关注者作品观看时长

可以从关注者对被关注者作品观看时间长度与取关率的统计结果（如图 14）看出，如果关注者并没有看过被关注者的作品，那么很有可能被关注者就不会引起关注者的过多注意，取消关注的可能也就比较小。当观看时间比较短的时候，取关率较大，可能是因为点进去看了不感兴趣，便取关了作者。随着观看时间变长，意味着关注者观看的时间变长，对之投入的情感、时间、经历也就越多，关注者与被关注者之间的联系也越来越紧密，取关率也逐渐降低。

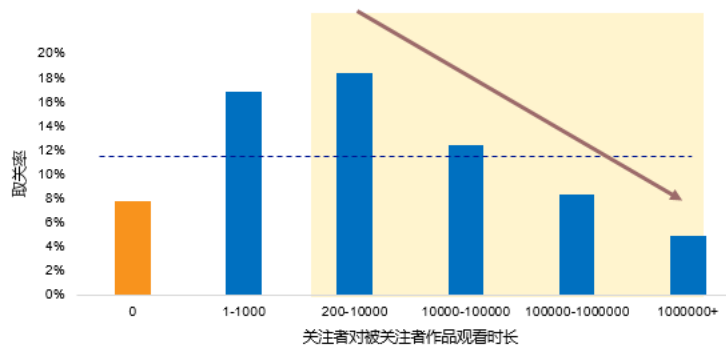


图 14: 关注者对被关注者作品观看时长（毫秒）对取关率的影响

(2) 关注者对被关注者作品评论数

关注者对被关注者作品的评论数与取关率的统计结果（如图 15）同上面的趋势类似，如果关注者从来未评论过被关注者的作品，取关率也基本与 10%持平，但是当评论数越来越大的时候，取关率确实处于下降的趋势。在作品下的留言很有可能得到回复，也增加了用户对被关注者的互动程度，互动得越频繁，可能心理的感觉就越亲密，于是就越不可能取消关注。

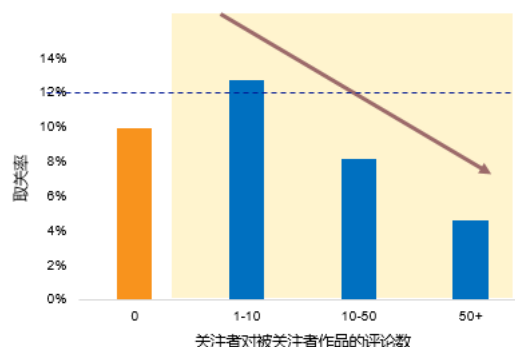


图 15：关注者对被关注者作品的评论数对取关率的影响

(3) 骗粉

在分析用户互动行为时，我们发现了一种异常的关注行为，经与行业老师讨论后得知，类似的行为称为“骗粉”。骗粉指的是为了让自己在短期迅速地增长粉丝、赢得更多流量与商业利益而进行的一系列活动，一般来说，骗粉者可能私下与被骗者达成协议，同时关注对方，为对方“涨粉”，之后骗粉者会取消关注被骗者从而欺骗更多的用户，积累更多的粉丝数。为了进一步分析这一现象，我们给出了互关时差的定义，即若 A 在 i 时刻关注 B，B 在 j 时刻关注了 A， i 与 j 的时间差便是互关时差。

由图 16 可以看出，当双方互相关注时，整体取关率都是偏低的。但当互相关注的时间间隔较小时，取关率还是处于一个较高水平。我们认为此处可能涉及到骗粉的问题。而如果双方互关时差较大，取关率也逐渐减小，骗粉的可能性也大大降低。

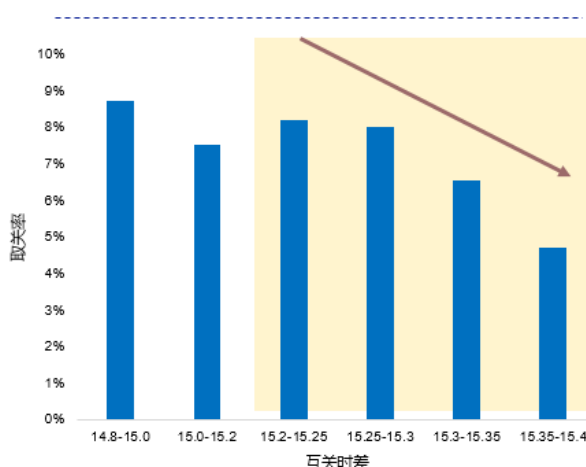


图 16：关注者与被关注者互关时差对取关率的影响

5. 案例分析

为了验证我们之前通过统计性分析得出的结果，我们针对 APP 关注上限和“骗粉识别”两个方面，从数据中找到了可以应证的案例进行详细分析。

(1) APP 关注上限

快手对于用户的关注人数设置了 1000 人上限，我们曾在前文讨论过，当关注人数超过 900 人时，取关率上升到 16%，远高于平均取关率 10%。在所有的用户信息中，我们也挑选了一位典型用户，其在 9 月 1 日关注了 20 人，总关注人数达到 998 人，这也意味着其很快就要达到关注人数上限。而对于被关注的 20 名用户，他在第二

天 7:41-7:47 时间段内一共取关 19 人。可以想象的一个场景是，此用户发现自己即将到达人数上限，便开始整理自己的关注列表，从而可以关注新的用户。

表 7：因关注人数接近上限而发生的集体取关行为

用户 ID	被关注者 ID	是否取关	取关时间
bOK5MWiKT JLAWc4eEkV8Q==	4X4w7l8XR49m+3I9apAR2A==	是	2018/9/2 7:42
	yLjqcnAjnsNvG1zsxxhn+A==	是	2018/9/2 7:42
	5fT949/1gMEBO/4P81j1bQ==	是	2018/9/2 7:41
	8s6d8R+RnMnXuvdL1J18/w==	是	2018/9/19 7:47
	E2qsrnqUIH9X/e/8LLzExQ==	是	2018/9/2 7:41
	qk5CYUMvctwbZ4ttfnJB9A==	是	2018/9/2 7:42
	07iRY3RIB8Je+0eUhzkr7A==	是	2018/9/2 7:42
	Oup5rA2hAUS44qDp17bYQw==	是	2018/9/2 7:41
	sDsBmmoln9DeLe+fbhaGCg==	是	2018/9/2 7:42
	7Rungs9Qth5qgZkKVKctyw==	是	2018/9/2 7:41
	PLH65vQ85GTmyL4vbXhRtw==	是	2018/9/2 7:41
	hyiBHG0KZIOGIMaUCfJmQ==	是	2018/9/2 7:41
	9vyGOj/n4mYmOgZ3ErMPGQ==	是	2018/9/2 7:41
	23y83Sn6gme/ZsW5U6ZRvg==	是	2018/9/2 7:41
	RH2U0eyrav7plKMBn46F8Q==	是	2018/9/2 7:41
	p9GxOUJFR5xSNfQzPZTLpQ==	是	2018/9/2 7:41
	tALMQnwSZ1CgO57GvzbteA==	是	2018/9/2 7:42
	FbOPQknQHM5/5P+dhngwZw==	是	2018/9/2 7:41
	fPU5gw5l5lb8IxRmQK70tQ==	是	2018/9/2 7:42
	lrzT/fHMNEyrh6YPkqJ9gw==	否	null

(2) 骗粉行为识别

在短视频 APP 中，骗粉是一个普遍存在的现象。骗粉者往往会与其他用户私下达成协议，互粉彼此，为彼此涨粉，然而用户之间不产生任何互动，不会互相浏览作品，且骗粉者在互粉之后很短时间就取消关注，从而进一步去骗其他人互粉，如此循环往复，达到为自己涨粉，让自己成为“大 V”的目的。

为了通过用户数据识别骗粉，我们给出了以下“骗粉”行为的标准：

- 双方互相关注，且互关时差不会太长
- 关注之后短时间内便会发生取消行为
- 关注之后双方没有互动，即关注者不会点击、观看、点赞、评论被关注者作品
- 关注途径极有可能是通过私信【6】、消息【17】、直播【7】、用户搜索【2】等，而热门【8】、用户推荐【1、18、19、20、21、22】³则比较少

基于以上四条标准，我们也试图识别了一些典型案例：

³ 【】内是数据中对于每一个路径的编码

表 8：骗粉行为案例

关注者 A 的 ID	dmALn6teQKC5np+x1nZXmQ==
被关注者 B 的 ID	0p9a8a6W/TXpCZ9ecKVEsg==
A 的粉丝数	943
B 的粉丝数	204
关注渠道	私信
A 对 B 的作品的关注	0 点击, 0 点赞, 0 评论
B 的作品受到的总关注	12 次点击, 6 个赞, 1 个评论
A 关注 B 的时间	2018/9/1 8:34 AM
B 关注 A 的时间	2018/9/1 8:34 AM
是否相互关注	是
A 取消关注 B 的时间	2018/9/3 1:37 AM

可以看出的是，A 本人粉丝较多，其与 B 在 2018 年 9 月 1 日 8:34 同一分钟内通过私信互相关注，这之前 B 的作品数据记录包含 12 次点击，6 个赞和 1 个评论，但是 A 没有观看过 B 的任何作品，也没有贡献过任何赞或评论。且值得关注的是，A 在关注了 B 两天之后就将 B 取关了。

为了进一步观察 A 的关注行为，我们特意抽出 A 互相关注的关注记录进行分析：

表 9：用户 A 其他的关注记录

ID	关注渠道	互关时差	取关时差	点击数
1	私信, 热门	00 00:02:00	未取关	9
2	私信, 消息	00 05:01:00	未取关	3
3	私信, 热门	00 00:21:00	未取关	4
4	私信, 热门	00 05:41:00	10 08:26:00	0
5	私信, 热门	13 11:55:00	10 07:49:00	0
6	私信, 热门	00 00:45:00	10 07:48:00	0
7	私信, 消息	00 00:00:00	01 17:03:00	0
8	关注页	00 00:00:00	01 05:30:00	0

可以看出的是，A 用户经常与其他用户在很短时间内发生互相关注，并且与被关注用户都无互动，在未来几天内就会将其取关，由此我们分析，A 可能会是恶意骗粉的用户。

六、预测模型

1. 基准 Logistic Regression 模型

为量化研究取关行为的发生，我们通过建立预测模型，进行进一步分析。数据的因变量为二元 0、1 变量，也就是关注者 f 在 9 月 1 日关注了被关注者 u 之后，在 9 月 30 日的状态为继续关注（0）还是取消关注（1）。自变量则为上述用户结构属性变量和用户行为属性变量。

由于数据量过于庞大，我们小组从全量数据中随机抽取 20000 条取关记录和 20000 条未取关记录，构建均衡样本，并对离散变量构建虚拟变量纳入自变量。基于此均衡样本，我们首先使用 Logistic Regression 作为基准模型，输出模型混淆矩阵。通过图 17 我们发现，Logistic 回归模型在训练集中的准确率为 70.1%，在测试集中的准确率为 70.4%，测准率为 68%，测误率为 72.8%，召回率为 71.4%。

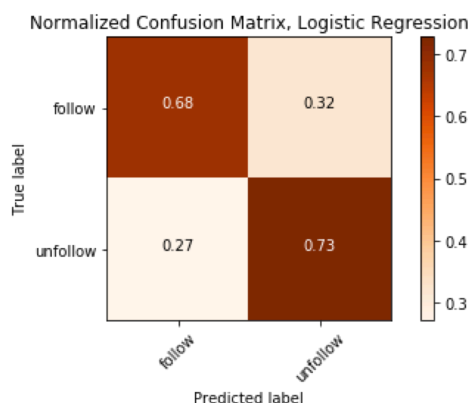


图 17: Logistic Regression 预测混淆矩阵

2. 更多机器学习分类器准确性比较

为比较不同机器学习分类器的预测准确性，并检测基准模型稳健性，我们这里使用 Python Sklearn 中的分类器，对基准模型中的数据分别进行训练和检验：包括线性 SVC（Linear SVC）、决策树（Decision Tree）、随机梯度下降（SGDC）、朴素贝叶斯（Naïve Bayes）和神经网络（Neural Network）。对于神经网络模型，我们分别使用了单隐层 256 个节点、单隐层 128 个节点、单隐层 64 个节点、单隐层 32 个节点、双隐层（128, 64）个节点和三隐层（128, 64, 32）个节点的神经网络等 11 种机器学习模型。

表 10 Panel A 中汇报了上述 11 种机器学习模型的表现。我们分别汇总了每个模型在训练样本中的准确率（Training Accuracy）、在测试集中的查准率（Testing Precision）、召回率（Testing Recall）和准确率（Testing Accuracy）。总体而言，不同模型在训练集和测试集中的准确率相差较小，这表明我们的模型不存在明显的过拟合问题。

通过比较我们发现以下三点，（1）相较于其它机器学习模型，神经网络总体而言具有更好的表现，在训练集和测试集中的准确率均高出 3-4%。（2）但是神经网络的深度对预测准确率的提高相对有限，不同深度神经网络之间的准确率相差小于 0.5%。这也与我们的直觉吻合，由于我们设计出的预测指标与取关行为存在较为直接的关系，更深的神经网络并不能发掘出更多的预测信息。（3）而对于单隐层的神经网络，更多的节点可以更好地捕获预测变量之间的关系，因此预测的准确率也就越高。

考虑到相同预测变量对于不同性质用户可能存在不同的预测作用，我们继续分别对普通用户（粉丝数小于 1,000 人）、大 V 用户（粉丝数在 1,000 人到 10,000 人之间）和超级大 V 用户（粉丝数大于 10,000 人）分别使用神经网络模型进行预测。该用户性质的划分标准为快手公司内部默认标准。

表 10 Panel B、Panel C 和 Panel D 分别汇报了对普通用户、大 V 用户和超级大 V 用户的取关预测模型的准确率。通过对人群细分，我们的子模型准确率总体而言相较于基准模型有 2-3% 的准确率提升。

表 10: 不同模型分类器样本内外准确率比较

Panel A: 全样本抽样数据对不同模型分类器的准确率比较

	Logistic Regression	Linear SVC	Decision Tree	SGD	Naïve Bayes	NN1 (256,)	NN1 (128,)	NN1 (64,)	NN1 (32,)	NN2 (128,64,)	NN3 (128,64,32,)	Ensemble Model
Training Acc.	70.1%	69.6%	62.0%	67.1%	59.6%	74.5%	74.2%	73.9%	73.4%	73.1%	73.8%	CV= 5
Testing Precision	68.0%	67.0%	63.4%	78.4%	51.2%	78.8%	78.1%	79.2%	79.4%	78.4%	79.2%	
Testing Recall	71.4%	71.6%	63.0%	63.8%	60.5%	72.7%	72.1%	71.5%	70.4%	71.1%	70.9%	
Testing Accuracy	70.4%	70.2%	63.1%	66.9%	58.9%	74.6%	73.9%	73.8%	73.0%	73.3%	73.4%	72.3%

Panel B: 粉丝数 1,000 以下抽样数据对不同模型分类器的准确率比较

	Logistic Regression	Linear SVC	Decision Tree	SGD	Naïve Bayes	NN1 (256,)	NN1 (128,)	NN1 (64,)	NN1 (32,)	NN2 (128,64,)	NN3 (128,64,32,)	Ensemble Model
Training Acc.	72.4%	72.3%	70.0%	72.4%	66.3%	79.1%	78.9%	78.4%	77.3%	77.5%	77.8%	CV= 5
Testing Precision	71.2%	71.2%	83.0%	79.7%	61.7%	82.6%	82.5%	82.4%	83.6%	83.1%	82.2%	
Testing Recall	71.5%	72.0%	65.4%	69.2%	68.1%	74.7%	74.4%	73.6%	73.2%	73.6%	73.9%	
Testing Accuracy	71.4%	71.7%	69.6%	72.1%	66.4%	77.3%	77.0%	76.5%	76.5%	76.7%	76.6%	76.4%

Panel C: 粉丝数 1,000 以上 10,000 以下抽样数据对不同模型分类器的准确率比较

	Logistic Regression	Linear SVC	Decision Tree	SGD	Naïve Bayes	NN1 (256,)	NN1 (128,)	NN1 (64,)	NN1 (32,)	NN2 (128,64,)	NN3 (128,64,32,)	Ensemble Model
Training Acc.	70.5%	70.3%	68.1%	70.4%	59.2%	74.2%	74.2%	74.0%	73.6%	73.9%	73.5%	CV= 5
Testing Precision	67.8%	66.6%	69.8%	85.1%	55.4%	75.9%	75.3%	74.9%	74.6%	75.6%	75.4%	
Testing Recall	72.0%	71.9%	67.5%	65.5%	60.1%	72.8%	72.5%	72.9%	72.4%	72.6%	71.5%	
Testing Accuracy	70.7%	70.3%	68.1%	70.1%	59.3%	73.8%	73.4%	73.6%	73.1%	73.5%	72.6%	72.9%

Panel D: 粉丝数 10,000 以上抽样数据对不同模型分类器的准确率比较

	Logistic Regression	Linear SVC	Decision Tree	SGD	Naïve Bayes	NN1 (256,)	NN1 (128,)	NN1 (64,)	NN1 (32,)	NN2 (128,64,)	NN3 (128,64,32,)	Ensemble Model
Training Acc.	69.9%	69.8%	65.2%	69.5%	59.0%	74.1%	73.7%	74.1%	73.2%	73.7%	73.6%	CV= 5
Testing Precision	65.4%	64.9%	45.4%	66.2%	53.1%	76.0%	76.5%	75.6%	76.4%	76.1%	77.2%	
Testing Recall	72.4%	72.3%	72.1%	71.5%	60.6%	72.4%	72.2%	72.4%	71.4%	72.3%	72.3%	
Testing Accuracy	70.3%	70.0%	63.9%	69.9%	59.3%	73.6%	73.5%	73.4%	72.9%	73.5%	73.8%	72.5%

Table10 报告了不同样本下不同分类器的准确率指标比较。Panel A 报告在全部 220 万数据中随机抽取 20000 条取关记录和 20000 条非取关记录构建得到均衡的数据集进行模型预测。

Panel B、Panel C、Panel D 则对用户粉丝数特征进行分层抽样，分别汇报不同模型下的准确率。NN1、NN2、NN3 分别为为单、双、三隐层全连接神经网络，括号中的数字为神经元的数量 Ensemble 模型为通过前 11 个模型构建得到的聚合模型，使用 hard voting 得到最终预测结果

图 18 比较了未分组和分组模型在 Logistic Regression, Linear SVC, Neural Network 三种分类器中的准确率。我们发现, 分组模型在普通用户人群

(fans<1000) 中有较为明显的提升(准确率提升约为 2%-3%)。这说明普通用户的行为更有可能是同质化的, 通过我们构建的指标可以更好地预测取关行为的发生; 而大 V 和超级大 V 的行为则更有可能是异质化的, 每个大 V 尝试着作出不同的改变, 从而吸引更多的粉丝, 因此对应的取关行为也就更加难以预测。

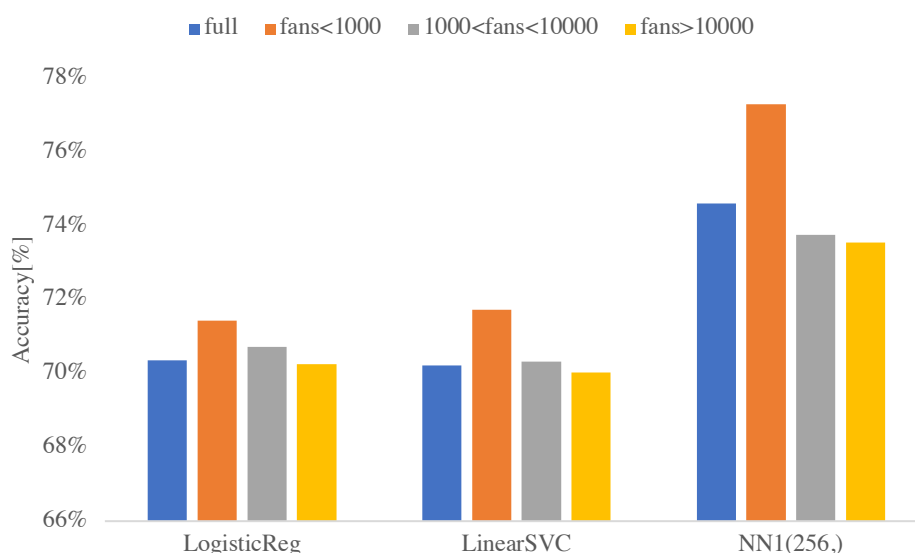


图 18: 分组预测模型准确性对比

3. 集成模型

为了进一步提高预测准确率, 我们尝试从模型优化的角度进行探究。为此我们绘制了特征重要性图(如图 19 所示)。从图中可以看出, 互相关注和共同好友在不同模型中都有非常重要的地位, 这一点和我们取关机制的分析结果一致。而从不同变量在同一模型的重要性表现来看, 朴素贝叶斯模型对于被关注用户的特征提取有更突出的表现。基于这一发现, 我们采用集成模型对用户的取关进行预测, 最终准确率为 72.3%, 相比基准模型提高了 2%。从结果来看, 预测准确率并没有明显提升, 我们分析认为, 这是由于选择的模型对于用户特征提取效果整体较为相似, 导致用户特征尚未被充分利用。在今后研究中, 可以尝试构建抽象变量和调试模型参数, 从而改善集成模型的效果。

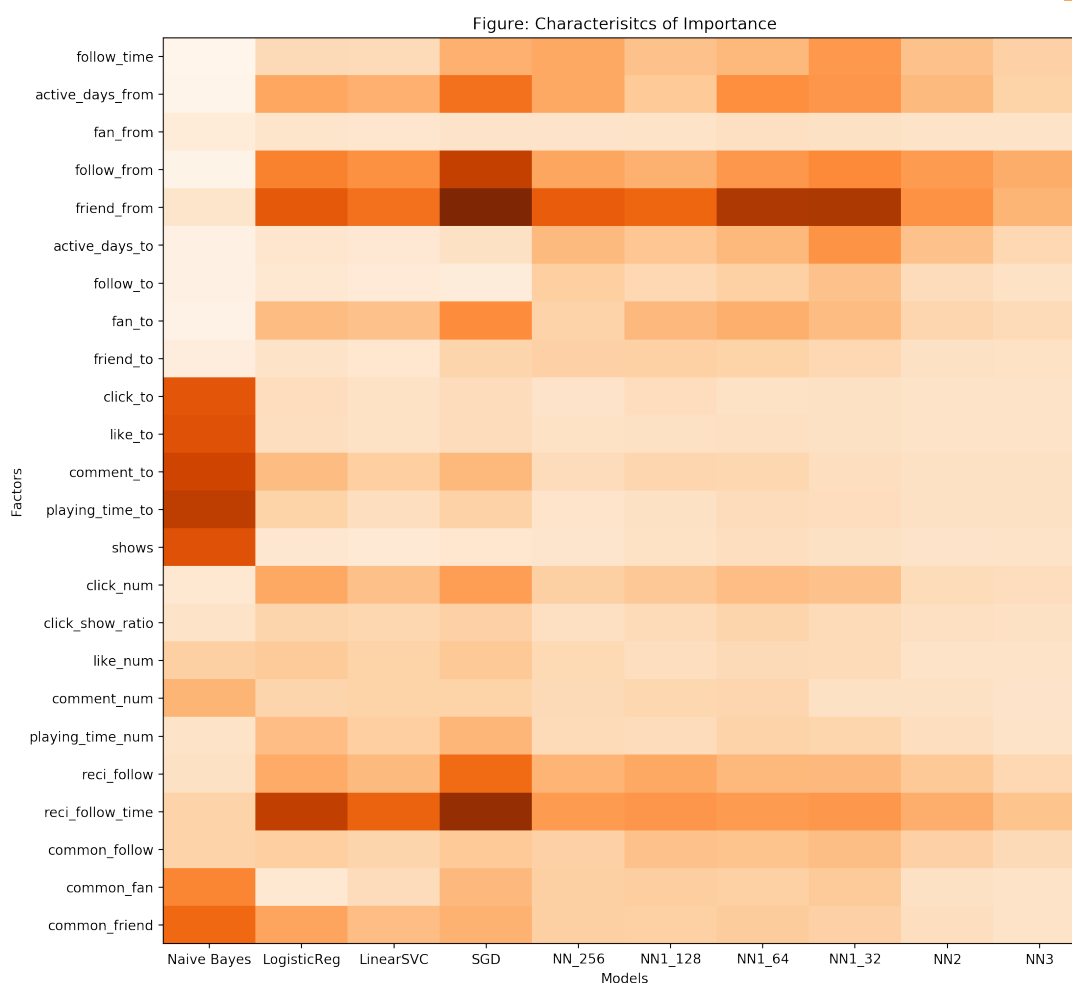


图 19：特征重要性图

七、Spark 平台全量数据分析

为避免抽样误差等因素对结果的影响，我们使用 9 月 1 号全量 220 万条数据进行模型训练。

快手提供了 Python/R + Hue + Spark 的大数据计算引擎解决方案。其允许每个人通过 Hue 平台，在线编写 Python 或 R 代码；通过调用 Spark 包，进行高层次的框架编程，并将数据和模型传输给后台 Spark 平台进行高速并行模型分析。

我们在使用 Hue 进行 SQL 操作的过程中，帮助快手发现了平台在数据前端与后端交互设置中的异常。通过与后台工程师积极交流，我们暂时只能通过直接读取数据库文件，生成 DataFrame 进行后续操作。我们也将该发现汇报给快手相关负责人，撰写了详细的公司内部 Wiki⁴，以指导快手的其他程序员避免在该问题上重蹈覆辙。

我们选择全部 22 万取关样本并随机抽取 22 万未取关样本，构建得到 1: 1 均衡训练数据集。这里我们选择 Logistic Regression with LBFGS 和 SVM with SGD 两种

⁴公司内部 Wiki 链接: <https://wiki.corp.kuaishou.com/pages/viewpage.action?pageId=111369162>

方法进行训练。由于（1）Spark 平台上支持并行训练的算法较少，仅包含 Logistic Regression, SVM 等线性方法；（2）特定 RDD 数据格式功能复杂；（3）Spark 平台惰性操作，代码错误较难调试；（4）全量数据可能存在更多噪音等原因，我们使用全量数据训练的结果相较于本地抽样训练结果并没有显著提升。Logistic Regression with LBFGS 的准确率为 60.5%，SVM with SGD 的准确率为 64.2%。

尽管我们的模型准确率没有显著提高，我们小组成员在这一过程中仔细学习了如何在 Spark 平台进行数据处理和逻辑操作，得到了非常宝贵的经验知识。

八、总结与展望

快节奏的生活、对虚拟社交的需求和碎片化的娱乐习惯，让短视频 APP 异军突起。为了让快手实现高质量的粉丝增长、提高用户社区质量、在激烈的竞争中保持领先地位，本小组从取消关注这一重要的用户行为入手，使用大数据分析方法，建立相关模型，探究用户取消关注背后的机制，提升快手对用户使用习惯的了解，优化包括推荐机制在内的算法，提高用户粘性。

基于公司提供的超过 220 万条用户数据，我们小组对取消关注这一行为进行了较为系统的分析。经过对过往社交理论、互联网产品理论和产品的真实使用体验，我们从四个角度对取关行为进行诠释：

（1）APP 本身机制

比如 APP 设置的关注人数上限，会导致人们在接近关注上限时产生大量取关行为。APP 内容分发机制可能带来的“刷屏”现象，也会导致“视觉疲劳”而促使用户取关。

（2）用户本身特质

用户关注路径对取关有非常显著的影响，通过推荐关注的用户取关率最低，通过作品评论页、直播和私信进行的关注，其对应的取关率明显较高，且可能涉及“骗粉”现象。

除此之外，双向关注中的关系往往更加牢固，其取关率普遍较低。当关注行为发生在粉丝规模不相当的用户之间时，取关率也会偏低。且当用户双方有了较多共同好友时，也更不容易取关对方。

（3）作品质量

我们用作品收获到的平均点击数、点赞数、评论数和播放时长来衡量作品的质量，发现一般来说作品的质量越高，取关率也就越低。但是如果作品没有被任何人看过，此用户可能会被忽视，取关率在数据显示上也比较低。

（4）用户互动

用户之间的互动行为会消耗情感、时间和精力，所以互动行为越多（比如用户之间的点赞、观看、评论），用户之间的取关率也就越低。

经过初步的统计性论证，我们证实了所提出的取消关注行为背后的四条机制的正确性，接下来，我们通过案例分析，从更加直观的角度验证了以上机制。

由于数据量过大，我们对数据进行了抽样，并用均衡样本对数据进行了建模，通过决策树、logistic 回归、Naïve Bayes、神经网络、Linear SVC 和 SGD 等模型对取关进行了预测，并通过 Ensemble Model 将这些模型进行整合，弥补各自的缺陷，提高了预测的准确率。

基于我们的研究结果，我们小组尝试着提出以下几点建议，以期提高产品的用户体验，降低取关率，提高用户留存：

（1）优化内容分发机制

根据我们之前的研究，“刷屏”现象会导致取关率的升高。所以除了根据用户的偏好给用户推送其感兴趣的内容之外，快手还应当注意减少“刷屏”现象的出现，避免同一作者的作品在短时间内密集出现在推荐页上，可以适当地设置同一作者的作品出现的最大频率。

（2）引导用户互动

用户之间的互动会增加用户彼此之间的黏性，也就降低了取关行为发生的可能性。通过引导用户参与互动，比如优化评论栏，方便用户在欣赏作品时随手评论。

其次，可以考虑设置“精品评论”，作品作者可以将自己喜欢的、专业的、有趣的评论“加精”，这样用户得到了关注者的认可，也加强了认同感和互动感。

另外，快手运营部门也可以引导话题性的评论或者转发，提高一些没有参与评论、互动意识的用户的参与意识，从而增加黏性，降低取关率。

最后，在用户决定取消关注某一用户时，可以弹出小窗，提示用户在关注的这一段时间内与此用户的互动，比如“2018 年 9 月 1 日至今，您观看了 xx 三则短视频、四次直播，点赞 3 次，评论 2 次”之类，提醒用户之前在此用户上投入的情感、时间与精力，从而可能会起到降低取关率、留存用户的效果。

（3）加强对骗粉的识别

骗粉行为的存在可能会带来较差的用户体验，公司可以加强对这方面行为的识别，如果某些用户此种行为过多，给其他用户带来的影响较大，则可以后台提醒。

实验设想⁵：

针对骗粉现象的识别，我们设计了两个小实验，如果未来条件允许，可以付诸实践，进一步验证我们的猜想：

1) 设置关注上限“软限制”：在 APP 中植入提醒机制，当用户关注的人数达到 900, 950, 990 时分别弹出提示框，提示用户关注人数将达到上限。如果我们的猜想成立，那么预计在关注数达到 900、950 和 990 时，用户会产生比较密集的取关记录；

2) 设置关注上限“硬限制”：对于部分内测用户改变其可以关注的人数上限为 800 和 1200。如果我们的猜想成立，那么预计在关注数达到 1000 左右的时候，用户不会产生如上的密集取关记录。

⁵ 感谢徐心教授为实验设想部分提出宝贵的建议

最后，我们组也在导师的帮助下用 Spark 平台对全量数据进行尝试分析，在未来可以利用更多数据，对某些特定人群进行追踪研究，深化案例分析；也可以对用户作品内容、标签等进行探索，观察不同话题对于用户的影响；由于目前时间限制，我们直接使用了默认的参数，未来可以对模型进行调参，构建更加精确的模型。

本项工作历经六周，不仅让小组成员对于短视频 APP 用户取关行为有了深刻的理解，也通过对大量数据的探索、分析，提高了 SQL、Python 的编程能力及大数据分析能力。同时也让我们深入体验了互联网公司工作的氛围和文化，让我们跳出学术的角度，培养商业分析思维，从企业实际需求的角度进行我们的研究实践。

喜欢快手的“老铁”们，喜欢周二周四相聚的龙吟阁，喜欢快手丰盛的午餐与下午甜点，也喜欢小组中彼此帮助、各自发挥所长的氛围！在此，我们不仅要感谢老师给我们提供这样宝贵的机会，也要感谢快手的廖闯、郭凡两位老师和助教学姐对我们的大力支持，让我们在轻松、严谨的氛围内，完成了这次商务实践，让我们每一个人都受益匪浅。