

国内互联网广告无效 流量白皮书

2018

CONTENT

01 背景介绍

02 2018 年无效流量分析

03 国双的监测方案

04 总结

01

背景介绍

背景介绍

网络时代里，人们把越来越多的时间花在了电脑和手机上，进入移动互联网时代以来，手机甚至成为了一个人体的延伸，大量的关注都投放在各类 APP 上。这种空前的关注度，也引起了各大广告主的青睐，纷纷把广告预算分拨到数字广告里。据 eMarketer 的预计，美国市场数字广告的预算将在 2019 年超过所有传统媒体，显示了数字广告迅猛的发展势头。

然而，2016 年发现的名为“MethBot”的大规模机器人流量源，每天都在攫取数百万美元的广告预算，近期，美国 FBI 联合 Google、White Ops 等科技公司一起捣毁了一个传播计算机病毒以谋取利益的窝点，更别提国内早已泛滥且体系成熟、分工明确的刷点击广告、刷留资、刷下载 APP、模仿正常用户玩游戏等一系列的作弊行为。相比传统广告，数字广告领域更加多样且恶劣的作弊行为，极大地侵害了广告主的利益，严重地破坏了整个行业生态的健康发展。

当前，行业内各方都在积极对抗无效流量，为获得真实数据而努力。中国广告协会作为中国广告业权威行业组织，一直致力于提升数字广告数据真实性和流量透明度。“数据、流量造假在数字广告行业的严重性已不容忽视。行业各方对数据、流量真实性的追求和对社会的责任感都对数字广告行业有重大影响。”中国广告协会会长张国华如是说道。

背景介绍

作为独立的第三方监测公司，国双是中国广告协会理事单位，同时也是中国广告协会互联网广告委员会常务委员单位。国双有义务，有责任为广告主和整个行业，检测并识别出这些作弊行为和无效流量，助力广告主掌控每一份广告预算的花费，还原广告的真实效果，并促进行业的健康发展。

为此，在中国广告协会的指导下，国双于 315 消费者权益保护日，制作并发布《国内互联网广告无效流量白皮书》，旨在揭示国内的互联网广告流量现状，为广告主、代理公司等行业参与者提供直观的数据参考，同时展示国双多年来总结并完善的监测方案，以让广告主在以后的数字营销中更加有底气。

02

2018 年无效流量分析

无效流量标准

本报告针对的无效流量，是参照了国内权威组织中国广告协会标准对于无效流量的定义和分类。这套标准是在欧美成熟市场的无效流量标准基础上，结合了国内实际的互联网技术环境，对部分标准做了针对性的调整之后形成的。

根据国内中国广告协会的标准以及国外 MRC/TAG 的标准，无效流量被分为两类——GIVT 和 SIVT：

GIVT 是 General Invalid Traffic 的缩写，即常规无效流量，是指能够通过应用多种名单或标准化参数等常规方式进行过滤的流量。

SIVT 是 Sophisticated Invalid Traffic 的缩写，即复杂无效流量，这种类型的流量无法通过简单的规则识别出来，一般需要通过高级分析，多方合作与协调，乃至人工干预等方法以及广告投放活动以外更大范围的数据信号才能分析和识别。

无效流量标准

中国广告协会对于无效流量的完整分类，具体如下：

GIVT/ 常规无效流量	SIVT/ 复杂无效流量
机器人和爬虫或其它伪装成合法用户的流量数据以及非浏览器用户代头或其它形式的未知浏览器带来的流量	劫持设备以及设备中的会话
超出频次、时间间隔等目标设定的流量数据	非法劫持广告创意和操纵流量
通过隐藏 / 堆叠 / 覆盖或其它方式导致用户无机会看到正常广告内容的流量	内容盗用、伪造、虚假展示
已知的来自数据中心的流量（指明显具有非人类访问广告所在的特定网络 IP 或 IP 段所产生的流量来源）	恶意修改、插入或删除 cookie 内容以改变用户访问记录
预获取或浏览器预览的广告流量	操纵或伪造位置数据以及相关属性
已知的来自高危或者作弊来源流量	无效代理流量（即来自中间代理设备的无效流量，包括通过代理设备操纵流量计数、创建 / 传输非人类流量或无法通过协议验证的流量）
基本信息缺失或不一致的流量（基本信息至少应包含事件类型、广告系列 ID、时间戳、IP、请求方式、用户代理 UA 字段）	

无效流量标准

MRC/TAG 对于无效流量的完整分类，具体如下：

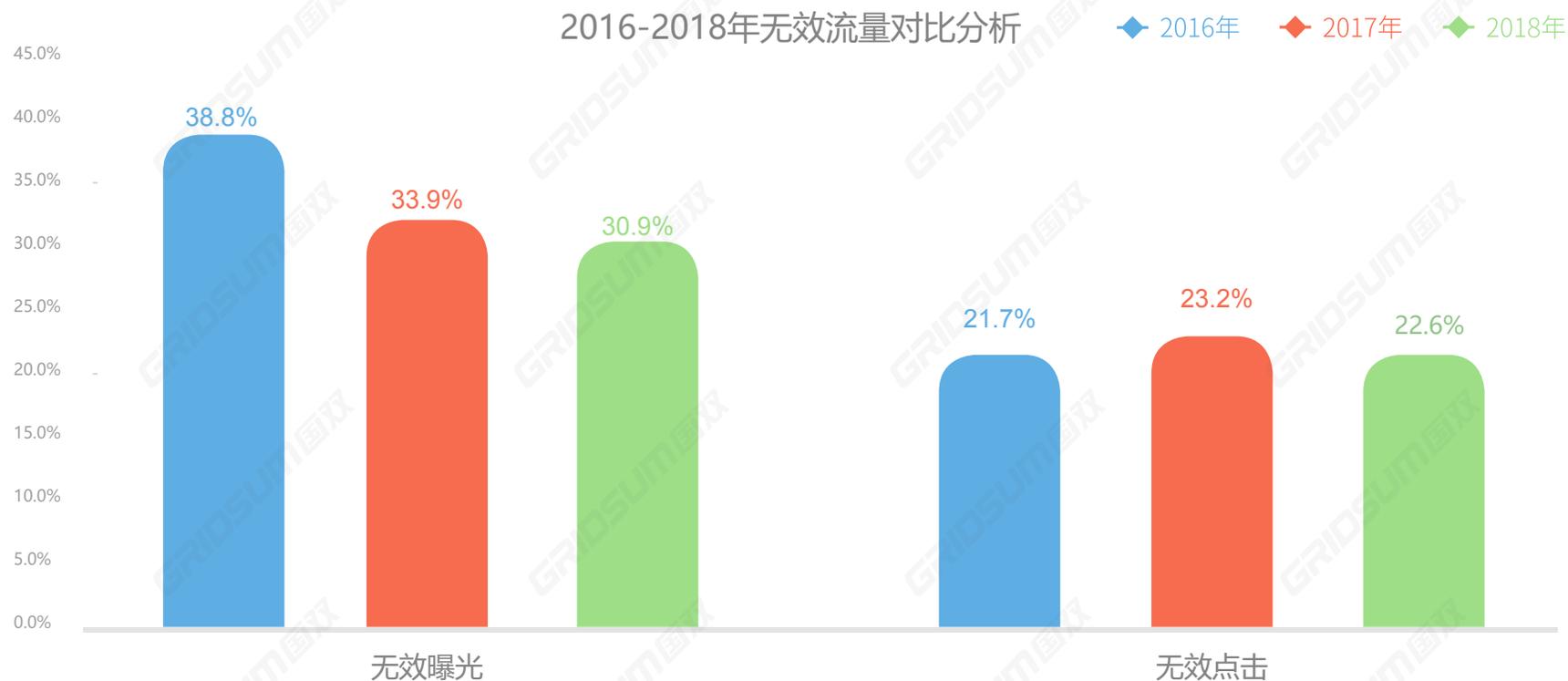
GIVT /General Invalid Traffic	SIVT/Sophisticated Invalid Traffic
Known data-center traffic (determined to be a consistent source of non-human traffic; not including routing artifacts of legitimate users or virtual machine legitimate browsing)	Bots and spiders or other crawlers masquerading as legitimate users
Bots and spiders or other crawlers	Hijacked devices; Hijacked sessions within hijacked devices; Hijacked ad tags; hijacked creative
Activity-based filtration using campaign or application data and transaction parameters from campaign or application data	Hidden/stacked/covered or otherwise intentionally obfuscated ad serving
Non-browser user-agent headers or other forms of unknown browsers and pre-fetch or browser pre-rendered traffic	Invalid proxy traffic
	Adware; malware
	Incentivized manipulation of measurements
	Misappropriated content
	Falsified viewable impression decisions; falsely represented sites or impressions
	Cookie stuffing
	Recycling or harvesting
	Manipulation or falsification of location data or related attributes
	Differentiating human and IVT traffic when originating from the same or similar source in certain closely intermingled circumstances

无效流量数据分析

通过对 2016—2018 年的数据进行全面地监测和分析后，结论如下。

全年整体分析：

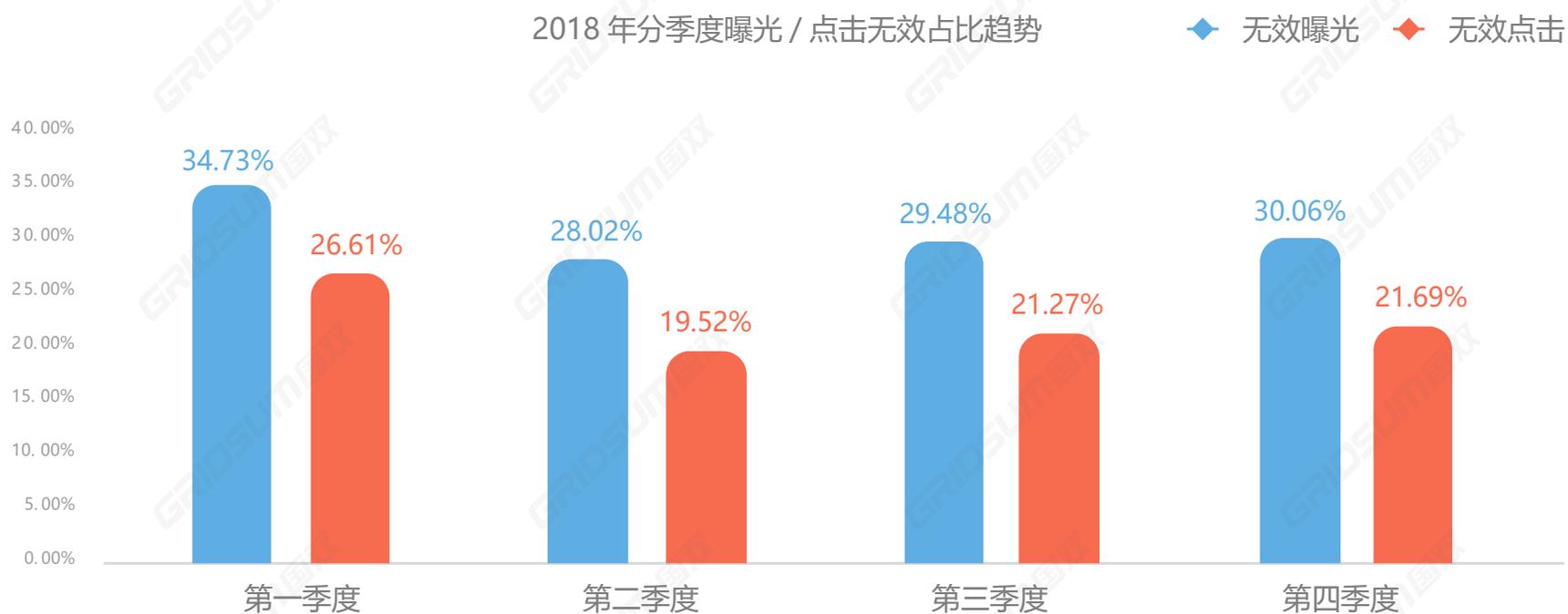
2018 年无效曝光有所下降，但占比仍然高于 30%，无效点击与前两年基本相差不大。



无效流量数据分析

2018 年分季度趋势分析：

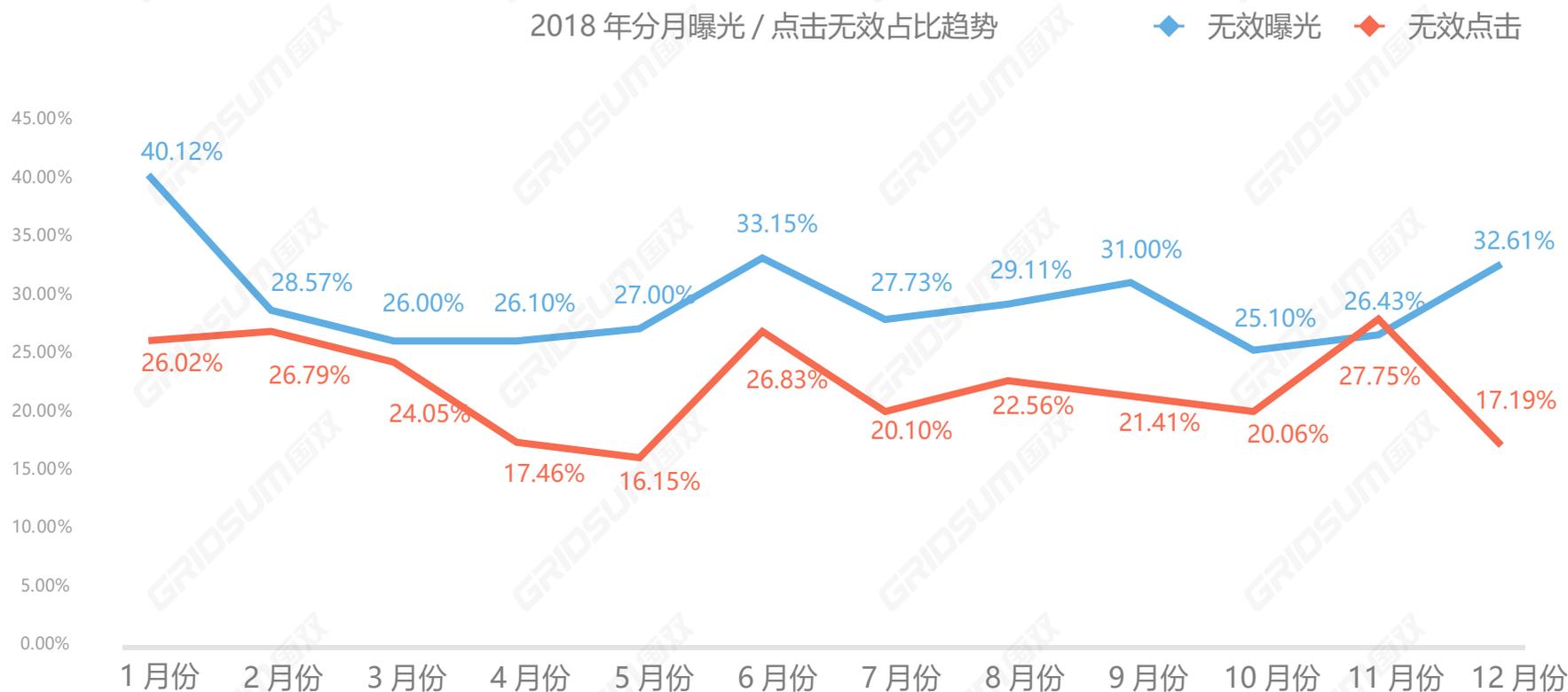
通过对数据的洞察我们发现无效曝光在第一季度较高，第二季度偏低，三、四季度趋于稳定。而无效点击方面，全年高峰同样出现在第一季度，达到了 26.61%。



无效流量数据分析

2018 年分月趋势分析：

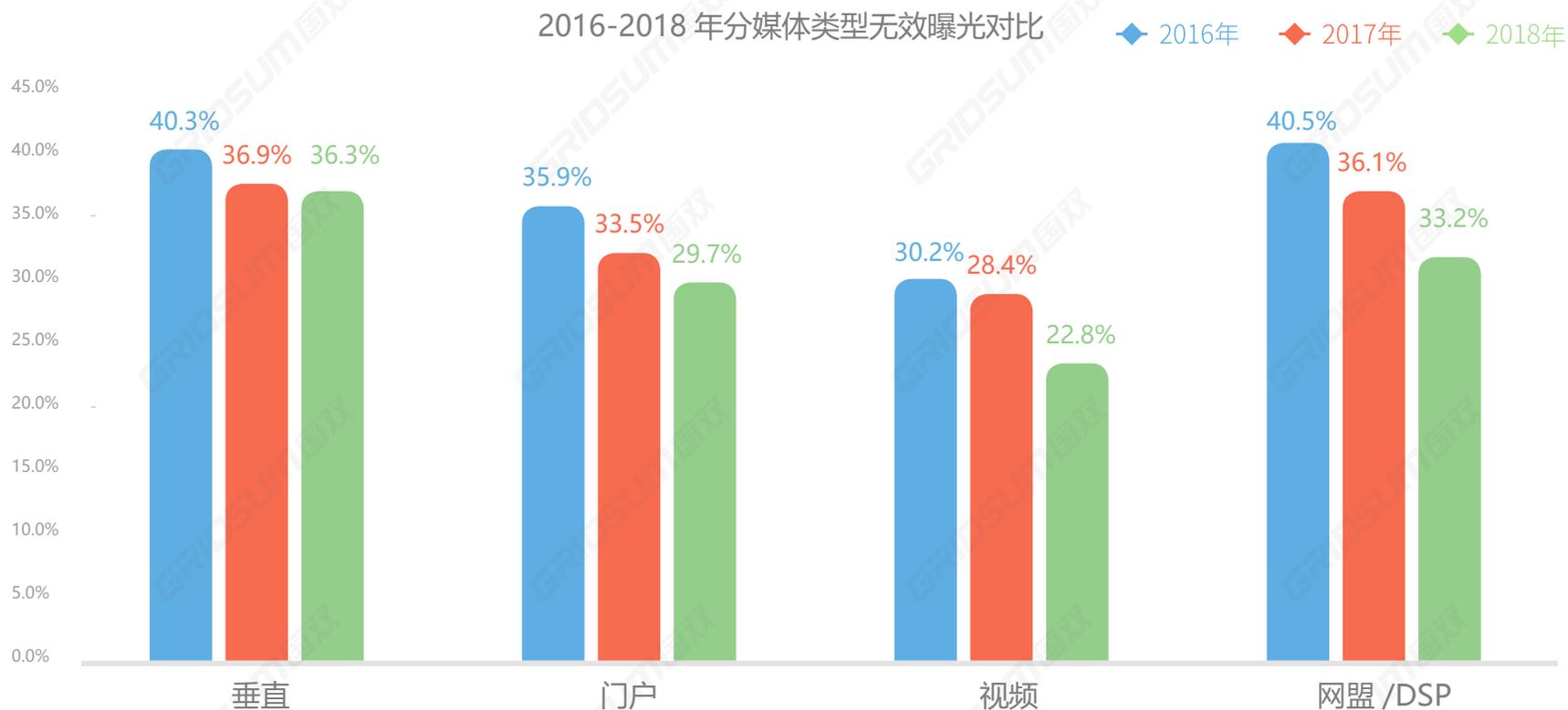
通过对数据的洞察我们发现 6 月和 11 月无效点击占比趋势呈现较为明显的波峰状态，而 6 月和 11 月分别有年中 618 和双 11 大型电商大促活动，各个广告主广告投放量都有所增加。



无效流量数据分析

分媒体类型——无效曝光：

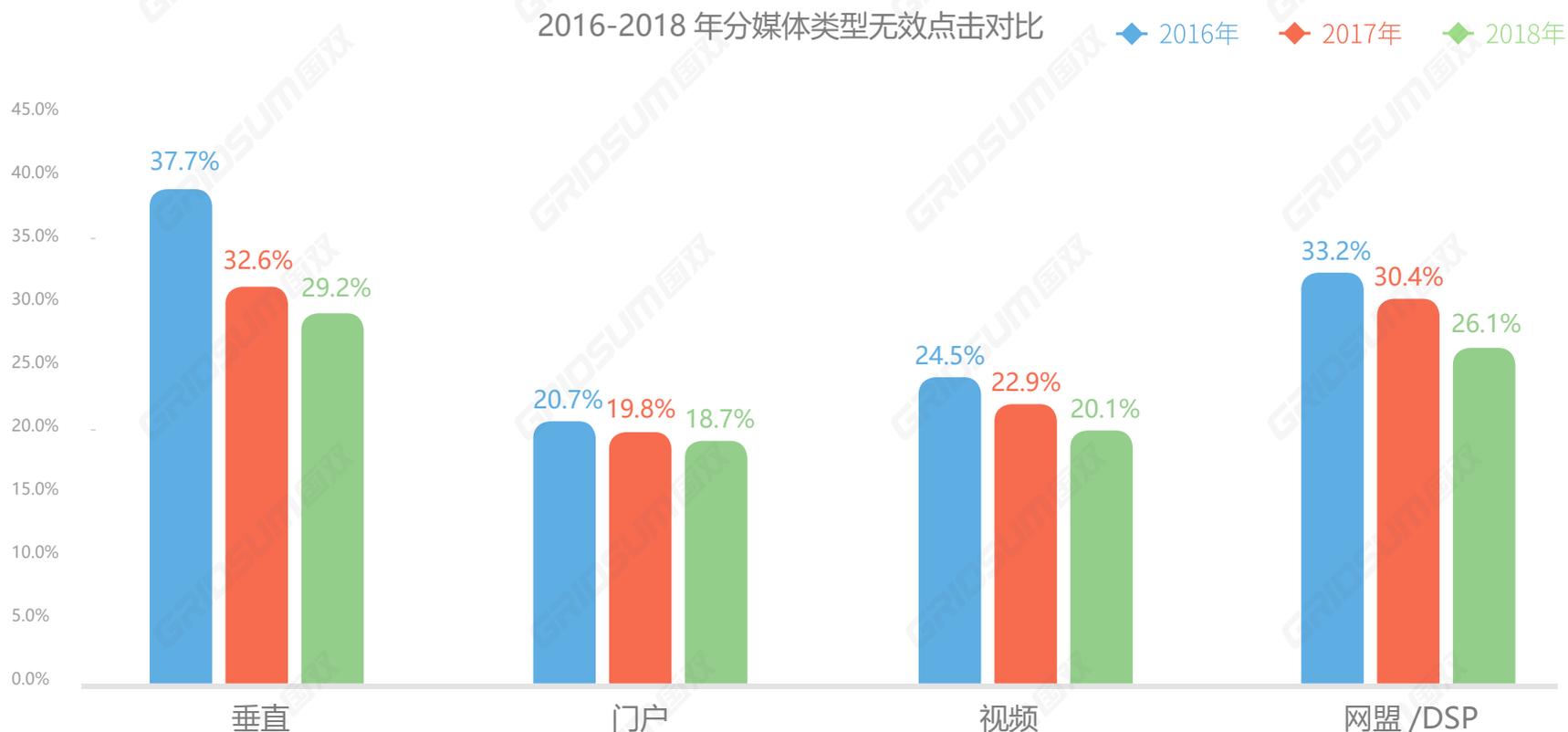
通过对数据的洞察我们重点分析了垂直、门户、视频以及网盟 /DSP 四个媒体类型，发现各媒体类型无效曝光占比排名维持不变，垂直类和网盟 /DSP 类的媒体无效曝光占比仍然是重灾区。另外，全部四类主要媒体类型都呈现出下降的趋势，视频广告曝光情况有较明显的改观，下降幅度超过 5%。



无效流量数据分析

分媒体类型——无效点击：

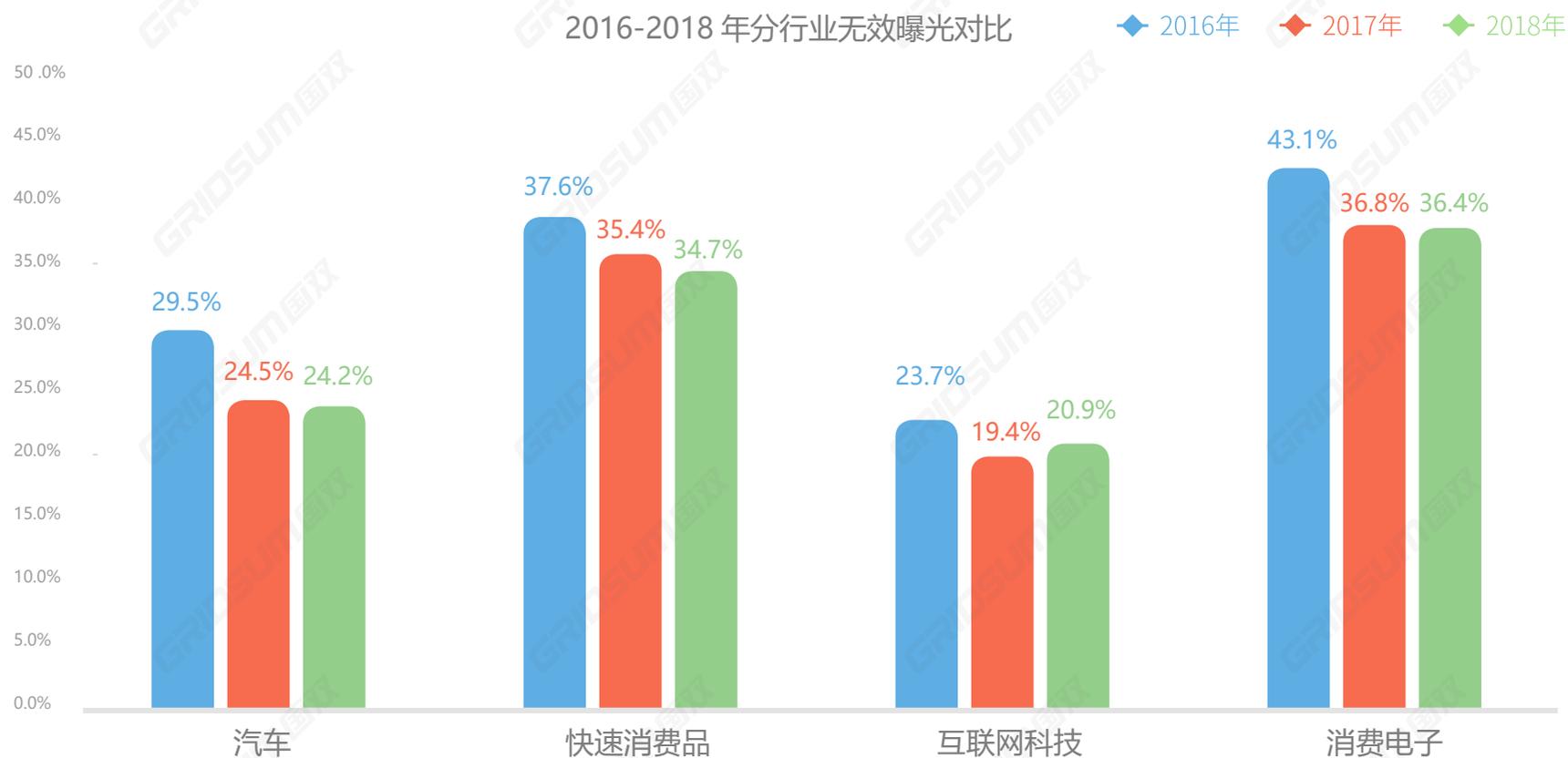
通过对数据的洞察我们重点分析了垂直、门户、视频以及网盟/DSP四个媒体类型，相较于2016年和2017年，2018年主要的4个媒体类型在无效点击方面，都有所下降。



无效流量数据分析

细分行业——无效曝光：

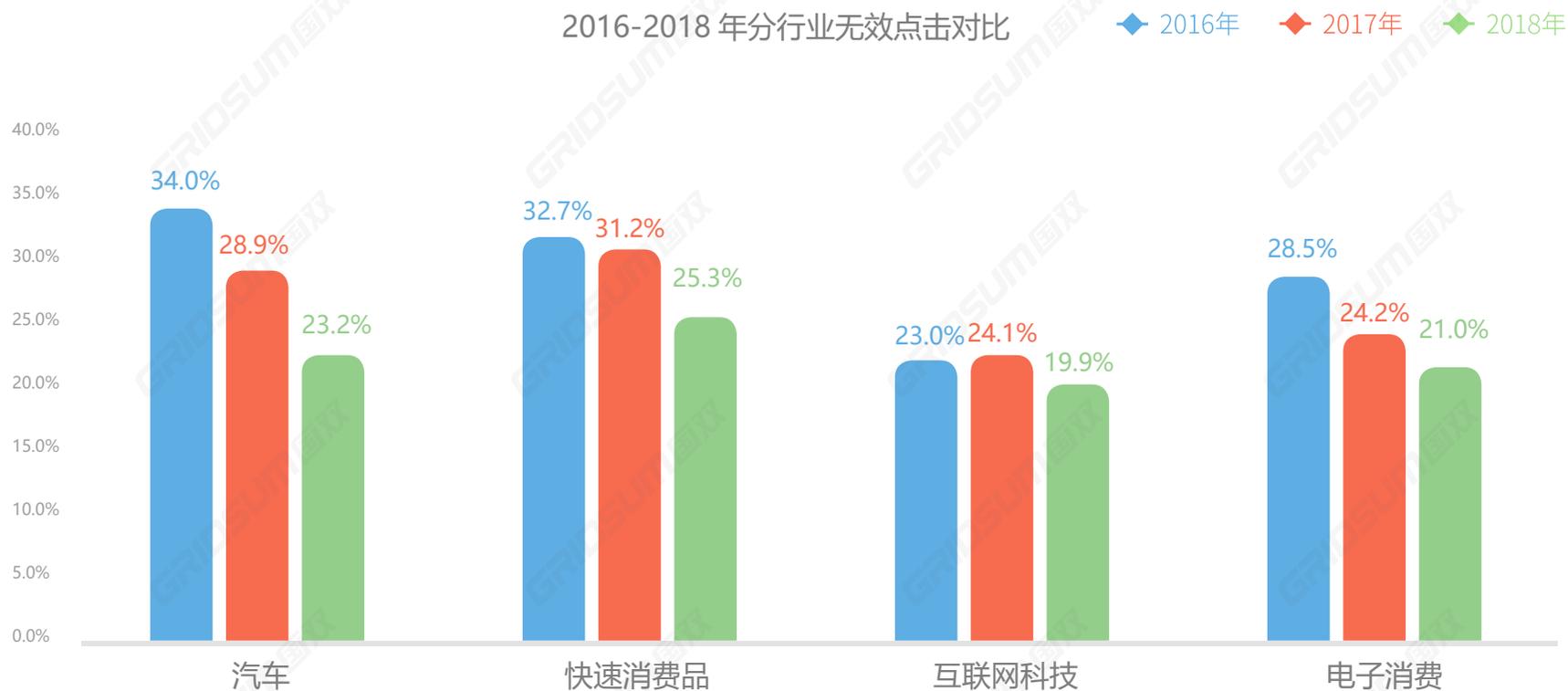
以汽车、快消、互联网科技、消费电子 4 个行业数据为例，通过对数据的洞察我们发现无效曝光情况整体有小幅下降趋势，互联网科技行业有小幅上升。



无效流量数据分析

细分行业——无效点击：

市场环境中汽车行业的无效占比约大于 30%。得益于国双在无效流量的研究及努力，国双客户的无效点击占比都有所降低，汽车行业降幅超过 5%。

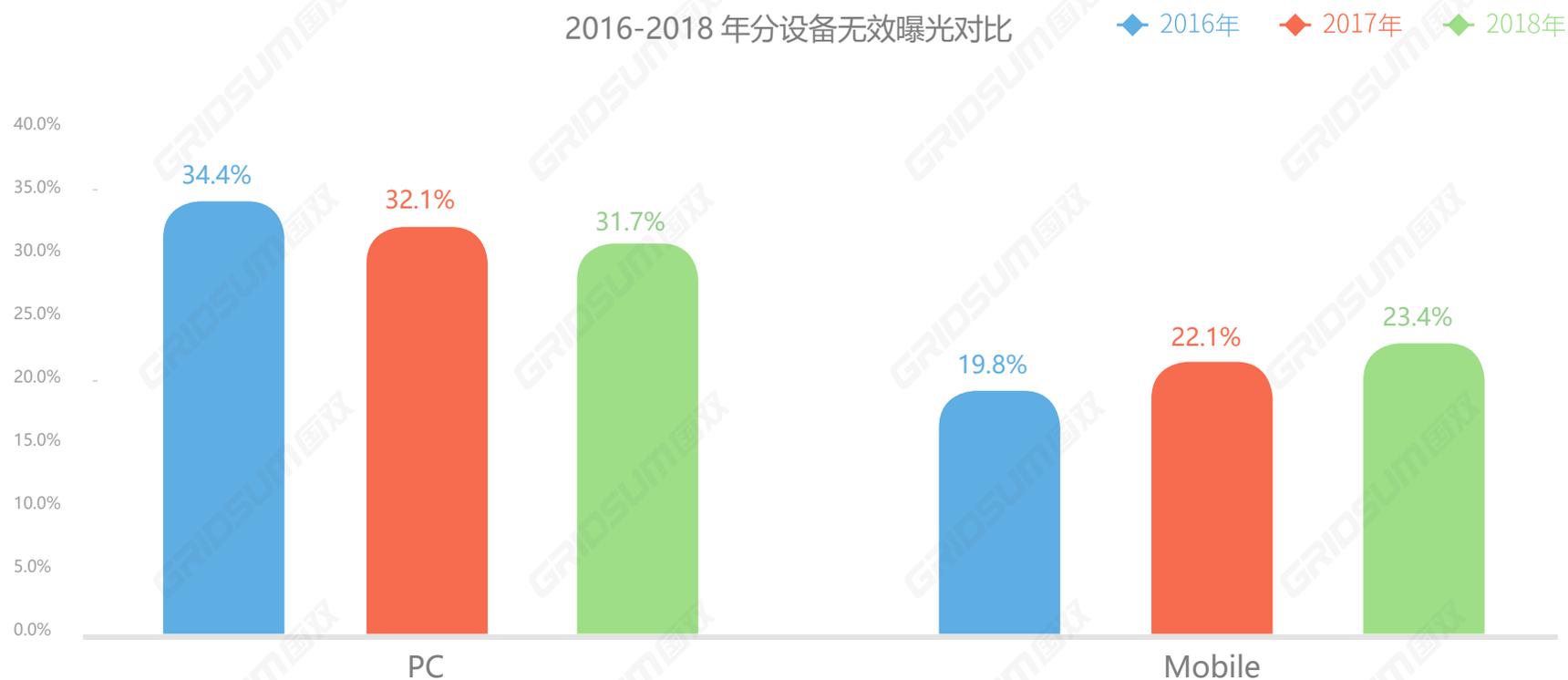


数据来源：2016年1月1日至2018年12月31日之间，国双监测的广告数据

无效流量数据分析

分设备——无效曝光：

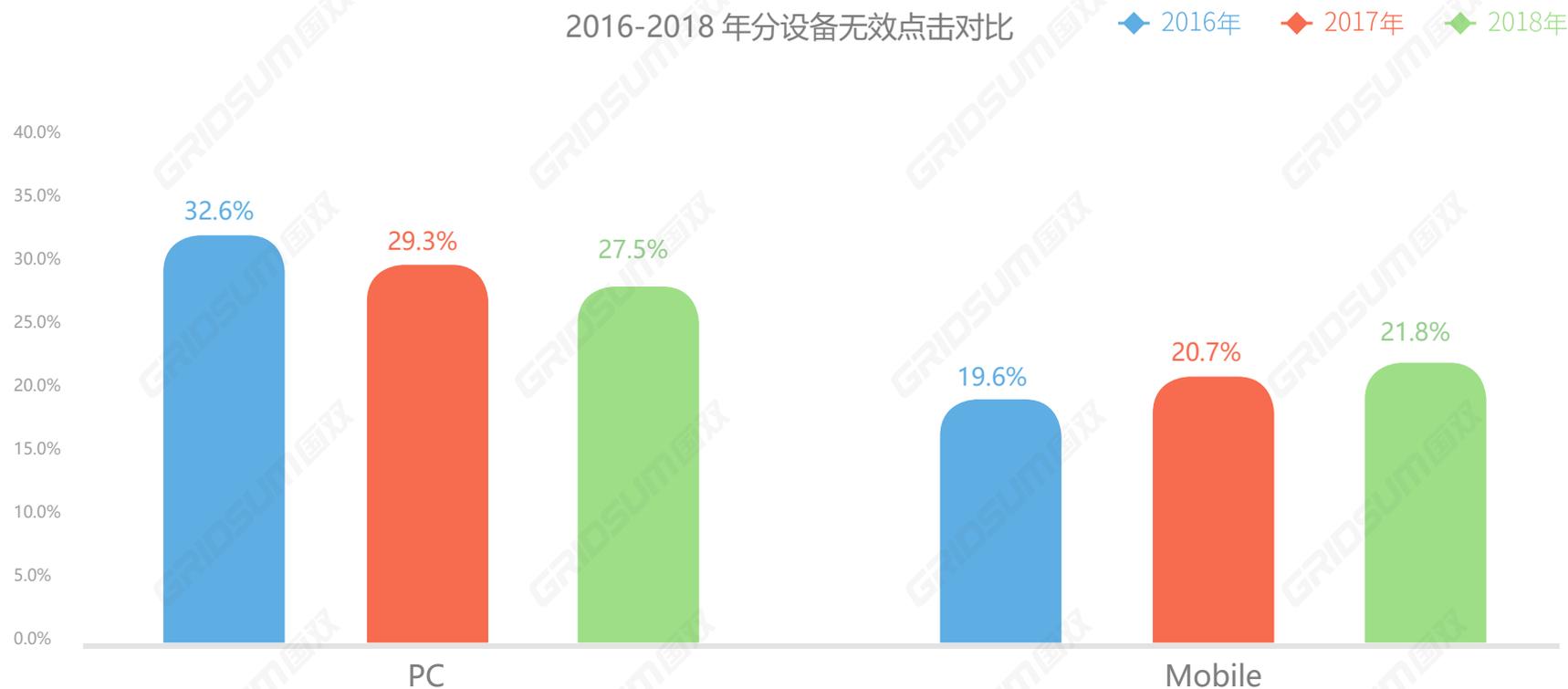
随着流量在不断向移动端迁移，移动端广告的问题也日益凸显，通过对 2016-2018 年的分设备流量进行分析后发现，虽然目前移动端的无效曝光占比较低，但有逐年上升的趋势，只是程度有一定减缓。



无效流量数据分析

分设备——无效点击：

与无效曝光类似，无效点击也呈现出了类似的趋势，PC 端占比虽高，但较上一年有小幅下降。而移动端却有了微弱的恶化。

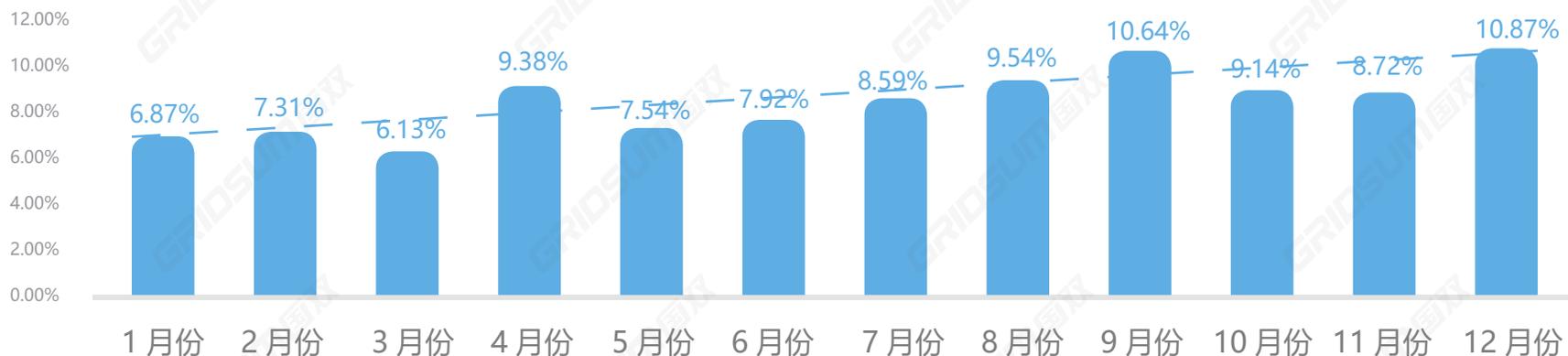


OTT 无效流量 (C2S)

随着 OTT 广告市场的成熟，中国首个针对 OTT 广告市场的监测标准——《MMA 中国无线营销联盟 OTT 广告监测标准》正式发布，这一标准是由 MMA 中国广告标准委员会委派，国双牵头联合 OTT 广告标准小组多家成员企业合作的结晶。市场上对于 OTT 的监测方式中，SDK 和 C2S (Client-to-Server) API 比 S2S (Server-to-Server) API 更为安全、也更容易监测流量异常情况。所以，以下我们仅针对 C2S 监测模式的流量进行无效流量的分析。

2018 年智能电视总体 C2S 的流量中无效流量平均占比为 8.65%，分月占比如下图，总体呈现小幅上升趋势。

2018 年智能电视平台无效流量状况

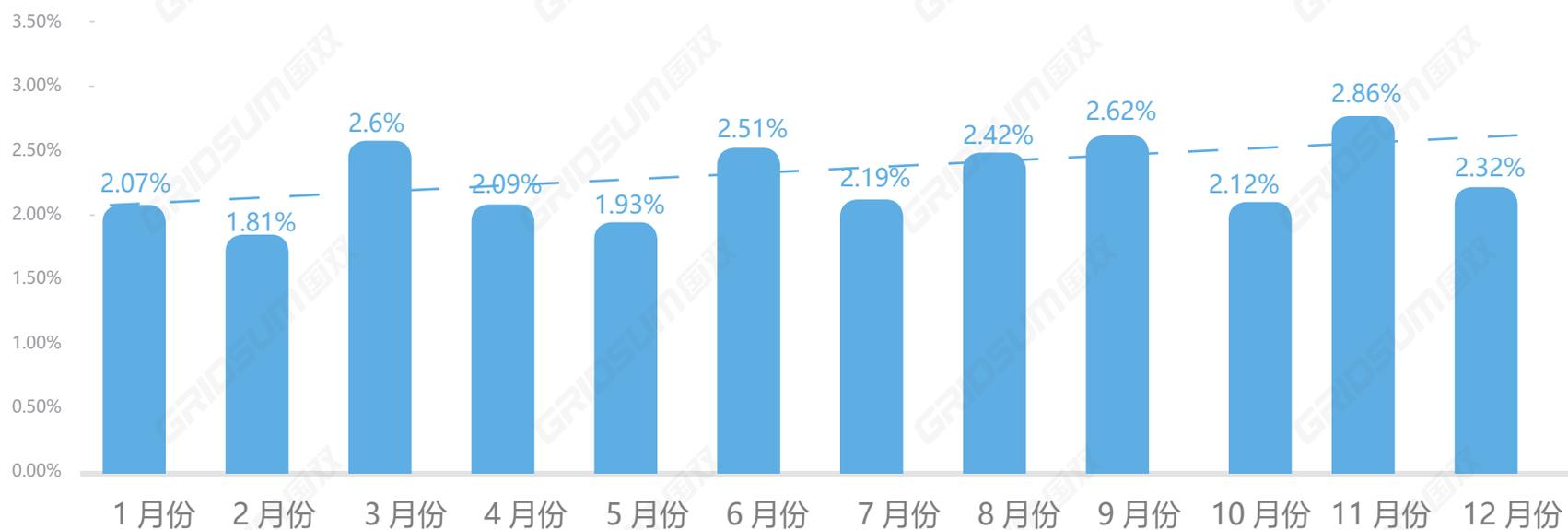


GIVT (常规无效流量)

GIVT :

2018年,国内数字广告GIVT(常规无效流量)平均占比为2.01%整体呈现上升趋势。

2018年智能电视平台无效流量状况



数据来源:2016年1月1日至2018年12月31日之间,国双监测的广告数据

GIVT (常规无效流量) 数据分析

GIVT :

2018年,国内数字广告GIVT(常规无效流量)分设备占比PC端占比最高2.35%,移动端1.99%,智能电视端0.61%。

常规无效流量分设备占比



GIVT（常规无效流量）数据分析

通过对数据的洞察发现，无效流量总体行业均值在 30% 左右，主要包含 GIVT 和 SIVT 两个部分，而 GIVT 的占比均值在 2% 左右，一方面是因为当前行业的黑名单量级还不足够，另一方面也是因为黑产各类作弊的手段越来越复杂越来越高明，简单常规的规则已经难以甄别出无效流量，我们需要更智能的技术手段有效判定出复杂无效流量。

03

国双的监测方案

如何取得信任？

为了能够更科学、更准确地评估互联网广告的投放效果，无效流量的识别和剔除是必不可少的，而这往往也是广告效果评估的第一步。

作为第三方验证公司，国双提供的技术方案和数据结果必须具有足够高的可信度，才能被广告主和媒体所认可。因此，国双很早就开始研究并参考欧美成熟市场的权威组织（主要有 MRC、IAB、MMA 和 TAG）对无效流量的定义和分类，来指导自己的无效流量识别。同时，国内市场环境也发展日趋成熟和稳定，国双也积极加入了中国广告协会和 MMA 中国等权威行业组织，并以核心会员的身份，推动相关无效流量标准在国内市场的落地。

国双的监测方案

经过多年努力，国双逐步研究并开发出了自己的一套无效流量监测体系，不仅包含广告数据的检验（如：黑名单、频次、定向投放异常等），还结合了网站数据的异常监测（如：热力图、多个维度构建的用户行为特征等），通过前后端广告数据打通，串联起广告流量的完整路径，有效识别异常情况。最后，我们还利用 AI 及机器学习，自动发现复杂的作弊特征，来甄别一般手段无法发现的无效流量。

作弊的方式在不断增多，作弊技术手段也一直在升级，但国双也在不断强化自己的监测体系，以达到对新型的异常场景，进行及时有效地识别和剔除；同时国双也在积极寻求外部合作，利用外部的庞大数据资源，更好地为客户服务。

国双广告反欺诈数据验证流程



原始监测日志

 **GIVT**

 **中国广告协会**
CHINA ADVERTISING ASSOCIATION

 **MMA**
CHINA 中国无线营销联盟

将监测到的 IP、Device ID 与 GIVT List 中的 IP、Device ID 进行对比，如果有匹配成功的，则说明该 IP 或 ID 为异常流量，需将其过滤掉

 **SIVT**

 **GRIDSUM 国双**
Empower your e-Performance

 **AdDissector**

对无效流量进行实时的 GIVT 粗过滤后，在每天结束的时候则会将历史数据及当天收到的信息进行综合分析判断，做国双 SIVT 复杂无效流量的过滤

 **客户定制**

客户

 **腾讯灯塔**
Beacon.tencent.com

 **GRIDSUM 国双**
Empower your e-Performance

客户定制属于自己的异常规则，由国双通过技术手段实现；与腾讯灯塔合作，和国双监测到的移动设备 ID 及行为路径进行 Mapping，实现精准分析，甄别异常流量



输出基于真实监测数据的广告投放分析报告

多重组合识别数据异常



- 90+ 维度异常判定技术
- 前后端打通分析每一次访问后的像素热力图、点击热力图，从中发现异常流量
- 监播实录技术有效规避视频类广告滥竽充数现象
- 实时的数据更新和一键式报告能够迅速发现问题
- 专业的代码监测工具及时发现人为错误
- 完善的历史数据库做参考

热力图强力识别验证

广告来源	访问量	综合浏览量	平均停留时间	跳出率	平均页面访问数	Sales leads 完成数	Sales leads 转化率
媒体 B	774,950	1,745,938	00:02:26	56.60%	2.25	1,741	0.22%

媒体 B 的各项基础数据及转化数据看似正常。但查看不同渠道在页面上的热力图鼠标点击分布，发现媒体 B 存在异常点击分布，符合程序点击发送数据包特征。

媒体 B 媒体 B 的实际效果需重新评估



媒体 A



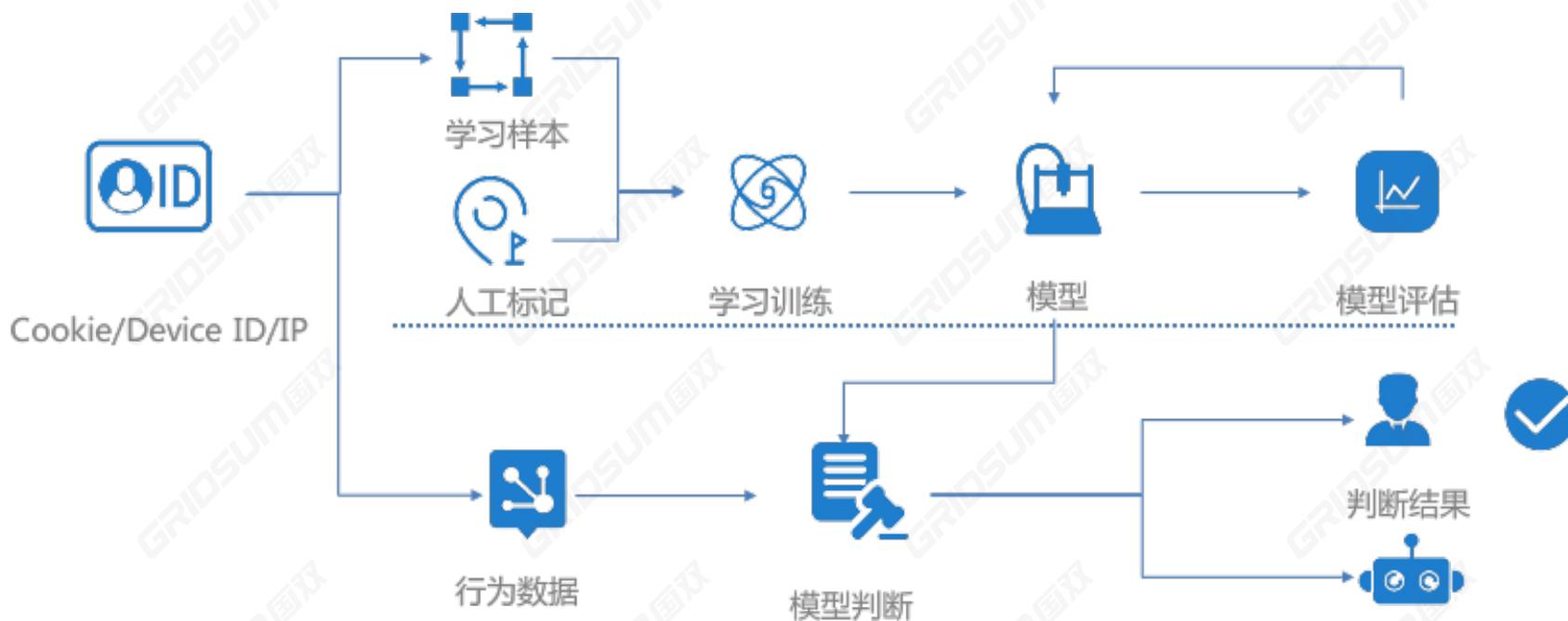
数据来源：2016年1月1日至2018年12月31日之间，国双监测的广告数据

异常流量形式及排查原则

更多异常流量判断方法	异常分析排查部分原则子
<ul style="list-style-type: none">• 连续曝光• Stable 人群占比• 时差异常• 周期性波动异常• 定向人群异常• 稳定率异常• 频次分布异常• 曝光不可见•	<h3>前端</h3> <ul style="list-style-type: none">▶ 曝光或点击的频次（ID/IP）超过设定的阈值▶ 在较长的时间段内，曝光或者点击数据呈现出剧烈的、周期性的上下波动▶ 曝光时间分布不符合网民活跃规律▶ 地域分布非定向，但与行业均值、历史分布相差非常大，表现为某一城市数据非常突出。▶ 同一广告曝光 / 点击来源于不同媒体，或者是相同媒体的不同页面上，一般是媒体在用不同页面的流量增加数据量数据非常突出。
	<h3>后端</h3> <ul style="list-style-type: none">✓ 热力图（点击热力图、像素热力图）✓ 停留时间分布、跳出率、二跳率、硬件厂商、机型、操作系统等维度指标✓ 行为路径相同、销售线索重复、用户行为特征✓

基于 AI 自动发现复杂作弊特征

基于固定规则的方法基本能够帮助我们过滤掉 60%-70% 的无效流量，剩下的部分是由于刷量技术不断升级绕开了规则。对于这些可以绕开规则的无效流量，国双采取基于机器学习的方式来反制。通过应用 AI 技术去自动发现作弊规律，不断学习改进模型，帮助我们识别那些不太容易被甄别出来的作弊。



04

总结

总结

通过以上的洞察分析不难发现，整个互联网广告市场的无效流量占比依然很高，与无效流量的斗争也依然艰巨。当前行业内数据割裂、数据权威性方面的问题依然存在，而随着作弊手段的不断升级，需要行业内各个参与者联合起来，共同应对。

一直以来国双都在中国广告协会的指导下积极的与行业内权威的组织和企业合作，同时遵循中国广告协会（CAA）、美国媒体评估委员会（MRC）、中国媒体评估委员会（CMAC）以及中国无线营销联盟（MMA China）等公认的行业标准，以公平、公正的原则切实维护广告主的合法权益。同时也呼吁行业内更多的参与者加入到抵制无效流量的队伍中来，不断净化互联网广告环境，推动广告数据的透明性，构建更为健康的数字营销生态。

关于国双

国双 (Nasdaq : GSUM) 是中国领先的企业级大数据和人工智能解决方案提供商。基于国双大数据平台独有的分布式数据架构和先进的实时、多维度关联性分析技术，国双的解决方案能够使客户充分洞悉数据间的复杂关系，获得全新的商业洞察，以利企业和政府客户作出更好的业务决策。“国双 (Gridsum) ” 其名，是分布式计算 (Grid) 与分析 (Sum) 的结合。作为数字智能化的先行者，国双致力于帮助企业和政府客户以新颖有效的方式使用数据，提高生产力。

更多详情，请访问官方网站：www.gridsum.com

欢迎关注国双官方微博 (@ 国双 Gridsum)、微信公众号 (ID : gridsumtech)。



国双官微



国双营销研究院