



用户分析 埋点最佳实践

白皮书





目录

• 为什么要埋点?	1
• 埋点面临的挑战	1
• 全埋点	2
• 可视化埋点	4
• 代码埋点: 最可控的埋点方式	5
• 标签管理器 (TAG MANAGER) : 低调的高手	6
• 趋势	8

为什么要埋点？

在愈发重视用户分析的时代，准确获取用户数据变得尤为重要。用户数据的获取一方面来自于用户主动提供，例如：通过调查来获取用户对某项服务的满意度，但这些主观调查通常参与度低，可靠度也有限。另一方面，仰仗对用户行为、以及行为产生的结果的监测和分析，则很好得弥补了这些不足。所以主观问询和客观监测就构成了用户数据获取的两个主要来源。其中客观监测就是通过埋点来实现的。

埋点面临的挑战

和日志分析不同，受限用户行为的不确定性，监测环境技术实现的不规范性，业务类型的多样性，用户数据的获取尚无法全自动来完成。这些不一致性构成了埋点的第一个挑战。同时，埋点往往具备一定的技术专业性，提出埋点需求的业务方往往是产品、运营、市场、销售甚至财务等不具备代码技术能力的部门，跨部门协作就构成了埋点的第二个挑战。另外，埋点的正确性需要在使用中来验证，测试环节想对埋点——验证无疑对产品发布上线过程增加了额外的工作量，而一旦有错误没有在测试环节被发现，想纠正就要重新进行一次产品发布，对于网页应用来说，即使这个动作很快，这期间的数据错误也难以纠正了。而对于移动应用来说，纠正的过程还需经历漫长的等待。验证难构成了埋点的第三个挑战。最后，埋点并不是一个一劳永逸的过程，而是需要根据业务的节奏，周期性、多次甚至频繁进行的。在业务快速变更中会出现大量新的埋点需求，也会让一些过往的埋点失去价值。埋点过程的重复性带来的管理问题构成了埋点的第四个挑战。

综合上述四个挑战，更易操作，正确率更高的埋点方式无疑会大大提高用户数据获取的能力，会让业务方和开发者能更加专注于用户分析，而不是数据获取上。

全埋点

全埋点，或者也被称为“无埋点”、“无痕埋点”以及“自动埋点”。是开发者对全自动的埋点方式的探索，希望能像日志分析一样将埋点过程完全自动化，但这条路目前是行不通的。

客户端行为一般分为三种层级：产品打开，页面访问，页内行为。其中产品打开和页面访问的标准化程度很高，我们能在捕捉到这个行为的同时知道这个行为相关联的几乎所有属性。我们就能了解这个动作的业务含义，所以我们会将这两类行为自动采集。

而对于页内行为，若想通过全埋点接近代码埋点的能力则需要至少解决以下问题：

1. 采集点击行为时，也将页面上的某些内容自动采集上来，以便后期作为属性关联。例如：我们在采集购买行为时，会将购买的商品名称、价格等相关属性一并上报。全埋点可以自动将点击动作采集上来，甚至也可以知道这个动作的名称是购买，但如果想将商品名称及价格上报到购买这个行为中，则需要进行至少一次手动关联。
2. 在页面功能改变后，为了能和过往版本的属性关联，需要能根据版本或者时间点来选择当时的属性。
3. 为了能在各种非标准化页面中采集到数据，用户在使用的时候还得能够精准地选择自己想查看的事件，很可能这个过程要比通过代码埋点需要更长的耗时，而且需要具备一定的页面开发知识，但关键的问题是没有办法能确定最终得到的数据是否准确。
4. 同样因为客户端的页面多样性实在太多，实现差异性太大，想确保将所有有效行为采集上报是难以保证的，所以全埋点无法得到准确的数据。



5. 多投入数十倍的数据存储成本，而这些数据中的绝大多数都会有未清洗的状态不会被用到。

由于页内行为远远不仅仅鼠标点击一种，还有滚动条滚动停留、鼠标框选、键盘输入等等。全埋点自动采集连点击都无法妥善实现，要是真能做到『全』甚至『无』目前来看是绝无可能的。

全埋点唯一的价值大约就是可回溯性了，毕竟一部分数据采集回来了。只不过采集回来的大部分数据是还需要进一步处理甚至压根就不会用到的罢了。

在我们看来全埋点不但没有为埋点提效，而且引入了更多的复杂度，最关键的是数据准确性根本无法保证。

这里总结全埋点优缺点如下：

优势	不足
可回溯	数据准确度难以保障
	分析前仍需定义事件，且难度大
	属性难以关联
	能自动采集的行为有限
	大量消耗存储资源
	难以对老版本事件定义

可视化埋点

数据分析，数准是前提。可视化埋点，并不像全埋点，一味追求简单而牺牲了数据的准确性。而是延续了代码埋点思路，只是通过所见即所得的方式让代码埋点变得人人都可以操作。

可视化埋点的部署原理

支持可视化埋点的 SDK 会在被监测的网站或移动应用被访问时向服务器校验是否有新的埋点，如果发现更新的埋点，则会从服务器下载并且立即生效。这样就能确保服务器收到最新的埋点后，所有客户端都能在下一次访问时得到部署了。

可视化埋点和全埋点有着对埋点和分析全然不同的追求。可视化埋点的理念是提升原工作流程的效率——依然要梳理需求、设计埋点；全埋点则是将工作流都进行了简化——反正数据会被采集回来，这两步的必要性就容易被忽视。这里不能说孰优孰劣，因为事先严谨的计划和事后发散的探索都是分析中的不同角度。况且这两种埋点也完全不是排他的，完全可以同时使用。

目前，可视化埋点也有其局限性。

首先，可视化埋点也只是针对可见元素的，其中可见元素最常见的就是点击行为了。对于点击操作的埋点也确实是目前可视化埋点的主攻点。但从实际情况看，复杂页面、不标准页面、动态页面都给可视化埋点增加不可用的风险，一旦遇到就还是只能代码埋点了。

其次，对于点击操作附带的业务属性，虽然也可通过进一步选取属性所在元素来获取属性信息，但国内厂商支持得好的就比较少了。

第三，为了确保埋点准确性，可视化埋点也逐步整合了更为复杂的高级设置，例如：“同页面”、“同版本”、“同层级”、“同文本”……，加上了这些复杂设置的可视化埋点还是那个为提效而生的可视化埋点吗？

代码埋点

代码埋点是最经典的帮助工程师了解用户是如何使用产品的埋点方式。因为是工程师人工将埋点结合到代码逻辑中，理论上只要是客户端种的操作，再复杂也能采集到。常见的如：页面停留时间，页面浏览深度，视频播放时长，用户鼠标轨迹，表单项停留及终止等等。尤其是一些非点击的、不可视的行为，是非要代码埋点来实现不可了。所以如果我们需要对埋点有更加精准的控制力，那么代码埋点是最好的选择。

也许你还分不清集成和埋点。为了进行埋点，厂商通常都提供一个代码包，可以理解为一个工具包，里面包含常用的工具。想埋点就要先有这个工具包，也就是集成 SDK。然后根据里面的说明书，再使用这个工具包制作出各种东西，也就是埋点了。

当然弊端也是很明显的，前文描述的那些苦恼几乎全是代码埋点相关的。为了能让埋点过程更高效，厂商们做了很多努力。

优势	不足
精准	需研发介入
几乎能解决所有客户端行为采集问题	验证难、纠正难
	无效埋点难以管理

标签管理器 (Tag manager)

大家可能对标签比较陌生，但用于采集网页数据的 SDK 大家已经不陌生，这些嵌入到网页中，能采集网页上、移动应用或者视频中的数据的，就是监测类的标签。但标签的用途远不止于此，通过在网站中嵌入代码，工程师可以对网站提供很多额外的能力。除了刚刚提到的数据监测，还可能为网站提供一些额外的功能，最常见的就是推送个性化的内容，例如：A/B 测试，消息推送，个性化广告等等。

假如网站或者移动应用借助标签的能力实现很多功能，那么就需要用到很多标签，而且标签可能也需要频繁更新或改动。同样网页还好，上线很容易，但移动应用可就难了，假如再出现了错漏，改正就要面临非常长的改正周期。这种情况下，标签管理器就派上了用场。

标签管理器提供了一个容器，工程师只需要在网页或移动应用中正确嵌入这个容器，之后不懂技术的团队也能通过在线管理的方式将后续各种标签发布到网页或移动应用中。这样就实现了技术人员和业务人员工作的各自为战。听起来是不是跟可视化埋点很像？是的，他们的原理是几乎一模一样的。只不过可视化埋点更倾向于针对客户端的用户点击行为提供了直观的方法，而标签管理器是代码层面的，能做的事情会更多一些。

标签管理器非常强大的地方在于能免去代码埋点而通过 DataLayer 就能获取到页面中的变量，如每个用户不同的用户 ID、用户等级、登录状态、购买的产品的名称以及价格等；而通过触发器能在这些变量符合一定的时才触发事件的上报。是不是非常厉害！

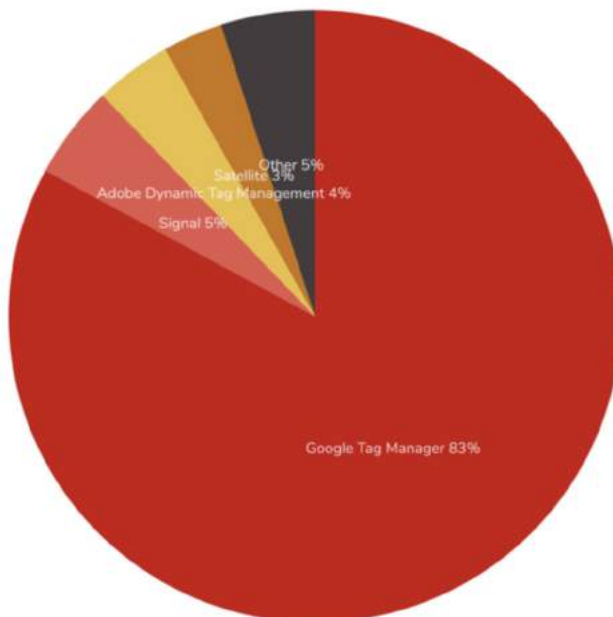
目前最著名的标签管理器是谷歌推出 Google Tag manager，简称 GTM，占据了 83% 的份额。个人版是免费的，但依然提供了极其强大的功能，一般团队用都足够了。想进一步地了解 GTM 的功能，可以阅读它的官网，里面有非常丰富的讲解和案例。



目前最著名的标签管理器是谷歌推出 Google Tag manager ,简称 GTM ,占据了 83% 的份额。个人版是免费的，但依然提供了极其强大的功能，一般团队用都足够了。想进一步地了解 GTM 的功能，可以阅读它的官网，里面有非常丰富的讲解和案例。

Tag Management Usage Distribution in the Top 1 Million Sites

Statistics for websites using Tag Management technologies



趋势

目前客户端中对用户数据的获取通过代码埋点和可视化埋点是更好的组合。未来，整合了标签管理器的某些特性的可视化埋点一定能更多地替代代码埋点，解决工作中常见的所有客户端行为采集需求。

就像早期论坛的编辑框，只能通过发布或者预览功能才能看到帖子的效果，但后来所见即所得的编辑器出现使得文字的编辑变得非常高效和愉悦。目前开源社区流行的 Markdown 格式依然沿用了这种方式，在诸多流行的 Markdown 编辑器中，依然是一侧编辑、一侧实时预览，或直接就以最终格式的方式来编辑。

随着 IoT 时代的带来，越来越多的用户界面会出现在电脑和手机之外，越来越多的内容是因人而异的。届时，未来越来越多的 SDK 集成后会自动采集更多标准的用户行为，而对于非标准以及业务含义强的，需要计算的，或者需要按照特定条件生效的埋点，则可以交给可视化埋点来完成。



开发者交流群
扫描二维码

精益成长

网址: www.analysys.cn

客户热线: 4006-010-231

微博: @Analysys 易观

Analysys 易观 易观出品
你要的数据分析