

神策分析 – 事件设计思路

文章主要内容

一、事件模型的由来

二、理解事件模型

三、事件梳理

四、事件梳理中应规避的问题

一、事件模型的由来

传统 Web 时代，通常使用 PageView 来衡量和分析一个网站的好坏，然而到了移动互联网时代，PV 已远远不能满足业务人员的分析需求。移动 APP 时代，衡量产品是否成功的指标已愈加复杂，不是简单的 PV，而是发帖量、视频播放量、交互点击量、订单量等等，这些都已远远超出 PV 所能衡量的。此外，对于更加深入的多维分析需求，PV 模型亦无法胜任。目前市面上的统计工具，如百度统计、谷歌分析等，均通过标准的 SDK 来采集客户端页面数据，即传统的 PV&Session 模型。

下面具体列举 PV&Session 为核心的数据模型不足：

- 不够全

行为数据不够全。PV 模型仅能采集线上页面浏览行为，而实际完整的用户行为却还包括注册、视频播放、Tab 切换、点击购买等丰富的交互行为，另外对于线下产生的用户行为，如拨打客服、系统退款等，就更是无能为力了。

- 不够细

维度数据不够细。任何想有洞察的分析，都依赖于丰富的维度数据，而传统的 PV 模型仅能采集如设备、终端、浏览器等简单维度，无法进行更多维度的下钻，来满足更加复杂的业务需求。如想对电商订单进行分析，就需要从订单金额、商品属性、优惠券使用、支付方式等角度去看，而这些订单特有属性，PV 模型无法获取。去进行分析，哪些和订单相关的特有属性，在传统的数据模型是获取不到的。

- 不够准

PV 模型依赖于前端 JavaScript SDK 或 APP SDK 进行数据采集，因此对于重要的业务数据，如订单数、交易金额等，受限于数据传输丢失，无法拿到精准数据。另外，PV 模型一般只使用 Cookie、设备 ID 等匿名 ID 来标记用户，这对于移动时代频繁的跨终端、跨产品、跨渠道的行为特征，无法进行精准的用户识别和打通。

总结：

由于 PV 模型在数据采集上的诸多不足，所以神策分析使用事件模型作为数据模型。事件模型在数据采集上支持更全面的数据源，更全的行为数据，更灵活的用户定义，更丰富的维度属性，进而满足使用者更加复杂的分析需求。

二、理解事件模型

1、事件模型

事件模型包含事件 (Event) 和用户 (User) 两个核心实体，即只有事件表 和用户表两张表。

2、Events 表、Users 表数据定义

2.1、Event 表

事件模型包含事件 (Event) 和用户 (User) 两个核心实体，即只有事件表 和用户表两张表。

Event 表记录用户的一系列行为，描述一个用户在某个时间、某个地点、以某种形式完成某个具体事情。即包含以下五个要素：

1. Who: 即事件的主体。对于未登录用户，可以是 Cookie、设备 ID 等匿名 ID；对于登录用户，可以使用后台分配的实际用户 ID。
2. When: 即事件发生时间。可以记录精确到毫秒的事件发生时间。
3. Where: 即事件发生地点。可以根据 IP 来解析相应省份和城市，也可以根据 GPS 定位等方式来获取地理位置信息。
4. How: 用户从事事件的方式。包括用户使用的设备、浏览器、APP 版本、操作系统等。
5. What: 描述事件具体内容。比如用户支付订单事件，即可以传输订单相关的如金额、商品属性、支付方式、优惠券等属性来描述该事件。

时间	用户	事件	渠道	搜索词	支付价格	订单ID
2017/4/11	小李	注册	百度			
2017/4/12	小李	登录				
2017/4/13	小李	搜索		苹果手机		
2017/4/13	小李	支付订单			3000	123441

2.2、Users 表

每个 User 对应一个用户实体，User 表可以通过用户 ID 与 Event 表进行关联，来明确什么样的用户在产品中产生了什么行为。User Profile 记录的是用户实体基本固定不变的属性，例如姓名、性别、出生年份、注册时间、注册渠道等，对于一些变化的用户属性（如：VIP 等级），我们也支持更新用户表的接口。

id	first_id	second_id	\$city	firstVisitOrigin	phone	\$province	firstVisitTime	\$id
-9223295567774550931	470b806d23f3b10e		河池市	baidu	13104053254	广西壮族自治区	20150322	-9223295
-9223020481274100324	67627271fc2a50ad		安庆市	sina	13935324145	安徽省	20160303	-9223020
-9222837296461035956	446f05587b4408ab		吉林市	Yahoo	13841842841	吉林省	20150322	-9222837
-9222834969425138593	672941969c095b30		长沙市	Yahoo	13841842841	湖南省	20160402	-9222834
-9222814415821140482	723c315041855269		沧州市	sina	13104053254	河北省	20160402	-9222814
-9222808940389790720	5b173e0ec1b7d86e		松原市	Yahoo	18631221853	吉林省	20160402	-9222808
-9222664470325185212	31bd04cb31b3dfd4		衡水市	baidu	13104053254	河北省	20160303	-9222664
-9222626604292912664	26682ecf0a9ad744		白银市	baidu	18631221853	甘肃省	20160303	-9222626

3、模型优势

- 事件模型减轻了 ETL 的压力
- 事件模型支持更灵活的数据采集方式，不局限于前端采集，还可以进行后端 采集
- 事件模型可以采集更多的用户行为数据
- 事件模型可以采集更多的自定义维度
- 事件模型可以跨屏识别、打通用户
- 事件模型对数据采集更为准确

三、事件梳理

1、Events 表梳理

1.1、明确业务流程

我们需要理解产品的业务流程，来确定需要设计的行为事件。对于任何产品来说，我们推荐利用增长黑客模型 AARRR 来梳理具体业务，即分为用户获取、用户活跃、用户留存、收入和传播推荐五个阶段。

以电商 APP 为例，用户首次启动可视为用户获取；注册、查看商品详情、参加活动等可视为用户活跃；用户不间断回访可视为用户留存；当用户完成购买支付，即实现了收入目标；用户在产品中对商品的分享、推荐，可视为传播。

1.2、根据业务流程梳理事件

确定业务流程后，就能明确用户在产品中会产生哪些行为。然后依据需求，把用户在产品中的行为，抽象成事件。下面我们以一个电商 APP 产品为场景举几个例子：

用户获取：

- 对于用户激活 APP 这样的行为：抽象成“激活 APP”事件

用户活跃：

- 用户在 APP 中发生了注册行为：抽象为“注册”事件
- 用户查看了商品的详情：抽象为“查看商品详情页”事件

收入方面：

- 用户在 APP 中成功提交了订单：抽象为“提交订单”事件
- 用户在 APP 中成功支付了订单：抽象为“支付订单”事件

用户传播：

- 用户在产品中的分享行为：可以抽象为“分享”事件

1.3、根据事件梳理相关维度

事件维度需要根据相应事件应有的特性及后续分析需求进行梳理，用以描述该事件。

例如以下事件应会有维度：

事件	属性		
“启动 APP” 和 “注册”	事件名称	事件属性	属性值格式
	启动APP	\$预置属性	字符串
		是否首日启动	BOOL
		是否首次启动	BOOL
	注册	\$预置属性	字符串
注册方式		字符串	
查看商品详情页	事件名称	事件属性	属性值格式
	点击查看商品详情	\$预置属性	字符串
		商品ID	字符串
		商品名称	字符串
		商品一级分类	字符串
		商品二级分类	字符串
		商品原价	数值
		商品现价	数值
参加活动	事件名称	事件属性	属性值格式
	参加活动	\$预置属性	字符串
		活动ID	字符串
		活动类型	字符串
		活动名称	字符串
活动持续时长		数值	
提交订单	事件名称	事件属性	属性值格式
	提交订单	\$预置属性	字符串
		订单ID	字符串
		订单金额	数值
		运费	数值
		优惠券名称	字符串
		优惠券金额	数值
		商品ID	字符串
		商品名称	字符串
		商品数量	数值
		商品总价	数值
		商品一级分类	字符串
商品二级分类		字符串	
成功支付订单	事件名称	事件属性	属性值格式
	成功支付订单	\$预置属性	字符串
		订单ID	字符串
		订单金额	数值
		运费	数值
		优惠券名称	字符串
		优惠券金额	数值
		商品ID	字符串
		商品名称	字符串
		商品数量	数值
		商品总价	数值
		商品一级分类	字符串
		商品二级分类	字符串
支付方式		字符串	
分享	事件名称	事件属性	属性值格式
	分享	\$预置属性	字符串
		商品ID	字符串
		商品名称	字符串
		商品类别	字符串
		商品价格	数值
		分享方式	字符串

每一个事件还会有预置属性：

ps：后端采集的事件如果想拥有前端预置属性，需要把预置属性数据传到后端做补充，再传到神策系统中

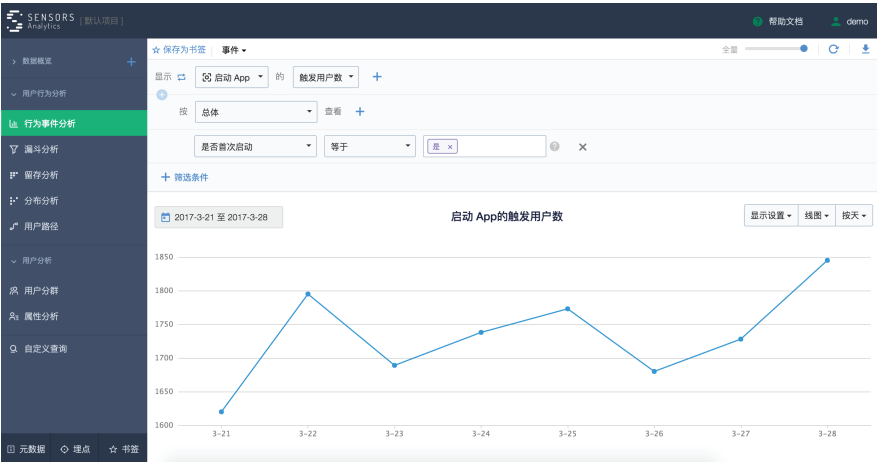
字段名称	类型	说明	JS SDK自动采集	iOS自动采集	Android自动采集	服务端SDK自动采集
\$app_version	字符串	应用的版本	N	Y	Y	N
\$country	字符串	国家	Y	Y	Y	N
\$city	字符串	城市	Y	Y	Y	N
\$province	字符串	省份	Y	Y	Y	N
\$lib	字符串	SDK类型，例如python、iOS等	Y	Y	Y	Y
\$lib_version	字符串	SDK版本	Y	Y	Y	Y
\$manufacturer	字符串	设备制造商，例如Apple	N	Y	Y	N
\$model	字符串	设备型号，例如iphone6	Y	Y	Y	N
\$os	字符串	操作系统，例如iOS	Y	Y	Y	N
\$os_version	字符串	操作系统版本，例如8.1.1	Y	Y	Y	N
\$screen_height	数值	屏幕高度，例如1920	Y	Y	Y	N
\$screen_width	数值	屏幕宽度，例如1080	Y	Y	Y	N
\$wifi	BOOL	是否使用wifi，例如true	N	Y	Y	N
\$browser	字符串	浏览器名，例如Chrome	Y	N	N	N
\$browser_version	字符串	浏览器版本，例如Chrome 45	Y	N	N	N
\$carrier	字符串	运营商名称，例如ChinaNet	N	Y	Y	N
\$network_type	字符串	网络类型，例如4G	N	Y	Y	N
\$utm_matching_type	字符串	iOS渠道追踪匹配模式， Cookie数据精准匹配/设备指纹模糊匹配	N	Y ¹	Y	N
\$utm_source	字符串	广告系列来源	N	Y ¹	Y	N
\$utm_medium	字符串	广告系列媒介	N	Y ¹	Y	N
\$utm_term	字符串	广告系列字词	N	Y ¹	Y	N
\$utm_content	字符串	广告系列内容	N	Y ¹	Y	N
\$utm_campaign	字符串	广告系列名称	N	Y ¹	Y	N

1.4、事件和指标之间的关系

上述内容梳理了事件及其维度，那指标和它们的关系是如何的？下面举一些典型示例进行解释：

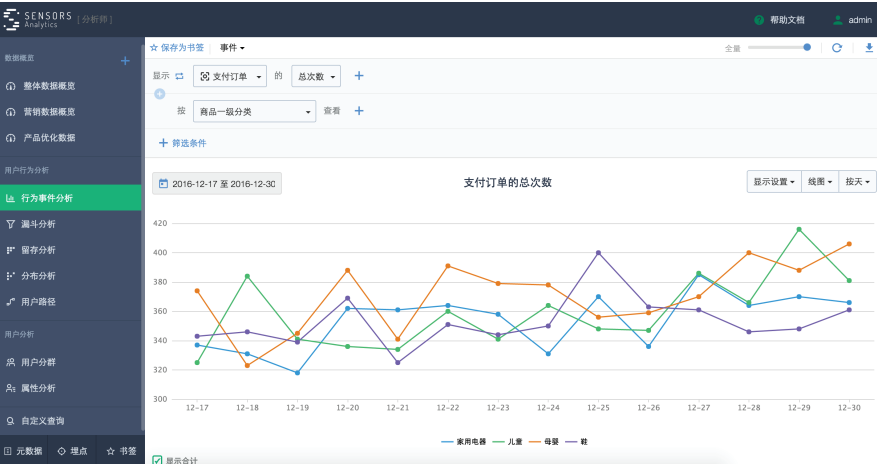
指标：新增用户数

解释：“启动 App”事件的触发用户数，且首次启动为是



指标：不同商品情况下的支付订单数

解释：以“支付订单”事件里的“商品一级分类”维度，进行分组查看该事件的总次数



指标：新用户留存率

解释：查看用户的初始事件为“激活 APP”，后续事件为“启动 APP”的情况，可以使用产品中现有的留存分析功能

用户先进行激活APP，后进行启动APP的7天留存分析

初始行为日期	总人数	1天之内	第2天	第3天	第4天	第5天	第6天	第7天
12-1(四)	1,981	310 15.65%	35 1.77%	20 1.01%	31 1.56%	25 1.26%	26 1.31%	28 1.41%
12-2(五)	1,924	288 14.97%	33 1.72%	24 1.25%	28 1.46%	34 1.77%	28 1.46%	25 1.3%
12-3(六)	1,914	289 15.1%	39 2.04%	29 1.52%	27 1.41%	25 1.31%	30 1.57%	40 2.09%
12-4(日)	1,975	314 15.9%	32 1.62%	20 1.01%	24 1.22%	32 1.62%	32 1.62%	32 1.62%
12-5(一)	1,960	333 16.99%	36 1.84%	31 1.56%	30 1.53%	22 1.12%	31 1.58%	36 1.84%
12-6(二)	1,956	326 16.67%	49 2.51%	30 1.53%	33 1.69%	25 1.28%	34 1.74%	32 1.64%

四、事件梳理中应规避的问题

事件梳理中需要注意的几种坑，分为三种，分别体现在 Events 表中、Users 表中还有一种是通用性的。

1、Events 表注意的问题

- 需求和技术沟通明白，清楚事件设计和业务关系，以及事件发生的条件

业务人员需明确事件所表达的业务含义，在何时会触发该事件，需要和技术表述清楚，从而有助于技术明确埋点方案和具体的采集方式。

- 事件和维度区分（重要）

严格理解事件和属性的区别，避免把属性设计成事件等，造成冗余。例：简单的统计三个按钮 A、B、C 的点击情况时，不需要做成“点击 A 按钮”、“点击 B 按钮”、“点击 C 按钮”三个事件，而是做成“点击按钮”事件，将 A、B、C 三个按钮以属性“按钮名称”进行传递。

- 公共行为和特征性行为

在电商产品中，“浏览页面事件”包含的属性一般都是比较通用的属性，如页面标题，页面地址等。这样的浏览行为我们称作为公共行为。而商品详情页的浏览，因为涉及属性较多，分析粒度较细，因此需要单独抽出做成单独事件。这样的浏览行为为特征性行为。

事件名称	事件属性	属性值格式
浏览页面	\$预置属性	字符串
	*是否首次访问	字符串
	网页Title	字符串
	网页URL	字符串
	前向网页路径（refer）	字符串
	自然流量来源域名	字符串
	自然流量来源URL	字符串
	*付费广告系列名称	字符串
	付费广告系列来源	字符串
	付费广告系列媒介	字符串
	付费广告系列关键词	字符串
	付费广告系列内容	字符串
	网页一级分类	字符串
	网页二级分类	字符串
点击查看商品详情	\$预置属性	字符串
	商品ID	字符串
	商品名称	字符串
	商品一级分类	字符串
	商品二级分类	字符串
	商品原价	数值
	商品现价	数值

• 同类事件，维度判断

电商产品中，类似于推荐位置的点击情况，和 Banner 位置的点击情况，虽然 是两个事件，但这两个事件以及他们的属性性质相近，因而可以合并成一个事件。

合并前

事件名称	事件属性	属性值格式
banner浏览	\$预置属性	字符串
	bannerID	字符串
	banner名称	字符串
推荐位浏览	\$预置属性	字符串
	推荐模块	字符串
	推荐位置	字符串

合并后

事件名称	事件属性	属性值取值	属性值格式
特定浏览	\$预置属性	字符串	字符串
	模块类型	banner, 推荐	字符串
	模块详细位置	bannerID1, bannerID2, 推荐位置1, 推荐位置2	字符串

虚拟事件的组合和拆分

虚拟事件的组合和拆分：神策分析分析界面提供虚拟事件功能，基于元事件的组合和拆解可以创建新事件，如对电商产品来说，可以定义一个活跃事件包含：“提交订单”、“支付订 单”、“浏览商品” 三个事件。

事件的组合

虚拟事件名

activeEvent

虚拟事件显示名

活跃事件

虚拟事件的组成

当以下事件中任意一个被触发时，视作该虚拟事件被触发

提交订单

+ 触发限制条件

支付订单

+ 触发限制条件

×

浏览商品

+ 触发限制条件

×

+ 添加事件

事件的拆解

虚拟事件名

viewPad

虚拟事件显示名

浏览平板电脑

虚拟事件的组成

当以下事件中任意一个被触发时，视作该虚拟事件被触发

浏览商品

+ 触发限制条件

商品类型

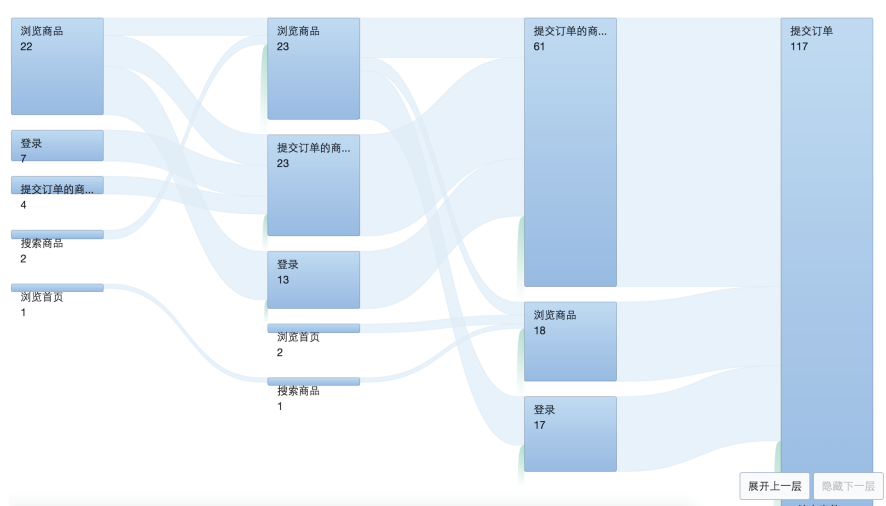
等于

平板电脑

+ 添加事件

• 来源问题，前向路径和渠道

神策分析提供灵活的用户路径分析功能，完整展现用户真实路径，因此对此类 分析需求，无需在事件属性中添加“前向路径”这样的属性。



2、Users 表注意的问题

• 单边，双边用户

单双边是针对产品有多个身份使用用户时才会进行区分。单边用户，即仅有一类用户的产品，如健身产品 Keep，聊天工具 QQ 等；双边用户如 O2O 产品，用户可能是普通消费者，也可能是商家用户。需要根据产品的不同，提前对用户识别和相应属性进行设计。

• 缓慢变化维

如果遇到一些会发生变化的属性，比如用户的 VIP 等级，不能只作为用户属性传进用户表中，还需在事件表中，记录一个“当前发生事件 VIP 等级”这个属性。因为当前会员等级的统计，和发生事件时用户的会员等级统计是两种情况。

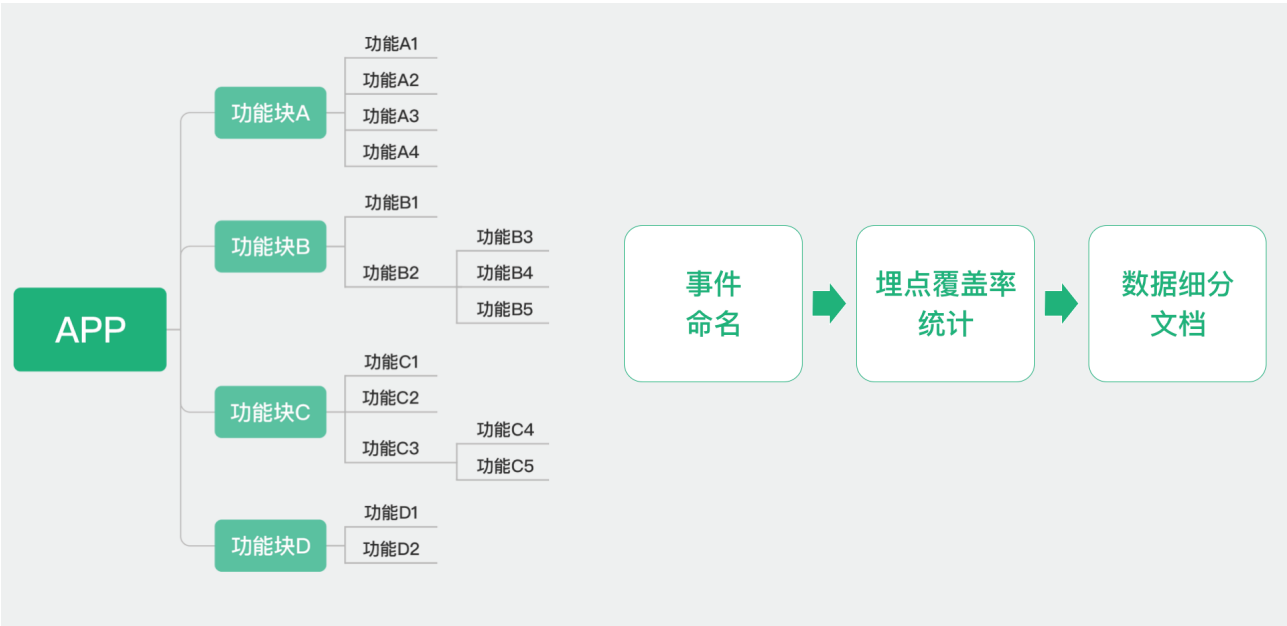
3、通用性问题

- 属性类型

神策分析会在上传第一条数据时确定该事件中所有属性的类型，所以需要提前做好属性类型再进行上传。

- 功能树

前期事件设计和埋点的规范性极其重要，如果对事件和属性的命名、埋点的覆盖范围没有一个好的规范文档，会对后期的分析需求和迭代造成混乱的现象。



神策分析 – 帮你实现数据驱动



联系我们

邮箱: contact@sensorsdata.cn

电话: 4006509827

地址: 北京市清华同方科技广场 D 座东楼