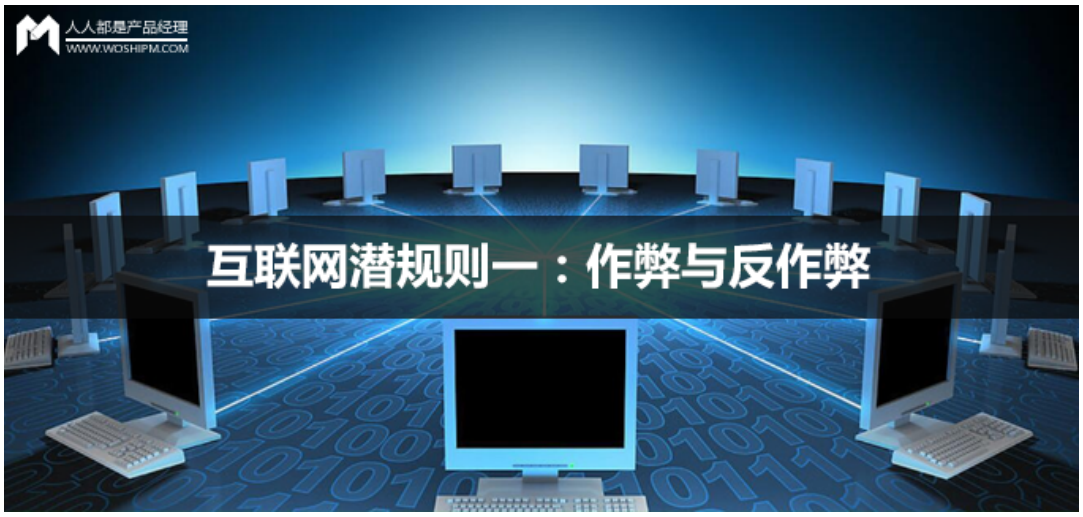
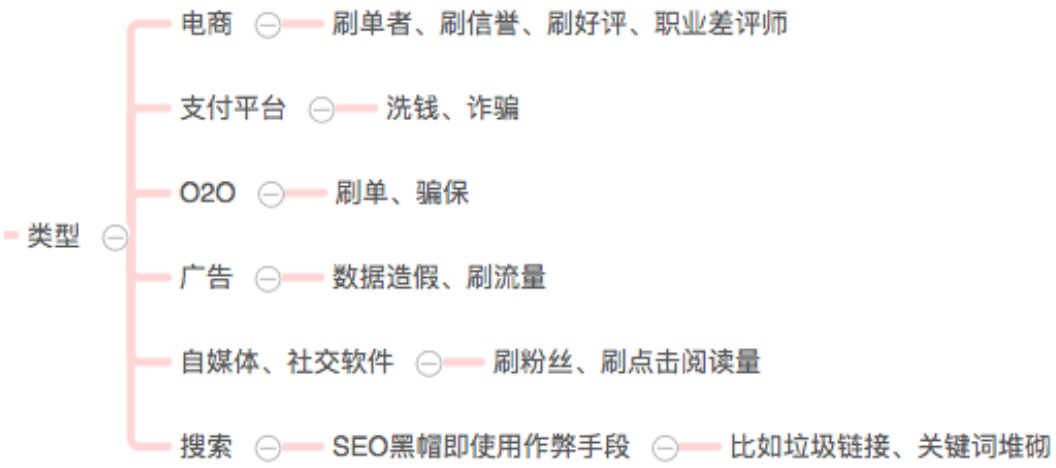
本文将阐述什么是互联网作弊，并以百度和淘宝看如何反作弊的行为。



互联网作弊是什么?

互联网作弊是一种很普遍的行为，就拿我们最熟悉的来说，有电商和O2O的刷单刷信誉行为、广告作弊等，具体分类如下图：



广告作弊与反作弊

1.背景：互联网广告成为主流

(1) 数字营销(互联网广告)分两类：

- 品牌广告：以品牌宣传为主，多以千次曝光的形式计费，广告主追求的是**长期的品牌溢价**;
- 效果广告：多以单次点击或单次行为的形式计费，更关注**短期转**

化和收益。

（2）投放方式：CPA、CPC、CPM每千人成本、CPP每购买成本等

2.现象：数据作弊

2016年上半年, AdMaster推出的《广告反欺诈白皮书》显示：2016上半年，AdMaster的广告反欺诈监测系统平均每天识别出高达 28% 的虚假流量。的确，中国的数字营销生态环境也正遭受着虚假流量的侵蚀。

3.作弊类型

- 曝光作弊：可能把广告展现在一些完全没有商业价值的垃圾流量上。
- 点击作弊：利用机器、人工或诱导用户点击，例如把广告换成一个美女图片，吸引完全不符合广告意图的点击。另外，竞争对手还可能进行恶意点击。
- 转化作弊：在注册、激活、下单等不同场景下通过自动化程序的模拟真人行为。

4.如何鉴别广告作弊

（1）初级作弊辨别：发现数据异常点。例如：

- 异常峰值
- 出现峰值时转化数据并没有增长
- 出现峰值时到站跳出率增长
- 投放的媒体属性和点击的地域属性不符

（2）中级作弊辨别：真人点击和机器模拟点击。例如：

- 点击请求的Headers异常
- 点击行为分析：机器点击具有一定的连续性，可以通过判断同IP同设备的连续点击、同IP段的大量点击、同IP连续点击间隔时间等进行判断。

——作者：岂安科技；出处：艾媒网（3）转化作弊辨别。例如：

- **行为频率、次数异常；**
- 注册者的URL访问轨迹：机器只访问注册URL**频繁注册；**
- 注册者是否查看了页面上的静态资源：机器注册在访问时**只关心**

网页上的文字；

- 不同账号**同密码注册；**
- 注册者从**到站到注册间的时间间隔**：真人在注册前会有较长时间的浏览过程，而机器行为则直奔主题。

以上整理自：http://www.gupowang.com/app/187.html，作者：姑婆；

5.如何反作弊?

- 目的：无限压缩作弊行为在正常商业行为中的比例，而非绝对根除。
- 最好的实现方法在于让作弊成本剧增。
- 思路：砌墙（不断的加限制条件）；拆台（使作弊行为的获利大幅度减少）

（1）排重：添加监测链接，通过Cookie、设备号或IP排重，如大量出现218.175.11.x这种相同C段的IP号。

（2）频度控制、SDK加密防护、人工介入监控

（3）点击有效期：限制点击的有效期，在有效期内，后续转化归属相应平台，如超时则不予计算。

（4）异常数据黑名单：对点击记录超过一定范围标记为黑名单，长期过滤。

（5）归因时间差监控：归因时间差即指从点击到下载激活的时间。一般作弊时，伪造点击与激活是并存的，所以往往在时间逻辑上是错误的。

（6）增加行为操作的复杂度，但可能伤害用户。

SEO反作弊——以百度为例

1.搜索引擎优化——SEO

具体来说，就是通过**站内优化**比如网站结构调整、内容建设、代码优化等，以及**站外优化**比如网站站外推广、品牌建设等，使网站满足搜索引擎收录排名需求，在搜索引擎中**提高关键词排名**，从而**吸引精准用户进入网站**，获得**免费流量**，产生直接销售或品牌推广。

2.百度怎么反作弊?

（1）绿萝算法：2013年2月上线的搜索引擎反作弊算法，主要**打击超链中介、出卖链接、购买链接**等超链作弊行为。通过综合**外链内容的相关性、A及B网站页面内容品质、更新频率、违规历史记录、总权重值**，从而判断外链的权重传递是否有效。

（2）石榴算法：针对低质量网站的进一步打击的升级版，将**重点整顿含有大量妨碍用户正常浏览的恶劣广告的页面**，尤其以弹出大量低质**弹窗广告、混淆页面主体内容的垃圾广告**页面为代表。

电商反作弊——以淘宝为例

淘宝搜索反作弊系统不仅**监控卖家行为，同时也监控买家行为，并且通过对买家ID的行为监控可倒推反证卖家作弊**。而且，该算法还可以**作为推荐算法使用。**反作弊手段大致划分为以下3种：**信任传播模型、不信任传播模型和异常发现模型**。

（以下来源于淘宝搜索技术内参，由薄言整理http://www.tao-sou.com/809.html）

1.信任传播模型

在海量的宝贝网页数据中，通过一定手段，**筛选出绝不会作弊的店铺、宝贝和ID（即白名单）**。算法以这些白名单内的页面作为出发点，**赋予白名单内的页面节点较高的信任度分值**，其他宝贝、买家、卖家是否作弊，要根据其和白名单内节点店铺或宝贝的成交关系来确定。**白名单内节点通过成交关系将信任度分值向外扩散传播，如果某个节点最后得到的信任度分值低于这一阈值**，那么该宝贝网页、买家或卖家则会被认为是有作弊嫌疑。

2.不信任传播模型

从大的技术框架上来讲，其和信任传播模型是相似的，最大的区别在于初始的页面子集合不是值得信任的店铺或宝贝页面节点，而是**确认存在作弊行为的页面或ID集合（即黑名单）。赋予黑名单内页面节点不信任分值，通过成交关系将这种不信任关系传播出去，如果最后页面节点的不信任分值大于设定的阈值**，则会被认为是作弊网页或有作弊嫌疑。

3.异常发现模型

先找到一些作弊或非作弊的集合，分析出其绝对特征有哪些，然后利用这些特征来识别作弊行为。具体来说，一种是直接从作弊行为包含的独特特征来构建算法；另一种是通过统计等手段分析正常的店铺、宝贝和ID应该具备哪些特征，如果不具备则被认为是作弊。这几种都是通过分析行为之间、物品之间的相似度或区别度，故也可以用来用于用户的个性化推荐，比如我们常见的“猜你喜欢”、“向你推荐”等。这是我了解的最有趣的一点。

作者：小乔，产品小白一枚。微信公众号：乱入花间花绿叶（qiaomaihexiaoqiao）