

<http://www.tao-sou.com/809.html>

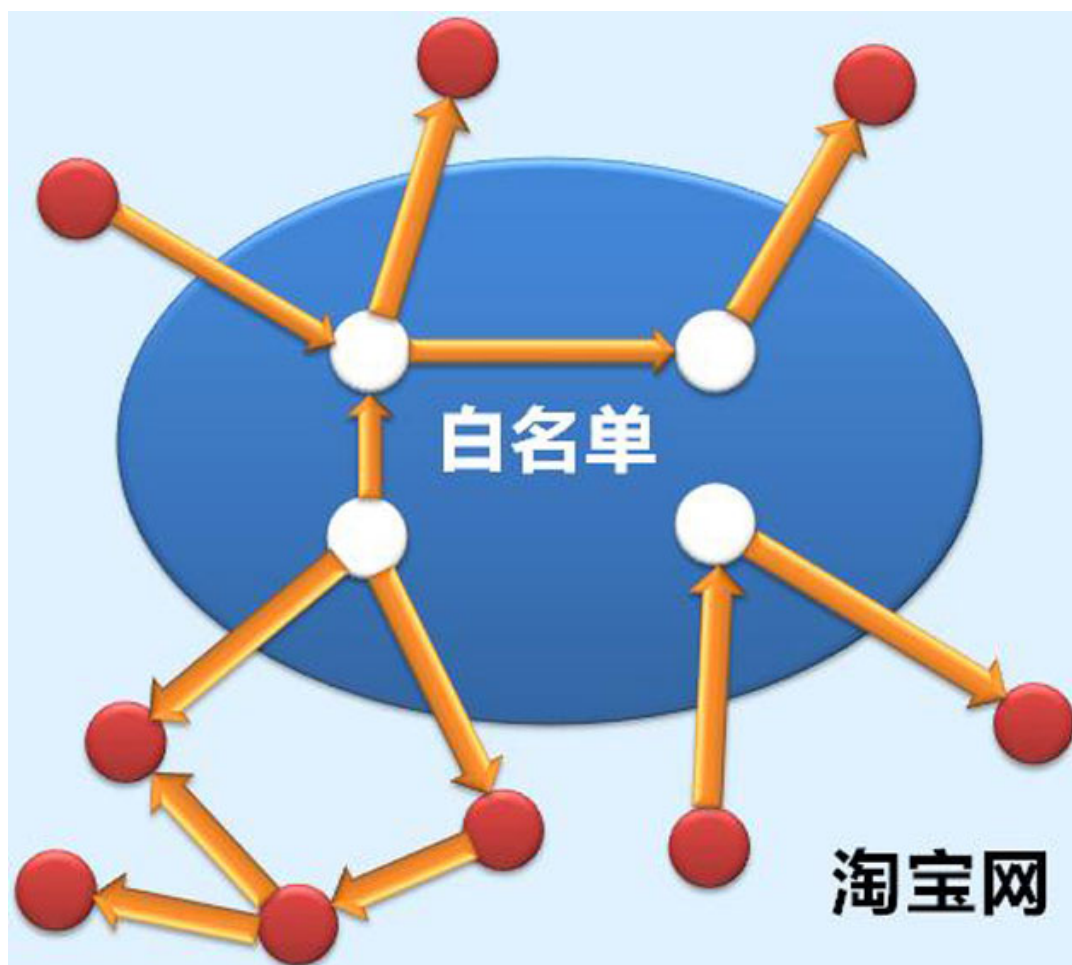
目前针对淘宝搜索引擎的作弊手段五花八门、层出不穷，作为应对方的淘宝搜索引擎，也相应调整技术思路，不断有针对性地提出反作弊的技术方案，所以如果整理反作弊技术方案，会发现技术方法很多，理清思路不易。

尽管如此，如果对淘宝反作弊技术深入分析，会发现在整体技术思路还是有规律可循的。从基本的思路角度看，可以将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。其中前两种技术模型可以进一步抽象归纳为子集传播模型，为了简化说明，此处不再赘述，而是直接将这两个子模型列出，将具体算法和这几个模型建立关系，有助于对反作弊算法的宏观思路和相互联系建立起清晰的概念。

在此需要强调的是，这三种模型不仅对于淘宝卖家适用，同时也对淘宝买家适用，换句话说，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。

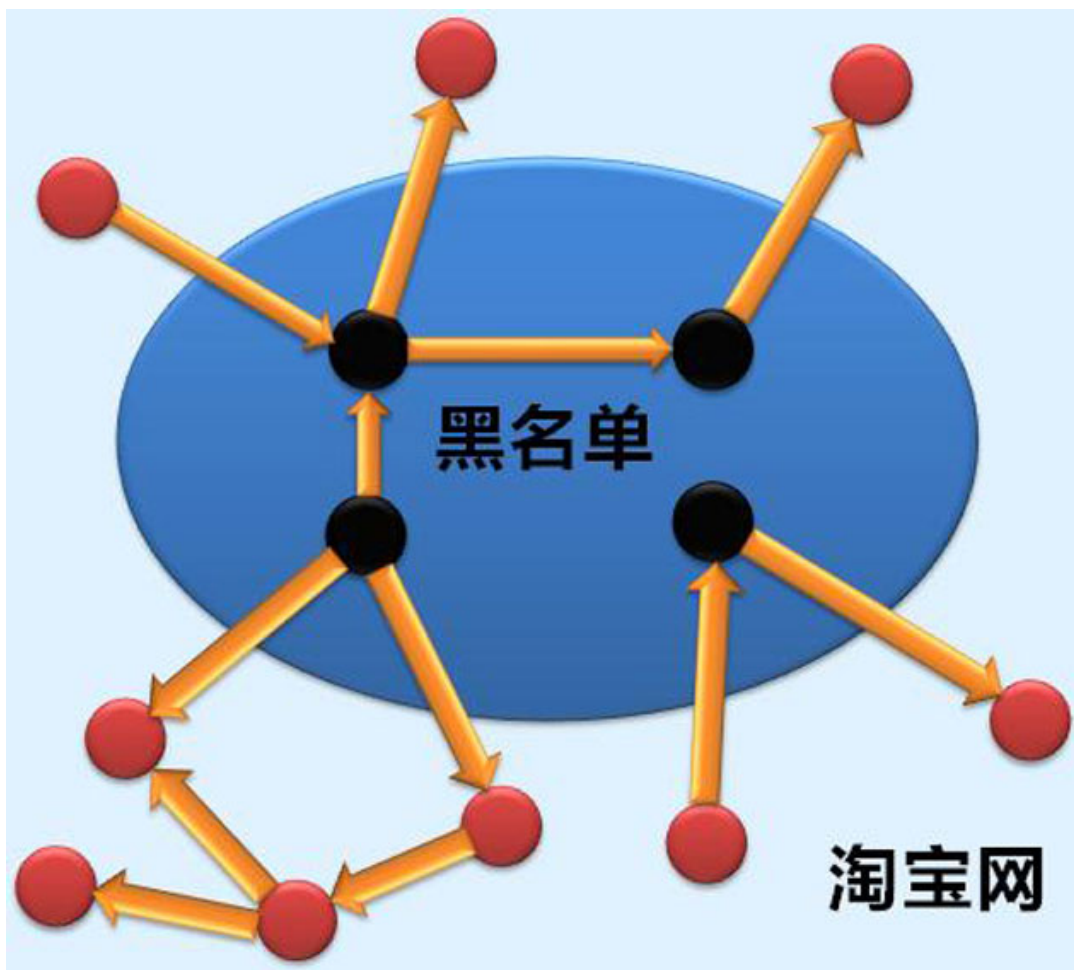
一、信任传播模型

下图展示了信任传播模型的示意图。所谓信任传播模型。基本思路如下：在海量的宝贝网页数据中，通过一定的技术手段或者人工半人工手段，从中筛选出部分完全值得信任的店铺或者宝贝页面，也就是肯定不会作弊的店铺、宝贝和ID（可以理解为白名单），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高于一定阈值，则认为没有问题，而低于这一阈值的宝贝网页、买家或卖家则会被认为是作弊或有作弊嫌疑。



二、不信任传播模型

下图展示了不信任传播模型的整体框架示意图。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，即不值得信任的集合（可以理解为黑名单）。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。



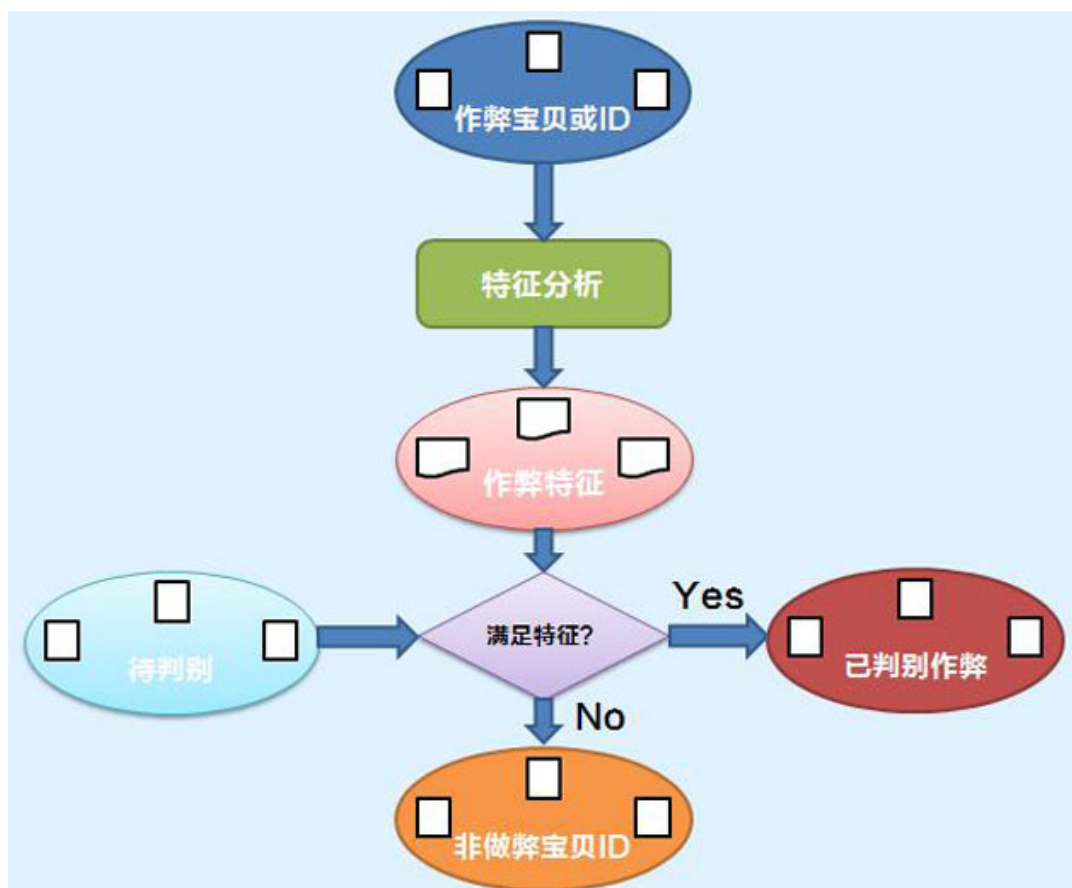
同样，很多算法可以归入这一模型框架，只是在具体实施细节方面有差异，整体思路基本一致。

三、异常发现模型

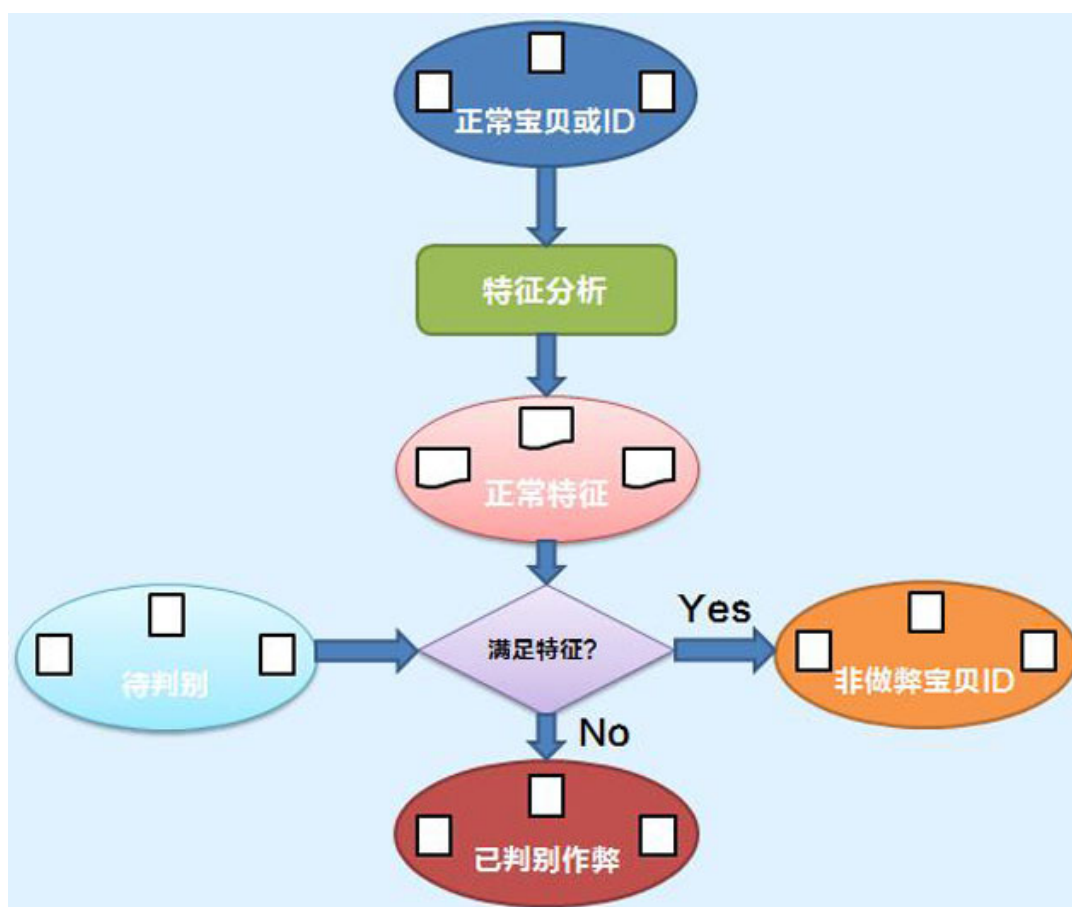
异常发现模型也是一个高度抽象化的算法框架模型。其基本假设认为：作弊店铺、卖家和买家必然存在有异于正常的特征，这种特征有可能是内容方面的，也有可能是成交关系方面的，而制定具体算法的流程往往是先找到一些作弊的集合，分析出其异常特征有哪些，然后利用这些异常特征来识别作弊行为。

具体来说，这个框架模型又可细分为两种子模型，这两种子模型在如何判断异常方面有不同的考虑角度。一种考虑角度比较直观，即直接从作弊行为包含的独特特征来构建算法；另外一种角度则认为不正常的

网页或ID即为作弊行为，也就是说，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，如果不具备这些正常的特征，则被认为是作弊。下面两图体现了这两种不同的思路。



尽管反作弊算法五花八门，但是不论采取哪种具体算法，其实都包含了一些基本假设，以上的三种模型就是经常被反作弊算法使用的基本假设，很多算法的基本思路都是从这些基本假设出发来构造的。



郑重声明：本站“淘宝搜索技术文档”版块所有内容均来自淘宝搜索技术内参，并由薄言亲自归纳整理，旨在将复杂的技术文档转化成直白容易理解的文字，以此来帮助淘宝卖家更好的了解淘宝搜索系统，为此耗费了巨大的时间和精力，请大家尊重薄言的劳动，不要转载和盗用，即使需要转载，请务必取得授权，并注明出处！谢谢！