实时序热搜图片作弊策略

# 一、背景介绍

作弊图片严重影响了搜索结果的呈现质量，尤其是在实时热搜的结果下。这类搜索用户多、流量大、影响用户体验。反作弊组已通过建立图片指纹库等方式来打压图片作弊。

但目前图文不相关的图片作弊仍然频繁，尤其是在特定时段的实时序，例如晚上7-9点、周末等。作弊的账号包括盗号、养号、运营账号作弊等。

具体的作弊方式、案例请参考下面链接：

<http://search.intra.weibo.com:8080/boards/123/topics/8119>

# 二、目标和方法

在AC端要做的事情就是从用户的发微博习惯上寻找作弊的行为特征，AC二次排序时直接将其删除。

首先，学习历史数据。尤其是反作弊日志，通过统计一年数据，来发现一些用户的一些行为特征，例如2015年6月至今只用一个终端发原创带图片微博的概率大于0.7等。具体见下文；

是学习作弊CASE，寻找作弊CASE的共性。案例主要来源是PM提供的作弊CASE以及对线上作弊CASE长期观察，尤其是实时热搜下的作弊CASE。目前总结出了一些作弊微博的特征：例如发布来源不一致、特定发布来源（某些浏览器）微博质量打分、发图图片个数偏多（多是6张以上）；

其次，线上策略验证。鉴于观察到的作弊CASE相对太少，作弊行为的验证需要到线上的实际环境中去测试。因此，目前的策略只写了相关日志，如果策略准确率和召回率都满意，则通过配置生效；

进一步，用户的行为特征是多维度的。随着反作弊日志的丰富和线上策略的验证，在保证准确率和召回率的基础上，后期会加大过滤力度。

# 三、策略迭代、CASE分析

## 3.1 数据分析

先统计HBase历史数据如下特征的UID列表：

1. 一定时间范围内（过短不能反映变化，过长空间复杂）发原创微博的频率；

2. 原创微博的是否带图片以及个数（非内容特征）；

3. 微博发布来源。

上面统计特征作为一个类似白名单的UID列表，对实时首页的带图片微博进行作弊识别，如果不在该UID列表，则被判断为图片作弊，二次排序之前直接从结果集移除。

计划是，先做特征统计，观察识别效果，迭代进行。

目前只做了用户发博终端类型分析。

数据来源：2015-6-11至今的反作弊日志

过滤条件：发 带图片 && 原创微博 && 非作弊

目的字段：uid, source, num

uid总数：

53,957,362

使用一种source发帖概率超过0.8的uid总数：

42,886,471

使用一种source发帖概率超过0.5的uid总数：

48,789,481

使用2个及以上source发帖的uid总数：

15,739,240

占总数比：0.292， 换言之超过70%的人只用一个source发微博 [发图片的用户]

使用3个及以上source发帖的uid总数：

6,001,817

也就是说用两个source的人有15739240-6001817 = 9737423

两个 占比（多个source发帖的uid）：0.618 [黄金分割。。。]

两个 占总数比：0.180

使用4个以上source发微博的用户数量：

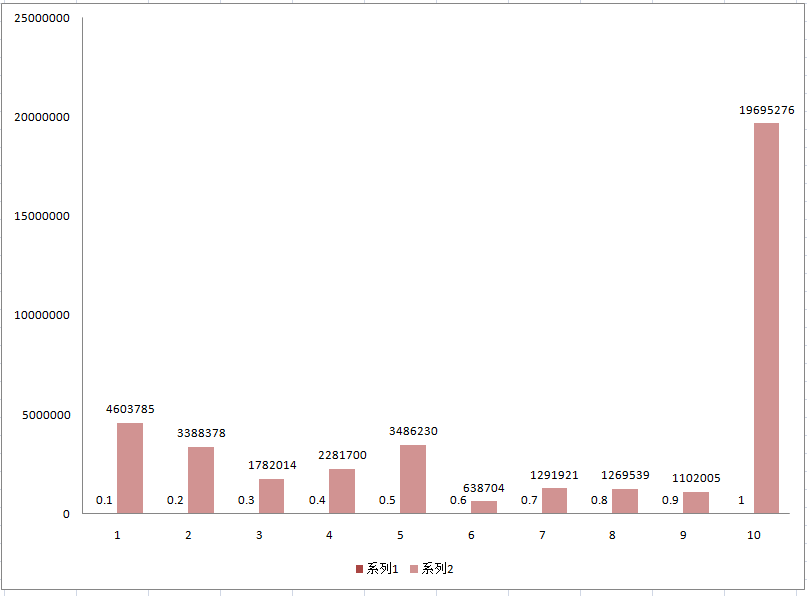
2,537,202

占比（多个source发帖的uid）：0.161

占总数比：0.047

**结论：70%以上的发图片用户都只用一种source，用多个source的用户中有60%以上只用两种source，80%以上的用户用两个或三个source。**

只用两个source的uid中使用某个source频率大于等于0.5的概率：0.9999



## 3.2 基于分析、观察的一些结论

现在有2类作弊非常严重：

主要是在周末的时候出现 平时稍微少点

1 之前的图片作弊 现在的热词已经非常少 基本一个微博 就一个热词

2 转发一个图片作弊的微博 转发的内容只有一个热词 其他完全正常

A pc端有聚合的，打算对原创没有召回的聚合在首页移除，

不使用聚合体最新的时间 使用其他的 只有转发时间都在首页的才删除

B 客户端 首页不展示转发微博

3. 一般图片作弊都在实时热搜第一条发布而且发布频率挺高，几分钟一次，转发和原创都有，第二条很少（PC端）；

4. 一般图片作弊的发得图片往往都很多，例如很多都是9副图片，3\*3排列，而正常实时热搜的结果没有这么多图片，个数也很随机；

## 3.3 初步策略

生成一个允许在热词 带图片 原创 首页 的uid列表

通过

1 统计最近3个月有发布过原创微博

2 有发布过一定数量的带图片的原创微博

3 统计发布这些微博的时候试用的来源（80%）

最后就是一个uid 来源的对应表

使用的时候

1 热词

2 原创 带图片

3 首页

必须在这个对应uid表里面

### Case分析

目前词典：3个月的（上周统计的），有图片，原创，非作弊，使用次数最多的source

query:努力努力再努力x （实时热搜第一条，测试时没有图片作弊微博）

DEBUG 2015-9-21 11:00:56 ac\_sort.cpp:hot\_pic\_cheat\_check:5271 T000# digit\_attr: 11, uid:5231509558, most\_used\_source:28, using source:776787

DEBUG 2015-9-21 11:00:56 ac\_sort.cpp:hot\_pic\_cheat\_check:5272 T000# delete:uid source cheat,mid:3889465816305479

DEBUG 2015-9-21 11:00:56 ac\_sort.cpp:hot\_pic\_cheat\_check:5271 T000# digit\_attr: 9, uid:1810975494, most\_used\_source:1370926, using source:12758

DEBUG 2015-9-21 11:00:56 ac\_sort.cpp:hot\_pic\_cheat\_check:5272 T000# delete:uid source cheat,mid:3889465397787531

DEBUG 2015-9-21 11:00:56 ac\_sort.cpp:hot\_pic\_cheat\_check:5271 T000# digit\_attr: 9, uid:3919982207, most\_used\_source:871267, using source:843300

DEBUG 2015-9-21 11:00:56 ac\_sort.cpp:hot\_pic\_cheat\_check:5272 T000# delete:uid source cheat,mid:3889465090802625

DEBUG 2015-9-21 11:00:56 ac\_sort.cpp:hot\_pic\_cheat\_check:5271 T000# digit\_attr: 9, uid:5049746569, most\_used\_source:28, using source:871267

DEBUG 2015-9-21 11:00:56 ac\_sort.cpp:hot\_pic\_cheat\_check:5272 T000# delete:uid source cheat,mid:3889463757384943

DEBUG 2015-9-21 11:00:56 ac\_sort.cpp:hot\_pic\_cheat\_check:5271 T000# digit\_attr: 9, uid:5551797178, most\_used\_source:1724548, using source:333536

DEBUG 2015-9-21 11:00:56 ac\_sort.cpp:hot\_pic\_cheat\_check:5272 T000# delete:uid source cheat,mid:3889463098715374

DEBUG 2015-9-21 11:00:56 ac\_sort.cpp:hot\_pic\_cheat\_check:5271 T000# digit\_attr: 11, uid:2956939237, most\_used\_source:1302240, using source:776787

DEBUG 2015-9-21 11:00:56 ac\_sort.cpp:hot\_pic\_cheat\_check:5272 T000# delete:uid source cheat,mid:3889463090006712

DEBUG 2015-9-21 11:00:56 ac\_sort.cpp:hot\_pic\_cheat\_check:5271 T000# digit\_attr: 9, uid:2168489791, most\_used\_source:28, using source:25

DEBUG 2015-9-21 11:00:56 ac\_sort.cpp:hot\_pic\_cheat\_check:5272 T000# delete:uid source cheat,mid:3889463064826964

误伤率太高，逐个检查case:

3889465816305479： 过去三个月发了大量带图片的原创微博，基本都是用华为Ascend P7，而这次使用的是 米公益应用， 根据策略被过滤掉了

3889465397787531： 过去三个月发了12条原创带图片微博，其中6条是weibo.com，另外6条是iPad客户端，而字典里只保留了weibo.com，这次的结果给过滤掉了

3889465090802625： 根据反作弊日志统计出来的结果是uid 3919982207 source 871267 条数 2 ，这个与事实不符，过去三个这个用户用andriod发了很多原创带图片微博，也不像是作弊的

3889463757384943： 结果同上，根据反作弊日志统计出来的结果不对

3889463098715374： 结果同上，根据反作弊日志统计出来的结果不对

3889463090006712： 结果同第一个

3889463064826964： source 28和25是不是同一个？ 这个uid只用了 [桃兔��iPhone] 发微博

上面的误伤原因：

最主要的是：根据反作弊日志统计出来的结果似乎与实际情况不同，一是发带图片原创微博条数不同，二是常用source不同。需要测试确认下原因；

第二个是：用户可能会有2个以上的常用source，但是这个case比较少；

第三个是：source数值是否是source的唯一标识？不同值是否对应同一个或同一类source？例如，上面的28和25在页面上都是[桃兔��iPhone]这个也需要进一步确认。

基于上面的问题：

一：确认日志统计不符合预期的原因，确定是分析的问题还是日志的问题；

二：增加微博数目过滤，暂定大于3条吧，太少增加误伤概率；

三：增加时间阈值，一周内的暂不统计；

四：考虑是否增加两个source最为筛选条件。

## 3.4 策略调整

对词典进行了优化：

词典内容为（uid,source1,source2），source1和source2为uid使用频率最高的两个source，且两个两个使用频率之和大于等于0.9；

每个source的发得原创带图片微博个数要大于3，同时对发图片个数也进行了限制：

在生成白名单的时候不考虑大于8张图片的微博，原因是正常用户发8张以下的很多，不会因为去掉而失准，作弊用户多发8张以上的，去掉会很大概率上防止其落入白名单

对上面的优化进行了验证，目前热搜词已经不再有误伤。但这个策略，不一定能抓到作弊的。现在热词“高圆圆”下的一个作弊案例：3902151091199254

1749940242 1142370|174

1749940242 1142370 10 0.4 174 13 0.52

现在发的这个就是使用174，频率为0.52，按我们的策略不会过滤掉

1 关于微博发布的来源，我观察了一下：很多用户的常用发布来源可能不止一个，2个或三个都有（例如weibo.com和iphone），所以source是不是有多个？ 通过阈值 >30%来确定合法source ？

2 如果有多个合法source的时候，最终生成的字典包括 uid+sourceids ，命中需要一次hash，多次比较，这样查询可能会慢一些 （如果只有一个source则key为uid\_source，一次Hash即可）

### Case分析

现在的词典包括历史数据，非作弊的、发原创带图片的超过3个。词典内容还是key:uid value:source

目前测试了几个热搜词：

朱七七

DEBUG 2015-9-23 15:16:23 ac\_sort.cpp:hot\_pic\_cheat\_check:5287 T000# digit\_attr: 1, uid:3909128977, most\_used\_source:174, using source:333536

DEBUG 2015-9-23 15:16:23 ac\_sort.cpp:hot\_pic\_cheat\_check:5288 T000# delete:uid source cheat,mid:3890254203405622 【误伤了】

DEBUG 2015-9-23 15:16:23 ac\_sort.cpp:hot\_pic\_cheat\_check:5287 T000# digit\_attr: 1, uid:2049246117, most\_used\_source:1252310, using source:174

DEBUG 2015-9-23 15:16:23 ac\_sort.cpp:hot\_pic\_cheat\_check:5288 T000# delete:uid source cheat,mid:3890254056525995 【作弊的】

赵丽颖 玉无心 【全部误伤了，没有作弊】

DEBUG 2015-9-23 15:22:47 ac\_sort.cpp:hot\_pic\_cheat\_check:5287 T000# digit\_attr: 1, uid:3087414264, most\_used\_source:717, using source:25

DEBUG 2015-9-23 15:22:47 ac\_sort.cpp:hot\_pic\_cheat\_check:5288 T000# delete:uid source cheat,mid:3890255340167340

DEBUG 2015-9-23 15:22:47 ac\_sort.cpp:hot\_pic\_cheat\_check:5287 T000# digit\_attr: 1, uid:3936675145, most\_used\_source:174, using source:1302240

DEBUG 2015-9-23 15:22:47 ac\_sort.cpp:hot\_pic\_cheat\_check:5288 T000# delete:uid source cheat,mid:3890254601267180

DEBUG 2015-9-23 15:22:47 ac\_sort.cpp:hot\_pic\_cheat\_check:5287 T000# digit\_attr: 1, uid:3194897354, most\_used\_source:425645, using source:576792

DEBUG 2015-9-23 15:22:47 ac\_sort.cpp:hot\_pic\_cheat\_check:5288 T000# delete:uid source cheat,mid:3890254240615404

DEBUG 2015-9-23 15:22:47 ac\_sort.cpp:hot\_pic\_cheat\_check:5287 T000# digit\_attr: 1, uid:5487499810, most\_used\_source:1370926, using source:1302240

DEBUG 2015-9-23 15:22:47 ac\_sort.cpp:hot\_pic\_cheat\_check:5288 T000# delete:uid source cheat,mid:3890253746115530

DEBUG 2015-9-23 15:22:47 ac\_sort.cpp:hot\_pic\_cheat\_check:5287 T000# digit\_attr: 9, uid:1764065424, most\_used\_source:317, using source:615949

DEBUG 2015-9-23 15:22:47 ac\_sort.cpp:hot\_pic\_cheat\_check:5288 T000# delete:uid source cheat,mid:3890246611602097

跟着习大大去美国 【没有误伤，没有作弊】

刘德华全家福【全部误伤了，有一个作弊没抓到】

DEBUG 2015-9-23 15:27:37 ac\_sort.cpp:hot\_pic\_cheat\_check:5287 T000# digit\_attr: 1, uid:5217866901, most\_used\_source:1302240, using source:1370926

DEBUG 2015-9-23 15:27:37 ac\_sort.cpp:hot\_pic\_cheat\_check:5288 T000# delete:uid source cheat,mid:3890252882102571 【误伤了】

DEBUG 2015-9-23 15:27:37 ac\_sort.cpp:hot\_pic\_cheat\_check:5287 T000# digit\_attr: 1, uid:1863392110, most\_used\_source:1302240, using source:174

DEBUG 2015-9-23 15:27:37 ac\_sort.cpp:hot\_pic\_cheat\_check:5288 T000# delete:uid source cheat,mid:3890227875663524 【误伤了】

DEBUG 2015-9-23 15:27:37 ac\_sort.cpp:hot\_pic\_cheat\_check:5287 T000# digit\_attr: 1, uid:1720071623, most\_used\_source:1370926, using source:871267

DEBUG 2015-9-23 15:27:37 ac\_sort.cpp:hot\_pic\_cheat\_check:5288 T000# delete:uid source cheat,mid:3890224058730198

误伤比之前要小一些，但是还是很高，也有作弊的抓不到。

## 3.5 策略调整

两个source的情况，误伤更低和准确率更高。

我们用一个source过滤，这个source的使用频率很高，假设设定阈值为90%，针对一次查询：

a. 该uid在词典中： a1 |与目标source相同——> 非作弊，放过 [决定准确率]

a2 |否则 ——> 作弊，删除 [决定误伤率]

b. 该uid不在词典中：无论是否作弊，放过

我们使用两个source过滤，同样的使用频率阈值 90%，但是source1和source2的使用频率之和，针对一次查询：

a. 该uid在词典中： a1 |与目标source1或source2相同——> | 非作弊，放过 [决定准确率]

a2 |否则 ——> 作弊，删除 [决定误伤率]

b. 该uid不在词典中：无论是否作弊，放过

使用两个source，同样的阈值，但是每个source的频率低了，所以更多的Uid会走a分支，准确率提高；进入a分支之后，因为两种可能，且每种的使用频率都较低，落入a1的概率更大，落入a2的频率减小，即误伤更小。

某些case抓不到，使我们策略本身的缺陷，例如，一个用户经常使用一个source发作弊图片，这个策略是没有办法的

### Case分析

抽样调查：只要不命中词典就是作弊的策略，抽样跑了一些热词，目前基本没有作弊，所以这些都是误伤。

移动偷跑流量 证据 2个

[hot\_pic\_cheat\_check:mid:3903326082967521] [hot\_pic\_cheat\_check:mid:3903326011271440] [hot\_pic\_cheat\_check:mid:3903325957053460]

移动偷跑流量 证据 7个

[hot\_pic\_cheat\_check:mid:3903327122972377] [hot\_pic\_cheat\_check:mid:3903327106311164] [hot\_pic\_cheat\_check:mid:3903327048230768] [hot\_pic\_cheat\_check:mid:3903326082967521]

[hot\_pic\_cheat\_check:mid:3903326011271440] [hot\_pic\_cheat\_check:mid:3903325957053460] [hot\_pic\_cheat\_check:mid:3903325793283497]

15岁产子 2个

[hot\_pic\_cheat\_check:mid:3903325835597498] [hot\_pic\_cheat\_check:mid:3903320567117116]

长得丑不像做好事的人 4个

[hot\_pic\_cheat\_check:mid:3903318000197082] [hot\_pic\_cheat\_check:mid:3903314670098523] [hot\_pic\_cheat\_check:mid:3903311038398414] [hot\_pic\_cheat\_check:mid:3903308307551976]

杨紫 4个

[hot\_pic\_cheat\_check:mid:3903327522080289] [hot\_pic\_cheat\_check:mid:3903323855816959] [hot\_pic\_cheat\_check:mid:3903314720645492] [hot\_pic\_cheat\_check:mid:3903314535967973]

小丸子谈恋爱 7个

[hot\_pic\_cheat\_check:mid:3903326854358114] [hot\_pic\_cheat\_check:mid:3903326528010981] [hot\_pic\_cheat\_check:mid:3903325114555118] [hot\_pic\_cheat\_check:mid:3903324857849792]

[hot\_pic\_cheat\_check:mid:3903317354642666] [hot\_pic\_cheat\_check:mid:3903315169033022] [hot\_pic\_cheat\_check:mid:3903314117189192]

杨洋 夜华 7个

[hot\_pic\_cheat\_check:mid:3903327303780072] [hot\_pic\_cheat\_check:mid:3903325294097361] [hot\_pic\_cheat\_check:mid:3903325059740203] [hot\_pic\_cheat\_check:mid:3903324501664036]

[hot\_pic\_cheat\_check:mid:3903323855523870] [hot\_pic\_cheat\_check:mid:3903323717023277] [hot\_pic\_cheat\_check:mid:3903322274279291]

移动偷跑流量 证据 4个

[hot\_pic\_cheat\_check:mid:3903329199203942] [hot\_pic\_cheat\_check:mid:3903327915908702] [hot\_pic\_cheat\_check:mid:3903327890548561] [hot\_pic\_cheat\_check:mid:3903327882332636]

15岁产子

[hot\_pic\_cheat\_check:mid:3903320567117116]

长得丑不像做好事的人 4个

[hot\_pic\_cheat\_check:mid:3903318000197082] [hot\_pic\_cheat\_check:mid:3903314670098523] [hot\_pic\_cheat\_check:mid:3903311038398414] [hot\_pic\_cheat\_check:mid:3903308307551976]

杨紫 4个

[hot\_pic\_cheat\_check:mid:3903323855816959] [hot\_pic\_cheat\_check:mid:3903314720645492] [hot\_pic\_cheat\_check:mid:3903314535967973] [hot\_pic\_cheat\_check:mid:3903304050002477]

# 四、下一步计划

之前的图片作弊我看了很多发现行为比较像被盗号

特征是这个账户之前很旧都不发微博

或者之前发微博都是在特定的来源【比如特定的手机等等】

所以之前我让立舟做了一个事情

统计一批在3个月---一周以前 发过原创带图片 总数大于阈值【3】的来源

当在检索热词的时候，对前20微博，检查来源是否一致

这样就是说 认为用户会有比较一致的行为

对于现在来源一致，是否之后可以扩展到发布地点？

大部分人应该不会到处在不同城市工作

能否用几个不太容易发生变化的特征，将用户分类

比如：发微博的来源，发微博的地点 等等

1 完全沉寂的，3个月内都没有特定行为

2 行为不定，3个月内的特定行为一直在变

3 行为一致，3个月内的特定行为比较固定【可能1-2个】

对应第一类，发的微博直接干预

第二类，发的微博 不处理

第三类，发的微博，必须符合之前的特定行为才允许，否则干预掉