爱奇艺安全的实时智能化实践

爱奇艺 资深安全算法专家 宗志远

爱奇艺-悦享品质

做一家以科技创新为驱动的伟大娱乐公司

爱奇艺，中国高品质视频娱乐服务提供者

2018.3.29 纳斯达克上市 代号:IQ

扫描二维码关注公共微信账号
【爱奇艺行业速递】
更及时帮助大家了解行业动态

爱奇艺

我们面对的威胁

会员

撞库盗号
帐号分享
批量注册
垃圾注册

视频

搜索爬虫
盗播盗看
广告屏蔽
推荐作弊
广告作弊
评论作弊
弹幕作弊
有效观看作弊
页面浏览作弊

活动

薅羊毛
抽奖作弊
投票作弊
拉新作弊
任务作弊

直播

挂站人气
恶意图文
抢红包

电商

恶意下单
订单欺诈
黄牛
虚拟商品套现

支付/金融

盗号盗卡
洗钱
恶意提现
恶意借款
代理中介
信息伪造
征信修饰

其他

钓鱼邮件
恶意爆破
短信轰炸
垃圾信息
渠道防刷

爱奇艺

黑产画像

黑产：网络技术+分工+非法牟利

据不完全统计，我国每年黑产从业人员在**百万级**以上，造成的损失**超过千亿**

https://www.leiphone.com/news/201511/UPC7tZTKIaPHncL7.html

http://www.weste.net/2015/09-15/105859.html

智能化安全的挑战

数据质量	多源	结构多样	对象缺乏唯一标识符
计算平台	海量数据	实时性要求高	关联化要求高
标注样本	依赖业务背景	样本匮乏	样本有效期差异大
性能评估	准确率要求高	安全认知不一致	黑灰白可分性差

爱奇艺智能风控

系统概览(代号：麦哲伦、哥伦布、郑和)

解决方案	羊毛党	盗号/盗卡	批量注册	拉新	刷量	秒杀	投票	刷单	
数据产品	IP信誉	用户信誉	设备信誉	安全画像	欺诈特征库	深度检测	统一清洗		
核心技术	白盒加密	设备指纹	生物探针	验证中心	安全盾APP				
	联防联控	超高频检测	大规模异常检测	社团挖掘	风险传播	智能评分卡	自动驾驶		
风控平台	策略编辑	策略灰度	策略仿真	上线审批	上线审批	一键发布	监控/报表	智能报警	案件管理
	实时服务	准实时分析	近实时分析	离线分析					
	弹性扩容	服务熔断	模型引擎	规则引擎	查询引擎	插件引擎			
核心引擎	强化学习引擎	模型训练引擎	模型推理引擎	异常检测引擎	特征检测引擎	图挖掘引擎	图计算引擎	UEBA	人机识别
内外部数据源	内外部攻击情报	业务黑白名单	商业情报	暗网情报	业务数据	行为数据	基础信息知识库(ip/手机号/身份证/银行卡)		

业务概览

覆盖：

passport、会员、播放、拉新、APP启动、泡泡、文学、短视频、直播、秀场、支付、电影票等40+业务大方向

几组数字：

50　　15000　　18500000　　800000000

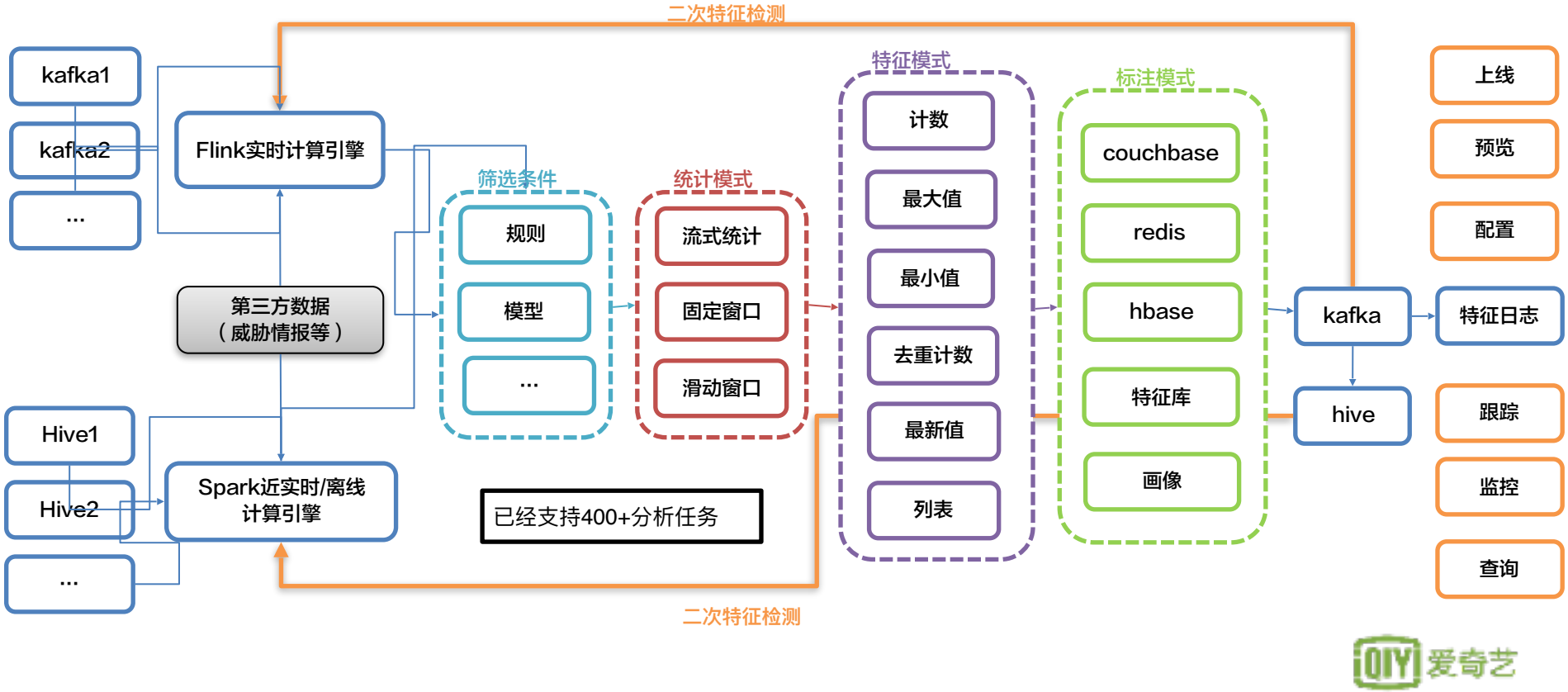
爱奇艺

智能风控演进

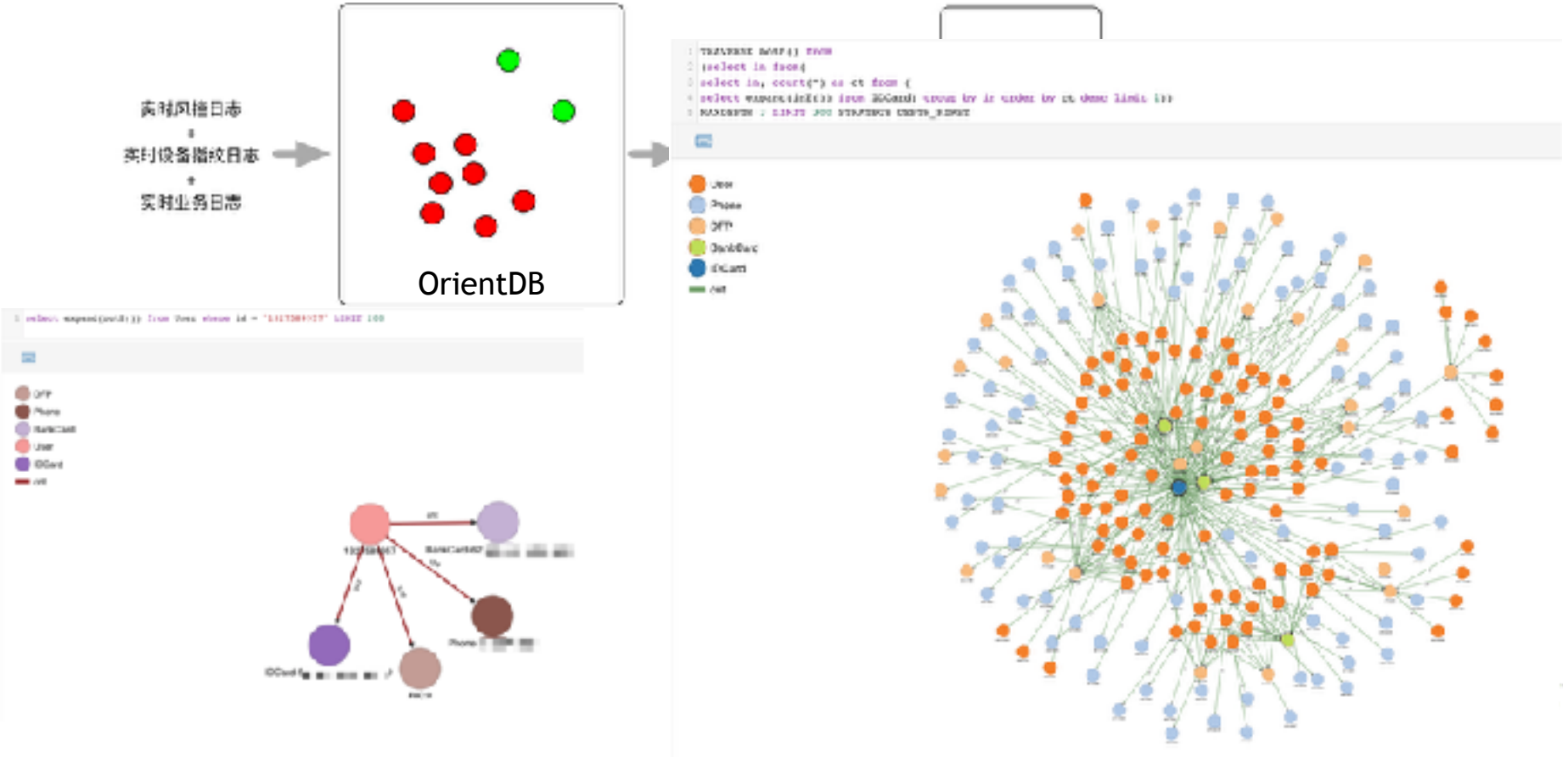
风控1.0	风控2.0	风控2.5	风控3.0
2015Q3-2016Q2	2016Q3-2017Q2	2017Q3-2018Q2	2018Q3-至今
风控上线：麦哲伦、哥伦布、郑和 黑名单与频次限制	近实时分析 离线分析 异常检测1.0 安全画像 评分卡 自研IP信誉库 简单ML模型 设备指纹落地	设备指纹反作弊 流量清洗 行为式反作弊 人机对抗 图片风格迁移 UEBA 奇流海 IOC（深度检测服务） ROC（运营中心）	反向爬虫验证 联防联控 设备信誉体系 账号信誉体系 策略自动驾驶 画像标注自动驾驶 关联关系与图计算 异常检测2.0

风控3.0技术

特征分析引擎(奇流海) - 多时间粒度+配置化

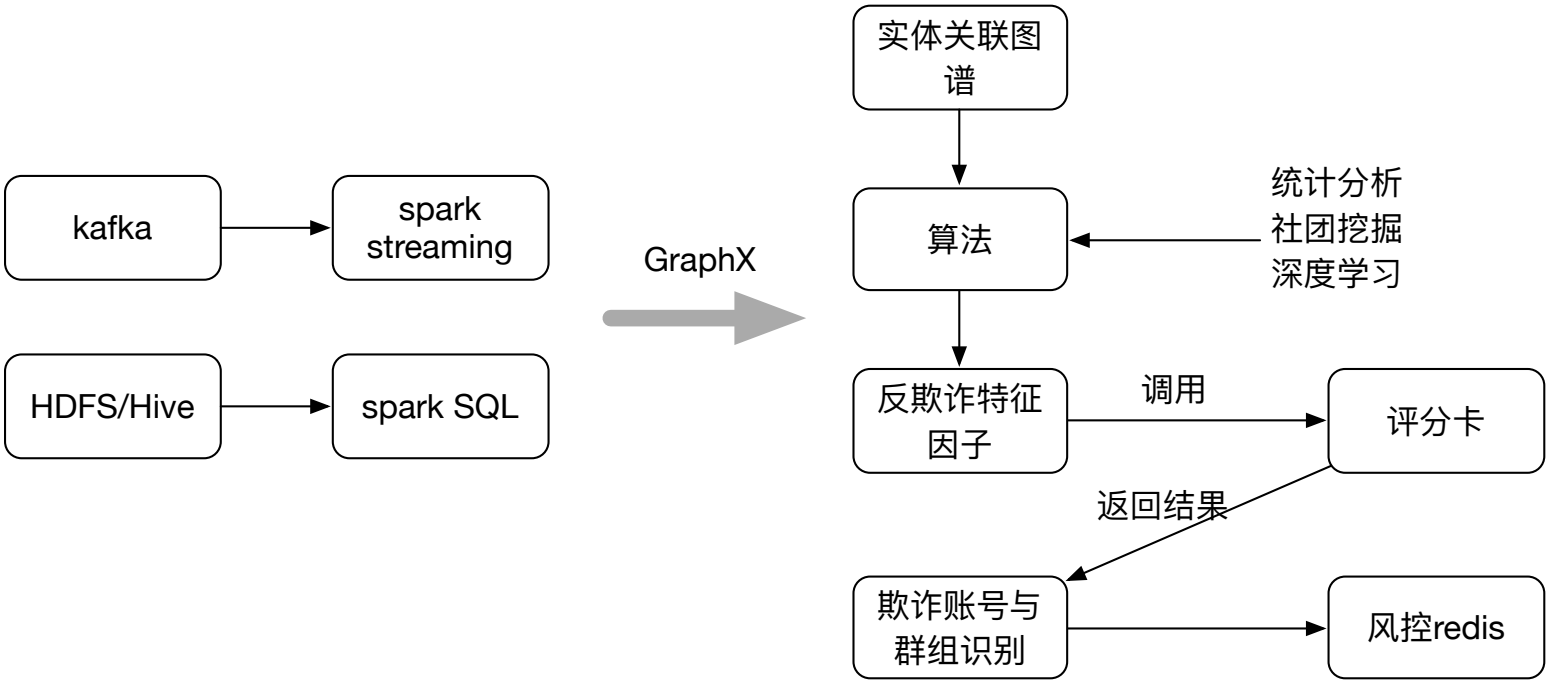


图计算引擎 - 探索分析+特征提取+策略配置



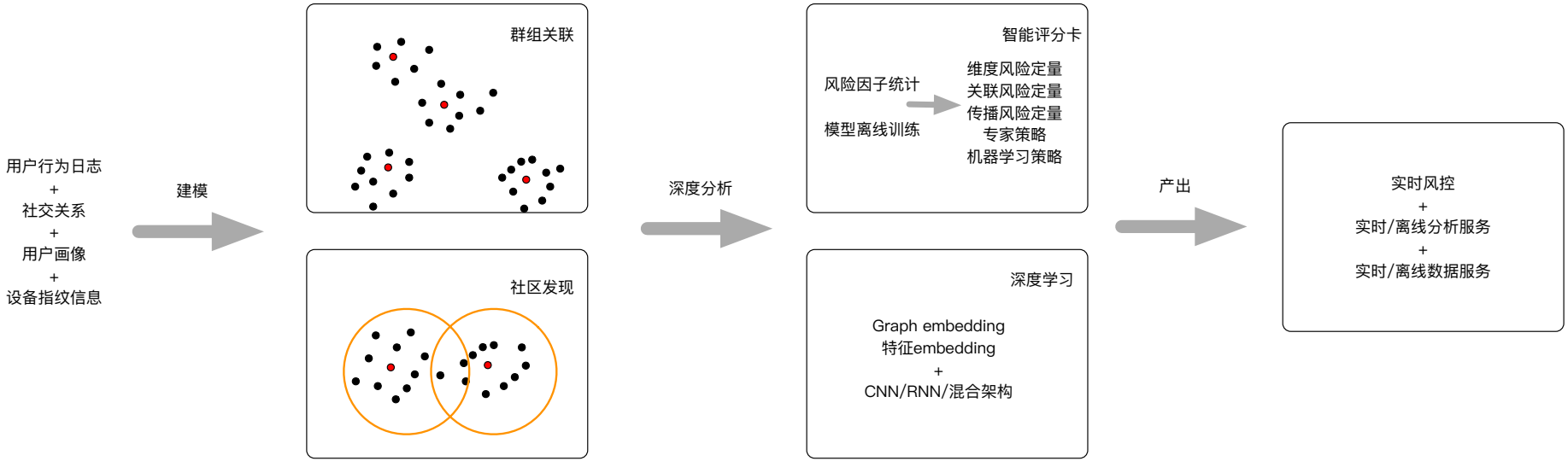
异常检测引擎2.0 - 多时间粒度图分析+自动化检测

工程化的图分析+异常检测



异常检测引擎2.0 - 示意图

采用**智能评分卡**技术与**深度学习(Graph Embedding)**技术完成“图To异常”的转换



实时关联分析

一对多惩罚	多级评分
风险扩散	一跳信誉传播+二跳信誉传播+……+N跳信誉传播
标签关联挖掘	一跳实体标签评分+二跳实体标签评分+……+N跳标签评分

Uid—>set{device_id1,device_id2,……}
device_id1—>set{uid1,uid2,……}

Uid—>{label1,label2,……}
device_id1—>set{label1,label2,……}

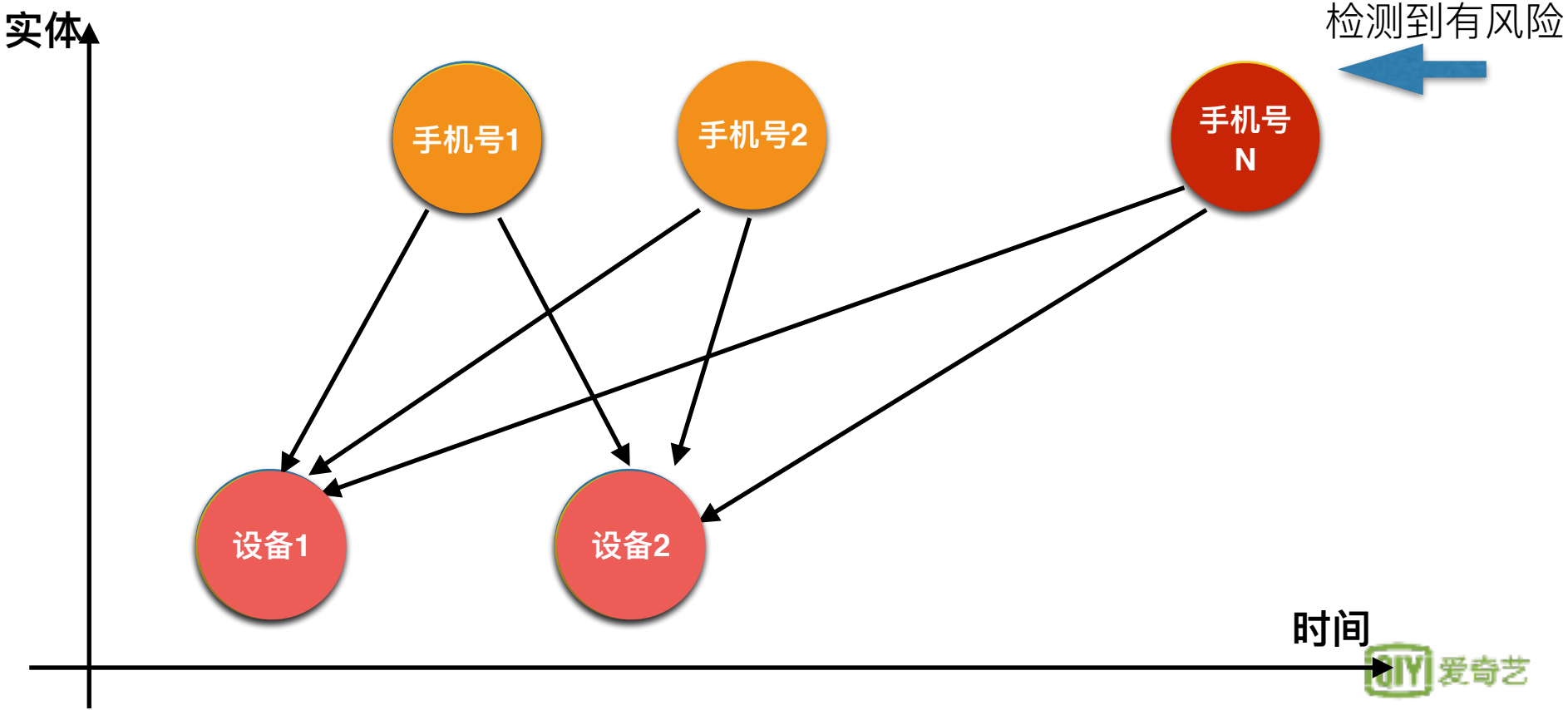
10亿k-v对

Couchbase

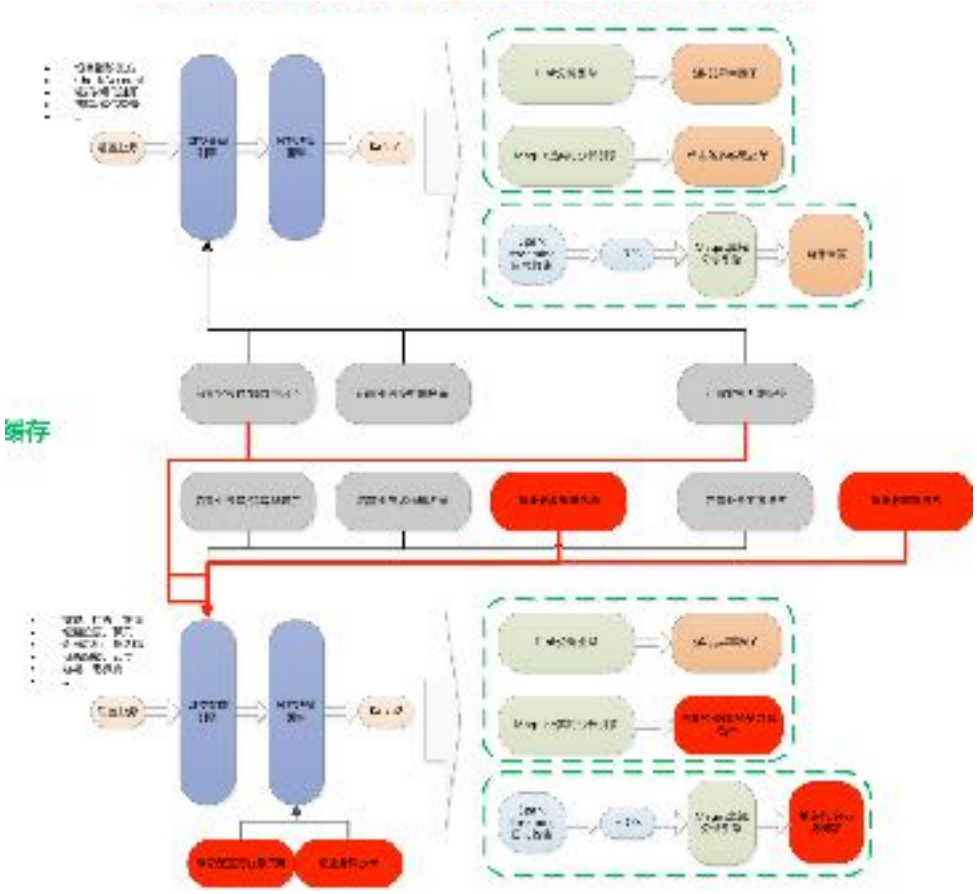
Couchbase5.0

爱奇艺

实时关联分析 - 示例



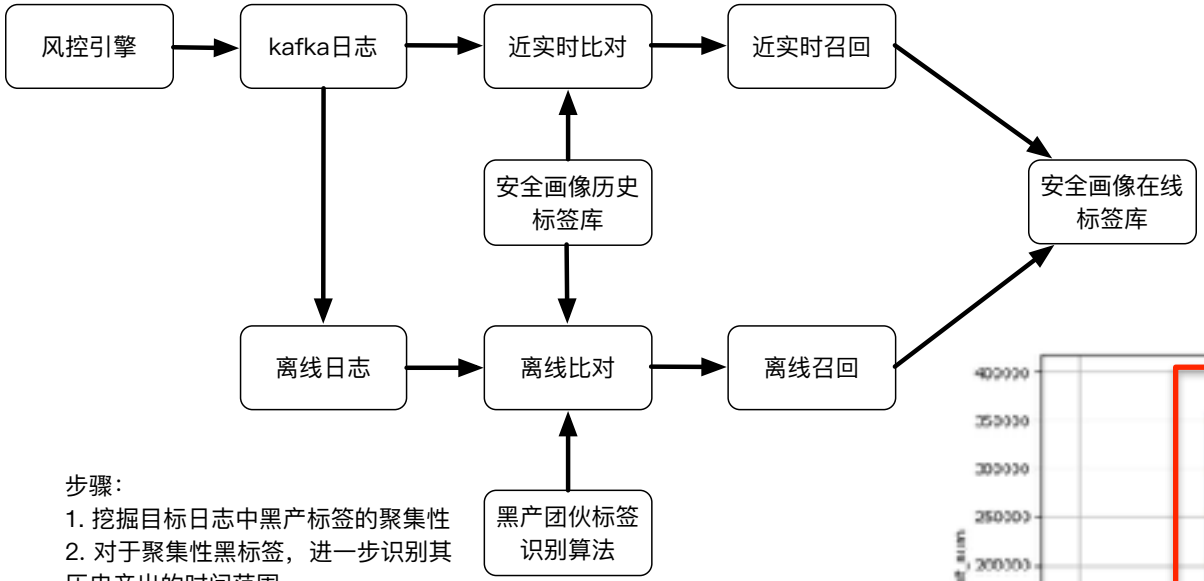
多业务联防联控



联防联控：前置为后置产出更加精准的跨业务黑名单与黑画像标签

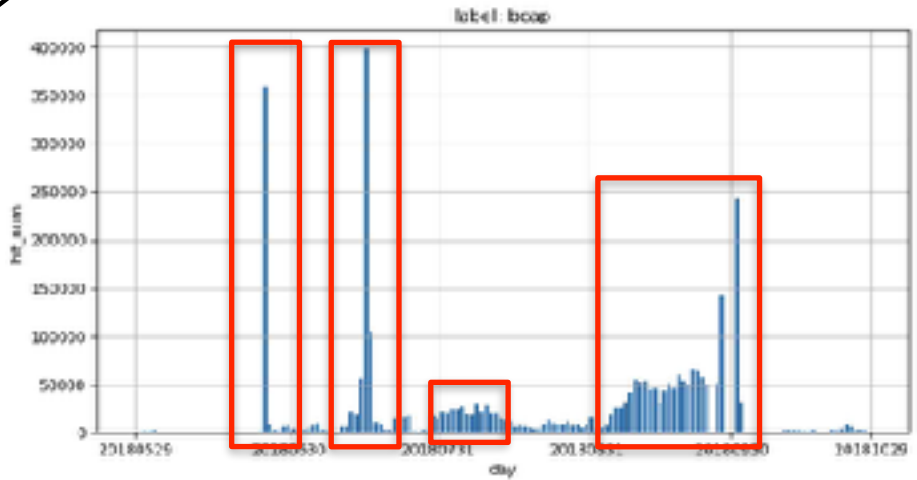
- 前置业务作用于后置业务，避免后置业务安全分析冷启动
- 多个前置业务弱因子，提升目标业务风险监测水平
- 较安全画像标签，实时性更好

自动驾驶 - 以标签自动召回为例

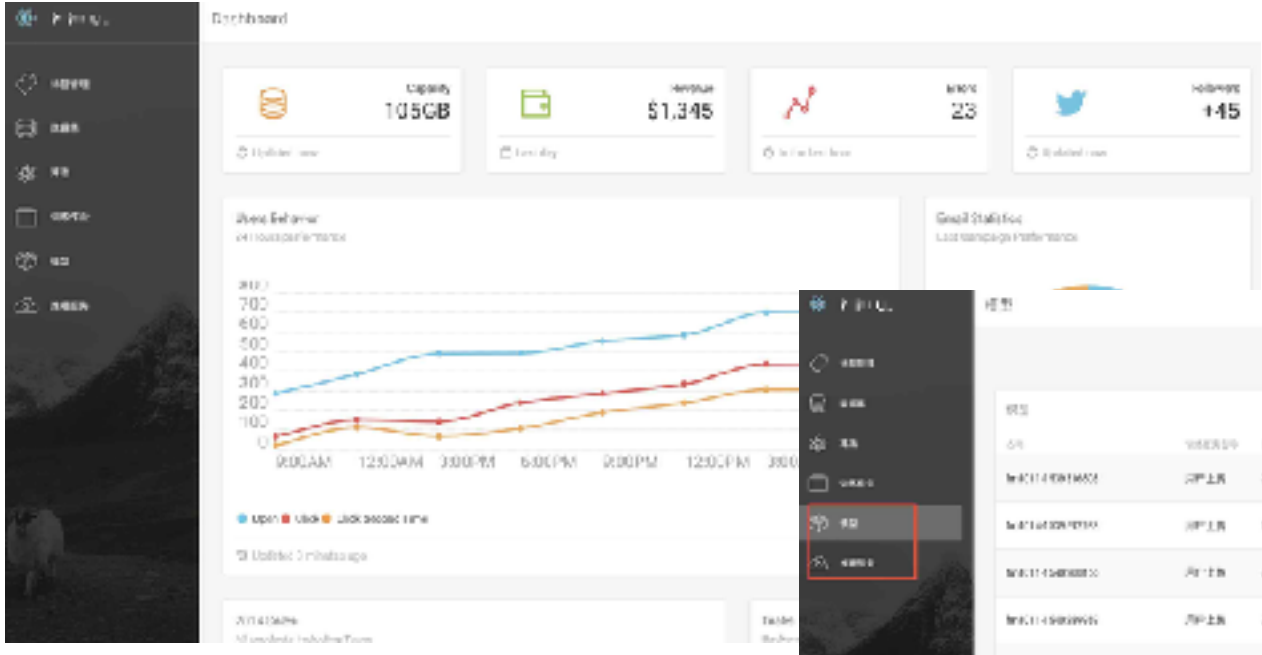


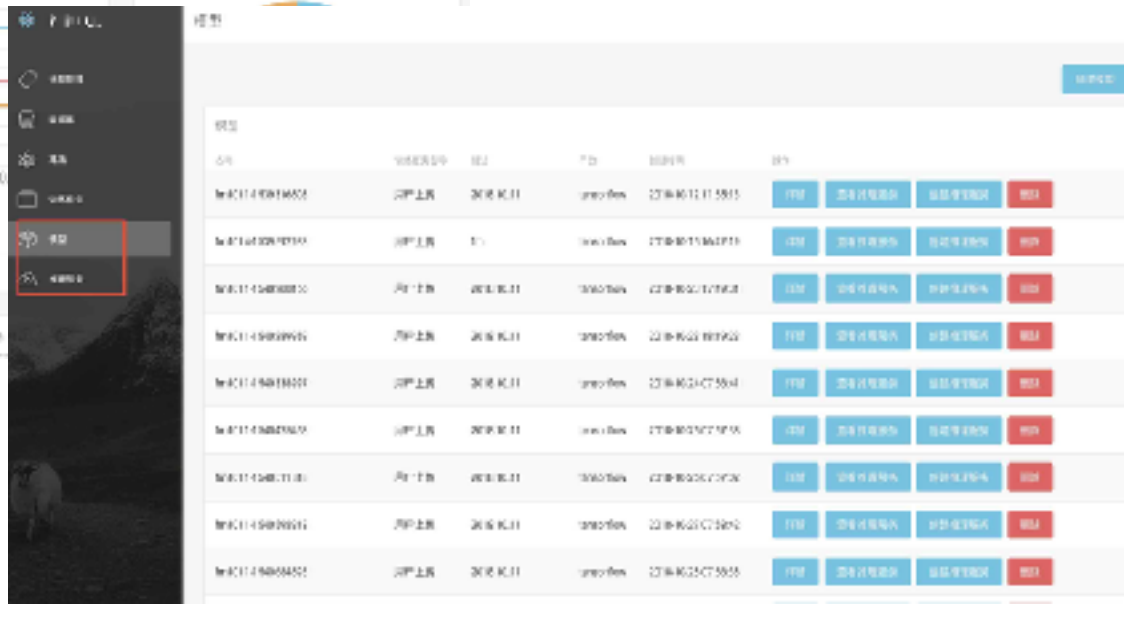
步骤：
1. 挖掘目标日志中黑产标签的聚集性
2. 对于聚集性黑标签，进一步识别其历史产出的时间范围

特点：
1. 近实时比对历史标签库，提升近实时精细化召回
2. 离线比对历史标签库，提升黑产批量/团伙识别，**大范围召回**



通用机器学习平台 - 支持训练、推理与部署

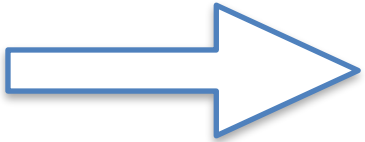


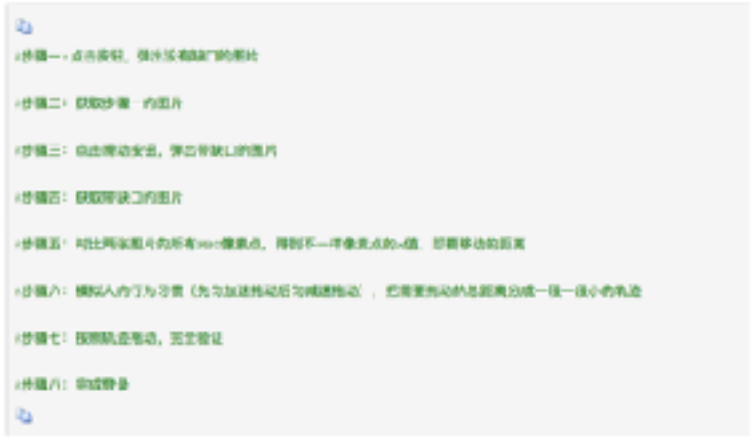


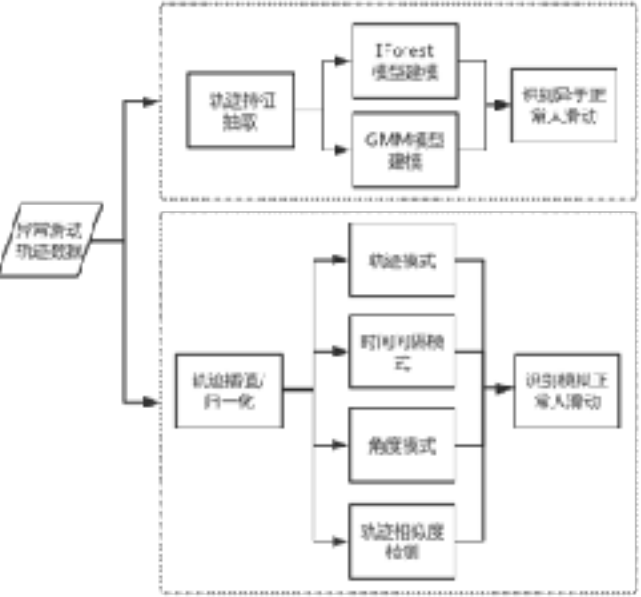
实时攻防对抗案例

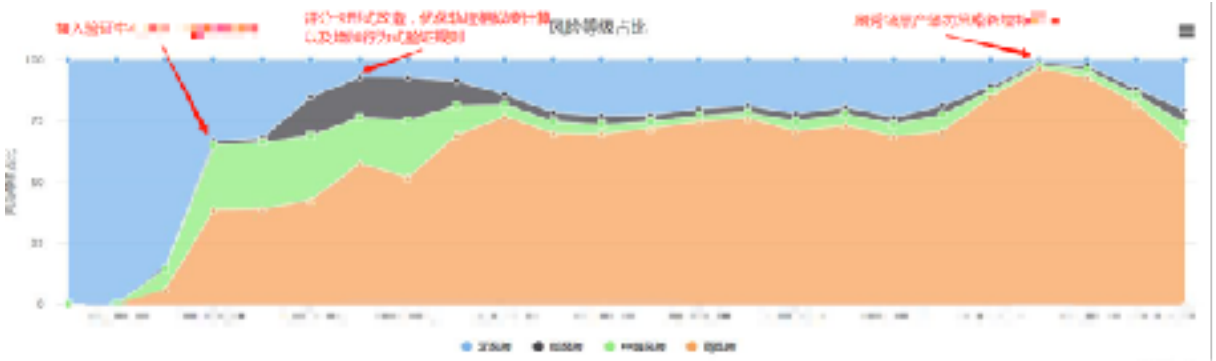
人机对抗

破解的固定套路









UGC上传实时反作弊 - 获利模式

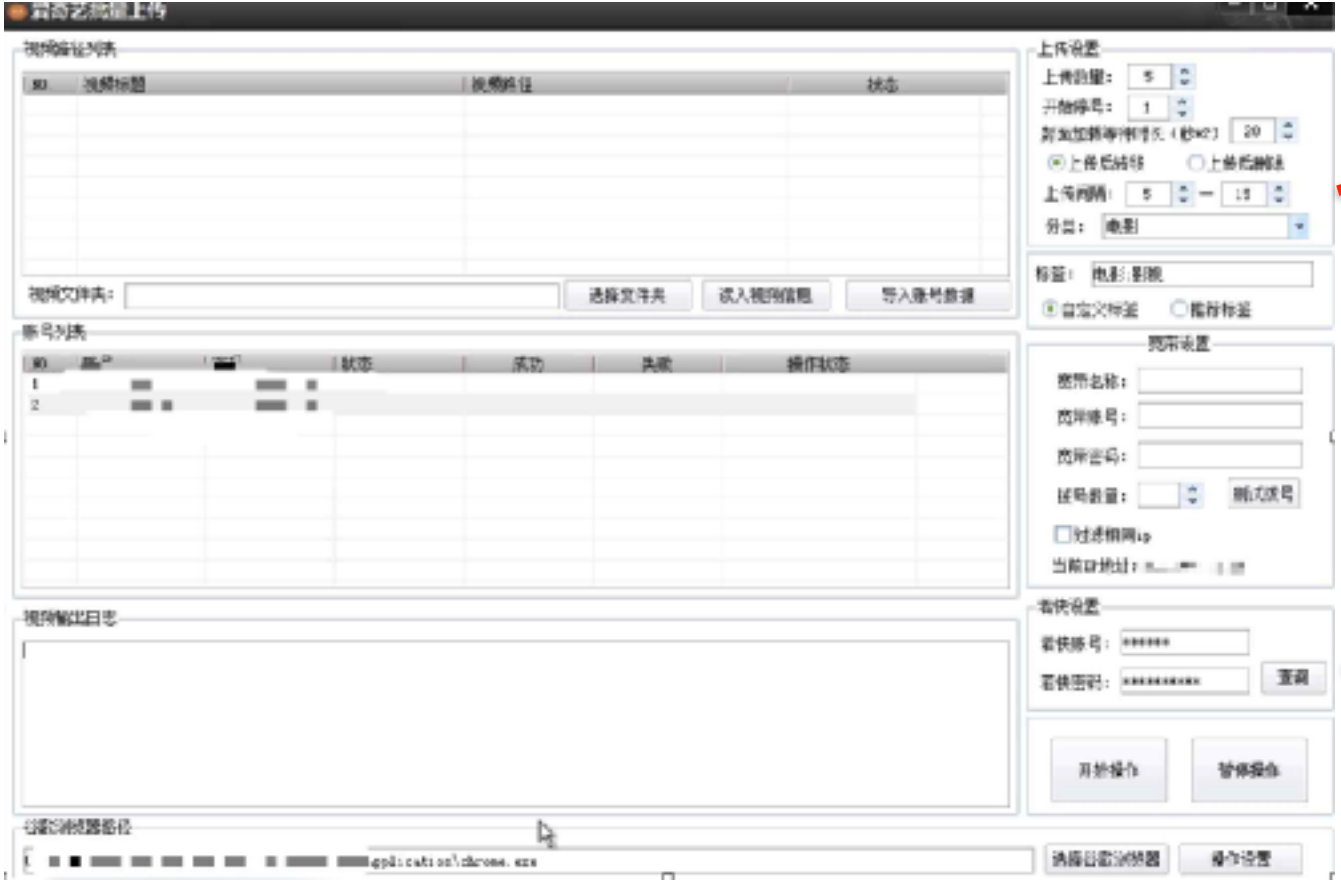
分成模式：

初始等级一，一般发完5篇视频后会升级；

三级以上就会有广告分成收益；

每个等级间有不同的权限，等级越高，收入来源越多

UGC上传实时反作弊 - 黑产软件

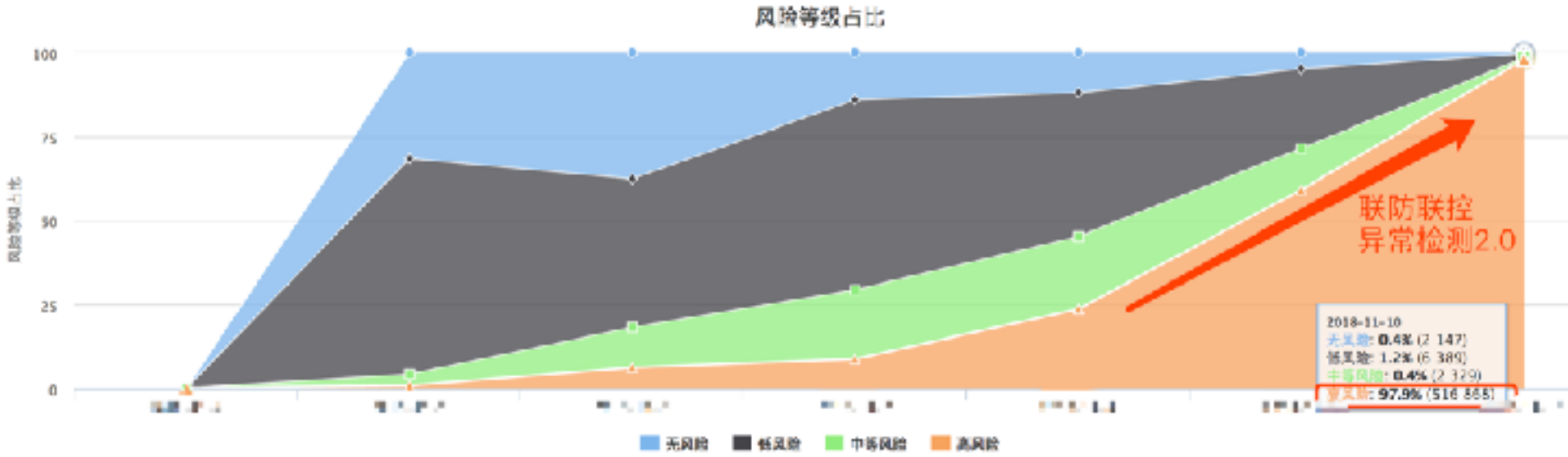


控制批量上传参数

控制拨号IP

外包验证码破解

UGC上传实时反作弊 - 技术对抗

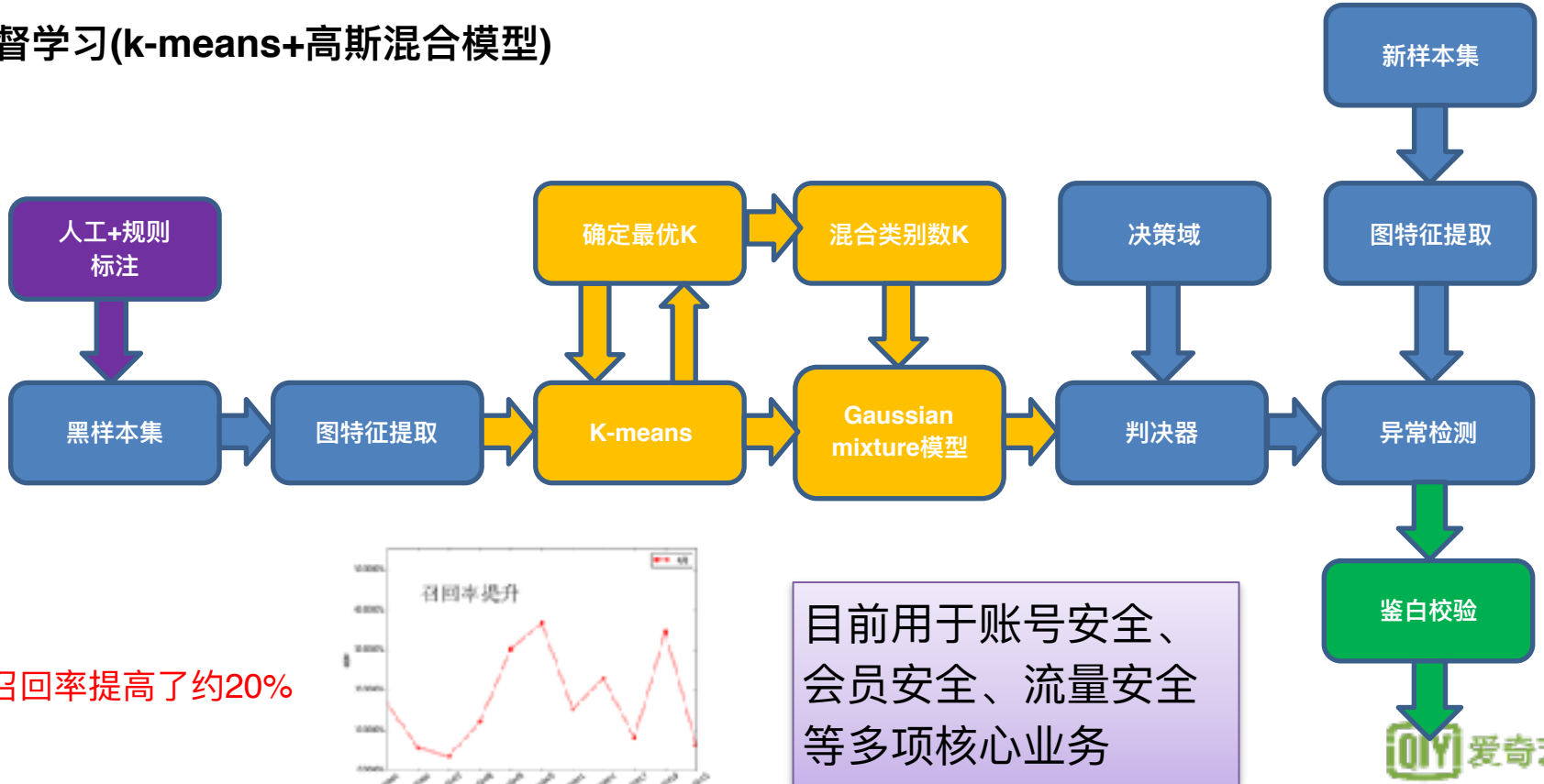


根据黑产软件特点，制定**联防联控**的整体解决方案，采用**Teemo+异常检测2.0**完成对黑产物料的实时识别，已将黑产恶意上传的拦截率提升至97.9%+，大大节省了审核资源。

爱奇艺

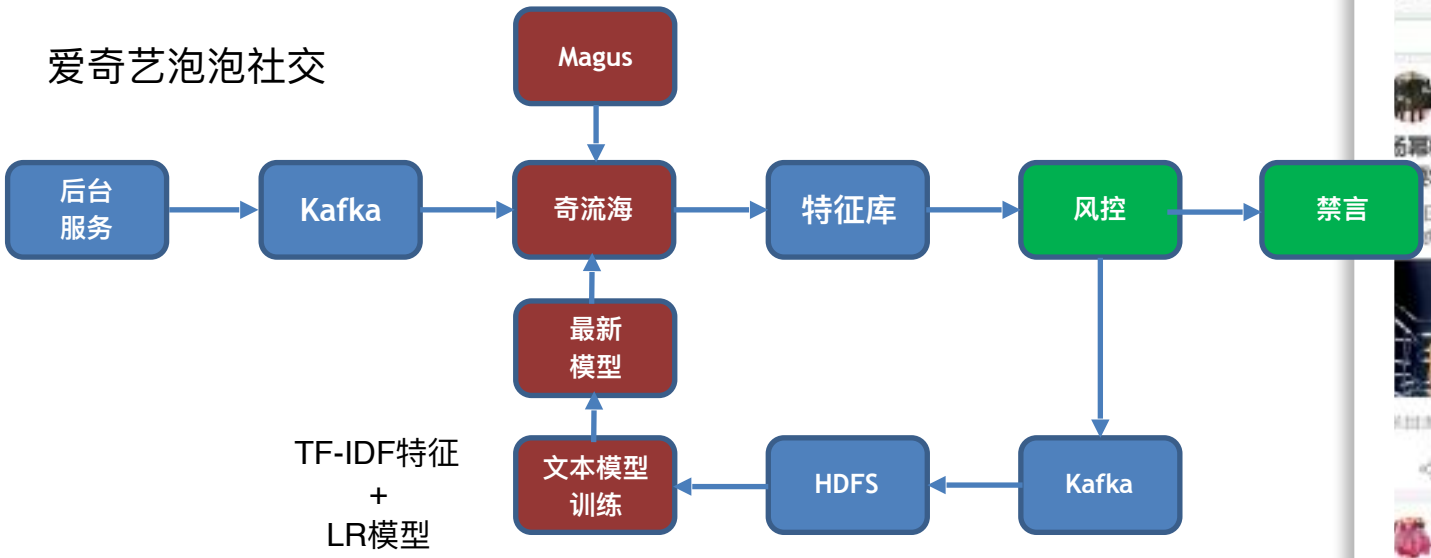
反批量注册

半监督学习(k-means+高斯混合模型)



社交反垃圾

爱奇艺泡泡社交



Highlight：

1. 通过多时间粒度的实时学习，生成规则进行实时拦截；
2. 解决文本反垃圾没有标注，以及垃圾文本动态变化的问题；
3. 通过文本反垃圾模型提升召回率35%(低频、冷启动等)；

爱奇艺

iQIYI 爱奇艺