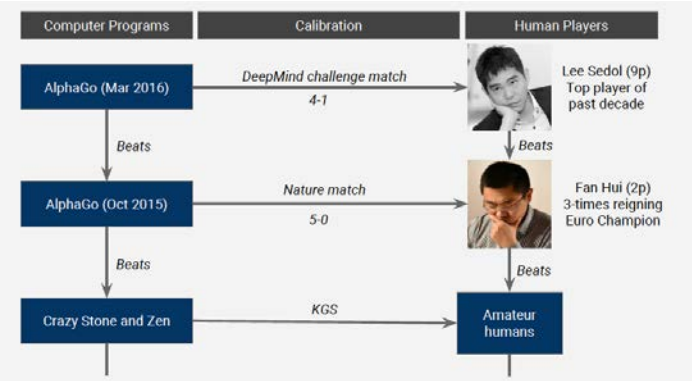
AI大数据时代电商攻防：AI对抗AI

苏志刚, Ph.D.

Head of JD Security Research Center
JD.COM

AlphaGo: 战胜世界冠军



Source: David Silver

深度学习的广泛应用

pcmag.com

theverge.com

攻击的规模和复杂性都在增加

账号安全: AI Empowered

JD.COM 京东

APP 安全 SDK

设备指纹

生物探针

注册和登陆环节的人机识别

异常检测

行为分析

黑产情报

订单风控

业务验证和反馈

欺诈行为

JD.COM 京东

- 机注
- 刷单
- 黄牛
- 代购和刷券
- 恶意订单
- 虚假评论

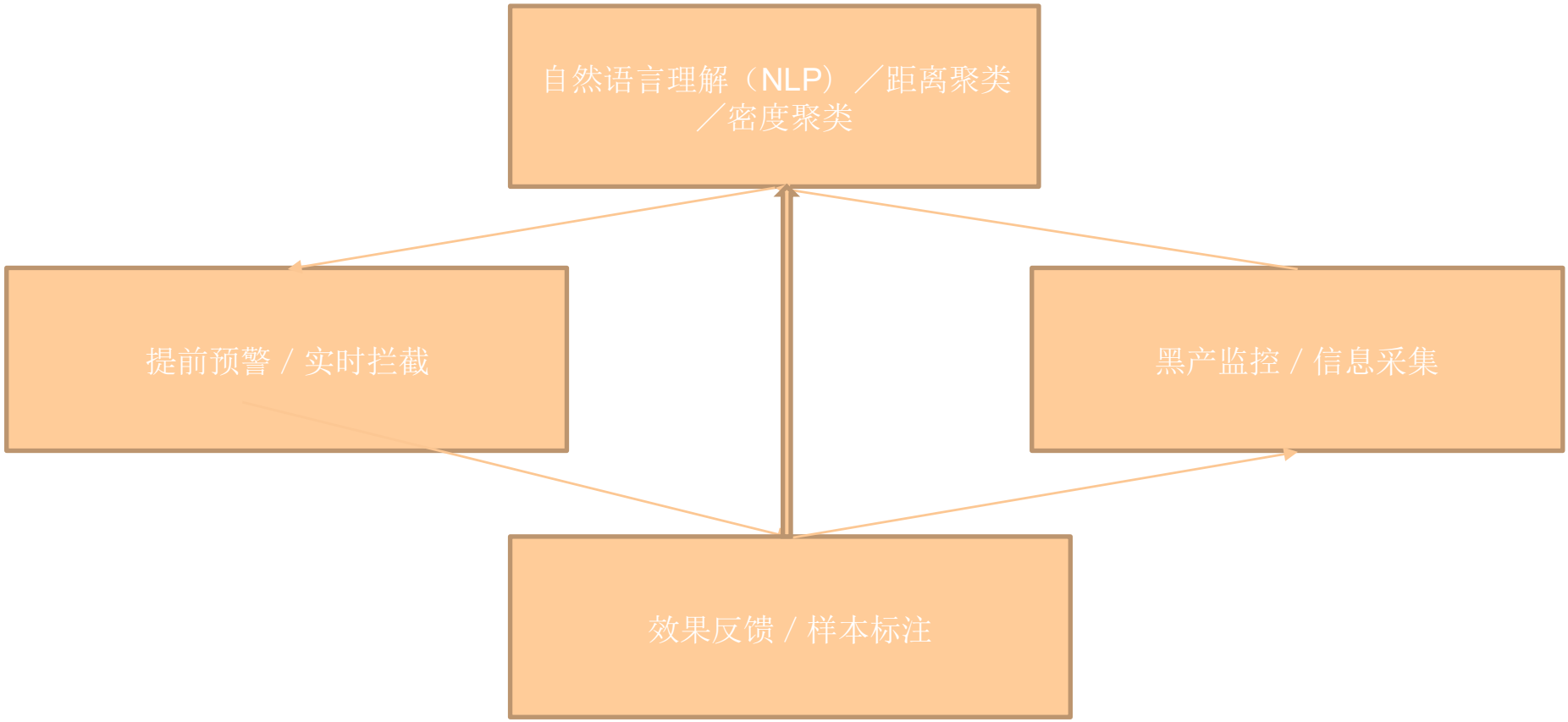
黄牛行为

攻：爬虫　　QQ群传递消息　　黑产软件登录　　打码平台图像识别

守：反爬虫　　NLP黑产活动监控　　黑产软件逆向　　高对抗性样本　　地址聚类

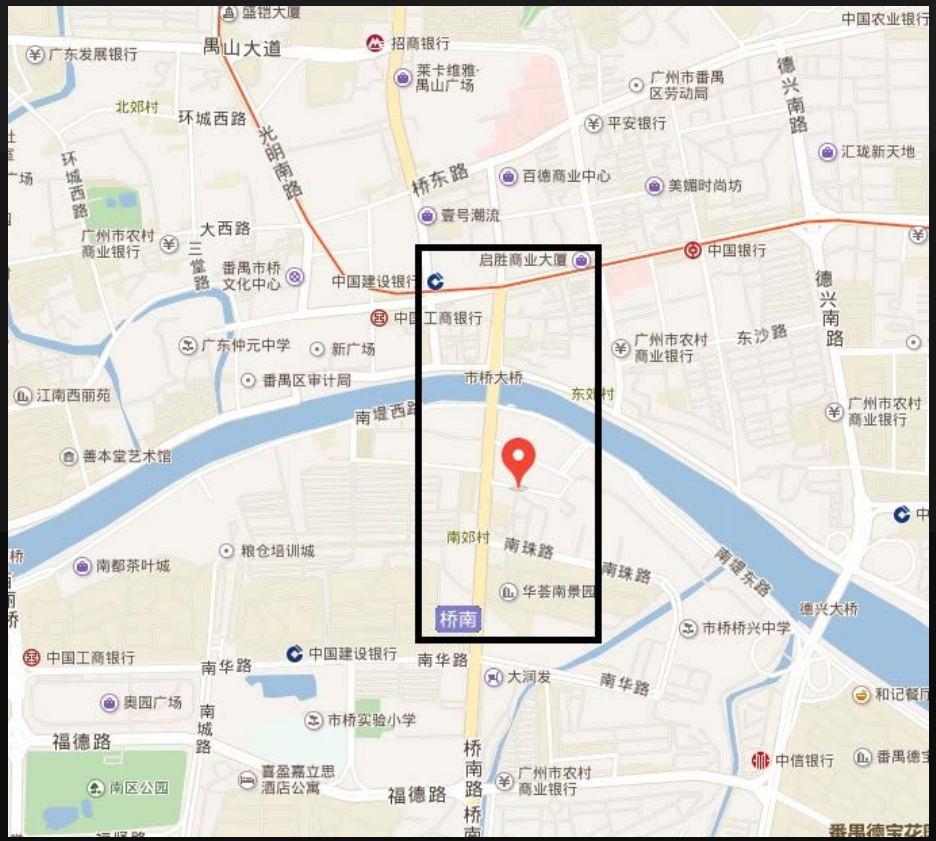
电商与黄牛的对抗

JD.COM 京东



JD.COM 京东

黄牛手机订单拦截



机注

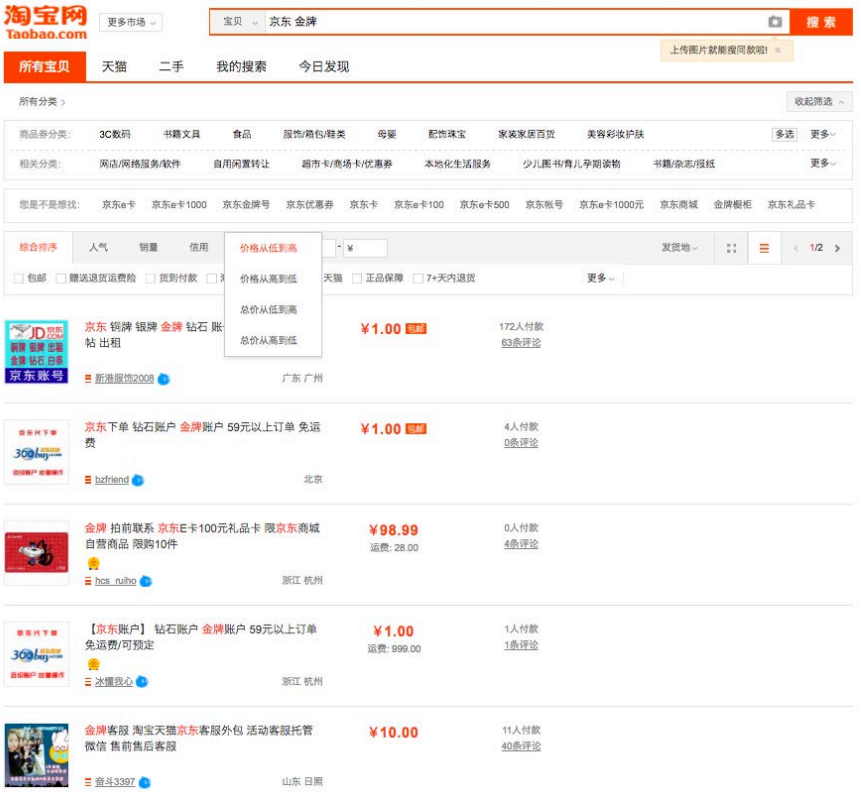
JD.COM 京东

- 成熟的工具：火牛和刀锋
- 黑灰产供应链
  - SIM卡
  - 解码平台
  - 短信验证服务
  - 身份证
- 检查特征
  - 机器行为
  - 虚假信息

账号销售

JD.COM 京东

- 账号分类销售
- 销售平台
  - 友商零售平台
  - 社区 IM (QQ，微信)
  - 批发网站



刷单

JD.COM 京东

- 利用虚假订单来提高第三方商家排名
- 刷单成本: 账号, 代购和物流.
- 组织行为
  - 代购仿照普通用户行为
  - 多个账号下单，但公用有限的支付账号。类似的收货地址
  - 部分使用虚假快递追踪编号

恶意订单

JD.COM 京东

- 第三方厂商之间的恶意竞争：破坏对手的促销活动
  - 打击第三方对手的库存量
  - DDOS 正常用户
  - 浪费物流资源
- 特征
  - 针对某个商家和产品
  - 货到付款

机注是万恶之源

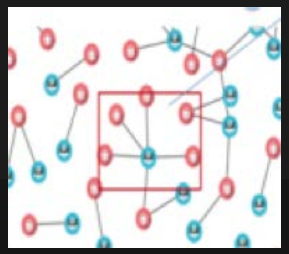
JD.COM 京东

- 批量机注账号
- 检查机注
  - 黑产情报：深挖黑产工具和策略
  - 无监督学习
  - 有监督学习
- 挑战: 低误判率

JD.COM 京东

打击机注

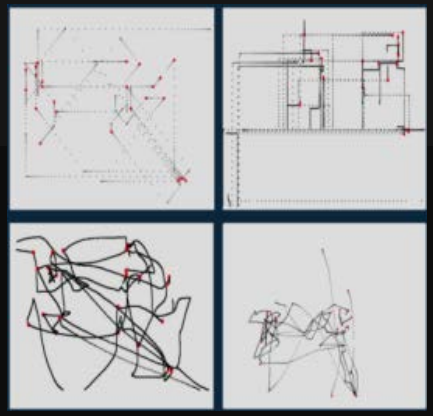
Device Info　User Behavior　Geography　Frequency

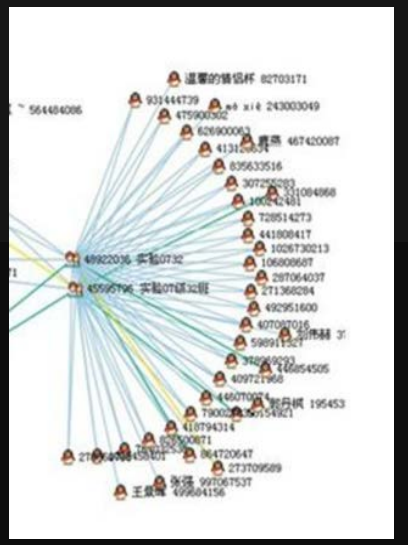


设备分析

无监督学习

行为分析



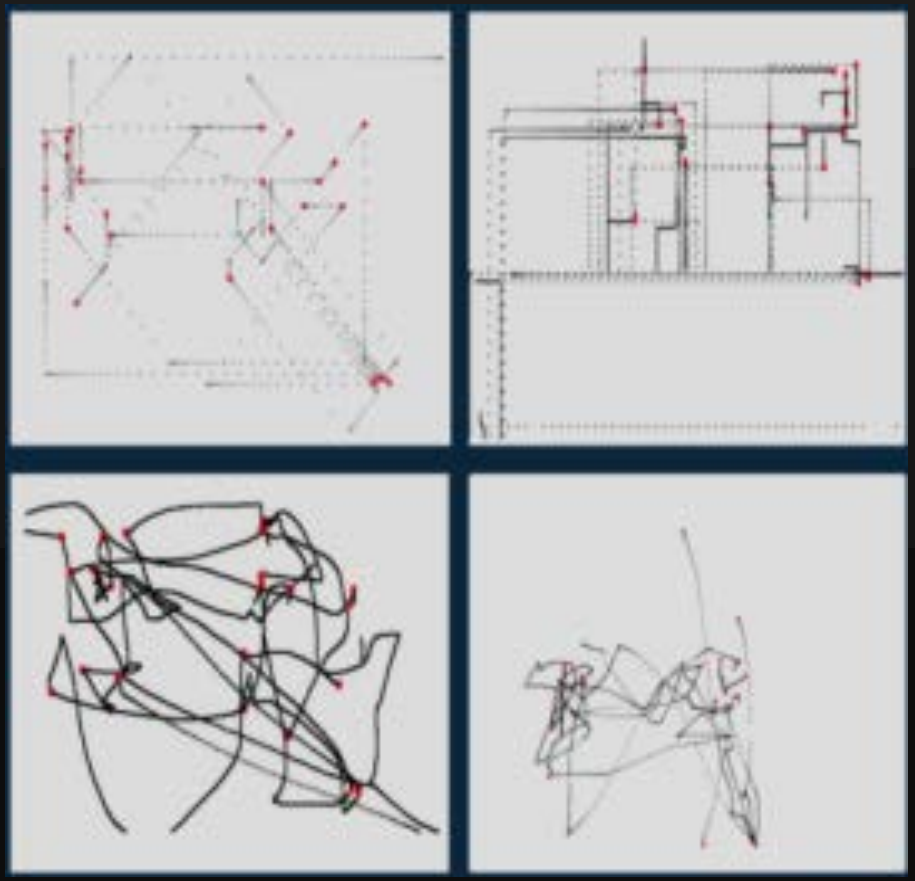


人机识别

- 京东有丰富的人机识别场景
  - 机注
  - 机器下单
  - 爬虫
- 使用键盘和鼠标的行为作为人机识别特征

POP QUIZ: 人工人机识别

人机识别

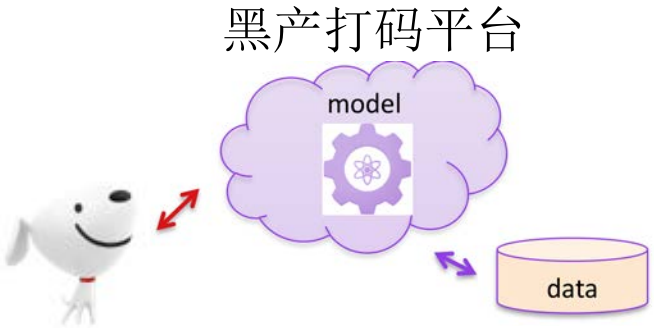


图片验证码对抗

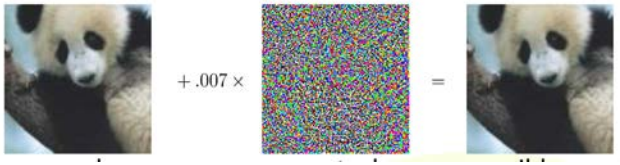
JD.COM 京东

AI的对抗与反制

模型提取：
使用尽可能少的查询去学习近似模型



使用对抗样本Adversarial example：
修改输入去击垮机器学习

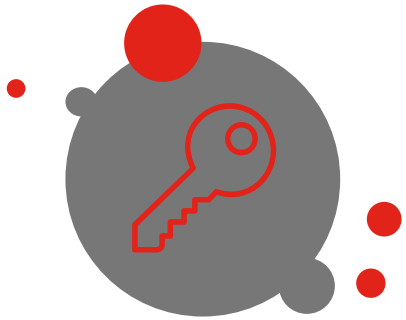


熊猫　　长臂猿

京东智慧物流系统安全

JD.COM 京东

安全挑战多源于恶意攻击(例如Mirai蠕虫攻击)
总结安全挑战类型可分为如下两类：



系统级别安全

设备终端　云端　通讯协议

人工智能安全

针对于AI系统本身的攻击和
相关安全性问题

无人仓

设备安全：**AGV自主路径规划**

系统安全：单一设备错误造成系统紊乱

无人机

人身安全：**降落环境识别**　**人体识别**

设备安全：**自主避让**　空中交通关系　**自主规划路径**

数据安全：GPS劫持　数据传输劫持

无人车

设备安全：车辆劫持

货物安全：被盗　**伪造客户（骗过人脸识别）**

数据安全：数据传输劫持

人身安全：**道路识别**　**人体识别**

无人超市

资产安全：伪造货物　**伪造用户（骗过人脸识别）**

隐私安全：用户数据　订单数据　人脸数据

京东

AI安全问题的后果

JD.COM 京东

- 无人车：无法正常识别路况，撞击其他人或正常行驶的交通工具；
- 无人机：自主规划路线违背规则，出现安全隐患；
- 无人仓：AGV协作出现问题，彼此碰撞损毁；损失财产和延误货物运送；
- 无人超市：无法正确识别购物人所购物品，并形成关联，造成财务损失；

不能按照既定功能运行

初级

- 无人车：被攻击者控制，变为工具，造成货物丢失；
- 无人机：被攻击者控制，变为工具，造成货物丢失或者更大危害；
- 无人仓：AGV出现故意扰乱其他AGV行走的状况，扰乱仓库秩序；
- 无人超市：出现货物被盗窃或者支付费用被转嫁其他购物者情形；用户数据被盗取训练；

产生攻击者想要的结果

中级

- 在特定时刻实现对其他系统的大规模毁坏性攻击；
- 消耗本系统所有资源，并导致其他系统瘫痪；

被潜伏和利用作为肉鸡网络

高级

京东

AI 系统安全

JD.COM 京东

京东信息安全部

京东

总结

JD.COM 京东

- 京东安全有广泛的AI应用场景
- AI应用高效运作的唯一途径
- 电商和黑产之间持续的高对抗性攻防