



TalkingData

TD收集 / 处理的各类数据

标准标签 UGC标签 模型探索与生产 安全合规的数据输出 原子标签 人口/设备 网络行为 位置相关 购物/消费 App相关 Wifi行为相关 属性 移动端数据采集 互联网数据采集 合作伙伴数据采集

32T/夫



≈每天700部 4K高清电影



这么多的数据,能干嘛?



标签1.0:人工标签



- 标签要怎么打?
- 1)性别+
- 2) 各类App:
 - 贝贝(母婴特卖)
 - 小时光
 - 小伴龙
 - 宝宝学颜色
 - 故事口袋读读,等等



标签1.0:人工标签



- 一共58个不同的App
- 历时2个礼拜



标签2.0:人工智能化标签







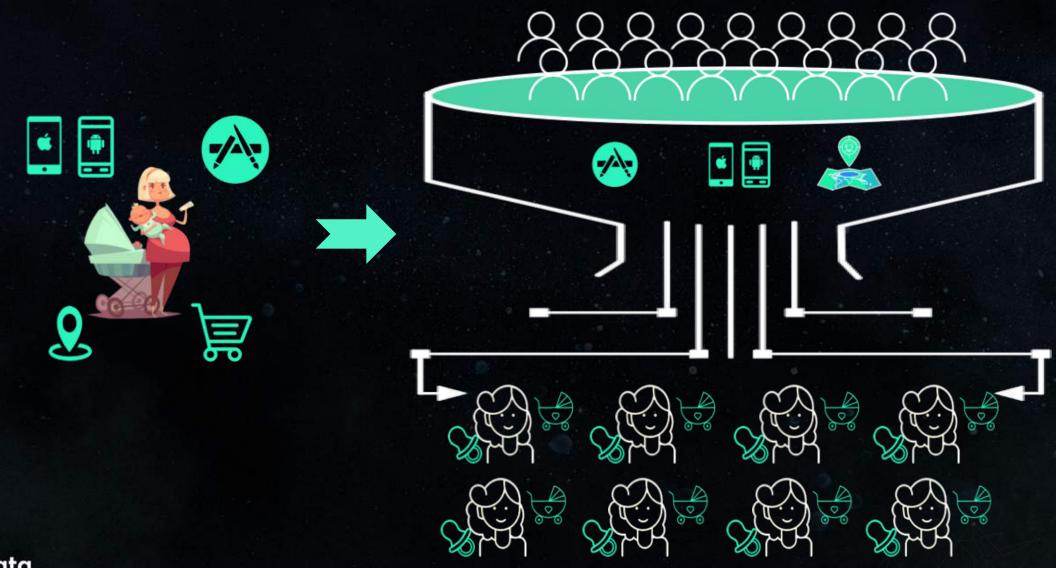








标签2.0:人工智能标签



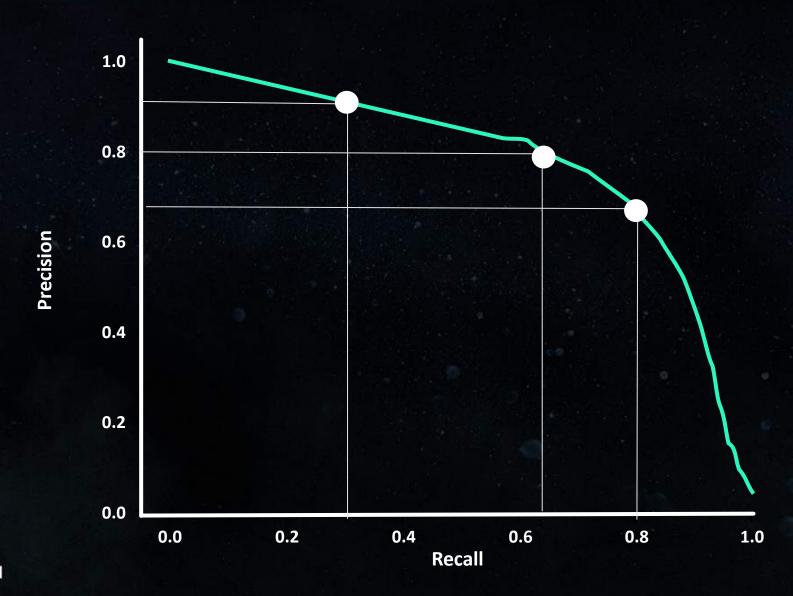


标签2.0:





标签2.0:数量和质量之间的定制化平衡





数量与质量的定制化平衡





标签2.0带来新的洞察







标签2.0带来新的洞察









规则为主 = 》需要专业知识 (Domain Knowledge)

好处:

- 对数据依赖小,适合冷启动
- 准确度比较好

缺点:

- 耗时长 = 》通用标签为主 = 》定制化能力弱
- 覆盖面有限,容易漏=》最害怕的就是做逻辑交集

标签2.0

数据方法,循环产生标签=》依赖数据

- 覆盖面大大增强,同时可以按需在数量和质量之间取得平衡
- 对专业知识的依赖大大降低,甚至可以不依赖
- 定制化的能力大大增强
- 速度快



标签2.0的要求

- 结果的可解释性不能保证
- · 需要对结果做验证及检查 = 》Tracking以及监控的重要性







Know who





知道对你来说最有价值的用户群





未来的价值才是最重要的价值





如何预测未来?





如何预测未来?

很多重要行为在一方数据里是缺失的



VS









价值的主体是人





各类场景下的标示归一化

同一设备 , 不同标识的归一化

不同类型标识之间的归一化: 手机相关标识, 金融支付标识, 各类帐号之间

不同设备, 同一人的归一化

同一家庭,同一栋楼宇 不同人之间的归一化



Know when





对用户场景有了解 根据场景做相应的触达









Know what

(Action to take)





根据 Who+When, 做最相关的行动(Action)





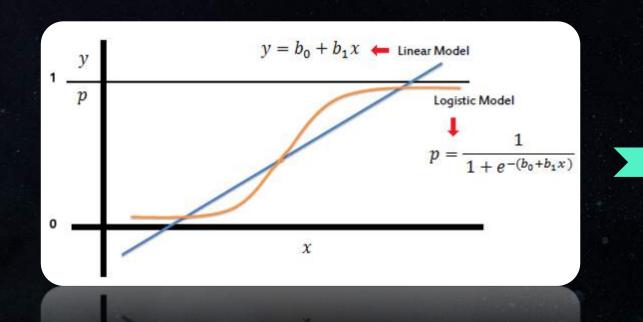
点击预测,广告行业里的经典Action场景

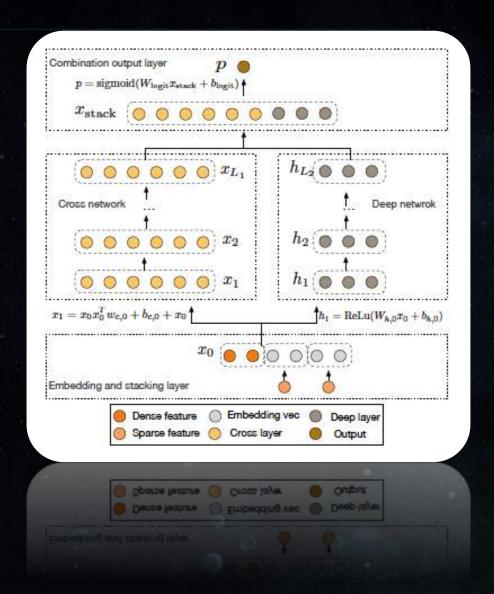






从逻辑回归到神经网络







智能时代的新重点

什么是未来 可能比较火的场景?







智能时代的新重点:反馈(Feedback)+调整(Adjustment)



