

如何做好推荐系统？

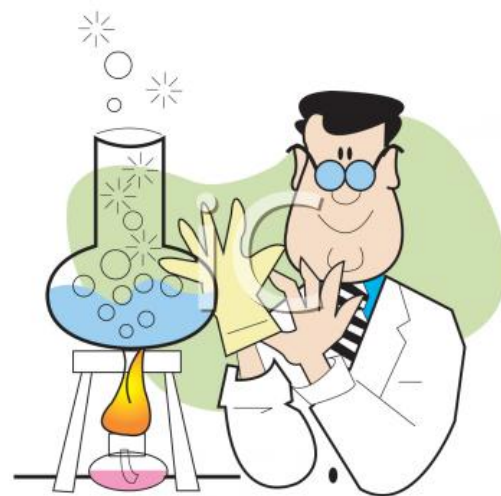
报告人：冯沁原

导师：代亚非

Qinyuan.feng@gmail.com

学术背景

- 2006.3-2008.9
 - 虚假投票反作弊
- 2008.10-2010.10
 - 个性化商品推荐



工业背景

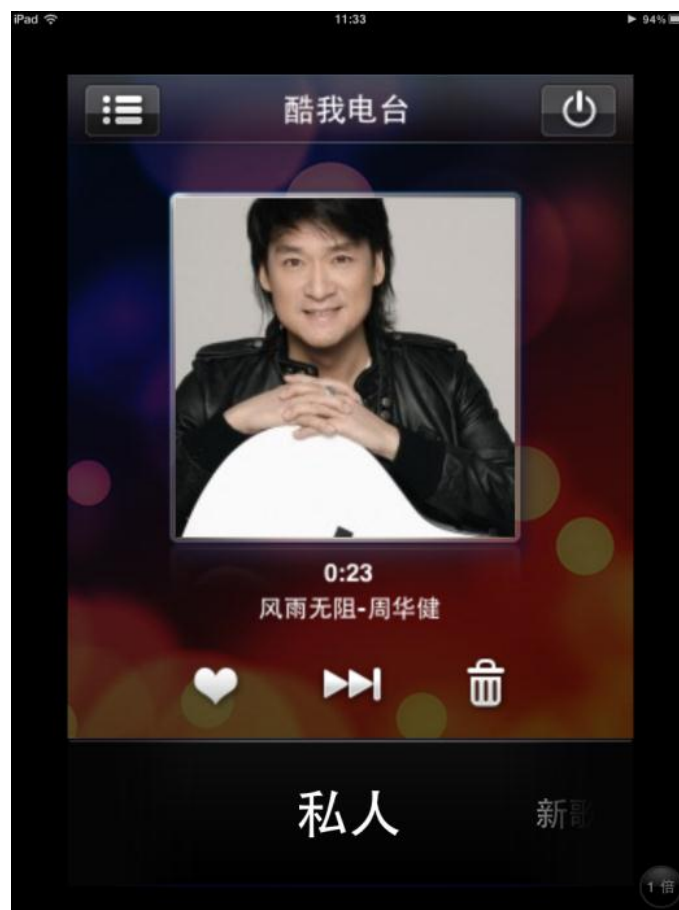
- 2010.11至今

– @  酷我音乐盒

– 设计了通用推荐服务平台

- 为每日千万在线用户
- 提供实时音乐和广告推荐

音乐推荐： 遇见你最喜欢的那首歌



广告推荐： 激发用户的参与

亲爱的用户，您好：

关闭



今日推荐活动：
劫商城！抢NPC！
《魔域》疯狂暑期，
火爆进行中！



点击体验

《魔域》暑期档

【注册抽话费、q币】

推荐系统的终极目标

- 在恰当的时候
- 把恰当的物品
- 用恰当的方式
- 提供给恰当的人



为什么那么多牛人研究了十几年算法
推荐系统还是不能令人满意呢？

从产品的角度看问题

Lesson1: 决策不要拍脑袋

- Say NO to HIPPO
 - Highest Influencing Person's Personal Opinion
- Say YES to TRUTH based on DATA

- 统计你的河马指数☺
=每天拍脑袋做决定的次数



Lesson2: 指标为用户服务

- 避免用户为指标服务
- 一个总目标
 - 易懂、全面、实用
- 一批子目标
 - 简约、专注、相关



Lesson3: 推荐 **不** 需要个性化

- 也许只需要做好分类
- 也许只需要最好的十个商品
 - 用户能够消化的商品数量有限
 - 存在一批绝大多数人都会喜欢的商品
- 即使需要个性化
- 也许简单的相似性算法就足够了

Lesson4: 平均是陷阱

	平均值	阈值比例
生活水平指标	北京人平均收入	收入大于1000的北京人比例

1万元!

1%

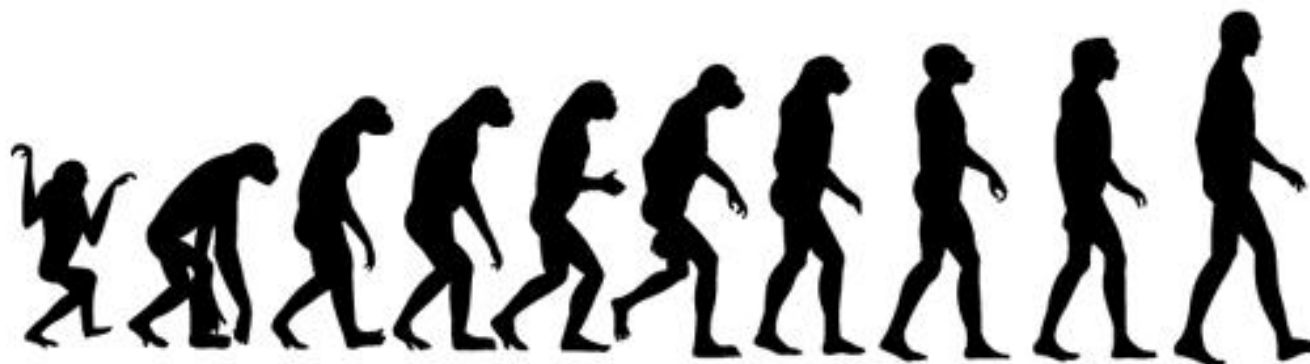
推荐系统指标	用户平均收藏的歌曲数	收藏歌曲超过5首的用户比例
--------	------------	---------------

1个北京人收入一百万

99个北京人收入一百元

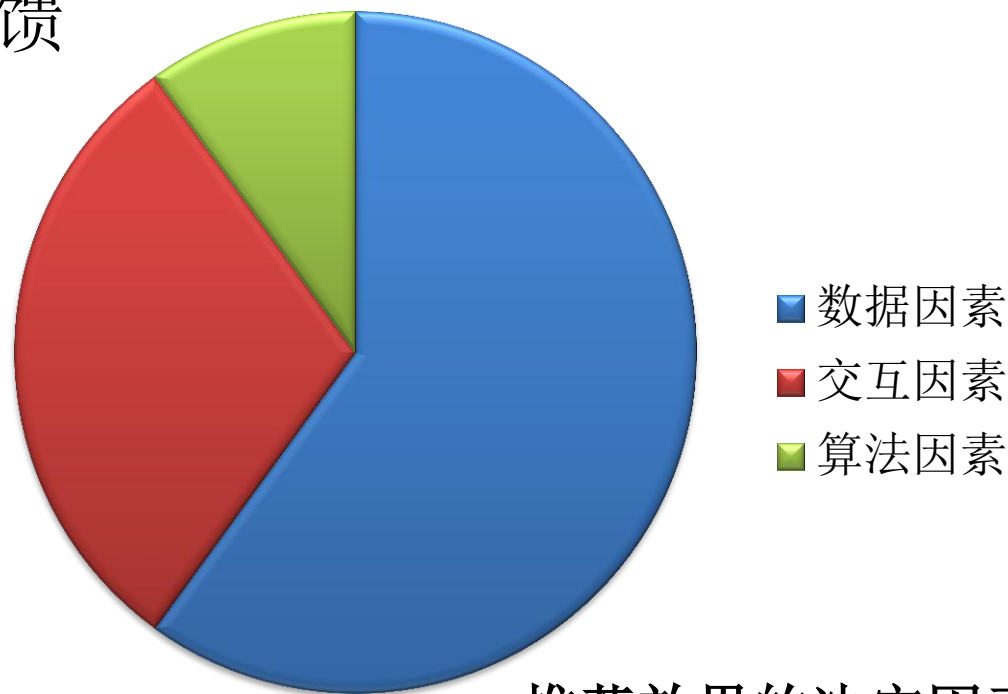
Lesson5: 精与惊

- 新用户
 - 推荐精品商品
- 老用户
 - 持续给予惊喜



Lesson6: 六成是数据

- 数据决定了60%的推荐效果
 - 宁缺勿滥
 - 引导用户主动反馈
 - 过滤用户虚假反馈



推荐效果的决定因素

Lesson7: 三成是交互

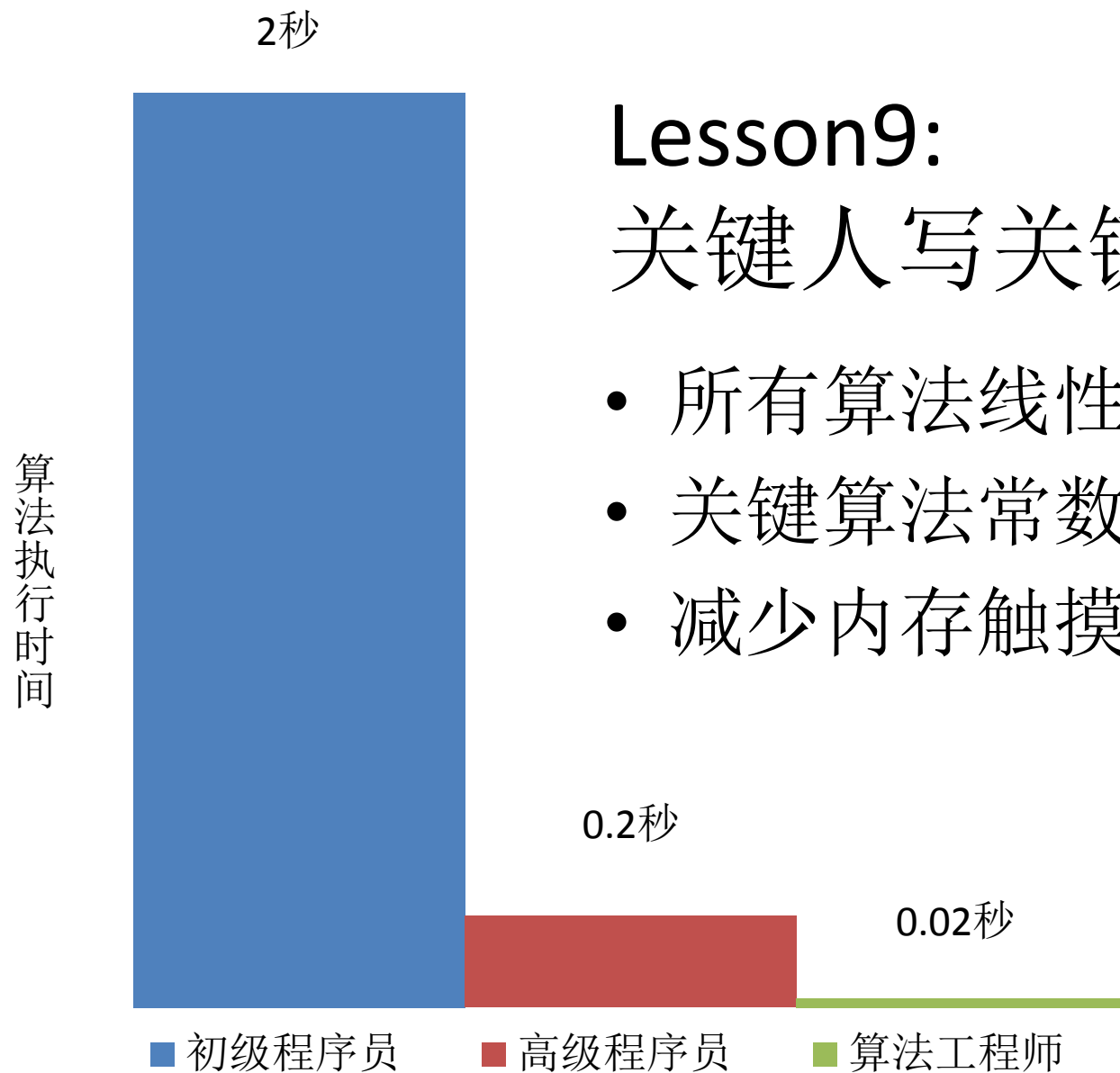
- 交互体验决定了30%的推荐效果
 - 不让用户思考
 - 简约清晰的设计
 - 引导用户思考
 - 给用户解释
 - 给用户20%的熟悉音乐




$$E=mc^2$$

Lesson8: 速度就是质量

- 一秒钟更新一个用户口味
- 一分钟部署一个新算法
- 一天验证一个新算法
- 一周打一套补丁
- 三个月重构一次架构



Lesson9:

关键人写关键算法

- 所有算法线性化
- 关键算法常数化
- 减少内存触摸数

Lesson10: 小心典型的非典型用户

- 他们很可能是非典型用户
 - 老板
 - 朋友
 - 同事
 - “热心用户”

—你自己！



- 他们的典型特征

—自以为是典型用户

Lesson10+1: 多走一步

- 可用与不可用之间
- 好用与不好用之间
- 也许只差一步



- 不懂产品的后台不是好测试

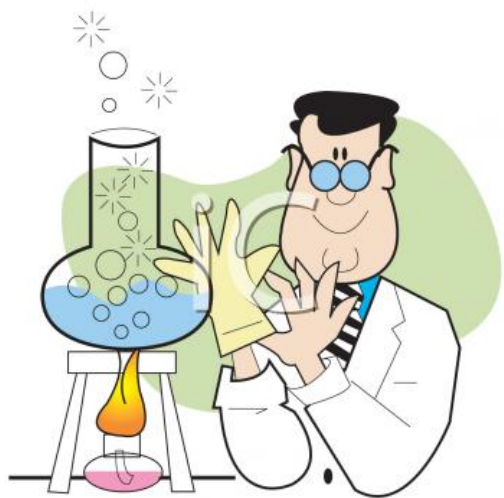
为什么那么多牛人研究了十几年算法
推荐系统还是不能令人满意呢？

学术界



工业界

VANCL 科研诚品



爱思考，爱发论文；
不喜欢实际系统，喜欢模拟实验；
不喜欢复杂系统，喜欢简约模型；
经常挑选实验结果；
偶尔伪造实验结果；
不关心用户是否满意；
只关心评审人是否满意；

我是一名伪科研工作者



- 联系方式
 - Qinyuan.feng@gmail.com
 - Weibo.com/fengqinyuan