

美丽联合业务升级下的 机器学习应用

Author: 吴海波 (花名吾加)

个人简介

- 2013-2014 华为2012实验室
- 2014-今 美丽联合（蘑菇街）
- 主要工作：
 - 数据挖掘

分类

韩版时尚套装



时尚丽人



百搭鞋柜



妆美出街



心机配角



潮男搭配



吃喝宅乐



瘦身助攻



辣妈潮宝

全部

限时快抢

00:15:05

¥49.90



团购特卖

甄选品质尖货

NEW



上衣 春款上新 不止5折



裤子 春款裤子7折尝鲜



裙子 0点上新 低至3折限时抢



首页



买买买



动态



购物车



我的



时尚套装

筛选

热卖精选

美裙套装

毛衣套装

外套穿搭

综合

销量

新品



两件套 时尚 九分 系带 新款

¥118.0

988 ☆



长裙套装 半身 松紧腰 飘逸

长袖 短款

¥69.0

最近上新



目录

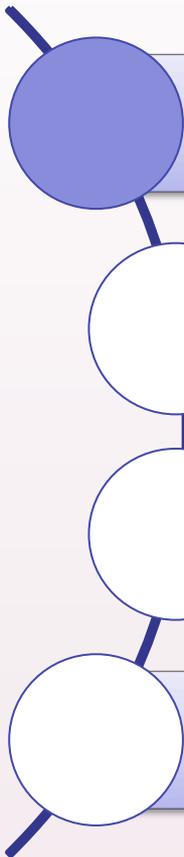
导购到电商：爆款模型vs公式

做大做强：转化率模型 + 个性化

品质升级：人群引导个性化 + EE

Q&A

目录



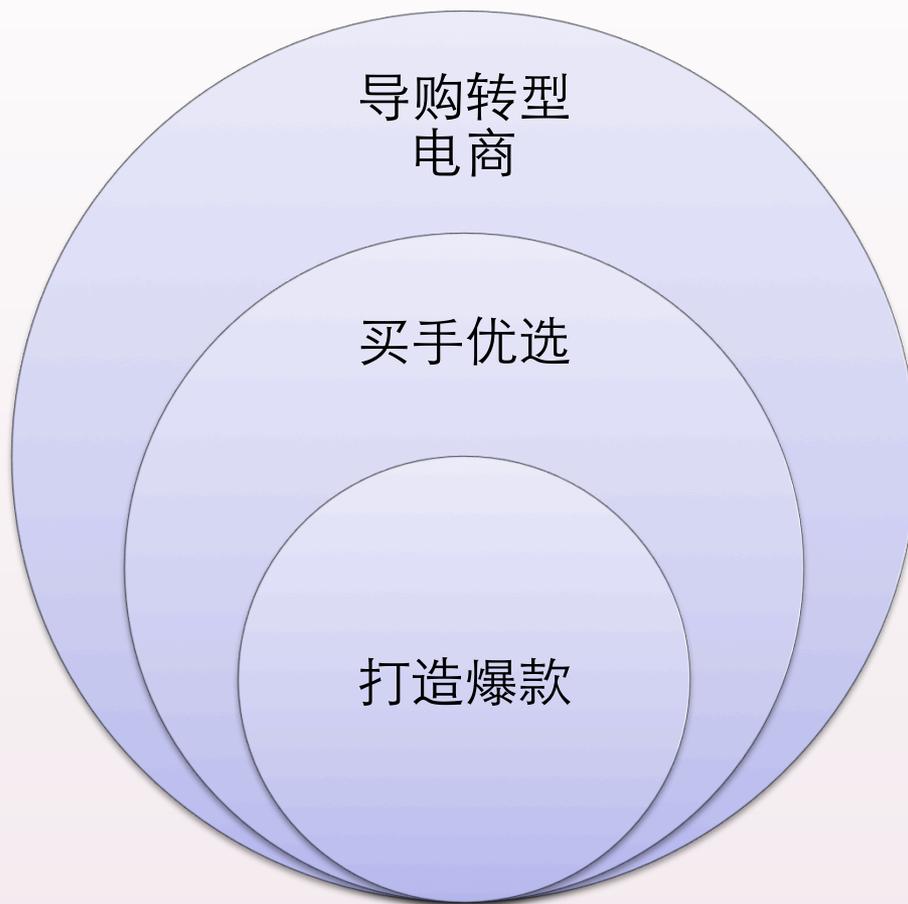
导购到电商：爆款模型vs公式

做大做强：转化率模型 + 个性化

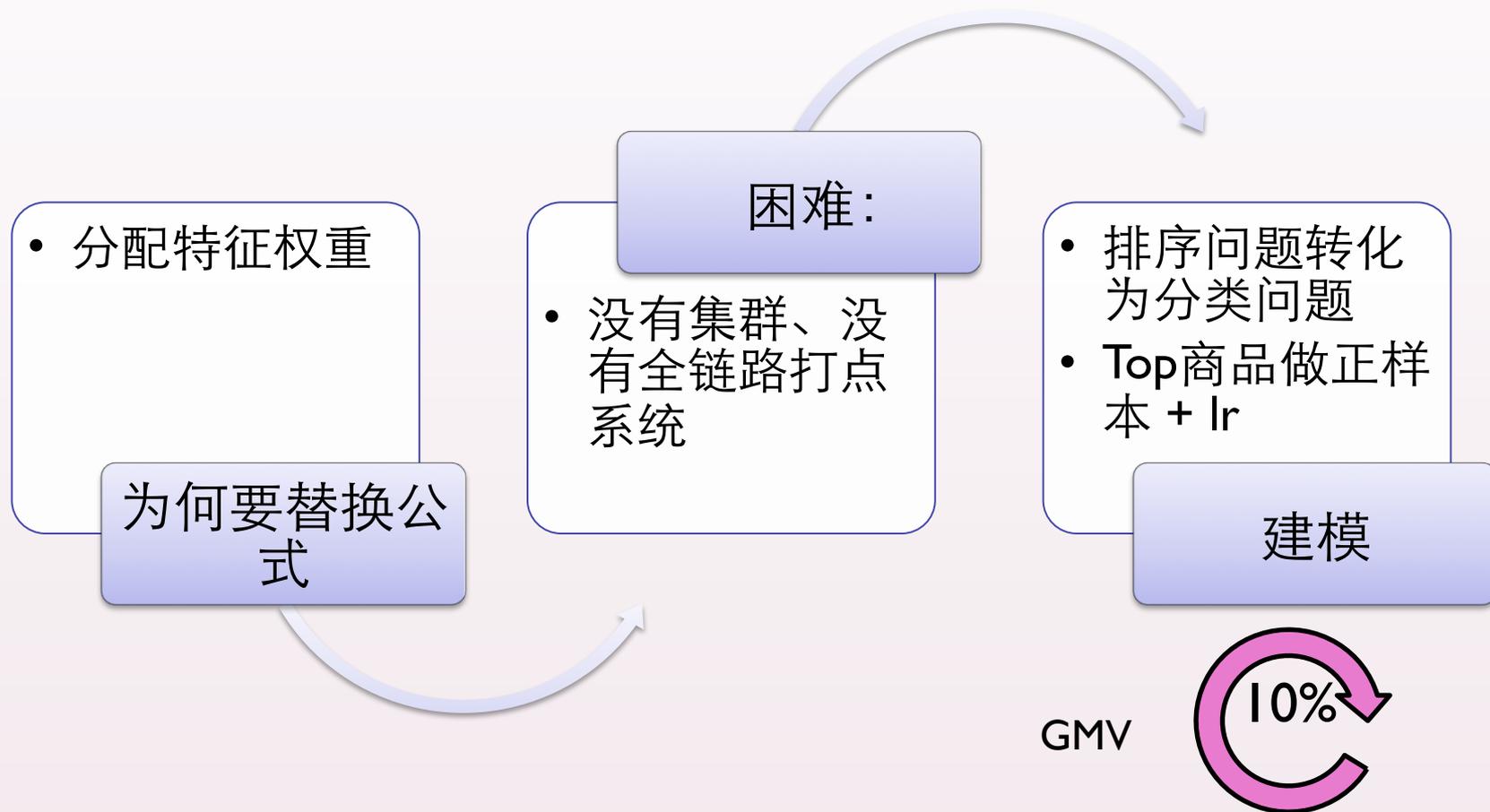
品质升级：人群引导个性化 + EE

Q&A

业务背景



小米加步枪



目录



导购到电商：爆款模型vs公式

做大做强：转化率模型 + 个性化

品质升级：人群引导个性化 + EE

Q&A

背景

- 业务：
 - 商家、商品、用户数快速增长。无法人工优选
 - 马太变大
- 模型：
 - Train auc: 0.95
 - 中间粒度商品没有区分度

问题分析

本质是类ctr预估问题，通常业界有两种：

	优势	劣势
高维离散特征 + lr	模型刻画细致 Memory能力强	对工程系统要求高 解释性差
中低维连续特征 + 非线性模型	减少特征工程工作量 可以与系统解耦	预估速度慢 调参困难 可解释性不强

样本

工业界常用的两种方式比较：

	优势	劣势
pointwise	样本构建直接 Loss function不用修改 预测和训练保持一致	模型刻画信息丢失
pairwise	问题建模更符合真实 用户行为	Loss function修改 训练和预测的特征构建不同 预测慢

特征

- 特征类型：商品、店铺、用户，三大维度



- 特征处理：log、平滑、交叉、离散化

模型

	优势	劣势
LR	可解释性强 容易调试	特征处理对模型的刻画能力影响很大
Xgboost + LR	相同的特征处理下，刻画能力更强 模型具有一定的特征挖掘能力	Robust差，会放大训练和预测数据分布的差异性 调试变困难

收益

LR模型带来的收益: 线下auc

+0.17

线上转化率

+8%

Xgboost + LR: 线下auc

+5%

线上转化率

+5%

系统收益: 个性化特征低成本加入

一些思考



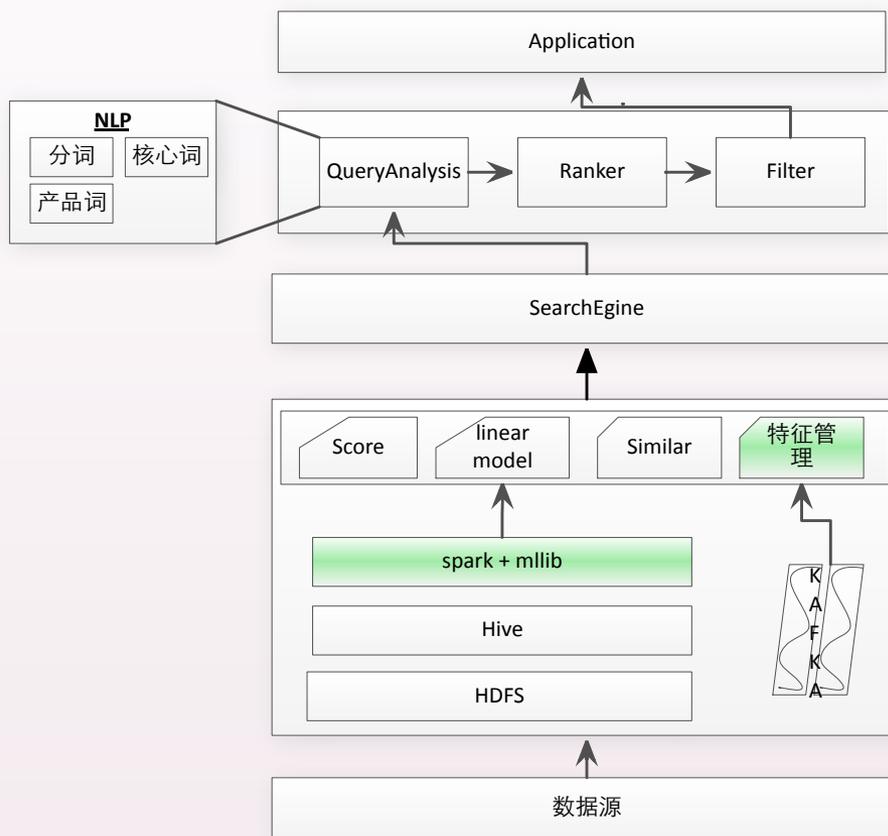
个性化方案比较

	优势	劣势
标签类个性化	可控，可解释 业务方合作场景多	覆盖率低 受限于生态的用户群， 能分的tag有限
行为粒度相似个性化	覆盖率高 刻画细致 与推荐场景类似，技术 迁移成本低	可解释差 业务控制力度低

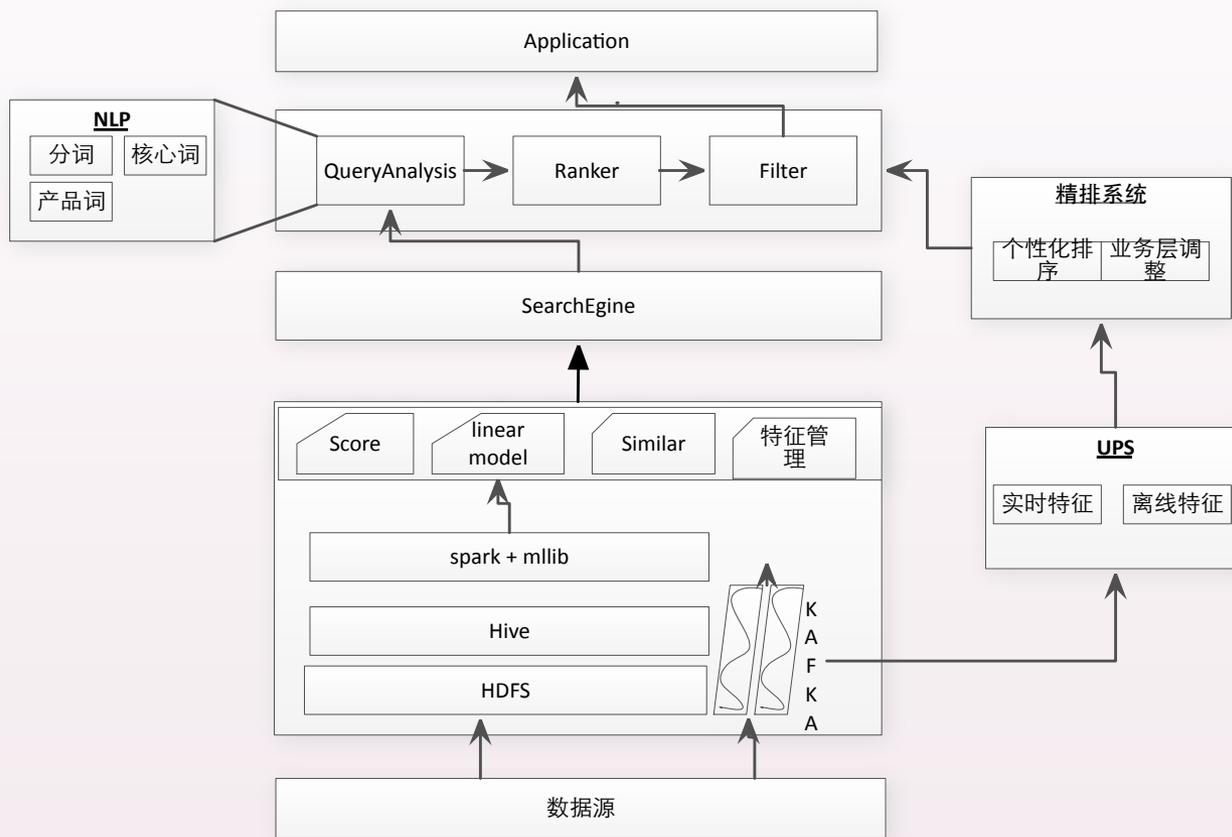
个性化 + 初排实践

	线下实践	线上实践
个性化因素作为初排之后的微调模型	单独计算偏好auc较高	gmv和转化率均有收益，多次迭代累计收益 >10%
个性化因素作为特征加入模型训练	特征空间增加很大，特征及其稀疏，整体权值较低	暂未上线

初排模型架构



个性化后架构



目录

导购到电商：爆款模型vs公式

做大做强：转化率模型 + 个性化

品质升级：人群引导个性化 + EE

Q&A

业务背景

- 生态初步形成：
 - 商家运营模式固化
 - 用户认知形成
- 短期ABtest的指标失效
- 业务方和策略组认知上的鸿沟

实时人群个性化引导

用户行为

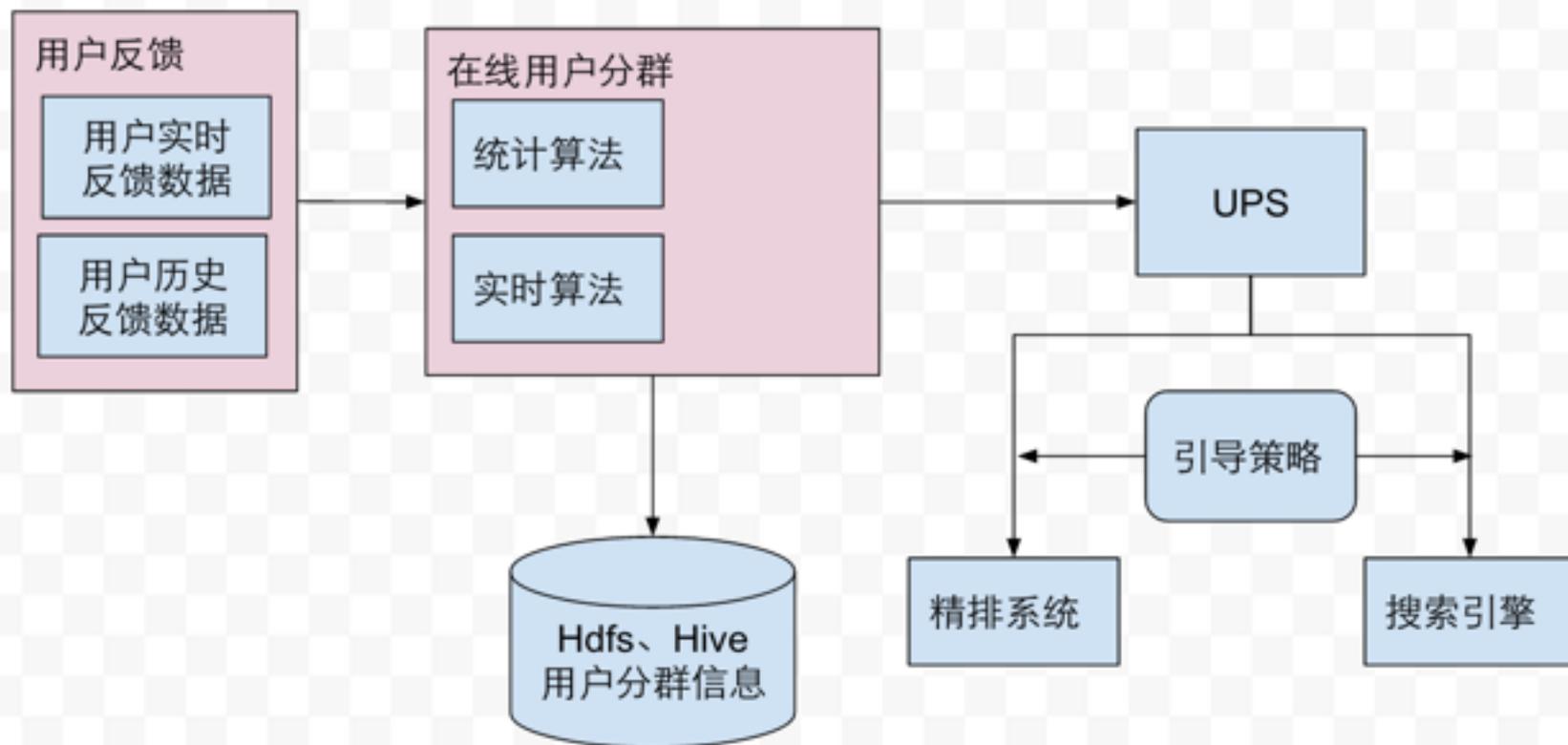
- 浏览
- 点击
- 购买
- 加购
- etc



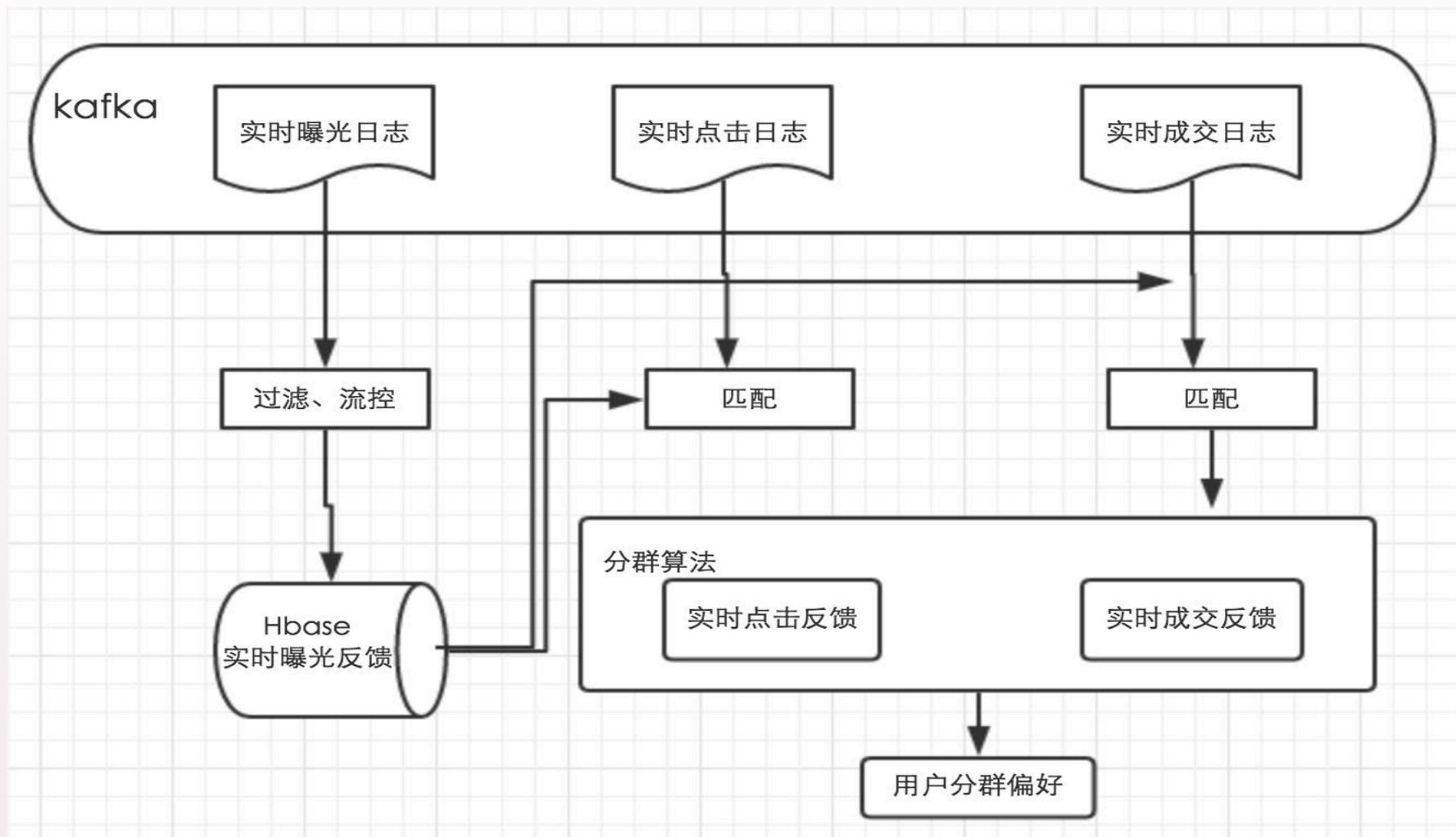
标签

- 购买力
- 新老用户
- 品质偏好
- 浏览型
- 购买型
- etc

系统设计



系统设计

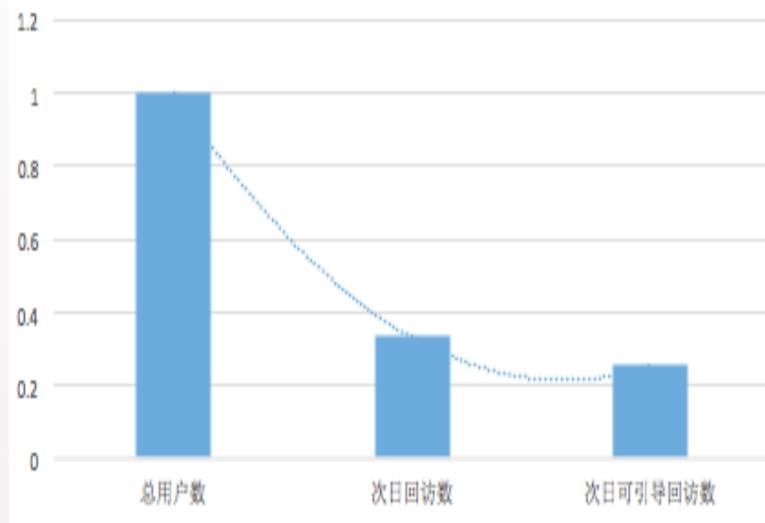


收益评估

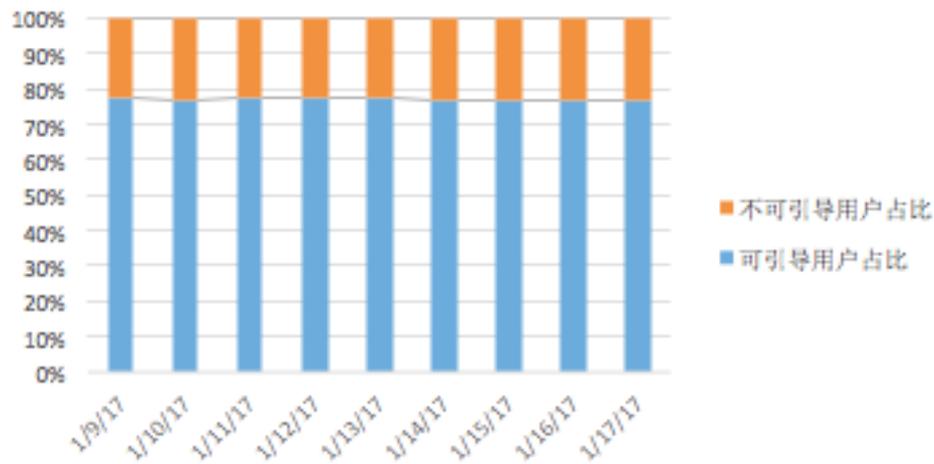
人群次日回访率



次日引导用户流失率



人群占比



在全局转化率微跌，
扶持目标流量增加40%.....

Explore&Exploit

- 小流量explore， 通过样本处理影响模型迭代
- 业务方通过EE去尝试新的业务， 比如新品、良品
- E-E策略：
 - 随机轮询
 - UCB

Future work

- dnn for ctr
- reinforcement learning

Thanks !
Q&A