

机器学习@美团

——吃喝玩乐中的算法问题

王 栋

wangdong07@meituan.com

概要

背景

- O2O行业及美团简介

算法应用

- 机器学习在用户端的应用
- 机器学习在商户端的应用

总结

- 算法应用八大“坑”

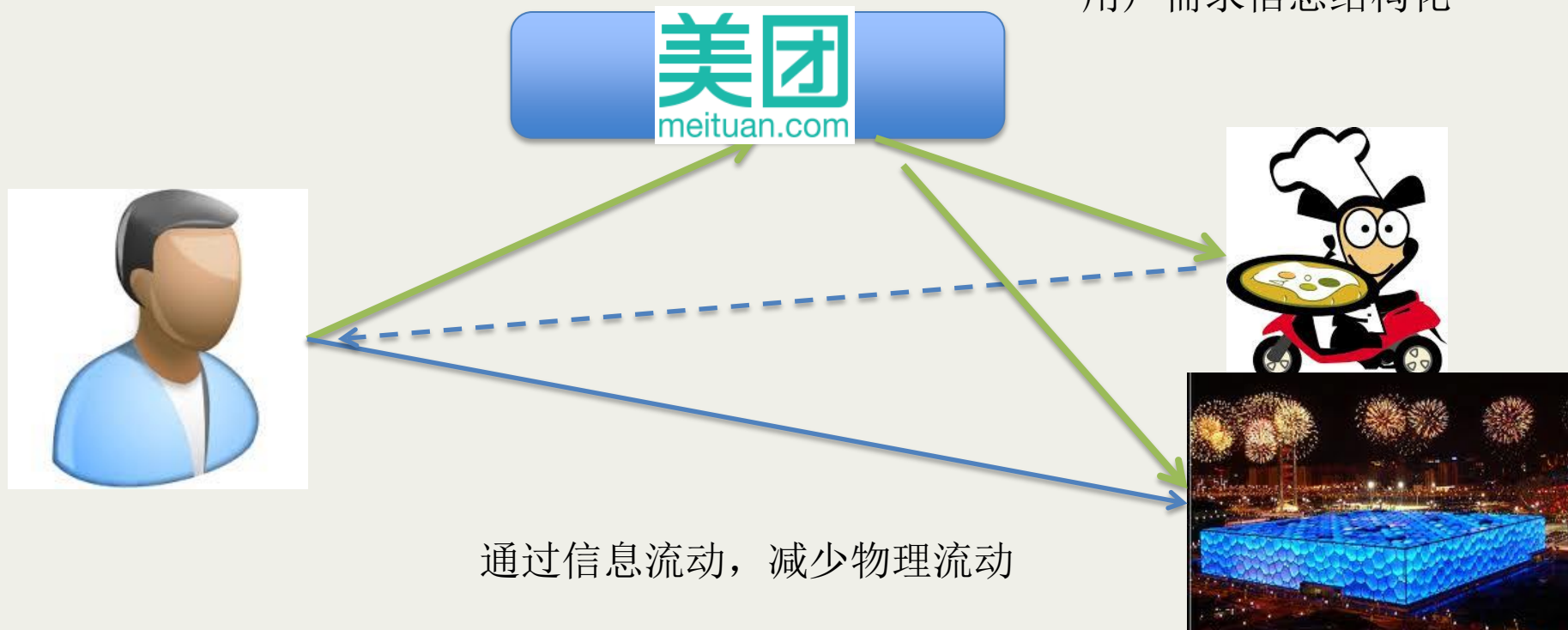
什么是O2O?

O2O (Online to Offline)

= 本地服务(LBS) + 线上连接



用户需求信息结构化

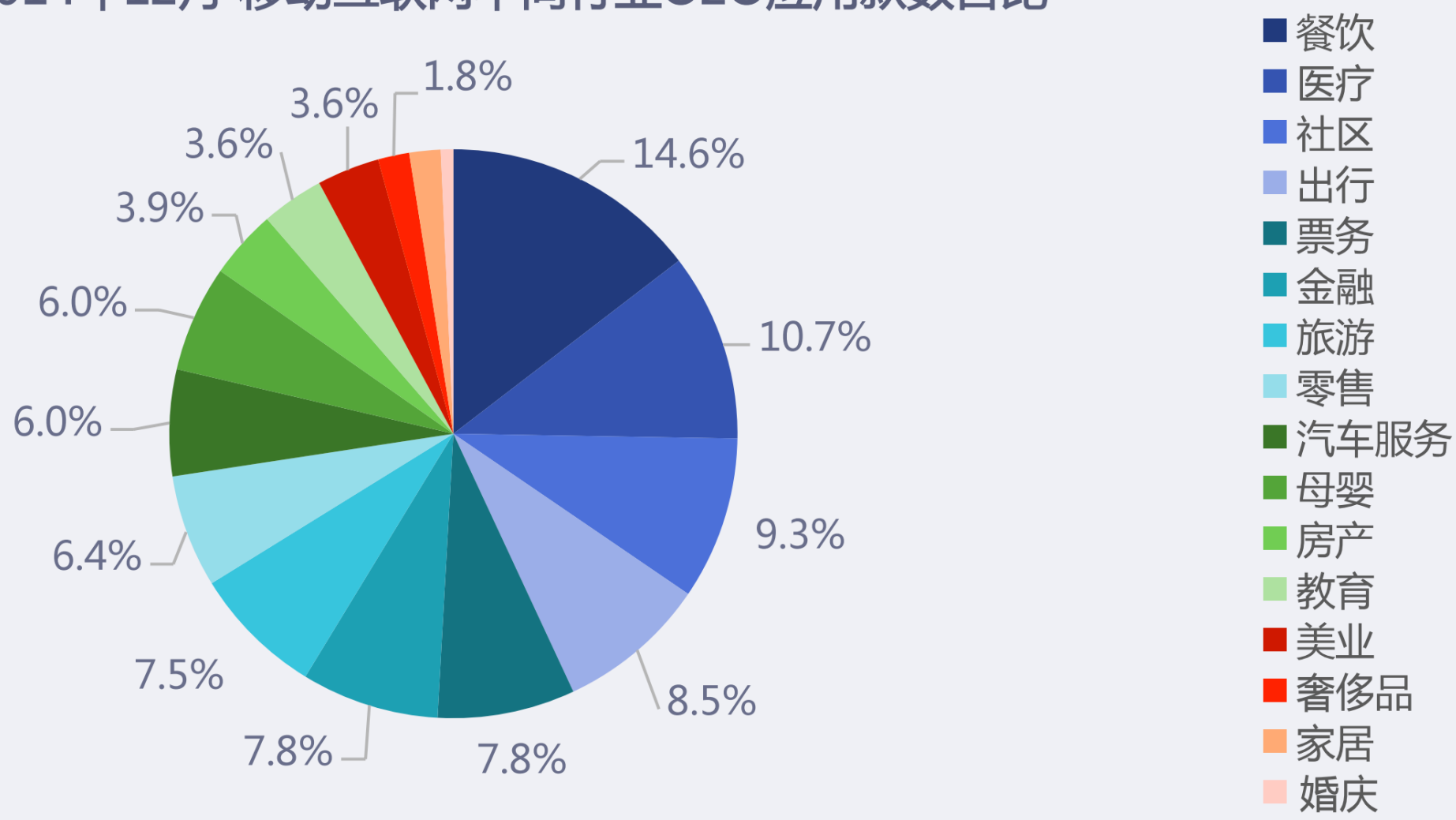


通过信息流动, 减少物理流动

O2O行业应用图谱



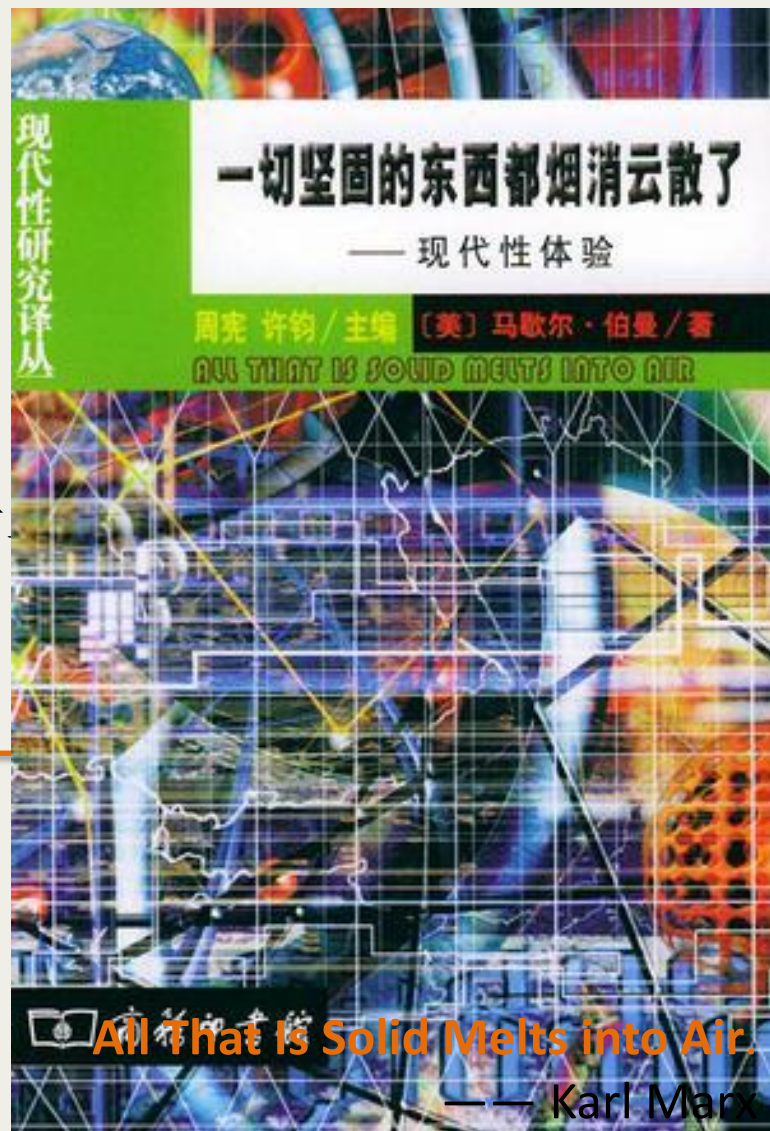
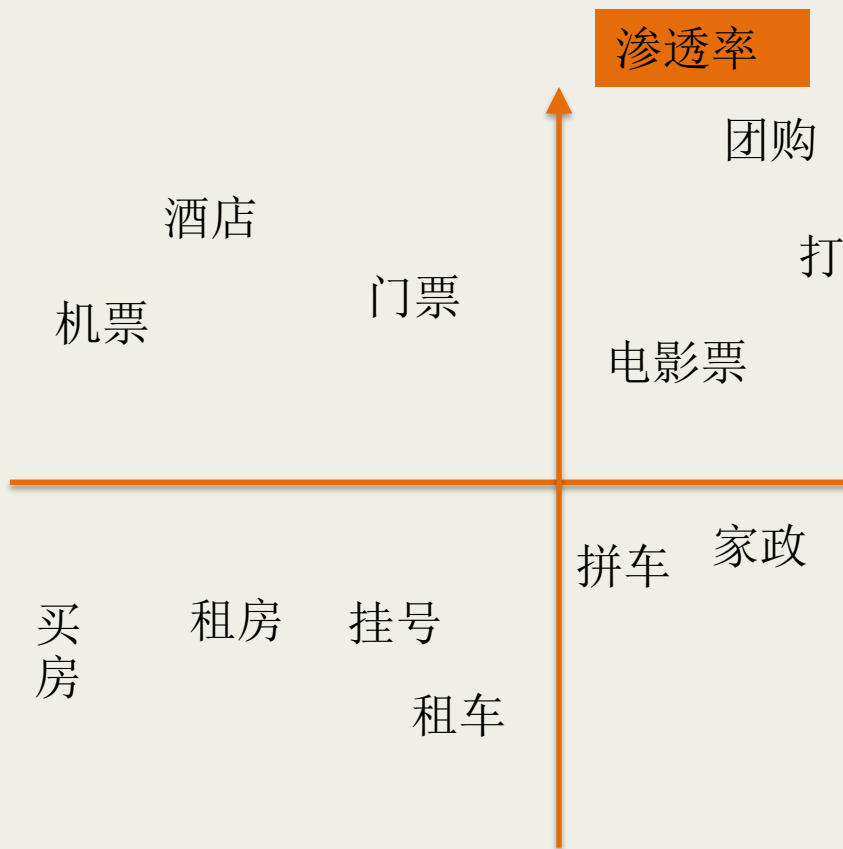
2014年12月 移动互联网不同行业O2O应用款数占比



垂直 vs. 水平（平台）

什么样的领域有**机会**？

用户获取成本 vs. 用户**价值**



美团简介



美团团购

国内最大的本地生活服务电商平台



猫眼电影

国内最大的电影分销商

美团酒店

国内第二大酒店分销商



美团外卖

国内最大的外卖平台

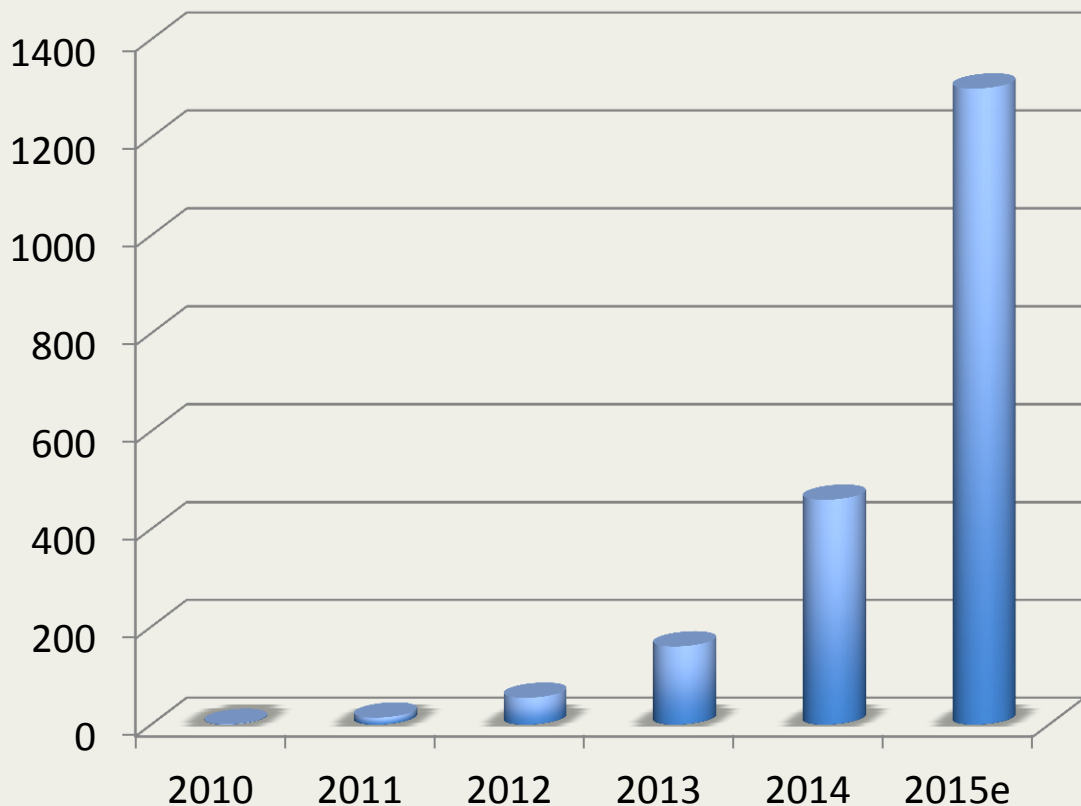
消费频度 vs 行业深度

O2O的特点



市场规模大、增长速度快

美团网交易额（亿元）



O2O vs 网络电商

共同：双边市场

差异：Location based,
消费距离受限

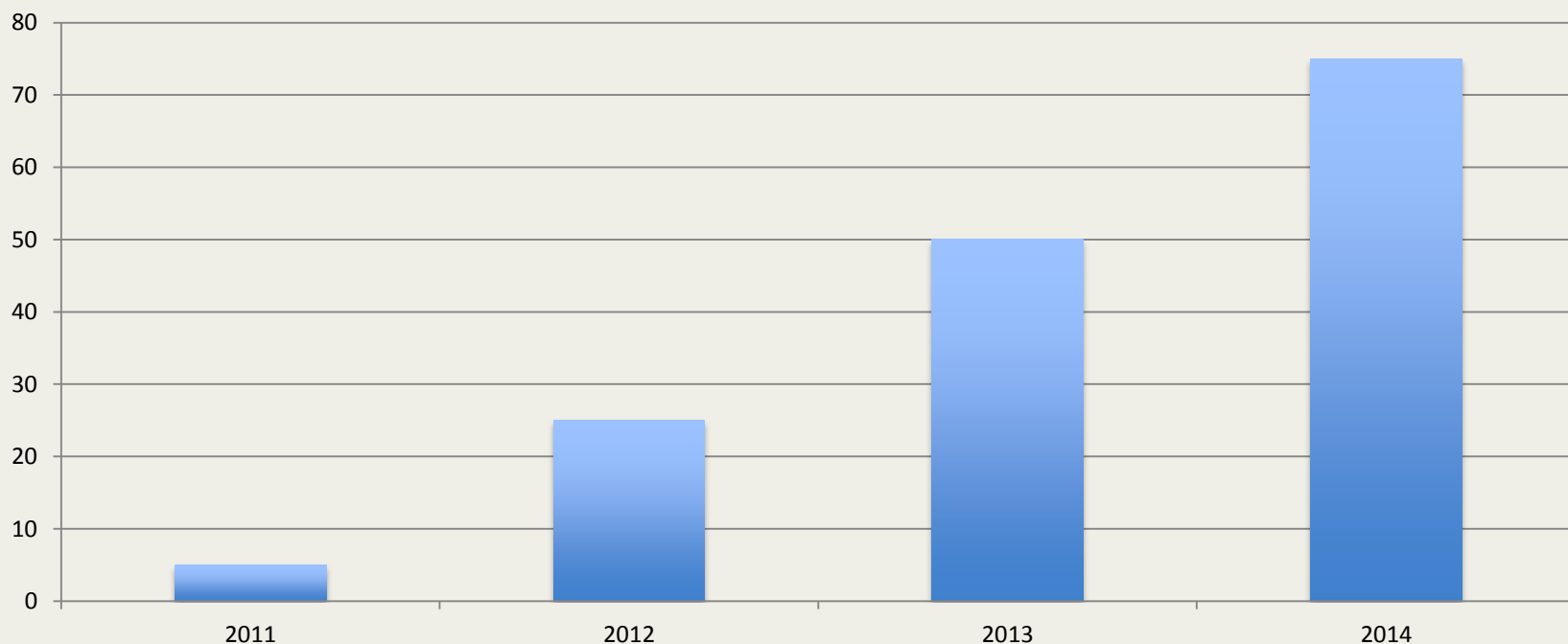
懒人经济：质优价廉

低价格
低成本
低毛利

高品质
高效率
高科技

移动化

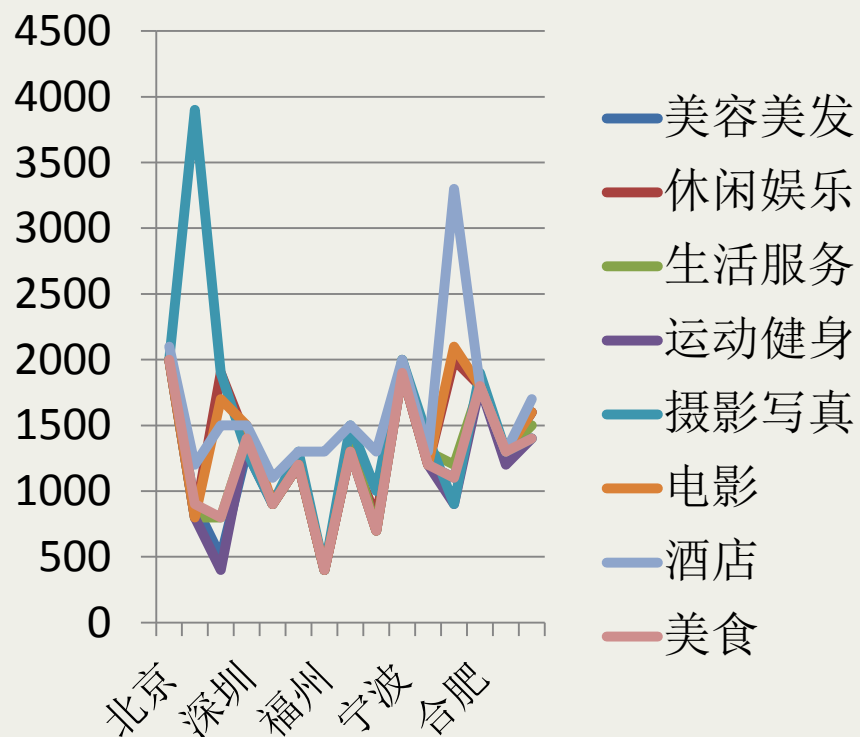
美团移动交易占比



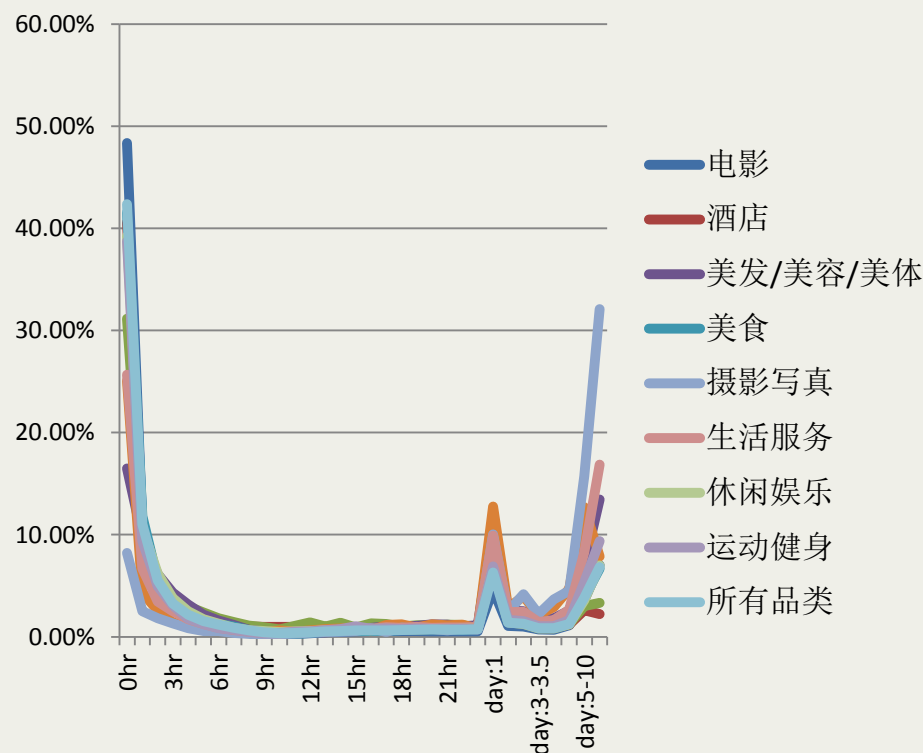
2014年12月：美团网拥有1亿多活跃移动用户，移动端贡献>90%的交易额

场景化、即时化

分品类持券80%距离



分品类持券时长



美团持券时长及消费距离

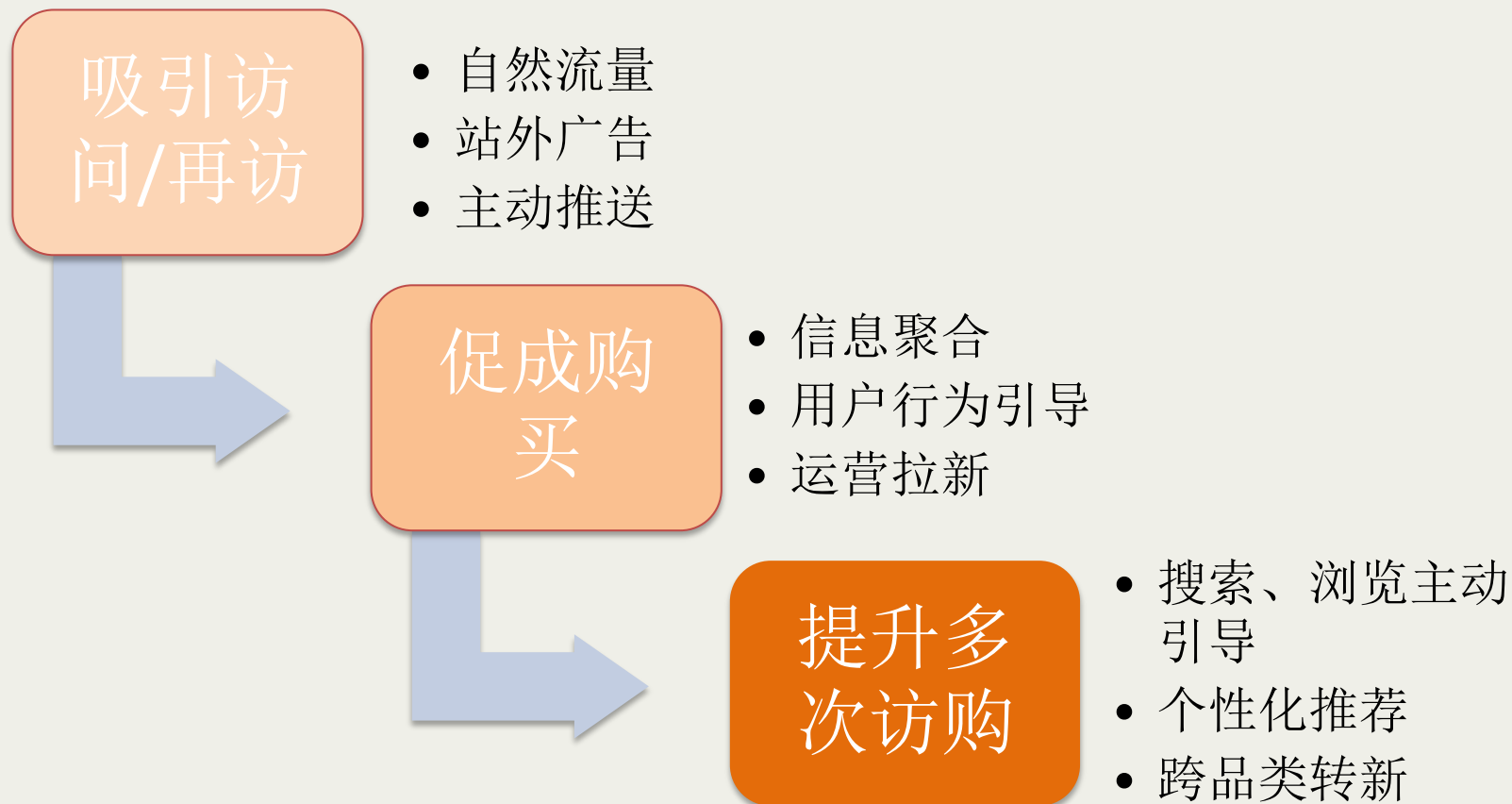
O2O平台



传统模型：边际收益递减

平台模型：双边市场的网络效应=》**边际收益递增**=》平台价值

平台： 用户端



平台： 商户端

商机发现

- 销售拜访路线优化

商家入住
及上单

- 预测可能入住商户
- 自动写单提高上单效率

运营

- 风控防刷单
- 销量预估

概要

背景

- O2O及美团介绍

算法应用

- 机器学习在用户端的应用
- 机器学习在商户端的应用

总结

- 算法应用八大“坑”

机器学习@美团

- 销量预估
- 电影票房预估
-

- 运营拉新
- 推送引导
- 流量转化

商户

- 动态触达用户
- 提供丰富产品

用户

- 个性化
- 场景化
- 即时化

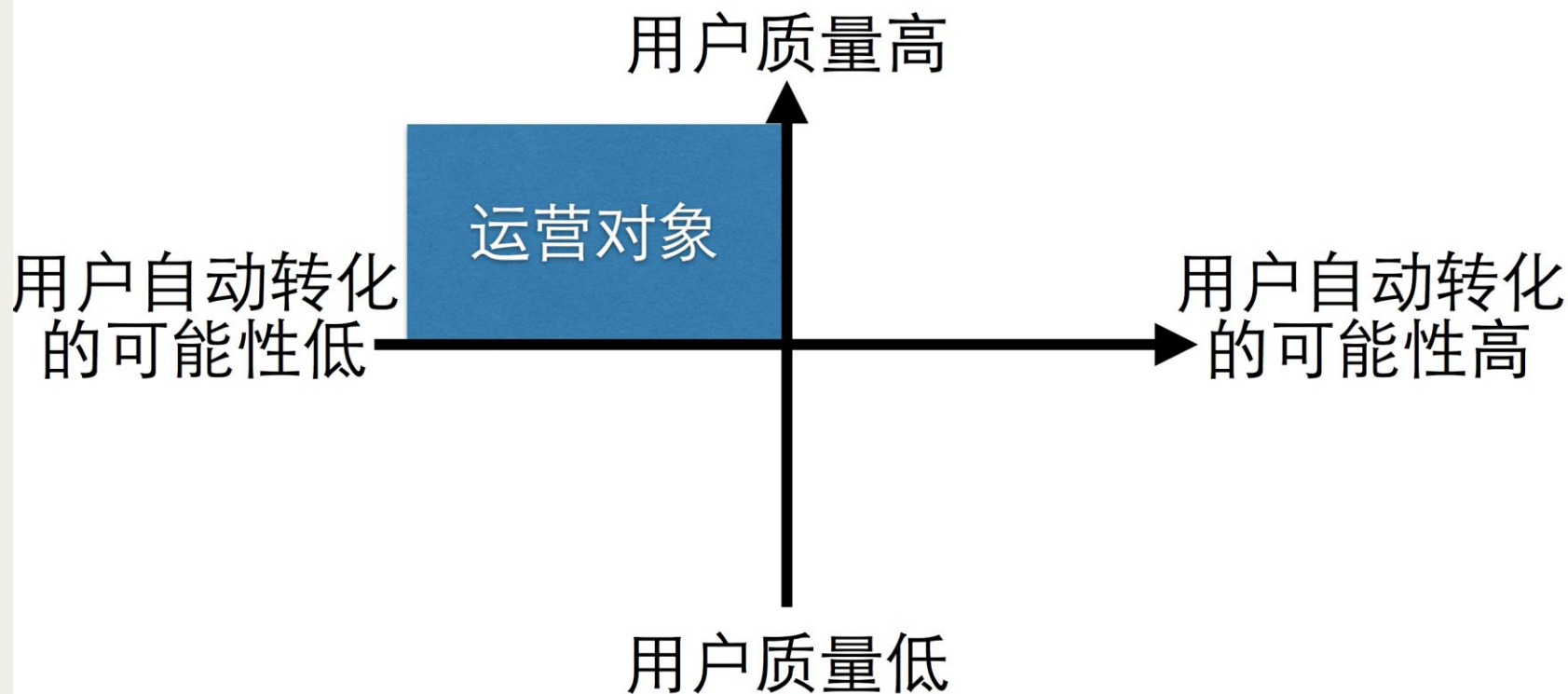
平台

- 获取用户/商户
- 匹配供给需求

用户端：运营拉新

背景：首购->留存，吸引已注册用户完成首购

目标：少花钱，多办事



用户端：运营拉新

方法：新用户画像 SVM分类

- 准确率：75%
- 召回率：68%

效果

- 拉新单位成本下降35%
单位用户营销成本=总花销/总新客数
- 总开支节省30%

• 用户画像：

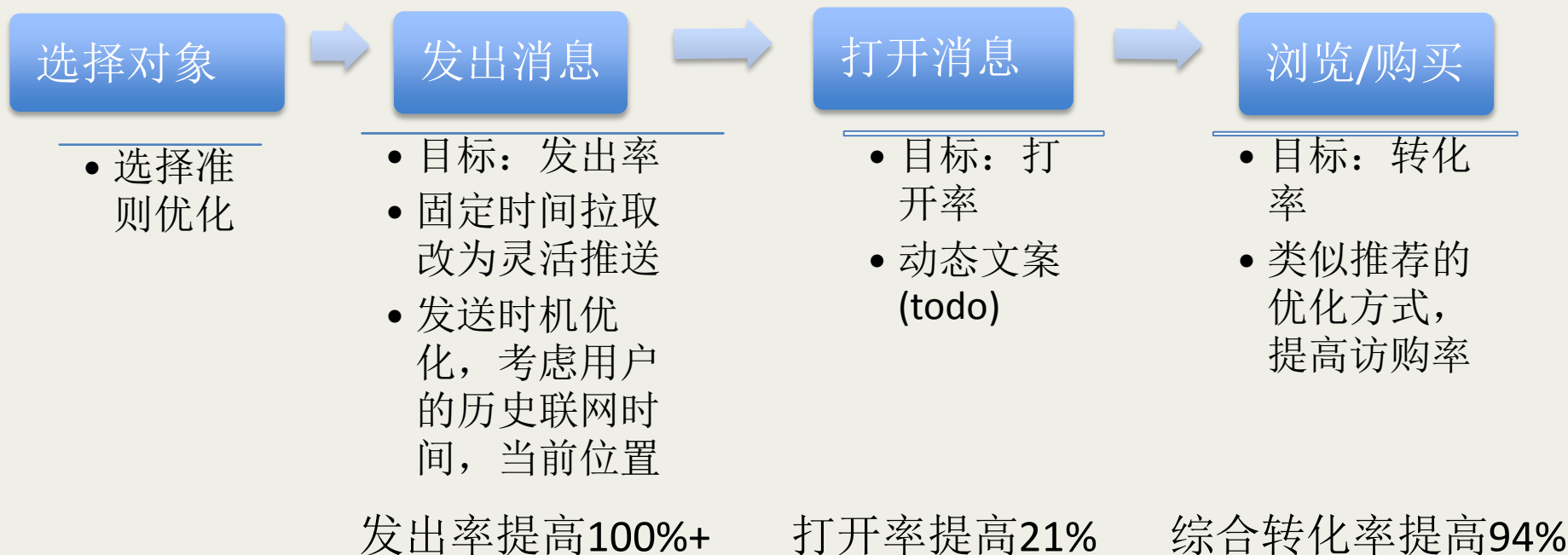


用户端：推送引导

背景：对不活跃用户发个性化推送消息，提升访问及购买

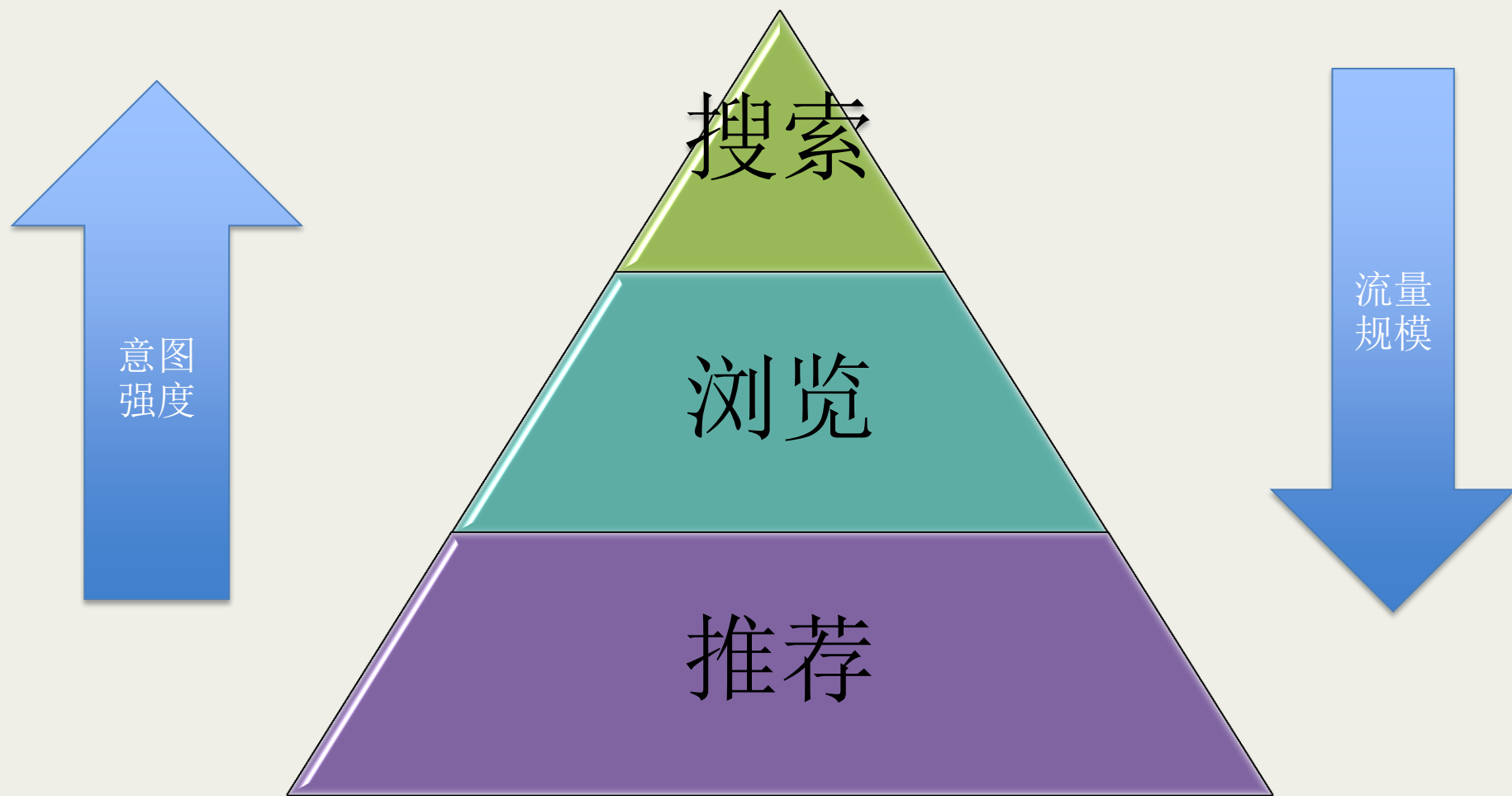
目标：提高用户转化，减少打扰用户

方法：



效果：有效降低了对用户的打扰。

用户端：流量转化



用户端：流量转化



快速交互

位置



用户偏好

场景

推荐和搜索框架的统一

基础信息挖掘

同义词挖掘
专名识别
新词识别
user profile 挖掘
deal/poi profile 挖掘
query/deal 的 topic model
.....

搜索推荐平台

应用层

Poill化列表浏览

首页推荐

Query补全

排序层

人工规则干预

排序机制

模型层

CTR model

Action model

准入层

相关性模型

业务规则

触发层

Query Retrieval

User Retrieval

数据层

通用数据

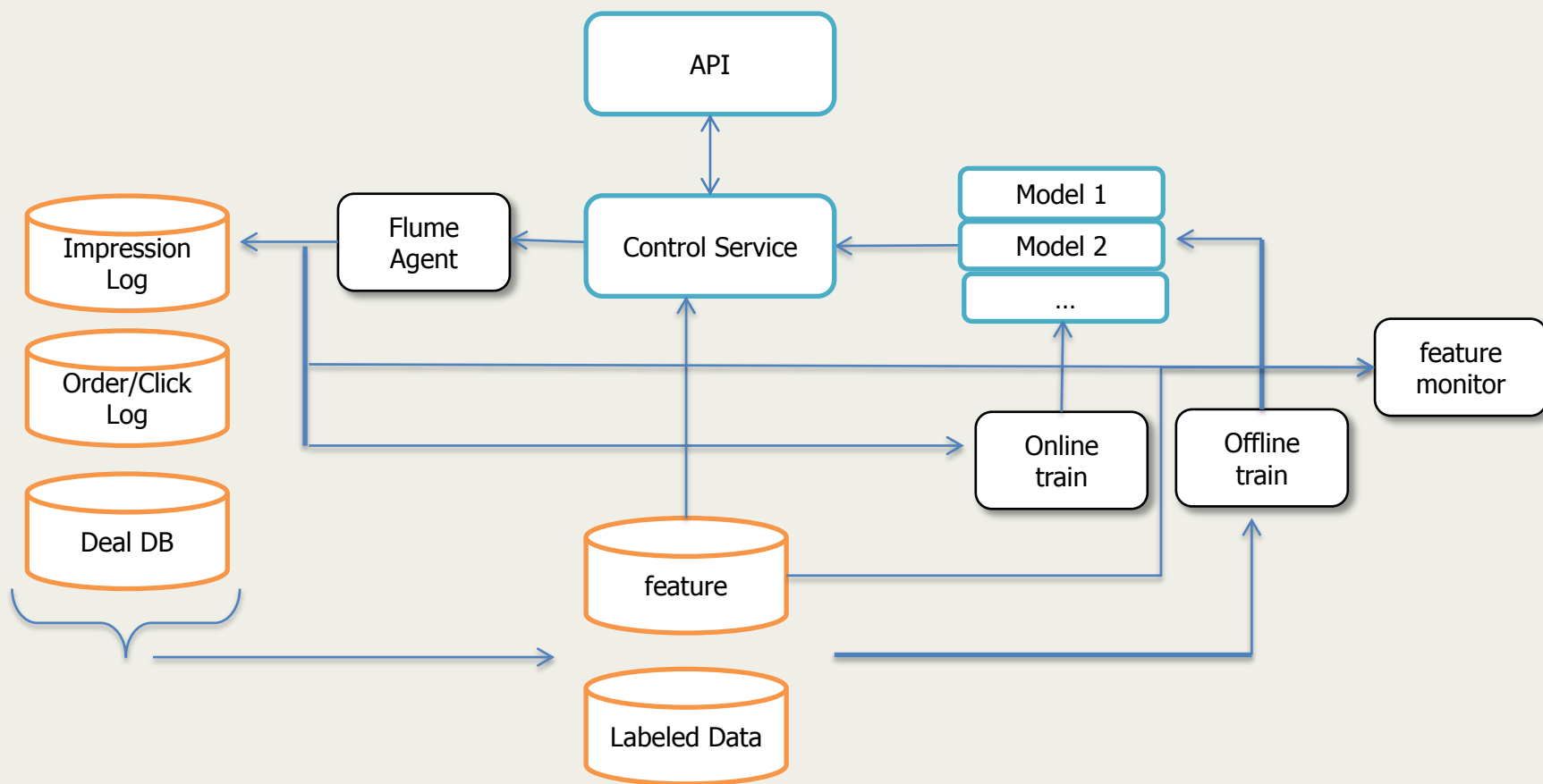
个性化数据

关系数据

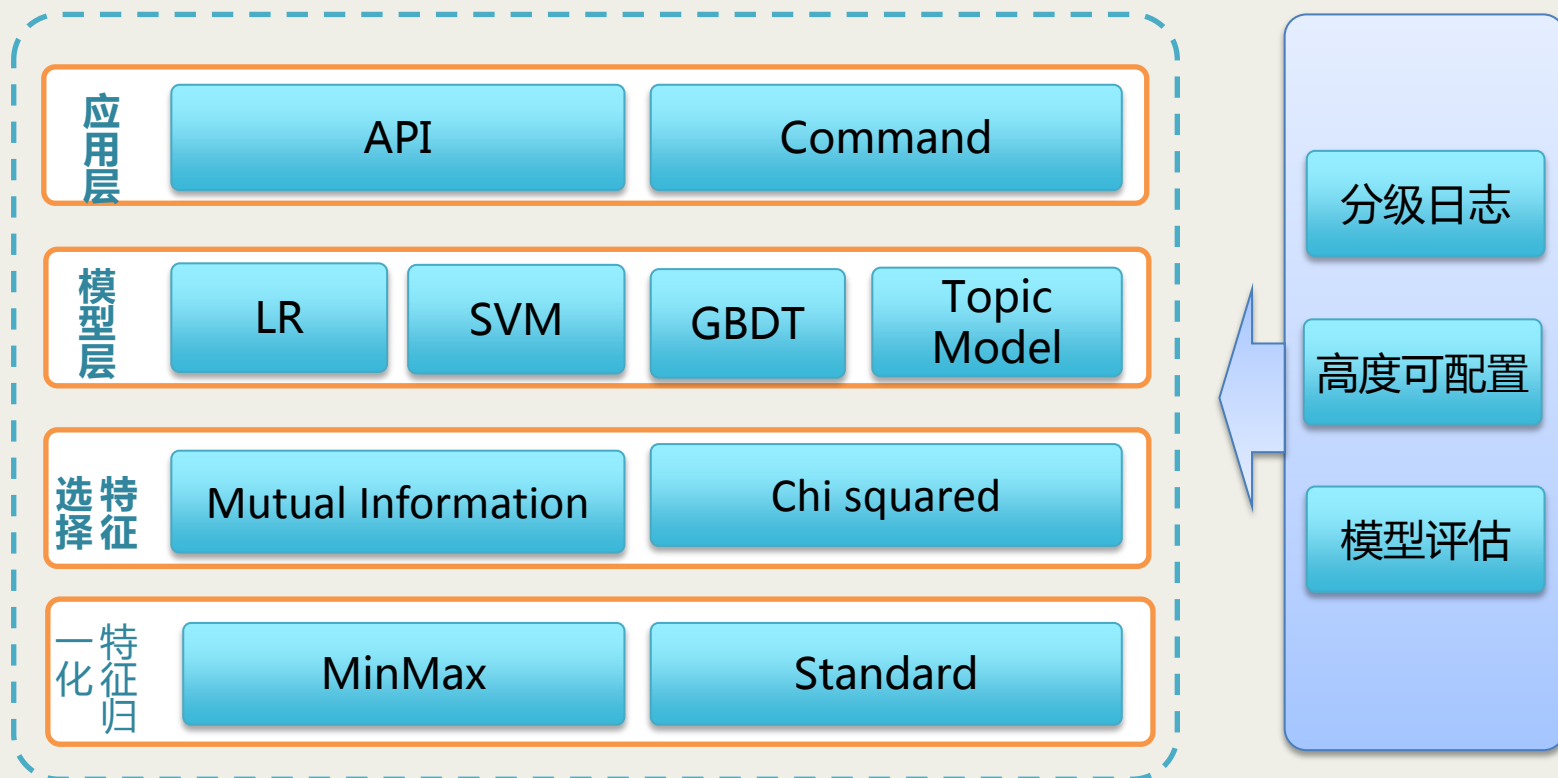
机器学习平台

评估框架
模型训练
数据清洗
特征抽取
数据流
.....

服务架构演进



基于spark的离线训练平台



美团搜索



查询补全

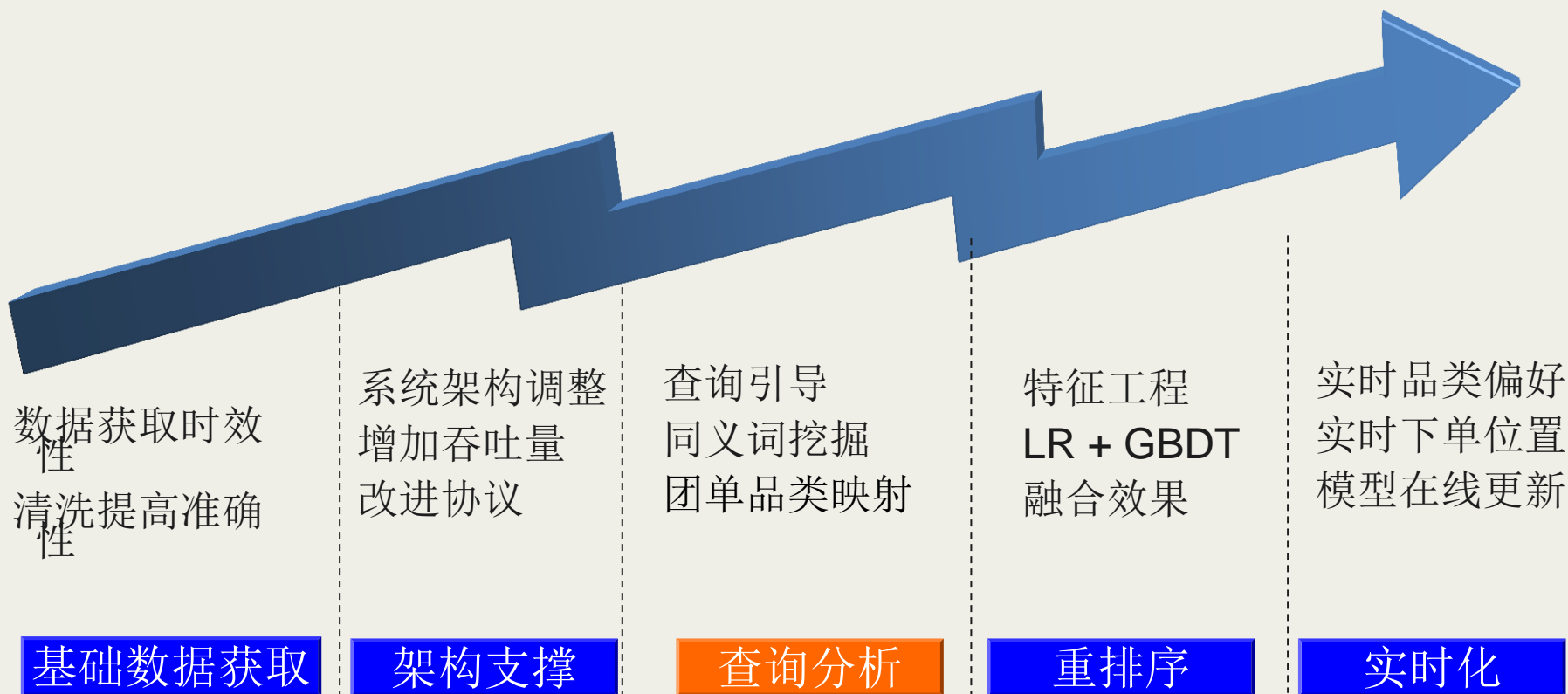


结果页



搜索系统实践

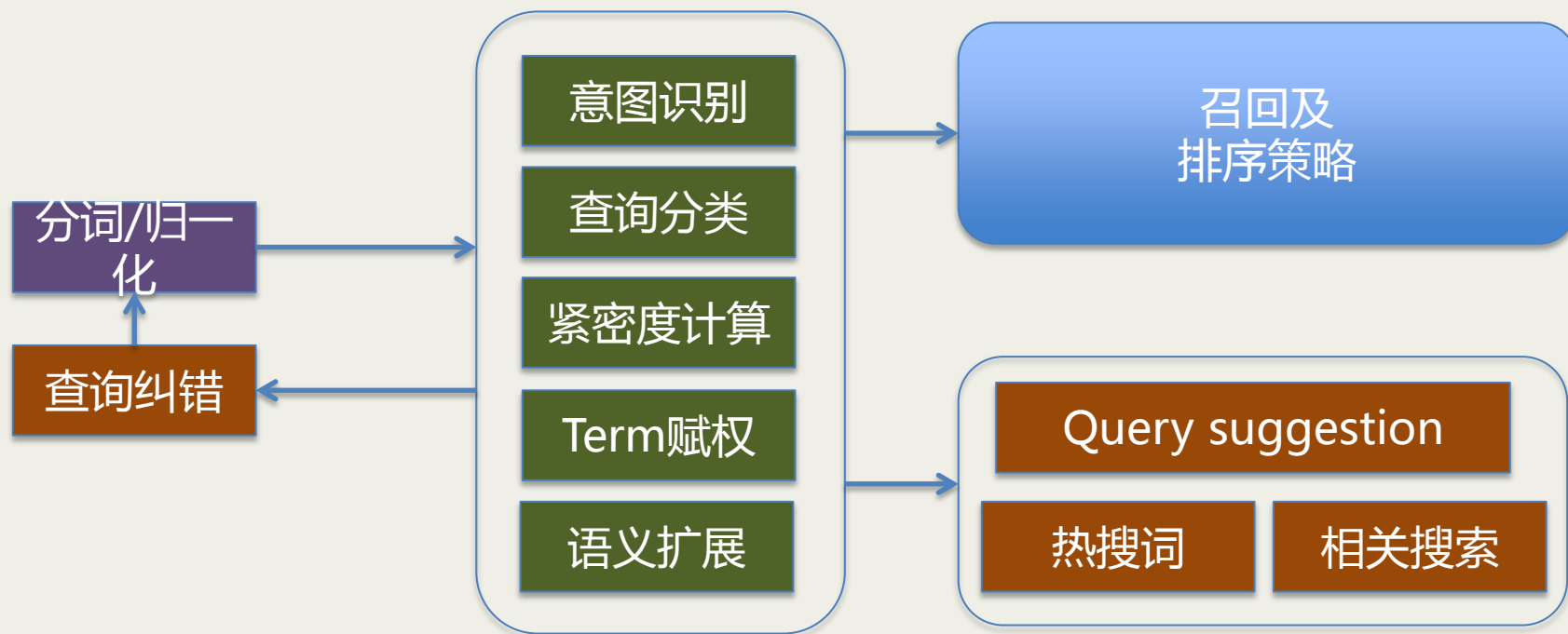
美团网
meituan.com



查询分析

查询分析面临的问题

- ✓ 用户意图多元化
 - 精确/模糊查询
 - 领域多：寻找吃喝玩乐，购物，旅游，住宿...
- ✓ 用户查询自然语言化
- ✓ 检索基于字面匹配，缺乏语义扩展
- ✓ 语义漂移

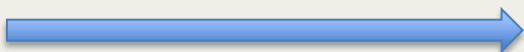


意图识别

搜索意图歧义示例及解决方案



找门票还是酒店？



通过意图分类，将“旅游”意图的POI排序提前



策略类型及占比	商家	品类	商圈	地标	酒店	旅游	商品	电影	复合意图	其他
改进前	12%	8%	-	3%	-	-	-	<1%	-	76%
改进后	21%	17%	7%	5%	6%	<2%	<1%	4%	9%	28%

团单品类细分

品类内差异对用户造成的困扰



“咖啡/酒吧/蛋糕”在后台配置为同一个类目下的品类

通过文本分类拆分子品类, 或者说打标签

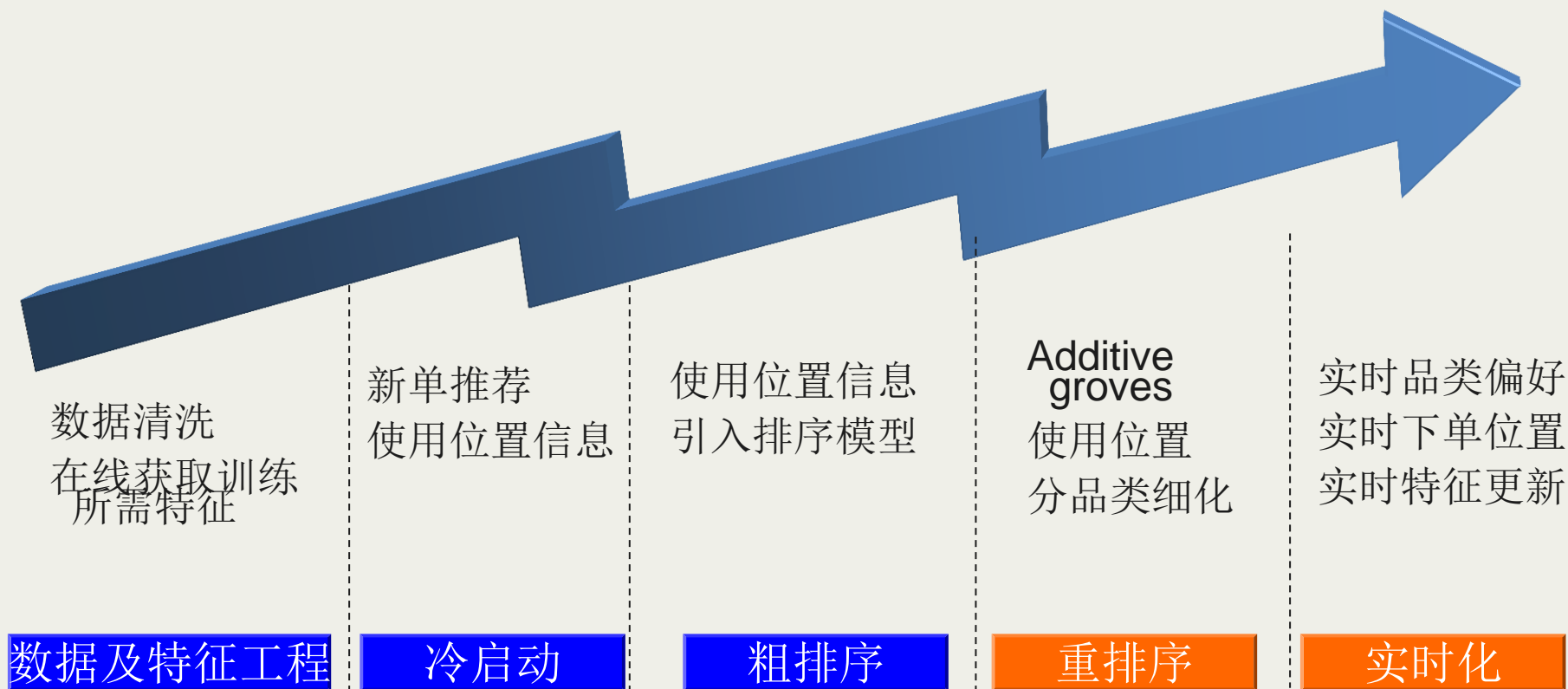


人工定义体系+SVM学习, 构建覆盖全面合理的品类类目, 更好满足用户精确品类检索的意图

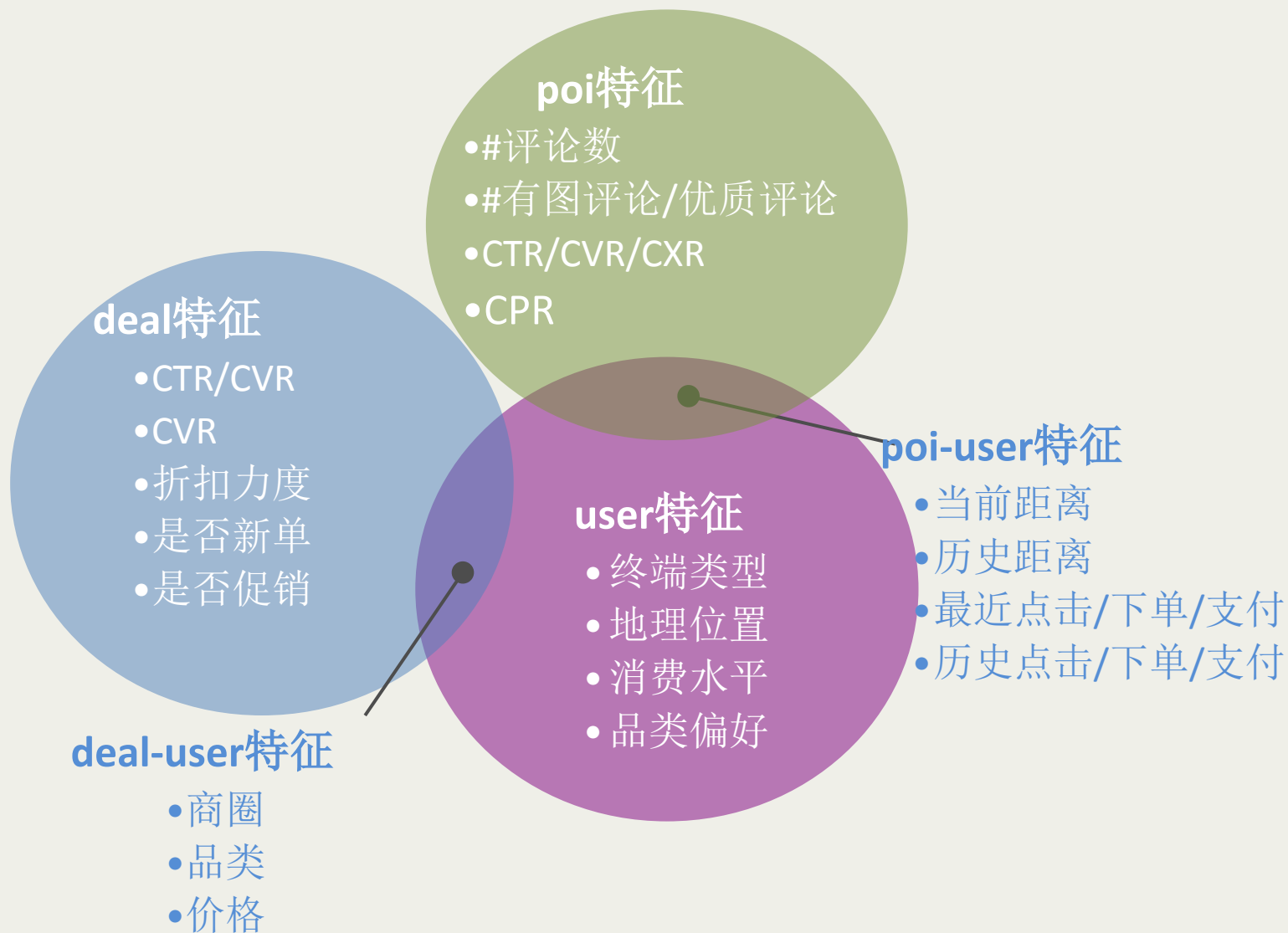
美团频道筛选



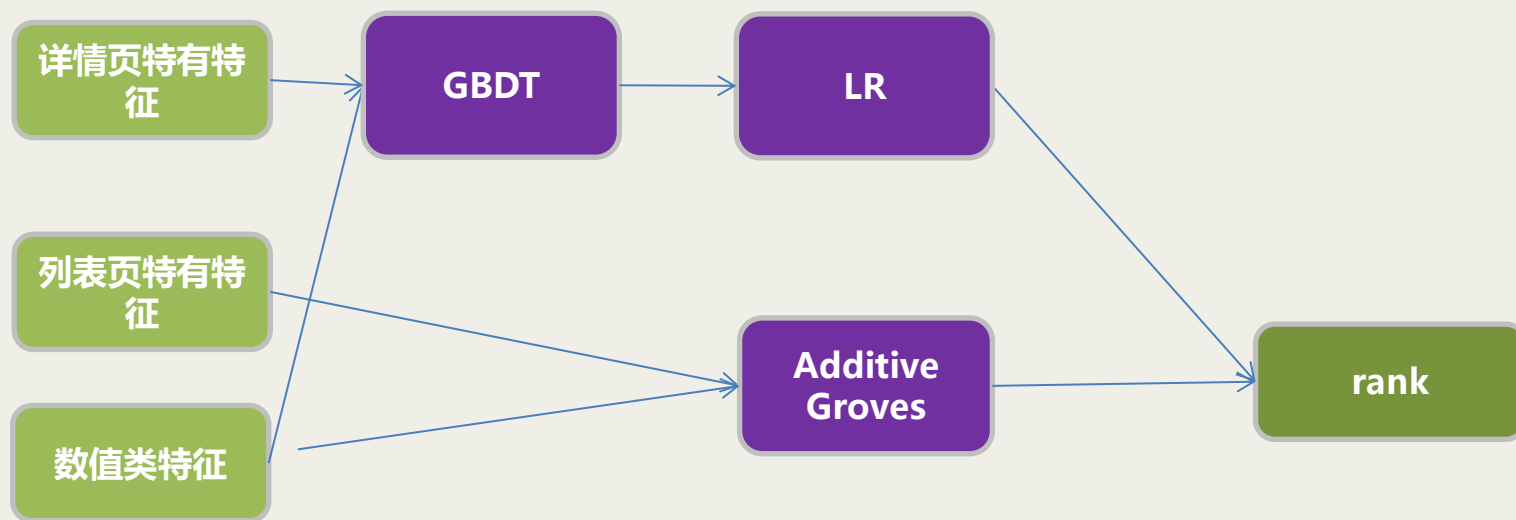
频道筛选实践



重排序：特征

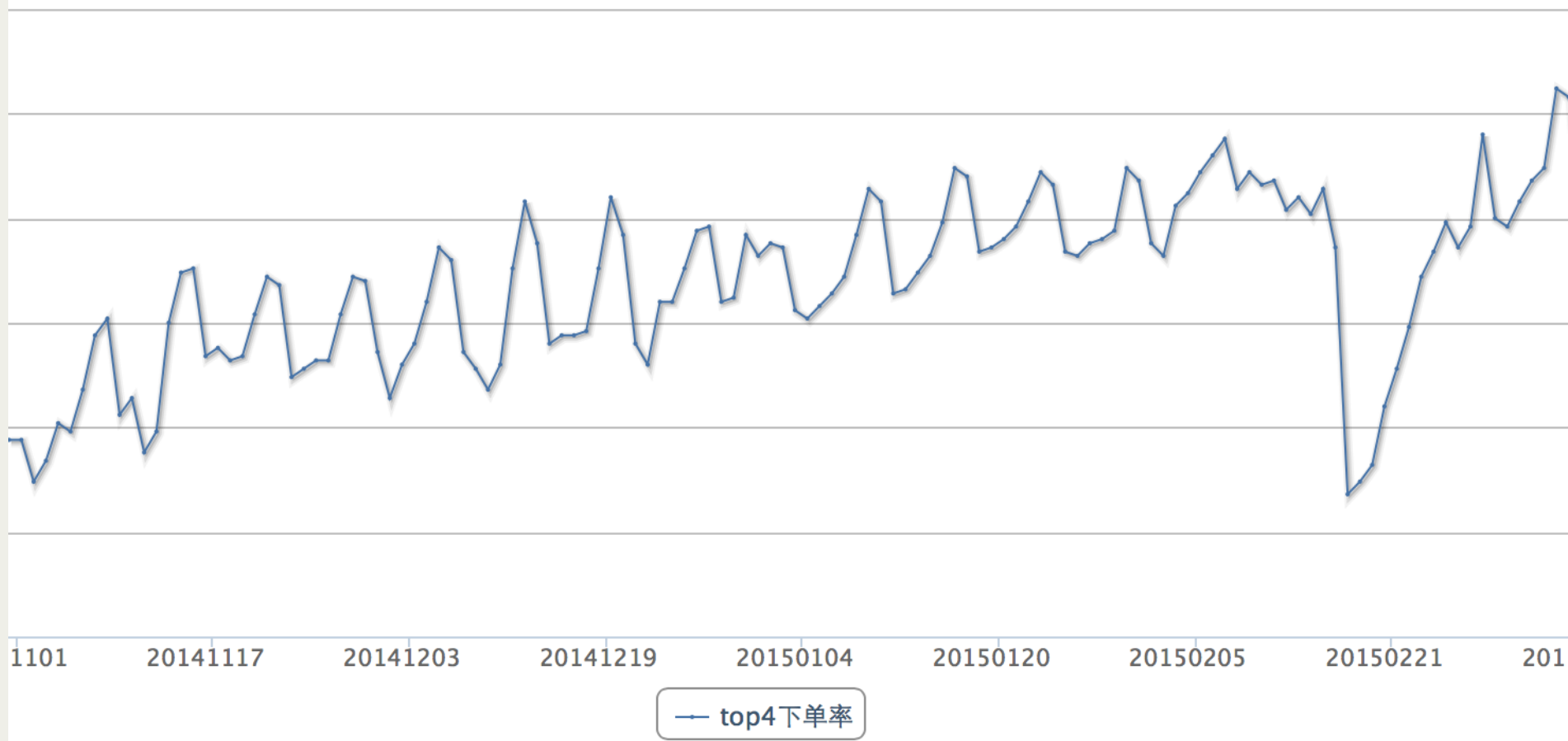


重排序：框架

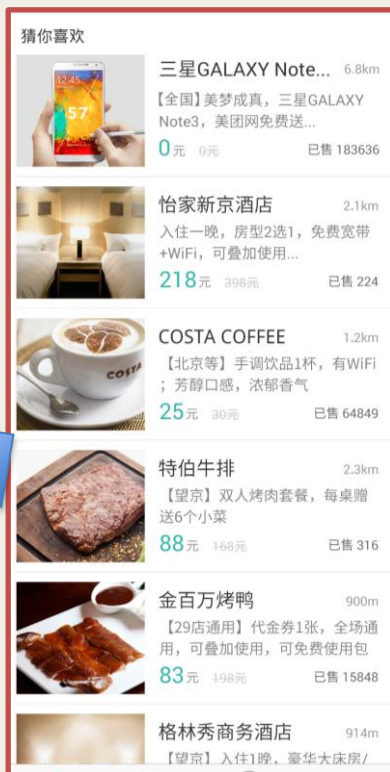
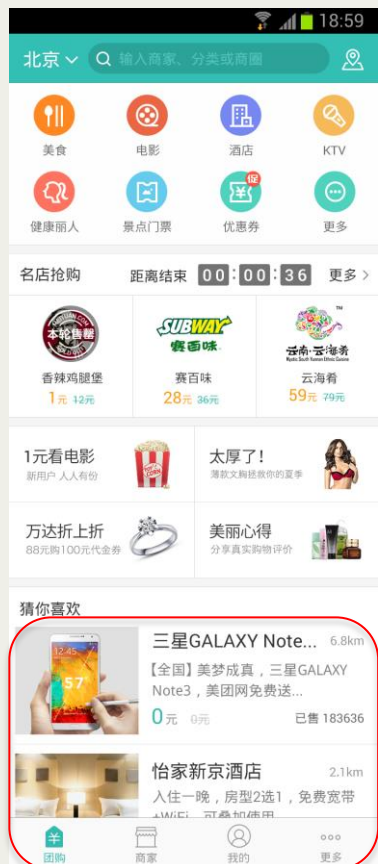


- 预测结果随数据更新而变化
 - 2小时销量数据
 - 2小时内数据预测用户实时偏好
 - 进入列表页请求用户地理位置，下单率提升3%
- 模型天级更新
 - Online learning继续尝试中

top4下单率

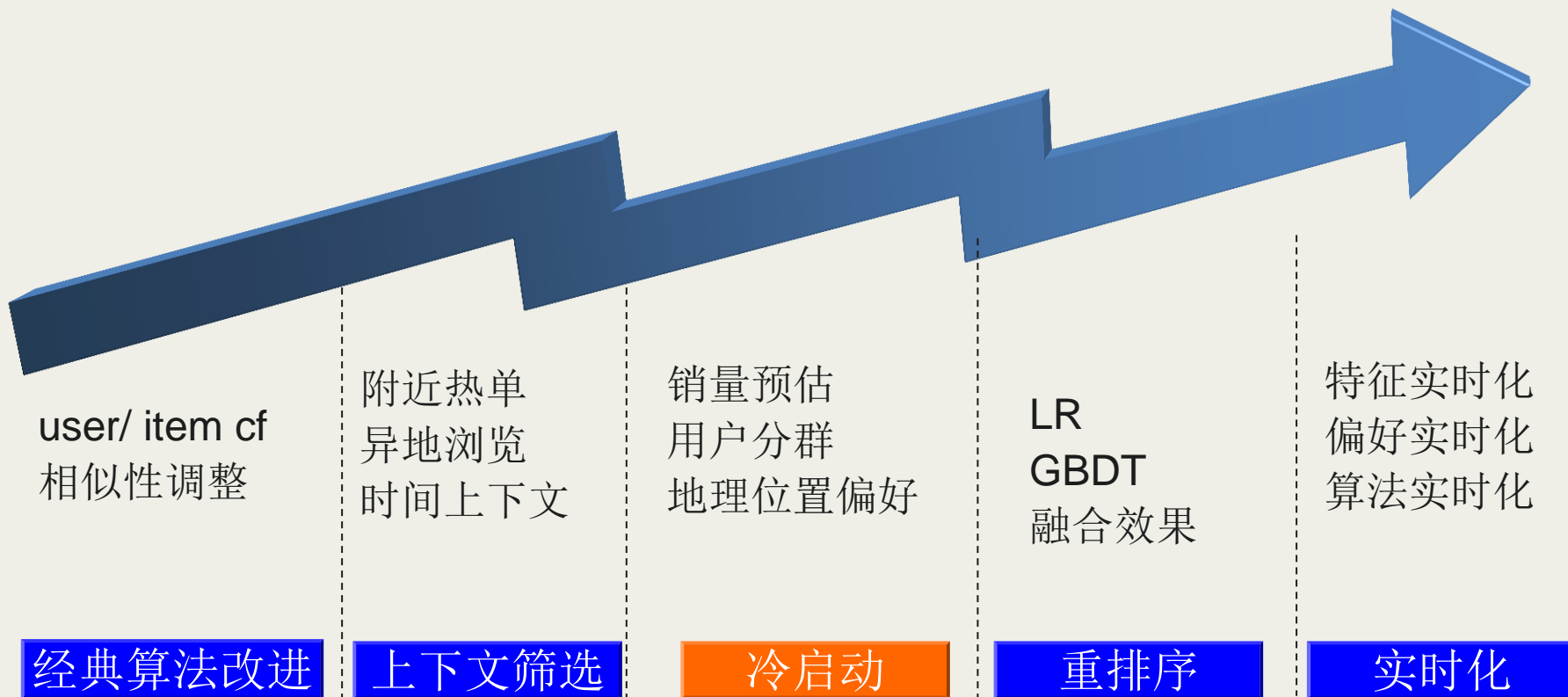


美团推荐



其他展位：
搜索无结果推荐，购买后推荐，评价后推荐，附近推荐

推荐系统实践

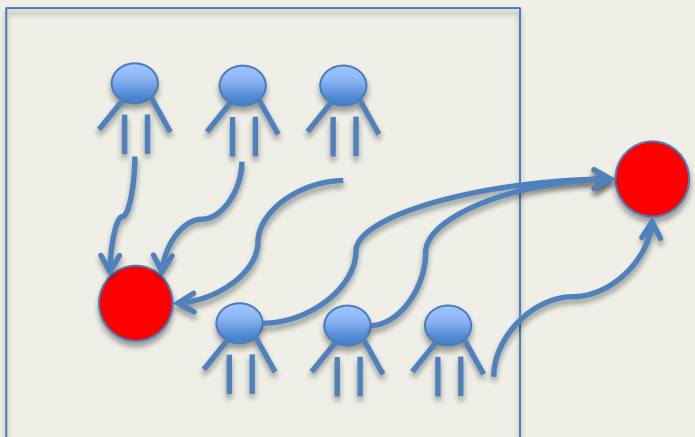
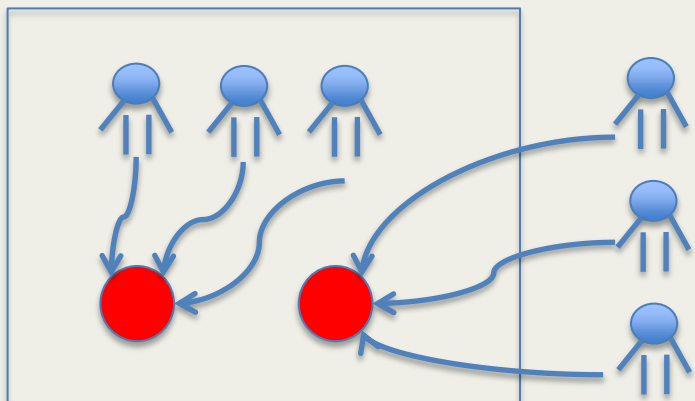


用户冷启动



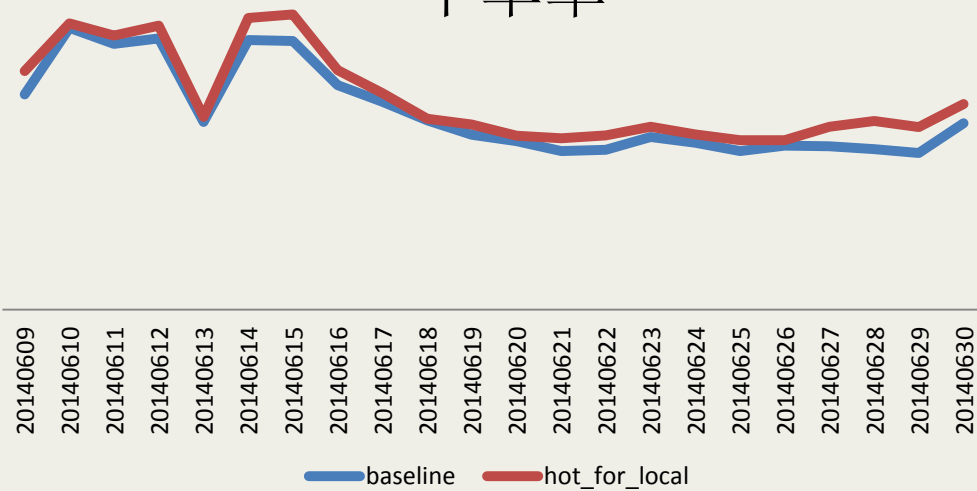
- 行为少，找不到有意义的相似用户
- Location, location, location
- 基于地理位置计算用户相似性：作为替补策略，效果一般
- 浏览地附近热单，作为上下文信息使用

加入context



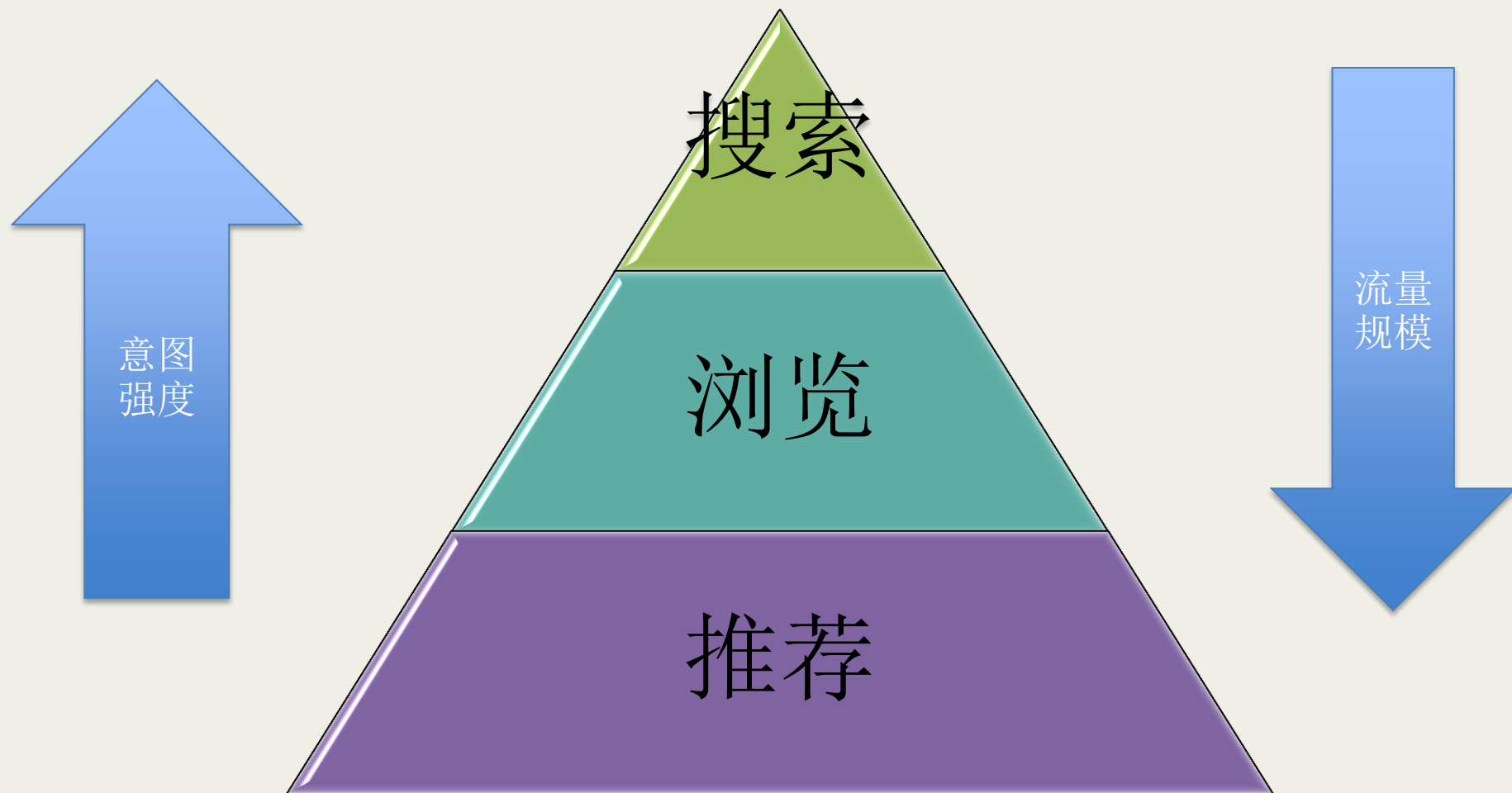
- 附近热单
 - “本地人热单”
 - “本地热单”

下单率



用户端：流量转化

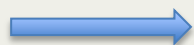
统一框架，快速迭代，经验共享



商户端：不仅是算法挑战

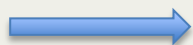


训练数据规模小



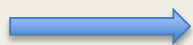
挖掘可扩展的特征
采用非线性、鲁棒模型

标注标准复杂主观



复杂目标进行模型分解
交叉验证减少噪音

单用户特征不可用



人机结合提取稳定因素
使用用户统计特征

多团队合作



协调各方预期和利益

团购单销售额预估

背景： 在线deal数量众多，需要关注deal的质量

目标： 评估deal的质量

方法：

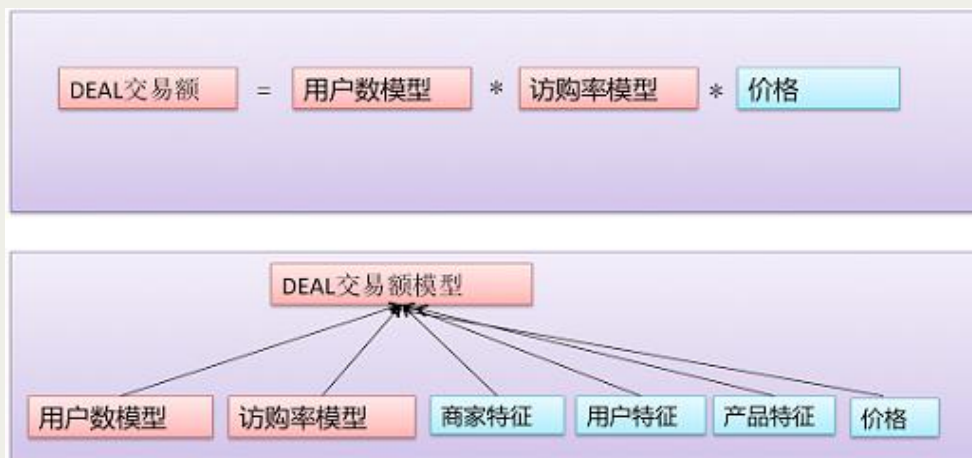
分解问题

- 以销售额来代表质量
- 再分解销售额

根据问题特征/模型并迭代

效果：

1. 访购率预测准确率(R²): 0.8
2. 价格特征权重符合产品预期



电影票房预测

背景：电影票房预测数据可以
指导片方发行/院线排片

目标：得到可用票房预测

机遇：
猫眼30%+市占的数据
用户其他行为及消费记录

挑战：
东西方人情世故不同
演员/导演的名气不靠谱
度量搜索/社交媒体效果不理想



概要

背景

- O2O及美团介绍

算法应用

- 机器学习在用户端的应用
- 机器学习在商户端的应用

总结

- 算法应用八大“坑”

学习算法应用八大

“坑”
确实需要上算法么？

星际可以穿越，但特征不能！

算法一定要酷炫

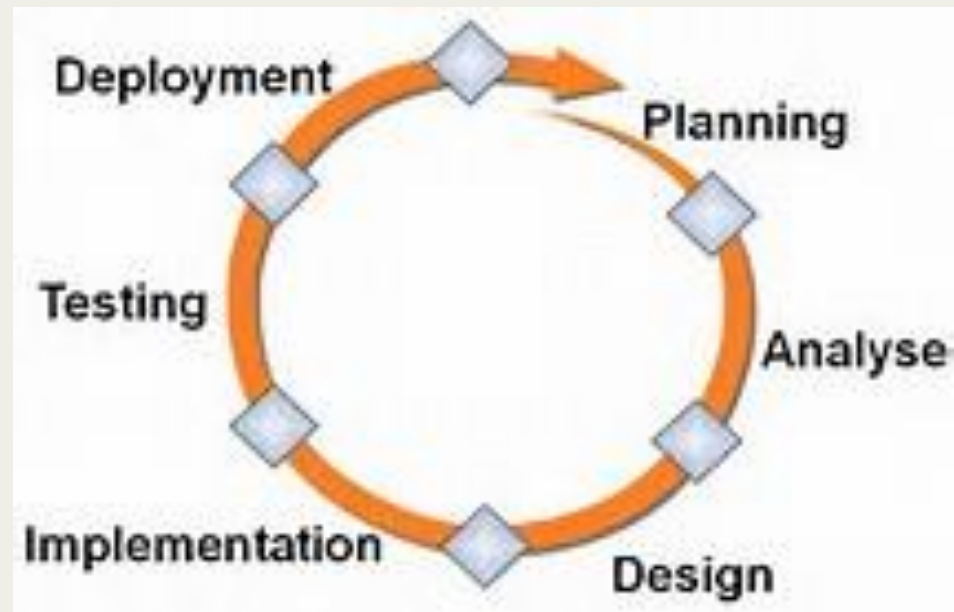
算法准确率不到90%就是要流氓？

重算法、轻速度（响应速度/迭代速度）

上线就是大功告成

线上AB定一切

统计分析实时化



Skepticism is the
first step towards
truth.

Denis Diderot

美团技术沙龙

筹办中，敬请关注



相关问题探讨及咨询，
个人微信号：dwang97

标题请注明 机器学习算法/架构工程师, 简历投递地址

hr.tech@meituan.co





谢谢!