

如何做好 本地生活服务推荐

美团网 推荐与个性化 沈国阳

Outline



- 美团的几个重要推荐产品
- 美团推荐产品的目标
- 美团推荐系统的整体框架
- 比较有特色的工作
 - 基于地理位置的冷启动
 - 排序的关键工作
 - Interleaving v.s. abtest
 - Online Learning

美团的几个重要推荐产品

美团网
meituan.com

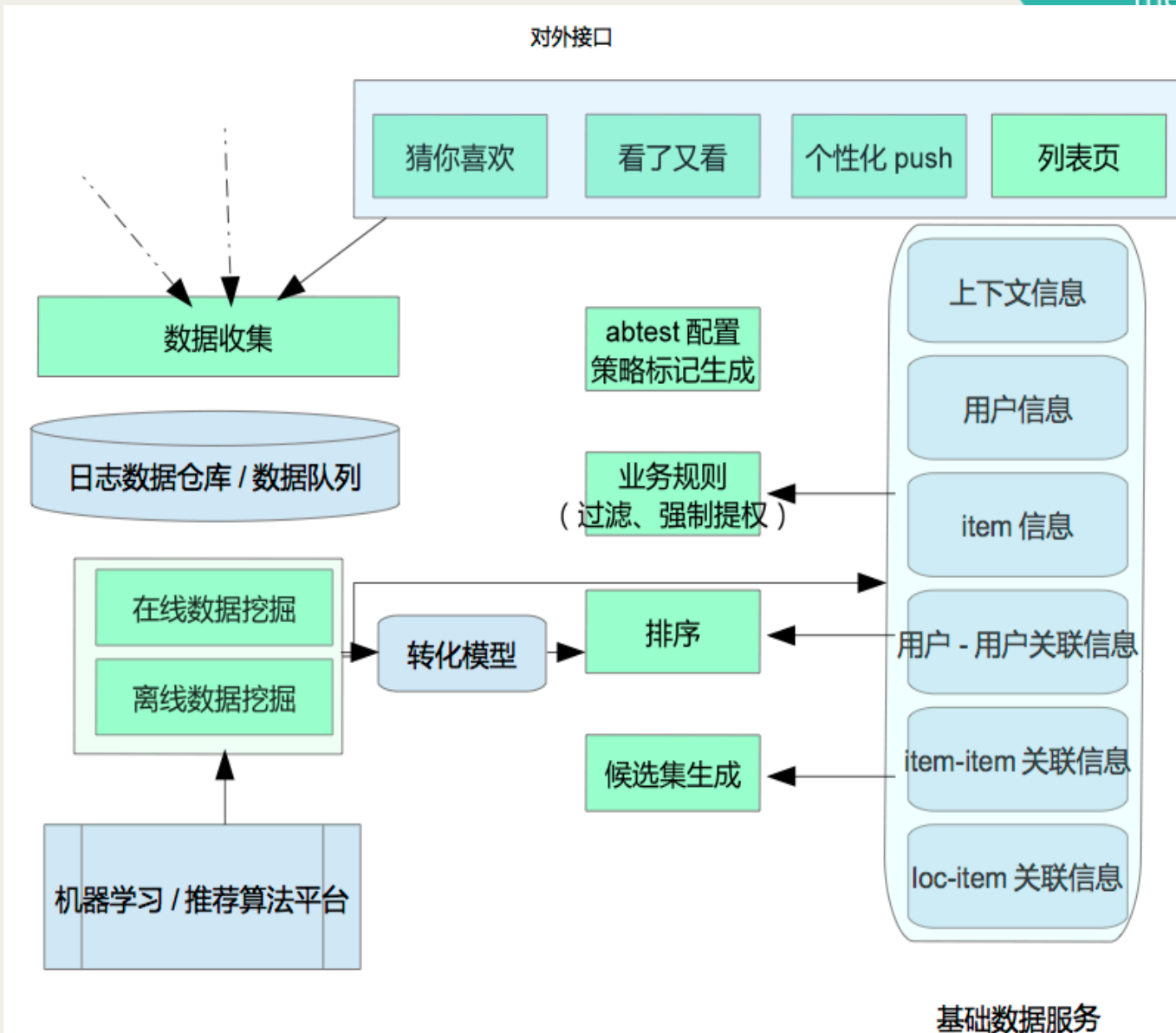


美团推荐产品的目标

- 快速找到所需
 - 高品质低价格（下单，消费后能够给出好评）
 - 品类丰富
- 推荐的核心指标
 - 下单率
 - 多样性

美团推荐系统框架

美团网
meituan.com



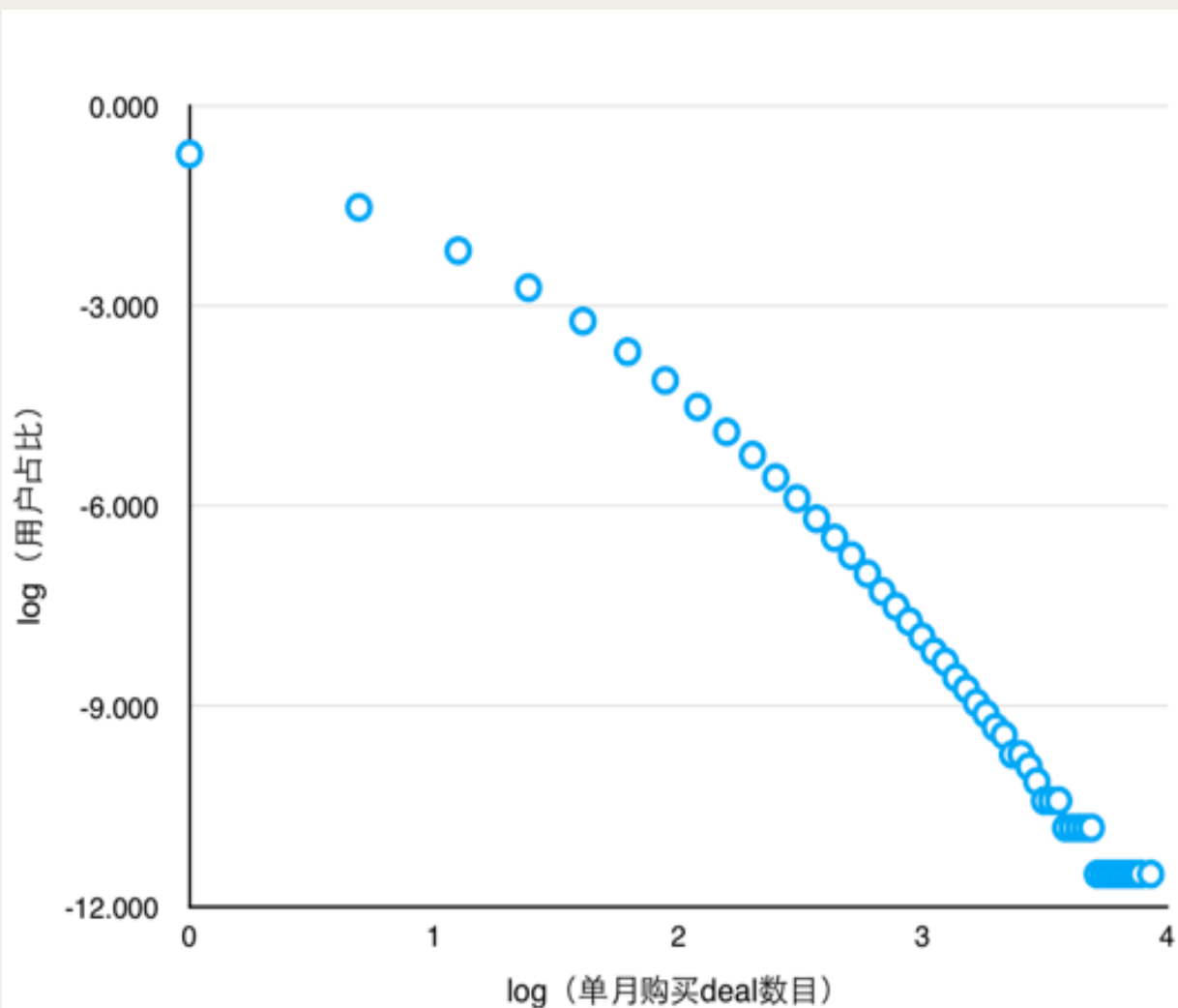
美团平台的特点

美团网
meituan.com

- 冷启动用户占比高
- 移动终端占比高
- 持券时间短
- 持券距离近

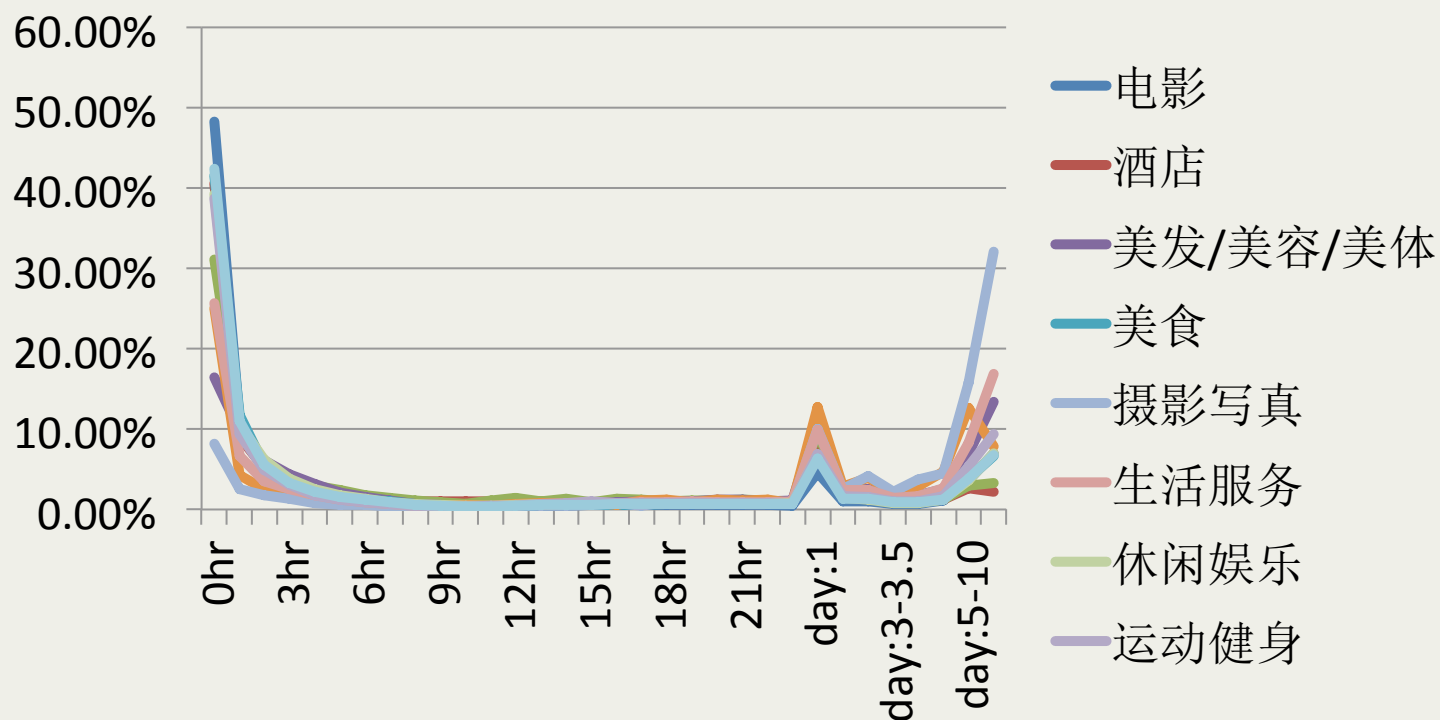
冷启动用户占比高

美团网
meituan.com

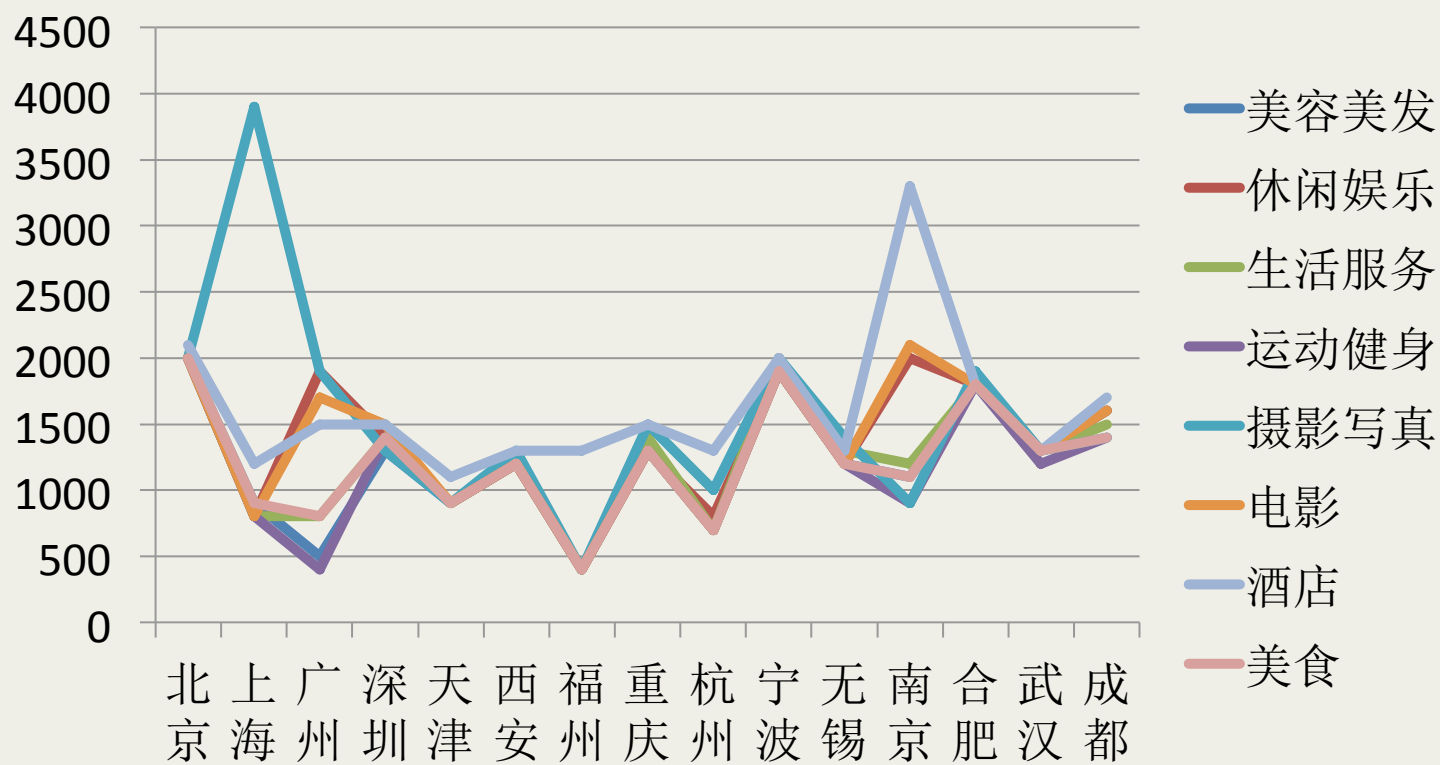


不同品类的持券时间分布

美团网
meituan.com



美团网
meituan.com



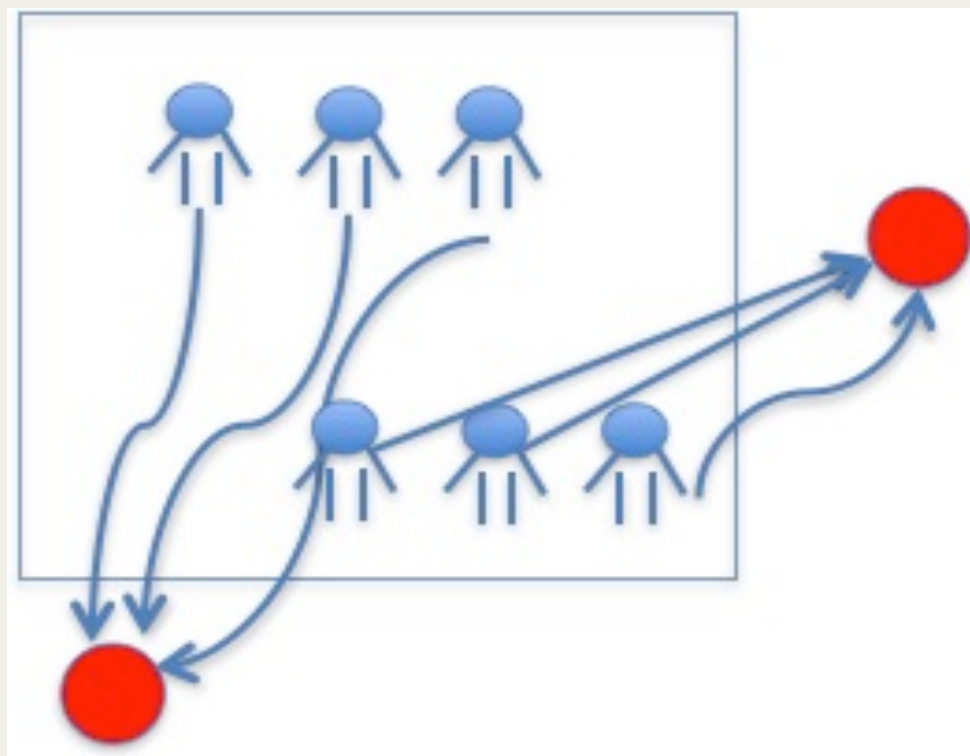
地理位置挖掘

- 地理位置特征
 - 实时地理位置
 - 周末/平时常去地理位置
 - 常去消费地
 - 工作地
 - 居住地
- 地理位置的粒度
 - 商圈

基于地理位置的冷启动

美团网
meituan.com

- 本地人热单



地理位置+时间上下文

- 计算时段之间的相似度
 - $\text{sim}(h[i], h[j])$
- 把相似时段的统计信息根据相似度加权统计到该时段中
 - $\text{newsum}(h[i], \text{item}, \text{loc}) = \text{sum}(h[i], \text{item}, \text{loc})$
 $+ \sum \text{sim}(h[i], h[j]) * \text{sum}(h[j], \text{item}, \text{loc})$

美团排序的关键工作

- 模型及建模
 - Additive Groves
 - 展示, 点击, 下单
- 样本采样及label处理
 - 点击: 下单 = 1:10
- 去除position bias
 - 位置p的平均下单率 cvr_p
 - item i在p位置的表现下单率 i_cvr_p
 - item i的实际下单率 - 这个值用来做特征
$$(i_cvr_0 + i_cvr_1 + \dots) / (cvr_0 + cvr_1 + \dots)$$

美团排序的关键工作

- 特征工程
 - 上下文特征：时间，地理位置，天气，温度等
 - Item特征：价格，销量，评分
 - 用户特征：年龄，性别，品类偏好，价格偏好等

Interleaving v.s. abtest

- abtest
 - 效果对比稳定性较差，一般需要一周的数据来验证效果
 - 可以获得定量数据
- Interleaving
 - 需要流量小，灵敏度高
 - 只能获得定性结论

Interleaving v.s. abtest

- Interleaving的列表生成方式

Rank	Input Ranking		Interleaved Rankings				
	A	B	Balanced		Team-Draft		
			A first	B first	AAA	BAA	ABA ...
1	a	b	a	b	a^A	b^B	a^A
2	b	e	b	a	b^B	a^A	b^B
3	c	a	e	e	c^A	c^A	e^B
4	d	f	c	c	e^B	e^B	c^A
5	g	g	d	f	d^A	d^A	d^A
6	h	h	f	d	f^B	f^B	f^B
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Interleaving v.s. abtest

- Interleaving的评估方法

$$\Delta_{AB} = \frac{wins(A) + \frac{1}{2}ties(A, B)}{wins(A) + wins(B) + ties(A, B)} - 0.5$$

wins(A) = 40%, wins(B) = 30%, tie = 30%

Result = 5% , A > B

Online Learning



- 互联网上的机器学习 vs 与针对静态物品的机器学习
- 用户群体的行为是在不断变化的
 - 季节
 - 天气, 空气质量
 - 社会潮流
 - 电视节目
 - 对算法的认知和反馈