

蘑菇街广告算法实践

—赵逸龙

蘑菇街
中国最大的女性时尚社交电商平台

提纲

- 蘑菇街业务概述
- 广告检索排序的技术架构
- 业务挑战及其背后的技术思考
- 未来展望

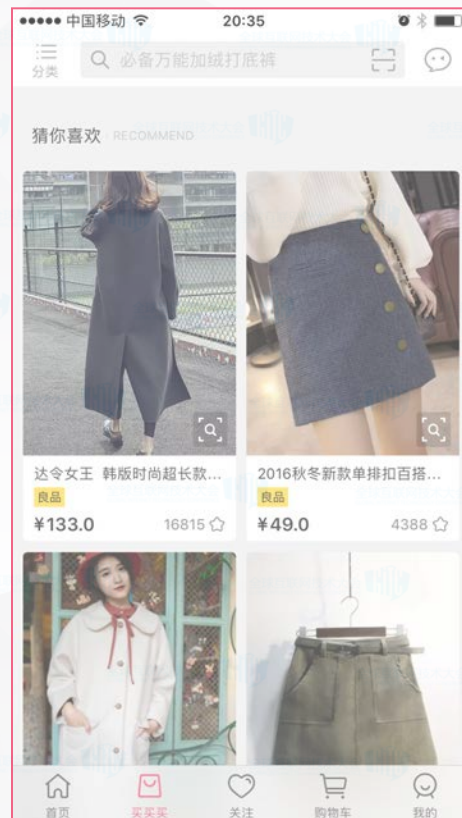
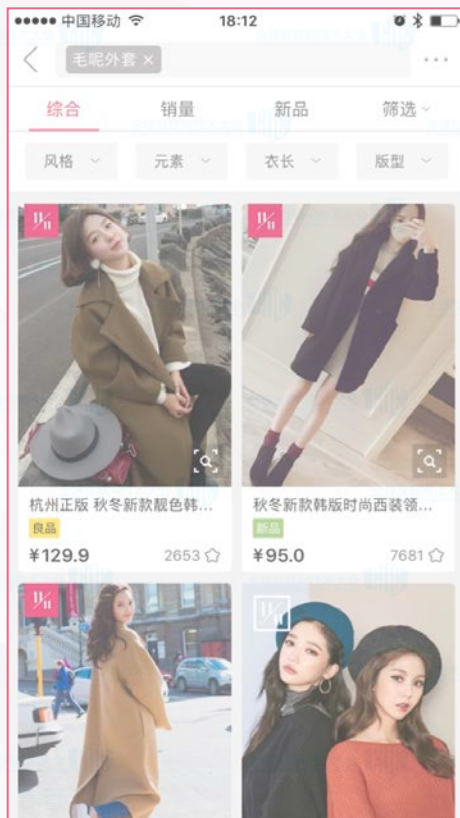
提纲

- 蘑菇街业务概述
- 广告检索排序的技术架构
- 业务挑战及其背后的技术思考
- 未来展望

蘑菇街的业务特性

- 用户“逛街”式访问占比高
- 移动端占比超过90%，用户浏览量大
- 商品存在明显的时效性和季节周期性
- 商家线下库存压力大

蘑菇街的广告业务



提纲

- 蘑菇街业务概述
- 广告检索排序的技术架构
- 业务挑战及其背后的技术思考
- 未来展望

广告检索排序的技术架构

A/B Test

统一投放

UPS

FlowCTL

ZooKeeper

Query
Rewriter

广告检索

Spark

ASAP

HBase

Hive

HDFS

Kafka

提纲

- 蘑菇街业务概述
- 广告检索排序的技术架构
- 业务挑战及其背后的技术思考
- 未来展望

业务挑战及其背后的技术思考

- 提升排序效率
- 缓解流量波动
- 维护生态健康

始于经验公式

- 统计平滑
 - 广告 / 自然
 - 搜索 / 类目
 - 商品 / 店铺
- 置信度区间

$$\frac{1}{1 + \frac{1}{n}z^2} \left[\hat{p} + \frac{1}{2n}z^2 \pm z \sqrt{\frac{1}{n}\hat{p}(1 - \hat{p}) + \frac{1}{4n^2}z^2} \right]$$

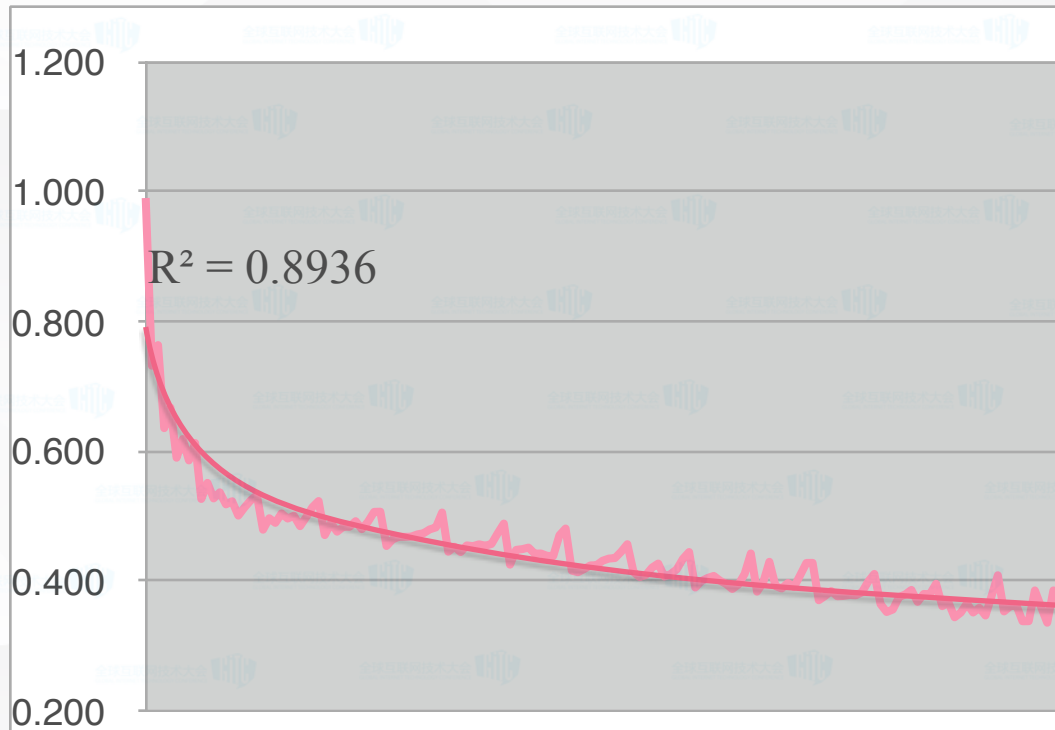
(Wikipedia: Binomial Proportion Confidence Interval)

迈向机器学习

- 击败统计经验公式
 - 统计特征 vs 泛化特征
 - 老广告 vs 新广告
- 从XGBoost到XGBoost on Spark
- 近线模型更新 + 在线模型预测

(XGBoost: A Scalable Tree Boosting System)

减轻位置偏置



业务挑战及其背后的技术思考

- 提升排序效率
- 缓解流量波动
- 维护生态健康

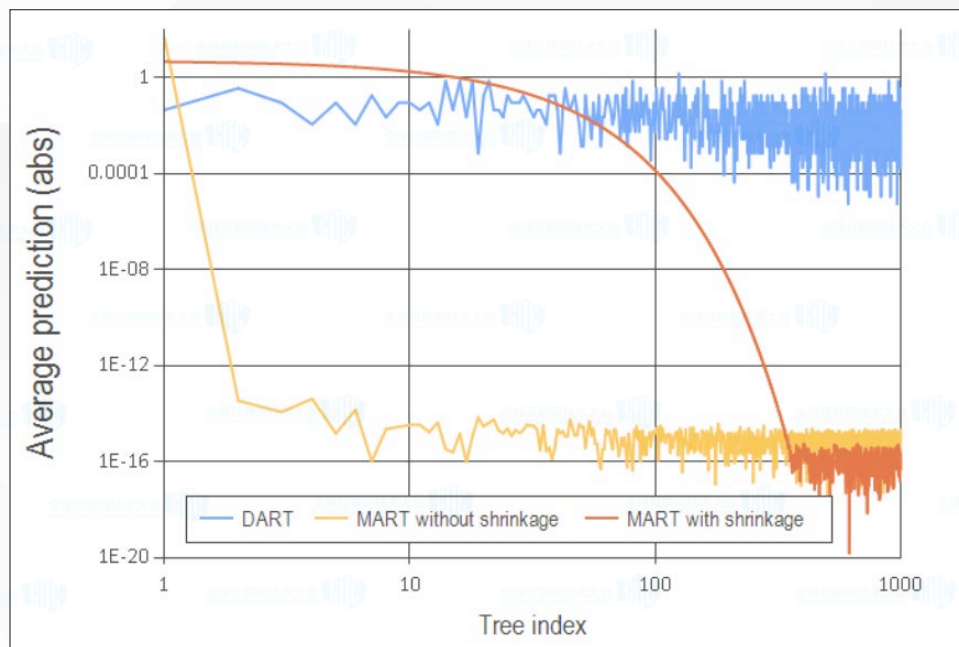
流量波动的原因

- 点击率的天然变化
- 点击率预估模型的波动
 - 广告从不置信到置信
 - 模型并不保证局部稳定性
- 训练数据异常
 - 作弊行为数据
 - 数据链路上的其他异常

一次突发的局部排序波动

- 症状
 - 部分商家反馈广告排序突然下滑
- 诊断
 - 特征分布 / 特征重要性排序正常
 - 训练和校验集上的AUC和RMSE等指标正常
 - 影响范围存在一定的随机性
- 病灶
 - 客户端打点逻辑变更，在微量灰度时引起局部特征值异常

Shrinkage参数

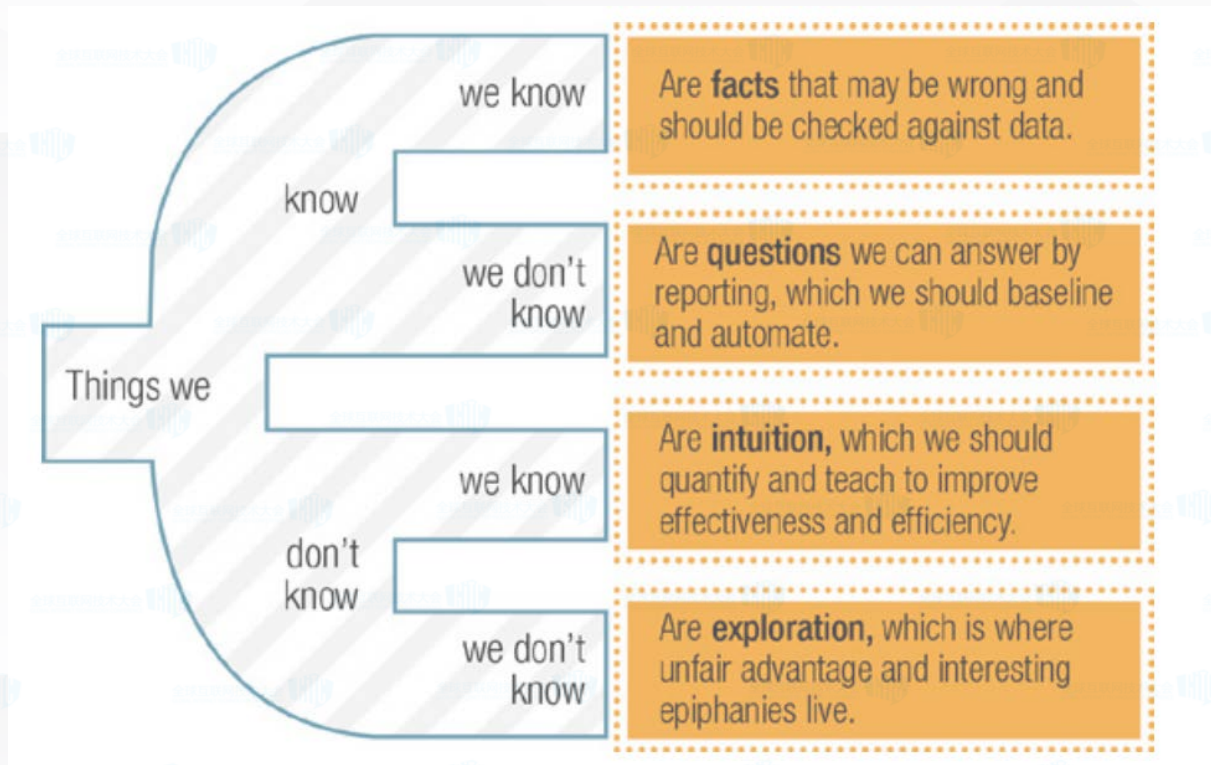


(DART: Dropouts meet Multiple Additive Regression Trees)

业务挑战及其背后的技术思考

- 提升排序效率
- 缓解流量波动
- 维护生态健康

一个从0到100的渐进思考框架



(Lean Analytics: Use Data to Build a Better Startup Faster)

关注系统可解释性与长期收益

- 以CTR的定义为例
 - $\text{CTR} = \text{全部UV的总点击} / \text{全部UV的总曝光}$
 - $\text{CTR} = \text{单个UV的点击率的平均值}$
- 以流量分配为例
 - 从商家侧来看，广告流量的分配和自然搜索对比有何异同
 - 从商品侧来看，新商品的冷启动速度和新商品的占比如何

(Data-Driven Metric Development for Online Controlled Experiments: Seven Lessons Learned)

探索与利用(Explore & Exploit)

- EE的收益和代价
- 反作弊
 - 进入和退出机制
 - Explore流量配额控制
- 效果评估

提纲

- 蘑菇街业务概述
- 广告检索排序的技术架构
- 业务挑战及其背后的技术思考
- 未来展望

未来展望

- 点击率模型的范化能力与时效性
- 图像特征的利用
- 预算均匀消耗
- 探索与利用机制的健全

A background image of two women sitting at a table in a cafe, laughing and clapping their hands. The table is set with two cups of coffee, a wicker basket, and plates of pastries. A large pink circle on the right side of the image contains the Mogujie logo and text.

Q & A

蘑菇街
mogujie.com
我的买手街