

天猫推荐业务与算法架构

张奇（得福） 天猫推荐算法团队

2013-07-14

天猫 TMALL.COM



个人介绍



张奇，花名：得福

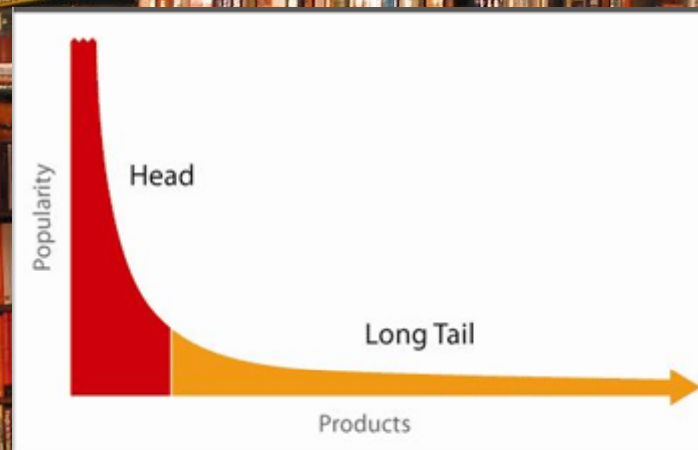
2010年博士毕业于中国科学技术大学

2010年至2012年，计算广告@阿里云

2012.4 至今，推荐系统@天猫



Search Engine



第二部分：user2item

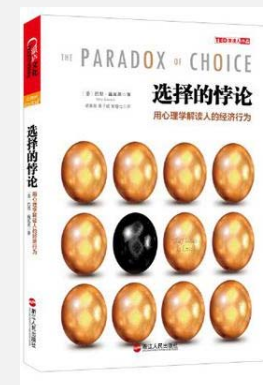
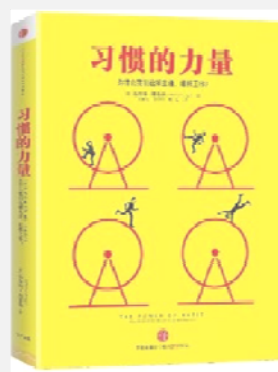
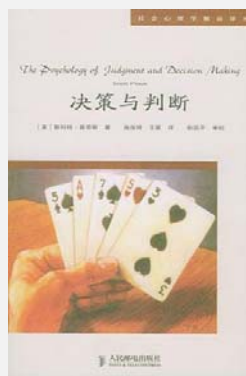
第三部分：推荐在天猫的应用

Recommendation

第一部分：item2item

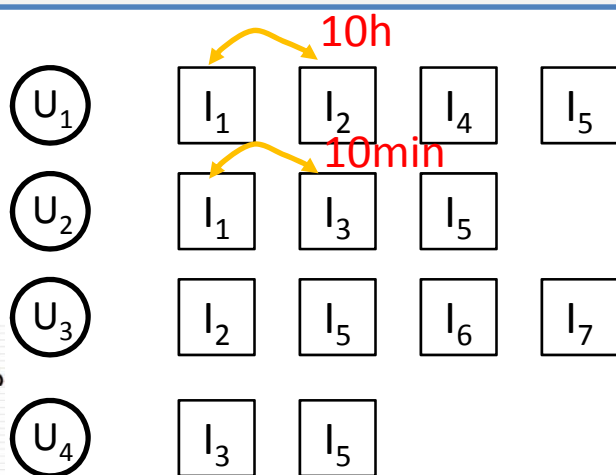
Begin with

浏览了该商品的用户还浏览了



天猫 TMALL.COM

Item based Recommendation



$$w_{ij} = \frac{N_i \cap N_j}{N_i}$$

$$w_{ij} = \frac{N_i \cap N_j}{|N_i|^{1-\alpha} |N_j|^\alpha}$$



浏览 (V2V) 购买 (B2B)

$$w_{ij} = \frac{\sum_{u \in U(I_i) \cap U(I_j)} W_u^2}{\sqrt{\sum_{u \in U(I_i)} W_u^2} \sqrt{\sum_{u \in U(I_j)} W_u^2}}$$

$$w_{ij} = \frac{\sum_{u \in U(I_i) \cap U(I_j)} (W_u^2 \times f(|t_{ui} - t_{uj}|))}{\sqrt{\sum_{u \in U(I_i)} W_u^2} \sqrt{\sum_{u \in U(I_j)} W_u^2}}$$

$$\text{TextSim}(I_i, I_j) = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

Item based 算法效果

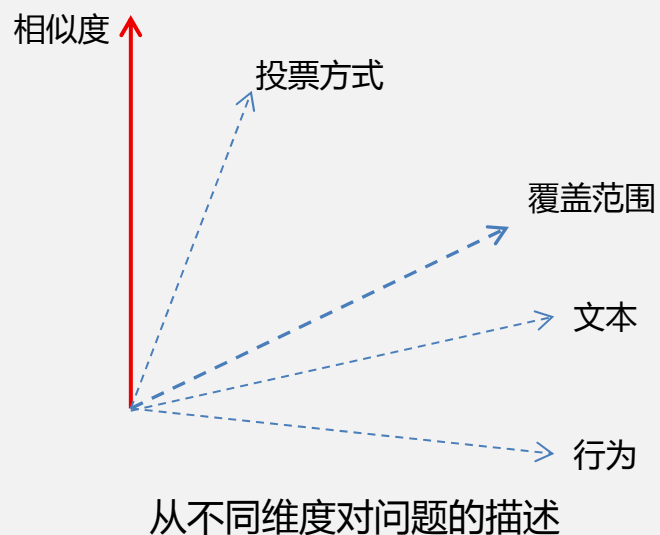
$$Jaccard = \frac{I_i \cap I_j}{I_i \cup I_j}$$

$$CosSim(I_i, I_j) = \frac{\sum_{u \in U(I_i) \cap U(I_j)} (W_u^2 \times f(|t_{ui} - t_{uj}|))}{\sqrt{\sum_{u \in U(I_i)} W_u^2} \sqrt{\sum_{u \in U(I_j)} W_u^2}}$$

	UV点击率	UV购买率	点击率*购买率
Jaccard	4.14%	2.42%	0.100%
CosSim	4.37%	2.45%	0.107%
提升比	5.53%	1.40%	7.00%



Model 融合算法



Model 融合问题

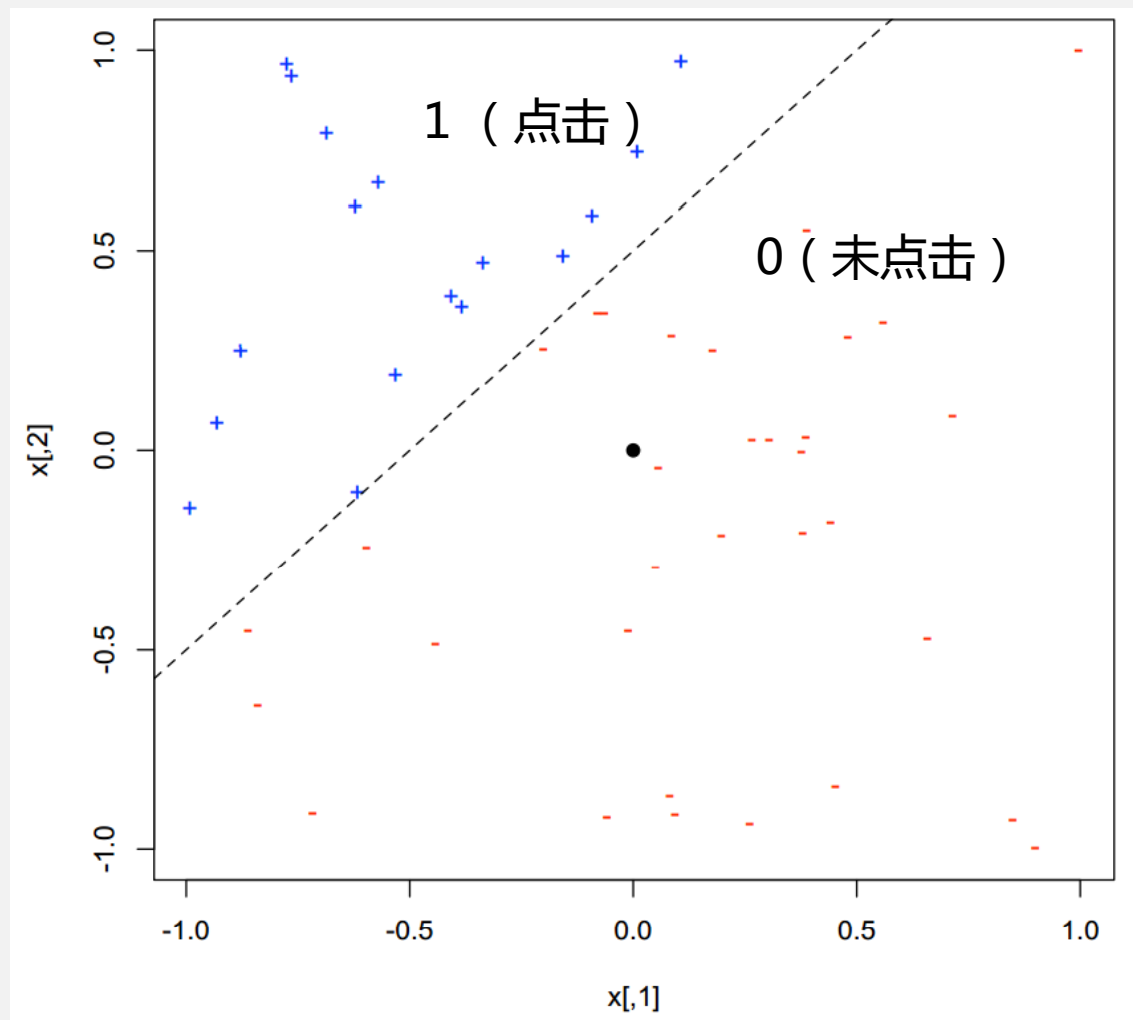


主商品	候选集	v2b	b2b	v2v(cf1)	v2v(cf2)	target
Item	Item ₁	0.021	0.51	0.47	...		0.53
	Item ₂			0.49			0.48
	Item ₃			0.33			0.45

	Item _n			0.58			0.21



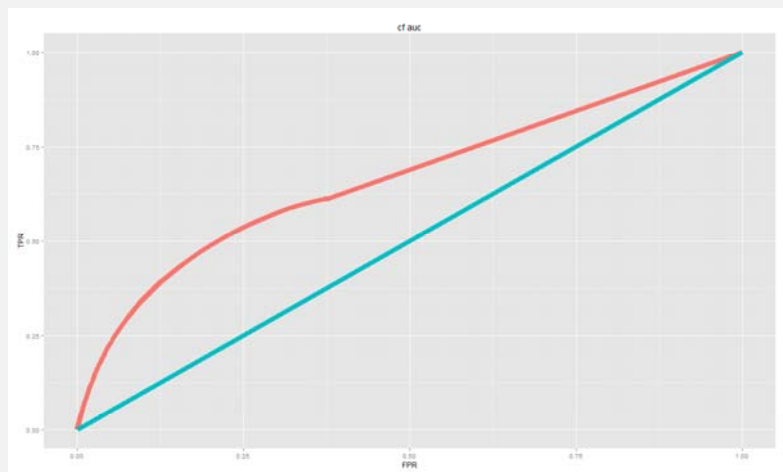
Model 融合算法-Machine Learning



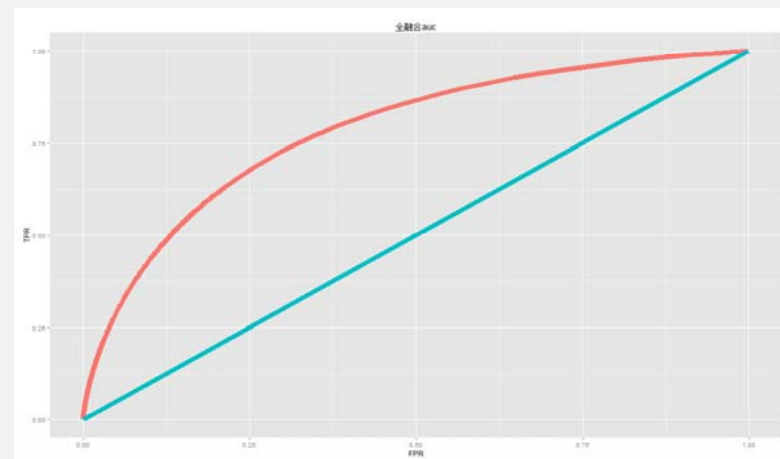
Logistic Regression @MPI



Model 融合算法-效果



单一CF模型 AUC= 0.6653

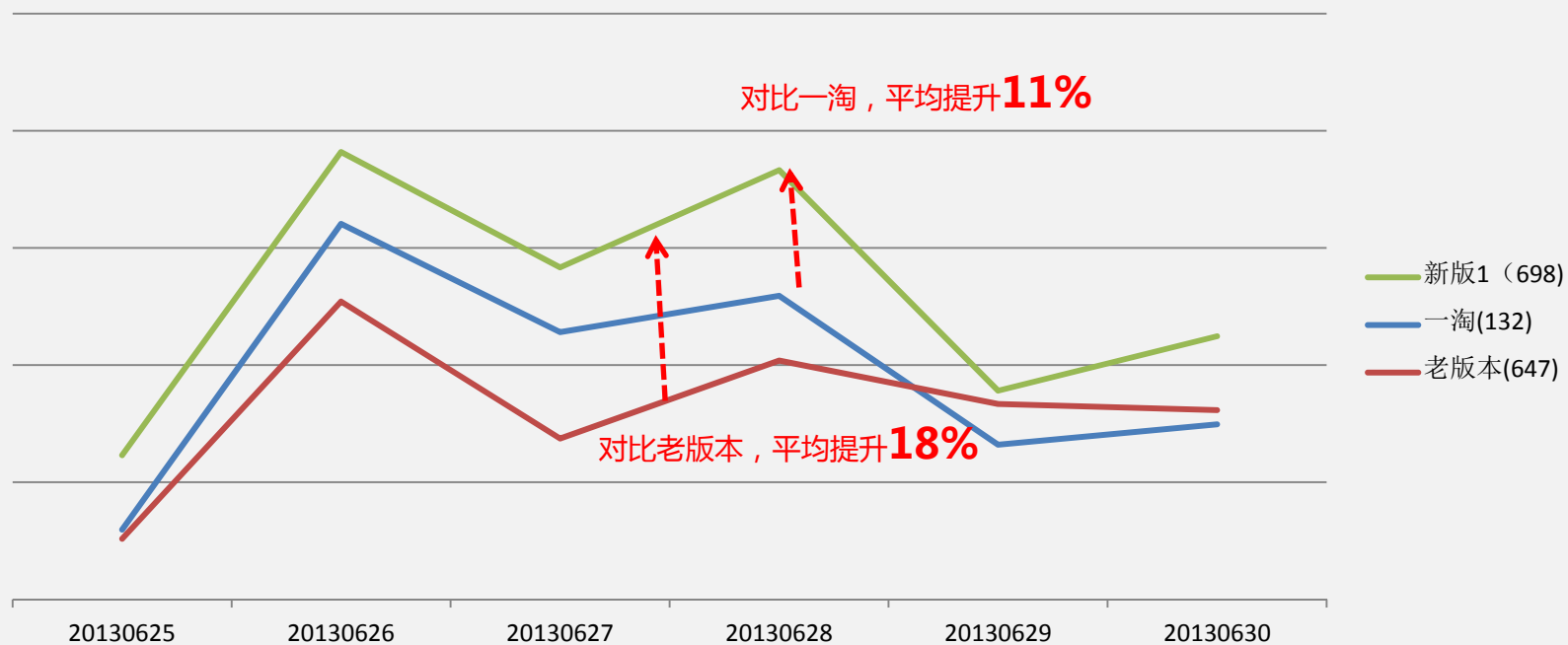


融合后模型 AUC= 0.7835



Model 融合算法-效果

Item2item 算法优化效果 (PV 曝光价值)



$$pv \text{ 曝光价值} = \frac{\text{支付宝成交金额}}{\text{总访问 PV}}$$



Part-1 小结

基础推荐实体关系库

商品关系库 品牌关系库 专辑关系库 活动关系库



离线算法系统

训练	训练数据准备		特征计算	参数调优
算法包	LR	MLR	LDA	NMF
集群	MPI 集群		Hadoop 集群	



猜你喜欢

?

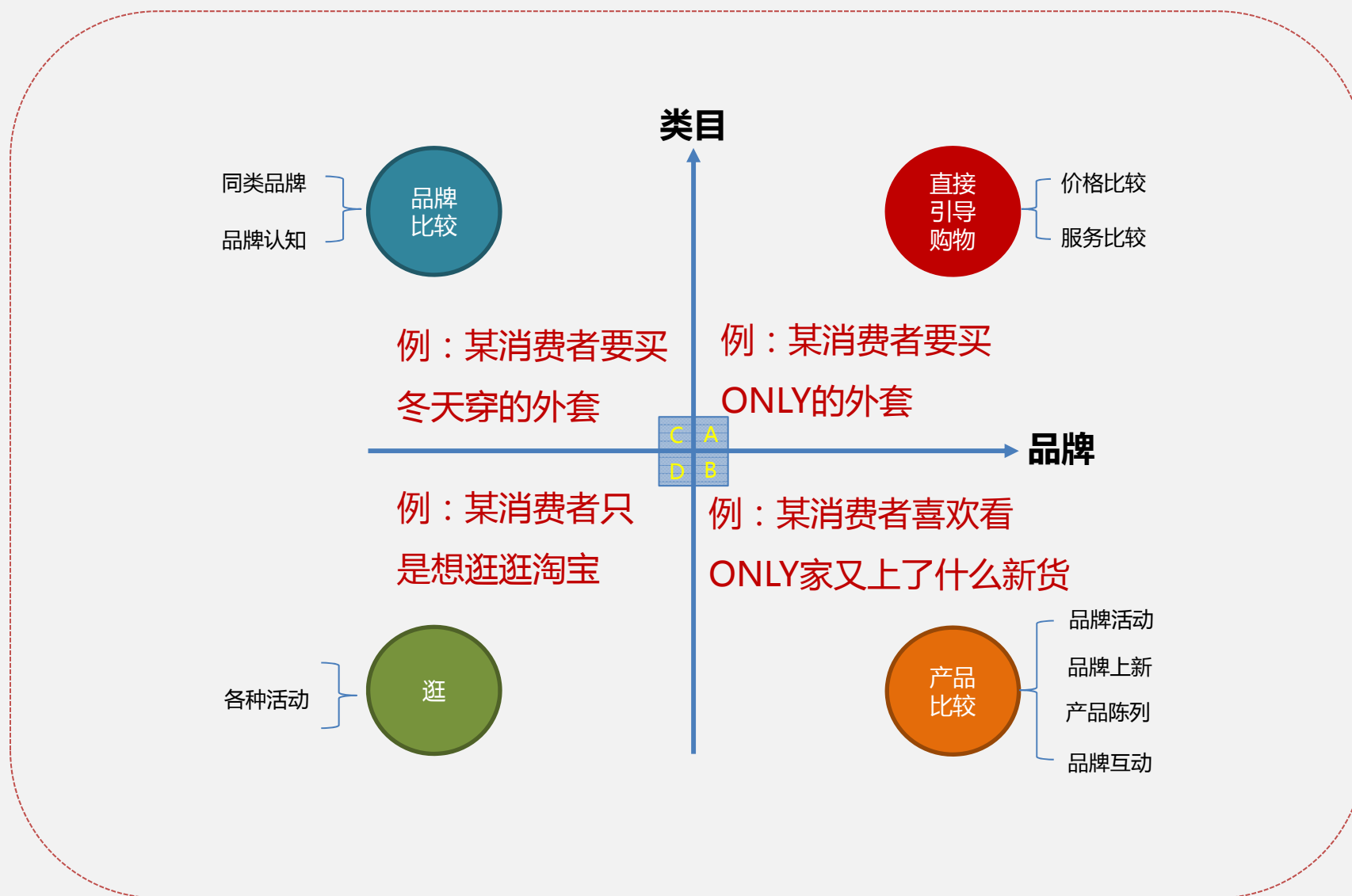
?

?

?



Personalization-用户购物意图模型



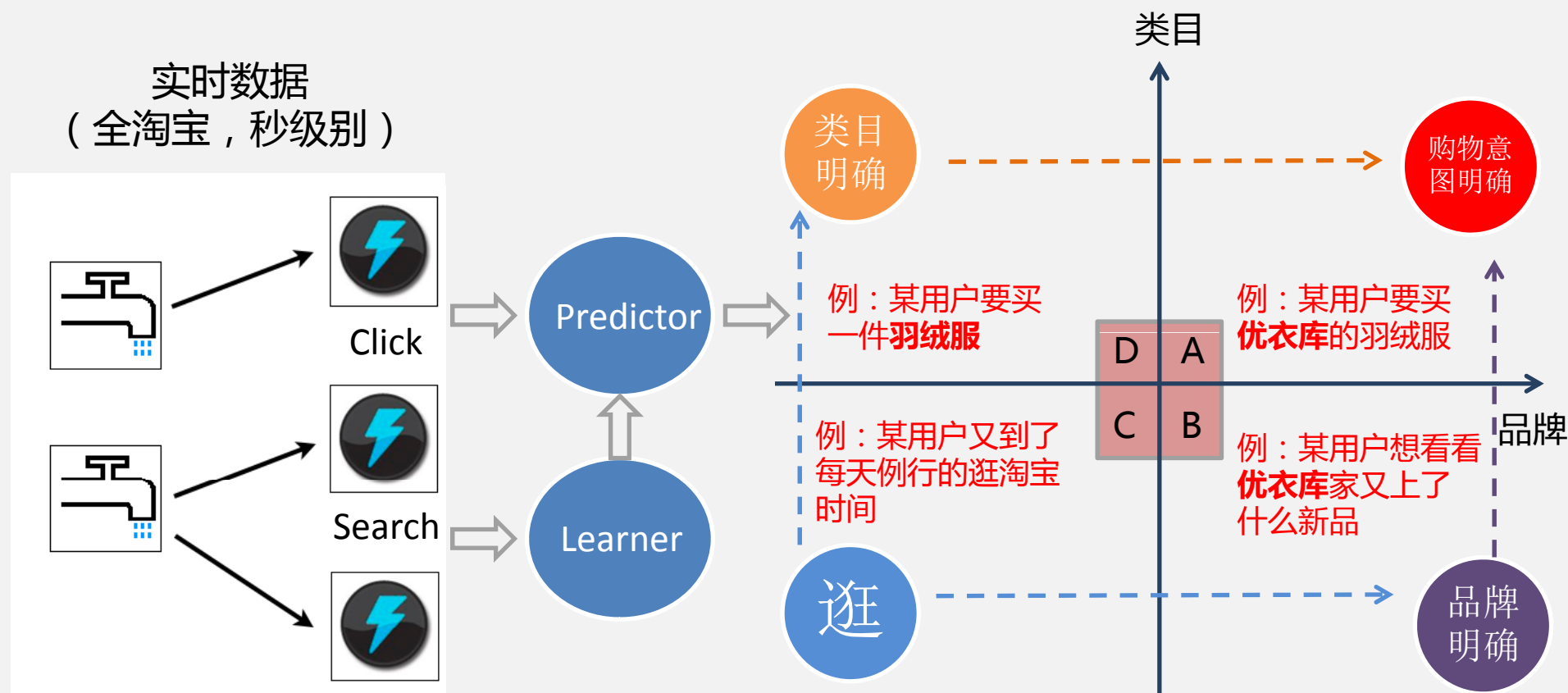
Personalization-场景引擎-模型

目标：基于用户实时行为和天猫的导购路径结构，判断用户当前的意图

意图建模（三维模型）：品牌意图、类目意图、单品意图

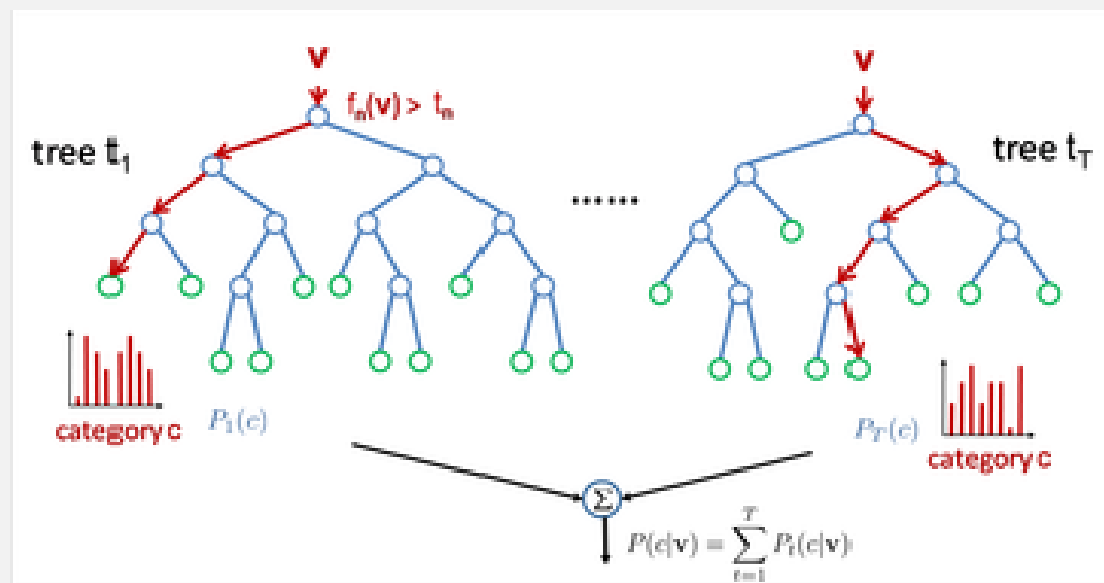
特征刻画：利用CPV、UIT 等数据刻画类目意图

技术框架：



Personalization-场景引擎-Paper

Making Recommendations Better: The Role of User Online Purchase Intention Identification, RecSys 2013 Hongkong (Submitted)



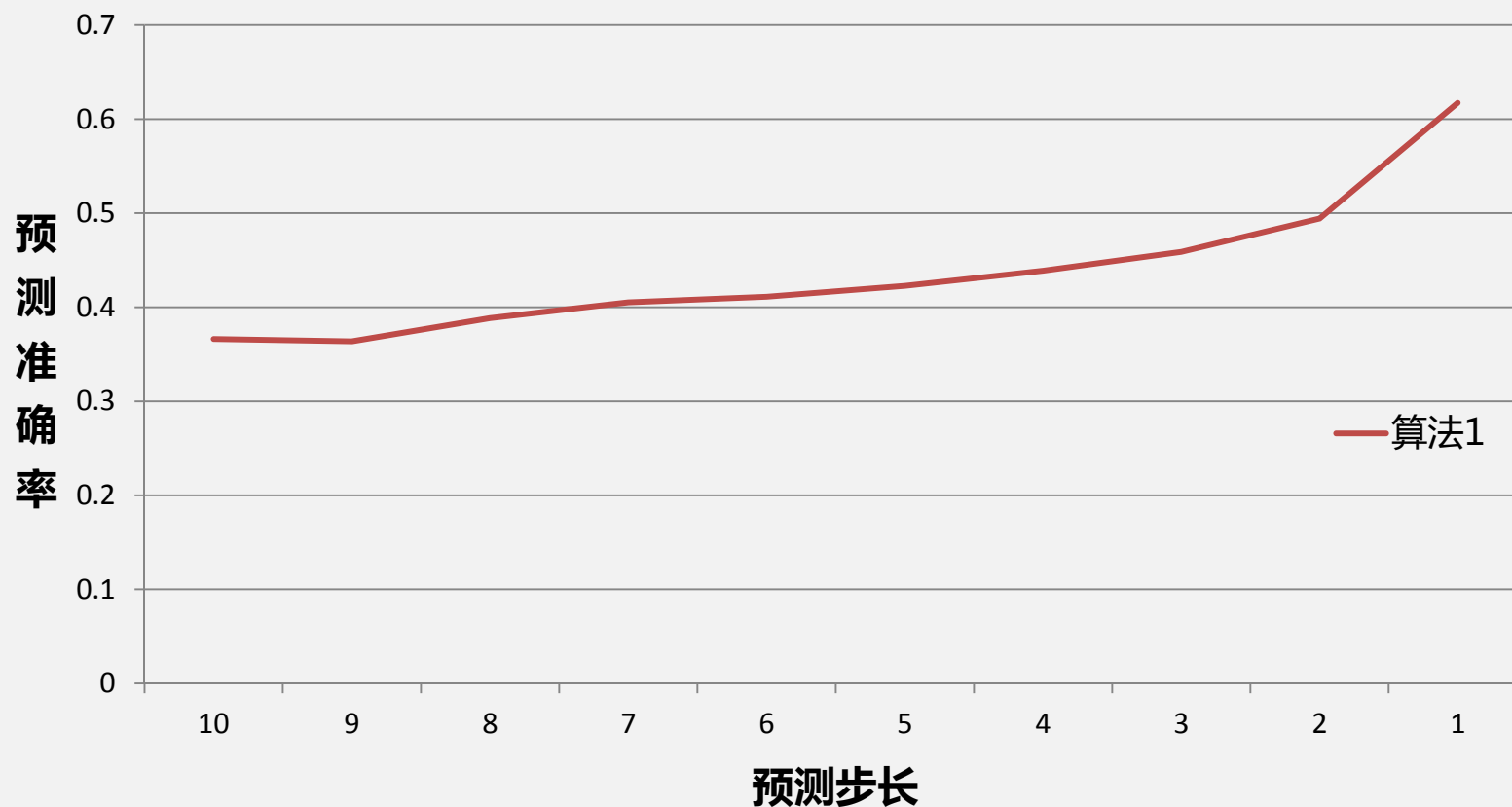
Random Forrest



天猫 TMALL.COM

Personalization-场景引擎-评估

场景引擎效果离线评估方法

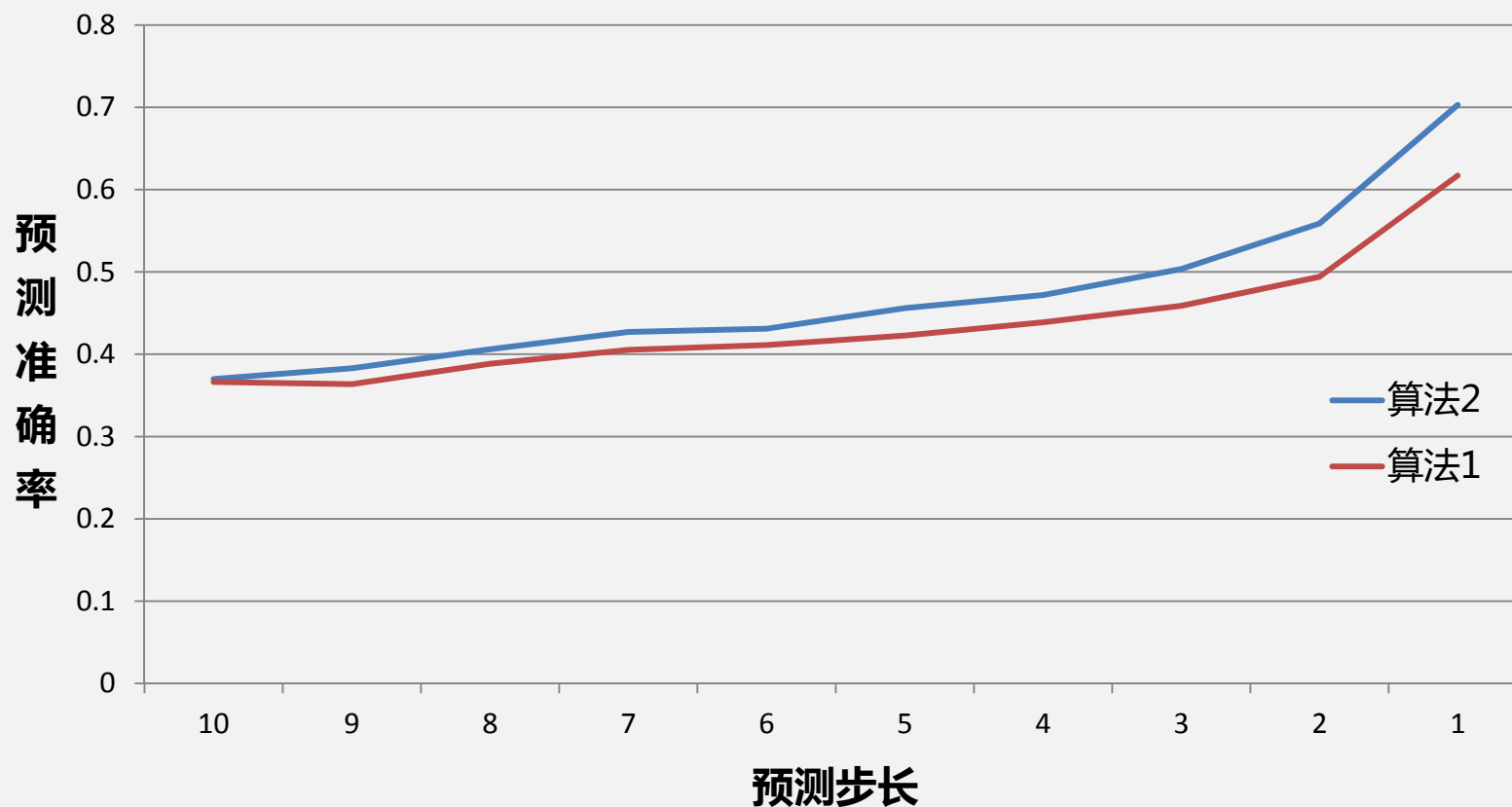


user session:



Personalization-场景引擎-评估

场景引擎效果离线评估方法



user session:



算法系统-场景引擎-DEMO

8231****

开始用户追踪

停止追踪

用户实时点击数据流

访问时间	商品标题	商品价格
2013-04-19 11:10:26	apple\苹果macbookairmd224ch\va11超薄笔记本电脑大陆行货	7225.56
2013-04-19 11:10:26	dell\戴尔ins15r-25182618经济版15rd-351815rr-3518笔记本	4331.00
2013-04-19 11:10:26	lenovo\联想z500a-ifi15-3230merazer异能者z500超薄笔记本	4440.91
2013-04-19 11:10:26	apple\苹果macbook promd101ch\va13笔记本电脑新款大陆行货	7810.95
2013-04-19 11:06:21	zimmur2013夏季新款时尚名媛气质修身欧根纱蕾丝连衣裙夏0121f0	259.01

实时意图判断

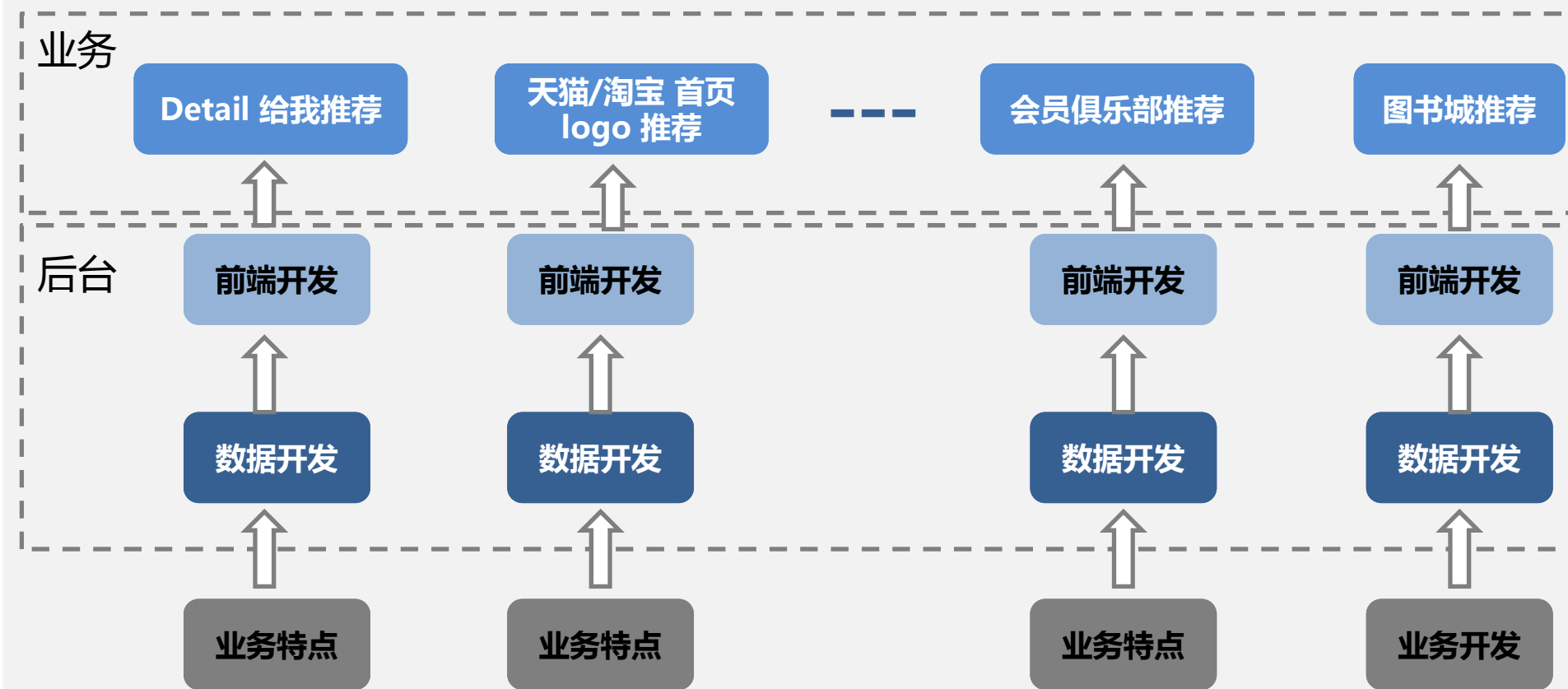
	场景	用户想买的类目(笔记本电脑), 但不确定具体的品牌
	品牌	笔记本电脑_品牌_Dell/戴尔 笔记本电脑_品牌_Apple/苹果 跑步鞋_品牌_New Balance/新百伦 跑步鞋_品牌_ANTA/安踏 跑步鞋_品牌_Asics/亚瑟士 连衣裙_品牌_Zimmur
	导购节点	TmallList<-- TmallDetail<-- TmallDetail<-- TmallDetail<-- TmallDetail<-- TmallDetail<--
	笔记本电脑	weight:100.00
	"笔记本电脑"的标签	苹果(0.30) 超薄(0.20) 联想(0.20) 大陆行货(0.19) dell(0.13)
	"笔记本电脑"的CPV	笔记本电脑_品牌_Dell/戴尔(0.80) 笔记本电脑_笔记本CPU_Intel Core/酷睿 i5(0.69) 笔记本电脑_品牌_Apple/苹果(0.60) 笔记本电脑_产品上市时间_2012年上半年(0.55)
	跑步鞋	weight:1.51
	"跑步鞋"的标签	安踏(0.40) 网面(0.21) 阿斯科斯(0.20) 专业缓冲(0.17) 全掌气垫(0.15)
	"跑步鞋"的CPV	跑步鞋_运动服性别_男性(0.66) 跑步鞋_颜色分类_黑色(0.58) 跑步鞋_上市年份_2012年春季(0.49) 跑步鞋_价格区间_501-800元(0.40)

天智 TRACON

算法组装



天猫推荐系统架构@2013之前



问题：业务和算法高度结合
高投入、低沉淀

天猫推荐系统架构@2013

Detail 给我推荐

天猫/淘宝 首页
logo 推荐

会员俱乐部推荐

图书城推荐

前端

匹配层

装配 (参数化)

Ranker

检索

场景引擎

用户实时意
图&长期兴
趣

在线实验
框架

推荐实体

商品推荐

同店 | 跨店

相似 | 搭配

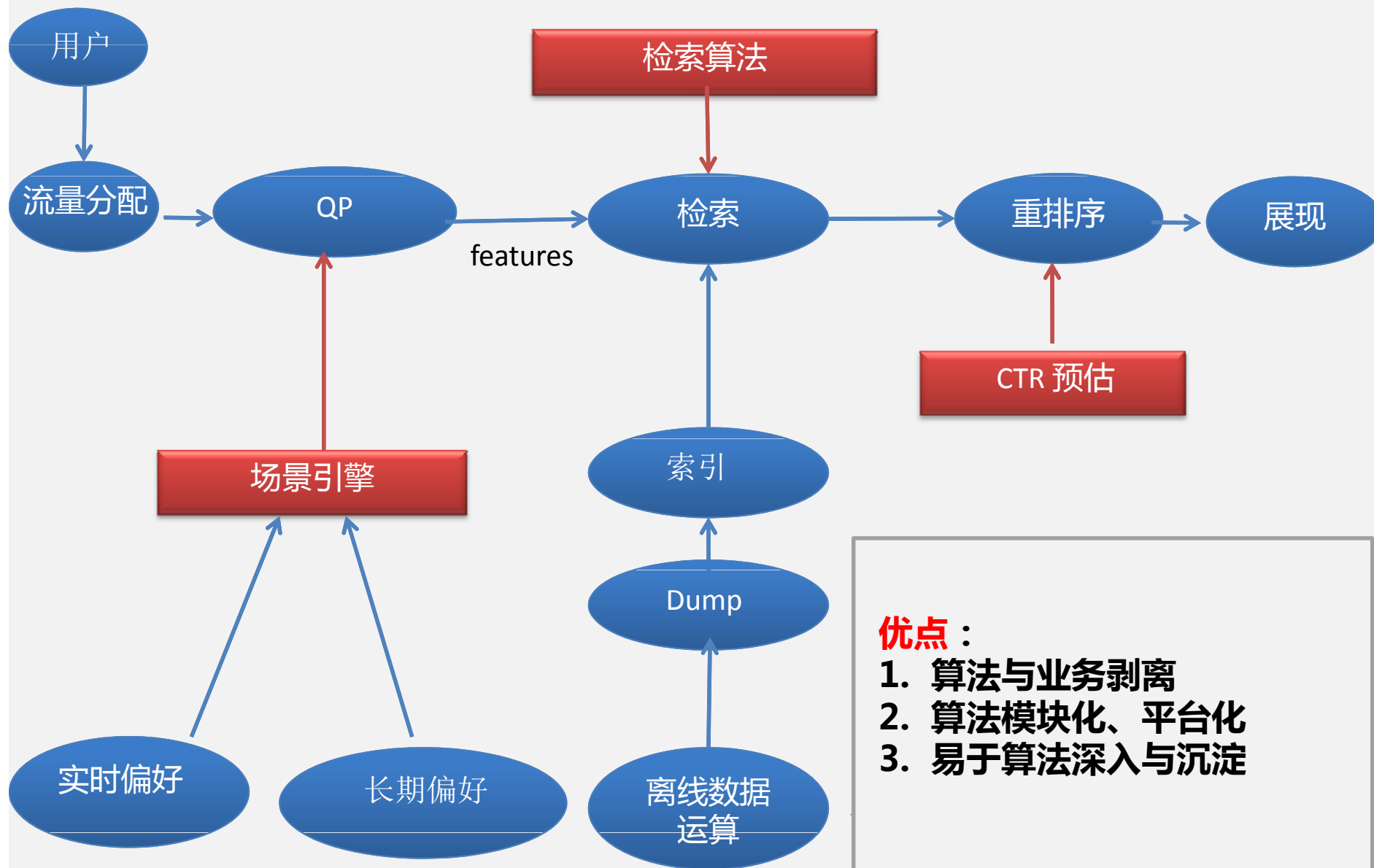
品牌推荐

相似

专辑推荐

离线实验
框架

天猫推荐流程图



优点：

1. 算法与业务剥离
2. 算法模块化、平台化
3. 易于算法深入与沉淀

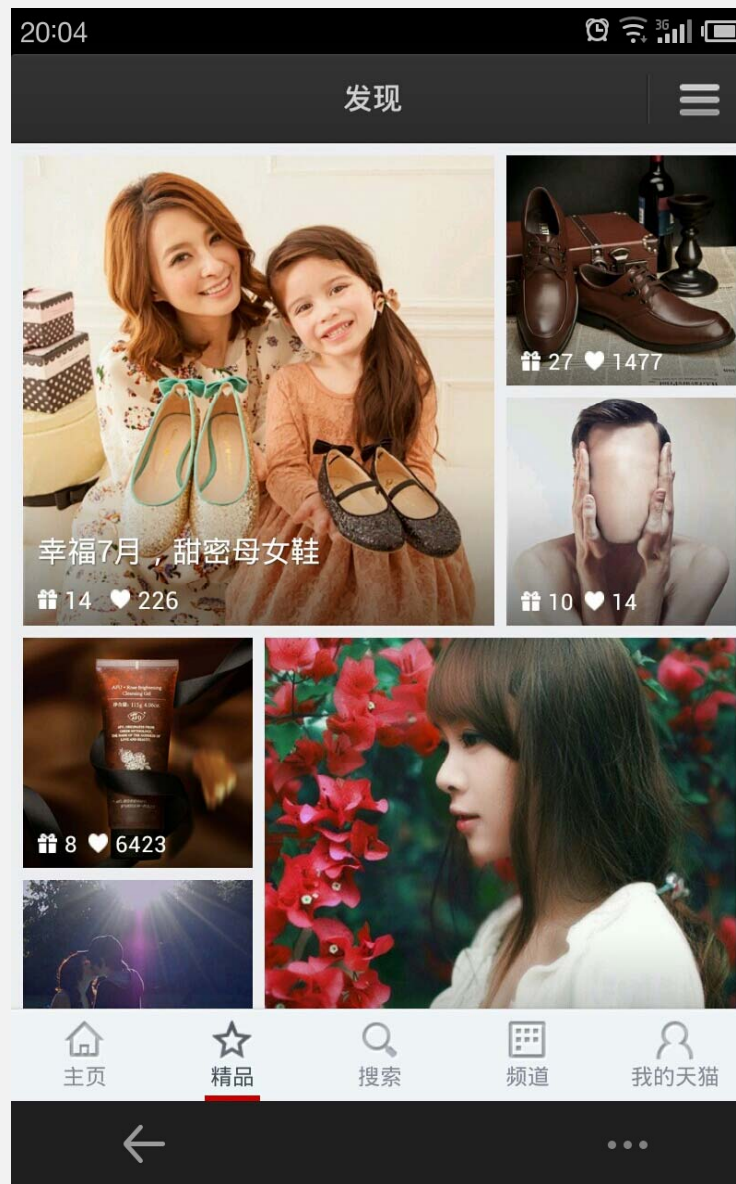
推荐应用@Tmall



天猫推荐业务一览图（PC端）

tmall改版品牌推荐
成功付款页 迎客松
搜索无结果商品推荐
左侧类目快捷导航推荐 List相似推荐
error页商品推荐 店铺宝贝推荐模块
专辑推荐 品牌导航品牌推荐
天猫购物车 旺旺 cspu页
list右侧热销店铺 确认收货页
“给我推荐”Tab 订单详情页
大家都喜欢（撕开层） 天猫首页楼层区活动推荐
detail新版瀑布流推荐
天猫首页小焦推荐 spu左侧SPU推荐
专辑相似推荐
猜你喜欢（旺铺模块） 已下架商品推荐

天猫推荐业务一览图（无线端）



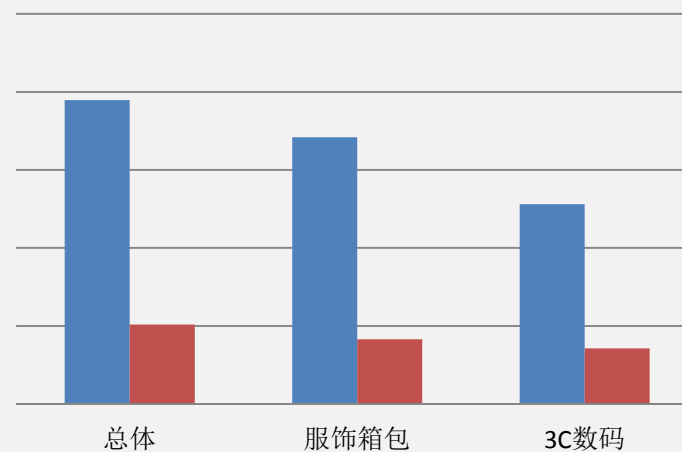
推荐对消费者的意义

让消费者更容易找到满意的商品

UV购买转化率



人均点击商品数

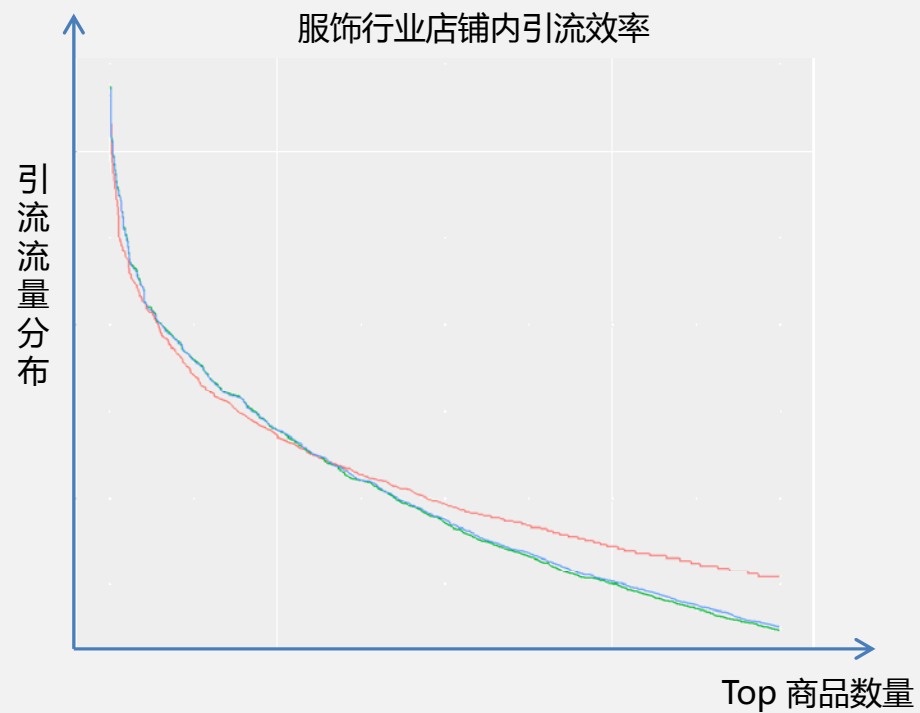
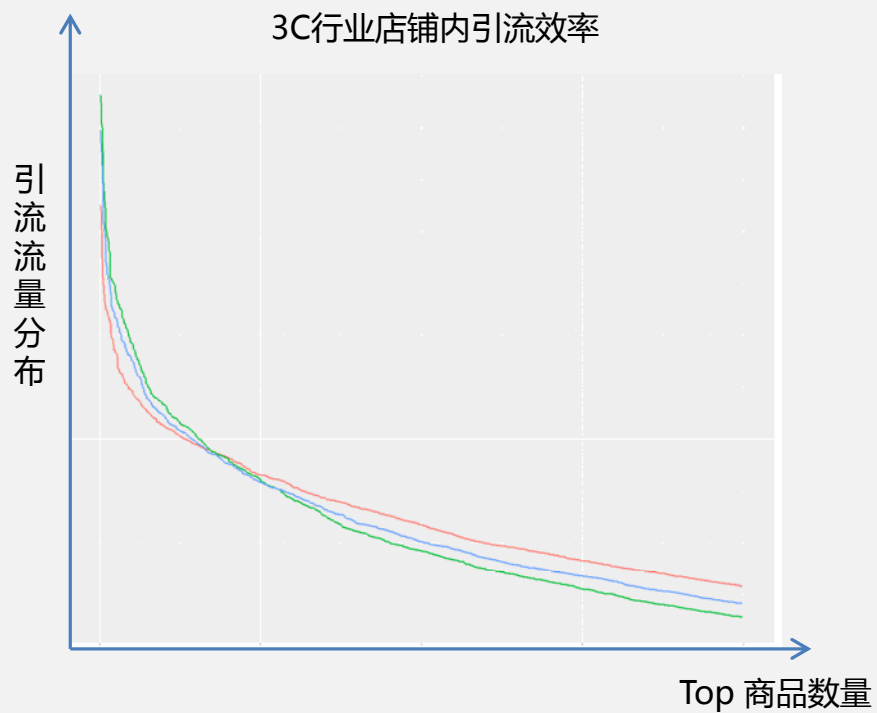


— 当日使用推荐的人群 — 当日未使用推荐的人群



推荐对卖家的意义

让流量的分配更公平（卖家）



— 推荐 — 搜索 — 总体



推荐对天猫的意义

15%



MPI 集群&机器学习算法包

- LR
- MLR

数据资源

- UIT
- TCIF

