

SACC

卓越 5周年 变迁

SequeMedia SACC2013 ChinaUnix IT PUB

2013中国系统架构师大会

SYSTEM ARCHITECT CONFERENCE CHINA 2013

大数据下的IT架构变迁

天猫推荐业务与算法架构

张奇（得福） 天猫推荐算法团队



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

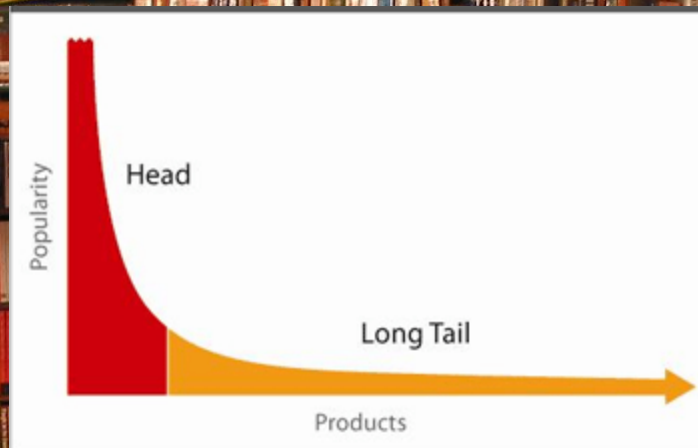
2012.4 至今，推荐算法@天猫

新浪微博：张奇_得福

邮箱：john.zhangq@tmall.com

SACC2013

Search Engine



第二部分：user2item

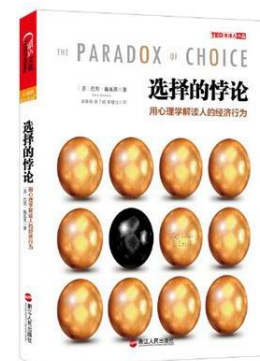
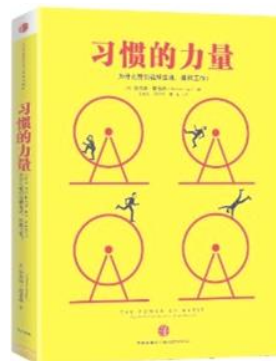
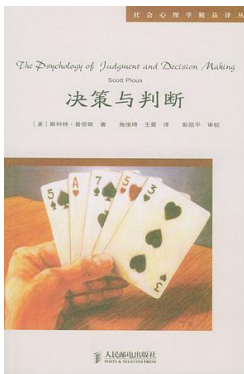
第三部分：推荐在天猫的应用

Recommendation

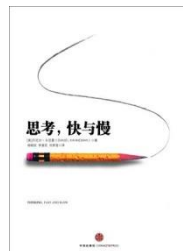
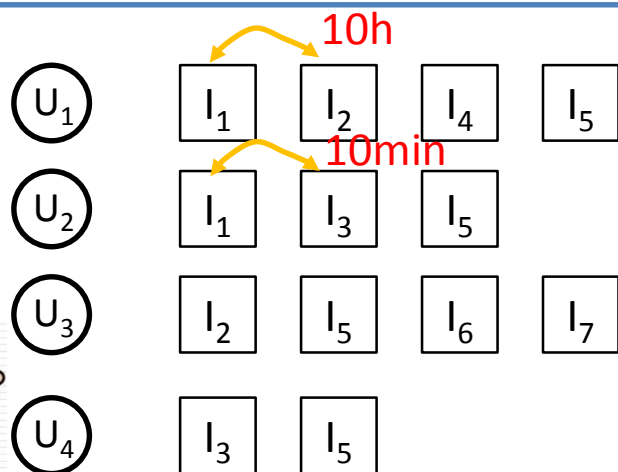
第一部分：item2item



浏览了该商品的用户还浏览了



Item based Recommendation



SACC2013 哈利·波特与死亡圣器



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

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

浏览 (V2V) 购买 (B2B)

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

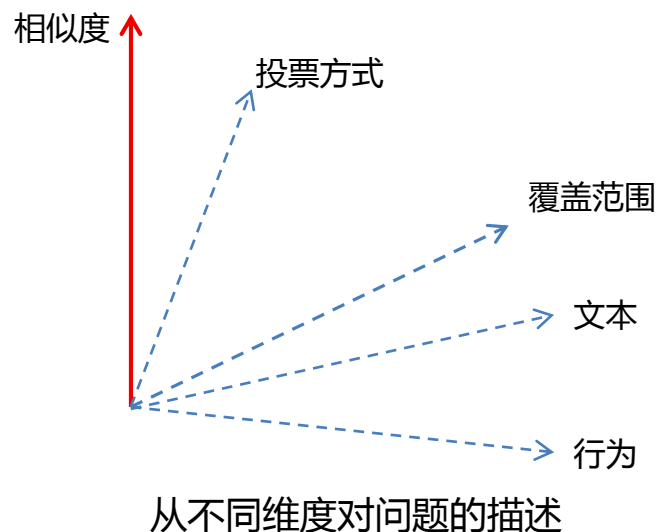
$$w_{ij} = \frac{\sum_{u \in U(I_i) \cap U(I_j)} (W_u^2 \times f(|t_{ui} - t_{uj}|))}{\sqrt{\sum_{u \in U(I_i)} W_u^2} \sqrt{\sum_{u \in U(I_j)} W_u^2}}$$

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

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

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

	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%

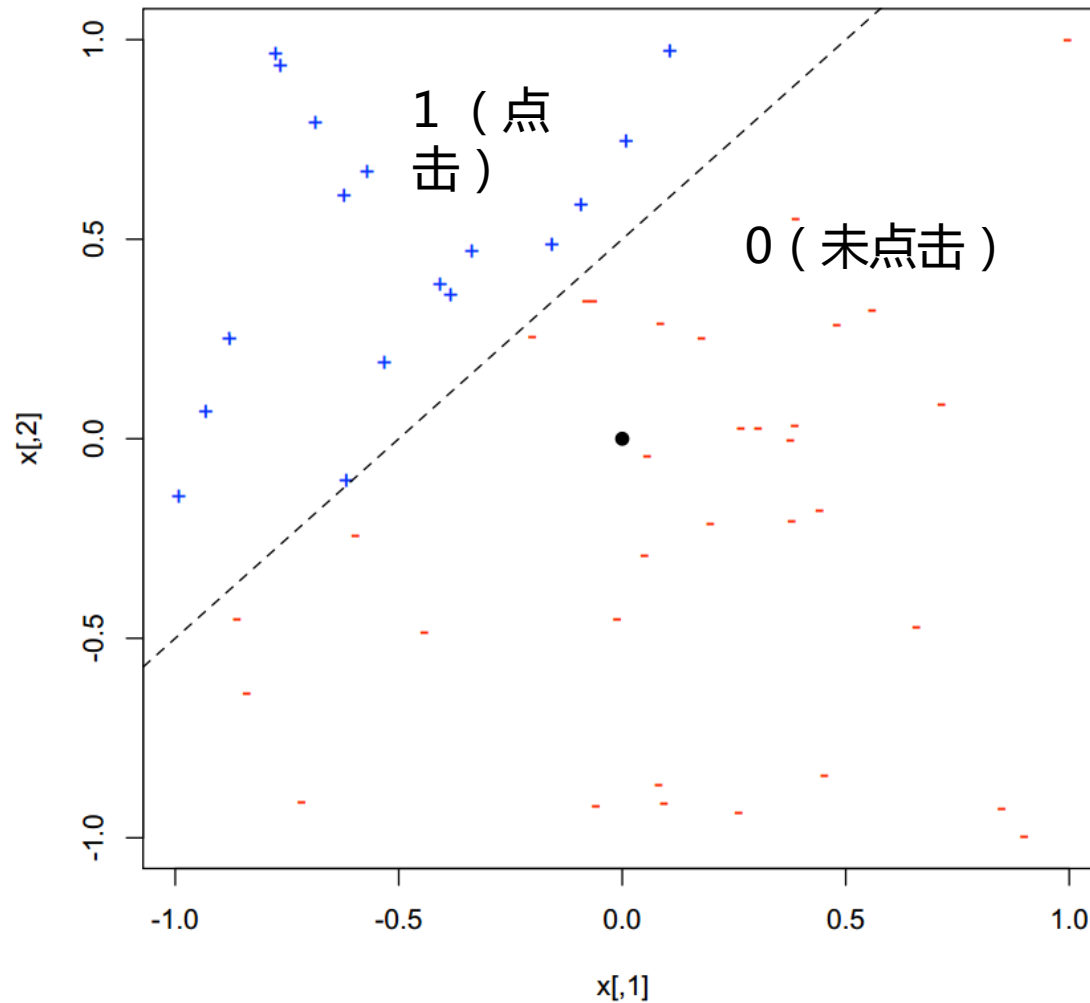


Model 融合问题

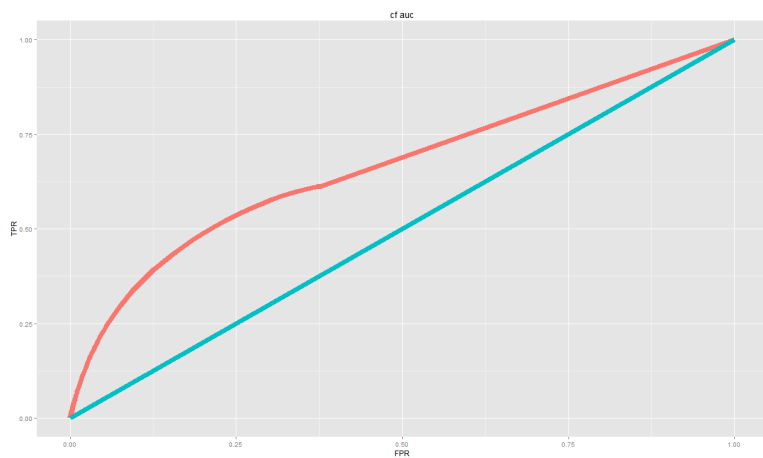


主商品	候选集	v2b	b2b	v2v(cf1)	v2v(cf2)	target
Item	Item ₁	0.021	0.51	0.47	...		0.53
	Item ₂			0.49			0.48
	Item ₃			0.33			0.45

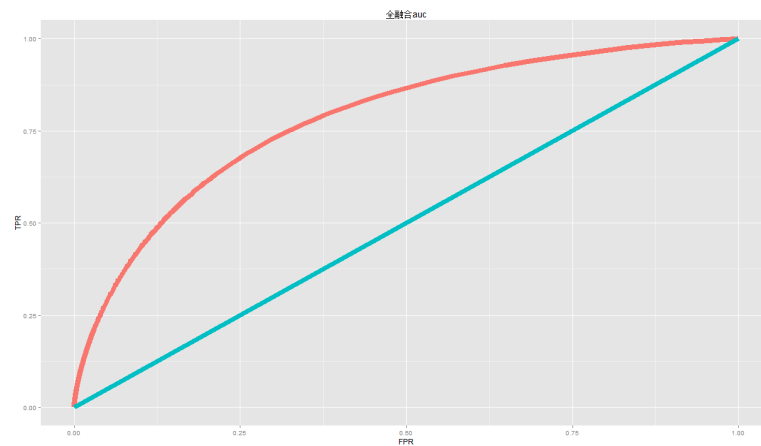
	Item _n			0.58			0.21



Logistic Regression @MPI

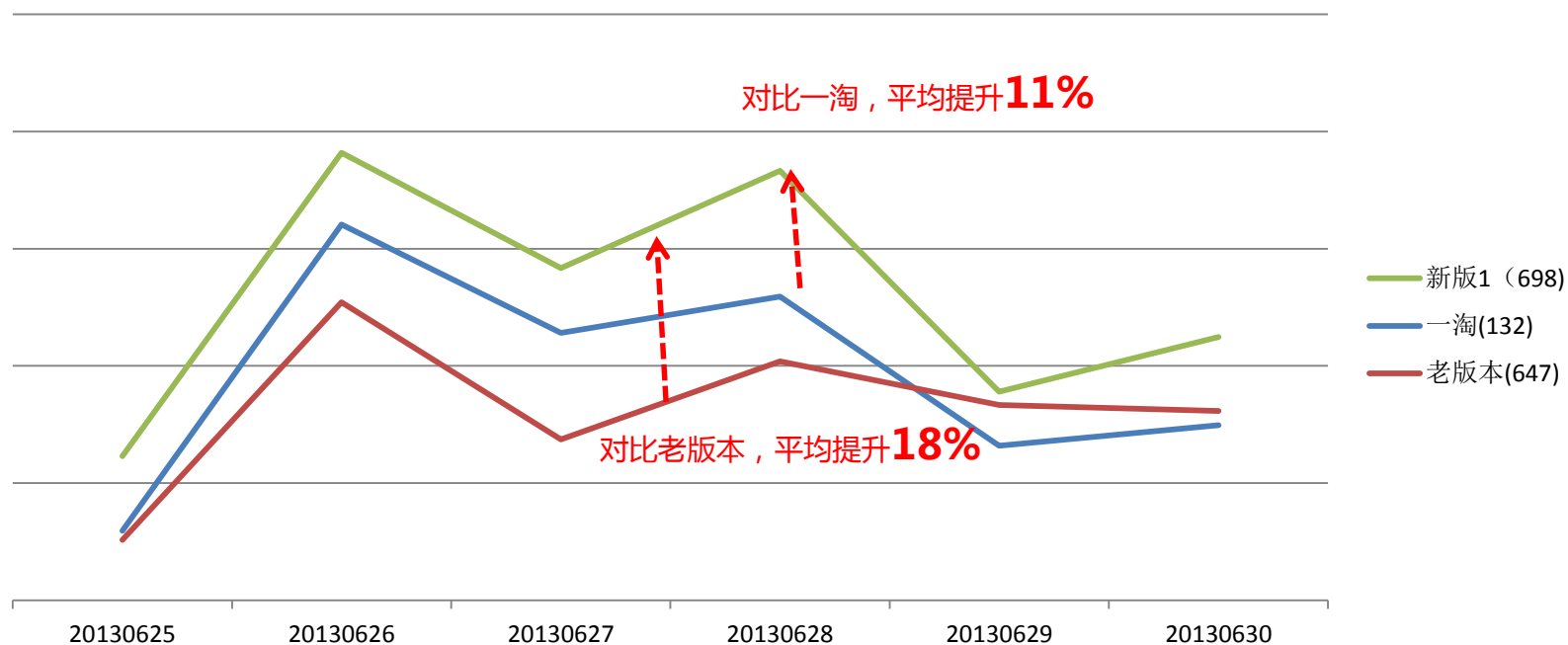


单一CF模型 AUC= 0.6653



融合后模型 AUC= 0.7835

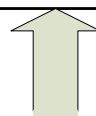
Item2item 算法优化效果 (PV 曝光价值)



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

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	
集群	MPI 集群		Hadoop 集群	

猜你喜欢

?

?

?

?



Personalization-场景引擎-模型

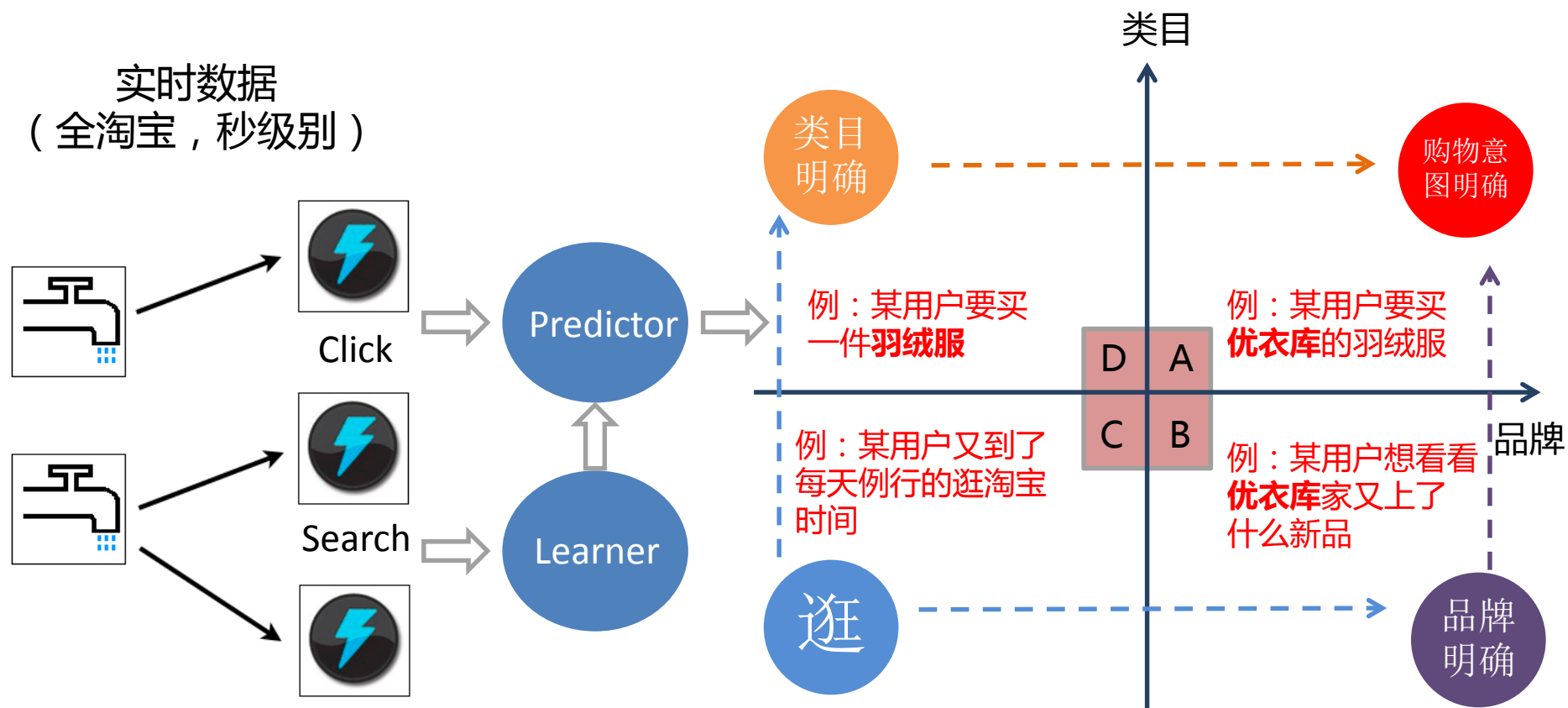
SAGC2013

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

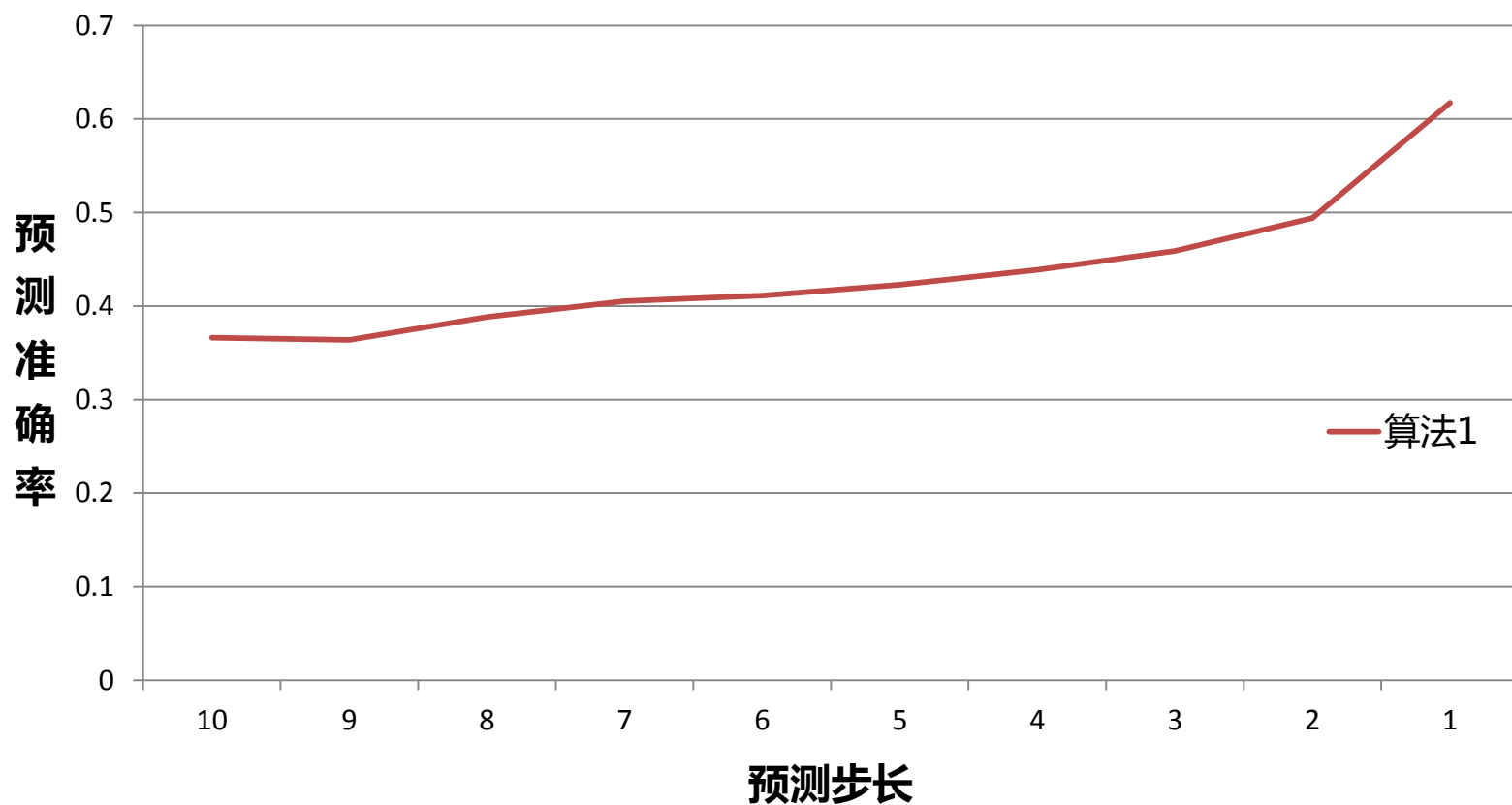
意图建模（三维模型）：品牌意图、类目意图

特征刻画：利用商品属性、商品标签等数据刻画类目意图

技术框架：



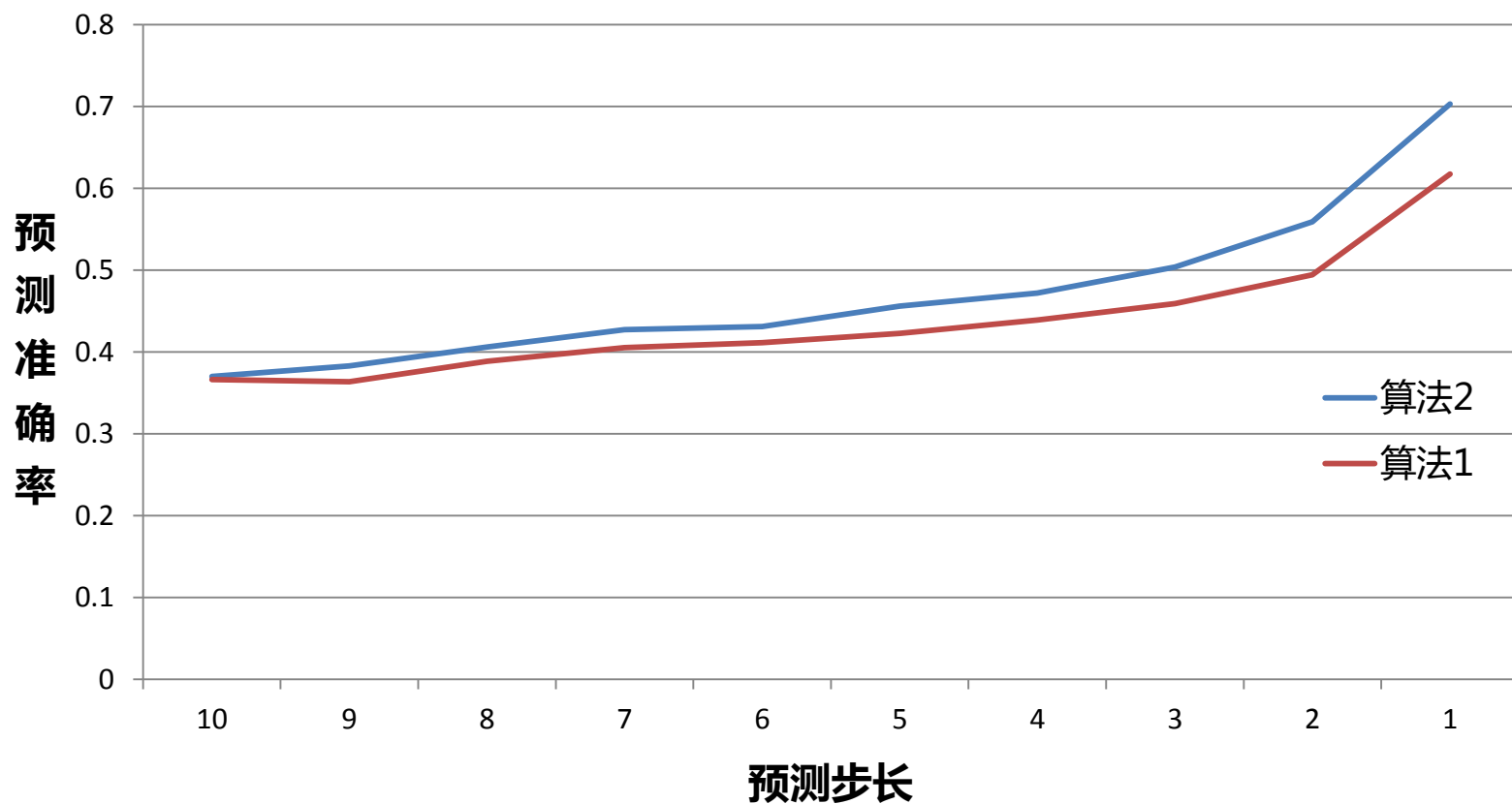
场景引擎效果离线评估方法



user session:



场景引擎效果离线评估方法



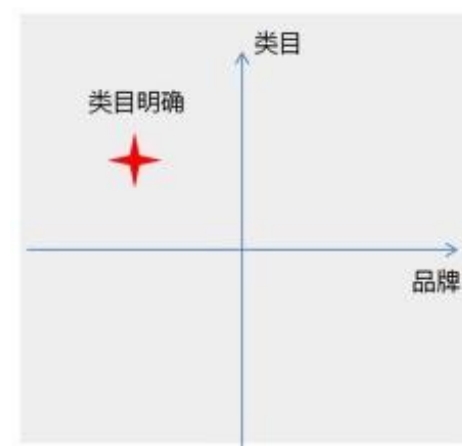
user session:



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

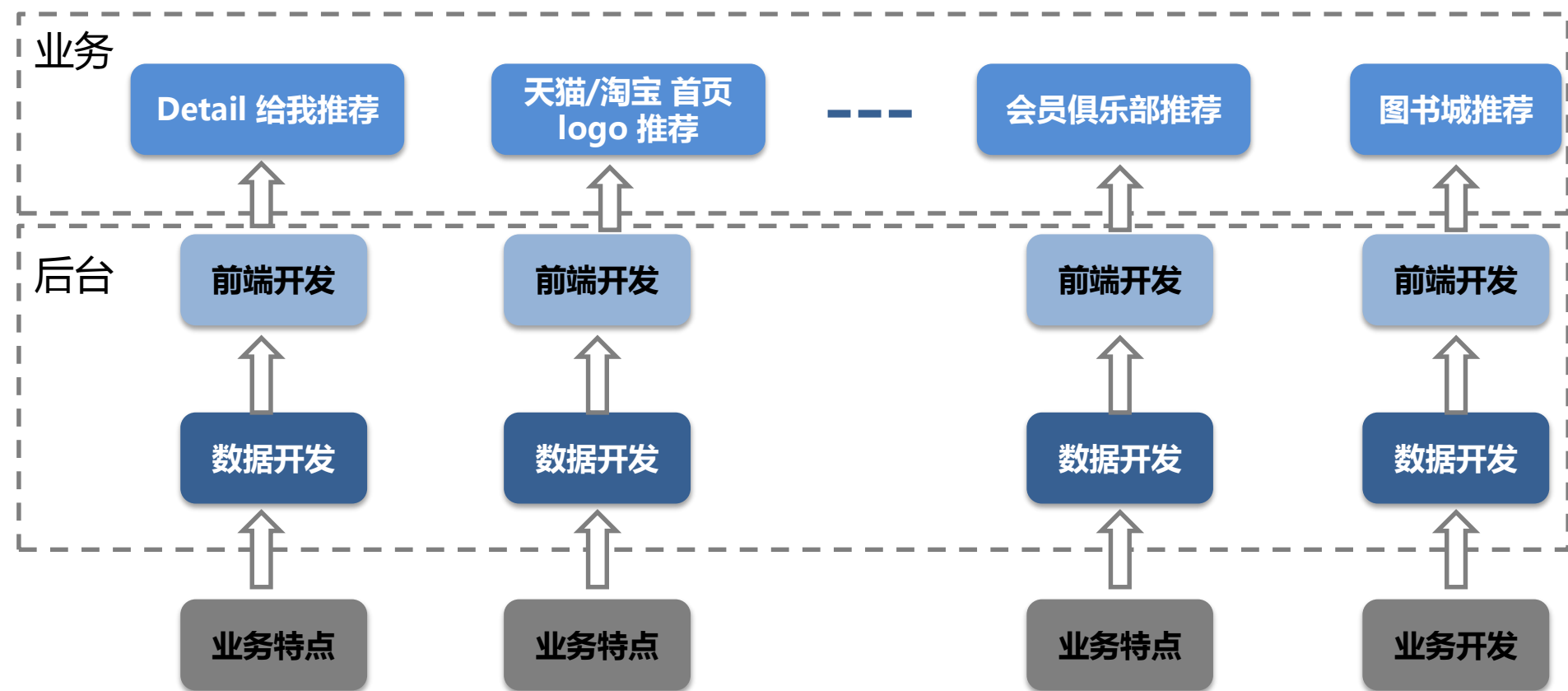


场景	用户想买的类目(笔记本电脑), 但不确定具体的品牌
品牌	笔记本电脑_品牌_Dell/戴尔 笔记本电脑_品牌_Apple/苹果 跑步鞋_品牌_New Balance/新百伦 跑步鞋_品牌_ANTA/安踏 跑步鞋_品牌_Asics/亚瑟士 连衣裙_品牌_Zimmur
导购节点	TmallList<-- TmallDetail<-- TmallDetail<-- TmallDetail<-- TmallDetail<-- TmallDetail<--
笔记本电脑	weight:100.00
"笔记本电脑"的标签	苹果(0.30) 超薄(0.20) 联想(0.20) 大陆行货(0.19) dell(0.13)
"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)

算法组装

天猫推荐系统架构@2013之前

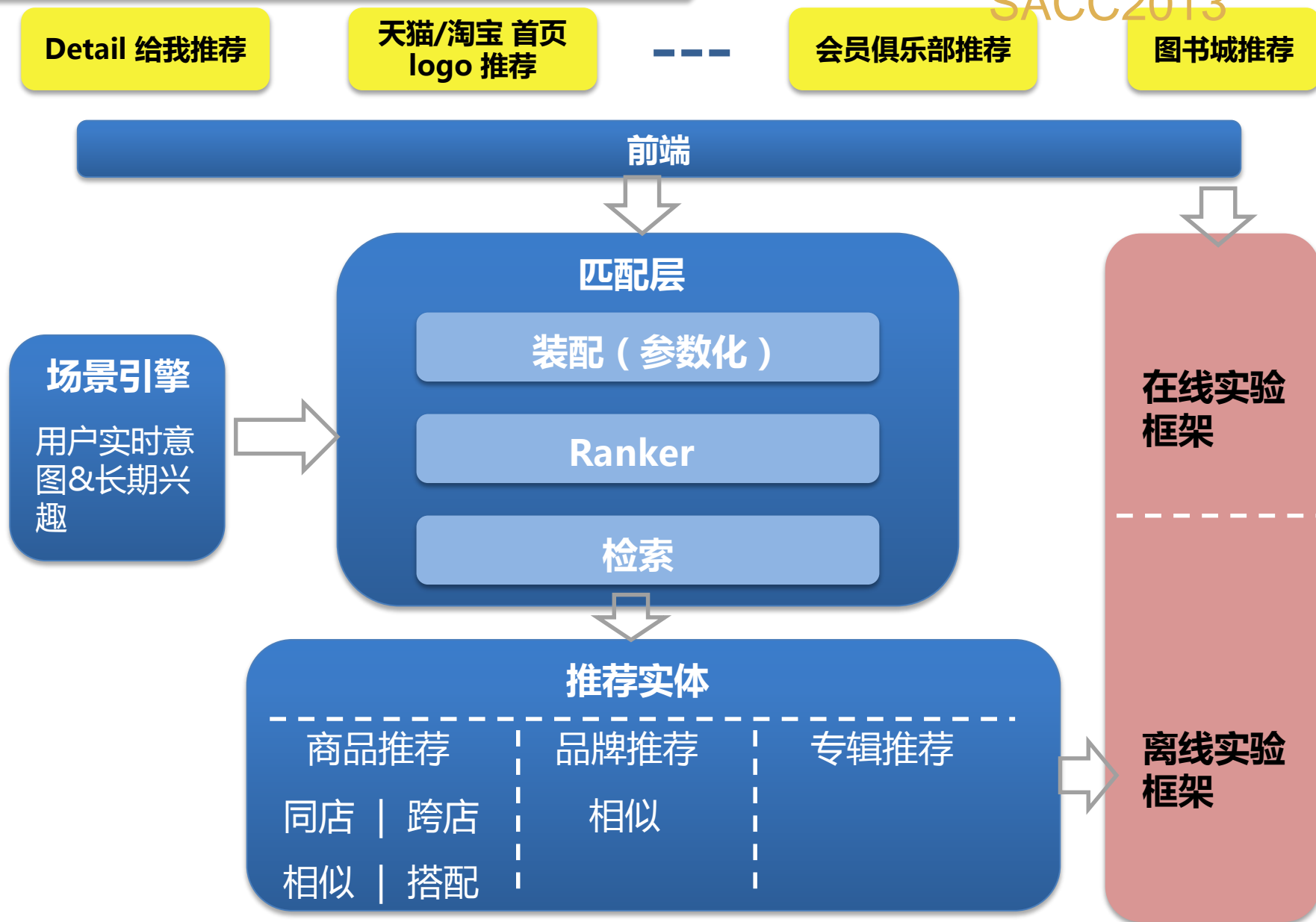
SACC2013

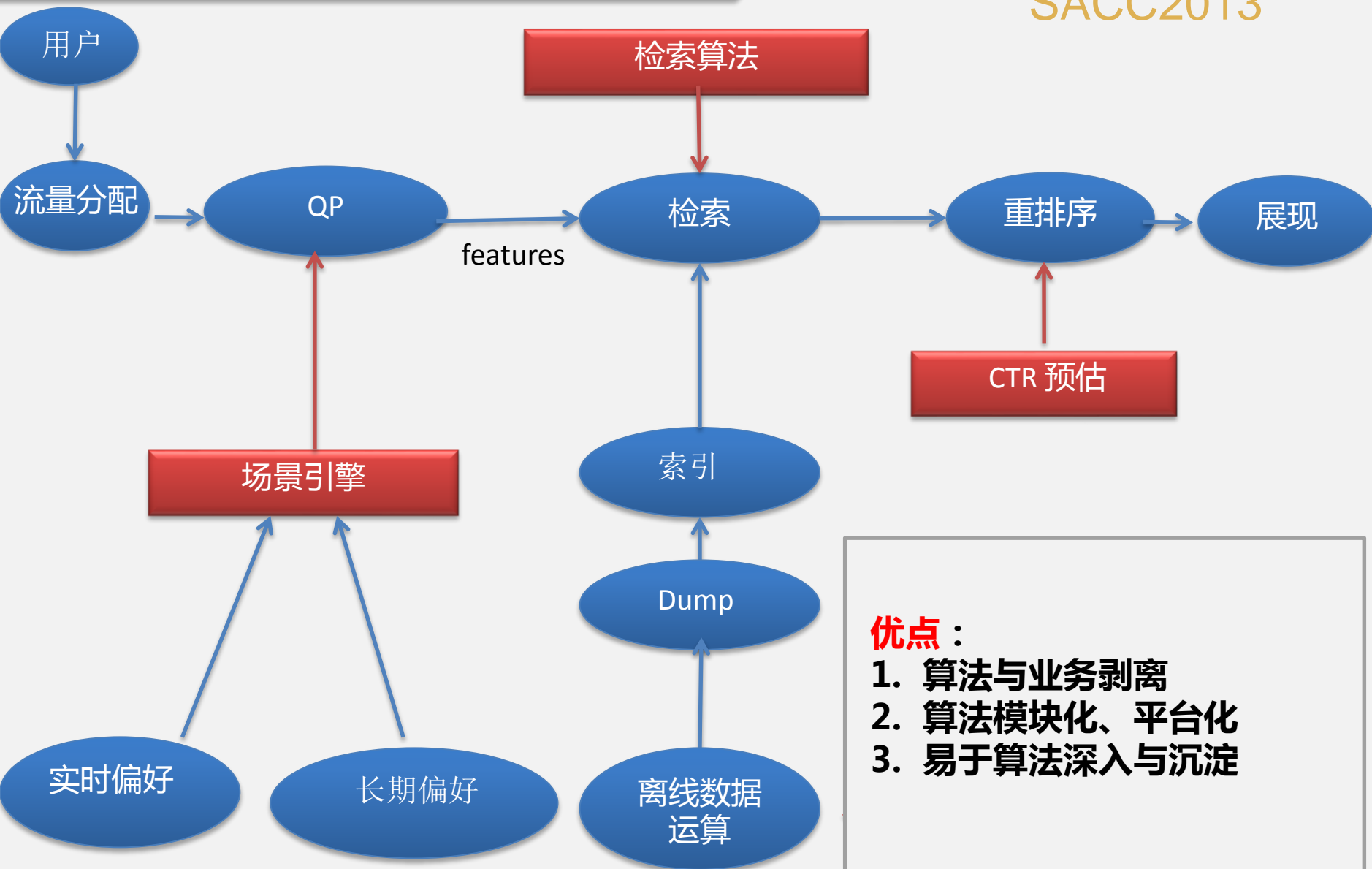


问题：业务和算法高度结合
高投入、低沉淀

天猫推荐系统架构@2013

SACC2013





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

$$\textit{prob}(\textit{click} \mid \textit{user}, \textit{item})$$

数据日志

用户数据								商品数据				
实时特征			长期特征			Demo						
Tag	类目偏好	...	品牌偏好	类目偏好	...	性别	...	类目	品牌	销量	标题	...

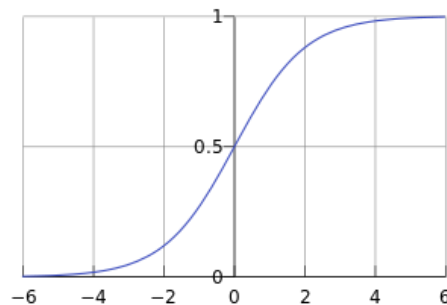


特征&训练数据

训练序列	类目Match	品牌Match	购买力Match	标签Match	类目	价格	销量	...
1	0.011	0.021	0	0.6	...			
2			...					



模型



Logistic_regression@MPI

Online CTR 预估-系统

SACC2013

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 (CTR 预估)

检索

1. 待排序商品列表
2. 用户实时特征

CTR 组件

效果提升13%

CTR Cache

推荐实体

商品推荐

品牌推荐

专辑推荐

同店 | 跨店

相似

相似 | 搭配

1. LR model
2. 商品特征
3. 用户历史特征

场景引擎

用户实时意
图&长期兴
趣

推荐应用@Tmall



斯凯奇skechers天猫正品gowalk情侣鞋2013新款男女鞋户外鞋13569
超轻 舒适 缓震 全店满480元包邮

价格 ¥ 599.00 (专柜价599.00 元, 什么是专柜价?)

促销 品牌总动员金秋疯狂购 ¥ **419.00**

超过30+ 推荐产品
每天服务超过6M用户

颜色分类



数量

库存934件

加入购物车

天猫积分

送209

支付

信用卡分期

快捷支付

服务承诺

商品: 按时发货

退款: T3极速退款

退货: T1退货保障卡

看了又看



¥499.02

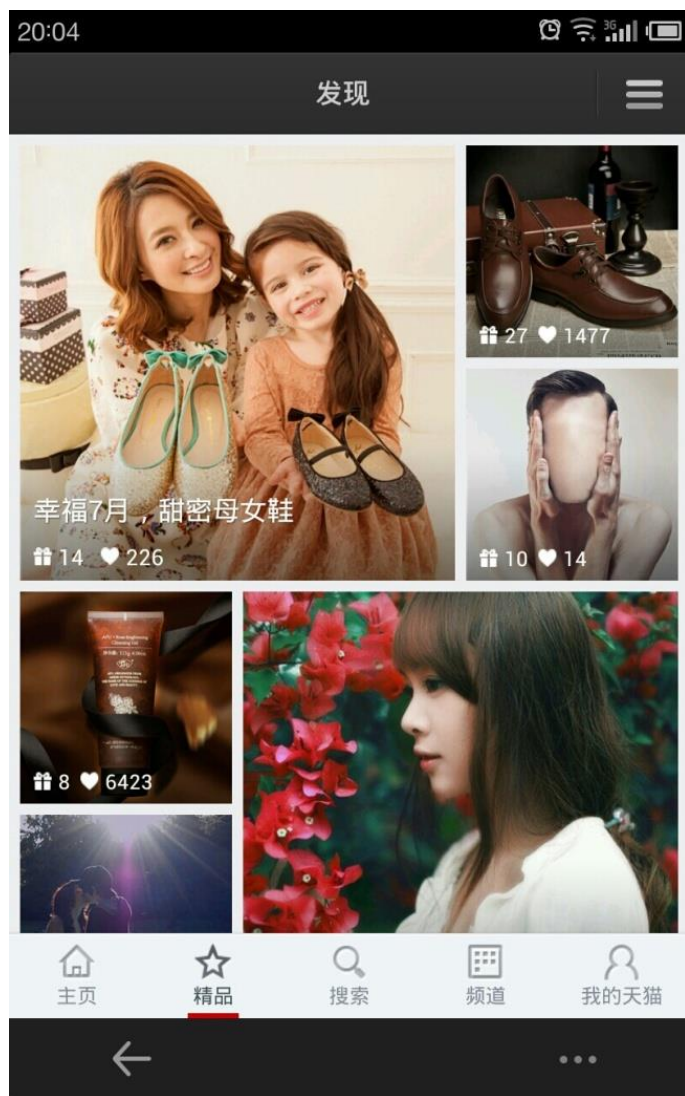


¥549.07

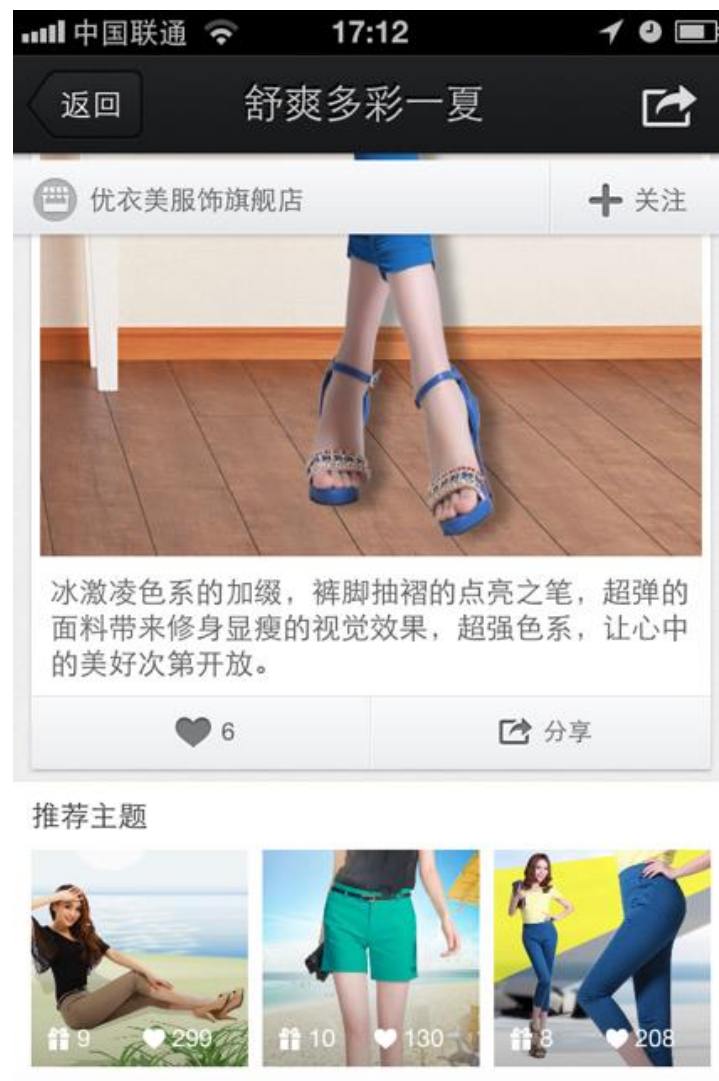


天猫推荐业务一览图（无线端）

SACC2013



精品频道个性化首页



相似主题推荐

让消费者更容易找到满意的商品

需解决问题：
所有人在会场看到相同的品牌列表



1000大牌满就减 心动价全场包邮

天猫新风尚

NEW 2013秋冬新品发布!

8月28日仅此一天!
领优惠券 立即抢购

Rule 活动规则 >

时尚无关价格，
而是您个人风格的表达。

我想逛流行街 GO >

Share 分享给朋友们 >
分享赢100积分
T3达人专享!

Teenie Weenie
2013新品全网首发

ochirly
欧时力
2013秋季新品上市

个性化版 VS 默认版

UV点击率提升：**6.32%**

UV转化率提升：**41.55%**

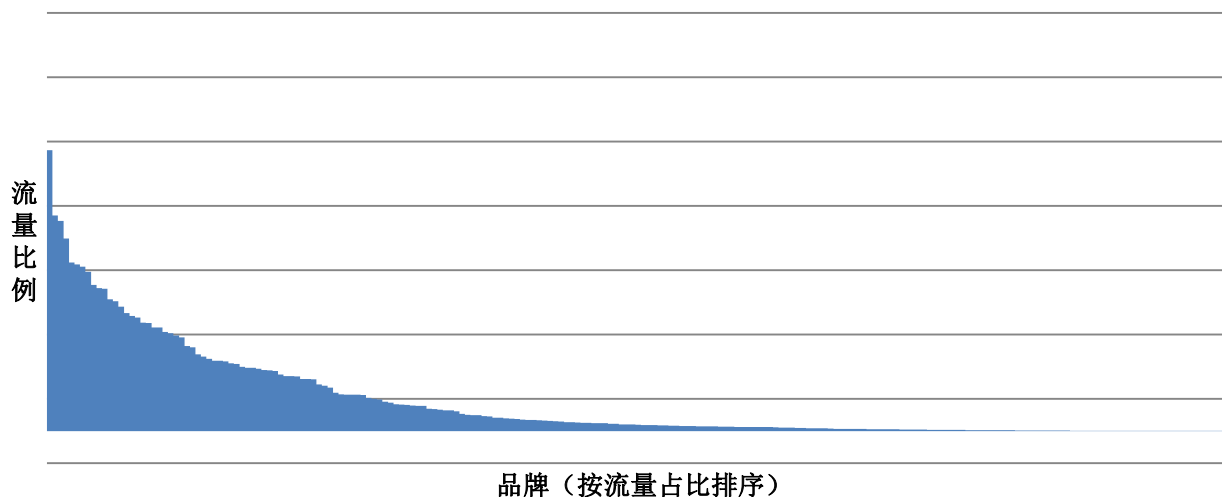
推荐对卖家的意义

SACC2013
让流量的分配
更公平（卖家）

默认版引导的品牌流量分布



个性化算法引导的品牌流量分布



15%

SACC2013

Thanks!

SequeMedia
盛拓传媒

IT168.com
www.it168.com

ChinaUnix.net

ITPUB