

Geekbang>

极客邦科技

全球领先的技术人学习和交流平台

扫我，码上开启新世界



Geekbang>

InfoQ | EGO NETWORKS | StuQ

InfoQ

专注中高端技术
人员的社区媒体

EGO NETWORKS

高端技术人员
学习型社交网络

StuQ

实践驱动的IT职业
学习和服务平台

促进软件开发领域知识与创新的传播



[上海] 2015年10月15-17日



[北京] 2015年12月18日-19日



关注InfoQ官方微信
及时获取ArchSummit演讲视频信息

ArchSummit全球架构师峰会 深圳站2015

京东数据驱动下的个性化推荐系统
刘尚堃

目录 CONTENT

1. 产品展示

2. 个性化推荐技术概览
3. 召回模型介绍
4. 排序模型介绍
5. 架构/ab测试
6. 未来优化方向

1 产品展示

- I. 提高网站销售能力，考核指标：转化率、GMV
- II. 定位和发现用户兴趣，让用户可以逛起来，考核指标：产品DAU、停留时间、访问深度、新老用户留存等



猜你喜欢 根据你的喜好精心为你推荐!

换一批



1.2 产品展示



移动端个性化推荐订单贡献已经达到13%。

目录 CONTENT

1. 产品展示
- 2. 个性化推荐技术概览**
3. 数据平台介绍
4. 召回模型介绍
5. 排序模型介绍
6. 架构/ab测试
7. 未来优化方向



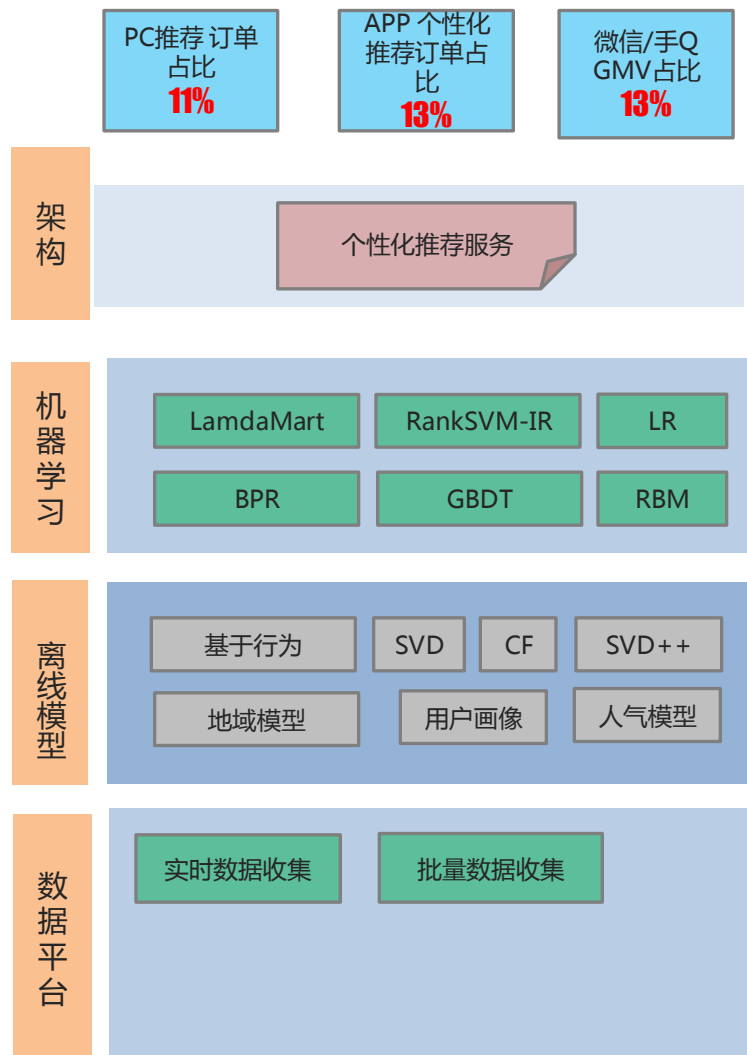
2、个性化推荐技术概览

✓跨体系、跨终端核心业务支撑

集团多项核心业务和战略项目均以推荐搜索为支撑，包括主页千人千面、移动端、微信、手Q，以及O2O；

✓个性化推荐

个性化推荐通过全方位的数据精准刻画用户的购买意图，有针对性给用户推荐用户购买意愿高的商品，为用户提供极致的产品体验，提升下单转化率，增强用户粘性。

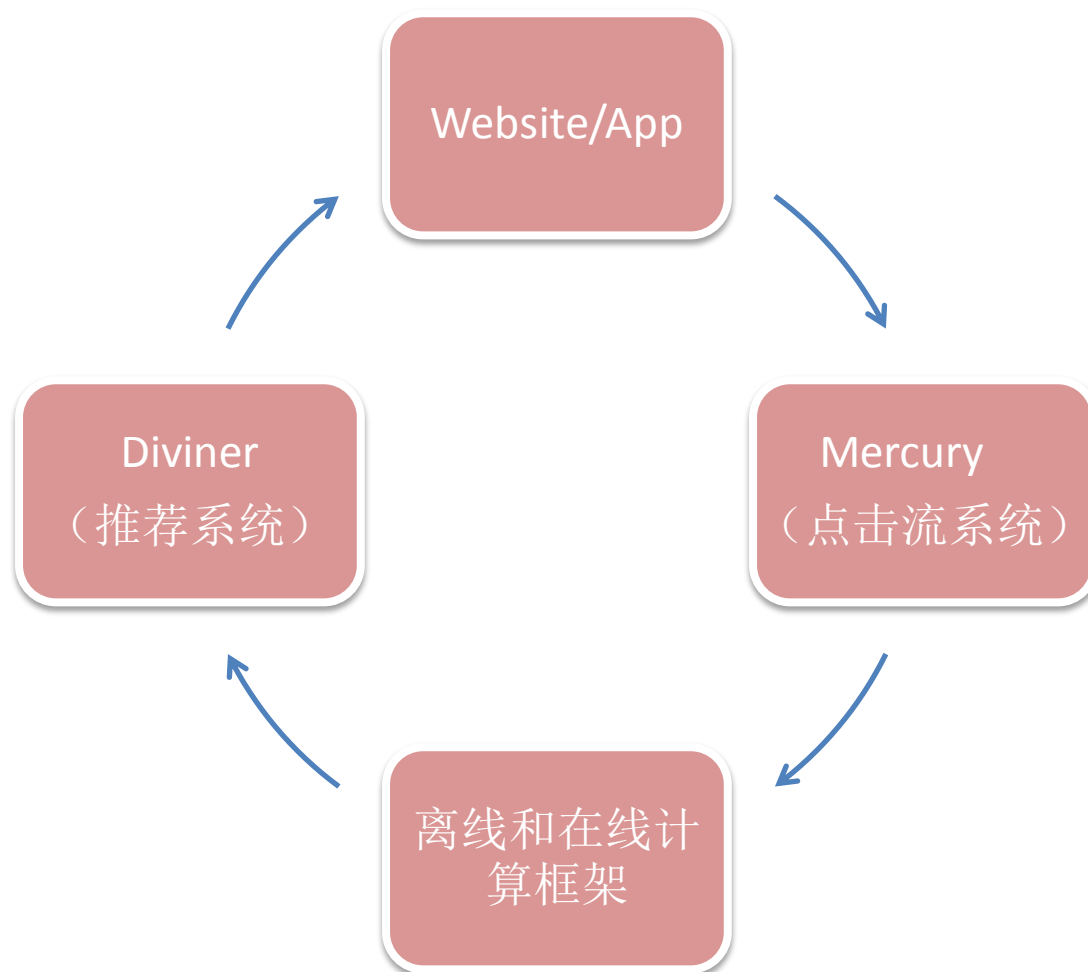


目录 CONTENT

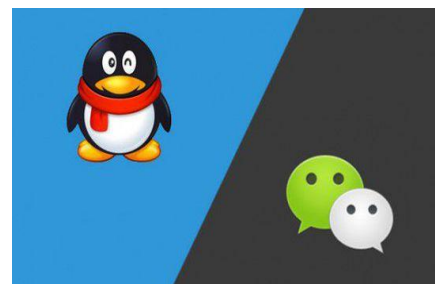
1. 产品展示
2. 个性化推荐技术概览
- 3. 数据平台介绍**
4. 召回模型介绍
5. 排序模型介绍
6. 架构/ab测试
7. 未来优化方向



3.1、数据平台——数据流程



3.2、数据平台——用户数据打通

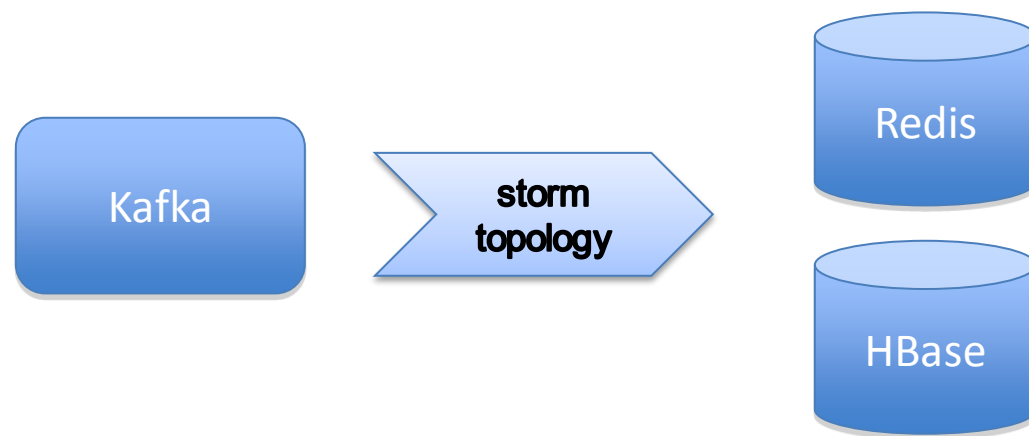


- 1、多屏互通
- 2、兴趣传递

3.3、数据平台——离线计算和在线计算框架



- I. 离线计算主要在Hadoop上运行Map Reduce
- II. 部分计算都使用Mahout , Spark
- III. 通过分布式任务调度系统对算法进行调度和监控



在线计算主要基于Storm , 实时消息基于kafka(150亿+)

目录 CONTENT

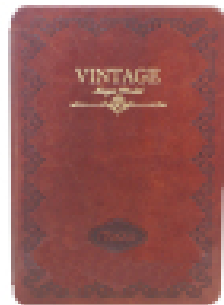
1. 产品展示
2. 个性化推荐技术概览
3. 数据平台介绍
- 4. 召回模型介绍**
5. 排序模型介绍
6. 架构/ab测试
7. 未来优化方向

4.1 召回模型（获取候选）——基于行为



股票投资入门与实战技巧
从零开始学炒股

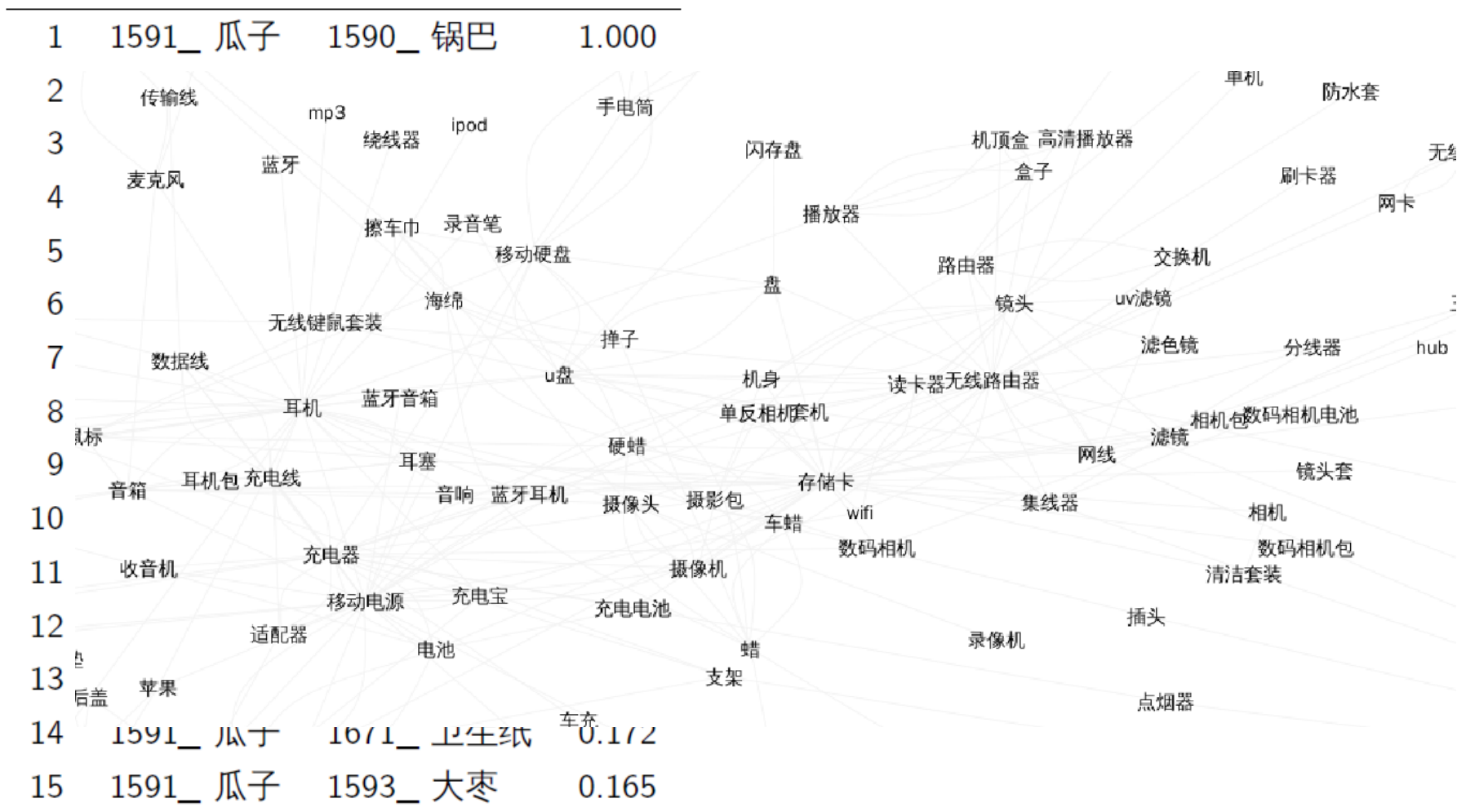
用户浏览相似



用户购买的搭配

4.1.1 召回模型——计算搭配商品的trick (1)

京东计算搭配商品除传统购物篮分析作用到单品外，还利用产品词到产品词的关系构建出网络。



4.1.1、召回模型——计算搭配商品的trick (2)

通过产品词之间的关系提升长尾商品的搭配覆盖

| | | | | | | | | |
|---|---|---|---|--|---|---|---|---|
|  | + |  | + |  | + |  | + |  |
| 斩浪海竿2.4 3.0 3.6米碳素海竿钓鱼 | | 平行线 D007 4.3寸 超大屏幕后视镜行 | | 芬美时恋 繁星蕾丝 全罩杯文胸薄杯 大 | | 【79秒杀】想川夏 季新款休闲鞋低帮 | | 平行线 P268 行车 记录仪双镜头HD高 |
| | | ¥758.00 | | ¥89.00 | | ¥79.00 | | ¥909.00 |

| | | | |
|--|--|---|--|
|  |  |  |  |
| 让保罗正品 支架鱼竿炮台筒 易支架钓鱼支架 渔具垂钓用 品特价鱼具 竿架2.1m | 猛攻伊势尼鱼钩鱼线鱼漂套 装组线鱼钩有倒刺散装进口油 钩渔具垂钓用品钓鱼钩特价 | 熊火1.7米不锈钢鱼竿炮台支 架杆架竿架钓竿支架钓椅支架 钓箱钓台炮台支架618好货钜 | 鱼钩 伊势尼盒装鱼钩 带倒刺 70枚 垂钓用品 渔具 鱼竿配 件 钓鱼配件 台钓竿配件 |
| 京东价: 29.00 市场价: 45.00 权重: 0.0076000000 浏览历史: | 京东价: 12.00 市场价: 45.00 权重: 0.0051000000 浏览历史: | 京东价: 35.00 市场价: 90.00 权重: 0.0049000000 浏览历史: | 京东价: 10.00 市场价: 12.00 权重: 0.0041000000 浏览历史: |

4.2 召回模型 —— 基于用户偏好

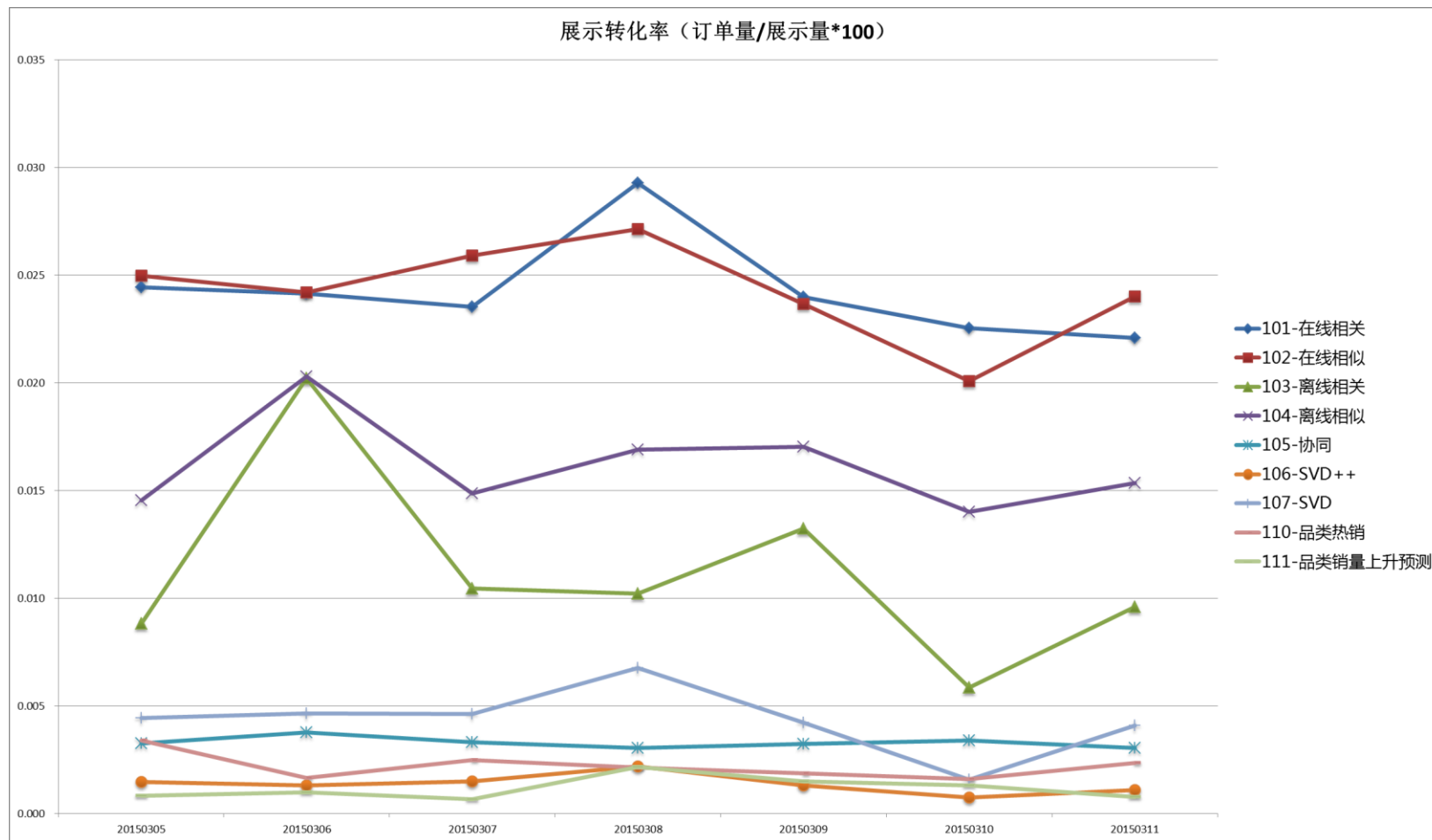


4.3、召回模型（获取候选）——基于地域

| | | | | | |
|---|--|---|--|--|--|
|  |  |  |  |  |  |
| 维康塑料扑克 | 摩斯维 手机套保护壳 适用于小米NOTE红米2 深宝蓝 送2张贴膜/2件减3元 | 摩斯维 高清贴膜（拍下会随手机壳型号配送相对贴膜）自动匹配型号 | 瑜伽服胸垫 女瑜伽服女环保海绵透气抹胸胸垫 加厚运动瑜伽服胸垫 插片胸垫 舞蹈服胸垫 均码 | 瑜伽袜 女瑜伽服女防滑袜 透气吸汗 环保胶粒 瑜伽袜 瑜伽服配套袜子 特价袜子 均码 | 罗美尔 西装衬衫羽绒大衣用 宽肩实木衣架 4只装 SM22/LF22 男款本色 |
|  |  |  |  |  |  |
| 台湾美保芒果汁490ml*6罐 | 魅动e族小巧巧克力蓝牙音箱立体声蓝牙音响 带NFC自动配对4.0音质 低音炮 巧克力 | 摩斯维 防尘塞 适用于小米/魅族/华为/HTC/索尼/OPPO/VIVO/三星/中兴 | 倍斯特 BST-0137松木 移动电源/充电宝 10400毫安 白色 | 倍斯特 移动电源保护袋 | 爱壳饰 适用于苹果ipad4/ipad3/ipad2保护套 带休眠皮套韩国超薄可爱外壳 iPad4/3/2通用纯色粉 |
|  |  |  |  |  |  |
| 6分子 Baby strong饮用山泉水 1L*6整箱装 | 三光云彩GLASSLOCK1500ML保鲜罐IP591 | 三光云彩glasslock长方形1100ml玻璃保鲜盒MCR110/RP518 | 东芝（TOSHIBA）隼系列（THUHYBS-016G）U盘 16G | 达尔优（dare-u）执法官 多媒体 游戏鼠标 | 漫步者（EDIFIER）H210 低频强劲 时尚入耳式耳塞 酷黑 |

4.4 召回模型——召回模型效率分析

模型展示转化率：订单量/展示量



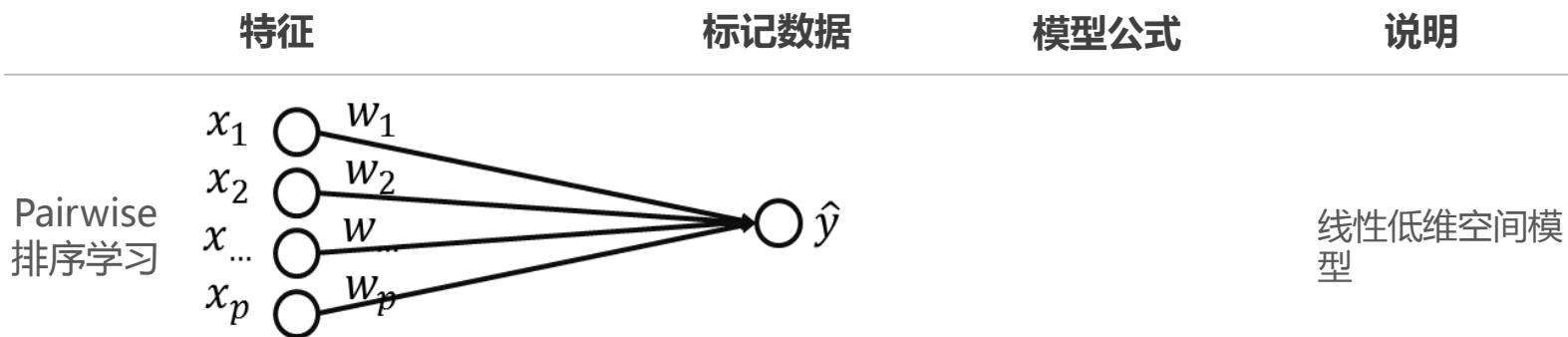
目录 CONTENT

1. 产品展示
2. 个性化推荐技术概览
3. 数据平台介绍
4. 召回模型介绍
- 5. 排序模型介绍**
6. 架构/ab测试
7. 未来优化方向

5.1、排序模型——排序学习

从用户的交互日志中通过模型训练特征权重；

离线监督学习



场景
恢复

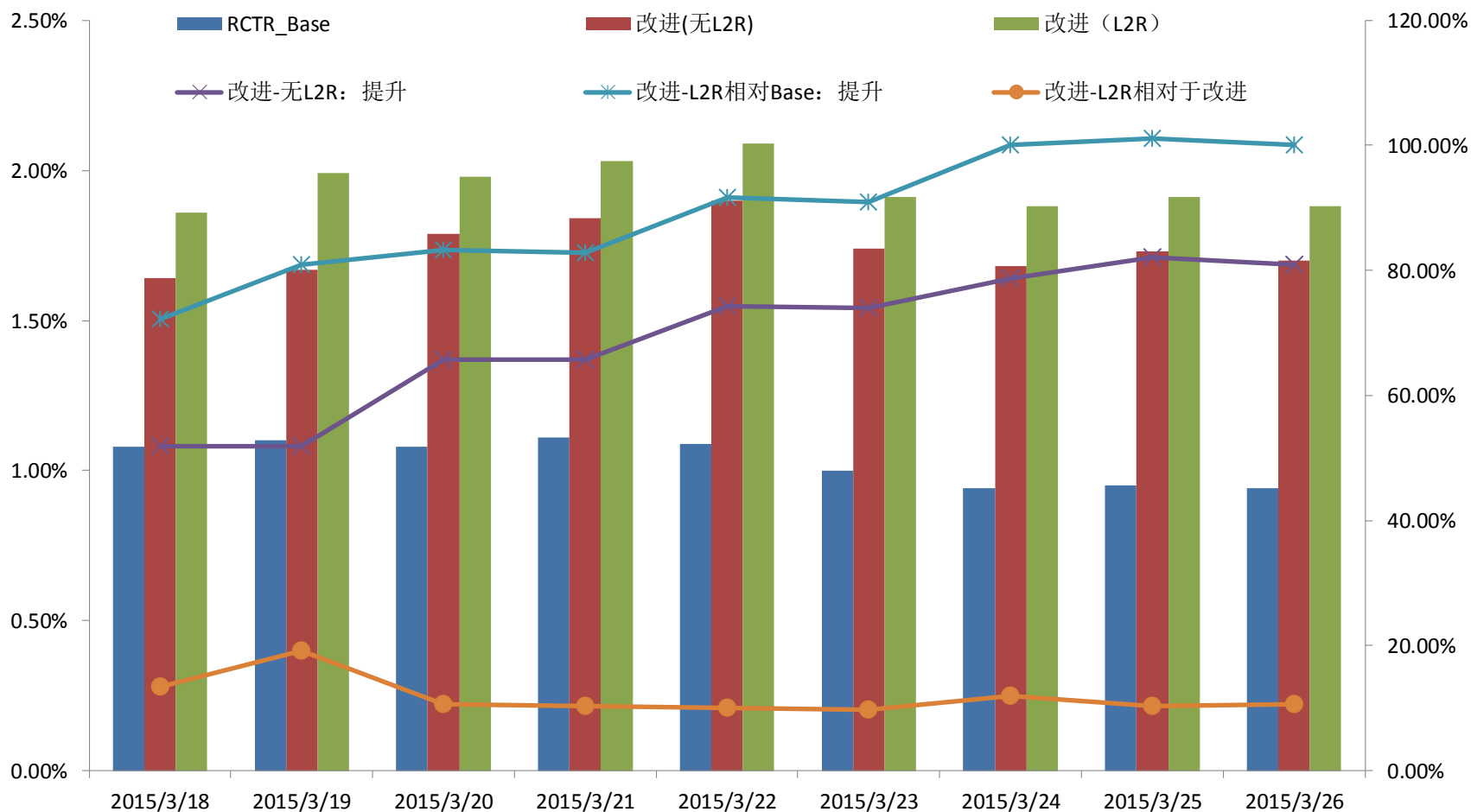
样本
标记

特征
计算

模型
训练

线上
实验

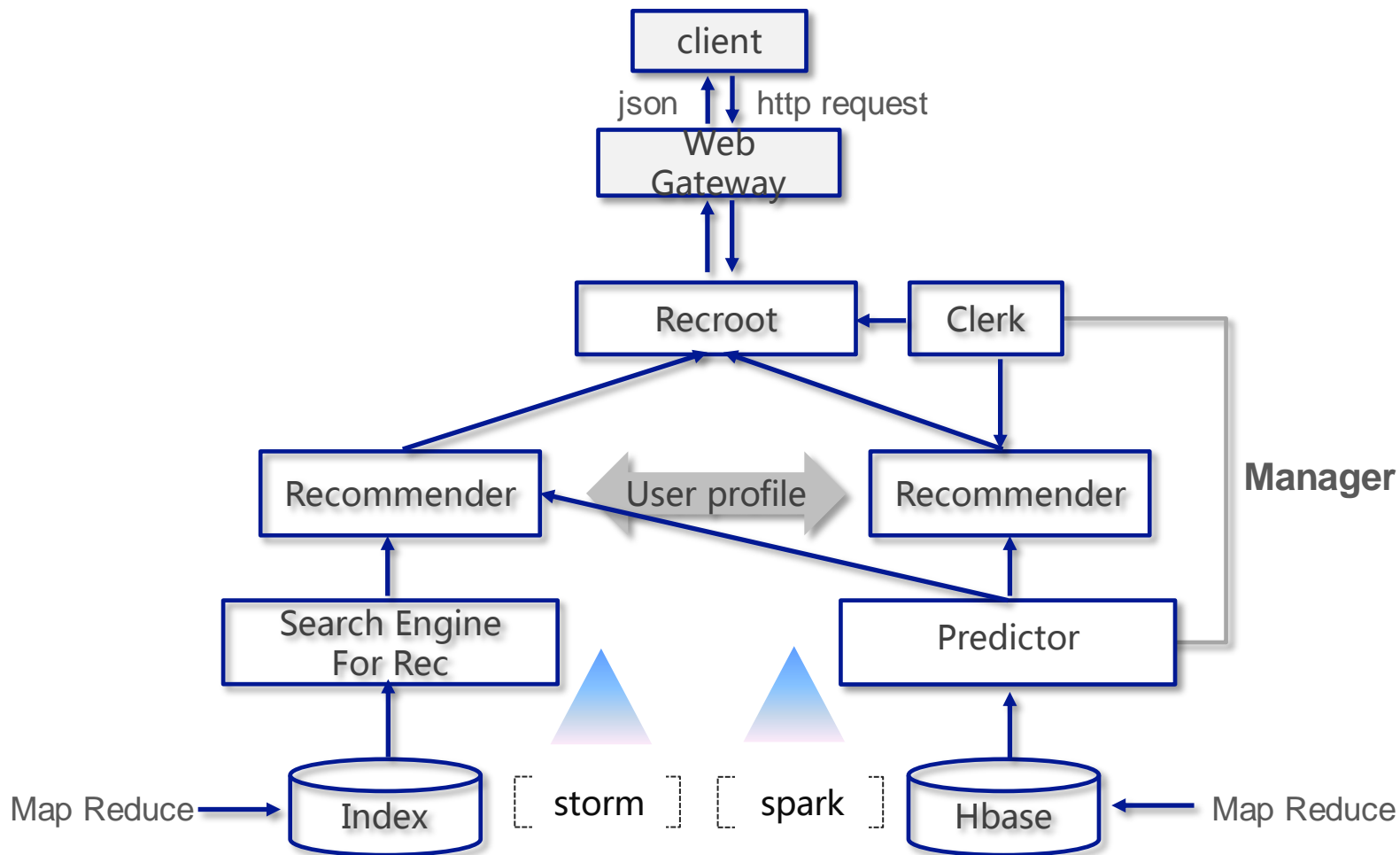
5.2、排序模型——排序学习效果改进



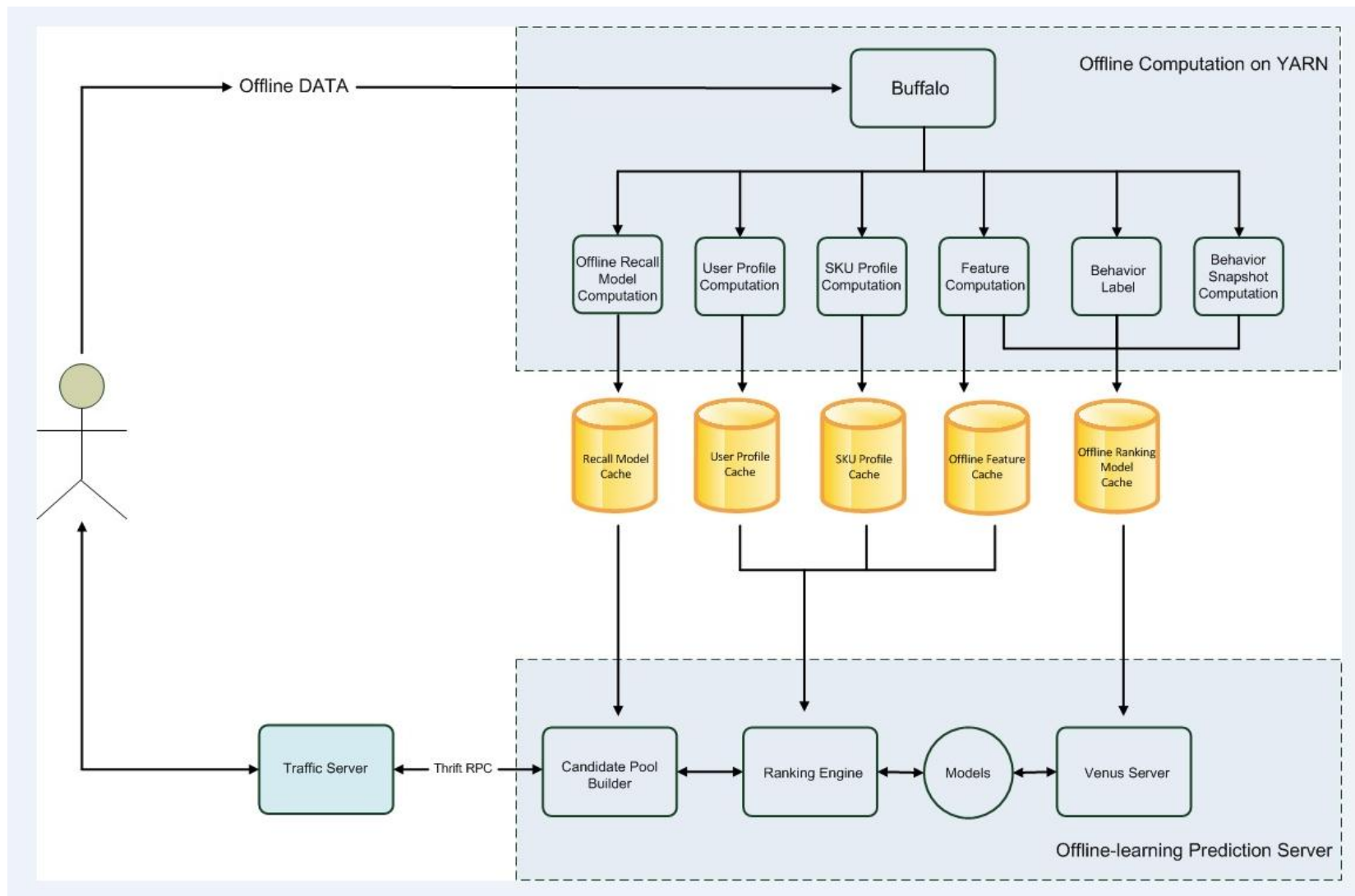
目录 CONTENT

1. 产品展示
2. 个性化推荐技术概览
3. 数据平台介绍
4. 召回模型介绍
5. 排序模型介绍
- 6. 架构/ab测试**
7. 未来优化方向

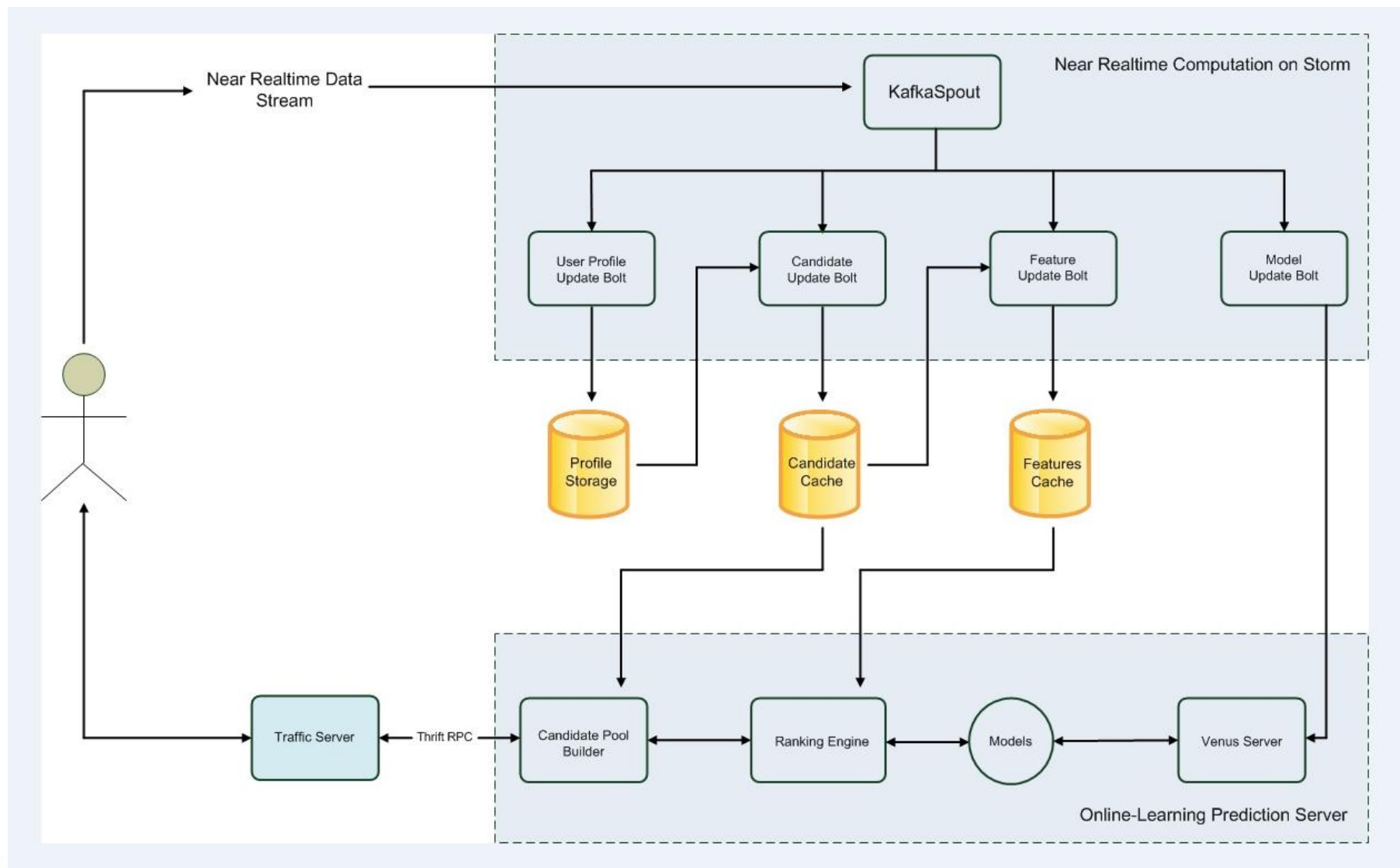
6.1 架构



6.1 架构 —— offline 架构



6.1 架构 —— oneline 架构



6.2 架构——ab测试

图形界面配置线上分流实验，配置即时生效，避免反复上线
增加实验，选择分流策略（随机、用户、品类）
调整流量，设置生效时间
配置追溯和回滚

▼ Diviner Configuration

Mercury Configuration

About

Help

Logged in as Logout

Diviner Configuration Edit

| Last Update | Least Effort | CDN Cache Time | Predictor Cache Time | Active Time | Version |
|-------------|--------------|----------------|----------------------|---------------------|---------|
| 3c | false | 5 | 7200 | 2013-11-19 11:35:19 | 364 |

Placements

104000

Basic Configuration

| Id | Name | Create Time | Update Time | |
|--------|--------|--------------------|---------------------|---------------------|
| 102002 | 104000 | 关联推荐:图书看了还看(图书商品页) | 2013-06-28 17:30:27 | 2013-11-18 13:48:45 |

Service Configuration

| Default Recommender | Traffic Strategy |
|------------------------------|-------------------|
| /si/diviner/recommender/hub2 | Partition By User |

Service Parameter

| # | Key | Value | |
|--------|-----|--------------------|-------|
| 102003 | 1 | use_cold_start | false |
| 102001 | 2 | no_order_filtering | false |
| 303000 | 3 | recommender_name | book |

Experiments

1040000113102126

Owner:

diviner@jd.com

Create Time:

2013-10-21 14:19:46

Update Time:

2013-11-10 20:42:46

Service Configuration

| Recommender | Experiment Name |
|------------------------------|-----------------|
| /si/diviner/recommender/hub2 | sku_browse_buy |

Traffic Configuration

| Traffic Setting |
|-----------------|
| 10.0 |

Experiment Parameter

| # | Key | Value |
|---|------------------|----------------|
| 1 | predictor_type | sku_browse_buy |
| 2 | recommender_name | book |

6.3 架构——traffic server

- I. 架构支持算法高速迭代：平均每周7个新的算法实验上线
- II. 决策机制：优先考虑转化率、UV价值。浏览发现类产品需要体现体验类指标提升。

目录 CONTENT

1. 产品展示
2. 个性化推荐技术概览
3. 数据平台介绍
4. 召回模型介绍
5. 排序模型介绍
6. 架构/ab测试
- 7. 未来优化方向**

7、未来优化方向

- I. 用户画像——家庭画像
- II. 加强用户兴趣的发现

Thanks!

