

计算广告

360商业产品首席架构师 刘鹏
微博：@北冥乘海生

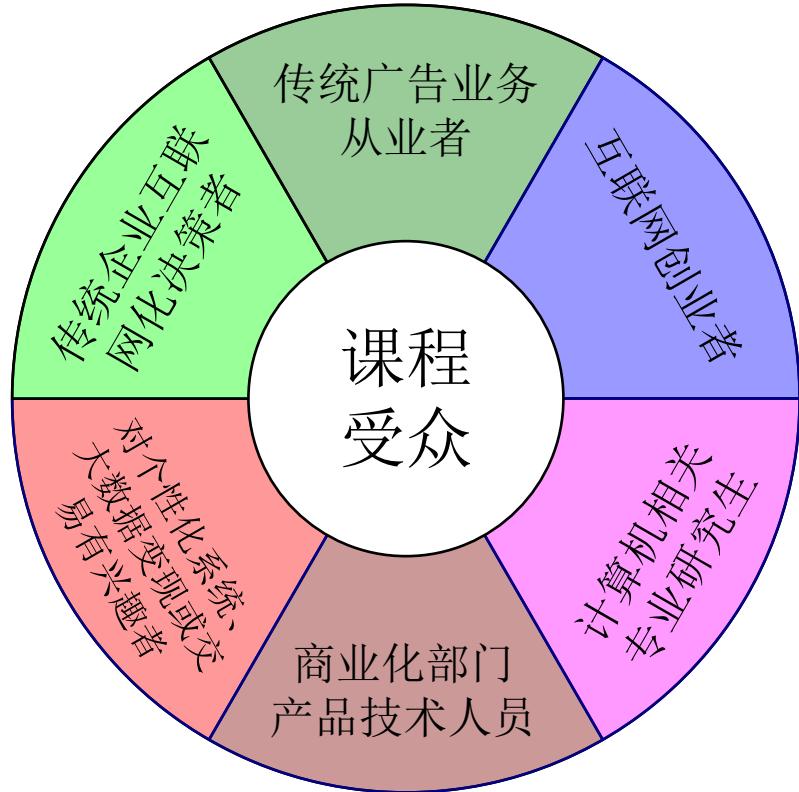
戏说互联网思维之“三个不要”

- “不要钱”
 - 免费倾销加后向变现的商业模式
 - 所有能够传播信息的商品，售价都会趋向其边际成本
- “不要脸”
 - 无底线迎合用户的产品与营销方式
- “不要命”
 - 用期权和价值观让程序猿在疯狂状态下全天候工作

与商业化相关的产品问题

- 商业模式探索，例如：电影是一种边际成本很低，同时信息传播量又很大的典型商品。是否可能探索一种售价很低，而充分利用其信息传播能力的电影行业发行模式？
- 流量变现，例如：互联网电视厂商除了销售收入，还可以获得用户流量。这些流量的性质如何，应如何变现？
- 数据变现，例如：室内导航技术是近年来快速发展的新型互联网应用。这类产品会得到什么样有价值的数据资产，又应该采用哪种具体的商业产品来变现？
- 商业产品建设和运营，例如：团购、游戏联运、返利购买这些推广模式与广告有什么内在联系？是否可以共用某些产品和技术平台？

哪些人需要了解商业化与计算广告？

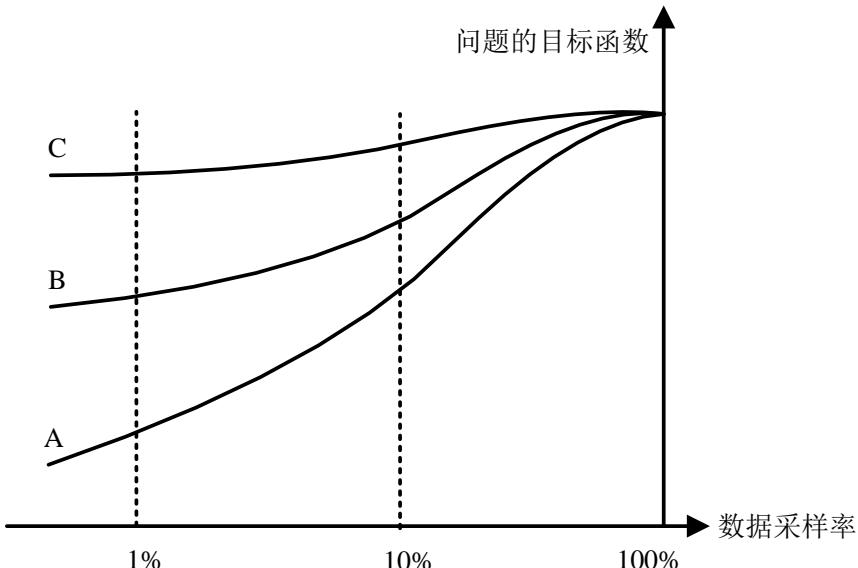


课程提纲

- 在线广告综述
- 计算广告基础
- 计算广告产品技术概览
- 合约广告
- 受众定向
- 搜索与竞价广告
- 程序化交易广告
- 移动互联与原生广告
- 在线广告产品技术实战
- 其他广告相关技术

在线广告综述

大数据与计算广告的关系



A: 典型大数据问题，解决问题的效果随着采样率降低显著降低，例如计算广告、个性化推荐等

C: 一般数据分析，非大数据问题，解决问题的效果在采样率降低时变化很小，例如各种洞察、单维度统计等

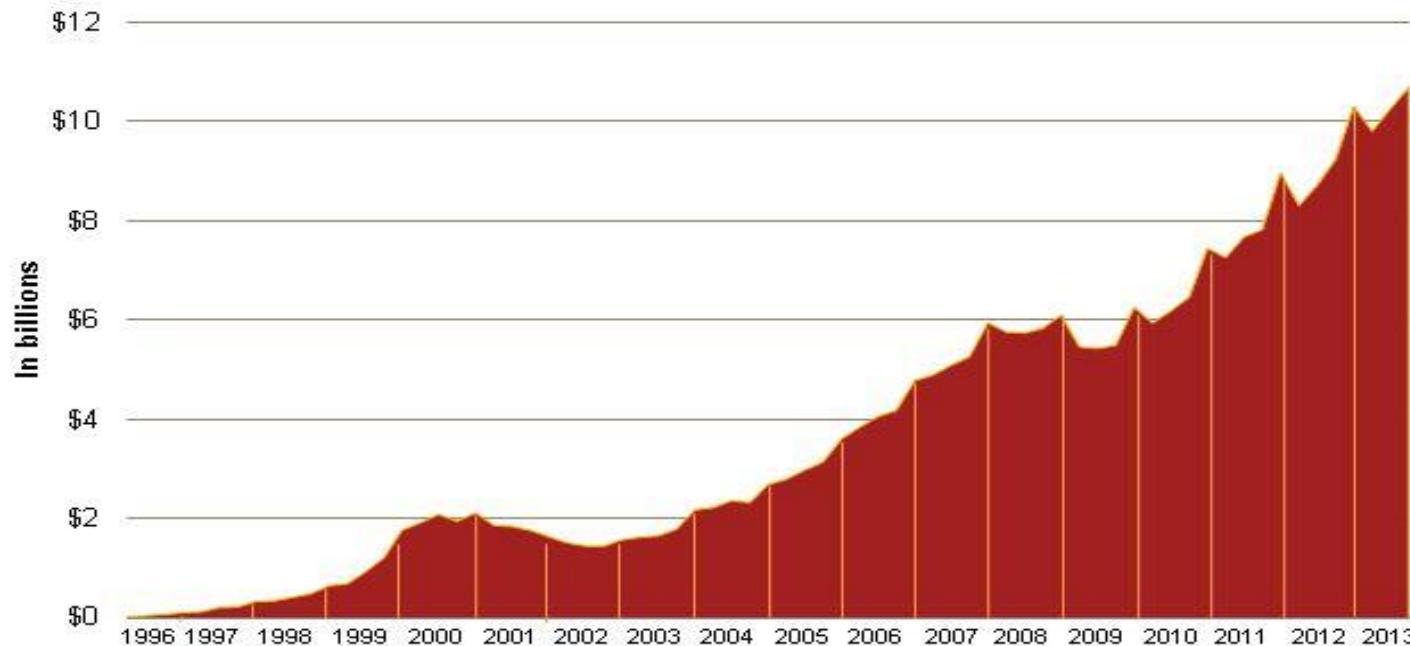
B: 过渡类型问题，解决问题的效果随采样率降低温和下降，例如文本主题分析等

关于在线广告

- 在线广告支撑了整个互联网行业的大半壁江山。不了解互联网广告，就不可能深入了解互联网。
- 在线广告是迄今为止，大数据领域唯一形成规模化营收的应用。
- 在线广告是结合了计算技术、心理学、经济学、营销学等的综合应用。

美国在线广告增长趋势

Quarterly revenue growth trends Q1 1996 – Q3 2013 (\$ billions)



Source: IAB/PwC 2013

中国在线广告增长趋势



中美主要广告市场变化趋势

		2007	2008	2009	2010	2011	2012	2013
中国	网络广告	17	27	33	52	83	122	179
	电视广告	97	114	127	153	182	207	212
美国	网络广告	212	234	226	260	317	366	428
	电视广告	719	394	359	401	685	721	745
	报纸广告	486	344	246	228	207	194	180

品牌广告(Brand Awareness)

- 创造独特良好的品牌或产品形象，目的在于提升较长时期内的离线转化率



效果广告(Direct Response)

- 有短期内明确用户转化行为诉求的广告。用户转化行为例如：购买，注册，投票，捐款等。

Round Table PIZZA

PALO ALTO/EAST PALO ALTO
263 University Avenue
(Downtown/Delivery to Stanford)
650-322-2893

We Deliver

20% OFF Any Order
Offer excludes beverages, Manager's Specials, Kids Meal or any promotional items.
Offer valid in-store, Carry-out, or Delivery. Limited delivery areas & hours. Minimum delivery fee may apply. Not valid with any other offer or discounts. Expires 7/15/06.

\$15.99
Any Large Specialty Pizza
Original or Skinny Crust only.
Offer valid in-store, Carry-out, or Delivery. Limited delivery areas & hours. Minimum delivery fee may apply. Not valid with any other offer or discounts. Expires 7/15/06.

\$11.99
Any Large 1-Topping Pizza
Original or Skinny Crust only.
Offer valid in-store, Carry-out, or Delivery. Limited delivery areas & hours. Minimum delivery fee may apply. Not valid with any other offer or discounts. Expires 7/15/06.

\$12.99
Any Large 2-Topping Pizza
Original or Skinny Crust only.
Offer valid in-store, Carry-out, or Delivery. Limited delivery areas & hours. Minimum delivery fee may apply. Not valid with any other offer or discounts. Expires 7/15/06.

\$5.00 OFF
ANY X-LARGE PIZZA
\$4.00 OFF
ANY LARGE PIZZA
\$2.00 OFF
ANY MEDIUM PIZZA
Offer valid in-store, Carry-out, or Delivery. Limited delivery areas & hours. Minimum delivery fee may apply. Not valid with any other offer or discounts. Expires 7/15/06.

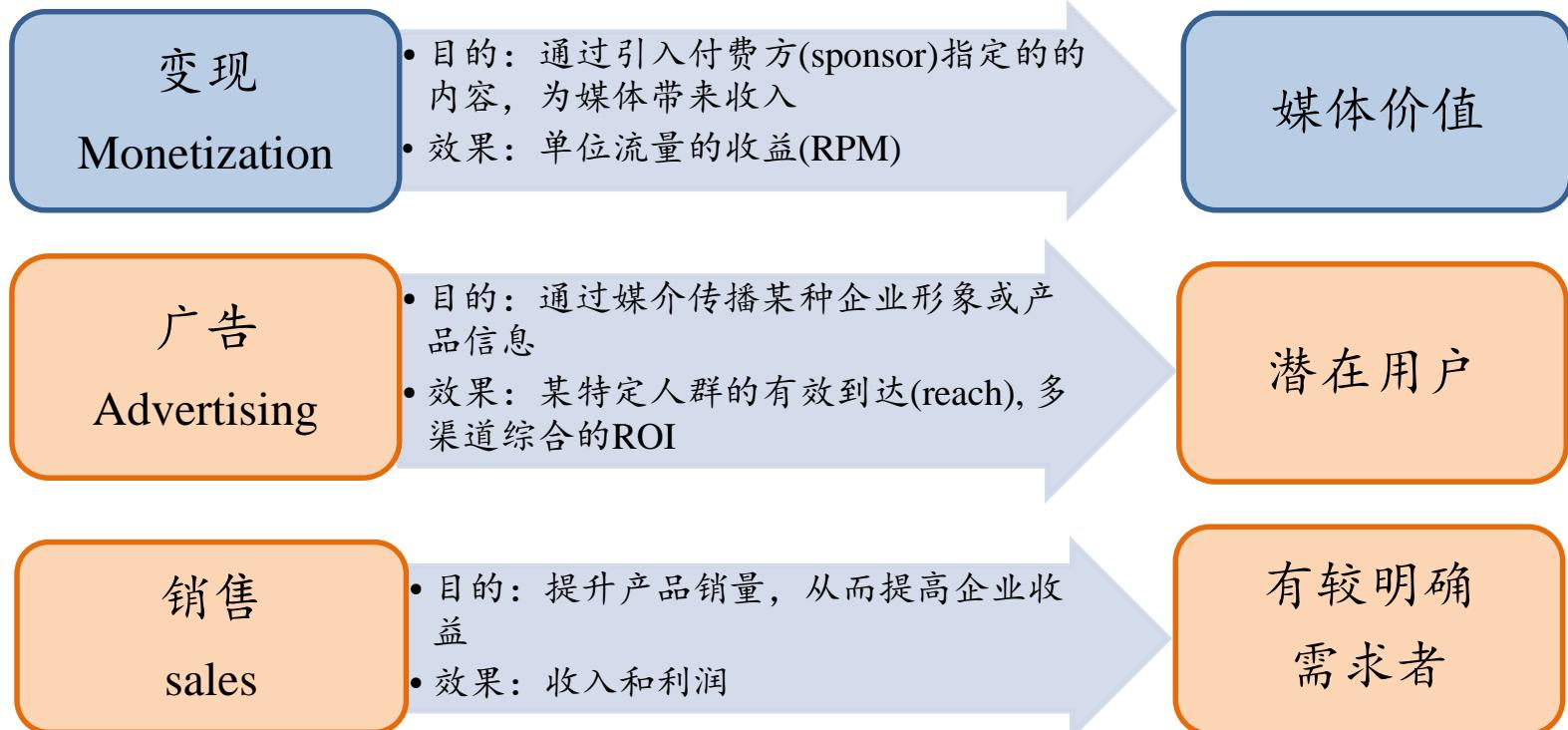
FREE
Medium 1 Topping Pizza
with the purchase of any Large or
X-Large Specialty at regular menu price.
Offer valid in-store, Carry-out, or Delivery. Limited delivery areas & hours. Minimum delivery fee may apply. Not valid with any other offer or discounts. Expires 7/15/06.

ADVISED BY MONEY MAILER OF PALO ALTO/LOS ALTOS/MEN IN VIEJO (800) 980-1231
226-07-0082 226-07-0082

H.O.T! Coupons Web Ad • 226-07-0082F

©2006 Money Mailer LLC
<http://www.hotcoupons.com>

概念：变现、广告与销售



广告的目的是什么？

- 传统的视角：
 - 广告的根本目的是广告主通过媒体达到低成本的用户接触
 - 广告的投入产出比(Return on Investment, ROI)相比于销售人员的劝服活动应该较高
- 互联网新视角：
 - 一切付费的信息、产品或服务的传播渠道，都是广告
 - 直接效果广告的ROI应该可衡量、可优化

在线广告创意类型（一）

- 条幅广告(Banner)



- 文字广告(Textual)

我司代理各报类文字广告,价格优惠,电话:400-831-3770 www.bj-hhtc.cn
华恒天成专业文字广告,资深专业,专注于服务,创造品牌.电话:87765716.

- 邮件营销>Email Direct Marketing, EDM)

		精品订阅	【139邮箱精品订阅】《意林》系列杂志5月里限时免费咯！读经典文摘，享...	5-23(四) 02:46	
		海信	海信电视“节能补贴收官，以旧换新重现”，领取340元折旧券！	5-22(三) 03:19	

在线广告创意类型（二）

- 视频广告(video)

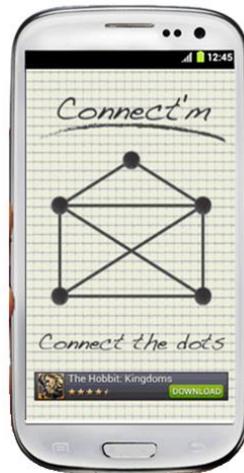


- 富媒体广告(Rich Media)



在线广告创意类型（三）

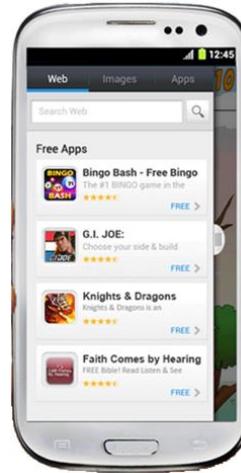
- 移动广告



条幅



开屏



推荐墙

- 社交信息流广告



泛广告商业产品—团购



【4店通用】松鹤楼

100元代金券1张，可叠加使用15张

¥ 80 立减18 立减 门店价 ¥100



【81店通用】必胜客

100元心意美食卡1张，全场通用，可叠加使用

¥ 91 立减18 立减 门店价 ¥100

泛广告商业产品—网址导航

网址管家						管理我的网址
 携程旅行网	 新浪微博	 网易新闻	 新浪财经	 京东商城	 + 添加网址	
好搜·视频	新浪·微博	 天猫 · 优惠精选	搜狐·视频	网易·邮箱	凤凰·军事	
360游戏	淘宝网	商城·团购	百度·热搜榜	360·手机卫士	人人·QQ空间	
爱淘宝	 360影视	土豆	360彩票	东方财富·理财	人民·新华	
京 京东商城	58 同城	太平洋电脑	世纪佳缘	同程旅游	中关村在线	
汽车之家	易车网	太平洋汽车	百合网	乐蜂网	苏宁易购	
糯米团购	央视网·直播	 赶集网	当当网	牛途牛旅游网	1 1号店	
珍爱婚恋网	梦芭莎	亚马逊	艺龙网	去哪儿网	猎聘网	
美团网	国美在线	唯品会 	聚美优品 	搜房网	美丽说	

计算广告基础

广告有效性模型



曝光效果突出的广告位

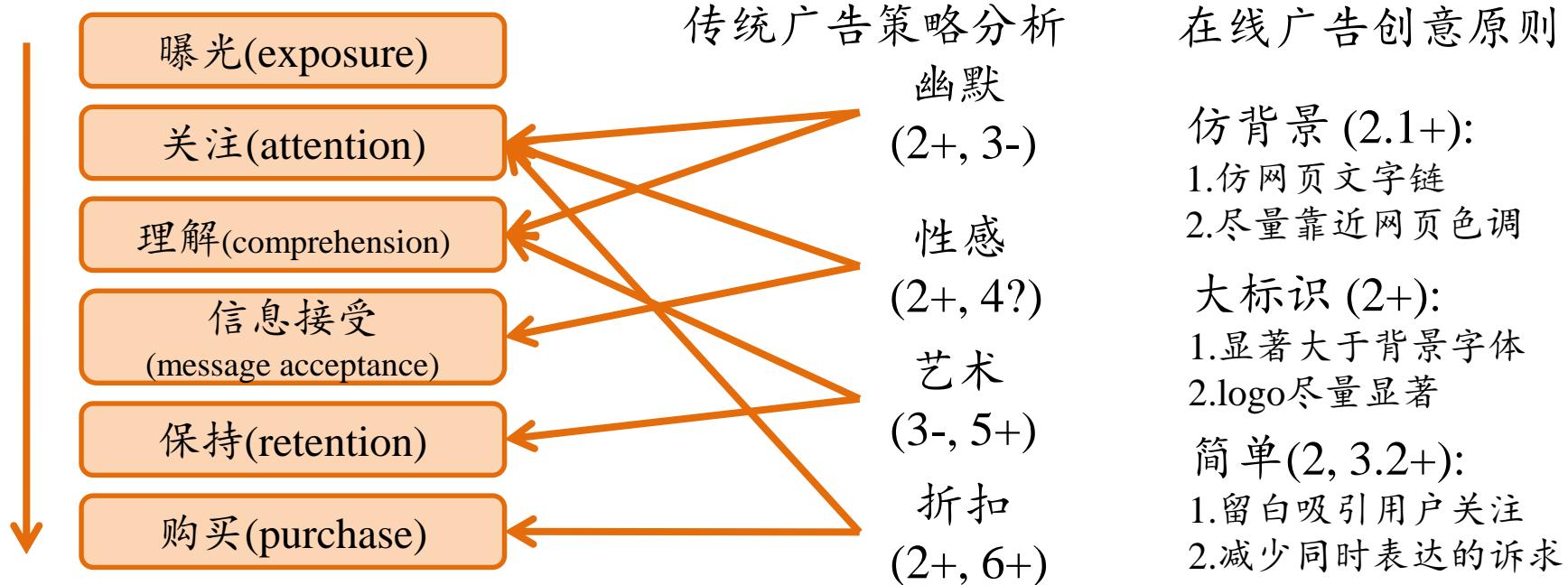


纽约时代广场



北京东三环京信大厦

一些广告策略的效果



计算广告核心挑战

- 计算广告的核心问题，是为一系列用户与环境的组合，找到最合适的广告投放策略以优化整体广告活动的利润。
- 优化问题描述：

$$\max_{a_1, \dots, T} \sum_{i=1}^T \{r(a_i, u_i, c_i) - q(a_i, u_i, c_i)\}$$

决策对象：一组广告展示 收入(eCPM) 成本

广告 用户 上下文

广告收入的分解



在线广告结算方式

结算方式	点击率估计	点击价值估计	优缺点	适用场景
CPT	需求方		<ol style="list-style-type: none">可以充分发挥橱窗效应。无法利用受众定向技术。	高曝光的品牌广告
CPM	需求方		<ol style="list-style-type: none">可以利用受众定向选择目标人群。合约售卖下，受众划分不能过细。	<ol style="list-style-type: none">有受众选择需求的品牌广告实时竞价广告交易
CPC	供给方	需求方	<ol style="list-style-type: none">可以非常精细地划分受众人群。比较合理的供给方和需求方分工。	竞价广告网络
CPS	供给方		<ol style="list-style-type: none">需求方无任何风险。供给方运营难度较大。	<ol style="list-style-type: none">效果类广告联盟效果类DSP

广告行业协会



- Interactive Advertising Bureau
 - 在线广告供给方的行业协会，推动数字化市场营销行业的发展
 - 制定市场效果衡量标准和在线广告创意的标准
 - 会员：Google, Yahoo, Microsoft, Facebook等
-



- American Association of Advertising Agencies
 - 主要的协议是关于广告代理费用的收取约定，以避免恶意竞争
 - 主要集中在创意和客户服务，在线业务是一部分
 - 会员：Ogilvy & Mather, JWT, McCann等，Dentsu等非4A会员的大公司但也被列为4A公司
-

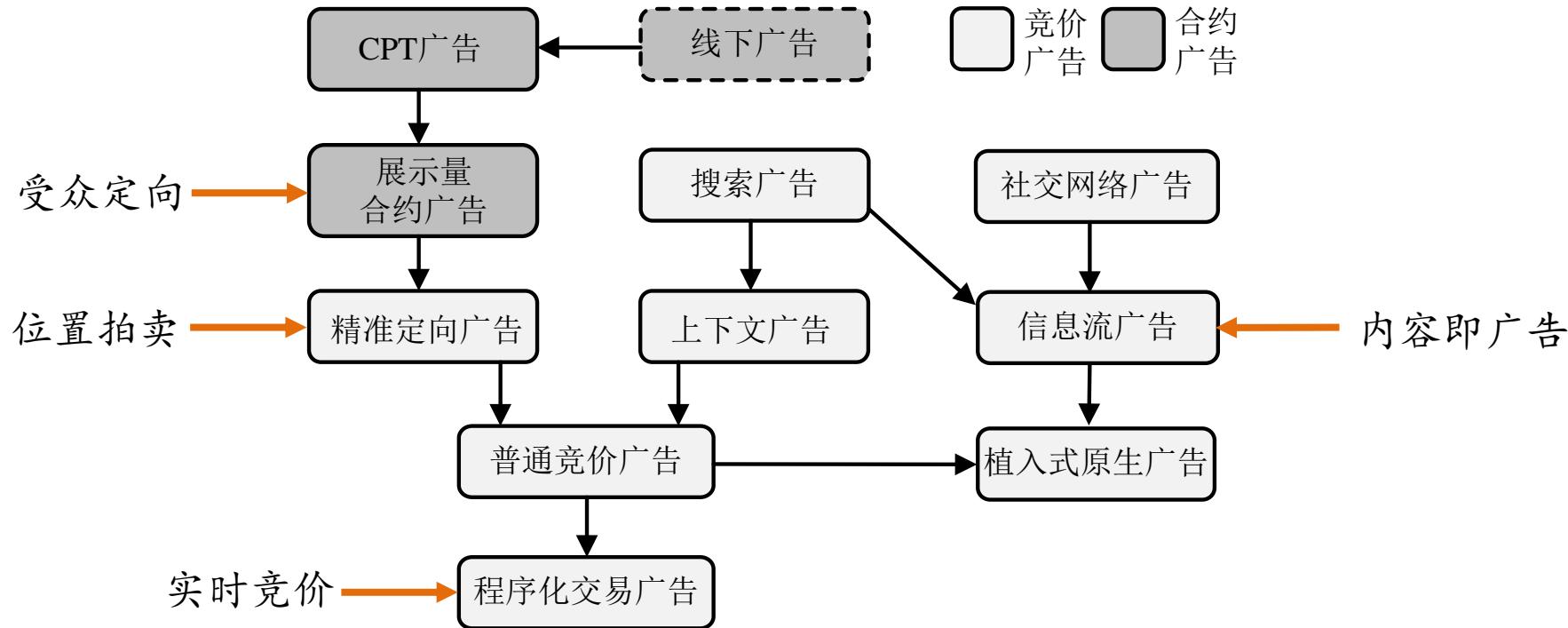


Leading the Marketing Community

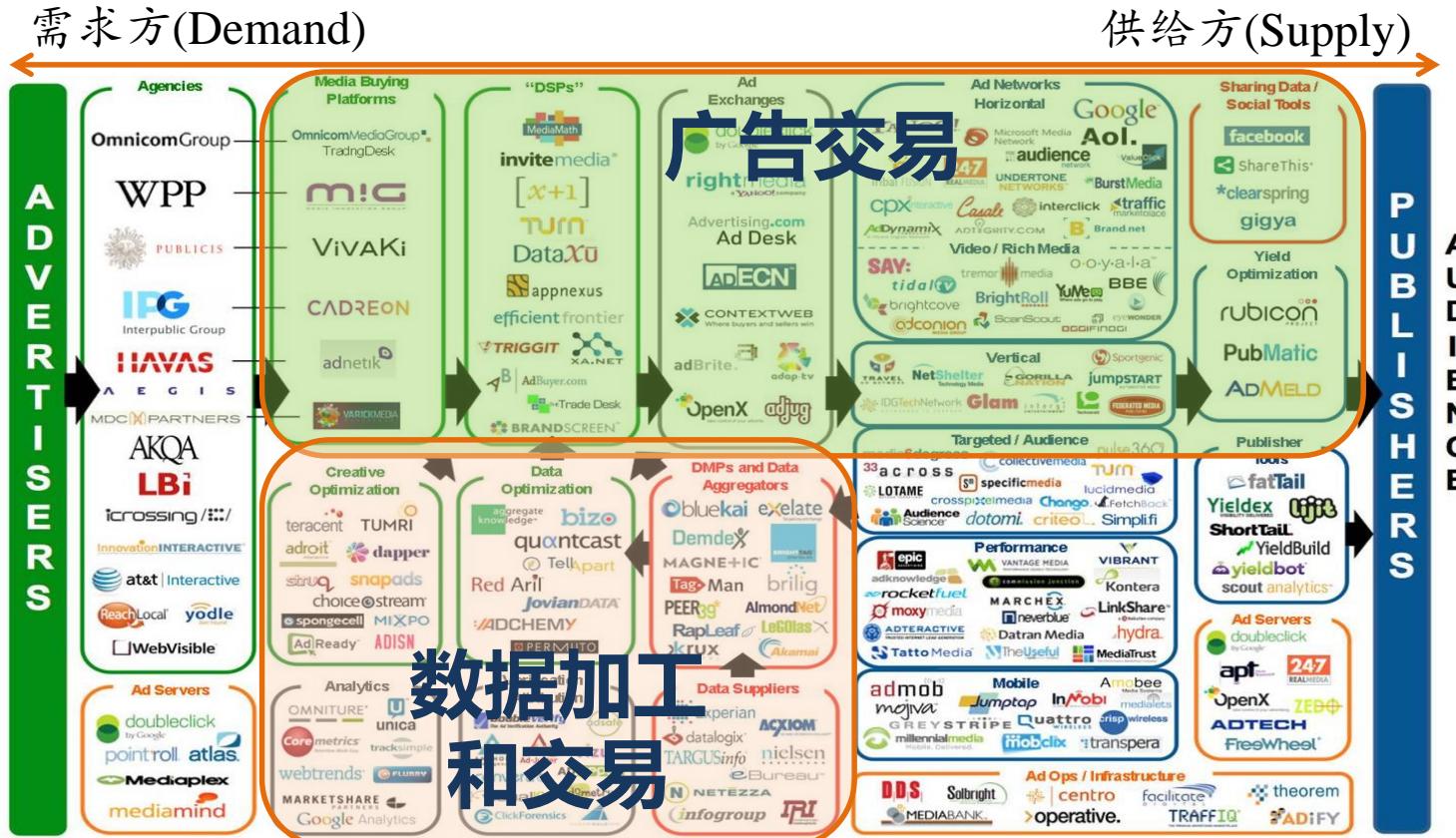
- Association of National Advertisers
- 主要代表广告需求方的利益(也有媒体和代理会员)
- 会员：AT&T, P&G, NBA等

计算广告 产品技术概览

在线广告产品历程



广告技术市场格局



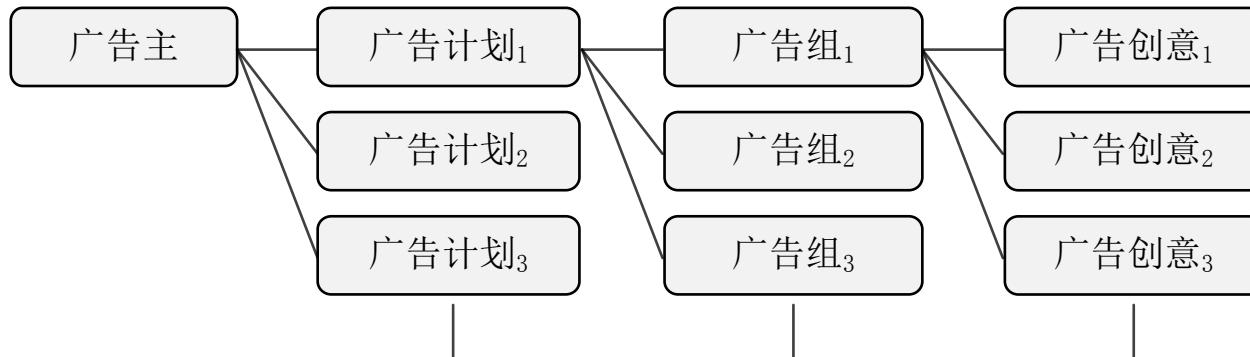
流量变现与数据变现

- 流量变现
 - 自有流量RPM水平的不断提升
 - 典型产品：搜索广告、导航广告、门户品牌广告、社交网站信息流广告
- 数据变现
 - 获取非自有流量，通过对用户行为的理解，提升其RPM水平从而获得差价
 - 是流量变现的护城河，不断吸引广告主的关键
 - 典型产品：Ad Network(网盟), DSP, DMP

商业产品设计运营原则

- 商业产品，即面向商业客户而非一般用户的产品，其中最典型的代表就是广告产品。
- 相对于产品功能，要特别关注产品中的策略部分。
- 要特别关注数据，让运营和产品优化形成闭环。所有产品特征和策略的成功与否，要严格根据数据的反馈来判断。
- 优化的是确定的商业目标，而非使用便捷性。

广告需求方层级组织



- **广告活动(Campaign)**概念上对应于广告主的一次投放合同，其中包括了预算、时间范围等基本信息。
- **广告组(Ad Group)**对应于一个具体的广告投放策略，主要是设定受众定向条件和出价。
- **广告创意(Creative)**则是最终展示出来的素材，可能在同一个组策略下有不同尺寸的创意存在。

广告计划、广告组设置示意

设置推广计划

* 推广计划名称 :

加载现有计划的设置 : 选择推广计划

* 选择投放网络 : 精选网络

标准展示广告 - 在PC端的流量中投放您的展示广告

关键词文字链广告 - 针对搜索过特定关键词的人群投放文字链广告

商品推荐广告 - 您的产品库没有符合条件的商品 [了解详情](#)

预算和时间

--预算--

* 每日预算 : 不限预算
不限预算时,每日投放量没有上限,设置预算后,当消费达到预算后即不再投放。

* 预算分配控制 : 标准 : 将每日预算均匀地分配到整个投放日程中
 加速 : 尽可能地将每日预算花费出去

--时间--

* 开始时间 :

结束时间 : 为空表示不限结束时间。

* 按小时投放 : 全日程展示广告 (当前排期时区: GMT+08:00) [自定义](#)

基本设置 :

* 推广组名称 :
建议您根据媒体、人群或活动内容来命名,以便于调整投放和报表查看。

加载现有推广组 : 选择推广组

哪些用户可以看到我的广告

-地域定向-

国家/地区 : 中国大陆 [修改](#)

地区选择 : 不限 选择省份 选择一二线城市

排除地区 :

-人群定向-

指定关键词 : 已添加关键词数量 : 0/5000 [修改](#)

用户在哪里看到我的广告

- 媒体定向 --

选择媒体 : 精选资源 指定广告位

高级设置

出价 (按第二高价和创意质量度计费, 将不会高于您的实际出价)

默认出价(¥): CNY/ CPC CPM

建议出价(¥) : 0.10-1.15

(推广组的出价将作为每个关键词的默认出价, 您也可以在关键词列表中单独修改每个关键词的出价)

计算广告技术特点

- 技术和计算导向
 - 数字媒体的特点使在线广告可以按人群投放
 - 广告决策和交易朝着计算驱动的方向发展
- 可衡量性
 - 广告的点击是效果的直接收集途径
 - 98至今，Banner点击率从10%降至0.1%
- 标准化
 - 技术投放和精准定向促进了在线广告标准化
 - **iab.**在标准化方面做了很多贡献



计算广告中的技术问题

$$\max_{a_1, \dots, T} \sum_{i=1}^T \{r(a_i, u_i, c_i) - q(a_i, u_i, c_i)\}$$

- 从优化角度来看
 - 特征提取: 受众定向
 - 微观优化: eCPM 估计
 - 宏观优化: 机制设计
 - 受限优化: 在线分配
 - 强化学习: 探索与利用
 - 个性化重定向: 推荐技术

- 从系统角度来看
 - 整体框架: 广告服务器
 - 候选查询: 实时索引
 - 特征存储: No-sql 技术
 - 离线学习: Hadoop
 - 在线学习: 流计算
 - 交易市场: 实时竞价

主要广告产品优化目标分解

	产出(r)		投入(p)	约束
	点击率(μ)	点击价值(v)		
GD	constant		constant	合约的展示量要求
Ad Network	$\mu(a, u, c)$	$\text{bid}_{\text{CPC}}(a)$	$\propto r$	无
Ad Exchange		$\text{bid}_{\text{CPM}}(a)$	$\propto r$	带宽、服务成本
DSP	$\mu(a, u, c)$	$v(a, u)$	$p(a, u, c)$	预算限制

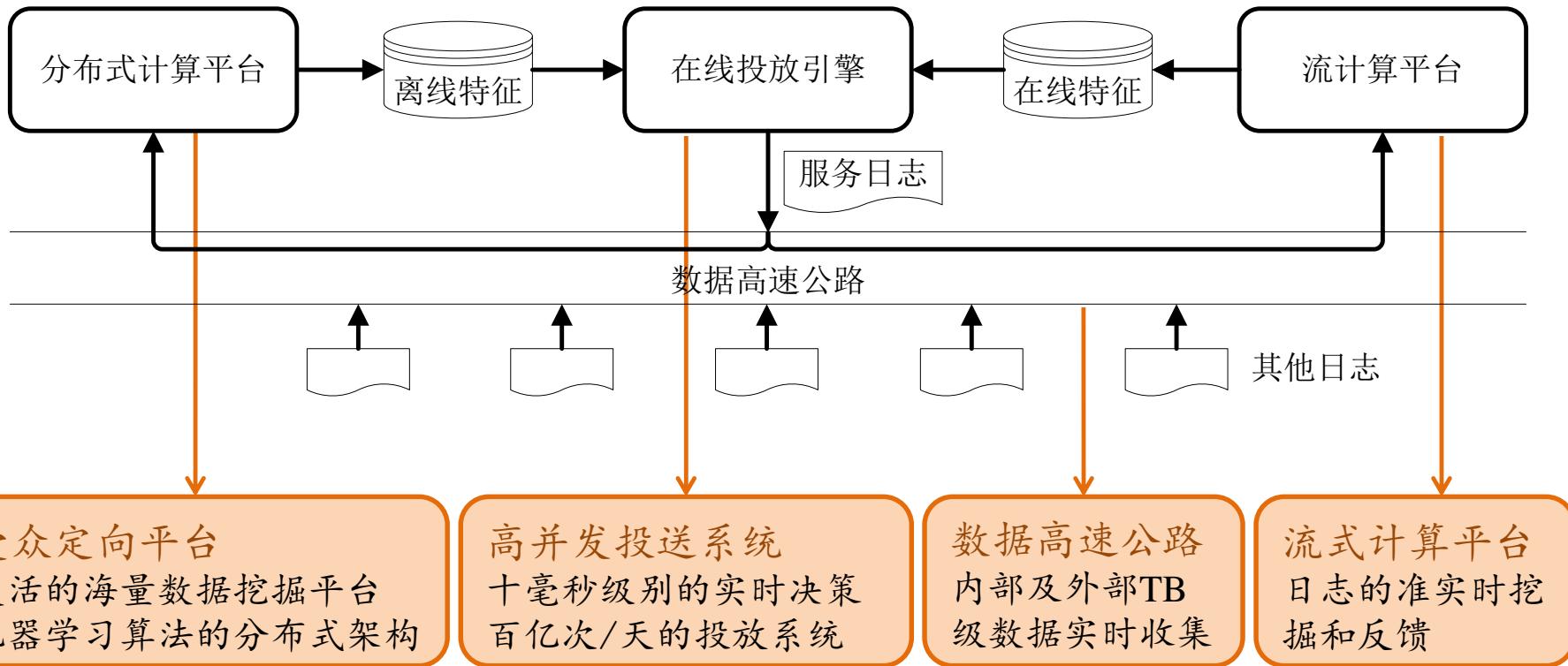
计算广告技术难点

- 大规模(Scale)
 - 百万量级的页面，十亿量级的用户，需要被分析处理
 - 高并发在线投放系统(例:ADX每天处理百亿次广告交易)
 - Latency的严格要求(例:RTB要求竞价在100ms内返回)
- 动态性(Dynamics)
 - 用户的关注和购物兴趣非常快速地变化
- 丰富的查询信息(Rich query)
 - 需要把用户和上下文中多样的信号一起用于检索广告候选
- 探索与发现(Explore & exploit)
 - 用户反馈数据局限于在以往投放中出现过的 (a, u, c) 组合，需要主动探索未观察到的领域，以提高模型正确性

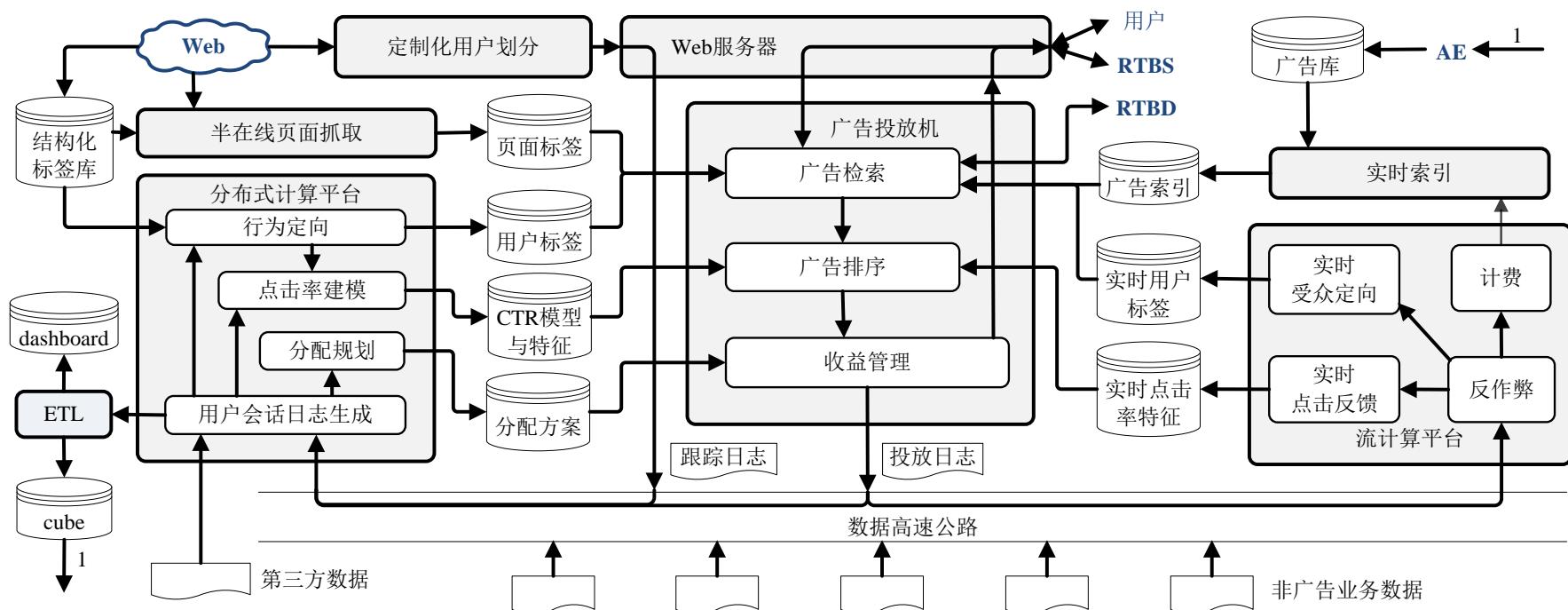
Web-scale技术问题比较

	搜索	搜索广告	显示广告	个性化推荐
主要准则	相关性	利润		用户兴趣
其他目标	垂直领域决定	质量、安全性		多样性、新鲜度
索引规模	十亿级	百万 千万级	百万级	百万级 亿级
个性化	较少的个性化需求		亿级规模用户上的个性化	
检索信号	较明确		较分散	
Downstream优化	不适用			适用

个性化系统一般框架

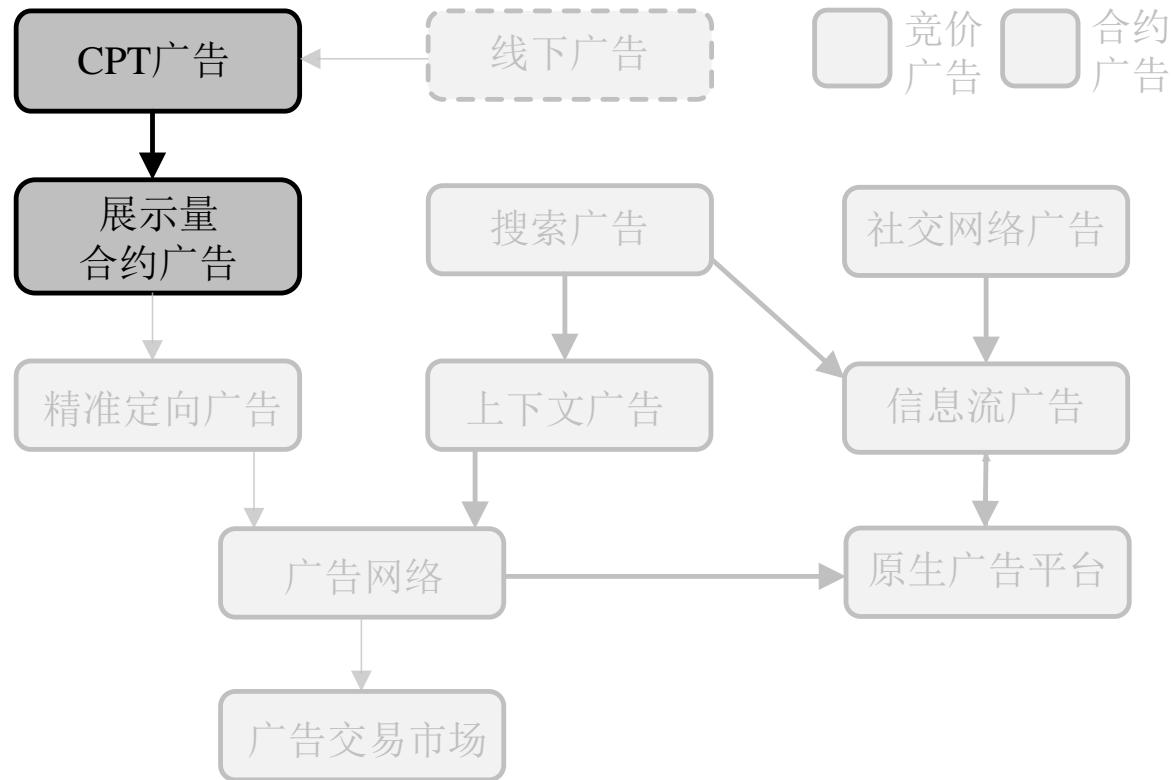


计算广告系统架构



合约广告

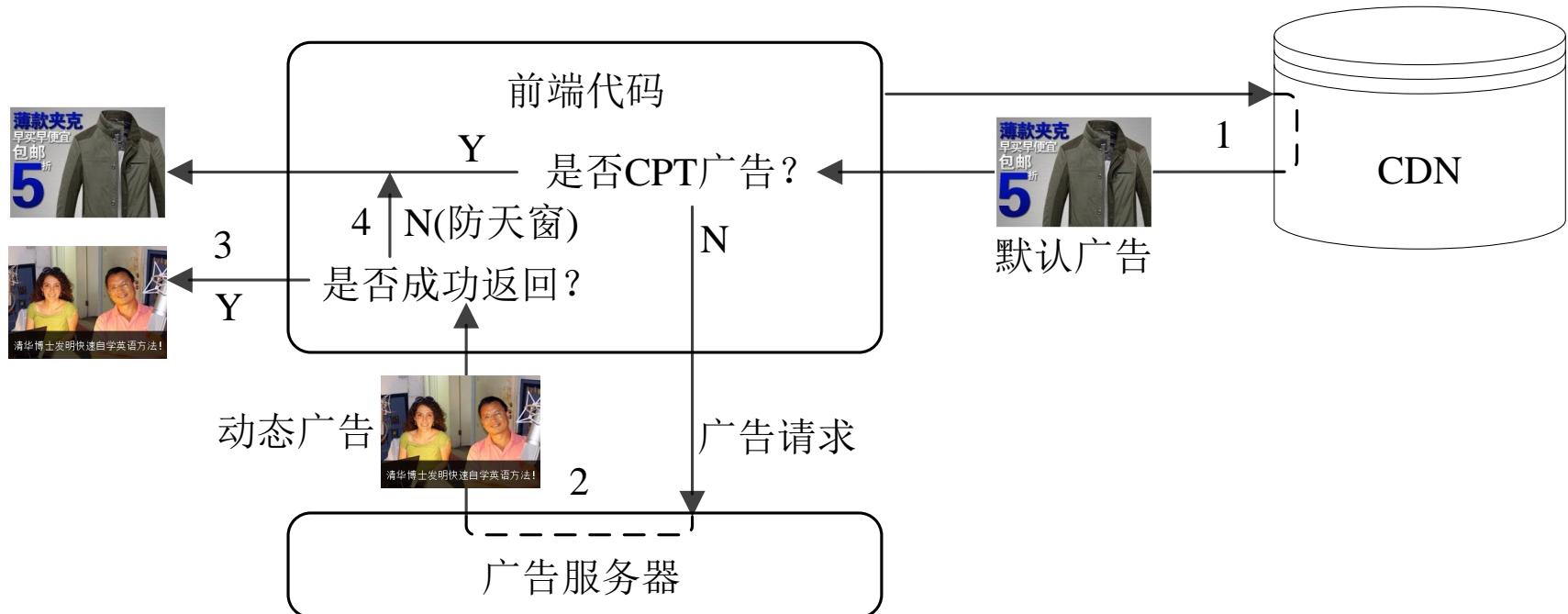
合约广告



广告位(CPT)合约

- 供给方：广告排期系统
 - 帮助媒体自动执行多个合同的排期
 - 不提供受众定向，可以将广告素材直接插入页面
- 需求方：代理商
 - 帮助广告商策划和执行排期
 - 用经验和人工满足广告商质和量的需求
- 典型场景
 - 高曝光品牌广告，移动应用开屏广告

排期与动态广告混合系统



展示量合约与担保式投送

- 展示量合约
 - 基于合约的广告机制，有时未完成约定量要向广告商补偿
 - 多采用千次展示付费(Cost per Mille, CPM)方式结算
- 担保式投送(Guaranteed Delivery, GD)引擎
 - CPM方式要求广告投送由服务器端完成决策
 - 三大技术基础：受众定向、流量预测、在线分配
- 典型场景
 - 视频广告、富媒体广告、品牌条幅广告

简化的流量分配模型

需求节点(Demand Nodes, 订单要求的定向标签组合)

A:

{gender=男,
200 mille}

{geo=广东,
200 mille}

{age=2, 1,000
mille}

E:

I:

{gender=男,
age=2,
400 mille}

{gender=男,
geo=上海,
age=2
400 mille}

{gender=男,
geo=广东,
age=2
100 mille}

{geo=广东,
age=2
100 mille}

{geo=北京,
age=2
500 mille}

{age=2
300 mille}

供给节点(Supply Nodes, 定向标签的最细组合)

假设: 节点内部的流量差异可以忽略

在线分配问题框架

供给节点*i*的总量

$$\max \sum_{(i,a) \in E} s_i x_{ia} r_{ia}$$

$$\text{s.t. } \sum_{a \in \Gamma(i)} x_{ia} \leq 1, \quad \forall i \in I$$

$$\sum_{i \in \Gamma(a)} s_i x_{ia} q_{ia} \leq d_a, \quad \forall a \in A$$

$$x_{ia} \geq 0, \quad \forall (i, a) \in E$$

优化变量 x_{ia} : 将供给节点*i*
分配给需求节点*a*的比例

将供给节点*i*连接到需求
节点*a*的单位流量惩罚

在线分配典型问题

GD问题：

$$\begin{aligned} \max \quad & C \\ \text{s.t.} \quad & \sum_{a \in \Gamma(i)} x_{ia} \leq 1, \quad \forall i \in I \\ & \sum_{i \in \Gamma(a)} s_i x_{ia} \geq d_a, \quad \forall a \in A \\ & x_{ia} \geq 0, \quad \forall (i, a) \in E \end{aligned}$$

Adwords问题：

$$\begin{aligned} \max \quad & \sum_{(i,a) \in E} q_{ia} s_i x_{ia} \\ \text{s.t.} \quad & \sum_{a \in \Gamma(i)} x_{ia} \leq 1, \quad \forall i \in I \\ & \sum_{i \in \Gamma(a)} q_{ia} s_i x_{ia} \leq d_a, \quad \forall a \in A \\ & x_{ia} \geq 0, \quad \forall (i, a) \in E \end{aligned}$$

约束优化的拉格朗日法

原问题(Primary problem)

$$\begin{aligned} & \min f(x) \\ \text{s.t. } & g(x) \leq 0, \quad h(x) = 0 \end{aligned}$$

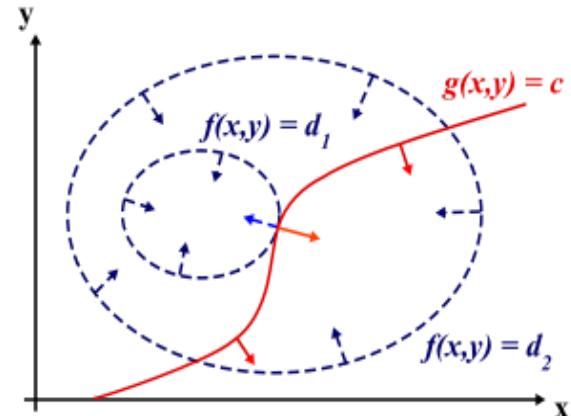
拉格朗日对偶函数
(Lagrangian dual function)

$$L(\lambda, \nu) = \inf_x [f(x) + \lambda^\top g(x) + \nu^\top h(x)]$$

对偶问题 (Dual problem)

$$\max L(\lambda, \nu) \quad \text{s.t. } \lambda \geq 0$$

- 等式约束下的几何意义见右，
KKT为保证此方法有效的条件
- 凸优化情形下满足KKT条件，
但注意一些非凸优化也满足



极限性能研究

- 极限性能：某策略与Oracle策略目标函数比例的下限， ε -complete
- Lagrangian: $\sum_{(i,a)} r_{ia} s_i x_{ia} + \sum_i \alpha_i \left[\sum_{a \in \Gamma(i)} s_i x_{ia} - s_i \right] + \sum_a \beta_a \left[\sum_{i \in \Gamma(a)} q_{ia} s_i x_{ia} - d_a \right] - \sum_{(i,a)} \gamma_{ia} s_i x_{ia}$
- 策略框架：
 - 初始化每个需求约束的对偶变量 $\beta_a \leftarrow 0$;
 - 当一次展示 i 到达时，令 $a' \leftarrow \operatorname{argmax}_a r_{ia} - \beta_a$; (如果该值对所有的广告都为负，则所有合约都不需要分配。)
 - 令 $x_{ia'} = 1$, 如果 a' 已被分配了 $d_{a'}$ 次展示，令 $i' \leftarrow \operatorname{argmin} x_{ia'}, x_{ia'} \leftarrow 0$;
 - 在对偶问题中，令 $\alpha_i = r_{ia'} - \beta_{a'}$, 并通过一定的规则来更新 $\beta_{a'}$ 。不同的规则对应了不同的分配算法，也会导致不同的极限性能。

不同更新规则对应极限性能

策略	描述	有效性
greedy	β_a 是分配给 a 的前 d_a 个高权重展示中最低的权重，也即 a 接受一个新的展示需要抛弃的权重。	$\frac{1}{2}$ -competitive
uniform weighting	β_a 是分配给 a 的前 d_a 个高权重展示的权重的算术平均。如果分配给 a 的展示少于 d_a 个， β_a 是这些展示总权重与 d_a 的比。	$\frac{1}{2}$ -competitive
exponential weighting	β_a 是分配给 a 的前 d_a 个高权重展示的权重的指数加权。即： 设 $r_1 \leq r_2 \leq \dots \leq r_{d_a}$ ，则： $\beta_a = d_a \left\{ \left[(1 + d_a^{-1})^{d_a} \right] \right\} \sum_{j=1}^{d_a} r_j (1 + d_a^{-1})^{j-1}$	$(1 - \frac{1}{e})$ -competitive

基于对偶算法的紧凑分配方案

- 紧凑分配方案(Compact Allocation plan)
 - 变量数正比于 $|A|$, 而非 $|E|$ 或 $|I|$
 - 分配策略最好是无状态的, 这样可以避免服务器之间的同步
- 由 α 恢复 β 和分配变量 x :

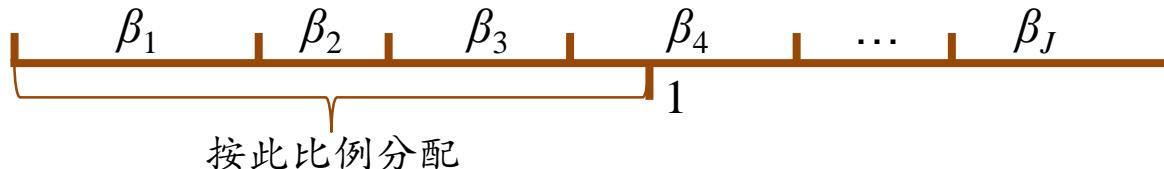
$$\sum_{a \in \Gamma(i)} r_{ia}^{-1} (\beta_a - \alpha_i) = 1$$

$$x_{ia}^* = \max \left\{ 0, r_{ia}^{-1} (-q_{ia} \sum_k \beta_k - \alpha_i) \right\}$$

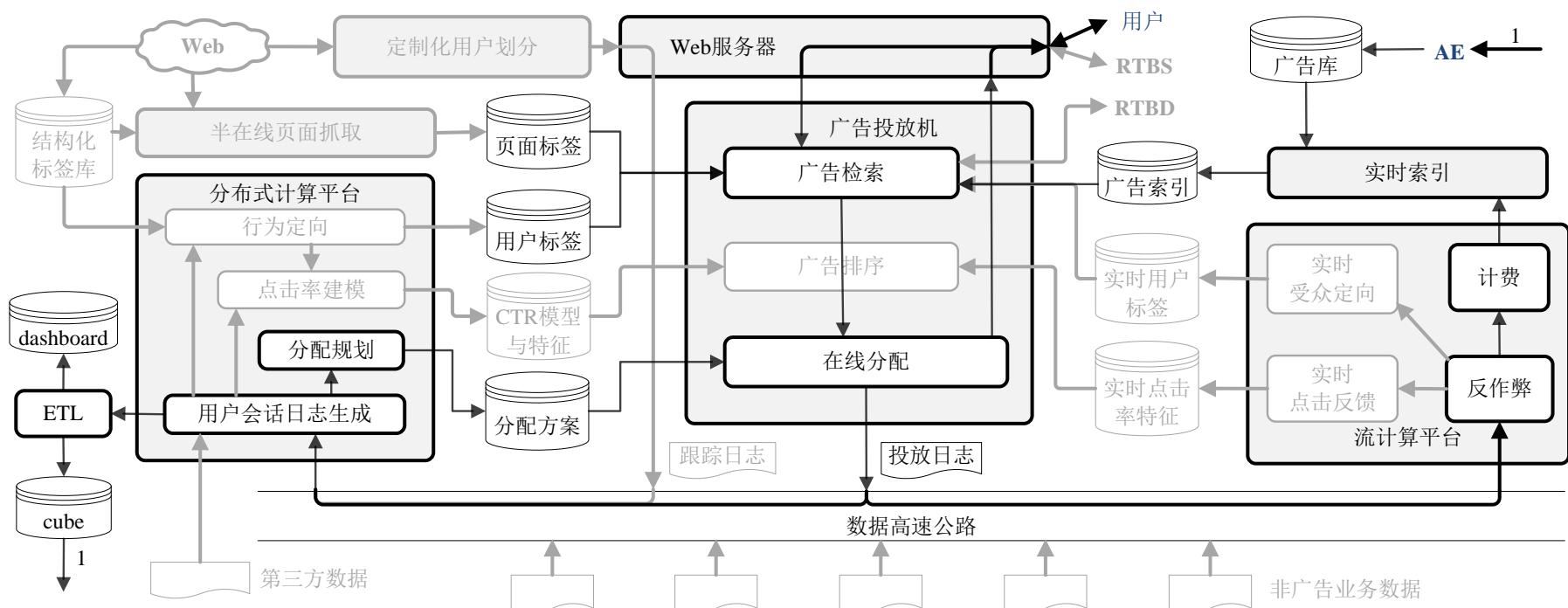
- 由历史数据求解上述问题规模太大, 需要对数据作一些采样以便更高效地得到分配方案。

启发式的High Water Mark算法

- 离线规划
 - 令每个供给节点*i*的剩余supply等于预测量 $r_i \leftarrow s_i$
 - 将每个*a*按照 $\Gamma(a)$ 中所有供给流量做升序排列，确定分配优先级
 - 按照分配优先级对每个*a*解下式得到其投放率 β_a ： $\sum_{i \in \Gamma(a)} \min\{r_i, s_i \beta_a\} = d_a$
 - 对 $\Gamma(a)$ 中的每个*i*, 令 $r_i \leftarrow r_i - \min\{r_i, s_i \beta_a\}$
- 在线分配
 - 对某次展示, $\{a_1, \dots, a_J\}$ 为按照分配优先级排序的所有满足要求的广告
 - 将上面的广告按投放率随机分配其展示机会



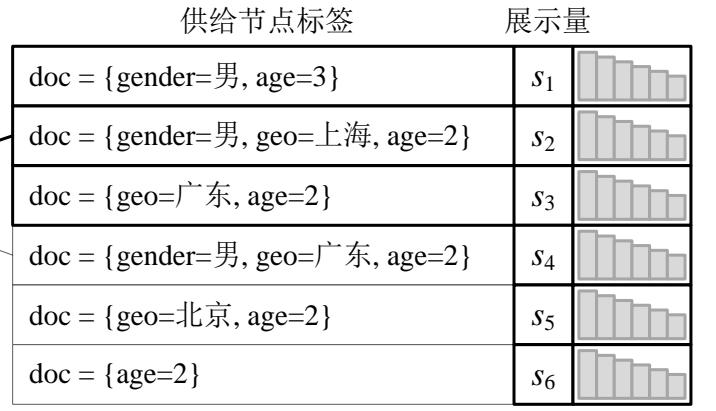
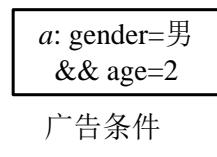
担保式投送系统架构



流量预测

- 可视为query为 a , 对流量进行检索的反向查询问题
- 供给节点(i)索引

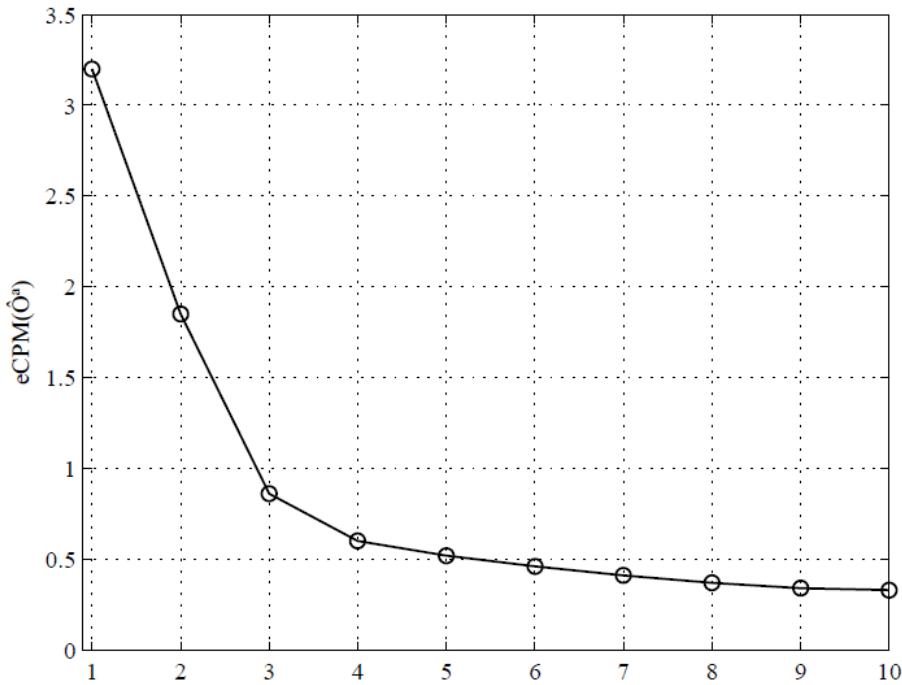
- i , s_i , Hist_i
- 供给 流量 eCPM分布



- 预测过程

- 给定 a , 首先通过 u 的索引找出所有符合条件 c 的集合
- 对每个 u 估计 $r(a, u)$, 并根据 $p_u(\text{eCPM})$ 得到 a 在 u 上胜出的百分比 $p(a, u)$, 并将 a 的流量累加 $p(a, u) \cdot \#\text{impression}_u$

频次控制



- 问题定义
 - 限制各 (a, u) 组合在一定时间周期的展示量
- 问题特性
 - 频次存储的规模是有上界的。 (a, u) 组合总数不会超过时间周期内的展示总数。
 - 生成用 (a, u) 对应的key时，不需要处理冲突，因此不需要哈希结构。
 - 专用的、轻量级的内存表比大多数No-SQL数据存储方案更实用。

YAHOO! 展示广告市场

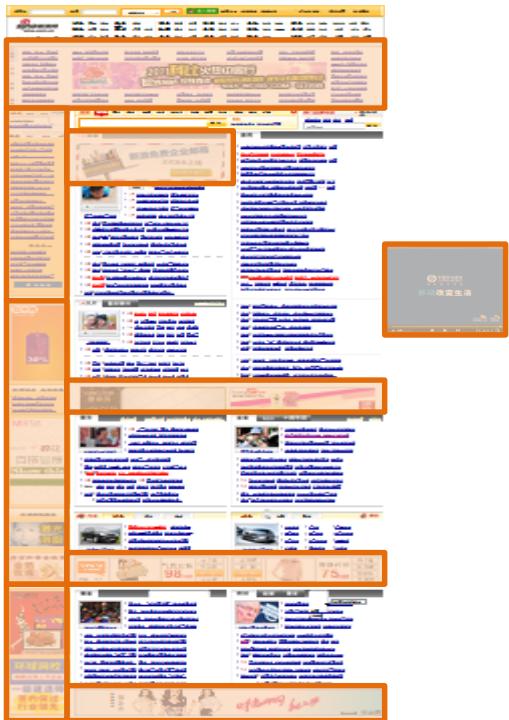
- 核心业务
 - GD, 无法分配的流量转接到NGD(non-guaranteed delivery, 即 Rightmedia exchange)进行变现
 - GD市场广告主数量为几千, 年收入为Billion量级
- 其他点评
 - 采用compact allocation plan完成线上决策
 - 提供地域、人口属性、行为定向(常用的仅有几十个标签)
 - 合约式销售中, 品牌广告主对曝光有独占要求

受众定向

中国互联网用户桌面



受众售卖与广告位售卖



YAHOO! Web Images Video Local Apps More ... Sign In New here? Sign Up Page Options

YAHOO! SITE S Mail Autos Dating Finance (Dow Jones) Flickr Games Health Horoscopes Jobs Messenger Movies omg! Real Estate Shine Shopping Sports Travel Weather (S) More Yahoo! Site S

TODAY - July 19, 2011

Murdoch faces grilling over hacking

The media baron and his son will appear before parliament in the wake of the growing scandal. Click for details... Murdoch faces grilling over hacking

6. Tour de France 7. Brain injury 8. Martin Sheen 9. Barbara Boxer 10. Tampa Bay Rays

Yahool Autos Find Your New Car Search

VIDEO PICKS City Hot Go to Video

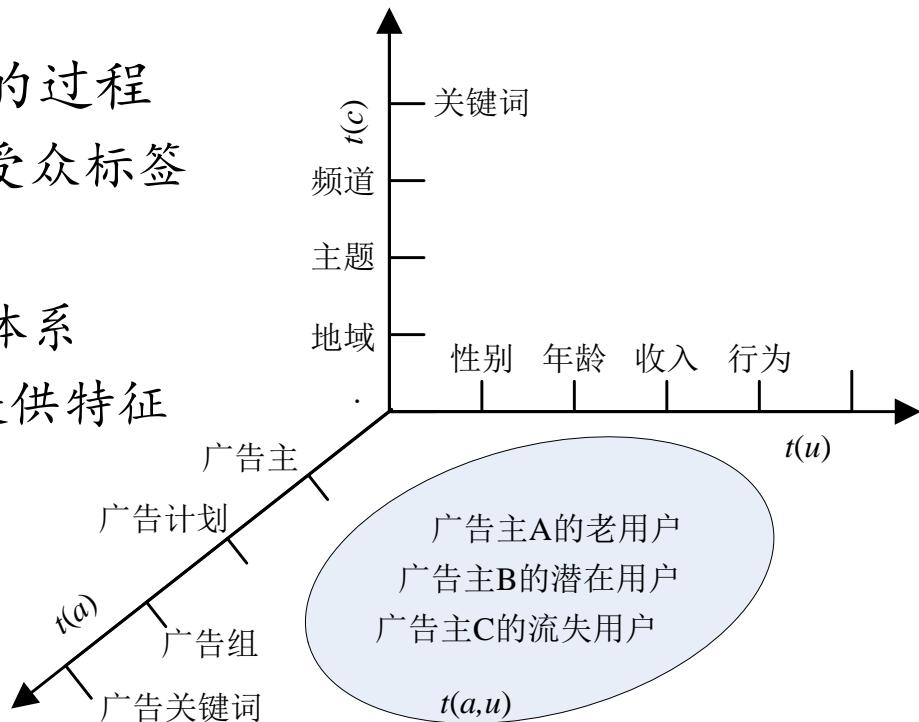
GPS devices help parents track kids Dramatic rescue of boy from rising water

Upgrade to Internet Explorer 9

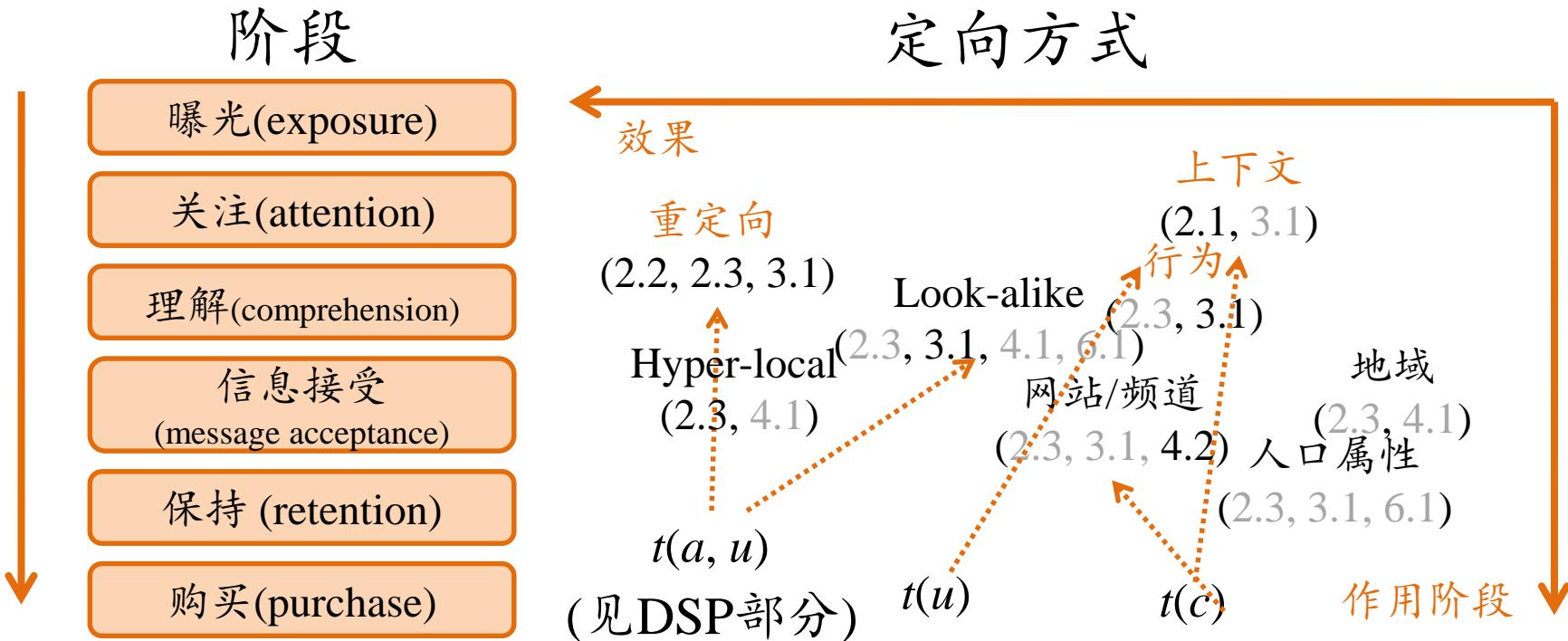
New Yahoo!-optimized Internet Explorer 9 offers a fast, elegant, and trusted way to enjoy your favorite Yahoo! sites.

受众定向方法分类

- 受众定向即为 (a, u, c) 打标签的过程
 - 上下文标签可以认为是即时受众标签
- 标签的两大主要作用
 - 建立面向广告主的流量售卖体系
 - 为各估计模块(如CTR预测)提供特征



常见受众定向方式



常用受众定向方式（一）

- 地域定向(Geo-targeting)
- 人口属性定向(Demographical targeting)
 - 年龄、性别、教育程度、收入水平
 - 可监测，难获得，对效果意义有限
- 频道定向(Channel targeting)
- 精确位置定向(Hyper-local Targeting)
 - 利用蜂窝信息或者GPS获得精准位置，适用于O2O广告主
- 上下文定向、行为定向(见后)

常用受众定向方式（二）

- 重定向(Retargeting)
 - 针对广告主老用户或相关用户进行再营销，属于定制化用户标签
 - 主要利用第一方数据
- 新客推荐(Look-alike)
 - 根据对广告主提供的种子用户，为其找到相关的新客
 - 同时利用第一方和第二方数据
- 团购(Group Purchase)
 - 并不是一种广告定向技术，但对价格敏感人群有筛选作用

受众定向标签体系

- 结构化标签体系
 - 按照某分类法(Taxonomy)制定一个层次标签体系，父节点与子节点在人群覆盖上是包含关系。
 - 主要用于面向品牌广告的受众定向，特别常用于GD系统中。
- 非结构化标签体系
 - 根据某类定向需求设置标签，标签并不能为同一个分类体系中所描述
 - 适用于多种目标、特别是效果目标并存的广告主的精准流量选择要求
- 关键词
 - 按照搜索或浏览内容的关键词划分人群
 - 非结构化，容易理解，但操作和优化不容易

YAHOO! GD受众定向标签

一级标签	二级标签
Finance	Bank Accounts, Credit Cards, Investment, Insurance, Loans, Real Estate, ...
Service	Local, Wireless, Gas & Electric, ...
Travel	Europe, Americas, Air, Lodging, Rail, ...
Tech	Hardware, Software, Consumer, Mobile, ...
Entertainment	Games, Movies, Television, Gambling, ...
Autos	Econ/Mid/Luxury, Salon/Coupe/SUV, ...
FMCG	Personal care, ...
Retail	Apparel, Gifts, Home, ...
Other	Health, Parenting, Moving, ...

 bluekai 受众定向标签

类别	描述	数据来源	用户规模
Intent	最近输入词表现出某种产品或服务需求的用户	Bluekai Intent	160+MM
B2B	职业上接近某种需求的用户	Bizo	90MM
Past Purchase	根据以往消费习惯判断可能购买某产品的用户	Addthis, Alliant	65+MM
Geo/Demo	地理上或人口属性上接近某标签的用户	Bizo, Datalogix, Expedia	
Interest/LifeStyle	可能喜欢某种商品，或某种生活风格的用户	Forbes, i360, IXI, ...	103+MM
Qualified Demo	多数据源上达成共识验证一致的人口属性	多数据源	90+MM
Estimated Financial	根据对用户财务状况的估计做的分类	V12	

上下文定向(Contextual targeting)

- 根据用户正在浏览的页面和其他信息投送广告

首页 » 婚礼策划方案发布区 » 婚礼策划方案发布区 » 琴缘·知音--婚礼策划方案

琴缘·知音--婚礼策划方案

时间：2014-01-03 09:49 作者：分享婚礼 阅读：21

【基本素材】

新郎小军，在区司法所工作。新娘小娜，是一位小学音乐老师。5年前，一位算命先生给他说：“今年你会遇到一个特别的人，并最终娶她，和她幸福的相伴一生。”在初春的某一天，小军追随着小巷里的琴声，遇到了小娜，并开启了他们的浪漫故事。在小娜的帮助和鼓励下，小军刻苦学习，终于通过了律师资格考试。他们将在金秋时间喜结连理，共享他们曾经在“梦中的婚礼”。

【思考过程】

他们音琴结缘，结为知心爱人，因此我把这场婚礼的主题确定为“琴缘·知音”。他们的相遇相识、相约相许和缔结婚姻，正好在三个不同的季节，于是，我考虑把婚礼也以季节分为3段，用灯光的颜色绿、白、黄来营造季节特征，给婚礼划分段落。

给他们充当媒人的乐曲正是那首《梦中的婚礼》，浪漫深情，我考虑整场婚礼就用这首曲子作为基调，作为主旋律，一方面，《梦中的婚礼》扣住婚礼主题，另一方面，一种基调的音乐贯彻婚礼始终，使婚礼

网 故事当基调 似乎一切都被赋予生命中 所以 我考虑在婚礼的开头和结尾 都用曾令先生的福音声

A vertical advertisement for "Wedding Planner Advanced Training". It features a man in a tuxedo and bow tie. To the right of the image, the text "婚礼督导师高级培训" is written vertically.

申请投稿

热门主题

相关主题

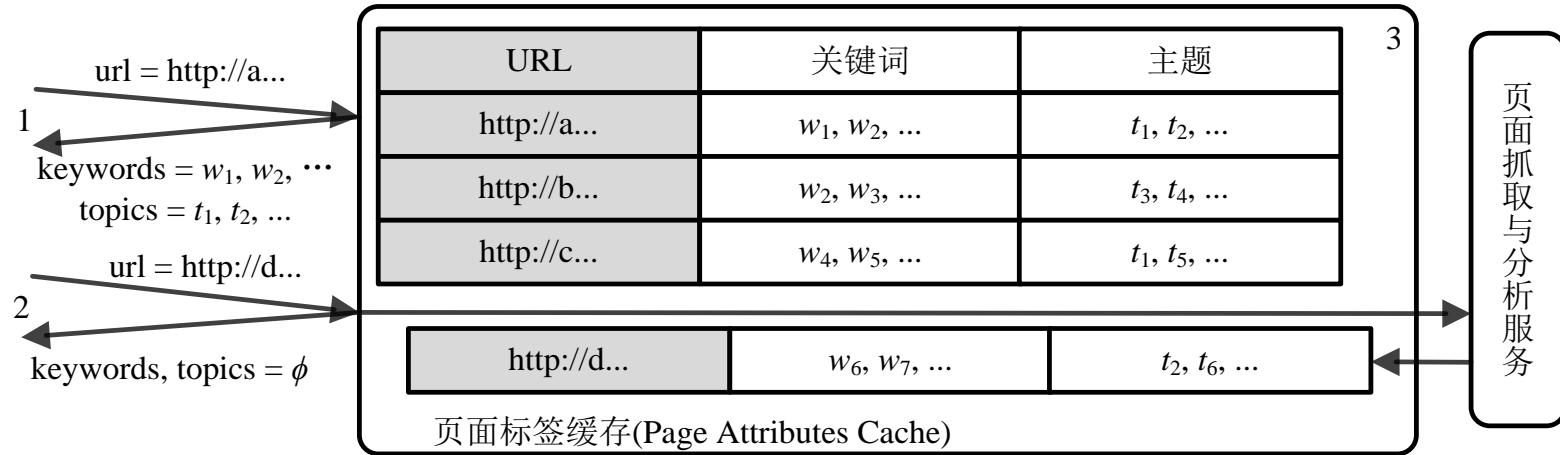
- » 小组婚礼策划 不要错过这个大好机会
- » 成为 婚礼督导师高级培训
- » 新郎
- » 在线
- » 真情
- » 女人

上下文定向主要方式

- 举例
 - 地域定向、频道/URL定向，页面内容定向
- 常用方法
 - 用规则将页面归类到一些频道或主题分类
 - 提取页面中的关键词
 - 提取页面入链锚文本中的关键词
 - 提取页面流量来源中的搜索关键词
 - 用主题模型将页面内容映射到语义空间的一组主题上

半在线(Near-line)抓取系统

- 用在线cache系统存储url -> 特征表以提供实时访问
- 不预先加载任何cache内容，对cache中不存在的url，立刻返回空特征，同时触发相应的页面爬虫和特征提取
- 设置cache系统合适的失效时间以完成特征自动更新



文档的表示与相似度量

- 词袋(Bag of Words, BoW)表示
 - 用关键词TFIDF组成的矢量来表示文档: $\mathbf{d} = (x_1, x_2, \dots, x_M)^\top$
- TF-IDF
 - TF(term frequency): 某文档中该词出现的次数
 - IDF(inverse document frequency): $\text{IDF}(m) = \log(N/\text{DF}(m))$
- 向量空间模型
 - 用余弦距离来衡量两个文档的相似度

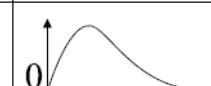
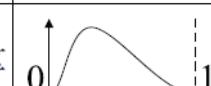
$$\cos(\mathbf{d}_1, \mathbf{d}_2) = \frac{\mathbf{d}_1^\top \mathbf{d}_2}{\|\mathbf{d}_1\| \cdot \|\mathbf{d}_2\|}$$

文本主题模型

- 问题
 - 发现一组文档中抽象的主题(topics)
- 输入
 - 一组文档: $\{d_1, d_2, \dots, d_N\}$
 - 文档 d_n 的BoW表示: $\{x_{n1}, x_{n2}, \dots, x_{nM}\}$
 - 文档矩阵: $\{x_{nm}\}_{N \times M}$
- 输出
 - 各文档得 T 个主题上的强度: $\{z_{n1}, z_{n2}, \dots, z_{nT}\}$

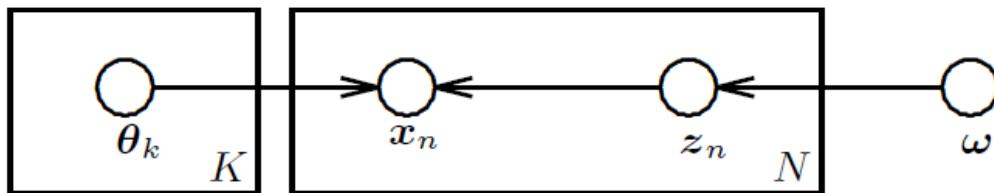
指数族分布（一）

- 归一化形式: $p(x|\theta) = h(x)g(\theta) \exp\{\theta^\top \mathbf{u}(x)\}$
- 若干重要指数族分布:

分布	$\mathbf{u}(x)$	解释	使用场景	示例
Gaussian	$\begin{bmatrix} x \\ x^2 \end{bmatrix}$	给定均值方差时熵最大的分布	一般实变量	
Gamma	$\begin{bmatrix} x \\ \ln x \end{bmatrix}$	给定均值方差, 且 $x > 0$ 时熵最大的分布	非负实变量	
Beta	$\begin{bmatrix} \ln x \\ \ln(1-x) \end{bmatrix}$	给定均值方差, 且 $x \in (0, 1)$ 时熵最大的分布	某区间内的实变量	
multinomial	\mathbf{x}	给定均值方差, 且 $\mathbf{x} \in \{0, 1\}^D$ 时熵最大的分布	离散变量	

指数族分布（二）

- 最大似然估计： $-\nabla \ln g(\theta_{\text{ML}}) = \frac{1}{N} \sum_{i=1}^N \mathbf{u}(x_i)$
- 混合模型： $p(x|\omega, \Theta) = \sum_{k=1}^K w_k h(x) g(\theta_k) \exp\{\theta_k^\top \mathbf{u}(x)\}$



- EM算法：

$$\max_{\omega, \Theta} Q(\omega, \Theta; \omega^{\text{old}}, \Theta^{\text{old}}) = \max_{\omega, \Theta} \sum_z p(z|X, \omega^{\text{old}}, \Theta^{\text{old}}) \ln p(X, \omega, \Theta|z)$$

常用文本主题模型（一）

- LSA(Latent Semantic Analysis)

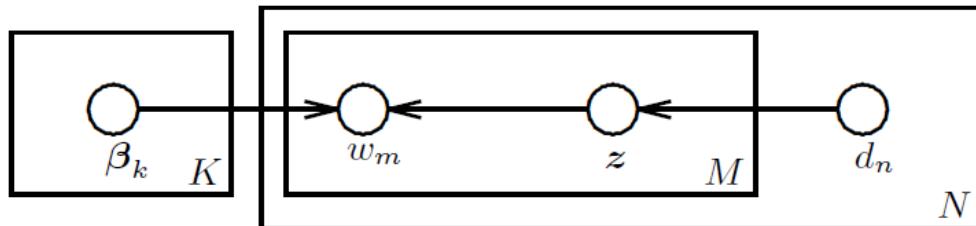
- 文档矩阵的奇异值分解

$$\mathbf{X} = (\alpha_1 \cdots \alpha_K) \cdot \text{diag}(s_1, \dots, s_K) \cdot (\beta_1 \cdots \beta_K)^\top$$

- 选择前 K 个奇异值进行近似:

$$\mathbf{X} \approx (\alpha_1 \cdots \alpha_T) \cdot \text{diag}(s_1, \dots, s_T) \cdot (\beta_1 \cdots \beta_T)^\top$$

- PLSI (Probabilistic Latent Semantic Indexing)



贝叶斯学习

- 贝叶斯公式：

$$\underbrace{P(\theta|X)}_{\text{posterior}} = \frac{\overbrace{p(X|\theta)}^{\text{likelihood}} \overbrace{p(\theta)}^{\text{prior}}}{\underbrace{p(X)}_{\text{evidence}}}$$

- 若干模型估计方法：

	参数估计	预测
ML	$\hat{\theta}_x^{\text{ML}} = \arg \max_{\theta} p(X \theta)$	$p(o X) = p(o \hat{\theta}_x^{\text{ML}})$
Bayes	$p(\theta X) = p(X \theta)p(\theta)$	$p(o X) = \int p(o \theta)p(\theta X)d\theta$
MAP	$\hat{\theta}_x^{\text{MAP}} = \arg \max_{\theta} p(\theta X)$	$p(o X) = p(o \hat{\theta}_x^{\text{MAP}})$

指数族贝叶斯学习

- 共轭先验：
 - 使先验分布与后验分布形式一致的先验分布
- 指数族分布共轭先验：
 - 一般形式： $p(\theta|\eta) = \exp\{\chi^\top \theta - \nu g(\theta) - b(\chi, \nu)\}$
 - $\eta = \{\chi, \nu\}$ 为超参数(hyperparameter)
- 指数族后验部分的超参数：

$$\begin{aligned}\tilde{\chi} &= \chi + \sum\nolimits_{i=1}^N \mathbf{u}(x_i) \\ \tilde{\nu} &= \nu + N\end{aligned}$$

经验贝叶斯 (Empirical Bayes)

- 用如下方法确定超参数 η :

$$\hat{\eta} = \arg \max_{\eta} \int \prod_{i=1}^K p(\mathbf{X}_i | \boldsymbol{\theta}_i) p(\boldsymbol{\theta}_i | \boldsymbol{\eta}) d\boldsymbol{\theta}_i$$

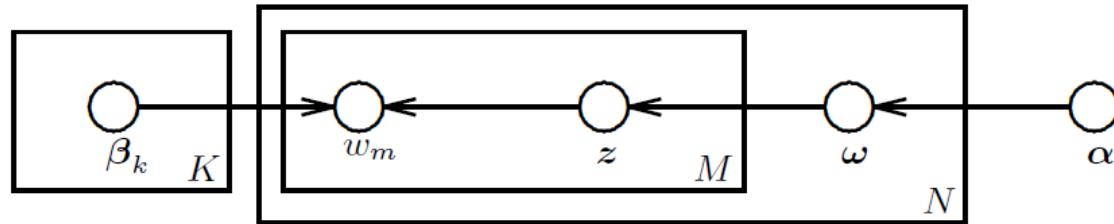
- 当 $p(x|\theta)$ 为指数族分布, $p(\theta|\eta)$ 为其共轭先验时可用EM求解:

$$\text{E-step: } \tilde{\chi}^{\text{old}} = \chi^{\text{old}} + \sum_{i=1}^N \mathbf{u}(x_i), \quad \tilde{\nu}^{\text{old}} = \nu^{\text{old}} + N$$

$$\text{M-step: } \langle \boldsymbol{\theta}, g(\boldsymbol{\theta}) \rangle_{p(\boldsymbol{\theta} | \boldsymbol{\eta}^{\text{new}})} = \frac{1}{K} \sum_{k=1}^K \langle \boldsymbol{\theta}, g(\boldsymbol{\theta}) \rangle_{p(\boldsymbol{\theta} | \tilde{\boldsymbol{\eta}}_k^{\text{old}})}$$

常用文本主题模型（二）

- LDA (Latent Dirichlet Allocation)



- PLSI的贝叶斯版本
- 优化目标函数

$$p(w|\alpha, \beta) = \int p(\omega|\alpha) \left(\prod_{m=1}^M \sum_z p(z|\omega) p(w_m|z, \beta) \right) d\omega$$

- 可用变分法或Gibbs sampling方法求解

行为定向(Behaviorial targeting)

- 根据用户历史上网记录和其他数据计算出用户兴趣，根据此投放广告



360 教育



The screenshot shows a search results page from Sogou. At the top, there's a navigation bar with links for '我的搜狐' (My Sohu), '用户名/邮箱/手机号' (Username/Email/Phone Number), '登录' (Login), '注册' (Register), '邮件' (Email), '相册' (Album), '说两句' (Say Something), and several news headlines. Below the navigation is a large advertisement for learning CNN, featuring a woman and the text '坚持1个月 每天只要 5分钟, 坚持30天, 听懂CNN 效果真的不一样! 立即行动'. The main search results are categorized into '汽车' (Cars), '降价' (Discount), '二手车' (Used Cars), '房产' (Real Estate), '家居' (Home Appliances), and '二手房' (Second-hand Real Estate). The '汽车' section lists various car models and their prices. The '房产' section lists real estate deals, including one for a 60 square meter apartment and another for a villa. The '家居' section lists items like a sofa set. The '二手房' section lists properties for sale, such as a 100 square meter house.

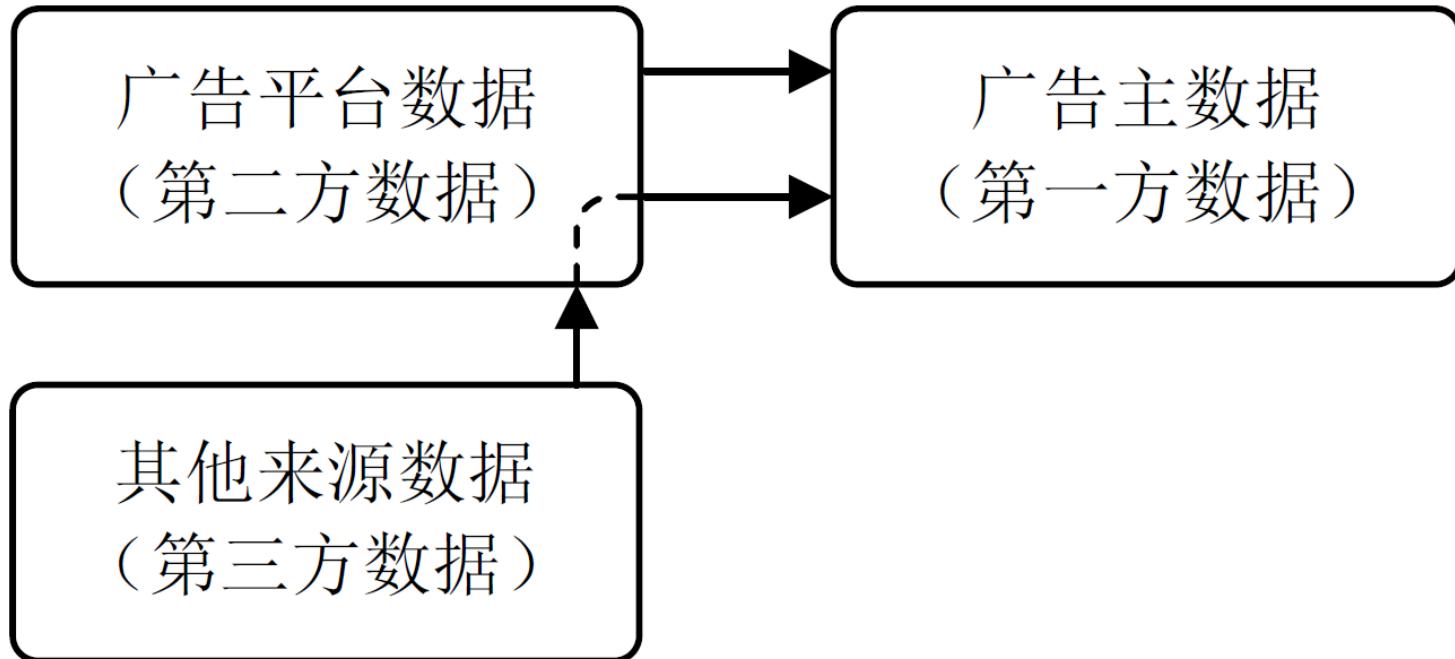
行为定向数据来源

- 决策行为
 - 转化(Conversion)、预转化(Pre-conversion)
 - 对应着非常明确的用户兴趣，价值最高
- 主动行为
 - 搜索(Search)、广告点击(Ad click)、搜索点击(Search click)
 - 在明确意图支配下主动产生的行为，价值也很高
- 半主动行为
 - 分享(Share)、网页浏览(Page View)
 - 量最大，用户意图较弱，但也有一定价值

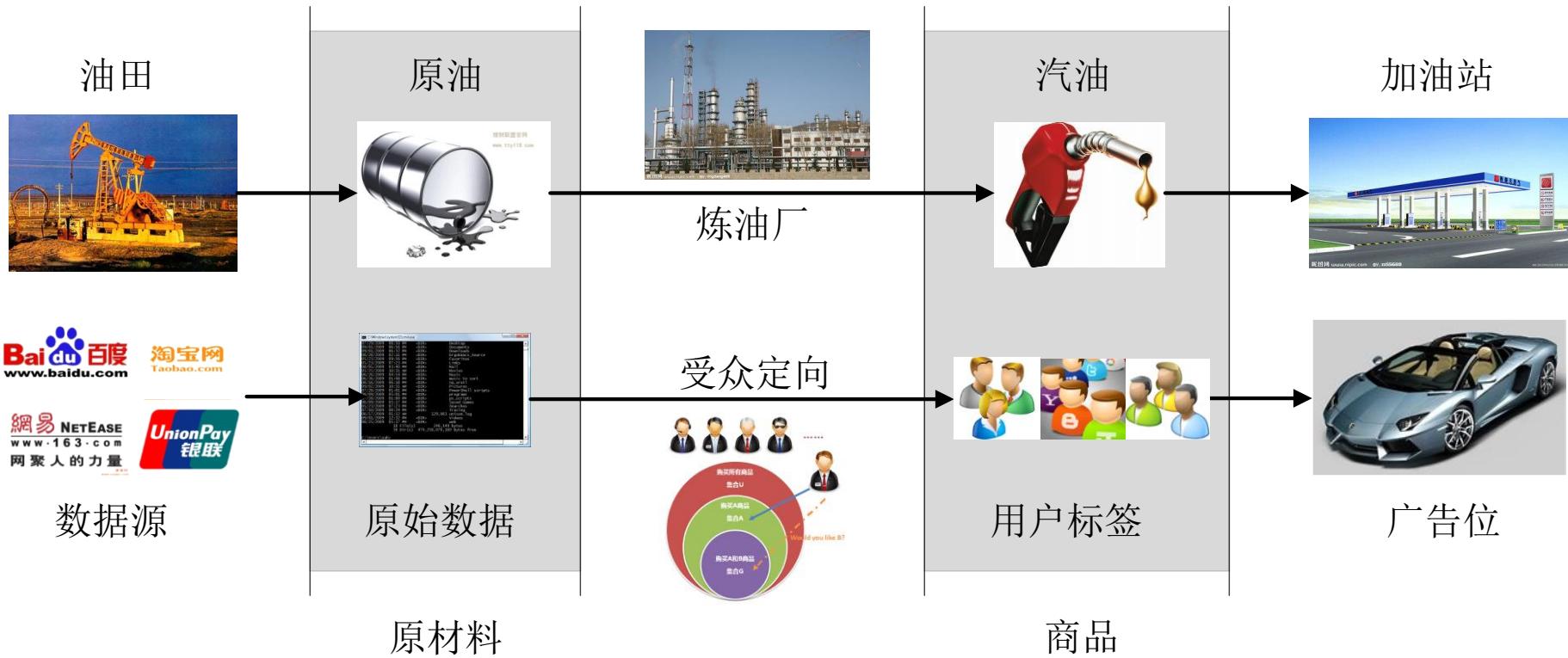
行为定向数据来源

- 被动行为
 - 广告浏览(Ad View)
 - 负面的加权因素
- 用户ID
 - 最重要的数据，一串0前面的那个1
 - 稳定、精确的用户ID能大幅提高行为数据使用效率
- 社交关系
 - 可以用与用户兴趣的平滑：当某个人的行为不足，无法进行精准的行为定向时，可以考虑借鉴其社交网络朋友的行为和兴趣。

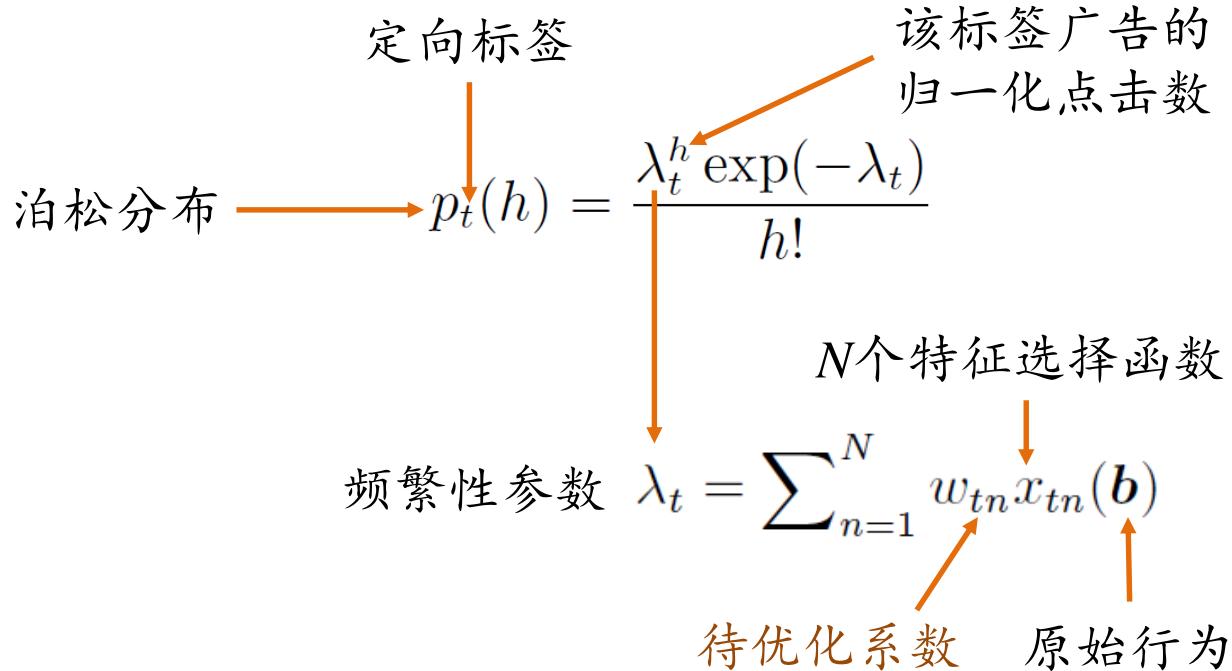
三方数据的概念



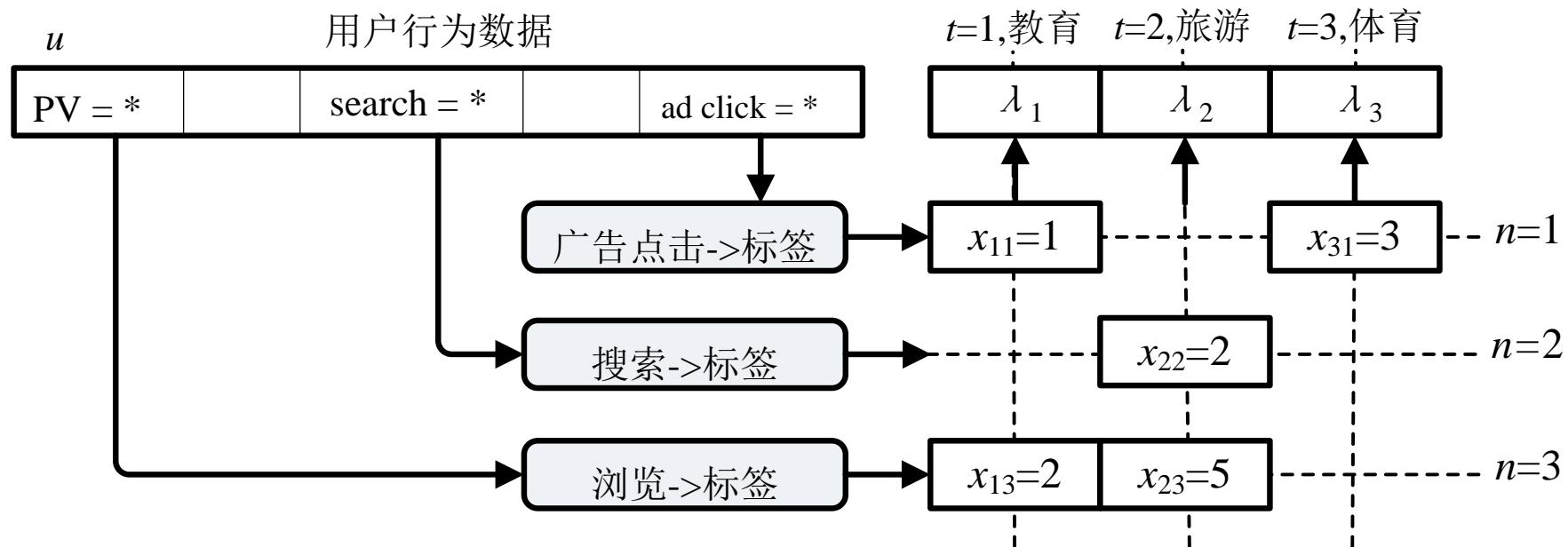
数据的价值与地位



行为定向建模



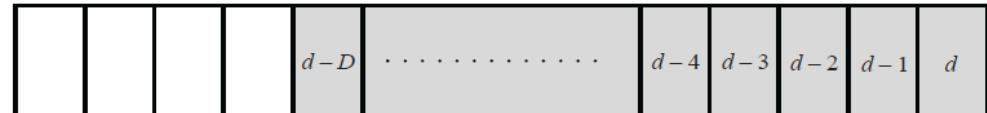
行为定向特征选择过程



行为定向数据组织

- Session log
 - 将各种行为日志整理成以用户ID为key的形式，作为各数据处理模块的输入源，可以将targeting变成局部计算
- 行为定向两种长期特征累积方式
 - 滑动窗方式

$$\tilde{x}(d) = \sum_{\delta=0}^D x(d - \delta)$$

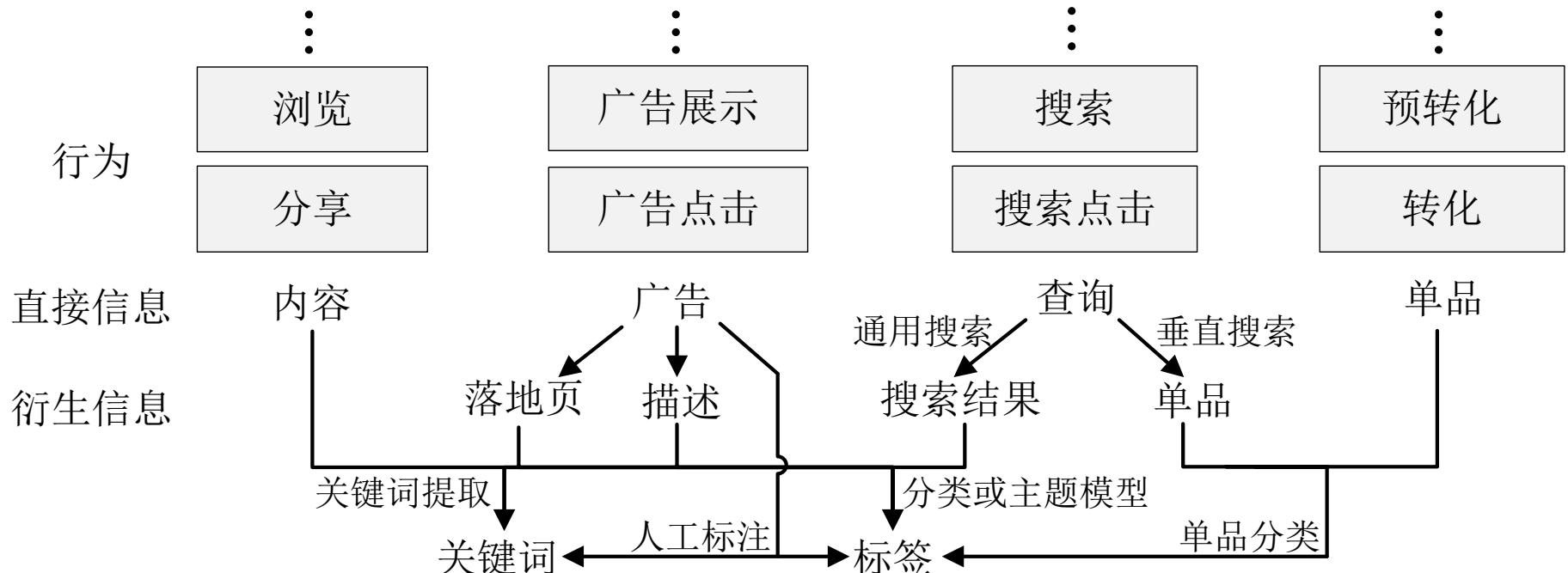


- 时间衰减方式

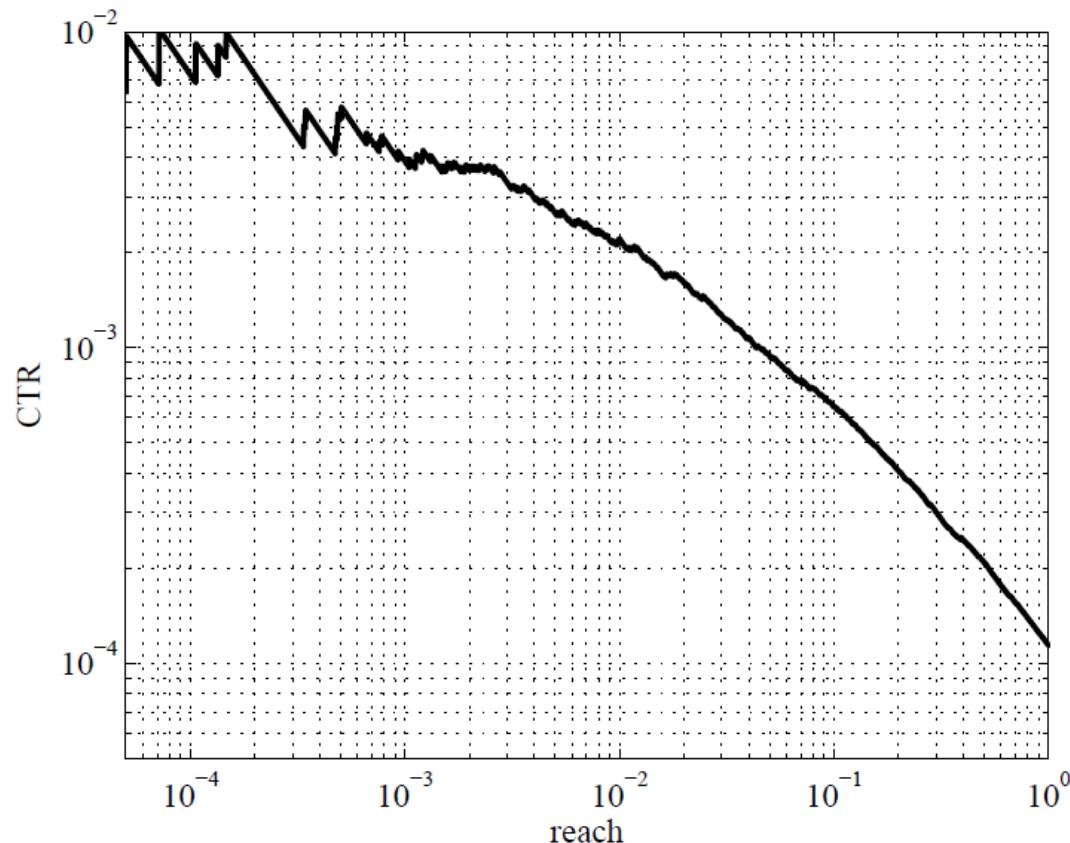
$$\tilde{x}(d) = \alpha \tilde{x}(d-1) + x(d)$$



各类行为的标签化方法



受众定向评测 - Reach/CTR 曲线



数据交易所(Data Exchange)

- 目的：
 - 聚合多种原始行为数据，加工成统一标签后，在公开市场上售卖
 - 聚合多种加工后标签数据，在公开市场上售卖
- 主要特征：
 - 主要负责第三方数据的收集、加工和流转，不一定直接从事广告交易
 - 多与广告交易平台、SSP等合作进行数据变现

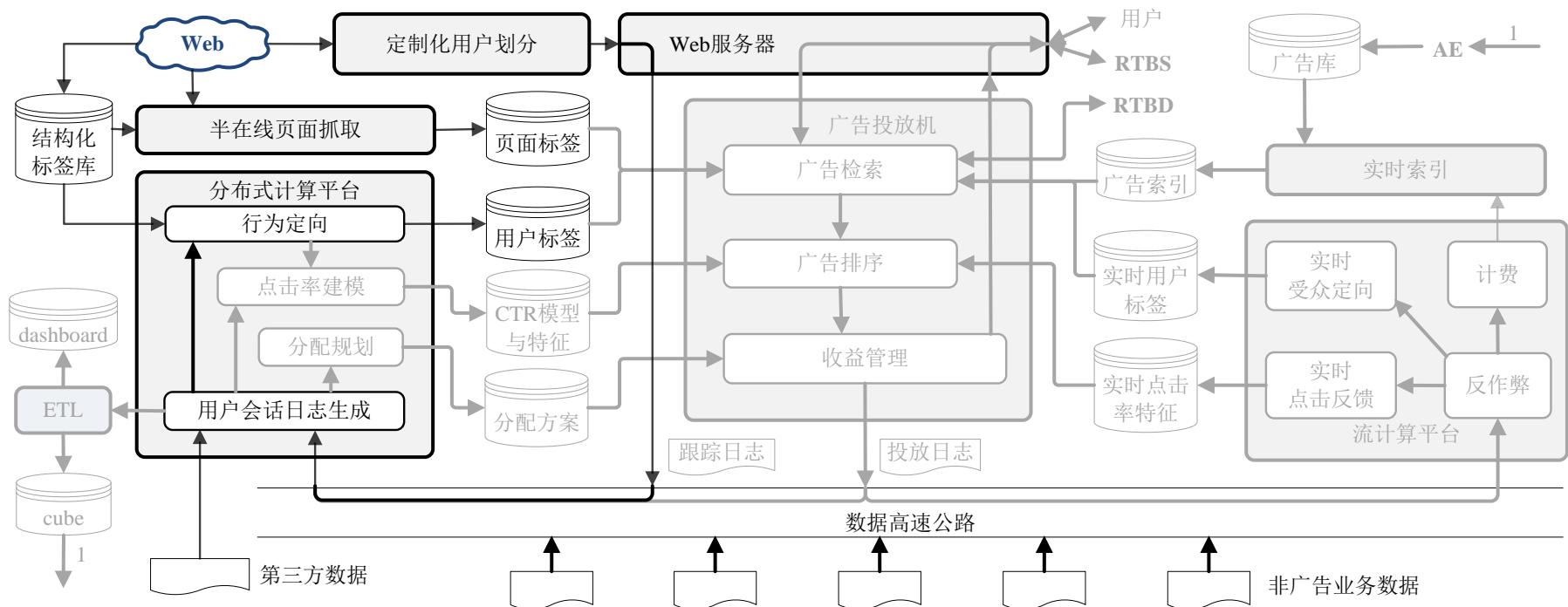
数据管理平台(Data Management Platform)

- 目的：
 - 为网站提供第一方数据加工和应用能力，
 - 结合公开市场第三方数据，加工跨媒体用户标签，支持网站业务运营和广告投放。
- 主要特征：
 - 第一方用户定制化划分能力
 - 统一的对外数据接口



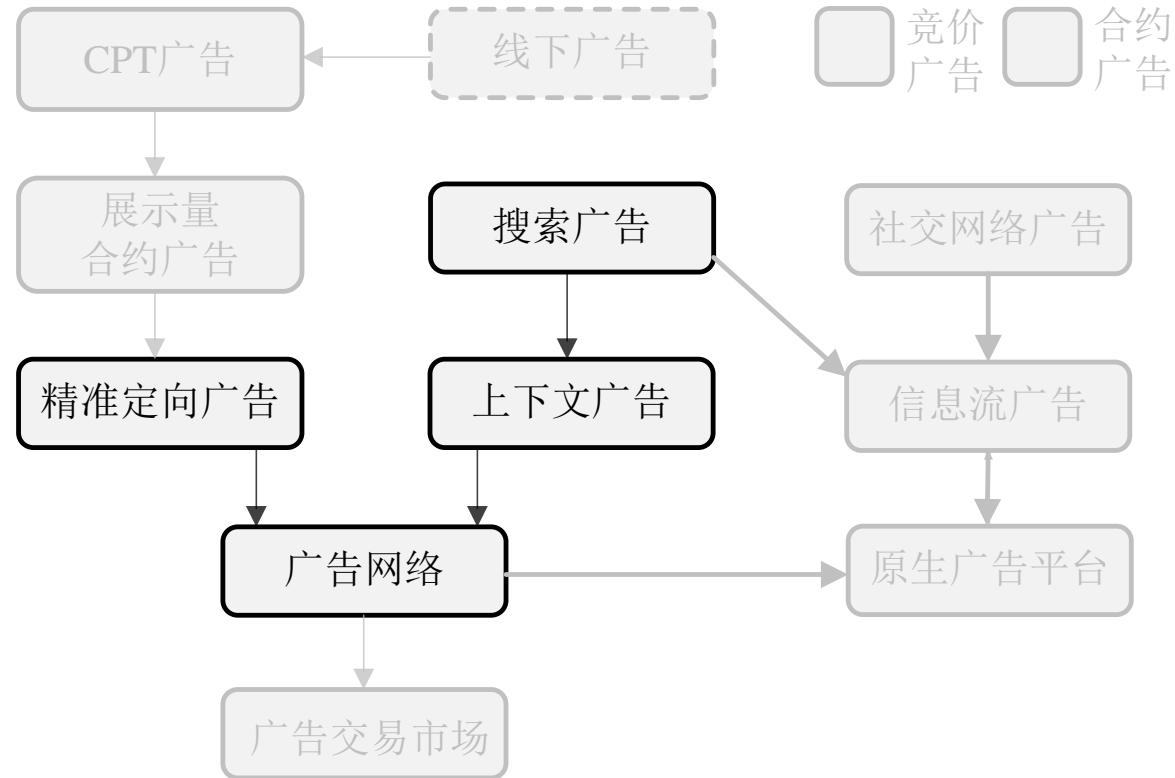
- 核心业务：
 - 主要提供面向publisher的数据加工服务
 - 直接运营ad network进行数据变现
- 其他点评：
 - 较早提出受众定向(audience targeting)的概念
 - 数据标签不像bluekai那样在市场上公开出售，
 - 使用标签创造的营收按照一定比例跟publisher分成

数据管理平台系统架构



搜索与竞价广告

搜索与竞价广告



搜索广告示例

- 分为北(North)、东(East)、南(South)三个区域
- 广告一般带有底色
- 位置次序为：
 - $\text{North}_1, \text{North}_2, \dots, \text{East}_1, \text{East}_2, \text{East}_3, \dots$
- 南区或重复北区、或重复东区

信用卡

百度一下

信用卡首选招商银行信用卡 积分可兑星巴克,1分钟办卡!

招商银行信用卡中心·官方24小时快速办理信用卡通道 信用卡积分可兑星巴克/DQI
积分永久有效 · 办卡送1500积分 · 增品换取背包 · 官方申请通道
ccclub.cmbchina.com

▲ 平安信用卡 在线办卡送100 方便快捷! www.pingan.com

平安信用卡 办卡成功立送100元,超长50天免息期,积分还可换油卡电话卡.

信用卡怎么申请_信用卡查看_信用卡官方申请_百度财富

北区(North)

发卡行： 不限 | 平安 | 兴业 | 民生 | 建设 | 招行 | 中信 | 交通 |
中行 | 农行 | 花旗 | 华夏 | 光大 | 清发 | 工商 | 广发



招商银行现金随地可取
100%的助効额度
新开送100购享3折
新开送100购享3折

[查看更多相关信用卡>](#)

caifu.baidu.com

信用卡_百度百科

信用卡（英语：Credit Card），又叫贷记卡。是一种非现金交易付款的方式，是简单的信贷服务。**信用卡**一般是长85.60毫米、宽53.98毫米、厚1毫米的具有消费信用的特制载体塑料卡片。是银行向个人和单位发行的，凭证向特约单位购物、消费和向银行存取现金，其形式是一张正面印有发卡银行名称、有效期、号码...
[产品起源](#) [产品种类](#) [主要特点](#) [持点对比](#) [申请流程](#) [更多>>](#)

baike.baidu.com/2014-10-15

信用卡论坛-我爱卡会员社区-中国最大最权威的信用卡论坛

信用卡论坛-我爱卡会员社区-中国最大的信用卡论坛,讲述申卡历程,谈谈权用卡经验,信用卡生活-从我爱卡开始
bbs.51credit.com/ 2014-10-12 - 百度直照 - 85%好评

信用卡首选招商银行信用卡 积分可兑星巴克,1分钟办卡! ccclub.cmbchina.com

推广链接
招商银行信用卡中心·官方24小时快速办理信用卡通道 信用卡积分可兑星巴克/DQI

▲ 平安信用卡 在线办卡送100 方便快捷! www.pingan.com

平安信用卡 办卡成功立送100元,超长50天免息期,积分还可换油卡电话卡.

推广链接

信用卡哪家比较好 上班族税
宜人贷o2o借贷平台 针对上班族,打卡工资4000元即可申请,0抵押0担保,额度高达50万.
cps.yirendai.com

信用卡申请平安 轻松借钱不用
信用卡首选平安普通贷款,活动时间有限,利率8折任你抢10抵押(担保)风险!
1010-0000.pingan.com

中信银行信用卡 中心小额度贷款
中信信用卡 中0抵押贷款,月入3000即可办理,贷款流程简单,放款快,专人上门服务.
www.ecitic.com

信用卡申请 贷款平台融360
融360贷款平台,超多借贷机构超多信贷员,信用卡申请
www.rong360.com

无需信用卡 1小时快速借贷!!
信用卡 押拍货新型个人网络借款方式,无抵押贷款,
www.ppdai.com

更多信用卡的百度推广结果>
关于**信用卡**百度为您推荐更多优质结果,放心搜索,有保障。
▶ 信用卡 **▶ 小额信用卡办理**
▶ 什么信用卡最好申请 **▶ 信用卡小额贷款**

▶ 想在此推广你的产品吗?
咨询热线: 400-800-8888
e.baidu.com

东区(East)

搜索广告产品新形式

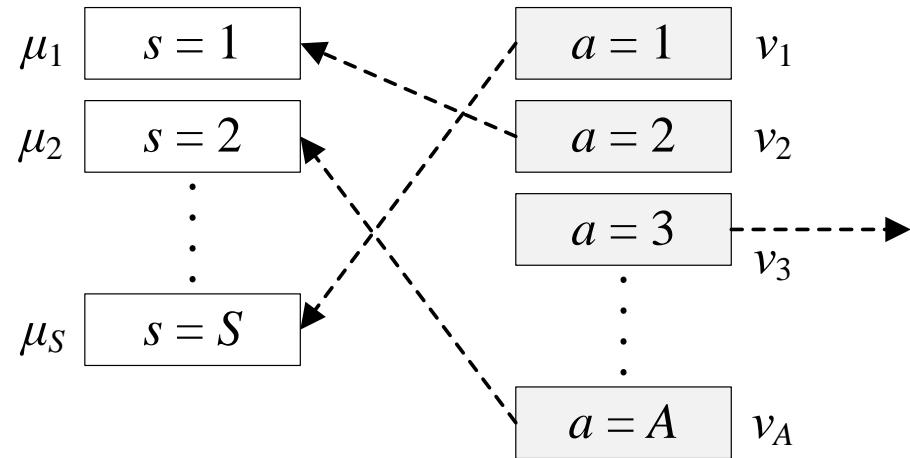
①、②：
原生化探索

③、④：
弱相关广告形式

The screenshot shows the Ctrip travel website homepage. At the top, there is a search bar with the placeholder "携程" (Ctrip) and a green "搜索一下" (Search) button. Below the search bar, there is a promotional banner for Ctrip's travel services. To the right of the banner, there is a link to "我的定制词" (My Customized Words). In the center, there is a large search form for flights, hotels, and trains. The flight section includes tabs for "机票" (Flight), "国内机票" (Domestic Flight), and "国际机票" (International Flight). It has input fields for "出发城市" (Departure City) and "到达城市" (Arrival City), and a date selector for "出发时间" (Departure Date) set to "2014-09-08". Below the search form, there are sections for "城市酒店预订" (City Hotel Booking), "品牌酒店预订" (Brand Hotel Booking), "1元门票预订" (1 Yuan Ticket Booking), "特价机票预订" (Special Price Flight Booking), and "特别优惠" (Special Offer). There are also lists of recommended hotels and travel packages. At the bottom left, there is a promotional banner for "携程订酒店机票上携程拿返现" (Get cashback when you book hotels and flights on Ctrip) with a URL "http://jiudian.ctrip.com". On the right side, there is a sidebar with links to other travel websites like LY.com, 马蜂窝 (Mafengwo), 穷游网 (Qiongyou.com), 驴妈妈 (Lumama), and Meituan.

位置拍卖(Position auctions)问题

- 将对象 $a = \{1, 2, \dots, A\}$ 排放到位置 $s = \{1, 2, \dots, S\}$
- 对象 a 的出价(bid)为 b_a , 而其对位置 s 的计价为 $r_{as} = \mu_s v_a$, ($\mu_1 > \mu_2 > \dots > \mu_S$)
- 将 v_a 视为点击价值, μ_s 视为点击率, 该模型可近似描述广告系统竞价问题(对显示广告 $S = 1$)



位置拍卖市场

- 位置拍卖(Position auctions)
 - 将对象 $a=\{1, 2, \dots, A\}$ 排放到位置 $s=\{1, 2, \dots, S\}$
 - 对象 a 的出价(bid)为 b_a , 而其对位置 s 的计价为 $u_{as} = v_a x_s$, ($x_1 > x_2 > \dots > x_S$)
 - 将 v_a 视为点击价值, x_s 视为点击率, 该模型可近似描述广告系统竞价问题(对显示广告, $S=1$)
- 对称纳什均衡(Symmetric Nash equilibrium)
 - $(v_s - p_s) x_s \geq (v_t - p_t) x_t$, 其中 $p_t = b_s + 1$
 - 寻找收入最大化且稳定的纳什均衡状态是竞价系统设计的关键

机制设计-定价

- 广义第二高价(Generalized second pricing)机制
 - 与VCG机制相比，会收取广告主更多的费用
 - 整体市场不是truth-telling的
 - 简单易行，为在线广告系统广泛采用
 - CPM情形下： $q_s = b_{s+1} + 1$
 - CPC情形下： $q_s = r_{s+1} / \mu_s + 1 = \mu_{s+1} b_{s+1} / \mu_s + 1$
- VCG(Vickrey–Clarke–Groves)机制
 - 某对象的收费等于给他人带来的价值损害：
 - 整体市场是truth-telling的

机制设计其他方法

- 市场保留价(Market Reserve Price, MRP)
 - 即“底价”，出价高于MRP才能参与竞价
 - 计费如果小于底价，按底价计费
 - 可以根据不同竞价标的设置不同的底价
- 价格挤压
 - $r = \mu \cdot v \rightarrow r = \mu^\kappa \cdot v$
 - κ 为正数：当其充分大时，只根据点击率排序；当其接近0时，只根据点击单价排序

整体计价过程

```
int auction(vector<int> & cands,
            vector<float> & ctrs, vector<float> & bids,
            float MRP, float squash, int slotNum,
            vector<int> & results, vector<float> & prices) {
    int candNum = cands.size();

    // 按照给定的squashing因子调整预估CTR
    for (int c = 0; c < candNum; c++)
        ctrs[c] = ctrs[c] ^ squash;

    // 计算调整后的eCPM
    vector<float> eCPMs;
    eCPMs.resize(candNum, 1e-10f);
    for (int c = 0; c < candNum; c++)
        if (bids[c] >= MRP) // 跳过那些出价小于市场保留价的候选
            eCPMs[c] = ctrs[c] * bids[c];
}
```

```
// 将所有候选按照eCPM排序
for (int c1 = 0; c1 < candNum; c1++)
    for (int c2 = c1 + 1; c2 < candNum; c2++)
        if (eCPMs[c1] < eCPMs[c2]) {
            SWAP(cands[c1], cands[c2]);
            SWAP(eCPMs[c1], eCPMs[c2]);
            SWAP(ctrs[c1], ctrs[c2]);
        }

// 得到各竞价结果并计算定价
results.clear();
prices.clear();
for (int c = 0; c < candNum - 1; r++) {
    if (eCPMs[c] <= 1e-10f)
        break;

    // 按照GSP计算定价
    float price = eCPMs[c + 1] / ctrs[c];
    if (price < MRP) price = MRP;

    results.push_back(cands[c]);
    prices.push_back(price);
}

}
```

GSP广告竞价示例

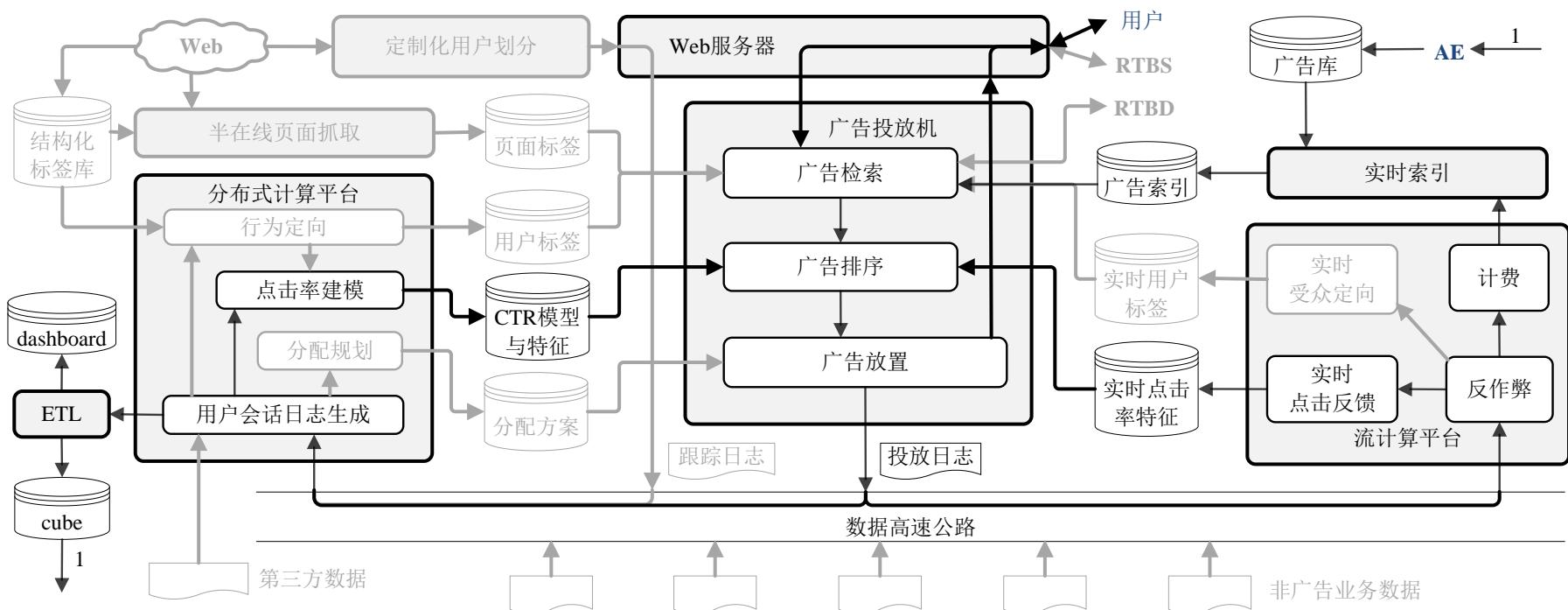
(货币单位: 元, MRP=0.25)

出价	点击率(%)	排序/定价($\kappa = 1.0$)	排序/定价($\kappa = 2.5$)	排序/定价($\kappa = 0.5$)
0.80	1.6	1/0.50	1/0.25	2/0.25
2.00	0.4	2/0.75	3/0.70	1/1.60
0.30	1.0	3/0.25	2/0.25	3/0.28
0.40	0.5	4/0.25	4/0.25	4/0.25

搜索广告计算概貌

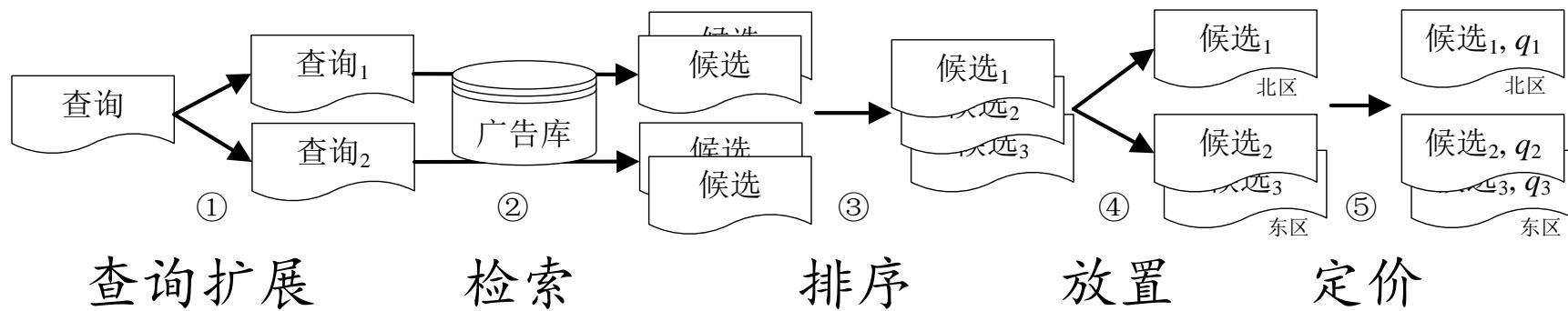
- 优化目标: $\max_{a_1, \dots, T} \sum_{i=1}^T \{\mu(a_i, c_i) \cdot \text{bid}_{\text{CPC}}(a_i)\}$
- 关键特点:
 - 搜索广告的变现能力即eCPM远高于一般显示广告
 - 搜索广告的受众定向标签，即是上下文的搜索查询
 - 搜索广告样式与自然结果的展示形式非常接近，有原生广告的意味
 - 从搜索广告发展起来的竞价交易模式，已经逐渐发展成为互联网广告最主流的交易模式

搜索广告系统架构



搜索广告产品策略

- 搜索广告决策过程：



- 关键产品策略：
 - 查询扩展、广告排序、广告放置

查询扩展主要模式

- 精确匹配
 - 英语培训 → {英语培训, 培训英语}
- 短语匹配
 - 英语培训 → {英语培训暑期班, 哪个英语培训机构好, 英语的培训, 英语相关培训, 培训英语, 英文培训, ...} 等
- 广泛匹配
 - 英语培训 → {外语学习班, 四级证书, ...}
- 否定匹配

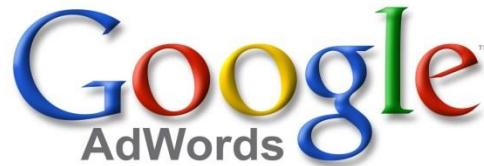
查询扩展主要方法

- 基于推荐的方法
 - 在 $\{\text{session}, \text{query}\}$ 矩阵上通过推荐技术来产生相关的关键词
 - 利用的是搜索的日志数据
- 基于主题模型的方法
 - 根据文档主题模型，对某个查询拓展出主题相似的其他查询
 - 利用的是一般的文档数据
- 基于历史效果的方法
 - 记录某些对特定广告主eCPM较高的关键词
 - 利用广告本身的历史eCPM数据

广告放置

- 广告放置(Ad Placement)问题
 - 广告候选完成排序以后，需要分别确定北区和东区的广告条数
- 用户体验指标
 - 北区广告的平均条数，即North Foot Print(NFP)或Average Serving Number(ASN)
- 广告放置策略优化

$$\begin{aligned} & \max \sum_{i=1}^T \left\{ \sum_{s=1}^{n_i} r(a_{is}, c_i, N_s) + \sum_{s=n_i+1}^{n_i+e_i} r(a_{is}, c_i, E_{s-n_i}) \right\} \\ & \sum_{i=1}^T n_i \leq TC \end{aligned}$$



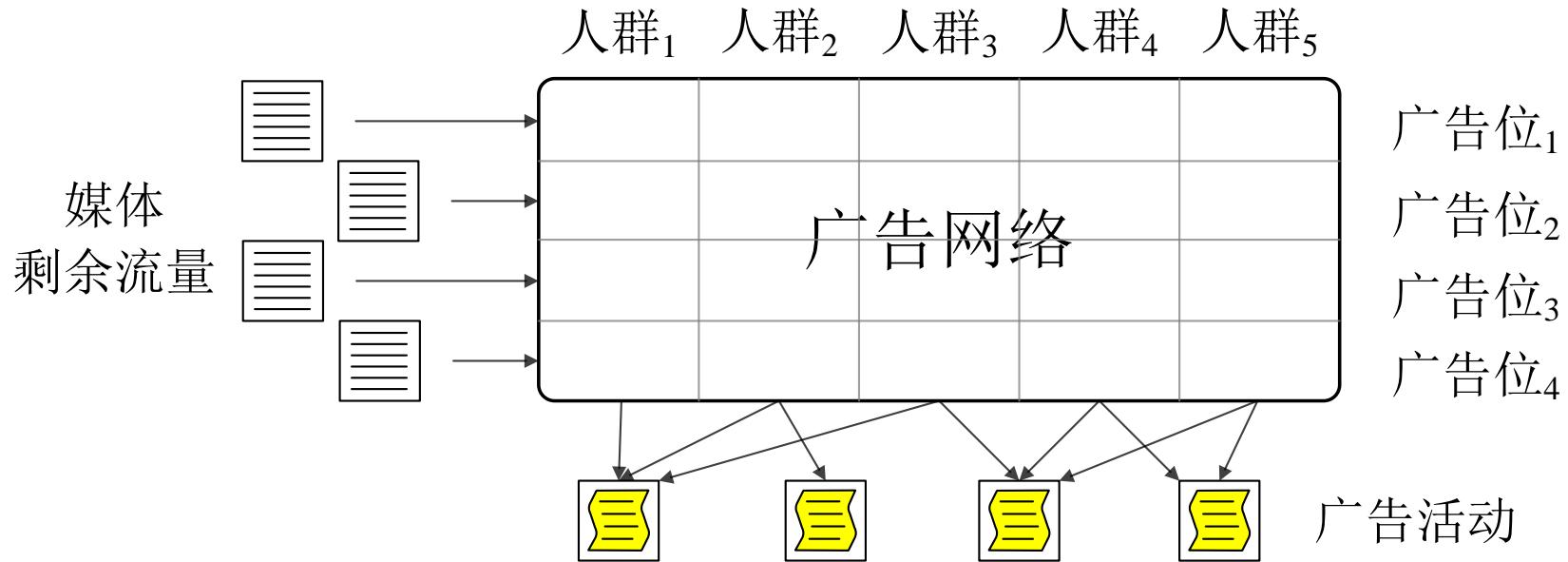
- 产品概况：
 - Google推出的关键词竞价广告产品
- 其他点评：
 - 最早按CPM售卖，销售不理想
 - 后来采用了Overture的CPC售卖，并创造性地引入点击率概念表达广告相关性，取得了非常高的营收
 - 2013年营收374亿美元



- 产品概况：
 - 淘宝专门服务于卖家的广告产品，在搜索结果等页面上展示竞价广告
- 其他点评：
 - 按CPC售卖，同样是按照eCPM排名
 - 广告主集中在电商行业，同时用户意图比通用搜索引擎更强烈
 - 与广告主是共生关系



竞价广告网络



聚合媒体剩余流量，以人群为标的物进行竞价售卖

广告网络广告示例

NEW

双核租用+高防护,仅需459元/月

腾佑科技
做质量拼服务

服务器租用/托管 - 认准 **腾佑科技** 百人团队为您服务!
全线服务器、满②年送①年、满③年送②年

当前位置：华彩软件站 → 文章资讯 → 软件相关 → 怎样截图整个网页？360浏览器给整个网页截图方法

超值热卖 HOT

- 日本牛人专线 ￥5878起
- 东京5日自由行 ￥2999起
- 冲绳4日自由行 ￥3368起

 遛牛旅游网 Tuniu.com

关闭

**① 土耳其最佳旅游
② 日本旅游**

怎样截图整个网页？360浏览器给整个网页截图方法

分享到： 百度搜索 新浪微博 QQ空间 腾讯微博 人人网 腾讯朋友 QQ收藏 百度贴吧 0

作者：佚名 来源：华彩软件站 发布时间：2014-03-10 13:30:24 | [减小字体](#) [增大字体](#)

1000万悬赏寻找民间操盘达人

 「EF英孚」寒暑假海外游学
出国学外语,开眼界,树自信,投资未来. 1965年成立于瑞典,49年留学游...

 贵金属模拟大赛赢大奖
月月真金白银大奖,边学边练习投资! 正规天津贵金属交易所会员,信誉...

360浏览器给整个网页截图方法：
首先我们要打开你的360浏览器，没有的自己去下载安装吧，下载地
址<http://www.huacolor.com/soft/77423.html>然在工具栏有一个应用盒子，打开它的下拉列表，看到有一个【获取更多应用】的选项，点击它

超值热卖 HOT

- 土耳其经典10日 ￥9839起
- 土耳其深度12日 ￥14499起
- 土耳其全景10日 ￥11299起

 遛牛旅游网 Tuniu.com

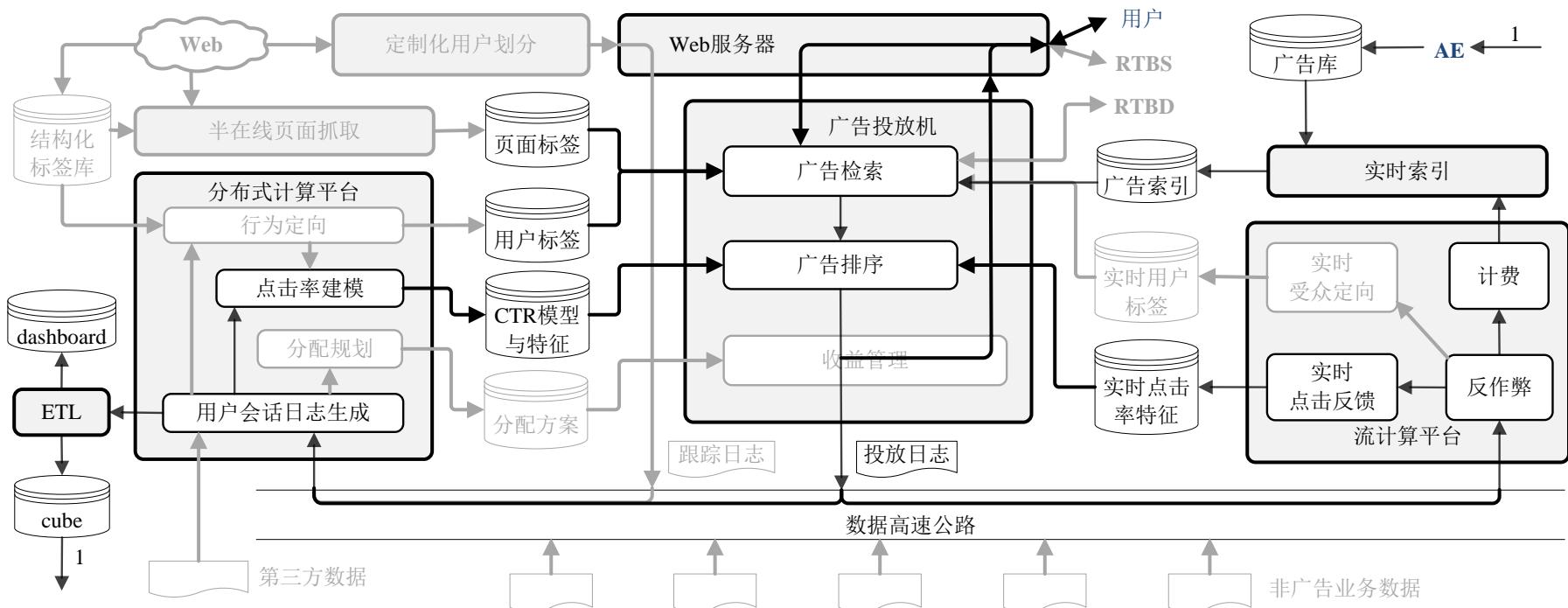
关闭

**① 日本旅游
② 土耳其最佳旅游**

广告网络计算概貌

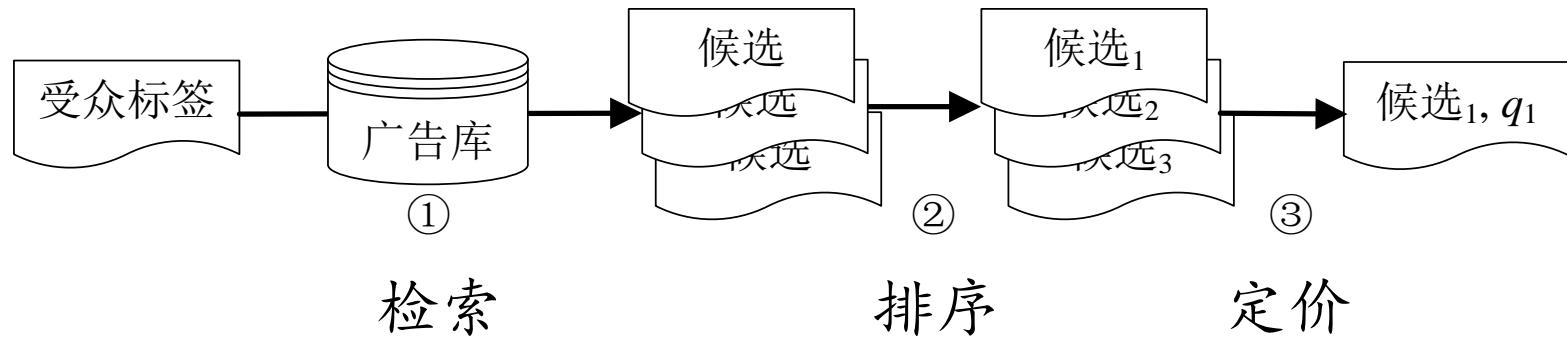
- 优化目标: $\max_{a_1, \dots, T} \sum_{i=1}^T \{\mu(a_i, u_i, c_i) \cdot \text{bid}_{\text{CPC}}(a_i)\}$
- 关键特点:
 - 竞价方式不向广告主做量的约定，而是根据变现能力，即eCPM，来决定每次展示分配给哪个广告主。
 - 按人群售卖，淡化媒体和广告位的概念
 - 无需再满足广告主品牌独占的要求
 - 采用分成结算，运营方的现金流状况大为改善

广告网络系统架构



广告网络产品策略

- 广告网络展示决策过程：



- 关键产品策略：
 - 广告检索、广告排序

倒排索引

- 文档集
 - D_1 = “谷歌地图之父跳槽Facebook”
 - D_2 = “谷歌地图创始人拉斯离开谷歌加盟Facebook”
 - D_3 = “谷歌地图创始人跳槽Facebook与Wave项目取消有关”
 - D_4 = “谷歌地图创始人拉斯加盟社交网站Facebook”
- 关键词(term)：
 - {谷歌, 地图, 之父, 跳槽, Facebook, ...}
- 倒排链：
 - 谷歌 → $\{D_1, D_2, D_3, D_4\}$, 地图 → $\{D_1, D_2, D_3, D_4\}$, 之父 → $\{D_1, D_3, D_4\}$, 跳槽 → $\{D_1, D_3\}$, Facebook → $\{D_1, D_2, D_3, D_4\}$, ...

布尔表达式检索

- 广告投放条件的过滤可以视为布尔表达式检索问题
- 布尔表达式检索的一些概念
 - Doc DNF: $(age \in \{3\} \wedge state \in \{NY\}) \vee (state \in \{CA\} \wedge gender \notin \{M\})$
 - Conjunction: $age \in \{3\} \wedge state \in \{NY\}, state \in \{CA\} \wedge gender \notin \{M\}$. 同一属性在某 Conjunction 里只出现一次
 - Assignment: $age \in \{3\}, state \in \{NY\}, state \in \{CA\}, \dots$
 - sizeof[Conjunction]: conjunction 包含非 \wedge 的 Assignment 个数
- 基本思想
 - 某查询满足conjunction，也就满足包含此conjunction的doc.
 - 维护两层倒排关系: Conjunction -> DocId, Assignment -> ConjunctionId
 - 如果 $sizeof(Conjunction)$ 大于 $sizeof(query)$ ，则无需考虑

布尔表达式检索 – index 算法

- 建立第一层 Index
 - 遍历文档DNF的Conjunction, 如果为新的, 则分配一个新ID(从0递增), 否则用之前分配的Conjunction ID; 文档分配Doc ID (从0递增); 写入conjunction到doc的倒排关系, 形成第一层Index
- 对于上步出现的新Conjunction, 建立第二层 Index:
 - 将Conjunction切成Assignment流, Term为(属性, 值), 例: $age \in \{3;4\}$ 切成两个Term: (age, 3), (age, 4); state $\notin \{CA;NY\}$ 也切成两个Term: (state, CA), (state, NY), \in 和 \notin 体现在倒排链表上
 - 计算Conjunction的size, 将size体现在Term中, 最终的Term的组成是 (sizeof[Conjunction], 属性, 值)
 - 对于size为0的Conjunction, 添加一个特殊的Term: Z, \in
 - 写入倒排关系, Term -> (ConjunctionID, $\in | \notin$)+

布尔表达式检索 – index示例

广告集合：

$$a_1 = (\text{age} \in \{3\} \cap \text{geo} \in \{\text{北京}\}) \cup (\text{geo} \in \{\text{广东}\} \cap \text{gender} \in \{\text{男}\})$$

$$a_2 = (\text{age} \in \{3\} \cap \text{gender} \in \{\text{女}\}) \cup (\text{geo} \notin \{\text{北京}, \text{广东}\})$$

$$a_3 = (\text{age} \in \{3\} \cap \text{gender} \in \{\text{男}\} \cap \text{geo} \notin \{\text{广东}\}) \cup (\text{state} \in \{\text{广东}\} \cap \text{gender} \in \{\text{女}\})$$

$$a_4 = (\text{age} \in \{3, 4\}) \cup (\text{geo} \in \{\text{广东}\} \cap \text{gender} \in \{\text{男}\})$$

$$a_5 = (\text{state} \notin \{\text{北京}, \text{广东}\}) \cup (\text{age} \in \{3, 4\})$$

$$a_6 = (\text{state} \notin \{\text{北京}, \text{广东}\}) \cup (\text{age} \in \{3\} \cap \text{state} \in \{\text{北京}\}) \cup (\text{state} \in \{\text{广东}\} \cap \text{gender} \in \{\text{男}\})$$

$$a_7 = (\text{age} \in \{3\} \cap \text{state} \in \{\text{北京}\}) \cup (\text{state} \in \{\text{广东}\} \cap \text{gender} \in \{\text{女}\})$$

Conjunctions:

$$a_1 = j_1 \cup j_4, a_2 = j_2 \cup j_6, a_3 = j_3 \cup j_7, a_4 = j_5 \cup j_4, \quad j_1 \rightarrow \{a_1, a_6, a_7\}, j_2 \rightarrow \{a_2\}, j_3 \rightarrow \{a_3\}, j_4 \rightarrow \{a_1, a_4, a_7\}, \\ a_5 = j_6 \cup j_5, a_6 = j_6 \cup j_1 \cup j_7, a_7 = j_1 \cup j_7 \qquad \qquad \qquad j_5 \rightarrow \{a_4, a_5\}, j_6 \rightarrow \{a_2, a_5, a_7\}, j_7 \rightarrow \{a_3, a_6\}$$

倒排索引：

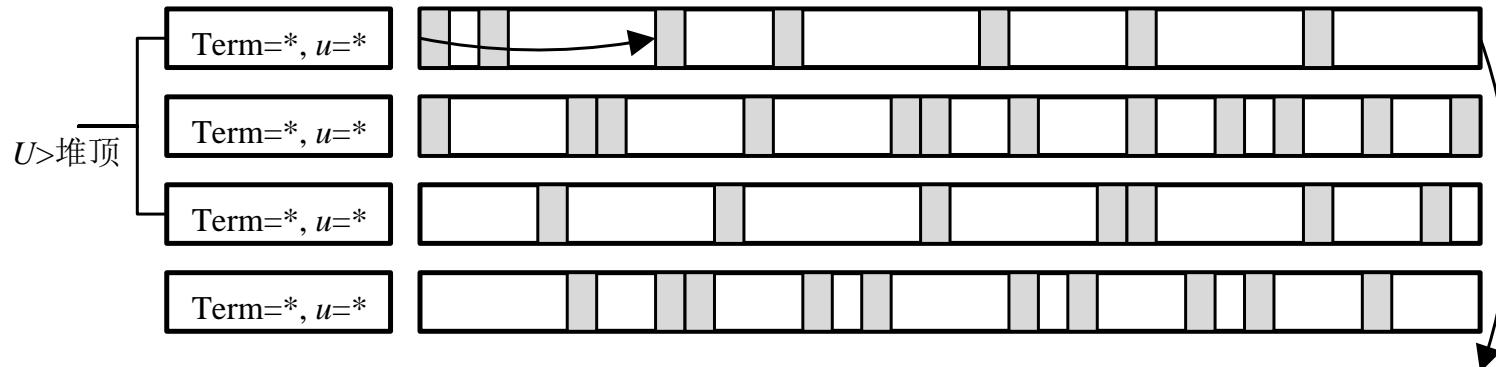
相关性检索

- 根据 (u, c) 检索广告时
 - query有可能比较长, 且子term都是should
 - 搜索引擎方案: term之间取或关系, 然后计算相关性并找到top- N , 在长query和大文档集时查询速度被巨大的计算量所限制
- 相关性检索
 - 在查找候选过程中做近似的评估, 切掉那些理论上不需要再考虑的文档, 只对进候选的文档进行相关性计算, 比Top-N最小堆最小值大的插入
 - 当相关性函数为线性时, 存在有效的剪枝算法

$$\text{score}(a, c) = \sum_{t \in \mathbf{F}(a) \cap \mathbf{F}(c)} \alpha_t v_t(a)$$

Weight-And (WAND) 检索算法

- 关键点：
 - 根据term贡献上界 u_t , 文档相关性上界 U_a 去掉无法胜出的候选
- 算法概要
 - step-1: 按doclist最前面的docID对terms排序
 - step-2: 迭代terms, 并累加UB直至大于堆顶, 设此时到达第n-1个term, 如果 $\text{terms}[0].\text{doc}$ 和 $\text{terms}[n-1].\text{doc}$ 一样, 逼出一个doc至最小堆; 如果不一样, 在前n个term挑选一个skip到 $\text{terms}[n-1].\text{doc}$, 跳转至step-1.



点击率预测问题

- 点击率预测，在 (a, u, c) 组合与点击间建立关系：

$$\mu(a, u, c) = p(h = 1 | a, u, c)$$

- Regression比Ranking合适一些

- 广告的实际排序是根据eCPM，因此需要尽可能准确估计CTR，而不仅仅是各候选的CTR排序正确

- 问题关键

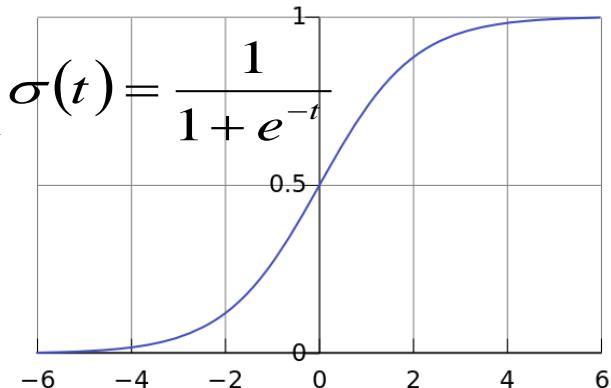
- 新广告的cold-start
 - 捕获点击率的动态特性

逻辑回归(Logistic Regression)

- 逻辑回归问题

$$p(h|a, u, c) = \sigma((2h - 1)\mathbf{w}^\top \mathbf{x}(a, u, c)) \rightarrow$$

参数 特征



- 实际带有正则化的优化问题

$$\min \left\{ C \sum_{i=1}^T \ln(1 + e^{-(2h_i - 1)\mathbf{w}^\top \mathbf{x}(a_i, u_i, c_i)}) + \frac{1}{2} \mathbf{w}^\top \mathbf{w} \right\}$$

- 可视为最大熵模型的特例，因此IIS方法适用

无约束优化问题一般思路

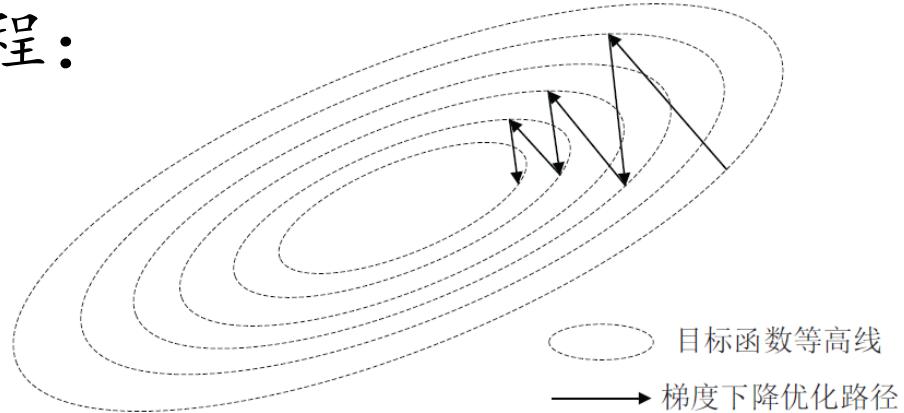
- 目标函数不可/不易求导
 - 下降单纯形法(Ameoba变形虫法)
- 目标函数易求导
 - 梯度下降法:

$$\boldsymbol{x} \leftarrow \boldsymbol{x} - \epsilon \nabla f(\boldsymbol{x}) \quad \nabla f(\boldsymbol{x}) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_D} \right)^\top$$

- 批处理模式: 训练集上的梯度分解为各个样本梯度的和, 可以并行实现
- 串行模式: 随机梯度下降法(Stochastic Gradient Descent, SGD)

批处理梯度法的问题与拟牛顿法

- 梯度法zigzag更新过程：



- 牛顿法：

$$\boldsymbol{x} \leftarrow \boldsymbol{x} + \epsilon [\nabla^2 f(\boldsymbol{x})]^{-1} \nabla f(\boldsymbol{x})$$

- 拟牛顿法：

- 用近似但正定的Hession阵 $\nabla^2 f(\boldsymbol{x})$ 确保稳定求解

BFGS和L-BFGS方法

- BFGS (Broyden, Fletcher, Oldfarb, and Shanno)
 - 拟牛顿法的一种，用函数值和特征的变化量来近似Hessian矩阵，以保证正定性，并减少计算量
 - Hessian拟合公式 (空间复杂度为 $O(n^2)$)：

$$H_{k+1} = H_k + \frac{s_k s_k^\top}{y_k^\top s_k} \left[\frac{y_k^\top H_k y_k}{y_k^\top s} + 1 \right] - \frac{1}{y_k^\top s_k} [s_k y_k^\top H_k + H_k y_k s_k^\top]$$
$$y_k = \nabla_{k+1} - \nabla_k$$
$$s_k = x_{k+1} - x_k$$

- L(imited memory)-BFGS
 - 将Hessian逆用右图方式加以近似
 - 空间复杂度降为将 $O(n \times k)$

$$B_k = \delta_k I + \begin{array}{c|c|c} & \square & \\ \times & \square & \times \\ & \{k \times k\} & \{k \times n\} \\ \hline & & \\ \{n \times k\} & & \end{array}$$

Trust-region方法

- 方法思想

- 不近似Hessian阵，但每次迭代将自变量限制在邻域内

$$\begin{aligned} \min_{\mathbf{s}} \quad & f(\mathbf{x}_k) + \nabla f(\mathbf{x}_k)^{\top} \mathbf{s} + \frac{1}{2} \mathbf{s}^{\top} \nabla^2 f(\mathbf{x}_k) \mathbf{s} \\ \text{s.t.} \quad & \|\mathbf{s}\|_2 \leq \delta_k \end{aligned}$$

- 先步长，后方向

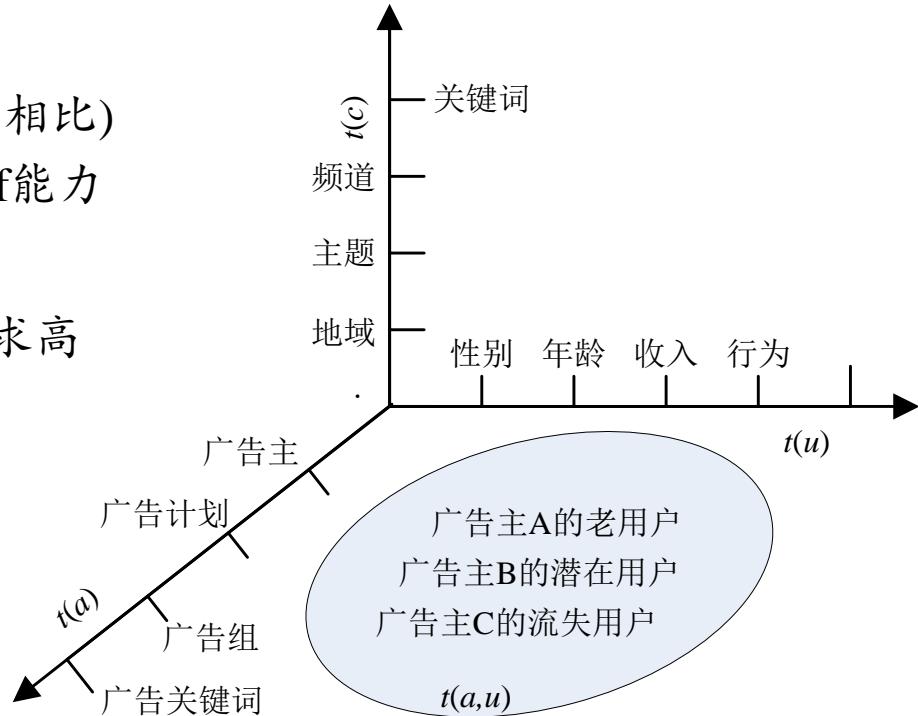
- 上述子问题虽非凸优化，但满足KKT条件，求解如下：

$$\begin{cases} (\mathbf{H}_k + \lambda \mathbf{I}) \hat{\mathbf{s}} = -\nabla f(\mathbf{x}_k) \\ \lambda(\delta_k - \|\hat{\mathbf{s}}\|) = 0 \\ (\mathbf{H}_k + \lambda \mathbf{I}) \geq 0 \end{cases}$$

- 对于LR模型收敛速度经常好于L-BFGS

动态特征 - 多层次点击反馈

- 在标签组合维度上聚合点击反馈统计作为CTR预测的特征
- 优势
 - 工程架构扩展性强(与在线学习相比)
 - 对新 (a, u, c) 组合有较强back-off能力
- 缺点
 - 在线特征的存储量大，更新要求高
- 组合维度举例
 - $\text{cookie}(u)$ and $\text{creative}(a)$
 - $\text{gender}(u)$ and $\text{topic}(c)$
 - $\text{category}(a)$ and $\text{category}(u)$
 - $\text{creative}(a), \text{gender}(u)$



归一化点击率 - COEC

- 有效展示
 - 可以通过Eye tracking测算
 - 工程上可以使用Expected Click (EC)来近似有效展示
- EC的计算
 - 实测法：实际随机流量测算
 - Bias模型法：只使用与广告决策无关的bias特征训练的CTR模型：
$$EC = p_{bias}(a, u, c)$$
- 归一化点击率 $COEC = \sum_i \text{click}_i / \sum_i EC_i$



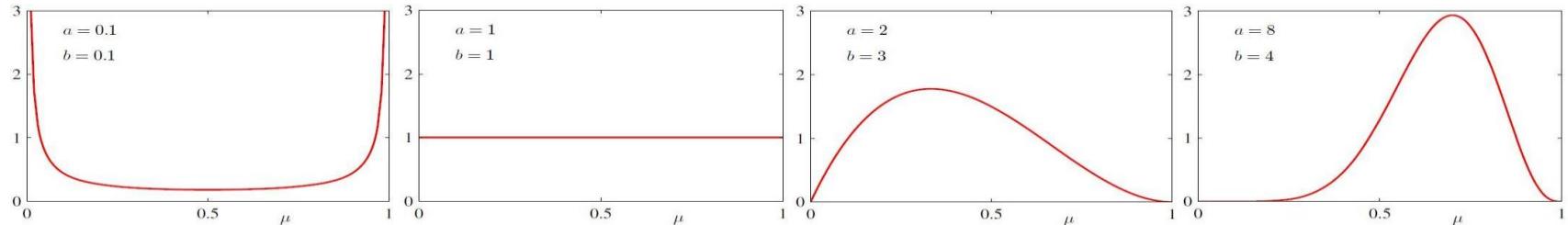
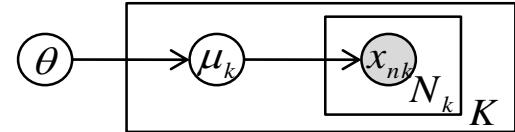
在线广告常见bias特征

- 广告位位置
 - 搜索广告: North 1, North 2, ..., East 1, East 2, ...
 - 显示广告: 相对页面的(x, y)
- 广告位尺寸
- 广告位类型
 - 门户首页, 频道首页, 内容页, 客户端, ...
- 操作系统和浏览器
- 广告投放延迟
- 日期和时间

点击反馈的平滑

- 问题：在数据稀疏的情况下较稳健地估计CTR或COEC
- 经验贝叶斯方案
 - 点击产生概率模型： $p(x | \mu) = \mu^x (1 - \mu)^{1-x}$
 - 视 μ 为随机变量，采用Beta分布共轭先验进行正则化：

$$p(\mu | \theta = \{a, b\}) = \text{Beta}(\mu; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \mu^{a-1} (1-\mu)^{b-1}$$

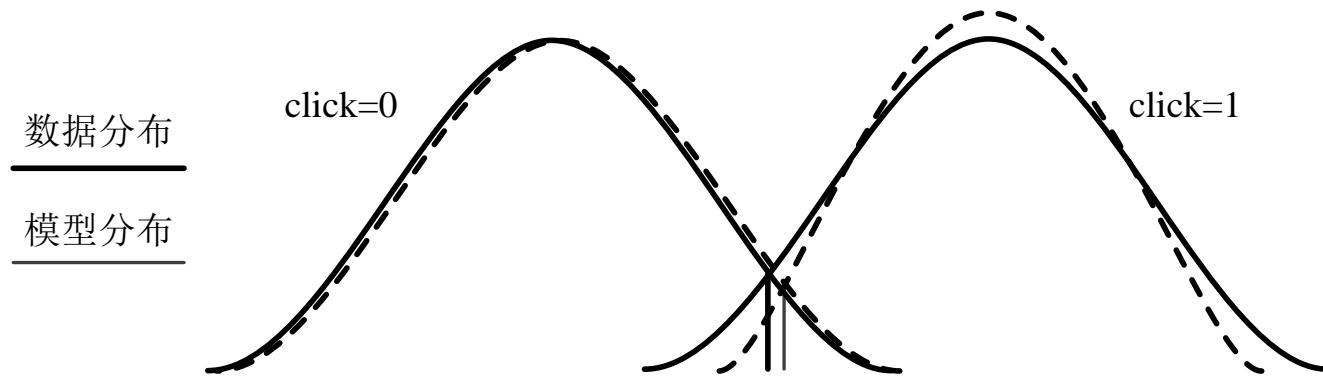


经验贝叶斯解

- E-step: $\tilde{\alpha}_k^{\text{old}} = \alpha^{\text{old}} + \sum_{i=1}^{N_k} h_{k,i}, \quad \tilde{\beta}_k^{\text{old}} = \beta^{\text{old}} + (N_k - \sum_{i=1}^{N_k} h_{k,i})$
- M-step: $\psi(\alpha^{\text{new}}) - \psi(\alpha^{\text{new}} + \beta^{\text{new}}) = \frac{1}{K} \sum_{k=1}^K \psi(\tilde{\alpha}_k^{\text{old}}) - \psi(\tilde{\alpha}_k^{\text{old}} + \tilde{\beta}_k^{\text{old}})$
 $\psi(\beta^{\text{new}}) - \psi(\alpha^{\text{new}} + \beta^{\text{new}}) = \frac{1}{K} \sum_{k=1}^K \psi(\tilde{\beta}_k^{\text{old}}) - \psi(\tilde{\alpha}_k^{\text{old}} + \tilde{\beta}_k^{\text{old}})$
- 点击率的平滑通常在每个维度组合上分别进行
- 物理意义
 - 数据不足时，更依赖于相应维度组合的点击率先验值
- 思考：对COEC的平滑可以采用什么模型？

点击率模型的校准

- 正负样本不均衡带来的估计误差



- 误差对一些广义线性模型都存在，包括LR
- 可以通过数据计算并校准

探索与利用 (E&E)



- 问题
 - 为长尾的(a, u, c)创造展示机会以积累统计量，从而更准确估计CTR
 - 提升整体的广告收入，即需要严格控制探索的量和有效性
- 方法思路
 - 通常描述为Multi-arm Bandit (MAB) 问题
 - 有限个arms(或称收益提供者) a , 每个有确定有限的期望收益 $E(r_{t,a})$
 - 在每个时刻 t , 我们必须从arms中选择一个, 最终目标是优化整体收益
 - 基本方法为 ε -greedy: 将 ε 比例的小部分流量用于随机探索
- 广告问题中的主要挑战
 - 海量的组合空间需要被探索
 - 各个arm的期望收益是动态变化的

E&E 算法 – Contextual Bandit



- 问题描述
 - 对每次展示，每个arm(广告) a 有一个对应的特征矢量 $\mathbf{x}(u,a)$
 - 用此特征矢量代替arm本身进行Bandit决策

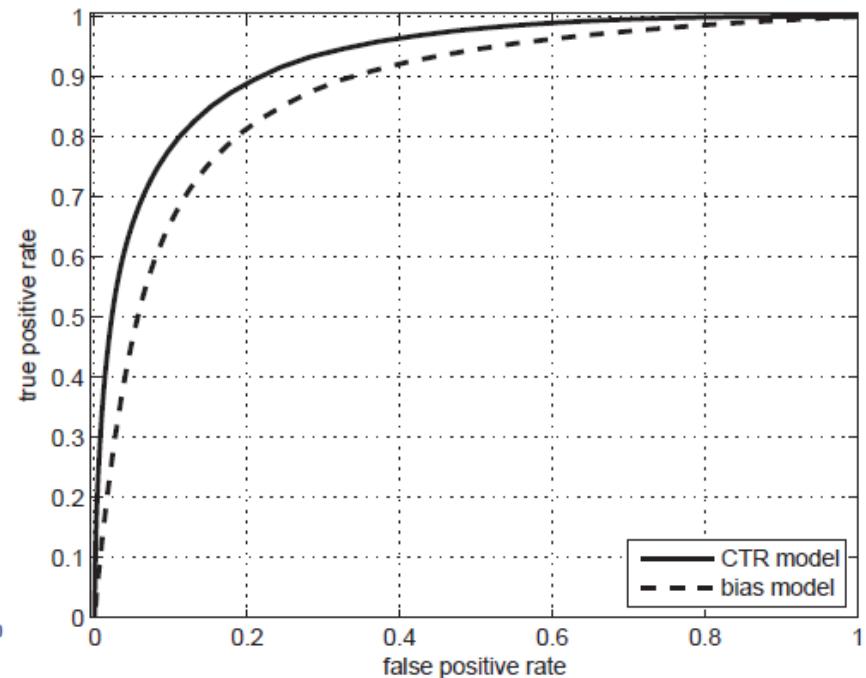
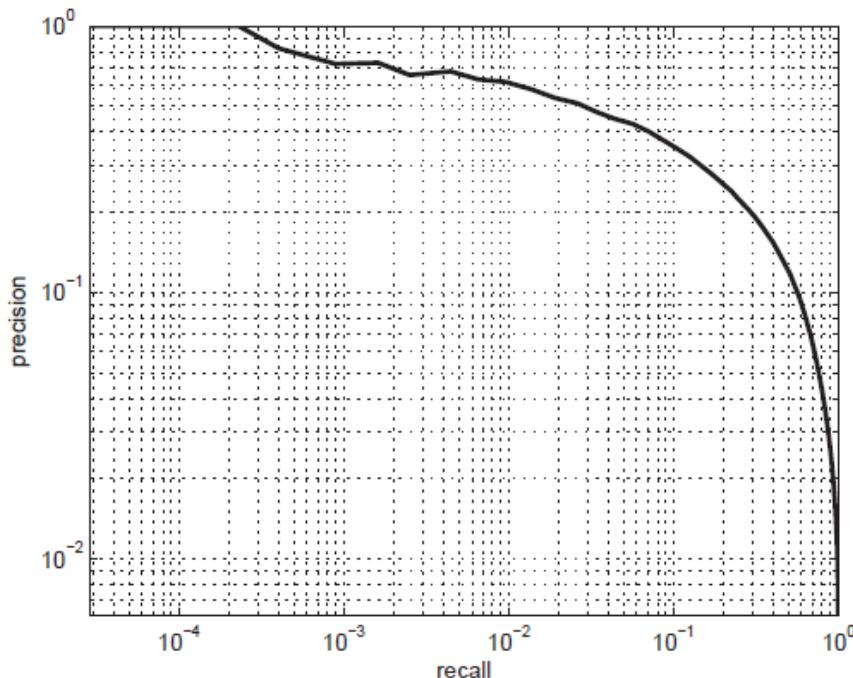
• LinUCB

- Disjoint linear model: $E(r_{t,a}|\mathbf{x}_{t,a}) = \mathbf{f}^\top(a, u_t, c_t) \boldsymbol{\theta}_a^*$
- 参数的岭回归(Ridge regression)解:

$$\hat{\boldsymbol{\theta}}_a = (\mathbf{D}_a^\top \mathbf{D}_a + \mathbf{I}_d)^{-1} \mathbf{D}_a^\top \mathbf{c}_a$$

- $\{\mathbf{D}_a\}_{\# \text{特征维度} \times \# \text{样本数}}$ 为观测样本， \mathbf{c}_a 为回报(广告中是点击)矢量
- 特征空间内的UCB策略: $a_t = \arg \max_a \left(\mathbf{x}_{t,a}^\top \hat{\boldsymbol{\theta}}_a + \alpha \sqrt{\mathbf{x}_{t,a}^\top (\mathbf{D}_a^\top \mathbf{D}_a + \mathbf{I}_d)^{-1} \mathbf{x}_{t,a}^\top} \right)$

评测指标 – PR曲线(左)与ROC曲线(右)

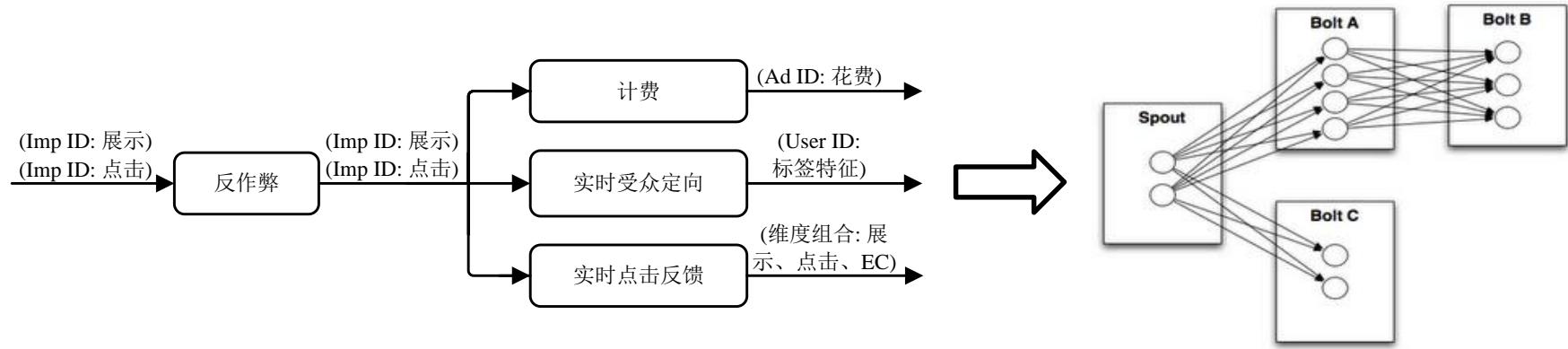


短时用户行为反馈

- 短时用户行为
 - 狹义: 用户在一个session内的行为
 - 广义: 用户在短时间(一般为一刀两天)内的行为
- 短时用户行为反馈
 - 短时受众定向: 根据短时行为为用户打上的标签
 - 短时点击反馈: 根据短时广告交互计算的动态特征
- 短时用户行为计算
 - 需要准实时(分钟级)对用户行为进行加工, 不适合在Hadoop上进行
 - 可以利用流式计算(stream computing)平台, 如S4, Storm等

流式计算平台 - Storm

- 大规模实时数据处理框架，自动完成数据分发和可靠性管理，开发者只需要关注处理逻辑。数据流基本在网络和内存进行。
- 计算逻辑类似Map/Reduce，区别在调度数据而非调度计算



Topology

Tasks

广告购买平台 (Trading Desk)

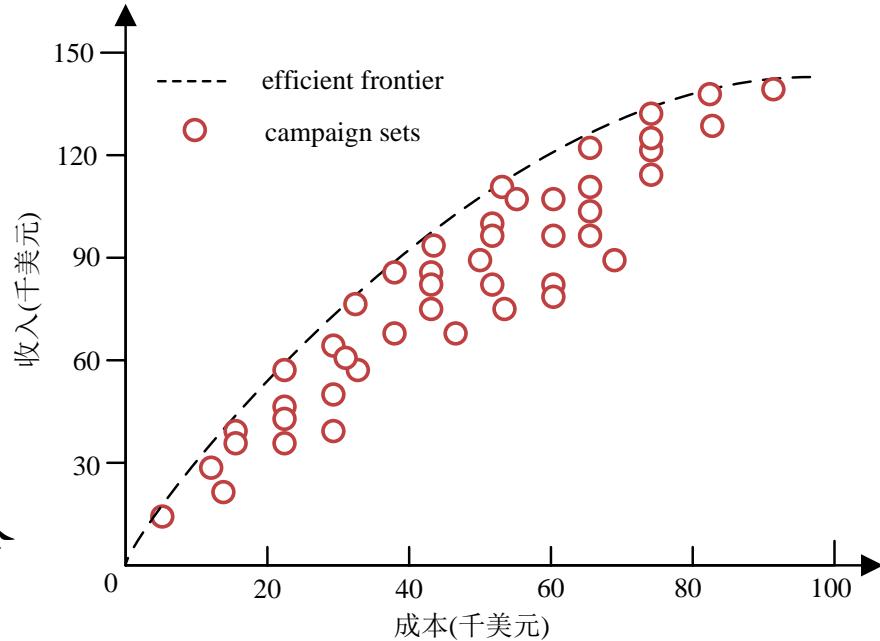
- 产品目标:
 - 帮助广告主按人群进行跨媒体和广告网络采买
- 关键特征:
 - 连接到不同媒体和广告网络，提供统一市场
 - 非实时竞价投放的ROI优化能力
 - 经常由代理公司孵化出来

非RTB流量的ROI优化

- 目标:
 - 给定总预算，在多广告网络中采买并优化ROI(例: SEM)
- 关键问题:
 - 在合适的流量segment上投放广告
 - SEM中的选词，显示广告网络中的标签组合选择
 - 在每个投放上合理地出价以优化ROI
 - 与RTB不同，采买方无法控制每次展示的出价
 - u, c 的取值未知，需要在各流量分割上估计其分布并合理出价

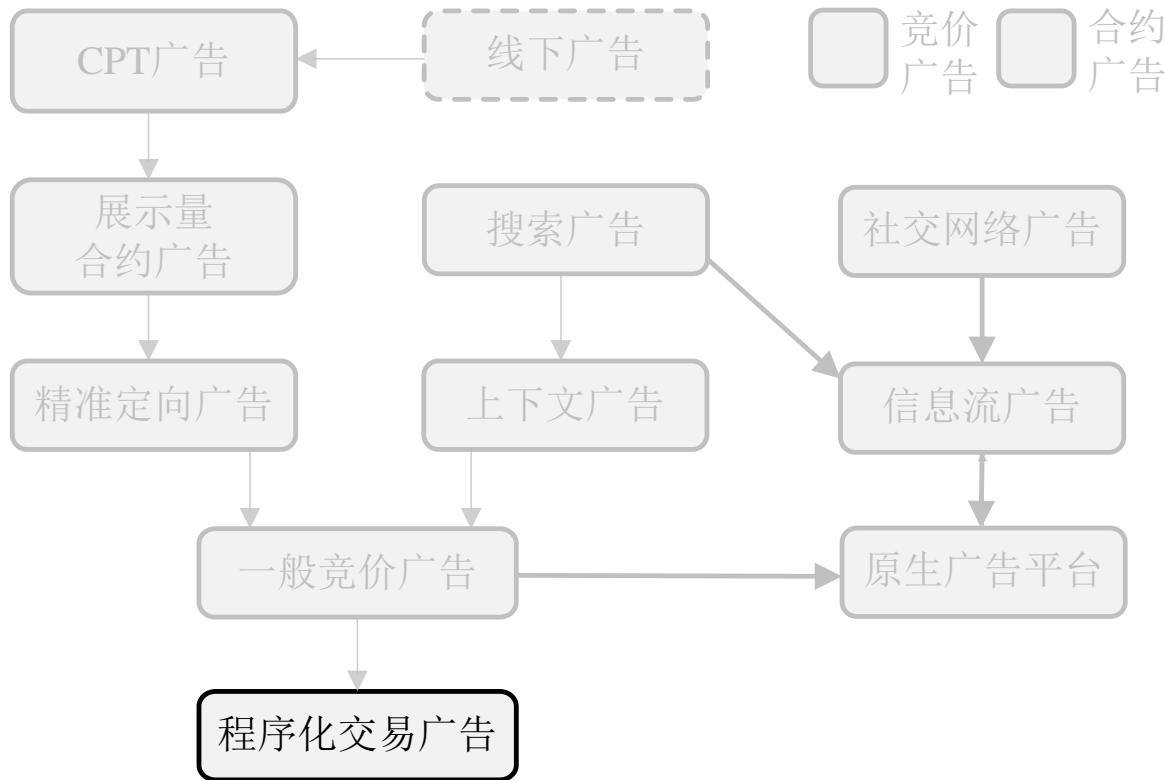


- 核心业务：
 - 为搜索广告主提供大量关键词情形下的ROI优化服务
 - 核心技术为Portfolio Optimization
- 其他点评：
 - 核心竞争力除了算法还来自于长时间数据积累
 - 被Adobe收购与Omniture数据结合，现改名Adobe Adlens

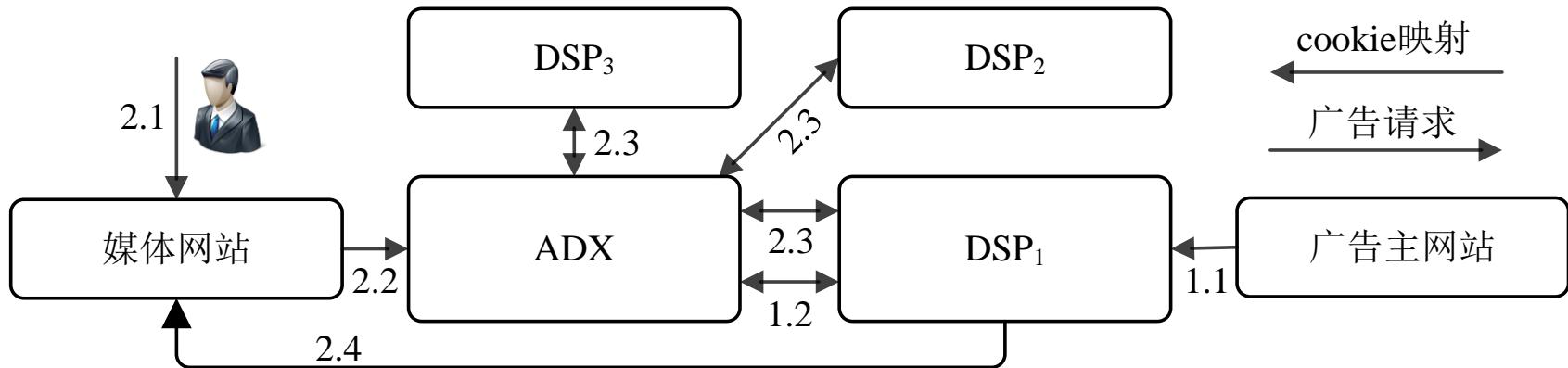


程序化交易广告

程序化交易广告



实时竞价过程示意

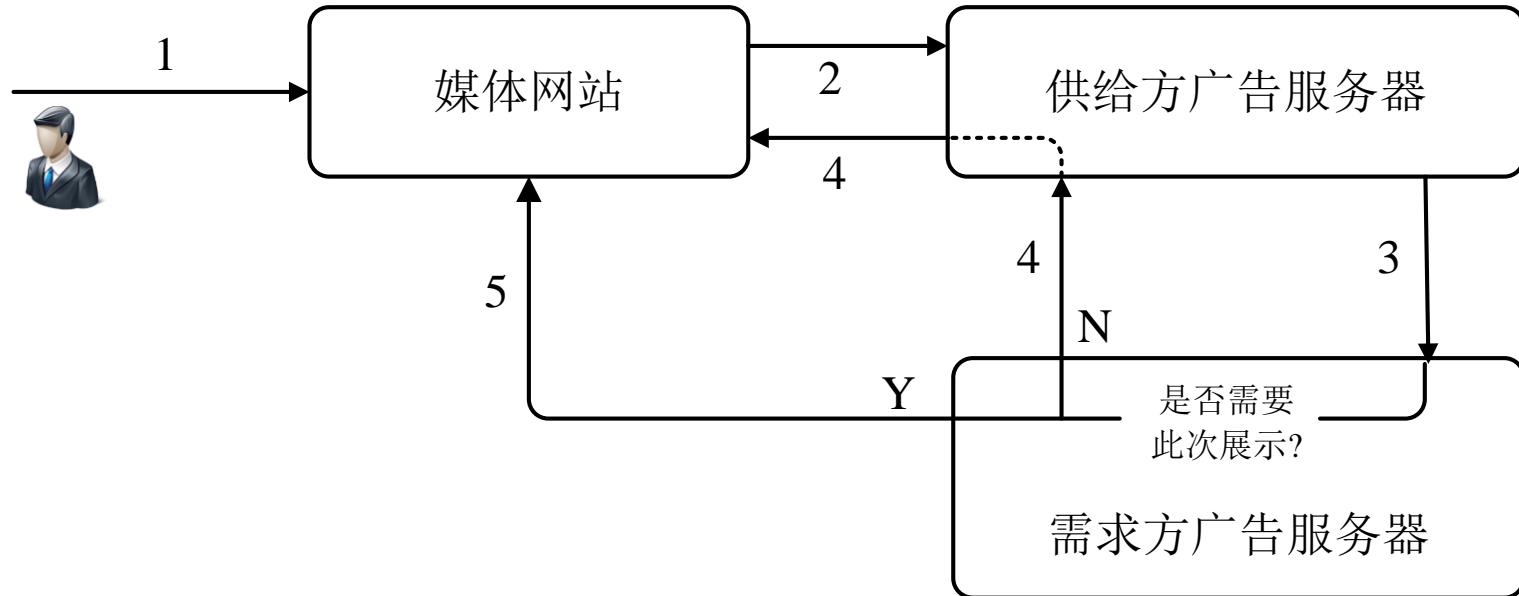


- Ad Exchange方实现的部分称为RTBD(RTB for Demand)
- 在DSP方实现的部分，我们称为RTBS(RTB for Supply)。

实时竞价特点

- 每次展示都有Ad Exchange服务器与多个DSP服务器的参与，这使得服务器与带宽成本大大增加；
- 在询价过程中，Ad Exchange要等待一个约定好的时间片(一般情况下为100ms)，这使得用户看到的广告延迟增加，对CTR有负面影响；
- 原理上DSP可以以极低的出价参与竞价，这样虽不能获得流量，却可以低成本得到在媒体网站上的用户行为数据，这里存在着潜在的信息泄露风险。

其他程序化交易 - 优选(Preferred Deal)



其他程序化交易 - 私有市场(PMP)

- 媒体为了保证广告主的质量，希望将拍卖限制在一些被邀请需求方的小范围内。这种程序化交易，叫做私有市场(Private Marketplace, PMP)
 - 兼顾了优选与实时竞价的好处
 - 广告主的质量可以由少量被邀请需求方很好地控制，这有利于确保媒体的价值不受伤害；
 - 在需求方之间保留了竞价了关系，有利于提升媒体的变现能力
- 交易接口的形式与一般RTB一致

在线广告交易方式

程序化交易				半程序化交易	优先销售	
供给方	RTBD	PMP	Preferred Deal	Network Optimization	CPM	CPT
中间平台	Ad Exchange			Ad Network	Ad Server	
需求方	RTBS	PMP	Preferred Deal	Portfolio Selection		

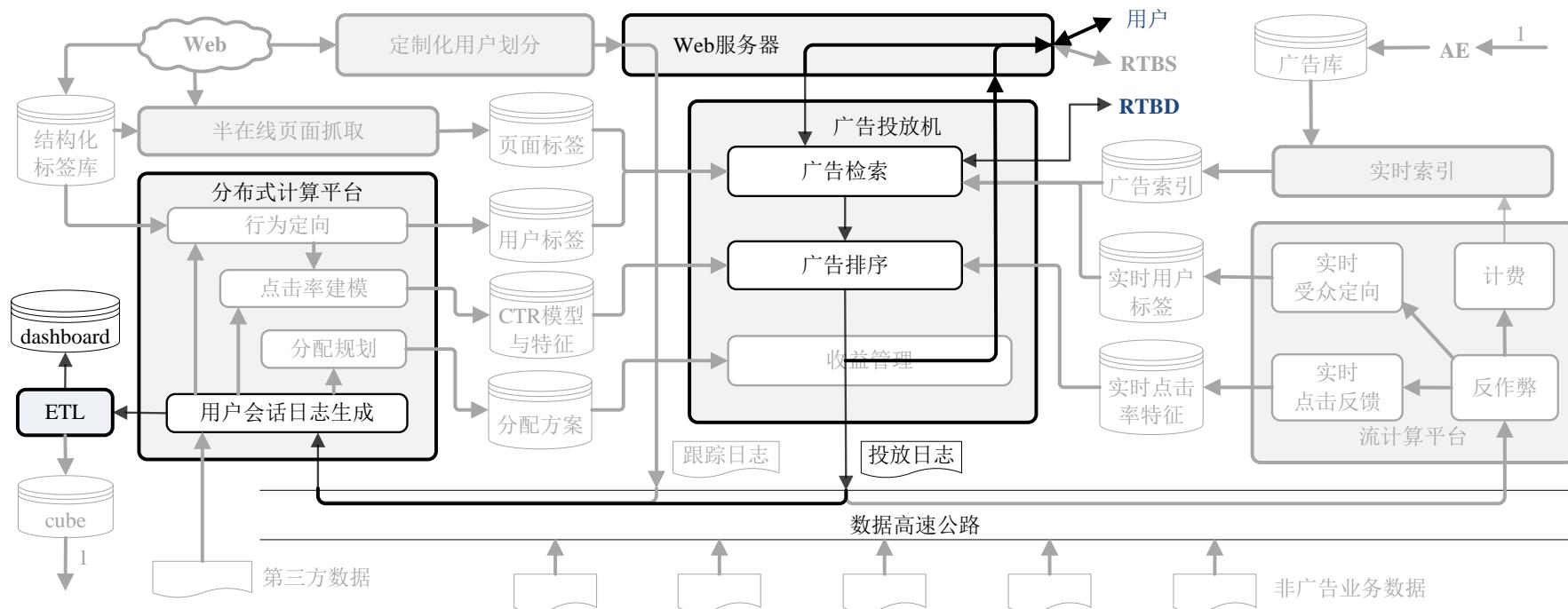
广告交易平台

- 关键特征：
 - 用实时竞价(RTB)方式连接广告和(上下文，用户)
 - 按照展示上的竞价收取广告主费用
- 类比：
 - 实时竞价 \leftrightarrow 股票实时交易
 - 广告交易市场 \leftrightarrow 证券交易所
 - 需求方平台 \leftrightarrow 券商

广告交易平台(Ad Exchange)概貌

- 优化目标: $\max_{a_1, \dots, T} \sum_{i=1}^T \text{bid}_{\text{CPM}}(a_i)$
- 关键特点:
 - 竞价方式不向广告主做量的约定，而是根据变现能力，即eCPM，来决定每次展示分配给哪个广告主。
 - 按人群售卖，淡化媒体和广告位的概念
 - 无需再满足广告主品牌独占的要求
 - 采用分成结算，运营方的现金流状况大为改善

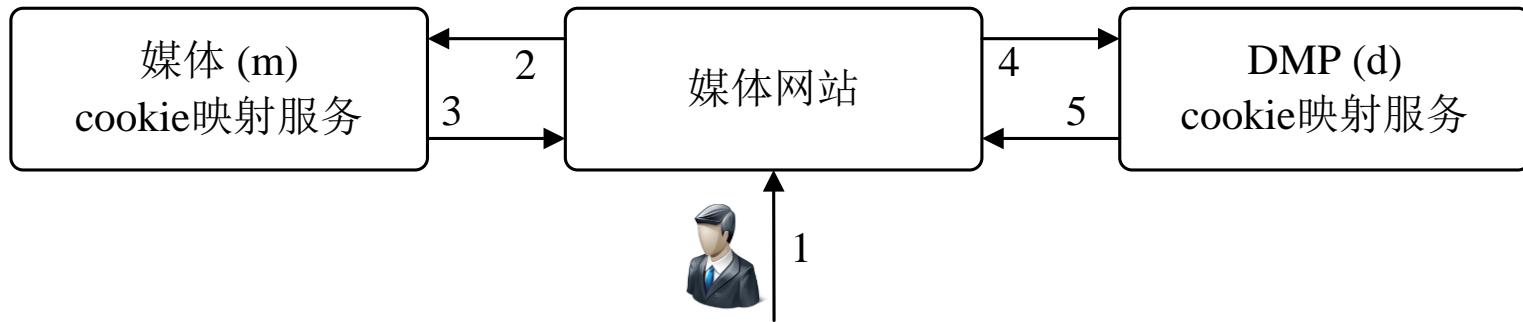
Ad Exchange 系统架构



Ad Exchange主要技术点

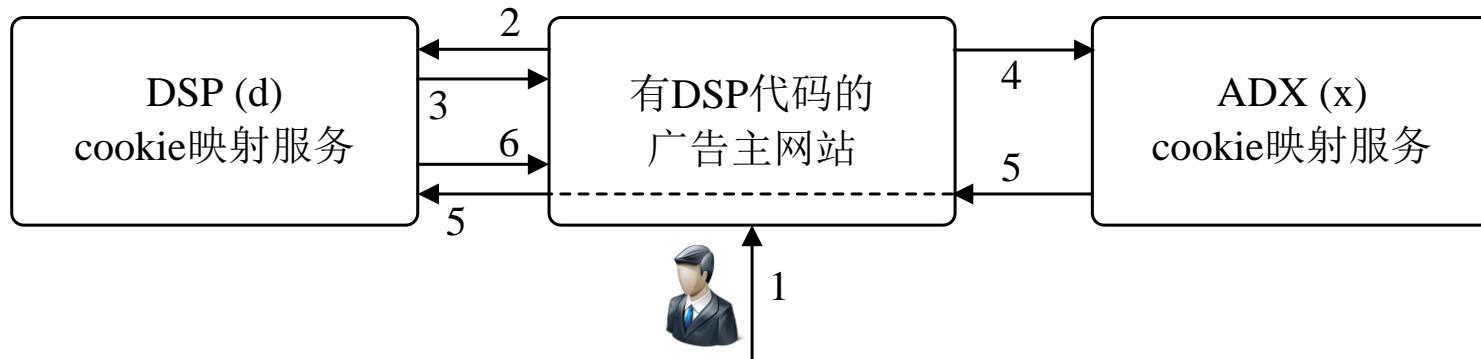
- 由于实时竞价的功能需求，广告交易市场解决供给方和需求方用户身份对应的问题，在web的广告环境下，这需要用到**cookie-mapping**的技术。
- 实践中当考虑到带宽和服务成本带来的约束时，希望用更少的询价请求完成尽可能高效的变现，在这种情况下，公式的优化问题会有所变化，这一问题称为询价优化。

两域名Cookie Mapping (例: 媒体与DMP)



1. 用户到达媒体页面；
2. 向媒体的服务器请求cookie mapping JS代码；
3. 媒体的cookie mapping服务返回该段JS代码；
4. JS代码判断需要mapping的话(最近已经做过则可以不做)，向DMP发起cookie mapping请求，并传送媒体的标识(mid) 及媒体方的cookie(mck)；
5. DMP返回一个 1×1 的beacon，并记录下媒体方cookie(mck)与己方cookie(dck)的对应关系。

三域名Cookie Mapping (例: ADX与DSP)



1. 用户访问广告主页面；
2. 选择性加载一个DSP域名下的iframe；
3. DSP判断需要mapping的话，返回包括多个beacon的动态html，此处多个beacon的目的是为了同时与多个ADX交换cookie；
4. 通过其中的某个beacon，向对应的ADX发送cookie-mapping请求，并带有ADX标识(xid)、DSP标识(did)和DSP cookie(dck)三个参数；
5. ADX通过302重定向向DSP返回ADX标识(xid)以及其域名下的cookie(xck)；
6. DSP返回beacon，并记录下ADX cookie(xid)与己方cookie(dck)对应关系

询价优化(Call Out Optimization)

- 问题：
 - 在带宽和服务成本的约束下获得更高的eCPM
- 在线分配框架：

$$\max \sum_{(i,a)} \sum_k k y_{iak} \quad \text{DSP } a \text{ 以出价 } k \text{ 赢得供给 } i \text{ 的展示的概率}$$

$$\text{s.t. } \sum_i x_{ia} \leq \rho_a \quad \text{可以给予DSP } a \text{ 的流量的上限}$$

$$x_{ia} \leq 1; y_{iak} \leq p_{iak} x_{ia} \quad \text{DSP } a \text{ 为供给 } i \text{ 的一次展示出价 } k \text{ 的概率}$$

$$\sum_{(i,a)} y_{iak} \leq 1; x_{ia}, y_{iak} \geq 0$$



- 核心业务：
 - 提供媒体网络和广告主网络之间的交易平台
 - 提供托管和RTB两种交易模式
- 其他点评：
 - 在很长一段时间以私有交易为主，并没有开放RTB模式，与 Doubleclick的公开市场模式有一些不同
 - 目前是北美最大的广告交易平台之一，每天100亿次展示，有相当部分流量来自yahoo的剩余流量
 - 与DMP合作以提高定向精准性和市场流动性

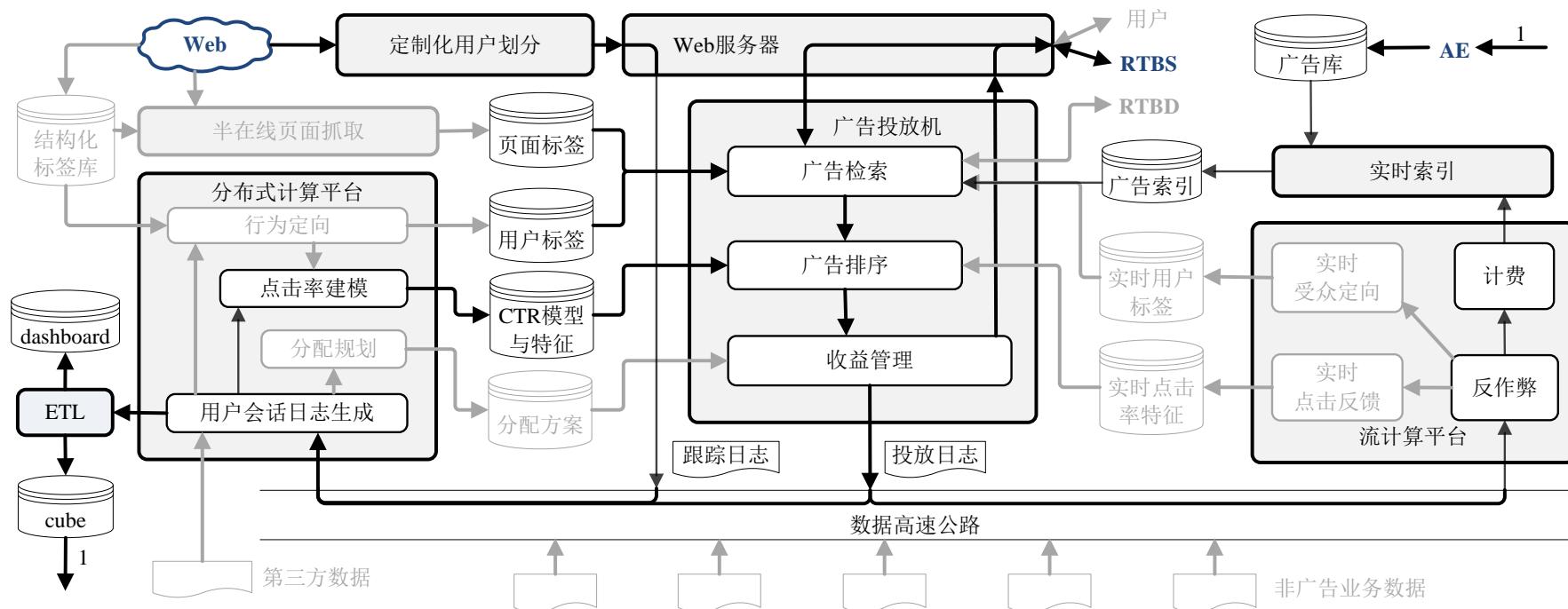


- 核心业务：
 - 提供公开RTB模式的交易市场，目前也支持PMP
 - 与DoubleClick其他产品构成完整系列，服务于供给方与需求方
 - 与Google收购的SSP正在整合同一个产品
- 其他点评：
 - 全球最大的广告交易平台
 - 有比较完善的数据交易、广告安全等附加第三方服务，以提高市场的流动性
 - 由于媒体和广告主种类众多，环境和规则比较复杂

需求方平台(Demand Side Platform)概貌

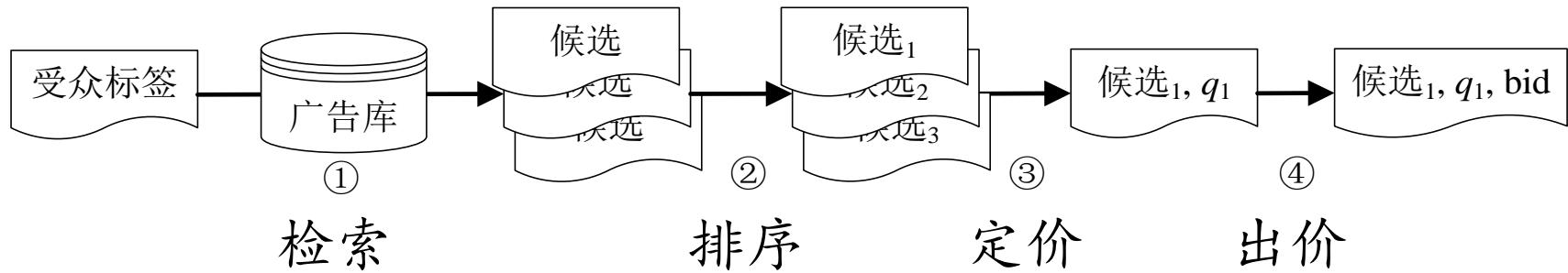
- 优化目标: $\max_{a_1, \dots, T} \sum_{i=1}^T \{\mu(a_i, u_i, c_i) \cdot \nu(a_i, u_i) - p(u_i, c_i)\}$
- 关键特点:
 - 需要支持定制化的用户划分能力。
 - 由于DSP是完全面向广告主的产品，需要在量的约束下投放，这产生了对于出价策略的需求。
 - 在按CPS等效果结算的DSP中，需要同时估计点击率点击价值。并且由于出价的要求，估计都要尽可能准确

DSP 系统架构



DSP产品策略

- DSP展示决策过程：

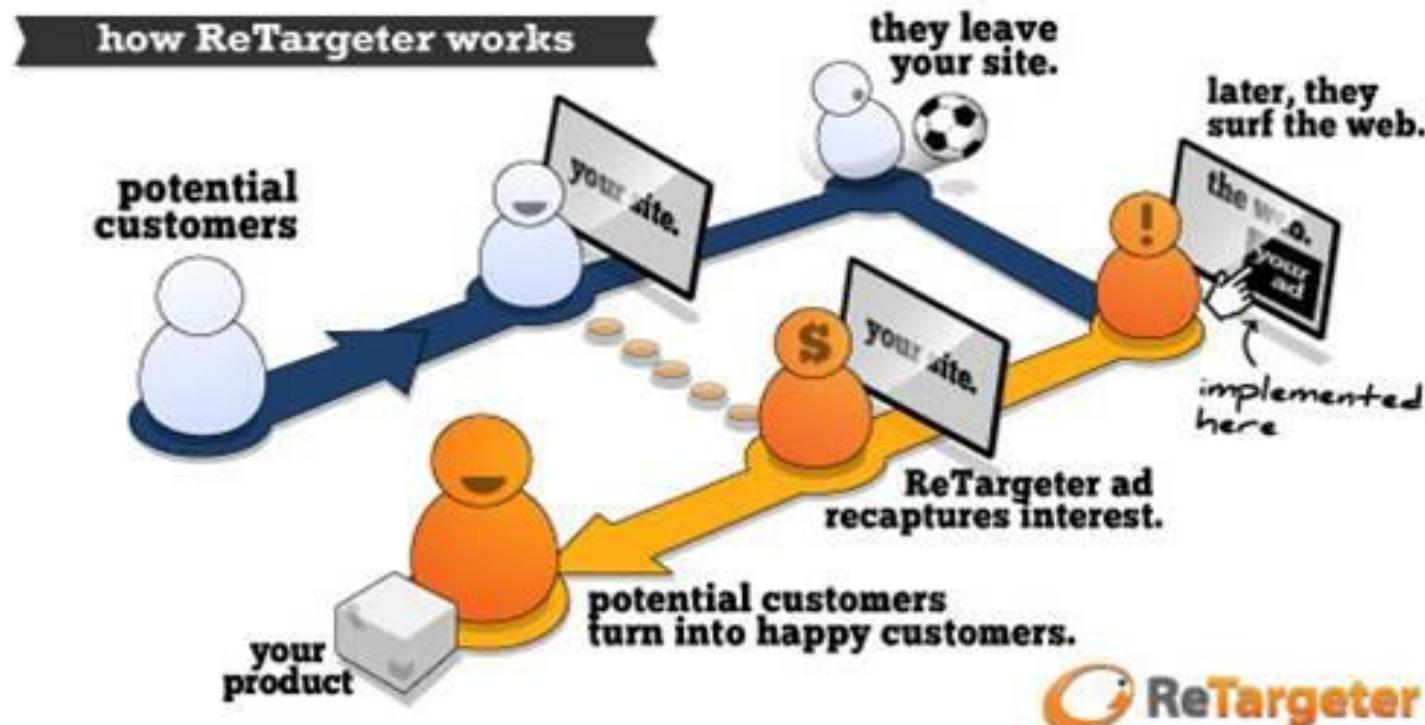


- 关键产品策略：
 - 点击价值估计、出价策略

定制化用户标签

- 定制化标签：
 - 根据广告主需求加工的标签体系
 - 往往要用到第一方数据
- 定制化标签典型方式：
 - 网站重定向→个性化重定向
 - 搜索重定向
 - 新客推荐(Look-alike)

重定向 (Retargeting) 示意



重定向分类

- 网站重定向(Site retargeting)
 - 根据用户在广告主网站上的行为进行重定向
- 搜索重定向(Search retargeting)
 - 根据用户与广告主相关的搜索行为进行重定向
- 个性化重定向(Personalized retargeting)
 - 根据用户在广告主网站上关注的具体产品和购买阶段，推送商品粒度的广告，可以视为一个站外推荐引擎



新客推荐(Look-alike)

- 问题：
 - 对于中小电商，仅对老用户定向营销远远不够
 - 对于某些类型的广告商，大多数用户无法通过重定向渠道捕捉，例如银行
- 新客推荐：
 - 由广告商提供一部分种子用户，DSP通过网络行为的相似性为其找到潜在用户
 - 是一种广告商自定义标签，可视为扩展重定向
 - 在同样reach水平下，效果应好于通用标签

新客推荐问题建模

- 问题:

$$p(y = 1 | x(a, u))$$

- y 不再是点击行为，而是表示用户能否成为广告主用户的二元变量
- 模型评估的是用户属性，因此与上下文 c 无关

- 训练集构建:

- 方法一：根据广告主提供的种子用户集，凡是出现在该种子用户集中的 u ，其对应的 y 标为1，否则标为0
- 方法二：根据广告投放的记录，将点击过该广告主广告一定次数(一般设为1)以上的用户，其对应的 y 标为1，否则标为0

点击价值估计

- 点击价值分解:

$$v(a, u, c) = h(a, c) \cdot c(a, u) \cdot t(a)$$

到达率 转化率 客单价

- 到达率:
 - 主要与媒体匹配程度、广告主网页加载速度相关
- 转化率:
 - 行为稀疏、定义与广告主类型密切相关

从电商角度看推荐



pv: ..., search: ...,
adc: ..., share: ...



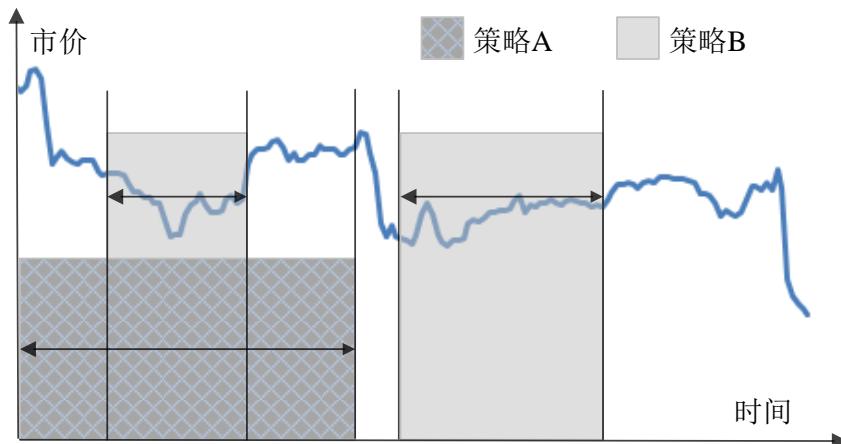
看起来很相似！

pv: ..., search: ...,
adc: ..., share: ...



出价策略

- 如果没有预算的限制：只要按照eCPM水平出价，就可以保证在第二高价的情况下每次展示都有利润。
- 在有预算约束的情况下，我们需要估计eCPM以及当前展示的市场价格，并在尽可能将出价集中在那些利润率较高的展示上。



问题框架：

$$\begin{aligned} & \max \sum_{(i,a) \in E} s_i x_{ia} (r_{ia} - m_i) \\ \text{s.t. } & \sum_{a \in \Gamma(i)} x_{ia} \leq 1, \quad \forall i \in I \\ & \sum_{i \in \Gamma(a)} s_i x_{ia} m_i \leq d_a, \quad \forall a \in A \\ & x_{ia} \geq 0, \quad \forall (i, a) \in E \end{aligned}$$



- 核心功能：
 - 基于站外推荐的个性化重定向系统
 - 动态创意技术
 - 与广告主商品库的准实时feed接口
- 其他点评：
 - 兼用RTB和优选方式购买流量
 - 在广告主端完全采用CPC结算方式
 - 不会将cross-site数据用于推荐



- 核心业务：
 - 传统业务是提供比较透明的Adx采买功能，并收取固定比例的佣金
 - 正在提供越来越深入的采买ROI优化服务
- 其他点评：
 - 与DFA, Adx已整合，可减少ck mapping带来的损失
 - 改称doubleclick bid manager，并提供更多的优化功能



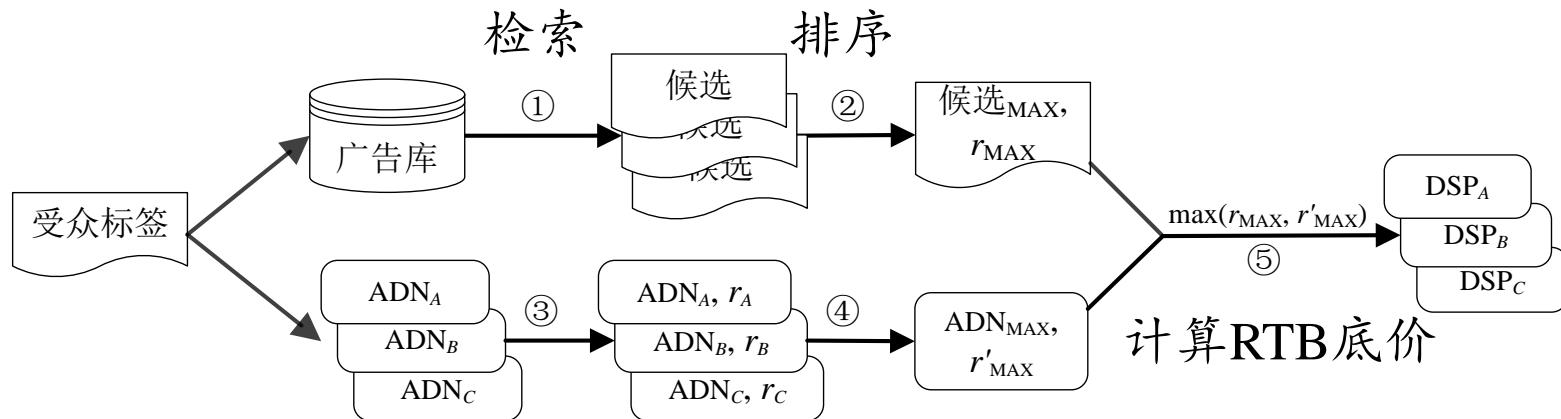
- 核心功能：
 - DSP+Ad Network模式，对广告主透明
 - 初期以电商为主要客户，与Criteo产品相近，服务中国绝大多数B2C，后与360对接后转向服务于中小广告主
 - 实现了部分移动与PC跨屏数据打通
- 其他点评：
 - 兼有数据交易市场“聚合数据平台”，但数据提供方不多

网络优化(Network Optimization)与SSP

- 网络优化问题:
 - 接入多个Ad Network, 动态决定某次展示分配给谁
 - 相当于DSP不主动出价的交易平台
- 供给方平台(Supply Side Platform, SSP)
 - 综合利用各种需求方预算, 在不伤害媒体品牌属性的前提下优化收入
 - 也会提供通过RTB对接DSP的接口
 - 与Ad Exchange的边界越来越模糊

SSP产品策略

- SSP展示决策过程—动态分配(Dynamic Allocation):

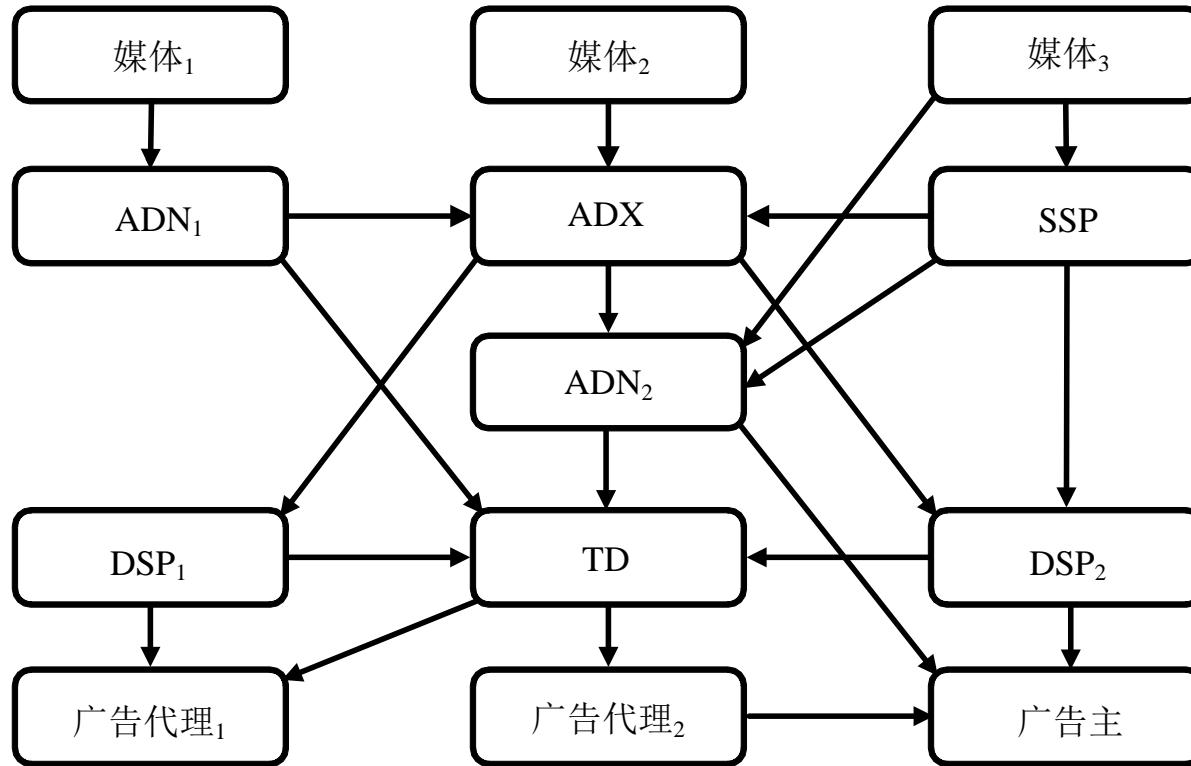


估算 Network 的 eCPM 对 Network 排序



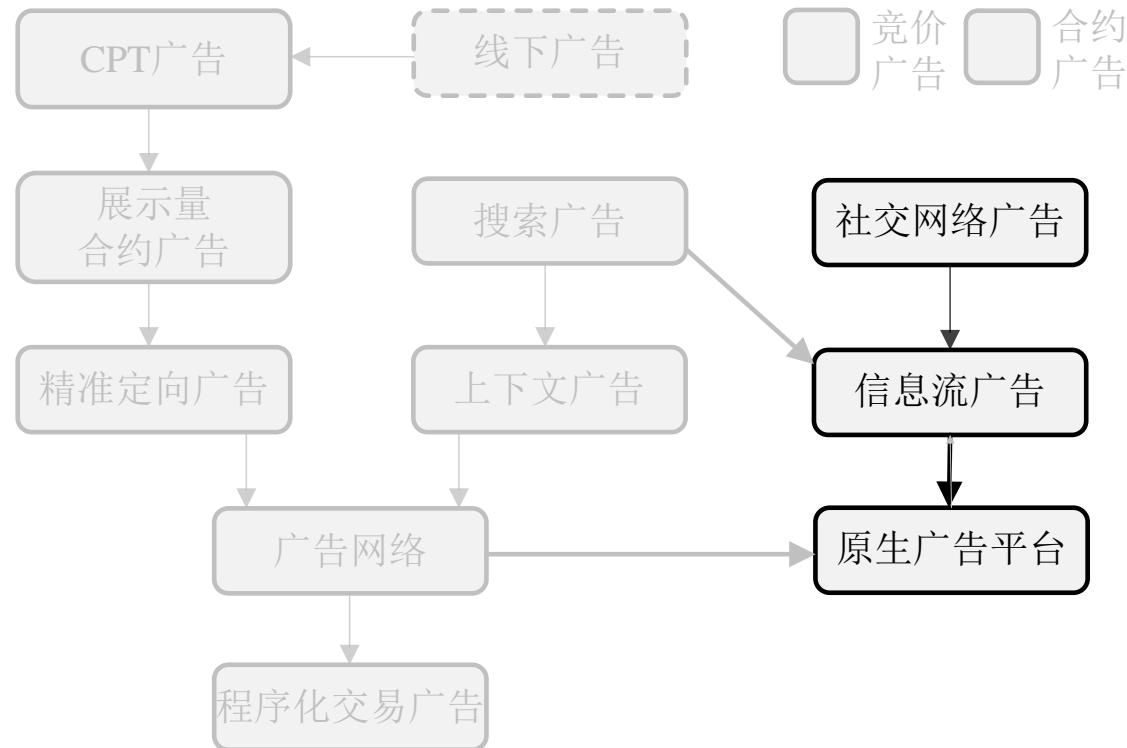
- 核心功能：
 - Yield Optimizer, 统一优化premium sales, network和RTB流量
 - 以优化媒体利益为目标
- 其他心得：
 - 起主要作用的部分除了RTB, 还有多network接入时eCPM的估计和选择。
 - 会从DMP购买标签数据用于优化广告效果
 - 主要进行广告位和时间维度上eCPM估计和流量切分

在线广告市场产品交互关系



移动互联与原生广告

原生广告



原生广告初级产品

• 信息流广告:



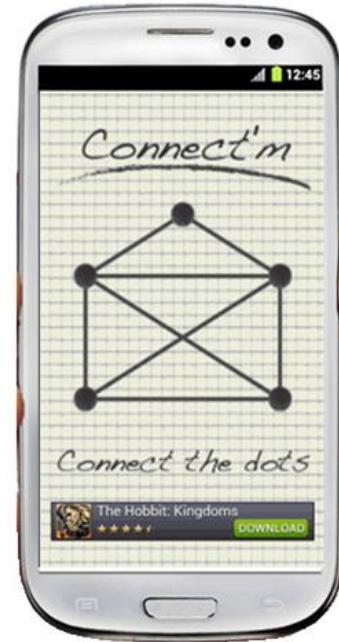
• 搜索广告:

• 软文广告

移动广告的特点

- 情境广告的可能性:
 - 移动设备与PC最大的不同，是可以对用户行为模式进行全天候的监测和分析。
 - 从地理位置、生活状态、需求意图等各方面都可能对用户有深入的理解。
- 大量潜在的O2O广告主
 - GPS、蜂窝、wifi等多种精确定位的手段，使得基于精确地理位置的O2O广告变得可行

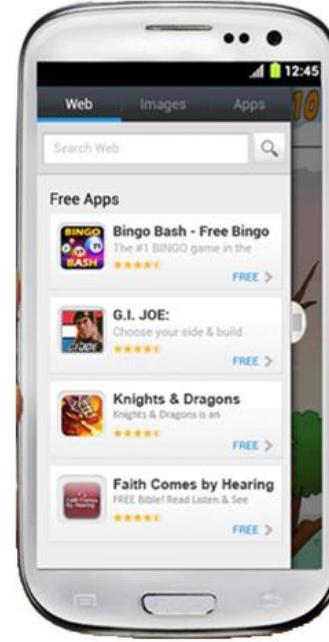
移动广告的创意形式



条幅(banner)



全屏



应用推荐、积分墙

移动广告的挑战

- 应用生态造成的行为数据割裂。
- 许多PC时代广告主移动化程度还不够，无法充分消化广告带来的流量

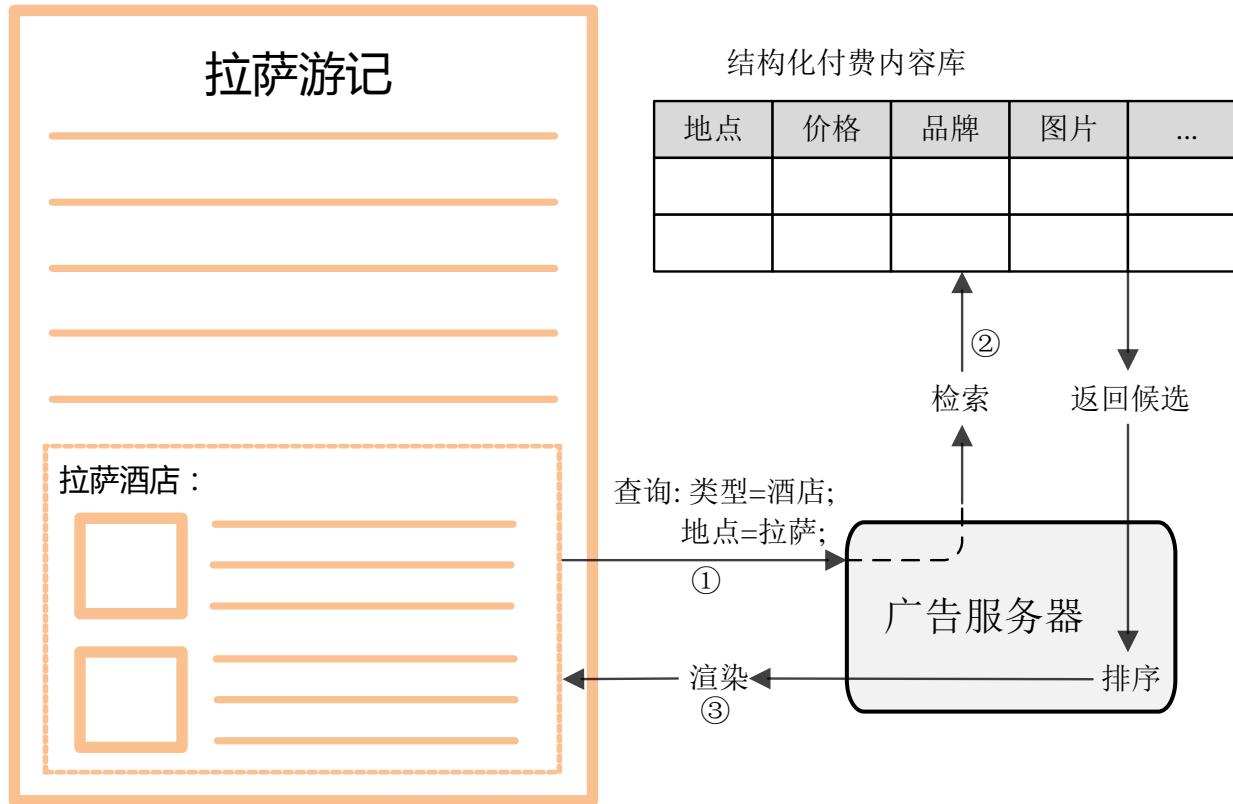


- 移动广告的产品形态需要一次革命

原生广告平台应该是什么样的？

- 媒体控制广告展示形式:
 - 原生广告的最重要的产品原则，是即内容与广告的展示形式要尽可能一致，而没有媒体的参与是不可能的
- 媒体明确提供广告需求:
 - 搜索广告的效果好，因为其投放是按照内容结果的展示原则进行的
- 广告平台提供结构化付费内容:
 - 核心的定向条件检索 + eCPM排序功能仍然没变

植入式原生广告

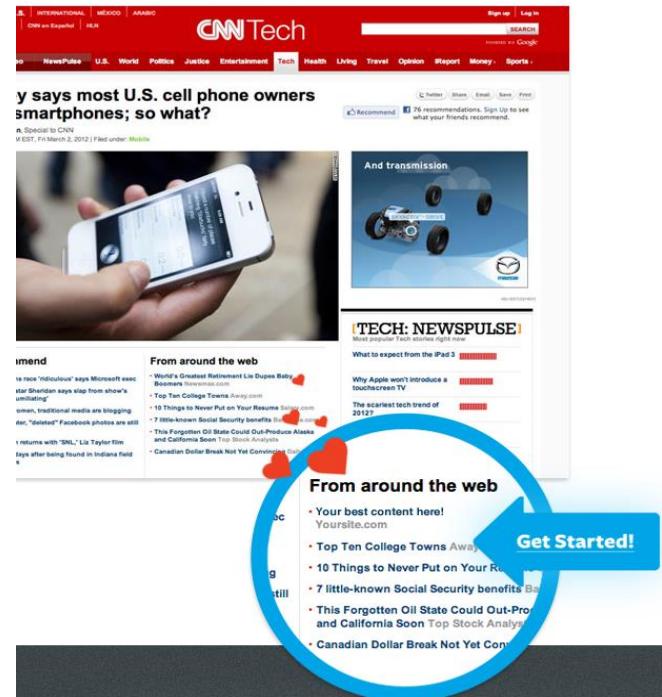


InMobi 原生广告形式举例



Outbrain 内容发现平台

- 网站将内容以RSS feed方式提供给Outbrain并给出点击单价，Outbrain将这些付费内容在其他网站上推荐给读者
- 推荐与广告的边界模糊化
- 将内容也按照eCPM最大的方式排序，这一思路也可以被站内的推荐系统，即流量塑形





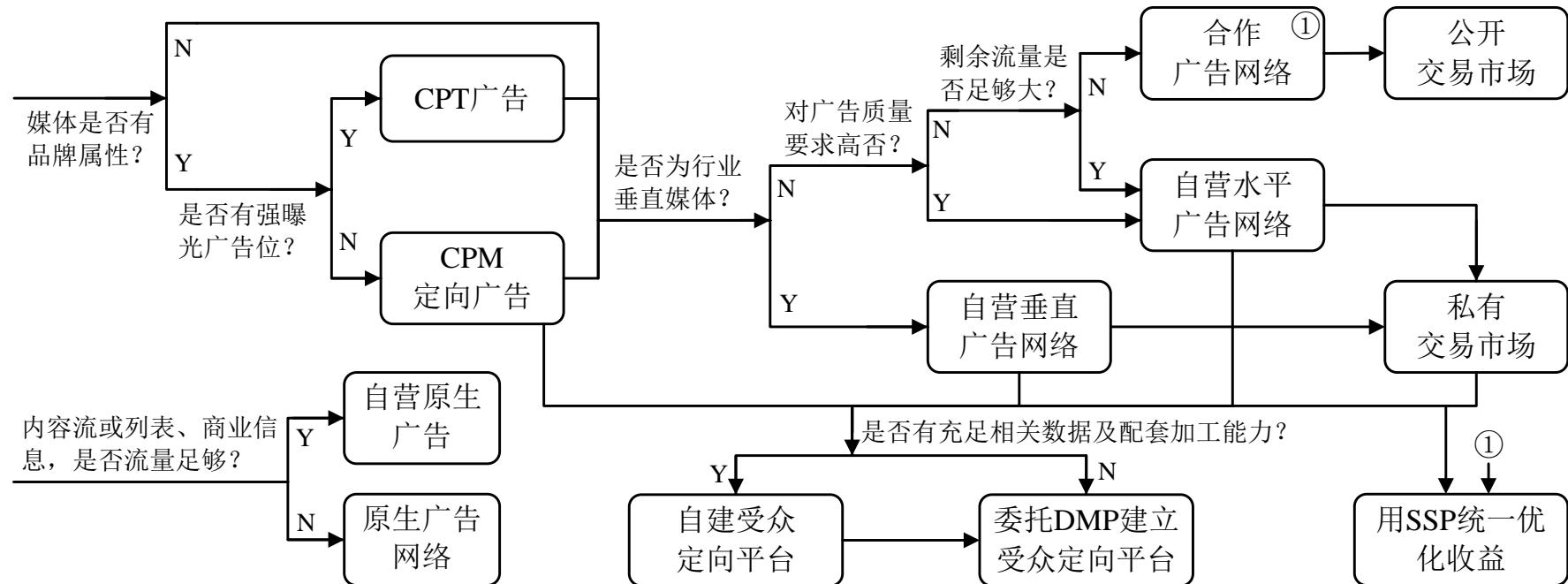
- 全球最大的移动App推广联盟（Mobile Affiliate Network）
- 采用跟广告主CPA结算的方式，在App中嵌入SDK以跟踪转化
- 主要面向海外市场，中国由于没有Google Play，移动应用市场较分散，此类业务不易开展。
- 并非严格意义上的广告技术公司，与多种异质渠道合作，核心竞争力是面向多个推广渠道时的优化能力。
- Banner并非广告业务的本质！

原生广告方向与挑战

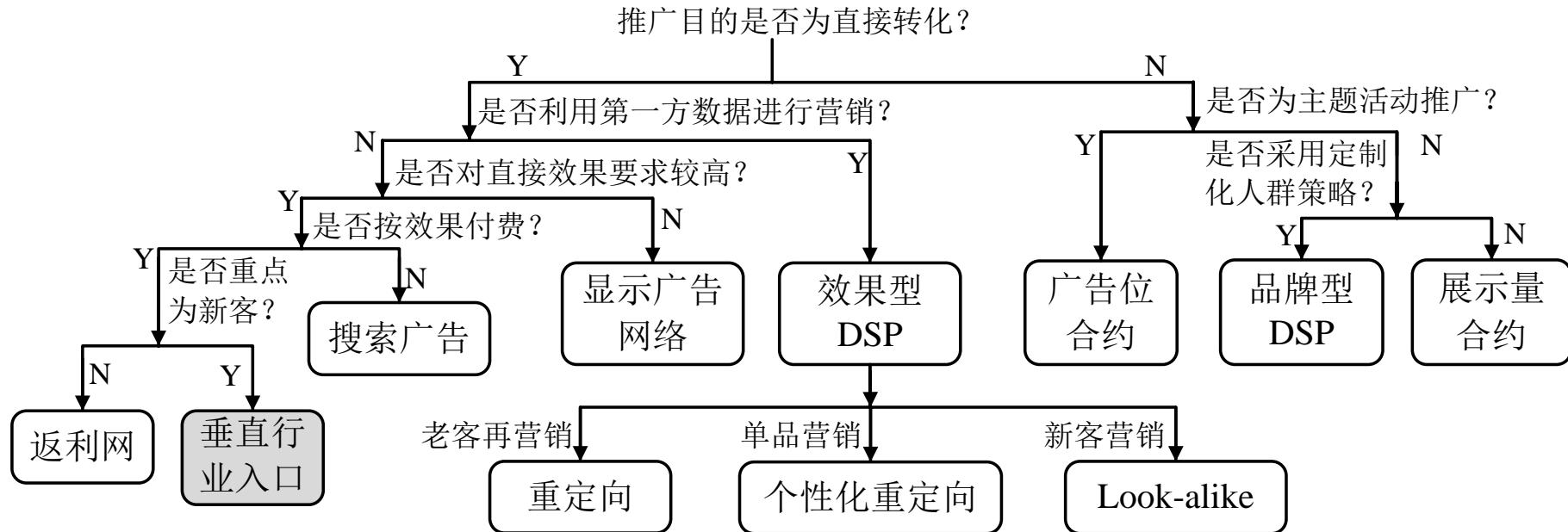
- 方向：
 - 让媒体深入参与广告交易，而不是像传统广告系统那样只在平台和需求方之间交易。
 - 将大数据驱动的受众定向展示广告，变成以媒体内容为导向的，不触及用户隐私的精准广告
 - 让高相关的付费内容代替广告成为媒体变现主要方式
- 挑战：
 - 媒体参与让广告多了个自由度，运营难度增大
 - 大量分行业、结构化广告信息的建立需要时间

在线广告产品技术实战

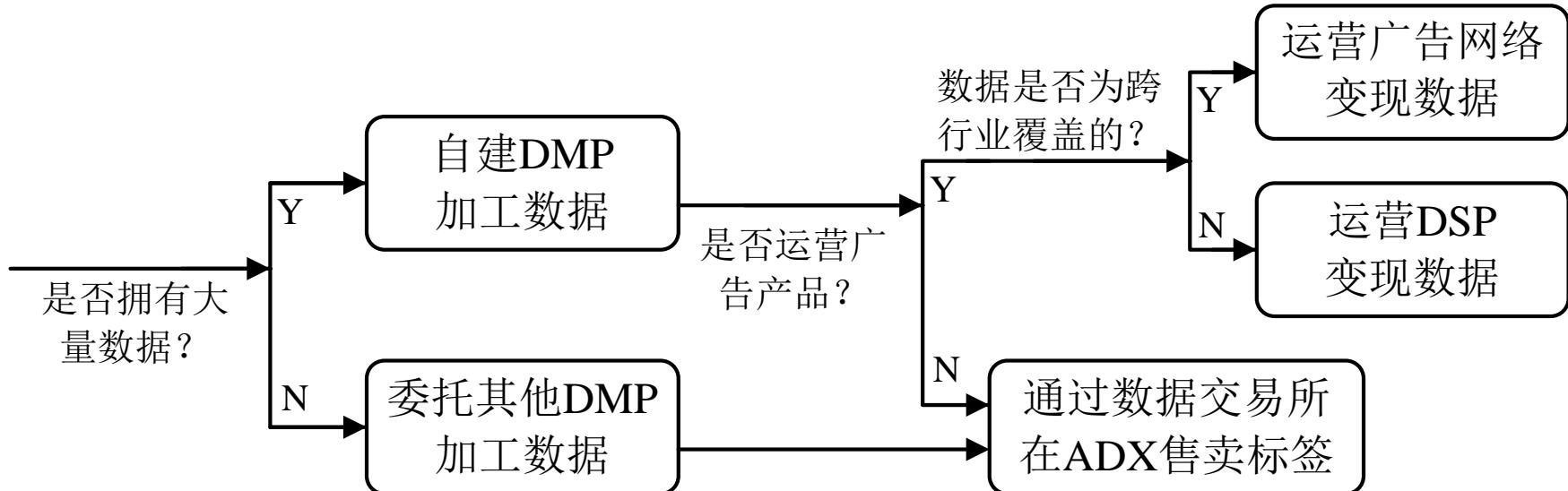
媒体方实战



广告主实战



数据方实战



广告系统有什么特点？

- 高并发、低延迟的要求
 - 同时关注QPS和Latency
- 数据处理的规模很大
 - (用户、环境、信息)三元组上的数据建模
- 数据处理的速度优先于精度
- 主流程的一致性要求不高

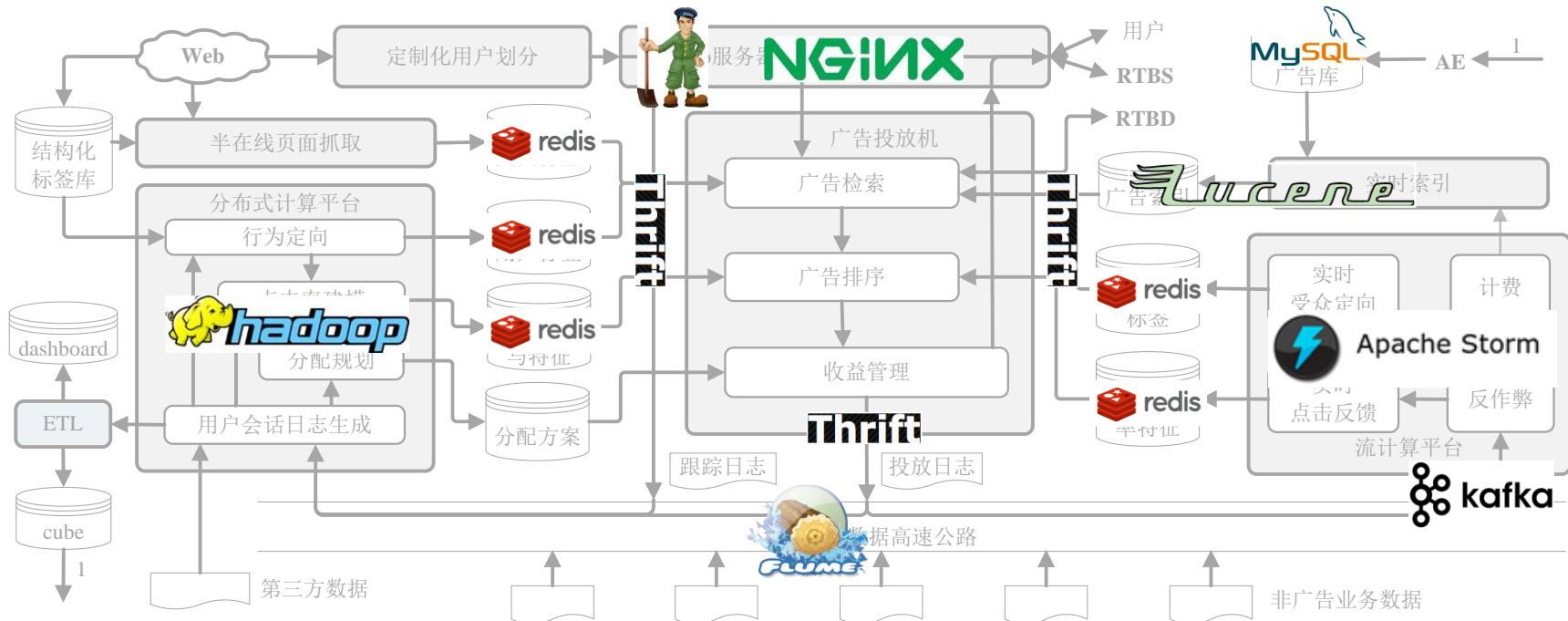
广告系统有什么设计原则？

- 建立弱一致性系统的设计思维方式
 - 例：Near-line page fetcher
- 大量数据尽量环形单向流动
 - 避免集中读写形成的单点性能瓶颈
- 在线时不要发生与关系型数据库的交互
 - 增删改查的传统需求实际上不存在
- 充分利用开源社区的成熟技术

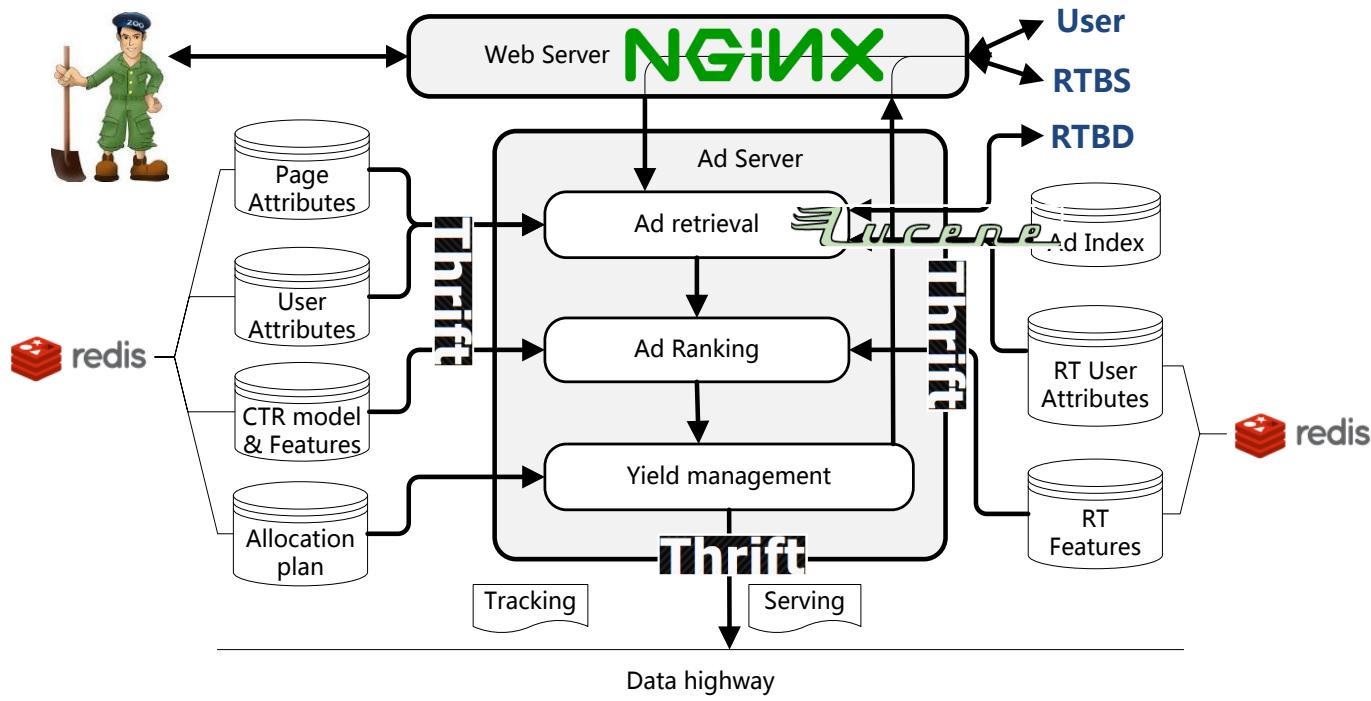
开源软件的优势与顾虑

- 优势
 - 大量细分使用场景都有开源方案
 - 大型互联网公司的开源产品经过充分测试
- 顾虑
 - 需要仔细甄别好的和不太好的开源项目
 - 在遇到深层次bug时无能为力
- 核心业务逻辑不应选择开源

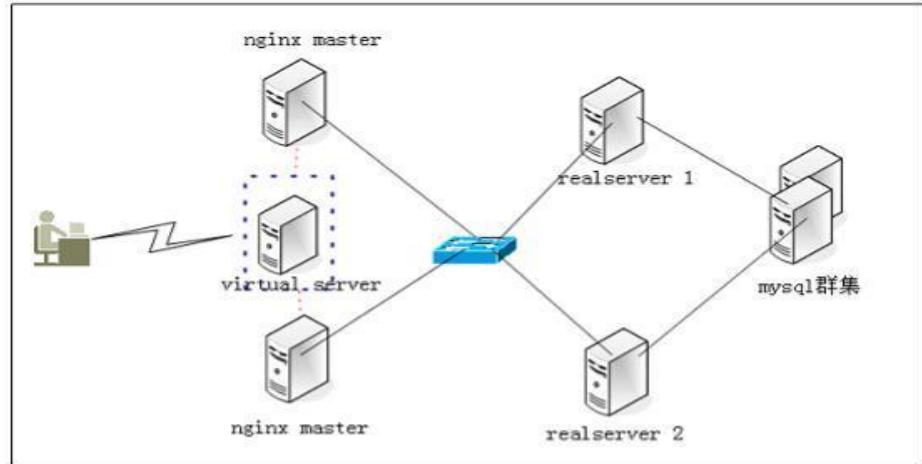
个性化系统的开源架构



在线投放引擎

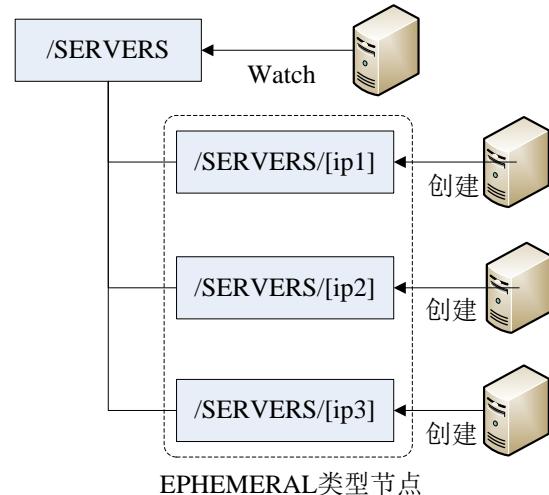


- 轻量级Web服务器 / 反向代理服务器
 - 专为性能优化而开发，可支持数万并发
 - 内存和CPU占用较少
- Nginx + FastCGI
 - 将主要业务逻辑在FastCGI中用C语言实现
- 其他产品选择：Apache



Zookeeper by YAHOO!

- 解决分布式应用中的一些数据管理问题
 - 集群管理、分布式应用配置项的管理等
 - 理论基础是著名的Paxos算法
- 集群管理
 - 将宕机、网络断链、新机器加入等迅速通知每台Server并重新分配流量
 - 用Zookeeper的Watch机制很容易实现



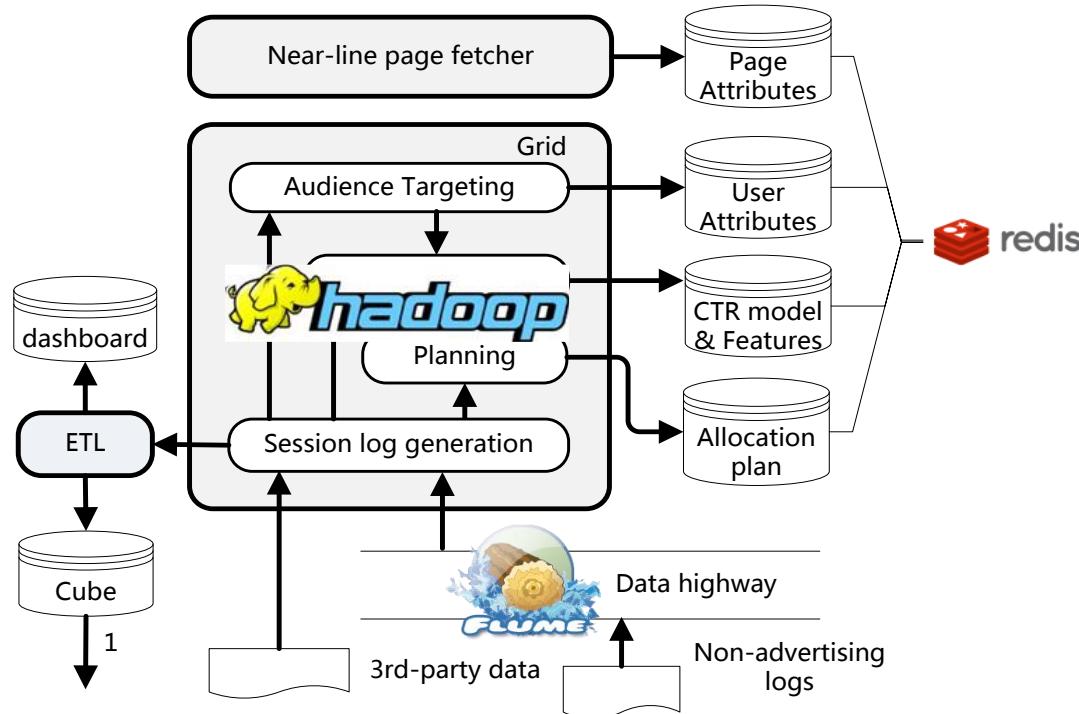


- 开源的全文检索引擎
 - 包括Index和Retrieval模块
 - 提供了简单易用的API，可以快速实现大量文档的检索
 - 采用分块索引提升Index响应速度
 - 实现了布尔查询、模糊查询等
- 不支持相关性检索
- 其他产品选择：Elastic Search



- 跨语言服务间通信
 - 不同的开源工具适合的、或不同程序员擅长的编程语言不同
 - 希望不用实现底层通信细节，专注于业务逻辑
- Thrift
 - 支持多种语言间的跨进程、跨机器通信
 - 用Thrift IDL定义通信接口
 - 某个接口版本更新后可以兼容旧的服务
- 其他开源产品：ProtoBuf

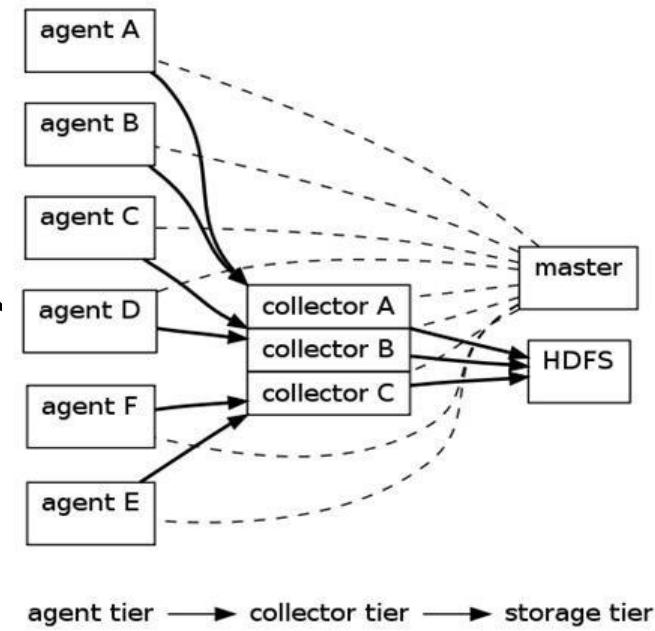
离线大数据平台





by **cloudera**

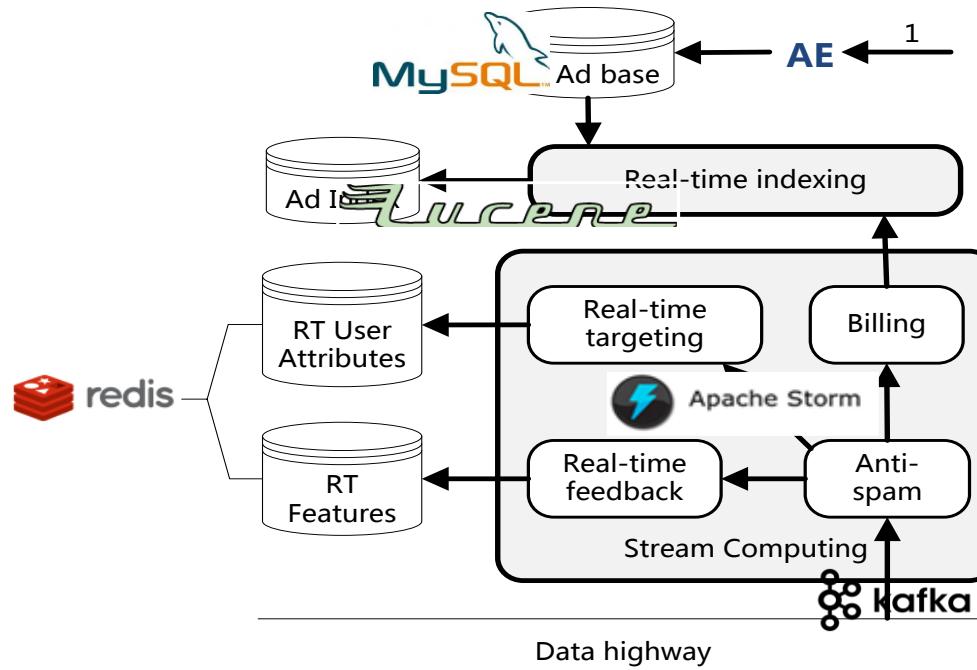
- 高可用的，高可靠的，分布式的海量日志采集、聚合和传输的系统
 - 数据源：console、text、Thrift-RPC、tail、syslog
 - 数据目的：console、dfs、Thrift-RPC、syslog TCP
- 其他产品选择
 - Scribe、Chukwa、KafKa





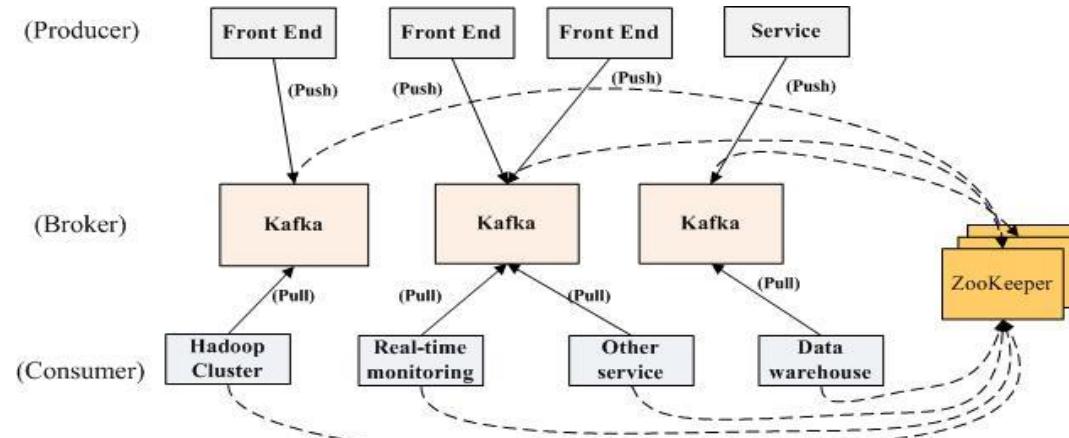
- 一个KV存储系统
 - Value可以支持string、list、set、zset、Hash
 - 支持push/pop、add/remove等操作
 - 高性能，使用简单，提供多种语言API
- 注意事项
 - 避免让批量写影响在线服务中的读性能
 - 尽量用自更新的方式减少数据输入
- 其他产品选择：Memcached

在线流计算平台





- 分布式消息系统, 数据消费方支持pull模式
- $O(1)$ 复杂度的持久化和很高的吞吐率
- 用Zookeeper来管理brokers

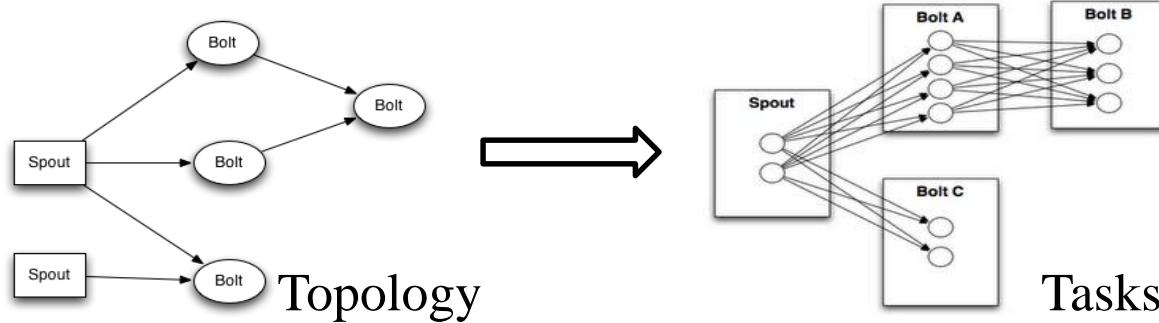




Apache Storm

by twitter

- 大规模实时数据处理框架, 自动完成数据分发和可靠性管理, 开发者只要关注处理逻辑.

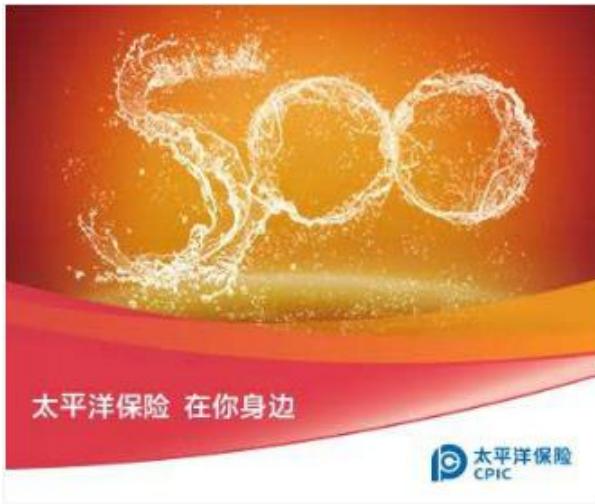


- “Storm on YARN” 可以与Hadoop共享计算资源
- 其他产品选择
 - S4 , PUMA , Spark Streaming

其他广告相关技术

创意优化

- 品牌性创意与效果性创意



- 在相近的广告诉求下，通过创意内容优化效果

创意优化主要思路

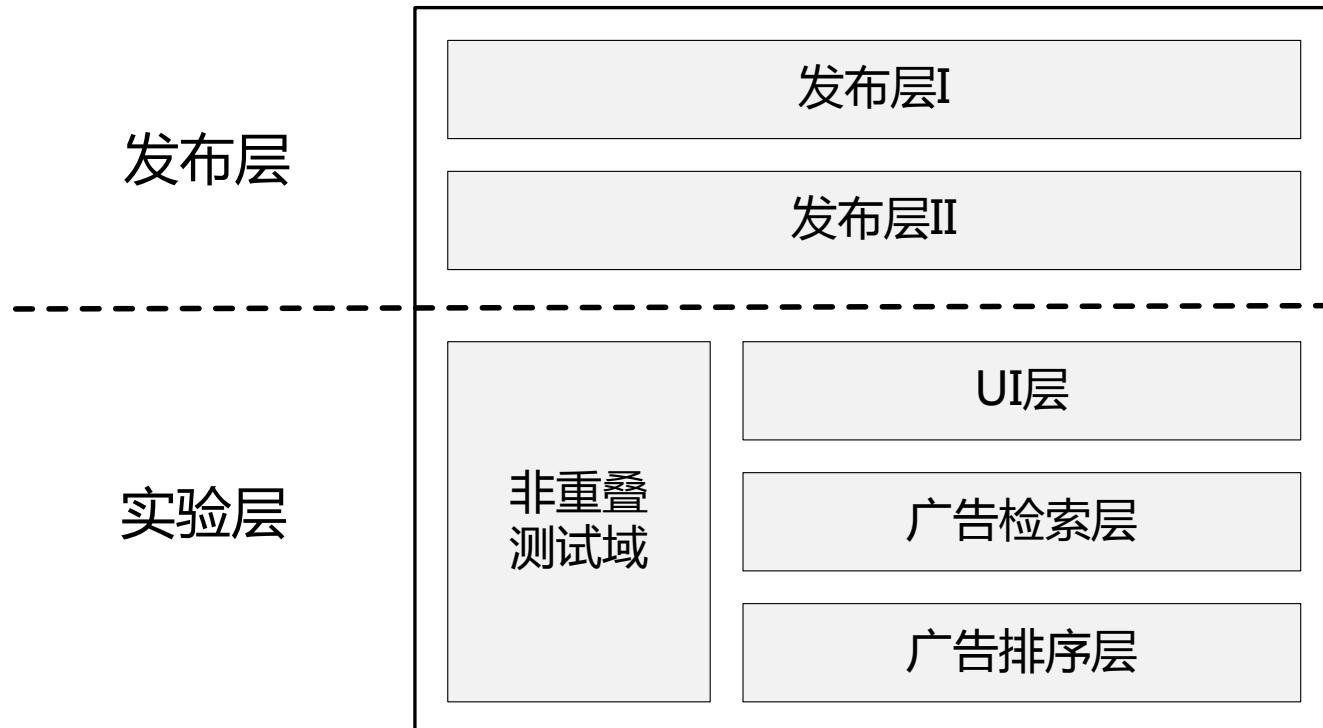
- 程序化创意
 - 网站重定向广告加入广告主 logo
 - 搜索重定向广告加入搜索框
 - 地域定向明示地点或本地电话
- 充分利用点击热点图
 - 如右所示，是否创意中的关键诉求确实吸引了用户的注意？



实验框架(Experimentation Framework)

- 哪些场景需要线上A/B测试?
 - 广告系统的升级
 - 广告算法的改进
 - 创意或定向策略的优化
- 实验框架的设计原则
 - 任何一组对比实验，其他的条件都应该相同
 - 线上流量容纳尽可能多的实验

分层实验框架



作弊目的和方式

- 媒体作弊
 - 将广告展示和点击代码放在非约定位置上
 - 非用户自然行为产生的流量
 - 将广告位与内容靠近来骗取点击的方式
- 广告平台作弊
 - 混入劣质流量、制造虚假点击
- 广告主竞争对手作弊
 - 通过技术手段大量消耗广告主预算
 - 频除cookie，改变IP，甚至通过木马控制多台电脑

点击热力图检查作弊



正常点击热力图



作弊点击热力图

流量劫持

- 信道弹窗
 - 在用户上网时强行向下行内容中插入弹窗广告创意
- 创意替换
 - 将某些网站上的广告创意直接替换为其他创意
- 搜索结果重定向
 - 将搜索流量导向其他搜索引擎，获得其收入分成
- 落地页来源劫持
 - 用户访问<http://mkt.mbaobao.com/a-hotalfshell1219>
 - 将其改成[http://mkt.mbaobao.com/a-hotalfshell1219?utm\ source=*](http://mkt.mbaobao.com/a-hotalfshell1219?utm_source=*>)

广告监测

- 产品目的
 - 广告主委托第三方公司，对实际发生的展示或点击数目进行核对
 - 主要适用于按CPM结算的合约广告
 - 竞价广告对此需求不大
- 难点
 - 人口属性、地域等的监测
 - 样本与总体之间的校正

广告安全

- 广告投放验证(Ad verification)
 - 确认品牌的安全性，防止广告出现在对品牌有伤害的页面上。（例：优衣库出现在钓鱼岛新闻上）
 - 页面内容不符合品牌诉求，则展示与品牌无关创意
 - 与广告监测不同，重点在于阻止不恰当展示的发生
 - 可以Pre-bid进行
- 可视性(Viewability)验证
 - 验证广告展示的曝光程度
 - 判断浏览器是否对广告创意发生了渲染过程

大数据隐私安全

- 隐私安全基本原则
 - A29: 欧盟负责隐私保护条例制定的委员会
 - A29原则
 - Personal Identifiable Information (PII) 不能使用
 - 用户可以要求系统停止记录和使用自己的行为数据
 - 不能长期保存和使用用户的行为数据
- Quasi-identifier 与 K-anonymity
 - Quasi-identifier: 朝阳区, 35岁, 在360上班
 - K-anonymity: 北京市, 30-40岁, 互联网行业

Ad Choices

OVERSTOCK CLEARANCE

AdChoices 

18 Piece Set Android 4GB Tablet PC  Price \$59 Save 76%	Dawgs Women's 9" Cozy Boots  Price \$19 Save 76%
---	---



Why Did I Get This Ad?

Based on your online experience for consumers, some of the ads you may receive from AOL Advertising and its networks are tailored to your interests based on your search behaviors/visits on this device.

Who Provided This Ad?

This ad has been brought to you by [AOL Advertising](#)

Where Can I Learn More About AOL Advertising?

The practices of AOL Advertising, including advertising on our network of websites, as well as advertising on other AOL properties, across other sites, products, and services that are offered by AOL Advertising are detailed in our [AOL Privacy Policy](#).

How to Opt-Out

[Opt-Out](#)

Clicking the Opt-Out button above will opt you out of behavioral advertising delivered by the AOL Advertising Network.

Please note that you will continue to receive ads; however, we will no longer tailor ads to your preferences and usage.

大数据隐私安全

- 稀疏行为数据的新挑战
 - 从一个人观影或购物记录，能否反推他是谁？
 - 实际案例：Netflix推荐大赛，有人从数据集里发现了自己的同事是同性恋
 - 理论研究：Robust De-anonymization of Large Sparse Datasets
- 深度个性化系统也有隐私安全风险！
 - 相关研究课题是差分隐私(Differential Privacy)
- 隐私是大数据头上的达摩克里斯之剑

需求方数据安全问题

