

# 在线广告与计算广告

## 基础概念

- 广告，即广而告之，是为了某种特定的目的，向公众传递信息的宣传手段。
- 在线广告，也称为网络广告、互联网广告，顾名思义，指的是在线媒体上投放的广告。
- 计算广告
- 2008年，第十九届ACM-SIAM学术讨论会上，雅虎研究院资深研究员 Andrei Broder首次提出了 计算广告学(Computational Advertising)的概念，他认为，计算广告学是一门由信息科学、统计学、计算机科学以及微观经济学等学科交叉融合的新兴分支学科。
- 研究在线广告中的计算问题
- 在线广告为什么需要计算
- 技术型投放:准确接触目标受众
- 数据驱动:大规模、自动化利用数据改善产品

## 大数据与广告的关系

- 确定目标函数的数据处理问题
- C类:采样显著降低数据处理的复杂程度，同时解决问题的效果(即目标函数)没有太大的下降
- 传统数据处理问题，统计报表，报告
- A类:不可能通过处理一小部分数据来达到全量数据所能达到的效果，数据采样率的下降，收益会快速下降
- 大数据问题:个性化推荐、广告
- B类:数据达到一定规模之后，效果提升不明显
- Topic Model
- 计算广告是典型的大数据应用
- 很长一段时间里，唯一充分商业化和规模化的大数据应用

## 广告的定义与目的:

- 广告由已确定的出资人通过各种媒介进行的有关产品(商品、服务)的，通常是有偿的、有组织的、综合的、劝服性的非人员的信息传播活动
- 广告活动的主动参与方
  - 出资人和媒体，即需求方和供给方

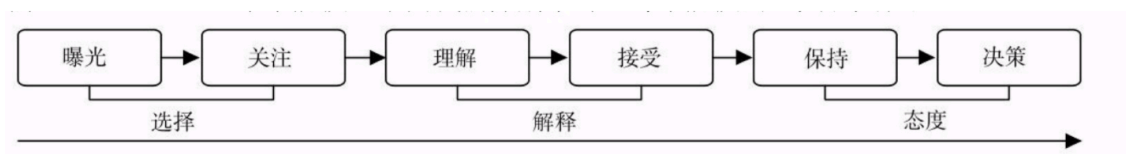
- 被动的参与方
  - 受众方
- 出资人、媒体、受众三者的利益博弈关系是广告活动永远的主线
- 目的
  - 接触大量用户，宣传品牌形象，提升中长期购买率与利润 => 品牌广告
  - 短期带来大量购买或其他转化行为 => 效果广告

## 在线广告创意类型

- 横幅广告
- 文字链广告
- 富媒体广告: 弹出式广告
- 视频广告
- 社交广告
  - 朋友圈广告
- 移动广告
- 邮件定向营销广告

## 广告有效性原理

- 选择: 曝光、关注
- 解释: 理解、接受
- 态度: 保持、决策



## 计算广告的技术特点

1. 技术和计算导向
2. 效果的可衡量性
3. 创意和投放方式的标准化
4. 媒体概念的多样化

# 计算广告的核心问题

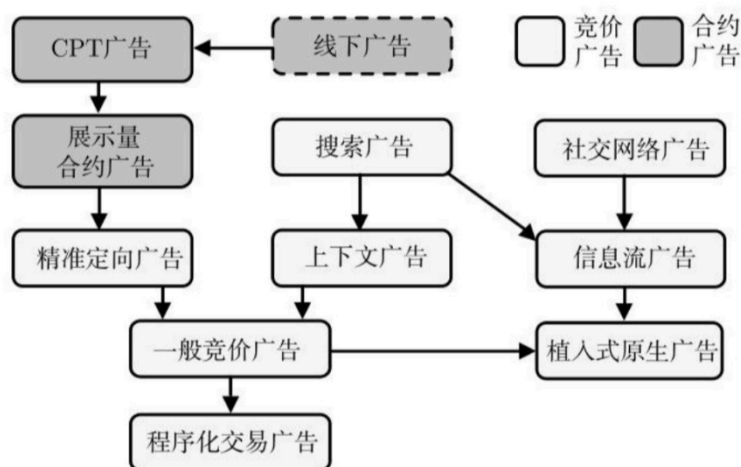
— 计算广告的核心问题，是为一系列用户与环境的组合找到最合适的广告投放策略以优化整体广告活动的利润。

— 表达式中的  $a$ 、 $u$ 、 $c$  三个变量，分别代表广告、用户与环境，即广告活动的三个参与主体，显然，广告展示的收入或成本与这三个因素都有关系

$$\max \sum_{i=1}^T (r_i - q_i)$$

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

# 在线广告产品进化



# 商业产品的设计原则

— 设计原则

—— 用户产品的设计原则总是朝着更简单、更直观、更快捷的方向努力。而相应的产品

设计重点也集中在关键功能的突出、操作过程的流畅等方面。

—— 商业产品一般都有一个明确的商业目标，而商业产品的使用者选用一款产品的动力

也是为了优化这个商业目标

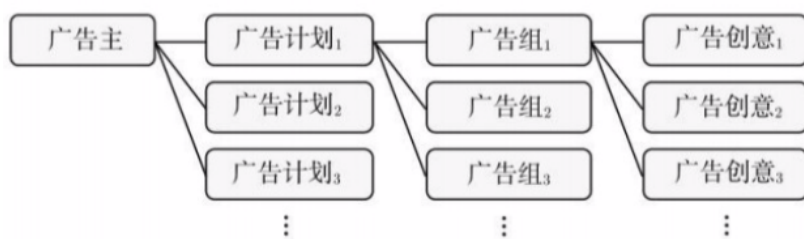
— 关键点

—— 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

## 需求方层级组织与接口

- 需求方提供的广告是分层次管理的
- 在市场上大多数的产品中，广告的层次：
  - 广告主
  - 广告(推广)计划(campaign)
  - 广告(推广)组(ad group)
  - 广告创意(creative)



## 供给方管理接口

添加、删除广告位以及查看各广告位的运营数据是主要的功能需求

广告位名称	状态	尺寸	创建日期	显示数	点击数	点击率	收入
	启用	65x65	2014-11-28	238,602,848	199,153	0.0835%	¥68,692.58
	启用	65x65	2014-11-28	0	0	0.0000%	¥0.00
	启用	65x65	2014-11-28	376,409,825	704,798	0.1872%	¥242,941.18
	启用	63x63	2014-11-28	1,331,921,299	944,687	0.0709%	¥321,429.39
	启用	64x64	2014-09-25	960,408,949	4,094,441	0.4263%	¥1,261,196.91
	启用	64x64	2014-07-18	127,600	104	0.0815%	¥34.76
	启用	64x64	2014-06-10	237,264	185	0.0780%	¥61.19
	启用	64x64	2014-06-10	295,444	266	0.0900%	¥91.78
	启用	64x64	2014-06-10	362,509	352	0.0971%	¥111.33
	启用	64x64	2014-06-10	435,934	447	0.1025%	¥147.65
	启用	64x64	2014-06-10	530,349	526	0.0992%	¥169.06

图 3-6 媒体广告位管理示意

## 在线广告产品分类

# 合约广告

- 概念

- CPT, Cost Per Time

- CPM, Cost Per Mille

- in Latin, French and Italian, mille means one thousand

- 广告位合约

- 展示量合约

- 受众定向

- 流量预测、塑形、分配

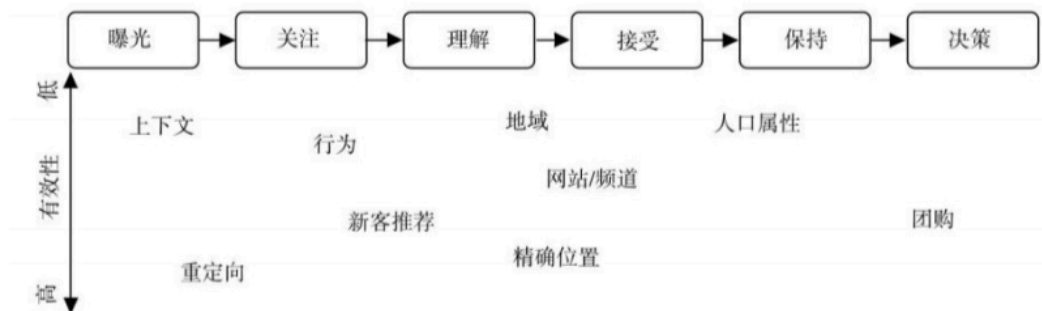


图 4-2 常见受众定向方法一览

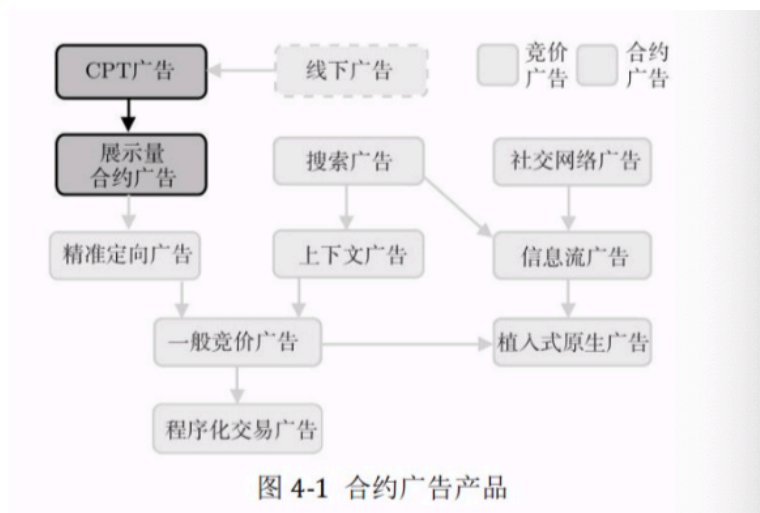


图 4-1 合约广告产品

## 产品案例

—百度信息流之合约投放篇

**Bai 信息流** | ①合约广告 | 百度首页 (APP/WAP)

Guaranteed Delivery

广告样式: 单图/大图/三图/下载/视频/图集/直播/电话/LBS/全景/随心互动

监测情况: 支持第三方曝光点击监测

**CPT** Cost Per Time, 即按时长计费广告

购买方式: 按天购买

展现位置: 推荐页信息流7楼&14楼

流量划分: 通投流量 (1-4刷) / 时段流量 (第1刷) / 地域流量 (第2刷)

**CPM** Cost Per Mille, 千次展示成本

购买方式: 按CPM购买 (1000CPM起售)

展现位置: 1. 推荐页: 百度APP 7楼&14楼

2. 详情页: 图文详情页底部banner、视频详情页后贴片

3. 百度app垂类频道 (汽车、旅游、居家、买呗): 频道feed广告-第5楼、频道金刚位

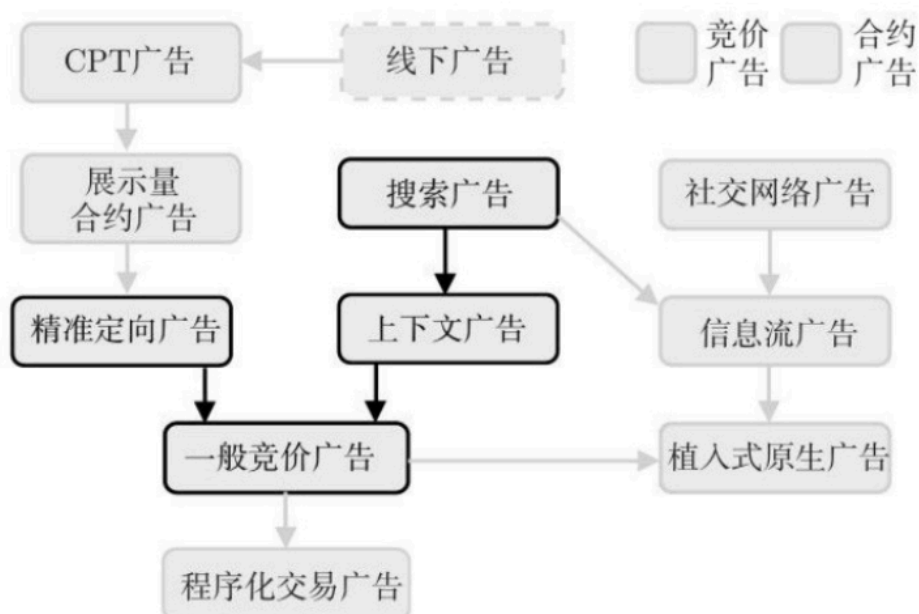
4. 百度通用频道 (热点、军事、体育、时尚、国际、本地、视频、好看6个频道): 百度app底部bar入口好看视频频道页4楼

定向方式: 地域 (核心城市、一线城市、二线城市) / 时段/网络环境/操作系统/人群属性 (性别、年龄、学历、兴趣)

百度首页



## 搜索与竞价广告



- 搜索广告几乎全是竞价广告, CPC结算, Cost Per Click
- 特点:
  - 变现效率高: 搜索广告的变现能力, 即 eCPM 远远高于一般的展示广告
  - 精准定向: 由于搜索词非常强地表征着用户的意图, 搜索广告可以进行非常精准的定向
  - 搜索广告的展示形式与自然结果的展示形式非常接近

- 竞价交易模式:根本上改变了广告的运营方法，并为其效果的快速提高释放出巨大的生产力

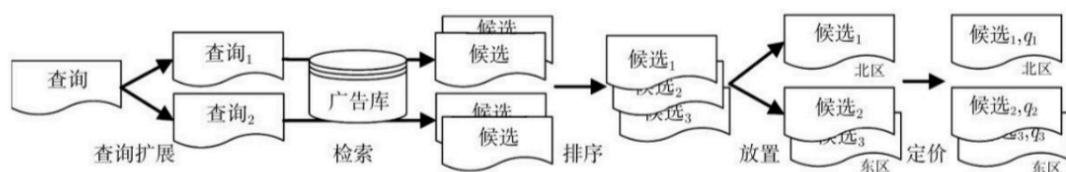


图 5-5 搜索广告决策过程示意

## 产品案例

—Baidu, Google, Tabao,.etc

## 信息流广告

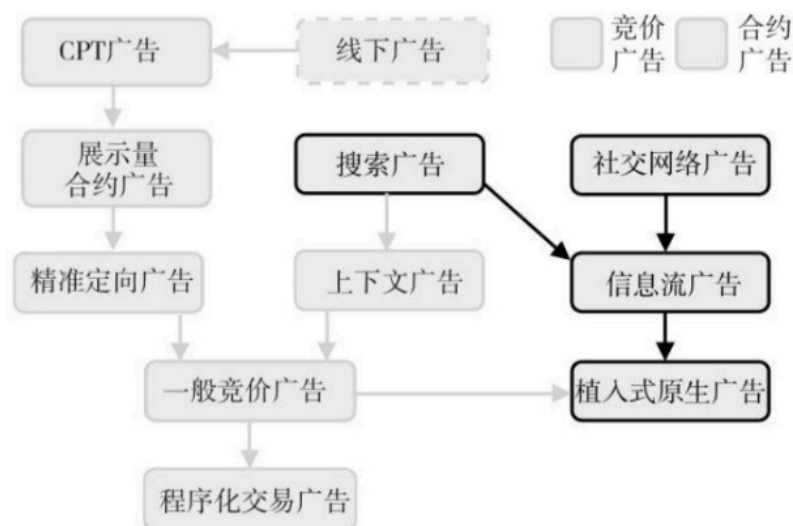


图 7-1 原生广告产品

### • 兴起

#### • 移动+社交

• 国外以Facebook为代表 • 国内以微信为代表

#### • 个性化推荐、短视频 • 今日头条

• 快手

#### • 特点

• 产品/售卖形式丰富

• 流量大/多样性强

• 可以承载几乎所有的广告类型，交易形式

# 程序化交易广告

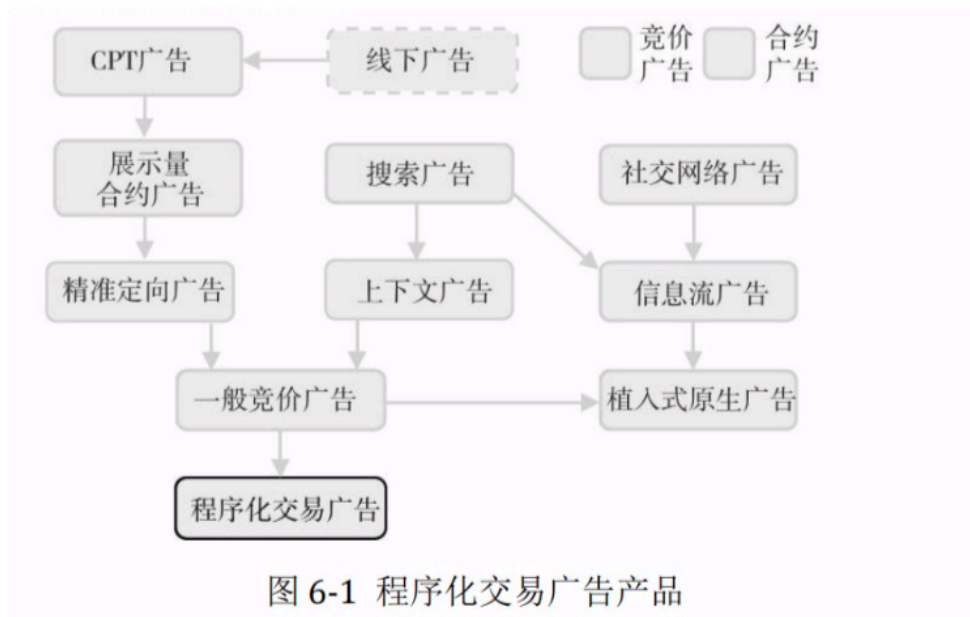


图 6-1 程序化交易广告产品

- 以实时竞价即 RTB 为核心的程序化交易市场
- RTB 的产生，使得广告市场向着透明的比价平台的方向发展，这样的平台就是广告交易平台，即 ADX，
- 其主要特征即是用 RTB 的方式实时得到广告候选，并按照其出价简单完成投放决策。
- RTB, Real Time Bidding
- ADX, Ad Exchange
- DSP, Demand-Side Platform

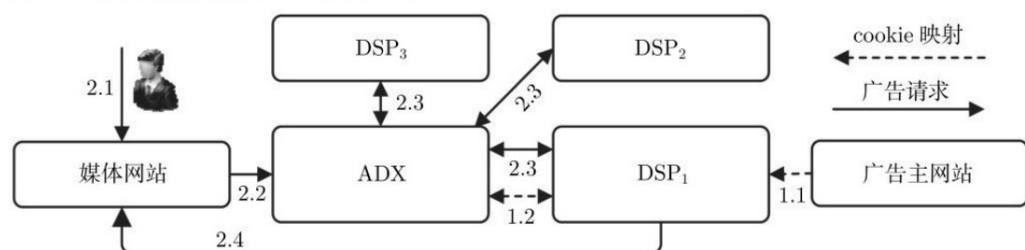


图 6-2 实时竞价（RTB）过程示意

## 文本挖掘之倒排索引与TF-IDF

**Example for inverted index:**



- 很多人都有Shakespeare's Collected Works(《莎士比亚全集》)这本大部头的书。假定你想知道其中的哪些剧本包含 Brutus 和 Caesar 但不包含 Calpurnia。
- 一种办法就是从头到尾阅读这本全集，对每部剧本都留心它是否包含 Brutus 和 Caesar且同时不包含Calpurnia。这种线性扫描就是一种最简单的计算机文档检索方式。
- 很多情况下只采用上述扫描方式是远远不够的，我们需要做更多的处理：
  - 大规模文档集条件下的快速查找
  - 有时我们需要更灵活的匹配方式
  - 需要对结果进行排序
- 倒排索引的数据结构

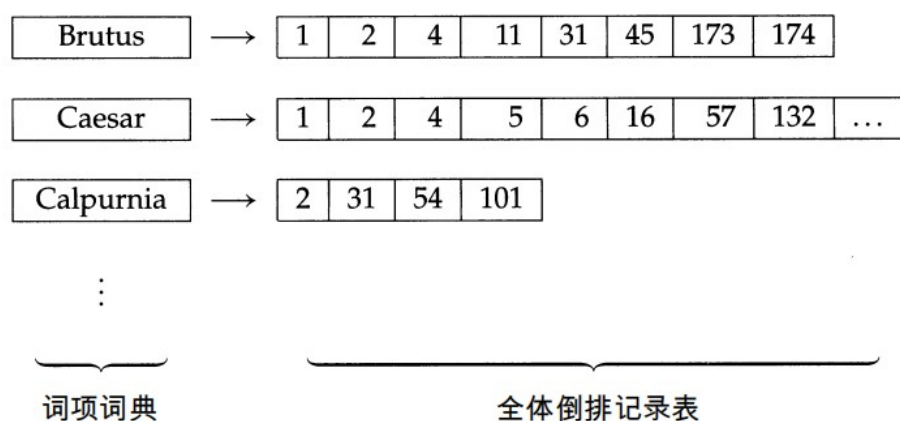


图 1-3 倒排索引的两个部分。词典部分往往放在内存中，而指针指向的每个倒排记录表则往往存放在磁盘上

How to implement inverted index:

## 词项集合的确定

—词条化 (Tokenization)

输入 : Friends, Romans, Countrymen, lend me your ears;

输出 : 

Friends	Romans	Countrymen	lend	me	your	ears
---------	--------	------------	------	----	------	------

—去停用词 (stopping words)

a	an	and	are	as	at	be	by	for	from
has	he	in	is	it	its	of	on	that	the
to	was	were	will	with					

图 2-5 Reuters-RCV1 语料库中的 25 个停用词

一词项归一化，是将看起来不完全一致的多个词条归纳成一个等价类，以便在它们之间进行匹配的过程。

查询词项	文档中应当匹配的词条
Windows	Windows
windows	Windows, windows, window
window	window, windows

图 2-6 能够对用户期望进行有效建模的不对称扩展的例

一词干还原和词形合并

am, are, is => be

car, cars, car's, cars' => car

## TF-IDF

首先，我们对于词项  $t$ ，根据其在文档  $d$  中的权重来计算它的得分。最简单的方式是将权重设置为  $t$  在文档中的出现次数。这种权重计算的结果称为词项频率 ( term frequency )，记为  $tf_{t,d}$ ，其中的两个下标分别对应词项和文档。

实际中，一个更常用到的因子是文档频率 ( document frequency )  $df_t$ ，它表示的是出现  $t$  的所有文档的数目。这是因为文档评分的目的是区分文档，所以最好采用基于文档粒度的统计量

由于  $df$  本身往往较大，所以通常需要将它映射到一个较小的取值范围中去。为此，假定所有文档的数目为  $N$ ，词项  $t$  的  $idf$  ( inverse document frequency, 逆文档频率 ) 的定义如下：

$$idf_t = \log \frac{N}{df_t} \quad (6-7)$$

## 排序

$$Score(q, d) = \sum_{t \in q} tf-idf_{t,d} \circ$$

对于每篇文档中的每个词项，可以将其 tf 和 idf 组合在一起形成最终的权重。tf-idf 权重机制对文档  $d$  中的词项  $t$  赋予的权重如下：

$$\text{tf-idf}_{t,d} = \text{tf}_{t,d} \times \text{idf}_t。 \quad (6-8)$$

换句话说， $\text{tf-idf}_{t,d}$  按照如下的方式对文档  $d$  中的词项  $t$  赋予权重：

- (1) 当  $t$  只在少数几篇文档中多次出现时，权重取值最大（此时能够对这些文档提供最强的区分能力）；
- (2) 当  $t$  在一篇文档中出现次数很少，或者在很多文档中出现，权重取值次之（此时对最后的相关度计算作用不大）；
- (3) 如果  $t$  在所有文档中都出现，那么权重取值最小。