

COMPUTATIONAL ADVERTISING

计算广告

互联网商业变现的市场与技术

刘鹏 王超 著



中国工信出版集团



人民邮电出版社
POSTS & TELECOM PRESS

还在督促自己每天进步一点吗？

还在坚持每天阅读的习惯吗？

还在为找不到自己喜欢的书籍烦恼吗？

那～

你愿意与我成为书友吗？

国内外当下流行书籍

各图书销量排行榜书籍

大量工具书籍

使我们受益终生的书籍

.....

海量电子版、纸质版书籍及音频课程

还有贴心的“学习管家”服务哦！

微信：shuyou055



目录

[封面](#)

[扉页](#)

[版权](#)

[对本书的点评](#)

[序一](#)

[序二](#)

[序三](#)

[前言](#)

[第一部分 计算广告关键技术](#)

[第1章 在线广告综述](#)

[1.1 大数据与广告的关系](#)

[1.2 广告的定义与目的](#)

[1.3 在线广告创意类型](#)

[1.4 在线广告简史](#)

[1.5 泛广告商业产品](#)

[1.6 延伸思考](#)

[第2章 计算广告基础](#)

[2.1 广告有效性原理](#)

[2.2 互联网广告的技术特点](#)

[2.3 计算广告的核心问题](#)

[2.3.1 广告收入的分解](#)

[2.3.2 结算方式与 eCPM 估计的关系](#)

[2.4 在线广告相关行业协会](#)

[2.4.1 交互广告局](#)

[2.4.2 美国广告代理协会](#)

[2.4.3 美国国家广告商协会](#)

[2.5 延伸思考](#)

[第二部分 在线广告产品逻辑](#)

[第3章 在线广告产品概览](#)

[3.1 商业产品的设计原则](#)

[3.2 需求方层级组织与接口](#)

[3.3 供给方管理接口](#)

[3.4 延伸思考](#)

[第4章 合约广告](#)

[4.1 广告位合约](#)

[4.2 受众定向](#)

[4.2.1 受众定向方法概览](#)

[4.2.2 受众定向标签体系](#)

[4.3 展示量合约](#)

[4.3.1 流量预测](#)

[4.3.2 流量塑形](#)

[4.3.3 在线分配](#)

[4.3.4 产品案例](#)

[4.4 延伸思考](#)

[第5章 搜索与竞价广告](#)

[5.1 搜索广告](#)

[5.1.1 搜索广告产品形态](#)

[5.1.2 搜索广告产品新形式](#)

[5.1.3 搜索广告产品策略](#)

[5.1.4 产品案例](#)

[5.2 位置拍卖与机制设计](#)

[5.2.1 定价问题](#)

[5.2.2 市场保留价](#)

[5.2.3 价格挤压](#)

[5.2.4 定价结果示例](#)

[5.3 广告网络](#)

[5.3.1 广告网络产品形态](#)

[5.3.2 广告网络产品策略](#)

[5.3.3 产品案例](#)

[5.4 竞价广告需求方产品](#)

[5.4.1 搜索引擎营销](#)

[5.4.2 媒体购买平台](#)

[5.4.3 产品案例](#)

[5.5 竞价广告与合约广告的比较](#)

[5.6 延伸思考](#)

[第6章 程序化交易广告](#)

[6.1 实时竞价](#)

[6.2 其他程序化交易方式](#)

[6.2.1 优选](#)

- [6.2.2 私有市场](#)
 - [6.2.3 广告交易方式谱系](#)
 - [6.3 广告交易平台](#)
 - [6.4 需求方平台](#)
 - [6.4.1 需求方平台产品策略](#)
 - [6.4.2 出价策略](#)
 - [6.4.3 重定向](#)
 - [6.4.4 新客推荐](#)
 - [6.4.5 产品案例](#)
 - [6.5 供给方平台](#)
 - [6.5.1 供给方平台产品策略](#)
 - [6.5.2 产品案例](#)
 - [6.6 数据加工与交易](#)
 - [6.6.1 有价值的数据来源](#)
 - [6.6.2 三方数据划分](#)
 - [6.6.3 数据管理平台](#)
 - [6.6.4 数据交易平台](#)
 - [6.6.5 产品案例](#)
 - [6.7 在线广告产品交互关系](#)
 - [6.8 延伸思考](#)
- [第7章 移动互联与原生广告](#)
 - [7.1 原生广告相关产品](#)
 - [7.1.1 信息流广告](#)
 - [7.1.2 搜索广告](#)
 - [7.1.3 软文广告](#)
 - [7.1.4 联盟](#)
 - [7.2 移动广告的现状与挑战](#)
 - [7.2.1 移动广告的特点](#)
 - [7.2.2 移动广告的创意形式](#)
 - [7.2.3 移动广告的挑战](#)
 - [7.3 原生广告平台](#)
 - [7.3.1 表现原生与意图原生](#)
 - [7.3.2 植入式原生广告](#)
 - [7.3.3 产品案例](#)
 - [7.4 原生广告与程序化交易](#)
 - [7.5 延伸思考](#)

[第8章 在线广告产品实践](#)

[8.1 媒体实战](#)

[8.1.1 变现方式和产品决策](#)

[8.1.2 数据支持方案决策](#)

[8.2 广告主实战](#)

[8.3 数据提供方实战](#)

[8.4 延伸思考](#)

[第三部分 计算广告关键技术](#)

[第9章 计算广告技术概览](#)

[9.1 个性化系统框架](#)

[9.2 各类广告系统优化目标](#)

[9.3 计算广告系统架构](#)

[9.3.1 广告投放引擎](#)

[9.3.2 数据高速公路](#)

[9.3.3 离线数据处理](#)

[9.3.4 在线数据处理](#)

[9.4 计算广告系统主要技术](#)

[9.5 用开源工具搭建计算广告系统](#)

[9.5.1 Web 服务器 Nginx](#)

[9.5.2 分布式配置和集群管理工具 ZooKeeper](#)

[9.5.3 全文搜索引擎 Lucene](#)

[9.5.4 跨语言通信接口 Thrift](#)

[9.5.5 数据高速公路 Flume](#)

[9.5.6 分布式数据处理平台 Hadoop](#)

[9.5.7 特征在线缓存 Redis](#)

[9.5.8 流计算平台 Storm](#)

[9.5.9 高效的迭代计算框架 Spark](#)

[9.6 延伸思考](#)

[第10章 基础知识准备](#)

[10.1 信息检索](#)

[10.1.1 倒排索引](#)

[10.1.2 向量空间模型](#)

[10.2 最优化方法](#)

[10.2.1 拉格朗日法与凸优化](#)

[10.2.2 下降单纯形法](#)

[10.2.3 梯度下降法](#)

[10.2.4 拟牛顿法](#)

[10.2.5 Trust-Region 法](#)

[10.3 统计机器学习](#)

[10.3.1 最大熵与指数族分布](#)

[10.3.2 混合模型和 EM 算法](#)

[10.3.3 贝叶斯学习](#)

[10.4 统计模型分布式优化框架](#)

[第11章 合约广告核心技术](#)

[11.1 广告排期系统](#)

[11.2 担保式投送系统](#)

[11.2.1 流量预测](#)

[11.2.2 频次控制](#)

[11.3 在线分配](#)

[11.3.1 在线分配问题](#)

[11.3.2 在线分配问题举例](#)

[11.3.3 极限性能研究](#)

[11.3.4 实用优化算法](#)

[11.4 延伸思考](#)

[第12章 受众定向核心技术](#)

[12.1 受众定向技术分类](#)

[12.2 上下文定向](#)

[12.2.1 半在线抓取系统](#)

[12.2.2 文本主题挖掘](#)

[12.3 行为定向](#)

[12.3.1 行为定向建模问题](#)

[12.3.2 行为定向特征生成](#)

[12.3.3 行为定向决策过程](#)

[12.3.4 行为定向的评测](#)

[12.4 人口属性预测](#)

[12.5 数据管理平台](#)

[12.6 延伸思考](#)

[第13章 竞价广告核心技术](#)

[13.1 竞价广告计价算法](#)

[13.2 搜索广告系统](#)

[13.2.1 查询扩展](#)

[13.2.2 广告放置](#)

[13.3 广告网络](#)

[13.4 广告检索](#)

[13.4.1 布尔表达式的检索](#)

[13.4.2 相关性检索](#)

[13.5 点击率预测](#)

[13.5.1 点击率预测模型](#)

[13.5.2 优化算法](#)

[13.5.3 点击率模型的校正](#)

[13.5.4 点击率模型的特征](#)

[13.5.5 点击率模型评测](#)

[13.5.6 智能频次控制](#)

[13.6 探索与利用](#)

[13.6.1 UCB 方法](#)

[13.6.2 考虑上下文的 bandit](#)

[13.7 延伸思考](#)

[第14章 程序化交易核心技术](#)

[14.1 广告交易平台](#)

[14.1.1 cookie 映射](#)

[14.1.2 询价优化](#)

[14.2 需求方平台](#)

[14.2.1 定制化用户标签](#)

[14.2.2 DSP 中的点击率预测](#)

[14.2.3 点击价值估计](#)

[14.2.4 出价策略](#)

[14.3 供给方平台](#)

[14.4 延伸思考](#)

[第15章 其他广告相关技术](#)

[15.1 创意优化](#)

[15.1.1 程序化创意](#)

[15.1.2 点击热力图](#)

[15.2 实验框架](#)

[15.3 流量保护和效果监测](#)

[15.3.1 反作弊](#)

[15.3.2 广告监测](#)

[15.3.3 广告安全](#)

[15.4 隐私保护和数据安全](#)

[15.4.1 隐私保护问题](#)

[15.4.2 程序化交易中的数据安全](#)

[15.5 延伸思考](#)

[第四部分 附录](#)

[附录A 主要术语及缩写索引](#)

[参考文献](#)

计算广告 互联网商业变现的市场与技术

刘鹏 王超 著

人民邮电出版社

北京

对本书的点评

以下点评分领域以点评人的姓氏笔画为序排列。

互联网公司管理层

在线广告市场是比较复杂的体系，它贯穿了互联网生态链的各种角色。刘鹏的著作既从商业角度介绍了在线广告，也深入到了广告的技术和算法层面，还包括对于一个大规模竞价市场在市场设计方面的相关原理和优化机会。这本书不仅是了解在线广告市场的途径，也是了解互联网商业和盈利模式设计的窗口，同时对于互联网产品设计会有很好的参考作用。

——王华（@超凡Derek），阿里巴巴副总裁，阿里妈妈负责人

这是一本非常系统、全面地介绍计算广告的书，一本在线广告专业人员必读的书，一本值得强烈推荐给想利用互联网力量的企业主和决策制定者和对大数据价值感兴趣的研究人员和工程师的书。虽然我亦曾亲眼目睹广告业在 20 世纪 90年代开始的革命，一直积极参与在线广告的演变，但阅读完这本书，我对计算广告整个图景以及很多细微之处有了更多更深刻的理解。

——毛建昌，微软Distinguished Engineer和Bing广告工程负责人，
前雅虎实验室广告科学副总裁

十多年的实践证明，互联网最有效的商业模式莫过于可以把流量直接变现的在线广告模式。从最初铺天盖地的横幅广告起步，到人群及兴趣精准定向的搜索广告与推荐引擎，直到与内容环境融为一体的原生广告，用户需求与口味的不断变迁促使着广告产品与技术持续不断地升级与发酵。本书最大的亮点在于，作者从中国互联网广告发展全过程亲历者的视角，极为系统地讲述了计算广告的产品设计思维与技术理论基础，涵盖从广告呈现到计价策略乃至算法实现并直接运用于互联网流量变现课题的方方面面。无论是产品经理还是工程师，如若准备投身于这一互联网最大的金矿领域，此书是手边必须常备的工具书与教材。

——刘子正（@刘子正），微博常务副总经理

本书内容全面且与时俱进，对核心技术的介绍深入浅出，是计算广告领域一本难得的好书。本书视野宽阔，涵盖了在线广告市场及核心技术的各个方面，除了主流技术以外，对一些其他著作很少涉及的方面，

如广告创意优化、反欺诈、隐私保护等也进行了介绍。本书内容新颖，把近年涌现出的一些新的广告形式和技术，如实时竞价的广告交易、原生广告等，都囊括其中。另外，本书行文流畅、逻辑清晰，对核心技术的介绍深入到位，包含了重要的算法细节以及理论探讨，对计算广告的从业者而言，是一本非常实用的参考书。

——刘铁岩（@刘铁岩），微软亚洲研究院首席研究员

在互联网深入改造传统行业的进程中，所谓“羊毛出在猪身上”的后向变现商业策略至关重要。市场上并不乏关于这一策略的推崇和讨论，不过本书从实战的角度出发，对其中关键的商业逻辑和产品结构作了全面的梳理，而这些对于商业化战略的落地至关重要。因此，我推荐所有正处在互联网化变革行业的从业者们阅读本书，结合各自的知识背景和行业问题，相信你们一定可以从本书中找到有指导意义的内容。

——陈彤（@老沉），小米内容运营和投资副总裁

刘鹏博士是我以前在微软时的同事，他博学睿智，融会文理，给我留下深刻印象。他将多年积累的相关经验与成果整理成国内第一部正式出版的计算广告著作，值得向大家强烈推荐。这本书将该领域的实际商业问题与技术解决方案结合起来，让读者对计算广告的理论与实践、应用与技术、系统与方法有全面深入的了解和认知。对于已经掌握了机器学习、数据挖掘技术的从业者、技术人员、教师和学生，如果想进入计算广告及其相关领域，这是一部极佳的指南。

——李航（@李航博士），华为诺亚方舟实验室主任

互联网业内人士都清楚流量的作用，流量之于互联网，正如血液之于人体。不同之处在于，互联网上的流量是趋利的，变现能力决定了流量的方向和价值。计算广告是流量变现最重要的方式之一。只有了解了互联网广告的精髓，才能真正懂得互联网上流量的奥秘，也才能懂得互联网的奥秘。本书作者对互联网广告的市场、产品和技术做了全面、深刻的剖析，为业内外人士理解并踏入这个领域提供了一条捷径。尽管本人从事互联网广告研发多年，也是第一次读到如此系统的著作，受益匪浅！

——沈抖（@沈抖），百度高级技术总监

我们的团队花了大量精力寻找和调研国内外的相关资料和文献，却一直苦于没有一套相对完整的知识体系来帮助我们的业务和技术人员加快学习、少走弯路。因此，我期待这样一本书的出现至少已经有两三年的时间。作为一门方兴未艾而发展迅猛的新兴产业，我相信刘鹏的这套科学且实际的知识体系，将对数字营销领域的同仁有非常大的指导价值。程序化营销、大数据应用是发展异常迅猛的产业，涉及大量的产

品、工程以及算法，也涉及相当多的商业逻辑的深刻洞察。而刘鹏通过对产业内在逻辑的诠释和推演贯穿程序化营销产业的发展，对我们真正把握大数据给各个产业带来的深刻变化和影响，具有深刻的指导意义。

——杨炯纬（@杨炯纬），360副总裁，聚效广告CEO

如果说当年Yahoo! 广告首席科学家Andrei Broder在斯坦福开设的课程第一次把计算广告学作为一门学科，那么这本《计算广告》堪称这门新兴学科的第一本教科书，因为这是首次全面系统地阐述计算广告学的著作，覆盖了商业逻辑、产品结构、关键技术、工程实践和应用实例。在内容结构编排上，本书由浅入深，从宏观背景到技术细节，从经典的搜索广告到最新的实时竞价，既适合作为从事在线广告的商务运营人员的参考书，也适合作为一线技术开发人员的实战指导。

——余凯（@余凯西二旗民工），百度研究院副院长，深度学习实验室主任

计算广告近年来特别热，全球大小互联网公司有大量的算法工程师、系统工程师、数据科学家在从事与此相关的工作。正如刘鹏在书中指出的，流量变现和数据变现是很多互联网公司商业模式的核心。虽然学术界和工业界有大量与计算广告相关的文章，但或侧重于算法，或侧重于系统，抑或侧重于商业逻辑，却很少能像本书一样把这几个维度融会贯通地串在一起，既有连贯性、有广度，又有足够的深度。刘鹏过去几年中花了大量的精力在清华大学、北京大学以及一些在线教育平台分享他对计算广告的深刻理解，影响了不少人，我过去和现在的团队都有他的粉丝、他的学生。现在刘鹏又把他对计算广告的深刻理解集结成书，能影响到更多的人。为刘鹏点赞！

——张小沛（@Joyce在路上），宜信CTO，前Hulu全球副总裁

计算广告学纷繁复杂，刘鹏以一个科学的实践家的态度抽丝剥茧，全面、系统地阐述了其技术架构与产品生态，为入门者普及了概念，为从业者开阔了眼界。

——罗征，腾讯广点通总经理

互联网广告在近十余年里一直保持着爆炸式的发展，支撑着谷歌、脸书、阿里、百度等数个百亿、千亿级互联网公司。或许其发展过于迅猛，或许其涉及领域过于宽泛，以至于近几年来一直没有一本优秀的书能够系统、全面地对互联网广告加以介绍。这本书把这件早就应该有人去做的事情漂亮地完成了。从业务模式到技术架构，从算法模型到工程实现，从理论基础到实现细节，从历史背景到最新动态，这本书都做了翔实、系统的讲解。相信对于每一位置身于互联网广告业的朋友，不论其在这纷繁复杂的行业里承担什么样角色，这本书真值得一读。

——顾大伟（@小米大伟），小米广告负责人

很高兴看到刘鹏博士把自己在互联网广告领域的多年经验和智慧整理成书，其中既有他对商业产品的理解，也有算法和工程实现的总结。本书不仅第一次全面梳理了互联网广告产品形态，针对每个产品描述了相应的核心算法和系统实现，而且全面描述了以媒体和广告主为核心的生态圈，以及技术如何一步步促使生态圈演化，不断创造出更大的商业价值。对于互联网广告从业者来说，相信读后一定获益良多。

——贾志峰（@贾志峰Michael），汽车之家技术副总裁

来自互联网创业者

如果你正从事或准备进入互联网广告行业，本书应该是你的职业生涯中必不可少的读物之一。无论在中国或是放眼世界，作者以更深、更广的视角向读者展示了当今互联网行业的市场与技术。在享受互联网大数据带来乐趣的同时，书中介绍的中西市场案例将助力中国互联网从业者（包括产品技术人员）开拓创新思维。本书为推动中国互联网发展赋予了更重要的意义。

——闫罍（@闫罍AdMaster），AdMaster创始人、CEO

今天，大数据浪潮正在席卷全球。数字技术正在改变我们的生活方式，同样也在驱动着商业、营销和广告业的未来。营销不再只是关于策略、创意和idea的，更和技术的发展紧紧地捆绑在一起。进入 2015年后，我们看到全行业正在迎接大数据的风口，从数据的沉积分析和管理到数据的真正打通，这是一场时代的演进。刘鹏博士的这本书正是对这个变革时代营销技术变化的实录，他对最新数字广告技术方方面面的精到剖析，不仅是对广告知识体系的实时更新，更是对未来新的数字营销体系架构的有益探索。希望今天我们在勇于探索和开拓的也正是未来被写进历史的故事。

——吴明辉，秒针系统创始人、董事长兼CEO

随着互联网的高速增长，广告开始往精细化发展，如何在有限的资源里获得最大化的广告综合收益是一个非常复杂、重要且有趣的问题，这也是计算广告研究的方向。刘鹏作为这个方向的专家，在本书中从计算广告问题的提出开始，介绍了计算广告的产品形态以及关键技术，非常适合互联网广告的从业者系统性地了解计算广告领域。

——周霖（@周霖-KCN），搜易贷联合创始人，前搜狐高级副总裁

这本书逻辑清晰，非常贴近实战，值得网络广告从业人员仔细阅读与思考。无论是媒体、广告代理还是广告主，谁能更好地获取数据、理解数据、应用数据，谁就能在日益激烈的市场竞争中脱颖而出。

——赵士路（@赵士路），WiseMedia创始人、CEO

互联网和移动互联网广告生态圈正在发生翻天覆地的变化，广告形式、产品形态、市场格局及产业链模式不断推陈出新，让人目不暇接。同时，基于营销大数据的计算广告技术也日益成熟。刘鹏博士的这本书将两者有机结合，既能帮助从业人员了解互联网广告全貌及流量变现的现状，也能帮助技术人员掌握计算广告的核心技术，是一本兼顾商业产品逻辑和技术实践的难得一见的佳作。

——唐健，智云众创始人、CEO

互联网广告是一个千亿级的市场，如果把互联网比作一辆车的话，互联网广告就是汽油，因为大多数网站都是依靠广告盈利。刘鹏博士的这本书涉及大量的基础知识、概念和商业模式，是目前此领域比较全面的一本广告技术著作。书的内容深入浅出，讲述了搜索广告、广告交易平台、广告基本算法以及开源系统等重要概念，也介绍了不少相关广告技术公司以及他们在互联网广告这个产业链上的各种典型产品，非常适合从业人员以及有兴趣进入这一阳光产业的同学学习。

——唐朝晖（@唐朝晖_adSage），艾德思奇创始人、CEO

过去5年是移动互联网发展最快的时期，开发者创造出如此多的应用和内容，用户行为习惯和数据积累发生如此惊人变化。在此过程中，在线广告作为最主要的变现形式，逐渐成为广大从业人员必须掌握的知识和技术。然而，由于此领域学习门槛较高，对于很多从业人员来说迷雾重重。大部分相关文章只是对于广告相关一些术语进行了罗列或介绍，无法让大家“知其然而知其所以然”。这本书的出版将弥补这一空白，它系统性地介绍在线广告的发展历史和逻辑，以及流行的程序化购买关键技术与算法。更为可贵的是，刘鹏在本书中融入了自己多年对于计算广告领域的理解和经验，使得整本书的思路和编排极为流畅。本书既适合想了解此领域的初学者或业务人员泛读，也适合专家以及产品人员对特定的知识点精读。本书将成为广大互联网从业人员必备的读物，特此重点推荐给大家。

——崔晓波（@崔晓波_TalkingData），TalkingData创始人、CEO

来自媒体与行业专家

本质上讲，互联网经济与广告经济都属于信息经济的具体技术形态或产业形态，核心要素是数据，经济学特征则是“所有能够传播信息的商品，其售价都会趋向其边际成本”。因此，确定数据商业化与广告产品化之间转换的逻辑、方法和路径极为重要，计算广告恰恰是这样一种经过多年实践的有效体系。刘鹏先生所著的这本书对相关的技术、创新与商业作了极好的刻画、梳理与论述。

——马旗戟（@马旗戟），原尼尔森高级副总裁

这本书于我而言，是打开了一扇窗，让我看见了在巨大的互联网广告产业后面蕴含的数学模型和算法基础。计算广告学中蕴含的各种方法让我想到了管理方法论中很著名的一句话：“If you can’t measure it, you can’t manage it!”量化的方法使得计算广告学成为计算机科学与工程的一个崭新和重要的方向。非常感谢作者的知识分享。

——陈怀临（@湾区评论），弯曲评论创始人

我有两个身份，既是从 20 世纪开始工作的广告主，同时又是大学老师，但面对的却是一样的问题和困惑。营销方法尤其是广告形式推陈出新，众多科技层出不穷，受限于自身的学问背景，不可能全部都了解。所以有拜读本书内容的机会我特别欣喜，终于有由业内专家执笔且技术含量特别高、非常实用的书了。广告主可以从中了解不同的展现方式，利用书中的广告主在线营销决策过程择善而为。媒体也可以凭借类似的广告变现决策，揭示未来的发展方向。专业人员可以进一步了解背后的技术，找出最有针对性的广告投放，提升推广成果。因此，我非常推崇本书的实用价值及参考价值。

——杨仕名（@楊仕名），香港大学SPACE中国商业学院副总监，
营销与传媒管理中心主任

广告带来的后向变现是互联网经济中核心的变现模式之一，也是互联网商业模式的重要根基，而本书对这一领域作了一次全面的总结。我们希望互联网企业、广告服务和技术公司，以及艾瑞这样的数据服务公司，以本书的出版为契机，认真探讨互联网商业模式上的分工协作，推动行业的变革与发展。

——杨伟庆（@杨伟庆），艾瑞咨询总裁

世界上有一种沟通是付费的，这就是广告的本质。然而，近几年互联网改变了整个广告生态的格局，目前世界上最叱咤风云的互联网公司几乎都依赖广告。刘鹏博士的这本书系统性地介绍了这种深层次的变化，以及整个产业链迸发出的各种技术手段与学问。此书由浅入深系统地介绍了几乎每个互联网广告的生态位置以及背后运作的机理，是我目前见过国内最系统的介绍计算广告的著作。此书对互联网、媒体、广告公司、市场营销人士，甚至消费者都是一本了解互联网广告的佳作。

——张迪（@广告技术流adexchanger），Adexchanger.cn创始人

有了互联网才有了计算广告学：计算广告学把传统的无法定向投放和无法度量的广告变得可以定向投放和可以量化度量效果。刘鹏博士在工作之余，把计算广告学的系统性知识和多年实战经验总结成书，对从事计算广告的工程师和想了解计算广告的工程师都非常有帮助。

——张栋（@张栋_机器学习），前Google研究员

对计算广告技术和商务人才的需求近两年迅速高涨，但计算广告是一个新兴交叉学科，一直缺乏全面系统的专著。这本书全面介绍了这一领域的商业背景知识、业务需求和详细的技术实现思路。本书一个重要特色是将该领域的商业挑战与技术的选择、应用、实现进行了融汇中西的系统化介绍，让不同知识背景的读者都能从中获得认识提升。此外，本书对于整个计算广告技术知识体系的梳理全面、准确，囊括了从业人士和学术研究需要关注和了解的主要知识点，对于已有一定基础和实践经验的读者也能从中温故知新和查遗补缺。此书的出版对于促进中国相关行业人才池的增长大有裨益。

——范秋华（@RTBChina），RTBChina创始人

互联互通正领跑，眼球经济网民包。创收多多靠广告，变现书籍好难找。计算广告学走俏，廿年蔚然成林梢。理论实践兼顾到，刘鹏此书及时抛。入门登堂先介绍，市场规模大蛋糕。产品技术两面刀，块块切尝大与小。搜索推广竞价搞，合约展现包推销。程序交易争分秒，移动平台憋新招。信息流起人社交，原生广告置混淆。探索利用平衡高，点击建模测验校。背景逻辑打夯牢，核心技术窥其奥。照葫芦可画出瓢，立竿见影编码跑。十载面壁勤思考，刘鹏功成发大招。油翁多年练广告，情不自禁拇指挑！

——洪涛（@zhazhaba），打油诗人，前百度高级科学家

序一

2009年11月，我在香港参加CIKM'09，听Andrei Broder等几位学者讲了一个导学课——Introduction to Computational Advertising（计算广告导论），耳目一新，觉得在我们的大学中应该有这样一门课。

回来后了解了一下周围青年教师的情况，没有发现能够开这种课的人。后来，好像首先是在微博上，知道了刘鹏是这方面的专家。一联系，果然如此。与其探讨在北大开一次这种课的可能性，他欣然应允。时间定在2013年夏天，我安排实验室的青年教师彭波做助教，目的之一，就是希望通过助教工作学会计算广告这一套知识，然后独立在北大开出课来。

刘鹏的课进行得很顺利，彭波每次都参加，我也去听过一次。2014年秋，彭波勇敢地开出课来了。课程结束后我问他感觉怎么样，他说内容太多，把握得还不好，而且没有教材，对老师学生都是个困难。

其实，最初我请刘鹏来上课的时候就谈到过教材的问题。他答应考虑，但因为他在公司里的工作很忙，所以需要比较长的时间。但他没有忘记！两年多过去了，一天他给我发邮件说书稿完成了，希望我能为他的书写个序，令我十分欣喜。

这本书不厚，但比较全面地覆盖了基于互联网服务的广告的市场背景、产品逻辑与关键技术，给出了一个宽阔的视野。作者基于多年的从业经验，从市场行为出发演绎对产品与技术的需求，而不是就技术讲技术，提高了本书的立意，因而也适合更广泛的读者群，包括计算机相关专业的研究生。应该说，这本书的风格不同于通常的教材，如果直接用于教学，对教师的要求会比较高，但不失为一本优秀的教学参考书。尤其是在其内容铺陈中展现出来的数据加工、利用与交易的思维主线，能让计算机专业的学生看到活生生的技术需求。而在互联网广告的背景下对数据的充分强调，让读者对大数据的意义有了一种更具体的体会。

国内大学中的计算机专业教育（尤其是高年级和研究生的）现在困难和问题还比较多。比较明显的一点就是，教学内容的时代感不够强。这一点在广度和深度上都有反应。跟不上业界的发展，一些重要的课程不能及时有效地开，我认为“计算广告”就是其中之一。这种情况和蓬勃发展的信息技术和产业是不相适应的。因此，我们欢迎业界中对技术和

产业有比较透彻理解且对教育有情怀的专家参与到大学教学活动中来，让我们的学生学到更多的真本事，适应产业发展的需要。刘鹏2013年在北大首开“计算广告”课程就是这样一种表率，他这本书的面世也是这个意义上的一种奉献，当予祝贺。

李晓明，北京大学计算机系教授

序二

所有互联网公司都对广告变现的地位和价值并不陌生。在每一个用户产品成长的各个阶段，除了认真解决需求痛点、优化用户体验，也应该不断地对流量和数据的价值进行评估，并积极探索商业变现的战略与产品。而在各种商业化产品当中，以计算为导向的广告变现无疑是最为重要的。

在产品选型、开发和运营的初期阶段，如果能对产品未来产生的数据和流量价值有正确的评估，并了解如何利用广告产品将这些资产变现，对于判断该产品的成长空间和商业价值非常重要。另外，早期的产品推广会用到许多广告营销产品，而对于计算广告原理的深入了解也将有利于高效地做好营销。

当产品得到市场认可，获得了一定的用户规模以后，积极制定系统性的商业化战略，用合理的变现方式获得现金流，从而支撑产品的快速发展，则是每一个互联网公司成长过程中必须经历的关键步骤。如果能洞悉互联网广告市场的产品技术全貌，无疑对此阶段的决策大有帮助。

虽然广告技术在互联网行业至关重要，长期以来，却只有一些只鳞片爪的专题文章，对业界系统架构与算法的介绍，从世界范围来看，都非常缺乏系统性的整理和总结。这一方面是由于广告市场发展迅速，从搜索竞价到程序化交易，再到移动互联网下的原生广告趋势，日新月异的产品进化速度让整个工业界来不及停下脚步做小结；另一方面是广告产品的内在逻辑不像用户产品那样直觉，要进行全面透彻的整理和剖析，需要兼有丰富的实践经验和相当的理论抽象能力。也正由于缺乏系统性的资料，互联网工业界在这方面的人才培养也不够系统，导致在广告产品技术这样一个重要的领域，人才一直是短缺的。

刘鹏博士曾经与我在搜狐集团有过一段时间的同事经历。从简短的几次接触中，我知道他在对媒体的流量变现和需求方广告产品方面都有丰富的实践经验，并曾在Yahoo! Labs对计算广告领域进行过系统性的研究，是对这一领域做全面总结的合适人选。如今，终于看到他不吝时间和精力，将计算广告领域的产品技术和商业逻辑整理成书，这将是令整个互联网工业界受益之举。

带着期待读完本书，我的第一印象是，其内容全面而富有条理：本

书既有计算广告全线产品的介绍，又有对其商业逻辑和原理的透彻解剖，还有对应的技术架构和关键算法的深入讨论。另外，除了受众定向、点击率预估、实时竞价等热点问题的讨论，还有详尽的周边产品和技术的介绍。相信认真读完此书的读者，一方面会对整个广告生态的全貌有全局性的了解，不会只见树木，不见森林；另一方面又可以按图索骥，再碰到各种实际问题时在本书中找到具体思路甚至解决方案。

当然，本书的另外一项重要意义就是，它是计算广告领域第一本系统性的正式出版物。非常希望以此为契机，从合理配置资源的角度出发，整个互联网领域能够在流量和数据变现上逐渐走向标准化与分工协作。这也许会从一个侧面促进中国互联网企业摆脱恶性竞争的囚徒困境，走向合作共赢之路。

最后，祝贺此书的出版，并希望它能够给你些许启示。

王小川，搜狗公司 CEO

序三

广告营销处在历史转折点，技术对传媒的驱动和融合趋势让数字营销充满变数，这变化颇有乐趣却又让人不安。原因很简单，一方面，技术驱动下的数字生态百花齐放，程序化营销渐入佳境；另一方面，数字世界各式各样的广告技术概念让市场营销者感到困惑。

不可否认，营销行业有专业的技术型人才和数据科学家帮助我们实践和创新这些技术，似乎有了需要的一切。可想象一下，某个下午时光，当和我们的客户坐在一起，他可能会问起这样的问题：未来你们能帮我们做什么？

如果说技术代表营销的未来，那么技术到底是什么呢？技术为什么存在？技术可以帮助做些什么？我常想，要拨开这些技术迷雾，营销人具备的知识背景应该能跨越技术理解的盲点，能洞察到真正关键且清晰的归因，把这些问题答案清晰和简单地传递给我们的客户。所以我期待行业中有人能把广告技术的真实情况和作用讲出来，无论是 **DSP**、**DMP** 或是**RTB**这些商业产品概念，还是“预测模型”“机器学习”和“人群定向”之类的技术名词。

带着这个期待，我阅读了刘鹏博士这本广告技术专著。我想说，我的这些疑惑在阅读这本书的过程中都得到了解答或者找到了线索。

刘鹏博士在互联网领域，特别是广告变现产品领域有着非常丰富的从业经验。从雅虎全球研发中心到微软研究院，再到今天作为360首席商业架构师，他既主持过需求方营销产品和供给方变现产品的设计开发，又兼有从产品到系统和算法的全面把握能力，而这些经验都成了本书丰富实用内容的基础。

一本好的广告书不会大谈趋势，而是会从细节观察出发，探知商业逻辑；一本好的技术书不会大谈常识，而是剖析实践领域的真知灼见。刘鹏博士编写的《计算广告》就是这样一本跨越领域、兼而有之的作品。

很愿意分享两点阅读感受。第一是繁纷复杂的数字生态和技术说辞，刘鹏对此做了系统的梳理和介绍，即使高度专业的产品概念、逻辑及算法应用，非技术背景的读者也能对这些概念建立统一的认识。第二是概念之外，书中列举了国际国内经典的广告平台产品，分析其形态、

技术、策略，描绘了商业和产品之间相互关联、相互促进的有趣演进。这些来自于作者多年从业实践和积累并给营销人带来“互联网+”的思考角度更难能可贵。而书中列举了很多详实的数据和图例，反映了刘鹏博士对技术和治学的严肃态度。

如果你需要了解在线广告的产品和技术，就应该马上行动，打开这本书，努力去学习和探索。

愿每位从事数字广告事业的营销人，都能读到此书。

李桂芬，安吉斯媒体集团大中华区首席执行官

前言

互联网的快速发展改变了整个世界。从门户网站到搜索引擎、从社交网络到电子商务，从免费Wi-Fi到应用市场，层出不穷的在线服务不仅方便了人们的生活，甚至颠覆了原有的产业。而且更为神奇的是，这些服务大多是免费的。在今天，“互联网思维”这个名词被越来越多互联网行业内的人们追捧和畏惧，而其中很多人的困惑都在于：这么多免费的服务是如何获得收入，乃至赚得盆盈钵满呢？实际上，如果把多样的互联网产品或服务看成各式硬币的正面，那么我们会发现，其中许多硬币的背面都有着一样的图案，这就是以广告为核心的后向变现体系。正面的免费服务是为了获得流量和数据，而背面的广告业务则是将这些流量和数据变成金钱，这就是互联网最关键的思维模式之一。



在能够获得充分的流量或高价值数据后，我们认为，所有能够传播信息的商品，其售价都会趋向其边际成本。这样的观点对许多传统行业商业模式的影响是深远的，也是我们认为大家应该在互联网时代深入了解广告、了解变现产品的原因。因此，本书的内容虽然以介绍互联网广告的产品和技术为核心，但并不是想让大家学会搭建一个广告系统。我们的核心目的，是让读者在清晰地了解互联网广告全貌的基础上，在遇到与后向变现相关的产品问题时，能够以合理的思维逻辑和背景知识来应对。实际上，在互联网时代，不论你身处哪个行业，只要用心留意，会发现这类问题可能比你想象得更为常见，也更为重要。对其中最

重要的几类问题，我们来看看下面几个具体的例子。

（1）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。那么现在电影的票价为什么这么高？能否探索一种售价很低，而充分利用其信息传播能力的电影行业发行模式，获得更高的经济效益和社会效益？

（2）流量变现。例如，互联网电视厂商除了硬件销售的回报以外，还可以获得一部分用户流量。这些流量的性质和价值如何，应该以什么方式变现？

（3）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。如果以向用户免费的方式运营室内导航产品，会得到什么有价值的数据资产，从而支撑相应的后向变现，又应该采用哪种具体的商业产品来支撑？

（4）商业产品建设和运营。例如，团购、游戏联运、返利购买、积分墙这些推广模式与一般的展示或搜索广告有什么内在联系？是否可以共用某些产品和技术平台？

这4类问题的典型性和价值不言而喻。不过，要回答这些问题，仅靠独立的深入思考是不够的，你还必须对当前互联网流量和数据变现市场的商业逻辑和产品现状有相当程度的了解，并在需要具体产品实施时有相应的方案可以参考和选择。而为读者提供这方面的帮助，正是本书希望能做到的。

从传统的视角看广告，会有人认为互联网服务中的广告破坏了用户体验，这实际上是一种观念上的误解。首先，互联网广告不再像线下广告那样，以宣教性的横幅为主，而是以各种自动决策的付费信息的方式存在，这其中既包括传统的创意形式，也包括游戏联运、团购、返利、原生广告等更加契合用户意图的新传播形式。虽然，从微观上看，部分不顾及媒体价值、盲目变现的广告产品确实存在这样的问题，但从宏观上看，恰恰是因为广告这一后向变现模式的存在，互联网产品的整体用户体验才达到了前所未有的高度。在传统的企业中，一般会根据产品线分设若干事业部，每个事业部在研发自己产品的同时还要对营收和利润负责。而在互联网企业或者按照互联网方式运营的企业当中，还存在另外一种组织方式，即面向用户的免费产品部门只负责优化产品体验，不对营收负责，而专门面向客户的商业产品部门通过广告等后向变现方式为企业创造营收。实践证明，在这样的组织方式下，用户产品部门往往能够心无旁骛，专心为了提高用户体验而努力。因此，我们会看到，比起传统软件企业，互联网企业的产品在把握用户需求、优化用户体验方面往往能够做得更加优秀。

因此，在互联网的世界里，广告不再只是广告公司的事，而是每一个互联网公司都要关心的事。从结果来看，在线广告实际上成为互联网最重要的发动机。从营收上看，它支撑着互联网业务的大半壁江山。当然，广告的概念本身在互联网业务中也已经发生了脱胎换骨的变化：首先，在互联网广告中，服务于中小商家、以直接销售为目的的广告取代品牌广告成为主流，这也创造了全新的巨大市场；其次，它的关键不再是创意、策略等人工服务，而是以数据支撑的流量规模化交易为典型特点。也就是说，机器和算法取代了人员与服务，成为在线广告最鲜明的特色。可以说，互联网广告的灵魂就在于数据与计算，因此，也就产生了“计算广告”这一名词以及后面复杂的产品与技术。

计算广告这个课题，逐渐成形于以Google AdWords 为代表的竞价广告业务产生以后，并且在展示广告进入程序化交易阶段以后愈加成熟。而将其整理成一个新的研究方向，则要归功于时任Yahoo! 广告首席科学家的Andrei Broder。他在斯坦福开设的“Computational Advertising”这门课，第一次全面而系统地介绍了在线广告中的计算挑战以及工业界实用的算法。既然有了计算广告的相关课程，为什么还要再整理这本计算广告的书呢？首先当然是因为这一领域变化太快，在“计算广告”这个词诞生后的几年里，它的内涵和外延都已经发生了重大的变化，而且这几年的变化使得这个领域逐渐完备起来。因此，有必要在此时对当下的计算广告领域做一个阶段性的小结。另外还有一个重要原因：那就是我们在几次计算广告的教学实践中发现，对于在校学生或者刚刚接触此领域的朋友们来说，最主要的理解障碍不在于算法和技术本身，而在于广告的商业逻辑和产品目标。以此为出发点，本书的组织方式将以广告产品为核心。在清楚地了解计算广告的产品逻辑与商业价值的基础上，我们再来有针对性地讨论其中的算法和架构问题。因此，商业逻辑驱动的在线广告产品和技术的升级将是本书最重要的一条主线。

另外，本书还有一条潜在的主线，即数据的加工、利用与交易。熟悉了计算广告业务和产品的读者会有认识，广告业务的收益只能来自于三个方面：数据、流量和品牌属性。其中后两点是媒体的专属，而大量的广告平台在做的事情，主要就是数据的加工与利用。不夸张地说，计算广告对于数据利用的广度和深度是空前的，而且产业的各环节也是比较完备的。在各行各业都在强调大数据思维与方法的今天，深入了解计算广告产品与技术具有特别强的范本意义。因此，本书在内容组织上特别强调数据这条线索，努力向读者解释清楚如何在广告产品进化过程中一步步地更有效地利用数据。

读者对象

既然在线广告不再只是广告公司的事，那么需要了解这一业务及其背后产品技术的人群也就相当广泛了。我们希望下面几类读者可以从本书中找到有价值的内容。

(1) 互联网公司商业化部门的产品、技术和运营人员。对互联网公司来说，商业化产品中最重要的就是广告产品，不过我们发现，囿于各公司具体的产品形态，仅仅从自己的业务中全貌地了解广告产品技术并不容易。因此，本书最主要面向的读者就是这些广告产品相关的人员，希望他们通过阅读此书，对互联网广告的全貌以及复杂的技术产品系列有整体的了解，避免只见树木不见森林。

(2) 对个性化系统、大数据变现或交易有兴趣者。计算广告在各种个性化系统中具有典型性，又因为其商业逻辑的存在而相对复杂；此外，计算广告还催生了对大规模数据利用和变现的直接市场。因此，推荐等个性化系统的产品技术人员以及大数据相关的产品技术人员都非常有必要通过了解计算广告的产品和技术，对个性化系统架构、约束下的效果优化、大数据变现和交易等诸多问题在实际工业界的落地有一定的理解。

(3) 传统企业互联网化进程的决策者。传统企业在互联网化的过程中需要借鉴的绝不仅仅是利用互联网的技术和产品，更重要的是按照互联网企业形成的高效运营和变现模式来改造传统业务。从这个意义上说，互联网企业以广告为基础的后向变现体系是整个互联网化过程中至关重要的一环。因此，在这样的传统企业中，互联网化进程的决策者对广告的原理和市场必须有一定程度的了解。

(4) 传统广告业务的从业者。传统广告业务与互联网广告业务既有密切的联系又存在着巨大的差别。以技术为导向、精准地面向受众的广告策略正深刻地影响这整个广告市场。并且随着互联网广告规模的迅速扩大，这样的策略越来越为广告主接受和青睐。因此，传统广告业务的从业者必须要顺应潮流，理解和运用计算广告的方法与策略，将线下资源与线上资源整合起来，才能更好地服务于广告主和媒体。

(5) 互联网创业者。我们接触过不少互联网行业的创业者，对他们来说，找到用户产品的痛点并漂亮地解决问题往往并不是十分困难。不过，一个企业最终需要的是利润，而许多对变现逻辑和思考方法不熟悉的创业者往往面对产品得到的流量和数据不知所措，而商业化进程的缓慢也会大大拖慢用户产品的运营进度，甚至因此错过企业的黄金发展机会。从这个意义上说，了解一些流量与数据变现的思路无疑会对创业方向的选择、创业过程的加速、创业果实的收获都有巨大的帮助。

(6) 计算机相关专业研究生。计算广告的人才在互联网行业相当稀缺，而目前学校对这样与工业界关系密切的实际问题在教育上是有些脱节的。我们整理此书的一个重要目的是希望为具有一定的计算机科学基础并且对工业界实际问题有兴趣的同学们提供一次指导旅行，让他们对思考和设计商业产品、运用技术解决产品问题形成正确的思考方法。

内容组织

前面说过，我们整理本书，并不是简单地为了介绍计算广告的产品和技术，更重要的目的是希望提供一个新的视角，让大家通过了解广告变现的内在逻辑，进而对互联网时代的用户产品如何将体验做到极致、将变现做到最高效有一个宏观的认识。在我们看来，如果不了解广告变现产品和市场，就谈不上真正透彻地了解互联网，也一定会在用户产品的设计和运营上有诸多掣肘。基于这样的目的，本书在内容上组织成三个部分。

(1) 第一部分介绍在线广告领域的一些基本问题和背景知识。虽然内容比较容易理解，但这部分是全书的基础，特别是对很多相关概念和术语的集中介绍，请不要略过。

(2) 第二部分主要面向产品、运营、销售等人员，以及互联网产品的宏观决策者，其内容重点在于介绍计算广告的市场结构、交易模式和主要产品。这部分内容将依在线广告产品发展的顺序展开，希望能帮助大家理解各种复杂的广告产品和交易机制产生的内在规律。

(3) 第三部分主要面向系统工程师、算法工程师和架构师。与前一部分的广告产品相对应，这部分也以在线广告产品发展的顺序，重点阐释实现各种广告产品的关键技术挑战，并提供基础的解决方案。

一般来说，对于那些想运营一项在线广告业务，或者想了解如何在线广告对用户产品怎样变现的读者来说，可以重点阅读第一部分和第二部分，并且对其中的产品与商业逻辑要深入理解；对于那些重点关注工程实现和收入优化的读者来说，在了解了前面两部分之后，还要花一些精力深入阅读第三部分，特别是其中与自己关注的广告产品相关的技术章节。计算广告这个领域的复杂性在于，对于任何一项产品或技术都需要放在相应的商业背景下去判断其合理性；而想要了解商业产品上能达到的目标，还需要对技术的现状和难点有相当的认识。因此，我们推荐的阅读方式还是尽可能地通读全书，对其中确实不相关或者知识背景上无法理解的部分简单跳过就可以了。另外，除第 10 章外，其他各章结束后我们都准备了若干开放性的延伸思考问题。这些问题往往并没有确定的标准答案，只是为了帮助大家进一步深入思考该章中的关键或有趣的问题。

在讨论在线广告市场的产品技术过程中，会涉及大量的术语和专业名词。对于对变现业务不太熟悉的读者来说，这些术语会给阅读带来一定的障碍。为了帮助读者检索和查找术语的相关内容，我们在附录中对主要术语及缩写给出了索引，以方便大家的阅读。

由于篇幅限制，本书中有些内容的细节或背景知识无法全面展开，在这种情形下会给出相应的参考文献。但是由于本书并非学术著作，在引用文献时并不会保证完备性，因此当正文足以说明观点和方法时其原始文献不一定还会列出，请读者谅解。另外，在本书的第三部分中，为帮助读者理解，会对一些比较关键的算法给出相应的代码片段。不过本书中的代码都是示例性代码，目的仅仅是为了更清楚地描述逻辑，而并非可以直接编译执行的程序，其中一些特别容易理解实现的子函数调用也可能会略去其具体实现。

本书的内容主要是由刘鹏在清华大学的公开课，以及在北京大学、北京航空航天大学研究生课程的内容整理加工而成的，并且在整理时针对更广泛的读者群体做了内容本身和顺序上的调整。在两位作者中，刘鹏为主要执笔者，负责主体部分的写作和内容的整体组织，王超负责其中产品案例和算法示例代码的部分。由于作者的水平有限，再加上时间仓促，书中难免出现错漏之处，敬请读者多多批评指正。此外，本书撰写的过程中，我们邀请了一些业内的专家和从业者帮忙对内容进行把关，得到了他们的很多有益的建议，这些建议使本书更加完备和实用，我们在此一并表示感谢，并将其中部分专家对本书的评价附在书中。

第一部分 计算广告关键技术

第1章 在线广告综述

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。与传统广告不同的是，在线广告在其短短十几年的发展过程中，已经形成了以人群为投放目标、以产品为导向的技术型投放模式。在线广告不仅为广告主带来了以准确接触目标受众为方法论的全新的营销渠道，也为互联网免费产品和媒体提供商们找到了规模化变现的手段。可以说，不论你在做一款用户产品还是商业产品，不深入了解在线广告，就不太可能全面地了解互联网业务。因此，所有互联网行业的从业者们花一些时间把现代的在线广告原理和产品搞清楚，不仅是有益的，而且是必须的。

从另外一个角度，即数据的角度来看，在线广告开启了大规模、自动化地利用数据改善产品和提高收入的先河。可以不夸张地说，在过去相当长的一段时期内，大数据（big data）这一方法论在实践中唯一形成规模化营收的落地行业就是在线广告，只不过当年大数据这个词还不那么流行罢了。即便在今天，计算广告仍然是大数据应用中最成熟、市场规模最大的行业。因此，对大数据感兴趣的读者认真研究在线广告发展过程中遇到的技术挑战和产品问题，会对探索其他的大数据应用有莫大的帮助。

熟悉在线广告市场的朋友都知道，这一领域的产品形态和业务逻辑相当复杂。为了对在线广告有宏观上的把握，我们在本章中将从两个方面来探讨：一是它的内涵，即这种商业活动的定义与目的；二是它的外延，即在线广告发展的简要历史和发展过程中产生的关键产品形态。对其中的许多概念和观点，读者未必能够马上形成清晰的印象，然而随着内容的展开，读者能够剥茧抽丝般层层递进地加深理解，这正是本书希望做到的。另外，本章还有另外一个目的，就是尽可能集中地介绍互联网广告的产品和技术术语，以方便后面的讨论。

相比传统的线下广告，在线广告的产品和创意形式由于互联网媒体形态、交互方式等方面存在非常大的差异，也呈现出各种各样的表现形式。我们也将对其中比较常见的创意形式作简要介绍，希望读者能对在线广告的具体表现有直观的了解。

本章的内容与计算基本无关，目的在于让读者在进入计算广告领域之初就建立起一些重要观念。首先，广告不完全等同于搜索或推荐，它

首先是一项商业活动，然后才是一项在互联网环境下需要技术优化的商业活动；其次，在这一商业活动中，广告主、媒体和用户的利益都需要被认真考虑和满足，这样才能达到整个市场的平衡和不断发展。在线广告市场所有产品和商业形式的演进，都是在这一主题下发生的。在商业逻辑的框架下思考和探索计算广告技术对理解本书中提到的产品、架构和算法非常关键。

1.1 大数据与广告的关系

近年来，大数据思维和技术渐成显学。然而，大数据这一概念至今为止并没有一个内涵上准确的界定。在参考文献[56]中，作者用 **Volume**（规模）、**Variety**（多样性）、**Velocity**（高速）和**Value**（价值），即所谓的**4V**特征来描述大数据问题的特性，但并没有给出这类问题的界定标准。然而，从实际操作的角度来看什么是大数据问题或许要比理论上的定义简单一些：如果有的数据处理问题无法通过数据采样的方法来降低处理的复杂程度，就必须利用一些专门为海量数据处理而设计的计算和存储技术（如 **MapReduce**、**NoSQL**数据库等）来实现。于是，这样的问题也就从工程上归为大数据问题，图1-1阐释了这一视角。

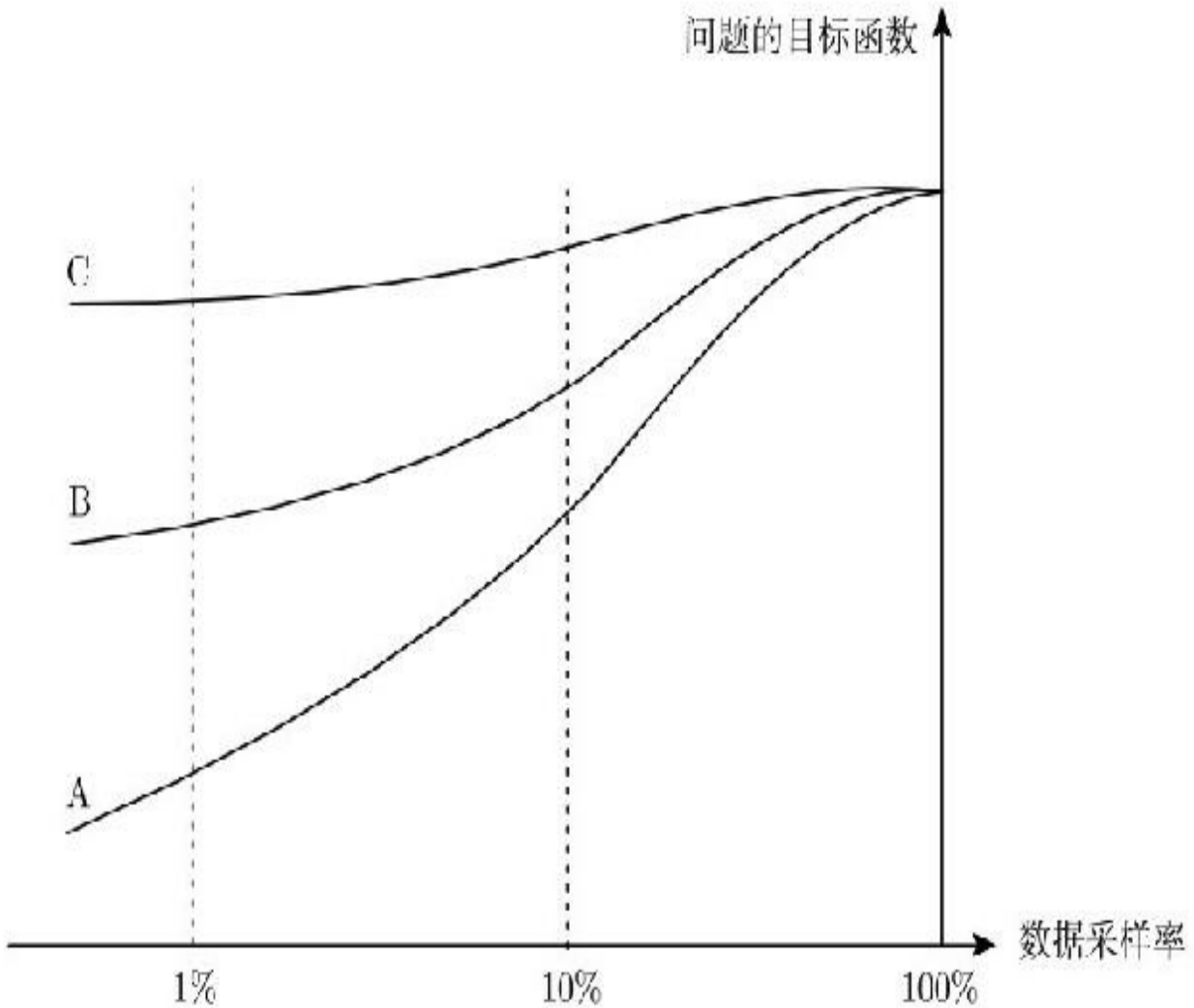


图1-1 大数据问题的特性示意

在图1-1中，我们考察的是某一个有确定目标函数的数据处理问题。图中的三条曲线是三类有代表性的数据问题。

(1) C类问题。从工程方便的角度来看，如果通过数据采样能够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，那么显然应该这样做。这类问题可以用图1-1 中的 C 曲线来示意。由于可以通过很低的采样率解决问题，并不需要大规模分布式的计算架构，用传统的数据方案就可以解决，因此，这类问题应该归为传统数据处理问题，而非大数据问题。一般的统计报表、报告等往往属于这类问题。

(2) A类问题。另外有一些数据问题基本上不可能通过只处理一小部分数据来达到处理全量数据所能达到的效果，或者说随着数据采样率的降低，解决问题的收益会快速下降，这类问题是典型的大数据问

题，用图1-1 中的 A 曲线来示意。由于需要处理大规模的全量数据，传统的存储和计算架构都不再合适，必须寻找新的方案，这实际上是推动大数据技术发展的原动力。个性化推荐（personalized recommendation）和计算广告（computational advertising）需要用到每一个人的行为进行定制化推送，而无法只采样其中的一部分人来处理，因此可以认为是典型的大数据问题。大数据问题由于无法利用传统的计算架构和数据仓库来处理，因此才会产生Hadoop等新的基础设施和NoSQL数据存储等技术。

（3）B类问题。当然，实践当中大数据问题和一般数据处理问题并不是泾渭分明的。有一些问题，其处理效果随着数据量的上升有一定提高，但当数据大到一定规模以后，再增加数据量价值就不大了，这类问题可以用图1-1 中的 B 曲线来示意。一个典型的例子是文本主题模型（topic model）。我们用 1000万文档往往会得到比 10万文档更稳定、更有意义的主题，然而用10亿文档和用1亿文档差别可能就会不那么明显。在解决这类问题时，往往是选取一个有较大规模但并非全量的数据集来处理。针对这种中等规模问题上的复杂算法，也产生了像Spark这样更加灵活高效的计算框架。

很显然，从以上观点出发，计算广告是非常典型的大数据应用。实际上，在以往相当长的一段时期里，我们认为唯一得到充分商业化和规模化的大数据应用就是计算广告。计算广告为各行各业大数据的落地提供了非常有价值的借鉴范本，下面几点尤其值得了解和关注。

（1）计算广告为规模化地将用户行为数据转化为可衡量的商业价值提供了完整产品线和解决方案，并且实际上创造了互联网行业大部分的营收。

（2）在线广告孕育和孵化了较为成熟的数据加工和交易产业链，并对其中的用户隐私边界有深入探讨，这值得所有涉及用户数据的互联网应用学习和借鉴。

（3）由于有了商业上的限制条件，计算广告的技术和产品逻辑比单纯的个性化系统更加复杂周密。因此，理解在线广告的产品和市场对于设计正确有效的商业产品大有益处。

由于以上这些原因，如果你是一位从事大数据或商业产品的产品经理、工程师或管理者，我们强烈建议你认真了解一下广告的产品和技术，相信你一定会有很大的收获，也会快捷地了解到这一领域真正有挑战的问题是什么。本着这样的目的，我们在本书后续部分中将广告市场的产品和技术演进作为一条明的主线，而将这一市场对数据的利用程度作为一条暗的主线来展开。希望读者能够通过阅读本书，具体地了解数

据是如何通过广告市场规模化地创造商业价值的。

1.2 广告的定义与目的

在了解计算广告之前，我们先从了解一般广告的目的开始。什么是广告？可以参考William F.Arens在《当代广告学》^[3]中给出的定义：

广告是由已确定的出资人通过各种媒介进行的有关产品（商品、服务和观点）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。

这一定义中有两个关键点。首先，它指出了广告活动的两个主动参与方——出资人（sponsor）和媒体（medium）。在数字广告这样更加复杂的市场结构中，我们可以用一般性的术语来描述它们：需求方（demand）和供给方（supply）。这里的需求方可以是广告主（advertiser）、代表广告主利益的代理商（agency）或其他技术形态的采买方；这里的供给方可以是媒体，也可以是其他技术形态的变现平台。另外，要特别注意的是，广告还有一个被动的参与方，即受众（audience）。请大家从现在开始就牢牢建立起这样的概念：出资人、媒体和受众这三者的利益博弈关系是广告活动永远的主线，这一主线将贯穿于商业和产品形态的整个演化过程。另外，该定义还阐明了广告必须是有偿的、非人员的信息传播活动。这两点限制，前者使得广告的目标变得明确，后者使得这一目标可以采用计算的方式来优化，而这些都是计算广告产生的基础。

广告这一商业行为，其本质目的是什么呢？在不同的时代，广告主与媒体对这一问题存在着不同的认知。在传统媒体时代，供给方与需求方在市场地位上有相当的距离，不论你运营的是电视台、机场或杂志，都与大多数广告主需要的转化行为之间有相当大的差距。因此，这一阶段广告的目的是希望借助媒体的力量来快速接触大量用户，以达到宣传品牌形象、提升中长期购买率与利润空间的目的。这种目的的广告称为品牌广告（brand awareness）。当然，也有许多广告商希望能利用广告手段马上带来大量的购买或其他转化行为，这种目的的广告称为直接效果广告（direct response），有时也简称为效果广告。

在传统广告产品中，大量投送和优化效果广告的能力显然是缺乏的。这是因为，对短期效果的追求要求广告精准地送达目标人群，而这在传统媒体上缺乏有效的技术手段。我们能够想起的以效果为目的的传统广告恐怕只有在写字楼下散发的快餐传单，而数字媒体的出现使得效

果广告空前蓬勃地发展起来。这主要有两方面的原因：一是数字媒体的特点可以让我们低成本地投送个性化广告；二是一些在线服务，如搜索、电子商务，由于可以更清楚地了解用户的意图，也就使广告效果的优化更加容易。

互联网广告兼有品牌和效果两方面的功能。不过要说明的是，到目前为止，互联网广告行业的高速发展主要是由于效果广告市场带来的巨大红利。从表1-1^[1]中可以看出，网络广告的市场规模发展迅猛；与此同时，传统广告渠道则增长乏力或快速下降。对比来看，网络广告的迅猛成长并没有直接带来电视广告的市场萎缩。这是因为网络广告的主要场景仍然集中在搜索引擎营销、效果类广告网络等直接效果类的广告活动上，而这部分相对于电视的品牌广告更多地是增量而非替代。另一方面，报纸广告则随着互联网的快速崛起而下降，这一方面是因为报纸占据的用户时间大幅被互联网抢走，另外也是由于报纸上一部分的分类信息广告与互联网效果类广告重叠较大。目前，随着数字媒体越来越多地占据了人们的时间以及在线视频等冲击力更强的媒体的普及，网络渠道也必将在品牌广告方面有更多的用武之地。

表1-1 中美主要广告市场规模（单位：亿美元）

国家	广告类型	2007 年	2008 年	2009 年	2010 年	2011 年	2012 年	2013 年
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

既然有品牌和效果这两种目标，究竟如何描述广告这种商业活动的根本目的呢，我们仍然借用《当代广告学》中的见解：

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。

也就是说，按某种市场意图接触相应的人群，进而影响其中的潜在用户，使他们选择广告主产品的几率增加，或者对产品性价比的苛求程度降低，这才是广告的根本目的。至于短期内的转化效果，由于市场意图或媒体性质的不同，并不是直接可比。换句话说，如果仅仅以转化效

果为目的来思考问题，可能会背离投放广告的正确方法论。举个例子，某感冒药广告商如果以短期效果为导向，那么最佳的策略是把广告投放给那些现在感冒的人，不过这显然是一个荒谬的决策；再比如，某汽车广告商为了提升自己的品牌形象，希望对自己竞品品牌的用户加强宣传，而对于这部分人群，广告的直接效果甚至有可能比随机投放还要差。认清这一概念，使得大家在遇到多种广告渠道的效果比较时能够避免偏颇的结论。关于这一点，有所谓整合营销（integrated marketing）的概念，即通过多种渠道的有机配合来达到整体投放效果的最优，这并非本书讨论的重点，有兴趣的读者可以参考其他文献。

广告的“低成本”是与那些由市场或销售人员完成的劝服活动成本相对而言的，实际上是广告搭了媒体流量和影响力的便车。要确定是否真的成本较低，需要用到投入产出比（Return On Investment, ROI）这一评价指标，即某次广告活动的总产出与总投入的比例。在实际中，广告活动的总投入容易确定，但总产出的确定却不那么容易，特别是在投放以中长期收益为目标的品牌广告时。因此，绝对的 ROI 有时难以计算，不过通过各个渠道之间的对比，我们仍然可以评估广告的成本是否令人满意。

需要说明，在互联网环境中，广告的本质虽然没有变化，但是由于大量直接效果需求的产生，其表现形式越来越丰富和灵活了。不论是与线下类似的横幅、搜索竞价排名，还是软文，甚至是表面上与广告并不相干的游戏联运，其本质都是付费的信息推广，从产品和技术的角度来看都可以归在广告的范畴下。因此，对于互联网广告，我们有如下的认识：

一切付费的信息、产品或服务的传播渠道，都是广告。

那么，在线广告主要有哪些表现形式呢？我们将在1.3节中介绍。

1.3 在线广告创意类型

在线广告除了产品有诸多形态，与用户接触的创意形式也有多种多样的选择，并且随着互联网产品的发展变得越来越丰富。诸多在线广告的创意类型在投送方式和用户交互方式等方面有不同的选择，下面我们对其中一些常见的类型进行简要介绍。

（1）横幅广告（banner ad）。这是展示广告中最传统也是最典型的形式。横幅广告一般是嵌入在页面中相对固定位置的图片，需要占据固定的版面，因此，这种广告一般要有底层的备选广告素材，以防没有合适的广告匹配时页面上开天窗。目前，横幅广告大多数也都不再是静

止的图片，而是由Flash或其他技术方式实现的动态素材。图1-2中给出了横幅广告的一个示例。



图1-2 横幅广告示例

(2) 文字链广告 (textual ad)。这种广告的素材形式是一段链接到广告主落地页的文字，在搜索广告中为主流形式，同时在展示广告中也被广泛采用。文字链广告有时像横幅广告那样占据固定的版面，有时也可以穿插在大量内容链接条目中。在后一种情形下，广告投放引擎可以灵活决定是否投出该文字链广告，以及投出的条目数。典型的例子就是搜索广告，如图1-3所示。

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.

图1-3 文字链广告示例

(3) 富媒体广告 (rich media ad)。这类广告往往是利用视觉冲击力较强的表现形式，在不占用固定版面位置的情况下，向用户侵入式地投送广告素材。富媒体广告常见的形式有弹窗、对联、全屏等。它比较适合在高质量的媒体做一些品牌性质比较强的广告投放，但是对用户的使用体验往往影响也较大。富媒体广告与横幅广告不同，在未售出的情况下可以静默，因而没有防天窗的问题。一些门户网站的首页有时会为某个品牌广告主提供专门定制的、交互形式很复杂的富媒体广告，这样的广告一般不采用按人群投放的逻辑，也主要强调创意的冲击力和交互形式的特色。图1-4中给出了一个富媒体广告中的弹窗广告示例。



图1-4 富媒体广告示例

(4) 视频广告 (video ad)。随着在线视频的快速发展，在视频流播放的间隙插入的广告也成为互联网广告的一种重要形式。根据插入位置的不同，视频广告又可以分为前插片、后插片、暂停等类型。视频广告由于载体的独特性质，其效果和广告创意比较类似于线下的电视广告。相应地，有关视频广告的效果评价，除了计算与横幅广告一样的点击率，还可以采用用户观看时长等更接近于用户印象的指标。

视频广告有两种最主要的形式：在视频内容播放之前的前插片广告以及视频播放暂停时的广告。图1-5 中给出了这两种视频广告形式的示例。前插片广告一般采用短视频的形式，创意的冲击力和表现力要远远强于普通的展示广告，因此价格往往也比较高；暂停广告则与普通的横幅广告区别不大。



图1-5 视频广告示例：前插片广告（左）暂停广告（右）

（5）社交广告（social ad）。社交网络的兴起给广告的传播渠道和能力都赋予了新的空间。在社交网络环境下嵌入的广告可以通称为社交广告。社交广告中最典型的形式是插入在社交网络信息流中的广告，这种方式最早见于Twitter，产品称为“Promoted Tweets”。这种方式力求在用户自然关注的交互过程中尽可能自然地插入广告，也被归于原生广告的范畴中。我们认为，“社交广告”与“社交网络中的广告”是两个不同的概念，如在社交网络页面上竞价售卖的文字链或横幅广告，其本质并不因为处于社交网络中有太大的变化。社交广告希望达到的效果是通过用户的扩散式传播获得更大的影响力和口碑，从这个意义上讲，在信息流的交互中挖掘价值前景光明。图1-6中给出了社交网络信息流广告的一个示例。



图1-6 社交网络信息流广告

(6) 移动广告 (mobile ad)。移动互联网在近几年爆发式地增长，并且大有取代桌面互联网之势。严格来说，移动互联网上的广告形式与桌面电脑上的广告没有本质的区别，不过由于移动设备上应用的大量普及，广告也由 Web 访问的页面上搬进了应用里。于是，也产生了在应用中插入广告的SDK和相应的广告网络。目前移动广告典型的形式有横幅、开屏、插屏、积分墙或推荐墙等。图1-7中给出了移动广告形式的一些示例，我们将在第7章中具体讨论与移动广告和原生广告相关的产品问题。

(7) 邮件定向营销广告 (E-mail Direct Marketing, EDM)。这是通过电子邮件的方式向目标用户传递推广信息的一种网络营销手段。与上面各种广告形式都不同，EDM是一种主动的广告形式，它不需要等到用户接触的机会出现时才被动地提供广告，而是可以随时向认为合适的用户发送推广信息。不过也正因为如此，EDM非常容易变成垃圾邮件的主要来源。因此，对EDM的运营者而言，精准地把握用户兴趣、

非常有节制地提供对用户用价值的相关信息是非常关键的。比起展示广告，EDM 中受众定向的利用更加直接，也更为重要。图1-8中给出了邮件营销广告的一些示例。另外，通过短信、iMessage等方式投放的广告与邮件营销广告也非常相似。



图1-7 移动广告形式示例

	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！欲经典文摘，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元折扣券！	5-22(三) 03:19	☆

图1-8 邮件定向营销广告（EDM）示例

当然，由于广告的本质是一切付费的信息、产品或服务的传播渠道，在线广告实际上的产品范畴远远不是上面介绍的这几种创意形式可以包罗的。除了上面这些以创意为载体的推广的狭义广告形式，还有很多的付费推广类商业产品也可以认为是广义的广告产品，其本质的产品

技术框架与普通广告非常类似，我们将在1.5节进一步介绍这些泛广告产品。

1.4 在线广告简史

在讨论广告技术之前，我们先浏览一下在线广告发展的历程。因为广告市场的概念、技术和术语繁多，如果不是对这些有基本的了解，很难深入探讨具体的产品。

回到 20 世纪末，那时的在线媒体（如 AOL、Yahoo! 等网站）刚刚产生不久。他们已经取得了不错的流量规模，可是投资人当然希望这些媒体也能够给他们带来真金白银。要对这些线上流量进行变现，最直接的方法就是把网站的 HTML 页面当成杂志的版面，在里面插入广告位。供给方有了，那么需求方呢？线下的广告代理公司也就把这些网站当成一本本新的杂志，按原来的思路和逻辑进行采买。这种在互联网上展示广告创意的产品形式称为展示广告（display advertising），也叫显示广告。这一阶段的展示广告售卖模式称为合约广告（agreement-based advertising），即采用合同约定的方式确定某一广告位在某一时间段为某特定广告主所独占，并且根据双方的要求，确定广告创意和投放策略。当然，这样的采买模式还没有对计算的需求，因为此时唯一需要的就是把广告主的创意作为一个 HTML 的片段插入到媒体的页面中。然而，情况变化很快，互联网媒体为了营收的增长，希望逐年甚至逐季提高自己的广告位报价。在流量快速增长或者数字广告逐渐为广告主认知的阶段，提价是比较容易被市场接受的。可是当媒体的流量和品牌认知度都相对稳定以后，又有什么新的提高收入的办法呢？

互联网广告运营者们经过探索，很快就发现了在线广告不同于传统媒体广告的本质特点：我们可以对不同的受众呈现不同的广告创意！在今天看来再平常不过的这个观念，实际上是在线广告的效果和市场规模不断发展的核心驱动力。认识到这一点，媒体找到了一条能使广告位报价继续提高的思路。例如，可以把某广告位的男性受众和女性受众卖给不同的广告主，比如对男性受众展示某剃须刀品牌的广告，而对女性受众展示某化妆品品牌的广告。这样的广告投放方式称为定向广告

（targeted advertising）。很显然，这样的广告系统已经对计算技术产生了两个具体需求：一是受众定向（audience targeting），即通过技术手段标定某个用户的性别、年龄或其他标签；二是广告投放（ad serving），即将广告投送由直接嵌入页面变为实时响应前端请求，并根据用户标签自动决策和返回合适的广告创意。由于从传统的品牌广告延

伸而来，此时的定向广告仍然以合约的方式进行。媒体向广告主保证某个投放量，并在此基础上确定合同的总金额以及投放量未完成情况下的赔偿方案。这种担保式投送（Guaranteed Delivery, GD）的交易方式逐渐成为互联网合约式广告的主要商业模式。一般来说，这样的合约仍然主要面向品牌广告主，并且遵循按千次展示付费（Cost per Mille, CPM）的计费方式。

合约广告系统中有一个重要的计算问题，即在满足各合约目标受众量要求的同时尽可能为所有广告商分配到质更好的流量。这一问题有两个难点：一是如何有效地将流量分配到各个合约互相交叉的人群覆盖上；二是要在在线的环境下实时地完成每一次展示决策。这个问题称为在线分配（online allocation）。如果将各合约的量看作约束条件，将某种度量下的质看作目标函数，可以利用带约束优化（constrained optimization）的数学框架来探索这一问题。为了得到在线环境下切实可行的解决方案，学术界和工业界的同仁在理论和工程方面进行了大量的研究，有一些高效且简便的实用方案已经为各媒体广泛采用。

需要注意，展示广告领域定向投放的最初动机是供给方为了拆分流量以获得更高的营收。如果一开始就提供非常精细的定向，反而会造成售卖率的下降。因此，最初的定向标签往往都设置在较粗的粒度上，最典型的是一些人口属性标签。受众定向显然更符合需求方的口味和利益——不要忘了，广告市场的钱全部是来自需求方的，他们的利益被满足得越好，市场的规模就会越大。因此，受众定向产生以后，市场向着精细化运作的方向快速发展。这一发展主要有两方面的趋势：一是定向标签变得越来越精准；二是广告主的数量不断膨胀。在这些趋势下，仍然按照合约的方式售卖广告会遇到越来越多的麻烦。首先，很难对这些细粒度标签组合的流量做准确预估；其次，当一次展示同时满足多个合约的时候，仅仅按照量约束下的在线分配策略进行决策有可能浪费掉了很多本来可以卖得更贵的流量。既然量的约束带来了这些麻烦，有没有可能抛弃量的保证而采用最唯利是图的策略来进行广告决策？这样的思路催生了计算广告历史上革命性的产品模式——竞价广告（auction-based advertising）。在这种模式下，供给方只向广告主保证质即单位流量的成本，但不再以合约的方式给出量的保证，换言之，对每一次展示都基本按照收益最高的原则来决策。

上面是从展示广告的发展看竞价产生的原因，实际历史的足迹却并非如此。竞价广告产生的最初场景是在互联网广告最主要的金矿——搜索广告（search ad）中。在以 Google 为代表的搜索引擎在技术成熟以后，迅速成为互联网新的入口。与门户网站不同，搜索引擎从一开始就

没有被当作媒体来看待，因此搜索流量的变现也采用了与服务自然结合的付费搜索（paid search 或 sponsored search）模式。从广告的视角来看，也可以把付费搜索看作一种定向广告，即根据用户的即时兴趣定向投送的广告，而即时兴趣的标签就是关键词。很显然，这种定向广告从一开始就直接达到了非常精准的程度，也就很自然地采用了竞价的方式售卖。

搜索广告产生了巨大的收益以后，搜索引擎开始考虑将这样的变现方式推广到其他互联网媒体上：将用户的即时兴趣标签由搜索词换成正在浏览页面中的关键词，可以将这套竞价广告系统从搜索结果页照搬到媒体页面上，这就产生了上下文广告（contextual advertising）。上下文广告的初期，创意的形式也是与搜索广告一样的文字链接。许多产品讨论将它与展示广告分开对待，不过当我们把它与上一段中精细化定向的展示广告对比来看时就知道，这样的区分实际上没有必要，或许从创意形式上把展示广告和文字链广告区别开更符合分类的逻辑。

从宏观市场上看，竞价广告与合约广告有很大的不同。没有了合约的保证，大量的广告主处在一个多方博弈的环境中。与直觉不同的是，在如何收取广告主费用这一点上，我们并非按照微观上最优的方案实施就可以达到整个市场最大的收益。关于定价机制的深入研究，产生了广义第二高价（Generalized Second Price, GSP）这一竞价重要的理论。

基于竞价机制和精准人群定向这两个核心功能，在线广告分化出了广告网络（ad Network, ADN）这种新的市场形态。它批量地运营媒体的广告位资源，按照人群或上下文标签售卖给需求方，并用竞价的方式决定流量分配。广告网络的结算以按点击付费（Cost per Click, CPC）的方式为主，这一点有数据和业务方面多层次的原因，在本书的后面会详细探讨。虽然我们不太能指这种方式的千次展示收益（Revenue per Mille, RPM）可以达到合约式品牌广告的水平，但它使得大量中小互联网媒体有了切实可行的变现手段：这些媒体有一定的流量，但还不值得建立自己的销售团队面向品牌广告商售卖，直接把自己的广告库存（inventory）托管给ADN，借助ADN的销售和代理团队为自己的流量变现。

ADN既然只通过出价接口提供价格约定，那么由谁来保证量呢？自然会有需求方的某种产品来完成。当ADN产生以后，代理公司当然也要对广告采买方式做出调整，因为此时的流量采买发生了几点显著变化：一是更多地面向受众而非媒体或广告位进行采买，这当然是与受众定向的流行有直接关系；二是需求方的代理需要采用技术的手段保证广告主量的要求，并在此基础上帮助广告主优化效果。这又是一个与在线分配

类似的带约束优化问题。但是实际上，这个问题有着本质的不同：由于只能在ADN定义好的定向标签组合上预先指定出价，而不能控制每一次展示的出价，因此，市场看起来像一个黑盒子，需求方只能靠选择合适的标签组合以及阶段性调整出价来间接控制效果。这种面向多个ADN或媒体按人群一站式采买广告并优化投入产出比的需求方产品，我们称为交易终端（Trading Desk，TD）。

在ADN中，核心的竞价逻辑是封闭的，这不能满足需求方越来越明确的利益要求。试想下面两个例子：（1）某电子商务网站需要通过一次广告投放来向它的忠实用户推广某产品；（2）某银行希望通过自己的信用卡用户在网络上找到类似的潜在用户群，并通过广告争取这批潜在用户。很显然，ADN很难直接为这两个需求提供人群标签。定制化需求催生了一种开放的竞价逻辑，让需求方按自己的人群定义来挑选流量，这就是实时竞价（Real Time Bidding，RTB）。它是将拍卖的过程由广告主预先出价，变成每次展示时实时出价。只要把广告展示的上下文页面URL以及访客的用户标识等信息传给需求方，它就有充分的信息来完成定制化的人群选择和出价。于是，市场上产生了大量聚合各媒体的剩余流量并采用实时竞价方式为他们变现的产品形态——广告交易平台（ad Exchange，ADX）。这个名称让我们很容易联想起股票交易所。事实上，如果我们把ADN的交易方式想象成场外交易市场（over-the-counter market），那么ADX与股票交易所确实有着类似的作用。

通过实时竞价的方式，按照定制化的人群标签购买广告，这样的产品就是需求方平台（Demand Side Platform，DSP）。由于实时竞价主要采用按展示次数计费的方式（本书后面部分会具体讨论其原因），DSP需要尽可能准确地估计每一次展示带来的期望价值。在这一点上，DSP比TD要方便多了，因为充分的环境信息使得深入的计算和估计成为可能。基于DSP的广告采买非常类似于股票市场上的程序交易，我们把这样的广告采买方式也叫作程序化交易（programmatic trade）。除了RTB以外，还有其他几种程序购买的交易方式，如优选（preferred deals）以及私有交易市场（Private Market Place，PMP）等。可以预见，在线广告中程序化交易的作用和地位将会不断加强，这是由数字广告的本质特点以及广告主利益最大化的趋势所决定的。

初次接触在线广告的读者可能对这部分中提到的大量概念和商业逻辑感到无所适从。不过没关系，上面所有用楷体字标出的关键概念在本书后面章节出现时会进行详细讨论。而我们写这一段的目的也仅仅是让大家对本书讨论的范畴有一个全局性的认识，从而在后面章节中接触到某个具体问题不会只见树木，不见森林。

由于在线广告存在着较为复杂的市场结构，LUMA Partners 对北美市场的主要代表公司进行非常全面的总结，并绘制成了图1-9中的“display LUMAscape”^[2]。这一图谱的骨架与上面我们介绍的在线广告简史有着非常紧密的联系，因此也是本书在广告产品方面重要的提纲。基本上可以说，这一图谱是从两端向中间逐渐发展和形成的：首先是合约阶段，广告主通过代理公司从媒体方采买广告，而媒体方的广告投放机则负责完成和优化各个广告主的合约；然后，市场进化出了竞价售卖方式，从而在靠近供给方产生了ADN这样的产品形态，而需求方的代理公司为了适应这一市场变化，孵化除了对应的媒介采买平台（media buying platform）；最后，当市场产生了实时竞价方式交易时，供给方进化出了 ADX，而需求方则需要用DSP与其对接来出价和投送广告。图中的下半部分，多是一些对这一骨干市场结构起支持作用的产品或者在细分领域的特异化产品。我们在后面介绍到相关部分时，将会给出相应的介绍。

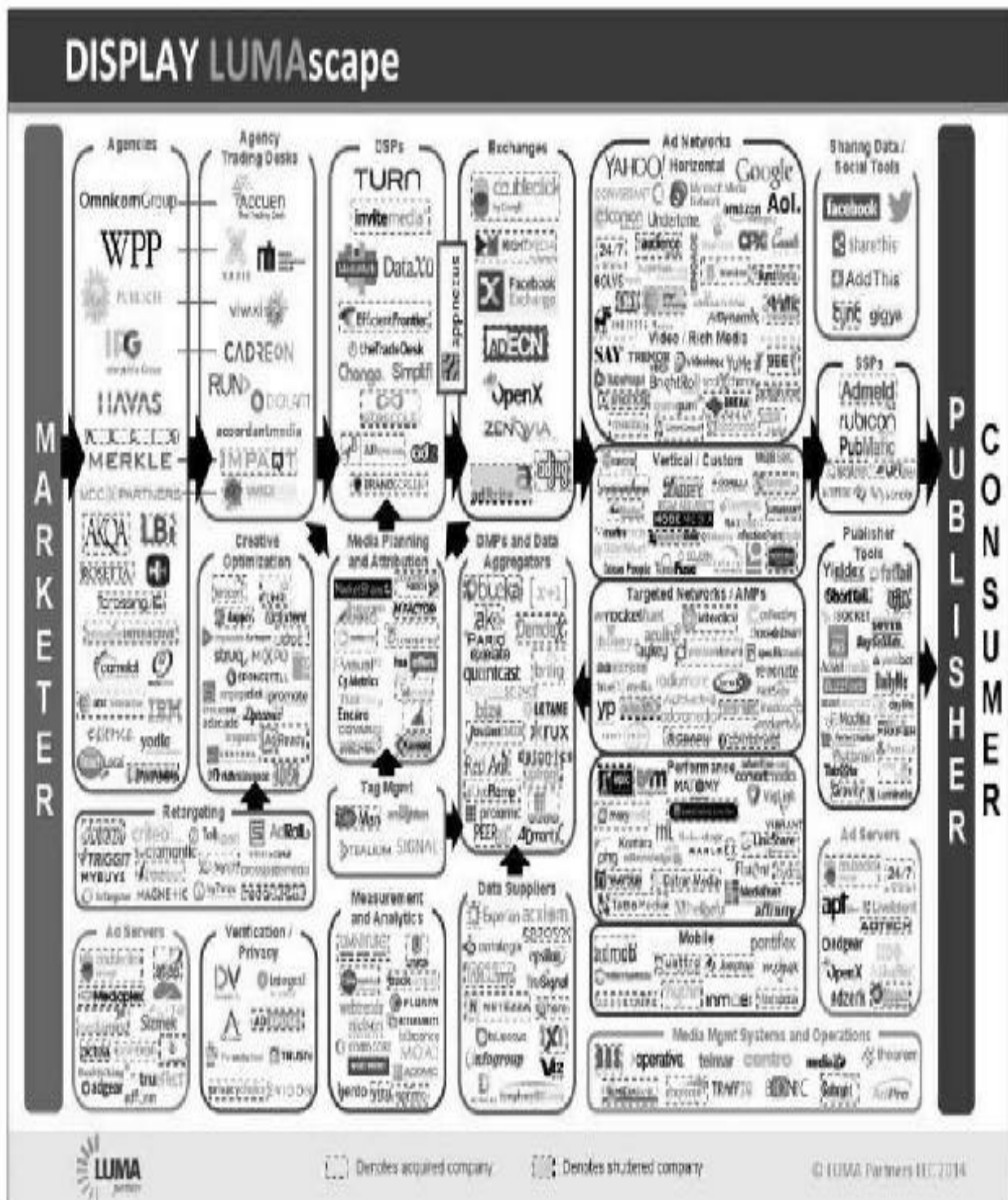


图1-9 LUMA总结的展示广告市场结构和代表公司图谱

图1-9中的个别术语与本书略有区别（如媒介采买平台），我们会在书中用相近似的概念TD来代替，请大家留意。

总结一下，在在线广告发展的历史上，定向技术和交易形式的进化

是一条主线。从最初的固定位置合约交易发展到进行受众定向、按展示量结算的合约交易，再到竞价交易方式，并最终发展成开放的实时竞价交易市场。这条主线的核心驱动力是让越来越多的数据源为广告决策提供支持，从而提升广告的效果。除了这条交易形态的主线，互联网广告产品还有另外一条发展线路，即产品展现逻辑上的发展：在展示广告的最初阶段，广告位作为与内容相对独立的单元来决策和运营，并且完全以优化收入为目标；但同时，人们从搜索广告和社交网络信息流广告中得到了启发——将内容与广告对立起来未必是一个好的选择。搜索广告和社交网络信息流广告这两种广告产品正是由于与内容的展现和触发逻辑有着高度的一致性，才使得它们的效果突出。沿着这样的思路，将内容与广告以某种方式统一决策或排序的广告产品——原生广告（Native AD）在近年来得到了工业界越来越多的关注。如何将原生的决策方式与前面介绍的广告市场已经非常成熟的规模化交易逻辑相结合是目前互联网广告产品发展的热点。需要特别说明的是，原生广告的思路在移动设备这样屏幕尺寸受限的环境下将会成为一个非常有前景的发展方向，这方面的进展我们将在第7章中具体讨论。

1.5 泛广告商业产品

下面要讨论的这些泛广告商业产品的本质都是付费推广。虽然这些付费推广模式的表现方式更加多样化，用户的感知程度和参与程度也大不相同，但是产品和销售模式却与狭义广告基本相同。希望读者在看完此书以后，能够认识到这些商业产品本质上也是在线广告，能够以统一的视角来理解和规划这些变现方式。



图1-10 团购产品示例

(1) 团购。团购本质上是一种按照效果付费的泛广告产品，其特殊性在于广告主除了付推广费用外，还向用户让利以获得转化。团购推广的主要广告主是一些本地化的店铺，主要目的是为了获得新客户。如图1-10所示，对团购平台来说，团购商品在一定环境下的排序与广告是一样的问题。不过团购销售很少采用竞价的方式，而多是预先约定价格。另外，团购的广告库中是付费信息而非创意，这有利于发展原生广告的推广方式，参见第7章。

(2) 游戏联运。游戏联运根据用户的最终游戏内消费在推广渠道和游戏开发商之间分成的商业产品，这仍然是一种按效果付费的泛广告产品。在页游和手游的推广中，联运是一种非常常见的发行模式，读者在各种产品中看到的图1-11所示的游戏下载专区，其背后的商业模式往往都是联运。实际上，在中国的各大Android应用市场中，游戏联运的收入远远超过其他类型的广告收入。如果将联运收入也算在广告收入中的话，互联网总体收入的广告占比还会提高不少。不同的联运渠道分成比例可能相差很大，经过我们调研，在Apple Store这样典型的国外市场中，渠道的分成比例一般为30%，但是在中国，有些强势的联运渠道分

成比例甚至可以达到 90%以上。游戏联运也完全可以按照广告产品思路来设计和运营，另外其广告库同样具有易于原生化的特点。



图1-11 游戏联运产品示例

(3) 固定位导航。这主要包括网址导航站的位置入口（如图1-12所示）、应用分发平台的推荐位置等付费推广位置。一般来说，这种产品的销售都采用按时间付费的固定位模式，而不是动态的竞价决策模式。这是因为广告主除了引流以外，往往更加关注这些入口位置的橱窗效应。这种广告的销售和运营模式与按天购买的合约广告相同。

(4) 返利购买。返利购买是电商行业常见的一种推广模式，它与团购有些类似，也是采用折扣或积分的方式激励用户购买。显然，这种方式可以获得非常高的ROI，但是也会带来一些老用户转而由返利网下单，因此实际的效果远没有那么好，特别是在获取新客户方面的价值值得考量。

需要特别说明，当一个公司同时运营普通广告和上述一种或多种泛广告产品时，它们之间甚至是和用户产品之间，经常会出现争夺广告位或其他入口资源的问题。面对这样的问题，最合理的分配方式是通过它们之间的竞价来决策，这是非常重要的内部流量货币化的运营理念。



图1-12 网址导航产品示例

1.6 延伸思考

- 1.考虑到品牌广告和直接效果广告的目的性差异，两者在创意设计、投放策略、媒介选择等方面应该有什么区别？
- 2.请你从自己熟悉的领域中找出几个泛广告产品的例子，并探讨其与典型广告产品的区别。

第2章 计算广告基础

从现在开始，我们开始接触在线广告中与“计算”有关的问题。在线广告中计算到底是为了解决什么问题，以及解决这些问题需要什么样的业务描述框架，将是本章重点关注的内容。

我们将先对传统广告中发展起来的广告有效性理论做简要的回顾。通过把广告产生效果的过程分解为若干阶段，并讨论其中各阶段关键的影响因素，可以对在线广告情形下受众定向、创意优化及其他有价值的技术点有感性认识。虽然这部分与具体的计算技术无关，却对计算广告中的一些根本原理有深入的揭示，希望读者有所了解。

在互联网广告中，计算之所以可以发挥巨大的作用，与它的一些根本技术特点有很大关系，这是本章的出发点。总的来说，可衡量的效果以及相应的计算优化是在线广告区别于线下广告的主要特点。在这些特色的基础上，我们对Andrei Broder提出的计算广告核心挑战稍做推广，得到贯穿本书的计算广告核心问题，即利润优化问题的概念性框架。

在大多数广告产品中，可以通过计算优化的主要是收入部分。而千次展示期望收入（expected Cost Per Mille, eCPM）正是计算广告中最为核心的量化指标之一。与广告的信息传达过程相关，eCPM又可以分解为点击率和点击价值的乘积，这两个指标是各种广告产品在计算过程中经常碰到的，也是产品运营需要深入理解和重点关注的。

同时，这样的收入分解方法还对在线广告产品市场结构和计费方式的理解很有帮助。大家将会看到，在线广告多种多样的计费方式实际上反映着市场结构的分工不同。具体来说，供给方和需求方如何分工估计点击率和点击价值，与整个市场的资源优化配置有关。对若干常见计费方式的深入理解，对于把握计算广告领域的核心问题，以及评估每个问题在特定情形下的难度有很重要的指导意义。

本章的最后，还介绍了若干广告和在线广告领域中重要的行业协会。了解这些协会在广告业务中代表的利益方以及他们对整个在线广告市场产品和技术形态的推动作用，对于有志从事广告技术和业务的读者来说也是必要的。

2.1 广告有效性原理

为了探讨用技术手段优化广告投放效果，我们先看看广告从产生用户接触开始是如何产生最终效果的。这一问题是广告领域一个传统重要的研究课题，我们直接借鉴了前人的研究成果，按照方便计算广告讨论的原则，用一个三段式信息传播模型来解剖广告由物理上产生到最终产生转化行为的全过程，这一模型如图2-1所示。

图2-1 所示这一有效性模型把广告的信息接收过程分为选择（selection）、解释（interpretation）与态度（attitude）三个大阶段，或者进一步分解为曝光（exposure）、关注（attention）、理解（comprehension）、接受（acceptance）、保持（retention）与决策（decision）6个子阶段。下面我们分别来讨论每个阶段的意义和关键点。

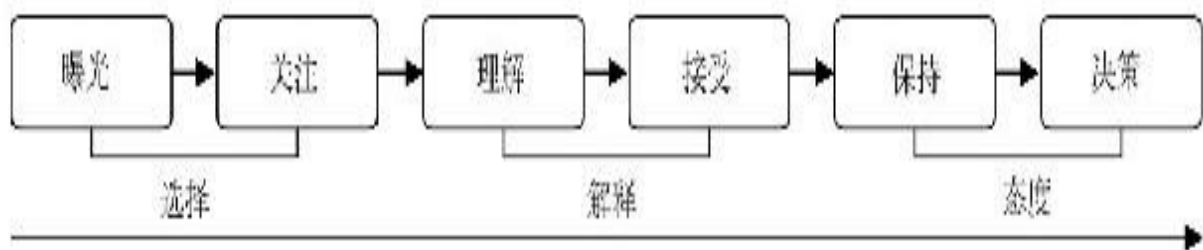


图2-1 广告效果产生过程示意

（1）曝光（exposure）阶段。这一阶段指的是广告物理上展现出来的过程，此阶段的有效程度往往与广告位的物理属性有关，并没有太多可以通过技术优化的空间。实际的广告实践中，曝光的有效性对最终结果的影响往往远远高于其他技术性因素，所以才会有传统广告中“位置为王”的说法。像纽约时代广场那组著名的广告牌（见图2-2中左图）以及北京东三环北端京信大厦外立面正对着东三环北路的广告牌（见图2-2中右图）就有非常好的曝光效果。在互联网广告中，位置的影响有时会更加显著，因此如何从算法上消除由此带来的点击率预估偏差，是一个非常重要的实际问题。



图2-2 曝光效果突出的广告位示例（左：纽约时代广场 右：北京东三环京信大厦）

（2）关注（attention）阶段。这一阶段指的是受众从物理上接触到广告到意识上注意到它的过程。对广告而言，曝光并不一定意味着关注。举个例子，有一天我的一位好友在浏览社交网站时，他的小儿子在旁边喊道：“爸爸，快看网页上的恐龙！”而这位朋友找了一分多钟都没有找到恐龙。实际上，恐龙就在网页上端最醒目的广告位上。这个例子说明，强曝光并不能等同于用户实际有效的关注。那么如何使得关注阶段的效率提高呢？有几个重要的原则。首先，尽量不要打断用户的任务。这一点是上下文相关的广告投送的原理基础，也是当今讨论原生广告产品的出发点之一。上面的例子也可以用这个原则来解释，当用户明确辨识出某个固定不变的广告位，并且不再认为它与自己当浏览网页的任务有关联时，他会下意识地屏蔽其中的内容。其次，明确传达向用户推送此广告的原因，这一点是受众定向广告创意优化的重要方向。另外，内容符合用户的兴趣或需求，这是行为定向的原理基础。

（3）理解（comprehension）阶段。用户关注到了广告的内容也并不意味着他一定能够理解广告传达的信息。再举一个例子，笔者有一次试玩了一款网页游戏，然后被某游戏广告定向到，并多次看到某宣

传“四维城战新模式”的游戏广告。应该说这样的定向是精准的，我也非常认真地关注了广告的内容，不过这些内容我确实不能直观理解，也就谈不上后续的转化。理解阶段有哪些原则呢？首先，广告内容要在用户能理解的具体兴趣范围内，这说明了真正精准的受众定向有多么必要。其次，要注意设定与关注程度相匹配的理解门槛。例如，在电视广告中，可以用有一定情节的短故事来宣传品牌；在路牌广告中，创意制作原则是将若干主要市场诉求都表达出来；而对于互联网广告，由于用户的关注程度非常低，我们应该集中强调一个主要诉求以吸引用户的注意力。

（4）接受（acceptance）阶段。受众理解了广告传达的信息，并不一定表示他认可这些信息。广告领域有一句名言：“我知道有一半的广告预算浪费了。”实际的情况还有可能更糟，如果表达的信息不适当，甚至有可能有三分之一的广告展示起到负面效果！在使用幽默、性感这样非常规的广告手段时，要特别注意这一点。广告的上下文环境对于广告的接受程度有着很大的影响，同一个品牌广告出现在某游戏社区上和门户网站首页上，用户会倾向于认为后者更具说服力，这也就是优质媒体的品牌价值。在定向广告越来越普遍的今天，如何让合适的广告出现在合适的媒体上，即广告安全（ad safety）的问题，正在引起大家越来越多的关注。

（5）保持（retention）阶段。对于不仅仅追求短期转化的广告商，当然希望广告传达的信息给用户留下长久的记忆，以影响他长时间的选择，因此品牌广告商在创意设计上了大量的精力提高此阶段效果。我们想想那些充满艺术性或浪漫气质的电视汽车广告，可以对此有直观的认识。

（6）决策（decision）阶段。成功广告的最终作用是带来用户的转化行为，虽然这一阶段已经离开了广告的业务范围，但好的广告还是能够为转化率的提高做好铺垫。特别是对于电商或团购业务，在创意上强调哪些信息以打动那些价格敏感的消费者是有相当的学问的。

定性地说，越靠前的阶段，其效果的改善对点击率的贡献越大；而越靠后的阶段，其效果的改善对转化率的贡献越大。但是以上各个阶段的划分绝非孤立的和绝对的，而某一项具体的广告策略或技术也往往会对几个阶段的效果同时发生影响。虽然这样的有关广告有效性模型的讨论多见于传统广告的研究中，然而其规律显然也对在线广告的产品方向有很强的指导作用。

2.2 互联网广告的技术特点

从前面的讨论中，大家一定已经发现了不少在线广告不同于传统广告的特点。在这些不同点当中，有一些对我们正确理解在线广告市场并探究合适的效果优化方案有着非常重要的指导意义。

(1) 技术和计算导向。数字媒体的特点使在线广告可以进行精细的受众定向，而技术又使得广告决策和交易朝着计算驱动的方向发展。实际上，受众定向这一思想在线下广告中也曾经被尝试过，比如试图把信用卡纸质账单背面的广告按照信用卡用户的年龄和性别做一些定制化，不过由于非数字的媒体上这么做的成本太高，因而无法规模化。在数字媒体上进行受众定向，其成本可以控制得非常低，这也直接催生了在线广告的计算革命。除了受众定向，由于在线广告存在着独特的竞价交易方式，因而广告效果精确的预估和优化能力也是非常重要的。

(2) 效果的可衡量性。在线广告刚刚产生之时，大家对这种广告最多的称道之处是它可以以展示和点击日志的形式直接记录广告效果，并且可以利用这些日志优化广告效果。不过，点击率这一指标从是否在绝对意义上能够反映广告效果是值得探讨的。从1998年到今天，横幅展示广告的点击率从10%一路降至0.1%，难道这说明广告的效果下降了两个数量级吗？快速增长的市场规模显然给出了否定的回答。我们认为，在不同的产品或时代中，点击率绝对值的比较并没有那么重要，而在一个特定时期不同广告和算法表现出来的差异才是更有意义的。从这一点来看，可衡量性仍然可以认为是在线广告的一个重要特点。

(3) 创意和投放方式的标准化。标准化的驱动力来自于受众定向与程序购买。既然需求方关心的是人群而非广告位，创意尺寸的统一化与一些关键接口的标准化非常关键。这些接口标准中，比较典型的有视频广告的 VAST 标准^[40]和实时竞价的 OpenRTB 标准^[41]等。实践表明，有越来越多的广告产品和平台愿意根据这些市场标准来设计自己的规范和接口，因为这样大家可以充分利用整个市场的流动性，更快地创造更多的价值。

(4) 媒体概念的多样化。随着Web 2.0和移动互联的普及，赋予了更多交互功能的互联网媒体与线下媒体已经有了本质差别。随着交互功能的不同，这些媒体与转化行为的距离也就不同。举个例子，对在线购物行业而言，门户网站、垂直网站、搜索引擎、电商网站、返利网，在转化链条上一个比一个更靠近购买行为。我们从直觉上就可以知道，越接近转化的媒体上的广告带来的流量一定可以达到越高的ROI，不过

离“引导潜在用户”这样的广告目的也就越远。因此我们在从需求方看在线广告时，应该注重各种性质媒体的配合关系，并从整合营销的角度去审视和优化整体的效果。试想，如果一家电商只用返利网作为线上广告渠道，ROI一定可以做到很高，可是这样的营销能给他带来大量潜在用户吗？^[3]

（5）数据驱动的投放决策。与工业革命时期机器化的根本驱动力——电力相类比，互联网化的根本驱动力可以认为是数据的深入加工和利用。这一点在大数据概念被广泛认知的今天已经成为老生常谈。前面提到的在线广告的计算技术在很大程度上也要依赖于对于数据的大规模利用。广泛收集用户的行为数据和广告反馈数据，利用云计算的基础设施对用户打上合适的标签，同样根据数据在多个广告竞争同一次展示时作出决策，再将投放的结果统计数据反馈给广告操作人员以调整投放策略，这已经成为在线广告的基本投放逻辑。因此可以认为，现代的在线广告系统就是一个大数据处理平台，而且其对数据处理的规模和响应速度的要求都相当高。可以说，从来没有任何传统广告形式像在线广告那样，需要大规模地收集并利用数据，而这正是在线广告最吸引人之处。

2.3 计算广告的核心问题

Andrei Broder在提出计算广告这一概念的同时也给出了该课题的核心研究挑战（注意是“核心挑战”而非“定义”）。对于这一核心挑战，他的表述是“Find the best match between a given user in a given context and a suitable advertisement”^[1]。我们结合近年来市场的发展以及实际业务中的一些体会，对此表述稍作加工，给出如下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。

与 Andrei Broder的表述相比较，我们主要进行的两方面的微调。首先，强调广告问题优化的是一组展示上的效果，而非孤立的某一次展示上的效果。这是由于广告活动中普遍存在着量的约束，在这一约束下进行利润优化，其最优解往往与每次展示独立决策时有很大的不同。其次，描述中去掉了“given”的字眼。这是由于在某些广告产品中，系统并不一定能拿到确定的用户或上下文唯一标识，但这并不意味着完全无法进行计算优化。同样地，我们也强调优化的结果是“广告投放策略”而不一定是具体的广告，这也是因为有些产品的策略并不是直接决定最后的展示。相信读完本书后面的部分，大家就能更深入地体会这些调整的原

因。

上面的计算广告核心挑战需要转化为数学上可以优化的目标才能利用计算技术来解决。把它用下面的最优化问题来表达：

$$\max \sum_{i=1}^T (r_i - q_i) \quad (2.1)$$

这里的 i 代表从第1次到第 T 次之间的某一次广告展示。我们优化的目标就是在这 T 次展示上的总收入（ r ）与总成本（ q ）的差，即广告活

动的利润。当某次广告活动的预算一定，即 $\sum_{i=1}^T q_i$ 是一个常数时，很容易验证优化公式2.1与优化另一个广告中更常见的目标投入产出

比，即 $ROI = \sum_i r_i / \sum_i q_i$ 也是一致的。进一步考虑收入与成本具体依赖的因素，上面的优化问题可以写成：

$$\max_{a_1, \dots, a_T} \sum_{i=1}^T \{r(a_i, u_i, c_i) - q(a_i, u_i, c_i)\} \quad (2.2)$$

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。实际上，对除了 DSP 以外的大多数广告产品来说，要么是自营或包断资源，要么按以收定支的方式与媒体分成，其成本也对应为常数或正比于收入，在这种情形下，成本部分可以从上面的优化公式中去掉。

注意，这里有一个隐含的假设，即整体的收入或成本可以被分解到每次展示上。显然，这一假设并不是十分合理的，但是考虑到实际线上决策时，必须对每次展示马上完成计算，所以，从实用出发我们仍然采用这一假设。在实际的系统中会采用频次控制、点击反馈等方法来对付多次展示之间效果相关性的问题。

在具体的广告产品中，优化公式2.2可能会省略掉一些内容或参数，或者增加一些约束条件，构成该广告产品独特的优化问题。在后面我们谈到若干广告产品的关键技术时，也会给出其具体形式。

2.3.1 广告收入的分解

下面，我们再来进一步分解广告收入，以便引出关于在线广告市场计费方式的重要分析。对一个广告市场中具体的产品形态，我们往往能

够主动优化的是收入而非成本，因此，可以主要关注收入优化的部分。在一次广告展示产生后，有可能发生哪些后续行为呢？参见图2-3：当用户在媒体页面的广告位上看到广告以后，如果产生兴趣，首先产生的是点击行为，广告点击与广告展现的比率称为点击率（Click Through Rate, CTR）；点击行为成功以后，将会打开广告主的落地页（landing page），落地页成功打开次数与点击次数的比例称为到达率，这是在广告主网站上发生的；如果用户从落地页开始，进一步完成下单等操作，则称为转化，转化次数与到达次数的比例称为转化率（Conversion Rate, CVR），这是在广告主网站上或线下发生的。



图2-3 在线广告产生效果的步骤

按照媒体网站和广告主网站上的行为段对回报 r 进行分解，是实践中比较合理且容易操作的方式：

$$eCPM = r(a, u, c) = \mu(a, u, c) \cdot \nu(a, u, c) \quad (2.3)$$

在后文中，我们都沿用这样的符号表示：用 μ 表示点击率，用 ν 表示点击价值（click value），即单次点击为广告产品带来的收益。其中前一部分描述的是发生在媒体上的行为，后一部分描述的是发生在广告主网站上的行为。而这两部分的乘积定量地表示了某次或若干次展示的期望 CPM 值，就是我们前面提到的 eCPM^[4]。请大家特别关注 eCPM 这个

指标，因为它是计算广告中最常被提及，也是最关键的定量评估收益的指标，本书的计算问题大都是围绕它展开的。在对多个检索候选进行排序时，是根据 eCPM 还是 CTR 排序也是区别广告产品和用户产品的重要策略特征。进一步，如果我们将所有的点击价值都等同起来，那么根据 eCPM 排序和根据 CTR 排序实际上将得到一样的结果。因此，可以认为根据 CTR 排序是根据 eCPM 排序的一种特例，这也使得将内容与广告统一排序的原生广告成为可能。

eCPM 一般指的是估计的千次展示收益，它有两个很相近的概念：如果讨论的是千次展示收入，往往用 RPM；如果讨论的是千次展示成本，往往用 CPM。这三个术语有时在实用中的区别并不明显，请大家注意。

根据图 2-3 所示的流程，点击价值还可以进一步分解为到达率、转化率和客单价的乘积。由于这部分的深入解剖与行业密切相关，而且更多地属于站内运营而非广告的范畴，因此在本书中将只在 14.2.3 节进行简要的讨论。

2.3.2 结算方式与 eCPM 估计的关系

对于大多数广告产品来说，需要计算给定 (a, u, c) 三元组的 eCPM 以进行决策。可是由于广告市场的协作关系复杂，并非每个广告产品都可以对 eCPM 中的两个步骤做出较准确的估计。根据 eCPM 的分解决定哪部分由谁来估计是广告市场各种计费模式产生的根本原因，也是广告市场中商业逻辑与产品架构衔接的关键一环。下面，我们来逐一分析下市场上主要的几种广告计费模式。

(1) CPM 结算，即按照千次展示结算。这种方式是供给方与需求方约定好千次展示的计费标准，至于这些展示是否能够带来相应的收益，由需求方来估计和控制其中的风险。对于品牌广告，由于目标是较长时期内的利益，很难通过对短期数据进行分析的方式直接计算点击价值，而点击率也因为对于用户接触的核心要求变得不是唯一重要的因素。在这种情况下，由需求方根据其市场策略与预算控制流量的单价并按 CPM 方式结算是比较合理的交易模式。实际上，在大多数互联网品牌广告，特别是视频广告中，CPM 都是主流的结算方式。

(2) CPC 结算，即按点击结算。这种方式最早产生于搜索广告，并很快为大多数效果类广告产品所普遍采用。在这种方式结算方式下，点击率的估计是由交给供给方（或者中间市场）完成的。点击价值的估计则由需求方完成，并通过点击出价的方式向市场通知自己的估价^[5]。这样的分工对于互联网广告，特别是以效果为导向的互联网广告而言，

有着清晰的合理性：供给方通过其收集的大量用户行为数据可以相对准确地估计点击率；而转化效果是广告商站内的行为，当然他们自己的数据分析体系也就能更准确地对其作出评估。因此，以CPC方式结算，在效果类广告市场中具有接近垄断的地位。

(3) CPS (cost per sale) /CPA (cost per action) /ROI结算，即按照销售订单数、转化行为数或投入产出比来结算。这些都是按照转化付费的一些变种。这是一种比较极端的结算方式，即需求方只按照最后的转化收益来结算，从而在最大程度上规避了风险。在这种结算方式下，供给方或中间市场除了估计点击率还要对点击价值作出估计，这样才能合理地决定流量分配。这里存在两个很明显的问题。一是转化行为为用户在广告商站内的行为，并非供给方能够直接监测和控制，因此无法进行准确的估计和优化。只有那些转化流程和用户体验类似的广告商组成的广告平台按转化付费才比较可行，典型的例子就是淘宝客广告。二是实际执行中，存在广告主故意扣单以降低转化率，从而低成本赚取大量品牌曝光的可能。因此，我们认为这种方式主要适合于一些垂直广告网络 (vertical ad network)。另外在DSP中，由于需要完全代表广告主利益出价和优化，因此也会出现一些跟广告主之间按照CPS计费的情形。总体而言，对于那些与广告主收益直接挂钩的需求方广告产品来说，CPS在一定条件下是可行的；但是对于普通的中间市场广告产品来说，CPS并不是一种趋势性的结算方式。而 CPA 广告在移动应用下载的场景下，由于转化流程统一在 Apple Store 或 Google Play中，且存在较完善的第三方转化监测，因而市场较为成熟。

(4) CPT (cost per time) 结算，这是针对大品牌广告主特定的广告活动，将某个广告位以独占式方式交给某广告主，并按独占的时间段收取费用的方式。严格来说，这是一种销售方式而非一种计费模式，因为价格是双方事先约定，无需计量。这种方式主要适用于一些强曝光属性，有一定定制性的广告位。在一般的展示广告中，这种方式在欧美市场市场并不经常采用，但在中国的门户网站广告中，CPT仍然是一种主流模式。CPT这样独占式的售卖虽然有一些额外的品牌效果和橱窗效应产生，但是非常不利于受众定向和程序交易的发展，因而从长期的角度来看，其比例会有下降的趋势。

表2-1展示了以上几种结算方式概要的对比。综合来看可以认为，对于效果广告，CPC计费方式最有利于发挥供给方和需求方的长处，因而在市场上被广泛接受；对于品牌广告，由于效果和目的有时不便于直接衡量，可以考虑按照 CPM的方式计费；而 CPS的计费方式只在一些特定的环境下才比较合理。

表2-1 在线广告结算方式比较

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		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		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	1. 有受众选择需求的品牌广告 2. 实时竞价广告交易
CPC	供给方	需求方	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/CPA/ROI	供给方		需求方无任何风险 供给方运营难度较大	1. 效果类广告联盟 2. 效果类 DSP

既然广告有计费的需求，也就同时产生了效果监测的需求。在CPM类品牌广告中，由于曝光在媒体上产生，广告主往往会委托第三方的广告监测公司对曝光量、点击量等指标作技术核实，并以此作为结算的依据。在CPC或CPS结算的广告交易中，由于计费的指标，即点击或转化在广告主的网站上产生，所以并不需要特别的监测服务。因此，可以认为广告监测主要服务的对象是品牌广告主。随着CPM广告定向方式越来越复杂，广告监测也从简单的展示和点击记数到频次、人口属性等信息的验证和计量。关于这方面的问题和技术，我们将在第15章中再作介绍。

2.4 在线广告相关行业协会

由于供给方和需求方的博弈关系，需要一些行业协会来约束和规范市场。关注这些行业协会的立场与使命对更清晰地认识广告的商业逻辑大有帮助。以最重要的北美市场为例，主要有三个行业协会需要了解。

2.4.1 交互广告局



交互广告局（Interactive Advertising Bureau, IAB, <http://www.iab.net>）成立于20世纪末，是在线广告领域最重要的行业协会，其使命是致力于推动在线广告市场的发展。IAB主要是站在供给方的长远利益上来研究和影响市场。换句话说，IAB主要关注的是在线广告供给方的利益。正如IAB在自己网站标题上声明的那样，这一组织存在的使命是“dedicated to the growth of interactive advertising market”，即致力于交互广告市场的壮大。而这个方向的受益者主要是各在线媒体与广告技术公司。因此，IAB的典型会员是Google、Facebook、Yahoo!、Microsoft这样的广告供给方以及AudienceScience、MediaMath这样的广告技术公司。广告技术公司和产品的发展是为了更好地服务广告主和提升在线广告效果，以利于更多的预算进入在线广告领域。从具体工作上看，IAB与互联网大量媒体和广告平台合作，制定了一系列意义重大的标准和规范，这些都极大地促进了在线广告行业的健康发展。其中包括如下几个重要的规范。

（1）横幅广告创意尺寸标准。2001年 IAB 公布的标准仅仅支持 7 种创意尺寸，即120×600（摩天大楼）、160×600（宽摩天大楼）、180×150（长方形）、300×250（中级长方形）、336×280（大长方形）、240×400（竖长方形）、250×250（正方形弹出）。创意尺寸的统一化，对于在线广告市场淡化广告位概念、推广受众定向有着非常根本的促进作用。中国市场与此对比，由于广告位尺寸非常复杂，因而各个网站之间的壁垒较高，非常不利于定向广告和程序采买的发展。

（2）视频广告标准VAST（digital video ad serving template）。由于视频广告创意和展示形式比较复杂，消耗资源也较多，IAB 制定了一套统一的 XML Schema用于向在线视频媒体投放视频流内的广告，并对其用户进行规范化的描述，这一标准实际上减少了进入视频广告领域的技术障碍，使得视频广告市场规模快速发展成为可能。

（3）通用实时竞价接口标准OpenRTB。实时竞价的技术我们将在第 6章中介绍，简单来说，这种采买方式是为了方便需求方按照自己的受众划分高精度地采买流量。假设各个广告交易平台的实时竞价接口不同，将意味着需求方需要付出几倍的技术成本以完成广泛的市场对接。于是，IAB制定了统一的 OpenRTB标准，将横幅广告、视频广告、移动广告情形下的实时竞价接口做了统一的规范。

2.4.2 美国广告代理协会



美国广告代理协会（American Association of Advertising Agencies, 4A）并不是一个专门从事互联网广告的组织，而是线上线下各种广告，特别是品牌广告的代理商在美国的行业协会。4A 公司向其会员代理公司约定，至少要向广告主收取一定比例的服务费用，这一方面是为了避免行业内的恶性竞争，另一方面也是确保广告代理公司能够站在广告主的利益角度考虑问题，而后一点对于市场的长期健康发展是有很大帮助的。4A公司的典型代表有奥美（Ogilvy&Mather）、智威汤逊（JWT）、麦肯（McCann）等。值得注意的是，由于4A是一个美国协会，因此严格意义上的4A公司都是美国公司，不过对于另外一些国际影响力较强、业务方式和准则与其类似的非美国广告代理公司，我们也往往都将其归为广义4A公司的行列，典型的例子如日本的电通（Dentsu）公司。

2.4.3 美国国家广告商协会



美国国家广告商协会（Association of National Advertisers, ANA）是一个广告主的协会，也是最彻底地代表需求方利益的组织。其会员多是AT&T、宝洁（P&G）以及NBA等这些拥有大量广告预算的广告主。ANA对广告主利益的维护可以从一件小事中得到体现：在微软宣布考虑在IE10支持限制第三方cookie滥用的“do not track (DNT)”协议时，是ANA明确声明对这样的计划表示反对，因为这样将会使得在线广告市场精确投送广告的能力受到很大影响，而这显然是与广告主的利益相违背的。

2.5 延伸思考

1.CPA/ROI的结算方式看似对广告主有利，实际也催生了一些变形的推广方式。请考察你接触到的CPA/ROI渠道，并研究其与CPM或CPC渠道的关键差异。

2.对广告产品而言，优化利润与优化ROI有什么区别？

3.在社交网络中发生的传播营销过程与典型的展示、点击、转化流程有很大不同，请对此建立一个合理的模型，并分析其中的关键量化指标。

注 释

[1].数据来源于 IAB (<http://www.iab.net>)、艾瑞 (<http://www.iresearch.cn>) 等机构的市场调研报告。

[2].资料来源于 <http://www.lumapartners.com/introducing-the-marketing-technology-lumascape/>。

[3].2012年年底，淘宝发布声明，宣布终止与以现金形式返利的淘宝客网站的合作，这正反映了市场对此问题的认知正在趋向一致和理性。

[4].由于 CPM 是千次展示的收益，因此 eCPM 实际上还要乘以 1000 才能与其相比较，为了表达简单起见，我们在本书中略去 1000 这一固定系数。

[5].当然需求方不会完全按照其点击价值来出价，而是会寻求更低的价格以获得套利空间。因此，如何在市场机制上避免广告主积极地调整出价，以促进市场竞争的激烈程度，是竞价体系设计的关键。我们将在第 5 章中讨论这一问题。

第二部分 在线广告产品逻辑

第3章 在线广告产品概览

在线广告不同于传统线下广告，其产品形式（这里说的并非创意形式）相当丰富。在1.4节中，我们对在线广告的整体发展历程有了初步了解，如果从产品的视角将这一过程提炼出来，可以用图3-1来示意。

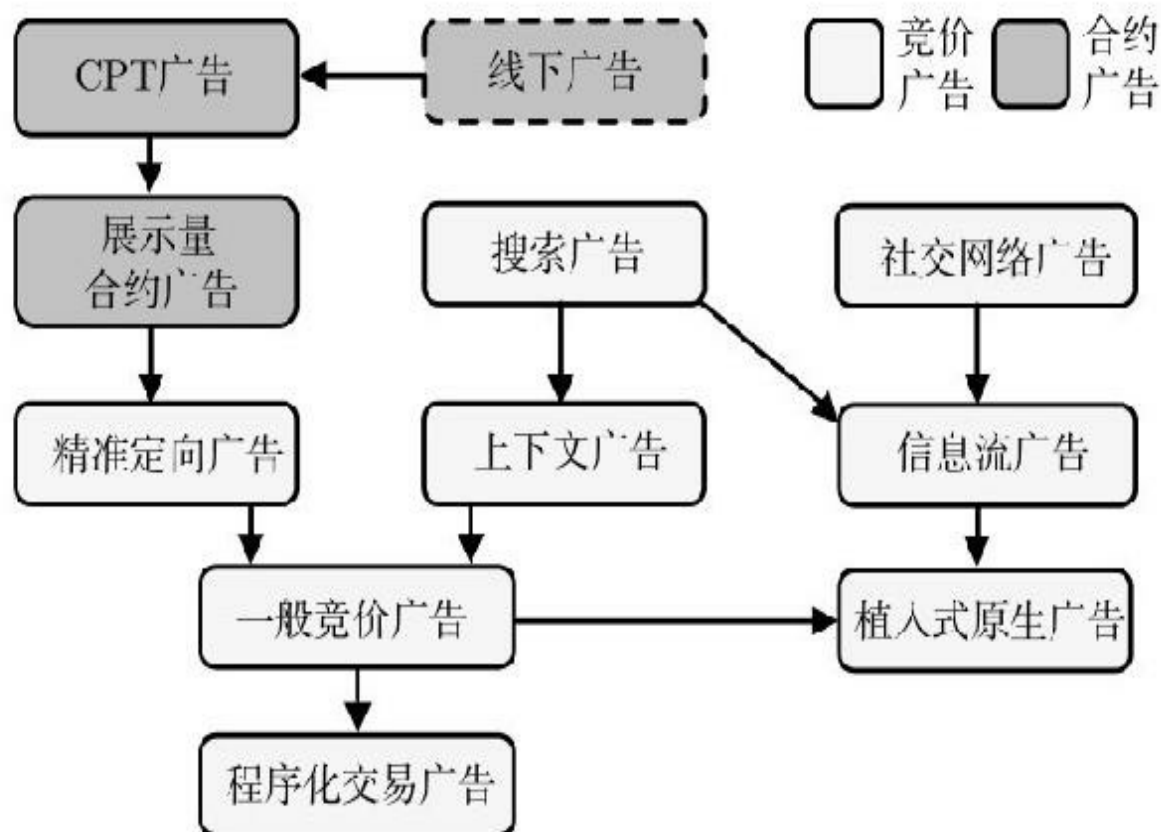


图3-1 在线广告产品进化示意

我们将图3-1中的产品分为4个部分，这些将在后面几章中分别介绍。

（1）合约广告产品：它由线下广告的交易形式衍生而来，又可以分为按照时段售卖的CPT广告和按照约定展示量售卖的CPM广告。这类广告产品主要服务于后续效果不宜直接衡量的品牌类广告主，在门户网站和视频网站较为常见。

（2）竞价广告产品：其最重要的形式是搜索广告，其产品形式为对搜索关键词的竞价。这种广告在拓展到站外展示广告流量时，演进成了对页面关键词或者用户标签竞价的产品形式，也就是ADN。竞价广

告的商业逻辑与合约广告完全不同，也是解决效果类广告需求的关键产品形式。

(3) 程序化交易广告产品：竞价广告的进一步发展催生了实时竞价的交易形式。实时竞价使得需求方可以更灵活地划分和选择自己的目标受众，也使得更广泛的数据使用和交易迅速发展起来。以实时竞价为核心的一系列交易方式逐渐演变为机器之间以程序化的方式完成广告交易决策，因此，这类产品称为程序化交易广告产品。

(4) 原生广告产品：广告的产品体系除了自身的演进，另一个重要课题是如何处理与非商业化内容的关系，让广告与内容尽可能以“原生”的方式共存。搜索广告和社交网络信息流广告对此作了非常有价值的探索，不过这样的原生广告在规模化和交易化方面也遇到了很多的问题。然而，随着近年来移动设备对于原生广告的强烈渴求，像标准ADN那样的原生广告平台正在快速发展中。

对于每一种广告的产品形态来说，都可能会有三个主要的组成部分：面向需求方的接口、面向供给方的接口、中间的投放系统及匹配策略。根据产品的不同，其中接口的形式可以是面向人工操作的界面，也可以是机器间通信的接口。这些广告中的产品环节，都属于商业产品的范畴，我们将先对这类产品设计的主要原则以及需求方和供给方界面的一般情形进行介绍，为后面几章中按上述的四个部分分别介绍广告产品做准备。

3.1 商业产品的设计原则

商业产品指的是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互联网广告产品。其他的一些面向客户的互联网产品，如客户关系管理（Customer Relation Management, CRM）、网站分析（Web Analytics, WA）以及后面要提到的数据管理平台（Data Management Platform, DMP）等，也属于商业产品的范畴。商业产品的设计和运营有着一定的共性原则，我们有必要在这里提一下。

互联网是一个产品驱动的行业。熟悉产品设计的读者应该了解，用户产品演进的根本驱动力是人们追求方便的天性。因此，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。

然而，如果有机会参与商业产品的设计和运营，你就会发现，其中关注的重点和运营的方式有相当大的区别。有时候良好的用户体验，并不一定能带来一款商业产品良好的口碑或市场上的成功，这是为什么

呢？简单来说，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。例如，广告，其使用者不论是媒体、代理或广告主，都是为了优化自己的利润。因此，对这类产品的选择标准是客观的，也是可衡量的。Google AdWords之所以为大量的中小广告主广泛采用，主要原因并不是因为 AdWords 的使用便捷性远超同类产品，而是因为其推广效果有目共睹。因此，商业产品的任何一项功能改进，只要能带来其对应商业目标的提升机会，即使在使用流程上引入一些不便，也是可以接受的。这样的产品原则带来的结果正如图1-9所示，整个广告市场的交易环节越来越复杂，使用门槛也越来越高，这与用户产品简化的大趋势是非常不同的。

在优化既定商业目标这一商业产品的总体原则下，我们在商业产品运营过程中有一些需要注意的关键点。

（1）相对于产品功能，要特别关注产品中的策略部分。策略本身是商业产品非常关键的环节，以广告为例，竞价中的机制设计、冷启动时的数据探索、受众定向的标签体系都是产品策略需要考虑的内容。策略上看似简单的调整往往能带来广告系统收入上巨大的变化。与一般产品不同，这些策略的制定既需要对于广告市场深入的了解，又需要许多扎实的基础知识。对于刚开始进入广告产品领域的读者，将关注点集中在这些“看不见的产品特征”上是需要下大工夫体会和实践的。

（2）要特别关注数据，让运营和产品优化形成闭环。由于商业产品的目标是确定和可优化的，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。同时，新产品功能的规划也要在洞察历史数据和其他用户产品数据的基础上进行。从数据分析开始，以数据结束，这样的闭环式迭代是最适合商业产品的开发模式。

（3）当然，在所有与使用者打交道的产品界面上，用户产品追求便捷性的设计原则依然非常重要。不过在商业产品中，实现功能以外过于新奇、炫酷的产品外观和交互模式是应当避免的。

由于本书探讨的是计算广告这一典型的商业产品，我们也会将重点放在广告投放、交易、策略、数据使用和交易等产品环节上，而对于广告系统与需求方或供给方的界面接口，将只在下面作简要的介绍。

[3.2 需求方层级组织与接口](#)

由于广告市场的复杂性，一个广告系统的需求方有可能是广告主、代理公司、TD或者DSP。无论面对哪种需求方，一般来说，我们都需要一个操作界面让其对预算、广告投放条件和其他策略做设置。

一般来说，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，广告的层次分为广告主、广告（推广）计划（campaign）、广告（推广）组（ad group）、广告创意（creative）等几个层级，参见图3-2。其中广告主层级管理一个广告主的通用信息，而其他三个层级则与具体的投放管理相关，我们来分别介绍一下。

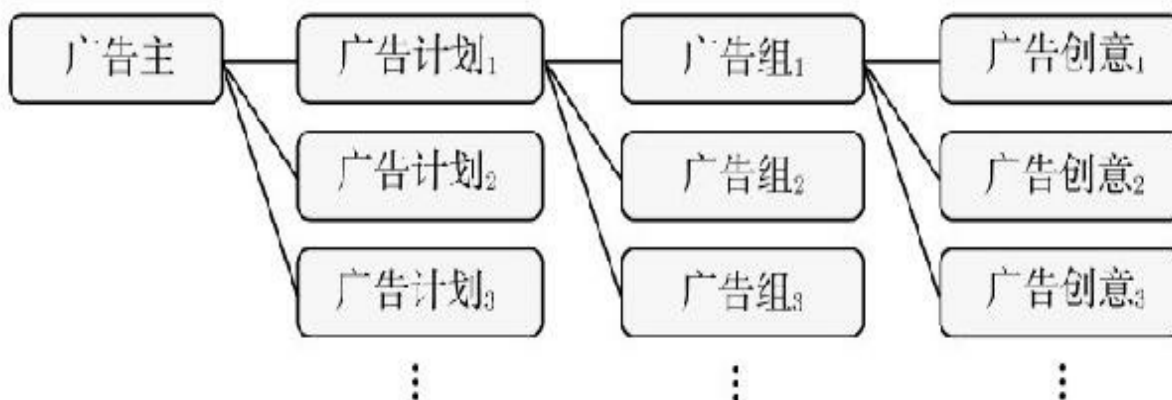


图3-2 广告层级关系示意

（1）广告计划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，参见图3-3中的示例。除了这些信息以外，图中还有另外的一些相关设置，例如，（a）在有多个广告产品可供选择时，要在广告计划中确定投放的是哪个产品，如图中的“选择投放网络”部分；（b）预算的分配策略（pacing），即图中的“预算分配控制”部分。对于大多数广告计划，较为均匀地分配预算可能比较合理，但是对于游戏开服、移动应用冲榜等类型的推广，预算的集中花费则很重要。

设置推广计划

* 推广计划名称：

加载现有计划的设置：

* 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： 2014-09-17

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](#)

图3-3 广告计划设置示意

(2) 广告组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，参见图3-4中的示例。广告组最重要的功能是设置各种各样的定向条件，因而是广告效果优化的关键层级。另外，对媒体的选择也可以认为是一种定向条件。值得注意的是出价——在后面我们将要介绍的竞价类广告产品中，出价是由广告主自行设置，而非预先约定。而对于出价与能获得流量的规模和质量的关系，广告主很难有直觉上的认识。因此，从产品的角度来说，往往需要给出一些有意义的提示，如根据当前出价做的流量预估，或者如图3-4中所示比较直接的“建议出价”。

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：

 ▼

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

用户在哪儿看到我的广告

--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）

图3-4 广告组设置示意

(3) 广告创意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，参见图3-5中的示例。对于文字链类型的创意，基本素材包括标题、描述等内容；而对于图片类型的创意，则直接上传图片素材。另外广告创意的必要设置还应包括展示和点击的监测地址等。为了方便广告投放人员直观地看到创意的展示效果，往往还会提供创意预览功能（如图3-5右侧的“预览区”）。

关于需求方设置管理的细节功能，在不同的产品中可能会有较大的差别，但是这样的4层级组织方式是比较通用的。除了业务管理的便捷性，这样的层级结构还为数据统计和建模提供了天然的、合理的层级结构，让新创意的冷启动问题变得容易一些。在后文介绍各类广告产品时，我们会重点关注其商业逻辑和产品策略，而对于界面上投放管理的功能不再展开讨论。

关键词创意

* 标题:

360合并创意.png

不超出50个字符,超出不显示

||

点击插入通配符

* 描述:

不超出80个字符,超出不显示

||

点击插入通配符

* 点击地址:

http://e.360.cn/

点击地址和显示地址应为同一个网站域名下的网址

* 显示地址:

不超出50个字符,超出不显示

显示URL的设置不会影响网民对网站的实际访问

提交

回到上一步

预览区 (以下预览效果仅提供示意,实际效果以最终投放为准)

360合并创意.png

聚合以最新值的网站资源实现全面覆盖, 众人群划分实现精准受众

http://ju.mediacv.com

图3-5 广告创意设置示意

对于投放规模很大的广告主或代理,所有的操作都通过类似上面的界面人工完成是非常低效的。对于这样的情形,广告平台一般会提供与界面功能相对应的API接口^[1],以便大的需求方用编程的方式进行批量投放和优化。不过,需求方也有可能大量滥用 API来做一些组合或测试投放,这会使得投放系统的压力大增,因此在实际运营中要对带宽或操作次数作一定的限制。

3.3 供给方管理接口

在目前的主流广告交易逻辑中,供给方即媒体对业务的控制比起

需求方来要弱很多。供给端的资源组织主要分媒体和广告位两个层次。其中媒体可以是网站，也可以是移动应用开发者。

媒体的操作比广告主方要简单，一般来说，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。这一接口的功能性示例见图3-6（广告位名称已经被隐去）。

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
	停用	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	0.0706%	¥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
	停用	64x64	2014-06-10	237,264	185	0.0780%	¥61.19
	停用	64x64	2014-06-10	295,444	266	0.0900%	¥91.78
	停用	64x64	2014-06-10	362,509	352	0.0971%	¥111.33
	停用	64x64	2014-06-10	435,934	447	0.1029%	¥147.65
	停用	64x64	2014-06-10	530,349	526	0.0992%	¥169.06

图3-6 媒体广告位管理示意

对于其中的某个具体广告位，根据产品功能的不同，需要的操作功能也不同。但一般来说，设定广告位尺寸、取广告投放代码或SDK以及设定该广告位对广告类型的要求是一些通用的需求。在广告管家或SSP之类的供给方产品中，往往还会有精细的流量分配功能。需要注意的是，广告平台会维护各个广告位对应的域名或应用名，以防其他域名的流量盗用广告位代码。

上面介绍的这种比较简单的供给方管理接口主要用于一般的 ADN 或 ADX，而在媒体需要深度参与的原生广告中，这样简单的对接方式不再适用，关于这部分内容参见第7章。

3.4 延伸思考

1.请研究你熟悉的某一个商业产品，并探讨其与用户产品在设计 and 运营方面的不同之处。

2.如果向需求方提供API来替代一般的UI投放方式，会对广告产品的运营产生什么样的影响？

样章到此结束

需要完整版

扫下面二维码



或加微信：shuyou055

免费领取

第15章 其他广告相关技术

在本书的前面章节中，我们以在线广告市场上产品形态发展的过程为主线，对按合约售卖的广告系统、非实时竞价的广告网络和实时竞价的程序化交易市场作了介绍，并对其中关键计算技术进行了深入讨论。除了这些骨干性的技术以外，在线广告中还有一些非常重要的外围问题，本章将对这些问题展开讨论，目的是为了让读者对实际广告系统的各个环节都有实际的认识。

在前面的章节中讨论的所有技术和算法，其核心都是为了优化广告效果。不过前面谈到的广告效果优化思路，基本上集中在受众的选择的角度，而在实际的在线广告中，还有一项对广告效果影响巨大的技术，那就是创意优化。创意优化与受众优化性质有所不同，因为创意的改变实际上也改变了广告要表达的诉求。如何在基本的宣传诉求可比的前提下，结合受众定向对创意做调整，这是广告系统不能不考虑的重要问题。

广告系统运营中另一个必须考虑的问题是建立一个灵活的实验框架（experimentation framework）。由于各种策略、算法、架构的调整，通过线下评测和模拟都很难完全反映线上的变化，因此，需要有一个线上的实验系统来确定其有效性。线上实验系统的原理很简单，无非是从实际流量中分出一定比例用于实验方案。不过，由于同时测试的方案个数可能比较多，如何在一个框架中进行更多的测试是工程中提高广告系统进化效率的关键。

还有两个与广告效果的度量相关的问题。一是如何在流量中去除那些恶意的和非主动的流量，这部分称为流量保护（Traffic Protection, TP）。其中去除恶意流量的反作弊问题，由于是一个“道高一尺、魔高一丈”的动态博弈过程，因此并无确定不变的技术和算法，不过也有一些原则和基础方法可以遵循。二是需求方站在自己的利益角度对广告效果的核实性度量，这称为广告监测。这两个问题其实有着相当深入的联系，在今天程序交易和受众定向大量被使用的在线广告市场中，这两个问题在一定程度上可以结合起来考虑，并催生了所谓广告安全的问题和相应技术。

受众定向和程序交易广告的另一个重要影响是用户的行为数据存在在不同的广告产品之间泄露的可能。因此，隐私保护技术与其对立面，

即去匿名化技术，从正面或者负面的角度都与在线广告有着密切的联系。关于隐私保护相关问题及其在广告中的可能应用，也将是本章涉及的内容。

15.1 创意优化

创意对于广告效果的影响无疑是巨大的，然而我们不能把调整创意带来的效果等同于受众定向产生的效果。因为随着创意的改变，广告表达的诉求已经发生了变化，其点击行为也就不再与其他创意完全可比。可以通过一个例子来理解这个问题：假如有一个保险类型的广告主，将一个宣传公司品牌和实力的品牌型创意变成一个用户填写车险申请的表单式创意，如图15-1所示。毫无疑问，后者的点击率会大幅度上升。但是实际上，这两个创意向用户传达的诉求有着相当大的区别：前者的目的是向潜在用户渗透性地宣传品牌的定位，以利于将来长期的用户转化和利润空间；而后者的目的则是短期内的转化效果，但对品牌特质的宣传有所不足。

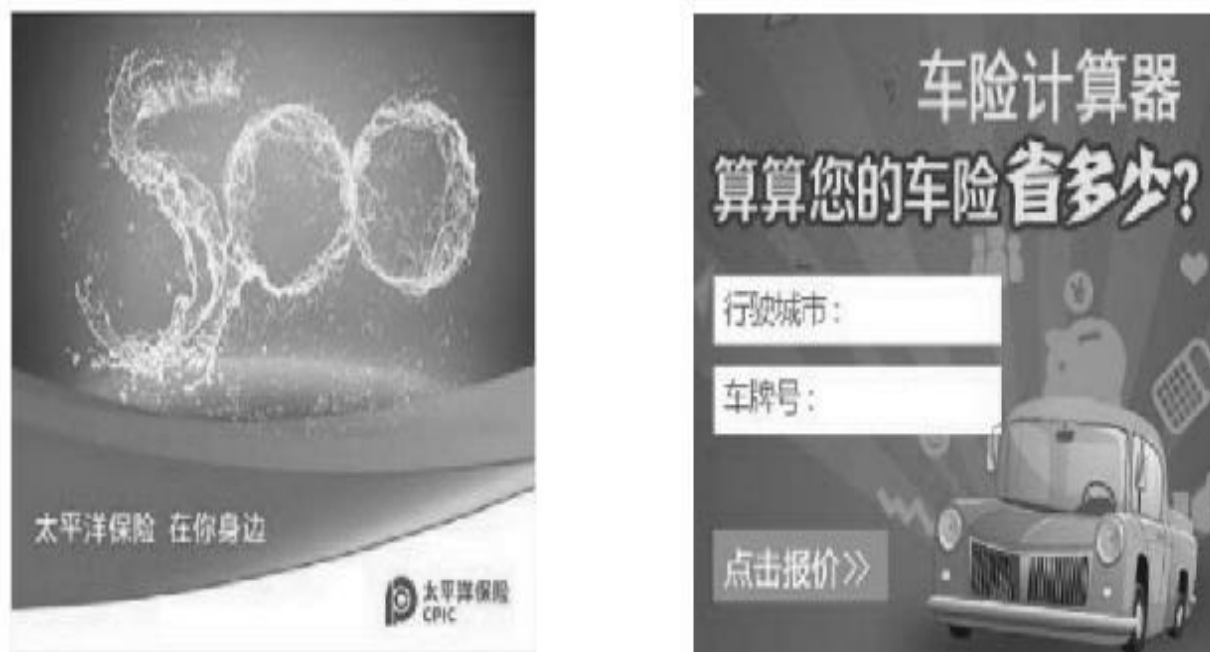


图15-1 品牌型创意（左）和效果型创意（右）

因此，我们重点讨论在广告的基本诉求保持相对稳定的前提下，如何调整创意以提高效果。

15.1.1 程序化创意

创意优化的一个重要原理是我们在第 2 章广告有效性模型中介绍过

的原则：为了提高用户的关注程度，需要将向用户推送此广告的关键原因在创意中明确表达出来。由于推荐原因众多，这样的创意优化往往需要用程序自动进行，而不是预先做好大量的素材。类比于程序化交易，我们把这类思路称为程序化创意。下面就程序化创意的思路给出一些示例性建议。

（1）地域型创意。地域定向是根据用户的地域信息投送相应的广告，如果能将明确地域指示性的内容体现在创意上，往往会对效果有非常直接的帮助。例如，如图15-2所示，对同样一个汽车广告，对北京和上海地域的受众，分别加上当地经销商的联系电话。显然，对每个城市制作一版独立的素材是不经济的，应该在投放时动态加入与地域相关的信息。



图15-2 同一汽车广告在北京（左）和上海（右）的地域型创意

（2）搜索重定向创意。根据用户的搜索行为提供的重定向图片广告，如果能明确标示搜索词，往往更容易唤起用户的注意力和兴趣。因此，可以采用图15-3所示的创意形式，将用户曾经的搜索词放在图片下方的搜索框中，现在这也需要投放系统在线自动完成。



图15-3 搜索重定向创意示例

(3) 个性化重定向创意。这种产品在6.4.3节已经做过具体介绍，其中展示的单品都是在线动态决定的（参见图6-7），而创意也是在线合成的，这也是一种程序化的创意。

在线广告的服务对象由传统的品牌广告向效果广告发生了倾斜，传统的由设计人员主导的、品牌形象驱动的创新生产模式也一定会越来越多地加入机器决策的、效果导向的内容。因此，程序化的创意优化模式应该得到足够的重视。

[15.1.2 点击热力图](#)

在找出创意设计的问题、优化效果等方面，点击热力图是一个非常重要的工具。点击热力图是将某一个创意各位置被点击的密度用热力图的方式呈现出来，帮助创意优化者直观地发现和解决其中的问题。一般来说，创意中的若干主要信息聚焦点应该会比较集中地吸引用户点击。如图15-4所示^[16]，在创意中人物的眼神发生变化时，用户关注和点击的热点也有很大的区别。显然，在这样的点击热力图指导下，创意的迭代优化可以半定量地进行，并且更加有目的性。



图15-4 创意中人物的眼神对点击热力图的影响

不过，在程序化创意的影响下，点击热力图的使用有一些障碍：由于在线时会对创意的部分内容作修改，这种情况下叠加在一起的热力图有时无法反映细节问题。不过对于创意中一些固定元素的优化或动态模块整体的效果评估，热力图还是很有帮助。

15.2 实验框架

无论是广告系统调整算法、架构，还是投放时调整创意、定向策略，都需要依赖线上的实际流量测试来确定其真实效果如何。切分部分流量用于测试并非难事，但是一个实用的实验框架需要尽可能多地同时容纳多组实验，以提高流量利用效率和产品进化速度。

设计这样一个实验系统的关键，是利用系统模块之间的相对独立性，用分层的结构来扩展实验容量。在参考文献 [69] 中，作者给出了比较典型的一种分层实验框架的架构，如图15-5所示。在这一架构中，不同的实验参数被分置于不同的实验层中，一般来说，可以按模块划分这

些实验层，例如在广告系统中，按检索、排序和展现将相应的参数划分成三层，每层都可以将流量切分成不同的测试子集或域。显然，在这种分层结构下，不同层上的实验是可以共享流量的，这样就大大提升了同时进行的实验数目。另外，系统还预留了一个非重叠测试域

（domain），目的是方便有时需要进行的、联合调整各层参数的一些特殊实验。除了实验层以外，此实验框架还涉及了专门的发布层，用于将实验通过的参数逐渐灰度发布到全流量上。同一个参数，只能出现在一个实验层和一个发布层中，而优先级关系是优先采用实验层参数，其次是发布层参数，最后是默认参数。这样的兼顾流量实验和灰度发布的实验框架在实践中能够满足大部分情形下的需求。

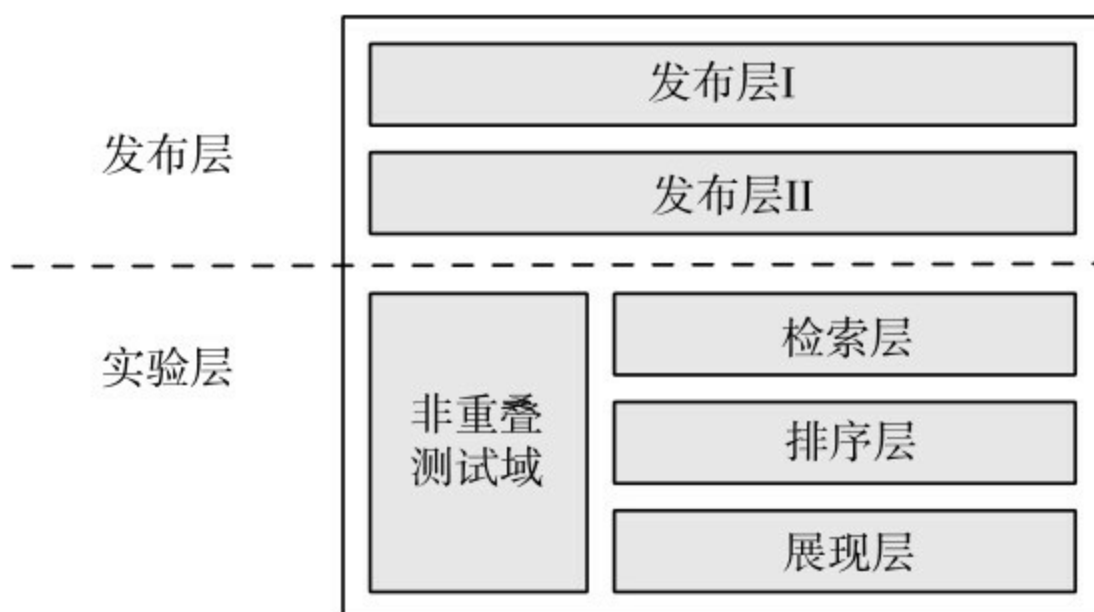


图15-5 分层实验框架架构示意

那么在每一层中，流量是如何随机被分到各个域中的呢？对于广告系统而言，按照每次展示做随机分配是不合适的，这是因为多次广告展示之间的相关性会对测试的结果产生影响。正确的做法是按用户划分，即每个用户的广告展示请求都被固定地发送到同一个域中。这样做的目的是使得一个广告策略的高阶或长期影响能够真实地表现出来。

15.3 流量保护和效果监测

流量保护包括两方面的内容：一是对非主动用户访问行为，主要是爬虫行为的去除；二是反作弊，即对于恶意展示或流量来源的去除。从技术难度上看，我们主要关注的是后一个问题。反作弊的检测方案，在

在线投放和离线统计中都有需求，并且是广告计价和其他数据统计的基础步骤。另一个在线完成的辅助任务是为广告主服务的广告监测服务，它为了完成广告的展示和点击计数，当然也需要一定的反作弊处理，因此这两个问题有着密切的联系。

15.3.1 反作弊

由于广告有很多的相关利益方，因此除了那些无恶意的、机器产生的流量，也会经常遇到欺骗性的展示或点击行为，即作弊。作弊的具体手段五花八门，无法一一列举，并且随着反作弊手段的进步而不断进步。要在面对广告作弊时快速找到思路，需要先搞清作弊者的目的。我们知道，广告活动是广告主、媒体与用户之间三方交互的行为，因此广告中的作弊行为主要有以下三种情况。

(1) 媒体作弊。媒体是广告活动的主要受益者，因而作弊的动力也最强。由于大多数广告网络与媒体之间是按照点击的价格来结算，因此点击作弊是最为常见的，当然也存在为了满足CPM订单量的需求而对展示进行作弊的情形。这种作弊的花样繁多，既有将广告展示和点击代码放在非约定位置上或非用户自然行为产生的流量上的方式，也有通过将广告位与内容靠得很近甚至相互重叠来骗取点击的方式。

(2) 广告平台作弊。广告网络或广告交易市场这样的广告平台也有制造虚假点击，以获取更多分成的目的。而 DSP这样的需求方广告产品，除了混入劣质流量的广告展示、制造虚假点击以外，还会通过一些作弊手段为广告主带来虚假转化，以满足效果考核的要求。

(3) 广告主竞争对手作弊。某些广告主的竞争对手，会通过技术手段大量消耗该广告主的预算，达到降低其广告效果的非正常竞争目的。与媒体作弊不同，广告主的竞争对手很难通过控制广告展示的方式来作弊，而是通过多次重复点击广告的形式来作弊。由于通过单一 IP或 cookie大量点击广告很容易被发现，作弊方往往会通过频繁清除 cookie，改变 IP，甚至通过木马控制多台用户电脑来达到作弊的目的，当然，这样的作弊手段也被媒体或广告平台所采用。

单一IP或cookie在大量展示或点击的作弊方式是最容易去除的，只需要一定时间段内的展示或点击设定合理的上限，进而发现那些显著超过上限的 IP 或 cookie 并加入黑名单即可。

对于更复杂的，通过控制多台电脑来产生假点击的作弊方式，上面提到的点击热力图也是一个很有用的反作弊工具：正常的用户点击，在创意上的位置分布往往呈现与创意关键区域相关的比较自然的分布；而机器产生的用户点击，其分布要么过于均匀，要么过于集中，很容易与

自然点击分布相区别。图15-6中给出了一个广告创意正常的点击热点分布与作弊的点击热点分布的示例，左侧是自然点击的热力图，右侧是有作弊行为的点击热力图。可以看出，除了自然点击区域外，还多了一些集中且均匀的点击分布，这些明显不符合正常用户的行为特征，可以认定为作弊行为。需要说明，认清作弊者的身份和动机对于用好点击热力图反作弊很有帮助，因为这关系到如何分解某个创意上的流量来绘制点击热力图，以发现明确的作弊信号。



图15-6 自然点击热点图（左）与机器作弊点击热点图（右）对比

除了cookie、IP级别的统计以及点击热点图这些思路以外，如果广告系统能在JavaScript代码或 SDK中收集到更多其他的物理信息，如展示时间、点击时间等，也会对于甄别作弊流量很有帮助。一般来说，在收集到比较充足的特征以后，可以建立一个反作弊的判断模型，用以过滤作弊行为。这样的模型需要有一个在线的实时计算版本，为在线计费和其他实时反馈模块做过滤；也需要有一个更加精细的离线版本，用于

每天处理广告日志，得到最终确认的财务结算数据。由于反作弊特征和模型是广告系统高度保密的模块，在此不展开讨论其细节。

流量劫持

除了制造虚假展示和点击的作弊行为，在广告市场上还存在着通过非法手段获得广告展示或点击的准作弊行为，而其中最典型的情形就是流量劫持。

所谓流量劫持，就是在无权投放广告的地方强行投放，或者改变广告创意甚至落地页的内容。一般来说，只有一些网络底层服务的提供商，如 DNS、CDN、电信运营商等，才有能力进行这种劫持。流量劫持并非互联网广告的新问题，在电视广告中也存在这种现象，如图15-7中强行加入的滚动字幕广告。尽管流量劫持不是合法的商业产品，但在中国及东南亚等地区，这确是一个不能忽视的广告渠道，在服务于效果类广告主时，必须要了解其客观存在并加以应对。



图15-7 电视广告流量劫持示例

下面通过几个例子了解流量劫持的具体手段。

(1) 信道弹窗。通过电信运营商对信道的控制能力，在用户上网时强行向下行内容中插入弹窗广告创意。这种形式在 PC和移动设备上

都存在，虽然 CCTV在 2013年“3·15”晚会对这种灰色广告渠道进行了曝光，但是至今仍然广泛存在，并且在移动设备上大有愈演愈烈之势。

（2）创意替换。创意替换仍然是通过电信运营商的信道，将某些网站上的广告创意直接替换为其他创意，显然，这是一种比信道弹窗更加粗暴的劫持行为。

（3）搜索结果重定向。由于搜索是高商业价值的流量，将搜索流量导向某些搜索引擎，可以获得其收入分成。因此，存在一种流量劫持手段，当用户在搜索引擎 A 输入某关键词以后，搜索的结果页会变成搜索引擎B提供的。或者在淘宝这类电商垂直搜索中，改变结果商品的排序或落地店铺。这虽然不是直接的广告劫持，但本质也是一样的。

（4）落地页来源劫持。这是最为简单粗暴的一种劫持方式，它并不投放广告，而是在用户访问广告主落地页时，直接在URL上加上广告来源代码。比如，当用户访问<http://mkt.mbaobao.com/a-hotalfshell1219>这个广告主页面时，将其修改成http://mkt.mbaobao.com/ahotalfshell1219?utm_source=*，这样广告主就会将其统计成某广告渠道带来的访问。

这些流量劫持手段中，前三种手段主要损害的是媒体利益，而对广告主来说，流量本身则是真实存在的。而第四种就是一种彻底的作弊行为了，损害的是广告主的利益。由于有这种非法劫持流量的存在，严重扰乱了正常的效果广告和程序化交易市场，客观上在中国形成了一个劣币驱逐良币的环境——毕竟努力提高技术和算法水平远不如去买低成本的非法流量效果来得快。因此，我们在这里呼吁大家，一起来抵制这种劫持流量，真正从产品技术方面扎扎实实地做一些有利于行业进步的事。

15.3.2 广告监测

在线广告区别于线下广告的重要特征就是可监测性。从需求方来看，存在切实的需要，委托某家第三方监测公司对实际发生的展示或点击数目进行核对，以确保自己的利益。不过监测的主要需求存在于按 CPT 或 CPM 结算的合约广告中。这是因为在竞价广告系统中，广告主与媒体之间并没有约定的价格，可以根据后续的效果来调整自己的出价，因此对展示或点击的监测并不是强需求。从这里我们可以看出，效果检测主要的服务对象是那些品牌广告主，一般会占有在线品牌广告投放1%左右的预算。

借助监测代码或者SDK实现广告展示或点击的计数并不困难。但是当广告投放基于受众进行时，监测就要困难多了。例如，某广告计划要求在男性用户流量上投放1 000千次的展示，那么如何才能确定投放的

结果满足这一条件呢？一般采用的方案是广告监测提供商用采样加付费的方式收集一个小比例人群上的真实用户属性，然后通过验证这个人群上性别的准确率来反推整体的投放数据。这一方法看起来十分简单，不过由于采样集一般规模不大，在人群分布上与投放人群相比可能存在较大的偏差，因此此方法的关键在于如何对数据进行有效的纠偏。另外，即便采用这样的方案，也只有部分基于人口属性信息的投放是可以验证的，而对于基于兴趣标签的投放，由于对同一个用户并不存在确定的标准答案，这样的监测意义不大。

由于实名社交网络的人口属性信息相对比较准确，又有很大的规模，因此现在的趋势是直接采用社交网络的数据作为标准来进行定向广告的监测。例如，尼尔森就与Facebook合作，推出了基于Facebook人口属性信息的广告监测服务。

当然，广告监测与反作弊有着密不可分的联系，所有展示或点击的计量都必须在去除了作弊流量的基础上进行。由于广告监测是代表需求方利益的服务，一般来说其反作弊规则比媒体方或广告平台更加有严格的动力。

[15.3.3 广告安全](#)

在越来越复杂的广告投放和交易逻辑当中，广告主已经很难像合约广告中那样非常明确地管理自己的投放媒体了。但是，实践中很多广告主又存在切实的需求，希望自己的广告不要出现在一些特定内容的媒体上。我们在第2章中曾经提及，广告预算除了被浪费的那部分，还存在一部分是会产生负面效果的，我们显然希望去掉这些有负面影响的展示。例如，汽车广告主不希望自己的广告出现在有关车祸内容的页面上，也不希望出现在一些低级庸俗内容的网站上。这样的需求，不是简单的广告可以完成的，而是需要专门的服务来保证广告主的品牌安全，这样的问题称为广告安全问题。与广告安全相关的关键技术是广告投放验证（ad verification）和可视性（viewability）验证。

广告投放验证

广告投放验证的作用是确认品牌的安全性，并保证展示的质量。其工作模式是当广告投放到页面以后，如果发现页面的内容不符合品牌安全的诉求，则停止展示广告主的创意，转而展示一个与品牌无关的创意。与广告监测不同，这里的重点不在于计量效果，而在于阻止不恰当展示的发生。

读者可能会疑惑，既然是在不安全的页面上禁投广告，在广告请求到来时直接根据URL过滤不就可以了吗？在实际的广告交易中，由于多

层 `iframe` 的嵌套，有时媒体会伪装自己的 URL 以达到流量以次充好的目的。比如，某些高质量媒体会将自己域名的 `iframe` 嵌套在其他小网站上以冒充自己的流量，从而获得高的溢价。因此，必须要在广告投放时，进行`iframe`穿透，实时判断投放页面的顶层URL，才能进行广告投放验证。而页面的内容则可以采用第12章中的半在线抓取系统来获得。

当有了一些历史经验以后，也可以采用投放前的验证方案，也就是对那些历史上发现不符合品牌安全策略的 URL 或广告位直接不参与广告交易，这样可以进一步节约服务成本。

可视性验证

品牌广告主的另一个常见诉求是广告展示的曝光程度。显然，出现在第二屏的广告位比第一屏的广告位曝光程度要差很多。这个问题也属于广告安全的范畴。

可视性验证的技术方案一般是判断浏览器是否对广告创意发生了渲染过程，如果没有，那么这次展示实际上不是可视的。解决可视性验证，需要对各种浏览器做充分的针对性测试，目前的技术水平已经可以做到对 95%以上的浏览器内广告流量进行可视性验证；但是在移动应用内广告中，目前还没有很好的检测办法。

可视性验证同样有投放前的方案，也就是对那些可视比例很低的广告位直接不参与广告交易。

15.4 隐私保护和数据安全

广告是一个典型的个性化系统，它需要大量使用用户的行为数据进行受众定向，同时，在广告市场中还存在着数据交易的产品。无论是受众定向还是数据交易，都需要谨慎地考虑对行为数据的使用是否会泄露用户的隐私；同时也要考虑拥有数据的利益方，特别是广告主，是否在广告市场中被平台或竞争对手获得和利用了自己的关键商业数据。

15.4.1 隐私保护问题


隐私问题讨论的是用户个人信息的安全性，不过对这个问题，市场上存在着一定的认识误区。实际上，隐私保护除了关心那些成批的用户资料泄露意外，更大的挑战是针对熟人的隐私窥探，即窥探者在了解被窥探者一些背景信息的基础上，即用这些背景信息进一步试图获取其更多的隐私信息。后面一种挑战由于可能是人工与机器相结合，而且对成本往往不敏感，给隐私带来的风险也最大。一个最生动的例子，可以参见“清华学生用自拍照推理出王珞丹住

址”（<http://news.cntv.cn/ent/20110819/105071.shtml>）这篇报道，在这个例子里，一名清华学生通过分析王珞丹的微博发帖和照片，准确地得到了其住址这一隐私信息。下面我们来具体看一下隐私保护的问题和原则。

1. 隐私保护基本原则

隐私保护在互联网个性化服务发展的很早阶段就得到了大家的重视，欧盟的 A29委员会也对此问题做过深入的研究和规范。目前，工业界有以下一些共识性的隐私保护原则。

(1) 要严格避免使用个人可辨识信息（Personal Identifiable Information, PII）。PII 是最为重要的隐私信息，它指的是那些被获取后可以被方便地定位到具体人的信息，例如身份证号、电话号码、电子邮件地址、家庭住址等。这些信息一旦被恶意获取，会给当事人带来非常大的不便和潜在风险，因此需要无条件地严格保护。需要说明，广告系统中经常使用的用户标识，如cookie、IMEI等，由于不具有方便地辨识人的作用，因此不属于PII。

(2) 用户有权要求系统停止跟踪和使用自己的行为数据。如图15-8所示，当向用户提供行为定向广告时，广告提供商应该给出明确的提示，如图中的广告创意右上角的“AdChoices”。如果用户对自己的行为被使用感到不满，可以通过此入口得到更多的详细说明，并且可以通过说明页面上的“Opt Out”操作通知系统停止跟踪和使用自己的行为数据。系统得到通知后，必须停止记录该用户的行为信息，也不再向其投放行为定向广告。这样的入口给了用户决定是否接受个性化广告的权利，对特定情形下的隐私保护非常重要。



The screenshot shows an AOL advertisement with two products: a tablet and socks. In the top right corner of the ad, there is a small 'AdChoices' icon. To the right of the ad, a larger panel titled 'Ad Choices: Why Did I Get This Ad?' provides more information. This panel includes sections for 'Who Provided This Ad?' (identifying AOL Advertising), 'Where Can I Learn More About AOL Advertising?' (linking to a privacy page), and 'How to Opt-Out' (showing an 'Opt-Out' button). The 'Opt-Out' button is highlighted with a red box. Below the button, text explains that clicking it will stop the user from seeing ads tailored to their interests.

图15-8 Ad Choices示例

(3) 不应长期保留和使用用户行为数据。即使用户同意接受行为定向广告，广告平台在数据的使用和存储上也应该有所节制，长期保留用户行为对受众定向价值有限，同时又加大了数据泄露的风险。因此应该只保存一段时期以内的行为数据，过期的数据如果并非与业务直接相关，物理上不应再存储。

(4) 工程上还需要特别注意权限的严格分配和最小数据访问的原则。工程师在调试程序时，最好是使用采样过的、关键信息被匿名化处理过的数据子集，而在生产系统中通过特别的密钥访问原始数据全集。而不需要开发数据处理程序的人，包括管理层，也不应当有数据访问的权限。

上面的这些原则相当基本，也非常重要，是广告系统、推荐系统在用户行为数据使用中首先要遵循的。不过，这些原则并不能解决一些深层次的数据隐私泄露问题，对此我们还需要更加深入地认识与并给出对策。

2.Quasi-Identifier 与 K-Anonymity

PII是可以辨识个人身份的隐私信息，那么是不是非 PII的信息就无法辨识身份了呢？举个例子，假设有这样一条用户信息：“姓名：XXX；手机号：XXX；年龄：36；工作地点：上海市携程大厦；性别：男；职位：测试工程师；爱好：羽毛球；月薪：15000元”。其中的“姓名”、“手机号”等 PII已经被隐藏。不过，如果此用户的一个朋友看到这条记录，根据“年龄、工作地点、性别、职位、爱好”这些非PII的组合，还是很容易得知是谁的信息，从而也就得到了“月薪”这一隐私信息。

在上面的例子里，“年龄、工作地点、性别、职位、爱好”这组信息虽然单独看来都无法确定一个人，但组合在一起有可能让熟人确定出对应的人，这样的信息称为“Quasi-Identifier”。由于有这样的Quasi-Identifier的存在，即使没有提供PII，仍然存在比较高的隐私泄露风险，这一点希望引起大家的注意。

有没有什么办法能够降低这一挑战带来的风险呢？简单的思路是将Quasi-Identifier 做一定程度的泛化。例如，将“年龄：36岁”泛化成“年龄：30~40岁”；将“工作地点：上海市携程大厦”泛化成“工作地点：上海市”。如果泛化的结果，使得数据集里的每一组Quasi-Identifier的实例都能找到 K 条与其相同的，那么我们就说实现了 K-Anonymity。显然，当K 的值取得比较合理时，隐私泄露的风险也就降低了不少。

Quasi-Identifier与 K-Anonymity并不是互联网隐私问题的产物，而

是在数据库领域就有的研究。它给了我们很大的启发：当背景信息充分，而这些信息又较为稀疏时，隐私问题的挑战会变得更大。而在以互联网广告、推荐等为代表的个性化系统相关的数据交易中，这样的风险变得空前严峻。

3.稀疏行为数据的挑战

在计算广告这样的个性化互联网应用中，对一个用户的描述不再限于上面例子中的基本信息，而是包括了其大量的行为数据。行为数据的特点是极为稀疏，换句话说，任何两个用户的行为数据都几乎不可能是相同的，也很难通过 K-Anonymity 的方案来解决。那么是否可以通过行为数据来反推用户的隐私呢？答案是肯定的，而且有实际的案例发生。

在著名的Netflix百万美元推荐大赛^[5]中，主办方公布了比赛用的数据库，其中的关键用户信息已经去除了PII，并做了K-Anonymity的处理。不过，用户的观影记录和打分由于是推荐使用的主要数据并未做处理。当数据公布后，恰好有一位关注者在数据中发现了一条记录，从其观看的影片和评价分数来看，集合可以确定无疑是自己的另一位朋友，而同样在这个用户观影记录中，还发现了一些同性恋题材的影片。读者显然明白这意味着什么，实际上，他的朋友不想为人所知的同性恋隐私被这样一个推荐大赛无意中泄露了。由于稀疏的行为数据很难通过简单技术处理模糊用户间的区别，再加上其他一些原因，这一大赛就没有继续举办下去。

在参考文献[60]中，作者对上面的隐私安全问题进行了更加系统的研究，发现通过简单的算法就可以将 Netflix 给出的数据与另一个 IMDb 给出的数据库进行用户身份上的对应，而且准确率相当高。抛开具体的方法不谈，这一研究向我们揭示的问题是：通过稀疏的行为数据，可以比较容易地定位自己熟悉的人，并进而获取其相关隐私信息。这一问题的发现使工业界对隐私安全问题的认识大大提高，也提醒我们在数据交易和洩漏过程中，要特别注意这方面巨大的风险。这也催生了与深度个性化系统中隐私安全相关的差分隐私（differential privacy）^[29]问题的研究。

坦率地讲，稀疏行为数据给隐私保护带来的巨大风险还并没有成熟的解决方案，这无疑将是大规模行为数据利用头上的达摩克里斯之剑。我们在实际的工业实践中，需要对此问题有一定程度的认识和理解，并尽可能考虑到自己的业务过程中与此相关的隐私安全风险。

15.4.2 程序化交易中的数据安全

程序化交易的产生使得在线广告市场可以综合利用需求方和供给方的数据来完成更加精准的广告决策。当然，这样的便利性也是一把双刃剑，在数据得到更加充分利用的同时，RTB中供给方和需求方对于数据安全性的顾虑和诉求也必须加以考虑。

1.供给方数据安全

我们先来看看供给方的数据安全性问题。由于在RTB过程中，ADX需要向参与竞价的DSP广播每次展示的URL和cookie，使得DSP理论上存在规模化监听媒体用户行为的可能。假设有某个恶意的DSP对于能够参与竞价的所有广告请求都以很低的价格参与竞价，目的不在于赢得流量，而在于收集媒体上的用户行为，这就产生了媒体数据的安全问题，我们将其称为供给方数据安全，第6章介绍RTB原理时也曾经提到过这一问题。

供给方的数据安全问题尽管在RTB中确实存在，但是并不是想象中那样严重。可以回顾一下14.1.2节介绍过的询价优化技术：由于带宽的限制，实际上在每次询价时，ADX应该尽可能只向那些最可能赢得竞价的DSP发送询价请求，而那些以恶意收集数据为目的的DSP，在理想情况下应该被挡在大部分的询价以外。

2.需求方数据安全

再来看看需求方的数据安全性问题。在RTB的环境下，由于定制化标签的引入，广告主的第一方数据也暴露在了广告交易的过程中，而这些数据有的是广告主的核心数据，需要认真考虑其安全性问题。为了表达更加清楚，我们用图15-9所示的例子来说明。假设有两个英语教育类广告主“英孚教育”和“华尔街英语”，两者都通过DSP进行重定向访客找回，那么他们分别利用RTB的方式接触到了自己的访客集合。需要注意的是，这里的顾客集合实际上是广告主的私有数据，也是特别具有商业价值的数据，然而，DSP、ADX和媒体都有可能在RTB过程中得到这些访客集合。如果DSP希望制造更加激烈的竞价环境，获得更高的利润，那么它实际上可以将这两个广告主的顾客集合合并在一起，并生产一个相应的用户标签吸引双方来对此标签竞价。这种做法的实质是在竞争对手之间倒卖顾客集合，并且可以通过比较模糊的标签名字（例如为上面两个广告主的访客集合打上“英语教育”的人群标签）非常隐蔽地操作。随着竞价激烈程度的增加，原本属于广告主的利润就向市场其他环节发生了转移，这个问题就是需求方数据安全性问题。

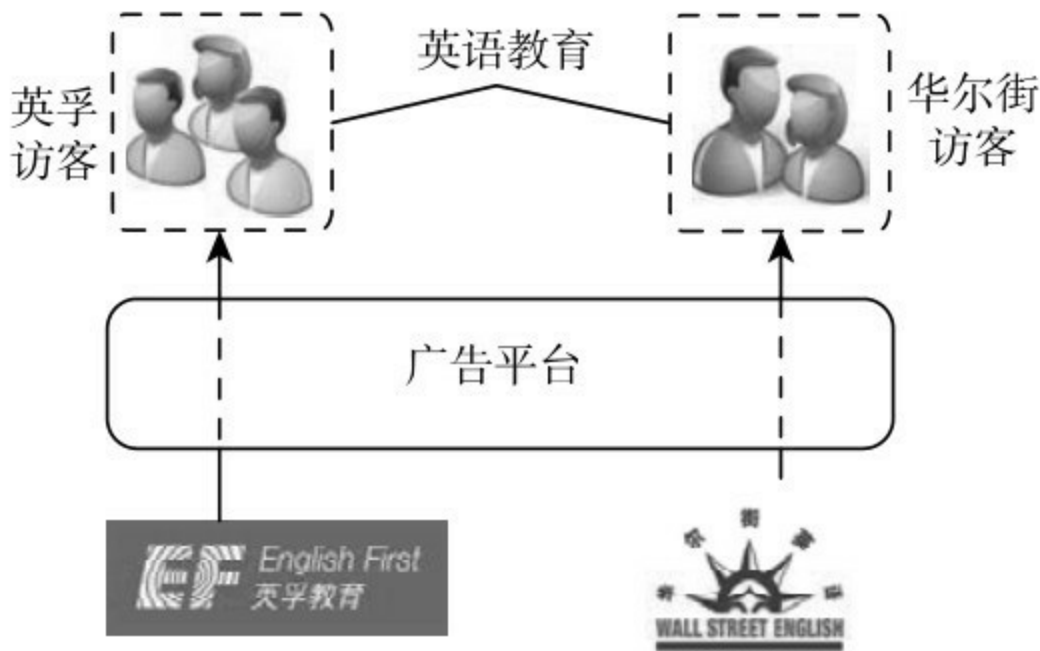


图15-9 需求方数据安全问题示例

需求方数据安全性在某种意义上比供给方数据安全性更加重要，因为这决定了广告主是否可以放心地通过 RTB 进行广告采买。坦率地讲，当前的广告交易市场，对这个问题的重视程度和解决方案都还很不充分。所以要提醒广告主，在广告交易中使用自己的第一方数据时，特别是面对强势的广告平台时，要特别留意数据安全性的问题。

15.5 延伸思考

1. 在移动广告中，流量保护可能会遇到哪些新问题？会有哪些新的方法和特征？

2. 调研一下移动应用下载类广告的转化流程以及相应的监测环节，并比较其与标准展示广告监测的不同之处。

3. 在一个电商平台上的某家网店中产生的用户行为数据，平台、网店和用户应该对这些数据的记录和使用分别拥有什么权利？

/p>

注 释

[1]. 关于为什么要这样做，可以参照 12.3 节中的讨论。

[2]. 不同的时间复杂度与索引的具体数据结构有关：采用哈希表索引时为 $O(1)$ ，采用树索引时为 $O(\log n)$

[3]. 参见参考文献 [13] 中的具体讨论。

[4]. 有时我们需要遍历所有数据或者使用很大内存才能得到目标函数的导数，这种情况实际上在工程中比问题不可导要更常遇到。

[5].关于 Wolfe 条件的几何意义和其他细节, 请参阅 [62] 中更详细的讨论。

[6].地域定向虽然是受众定向的一种, 但由于逻辑比较简单, 在CPT广告中也会售卖, 并且可以采用在前端代码中实现逻辑的方式。

[7].这里没有考虑广义第二高价等定价机制的影响。

[8].Free Disposal指的是给某个合约投送的量超过合约要求是无收益也无损失的, 这一点符合大多数广告合约的实际情况。

[9].多维 y 分布随机变量在归一化以后服从狄利克雷分布, 也就是多项式分布的共轭先验形式。

[10].参考文献[22]用一个通过历史行为决定参数的泊松分布来建模展示数量, 我们认为, 广告并非用户主动行为, 因此不宜用历史行为来预测。

[11].更快地利用用户行为反馈还有其他系统方面的需求, 参见 13.3 节。

[12].实际上, 在排序阶段, 我们都是按照这个参数等于 N_1 来处理的。

[13].搜索广告中查询的过分泛化会对相关性有较大负面影响。因此, 我们不提倡在搜索广告检索阶段利用短时用户标签, 不过在排序阶段可以利用短时动态特征来加权那些用户更倾向于选择的结果。

[14].不过, 由于 EC 计算的不确定性和难以解释, 这样的方案在实际品牌广告中可操作性并不强。

[15].如果 DSP 不能确定代表广告主的利益, 这种深度合作有可能会带来广告主高商业价值数据的安全性风险, 参见15.4.2.2 节中的讨论。

[16].来源: <http://site.douban.com/106407/widget/notes/335509/note/252343905/>。此图是眼球追踪的热点图, 不过考虑到其与点击热力图的一致性, 我们仍用它来说明问题。

第四部分 附录

附录A 主要术语及缩写索引

缩写	英文	中文	参考章节
4A	American Association of Advertising Agencies	美国广告代理协会	2.4.2
ADX	AD eXchange	广告交易平台	6.3, 14.1
	ad group	广告组	3.2
ADN	AD Network	广告网络	5.3, 13.3
	ad placement	广告放置	5.1.3, 13.2.2
	ad safety	广告安全	15.3.3
	affiliate	联盟	5.3, 7.1.4
	agreement-based advertising	合约广告	4, 11
AUC	Area Under Curve	曲线下面积	13.5.5
	auction-based advertising	竞价广告	5, 13
	audience targeting	受众定向	4.2, 12
ASN	Average Show Number		5.1.3
BoW	Bag of Words	词袋	10.1.2
	banner ad	横幅广告	1.3
	bayesian learning	贝叶斯学习	10.3.3
BT	Behaviorial Targeting	行为定向	12.3
	bid term	竞价关键词	5.1.1
	brand awareness	品牌广告	1.2
	call out optimization	询价优化	14.1.2
	campaign	广告计划	3.2
CoEC	Click on Expected Click		13.5.4
CoPC	Click on Predicted Click		14.2.2
CTR	Click Through Rate	点击率	2.3.1

续表

缩写	英文	中文	参考章节
	click value	点击价值	2.3.1
CF	Collaborative Filtering	协同过滤	13.2.1
	compact allocation plan	紧凑分配方案	11.3.1
	constrained optimization	带约束优化	10.2.1
	content as ad	内容即广告	7
CDN	Content Delivery Network	内容分发网络	11.1
	contextual advertising	上下文广告	1.4
	contextual targeting	上下文定向	4.2.1, 12.2
CVR	Conversion Rate	转化率	2.3.1
CPA	Cost per Action	按转化付费	2.3.2
CPC	Cost per Click	按点击付费	2.3.2
CPM	Cost per Mille	按千次展示付费	2.3.2
CPS	Cost per Sale	按销售额付费	2.3.2
CPT	Cost per Time	按时间付费	2.3.2
	creative	广告创意	3.2
CRM	Customer Relation Management	客户关系管理	3.1
	customized audience segmentation	定制化用户标签	6.1, 14.2.1
	data exchange	数据交易平台	6.6.4
DMP	Data Management Platform	数据管理平台	3.1, 6.6.3, 12.5
	demand	需求方	1.2
	demand constraint	需求约束	11.3.1
DSP	Demand Side Platform	需求方平台	6.4, 14.2
	demographical targeting	人口属性定向	14.2.1, 12.4
	direct response	直接效果广告	1.2
	display advertising	展示广告	1.4
	downhill simplex method	下降单纯形法	10.2.2
	dynamic allocation	动态分配	6.5
EDM	E-mail Direct Marketing	邮件定向营销广告	1.3
EM	Expectation-Maximization	最大期望	10.3.2
EC	Expected Click	期望点击	13.5.4
eCPM	Expected Cost per Mille	千次展示期望收入	2.3.1

续表

缩写	英文	中文	参考章节
	experimentation framework	实验框架	15.2
E&E	Explore and Exploit	探索与利用	13.6
	frequency capping	频次控制	11.2.2, 13.5.6
GalP	Gamma-Poisson	γ 泊松	12.2.2
GLM	Generalized Linear Model	广义线性模型	12.3
GSP	Generalized Second Price	广义第二高价	5.2.1
	geo-targeting	地域定向	4.2.1
	gradient descent	梯度下降法	10.2.3
GD	Guaranteed Delivery	担保式投送	4.3, 11.2
HWM	High Water Mark	高水位算法	11.3.4
	hyper-local targeting	精确位置定向	4.2.1
IDFA	Identifier for Advertising	广告专用用户标识符	6.6.1
IIS	Improved Iterative Scaling	改进的迭代缩放	13.5.2
IR	Information Retrieval	信息检索	10.1
	integrated marketing	整合营销	1.2
IMEI	International Mobile Equipment Identity		6.6.1
IAB	Interactive Advertising Bureau	交互广告局	2.4.1
	inventory	库存	1.4
IDF	Inverse Document Frequency	倒数文档频率	10.1.2
	landing page	落地页	2.3.1
LDA	Latent Dirichlet Allocation	潜在狄利克雷分配	12.2.2
LSA	Latent Semantic Analysis	潜在语义分析	12.2.2
L-BFGS	Limited-memory BFGS	有限内存 BFGS	13.5.2
LR	Logistic Regression	逻辑回归	13.5.1
	look-alike	新客推荐	4.2.1, 6.4.1, 6.4.4, 14.2.1
ML	Machine Learning	机器学习	10.3
MRP	Market Reserve Price	市场保留价	5.2.2
MRA	Maximal Representative Allocation	最大代表性分配	11.3.2
MAP	Maximum a Posterior	最大后验概率	10.3.3
ME	Maximum Entropy	最大熵	10.3.1

续表

缩写	英文	中文	参考章节
	mechanism design	机制设计	5.2
	media buying platform	媒介采买平台	1.4
ML	Mixture Model	混合模型	10.3.2
MoG	Mixture of Gaussians	高斯混合模型	10.3.2
	mobile ad	移动广告	7.2
MAB	Multi-Arm Bandit	多臂老虎机	13.6
	native ad	原生广告	7.1, 7.3
	network optimization	网络优化	6.5.1, 14.3
NFP	North Foot Print	北区广告平均条数	5.1.3
	off-site recommendation	站外推荐	6.4.3
	online allocation	在线分配	4.3.3, 11.3
PII	Personal Identifiable Information	个人可辨识信息	15.4
	personalized recommendation	个性化推荐	13.2.1
	personalized retargeting	个性化重定向	6.4.3
	position auction	位置拍卖	5.2
PR	Precision/Recall		13.5.5
	preferred deals	优选	6.2
	premium sales	优先销售	6.2
	pricing	定价	5.2.1
PMP	Private Marketplace	私有交易市场	6.2
PLSI	Probabilistic Latent Semantic Indexing		12.2.2
	programmatic trade	程序化交易	6, 14
	query	查询	13.2
	reach	到达率	2.3.1
RTB	Real Time Bidding	实时竞价	6.1
ROC	Receive Operating Characteristic	接收机操作特性	13.5.5
	remnant inventory	剩余流量	5.3
	retargeting	重定向	6.4.3
ROI	Return on Investment	投入产出比	1.2
RPM	Revenue per Mille	千次展示收益	1.4
	rich media ad	富媒体广告	1.3

续表

缩写	英文	中文	参考章节
	search ad	搜索广告	5.1
SEM	Search engine marketing	搜索引擎营销	5.4
	search retargeting	搜索重定向	6.4.3
	second price	第二高价	5.2.1
SVD	Singular Value Decomposition	奇异值分解	12.2.2
	site retargeting	网站重定向	6.4.3
	social ad	社交广告	1.3, 7.1.1
	sponsored search	付费搜索	5.1
SGD	Stochastic Gradient Descent	随机梯度下降	10.2.3
	stream computing	流计算	9.1, 13.3
	sufficient statistics	充分统计量	10.3.1
	supply	供给方	1.2
	supply constraint	供给约束	11.3.1
SSP	Supply Side Platform	供给方平台	6.5, 14.3
	targeted advertising	定向广告	1.4
TD	Trading Desk	交易终端	5.4.2
TF	Term Frequency	词频	10.1.3
	textual ad	文字链广告	1.3
	topic model	文本主题模型	12.2.2
	traffic forecasting	流量预测	4.3.1
TP	Traffic Protection	流量保护	15.3
	traffic shaping	流量整形	4.3.2
UCB	Upper Confidence Bound	置信上界	13.6.1
VSM	Vector Space Model	向量空间模型	10.1.2
	vertical ad network	垂直广告网络	5.3.3
VCG	Vickrey - Clarke - Groves		5.2.1
	video ad	视频广告	1.3
WA	Web Analytics	网站分析	3.1

参考文献

- [1] AKELLA R, BRODER A, JOSIFOVSKI V. Introduction to computational advertising[C]. Meet-ing of the Association for Computational Linguistics, 2011. <https://classes.soe.ucsc.edu/ism293/Spring09/material/Lecture%201>
- [2] ANDERSON Q. Storm实时数据处理[M]. 卢誉声, 译. 北京: 机械工业出版社, 2014.
- [3] ARENSW. 当代广告学[M]. 丁俊杰, 程坪, 译. 北京: 人民邮电出版社, 2005.
- [4] AUDIBERT J Y, MUNOS R, SZEPESVARI C. Use of variance estimation in the multi-armed bandit problem[C]. Proceedings of NIPS'06, 2006.
- [5] BENNETT J, LANNING S. The Netflix Prize[C]. Kdd Cup & Workshop in Conjunction with Kdd, 2007.
- [6] BERGER A, PIETRA S AD, PIETRA V JD. A maximum entropy approach to natural language processing[J]. Computational Linguistics, 1996, 22 (1) :39-71.
- [7] BERGER A. The Improved Iterative Scaling Algorithm, a Gentle Introduction[M]. Pittsburgh: Carnegie Mellon University Press, 1997.
- [8] BHARADWAJ V, CHEN P, MA W, et al. SHALE: an efficient algorithm for allocation of guaranteed display advertising[C]. Proceedings of KDD, 2012:1195-1203.
- [9] BISHOP C M. Pattern recognition and machine learning. Springer[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 2006, 16 (4) :049901.
- [10] BLEI D M, NG A Y, JORDAN M I. Latent dirichlet allocation[J]. Journal of Machine Learning Research, 2003, 3:993-1022.
- [11] BLEI D M, MCAULIFFE J D. Supervised Topic Models[C]. Proceedings of NIPS, 2007:993-1022.
- [12] BOTTOU L. Large-scale machine learning with stochastic

gradient descent[C].Proceedings ofCOMPSTAT'10.2010.

[13] BOYD S, VANDENBERGHE L.Convex Optimization[M].Cambridge: Cambridge University Press, 2004.

[14] BOYD S, PARIKH N, CHU E, et al.Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers[J].Foundations&Trends in Machine Learning, 2010, 3 (1) :883-898 (16) .

[15] BRODER A Z, CARMEL D, HERSCOVICI M, et al.Efficient Query Evaluation Using a Twolevel-Retrieval Process[C].Proceedings of CIKM'03, 2003: 426-434.

[16] BURGESS C JC.A Tutorial on Support Vector Machines for Pattern Recognition.Data Mining and Knowledge Discovery[J].Data Mining&Knowledge Discovery, 1998, 2 (2) :121-167.

[17] BURROWS M.The Chubby lock service for loosely-coupled distributed systems[C].Proceedings ofOSDI'06, 2006.

[18] CANNY J.GaP:a factor model for discrete data[C].Proceedings of SIGIR.2004:122-129.

[19] CHAKRABORTY T, EVEN-DAR E, GUHA S, et al.Internet and Network Economics[C].Pro-ceedings of WINE'10, 2010:145-157.

[20] CHANG F, DEAN J , GHEMAWAT S, et al.Bigtable: A Distributed Storage System For Struc-tured Data[J].ACM Transactions on Computer Systems, 2008, 26 (2) .

[21] 陈超.Spark大数据平台[M].北京, 2014.<http://www.chinahadoop.cn/course/7>.

[22] CHENYE, PAVLOV D, CANNY J F.Large-scale behavioral targeting[C].Proceedings of KDD, 2009:426-434.

[23] CHEN PEIJI, MA WENJING, MANDALAPU S, et al.Ad Serving Using a Compact Allocation-Plan[M].[S.I]:[s.n], 2012.

[24] CLARKE E H.Multipart pricing of public goods[J].Public Choice, 1971, 11 (1) :17-33.

[25] CUI YING, ZHANG RUOFEI, LI WEI, et al.Bid landscape forecasting in online ad exchange marketplace[J].Proceedings of KDD'11, 2011:265-273.

[26] DEAN J, GHEMAWAT S.MapReduce: Simplified Data Processing on Large Clusters[CJ].Pro-ceedings of OSDI'04, 2004.

[27] DEERWESTER S, DUMAIS S T, FURNAS G W, et al. Richard Harshman Indexing by Latent Semantic Analysis[J]. Journal of the American Society for Information Science, 1990, 41 (6) :391-407.

[28] DEMPSTER A.P, LAIRD N.M, RUBIN D.B. Maximum likelihood from incomplete data via the EM algorithm[J]. Journal of the Royal Statistical Society, 1977, 39 (1) :1-38.

[29] DWORK C. Differential Privacy: A Survey of Results[J]. Theory and Applications of Models of Computation, 2008, 4978:1–19.

[30] EDELMAN B, OSTROVSKY M, SCHWARZ M, et al. Internet Advertising and the Generalized Second Price Auction: Selling Billions of Dollars Worth of Keywords. Working paper 2005[J]. American Economic Review, 2005, 97 (1) :242-259.

[31] FELDMAN J, KORULA N, MIRROKNI V, et al. Online Ad Assignment with Free Disposal[C]. Proceedings of WINE, 2009:374-385.

[32] FREUND Y, SCHAPIRE R E. A Short Introduction to Boosting (“Recent Developments in the Theory and Applications of Machine Learning”) [J]. Journal of Japanese Society for Artificial Intelligence, 1999, 14 (5) :771-780.

[33] GATES A F. Pig: A Structured, High-Level Dataflow System for Hadoop[M]. [S.I]:[s.n.].

[34] GHEMAWAT S, GOBIOFF H, LEUNG S T. The Google File System[C]. Proceedings of SOSP, 2003.

[35] GHOSH A, MCAFEE P, PAPINENI K, VASSILVITSKII S. Representative Allocations for Guaranteed Display Advertising[J]. Manual script, 2009.

[36] GILL J. Generalized Linear Model: a Unified Approach[M]. London: SAGE publications, 2000.

[37] GITTINS J. Bandit processes and dynamic allocation indices[J]. Journal of the Royal Statistical Society. 1979:148-177.

[38] GROVES T. Incentives in teams[C]. //Econometrica. 1973:617-31.

[39] HOFMANN T. Probabilistic Latent Semantic Indexing[C]. Proceedings of Proceedings of SIGIR, 2009:0-57.

- [40] IAB.Digital Video Ad Serving Template (VAST) 3.0[EB/OL].北京.2014.<http://www.iab.net/media/file/VASTv3.0.pdf>.
- [41] IAB.OpenRTB API Specification Version 2.2.[EB/OL].北京.2014.<http://www.iab.net/media/file/OpenRTBAPISpecificationVersion2.2.pdf>.
- [42] JANSEN B J, MULLEN T.Sponsored search: an overview of the concept, history, and technology[J].International Journal of Electronic Business, 2008.
- [43] JORDAN M I., GHAHRAMANI Z, JAAKKOLA T S., et al.An Introduction to Variational Methods for Graphical Models[J].Machine Learning, 1999, 37 (2) :183-233.
- [44] JUNQUEIRA F, REED B.Zookeeper: Distributed Process Coordination[M].Oreilly Media, 2013.
- [45] KARAUH.Spark快速数据处理[M].余璜, 张磊, 译.北京: 机械工业出版社, 2014.
- [46] KOREN Y.Factorization meets the neighborhood: a multifaceted collaborative filtering model[C].Proceedings of KDD'08, 2008:426-434.
- [47] KING G, ZENG L.Logistic Regression in Rare Events Data[J].Political Analysis, 2002, 9 (2) :137-163.
- [48] KRUGMAN H E.The Impact of Television Advertising: Learning Without Involvement [J].Public Opinion Quarterly, 1965, 29 (3) :349-356.
- [49] LAMPORT L.Paxos Made Simple[N].SIGACT News.2001:32 (4) 51-58.
- [50] LIHONG LI, WEI CHU, LANGFORD J, et al.A Contextual-Bandit Approach to Personalized News Article Recommendation[C].Proceedings of WWW'10.2010:661-670.
- [51] 李子骅.Redis入门指南[M].北京: 人民邮电出版社, 2013.
- [52] LIN C J, WENG R C, KEERTHI S S.Trust Region Newton Method for Large-Scale Logistic Regression[J].Journal of Machine Learning Research, 2008, 9 (2) :627-650.
- [53] 刘奕群, 马少平, 洪涛, 等.搜索引擎技术基础[M].北京: 清华大学出版社, 2010.
- [54] MALOUF R.A Comparison Of Algorithms For Maximum Entropy Parameter Estimation[J].Proceedings of 6th Conference on

Natural Language Learning, 2002:1-7.

[55] MARKOWITZ H M. Portfolio Selection: Efficient Diversification of Investment[R]. London: John Wiley & Sons, Inc, 1959.

[56] MAYERS. V, CUKIER K. Big Data: A Revolution That Transforms How we Work, Live, and Think[J]. Houghton Mifflin Harcourt, 2012.

[57] MCCANDLESS M, HATCHER E, GOSPODNETICO. Lucene 实战[M]. 牛长流, 肖宇, 译. 2版. 北京: 人民邮电出版社, 2011.

[58] MELVILLE P, SINDHWANI V. Recommender Systems. Encyclopedia of Machine Learning[M]. Berlin: Springer, 2010.

[59] MEHTA A, SABERI A, VAZIRANI U, et al. Adwords And Generalized Online Matching[J]. Journal of ACM, 2007, 54 (5) :264-273.

[60] NARAYANAN A, SHMATIKOV V. Robust De-anonymization of Large Sparse Datasets[C]. Proceedings of IEEE Symposium on Security and Privacy. IEEE, 2008:111-125.

[61] NEWMAN D, ASUNCION A, SMYTH P, et al. Distributed Inference for Latent Dirichlet Allocation[C]. Proceedings of NIPS'08. 2008.

[62] NOCEDAL J, WRIGHT S J. Numerical Optimization[J]. Springer. 1999.

[63] OLSTON C, REED B, SRIVASTAVA U, et al. PigLatin: a not-so-foreign language for data processing[C]. Proceedings of SIGMOD'08, 2008:1099-1110.

[64] OSTROVSKY M, SCHWARZ M. Reserve Prices in Internet Advertising Auctions: A Field Experiment[J]. Stanford University Graduate School of Business Research., 2009, (2054) :59-60.

[65] PEROTTE A, BARTLETT N, ELHADAD N, et al. Hierarchically Supervised Latent Dirichlet Allocation[C]. Proceedings of NIPS'11, 2011.

[66] PRESS W H. Numerical Recipe [M]. 3rd ed. Cambridge: Cambridge University Press, 2007.

[67] SLEE M, AGARWAL A, KWIATKOWSKI M. Thrift: Scalable cross-language services implementation[J]. Facebook White

Paper, 2007.

[68] SU XIAOYUAN, KHOSHGOFTAAR T M.A Survey of Collaborative Filtering Techniques[J].Advances in Artificial Intelligence, 2009.

[69] TANG D, AGARWAL A, BRIEN D O, et al.Meyer M.Overlapping experiment infrastructure:more, better, faster experimentation[C].Proceedings of KDD'10, 2010.

[70] 陶辉.深入理解Nginx: 模块开发与架构解析[M].北京: 机械工业出版社, 2013.

[71] TURNEY P D, PANTEL P.From Frequency to Meaning: Vector Space Models of Semantics[J].Journal of Artificial Intelligence Research, 2010:141-188.

[72] VARIAN H R.Position auctions[J].International Journal of Industrial Organization, 2007, 25 (6) :1163-1178.

[73] VEE E, VASSILVITSKII S, SHANMUGASUNDARAM J.Optimal online assignment with fore-casts[C].Proceedings of the ACM Conference on Electronic Commerce, 2010.

[74] VICKREY W.Counterspeculation, Auctions, and Competitive Sealed Tenders.[J].Journal of Finance.1961:8-37.

[75] WANG XUERUI, BRODER A, FONTOURA M, et al.A Search-Based Method For Forecasting Ad Impression In Contextual Advertising[C].Proceedings of WWW'09.2009: 491-500.

[76] WHANG S E, MOLINA H G.Indexing Boolean Expressions. [C].Proceedings of VLDB, 2009, 2 (1) .

[77] WHITET.Hadoop权威指南[M].周敏奇, 钱卫宁, 金澈清, 等, 译.2版.北京: 清华大学出版社, 2011.

[78] ZAHARIA M, CHOWDHURY M, MA J, et al.Spark: cluster computing with working sets[M].[S.l.]:[s.n.], 2010.

[79] FORUM M P.MPI:A message-passing interface standard [J].Proc ACM International Conference on Supercompting, 1994, 20 (2) : 179.

[80] 赵必厦, 程丽明.从零开始学Storm.[M].北京: 清华大学出版社, 2014.

[81] YI ZHU, WILBUR K C.Hybrid Advertising Auctions[J].Social Science Electronic Publishing, 2011, 30 (2) :249-273.

[82] ZINKEVICH M A, SMOLA A, WEIMER M, et al.Parallelized Stochastic Gradient Descent[C].Proceedings of NIPS'10, 2010.

[83] ZOBEL J, MOFFAT A.Inverted Files For Text Search Engines[J].ACM Computing Surveys Surveys, 2006, 38 (4) .

Table of Contents

[目录](#)

[扉页](#)

[版权](#)

[对本书的点评](#)

[序一](#)

[序二](#)

[序三](#)

[前言](#)

[第一部分 计算广告关键技术](#)

[第1章 在线广告综述](#)

[1.1 大数据与广告的关系](#)

[1.2 广告的定义与目的](#)

[1.3 在线广告创意类型](#)

[1.4 在线广告简史](#)

[1.5 泛广告商业产品](#)

[1.6 延伸思考](#)

[第2章 计算广告基础](#)

[2.1 广告有效性原理](#)

[2.2 互联网广告的技术特点](#)

[2.3 计算广告的核心问题](#)

[2.3.1 广告收入的分解](#)

[2.3.2 结算方式与 eCPM 估计的关系](#)

[2.4 在线广告相关行业协会](#)

[2.4.1 交互广告局](#)

[2.4.2 美国广告代理协会](#)

[2.4.3 美国国家广告商协会](#)

[2.5 延伸思考](#)

[第二部分 在线广告产品逻辑](#)

[第3章 在线广告产品概览](#)

[3.1 商业产品的设计原则](#)

[3.2 需求方层级组织与接口](#)

[3.3 供给方管理接口](#)

[3.4 延伸思考](#)

[第4章 合约广告](#)

[4.1 广告位合约](#)

[4.2 受众定向](#)

[4.2.1 受众定向方法概览](#)

[4.2.2 受众定向标签体系](#)

[4.3 展示量合约](#)

[4.3.1 流量预测](#)

[4.3.2 流量塑形](#)

[4.3.3 在线分配](#)

[4.3.4 产品案例](#)

[4.4 延伸思考](#)

[第5章 搜索与竞价广告](#)

[5.1 搜索广告](#)

[5.1.1 搜索广告产品形态](#)

[5.1.2 搜索广告产品新形式](#)

[5.1.3 搜索广告产品策略](#)

[5.1.4 产品案例](#)

[5.2 位置拍卖与机制设计](#)

[5.2.1 定价问题](#)

[5.2.2 市场保留价](#)

[5.2.3 价格挤压](#)

[5.2.4 定价结果示例](#)

[5.3 广告网络](#)

[5.3.1 广告网络产品形态](#)

[5.3.2 广告网络产品策略](#)

[5.3.3 产品案例](#)

[5.4 竞价广告需求方产品](#)

[5.4.1 搜索引擎营销](#)

[5.4.2 媒体购买平台](#)

[5.4.3 产品案例](#)

[5.5 竞价广告与合约广告的比较](#)

[5.6 延伸思考](#)

[第6章 程序化交易广告](#)

[6.1 实时竞价](#)

[6.2 其他程序化交易方式](#)

[6.2.1 优选](#)

[6.2.2 私有市场](#)

[6.2.3 广告交易方式谱系](#)

[6.3 广告交易平台](#)

[6.4 需求方平台](#)

[6.4.1 需求方平台产品策略](#)

[6.4.2 出价策略](#)

[6.4.3 重定向](#)

[6.4.4 新客推荐](#)

[6.4.5 产品案例](#)

[6.5 供给方平台](#)

[6.5.1 供给方平台产品策略](#)

[6.5.2 产品案例](#)

[6.6 数据加工与交易](#)

[6.6.1 有价值的数据来源](#)

[6.6.2 三方数据划分](#)

[6.6.3 数据管理平台](#)

[6.6.4 数据交易平台](#)

[6.6.5 产品案例](#)

[6.7 在线广告产品交互关系](#)

[6.8 延伸思考](#)

[第7章 移动互联与原生广告](#)

[7.1 原生广告相关产品](#)

[7.1.1 信息流广告](#)

[7.1.2 搜索广告](#)

[7.1.3 软文广告](#)

[7.1.4 联盟](#)

[7.2 移动广告的现状与挑战](#)

[7.2.1 移动广告的特点](#)

[7.2.2 移动广告的创意形式](#)

[7.2.3 移动广告的挑战](#)

[7.3 原生广告平台](#)

[7.3.1 表现原生与意图原生](#)

[7.3.2 植入式原生广告](#)

[7.3.3 产品案例](#)

[7.4 原生广告与程序化交易](#)

[7.5 延伸思考](#)

[第8章 在线广告产品实践](#)

[8.1 媒体实战](#)

[8.1.1 变现方式和产品决策](#)

[8.1.2 数据支持方案决策](#)

[8.2 广告主实战](#)

[8.3 数据提供方实战](#)

[8.4 延伸思考](#)

[第三部分 计算广告关键技术](#)

[第9章 计算广告技术概览](#)

[9.1 个性化系统框架](#)

[9.2 各类广告系统优化目标](#)

[9.3 计算广告系统架构](#)

[9.3.1 广告投放引擎](#)

[9.3.2 数据高速公路](#)

[9.3.3 离线数据处理](#)

[9.3.4 在线数据处理](#)

[9.4 计算广告系统主要技术](#)

[9.5 用开源工具搭建计算广告系统](#)

[9.5.1 Web 服务器 Nginx](#)

[9.5.2 分布式配置和集群管理工具 ZooKeeper](#)

[9.5.3 全文检索引擎 Lucene](#)

[9.5.4 跨语言通信接口 Thrift](#)

[9.5.5 数据高速公路 Flume](#)

[9.5.6 分布式数据处理平台 Hadoop](#)

[9.5.7 特征在线缓存 Redis](#)

[9.5.8 流计算平台 Storm](#)

[9.5.9 高效的迭代计算框架 Spark](#)

[9.6 延伸思考](#)

[第10章 基础知识准备](#)

[10.1 信息检索](#)

[10.1.1 倒排索引](#)

[10.1.2 向量空间模型](#)

[10.2 最优化方法](#)

[10.2.1 拉格朗日法与凸优化](#)

[10.2.2 下降单纯形法](#)

	10.2.3 梯度下降法
	10.2.4 拟牛顿法
	10.2.5 Trust-Region 法
10.3 统计机器学习	
	10.3.1 最大熵与指数族分布
	10.3.2 混合模型和 EM 算法
	10.3.3 贝叶斯学习
10.4 统计模型分布式优化框架	
第11章 合约广告核心技术	
	11.1 广告排期系统
	11.2 担保式投送系统
	11.2.1 流量预测
	11.2.2 频次控制
	11.3 在线分配
	11.3.1 在线分配问题
	11.3.2 在线分配问题举例
	11.3.3 极限性能研究
	11.3.4 实用优化算法
	11.4 延伸思考
第12章 受众定向核心技术	
	12.1 受众定向技术分类
	12.2 上下文定向
	12.2.1 半在线抓取系统
	12.2.2 文本主题挖掘
	12.3 行为定向
	12.3.1 行为定向建模问题
	12.3.2 行为定向特征生成
	12.3.3 行为定向决策过程
	12.3.4 行为定向的评测
	12.4 人口属性预测
	12.5 数据管理平台
	12.6 延伸思考
第13章 竞价广告核心技术	
	13.1 竞价广告计价算法
	13.2 搜索广告系统
	13.2.1 查询扩展

- [13.2.2 广告放置](#)
 - [13.3 广告网络](#)
 - [13.4 广告检索](#)
 - [13.4.1 布尔表达式的检索](#)
 - [13.4.2 相关性检索](#)
 - [13.5 点击率预测](#)
 - [13.5.1 点击率预测模型](#)
 - [13.5.2 优化算法](#)
 - [13.5.3 点击率模型的校正](#)
 - [13.5.4 点击率模型的特征](#)
 - [13.5.5 点击率模型评测](#)
 - [13.5.6 智能频次控制](#)
 - [13.6 探索与利用](#)
 - [13.6.1 UCB 方法](#)
 - [13.6.2 考虑上下文的 bandit](#)
 - [13.7 延伸思考](#)
- [第14章 程序化交易核心技术](#)
 - [14.1 广告交易平台](#)
 - [14.1.1 cookie 映射](#)
 - [14.1.2 询价优化](#)
 - [14.2 需求方平台](#)
 - [14.2.1 定制化用户标签](#)
 - [14.2.2 DSP 中的点击率预测](#)
 - [14.2.3 点击价值估计](#)
 - [14.2.4 出价策略](#)
 - [14.3 供给方平台](#)
 - [14.4 延伸思考](#)
- [第15章 其他广告相关技术](#)
 - [15.1 创意优化](#)
 - [15.1.1 程序化创意](#)
 - [15.1.2 点击热力图](#)
 - [15.2 实验框架](#)
 - [15.3 流量保护和效果监测](#)
 - [15.3.1 反作弊](#)
 - [15.3.2 广告监测](#)
 - [15.3.3 广告安全](#)

15.4 隐私保护和数据安全

15.4.1 隐私保护问题

15.4.2 程序化交易中的数据安全

15.5 延伸思考

第四部分 附录

附录A 主要术语及缩写索引

参考文献