

Online Advertising

# 在线广告

互联网广告系统的架构及算法

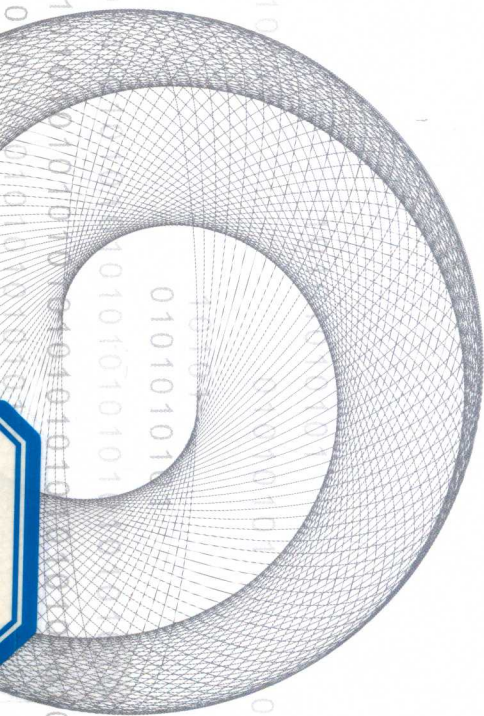
张亚东 © 编著

▪ 架构

品牌广告、搜索类广告、社交类广告、视频类广告等

▪ 算法

用户数据和定向算法、点击率预估与排序算法、在线匹配、机制设计、广告中低质量和敏感控制、实验架构和调参，以及数据监控和效果衡量等



清华大学出版社

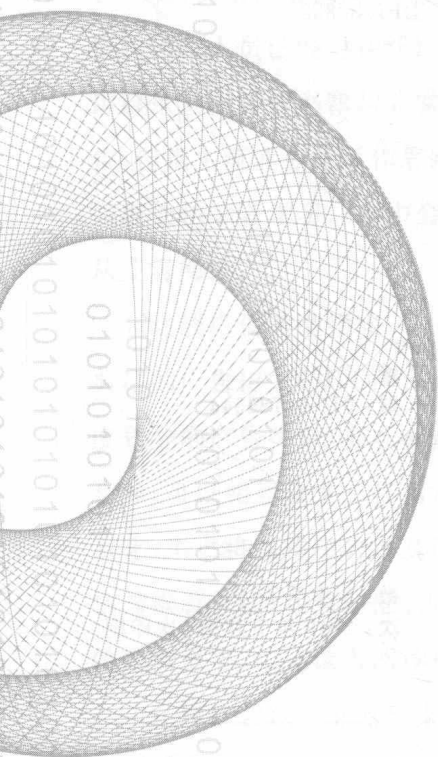


Online Advertising

# 在线广告

互联网广告系统的架构及算法

张亚东 © 编著



清华大学出版社

北京

## 内 容 简 介

在线广告是一个多学科交融的领域,本书力求系统地讲解在线广告的架构和算法,让读者对在线广告有一个整体的认识。全书共 15 章,第 1~3 章介绍在线广告的发展简史、样式与创意以及广告系统的架构流程等基础知识;第 4~7 章从品牌广告、搜索类广告、社交类广告和视频类广告 4 种典型的广告出发,详细讲解广告系统的更多设计细节;第 8~14 章分类讲述在线广告中的重要研究领域,包括用户数据和定向算法、点击率预估与推荐算法、在线匹配、机制设计、低质量和敏感控制、实验架构和调参,以及数据监控和效果衡量;第 15 章简要介绍在线广告的发展趋势。

本书可作为对在线广告感兴趣的初学者的入门书籍,也可供在线广告相关领域的从业者阅读参考。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

### 图书在版编目(CIP)数据

在线广告:互联网广告系统的架构及算法/张亚东编著. —北京:清华大学出版社,2019(2019.9 重印)  
ISBN 978-7-302-52652-0

I. ①在… II. ①张… III. ①网络广告—算法设计—研究 IV. ①F713.852

中国版本图书馆 CIP 数据核字(2019)第 047108 号

责任编辑:黄 芝 张爱华

封面设计:刘 键

责任校对:梁 毅

责任印制:杨 艳

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座

邮 编:100084

社 总 机:010-62770175

邮 购:010-62786544

投稿与读者服务:010-62776969, [c-service@tup.tsinghua.edu.cn](mailto:c-service@tup.tsinghua.edu.cn)

质量反馈:010-62772015, [zhiliang@tup.tsinghua.edu.cn](mailto:zhiliang@tup.tsinghua.edu.cn)

课件下载:<http://www.tup.com.cn>, 010-62795954

印 装 者:北京鑫海金澳胶印有限公司

经 销:全国新华书店

开 本:185mm×230mm 印 张:20.25

字 数:354 千字

版 次:2019 年 8 月第 1 版

印 次:2019 年 9 月第 2 次印刷

印 数:2001~3500

定 价:69.80 元

产品编号:082147-01

# 前言

FOREWORD

在线广告是一个多学科交融的领域,也是互联网公司最主要的营利模式之一,涉及内容广泛,实用价值极高。伴随着互联网公司线上业务的不断更迭,相关领域从业者和学术机构研究人员已经对在线广告的各种业务痛点进行了大量深入的分析研究,但市面上系统讲解这些内容的书籍却非常少。本书尽可能以通俗易懂的语言系统地讲解互联网在线广告,包括主流业务的运作思路、常见的系统架构以及实践中的核心算法,适合对在线广告感兴趣的初学者学习。书中会对真实业务中实施的部分算法进行深入讲解,适合相关领域的从业者阅读。

在线广告类型很多,本书从品牌广告、搜索类广告、社交类广告和视频类广告4种较为典型的广告出发,从不同角度解读在线广告的业务思路和相关系统,它们并不是从一个维度划分的广告类型,会互有重复,但非常具有代表性,其他类型广告的业务和架构流程大同小异,可以相互参考。目前对在线广告核心内容清晰归类的书籍很少,为了使内容更加体系化,本书将在线广告涉及的主要内容及核心算法划分为用户数据和定向算法、点击率预估与推荐算法、在线匹配、机制设计、低质量和敏感控制、实验架构和调参以及数据监控和效果衡量等,希望读者阅读后能对在线广告有相对系统的认识。

在线广告内容非常庞杂且在快速发展中,实践中仍会不断涌现出很多新问题、新方法,需要持续地探索分析。由于作者能力有限,不能对所有问题给出全面、准确的描述,加之受

时间和精力限制,书中难免有疏漏之处,若读者不吝赐教,将感激不尽。希望经过不懈的努力,使本书成为一本对初学者系统了解在线广告,对相关从业者实战有所帮助的书籍。读者可以关注公众号 Chinabigdata 或者关注微博“亚东爱阅读”,后期增补内容会在公众号中更新。

编者

2019年6月

# 目录

CONTENTS

第 1 章 在线广告发展简史 .....	1
1.1 在线广告发展简介 .....	2
1.1.1 中国古代的广告 .....	2
1.1.2 在线广告的诞生 .....	3
1.1.3 搜索广告的诞生和发展 .....	4
1.1.4 社交和视频类广告 .....	6
1.1.5 Ad Network 的诞生 .....	6
1.1.6 Ad Exchange 和 TradingDesk .....	7
1.2 移动广告的发展 .....	8
1.3 搜索广告和定价模式 .....	11
1.4 社交媒体广告 .....	15
1.5 视频广告 .....	17
1.6 在线广告优势 .....	19
1.7 在线广告规模 .....	21
参考文献 .....	21

第2章 广告样式与创意 .....	25
2.1 主流广告样式 .....	26
2.1.1 PC端广告样式 .....	26
2.1.2 移动端广告样式 .....	27
2.2 技术驱动营销 .....	31
2.3 广告样式发展趋势 .....	32
2.4 程序化创意 .....	34
2.4.1 程序化创意的缘由 .....	34
2.4.2 程序化创意 .....	35
2.5 动态创意优化 .....	37
参考文献 .....	39
第3章 广告系统架构流程 .....	41
3.1 投放引擎架构流程 .....	42
3.1.1 广告投放引擎架构 .....	42
3.1.2 客户系统 .....	43
3.1.3 内部管理平台 .....	44
3.1.4 基础架构及相关模块 .....	45
3.1.5 广告投放引擎内部模块 .....	46
3.1.6 工具和测试平台 .....	48
3.2 收入分解 .....	48
3.3 程序化广告技术生态 .....	50
3.4 Ad Network .....	52
3.4.1 工作流程 .....	52
3.4.2 分类 .....	53
3.4.3 定向方式 .....	54
3.4.4 优势 .....	54

3.4.5 移动广告网络 .....	55
3.5 Ad Exchange .....	56
3.5.1 产生背景 .....	56
3.5.2 工作流程 .....	57
3.5.3 与 Ad Network 的不同 .....	58
3.5.4 国内 Ad Exchange 的发展 .....	58
3.6 程序化售卖方式 .....	58
3.7 其他机制 .....	61
3.7.1 匿名设置 .....	61
3.7.2 Reserve Price .....	61
3.7.3 Pre-Targeting .....	63
参考文献 .....	64
<b>第 4 章 品牌广告 .....</b>	<b>65</b>
4.1 品牌推广的意义 .....	66
4.2 品牌广告简介 .....	67
4.2.1 品牌广告 .....	67
4.2.2 品牌广告常见形式 .....	68
4.2.3 计费 and 购买方式 .....	70
4.2.4 样式和创意 .....	70
4.3 品牌广告的有效性 .....	71
4.4 品牌广告效果评估指标 .....	73
4.5 Benchmark .....	75
参考文献 .....	77
<b>第 5 章 搜索类广告 .....</b>	<b>81</b>
5.1 搜索广告简介 .....	82
5.1.1 搜索广告的模式 .....	83



5.1.2	广告投放及相关问题	83
5.1.3	搜索广告的优势	84
5.2	常见产品形态	85
5.2.1	综合搜索	85
5.2.2	定制类搜索	86
5.2.3	图片类搜索	87
5.2.4	内容定向	87
5.2.5	电商类搜索	88
5.2.6	应用商店搜索	89
5.2.7	其他	89
5.3	系统架构和重要模块	89
5.3.1	广告架构	90
5.3.2	广告账户组织结构	90
5.3.3	广告检索流程	91
5.3.4	预算控制	94
5.3.5	在线匹配	95
5.3.6	机制设计	96
5.3.7	计费流程	96
5.3.8	准入	97
5.4	主流竞价机制	98
5.4.1	GFP 机制	98
5.4.2	GSP 机制	99
5.4.3	VCG 机制	100
5.5	搜索生态	100
5.6	GSP 优化	101
5.6.1	Weighted GSP	102
5.6.2	Squashing	102
5.6.3	UWR	103

5.6.4	QWR .....	103
5.6.5	Anchoring .....	104
5.6.6	模型对比 .....	104
5.6.7	Hidden Cost .....	105
5.7	长尾查询 .....	105
5.8	市场规模 .....	107
	参考文献 .....	107
<b>第6章</b>	<b>社交类广告 .....</b>	<b>109</b>
6.1	社交媒体 .....	110
6.1.1	社交网络国度 .....	110
6.1.2	社交网络的特点 .....	111
6.1.3	常见的社交应用 .....	111
6.1.4	社交网络影响购买行为 .....	112
6.2	社交广告 .....	112
6.2.1	常见广告类型 .....	112
6.2.2	定向方式 .....	113
6.3	基于社交关系的算法 .....	114
6.3.1	社交内容推荐算法 .....	114
6.3.2	社区分割算法 .....	115
6.3.3	社交内容扩散算法 .....	117
6.4	社交网络营销 .....	119
	参考文献 .....	120
<b>第7章</b>	<b>视频类广告 .....</b>	<b>123</b>
7.1	视频广告简介 .....	124
7.1.1	常见的广告类型 .....	124
7.1.2	售卖方式 .....	125

7.1.3	广告时长	126
7.2	视频广告生态和投放流程	127
7.2.1	视频广告生态	127
7.2.2	广告投放流程	128
7.3	流量预估	131
7.3.1	优化目标	131
7.3.2	模型特征	131
7.3.3	特征平滑处理	132
7.3.4	流量预估函数	133
7.3.5	模型评估方法	134
7.4	库存分配问题	134
7.5	库存分配算法	138
7.5.1	HWM	139
7.5.2	优化调整	140
7.5.3	反馈机制	141
7.5.4	SHALE	142
7.6	Pacing	144
7.7	市场规模	146
	参考文献	146
<b>第 8 章</b>	<b>用户数据和定向算法</b>	<b>149</b>
8.1	用户识别	150
8.1.1	Cookie	150
8.1.2	Cookie Matching	150
8.1.3	移动端用户识别	152
8.1.4	跨屏识别	154
8.2	用户画像	154
8.3	定向方式	157

8.4 经营状况评估和优化 .....	158
8.4.1 评估指标 .....	158
8.4.2 CLV 优化 .....	159
8.4.3 客户关系管理和使用 .....	161
8.5 Lookalike .....	162
8.5.1 特征提取和建模 .....	162
8.5.2 扩展方式 .....	163
8.5.3 最近邻选择 .....	164
8.5.4 离线扩展流程 .....	165
8.5.5 node2vec .....	165
8.5.6 实战 .....	167
8.6 竞价环境预估 .....	168
8.7 超级用户 .....	173
参考文献 .....	173
<b>第 9 章 点击率预估与推荐算法 .....</b>	<b>175</b>
9.1 点击率预估简介 .....	176
9.2 点击率预估特征 .....	177
9.2.1 相同竞价词下其他订单的特征 .....	177
9.2.2 相关竞价词的 CTR .....	178
9.2.3 广告质量相关特征 .....	178
9.2.4 订单竞价词相关特征 .....	179
9.2.5 外部相关特征 .....	179
9.2.6 特征预处理 .....	179
9.3 预估模型 .....	180
9.3.1 基础模型 .....	180
9.3.2 $L_2$ -TreeBoost+LR 模型 .....	181
9.3.3 回归树 .....	182

9.3.4	Gradient Boosting	183
9.3.5	$L_2$ -TreeBoost	184
9.3.6	特征组合	185
9.3.7	Freshness	186
9.3.8	数据采样	186
9.4	模型评估方法	186
9.4.1	KL 离散算法	186
9.4.2	AUC	187
9.4.3	NE	188
9.5	Bandit	189
9.5.1	Bandit 问题	189
9.5.2	$\epsilon$ -Greedy 方法	190
9.5.3	Thompson Sampling	190
9.5.4	UCB	192
9.5.5	LinUCB	193
9.6	在线学习方法	194
9.6.1	梯度下降方法	195
9.6.2	BGD	195
9.6.3	SGD	196
9.6.4	MBGD	196
9.6.5	简单截断法	198
9.6.6	截断梯度法	198
9.6.7	FOBOS	198
9.6.8	RDA	200
9.6.9	$L_1$ -FOBOS 和 $L_1$ -RDA 的对比	201
9.6.10	FTRL	202
9.7	推荐算法	203
9.8	基于协同过滤的推荐	204

9.8.1	基于用户的协同过滤算法	204
9.8.2	基于物品的协同过滤算法	206
9.8.3	其他相似度计算方法	206
9.8.4	应用	207
9.9	基于矩阵分解的推荐	207
9.9.1	矩阵分解	207
9.9.2	正则化	209
9.9.3	隐性特征	209
9.10	基于深度学习的推荐	210
9.10.1	推荐流程	210
9.10.2	排序	212
9.11	广告排序性能优化	212
	参考文献	213
<b>第 10 章</b>	<b>在线匹配</b>	<b>215</b>
10.1	图论基础知识	216
10.2	在线匹配类型	221
10.3	在线二部图匹配	223
10.3.1	Greedy 算法	224
10.3.2	Random 算法	225
10.3.3	Ranking 算法	225
10.4	加权的在线二部图匹配	227
10.5	Adwords	229
10.5.1	Greedy 算法	229
10.5.2	Balance 算法与 Greedy 算法对比	230
10.5.3	MSVV 算法	231
10.5.4	一般情况的证明	233
10.6	基于原始对偶的匹配	234

10.6.1	原始对偶问题 .....	234
10.6.2	互补松弛性 .....	236
10.6.3	Greedy 算法实现 .....	237
10.6.4	更优算法 .....	239
10.7	现实系统中的匹配算法 .....	240
	参考文献 .....	241
<b>第 11 章</b>	<b>机制设计 .....</b>	<b>243</b>
11.1	机制设计概述 .....	244
11.2	经典案例 .....	245
11.2.1	囚徒困境 .....	245
11.2.2	二难问题 .....	246
11.2.3	无怨算法 .....	246
11.2.4	TureView 广告 .....	246
11.2.5	策略性投票 .....	247
11.3	激励兼容 .....	247
11.3.1	投票悖论 .....	247
11.3.2	阿罗不可能定理 .....	248
11.4	引入金钱的机制 .....	249
11.4.1	拍卖机制 .....	250
11.4.2	VCG 机制 .....	251
11.5	激励兼容的特性 .....	251
11.6	贝叶斯纳什均衡 .....	253
11.7	竞价机制分析 .....	255
11.7.1	临界条件分析 .....	255
11.7.2	VCG 机制 .....	255
11.7.3	Simplest GSP 机制 .....	256
11.7.4	Weighted GSP 机制 .....	257

11.8 拥挤控制 .....	257
参考文献 .....	261
<b>第 12 章 低质量和敏感控制 .....</b>	<b>263</b>
12.1 作弊背景 .....	264
12.1.1 作弊参与者 .....	264
12.1.2 作弊动机 .....	265
12.2 广告作弊方法 .....	265
12.2.1 单机作弊 .....	266
12.2.2 黑客作弊 .....	266
12.2.3 有组织的网络作弊 .....	266
12.2.4 有组织的人工作弊 .....	267
12.2.5 基于大流量平台的作弊 .....	267
12.3 广告反作弊 .....	267
12.3.1 反作弊架构 .....	267
12.3.2 反作弊算法分类 .....	268
12.4 广告质量 .....	269
12.5 数据安全 .....	270
参考文献 .....	272
<b>第 13 章 实验架构和调参 .....</b>	<b>273</b>
13.1 A/B testing .....	274
13.2 分层实验 .....	275
13.2.1 分层实验方案 .....	276
13.2.2 实验平台 .....	277
13.3 实验设计和分析 .....	277
13.3.1 置信度 .....	278
13.3.2 置信区间 .....	278



13.3.3	最少样本数 .....	279
13.3.4	逐步放量 .....	280
13.3.5	50% vs 50% .....	280
13.3.6	其他因素 .....	280
13.3.7	对比实验局限 .....	281
13.3.8	参数化 .....	281
13.4	自动化调参 .....	282
	参考文献 .....	286
<b>第 14 章</b>	<b>数据监测和效果衡量 .....</b>	<b>289</b>
14.1	第三方监测 .....	290
14.2	效果跟踪 .....	291
14.2.1	归因模型 .....	292
14.2.2	增效测试 .....	293
	参考文献 .....	294
<b>第 15 章</b>	<b>在线广告的发展趋势 .....</b>	<b>295</b>
15.1	网络带来的变化 .....	296
15.2	未来发展趋势 .....	297
15.2.1	流量入口 .....	297
15.2.2	需求和市场 .....	298
<b>附录 A</b>	<b>单词表 .....</b>	<b>299</b>