

Web Data Mining and Analysis

Systematic Approach and
Business Practice

网站数据挖掘与分析

系统方法与商业实践

宋天龙◎著

在线数据分析领域巨擘Webtrekk官方资深数据分析专家撰写，目前最具系统性、深度和商业指导价值的Web数据挖掘与分析专著

技术层面，系统讲解网站数据分析的各种指标、工具、技术和方法；应用层面，通过10余个商业案例还原真实的数据分析场景；管理层面，深入探讨如何进行数据风险、数据质量、数据投入和产出、数据流程和落地管理



机械工业出版社
China Machine Press

Web Data Mining and Analysis

Systematic Approach and
Business Practice

网站数据挖掘与分析 系统方法与商业实践

宋天龙◎著



机械工业出版社
China Machine Press

图书在版编目(CIP)数据

网站数据挖掘与分析:系统方法与商业实践 / 宋天龙著. —北京:机械工业出版社, 2015.1
(数据分析技术丛书)

ISBN 978-7-111-49059-3

I. 网… II. 宋… III. 数据采集 IV. TP274

中国版本图书馆 CIP 数据核字 (2014) 第 312536 号



网站数据挖掘与分析 系统方法与商业实践

出版发行: 机械工业出版社 (北京市西城区百万庄大街 22 号 邮政编码: 100037)

责任编辑: 姜 影

责任校对: 殷 虹

印 刷: 中国电影出版社印刷厂

版 次: 2015 年 2 月第 1 版第 1 次印刷

开 本: 186mm×240mm 1/16

印 张: 29.75

书 号: ISBN 978-7-111-49059-3

定 价: 89.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

客服热线: (010) 88378991 88361066

投稿邮箱: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzjsj@hzbook.com

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问: 北京大成律师事务所 韩光 / 邹晓东

Praise 本书赞誉

宋天龙是中国新生代的网站数据分析界“大拿”的代表，其对网站数据分析的理解是行业中数一数二的。他把他对网站数据分析的阐释和领会一览无余地吐露在本书中，并分四篇深入浅出地讲述了从企业数据体系建设之初，到网站分析工具的选择与部署，再到实际案例分析，最后收尾于其对网站数据分析的升华与提高，尤其是引人入胜的第15章对网站数据分析的投入与产出的分析，更是充分展示了宋天龙对网站数据分析游刃有余的拿捏。据此，我推荐任何刚入门及所有对网站数据分析感兴趣的朋友来学习和阅读，如果可能，与宋天龙一起切磋和提高。

张姝 Webtrekk 亚洲区市场总监

这是一本写网站数据分析的专业书，有思路，有分析方法，有分析工具讲解和案例剖析。推荐数据分析、网站运营等人士阅读。

黄成明 (@数据化管理) 数据化管理的咨询顾问及培训师

很早就认识天龙，第一次正式认识是通过我的一篇博文，他的评论来源于有见地的实践。与其他网站数据书籍不同的是，本书既强调工具和技术，又重视意识和应用，尤其融入了很多大型企业的宝贵经验。他把统计学、数据挖掘、自动化应用等方法应用其中，这是对网站数据工作方法的延伸；同时，对于数据工作的完整认知和数据管理意识，更是丰富了网站数据工作的广度与深度。因此，这是一本不可多得的佳作。

宋星 阳狮锐奇 (Publicis Groupe VivaKi) 数据解决方案总经理

数据分析 = 技术 + 市场 + 战略，这就决定了数据分析在企业中的重要性，我们需要借助数据来指引决策，而不是随机拍脑袋！传统行业的决策过多依赖于管理者的眼光和洞察力，

而数据分析人要做的事情，就是把这些眼光和洞察力转化为可量化的数字信息！同时数据的竞争在于数据采集的竞争，在于数据维度的竞争以及玩转数据的人，再好的数据没有优秀的数据人员进行经营及挖掘，其价值终究无法展现。推荐阅读天龙撰写的此书，本书很好地阐述了如何进行企业级数据运营及数据分析工作，对于准备了解及从事数据分析工作的人有非常大的帮助。

田学峰 萝卜教育网创始人

如何定位及发现数据中的价值是网站分析中最受关注的问题。宋天龙通过多年的从业经验给出了答案。本书从数据的认知到数据的应用，以丰富的案例由浅入深地讲解了数据在商业中的作用及价值。无论你是数据行业的新人，还是有一定实践经验的从业者，这都是一本不应该错过的好书。

王彦平 网站分析爱好者《网站分析实战》作者之一

对于每一位网站分析师来讲，这是本必须买和必须认真读的书。

第一次读本书时，就被书中宽广的知识视角和清晰的阐述所折服，这完全是源于作者极其丰富的实战经验，以及对产品和技术孜孜不倦的追求。

宫鑫 射手学院创始人

中国网站分析的圈子其实不大，我和天龙也一直相望于江湖。他从业时间很长，使用过各家主流的分析工具，亦有着从事大型电商数据管理工作的经验，实属不多见的实战派人才。天龙是一个严谨的网站分析工作者，正如他所理解的那样，在这个领域里很多人是一只翅膀的鸟儿，要么执着地偏向分析技术，要么急躁地追求数据价值。网站数据分析是一条漫长的路，它隶属于一个细分研究领域，却涉猎广博、变化极快。希望这本书能够带给你一个新的角度，从而帮助你站到一个新的高度。

王晓东 TrueMetrics 创始人

互联网的发展日新月异，企业要想有所作为，必须打造完善的数据跟踪、分析和优化体系。网站数据分析区别于传统的业务分析，会涉及数据跟踪机制、数据分析建模层面，也区别于数据挖掘，会有数据与业务结合的层面。此书深入浅出地介绍了网站数据分析的原理和实际案例，是行业内不可多得的开山之作。无论是对于互联网数据分析相关从业人员，数据分析专业人员，还是管理者，本书都是不可多得的好书。优秀的人出优秀的书，强烈推荐！

胡力 Netconcepts 深圳分公司总经理《流量的秘密（第3版）》审校

无论是互联网还是传统公司都越来越意识到数据资产和数据分析的重要性，但很多公司对如何利用数据创造价值和基于数据做决策存在一些疑惑。

本书作者基于自身的行业经验和理论，从数据的价值、如何收集数据，到如何形成数据决策和数据驱动的体系一层层地回答了企业线上营销和线上运营等维度的数据分析问题。无论是对职业的数据分析师、数据分析入门者，还是企业管理人员，本书都不失为既有理论深度又有较高实践性的优秀之作。

李俊 百度大数据部高级产品经理

前言 Preface

为什么要写这本书

随着中国商业精细化运营价值的凸显以及企业对数据价值认可度的提高，网站数据分析正变得炙手可热，尤其在互联网企业中，网站数据分析已经成为从业人员必备的一项职业技能。

但在对网站数据进行分析的过程中，我发现企业中普遍存在三类问题，本书就致力于帮助读者解决这三类问题。

第一类是数据工作者的认知问题。

纵观当前与网站数据相关的从业者，或多或少都会存在以下两种认知：第一种是技术论，这种观点的核心是关注数据部署和采集、数据工具、技术、模型的重要性而忽略了应用场景；第二种是业务论，这种观点只关注应用层面的业务问题，由于缺乏对数据前端处理的把握和专业技术、工具的支持，导致后期数据质量和应用都缺乏可靠依据，最终影响了数据价值的提炼及应用效果的提升。

以数据价值为导向的数据分析师应该具备以下素质，这也是贯穿本书的核心思想：

- 一是立足于数据本身的追本溯源。数据分析师需要了解数据的整个工作链，从数据的产生、采集、存储、提取、挖掘、分析、展现到集成应用，并能在各个环节有独到的见解。
- 二是着眼于数据应用价值的研究。研究数据如何能更智能化、可视化、自动化，以及如何更有价值地解决业务问题并带来业务价值的直接提升。

不得不说，技术是实现商业理解的必要保证。网站数据分析的传统方法是趋势、细分和

转化,但仅有这些方法还不够,很多深层次的问题需要借助其他方法来实现,例如数据挖掘、统计学、人工智能、商业智能等。我从来不认为网站数据分析与数据工作是割裂的,它是数据工作的一部分,所有关于数据的工作方法都可以和网站数据结合使用。但可惜的是,当前将网站分析与其他数据工作方法结合起来的较少,因此,我在本书中用大量的篇幅介绍数据挖掘在网站分析中的应用案例。

第二类是数据价值的认知问题。

对于任何一个企业来说,数据工作都不是企业发展的必需条件,最起码在企业运作初期没有大量数据的情况下企业同样可以快速发展。这时我开始思考数据的价值到底是什么?数据到底能给企业带来什么?如果没有数据企业又会损失什么?归根结底,数据存在的意义是用来解决商业问题的,换句话说数据能给企业带来多少价值,以及这些价值是如何体现在企业的利润报表里面的。作为网站数据分析应该如何带动企业的业务成长,或者如何以单独的形态与业务结成依存关系最终实现自我价值。这些问题是需要讨论的。未来,数据的作用将主要着眼于基于数据驱动和系统智能工作机制,而辅助决策工作将成为数据的一个非主要应用。所以本书在案例篇中重点介绍了基于数据驱动的营销和运营应用,其目的便在于此。

第三类是如何从企业的角度做数据工作管理的问题。

作为初、中级分析师,主要工作职责是把数据本身或数据项目工作做好;但作为管理层的高级分析师或管理者,需要思考的问题不仅是如何完成工作,还包括如何建立企业数据架构、数据工作流程、数据应用体系、数据风险以及质量管理体系,这是站在企业的高度来思考数据的定位及布局的必经之路。

基于以上三类问题,我萌生了写本书的想法,目的是希望读者能够放开眼界,首先破除网站数据的局限性,其次破除数据的局限性,最终站在企业的角度思考问题。作为一本接地气的书,书中列举了大量案例并通过对每个案例的详细介绍来帮助读者进行案例式的学习,希望能带给读者一些新的理解、观念和应用思路,使其无论是在工作机会的选择上还是收入上都能获得较大的帮助。

读者对象

本书适合以下几类从业者阅读。

- 对数据研究感兴趣的在职人员。无论你从事什么工作,如果你能够将数据的思路、价值和应用方法结合到你的工作实践中,一定会对你现有的工作有所帮助。数据化思维和工作能力已经成为每个在职人员的加分项。

□ 刚入数据行业的新人。如果你是一位刚入行的新人，一定希望能够有一本兼具实战和理论高度的书籍，从全局到局部的每个细节为你理清工作思路并明确职业成长方向。如果你要了解数据在企业内的价值、工作流程，同时想快速融入企业并得到领导的赏识，那么本书绝对适合你。

□ 已经具备一定实践经验的数据从业者。对于已经在数据方面工作 1 ~ 3 年的从业者，相信你们会面临一些瓶颈，并想要在原有数据思维的基础上获得更有效的工作方法和工作价值的提升。本书中丰富的应用案例可以帮助你拨开云雾见青天。

□ 已经具有丰富工作经验的数据从业者。当数据从业者工作 3 年或 3 年以上时，就已经有机会从执行层走向管理层了。机会总是留给有准备的人的，作为管理者如何从数据工作流程、制度、风险和绩效方向进行思考并开展工作呢？相信本书会给你满意的答案。

如何阅读本书

本书的正文内容分为四篇，按照数据工作的成长思路来撰写，每篇都对应着一个成长阶段。

认知篇，介绍网站数据分析在企业内的价值体现，以及如何建立个人的数据成长体系、企业数据职能架构与数据价值最大化的定位。本篇的目的是帮助读者快速了解企业内部的整体数据架构、网站数据分析所扮演的角色以及个人在企业中如何进行自我定位和发展。

基础篇，分别介绍了网站数据的采集和配置、网站分析工具的选择、网站数据整合的方法、数据监测与评估指标以及数据分析场景和方法。这些知识是进行数据分析的基本前提，其中的第 7 章和第 8 章直接以业务场景为切入点，这部分知识可直接应用到实际工作中。

案例篇，以与网站分析结合最为紧密的营销和网站运营为例，分别介绍了网站数据的辅助决策以及数据驱动的工作内容，通过十多个案例还原数据分析的实际场景，这些案例可以套用到实际工作中并直接发挥作用。

提高篇，从数据管理者和领导者的角度介绍数据风险、数据质量、数据投入与产出、数据流程与落地管理，这些都是作为数据管理者自我提升的必备知识。

除正文内容外，本书还提供了两个附录。

附录 A 是关于网站分析工具中三个典型工具的特性的具体介绍，包括 Webtrekk（主要是 Q3）和 Adobe Analytics（主要是 Sitecatalyst）的默认报表和指标，以及 Universal Analytics 的通用和自定义代码的部署示例。

附录 B 总结了关于网站数据工作的局限与发展，从网站数据的价值、认知和技术局限性来阐述其所面临的挑战，展望了未来数据发展的三个方向：整合化、智能化、可视化。

勘误和支持

由于作者的水平有限以及编写时间仓促，书中难免会出现一些错误或者不准确的地方，恳请读者批评指正。为此，作者特意创建一个 QQ 群以方便大家讨论与沟通（群号：127586352），读者可以将书中的错误发布到群中；如果你遇到任何问题，也可以访问 <http://www.searchmarketingart.com/> 并在“书籍讨论区”进行提问，我将尽量在线上为读者提供最满意的解答。书中的全部源文件除可以从该网站“书籍讨论区”下载外，还可以从华章网站（www.hzbook.com）下载，我会将相应的功能更新及时在该网站发布出来。如果你有更多的宝贵意见，还欢迎发送邮件至邮箱：beijingtl@gmail.com，期待能够得到大家的真诚反馈。

致谢

首先要感谢机械工业出版社华章公司的杨福川老师，是他鼓励我写这本书并给予了详细的思路和专业指导，其次感谢全程参与审核、校验等工作的姜影编辑以及其他支持本书出版的相关工作者。

感谢我的领导彭亮以及我的良师益友田学峰，是他们在我的职业生涯中给我非常多的启发和知识拓展的机会，我在工作过程中接触到的各个世界级的网站数据解决方案以及数据工作知识都是基于前辈们已有的宝贵经验。另外，在从网站数据到企业级数据的角色转变过程中，我的几个好朋友、好同事提供了许多帮助，在他们的共同努力下，我才能获得较多关于企业级数据价值的资料，在此对他们也表示感谢，他们是姜继浩、庞程程、吕兆星、赵光娟、郑传峰等。除此之外，还有很多一起奋斗过的小伙伴，如徐子东、陈林、郭靖等，在此就不一一列举了，同样感谢他们对我工作的支持。

感谢我的领导 Webtrekk 亚太区总监张姝，是她给了我更多学习世界级网站数字智能解决

方案的机会并支持我撰写关于 Webtrekk 的相关内容。

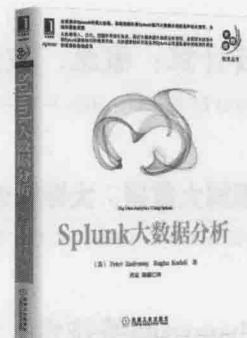
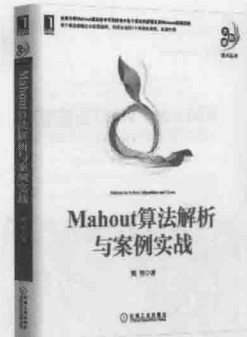
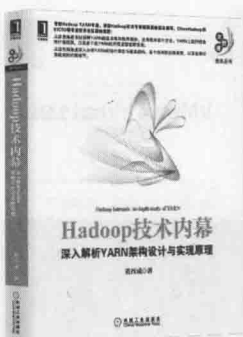
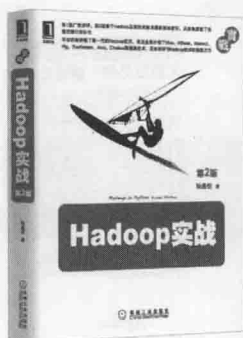
最后感谢我的爸爸、妈妈将我培养成人，并教我正确做人、做事的方法！感谢我的夫人，是她在我写作本书期间默默打理着家里的一切事务，使我有精力、有时间完成本书的全部撰写工作。

谨以此书献给我最亲爱的家人，以及众多热爱数据工作并努力为之奋斗的朋友们！

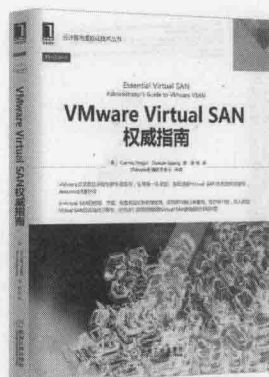
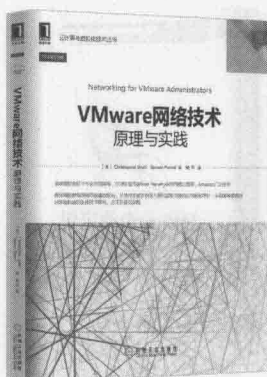
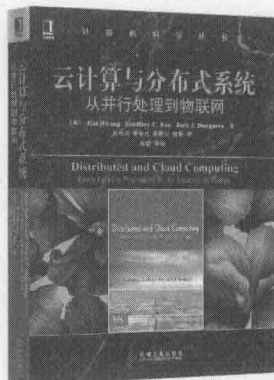
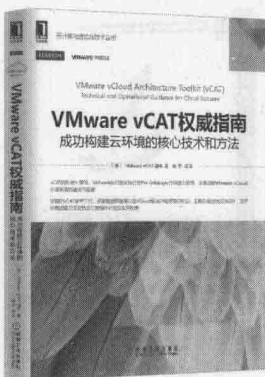
宋天龙 (Tony Song)

北京

推荐阅读



推荐阅读



云计算：概念、技术与架构

作者：Thomas Erl 等 ISBN：978-7-111-46134-0 定价：69.00元

云计算与分布式系统：从并行处理到物联网

作者：Kai Hwang 等 ISBN：978-7-111-41065-2 定价：85.00元

深入理解大数据：大数据处理与编程实践

作者：黄宣华 ISBN：978-7-111-47325-1 定价：79.00元

VMware vCAT权威指南：成功构建云环境的核心技术和方法

作者：VMware vCAT 团队 ISBN：978-7-111-48228-4 定价：119.00元

VMware网络技术：原理与实践

作者：Christopher Wahl 等 ISBN：978-7-111-47987-1 定价：59.00元

VMware Virtual SAN权威指南

作者：Cormac Hogan 等 ISBN：978-7-111-48023-5 定价：59.00元

Contents 目 录

本书赞誉

前 言

认知篇

第1章 科学地认识网站的数据分析 2

1.1 企业为什么要对网站的数据进行分析 2

1.1.1 网站数据分析能为“谁”解决问题 2

1.1.2 网站数据分析能解决哪些问题 3

1.1.3 常见的几种“分析”概念 5

1.2 网站数据分析的5个误区 7

1.3 识别网站数据具有欺骗性的3种形态 11

1.4 辅助决策与数据驱动的争议 15

1.4.1 辅助决策 15

1.4.2 数据驱动 16

1.4.3 辅助决策与数据驱动差异点 17

1.5 本章小结 17

第2章 从零开始建立企业数据体系 18

2.1 数据价值最大化的定位 18

2.1.1 数据价值定位的基本原则 18

2.1.2 数据价值的4种常见定位 20

2.2 企业数据的职能架构与组成 21

2.2.1 企业内部的职能架构 22

2.2.2 企业外部的职能架构 25

2.3 企业数据技术架构与组成 26

2.3.1 数据收集层 27

2.3.2 数据存储层 28

2.3.3 数据计算层 29

2.3.4 数据管理层 34

2.3.5 数据应用层 35

2.4 本章小结 36

第3章 从零开始建立数据分析师个人成长体系 37

3.1 数据分析师的完整知识结构 37

3.2 对数据分析师的职能素质要求 40

3.2.1 工作方向划分	40
3.2.2 工作职位划分	41
3.3 数据分析师成长的 4 个阶段	42
3.4 给数据分析师的 5 点建议	45
3.5 本章小结	47

基础篇

第 4 章 网站数据采集和配置

4.1 网站分析系统的数据工作机制	50
4.1.1 数据采集	50
4.1.2 数据处理	53
4.1.3 数据报告	55
4.2 网站代码部署	55
4.2.1 通用全局的脚本部署	56
4.2.2 通用页面的脚本部署	59
4.2.3 特定页面的脚本部署	62
4.3 系统功能配置	63
4.3.1 数据安全设置	63
4.3.2 数据处理设置	64
4.3.3 数据转化设置	79
4.3.4 数据整合设置	81
4.3.5 数据智能工作设置	82
4.4 本章小结	86

第 5 章 网站分析工具的选择

5.1 网站分析工具	87
5.1.1 Adobe Analytics	89
5.1.2 Webtrekk	106
5.1.3 Webtrends	112
5.1.4 Google Analytics	114

5.1.5 IBM Coremetrics	122
5.1.6 Piwik	125
5.1.7 百度统计	127
5.2 移动分析工具	128
5.2.1 Flurry	128
5.2.2 友盟	134
5.3 如何选择网站分析工具	135
5.3.1 整体解决方案的能力	135
5.3.2 产品易用性	136
5.3.3 功能丰富性	137
5.3.4 增值服务价值	140
5.3.5 价格和费用	141
5.4 本章小结	142

第 6 章 网站数据整合的方法

6.1 网站数据整合的意义	143
6.2 网站数据整合的范畴	144
6.2.1 业务数据整合	144
6.2.2 IT 数据整合	153
6.3 网站数据整合的方法	156
6.3.1 在线数据整合	156
6.3.2 本地数据整合	163
6.4 本章小结	170

第 7 章 数据监测与评估指标

7.1 业务效果流指标	171
7.1.1 站外推广类指标	171
7.1.2 网站运营类指标	172
7.1.3 企业会员类指标	182
7.1.4 呼叫中心类指标	186
7.1.5 仓储配送类指标	188
7.2 成本控制流指标	193

7.3 收益控制流指标	197
7.4 本章小结	201

第8章 数据分析场景和方法 202

8.1 以效果预测为目的的数据分析	202
8.1.1 效果预测是什么	202
8.1.2 效果预测的两种类型	203
8.1.3 效果预测的应用场景	203
8.1.4 预测结果的常用方法	204
8.2 以结论定义为目的的数据分析	211
8.2.1 结论定义是什么	211
8.2.2 结论定义的4种方向	212
8.2.3 结论定义的3个误区	213
8.2.4 下结论的常用方法	215
8.3 以数据探究为目的的数据分析	218
8.3.1 数据探究是什么	218
8.3.2 数据探究的两种类型	218
8.3.3 探究原因的分析方法	219
8.4 以业务执行为目的的数据分析	222
8.4.1 业务执行是什么	222
8.4.2 业务执行的两种类型	222
8.4.3 提取业务执行建议的常用 方法	223
8.5 正确的数据模型与算法选择观	228
8.6 本章小结	229

案例篇

第9章 网站数据的营销辅助应用 232

9.1 网站营销分析辅助决策报告 矩阵	232
------------------------------	-----

9.2 三种常用的网站营销分析 场景	239
9.2.1 营销前的媒体规划与效果 预测	239
9.2.2 营销时的异常检测与及时 反馈	241
9.2.3 营销结果总结与项目分析	243
9.3 常用的网站营销分析维度	245
9.3.1 目标端	245
9.3.2 媒体端	245
9.3.3 用户端	249
9.3.4 网站端	249
9.3.5 竞争端	251
9.3.6 其他因素	252
9.4 网站营销辅助决策四大案例	253
9.4.1 恶意流量分析	253
9.4.2 多渠道订单归因分析	259
9.4.3 渠道效果聚类	274
9.4.4 营销效果分析	286
9.5 本章小结	295

第10章 数据驱动下的数字营销

应用 297

10.1 数字营销的概念和范围	297
10.2 数字营销发展的三个阶段	298
10.3 个性化媒体投放的价值	298
10.4 个性化媒体投放的技术架构	299
10.4.1 数据层	300
10.4.2 算法层	301
10.4.3 API层	304
10.4.4 应用层	305

10.5	个性化媒体投放的实现	306
10.5.1	个性化媒体投放的实现方式	306
10.5.2	如何选择优秀的服务提供商	307
10.6	个性化媒体投放的问题	307
10.7	本章小结	309

第11章 网站数据的运营辅助应用

11.1	网站运营分析辅助决策报告矩阵	310
11.2	三类常见的网站运营分析场景	311
11.2.1	点：面向单体坑位的辅助分析	311
11.2.2	线：面向站内流程的优化与提高	314
11.2.3	面：面向整体网站资源的价值最大化	318
11.3	常用的网站运营分析维度	319
11.3.1	目标端	319
11.3.2	运营端	320
11.3.3	用户端	322
11.3.4	网站端	323
11.3.5	竞争端	324
11.3.6	其他因素	325
11.4	网站运营辅助决策四大案例	325
11.4.1	站内广告位效果标杆管理	326
11.4.2	网站用户调研	330

11.4.3	站内活动分析	338
11.4.4	商品销售诊断	354
11.5	本章小结	362

第12章 数据驱动下的个性化运营应用

12.1	网站运营的概念和范围	363
12.2	网站运营发展的三种形态	364
12.3	个性化网站运营的价值	364
12.4	个性化网站运营的应用	365
12.4.1	个性化网站运营的在线应用	365
12.4.2	个性化网站运营的离线应用	368
12.5	个性化网站运营的实现	369
12.6	个性化网站运营遇到的问题	371
12.7	本章小结	372

提高篇

第13章 数据风险管理与控制

13.1	数据风险管理的概念	374
13.2	数据风险管理的类型	375
13.3	数据风险管理的原则	378
13.4	数据风险管理与控制	379
13.5	本章小结	383

第14章 数据质量把控与建设

14.1	数据质量建设的内涵	384
------	-----------	-----