

数据化运营

系统方法与实践案例

赵宏田 江丽萍 李宁◎著



DATA-DRIVEN OPERATION

Systematic Methods and Practical Cases



机械工业出版社
China Machine Press

数据化运营

系统方法与实践案例

赵宏田 江丽萍 李宁◎著



DATA-DRIVEN OPERATION

Systematic Methods and Practical Cases



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

数据化运营：系统方法与实践案例 / 赵宏田，江丽萍，李宁著. —北京：机械工业出版社，2018.7

(产品管理与运营系列丛书)

ISBN 978-7-111-60451-8

I. 数… II. ①赵… ②江… ③李… III. 电子商务—企业管理—产品管理
IV. F713.36

中国版本图书馆 CIP 数据核字 (2018) 第 156325 号

数据化运营：系统方法与实践案例

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：李 艺

责任校对：李秋荣

印 刷：北京市荣盛彩色印刷有限公司

版 次：2018 年 8 月第 1 版第 1 次印刷

开 本：170mm×230mm 1/16

印 张：21

书 号：ISBN 978-7-111-60451-8

定 价：79.00 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) 88379426 88361066

投稿热线：(010) 88379604

购书热线：(010) 68326294 88379649 68995259

读者信箱：hzit@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

为什么要写这本书

作为互联网大数据行业的从业者，笔者将自己在这个行业中所掌握的知识、在日常工作中遇到的问题以及积累的项目经验整理成书，在这里和大家分享。希望通过本书能给读者一些新的理解和应用的思路，如果书中的内容能让你有些许收获，能够解决你工作中的一两个问题，那将是笔者的荣幸。

本书特色

本书从实践出发，结合工作中数据运营经验，以应用案例为主线，通过业务分析+代码实践这种更“接地气”的方式讲述数据的应用。书中对搭建数据监控指标体系、数据分析、数据挖掘、AB Test、埋点策略、用户画像建模等常见数据运营方式做了详细介绍。

根据读者的阅读习惯，本书由浅入深地分为基础篇、应用篇和提高篇三部分。

- 基础篇讲解数据运营常见场景、运营方式与数据运营人员的工作职责。

- 应用篇讲解数据分析方法与应用方式，并通过 5 个实际数据分析挖掘案例帮助读者深入理解数据运营方式。在内容编排上先提出案例应用背景和目标，再阐述分析方法和建模流程，最后完成数据的处理和业务上的应用。
- 提高篇讲解当下热门的“用户画像建模”，从建模流程、标签开发到画像应用，并辅以案例，讲解如何用 HQL 语言建立一个图书电商场景的用户标签体系，以帮助读者更深入地理解用户画像是如何建模打标签的。

本书没有过多复杂的理论公式，所讲案例操作步骤详细，可作为数据分析运营人员在解决实际问题中参考的“action book”。

本书适用对象

- 对数据分析及数据挖掘感兴趣的大专院校师生及其他初学者
- 对互联网行业数据分析、用户画像建模感兴趣的数据运营人员
- 互联网行业的产品运营人员及产品经理
- 各行各业的数据分析师

如何阅读本书

本书以互联网企业中常见的数据运营场景为切入点，以工作中实际面临的问题为案例，从方法、技术、业务、实践 4 个维度讲述数据运营的场景及应用方式。全书共分 10 章，各章的主要内容如下：

- 第 1 章介绍企业中数据的应用方式、企业数据职能架构与组成、数据运营人员的工作职责和应掌握的技能。
- 第 2 章介绍数据运营规划，常见的运营场景以及如何结合数据展开用户运营和流量运营工作。
- 第 3 章讲述如何结合业务搭建数据监控指标体系，从搭建模板到自动化数据报表，做好日报、周报、月报、专题分析报告等日常数据运营工作。
- 第 4 章是本书的重点章节，讲述了数据分析中常用的方法及其应用场

景，包括常见营销理论、数据分析思路、AB Test 分析以及埋点策略与分析。对数据分析感兴趣的人员可着重看此章节。

□ 第 5~9 章是数据分析挖掘中的几个具体案例，从数据运营的实际应用场景出发，以案例的形式讲述了如何在某些常见的业务需求背景下，分析项目需求，厘清思路，展开数据分析并输出报告和结论。案例中的关键步骤都附上了详细的代码说明。

□ 第 10 章是本书的重点章节，先介绍了什么是用户画像、应用场景、开发流程，然后以案例的形式讲解了用户画像建模过程中的需求分析、建立模型、打标签、计算标签权重、画像数据管理、用户画像应用等环节，并为案例附上了详细的代码说明。想要了解如何在用户画像建模过程中建模打标签的人员可着重看此章节。

勘误和支持

由于笔者水平有限，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，读者可通过邮箱（892798505@qq.com）或微信（administer-00001）反馈有关问题，笔者将尽全力为读者提供解答。书中的源代码可以从 Github 网站（<https://github.com/HunterChao/book>）下载。

致谢

在本书的编写过程中，得到了许多朋友的帮助，感谢天善智能创始人梁勇的支持与帮助，感谢为本书撰写推荐的朋友们，感谢你们的支持及专业的建议。

感谢机械工业出版社华章公司的杨福川副总编，本书从 2017 年 4 月开始筹划，从确定基本框架到后期的写作，杨总编不断地给笔者以指导，感谢杨总编的帮助与支持，与他的合作总是十分愉快！感谢机械工业出版社编辑李艺老师，李艺老师用严谨的态度孜孜不倦地帮助我们修改稿件。

谨以此书献给众多互联网运营人员和数据分析师们！

| 目 录 |

前 言

基 础 篇

第1章	概述：数据运营基础	002
	1.1 大数据时代	002
	1.2 企业数据应用方式	004
	1.3 数据运营的岗位职责	007
	1.4 数据运营应掌握的技能	009
	1.5 本章小结	013
第2章	业务：数据驱动运营	014
	2.1 如何用数据驱动运营	014
	2.1.1 定义数据分析目标	014
	2.1.2 目标分解与聚焦	016
	2.1.3 数据运营重点	019
	2.2 流量运营分析	021

2.2.1	流量运营规划	021
2.2.2	流量分析	023
2.2.3	解读 PV、UV	027
2.2.4	跳出率分析	029
2.2.5	漏斗图分析	030
2.2.6	A/B 测试	032
2.3	用户运营分析	033
2.3.1	用户分群	034
2.3.2	用户行为分析	040
2.3.3	用户生命周期价值	047
2.4	本章小结	051
第3章	报表：数据管理模板	052
3.1	个性化数据管理报告——Excel	054
3.1.1	创建报告的准备工作	054
3.1.2	报告自动化步骤	055
3.1.3	从数据源表到数据转化表	056
3.1.4	报告正文展示	062
3.1.5	自动化报表脚本	064
3.2	搭建数据分析报告模板——PPT	066
3.2.1	业务指标梳理（搭建运营监控指标体系）	067
3.2.2	分析思路与框架	078
3.2.3	图表展现	079
3.2.4	数据与结论	080
3.2.5	报告布局与排版	081
3.2.6	PPT 随 Excel 模板自动更新	084
3.3	本章小结	085

应 用 篇

第4章 理论：数据分析方法	088
4.1 数据分析理论模型	088
4.1.1 4P 营销理论	089
4.1.2 5W2H 分析法	090
4.1.3 PEST 分析方法	092
4.1.4 SWOT	093
4.1.5 逻辑树	095
4.2 数据分析方法与运用场景	095
4.2.1 多维分析	095
4.2.2 趋势分析	097
4.2.3 综合评价法	101
4.2.4 转化分析	103
4.2.5 数据挖掘方法	106
4.3 可视化：常用图表的特点及适用场合	106
4.3.1 环形图	107
4.3.2 矩阵图	108
4.3.3 组合图	112
4.3.4 文字云	118
4.4 AB Test 的原理与实现	125
4.4.1 AB Test 的原理	126
4.4.2 AB Test 的埋点与报表部署	128
4.4.3 AB Test 的分析方法	129
4.4.4 AB Test 的常见误区	132
4.5 埋点策略与实现	134
4.5.1 utm 来源埋点	135

4.5.2	页面 PV 埋点	137
4.5.3	单击埋点 native	139
4.5.4	单击埋点 hybrid	141
4.5.5	业务埋点	142
4.5.6	曝光埋点	144
4.5.7	埋点常见问题	145
4.6	本章小结	146
第5章	案例：竞品数据对标分析	148
5.1	网络爬虫基础知识	148
5.1.1	开发环境准备	149
5.1.2	Web 前端基础	149
5.1.3	解析网页	152
5.1.4	数据存储	159
5.2	网站结构分析	166
5.3	Scrapy 爬虫架构	168
5.3.1	items 模块	170
5.3.2	pipelines 模块	171
5.3.3	settings 模块	172
5.3.4	爬虫模块	173
5.4	数据爬取与解析	174
5.5	项目优化与改进	177
5.5.1	爬虫脚本部署在服务器端	178
5.5.2	分布式爬虫的实现	178
5.6	反爬手段及应对机制	179
5.6.1	禁止 IP 请求	180
5.6.2	禁止非浏览器访问	180

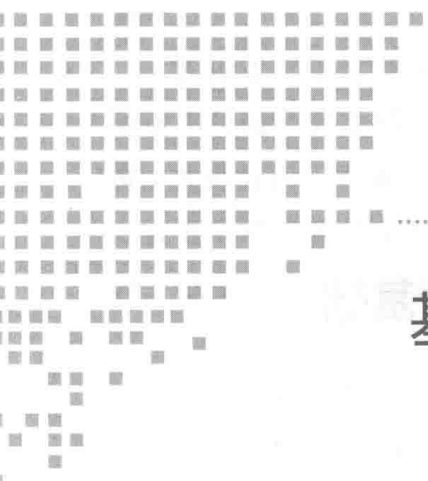
5.6.3	ajax 加载目标数据	181
5.6.4	需要登录后才能访问	182
5.6.5	手机 App 页面数据抓取	182
5.7	本章小结	184
第6章	案例：某互联网医疗产品用户特征分析	185
6.1	应用背景与分析维度	185
6.2	基于用户细分的行为分析	186
6.3	用户来源渠道分析	190
6.4	基于前端展示的用户行为分析	191
6.5	产品改进与运营建议	195
6.6	本章小结	195
第7章	案例：RFM用户价值模型应用	196
7.1	应用背景与目标	196
7.2	基于规则的划分	198
7.3	基于聚类方法的划分	203
7.4	本章小结	209
第8章	案例：用户流失分析与预测	210
8.1	应用背景与目标	210
8.2	问题分析与模型构建	211
8.3	数据处理与结果	212
8.3.1	确定用户流失周期	212
8.3.2	抽取训练数据建立决策树模型	214
8.3.3	线上部署脚本定期监测流失用户	221

8.3.4	流失用户分析	224
8.4	问题定位与解决方案	226
8.5	本章小结	229
第9章	案例：站内文章自动分类打标签	230
9.1	应用背景与目标	230
9.2	问题分析与模型构建	231
9.3	案例中主要应用的技术	232
9.3.1	数据预处理	232
9.3.2	TF-IDF 词空间向量转换	233
9.3.3	文章关键词提取	234
9.3.4	朴素贝叶斯分类	235
9.4	数据处理与模型检验	235
9.4.1	文本分词处理（数据分类与数据预处理）	236
9.4.2	数据结构处理	238
9.4.3	计算文本的 TF-IDF 权重矩阵	240
9.4.4	用朴素贝叶斯方法分类文章	242
9.5	本章小结	245

提 高 篇

第10章	应用：用户画像建模	248
10.1	用户画像简介	248
10.1.1	什么是用户画像	249
10.1.2	用户画像模型及应用场景	250
10.1.3	数仓架构及项目流程	254
10.2	用户画像管理	257

10.2.1	模块化开发	257
10.2.2	存储方式	259
10.2.3	更新机制	259
10.3	业务背景	262
10.3.1	案例背景介绍	262
10.3.2	数据仓库相关表介绍	262
10.4	用户画像建模	267
10.4.1	业务需求梳理	267
10.4.2	用户标签体系及开发内容	268
10.4.3	用户画像开发流程	274
10.4.4	时间衰减系数	279
10.4.5	标签权重配置	280
10.5	用户画像数据开发	282
10.5.1	建立用户属性画像	283
10.5.2	建立用户行为画像	289
10.5.3	建立用户偏好画像	303
10.5.4	建立群体用户画像	308
10.5.5	画像效果验收	313
10.5.6	画像数据质量管理	314
10.6	用户画像应用方式	317
10.6.1	业务精细化运营	317
10.6.2	数据分析	319
10.6.3	精准营销	319
10.6.4	用户个性化推荐	322
10.7	本章小结	323



基础篇

- 第1章 概述：数据运营基础
 - 第2章 业务：数据驱动运营
 - 第3章 报表：数据管理模板
- 

第 1 章

概述：数据运营基础

1.1 大数据时代

随着信息化时代的来临和全球经济化的发展，数据存储成本的不断下降，企业数据的总量正在以惊人的速度增长，这些数据是企业的重要资源。随着近几年移动互联网和物联网的兴起，信息传播技术和传播渠道得到了极大的发展，数据的涌现呈“爆炸”形式增长。如图 1-1 所示，从百度指数中搜索“大数据”可以看到该关键词自 2013 年起的关注度呈逐年上升趋势，各行业已逐渐意识到数据在企业经营发展中的重要作用。

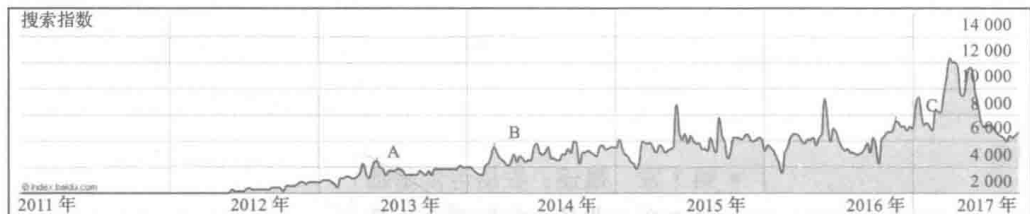


图 1-1 百度指数 - 大数据趋势图

注：图截取自百度指数，图中相关内容的著作权归原著作人所有

从电商网站的“看了又看”“猜你喜欢”等个性化推荐，商家促销活动的

精准营销，到智慧城市建设的一站式服务等应用场景，大数据已逐渐渗透、应用、影响到我们的工作、生活中，成为提高我们生活、工作质量的重要组成部分。

各行业产生的海量数据只有通过技术手段进行存储、分析、挖掘才能创造出价值。大数据领域涵盖了数据采集、数据存储、数据处理、数据挖掘、数据分析、数据呈现等一系列的技术手段，随着技术的不断成熟，并逐渐被更多富数据的行业和企业所应用，各个行业在得到大数据带来收益的同时，也在推动大数据技术的快速发展。下面简要介绍几个行业中大数据的应用。

- 互联网行业：互联网行业的主要特征之一就是各种类型的数据都呈爆炸式增长。用户在互联网上的丰富行为都能被网站日志所记录，网站可以利用大数据技术从海量用户数据中挖掘出有价值的信息，建立用户模型，针对性提供产品和服务，提高用户体验。
- 零售行业：零售企业需要根据销售有特色的本地化商品并增加流行款式和生命周期短的产品，零售企业需要对大量的用户消费行为进行分析，预测出未来的消费需求，迅速提供有针对性的个性化服务。零售行业需要增强产品流转率，实现快速营销。
- 保险行业：保险机构依托互联网和移动通信等技术，通过自营网络平台和第三方平台等订立保险合同，提供保险业务服务。依托互联网大数据的优势实现差异化产品和服务，在提升用户黏性和品牌认知度的同时促进了客户价值转化。
- 制造业：采集产品研发、投放、销售、购买、评论等全流程数据，在融合内外部数据的基础上建立用户画像，让用户需求成为产品设计导向，使新产品更符合用户习惯和期望，实现大数据驱动的产销模式。

本书主要站在互联网行业大数据处理、分析、运用的角度，以分析方法、实践案例、辅助代码详解的方式讲述了大数据在企业运营与决策中的应用方式。

1.2 企业数据应用方式

数据激增是当今企业的一大特性，如何有效地利用企业经营过程中产生的数据，从海量信息中提取出有用的模式并对其进行分析、挖掘、应用已成为人们的迫切需求。企业正不断将数据分析、数据挖掘视为重要部分，将数据转化为商业智能，提高企业的核心竞争力。企业中数据从产生到应用依次需要经过数据源层、数据仓库层、数据建模层，最后到数据应用层，经过层层加工，从原始的海量数据经过各层的清洗、建模、挖掘之类的加工后逐渐支持到上游的应用环节。企业数据应用流程如图 1-2 所示。

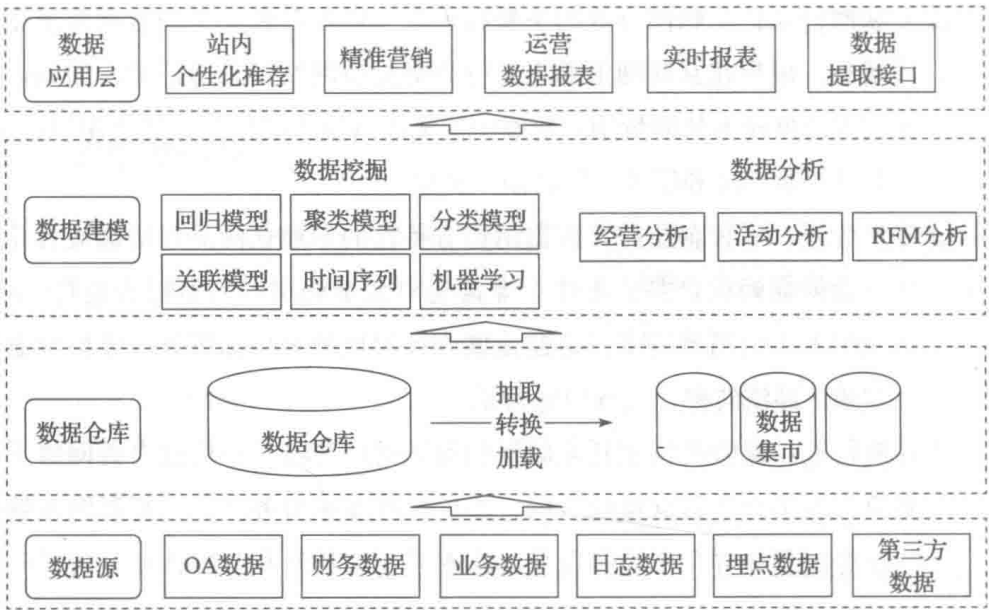


图 1-2 企业数据应用流程

1. 数据源

数据是商务智能的基础，数据种类通常包括企业内部的 OA 数据、财务数据、BD 数据、业务数据、日志数据、埋点数据和外部第三方数据。数据需要经过抽取、转换和装载，即经过 ETL 后才可以存储在数据仓库中心，为数据分析奠定基础。