

<<数据挖掘与数据化运营实战>>

图书基本信息

书名：<<数据挖掘与数据化运营实战>>

13位ISBN编号：9787111426509

10位ISBN编号：7111426509

出版时间：2013-6

出版时间：卢辉 机械工业出版社 (2013-06出版)

作者：卢辉

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：<http://www.tushu007.com>

<<数据挖掘与数据化运营实战>>

前言

前言为什么要写这本书自从2002年第一次接触“数据挖掘”（Data Mining）这个新名词以来，转眼之间我已经在数据挖掘商业应用相关领域度过了11年。

这11年里我既见识了国外数据挖掘商业应用如火如荼地开展；又经历了从21世纪开始，国内企业在数据挖掘商业应用中的摸索起步，到如今方兴未艾的局面；更有幸在经历了传统行业的数据挖掘商业应用之后，投身到互联网行业（当今数据分析商业应用热火朝天、发展最快，并且对数据和数据挖掘的商业应用依赖性最强的行业）的数据挖掘商业实践中。

这11年是我职业生涯中最为重要的一段时光，从个人生存的角度来说，我找到了谋生和养家糊口的饭碗—数据挖掘工作；从个人归属的角度来说，我很幸运地碰到了职业与兴趣的重合点。

在国内，“数据挖掘”作为一门复合型应用学科，其在商业领域的实践应用及推广只有十几年的时间，在此期间，国内虽然陆续出版了一些相关的书籍，但是绝大多数都是基于理论或者国外经验来阐述的，少有针对国内企业相关商业实战的详细介绍和分享，更缺乏从数据分析师的角度对商业实战所进行的总结和归纳。

因此，从商业应用出发，基于大量的商业实战案例而不是基于理论探讨的数据挖掘应用书籍成为当今图书市场和广大“数据挖掘”学习者的共同需求。

同时，在有幸与数据挖掘商业实践相伴11年之后，我也想稍微放慢些脚步，正如一段长途跋涉之后需要停下脚步，整理一路经历的收获和感悟一样，我希望将自己一路走来的心得与体会、经验与教训、挫折与成绩整理出来。

基于以上原因，我决定从数据挖掘的商业需求和商业实战出发，结合我10多年来在不同行业（尤其是最近4年在互联网行业）的大量数据挖掘商业实战项目，将自己这些年来积累的经验和总结分享出来，希望能够起到抛砖引玉的作用，为对数据挖掘商业实践感兴趣的朋友、爱好者、数据分析师提供点滴的参考和借鉴。

同时，鉴于“数据化运营”在当今大数据时代已经成为众多（以后必将越来越多）现代企业的普遍经营战略，相信本书所分享的大量有关数据化运营的商业实践项目也可以为企业的管理层、决策层提供一定程度的参考和借鉴。

我相信，本书总结的心得与体会，可以推动自己今后的工作，会成为我的财富；同时，这些心得与体会对于部分数据分析师来说也可以起到不同程度的参考和借鉴作用；对于广大对数据挖掘商业应用感兴趣的初学者来说也未尝不是一种宝贵经验。

我是从机械制造工艺与设备这个与“数据挖掘”八竿子打不着的专业转行到数据挖掘商业应用行业的，这与目前国内绝大多数的数据分析挖掘专业人士的背景有较大差别（国内绝大多数数据分析挖掘专业人士主要来自统计专业、数学专业或者计算机专业）。

我的职业道路很曲折，之所以放弃了自己没兴趣的机械制造工艺与设备专业，是因为自己喜欢市场营销。

有幸在国外学习市场营销专业时了解并亲近了国外市场营销中的核心和基石—市场营销信息学（Marketing Informatics）。

当然，这是国外10多年前的说法，换成行业内与时俱进的新说法，就是时下耳熟能详的“数据分析挖掘在市场营销领域的商业实践应用”）。

说这么多，其实只是想告诉有缘的对数据挖掘商业实践感兴趣的朋友，“以业务为核心，以思路为重点，以挖掘技术为辅佐”就是该领域的有效成长之路。

很多初学者总以为掌握了某些分析软件，就可以成为数据分析师。

其实，一个成功的数据挖掘商业实践，核心的因素不是技术，而是业务理解和分析思路。

本书自始至终都在力图用大量的事实和案例来证明“以业务为核心，以思路为重点，以挖掘技术为辅佐”才是数据挖掘商业实践成功的宝典。

另外，现代企业面对大数据时代的数据化运营绝不仅仅是数据分析部门和数据分析师的事情，它需要企业各部门的共同参与，更需要企业决策层的支持和推动。

读者对象对数据分析和数据挖掘的商业实践感兴趣的大专院校师生、对其感兴趣的初学者。

<<数据挖掘与数据化运营实战>>

互联网行业对数据分析挖掘商业实践感兴趣的运营人员以及其他专业的人士。

实施数据化运营的现代企业的运营人员以及其他专业的人士，尤其是企业的管理者、决策者（数据化运营战略的制定者和推动者）。

各行各业的数据分析师、数据挖掘师。

勘误和支持由于作者水平和能力有限，编写时间仓促，不妥之处在所难免，在此恳请读者批评指正。作者有关数据挖掘商业实践应用的专业博客“数据挖掘人在旅途”，欢迎读者和数据挖掘商业实践的爱好者不吝赐教。

另外，如果您有关于数据挖掘商业实践的任何话题，也可以发送邮件到邮箱，期待你们的反馈意见。如何阅读本书本书分为19章。

第1~4章为基础和背景部分，主要介绍数据分析挖掘和数据化运营的相关背景、数据化运营中“协调配合”的本质，以及实践中常见的分析项目类型。

第6~13章是数据分析挖掘中的具体技巧和案例分享部分，主要介绍实践中常见的分析挖掘技术的实用技巧，并对大量的实践案例进行了全程分享展示。

第5章，第14~19章是有关数据分析师的责任、意识、思维的培养和提升的总结与探索，以及一些有效的项目质控制度和经典的方法论。

本书几乎每章都会用至少一个完整翔实的实战案例来进行说明、反复强化“以业务为核心，以思路为重点，以挖掘技术为辅佐”，希望能给读者留下深刻印象，因为这是数据挖掘商业实践成功的宝典。致谢首先要感谢机械工业出版社华章公司的杨绣国（Lisa）编辑，没有您的首倡和持续的鼓励，我不会想到要写这样一本来自实践的书，也不会顺利地完成这本书。

写作过程中，您的帮助让我对“编辑”这个职业有了新的认识，编辑就是作者背后的无名英雄。

在本书出版之际，我向Lisa表达我深深的感谢和祝福。

同时感谢朱秀英编辑在本书后期编辑过程中付出的辛劳，您的专业、敬业和细心使得书稿中诸多不完善之处得以修正和提高。

作为一名30多岁才从机械工程师转行，进入数据挖掘及其商业实践的迟到者，我在数据挖掘的道路上一路走来，得到了无数贵人的帮助和提携。

感谢我的启蒙导师，加拿大Dalhousie University的数据挖掘课程教授Tony Schellinck。

他风趣幽默的授课风格，严谨扎实的专业功底，随手拈来的大量亲身经历的商业实战案例，以及对学生的耐心和热情，让我作为一名外国学生能有效克服语言和生活环境的挑战，比较顺利地进入数据挖掘的职业发展道路。

感谢回国后给我第一份专业工作机会的前CCG集团（Communication Central Group）商业智能应用事业部总经理Justin Jencks。

中国通Justin在我们一起共事的那段日子里，果敢放手让我尝试多个跨行业的探索性商业应用项目，给了我许多宝贵的机会，使我迅速熟悉本土市场，积累了不同行业的实战案例，这些对我的专业成长非常重要。

感谢4年前给我机会，让我得以从传统行业进入互联网行业的阿里巴巴集团ITBU事业部的前商业智能部门总监李红伟（菠萝）。

进入互联网行业之后，我才深深懂得作为一名数据分析师，相比传统行业来说，互联网行业有太多的机会可以去尝试不同的项目，去亲历数不清的“一竿子插到底”的落地应用，去学习面对日新月异的需求和挑战。

在本书的编写过程中，得到了淘宝网的商品推荐高级算法工程师陈凡和阿里巴巴B2B的数据仓库专家蒿亮热情而专业的帮助和支持。

陈凡友情编写了本书的3.11节，蒿亮友情编写了本书的1.4.1节和13.1节。

感谢一路走来，在项目合作和交流中给我帮助和支持的各位前辈、领导、朋友和伙伴，包括：上海第一医药连锁经营有限公司总经理顾咏晟先生、新华信国际信息咨询北京有限公司副总裁欧万德先生（Alvin）、上海联都集团的创始人冯铁军先生、上海通方管理咨询有限公司总经理李步峰女士和总监张国安先生、鼎和保险公司的张霖霏先生、盛大文学的数据分析经理张仙鹤先生、途牛网高级运营专家焦延伍先生，以及来自阿里巴巴的数据分析团队的领导和伙伴（资深总监车品觉先生、高级专家范

<<数据挖掘与数据化运营实战>>

国栋先生、资深经理张高峰先生、数据分析专家樊宁先生、资深数据分析师曹俊杰先生、数据分析师宫尚宝先生，等等，尤其要感谢阿里巴巴数据委员会会长车品觉老师在百忙中热情地为本书作推荐序，并在序言里为广大读者分享了数据分析师当前面临的最新机遇和挑战），以及这个仓促列出的名单之外的更多前辈、领导、朋友和伙伴。

感谢我的父母、姐姐、姐夫和外甥，他们给予了我一贯的支持和鼓励。

我将把深深的感谢给予我的妻子王艳和女儿露璐。

露璐虽然只是初中一年级的学生，但是在本书的写作过程中，她多次主动放弃外出玩耍，帮我改稿，给我提建议，给我鼓励，甚至还为本书设计了一款封面，在此向露璐同学表达我衷心的感谢！

而我的妻子，则将家里的一切事情打理得井井有条，使我可以将充分的时间和精力投入本书的写作中。

谨以此书献给她们！

卢辉中国 杭州

<<数据挖掘与数据化运营实战>>

内容概要

《数据挖掘与数据化运营实战:思路、方法、技巧与应用》是目前有关数据挖掘在数据化运营实践领域比较全面和系统的著作，也是诸多数据挖掘书籍中为数不多的穿插大量真实的实践应用案例和场景的著作，更是创造性地针对数据化运营中不同分析挖掘课题类型，推出一一对应的分析思路集锦和相应的分析技巧集成，为读者提供“菜单化”实战锦囊的著作。

作者结合自己数据化运营实践中大量的项目经验，用通俗易懂的“非技术”语言和大量活泼生动的案例，围绕数据分析挖掘中的思路、方法、技巧与应用，全方位整理、总结、分享，帮助读者深刻领会和掌握“以业务为核心，以思路为重点，以分析技术为辅佐”的数据挖掘实践应用宝典。

《数据挖掘与数据化运营实战:思路、方法、技巧与应用》共19章，分为三个部分：基础篇（第1~4章）系统介绍了数据分析挖掘和数据化运营的相关背景、数据化运营中“协调配合”的核心，以及实践中常见分析项目类型；实战篇（第6~13章）主要介绍实践中常见的分析挖掘技术的实用技巧，并对大量的实践案例进行了全程分享展示；思想意识篇（第5章，第14~19章）主要是有关数据分析师的责任、意识、思维的培养和提升的总结和探索，以及一些有效的项目质控制度和经典的方法论介绍。

海报：

<<数据挖掘与数据化运营实战>>

作者简介

卢辉，阿里巴巴商业智能部数据分析专家，从事数据库营销和数据化运营分析多年，曾在不同行业以商务拓展(BD)经理、项目经理、市场营销部经理、高级咨询顾问、数据分析专家的身份亲历大量的数据库营销和互联网行业数据化运营应用项目。

目前在阿里巴巴主要从事数据化运营的数据挖掘规划、项目管理、实施，拥有比较丰富的互联网行业数据化运营项目经验。

关注数据化运营的规划和数据挖掘项目的管理。

书籍目录

推荐序前言第1章 什么是数据化运营 11.1 现代营销理论的发展历程 21.1.1 从4P到4C 21.1.2 从4C到3P3C
31.2 数据化运营的主要内容 51.3 为什么要数据化运营 71.4 数据化运营的必要条件 81.4.1 企业级海量数
据存储的实现 81.4.2 精细化运营的需求 101.4.3 数据分析和数据挖掘技术的有效应用 111.4.4 企业决策层
的倡导与持续支持 111.5 数据化运营的新现象与新发展 121.6 关于互联网和电子商务的最新数据 14第2
章 数据挖掘概述 152.1 数据挖掘的发展历史 162.2 统计分析与数据挖掘的主要区别 162.3 数据挖掘的主
要成熟技术以及在数据化运营中的主要应用 182.3.1 决策树 182.3.2 神经网络 192.3.3 回归 212.3.4 关联规
则 222.3.5 聚类 232.3.6 贝叶斯分类方法 242.3.7 支持向量机 252.3.8 主成分分析 262.3.9 假设检验 272.4 互
联网行业数据挖掘应用的特点 28第3章 数据化运营中常见的数据分析项目类型 303.1 目标客户的特征分
析 313.2 目标客户的预测(响应、分类)模型 323.3 运营群体的活跃度定义 333.4 用户路径分析 343.5 交
叉销售模型 353.6 信息质量模型 373.7 服务保障模型 393.8 用户(买家、卖家)分层模型 403.9 卖家(买
家)交易模型 443.10 信用风险模型 443.11 商品推荐模型 453.11.1 商品推荐介绍 453.11.2 关联规则
453.11.3 协同过滤算法 503.11.4 商品推荐模型总结 543.12 数据产品 553.13 决策支持 56第4章 数据化运营
是跨专业、跨团队的协调与合作 574.1 数据分析团队与业务团队的分工和定位 584.1.1 提出业务分析需
求并且能胜任基本的数据分析 584.1.2 提供业务经验和参考建议 604.1.3 策划和执行精细化运营方案
604.1.4 跟踪运营效果、反馈和总结 614.2 数据化运营是真正的多团队、多专业的协同作业 624.3 实例示
范数据化运营中的跨专业、跨团队协调合作 62第5章 分析师常见的错误观念和对治的管理策略 675.1 轻
视业务论 685.2 技术万能论 695.3 技术尖端论 715.4 建模与应用两段论 725.5 机器万能论 735.6 幸福的家
庭都是相似的,不幸的家庭各有各的不幸 74第6章 数据挖掘项目完整应用案例演示 766.1 项目背景和业
务分析需求的提出 776.2 数据分析师参与需求讨论 786.3 制定需求分析框架和分析计划 796.4 抽取样本
数据、熟悉数据、数据清洗和摸底 816.5 按计划初步搭建挖掘模型 816.6 与业务方讨论模型的初步结论
,提出新的思路和模型优化方案 836.7 按优化方案重新抽取样本并建模,提炼结论并验证模型 846.8 完
成分析报告和落地应用建议 866.9 制定具体的落地应用方案和评估方案 866.10 业务方实施落地应用方
案并跟踪、评估效果 866.11 落地应用方案在实际效果评估后,不断修正完善 886.12 不同运营方案的评
估、总结和反馈 886.13 项目应用后的总结和反思 89第7章 数据挖掘建模的优化和限度 907.1 数据挖
掘模型的优化要遵循有效、适度的原则 917.2 如何有效地优化模型 927.2.1 从业务思路优化 927.2.2 从建
模的技术思路优化 947.2.3 从建模的技术技巧上优化 957.3 如何思考优化的限度 967.4 模型效果评价的
主要指标体系 967.4.1 评价模型准确度和精度的系列指标 977.4.2 ROC曲线 997.4.3 KS值 1007.4.4 Lift值
1027.4.5 模型稳定性的评估 104第8章 常见的数据处理技巧 1058.1 数据的抽取要正确反映业务需求
1068.2 数据抽样 1078.3 分析数据的规模有哪些具体的要求 1088.4 如何处理缺失值和异常值 1098.4.1 缺
失值的常见处理方法 1098.4.2 异常值的判断和处理 1118.5 数据转换 1128.5.1 生成衍生变量 1138.5.2 改
善变量分布的转换 1138.5.3 分箱转换 1148.5.4 数据的标准化 1158.6 筛选有效的输入变量 1158.6.1 为
什么要筛选有效的输入变量 1168.6.2 结合业务经验进行先行筛选 1168.6.3 用线性相关性指标进行初步
筛选 1178.6.4 R平方 1188.6.5 卡方检验 1198.6.6 IV和WOE 1208.6.7 部分建模算法自身的筛选功能
1218.6.8 降维的方法 1228.6.9 最后的准则 1228.7 共线性问题 1238.7.1 如何发现共线性 1238.7.2 如
何处理共线性 123第9章 聚类分析的典型应用和技术小窍门 1259.1 聚类分析的典型应用场景 1269.2 主
要聚类算法的分类 1279.2.1 划分方法 1279.2.2 层次方法 1289.2.3 基于密度的方法 1289.2.4 基于网
格的方法 1299.3 聚类分析在实践应用中的重点注意事项 1299.3.1 如何处理数据噪声和异常值 1299.3.2
数据标准化 1309.3.3 聚类变量的少而精 1319.4 聚类分析的扩展应用 1329.4.1 聚类的核心指标与非聚
类的业务指标相辅相成 1329.4.2 数据的探索 and 清理工具 1329.4.3 个性化推荐的应用 1339.5 聚类分
析在实际应用中的优势和缺点 1349.6 聚类分析结果的评价体系和评价指标 1359.6.1 业务专家的评估
1359.6.2 聚类技术上的评价指标 1369.7 一个典型的聚类分析课题的案例分享 1379.7.1 案例背景
1379.7.2 基本的数据摸底 1379.7.3 基于用户样本的聚类分析的初步结论 138第10章 预测响应(分
类)模型的典型应用和技术小窍门 14010.1 神经网络技术的实践应用和注意事项 14110.1.1 神经网络
的原理和核心要素 14110.1.2 神经网络的应用优势 14310.1.3 神经网络技术的缺点和注意事项 14310.2
决策树技术的实践应用和注意事项 14410.2.1 决策树的原理和核心要素 14410.2.2 CHAID算法 14510.2.3
CART算法 14510.2.4 ID3算法 14610.2.5 决策树的应用优势 14610.2.6

<<数据挖掘与数据化运营实战>>

决策树的缺点和注意事项 14710.3 逻辑回归技术的实践应用和注意事项 14810.3.1 逻辑回归的原理和核心要素 14810.3.2 回归中的变量筛选方法 15010.3.3 逻辑回归的应用优势 15110.3.4 逻辑回归应用中的注意事项 15110.4 多元线性回归技术的实践应用和注意事项 15210.4.1 线性回归的原理和核心要素 15210.4.2 线性回归的应用优势 15310.4.3 线性回归应用中的注意事项 15310.5 模型的过拟合及对策 15410.6 一个典型的预测响应模型的案例分享 15610.6.1 案例背景 15610.6.2 基本的数据摸底 15610.6.3 建模数据的抽取和清洗 15810.6.4 初步的相关性检验和共线性排查 15910.6.5 潜在自变量的分布转换 16010.6.6 自变量的筛选 16110.6.7 响应模型的搭建与优化 16210.6.8 冠军模型的确定和主要的分析结论 16210.6.9 基于模型和分析结论基础上的运营方案 16410.6.10 模型落地应用效果跟踪反馈 165第11章 用户特征分析的典型应用和技术小窍门 16611.1 用户特征分析所适用的典型业务场景 16711.1.1 寻找目标用户 16711.1.2 寻找运营的抓手 16811.1.3 用户群体细分的依据 16911.1.4 新品开发的线索和依据 16911.2 用户特征分析的典型分析思路和分析技术 17011.2.1 3种划分的区别 17011.2.2 RFM 17111.2.3 聚类技术的应用 17211.2.4 决策树技术的应用 17311.2.5 预测（响应）模型中的核心自变量 17311.2.6 假设检验的应用 17411.3 特征提炼后的评价体系 17411.4 用户特征分析与用户预测模型的区别和联系 17511.5 用户特征分析案例 176第12章 运营效果分析的典型应用和技术小窍门 17712.1 为什么要做运营效果分析 17812.2 统计技术在数据化运营中最重要最常见的应用 17912.2.1 为什么要进行假设检验 17912.2.2 假设检验的基本思想 17912.2.3 T检验概述 18012.2.4 两组独立样本T检验的假设和检验 18112.2.5 两组独立样本的非参数检验 18212.2.6 配对差值的T检验 18312.2.7 配对差值的非参数检验 18412.2.8 方差分析概述 18612.2.9 单因素方差分析 18712.2.10 多个样本组的非参数检验 19012.2.11 卡方检验 19012.2.12 控制变量的方法 19112.2.13 AB Test 192第13章 漏斗模型和路径分析 19313.1 网络日志和布点 19413.1.1 日志布点 19513.1.2 日志采集 19513.1.3 日志解析 19513.1.4 日志分析 19513.2 漏斗模型与路径分析的主要区别和联系 19613.3 漏斗模型的主要应用场景 19713.3.1 运营过程的监控和运营效率的分析与改善 19713.3.2 用户关键路径分析 19813.3.3 产品优化 19813.4 路径分析的主要应用场景 19813.5 路径分析的主要算法 19913.5.1 社会网络分析方法 19913.5.2 基于序列的关联分析 20013.5.3 最朴素的遍历方法 20113.6 路径分析案例的分享 20313.6.1 案例背景 20313.6.2 主要的分析技术介绍 20313.6.3 分析所用的数据概况 20313.6.4 主要的结论和业务解说 20313.6.5 主要分析结论的落地应用跟踪 206第14章 数据分析师对业务团队数据分析能力的培养 20814.1 培养业务团队数据分析意识与能力的重要性 20914.2 数据分析师在业务团队数据分析意识能力培养中的作用 21014.3 数据分析师如何培养业务团队的数据分析意识和能力 21014.4 数据分析师培养业务团队数据分析意识能力的案例分享 21214.4.1 案例背景 21214.4.2 过程描述 21214.4.3 本项目的效果跟踪 214第15章 换位思考 21615.1 为什么要换位思考 21715.2 从业务方的角度换位思考数据分析与挖掘 21815.3 从同行的角度换位思考数据分析挖掘的经验教训 220第16章 养成数据分析师的品质和思维模式 22216.1 态度决定一切 22316.1.1 信念 22316.1.2 信心 22416.1.3 热情 22516.1.4 敬畏 22616.1.5 感恩 22716.2 商业意识是核心 22816.2.1 为什么商业意识是核心 22816.2.2 如何培养商业意识 22916.3 一个基本的方法论 23016.4 大胆假设，小心求证 23116.5 20/80原理 23316.6 结构化思维 23316.7 优秀的数据分析师既要客观，又要主观 234第17章 条条大道通罗马 23617.1 为什么会条条大道通罗马 23717.2 条条大道有侧重 23817.3 自觉服从和积极响应 23917.3.1 自觉服从 23917.3.2 积极响应 24017.4 具体示例 242第18章 数据挖掘实践的质量保障流程和制度 24318.1 一个有效的质量保障流程制度 24418.1.1 业务需求的收集 24518.1.2 评估小组评估需求的优先级 24618.1.3 课题组的成立及前期摸底 24718.1.4 向业务方提交正式课题（项目）计划书 24718.1.5 数据分析挖掘的课题展开 24818.1.6 向业务方提交结论报告及业务落地应用建议 24818.1.7 课题（项目）的落地应用和效果监控反馈 24818.2 质量保障流程制度的重要性 24918.3 如何支持与强化质量保障流程制度 250第19章 几个经典的数据挖掘方法论 25119.1 SEMMA方法论 25219.1.1 数据取样 25319.1.2 数据探索 25319.1.3 数据调整 25319.1.4 模式化 25419.1.5 评价 25419.2 CRISP-DM方法论 25419.2.1 业务理解 25519.2.2 数据理解 25619.2.3 数据准备 25619.2.4 模型搭建 25619.2.5 模型评估 25619.2.6 模型发布 25619.3 Tom Khabaza的挖掘9律 256

<<数据挖掘与数据化运营实战>>

编辑推荐

《数据挖掘与数据化运营实战:思路、方法、技巧与应用》编辑推荐：阿里巴巴资深数据分析专家卢辉撰写，多年数据挖掘应用实践的经验结晶。实战性强，从数据分析师的角度对商业实战进行了总结和归纳，以大量事实和案例展现了“以业务为核心，以思路为重点，以挖掘技术为辅佐”的数据挖掘商业实践。

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:<http://www.tushu007.com>