



高等学校经济与管理类系列教材



商业分析

——商业数据的分析、挖掘和应用

王仁武 蔚海燕 范并思 ◇ 编 著



华东师范大学出版社



高校教材

高等学校经济与管理类系列教材



商业分析

——商业数据的分析、挖掘和应用

王仁武 蔚海燕 范并思 ◇ 编 著



华东师范大学出版社

图书在版编目(CIP)数据

商业分析:商业数据的分析、挖掘和应用/王仁武,蔚海燕,范并思编著. —上海:华东师范大学出版社,2014.6
ISBN 978-7-5675-2209-1

I. ①商… II. ①王…②蔚…③范… III. ①商业信息学—高等学校—教材 IV. ①F713.51

中国版本图书馆 CIP 数据核字(2014)第 141327 号

商业分析

——商业数据的分析、挖掘和应用

编 著 王仁武 蔚海燕 范并思
项目编辑 孙小帆
审读编辑 陈 震
装帧设计 孔薇薇

出版发行 华东师范大学出版社
社 址 上海市中山北路 3663 号 邮编 200062
网 址 www.ecnupress.com.cn
电 话 021-60821666 行政传真 021-62572105
客服电话 021-62865537 门市(邮购)电话 021-62869887
地 址 上海市中山北路 3663 号华东师范大学校内先锋路口
网 店 <http://hdsdcbs.tmall.com>

印 刷 者 上海丽佳制版印刷有限公司
开 本 787×1092 16 开
印 张 22.5
字 数 500 千字
版 次 2014 年 9 月第 1 版
印 次 2014 年 9 月第 1 次
书 号 ISBN 978-7-5675-2209-1/F·282
定 价 45.00 元

出 版 人 王 焰

(如发现本版图书有印订质量问题,请寄回本社客服中心调换或电话 021-62865537 联系)

序

企业在经营决策和商业竞争中,首先要了解、掌握相关的信息并加以适当的分析。以前的信息资料及其分析手段、分析过程都相对比较简单。在当今大数据时代的背景下,面对海量的数据和信息,分析技术就显得尤为重要。简单地来说,对商业大数据以及由数据刻画的各类商业信息的归集、甄别和探寻其中的某些规律性现象进而发现商业上的可用性就是所谓的商业分析。

目前,人们对于商业分析谈论得很多,但国内许多这方面的谈论还停留在或散见于各种媒体上,系统性深入阐述的并不多。这本《商业分析》是作者在该前沿领域努力探索的结晶,相信它的出版会对学术界商业分析的发展起到一定的推动作用。

商业分析首先是一种数据分析,一种信息分析。在解放战争时期的一场战斗中,我军指挥员从战报中发现,缴获的敌人的短枪与长枪的比率较往常战斗偏高,缴获的小车与大车的比率也较往常战斗偏高,因此判断这次面对的是敌人的指挥部及其护卫部队,于是便调集部队将之围歼。这虽是军事上的数据分析,但也显示了商业分析中最基本的某些逻辑性特征。当然,迅速发展中的商业分析决非如此简单,如此直观。而且,用于商业分析的数据还有一个甄别的问题。譬如,某人在网络视频上多次观看了电影《断背山》,但商家如果凭此数据信息向他推荐另一部同性恋题材的电影则可能是错误的,因为这个人很可能只是欣赏《断背山》的艺术性而已,并非偏好这一题材。当然,另外一个人如果多次观看电影《断背山》,也有可能他是偏好这一题材的,商家向他推荐其他同性恋题材的艺术作品是对商机的挖掘和利用。这其中的甄别、判断等需要不同信息和不同分析技术手段的相互补充。总之,商业分析是有趣的,但也是非常复杂的。

在市场竞争日益激烈的当今时代,商业分析无疑日显重要。在高等院校,商业分析的教育和科研也越来越受到重视。同以往专门介绍各类统计分析方法、数据与模型、数据挖掘等的教材相比,本书是从业务导向和需求导向来组织商业信息和商业数据的采集、存储、组织和分析过程,而不是单纯地介绍分析方法和工具,系统、清晰地剖析了商业分析的整个流程。这无论对理论研究还是实践应用都具有一定的指导意义和操作价值。

叶德磊

2014. 4. 12.

前言

大数据时代的到来,企业获取数据和信息变得容易,分析和利用难度却加大了。数据逐渐成为企业重要的数字资产,成为现代商业社会的核心竞争力。尤其是大数据的到来让企业重新审视和利用自己的数字资产,并将其作为对企业各项业务的支持。商业分析正是在此背景下受到学者和专业人士的重视。

2011年的 *IBM Tech Trends Report* 指出,商业分析是商业竞争中用于自动化流程和分析不断增长的海量数据中最常用的新技术;2011年的 *Bloomberg Businessweek* 在对商业分析现状调查中发现,在收入超过10亿美元的大公司中,97%都在使用某种形式的商业分析;贝恩咨询公司的《管理工具与趋势》(2013)报告将大数据和分析评为企业最流行的25种管理工具之一;2011年的麦肯锡全球机构报告预测,到2018年,仅仅美国一个国家将会缺乏具备高级分析技能的专业人才约14万至19万,缺乏懂得通过分析大数据来作出有效决策的有数据悟性的人才约1500万。

全球的多个研究报告均指出商业分析对企业未来业务发展、企业决策的重要性,也显示出企业对商业分析人才以及教育的需求。本书正是在这样的背景下产生的,我们力图去剖析商业分析的流程,搭建商业分析的框架,展现商业分析的各个步骤,最终将重点放在商业数据的分析方法和数据挖掘、数据可视化的原理和应用上。同时,注重运用大量案例来阐述问题。

本书可作为商业分析相关专业或课程的教材,也可作为信息管理与信息系统、企业管理等专业的本科生、研究生的教材或参考用书。书中加※号的内容,主要供研究生学习使用,也可供有学习潜力的本科生使用。教材中使用的数据和代码可从出版社网站上下载。

全书由范并思老师提出大纲,大家分头编写,最后由王仁武、蔚海燕来统稿。各章节的编写分工如下:第1、3、4、6章由范并思和蔚海燕撰写,第2、5章由阮光册撰写,第7、8、9章由王仁武撰写。此外,多名研究生也参与了资料的收集和整理,他们是边文冰、曹芸、常晓立、郭恒志、郭玉冰、贾喆、姜雯、梁婕、刘甜、罗银洱、粘祎行、任英剑、山珊、孙智冉、徐维娟、叶美麟、袁旭萍、翟伯荫、张晓婕、甄瑞辰、郑丽娟(学生姓名按姓氏字母排序),谢谢你们的帮助与支持!

本书的出版得到了华东师范大学研究生院重点课程项目“商业分析”的支持,也得到了华东师范大学出版社的大力支持,在此一并表示感谢!

本书在写作过程中参考了大量的国内外文献,谨对有关作者表示衷心的感谢。书中难免有疏漏错误之处,敬请读者批评指正。

编者

2014年4月

目录

第 1 章 商业分析概论	1
本章导学	3
1.1 商业分析概述	4
1.2 商业分析的发展历程	8
1.3 商业分析的基本思想	11
1.4 商业分析对企业的影响	13
习题	16
参考文献	16
第 2 章 商业分析的流程、方法与工具	17
本章导学	19
2.1 商业分析的流程与步骤	20
2.2 商业分析的主要方法	28
2.3 商业分析的主要工具	36
习题	38
参考文献	39
第 3 章 商业问题的分析与建模	41
本章导学	43
3.1 商业分析的框架	44
3.2 商业数据分析的基本步骤	47
3.3 商业数据建模	49
3.4 典型商业问题分析	53
习题	58
参考文献	58
第 4 章 商业信息的采集	59
本章导学	61
4.1 商业信息的类型	62

4.2 商业信息源	65
4.3 商业信息的采集方法	67
4.4 商业信息的采集途径与工具	69
习题	72
参考文献	72

第5章 商业信息的管理 73

本章导学	75
5.1 数据库与数据仓库	76
5.2 数据仓库设计	81
5.3 数据集市	87
5.4 商业信息仓储设计	91
5.5 商业智能	93
习题	97
参考文献	97

第6章 商业数据分析方法 99

本章导学	101
6.1 回归分析法	103
6.2 Logistic 回归	113
6.3 时间序列分析	120
6.4 交叉影响分析法	131
习题	137
参考文献	137

第7章 商业数据挖掘方法 139

本章导学	141
7.1 数据挖掘	142
7.2 决策树	146
7.3 关联规则	159
7.4 聚类分析	174
7.5 高级应用	188
7.6 小结	207
习题	208
参考文献	209

第 8 章 商业中的非结构化信息分析 **211**

本章导学	213
8.1 内容分析法	214
8.2 文本挖掘	219
8.3 社会网络分析	242
8.4 多维、异构信息分析	254
8.5 小结	267
习题	268
参考文献	268

第 9 章 商业信息可视化 **271**

本章导学	273
9.1 概述	274
9.2 基本信息可视化与应用	274
9.3 多维信息的可视化分析	324
9.4 文本与文档可视化	332
习题	338
参考文献	338

附录 **341**

1. Clementine 概述	343
2. R 语言概述	345



第 1 章 商业分析概论

本章导学

互联网时代,信息变得唾手可得。管理信息系统的广泛使用也使企业积累了大量的数据。如何管理和进一步利用这些信息和数据来支持企业的决策成为企业的迫切需求,甚至成为企业的核心竞争力所在。正如 Davenport 在其著作 *Competing on Analytics* 中将分析竞争力定义为企业赢之道。企业的决策需要哪些情报支持,这些情报来源于企业内外部的哪些数据和信息,而这些数据和信息又以怎样的分析方法、工具来成为决策支持情报,是研究人员和分析专业人员面临的重要问题。大数据时代的到来,其 4V 特征——规模性、高速性、多元性和价值性更加凸显了商业数据的价值,不仅可为企业决策作支持,甚至可以引导企业的业务发展。应运而生的商业分析(business analytics, BA)将企业的业务、分析、IT 联系起来,服务于企业的战略和决策。

本章主要内容

- 商业分析概述
- 商业分析的发展历程
- 商业分析的基本思想
- 商业分析对企业的影响

1.1 商业分析概述

1.1.1 商业分析的定义

商务分析(business analytics, BA)这一术语发源于19世纪,但其在近年来才真正引起关注。一方面是因为明智的商业决策通常建立在数据分析上,当今的数据量正以飞快的速度增长,相关的信息已经可以纳入企业决策考量,并以原始数据的形式存储下来。具有非凡价值的数据分析,已经由曾经只能被政府和大型企业所掌握,到如今各个企业都能够负担和获得,大数据环境是商业分析发展的温床;另一方面,当今相互关联的全球数据在云计算环境下存储和利用,先进的计算机软硬件也能够让各个企业以惊人的速度进行数据处理和分析,数据分析、预测、挖掘的算法、工具的产生和应用推动了商业分析的落地。

传统的商业分析是指以统计学以及概率论的内容为基础,研究、分析、处理当前经济管理领域中的一些实际问题,这些问题涉及统计报表分析、财务预测、投资风险预测、市场预测、质量管理的评估和投资效益的分析与决策等。传统商业分析的目的是通过分析数据,反映市场对企业产品的需求、资源的成本和需求情况、市场条件以及整体经济运行情况,达到强化企业经营管理、不断提高企业经济效益的效果。

传统商业分析在上个世纪60年代决策支持系统产生后,开始获得更多的关注,并开始和ERP(企业资源计划)系统、数据仓库、数据挖掘、商业智能以及其他的硬件工具和应用一起发展。而随着数据量的加大,尤其是大数据的出现,使商业分析也更多地与商业智能(business intelligence, BI)联系在一起,如Chiang(2012)、Hsinchun Chen(2012)等在讨论商业分析的教育时,将商业智能与分析连在一起称之为BI&A。

Shmueli G.、Koppius O.(2007)认为商业分析是使用大量的数据、统计和定量分析、解释和预测模型,并基于事实的管理方法来推动决策,其中涉及一系列的技能、技术、应用和实践^①。在*Competing on Analytics*中,Davenport与Harris把商业分析形容为“数据的延伸引用、统计与量化分析、说明性与预测性的模型以及驱动决策与实施的基于事实的管理”^②。美国运筹学与管理学研究协会(INFORMS)认为:商业分析通过数据分析趋势、构建预测模型、优化企业流程,提升企业绩效,以促进商业目的的实现。开设了商业分析专业的各大高校也分别对商业分析进行了定义。伊利诺伊大学芝加哥分校认为商业分析是不断地探索、调研商业业绩数据的过程。探索的过程中要使用相关的技能、技术、应用和经验,会产生对现在和未来商业计划的一些新的定性或定量的见解。商业分析关注统计模型和大数据,因为它们会被应用于包括医疗、汽车和体育等每一个行业的分析^③。宾夕

① Shmueli G, Koppius O. Predictive vs. explanatory modeling in IS research [J]. Joint work, University of Maryland, College Park, and Rotterdam School of Management, Erasmus University. <http://www.citi.uconn.edu/cist07/5c.pdf>, 2007.

② Davenport, Thomas H.; Harris, Jeanne G. (2007). *Competing on Analytics: The New Science of Winning*. Boston, Mass.; Harvard Business School Press. ISBN 978-1-4221-0332-6.

③ UIC Business. Business Analytics. [EB/OL]. <http://business.uic.edu/home-uic-business/faculty-departments-research/department-of-ids/programs/business-analytics>. [2014-3-25].

法尼亚大学对商业分析的定义是,商业分析是谨慎地用分析技术来解决业务问题。这些分析技术来自于多个学科领域,如运筹学、统计学、机器学习、数据挖掘、信息科学、计算机科学、市场营销和博弈论。商业分析不是一项独特的管理功能,如财务或者会计;相反地,它被广泛认为是一项技能的集合,可以为每个管理过程增加价值^①。

美国运筹学与管理学研究协会(INFORMS)提出商业分析有以下三种类型:

一是描述性分析:描述分析以前发生了什么。二是预测性分析:预测将来会发生什么。三是规范性分析:描述将产生最佳结果的最佳方法。

综合上述概念,本书认为商业分析就是在理解业务问题的基础上,利用IT技术进行数据的处理和呈现,应用各种分析模型进行数据分析和挖掘,描述业务现状,预测业务未来,提出业务解决方法。

商业分析的分析对象很宽泛,不仅仅包含数据库和数据仓库中的结构化的数据,文本信息、网络信息等非结构化数据也是其重要对象,且分析的方法手段也不仅仅局限在数据分析和挖掘的方法,还包括对文本、音频、视频等信息的分析方法。在商业分析的发展过程中,商业智能和其息息相关。

1.1.2 商业分析与商业智能

1. 商业智能的定义

商业智能(business intelligence, BI)的概念最早是 Gartner Group 于 1996 年提出的。当时将商业智能定义为一类由数据仓库或数据集市、查询报表、数据分析、数据挖掘、数据备份等部分组成的,以帮助企业决策为目的的技术及其应用。商业智能是对商业信息的搜集、管理和分析过程,目的是使企业的各级决策者获得知识或洞察力,促使他们做出对企业更有利的决策。从系统的观点来看,商业智能的过程首先是从不同的数据源收集的数据中提取有用的数据,对数据进行清理以保证数据的正确性,将数据经转换、重构后存入数据仓库(这时数据变为信息),然后利用合适的查询和分析工具,如 OLAP 工具对信息进行处理(这时信息变为辅助决策的知识),最后将知识呈现于用户面前,来支持决策。可以看出,企业信息化是商业智能应用的基础,商业智能最大限度地利用了企业信息化中的数据,将企业数据整理为信息,再升华为知识,从而对用户 provide 最大程度的支持。

2. 商业分析与商业智能的区别

虽然商业分析与商业智能的关系密切,且很多学者也经常将这两个概念连在一起使用,但两者之间还是存在着区别。

(1) BA 和 BI 分析对象不同

商业智能的分析和挖掘对象为结构化的数据,采用数据挖掘的一些算法和软件等挖掘出数据中存在的隐性规则。商业分析的内容更宽泛,不仅面对的分析对象要宽泛,突破

^① Business Analytics: A practioner's Guide. [EB/OL]. <https://opimweb.wharton.upenn.edu/programs/undergraduate/program-information/business-analytics-track/>. [2014-3-25].

了结构化的数据,非结构化数据也是其重要的分析对象,且分析的方法手段也不仅仅局限在数据分析和挖掘的方法,包括对文本、视频等信息的分析与挖掘。

(2) BI 侧重数据,BA 与业务结合更紧密

商业智能偏重于衡量业务结果,而商业分析更加偏重于业务流程中的分析,借助于预测性分析工具和软件,如时间序列和 SAS 预测软件等为管理者和员工团队提供更具有指导性意义的有效信息,帮助企业更好地完成分析和决策,全面提高企业的绩效。

随着数据爆炸式增长,BI 无法处理日益复杂的数据,将 BI 扩展至 BA 将成为未来竞争的焦点。BI 和 BA 虽然只有一字之差,但其内涵却大不相同。BI 只是从数据出发,侧重技术挖掘的过程,而 BA 则是以业务为出发点,提出有针对性的业务问题,使 BI 平台中的数据价值得到充分利用,也让 BI 更深入业务。从 BI 到 BA,是用户的强烈需求,也是商业智能发展的必然。SAS 公司大中华区总裁黄永恒谈到,以前 BI 的概念有些狭隘,是从数据出发,侧重技术挖掘过程,而与业务脱节,往往得出的数据和结论并不是业务真正需要的。

从这个角度看,BA 能够通过预测性分析为管理者和员工团队提供所需要信息,帮助他们更好地完成决策和分析,更能帮助企业解决实际业务问题,也能更好地发挥 BI 的商业价值。

(3) BI 是 BA 的一部分

商业分析是相对于商业智能更高层次的概念,注重在分析过程中对不同的业务运用不同的分析、预测模型,形成有利于业务的决策方案。

商业分析主要由数据仓库平台和绩效管理工具构成。数据仓库平台包含数据仓库管理工具和数据质量工具等。绩效管理工具包括基于数据仓库或统一数据平台之上的分析功能,如财务绩效和战略管理、供应链分析、劳动力分析、客户关系分析等^①。商业分析之所以比商业智能更加高端,是因为它需要有分析应用的解决方案来支撑。而分析应用里面最关键的有两个部分:一个部分是分析,另一个部分是智能,即 BI,这两者同时支持分析应用。

从本质上看,商业智能更注重工具的应用,使用统一指标来衡量企业绩效或分析报表。而商业分析不仅包括统计、因果和预测模型等要素,而且其应用还可囊括所有的简单报表和复杂的系统。

如此可以看出,BI 只是 BA 的一部分。

(4) BA 和 BI 追求的目标不同

商业智能是基于关键绩效指标(KPI)来确定企业的追求目标^②。KPI 指标通常可以继续分解成多个衡量指标,即多个事实(fact)。这些事实可以直接从业务系统中获得,也可以通过累加或指标聚合而得出。因而在商业智能中,数据仓库的粒度设计非常关键,这样才能保证商业智能系统对 KPI 指标的支持。

^① 陈雨田. 商业分析:挖掘数据的最佳工具[N]. 中国城乡金融报,2011-7-28(A03).

^② Robert Laberge 著,祝洪凯等译. 数据仓库应用指南:数据仓库与商务智能最佳实践[M]. 北京:机械工业出版社,2012.

商业分析的目标是通过信息分析、数据挖掘与业务结合起来,为企业提供最优化的问题解决方案;通过对信息的运用和分析,充分挖掘其中的价值,以提高决策的科学性和准确性、发现问题的及时性和快捷性。商业分析并不局限于支持企业绩效,还可以为企业带来业务流程和企业文化的变革以及新鲜的科学管理思维。

1.1.3 商业分析的三个视角^①

商业分析为企业的业务决策服务,关乎企业的价值创造。商业分析是从理解业务问题出发,从分析视角看数据,利用数据分析、挖掘模型,基于数据库、数据仓库等 IT 工具,结合分析人员的智慧来完成。

1. 业务:决策与执行视角

企业人员通常把自己看作是分析的“消费者”,期望分析师构造一种模型有助于“企业成长”或者“提升业务”;但企业决策的过程有时很难有效传达给分析师。普遍认为分析师理解商业运作的目的就是为了构建有效的模型。但如果分析师在模型中投入少甚至不投入,则会使企业怀疑分析能力所产生的价值。

如果管理人员和分析团队之间的距离较大,商业分析只能产生基本的报表,这些基本的报告和图表不能满足用户的真正需求。同样情况也适用于如何使业务部门与 IT 部门合作。IT 部门常常被认为只提供系统,可靠的分析系统必须随着业务、客户、竞争对手、员工、供应商和市场的每日变化不断发展。在组织中的业务部门的人需要学习与分析师、IT 团队合作。

2. 分析:技术视角

分析师经常认为自己是“数据和数学专家”,他们构建了成熟的技术模型。如果模型支持的决策过程不能被完全理解,企业用户可能会提出“这很不错,但如何有效应用”这类疑问;另一个问题在于,由于缺乏数据(质量和数量)能够建立理想的分析模型,产生的商业决策难以达到最佳状态。

通常 IT 团队被分析团队看作分析基础设备的提供者,但分析团队与 IT 团队合作解决业务问题的情况则很少。由于这种情况,当商业用户需要商业分析时,他们会直接聘用 IT 团队,忽视分析团队。有效的三方合作消减了。分析师需要发展与业务合作同行和 IT 团队有效的合作方法。

3. IT:工具和系统的视角

IT 通常认为是提供商业智能和数据仓库(data warehouse, DW)等基础设施和工具来支持分析师和商业用户的。为了应对企业分析功能的需求开发项目,经常开发巨大的

^① Business Analytics: A practitioner's Guide. [EB/OL]. <https://opimweb.wharton.upenn.edu/programs/undergraduate/program-information/business-analytics-track/>. [2014-3-25].

数据仓库存储数据或开发多个工具,如分析报表和分析用的统计工具。IT的角色无疑是提供商业需求中的有效数据的工具。当商业管理者和分析师提出了不使用如数据仓库、报告等昂贵设备时,BI和DW就是一项失败的投资。IT需要增加与业务和分析部门的有效合作。

1.2 商业分析的发展历程

1.2.1 商业分析的历史

分析业务有着很长的历史,其根源可追溯到从企业的科学管理时代泰勒从观察到决策使用分析伊始。后来咨询公司开始作为第三方机构为企业提供分析业务的服务。商业分析师受雇来协助经理人扮演分析角色,主要负责完成分析报告。一些分析师通过工业

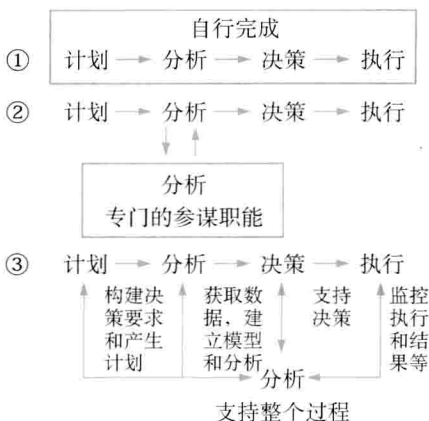


图 1-1 分析的发展历程

图 1-1①中,分析工作由决策者自己完成。这种方法使人们能够应用分析贯穿从计划到执行整个环节的始末。其间要求人们在常规工作中保持理性思考并使用合理数据,也就是说,把分析性思考看作是一种分析性活动。这种方法仍只是在小范围内适用——专业团队专注的有限结果。

图 1-1②中,分析作为一项专门的业务职能。当大多数企业想要利用分析达到企业目标时,分析团队作为如财务、运营、营销等传统员工的职能应运而生。这种方法是基于大家普遍认可的经济规模下提供的商业分析业务——“分析师”。

图 1-1③中,分析支持从计划到执行的整个过程。它集中了分析人才和工具,并将其整合成一项专业化功能,从而关联和实现企业的各种商业需求。目前分析业务领域中,分析专家的需求大幅提升,分析专家的工作正在尽可能全方位延伸以满足商业需求。

1.2.2 商业分析的研究进展

在大数据背景下,产业界对商业分析的需求超越了以往任何一个阶段。学术界也开始转