



商业分析

Business Analytics

# 商业分析概论

蔚海燕 许鑫◎编著



华东师范大学出版社

商业分析

Business Analytics

# 商业分析概论

蔚海燕 许 鑫◎编著

## 图书在版编目(CIP)数据

商业分析概论/蔚海燕,许鑫编著. —上海:华东师范大学出版社,2015.7

(商业分析丛书)

ISBN 978-7-5675-3844-3

I. ①商… II. ①蔚…②许… III. ①商业信息学  
IV. ①F713.51

中国版本图书馆CIP数据核字(2015)第161221号

## 商业分析概论

编 著 蔚海燕 许 鑫

策划组稿 孙小帆

项目编辑 孙小帆

审读编辑 丁 倩

装帧设计 卢晓红

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](http://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 苏州美柯乐制版印务有限公司

开 本 787×1092 16开

印 张 14

字 数 268千字

版 次 2015年9月第1版

印 次 2015年9月第1次

书 号 ISBN 978-7-5675-3844-3/F·333

定 价 36.00元

出 版 人 王 熠

(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)

# 本书简介

## 目标：

通过对商业分析(商业数据分析)领域相关知识点的介绍,帮助读者建立商业分析学(Business Analytics)的整体知识框架,初步了解商业分析的流程与步骤,即商业环境中的数据采集、数据处理、数据存储、数据分析环境搭建以及相关的分析方法。

## 内容组织：

本书分为基础编、数据编和方法编三部分。

基础编包括如何理解商业分析(第1章)、商业问题建模和商业数据分析(第2章),重点帮助大家了解商业分析的历史、发展、基本思想、主要流程等知识。

数据编包括商业信息采集(第3章)、数据库与数据仓库(第4章)、商业数据分析环境的搭建(第5章)、联机分析处理(第6章),在商业数据采集的概括性介绍基础上,分析比较了企业环境下数据库到数据仓库的发展,关注实际商业环境中数据分析环境的搭建问题,重点介绍了数据仓库的构建和基于数据仓库的OLAP。

方法编介绍了最主要的几类商业数据分析方法,包括常用统计分析方法(第7章)、常用数据挖掘方法(第8章)、文本挖掘(第9章)、Web挖掘(第10章)、复杂数据挖掘(第11章)和商业信息可视化方法(第12章),帮助读者了解如何在商业信息采集和分析环境构建的基础上利用合适的方法开展商业分析工作。

## 体例特点：

本书以概括性介绍为主,并未涉及太多具体的算法和实例,主要目的是构建一个相对完整的商业分析学体例。每章结尾部分均有本章小结,有助于帮助读者进一步总结和思考。

# 目录

本书简介 1

## 基础编 1

- 1 如何理解商业分析 3
  - 1.1 什么是商业分析 3
  - 1.2 商业分析的发展历程 7
  - 1.3 商业分析的基本思想 10
  - 1.4 商业分析的主要流程 15
  - 1.5 更深地理解商业分析 23
- 本章小结 28
- 2 商业问题建模和商业数据分析 30
  - 2.1 业务理解与需求定义 30
  - 2.2 数据准备与数据理解 31
  - 2.3 模型构建与模型优化 31
  - 2.4 模型评估与部署应用 33
  - 2.5 商业数据分析示例 34
- 本章小结 39

## 数据编 41

- 3 商业信息采集 43
  - 3.1 商业信息源 43

- 3.2 商业信息采集方法 50
- 3.3 商业信息常用采集途径 55
- 3.4 商业信息新兴采集工具 56
- 本章小结 59
- 4 数据库与数据仓库 61
  - 4.1 从传统数据库到数据仓库 61
  - 4.2 数据仓库的相关概念 68
  - 4.3 数据仓库的相关模型 70
  - 4.4 数据仓库的体系结构 75
  - 4.5 数据仓库与数据挖掘 76
  - 本章小结 77
- 5 商业数据分析环境的搭建 78
  - 5.1 数据仓库的数据组织 78
  - 5.2 数据仓库设计步骤 80
  - 5.3 数据集市及其应用 85
  - 5.4 商业信息仓储设计 89
  - 本章小结 92
- 6 联机分析处理(OLAP) 93
  - 6.1 OLAP 的提出 93
  - 6.2 OLAP 的基本操作 98
  - 6.3 基于 OLAP 的多维分析 99
  - 6.4 OLAP 的评价标准 106
  - 本章小结 109

## 方法编 111

- 7 常用统计分析方法 113
  - 7.1 描述性统计分析 113
  - 7.2 回归分析法 114
  - 7.3 Logistic 回归 121

- 7.4 时间序列分析 123
- 7.5 探索性分析方法 125
- 7.6 常用统计分析工具 129
- 本章小结 130
- 8 常用数据挖掘方法 131
  - 8.1 数据挖掘概述 131
  - 8.2 分类 133
  - 8.3 聚类分析 136
  - 8.4 关联规则 139
  - 8.5 统计方法与数据挖掘 141
  - 本章小结 143
- 9 文本挖掘 144
  - 9.1 文本挖掘概述 144
  - 9.2 文本的预处理 146
  - 9.3 文本分类 150
  - 9.4 文本聚类 153
  - 9.5 自动摘要 157
  - 9.6 信息过滤 160
  - 本章小结 162
- 10 Web 挖掘 163
  - 10.1 Web 挖掘概述 163
  - 10.2 数据源及任务分类 165
  - 10.3 Web 内容挖掘 168
  - 10.4 Web 结构挖掘 170
  - 10.5 Web 使用挖掘 171
  - 本章小结 173
- 11 复杂数据挖掘 174
  - 11.1 复杂数据挖掘概述 174
  - 11.2 多媒体数据的存储 176
  - 11.3 多媒体数据标引与检索 179

11.4 多媒体数据的挖掘分析 182

11.5 空间数据库挖掘 185

本章小结 194

## 12 商业信息可视化方法 195

12.1 商业信息可视化概述 195

12.2 不同分类的绘图 196

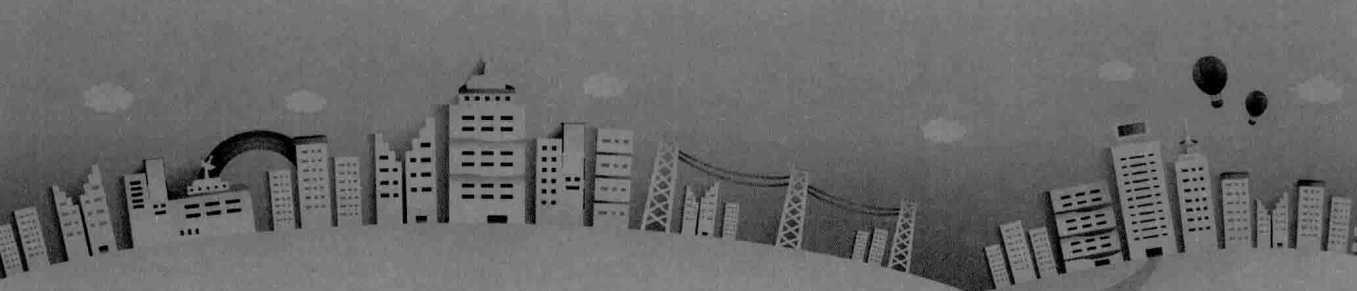
12.3 多维信息的可视化分析 203

12.4 文本与文档可视化 210

本章小结 214

参考文献 215





# 基 础 编



# 1 如何理解商业分析

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。正如达文波特(Davenport)在其著作 *Competing on Analytics* 中将分析竞争力定义为企业赢之道。企业的决策需要有力的情报支持,这些情报来源于企业内外部的信息和数据。如何收集、整理、分析信息和数据,使其成为情报,为决策提供支持,成为研究人员和专业的分析人员面临的研究问题。过去的 40 多年里,分析和理解数据意义的活动也从决策支持发展到联机分析处理和商业智能等多个领域。尤其是大数据时代的到来,更增加了这一问题的重要性与紧迫性。商业分析(Business Analytics,简称 BA)成为将商业数据与信息转化为商业情报的重要环节,它将企业的业务、分析、IT 联系起来,服务于企业的战略和决策。

## 1.1 什么是商业分析

### 1.1.1 商业分析的概念

商业分析这一术语发源于 19 世纪,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,当今的数据量飞速增长,相关的信息已经可以被纳入企业决策考量,并以原始数据的形式存储下来。具有非凡价值意义的数据分析,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获得和负担;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分析。这样的高速处理为预测数据结果或发现数据模式的计算机算法的产生和应用奠定了基础。大数据环境是商业分析发展的温床。

传统的商业分析是指以统计学及其有关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、销售预测、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情况、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 20 世纪 60 年代计算机应用于决策支持系统后,开始获得更多的关注。自从决策支持系统出现后,商业分析开始和 ERP(企业资源管理计划)系统、数据仓库、数据挖掘、商业智能以及其他软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,商业分析更多地与商业智能联系在一起,有学者在其讨论商业分析的文章中,将商业智能与分析连在一起称为 BI&A(Business Intelligence and Analytics)。达文波特等将商业数据分析的相关术语的发展历程做了总结,如表 1-1 所示:

表 1-1 使用和分析数据的术语

术 语	时 间	具 体 意 义
决策支持	1970—1985	使用数据分析来支持决策的制定
高管支持	1980—1990	专指为高级管理人员的决策做数据分析
联机分析处理	1990—2000	用软件分析多维数据表
商业智能	1989—2005	利用工具来支持由数据驱动的决策,并强调报表
分 析	2005—2010	专注于用统计和数学分析支持决策
大数据	2010—现在	专注于海量的、非结构化的、快速变化的数据

G. Shmueli 和 O. Koppius 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。在 *Competing on Analytics* 这本书中,Davenport 与 Harris(2007)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”。美国运筹学与管理学研究协会(The Institute for Operations Research and the Management Science,简称 INFORMS)将商业分析定义为: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都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,因为他们会被应用于包括医疗、汽车和体育等每一个行业的分析。宾夕法尼亚大学对商业分析的定义是,商业分析即谨慎地用分析技术来解决业务问题。这些分析技术来自多个学科,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。美国运筹学与管理学研究协会提出商业分析有以下三种类型:① 描述性分析:描述分析以前发生了什么。② 预测性分析:预测将来会发生什么。③ 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围绕企业的业务需求,运用定性或定量的分析模型将信息、数据转化为有用的信息来帮助做出更好的商业决策。在商业分析的过程中需要运用现代信息分析方法与计算机技术,来对企业数据进行整合、分析,描述企业的现状,预测未来趋势和状态,制订企业解决方案。

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。在商业分析的发展历程中,商业智能和其息息相关。

### 1.1.2 商业分析与商业智能的区别

商业智能(Business Intelligence,简称 BI)的概念最早是 Gartner Group 于 1996 年提出来的。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,商业智能应用的过程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),然后寻找合适的查询和分析工具如联机分析处理(Online Analytical Processing,简称 OLAP)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,转变为决策。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 最大限度的支持。

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) 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。

商业分析系统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工具,如财务绩效和战略管理工具、供应链分析工具、劳动力分析工具、客户关系分析工具等。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。

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,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。

#### (2) BI 侧重数据,BA 与业务结合更紧密

商业智能侧重衡量业务结果,而商业分析更加侧重业务流程的分析。借助预测性分析工具和软件如时间序列和 SAS 预测软件等,商业分析可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,帮助他们更好完成决策和分析,更能帮助企业解决实质业务问题,也能更好地发挥 BI 的商业价值。

#### (3) BA 和 BI 追求的目标不同

商业智能基于关键绩效指标(KPI)来确定企业的追求目标。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。

商业分析的目标是将信息分析、数据挖掘与业务结合起来,为企业提供最佳化的问题解决方案;通过对信息的运用和分析,充分挖掘其中的价值,为企业带来科学管理思维,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,带来新鲜的科学管理思维。

#### (4) 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更宽泛,突破了结构化的数据,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、视频等信息的分析。

## 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,其根源可追溯到企业的科学管理时代,即始于泰勒从观察到决策中使用分析。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者的角色,主要负责完成分析报告。一些分析师通过工业工程和质量控制的工具和技术、统计和运筹学,为企业提供服务。而企业的 IT 团队看到机会,也希望为管理者提供报告,由此管理信息系统(MIS)的概念产生了。这些系统的应用使得 IT 团队以报告和图表的形式为组织机构提供分析业务,目的是将正确的信息在合适的时间提供给正确的客户。在 IT 部门中的商务智能团队和数据仓库(Data Warehouse,简称 DW)延续了这一传统。计划,决策,指导,激励,监督和控制是管理者的职责。图 1-1 展示了管理中从计划到执行的理性决策过程,以及分析在其中起的不同作用。

图 1-1 中,① 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——专业团队专注有限的结果。② 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。③ 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,关联和实现企业的各种商业需求。目前分析业务领域分析专家的需求大幅提升,分析专家的工作尽可能全方位延伸以满足商业需求。

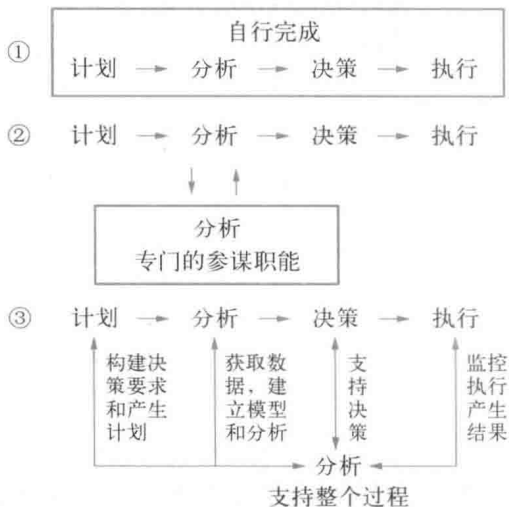


图 1-1 分析的发展历程

### 1.2.2 商业分析的发展演化

2011 年的 IBM 科技趋势(发展)报告认为,在 2010 年代商业分析将是四大主要科技趋势之

一;2011年的彭博商业周刊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;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大约14万至19万的深入分析技能人才以及1500万的懂得如何分析大数据来做出有效决策的有数据悟性的人才。

Chen Hsinchun等提出了BI&A从1.0、2.0到3.0的演化过程。具体见图1-2。

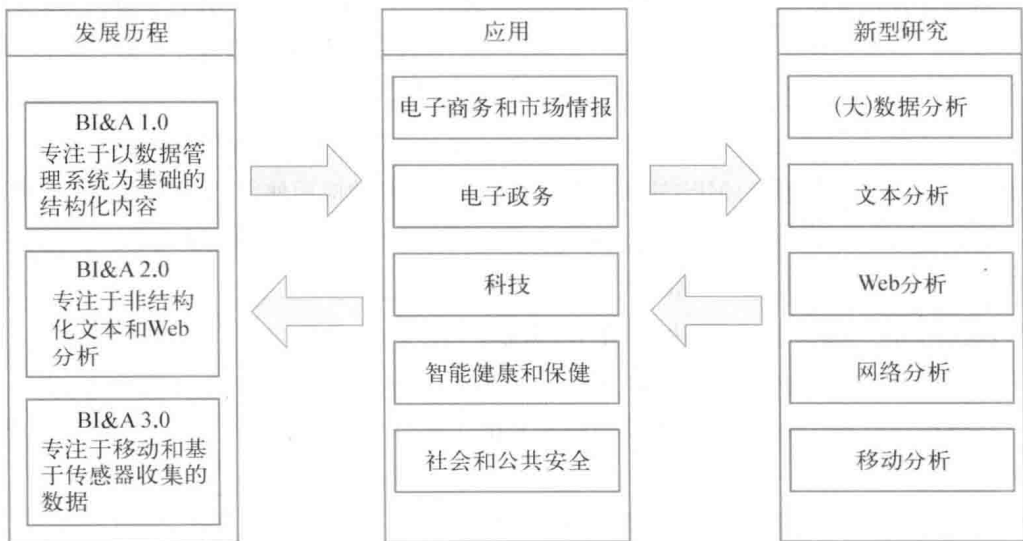


图1-2 BI&A的演化过程

同时,该研究中还进一步分析了BI&A三个阶段的特征:

#### (1) BI&A 1.0

现在正被行业所采用的商业智能与分析技术及应用被认为是BI&A 1.0版。此时,大部分的数据是结构化的,由公司多种传统系统收集的,而且通常被储存在与商业相关的数据管理系统(RDBMS)中。数据管理和数据仓库被认为是商业智能与分析1.0的基础。数据集市的设计和抽取、数据转换及数据加载(ETL)的工具对于企业特定领域数据的转变和集成是非常必要的。数据库查询,联机分析处理(OLAP),基于直觉而又简单的图表报告工具被用来探索重要的数据特征。企业绩效管理使用记分卡和仪表盘帮助分析和可视化多种多样的绩效考量。除了这些稳定使用的商业报告功能,统计分析和数据挖掘技术也在多种多样的商业应用程序中被用于关联分析、数据分割、聚类、分类、回归分析、异常检测和预测模型。这些数据处理和分析技术中的大多数已经被融合进先进的商业智能平台。这些平台是由包括微软、IBM、Oracle和SAP在内的主要的IT供应商提供的。



根据 Sallam 等人在 2011 年的 Gartner report 中所说,在商业智能平台所必需的 13 种功能中共有 9 项被认为是商业智能与分析 1.0 的功能: 报表、仪表、即席查询、基于检索的商业智能、OLAP、交互型的可视化、记分卡、预测模型和数据挖掘。

### (2) BI&A 2.0

BI&A 2.0 领域专注于非结构化文本和 Web 分析。大量的公司、行业、产品和顾客信息都可以通过网络收集,并通过多种文本和 Web 挖掘进行组织与可视化。诸如 Google Analytics 这样的 Web 分析工具就可以提供用户在线行为踪迹和揭示用户浏览和购买模型。网站设计、产品摆放操作、顾客交易分析、市场结构分析和产品推荐都可以通过 Web 分析来完成。2004 年以后开发出来的许多 Web 2.0 应用程序,已经通过多种在线社交媒体比如论坛、在线群体、网页日志、社交网站、社交多媒体网站(照片和视频),甚至是虚拟世界和社交游戏,创造出来大量的用户产生内容。除了抓取这些媒体中的名人每日发布的信息、日常事件的信息和社会政治情绪,Web 2.0 应用程序能够非常有效地从不同种类业务的不同客户人群中获得大量的及时反馈和意见。未来的 BI&A 2.0 系统将需要对文本挖掘(比如信息抽取、主题识别、观点挖掘、问答)、Web 挖掘、社交网络分析和现在已经存在的基于数据库管理系统的商业智能与分析 1.0 系统的时空分析方面的成熟而又可扩展技术等各方面进行集成。

### (3) BI&A 3.0

随着拥有移动手机和平板电脑的人群数量超过了拥有手提电脑和 PC 机的人群数量,人们正在针对利用 RFID、条形码、无线标签技术的基于传感器可联网的设备开发一些激动人心的具有创新性的应用程序。像这样的移动和可联网设备能够支持高度移动、地理定位和以人为中心、文本相关操作以及交易的能力将在 2010 年代继续带来独特的研究挑战和机遇。移动接口、可视化和人机交互设计也都是很有希望的研究领域。尽管 Web 3.0(移动和基于传感器)时代肯定会到来,但用于数据收集、处理、分析和可视化如此大规模和流动的移动和传感器数据的底层移动分析、定位和文本感知技术仍未成形。因为 BI 2.0 和 3.0 的方法和技术仍处在探索和研究阶段,还不够成熟,因此本书的内容架构上仍然以 BI 1.0 的内容为主。

## 1.2.3 大数据时代的商业分析特征

大数据时代的商业分析与传统的商业信息管理与分析有着显著的不同。传统的商业信息管理与分析产生的研究报告主要是在内部决策中给予高级管理人员建议,而大数据科学家需要普遍地面向顾客提供产品和服务。

大数据时代的商业分析在支持决策的时间上有了更多的要求,如要求实时分析或接近实时