



普通高等教育“十一五”国家级规划教材

数字经济高质量人才培养系列 · 大数据管理

商务智能

原理与方法

/ 第3版 /

陈国青 卫强 张瑾◎编著

电子工业出版社

Publishing House of Electronics Industry

北京 · BEIJING

内 容 简 介

处在大数据时代发展的重要节点上,面对数字化生活的新需求、数字化转型的新格局及数字经济的新业态,商务智能正在扮演着越来越重要的角色。商务智能通过数据挖掘和机器学习技术从海量多模态数据中发现潜在、新颖和有用的知识,以支持管理与决策。它是内在技术的,同时也是面向管理决策问题的。

本书旨在把握前沿趋势,以基础篇、方法篇、专题篇三大板块的形式,为读者提供一个技术与管理的融合视角,介绍和阐释商务智能领域的主要知识内涵,包括面向管理决策的商务智能基本原理、主流方法、应用情境和发展前景,帮助读者理解如何通过商务智能进行大数据/人工智能分析和赋能,从而提升组织和个体的核心能力及其竞争优势。

本书既可以作为高等学校信息管理与信息系统、大数据管理与应用、电子商务、管理科学与工程以及工商管理、计算机应用等相关学科专业的高年级本科生和研究生的专业课教材,也可以作为财经类或其他工程类专业学生的专业课/选修课教材,还可以为社会各领域信息化培训和相关管理决策人员提供参考。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究。

图书在版编目(CIP)数据

商务智能原理与方法 / 陈国青, 卫强, 张瑾编著. —3 版. —北京: 电子工业出版社, 2023.4

ISBN 978-7-121-45388-5

I. ①商… II. ①陈… ②卫… ③张… III. ①电子商务—高等学校—教材 IV. ①F713.36

中国国家版本馆 CIP 数据核字(2023)第 062479 号

责任编辑: 章海涛

特约编辑: 刘宪兰

印 刷:

装 订:

出版发行: 电子工业出版社

北京市海淀区万寿路 173 信箱 邮编: 100036

开 本: 787×1092 1/16 印张: 23.75 字数: 600 千字

版 次: 2009 年 9 月第 1 版

2023 年 4 月第 3 版

印 次: 2023 年 4 月第 1 次印刷

定 价: 78.00 元

凡所购买电子工业出版社图书有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系,联系及邮购电话:(010) 88254888, 88258888。

质量投诉请发邮件至 zltz@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式: 192910558 (QQ 群)。

前言

不断动态涌现、交融渗透的技术革命与管理变革,给我国经济社会生活持续提供了难得的动力、多样性和探索空间,同时带来了知识的更迭、复杂性和诸多挑战。处在大数据时代发展的重要节点上,面对数字化生活的新需求、数字化转型的新格局,以及数字经济中的新业态,本书旨在把握前沿趋势,为读者提供一个技术与管理的融合视角,介绍和阐释商务智能领域的主要知识内涵,包括面向管理决策的商务智能基本原理、主流方法、应用情境和发展前景。

作为一个相对新兴的领域,商务智能(Business Intelligence, BI)在近些年来得到了飞速发展。首先,商务智能是不断拓新的理论、方法、技术发展的产物,是信息科学、管理学、决策科学等学科的交叉领域。其次,商务智能通过数据挖掘和机器学习技术从海量多模态数据中发现潜在、新颖和有用的知识,以支持组织的管理与决策。它是内在技术的,同时也是面向管理决策问题的。最后,商务智能包括一系列高级商务分析(Business Analytics, BA)方法,具有广阔的应用前景。特别值得一提的是,商务智能的应用既可以建立在组织中的全面信息化实施框架基础之上,也可以仅针对具体业务问题在局部业务单元和环节上展开。此外,在大数据环境中,内外数据融合成为商务分析的基本特征。也就是说,商务智能的视角也从传统的组织内部拓展到组织外部,并更具有社会化、移动化、智能化的特点。

由于商务智能已经形成了比较成熟的知识体系框架,而在知识内涵上又与信息技术革新和管理应用需求保持联动,所以从数字化演进的角度出发,商务智能具有鲜明的动态性和时代感。鉴于此,开展商务智能知识教育和人才培养是我国现代化建设乃至数字经济发展的需要,也是我国高校相关学科(如管理科学与工程、工商管理等)和相关专业(如信息管理与信息系统、大数据管理与应用、电子商务、计算金融、管理工程等)建设的重要内容。

本书仍然采用与前两版(《商务智能原理与方法》,电子工业出版社,2009年、2014年版)一致的主体构架,包括基础篇、方法篇、专题篇三大知识板块。整体上沿着商务智能应用、方法和发展的总脉络,试图较全面地涵盖商务智能领域的基础知识和主流内容,同时融入若干前沿成果以体现新兴特征与前瞻性,如机器学习、大数据、人工智能等方面的新内容。

基础篇着重介绍商务智能的概念和用途、应用领域和环境,以及对于组织管理决策的影响。其中更新了各章相关内容并注重体现大数据特征、用户生成内容(User Generated Content, UGC)、移动商务和社会商务应用等知识点,同时新增了企业知识

发现、数据平台、新型商务智能系统等方面的内容。

方法篇详细讨论商务智能领域和业务管理决策中的重要机器学习和商务分析方法。其中更新了原有关联、分类、聚类、社会网络分析等主要知识形式和挖掘方法的相关内容,注重体现大数据、人工智能等知识点,同时新增了概率图模型、神经元网络、多模态数据表征等内容。

专题篇重点关注商务智能领域的若干课题。其中更新了关联分类、不确定性知识发现、推荐方法、商务智能的经济社会影响等相关内容,新增了信息提取与洞察(如“大数据—小数据”问题)、智能推荐(如多模态信息推荐、序列推荐、捆绑推荐)、人工智能赋能、数据治理等内容。

本书既可以作为高等学校信息管理与信息系统、大数据管理与应用、电子商务、管理科学与工程以及工商管理、计算机应用等相关学科专业的高年级本科生和研究生的专业课教材,也可以作为财经类或其他工程类专业学生的专业课/选修课教材。教学中可以单独或组合性地使用本书知识板块的内容。此外,本书内容可以为社会各领域信息化培训和相关管理决策人员提供参考。

本书内容反映了我们多年在教学实践和素材上的积累,也部分反映了我们相关研究团队的若干成果。我们衷心感谢清华大学经济管理学院(暨电子商务实验室、人工智能管理研究中心、中国零售研究中心)、清华大学现代管理研究中心、中国人民大学商学院对于相关研究和本书编写提供的工作条件,衷心感谢教育部人文社会科学重点研究基地项目的资助(17JJD630006)和国家自然科学基金项目的资助(72172070、72072177、71772177)。本书从第1版的立意、写作到完稿及之后的各版修编过程都得到了许多专家学者的关注和支持,包括国家信息化专家咨询委员会和信息化与信息社会丛书编委会的曲维枝主任、周宏仁副主任等及大连理工大学的王众托院士、哈尔滨工业大学的李一军教授等专家学者的意见和建议,对此我们深表谢意。除了本书编著者参与的研究工作以及负责本书各版的整体框架和知识模块设计、各章修编和统纂,感谢本书作者所在研究团队其他部分成员在相关领域的学术贡献以及本书各版编写团队的高水平工作。特别地,在本书第3版修订过程中,清华大学博士生金恩泽、田智航、王涵菲、庄园等参与了素材整理及部分初编工作,做出了辛勤努力和出色贡献。同时,我们由衷感谢电子工业出版社及刘宪兰编辑在本书策划、协调、修编和出版发行中所做的大量杰出工作。

由于时间和水平有限,本书难免存在不足之处,恳请广大读者批评指正。

陈国青 卫强 张瑾

2022年8月于北京

目 录

基 础 篇

第 1 章 引言	003
1.1 商务智能简介	004
1.2 商务智能与信息社会	007
1.2.1 信息技术提升信息社会发展水平	007
1.2.2 商务智能是信息社会的产物	010
1.2.3 商务智能是信息社会繁荣的推动力	011
1.3 商务智能与企业管理	013
1.3.1 商务智能在企业管理中的作用	013
1.3.2 商务智能协助企业管理的方式	014
1.3.3 商务智能的商业价值	015
1.4 商务智能的方法	016
1.5 商务智能的数据	019
小结	020
思考与练习	021
第 2 章 商务智能应用	022
2.1 制造领域应用	023
2.2 金融领域应用	026
2.3 通信领域应用	029
2.4 生物和医药领域应用	031
2.5 零售和营销领域应用	033
2.6 移动商务应用	036
2.7 社会化商务应用	037
小结	039
思考与练习	040
第 3 章 商务智能过程	041
3.1 数据库与事务处理	043
3.1.1 数据库与数据库管理系统	043
3.1.2 在线事务处理	044
3.2 数据仓库与在线分析处理	045
3.2.1 从事务处理到分析处理	045
3.2.2 数据仓库	047

3.3	企业知识发现	050
3.3.1	OLAP 与知识发现	051
3.3.2	企业内部知识发现	052
3.3.3	企业外部知识发现	052
小结		054
思考与练习		054
第 4 章	数据平台	055
4.1	数据处理技术演进	056
4.2	数据仓库过程与体系结构	058
4.3	数据集成、提取与转换	060
4.3.1	数据提取	060
4.3.2	数据转换	062
4.3.3	数据加载	063
4.3.4	ETL 设计与开发	063
4.4	数据仓库开发、管理与安全	064
4.4.1	数据仓库开发模式	064
4.4.2	数据仓库设计	066
4.4.3	数据仓库的逻辑数据模型	068
4.4.4	元数据	070
4.4.5	数据仓库的安全	072
4.5	分布式数据平台	074
4.5.1	分布式数据平台概念	074
4.5.2	分布式数据平台与功能组件	075
4.6	云数据平台	078
4.6.1	云数据平台概念	078
4.6.2	云数据平台与商务智能	079
小结		081
思考与练习		081
第 5 章	构建商务智能环境	083
5.1	商务智能环境	084
5.1.1	确定数据可用的能力	084
5.1.2	数据挖掘的能力	085
5.1.3	用户与系统交互的能力	085
5.2	商务智能组织	086
5.2.1	外包商务智能	088
5.2.2	内给商务智能	089
5.2.3	商务智能组织成员	090
5.3	商务智能系统	090

5.3.1 商务智能基础设施	090
5.3.2 商务智能系统软件	091
5.3.3 商务智能系统产品	094
小结	096
思考与练习	096

方 法 篇

第 6 章 数据预处理	099
-------------------	-----

6.1 数据预处理简介	100
6.1.1 数据预处理的原因	100
6.1.2 数据预处理的目的	102
6.1.3 数据预处理的方法	102
6.2 数据清洗	103
6.2.1 缺失数据处理	103
6.2.2 噪声数据处理	105
6.3 数据集成、规范与归纳	107
6.3.1 数据集成处理	107
6.3.2 数据规范化处理	108
6.3.3 数据归纳处理	110
6.4 数据消减	115
6.4.1 数据冗余清除	115
6.4.2 数据采样	116
6.4.3 数据立方合计	117
6.4.4 数据属性选取与生成	118
6.4.5 数据压缩	120
6.4.6 数据离散化与概念分层	121
小结	124
思考与练习	125

第 7 章 关联规则	126
------------------	-----

7.1 关联规则简介	127
7.2 关联规则挖掘方法	130
7.3 关联规则兴趣性	133
7.4 关联规则知识形式扩展	136
7.4.1 广义关联规则	136
7.4.2 数量关联规则	138
7.4.3 时态关联规则	139
7.5 简单关联规则	140
小结	143
思考与练习	144

第 8 章	分类分析	145
8.1	分类分析简介	146
8.2	决策树分类	147
8.2.1	决策树构建	148
8.2.2	决策树剪枝	151
8.3	贝叶斯分类	153
8.3.1	贝叶斯定理	153
8.3.2	简单贝叶斯分类器	153
8.3.3	贝叶斯信念网络	155
8.4	其他分类方法	157
8.4.1	神经网络分类	157
8.4.2	支持向量机分类	157
8.4.3	懒惰型分类器	158
8.5	分类准确率	159
8.5.1	分类准确率比较与评估	159
8.5.2	提高分类器的准确率	164
	小结	165
	思考与练习	165
第 9 章	聚类分析	167
9.1	聚类分析简介	168
9.2	相似度与距离测度	169
9.3	聚类分析方法	172
9.3.1	划分方法	172
9.3.2	层次方法	173
9.3.3	基于密度的方法	175
9.3.4	基于网格的方法	175
9.3.5	基于模型的方法	176
9.4	k-means 方法	176
9.5	DBSCAN 方法	179
	小结	184
	思考与练习	185
第 10 章	社会网络分析	186
10.1	社会网络的中心性	187
10.1.1	度中心性	188
10.1.2	贴近中心性	188
10.1.3	中介中心性	189
10.2	社会网络的权威	190
10.2.1	度权威	190

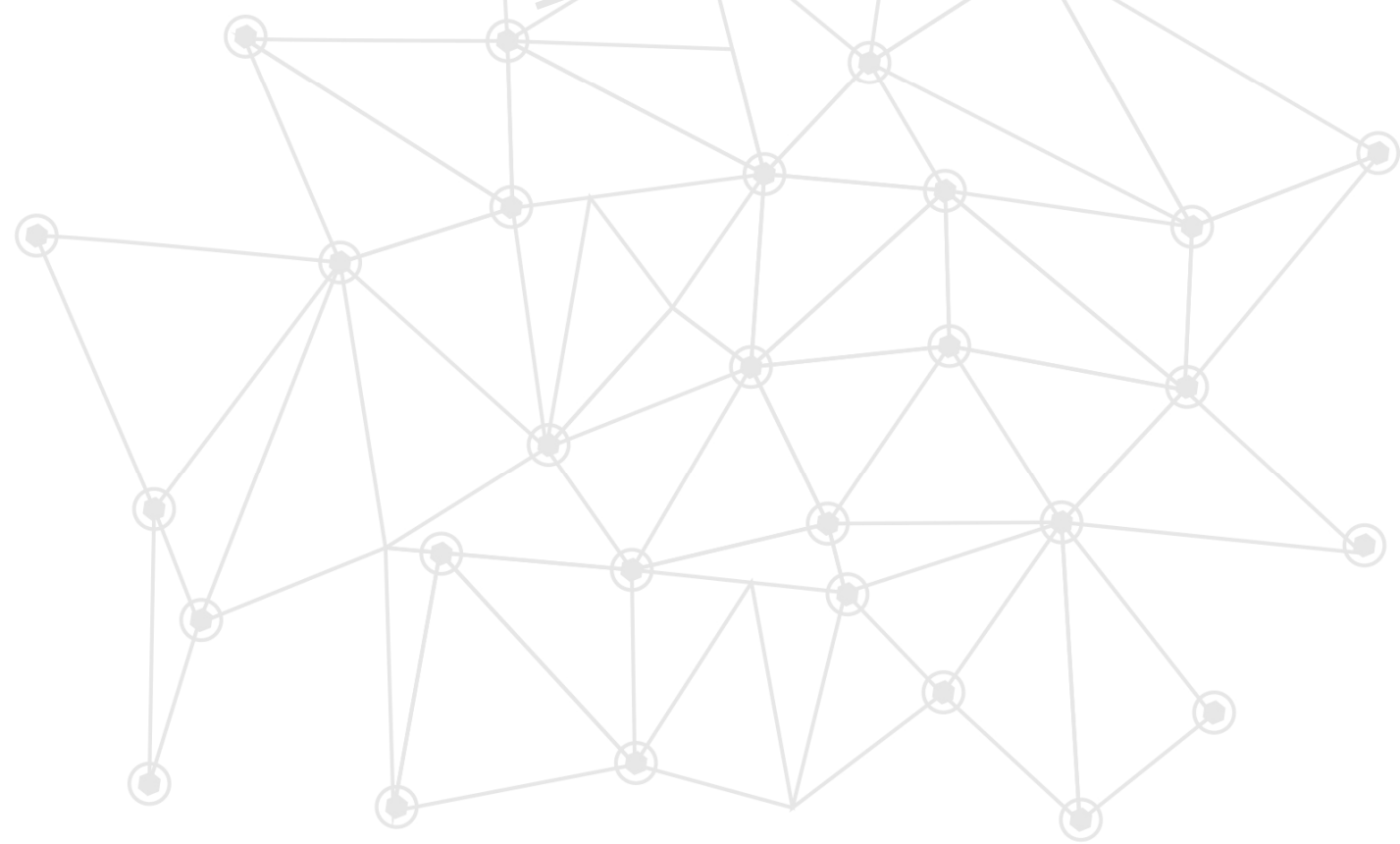
10.2.2	邻近权威	191
10.2.3	等级权威	192
10.3	引用社会网络	192
10.3.1	同引分析	192
10.3.2	引文耦合	193
10.4	社会网络的链接分析	193
10.4.1	PageRank 算法	194
10.4.2	HITS 算法	196
10.5	社会网络中的社区	198
	小结	199
	思考与练习	200
第 11 章	概率图模型	201
11.1	概率图模型简介	202
11.2	朴素贝叶斯模型	203
11.3	隐马尔可夫模型	205
11.3.1	马尔可夫过程	205
11.3.2	隐马尔可夫建模与处理	206
11.4	高斯混合模型	209
11.5	LDA 模型	211
	小结	214
	思考与练习	215
第 12 章	神经网络	216
12.1	神经网络简介	217
12.2	前馈神经网络	218
12.2.1	神经元	218
12.2.2	激活函数	218
12.2.3	前馈神经网络	220
12.3	卷积神经网络	222
12.3.1	卷积层	222
12.3.2	汇聚层	224
12.3.3	卷积神经网络结构	225
12.3.4	典型卷积神经网络	225
12.4	循环神经网络	226
12.4.1	循环神经网络基础结构	227
12.4.2	长/短期记忆网络	228
12.4.3	门控循环单元网络	230
12.5	注意力机制	231
12.5.1	多头注意力模型	232

12.5.2	自注意力模型	232
小结		234
思考与练习		234
第 13 章	多模态数据表征	236
13.1	文本表征	237
13.1.1	词袋模型	237
13.1.2	Word2vec 模型	239
13.1.3	Doc2vec 模型	240
13.1.4	主题模型	241
13.2	图像表征	243
13.3	音频表征	244
13.4	视频表征	247
小结		248
思考与练习		248
专 题 篇		
第 14 章	信息提取与洞察	251
14.1	“大数据—小数据”问题概述	252
14.2	代表性信息提取	254
14.2.1	代表性评估测度	254
14.2.2	代表性信息提取方法	255
14.3	一致性信息提取	257
14.3.1	一致性评估测度	258
14.3.2	一致性信息提取方法	259
14.4	多样性信息提取	260
14.4.1	多样性评估测度	261
14.4.2	多样性信息提取方法	262
小结		263
思考与练习		264
第 15 章	关联分类	265
15.1	生成分类关联规则	266
15.2	分类关联规则剪枝	269
15.2.1	后剪枝方式	269
15.2.2	先剪枝方式	272
15.3	构建分类器	275
15.3.1	单一规则分类器	275
15.3.2	多规则分类器	277
15.4	混合型关联分类	278

15.5	GARC 方法解析	278
15.5.1	GARC 思路与算法框架	279
15.5.2	数据实验与方法比较	282
	小结	285
	思考与练习	286
第 16 章	不确定性知识发现	288
16.1	不确定性信息表达	289
16.2	分区中的边界问题	296
16.3	数据间的部分隶属性	300
16.4	不完整数据依赖	303
	小结	307
	思考与练习	307
第 17 章	智能推荐	309
17.1	信息推荐方法	310
17.1.1	推荐系统概述	310
17.1.2	协同过滤推荐方法	312
17.1.3	推荐系统评测指标	314
17.1.4	消费者信息搜索过程中的推荐	316
17.2	多模态信息推荐	319
17.2.1	多模态信息表征	320
17.2.2	多视图信息整合	322
17.2.3	商品推荐场景	322
17.3	序列推荐	323
17.4	捆绑推荐	325
	小结	328
	思考与练习	328
第 18 章	商务智能的经济社会影响与发展	329
18.1	商务智能与管理决策	330
18.2	商务智能的发展趋势	332
18.2.1	移动商务智能	332
18.2.2	人工智能赋能	333
18.2.3	数据安全性与隐私保护	334
18.2.4	数据治理	336
	小结	337
	思考与练习	338
	参考文献	339
	索引词检索目录	362

基础篇

BI



第 1 章

引 言

BI

回眸漫漫的历史长河，人类习惯用最具代表性的生产工具来描述文明划过的轨迹：从石器时代到青铜器时代，再到铁器时代和蒸汽时代。用这种思维模式来看待过去的一百年，我们会说，人类已从电气时代步入了信息时代。世界上第一台电子计算机 ENIAC 的神奇并未因第二次世界大战炮声的远去而消退。没有硝烟的商业战场为信息通信技术的蓬勃发展和广泛应用提供了靓丽的舞台；Apple 计算机的出现、Microsoft 软件的盛行、Google 搜索的普及、智能手机的风靡、淘宝平台的繁荣、微信的流行等，不仅彻底改变了商务活动的形式与内容，更潜移默化地影响着人类日常生活中更广阔的领域。信息量、信息传播速度几何级数般的增长也对信息处理能力及应用手段不断提出新的需求。当获取信息越发容易时，有效分析数据并支撑管理决策的能力逐渐成为企业逐鹿商场的核心竞争力，使之成为“智慧”的企业。这就是本书将讨论的商务智能。

本章学习重点

- ❖ 商务智能与信息社会
- ❖ 商务智能与企业管理
- ❖ 商务智能与数据科学
- ❖ 商务智能与新技术融合

随着现代信息技术（Information Technology, IT）的飞速发展，计算机系统、互联网（Internet）和移动终端（如智能手机、平板电脑等）的广泛应用使得社会活动、组织运作和个人生活日趋数字化。一方面，作为日常经济活动的主体，企业收集和存储了大量数据，如市场情况、客户资料、订单、库存状态、生产过程、交易账目、通信记录等。这些数据反映了企业内部机制和运作管理状态的重要内容，称为企业内部数据。另一方面，企业的运作管理受到企业外部活动的影响，而且这种影响在社会化媒体和用户广泛参与的背景下日趋强烈，如评论、口碑、流言、商誉和舆情等，构成了企业外部数据。只有对这些企业内/外部数据进行充分、及时的分析和利用，才能增进对企业业务、用户行为、发展潜力、社会形象的理解，提高管理决策水平，有助于发现新的市场机会和创新模式，赢得可持续的生存和竞争优势。这种对海量数据进行分析利用的迫切需求促成了商务智能（Business Intelligence, BI）的产生，并且成为三十多年来学界和业界研究和应用的热门课题之一。纵观信息技术创新和应用的发展趋势，结合信息化进程的移动性、虚拟性、个性化、社会性、极端数据等主流特征，以及的物联网、云计算、大数据、人工智能为标志的新兴计算范式变革，商务智能的研究和应用将具有更加广阔的前景，并将在企业决策和高质量发展中发挥更加重要的作用。

1.1 商务智能简介

在经济管理活动中，数据代表着对现实世界及业务活动的事实性记录，信息可以被理解为以数据的形式存在的、对现实世界语义的反映，知识代表着被组织起来用于解决问题的信息，智能则意味着利用知识以获取效益的能力。图 1-1 显示了从数据到智能的层次性联系。

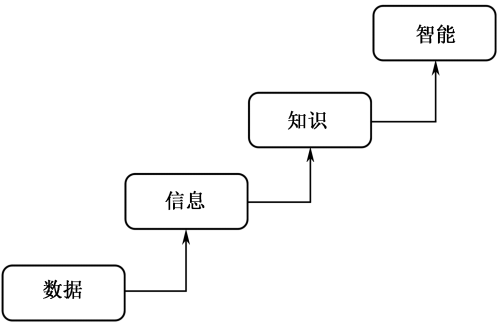


图 1-1 从数据到智能的层次性联系

企业中存储、流动、转换、展示的内部数据以及与运作管理相关的外部数据构成了企业的信息资源。对数据的有效管理与利用是信息资源管理的基础性任务^[1]。随着信息技术创新和应用的不断发展及数字化进程的不断深入，信息资源日益成为企业的战略性资源。对信息资源的开发与利用已经成为企业在竞争中生存并获取优势的必要途径^[2]。

信息资源的开发与利用包含两层含义：首先，数据只有得到有效的获取、管理和组织，才能通过系统化的应用服务于企业和组织的管理和决策；其次，对信息资源的利用存在着一个由浅入深、由单一到综合的提升过程^[1]。如图 1-2 所示，从企业对信息应用的角度，也就是根据企业对于得到的数字化信息进行利用以支持决策的程度，信息系统可以分为在线事务处理（OnLine Transaction Processing, OLTP）、在线分析处理（OnLine Analytical Processing, OLAP）和知识发现（Knowledge Discovery in Databases, KDD）三个层次。

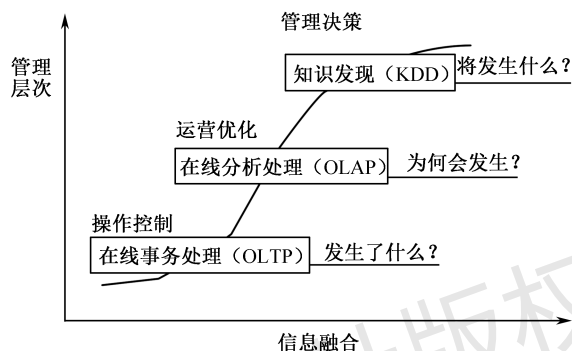


图 1-2 信息融合与管理支持层次

事务处理指的是围绕企业的基本业务和生产过程的自动化，对数据和信息进行加工和处理。在管理中，它主要回答“发生了什么”的问题。分析处理指的是围绕企业的分析和控制功能，对数据和信息进行回溯、分维、切片和 what-if 分析，从而回答“为何会发生”的问题。知识发现则是围绕企业的经营策略和竞争优势，对数据和信息进行挖掘和整理以求获得支持决策的知识，进而回答“将发生什么”的问题。

这三个层次逐级提升，是围绕着对信息资源的开发、管理和利用展开的。事务处理层要求实现数据的电子化采集、交换和处理；分析处理层要求对数据进行多维度的综合整理；知识发现层要求对数据进行深入的、智能化的挖掘分析，以寻找潜在的新颖知识。换句话说，知识发现是信息资源利用从数据和信息层次上升到知识和智能层次的关键纽带。当代信息技术具有的强大的计算能力，使得人们能够对数据进行多层次和多角度的处理，以获取新颖的、具有潜在有用性的知识。由于这些知识是事先未知的，也难以通过已有的规则或模式推断得到，因此，如果它反映了业务运作中潜在的某种规律性并被及时利用和把握，将对企业竞争优势的获取产生不容忽视的作用^[3]。

上述信息资源利用的三个层次在很大程度上代表着现代企业竞争力水平的三个台阶。当信息技术与信息系统应用推动着企业的信息融合水平从基本的事务处理发展到智能化分析的层次时，企业的管理者才能真正站在信息资源的瞭望塔上获得全局性视野，全面性的企业绩效管理和竞争决策才成为可能。正如耶鲁大学伊恩·艾尔斯（Ian Ayres）所说：“在当今时代，顶尖级别的组织机构每天都在以闪电般的速度分析大量的数据，以获取对人类行为的深层认知……这种新一代的决策者正在成为当今社会的掌舵者。”^[4]

当前,企业数字化的发展呈现出集成化和智能化两大趋势。在集成化方面,从传统的物料需求计划系统(Material Requirement Planning, MRP)、制造资源计划系统(Manufacturing Resources Planning, MRP II)中发展出企业资源计划系统(Enterprise Resource Planning, ERP),将采购、生产、销售、财务和人力资源等信息系统基本功能连为一体,形成了整个组织范围内的集成化信息系统。同时,电子数据交换技术和电子商务的发展不断推动着企业间信息系统的集成,其典型代表包括供应链管理系统、客户关系管理系统及电子商务交易平台系统等。而智能化方向的发展使得信息系统应用在全面集成的同时,逐步将其支撑力融合在企业的核心竞争目标绩效的提升上。因此,商务智能可以视为企业信息化达到一定高度的产物。

在这样的背景下,企业信息系统市场与商务智能市场出现了大规模的整合趋势。全球三大商务智能厂商 Business Objects、Cognos 和 Hyperion 分别被 SAP、IBM 和 Oracle 公司收购,也从侧面反映出在企业集成化信息系统的基础上,融合具有绩效管理和决策支持能力的商务智能应用正在成为提升企业竞争力水平的一个新的突破点。

商务智能是一个涉及信息科学、管理学、决策科学、数据科学等多学科知识的交叉领域,又具有很强的与应用背景密切融合的特征,人们往往从理论与实践、方法与技术等不同的角度对“商务智能”概念给出不同的解读。

1989年,Gartner Group 的 Howard Dresner 在提出“商务智能”时认为,它是通过应用基于事实的支持系统进行商业决策支持的一系列概念和方法^[5]。IBM 公司认为,商务智能是基于数据仓库、数据挖掘和决策支持中的先进技术等,收集相关的信息并加以分析,以发现商业机会和针对客户需求制定相应战略^[6]。按照 Oracle 公司的观点,“商务智能”是在合适的时间提供合适的数据访问以制定正确的决策^[7]。DM Review 杂志^①将“商务智能”理解为一个过程,一个依赖人、过程、数据、工具和方法论来制定更好的决策的过程^[8]。数据仓库与商务智能专家 Larissa Moss 认为,“商务智能”不是产品,也不是系统,而是一个体系,是一系列集成的应用,提供商业数据的查询和决策支持^[9]。而 Brian Larson 认为,“商务智能”是一种行为,即在一定的时间内向决策人员提供准确、有用的信息,以有效地支持决策活动^[10]。

概括来说,商务智能是一个从大规模(海量)数据中发现潜在的、新颖的、有用的知识的过程,旨在支持组织的业务运作和管理决策。数据挖掘(Data Mining, DM)和机器学习(Machine Learning, ML)是商务智能的核心技术,以一系列商务智能的理论和方法为基础。

商务智能概念中蕴含如下四要素。

(1) 大规模数据。商务智能是从海量数据中发现知识,因此数据挖掘/机器学习的效率是一个重要考量标准。许多传统的数据分析和信息查询方法面对海量数据时往往显得力不从心,因而对方法的扩展和创新成为必然。

(2) 数据驱动(Data-driven)。由于数据规模的缘故,传统的建模假设很难人工穷举数据

① 美国著名的《商业智能和数据仓库》杂志。

变量和属性组合，而商务智能方法通常自动地考虑数据变量和属性之间的组合关系，以获得相应的知识模式。

(3) 潜在新颖性。商务智能通过数据挖掘/机器学习获得的知识不是显见的，而是新颖的。这就使得这些知识用于决策时往往能够帮助企业获得竞争对手不曾掌握的信息，进而取得竞争优势。从此意义上，数据挖掘技术也称为优势提升技术。

(4) 知识有用性。通过商务智能得到的知识是用于业务运作和管理决策的。这就决定了商务智能的应用是对背景依赖 (Context-dependent) 的，也就是说，所发现的知识的有用性通常与应用环境 (包括应用问题、环境条件、决策者等) 密切相关。

特别值得指出的是，自 20 世纪 80 年代末商务智能被提出以来，“商务智能”和“商业智能”成为中国业界和学界惯常的用法，也是多个国际主流商务智能系统在中国市场上的叫法。这其中的一个可能的原因是早期的一些经典商务智能应用发生在商务/商业领域，如超市购物关联问题、银行信用分类问题等。在此，我们将沿用商务智能的叫法，但其应用不限于商务/商业领域，而是广义地面向组织 (如企事业单位等) 活动的所有业务领域。再者，Business Analytics (BA) 是近年来在业界和学界常用的一个概念，我们称之为深度商务分析 (或深度业务分析)。一般，深度商务分析是现阶段 (如大数据/人工智能背景下) 商务智能的概念，所以经常互用。具体说来 (见图 1-2)，商务智能的大量应用正在从前期的“OLAP 为主—KDD 为辅”模式转向“OLAP—KDD 并重”模式。这也是“深度商务分析”中“深度”的含义所在。

此外，为方便起见，下文将习惯性地使用“企业”作为“组织”的广义用法。当然，在必要的时候，将区分说明企业与其他组织形式在相关问题上的差别。

1.2 商务智能与信息社会

以摩尔定律速度发展起来的信息技术在当前的信息社会呈现了高加速度的快速发展态势，信息技术的形态多样，技术特征迭代创新，形成了信息社会的数字化支撑，丰富了信息社会的数字化内涵。信息技术在促进信息社会不断向前发展的同时，也创造了商务智能的发展沃土，商务智能进一步提升了信息社会的繁荣水平。

1.2.1 信息技术提升信息社会发展水平

自从个人计算机出现以来，信息技术一直处于持续而迅猛的发展之中。各种新兴技术层出不穷，推动着人类社会、经济、生活一波又一波的变革浪潮。20 世纪末期以来，互联网得到了广泛而深入的普及，从而触发了以社会性、虚拟性、移动性、个性化和大数据为典型特征的新一代信息技术的应用形式。新兴技术和新兴应用不断涌现，创造了一个动态变化的商

务环境。在这样的形势下，企业管理信息系统、电子商务系统、知识管理系统、Web 2.0 等技术和应用融合在一起，日益显示出其强大的生命力和巨大的商业价值，进一步推动信息经济和信息社会的大繁荣。以 5G（5th Generation Mobile Communication Technology，第五代移动通信技术）、物联网、云计算、人工智能等为代表的新兴技术也为信息社会的发展注入了新的时代元素。

相较于 4G，5G 具有数据传输速率更高、时延更低、带宽更高、设备连接密度更高、能耗更低等诸多优势^[11]。5G 正在重新定义世界连接的方式，未来数以亿计 5G 设备将产生新一轮更大的数据浪潮，更多的数据可以被应用在商务智能和商务分析中，进而发掘有效信息，推动企业的高质量发展。同时，各类行业的产品、服务、业务和运营模式都可能因 5G 赋能而产生更为多元的变化。例如，在零售业中，基于 5G 的增强现实和虚拟现实应用为消费者创造随时随地自由可视化体验产品的可能，提高消费者的满意度与购买意愿；在媒体娱乐行业中，依托 5G 技术，移动设备可以支持更高质量的流媒体；在制造业中，5G 技术使得利用高分辨率视频、触觉反馈、温度或其他传感器构建大型传感器网络成为可能，可用于工厂车间机器/机器人的预测性维护，识别和跟踪供应链中的货物并进行远程质量检查和诊断等。

物联网（Internet of Things, IoT）是一个由相互连接的智能设备和传感器组成的系统网络，这些微型智能设备和传感器实时采集需要监控的物体或过程信息，并将收集到的数据用于进一步的存储、处理和分析。物联网为企业描绘了跨时间、跨地点和跨功能的综合业务环境的整体视图，并为后续商务智能分析提供了信息数据来源^[12]。商务智能大数据分析解决方案使企业能够从物联网设备生成的海量数据中发掘商业洞察和管理启示，帮助企业提高运营效率、开发新产品、强化服务质量、提升风险管理水平，真正实现物联网的宝贵价值。

云计算（Cloud Computing）旨在通过一系列动态和虚拟化的整合优化计算资源技术，以类似效能（Utility）的方式（如电力供应形态），提供“随时获取，按需使用”的计算服务。云以透明的方式为用户提供基础设施访问、计算平台、软件、数据乃至业务流程的标准化服务，派生了 IaaS（Infrastructure as a Service）、PaaS（Platform as a Service）、SaaS（Software as a Service）、DaaS（Data as a Service）和 BPaaS（Business Process as a Service）等产品和服务形态^[13]。企业希望通过使用创新的商务智能解决方案在市场上获得竞争优势。一般而言，商务智能解决方案的集成涉及大量的人力和财务资本，对于许多中小型企业而言，对传统商务智能解决方案的大量投资通常是难以支撑的。此时，基于云计算的云 BI 解决方案应运而生。对于希望提高敏捷性、降低 IT 成本并利用商务智能和云计算优势的企业而言，云 BI 是一个合适的选择。云 BI 可以帮助企业在无须大量资本支出的情况下提高单位成本效益，在保持业务灵活性和扩展性的同时，提高数据存储的可靠性，增强企业数据分享和业务运营能力。随着云变得更加安全、可靠和经济实惠，商务智能与云计算的结合也将进一步提升。

人工智能（Artificial Intelligence, AI）泛指可模仿人类智能来执行工作任务，并基于收集的大规模信息对自身进行迭代式改进的系统和机器，可以通过自主式学习、分析复杂数据，

并通过一定的自然对话方式与人类进行交互，从而达到增强人类认知、分析和决策的能力，甚至在特定场景下取代人类的决策与执行，如在围棋比赛中大杀四方的 AlphaGo、移动支付中的快速人脸识别模块、客服系统中的聊天机器人等。人工智能技术在 20 世纪 50 年代提出后，经过几轮起伏不断发展，随着高速计算和大规模并行计算技术、大数据技术和以深度学习为代表的高性能算法的进步和助力，在近十年形成一些“质变”，特别是在自然语言处理、语音识别与合成、图像识别和分析、生成式人工智能（AIGC）等领域已出现一些较为成熟的工具和应用。这些人工智能方法、模型和算法，与商务智能系统采用的数据挖掘和机器学习技术一脉相承，越来越多地在商务智能系统中不断得到整合和集成，提升了商务智能系统的“智能化”决策支持水平。

以 5G、物联网、云计算、人工智能等为代表的新兴信息技术催生计算范式（Paradigm of Computing）向外部性和集中式转变，体现为物理上分散、逻辑上集中的特征。5G、物联网、云计算和人工智能强调物体间、虚拟资源间、人机间的互连，扩展了信息技术应用的疆界。这种外部性使得人一物信息交换、虚拟资源配置、内/外部数据融合成为新的计算范式特征。无处不在的探测感应装置、互联网络、移动终端、社会化媒体使得数字世界和计算能力处于物理上高度分散的分布式状态。与此同时，5G、物联网、云计算、人工智能等技术应用强调信息整合、数据平台、全局视图及高阶智能应用，呈现出了一种新型集中式智能计算范式。

信息技术的飞速发展和深入应用不断推进社会的进化。我们所处当前社会的一个重要的标志性特征就是技术融合（或称为 IT 融合）。也就是说，当前所处的信息社会与以往的工业社会、农业社会相比，处于一个前所未有的技术大融合的阶段。这体现在两方面：技术透明性和技术渗透性。技术透明性是指，由于技术水平的提高和广泛的业务应用，在越来越多的产品和服务中，顾客往往感觉不到（也无须过多了解）内在的技术细节。例如，手机用户无须了解无线通信基站的工作方式和信号发送指标，网络用户也无须了解网络通信的底层协议结构。最终用户只需了解技术呈现出的效用即可，因此技术对于用户来说具有透明性。这种技术透明性在信息社会中尤为明显。技术渗透性是指技术对人类社会和生活的方方面面的影响深度。在工业社会早期（18 世纪），虽然在工厂中大量的蒸汽机及机器设备得到应用，而人们在生活中仍然很大程度上保持着传统的生活方式，如交通还是靠人力或畜力。到 19 世纪，蒸汽火车等交通工具才得到发明，但并不是所有的普通民众可以轻易享用的。这个技术渗透的过程经过了约 100 年。而在信息社会，新技术从发明到获得普通民众使用的速度大大加快，对民众生活方方面面的影响也更深入。例如，从第一封 E-mail 发出到 E-mail 技术在全球范围内得到普及并成为人们日常生活中最常用的信息技术之一，也就经过了约 20 年时间。又如，录音机从出现到获得 5000 万受众用了 38 年时间，而互联网获得同样规模的受众仅用了 4 年^[356]。

近年来，随着信息技术不断进步与社会经济生活的数字化改造，全球数据量正呈现出前所未有的爆发式增长态势^[14]。国际数据公司（IDC）的数字宇宙研究报告称：2020 年，全球

被创建和被复制的数据总量达到 64.2 ZB, 且增长趋势大约每 2 年翻一番, 预计 2025 年将超过 180 ZB^[15]。与此同时, 数据复杂性急剧增长, 其多样性 (多源、异构和富媒体等)、低价值密度 (大量不相关信息和知识“提纯”难度高)、实时性 (流数据, 需实时生成、存储和分析) 等复杂特征日益显著。“大数据” (以超规模、多样性、低价值密度和实时性为显著特征的数据) 成为重要话题, 并在业界和学界引起广泛关注^[16-20]。

信息技术的进步和融合对于人们的社会生活和经济活动产生着日益深远的影响。一方面, 企业中许多传统的业务决策问题逐渐变成了基于信息分析的决策问题。例如, 通过产品浏览日志和评论内容分析以获得网络购物用户的行为偏好, 通过社会传播和关系强度分析以辨识舆论领袖和群体联系, 通过对搜索引擎日志和广告内容的分析和匹配来推荐适合的关键词广告, 通过对股票交易数据的深度模式挖掘来设计程序化高频交易策略, 通过特征提取和相似关系构建目标人群和业务策略等。另一方面, 随着技术创新信息产品 (如软件和 IT 服务等) 及其应用 (如虚拟体验、个性推荐、社会网络和网络搜索等) 呈现出越来越丰富的形态和特征, 催生出新模式 (如线上线下互动的 O2O (Online-to-Offline 或 Offline-to-Online) 运作模式、融合渠道 (Omnichannel) 业务模式)、新业态 (如众包 (Crowdsourcing)、数据标注、视频直播电商等生产组织和商业模式)、新人群 (如生活在赛博空间 (Cyberspace) 中的人群及其行为和需求) 等。在这两方面, 商务智能之数据挖掘/机器学习技术均扮演着重要角色。

1.2.2 商务智能是信息社会的产物

信息社会的技术融合使技术透明性和渗透性非常高, 这也为商务智能的产生和发展提供了肥沃的土壤, 使得企业和广大用户可以利用强大的信息工具和智能分析方法从业务运作和数字化生活中提取信息和知识, 进一步为企业决策和生活需求提供有效支持, 使企业获得可持续竞争能力、广大用户获得高质量的服务。

具体说来, 信息社会中技术融合的深化使企业和个人面临着数据激增的问题。特别是, 相对于数据海洋的浩瀚流量, 与个体企业或个人相关的价值信息显得相对微小, 这就是所谓的大数据时代的低价值密度特征。而传统分析工具的功能有限, 传统的业务报告和面向用户的浏览和提取服务无论在内容上还是时效性上都难以满足需要。

1. 数据充足而知识相对匮乏

现代信息技术的发展产生了以数字业态为代表的数字经济。在这一新型的经济和社会形态中, 企业活动面临日新月异的市场环境和挑战, 信息成为关乎企业兴衰成败的战略性资源; 人们在社会生活中对于数字内容的依赖和参与感, 使得信息成为关乎现代生活质量的基础性资源。企业中和社会上普遍存在数据充足而知识相对匮乏的矛盾, 迫切需要将大量数据转化

为有价值知识的技术和能力。这种对能够处理和控制信息的新技术的强烈需求不断催生着商务智能技术与应用。

2. 传统分析工具的整合能力有限

在过去的几十年，许多组织耗费大量财力和人力建立了联机事务处理（OLTP）系统和企业资源规划（ERP）系统等，这些系统在极大地改善数据自动处理能力的同时，也存在着获取和连接与企业（或部门）活动密切相关的外部数据的局限性，包括与客户和供应商数据的连接、与社会网络和企业舆情数据的连接、与电子商务生态圈数据的连接等，这给从全局的角度观察和分析企业运作、价值链管理和竞争态势带来了很大困难。用户被限定在有限的数据对象中，无法连接不同的信息源，更无法进行整合和进一步的分析，大大降低了解决方案的实用性和灵活性。

3. 传统报告不能满足用户需要

由于传统报告以业务支持为导向，而不是以分析为导向，因此传统报告的弊端之一在于数据细节水平与实际需求的不匹配，即决策者不能得到所需的有关现状、异常情况或未来趋势的有价值信息；弊端之二，在传统报告格式固定、时效性差、流动性弱，处理这样的业务数据时，在时间和资金上都会造成极大的浪费；弊端之三，在于传统报告对问题的分析和解答不够全面和透彻，尤其是涉及多部门、多角度的深层次的问题时，传统报告显得束手无策。

商务智能在有效解决上述问题方面可以发挥重要作用。商务智能是分等级的，可以提供高层次的汇总，也可根据需要挖掘细节；商务智能是多维的，给人们提供了从不同角度分析数据的能力。

4. 信息技术及应用的推动

基于互联网的企业运作模式、电子商务、电子支付、移动和位置服务、社会化媒体等应用的发展普及，为商务智能实践提供了有利环境。另外，信息技术的进步和创新，如物联网技术、移动通信技术、大容量存储技术、数据仓库管理工具、数据挖掘算法、社会网络分析、云计算平台、大数据分析与人人工智能等的发展，使得商务智能在理论和方法上不断丰富和发展。网络技术促进了供应商、分销商、合作伙伴、零售商与生产者之间的数据访问和共享，物联网和云计算平台进一步支撑了人们数字化生活的需求和服务整合。特别是在大数据和人工智能背景下，数智化及其赋能^[21]日益成为企业竞争生存、创新发展的核心能力，使得商务智能理论与方法既面临新的挑战，也具有广阔的发展前景。

1.2.3 商务智能是信息社会繁荣的推动力

商务智能为企业提供有效的决策支持技术和工具。从最初的基本报表到多维分析和知识

发现,帮助企业改善内部的经营,使企业能够创造更大的社会价值;同时,商务智能技术赋予了企业强大的创新动力,这些又成为推动信息社会和数字经济繁荣的重要动力。

尽管商务智能最初在企业中得到应用,但它的应用领域并不局限于企业。只要一个组织有大量的数据存储,或面对着大量相关的外部数据,就意味着具有进行智能分析的需求和必要性,商务智能就有用武之地。实际上,商务智能正在渗透到各行各业及各种职能中。例如,在银行业中,商务智能可以帮助工作人员更为精准地识别可能的违约者,从而实现贷款申请流程的自动化审批,监测信用卡和网上银行交易欺诈行为,预测客户偏好从而推荐相应产品服务实现客户价值最大化等。在零售物流行业中,商务智能可以预测特定地点、时间条件下不同产品的销量,从而实现库存调整、物流优化,结合市场销售情况分析识别不同产品之间的销售关联,进而优化促销策略等。在证券行业中,商务智能能够预测股票和债券价格的波动方向与幅度,评估特定问题和事件对于市场整体交易走势的影响,识别和防范证券交易中的欺诈行为等。在旅游业中,商务智能可以预测不同产品服务(如预订机票、酒店、车辆)的销售情况,以优化产品服务价格实现收益管理,定位目标客户并提供个性化服务。在医疗行业,商务智能也能够挖掘症状、疾病及治疗方法之间的联系,预测手术成功率,辅助医务人员做出正确、合理的决定。在市场营销领域,商务智能能够为目标客户提供更有价值的营销信息,在对与客户、产品、环境相关的海量数据进行智能化分析的基础上,实现对市场客户的精准细分,针对不同客户的不同特点,以最有效的方式和渠道去推介最适合的产品。像 Amazon 这样的网上零售商,凭借强大的数据挖掘管理,往往能在客户进入网站的时刻就向他们展示出最有吸引力的商品。除此以外,商务智能还被广泛运用于制造、保险、文体娱乐等行业,解决具体问题或探索新兴的商业机会,为使用者持续创造竞争优势。

近年来,商务智能渐渐进入政府机构,扮演日趋重要的角色。传统的政府机构管理中同样存在着数据孤岛、信息整合能力有限等问题,政府机构由于各种原因无法及时掌握下级机关的财务、绩效等情况,给政府机关实现科学化、精细化的管理和决策造成障碍。商务智能将来自业务支撑系统的财务、医疗、人事等结构化数据与协同办公、档案管理等行政办公系统中的报表、公文、档案等非结构化数据信息组织起来,同时结合社会舆情和民生热点进行信息分析,形成整个政府业务和公共管理的全局视图。在此基础上加以科学的分析,将其转化为辅助管理者决策的知识,及时解决政策执行中存在的各种问题。同时,商务智能软件提供的可信和及时的分析结果,可以帮助政府部门加快决策进程,缩短办事周期,提高整个机构的透明度,促进公共服务水平的提升。

此外,商务智能的信息提取和分析手段将大大帮助社会公众和用户有效获取生活资讯和相应支持(如关键词查询搜索、购物产品推荐、精准广告推送、评论概述提取、出行路径优化、好友和关系辨识等),并支持大量社会公众和用户参与生成的数据和信息(公众生成内容(Public Generated Content, PGC)或用户生成内容(User Generated Content, UGC),如博文、评论、意见、消息、情绪和创意等)的获取、分析和共享。

1.3 商务智能与企业管理

商务智能在企业管理中扮演着越来越重要的角色，本节以其在企业管理中的作用、方式及商业价值三方面进行说明。

1.3.1 商务智能在企业管理中的作用

商务智能对战略决策的支持表现在对公司战略、业务战略和职能战略等方面。在公司战略决策支持上，商务智能可以根据公司各战略业务单元的经营业绩和经营定位来选择合理的投资组合战略。在业务战略决策支持上，商务智能可以进行企业外部因素分析，如外部环境分析、行业状况分析、竞争对手分析、企业舆情分析等，以帮助各战略业务单元制定竞争策略。在职能战略决策支持上，商务智能可以在分析企业内部因素（如技术实力、资源能力、资本实力、投资方案、竞争实力和优/劣势等）的基础上为职能战略的决策提供科学的依据。

商务智能可以在企业各层面和职能中发挥作用，如营销管理、销售管理、客户关系管理、财务分析、供应链管理、人力资源管理 etc 能够利用商务智能来提高决策水平和改善业务绩效，增强企业核心能力。

(1) 在营销领域，商务智能可以使用分类和预测等技术，分析长/短期需求、市场机会和企业利益增长点，分析和预测市场容量、占有率、细分情况和风险程度，策划和评价市场营销策略等，据此调整和优化其市场营销策略，以获得最大的成功。

(2) 在销售管理方面，商务智能的作用可以体现在产品销售情况分析和服务信息反馈、新产品开发预算和销售预测等方面，基于销售、库存、财务和人事等多种基础数据进行多角度分析，给出销售情况分析和趋势预测等辅助决策信息。

(3) 在客户关系管理方面，商务智能可以根据消费记录及客户档案资料进行客户群分析，以及对他们的消费能力、消费习惯、消费周期、忠诚度、盈利能力、客户促销分析等进行分析，从而为企业更准确地理解客户的行为和趋势、挖掘潜在客户、制定相应的促销和服务策略、争取和保持客户等提供支持。

(4) 在财务分析上，商务智能可以满足企业管理者对费用支出情况的查询、应收账款和应付账款的分析等方面的财务分析，有助于提高在现金流量、资产负债和资金回收率等方面的决策科学化水平。此外，根据公众/用户生成内容（PGC/UGC）数据（如关键词搜索轨迹、事件评论热度和话题内容演化走向等），可以进行财务状况关联分析和股价走势预测，为管理决策提供支持。

(5) 在供应链管理上，商务智能可以进行流程分析、供应商评价、分销商管理和物流成本分析等，既可以满足对物流、供应和销售相关情况的基本层次的查询，也可以辅助对深层

次的问题的决策，如进行库存调整及优化，以保证有效、合理周转等。

(6) 在人力资源管理方面，商务智能可以基于翔实的数据进行多视角的人力资源统计分析，基于企业内岗位需求和市场上人才大数据进行智能分析和匹配，来挖掘潜在人才。商务智能的人力规划分析也可以实现不同角度的员工工资查询和分析，结合完成的工作量，提高员工利益分配的科学性。

归纳起来，商务智能的作用至少有以下四方面。

(1) 理解业务。商务智能可以用来对各项业务进行准确的评估，帮助理解业务的驱动因素，识别对业务产生影响的关键因素，积极推动业务发展，培养良好发展态势，规避潜在的风险。具有前瞻性的商务智能将帮助企业有效地解决所面临的各种问题和挑战。

(2) 衡量绩效。商务智能能够从企业各种应用系统中提取各种基础绩效指标与关键绩效指标（KPI），对员工的工作绩效进行追踪、衡量和评价，也可以进行其他各种绩效（财务的和非财务的、前台的和后台的、企业内的和供应链内的、组织的和个人的）的跟踪和管理，充分分析现有系统的执行力，从而完成业务流程的调整和优化。以商务智能为基础的企业绩效管理已成为欧美企业目前最热门的管理和信息技术课题之一。

(3) 改善关系。商务智能可以提供有关业务状况的有用信息、提高企业知名度、改善整个信息链的效率。商务智能也有助于加强客户忠诚度，一个参与其中并掌握充分信息的顾客更有可能购买你的产品或服务。商务智能有利于企业维持各种良好的关系，在问题转化成危机之前及时采取措施。

(4) 创造商机。掌握各种商务数据和信息的企业可以出售这些信息从而获得利润，但是企业需要发现信息的买主并找到合适的传递方式。通过构建基于信息分析的能力是当前许多企业成功进行转型升级的关键，也是新企业寻找机遇、创新发展的关键。

总之，商务智能可以帮助企业从大量内/外部数据中得到结论性的、基于事实的和具有实施性的信息和知识，使企业更快、更容易地做出高水平的决策，使企业管理者和决策者更清晰地了解业务状况，建立良好的关系，从而提高工作效率、增加利润并迅速发现商业机会。

1.3.2 商务智能协助企业管理的方式

商务智能为决策者提供可以用于决策的量化信息和知识，以提高决策水平和决策效率。商务智能主要通过以下五种方式协助企业管理。

(1) 基于目标的管理。一个企业或组织可能有众多的绩效目标。在横跨整个企业的信息系统上，结合外部的数据和信息，商务智能能够计算跨组织的绩效目标，与同行业或工业标准相比，便于企业明确自身的竞争优势。

(2) 基于异常的管理。商务智能具备即时计算各种绩效目标的能力，从而可以检测实际指标与计划目标之间的偏差。当偏差过大时，系统将迅速通知相关管理人员，便于及时采取

措施。而将例外管理与 workflow 技术相结合,可以使整个例外管理更加自动化。

(3) 基于事实的管理。企业的 ERP 系统在日常交易和处理中积累了大量的事实和数据。商务智能将企业目标与事实结合,使管理人员进一步分析原因或趋势,查询并探测相关信息。

(4) 基于外部数据的管理。在大数据背景下,企业外部数据(如社会网络数据、PGC/UGC 数据等)在企业运作和价值创造过程中发挥着越来越重要的作用。商务智能能通过获取、关联和分析企业外部数据,获得企业状态、竞争环境、社会关系和品牌形象等重要信息,便于管理人员进行及时响应和科学决策。

(5) 基于智能协同的管理。商务智能可实现企业外部资源与内部资源的协同集成,使管理人员不仅可以充分协同调度企业内部资源,还可以与客户、供应商及业务环境协同同步,改进商务协作及决策方式,有效地解决和避免目前信息系统中存在的诸多资源管理失调问题。

需要说明的是,商务智能不仅涉及相关的软件与工具,也承载着特定的业务和决策模式。商务智能的良好应用往往不仅是技术问题,更包括与业务和决策融合的问题,这样才能有效、充分地发挥其作用。

1.3.3 商务智能的商业价值

信息技术给企业带来的回报可以总结为三层。第一层,信息技术可以提高速率,即加速现有流程、提升业务响应能力;第二层,信息技术可以提高生产率,即以较少的人力、物力完成较多的工作;第三层,信息技术可以改变或完善企业的经营模式,从而取得竞争优势。

据市场调研机构 IDC 对欧洲和北美 62 家采用商务智能的企业的调查分析^[22]显示,这些企业的 3 年平均投资回报率为 401%,其中 25%的企业投资回报率超过 600%。调查结果还显示,一个企业要想在复杂的环境中获得成功,高层管理者必须能够控制极其复杂的商业结构,若没有翔实的事实和数据支持,是很难办到的。那些在商务智能领域投资并已得到回报的企业是最好的例子。日本最大的 CD 和视频零售商 Tsutaya,运用客户关系管理系统、数据仓库,通过下属的 TOL (Tsutaya 在线)为客户提供服务。TOL 在当时日本经济状况不佳的情况下,使 Tsutaya 利润增长了 48%。

需要说明的是,商务智能作为对信息的提炼和知识的积累,是企业的一项重要的隐性资产,不能只用传统的指标来衡量。麻省理工学院 (MIT) 的信息管理研究机构首席研究员 Eric Brynjolfsson 花了近 10 年的时间研究信息技术成本和收益,他的研究表明,信息技术投资中高达 90%的成本和收益都是隐性的^[23]。企业在进行信息技术投资评估时,必须公正地分析显性和隐性两方面的信息,从而获得适当的平衡。商务智能作为一项信息技术投资,其总体价值也包括显性和隐性两方面,因而需要在评估的时候做好平衡。

市场研究集团 Gartner 建立了 VOI (Value on Investment) 投资价值的理论框架,用来衡量隐性资产投资所创造的价值和财富,特别适合衡量企业从商务智能中获得的总体价值^[24]。由

美国商务智能专家莫斯和阿特瑞提出的分析框架列举了收入增加、利润增加、客户满意度提高、节省和市场份额增加五个收益类别，大多数是商务智能的显性收益，但也包括了像客户满意度这样的隐性价值。此外，BASE 咨询公司、Nucleus 研究公司都提出了商务智能的成本和投资回报的计算方法^[55]。

当前，大数据技术和服务市场呈现快速增长态势。以麦肯锡评估西方产业数据为例，大数据将能使欧洲发达国家政府节省至少 1000 亿欧元（约 7280 亿元人民币）的运作成本；使美国医疗保健行业降低 8% 的成本（每年约 2 万亿元人民币）；并使得大多数零售商的营业利润率提高 60% 以上^[16]。据 IDC 调研，2021 年，全球大数据和业务分析解决方案的支出达到 2157 亿美元，与之相比 2010 年的数额为 32 亿美元，并预计 2021—2025 年年度复合增长率将高达 12.8%^①。这里，大数据市场价值的重要成分是通过大数据分析应用实现的，其中商务智能扮演着主要角色。

1.4 商务智能的方法

商务智能的关键是要通过深度数据分析获取知识以支持决策，其中数据挖掘和机器学习是商务智能的核心方法。本节将对这两种方法进行简要介绍。

传统意义上，数据挖掘主要针对数据库数据进行知识发现。随着大数据时代的到来，数据已经不仅局限于传统数据库的框架中，相应的数据挖掘和知识发现技术能处理和分析的数据范围也不断扩大，包括结构化的数据库数据和非结构化的大规模数据。简而言之，数据挖掘主要基于人工智能、数据库、统计学等技术，高度自动化地分析大规模数据，做出归纳性的推理，从中挖掘出潜在的模式，预测客户的行为，帮助企业决策者调整经营策略，减小风险，做出正确的决策。

从认识层面，数据挖掘的基本目标是预测（Prediction）和描述（Description）。所谓预测，就是利用数据中已知的变量和字段来确定一些感兴趣的未知或未来的值。而描述集中于寻找一种人类能够解释的模式来对数据进行刻画。

一般来说，根据知识类型，数据挖掘可以分为以下 6 类。

(1) 概念描述（归纳或简约），包括概念的识别和比较。它通过将数据进行一般化、汇总或将可能矛盾的数据特征进行说明，来寻求对一个数据子集的简约描述。例如，将“湿”和“干”的概念进行数据上的说明。

(2) 聚类，将一组个体按照某种标准进行汇总，形成新的类。它的目的是使属于同一类别的个体之间的差异尽可能地小，而属于不同类别的个体间的差异尽可能地大。例如，将房

① 来源 IDC 网站。

屋进行聚类分析以寻找房屋分布模式。从机器学习的角度，这是一种非监督学习方式。

(3) 关联规则，发现数据之间的关联性、相关性和因果关系。例如，发现形如“在购买面包和黄油的顾客中，有 90%的人同时买了牛奶”的关联模式。关联规则发现的思路还可以用于序列模式发现，即不同对象在时间或序列上的规律。

(4) 分类和预测，一方面包括按不同的属性值将数据进行分类，挖掘出关于每类数据的描述或模型，如按气候特点将国家进行分类；另一方面包括根据已有的信息和模式来预测未来的或未知的属性值。从机器学习的角度，这是一种监督学习方式。

(5) 时序数据分析，这是统计方法的直接应用，主要包括趋势和偏差分析、用户定义的模式匹配分析及周期数据分析。

(6) 其他的模式识别和统计分析方法，如回归分析、相关分析。

值得说明的是，上述这些方法多为基础性方法，可以直接应用于商务智能任务，也可以与其他方法一起组合应用，面向各种商务智能任务。诸如关键词广告推荐、网上购物推荐、代表性信息提取、模式关联分析等都是组合应用的例子。

典型的数据挖掘系统一般如图 1-3 所示。

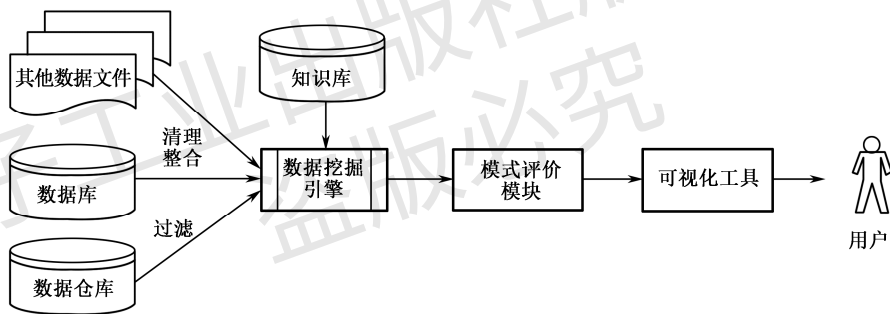


图 1-3 典型的数据挖掘系统

(1) 数据库、数据仓库和其他数据文件。数据挖掘的对象不仅是前面提到的数据仓库，还可以是数据库或其他格式的数据文件。数据挖掘技术可以用于分析各种类型的数据，如结构化数据、半结构化数据及非结构化数据、静态的历史数据和动态数据流数据等。

(2) 知识库。数据挖掘中常常会用到领域知识，它们被存储到知识库中，并用于引导挖掘，或者评价挖掘结果的兴趣性。这种知识可能包括用于在不同层次上组织属性的概念层次（Concept Hierarchies）知识，可以用于评价挖掘模式是否具有新奇性的用户信念知识，以及其他知识，如约束、阈值、元数据等。

(3) 数据挖掘引擎。这是数据挖掘系统的关键部分，包括一系列功能模块，如特征提取模块、关联规则分析模块、分类模块等。每个模块还可能有多多种备选算法。

(4) 模式评价模块，通常是通过应用兴趣度标准来精练、聚焦发现的模式。模式评价模块还会评价存储在知识库中的阈值，有时集成到数据挖掘模块中。

(5) 可视化工具。这是数据挖掘系统与用户的通信接口。用户通过它制定数据挖掘的计

划，提供挖掘所需的信息，浏览数据挖掘的结果，评价挖掘到的模式。

虽然数据挖掘与其他决策分析工具一样，也具有对数据进行归纳和分析的功能，但是它具有传统决策分析工具所不具有的特点，因此成为商务智能的核心技术，能够得到传统分析工具无法得到的新颖的、潜在有用的管理知识和业务规则。

企业管理决策人员通常会有许多信息需求（见图 1-2）。他需要了解业务过程中发生了什么？接下来要了解它为什么发生？顾客和市场的行为是怎样的？他还要了解将会发生什么及可以采取什么样的行动？当给出的分析结果是可以采取行动的建议时，管理决策人员获得分析结果的价值就会高。在传统决策分析技术中，这些先由决策人员事先给出逻辑假设，然后在数据中进行检验。但是，在大规模的数据中，如果仅根据某个或某几个逻辑假设来进行验证，就可能丧失对一些潜在的而事先未知的模式进行检验的机会，就不会得到新颖的模式。

数据挖掘是数据驱动的，并不始于一个有待证明的具体逻辑模式，而始于纷繁复杂的海量数据，利用强大的数据分析工具和特定的知识提取方法，从数据出发，对各种模式进行匹配，经过筛选，获得潜在的、新颖的和有用的知识。

对传统的决策分析工具和数据挖掘的不同驱动方式，即数据挖掘与传统数据分析的本质区别在于，数据挖掘是在没有明确假设的前提下去挖掘信息和发现知识，得到的知识具有先前未知、潜在有效和可实用等特征，如图 1-4 所示。

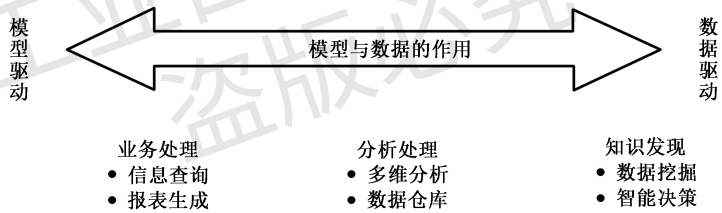


图 1-4 模型驱动与数据驱动

商务智能的另一项重要技术是机器学习。机器学习赋予了计算机在没有明确编程情况下的学习能力，从数据中进行探索学习并预测。从这个意义上，上述数据挖掘基本方法也在不同程度上具有机器学习的特征。机器学习通过根据样本的输入构建模型，进行数据驱动的预测或决策，克服了严格静态的程序指令遵循。机器学习算法被用于不同的领域，如电子邮件过滤、语音识别和计算视觉等^[25]。

机器学习方法传统上可以分为以下三大类，具体取决于学习系统可用的“信号”或“反馈”的性质。

(1) 有监督学习。有监督学习算法根据训练集中的输入/输出对将输入映射到输出，训练集中每个示例都是一对，由输入对象（通常是向量）和输出值（也称为监督信号）组成。有监督学习算法分析训练数据并生成推断函数，该函数可用于映射新示例。常见的有监督学习算法包括支持向量机（Support Vector Machines）、线性回归（Linear Regression）、逻辑回归（Logistic Regression）、朴素贝叶斯（Naive Bayes）、决策树（Decision Trees）、线性判别分析

(Linear Discriminant Analysis) 等。

(2) 无监督学习。无监督学习算法没有为训练数据提供任何预先分配的标签,从数据中学习潜在的模式或规则,而不以预先定义的真值作为基准。无监督学习包括聚类 (Clustering) 和主成分分析 (Principal Component Analysis) 等。

(3) 强化学习。不同于有监督学习和无监督学习方法,强化学习描述和解决的是,智能体在与环境的交互过程中通过学习策略达到回报最大化或实现特定目标的任务。常见的强化学习模型是标准的马尔可夫决策过程 (Markov Decision Process, MDP) [26]。

近年来,深度学习作为机器学习的一个分支获得了越来越多的应用。大多数深度学习方法具有多层和大量参数的神经网络架构,因此也被称为深度神经网络。深度学习使用多层非线性处理单元的级联来进行特征提取和转换,在靠近数据输入的较低层学习简单特征,而在较高层学习从较低层特征派生的更复杂特征。该架构形成了分层且强大的功能表示形式,意味着深度学习适用于从大量数据和从不同来源收集的数据中分析和提取有用知识[25]。深度学习算法被广泛应用于语音识别、图像识别、自然语言处理、客户关系管理、推荐系统、移动广告投放等与商务智能息息相关的场景中。

1.5 商务智能的数据

没有充足的、高质量的数据作为基础,商务智能很难有效发挥其威力。本章最后将从数据来源和数据类型的角度简要探讨商务智能中的数据。

从数据来源角度,商务智能的数据来源主要有企业内部数据和企业外部数据两种。企业内部数据的种类十分丰富,如企业的交易数据和 POS 信息、客户关系管理 (Customer Relationship Management, CRM) 数据、内部各类文档表格、历史存档数据,以及其他内部应用程序 (包括项目管理、市场营销、企业资产管理、人力资源管理、成本管理等自动化应用程序) 产生的数据。企业外部数据的常见来源如社交媒体和政府开放数据。通过社交媒体平台,企业可以获得较为真实的用户反馈,以便更好地了解当前和潜在客户,如社交媒体上的数据包括用户行为数据 (用户点击习惯、搜索记录、业务流量等)、用户生成内容数据 (用户评论、互动交流及注册产生的资料数据等) 和用户关系数据 (粉丝、好友等关系网数据) 等。除了社交媒体平台,政府也会提供一些公开的信息资源,可以帮助企业更好地了解公众。

从数据类型角度,商务智能常常需要关注数据的结构化程度,结构化程度是衡量信息和数据能够使用统一的数据结构加以表示的程度。根据结构化程度,商务智能数据可以分为结构化数据、半结构化数据和非结构化数据[27]。结构化数据指可以在记录文件里以固定格式存在的数据,通常可以通过关系型数据库表示、存储和管理。常见的结构化数据如日期、电话

号码、信用卡号码等数据。非结构化数据不符合任何预定义的数据模型，不易组织和格式化。非结构化数据通常涉及较多不规则和模糊的数据信息，与数据库中以字段形式存储的结构化数据相比，对于计算机程序来说更难理解和处理。常见的非结构化数据如文本文档、PowerPoint 演示文稿、JPEG 和 GIF 图像文件、视频文件、网页等类型的数据。部分非结构化数据中存在一定的有迹可循的内部结构，为了更准确地描述这种差别，半结构化数据的概念被提出^[8]。半结构化数据介于结构化数据和完全非结构化数据之间，不遵循与关系数据库或其他形式的数据表关联的表格结构，但仍然包含一些可以用于分离语义元素或划分数据中记录和字段层次结构的标签或其他标记。简单来说，半结构化数据是无法在关系数据库中组织或没有严格结构框架的数据，但确实具有一些结构属性或松散的组织框架。常见的半结构化数据有 XML 文档、JSON 文档、E-mail 等。

小 结

在经济管理活动中，数据代表着对现实世界及业务活动的事实性记录，信息可以被理解为以数据的形式存在的、对现实世界语义的反映，知识代表着被组织起来用于解决问题的信息，智能则意味着利用知识来获取效益的能力。

商务智能是随着信息技术的高速发展和企业信息化和数字化实践的不断深化而产生的，是信息社会高度发展的产物；同时，商务智能的出现和发展极大地推动了信息经济的发展和信息社会的进步。

商务智能是从大量数据中提取出未知的、潜在有用的和新颖的知识的過程，旨在支持组织的业务运作和管理决策。数据挖掘、机器学习、深度学习都是商务智能中常用的技术方法。常见的数据挖掘的基本方法有概念描述、聚类、关联规则、分类和预测、时序数据分析和和其他一些模式识别和统计分析方法。常见的机器学习方法有支持向量机、线性回归、逻辑回归、朴素贝叶斯、线性判别分析、决策树等。机器学习的一个重要分支是深度学习（如深度神经网络）。

商务智能应用已经对商务环境及现代企业的经营、管理活动产生了深刻的影响。近年来，商务智能领域的创新不断加速，并越来越紧密地与企业系统、电子商务系统、知识管理系统、Web 2.0 等类型的技术及应用融合，日益显示出其强大的生命力和巨大的商业价值。

以 5G、物联网、云计算、大数据和人工智能为标志的计算范式变化具有外部性和集中式的特征。在大数据背景下，企业外部数据（包括公众/用户生成内容（PGC/UGC）等）和社会化媒体将受到更多关注。商务智能应用将扮演着越来越重要的角色，其理论和方法创新也面临着新的机遇和挑战。

思考与练习

- 1.1 什么是商务智能？商务智能概念中的要素有哪些？
- 1.2 技术融合的主要含义是什么？请举例说明。
- 1.3 试举例说明商务智能在企业管理各领域中的作用。
- 1.4 试举例说明利用商务智能方法对用户生成内容（UGC）进行分析的必要性。
- 1.5 试从不同数据来源、不同数据类型角度分别举出若干商务智能数据的例子。
- 1.6 数据挖掘/机器学习与商务智能的关系是什么？
- 1.7 试描述模型驱动与数据驱动的特点和区别。
- 1.8 试说明商务智能在与新技术融合中的若干应用领域和实例。

电子工业出版社版权所有
盗版必究