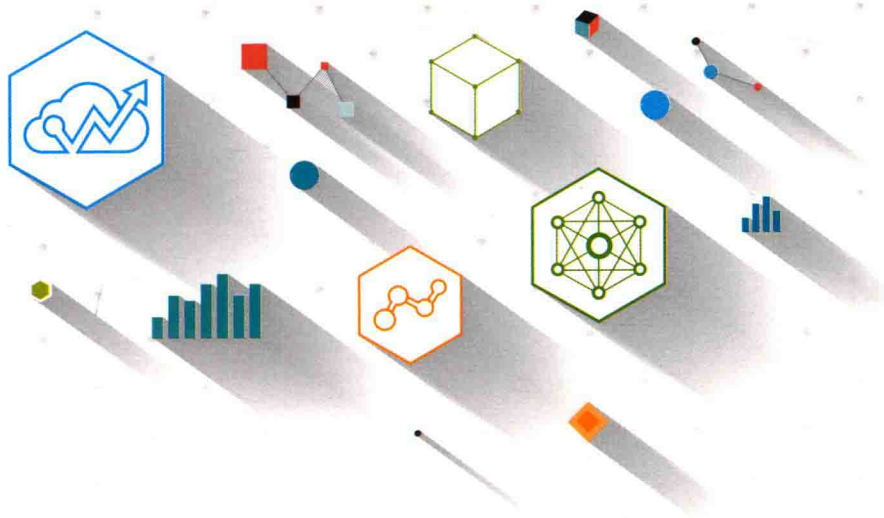




大数据技术与应用专业规划教材

商务智能 与 数据挖掘 (第2版)

◎ 蔡晓妍 杨黎斌 张晓婷 李梅 编著



清华大学出版社

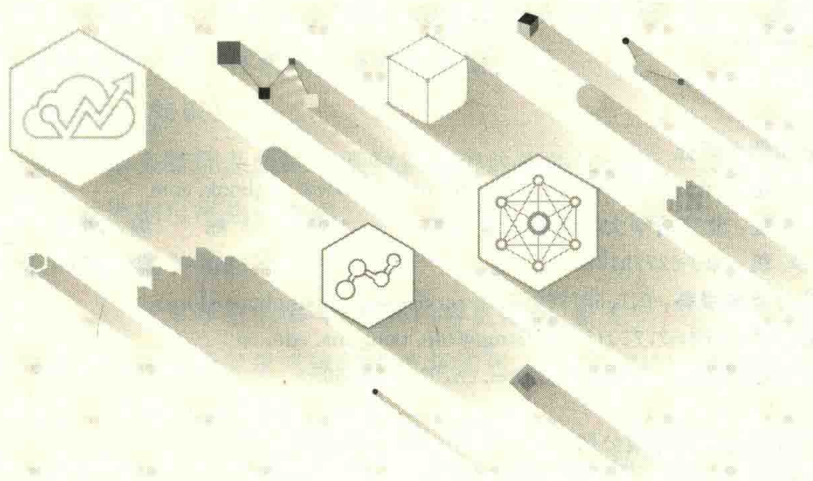




大数据技术与应用专业规划教材

商务智能与 数据挖掘 (第2版)

◎ 蔡晓妍 杨黎斌 张晓婷 李梅 编著



清华大学出版社

北京

内 容 简 介

商务智能与数据挖掘是近年来企业信息化的热点研究内容。本书共分为9章,包括商务智能概述、商务智能中的核心技术、数据挖掘概述、分类分析、关联分析、聚类分析、深度学习、Web挖掘技术、数据挖掘在电子商务中的应用等内容,它汇集了统计学、机器学习、数据库、人工智能等学科,具有多学科交叉以及技术与管理融合等特点。

本书内容全面,案例丰富,适合作为计算机应用、软件工程、信息管理、电子商务和管理科学等相关专业的本科生和研究生教材,也可作为企事业单位、政府部门、研究机构等从事商务智能理论研究工作的相关人员的参考资料。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

商务智能与数据挖掘/蔡晓妍等编著.—2版.—北京:清华大学出版社,2018(2018.3重印)
(大数据技术与应用专业规划教材)

ISBN 978-7-302-48631-2

I. ①商… II. ①蔡… III. ①电子商务—高等学校—教材 ②数据采集—高等学校—教材
IV. ①F713.36 ②TP274

中国版本图书馆CIP数据核字(2017)第261780号

责任编辑:郑寅堃 薛 阳

封面设计:刘 键

责任校对:梁 毅

责任印制:刘海龙

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦A座 邮 编:100084

社总机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

课件下载: <http://www.tup.com.cn>, 010-62795954

印 装 者:三河市铭诚印务有限公司

经 销:全国新华书店

开 本:185mm×260mm 印 张:13.25

字 数:321千字

版 次:2016年4月第1版 2018年3月第2版

印 次:2018年3月第1次印刷

印 数:1~1500

定 价:39.50元

产品编号:075936-01

当今社会信息技术飞速发展,经济全球化趋势日益明显,市场竞争激烈。生存在这样一个“信息爆炸”的时代,企业管理者能否利用信息进行快速而有效的决策已直接关系到企业的生死存亡。越来越多的企业提出对商务智能的需求,商务智能的出现和飞速发展已成为必然趋势。但随着商务智能在企业活动中的不断进行,海量的、具有潜在价值知识的信息便混同一般信息积累在企业商务智能系统中,尽管企业希望能够提取出有价值的信息,应用各种技术进行数据分析,以期将分析结果用于科学研究、商业决策或企业管理决策等方面,但目前市场数据分析工具种类繁多,若选择不当就很难对数据进行深层次的处理,无法真正提炼出信息的价值,使得企业管理层只能望“数”兴叹。除此之外,电子商务数据类型繁多、模型复杂,以及应用系统的分布性和异构性都会增加这些数据的分析难度,由此数据挖掘技术应运而生。

本教材自2016年4月出版,经过一年多的使用,根据学生和教师的反馈,编者对教材中的相关内容进行了改动。第2版与第1版相比增加了新的知识,删除了一些陈旧或使用较少的知识,使其更加简练。具体体现在以下几个方面。

(1) 增加了新的技术知识。

- ① 增加了第1章商务智能系统框架及数据流程的相关知识。
- ② 增加了第2章数据可视化的内容。
- ③ 增加了第7章深度学习的内容。

(2) 对原第4、5、10章重新组织编写,使其内容更贴切、更充实。

① 第4章对决策树和支持向量机进行了重新描述,去掉了BP神经网络算法,增加了粗糙集的阐述。

② 第5章对Apriori算法、Apriori改进算法以及FP增长算法进行了重新描述。

③ 去掉了原第7章。

④ 原第10章改为现第9章,重新举例并深入阐述了数据挖掘在电子商务中的典型应用。

本书各章编写分工如下:蔡晓妍编写第1、3章;杨丽丽编写第2章;朱珊珊编写第4章;李梅编写第5章;梁春泉编写第6章;杨黎斌编写第7、9章;张晓婷编写第8章。蔡晓

妍负责全书的策划、大纲的制定和统筹工作。

本书在编写过程中,参考了一些优秀教材和论文,在此对所有被引用文献的原作者表示衷心的感谢。研究生郭蓝天、何健、张野和吕鑫分别对本书进行了校对并提出了宝贵的建议,特在此表示感谢。由于编者水平所限,书中如有不妥之处,欢迎读者批评指正。

编者
2017年8月

| | |
|-------------------------------|----|
| 第 1 章 商务智能概述 | 1 |
| 1.1 商务智能的概念 | 1 |
| 1.1.1 数据、信息与知识 | 1 |
| 1.1.2 商务智能的定义..... | 3 |
| 1.1.3 商务智能的特点..... | 4 |
| 1.1.4 商务智能的过程..... | 5 |
| 1.2 商务智能的产生与发展 | 5 |
| 1.2.1 商务智能的产生和发展过程..... | 5 |
| 1.2.2 商务智能与其他系统的关系..... | 6 |
| 1.3 商务智能的研究内容 | 9 |
| 1.4 商务智能系统的支撑技术..... | 10 |
| 1.5 商务智能系统框架及数据流程..... | 11 |
| 1.5.1 商务智能系统框架 | 12 |
| 1.5.2 商务智能系统的数据流程 | 13 |
| 1.6 主流商务智能产品..... | 14 |
| 1.6.1 主流商务智能产品简介 | 14 |
| 1.6.2 商务智能的抉择 | 15 |
| 1.7 商务智能的未来发展趋势..... | 16 |
| 1.8 商务智能的应用..... | 17 |
| 小结 | 19 |
| 习题 | 20 |
| 第 2 章 商务智能中的核心技术 | 21 |
| 2.1 数据仓库..... | 22 |
| 2.1.1 数据仓库的产生与发展 | 22 |
| 2.1.2 数据仓库的概念与特征 | 22 |
| 2.1.3 ETL | 23 |
| 2.1.4 数据集市 | 25 |
| 2.1.5 数据仓库的数据组织 | 26 |

| | | |
|--------------|-----------------------|-----------|
| 2.1.6 | 数据仓库的体系结构 | 26 |
| 2.1.7 | 数据仓库的开发步骤 | 27 |
| 2.2 | 在线分析处理 | 30 |
| 2.2.1 | OLAP 简介 | 30 |
| 2.2.2 | OLAP 的定义和相关概念 | 31 |
| 2.2.3 | OLAP 与 OLTP 的区别 | 32 |
| 2.2.4 | OLAP 的分类 | 33 |
| 2.2.5 | OLAP 多维数据分析 | 34 |
| 2.3 | 数据可视化 | 36 |
| 2.3.1 | 什么是数据可视化 | 36 |
| 2.3.2 | 数据可视化的优势 | 37 |
| 2.3.3 | 数据可视化工具 | 37 |
| 小结 | | 41 |
| 习题 | | 41 |
| 第 3 章 | 数据挖掘概述 | 42 |
| 3.1 | 数据挖掘的起源与发展 | 42 |
| 3.1.1 | 数据挖掘的起源 | 42 |
| 3.1.2 | 数据挖掘的发展 | 43 |
| 3.2 | 数据挖掘所要解决的问题 | 45 |
| 3.3 | 数据挖掘的定义 | 46 |
| 3.4 | 数据挖掘的过程 | 46 |
| 3.5 | 数据挖掘系统 | 49 |
| 3.5.1 | 数据挖掘系统的分类 | 49 |
| 3.5.2 | 数据挖掘系统的发展 | 50 |
| 3.6 | 数据挖掘的功能和方法 | 51 |
| 3.6.1 | 数据挖掘的功能 | 51 |
| 3.6.2 | 数据挖掘的方法 | 53 |
| 3.7 | 数据挖掘的典型应用领域 | 55 |
| 3.8 | 数据挖掘的发展趋势 | 57 |
| 小结 | | 59 |
| 习题 | | 59 |
| 第 4 章 | 分类分析 | 60 |
| 4.1 | 预备知识 | 60 |
| 4.2 | 解决分类问题的一般方法 | 62 |
| 4.3 | 分类算法 | 63 |
| 4.3.1 | 贝叶斯分类器 | 63 |
| 4.3.2 | 贝叶斯信念网络 | 68 |

| | | |
|--------------|--------------------|------------|
| 4.3.3 | 决策树 | 72 |
| 4.3.4 | 支持向量机 | 77 |
| 4.3.5 | 粗糙集 | 80 |
| 4.3.6 | 其他分类算法 | 83 |
| 4.4 | 评估分类器的性能 | 84 |
| 4.4.1 | 保持方法 | 84 |
| 4.4.2 | 随机二次抽样 | 85 |
| 4.4.3 | 交叉验证 | 85 |
| 4.4.4 | 自助法 | 85 |
| 小结 | | 86 |
| 习题 | | 86 |
| 第 5 章 | 关联分析 | 89 |
| 5.1 | 引言 | 89 |
| 5.2 | 基本概念 | 91 |
| 5.3 | 关联规则的种类 | 93 |
| 5.4 | 关联规则的研究现状 | 94 |
| 5.5 | 关联规则挖掘算法 | 95 |
| 5.5.1 | Apriori 算法 | 95 |
| 5.5.2 | Apriori 改进算法 | 100 |
| 5.5.3 | FP 增长算法 | 102 |
| 5.6 | 改善关联规则挖掘质量问题 | 104 |
| 5.6.1 | 用户主观层面 | 104 |
| 5.6.2 | 系统客观层面 | 105 |
| 5.7 | 约束数据挖掘问题 | 105 |
| 小结 | | 106 |
| 习题 | | 107 |
| 第 6 章 | 聚类分析 | 109 |
| 6.1 | 聚类的概念 | 109 |
| 6.1.1 | 聚类概念及应用 | 109 |
| 6.1.2 | 聚类算法要求 | 110 |
| 6.1.3 | 聚类技术类型划分 | 111 |
| 6.2 | 聚类分析的统计量 | 113 |
| 6.2.1 | 模型定义 | 113 |
| 6.2.2 | 相似性度量 | 114 |
| 6.3 | 常用聚类算法 | 118 |
| 6.3.1 | k 均值算法 | 118 |
| 6.3.2 | k-medoids 算法 | 120 |

| | | |
|------------|-----------------------|------------|
| 6.3.3 | 凝聚层次聚类算法 | 122 |
| 6.3.4 | DBSCAN 算法 | 126 |
| 6.3.5 | STING 算法 | 129 |
| 6.3.6 | CLIQUE 算法 | 130 |
| 6.4 | 簇评估 | 131 |
| 6.4.1 | 概述 | 132 |
| 6.4.2 | 非监督簇评估:使用凝聚度和分离度 | 132 |
| 6.4.3 | 非监督簇评估:使用邻近度矩阵 | 136 |
| 6.4.4 | 层次聚类的非监督评估 | 137 |
| 6.4.5 | 确定正确的簇个数 | 137 |
| 6.4.6 | 聚类趋势 | 138 |
| 6.4.7 | 簇有效性的监督度量 | 138 |
| 6.5 | 聚类与分类比较 | 141 |
| | 小结 | 141 |
| | 习题 | 141 |
| 第7章 | 深度学习 | 143 |
| 7.1 | 深度学习的由来 | 144 |
| 7.1.1 | 深度学习的神经学启示 | 144 |
| 7.1.2 | 浅层结构函数表示能力的局限性 | 144 |
| 7.1.3 | 特征提取的需要 | 145 |
| 7.2 | 深度学习的经典方法 | 146 |
| 7.2.1 | 深度学习表示模型和网络结构 | 146 |
| 7.2.2 | 自动编码器 | 146 |
| 7.2.3 | 受限玻尔兹曼机 | 148 |
| 7.2.4 | 卷积神经网络 | 149 |
| 7.3 | 深度学习的应用 | 150 |
| 7.3.1 | 深度学习在语音识别、合成及机器翻译中的应用 | 150 |
| 7.3.2 | 深度学习在图像分类及识别中的应用 | 151 |
| 7.3.3 | 深度学习在视频分类及行为识别中的应用 | 152 |
| 7.4 | 深度学习的研究近况及未来研究方向 | 153 |
| 7.4.1 | 研究近况 | 153 |
| 7.4.2 | 未来研究方向 | 153 |
| | 小结 | 155 |
| | 习题 | 156 |
| 第8章 | Web 挖掘技术 | 157 |
| 8.1 | Web 数据挖掘概述 | 157 |
| 8.1.1 | Web 数据挖掘的概念 | 157 |

| | | |
|--------------|-----------------------------|------------|
| 8.1.2 | Web 数据挖掘的特点 | 158 |
| 8.1.3 | Web 数据挖掘的处理流程 | 159 |
| 8.1.4 | Web 数据挖掘与信息检索、信息抽取的区别 | 159 |
| 8.2 | Web 数据挖掘分类 | 160 |
| 8.2.1 | Web 内容挖掘概述 | 160 |
| 8.2.2 | Web 结构挖掘概述 | 161 |
| 8.2.3 | Web 使用挖掘概述 | 161 |
| 8.3 | Web 内容挖掘 | 162 |
| 8.3.1 | 特征提取和特征表示 | 162 |
| 8.3.2 | 自动摘要 | 163 |
| 8.3.3 | 文本分类 | 163 |
| 8.3.4 | 文本聚类 | 164 |
| 8.4 | Web 结构挖掘 | 164 |
| 8.4.1 | 超链和页面内容的关系 | 165 |
| 8.4.2 | 不同挖掘阶段的分析 | 165 |
| 8.4.3 | PageRank | 166 |
| 8.4.4 | HITS | 172 |
| 8.4.5 | 两种算法的比较 | 176 |
| 8.4.6 | Web 结构挖掘应用 | 177 |
| 8.5 | Web 使用挖掘 | 178 |
| 8.5.1 | Web 使用挖掘数据预处理 | 178 |
| 8.5.2 | Web 使用挖掘模式发现 | 181 |
| 8.5.3 | Web 使用挖掘模式分析 | 183 |
| 8.5.4 | Web 使用挖掘模式应用 | 184 |
| | 小结 | 185 |
| | 习题 | 185 |
| 第 9 章 | 数据挖掘在电子商务中的应用 | 187 |
| 9.1 | 网站结构优化 | 187 |
| 9.2 | 智能搜索引擎 | 189 |
| 9.2.1 | 网络机器人 | 189 |
| 9.2.2 | 文本分析 | 190 |
| 9.2.3 | 搜索条件的获取和分析 | 191 |
| 9.2.4 | 信息的搜索和排序 | 191 |
| 9.3 | 移动商务智能 | 192 |
| 9.4 | 客户关系管理 | 193 |
| 9.4.1 | 营销 | 193 |
| 9.4.2 | 销售 | 194 |
| 9.4.3 | 客户服务 | 194 |

| | | |
|-------|-----------------|-----|
| 9.4.4 | 客户保持 | 194 |
| 9.4.5 | 风险评估和欺诈识别 | 195 |
| 9.5 | 客户分类 | 195 |
| 9.5.1 | 传统的客户分类理论 | 196 |
| 9.5.2 | 基于客户行为的客户分类 | 196 |
| 9.5.3 | 基于客户生命周期的客户分类 | 196 |
| 9.5.4 | 基于客户生命周期价值的客户分类 | 196 |
| | 小结 | 197 |
| | 习题 | 197 |
| | 参考文献 | 198 |

第 1 章

商务智能概述

随着世界经济全球化的迅猛发展,生产国际化的趋势不断加强,企业必须能够在瞬息万变的环境下及时做出反应。为了迎接市场的挑战,企业需要对市场有准确的把握,分析顾客的消费趋势,找出企业经营中出现的问题,加强与供应链合作伙伴的关系,挖掘新的商业机会,并能够对未来进行预测。随着企业信息化的发展,各种应用系统产生的数据量平均 18 个月就翻一番。如何充分利用这些数据资产,挖掘出决策者需要的信息,做出高质量的决策是企业管理者需要考虑的问题。近年来,数据集成、数据分析、大容量数据存储与并行处理等技术不断成熟,成本不断下降,企业各种应用软件积累了大量的数据。这些因素促进了商务智能的发展。商务智能(Business Intelligence, BI)可以将各种数据及时地转换为支持决策的信息和知识,帮助企业管理者了解顾客的需求与消费习惯,预测市场的变化趋势以及行业的整体发展方向,进行有效的决策,从而在竞争中占据有利地位。

1.1 商务智能的概念

商务智能越来越受到学术界和产业界的青睐,逐渐成为目前国内外企业界和软件开发界备受关注的研究热点。作为一项新兴的技术,在过去的十多年间,围绕商务智能的理论、方法、技术等的应用已经取得了许多令人瞩目的成就。

目前,不少企业积累的海量数据不仅没能给企业带来财富,相反却使得企业淹没于数据之中,形成一个个信息孤岛和数据坟墓。企业面临着由于数据库变得越来越庞大而带来的对数据管理的困难。如何充分利用这些数据,为企业的经营决策服务?这就需要一种合适的数据处理和数据分析工具。

1.1.1 数据、信息与知识

在信息时代,数据是宝贵的财富,但只有充分利用这种财富,识别信息,获取知识,辅助

商业决策才能从中获得价值。

1. 数据

数据是用来记录、描述和识别事物的按一定规则排列组合的物理符号,是一组表述数量、行动和目标的非随机的可鉴别的符号,是客观事物的属性、数量、位置及其相互关系等的抽象表示,它以适合于用人工或自然的方式进行保存、传递和处理。它既可以是数字、文字、图形、图像、声音或者味道,也可以是计算机代码。在计算机科学中,数据是指所有能输入到计算机中具有一定意义的数字、字母、符号和模拟量等并能够被计算机程序处理的符号介质的总称,是计算机能够识别的二进制数的形式。

数据本身是孤立的、互不关联的客观事实、文字、数字和符号,没有上下文和解释。数据表达的仅仅是一个描述,如 20140112,我们只知道这是一个数字,或者可以看作是日期,对于这个数字来说,它就是数字,不表示别的任何含义。

数据是用属性描述的,属性也称变量、特征、字段或维。数据经过处理仍然是数据,只有经过解释,数据才有意义,才能成为信息。

2. 信息

信息是指人们对数据进行系统的收集、整理、管理和分析的结果,是经过一系列的提炼、加工和集成后的数据。信息是对客观世界各种事物特征的反映。数据是信息的符号表示,或称载体,数据不经加工只是一种原始材料,其价值只是在于记录了客观数据的事实。信息是数据的内涵,是对数据的解释。信息可以是完整的,也可以是片段的;可以是关于过去的,或者关于现在的,也可以是涉及未来的。目前天气很热,气温高达 35℃,这条信息描述的是现在的天气状况。参考过去连续三年的气温记录,每年这一天的历史温度都高于 37℃,这是关于过去的信息。如果根据这两天信息预测明天的气温至少为 37℃,那么这是涉及未来的信息。尽管明天高温天气是有可能的,甚至是必然的,但这种预测未来的信息多少会带有不确定性,为了减少不确定性,提高置信度,必须对信息进行提炼、加工和集成。

3. 知识

所谓知识,就它反映的内容而言,是客观事物的属性与联系的反映,是客观世界在人脑中相对正确的反映。就它反映的活动形式而言,有时表现为主体对事物的感性直觉或表象,属于感性知识,有时表现为关于事物的概念或规律,属于理性知识。知识是在实践活动中获得的关于世界的最本质的认识,是对信息的提炼、比较、挖掘、分析、概括、判断和推论。

一般而言,知识具有共享性、传递性、非损耗性(可以反复使用,其价值不会减小)及再生性等特点。

按知识的复杂性可将知识划分为显性(Explicit)知识和隐性(Tacit)知识,它是知识最基本和最重要的划分结构。显性知识是用系统、正式的语言传递的知识,可以编码和度量,可以清晰地表达出来,易于传播,可以在人与人之间进行直接的交流,通常以语言文字(如书籍、文件、网页、电子邮件等)形式存在。显性知识的处理可以用计算机实现。隐性知识是存在于人脑中的、非结构化的、与特定语境相关的知识,很难编码和度量。隐性知识是人们在实践中不断摸索和反复体验形成的,通常以直觉、价值观、推断、经验、技能等形式表现出来。它难以表述,但却是个人的直接表现且更为宝贵。隐性知识的处理只能通过人脑实现,一般要通过言传身教和师徒等形式传播。

数据、信息和知识之间的关系为：从数据中提取信息，从信息中挖掘知识，如图 1-1 所示。



图 1-1 数据、信息和知识的区别

1.1.2 商务智能的定义

1989年，美国加特纳公司的分析师 Howard Dresner 首次提出“商务智能”的概念。商务智能可以定义为一组数学模型和分析方法，它们系统地开发可以利用的数据，以便检索对支持复杂决策过程有用的信息和知识。出于帮助企业所有者打造更好的商业的目的，商务智能是涉及收集、提供、存取及分析数据的一大类别的应用与技术。

商务智能从产生以来一直发展较快，但目前还不成熟，企业界和学术界对商务智能存在着或多或少不同的理解，这里分别列举几个比较全面且较新的定义。

IBM 对 BI 的定义：“商务智能是一系列技术支持的简化信息收集、分析的策略集合。通过使用企业的数据资产来制定更好的商务决策。企业的决策人员以数据仓库为基础，通过各种查询分析工具、联机分析处理或者是数据挖掘，加上决策人员的行业知识，从数据仓库中获得有利的信息，进而帮助企业提高利润，增加生产力和竞争力。”

Business Object 公司对 BI 的定义：“商务智能是一种基于大量数据的信息提炼的过程，这个过程与知识共享和知识创造紧密结合，完成了从信息到知识的转变，最终为商家提供网络时代的竞争优势和实实在在的利润。”

Microsoft 公司对 BI 的定义：“商务智能是任何尝试获取、分析企业数据以便更清楚地了解市场和顾客，改进企业流程，更有效地参与竞争的过程。”

IDC 公司对 BI 的定义：“商务智能是下列软件工具的集合：终端用户查询和报告工具、在线分析处理工具、数据挖掘软件、数据集市、数据仓库产品和主管信息系统。”

Oracle 公司对 BI 的定义：“商务智能是一种商务战略，能够持续不断地对企业经营理念、组织机构和业务流程进行重组，实现以顾客为中心的自动化管理。”

Data Warehouse Institute 对 BI 的定义：“商务智能是把数据转换成知识并把知识应用到商业运营的一个过程。”

商务智能专家王茁在总结了商务智能的众多版本之后给商务智能下的定义：“商务智能是企业利用现代信息技术收集、管理和分析结构化和非结构化的商务数据和信息，创造和积累商务知识和见解，改善商务决策水平，采取有效的商务行动，完善各种商务流程，提升各

方面商务绩效,增强综合竞争力的智慧和能力。”

利用现代信息技术——这是这一定义中的关键之一。现代信息技术的发展催生了信息经济和信息社会,在这一新型的经济和社会形态中,信息的爆炸式增长又产生了对能够处理和
控制信息的技术的强烈需求,商务智能正是新的信息技术在商务分析中的有效应用。

总结上述观点,商务智能是融合了先进信息技术与创新管理理念的结合体,它集成了企业内外的数据,进行加工并从中提取能够创造商业价值的信息,面向企业战略并服务于管理层、业务层,指导企业经营决策,提升企业竞争力,涉及企业战略、管理思想、业务整合和技术体系等层面,促进信息到知识再到利润的转化,从而实现更好的绩效。事实上,商务智能应用的核心不在其功能,而在于对业务的优化,IBM公司更强调数据集成和数据分析基础上的业务分析和优化(Business Analytics and Optimization, BAO)。目前,商务智能的应用已延伸到了非商业领域,政府和教育部门等也成为了商务智能的应用领域。

1.1.3 商务智能的特点

商务智能具有以下主要特点,了解这些特点有助于更好地理解商务智能的内涵。

1. 商务智能服务企业战略

商务智能能够对企业的内外部数据进行分析,支持企业战略管理。哈佛商学院的迈克尔·波特博士在 *Harvard Business Review* 一文中把战略分为三个方面:定位、取舍和配称(各项运营活动之间如何关联),而商务智能可以通过数据分析帮助企业对这些方面进行规划。

2. 商务智能提升企业绩效

商务智能更多地是用来解决管理问题。通过商务智能能从企业多年运营的数据中,挖掘有效的模式辅助管理决策。随着商务智能应用的发展,商务智能离业务越来越近。商务智能在企业绩效管理中扮演着重要的角色,而商务智能相关的产品在管理角色和方法、管理职能和过程等方面烙印渐深,并且融合了越来越多的企业管理的理念。例如,Business Objects(SAP)在2007年“商业智能 点亮明天”商务智能解决方案研讨会展示的绩效管理套件中包含管理仪表盘、计分卡等工具。这些工具不再是色彩和图形的结合,而是包含大量的业务逻辑关系和线性规划的运算模型。

3. 商务智能是“数据炼油厂”

商务智能可看作“数据炼油厂”,根据业务需要收集数据,并进行提炼和加工,最终产生对企业有价值的知识,提高企业的绩效。商务智能需要整合企业的业务系统数据,从而保证足够的“原料补给”。商务智能对ERP、CRM和SCM等业务系统中生成的运营数据进行分析,并给出报告,帮助管理者认识企业和市场的现状,预测发展趋势,做出正确的决策。

4. 商务智能是多项技术的综合应用

随着信息化的发展,商务智能已成为企业充分利用数据资产的重要方法。它从不同的数据源中提取有用的数据,通过数据仓库、在线分析处理和数据挖掘等技术实现企业的决策、考核、分析有机结合和量化以达到为企业提供经营管理、决策支持的目的。最新的商务智能还涉及其他一些新技术,例如,内存中的分析处理、面向服务的软件架构(Service

Oriented Architecture, SOA)、文本挖掘和元数据存储等。商务智能在这些技术的支持下,发现数据背后隐藏的商机或威胁,洞察企业和市场的现状,把握趋势,识别异常情况,理解企业业务是推动力量,认清正在对企业的业务产生影响的行为及影响程度。

5. 商务智能用户的多样性

商务智能服务于各类企业决策者。传统应用中,商务智能主要支持中、高层管理人员决策。目前,商务智能平台的用户包括一线业务人员、各级管理者,甚至外部的顾客和商业伙伴。这是因为业务经营决策的范围发生了扩展,包括操作层、战术层和战略层的决策。

1.1.4 商务智能的过程

商务智能的过程是:首先需要准备正确可用的数据,其次要将这些数据转化为有价值的信息,再用于指导商业实践(智慧)。该过程就包括数据抽取、分析和挖掘三个主要环节,分别由 DW、OLAP、DM 技术来完成。

DW 是商务智能的基础和核心,存储按照商务智能的要求重新组织的、来自业务系统的数据。

OLAP 和 DM 在 DW 的基础上进行分析,提供给最终用户灵活自主的信息访问途径、丰富的数据分析和报表功能。

从系统的观点来看,商务智能中信息处理的过程可以归结为以下三个子过程:数据获取和管理、数据整理和分析、信息展现和表示。

(1) 数据获取和管理:从不同的数据源获取有用的数据,对数据进行清理以保证数据的正确性,将数据经过转换、重构后存入数据仓库或数据集市(这时数据变为信息)。

(2) 数据整理和分析:通过合适的查询和分析工具、数据挖掘工具,对信息进行处理(这时信息变为辅助决策的知识)。

(3) 信息展现和表示:将管理与决策所需要的知识呈现于用户面前,支持管理与决策。

1.2 商务智能的产生与发展

商务智能的概念最早是由美国加特纳公司的 Howard Dresner 于 1996 年提出来的,对商务智能所涉及的一系列的概念和方法进行了详细的描述,通过应用基于数据分析的支持系统来辅助商业决策的制定。BI 技术提供帮助企业迅速分析数据的技术和方法,包括收集、管理和分析数据,将这些数据转化为有用的信息,然后分发到企业各处。近年来,商务智能技术日趋成熟,越来越多的企业决策者意识到需要商务智能来保持和提升企业竞争力。在美国,500 强企业中已经有 90% 以上的企业利用企业管理和商务智能软件帮助管理者做出决策。国外已经有很多成功实施商务智能的案例。目前我国的商务智能尚处于起步阶段,无论是商务智能应用的程度还是其实际效果与国外企业都有很大差距。近年来,国内外商务智能供应商和高等院校都开展了广泛的商务智能的基础研究和应用研究。

1.2.1 商务智能的产生和发展过程

商务智能是随着 Internet 的高速发展和企业信息化的不断深入而产生的。其发展也是

一个渐进的、复杂的演变过程,而且目前仍然处于发展之中。它经历了事物处理系统(Transaction Processing System, TPS)、高级管理人员信息系统(Executive Information System, EIS)、管理信息系统(Management Information System, MIS)和决策支持系统(Decision Support System, DSS)等几个不同阶段,最终演变成今天的企业商务智能系统(BIS)。它是一个可包含企业所有知识的系统,服务于管理决策层或部门执行经理,帮助其进行分析和决策。

TPS 是以计算机处理代替某些手工操作,这个时期主要是借助计算机的运算能力将人力从大量的计算和重复性的工作中解脱出来。TPS 的优点:效率提高,耗用时间缩短;缺点:完全“就事论事”,只完成了信息的“一次利用”,舍弃了信息的更高价值的利用。TPS 主要是在服务行业中直接针对广大顾客的某种专门性事物处理要求而提供的一种服务性系统,因此,TPS 的设计目标主要是方便顾客,为顾客提供高质量、高效率以及安全、可靠的各种专门性事物服务。

EIS 指为了满足无法专注于计算机技术的领导人员的信息查询需求,而特意制定的以简单的图形界面访问数据仓库的一种应用。EIS 根据预先定义的查询,以报表或图标的形式向使用者提供商业活动情况的相关数据。这些信息通常是一定时期内的总销售额、每种产品的销售额、销售数量等。EIS 的优点:使决策者在一定程度上掌握企业的业务运营状况,不至于完全靠“拍脑袋”决策;EIS 的缺点:其应用面太窄,仅限于高、中层管理人员的管理活动,依然得不到信息支持。

MIS 的出现部分地解决了 EIS 的问题,它面向所有的管理人员,覆盖了企业所有的业务内容,能够帮助管理人员了解企业的日常业务,并进行高效的控制、组织和计划。MIS 的优点:能处理日常事务,尤其对中、低层管理人员更为有效;缺点:对于高层决策者而言,无法从全局的、战略的高度给予很大的支持。

DSS 是辅助决策者通过数据、模型和知识,以人机交互方式进行半结构化或非结构化决策的计算机应用系统。它是管理信息系统向更高一级发展而产生的先进信息管理系统。它为决策者提供分析问题、建立模型、模拟决策过程和方案的环境,调用各种信息资源和分析工具,帮助决策者提高决策水平和质量。

结构化决策是指对某一决策过程的环境及规则,能用确定的模型或语言描述,以适当的算法产生决策方案,并能从多种方案中选择最优解的决策。

非结构化决策是指决策过程复杂,不可能用确定的模型和语言来描述其决策过程,更无所谓最优解的决策。

半结构化决策是介于以上两者之间的决策,这类决策可以建立适当的算法产生决策方案,使决策方案得到较优的解。

非结构化和半结构化决策一般用于一个组织的中、高层管理,其决策者一方面需要根据经验进行分析判断,另一方面也需要借助计算机为决策提供各种服务信息,及时做出正确有效的决策。

BIS、DSS、MIS、EIS 和 TPS 之间的区别如图 1-2 所示。

1.2.2 商务智能与其他系统的关系

商务智能作为一种企业信息集成解决方案,是一个分析型系统。一方面,它为企业不同