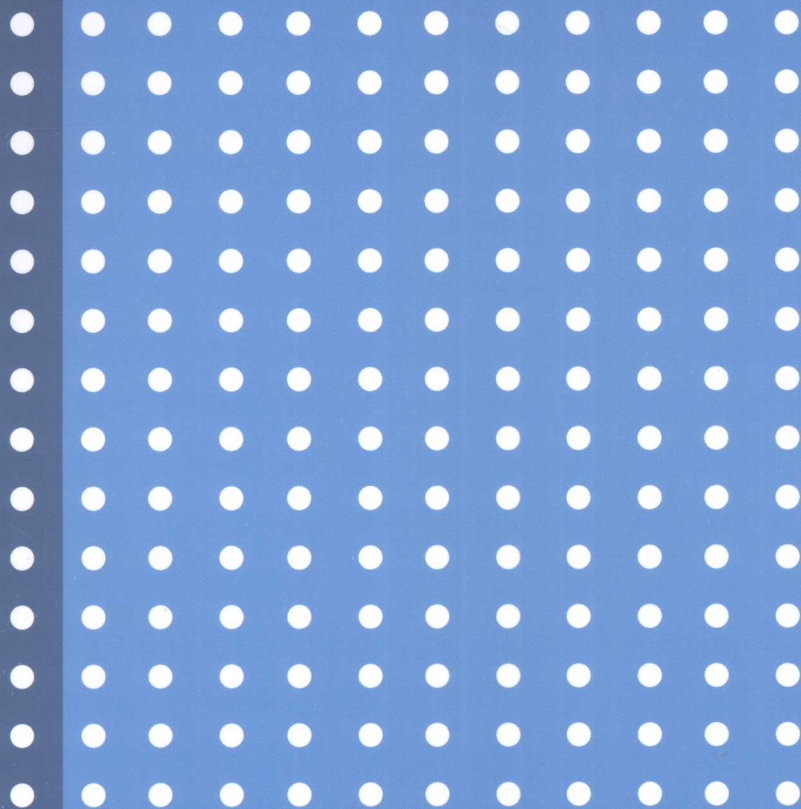


商务智能

赵卫东 主编

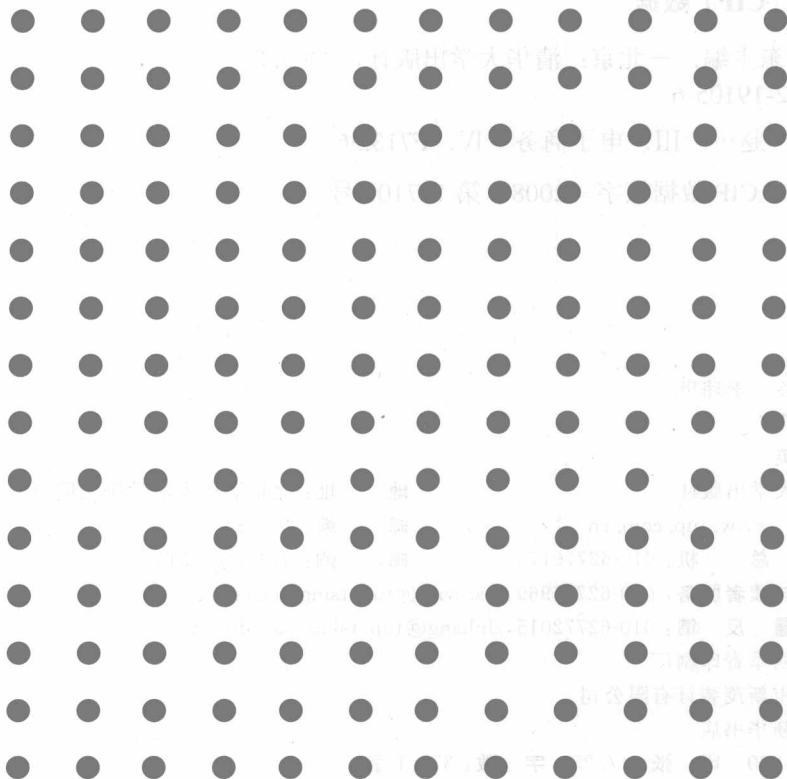


清华大学出版社



商务智能

赵卫东 主编



清华大学出版社
北京

内 容 简 介

商务智能是近年来企业信息化的热点,有着广阔的应用前景。本书首先介绍了商务智能的基本概念、商务智能系统的架构以及数据仓库、OLAP、数据挖掘等核心技术。在此基础上,讨论了商务智能在电子商务、移动商务、知识管理、Web 挖掘、企业绩效管理和流程管理等领域的应用。此外,还分析了商务智能在国内外的发展趋势。

本书内容比较新颖、全面,案例丰富,适合计算机应用、软件工程、信息管理、电子商务和管理科学等相关专业本科生和研究生的教材,也可作为从事商务智能的信息化人员的参考资料。

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/赵卫东主编. —北京:清华大学出版社,2009.3
ISBN 978-7-302-19105-6

I. 商… II. 赵… III. 电子商务 IV. F713.36

中国版本图书馆CIP数据核字(2008)第197103号

责任编辑:丁 岭 李玮琪

责任校对:李建庄

责任印制:李红英

出版发行:清华大学出版社

地 址:北京清华大学学研大厦A座

<http://www.tup.com.cn>

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

印 刷 者:北京四季青印刷厂

装 订 者:三河市新茂装订有限公司

经 销:全国新华书店

开 本:185×230 印 张:17.25 字 数:371千字

版 次:2009年3月第1版 印 次:2009年3月第1次印刷

印 数:1~3000

定 价:28.00元

本书如存在文字不清、漏印、缺页、倒页、脱页等印装质量问题,请与清华大学出版社出版部联系
退换。联系电话:(010)62770177 转 3103 产品编号:031078-01

序

在日益激烈的竞争环境下，商务智能已经成为很多企业提高竞争力的重要手段。简单地讲，商务智能是利用先进的数据库存储、数据集成、数据分析技术，把海量的业务数据转化成有价值的信息和知识，辅助管理者进行决策。因此，商务智能在市场营销、客户关系管理、风险管理、绩效管理等重要的商业应用领域中发挥着越来越大的作用，已被银行、电信、保险、证券、零售、制造、社保等企业和政府部门广泛采用。

商务智能技术与应用在中国起步较晚，但发展迅速。1999年4月第三届亚太地区知识发现与数据挖掘国际会议（PAKDD）在北京召开，我国的学术界迅速掀起商务智能与数据挖掘的研究与应用热潮。国家863计划和自然科学基金多次为商务智能与数据挖掘立项，具有自主知识产权的商务智能技术和软件产品不断出现。商务智能的应用近年来也有了很大进展，企业界已经逐步认识到商务智能对提高企业竞争力的重要性，特别是电信、银行、保险、税务等信息化水平较高的行业，在完成数据集中整合后都把商务智能作为新的应用重点，从而带动了商务智能技术和水平的提高。

目前，中国的商务智能技术和应用市场正快速增长。商务智能的应用正向制造业、电子商务、物流和政府机构扩展，各行各业对商务智能技术和人才的需求也在不断扩大。很多大学的计算机学院、软件学院和商学院等都开设了商务智能相关课程，培养各类商务智能技术和应用人才，以满足商务智能人才市场的需求。

商务智能涉及的内容很多，既包括很多技术，例如ETL、OLAP、数据仓库和数据挖掘，又包括很多行业应用，如市场营销、客户关系管理、风险管理、绩效管理等。目前有关数据仓库、OLAP和数据挖掘的专业书籍很多，但多来自国外，而且缺少全面、系统地介绍商业智能技术和应用的教材。

赵卫东博士编写的《商务智能》教材，从商务智能的基本概念和应用开篇，全面介绍了商务智能的核心技术，包括商务智能系统架构、数据仓库、OLAP和数据挖掘。然后阐述了商务智能在移动商务、知识管理、Web挖掘、绩效管理、电子商务、业务流程优化等领域的应用。最后展望了商务智能在未来一段时间国内外发展趋势。

这是一本技术与应用相辅相成，内容广泛，案例丰富，深入浅出的书籍。既可作为研究生和本科生的教材，也可作为从事商务智能技术开发和应用的企业人员参考书。

黄哲学

2008年12月于香港大学

前言



从 20 世纪 80 年代以来，企业进入了信息时代，市场全球化，顾客需求多样化、个性化，变化频率加快，竞争范围和激烈程度逐渐加大和加剧。在这种快鱼吃慢鱼的商业环境中，企业为了生存，就必须迅速反应，实施管理信息化和决策智能化。Internet、各种管理应用系统的广泛使用，为企业打通了数字神经，减少了企业运营成本，提高了企业的效率。另一方面，企业在提高效率的同时，也要考虑其本身的效益，这就要求企业决策者及时掌握运营过程中的各种信息、知识，而不是拍脑袋解决问题。信息、知识已成为企业最基本、最重要的生产要素，而这些信息、知识是各种管理应用系统难以提供的。在信息化提高企业竞争力的同时，各种管理应用系统也积累大量的数据。这些数据是企业的重要资产，其中蕴涵了许多有价值的信息、知识。事实上，日益积累的数据利用率还相当低，如何从中充分挖掘有价值的信息和知识，提高管理人员的决策水平，满足不同层次、不同部门和行业应用的需求，已成为业界和学术界关注的问题。

新经济时代的赢家是那些把顾客、供应商等相关的运营数据整合、分析和共享，转化为信息，并进一步分析得到知识，提高企业智能从而保持赢利的企业。面对激烈的竞争，传统的决策支持系统（Decision Support System, DSS）已难以支撑，而作为 ERP 应用之后的企业信息化亮点，商务智能，也称为商业智能为企业提供了这样的一种利器。2002 年 IDC 的研究表明，一些商务智能项目在一年多的时间就会获得 430% 的回报。商务智能具有传统 DSS、主管信息系统（Executive Information System, EIS）等不具备的强大数据管理、数据分析和知识发现的能力，已成为企业差异化的重要因素，它对改善商务决策水平，采取有效的商务行动，提升企业绩效是非常有效的，因此在竞争比较激烈、信息化基础比较扎实的一些行业，例如银行、电信、零售、保险和制造等受到了重视，已成为信息化领域继 ERP、CRM 和 SCM 等应用软件之后的新热点。近几年来，随着竞争的进一步加剧，越来越多的企业青睐商务智能，希望获得先机以抢占市场有利位置。欧美的国家在商务智能方面的投资逐年上涨，商务智能已形成一个产业。据 IDC 预测，商务智能软件在中国内地的年销售额平均增长至少为 65.6%。商务智能需求的增长，也促进了商务智能厂商不断进行技术创新，以抢占尚不成熟、处于高速发展阶段的商务智能市场。最近 Oracle、SAP 和 IBM 等公

司的并购案例,充分说明了未来商务智能市场的光明前途。

商务智能的成功需要人们对商务智能有一定的理解,目前熟悉商务智能的人才还很缺乏。从内容上来讲,商务智能包括数据仓库、在线分析处理、数据挖掘及其应用。目前国内在商务智能的教学才刚起步,尤其是教材基本还集中在数据挖掘方面,而难度适中、综合介绍商务智能的教材极少。针对这种情况,作者参阅了大量国内外最新的商务智能资料,编写了一本较全面反映商务智能的教材,并在复旦大学软件工程专业的研究生、本科生中多次使用。本书不是一本深入探讨商务智能技术的学术性书籍,而是对商务智能的基本问题进行系统的介绍,为读者对商务智能的深入学习打下基础。为增强读者的感性认识,本书的讨论配有很多实例,每章最后还附有主要的参考文献以供读者更深入地学习。

全书共 13 章,分为 5 个部分。第一部分是引言,讨论商务智能基础,主要介绍了商务智能的发展、概念、价值以及目前在一些领域的应用情况,使读者对商务智能有一个概要的认识。第二部分介绍商务智能的核心技术,由第 2~5 章组成,涉及商务智能系统架构、数据仓库、在线分析处理和数据挖掘等核心技术,这些内容是商务智能的理论基础,此部分内容不刻意介绍复杂的、扩展的数据分析算法,而是强调基本内容的应用。对于有一定基础的读者,可以在学习这些内容后选择专门介绍数据仓库和数据挖掘的书籍深入学习。第三部分由第 6~11 章组成,涉及商务智能最新的一些前沿问题,总结了商务智能在移动商务、商务智能与知识管理、Web 挖掘、企业绩效管理、电子商务和流程挖掘等领域的最新应用。此部分内容是提高性的,供有一定基础的读者阅读。第 12 章针对目前商务智能的进展,对国内外商务智能的前景进行了展望。最后,第 13 章介绍了香港大学电子商务科技研究所(ETI)开发的数据挖掘系统 AlphaMiner,可供学生实验时使用。

本书在编写过程中得到了香港大学黄哲学教授和 Business Objects (SAP) 公司鲁百年博士的指导,研究生范力、王安华、林涵溪、周佶平、曹焯和吴海峰等同学在案例分析等方面做了许多工作,特表示感谢。由于作者理论水平和实践经验有限,书中难免有不当和疏漏之处,望广大读者指正。

赵卫东

2008 年 12 月于复旦大学

目 录

第一部分 商务智能基础

第 1 章 商务智能概论	3
1.1 商业决策需要商务智能	3
1.1.1 数据、信息与知识	4
1.1.2 管理就是决策	5
1.1.3 决策需要信息和知识	5
1.1.4 智能型企业	5
1.1.5 商务智能支持商业决策	6
1.2 商务智能简介	7
1.2.1 商务智能概念	7
1.2.2 商务智能的发展	10
1.2.3 商务智能的价值	11
1.3 商务智能系统的功能	14
1.4 商务智能的应用	16
思考题	22
本章参考文献	22

第二部分 商务智能核心技术

第 2 章 商务智能系统架构	27
2.1 商务智能系统的组成	27
2.2 数据集成	29
思考题	31
本章参考文献	32
第 3 章 数据仓库	34
3.1 从数据库到数据仓库	34

3.2	数据仓库的概念	36
3.3	数据集市	38
3.4	元数据	39
3.5	ETL	42
3.6	操作数据存储	43
3.7	数据仓库模型	44
3.8	数据挖掘查询语言	48
3.9	医保数据仓库设计	50
	思考题	58
	本章参考文献	58
第4章	在线分析处理	59
4.1	OLAP 简介	59
4.2	OLAP 与 OLTP	62
4.3	OLAP 操作	62
4.4	OLAP 分类	66
4.5	OLAP 操作语言	69
4.6	流行的 OLAP 工具	72
	思考题	76
	本章参考文献	76
第5章	数据挖掘	78
5.1	数据挖掘基础	78
5.1.1	数据挖掘的概念	78
5.1.2	数据挖掘的发展	80
5.1.3	数据挖掘的过程	81
5.1.4	数据挖掘原语与语言	84
5.1.5	基于组件的数据挖掘	86
5.1.6	可视化技术	88
5.1.7	数据挖掘的隐私保护	91
5.2	数据挖掘的典型应用领域	92
5.3	数据预处理	93
5.4	聚类分析	99
5.4.1	聚类的概念	99
5.4.2	聚类分析的统计量	99
5.4.3	常用聚类算法	102
5.4.4	其他聚类方法	109

5.4.5 离群点检测	110
5.5 分类分析	111
5.5.1 贝叶斯分类器	112
5.5.2 决策树	115
5.5.3 支持向量机	123
5.5.4 BP 神经网络	128
5.5.5 其他分类方法	131
5.6 关联分析	133
5.6.1 关联规则	134
5.6.2 Apriori 算法	136
5.6.3 FP 增长算法	139
5.6.4 其他关联规则挖掘算法	141
5.7 序列模式挖掘	142
5.7.1 基本概念	142
5.7.2 类 Apriori 算法	143
5.8 回归分析	145
5.8.1 一元回归分析	146
5.8.2 多元线性回归分析	149
5.8.3 其他回归分析	153
5.9 时间序列分析	153
5.10 数据挖掘技术与应用的发展方向	155
思考题	157
本章参考文献	160
第三部分 商务智能应用	
第 6 章 移动商务智能	165
6.1 移动商务	165
6.2 商务智能在移动商务中的应用	166
思考题	173
本章参考文献	174
第 7 章 商务智能与知识管理	175
7.1 知识管理	175
7.2 知识管理与商务智能的关系	176
7.2.1 商务智能与知识管理的区别	176
7.2.2 商务智能与知识管理的联系	177

思考题	179
本章参考文献	179
第 8 章 Web 挖掘	181
8.1 Web 挖掘基础	181
8.2 Web 内容挖掘	183
8.3 Web 结构挖掘	188
8.4 Web 日志挖掘	192
思考题	196
本章参考文献	197
第 9 章 商务智能在企业绩效管理中的应用	199
9.1 企业绩效管理的层次	200
9.2 商务智能贯穿企业绩效管理的闭环流程	201
9.3 商务智能在企业绩效管理中的应用	203
9.4 商务智能给企业绩效管理带来的价值	206
9.5 企业绩效管理的主要工具	207
思考题	210
本章参考文献	210
第 10 章 数据挖掘在电子商务中的应用	211
10.1 电子商务需要数据挖掘	211
10.2 顾客管理	212
10.3 网站结构优化	215
10.4 智能搜索引擎	217
10.5 异常事件确定	219
思考题	220
本章参考文献	221
第 11 章 workflow 挖掘	222
11.1 workflow 挖掘的发展	222
11.2 workflow 挖掘的概念与作用	223
11.3 workflow 挖掘的内容	225
11.3.1 workflow 模型的重构	225
11.3.2 workflow 的监控与 workflow 挖掘的评价	226
11.3.3 组织视图挖掘	228
11.4 workflow 挖掘的应用	229
思考题	231
本章参考文献	231

第四部分 商务智能发展

第 12 章 商务智能进展	235
12.1 商务智能应用趋势	235
12.2 商务智能在中国的发展	239
12.3 商务智能动态	242
思考题	247
本章参考文献	247

第五部分 数据挖掘实验

第 13 章 AlphaMiner 数据挖掘系统	251
13.1 AlphaMiner 简介	251
13.2 AlphaMiner 的应用	254
思考题	260
本章参考文献	260



第一部分 商务智能基础

商务智能概论

随着信息化时代的来临，经济全球化的发展，信息数据存储成本的不断下降，企业数据正在以惊人的速度增长。这些数据是企业的重要资源，但目前大多数的企业并未对其进行进一步的利用。统计表明，目前国内企业数据有效利用率不足 7%，许多决策是在没有充分信息支持的情况下做出的。为应对日益激烈的竞争，企业需要有灵敏的感觉和快速反应能力，提高反应速度和决策的准确性。美国市场调查公司 IDC 的调查表明，37% 的业务决策仍然主要依赖于“直觉”或本能^{*}。如何充分利用这些隐藏着巨大商业价值的数据资产，提炼出有价值的信息、知识，对提高企业的智能至关重要，依赖直觉制定决策使企业管理水平远远落后于投资商务智能的领先企业，商务智能已成为领先企业与传统企业产生差异化的重要因素，而且商务智能开始扩展到业务运营将使这种差异更加明显。正如前 GE 首席执行官 Jack Welch 所说：“一个组织获取知识以及把知识快速转化为行动的能力决定其最终的竞争优势”。在这种背景下，商务智能（Business Intelligence, BI）逐渐得到学术界和企业界的重视，它被认为是继 ERP 之后企业信息化的又一个新热潮。商务智能能够帮助企业整合数据，并把数据转换成有用的信息，从数据中获取知识，提高企业理性的管理决策的能力。“亿万富翁商务成功的秘密是知道一些其他人不知道的东西”，商务智能的最终目的是帮助管理者实现更有效的企业管理，做出更明智的决策，取得更大的收益。

本章从商务智能的产生背景、发展历程开始，讨论商务智能的定义、作用等基础知识，然后对商务智能的功能和应用领域进行讨论，通过案例阐述商务智能的价值。

1.1 商业决策需要商务智能

在信息时代，企业的智能已成为生存之本。智能企业的发展会经历以下阶段：数据和

* IDC 发布最新报告呼吁企业积极运用商业智能。 <http://www.chinabi.net/Article/binews/200712/888.html>。

应用集成、把数据转化为信息、信息转化成知识以及知识转化为行动。

1.1.1 数据、信息与知识

数据是记录、描述和识别事物的符号，通过有意义的组合来表达现实世界中某种实体的特征。数据多表现为简单的事实，例如超市 POS 终端输入的一条商品销售记录。数据也称为记录、案例、样本等。数据用属性描述，属性也称为变量、特征、字段或维等。

数据经过解释后可以转换为有用的信息。信息是经过某种提炼、加工和集成后的数据。信息是可以被人们理解和解释的，对不同的人可能价值不同。公司员工把销售报表交给企业高层查看，这个报表文件可以被双方解读，传递的就是信息。

数据和信息虽然并不等同，却也是密不可分的。概括地讲，数据是信息的载体，而信息是对数据的解释。举个更通俗的例子，一张纸条上面写着“1982年11月18日”。“1982年11月18日”是数据，可是对于某个人来说，这可能正是某个好朋友的生日，这个日期对他来说就是信息。

目前企业缺少的不是数据，而是如何得到正确的信息辅助决策。信息可以是完整的，也可以是片断的；可以是关于过去，或者关于现在的，也可能是涉及未来的。目前天很热，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。如果根据这两条信息预测明天气温至少 37℃，那么这是有关未来的信息。尽管明天高温是可能甚至必然的，但这种预测未来的信息多少带有不确定性。为了减少不确定性，提高内容置信度，必须对信息进行提炼、加工和集成。

知识就是对信息内容进行的提炼、比较、挖掘、分析、概括、判断和推论。知识作为一种资源，其重要性越来越受到重视。知识分为事实性知识和经验知识。事实性知识是人类对于客观事物和现象的认识结果。经验知识多是一种隐性知识，是存储在人们大脑中的经历、经验、技巧、体会和感悟等尚未公开的知识。例如医生给病人看病，首先会做病体检查以了解病情，然后对病人的家族史和病史了解分析，再结合自身的临床经验进行诊断。病人主诉、病体检查、家族病史是事实性知识，而医生的临床经验则属于经验知识。超市经理把 POS 中的数据进行统计分析，发现几种商品被同时购买的比例很高，分析影响销售的关键因素或建立回归模型对下一阶段的销量进行预测，这些有助于业务决策的分析结果都是知识。

数据是宝贵的财富，只有充分利用这种财富，识别信息，获取知识，辅助商务决策，才能从中获得价值。图 1.1 展示了数据、信息、知识和决策之间的关系。

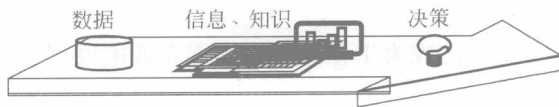


图 1.1 数据、信息、知识和决策的关系

1.1.2 管理就是决策

管理学家西蒙指出：管理就是决策。决策是企业管理的核心，贯穿管理的全过程。根据决策活动特点，企业管理可分为战略层、中间管理层和运营层三个层次。企业的各个层次都需要做决策。

(1) 战略层是企业决策中的最高层，负责管理、控制和协调整个企业的正常运行。厂址选择、资金分配计划、管理体制确定等都是战略层的决策范围。

(2) 中间管理层处于企业决策的中间层，主要包括销售、财务、生产、人力资源等部门，是实施战略层计划、管理控制运营层业务的关键所在，例如作业计划、作业调度、广告部署等。又如，某大型企业的区域销售主管把工作重心放在正确推断业务的发展方向上，用以确定该地区未来的业务重点。

(3) 运营层也称为业务操作层。随着科技的发展，员工素质的提高和组织的日趋扁平化，决策权逐渐下放。超市货品怎样摆放能增加销量，银行向哪一类顾客推广业务能得到最大反馈率，电子商务网站如何布局使浏览者变为购买者，这些都取决于业务操作层的决策。好的决策不但可以提高运营层的绩效，而且有助于中高层获得有价值的信息和知识，实现企业资源的优化配置。

1.1.3 决策需要信息和知识

决策不是简单的事情，直觉式的决策并不一定可靠。根据 Microsoft 公司统计，超过 74% 的商业决策落后于预定计划或以失败告终，每年损失 740 多亿美元。因决策失误造成重大损失的案例很多。2005 年世界第二大零售商 Carrefour（家乐福）败走日本是决策失误的后果。日本消费者的消费习惯和欧美不同。欧美国家的许多家庭在周末会驱车到郊区的大型超市采购价格便宜的食品和用品存放在家中。但日本人的饮食十分讲究新鲜，所以日本的超市一般都设在交通流量大的车站附近或者居民比较集中的住宅区。而 Carrefour 照搬在欧美国家的经验，在日本开设的超市位于城市的远郊区，没有根据日本不同的商业文化和消费习惯来调整经营策略，导致了 Carrefour 在日本“水土不服”。

企业各层级决策人员都必须有足够的信息，越往高层，决策需要的信息力度越大，而信息对决策的影响也越大。决策是在充分的信息和知识基础上的判断，决策也是有风险的：决策 = 信息 + 知识（经验） + 冒险。而充分的有价值的信息、知识以及经验可以降低决策风险。

1.1.4 智能型企业

企业有智能吗？智能是生物体获取知识、利用知识的能力，从仿生学的角度而言，企业