

管理科学博士论丛

决策支持与知识发现

■ 马尚才 李爱军 石洪波 编著



中国科学技术出版社

C934
70

管理科学博士论丛

决策支持与知识发现

马尚才 李爱军 石洪波 编著

中国科学技术出版社

· 北 京 ·

图书在版编目 (CIP) 数据

决策支持与知识发现/马尚才, 李爱军, 石洪波编著.
北京: 中国科学技术出版社, 2005. 3

ISBN 7-5046-4013-1

I. 决... II. ①马...②李...③石... III. 决策支
持系统—研究 IV. TP399

中国版本图书馆 CIP 数据核字 (2005) 第 024596 号

责任编辑: 郑洪炜

封面设计: 李瑞峰

责任校对: 林 华

责任印制: 王 沛

中国科学技术出版社出版

北京市海淀区中关村南大街 16 号 邮政编码: 100081

电话: 010-62103210 传真: 010-62183872

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

科学普及出版社发行部发行

山西荣博印业有限责任公司印刷

*

开本: 850 毫米×1168 毫米 1/32 印张: 9.75 字数: 262 千字

2005 年 3 月第 1 版 2005 年 3 月第 1 次印刷

印数: 1—1000 册 定价: 30.00 元

(凡购买本社的图书, 如有缺页、倒页、
脱页者, 本社发行部负责调换)

前 言

决策是人类社会生产与生活中无处不在的管理活动。为了使决策更加科学、更加准确,人类一直寻求科学决策的方法和技术,特别是计算机的诞生,为人们提供了十分强大的决策支持工具。从最初的单项处理系统、管理信息系统到高级形式的决策支持系统,人们把计算机技术与管理科学相结合,不断地为决策者提供先进的决策支持技术。尤其是决策支持系统从本质上脱离了单纯管理数据的决策支持理念,通过灵活的二级管理系统对数据资源进行整合和利用,更重要的是提供了灵活多样的模型管理和知识系统。但由于近年来计算机的日益普及,特别是互联网的普及,全球经济趋于一体化,使得现代决策更加复杂、对决策的时间要求越来越短,传统的决策支持系统的理论和框架需要进一步拓展,以适应现代决策的要求。知识发现就是决策系统模型和知识拓展的更趋于智能和人性化的决策支持技术,而数据仓库技术则拓展决策支持系统的数据管理技术,形成了新的数据管理理念。

知识发现是从数据集中识别出有效的、新颖的、潜在有用的可支持决策者的决策活动的知识。实现知识发现的数据管理基础是数据仓库,知识发现的过程是建立在管理完善的数据集上的。数据仓库用于存储大规模的数据集,它采用不同于传统数据库存储管理数据的方法,利用主题组织数据,具有连续稳定的各层次数据。而知识发现本身也不同于传统决策支持系统的建模和建立知识系统的原理方法。它集数据预处理、规则归纳、模式识别、数据分析预测等多种过程于一身,是统计学、模式识别、人工智能、机器学习、神经与进化计算等多学科的研究领域。知识发现技术的研究为人们在巨复杂系统中的决策提供更科学、更方便的支持。

本书全面介绍了决策支持与知识发现的方法和技术,全书共

分九章。第一章讨论了传统决策支持系统的概念、理论框架和技术特点，指出了传统决策支持系统需要改进的方向。第二章则介绍了知识发现的概念、知识发现的过程和目的以及与决策支持的关系。第三章至第六章介绍知识发现的数据基础——数据仓库，重点阐述了数据仓库的概念、数据模型以及数据仓库的数据集成和维护。第七章至第九章详细讨论了知识发现的方法和技术，包括 OLAP 技术、智能计算方法、统计方法和机器学习方法，给出了最新的知识发现算法及其应用范围。

本书由山西财经大学信息管理学院的马尚才教授、李爱军副教授和石洪波博士合作完成。其中马尚才教授撰写了第一、二章，石洪波博士撰写了第三、四、五章，李爱军副教授撰写了第六、七、八、九章。

山西财经大学硕士研究生殷召、付永贵、杨慧等同学参加了本书的校对修订工作，谨向他们表示诚挚的谢意。

此外，本书的出版得到了山西财经大学领导和同事们，特别是管理科学与工程学院院长张所地教授的支持和帮助，在此致以衷心的感谢！

本书既想对决策支持和知识发现的概念和方法进行系统的阐述，又力求反映最新的知识发现和决策支持的新技术和新思想，还要考虑篇幅的限制。这项对于作者的水平来说，有一定难度，加之成书仓促，不妥之处难免，希望得到更多的建设性意见。

作者 马尚才

2004 年 12 月

目 录

第一章 传统的决策支持系统	1
第一节 决策支持系统的概念	2
一、决策支持系统的定义和特征	3
二、决策支持系统的功能	5
三、决策支持系统的分类	7
第二节 决策支持系统的系统结构	8
一、决策支持系统的基本组成	9
二、决策支持系统的体系结构	28
三、决策支持系统的三个技术层次	36
第三节 传统决策支持系统存在的问题	38
一、数据管理问题	38
二、模型分析问题	40
第二章 知识发现：决策支持的新思路	42
第一节 知识发现的概念	43
一、知识发现的定义和基本概念	43
二、知识发现的处理过程	45
三、知识发现与数据挖掘	51
第二节 知识发现的数据基础	51
一、数据的组织存储方式	52
二、数据的类型	52
第三节 知识发现的目标	53
一、广义知识	55
二、分类和预测	55
三、聚类知识	57

四、关联规则	58
五、偏差分析	59
第四节 知识发现的方法	59
一、数据库访问技术	60
二、统计方法	61
三、机器学习	63
四、智能计算	65
第五节 决策支持与知识发现	66
第三章 数据仓库原理	68
第一节 数据仓库的基本概念	68
一、数据仓库的定义	68
二、数据仓库的特征	69
第二节 数据仓库的体系结构	72
一、数据仓库中的数据组织	72
二、数据仓库中的体系结构	75
第三节 数据仓库中的数据存储与组织形式	77
一、数据的组织方式	77
二、数据的存储方式	79
三、元数据	81
第四节 数据仓库的粒度和数据分割	88
第四章 数据仓库的数据模型	92
第一节 数据仓库建模基本原则	92
第二节 数据仓库中的基本概念	94
一、维	94
二、数据立方体	95
三、视图	96
第三节 高层模型	98
一、E-R 图	98
二、信息包图	99

第四节	中层模型	103
一、事实表和维表	103	
二、三种中层模型	106	
第五节	低层模型	108
第五章	数据仓库数据集成和维护	113
第一节	数据集成概述	113
一、数据集成的概念	113	
二、数据集成的步骤	114	
三、影响数据集成的关键因素	115	
四、数据集成的作用	116	
第二节	数据抽取	117
一、数据抽取的任务	117	
二、数据抽取时应考虑的问题	119	
第三节	数据转换	120
一、数据转换的任务	120	
二、如何实施转换	122	
三、数据转换时应考虑的问题	122	
第四节	数据清理	123
第五节	数据装载	126
一、数据装载的任务	126	
二、数据装载应考虑的问题	128	
第六节	数据集成工具	129
第七节	数据维护策略	132
第六章	OLAP 技术	135
第一节	OLAP 的技术概念	135
一、OLAP 的定义和准则	135	
二、OLAP 的特性	138	
三、OLAP 和 OLTP 的区别	139	
第二节	OLAP 多维分析	140

第三节	OLAP 的实现方式.....	145
一、	基于多维数据库的 OLAP.....	145
二、	基于关系数据库的 OLAP.....	149
三、	混合型的 OLAP.....	151
第七章	智能计算	153
第一节	人工神经网络	153
一、	人工神经元及感知机模型	155
二、	前馈神经网络	158
三、	径向基函数神经网络	164
四、	自组织特征映射神经网络	173
五、	神经网络集成	179
第二节	遗传算法	185
一、	遗传算法的一般框架	186
二、	遗传算法的实现技术	189
第八章	统计方法	194
第一节	关联分析	194
一、	基本概念	194
二、	关联规则挖掘算法	196
三、	关联规则价值衡量方法	202
第二节	聚类分析	203
一、	聚类原理	204
二、	经典聚类算法	209
三、	概念聚类	212
第三节	支持向量机	215
一、	支持向量机的基本思想	215
二、	模式分类的支持向量机	221
三、	非线性回归支持向量机	223
第四节	粗糙集	224
一、	粗糙集基本概念	225

目 录

二、粗糙集的知识约简	229
三、粗糙集的分类算法	232
第九章 机器学习	235
第一节 决策树	235
一、决策树学习	238
二、选择最佳分类属性	238
三、控制决策树规模	245
第二节 规则归纳	250
一、序列覆盖算法(sequential covering) ...	250
二、AQ 学习	255
三、学习谓词逻辑形式规则	259
四、反转归并	264
第三节 基于案例的学习	267
一、最近邻方法	268
二、基于案例的推理	269
第四节 贝叶斯学习	273
一、贝叶斯学习理论	273
二、贝叶斯网络	276
三、贝叶斯网络学习	278
四、贝叶斯分类器	286
参考文献	291

第一章 传统的决策支持系统

由于现代社会的生产力发展和科学技术的发展，各组织管理者们决策的科学性成为事业成功的关键。由于经济全球化的趋势和互联网的普及，使得管理者的经营活动呈现如下的特点。

① 第一个特点是社会活动日益复杂，决策呈现多目标、多因素特征。整个组织的活动涉及到大科学、大工程越来越多，一个决策问题涉及的方面和运用的科学越来越多，因此现代决策规模大、结构复杂、变量多。比如我国宏观经济决策的目标就有两个：一是要完成国家任务，二是要满足人民需要。宏观经济决策不仅要考虑国内经济环境，还要考虑政治、文化、自然资源因素。还要考虑国际环境。随着 Internet 越来越广泛的应用，企业规模越来越大，涉及业务越来越多。

② 第二个特点是社会活动日益多变，决策具有时间紧迫性的特征。社会活动日益多变主要从两个方面来体现，一是科学发明转化为生产力的周期越来越短，比如美国在第一次世界大战前，转化周期大约为 30 年，第一次世界大战后到第二次世界大战期间，转化周期大约为 15~16 年，而到第二次世界大战后，则缩短为 8~9 年，发展到日新月异的今天，转化周期还在缩短。二是新技术更新快，新的科学技术不断产生，据统计，最近 40 年产生的新科技超过了以往 2000 年的总和。

③ 第三个特点是现代决策的影响范围比较大。由于组织的开放性和全球一体化的综合业务发展趋势，现代决策影响范围较大，一个决策可能会影响到多个企业、一个行业，甚至是几个行业，这就要求决策要科学、准确。

这样导致了现代决策日趋复杂且越来越难以用程序化的程式

描述决策的过程。面对现代决策要做到快而准，这就要求管理者能做出科学决策。

决策过程是一个具有明确目标、由许多相互制约、相互联系的要素构成的有机整体，必须应用系统理论分析、研究问题，而这些研究必须建立在信息基础之上。只有掌握大量信息，才能系统地对信息进行归纳、比较、选择，提炼出对决策有效的信息。决策成功与否与决策事件面临的主、客观条件密切相关，因此，决策过程中，不仅要考虑需要，还应考虑可能，要估计不利条件和失败的风险。对决策方案的评判显然也不能追求其最优，而应当追求“足够好”或“令人满意”的效果。掌握科学的决策技术和方法是决策科学化的重要方面。科学决策的技术和方法有很多，但一般把其分为“硬”技术和“软”技术。所谓“硬”技术，是指建立在数学模型的基础上，运用计算机辅助决策。比如运筹学、计量经济学等。而“软”技术是指建立在心理学、社会学、行为科学等基础之上的“专家创造力”技术。这种技术可弥补“硬”技术对某些政治、社会及人文因素无法定量测算分析的缺陷。对于一个经营管理者来讲，一个人的思维与经验均是有限的，他们需要能够帮助他们进行决策的计算机系统。决策支持系统（decision support system, DSS）就应运而生。本章对于传统决策支持系统的概念、方法和技术做一个阐述。

第一节 决策支持系统的概念

决策支持系统（decision support system, DSS）在 20 世纪 80 年代末引入我国。它是一门集管理科学、计算机技术、信息科学、系统工程于一身的综合边缘学科。在近三十年的发展中，决策支持系统无论从理论还是应用都取得了很大进展。决策支持系统在社会各领域的管理决策中，起到越来越重要的作用，它的开

发与研制对我国管理决策水平的提高将起到巨大的促进作用。本节介绍决策支持系统的理论概念。

一、决策支持系统的定义和特征

1. DSS 的定义

由于决策支持系统涉及的范围比较广，一直没有一个公认的定义，为了决策支持系统的进一步研究和发展，我们在此介绍两种经常被学者引用的定义。

定义 1.1（DSS 的狭义定义）：DSS 是具有交互式计算机系统的特征，帮助决策者利用数据和模型去解决非结构化的问题的系统。

该定义是 Scott Morton M. S 在 20 世纪 70 年代初题为“管理决策系统”一文中首先提出来的。这个定义一般被人们称之为狭义定义，主要是这个定义在业界一直存在着分歧。有些学者认为 DSS 这个定义不严密，用一些尚未定义过的名词和术语（比如象非结构化、半结构化这些词本身没有确切的定义）来定义另一个名词。还有学者认为这种定义太严格，只有少数实际系统能满足它，在实际的应用中很难做出完全满足上述定义的系统。因此，有些学者提出了一种广泛的定义，以利于决策系统的实际构造。

定义 1.2（DSS 的广义定义）：任何（以任何方式）支持决策的系统都称作为决策支持系统。

该定义刻画的 DSS 范围很广，没有对 DSS 技术特征以及与其他计算机系统的区别加以描述。

还有一些其它的定义方法，但均未得到严格的论证，目前尚未有绝对认可的结论。而且无论是狭义定义还是广义定义都不能够把 DSS 的价值技术要求和研制方法表述出来，使人对 DSS 系统的这几个方面有所了解。更重要的是不同研究者对 DSS 的看法颇

有差别,管理者和计算机科学家,很难以相同观点看待同一事物。1977年在加利福尼亚举行了DSS讨论会,讨论了可以称之为DSS的系统,1981年在美国举行了DSS国际讨论会,又对DSS系统进行了进一步的讨论,为了补充决策支持系统定义的不足,使大家对DSS及其潜力有更深刻的理解,Keen在1986年的一次DSS讨论会上提出了DSS特征,对DSS进行了更详尽的描述。

2. DSS 特征

我们经过对DSS的实例研究,并参考了许多学者对DSS的特征的解释,总结出DSS具有下面的特征:

(1) DSS是要解决高层管理者遇到的半结构化的决策问题

DSS的服务对象是面临某种决策问题的管理人员(包括领导人员和一般工作人员)。没有决策的需要和可能,DSS就是无的放矢。DSS能够帮助的是半结构化问题。所谓结构化程度是指对某一决策过程的环境及原则能否用明确的语言(数学的或逻辑的、定量的或定性的)加以说明与描述。如果能描述清楚,则称之为结构化决策问题;如果不能,则为非结构化问题。介于两者之间的,即有所了解但不全面;有所分析但不确切;有所估计,但不确定的问题则称之为半结构化问题。对于结构化问题完全可以按例行规定方法处理,无须DSS帮助,对于半结构化和非结构化问题是我们DSS所要解决的问题。

(2) DSS是把模型与分析技术的使用与传统的数据存贮和检索功能结合起来

在管理者进行决策时除了一些历史和实时信息外还需要一些预测信息及各种方案的分析比较等,这样就需要一个比较复杂的计算机系统。

(3) DSS着眼于让非计算机人员方便地使用

由于解决的非结构化问题不能象传统的计算机系统那样,全部工作都由计算机来完成,而需要决策者与系统交互使用,把决

策者的思想与方法融于系统中，使系统能更好地帮助决策者，这就要求有相当水平的人机交互界面。DSS 就是要使各类决策人员都可以比较方便地操作系统。

(4) DSS 注重适应性和灵活性，以便能适应环境和用户的决策手段的变化

高层次的决策问题是一个复杂的过程。在不同的时期、不同的环境，决策者所面临的决策问题不同，DSS 要随着这些因素的变化提供不同形式的帮助，而不象早期的简单的应用软件那样，只能在某一阶段、某一工作中，按固定的算法给出一个孤立的数据作为结果。

(5) DSS 是帮助决策者而不是代替决策者

在决策过程中，决策结果的好坏取决于决策者的经验、知识、理论观点、思想方法以及个人偏好等等主观因素，但一个人的这几个方面的能力都是有限的，要想做出正确的决策，提供的素材越多、方法越多、知识越多，这个正确度就会提高。DSS 就是要起到一个决策者的智囊团的作用，而不是要代替决策者做决策。

建立高效的决策支持系统是设计者必须致力的方向。我们学习 DSS 的理论和技術，就是为了建立适合组织需要的、能达到上述要求的、有效的 DSS 这一工具。

二、决策支持系统的功能

由于半结构化、非结构化问题本身存在不确定性，因此，不能严格地规定 DSS 功能，但要能支持半结构化或非结构化的决策问题，一个 DSS 应该具有以下最基本的功能：

1. 能整理、收集并及时提供本系统内外与决策有关的数据

它包括内部各种数据，如财务数据、销售、生产、资金、设备、技术等业务数据，这些信息一般可以从已有的管理信息系统

的业务数据库中获得；与决策有关的还有外部信息，如市场需求、行业状况、原材料、新技术等，这些信息的获得较难，主要靠调查和从有关资料、文献中摘取，在目前全球信息化的大趋势下，将可以由信息服务业提供联机查询。

2. 及时提供各项活动的反馈信息

决策制订和实施过程中产生的各种信息是控制决策实施的重要的资源，DSS 系统应当能及时收集反馈信息与决策方案预期的预期目标项比较，包括不同阶段、不同环节产生的各种有用信息。

3. 能存贮提供与决策有关的各种技术模型及分析方法

能用一定方式存贮提供与决策有关的各种技术模型，包括各种常用模型特别是数学模型与运筹学的分析方法；能存贮支持决策的各种分析、决策、预测模型，比如运筹学、统计学、计量经济学等模型，并能支持这些模型间的联系。

4. 能使用户方便地修改数据、模型、知识、方法等

对于非计算机用户能方便地存取、查询数据，还要能方便地建立模型（专用）、使用模型，对必要的模型能方便地存取，并能和基本数据、常用方法、专家知识等子系统灵活地传递信息，并以非常人性化的方式表示、接收。

5. 能够灵活应用知识、模型和方法对数据进行加工、预测

为非计算机用户提供灵活的操作使用方式，使用户能用菜单、窗口等简单易学的界面对 DSS 中的知识、模型、数据进行加工，形成支持决策的模型、方案、报告。

6. 具有数据表格、图形、通信等传输功能

现代决策涉及的问题范围大，往往一个决策者无法把握决策的效果，需要小组（群体）进行决策，而群体决策中一个最主要的一个问题是决策者之间的信息交流，因此，DSS 系统应当能表示多种媒体信息，并具有异地传输信息的功能。

7. 具有很快的加工速度与反应时间

系统应能很快地做出反应，包括运算速度、组织数据、构造

模型等方面。用户需要具有与自己同步思维的 DSS 来支持决策。

为帮助大家理解和记忆，上述功能可归纳如下：

收集数据，提供方法，发展数据，分析比较；

灵活迅速，易用易改，面向用户，支持决策。

DSS 系统的中心就是支持决策，决策的需要决定了系统应该具有哪些功能。

三、决策支持系统的分类

自从 DSS 产生以来，人们对其作了各种研究，为了总结 DSS 的理论和技術，对 DSS 系统进行了分类。不同的 DSS 有不同的功能特征和应用领域，按其应用范围分，可分为专用 DSS 与通用 DSS；按其发展的进程，可分为传统 DSS 与职能 DSS；按其支持的决策层次有战略 DSS、战术 DSS、操作控制 DSS；按其使用者分为个人 DSS 和群体 DSS；本书介绍 S. L. Alter 的分类方法。他在对 56 个 DSS 进行分析后，按 DSS 的功能和应用范围，提出了将 DSS 划分为七类。

1. 文件检索系统

对各种数据文件进行联机查阅检索。例如，库存控制系统，航空订票系统。

2. 数据分析系统

利用有关操作对各种业务数据进行分析，产生分析报告和预测结果等。例如银行数据分析系统，经济预测预警系统。

3. 信息分析系统

通过使用决策数据库和模型库的各种简单模型，对管理决策的数据进行分析。例如，市场分析系统。

4. 统计模型系统

主要使用统计模型和会计中的一些方法求出最佳结果。例如