

曾黄麟 著

关于粗集理论、模糊逻辑、
神经网络的理论及其应用

智能计算

重庆大学出版社

内 容 简 介

智能计算体现了智能信息处理新一代计算风范。本书主要介绍智能计算研究领域近年来迅速发展的粗集理论、模糊逻辑、人工神经网络理论及其应用。这些理论在研究不精确、不完整、不确定的真实世界中数据的知识表达、学习、挖掘和归纳等方面提供了有效的处理技术和方法,把人们带到一个科学逻辑思维和智能信息处理的新天地。

为了阐明这些理论及技术的基本问题,突出这些新方法的应用,全书共分为四部分,第1部分介绍粗集理论及其应用,突出从数据中挖掘知识、评价系统参数的重要性和知识简化的特点。第2部分介绍模糊逻辑及其应用,重点在于知识的模糊化概念、隶属度特征函数的构建及模糊推理方法。第3部分介绍人工神经网络理论及其应用,主要讨论人工神经网络的结构和学习方法,突出优化计算和非线性建模思路。第4部分介绍综合智能信息处理方法及其应用,讨论粗集理论、模糊逻辑和人工神经网络各自的特点,探讨这些智能计算的有机结合,提出综合智能信息处理及其应用方法。

本书是有关智能计算方法及其应用的著作,为了便于学习和供计算机、人工智能和信息处理等学科的大专院校高年级本科生、研究生作为教材,本书重在智能计算方法的介绍,省略了一些繁琐的理论推导过程,并用较大的例题、习题和应用实例来阐明其概念和方法,有些问题本身就是从科研课题中提取出来的,既可供学生作为毕业设计题目,也可供计算机学科、人工智能学科和信息处理学科等领域的大专院校师生,工作者,工程技术人员作为阅读参考和深入研究的课题。

图书在版编目(CIP)数据

智能计算:关于粗集、模糊逻辑、神经网络的理论及其应用/曾黄麟著. —重庆:重庆大学出版社,2004.6
ISBN 7-5624-3175-2

I. 智... II. 曾... III. 人工智能—计算 IV. TP183

中国版本图书馆CIP数据核字(2004)第045787号

智能计算

——关于粗集理论、模糊逻辑、神经网络
的理论及其应用

曾黄麟 著

责任编辑:彭宁 版式设计:彭宁
责任校对:廖应碧 责任印制:张立全

*

重庆大学出版社出版发行

出版人:张鸽盛

社址:重庆市沙坪坝正街174号重庆大学(A区)内

邮编:400030

电话:(023) 65102378 65105781

传真:(023) 65103686 65105565

网址:<http://www.cqup.com.cn>

邮箱:fxk@cqup.com.cn(市场营销部)

全国新华书店经销

重庆科情印务有限公司印刷

*

开本:787×1092 1/16 印张:20 字数:499千

2004年6月第1版 2004年6月第1次印刷

印数:1—3 000

ISBN 7-5624-3175-2/TP·492

定价:28.00元

本书如有印刷、装订等质量问题,本社负责调换

版权所有 翻印必究

前言

近年来,智能信息处理已成为众多学科领域研究的热点。随着科学技术的高速发展,系统越来越复杂,数据积累成倍增大,需要处理的信息量迅速增加,要把大量的数据作为信息,从数据中发现出有用的知识,把知识变成指导工作、生产、运营和销售的策略,等等这些已使一些传统的方法感到无能为力。特别是从不精确、不完整、不确定的数据信息中挖掘出科学的规律,归纳、推理知识,迫切需要创建新一代计算风范来实现。因此,智能计算^[1] (Intelligent Computing) 作为创建计算智能系统的一种新颖的计算方法,与其说它为人们开拓了一个科学逻辑思维和智能信息处理的新天地,倒不如说它是人工智能、机器学习与数据库技术相结合的必然产物。

智能(Intelligence)是人们认识客观世界并运用知识解决实际问题的能力,它集中表现在反映客观事物深刻、正确、完全的程度,以及应用知识解决问题的速度和质量上,往往通过观察、记忆、判断、联想和创造等表现出来。因此,利用智能计算来实现的智能信息处理是与传统信息处理方法完全不同的一种新颖的软计算方法(Soft Computing),智能信息处理科学家 S. K. Pal 指出^[61] “Soft computing is a consortium of methodologies that works synergetically and provides in one form or another flexible information processing capability for handling real life ambiguous situations. Its aim is to exploit the tolerance for imprecision, uncertainty, approximate reasoning, and partial truth in order to achieve tractability, robustness, and low-cost solution. The guiding principle is to devise methods of computation that lead to an acceptable solution at low cost by seeking for an approximate solution to an imprecisely/precisely formulated problem”。

因此智能计算具有以下特点:

1. 它能处理一些真实世界的现实信息,这些信息可能是定性、定量描述,也可能是具有不完整、不精确或不确定性的数据。

2. 以对观察和测量所得数据进行分类的能力为基础,依赖于强化数值计算,从中发现、推理知识或分辨、预测系统的某些特点、过程或结果对象等。

3. 受感于生物的计算机理,通过学习、自组织等方式对信息进行综合、归纳或推理,从而建立符号主义(Symbolism)、连接主义(Connectionism)或行为主义(Behaviourism)的数学模型。

4. 建立的数学模型具有对信息进行综合或并行处理能力,具有适应外部变化情况的自主控制能力、具有自扩展性和系统的稳健性。

利用智能计算来实现的智能信息处理一般有如下过程:

1. 理解该领域知识和相关的先验知识,明确系统目标。

2. 创建相关的数据库,这里包括确定系统的特征参数或变量,选择系统的训练学习样本,选择系统的测试样本。

3. 数据整理和预处理,这里包括不完整、不精确数据或定量信息、语意定性描述以及信息离散化、模糊化、特征提取等预处理。

4. 利用某种智能计算方法,进行数据简化,确定系统的有用特征参数或变量,化简系统,建立相应的数学模型或逻辑规则。

5. 测试从数据中挖掘或建立的数学模型或逻辑规则,并解释所得到的模式结果。进一步加工、综合系统,运用挖掘的知识解决客观世界的实际问题。

通过智能计算,从数据中挖掘知识,一般将产生下列类型的信息:

关联信息,若干个事件相关联的信息;

时序模式,发生概率最高的模式;

聚类信息,对数据进行聚类;

分类信息,进行分类的特性描述;

偏差信息,反映异常情况;

预测信息,从现有的数据中建立数学模型,计算未来值。

基于上述从数据中挖掘知识的智能信息处理的特点和重要任务,目前人们广泛采用的智能计算方法主要有粗集理论方法、模糊逻辑方法、人工神经网络方法、概率统计方法、聚类分析方法、回归优化和遗传算法等^[1-4,61-62,47-49,81-82]。特别是对于从不准确、不完整、不确定的数据中发现知识,粗集理论方法、模糊逻辑方法和人工神经网络方法更显示了它们无穷的魅力,所以本书主要介绍粗集理论、模糊逻辑和人工神经网络的基本原理以及它们从数据中挖掘知识的智能计算方法。

从数据中挖掘知识的智能计算方法的角度看,虽然三者都可以进行信息的并行处理,从输入-输出信息中实现一种非线性映射的数学关系。然而,它们的思路和方法都各有特点,粗集方法模拟人类的抽象逻辑思维,模糊理论着眼于现实世界的精确和不完整的信息传感,神经网络方法模拟形象直觉思维。因而,粗集理论的核心是基于不可分辨性的思想和知识简化的方法,从数据中推理逻辑规则作为知识表达系统的数学模型;模糊逻辑的基石是通过定义的隶属度特征函数表达模糊性,利用模糊推理规则,从数据中挖掘知识表达的逻辑关系;神经网络的关键则是利用非线性映射的思想和并行处理的方法,用神经网络本身结构表达输入与输出关联知识的隐函数编码。

智能计算不仅为人工智能和认知科学提供了新的科学逻辑和研究方法,而且为信息科学提供了有效的处理技术。正如模糊逻辑本身并不模糊,粗集理论本身不但不粗糙,而且还是一件去粗取精、去伪存真的锐利武器一样,智能计算的概念在理论上的深远意义和应用中的巨大潜力正吸引着世界各国许多专家学者的广泛注意。

本书的主旨就是向计算机学科、人工智能学科和信息处理学科等领域的人员介绍当代科学技术的这些新的研究方法,衷心希望这些理论方法成为科技工作者在智能信息处理中使用的有力工具。

为了阐明这些理论及技术的基本问题,突出这些新方法的应用,全书共分为四部分。第1部分介绍粗集理论及其应用,突出从数据中挖掘知识、评价系统参数的重要性和知识简化的特点。第2部分介绍模糊逻辑及其应用,重点在于知识的模糊化概念、隶属度特征函数的构建及模糊推理方法。第3部分介绍人工神经网络理论及其应用,主要讨论人工神经网络的结构和学习方法,突出优化计算和非线性建模思路。第4部分介绍综合智能信息处理方法及其应用。讨论粗集理论方法、模糊逻辑方法和人工神经网络方法各自的特点以及它们的一些不足之处,例如,神经网络可以实现有导师、无导师学习,但不能确定哪些知识是冗余的,哪些属性是有用的;粗集理论方法可以描绘知识表达中不同属性的重要性,进行知识表达空间简化,但它往往要求从训练数据中通过有导师学习推理规则;模糊理论可以利用领域中的专家知识,解决病态问题,建立便于有效实现的控制系统;但它像粗集理论一样,对适应变化的环境和系统本身的容错性都不如人工神经网络。因此,有必要进而探讨它们的有机结合方法。由于该部分(以*标注)有些内容目

前尚属研究前沿,理论和方法都不太成熟,作为大专院校高年级本科生、研究生可以不选修其中部分的内容。这里为智能信息处理领域的研究人员提出来,抛砖引玉,希望能推动智能计算的蓬勃发展,在智能信息处理领域引起更广泛地重视,以期获得更成功的应用。

全书共 13 章,具体内容安排如下:

第 1 部分 粗集理论及其应用

第 1 章介绍智能计算中的数据整理和预处理以及知识系统表达,包括不完整、不精确数据或定性、定量信息以及信息离散化等处理等方法。

第 2 章基于知识与分类相联系的观点,简明地定义智能计算中,知识的分类、等价、集合近似、近似分类与粗集的概念。

第 3 章是粗集理论的主要内容,介绍粗集理论关于知识简化、系统参数的重要性评价、决策系统条件属性简化、范畴简化、决策规则简化、推理学习以及知识系统最简逻辑表达等方法。

第 4 章讨论粗集理论在智能数据处理中的应用。从基于粗集理论的概念的应用、基于粗集理论方法的系统简化、从数据中挖掘决策规则等几个方面,介绍粗集理论在智能计算、数据处理中的应用思路和实例。

第 2 部分 模糊逻辑及其应用

第 5 章介绍模糊集合理论基础,重点阐明模糊隶属函数的定义及构建方法、模糊集合的基本运算律等概念。

第 6 章首先介绍精确数据的模糊化概念、模糊逻辑关系、模糊规则及模糊逻辑推理方法,并讨论了几种典型的模糊推理系统的特点、性质及建立方法等。

第 7 章讨论模糊逻辑在智能计算中的应用。这些应用主要包括:模糊综合评判系统,模糊推理与系统建模,模糊系统混沌时间序列预测,模糊划分及模糊聚类方法等。

第 3 部分 人工神经网络理论及其应用

第 8 章从人工神经网络基础知识入手,首先介绍人工神经网络的基本特性和人工神经网络的模拟方法。

第 9 章主要讨论人工神经网络结构及学习方法,按反馈互联神经网络和无反馈的前向式神经网络两大类来划分,介绍了目前这两大类网络中的主要典型网络的结构形式、特点、动态特性、网络综合、学习方法等。

第 10 章讨论人工神经网络在智能计算和信息处理的应用。这些应用主要包括:神经网络优化计算,神经网络非线性系统建模,神经网络分类及模式识别,和神经网络信号处理等。

第4部分 综合智能信息处理

第11章探讨粗集与模糊逻辑、概率统计方法的结合,从基于不确定性问题研究方法评价系统参数的重要性,讨论了基于概率统计定义的粗集方法、基于粗集的概率统计推理方法,和粗集理论与模糊逻辑的结合方法。

第12章探讨模糊逻辑与神经网络、神经网络与模糊推理结合的综合智能信息处理方法。主要讨论了:模糊神经网络分类器,模糊逻辑神经网络,利用神经模糊系统从数据中挖掘模糊逻辑,自适应神经模糊控制系统,自组织神经模糊推理专家系统等。

第13章探讨粗集理论和神经网络的综合智能信息处理。主要探讨了:信息近似化与粗神经网络,粗集理论与神经网络结合的推理决策系统,基于粗集预处理神经网络系统,和粗集、模糊与神经网络结合的综合信息处理系统等。

本书的安排将使读者不仅系统地了解粗集理论、模糊逻辑和人工神经网络的基本理论和技术知识,而且使读者从中领悟到智能计算的科学逻辑和应用方法论。

本书是作者在悉心阅读大量有关粗集理论、模糊逻辑、人工神经网络理论及智能信息处理的有关著作,加以综合分析和归纳,同时融入作者多年来的学习、研究成果编著的。由于智能计算涉及的粗集理论、模糊逻辑、人工神经网络理论等本身所具有的科学前沿的性质,还处于不断发展的阶段,加上作者水平有限和成稿时间仓促,书中一定存在许多缺陷和错误,诚祈专家和读者们指正。

在美国圣地亚哥州立大学高访期间,著名的智能信息处理科学家、美国圣地亚哥州立大学 R. Swiniarski 教授对本书的写作给予了热情指导;此外,本书中的一些研究工作的完成曾得到美国 RSCA 科研项目基金、国家自然科学基金和四川省科委重点科研项目基金的部分支持。值得一提的是我的母亲、妻子和女儿对我工作的理解、鞭策和支持,以及原四川轻化工学院的领导、老师的关怀与鼓励。对于所有这些亲人、老师和朋友的帮助,我在此谨表示深深地感谢!

曾黄麟
2004年2月

符号注释

$\sim$ 否定,(非)	$\cup P$ P 的并集
A 矩阵	$\nabla^2 E(x)$ 函数的 Hessian 矩阵
$\vee$ 分解,(或)	$\cap P$ P 的交集
A^T 矩阵的转置	$card(U) = U $ 集合 U 的基数
$\wedge$ 合并,(与)	$=df$ 表示定义真伪
A^{-1} 逆矩阵	$[X]_R$ 表示基于 R 的 X 的等价类
$\rightarrow$ 蕴含,(当……则)	$ind(P)$ 表示 P 上的不可分辨关系
A^+ 矩阵的伪逆	$P \Rightarrow Q$ 表示 Q 依赖于 P
$\equiv$ 等价,(当且仅当)	$R^-(X)$ R 上近似集
$\ A\ $ 矩阵的范数	$U \mid ind(R)$ 基于 R, U 中构成的所有等价类
$\Rightarrow$ 依赖	$R_-(X)$ R 下近似集
$\ A\ _p$ 矩阵的 L_p 范数	$BN_R(X)$ X 的 R 边界
(U, R) 基于 R, U 上划分的等价空间	$POS_R(X)$ X 的 R 正域
w_{ij} 突触加权值	$NEG_R(X)$ X 的 R 负域
$U \mid R$ 基于 R, U 中的对象构成的所有等价类族	$x \in _R(X)$ x 为是 R 的下成员关系
W 加权矩阵	$x \in ^R(X)$ x 为是 R 的上成员关系
$x \in X$ 元素 x 属于集合 X	$(X)_R(Y)$ X 和 Y 称为 R 下等价
R_{xy} 相关矩阵	$(X)^R(Y)$ X 和 Y 称为 R 上等价
$x \notin X$ 元素 x 不属于集合 X	$(X)R(Y)$ X 和 Y 称为 R 等价
$\forall$ 对于所有的	$(X)C_-(Y)$ X 为 R 下包含于 Y
$P \subset R$ P 为 R 的真子集	$(X)C^-(Y)$ X 为 R 上包含于 Y
$L(x, \lambda)$ 拉格朗日函数	$(X)R(Y)$ X 为 R 包含于 Y
$P \subseteq R$ R 包含 P	$d_R(X)$ R 描述的近似精度
R^n n 维实数空间	$P_R(X)$ R 粗糙度
$X \not\subset U$ 子集 X 不属于 U	$\alpha_R(X)$ 系统参数的重要性
$\nabla E(x)$ 函数一阶梯度	$red(P)$ 属性 P 的简化
	$core(P)$ $red(P)$ 的交称为 P 的核

$POS_P(S)$ S 的 P 正域

$red_s(P)$ 属性 P 相对于 S 的简化

$core_s(P)$ $red_s(P)$ 的交称为 P 相对于 S 的核

$red(F)$ 集合的简化族

$core(F)$ 简化集合族的交称为 F 的核

$red_Y(F)$ 集合 F 相对于 Y 的简化

$core_Y(F)$ 简化集合族的交称为 F 相对于 Y 的核

$r_P(S)$ 知识 S 和 P 之间的依赖度

V_a 属性 a 的属性值

d_x 定义决策规则函数

$|\theta|_s$ 表示模型 S 中公式 θ 的含义

$\text{Max}(k, L)$ 取 (k, L) 中的最大

$\text{min}(k, L)$ 取 (k, L) 中的最小

$\text{sup}(H)$ 取 H 的最小上确界

$\text{inf}(H)$ 取 H 的最大下确界

$\sum(P)$ 表示模型 S 中所有满足公式 P 的分解

F_ψ 基本算法 F 中具有相同后继 Ψ 的全部基本规则的集合

p_ψ 属于 F_ψ 的决策规则的全部前代的集合

A_λ 模糊子集 A 的 λ 截集

$\text{supp}(A)$ 模糊子集 A 的支集

$\bar{A}, A_c$ 模糊子集 A 的补集

$\mu_A(x)$ 模糊子集 A 的隶属度

$\mu_{R^-(F)}(x)$ 模糊集合 F 的 R 上近似中 x 的隶属度

$\mu_{R_-(F)}(x)$ 模糊集合 F 的 R 下近似中 x 的隶属度

$\omega(x)$ 特征函数

$A \cdot B$ 模糊子集 A 和模糊子集 B 的代数积

$A + B$ 模糊子集 A 和模糊子集 B 的代数和

$A \odot B$ 模糊子集 A 和模糊子集 B 的环和

目 录

符号注释	1
第 1 部分 粗集理论及其应用	1
第 1 章 智能数据预处理及知识系统表达	2
1.1 数据表知识表达系统	2
1.2 不完整、不精确数据预处理	4
1.3 属性值的离散归一化处理	8
习题与思考题	10
第 2 章 知识与分类、近似与粗集的基本概念	14
2.1 知识与分类	14
2.2 集合近似与粗集概念	18
2.3 集合近似及分类近似的度量	24
习题与思考题	28
第 3 章 知识系统的简化及逻辑表达	31
3.1 知识的简化	31
3.2 知识的相对简化	34
3.3 范畴的简化	35
3.4 范畴的相对简化	37
3.5 知识的依赖性	38
3.6 知识表达系统数据的协调性	42
3.7 知识表达系统属性的简化	45
3.8 知识表达系统决策规则的简化	50
3.9 推理学习	65
习题与思考题	70
第 4 章 粗集理论在智能计算中的应用	78
4.1 基于粗集理论的概念的应用	78
4.2 基于粗集理论方法的系统简化	81
4.3 从数据中挖掘决策规则	89
第 2 部分 模糊逻辑及其应用	98
第 5 章 模糊集合理论基础	99
5.1 集合论基础	99

5.2	模糊集合与隶属函数	102
5.3	模糊集合的基本运算	106
	习题与思考题	107
第6章	模糊规则与模糊逻辑推理系统	109
6.1	扩张原理及模糊关系	109
6.2	语言术语变量及模糊逻辑规则	114
6.3	模糊逻辑推理	118
6.4	模糊逻辑推理系统	123
	习题与思考题	138
第7章	模糊逻辑在智能计算中的应用	142
7.1	模糊综合评判系统	142
7.2	模糊推理与系统建模	144
7.3	模糊系统混沌时间序列预测	151
7.4	模糊划分及模糊聚类方法	156
第3部分	人工神经网络理论及其应用	163
第8章	人工神经网络基础	164
8.1	神经网络基本特性	164
8.2	人工神经网络的模拟	166
	习题与思考题	173
第9章	人工神经网络结构及学习方法	177
9.1	反馈互联神经网络	177
9.2	前向式神经网络	182
	习题与思考题	195
第10章	人工神经网络在智能计算、信息处理中的应用	199
10.1	神经网络优化计算	199
10.2	神经网络非线性系统建模	205
10.3	神经网络分类及模式识别	214
10.4	神经网络信号处理	217
第4部分	综合智能信息处理	221
第11章	粗集理论与模糊逻辑、概率统计方法的综合 智能信息处理	222
11.1	基于不确定性问题研究方法评价系统参数的重要 重要性	222
11.2	基于概率统计定义的粗集方法	228
11.3	基于粗集的概率统计推理方法	232

11.4	粗集理论和模糊逻辑的综合智能信息处理	236
第12章	模糊逻辑与神经网络的综合智能信息处理	243
12.1	模糊神经网络分类器	243
12.2	模糊逻辑神经网络	247
12.3	利用神经模糊系统从数据中挖掘模糊逻辑	249
12.4	自适应神经模糊控制系统	255
12.5	自组织神经模糊推理专家系统	259
第13章	粗集理论与神经网络的综合智能信息处理	264
13.1	信息近似化与粗神经元网络	264
13.2	粗集理论方法与神经网络结合的推理 决策系统	267
13.3	基于粗集方法预处理的神经网络系统	270
13.4	粗集、模糊与神经网络的综合信息处理系统	274
	中英文名词对照表	296
	参考文献	299

第 **I** 部分

粗集理论及其应用

随着人工智能从研究内容到研究方法经历的发展与变化,在对于高层次智能行为的研究中,大多数研究不仅集中于知识表示和符号推理,而是更加重视知识与大量观察和实验数据的处理、归纳、分类相联系。因而如何对不完整、不精确数据或定性、定量信息进行分析、推理,发现数据间的关系,提取有用特征,简化信息处理,研究不完整、不精确、不确定知识的表达、学习、归纳方法等已成为智能信息处理中的重要研究课题。

近年来,波兰华沙理工大学 Z. Pawlak 教授等一批科学家提出了用粗集理论(Rough Sets),研究不完整数据、不精确知识的表达、学习、归纳等方法^[2-5],这套方法与这一领域传统的不完整、不精确、不确定知识的处理方法不相同。它是根据对一个系统的观察和测量所得的现实数据信息,从分类的观点,以集合近似、近似分类与不可分辨性的概念为基础,通过粗集理论简化方法,从中发现、推理知识和分辨系统的某些特点、过程、预测系统的结果等。

自 1992 年(波兰)第一届粗集理论国际研讨会召开以来,几乎每年都有这方面的国际研讨会,报道粗集理论在知识发现与数据挖掘方面的成果。近十年来,粗集理论作为智能计算的科学研究,无论是在理论方面还是在应用实践方面都取得了很大的进展,展示了它光明的前景。粗集理论不仅为信息科学和认知科学提供了新的科学逻辑和研究方法,而且为智能信息处理提供了有效的处理技术。

第 1 章

智能数据预处理及知识系统表达

1.1 数据表知识表达系统

在智能计算系统中,常会碰到要处理的对象可能是用语言方式表达,也可能是用数据表达,可能是精确的数据,也可能是不精确的数据,甚至可能会有一些缺省的信息或者相互矛盾的信息,这些要通过人们的智能才能处理的数据,把它叫作智能数据。

在现实生活中,不管是考察一件事情,还是分析一个系统的工作状态,往往是通过观察、测量得到一个记录表,这个记录表就是一个知识表达系统 S 的表格表达形式。这样的数据表可称为知识表达系统 (knowledge representation system), 或简称为 KRS, 有时也称为信息系统属性值表。

因为知识系统的表格表达法可以看做一种特殊的形式语言,在知识表达系统数据表中,列表示属性 (attributes), 表中的数据就是各属性值;行表示研究对象 (objects) (如状态,过程等), 并且每行表示该对象的一条信息。显然,知识系统的表格表达的数据不仅可以通过观察、测量得到,而且数据间的关系清楚明晰。容易看出,一个表可以看做由属性定义的一族等价关系,它构成一个表达系统的知识库,也就是说,一个数据表代表了一个系统的性质。

为了处理智能数据,需要进行知识的公式化表达。因为知识表达系统的基本成分是研究对象的集合,关于这些对象的知识是通过指定对象的基本特征 (属性) 和它们的特征值 (属性值) 来描述的,故一个数据表也可以用公式表达为知识系统 S :

$$S = \langle U, C, D, V, f \rangle \quad (1.1)$$

这里 U 是研究论域 (universe), 即研究对象的集合, $C \cup D = R$ 是属性集合, 子集 C 和 D 分别称为条件属性集和结果属性集, $V = \cup v_r, r \in R$ 是属性值的集合, v_r 表示了某一个属性 $r \in R$ 的属性值范围, f 定义一个信息函数, 即 $f: U \times R \rightarrow V$, 它指定 U 中每一对象 x 的属性值。所以, 一个知识系统 S 的表格表达形式的数据表有时又可称为决策表 (decision table)。

下面给出一些表格法表示知识表达系统的例子,使概念更清晰。

例 1.1 用表格表达计算器的七段显示管数码信息系统,该信息系统的研究对象是数字 $0, 1, \dots, 9$, 系统属性由显示器的元件 a, b, c, d, e, f, g 来表征,如图 1.1 所示。

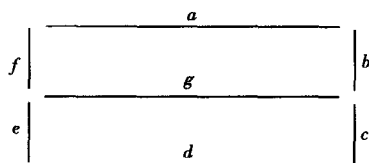


图 1.1 计算器的七段显示管数码信息系统

在该表格表达计算器的七段显示管数码信息系统中, a, b, c, d, e, f, g 7 个元素是条件属性, 描述的 0~9 数字既是研究对象, 又是结果属性。该信息系统的决策表如下表 1.1 所示。

表 1.1 计算器的七段显示管数码信息知识表达系统

U	a	b	c	d	e	f	g
0	1	1	1	1	1	1	0
1	0	1	1	0	0	0	0
2	1	1	0	1	1	0	1
3	1	1	1	1	0	0	1
4	0	1	1	0	0	1	1
5	1	0	1	1	0	1	1
6	1	0	1	1	1	1	1
7	1	1	1	0	0	0	0
8	1	1	1	1	1	1	1
9	1	1	1	1	0	1	1

例 1.2 一个数字电路中二位全加器的真值表如下表 1.2 所示, 它表达了该数字系统的逻辑功能。表中研究对象是加法器的状态 0, 1, ..., 7, 系统属性由数位 a_i, b_i 、求和 S_i 、上次进位 C_{i-1} 和现行进位 C_i 属性来表征。

表 1.2 数字电路中二位全加器的真值表知识表达系统

U	数位 a_i	数位 b_i	上次进位 C_{i-1}	和 S_i	现行进位 C_i
0	0	0	0	0	0
1	0	0	1	1	0
2	0	1	0	1	0
3	0	1	1	0	1
4	1	0	0	1	0
5	1	0	1	0	1
6	1	1	0	0	1
7	1	1	1	1	1

例 1.3 考虑下面的一个知识表达系统

表 1.3 某一知识表达系统

U	a	b	c	d	e
1	1	0	2	2	0
2	0	1	1	1	2
3	2	0	0	1	1
4	1	1	0	2	2
5	1	0	2	0	1
6	2	2	0	1	1
7	2	1	1	2	2
8	0	1	1	0	1

论域由编号为 $1, 2, \dots, 8$ 的8个元素组成, 属性集 $A = \{a, b, c, d, e\}$, 其中子集 $C = \{a, b, c\}$ 称为条件属性集, $D = \{d, e\}$ 称为结果属性集, 而 $v = v_a = v_b = v_c = v_d = v_e = \{0, 1, 2\}$, 根据表中的情况, 可以构成各种分类。

由于是对知识表达系统方法的研究, 对系统的实际语义内容不感兴趣, 以后像例 1.3 一样, 只把它当做数据表来研究, 用完全公式化的方法处理。

1.2 不完整、不精确数据预处理

要从记录数据中获取有用的知识, 必须对所获得的数据进行预处理, 这里的预处理主要介绍不完整、不精确数据的完整化和属性值的离散归一化。

在很多情况下, 得到的数据表是不完整的, 如表 1.4 所示, 包含不完整信息的某一知识表达系统, 表中的数据被丢失了, 无从知道其原始值。

表 1.4 包含不完整信息的某一知识表达系统

研究对象	大小	颜色	年代	等级	销售
1	3.14	红	1970	1.0	N
2	2.17	绿	1492	1.5	Y
3	10.66	红	1814	2.0	Y
4	0.99	红			N
5	0.2	蓝	1776	3.5	N
6		黄	1865	2.5	N
7	4 925.6		1968	6.0	Y

对于包含不完整信息的知识表达系统,可采用下述方法完整化信息表中的不完整数据。

(1) 删除法

根据记录表,简单地将具有空缺属性值的一行删除,从而得到一个完整的信息表。例如,对于表 1.4 所示的包含不完整信息的某一知识系统,通过删除,得到一个表 1.5 所示完整的信息表。

表 1.5 删除后的一个完整信息的知识表

研究对象	大小	颜色	年代	等级	销售
1	3.14	红	1970	1.0	N
2	2.17	绿	1492	1.5	Y
3	10.66	红	1814	2.0	Y
5	0.2	蓝	1776	3.5	N

显然,这种方法的优点是简单方便,在信息表数据量很大,不完整信息的对象数量远远小于完整信息数据数量时,可以采用这种方法。当信息表记录数据有限,不完整信息的对象数量较多时,不可以采用这种方法。

(2) 补偿法

对于包含不完整信息的知识表达系统,补齐一般有三种途径:

a) 将不完整(例如空缺)的属性值作为一种特殊的属性值处理,这将根据具体系统确定特殊的属性值处理方法。

b) 利用统计原理,根据包含不完整信息的知识表中,空缺的属性在其余实例中其取值的分布,按某一种方法,对空缺的属性值进行估计值补充。

均值补偿法(Mean Completer):

均值补偿法就是根据包含不完整信息的知识表达系统中,空缺的属性值由该属性在其他所有实例中的取值的平均值来补充空缺的属性值;如果空缺的属性值是非数值的,就用该属性在其他所有实例中的取值次数最多的属性值来补充空缺的属性值。

例如,对于表 1.4 所示的包含不完整信息的某一知识系统,通过均值补偿,得到一个表 1.6 所示完整的信息表。

表 1.6 通过均值补齐的某一知识表达系统

研究对象	大小	颜色	年代	等级	销售
1	3.14	红	1970	1.0	N
2	2.17	绿	1492	1.5	Y
3	10.66	红	1814	2.0	Y
4	0.99	红	1814	2.8	N
5	0.2	蓝	1776	3.5	N
6	823.88	黄	1865	2.5	N
7	4 925.6	红	1968	6.0	Y