粗糙集对分析理论与决策模型

刘保相　著

科学出版社

粗糙集对分析理论与决策模型

刘保相 著

科学出版社
北京

内 容 简 介

本书共8章，主要内容包括集对分析理论、粗糙集理论、信息系统基本理论、不完备信息系统的集对粗集模型、SPA模糊聚类与决策、SPA格序决策模型、粗糙概念格扩展模型和动态粗决策模型.

本书可供从事粗糙集理论、集对分析理论、信息科学、决策系统和模式识别的相关研究人员及高等院校相关专业的师生参考.

图书在版编目(CIP)数据

粗糙集对分析理论与决策模型/刘保相著. —北京：科学出版社，2010

ISBN 978-7-03-027613-1

Ⅰ.①粗… Ⅱ.①刘… Ⅲ.①集论—研究 Ⅳ.①O144

中国版本图书馆CIP数据核字(2010)第089279号

责任编辑：童安齐 赖文华 王纯刚/责任校对：刘玉靖

责任印制：吕春珉/封面设计：耕者设计工作室

科学出版社 出版

北京东黄城根北街16号

邮政编码：100717

http://www.sciencep.com

双青印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2010年11月第 一 版 开本：B5(720×1000)

2010年11月第一次印刷 印张：12 1/4

印数：1—1 500 字数：200 000

定价：45.00元

(如有印装质量问题，我社负责调换〈双青〉)

销售部电话 010-62134988 编辑部电话 010-62137026(BA08)

前　　言

波兰学者 Pawlak 提出的粗糙集理论是描述和处理模糊和不确定性知识的有效数学工具，它在人工智能、模式识别与智能信息处理等众多领域都有成功的应用. 经典的粗糙集理论的研究对象是具有离散属性值的完备信息系统，即信息系统中的所有属性值都是已知的，但在现实世界中由于多种原因使得在数据获取时往往面临的是不完备的信息系统，即可能存在部分对象的一些属性值未知的情况. 针对不完备信息系统，人们提出了很多处理方法，如删除含有未知值的对象或采用填补的方法用一个已知的值去代替未知值或在不改变原有信息系统的前提下直接处理等. 现有一些学者将集对分析中的联系度引入到不完备信息系统之中，建立了基于联系度的粗糙集模型，但这种模型将信息系统中的所有未知属性值默认为遗漏型来处理的方式需进一步探讨改进.

近年来作者一直从事集对分析、粗糙集理论和模糊数学理论的学习和研究，在《计算机应用研究》、《模糊系统与数学》等期刊上发表"基于 SPA 的不完备信息系统的双向 S-粗集模型"、"基于 SPA 的双枝模糊决策分析"等多篇论文. 2009 年完成了河北省自然科学基金项目"动态 S-粗集决策方法的研究". 本书系统总结了上述研究工作，主要目的是介绍集对分析、粗糙集理论的研究发展及应用状况，探索集对分析和粗糙集的融合，将粗糙集理论与集对论相互嫁接、相互渗透，提出集对分析中的粗糙集方法，用粗集中的上下近似集来定义集对中的同异反联系度，为确定不确定系统的数据分析、数据挖掘、知识发现等提供一种新的理论和方法. 希望本书能为从事粗糙集理论、集对分析理论、信息科学、决策系统和模式识别的有关研究人员提供帮助.

本书在阐述基本概念和方法时，力求概念清晰、内容组织合理、论证严谨、深入浅出、通俗易懂，着力体现出内容广泛、学术思想浓厚和学术观点新颖的特点.

全书共 8 章. 第 1 章介绍了集对分析理论的基本概念和理论基础，较详细地论述了该理论的最新研究成果与应用进展情况；第 2 章详述了粗糙集的基本理论及应用研究；第 3 章总结了动态信息系统的特征及粗糙集理论下的知识表示系统；第 4 章讨论了集对分析和不完备信息系统的相似关系，给出了基于 SPA 的不完备信息系统粗糙集模型；第 5 章推广了双枝模糊决策模型，定义了 IDO 结构联系度，讨论了 IDO 态势序结构在动态决策中的应用；第 6 章给出了格序决策基本方法及 SPA 格序决策模型；第 7 章讨论了粗糙概念格扩展模型的构造，构建了一种格序决策模型；第 8 章提出了一种动态粗决策系统的理论框架，并讨论了其应用前景.

本书第 5、6 章的部分内容来自作者与张春英教授共同研究、发表的学术论文，第 7、8 章的部分内容来自作者与其研究生谢玉婧共同研究的成果；闫红灿教授和李丽红老师为本书做了大量的整理工作；应用数学专业研究生杨亚锋、封丽等校对了全书。在此作者一并对相关人员表示诚挚的谢意.

作者感谢河北省自然科学基金委员会对相关研究工作的支持；感谢河北理工大学以及科学出版社为本书出版提供的的大力帮助.

由于作者水平有限，书中的疏漏之处在所难免，敬请读者给予批评指正.

刘保相

2010 年 2 月

目　　录

第 1 章　集对分析理论

集对分析(set pair analysis, SPA)理论(简称集对论)是一种新型的处理模糊和不确定知识的数学工具,能有效地分析和处理不精确、不一致、不完整等各种不确定信息,并从中发现隐含的知识,揭示潜在的规律[1].

1.1　集对分析的基本概念

所谓集对,就是具有一定联系的两个集合所组成的对子. 集对分析从两个集合的同一性、差异性和对立性三个方面来研究系统的不确定性,其核心思想是:认为任何系统都是由确定性和不确定性信息构成的. 在这个系统中,确定性与不确定性相互联系、相互影响、相互制约,甚至在一定条件下可以相互转化,并用联系度表达式来统一描述.

1.1.1　联系度

定义 1.1.1　给定两个集合 A 和 B,并设这两个集合组成集对 $H=(A,B)$,在某个具体的问题背景 W 下对集对 H 展开分析,共得到 N 个特性,其中,S 个特性为集对 H 中的两个集合 A 和 B 所共同具有,在 P 个特性上集对中两个集合 A 和 B 相对立,在其余的 $F=N-S-P$ 个特性上既不相互对立,又不为这两个集合共同具有,则称比值:S/N 为这两个集合在问题 W 下的同一度,简称同一度;F/N 为这两个集合在问题 W 下的差异度,简称差异度;P/N 为这两个集合在问题 W 下的对立度,简称对立度,并用式子

$$\mu(W)=\frac{S}{N}+\frac{F}{N}i+\frac{P}{N}j \tag{1.1.1}$$

加以统一的表示,式中的 μ 称为集合 A 和 B 的联系度,这个式子称为联系度定义式. 若令 $a=\frac{S}{N}$,$b=\frac{F}{N}$,$c=\frac{P}{N}$,则式(1.1.1)可简写为

$$\mu(W)=a+bi+cj \tag{1.1.2}$$

在不至于引起混淆的情况下,式(1.1.2)可进一步简写成

$$\mu=a+bi+cj \tag{1.1.3}$$

根据联系度的定义以及式(1.1.2)和式(1.1.3)可知 $0\leqslant a,b,c\leqslant 1$,且满足归一化条件,即有关系式

$$a+b+c=1 \tag{1.1.4}$$

据此，有时又可把式(1.1.3)改写为

$$\mu'=a+bi \tag{1.1.5}$$

或

$$\mu'=a+cj \tag{1.1.6}$$

$$\mu'=bi+cj \tag{1.1.7}$$

注 在联系度表达式中，i 和 j 有双重含义：

第一个含义是 i 和 j 分别作为差异度 F/N 和对立度 P/N 的系数. 规定，i 在 $[-1,1]$ 区间视不同情况不确定取值；j 一般情况下规定其取值为 -1，以表示 P/N 与同一度 S/N 相反.

第二个含义是不计较 i 和 j 的取值情况，此时仅起标记的作用，即表示 F/N 是差异度，P/N 是对立度，并以这两个标记与同一度相区别.

在实际分析中，i 和 j 的上述双重含义常常同时起作用，从而使分析或计算过程能方便地进行. 例如，当我们不需要对研究对象作更精细刻画或不去计较不确定系数 i 取什么值时，可把 i 作为差异度的标记处理；当要充分考虑差异度对同一度和对立度的影响时则把 i 作为差异度的系数处理，这时要着重讨论 i 的取值以及由此产生的效应. 上述原则同样适用于对 j 的处理.

1.1.2 联系数

定义 1.1.2 称形如 $a+bi+cj$，$a+bi$，$a+cj$，$bi+cj$ 的数为联系数，其中 a,b,c 为任意正数，$j=-1$，$i\in[-1,1]$且不确定取值.

引进联系度的最初目的是为了应用上的方便，但其理论意义则在于拓广了数的概念. 联系数的意义一方面在于它把可确定数与所在范围相结合，把数与值联系了起来；另一方面联系数把宏观层次上的确定量与微观层次上的不确定量联系起来，构成了一个不确定系统.

1.1.3 联系变量与联系函数

定义 1.1.3 称在宏观层次上随时间 t 等因素变动着的联系数为联系变量. 联系变量可以用 $\mu(t)$，$\mu(x)$，$\mu(y)$加以表示.

定义 1.1.4 如果一个联系变量 $\mu(y)$由另一个联系变量 $\mu(x)$引起，则称$\mu(y)$是 $\mu(x)$的联系函数. 其中 $\mu(x)$称为自变不确定量，$\mu(y)$是因变不确定量.

1.2 联系度的四则运算

联系度 μ 可以根据不同问题的求解要求从不同的角度进行各种运算，以下给

出联系度 μ 的加减乘除四则运算规律.

1.2.1　联系度的加法运算

定义 1.2.1　(加法)设有两个联系度,$\mu_1=a_1+b_1i+c_1j$,$\mu_2=a_2+b_2i+c_2j$,则它们之和

$$\mu_1+\mu_2=(a_1+a_2)+(b_1+b_2)i+(c_1+c_2)j \tag{1.2.1}$$

联系度的加法运算满足交换律和结合律.

$$\mu_1+\mu_2=\mu_2+\mu_1$$

$$(\mu_1+\mu_2)+\mu_3=\mu_1+(\mu_2+\mu_3)=(\mu_1+\mu_3)+\mu_2$$

1.2.2　平均联系度

n 个联系度之和,一般可表示成 n 与这 n 个联系度的算术平均值相乘的形式,即

$$\begin{aligned}\mu_1+\mu_2+\cdots+\mu_n&=n\left(\frac{\mu_1+\mu_2+\cdots+\mu_n}{n}\right)\\&=n\left(\frac{a_1+a_2+\cdots+a_n}{n}+\frac{b_1+b_2+\cdots+b_n}{n}i+\frac{c_1+c_2+\cdots+c_n}{n}j\right)\\&=n(\bar{a}+\bar{b}i+\bar{c}j)\end{aligned}$$

其中$\bar{a}$,$\bar{b}$,$\bar{c}$分别为 n 个同一度、差异度、对立度的平均值,$\bar{a}+\bar{b}i+\bar{c}j$ 称为平均联系度.

1.2.3　联系度的减法运算

定义 1.2.2　(减法)设有两个联系度,$\mu_1=a_1+b_1i+c_1j$,$\mu_2=a_2+b_2i+c_2j$,则它们之差

$$\mu_1-\mu_2=(a_1-a_2)+(b_1-b_2)i+(c_1-c_2)j \tag{1.2.2}$$

在 n 个联系度的和中减去其中一个联系度时,只要按普通多项式相减法则在对应项上相减即可,即

$$\begin{aligned}&n(\bar{a}+\bar{b}i+\bar{c}j)-(a+bi+cj)\\&=(n\bar{a}-a)+(n\bar{b}-b)i+(n\bar{c}-c)j\\&=(n-1)\left[\left(\frac{n\bar{a}-a}{n-1}\right)+\left(\frac{n\bar{b}-b}{n-1}\right)i+\left(\frac{n\bar{c}-c}{n-1}\right)j\right]\end{aligned} \tag{1.2.3}$$

1.2.4　联系度的乘法运算

定义 1.2.3　(乘法)设有两个联系度 $\mu_1=a_1+b_1i+c_1j$,$\mu_2=a_2+b_2i+c_2j$,则它们之积仍是一个联系度 $\mu=a+bi+cj$,记作 $\mu=\mu_1\times\mu_2$,其中

$$a=a_1a_2+b_1b_2+c_1c_2,\ b=a_1b_2+a_2b_1+b_1c_2+b_2c_1,\ c=a_1c_2+a_2c_1$$

用矩阵表示则为

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix}=\begin{pmatrix} a_1 & b_1 & c_1 \\ b_1 & a_1+c_1 & b_1 \\ c_1 & 0 & a_1 \end{pmatrix}\begin{pmatrix} a_2 \\ b_2 \\ c_2 \end{pmatrix}$$

联系度的乘法运算满足乘法交换律.

$$\mu_1\mu_2=\mu_2\mu_1$$

$$\mu_1\mu_2\mu_3=\mu_3\mu_1\mu_2=\mu_2\mu_3\mu_1$$

1.2.5 联系度的加法和乘法混合运算

n 个联系度的加乘混合运算满足分配律.

$$\mu_1(\mu_2+\mu_3)=\mu_1\mu_2+\mu_1\mu_3 \tag{1.2.4}$$

n 个联系度乘积运算以及加乘混合运算可以采取“先简化 μ,再进行运算,并把结果还原”的做法.

例 1.2.1 已知 $\mu_1=0.2+0.3i+0.5j$,$\mu_2=0.8+0.1i+0.1j$,$\mu_3=0.3+0.2i+0.5j$,求 $\mu_1(\mu_2+\mu_3)$.

解 第一步,将 μ_1,μ_2,μ_3 先化简,分别记作

$$\mu'_1=0.2+0.5j$$

$$\mu'_2=0.8+0.1j$$

$$\mu'_3=0.3+0.5j$$

于是原式简化成

$$\mu'_1(\mu'_2+\mu'_3)=(0.2+0.5j)[(0.8+0.1j)+(0.3+0.5j)]$$

第二步,计算

$$\begin{aligned}\mu'_1(\mu'_2+\mu'_3)&=(0.2+0.5j)[(0.8+0.1j)+(0.3+0.5j)]\\&=(0.2+0.5j)[2(0.55+0.3j)]\\&=2(0.2+0.5j)(0.55+0.3j)\\&=2(0.26+0.335j)\end{aligned}$$

第三步,把 $0.26+0.335j$ 还原成 $a+bi+cj$ 的形式,这时,$b=1-0.26-0.335=0.405$,所以 $\mu_1(\mu_2+\mu_3)=2(0.26+0.405i+0.335j)$.

1.2.6 联系度的除法运算

用 n 个联系度的乘积除一个联系度时,也采取“先简化 μ,再进行运算,并把结果还原”的做法.

要注意的是,这里涉及到方程组的求解.对于两个同反联系度相乘,其一般形式可记为

$$(a_1+c_1j)(a_2+c_2j)=a+cj \tag{1.2.5}$$

展开后得

$$(a_1a_2+c_1c_2)+(a_1c_2+a_2c_1)j=a+cj \tag{1.2.6}$$

比较等式两边对应项的系数得方程组为

$$\begin{cases} a_1a_2+c_1c_2=a \\ a_1c_2+a_2c_1=c \end{cases} \tag{1.2.7}$$

于是当方程组中 a,c,a_1,c_1 为已知时,只要代入 a,c,a_1,c_1 的具体值,即可解出 a_2,c_2 的值,由此可给出有关联系度除法的一般步骤是:

第一步,对给出的被除联系度和作为除式的联系度根据 $a+b+c=1$ 这个归一化条件简化成只含两项的同反式.

第二步,利用方程组(1.2.7)解得商式中的 a_2 和 c_2,其中 a,c 为被除式的同一度与对立度,a_1 和 c_1 是除式中的同一度和对立度.

第三步,按 a_2,b_2,c_2 的归一化条件求得 b_2,并把商式写成 $a+bi+cj$ 的形式.

1.3 集对分析在科学研究和工程技术中的应用

集对分析理论是一种较新的软计算方法,可有效地分析和处理不确定信息. 近年来,该理论日益受到学术界的重视,已经在决策、预测、数据融合、不确定性推理、产品设计、网络计划、综合评价等领域得到较为成功的应用. 本节较详细地论述了该理论的最新研究成果与应用进展情况,并指出可能的发展趋势和研究方向.

1.3.1 集对分析在决策中的应用

彭飞等将集对分析引入模糊、灰色物元空间(FHW)决策支持系统中,建立了基于集对分析和模糊、灰色物元空间决策支持系统的方案评价决策方法[2].

假设有 $m(m>0)$ 个专家 $(e_1,e_2,\cdots,e_m)$ 对一个含有 n 个指标 $(h_1,h_2,\cdots,h_n)$ 的方案 H 进行评价[假设专家权重分别为 $(\omega_1,\omega_2,\cdots,\omega_m)$].

若第 k 个专家 e_k 对方案 H 的 n 个指标中有 s_k 个认为“好”,f_{1k} 个认为“较好”,f_{2k} 个认为“一般”,f_{3k} 个认为“较差”,$p_k=n-s_k-f_{1k}-f_{2k}-f_{3k}$ 个认为“差”,则第 k 个专家对方案 H 的同异反决策联系度为

$$\mu(k-H)=\frac{s_k}{n}+\frac{f_{1k}}{n}i_1+\frac{f_{2k}}{n}i_2+\frac{f_{3k}}{n}i_3+\frac{p_k}{n}j$$

简记为

$$\mu(k-H)=a_k+b_{1k}i_1+b_{2k}i_2+b_{3k}i_3+c_kj.$$

从而 m 个专家对方案 H 的评价就形成了一个方案评价决策矩阵. 进一步考虑专家权重,可用加权平均法求解总的同一度 a,总的差异度 b_1,b_2,b_3 和总的对立度 c,从而得到其同异反联系度表达式为

$$\mu(k-H)_{sum}=a+b_1i_1+b_2i_2+b_3i_3+cj$$

最后,用“过半决策准则”或其他常用方法进行评价.

实际应用表明,集对分析可以把专家对方案所做评价的同、异、反各方面因素综合考虑,并将之量化,使不确定性也能够定量地表达;辅之以权重分析系统,以期精确表示各方面因素;最终根据方案评价决策矩阵和相关评价准则来判断方案的优劣.

程启月等给出了集对贴近度(ρ)、集对摆动度(γ)、集对尖锐度(κ)的概念及计算方法以及具有实际意义的联系度公式,给出了同异反决策模型,并应用此决策模型对“射击准备方案”进行定性定量分析,通过给出的理想方案的测度计算各子系统的联系度和系统的联系度,由此排序,达到择优“射击准备方案”的目的[3].

王坚强在分析模糊数学和灰色关联分析方法评价多指标、多水平问题的不足的基础上,针对多指标、多水平系统决策特点,提出了一种新的基于集对分析的多指标、多水平决策模型,并将该模型成功应用于经济决策问题[4].此外,黄良骥则研究了集对分析在交通指挥控制决策支持系统中的应用[5].

1.3.2 集对分析在预测中的应用

高洁等提出了集对分析聚类预测法,该方法融合了集对分析中的同异反模式识别的“择近原则”和聚类分析的基本思路[6].具体步骤为(设待预测的事物为 N,相应的待预测系统为 B):

(1) 确定事物 N 的分类模式系统.

(2) 建立描述事物 N 的分类模式系统与参照系统的同异反联系向量.

(3) 建立描述事物 N 的待预测系统 B 与参照系统的同异反联系向量.

(4) 计算同异反距离.

(5) 确定待预测系统 B 所属的类别.

文献[6]将该方法应用于邮电业务总量预测的研究.研究中考虑了邮电业务总量和第一、二、三产业的国内生产总值之间的关系,利用邮电业务总量和三个产业的生产总值的历史数据,建立了邮电业务量水平聚类预测模型,从而得到邮电业务总量预测结果.利用我国某地区的实际数据进行分析计算,并与其它方法比较,结果表明该预测方法是行之有效的.

1.3.3 集对分析在数据融合中的应用

雷达与电子支援措施(electronic support measurement,ESM)的数据关联是数据融合领域中的主要研究内容之一,具有重要的军事应用价值.由于雷达可以测量目标的完整位置信息(包括距离和方位),而 ESM 通常采用被动式传感器,只能提供关于目标的属性信息和位置中的方位信息,不能提供距离信息,且现代电子战

环境中存在的干扰及其他因素影响，使得在雷达与ESM的数据关联中面临着很大的不确定性. 文献[7]给出了基于模糊集理论的雷达-ESM数据关联算法，该算法的缺点是阈值的选取比较困难. 文献[8]基于集对论和多假设检验(multiple hypothesis testing, MHT)理论研究了在各雷达目标航迹样本容量不相等的情况下利用方位信息进行雷达与ESM航迹关联的问题，给出了新的雷达-ESM数据关联算法. 其核心思想是建立雷达航迹和ESM航迹之间的同异反联系度，并以此来测量雷达航迹和ESM航迹的关联程度. 该算法克服了时间效率低和阈值选取困难的缺点. 仿真结果表明算法特别适用于大数目雷达和ESM航迹样本的情况.

1.3.4　集对分析在不确定性推理中的应用

文献[9]提出一种把确定性推理与不确定性推理结合起来的推理方法—同异反推理法. 文献[10]把联系度μ的乘法运算和i的取值与同异反推理结合起来，给出同异反推理的一种定量推理模式，同时指出：同异反定量推理中选择合适的推理模式相当重要，它直接影响推理精度. 文献[11]则把这一问题归结为“推理路径”的选择，给出了路径选择的一般规则，并用推理实例表明了这一思路的正确性. 文献[11]同时指出存在推理的类同性、由样本推理推断总体特征等需要进一步研究的问题.

1.3.5　集对分析在产品设计中的应用

李志辉等提出了一种基于实例推理，CBR(case based reasoning)的同异反产品设计方法. 该方法的核心在于设计过程中领域专家的经验知识和人类创新思维的信息交合. 这种信息交合是以人类专家为主体，在相应的数学基础上建立系统模型而得到的. 该方法的基本思路是：从现代设计方法论的基本问题入手，围绕三个统一(技术与艺术的统一、功能与形式的统一、微观与宏观的统一)和以人为核心的设计价值观，利用由SPA支撑的CBR系统产生参考案例；然后对其展开同异反分析，在定性描述的基础上用联系数μ定量刻画；进而建立同异反决策矩阵来评价方案的优劣，最终形成产品设计方案[12].

已发表的关于集对论应用的论文已超过400篇. 集对论的应用领域还包括网络计划、综合评价、农业、林业、教育、体育、经济、环境、科研管理、知识创新、军事、交通、电力系统、聚类分析和其他领域[13~39]. 此外，有学者提出了针对小数据集的基于SPA的知识发现算法，但由于现实中的数据库越来越大，有必要研究面向海量数据库的基于SPA的知识发现算法.