第3章

CHAPTER 3

智能计算与识别理论基础

本章给出多源信息融合赖以发展的智能计算与识别理论基础，大部分内容比较新。

3.1 概述

3.1.1 模式识别的一般概念

模式识别是人类自然智能的一种基本形式。所谓“模式”，就是指人类按照时间和空间中可观察的自然属性和认识属性对客观事物的划分，所谓“模式识别”就是依据某些观测获得的属性把一类事物和其他类型的事物区分的过程。例如，人们依据视觉、听觉、触觉等感官所接受的信息，能够正确认识外界事物，并能把一类事物与其他事物正确区分。

然而，我们现在研究的“模式识别”主要属于人工智能的范畴。随着20世纪40年代计算机的出现以及50年代人工智能的兴起，一种以计算机为工具进行“模式识别”的理论和方法便发展起来了，并迅速发展并成为一门新的学科。

目前，模式识别的理论研究主要集中在两方面，一是关于生物体（包括人）感知客观事物机理的研究，这是生物学家、生理学家、心理学家、神经生理学家等的研究内容，属于认知科学的范畴；二是针对给定的任务，关于计算机实现模式识别的有关理论和方法研究，这是数学家、信息学家和计算机科学家的研究内容，属于应用信息科学的范畴。

用计算机实现的模式识别系统，一般由三个相互关联而又有明显区别的过程组成，即数据生成、模式分析和模式分类。所谓数据生成是将原始信息转换为向量，成为计算机易于处理的形式；所谓模式分析是对数据进行加工，包括特征选择、特征提取、数据维数压缩和决定可能存在的类别等；模式分类则是利用模式分析所获得的信息，对计算机进行训练，从而根据判别准则，以实现模式的分类。

现在通用的模式识别方法有两种，即统计模式识别方法和结构（句法）模式识

别方法。所谓统计模式识别方法,就是对模式的统计分类方法,即结合统计概率论的 Bayes 决策系统进行模式识别的技术,又称决策理论识别方法。利用模式与子模式分层结构的树状信息所完成的模式识别工作,就是结构模式识别或句法模式识别。

模式识别系统已经成功地应用于字符识别、语音识别、指纹识别、人脸识别、医学图像分析、工业产品检测等许多方面。然而,模式识别方法的研究目前仍然是针对具体应用对象的,至今还难以发展成为一种统一的模式识别理论与方法。

3.1.2 智能学习与统计模式识别

统计模式识别的基本原理是,根据先验知识,定义**模式空间**为

$$\Omega = \{\omega_1, \omega_2, \cdots, \omega_M\} \tag{3-1-1}$$

其中 ω_k 表示一个模式类;同时可定义**样本特征向量**,或称为待识别模式为

$$X_i = [x_{i1}, x_{i2}, \cdots, x_{id}]^{\mathrm{T}} \in X \subseteq \mathbb{R}^d, \quad i = 1, 2, \cdots, N \tag{3-1-2}$$

其中 X 表示样本特征向量空间,上标 T 表示转置;N 为样本点数;x_{ij} 是一个样本特征,d 为样本特征数;所谓**统计模式识别方法**,就是根据属性或特征来定义某个距离函数以判别样本 X_i属于哪个 ω_k。或者说,把有特征相似性的样本在模式空间中划分为互相接近的若干模式类,并以特征进行“聚类”,从而达到模式分类的目的。

统计模式识别的主要方法有:判别函数法、k 近邻分类法、非线性映射法、特征分析法、主因子分析法等。在统计模式识别中,Bayes 决策规则从理论上解决了最优分类器的设计问题,但其实施却必须首先解决更困难的概率密度估计问题。反向传播(BP)神经网络直接从观测数据(训练样本)学习,是更简便有效的方法,因而获得了广泛的应用。但它是一种启发式技术,缺乏坚实的理论基础。统计推断理论研究所取得的突破性成果导致现代统计学习理论——Vapnik Chervonenkis(VC)理论的建立,该理论不仅在严格的数学基础上圆满地回答了人工神经网络中出现的理论问题,而且导出了一种新的学习方法——支持向量机方法。

所谓智能学习,就是从样本获得分类的一类数学方法,这些方法具有自学习的智能。

20 世纪 70 年代,波兰学者 Pawlak 等提出了**粗糙集(rough sets)**理论,此后,许多科学家在粗糙集的理论和应用方面做了大量的研究工作。至今,粗糙集理论已经成为智能学习的一类标准方法。粗糙集理论是处理模糊和不确定性的一个新的数学工具,用于数据分类分析。粗糙集理论分析的主要目的在于由获取的数据综合出近似的概念。本章 3.2 节将详细讨论粗糙集理论及应用。然而,Pawlak 的粗糙集理论是基于等价关系建立起来的,也就是说分类是严格的,但实际问题往往并不能进行严格分类,于是就引入**广义粗糙集(generalized rough sets)**的概念,并在此基础上建立容差关系的分类方法。70 年代初,Dempster 和 Shafer 建立的一套数学理论,称为 **D-S 证据理论(evidence theory)**,这一理论的核心是超越了概率统计推断的理论框架,可以适应于专家系统、人工智能、模式识别和系统决策等领域的实际问题。而且这一理论的发展很快成为智能学习和多源信息融合的重要组成部分。本章 3.3 节将对 D-S 证据理论进行详细研究。

G. Matheron 于 1975 年出版了《随机集与积分几何学》一书,从而提出了所谓**随机集(random sets)**理论,现在已经成为智能学习的又一重要研究方向。随机集处理的是随机集合问题,集值的引入带来许多令人意想不到的奇特结果,有可能发展成为一种更有力的工具。因为它对异类多源信息融合有特别的意义,本章 3.4 节将专门讨论随机集理论。进而,

我们还将介绍随机有限集理论的基本知识。

20 世纪 90 年代中期，统计学习理论和**支持向量机**（**support vector machine**）算法的深入研究和成功应用引起了广泛的关注。支持向量机算法具有比较坚实的理论基础和良好的推广能力，并在手写数字识别和文本分类等方面取得了良好的应用效果。特别是利用满足 Mercer 条件的核函数实现非线性分类，对于模式识别而言具有划时代的意义。

Bayes 网络是另外一种用于智能推断的工具，本章 3.7 节也将进行简单的讨论。

模式识别发展至今，人们已经建立了各种针对具体应用问题的模型和方法，但仍然没有得到对所有模式识别问题都适用的统一模型和统一方法。我们现在拥有的只是一个工具箱，所要完成的工作就是结合具体的问题，把统计模式识别或句法模式识别结合起来，把模式识别方法与人工智能中的启发式搜索方法结合起来，把模式识别方法与支持向量机的机器学习方法结合起来，把人工神经元网络与各种已有技术以及人工智能中的专家系统、不确定推理方法结合起来，深入掌握各种工具的效能和应有的可能性，互相取长补短，不断开创模式识别应用的新局面。

3.2　粗糙集理论基础

粗糙集理论由 Zdzislaw Pawlak 在 20 世纪 80 年代早期提出[3,4]，用于数据分类分析。这些数据可以来自于量测，也可以来自于人类专家。粗糙集理论分析的主要目的在于由获取的数据综合出近似的概念。

3.2.1　信息系统的一般概念

定义 3.2.1　一个数据表 $\mathcal{A}=(U,A)$ 称为是一个**信息系统**（**information system**），其中 $U=\{x_1,x_2,\cdots,x_n\}$ 称为**对象集**（**set of objects**），$A=\{a_1,a_2,\cdots,a_m\}$ 称为**属性集**（**set of attributes**），都是非空有限集合；而 $a: U\to V_a$，$\forall a\in A$ 描述了对应关系，其中集合 V_a 是属性 a 的**值集**（**value set**）。

例题 3.2.1　设有一个信息系统，示于表 3-2-1，其中有七个对象，两个属性。

表 3-2-1　信息系统 $\mathcal{A}=(U,A)$

案例(U)	年龄段(a_1)	近三年患病次数(a_2)
x_1	16～30	50
x_2	16～30	0
x_3	31～45	1～25
x_4	31～45	1～25
x_5	46～60	26～49
x_6	16～30	26～49
x_7	46～60	26～49

注意，表中案例 3 和 4，以及案例 5 和 7 具有完全相同的条件值。在利用这些属性的前提下，这些案例就是成对难识别的案例。　□

定义 3.2.2　一个**决策系统**（**decision system**）就是形式如 $\mathcal{A}=(U,A\cup\{d\})$ 的信息系统，

其中 $d \notin A$ 称为**决策属性**(**decision attribute**),相应地把 A 中的元素称为**条件属性**(**conditional attribute**),或简称为**条件**(**condition**)。

决策属性通常取值为二值,即"是"或"否"(1 或 0),但也可以取多值。

例题 3.2.2 针对例题 3.2.1,增加决策属性,示于表 3-2-2,其中决策是决定是否复查,所以决策属性取二值。

表 3-2-2 决策系统 $\mathcal{A}=(\boldsymbol{U},\mathbf{A}\cup\{\boldsymbol{d}\})$

案例(U)	年龄段(a_1)	近三年患病次数(a_2)	复查(d)
x_1	16～30	50	是
x_2	16～30	0	否
x_3	31～45	1～25	否
x_4	31～45	1～25	是
x_5	46～60	26～49	否
x_6	16～30	26～49	是
x_7	46～60	26～49	否

请注意,表中案例 3 和 4,以及案例 5 和 7 仍然具有完全相同的条件值。但第一对取相反的决策值,而第二对取相同的决策值。 □

由决策表就可以综合出如下规则:"如果年龄在 16～30 年龄段,近 3 年来患病 50 次,身体复查是必需的"。

在许多应用中,有一种分类结果是已知的,就可以把这种后验知识用决策属性来表示,把这一过程就说成是**有监督学习过程**。

3.2.2 决策系统的不可分辨性

信息系统的知识发现问题本质上是按照属性特征给对象进行分类的问题。为此引入有关关系的概念[1]。

定义 3.2.3 设有对象集 U,其笛卡儿积的子集 $R\subseteq U\times U$,就称为 U 上的一个二元关系 R。进而,如果这个二元关系还具有如下特性

(1) 自返性,如果 $\forall x\in U \Rightarrow (x,x,)\in R$;

(2) 对称性,如果 $(x,y)\in R\Rightarrow(y,x)\in R$;

(3) 传递性,如果 $(x,y),(y,z)\in R\Rightarrow(x,z)\in R$,则称其为**等价关系**。

定义 3.2.4 设 R 是定义在对象集 U 上的一个等价关系,$x\in U$ 的一个**等价类**定义为

$$R(x)\triangleq\{y\in U:(x,y)\in R\} \tag{3-2-1}$$

定义 3.2.5 对象集 U 的某些子集形成的集类 $\mathbb{U}=\{U_i\subseteq U:\ i\in I\}$,如果

(1) $\bigcup\limits_{i\in I} U_i = U$,$I$ 是指标集合;

(2) 对于任意 $i\neq j\Rightarrow U_i\cap U_j=\varnothing$,则称 $\mathbb{U}$ 为 U 的一个**划分**。

引理 1 设 R 是定义在对象集 U 上的一个等价关系,$R(x)$ 是 $x\in U$ 的等价类,如果 $y\in R(x)$,则必然有 $R(y)=R(x)$。

证明 见参考文献 1。 ■

引理 2 设 R 是定义在对象集 U 上的一个等价关系,对于 $\forall x,y\in U$,如果 $R(y)\cap$

$R(x)\neq\varnothing$，则必然有 $R(y)=R(x)$。

证明　见参考文献 1。■

引理 3　设 R 是定义在对象集 U 上的一个等价关系，$R(x)$ 是 $x\in U$ 的等价类，则

$$\bigcup_{x\in U} R(x)=U \tag{3-2-2}$$

证明　见参考文献 1。■

定理 3.2.1　定义在对象集 U 上的一个等价关系 R，确定了 U 对等价类的一个划分；反之，U 的任一划分也定义了 U 上的一个等价关系。

证明　见参考文献 1。■

定义 3.2.6　对象集 U 按等价关系 R 形成的等价类所构成的集类称为 U 关于 R 的**商集**，记为 U/R。

根据定理 3.2.1，商集 U/R 实际上就是 U 按 R 的一个划分。

例题 3.2.3　设 $U=\mathbb{Z}^+$ 表示全体非负整数集合，其上定义了等价关系 R 如下

$$(x,y)\in R\Rightarrow x-y \text{ 能被 3 整除}，\quad \forall x,y\in U$$

容易验证 R 是等价关系：(1)自返性，$\forall x\in U\Rightarrow x-x=0\Rightarrow(x,x)\in R$；(2)对称性，$\forall x,y\in U,(x,y)\in R\Rightarrow x-y$ 能被 3 整除$\Rightarrow y-x$ 能被 3 整除，即$(y,x)\in R$；(3)传递性，$(x,y),(y,z)\in R\Rightarrow x-y,y-z$ 能被 3 整除$\Rightarrow x-z=(x-y)+(y-z)$能被 3 整除，即$(x,z)\in R$。于是可以得到三个等价类

$$R(1)=\{1,4,7,10,\cdots\},\quad R(2)=\{2,5,8,11,\cdots\},\quad R(3)=\{0,3,6,9,\cdots\}$$

而$\{R(1),R(2),R(3)\}$就是 U 按 R 的一个划分，即 $U/R=\{R(1),R(2),R(3)\}$。□

一个决策系统(即决策表)表达了有关模型的所有知识。这个决策表的一部分可能没有必要那样大，因为至少按两种方式衡量有多余，即相同的或不可识别的对象可能被表示了几次，或某些属性可能是多余的。

定义 3.2.7　设$\mathcal{A}=(U,A)$是一个信息系统，与任意 $B\subseteq A$ 相联系的等价关系定义为

$$\mathrm{IND}_{\mathcal{A}}(B)\triangleq\{(x,x')\in U\times U:\ \forall a\in B,a(x)=a(x')\} \tag{3-2-3}$$

称其为 ***B*-不可分辨关系(*B*-indiscernibility relation)**。如果$(x,x')\in\mathrm{IND}_{\mathcal{A}}(B)$，那么按 B 的属性，x 和 x'是不可分辨的。***B*-不可分辨关系的等价类**记为$[x]_B$。

如果不引起混淆的话，B-不可分辨关系的下标$\mathcal{A}$可以略去。

例题 3.2.4　对于表 3-2-1，条件属性 A 的非空子集有 $B_1=\{a_1\}=\{$年龄段$\}$、$B_2=\{a_2\}=\{$近三年患病次数$\}$，以及 $B_3=\{a_1,a_2\}=\{$年龄段，近三年患病次数$\}$，从而有如下三个 B-不可分辨关系

$$\mathrm{IND}(B_1)=\{\{x_1,x_2,x_6\},\{x_3,x_4\},\{x_5,x_7\}\}$$
$$\mathrm{IND}(B_2)=\{\{x_1\},\{x_2\},\{x_3,x_4\},\{x_5,x_6,x_7\}\}$$
$$\mathrm{IND}(B_3)=\{\{x_1\},\{x_2\},\{x_3,x_4\},\{x_5,x_7\},\{x_6\}\}$$

而 $[x_1]_{B_1}=\{x_1,x_2,x_6\}$，$[x_1]_{B_2}=\{x_1\}$；$[x_6]_{B_2}=\{x_5,x_6,x_7\}$，$[x_6]_{B_3}=\{x_6\}$，等等。□

3.2.3　集合近似

定义 3.2.8　设$\mathcal{A}=(U,A)$是一个信息系统，$B\subseteq A$，$X\subseteq U$，定义 U 的子集合

$$\underline{B}(X)\triangleq\{x\in U:[x]_B\subseteq X\} \tag{3-2-4}$$

称为集合 X 的 **B-下近似(*B*-lower approximation)**；而 U 的另一子集合

$$\overline{B}(X) \triangleq \{x \in U : [x]_B \cap X \neq \varnothing\} \tag{3-2-5}$$

称为集合 X 的 **B-上近似(*B*-upper approximation)**。

$\underline{B}(X)$中的对象是基于 B 中的知识确定性地分类成 X 中的元；而 $\overline{B}(X)$中的对象是基于 B 中的知识仅仅分类成 X 中可能的元。

定义 3.2.9 集合

$$\mathrm{BN}_B(X) \triangleq \overline{B}(X) - \underline{B}(X) \tag{3-2-6}$$

称为 X 的 **B-边界区域(*B*-boundary region)**,是由不能基于 B 中的知识确定性地分类成 X 中元的对象组成。而

$$\mathrm{OT}_B(X) \triangleq U - \overline{B}(X) \tag{3-2-7}$$

称为 X 的 **B-外部区域(*B*-outside region)**,是由基于 B 中的知识确定性地分类成不属于 X 中元素的对象组成。

如果对象集合 U 一个子集 X 的 B-边界区域非空,则称其为**粗糙集(rough set)**,或简称为**粗集**；否则,如果其 B-边界区域是空集,则称其为**明晰集(crisp set)**。

例题 3.2.5 最常见的情况是按照条件属性综合决策类。在表 3-2-2 中,设结果集是

$$W = \{x \in U: d(x) = \{是\}\} = \{x_1, x_4, x_6\} \subset U$$

于是得到：$\underline{A}(W) = \{x_1, x_6\}$是集合 W 的 A-下近似,表示其等价类在 W 中的对象集合；$\overline{A}(W) = \{x_1, x_3, x_4, x_6\}$是集合 W 的 A-上近似,表示其等价类与 W 相交非空的对象集合；$\mathrm{BN}_A(W) = \{x_3, x_4\}$是 W 的 A-边界区域,是由不能基于 A 中的知识确定性地分类成 W 中元素的对象组成；$\mathrm{OT}_A(W) = \{x_2, x_5, x_7\}$,$W$ 的 A-外部区域,是由基于 A 中的知识确定性地分类成不属于 W 中元素的对象组成。所以 W 是一个粗集,因为它的边界区域非空。 □

集合近似图形表示如图 3-2-1 所示。

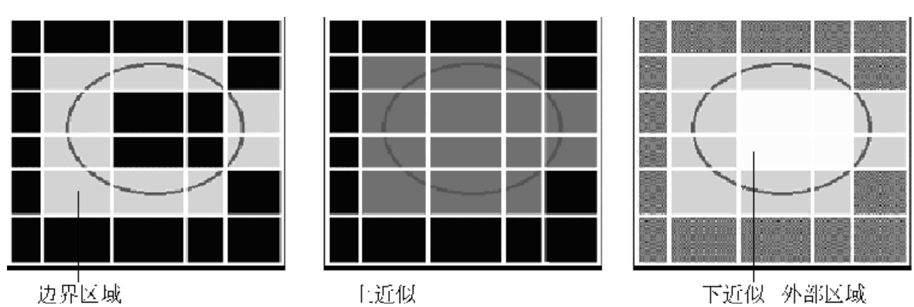

图 3-2-1 集合近似的图形表示

定理 3.2.2 设$\mathcal{A} = (U, A)$是一个信息系统,$B \subseteq A$,$X \subseteq U$,则有：

(1) $\underline{B}(X) \subseteq X \subseteq \overline{B}(X)$ (3-2-8)

(2) $\underline{B}(\varnothing) = \overline{B}(\varnothing) = \varnothing$, $\underline{B}(U) = \overline{B}(U) = U$ (3-2-9)

(3) $\overline{B}(X \cup Y) = \overline{B}(X) \cup \overline{B}(Y)$, $\underline{B}(X \cap Y) = \underline{B}(X) \cap \underline{B}(Y)$ (3-2-10)

(4) $X \subseteq Y \Rightarrow \underline{B}(X) \subseteq \underline{B}(Y)$ 且 $\overline{B}(X) \subseteq \overline{B}(Y)$ (3-2-11)

(5) $\underline{B}(X \cup Y) \supseteq \underline{B}(X) \cup \underline{B}(Y)$, $\overline{B}(X \cap Y) \subseteq \overline{B}(X) \cap \overline{B}(Y)$ (3-2-12)

(6) $\underline{B}(U - X) = U - \overline{B}(X)$, $\overline{B}(U - X) = U - \underline{B}(X)$ (3-2-13)

(7) $\underline{B}(\underline{B}(X)) = \overline{B}(\underline{B}(X)) = \underline{B}(X)$, $\overline{B}(\overline{B}(X)) = \underline{B}(\overline{B}(X)) = \overline{B}(X)$ (3-2-14)

证明　我们只证明式(3-2-8)和式(3-2-10)，其他可以类似证明。

(1) $x\in\underline{B}(X)\Rightarrow[x]_B\subseteq X\Rightarrow x\in X\Rightarrow[x]_B\cap X\neq\varnothing\Rightarrow x\in\overline{B}(X)$

(3) $x\in\overline{B}(X\cup Y)\Rightarrow\exists x\in U,[x]_B\cap(X\cup Y)\neq\varnothing\Rightarrow[x]_B\cap X\neq\varnothing$ 或 $[x]_B\cap Y\neq\varnothing$

$\Rightarrow x\in\overline{B}(X)$　或　$x\in\overline{B}(Y)\Rightarrow x\in\overline{B}(X)\cup\overline{B}(Y)$

$x\in\underline{B}(X\cap Y)\Rightarrow\exists x\in U,[x]_B\subseteq(X\cap Y)\Rightarrow[x]_B\subseteq X$ 且 $[x]_B\subseteq Y\Rightarrow x\in\underline{B}(X)$，且 $x\in\underline{B}(Y)\Rightarrow x\in\underline{B}(X)\cap\underline{B}(Y)$ ■

显然，集合的下近似和上近似分别是由不可分辨关系产生的拓扑上集合的内部和闭包。图 3-2-2 给出了表 3-2-2 决策系统集合近似特性的图形表示。

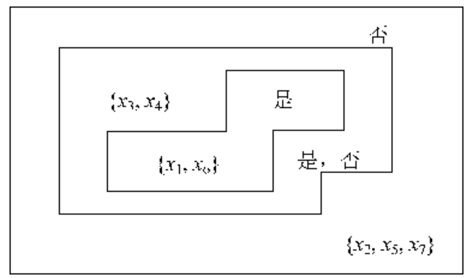

图 3-2-2　表 3-2-2 决策系统集合近似特性的图形表示

定义 3.2.10　可以定义如下四种粗集：

(1) X 是**粗糙 *B*-可定义的**(**roughly *B*-definable**)，当且仅当 $\underline{B}(X)\neq\varnothing$ 且 $\overline{B}(X)\neq U$；

(2) X 是**内部 *B*-不可定义的**(**internally *B*-undefinable**)，当且仅当 $\underline{B}(X)=\varnothing$ 且 $\overline{B}(X)\neq U$；

(3) X 是**外部 *B*-不可定义的**(**externally *B*-undefinable**)，当且仅当 $\underline{B}(X)\neq\varnothing$ 且 $\overline{B}(X)=U$；

(4) X 是**完全 *B*-不可定义的**(**totally *B*-undefinable**)，当且仅当 $\underline{B}(X)=\varnothing$ 且 $\overline{B}(X)=U$。

X 是粗糙 B-可定义的集合，就意味着借助于集合 B，就能确定 U 中的某些元属于 X，以及 U 中的另外一些元属于 $U-X$。

X 是内部 B-不可定义的集合，就意味着利用集合 B，我们就能确定 U 中的某些元属于 $U-X$，但不能确定 U 中的任何元是否属于 X。

X 是外部 B-不可定义的集合，就意味着利用集合 B，我们就能确定 U 中的某些元属于 X，但不能确定 U 中的任何元是否属于 $U-X$。

X 是完全 B-不可定义的集合，就意味着利用集合 B，我们既不能确定 U 中的任何元是否属于 X，也不能确定 U 中的任何元是否属于 $U-X$。

定义 3.2.11　粗集也可以用如下系数进行数值表示：

$$\alpha_B(X)=\frac{|\underline{B}(X)|}{|\overline{B}(X)|}\tag{3-2-15}$$

称为**近似精度**(**accuracy of approximation**)，其中 $|X|$ 表示集合 X 的势(当其为有限集时，就是所包含元的个数)；而

$$1-\alpha_B(X)\tag{3-2-16}$$

称为**近似粗糙度**(**roughness of approximation**)。

显然，$0\leqslant\alpha_B(X)\leqslant 1$。如果 $\alpha_B(X)=1$，X 关于属性集合 B 是明晰集，或称关于 B 是**精确的**(**precise**)；否则，$\alpha_B(X)<1$，X 关于属性集合 B 是粗集，或称关于 B 是**含糊的**(**vague**)。

3.2.4　属性约简

前一节我们研究了等价类的数据约简问题，即利用给定的属性，在对象集中建立不可分辨的等价类，即完成聚类。我们面临的另外一类问题是尽可能使条件属性集变小，并保留原

有属性集的分类能力,这就是**属性约简**(**attribute reduction**)的问题。然而,计算等价类是容易的,求得最小约简子集(即在所有约简子集中具有最小集合势的约简子集)却是一个 NP 难题。事实上,这正是把粗集理论应用于实际的一个瓶颈问题。已经发展了一些启发式算法,除了属性集合很大的情况外,可以在能接受的时间范围内计算充分多的约简子集。

例题 3.2.6 考虑表 3-2-3 的决策表,$\mathcal{A}'=(U,\{$文凭,经历,外语,介绍人意见$\}\cup\{$决策$\})$,我们只考虑条件属性,即信息系统$\mathcal{A}=(U,\{$文凭,经历,外语,介绍人意见$\})$。为了简单起见,每个等价类只包含一个元。

表 3-2-3 招聘问题:一个不可约简的决策表

对象(U)	文凭(d)	经历(e)	外语(f)	介绍人意见(r)	决策(dc)
x_1	MBA	中	通过	很好	接收
x_2	MBA	低	通过	不明确	拒绝
x_3	MCE	低	通过	好	拒绝
x_4	MSc	高	通过	不明确	接收
x_5	MSc	中	通过	不明确	拒绝
x_6	MSc	高	通过	很好	接收
x_7	MBA	高	否	好	接收
x_8	MCE	低	否	很好	拒绝

注:MBA:工商硕士;MCE:土木工程硕士;MSc:硕士。

似乎有一个最小的属性集$\{e,r\}$,用它识别目标与利用全部属性集识别目标是一样的。不妨利用全部属性集和利用属性集$\{e,r\}$来建立不可分辨关系,二者结果是一样的。具有这种特性的最小属性集的结构马上就显现出来了。 □

定义 3.2.12 给定一个信息系统$\mathcal{A}=(U,A)$,它的一个**约简**(**reduct**),就是一个最小属性集 $B\subseteq A$,使得 $\mathrm{IND}_{\mathcal{A}}(B)=\mathrm{IND}_{\mathcal{A}}(A)$。

换句话说,一个约简集就是属性集 A 中能保留对总对象集合划分特性的最小属性集合。因此,约简集的分类能力与总属性集的分类能力相同。

定义 3.2.13 给定一个具有 n 个对象的信息系统$\mathcal{A}=(U,A)$,它的**可分辨矩阵**(**discernibility matrix**)就是一个 $n\times n$ 阵 $\boldsymbol{C}=\{c_{ij}\}$,且

$$c_{ij}=\{a\in A:a(x_i)\neq a(x_j)\},\quad i,j=1,2,\cdots,n \tag{3-2-17}$$

其中每个元都由一组属性构成,取决于对象 x_i 与 x_j 属性的不同。

定义 3.2.14 给定信息系统$\mathcal{A}=(U,A)$,它的**可分辨函数** $f_{\mathcal{A}}$ 定义为 m 个 Boolean 变量 $a_1^*,a_2^*,\cdots,a_m^*$(相应于属性 $a_1,a_2,\cdots,a_m$)的 Boolean 函数,定义为

$$f_{\mathcal{A}}(a_1^*,a_2^*,\cdots,a_m^*)=\wedge\{\vee c_{ij}^*:1\leqslant j\leqslant i\leqslant n,c_{ij}\neq\varnothing\} \tag{3-2-18}$$

其中 $c_{ij}^*=\{a^*:a\in c_{ij}\}$;而$\wedge$表示逻辑"与"运算,$\vee$表示逻辑"或"运算。

注 $f_{\mathcal{A}}$ 的所有**主蕴涵**(**prime implicant**)构成的集合,确定了$\mathcal{A}$的所有约简构成的集合。所谓 Boolean 函数 f 的一个蕴涵,就是字符(变量或者其"非")间的某种关联,使得在变量的任意值 v 条件下字符取值为"真"时,函数 f 在 v 条件下的取值也为"真"。主蕴涵就是最小的蕴涵。因此,我们只对单调 Boolean 函数的蕴涵感兴趣,即这些函数不是由"非"来确定。

例题 3.2.7 仍考虑表 3-2-3 的决策表所定义的信息系统$\mathcal{A}=(U,A)$,它的可分辨矩阵如表 3-2-4 所示。

表 3-2-4　可分辨矩阵 $C=\{c_{ij}\}$

	$[x_1]$	$[x_2]$	$[x_3]$	$[x_4]$	$[x_5]$	$[x_6]$	$[x_7]$	$[x_8]$
$[x_1]$	$\varnothing$							
$[x_2]$	e,r	$\varnothing$						
$[x_3]$	d,e,r	d,r	$\varnothing$					
$[x_4]$	d,e,r	d,e	d,e,r	$\varnothing$				
$[x_5]$	d,r	d,e	d,e,r	e	$\varnothing$			
$[x_6]$	d,e	d,e,r	d,e,r	r	e,r	$\varnothing$		
$[x_7]$	e,f,r	e,f,r	d,e,f	d,f,r	d,e,f,r	d,f,r	$\varnothing$	
$[x_8]$	d,e,f	d,f,r	f,r	d,e,f,r	d,e,f,r	d,e,f	d,e,r	$\varnothing$

而可分辨函数是

$$\begin{aligned}f_{\mathcal{A}}(d,e,f,r)=&(e\vee r)(d\vee e\vee r)(d\vee e\vee r)(d\vee r)(d\vee e)(e\vee f\vee r)(d\vee e\vee f)\\&(d\vee r)(d\vee e)(d\vee e)(d\vee e\vee r)(e\vee f\vee r)(d\vee f\vee r)\\&(d\vee e\vee r)(d\vee e\vee r)(d\vee e\vee r)(d\vee e\vee f)(f\vee r)\\&(e)(r)(d\vee f\vee r)(d\vee e\vee f\vee r)\\&(e\vee r)(d\vee e\vee f\vee r)(d\vee e\vee f\vee r)\\&(d\vee f\vee r)(d\vee e\vee f)\\&(d\vee e\vee r)\end{aligned}$$

其中每个字符相应于每个属性，而每个括号内就是 Boolean 表达式的联结；而“与”运算符号均省略。经过函数的简化计算，结果就是 $f_{\mathcal{A}}(d,e,f,r)=er$（即 $e\wedge r$）。这与前述直观的约简结果是一致的。

注意，可分辨函数中的每一行相应于可分辨矩阵的每一列。而可分辨矩阵是一个对称阵，对角元为空集。于是，倒数第二行就意味着第六个对象（更确切地说是第六个等价类），可以根据文凭（d）、外语（f）和介绍人意见（r）中的任意一个属性与第七个对象进行识别；可以根据文凭（d）、经历（e）和外语（f）中的任意一个属性与第八个对象进行识别。 □

定义 3.2.15　如果在构造 Boolean 函数时，把 Boolean 变量的联结限制在可分辨矩阵中的 k 列，而不是所有的列，得到的函数称为 **k-相对可分辨函数**（**k-relative discernibility function**）；而由 k-相对可分辨函数得到的主蕴涵集，就确定了 $\mathcal{A}=(U,A)$ 的所有 **k-相对约简**（**k-relative reducts**）集。

这样的约简展现了由其他对象识别 $x_k\in U$（更确切地说是 $[x_k]\subseteq U$）所需要的最小信息量。

利用上述概念，有监督学习问题（即分类结果已知的问题）就是求取决策 d 的值，这个值应当赋予新的对象，而这个对象的描述借助于条件属性。我们经常要求用来定义对象的属性集达到最小，例如，表 3-2-3 中出现了两个最小属性集｛经历（e），介绍人意见（r）｝和｛文凭（d），经历（e）｝，都唯一地定义了对象隶属的决策类。相应的可分辨函数是相对于决策的。现在给出这个概念的形式化描述。

定义 3.2.16　设决策系统 $\mathcal{A}=(U,A\cup\{d\})$ 给定，映像 $d(U)=\{k:d(x)=k,x\in U\}$ 的势称为 d 的**秩**（**rank**），记为 $r(d)$，此处假定决策 d 的值集为 $V_d=\{v_d^{(1)},v_d^{(2)},\cdots,v_d^{(r(d))}\}$。

相当多的决策系统中 d 的秩是 2，例如，表 3-2-3 中的决策属性中就是｛是，否｝或｛接

收，拒绝}。然而 d 的秩可以是任意自然数。例如，表 3-2-3 中的决策属性中可以是{接收，保留，拒绝}，则 d 的秩就成为 3。

定义 3.2.17 决策集 d 确定了对象集 U 的一个划分

$$\mathrm{CLASS}_{\mathcal{A}} = \{X_{\mathcal{A}}^{(1)}, X_{\mathcal{A}}^{(2)}, \cdots, X_{\mathcal{A}}^{(r(d))}\}$$

其中等价类 $X_{\mathcal{A}}^{(k)} = \{x \in U: d(x) = v_d^{(k)}\}, k=1,2,\cdots,r(d)$；这个划分被称为决策系统**按决策的对象分类（classification of objects determined by the decision）**；相应地称 $X_{\mathcal{A}}^{(k)}$ 为$\mathcal{A}$**的第 k 个决策类（k-th decision class）**，且等价类 $X_{\mathcal{A}}(u) = \{x \in U: d(x) = d(u)\}, \forall u \in U$。

例题 3.2.8 考虑表 3-2-2 的决策表，有如下按决策的对象分类

$$\mathrm{CLASS}_{\mathcal{A}} = \{X_{\mathcal{A}}^{(是)}, X_{\mathcal{A}}^{(否)}\}, \quad X_{\mathcal{A}}^{(是)} = \{x_1, x_4, x_6\}, \quad X_{\mathcal{A}}^{(否)} = \{x_2, x_3, x_5, x_7\}$$

再考虑表 3-2-3 的决策表所定义的决策系统，则有如下按决策的对象分类

$$\mathrm{CLASS}_{\mathcal{A}} = \{X_{\mathcal{A}}^{(收)}, X_{\mathcal{A}}^{(拒)}\}, \quad X_{\mathcal{A}}^{(收)} = \{x_1, x_4, x_6, x_7\}, \quad X_{\mathcal{A}}^{(拒)} = \{x_2, x_3, x_5, x_8\}$$

□

定义 3.2.18 设$\{X_{\mathcal{A}}^{(1)}, X_{\mathcal{A}}^{(2)}, \cdots, X_{\mathcal{A}}^{(r(d))}\}$是决策系统$\mathcal{A}$的决策类，则集合

$$\underline{B}(X_{\mathcal{A}}^{(1)}) \cup \underline{B}(X_{\mathcal{A}}^{(2)}) \cup \cdots \cup \underline{B}(X_{\mathcal{A}}^{(r(d))}) \tag{3-2-19}$$

称为是$\mathcal{A}$**的 B-正区域（B-positive region）**，记为$\mathrm{POS}_B(d)$。

例题 3.2.9 考虑决策表 3-2-2 所定义的决策系统$\mathcal{A}=(U, A\cup\{d\})$，则有如下结果：$\underline{A}(X_{\mathcal{A}}^{(是)}) \cup \underline{A}(X_{\mathcal{A}}^{(否)}) \neq U$，这是因为对于对象 x_3, x_4，决策并不唯一。再考虑决策表 3-2-3 所定义的决策系统$\mathcal{A}=(U, A\cup\{d\})$，则有 $\underline{A}(X_{\mathcal{A}}^{(收)}) \cup \underline{A}(X_{\mathcal{A}}^{(拒)}) = U$，因为此时所有的决策都是唯一的。 □

定义 3.2.19 决策系统的这一重要特性可以形式化描述如下：设$\mathcal{A}=(U, A\cup\{d\})$是一个决策系统，而$\mathcal{A}$中的**广义决策（generalized decision）**定义为函数$\partial_A: U \rightarrow P(V_d)$，而

$$\partial_A(x) = \{i: \exists x' \in U, 使得(x, x') \in \mathrm{IND}(A), 且\ d(x) = i\} \tag{3-2-20}$$

决策系统$\mathcal{A}$称为是**相容的（确定性的）**，如果对任意 $x \in U$，总有$|\partial_A(x)|=1$；否则，决策系统$\mathcal{A}$称为是**不相容的（不确定性的）**。

容易看出，一个决策系统$\mathcal{A}$是相容的，当且仅当 $\mathrm{POS}_A(d)=U$。进而，对于任意一对非空集合 $B, B' \subseteq A$，如果$\partial_B = \partial_{B'}$，则 $\mathrm{POS}_B(d) = \mathrm{POS}_{B'}(d)$。

例题 3.2.10 考虑决策表 3-2-2 所定义的决策系统$\mathcal{A}=(U, A\cup\{d\})$，其 **A**-正区域是 U 的一个子集；而表 3-2-3 所定义的决策系统$\mathcal{A}=(U, A\cup\{d\})$，其 **A**-正区域是整个 U。前者是非确定性的，而后者是确定性的。 □

前面我们已经引入了 k-相对可分辨函数的概念，因为决策属性是如此重要，对此引入专门的定义是非常有用的。

定义 3.2.20 设$\mathcal{A}=(U, A\cup\{d\})$是一个相容的决策系统，而 $\boldsymbol{M}(\mathcal{A})=\{c_{ij}\}$是其可分辨矩阵，构造一个新的矩阵

$$\boldsymbol{M}^d(\mathcal{A}) = \{c_{ij}^d\} \tag{3-2-21}$$

并假定

$$c_{ij}^d = \begin{cases} \varnothing, & 如果\ d(x_i) = d(x_j) \\ c_{ij} - \{d\}, & 其他 \end{cases} \tag{3-2-22}$$

矩阵 $\boldsymbol{M}^d(\mathcal{A})$称为系统$\mathcal{A}$的**决策相对可分辨矩阵（decision-relative discernibility matrix）**。与

由可分辨矩阵构造可分辨函数相类似，可以由决策相对可分辨矩阵构造**决策相对可分辨函数**（**decision-relative discernibility function**）$f_M^d(\mathcal{A})$。

有关文献已经证明，$f_M^d(\mathcal{A})$的主蕴涵定义了$\mathcal{A}$的所有**决策相对约简**（**decision-relative reducts**）的集合。

例题 3.2.11　仍考虑表 3-2-3 所示的决策，重排后如表 3-2-5 所示。该表用来说明如何构造$\mathcal{A}$的决策相对可分辨矩阵和决策相对可分辨函数。为方便起见，把所有接收的行重排在上部，而所有拒绝的行重排在下部。

表 3-2-5　重排的决策表

	文凭	经历	外语	介绍人意见	决策
x_1	MBA	中	通过	很好	接收
x_4	MSc	高	通过	不明确	接收
x_6	MSc	高	通过	很好	接收
x_7	MBA	高	否	好	接收
x_2	MBA	低	通过	不明确	拒绝
x_3	MCE	低	通过	好	拒绝
x_5	MSc	中	通过	不明确	拒绝
x_8	MCE	低	否	很好	拒绝

由这个决策相对可分辨矩阵得到的简化决策相对可分辨函数是 $f_M^d(\mathcal{A})=ed\vee er$。由决策相对可分辨矩阵的定义知，选择其中的一个列，如相应于$[x_1]$的列，简化并给出最小函数，以识别$[x_1]$所属的决策类。例如，第一列给出了 Boolean 函数是：$(e\vee r)(d\vee e\vee r)(d\vee r)(d\vee e\vee f)$，经化简后变为 $ed\vee rd\vee re\vee rf$。读者可以检验："如果介绍人意见是很好，外语是通过，则接收"，这就是 x_1 的情况；其实，对于别的对象如 x_6 也是同样情况。

相应的决策相对可分辨矩阵示于表 3-2-6，这是一个对称矩阵，其中对角元素为空集，而所有决策相同的元素也是空集。

表 3-2-6　决策相对可分辨矩阵

	$[x_1]$	$[x_4]$	$[x_6]$	$[x_7]$	$[x_2]$	$[x_3]$	$[x_5]$	$[x_8]$
$[x_1]$	$\varnothing$							
$[x_4]$	$\varnothing$	$\varnothing$						
$[x_6]$	$\varnothing$	$\varnothing$	$\varnothing$					
$[x_7]$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$				
$[x_2]$	e,r	d,e	d,e,r	e,f,r	$\varnothing$			
$[x_3]$	d,e,r	d,e,r	d,e,r	d,e,f	$\varnothing$	$\varnothing$		
$[x_5]$	d,r	e	e,r	d,e,f,r	$\varnothing$	$\varnothing$	$\varnothing$	
$[x_8]$	d,e,f	d,e,f,r	d,e,f	d,e,r	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$

□

3.2.5　粗糙隶属度

定义 3.2.21　一个具有特征函数的集合称为**明晰集**（**crisp set**），即 U 是对象集合（或论域），$S\subseteq U$ 是子集，$\chi_S: U\to\{0,1\}$是**特征函数**（**characteristic function**），则