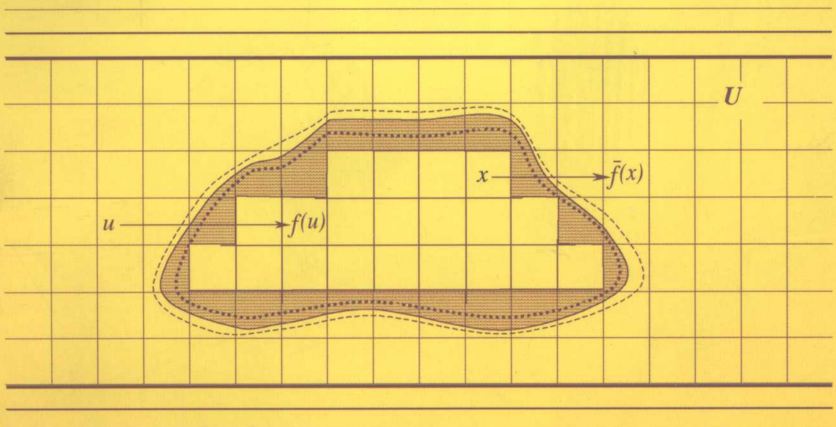
S-粗集与粗决策

史开泉 崔玉泉 著

国 家 自 然 科 学 基 金 资 助

山 东 大 学 聊 城 大 学 出 版 基 金 资 助



科学出版社
www.sciencep.com

S-粗集与粗决策

史开泉　崔玉泉　著

国家自然科学基金资助

山东大学
聊城大学　出版基金资助

科学出版社

北京

内 容 简 介

本书给出了S-粗集的两类形式：单向S-粗集和双向S-粗集以及S-粗集的遗传、记忆特征，提出S-粗决策，给出S-决策分析模型；S-粗集与生命科学、人工智能系统交叉，进行了多视角的讨论. 提出了函数S-粗集的两类形式：函数单向S-粗集和函数双向S-粗集，并给出函数S-粗集在系统（金融系统，投资系统）中规律挖掘、规律发现的讨论，这些讨论是开展粗系统理论与应用研究的理论基础. 该书的特点是视野宽，视角新，学科渗透性强.

本书适合数学、经济、金融、管理、系统科学等专业的大学生、研究生、教师以及广大工程技术研究人员阅读参考.

图书在版编目（CIP）数据

S-粗集与粗决策/史开泉，崔玉泉著. —北京：科学出版社，2006
ISBN 7-03-016544-6

Ⅰ. S… Ⅱ. ①史…②崔… Ⅲ. 粗决策 Ⅳ. O175

中国版本图书馆CIP数据核字（2005）第141450号

责任编辑：张 扬 姚庆爽/责任校对：钟 洋
责任印制：安春生/封面设计：陈 敬

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
*
2006年5月第 一 版 开本：B5（720×1000）
2006年5月第一次印刷 印张：13
印数：1—3 000 字数：241 000

定价：32.00元

序　　言

在经济、金融、管理、工程等若干领域中，人们经常接收到这样的信息，这类信息不能用精确的集合表示，从它们的特征上看，这类信息是粗糙的，或称为粗信息.因为不能用精确的数学方法表示这类信息，使得人们对它的认识能力与对它的开发使用能力降低，人们为此而丢失了这类重要的信息资源.在计算机与计算技术广泛使用的今天，因为没有合适的数学方法描述这类信息，利用计算机去识别这类信息遇到了困难，一些重要的应用研究因此而搁浅.波兰数学家 Z. Pawlak 教授 1982 年提出粗集(rough sets)，给出粗集的一般性研究. Z. Pawlak 教授的工作给人们研究粗信息和它的特性提供了理论支持.由此，对粗信息的理论研究与应用研究得到广泛地开展并取得了很多优秀的成果.粗集的应用涉猎到下列一些领域:系统管理、系统识别、系统状态分析、系统决策、金融系统、风险投资分析系统、数据挖掘、知识发现等.粗集已成为粗系统理论与应用研究的基石.

先观察 Z. Pawlak 教授提出的粗集$(R_-(X),R^-(X))$，这里 X 是元素论域 U 上的元素集合，$X\subset U$，$R_-(X)$是 $X\subset U$ 的下近似，$R^-(X)$是 $X\subset U$ 的上近似，R 是元素等价关系.我们从 Z. Pawlak 粗集的结构能够得到:①如果给定集合 $X\subset U$，R-元素等价类，则 $R_-(X)$，$R^-(X)$就确定，粗集$(R_-(X),R^-(X))$就确定.②不允许集合 $X\subset U$(或 R-元素等价类$[x]_R$)之内的元素离开集合 X；也不允许集合 $X\subset U$(或R-元素等价类$[x]_R$)之外的元素进入集合 X.容易看到:Z. Pawlak 粗集是具有静态特性的集合 X 的粗集.因此，Z. Pawlak 粗集的应用范围是十分有限的.特别是对于动荡不定的金融系统、投资系统、智能系统、目标多变的识别-分析系统、动态数据挖掘、动态知识发现的研究遇到困难.

下面是一个通俗的例子:假期中，$[x]_A$，$[x]_B$，$[x]_C$ 是赴 A，B，C 三地旅游的人组成的集合($[x]_A$，$[x]_B$，$[x]_C$可以看成是赴 A，B，C 三地的等价类).因为受到诱惑，一些不想赴 A 地旅游的人，加入到$[x]_A$ 中，使得$[x]_A$ 的边界向外扩张(card$([x]_A)$变大).因为一些原因，$[x]_B$ 中的某些人取消了本次旅游的打算，这些人离开$[x]_B$，使得$[x]_B$ 的边界向内收缩(card$([x]_B)$变小).因为某些原因，$[x]_C$ 中的某些人取消了赴 C 地旅游的打算，一些不打算去 C 地旅游的人改变了主意，这些人参加到$[x]_C$ 中，使得$[x]_C$ 的边界向内收缩又向外扩张，或者 card$([x]_C)$变小同时 card$([x]_C)$变大.前两者使得集合 $X\subset U$ 的边界具有了单向动态特性($[x]_A$ 向外扩张，$[x]_B$ 向内收缩)，后者使得集合 $X\subset U$ 的边界具有了双向动态特性($[x]_C$ 向内收缩又向外扩张).如果集合 $X\subset U$ 具有了动态特性(单向动态特性，双向动态特性)，那么具有动态特性的集合 $X\subset U$ 具有粗集吗?如果有，这个粗集是一个什么

样子？利用 Z. Pawlak 教授的工作，2002 年史开泉教授给出动态集合的描述，以此提出 S-粗集（singular rough sets）. S-粗集具有两类形式：单向 S-粗集（one direction singular rough sets）和双向 S-粗集（two direction singular rough sets），本书给出 S-粗集的一般性讨论. 因为 S-粗集具有动态特性，对于动态数据挖掘、动态知识发现、系统动态粗特性的研究，S-粗集提供了理论支持. 人们自然要问：Z. Pawlak 粗集、S-粗集能够应用于系统中的规律挖掘、规律发现研究吗？回答是否定的. 这是因为 Z. Pawlak 粗集、S-粗集的共同特征是：它们都是以 R-元素等价类 $[x]$ 定义的，R-元素等价类 $[x]$ 不具有规律特征. **什么是规律？一个函数就是一个规律**. 利用 Z. Pawlak 粗集，S-粗集对数据挖掘、知识发现的研究固然有重要的理论意义与应用意义；过热的投资系统中隐藏着“投资破产的杀机”，令人沮丧的规律一旦从系统中挖掘出来并被人们认识，将减少人间悲剧的发生. 在 S-粗集的基础上，2005 年史开泉教授利用动态函数集合 $Q\subset\mathscr{D}(Q=Q^{\circ}, Q=Q^{*})$ 提出函数 S-粗集（function singular rough sets），函数 S-粗集具有两类形式：函数单向 S-粗集（function one direction singular rough sets）和函数双向 S-粗集（function two direction singular rough sets），本书给出函数 S-粗集的一般性讨论. 函数 S-粗集对于变化多端的金融系统、风险投资系统、系统状态识别系统、故障诊断系统等诸多系统中的规律识别，系统状态的规律生成及规律挖掘、规律发现的研究，提供了理论支持和新的研究思想.

如果浓缩本书的内容和研究思想，读者能够得到这样的认识：Z. Pawlak 粗集是 S-粗集的特例，S-粗集是函数 S-粗集的特例；反之，函数 S-粗集是 S-粗集的一般形式，S-粗集是 Z. Pawlak 粗集的一般形式.

应该回答读者一个问题：在 Z. Pawlak 粗集 $(R_{-}(X), R^{-}(X))$ 中，集合 $X\subset U$ 是静态的. 静态集是动态集的特例. 在工程、经济、金融诸多系统中，人们遇到的集合 $X\subset U$（信息集合）多是动态的. 为了找回被静态集合丢掉的动态特性并把动态特性返还给静态集合 $X\subset U$，这里用“singular”一词表示这种“返还”，因为这个原因，才有本书中的 S-粗集、函数 S-粗集这两个名称与概念.

本书内容分作 8 章，第 1 章简单介绍 Z. Pawlak 粗集，作为本书的概念、研究思路的导引；从第 2 ~ 7 章展开了对 S-粗集，函数 S-粗集的讨论，讨论了 S-粗集的结构、S-粗集的分解-还原、S-粗集的遗传、S-粗集的记忆、函数 S-粗集的结构与生成、函数 S-粗集的应用、S-粗决策等；第 8 章给出 S-粗集与其他学科交叉、渗透、嫁接的讨论. 这些内容对于粗集的理论研究与应用研究具有启发性、可借鉴性. 本书的内容对于工程系统中的诸多领域的应用工作者具有重要的参考价值，对于那些正在攻读各个领域的硕士学位、博士学位的年轻一代具有启迪性. 这是因为，作者在每一章的开头都给出一段文字作为读该章的导引，引导读者去想问题，去讨论问题. 作者试图利用本书的内容搭建沟通智能系统、生命系统、经济系统、管理系统、识别系统之间的桥梁，因为这些系统都具有遗传、记忆特征. 作者试图为这些系统的研

究提供一个新思路、新手段. 如果本书能为这些系统的研究帮一点忙，作者将深感欣慰.

本书仅是对 S-粗集、函数 S-粗集的理论与应用研究的开始，书中涉猎的内容具有良好的后继理论研究，后继应用研究的前景. 或许利用本书的讨论，读者将对粗集概念的内涵有更深的理解，将获得一些更新的、富有创意的成果，我们祈盼着这些新成果的问世. 本书的取材是我们在最近几年发表与待发表的 50 余篇学术论文，我们的学生王洪凯博士、胡海清博士、刘华文博士、管延勇博士、张萍硕士等发表的论文也纳入本书中. 为了能把 Z. Pawlak 粗集用最通俗的语言进行介绍又能把国内的一些重要研究介绍给读者，本书中引用国内一些朋友的研究，这里向他们致以深深的感谢！

在本书撰著过程中，我们的学生胡海清博士、王洪凯博士、张萍硕士、崔明辉硕士等不辞劳苦地为本书排版与打印，向他们表示感谢. 感谢国家自然科学基金委的支持. 山东大学数学与系统科学学院院长、长江学者、博士生导师刘建亚教授对本书的出版给予了热情的支持，这里表示深深的感谢！感谢聊城大学科研处、数学院为完成本书终稿提供帮助. 感谢科学出版社的编辑们为本书的问世提供了多方面的支持与帮助.

因为作者的学识浅陋，书中的不足与错误请同行给予批评、指正.

史开泉　崔玉泉

2005 年 9 月

目　　录

第 1 章　Z. Pawlak 粗集的概念与应用

§1.1　Z. Pawlak 粗集与它的结构

设 U 是一个有限元素论域,X 是 U 上的元素集合,$X\subset U$,R 是 U 上的元素等价关系,$[x]$是 R-元素等价类,如图 1.1 所示.

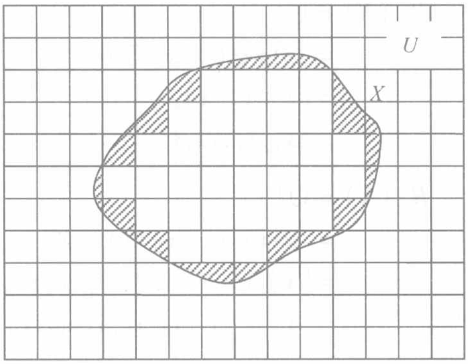


图 1.1

图中的每一个小方块是 R-元素等价类$[x]$,阴影中的白色方块是 $X\subset U$ 的下近似 $R_-(X)$

定义 1.1.1　称 $R_-(X)$是集合 $X\subset U$ 的下近似,而且

$$\begin{aligned} R_-(X) &= \cup[x] \\ &= \{x \mid x\in U,[x]\subseteq X\} \end{aligned} \tag{1.1.1}$$

定义 1.1.2　称 $R^-(X)$是集合 $X\subset U$ 的上近似,而且

$$\begin{aligned} R^-(X) &= \cup[x] \\ &= \{x \mid x\in U,[x]\cap X\neq\varnothing\} \end{aligned} \tag{1.1.2}$$

定义 1.1.3　由 $R_-(X)$,$R^-(X)$构成的集合对,称作 $X\subset U$ 的 R-粗集,简称 $X\subset U$ 的粗集,而且

$$(R_-(X),R^-(X)) \tag{1.1.3}$$

定义 1.1.4　称 $B_{nR}(X)$是 $X\subset U$ 的 R-边界,而且

$$B_{nR}(X)=R^-(X)-R_-(X) \tag{1.1.4}$$

例　设论域 $U=\{x_0,x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8,x_9,x_{10}\}$,$U$ 上的 R-元素价类

$[x]_1,[x]_2,[x]_3,[x]_4,[x]_5$,而且

$$\begin{aligned}
[x]_1 &= \{x_0,x_1\}\\
[x]_2 &= \{x_2,x_6,x_9\}\\
[x]_3 &= \{x_3,x_5\}\\
[x]_4 &= \{x_4,x_8\}\\
[x]_5 &= \{x_7,x_{10}\}
\end{aligned} \tag{1.1.5}$$

取 U 上的子集 $X=\{x_0,x_3,x_4,x_5,x_8,x_{10}\}\subset U$,则有 $X\subset U$ 的下近似 $R_-(X)$,上近似 $R^-(X)$;而且

$$\begin{aligned}
R_-(X) &= \cup[x]\\
&= \{x \mid x\in U,[x]\subseteq X\}\\
&= [x]_3\cup[x]_4\\
&= \{x_3,x_4,x_5,x_8\}
\end{aligned} \tag{1.1.6}$$

$$\begin{aligned}
R^-(X) &= \cup[x]\\
&= \{x \mid x\in U,[x]\cap X\neq\varnothing\}\\
&= [x]_1\cup[x]_3\cup[x]_4\cup[x]_5\\
&= \{x_0,x_1,x_3,x_4,x_5,x_7,x_8,x_{10}\}
\end{aligned} \tag{1.1.7}$$

$X\subset U$ 的粗集$(R_-(X),R^-(X))$是

$$(R_-(X),R^-(X))=\{\{x_3,x_4,x_5,x_8\},\{x_0,x_1,x_3,x_4,x_5,x_7,x_8,x_{10}\}\} \tag{1.1.8}$$

为了容易接受 Z. Pawlak 粗集的概念,这里利用一个通俗的例子给出解释.

图 1.2 是一块长方形的牛皮,它们由许多块小长方形组成.

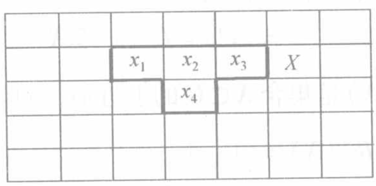


图 1.2

具有规则边界的集合 X

如果图 1.2 中的每一块长方形的牛皮可做一双皮鞋,图 1.2 中粗线所包围的牛皮可做 4 双皮鞋,它们分别用 x_1,x_2,x_3,x_4 表示,如图 1.2 所示. 所做成的皮鞋用集合 X 表示,则有 $X=\{x_1,x_2,x_3,x_4\}$. 这里的集合 X 是我们在数学分析、高等代数、高等数学等课程中经常遇到的,它是一个精确集,换句话说,粗线所包围的牛皮能做而且只能做 4 双皮鞋,所用的皮革不多不少. 在生活与实际中,我们见到的牛皮

是否都是方方正正的？回答是否定的. 我们给出图 1.3.

图 1.3 是一块自然形状的牛皮.

如果制作皮鞋的用皮尺寸不变，图 1.3 中曲线围成的牛皮（白色方块）只能做 4 双皮鞋，它们用 x_1，x_2，x_3，x_4 表示（长方形），如图 1.3 所示；图中带有阴影的部分是做 4 双皮鞋剩下来的牛皮（俗称边角料）. 人们自然想到，阴影部分可以通过拼接的方式，使它成为长方形的牛皮，它们也可以做皮鞋使用，例如，阴影部分，通过拼接，大约可做 2.53 双皮鞋. 因此，图 1.3 中的曲线边界围成的牛皮大约可做 6.53 双皮鞋.

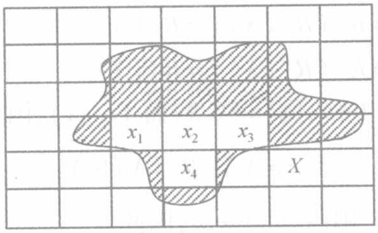


图 1.3

具有非规则边界的集合 X

比较图 1.2 和图 1.3，容易得到这样的事实：图 1.2 能用精确的整数表示皮鞋的数目（4 双）；图 1.3 不能用精确的整数表示皮鞋的数目（6.53 双）. 图 1.2 能用普通集（精确集）表示，或者 $X=\{x_1,x_2,x_3,x_4\}$；图 1.3 不能用普通集（精确集）表示. 这个通俗的例子，一般人都能接受.

如果我们把图 1.2 或图 1.3 中的每一个长方形看成是一个由有限点（元素）构成的 R-元素等价类 $[x]$，则图 1.2 可以用 R-元素等价类 $[x]$ 精确表示，图 1.3 不可以用 R-元素等价类 $[x]$ 精确表示. 等价类是数学中的一个普通概念.

对于图 1.3，如何用集合的概念来描述；或者，用集合的概念如何去表达图 1.3？波兰数学家 Z. Pawlak 教授于 1982 年提出粗集（rough sets）[1] 的概念. 由此，数学中又增加了一个新的理论与应用分支. 粗集理论及其应用在最近几年中，在国际、国内成为研究的热点，有很多重要的理论与应用研究问世.

§1.2　集合 X 的下近似与上近似关系[2,3]

给定集合 $X,Y\subset U$，$R_-(X)$，$R^-(X)$ 分别是 $X\subset U$ 的下近似，上近似；$R_-(Y)$，$R^-(Y)$ 分别是 $Y\subset U$ 的下近似，上近似；则有

1°　$R_-(X)\subseteq X\subseteq R^-(X)$　　(1.2.1)

2°　$R_-(\varnothing)=R^-(\varnothing)=\varnothing$，$R_-(U)=R^-(U)=U$　　(1.2.2)

3°　$R^{-}(X\cup Y)=R^{-}(X)\cup R^{-}(Y)$　(1.2.3)

4°　$R_{-}(X\cap Y)=R_{-}(X)\cap R_{-}(Y)$　(1.2.4)

5°　$X\subseteq Y\Rightarrow R_{-}(X)\subseteq R_{-}(Y)$　(1.2.5)

6°　$X\subseteq Y\Rightarrow R^{-}(X)\subseteq R^{-}(Y)$　(1.2.6)

7°　$R_{-}(X\cup Y)\supseteq R_{-}(X)\cup R_{-}(Y)$　(1.2.7)

8°　$R^{-}(X\cap Y)\subseteq R^{-}(X)\cap R^{-}(Y)$　(1.2.8)

9°　$R_{-}(\sim X)=\sim R^{-}(X)$　(1.2.9)

10°　$R^{-}(\sim X)=\sim R_{-}(X)$　(1.2.10)

11°　$R_{-}(R_{-}(X))=R^{-}(R_{-}(X))=R_{-}(X)$　(1.2.11)

12°　$R^{-}(R^{-}(X))=R_{-}(R^{-}(X))=R^{-}(X)$　(1.2.12)

证明　1°　若 $x\in R_{-}(X)$,则$[x]\subseteq X$;而 $x\in[x]$,因此

$$x\in X \text{ 且 } R_{-}(X)\subseteq X$$

若 $x\in X$,因为 $x\in[x]\cap X$,则$[x]\cap X\neq\varnothing$,因此

$$x\in R_{-}(X) \text{ 且 } X\subseteq R_{-}(X)$$

2°　因为空集包含在每一个集合中,$R_{-}(\varnothing)\subseteq\varnothing$而且$\varnothing\subseteq R_{-}(\varnothing)$,所以

$$R_{-}(\varnothing)=\varnothing$$

假设 $R^{-}(\varnothing)\neq\varnothing$. 则存在 x,而且 $x\in R^{-}(\varnothing)$,则有

$$[x]\cap\varnothing\neq\varnothing$$

因为$[x]\cap\varnothing=\varnothing$,与假设矛盾,所以

$$R^{-}(\varnothing)=\varnothing$$

由 1°,$R_{-}(U)\subseteq U$,若使 $U\subseteq R_{-}(U)$,对于 $x\in U$,因$[x]\subseteq U$,则 $x\in R_{-}(U)$,所以

$$R_{-}(U)=U$$

由 1°,$R^{-}(U)\supset U$,但 $R^{-}(U)\subseteq U$,所以

$$R^{-}(U)=U$$

3°　若 $x\in R^{-}(X\cup Y)$,当且仅当$[x]\cap(X\cup Y)\neq\varnothing\Rightarrow([x]\cap X)\neq\varnothing\vee[x]\cap Y\neq\varnothing\Rightarrow x\in R^{-}(X)\vee x\in R^{-}(Y)\Rightarrow x\in R^{-}(X)\cup R^{-}(Y)\Rightarrow R^{-}(X\cup Y)=R^{-}(X)\cup R^{-}(Y)$.

4°　若 $x\in R_{-}(X\cap Y)$,当且仅当$[x]\subseteq X\cap Y\Rightarrow[x]\subseteq X\wedge[x]\subseteq Y\Rightarrow x\in R_{-}(X)\cap R_{-}(Y)\Rightarrow R_{-}(X\cap Y)=R_{-}(X)\cap R_{-}(Y)$.

5°　当且仅当 $X\cap Y=X$,由 4°得到 $R_{-}(X\cap Y)=R_{-}(X)$. 因为 $R_{-}(X)\cap R_{-}(Y)=R_{-}(X)\Rightarrow R_{-}(X)\subseteq R_{-}(Y)$.

6°　当且仅当 $X\cup Y=Y, X\subseteq Y\Rightarrow R^-(X\cup Y)=R^-(Y)$. 由 3°知 $R^-(X)\cup R^-(Y)=R^-(Y)\Rightarrow R^-(X)\subseteq R^-(Y)$.

7°　因为 $X\subseteq X\cup Y$ 而且 $Y\subseteq X\cup Y\Rightarrow R_-(X)\subseteq R_-(X\cup Y)$, 又因为 $R_-(Y)\subseteq R_-(X\cup Y)\Rightarrow R_-(X)\cup R_-(Y)\subseteq R_-(X\cup Y)$.

8°　因为 $X\cap Y\subseteq X$ 而且 $X\cap Y\subseteq Y\Rightarrow R^-(X\cap Y)\subseteq R^-(X)$, 又因为 $R^-(X\cap Y)\subseteq R^-(Y)\Rightarrow R^-(X\cap Y)\subseteq R^-(X)\cap R^-(Y)$.

9°　当且仅当 $[x]\subseteq X, [x]\cap(\sim X)=\varnothing\Rightarrow x\bar{\in} R^-(\sim X)$ 而且 $x\in R_-(X)\Rightarrow x\in\ \sim R^-(\sim X)\Rightarrow R_-(X)=\ \sim R^-(\sim X)$.

10°　在 9°中用 $\sim X$ 代替 X, 则有 $R^-(X)=\ \sim R_-(\sim X)$.

11°　由 1°得到 $R_-(R_-(X))\subseteq R_-(X)$, $x\in R_-(X)\Rightarrow[x]\subseteq X\Rightarrow R_-([x])\subseteq R_-(X)$, 因为 $R_-([x])=[x]\Rightarrow[x]\subseteq R_-(X)$ 而且 $x\in R_-(R_-(X))\Rightarrow R_-(X)\subseteq R_-(R_-(X))$. 由 1°得到 $R_-(X)\subseteq R^-(R_-(X))$, 当 $x\in R^-(R_-(X))\Rightarrow[x]\cap R_-(X)\neq\varnothing$, 存在 $y\in[x]$, $y\in R_-(X)\Rightarrow[y]\subseteq X$, 因此 $[x]=[y]\Rightarrow[x]\subseteq X$ 而且 $x\in R_-(X)\Rightarrow R^-(R_-(X))\subseteq R_-(X)$.

12°　若 $x\in R^-(R^-(X))$, $[x]\cap R^-(X)\neq\varnothing$, 对于某些 $y\in[x]$, $y\in R^-(X)\Rightarrow[y]\cap X\neq\varnothing$, 而且 $[y]=[x]\Rightarrow[x]\cap X\neq\varnothing\Rightarrow x\in R^-(X)\Rightarrow R^-(X)\supset R^-(R^-(X))$ 且 $R^-(X)\subset R^-(R^-(X))\Rightarrow R^-(R^-(X))=R^-(X)$.

由 1°得到 $R_-(R^-(X))\subseteq R^-(X)$, 若 $R_-(R^-(X))\supset R^-(X)$ 成立, 当 $x\in R^-(X)$, $[x]\cap X\neq\varnothing\Rightarrow[x]\subseteq R^-(X)$. 因为, 若 $y\in[x]$, 则 $[y]\cap X=[x]\cap X\neq\varnothing$, 即 $y\in R^-(X)\Rightarrow x\in R_-(R^-(X))$, 因此 $R_-(R^-(X))\supset R^-(X)$.

§1.3　知识的属性依赖发现

Z. Pawlak 粗集是以 R-元素等价类 $[x]$ 定义的, R 是属性集. 例如, 属性 α_1 = 红色, α_2 = 甜味, 属性集 $R=\{\alpha_1,\alpha_2\}$; 具有属性 α_1,α_2 的苹果 x_1,x_2,x_3,x_4 构成关于属性 α_1,α_2 的 R-元素等价类 $[x]_{(\alpha_1,\alpha_2)}=\{x_1,x_2,x_3,x_4\}$; 元素 x_1,x_2,x_3,x_4 关于属性 α_1,α_2 不可分辨, 记作

$$\underset{\alpha_1,\alpha_2}{\mathrm{IND}}(\{x_1,x_2,x_3,x_4\}) \tag{1.3.1}$$

因此, 等价关系 R 称作不可分辨关系.

如果在属性集 R 中再增加一个属性 α_3 = 产地山东, 则有属性集 $R'=\{\alpha_1,\alpha_2,\alpha_3\}$, R'-元素等价类 $[x]_{(\alpha_1,\alpha_2,\alpha_3)}=\{x_2,x_3\}$, 或者

$$\underset{\alpha_1,\alpha_2,\alpha_3}{\mathrm{IND}}(\{x_2,x_3\}) \tag{1.3.2}$$

如果属性集 $R'=\{\alpha_1,\alpha_2,\alpha_3\}$ 中删除属性 α_2,α_3, 则有属性集 $R''=\{\alpha_1\}$, R''-元素等

价类$[x]_{(\alpha_1)}=\{x_2,x_2,x_3,x_4,x_5,x_6,x_7,x_8\}$,或者

$$\underset{\alpha_1}{\mathrm{IND}}(\{x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8\}) \tag{1.3.3}$$

这个简单的事实告诉我们:如果属性集 R 中的属性减少,则 R-元素等价类$[x]$中的元素个数增加;如果属性集 R 中的属性增多,则 R-元素等价类$[x]$的元素个数减少. 显然,属性集 R 中的属性个数变化,引起 R-元素等价类$[x]$中元素个数的变化. 因此,R-元素等价类$[x]$具有颗粒特征. 一个 R-元素等价类$[x]$称作一个知识$[x]$. 因此,R-元素等价类$[x]$与知识$[x]$是两个等价概念,知识$[x]$具有颗粒特征.

容易得到:

命题1　知识$[x]$具有颗粒特征. 依赖于属性集 R 中的属性增加,知识$[x]_{R'}$从知识$[x]_R$中被挖掘,$R\subset R'$.

命题2　依赖于属性集 R 中的属性减少,知识$[x]_{R''}$依赖于知识$[x]_R$被发现,$R''\subset R$.

例　设$[x]_R=\{x_1,x_2,x_3,x_4,x_5\}$是 U 上的知识,$R=\{\alpha_1,\alpha_2,\alpha_3\}$是知识$[x]_R$的属性集,若存在属性 α_4,而且 $R'=R\cup\{\alpha_4\}=\{\alpha_1,\alpha_2,\alpha_3,\alpha_4\}$是知识$[x]_{R'}$的属性集,而且$[x]_{R'}=\{x_2,x_4,x_5\}$. 显然$[x]_{R'}\subset[x]_R$. 知识$[x]_{R'}$是依赖于对属性集 R 的属性补充得到的. 在未对属性集 R 进行属性补充之前,$\{x_2,x_4,x_5\}$是潜藏在$\{x_1,x_2,x_3,x_4,x_5\}$中,因为对属性集的属性补充,使得$\{x_2,x_4,x_5\}$从$\{x_1,x_2,x_3,x_4,x_5\}$中被挖掘出来. 这个简单的事实告诉人们如何从庞大的数据中去寻找人们所需要的数据.

在§1.4~§1.8中,引入苗夺谦先生的工作[4].

§1.4　知识的颗粒特征

近似空间

近似空间是指一个二元序对 $A=(U,R)$,U 是一个非空有限集,称作论域,R 是一个二元等价关系,$R\subseteq U\times U$,也称作 U 上的一个不可分辨关系. $U/R=\{[u]_R \mid u\in U\}$,表示 R 在 U 上的一个划分,由 U 中每个对象 u 所在的 R-等价类$[u]_R$组成.

知识库

$K=(U,R)$称作知识库,其中 R 是一等价关系. 对于$\forall X\subseteq R$, 由 X 产生的等价关系记作 $\mathrm{IND}(X)$,$\mathrm{IND}(X)=\bigcap_{R\in X}R$. 它表达了智能体利用知识库中的一部分知识 X 所能达到的最高的认知程度. $\mathrm{IND}(R)$表达了知识库 $K=(U,R)$的最高的分辨

表示程度.

知识的粒度

设 $K=(U,R)$ 是一知识库, $R\subseteq\mathscr{R}$ 是一等价关系, R 称作知识; $R\subseteq U\times U$.

定义 1.4.1 称 $\mathrm{GD}(R)$ 是知识 $R\subseteq\mathscr{R}$ 的粒度,如果

$$\mathrm{GD}(R)=|R|/|U|^2 \tag{1.4.1}$$

其中 $|R|$ 表示 $R\subseteq U\times U$ 的基数.

当 R 是相等关系,即 $R=\omega$ 时, R 的粒度达到最小值 $|U|/|U|^2=1/|U|$.

当 R 是论域关系,即 $R=\delta$ 时, R 的粒度达到最大值 $|U|^2/|U|^2=1$.

一般情况下, $1/|U|\leqslant\mathrm{GD}(R)\leqslant 1$. 知识的粒度能够表达对知识的分辨能力, $\mathrm{GD}(R)$ 越小,分辨能力越强. 当 $(u,v)\in R$ 时,表明对象 u,v 在 R 下不可分辨,属于 R 的同一个等价类. 否则,它们可分辨,属于不同的 R-等价类. 因此 $\mathrm{GD}(R)$ 表示在 U 中随机选择两个对象,这两个对象 R-不可分辨的可能性的大小. 可能性越大,即 $\mathrm{GD}(R)$ 越大,表明 R 的分辨能力越弱,否则越强.

定义 1.4.2 称 $\mathrm{DIS}(R)$ 是知识 R 的分辨度,如果

$$\mathrm{DIS}(R)=1-\mathrm{GD}(R) \tag{1.4.2}$$

显然 $0\leqslant\mathrm{DIS}(R)\leqslant 1-1/|U|$.

命题 1 若 R 是知识库 $K=(U,R)$ 中的知识,而且 $U/R=\{X_1,X_2,\cdots,X_n\}$,则

$$\mathrm{GD}(R)=\sum_{i=1}^{n}|X_i|^2/|U|^2 \tag{1.4.3}$$

证明 $\forall(u,v)\in R$,有 $u,v\in$ 某一 X_i,故 R 的元是通过在 U/R 的每个等价类 X_i 中任取二元(可以相同)组成的,自然有 $|R|=\sum_{i=1}^{n}|X_i|^2$,所以 $\mathrm{GD}(R)=|R|/|U|^2=\sum_{i=1}^{n}|X_i|^2/|U|^2$.

由命题 1 得到: $\mathrm{DIS}(R)=1-\mathrm{GD}(R)=1-\sum_{i=1}^{n}|X_i|^2/|U|^2$. 分辨度 $\mathrm{DIS}(R)$ 的大小直接反映了对知识的分辨能力.

知识粒度、分辨度与熵的关系

我们知道,熵值也是知识颗粒状的一种度量. 设 $K=(U,R)$ 是一知识库, $R\subseteq\mathscr{R}$ 是一知识,在 R 对 U 形成均匀划分的情况下, R 的熵值 $H(R)$ 较大. 而此时,知识的粒度 $\mathrm{GD}(R)$ 较小(这相当于和一定的几个自然数,当它们彼此接近时平方之和较小),分辨度 $\mathrm{DIS}(R)$ 较大.

表 1.1 说明,当 R 由最粗的论域关系 δ 变为最细的相等关系 ω 时,熵值 $H(R)$ 由 0 增大到 $\log_2(|U|)$,而分辨度由 0 增大到 $1-1/|U|$.

表 1.1　$H(R)$,GD(R)与 DIS(R)的比较

R	δ	ω	一般的 R
$H(R)$	0	$\log_2(\|U\|)$	$0\leqslant H(R)\leqslant\log_2(\|U\|)$
GD(R)	0	$1/\|U\|$	$1/\|U\|\leqslant \mathrm{GD}(R)\leqslant 1$
DIS(R)	0	$1-1/\|U\|$	$0\leqslant \mathrm{DIS}(R)\leqslant 1-1/\|U\|$

§1.5　知识粒度与属性的依赖

信息系统与粒度

称序对 $I=(U,A)$ 是信息系统,其中 U 是一有限对象集,称作论域. A 是有限属性集. $\forall a\in A$,定义 U 上的一个等价关系 θ_a: $u\theta_a v\Leftrightarrow a(u)=a(v)$,其中 $a(u)$ 表示对象 $u\in U$关于属性 $a(\in A)$的值. 容易验证:这样定义的二元关系是一等价关系. θ_a 在 U 上产生的划分是 $U/\theta_a=\{[u]_{\theta_a}|u\in U\}$. 为方便计,常将 U/θ_a 记为 U/a. 因此,$[u]_{\theta_a}$也可写为$[u]_a=\{v\in U|a(u)=a(v)\}$,由 U 中所有与 u 的 a-不可分辨的对象 v 组成. 设 $X\subseteq A$ 是一个属性子集,由 X 产生的不可分辨关系是 $\mathrm{IND}(X)=\bigcap_{x\in X}\theta_x$,将 $U/\mathrm{IND}(X)$简记为 U/X. 因此,在以下的论述中,等价关系、属性、特征、知识、划分等概念不加区别地直接使用它们.

命题 1　设 $I=(U,A)$是一信息系统,$X,Y\subseteq A$,则有:

1° 若 $X\to Y$,则 $\mathrm{GD}(X)\leqslant\mathrm{GD}(Y)$.　　(1.5.1)

2° 若 $X\leftrightarrow Y$,则 $\mathrm{GD}(X)=\mathrm{GD}(Y)$.　　(1.5.2)

证明　1°由 $X\to Y$ 知 $\mathrm{IND}(X)\subseteq\mathrm{IND}(Y)$,$|\mathrm{IND}(X)|\leqslant|\mathrm{IND}(Y)|$,而 $\mathrm{GD}(X)=\mathrm{GD}(\mathrm{IND}(X))=|\mathrm{IND}(X)|/|U|^2$,$\mathrm{GD}(Y)=\mathrm{GD}(\mathrm{IND}(Y))=|\mathrm{IND}(Y)|/|U|^2$,从而有 $\mathrm{GD}(X)\leqslant\mathrm{GD}(Y)$.

2° 由 $X\leftrightarrow Y$ 知,$X\to Y$ 且 $X\leftarrow Y$,因此 $\mathrm{GD}(X)\leqslant\mathrm{GD}(Y)$且 $\mathrm{GD}(Y)\leqslant\mathrm{GD}(X)$,从而有 $\mathrm{GD}(X)=\mathrm{GD}(Y)$.

利用命题 1 得到:

命题 2　设 $I=(U,A)$是一信息系统,$X,Y\subseteq A$,则有:

1° 若 $X\to Y$,则 $\mathrm{DIS}(X)\geqslant\mathrm{DIS}(Y)$.　　(1.5.3)

2° 若 $X\leftrightarrow Y$,则 $\mathrm{DIS}(X)=\mathrm{DIS}(Y)$.　　(1.5.4)

特别地,若 $X\subseteq Y\subseteq A$,$Y\to X$,从而有 $\mathrm{GD}(Y)\leqslant\mathrm{GD}(X)$且 $\mathrm{DIS}(Y)\geqslant\mathrm{DIS}(X)$,说明对于 A 的属性子集,随着属性的增加粒度减小,分辨度增加.

属性重要度

设 $I=(U,A)$是一信息系统,$X\subseteq A$ 是一属性子集,$x\in A$ 是一属性,我们考虑 x