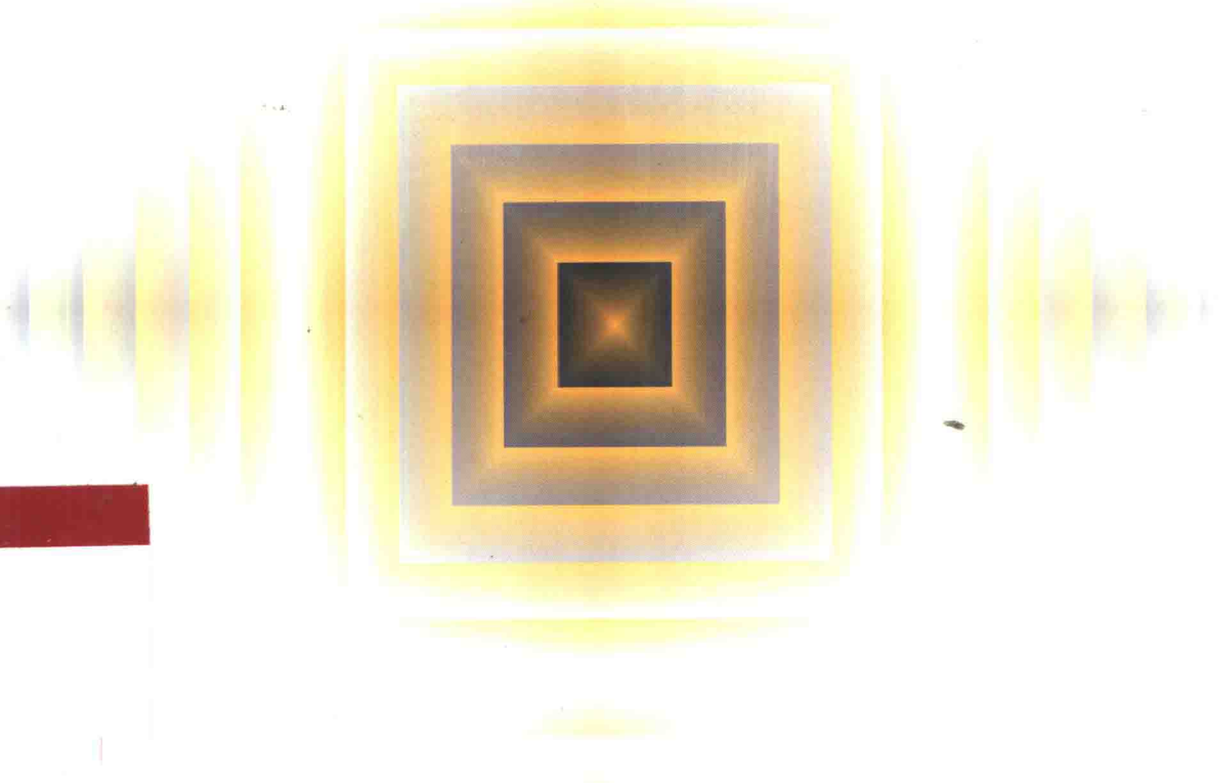
粗糙集理论与方法

西安交通大学数学研究生教学丛书

张文修 吴伟志
梁吉业 李德玉 编著



科学出版社

西安交通大学数学研究生教学丛书

粗糙集理论与方法

张文修　吴伟志
梁吉业　李德玉　编著

科学出版社
2001

内 容 简 介

本书系统地介绍了粗糙集理论的基本内容与方法，力图概括国内外最新成果. 主要内容有：粗糙集的基本概念，粗糙计算方法，粗糙集的代数性质与粗糙逻辑，粗糙集的各种推广模型，粗糙集与其他处理不确定或不精确问题理论的联系以及不完备信息系统下的粗糙集方法.

本书可作为计算机科学、应用数学、自动控制、信息科学和管理工程等专业的高年级学生及研究生的教材，也可作为研究粗糙集理论与方法的科技人员的参考书.

图书在版编目(CIP)数据

粗糙集理论与方法/张文修等编著. -北京：科学出版社，2001

(西安交通大学数学研究生教学丛书)

ISBN 7-03-008798-4

Ⅰ. 粗… Ⅱ. 张… Ⅲ. 粗糙集 Ⅳ. O144

中国版本图书馆 CIP 数据核字(2000)第 69236 号

科学出版社 出版

北京东黄城根北街16号

邮政编码：100717

新蕾印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2001年7月第 一 版 开本：B5(720×1000)

2001年7月第一次印刷 印张：14 3/4

印数：1-3 000 字数：261 000

定价：22.00 元

(如有印装质量问题，我社负责调换〈新欣〉)

前　言

当今,社会已经进入了网络信息时代,计算机与网络信息技术的飞速发展使得各个领域的数据和信息急剧增加(信息爆炸),并且由于人类的参与使数据与信息系统中的不确定性更加显著(复杂系统).如何从大量的、杂乱无章的、强干扰的数据(海量数据)中挖掘潜在的、有利用价值的信息(有用知识),这给人类的智能信息处理能力提出了前所未有的挑战.由此产生了人工智能研究的一个崭新领域——数据挖掘(DM)和数据库知识发现(KDD).

在DM和KDD诸多方法中,粗糙集理论与方法对于处理复杂系统不失为一种较为有效的方法,因为它与概率方法、模糊集方法和证据理论方法等其他处理不确定性问题理论的最显著的区别是它无需提供问题所需处理的数据集合之外的任何先验信息.当然,由于该理论未能包含处理不精确或不确定原始数据的机制,所以与其他处理不确定性问题的理论有很强的互补性.

粗糙集理论是波兰数学家Z. Pawlak于1982年提出的一种数据分析理论.由于最初关于粗糙集理论的研究主要集中在波兰,因此当时并没有引起国际计算机界和数学界的重视,研究地域仅局限于东欧一些国家.直到1990年前后,由于该理论在数据的决策与分析、模式识别、机器学习与知识发现等方面的成功应用,才逐渐引起了世界各国学者的广泛关注.1991年Z. Pawlak的专著《粗糙集——关于数据推理的理论》(Rough Sets——Theoretical Aspects of Reasoning about Data)的问世,标志着粗糙集理论及其应用的研究进入了活跃时期.1992年在波兰召开了关于粗糙集理论的第一届国际学术会议.1995年ACM Communication将粗糙集列为新浮现的计算机科学的研究课题.目前,粗糙集理论已成为信息科学最为活跃的研究领域之一.同时,该理论还在医学、化学、材料学、地理学、管理科学和金融等其他学科得到了成功的应用.

本书的目的是介绍粗糙集的基本理论与方法以及这一理论的研究发展状况.为了阅读方便,本书对国内外已发表的文章进行了系统化处理,规范了数学概念与符号,在统一的框架下叙述了粗糙集理论的最新研究成果,同时也包含了作者的某些新成果,期望为从事粗糙集理论研究人员和研究生进入这一新领域提供捷径.

鉴于我们从事该领域的研究工作时间较短,加之自身知识的局限性,错误与不妥之处在所难免,热忱欢迎广大同仁批评、指正.

作　者

2000年8月

目　录

第一章　粗糙集理论的基本概念 …… 1
§1.1　知识与知识库 …… 1
§1.2　不精确范畴，近似与粗糙集 …… 3
§1.3　知识约简 …… 12
§1.4　知识的依赖性 …… 16
§1.5　知识表达系统 …… 17
§1.6　决策表 …… 19
§1.7　区分矩阵与区分函数 …… 22
第二章　粗糙集模型的算法 …… 26
§2.1　信息系统和决策表 …… 26
§2.2　简单分类 …… 27
§2.3　支持子集 …… 30
§2.4　决策属性的支持度 …… 32
§2.5　交的计算 …… 33
§2.6　多个条件的支持度 …… 34
§2.7　函数依赖 …… 35
§2.8　恒等依赖 …… 36
§2.9　重要性和核 …… 36
§2.10　属性依赖性 …… 39
§2.11　约简 …… 39
第三章　一般关系下的粗糙集模型 …… 41
§3.1　二元关系与邻域算子 …… 41
§3.2　二元关系与粗糙近似算子 …… 43
§3.3　近似算子的其他定义形式与比较 …… 48
§3.4　近似算子的表示 …… 53
§3.5　程度粗糙集模型 …… 55
第四章　粗糙集代数的公理化方法 …… 57
§4.1　粗糙集理论的构造性方法 …… 57
§4.2　粗糙集理论的公理化方法 …… 59
§4.3　构造性方法与公理化方法的关系 …… 62
§4.4　特殊类型的粗糙集代数 …… 64
第五章　粗糙集系统的代数结构 …… 72

§5.1 粗糙集的 Stone 代数 …… 72
§5.2 粗糙近似空间 …… 76
§5.3 粗糙集和 Nelson 代数 …… 77
§5.4 粗糙概念的代数刻画 …… 85
§5.5 半群中的粗理想 …… 93
第六章 粗糙逻辑与决策 …… 98
§6.1 基于完备信息系统的粗糙逻辑 …… 98
§6.2 决策逻辑与决策 …… 100
§6.3 基于不完备信息系统的模态逻辑 …… 115
第七章 变精度粗糙集模型 …… 123
§7.1 多数包含关系 …… 123
§7.2 变精度粗糙集模型中的近似集 …… 124
§7.3 集合的相对可辨别性 …… 126
§7.4 β 近似的性质 …… 128
§7.5 属性的近似依赖性 …… 129
§7.6 近似约简 …… 130
第八章 概率粗糙集模型 …… 132
§8.1 有限论域上概率测度的基本知识 …… 132
§8.2 信息熵 …… 133
§8.3 概率粗糙集模型 …… 135
§8.4 概率粗糙集模型的其他形式 …… 139
§8.5 Bayes 决策与粗糙近似 …… 142
§8.6 粗糙隶属函数与概念的联合 …… 148
§8.7 知识的不确定性度量 …… 151
§8.8 概率粗糙集模型和确定性粗糙集模型的比较 …… 155
第九章 模糊粗糙集模型 …… 158
§9.1 模糊集的基本概念 …… 158
§9.2 模糊关系 …… 160
§9.3 模糊粗糙集 …… 161
§9.4 基于三角模的模糊粗糙集模型 …… 168
§9.5 基于包含度的粗糙集模型 …… 178
§9.6 修正型模糊粗糙集模型 …… 182
§9.7 粗糙集与模糊集的比较 …… 185
第十章 基于随机集的粗糙集模型 …… 187
§10.1 随机集与容度泛函 …… 187
§10.2 信任函数与似然函数 …… 188
§10.3 基于随机集的粗糙集模型 …… 194
§10.4 近似算子与可能性测度 …… 201

第十一章　不完备信息系统的粗糙集方法 …… 206
§11.1　不完备信息系统 …… 206
§11.2　近似集 …… 207
§11.3　决策表,决策规则和知识约简 …… 208
§11.4　区分函数与约简的计算 …… 211
参考文献 …… 213
记号表 …… 223

第一章　粗糙集理论的基本概念

粗糙集理论是一种新的处理模糊和不确定性知识的数学工具.其主要思想就是在保持分类能力不变的前提下,通过知识约简,导出问题的决策或分类规则.目前,粗糙集理论已被成功地应用于机器学习、决策分析、过程控制、模式识别与数据挖掘等领域.本章介绍标准粗糙集理论(Pawlak 粗糙集模型)的基本概念,作为后面各章节的基础.

§1.1　知识与知识库

设 $U\neq\varnothing$ 是我们感兴趣的对象组成的有限集合,称为论域.任何子集 $X\subseteq U$,称为 U 中的一个概念或范畴.为规范化起见,我们认为空集也是一个概念.U 中的任何概念族称为关于 U 的抽象知识,简称知识.本书主要是对在 U 上能形成划分的那些知识感兴趣.一个划分 $\mathscr{C}$ 定义为:$\mathscr{C}=\{X_1,X_2,\cdots,X_n\}$;$X_i\subseteq U$,$X_i\neq\varnothing$,$X_i\cap X_j=\varnothing$,对于 $i\neq j$,$i,j=1,2,\cdots,n$;$\bigcup\limits_{i=1}^{n}X_i=U$.

U 上的一族划分称为关于 U 的一个知识库(knowledge base).

设 R 是 U 上的一个等价关系,U/R 表示 R 的所有等价类(或者 U 上的分类)构成的集合,$[x]_R$ 表示包含元素 $x\in U$ 的 R 等价类.一个知识库就是一个关系系统 $K=(U,\mathbf{R})$,其中 U 为非空有限集,称为论域,$\mathbf{R}$ 是 U 上的一族等价关系.

若 $\mathbf{P}\subseteq\mathbf{R}$,且 $\mathbf{P}\neq\varnothing$,则 $\bigcap\mathbf{P}$($\mathbf{P}$ 中所有等价关系的交集)也是一个等价关系,称为 $\mathbf{P}$ 上的不可区分(indiscernibility)关系,记为 $\mathrm{ind}(\mathbf{P})$,且有

$$[x]_{\mathrm{ind}(\mathbf{P})}=\bigcap_{R\in\mathbf{P}}[x]_R. \tag{1.1}$$

这样,$U/\mathrm{ind}(\mathbf{P})$(即等价关系 $\mathrm{ind}(\mathbf{P})$ 的所有等价类)表示与等价关系族 $\mathbf{P}$ 相关的知识,称为 K 中关于 U 的 $\mathbf{P}$ 基本知识($\mathbf{P}$ 基本集).为简单起见,我们用 $U/\mathbf{P}$ 代替 $U/\mathrm{ind}(\mathbf{P})$,$\mathrm{ind}(\mathbf{P})$ 的等价类称为知识 $\mathbf{P}$ 的基本概念或基本范畴.特别地,如果 $Q\in\mathbf{R}$,则称 Q 为 K 中关于 U 的 Q 初等知识,Q 的等价类为知识 $\mathbf{R}$ 的 Q 初等概念或 Q 初等范畴.

事实上,$\mathbf{P}$ 基本范畴是拥有知识 $\mathbf{P}$ 的论域的基本特性.换句话说,它们是知识的基本模块.

同样,我们也可定义:当 $K=(U,\mathbf{R})$ 为一个知识库,$\mathrm{ind}(K)$ 定义为 K 中

所有等价关系的族,记作 $\mathrm{ind}(K)=\{\mathrm{ind}(\mathbf{P})\mid\varnothing\neq\mathbf{P}\subseteq\mathbf{R}\}$.

例 1.1 给定一玩具积木的集合 $U=\{x_1,x_2,\cdots,x_8\}$,并假设这些积木有不同的颜色(红、黄、蓝),形状(方、圆、三角),体积(小,大).因此,这些积木都可以用颜色、形状、体积这些知识来描述.例如一块积木可以是红色、小而圆的,或黄色、大而方的等.如果我们根据某一属性描述这些积木的情况,就可以按颜色、形状、体积分类.

按颜色分类:

x_1,x_3,x_7——红;

x_2,x_4——蓝;

x_5,x_6,x_8——黄.

按形状分类:

x_1,x_5——圆;

x_2,x_6——方;

x_3,x_4,x_7,x_8——三角.

按体积分类:

x_2,x_7,x_8——大;

x_1,x_3,x_4,x_5,x_6——小.

换言之,我们定义三个等价关系(即属性):颜色 R_1,形状 R_2 和体积 R_3,通过这些等价关系,可以得到下面三个等价类:

$$U/R_1=\{\{x_1,x_3,x_7\},\{x_2,x_4\},\{x_5,x_6,x_8\}\},$$
$$U/R_2=\{\{x_1,x_5\},\{x_2,x_6\},\{x_3,x_4,x_7,x_8\}\},$$
$$U/R_3=\{\{x_2,x_7,x_8\},\{x_1,x_3,x_4,x_5,x_6\}\}.$$

这些等价类是由知识库 $K=(U,\{R_1,R_2,R_3\})$中的初等概念(初等范畴)构成的.

基本范畴是初等范畴的交集构成的,例如下列集合:

$$\{x_1,x_3,x_7\}\cap\{x_3,x_4,x_7,x_8\}=\{x_3,x_7\},$$
$$\{x_2,x_4\}\cap\{x_2,x_6\}=\{x_2\},$$
$$\{x_5,x_6,x_8\}\cap\{x_3,x_4,x_7,x_8\}=\{x_8\}.$$

它们分别为$\{R_1,R_2\}$的基本范畴,即:红色三角形,蓝色方形,黄色三角形.

下列集合:

$$\{x_1,x_3,x_7\}\cap\{x_3,x_4,x_7,x_8\}\cap\{x_2,x_7,x_8\}=\{x_7\},$$
$$\{x_2,x_4\}\cap\{x_2,x_6\}\cap\{x_2,x_7,x_8\}=\{x_2\},$$
$$\{x_5,x_6,x_8\}\cap\{x_3,x_4,x_7,x_8\}\cap\{x_2,x_7,x_8\}=\{x_8\}.$$

它们分别为$\{R_1,R_2,R_3\}$的基本范畴,即:红色大三角形,蓝色大方形,黄色大

三角形.

下列集合:

$$\{x_1,x_3,x_7\}\cup\{x_2,x_4\}=\{x_1,x_2,x_3,x_4,x_7\},$$
$$\{x_2,x_4\}\cup\{x_5,x_6,x_8\}=\{x_2,x_4,x_5,x_6,x_8\},$$
$$\{x_1,x_3,x_7\}\cup\{x_5,x_6,x_8\}=\{x_1,x_3,x_5,x_6,x_7,x_8\}.$$

它们分别为$\{R_1\}$的范畴,即:红或蓝(非黄),蓝或黄(非红),红或黄(非蓝).

注 有些范畴在这个知识库中是无法得到的,例如集合:

$$\{x_2,x_4\}\cap\{x_1,x_5\}=\varnothing,$$
$$\{x_1,x_3,x_7\}\cap\{x_2,x_6\}=\varnothing$$

为空集,也就是说,在我们的知识库中不存在蓝色圆形和红色方形的范畴,即为空范畴.

下面讨论两个知识库之间的关系.

令 $K=(U,\mathbf{P})$和 $K'=(U,\mathbf{Q})$为两个知识库.若 $\mathrm{ind}(\mathbf{P})=\mathrm{ind}(\mathbf{Q})$,即$U/\mathbf{P}=U/\mathbf{Q}$,则称 K 和K'($\mathbf{P}$ 和 $\mathbf{Q}$)是等价的,记作 $K\simeq K'$($\mathbf{P}\simeq\mathbf{Q}$).因此,当 K 和 K'有同样的基本范畴集时,知识库 K 和K'中的知识都能使我们确切地表达关于论域的完全相同的事实.

这个概念意味着可以用不同的属性集对对象进行描述,以表达关于论域的完全相同的事实.

对于 $K=(U,\mathbf{P})$和 $K'=(U,\mathbf{Q})$两个知识库,当 $\mathrm{ind}(\mathbf{P})\subset\mathrm{ind}(\mathbf{Q})$时,我们称知识 $\mathbf{P}$(知识库 K)比知识 $\mathbf{Q}$(知识库 K')更精细,或者说 $\mathbf{Q}$ 比 $\mathbf{P}$ 更粗糙.当 $\mathbf{P}$ 比 $\mathbf{Q}$ 更精细时,我们也称 $\mathbf{P}$ 为 $\mathbf{Q}$ 的特化,$\mathbf{Q}$ 为 $\mathbf{P}$ 的推广.这意味着,推广是将某些范畴组合在一起,而特化则是将范畴分割成更小的单元.

§1.2 不精确范畴,近似与粗糙集

令 $X\subseteq U$,R 为U 上的一个等价关系.当 X 能表达成某些R 基本范畴的并时,称 X 是R 可定义的;否则称 X 为R 不可定义的.

R 可定义集是论域的子集,它可在知识库 K 中精确地定义,而 R 不可定义集不能在这个知识库中定义.R 可定义集也称作R 精确集,而 R 不可定义集也称为R 非精确集或R 粗糙集(rough set).

当存在等价关系 $R\in\mathrm{ind}(K)$且 X 为R 精确集时,集合 $X\subseteq U$ 称为K 中的精确集;当对于任何 $R\in\mathrm{ind}(K)$,X 都为R 粗糙集,则 X 称为K 中的粗糙集.

对于粗糙集可以近似地定义,我们使用两个精确集,即粗糙集的上近似

(upper approximation)和下近似(lower approximation)来描述.

给定知识库 $K=(U,\mathbf{R})$,对于每个子集 $X\subseteq U$ 和一个等价关系 $R\in \mathrm{ind}(K)$,定义两个子集:

$$\underline{R}X=\bigcup\{Y\in U/R \mid Y\subseteq X\}, \tag{1.2}$$

$$\overline{R}X=\bigcup\{Y\in U/R \mid Y\cap X\neq\varnothing\}. \tag{1.3}$$

分别称它们为 X 的 R 下近似集和 R 上近似集.

下近似、上近似也可用下面的等式表达:

$$\underline{R}X=\{x\in U \mid [x]_R\subseteq X\}, \tag{1.4}$$

$$\overline{R}X=\{x\in U \mid [x]_R\cap X\neq\varnothing\}. \tag{1.5}$$

集合 $\mathrm{bn}_R(X)=\overline{R}X-\underline{R}X$ 称为 X 的 R 边界域;$\mathrm{pos}_R(X)=\underline{R}X$ 称为 X 的 R 正域;$\mathrm{neg}_R(X)=U-\overline{R}X$ 称为 X 的 R 负域.显然:$\overline{R}X=\mathrm{pos}_R(X)\cup\mathrm{bn}_R(X)$.

$\underline{R}X$ 或 $\mathrm{pos}_R(X)$是由那些根据知识 R 判断肯定属于 X 的 U 中元素组成的集合;$\overline{R}X$ 是那些根据知识 R 判断可能属于 X 的 U 中元素组成的集合;$\mathrm{bn}_R(X)$是那些根据知识 R 既不能判断肯定属于 X 又不能判断肯定属于 $\sim X$(即 $U-X$)的 U 中元素组成的集合;$\mathrm{neg}_R(X)$是那些根据知识 R 判断肯定不属于 X 的 U 中元素组成的集合.

下列性质是显而易见的.

定理 1.1

(1) X 为 R 可定义集当且仅当$\underline{R}X=\overline{R}X$.

(2) X 为 R 粗糙集当且仅当$\underline{R}X\neq\overline{R}X$.

我们也可将 $\underline{R}X$ 描述为 X 中的最大可定义集,将 $\overline{R}X$ 描述为含有 X 的最小可定义集.

这样,范畴就是可以用已知知识表达的信息项.换句话说,范畴就是用我们的知识可表达的具有相同性质的对象的子集.一般地说,在一给定的知识库中,并不是所有对象子集都可以构成范畴(即用知识表达的概念).因此,这样的子集可以看作为粗范畴(即不精确或近似范畴),它只能用知识通过两个精确范畴,即上、下近似集,粗略地定义.

从近似的定义,我们可以直接得到 R 下近似集和上近似集的下列性质.

定理 1.2

(1) $\underline{R}X\subseteq X\subseteq\overline{R}X$.

(2) $\underline{R}\varnothing=\overline{R}\varnothing=\varnothing$, $\underline{R}U=\overline{R}U=U$.

(3) $\overline{R}(X\cup Y)=\overline{R}X\cup\overline{R}Y$.

(4) $\underline{R}(X\cap Y)=\underline{R}X\cap\underline{R}Y$.

(5) $X\subseteq Y\Rightarrow\underline{R}X\subseteq\underline{R}Y$.

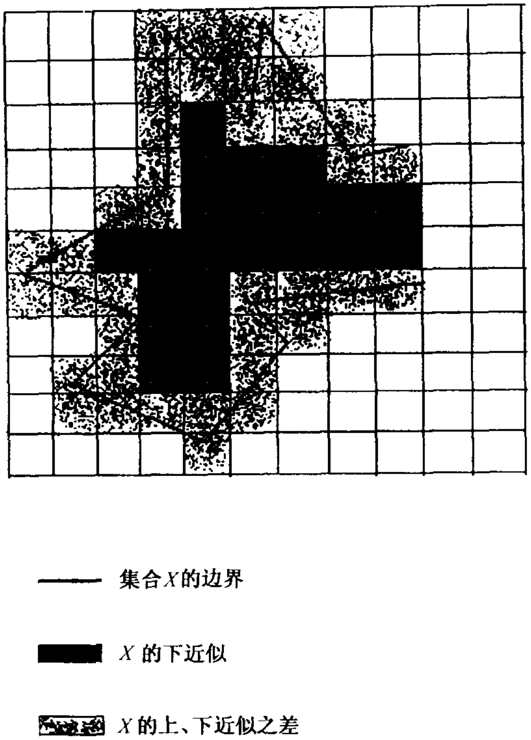


图 1.1 粗糙近似

(6) $X \subseteq Y \Rightarrow \overline{R}X \subseteq \overline{R}Y$.

(7) $\underline{R}(X \cup Y) \supseteq \underline{R}X \cup \underline{R}Y$.

(8) $\overline{R}(X \cap Y) \subseteq \overline{R}X \cap \overline{R}Y$.

(9) $\underline{R}(\sim X) = \sim \overline{R}X$.

(10) $\overline{R}(\sim X) = \sim \underline{R}X$.

(11) $\underline{R}(\underline{R}X) = \overline{R}(\underline{R}X) = \underline{R}X$.

(12) $\overline{R}(\overline{R}X) = \underline{R}(\overline{R}X) = \overline{R}X$.

证

(1a)设 $x \in \underline{R}X$,则有$[x] \subseteq X$;而 $x \in [x]$,所以 $x \in X$.因此,$\underline{R}X \subseteq X$.

(1b)设 $x \in X$,则$[x] \cap X \neq \varnothing$,所以,$x \in \overline{R}X$.因此 $X \subseteq \overline{R}X$.

(2a)由(1)知,$\underline{R}\varnothing \subseteq \varnothing$,而$\varnothing \subseteq \underline{R}\varnothing$,因此 $\underline{R}\varnothing = \varnothing$.

(2b)假设 $\overline{R}\varnothing \neq \varnothing$,则存在 x 使得$x \in \overline{R}\varnothing$,即$[x] \cap \varnothing \neq \varnothing$,而$[x] \cap \varnothing = \varnothing$,与假设矛盾.因此,$\overline{R}\varnothing = \varnothing$.

(2c)由(1)知,$\underline{R}U \subseteq U$.又因为当 $x \in U$,有$[x] \subseteq U$,所以 $x \in \underline{R}U$,即

$U \subseteq \underline{R}U$,因此,$\underline{R}U = U$.

(2d)由(1)知 $\overline{R}U \supseteq U$,但 $\overline{R}U \subseteq U$,因此 $\overline{R}U = U$.

(3) $x \in \overline{R}(X \cup Y) \Leftrightarrow [x] \cap (X \cup Y) \neq \varnothing$

$$\begin{aligned}
&\Leftrightarrow ([x] \cap X) \cup ([x] \cap Y) \neq \varnothing \\
&\Leftrightarrow [x] \cap X \neq \varnothing \vee [x] \cap Y \neq \varnothing \\
&\Leftrightarrow x \in \overline{R}X \vee x \in \overline{R}Y \\
&\Leftrightarrow x \in \overline{R}X \cup \overline{R}Y,
\end{aligned}$$

因此 $\overline{R}(X \cup Y) = \overline{R}X \cup \overline{R}Y$.

(4) $x \in \underline{R}(X \cap Y)$

$$\begin{aligned}
&\Leftrightarrow [x] \subseteq X \cap Y \\
&\Leftrightarrow [x] \subseteq X \wedge [x] \subseteq Y \\
&\Leftrightarrow x \in \underline{R}X \cap \underline{R}Y.
\end{aligned}$$

因此 $\underline{R}(X \cap Y) = \underline{R}X \cap \underline{R}Y$.

(5)设 $X \subseteq Y$,则 $X \cap Y = X$,所以 $\underline{R}(X \cap Y) = \underline{R}X$. 由(4)知,$\underline{R}X \cap \underline{R}Y = \underline{R}X$,因此 $\underline{R}X \subseteq \underline{R}Y$.

(6)设 $X \subseteq Y$,则 $X \cup Y = Y$,所以 $\overline{R}(X \cup Y) = \overline{R}Y$. 由(3)知,$\overline{R}X \cup \overline{R}Y = \overline{R}Y$,因此 $\overline{R}X \subseteq \overline{R}Y$.

(7)因为 $X \subseteq X \cup Y$, $Y \subseteq X \cup Y$,所以 $\underline{R}X \subseteq \underline{R}(X \cup Y)$, $\underline{R}Y \subseteq \underline{R}(X \cup Y)$,故 $\underline{R}X \cup \underline{R}Y \subseteq \underline{R}(X \cup Y)$.

(8)因为 $X \cap Y \subseteq X$, $X \cap Y \subseteq Y$,所以 $\overline{R}(X \cap Y) \subseteq \overline{R}X$, $\overline{R}(X \cap Y) \subseteq \overline{R}Y$,故 $\overline{R}(X \cap Y) \subseteq \overline{R}X \cap \overline{R}Y$.

(9)因为 $x \in \underline{R}X \Leftrightarrow [x] \subseteq X \Leftrightarrow [x] \cap \sim X = \varnothing$

$$\Leftrightarrow x \notin \overline{R}(\sim X) \Leftrightarrow x \in \sim \overline{R}(\sim X),$$

所以 $\underline{R}X = \sim \overline{R}(\sim X)$.

(10)在(9)中用 $\sim X$ 代替 X,即得 $\overline{R}X = \sim \underline{R}(\sim X)$.

(11a)由(1)知,$\underline{R}(\underline{R}X) \subseteq \underline{R}X$. 又当 $x \in \underline{R}X$ 时,有 $[x] \subseteq X$,因此,$\underline{R}[x] \subseteq \underline{R}X$. 但 $\underline{R}[x] = [x]$,于是 $[x] \subseteq \underline{R}X$,所以 $x \in \underline{R}(\underline{R}X)$,即 $\underline{R}X \subseteq \underline{R}(\underline{R}(X)$. 故 $\underline{R}(\underline{R}X) = \underline{R}X$.

(11b)由(1)知,$\underline{R}X \subseteq \overline{R}(\underline{R}X)$. 又当 $x \in \overline{R}(\underline{R}X)$,有 $[x] \cap \underline{R}X \neq \varnothing$,即存在 $y \in [x]$ 使得 $y \in \underline{R}X$,所以 $[y] \subseteq X$,但 $[x] = [y]$,这样就有 $[x] \subseteq X$,即 $x \in \underline{R}X$,因此,$\underline{R}X \supseteq \overline{R}(\underline{R}X)$. 故 $\overline{R}(\underline{R}X) = \underline{R}X$.

(12a)由(1)知,$\overline{R}X \subseteq \overline{R}(\overline{R}X)$,又当 $x \in \overline{R}(\overline{R}X)$ 时,有 $[x] \cap \overline{R}X \neq \varnothing$. 因此存在 $y \in [x]$ 且 $y \in \overline{R}X$,所以 $[y] \cap X \neq \varnothing$. 但 $[x] = [y]$,所以 $[x] \cap X \neq \varnothing$,即 $x \in \overline{R}X$,这样就有 $\overline{R}X \supseteq \overline{R}(\overline{R}X)$. 故 $\overline{R}(\overline{R}X) = \overline{R}X$.

(12b)由(1)知，$\underline{R}(\overline{R}X)\subseteq\overline{R}X$. 又当 $x\in\overline{R}X$ 时有$[x]\cap X\neq\varnothing$，所以$[x]\subseteq\overline{R}X$ （因为任取 $y\in[x]$时，则有$[y]\cap X=[x]\cap X\neq\varnothing$，即 $y\in\overline{R}X$），即 $x\in\underline{R}(\overline{R}X)$，这样就有 $\underline{R}(\overline{R}X)\supseteq\overline{R}X$. 故 $\underline{R}(\overline{R}X)=\overline{R}X$. □

集合近似的概念导致了一个新的概念——成员关系. 因为在我们的方法中，对于一个集合的定义是与集合的知识相联系的，所以成员关系一定也和知识有关，可以形式地定义为

$$x\ \underline{\in}_R X \quad 当且仅当 \quad x\in\underline{R}X;$$

$$x\ \overline{\in}_R Y \quad 当且仅当 \quad x\in\overline{R}X;$$

这里$\underline{\in}_R$ 表示根据R，x 肯定地属于X；$\overline{\in}_R$ 表示根据R，x 可能属于X. 分别称$\underline{\in}_R$ 和$\overline{\in}_R$ 为下和上成员关系.

成员关系依赖于我们的知识，即一个对象是否属于一个集合依赖于我们的知识，并且这不是绝对特性.

由定理 1.2 我们可以得到成员关系的性质：

定理 1.3

(1) $x\underline{\in}X$ 蕴涵$x\in X$ 蕴涵$x\overline{\in}X$；

(2) $X\subseteq Y$ 蕴涵($x\underline{\in}X$ 蕴涵$x\underline{\in}Y$ 且$x\overline{\in}X$ 蕴涵$x\overline{\in}Y$)；

(3) $x\overline{\in}(X\cup Y)$当且仅当 $x\overline{\in}X$ 或$x\overline{\in}Y$；

(4) $x\underline{\in}(X\cap Y)$当且仅当 $x\underline{\in}X$ 且$x\underline{\in}Y$；

(5) $x\underline{\in}X$ 或$x\underline{\in}Y$ 蕴涵$x\underline{\in}(X\cup Y)$；

(6) $x\overline{\in}(X\cap Y)$蕴涵 $x\overline{\in}X$ 且$x\overline{\in}Y$；

(7) $x\underline{\in}(\sim X)$当且仅当 $x\overline{\in}X$ 不成立；

(8) $x\overline{\in}(\sim X)$当且仅当 $x\underline{\in}X$ 不成立.

为简单起见，我们省略了以上式子中的下标 R.

集合(范畴)的不精确性是由于边界域的存在而引起的. 集合的边界域越大，其精确性则越低. 为了更准确地表达这一点，我们引入精度的概念. 由等价关系 R 定义的集合X 的近似精度为

$$\alpha_R(X)=\frac{|\underline{R}X|}{|\overline{R}X|}, \tag{1.6}$$

其中 $X\neq\varnothing$，$|X|$表示集合 X 的基数.

精度 $\alpha_R(X)$用来反映我们对于了解集合 X 的知识的完全程度. 显然，对每一个 R 和$X\subseteq U$ 有 $0\leqslant\alpha_R(X)\leqslant1$. 当 $\alpha_R(X)=1$ 时，X 的R 边界域为空集，集合 X 为R 可定义的；当 $\alpha_R(X)<1$ 时，集合 X 有非空R 边界域，集合 X

为 R 不可定义的.

当然,其他一些量度同样可用来定义集合 X 的不精确程度.

例如,可用 $\alpha_R(X)$的一种变形,即 X 的R 粗糙度$\rho_R(X)$来定义:

$$\rho_R(X) = 1 - \alpha_R(X).$$

X 的R 粗糙度与精度恰恰相反,它表示的是集合 X 的知识的不完全程度.

我们可以看到,与概率论和模糊集合论不同,不精确性的数值不是事先假定的,而是通过表达知识不精确性的概念近似计算得到的,这样不精确性的数值表示的是有限知识(对象分类能力)的结果,这里我们不需要用一个机构来指定精确的数值去表达不精确的知识,而是采用量化概念(分类)来处理.不精确的数值特征用来表示概念的精确度.

除了用数值(近似程度的精度)来表示粗糙集的特征外,也可根据上、下近似的定义来表达粗糙集的另一个有用的特征,即拓扑特征.

下面定义四种不同的重要粗糙集:

(1)如果 $\underline{R}X\neq\varnothing$且 $\overline{R}X\neq U$,则称 X 为R 粗糙可定义.

(2)如果 $\underline{R}X=\varnothing$且 $\overline{R}X\neq U$,则称 X 为R 内不可定义.

(3)如果 $\underline{R}X\neq\varnothing$且 $\overline{R}X=U$,则称 X 为R 外不可定义.

(4)如果 $\underline{R}X=\varnothing$且 $\overline{R}X=U$,则称 X 为R 全不可定义.

这个划分的直观意义是这样的:

如果集合 X 为R 粗糙可定义,则意味着我们可以确定 U 中某些元素属于X 或$\sim X$.

如果 X 为R 内不可定义,则意味着我们可以确定 U 中某些元素是否属于$\sim X$,但不能确定 U 中的任一元素是否属于X.

如果 X 为R 外不可定义的,则意味着我们可以确定 U 中某些元素是否属于X,但不能确定 U 中的任一元素是否属于$\sim X$.

如果 X 为R 全不可定义,则意味着我们不能确定 U 中任一元素是否属于X 或$\sim X$.

下面我们给出集合拓扑划分的一个有用性质.

定理 1.4

(1)集合 X 为R 粗糙可定义(或 R 全不可定义)当且仅当$\sim X$ 为R 粗糙可定义(或 R 全不可定义);

(2)集合 X 为R 外(内)不可定义当且仅当$\sim X$ 为R 内(外)不可定义.

证　(1)设 X 为R 粗糙可定义,则 $\underline{R}X\neq\varnothing,\overline{R}X\neq U$.

$\underline{R}X\neq\varnothing\Leftrightarrow$存在 $x\in X$,使得$[x]_R\subseteq X$

$\Leftrightarrow[x]_R\cap\sim X=\varnothing\Leftrightarrow\overline{R}(\sim X)\neq U$;

类似地,

$\overline{R}X \neq U \Leftrightarrow$ 存在 $y \in U$,使得$[y]_R \cap \overline{R}X = \varnothing$

$\Leftrightarrow [y]_R \subseteq \sim \overline{R}X \Leftrightarrow [y]_R \subseteq \underline{R}(\sim X) \Leftrightarrow \underline{R}(\sim X) \neq \varnothing$.

同理可证其余情况. □

至此,我们已经介绍了两种刻画粗糙集的方法.其一为用近似程度的精度来表示粗糙集的数字特征;其二为用粗糙集的分类表示粗糙集的拓扑特征.粗糙集的数字特征表示了集合边界域的大小,但没有说明边界域的结构;而粗糙集的拓扑特征没有给出边界域大小的信息,它提供的是边界域的结构.

此外,粗糙集的数字特征和粗糙集的拓扑特征之间存在一种关系.首先,如果集合为内不可定义或全不可定义,则其精度为0;其次,当集合为外不可定义或全不可定义,则它的补集的精度为0.这样即使知道了集合的精度,我们也不能确定它的拓扑结构;反过来,集合的拓扑结构也不具备精度的信息.

因此在粗糙集的实际应用中,我们需要将边界域的两种信息结合起来,既要考虑精度因素,又要考虑到集合的拓扑结构.

例 1.2 给定一知识库 $K=(U,\mathbf{R})$和一个等价关系 $R \in \mathrm{ind}(K)$,其中 $U=\{x_0,x_1,\cdots,x_{10}\}$,且有 R 的下列等价类:

$$E_1 = \{x_0,x_1\},$$

$$E_2 = \{x_2,x_6,x_9\},$$

$$E_3 = \{x_3,x_5\},$$

$$E_4 = \{x_4,x_8\},$$

$$E_5 = \{x_7,x_{10}\}.$$

集合 $X_1=\{x_0,x_1,x_4,x_8\}$为 R 可定义集,因为

$$\underline{R}X_1 = \overline{R}X_1 = E_1 \cup E_4.$$

集合 $X_2=\{x_0,x_3,x_4,x_5,x_8,x_{10}\}$为 R 粗糙可定义集.

X_2 的近似集,边界和精度为

$\underline{R}X_2 = E_3 \cup E_4 = \{x_3,x_4,x_5,x_8\}$,

$\overline{R}X_2 = E_1 \cup E_3 \cup E_4 \cup E_5 = \{x_0,x_1,x_3,x_4,x_5,x_7,x_8,x_{10}\}$,

$\mathrm{bn}_R(X_2) = E_1 \cup E_5 = \{x_0,x_1,x_7,x_{10}\}$,

$\alpha_R(X_2) = 4/8 = 1/2$.

集合 $X_3=\{x_0,x_2,x_3\}$为 R 内不可定义.因为

$$\underline{R}X_3 = \varnothing,$$

$$\overline{R}X_3 = E_1 \cup E_2 \cup E_3 = \{x_0,x_1,x_2,x_3,x_5,x_6,x_9\} \neq U.$$

集合 $X_4=\{x_0,x_1,x_2,x_3,x_4,x_7\}$为 R 外不可定义.

X_4 的近似集,边界和精度为

$$\underline{R}X_4 = E_1 = \{x_0, x_1\},$$
$$\overline{R}X_4 = U,$$
$$\mathrm{bn}_R(X_4) = E_2 \cup E_3 \cup E_4 \cup E_5 = \{x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}\},$$
$$\alpha_R(X_3) = 2/11.$$

集合 $X_5 = \{x_0, x_2, x_3, x_4, x_7\}$ 为 R 全不可定义,因为

$$\underline{R}X_1 = \varnothing, \qquad \overline{R}X_5 = U.$$

下面讨论近似分类问题.

令 $\mathscr{F} = \{X_1, X_2, \cdots, X_n\}$ 是 U 的一个分类或划分,这个分类独立于知识 R. 例如,它可能由一个专家为解决一个分类问题所给出. 子集 $X_i (i = 1, 2, \cdots, n)$ 是划分 $\mathscr{F}$ 的类. $\mathscr{F}$ 的 R 下近似和上近似分别定义为: $\underline{R}\mathscr{F} = \{\underline{R}X_1, \underline{R}X_2, \cdots, \underline{R}X_n\}$ 和 $\overline{R}\mathscr{F} = \{\overline{R}X_1, \overline{R}X_2, \cdots, \overline{R}X_n\}$.

我们定义两个量度来描述近似分类的不精确性.

第一个量度为根据 R, $\mathscr{F}$ 的近似分类精度:

$$\alpha_R(\mathscr{F}) = \frac{\sum_{i=1}^{n} |\underline{R}X_i|}{\sum_{i=1}^{n} |\overline{R}X_i|}. \tag{1.7}$$

第二个量度为根据 R, $\mathscr{F}$ 的近似分类质量:

$$\gamma_R(\mathscr{F}) = \frac{\sum_{i=1}^{n} |\underline{R}X_i|}{|U|}. \tag{1.8}$$

近似分类的精度描述的是当使用知识 R 对对象分类时,可能的决策中正确决策的百分比;分类的质量表示的是应用知识 R 能确切地划入 $\mathscr{F}$ 类的对象的百分比.

粗糙集概念和通常的集合有着本质的区别. 在集合的相等上也有一个重要的区别. 在普通集合里,如果两个集合有完全相同的元素,则这两个集合相等. 在粗糙集理论中,我们需要另一个集合相等的概念,即近似(粗糙)相等. 两个集合在普通集合里不是相等的,但在粗糙集里可能是近似相等的. 因为两个集合是否近似相等是根据我们得到的知识判断的.

现在我们介绍集合的三种近似相等的定义.

令 $K = (U, \mathbf{R})$ 是一个知识库, $X, Y \subseteq U$ 且 $R \in \mathrm{ind}(K)$.

(1)若 $\underline{R}X = \underline{R}Y$,则称集合 X 和 Y 为 R 下粗相等,记作 $X \eqsim_R Y$.

(2)若 $\overline{R}X = \overline{R}Y$,则称集合 X 和 Y 为 R 上粗相等,记作 $X \simeq_R Y$.