基于粗糙集与概念格的知识系统模型

李进金　李克典　吴端恭　编著

科学出版社

基于粗糙集与概念格的知识系统模型

李进金　李克典　吴瑞恭　编著

科学出版社
北京

内 容 简 介

本书依据粗糙集与概念格理论的基本内容和方法，系统地介绍了各种知识系统模型. 书中融入国内外学者研究的许多新成果，主要内容有：粗糙集与知识系统的基本概念，覆盖粗糙集模型，不完备信息系统，变精度粗糙集模型，粗糙集理论中的优化问题与属性重要性的刻画，格值信息系统，形式背景与概念格，决策形式背景.

本书可作为计算机科学、管理科学、系统科学、信息科学和数学与应用数学等专业高年级本科生、研究生的教材及教师的科研和教学的参考书，也可作为决策管理人员及研究工作者的参考书.

图书在版编目(CIP)数据

基于粗糙集与概念格的知识系统模型/李进金，李克典，吴端恭编著. —北京：科学出版社，2013. 3

ISBN 978-7-03-036538-5

Ⅰ. ①基… Ⅱ. ①李… ②李… ③吴… Ⅲ. ①知识系统－系统模型－研究 Ⅳ. ①G302

中国版本图书馆 CIP 数据核字(2013)第 016874 号

责任编辑：王丽平 唐保军／责任校对：宋玲玲
责任印制：钱玉芬／封面设计：王 浩

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

骏杰印刷厂 印刷
科学出版社发行 各地新华书店经销
*
2013 年 3 月第 一 版 开本：B5(720 × 1000)
2013 年 3 月第一次印刷 印张：11 1/4
字数：230 000

定价：48.00 元

(如有印装质量问题，我社负责调换)

序

知识是对某些客观对象的认识, 并通过属性来刻画这种认识, 人类认识的过程即是知识发现的过程. 随着科学技术的发展和人类社会的进步, 人们对自然界和客观事物演化规律的认识逐步深化. 新知识来源于新信息、新分析、新归纳或新关系的觉识及关系应用途径的新开发. 知识系统模型包括: 知识点 (概念) 模型的形成、进化, 知识集的形成, 知识集内结构的引进. 知识有内容与形式两个方面, 概念有内涵与外延两个方面的描述、确定, 关系也有内涵与外在形式的描述 (解释) 与表达. 知识形成过程的模型也是知识系统模型的重要组成, 是元知识系统的模型问题, 是一个重要级别 (层次) 的知识系统的模型, 知识有分层次, 有分段, 有时空特性. 概念在粗糙集理论中被简单地看成论域 U 上的子集, 进而与属性集 A 发生联系, 引入了概念 (类) 的描述, 内涵描述与外延描述的匹配在概念格的形成概念中达到一种完备性, 而决策规则形成的多样化, 规则内涵的泛化, 决策规则集的内在关系等的深入研究也丰富了知识系统的模型. 知识系统的模型可以有许多层次, 有知识的层次 (对论域 U 的深入、获取知识的不同层次), 有认识水平的层次 (广泛性、针对性、精确性、有效性等). 可以有等价不等效, 可以有偏序关系等. 对知识系统的比较、评价也是知识系统模型关注的重要内容. 知识系统的结构与功能, 知识系统的开发应用, 是知识系统研究的根本目标所在.

20 世纪 80 年代波兰数学家 Z.Pawlak 教授创立的粗糙集理论是一种不确定性系统研究的有效方法, 有着广泛的应用前景和发展方向. 李进金教授等撰写的《基于粗糙集与概念格的知识系统模型》, 在粗糙集理论与概念格的研究及其应用方面做了大量的工作, 建立的知识系统模型具有强大的数据挖掘、知识发现和规则提取等功能. 该书不仅把读者带到不确定性理论研究的前沿, 而且还拓宽了粗糙集理论的应用范围. 相信该书的出版不论对于从事粗糙集理论、概念格等不确定性理论及其应用研究的专家, 还是在读研究生都是有趣和有用的, 对于促进粗糙集与概念格的理论研究和应用具有积极的意义.

王国俊

2012 年 4 月于陕西师范大学

前　言

粗糙集理论是由波兰数学家 Z. Pawlak 于 1982 年提出的一种处理不精确、不确定和不完全数据的有效方法, 它无需提供所处理的数据集合之外的任何先验信息便可以分析不精确、不一致和不完全等各种不完备的信息. 粗糙集理论作为数据处理的强有力工具已被成功地广泛应用于医学、化学、材料科学、地理学、管理决策科学和金融经济等学科的知识获取、决策分析、预测、专家系统和数据库中的知识发现. 因此, 粗糙集理论已成为信息科学最为活跃的研究领域之一.

本书依据粗糙集与概念格理论的基本内容和方法, 系统地介绍了各种知识系统模型，融入了国内外学者研究的许多最新成果. 全书共八章, 第 1 章介绍粗糙集与知识系统的基本概念, 第 2 章介绍覆盖粗糙集模型, 第 3 章介绍不完备信息系统, 第 4 章介绍变精度粗糙集模型, 第 5 章介绍粗糙集理论中的优化问题与属性重要性的刻画, 第 6 章介绍格值信息系统, 第 7 章介绍形式背景与概念格, 第 8 章介绍决策形式背景. 本书可作为计算机科学、管理科学、系统科学、信息科学和数学与应用数学等专业高年级本科生、研究生的教材及教师的科研和教学的参考书, 也可作为决策管理人员及研究工作者的参考书.

本书的第 1 与第 2 章及第 8 章由李进金编写, 第 3 与第 4 章由李克典编写, 第 5~7 章由吴端恭编写, 全书由李克典统一整理. 本书是作者在 2006 年开始招收粗糙集理论及其应用方向的硕士研究生以来所使用教材的基础上, 结合多年对粗糙集理论及其应用的部分研究成果编著而成. 书稿中的大部分内容曾在漳州师范学院数学与信息科学系研究生中讲授过, 并在粗糙集理论及其应用讨论班使用. 书稿的写作与出版始终得到漳州师范学院研究生处、科研处的支持, 书稿的编辑和排版得到南京大学师维学教授、漳州师范学院特聘教授林寿教授、漳州师范学院粗糙集理论及其应用专业研究生和讨论班的老师们的帮助, 感谢陕西师范大学王国俊教授和漳州师范学院“闽江学者”特聘教授祝峰教授在百忙之中认真审阅了全部书稿, 感谢王国俊教授为本书作序. 本书出版工作得到国家自然科学基金项目“覆盖的约简理论及其在中医方剂配伍规律研究中的应用”(项目编号：10971186)、“基本拓扑方法及其在广义度量空间和覆盖粗糙集理论中的应用”(项目编号：11061004)、“覆盖决策信息系统理论及其在复杂系统决策中的应用”(项目编号：71140004) 和漳州师范学院学术出版基金的资助. 对所有关心本书编写和出版的同行们, 及漳州师范学院的研究生们、读书班的老师们给予的支持与帮助, 在此一并表示衷心的感谢.

目　　录

第 1 章 粗糙集与知识系统的基本概念

1.1 知识系统的一个简单模型

1.1.1 一个知识系统

设 $U \neq \varnothing$ 是我们感兴趣的对象组成的有限集合, 称为论域. 论域 U 的任何子集 (包括空集) X 称为 U 中一个概念或范畴. U 中的任何一个概念族 (即 U 的幂集 $\mathcal{P}(U)$ 的任何一个子族) 称为关于 U 的一个知识系统, 简称知识.

定义 1.1.1 设 U 是论域, 如果 $X_i \in \mathcal{P}(U)$ $(i \leqslant n)$, 且 $X_i \neq \varnothing$ $(i \leqslant n)$, $X_i \cap X_j = \varnothing$ $(i \neq j)$, $\bigcup\limits_{i=1}^{n} X_i = U$, 则称 $\{X_1, X_2, \cdots, X_n\}$ 为论域 U 的一个划分(或剖分), 记为 $\pi = \{X_1, X_2, \cdots, X_n\}$.

论域 U 的一个划分即是 U 的一个分类, 把研究的对象分成不同类, 每一个对象都在某一类中, 不同的类之间互不相交. 这是知识系统内部结构的一种特殊情况, 即一个概念族构成论域 U 的一个划分. 这样的知识系统可以表示为 $K = (U, \pi)$, 其中 π 是论域 U 的一个划分. 理解这个知识系统 (作为一个概念族), 不止含有 X_i $(1 \leqslant i \leqslant n)$, 而且含有若干个 X_i $(1 \leqslant i \leqslant n)$ 的并这类概念. 划分 π 中各成员称为该知识系统的基本范畴, 知识系统中其他概念称为初等范畴, 约定 $\varnothing \in K$, 并且显然 $U \in K$.

定义 1.1.2 设 U 是论域, 如果 $R \subseteq U \times U$, 则称 R 为 U 上的一个关系. 对于 $x, y \in U$, 当 $(x, y) \in R$ 时, 则称 x 与 y 有关系 R, 记为 xRy; 当 $(x, y) \notin R$ 时, 则称 x 与 y 没有关系 R.

定义 1.1.3 设 R 是论域 U 上的一个关系. 如果 R 满足以下性质 (E1) ~ (E3), 则称 R 是 U 上的等价关系:

(E1) 自反性: 任意 $x \in U$, 有 $(x, x) \in R$;

(E2) 对称性: 若 $(x, y) \in R$, 则 $(y, x) \in R$ $(x, y \in U)$;

(E3) 传递性: 若 $(x, y) \in R$ 且 $(y, z) \in R$, 则 $(x, z) \in R$ $(x, y, z \in U)$.

设 R 是 U 上的等价关系, 对于 $x \in U$, $[x]_R = \{y \in U | (x, y) \in R\}$ 表示 x 的 R 等价类; U/R 表示 R 的所有等价类构成的集合, 即 $U/R = \{[x]_R | x \in U\}$.

如果 R 仅满足性质 (E1) 与 (E2), 则称 R 是 U 上的相似关系.

定理 1.1.1 论域 U 上的等价关系必然产生 U 的一个划分; 论域 U 的一个

划分又对应于 U 上的一个等价关系导出的等价类集合.

若划分 π 对应的是 U 上的等价关系 R 的等价类全体, 知识系统也可表示为 $K=(U,R)$.

把知识系统 $K=(U,R)$ 表示法推广, $K=(U,\mathbf{R})$ 也表示一个知识系统, 其中 $\mathbf{R}$ 是 U 上一个等价关系的有限族. 知识系统 $K=(U,\mathbf{R})$ 对应 U 上的分划 π, 也是 U 上的一个等价关系 P 所对应的等价类的全体, 而等价关系 P 是等价关系族 $\mathbf{R}$ 的交 (即 $U\times U$ 上的一族子集的交). 容易验证, U 上两个等价关系之交仍是 U 上的等价关系, 从而可知 P 也是等价关系.

定义 1.1.4 设 $\mathbf{R}$ 是 U 上一个等价关系有限族, 如果 $\mathbf{P}\subset\mathbf{R}$ 且 $\mathbf{P}\neq\varnothing$, U 上的关系 $\cap\{R|R\in\mathbf{P}\}$, 则称为 $\mathbf{P}$ 的不可分辨关系, 记为 $\mathrm{ind}(\mathbf{P})$.

容易验证, $\mathbf{P}$ 的不可分辨关系 $\mathrm{ind}(\mathbf{P})$ 是 U 上的一个等价关系, 且 $[x_i]_{\mathbf{P}}=\bigcap\limits_{R\in\mathbf{P}}[x_i]_R$.

设 $K=(U,\mathbf{R})$ 是知识系统, $U/\mathrm{ind}(\mathbf{R})$ 表示与等价关系族 $\mathbf{R}$ 相关知识, 称为 K 中关于 U 的 $\mathbf{R}$ 基本知识或基本范畴, 其中基本知识或基本范畴为 $[x_i]_{\mathbf{R}}=\bigcap\limits_{R\in\mathbf{R}}[x_i]_R$. 为了方便起见, 可用 $U/\mathbf{R}$ 代替 $U/\mathrm{ind}(\mathbf{R})$.

值得注意的是, U 上两个等价关系的并不一定是 U 上的等价关系.

定理 1.1.2 如果 R_1 与 R_2 都是 U 上的等价关系,

$$U/R_1=\{X_1,X_2,\cdots,X_{n_1}\},\quad U/R_2=\{Y_1,Y_2,\cdots,Y_{n_2}\},$$

则 $R_1\cup R_2$ 是 U 上的等价关系 $\Leftrightarrow$ 当 $X_i\cap Y_j\neq\varnothing$ 时, 有 $X_i\supseteq Y_j$, 或 $X_i\subseteq Y_j$.

一个数据库, 可以看成现在所讨论的这种类型的知识系统. 数据库的行 (即记录) 看成论域的对象, 全体记录 (数据库的行的全体) 构成有限论域 U; 而数据库的列 (项) 看成为属性, 所有列构成属性集 A. 可以利用知识系统给出数据库的抽象描述.

定义 1.1.5 称 (U,A,F,V) 一个为知识系统, 其中 U 是对象集, 即

$$U=\{x_1,x_2,\cdots,x_n\},$$

U 中的每一个元素 x_j $(1\leqslant j\leqslant n)$ 称为一个对象. A 是属性集, 即

$$A=\{a_1,a_2,\cdots,a_m\},$$

A 中的每一个元素 a_i $(1\leqslant i\leqslant m)$ 称为一个属性. F 为 U 与 A 之间的关系集, 即

$$F=\{f_{a_1},f_{a_2},\cdots,f_{a_m}\},$$

每一个属性 $a_i \in A$, f_{a_i} 是从 U 到 V_{a_i} 的一个映射 (它是 U 到 V_{a_i} 的一个关系), 即

$$f_{a_i} : U \to V_{a_i},$$

V_{a_i} 是属性 a_i $(1 \leqslant i \leqslant m)$ 的值域, $V = \{(v_{a_1}, v_{a_2}, \cdots, v_{a_m}) | v_{a_i} \in V_{a_i}, 1 \leqslant i \leqslant m\}$.

对于每一个属性 $a_i \in A$ 对应 U 上一个等价关系 $R_{a_i}(1 \leqslant i \leqslant m)$, 即

$$R_{a_i} = \{(x, y) | x, y \in U, f_{a_i}(x) = f_{a_i}(y)\}.$$

令 $R = \bigcap\limits_{1 \leqslant i \leqslant m} R_{a_i}$, 则一个数据库被看成一个知识系统 $K = (U, R)$, 这一知识系统也被表示为 $K = (U, A, F, V)$.

1.1.2 两个知识系统的关系

论域 U 上的知识系统多种多样, 自然引起人们对不同知识系统的比较研究, 我们讨论两个知识系统的关系.

设 $K = (U, A, F, V)$ 为知识系统, 记

$$R_a = \{(x, y) | x, y \in U, f_a(y) = f_a(x)\}, \quad \mathbf{R} = \{R_a | \forall a \in A\},$$

这里 $\mathbf{R}$ 对应 U 上一个属性集 A. 由定义 1.1.4, $\mathrm{ind}(\mathbf{R})$ 表示知识系统 $K = (U, A, F, V)$ 中对象之间的一个不可分辨关系, 也是一个等价关系. 对于每一个 $x \in U$, $\mathrm{ind}(\mathbf{R})$ 的等价类 $[x]_{\mathrm{ind}(\mathbf{R})}$ 是依属性集 A 与 x 无法分辨的对象的全体. 即

$$[x]_{\mathrm{ind}(\mathbf{R})} = \{y \in U | f_a(y) = f_a(x), \forall a \in A\}.$$

此时, $\mathrm{ind}(\mathbf{R})$ 也可记为 $\mathrm{ind}(A)$.

定义 1.1.6 设 $K = (U, \mathbf{P})$ 和 $K' = (U, \mathbf{Q})$ 为两个知识系统.

(i) 如果 $\mathrm{ind}(\mathbf{P}) = \mathrm{ind}(\mathbf{Q})$, 即 $U/\mathbf{P} = U/\mathbf{Q}$, 则称知识系统 K 和知识系统 K' (或 $\mathbf{P}$ 和 $\mathbf{Q}$) 是等价的, 记作 $K \simeq K'$ (或 $\mathbf{P} \simeq \mathbf{Q}$).

(ii) 如果 $\mathrm{ind}(\mathbf{P}) \subset \mathrm{ind}(\mathbf{Q})$ (这里的包含 “$\subset$” 是等价关系的包含关系, 即 $\mathrm{ind}(\mathbf{P})$ 与 $\mathrm{ind}(\mathbf{Q})$ 作为 $U \times U$ 的子集之间所存在的包含关系), 则称知识系统 K (知识 $\mathbf{P}$) 比知识系统 K' (知识 $\mathbf{Q}$) 更精细, 或者说知识系统 K' (知识 $\mathbf{Q}$) 比知识系统 K (知识 $\mathbf{P}$) 更粗糙.

(iii) 如果 $\mathrm{ind}(\mathbf{P})$ 与 $\mathrm{ind}(\mathbf{Q})$ 既不相等也不相互包含, 则称两个知识系统 K 和 K' 不可比较.

由定义 1.1.6 可知, 当知识系统 K 和知识系统 K' 等价时, K 和 K' 有同样的基本范畴. 关于论域 U 的概念, K 和 K' 的表达能力是等同的. 对于这一情况, 引申出的问题是: 功能相同的前提下, 结构上尽可能优化. 对论域 U 所对应的不同属

性集 A 与 B, 可能知识系统 $K=(U,A,F,V)$ 与知识系统 $K'=(U,B,F',V')$ 等价, 在等价的知识系统中, 寻求结构表达上尽可能简单. 例如, 属性集尽可能简单.

当知识系统 K 比知识系统 K' 更精细, 即知识 $\mathbf{P}$ 比知识 $\mathbf{Q}$ 更精细时, 知识 $\mathbf{P}$ 为知识 $\mathbf{Q}$ 的特化, 或者知识 $\mathbf{Q}$ 为知识 $\mathbf{P}$ 的推广. 这里推广意味着 $\mathrm{ind}(\mathbf{Q})$ 的基本范畴可以表示为 $\mathrm{ind}(\mathbf{P})$ 的某些基本范畴的合并, 而知识 $\mathbf{P}$ 为知识 $\mathbf{Q}$ 的特化则意味着 $\mathrm{ind}(\mathbf{P})$ 的基本范畴细分 $\mathrm{ind}(\mathbf{Q})$ 的基本范畴. 至于细化好还是粗化好, 这要根据需要来确定. 实际工作中, 可以由粗求细, 也可以由细求粗.

当两个知识系统 K 和 K' 不可比较时, 有些可以由 $\mathrm{ind}(\mathbf{P})$ 的基本范畴精确表示的概念不一定能用 $\mathrm{ind}(\mathbf{Q})$ 的基本范畴精确表示; 反过来, 有些可以由 $\mathrm{ind}(\mathbf{Q})$ 的基本范畴精确表示的概念也不一定能用 $\mathrm{ind}(\mathbf{P})$ 的基本范畴精确表示. 用已有的知识背景, 来考虑新的概念的表示问题, 这是常见的. 如果新概念属于知识系统 K', 已有知识系统为 K, 而 K 与 K' 不可比较, 寻求知识系统 K' 中的概念如何用 K 来表示, 对于不能精确表示的概念, 如何近似表示的问题, 是粗集理论着重讨论的问题.

1.1.3　论域 U 上知识系统全体的结构

定义 1.1.7　设 R 是论域 U 上的一个关系. 如果 R 满足以下性质 (P1) ~ (P3), 则称 R 是 U 上的偏序关系:

(P1) 自反性: 任意 $x\in U$, 有 $(x,x)\in R$;

(P2) 反对称性: 若 $(x,y)\in R$, 且 $(y,x)\in R$, 则 $x=y$ $(x,y\in U)$;

(P3) 传递性: 若 $(x,y)\in R$ 且 $(y,z)\in R$, 则 $(x,z)\in R$ $(x,y,z\in U)$.

如果 R 仅具有性质 (P1) 与 (P3), 则称 R 是 U 上的拟序关系.

设 U 是论域, 记

$$\mathcal{K}(U)=\{\mathcal{A}\mid \mathcal{A}\text{ 是 }U\text{ 的划分}\}.$$

在 $\mathcal{K}(U)$ 上按照的子集之间的包含关系, 可以定义以下类型的偏序关系.

定义 1.1.8　设 U 是论域, 且 $\mathcal{A},\mathcal{B}\in\mathcal{K}(U)$.

(1) 如果 $K_1=(U,\ \mathcal{A}), K_2=(U,\ \mathcal{B}), K_1\supseteq K_2$, 则称 K_1 比 K_2 精细.

(2) 如果对于任意的 $Y_j\in\mathcal{B}$, 存在 $X_i\in\mathcal{A}$, 使 $X_i\subseteq Y_j$, 则称 $\mathcal{A}$ 细于 $\mathcal{B}$, 记作 $\mathcal{A}\leqslant\mathcal{B}$.

(3) 如果对于任意的 $X_i\in\mathcal{A}$, 都存在 $Y_j\in\mathcal{B}$ 使得 $X_i\subseteq Y_j$, 则称 $\mathcal{A}$ 加细 $\mathcal{B}$, 记为 $\mathcal{A}\leqslant^*\mathcal{B}$.

定理 1.1.3　设 U 是论域, 则 $(\mathcal{K}(U),\ \leqslant)$ 与 $(\mathcal{K}(U),\ \supseteq)$ 都是偏序集.

证明　仅证明 $(\mathcal{K}(U),\ \leqslant)$ 是偏序集. 类似地可以证明 $(\mathcal{K}(U),\ \supseteq)$ 也是偏序集.

(1) 自反性. 显然成立.

(2) 反对称性. 若 $\mathcal{A} \leqslant \mathcal{B}$ 且 $\mathcal{B} \leqslant \mathcal{A}$, 任取 $Y_j \in \mathcal{B}$, 有 $X_i \in \mathcal{A}$, 使 $X_i \subseteq Y_j$, 但 $\mathcal{B} \leqslant \mathcal{A}$, 对 X_i, 有 $Y_{j'} \in \mathcal{B}$ 使 $Y_{j'} \subseteq X_i \subseteq Y_j$. 于是 $Y_{j'} \cap Y_j \neq \varnothing$, 所以 $Y_{j'} = Y_j$, 且 $X_i = Y_j$, 从而 $\mathcal{A} = \mathcal{B}$.

(3) 传递性. 若 $\mathcal{A}_1 \leqslant \mathcal{A}_2, \mathcal{A}_2 \leqslant \mathcal{A}_3$. 任取 $X_k \in \mathcal{A}, \exists Y_j \in \mathcal{A}_2$, 使 $Y_j \subseteq X_k$. 又由 $\mathcal{A}_1 \leqslant \mathcal{A}_2, \exists X_i \in \mathcal{A}_1$, 使 $X_i \subseteq Y_j \subseteq X_k$, 所以 $\mathcal{A}_1 \leqslant \mathcal{A}_3$. □

注 $(\mathcal{K}(U), \supseteq)$ 与 $(\mathcal{K}(U), \leqslant)$ 有区别.

定理 1.1.4 设 U 是论域, 且 $\mathcal{A}, \mathcal{B} \in \mathcal{K}(U)$, 若 $K_1 = (U, \mathcal{A}) \supseteq K_2 = (U, \mathcal{B})$, 则必有 $\mathcal{A} \leqslant \mathcal{B}$.

定理 1.1.4 的逆命题不一定成立.

例 1.1.1 设论域 $U = \{x_1, x_2, x_3, x_4, x_5, x_6\}$, $X_1 = \{x_1, x_2\}$, $X_2 = \{x_3, x_4, x_5\}$, $X_3 = \{x_6\}$; $Y_1 = \{x_1, x_2, x_3, x_4\}$, $Y_2 = \{x_5, x_6\}$, 则 $\mathcal{A} = \{X_1, X_2, X_3\}$, $\mathcal{B} = \{Y_1, Y_2\}$ 是 U 上的两个划分.

显然, $\mathcal{A} \leqslant \mathcal{B}$, 但 $K_1 = (U, \mathcal{A}) \not\supseteq K_2 = (U, \mathcal{B})$. 因为 $X_1 \subset Y_1, X_3 \subset Y_2$, 而 Y_1 与 Y_2 都不能由 X_i 的并表出, 所以 $K_1 \not\supseteq K_2$.

定义 1.1.9 设 $(A, \leqslant_A), (B, \leqslant_B)$ 都是偏序集. 如果存在一一映射 $f : A \to B$, 对于任意 $a_1, a_2 \in A$ 满足:

$$a_1 \leqslant_A a_2 \text{ 当且仅当 } f(a_1) \leqslant_B f(a_2),$$

则称 $(A, \leqslant_A)$ 与 $(B, \leqslant_B)$ 序同构, 记为

$$(A, \leqslant_A) \cong (B, \leqslant_B).$$

定理 1.1.5 设 U 是论域, 若 $\mathcal{A}, \mathcal{B} \in \mathcal{K}(U)$, 则 $(\mathcal{K}(U), \leqslant^*) \cong (\mathcal{K}(U), \subseteq)$.

1.2 粗糙集的基本概念与性质

1.2.1 精确集与粗糙集

定义 1.2.1[24] 若 R 是论域 U 上的等价关系, 则称 (U, R) 是近似空间, 也称为 Pawlak 近似空间. 由等价关系 R 产生的等价类为

$$U/R = \{[x_i]_R | x_i \in U\},$$

其中 $[x_i]_R = \{x_j \in U | (x_i, x_j) \in R\}$.

知识系统 $K = (U, R)$ 也称近似空间. 考察知识系统 $K = (U, R)$, R 为 U 上的等价关系. 将待表示的知识系统 K' 取为 $\mathcal{P}(U)$ 本身, 这是论域 U 上最精细的知识

系统, K' 用 K 表示就归结为 U 的任意子集 X 用 U/R 中元素的表示问题. 由此提出了粗糙集中的最重要的两个概念, 下近似和上近似.

定义 1.2.2[24]　设 $K=(U,R)$ 是一个知识系统, 对于任意 $X\subseteq U$, 称

$$\underline{R}(X)=\cup\{X_i|X_i\in U/R, X_i\subseteq X\}$$

为 X 关于 R 的下近似. 称

$$\overline{R}(X)=\cup\{X_i|X_i\in U/R, X_i\cap X\neq\varnothing\}$$

为 X 关于 R 的上近似.

如果 $\underline{R}(X)=\overline{R}(X)$, 则称 X 为 R 可定义集或精确集, 否则称 X 为 R 不可定义集或粗糙集(rough set).

如果 $X\notin K$, 则 $\underline{R}(X)\subset X\subset\overline{R}(X)$. 此时, 称

$$\mathrm{Bn}_R(X)=\overline{R}(X)-\underline{R}(X)$$

为 X 的边界, 这可反映 X 用 R 基本范畴的近似表示的一种误差界, 边界的基数 $|\mathrm{Bn}_R(X)|$ 反映了 X 的近似的绝对误差. 当 $|X|\neq 0$ 时, $\dfrac{|\mathrm{Bn}_R(X)|}{|X|}$ 反映了 X 的近似的相对误差.

由定义 1.2.2, 可直接得到下面的定理.

定理 1.2.1　设 (U,R) 是一个知识系统, 对于任意 $X\subseteq U$, 则

(1) X 为 R 可定义集当且仅当 X 能表示成某些 R 基本范畴的并;

(2) X 为 R 粗糙集当且仅当 $\underline{R}(X)\neq\overline{R}(X)$, 即 $\mathrm{Bn}_R(X)\neq\varnothing$.

定义 1.2.3　设 (U,R) 是一个知识系统, 对于任意 $X\subseteq U$, 由等价关系 R 确定的集合 X 的近似精度定义为

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

其中 $X\neq\varnothing$, $|X|$ 表示集合 X 的基数. 显然, $0\leqslant\alpha_R(X)\leqslant 1$.

定理 1.2.2　设 (U,R) 是一个知识系统, $X\subseteq U$, 则

(1) X 为 R 可定义集当且仅当 $\alpha_R(X)=1$;

(2) X 为 R 粗糙集当且仅当 $\alpha_R(X)<1$.

定义 1.2.4　设 (U,R) 是一个知识系统, 对于任意 $X\subseteq U$, X 的 R 粗糙度定义为 $\rho_R(X)$:

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

粗糙集又可引入下述区分:

(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.

设 $X \subseteq U$, 集合 $\{x_i \in U | x_i \notin X\}$ 称为 X 的补集, 记为 $\sim X$ 或 X^{c}. 例如, $X - Y = X \cap (\sim Y)$, 或 $X - Y = X \cap Y^{\mathrm{c}}$.

定理 1.2.3 设 (U, R) 是一个知识系统, $X \subseteq U$, 则

(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.

1.2.2 知识系统 $(U, \mathcal{P}(U))$ 中的粗关系

在 $\mathcal{P}(U)$ 的元素引入近似表示之后, 在知识系统的研究中就有新的关系值得注意. 粗糙集理论中还可给出下述三种近似相等的定义.

定义 1.2.5 设 $K = (U, R)$ 是知识系统, $X, Y \subseteq U$,

(1) 如果 $\underline{R}(X) = \underline{R}(Y)$, 则称 X 和 Y 为 R 下粗相等, 记作 $X \eqsim_R Y$;

(2) 如果 $\overline{R}(X) = \overline{R}(Y)$, 则称 X 和 Y 为 R 上粗糙相等, 记作 $X \simeq_R Y$;

(3) 如果 $X \eqsim_R Y$ 且 $X \simeq_R Y$, 则称 X 和 Y 为 R 粗糙相等, 记作 $X \approx_R Y$.

如果 R 为 U 上的等价关系, 则 $\eqsim_R$, $\simeq_R$ 和 $\approx_R$ 均为 $\mathcal{P}(U)$ 上的等价关系.

$(\mathcal{P}(U), \approx_R)$ 中的等价类是由 $K = (U, R)$ 中有共同边界的那些粗糙集组成, 而 R 精确集所在的等价类为单元素集.

$(\mathcal{P}(U), \simeq_R)$ 中的等价类是由 $K = (U, R)$ 中有共同内部的那些粗糙集组成, 这些等价类中的最小者就是它们的共同内部, 即 $X \in [X]_{\simeq_R}$, 有

$$\underline{R}(X) = \cap [X]_{\simeq_R} = \min [X]_{\simeq_R}.$$

比集合相等更一般的关系是集合的相互包含关系. 对于知识系统 $K = (U, R)$, $X, Y \subseteq U$ 作为粗糙集, 并不能完全用 R 基本范畴明确表示出来, 能表示的仅是 $(\underline{R}(X), \overline{R}(X))$ 与 $(\underline{R}(Y), \overline{R}(Y))$. 引入 U 子集的上近似与下近似后, 不仅可以引入 U 的子集之间的粗相等, 还可以定义 U 的子集之间的粗包含关系.

定义 1.2.6 设 $K = (U, R)$ 为知识系统, $X, Y \subseteq U$,

(1) 如果 $\underline{R}(X) \subseteq \underline{R}(Y)$, 则称集合 X 为 R 下包含于 Y, 记作 $X \underset{\sim R}{\subset} Y$;

(2) 如果 $\overline{R}(X) \subseteq \overline{R}(Y)$, 则称集合 X 为 R 上包含于 Y, 记作 $X \widetilde{\subset}_R Y$;

(3) 如果 $X \widetilde{\subset}_R Y$ 且 $X \underset{\sim R}{\subset} Y$, 则称集合 X 为 R 包含于 Y, 记作 $X \underset{\sim R}{\widetilde{\subset}} Y$.

容易验证, $\underset{\sim R}{\subset}$, $\widetilde{\subset}_R$ 和 $\underset{\sim R}{\widetilde{\subset}}$ 均为 $\mathcal{P}(U)$ 上的拟序关系. 若引入粗偏序集, 利用粗

反对称性, 即以粗相等替代反对称性中的相等条件引入粗结构产生粗系统. 而粗偏序关系可以引入粗规则的研究, 粗规则的形式表达, 粗规则的提取过程, 变精度的粗规则模型研究.

为叙述简单起见, 下面的定理中用 $\underset{\sim}{\subset}$, $\overset{\sim}{\subset}$ 和 $\underset{\sim}{\overset{\sim}{\subset}}$ 分别代替 $\underset{\sim R}{\subset}$, $\overset{\sim}{\subset}_R$ 和 $\underset{\sim R}{\overset{\sim}{\subset}}$.

定理 1.2.4　设 $K=(U,R)$ 为知识系统, 则 $\underset{\sim}{\subset}$, $\overset{\sim}{\subset}$ 和 $\underset{\sim}{\overset{\sim}{\subset}}$ 均为 $\mathcal{P}(U)$ 上的偏序关系, 且对于 $X,Y,Z\subseteq U$ 有下列性质:

(1) 如果 $X\subseteq Y$, 则 $X\underset{\sim}{\subset}Y$, $X\overset{\sim}{\subset}Y$ 且 $X\underset{\sim}{\overset{\sim}{\subset}}Y$;

(2) 如果 $X\underset{\sim}{\subset}Y$ 且 $Y\underset{\sim}{\subset}X$, 则 $X\eqsim Y$;

(3) 如果 $X\overset{\sim}{\subset}Y$ 且 $Y\overset{\sim}{\subset}X$, 则 $X\simeq Y$;

(4) 如果 $X\underset{\sim}{\overset{\sim}{\subset}}Y$ 且 $Y\underset{\sim}{\overset{\sim}{\subset}}X$, 则 $X\approx Y$;

(5) $X\overset{\sim}{\subset}Y$ 当且仅当 $X\cup Y\simeq Y$;

(6) $X\underset{\sim}{\subset}Y$ 当且仅当 $X\cap Y\eqsim Y$;

(7) 如果 $X\subseteq Y$, $X\eqsim X'$ 且 $Y\eqsim Y'$, 则 $X'\underset{\sim}{\subset}Y'$;

(8) 如果 $X\subseteq Y$, $X\simeq X'$ 且 $Y\simeq Y'$, 则 $X'\overset{\sim}{\subset}Y'$;

(9) 如果 $X\subseteq Y$, $X\approx_R X'$ 且 $Y\approx_R Y'$, 则 $X'\underset{\sim}{\overset{\sim}{\subset}}Y'$;

(10) 如果 $X'\overset{\sim}{\subset}X$ 且 $Y'\overset{\sim}{\subset}Y$, 则 $X'\cup Y'\overset{\sim}{\subset}X\cup Y$;

(11) 如果 $X'\underset{\sim}{\subset}X$ 且 $Y'\underset{\sim}{\subset}Y$, 则 $X'\cap Y'\underset{\sim}{\subset}X\cap Y$;

(12) 如果 $X\underset{\sim}{\subset}Y$ 且 $X\eqsim Z$, 则 $Z\underset{\sim}{\subset}Y$;

(13) 如果 $X\overset{\sim}{\subset}Y$ 且 $X\simeq Z$, 则 $Z\overset{\sim}{\subset}Y$;

(14) 如果 $X\underset{\sim}{\overset{\sim}{\subset}}Y$ 且 $Z\approx Y$, 则 $Z\underset{\sim}{\overset{\sim}{\subset}}Y$.

粗糙集理论表明, 一个元素是否属于某一集合, 不只是该元素具有的客观性质, 还与我们对它的了解程度有关. 同样, 集合的相等和包含不仅是一种客观性质, 与我们对所研究问题中的集合的了解程度有关. 也就是说对客观性质的了解受到主观的历史状况的约束、限制. 面对这一事实, 我们就应该把这种局限性明确地表示出来, 这就是近似表示的客观背景.

知识系统 $K=(U,R)$ 中的粗糙集是指那些不能精确表示的子集 X, 正是由于它们不能被精确地具体表示, 对它们的认识仅是抽象的. 认识到的只是对 X 的一种表示方式 $\underline{R}(X)$ 与 $\overline{R}(X)$, 即通过 $(\underline{R}(X),\overline{R}(X))$ 来获得关于 X 的认识. 因此, 与其说粗糙集是 X, 不如说是 $(\underline{R}(X),\overline{R}(X))$. 间接从 $(\underline{R}(X),\overline{R}(X))$ 得到关于 X 的信息就带有粗糙性, 即不精确性.

1.2.3 粗糙系统

在粗糙集组成的集族上, 引入粗关系就成为一种粗糙系统.

粗糙系统 $(\mathcal{P}(U), \eqsim_R)$ 是在 $\mathcal{P}(U)$ 上引入 R 下粗相等关系. 容易验证, R 下粗相等关系是 $\mathcal{P}(U)$ 上的等价关系. 从而, R 下粗相等的等价类构成 $\mathcal{P}(U)$ 的粗划分.

粗糙系统 $(\mathcal{P}(U), \overset{\subset}{\sim}_R)$ 是在 $\mathcal{P}(U)$ 上引入 R 下粗包含关系. 由定义 1.2.6 (1), $\mathcal{P}(U)$ 上的 R 下粗包含关系是一个拟序关系. 可以进一步指出, $(\mathcal{P}(U), \underset{\sim R}{\subset})$ 是一种粗偏序集. 事实上, 当粗包含与粗相等被同时引入时, R 下粗包含关系是 $(\mathcal{P}(U), \underset{\sim R}{\subset})$ 上的粗偏序关系. 自反性显然; 反对称性由定理 1.2.4 中的 (2) 推出; 传递性, 即若 $X \underset{\sim R}{\subset} Y$, 且 $Y \underset{\sim R}{\subset} Z$, 则 $X \underset{\sim R}{\subset} Z$.

因此, $(\mathcal{P}(U), \underset{\sim R}{\subset})$ 构成一个粗偏序系统.

粗糙系统中可能成员粗而关系不粗, 也可能成员不粗而关系粗, 也可能成员及关系两者皆粗. 粗糙性这一概念的泛化, 不限于集合的粗, 而推广为各种数学对象的粗, 而且 "粗" 成为近似方法的一种特定形式推广, 其研究领域是相当宽广的.

1.3 近似空间与近似算子

1.3.1 近似空间与近似算子

粗糙集理论可以看成是近似方法的一种拓广. 从数的近似、函数的近似、曲线的近似、曲面的近似到空间的近似、算子的近似, 近似表示是一个广泛课题. 我们将讨论的是概念的近似. 近似方法必须选择适当的近似工具 (如函数近似中多项式空间、三角多项式空间等), 还要选择适当的近似算子, 特别是不足近似与过剩近似 (如区域的内接多边形与外切多边形) 配合使用, 便于获得误差估计. 近似算子是从近似表示的对象集到近似工具集中的映射, 这也可以有多种可能选择. 例如, 函数近似表示中可选择插值算子 (有 L 插值与 H 插值等区分)、拟合算子或最佳逼近算子等. 近似空间的选择对近似表示的效果有直接的影响, 这是近似表示方法中首先要讨论的问题.

定义 1.3.1 设 U 为论域, 若 $\mathcal{F} \subseteq \mathcal{P}(U)$ 满足

(1) $U \in \mathcal{F}$;

(2) 如果 $A, B \in \mathcal{F}$, 则 $A \cup B \in \mathcal{F}$;

(3) 如果 $A \in \mathcal{F}$, 则 $\sim A \in \mathcal{F}$,

则称 $\mathcal{F}$ 为集代数, 其中 $\sim$ 表示集合的补运算.

容易证明, 若 $\mathcal{F}$ 为集代数, 则下列条件 $(1)'$ 与 $(2)'$ 分别等价于定义 1.3.1 的 (1) 与 (2).

(1)′ $\varnothing \in \mathcal{F}$;

(2)′ 如果 $A, B \in \mathcal{F}$, 则 $A \cap B \in \mathcal{F}$.

定理 1.3.1[134] 设 R 是论域 U 上的等价关系, 则

$$\sigma(U/R) = \left\{ \bigcup_{x_i \in X} [x_i]_R | X \subseteq U \right\}$$

是 $\mathcal{P}(U)$ 中的集代数, 即满足

(1) $U \in \sigma(U/R)$;

(2) 如果 $X, Y \in \sigma(U/R)$, 则 $X \cup Y \in \sigma(U/R)$;

(3) 如果 $X \in \sigma(U/R)$, 则 $\sim X \in \sigma(U/R)$.

知识表示的近似空间. 首先考察知识系统 $K = (U, R)$, R 为 U 上的等价关系. 将待表示的知识系统 K' 取为 $\mathcal{P}(U)$ 本身, 即 $K' = (U, \mathcal{P}(U))$ 为论域 U 上最精细的知识系统, K' 中的知识用 K 基本范畴表示就归结为 U 的任意子集 X 用 U/R 表示的问题.

定义 1.3.2 设 $K = (U, R)$ 是知识系统, $\mathcal{F} \subseteq \sigma(U/R)$, 对于 $X, Y \in \mathcal{P}(U)$,

若 $L : \mathcal{P}(U) \to \mathcal{F}$, 满足:

(L1) $L(X) \subseteq X$;

(L2) $L(U) = U$;

(L3) $L(X \cap Y) = L(X) \cap L(Y)$,

则称 L 为必然算子, $L(X)$ 称为 X 的必然知识.

若 $H : \mathcal{P}(U) \to \mathcal{F}$, 满足:

(H1) $X \subseteq H(X)$;

(H2) $H(\varnothing) = \varnothing$;

(H3) $H(X \cup Y) = H(X) \cup H(Y)$,

则称 H 为可能算子, $H(X)$ 称为 X 的可能知识.

若对于每一个 $X \in \mathcal{P}(U)$, 有 $H(\sim X) = \sim L(X)$, 则称算子 L 与 H 是对偶的.

若 $H(X) = X$ (或 $L(X) = X$), $X \in \mathcal{P}(U)$, 则称 X 为算子 H (或 L) 的不动点.

定义 1.2.2 中的 $\underline{R}$ 与 $\overline{R}$ 分别称为 $\mathcal{P}(U)$ 上的下近似算子与上近似算子.

容易证明, $\underline{R}$ 与 $\overline{R}$ 分别是 $\mathcal{P}(U)$ 上的必然算子与可能算子. 它们都是粗糙集理论的基本近似算子.

命题 1.3.1 设 $K = (U, R)$ 是知识系统, 对于 $X \subseteq U$, 则下列各条等价:

(1) $\underline{R}(X) = X$;

(2) $X \in K$;

(3) $\overline{R}(X) = X$;