

实数域上的粗糙函数及其 Galois 格构建

郭显娥¹, 王俊红²

(1. 山西大同大学数学与计算机科学学院, 大同 037009; 2. 山西大学计算机与信息技术学院, 太原 030006)

摘 要: 传统粗糙集理论源于集合论平台, 其上、下近似算子在描述函数方面存在缺陷。针对该问题, 利用定义在整数轴上能严格划分出单调实函数的标度工具, 提出上、下粗糙函数概念, 形成实数域上的粗糙函数模型。构建与其匹配的 Galois 格, 并通过可辨识矩阵对其概念格进行了知识约简。

关键词: 标度; Galois 格; 粗糙函数

Rough Function in Real Domain and Its Establishment of Galois Lattice

GUO Xian'e¹, WANG Jun-hong²

(1. School of Mathematics and Computer Science, Shanxi Datong University, Datong 037009;

2. School of Computer and Information Technology, Shanxi University, Taiyuan 030006)

【Abstract】 The traditional rough set theory is based on the set theory. It can be found that upper and lower approximation operators have some shortcomings in the description of function. This paper puts forward the definition of upper and lower rough function in view of scale, which can confirm strict partition of a monotone real function defined in the integer, then form rough function model of real domain. It establishes the corresponding Galois concept lattice in light of the analysis of rough function model. It simplifies the property of lattice concept by discernibility matrix.

【Key words】 scale; Galois lattice; rough function

1 概述

概念格和粗糙集是 20 世纪 80 年代初期产生的 2 个数据分析方法, 并在模式识别、机器学习、数据挖掘、信息检索等领域获得了广泛应用^[1]。概念格与粗糙集的研究已有了很多成果, 如: 通过引入实集的概念, 提出基于实集和实关系的集合理论, 将 Galois 连接扩展到了实二元关系等。已有方法可以解决区间值的概念格构建问题, 但其只针对数值闭区间, 含义较为狭窄。本文对文献[2]中粗糙集与粗糙函数的概念进行扩展, 提出基于实数域上粗糙函数的形式概念分析, 通过粗糙函数处理实数域上的数据, 完成其 Galois 格的构建, 并用粗糙函数的可辨识矩阵实现了概念格的约简。

2 实数域上粗糙集的基本理论

2.1 相关概念

实数域上粗糙集的相关概念是基于文献[3]的标度概念。

定义 1^[3] 令 $[n] = \{0, 1, \dots, n\}$ 为一有限整数集, R 为实数集。若单调增函数 $d: [n] \rightarrow R$ 满足: $\forall i, j \in [n], i < j$, 有 $d(i) < d(j)$, 则 d 为标度(Scale)。

设任一 $d: [n] \rightarrow R$, 确定一有限实数列 $Z = \{x_0, x_1, \dots, x_n\}$, 满足 $x_0 < x_1 < \dots < x_n$, 其中, $x_i = d(i), i \in [n]$, 易证标度满足下列定理。

定理 1^[3](标度双射定理) 标度 d 是 $[n] \rightarrow R$ 的双射, 其中, $d(i) = [x_0, x_n] \in R_n, i \in [n]$ 。

标度 d 是 $[n] \rightarrow Z$ 的双射, 其中, $d(i) = x_i \in Z, i \in [n]$ 。

定义 2^[4] 给定标度 $d: [n] \rightarrow R, \forall x \in R_n, x$ 的 d -标度下近似

和 d -标度上近似分别定义为 $d_*(x) = \max\{i \mid i \in [n], x_i \leq x\}$ 与 $d^*(x) = \min\{i \mid i \in [n], x_i \geq x\}$, 称 $\delta_d(x) = d^*(x) - d_*(x)$ 为 x 的 d 标度近似误差。

定义 3^[4] $\forall x \in R_n$, 定义 R_n 上对应于 d 的不可分辨关系 I_d :

$$I_d = \{(x, y) \mid x, y \in R_n \wedge d_*(x) = d_*(y) \wedge d^*(x) = d^*(y)\}$$

$\forall x \in R_n$, 不可分辨关系 I_d 在 R_n 上生成等价类 $[x]_d$ 为

$$[x]_d = I_d(x) = \{y \mid y \in R_n, d_*(x) = d_*(y) \wedge d^*(x) = d^*(y)\}$$

简写为

$$[x]_d = I_d(x) = \{y \mid y \in R_n, *d(x) = *d(y)\}$$

为方便起见, 用 $*d$ 表示 d^* 和 d_* 两者。 $[x]_d$ 构成 R_n 的一个划分,

$$[x]_d = \{\{x_0\}, (x_0, x_1), \{x_1\}, (x_1, x_2), \{x_2\}, \dots, (x_{n-1}, x_n), \{x_n\}\}$$

当 $x_i \leq x < x_{i+1}$ 时, $[x]_d = (x_i, x_{i+1})$; 当 $x = x_i$ 时, $[x]_d = \{x_i\}$ 。

定义 4 给定标度 d 及实数集 $Z = \{x_0, x_1, \dots, x_n\}$, 对 $\forall x \in R_n$, 定义 x 的 d -实数域下近似与 d -实数域上近似为

$$(I_d)_*(x) = \max\{x_i \mid x_i \in Z, x_i \leq x\}$$

$$(I_d)^*(x) = \min\{x_i \mid x_i \in Z, x_i \geq x\}$$

基金项目: 山西省国际科技合作基金资助项目(2010081005); 天津市自然科学基金重大基金资助项目(07JCZDJ06500)

作者简介: 郭显娥(1964—), 女, 副教授、硕士, 主研方向: 概念格, 知识发现; 王俊红, 讲师、博士

收稿日期: 2010-03-01 **E-mail:** scjguo@yahoo.com.cn

称 $\delta_{id}(x) = (I_d)^*(x) - (I_d)_*(x)$ 为 x 的 d -实数域近似误差。

定义 5 $\forall x \in R_n$, 若 $\delta_{id}(x) = (I_d)^*(x) - (I_d)_*(x) \neq 0$, 称 x 为 d -实数域上的粗糙数, 否则称为精确数。

例: 设 $x, y \in (x_2, x_3)$, 分别求出 x, y 的 d -标度上、下近似和 d -实数域上、下近似, 并判断 x, y 是否为粗糙数。

解: $d_*(x) = d_*(y) = 2, d^*(x) = d^*(y) = 3, (I_d)_*(x) = (I_d)_*(y) = x_2, (I_d)^*(x) = (I_d)^*(y) = x_3$, 显然 $(I_d)^*(x) \neq (I_d)_*(x)$, 故知, x, y 都是粗糙数。

推论 1

(1) $(I_d)_*(x) = d(d_*(x)), (I_d)^*(x) = d(d^*(x))$;

(2) $x, y \in R_n, x I_d y \Leftrightarrow *(I_d)(x) = *(I_d)(y)$, 其中, $*I_d$ 表示 $(I_d)_*$ 与 $(I_d)^*$ 两者。

证明:

(1) 据定义 2, $d_*(x) = \max\{i \mid i \in [n], x_i \leq x\}, d_*(x) \in [n]$, 对 $d_*(x)$ 作 d 运算得, $d(d_*(x)) = \max\{x_i \mid x_i \in Z, x_i \leq x\}$, 据定义 4, $(I_d)_*(x) = d(d_*(x))$ 成立, 同理可证 $(I_d)^*(x) = d(d^*(x))$ 也成立。

(2) 的证明由定义 3 易得, 略。

2.2 实数域上粗糙集的性质

由上述定义的近似算子 $*d$ 与 $*I_d$, 可推导出如下性质, 证明详见文献[4]。

(1) $(I_d)_*(x) = d(d_*(x))$;

(2) $(I_d)^*(x) = d(d^*(x))$;

(3) $(I_d)_*(x) \leq x \leq (I_d)^*(x)$;

(4) $x \leq y \Rightarrow *I_d(x) \leq *I_d(y)$;

(5) $*I_d(x \vee y) = *I_d(x) \vee *I_d(y)$;

(6) $*I_d(x \wedge y) = *I_d(x) \wedge *I_d(y)$ 。

由上面的内容, 得到如下结论: 任一标度唯一确定实数域上的一个划分, 唯一确定实数域上的一个等价关系, 即不可分辨关系, 因此, 在给定标度下实数可由整数近似代替。

3 实数域上的粗糙函数模型

给定 2 个标度 d 和 e , $d: [n] \rightarrow R, e: [m] \rightarrow R$, 令实函数 $f: R_n \rightarrow R_m, R_n = [x_0, x_n], R_m = [y_0, y_m]$, 将定义 2、定义 4 中的近似算子推广到二维平面中, 给出实函数 f 对于标度 d 和 e 的一元粗函数的定义。

定义 6 函数 $f: R_n \rightarrow R_m$ 的 e -上粗函数和下粗函数定义为 $\bar{f}_*(x) = I_{e*}(f(x)), \bar{f}^*(x) = I_e^*(f(x)), \delta_{id}^f(x) = \bar{f}^*(x) - \bar{f}_*(x)$ 称为 f 的 d -粗函数近似误差。

定义 7 对于函数 $f: R_n \rightarrow R_m$, 当 $\delta_{id}^f(x) = \bar{f}^*(x) - \bar{f}_*(x) \neq 0$ 时, 称 f 是 x 点处的粗糙函数, 否则称为是精确函数。

4 基于实数域粗糙函数的 Galois 格

定义 8 给定粗糙函数信息系统 $IS = (G, M, V, \bar{f})$, G 为实数域集, M 为属性集, V 为值集, $\bar{f}$ 为实数域上的粗糙函数, 且 $\bar{f}: G \times M \rightarrow V$, 设 $x \in X \subseteq G, a \in M$, 则 $\bar{f}(x, a)$ 表示实数 x 在属性 a 上的值, 且:

$$\bar{f}(x, a) = [(I_e)_*(f(x, a)), (I_e)^*(f(x, a))] = [\bar{f}_*(x, a), \bar{f}^*(x, a)]$$

将粗糙函数信息系统 IS 简称为粗糙信息系统。

推论 2 在粗糙信息系统 $IS = (G, M, V, \bar{f})$ 上, 有序实数对

$$(x, a) \in G \times M, \bar{f}_*(x, a) = \inf V_a, \bar{f}^*(x, a) = \sup V_a。$$

推论 2 可由定义 4、定义 6、定义 8 证得。

定义 9 粗糙信息系统 $IS = (G, M, V, \bar{f})$, $M = \{a_1, a_2, \dots, a_m\}$, 任意 $x, y \in G$ 对应的 M 值属性集合分别为

$$V_x = \{\bar{f}(x, a_1), \bar{f}(x, a_2), \dots, \bar{f}(x, a_m)\}$$

$$V_y = \{\bar{f}(y, a_1), \bar{f}(y, a_2), \dots, \bar{f}(y, a_m)\}$$

定义上述集合元素的交、并与蕴含关系如下:

$$V_x \wedge V_y = \{\bar{f}(x, a_1) \wedge \bar{f}(y, a_1), \bar{f}(x, a_2) \wedge$$

$$\bar{f}(y, a_2), \dots, \bar{f}(x, a_m) \wedge \bar{f}(y, a_m)\}$$

$$V_x \vee V_y = \{\bar{f}(x, a_1) \vee \bar{f}(y, a_1), \bar{f}(x, a_2) \vee$$

$$\bar{f}(y, a_2), \dots, \bar{f}(x, a_m) \vee \bar{f}(y, a_m)\}$$

$$V_x \rightarrow V_y = \{\bar{f}(x, a_1) \rightarrow \bar{f}(y, a_1),$$

$$\bar{f}(x, a_2) \rightarrow \bar{f}(y, a_2), \dots, \bar{f}(x, a_m) \rightarrow \bar{f}(y, a_m)\}$$

$$V_x \Leftrightarrow V_y \Leftrightarrow V_x \rightarrow V_y \wedge V_y \rightarrow V_x$$

其中, 运算符 “ $\rightarrow$ ” 表示蕴含关系。

定义 10 将 G 中对象对应的 M 值属性集与其通过交、并运算后得到的所有子集记为 Ω , 也即:

$$\bigcup_{x \in G} \{\bar{f}(x, a_1), \bar{f}(x, a_2), \dots, \bar{f}(x, a_m)\} \subseteq \Omega$$

$$\bar{f}(x_p, a_i) \wedge \bar{f}(x_q, a_i) \subseteq \Omega$$

$$\bar{f}(x_p, a_i) \vee \bar{f}(x_q, a_i) \subseteq \Omega, x_p, x_q \in G, a_i \in M$$

定义 11 在粗糙信息系统 $IS = (G, M, V, \bar{f})$ 上, 对于 $X \subseteq G, Y \in \Omega$, 定义 2 个映射:

$$\tau: P(G) \rightarrow \Omega, \tau(X) = \{\bigwedge Y_x \mid x \in X\},$$

$$\mu: \Omega \rightarrow P(G), \mu(Y) = \{x \in G \mid Y \rightarrow Y_x\}$$

称 $\mu(Y)$ 是 Y 的外延, $\tau(X)$ 是 X 的内涵, 则序对 (τ, μ) 称为一个 Galois 连接。

定义 12 在 $IS = (G, M, V, \bar{f})$ 上, 对于 $X \subseteq G, Y \in \Omega$, 序对 (X, Y) 满足 $\tau(X) = Y \wedge \mu(Y) = X$, 则称 (X, Y) 为一个概念, 所有这些概念构成 IS 的概念格, 记为 LR , 也称为实数域粗糙函数的 Galois 格。 LR 上的序关系 $(X, Y) \leq (X', Y') \Leftrightarrow X \subseteq X' \wedge Y' \subseteq Y$, 显然, “ $\leq$ ” 为 LR 上的一个偏序。

定理 2 偏序集 $(LR, \leq)$ 是一个完备格, 其中下确界和上确界定义为

$$\bigwedge_{t \in T} (A_t, B_t) = (\bigcap_{t \in T} A_t, \bigvee_{t \in T} B_t)$$

$$\bigvee_{t \in T} (A_t, B_t) = (\bigcup_{t \in T} A_t, \bigwedge_{t \in T} B_t)$$

其中, T 为指标集; $t \in T, A_t \subseteq G, B_t \in \Omega$ 。

证明: 先证明 $(\bigcap_{t \in T} A_t, \bigvee_{t \in T} B_t) \in LR$ 。因为 $(A_t, B_t) \in LR$, 所以 $\tau(A_t) = B_t, \mu(B_t) = A_t$, 而 $\tau(\bigcap_{t \in T} A_t) = \bigvee_{t \in T} \tau(A_t) = \bigvee_{t \in T} B_t, \mu(\bigvee_{t \in T} B_t) = \bigcap_{t \in T} \mu(B_t) = \bigcap_{t \in T} A_t$, 因此, $(\bigcap_{t \in T} A_t, \bigvee_{t \in T} B_t) \in LR$, 同理有: $(\bigcup_{t \in T} A_t, \bigwedge_{t \in T} B_t) \in LR$ 。再证明: $(\bigcup_{t \in T} A_t, \bigwedge_{t \in T} B_t)$ 为集合 $\{(A_t, B_t) \mid t \in T\}$ 的上确界, 下确界用同样方法证得。因为 $\bigcup_{t \in T} A_t \subseteq \mu\tau(\bigcup_{t \in T} A_t)$, $\mu\tau(\bigcup_{t \in T} A_t)$ 是包含 $\bigcup_{t \in T} A_t$ 的最小闭集, 而 $\tau\mu\tau(\bigcup_{t \in T} A_t) = \tau(\bigcup_{t \in T} A_t) =$

$\bigwedge_{t \in T} B_t$ ，所以 $(\bigcup_{t \in T} A_t, \bigwedge_{t \in T} B_t)$ 为集合 $\{(A_t, B_t) | t \in T\}$ 的上确界。

5 实例分析

基于实数域粗糙函数 Galois 格的构建，可以采用文献[5]中利用区分矩阵求核约简的算法，并结合文献[6-8]用矩阵算法构造 Galois 格等方法即可实现，算法描述如下：首先计算粗糙函数信息系统 $IS = (G, M, V, \bar{f})$ 的可辨识矩阵，并根据可辨识矩阵求出可辨识函数；删除重复元素，利用吸取合取进行处理；最后计算核属性集，得到约简后的 Galois 格。

粗糙信息系统 IS 的数据如表 1 所示。

	a_1	a_2	a_3	a_4
x_1	[0.825, 1.000]	[0.721, 0.944]	[0.499, 0.963]	[0.588, 0.822]
x_2	[0.609, 0.899]	[0.566, 0.888]	[0.367, 0.858]	[0.424, 0.666]
x_3	[0.544, 0.855]	[0.299, 0.503]	[0.500, 1.000]	[0.645, 1.000]
x_4	[0.522, 0.700]	[0.466, 0.555]	[0.323, 0.756]	[0.489, 0.744]
x_5	[0.522, 0.700]	[0.308, 0.499]	[0.123, 0.299]	[0.343, 0.500]

根据表 1，构建其约简后的 Galois 格。操作步骤如下：

(1) 求初始化矩阵，矩阵行数为 $\frac{|G| \times (|G| - 1)}{2}$ ，列数为 $|M|$ ，矩阵元素为 0 或 1，计算公式如下：

$$\lambda((p, q), i) = \begin{cases} 1 & \bar{f}(x_p, a_i) \wedge \bar{f}(x_q, a_i) \neq \phi, \quad p < q \\ 0 & \text{others} \end{cases}$$

本例中矩阵的行数为 10，列数为 4，初始化矩阵为

0	0	0	0
0	1	0	0
1	1	0	0
1	1	1	1
0	1	0	0
0	1	0	0
0	1	1	0
0	0	0	0
0	0	1	1
0	0	1	0

(2) 求可辨识矩阵，即得候选约简如下：

	x_1	x_2	x_3	x_4	x_5
x_1	ϕ	ϕ	ϕ	ϕ	ϕ
x_2	ϕ	ϕ	ϕ	ϕ	ϕ
x_3	a_2	$a_2 a_4$	ϕ	ϕ	ϕ
x_4	$a_1 a_2$	a_2	ϕ	ϕ	ϕ
x_5	$a_1 a_2 a_3$	$a_2 a_3$	$a_3 a_4$	a_3	ϕ

(3) 求可辨识函数，先吸取再合取，得到核属性集：

$$a_2 \wedge (a_1 \vee a_2) \wedge (a_1 \vee a_2 \vee a_3) \wedge (a_2 \vee a_4) \wedge a_2 \wedge (a_2 \vee a_3) \wedge (a_3 \vee a_4) \wedge a_3 = a_2 \wedge a_3$$

其中，核为 $\{a_2\}, \{a_3\}$ 。

(4) 得到约简后的 Galois 格如图 1 所示，其中格结点只写出其外延。

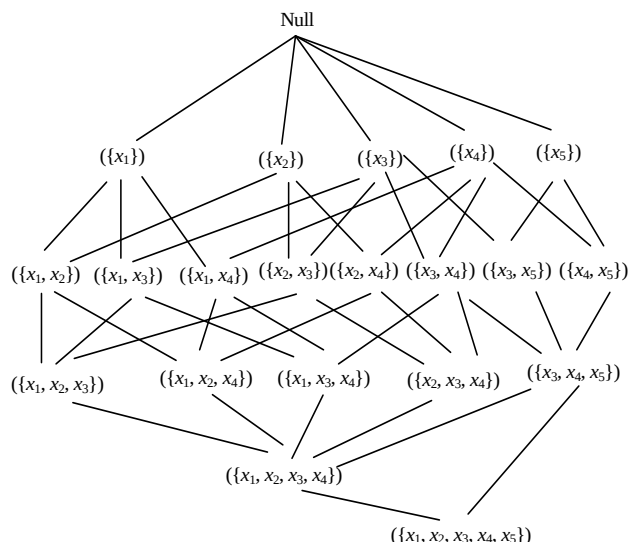


图 1 粗糙信息系统 IS 约简后的 Galois 格

6 结束语

粗糙集作为处理不确定性问题的数学工具，可以从不同的角度对模糊和不确定性信息进行处理。本文以实数域上的粗糙函数构建实数函数 Galois 格，与一般区间数概念格的不同之处在于：(1) 数据处理是完备的；(2) 更适合处理含有区间实数的信息表。基于粗糙函数运算定义的实数 Galois 格具有更强的适应性和扩展性。

参考文献

- [1] Wille R. Restructuring Lattice Theory: An Approach Based on Hierarchies of Concepts[M]. [S. l.]: Kluwer Academic Publishers, 1982.
- [2] Wang Yun, Guan Yanyong, Shi Kaiqun. Basic Theory in the New Real Line-scale Rough Function Model[J]. Journal of Information and Computing Science, 2007, 2(1): 41-47.
- [3] Pawlak Z. Rough Sets, Rough Function and Rough Calculus[M]// Rough-fuzzy Hybridization: A New Trend in Decision-making. [S. l.]: Springer-Verlag, 1999.
- [4] 王 耘. 粗糙集与粗糙函数模型研究[D]. 济南: 山东大学, 2008.
- [5] 王志海, 胡可云, 刘宗田, 等. 概念格上粗糙集合运算与函数依赖生成[J]. 清华大学学报: 自然科学版, 1998, 8(S2): 1-4.
- [6] 杨 丽, 徐 扬. 基于形式背景的概念格约简及其修复[J]. 计算机工程, 2008, 34(9): 22-24.
- [7] 余 远, 钱 旭, 钟 锋, 等. 基于最大概念的概念格增量构造算法[J]. 计算机工程, 2009, 35(21): 62-64.
- [8] Wu Qiang, Liu Zongtian, Shi Baisheng. Matrix Computation for Concept Lattices[C]//Proc. of the 5th International Conference on Cognitive Informatics. Beijing, China: [s. n.], 2006.

编辑 顾姣健