

文章编号:1671-9352(2007)10-0118-05

函数 S-粗集与规律积分度量

谢云龙¹, 杜英玲²

(1. 杭州师范大学 数学系, 浙江 杭州 310036; 2. 山东大学 控制科学与工程学院 山东 济南 250061)

摘要:利用函数单向 S-粗集和函数单向 S-粗集对偶, 给出 f -规律知识, \bar{f} -规律知识, f -规律和 \bar{f} -规律的概念, 利用这些概念, 给出规律知识生成的规律之间的关系和积分度量.

关键词:函数 S-粗集; f -规律; \bar{f} -规律; 积分度量

中图分类号: O159 **文献标志码:** A

Function S-rough sets and integral metric of laws

XIE Yun-long¹, DU Ying-ling²

(1. Department of Mathematics, Hangzhou Normal University, Hangzhou 310036, Zhejiang, China;
2. School of Control Science and Engineering, Shandong University, Jinan 250061, Shandong, China)

Abstract: Using function one direction singular rough sets and dual of function one direction singular rough sets, the concepts of f -law knowledge, \bar{f} -law knowledge, f -law and \bar{f} -law were given. With these concepts, the relation of laws generated by law knowledge, and the integral metric were both discussed.

Key words: function S-rough sets; f -law; \bar{f} -law; integral metric

0 引言

波兰数学家 Z. Pawlak 教授于 1982 年提出粗集^[1], 粗集得到了广泛的应用. 2005 年史开泉教授提出了函数 S-粗集理论^[2], 函数 S-粗集是改进了 S-粗集^[3-5]得到的, S-粗集是改进了 Z. Pawlak 粗集得到的, 将粗集应用从静态推广到动态, 从数据挖掘推广到规律挖掘. 函数 S-粗集(函数单向 S-粗集, 函数单向 S-粗集对偶, 函数双向 S-粗集)是以 α -函数等价类 $[u]$ 来定义的, $\forall u_i \in [u]$ 是一个函数, 一个函数是一个规律^[5,6]. α -函数等价类 $[u] = \{u_1, u_2, \dots, u_n\}$ 具有这样的特性: 如果在 $[u]$ 的属性集 α 中增加属性, 则 $\text{card}([u])$ 减少, $[u]$ 生成的规律曲线随着 α 中的属性增加而下移; 反之, 如果在 $[u]$ 的属性集 α 中减少属性, 则 $\text{card}([u])$ 增加, $[u]$ 生成的规律曲线随着 α 中的属性减少而上移. 函数 S-粗集的粗规律曲线是动态变化的, 这些规律之间存在着怎样的积分关系又具有怎样的积分度量特征本文给出讨论.

函数单向 S-粗集和函数单向 S-粗集对偶的结构与更多的概念, 见[3-11].

1 规律知识与规律生成

约定: 在下面的讨论中, α -函数等价类 $[u(x)]$ 记作 $[u]$; α -函数等价类 $[u]$ 与规律知识 $[u]$ 是同一个概念, 对它们不加区分直接使用; 规律知识简称知识; 特征值分量取值均为非负.

定义 1.1 设 $[u] = \{u_1, u_2, \dots, u_n\}$ 是一知识集, $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_\lambda\}$ 是 $[u]$ 的属性集, 称 $[u]^f$ 是 $[u]$ 的 f -知识, 如果 $[u]^f$ 的属性集 α^f 与 $[u]$ 的属性集 α 满足

$$\alpha \subseteq \alpha^f. \quad (1.1)$$

其中 $\alpha^f = \alpha \cup \{f(\beta)\}$, $\beta \in V$, $\beta \bar{\in} \alpha$, $f(\beta) = \alpha' \in \alpha$, 这里: $f \in F$ 是 V 上的属性迁移^[6,7], V 是属性论域.

显然有 $\text{card}([u]^f) \leq \text{card}([u])$.

定义 1.2 称 $[u]^{\bar{f}}$ 是 $[u]$ 的 \bar{f} -知识, 如果 $[u]^{\bar{f}}$ 的属性集 $\alpha^{\bar{f}}$ 与 $[u]$ 的属性集 α 满足

$$\alpha^{\bar{f}} \subseteq \alpha. \quad (1.2)$$

其中 $\alpha^{\bar{f}} = \alpha \setminus \{\bar{f}(\alpha_i)\}$, $\alpha_i \in \alpha$, $\bar{f}(\alpha_i) = \beta_i \bar{\in} \alpha$. 这里: $\bar{f} \in \bar{F}$ 是 V 上的属性迁移^[7,11].

显然有 $\text{card}([u]) \leq \text{card}([u]^{\bar{f}})$.

定义 1.3 给定 $[u] = \{u_1, u_2, \dots, u_n\}$, 设 u_1, u_2, \dots, u_n 分别具有特征值

$$\begin{aligned} x_1 &= (x_1(1), x_1(2), \dots, x_1(n)), \\ x_2 &= (x_2(1), x_2(2), \dots, x_2(n)), \\ &\vdots \\ x_n &= (x_n(1), x_n(2), \dots, x_n(n)). \end{aligned} \quad (1.3)$$

利用 $y(k) = \sum_{i=1}^n x_i(k)$, $k = 1, 2, \dots, n$, 得到

$$y = (y(1), y(2), \dots, y(n)), \quad (1.4)$$

$\forall y(k) \in R^+$, $k = 1, 2, \dots, n$; 令 $y_j = y(j)$, $j = 1, 2, \dots, n$, 则有数据点

$$(1, y_1), (2, y_2), \dots, (n, y_n), \quad (1.5)$$

利用 $p(x) = \sum_{j=1}^n y_j \prod_{i \neq j} \frac{x - x_i}{x_j - x_i}$ 得到 $[u]$ 生成的多项式函数 $p(x)$:

$$p(x) = a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \dots + a_1x + a_0. \quad (1.6)$$

多项式函数 $p(x)$ 称作知识 $[u]$ 的生成规律, 简称 $p(x)$ 是 $[u]$ 的规律.

定义 1.4 给定 $[u]^f = \{u_1, u_2, \dots, u_m\}$, $m \leq n$, 设 u_1, u_2, \dots, u_m 分别具有特征值

$$\begin{aligned} x_1 &= (x_1(1), x_1(2), \dots, x_1(n)), \\ x_2 &= (x_2(1), x_2(2), \dots, x_2(n)), \\ &\vdots \\ x_m &= (x_m(1), x_m(2), \dots, x_m(n)), \end{aligned} \quad (1.7)$$

利用 $y^f(k) = \sum_{i=1}^m x_i(k)$, $k = 1, 2, \dots, n$, 得到

$$y^f = (y^f(1), y^f(2), \dots, y^f(n)), \quad (1.8)$$

$\forall y^f(k) \in R^+$, $k = 1, 2, \dots, n$; 令 $y_j^f = y^f(j)$, $j = 1, 2, \dots, n$, 则有下列数据点

$$(1, y_1^f), (2, y_2^f), \dots, (n, y_n^f), \quad (1.9)$$

相似地得到知识 $[u]^f$ 的生成规律 $p(x)^f$:

$$p(x)^f = b_{n-1}x^{n-1} + b_{n-2}x^{n-2} + \dots + b_1x + b_0. \quad (1.10)$$

简称 $p(x)^f$ 是 $[u]^f$ 的 f -规律.

定义 1.5 给定 $[u]^{\bar{f}} = \{u_1, u_2, \dots, u_t\}$, $t \geq n$, 设 u_1, u_2, \dots, u_t 分别具有特征值

$$\begin{aligned} x_1 &= (x_1(1), x_1(2), \dots, x_1(n)), \\ x_2 &= (x_2(1), x_2(2), \dots, x_2(n)), \\ &\vdots \\ x_t &= (x_t(1), x_t(2), \dots, x_t(n)). \end{aligned} \quad (1.11)$$

利用 $y^{\bar{f}}(k) = \sum_{i=1}^t x_i(k)$, $k = 1, 2, \dots, n$, 得到

$$y^{\bar{f}} = (y^{\bar{f}}(1), y^{\bar{f}}(2), \dots, y^{\bar{f}}(n)), \quad (1.12)$$

$\forall y^{\bar{f}}(k) \in R^+, k = 1, 2, \dots, n$; 令 $y_j^{\bar{f}} = y^{\bar{f}}(j), j = 1, 2, \dots, n$, 则有数据点

$$(1, y_1^{\bar{f}}), (2, y_2^{\bar{f}}), \dots, (n, y_n^{\bar{f}}), \quad (1.13)$$

相似地得到知识 $[u]^{\bar{f}}$ 的生成规律 $p(x)^{\bar{f}}$:

$$p(x)^{\bar{f}} = c_{n-1}x^{n-1} + c_{n-2}x^{n-2} + \dots + c_1x + c_0. \quad (1.14)$$

简称 $p(x)^{\bar{f}}$ 是 $[u]$ 的 \bar{f} -规律.

由定义 1.1 ~ 1.5 得到:

定理 1.1 若 $p(x)$ 是 $[u]$ 的规律, $p(x)^f$ 是 $[u]$ 的 f -规律, 则在给定的规律区间 $[a, b]$ 上, $p(x)^f$ 与 $p(x)$ 满足

$$p(x)^f \leq p(x). \quad (1.15)$$

证明 由 $[u]^f$ 是 $[u]$ 的 f -知识可知 $\text{card}([u]^f) \leq \text{card}([u])$, 则 $[u]^f$ 的特征值对应点的离散数据叠加个数 $\leq [u]$ 的特征值对应点的离散数据叠加个数, 即 $\sum_{i=1}^m x_i(k) \leq \sum_{i=1}^n x_i(k)$, 故 $p(x)^f$ 的规律曲线在 $p(x)$ 的规律曲线的下方, 所以 $p(x)^f \leq p(x)$.

定理 1.2 若 $p(x)$ 是 $[u]$ 的规律, $p(x)^{\bar{f}}$ 是 $[u]$ 的 \bar{f} -规律, 则在给定的规律区间 $[a, b]$ 上, $p(x)^{\bar{f}}$ 与 $p(x)$ 满足

$$p(x) \leq p(x)^{\bar{f}}. \quad (1.16)$$

证明类似定理 1.1, 略.

利用定理 1.1, 1.2 直接得到:

定理 1.3 若 $p(x), p(x)^f, p(x)^{\bar{f}}$ 分别是 $[u]$ 的规律, $[u]$ 的 f -规律, $[u]$ 的 \bar{f} -规律, 则在给定的规律区间 $[a, b]$ 上, 有

$$p(x)^{\bar{f}} \leq p(x) \leq p(x)^f. \quad (1.17)$$

2 f -规律与 \bar{f} -规律的积分度量关系

约定: 在这一节中, $p(x), p(x)^f, p(x)^{\bar{f}}$ 是区间 $[a, b]$ 上的规律.

定理 2.1 (规律 $p(x)$ 与规律 $p(x)^f$ 的积分度量 f -关系) 若 $p(x)$ 是 $[u]$ 的规律, $p(x)^f$ 是 $[u]$ 的 f -规律, 则

$$\int_a^b p(x)^f dx \leq \int_a^b p(x) dx. \quad (2.1)$$

证明 利用定理 1.1 直接得到, 略.

定理 2.2 给定规律序列 $p(x)_1^f, p(x)_2^f, \dots, p(x)_n^f$, 若它们的属性集 $\alpha_1^f, \alpha_2^f, \dots, \alpha_n^f$ 满足

$$\alpha_n^f \subseteq \alpha_{n-1}^f \subseteq \dots \subseteq \alpha_1^f, \quad (2.2)$$

则下列不等式成立, 而且

$$\int_a^b p(x)_1^f dx \leq \int_a^b p(x)_2^f dx \leq \dots \leq \int_a^b p(x)_n^f dx. \quad (2.3)$$

证明 由 $\alpha_n^f \subseteq \alpha_{n-1}^f \subseteq \dots \subseteq \alpha_1^f \Rightarrow \text{card}([u]_1^f) \leq \text{card}([u]_2^f) \leq \dots \leq \text{card}([u]_n^f) \Rightarrow p(x)_1^f \leq p(x)_2^f \leq \dots \leq p(x)_n^f \Rightarrow \int_a^b p(x)_1^f dx \leq \int_a^b p(x)_2^f dx \leq \dots \leq \int_a^b p(x)_n^f dx$.

推论 1 设 $\{p(x)_1^f, p(x)_2^f, \dots, p(x)_n^f\}$ 是 $[u]$ 的 f -规律集, 任取 $p(x)_k^f \in \{p(x)_1^f, p(x)_2^f, \dots, p(x)_n^f\}$, 则

$$\int_a^b p(x)_k^f dx \leq \int_a^b p(x) dx. \quad (2.4)$$

定理 2.3 (规律 $p(x)$ 与规律 $p(x)^{\bar{f}}$ 的积分度量 \bar{f} -关系) 若 $p(x)$ 是 $[u]$ 的规律, $p(x)^{\bar{f}}$ 是 $[u]$ 的 \bar{f} -规律, 则

$$\int_a^b p(x) dx \leq \int_a^b p(x)^{\bar{f}} dx. \quad (2.5)$$

证明 利用定理 1.2 直接得到, 略.

定理 2.4 给定规律序列 $p(x)_1^f, p(x)_2^f, \dots, p(x)_n^f$, 若它们的属性集 $\alpha_1^f, \alpha_2^f, \dots, \alpha_n^f$ 满足

$$\alpha_n^f \subseteq \alpha_{n-1}^f \subseteq \dots \subseteq \alpha_1^f, \quad (2.6)$$

则下列不等式成立, 而且

$$\int_a^b p(x)_1^f dx \leq \int_a^b p(x)_2^f dx \leq \dots \leq \int_a^b p(x)_n^f dx. \quad (2.7)$$

证明类似定理 2.2, 略.

推论 2 设 $\{p(x)_1^f, p(x)_2^f, \dots, p(x)_n^f\}$ 是 $[u]$ 的 \bar{f} -规律集, 任取 $p(x)_i^f \in \{p(x)_1^f, p(x)_2^f, \dots, p(x)_n^f\}$, 则

$$\int_a^b p(x) dx \leq \int_a^b p(x)_i^f dx. \quad (2.8)$$

利用定理 2.1, 2.3 直接得到:

定理 2.5 若 $p(x)$ 是 $[u]$ 的规律, $p(x)^f$ 是 $[u]$ 的 f -规律, $p(x)^{\bar{f}}$ 是 $[u]$ 的 \bar{f} -规律, 则

$$\int_a^b p(x)^f dx \leq \int_a^b p(x) dx \leq \int_a^b p(x)^{\bar{f}} dx. \quad (2.9)$$

定理 2.6 设 $p(x), p(x)^f, p(x)^{\bar{f}}$ 分别是 $[u]$ 的规律, $[u]$ 的 f -规律, $[u]$ 的 \bar{f} -规律, 若

$$p(x) - p(x)^f = p(x)^{\bar{f}} - p(x), \quad (2.10)$$

则

$$\int_a^b p(x) dx = \frac{1}{2} \left(\int_a^b p(x)^f dx + \int_a^b p(x)^{\bar{f}} dx \right). \quad (2.11)$$

证明 由(2.10)得到 $p(x) = \frac{1}{2}(p(x)^f + p(x)^{\bar{f}})$, 然后将这个等式的两边在区间 $[a, b]$ 上积分可以得到

$$\int_a^b p(x) dx = \frac{1}{2} \left(\int_a^b p(x)^f dx + \int_a^b p(x)^{\bar{f}} dx \right).$$

推论 3 设 $p(x), p(x)^f, p(x)^{\bar{f}}$ 分别是 $[u]$ 的规律, $[u]$ 的 f -规律, $[u]$ 的 \bar{f} -规律, 且

$$\begin{aligned} p(x) &= a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \dots + a_1x + a_0, \\ p(x)^f &= b_{n-1}x^{n-1} + b_{n-2}x^{n-2} + \dots + b_1x + b_0, \\ p(x)^{\bar{f}} &= c_{n-1}x^{n-1} + c_{n-2}x^{n-2} + \dots + c_1x + c_0. \end{aligned}$$

若

$$a_i - b_i = c_i - a_i, \quad i = 0, 1, \dots, n-1, \quad (2.12)$$

则

$$\int_a^b p(x) dx = \frac{1}{2} \left(\int_a^b p(x)^f dx + \int_a^b p(x)^{\bar{f}} dx \right). \quad (2.13)$$

定理 2.7 设 $p(x), p(x)^f, p(x)^{\bar{f}}$ 分别是 $[u]$ 的规律, $[u]$ 的 f -规律, $[u]$ 的 \bar{f} -规律, 若在规律区间 $[a, b]$ 上存在三点 θ, η, τ 分别满足

$$\begin{aligned} p(\theta) \cdot (b-a) &= \int_a^b p(x) dx, \\ p(\eta)^f \cdot (b-a) &= \int_a^b p(x)^f dx, \end{aligned} \quad (2.14)$$

$$p(\tau)^{\bar{f}} \cdot (b-a) = \int_a^b p(x)^{\bar{f}} dx,$$

且

$$p(\theta) - p(\eta)^f = p(\tau)^{\bar{f}} - p(\theta). \quad (2.15)$$

则

$$\int_a^b p(x) dx = \frac{1}{2} \left(\int_a^b p(x)^f dx + \int_a^b p(x)^{\bar{f}} dx \right). \quad (2.16)$$

证明 由(2.15)得到 $p(\theta) \cdot (b-a) - p(\eta)^f \cdot (b-a) = p(\tau)^{\bar{f}} \cdot (b-a) - p(\theta) \cdot (b-a)$, 结合(2.14)有 $\int_a^b p(x) dx - \int_a^b p(x)^f dx = \int_a^b p(x)^{\bar{f}} dx - \int_a^b p(x) dx$, 所以 $\int_a^b p(x) dx = \frac{1}{2} \left(\int_a^b p(x)^f dx + \int_a^b p(x)^{\bar{f}} dx \right)$.

3 应用举例

考虑到用插值多项式计算的复杂性, 在本节我们用简单的几个数据来说明规律的生成以及各规律之间的积分度量关系.

设系统 m 的稳定状态在终端输出的是由属性集 $\alpha = \{\alpha_1, \alpha_2, \alpha_3\}$ 对应的知识 $[u] = \{u_1, u_2, u_3\}$, 系统 m 在运行中受到属性迁移的干扰, 在终端输出的是由属性集 $\alpha^f = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ 对应的知识 $[u]^f = \{u_1, u_2\}$ 或者是由属性集 $\alpha^{\bar{f}} = \{\alpha_1, \alpha_2\}$ 对应的知识 $[u]^{\bar{f}} = \{u_1, u_2, u_3, u_4\}$, 这两种属性迁移生成的知识我们并在一起讨论, 其中 $u_1 = \{1.4, 1.8, 1.5, 1.1, 1.7\}$, $u_2 = \{1.1, 1.7, 1.8, 1.3, 1.4\}$, $u_3 = \{1.2, 1.8, 1.9, 1.1, 1.4\}$, $u_4 = \{1.8, 1.1, 1.9, 2.1, 1.7\}$. 利用第一节内容得到由知识 $[u], [u]^f, [u]^{\bar{f}}$ 生成的数据 $y = (3.7, 5.3, 5.2, 3.5, 4.5)$, $y^f = (2.5, 3.5, 3.3, 2.4, 3.1)$, $y^{\bar{f}} = (5.5, 6.4, 7.1, 5.6, 6.2)$, 则有三组数据点 $(1, 3.7), (2, 5.3), (3, 5.2), (4, 3.5), (5, 4.5); (1, 2.5), (2, 3.5), (3, 3.3), (4, 2.4), (5, 3.1); (1, 5.5), (2, 6.4), (3, 7.1), (4, 5.6), (5, 6.2)$. 利用插值多项式 $p(x) = \sum_{j=1}^n y_j \prod_{i \neq j} \frac{x - x_i}{x_j - x_i}$ 得到知识 $[u], [u]^f, [u]^{\bar{f}}$ 的生成规律:

$$\begin{aligned} p(x) &= 0.1750x^4 - 1.7334x^3 + 5.1750x^2 - 4.4167x + 4.5000, \\ p(x)^f &= 0.0750x^4 - 0.6667x^3 + 1.5250x^2 - 0.0333x + 1.6000, \\ p(x)^{\bar{f}} &= 0.2625x^4 - 2.9583x^3 + 11.0875x^2 - 15.5917x + 12.7000. \end{aligned}$$

根据三组数据点的相互比较可以看出 $p(x)^f \leq p(x) \leq p(x)^{\bar{f}}$, 也就是说曲线 $p(x)^f$ 在曲线 $p(x)$ 的下方, 曲线 $p(x)^{\bar{f}}$ 在曲线 $p(x)$ 的上方.

在任意的 $[0, t]$ 时刻, 不妨令 $t = 4$, 在 $[0, 4]$ 上我们对 $p(x), p(x)^f, p(x)^{\bar{f}}$ 分别积分, 经计算得出 $\int_0^4 p(x) dx = 17.9688$, $\int_0^4 p(x)^f dx = 11.3581$, $\int_0^4 p(x)^{\bar{f}} dx = 27.0285$, 显然 $\int_0^4 p(x)^f dx < \int_0^4 p(x) dx < \int_0^4 p(x)^{\bar{f}} dx$. 由于系统 m 在终端输出的知识由其本身的属性集控制, 所以规律知识的属性集起着决定性作用. 在本文中, 各知识的属性集是包含与被包含的关系, 则对应的知识也是包含与被包含的关系. 属性集里的属性越多, 对应知识的基数越小; 属性集里的属性越少, 对应知识的基数越大. 知识本身具有的特征值是不变的, 所以基数越大的知识在特征值对应点的离散数据叠加个数就越多, 生成的规律曲线会向上移动变化, 反之则向下移动变化, 因而就有了明显的积分度量关系, 特征值的取值不会对积分度量构成影响.

4 讨论

在军事, 通讯, 经济, 金融等诸多系统中, 会有大量动态, 不确定的规律, 这些规律都是以某个范围内变化的离散数据构成的离散函数. 具有属性集 α 的多个离散函数构成了函数 S-粗集中的 α -函数等价类 $[u]$, 其特性等价地表征了这些系统的运动规律, 在属性集 α 中增加或减少属性, 生成的知识又等价地表征了这些运动规律的动态特征.

参考文献:

- [1] PAWLAK Z. Rough sets [J]. International Journal of Computer and Information Sciences, 1982, 11: 341-356.
- [2] SHI Kaiquan. Function S-rough sets and function transfer [J]. International Journal Advances in Systems Sciences and Applications, 2005, 5(1): 1-8.
- [3] SHI Kaiquan. S-rough sets and its applications in diagnosis-recognition for disease [J]. IEEE Proceedings of the First International Conference on Machine Learning and Cybernetics, 2002(1): 50-54.
- [4] SHI Kaiquan, CHANG Tingcheng. One direction S-rough sets [J]. International Journal of Fuzzy Mathematics, 2005(2): 319-334.
- [5] SHI Kaiquan. Two direction S-rough sets [J]. International Journal of Fuzzy Mathematics, 2005(2): 335-349.
- [6] 史开泉, 姚炳学. 函数 S-粗集与系统规律挖掘 [M]. 北京: 科学出版社, 2007: 51-59.
- [7] 史开泉. 函数 S-粗集 [J]. 山东大学学报: 理学版, 2005, 40(1): 1-10.
- [8] CUI Yuquan, SHI Kaiquan. Function S-rough sets and applications [J]. Journal of Systems Engineering and Electronics, 2006, 17(3): 434-442.
- [9] SHI Kaiquan, YAO Bingxue. Function S-rough sets and recognition of financial risk laws [J]. The First International Conference on Rough sets and Knowledge Technology, 2006(1): 247-253.
- [10] SHI Kaiquan. Function S-rough sets and mining-discovery of rough law in system [J]. Journal of Systems Engineering and Electronics, 2006, 17(4): 919-926.
- [11] 史开泉, 崔玉泉. S-粗集与粗决策 [M]. 北京: 科学出版社, 2006: 39-165.