

层次粒结构下粗糙模糊集的不确定性度量

杨 洁^{1,2} 王国胤¹ 张清华¹ 冯 林³

(重庆邮电大学计算智能重庆市重点实验室 重庆 400065)¹

(遵义师范学院物理与电子科学学院 贵州 遵义 563002)² (四川师范大学计算机科学学院 成都 610000)³

摘 要 众所周知,经典粗糙集的不确定性来自于边界域,但是对于粗糙模糊集来说,其正域和负域中的元素存在不确定性,从而导致粗糙模糊集的不确定性不仅来自于边界域,还来自于正域和负域。另外,在粗糙模糊集中,一个模糊概念可以通过层次粒结构中不同的粗糙近似空间进行刻画,随着粒度的变化,模糊概念的不确定性的变化规律如何?对此,文中提出一种基于模糊度的不确定性度量公式,并基于均值模糊集分析了粗糙模糊集模型,得出粗糙模糊集不确定性度量的模型同样适合于度量概率粗糙集的不确定性的结论。其次,采用基于模糊度的不确定性度量方法,揭示了分层递阶的多粒度空间下粗糙模糊集不确定性的变化规律。然后,分析了 3 个域(正域、边界域和负域)的不确定性,并揭示了它们在分层递阶的多粒度空间下的变化规律。最后,通过实验验证了所提不确定性度量理论的有效性。

关键词 粗糙模糊集,不确定性度量,模糊度,层次粒结构

中图法分类号 TP311 文献标识码 A DOI 10.11896/j.issn.1002-137X.2019.01.007

Uncertainty Measure of Rough Fuzzy Sets in Hierarchical Granular Structure

YANG Jie^{1,2} WANG Guo-yin¹ ZHANG Qing-hua¹ FENG Lin³

(Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)¹

(School of Physics and Electronic, Zunyi Normal University, Zunyi, Guizhou 563002, China)²

(School of Computer Science, Sichuan Normal University, Chengdu 610000, China)³

Abstract There has been a consensus that the uncertainty of Pawlak's rough sets model is rooted in the objects contained in the boundary region of the target concept, while the uncertainty of rough fuzzy sets results from three regions, because the objects in the positive or negative regions are probably uncertain. Moreover, in rough fuzzy sets model, a fuzzy concept can be characterized by different rough approximation spaces in a hierarchical granular structure, so how will the uncertainty of a fuzzy concept change with granularity? This paper firstly proposed a fuzziness-based uncertainty measure, analyzed the rough fuzzy set model through the average fuzzy sets and drew a conclusion, that is the uncertainty measure for rough fuzzy sets is also suitable for probabilistic rough sets. Based on the fuzziness-based uncertainty measure, this paper revealed the change rules of their uncertainty of rough fuzzy sets in a hierarchical granular structure. Then, it discussed the uncertainties of the three regions (positive region, boundary region and negative region) and revealed the change rules of their uncertainty in a hierarchical granular structure. Finally, experimental results demonstrate the effectiveness of the proposed uncertainty measure theory.

Keywords Rough fuzzy sets, Uncertainty measure, Fuzziness, Hierarchical granular structure

1 引言

作为 Pawlak 粗糙集^[1-2]的扩展模型之一,粗糙模糊集^[3-5]常用于处理目标概念为模糊的情形。在粗糙模糊集模型中,通过定义上、下近似集来描述一个不确定目标概念(目标模糊集),然后利用它们来进行不确定性度量或提取模糊决策规

则。粗糙模糊集通过一对阈值来控制分类精度,并通过等价类的均值隶属度将论域划分为正域、负域和边界域。

粗糙集的不确定性度量在知识获取中扮演着非常重要的角色。Wierman^[6]首次给出了一个合理的粗糙集理论的不确定性度量及其公理的推导。梁吉业等^[7-9]从信息熵、条件熵、互信息和知识粒度的角度分析了粗糙集的不确定性,并给出

到稿日期:2018-06-10 返修日期:2018-07-04 本文受国家自然科学基金(61572091,61472056,61772096),高端人才项目(RC2016005),贵州省联合基金项目(黔科合 LH 字[2017]7075 号),贵州省教育厅青年科技人才成长项目(黔教合 KY 字[2018]318 号)资助。

杨 洁(1987-),男,博士生,副教授,主要研究方向为粗糙集、粒计算、机器学习;王国胤(1970-),男,博士,教授,CCF 杰出会员,主要研究方向为粗糙集、粒计算、认知计算、智能信息处理、数据挖掘,E-mail: wanggy@ieee.org(通信作者);张清华(1974-),男,博士,教授,主要研究方向为粗糙集、粒计算;冯 林(1972-),男,博士,教授,主要研究方向为粗糙集、粒计算。

了一种新的粗糙集的粗糙熵。在知识距离的基础上, Qian等^[10]提出了模拟人类观察问题时的多粒度思维的方法,即通过知识距离计算不同粒度知识的相似性,从而对知识进行自动分组。Hu等^[11]基于香农熵提出的不确定性度量公式研究了模糊概率空间的不确定性,以有效地处理数据中的不一致性和噪声。Zhang等^[12-13]研究了概率粗糙集模型的不确定性度量方法,并揭示了其不确定性变化规律。王国胤等^[14]从属性空间的角度讨论了不同知识粒度的近似空间下粗糙集的模糊度的变化问题,提出了一种基于信息熵的粗糙集的模糊度度量方法,从而弥补了粗糙度和粗糙熵对粗糙集不确定性度量的不足。Yao等^[15]提出了粒度度量的统一框架,并指出当前存在的大多数不确定性度量方法都属于该框架。除此之外,当前也有许多粗糙模糊集的不确定性研究工作。郭增晓等^[16]基于条件信息熵提出了一种粗糙模糊集的模糊度度量。结合粗糙度和粗糙熵, Qin等^[17]提出了一种新的粗糙熵,其可以很好地反映粗糙模糊集的不确定性。Hu等^[18]从距离的角度提出了一种粗糙模糊集的粗糙度度量,并将它应用于不完备的模糊决策信息表。Sun等^[19]基于香农熵提出了一种广义粗糙模糊集的不确定性度量方法,其可以有效地评估广义粗糙模糊集的精度和粗糙度。但是,这些工作都是在单个粒度空间上研究粗糙模糊集的不确定性度量。在多粒度空间下,随着粒度的变化,粗糙模糊集的总不确定性以及3个决策域(正域、负域、边界域)的不确定性具有什么样的变化规律仍缺乏相关研究。

粒计算^[20-23]是一种新的模拟人类思维机制的方法论,从粒计算的观点来看,知识的确定性与不确定性在一定粒度层次上可以相互转化^[24]。模糊集^[25]、粗糙集^[1]、商空间^[26]和云模型^[27]从不同的角度实现了不确定知识的表示与转化。分析粗糙模糊集3个域(正域、负域、边界域)的不确定性有助于提高分类质量。不同于经典粗糙集,粗糙模糊集通过计算各个等价类的均值隶属度将论域划分为3个域,类似于概率粗糙集,粗糙模糊集有较强的容错能力。在粗糙模糊集中,一个模糊概念可以通过分层递阶的多粒度空间下的不同粗糙近似空间进行刻画,其不确定性来自于3个域,而且边界域随着粒度的细化不具有单调递减性。在许多应用领域中,如医疗诊断^[28]、质量评估^[29]、风险预测^[30]这3个域的不确定性的分析将有助于提高最终的决策质量。针对以上问题,本文首先提出了一种基于模糊度的不确定性度量模型,然后在此基础上讨论了分层递阶的多粒度知识空间下粗糙模糊集的不确定性的变化规律,并分析了3个域的不确定性变化规律。

本文第2节简要介绍了相关的基本概念,第3节基于模糊度提出了粗糙模糊集的不确定性度量公式;第4节讨论了多粒度空间下粗糙模糊集的不确定性变化规律以及3个域的变化规律;第5节通过实验验证了本文提出的不确定性度量理论;最后总结全文。

2 相关定义

在介绍粗糙模糊集的不确定性度量之前,先回顾一些基本概念。

定义1(粗糙集)^[1-2] 设一个信息系统 $S=(U, C \cup D, V,$

$f)$, 其中 $R \subseteq C, X \subseteq U$, 那么 X 的上、下近似集定义如下:

$$\bar{R}(X) = \{x \in U \mid [x]_R \cap X \neq \emptyset\}$$

$$\underline{R}(X) = \{x \in U \mid [x]_R \subseteq X\}$$

其中, $[x]_R$ 代表由等价关系 U/R 诱导的等价类, 即 $U/R = \{[x]_R\} = \{[x]_1, [x]_2, \dots, [x]_m\}$ 。

本文中, 一个划分空间 U/R 通常也叫作一个知识空间或粒度空间。简单而言, 为了防止混淆, 我们假设 $[x]_R \stackrel{\Delta}{=} [x]$ 。如果 $\bar{R}(X) = \underline{R}(X)$, 则 X 是一个可定义集, 否则 X 是一个粗糙集。在粗糙集中, 论域 U 通常被划分为正域、负域和边界域, 分别定义如下:

$$POS_R(X) = \underline{R}(X)$$

$$NEG_R(X) = U - \bar{R}(X)$$

$$BND_R(X) = \bar{R}(X) - \underline{R}(X)$$

定义2(粗糙模糊集)^[3-4] 设一个信息系统 $S=(U, C \cup D, V, f)$, 其中 $R \subseteq C$ 和 X 是一个在 U 上的模糊集, 那么 X 的上、下近似集可以定义为一对模糊集, 其隶属度分别定义如下:

$$\mu_{\bar{R}}(x) = \sup\{\mu(y) \mid y \subseteq [x]_R\}$$

$$\mu_{\underline{R}}(x) = \inf\{\mu(y) \mid y \subseteq [x]_R\}$$

其中, 如果 $\bar{R}(X) = \underline{R}(X)$, 那么 X 是一个可定义的模糊集, 否则称 X 为一个粗糙模糊集。

定义3(阶梯模糊集)^[31] 设一个信息系统 $S=(U, C \cup D, V, f)$, 其中 C 代表一个属性集, $R \subseteq C$ 和 X 是一个在 U 上的模糊集。 $U/R = \{[x]_1, [x]_2, \dots, [x]_L\}$ 是一个划分空间, 其中, $[x]_i = \{x_{i1}, x_{i2}, \dots, x_{i_{t_i}}\} (i=1, 2, \dots, L), t_1 + t_2 + \dots + t_L = |U|$, 且 X 是一个在 U 上的模糊集。如果 $\mu_X(x_{i1}) = \mu_X(x_{i2}) = \dots = \mu_X(x_{i_{t_i}}) = c_i (0 \leq c_i \leq 1, i=1, 2, \dots, L)$, 那么 X 是一个在等价关系 U/R 上的阶梯模糊集。

定义4(均值隶属度)^[31] 设一个信息系统 $S=(U, C \cup D, V, f)$, 其中 C 代表一个属性集, $R \subseteq C$ 和 X 是一个在 U 上的模糊集。 $U/R = \{[x]_1, [x]_2, \dots, [x]_L\}$ 是一个划分空间, $\forall x \in [x]_i$, 其中 $i=1, 2, \dots, L, \bar{\mu}(x) = \bar{\mu}([x]_i) = \frac{\sum_{x \subseteq [x]_i} \mu(x)}{|[x]_i|}$,

那么称 $\bar{\mu}(x)$ 为均值隶属度。

定义4中, $\bar{\mu}(x)$ 代表 $[x]_i$ 对于模糊概念 X 的隶属度, $1 - \bar{\mu}(x)$ 代表 $[x]_i$ 对于模糊概念 X^C 的隶属度。从概率统计的角度来说, $\bar{\mu}(x)$ 可以理解任意 $x \in [x]$ 属于模糊概念 X 的概率; 同理, $1 - \bar{\mu}(x)$ 可以理解任意 $x \in [x]$ 不属于模糊概念 X^C 的概率。

定义5(均值模糊集)^[31] 设一个信息系统 $S=(U, C \cup D, V, f)$, 其中 C 代表一个属性集, $R \subseteq C$ 和 X 是一个在 U 上的模糊集。 $U/R = \{[x]_1, [x]_2, \dots, [x]_L\}$ 是一个划分空间, 其中, $[x]_i = \{x_{i1}, x_{i2}, \dots, x_{i_{t_i}}\} (i=1, 2, \dots, L), t_1 + t_2 + \dots + t_L = |U|$, 且 X 是一个在论域 U 上的模糊集。如果 $\bar{\mu}_{X^*}(x_{i1}) = \bar{\mu}_{X^*}(x_{i2}) = \dots = \bar{\mu}_{X^*}(x_{i_{t_i}}) = \bar{\mu}([x]_i) = \frac{\sum_{x \subseteq [x]_i} \mu(x)}{|[x]_i|}, i=1, 2, \dots, L$, 其中 X^* 是一个在 U 上的模糊集, 那么称 X^* 为 X 的均值模糊集。

定义 6(概率粗糙模糊集)^[19] 设一个信息系统 $S=(U, C \cup D, V, f)$, 其中, 阈值为 $\alpha, \beta(0 \leq \beta \leq \alpha \leq 1), R \subseteq C$ 和 X 是一个在 U 上的模糊集, 那么 X 的上、下近似集定义如下:

$$\bar{R}^{(\alpha, \beta)}(X) = \{x \in U \mid \bar{\mu}([x]) > \beta\}$$

$$\underline{R}^{(\alpha, \beta)}(X) = \{x \in U \mid \bar{\mu}([x]) \geq \alpha\}$$

论域 U 通常被划分为正域、负域和边界域, 分别定义如下:

$$POS_R^{(\alpha, \beta)}(X) = \{x \in U \mid \bar{\mu}([x]) \geq \alpha\} = \underline{R}^{(\alpha, \beta)}(X)$$

$$NEG_R^{(\alpha, \beta)}(X) = \{x \in U \mid \bar{\mu}([x]) \leq \beta\} = U - \bar{R}^{(\alpha, \beta)}(X)$$

$$BND_R^{(\alpha, \beta)}(X) = \{x \in U \mid \beta < \bar{\mu}([x]) < \alpha\} = \bar{R}^{(\alpha, \beta)}(X) - \underline{R}^{(\alpha, \beta)}(X)$$

由于阈值的存在, 与 Pawlak 粗糙集相比, 概率粗糙模糊集的边界域变得更小。

定义 7(模糊度)^[32] 设 $U=(x_1, x_2, \dots, x_n)$ 是一个有限论域, A 和 B 是两个在 U 上的模糊集, 如果存在映射 $H: F(U) \rightarrow [0, 1]$ 满足以下条件:

(1) 有且仅有 $A \in P(U)$ 时, $H(A) = 0$;

(2) 有且仅有 $\forall x_i \in U$ 和 $A(x_i) = \frac{1}{2}$ 时, $H(A) = 0$;

(3) $\forall x_i \in U$, 如果 $B(x_i) \leq A(x_i) \leq \frac{1}{2}$ 或 $B(x_i) \geq A(x_i) \geq \frac{1}{2}$, 那么 $H(B) \leq H(A)$;

(4) 对于每一个 $A \in F(U)$, $H(A) = H(A^c)$ 。

那么, $H(*)$ 被称为模糊子集的模糊度。

3 粗糙模糊集的不确定性度量

根据定义 7, 当前有许多研究工作提出了不同的模糊度的公式。本文选择以下模糊度公式^[30]来建立粗糙模糊集的不确定性度量模型:

$$H(A) = \frac{4}{n} \sum_{i=1}^n \mu_A(x_i) (1 - \mu_A(x_i)) \quad (1)$$

基于文献[30]中的模糊度公式, 提出如下的粗糙模糊集的不确定性度量公式:

$$H_R^{(\alpha, \beta)}(X) = \frac{4}{n} \sum_{i=1}^n \mu_R(x_i) (1 - \mu_R(x_i)) \quad (2)$$

其中, $\mu(x_i) = \bar{\mu}([x]) = \frac{\sum_{x \in [x]} \mu(x)}{|[x]|}$ 。

通过第 1 节的分析, 粗糙模糊集中目标概念的不确定性主要来自于正域、负域和边界域, 即:

$$\begin{aligned} H_R^{(\alpha, \beta)}(X) &= \frac{4}{n} \sum_{i=1}^n \mu_R(x_i) (1 - \mu_R(x_i)) \\ &= \frac{4}{n} \sum_{x \in POS_R^{(\alpha, \beta)}(X)} (\bar{\mu}(x) (1 - \bar{\mu}(x))) + \frac{4}{n} \sum_{x \in NEG_R^{(\alpha, \beta)}(X)} (\bar{\mu}(x) (1 - \bar{\mu}(x))) \\ &\quad + \frac{4}{n} \sum_{x \in BND_R^{(\alpha, \beta)}(X)} (\bar{\mu}(x) (1 - \bar{\mu}(x))) \\ &= H(POS_R^{(\alpha, \beta)}(X)) + H(NEG_R^{(\alpha, \beta)}(X)) + H(BND_R^{(\alpha, \beta)}(X)) \end{aligned}$$

显然, 已知一个信息系统 $S=(U, C \cup D, V, f)$, 对于任意的 $0 \leq \beta \leq \alpha \leq 1, R \subseteq C, X$ 是一个在 U 上的模糊集, $H_R^{(\alpha, \beta)}(X)$

具有如下性质:

$$(1) 0 \leq H_R^{(\alpha, \beta)}(X) \leq 1;$$

$$(2) H_R^{(\alpha, \beta)}(X) \text{ 不随 } \alpha \text{ 和 } \beta \text{ 的变化而变化。}$$

图 1 给出了粗糙模糊集不确定性度量的流程图, 其中虚线框代表概率粗糙集不确定性度量的流程图。由图 1 可知, 概率粗糙集是概率粗糙模糊集的特殊形式。

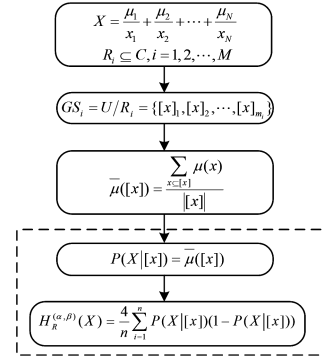


图 1 粗糙模糊集的不确定性度量流程

Fig. 1 Process of uncertainty measure for rough fuzzy sets

引理 1 设一个信息系统 $S=(U, C \cup D, V, f), R_1 \subseteq R_2 \subseteq C, X$ 是一个在 U 上的模糊集, $U/R_1 = \{p_1, p_2, \dots, p_l\}$, $U/R_2 = \{q_1, q_2, \dots, q_m\}$ 。如果仅有等价类 $p_i (i=1, 2, \dots, l)$ 通过属性集 $\Delta R = R_2 - R_1$ 细分为 $q_{i1}, q_{i2}, \dots, q_{in}$, 且 $\bar{\mu}(q_{i1}) = \bar{\mu}(q_{i2}) = \dots = \bar{\mu}(q_{in}) = a$, 则 $\bar{\mu}(p_i) = a$ 。

证明: 显而易见, $\bar{\mu}(p_1) = \frac{|q_1|}{|p_1|} \bar{\mu}(q_1) + \frac{|q_2|}{|p_1|} \bar{\mu}(q_2) + \dots +$

$\frac{|q_m|}{|p_1|} \bar{\mu}(q_m)$, 如果 $\bar{\mu}(q_1) = \bar{\mu}(q_2) = \dots = \bar{\mu}(q_m) = a$, 由于 $|p_1| =$

$|q_1| + |q_2| + \dots + |q_m|$, 那么 $\bar{\mu}(p_1) = \frac{|q_1|}{|p_1|} a + \frac{|q_2|}{|p_1|} a + \dots +$

$\frac{|q_m|}{|p_1|} a = a$ 。

由引理 1 可知, 如果一个等价类细分以后的子等价类对于目标概念的隶属度相同, 那么等价类也具有与它们相同的隶属度。通过引理 1, 我们可以得到如下定理。

定理 1 设一个信息系统 $S=(U, C \cup D, V, f), R_1 \subseteq R_2 \subseteq C, X$ 是一个在 U 上的模糊集, $U/R_1 = \{p_1, p_2, \dots, p_l\}$, $U/R_2 = \{q_1, q_2, \dots, q_m\}$ 。如果仅有等价类 $p_i (i=1, 2, \dots, l)$ 通过属性集 $\Delta R = R_2 - R_1$ 细分为 $q_{i1}, q_{i2}, \dots, q_{in}$, 且 $\bar{\mu}(q_{i1}) = \bar{\mu}(q_{i2}) = \dots = \bar{\mu}(q_{in}) = a$, 那么 $H_R^{(\alpha, \beta)}(X) = H_{R_2}^{(\alpha, \beta)}(X)$ 。

通过定理 1 可知: 如果较粗粒度中的等价类等比例细分为较细粒度上的多个子等价类, 那么其不确定性不发生变化。

4 层次粒结构下的不确定性度量

通过第 1 节的分析可知, 随着信息的增加, 3 个域(正域、负域和边界域)的不确定性可能会发生变化。从粗糙集的角度来说, 当知识粒度发生细分时, 3 个域各自的不确定性将会发生变化。下面我们将进行相关的分析和讨论。

定理 2 设一个信息系统 $S=(U, C \cup D, V, f), R_1 \subseteq C, R_2 \subseteq C, X$ 是一个在 U 上的模糊集。如果 $R_1 \subseteq R_2$, 那么 $H_{R_1}^{(\alpha, \beta)}(X)$

$(X) \geq H_{R_2}^{(\alpha, \beta)}(X)$ 。

证明: 设 $U = \{x_1, x_2, \dots, x_n\}$ 是一个非空有限论域, $U/R_1 = \{p_1, p_2, \dots, p_l\}$, $U/R_2 = \{q_1, q_2, \dots, q_m\}$ 。

由于 $R_1 \subseteq R_2$, 因此 $U/R_2 \leq U/R_1$ 。由条件可知, 为了简单起见, 假设只有一个等价类 p_1 通过属性集 $\Delta R = R_2 - R_1$ 细分为两个子等价类 q_1 和 q_2 (其他复杂的情况均可以转化为这种情况, 这里不再重复)。

假设 $p_1 = q_1 \cup q_2, p_2 = q_3, p_3 = q_4, p_l = q_m (m = l + 1)$, 换言之, $U/R_2 = \{q_1, q_2, p_2, p_3, \dots, p_m\}$ 。

$$\begin{aligned} \Delta H_{R_1 - R_2} &= H_{R_1}^{(\alpha, \beta)}(X) - H_{R_2}^{(\alpha, \beta)}(X) \\ &= \frac{4}{n} (\bar{\mu}(p_1)(1 - \bar{\mu}(p_1)) |p_1| - \bar{\mu}(q_1)(1 - \bar{\mu}(q_1)) \\ &\quad |q_1| - \bar{\mu}(q_2)(1 - \bar{\mu}(q_2)) |q_2|) \\ &= \frac{4}{n} (|q_1| \bar{\mu}(q_1)^2 + |q_2| \bar{\mu}(q_2)^2 - |p_1| \bar{\mu}(p_1)^2) \end{aligned}$$

由于 $\bar{\mu}(p_1) = \frac{|q_1|}{|p_1|} \bar{\mu}(q_1) + \frac{|q_2|}{|p_1|} \bar{\mu}(q_2)$, 则:

$$\bar{\mu}(p_1)^2 = \frac{|q_1|^2}{|p_1|^2} \bar{\mu}(q_1)^2 + 2 \frac{|q_1| |q_2|}{|p_1|^2} \bar{\mu}(q_1) \bar{\mu}(q_2) + \frac{|q_2|^2}{|p_1|^2} \bar{\mu}(q_2)^2$$

那么

$$\begin{aligned} \Delta H_{R_1 - R_2} &= \frac{4}{n} \left(\frac{|q_1| |q_2|}{|p_1|^2} (\bar{\mu}(q_1)^2 - 2 \bar{\mu}(q_1) \bar{\mu}(q_2) + \bar{\mu}(q_2)^2) \right) \\ &= \frac{4}{n} \frac{|q_1| |q_2|}{|p_1|^2} (\bar{\mu}(q_1) - \bar{\mu}(q_2))^2 \geq 0 \end{aligned}$$

因此, $H_{R_1}^{(\alpha, \beta)}(X) \geq H_{R_2}^{(\alpha, \beta)}(X)$ 。其中, 当 $\bar{\mu}(q_1) = \bar{\mu}(q_2)$ 时, $H_{R_1}^{(\alpha, \beta)}(X) = H_{R_2}^{(\alpha, \beta)}(X)$ 。

通过定理 2 可知, 随着粒度的细化, 从模糊度的角度, 粗糙模糊集的不确定性单调递减。经典粗糙集中, 随着粒度的细化, 边界域的不确定性将单调递减。不同于经典粗糙集, 在粗糙模糊集模型中, 在层次粒结构的多粒度空间下, 随着粒度的细化, 不仅是边界域的元素有可能被重新分类, 正域和负域中的元素也会被重新分类。因此, 对于粗糙模糊集来说, 其不确定性来自于 3 个域, 而且边界域的不确定性随着粒度的细化不具有单调性。下面我们将分析粗糙模糊集的边界域在分层递阶的多粒度空间下的不确定性的变化情况。

定理 3 设一个信息系统 $S = (U, C \cup D, V, f), R_1 \subseteq C, R_2 \subseteq C, X$ 是一个在 U 上的模糊集。如果 $R_1 \subseteq R_2$ 以及仅有 $NEG_{R_1}^{(\alpha, \beta)}(X)$ 中的等价类通过属性集 $\Delta R = R_2 - R_1$ 发生细分, 那么 $H(BND_{R_1}^{(\alpha, \beta)}(X)) \leq H(BND_{R_2}^{(\alpha, \beta)}(X))$ 。

证明: 为了简单起见, 这里基于定理 2 中证明过程的假设条件。由于 $\bar{\mu}(p_1) \leq \beta$, 我们将通过以下 3 种情形证明该定理。

(1) 如果 $\bar{\mu}(q_1) \leq \beta, \bar{\mu}(q_2) \leq \beta$, 即 $q_1 \subseteq NEG_{R_2}^{(\alpha, \beta)}(X), q_2 \subseteq NEG_{R_2}^{(\alpha, \beta)}(X)$, 那么 $BND_{R_1}^{(\alpha_1, \beta_1)}(X) = BND_{R_2}^{(\alpha_2, \beta_2)}(X)$ 。因为 $p_2 = q_3, p_3 = q_4, p_l = q_m (m = l + 1)$, 所以 $H(BND_{R_1}^{(\alpha, \beta)}(X)) = H(BND_{R_2}^{(\alpha, \beta)}(X))$ 。

(2) 如果 $\bar{\mu}(q_1) \geq \alpha, \bar{\mu}(q_2) \leq \beta$, 即 $q_1 \subseteq POS_{R_2}^{(\alpha, \beta)}(X), q_2 \subseteq NEG_{R_2}^{(\alpha, \beta)}(X)$, 那么 $BND_{R_1}^{(\alpha_1, \beta_1)}(X) = BND_{R_2}^{(\alpha_2, \beta_2)}(X)$ 。因为 $p_2 = q_3, p_3 = q_4, p_l = q_m (m = l + 1)$, 所以 $H(BND_{R_1}^{(\alpha, \beta)}(X)) = H(BND_{R_2}^{(\alpha, \beta)}(X))$ 。

(3) 如果 $\beta < \bar{\mu}(q_1) < \alpha, \bar{\mu}(q_2) \leq \beta$, 即 $q_1 \subseteq BND_{R_2}^{(\alpha, \beta)}(X), q_2 \subseteq NEG_{R_2}^{(\alpha, \beta)}(X)$, 那么 $BND_{R_1}^{(\alpha_1, \beta_1)}(X) = BND_{R_1}^{(\alpha_2, \beta_2)}(X) \cup q_1$, $H(BND_{R_2}^{(\alpha, \beta)}(X)) = H(BND_{R_1}^{(\alpha, \beta)}(X)) + H(q_1) > H(BND_{R_2}^{(\alpha, \beta)}(X))$ 。因此, $H(BND_{R_1}^{(\alpha, \beta)}(X)) < H(BND_{R_2}^{(\alpha, \beta)}(X))$ 。

定理 4 设一个信息系统 $S = (U, C \cup D, V, f), R_1 \subseteq C, R_2 \subseteq C, X$ 是一个在 U 上的模糊集。如果 $R_1 \subseteq R_2$ 以及仅有 $POS_{R_1}^{(\alpha, \beta)}(X)$ 中的等价类通过属性集 $\Delta R = R_2 - R_1$ 发生细分, 那么 $H(BND_{R_1}^{(\alpha, \beta)}(X)) \leq H(BND_{R_2}^{(\alpha, \beta)}(X))$ 。

证明: 与定理 3 的证明过程相似, 这里不再重复。

由定理 3 和定理 4 可知, 当只有正域或负域的等价类随着粒度的细化发生细分时, 边界域的不确定性将随着粒度的降低而逐渐增大, 这在一定程度上与人类的认知并不相符。

定理 5 设一个信息系统 $S = (U, C \cup D, V, f), R_1 \subseteq C, R_2 \subseteq C, X$ 是一个在 U 上的模糊集。如果 $R_1 \subseteq R_2$ 以及仅有 $BND_{R_1}^{(\alpha, \beta)}(X)$ 中的等价类通过属性集 $\Delta R = R_2 - R_1$ 发生细分, 那么 $H(BND_{R_1}^{(\alpha, \beta)}(X)) \geq H(BND_{R_2}^{(\alpha, \beta)}(X))$ 。

证明: 为了简单起见, 这里基于定理 2 中证明过程的假设条件。由于 $\beta < \bar{\mu}(p_1) < \alpha$, 我们将通过以下 4 种情形来证明该定理。

(1) 如果 $\beta < \bar{\mu}(q_1) < \alpha, \beta < \bar{\mu}(q_2) < \alpha$, 即 $q_1 \subseteq BND_{R_2}^{(\alpha, \beta)}(X), q_2 \subseteq BND_{R_2}^{(\alpha, \beta)}(X)$, 那么 $BND_{R_1}^{(\alpha_1, \beta_1)}(X) = BND_{R_2}^{(\alpha_2, \beta_2)}(X)$ 。因为 $p_2 = q_3, p_3 = q_4, p_l = q_m (m = l + 1)$, 所以 $H(BND_{R_1}^{(\alpha, \beta)}(X)) = H(BND_{R_2}^{(\alpha, \beta)}(X))$ 。

(2) 如果 $\bar{\mu}(q_1) \geq \alpha, \bar{\mu}(q_2) \leq \beta$, 即 $q_1 \subseteq POS_{R_2}^{(\alpha, \beta)}(X), q_2 \subseteq NEG_{R_2}^{(\alpha, \beta)}(X)$, 那么 $BND_{R_1}^{(\alpha_1, \beta_1)}(X) > BND_{R_2}^{(\alpha_2, \beta_2)}(X)$ 。因为 $p_2 = q_3, p_3 = q_4, p_l = q_m (m = l + 1)$, 所以 $H(BND_{R_1}^{(\alpha, \beta)}(X)) > H(BND_{R_2}^{(\alpha, \beta)}(X))$ 。

(3) 如果 $\beta < \bar{\mu}(q_1) < \alpha, \bar{\mu}(q_2) \geq \alpha$, 即 $q_1 \subseteq BND_{R_2}^{(\alpha, \beta)}(X), q_2 \subseteq POS_{R_2}^{(\alpha, \beta)}(X)$, 那么 $BND_{R_1}^{(\alpha_1, \beta_1)}(X) > BND_{R_2}^{(\alpha_2, \beta_2)}(X)$ 。因为 $p_2 = q_3, p_3 = q_4, p_l = q_m (m = l + 1)$, 所以 $H(BND_{R_1}^{(\alpha, \beta)}(X)) > H(BND_{R_2}^{(\alpha, \beta)}(X))$ 。

(4) 如果 $\beta < \bar{\mu}(q_1) < \alpha, \bar{\mu}(q_2) \leq \beta$, 即 $q_1 \subseteq BND_{R_2}^{(\alpha, \beta)}(X), q_2 \subseteq NEG_{R_2}^{(\alpha, \beta)}(X)$, 那么 $BND_{R_1}^{(\alpha_1, \beta_1)}(X) > BND_{R_2}^{(\alpha_2, \beta_2)}(X)$ 。因为 $p_2 = q_3, p_3 = q_4, p_l = q_m (m = l + 1)$, 所以 $H(BND_{R_1}^{(\alpha, \beta)}(X)) > H(BND_{R_2}^{(\alpha, \beta)}(X))$ 。

由定理 5 可知, 当只有边界域的等价类随着粒度的细化而发生细分时, 边界域的不确定性才会随着粒度的降低逐渐降低, 这在一定程度上与人类的认知相符。

定理 6 设一个信息系统 $S = (U, C \cup D, V, f), R_1 \subseteq C, X$ 是一个在 U 上的模糊集。如果 $0 \leq \beta_1 \leq \beta_2 \leq \alpha_2 \leq \alpha_1 \leq 1$, 那么以下性质成立:

$$(1) H(BND_{R_1}^{(\alpha_1, \beta_1)}(X)) \geq H(BND_{R_1}^{(\alpha_2, \beta_2)}(X));$$

$$(2) H(POS_{R_1}^{(\alpha_2, \beta_1)}(X)) \leq H(POS_{R_1}^{(\alpha_1, \beta_1)}(X));$$

$$(3) H(NEG_{R_1}^{(\alpha_1, \beta_2)}(X)) \geq H(NEG_{R_1}^{(\alpha_1, \beta_1)}(X))。$$

证明: 在概率粗糙模糊集中, 当 α 增大时, 正域中的等价类可能会被分配到边界域或负域, 导致正域减小, 即 $POS_{R_1}^{(\alpha_2, \beta_1)}(X) \leq POS_{R_1}^{(\alpha_1, \beta_1)}(X)$, 则 $H(POS_{R_1}^{(\alpha_2, \beta_1)}(X)) \leq H(POS_{R_1}^{(\alpha_1, \beta_1)}(X))$; 同理, 当 β 减小时, 负域中的等价类可能会

被分配到边界域或正域,导致负域减小,即 $NEG_{R_1}^{(\alpha_1, \beta_2)}(X) \geq NEG_{R_1}^{(\alpha_1, \beta_1)}(X)$, 则 $H(NEG_{R_1}^{(\alpha_1, \beta_2)}(X)) \geq H(NEG_{R_1}^{(\alpha_1, \beta_1)}(X))$; 当 α 增大且 β 减小时,边界域面积减小,即 $BND_{R_1}^{(\alpha_1, \beta_1)}(X) \geq BND_{R_1}^{(\alpha_2, \beta_2)}(X)$, 则 $H(BND_{R_1}^{(\alpha_1, \beta_1)}(X)) \geq H(BND_{R_1}^{(\alpha_2, \beta_2)}(X))$ 。

通过定理 6 可知,粗糙模糊集中正域的不确定性随着参数 α 的减小而增大,负域的不确定性随着参数 β 的减小而减小,边界域的不确定性随着参数 α 的增大以及参数 β 的减小而单调增大。

5 实验仿真分析

本实验的硬件配置为 Intel i5-2430M 的 CPU, 8 GB 内存, 操作系统为 Windows7 64bit OS 的台式机, 采用 MATLAB2014

软件进行仿真。表 1 列出了本文所用的 UCI 数据集^[33]。

表 1 数据集描述
Table 1 Description of data sets

Dataset	Attribute Characteristics	Instances	Condition Attributes
Air Quality	Real	9358	12
Concrete	Real	1030	8
Breast	Integer	699	9
ENB2012	Real	768	8

图 2 给出了层次粒结构中总不确定性的变化趋势,其中,横坐标为粒度(属性集合形成的划分),纵坐标为总不确定性。从图中可以看出,随着粒度的细化,不确定性单调递减。

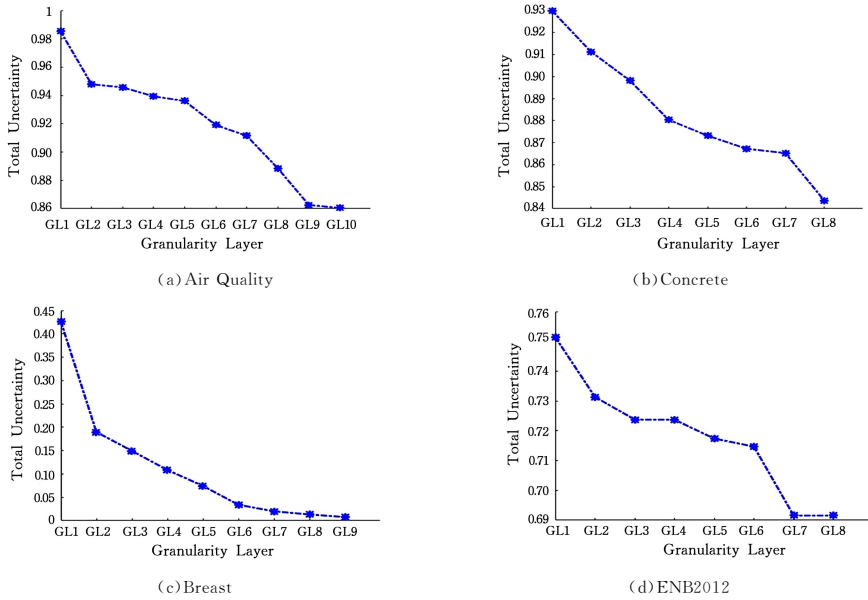


图 2 层次粒结构中不确定性的变化趋势图

Fig. 2 Trend diagram of uncertainty in hierarchical granular structure

图 3 给出了不同阈值时 3 个域的不确定性,其中横坐标代表阈值,表 2 列出了各个横坐标所表示的阈值。

表 2 阈值对信息表

Table 2 Information table of threshold pair

x-coordinate	α	β
1	0.7	0.3
2	0.7	0.4
3	0.6	0.4
4	0.6	0.3

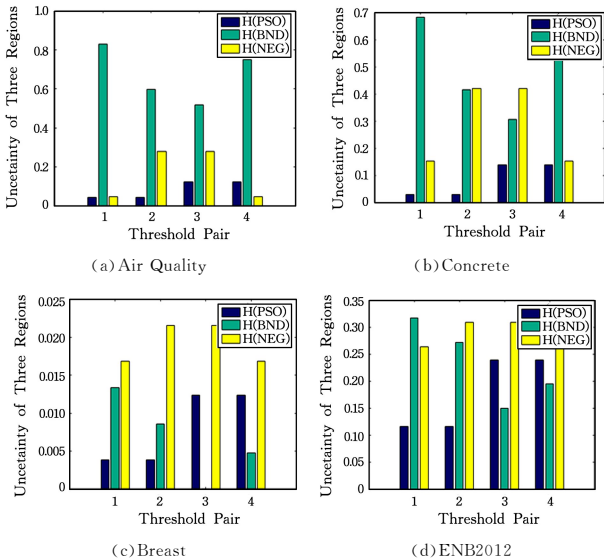


图 3 不同阈值下 3 个域的不确定性

Fig. 3 Uncertainty of three regions under different thresholds

我们可以发现,正域的不确定性随着参数 α 的减小而增大,负域的不确定性随着参数 β 的减小而减小,边界域的不确定性随着参数 α 的增大以及参数 β 的减小而单调增加。

结束语 经典粗糙集的不确定性度量方法往往不适用于粗糙模糊集的不确定性度量,本文首先基于模糊度提出了一种粗糙模糊集的不确定性度量。从模糊度的角度分析了粗糙模糊集的不确定性在多粒度空间中的变化规律,即随着粒度的细化,不确定性逐渐减小。在此基础上,讨论了粗糙模糊集边界域的不确定性变化规律,发现在仅有正域或负域中的等价类随着粒度的细化发生细分时,边界域的不确定性随着粒度的降低反而逐渐增大,这在一定程度上与人类的认知不相符。仅有边界域中的等价类随着粒度的细化发生细分时,边界域的不确定性随着粒度的细化逐渐降低,这在一定程度上

与人类的认知相符。另外,我们还讨论了3个域的不确定性随阈值 α 和 β 的变化规律。这些工作对于粗糙集的不确定性理论具有很好的促进作用。

参考文献

- [1] PAWLAK Z. Rough sets[J]. International Journal of Computer Information Sciences, 1982, 11(5): 341-356.
- [2] PAWLAK Z. Rough classification [J]. International Journal of Man-Machine Studies, 1984, 20(5): 469-483.
- [3] DUBOIS D, PRADE H. Rough fuzzy sets and fuzzy rough sets [J]. International Journal of General System, 1990, 17(2-3): 191-209.
- [4] DUBOIS D, PRADE H. Putting rough sets and fuzzy sets together[M] // Slowin' ski R, Ed. Intelligent decision support: Handbook of applications and advances of the rough sets theory. Dordrecht: Kluwer Academic Publishers, 1992: 203-232.
- [5] DUBOIS D, PRADE H. Twofold fuzzy sets and rough sets—Some issues in knowledge representation[J]. Fuzzy Sets & Systems, 1987, 23(1): 3-18.
- [6] WIERMAN M J. Measureing uncertainty in rough set theory [J]. International Journal of General Systems, 1999, 28(4-5): 283-297.
- [7] LIANG J Y, CHIN K S, DANG C Y. A new method for measuring uncertainty and fuzziness in rough set theory[J]. International Journal of General Systems, 2002, 31(4): 331-342.
- [8] 梁吉业, 李德玉. 信息系统中的不确定性与知识获取[M]. 北京: 科学出版社, 2005.
- [9] LIANG J Y, SHI Z Z. The information entropy, rough entropy and knowledge granulation in rough set theory[J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2004, 12(1): 37-46.
- [10] QIAN Y H, CHENG H, WANG J, et al. Grouping granular structures in human granulation intelligence [J]. Information Sciences, 2017, 382-383: 150-169.
- [11] HU Q H, YU D, XIE Z, et al. Fuzzy probabilistic approximation spaces and their information measures[J]. IEEE Transactions on Fuzzy Systems, 2006, 14(2): 191-201.
- [12] ZHANG Q H, ZHANG Q, WANG G Y. The uncertainty of probabilistic rough sets in multi-granulation spaces[J]. International Journal of Approximate Reasoning, 2016, 77: 38-54.
- [13] ZHANG Q H, YANG S, WANG G Y. Measuring Uncertainty of Probabilistic Rough Set Model From Its Three Regions[J]. IEEE Transactions on Systems Man & Cybernetics Systems, 2016, PP(99): 1-11.
- [14] WANG G Y, ZHANG Q H. Research on the uncertainty of rough sets under different knowledge granularity[J]. Chinese Journal of Computers, 2008, 31(9): 1588-1598. (in Chinese)
王国胤, 张清华. 不同知识粒度下粗糙集的不确定性研究[J]. 计算机学报, 2008, 31(9): 1588-1598.
- [15] YAO Y Y, ZHAO L. A measurement theory view on the granularity of partitions [J]. Information Sciences, 2012, 213(23): 1-13.
- [16] GUO Z X, MI J S. An Uncertainty Measure in Rough Fuzzy Sets [J]. Fuzzy Systems & Mathematics, 2005, 19(4): 135-140. (in Chinese)
郭增晓, 米据生. 粗糙模糊集的模糊性度量[J]. 模糊系统与数学, 2005, 19(4): 135-140.
- [17] QIN H N, LUO D R. New uncertainty measure of rough fuzzy sets and entropy weight method for fuzzy-target decision-making tables[J]. Journal of Applied Mathematics, 2014, 42(6): 1-7.
- [18] HU J, PEDRYCZ W, WANG G Y. A roughness measure of fuzzy sets from the perspective of distance [J]. International Journal of General Systems, 2016(3): 1-16.
- [19] SUN B Z, MA W M. Uncertainty measure for general relation-based rough fuzzy set[J]. Kybernetes, 2013, 42(6): 979-992.
- [20] PEDRYCZ W, SKOWRON A, KREINOVICH V. Handbook of Granular Computing [M]. Wiley-Interscience, 2008: 719-740.
- [21] PEDRYCZ W, ALHMOUZ R, MORFEQ A, et al. The design of free structure granular mappings; the use of the principle of justifiable granularity [J]. IEEE Transactions on Cybernetics, 2013, 43(6): 2105-2113.
- [22] WANG G Y, YANG J, XU J. Granular computing: from granularity optimization to multi-granularity joint problem solving [J]. Granular Computing, 2017, 2(3): 1-16.
- [23] YAO Y Y. Perspectives of granular computing[C]// IEEE International Conference on Granular Computing. IEEE, 2005: 85-90.
- [24] 李德毅, 刘常昱, 杜鹤, 等. 不确定性人工智能[M]. 北京: 国防工业出版社, 2005.
- [25] ZADEH L A. Fuzzy sets [J]. Information & Control, 1965, 8(3): 338-353.
- [26] ZHANG B, ZHANG L. Theory and Applications of Problem Solving[M]. Elsevier Science Inc. , 1992.
- [27] LI D Y, MENG H J, SHI X M. Membership cloud and affiliated cloud generator[J]. Journal of Computer Research and Development, 1995, 32(6): 15-20. (in Chinese)
李德毅, 孟海军, 史雪梅. 隶属云和隶属云发生器[J]. 计算机研究与发展, 1995, 32(6): 15-20.
- [28] STRASZECKA E. Combining uncertainty and imprecision in models of medical diagnosis [J]. Information Sciences, 2006, 176(20): 3026-3059.
- [29] DAI J, WANG W, XU Q. An Uncertainty Measure for Incomplete Decision Tables and Its Applications[J]. IEEE Transactions on Cybernetics, 2013, 43(4): 1277-1289.
- [30] YAGER R R. Decision making under measure-based granular uncertainty[J]. Granular Computing, 2018, 3(4): 345-353.
- [31] ZHANG Q H, WANG J, WANG G Y. Approximate representation of rough fuzzy sets [J]. Chinese Journal of Computers, 2015, 38(7): 1484-1496. (in Chinese)
张清华, 王进, 王国胤. 粗糙模糊集的近似表示[J]. 计算机学报, 2015, 38(7): 1484-1496.
- [32] CHAKRABARTY K, BISWAS R, NANDA S. Fuzziness in rough sets[J]. Fuzzy Sets & Systems, 2000, 110(2): 247-251.
- [33] UCIrvine machine learning repository[EB/OL]. <http://archive.ics.uci.edu/ml>