

DOI: 10.11992/tis.201904037

基于知识距离的粗糙粒结构的评价模型

杨洁^{1,2}, 王国胤^{1,2,3}, 张清华^{1,2,3}

(1. 重庆邮电大学 计算智能重庆市重点实验室, 重庆 400065; 2. 重庆邮电大学 计算机科学与技术学院, 重庆 400065; 3. 重庆邮电大学 人工智能学院, 重庆 400065)

摘要: 在粒计算理论中, 通过不同的粒计算机制可以生成不同的粒结构。在粗糙集中, 对于同一个信息表而言, 通过不同的属性添加顺序可以得到由不同的序贯层次结构, 即粗糙粒结构。在粗糙粒结构中, 不同的属性获取顺序导致了对不确定性问题求解的不同程度。因此, 如何有效评价粗糙粒结构是一个值得研究的问题。本文将从知识距离的角度研究这个问题。首先, 在前期工作所提出的知识距离框架上提出了一种粗糙近似空间距离, 用于度量粗糙近似空间之间差异性。基于提出的知识距离, 研究了粗糙粒结构的结构特征。在粗糙粒结构中, 在对不确定性问题进行求解时, 本文希望在约束条件下可以利用尽可能少的知识空间使不确定性降低达到最大化。基于这个思想并利用以上得出的结论, 在属性代价约束条件下, 引入了一个评价参数 λ , 并在此基础建立了一种粗糙粒结构的评价模型, 该方法实现了在属性代价约束条件下选择粗糙粒结构的功能。最后, 通过实例验证了本文提出的模型的有效性。

关键词: 粗糙粒结构; 知识距离; 不确定性度量; 评价模型; 粒计算; 粗糙集; 约束条件; 不确定性度量

中图分类号: TP311 **文献标志码:** A **文章编号:** 1673-4785(2020)01-0166-09

中文引用格式: 杨洁, 王国胤, 张清华. 基于知识距离的粗糙粒结构的评价模型 [J]. 智能系统学报, 2020, 15(1): 166-174.

英文引用格式: YANG Jie, WANG Guoyin, ZHANG Qinghua. Evaluation model of rough granular structure based on knowledge distance[J]. CAAI transactions on intelligent systems, 2020, 15(1): 166-174.

Evaluation model of rough granular structure based on knowledge distance

YANG Jie^{1,2}, WANG Guoyin^{1,2,3}, ZHANG Qinghua^{1,2,3}

(1. Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; 2. School of computer science and technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; 3. College of Artificial Intelligence, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: In the theory of granular computing (GrC), different granular structures are generated by various grain calculation mechanisms. In rough sets, for the same information table, different attribute adding sequence produces different sequential hierarchical structure, namely the rough granular structure. In rough granular structure, various order of attribute acquisition leads to different effects of solving uncertain problems. This leads to an interesting research topic: how to effectively evaluate the rough granular structures. This problem is solved from the perspective of knowledge distance in the paper. Firstly, the knowledge distance mentioned in our previous works is introduced and then a rough approximation space distance (RASD) is proposed to measure the difference between rough approximate space. On the basis of the knowledge distance mentioned above, the characteristic of rough granular structure (RGS) is investigated. In the rough granular structure, when solving uncertain problem, we expect to maximize the uncertainty reduction as much as possible by using smaller knowledge space. Then, based on this idea and the above conclusions, an evaluation parameter λ is introduced under the constraint of attribute cost, and further, an evaluation model of rough granular structure is established. This achieves a way to select the rough granular structure according to the constraint. Finally, the effectiveness of this method is verified by an example.

Keywords: rough granular structure; knowledge distance; uncertainty measure; evaluation model; granular computing; rough sets; constraint condition; uncertainty measure

收稿日期: 2019-04-16.

基金项目: 国家自然科学基金项目 (61572091); 贵州省教育厅科技人才成长项目 (KY(2018)No.318).

通信作者: 杨洁. E-mail: 530966074@qq.com.

早在 1997 年, Zadeh 教授^[1]就提出了粒计算是模糊信息粒化、粗糙集理论和区间计算的超

集, 是粒数学的子集。粒计算^[2-5]是一种新的模拟人类思维机制的方法论, 在粒计算的“大伞”之下, 包含了很多具体的模型, 如: 模糊集^[6]、粗糙集^[7]、商空间^[8]、云模型^[9]等。粒计算是研究基于多层次粒结构的思维方式、问题求解方法、信息处理模式的理论^[10]。

基于粒计算理论, Yao^[12]提出了序贯三支决策理论 (sequential three-way decisions, S3WD)。从多粒度的角度来说^[13-14], 随着信息系统中属性或信息的增加, S3WD 中等价类将逐渐变小。S3WD 是一种利用粗糙近似空间的渐近式问题求解模型, 本质上是粗糙粒结构, 即通过由粗到细的切换粒度, 实现问题的逐步求解。为了实现渐近式的最优尺度的选择和属性约简, 文献 [15] 提出了一种局部规则提取方法, 本质上运用了序贯三支决策的思想。在现实的决策分析中, S3WD 为求解复杂问题提供了一种模拟人类的多粒度思维。从层次商空间结构 (hierarchical quotient space structure, HQSS)^[16-18]的角度来说, 等价类将会随着粒度的细化逐渐细分成更细的等价类, 这意味着在 S3WD 模型中, 随着属性或信息的增加, 一个目标概念可以通过不同粒度的知识进行描述。再者, 由于较细粒度的知识空间上包含了更多的信息, 所以在较细粒度的知识空间上可以更准确地刻画目标概念。但是, 构建较细粒度的知识空间需要更多的代价以及在较细粒度上进行问题求解需要更多的时间消耗。

通过以上分析, 在 S3WD 模型中, 本文希望在约束条件下可以利用尽可能少的知识空间最大化降低问题的不确定性, 从而对问题进行求解, 并且针对不同的约束条件评价并选择不同的粗糙粒结构进行问题求解也是需要解决的问题。

基于这个思想, 本文从知识距离的角度研究粗糙粒结构的评价模型 (为了简单起见, 本文仅考虑属性代价为约束条件下粗糙粒结构进行评价选择)。首先, 基于前期工作中提出的知识距离框架进一步提出了一种粗糙近似空间距离并讨论了其相关性质。其次, 基于提出的粗糙近似空间距离研究了层次粒结构对目标概念的刻画能力, 并得到同一个粗糙粒结构中的两个知识空间的不确定性差异等于它们之间的知识距离的结论, 并通过实例表验证了本文的评价模型的有效性。此外, 本文的评价模型具有可扩展性, 针对不同的粒结构设计合适的知识距离, 可用于评价不同类型的粒结构, 例如层次聚类、多尺度图像分割以及多粒度社区发现等问题。

1 基本概念

一个信息系统是一个四元组, 可表示为 $S = (U, C \cup D, V, f)$, 其中 U 为一个非空论域, C 为条件属性集合, D 为决策属性集, V 为对每个对象对应的属性值集合, $f: U \times C$ 是一个信息函数。

定义 1 粗糙集^[7]。设一个信息系统 $S = (U, C \cup D, V, f)$, 其中 $R \subseteq C$ 和 $X \subseteq U$, 那么 X 的上、下近似集定义如下:

$$\begin{aligned}\underline{R}(X) &= \{x \in U | [x]_R \subseteq X\} \\ \overline{R}(X) &= \{x \in U | [x]_R \cap X \neq \emptyset\}\end{aligned}$$

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

本章中, 仅从划分的角度, 一个知识空间 U/R 通常叫做一个商空间, 当利用 U/R 对目标概念 X 进行粗糙近似描述时, 通常将 U/R 表示为 $U/R(X)$, 称为粗糙近似空间。简单而言, 为了防止混淆本文假设 $[x]_R \triangleq [x]$ 。如果 $\overline{R}(X) = \underline{R}(X)$, 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:

$$\begin{aligned}\text{POS}_R(X) &= \underline{R}(X) \\ \text{BND}_R(X) &= \overline{R}(X) - \underline{R}(X) \\ \text{NEG}_R(X) &= U - \overline{R}(X)\end{aligned}$$

定义 2 粒度度量^[19-21]。假设 U 为一个有限非空论域, 一个函数 $2^U \rightarrow \mathbb{R}$ 对于任意 $P, Q \in 2^U$, 如果满足以下条件:

- 1) $m(x) \geq 0$;
- 2) $P \subset Q \Rightarrow m(P) < m(Q)$;
- 3) $P \equiv_s Q \Rightarrow m(P) = m(Q)$;

则它是一个粒度度量。

定义 3 距离度量^[22]。假设 U 为一个有限非空论域, A, B, C 是 3 个有限集合, $\forall A, B, C \subseteq U$ 。如果满足以下 3 个条件:

- 1) 正定性: $V(A, B) \geq 0$;
- 2) 对称性: $V(A, B) = V(B, A)$;
- 3) 三角不等式: $V(A, B) + V(B, C) \geq V(A, C)$;

则函数 $V(\cdot, \cdot)$ 是一个真正的距离度量。

定义 4^[1, 2, 3] 假设 $P = \{(p_1, w_{p_1}), (p_2, w_{p_2}), \dots, (p_l, w_{p_l})\}$ 和 $Q = \{(q_1, w_{q_1}), (q_2, w_{q_2}), \dots, (q_m, w_{q_m})\}$ 分别是两个具有 l 和 m 个类簇的分布。 p_i 和 q_j 代表 P 和 Q 类簇, w_{p_i} 和 w_{q_j} 分别 p_i 和 q_j 的权重。 c_{ij} 代表 p_i 和 q_j 之间的差异性度量, 例如, 欧式距离。EMD 的目标是寻找一个流量矩阵 $F = [f_{ij}]$ 最小化总代价, 其中 f_{ij} 代表 p_i 到 q_j 的流量。

$$\text{work}(P, Q) = \sum_{i=1}^l \sum_{j=1}^m c_{ij} f_{ij}$$

$$\text{s.t.} \quad f_{ij} \geq 0; 1 \leq i \leq n, 1 \leq j \leq m \quad (1)$$

$$\sum_{j=1}^m f_{ij} \leq w_i; 1 \leq i \leq n \quad (2)$$

$$\sum_{i=1}^l f_{ij} \leq w_j; 1 \leq j \leq m \quad (3)$$

$$\sum_{i=1}^l \sum_{j=1}^m f_{ij} = \min \left\{ \sum_{i=1}^l w_i, \sum_{j=1}^m w_j \right\} \quad (4)$$

式 (1) 限制流量传递过程为 P 到 Q , 不能反向; 式 (2) 限制从 P_i 运出去的总流量不能超过自身的总流量; 式 (3) 限制 Q_j 接收的总流量不能超过自身的总流量; 式 (4) 限制总传递流量的上限是 P 和 Q 的流量的最小值。当这个运输问题解决时, 这个最优的流量分配即可求得, 即 EMD 被定义为由总流量归一化的结果:

$$\text{EMD}(P, Q) = \frac{\sum_{i=1}^l \sum_{j=1}^m c_{ij} f_{ij}}{\sum_{i=1}^l \sum_{j=1}^m f_{ij}} \quad (5)$$

一般来说, 式 (5) 中的 $c(\cdot, \cdot)$ 可以是任意差异性度量, 可根据具体问题来选取。

定理 1^[23-24] 如果 $c(\cdot, \cdot)$ 是一个真正的度量, 且两个分布的总权重相同, 即 $\sum_{i=1}^n w_i = \sum_{j=1}^m w_j$, 那么 EMD 为一个真正的距离度量。

在粒计算中, 信息粒化旨在建立基于外部世界的有效的并以用户为中心的概念, 同时简化物理世界和虚拟世界的认识, 可以高效地提供“实用”的非精确解。信息粒化首先需要研究信息粒、粒层和整个粒结构的表示, 然后针对表示方法进行构建^[11]。同一粒化准则对应一个知识空间, 不同粒化准则对应多个知识空间, 不同知识空间 (相邻两知识空间或跨知识空间) 中信息粒之间的相互关系构成一个结构, 由此得到的多个知识空间及知识空间之间的关系称为粒的拓扑空间, 简称为粒结构, 如图 1 所示, 为一个粒结构的整体示意图。

在 S3WD 模型中, 随着粒度的细分, 一方面, 如果知识空间之间的不确定性差异太小, 意味着新增加的属性对于不确定性的降低几乎没有贡献, 因此原先在边界域的样本有可能仍然在边界域, 无法对它们进行决策; 另一方面, 如果知识空间之间的不确定性差异太大, 意味着获取新增加的属性需要很大的代价, 因此有可能无法满足用户的时限约束条件。如图 2 所示, 图 2(a) 和图 2(b)

分别表示了上面分析的两种情况, 图 2(c) 表示两个知识空间之间不确定性差异适中的情况。

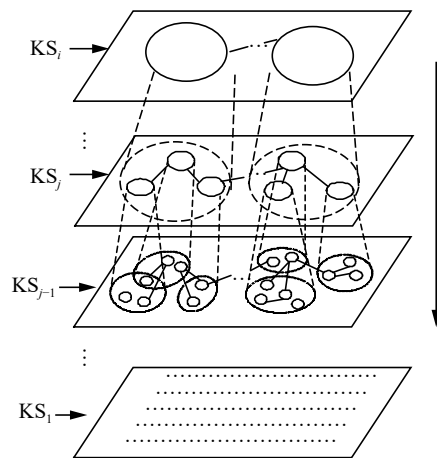


图 1 粒结构示意图

Fig. 1 Schematic diagram of granular structure

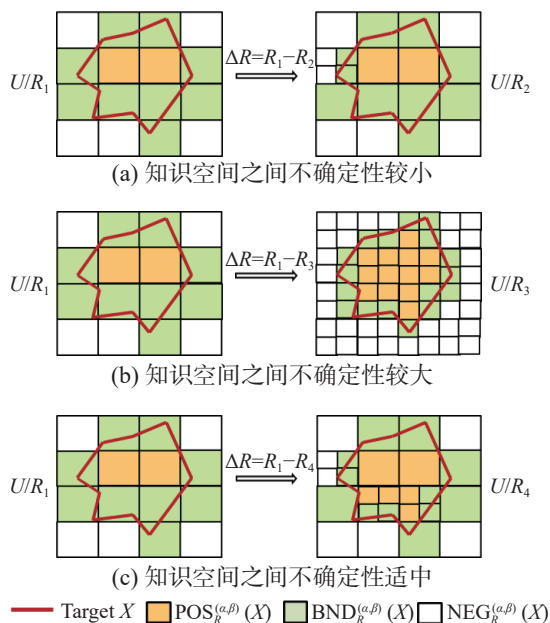


图 2 3 种不同程度的知识空间细分情况

Fig. 2 Three different levels of knowledge spaces subdivision

2 知识距离

从粒计算的角度来说, 无论是信息熵还是知识粒度都无法刻画粒结构中知识空间之间的差异性。在粗糙集模型中, 不确定性度量用于度量粗糙近似空间刻画目标概念的不确定性程度。虽然当前存在许多不确定性度量方法^[25-29], 但是这些方法仍然不够准确, 即具有相同不确定性的两个知识空间不一定等价, 它们的差异性如何刻画? 为了解决这个问题, 钱宇华^[30-32]首次提出了基于邻域信息粒的知识距离, 并建立了同一知识基中完备粒空间与不完备粒空间之间的关系。杨习

贝^[33-34]将知识距离应用到多粒度空间, 通过知识距离代数格导出一个偏序关系, 可用于描述多粒度空间中的层次结构。但是, 当前的知识距离模型缺乏清晰的物理解释和理论背景, 并且不具有扩展性。在前期工作中^[35-36], 本文基于 EMD 提出了一种知识距离框架 (knowledge distance framework, KDF) 并给出了知识距离的物理意义, 在此基础上进一步构建了一种商空间知识距离 (quotient space distance, QSD)。

定义 5 知识距离框架^[35]。假设 $K = (U, \mathfrak{K})$ 是一个知识基, $K_P = \{p_1, p_2, \dots, p_l\}$ 和 $K_Q = \{q_1, q_2, \dots, q_m\}$ 是两个分别由等价关系 P 和 Q 产生的知识空间, $d_{ij} = d(p_i, q_j)$ 代表信息粒 p_i 和 q_j 之间的距离, f_{ij} 代表 p_i 和 q_j 之间的流量, 其中 $f_{ij} = |p_i \cap q_j|$ 。 K_P 和 K_Q 之间的知识距离框架有如下定义:

$$\text{KDF}(K_P, K_Q) = \frac{1}{|U|} \sum_{i=1}^l \sum_{j=1}^m d_{ij} f_{ij} \quad (6)$$

由于对同一论域进行划分的不同知识空间的对象数是相同的, 故 $\sum_{i=1}^l |p_i| = \sum_{j=1}^m |q_j| = |U|$, 式 (6) 符合约束条件式 (4)。同样, 式 (7) 符合约束条件式 (1)~(3)。KDF 表示的是同一个知识基的任意两个知识空间之间匹配对象的最小成本。显然, 通过采用不同的信息粒距离, 可以产生一组基于划分的知识距离。基于 EMD 的优点, KDF 可以实现一个多对多的匹配, 更符合人类的认知, 可以更准确地刻画两个知识空间的差异性。因此, KDF 是一个有效直观的描述知识空间之间差异性的数学工具, 可扩展到度量任何形式的知识空间之间的差异性 (或相似性)。

定义 6 商空间知识距离^[35]。设一个信息系统 $S = (U, C \cup D, V, f)$, $R_1, R_2 \subseteq C$, $U/R_1 = \{p_1, p_2, \dots, p_l\}$ 和 $U/R_2 = \{q_1, q_2, \dots, q_m\}$ 是 U 上的两个知识空间, $d_{ij} = d(p_i, q_j)$ 代表信息粒 p_i 和 q_j 之间的距离, f_{ij} 代表 p_i 和 q_j 之间的流量, 其中 $f_{ij} = |p_i \cap q_j|$ 。 $U/R_1 = \{p_1, p_2, \dots, p_l\}$ 和 $U/R_2 = \{q_1, q_2, \dots, q_m\}$ 之间的知识距离框架有如下定义:

$$\text{QSD}(U/R_1, U/R_2) = \frac{1}{|U|} \sum_{i=1}^l \sum_{j=1}^m \frac{|p_i \oplus q_j|}{|U|} |p_i \cap q_j|$$

式中 $|p_i \oplus q_j| = |p_i \cup q_j| - |p_i \cap q_j|$ 。

例 1 设一个信息系统 $S = (U, C \cup D, V, f)$, $U = \{x_1, x_2, \dots, x_9\}$, $R_1, R_2 \subseteq C$, $X = \frac{0}{x_1} + \frac{0}{x_2} + \frac{0}{x_3} + \frac{1}{x_4} + \frac{1}{x_5} + \frac{1}{x_6} + \frac{1}{x_7} + \frac{0}{x_8} + \frac{1}{x_9}$ 是 U 上的一个目标概念, 即 $X = \{x_4, x_5, x_6, x_7, x_9\}$, $U/R_1 = \{\{x_1, x_2, \dots, x_6\}, \{x_7, x_8, x_9\}\}$ 和 $U/R_2 = \{\{x_1, x_2, x_3, x_4\}, \{x_5, x_6, x_7\}, \{x_8, x_9\}\}$ 是 U 上的两个知识空间。

$$\begin{aligned} \text{QSD}(U/R_1, U/R_2) &= \frac{1}{|U|} \sum_{i=1}^2 \sum_{j=1}^3 \frac{|p_i \oplus q_j|}{|U|} |p_i \cap q_j| = \\ &= \frac{\frac{2}{9} \times 4 + \frac{5}{9} \times 2 + \frac{4}{9} \times 1 + \frac{1}{9} \times 2}{9} = 0.296 \end{aligned}$$

通过多对多匹配计算, QSD 实现了两个知识空间之间的差异性度量。但是, 对于粗糙集来说, 当前知识距离的工作 (包括 QSD) 仅仅是从知识空间的角度出发, 反映的是两个知识空间划分差异性, 并没有考虑它们对目标概念刻画时的差异性, 有可能导致同一目标概念在两个不同粒度的知识空间中具有相同的不确定性度量结果, 从而无法区分它们对目标概念的近似能力。为了解决以上问题, 本文提出了一种粗糙近似空间知识距离 (rough approximation space distance, RASD)。如图 3 所示, 为 QSD 和 RASD 两种类型的知识距离的示意图, 与 QSD 相比, RASD 可以体现两个知识空间对目标概念刻画能力的差异性。

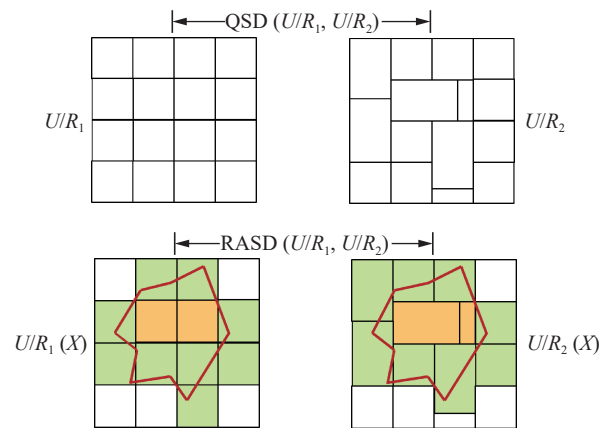


图 3 两种类型的知识距离

Fig. 3 Two types of knowledge distance

2.1 粗糙近似空间距离

为了反映不同知识空间对目标概念的刻画能力的差异性, 本文基于知识距离框架进一步提出了一种粗糙近似空间距离。

定义 7 模糊集合距离。假设 U 为一个非空有限论域, X 是 U 上的一个目标概念。如果 \tilde{A}, \tilde{B} 是 U 上的两个有限集合, A 和 B 之间的距离为

$$\delta(\tilde{A}, \tilde{B}) = \frac{\sum_{x \in \tilde{A} \cup \tilde{B}} \mu(x) - \sum_{x \in \tilde{A} \cap \tilde{B}} \mu(x)}{|U|}$$

由定义 7 可知, 无论 X 为 U 上的一个清晰集还是模糊集, $\delta(\tilde{A}, \tilde{B})$ 都可以刻画两个集合 (清晰集或模糊集) 之间的距离。

定理 2 假设 U 为一个非空有限论域, $\delta(\cdot, \cdot)$ 是 U 上的一个距离度量。

证明 假设 \tilde{A} 、 \tilde{B} 、 \tilde{C} 是 U 上的 3 个有限集合。显然, $\delta(\cdot, \cdot)$ 满足正定性和对称性。通过文献 [36] 可知:

$$\left(\sum_{x \in \tilde{A} \cup \tilde{B}} \mu(x) - \sum_{x \in \tilde{A} \cap \tilde{B}} \mu(x)\right) + \left(\sum_{x \in \tilde{B} \cup \tilde{C}} \mu(x) - \sum_{x \in \tilde{B} \cap \tilde{C}} \mu(x)\right) \geq$$

$$\left(\sum_{x \in \tilde{A} \cup \tilde{C}} \mu(x) - \sum_{x \in \tilde{A} \cap \tilde{C}} \mu(x)\right)$$

显然

$$\frac{\sum_{x \in \tilde{A} \cup \tilde{B}} \mu(x) - \sum_{x \in \tilde{A} \cap \tilde{B}} \mu(x)}{|U|} + \frac{\sum_{x \in \tilde{B} \cup \tilde{C}} \mu(x) - \sum_{x \in \tilde{B} \cap \tilde{C}} \mu(x)}{|U|} \geq$$

$$\frac{\sum_{x \in \tilde{A} \cup \tilde{C}} \mu(x) - \sum_{x \in \tilde{A} \cap \tilde{C}} \mu(x)}{|U|}$$

则

$$\delta(\tilde{A}, \tilde{B}) + \delta(\tilde{B}, \tilde{C}) \geq \delta(\tilde{A}, \tilde{C})$$

因此, $\delta(\cdot, \cdot)$ 是 U 上的一个距离度量。

例 2 假设 $U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$, $X = \frac{0.3}{x_1} + \frac{0.5}{x_2} + \frac{0.7}{x_3} + \frac{0.9}{x_4} + \frac{0.8}{x_5} + \frac{0.5}{x_6} + \frac{0.2}{x_7}$ 是 U 上的一个目标概念, $A = \{x_1, x_2, x_3, x_4\}$, $B = \{x_4, x_5, x_6, x_7\}$, 则 $\tilde{A} = \frac{0.3}{x_1} + \frac{0.5}{x_2} + \frac{0.7}{x_3} + \frac{0.9}{x_4}$, $\tilde{B} = \frac{0.9}{x_4} + \frac{0.8}{x_5} + \frac{0.5}{x_6} + \frac{0.2}{x_7}$ $\sum_{x \in X} \mu(x) = 3.9$, $\sum_{x \in \tilde{A} \cap \tilde{B}} \mu(x) = 0.7 + 0.9 = 1.6$, $\sum_{x \in \tilde{A} \cup \tilde{B}} \mu(x) = 0.3 + 0.5 + 0.7 + 0.9 + 0.8 + 0.5 = 3.7$, 因此, $\delta(\tilde{A}, \tilde{B}) = \frac{3.7 - 1.6}{7} = 0.3$ 。

定义 8 设一个信息系统 $S = (U, C \cup D, V, f)$, $R_1, R_2 \subseteq C$, X 是 U 上的一个目标概念。 $U/R_1 = \{p_1, p_2, \dots, p_l\}$ 和 $U/R_2 = \{q_1, q_2, \dots, q_m\}$ 是 U 上的两个知识空间。对于描述 X 、 U/R_1 和 U/R_2 的知识距离定义为

$$\text{RASD}(U/R_1(X), U/R_2(X)) = \frac{\sum_{i=1}^l \sum_{j=1}^m \delta_{ij} f_{ij}}{|U|} \quad (7)$$

式中: $\delta_{ij} = \delta(\tilde{p}_i, \tilde{q}_j)$, \tilde{p}_i 和 \tilde{q}_j 分别代表 p_i 和 q_j 对应的模糊集; $f_{ij} = |p_i \cap q_j|$ 。

由于对同一论域进行划分的不同知识空间的对象数是相同的, 即 $\sum_{i=1}^l |p_i| = \sum_{j=1}^m |q_j| = |U|$, 式 (7) 符合约束条件式 (4)。同样, 式 (7) 符合约束条件式 (1)~(3)。

定理 3 假设 U 为一个非空有限论域, RASD 是 U 上的一个距离度量。

证明 假设 $U/R_1 = \{p_1, p_2, \dots, p_l\}$ 和 $U/R_2 = \{q_1, q_2, \dots, q_m\}$ 是两个 U 上的知识空间。对于同一个论域而言, 式 (7) 中 $\sum_{i=1}^l |p_i| = \sum_{j=1}^m |q_j| = |U|$ 。由定

理 2 可知, $\delta(\cdot, \cdot)$ 是 U 上的距离度量。通过定理 1 可知, 类似于 EMD, RASD 也是 U 上的距离度量。

由定义 7 和定义 8 可知, 无论 X 为 U 上的一个清晰集还是模糊集, RASD 都可以刻画两个知识空间之间的对 X 近似刻画能力的差异, 即 RASD 不仅适用于经典粗糙集, 同样适用于粗糙模糊集。由于继承了 EMD 的优点, RASD 能够有效和直观的刻画不同知识空间对目标概念的描述能力的差异性。

例 3(续例 1)

$$\text{RASD}(U/R_1(X), U/R_2(X)) =$$

$$\frac{\sum_{i=1}^2 \sum_{j=1}^3 \delta_{ij} f_{ij}}{|U|} =$$

$$\frac{\frac{2}{9} \times 4 + \frac{2}{9} \times 2 + \frac{3}{9} \times 1 + \frac{1}{9} \times 2}{9} = 0.209$$

2.2 粗糙粒结构的结构特征

定理 4 设一个信息系统 $S = (U, C \cup D, V, f)$, $R_1, R_2, R_3 \subseteq C$, X 是 U 上的一个目标概念。如果 $R_1 \subseteq R_2 \subseteq R_3$, 则 $\text{RASD}(U/R_1(X), U/R_3(X)) = \text{RASD}(U/R_1(X), U/R_2(X)) + \text{RASD}(U/R_2(X), U/R_3(X))$ 。

证明 假设 $U = \{x_1, x_2, \dots, x_n\}$ 是一个非空论域, $U/R_1 = \{p_1, p_2, \dots, p_l\}$, $U/R_2 = \{q_1, q_2, \dots, q_m\}$ 和 $U/R_3 = \{r_1, r_2, \dots, r_s\}$ 是 U 上的 3 个知识空间。由于 $R_1 \subseteq R_2 \subseteq R_3$, 故 $U/R_1 \leq U/R_2 \leq U/R_3$ 。为了简单化, 本文假设仅有一个信息粒 $p_1 (p_1 \in U/R_1)$ 细分为两个更细的信息粒 $q_1, q_2 (q_1, q_2 \in U/R_2)$, 仅有一个信息粒 q_1 细分为两个更细的信息粒 r_1, r_2 (其他复杂情形均可转化为这种情形, 这里不再重复)。则 $p_1 = q_1 \cup q_2$, $p_2 = q_3, \dots, p_n = q_m$ ($m = l + 1$), $q_1 = r_1 \cup r_2$, $q_2 = r_3, \dots, q_n = r_s$ ($s = m + 1$), 那么 $U/R_2 = \{q_1, q_2, p_2, p_3, \dots, p_l\}$ 以及 $U/R_3 = \{r_1, r_2, q_2, q_3, \dots, q_l\}$ 。通过公式, 可得:

$$\text{RASD}(U/R_1(X), U/R_2(X)) =$$

$$\frac{1}{|U|} \sum_{i=1}^l \sum_{j=1}^m \frac{(\sum_{x \in \tilde{p}_i} \mu(x) - \sum_{x \in \tilde{q}_j} \mu(x)) \sum_{x \in \tilde{q}_j} \mu(x)}{|U|}$$

$$\text{RASD}(U/R_1(X), U/R_3(X)) =$$

$$\frac{1}{|U|} \sum_{i=1}^l \sum_{k=1}^s \frac{(\sum_{x \in \tilde{p}_i} \mu(x) - \sum_{x \in \tilde{r}_k} \mu(x)) \sum_{x \in \tilde{r}_k} \mu(x)}{|U|}$$

由于 $p_1 = q_1 \cup q_2$ 和 $q_1 = r_1 \cup r_2$, 故 $\mu_{p_1} = \mu_{q_1} + \mu_{q_2}$ 和 $\mu_{q_1} = \mu_{r_1} + \mu_{r_2}$ 。

$$\text{RASD}(U/R_1(X), U/R_2(X)) =$$

$$\frac{1}{|U|} \cdot \frac{\mu_{p_1} (\mu_{q_1} + \mu_{q_2}) - (\mu_{q_1}^2 + \mu_{q_2}^2)}{|U|}$$

$$\begin{aligned} \text{RASD}(U/R_1(X), U/R_3(X)) = \\ \frac{1}{|U|} \cdot \frac{\mu_{p_1}(\mu_{r_1} + \mu_{r_2} + \mu_{r_3}) - (\mu_{r_1}^2 + \mu_{r_2}^2 + \mu_{r_3}^2)}{|U|} = \\ \frac{1}{|U|} \cdot \frac{2(\mu_{r_1}\mu_{r_2} + \mu_{r_2}\mu_{r_3} + \mu_{r_1}\mu_{r_3})}{|U|} \end{aligned}$$

同理, $\text{RASD}(U/R_2(X), U/R_3(X)) =$

$$\begin{aligned} \frac{1}{|U|} \cdot \frac{(\mu_{q_1} - \mu_{r_1})\mu_{r_1} + (\mu_{q_1} - \mu_{r_2})\mu_{r_2}}{|U|} = \\ \frac{1}{|U|} \cdot \frac{(\mu_{q_1}\mu_{r_1} + \mu_{q_1}\mu_{r_2}) - (\mu_{r_1}^2 + \mu_{r_2}^2)}{|U|} \end{aligned}$$

由于 $p_1 = q_1 \cup q_2$ 和 $q_1 = r_1 \cup r_2$, 故 $\mu_{p_1} = \mu_{q_1} + \mu_{q_2}$

和 $\mu_{q_1} = \mu_{r_1} + \mu_{r_2}$, 则

$$\begin{aligned} \text{RASD}(U/R_1(X), U/R_2(X)) + \text{RASD}(U/R_2(X), U/R_3(X)) = \\ \frac{1}{|U|} \cdot \frac{2(\mu_{r_1}\mu_{r_2} + \mu_{q_1}\mu_{q_2})}{|U|} = \\ \frac{1}{|U|} \cdot \frac{2(\mu_{r_1}\mu_{r_2} + \mu_{r_1}\mu_{r_3} + \mu_{r_2}\mu_{r_3})}{|U|} = \\ \text{RASD}(U/R_1(X), U/R_3(X)) \end{aligned}$$

由定理 4 可知, 同一个粗糙粒结构中的两个粗糙近似空间的不确定性差异等于它们之间的知识距离。

例 4 设一个信息系统 $S = (U, C \cup D, V, f)$, $U =$

$\{x_1, x_2, \dots, x_9\}$, $R_1, R_2, R_3 \subseteq C$, $X = \frac{0.4}{x_1} + \frac{0.6}{x_2} + \frac{0.8}{x_3} + \frac{0.9}{x_4} + \frac{0.8}{x_5} + \frac{0.5}{x_6} + \frac{0.4}{x_7} + \frac{0.4}{x_8} + \frac{0.2}{x_9}$ 是 U 上的一个目标概念, $U/R_1 = \{\{x_1, x_2, \dots, x_6\}, \{x_7, x_8, x_9\}\}$, $U/R_2 = \{\{x_1, x_2, x_3, x_4\}, \{x_5, x_6\}, \{x_7, x_8, x_9\}\}$, 以及 $U/R_3 = \{\{x_1, x_2, x_3\}, \{x_4\}, \{x_5, x_6\}, \{x_7\}, \{x_8, x_9\}\}$ 是 U 上的 3 个知识空间。

$$\text{RASD}(U/R_1(X), U/R_2(X)) = \frac{\frac{1.3}{5} \times 2.7 + \frac{2.7}{5} \times 1.3}{9} = \frac{10.6}{81}$$

类似地

$$\text{RASD}(U/R_2(X), U/R_3(X)) = \frac{5.9}{81}$$

$$\text{RASD}(U/R_1(X), U/R_3(X)) = \frac{16.5}{81}$$

因此

$$\begin{aligned} \text{RASD}(U/R_1, U/R_3) = \\ \text{RASD}(U/R_1(X), U/R_2(X)) + \\ \text{RASD}(U/R_2(X), U/R_3(X)) = \frac{16.5}{81} \end{aligned}$$

在本文中, 将一个信息系统中的最粗知识空间和最细知识空间分别表示为 σ 和 ω 。通过定理 4, 本文可得到以下推论。

推论 1 设一个信息系统 $S = (U, C \cup D, V, f)$, $R_1, R_2 \subseteq C$, X 是 U 上的一个目标概念。如果 $R_1 \subseteq R_2$, 则:

$$\begin{aligned} \text{RASD}(U/R_1(X), U/R_2(X)) = \\ \text{RASD}(U/R_1(X), \omega) - \text{RASD}(U/R_2(X), \omega) \end{aligned}$$

推论 2 设一个信息系统 $S = (U, C \cup D, V, f)$, $R_1, R_2 \subseteq C$, X 是 U 上的一个目标概念。如果 $R_1 \subseteq R_2$, 则:

$$\begin{aligned} \text{RASD}(U/R_1(X), U/R_2(X)) = \\ \text{RASD}(U/R_2(X), \sigma) - \text{RASD}(U/R_1(X), \sigma) \end{aligned}$$

定理 5 设一个信息系统 $S = (U, C \cup D, V, f)$, $R_1 \subseteq C$, X 是 U 上的一个目标概念, $\text{RASD}(U/R_1(X), \omega)$ 是一个粒度度量。

证明 1) 由定理 3 可知, $\text{RASD}(U/R_1(X), \omega) \geq 0$;

2) 假设 $R_2 \subseteq C$, 当 $R_1 = R_2$ 时, 可得:

$$\text{RASD}(U/R_1(X), \omega) = \text{RASD}(U/R_2(X), \omega)$$

3) 由推论 1 可知, 如果 $R_1 \subseteq R_2$, 那么 $\text{RASD}(U/R_2(X), \omega) \leq \text{RASD}(U/R_1(X), \omega)$ 。

定理 6 设一个信息系统 $S = (U, C \cup D, V, f)$, $R_1 \subseteq C$, X 是 U 上的一个目标概念, $\text{RASD}(U/R_1(X), \sigma)$ 是一个信息度量。

证明 1) 由定理 3 可知, $\text{RASD}(U/R_1(X), \sigma) \geq 0$;

2) 假设 $R_2 \subseteq C$, 当 $R_1 = R_2$ 时, 可得:

$$\text{RASD}(U/R_1(X), \sigma) = \text{RASD}(U/R_2(X), \sigma)$$

3) 由推论 1 可知, 如果 $R_1 \subseteq R_2$, 那么 $\text{RASD}(U/R_1(X), \sigma) \leq \text{RASD}(U/R_2(X), \sigma)$ 。

通过推论 1 和定理 5 可知, 对于一个目标概念, 粗糙粒结构中的任意两个粗糙近似空间的 RASD 等于它们之间的粒度度量差异。通过推论 2 和定理 6 可知, 对于一个目标概念, 粗糙粒结构中的任意两个粗糙近似空间的 RASD 等于它们之间的信息度量差异。结合定理 4, 本文以粗糙粒结构中的 ω 为原点, 可以将粗糙近似空间通过 RASD 映射到一维坐标上。

3 粗糙粒结构的评价模型

在粗糙粒结构中, 对不确定性问题进行求解时, 通常希望在约束条件下可以利用尽可能少的知识空间使不确定性降低最大化。为了简单化, 本文仅考虑约束条件为属性代价时的情形, 通过第 2 节定义的粗糙知识距离以及在粗糙粒结构上得出的结论, 提出了一种粗糙粒结构的评价模型。

定义 9 知识距离序列。假设一个信息系统 $S = (U, C \cup D, V, f)$, 其中 $R \subseteq C$ 和 $X \subseteq U$, $\text{GS} = (\text{KS}_1, \text{KS}_2, \dots, \text{KS}_\xi)$ 是一个由 ξ 个知识空间组成的粗糙粒结构, 其中 $\text{KS}_i = U/R_i$, $i = 1, 2, \dots, \xi$ 。 R_i 为一个条件属性集, $R_\xi \subset \dots \subset R_2 \subset R_1 \subseteq C$, 知识距离序列可表示为 $L_{\text{GS}} = (\Delta_1, \Delta_2, \dots, \Delta_{\xi-1})$, 其中 $\Delta_i = \text{RASD}(U/R_i(X), U/R_{i+1}(X))$, $i \in [1, \xi - 1]$ 。

基于定义 9, 本文提出了一种属性约束条件下的粗糙粒结构的评价模型, 首先, 本文在定义 10 中定义了一个评价参数 λ 。

定义 10 评价参数。假设一个信息系统 $S = (U, C \cup D, V, f)$, 其中 $R \subseteq C$ 和 $X \subseteq U$, $\text{GS} = (\text{KS}_1, \text{KS}_2, \dots,$

KS_{ξ}) 是一个由 ξ 个知识空间组成的粗糙粒结构, $L_{GS} = (\Delta_1, \Delta_2, \dots, \Delta_{\xi-1})$, 则

$$\lambda = \frac{\sum_{i=1}^{\psi-1} \Delta_i}{\psi} \quad (8)$$

其中,

$$\begin{aligned} \psi &= \max_{1 \leq k \leq \xi} k \\ \text{s.t.} \quad & TC_{KS_k} \leq TC_{user} \end{aligned}$$

通过式 (8) 可知, λ 越大, 说明对应的粗糙粒结构就越好, 即该粒结构能够在约束条件下利用尽可能少的知识空间实现不确定性的降低最大化。

如图 4 所示, 为整个粗糙粒结构评价模型的流程图。首先, 通过粗糙近似空间距离计算出不同的粗糙粒结构对应的知识距离序列, 然后, 通过式 (8) 并结合不同的约束条件计算得到对应的评价参数, 从而实现不同约束条件下的粗糙粒结构的评价和选择。

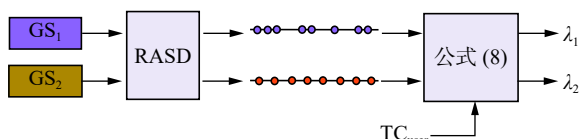


图 4 粗糙粒结构的评价模型

Fig. 4 Evaluation model of rough granular structure

例 5 如图 5 所示, 设一个信息系统 $S = (U, C \cup D, V, f)$, $X = \frac{0}{x_1} + \frac{0}{x_2} + \frac{1}{x_3} + \frac{0}{x_4} + \frac{1}{x_5} + \frac{1}{x_6} + \frac{1}{x_7} + \frac{1}{x_8} + \frac{0}{x_9}$ 是 U 上的一个目标概念, $GS_1 = (KS_1^1, KS_2^1, KS_3^1, KS_4^1)$ 和 $GS_2 = (KS_1^2, KS_2^2, KS_3^2, KS_4^2)$ 是两个粗糙粒结构, 它们的属性代价序列分别为 $TC_1 = (2, 13, 27, 31)$ 和 $TC_2 = (2, 7, 22, 28)$, 其中属性代价序列为 $TC_{user} = (10, 20, 30, 40)$ 。

$$\begin{aligned} KS_1^1 &= \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9\} & KS_2^1 &= \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9\} \\ KS_1^2 &= \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9\} & KS_2^2 &= \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9\} \\ KS_1^3 &= \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9\} & KS_2^3 &= \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9\} \\ KS_1^4 &= \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9\} & KS_2^4 &= \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9\} \end{aligned}$$

图 5 两个粗糙粒结构中的知识空间

Fig. 5 Knowledge spaces in two rough granular structures

通过式 (7) 可计算出 GS_1 和 GS_2 的知识距离序列分别为 $L_{GS_1} = \left(\frac{10}{81}, \frac{25}{81}, \frac{3}{81}\right)$ 和 $L_{GS_2} = \left(\frac{5}{81}, \frac{18}{81}, \frac{8}{81}\right)$, 结合已知条件, 由式 (8) 可得, GS_1 中的 KS_1^1 和 GS_2 中的 KS_1^2 、 KS_2^2 属性代价小于 $TC_{user} = 10$, 那么, $\lambda_1 = 0$, $\lambda_2 = \frac{5}{81}$, 此时, $BND_1(X) = \{x_1, x_2, \dots, x_9\}$, $BND_2(X) = \{x_1, x_2, \dots, x_8\}$ 。

因此, $\lambda_1 > \lambda_2$ 和 $|BND_1(X)| > |BND_2(X)|$ 。由于 GS_1 在约束条件下只有一个知识空间可利用, 无法进一步降低不确定性, 不能进行渐近式计算, 而 GS_2 在约束条件下利用知识空间进一步降低了不确定性, 所以在约束条件 $TC_{user} = 10$ 下应该选择粗糙粒结构 GS_2 进行问题求解。

GS_1 中的 KS_1^1 、 KS_2^1 和 GS_2 中的 KS_1^2 、 KS_2^2 、 KS_3^2 属性代价小于 $TC_{user} = 20$, 那么:

$$\lambda_1 = \frac{10}{\frac{81}{2-1}} = \frac{10}{81}, \lambda_2 = \frac{5}{\frac{81}{3-1}} + \frac{18}{81} = \frac{23}{162}$$

此时, $BND_1(X) = \{x_3, x_4, \dots, x_9\}$, $BND_2(X) = \{x_1, x_2, \dots, x_6\}$ 。

因此, $\lambda_1 < \lambda_2$ 和 $|BND_1(X)| > |BND_2(X)|$, 虽然 GS_2 在约束条件下利用了更多的知识空间个数, 但是总体上却更大程度的降低了不确定性, 所以在约束条件 $TC_{user} = 20$ 下应该选择粗糙粒结构 GS_2 进行问题求解。

同理, GS_1 中的 KS_1^1 、 KS_2^1 、 KS_3^1 和 GS_2 中的 KS_1^2 、 KS_2^2 、 KS_3^2 、 KS_4^2 属性代价小于 $TC_{user} = 30$, 那么:

$$\lambda_1 = \frac{10}{\frac{81}{3-1}} + \frac{25}{81} = \frac{35}{162}, \lambda_2 = \frac{5}{\frac{81}{4-1}} + \frac{18}{81} + \frac{8}{81} = \frac{31}{243}$$

此时, $BND_1(X) = \{x_3, x_4\}$, $BND_2(X) = \{x_1, x_2, x_3, x_4\}$ 。

因此, $\lambda_1 > \lambda_2$ 和 $|BND_1(X)| < |BND_2(X)|$ 。由于 GS_1 在约束条件下利用了更少的知识空间个数降低了更多的不确定性, 所以在约束条件 $TC_{user} = 30$ 下应该选择粗糙粒结构 GS_1 进行问题求解。

GS_1 中的 KS_1^1 、 KS_2^1 、 KS_3^1 、 KS_4^1 和 GS_2 中的 KS_1^2 、 KS_2^2 、 KS_3^2 、 KS_4^2 属性代价小于 $TC_{user} = 40$, 那么:

$$\lambda_1 = \frac{10}{\frac{81}{4-1}} + \frac{25}{81} + \frac{3}{81} = \frac{38}{243}, \lambda_2 = \frac{5}{\frac{81}{4-1}} + \frac{18}{81} + \frac{8}{81} = \frac{31}{243}$$

此时, $BND_1(X) = \emptyset$, $BND_2(X) = \{x_1, x_2, x_3, x_4\}$ 。

因此, $\lambda_1 > \lambda_2$ 和 $|BND_1(X)| < |BND_2(X)|$ 。由于 GS_1 在相同的知识空间个数上降低了更多得不确定性, 所以在约束条件 $TC_{user} = 40$ 下应该选择粗糙粒结构 GS_1 进行问题求解。

4 结束语

不同的粒化机制可以形成不同的粒结构, 在粗糙集的序贯三支决策模型中, 通过不同的属性集添加顺序可形成不同的粗糙粒结构。如何建立对这些粗糙粒结构的评价方法是本文的主要研究内容。在前期工作的基础上, 提出了一种粗糙近似空间距离, 并利用该距离研究了粗糙粒结构的结构特征, 得到了粗糙粒结构中的两个知识空间的不确定性差异等于它们之间知识距离的结论。

利用该结论,本文建立了一种在属性代价约束条件下的粗糙粒结构的评价模型,解决了在不同约束条件下评价并选择不同粗糙粒结构的问题。由于本文采用的是基于 EMD 的知识距离框架,因此,本文的方法具有扩展性,可以针对不同类型的粒结构设计评价模型,例如,对不同层次聚类算法、多尺度图像分割算法以及多粒度社区发现算法的评价和选择。但是,本文的工作还处于初步阶段,仅研究了如何评价粗糙粒结构,并没有考虑粗糙粒结构的构建问题,并且只考虑了属性代价为约束条件,评价模型的目标函数还有待进一步改进,尤其如何根据不同的约束条件设计目标函数是下一步工作的主要目标。

参考文献:

- [1] ZADEH L A. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic[J]. *Fuzzy sets and systems*, 1997, 90(2): 111–127.
- [2] PEDRYCZ W, SKOWRON A, KREINOVICH V. Handbook of granular computing[M]. Chichester: John Wiley & Sons, 2008: 719–740.
- [3] PEDRYCZ W, AL-HMOUZ 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.
- [4] WANG Guoyin, YANG Jie, XU Ji. Granular computing: from granularity optimization to multi-granularity joint problem solving[J]. *Granular computing*, 2017, 2(3): 105–120.
- [5] YAO Yiyu. Perspectives of granular computing[C]//Proceedings of 2015 IEEE International Conference on Granular Computing. Beijing, China, 2005: 85–90.
- [6] ZADEH L A. Fuzzy sets[J]. *Information and control*, 1965, 8(3): 338–353.
- [7] PAWLAK Z. Rough sets[J]. *International journal of computer & information sciences*, 1982, 11(5): 341–356.
- [8] 张钹, 张铃. 问题求解理论及应用 [M]. 北京: 清华大学出版社, 1990.
- [9] 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(6): 15–20.
LI Deyi, MENG Haijun, SHI Xuemei. Membership clouds and membership cloud generators[J]. *Journal of computer research and development*, 1995(6): 15–20.
- [10] 苗夺谦, 李德毅, 姚一豫, 等. 不确定性与粒计算 [M]. 北京: 科学出版社, 2011.
- [11] 徐计, 王国胤, 于洪. 基于粒计算的大数据处理 [J]. *计算机学报*, 2015(8): 1497–1517.
- XU Ji, WANG Guoyin, YU Hong. Review of big data processing based on granular computing[J]. *Chinese journal of computers*, 2015(8): 1497–1517.
- [12] YAO Yiyu. Three-way decisions with probabilistic rough sets[J]. *Information sciences*, 2010, 180(3): 341–353.
- [13] QIAN Yuhua, LIANG Jiye, YAO Yiyu, et al. MGRS: a multi-granulation rough set[J]. *Information sciences*, 2010, 180(6): 949–970.
- [14] QIAN Yuhua, LIANG Jiye, DANG Chuangyin. Incomplete multigranulation rough set[J]. *IEEE transactions on systems, man, and cybernetics-part A: systems and humans*, 2010, 40(2): 420–431.
- [15] QIAN Yuhua, CHENG Honghong, WANG Jieting, et al. Grouping granular structures in human granulation intelligence[J]. *Information sciences*, 2017, 382–383: 150–169.
- [16] 张钹, 张铃. 问题求解理论及应用 [M]. 北京: 清华大学出版社, 1990.
- [17] ZHANG Ling, ZHANG Bo. The structure analysis of fuzzy sets[J]. *International journal of approximate reasoning*, 2005, 40(1/2): 92–108.
- [18] ZHANG Ling, ZHANG Bo. Fuzzy reasoning model under quotient space structure[J]. *Information sciences*, 2005, 173(4): 353–364.
- [19] WIERMAN M J. Measuring uncertainty in rough set theory[J]. *International journal of general systems*, 1999, 28(4/5): 283–297.
- [20] LIANG Jiye, CHIN K S, DANG Chuangyin, et al. A new method for measuring uncertainty and fuzziness in rough set theory[J]. *International journal of general systems*, 2002, 31(4): 331–342.
- [21] YAO Yiyu, ZHAO Liquan. A measurement theory view on the granularity of partitions[J]. *Information sciences*, 2012, 213: 1–13.
- [22] LAWVERE F W. Metric spaces, generalized logic, and closed categories[J]. *Rendiconti del seminario matematico e fisico di milano*, 1973, 43(1): 135–166.
- [23] RUBNER Y, GUIBAS L, TOMASI C. The earth mover's distance, multi-dimensional scaling, and color-based image retrieval[C]//Proceedings of the ARPA Image Understanding Workshop. 1997: 661–668.
- [24] RUBNER Y, TOMASI C, GUIBAS L J. The earth mover's distance as a metric for image retrieval[J]. *International journal of computer vision*, 2000, 40(2): 99–121.
- [25] 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.

- GUO Zengxiao, MI Jusheng. An uncertainty measure in rough fuzzy sets[J]. *Fuzzy systems and mathematics*, 2005, 19(4): 135–140.
- [26] QIN Huani, LUO Darong. New uncertainty measure of rough fuzzy sets and entropy weight method for fuzzy-target decision-making tables[J]. *Journal of applied mathematics*, 2014, 2014: 487036.
- [27] HU J, PEDRYCZ W, WANG G. A roughness measure of fuzzy sets from the perspective of distance[J]. *International journal of general systems*, 2016, 45(3): 352–367.
- [28] SUN Bingzhen, MA Weimin. Uncertainty measure for general relation-based rough fuzzy set[J]. *Kybernetes*, 2013, 42(6): 979–992.
- [29] DAI Jianhua, WANG Wentao, XU Qing. An uncertainty measure for incomplete decision tables and its applications[J]. *IEEE transactions on cybernetics*, 2013, 43(4): 1277–1289.
- [30] QIAN Y H, LIANG J Y, DANG C Y, et al. Knowledge granulation and knowledge distance in a knowledge base[J]. *International journal of approximate reasoning*, 2011, 19(2): 263–264.
- [31] QIAN Yuhau, LI Yebin, LIANG Jiye, et al. Fuzzy granular structure distance[J]. *IEEE transactions on fuzzy systems*, 2015, 23(6): 2245–2259.
- [32] LIANG Jiye, LI Ru, QIAN Yuhau. Distance: a more comprehensible perspective for measures in rough set theory[J]. *Knowledge-based systems*, 2012, 27: 126–136.
- [33] YANG Xibei, QIAN Yuhua, YANG Jingyu. On characterizing hierarchies of granulation structures via distances[J]. *Fundamenta informaticae*, 2013, 123(3): 365–380.
- [34] YANG Xibei, SONG Xiaoning, SHE Yanhong, et al. Hierarchy on multigranulation structures: a knowledge distance approach[J]. *International journal of general systems*, 2013, 42(7): 754–773.
- [35] YANG Jie, WANG Guoyin, ZHANG Qinghua. Knowledge distance measure in multigranulation spaces of fuzzy equivalence relations[J]. *Information sciences*, 2018, 448-449: 18–35.
- [36] YANG Jie, WANG Guoyin, LI Xukun. Multi-granularity similarity measure of cloud concept[C]//International Joint Conference on Rough Sets. Santiago, Chile, 2016: 318–330.

作者简介:



杨洁, 副教授, 博士, 主要研究方向为粒计算、数据挖掘、粗糙集。参与国家自然科学基金项目 2 项, 授权国家发明专利 5 项。发表学术论文 20 余篇。



王国胤, 教授, 博士生导师, 博士, 重庆邮电大学研究生院院长、人工智能学院院长, 主要研究方向为粗糙集、粒计算、认知计算、数据挖掘、智能信息处理。主持国家自然科学基金项目 10 余项。入选“重庆高校创新团队”、“重庆市首席专家工作室”和“国家级教学团队”; 获国家级教学成果二等奖、重庆市自然科学一等奖、重庆市科技进步一等奖、重庆市教学成果一等奖和吴文俊人工智能科学技术奖科技进步一等奖等多项国家级/省部级教学和科技成果奖励。出版著作 10 余部, 发表学术论文 300 余篇。



张清华, 教授, 博士生导师, 博士, 主要研究方向为粗糙集、模糊集、粒计算、三支决策。主持国家自然科学基金项目 2 项。发表学术论文 40 余篇。