

模糊决策表中基于 OWA 算子的三支属性约简

杨霖琳¹ 张贤勇² 唐孝²

(1. 四川师范大学基础教学学院, 成都, 610068; 2. 四川师范大学数学与软件科学学院, 成都, 610068)

摘要: 在模糊决策表中, 基于有序加权平均(Ordered weighted averaging, OWA)算子建立的相容关系中, OWA 算子是一个信息集成工具。本文利用 OWA 算子及其诱导的截集相容关系, 分别讨论基于正域、负域与边界域的属性约简。首先, 利用 OWA 算子中的权重给出属性区分对象的贡献度; 然后, 根据属性贡献度定义每个属性被约简的可能性; 从而给出模糊决策表的一种启发式三支属性约简方法, 该方法可以减少属性约简的搜索空间并避免属性约简的盲目性; 最后, 利用实例来分析影响该三支区域属性约简方法的主要因素, 并说明该启发式属性约简方法的合理性和有效性。本文提出的属性贡献度度量及启发式三支属性约简方法将减少属性约简的搜索空间, 有益于模糊决策表特征选取的应用。

关键词: 三支决策; 粗糙集; 属性约简; OWA 算子; 相容关系

中图分类号: TP18 **文献标志码:** A

Three-Way Attribute Reductions Based on OWA Operator in Fuzzy Decision Table

Yang Jilin¹, Zhang Xianyong², Tang Xiao²

(1. College of Fundamental Education, Sichuan Normal University, Chengdu, 610068, China; 2. College of Mathematics and Software Science, Sichuan Normal University, Chengdu, 610068, China)

Abstract: In the fuzzy decision table, the tolerance relation is built based on the ordered weighted averaging (OWA) operator, and OWA operator is a basic tool of information integration. Based on the OWA operator and its induced tolerance relation of cut sets, attribute reductions of the positive region, negative region and boundary region are discussed. Firstly, an attribute contribution degree distinguishing objects is defined by the weight of the OWA operator, to describe the reduction possibility of attributes. Consequently, a method of three-way attribute reduction based on the reduction possibility is proposed in the fuzzy decision table. The method reduces the search space and avoids reduction blindness. Finally, an example is analyzed to show the influencing causes of attribute reductions and the validity of the proposed method. The proposed measurement method of attribute contribution degree and the heuristic method of attribute reduction based on the three regions will reduce the searching space of attribute reduction, and thus they are benefit for applications of the fuzzy decision table.

Key words: three-way decisions; rough set; attribute reduction; OWA operator; tolerance relation

引言

粗糙集理论是一种处理模糊和不确定性知识的数学工具^[1],其主要思想是以等价关系为基础,利用已有的知识库来表示不确定或不精确的知识,在保持分类能力不变的前提下,通过知识约简导出问题的决策规则。这种模型在完备的信息系统中得到了成功运用^[1,2]。

在模糊决策表中,数据往往被认为是一个模糊概念,于是研究者们对经典的粗糙集理论进行扩充^[3,4]。许多学者在构造各种不同区分关系的基础上,讨论各种属性约简的理论和方法,并进行了相关应用研究^[5,6]。管涛^[7]等基于模糊集合的贴近度,构造模糊相似关系,取其截集后得到不可区分关系,利用水平集粗糙成员函数给出分布约简与分配约简;Jensen 和 Shen 提出以依赖度函数作为启发信息的相对约简算法^[8,9],但 Bhatt^[10]指出该算法在许多实际分析中不收敛;张慧哲^[11]等提出一种基于海明距离定义的相似度系数,通过定义模糊相似矩阵和不一致程度矩阵,给出属性约简的定义及算法;曾雪兰^[12]等提出一种全序优势关系,该全序优势关系是 Greco^[13]提出的优势关系的扩充,并进而给出关于对象的相对上下近似约简计算方法;胡清华等^[14,15]提出邻域粗糙集模型,其中数值型属性通过欧式距离表示为模糊相似关系,以决策属性对条件属性的依赖度为标准给出属性约简的概念与约简方法;张家录^[16]等提出基于模糊包含的模糊粗糙集模型,通过计算模糊信任测度、模糊似然测度来进行随机模糊信息系统的属性约简;赵涛^[17]引入模糊随机变量,提出期望相关关系,并基于该关系讨论属性约简方法;黄兵等^[18]将区分函数引入模糊信息系统,以依赖度为约简标准,给出相应的知识约简方法。冯楠坪等^[19]通过模糊相似关系建立条件相似度与决策相似度之间的相对比较矩阵,给出一种基于相似度比较的模糊属性约简方法。在模糊决策表中,利用区分函数完成属性约简,算法复杂度较高,并且在实际应用中,往往不需要找到所有的属性约简。因此,以上文献针对模糊决策表,都是在基于特定的不可区分关系之上,构造相应的属性约简标准,建立其属性约简方法。本文将借鉴此理论分析方法,在模糊决策表中,基于有序加权平均(Ordered weighted averaging, OWA)算子建立的相容关系,给出属性重要度,构造属性约简标准,建立一种启发式属性约简算法。

三支决策是一种处理不确定性信息决策的智能理论,特别适用于粗糙集及其属性约简^[20]。在三支决策区域中,决策类的正域、边界域和负域分别对应接受、不承诺和拒绝决策。在模糊决策表中,以上文献大都基于正域对属性约简进行讨论。随着三支决策的发展,三支决策区域也用来构建三支属性约简^[21]。因此,本文将在正域属性约简的基础上,扩展到三支区域,分别讨论正域、负域和边界域的属性约简。前期研究工作中,在模糊信息系统中 OWA 算子能够诱导相容关系^[22,23]。在此基础上,本文进一步讨论属性约简方法。

1 基于 OWA 算子的粗糙集模型

1.1 基于 OWA 算子的相容关系

定义 1^[4] 设模糊信息系统 $\Omega = (U, A, V, f)$, $U = \{x_1, x_2, \dots, x_n\}$ 是非空有限对象集, $A = \{a_1, a_2, \dots, a_m\}$ 是属性集, $V = \{V_a \mid a \in A\}$ 是属性值集,此时 V_a 可以是一个隶属函数值,则对象 x 在条件属性 a 下属性值可以表示为 $\mu_a(x) \in [0, 1]$, f 表示一个映射, $f: U \times A \rightarrow V_a$, 即 $(x, a) \rightarrow \mu_a(x)$ 。

在模糊信息系统中, $\mu_a(x) \in [0, 1]$ 体现了数据的不确定性,但经典粗糙集的等价关系很难再成立。因此,经典粗糙集模型被许多学者做了进一步推广。在前期研究中,利用 OWA 算子聚合每个属性上在对象间区分时的差异,得到对象的相似度,从而利用对象相似度建立了 λ 截集的相容关系^[22]。

定义 2^[24] 设 $F: R^m \rightarrow R$, 有一与 F 相关联的 m 维加权向量, $w = (\omega_1, \omega_2, \dots, \omega_m)$, $\omega_i \in [0, 1]$, $1 \leq i \leq m$, 且 $\sum_{i=1}^m \omega_i = 1$, 使得 $F(a_1, a_2, \dots, a_m) = \sum_{i=1}^m \omega_i b_i$, 其中元素 b_i 是 (a_1, a_2, \dots, a_m) 中第 i 个最大元

素,则称 F 为 m 维 OWA 算子。

在 OWA 算子中,根据聚合要求,加权向量一般可通过模糊量词 Q 确定^[24],即模糊量词 Q 表示为

$$Q(r) = \begin{cases} 0 & 0 \leq r < \alpha \\ \frac{r-\alpha}{\beta-\alpha} & \alpha \leq r \leq \beta \\ 1 & \beta < r \leq 1 \end{cases} \quad (1)$$

式中: $\alpha, \beta \in [0, 1]$, (α, β) 有多种取值方式,如参数 (α, β) 最常用的是 3 种取值: $(0.3, 0.8)$, $(0, 0.5)$, $(0.5, 1)$, 它们分别表示模糊量词“大多数”,“至少一半”和“尽可能多”^[22]。相应地,有序加权向量 $w = (\omega_1, \omega_2, \dots, \omega_m)$ 可如下确定

$$\omega_i = Q\left[\frac{i}{m}\right] - Q\left[\frac{i-1}{m}\right] \quad (2)$$

定义 3^[22] 设 $\Omega = (U, A, V, f)$ 是一个模糊信息系统, $\forall x, y \in U$ 在属性集 A 上的相似度为

$$s_A(x, y) = 1 - F_A(\mathbf{T}_A) = 1 - \mathbf{H}_A(\mathbf{E}_A)^T \quad (3)$$

式中: F_A 是 OWA 算子, $\mathbf{T}_A = (\mu_{a_1}(x, y), \mu_{a_2}(x, y), \dots, \mu_{a_m}(x, y))$, $\mu_{a_j}(x, y) = |\mu_{a_j}(x) - \mu_{a_j}(y)|$ ($j=1, 2, \dots, m$), 是对象 x 和 y 在属性 a_j 上的差异。 $\mathbf{H}_A = (\omega_1, \omega_2, \dots, \omega_m)$ 根据式(1)和式(2)获得,它是模糊信息系统属性集 A 中各属性对应的权重。 \mathbf{T}_A 中各对象 $\mu_{a_j}(x, y)$ 按值从大到小排序后,得到 $\mathbf{E}_A = (\mu_{a^{(1)}}(x, y), \mu_{a^{(2)}}(x, y), \dots, \mu_{a^{(m)}}(x, y))$ 且满足 $\forall l \in \{1, 2, \dots, m\}$, 有 $\mu_{a^{(l)}}(x, y) \geq \mu_{a^{(l+1)}}(x, y)$ 。

定义 4^[22] 设 $\Omega = (U, A, V, f)$ 是一个模糊信息系统, $\forall x, y \in U$, 在属性集 $B \subseteq A$ 下,不可区分关系定义为 $R_\lambda: U \times U \rightarrow [0, 1]$, $xR_\lambda y = \{(x, y) \in U \times U \mid s_B(x, y) \geq \lambda\}$ 。其中, $s_B(x, y)$ 是对象 x 和 y 在属性集 B 上的相似度, $\lambda \in [0, 1]$ 是阈值,可根据具体问题设置。显然, R_λ 满足自反性和对称性,但不一定满足传递性,因此 R_λ 是相容关系。

定义 5^[22] 设 $\Omega = (U, A, V, f)$ 是一个模糊信息系统, $\forall x \in U$, 在属性集 A 下,其相容类定义为

$$[x]_{R_\lambda} = \{y \in U \mid s_A(x, y) \geq \lambda\} \quad (4)$$

$[x]_{R_\lambda}$ 是一个自反、对称的信息粒。基于 OWA 算子,利用对象相似度建立的相容关系主要受两个因素的影响^[22]: 一是判定两个对象相似度的阈值 λ 的取值;二是 OWA 算子中模糊量词参数 (α, β) 的选取。这两个因素直接影响模糊信息系统中论域不同的划分,即不同粒度的表示。模糊量词参数 (α, β) 的选取对粗糙集模型的影响已在前期工作中详细分析^[23]。在实际问题中,根据对粒度要求的粗细,可灵活选取 λ 和 (α, β) 的值,使得建立的信息粒更合理更实用。

1.2 基于 OWA 算子的粗糙集模型

在模糊决策表中,基于 OWA 算子建立的 λ 截集相容关系建立粗糙集模型。

定义 6 设 $\Omega = (U, A \cup D, V, f)$ 是一个模糊决策表, $U = \{x_1, x_2, \dots, x_n\}$ 是对象集合, $A = \{a_1, a_2, \dots, a_m\}$ 是条件属性集合, $V_A = \{V_a \mid a \in A\}$ 是属性值集, $\mu_a(x) \in [0, 1]$, $f_A: U \times A \rightarrow V_a$, 即有 $(x, a) \rightarrow \mu_a(x)$ 。 D 为决策属性集,有 $U/D = \{D_1, D_2, \dots, D_r\}$ 。

在同一模糊决策表中,约定当各条件属性的权重由 OWA 算子确定后将不再更改。因此若 $B \subseteq A$, 则相应的对象相似度定义如下。

定义 7 设 $\Omega = (U, A \cup D, V, f)$ 是一个模糊决策表,若 $B \subseteq A$, 则 $\forall x, y \in U$ 在属性集 B 上的相似度为

$$s_B(x, y) = 1 - F_A(\mathbf{T}_B) = 1 - \mathbf{H}_A(\mathbf{E}_B)^T \quad (5)$$

其中 $\mathbf{T}_B = (\mu_{a_1}(x, y), \mu_{a_2}(x, y), \dots, \mu_{a_n}(x, y))$, 若 $B \subset A$, $\exists a \in A$, 且 $a \notin B$, 则有 $\mu_a(x, y) \in \mathbf{T}_B$, 同时 $\mu_a(x, y) = 0$ 。 $\mathbf{H}_A = (\omega_1, \omega_2, \dots, \omega_m)$ 是属性集 A 中各属性对应的权重。

性质 1 在模糊决策表 $\Omega = (U, A \cup D, V, f)$ 中,若 $B \subseteq A$, $\forall x, y \in U$, 有 $s_B(x, y) \geq s_A(x, y)$ 。

证明 当 $B=A$, 则 $s_B(x, y) = s_A(x, y)$; 当 $B \subset A$, 则 $\exists a \in A, a \notin B$, 且 $\mu_a(x, y) = 0$, 则 $\mathbf{E}_B = (\mu_{a^{(1)}}(x, y), \mu_{a^{(2)}}(x, y), \dots, \mu_a(x, y)) = (\mu_{a^{(1)}}(x, y), \mu_{a^{(2)}}(x, y), \dots, 0)$ 而 $\mathbf{E}_A = (\mu_{a^{(1)}}(x, y), \mu_{a^{(2)}}(x, y), \dots, \mu_{a^{(n)}}(x, y))$, $\mu_{a^{(n)}}(x, y) \geq 0$, 因此 $\mathbf{H}_A(\mathbf{E}_B)^T \leq \mathbf{H}_A(\mathbf{E}_A)^T$, 则 $1 - \mathbf{H}_A(\mathbf{E}_B)^T \geq 1 - \mathbf{H}_A(\mathbf{E}_A)^T$, 根据定义 3 和定义 4, 即有 $s_B(x, y) \geq s_A(x, y)$ 。因此, 若 $B \subseteq A$, 有 $s_B(x, y) \geq s_A(x, y)$ 。

推论 1 在模糊决策表 $\Omega = (U, A \cup D, V, f)$ 中, 若 $B \subseteq A$, 当 λ 值相同时, $\forall x \in U$, 有 $[x]_{R_\lambda^B} \subseteq [x]_{R_\lambda^A}$ 。

证明 根据定义 5 和性质 1, 易证。

定义 8 设 $\Omega = (U, A \cup D, V, f)$ 是一个模糊决策表, $U/D = \{D_1, D_2, \dots, D_r\}$ 是根据决策属性 D 建立的等价类, 对任意 $B \subseteq A$, D_t 相对于条件属性 B 的正域、边界域和负域定义为: $\text{POS}_B(D_t) = \{x \in U \mid [x]_{R_\lambda^B} \subseteq D_t\}$; $\text{BND}_B(D_t) = \{x \in U \mid [x]_{R_\lambda^B} \cap D_t \neq \emptyset \wedge [x]_{R_\lambda^B} \not\subseteq D_t\}$; $\text{NEG}_B(D_t) = \{x \in U \mid [x]_{R_\lambda^B} \cap D_t = \emptyset\}$ 。其中, $t=1, 2, \dots, r$, $[x]_{R_\lambda^B}$ 是在条件属性集 B 下的相容类。

性质 2 在模糊决策表 $\Omega = (U, A \cup D, V, f)$ 中, 若 $B \subseteq A$, 则

- (1) $\text{POS}_B(D_t) \subseteq \text{POS}_A(D_t)$; (2) $\text{BND}_B(D_t) \supseteq \text{BND}_A(D_t)$; (3) $\text{NEG}_B(D_t) \subseteq \text{NEG}_A(D_t)$ 。

证明 当 $B \subseteq A$ 时, 根据推论 2, $\forall x \in U$, 有 $[x]_{R_\lambda^B} \subseteq [x]_{R_\lambda^A}$ 。因此, 根据三支区域的定义, 易证三支区域分别有 $\text{POS}_B(D_t) \subseteq \text{POS}_A(D_t)$, $\text{BND}_B(D_t) \supseteq \text{BND}_A(D_t)$ 和 $\text{NEG}_B(D_t) \subseteq \text{NEG}_A(D_t)$ 。

定义 8 给出了一个决策类 D_t 的三支区域定义, 进而所有决策类集合 D 相对于条件属性集 B 的三支区域系统可以定义如下。

定义 9 设 $\Omega = (U, A \cup D, V, f)$ 是一个模糊决策表, 对任意 $B \subseteq A$, D 相对于条件属性 B 的正域、边界域和负域定义为: $\text{POS}_B(D) = \bigcup_{t=1}^r \text{POS}_B(D_t)$, $\text{BND}_B(D) = \bigcup_{t=1}^r \text{BND}_B(D_t)$, $\text{NEG}_B(D) = U - \text{POS}_B(D) - \text{BND}_B(D)$ 。

性质 3 在模糊决策表 $\Omega = (U, A \cup D, V, f)$ 中, 若 $B \subseteq A$, 则

- (1) $\text{POS}_B(D) \subseteq \text{POS}_A(D)$; (2) $\text{BND}_B(D) \supseteq \text{BND}_A(D)$ 。

证明 根据性质 2 和定义 9 易证。

2 基于 OWA 算子的三支属性约简

根据 Pawlak 粗糙集理论的经典思想, 属性约简是信息系统保持分类能力不变的条件下, 去掉冗余属性。将这种思想引入到模糊决策表中, 基于三支区域(即正域、边界域和负域)给出三支属性约简如下。

定义 10 $\Omega = (U, A, V, f)$ 是一个模糊决策表, $\forall x \in U, B \subseteq A$, 有 $\text{POS}_B(D) = \text{POS}_A(D)$, 且 $\forall b \in B$, 有 $\text{POS}_{B-(b)}(D) \neq \text{POS}_B(D)$, 则称 B 是一个正域属性约简。

定义 11 $\Omega = (U, A, V, f)$ 是一个模糊决策表, $\forall x \in U, B \subseteq A$, 有 $\text{BND}_B(D) = \text{BND}_A(D)$, 且 $\forall b \in B$, 有 $\text{BND}_{B-(b)}(D) \neq \text{BND}_B(D)$, 则称 B 是一个边界域属性约简。

定义 12 $\Omega = (U, A, V, f)$ 是一个模糊决策表, $\forall x \in U, B \subseteq A$, 有 $\text{NEG}_B(D) = \text{NEG}_A(D)$, 且 $\forall b \in B$, 有 $\text{NEG}_{B-(b)}(D) \neq \text{NEG}_B(D)$, 则称 B 是一个负域属性约简。

2.1 属性贡献度与属性约简可能性

在模糊决策表 $\Omega = (U, A \cup D, V, f)$ 中, 基于 OWA 算子的相容关系 R_λ 是根据对象间的相似度 $s_A(x, y)$ 确定的, 而 $s_A(x, y)$ 是通过聚合对象 x 和 y 在每个属性 a_j 上的差异 $\mu_{a_j}(x, y)$ 而得到的。在聚合时, 每个属性 $\mu_{a_j}(x, y)$ 所对应的权重有可能不一样。直观地, 属性对应的权重越大, 对 x 和 y 的区分贡献越大, 即该属性相对越重要; 相反地, 属性对应的权重越小, 其对 x 和 y 的区分贡献越小, 即该属性的重要性相对越小。因此, 在聚合对象 x 和 y 在每个属性上的差异时, 其每个属性 a_j 对应权重的大小决

定了 a_j 在对 x 和 y 进行区分时贡献度的大小。

定义 13 设 $\Omega = (U, A \cup D, V, f)$ 是一个模糊决策表, $\forall x, y \in U$, 其每个属性的相对贡献度为

$$C_{a_j^{\sigma(l)}}(x, y) = \begin{cases} \omega_l & \mu_{a_j^{\sigma(l)}}(x, y) \neq \mu_{a_q^{\sigma(l)}}(x, y) \\ \{\omega_l, \omega_l\} & \mu_{a_j^{\sigma(l)}}(x, y) = \mu_{a_q^{\sigma(l)}}(x, y) \end{cases} \quad (6)$$

式中: $\forall a_j^{\sigma(l)} \in A, j, l, q, t \in \{1, 2, \dots, m\}; \omega_l \in [0, 1]$ 由式(2)计算得到, 是 $a_j^{\sigma(l)}$ 所对应的权重, 即 $C_{a_j^{\sigma(l)}}(x, y) \in [0, 1]$ 。

由定义 13 可知, 每个属性的相对贡献度分为两种情况: (1) 当对象 x 和 y 在 $\forall a_j^{\sigma(l)} \in A$ 上的差异都不相等时, 其贡献度的值唯一, 即 $C_{a_j^{\sigma(l)}}(x, y) = \omega_l$; (2) 当 $\exists a_j^{\sigma(l)}, a_q^{\sigma(l)} \in A$, 对象 x 和 y 有 $\mu_{a_j^{\sigma(l)}}(x, y) = \mu_{a_q^{\sigma(l)}}(x, y)$ 时, 即在两个属性上的区分差异相同, 则属性 $a_j^{\sigma(l)}$ 和 $a_q^{\sigma(l)}$ 相对于 x 和 y 的贡献度的值并不唯一, 有 $C_{a_j^{\sigma(l)}}(x, y) = C_{a_q^{\sigma(l)}}(x, y) = \{\omega_l, \omega_l\}$, 即属性 a_j 和 a_q 的相对于 x 和 y 的贡献度可以是 ω_l , 也可以是 ω_l 。

定义 14 设 $\Omega = (U, A \cup D, V, f)$ 是一个模糊决策表, 对象集 $U = \{x_1, x_2, \dots, x_n\}$, 属性集 $A = \{a_1, a_2, \dots, a_m\}$ 。 $\forall x, y \in U$, 属性集中相对于对象 x 和 y 的最小贡献度属性为

$$a_{\min}(x, y) = \{a_j^{\sigma(l)} \in A \mid \omega_{\min} \in C_{a_j^{\sigma(l)}}(x, y)\} \quad (7)$$

式中: $j, l \in \{1, 2, \dots, m\}; \omega_{\min} = \min(\omega_1, \omega_2, \dots, \omega_m)$ 是权重向量中的最小值。

性质 4 设 $\Omega = (U, A \cup D, V, f)$ 是一个模糊决策表, $\forall x, y \in U$, 有

(1) $a_{\min}(x, x) = \emptyset$; (2) $a_{\min}(x, y) = a_{\min}(y, x)$; (3) $|a_{\min}(x, y)| \geq 1$ 。

证明 在基于 OWA 算子的相容关系 R_λ 中, R_λ 满足自反性和对称性, 有 $R_\lambda(x, x) = 1, R_\lambda(x, y) = R_\lambda(y, x)$, 因此性质(1)和(2)成立。根据定义 13, $C_{a_j^{\sigma(l)}}(x, y)$ 的值有可能并不唯一, $|C_{a_j^{\sigma(l)}}(x, y)| \geq 1$, 即最小权重有可能对应两个及以上的属性, 因此 $a_{\min}(x, y)$ 所对应的属性可以并不唯一, 则性质(3)成立。

根据定义 14 可以找到任意对象 x 和 y 进行区分时的最小贡献度属性。直观地, $\forall x, y \in U$, 若属性 a 在所有 $a_{\min}(x, y)$ 中出现的频率越高, 即 a 对较大多数对象间的区分贡献度越小, 其被约掉的可能性越大; 若属性 a 在所有 $a_{\min}(x, y)$ 出现的频率越低, 即 a 对较大多数对象间的区分贡献度越大, 其被约掉的可能性也越小。因此, 属性集中各属性被约简的可能性可形式化如下。

定义 15 设 $\Omega = (U, A \cup D, V, f)$ 是一个模糊决策表, $\forall a \in A$, 则属性被约简的可能性为

$$P(a) = \frac{|M(a)|}{n \times n} \quad (8)$$

式中: $\forall x, y \in U, M(a) = \{a_{\min}(x, y) \mid a \in a_{\min}(x, y)\}$, $|M(a)|$ 为集合的势, n 为 U 中对象的个数。 $0 \leq P(a) < 1$, 由于 $a_{\min}(x, x) = \emptyset$, 且 $a_{\min}(x, y) = a_{\min}(y, x)$, 因此只统计 $a_{\min}(x, y)$, 有 $P(a) < 1$ 。

显然, $P(a)$ 的值越小, 该属性被约简的可能性越小; 反之, $P(a)$ 的值越大, 该属性被约简的可能性越大。例如: 若 $P(a) = 0$, 即 $\forall x, y \in U$, 有 $a \notin a_{\min}(x, y)$, 表示属性 a 对每个对象间区分的贡献度都较大, 显然 a 被约掉的可能性很小; 若当 $P(a)$ 越接近 1 时, $\forall x, y \in U$, 若有 $a \in a_{\min}(x, y)$, 表示属性 a 对每个对象间区分的贡献度都最小, 显然 a 是冗余的, 被约掉的可能性最大。

2.2 基于 OWA 算子的属性约简算法

直观地, 定义 15 的属性约简可能性给出了一个属性约简顺序, 即在约简中, 首先考虑约简可能性大的属性是否可以被约掉。因此, 本小节利用属性重要度进行启发, 构建基于 OWA 算子的三支属性约简算法。下面主要以正域属性约简算法(即如下算法 1)的构建为例。

算法 1 基于 OWA 算子的启发式正域属性约简算法

输入: 模糊决策表 $\Omega = (U, A \cup D, V, f)$;

输出: 模糊决策表的一个正域属性约简。

步骤 1 在 OWA 算子中,根据式(1)和(2)

- (1) 设定 α, β 的值;
- (2) 计算权重向量 $w_i (i=1, 2, \dots, m)$ 。

步骤 2 $\forall x \in U$, 相容关系的建立,根据式(3,4)和定义 8,9

- (1) 计算相似度 $s_A(x, y)$;
- (2) 设定 λ 的值,计算相容类 $[x]_{R_A^\lambda}$ 和正域 $POS_A(D)$ 。

步骤 3 $\forall x, y \in U, \forall a \in A$, 根据式(6—8)

- (1) 计算贡献度 $C_{a^{no}}(x, y)$ 及最小贡献度属性 $a_{\min}(x, y)$;
- (2) 计算每个属性的约简可能性 $P(a)$;
- (3) 根据 $P(a)$ 的值,得到启发式属性约简的顺序 $A' = (a^1, a^2, \dots, a^m)$, 其中 $P(a^1) \geq P(a^2) \geq \dots \geq P(a^m)$ 。

步骤 4 属性约简判断

- (1) 初始化 $k=1, C=\emptyset$;
- (2) $A^k = A - \{a^k\}, \forall x, y \in U$, 令 $\mu_{a^k}(x, y) = 0$, 计算 $s_{A^k}(x, y)$ 和 $POS_{A^k}(D)$;
- (3) 若 $POS_{A^k}(D) = POS_A(D)$, 则 $C = A^k$, 跳转到步骤 4(5), 否则继续;
- (4) 若 $k < m$, 则 $k = k + 1$, 跳转到步骤 4(2), 否则该信息系统无属性约简, 结束;
- (5) 若 $k < m$, 则 $k = k + 1, A = C$, 跳转到步骤 4(2); 否则, 输出 C , 结束。

基于以上步骤可得到模糊决策表 $\Omega = (U, A \cup D, V, f)$ 的一个正域属性约简。类似地,根据算法 1, 将步骤 2 计算正域 $POS_A(D)$, 分别改为计算边界域 $BND_A(D)$ 或负域 $NEG_A(D)$, 在步骤 4(3)中, 将判定条件对应改为 $BND_{A^k}(D) = BND_A(D)$ 或 $NEG_{A^k}(D) = NEG_A(D)$, 则将得到模糊决策表的一个边界域约简或负域约简。

该启发式属性约简算法在三支区域属性约简的过程中,根据属性间顺序(属性被约简的可能性)进行约简,可以减少约简的搜索空间,同时尽可能避免约简时的盲目性。

3 实例分析

下面利用 1 个模糊决策表实例,分析本文启发式属性约简方法进行三支区域属性约简的合理性及有效性。

例 1 $\Omega = (U, A \cup D, V, f)$ 是一个模糊决策表(如表 1)。其中对象集合 $U = \{x_1, x_2, \dots, x_8\}$, 条件属性集合 $A = \{a, a_2, a_3\}$, 决策属性集 $D = \{d\}, V_d = \{0, 1, 2\}$ 。

表 1 $\Omega = (U, A \cup D, V, f)$ 模糊决策表

Tab. 1 Fuzzy decision table $\Omega = (U, A \cup D, V, f)$

Ω	a_1	a_2	a_3	d
x_1	0.1	0.8	0.9	1
x_2	1.0	0.5	0.3	2
x_3	0.6	0.3	0.8	0
x_4	0.3	0.8	0.9	0
x_5	0.2	0.9	0.9	1
x_6	0.5	0.3	0.6	0
x_7	0.8	0.7	0.3	2
x_8	1.0	0.6	0.2	2

根据表 1, $U/D = \{D_1, D_2, D_3\}, D_1 = \{x_1, x_5\}, D_2 = \{x_2, x_7, x_8\}, D_3 = \{x_3, x_4, x_6\}$ 。

在算法 1 中,根据步骤 1,设 OWA 算子模糊量词参数为 $(\alpha, \beta) = (0, 0.5)$, 则 $w = \{\frac{2}{3}, \frac{1}{3}, 0\}$ 。

根据步骤 2,在属性集 A 下,以对象 x_1 和 x_2 为例,有 $\mu_{a_1}(x_1, x_2) = 0.9, \mu_{a_2}(x_1, x_2) = 0.3, \mu_{a_3}(x_1, x_2) = 0.6$, 则 x_1 和 x_2 的相似度为 $s_A(x_1, x_2) = 1 - \mathbf{H}_A(\mathbf{E}_A)^T = 1 - (\frac{2}{3}, \frac{1}{3}, 0)(0.9, 0.6, 0.3)^T = 1 - 0.8 = 0.2$ 。同理,可计算所有对象间的相似度如表 2。

表 2 相似度 $s_A(x, y)$ 数据表
Tab. 2 Similarity degree $s_A(x, y)$

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	1							
x_2	0.20	1						
x_3	0.50	0.53	1					
x_4	0.87	0.33	0.57	1				
x_5	0.90	0.27	0.47	0.90	1			
x_6	0.53	0.57	0.83	0.57	0.50	1		
x_7	0.33	0.80	0.53	0.43	0.40	0.63	1	
x_8	0.17	0.90	0.47	0.30	0.23	0.53	0.83	1

令 $\lambda = 0.85$, 则相容类有 $[x_1]_{R_1^A} = [x_4]_{R_1^A} = [x_5]_{R_1^A} = \{x_1, x_4, x_5\}, [x_2]_{R_1^A} = [x_8]_{R_1^A} = \{x_2, x_8\}, [x_3]_{R_1^A} = \{x_3\}, [x_6]_{R_1^A} = \{x_6\}, [x_7]_{R_1^A} = \{x_7\}$ 。则论域 U 在属性集 A 下被划分为: $U/A = \{\{x_1, x_4, x_5\}, \{x_2, x_8\}, \{x_3\}, \{x_6\}, \{x_7\}\}$, 根据定义 8, $POS_A(D_1) = \emptyset, POS_A(D_2) = \{x_2, x_7, x_8\}, POS_A(D_3) = \{x_3, x_6\}$, 因此,在属性集 A 下 $POS_A(D) = \{x_2, x_3, x_6, x_7, x_8\}$ 。

根据步骤 3,以对象 x_1 和 x_2 为例, a_1, a_2 和 a_3 相对于 x_1 和 x_2 的贡献度分别为 $C_{a_1^{(v)}}(x_1, x_2) = \frac{2}{3}, C_{a_2^{(v)}}(x_1, x_2) = 0, C_{a_3^{(v)}}(x_1, x_2) = \frac{1}{3}$, 显然,对于 x_1 和 x_2 的最小贡献度属性 $a_{\min}(x_1, x_2) = a_2$, 于是在属性集 A 中,关于所有对象之间的区分,其相对最小贡献度属性如表 3。

表 3 对象间最小贡献度属性 $a_{\min}(x, y)$ 数据表
Tab. 3 The minimum contribution attribute $a_{\min}(x, y)$

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	\emptyset							
x_2	a_2	\emptyset						
x_3	a_3	a_2	\emptyset					
x_4	a_2, a_3	a_2	a_3	\emptyset				
x_5	a_3	a_2	a_3	a_3	\emptyset			
x_6	a_3	a_2	a_2	a_1	a_1, a_3	\emptyset		
x_7	a_2	a_3	a_1	a_2	a_2	a_1, a_3	\emptyset	
x_8	a_2	a_1	a_2	a_2	a_2	a_2	a_2, a_3	\emptyset

进一步,根据表 3 可计算属性 a_1, a_2 和 a_3 被约简的可能性分别为 $P(a_1) = \frac{5}{64}, P(a_2) = \frac{16}{64}, P(a_3) =$

$\frac{11}{64}$ 。根据 $P(a)$ 的值,得到启发式属性约简的顺序为 a_2, a_3, a_1 。

根据步骤 4,选取 a_2 作为最有可能被约掉的属性,初始化 $k = 1$, 于是 $A^1 = A - \{a_2\} = \{a_1, a_3\}$, 仍以对象 x_1 和 x_2 为例,有 $\mu_{a_1}(x_1, x_2) = 0.9, \mu_{a_2}(x_1, x_2) = 0, \mu_{a_3}(x_1, x_2) = 0.6$, 则 x_1 和 x_2 在属性集 A^1 的相似度为

$$s_{A^1}(x_1, x_2) = 1 - \mathbf{H}_A(\mathbf{E}_{A^1})^T = 1 - (\frac{2}{3}, \frac{1}{3}, 0)(0.9, 0.6, 0)^T = 1 - 0.8 = 0.2 \quad (9)$$

则对象相似度 $s_{A^1}(x_1, x_2) = s_A(x_1, x_2)$ 。计算所有的对象在属性集 A^1 下的相似度,结果如表 4 所示。

表 4 相似度 $s_{A^1}(x, y)$ 数据表
Tab. 4 Similarity degree $s_{A^1}(x, y)$

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	1							
x_2	0.200	1						
x_3	0.633	0.533	1					
x_4	0.867	0.333	0.767	1				
x_5	0.933	0.267	0.700	0.933	1			
x_6	0.633	0.567	0.833	0.733	0.700	1		
x_7	0.333	0.867	0.600	0.433	0.400	0.700	1	
x_8	0.167	0.933	0.467	0.300	0.233	0.533	0.833	1

对比表 2 和表 4, $A^1 \subseteq A, \forall x, y \in U$, 有 $s_{A^1}(x_1, x_2) \geq s_A(x_1, x_2)$ 。当 λ 的值不变, 即 $\lambda=0.85$, 在属性集 A^1 下, 相容类为: $[x_1]_{R_i^1} = [x_4]_{R_i^1} = [x_5]_{R_i^1} = \{x_1, x_4, x_5\}, [x_2]_{R_i^1} = [x_7]_{R_i^1} = [x_8]_{R_i^1} = \{x_2, x_7, x_8\}, [x_3]_{R_i^1} = \{x_3\}, [x_6]_{R_i^1} = \{x_6\}$ 。显然 $\forall x \in U$, 有 $[x]_{R_i^1} \subseteq [x]_{R_i^0}$, 则 U 在 A^1 下被划分为: $U/A^1 = \{\{x_1, x_4, x_5\}, \{x_2, x_7, x_8\}, \{x_3\}, \{x_6\}\}$, 计算正域 $POS_{A^1}(D) = \{x_2, x_3, x_6, x_7, x_8\}$, 有 $POS_{A^1}(D) = POS_A(D)$, 则 $C = A^1 = \{a_1, a_3\}$ 。

判断 $k=1 < 3$, 继续, 令 $k=2, A^2 = A^1 - \{a_3\} = \{a_1\}$, 仍以 x_1 和 x_2 为例, 有 $\mu_{a_1}(x_1, x_2) = 0.9, \mu_{a_2}(x_1, x_2) = 0, \mu_{a_3}(x_1, x_2) = 0$, 则在属性集 A^2 下, 相似度为 $s_{A^2}(x_1, x_2) = 1 - \mathbf{H}_A(\mathbf{E}_{A^2})^T = 1 - (\frac{2}{3}, \frac{1}{3}, 0)(0.9, 0, 0)^T = 1 - 0.6 = 0.4$, 有 $s_{A^2}(x_1, x_2) \geq s_{A^1}(x_1, x_2) = s_A(x_1, x_2)$ 。计算所有的对象在属性集 A^2 下的相似度, 结果如表 5。

表 5 相似度 $s_{A^2}(x, y)$ 数据表
Tab. 5 Similarity degree $s_{A^2}(x, y)$

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	1							
x_2	0.400	1						
x_3	0.667	0.733	1					
x_4	0.867	0.533	0.800	1				
x_5	0.933	0.467	0.733	0.933	1			
x_6	0.733	0.667	0.933	0.867	0.800	1		
x_7	0.533	0.867	0.867	0.667	0.600	0.800	1	
x_8	0.400	1.00	0.733	0.533	0.467	0.667	0.867	1

当 $\lambda=0.85$, 在属性集 A^2 下, 相容类为: $[x_1]_{R_i^2} = [x_5]_{R_i^2} = \{x_1, x_4, x_5\}, [x_2]_{R_i^2} = [x_8]_{R_i^2} = \{x_2, x_7, x_8\}, [x_3]_{R_i^2} = \{x_3, x_6, x_7\}, [x_4]_{R_i^2} = \{x_1, x_4, x_5, x_6\}, [x_6]_{R_i^2} = \{x_3, x_4, x_6\}, [x_7]_{R_i^2} = \{x_2, x_3, x_7, x_8\}$ 。 $\forall x \in U$, 有 $[x]_{R_i^2} \subseteq [x]_{R_i^1} \subseteq [x]_{R_i^0}$ 。计算正域 $POS_{A^2}(D) = \{x_2, x_6, x_8\}$, 显然有 $POS_{A^2}(D) \neq POS_A(D)$, 所以, 属性集 $A^1 = \{a_1, a_3\}$ 是模糊决策表的一个属性约简。

类似地, 可以计算在属性集 A 下, 边界域和负域分别为 $BND_A(D) = \{x_1, x_4, x_5\}, NEG_A(D) = \emptyset$; 在属性集 A^1 下分别为 $BND_{A^1}(D) = \{x_1, x_4, x_5\}, NEG_{A^1}(D) = \emptyset$; 在属性集 A^2 下分别为 $BND_{A^2}(D) = \{x_1, x_3, x_4, x_5, x_7\}, NEG_{A^2}(D) = \emptyset$ 。显然, 有 $BND_{A^2} \neq BND_{A^1} = BND_A(D), NEG_{A^2}(D) = NEG_{A^1}(D) = NEG_A(D)$ 。因此, $A^1 = \{a_1, a_3\}$ 也是模糊决策表的一个边界域属性约简。因为 $NEG_{A^2}(D) = NEG_{A^1}(D) = NEG_A(D)$ 。

\emptyset , 而 $A^2 = \{a_i\}$ 为单元集, 故 A^2 为模糊决策表的一个负域约简。

根据以上实例计算和分析, 基于 OWA 算子的模糊粗糙集模型中, 利用 OWA 算子的权重定义属性的贡献度, 从而给出属性约简的顺序, 建立一种启发式约简方法, 对于模糊粗糙集模型中基于正域、负域和边界域的三支属性约简都是合理有效的。属性贡献度的定义为属性重要性度量给出了一种新的思考方法, 同时该启发式属性约简方法能减少搜索空间, 有效地寻找到分别基于三支的属性约简。

OWA 算子中的权重向量对本文启发式属性约简方法有直接影响, 而 OWA 算子中的权重向量会随着模糊量词参数 (α, β) 值的不同而变化。因此, OWA 算子中模糊量词参数 (α, β) 值的选取主要影响有两个方面: (1) 对对象相似度 $s_A(x, y)$, 相容类 $[x]_{R_i}$ 的影响, 从而影响三支区域的建立, 已在前期工作中进行了讨论^[23]。(2) 在本文启发式约简算法中, 对属性的贡献度, 属性被约简的可能性, 也即是属性约简的顺序有直接影响。

在模糊决策表中, 利用本文方法进行属性约简时, 可灵活选取 (α, β) 和 λ 的值。一般来讲, 选取 (α, β) 的值, 以保证聚合算子中权重向量 $w = (w_1, w_2, \dots, w_n)$, 有 $w_1 > w_2 > \dots > w_n$, 即在聚合对象间区分差异时, 属性值差异较大的属性对应较大的权重, 从而该属性获得较大的属性贡献度; 属性值差异较小的属性对应较小的权重, 从而该属性获得较小的属性贡献度。直观地, 这在属性约简过程中是合理的, 实例也证明了其合理性和可行性。

4 结束语

在模糊决策表中, 基于 OWA 算子的 λ 截集相容关系, 本文定义了三支区域以及基于三支区域的属性约简标准, 讨论了基于三支区域的属性约简。通过利用 OWA 算子中的权重给出每个属性的贡献度, 进一步定义了属性约简的可能性, 即给出了一个属性约简的顺序, 从而给出了模糊决策表中一种启发式三支区域属性约简方法。实例说明该约简方法可以在模糊决策表中找到基于正域、边界域和负域的一个属性约简。在启发式约简算法过程中, 利用 OWA 权重定义贡献度为属性重要性度量给出了一种新的思考方法, 同时, 该启发式约简算法在能减少属性约简的搜索空间, 更有利于模糊决策表在实际特征选取中的应用。在今后的研究工作中, 可以考虑该分别和同时基于三支区域属性约简, 它们之间的联系和相关性质, 以及利用实际应用数据来优化 OWA 算子参数的选取, 实现不同程度的属性约简。

参考文献:

- [1] Pawlak Z. Rough set[J]. *International Journal of Computer and Information Sciences*, 1982, 11: 341-356.
- [2] 张文修, 梁柏, 吴伟志. 信息系统与知识发现 [M]. 北京: 科学出版社, 2003: 96-125.
Zhang Wenxiu, Liang Yi, Wu Weizhi. Information system and knowledge discovery [M]. Beijing: Science Press, 2003: 96-125.
- [3] Dubois D, Prade H. Rough fuzzy sets and fuzzy rough sets [J]. *International Journal of General Systems*, 1990, 17: 191-209.
- [4] 胡寿松, 何亚群. 粗糙决策理论与应用 [M]. 北京: 北京航空航天大学出版社, 2006: 232-239.
Hu Shousong, He Yaqun. Theory and application of rough decision [M]. Beijing: Beihang University Press, 2006: 232-239.
- [5] 徐久成, 李涛, 孙林, 等. 基于信噪比与邻域粗糙集的特征基因选择方法[J]. *数据采集与处理*, 2015, 30 (5): 973-981.
Xu Jiucheng, Li Tao, Sun Lin, et al. Feature gene selection based on SNR and neighborhood rough set[J]. *Journal of Data Acquisition and Processing*, 2015, 30 (5): 973-981.
- [6] 崔建国, 宋博翰, 董世良, 等. 基于邻域粗糙集的航空发电机健康诊断方法[J]. *数据采集与处理*, 2012, 27(1): 80-84.
Cui Jianguo, Song Bohan, Dong Shiliang, et al. Health diagnosis of aero-generator based on neighborhood rough sets theory [J]. *Journal of Data Acquisition and Processing*, 2012, 27(1): 80-84.
- [7] 管涛, 冯博琴. 模糊目标信息系统上的知识约简方法 [J]. *软件学报*, 2004, 15(10): 1470-1478.
Guan Tao, Feng Boqin. Knowledge reduction methods in fuzzy objective information systems [J]. *Journal of Software*, 2004, 15(10): 1470-1478.
- [8] Shen Qiang, Jensen R. Selecting informative features with fuzzy-rough sets and its application for complex systems monito-

- ring [J]. Patter Recognition, 2004, 37(7): 1351-1363.
- [9] Jensen R, Shen Qiang. Fuzzy-rough attributes reduction with application to web categorization [J]. Fuzzy Sets and Systems, 2004, 141(3): 469-485.
- [10] Bhatt R B, Gopal M. On fuzzy rough sets approach to feature selection [J]. Pattern Recognition Letters, 2005, 26(7): 965-975.
- [11] 张慧哲,王坚,梅宏标. 一种变相似度的模糊粗糙集属性约简 [J]. 模式识别与人工智能, 2009, 22(3): 393-399.
Zhang Huizhe, Wang Jian, Mei Hongbiao. Attribute reduction of fuzzy rough sets based on variable similar degree [J]. Pattern Recognition and Artificial Intelligence, 2009, 22(3): 393-399.
- [12] Zeng Xuelan, Sun Xingxing, Yu Yingying. Incomplete fuzzy information system based on total order dominance relation and its attribute reduction [J]. Journal of Computer Applications, 2012, 32(5): 1303-1306.
- [13] Greco S, Matarazzo B, Slowinski R. Rough approximation by dominance relations [J]. International Journal of Intelligent Systems, 2002, 17(2): 153-171.
- [14] 胡清华, 于达仁, 谢宗霞. 基于邻域粒化和粗糙逼近的数值属性约简 [J]. 软件学报, 2008, 19(3): 640-649.
Hu Qinghua, Yu Daren, Xie Zongxia. Numerical attribute reduction based on neighborhood granulation and rough approximation [J]. Journal of Software, 2008, 19(3): 640-649.
- [15] Hu Q H, Xie Z X, Yu D R. Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation [J]. Pattern Recognition, 2007, 40(12): 3509-3521.
- [16] 张家录, 赵晓东. 基于模糊包含的粗糙集模型与模糊信任测度 [J]. 模式识别与人工智能, 2010, 23(4): 531-538.
Zhang Jialu, Zhao Xiaodong. Rough set models based on fuzzy inclusion and fuzzy belief measures [J]. Pattern Recognition and Artificial Intelligence, 2010, 23(4): 531-538.
- [17] 赵涛, 秦克云. 模糊随机信息系统及其属性约简 [J]. 计算机工程与应用, 2012, 48(22): 147-150.
Zhao Tao, Qin Keyun. Fuzzy random information system and attribute reduction [J]. Computer Engineering and Applications, 2012, 48(22): 147-150.
- [18] 黄兵, 胡作进, 周献中. 模糊信息系统知识约简的分辨函数法 [J]. 系统工程与电子技术, 2008, 30(11): 2164-2166.
Huang Bing, Hu Zuojin, Zhou Xianzhong. Knowledge reduction method for fuzzy information systems based on discernibility function [J]. Systems Engineering and Electronics, 2008, 30(11): 2164-2166.
- [19] 冯楠坪, 周磊. 一种基于相似度比较的模糊属性约简方法 [J]. 模糊系统与数学, 2014, 28(4): 164-170.
Feng Nanping, Zhou Lei. One method of fuzzy attribute reduction based on similarity degree comparison [J]. Fuzzy Systems and Mathematics, 2014, 28(4): 164-170.
- [20] Sun Bingzhen, Ma Weimin, Xiao Xia. Three-way group decision making based on multigranulation fuzzy decision-theoretic rough set over two universes [J]. International Journal of Approximate Reasoning, 2017, 81: 87-102.
- [21] Chen Yunming, Zeng Zhiqiang, Zhu Qingxin, et al. Three-way decision reduction in neighborhood systems [J]. Applied Soft Computing, 2016, 38: 942-954.
- [22] 杨霖琳, 秦克云. 模糊信息系统中一种改进的模糊相容关系 [J]. 小型微型计算机系统, 2014, 35(9): 2131-2135.
Yang Jilin, Qin Keyun. Improved fuzzy tolerance relation in the fuzzy information system [J]. Journal of Chinese Computer Systems, 2014, 35(9): 2131-2135.
- [23] 杨霖琳, 张贤勇, 唐孝. 基于三支决策的模糊信息系统 OWA 算子参数选择 [J]. 数据采集与处理, 2016, 31(6): 1156-1163.
Yang Jilin, Zhang Xianyong, Tang Xiao. Three-way decisions based parameter selection of OWA operators in fuzzy information system [J]. Journal of Data Acquisition and Processing, 2016, 31(6): 1156-1163.
- [24] Yager R. Generalized OWA aggregation operators [J]. Fuzzy Optimization and Decision Making, 2004, 3(1): 93-107.

作者简介:



杨霖琳 (1981-), 女, 博士, 副教授, 研究方向: 粗糙集、模糊集和粒计算, E-mail: yjl524@163.com.



张贤勇 (1978-), 男, 博士, 教授, 研究方向: 粗糙集、粒计算、数据挖掘。



唐孝 (1981-), 男, 博士, 副教授, 研究方向: 不确定性分析、数据挖掘。