

知识粒度视角下区间值决策信息系统的增量式属性约简

邱利芹^{1,2}, 王磊^{1,2*}, 于越^{1,2}, 孙雅慧^{1,2}

(1.江西水利电力大学信息工程学院,江西南昌330099;2.智慧水利江西省重点实验室(江西水利电力大学),江西南昌330099)

摘要:鉴于以区间值决策信息系统为对象的非增量式属性约简效率不高,现将知识粒度的概念推广到区间值决策信息系统中,从知识粒度的视角系统地研究区间值决策信息系统中的增量式属性约简方法。在区间值决策信息系统中引入相容度的概念并由此改进区间值相容度的度量方法;根据相容度确定相容关系并构造出相应的相容关系矩阵,以此得出一种基于矩阵的区间值决策信息系统中知识粒度的计算方法;探讨在对象集发生变化条件下知识粒度的更新机制,在此基础上用知识粒度来表示属性重要度,以属性重要度为启发式信息构建出增量式属性约简算法。在6个精选的UCI数据集上实施增量式属性约简算法的试验,试验结果表明:在不影响属性约简结果精度条件下,增量式属性约简方法较非增量式属性约简方法消耗时间更少,增量式属性约简方法更加高效。

关键词:属性约简;相容度;区间值决策信息系统;知识粒度;增量学习

中图分类号:TP391 **文献标志码:**A

引用格式:邱利芹,王磊,于越,等.知识粒度视角下区间值决策信息系统的增量式属性约简[J].山东大学学报(工学版),2025,55(6):45-57.

QIU Liqin, WANG Lei, YU Yue, et al. Incremental attribute reduction of interval-valued decision-making information systems from the perspective of knowledge granularity[J]. Journal of Shandong University (Engineering Science), 2025, 55(6):45-57.

Incremental attribute reduction of interval-valued decision-making information systems from the perspective of knowledge granularity

QIU Liqin^{1,2}, WANG Lei^{1,2*}, YU Yue^{1,2}, SUN Yahui^{1,2}

(1. School of Information Engineering, Jiangxi University of Water Resources and Electric Power, Nanchang 330099, Jiangxi, China; 2. Jiangxi Province Key Laboratory of Smart Water Conservancy (Jiangxi University of Water Resources and Electric Power), Nanchang 330099, Jiangxi, China)

Abstract: In view of the inefficiency of non-incremental attribute reduction in interval-valued decision-making information systems, the concept of knowledge granularity was extended to interval-valued decision-making information systems, and incremental attribute reduction in interval-valued decision-making information systems was systematically investigated from the viewpoint of knowledge granularity. The concept of tolerance degree was introduced into interval-valued decision-making information systems and the measurement method of interval-valued tolerance degree was improved; the tolerance relation was determined according to the tolerance degree and the corresponding tolerance relation matrix was constructed. This led to an calculation method of the knowledge granularity in the interval-valued decision-making information system based on the matrix; The updating mechanism of knowledge granularity under the condition of change in the object set was explored, based on which the attribute importance was represented by knowledge granularity, and an incremental attribute reduction algorithm was constructed with the attribute importance as the heuristic information. The experiments of the incremental attribute reduction algorithm were implemented on 6 selected UCI datasets. The

收稿日期:2024-05-23

基金项目:国家自然科学基金资助项目(61562061);江西省教育厅科技资助项目(GJJ211920)

第一作者简介:邱利芹(1999—),男,江西赣州人,硕士研究生,主要研究方向为粗糙集理论等。E-mail:739149323@qq.com

* 通信作者简介:王磊(1967—),男,湖北鄂州人,教授,硕士生导师,博士,主要研究方向为数据挖掘与知识发现,粗糙集与粒计算等。

E-mail:ezhoulei@163.com

experimental results demonstrated that the incremental attribute reduction method consumed less time than the non-incremental method without compromising the accuracy of the reduction results, indicating that the incremental method was more efficient.

Keywords: attribute reduction; tolerance degree; interval-valued decision-making information systems; knowledge granularity; incremental learning

0 引言

粗糙集理论^[1]是由波兰数学家 Pawlak 提出的一种处理不确定性、不精确性和模糊数据的有效数学工具,它可以挖掘隐藏的知识和揭示潜在的规则。如今,基于粗糙集理论的模型应用广泛,例如决策树模型应用医疗诊断^[2]、预测模型应用于机器学习^[3]、逻辑回归模型应用于模式识别^[4]、三支决策模型应用于决策分析^[5]等。

决策信息系统是通过等价关系来处理具有特定类别属性的单值信息系统,但由于诸多因素的影响,某些数据是以区间值的形式存在的,譬如一天的气温变化、一段时间内商品的价格等。区间值是一个取值范围而非单值,这使得区间值决策信息系统的应用场景非常丰富。为此,已有许多学者对区间值决策信息系统进行了一系列的研究,尤其是在属性约简领域。文献[6]给出不完备区间值决策信息系统中的区间值模糊最小-最大相似关系概念并基于区间值模糊可辨识对模型提出两种属性约简算法。文献[7]通过提出的相容关系定义粗糙近似值及引入 α -广义决策的概念,提出不完备区间值决策信息系统中基于熵测度与粗糙近似的两种属性约简方法。文献[8]提出一种区间值有序决策系统中基于优势关系的属性约简方法。文献[9]定义模糊 α -相似关系并提出 α -等价关系的概念,由此研究不完备区间值信息系统基于条件熵的属性约简方法。文献[10]基于二元关系研究区间值决策信息系统中提取置信度的决策规则并基于组合优化的算法提出一种区间值信息系统的属性约简方法。文献[11]基于改进的三种新条件熵及一个新的相似度,给出一种区间值决策信息系统的属性约简方法。文献[12]等通过构建 β 分布约简的差别矩阵和差别函数,提出区间值信息系统基于 β 分布的属性约简算法。文献[13]等引入区间相对知识粒度等概念,针对区间值决策信息系统,提出一种基于相对知识粒度的属性约简方法。

实际应用中区间值决策信息系统的的数据并不是一成不变的,它因新信息的收集而增加、信息的丢弃而减少,这导致属性约简的结果也随之更新。如果使用非增量式的更新方法则需要从头开始计算,造成时间浪费;而增量式算法是在原有结果的基础上进行,避免重复计算,所以对增量式属性约简算法进行系统的研究尤为重要。目前已有许多学者对区间值决策信息系统中数据动态变化情况进行深入研究。文献[14]提出一种在区间值信息系统中属性发生变化时,粗糙近似的增量机制,基于所提出的更新机制设计出相应的增量更新算法。文献[15]提出一种新的模糊邻域粗糙集模型,在区间值模糊决策信息系统中对象集变化时,构建增量更新近似机制并设计相应的更新算法。文献[16]提出 λ -模糊相似度自信息的概念,在对象集增或删时,为区间值模糊决策信息系统构建增量式属性约简方法。文献[17]给出区间值邻域关系及邻域的定义,在区间值信息系统中对象集发生变化时,提出一种极大相容类的增量更新算法。上述研究推动区间值信息系统中数据动态变化的发展,但是当对象集发生动态变化时,以矩阵为表达和运算工具,从知识粒度的视角针对区间值决策信息系统的增量式属性约简算法方面的研究还相对较少。

本研究以区间值决策信息系统为对象,研究一种对象集发生变化时基于知识粒度的增量式属性约简方法。由于区间值决策信息系统的特殊性,需要引入相容度的概念判断对象之间的关系,所以相容度的度量十分重要。为便于理解及简化计算步骤,本研究改进一种相容度量方法,考虑到论域上的相容关系不一定满足传递性,而用矩阵的方式来计算知识粒度可以有效地处理没有传递性的情况,并且当有对象集发生变化时,矩阵的变化是有规律的,利用这个规律可以更快地计算出对象变化后的知识粒度,进而提高属性约简效率,最后通过在 UCI 数据集上的测试,试验证明本研究所提增量式属性约简算法的可行性以及高效性。

1 基础概念

1.1 区间值决策信息系统

定义 1 设 X 为一个区间值,则 $X=[x^-,x^+]$,且满足 $x^+>x^-$,其中, x^- 、 x^+ 分别称为 X 的区间左端点和区间右端点,如果 $x^+=x^-$,那它就是一个单值。

定义 2^[18] 决策信息系统是一个四元组 $DIS=(U,A=C\cup D,V,f)$,其中 U 是非空有限的对象集, A 为非空有限属性集。即描述对象的所有属性的集合, C 为条件属性集, D 为决策属性集。对于任意 $a\in A$,都会存在一个 U 上所有对象的属性值集 V_a , $V=\cup V_a$ 为信息函数的值域。 $f:U\times C\cup D\rightarrow V$ 是一个信息函数,当决策信息系统中的条件属性值以区间值的形式存在且决策属性值为单值时,决策信息系统就转化成区间值决策信息系统,即区间值决策信息系统 $IVDS=(U,A=C\cup D,V,f)$ 。

决策信息系统可借助等价关系划分论域,而用等价关系划分论域并不适用于区间值决策信息系统,在该系统中只能先确定对象间是否存在相容关系,然后由相容关系构造的相容类来形成对论域的覆盖。

定义 3^[12] 设区间值决策信息系统 $IVDS=(U,A=C\cup D,V,f)$,给定 $\theta\in[0,1]$ 和属性子集 $B\subseteq C$,则相容关系为

$$T_B^\theta = \{(x_i, x_j) \mid T_{ij}^a \geq \theta, \forall a \in B, x_i, x_j \in U\}, \tag{1}$$

式中, T_B^θ 为条件属性子集 B 中的相容关系, θ 为相容度阈值, a 为属性子集 B 中的属性, x_i, x_j 为论域 U 中的任意两个对象, T_{ij}^a 为属性 a 对应的两个对象的相容度。

当 $T_{ij}^a \geq \theta$ 时,则认为对象 x_i, x_j 间存在相容关系,否则认为不存在相容关系。

由此可定义对象 $x_i \subseteq U$ 关于 B 的相容类^[12]

$$S_B^\theta(x_i) = \{x_j \mid x_j \in U, (x_i, x_j) \in T_B^\theta\}, \tag{2}$$

式中, $S_B^\theta(x_i)$ 为条件属性子集 B 中的相容类,当对象 $x_i, x_j \in T_B^\theta$ 时,即它们在条件属性子集 B 上具有相容关系时,将对象 x_i, x_j 视为同一相容类。

1.2 改进两区间值相容度定义

当决策信息系统转变成区间值决策信息系统时,等价关系已不再适用,此时需要确定两个区间值之间是否具有相容关系。而任意两个区间值之间相容程度的度量方法有多种,受文献[18]中计算两个区间值相容度方法的启示,本研究将相容度的计算方法简化成一个定义。同时在两区间值存在包含关系时,根据对较短线段长度占较长线段长度比例这一物理量的理解,进一步分析两区间值之间相容度的含义,并给出新的相容度计算方法。

定义 4 设 $A=[a^-,a^+]$, $B=[b^-,b^+]$ 为两个区间值,则两个区间值之间的相容度为

$$T_{AB} = \begin{cases} 0, & a^+ \leq b^- \text{ or } b^+ \leq a^- \\ \frac{\min\{a^+, b^+\} - \max\{a^-, b^-\}}{\max\{a^+, b^+\} - \min\{a^-, b^-\}}, & (b^+ > a^+ \text{ and } b^- < a^-) \text{ or } (a^+ > b^+ \text{ and } a^- < b^-) \\ \frac{2 \times (\min\{a^+, b^+\} - \max\{a^-, b^-\})}{(a^+ - a^-) + (b^+ - b^-)}, & (b^+ > a^- \text{ and } b^- < a^+) \text{ or } (a^+ > b^- \text{ and } a^- < b^+) \end{cases}, \tag{3}$$

式中, T_{AB} 为区间值 A, B 之间的相容度, a^-, a^+ 分别为区间值 A 的区间左端点和区间右端点, b^-, b^+ 分别为区间值为 B 的区间左端点和区间右端点。

改进的相容度定义本质上表示的是两个区间值的相似程度。由定义 4 可知,当 $a^+ \leq b^-$ or $b^+ \leq a^-$ 时,即区间值 A 与区间值 B 不存在交集时, T_{AB} 的取值为 0,而当 $a^- = b^-$ 且 $a^+ = b^+$ 时,即区间值 A 与区间值 B 是完全相同的两个区间值时, T_{AB} 的取值为 1,所以 T_{AB} 区间为 $[0, 1]$ 。 T_{AB} 越大,则表明区间值 A 与区间值 B 相似程度越高。

特别地,对以下三种特殊的情形进行讨论与分析。

(1) 当 $a^+ = b^-$ or $b^+ = a^-$,即区间值 A 与区间值 B 的交集为一个点时,根据定义 1 可知 $A \cap B$ 为一个单值,由定义 4 可知, $T_{AB} = 0$,表明这两个区间值完全不相似。

(2) 当 $(b^+ > a^+ \text{ 且 } b^- < a^-)$ 或 $(a^+ > b^+ \text{ 且 } a^- < b^-)$ 时, 即 $B \supseteq A$ 或 $A \supseteq B$ 时, $A \cap B$ 为 A 或 B , 由定义 4 可知,

$T_{AB} = \frac{\min\{a^+, b^+\} - \max\{a^-, b^-\}}{\max\{a^+, b^+\} - \min\{a^-, b^-\}}$, 表示区间值 A 长度占区间值 B 长度的比例或区间值 B 长度占区间值 A 长度的比例。

(3) 当区间值 A 与区间值 B 的区间左右端点均相等时, 即 A 与 B 是相同的两个区间值, 此时 $a^- = b^-$ 且 $a^+ = b^+$, 有 $A \cap B = [a^-, a^+] = [b^-, b^+]$, 根据定义 4 可知, $T_{AB} = 1$, 表示 A 与 B 完全相同。

这三种情形均与实际一致。

例 1 设 $A = [-3, 3]$, $B = [-7, 7]$, 试求 A 与 B 的相容度。

解: 显然 $A \cap B = A$, 根据定义 4 可知

$$T_{AB} = \frac{\min\{a^+, b^+\} - \max\{a^-, b^-\}}{\max\{a^+, b^+\} - \min\{a^-, b^-\}} = \frac{3}{7}。$$

根据定义 4 可以计算出两区间值之间的相容度, 从而可以确定它们是否存在相容关系。进而通过相容关系在区间值决策信息系统中构造出相应的矩阵。

1.3 基于知识粒度的启发式属性约简方法

由于使用矩阵的方法计算知识粒度可以处理区间值决策信息系统中相容关系不一定满足传递性的情况, 且当对象集发生变化时, 根据矩阵的变化规律, 可以得出知识粒度的更新机制, 进而提高属性约简的效率。因此本小节将文献[19]中知识粒度的矩阵定义及计算方法推广到区间值决策信息系统中, 并给出区间值决策信息系统中相容关系矩阵的定义, 以此计算知识粒度。

定义 5 设区间值决策信息系统 $IVDS = (U, A = C \cup D, V, f)$, $U = \{x_1, x_2, \dots, x_n\}$, 对于 $\forall B \subseteq C$, 给定 $\theta \in [0, 1]$, 则相容关系矩阵表示为 $(M_U^{T_B^\theta})_{n \times n}$, m_{ij} 表示矩阵 $(M_U^{T_B^\theta})_{n \times n}$ 的第 i 行、第 j 列元素, 且 $1 \leq i \leq n$, $1 \leq j \leq n$, 那么 m_{ij} 为

$$m_{ij} = \begin{cases} 1, & T_{ij}^b \geq \theta, \forall b \in B \\ 0, & T_{ij}^b < \theta, \exists b \in B \end{cases} \quad (4)$$

式中, T_{ij}^b 为两对象间的相容度。

定义 6 设区间值决策信息系统 $IVDS = (U, A = C \cup D, V, f)$, $U = \{x_1, x_2, \dots, x_n\}$, $\forall B \subseteq C$, $\forall b \in B$, 即 B 为条件属性集 C 的任意子集, 给定 $\theta \in [0, 1]$, 决策属性 $D = \{d\}$, 则相容关系矩阵为 $(M_U^{T_{B \cup D}^\theta})_{n \times n}$, M_{ij} 表示矩阵 $(M_U^{T_{B \cup D}^\theta})_{n \times n}$ 的第 i 行、第 j 列元素, 且 $1 \leq i \leq n$, $1 \leq j \leq n$, 那么 M_{ij} 为

$$M_{ij} = \begin{cases} 1, & T_{ij}^b \geq \theta \text{ 且 } f(x_i, d) = f(x_j, d) \\ 0, & T_{ij}^b < \theta \text{ 或 } f(x_i, d) \neq f(x_j, d) \end{cases} \quad (5)$$

式中, b 为条件属性子集 B 中的任意属性。

定义 7 设区间值决策信息系统 $IVDS = (U, A = C \cup D, V, f)$, $\forall B \subseteq C$, 给定 $\theta \in [0, 1]$, 那么 B 在 U 上的知识粒度为

$$K_U^\theta(B) = \frac{\sum_{i=1}^n \sum_{j=1}^n m_{ij}}{|U|^2} = \overline{(M_U^{T_B^\theta})_{n \times n}} \quad (6)$$

式中, $\overline{(M_U^{T_B^\theta})_{n \times n}}$ 为相容关系矩阵 $(M_U^{T_B^\theta})_{n \times n}$ 算术平均值。

定义 8 设区间值决策信息系统 $IVDS = (U, A = C \cup D, V, f)$, $\forall B \subseteq C$, 给定 $\theta \in [0, 1]$, 则决策属性 D 关于条件属性 B 的相对知识粒度为

$$K_U^\theta(D|B) = K_U^\theta(B) - K_U^\theta(B \cup D) = \overline{(M_U^{T_B^\theta})_{n \times n}} - \overline{(M_U^{T_{B \cup D}^\theta})_{n \times n}} \quad (7)$$

定义 9 设区间值决策信息系统 $IVDS = (U, A = C \cup D, V, f)$, $\forall B \subseteq C$, $\forall a \in B$, 给定 $\theta \in [0, 1]$, 对于属性 a 相对于属性集 B 的内部属性重要度为

$$S_U^\theta(a, B, D) = K_U^\theta(D|B - \{a\}) - K_U^\theta(D|B) \quad (8)$$

定义 10 设区间值决策信息系统 $IVDS = (U, A = C \cup D, V, f)$, $\forall B \subseteq C$, $\forall a \in C - B$, 给定相容度 $\theta \in$

$[0, 1]$, 对于属性 a 相对于属性集 B 的外部重要度定义为

$$S_U^\theta(a, B, D) = K_U^\theta(D|B) - K_U^\theta(D|B \cup \{a\})。 \quad (9)$$

定义 11 已知区间值决策信息系统 $IVDS = (U, A = C \cup D, V, f)$, 对于 $\forall B \subseteq C$, 属性集 B 是系统的相对属性约简当且仅当满足以下两个条件:

- (1) $K_U^\theta(D|B) = K_U^\theta(D|C)$;
- (2) $\forall a \in B, K_U^\theta(D|B - \{a\}) \neq K_U^\theta(D|B)$ 。

根据定义 11 的两个条件给出区间值决策信息系统中用知识粒度表示属性重要度, 进而以属性重要度为启发式信息构建属性约简算法, 该算法的具体描述见算法 1。

算法 1

输入: $IVDS = (U, A = C \cup D, V, f)$, 阈值 θ 。

输出: 一个属性约简 R_U 。

步骤 1 初始化 $R_U \leftarrow \emptyset$ 。

步骤 2 对于 $\forall a_i \in C$, 计算内部重要度 $S_U^l(a_i, B, D)$, 如果 $S_U^l(a_i, B, D) > 0$, 则将其存入 R_U 中, 即: $R_U \leftarrow R_U \cup \{a_i\}$ 。

步骤 3 $B \leftarrow R_U$ 。

步骤 4 计算相对知识粒度 $K_U^\theta(D|B)$ 和 $K_U^\theta(D|C)$, 如果相等, 则跳转到步骤 9; 否则跳转到步骤 5。

步骤 5 对于 $\forall a \in (C - B)$, 计算 $S_U^\theta(a, B, D)$, 依次取 $a = \max \{S_U^\theta(a, B, D)\}$ 添加到属性集 B 中。

步骤 6 $B \leftarrow B \cup \{a\}$ 。

步骤 7 计算更新后的 $K_U^\theta(D|B)$ 和 $K_U^\theta(D|C)$, 如果 $K_U^\theta(D|B) = K_U^\theta(D|C)$, 则转到步骤 8; 否则继续循环直到满足条件 $K_U^\theta(D|B) = K_U^\theta(D|C)$ 为止。

步骤 8 向前推移冗余属性, 对于 $\forall a_j \in B$, 如果 $K_U^\theta(D|B - \{a_j\}) = K_U^\theta(D|C)$, 则 a_j 为冗余属性, 需要将其删除; 否则 B 保持不变。

步骤 9 $R_U \leftarrow B$, 返回属性约简结果 R_U 。

2 对象集变化时增量式属性约简

2.1 向论域添加对象时知识粒度的更新机制

为了提高属性约简更新的效率, 本研究提出一种当对象集发生变化时, 用知识粒度表示属性重要度, 进而以属性重要度为启发信息的区间值决策信息系统增量式属性约简。

当有新对象集添加至区间值决策信息系统中时, 相容关系矩阵会发生变化, 此时相容关系矩阵可以分成四块区域, 而本研究将增加对象集后相容关系矩阵增加的三个区域构造造成一个新的相容关系矩阵, 以此对知识粒度进行更新。

定义 12 设区间值决策信息系统 $IVDS = (U, A = C \cup D, V, f)$, $B \subseteq C$, $U = \{x_1, x_2, \dots, x_n\}$, 假设新添加的对象集为 $\Delta U = \{Y_1, Y_2, \dots, Y_m\}$, 给定 $\theta \in [0, 1]$, T_C^θ 是集合 $U \cup \Delta U$ 上的相容关系, 则增量关系矩阵表示为 $(M_{U \cup \Delta U}^{T_C^\theta})_{m \times (n+m)}$, a_{ij} 表示矩阵 $(M_{U \cup \Delta U}^{T_C^\theta})_{m \times (n+m)}$ 的第 i 行、第 j 列元素, 且 $1 \leq i \leq m, 1 \leq j \leq n+m$, 则 a_{ij} 为

$$a_{ij} = \begin{cases} 2, & (y_i, x_j) \in T_C^\theta, 1 \leq i \leq m, 1 \leq j \leq n \\ 1, & (y_i, y_{(j-n)}) \in T_C^\theta, 1 \leq i \leq m, n \leq j \leq n+m \\ 0, & \text{其他} \end{cases}。 \quad (10)$$

即当增加的对象集为 ΔU 时, 对象集更新为 $U \cup \Delta U$, 则此时相容关系矩阵为 $(M_{U \cup \Delta U}^{T_C^\theta})_{(n+m) \times (n+m)}$, 且

$(M_{U \cup \Delta U}^{T_C^\theta})_{(n+m) \times (n+m)} = \begin{bmatrix} (M_U^{T_C^\theta})_{n \times n}, & (M_{U \cup \Delta U}^{T_C^\theta})_{n \times m}^T \\ (M_{U \cup \Delta U}^{T_C^\theta})_{m \times n}, & (M_{\Delta U}^{T_C^\theta})_{m \times m} \end{bmatrix}$, 其中 $(M_U^{T_C^\theta})_{n \times n}$ 为增加对象集前所构成的相容关系矩阵, 它是已知的, 因此只需要求出另外三个矩阵, 而本研究将这三个矩阵构造造成一个 $m \times (n+m)$ 阶矩阵, 即 $(M_{U \cup \Delta U}^{T_C^\theta})_{m \times (n+m)}$, 进而对知识粒度进行更新。显然 $(M_{U \cup \Delta U}^{T_C^\theta})_{m \times n}^T$ 为 $(M_{U \cup \Delta U}^{T_C^\theta})_{n \times m}$ 的转置矩阵, 因此当增加的对象

与原来的对象存在相容关系时,本研究将 $m \times (n+m)$ 阶矩阵值取 2, 此时 a_{ij} 取值为 2。当增加的对象集之间存在相容关系时,将 $m \times (n+m)$ 阶矩阵值取 1, 此时 a_{ij} 取值为 1。

定理 1 设区间值决策信息系统 $IVDS = (U, A = C \cup D, V, f)$, 且 $\Delta U = \{Y_1, Y_2, \dots, Y_m\}$ 是新添加的对象集, 给定 $\theta \in [0, 1]$, 则增加对象集 ΔU 后的知识粒度更新为

$$K_{U \cup \Delta U}^\theta(B) = \frac{\sum_{i=1}^n \sum_{j=1}^n m_{ij} + \sum_{i=1}^m \sum_{j=1}^{n+m} a_{ij}}{|U \cup \Delta U|^2}, \quad (11)$$

式中: m_{ij} 为增加对象集前所构成的相容关系矩阵的元素, 即 $(M_U^{\theta})_{n \times n}$ 中的元素; a_{ij} 为增量关系矩阵中的元素, 即 $(M_{U \cup \Delta U}^{\theta})_{m \times (n+m)}$ 中的元素。

证明 令 $U = U \cup \Delta U$, 则 $K_U^\theta(B) = \frac{\sum_{i=1}^{n+m} \sum_{j=1}^{n+m} m_{ij}}{|U \cup \Delta U|^2}$, 将 $(n+m) \times (n+m)$ 矩阵分块为 4 个子矩阵求和 $(n+m) \times (n+m) = n^2 + m \times n + n \times m + m^2$, 即 $(M_{U \cup \Delta U}^{\theta})_{(n+m) \times (n+m)} = (M_U^{\theta})_{n \times n} + (M_{U \cup \Delta U}^{\theta})_{m \times n} + (M_{U \cup \Delta U}^{\theta})_{n \times m} + (M_{U \cup \Delta U}^{\theta})_{m \times m}$, 又因为 $m \times (n+m) = m \times n + m^2$ 。由定义 12 知, 第一种条件的矩阵值为 2, 因此 $m \times (n+m) = 2 \times m \times n + m^2 = m \times n + n \times m + m^2$, 即 $(M_{U \cup \Delta U}^{\theta})_{m \times (n+m)} = (M_{U \cup \Delta U}^{\theta})_{m \times n} + (M_{U \cup \Delta U}^{\theta})_{n \times m} + (M_{U \cup \Delta U}^{\theta})_{m \times m}$, 所以 $K_{U \cup \Delta U}^\theta(B) = \frac{\sum_{i=1}^n \sum_{j=1}^n m_{ij} + \sum_{i=1}^m \sum_{j=1}^{n+m} a_{ij}}{|U \cup \Delta U|^2}$ 。

定理 2 设区间值决策信息系统 $IVDS = (U, A = C \cup D, V, f)$, 新增加的对象集为 ΔU , 给定 $\theta \in [0, 1]$, 则增加对象集 ΔU 后相对知识粒度更新为

$$K_{U \cup \Delta U}^\theta(D|B) = K_{U \cup \Delta U}^\theta(B) - K_{U \cup \Delta U}^\theta(B \cup D)。 \quad (12)$$

证明:

$$K_{U \cup \Delta U}^\theta(D|B) = (M_{U \cup \Delta U}^{\theta})_{m \times (n+m)} - (M_{U \cup \Delta U}^{\theta})_{m \times (n+m)} = \frac{\sum_{i=1}^n \sum_{j=1}^n m_{ij} + \sum_{i=1}^m \sum_{j=1}^{n+m} a_{ij}}{|U \cup \Delta U|^2} - \frac{\sum_{i=1}^n \sum_{j=1}^n M_{ij} + \sum_{i=1}^m \sum_{j=1}^{n+m} A_{ij}}{|U \cup \Delta U|^2} = K_{U \cup \Delta U}^\theta(B) - K_{U \cup \Delta U}^\theta(B \cup D)$$

式中, m_{ij} 为矩阵 $(M_U^{\theta})_{n \times n}$ 的元素, M_{ij} 为矩阵 $(M_U^{\theta})_{n \times n}$ 的元素, a_{ij} 为矩阵 $(M_{U \cup \Delta U}^{\theta})_{m \times (n+m)}$ 的元素, A_{ij} 为矩阵 $(M_{U \cup \Delta U}^{\theta})_{m \times (n+m)}$ 的元素

定理 2 得证。

2.2 从论域中删除对象时知识粒度的更新机制

定义 13 设区间值决策信息系统 $IVDS = (U, A = C \cup D, V, f)$, $B \subseteq C$, $U = \{x_1, x_2, \dots, x_n\}$, 假设删除的对象集为 $\Delta U = \{Y_1, Y_2, \dots, Y_m\}$, 在给定 $\theta \in [0, 1]$ 的情况下, T_C^θ 是集合 U 上的相容关系, 则增量关系矩阵 $(M_{U-\Delta U}^{\theta})_{m \times (n+m)}$, 其中 d_{ij} 表示矩阵 $(M_{U-\Delta U}^{\theta})_{m \times (n+m)}$ 第 i 行、第 j 列元素, 且 $1 \leq i \leq m$, $1 \leq j \leq n+m$, d_{ij} 表示为

$$d_{ij} = \begin{cases} 2, & (y_i, x_j) \in T_C^\theta, 1 \leq i \leq m, 1 \leq j \leq n \\ 1, & (y_i, y_{(j-n)}) \in T_C^\theta, 1 \leq i \leq m, n \leq j \leq n+m \\ 0, & \text{其他} \end{cases} \quad (13)$$

当有对象集 ΔU 从论域中删除时, 对象集更新为 $U - \Delta U$, 则此时相容关系矩阵为 $(M_{U-\Delta U}^{\theta})_{(n-m) \times (n-m)}$ 删除

对象集前的相容关系矩阵为 $(M_U^{\theta})_{n \times n} = \begin{bmatrix} (M_{U-\Delta U}^{\theta})_{(n-m) \times (n-m)}, & (M_{U-\Delta U}^{\theta})_{m \times (n-m)}^T \\ (M_{U-\Delta U}^{\theta})_{(n-m) \times m}, & (M_{U-\Delta U}^{\theta})_{m \times m} \end{bmatrix}$, 它是已知的。由于用非增量

式直接计算 $(M_{U-\Delta U}^{\theta})_{(n-m) \times (n-m)}$ 效率很低, 所以本研究通过计算另外三个矩阵, 将它们合成一个矩阵 $(M_{U-\Delta U}^{\theta})_{m \times (n+m)}$ 之后, 间接计算出 $(M_{U-\Delta U}^{\theta})_{(n-m) \times (n-m)}$, 进而再对知识粒度更新。

定理 3 设区间值决策信息系统 $IVDS = (U, A = C \cup D, V, f)$, ΔU 是删除的对象集,且 $\Delta U = \{Y_1, Y_2, \dots, Y_m\}$, 在给定相容度 $\theta \in [0, 1]$ 的情况下,则删除对象后 B 在 U 上的知识粒度更新为

$$K_{U-\Delta U}^\theta(B) = \frac{\sum_{i=1}^n \sum_{j=1}^n m_{ij} - \sum_{i=1}^m \sum_{j=1}^{n+m} d_{ij}}{|U - \Delta U|^2}, \quad (14)$$

式中: m_{ij} 为删除对象集前所构成的相容关系矩阵的元素,即 $(M_U^{\theta})_{n \times n}$ 中元素; d_{ij} 为增量关系矩阵中的元素,即 $(M_{U-\Delta U}^{\theta})_{m \times (n+m)}$ 中元素。

证明 令 $U = U - \Delta U$, 则 $K_U^\theta(B) = K_{U-\Delta U}^\theta(B) = \frac{\sum_{i=1}^{n-m} \sum_{j=1}^{n-m} m_{ij}}{|U - \Delta U|^2}$, 而由定义 13 可知: $\Delta \text{Sum}(M_{U-\Delta U}^{\theta}) = (M_{U-\Delta U}^{\theta})_{m \times (n+m)}$, 即矩阵变化的值为 $\sum_{i=1}^m \sum_{j=1}^{n+m} d_{ij}$, 而原本的矩阵值为 $\sum_{i=1}^n \sum_{j=1}^n m_{ij}$, 故 $K_{U-\Delta U}^\theta(B) = \frac{\sum_{i=1}^n \sum_{j=1}^n m_{ij} - \sum_{i=1}^m \sum_{j=1}^{n+m} d_{ij}}{|U - \Delta U|^2}$ 。

定理 4 设区间值决策信息系统 $IVDS = (U, A = C \cup D, V, f)$, 待删除的对象集为 ΔU , 且 $\Delta U = \{Y_1, Y_2, \dots, Y_m\}$, 在给定相容度 $\theta \in [0, 1]$ 的情况下,可得

$$K_{U-\Delta U}^\theta(D|B) = K_{U-\Delta U}^\theta(B) - K_{U-\Delta U}^\theta(B \cup D). \quad (15)$$

证明 定理 4 的证明与定理 2 相类似。

在区间值决策信息系统中的对象集发生变化时,分别构建了以下两个增量式属性约简算法,算法 2 是增加区间值决策信息系统中的对象集时,以知识粒度表示属性重要度,进而以属性重要度为启发式信息的增量属性约简算法。

算法 2

输入: $IVDS = (U, A = C \cup D, V, f)$, 增加对象集前的属性约简 R_U , 新增加的对象集 ΔU , 阈值 θ 。

输出: 增加对象集后的属性约简 $R_{U \cup \Delta U}$ 。

步骤 1 初始化 $B \leftarrow R_U$ 。

步骤 2 计算增量关系矩阵 $(M_{U \cup \Delta U}^{\theta})_{m \times (n+m)}$ 、 $(M_{U \cup \Delta U}^{\theta})_{m \times (n+m)}$ 、 $(M_{U \cup \Delta U}^{\theta})_{m \times (n+m)}$ 、 $(M_{U \cup \Delta U}^{\theta})_{m \times (n+m)}$ 。

步骤 3 计算相对知识粒度 $K_{U \cup \Delta U}^\theta(D|B)$ 和 $K_{U \cup \Delta U}^\theta(D|C)$, 如果 $K_{U \cup \Delta U}^\theta(D|B) = K_{U \cup \Delta U}^\theta(\theta; D|C)$, 则跳转到步骤 7。

步骤 4 对于 $\forall a \in (C - B)$, 计算 $S_{U \cup \Delta U}^O(a, B, D)$, 依次取 $a = \max \{S_{U \cup \Delta U}^O(a, B, D)\}$ 添加到属性集 B 中。即 $B \leftarrow B \cup \{a_i\}$ 。

步骤 5 计算更新后的相对知识粒度 $K_{U \cup \Delta U}^\theta(D|B)$, 如果 $K_{U \cup \Delta U}^\theta(D|B) = K_{U \cup \Delta U}^\theta(\theta; D|C)$, 则转到步骤 6; 否则继续循环直到满足 $K_{U \cup \Delta U}^\theta(D|B) = K_{U \cup \Delta U}^\theta(\theta; D|C)$ 为止。

步骤 6 向前推移冗余属性, 对于 $\forall a_j \in B$, 如果 $K_{U \cup \Delta U}^\theta(D|B - \{a_j\}) = K_{U \cup \Delta U}^\theta(D|C)$, 则 a_j 为冗余属性, 需要将它删除; 否则 B 保持不变。

步骤 7 $R_{U \cup \Delta U} \leftarrow B$ 。

算法 3 是当区间值决策信息系统中有对象集删除时,以知识粒度表示属性重要度,进而以属性重要度为启发式信息的增量属性约简算法。

算法 3

输入: $IVDS = (U, A = C \cup D, V, f)$, 增加对象前的属性约简 R_U , 待删除的对象集 ΔU , 阈值 θ 。

输出: 删除对象集后的属性约简 $R_{U-\Delta U}$ 。

步骤 1 初始化 $B \leftarrow R_U$ 。

步骤 2 计算 $K_{U-\Delta U}^\theta(D|C)$ 。

步骤 3 对于 $\forall a_i \in B$, 计算 $K_{U-\Delta U}^\theta(D|B - \{a_i\})$, 如 $K_{U-\Delta U}^\theta(D|B - \{a_i\}) = K_{U-\Delta U}^\theta(D|C)$, 那么 $B \leftarrow (B -$

$\{a_i\}$), 否则继续执行步骤 3。

步骤 4 $R_{U-\Delta U} \leftarrow B$ 。

步骤 5 return $R_{U-\Delta U}$ 。

2.3 分析与比较算法时间复杂度

表 1 给出了对象集增加时算法 1 与算法 2 的时间复杂度对比。

表 1 对象增加时两种算法的时间复杂度比较
Table 1 Comparison of time complexity of two algorithms for object increase

算法名称	时间复杂度
算法 1	$O(C ^2 U+\Delta U ^2+n \cdot C U+\Delta U ^2+m^2 \cdot U+\Delta U ^2)$
算法 2	$O(k \cdot n \cdot C \Delta U \Delta U+U +m^2 \cdot \Delta U \Delta U+U)$

表 2 给出了对象集删除时算法 1 与算法 3 的时间复杂度对比。

表 2 对象删除时两种算法的时间复杂度比较
Table 2 Time complexity comparison of two algorithms for object deletion

算法名称	时间复杂度
算法 1	$O(C ^2 U-\Delta U ^2+n \cdot C U-\Delta U ^2+m^2 \cdot U-\Delta U ^2)$
算法 3	$O(C \Delta U U+\Delta U +m^2 \cdot \Delta U U+\Delta U)$

算法 1 中, 步骤 1~4 时间复杂度为 $O(|C| |U|^2+|C|^2 |U|^2+n \cdot |U|^2)$, 步骤 5~7 的时间复杂度为 $O(n \cdot |C| |U|^2+k \cdot n \cdot |U|^2)$, 步骤 8~9 的时间复杂度为 $O(m^2 \cdot |U|^2)$, 所以非增量式方法总的时间复杂度为 $O(|C|^2 |U|^2+n \cdot |C| |U|^2+m^2 \cdot |U|^2)$ 。其中 n 为原属性约简结果的个数, k 为去除核属性之后的条件属性个数, m 为去除冗余前属性集中属性的个数。

当有新对象添加至论域中时, 算法 2 的时间复杂度分析如下: 步骤 1~3 求更新后的相对知识粒度, 时间复杂度为 $O(|C| |\Delta U| |\Delta U+U|)$; 步骤 4~5 是将每次迭代后最重要的外部属性, 即对整个区间值决策信息系统影响最大的外部属性加入到约简集中, 其时间复杂度为 $O(n \cdot |C| |\Delta U| |\Delta U+U|+k \cdot n \cdot |\Delta U| |\Delta U+U|)$; 步骤 6~7 是判断约简结果中是否有冗余的时间复杂度为 $O(m^2 \cdot |\Delta U| |\Delta U+U|)$, 所以算法 2 总的时间复杂度为 $O(k \cdot n \cdot |C| |\Delta U| |\Delta U+U|+m^2 \cdot |\Delta U| |\Delta U+U|)$ 。其中 n 为原属性约简结果的个数, k 为去除核属性之后的条件属性个数, m 为去除冗余前属性集中属性的个数。

算法 3 中, 步骤 1~2 计算删除对象集后的相对知识粒度的时间复杂度为 $O(|C| |\Delta U| |U+\Delta U|)$, 步骤 3~5 判断是否有冗余的时间复杂度为 $O(m^2 \cdot |\Delta U| |U+\Delta U|)$, 所以算法 3 总的时间复杂度为 $O(|C| |\Delta U| |U+\Delta U|+m^2 \cdot |\Delta U| |U+\Delta U|)$ 。其中, m 为去除冗余前属性集中属性的个数。

通过上述算法时间复杂度的分析, 可清晰地看出增量式约简算法的时间复杂度比非增量约简算法的时间复杂度低, 由此可知增量式属性约简算法较非增量式属性约简算法可节省大量的时间。

3 试验分析

为了证实算法 2、算法 3 的高效性和准确性, 本研究从 UCI 机器学习数据库 (UCI Machine Learning Repository) 中选取了 6 个 UCI 数据集进行算法性能测试。

3.1 试验数据集

数据集的选取原则: 选取的 6 个 UCI 数据集的属性值均为单值且无缺失值, 数据集是一致的。各个 UCI 数据集的相关信息见表 3。

表 3 试验所使用的 UCI 数据集
Table 3 Datasets used in the experiment

序号	数据集名称	对象数目	属性数目	类别数目
1	威斯康星州乳腺癌(诊断)	569	30	2
2	孕产妇健康风险	1 014	6	3
3	避孕方法选择	1 473	9	3
4	根据饮食习惯和身体状况估计肥胖水平	2 111	16	6
5	大米(物种 Cammeo 和物种 Osmancik)	3 810	7	2
6	文档页面的区块分类	5 473	10	5

硬件环境:处理器 Intel(R) Core(TM) i7-8750H CPU @ 2.20 GHz 主频 2.21 GHz;内存 16.0 GB。
软件平台:PyCharm Community Edition 2023.1.2。

3.2 试验方案

3.2.1 数据的预处理

由于原始的 UCI 数据集的属性值属于非区间值类型,故本研究提出了一种将单值数据转换成区间值数据的方法,具体转化步骤如图 1 所示。

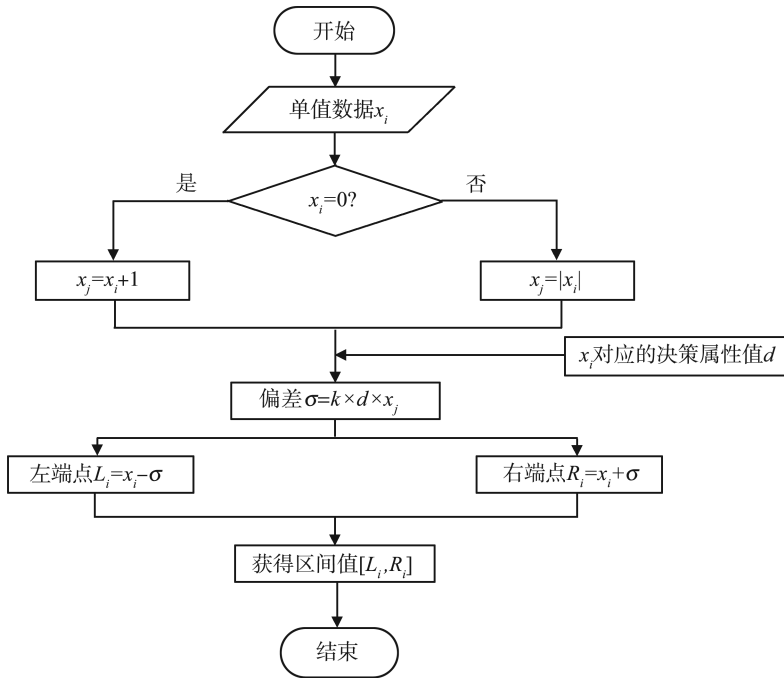


图 1 单值转换成区间值的流程图

Fig.1 Flowchart for converting single values to interval values

图 1 中, x_i 为待转换的单值数据, x_j 为根据 x_i 值所设置的变量, L_i 为将 x_i 转换后的区间值的左端点, R_i 为将 x_i 转换后的区间值的右端点, d 为 x_i 对应的决策属性值, k 为常数。

偏差 σ 的取值原则如下:为了使转换后的数据更加客观,偏差 σ 不能取为固定值,偏差 σ 需要根据不同的 x_i 和 d 取不同的值;考虑到转换后的区间值不能为 0 以及区间左端点值不能大于区间右端点值的情况,所以在 x_i 为 0 时,将变量 x_j 赋值为 1,而在 x_i 不为 0 时,将 x_i 的绝对值赋值给变量 x_j ;为了避免出现左端点出现负数而影响转换后数据集的客观性, σ 的取值不能超过 x_i ,故有 $k * d < 1$ 这个限制条件,根据所选 UCI 数据集的决策类数目,常数项 k 取值为 0.1。

通过上述方法可将单值数据转换成区间值数据,从而获得了区间值数据集。

3.2.2 非增量式属性约简方法与增量式属性约简方法的比较

为了非增量式属性约简与增量式属性约简作直观的比较,本研究拟从实施这两种算法程序的时间消耗和所得属性约简结果两个方面探讨。为了比较两种方法的运行时间,设计了以下的方案。

(1)将预处理后的数据集分成两等份,第一部分作为基础数据集使用,再将第二部分平均分为 5 份,依次将每份增加至基础数据集中,直到第二部分的数据取完,记录算法 1 和算法 2 每次运行的时间和属性约简结果。

(2)选取预处理后的数据集的 10%作为待删除的数据集使用,且按 10%的比例依次递增,直到取完整个数据集的 50%,记录算法 1 和算法 3 每次运行的时间和属性约简结果。

3.3 试验结果及分析

文献[20]将 θ 值取为 0.6,6 个 UCI 数据集在对象集增加情况下,增量式属性约简方法与非增量式属性约简方法所消耗的时间对比如图 2 所示,6 个 UCI 数据集在有对象集删除时,增量式与非增量式属性约简所消耗的时间对比如图 3 所示。

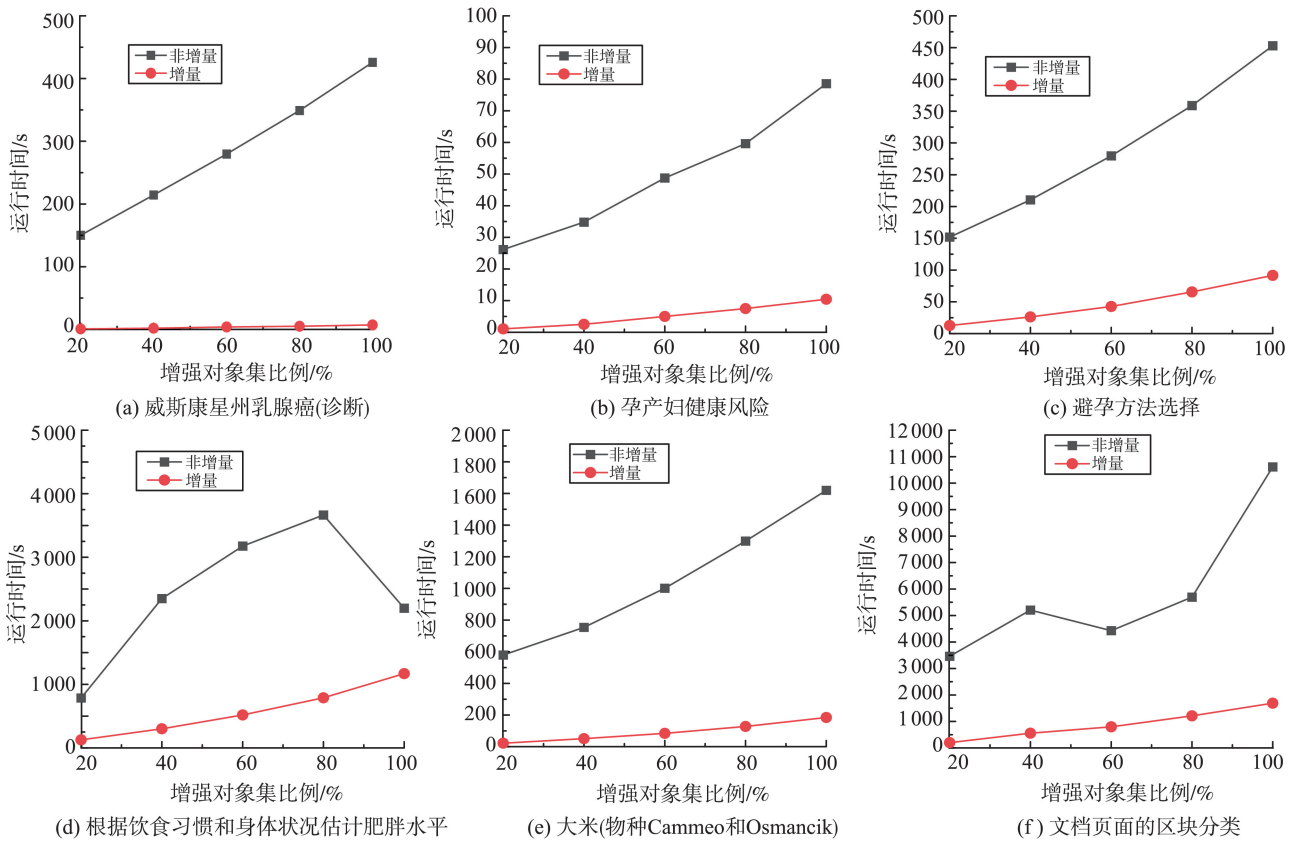


图2 两种算法在数据集上每增加一组数据的时间比较

Fig.2 Comparison of the time taken by the two algorithms for each additional set of data on the dataset

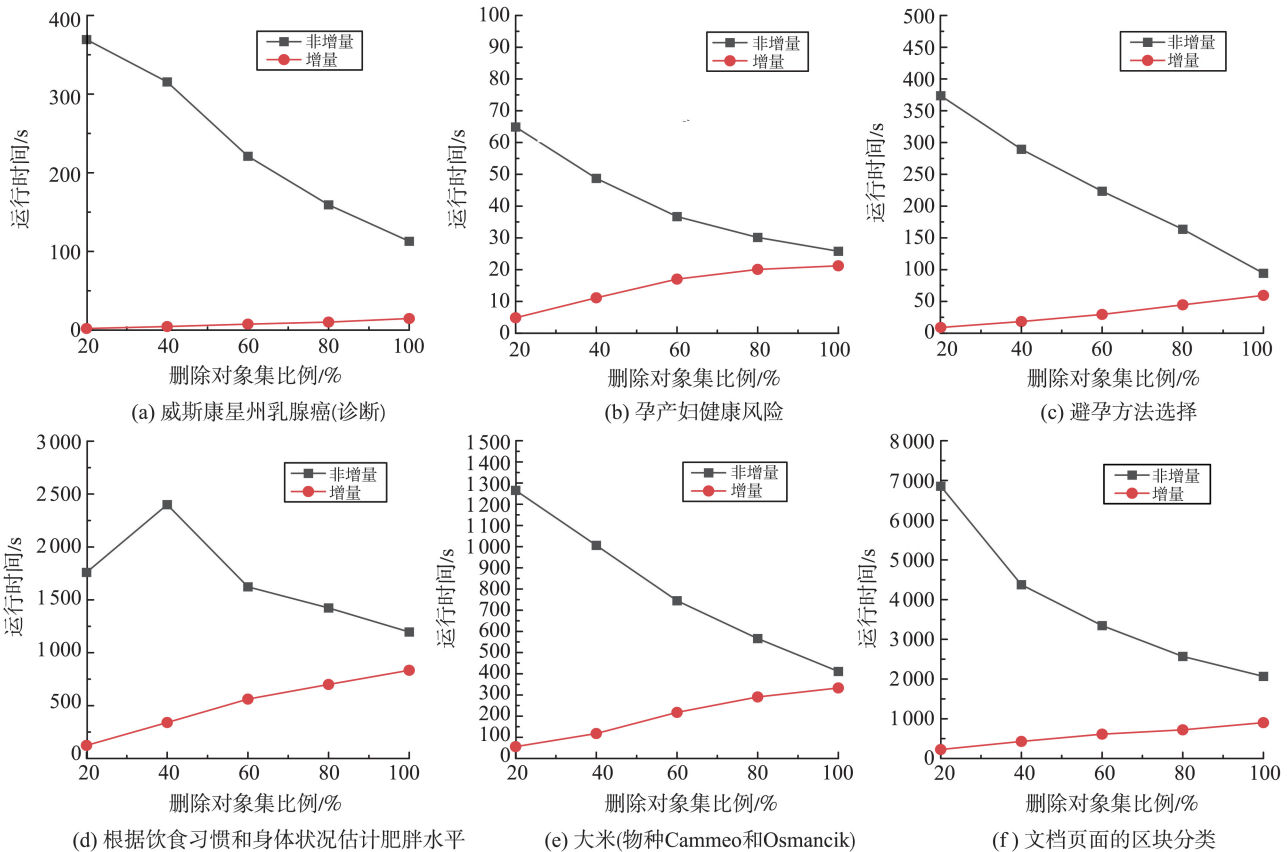


图3 两种算法在数据集上每删除一组数据的时间比较

Fig.3 Comparison of the time taken by each algorithm to delete a group of data on the dataset

图 2 和图 3 中各子图的试验结果分析如下。

(1) 当区间值决策信息系统中的对象集发生变化时,增量式属性约简方法比非增量式属性约简方法消耗的时间要少。

(2) 由于满足算法 1 中的步骤 4 的条件,导致图 2(d)(f)与图 3(d)非增量式属性约简算法中出现数据量多,时间消耗却花费更少的情况。

(3) 数据集的对象集越大、增加的对象越多,增量式属性约简方法的优势越明显。而删除的数据对象越多,尽管增量式属性约简方法的时间消耗随着数据的减少会有所上升,但仍然比非增量式属性约简方法节省时间。

(4) 非增量属性约简方法消耗时间的变化幅度较大,而增量方法消耗时间的变化幅度较平缓。

(5) 本研究确定阈值 θ 取为 0.6,依据是由经验得来的,不够充分,阈值 θ 如何确定应予以考虑,应依据试验的结果来确定。

3.4 属性约简结果及分析

为了验证增量式属性约简算法不会影响属性约简的分类精度,每个数据集增加到 100%时,两种算法的属性约简结果对比如表 4 所示。

表 4 两种算法的属性约简结果对比
Table 4 Comparison of attribute reduction results between two algorithms

数据集	属性约简结果	
	非增量算法	增量算法
威斯康星州乳腺癌(诊断)	7,6,27,26	7,6,27,26
孕产妇健康风险	0,1,2,3,5	0,1,2,3,5
避孕方法选择	8,7,2,1	8,7,2,1
根据饮食习惯和身体状况估计肥胖水平	0,1,3,5,6,7,8,9,10,12,13,14,15	0,9,10,12,13,3,14,15,6,7,1,5
大米(物种 Cammeo 和物种 Osmancik)	0,1,2,3,5,6	0,1,2,3,5,6
文档页面的区块分类	3,4,5,8,9,6,2	3,4,5,8,9,6

使用 KNN 分类器进一步对属性约简结果进行分类精度分析(其中 k 取 3),随机选取数据集中每类对象的 66%作为训练集,剩下的 34%作为测试集,结果如表 5 所示。

表 5 两种算法属性约简结果分类精度对比
Table 5 Comparison of classification accuracy of attribute reduction results between two algorithms

数据集	分类精度/%	
	非增量算法	增量算法
威斯康星州乳腺癌(诊断)	83.937 8	83.937 8
孕产妇健康风险	79.130 4	79.130 4
避孕方法选择	45.109 8	45.109 8
根据饮食习惯和身体状况估计肥胖水平	76.880 2	75.208 9
大米(物种 Cammeo 和物种 Osmancik)	87.335 9	87.335 9
文档页面的区块分类	95.056 4	95.002 7

通过对表 5 的分类精度分析可知,使用增量式算法与非增量式算法所得属性约简结果分类精度很接近,进一步说明了所提出的增量式算法是可行且有效的。

3.5 与其他增量式方法的对比

选取文献[21-23]中的增量式属性约简方法与本研究方法相对比,试验的软硬件环境及所选数据集与前文一致,本次试验选取数据集的 20%作为增加或删除的对象集,记录每次试验运行的时间,试验结果如表 6 所示。

表 6 四种算法运行所消耗的时间对比
Table 6 Comparison of time consumed to run the four algorithms

数据集	单位:s							
	文献[21]运行时间		文献[22]运行时间		文献[23]运行时间		本研究运行时间	
	增加	删除	增加	删除	增加	删除	增加	删除
威斯康星州乳腺癌(诊断)	5.36	4.54	6.29	5.37	5.97	9.55	3.05	5.51
孕产妇健康风险	8.48	15.68	11.62	15.62	15.79	29.36	3.75	12.40
避孕方法选择	30.76	132.66	62.34	165.34	53.26	163.85	10.86	17.36
根据饮食习惯和身体状况估计肥胖水平	302.50	426.58	537.6	629.86	639.85	1891.29	293.32	1312.82
大米(物种 Cammeo 和物种 Osmancik)	532.91	892.35	563.42	729.12	1 257.26	2 807.20	79.21	603.92
文档页面的区块分类	579.52	1 648.32	738.3	2 629.61	2 269.35	2 378.54	121.67	1 082.08

由表6的时间消耗结果对比可以看出,本研究增量式属性约简算法有一定的优势,特别是在对象集增加时,优势更加的明显。

4 结论

本研究以区间值决策信息系统为对象进行属性约简的研究,重新给出相容度的定义并以此构造出相容关系矩阵,以矩阵为工具分析了当区间值决策信息系统的对象集发生变化时相对知识粒度的更新机理,由此提出了一种以知识粒度表示属性重要度,进而以属性重要度为启发式信息的增量式更新属性约简的方法。最后在选取的6个UCI数据集上实施算法的试验,以验证增量式算法的有效性。试验结果表明:本研究提出的增量式属性约简方法在时间消耗上明显低于非增量式属性约简方法,而且这两种方法所得到的属性约简结果的分类精度十分接近。本研究的研究仅考虑了从对象集发生变化的情况,未来工作的重点是考虑选取阈值 θ 的方法以及在属性集发生变化时,区间值决策信息系统中基于知识粒度的增量式属性约简方法。

参考文献:

- [1] PAWLAK Z. Rough sets[J]. International Journal of Computer & Information Sciences, 1982, 11(5): 341-356.
- [2] 黄锦静, 陈岱, 李梦天. 基于粗糙集的决策树在医疗诊断中的应用[J]. 计算机技术与发展, 2017, 27(12): 148-152.
HUANG Jinjing, CHEN Dai, LI Mengtian. Application of decision tree based on rough set in medical diagnosis[J]. Computer Technology and Development, 2017, 27(12): 148-152.
- [3] TIEW S T, CHEW Y E, LEE H Y, et al. A fragrance prediction model for molecules using rough set-based machine learning [J]. Chemie Ingenieur Technik, 2023, 95(3): 438-446.
- [4] 王江荣, 黄建华, 罗资琴, 等. 基于粗糙集的 Logistic 回归模型在矿井突水模式识别中的应用[J]. 煤田地质与勘探, 2015, 43(6): 70-74.
WANG Jiangrong, HUANG Jianhua, LUO Ziqin, et al. Application of Logistic regression model based on rough set in recognition of mine water inrush pattern[J]. Coal Geology & Exploration, 2015, 43(6): 70-74.
- [5] QI Z H, LI H, LIU F, et al. Fusion decision strategies for multiple criterion preferences based on three-way decision[J]. Information Fusion, 2024, 108: 102356.
- [6] DAI J H, WANG Z Y, HUANG W Y. Interval-valued fuzzy discernibility pair approach for attribute reduction in incomplete interval-valued information systems[J]. Information Sciences, 2023, 642: 119215.
- [7] CHEN Y Y, LI Z W, ZHANG G Q. Attribute reduction in an incomplete interval-valued decision information system[J]. IEEE Access, 2021, 9: 64539-64557.
- [8] LI W T, ZHOU H X, XU W H, et al. Interval dominance-based feature selection for interval-valued ordered data[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(10): 6898-6912.
- [9] LIU X F, DAI J H, CHEN J L, et al. A fuzzy α -similarity relation-based attribute reduction approach in incomplete interval-valued information systems[J]. Applied Soft Computing, 2021, 109: 107593.
- [10] ZHANG X, MEI C L, CHEN D G, et al. Multi-confidence rule acquisition and confidence-preserved attribute reduction in interval-valued decision systems[J]. International Journal of Approximate Reasoning, 2014, 55(8): 1787-1804.
- [11] CHEN B W, ZHANG X Y, YANG J L. Feature selections based on three improved condition entropies and one new similarity degree in interval-valued decision systems[J]. Engineering Applications of Artificial Intelligence, 2023, 126: 107165.
- [12] 李磊涛, 张楠, 童向荣, 等. 基于差别矩阵的区间值决策系统 β 分布约简[J]. 计算机应用, 2021, 41(4): 1084-1092.
LI Leitao, ZHANG Nan, TONG Xiangrong, et al. β -distribution reduction based on discernibility matrix in interval-valued decision systems[J]. Journal of Computer Applications, 2021, 41(4): 1084-1092.
- [13] 唐鹏飞, 莫智文, 谢鑫. 区间值决策表中基于相对知识粒度的属性约简[J]. 重庆理工大学学报(自然科学), 2021, 35(11): 286-292.
TANG Pengfei, MO Zhiwen, XIE Xin. Attribute reduction based on relative knowledge granularity in interval-valued decision table[J]. Journal of Chongqing University of Technology (Natural Science), 2021, 35(11): 286-292.
- [14] ZHANG Y Y, LI T R, LUO C, et al. Incremental updating of rough approximations in interval-valued information systems

- under attribute generalization[J]. *Information Sciences*, 2016, 373: 461-475.
- [15] YANG L, QIN K Y, SANG B B, et al. Dynamic fuzzy neighborhood rough set approach for interval-valued information systems with fuzzy decision[J]. *Applied Soft Computing*, 2021, 111: 107679.
- [16] ZHANG X Y, LI J R. Incremental feature selection approach to interval-valued fuzzy decision information systems based on λ -fuzzy similarity self-information[J]. *Information Sciences*, 2023, 625: 593-619.
- [17] 张迎迎, 李天瑞. 区间值信息系统中一种极大相容类的增量更新算法[J]. *小型微型计算机系统*, 2017, 38(7): 1573-1579.
- ZHANG Yingying, LI Tianrui. Incremental algorithm for updating maximal consistent classes in interval-valued information system[J]. *Journal of Chinese Computer Systems*, 2017, 38(7): 1573-1579.
- [18] DAI J H, WANG W T, XU Q, et al. Uncertainty measurement for interval-valued decision systems based on extended conditional entropy[J]. *Knowledge-Based Systems*, 2012, 27: 443-450.
- [19] 王磊, 李天瑞. 一种基于矩阵的知识粒度计算方法[J]. *模式识别与人工智能*, 2013, 26(5): 447-453.
- WANG Lei, LI Tianrui. A matrix-based approach for calculation of knowledge granulation[J]. *Pattern Recognition and Artificial Intelligence*, 2013, 26(5): 447-453.
- [20] SANG B B, CHEN H M, YANG L, et al. Feature selection for dynamic interval-valued ordered data based on fuzzy dominance neighborhood rough set[J]. *Knowledge-Based Systems*, 2021, 227: 107223.
- [21] 焦玉清, 张勇. 基于区间值信息系统的信息熵增量式属性约简算法[J]. *绥化学院学报*, 2021, 41(9): 141-147.
- JIAO Yuqing, ZHANG Yong. Incremental attribute reduction of information entropy in interval-valued information system [J]. *Journal of Suihua University*, 2021, 41(9): 141-147.
- [22] 鲍迪, 张楠, 童向荣, 等. 区间值决策表的正域增量式属性约简算法[J]. *计算机应用*, 2019, 39(8): 2288-2296.
- BAO Di, ZHANG Nan, TONG Xiangrong, et al. Incremental attribute reduction algorithm of positive region in interval-valued decision tables[J]. *Journal of Computer Applications*, 2019, 39(8): 2288-2296.
- [23] JING Y G, LI T R, LUO C, et al. An incremental approach for attribute reduction based on knowledge granularity[J]. *Knowledge-Based Systems*, 2016, 104: 24-38.

(编辑:熊小原)

(上接第44页)

- LI Hao, LIU Yongjian, XIE Qing, et al. Distant supervision relation extraction model based on multi-level attention mechanism [J]. *Computer Science*, 2019, 46(10): 252-257.
- [19] YANG X N, XIAO Y L. Named entity recognition based on BERT-MBiGRU-CRF and multi-head self-attention mechanism [C]//2022 4th International Conference on Natural Language Processing (ICNLP). Xi'an, China: IEEE, 2022: 178-183.
- [20] ZHANG Q H, WU M, LV P T, et al. Research on named entity recognition of Chinese electronic medical records based on multi-head attention mechanism and character-word information fusion [J]. *Journal of Intelligent & Fuzzy Systems*, 2022, 42(4): 4105-4116.
- [21] YANG Z W, MA J, CHEN H C, et al. Context-aware attentive multilevel feature fusion for named entity recognition[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 35(1): 973-984.
- [22] MA X D, YU R L, GAO C X, et al. Research on named entity recognition method of marine natural products based on attention mechanism[J]. *Frontiers in Chemistry*, 2023, 11: 958002.
- [23] KIM J, KO Y, SEO J. A bootstrapping approach with CRF and deep learning models for improving the biomedical named entity recognition in multi-domains [J]. *IEEE Access*, 2019, 7: 70308-70318.
- [24] 何晨, 苑迎春, 王克俭, 等. 高校学业文本命名实体识别及数据集构建研究[J]. *计算机工程与应用*, 2023, 59(22): 322-328.
- HE Chen, YUAN Yingchun, WANG Kejian, et al. Research on college academic text named entity recognition and dataset construction [J]. *Computer Engineering and Applications*, 2023, 59(22): 322-328.

(编辑:熊小原)