

·信息工程·

DOI:10.12454/j.jsuese.202400187



本刊网刊

基于优势邻域相对决策熵的增量式属性约简算法

陈宝国¹, 陈磊^{1*}, 邓明¹, 李晓燕¹, 陈金林^{1,2}

(1. 淮南师范学院 计算机学院, 安徽 淮南 232038; 2. 南京航空航天大学 电子与信息工程学院, 江苏 南京 211106)

摘要:在大数据环境下,数据处于不断动态更新之中,这使得传统的机器学习算法面临着一定的局限和挑战。增量式学习是一种在已有模型学习结果的基础上,只针对变化数据进行学习的一个过程,可以大幅度提升数据更新过程的学习性能。增量式属性约简是在动态数据集环境下进行高效属性约简的一种常用方法和策略,然而,针对对象动态更新变化的动态有序信息系统中,现有的增量式属性约简方法忽略了粗糙近似集可以为属性约简提供分类信息这一作用。因此,在动态混合有序信息系统中提出了一种基于优势邻域相对决策熵的增量式属性约简算法。首先,针对混合有序信息系统提出了优势邻域相对决策熵,并设计了相应的矩阵表示形式和非增量式属性约简算法;然后,针对混合有序信息系统对象动态增加和动态减少的两种场景,分别研究了优势邻域相对决策熵的增量式更新,并设计出了对应的增量式属性约简算法;最后,在公共数据集上进行比较实验。结果证明了所提出增量式属性约简算法的有效性,并且其在分类精度和算法效率方面均优于现有的同类型增量式算法。

关键词:属性约简;对象变化;增量式;有序信息系统;优势邻域相对决策熵

中图分类号:TP181

文献标志码:A

文章编号:2096-3246(2026)02-0372-15

属性约简又称特征选择^[1-2],是机器学习和数据挖掘领域中最重要数据预处理技术之一。目前,粗糙集理论作为属性约简的常用工具,受到了研究者的广泛关注^[2]。在实际应用中,数据呈现出随时间变化而变化的动态特性,增量式属性约简方法可以有效地从动态数据中获得新的约简结果,提升了动态数据的约简性能和效率,使得增量式属性约简成为了当前的研究热点^[3-4]。

有序信息系统是一种常见的数据集类型,一般用于单调性分类场景,优势粗糙集是处理有序信息系统的一种重要的粗糙集模型^[5-7],Chen等^[8]针对混合型有序信息系统,进一步提出了优势邻域粗糙集模型。针对有序信息系统的动态更新,研究者研究提出了多种增量式属性约简方法。Sang等^[9]针对离散型有序信息系统,将优势条件熵作为启发式函数构造了一种基于矩阵的增量式属性约简,然而这种属性约简方法不适用于数值型和混合型的有序信息系统。同时,Sang等^[10]考虑数值型和离散型混合的有序信息系统,又进

一步提出了优势邻域条件熵,并构造基于矩阵形式的增量式属性约简方法,然而,这种增量式方法仅从熵的角度评估有序信息系统中属性的分类性能,忽视了优势粗糙近似的确定性。考虑到优势邻域粗糙集模型不能很好地处理包含属性噪声的有序信息系统,Sang等^[11]提出了模糊优势邻域粗糙集模型,利用模糊优势邻域条件熵作为启发式函数设计了增量式属性约简算法,同样地,这种方法也未从优势粗糙近似的角度对优势类的不确定性进行评估。由于传统的优势粗糙集忽略了信息系统每个属性之间的权重差异,Pan等^[12]提出了基于加权的优势邻域粗糙集模型,并对加权优势邻域条件熵进行增量式更新,构造了一种对象动态变化时的增量式属性约简算法,然而,这种方法需要预先学习有序信息系统中属性的权重,会消耗大量的时间,尤其对于大规模的数据,时效性差。Yang等^[13]从邻域自信息的视角出发,提出了量化优势邻域自信息,并利用该度量方法设计和构建出一种对象更新的增量式属性约简算法,然而这种增量式属性约简

收稿日期:2024-03-20 修回日期:2024-11-06 网络出版日期:2024-12-17

基金项目:安徽省高校理工科教师赴企业挂职实践计划项目(2024jsqyg284);安徽省省级质量工程项目(2023jyxm0780)

作者简介:陈宝国(1978—),男,副教授。研究方向:数据挖掘、粗糙集与粒计算。E-mail: bgchen0706@163.com

* 通信作者:陈磊,教授, E-mail: 13855491762@163.com

https://jsuese.scu.edu.cn

算法也同样未考虑优势粗糙近似的不确定性,同时也不适用于离散型和数值型混合的有序信息系统。

通过分析已提出的有序信息系统增量式属性约简,可以发现这些算法基本上以信息熵为基础进行属性评估和启发式搜索,只考虑了信息系统以熵形式所提供的确定性分类信息^[14-15],然而,粗糙集模型的上下近似集也可以为属性约简提供分类信息^[3,15-17]。针对这一问题,Jiang等^[15]利用上下近似集的粗糙度和依赖度,联合信息熵提出了相对决策熵模型,该熵模型既考虑了属性上下近似集的分类信息,也同时考虑了属性信息熵视角的分类信息;Zhang^[18]和Thuy^[19]等也提出了类似的方法,使其具有更优的属性不确定性度量效果和属性约简性能。

由于相对决策熵具有更好的属性不确定性度量效果,同时,相对决策熵的基本计算单元为属性的上下近似集,当对象变化时,无需对更新的对象进行信息粒的计算^[15],因此,本文将相对决策熵推广至混合型有序信息系统环境,提出优势邻域相对决策熵模型,并设计出一种增量式属性约简算法。首先,针对优势邻域关系,提出了优势邻域相对决策熵,利用矩阵的形式重构了优势邻域相对决策熵,并设计了混合型有序信息系统的非增量式属性约简算法;然后,对于混合型有序信息系统对象增加和对象减少的两种变化场景,分析和研究了优势邻域相对决策熵的矩阵形式增量式更新,并以这种更新机制分别构建了增量式属性约简算法;最后,在8个公共数据集上进行仿真实验,并与非增量式算法的实验结果相比,验证了所提出增量式算法的有效性,同时与现有的增量式算法相比,验证了所提出增量式算法的优越性。

1 基本理论

决策信息系统通常由4元组 $I(I=(U, T, V, f))$ 表示,其中: U 称为论域, $U=\{u_1, u_2, \dots, u_n\}$; $T=CUD$, C 为条件属性集, D 为决策属性集; $V=\bigcup_{a \in T} V_a$, V_a 为属性 a 的值域; f 为对象与属性到属性值域的映射, $f: U \times T \rightarrow V$, $f(x, a) \in V_a, \forall a \in T$, 如果 V_a 为一个递增或递减的偏序,那么属性 a 也被称为一个准则,该决策信息系统也称为有序决策信息系统^[5],表示为 $I^{\pm}=(U, T, V, f)$ 。

针对离散型和数值型混合的信息系统,Chen等^[8]提出了优势邻域粗糙集模型,该模型中,决策信息系统的条件属性集满足 $C=C^c \cup C^n, C^c \cap C^n = \emptyset, C^c$ 和 C^n 分别表示离散型和数值型属性子集。

定义1^[8] 对于有序决策信息系统 $I^{\pm}=(U, T, V, f)$, 设属性子集 $A \subseteq T$ 且 $A=A^c \cup A^n, A^c$ 和 A^n 分别表示离散型和数值型属性子集,邻域半径为 δ ,由属性子集 A 确

定的优势邻域关系 N_A 定义为:

$$N_A = \{(u_i, u_j) \in U \times U \mid \Delta_{ij}(a^n) \geq \delta \wedge f(u_i, a^n) \succ_r f(u_j, a^n) \wedge f(u_i, a^c) \succ_r f(u_j, a^c), \forall a^n \in A^n, \forall a^c \in A^c\} \cup \{(u_i, u_i) \mid u_i \in U\} \quad (1)$$

式中, $\Delta_{ij}(a^n)$ 表示对象 u_i 和 u_j 在属性 a^n 下的距离度量,定义为 $\Delta_{ij}(a^n) = |f(u_i, a^n) - f(u_j, a^n)|$ 。在不引起混淆的情形下,下文中省略邻域半径 δ 的标记。

对于 $\forall u_i \in U$ 和优势邻域关系 N_A ,可以诱导出对象 u_i 的两种邻域类,分别为对象 u_i 优势邻域类 $n_A^+(u_i)$ ($n_A^+(u_i) = \{u_j \in U \mid (u_j, u_i) \in N_A\}$)和对象 u_i 劣势邻域类 $n_A^-(u_i)$ ($n_A^-(u_i) = \{u_j \in U \mid (u_i, u_j) \in N_A\}$)。

对于有序决策信息系统 $I^{\pm}=(U, T, V, f), T=CUD$,论域 U 在决策属性集 D 上诱导出一组决策类划分 $U/D, U/D = \{I_1^{\pm}, I_2^{\pm}, \dots, I_r^{\pm}\}, r$ 为元素个数,同时这些决策类满足偏序关系,因此对于决策类 $\forall I_s \in U/D$,上联合集 I_s^{\pm} 定义为 $I_s^{\pm} = \bigcup_{r' \geq s} I_{r'}^{\pm}$,下联合集 I_s^{\pm} 定义为 $I_s^{\pm} = \bigcup_{r' \leq s} I_{r'}^{\pm}, s, r'$ 为元素编号。

定义2^[8] 对于有序决策信息系统 $I^{\pm}=(U, T, V, f)$,上联合集 I_s^{\pm} 和下联合集 I_s^{\pm} 关于属性子集 $A(A \subseteq T)$ 的下近似集 $\underline{N}_A(I_s^{\pm})$ 和上近似集 $\overline{N}_A(I_s^{\pm})$ 分别定义为:

$$\begin{aligned} \underline{N}_A(I_s^{\pm}) &= \{u \in U \mid n_A^+(u) \subseteq I_s^{\pm}\}, \\ \overline{N}_A(I_s^{\pm}) &= \{u \in U \mid n_A^+(u) \cap I_s^{\pm} \neq \emptyset\}, \\ \underline{N}_A(I_s^{\pm}) &= \{u \in U \mid n_A^-(u) \subseteq I_s^{\pm}\}, \\ \overline{N}_A(I_s^{\pm}) &= \{u \in U \mid n_A^-(u) \cap I_s^{\pm} \neq \emptyset\}. \end{aligned}$$

同时,上联合集 I_s^{\pm} 关于属性子集 $A(A \subseteq T)$ 的近似精度定义为 $\mu_A(I_s^{\pm}), \mu_A(I_s^{\pm}) = \frac{|\underline{N}_A(I_s^{\pm})|}{|\overline{N}_A(I_s^{\pm})|}$;上联合集 I_s^{\pm} 关于属性子集 $A(A \subseteq T)$ 的近似粗糙度定义为 $\rho_A(I_s^{\pm}), \rho_A(I_s^{\pm}) = 1 - \mu_A(I_s^{\pm})$;上联合集 I_s^{\pm} 关于属性子集 $A(A \subseteq T)$ 的近似依赖度定义为 $\gamma_A(I_s^{\pm}), \gamma_A(I_s^{\pm}) = \frac{|\underline{N}_A(I_s^{\pm})|}{|U|}$; $|\cdot|$ 代表集合的基数。

2 基于优势邻域相对决策熵的属性约简

2.1 优势邻域相对决策熵

Jiang等^[15]同时考虑了决策正区域和决策边界对信息系统不确定性的影响,提出了相对决策熵模型。接下来在优势邻域关系下,定义一种优势邻域相对决策熵度量。

定义3 对于有序决策信息系统 $I^{\pm}=(U, T, V, f)$,上联合集 I_s^{\pm} 关于属性子集 $A(A \subseteq T)$ 的优势邻域相对决策熵 $E_A(I_s^{\pm})$ 定义为:

$$E_A(I_s^{\pm}) = (1 - \gamma_A(I_s^{\pm})) \cdot \rho_A(I_s^{\pm}) \lg(\rho_A(I_s^{\pm}) + 1).$$

同时,决策属性 D 关于属性子集 $A(A \subseteq T)$ 的优势邻域相对决策熵定义为 $E_A(D)$, $E_A(D) = \sum_{\forall l_i^c \in U/D} E_A(l_i^c)$ 。根据定义 2 可以得到 $0 \leq E_A(l_i^c) \leq 1$, 因此, $0 \leq E_A(D) \leq |U/D|$ 。

粗糙度是信息系统不确定性度量的常用方法^[15,20-21],它在不确定性度量的过程中提供了近似集边界域的信息,而依赖度提供了近似集的正区域信息,因此,定义 3 中的优势邻域相对决策熵以熵的形式联合了粗糙度和依赖度度量,可以综合评估信息系统的确定性。

性质 1 对于有序决策信息系统 $I^c=(U, T, V, f)$,属性子集 $A_1 \subseteq A_2 \subseteq T$ 满足 $E_{A_2}(D) \leq E_{A_1}(D)$ 。

证明:见附录 A 性质 1 证明。

2.2 优势邻域相对决策熵的矩阵表示

矩阵是表示粗糙集模型的一种重要方法^[4,9-11,13,22-25],本文将提出优势邻域相对决策熵的矩阵表示方法。

定义 4 对于有序决策信息系统 $I^c=(U, T, V, f)$, $U=\{u_1, u_2, \dots, u_n\}$, $T=CUD$,其中 $U/D=\{l_1^c, l_2^c, \dots, l_r^c\}$,那么决策属性 D 对应的决策矩阵 $D(D=[d_{ij}]_{n \times r}, 1 \leq i \leq n, 1 \leq j \leq r)$ 的元素 d_{ij} 定义为:

$$d_{ij} = \begin{cases} 1, & u_i \in l_j^c; \\ 0, & u_i \notin l_j^c \end{cases} \quad (2)$$

通过定义 4 可以看出,决策矩阵 D 的第 j 个列向量即为决策类 $l_j^c(l_j^c \in U/D)$ 对应的特征向量。

定义 5 对于有序决策信息系统 $I^c=(U, T, V, f)$, $U=\{u_1, u_2, \dots, u_n\}$, $A \subseteq T$,优势邻域关系 N_A 对应的优势邻域关系矩阵 $N_A(N_A=[p_{ij}^A]_{n \times n}, 1 \leq i, j \leq n)$ 的元素 p_{ij}^A 定义为:

$$p_{ij}^A = \begin{cases} 1, & (u_i, u_j) \in N_A; \\ 0, & (u_i, u_j) \notin N_A \end{cases} \quad (3)$$

通过定义 5 可以看出,优势邻域关系矩阵 N_A 的第 j 个列向量即为优势邻域类 $N_A^+(u_j)$ 的特征向量。

定义 6 对于有序决策信息系统 $I^c=(U, T, V, f)$,给定决策矩阵 D 和优势邻域关系矩阵 N ,那么 $U/D(U/D=\{l_1^c, l_2^c, \dots, l_r^c\})$ 关于优势邻域关系 N_A 的下近似集矩阵 $L_A(L_A=[\varphi_{ij}]_{n \times r})$ 和上近似集矩阵 $U_A(U_A=[\phi_{ij}]_{n \times r})$ 的元素分别定义为:

$$\varphi_{ij} = \bigwedge_{k=1}^n [(1 - p_{ki}^A) \vee d_{kj}], 1 \leq i \leq n, 1 \leq j \leq r;$$

$$\phi_{ij} = \bigvee_{k=1}^n [p_{ki}^A \wedge d_{kj}], 1 \leq i \leq n, 1 \leq j \leq r.$$

在定义 6 中,对于决策上联合集 l_s^c 的下近似集 $\underline{N}_A(l_s^c)$ 和上近似集 $\bar{N}_A(l_s^c)$,若 $u_e \in \underline{N}_A(l_s^c)$, $u_f \notin \underline{N}_A(l_s^c)$,那么 $\varphi_{es}=1, \varphi_{fs}=0$;同理若 $u_g \in \bar{N}_A(l_s^c)$, $u_h \notin \bar{N}_A(l_s^c)$,那

么 $\phi_{gs}=1, \phi_{hs}=0$;其中, e, f, g, h 为不同对象编号。

可以看出,下近似集矩阵的第 i 列刚好是决策类 l_i^c 下近似集的特性向量形式,上近似集矩阵也是如此。

定义 7 给定矩阵 $M=[m_{ij}]_{n \times n}$,定义其函数 $S(\cdot)$ 运算结果 $S(M)$ 为: $S(M)=[\omega_{ij}]_{n \times n}$,定义该行向量元素

$$\omega_{ij} \text{ 为 } \omega_{ij} = \sum_{k=1}^n m_{kj}, 1 \leq j \leq n.$$

综合定义 4 至定义 7 的矩阵计算方式,可以得到优势邻域相对决策熵的矩阵形式表达。

定理 1 对于有序决策信息系统 $I^c=(U, T, V, f)$,决策属性 D 关于 $A(A \subseteq T)$ 的优势邻域相对决策熵的矩阵形式 $E_A(D)(E_A(D)=[\sigma_{ij}]_{1 \times r})$ 表示为:

$$E_A(D) = (1 - S(L_A) ./ n) \cdot (1 - S(L_A) ./ S(U_A)) \cdot \text{lb}(2 - S(L_A) ./ S(U_A)) \quad (4)$$

式中:“./”代表矩阵的点除,即矩阵对应元素相除; $S(L_A) ./ n$ 代表 $S(L_A)$ 每个元素除以常数 n ; $1 - S(L_A) ./ n$ 代表元素全为 1 的矩阵减去矩阵 $S(L_A) ./ n$ 。

证明:见附录 A 定理 1 证明。

在定理 1 中,可以看出优势邻域相对决策熵的矩阵形式 $E_A(D)$ 是一个大小为 $1 \times r$ 的向量,向量第 i 个元素即为 $E_A(l_i^c)$ 的值,因此对 $E_A(D)$ 中所有元素进行累加便得到最终的 $E_A(D)$ 结果。

根据优势邻域相对决策熵的矩阵表达,可以得到如下性质。

性质 2 对于有序决策信息系统 $I^c=(U, T, V, f)$,属性子集 $P, Q(P \subseteq T, Q \subseteq T)$ 对应的优势邻域关系矩阵分别为 $N_P(N_P=[p_{ij}^P]_{n \times n})$ 和 $N_Q(N_Q=[q_{ij}^Q]_{n \times n})$,那么优势邻域关系 $N_{P \cup Q}$ 的优势邻域关系矩阵 $N_{P \cup Q}([m_{ij}^{P \cup Q}]_{n \times n})$ 的元素 $m_{ij}^{P \cup Q}$ 可表示为:

$$m_{ij}^{P \cup Q} = \begin{cases} 1, & p_{ij}^P = 1 \wedge q_{ij}^Q = 1; \\ 0, & \text{否则} \end{cases} \quad (5)$$

性质 3 对于有序决策信息系统 $I^c=(U, T, V, f)$,属性子集 $P, Q(P \subseteq Q \subseteq T)$ 对应的优势邻域关系矩阵分别为 N_P 和 N_Q ,那么优势邻域关系 N_{Q-P} 的优势邻域关系矩阵 $N_{Q-P}(N_{Q-P}=[m_{ij}^{Q-P}]_{n \times n})$ 的元素 m_{ij}^{Q-P} 可表示为:

$$m_{ij}^{Q-P} = \begin{cases} 1, & q_{ij}^Q = 1; \\ 0, & p_{ij}^P = 1 \wedge q_{ij}^Q = 0; \\ 1, & (u_i, u_j) \in N_{Q-P}; \\ 0, & \text{否则} \end{cases} \quad (6)$$

性质 4 对于有序决策信息系统 $I^c=(U, T, V, f)$,属性子集 $P \subseteq T, Q \subseteq T$,优势邻域关系 N_P 的下近似集矩阵定义为 $L_P, L_P=[\varphi_{ij}]_{n \times r}$;上近似集矩阵定义为 $U_P, U_P=[\phi_{ij}]_{n \times r}$;优势邻域关系 $N_{P \cup Q}$ 的优势邻域关系矩阵

定义为 $N_{P \cup Q} = [m_{ij}^{P \cup Q}]_{n \times n}$, 那么 $N_{P \cup Q}$ 的下近似集矩阵 $L_{P \cup Q}$ ($L_{P \cup Q} = [\phi'_{ij}]_{n \times r}$) 和上近似矩阵 $U_{P \cup Q}$ ($U_{P \cup Q} = [\phi'_{ij}]_{n \times r}$) 的元素分别表示为:

$$\phi'_{ij} = \begin{cases} 1, \phi_{ij} = 1; \\ \bigwedge_{k=1}^n ((1 - m_{ki}^{P \cup Q}) \vee d_{kj}), \text{ 否则} \end{cases} \quad (7)$$

$$\phi'_{ij} = \begin{cases} 0, \phi_{ij} = 0; \\ \bigvee_{k=1}^n (m_{ki}^{P \cup Q} \wedge d_{kj}), \text{ 否则} \end{cases} \quad (8)$$

证明: 见附录 A 性质 4 证明。

性质 5 对于有序决策信息系统 $I^c = (U, T, V, f)$, 属性子集 $P \subseteq Q \subseteq T$, 优势邻域关系 N_Q 的下近似集矩阵定义为 $L_Q, L_Q = [\phi_{ij}]_{n \times r}$; 上近似集矩阵定义为 $U_Q, U_Q = [\phi_{ij}]_{n \times r}$; 优势邻域关系 N_{Q-P} 的优势邻域关系矩阵定义为 $N_{Q-P} = [m_{ij}^{Q-P}]_{n \times n}$, 那么 N_{Q-P} 的下近似集矩阵 L_{Q-P} ($L_{Q-P} = [\phi'_{ij}]_{n \times r}$) 和上近似矩阵 U_{Q-P} ($U_{Q-P} = [\phi'_{ij}]_{n \times r}$) 的元素分别表示为:

$$\phi'_{ij} = \begin{cases} 0, \phi_{ij} = 0; \\ \bigwedge_{k=1}^n ((1 - m_{ki}^{Q-P}) \vee d_{kj}), \text{ 否则} \end{cases} \quad (9)$$

$$\phi'_{ij} = \begin{cases} 1, \phi_{ij} = 1; \\ \bigvee_{k=1}^n (m_{ki}^{Q-P} \wedge d_{kj}), \text{ 否则} \end{cases} \quad (10)$$

证明: 见附录 A 性质 5 证明。

性质 2~5 表明了信息系统增加和减少属性时下近似集矩阵和上近似矩阵的变化关系, 利用此性质可进一步实现混合有序信息系统矩阵方法的属性约简。

2.3 优势邻域相对决策熵的非增量式属性约简

定义 8 对于有序决策信息系统 $I^c = (U, T, V, f)$, $T = CUD$, 若 c 为信息系统的一个属性约简集, 则需同时满足:

- 1) $E_c(D) = E_C(D)$;
- 2) $\forall a \in c, E_{c-\{a\}}(D) > E_c(D)$ 。

根据定义 8 可得到矩阵方法的属性重要度定义。

定义 9 对于有序决策信息系统 $I^c = (U, T, V, f)$, $T = CUD$ 且 $A \subseteq C, \forall a \in A$ 的内部属性重要度矩阵 $G_{\text{inn}}(a, A, D)$ 定义为: $G_{\text{inn}}(a, A, D) = E_{A-\{a\}}(D) - E_A(D)$ 。

$\forall b \in C - A$ 的外部属性重要度矩阵 $G_{\text{out}}(b, A, D)$ 定义为: $G_{\text{out}}(b, A, D) = E_A(D) - E_{A \cup \{b\}}(D)$ 。

算法 1 基于优势邻域相对决策熵的非增量式属性约简算法。

输入: 有序决策信息系统 $I^c = (U, T, V, f), T = CUD$, 邻域半径 δ 。

输出: 属性约简集 c 。

初始化 $c \leftarrow \emptyset$; /* 步骤 1 */

/* 步骤 2 */

for $\forall a \in C$

计算内部属性重要度 $G_{\text{inn}}(a, C, D) = E_{C-\{a\}}(D) - E_C(D)$;

if $G_{\text{inn}}(a, C, D) > 0$

$c \leftarrow c \cup \{a\}$;

end if

end for

/* 步骤 3 */

while $E_c(D) \neq E_C(D)$

if $E_c(D) > E_C(D)$

for $\forall a \in C - c$

计算外部属性重要度 $G_{\text{out}}(a, c, D) = E_c(D) - E_{c \cup \{a\}}(D)$;

end for

$a_{\text{max}} = \arg \max_{\forall a \in C - c} (G_{\text{out}}(a, c, D))$;

$c \leftarrow c \cup \{a_{\text{max}}\}$;

end if

end while

/* 步骤 4 */

for $\forall b \in c$

计算内部属性重要度 $G_{\text{inn}}(b, c, D) = E_{c-\{b\}}(D) - E_c(D)$;

if $G_{\text{inn}}(b, c, D) = 0$

$c \leftarrow c - \{b\}$;

end if

end for

return c ; /* 步骤 5 */

根据启发式属性约简的算法架构^[15], 算法 1 的时间复杂度为 $O(|C|^2 \cdot |U|^2)$ 。

3 优势邻域相对决策熵的增量式属性约简

动态变化的信息系统中采用算法 1 进行属性约简非常耗时, 因此, 针对信息系统对象动态变化场景, 提出一种优势邻域相对决策熵的增量式属性约简算法。

3.1 对象动态增加时的增量式属性约简

性质 6 对于有序决策信息系统 $I^c = (U, T, V, f)$, 论域 U 增加对象集 $U^+ (U^+ = \{u_{n+1}, u_{n+2}, \dots, u_{n+n^+}\})$ 、 $U' (U' = U \cup U^+)$, $U/D (U/D = \{I_1^c, I_2^c, \dots, I_r^c\})$, 对应的决策矩阵 $D = [d_{ij}]_{n \times r}$, 令 $U'/D = \{I_1'^c, I_2'^c, \dots, I_r'^c\}$, 那么对应的决策矩阵 $\hat{D} (\hat{D} = [\hat{d}_{ij}]_{(n+n^+) \times r})$ 的元素更新为:

$$\hat{d}_{ij} = \begin{cases} d_{ij}, 1 \leq i \leq n, 1 \leq j \leq r; \\ 1, u_{n+i} \in I_j^c, 1 \leq i \leq n^+, 1 \leq j \leq r; \\ 0, u_{n+i} \notin I_j^c, 1 \leq i \leq n^+, 1 \leq j \leq r \end{cases} \quad (11)$$

证明:根据定义 4 可以直接得到性质 6 成立。

性质 7 对于有序决策信息系统 $I^c=(U, T, V, f)$, 优势邻域关系 N_A 对应的关系矩阵 N_A 定义为 $N_A=[p_{ij}^A]_{n \times n}$, 论域 U 增加对象集 U^+ 后, 新的优势邻域关系 N'_A 对应的关系矩阵 \hat{N}_A ($\hat{N}_A=[\hat{p}_{ij}^A]_{(n+n^+) \times (n+n^+)}$) 的元素更新为:

$$\hat{p}_{ij}^A = \begin{cases} p_{ij}^A, 1 \leq i, j \leq n; \\ 1, (u_i, u_{n+j}) \in N'_A, 1 \leq i \leq n, 1 \leq j \leq n^+; \\ 1, (u_{n+i}, u_j) \in N'_A, 1 \leq i \leq n^+, 1 \leq j \leq n; \\ 1, (u_{n+i}, u_{n+j}) \in N'_A, 1 \leq i, j \leq n^+; \\ 0, \text{否则} \end{cases} \quad (12)$$

证明:根据定义 5 可以直接得到性质 7 成立。

性质 8 对于有序决策信息系统 $I^c=(U, T, V, f)$, U/D 关于优势邻域关系 N_A 的下近似集矩阵和上近似集矩阵分别为 L_A 和 U_A , 论域 U 增加 \hat{L}_A 对象集 U^+ 后, U^+/D 关于优势邻域关系 N'_A 的下近似集矩阵 \hat{L}_A ($\hat{L}_A=[\hat{\phi}_{ij}]_{(n+n^+) \times r}$) 和上近似集矩阵 \hat{U}_A ($\hat{U}_A=[\hat{\phi}_{ij}]_{(n+n^+) \times r}$) 分别更新为:

$$\hat{\phi}_{ij} = \begin{cases} 0, \phi_{ij} = 0; \\ \bigwedge_{k=n+1}^{n+n^+} ((1 - \hat{p}_{ki}^A) \vee \hat{d}_{kj}), \phi_{ij} = 1; \\ \bigwedge_{k=1}^{n+n^+} ((1 - \hat{p}_{ki}^A) \vee \hat{d}_{kj}), n+1 \leq i \leq n+n^+ \end{cases} \quad (13)$$

$$\hat{\phi}_{ij} = \begin{cases} 1, \phi_{ij} = 1; \\ \bigvee_{k=n+1}^{n+n^+} (\hat{p}_{ki}^A \wedge \hat{d}_{kj}), \phi_{ij} = 0; \\ \bigvee_{k=1}^{n+n^+} (\hat{p}_{ki}^A \wedge \hat{d}_{kj}), n+1 \leq i \leq n+n^+ \end{cases} \quad (14)$$

证明:见附录 A 性质 8 证明。

根据性质 6 至性质 8, 可以进一步得到信息系统论域增加对象集后优势邻域相对决策熵的矩阵形式增量式更新结果。对于有序决策信息系统 $I^c=(U, T, V, f)$, 论域增加对象集 U^+ 后, 决策属性 D 关于 $A(A \subseteq T)$ 的优势邻域相对决策熵的矩阵形式 $E_A(D)$ ($E_A(D)=[\hat{\sigma}_{ij}]_{1 \times r}$) 增量式更新为:

$$E_A(D) = (1 - S(\hat{L}_A)) / (n + n^+).$$

$$(1 - S(\hat{L}_A)) / S(\hat{U}_A) \cdot 1b(2 - S(\hat{L}_A)) / S(\hat{U}_A).$$

根据优势邻域相对决策熵的增量式更新方法, 可以进一步构建有序决策信息系统增加对象集后的增量式属性约简算法。

算法 2 基于对象增加时的优势邻域相对决策熵增量式属性约简算法。

输入: 1) 有序决策信息系统 $I^c=(U, T, V, f)$, $T=C \cup D$, 属性约简集 c , 决策矩阵 $D=[d_{ij}]_{n \times r}$, 优势邻域关系矩阵 $N_c=[p_{ij}^c]_{n \times n}$ 和 $N_C=[p_{ij}^C]_{n \times n}$, 下近似集矩阵 $L_c=$

$[\phi_{ij}]_{n \times r}$ 和 $L_C=[\phi_{ij}]_{n \times r}$, 上近似集矩阵 $U_c=[\phi_{ij}]_{n \times r}$ 和 $U_C=[\phi_{ij}]_{n \times r}$; 2) 论域 U 增加对象集 $U^+ = \{u_{n+1}, u_{n+2}, \dots, u_{n+n^+}\}$, $U' = U \cup U^+$ 。

输出: 新信息系统的属性约简集 c' 。

初始化 $c' \leftarrow c$; /* 步骤 1 */

/* 步骤 2 */

增量式更新决策矩阵 \hat{D} ;

增量式更新优势邻域关系矩阵 \hat{N}_c 和 \hat{N}_C ;

增量式更新下近似集矩阵 \hat{L} 和 \hat{L}_C ;

增量式更新上近似集矩阵 \hat{U}_c 和 \hat{U}_C ;

/* 根据性质 6 至 8, $\hat{D}=[\hat{d}_{ij}]_{(n+n^+) \times r}$, $\hat{N}_c'=[$

$\hat{p}_{ij}^c]_{(n+n^+) \times (n+n^+)}$, $\hat{N}_C'=[\hat{p}_{ij}^C]_{(n+n^+) \times (n+n^+)}$, $\hat{L}_c'=[\hat{\phi}_{ij}]_{(n+n^+) \times r}$,

$\hat{L}_C'=[\hat{\phi}_{ij}]_{(n+n^+) \times r}$, $\hat{U}_c'=[\hat{\phi}_{ij}]_{(n+n^+) \times r}$, $\hat{U}_C'=[\hat{\phi}_{ij}]_{(n+n^+) \times r}$ */

/* 步骤 3 */

增量式计算优势邻域相对决策熵 $E_{c'}(D)$ 和 $E_C(D)$;

/* 步骤 4 */

while $E_{c'}(D) \neq E_C(D)$

for $\forall a \in C - c'$

计算优势邻域关系矩阵 $\hat{N}_{c' \cup \{a\}}$;

更新计算 $\hat{L}_{c' \cup \{a\}}$ 和 $\hat{U}_{c' \cup \{a\}}$;

/* 性质 2 和 4, $\hat{N}_{c' \cup \{a\}}=[\hat{p}_{ij}^{c' \cup \{a\}}]_{(n+n^+) \times (n+n^+)}$,

$\hat{L}_{c' \cup \{a\}}=[\hat{\phi}_{ij}]_{(n+n^+) \times r}$, $\hat{U}_{c' \cup \{a\}}=[\hat{\phi}_{ij}]_{(n+n^+) \times r}$ */

更新计算 $E_{c' \cup \{a\}}(D)$;

计算外部属性重要度 $G_{\text{out}}(a, c', D)$;

/* $G_{\text{out}}(a, c', D) = E_{c'}(D) - E_{c' \cup \{a\}}(D)$ */

end for

$a_{\max} = \arg \max_{a \in C - c'} (G_{\text{out}}(a, c', D))$; /* a_{\max} 对应运

算最大值的属性 */

if $G_{\text{out}}(a_{\max}, c', D) > 0$

$c' \leftarrow c' \cup \{a_{\max}\}$;

else

break;

end if

end while

/* 步骤 5 */

for $\forall b \in c'$

计算优势邻域关系矩阵 $\hat{N}_{c' - \{b\}}$;

更新计算 $\hat{L}_{c' - \{b\}}$ 和 $\hat{U}_{c' - \{b\}}$;

/* 性质 3 和 5, $\hat{N}_{c' - \{b\}}=[\hat{p}_{ij}^{c' - \{b\}}]_{(n+n^+) \times (n+n^+)}$, $\hat{L}_{c' - \{b\}}=[$

$\hat{\phi}_{ij}]_{(n+n^+) \times r}$, $\hat{U}_{c' - \{b\}}=[\hat{\phi}_{ij}]_{(n+n^+) \times r}$ */

```

更新计算  $E_{c'-\{b\}}(D)$ ;
计算内部属性重要度  $G_{\text{inn}}(b, c', D)$ ;
/*  $G_{\text{inn}}(b, c', D) = E_{c'-\{b\}}(D) - E_{c'}(D)$  */
if  $G_{\text{inn}}(b, c', D) = 0$ 
     $c' \leftarrow c' - \{b\}$ ;
end if
end for
return  $c'$ ;

```

在算法 2 中,步骤 2 和步骤 3 的时间复杂度为 $O(|C| \cdot |U|^2 + |C| \cdot |U| \cdot |U|)$,步骤 4 至步骤 5 搜索属性和剔除属性的时间复杂度为 $O(|C| \cdot |U|^2)$,因此整个算法 2 的时间复杂度为 $O(|C| \cdot (|U|^2 + |U|^2 + |U| \cdot |U|))$ 。

3.2 对象动态减少时的增量式属性约简

性质 9 对于有序决策信息系统 $I^c = (U, T, V, f)$, 设论域 U 被删除的对象集为 $U^- = \{u_{n-n^-+1}, u_{n-n^-+2}, \dots, u_n\}$, $U' = U - U^-$, $U/D = \{I_1^c, I_2^c, \dots, I_r^c\}$, 对应的决策矩阵 $D = [d_{ij}]_{n \times r}$, 令 $U'D = \{I_1^{c'}, I_2^{c'}, \dots, I_r^{c'}\}$, 那么对应的决策矩阵 $\hat{D} (\hat{D} = [\hat{d}_{ij}]_{(n-n^-) \times r})$ 的元素更新为 $\hat{d}_{ij} = d_{ij}$, $1 \leq i \leq n - n^-$, $1 \leq j \leq r$ 。

证明:根据定义 4 可以直接得到性质 9 成立。

性质 10 对于有序决策信息系统 $I^c = (U, T, V, f)$, 优势邻域关系 N_A 对应的关系矩阵为 $N_A (N_A = [p_{ij}^A]_{n \times n})$, 论域 U 删除对象集 U^- 后, 新的优势邻域关系 N'_A 对应的关系矩阵 $\hat{N}_A (\hat{N}_A = [\hat{p}_{ij}^A]_{(n-n^-) \times (n-n^-)})$ 的元素更新为: $\hat{p}_{ij}^A = p_{ij}^A$, $1 \leq i, j \leq n - n^-$ 。

证明:根据定义 5 可以直接得到性质 10 成立。

性质 11 对于有序决策信息系统 $I^c = (U, T, V, f)$, U/D 关于优势邻域关系 N_A 的下近似集矩阵和上近似集矩阵分别为 L_A 和 U_A , 论域 U 删除对象集 U^- 后, U'/D 关于优势邻域关系 N'_A 的下近似集矩阵 $\hat{L}_A (\hat{L}_A = [\hat{\phi}_{ij}]_{(n-n^-) \times r})$ 和上近似集矩阵 $\hat{U}_A (\hat{U}_A = [\hat{\phi}_{ij}]_{(n-n^-) \times r})$ 的元素分别更新为:

$$\hat{\phi}_{ij} = \begin{cases} 1, \phi_{ij} = 1, 1 \leq i \leq n - n^-, 1 \leq j \leq r; \\ \bigwedge_{k=1}^{n-n^-} ((1 - \hat{p}_{ki}^A) \vee \hat{d}_{kj}), \phi_{ij} = 0, 1 \leq i \leq n - n^-, 1 \leq j \leq r \end{cases} \quad (15)$$

$$\hat{\phi}_{ij} = \begin{cases} 0, \phi_{ij} = 0, 1 \leq i \leq n - n^-, 1 \leq j \leq r; \\ \bigvee_{k=1}^{n-n^-} (\hat{p}_{ki}^A \wedge \hat{d}_{kj}), \phi_{ij} = 1, 1 \leq i \leq n - n^-, 1 \leq j \leq r \end{cases} \quad (16)$$

证明:见附录 A 性质 11 证明。

根据性质 9 至 11, 可以进一步得到信息系统论域删除对象集后优势邻域相对决策熵的矩阵形式增量式更新。对于有序决策信息系统 $I^c = (U, T, V, f)$, 论域删

除对象集 U^- 后, 决策属性 D 关于 $A \subseteq T$ 的优势邻域相对决策熵的矩阵形式 $E_A(D) = [\hat{\sigma}_{ij}]_{1 \times r}$ 的增量式更新为:

$$E_A(D) = (1 - S(\hat{L}_A) / (n - n^-)) \cdot (1 - S(\hat{L}_A) / S(\hat{U}_A)) \cdot \text{lb}(2 - S(\hat{L}_A) / S(\hat{U}_A))$$

根据优势邻域相对决策熵的增量式更新方法, 接下来可以进一步得到有序决策信息系统删除对象集后的增量式属性约简算法。

算法 3 基于对象删除时的优势邻域相对决策熵增量式属性约简算法

输入: 1) 有序决策信息系统 $I^c = (U, T, V, f)$, $T = C \cup D$, 属性约简 c , 决策矩阵 $D = [d_{ij}]_{n \times r}$, 优势邻域关系矩阵 $N_c = [p_{ij}^c]_{n \times n}$ 和 $N_C = [p_{ij}^C]_{n \times n}$, 下近似集矩阵 $L_c = [\phi_{ij}]_{n \times r}$ 和 $L_C = [\phi_{ij}]_{n \times r}$, 上近似集矩阵 $U_c = [\phi_{ij}]_{n \times r}$ 和 $U_C = [\phi_{ij}]_{n \times r}$ 。

2) 论域 U 删除对象集 $U^- = \{u_{n-n^-+1}, u_{n-n^-+2}, \dots, u_n\}$, $U' = U - U^-$ 。

输出: 新信息系统的属性约简 c' 。

初始化 $c' \leftarrow c$; /* 步骤 1 */

/* 步骤 2 */

增量式更新矩阵 $\hat{D}, \hat{N}_C, \hat{N}_C, \hat{L}_C, \hat{L}_C, \hat{U}_C, \hat{U}_C$;

/* 性质 9 至 11, $\hat{D} = [\hat{d}_{ij}]_{(n-n^-) \times r}, \hat{N}_C = [\hat{p}_{ij}^C]_{(n-n^-) \times (n-n^-)}$,

$\hat{N}_C = [\hat{p}_{ij}^C]_{(n-n^-) \times (n-n^-)}, \hat{L}_C = [\hat{\phi}_{ij}]_{(n-n^-) \times r}, \hat{L}_C = [\hat{\phi}_{ij}]_{(n-n^-) \times r}$,

$\hat{U}_C = [\hat{\phi}_{ij}]_{(n-n^-) \times r}, \hat{U}_C = [\hat{\phi}_{ij}]_{(n-n^-) \times r}$ */

增量式计算 $E_{c'}(D)$ 和 $E_C(D)$; /* 步骤 3 */

/* 步骤 4 */

while $E_{c'}(D) \neq E_C(D)$

for $\forall a \in C - c'$

计算 $\hat{N}_{c' \cup \{a\}}$;

更新计算 $\hat{L}_{c' \cup \{a\}}$ 和 $\hat{U}_{c' \cup \{a\}}$;

/* 性质 2 和 4, $\hat{N}_{c' \cup \{a\}} = [\hat{p}_{ij}^{c' \cup \{a\}}]_{(n-n^-) \times (n-n^-)}$,

$\hat{L}_{c' \cup \{a\}} = [\hat{\phi}_{ij}]_{(n-n^-) \times r}, \hat{U}_{c' \cup \{a\}} = [\hat{\phi}_{ij}]_{(n-n^-) \times r}$ */

更新计算 $E_{c' \cup \{a\}}(D)$;

计算外部属性重要度 $G_{\text{out}}(a, c', D)$;

/* $G_{\text{out}}(a, c', D) = E_{c'}(D) - E_{c' \cup \{a\}}(D)$ */

end for

$a_{\text{max}} = \arg \max_{a \in C - c'} (G_{\text{out}}(a, c', D))$;

if $G_{\text{out}}(a_{\text{max}}, c', D) > 0$

$c' \leftarrow c' \cup \{a_{\text{max}}\}$;

else

break;

end if

```

end while
/*步骤 5*/
for  $\forall b \in c'$ 
    计算  $\hat{N}_{c'-\{b\}}$ ;
    进一步更新  $\hat{L}_{c'-\{b\}} = \text{和 } \hat{U}_{c'-\{b\}} = [\hat{\phi}_{ij}]_{(n-n^-) \times r}$ ;
/*性质 3 和 5,  $\hat{N}_{c'-\{b\}} = [\hat{\rho}_{ij}^{c'-\{b\}}]_{(n-n^-) \times (n-n^-)}$ ,  $\hat{L}_{c'-\{b\}} =$ 
 $[\hat{\phi}_{ij}]_{(n-n^-) \times r}$ ,  $\hat{U}_{c'-\{b\}} = [\hat{\phi}_{ij}]_{(n-n^-) \times r}$ ;*/
    更新计算  $E_{c'-\{b\}}(D)$ ;
    计算  $G_{\text{inn}}(b, c', D)$ ;
/*  $G_{\text{inn}}(b, c', D) = E_{c'-\{b\}}(D) - E_{c'}(D)$ */
    if  $G_{\text{inn}}(b, c', D) = 0$ 
         $c' \leftarrow c' - \{b\}$ ;
    end if
end for
return  $c'$ ; /*步骤 6*/

```

在算法 3 中, 步骤 2 和步骤 3 的时间复杂度为 $O(|C| \cdot |U|^2 + |C| \cdot |U| \cdot |U|)$, 步骤 4 至步骤 8 搜索属性和剔除属性的时间复杂度为 $O(|C| \cdot |U|^2)$, 因此整个算法 3 的时间复杂度为 $O(|C| \cdot (|U|^2 + |U|^2 + |U| \cdot |U|))$ 。

4 实验分析

本文实验从加州大学欧文分校机器学习数据集库 (<https://archive.ics.uci.edu/>) 中下载了 8 个公共数据集, 信息如表 1 所示, 其中数据集的数值型条件属性值采用极差正规化方法标准化至 $[0, 1]$ 范围内。为保证数据集满足类别有序性, 计算每个对象所有条件属性值之和, 并按照其大小顺序标记新类别标签^[13]。所有实验环节运行的硬件环境如下: CPU 型号为 Intel(R) Core (TM) i5-8265U 1.8 GHz, 内存为 8 GB, 实验的软件为 Matlab2017b。

表 1 实验数据集

Tab. 1 Experimental datasets

编号	数据集名称	对象数量	属性数量	类别数量	类型
1	Cylinder	512	40	3	混合型
2	Wdbc	569	30	2	数值型
3	Credit	1 000	20	2	数值型
4	Abalone	4 177	8	29	混合型
5	Mushroom	5 644	22	3	离散型
6	Ticdata2000	5 822	85	3	离散型
7	Musk	6 598	166	2	数值型
8	Thyroid	9 172	29	2	混合型

4.1 对象增加时增量式与非增量式算法的性能比较

对表 1 中的每个数据集, 随机选择论域的 50% 作为初始数据集, 剩余的 50% 作为动态增加数据集, 然

后利用本文的算法 2(对象增加时的增量式算法)和算法 1(非增量式算法)分别进行属性约简计算, 邻域半径 δ 设为 0.1^[11,26-28]。表 2 为两类算法的属性约简集长度, 对于该约简结果分别采用支持向量机分类器(SVM)和朴素贝叶斯分类器(NB)进行分类训练, 结果如表 3 所示。同时, 分别记录算法 2 和算法 1 的属性约简运行时间, 评估算法效率, 重复上述实验 10 次, 其属性约简的平均运行时间如表 4 所示。

表 2 对象增加时增量式与非增量式算法属性约简集长度
Tab. 2 Length of attribute reduction sets for incremental and non-incremental algorithms when objects increase

数据集	算法 1	算法 2
Cylinder	13	13
Wdbc	15	16
Credit	9	9
Abalone	6	6
Mushroom	14	12
Ticdata2000	23	21
Musk	27	28
Thyroid	11	12

表 3 对象增加时增量式与非增量式算法属性约简集分类精度

Tab. 3 Classification accuracy of attribute reduction sets for incremental and non-incremental algorithms when objects increase

数据集	%			
	算法 1		算法 2	
	SVM	NB	SVM	NB
Cylinder	87.37	88.35	87.37	88.35
Wdbc	93.24	92.41	94.18	91.79
Credit	83.48	85.69	83.48	85.69
Abalone	90.32	91.53	90.32	91.53
Mushroom	79.65	81.41	82.85	80.56
Ticdata2000	85.84	84.96	84.46	85.75
Musk	90.47	92.56	91.31	92.59
Thyroid	78.64	80.85	79.25	81.72

表 4 对象增加时增量式与非增量式算法运算时间比较

Tab. 4 Comparison of computational time for incremental and non-incremental algorithms when objects increase

数据集	s	
	算法 1	算法 2
Cylinder	9.22	0.34
Wdbc	6.41	0.32
Credit	8.80	0.66
Abalone	24.56	4.60
Mushroom	339.18	23.12
Ticdata2000	5 387.71	95.07
Musk	26 391.46	238.47
Thyroid	1 556.49	80.50

由表2~4可知:使用本文算法2和算法1得到的属性约简集长度非常接近,其中数据集 Cylinder、Credit和 Abalone的属性约简集长度完全相同;两种算法采用SVM和NB分类的精度也很接近,部分数据集下的分类精度相同;但各个数据集下算法2的运行时间仅占算法1的1.3%左右,大幅度少于算法1。以上结果表明:本文算法2(增量式算法)在保持了属性约简长度和属性约简分类精度的同时,降低了属性约简的运行时间成本,这主要是由于算法2的属性约简与算法1采用了相同的启发式函数,因此,具有相似的约简集和分类精度;而算法2基于增量式的方法计算对象增加时的属性约简,算法1对对象增加更新后的完整数据进行属性约简,因此算法2的效率大幅度高于算法1。

4.2 对象减少时增量式与非增量式算法的性能比较

将表1的每个完整数据集分别作为初始数据集,随机选择论域的50%作为被删除的对象数据,利用本文的算法3(对象减少时的增量式算法)和算法1(非增量式算法)分别进行属性约简计算,邻域半径 δ 设为0.1。表5为两类算法的属性约简集长度,表6为SVM和NB的分类精度结果。同时,分别记录算法3和算法1的属性约简运行时间,重复上述实验10次,平均运行时间如表7所示。由表5~7可知:算法3和算法1的属性约简集长度结果非常接近或完全相同,且两种算法采用SVM和NB分类的精度也非常接近,部分数据集下分类精度完全相同;同样地,算法3的运行时间大幅少于算法1,算法3的运行时间仅占算法1的3.5%左右。以上结果表明,算法3在保持了属性约简长度和属性约简分类精度的同时,降低了约简的运行时间成本,这是由于算法3与算法1采用了相同的启发式函数,因此有相近的属性约简集和分类精度;而算法3基于旧信息系统的约简结果增量式计算对象减少后的属性约简,算法1则对更新后的完整数据进行属性约简,因此算法3的效率大幅度高于算法1。

表5 对象减少时增量式与非增量式算法属性约简集长度
Tab. 5 Length of attribute reduction sets for incremental and non-incremental algorithms when objects decrease

数据集	算法1	算法3
Cylinder	14	14
Wdbc	16	16
Credit	9	9
Abalone	7	6
Mushroom	14	14
Ticdata2000	24	23
Musk	29	29
Thyroid	15	14

表6 对象减少时增量式与非增量式算法属性约简集分类精度

Tab. 6 Classification accuracy of attribute reduction sets for incremental and non-incremental algorithms when objects decrease

数据集	算法1		算法3	
	SVM	NB	SVM	NB
Cylinder	85.90	87.05	86.71	89.08
Wdbc	93.48	91.66	93.76	91.87
Credit	84.24	86.48	82.90	86.49
Abalone	91.14	91.50	90.86	90.87
Mushroom	79.52	82.08	83.33	81.30
Ticdata2000	86.10	84.17	85.04	86.49
Musk	90.79	93.03	91.75	92.39
Thyroid	78.88	80.31	79.57	80.95

表7 对象减少时增量式与非增量式算法运算时间比较

Tab. 7 Comparison of computational time for incremental and non-incremental algorithms when objects decrease

数据集	算法1	算法3
Cylinder	2.30	0.23
Wdbc	1.60	0.21
Credit	2.20	0.44
Abalone	6.14	3.07
Mushroom	84.79	15.41
Ticdata2000	1 346.92	63.38
Musk	6 597.86	158.98
Thyroid	389.12	53.67

4.3 本文增量式算法与其他增量式算法的性能比较

为了进一步验证本文增量式算法的优越性,选择了4种同类型的增量式属性约简算法进行比较,参与对比算法的详细信息如表8所示。实验评价主要分为3个方面,分别为增量式属性约简效率、属性约简集长度和属性约简分类精度。表8中,对比增量式算法2只适用于离散型的有序信息系统,该算法进行实验时需将表1各个数据集的数值型属性进行离散化处理。

表8 实验参与对比的增量式属性约简算法

Tab. 8 Incremental attribute reduction algorithms involved in the experimental comparison

算法名称	属性度量方法	简称
基于优势邻域条件熵的动态有序数据增量式属性约简算法 ^[10]	优势邻域条件熵	对比增量式算法1
基于优势条件熵的动态有序数据增量式属性约简算法 ^[9]	优势条件熵	对比增量式算法2
基于加权优势邻域条件熵的动态有序数据增量式属性约简算法 ^[12]	加权优势邻域条件熵	对比增量式算法3
基于量化优势邻域自信息的动态有序数据增量式属性约简算法 ^[13]	量化优势邻域自信息	对比增量式算法4

4.3.1 对象增加时算法性能比较

对表 1 中各数据集,随机选择论域的 50% 对象作为初始对象集,然后将剩余 50% 对象中的 20%、40%、60%、

80% 和 100% 分别添加到初始对象集中,以实现数据集不同量对象的动态增加。图 1 为各数据集在不同对象数据增加比例下不同实验算法的增量式属性约简运行时间。

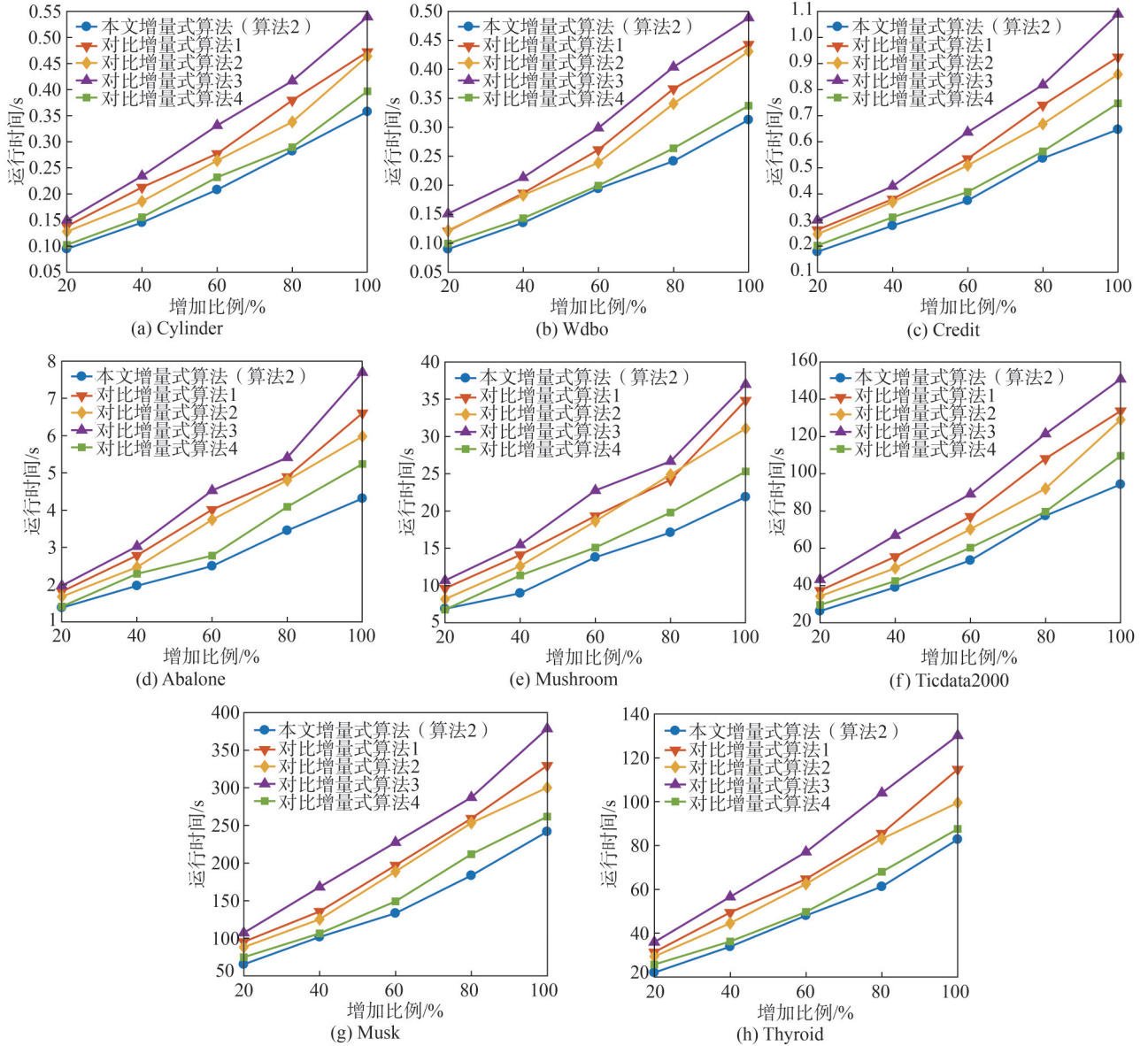


图 1 对象增加时各个增量式算法运行时间比较

Fig. 1 Comparison of execution time for various incremental algorithms with increasing objects

分析图 1 可以发现:

1)随着增加对象的百分比不断增加,所有实验算法的运行时间整体上呈现出逐渐增长的趋势。

2)对比增量式算法 3 在各个数据集下的约简用时最长,主要由于该算法在增量式计算过程中进行了属性权重的计算,因而增加了额外的开销。

3)本文的增量式算法(算法 2)有着最少的运行时间,说明本文算法有着最高的约简效率,这主要是由于优势邻域相对决策熵主要以优势邻域上下近似集为基础,本文算法采用矩阵的策略直接去增量式更新上下近似集,而无需对每个优势邻域类进行更新计

算,相比较于对比算法具有更高的计算效率,因而运算耗时更少。

表 9 为初始数据集增加 100% 剩余对象集后所有实验算法的属性约简集结果,表 10 为属性约简集下 SVM 和 NB 分类精度结果。

由表 9 可知:本文增量式算法在数据集 Credit、Abalone 和 Thyroid 的约简集长度最小;对比增量式算法 2 有着最大的约简集长度,这主要是由于该算法仅适用于离散型的信息系统,而原始数据集的离散化处理丢失了一定的数值信息,因此需要更多的属性去支持分类信息的一致性^[9]。

表9 对象增加时各个增量式算法属性约简集长度比较

Tab. 9 Comparison of attribute reduction set lengths for various incremental algorithms when objects increase

数据集	对比增量式算法1	对比增量式算法2	对比增量式算法3	对比增量式算法4	本文增量式算法(算法2)
Cylinder	13	15	12	13	13
Wdbc	17	19	15	16	16
Credit	9	10	9	9	9
Abalone	6	8	6	6	6
Mushroom	11	14	11	12	12
Ticdata2000	21	24	20	22	21
Musk	28	29	26	27	28
Thyroid	12	15	12	12	12

由表 10 可知:本文增量式算法整体上具有最高的精度结果,而对比增量式算法 3 的分类精度结果接近或仅次于本文增量式算法,这主要是由于对比增量式算法 3 采用属性加权的方法进行属性约简,在约简的选择过程中进行了属性权重的排序和过滤,因此,其选择的约简结果同样具有较高的分类性能。对比增量式算法 4 在少部分数据集下有着比本文算法更高的分

类精度,例如在数据集 Wdbc 下采用 SVM 分类时的精度、在数据集 Abalone 下采用 NB 分类时的精度,这主要是由于对比增量式算法 4 采用了一种量化优势关系的邻域自信息方法进行属性约简,对数据之间的优势程度进行了量化,在一些优势关系不明显的数据集上会有着更好的属性选择结果,从而使得到的属性约简具有更优的分类效果。

表 10 对象增加时各个增量式算法属性约简分类精度比较

Tab. 10 Comparison of classification accuracy for attribute reduction sets for various incremental algorithms when objects increase %

数据集	对比增量式算法1		对比增量式算法2		对比增量式算法3		对比增量式算法4		本文增量式算法(算法2)	
	SVM	NB	SVM	NB	SVM	NB	SVM	NB	SVM	NB
Cylinder	90.11	84.79	84.11	84.34	90.91	86.37	87.37	88.35	90.98	91.27
Wdbc	92.09	90.11	90.38	87.51	94.62	93.57	95.68	91.79	95.42	95.92
Credit	87.29	85.16	80.62	84.67	87.36	87.96	83.48	85.69	87.40	88.21
Abalone	90.18	91.03	87.08	88.66	93.03	91.43	90.32	93.13	94.44	92.86
Mushroom	85.27	82.00	82.20	78.75	86.29	82.24	82.85	80.56	86.65	82.61
Ticdata2000	85.77	82.85	80.53	82.48	87.40	85.73	84.46	85.75	88.12	87.07
Musk	92.94	91.11	90.32	90.03	93.53	93.37	91.31	92.59	93.66	96.84
Thyroid	76.64	81.76	75.53	79.71	80.48	83.34	79.25	81.72	80.88	83.77

4.3.2 对象减少时算法性能比较

将表 1 中的每个完整数据集作为初始对象集,然后随机分别选择论域的 50% 对象中的 20%、40%、60%、80% 和 100% 从初始对象集中删除,以实现不同量对象的动态减少。

图 2 为各数据集在不同对象数据减少比例下所有实验算法的增量式属性约简运行时间。由图 2 可知:1)随着减少对象的百分比不断增加,所有实验算法的运行时间整体上呈现出逐渐下降趋势;2)对比增量式算法 3 在各数据集下增量式属性约简用时最长,主要是由于该算法需要进行属性权重的计算,增加额

外开销;3)本文的增量式算法有着最少的运行时间,同样是由于本文算法直接增量式更新上下近似集,提升了动态更新约简的效率。

表 11 为初始数据集减少 50% 对象集后所有实验算法的增量式属性约简集长度比较,表 12 为属性约简集结果采用 SVM 和 NB 分类的精度对比。由表 11 和 12 可知:本文增量式算法(算法 3)整体上仍然具有较少的约简属性数量,其分类精度也是整体上最高;对比增量式算法 3 的分类精度接近或仅次于本文增量式算法,其原因同样是由于对比增量式算法 3 进行了属性权重的排序和过滤,因此,选择出的约简子集具有较高的分类精度。

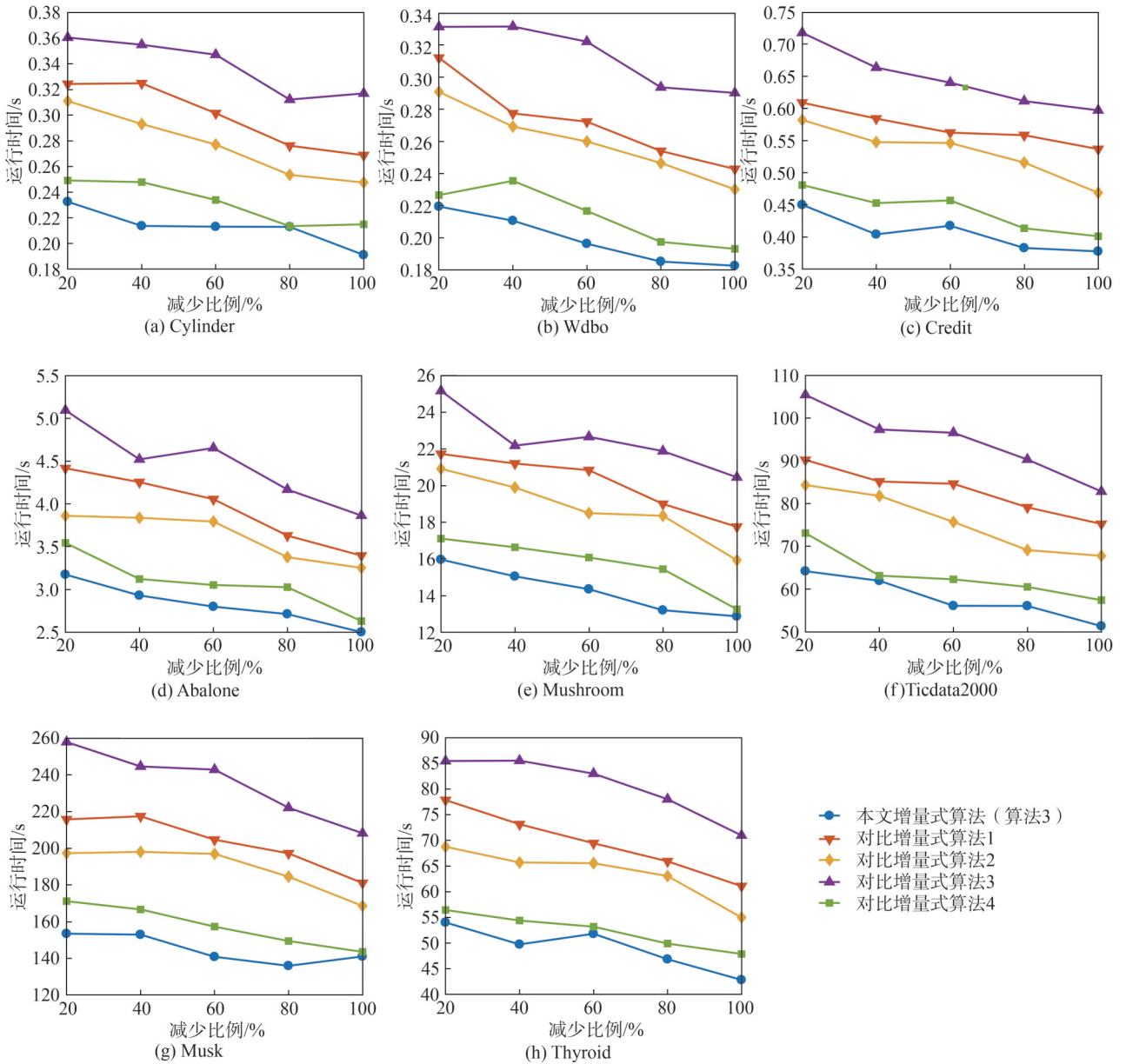


图2 对象减少时各个增量式算法运行时间比较

Fig. 2 Comparison of execution time for various incremental algorithms when objects decrease

表 11 对象减少时各个增量式算法属性约简集长度比较

Tab. 11 Comparison of attribute reduction set lengths for various incremental algorithms when objects decrease

数据集	对比增量式算法1	对比增量式算法2	对比增量式算法3	对比增量式算法4	本文增量式算法(算法3)
Cylinder	15	16	12	14	14
Wdbc	16	17	14	16	16
Credit	9	9	9	9	9
Abalone	6	7	6	6	6
Mushroom	14	15	14	15	14
Ticdata2000	24	26	22	23	23
Musk	29	32	26	28	29
Thyroid	15	25	13	14	14

表 12 对象减少时各个增量式算法属性约简分类精度比较

Tab. 12 Comparison of classification accuracy for attribute reduction sets for various incremental algorithms when objects decrease %

数据集	对比增量式算法 1		对比增量式算法 2		对比增量式算法 3		对比增量式算法 4		本文增量式算法(算法 3)	
	SVM	NB	SVM	NB	SVM	NB	SVM	NB	SVM	NB
Cylinder	87.11	87.62	83.57	84.73	90.09	88.80	86.71	89.08	90.69	89.69
Wdbc	91.45	90.13	90.47	88.76	91.48	92.13	93.76	91.87	96.95	94.57
Credit	81.65	84.43	80.77	83.59	85.50	87.37	82.90	86.49	86.45	87.82
Abalone	88.59	90.42	88.10	87.08	90.61	92.58	90.86	90.87	91.91	93.12
Mushroom	83.74	79.09	82.09	78.47	84.04	83.94	83.33	81.30	86.08	84.66
Ticdata2000	87.19	85.99	83.21	82.84	87.22	86.82	85.04	86.49	88.58	90.78
Musk	87.85	88.75	87.65	87.81	90.65	94.92	91.75	92.39	92.37	96.65
Thyroid	79.81	80.13	76.62	77.58	81.79	83.51	79.57	80.95	83.02	83.93

4.4 本文增量式算法适用性讨论

进一步分析讨论本文增量式算法的适用性。从 UCI 数据集库中选择了 4 个高维数据集,如表 13 所示。在这些数据集中,其属性的数量都远大于对象的数量,通过这些数据集进一步验证本文增量式算法的实验性能。

表 13 实验数据集 2

Tab. 13 Experimental dataset 2

数据集名称	对象数量	属性数量	类别数量	类型
DrivFace	606	6 400	2	混合型
Arcene	900	10 000	3	混合型
Gas sensor	180	15 000	3	混合型
TCGA Kidney Cancers	1 020	60 660	5	混合型

按照第 4.1、4.2 节的方法构造数据集对象的动态增加和动态减少,将本文的非增量式算法与增量式算法分别进行属性约简,重复实验 10 次,记录每个实验的算法用时,表 14 和 15 分别为高维数据集下对象增加、减少时属性约简的平均用时比较。

表 14 高维数据集下对象增加时增量式与非增量式算法运算时间比较结果

Tab. 14 Comparison of computational time for incremental and non-incremental algorithms when objects increase under high-dimensional datasets

数据集	算法 1	算法 2
DrivFace	2 256	1 540
Arcene	5 215	4 047
Gas sensor	1 093	829
TCGA Kidney Cancers	13 758	11 083

对比表 14 与 4 可知,表 14 中非增量式算法和增量式算法的属性约简用时相差不是特别大,处于相同数量级水平,说明对于高维数据集下对象增加场景,增量式算法的效率相较于非增量式算法并没有提升太多。由表 15 可知,增量式算法的属性约简用时反而要

表 15 高维数据集下对象减少时增量式与非增量式算法运算时间比较结果

Tab. 15 Comparison of computational time for incremental and non-incremental algorithms when objects decrease under high-dimensional datasets

数据集	算法 1	算法 3
DrivFace	476	1 104
Arcene	2 025	6 100
Gas sensor	282	729
TCGA Kidney Cancers	3 707	7 783

多于非增量式算法,这说明对于高维数据集下对象减少场景,增量式算法不仅没有提升效率,反而增加了属性约简的复杂度,消耗了更多的计算时间。因此,综合表 14 和 15 的结果可知,本文所提出的增量式属性约简算法不适合用于高维数据集环境,其增量式的性能没有明显提升,同时可能会出现效率变差的情况。

5 结 论

本文利用优势邻域相对决策熵方法提出了一种有序决策信息系统的增量式属性约简方法。本文的贡献主要包括以下几个方面。首先,将传统的相对决策熵在优势邻域关系下进行推广,提出了优势邻域相对决策熵模型,并利用优势邻域相对决策熵评估属性重要度;然后,研究了优势邻域相对决策熵的矩阵计算方法以及相关性质,并利用优势邻域相对决策熵的矩阵形式设计了一种启发式属性约简算法;最后,分别研究了对象动态增加和动态减少场景下的优势邻域相对决策熵增量式属性约简算法。通过一系列的对比实验,验证了该方法的有效性。在将来的工作和研究中,将进一步探索属性变化以及对象和属性同时变化的增量式属性约简问题。

附录见本刊网络版,扫描标题旁的二维码可阅读网络全文。

参考文献:

- [1] Chen Baoguo, Chen Lei, Deng Ming, et al. Incremental attribute reduction algorithm based on incomplete hybrid order information system[J]. *Advanced Engineering Sciences*, 2024, 56(1): 65–81. [陈宝国, 陈磊, 邓明, 等. 基于不完备混合序信息系统的增量式属性约简[J]. *工程科学与技术*, 2024, 56(1): 65–81.]
- [2] Qu Kanglin, Xu Jiucheng, Hou Qincheng, et al. Feature selection using Information Gain and decision information in neighborhood decision system[J]. *Applied Soft Computing*, 2023, 136: 110100.
- [3] Yang Xin, Yang Yuxuan, Luo Junfang, et al. A unified incremental updating framework of attribute reduction for two-dimensionally time-evolving data[J]. *Information Sciences*, 2022, 601: 287–305.
- [4] Huang Yanyong, Guo Kejun, Yi Xiuwen, et al. Matrix representation of the conditional entropy for incremental feature selection on multi-source data[J]. *Information Sciences*, 2022, 591: 263–286.
- [5] Sang Binbin, Yang Liuzhong, Chen Hongmei, et al. Incremental attribute reduction algorithm in dominance-based rough set[J]. *Computer Science*, 2020, 47(8): 137–143. [桑彬彬, 杨留中, 陈红梅, 等. 优势关系粗糙集增量属性约简算法[J]. *计算机科学*, 2020, 47(8): 137–143.]
- [6] Qamar U, Raza M S. A computationally efficient approximation calculation method in dominance-based rough set approach[J]. *Applied Soft Computing*, 2023, 148: 110926.
- [7] Yang Shuyun, Zhang Hongying, Shi Guang, et al. Attribute reductions of quantitative dominance-based neighborhood rough sets with A-stochastic transitivity of fuzzy preference relations[J]. *Applied Soft Computing*, 2023, 134: 109994.
- [8] Chen Hongmei, Li Tianrui, Cai Yong, et al. Parallel attribute reduction in dominance-based neighborhood rough set[J]. *Information Sciences*, 2016, 373: 351–368.
- [9] Sang Binbin, Chen Hongmei, Yang Lei, et al. Incremental attribute reduction approaches for ordered data with time-evolving objects[J]. *Knowledge-Based Systems*, 2021, 212: 106583.
- [10] Sang Binbin, Chen Hongmei, Li Tianrui, et al. Incremental approaches for heterogeneous feature selection in dynamic ordered data[J]. *Information Sciences*, 2020, 541: 475–501.
- [11] Sang Binbin, Chen Hongmei, Yang Lei, et al. Incremental feature selection using a conditional entropy based on fuzzy dominance neighborhood rough sets[J]. *IEEE Transactions on Fuzzy Systems*, 2022, 30(6): 1683–1697.
- [12] Pan Yanzhou, Xu Weihua, Ran Qinwen. An incremental approach to feature selection using the weighted dominance-based neighborhood rough sets[J]. *International Journal of Machine Learning and Cybernetics*, 2023, 14(4): 1217–1233.
- [13] Yang Lei, Qin Keyun, Sang Binbin, et al. A novel incremental attribute reduction by using quantitative dominance-based neighborhood self-information[J]. *Knowledge-Based Systems*, 2023, 261: 110200.
- [14] Zhang Qinli, Chen Yiyang, Zhang Gangqiang, et al. New uncertainty measurement for categorical data based on fuzzy information structures: An application in attribute reduction[J]. *Information Sciences*, 2021, 580: 541–577.
- [15] Jiang Feng, Sui Yuefei, Zhou Lin. A relative decision entropy-based feature selection approach[J]. *Pattern Recognition*, 2015, 48(7): 2151–2163.
- [16] Xing Jinming, Gao Can, Zhou Jie. Weighted fuzzy rough sets-based tri-training and its application to medical diagnosis[J]. *Applied Soft Computing*, 2022, 124: 109025.
- [17] Huang Haixin, Li Zhaowen, Liu Fang, et al. New uncertainty measurement for hybrid data and its application in attribute reduction[J]. *Information Sciences*, 2024, 664: 120334.
- [18] Zhang Chucai, Lu Zhengxiang, Dai Jianhua. Incremental attribute reduction for dynamic fuzzy decision information systems based on fuzzy knowledge granularity[J]. *Information Sciences*, 2025, 689: 121467.
- [19] Thuy N N, Wongthanavas S. Attribute reduction with fuzzy divergence-based weighted neighborhood rough sets[J]. *International Journal of Approximate Reasoning*, 2024, 173: 109256.
- [20] Vo B K, Nguyen H S. An attribute ranking method based on rough sets and interval-valued fuzzy sets[J]. *International Journal of Approximate Reasoning*, 2024, 170: 109215.
- [21] Zhao Jie, Wu Daiyang, Wu Jiabin, et al. Consistency approximation: Incremental feature selection based on fuzzy rough set theory[J]. *Pattern Recognition*, 2024, 155: 110652.
- [22] Hu Chengxiang, Zhang Li, Huang Xiaoling, et al. Matrix-based approaches for updating three-way regions in incomplete information systems with the variation of attributes[J]. *Information Sciences*, 2023, 639: 119013.
- [23] Wen Haotong, Xu Yi, Liang Meishe. Parallel attribute reduction algorithm for unlabeled data based on fuzzy discernibility matrix and soft deletion behavior[J]. *Information Sciences*, 2025, 689: 121472.
- [24] Xu Weihua, Bu Qinyuan. Matrix-based incremental feature selection method using weight-partitioned multigranulation rough set[J]. *Information Sciences*, 2024, 681: 121219.
- [25] Zhang Xiaoyan, Wang Jinghong, Hou Jianglong. Matrix-based approximation dynamic update approach to multi-granulation neighborhood rough sets for intuitionistic fuzzy ordered datasets[J]. *Applied Soft Computing*, 2024, 163: 111915.
- [26] Xie Jingjing, Hu Baoqing, Jiang Haibo. A novel method to at-

tribute reduction based on weighted neighborhood probabilistic rough sets[J].International Journal of Approximate Reasoning,2022,144:1–17.

[27] Zhang Tengfei,Zhang Yudi,Ma Fumin.Improved neighborhood space based feature selection algorithm for high-dimensional mixed data[J].Control and Decision,2024,39(3):

929–938.[张腾飞,张宇迪,马福民.基于改进邻域空间的高维混合数据特征选择算法[J].控制与决策,2024,39(3):929–938.]

[28] Zhang Di,Zhu Ping.A novel approach to discretizing information systems associated with neighborhood rough sets[J].International Journal of Approximate Reasoning,2024,164:109062.

Incremental Attribute Reduction Algorithm for Dominance-based Neighborhood Relative Decision Entropy

CHEN Baoguo¹, CHEN Lei^{1*}, DENG Ming¹, LI Xiaoyan¹, CHEN Jinlin^{1,2}

(1.School of Computer Science, Huainan Normal University, Huainan 232038, China;

2.College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

Abstract:

Objective In a big data environment, data is continuously and dynamically updated, which imposes significant limitations and challenges on traditional machine learning algorithms. Incremental learning is a learning paradigm that focuses exclusively on newly changed data based on the learning outcomes of existing models, substantially improving learning efficiency during data update processes. Incremental attribute reduction is a widely adopted method and strategy for achieving efficient attribute reduction in dynamic dataset environments. However, in ordered information systems with dynamically updated and evolving objects, existing incremental attribute reduction methods neglect the potential classification information provided by rough approximation sets, which limits their effectiveness in supporting attribute reduction.

Methods In existing research results, scholars jointly proposed a relative decision entropy model by utilizing the roughness and dependency of upper and lower approximation sets to combine information entropy. This entropy model considered both the classification information of the upper and lower approximation sets of attributes and the classification information from the perspective of attribute information entropy, and it demonstrated better performance in attribute uncertainty measurement and attribute reduction. At the same time, the basic calculation unit of relative decision entropy was the upper and lower approximation sets of attributes. When the object changed, there was no need to calculate the information granules of the updated object. Therefore, this study extended relative decision entropy to mixed ordered information system environments. First, the relative decision entropy model of dominance-based neighborhoods was proposed. Then, the relative decision entropy of dominance-based neighborhoods was constructed based on the dominance-based neighborhood relation. The relative decision entropy of dominance-based neighborhoods was reconstructed in the form of a matrix, and a non-incremental attribute reduction algorithm for hybrid ordered information systems was designed. Finally, for the two scenarios of increasing and decreasing objects in a hybrid ordered information system, the matrix-form incremental update of the relative decision entropy of the dominance-based neighborhood was analyzed and studied, and incremental attribute reduction algorithms were constructed using this update mechanism.

Results and Discussions In the experimental stage, 8 public datasets were selected for simulation experiments to compare the incremental algorithm with non-incremental algorithms. In the case of increasing the number of objects in the dataset, the number of reduced attributes for the non-incremental and incremental algorithms was basically similar across the 8 datasets, with averages of 14.75 and 14.62 attributes, respectively. The classification accuracy of the non-incremental algorithm and the incremental algorithm was also basically similar across the 8 datasets, with average SVM classification accuracies of 86.12% and 86.65%, and average NB classification accuracies of 87.22% and 87.24%, respectively. The processing time of the incremental algorithms was significantly shorter than that of the non-incremental algorithms. The processing time of the non-incremental algorithms on the 8 datasets was 4 215.13 seconds, whereas the processing time of the incremental algorithms on the 8 datasets was only 55.38 seconds. The performance of the incremental algorithms was significantly superior. In the case of reducing dataset objects, the number of reduced attributes for the non-incremental and incremental algorithms was basically similar across the 8 datasets, with averages of 16.00 and 15.62, respectively. The classification accuracy of the non-incremental algorithm and the incremental algorithm was basically similar across the 8 datasets, with average SVM classification accuracies of 86.25% and 86.74%, and average NB classification accuracies of 87.03% and 87.48%, respectively. The processing time of the incremental algorithms was significantly shorter than that of the non-incremental algorithms, with the non-incremental algorithms consuming 1 053.56 seconds and the incremental algorithms consuming only 36.92 seconds. The effectiveness of the incremental algorithm was verified through these experimental results. An experimental comparison was conducted between the incremental algorithm proposed in this study and four superior comparative algorithms to verify the superiority of the algorithm. In the case of increasing the number of objects in the dataset, the number of reduced attributes for the four comparative algorithms and the algorithm proposed in this study were 14.76, 16.75, 14.87, 14.62, and 14.62, respectively. The SVM classification accuracies of the four comparative algorithms and the al-

gorithm proposed in this study were 87.53%, 83.84%, 89.20%, 86.84%, and 89.69%, respectively. The NB classification accuracies of the four comparative algorithms and the algorithm proposed in this study were 86.10%, 84.51%, 88.00%, 87.44%, and 89.81%, respectively. In terms of processing time, the proposed algorithm was 35%, 32%, 40%, and 7% faster than the four compared algorithms, respectively. In the case of reducing dataset objects, the number of reduced attributes for the four comparative algorithms and the algorithm proposed in this study were 16.00, 18.37, 15.90, 15.82, and 15.62, respectively. The SVM classification accuracies of the four comparative algorithms and the algorithm proposed in this study were 85.92%, 84.06%, 87.67%, 86.74%, and 89.50%, respectively. The NB classification accuracies of the four comparative algorithms and the algorithm proposed in this study were 85.82%, 83.85%, 88.75%, 87.43%, and 90.15%, respectively. In terms of processing time, the proposed algorithm was 55%, 46%, 63%, and 12% faster than the four compared algorithms, respectively. Comparing these results demonstrated that the incremental algorithm in this study selected fewer attributes, achieved higher classification accuracy, and significantly outperformed the comparative algorithms in terms of efficiency.

Conclusions The experimental results demonstrate that the proposed incremental algorithm exhibits superior attribute reduction performance on dynamic datasets, significantly enhancing the efficiency of dynamic attribute reduction while maintaining the number of attribute selections and classification accuracy. At the same time, the proposed incremental algorithm selects fewer attributes and achieves higher classification accuracy compared to similar algorithms. Most importantly, the algorithm demonstrates higher computational performance.

Key words: attribute reduction; object changes; incremental; ordered information system; dominance-based neighborhood relative decision entropy

(编辑 吴芝明)

引用格式: Chen Baoguo, Chen Lei, Deng Ming, et al. Incremental attribute reduction algorithm for dominance-based neighborhood relative decision entropy[J]. *Advanced Engineering Sciences*, 2026, 58(2): 372–386. [陈宝国, 陈磊, 邓明, 等. 基于优势邻域相对决策熵的增量式属性约简算法[J]. *工程科学与技术*, 2026, 58(2): 372–386.]