

文章编号: 1000-5641(2018)03-0097-12

近似边界精度信息熵的属性约简

梁宝华^{1,2}, 吴其林^{1,2}

(1. 巢湖学院 信息工程学院, 合肥 238000;
2. 巢湖学院 网络与分布式研究所, 合肥 238000)

摘要: 针对信息观只考虑知识粒度的大小, 不能客观、全面度量属性重要性的不足, 首先从代数观出发, 提出近似边界精度的定义; 其次, 根据相对模糊熵的定义, 提出相对信息熵及增强信息熵概念, 与相对模糊熵相比具有明显的放大作用; 再次, 将近似边界精度融合相对信息熵和增强信息熵, 提出两种新的属性约简方法, 在求 $U/(B \cup b)$ 时充分利用 U/B 的结果, 可大大减少系统的时间开销; 最后, 通过实验分析和比较, 本文算法在约简质量、分类精度上的可行性和有效性得到了验证.

关键词: 粗糙集; 属性约简; 近似边界精度; 相对信息熵; 增强信息熵

中图分类号: TP18 文献标志码: A DOI: 10.3969/j.issn.1000-5641.2018.03.011

Attribute reduction based on information entropy of approximation boundary accuracy

LIANG Bao-hua^{1,2}, WU Qi-lin^{1,2}

(1. College of Information Engineering, Chaohu University, Hefei 238000, China;
2. Institute of Networks and Distributed System, Chaohu University, Hefei 238000, China)

Abstract: From an information point of view, only the size of knowledge granularity is taken into account, while the importance of attributes cannot be objectively and comprehensively measured. First, starting from the perspective of algebra, the concept of approximate boundary accuracy is proposed. Afterwards, according to the definition of relative fuzzy entropy, this paper proposes two new concepts for relative information entropy and enhanced information entropy. Compared with relative fuzzy entropy, the proposed information entropy has an obvious magnification effect. Two new methods of attribute reduction are subsequently proposed by incorporating approximate boundary accuracy into relative information entropy and enhanced information entropy. Computing $U/(B \cup b)$ while making full use of U/B can greatly reduce the computational overhead on time. Finally, through the experimental analysis and comparison, it is validated that the proposed algorithm has feasibility and effectiveness in both reduction quality and

收稿日期: 2017-04-21

基金项目: 安徽省自然科学基金(1308085MF101); 安徽省高校自然科学重点研究项目(KJ2016A502);
安徽省高校优秀青年国内外访学研修项目(gxfx2017100)

作者简介: 梁宝华, 男, 硕士, 副教授, 研究方向为粗糙集、数据挖掘. E-mail: liangbh426@126.com.
吴其林, 男, 博士, 副教授, 主要研究方向为无线网络. E-mail: lingqiw@126.com.

classification accuracy.

Keywords: rough set; attribute reduction; approximation boundary accuracy; relative information entropy; enhanced information entropy

0 引言

粗糙集理论^[1-2]是模糊信息系统中处理不精确、非完备、非协调信息等典型的数学工具。目前,已成功应用于数据挖掘、机器学习、规则生成、故障诊断、风险预测等领域^[3-7]。

属性约简是粗糙集理论研究的核心内容之一,旨在保持区分能力不变的情况下,剔除冗余的属性,用最少的属性代替原有样本,从而达到约简的效果。近年来,粗糙集研究者们提出了多种属性重要性的度量方法,并以此为基础设计很多启发式属性约简算法。对于启发式属性约简算法可归纳为两类:代数观和信息观。前者以不可分辨关系为基础,研究属性对论域中确定分类子集的影响,以代数式方式定义属性重要性,主要算法有基于正区域算法^[8-10]和基于差别矩阵算法^[11-13]。后者从信息粒度考虑,研究属性对论域中不确定分类子集的影响、有效度量知识的粒度大小。

信息熵是由香农于1948年提出的度量物质状态变化的概念^[14],后来逐渐被研究者们引入粗糙集中,用于不确定性信息度量。相应地,基于熵的属性约简也应运而生^[15-21]。王国胤^[15]给出了基于条件信息熵的属性约简算法;杨明^[16]提出了“基于条件信息熵的近似约简”,算法依据实际应用的需要有效地对冗余属性进行取舍,可有效增强抗噪声、抗干扰的能力,对基于条件信息熵的决策表属性约简起到了完善和补充作用;张清华^[17]根据模糊熵及模糊度量不确定性的特点,提出了在决策系统中修正条件信息熵和相对模糊熵的概念,根据熵在知识粒度细分过程中表现的单调性,证明两种熵的合理性,利用向前添加算法得到熵的约简算法,为其他信息熵约简质量提供了参考。

现有的基于熵的属性约简方法,多数仅从信息论的角度度量属性重要性,很少有人有效结合代数观来定义属性重要性。王国胤在文献[15]中指出并证明了代数观与信息论定义的属性重要性具有较强的互补性:代数观考虑确定分类子集的影响,而信息论考虑不确定分类子集的影响。属性重要性度量方法的客观性,直接影响到属性约简结果及系统的时间消耗。为了快速、准确找到最重要的属性,有必要综合考虑多种因素,有机结合代数观及信息论的特点,定义一种更加全面的属性重要性度量方法。另外,现有的基于熵的属性约简算法有较高的时间复杂度,还有待于进一步改进。

针对上述问题,文献[19]提出了近似决策熵,有效结合代数观下的近似精度与信息论下的条件熵,得到了一种较全面的属性重要性度量方法。为进一步探究客观、全面的度量方法,本文提出了一种新的代数观下的信息熵模型——近似边界精度信息熵,有效地结合了代数观与信息论的特点。首先在代数观下计算正区域对边界精度的影响,提出并证明了近似边界精度与近似精度具有相同单调性,同时还提出了有放大作用的“修正的近似边界精度”;其次,从信息论角度考虑,以文献[17]的相对模糊熵为基础,提出了相对信息熵和增强信息熵概念,证明了本文提出的两种信息熵的单调性与合理性,同时比相对模糊熵具有放大属性重要的作用;然后,将修正近似边界精度与两种熵相融合,产生了新的约简方法;最后,通过实验与6种同类算法在约简时间和约简剩余属性数方面进行了比较,并将约简结果在SVM分类器上利用径向基核进行了相关参数设定,通过十折交叉法进行了分类精度的验证。

本文基于近似边界精度信息熵提出相对信息熵约简算法(Approximate Boundary Accu-

racy Entropy, ABAE) 及增强信息熵约简算法 (ABAE'). 在约简过程中, 由于近似边界精度相对信息熵及增强信息熵呈现非严格单调情况, 从而为算法的合理性提供依据. 同时, 新模型在同等条件下衡量属性重要性时, 比文献 [17] 的相对模糊熵有明显的放大作用, 为新算法能够快速选择出最重要的属性提供有效的途径. 由于本文算法充分考虑了确定分类和不确定分类对约简的影响, 所以新的信息熵模型更全面; 本文的实验数据分析有力证实了新模型下的约简算法的可行性、有效性. 另外, 根据实验还可得知, 新的信息熵模型下的优势(时间性能及约简质量)及存在的不足(受离群率高低的影响较大), 可为属性约简的算法选择提供合理的参考.

1 粗糙集相关概念

定义1 给定决策表 $DT = (U, A, V, f)$, 其中 U 是一组非空有限的对象集, $A = C \cup D$, C 为条件属性集, D 为决策属性集, $V = \bigcup_{a \in C \cup D} V_a$, V_a 是属性 a 的值域, $f : U \times A \rightarrow V$ 是一个信息函数, 即 $\forall a \in A, x \in U$ 有 $f(x, a) \in V_a$; 属性子集 $B \subseteq A$ 决定一个不可区分关系

$$\text{ind}(B) = \{(x, y) \in U^2 | \forall b \in B, f(x, b) = f(y, b)\}, \quad (1.1)$$

其中, 关系 $\text{ind}(B)$ 将样本 U 按属性 B 划分多个等价类, 可简记为 U/B .

性质1 给定决策表 $DT = (U, C \cup D, V, f)$, 其中 U 是非空有限的对象集, C 为条件属性, D 为决策属性, $B \subseteq C, b \in C - B$, 若 $U/B = \{B_1, B_2, \dots, B_m\}$, 则 $U/(B \cup b) = \bigcup_{i=1}^m B_i/b$, 且 $U/(B \cup b) \subseteq U/B$.

证明 (1) 若 $\forall (x, y) \in U/(B \cup b)$, 根据定义 1 可得, $\forall a \in B \cup b, f(x, a) = f(y, a)$, 即 $\forall a \in B, f(x, a) = f(y, a)$ 且 $f(x, b) = f(y, b)$, 则 $\exists B_i \subseteq U/B$ 使 $(x, y) \in B_i \times B_i$, 又因为 $f(x, b) = f(y, b)$, 所以 $(x, y) \in B_i^2$ 且 $(x, y) \in U/b$, 由于 $U/B = \{B_1, B_2, \dots, B_m\}$, 故 $U/(B \cup b) \subseteq \bigcup_{i=1}^m B_i/b$.

(2) 另外, 假设 $\forall (x, y) \in B_i/b, B_i$ 中所有样本在属性 B 上不可区分, B_i/b 表示将数据集 B_i 上的元素再按属性 b 细划, 因 $(x, y) \in B_i/b$, 所以 $f(x, B \cup b) = f(y, B \cup b)$, 根据定义 1 可得, $(x, y) \in U/(B \cup b)$, 所以 $B_i/b \subseteq U/(B \cup b)$, 因 (x, y) 具有任意性, 故 $\bigcup_{i=1}^m B_i/b \subseteq U/(B \cup b)$.

综合(1)(2)所得, $U/(B \cup b) = \bigcup_{i=1}^m B_i/b$.

(3) U/B 表示将样本 U 按属性 B 进行划分, $U/(B \cup b)$ 将样本 U 按属性 $B \cup b$ 划分, 由于 $B \subseteq (B \cup b)$, 所以 $U/(B \cup b)$ 划分是 U/B 划分的加细, 即 $U/(B \cup b) \subseteq U/B$, 也即 $U/(B \cup b)$ 与 U/B 有非严格的单调性.

定义2 给定决策表 $DT = (U, C, D, V, f)$, $\forall B \subseteq C \cup D, \forall X \subseteq U$, 集合 X 关于属性集 B 的下近似 $\underline{B}(X)$ 、上近似 $\overline{B}(X)$ 、正区域 $POS_B(X)$ 、边界域 $BND(X)$ 、负区域 $NEG(X)$, 有

$$\overline{B}(X) = \{x \in U | [x]_B \cap X \neq \emptyset\} = \bigcup \{[x]_B \in U/B | [x]_B \cap X \neq \emptyset\}, \quad (1.2)$$

$$\underline{B}(X) = \{x \in U | [x]_B \subseteq X\} = \bigcup \{[x]_B \in U/B | [x]_B \subseteq X\}. \quad (1.3)$$

显然, 粗糙集的上、下近似集将论域分割成正区域、负区域及边界域, 其中, 正区域 $POS_B(X) = \underline{B}(X)$, 边界域 $BND(X) = \overline{B}(X) - \underline{B}(X)$, 负区域 $NEG_B(X) = U - \overline{B}(X)$.

定义 3 给定决策表 $DT = (U, C, D, V, f)$, $\forall B \subseteq C \cup D$, $\forall X \subseteq U$, 集合 X 在 $\text{ind}(B)$ 下的近似精度定义为^[19]

$$\alpha_B(X) = \frac{|B(X)|}{|\overline{B(X)}|}. \quad (1.4)$$

定义 4 给定决策表 $DT = (U, C, D, V, f)$, $\forall B \subseteq C \cup D$, $\forall X \subseteq U$, 集合 X 在 $\text{ind}(B)$ 下的近似边界精度定义为

$$\beta_B(X) = \frac{|\text{POS}_B(X)|}{|\text{BND}(X)|}. \quad (1.5)$$

性质 2 给定决策表 $DT = (U, C, D, V, f)$, $\forall B \subseteq C \cup D$, $\forall X \subseteq U$, 近似精度 $\alpha_B(X)$ 与近似边界精度 $\beta_B(X)$ 具有相同的单调性, 且 $\beta_B(X) \geq \alpha_B(X)$.

证 明

$$\beta_B(X) = \frac{|\text{POS}_B(X)|}{|\text{BND}(X)|} = \frac{|B(X)|}{\left|\overline{B(X)}\right| - |B(X)|} = \frac{\frac{|B(X)|}{|B(X)|}}{1 - \frac{|B(X)|}{|B(X)|}} = \frac{\alpha_B(X)}{1 - \alpha_B(X)}.$$

(1) 当 $\alpha_B(X)$ 增大时, $1 - \alpha_B(X)$ 减小, 则 $\frac{\alpha_B(X)}{1 - \alpha_B(X)}$ 增大; 当 $\alpha_B(X)$ 减少时, $1 - \alpha_B(X)$ 增大, 则 $\frac{\alpha_B(X)}{1 - \alpha_B(X)}$ 减小, 所以 $\beta_B(X)$ 与 $\alpha_B(X)$ 单调性一致.

$$(2) \beta_B(X) - \alpha_B(X) = \frac{\alpha_B(X)}{1 - \alpha_B(X)} - \alpha_B(X) - \frac{\alpha_B(X)^2}{1 - \alpha_B(X)} \geq 0$$

综合(1)(2), $\alpha_B(X)$ 与 $\beta_B(X)$ 具有相同的单调性, 且 $\beta_B(X) \geq \alpha_B(X)$, 证毕.

为了计算范围统一, 且控制在 [0–1] 之间, 将近似边界精度修正, 定义如下.

定义 5 给定决策表 $DT = (U, C, D, V, f)$, $\forall B \subseteq C \cup D$, $\forall X \subseteq U$, 集合 X 在 $\text{ind}(B)$ 修正近似边界精度 $\beta'_B(X) = \alpha_B(X)(2 - \alpha_B(X))$, 很显然 $\beta'_B(X) = -(\alpha_B(X) - 1)^2 + 1$, 当 $\alpha_B(X) \in [0, 1]$ 时, $\beta'_B(X) \in [0, 1]$.

性质 3 给定决策表 $DT = (U, C, D, V, f)$, $\forall B \subseteq C \cup D$, $\forall X \subseteq U$, 近似精度 $\alpha_B(X)$ 与修正近似边界精度 $\beta'_B(X)$ 具有相同的单调性, 且 $\beta'_B(X) \geq \alpha_B(X)$.

证 明 (1) 因 $\beta'_B(X) = -(\alpha_B(X) - 1)^2 + 1$, 当 $\alpha_B(X)$ 时, 若 $\alpha_B(X)$ 增大, 则 $(\alpha_B(X) - 1)^2$ 减小, $-(\alpha_B(X) - 1)^2$ 会增大, 即 $\beta'_B(X)$ 增大; 同理, 当 $\alpha_B(X)$ 减小时, $\beta'_B(X)$ 减小, 所以 $\alpha_B(X)$ 与 $\beta'_B(X)$ 具有相同的单调性.

(2) $\beta'_B(X) - \alpha_B(X) = \alpha_B(X)(2 - \alpha_B(X)) - \alpha_B(X) = \alpha_B(X) - \alpha_B(X)^2 = \alpha_B(X)(1 - \alpha_B(X))$. 因 $\alpha_B(X) \in [0, 1]$, $\alpha_B(X)(1 - \alpha_B(X)) \geq 0$, 所以 $\beta'_B(X) - \alpha_B(X) \geq 0$ 即 $\beta'_B(X) \geq \alpha_B(X)$, 证毕.

2 近似边界精度信息熵

2.1 相对信息熵、增强信息熵

定义 6^[17] 给定决策表 $DT = (U, C, D, V, f)$, 令 U 在决策属性 D 上的划分为 $U/D = \{Y_1, Y_2, \dots, Y_m\}$, 对于 $\forall B \subseteq C$, 令 U 在条件属性 B 上的划分为 $U/B = \{X_1, X_2, \dots, X_n\}$, 决策概念集相对知识 B 的修正条件信息熵为 $H(D_B) = -\sum_{j=1}^m \sum_{i=1}^{|U|} [\mu_{Y_j}^B(x_i) \ln \mu_{Y_j}^B(x_i)] + (1 -$

$\mu_{Yj}^B(x_i))lb(1 - \mu_{Yj}^B(x_i))]$, 相对模糊熵为 $E(D_B) = \sum_{j=1}^m \sum_{i=1}^{|U|} \mu_{Yj}^B(x_i)(1 - \mu_{Yj}^B(x_i))$, 其中, $\mu_Y^B(x) = \frac{|\{x\}_P \cap Y|}{|\{x\}_P|}$ 表示知识 B 概念下的粗糙隶属度函数. 相对模糊熵的等价形式为

$$E'(D_B) = \sum_{j=1}^n \sum_{i=1}^m |X_i| \frac{|X_i \cap Y_j|}{|X_i|} \left(1 - \frac{|X_i \cap Y_j|}{|X_i|}\right), \quad (2.1)$$

修正条件信息熵的等价形式为

$$\begin{aligned} H'(D_B) = & - \sum_{i=1}^n |X_i| \times \sum_{j=1}^m \left[\frac{|X_i \cap X_j|}{|X_i|} lb \frac{|X_i \cap X_j|}{|X_i|} \right. \\ & \left. + \left(1 - \frac{|X_i \cap X_j|}{|X_i|}\right) lb \left(1 - \frac{|X_i \cap X_j|}{|X_i|}\right) \right]. \end{aligned} \quad (2.2)$$

定义7 给定决策表 $DT = (U, C, D, V, f)$, 对于 $\forall B \subseteq C$, 令 U 在条件属性 B 上的划分为 $U/B = \{X_1, X_2, \dots, X_n\}$, 令 U 在决策属性 D 上的划分为 $U/D = \{Y_1, Y_2, \dots, Y_m\}$, 称

$$AE(D_B) = \begin{cases} \sum_{j=1}^m \sum_{i=1}^n |X_i| \left[1 - P(Y_j|X_i) / \overline{P}(Y_j|X_i) \right], & 0 \leq P(Y_j|X_i) \leq \frac{1}{2}, \\ \sum_{j=1}^m \sum_{i=1}^n |X_i| \left[1 - \overline{P}(Y_j|X_i) / P(Y_j|X_i) \right], & \frac{1}{2} < P(Y_j|X_i) \leq 1 \end{cases} \quad (2.3)$$

为决策属性 D 相对于条件属性 B 的相对信息熵, 其中, $AE(Amplify Entropy)(D_B)$ 是扩大信号重要性, $P(Y_j|X_i) = \frac{|X_i \cap Y_j|}{|X_i|}$, $\overline{P}(Y_j|X_i) = 1 - \frac{|X_i \cap Y_j|}{|X_i|}$, $AE(D_B)$ 的等价形式为

$$AE(D_B) = \begin{cases} \sum_{j=1}^m \sum_{i=1}^n |X_i| \left[1 - \frac{|X_i \cap Y_j|}{|X_i|} / \left(1 - \frac{|X_i \cap Y_j|}{|X_i|}\right) \right], & 0 \leq \frac{|X_i \cap Y_j|}{|X_i|} < \frac{1}{2}, \\ \sum_{j=1}^m \sum_{i=1}^n |X_i| \left[1 - \left(1 - \frac{|X_i \cap Y_j|}{|X_i|}\right) / \frac{|X_i \cap Y_j|}{|X_i|} \right], & \frac{1}{2} \leq \frac{|X_i \cap Y_j|}{|X_i|} \leq 1. \end{cases} \quad (2.4)$$

定理1 给定决策表 $DT = (U, C, D, V, f)$ 中, 在定义7的条件下, 若 $U/B_{i+1} \prec U/B_i$ (其中, 则 $B_i, B_{i+1} \subseteq C$), 则 $AE(D_{B_{i+1}}) \leq AE(D_{B_i})$.

证 明 设某等价关系在 U 上形成的划分为 $U/B_{i+1} = \{X_1, X_2, \dots, X_n\}$, 而 $U/B_i = \{X_1, X_2, \dots, X_{k-1}, X_{k+1}, \dots, X_{l-1}, X_{l+1}, \dots, X_n, X_k \cup X_l\}$ 是将划分 B_{i+1} 中的某两个划分 X_k 与 X_l 合并为 $X_k \cup X_l$ 得到的新划分, $D = \{Y_1, Y_2, \dots, Y_m\}$ 也是 U 上的一个划分.

(1) 当 $0 \leq \frac{|X_i \cap Y_j|}{|X_i|} < \frac{1}{2}$ 时,

$$\begin{aligned} AE(D_{B_i}) = & AE(D_{B_{i+1}}) + |X_k \cup X_l| \times \sum_{j=1}^m \left[1 - \frac{|(X_k \cup X_l) \cap Y_j|}{|X_k \cup X_l|} / \left(1 - \frac{|(X_k \cup X_l) \cap Y_j|}{|X_k \cup X_l|}\right) \right] \\ & - |X_k| \sum_{j=1}^m \left[1 - \frac{|X_k \cap Y_j|}{|X_k|} / \left(1 - \frac{|X_k \cap Y_j|}{|X_k|}\right) \right] - |X_l| \sum_{j=1}^m \left[1 - \frac{|X_l \cap Y_j|}{|X_l|} / \left(1 - \frac{|X_l \cap Y_j|}{|X_l|}\right) \right], \end{aligned}$$

$$AE_\Delta = AE(D_{B_i}) - AE(D_{B_{i+1}})$$

$$= \sum_{j=1}^m \left[|X_k \cup X_l| \times \left(1 - \frac{|(X_k \cup X_l) \cap Y_j|}{|X_k \cup X_l|} / \left(1 - \frac{|(X_k \cup X_l) \cap Y_j|}{|X_k \cup X_l|}\right)\right) \right]$$

$$\begin{aligned}
& -|X_k| \times \left(1 - \frac{|X_k \cap Y_j|}{|X_k|} \right) / \left(1 - \frac{|X_k \cap Y_j|}{|X_k|} \right) \\
& -|X_l| \times \left(1 - \frac{|X_l \cap Y_j|}{|X_l|} \right) \left(1 - \frac{|X_l \cap Y_j|}{|X_l|} \right) \\
= & |X_k \cup X_l| \times \left(1 - \frac{|(X_k \cup X_l) \cap Y_j|}{|X_k \cup X_l| - |(X_k \cup X_l) \cap Y_j|} \right) \\
& -|X_k| \times \left(1 - \frac{|X_k \cap Y_j|}{|X_k| - |X_k \cap Y_j|} \right) -|X_l| \times \left(1 - \frac{|X_l \cap Y_j|}{|X_l| - |X_l \cap Y_j|} \right).
\end{aligned}$$

因为 $X_k \cap X_l = \emptyset$, 所以 $|X_k \cup X_l| = |X_k| + |X_l|$, $|(X_k \cup X_l) \cap Y_j| = |X_k \cap Y_j| + |X_l \cap Y_j|$.

设 $|X_k| = x$, $|X_l| = y$, $|X_k \cap Y_j| = ax$, $|X_l \cap Y_j| = by$, 显然有 $x > 0$, $y > 0$, $0 \leq a < \frac{1}{2}$, $0 \leq b < \frac{1}{2}$. 所以,

$$\begin{aligned}
\text{AE}_\Delta &= \sum_{j=1}^m \left[-(x+y) \times \frac{ax+by}{x+y-ax-by} + x \times \frac{ax}{x-ax} + y \times \frac{by}{y-by} \right] \\
&= \sum_{j=1}^m \frac{(ax+by)[xb(1-a)+ya(1-b)]}{(x+y-ax-by) \cdot (1-a) \cdot (1-b)} \geq 0.
\end{aligned}$$

(2) 当 $\frac{1}{2} \leq \frac{|X_i \cap Y_j|}{|X_i|} \leq 1$ 时,

$$\begin{aligned}
\text{AE}(D_{B_{i+1}}) &= \sum_{j=1}^m \sum_{i=1}^n |X_i| \left[1 - \left(1 - \frac{|X_i \cap Y_j|}{|X_i|} \right) \right] / \frac{|X_i \cap Y_j|}{|X_i|} = \sum_{j=1}^m \sum_{i=1}^n |X_i| \left(2 - \frac{|X_i|}{|X_i \cap Y_j|} \right), \\
\text{AE}(D_B) &= \text{AE}(D_{B_{i+1}}) + |X_k \cup X_l| \times \sum_{j=1}^m \left(2 - \frac{|X_k \cup X_l|}{|(X_k \cup X_l) \cap Y_j|} \right) \\
&\quad - |X_k| \sum_{j=1}^m \left(2 - \frac{|X_k|}{|X_k \cap Y_j|} \right) - |X_l| \sum_{j=1}^m \left(2 - \frac{|X_l|}{|X_l \cap Y_j|} \right), \\
\text{AE}_\Delta &= \text{AE}(D_{B_i}) - \text{AE}(D_{B_{i+1}}) \\
&= \sum_{j=1}^m \left[\frac{|X_k|^2}{|X_k \cap Y_j|} + \frac{|X_l|^2}{|X_l \cap Y_j|} - \frac{|X_k \cup X_l|^2}{|(X_k \cup X_l) \cap Y_j|} \right].
\end{aligned}$$

设 $|X_k| = x$, $|X_l| = y$, $|X_k \cap Y_j| = ax$, $|X_l \cap Y_j| = by$, 显然 $x > 0$, $y > 0$, $\frac{1}{2} \leq a \leq 1$, $\frac{1}{2} \leq b \leq 1$,

$$\text{AE}_\Delta = \sum_{j=1}^m \left[\frac{x}{a} + \frac{y}{b} - \frac{x+y}{ax+by} \right] = \sum_{j=1}^m \frac{xy(a-b)^2}{ab(ax+by)} \geq 0.$$

综合(1)(2)得, $\text{AE}(D_{B_{i+1}}) \leq \text{AE}(D_{B_i})$, 证毕.

由于上述在求相对信息熵时, 每次要判断 $\frac{|X_i \cap Y_j|}{|X_i|}$ 的取值范围, 增加系统开销, 下面对相对信息熵进行改进.

定义 8 给定决策表 $DT = (U, C, D, V, f)$, 对于 $\forall B \subseteq C$, 令 U 在条件属性 B 上的划分为 $U/B = \{X_1, X_2, \dots, X_n\}$, 令 U 在决策属性 D 上的划分为 $U/D = \{Y_1, Y_2, \dots, Y_m\}$, 称

$$\text{AE}'(D_B) = \sum_{j=1}^m \sum_{i=1}^n |X_i| \left[1 - \frac{|X_i \cap Y_j|}{2|X_i|} \right] / \left(1 - \frac{|X_i \cap Y_j|}{2|X_i|} \right) \quad (2.5)$$

为决策属性 D 相对于条件属性 B 的增强信息熵.

定理2 在论域 U 中, 属性集 B_{i+1} 在 U 上的划分为 $U/B_{i+1} = \{X_1, X_2, \dots, X_n\}$, 属性集 B_i 在 U 的划分 $U/B_i = \{X_1, X_2, \dots, X_{k-1}, X_{l+1}, \dots, X_n, X_k \cup X_l\}$, 划分 U/B_{i+1} 中的两个等价子划分 X_k, X_l 合并得到新的等价子划分 $X_k \cup X_l$. $U/D = \{Y_1, Y_2, \dots, Y_m\}$ 是决策属性 D 在 U 上的划分, 则 $\text{AE}'(D_{B_{i+1}}) \leq \text{AE}'(D_{B_i})$.

证 明

$$\begin{aligned} \text{AE}'(D_B) &= \text{AE}'(D_{B_{i+1}}) + |X_i \cup Y_j| \times \sum_{j=1}^m \left[1 - \frac{|X_i \cap Y_j|}{2|X_i|} / \left(1 - \frac{|X_i \cap Y_j|}{2|X_i|} \right) \right] \\ &\quad - |X_k| \sum_{j=1}^m \left[1 - \frac{|X_k \cap Y_j|}{2|X_k|} / \left(1 - \frac{|X_k \cap Y_j|}{2|X_k|} \right) \right] \\ &\quad - |X_l| \sum_{j=1}^m \left[1 - \frac{|X_l \cap Y_j|}{2|X_l|} / \left(1 - \frac{|X_l \cap Y_j|}{2|X_l|} \right) \right], \\ \text{AE}'_\Delta &= \sum_{j=1}^m \left[|X_k| \left(\frac{|X_k \cap Y_j|}{2|X_k|} / \left(1 - \frac{|X_k \cap Y_j|}{2|X_k|} \right) \right) + |X_l| \left(\frac{|X_l \cap Y_j|}{2|X_l|} / \left(1 - \frac{|X_l \cap Y_j|}{2|X_l|} \right) \right) \right. \\ &\quad \left. - \frac{|(X_k \cup X_l) \cap Y_j| \cdot |X_k \cup X_l|}{2|X_k \cup X_l|} / \left(1 - \frac{|(X_k \cup X_l) \cap Y_j|}{2|X_k \cup X_l|} \right) \right]. \end{aligned}$$

设 $|X_k| = x, |X_l| = y, |X_k \cap Y_j| = ax, |X_l \cap Y_j| = by$, 显然 $x > 0, y > 0, 0 \leq a \leq 1, 0 \leq b \leq 1$,

$$\text{AE}'_\Delta = \sum_{j=1}^m \left(\frac{ax}{2-a} + \frac{by}{2-b} - \frac{(x+y)(ax+by)}{2x+2y-ax-by} \right) = \sum_{j=1}^m \frac{2xy(a-b)^2}{(2x+2y-ax-by)(2-a)(2-b)} \geq 0.$$

所以, $\text{AE}'(D_{B_{i+1}}) \leq \text{AE}'(D_{B_i})$, 证毕.

性质4 给定决策表 $DT = (U, C, D, V, f)$, 对于 $\forall B \subseteq C$, 令 U 在条件属性 B 上的划分为 $U/B = \{X_1, X_2, \dots, X_n\}$, 令 U 在决策属性 D 上的划分为 $U/D = \{Y_1, Y_2, \dots, Y_m\}$, 增强信息熵比文献[17]的相对模糊熵的信号强, 即 $\text{AE}'(D_B) \geq E(D_B)$.

证 明

$$\begin{aligned} \text{AE}'(D_B) - E(D_B) &= \sum_{j=1}^m \sum_{i=1}^n |X_i| \left[1 - \frac{|X_i \cap Y_j|}{2|X_i|} / \left(1 - \frac{|X_i \cap Y_j|}{2|X_i|} \right) \right] \\ &\quad - \sum_{j=1}^m \sum_{i=1}^n |X_i| \cdot \frac{|X_i \cap Y_j|}{|X_i|} \left(1 - \frac{|X_i \cap Y_j|}{|X_i|} \right) \\ &= \sum_{j=1}^m \sum_{i=1}^n |X_i| \left(\left[1 - \frac{|X_i \cap Y_j|}{2|X_i|} / \left(1 - \frac{|X_i \cap Y_j|}{2|X_i|} \right) \right] \right. \\ &\quad \left. - \frac{|X_i \cap Y_j|}{|X_i|} \left(1 - \frac{|X_i \cap Y_j|}{|X_i|} \right) \right). \end{aligned}$$

设 $\frac{|X_i \cap Y_j|}{|X_i|} = x$ 且 $0 \leq x \leq 1$, 则 $\text{AE}'(D_B) - E(D_B) = \frac{2-2x}{2-x} - x(1-x) = \frac{(1-x)[(x-1)^2+1]}{2-x} \geq 0$, 证毕.

2.2 近似边界精度信息熵

定义9 给定决策表 $DT = (U, C, D, V, f)$, 令 U 在条件属性 B 上的划分为 $U/B = \{X_1, X_2, \dots, X_n\}$ 对于 $\forall B \subseteq C$, U 在决策属性 D 上的划分为 $U/D = \{Y_1, Y_2, \dots, Y_m\}$, 决策属性 D 相对于条件属性 B 的近似边界精度近似边界精度相对信息熵为

$$\text{ABAE}(D|B) = \begin{cases} \sum_{j=1}^m \sum_{i=1}^n [1 - \beta'_B(Y_j)] \cdot |X_i| \cdot \left[1 - \frac{|X_i \cap Y_j|}{|X_i|} / \left(1 - \frac{|X_i \cap Y_j|}{|X_i|} \right) \right], & 0 \leq \frac{|X_i \cap Y_j|}{|X_i|} < \frac{1}{2}, \\ \sum_{j=1}^m \sum_{i=1}^n [1 - \beta'_B(Y_j)] \cdot |X_i| \cdot \left[1 - \left(1 - \frac{|X_i \cap Y_j|}{|X_i|} \right) / \frac{|X_i \cap Y_j|}{|X_i|} \right], & \frac{1}{2} \leq \frac{|X_i \cap Y_j|}{|X_i|} \leq 1. \end{cases} \quad (2.6)$$

近似边界精度增强信息熵为

$$\text{ABAE}'(D|B) = \sum_{j=1}^m \sum_{i=1}^n [1 - \beta'_B(Y_j)] \cdot |X_i| \cdot \left[1 - \frac{|X_i \cap Y_j|}{2|X_i|} / \left(1 - \frac{|X_i \cap Y_j|}{2|X_i|} \right) \right], \quad (2.7)$$

其中, $\beta'_B(Y_j)$ 为 Y_j 在 $\text{ind}(B)$ 下的修正边界精度, $1 \leq i \leq n$, $1 \leq j \leq m$.

在属性约简中, 有基于代数观的算法, 也有基于信息观的算法. 代数观的算法能有效度量知识的完备性, 但对知识粒度的大小度量效果不理想; 相反, 信息观的方法能有效度量知识粒度的大小, 对于知识的完备性却束手无策. 显然, 只从代数观或信息观某一方面研究属性约简, 均具有片面性. 由定义9可知, 近似边界精度增强信息熵有效结合了信息观与代数观的特点, 相对于传统的信息熵模型而言, 既考虑了完备性, 又考虑了知识粒度的大小, 从而能够全面、客观地度量粗糙集中的不确定性问题.

定理3 在给定的决策系统 $S = (U, C, D, V, f)$ 中, 对于 $\forall B \subseteq C$ 和 $a \in C - B$, 有

$$\text{ABAE}'(D|B) \geq \text{ABAE}'(D|(B \cup a)), \text{ABAE}(D|B) \geq \text{ABAE}(D|(B \cup a)).$$

证 明 因为近似边界精度增强信息熵的定义是由近似边界精度与增强信息熵组合而成, 所以下面从这两个方面证明.

(1) 由 U/B 定义及偏序关系 \preceq 可得, $U/(B \cup a) \preceq U/B$, 对于有 $\forall Y \subseteq U$, $|\overline{Y_B}| \geq |\overline{Y_{B \cup a}}|$ 且 $|\underline{Y_B}| \leq |\underline{Y_{B \cup a}}|$, 由定义5可知, $\beta'_B(Y) = \frac{|\underline{Y_B}|/|\overline{Y_B}|}{2 - |\underline{Y_B}|/|\overline{Y_B}|} \leq \frac{|\underline{Y_{B \cup a}}|/|\overline{Y_{B \cup a}}|}{2 - |\underline{Y_{B \cup a}}|/|\overline{Y_{B \cup a}}|} = \beta'_{B \cup a}(Y)$. 所以, $1 - \beta'_{B \cup a}(Y) \leq 1 - \beta'_B(Y)$

(2) 设 U 在决策 D 上的划分 $U/D = \{Y_1, Y_2, \dots, Y_m\}$, 由定理2可得

$$\sum_{i=1}^n \sum_{X \in U/(B \cup a)} \left[1 - \frac{|X \cap Y_i|}{2|X|} \left(1 - \frac{|X \cap Y_i|}{2|X|} \right) \right] \leq \sum_{i=1}^n \sum_{X \in U/B} \left[1 - \frac{|X \cap Y_i|}{2|X|} \left(1 - \frac{|X \cap Y_i|}{2|X|} \right) \right],$$

即 $\text{AE}'(D|B) \geq \text{AE}'(D|(B \cup a))$.

同理, 由定理1可得, $\text{AE}(D|B) \geq \text{AE}(D|(B \cup a))$

综合(1)(2)可得, $\text{ABAE}'(D|B) \geq \text{ABAE}'(D|(B \cup a)), \text{ABAE}(D|B) \geq \text{ABAE}(D|(B \cup a))$.

3 属性约简算法(ABAE)

由定理3可知, 近似边界精度增强信息熵(或相对信息熵)是一个具有非严格单调性规律的函数, 因此可以借助近似边界精度增强信息熵(或相对信息熵)的特点, 来度量属性重要性, 并以此为依据, 计算约简集.

定义10 给定的决策表 $DT = (U, C \cup D, V, f)$, 对于 $\forall B \subseteq C$, 若存在 $ABAE'(D|B) = ABAE'(D|C)$, 且对于 $\forall a \in B$, 有 $ABAE'(D|B - \{a\}) \neq ABAE'(D|C)$, 则称 B 为 C 相对于决策属性 D 的一个近似边界精度增强信息熵的约简.

定义11 给定的决策表 $DT = (U, C \cup D, V, f)$, 对于 $\forall B \subseteq C$, $\forall b \in C - B$, 则属性 b 在 S 中, 相对于属性集 B 在决策属性 D 下的相对信息熵重要性为

$$\text{Sig}(b, B, D) = ABAE(D|B) - ABAE(D|B \cup \{b\}), \quad (3.1)$$

增强信息熵重要性为

$$\text{Sig}'(b, B, D) = ABAE'(D|B) - ABAE'(D|B \cup \{b\}). \quad (3.2)$$

因为在计算信息熵过程中, 存在大量的计算 $|X_i \cap Y_j|$, X_i 为划分 U/B 的子划分, Y_i 为 U/D 的子划分. 求这两个子划分的交集伪代码为如下.

```
char[ ] intersect(string X, string Y)
{
    Char[ ] Z;
    int i = 0, j = 0, k = 0;
    while( i < X.length && j < Y.length)
    {
        if ( X[i] == y[j])
            { Z[k ++] = X[i]; i ++; j ++; }
        else if ( X[i] > Y[j] )
            j++;
        else
            i++;
    }
    return Z;
}
```

由 `intersect(string X, string Y)` 思想可知, 本文使两个集合的交集时间复杂度由 $|X||Y|$ 降为 $|X| + |Y|$, 其中 $|X|$ 为集合 X 的基数, $|Y|$ 为集合 Y 的基数.

在求相对信息熵的过程中, 存在着大量的求 $U/(B \cup b)$ 步骤, 根据性质1可得, 利用 U/B 的结果, 再对 U/B 每个子划分按属性 b 的取值不同求加细, 这样可大大节省系统开销. 具体见算法1、算法2.

算法1 近似边界精度增强信息熵及相对信息熵的计算

输入: 决策表 $DT = (U, C, D, V, f)$, 划分
 $U/B = \{X_1, X_2, \dots, X_n\}$, $U/D = \{Y_1, Y_2, \dots, Y_m\}$, 且 $B \subseteq C$
输出: 近似边界精度增强信息熵 $ABAE'(D|B)$, 相对信息熵为 $ABAE(D|B)$

```
int ABAE' = ABAE = 0;
int type;
for ( i from 1 to m ) //m 为 U/D 子划分的数目
{
    int t1=t2=t3=t4=count=0;
    input type;
    for ( j from 1 to n ) // n 为 U/B 子划分的数目
    {
        count = count + |X_j|;
        temp = intersect (X_j, Y_i).length; //统计下近似元素的个数
        if (temp == |X_j|) then
            t1 = t1 + |X_j|;
        else if (temp == 0)
            t2 = t2 + |X_j|;
    }
    else { switch( type ) {
```

```

case 1: //计算增强信息熵
     $t3 = t3 + |X_j| \cdot \frac{2\left(1 - \frac{\text{temp}}{|X_j|}\right)}{2 - \frac{\text{temp}}{|X_j|}}$ ; break;
case 2: //计算相对信息熵
    if  $\left(0 \leq \frac{\text{temp}}{|X_j|} < \frac{1}{2}\right)$   $t4 = t4 + |X_j| \cdot \left[1 - \frac{\text{temp}}{|X_j|} / \left(1 - \frac{\text{temp}}{|X_j|}\right)\right]$ ;
    else  $t4 = t4 + |X_j| \cdot \left[1 - \left(1 - \frac{\text{temp}}{|X_j|}\right) / \frac{\text{temp}}{|X_j|}\right]$ ; }
break; }

t2 = count - t2; //统计上近似元素个数
 $\text{ABAE} = \text{ABAE} + \left(1 - \frac{\frac{t_1}{t_2}}{2 - \frac{t_1}{t_2}}\right) \cdot t4; \quad \text{ABAE}' = \text{ABAE}' + \left(1 - \frac{\frac{t_1}{t_2}}{2 - \frac{t_1}{t_2}}\right) \cdot t3;$ 
}

```

算法 2 基于近似边界精度增强信息熵约简及相对信息熵约简

输入: DT = (U, C, D, V, f) ;
输出: R.

```

step 1  $R = \emptyset, C' = C$ ;
step 2 计算  $U/C, U/D$ ;
step 3 计算条件属性 C 相对决策属性 D 的近似边界增强信息熵 ABAE'(D|C), 相对信息
熵 ABAE(D|C);
step 4 while ( 1 )
{
    计算  $\text{Sig}'(b_k, R, D) = \max_{b_i \in C'} \text{Sig}'(b_i, R, D)$  或  $\text{Sig}(b_k, R, D) = \max_{b_i \in C'} \text{Sig}(b_i, R, D)$ ,  $1 \leq i \leq |C'|$ .
    若有多个属性的重要性都达到最大值, 则从中任选一个  $b_k$ ;
     $R = R \cup b_k; \quad C' = C' - b_k$ ;
    if (ABAE'(R|C) = ABAE'(D|C))
        输出增强信息熵的约简 R;
    if (ABAE(R|C) = ABAE(D|C))
        输出相对信息熵的约简 R;
}

```

算法主要的时间开销在于求信息熵, 每加入一个属性时, 要重新计算近似边界精度及信息熵. 由于巧妙利用 U/B 求 $U/(B \cup b)$, 使得计算 ABAE' 的时间降为 $O(|U|)$; 在求属性约简时, 每次选择一个最重要的属性加入约简集, 算法最坏的时间复杂度为 $O(|C|^2|U|)$.

4 实验分析

为进一步验证本文近似边界精度增强信息熵、相对信息熵约简算法与其他同类算法的性能, 利用 Windows7 系统, 开发工具为 VS2012, CPU 2.3 G, 内存 4 GB 环境下编程. 本文选用 UCI 机器学习数据库中的 6 个数据集测试: ① Tic-Tac-Toe Endgame (Tic) 数据集, 958 个样例数, 10 个属性, 2 个模式类; ② Zoo 数据集, 101 个样例, 17 个属性, 7 个模式类; ③ Ecoli 数据集, 336 个样例, 8 个属性, 8 个模式类; ④ Lymphography 数据集 (Lymph), 148 个样例, 19 个属性, 4 个模式类; ⑤ Chess King-Rook vs. King-Pawn (Chess) 数据集, 3196 个样例, 37 个属性, 2 个模式类; ⑥ Mushroom (Mush) 数据集, 8124 个样例, 23 个属性, 2 个模式类.

本文与以下具有代表性的几个算法进行比较: ① 基于条件熵的算法 (Conditional Entropy, CE)^[15]; ② 相对模糊熵 (Fuzzy Entropy, FE)^[17]; ③ 基于近似决策熵的算法 (Approximate Decision Entropy Attribute Reduction, ADEAR)^[19]; ④ 组合条件熵 (Combine Conditional Entropy, CCE)^[20].

实验从约简结果、分类精度、运行时间上比较. 表 1 给出了不同算法在上述 6 个数据集上

的约简情况, 其中RN为最小属性约简的基数。通过表1可知, CE、FE、CCE3种算法计算属性重要性时, 只是围绕信息观下的信息熵研究, 考虑信息的粒度, 影响约简效果。但ADEAR将近似精度融入信息熵、ABAE将近似边界精度融入相对信息熵、ABAE'将近似边界精度融入增强信息熵计算, 有效结合代数观、信息观的方法, 全面、客观地度量属性重要性, 在表1的6个数据集中, 均有较好的约简效果。

表1 约简结果比较(剩余属性数)

Tab. 1 Comparison of reduction result(residue number)

DataSet	RN	CE	FE	CCE	ADEAR	ABAE	ABAE'
Tic	8	8	8	8	8	8	8
Zoo	5	9	5	5	5	5	5
Ecoli	5	6	6	6	5	5	5
Lymph	6	8	6	7	6	6	6
Chess	29	30	30	29	29	29	29
Mush	4	5	5	5	5	5	4

运行时间上, 本文算法ABAE、ABAE'与算法CE、FE、CCE的比较, 如表2所示。

表2 运行时间比较

Tab. 2 Comparison of run time

DataSet	CE/s	FE/s	CCE/s	ADEAR/s	ABAE/s	ABAE'/s
Tic	8.73	6.76	0.36	0.101	0.103	0.089
Lymph	3.124	1.36	0.126	0.053	0.054	0.048
Mush	300.22	166.92	24.88	1.80	1.88	1.67

由表2可知, 针对相同的数据约简, ABAE算法的运行时间远小于CE、FE, 由于ADEAR与ABAE方法在计算U/B时利用高效方法, 节约大量的系统时间开销, 且对属性重要性计算的计算量相当, 所以时间性能上没有多大差异。ABAE'比ABAE方法少了判断步骤, 且通过利用信息熵扩大的特点, 快速找到最重要的属性, 加快约简的速度, 所以时间较优越。

为进一步验证基于信息熵算法、基于正区域算法^[21](简称FBS)及本文算法的分类准确性, 将各算法得到的约简结果作为SVM分类器的输入。为提高SVM分类器对样本的分类精度, 以20%数据作测试集, 80%为训练集, 核函数采用“径向基核”, 调整拉格朗日乘子上界及径向基核参数, 利用十折交叉验证的方法来评估所有的约简分类精度, 如图1所示。

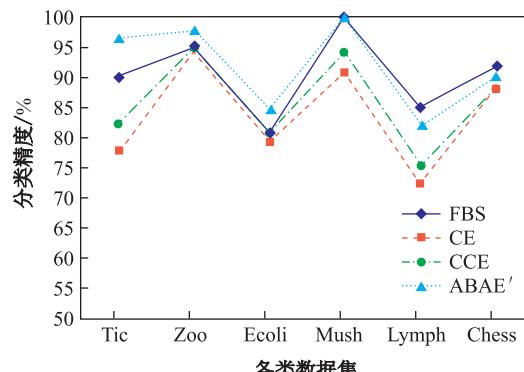


图1 分类精度比较

Fig. 1 Comparison of classification accuracy

通过图1的分类效果可知, ABAE'在数据集Tic、Zoo、Ecoli、Mush上均能获得较高分类

精度的约简结果; 基于单一信息熵的约简算法 CE, 约简结果的分类准确率不高, 由于只单纯考虑信息熵的方法计算属性重要性; FBS 算法的约简分类准确率, 性能较稳定, 但多数情况下较 ABAE' 的分类精度低, 由于在度量属性重要性时, 只考虑决策表中正域样本的区分效果, 未考虑负区域样本间的区分能力。特别是在利用正区域方法计算属性重要性时, 常会出现多个条件属性的属性重要性值相同的情况, 传统的算法是随机选择一个。由于每次计算属性重要性都是计算相对重要性, 所以随机选择的属性会影响下一次属性的选择。若采用本文的两种放大信息熵的方法, 并结合代数观下的近似边界精度计算属性重要性, 可以有效度量样本间的细微差异, 从而使得在后续的数据分类工作中, 展现出本文算法在分类效果上的优势。

为了研究离群点对算法的约简影响, 利用欧式距离计算各样本对象间的空间距离, 再利用拉依达准则计算离群点个数。上述 6 个数据集计算得到的离群点个数: Tic 为 0; Zoo 为 2; Ecoli 为 4; Lymph 为 5; Chess 为 188; Mush 为 156。设离群点数占样本总数的比例为离群率, 则各数据集的离群率: Tic 为 0%; Zoo 为 19.8%; Ecoli 为 11.9%; Lymph 为 33.8%; Chess 为 58.8%; Mush 为 19.2%。上述 6 个数据集中, ABAE' 在数据集 Chess 及 Lymph 上分类精度低于 FBS 算法, 因为数据集 Chess、Lymph 有较高的离群率, 由于离群率的存在影响了 ABAE' 的分类性能。相比之下, FBS 算法受离群点率的影响较小, 因此其分类精度高于 ABAE'。针对离群率问题, 要提高 ABAE' 分类性能, 必须考虑在属性约简之前, 尽量采用离群点检测方法, 找出离群点, 以减少离群点对属性约简的影响。

5 结 论

属性约简是粗糙集理论研究的热点之一, 广大研究者们多数只从代数观或者只从信息观设计相关属性约简算法。本文融合代数观和信息观的特点, 设计一种计算属性重要性的方法, 使分类精度更高。另外, 由于在计算 $U/(B \cup b)$ 时, 充分利用 U/B 的信息, 使求等价类的时间大大减少, 节省系统开销。

下一步工作, 考虑如何在保持高精度分类的情况下, 减少算法的计算量, 更进一步提高算法的时间性能, 以及考虑如何将计算与现代的多核处理器结合, 降低算法的时间复杂度。

[参 考 文 献]

- [1] PAWLAK Z, SKOWRON A. Rough sets: Some extensions [J]. Information Sciences, 2007, 117(1): 28-40.
- [2] PAWLAK Z. Rough sets [J]. Int J of Computer and Information Sciences, 1982, 11(5): 341-356.
- [3] 王熙熙, 王婷婷, 翟俊海. 基于样例选取的属性约简算法 [J]. 计算机研究与发展, 2012, 49(11): 2305-2310.
- [4] KIM K, CHU Y Y, WATADA J Z, et al. A DNA-based algorithm for minimizing decision rules: A rough sets approach [J]. IEEE Trans Nanobiosci, 2011, 10(3): 139-151.
- [5] HUANG C C, TSENG T L, JIANG F H, et al. Rough set theory: A novel approach for extraction of robust decision rules based on incremental attributes [J]. Annals of Operations Research, 2014, 216(1): 163-189.
- [6] 荆涛, 王家林, 石旭东, 等. 基于粗糙集约简的飞机发电机故障诊断决策研究 [J]. 计算机应用研究, 2017, 34(4): 1101-1104.
- [7] 朱庆, 苗双喜, 丁雨淋, 等. 知识引导的滑坡监测数据粗差定位与剔除方法 [J]. 武汉大学学报(信息科学版), 2017, 42(4): 496-502.
- [8] 梁宝华, 汪世义. 行式存储的快速属性约简算法 [J]. 模式识别与人工智能, 2015, 8(9): 795-801.
- [9] 赵洁, 梁俊杰, 董振宁, 等. 位运算和核属性快速识别下的粗糙集属性约简算法研究 [J]. 小型微型计算机系统, 2015, 36(2): 316-321.
- [10] 史博文, 李国和, 吴卫江, 等. 基于强化正域的属性约简方法 [J]. 计算机应用研究, 2017, 34(1): 107-109.
- [11] 周建华, 徐章艳, 章晨光. 改进的差别矩阵的快速属性约简算法 [J]. 小型微型计算机系统, 2014, 35(4): 831-834.
- [12] 蒋瑜. 基于差别信息树的 rough set 属性约简算法 [J]. 控制与决策, 2015, 30(8): 1531-1536.
- [13] 龙浩, 徐超. 基于改进差别矩阵的属性约简增量式更新算法 [J]. 计算机科学, 2015, 42(6): 251-255.
- [14] SHANNON C E. The mathematical theory of communication [J]. Bell System Technical J, 1948, 27(3/4): 373-423.

(下转第 156 页)

- [11] SUN B, REALE A, SEIDEL D J, et al. Comparing radiosonde and COSMIC atmospheric profile data to quantify differences among radiosonde types and the effects of imperfect collocation on comparison statistics[J]. *Journal of Geophysical Research Atmospheres*, 2010, 115(D23): 6696-6705.
- [12] PU Z, ZHANG L. Validation of Atmospheric Infrared Sounder temperature and moisture profiles over tropical oceans and their impact on numerical simulations of tropical cyclones[J]. *Journal of Geophysical Research*, 2010, 115(D24): 9-12.
- [13] CHAHINE M T, PAGANO T S, AUMANN H H, et al. AIRS: Improving weather forecasting and providing new data on greenhouse gases[J]. *Bulletin American Meteorological Society*, 2006, 87(7): 911-926.
- [14] 刘辉, 董超华, 张文建, 等. AIRS 晴空大气温度廓线反演试验[J]. 气象学报, 2008, 66(4): 513-519.
- [15] SUSSKIND J, BARNET C D, BLAISDELL J M. Accuracy of geophysical parameters derived from Atmospheric Infrared Sounder/Advanced Microwave Sounding Unit as a function of fractional cloud cover[J]. *Journal of Geophysical Research Atmospheres*, 2006, 111(D9): 9-17.

(责任编辑: 李万会)

(上接第 108 页)

- [15] 王国胤, 于洪, 杨大春. 基于条件信息熵的决策表约简 [J]. *计算机学报*, 2002, 25(7): 759-776.
- [16] 杨明. 决策表中基于条件信息熵的近似约简 [J]. *电子学报*, 2007, 35(11): 2156-2160.
- [17] 张清华, 肖雨. 新的信息熵属性约简 [J]. *计算机科学与探索*, 2013, 7(4): 359-367.
- [18] 潘瑞林, 李园沁, 张洪亮, 等. 基于 α 信息熵的模糊粗糙集属性约简方法 [J]. *控制与决策*, 2017, 32(2): 340-348.
- [19] 江峰, 王莎莎, 杜军威, 等. 基于近似决策熵的属性约简 [J]. *控制与决策*, 2015, 30(1): 65-70.
- [20] QIAN Y H, LIANG J Y, PEDRYCZ W, et al. Positive approximation: An accelerator for attribute reduction in rough set theory [J]. *Artificial Intelligence*, 2010, 174(9): 597-618.
- [21] 蒋瑜, 刘胤田, 李超. 基于 Bucket Sort 的快速属性约简算法 [J]. *控制与决策*, 2011, 26(2): 207-212.

(责任编辑: 李艺)

(上接第 135 页)

- [19] YANG X L, YANG J X, CHEN G, et al. Magneto-impedance effect in field and stress-annealed Fe-based nanocrystalline alloys [J]. *Journal of Magnetism and Magnetic Materials*, 1997, 175(3): 285-289.
- [20] RODIONOVA V, KUDINOV N, ZHUKOV A, et al. Interaction of bistable glass-coated microwires in different positional relationship[J]. *Physica B: Condensed Matter*, 2012, 407(9): 1438-1441.
- [21] 何济洲, 缪贵玲. 两个磁偶极子间的相互作用[J]. *南昌大学学报*, 1996, 18(3): 96-98.
- [22] 易建宏, 李丽娅, 彭元东. 微波场作用下非晶合金 $Fe_{73.5}Cu_1Nb_3Si_{13.5}B_9$ 的纳米晶化[J]. *材料研究学报*, 2007, 21(6): 632-636.
- [23] 王宗篪, 杨燮龙, 宫峰飞, 等. 纵向驱动巨磁阻抗效应的相位特性[J]. *华东师范大学学报(自然科学版)*, 2000(2): 49-53.
- [24] AMALOUA F, GIJS M A M. Giant magnetoimpedance in trilayer structures of patterned magnetic amorphous ribbons [J]. *Applied Physics Letters*, 2002, 81(9): 1654-1656.
- [25] 林宏, 袁望治, 赵振杰. $FeCuNbSiB$ 多层膜的磁特性和巨磁阻抗效应[J]. *功能材料与器件学报*, 2007, 13(6): 666-670.

(责任编辑: 李艺)