

## 基于邻域粗糙互信息熵的非单调性属性约简

姚 晟, 徐 风<sup>†</sup>, 吴照玉, 陈 菊, 汪 杰, 王 维

(1. 安徽大学 计算智能与信号处理教育部重点实验室, 合肥 230601;  
2. 安徽大学 计算机科学与技术学院, 合肥 230601)

**摘 要:** 属性约简是粗糙集理论一项重要的应用, 目前已广泛运用于机器学习和数据挖掘等领域, 邻域粗糙集是粗糙集理论中处理连续型数据的一种重要方法. 针对目前邻域粗糙集模型中属性约简存在的缺陷, 构造一种基于邻域粗糙集的邻域粗糙熵模型, 并基于此给出邻域粗糙联合熵、邻域粗糙条件熵和邻域粗糙互信息熵等概念. 邻域粗糙互信息熵是评估属性集相关性的一种重要的方法, 具有非单调性变化的特性, 对此, 提出一种基于邻域粗糙互信息熵的非单调性属性约简算法. 实验分析表明, 所提出算法不仅比目前已有的单调性属性约简算法具有更优越的属性约简结果, 而且具有更高的约简效率.

**关键词:** 邻域粗糙集; 邻域粗糙熵; 邻域粗糙条件熵; 邻域粗糙互信息熵; 非单调性; 属性约简

中图分类号: TP18

文献标志码: A

## Non-monotonic attribute reduction based on neighborhood rough mutual information entropy

YAO Sheng, XU Feng<sup>†</sup>, WU Zhao-yu, CHEN Ju, WANG Jie, WANG Wei

(1. Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, Anhui University, Hefei 230601, China; 2. College of Computer Science and Technology, Anhui University, Hefei 230601, China)

**Abstract:** Attribute reduction is an important application in rough set theory, and it has been widely used in such areas as machine learning and data mining so far. Neighborhood rough set is a vital method for processing continuous data in rough set theory. For the existed detects of attribute reduction in the current neighborhood rough set model, the model of neighborhood rough entropy based on a neighborhood rough set is defined, meanwhile, the concepts of neighborhood rough combination entropy, neighborhood rough conditional entropy and neighborhood rough mutual information entropy are given, where the neighborhood rough mutual information entropy is an important method for evaluating the correlation of attribute sets, and at the same time, the neighborhood rough mutual information entropy is also proved to have a property of non-monotonic changing, therefore a non-monotonic attribute reduction algorithm based on neighborhood rough mutual information entropy is proposed. The experimental analysis show that the proposed algorithm has not only better results but also higher reduction efficiency than existing monotonic algorithms in attribute reduction.

**Keywords:** neighborhood rough set; neighborhood rough entropy; neighborhood rough conditional entropy; neighborhood rough mutual information entropy; non-monotonic; attribute reduction

### 0 引 言

属性约简<sup>[1]</sup>又称为特征选择, 是粗糙集理论<sup>[1]</sup>在机器学习和数据挖掘等领域中一项重要的数据预处理方法. 在当今大数据环境下, 海量高维是数据的一个典型的特征, 并且部分数据内部包含着大量的噪声数据, 对数据的知识发现产生很大的影响; 因此, 对其

进行属性约简至关重要. 目前属性约简已成为大数据中一项重要的数据处理技术<sup>[2]</sup>.

传统的粗糙集理论是建立在等价关系的基础上, 适用于离散型的数据集, 学者们提出了众多的关于离散型数据的属性约简算法<sup>[1-2]</sup>. 由于现实中很多的数据包含连续型的属性, 为了解决这一问题, Hu 等<sup>[3-4]</sup>

收稿日期: 2017-08-11; 修回日期: 2017-11-03.

基金项目: 国家自然科学基金项目(61602004, 61300057); 安徽省自然科学基金项目(1508085MF127); 安徽省高等学校自然科学研究重点项目(KJ2016A041, KJ2017A011); 安徽大学信息保障技术协同创新中心公开招标项目(ADXXBZ2014-5, ADXXBZ2014-6); 安徽大学博士科研启动基金项目(J10113190072).

责任编委: 孙秋野.

作者简介: 姚晟(1979—), 女, 副教授, 博士, 从事智能计算、大数据等研究; 徐风(1993—), 男, 硕士生, 从事智能计算、机器学习的研究.

<sup>†</sup>通讯作者. E-mail: xuf\_2015@163.com.

在连续型数据集上建立了邻域关系,并提出了邻域粗糙集模型,该模型可以对连续型数据直接进行处理;同时Hu等<sup>[4-6]</sup>提出了连续型数据的正区域属性约简方法.在Hu等研究的基础上,学者们提出了大量的基于邻域粗糙集的属性约简算法,例如Liu等<sup>[7]</sup>提出了邻域粗糙集的快速属性约简,Wang等<sup>[8-9]</sup>提出了基于模糊邻域粗糙集模型的属性约简和基于邻域区分度的属性约简,Li等<sup>[10]</sup>提出了基于邻域决策理论粗糙集的属性约简,姚晟等<sup>[11-12]</sup>提出了邻域量化容差关系的条件熵属性约简和基于改进邻域粒化的模糊熵属性约简,何松华等<sup>[13]</sup>提出了邻域组合测度的属性约简方法,段洁等<sup>[14]</sup>提出了基于邻域粗糙集的多标记特征选择算法.这些算法的提出使得邻域粗糙集成为连续型数据属性约简的一种强有力的工具.

目前,基于邻域粗糙集模型的属性约简算法有一个共同点,都是通过邻域粗糙集的邻域构造评估函数,常见的评估函数有正区域<sup>[4-5]</sup>、知识粒度<sup>[13]</sup>、信息熵<sup>[11-12]</sup>等,这些评估函数都满足单调性原则,基于评估函数这种单调性去启发式地选择属性,直到属性子集的评估函数度量值与属性全集的值保持一致,便获得了最终的属性约简结果<sup>[4-14]</sup>.Li等<sup>[15]</sup>指出,这种单调性的属性约简存在一定的缺陷,例如当原始数据集的分类性能较差时,对应的评估函数度量值比较低,这种单调性的属性约简不能得到更好的属性约简结果.为了改善这种缺陷,Li等提出了一种非单调性的属性约简算法,并通过实验表明了其有效性.然而,Li等提出的属性约简算法只适用于离散型数据.

为了对连续型的数据集达到更好的属性约简效果,本文首先在邻域粗糙集模型中提出邻域粗糙熵的概念;然后,基于邻域粗糙熵定义邻域粗糙联合熵、邻域粗糙条件熵和邻域粗糙互信息熵,邻域粗糙互信息熵是评估属性集相关性的一种重要方法,它满足属性集度量的非单调性;最后,利用邻域粗糙互信息熵的非单调性结合Li等的观点,提出一种基于邻域粗糙互信息熵的非单调性属性约简算法.实验结果验证了所提出方法具有更好的属性约简效果.

## 1 基本理论

在粗糙集理论<sup>[1]</sup>中,所研究的数据被描述成信息系统的形式,一个信息系统可表示为 $IS = (U, AT, V)$ .其中: $U$ 是一个非空有限集合,被称为论域; $AT$ 为信息系统的全体属性集,又称为特征集; $V$ 为全体属性集的值域, $V_a$ 表示属性 $a$ 所有取值的集合.对象 $x \in U$ 在属性 $a$ 下的值可表示为 $a(x)$ ,即 $a(x) \in V_a$ .

另外,当全体属性集 $AT = C \cup D$ ,且 $C \cap D = \emptyset$ 时,此信息系统又称为决策信息系统(DIS),其中 $C, D$ 分别表示为信息系统的条件属性集和决策属性.当该信息系统的条件属性的取值均为连续数值时,此信息系统又称为邻域决策信息系统(NDIS).

**定义1** 对于信息系统 $IS = (U, AT, V)$ ,设 $\forall B \subseteq AT$ ,定义

$$R_B = \{(x, y) \mid a(x) = a(y), \forall a \in B\}. \quad (1)$$

$R_B$ 是 $U$ 上的一个等价关系,论域 $U$ 在 $R_B$ 上可以诱导出一个划分 $U/R_B$ ,同时对象 $x \in U$ 在 $R_B$ 上所产生的等价类定义为 $[x]_B = \{y \mid (x, y) \in R_B\}$ .

在邻域信息系统中,由于属性的取值为连续型,邻域信息系统中的等价关系无法构建,Hu等<sup>[3-4]</sup>运用邻域关系代替等价关系,并提出了基于邻域关系的粗糙集模型,简称邻域粗糙集模型.

**定义2** 设邻域信息系统表示为 $NIS = (U, AT, V)$ , $\Delta B \subseteq AT$ 在论域 $U$ 上诱导的邻域关系定义为

$$N_B = \{(x, y) \mid \Delta_B(x, y) \leq \delta, \forall x, y \in U\}. \quad (2)$$

其中: $\delta$ 为一个非负常数,被称为邻域半径; $\Delta_B(x, y)$ 为对象 $x, y$ 之间的距离,通常的距离计算采用如下度量公式:

$$\Delta_B(x, y) = \left( \sum_{\forall a \in B} |a(x) - a(y)|^p \right)^{1/p}. \quad (3)$$

**定义3** 对于邻域信息系统 $NIS = (U, AT, V)$ ,邻域半径为 $\delta$ , $B \subseteq AT$ 诱导的邻域关系 $N_B$ 可以将论域 $U$ 粒化成一组邻域粒 $U/N_B$ ,表示为

$$U/N_B = \{n_B^\delta(x_i) \mid x_i \in U, 1 \leq i \leq |U|\}. \quad (4)$$

这里的 $n_B^\delta(x_i)$ 表示对象 $x \in U$ 在邻域关系 $N_B$ 下的邻域粒,定义为

$$n_B^\delta(x) = \{y \in U \mid (x, y) \in N_B\}. \quad (5)$$

对于 $P, Q \subseteq AT$ 诱导的两个邻域关系分别为 $N_P$ 和 $N_Q$ ,对论域 $U$ 的粒化分别表示为 $U/N_P$ 和 $U/N_Q$ ,如果对于 $\forall x \in U$ ,都满足 $n_P^\delta(x) \subseteq n_Q^\delta(x)$ ,则称 $U/N_P$ 比 $U/N_Q$ 邻域粒化得更加精细,称 $U/N_Q$ 比 $U/N_P$ 邻域粒化得更加粗糙,表示为 $U/N_P \preceq U/N_Q$ ,即邻域关系 $N_P$ 和 $N_Q$ 满足偏序关系.特别地,对于 $\forall x \in U$ ,都满足 $n_P^\delta(x) \subset n_Q^\delta(x)$ ,则称 $U/N_P$ 比 $U/N_Q$ 严格精细, $U/N_Q$ 比 $U/N_P$ 严格粗糙,表示为 $U/N_P \prec U/N_Q$ .

**定义4** 对于邻域信息系统 $NIS = (U, AT, V)$ ,邻域半径为 $\delta$ , $\forall B \subseteq AT$ 所诱导的邻域关系为 $N_B$ ,对于论域上的对象集 $X \subseteq U$ ,关于邻域关系 $N_B$ 的邻域下近似和邻域上近似分别定义为

$$\underline{N}_B(X) = \{x \in U \mid n_B^\delta(x) \subseteq X\}, \quad (6)$$

$$\overline{N}_B(X) = \{x \in U \mid n_B^\delta(x) \cap X \neq \emptyset\}, \quad (7)$$

同时称

$$\text{POS}_{N_B}(X) = \underline{N}_B(X), \quad (8)$$

$$\text{BUN}_{N_B}(X) = \overline{N}_B(X) - \underline{N}_B(X), \quad (9)$$

$$\text{NEG}_{N_B}(X) = U - \overline{N}_B(X) \quad (10)$$

为  $X$  关于邻域关系  $N_B$  的邻域正区域、邻域边界域和邻域负区域。

## 2 邻域粗糙熵模型

近年来, Liang等<sup>[16]</sup>在信息系统中定义了信息熵理论, 并用于信息系统的不确定度量. 随后学者们提出了多种的变形和推广, 如联合熵<sup>[17]</sup>、粗糙熵<sup>[18]</sup>和粒度度量<sup>[19]</sup>等. 本节主要在前人研究的基础上, 将粗糙熵推广到邻域粗糙集模型中.

**定义5** 对于邻域信息系统  $\text{NIS} = (U, \text{AT}, V)$  且  $|U| = n$ , 设  $\forall B \subseteq \text{AT}$ , 邻域半径为  $\delta$ , 由  $B$  在论域  $U$  上诱导出的邻域粒化为  $U/N_B = \{n_B^\delta(x_1), n_B^\delta(x_2), \dots, n_B^\delta(x_n)\}$ , 则定义邻域信息系统基于  $B$  的邻域粗糙熵为

$$E_\delta(B) = -\frac{1}{n} \sum_{i=1}^n \log \frac{1}{|n_B^\delta(x_i)|}. \quad (11)$$

下文中, 在不引起混淆的情况下, 适当省略  $\delta$  的标记.

**性质1** 对于邻域信息系统  $\text{NIS} = (U, \text{AT}, V)$ , 设  $P, Q \subseteq \text{AT}$ , 若  $P$  和  $Q$  在论域上的邻域粒化满足  $U/N_P = U/N_Q$ , 则有  $E(P) = E(Q)$ .

**证明**  $P, Q$  在论域上的邻域粒化满足  $U/N_P = U/N_Q$ , 即  $\forall x \in U$  有  $n_P(x) = n_Q(x)$ , 因此根据邻域粗糙熵的定义有  $E(P) = E(Q)$ .  $\square$

**性质2** 对于邻域信息系统  $\text{NIS} = (U, \text{AT}, V)$  且  $|U| = n$ , 设  $P, Q \subseteq \text{AT}$ , 若  $P$  和  $Q$  在论域上的邻域粒化满足  $U/N_P \preceq U/N_Q$ , 则有  $E(P) \leq E(Q)$ .

**证明**  $P, Q$  在论域上的邻域粒化满足  $U/N_P \preceq U/N_Q$ , 即  $\forall x \in U$  有  $n_P(x) \subseteq n_Q(x)$ , 则有  $\forall x \in U$  满足  $\log |n_P(x)| \leq \log |n_Q(x)|$ , 所以  $\sum_{i=1}^n \log |n_P(x_i)| \leq \sum_{i=1}^n \log |n_Q(x_i)|$ , 即满足关系  $E(P) \leq E(Q)$ .  $\square$

性质2表明, 随着邻域关系对论域的粒化逐渐变得精细, 其邻域粗糙熵是单调不增的, 这表明邻域粗糙熵可以作为邻域信息系统粒化程度的一种度量指标.

在邻域粗糙熵中, 邻域半径  $\delta$  影响着对象邻域的大小, 因此接下来分析邻域半径与邻域粗糙熵之间的关系.

**性质3** 对于邻域信息系统  $\text{NIS} = (U, \text{AT}, V)$

且  $|U| = n, B \subseteq \text{AT}$ , 设邻域半径  $\delta_1$  和  $\delta_2$  满足  $\delta_1 \leq \delta_2$ , 则邻域粗糙熵有  $E_{\delta_1}(B) \leq E_{\delta_2}(B)$ .

**证明** 由于  $\delta_1 \leq \delta_2$ , 根据定义3关于对象邻域粒的定义, 对于  $\forall x \in U$  满足  $n_B^{\delta_1}(x) \subseteq n_B^{\delta_2}(x)$ , 可以得到  $\sum_{i=1}^n \log |n_B^{\delta_1}(x_i)| \leq \sum_{i=1}^n \log |n_B^{\delta_2}(x_i)|$ , 即满足关系  $E_{\delta_1}(B) \leq E_{\delta_2}(B)$ .  $\square$

定义5中的邻域粗糙熵是针对一个邻域关系导出的, 接下来针对多个邻域关系给出信息系统的邻域粗糙联合熵.

**定义6** 对于邻域信息系统  $\text{NIS} = (U, \text{AT}, V)$  且  $|U| = n$ , 设  $P, Q \subseteq \text{AT}$ , 由  $P$  和  $Q$  在论域  $U$  上诱导出的邻域粒化分别为

$$U/N_P = \{n_P(x_1), n_P(x_2), \dots, n_P(x_n)\},$$

$$U/N_Q = \{n_Q(x_1), n_Q(x_2), \dots, n_Q(x_n)\},$$

则邻域信息系统基于  $P$  和  $Q$  的邻域粗糙联合熵定义为

$$E(P \cup Q) = -\frac{1}{n} \sum_{i=1}^n \log \frac{1}{|n_P(x_i) \cap n_Q(x_i)|}. \quad (12)$$

类似地, 邻域联合粗糙熵满足如下性质.

**性质4** 对于邻域信息系统为  $\text{NIS} = (U, \text{AT}, V)$  且  $|U| = n$ , 设  $P, Q \subseteq \text{AT}$ , 邻域粗糙联合熵满足  $E(P \cup Q) \leq E(P), E(P \cup Q) \leq E(Q)$ .

**证明** 因为  $|n_P(x_i) \cap n_Q(x_i)| \leq |n_P(x_i)|$ , 有  $\sum_{i=1}^n \log |n_P(x_i) \cap n_Q(x_i)| \leq \sum_{i=1}^n \log |n_P(x_i)|$ , 所以性质4成立.  $\square$

**性质5** 对于邻域信息系统  $\text{NIS} = (U, \text{AT}, V)$  且  $|U| = n$ , 设  $P, Q \subseteq \text{AT}$ , 如果  $P$  和  $Q$  在论域  $U$  上的邻域粒化满足  $U/N_P \preceq U/N_Q$ , 则有  $E(P \cup Q) = E(P)$ .

**证明**  $P, Q$  在论域  $U$  上的邻域粒化满足  $U/N_P \preceq U/N_Q$ , 对于  $\forall x \in U$  有  $n_P(x) \subseteq n_Q(x)$ , 即  $n_P(x) \cap n_Q(x) = n_P(x)$ , 所以

$$\sum_{i=1}^n \log |n_P(x_i) \cap n_Q(x_i)| = \sum_{i=1}^n \log |n_P(x_i)|,$$

因此  $E(P \cup Q) = E(P)$ .  $\square$

性质5表明, 当两个邻域关系满足偏序关系时, 邻域联合粗糙熵的值是等于粒化更为精细的邻域关系对应的粗糙熵值, 也就是说邻域联合粗糙熵满足偏序不变性. 基于邻域粗糙熵和邻域联合粗糙熵, 接下来可以进一步导出邻域粗糙条件熵.

**定义7** 对于邻域信息系统  $\text{NIS} = (U, \text{AT}, V)$  且  $|U| = n$ , 设  $P, Q \subseteq \text{AT}$ , 由  $P$  和  $Q$  在论域  $U$  上诱导出的邻域粒化分别为  $U/N_P, U/N_Q$ . 定义邻域关系  $Q$

关于  $P$  的邻域粗糙条件熵为

$$E(Q|P) = -\frac{1}{n} \sum_{i=1}^n \log \frac{|n_P(x_i)|}{|n_P(x_i) \cap n_Q(x_i)|}. \quad (13)$$

同时,对于邻域决策信息系统  $NDIS = (U, C \cup D, V)$ , 决策属性  $D$  关于  $P$  的邻域粗糙条件熵为

$$E(D|P) = -\frac{1}{n} \sum_{i=1}^n \log \frac{|n_P(x_i)|}{|n_P(x_i) \cap [x_i]_D|}. \quad (14)$$

**性质6** 对于邻域信息系统为  $NIS = (U, AT, V)$  且  $|U| = n$ , 设  $P, Q \subseteq AT$ , 邻域粗糙条件熵满足  $E(Q|P) = E(P \cup Q) - E(P)$ .

**证明** 根据定义7有

$$\begin{aligned} E(Q|P) &= -\frac{1}{n} \sum_{i=1}^n \log \frac{|n_P(x_i)|}{|n_P(x_i) \cap n_Q(x_i)|} = \\ &= -\frac{1}{n} \sum_{i=1}^n (\log |n_P(x_i)| - \log |n_P(x_i) \cap n_Q(x_i)|) = \\ &= -\frac{1}{n} \sum_{i=1}^n \log |n_P(x_i)| + \frac{1}{n} \sum_{i=1}^n \log |n_P(x_i) \cap n_Q(x_i)| = \\ &= E(P \cup Q) - E(P). \quad \square \end{aligned}$$

**性质7** 对于邻域信息系统  $NIS = (U, AT, V)$  且  $|U| = n$ , 设  $P, Q \subseteq AT$ , 如果  $P$  和  $Q$  在论域  $U$  上的邻域粒化满足  $U/N_P \preceq U/N_Q$ , 则有  $E(Q|P) = 0$ .

性质7的证明可由性质5和性质6直接得到, 证明略。

由性质6可以看出, 邻域粗糙条件熵可以看成是单个邻域关系与联合邻域关系之间的信息增量, 由性质7可以得到, 如果两个邻域关系满足偏序关系, 则这种信息的增量为零。

**定义8** 对于邻域信息系统  $NIS = (U, AT, V)$  且  $|U| = n$ , 设  $P, Q \subseteq AT$ , 由  $P$  和  $Q$  在论域  $U$  上诱导出的邻域粒化分别为  $U/N_P, U/N_Q$ . 定义邻域关系  $Q$  关于  $P$  的邻域粗糙互信息熵为

$$I(Q; P) = -\frac{1}{n} \sum_{i=1}^n \log \frac{|n_P(x_i) \cap n_Q(x_i)|}{|n_P(x_i)| \cdot |n_Q(x_i)|}. \quad (15)$$

同时,对于邻域决策信息系统  $NDIS = (U, C \cup D, V)$ , 决策属性  $D$  关于  $P$  的邻域粗糙互信息熵为

$$I(D; P) = -\frac{1}{n} \sum_{i=1}^n \log \frac{|n_P(x_i) \cap [x_i]_D|}{|n_P(x_i)| \cdot |[x_i]_D|}. \quad (16)$$

**性质8** 对于邻域信息系统为  $NIS = (U, AT, V)$  且  $|U| = n$ , 设  $P, Q \subseteq AT$ , 邻域粗糙互信息熵满足:

- 1)  $I(Q; P) = I(P; Q)$ ;
- 2)  $I(Q; P) = E(P) - E(P|Q) = E(Q) - E(Q|P)$ ;
- 3)  $I(Q; P) = E(Q) + E(P) - E(P \cup Q)$ .

**证明** 由定义8显然有  $I(Q; P) = I(P; Q)$ , 因此式(1)成立. 因为

$$\begin{aligned} I(Q; P) &= -\frac{1}{n} \sum_{i=1}^n \log \frac{|n_P(x_i) \cap n_Q(x_i)|}{|n_P(x_i)| \cdot |n_Q(x_i)|} = \\ &= -\frac{1}{n} \sum_{i=1}^n \left( \log \frac{1}{|n_P(x_i)|} - \log \frac{|n_Q(x_i)|}{|n_P(x_i) \cap n_Q(x_i)|} \right) = \\ &= -\frac{1}{n} \sum_{i=1}^n \log \frac{1}{|n_P(x_i)|} - \\ &= \left[ -\frac{1}{n} \sum_{i=1}^n \log \frac{|n_Q(x_i)|}{|n_P(x_i) \cap n_Q(x_i)|} \right] = \\ &= E(P) - E(P|Q), \end{aligned}$$

同理有  $I(P; Q) = E(Q) - E(Q|P)$ , 所以式(2)成立。

由性质6有  $E(Q|P) = E(P \cup Q) - E(P)$ , 所以  $I(Q; P) = E(Q) + E(P) - E(P \cup Q)$ .  $\square$

在性质8的条件2)中,  $P$  和  $Q$  的邻域粗糙互信息熵为  $P$  的邻域粗糙熵减去  $P$  关于  $Q$  的邻域粗糙条件熵, 因此  $P$  和  $Q$  的邻域粗糙互信息熵反应了  $N_P$  和  $N_Q$  两个邻域关系公共的邻域粗糙信息量, 体现了两个属性集  $P$  和  $Q$  之间相关性程度, 这与其他形式互信息熵的结论是完全一致的<sup>[9]</sup>.

### 3 邻域粗糙互信息熵的非单调性属性约简算法

在目前基于邻域粗糙集模型的属性约简算法中, 首先是构建相应的属性集评估函数, 然后基于评估函数的单调性来启发式搜索约简集<sup>[4-14]</sup>. 近年来, 学者提出了一种非单调性的属性约简方法. 例如Li等<sup>[15]</sup>指出决策理论粗糙集模型中的决策属性正区域不满足单调性, 也就是说, 随着条件属性的增加, 决策属性的正区域可能会减小, 因此无法通过正区域的一致性来构造属性约简方法, 但是如果原始数据集的分类性能较差, 决策属性的正区域比较小, 属性约简应该得到正区域较大的约简集, 这样才能提高数据的分类性能. 因此, Li等<sup>[15]</sup>运用决策理论粗糙集模型中正区域的非单调性构造出了相应的属性约简算法, 该算法得到约简集的正区域不低于原始数据集的正区域, 从而提高了分类性能, 更具一定的优越性。

**定理1** 对于邻域决策信息系统  $NDIS = (U, C \cup D, V)$ ,  $P, Q \subseteq C$ , 并且  $P$  和  $Q$  在论域  $U$  上诱导出的邻域粒化分别为  $U/N_P, U/N_Q$ . 若  $U/N_P \preceq U/N_Q$ , 则邻域粗糙互信息熵  $I(D; P)$  和  $I(D; Q)$  之间的大小关系无法确定。

**证明** 若  $U/N_P \preceq U/N_Q$ , 则  $\forall x \in U$  有  $n_P(x) \subseteq n_Q(x)$ ,  $x$  的决策类为  $[x]_D$ . 令  $|n_P(x)| = a, |n_Q(x)| = b, |n_P(x) \cap [x]_D| = c, |n_Q(x) \cap [x]_D| = d$ . 显然有  $0 < a \leq b, 0 < c \leq d$ . 下面根据这一条件分析  $a/c$  与  $b/d$  之间的关系. 设  $\lambda = a/c - b/d$ , 即  $\lambda = (ad - bc)/(cd)$ ,

这里的  $cd > 0$ , 但是  $ad - bc$  的正负无法确定, 例如, 当  $a = b = c = d$  时, 有  $\lambda = 0$ . 若  $a = 3, b = 5, c = 1, d = 2$ , 则  $\lambda > 0$ ; 若  $a = 3, b = 7, c = 1, d = 2$ , 则  $\lambda < 0$ . 因此根据已知条件无法判断  $a/c$  与  $b/d$  的大小关系, 即  $|n_P(x)|/(|n_P(x) \cap [x]_D|)$  与  $|n_Q(x)|/(|n_Q(x) \cap [x]_D|)$  大小无法判断. 根据定义 8, 同样也无法判断  $I(D; P)$  和  $I(D; Q)$  之间的大小关系, 即邻域粗糙互信息熵也不满足单调性.  $\square$

在性质 8 中, 对于邻域决策信息系统  $\text{NDIS} = (U, C \cup D, V)$ , 设  $P \subseteq C$ , 邻域粗糙互信息熵  $I(D; P)$  表示的是属性子集  $P$  和决策属性  $D$  之间公共的邻域粗糙信息量, 如果  $I(D; P) > I(D; C)$ , 则说明属性子集  $P$  比属性全集  $C$  具有更高的决策属性公共邻域粗糙信息量, 也就是说, 属性子集  $P$  与决策属性  $D$  之间的相关性比属性全集  $C$  更高, 因此属性集  $P$  对该邻域信息系统具有更高的分类性能, 特别是对于属性全集  $C$  和决策属性  $D$  之间相关性较小的邻域信息系统,  $P$  作为该邻域信息系统的一个相对约简就更加合适. 类似于 Li 等<sup>[15]</sup> 的观点, 这里给出基于邻域粗糙互信息熵的非单调性属性约简算法.

**定义 9** 对于邻域决策信息系统  $\text{NDIS} = (U, C \cup D, V)$ , 邻域半径为  $\delta$ , 若  $B \subseteq C$  是该信息系统的约简当且仅当以下两点同时成立:

$$I(D; B) \geq I(D; C), \quad (17)$$

$$I(D; B - \{a\}) < I(D; B). \quad (18)$$

在定义 9 中, 式 (17) 表示约简集  $B$  的邻域粗糙互信息熵不能低于全属性集的邻域粗糙互信息熵, 式 (18) 限制了约简集的极小性.

**定义 10** 对于邻域决策信息系统  $\text{NDIS} = (U, C \cup D, V)$ , 邻域半径为  $\delta$ ,  $\forall a \in B \subseteq C$  关于  $B$  的属性重要度定义为

$$\text{sig}_{\text{in}}(a, B, D) = I(D; B) - I(D; B - \{a\}); \quad (19)$$

$\forall a \in (C - B)$  关于  $B$  的属性重要度定义为

$$\text{sig}_{\text{out}}(a, B, D) = I(D; B \cup \{a\}) - I(D; B). \quad (20)$$

在定义 10 中, 对于  $a \in B$ , 若  $a$  关于  $B$  是内部重要的, 则邻域粗糙互信息熵满足关系  $I(D; B) > I(D; B - \{a\})$ , 因此有  $\text{sig}_{\text{in}}(a, B, D) > 0$ , 反之  $\text{sig}_{\text{in}}(a, B, D) \leq 0$ . 对于  $a \in (C - B)$ , 若  $a$  关于  $B$  是外部重要的, 则  $I(D; B \cup \{a\}) > I(D; B)$ , 因此有  $\text{sig}_{\text{out}}(a, B, D) > 0$ , 反之  $\text{sig}_{\text{out}}(a, B, D) \leq 0$ . 将这两种属性重要度作为启发式函数, 本文提出一种启发式的属性约简算法.

**算法 1** 邻域粗糙互信息熵的非单调性属性约简算法.

输入:  $\text{NDIS} = (U, C \cup D, V)$ , 邻域半径  $\delta$ ;

输出: 约简集  $\text{red}$ .

Step 1: 初始化  $\text{red} = \emptyset$ .

Step 2: 对于  $\forall a \in C$ , 计算  $a$  的属性重要度  $\text{sig}_{\text{in}}(a, C, D)$ , 对于满足  $\text{sig}_{\text{in}}(a, C, D) > 0$  的属性进行  $\text{red} \leftarrow \text{red} \cup \{a\}$ .

Step 3: 根据定义 10 分别计算约简集  $\text{red}$  和全集  $C$  关于决策属性  $D$  的邻域粗糙互信息熵, 分别表示为  $I(D; \text{red}), I(D; C)$ . 若满足关系  $I(D; \text{red}) \geq I(D; C)$ , 则进入 Step 5; 否则进入 Step 4.

Step 4: 对于  $\forall a \in (C - \text{red})$ , 计算每个属性的属性重要度  $\text{sig}_{\text{out}}(a, \text{red}, D)$ , 并选择属性重要度最大的属性  $a'$  进行  $\text{red} = \text{red} \cup \{a'\}$ , 并进入 Step 3.

Step 5: 令  $\Omega = \emptyset$ , 对于  $\forall a \in \text{red}$ , 若属性  $a$  满足  $I(D; \text{red} - \{a\}) = I(D; \text{red})$ , 则  $\Omega \leftarrow \Omega \cup \{a\}$ .

Step 6: 判断  $\Omega$  是否为空集, 如果为空, 则直接进入 Step 7; 否则, 选取  $\Omega$  中任意一个属性  $b$  进行  $\text{red} \leftarrow \text{red} - \{b\}$ , 并进入 Step 5.

Step 7: 返回  $\text{red}$ .

在算法 1 中, 算法的输入包含邻域半径  $\delta$ , 不同的邻域半径将影响到邻域粗糙互信息熵的结果, 为了简洁, 在算法中省略了  $\delta$  的标记.

在整个算法的流程中, Step 2 是一个启发式搜索的过程, 主要是在属性全集  $C$  中选取对分类具有相关作用的属性. 该步骤的计算量主要集中在邻域粗糙互信息熵方面, 而邻域粗糙互信息熵主要是对象邻域的计算, 采用文献 [7] 所提出的快速计算邻域的算法, 则 Step 2 的时间复杂度为  $O(|C||U| \log |U|)$ . Step 3 是一个约简集的评估过程, 如果邻域粗糙互信息熵低于属性全集的熵值, 则需要在 Step 4 中对剩余的属性集进行属性搜索, 直到满足终止条件, 这两步骤的时间复杂度较低. 至此, 算法得到的约简集满足了  $I(D; \text{red}) \geq I(D; C)$ , 但是这时的  $\text{red}$  并不一定满足式 (18), 这主要是由于约简集  $\text{red}$  中还可能会包含一些冗余属性, 这时需要通过 Step 5 和 Step 6 进行反向冗余剔除, 最终保证了约简集能够同时满足式 (17) 和 (18). 算法 1 的计算量主要集中在 Step 2, 所以整个算法 1 的时间复杂度为  $O(|C||U| \log |U|)$ .

## 4 实验分析

本节将进行一系列实验来验证本文所提出算法的属性约简性能. 在进行实验之前, 首先需要选取参与对比的算法和进行实验的数据集, 根据本文所提出的属性约简的所属类型, 选取 8 种相关的属性约简算法: 1) 基于决策理论粗糙集模型的非单调属性约简<sup>[15]</sup>, 记为算法 A; 2) 一种新的模糊粗糙集信息

熵的特征选择算法<sup>[20]</sup>,记为算法B;3)基于邻域决策理论粗糙集的属性约简算法<sup>[10]</sup>,记为算法C;4)基于邻域区分度量的特征选择算法<sup>[9]</sup>,记为算法D;5)基于邻域粗糙集的正区域特征选择算法<sup>[4]</sup>,记为算法E;6)基于邻域组合测度的属性约简方法<sup>[13]</sup>,记为算法F;7)基于邻域粒化和粗糙逼近的数值属性约简<sup>[5]</sup>,记为算法G;8)基于高斯核的模糊粗糙集特征选择算法<sup>[6]</sup>,记为算法H.记本文所提出的邻域粗糙互信息熵的非单调属性约简为算法I.

本实验从UCI机器学习数据库和Keng Ridge数据库中总共选取了8个数据集,具体如表1所示.所有算法均采用Matlab2010b进行编程实现,实验运行的硬件环境为Intel(R)Core(TM)i7-6700K CPU和32GB RAM.本实验总共分为4个部分:第1部分为9种算法对表1所示各个数据集的属性约简结果比较;第2部分为9种算法对应属性约简结果的分类精度比较;第3部分为9种算法的属性约简计算时间比较;第4部分主要探究本文所提出的算法I中邻域半径 $\delta$ 对属性约简结果的影响以及邻域半径 $\delta$ 的选取问题.

表1 数据集信息

编号	数据集名称	对象	属数	类
1	iono	351	34	2
2	sonar	208	60	2
3	wdbc	569	31	2
4	libras	360	90	15
5	sick	2 800	29	2
6	gerbox	1 603	72	4
7	prostate	136	12 600	2
8	tumors	327	12 588	7

### 4.1 属性约简结果比较

表1中所示均为连续型数据集,为了消除属性量纲的影响,对其进行归一化处理.而在参与比较的9种算法中,算法A只能处理离散型的数据集,因此算法A在进行属性约简之前,需要将这8个数据集进行离散化处理,这里直接采用WEKA软件中的等频离散化法进行数值型数据离散化.对于本文所提出的算法I以及目前已提出的算法C~算法G,邻域半径 $\delta$ 是一个非常重要的参数,它的取值不同对最终的结果具有很大的影响,因此通过对邻域半径 $\delta$ 取不同值分别进行实验.由于各个数据集的属性数量并不一样,因而很难对邻域半径 $\delta$ 进行统一选取.为了解决这一问题,需重新定义邻域的选取方法,即定义邻域半径 $\delta = \Delta_{\min} + k(\Delta_{\max} - \Delta_{\min})$ ,其中 $k \in [0, 1]$ , $\Delta_{\min}$ 表示数据集所有对象之间距离的最小值, $\Delta_{\max}$ 表示数据集所有对象之间距离的最大值,这样即可通过选取 $k$ 值间接选取 $\delta$ .通过选取不同的 $k$ 进行实验,便得到了对应的属性约简结果.通过分类器CART和SVM分别对不同 $k$ 值得到的约简集进行十折交叉分类精度计算,由于CART和SVM分类精度最高时的约简集往往不相同,算法C~算法G选取二者分类精度最高时的约简集作为最终的属性约简结果.而对于本文所提出的算法I,将两种分类器各自分类精度达到最高时的属性约简结果展示出来,具体如表2所示,其中的数值表示的是约简结果的属性序号,圆括号里的值表示约简结果的大小.表3所示的是9种算法在每个数据集下属性约简大小的比较,其中圆括号里的值表示的是该约简结果对应的 $k$ 值.

表2 算法I的属性约简结果

数据集	CART	SVM
iono	1, 3, 4, 5, 7, 8, 12, 14, 19, 25, 30, 34, (12)	1, 4, 5, 14, 16, 17, 19, 25, 26, 34, (10)
sonar	1, 12, 13, 17, 21, 26, 42, 48, 55, (9)	1, 6, 12, 14, 21, 26, 42, 48, 55, (9)
wdbc	7, 10, 15, 22, 23, 26, 28, (7)	6, 9, 15, 16, 23, 25, (6)
libras	7, 12, 14, 15, 21, 27, 36, 46, 53, 58, 65, 67, (12)	12, 14, 19, 25, 27, 36, 46, 49, 53, 67, (10)
sick	1, 2, 3, 6, 10, 17, 18, 19, 20, 22, 24, 26, 29, (13)	1, 2, 10, 15, 16, 17, 19, 20, 22, 24, 29, (11)
gerbox	2, 11, 15, 17, 25, 29, 33, 48, 56, 63, 67, (11)	2, 10, 11, 15, 17, 25, 27, 29, 36, 56, 63, 67, (12)
prostate	3 462, 4 483, 6 257, 6 984, 8 623, 8 854, 9 527, 11 321, 12 067, (9)	857, 3 462, 4 483, 5 127, 6 984, 8 623, 9 071, 9 527, 12 067, (9)
tumors	1 573, 2 132, 3 264, 4 300, 5 411, 6 079, 6 321, 6 452, 8 319, 8 674, 8 825, 9 875, 10 784, (13)	2 132, 2 782, 3 588, 4 300, 5 411, 6 325, 6 764, 7 542, 8 574, 8 676, 10 750, 12 369, (12)

表3 9种算法的约简集大小比较

数据集	算法A	算法B	算法C	算法D	算法E	算法F	算法G	算法H	算法I+CART	算法I+SVM
iono	15	13	9 (0.16)	11 (0.13)	12 (0.18)	14 (0.28)	14 (0.25)	11	12 (0.25)	10 (0.09)
sonar	14	10	12 (0.22)	11 (0.24)	10 (0.12)	12 (0.22)	13 (0.29)	10	9 (0.23)	9 (0.13)
wdbc	13	10	11 (0.21)	7 (0.09)	8 (0.08)	9 (0.15)	12 (0.23)	10	7 (0.22)	6 (0.14)
libras	12	12	11 (0.15)	13 (0.19)	10 (0.13)	12 (0.25)	12 (0.21)	12	12 (0.21)	10 (0.12)
sick	12	12	13 (0.31)	10 (0.17)	9 (0.12)	11 (0.24)	12 (0.18)	10	13 (0.17)	11 (0.26)
gerbox	13	10	10 (0.14)	12 (0.23)	11 (0.18)	10 (0.13)	11 (0.17)	11	11 (0.14)	12 (0.21)
prostate	12	10	9 (0.06)	10 (0.16)	12 (0.19)	10 (0.17)	11 (0.10)	10	9 (0.15)	9 (0.18)
tumors	16	16	13 (0.18)	15 (0.12)	14 (0.11)	15 (0.31)	15 (0.25)	14	13 (0.22)	12 (0.14)

在表3中,“算法I+**CART**”表示的是采用**CART**分类器得到的约简集,“算法I+**SVM**”表示的是采用**SVM**分类器得到的约简集.由表3可以看出,经过约简后的属性子集均小于原始数据的属性集,说明对数据集进行约简降维很有必要.比较这9种算法的属性约简结果,本文所提出的算法I在部分数据集的约简集更小一些,例如**CART**分类器下的sonar, wdbc, prostate和tumors数据集,**SVM**分类器下的iono, sonar, wdbc, prostate和tumors数据集,算法C在数据集iono上拥有更小约简集,算法E在sick和libras数据集上拥有更小约简集.而算法A在大部分数据集具有更大的约简集,这主要是由于算法A所运行的数据集是经过离散化的,离散化的过程可能会丢

失数据的分类信息,影响了属性约简的结果.对于数值型的数据集,直接对其进行属性约简具有更好的约简效果.

4.2 分类精度结果比较

在4.1节中给出了9种算法的属性约简结果比较,但是小的约简集并不代表有好的分类性能,同样地,大的约简集也并不代表有差的分类性能.因此,运用**CART**和**SVM**分类器分别对表3的8个数据集的属性约简结果进行十折交叉分类训练,得到了对应的分类精度值,具体如表4和表5所示.其中表4所示的是**CART**分类器下分类精度比较,表5所示的是**SVM**分类器下的分类精度比较.

表4 CART分类精度结果比较

数据集	算法A	算法B	算法C	算法D	算法E	算法F	算法G	算法H	算法I
iono	88.42	84.48	85.36	86.12	85.65	86.32	85.27	87.35	89.62√
sonar	74.26	73.94	72.82	74.62	74.55	73.84	73.54	73.64	74.67√
wdbc	92.88√	91.63	88.37	88.45	89.52	87.26	90.38	89.73	90.24
libras	77.28	78.37√	75.59	77.82	74.95	76.12	76.18	75.36	77.35
sick	83.05	81.82	82.63	80.75	81.69	81.17	82.43	82.07	84.75√
gerbox	73.56	72.15	74.25	73.03	73.56	72.95	74.35√	72.68	74.35√
prostate	75.15	73.69	72.95	78.24√	75.06	74.85	75.32	74.15	75.65
tumors	68.78	65.26	64.28	65.82	63.59	66.56	66.75	64.28	69.39√

表5 SVM分类精度结果比较

数据集	算法A	算法B	算法C	算法D	算法E	算法F	算法G	算法H	算法I
iono	88.56	86.58	85.73	89.42	88.43	87.63	88.45	86.75	90.45√
sonar	76.24	76.08	77.85√	75.64	72.46	74.25	72.35	74.86	75.52
wdbc	90.47	88.36	87.72	86.68	85.96	89.76	85.67	86.74	92.48√
libras	81.27	78.46	80.56	82.45	77.58	79.18	80.17	82.45√	81.57
sick	86.79	85.76	83.27	85.48	84.09	84.38	82.47	85.39	87.78√
gerbox	77.45√	73.53	74.82	72.93	73.36	74.54	74.63	72.40	75.73
prostate	76.08	73.82	74.85	75.19	74.32	75.49	75.24	73.68	77.28√
tumors	68.57	64.47	65.19	66.46	68.47	67.48	66.04	68.56	70.32√

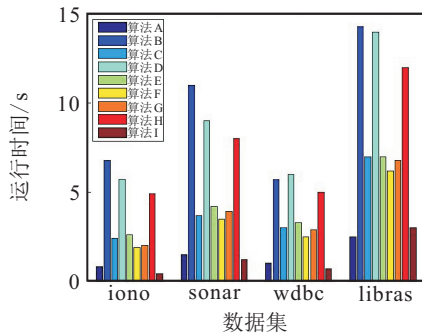
在表4和表5中,“√”标记表示的是9种算法分类精度的最大值.观察表4可以发现,在大部分数据集中,算法I的属性约简结果具有更高的分类精度,算法A约简结果的性能也比较高,而其他8种算法只有在少部分数据集拥有更高的分类精度,如prostate和libras数据集,这主要是由于算法B~算法H均是基于单调性的启发式属性约简算法,虽然采用了不同的属性集评估函数,如算法B采用了基于模糊粗糙集的信息熵度量<sup>[20]</sup>,算法D采用了邻域区分度度量<sup>[9]</sup>,算法F采用了邻域组合度量<sup>[13]</sup>,算法G采用依赖度量<sup>[5]</sup>,以及算法H采用高斯核的模糊粗糙集的方法<sup>[6]</sup>,但是在进行属性约简的过程中,各种算法都是以最终约简集的度量结果和属性全集的度量结果保持一致作为终止条件,如果原始数据集的分类性

能不是很理想,则得到的约简集也就不会有很好的分类性能.因此,对于原始分类性能不是很理想的数据集,基于单调性的属性约简不能够得到较优的属性约简结果.相反,对于算法A和算法I的非单调性属性约简算法,在属性约简的过程中,其约简集的分类性能度量值控制不低于原始数据集的度量值,最终的约简集具有更好的分类性能.这就解释了算法A和算法I的约简结果具有更高分类精度的原因.相比算法I,算法A的分类精度较低一点,主要是由于数据集离散化后对属性约简的影响.同样地,表5所示的**SVM**分类精度结果同样能得到相同的结论.

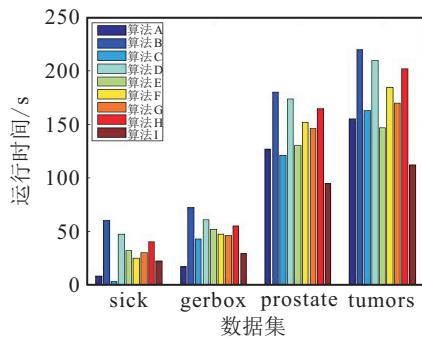
4.3 属性约简运行时间比较

图1所示为9种算法在各个数据集上的属性约简用时比较,其中运行时间的结果取多次约简用时的

平均值. 观察图1可发现,对于规模较小的数据集,9种算法的运行时间相接近,当数据集的规模逐渐增大时,9种算法属性约简的运行时间出现了明显的差别.



(a) 部分数据集I



(b) 部分数据集II

图1 9种算法运行时间比较

算法A的时间复杂度为 $O(|C|^2|U|)$ <sup>[15]</sup>,由于算法A针对的是符号型数据集,算法A在大部分数据集的约简运行时间较低. 算法B、算法D和算法H的时间复杂度为 $O(|C|^2|U|^2)$ <sup>[6,9,20]</sup>,算法C、算法E、算法F和算法G的时间复杂度为 $O(|C|^2|U| \log |U|)$ <sup>[4-5,10,13]</sup>,本文所提出的算法I的时间复杂度为 $O(|C||U| \log |U|)$ . 因此,大部分数据集拥有较低的约简耗时,特别是对一些规模较大的数据集,例如 prostate 和 tumors,算法I和其他算法差距还是相当明显的.

通过9种算法对各个数据集的属性约简结果、分类精度和约简用时进行了比较,验证了本文所提出的属性约简算法具有较好的性能.

#### 4.4 邻域半径 $\delta$ 的探究

由于邻域半径  $\delta$  对本文所提出的算法I的约简结果具有很重要的影响,接下来将探究邻域半径  $\delta$  和约简结果之间的变化关系. 首先,同样采用4.1节中利用  $k$  值来选取邻域半径的方式,让  $k$  从0至1以0.05作为间隔进行赋值;然后将对应的邻域半径进行属性约简;最后对得到的属性约简结果分别计算出CART和SVM分类精度. 图2所示的是部分数据集约简集大小和两种分类精度关于  $k$  值的关系曲线.

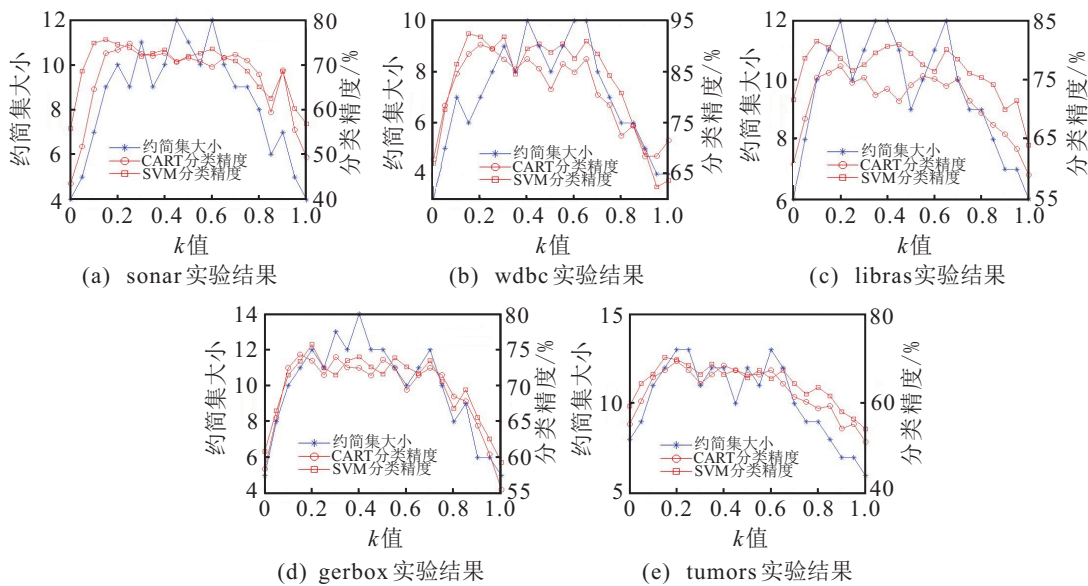


图2 部分数据集实验结果

由图2中部分数据集的实验结果可以看出: $k$ 从0至1逐渐增加的过程中,约简集的大小开始逐渐增大,当达到一定值后趋于稳定,当 $k$ 值增加到一定程度后其约简集大小逐渐降低;对应的CART和SVM分类精度一开始也是逐渐增大的,接着便趋于稳定,最后逐渐降低. 这主要是由于开始时邻域半径选取的过小,每个对象的邻域类比较小,较小的邻域类不

能够具有较好的属性评估效果. 当邻域半径选取过大时,较大的邻域类对对象之间的区分能力较低,同样不能具有较好的属性评估效果,从而会降低分类精度. 另外,综合图2部分数据集的结果可以看出,当 $k$ 取 $[0.1, 0.3]$ 时,约简结果具有较高的分类精度,因此对于本文所提出的属性约简算法,其 $k$ 值取 $[0.1, 0.3]$ 是一个较为合适的范围.

## 5 结论

本文针对目前邻域粗糙集模型中单调性属性约简存在的缺陷,首先在邻域粗糙集模型中定义了邻域粗糙熵的概念;然后给出了邻域粗糙联合熵、邻域粗糙条件熵和邻域粗糙互信息熵,邻域粗糙互信息熵是对属性集相关性大小的一种重要体现.文中证明了邻域粗糙互信息熵是随属性集增加而不是单调变化的,利用这一性质,提出了一种非单调性的属性约简算法,UCI实验结果验证了该算法相比其他算法具有更好的属性约简性能.在本文所提出的算法中,如何得到较优的约简结果,与邻域半径的大小有着很直接的关系,而实验中只给出一个合适的邻域半径选取区间,因而,通过理论的方法选择出更精准的邻域半径将是接下来的研究课题.

### 参考文献(References)

- [1] Pawlak Z. Rough sets[J]. *Int J of Parallel Programming*, 1982, 11(5): 341-356.
- [2] 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(9/10): 597-618.
- [3] Hu Q H, Yu D R, Xie Z X. Neighborhood classifiers[J]. *Expert Systems with Applications*, 2008, 34(2): 866-876.
- [4] Hu Q H, Yu D R, Liu J F, et al. Neighborhood rough set based heterogeneous feature subset selection[J]. *Information Sciences*, 2008, 178(18): 3577-3594.
- [5] 胡清华, 于达仁, 谢宗霞. 基于邻域粒化和粗糙逼近的数值属性约简[J]. *软件学报*, 2008, 19(3): 640-649. (Hu Q H, Yu D R, Xie Z X. Numerical Attribute Reduction Based on Neighborhood Granulation and Rough Approximation[J]. *J of Software*, 2008, 19(3): 640-649.)
- [6] Hu Q H, Zhang L, Chen D G, et al. Gaussian kernel based fuzzy rough sets: Model, uncertainty measures and applications[J]. *Int J of Approximate Reasoning*, 2010, 51(4): 453-471.
- [7] Liu Y, Huang W L, Jiang Y L, et al. Quick attribute reduct algorithm for neighborhood rough set model[J]. *Information Sciences*, 2014, 271(7): 65-81.
- [8] Wang C Z, Shao M W, He Q, et al. Feature subset selection based on fuzzy neighborhood rough sets[J]. *Knowledge-Based Systems*, 2016, 111(1): 173-179.
- [9] Wang C Z, Hu Q H, Wang X Z, et al. Feature selection based on neighborhood discrimination index[J]. *IEEE Trans on Neural Networks and Learning Systems*, 2018, 29(7): 2986-2999.
- [10] Li W W, Huang Z Q, Jia X Y, et al. Neighborhood based decision-theoretic rough set models[J]. *Int J of Approximate Reasoning*, 2016, 69: 1-17.
- [11] 姚晟, 徐风, 赵鹏, 等. 基于邻域量化容差关系粗糙集模型的特征选择算法[J]. *模式识别与人工智能*, 2017, 30(5): 416-428. (Yao S, Xu F, Zhao P, et al. Feature selection algorithm based on neighborhood valued tolerance relation rough set model[J]. *Pattern Recognition and Artificial Intelligence*, 2017, 30(5): 416-428.)
- [12] 姚晟, 徐风, 赵鹏, 等. 基于改进邻域粒的模糊熵特征选择算法[J]. *南京大学学报: 自然科学版*, 2017, 53(4): 802-814. (Yao S, Xu F, Zhao P, et al. Fuzzy entropy feature selection algorithm based on improved neighborhood granule[J]. *J of Nanjing University: Natural Science*, 2017, 52(4): 802-814.)
- [13] 何松华, 康婵娟, 鲁敏, 等. 基于邻域组合测度的属性约简方法[J]. *控制与决策*, 2016, 31(7): 1225-1230. (He S H, Kang C J, Lu M, et al. Attribute reduction method based on neighborhood combination measure[J]. *Control and Decision*, 2016, 31(7): 1225-1230.)
- [14] 段洁, 胡清华, 张灵均, 等. 基于邻域粗糙集的多标记分类特征选择算法[J]. *计算机研究与发展*, 2015, 52(1): 56-65. (Duan J, Hu Q H, Zhang L J, et al. Feature selection for multi-Label classification based on neighborhood rough sets[J]. *J of Computer Research and Development*, 2015, 52(1): 56-65.)
- [15] Li H X, Zhou X Z, Zhao J B, et al. Non-monotonic attribute reduction in decision-theoretic rough sets[J]. *Fundamenta Informaticae*, 2013, 126(4): 415-432.
- [16] Liang J Y, Shi Z Z, Li D Y, et al. Information entropy, rough entropy and knowledge granulation in incomplete information systems[J]. *Int J of General Systems*, 2006, 35(6): 641-654.
- [17] Qian Y H, Liang J Y. Combination entropy and combination granulation in incomplete information system[C]. *The 1st Int Conf on Rough Sets and Knowledge Technology*, Chongqing, 2006: 184-190.
- [18] Qian Y H, Liang J Y, Wu W Z, et al. Information granularity in fuzzy binary GrC model[J]. *IEEE Trans on Fuzzy Systems*, 2011, 19(2): 253-264.
- [19] 徐风, 姚晟, 纪霞, 等. 基于模糊邻域粗糙集的信息系统不确定性度量方法[J]. *南京大学学报: 自然科学版*, 2017, 53(5): 926-936. (Xu F, Yao S, Ji X, et al. Uncertainty measurement method for information system based on fuzzy neighborhood rough set[J]. *J of Nanjing University: Natural Science*, 2017, 53(5): 926-936.)
- [20] Zhang X, Mei C L, Chen D G, et al. Feature selection in mixed data: A method using a novel fuzzy rough set-based information entropy[J]. *Pattern Recognition*, 2016, 56(1): 1-15.

(责任编辑: 孙艺红)