

引用格式:徐岩柏,景运革. 分布决策信息系统增量属性约简算法[J]. 微电子学与计算机,2020,37(9):31-36. [XU Y B,JING Y G. An incremental attribute reduction algorithm for distributed decision information system[J]. Microelectronics & Computer,2020,37(9):31-36.]

# 分布决策信息系统增量属性约简算法

徐岩柏,景运革

(运城学院,数学与信息技术学院,山西 运城 044000)

**摘 要:** 为了解决数据动态变化后如何快速更新属性约简的问题,一些增量属性约简的算法被提出.然而,对于分布决策信息系统增量属性约简的算法研究却甚少.为了有效解决动态分布决策信息系统约简更新的问题,提出了分布决策信息系统增量属性约简的方法.首先,本文介绍了分布决策系统的相关概念,然后,给出了分布决策信息系统基于矩阵方法的知识粒度增量更新机制,并设计了分布决策信息系统基于矩阵方法的增量属性约简算法.最后,从 UCI 机器学习库下载一些数据集,分别用增量和非增量属性约简算法进行对比实验,实验结果说明了增量属性约简算法比非增量约简算法在计算时间上有较强的优势.

**关键词:** 分布决策信息系统;属性约简;增量学习;知识粒度

**中图分类号:** TP311

**文献标识码:** A

**文章编号:** 1000—7180(2020)09—0031—06

## An incremental attribute reduction algorithm for distributed decision information system

XU Yan-bai,JING Yun-ge

(Maths & Information Technology School, Yuncheng University, Yuncheng 044000, China)

**Abstract:** How to effectively carry out attribute reduction is an important task when some data vary dynamically in the decision system, many incremental attribute reduction methods have been proposed. Whereas there is less work on incremental attribute reduction method for distributed decision information system. In order to address this problem, a matrix-based incremental attribute reduction algorithm is proposed for distributed data. In this paper, we first introduce some definitions and conceptions of distributed decision information system. Then, the matrix-based incremental mechanisms for knowledge granularity are discussed and the corresponding matrix-based incremental attribute reduction algorithm for updating reduct is developed. Finally, some data sets is downloaded from the UCI machine learning libraries and some numerical experiments demonstrate that the incremental attribute reduction algorithm is effective and efficient compared to non-incremental attribute reduction algorithm.

**Key words:** distributed decision information system; attribute reduction; incremental learning; knowledge granularity

### 1 引言

由于粗糙集理论在处理不确定性和不精确性复

杂问题具有较大的优势,因此它已经被应用到数据挖掘和模式识别等领域中.近年来,许多研究者提出了基于粗糙集的知识发现方法<sup>[1]</sup>.但是,这些方法主

要是处理静态数据知识发现问题. 而对于动态数据知识发现问题, 需要重新计算, 不能利用原有的运算结果而浪费了大量的运行时间, 导致计算效率不是很高, 特别是对于规模较大数据集时, 静态计算方法不适合解决动态数据集知识发现的问题.

增量学习技术主要是解决动态数据数据挖掘和知识发现的方法, 该方法可以充分利用原有的运算结果, 避免了对整个数据集进行重新计算, 降低运算时间并提高运算效率<sup>[2]</sup>. 近年来, 为了有效解决动态数据知识发现问题, 已经有一些学者把增量学习技术应用到粗糙集理论研究中<sup>[3-4]</sup>.

目前信息决策信息系统的增量属性约简算法主要分为三类, 基于对象集<sup>[5-8]</sup>、属性集<sup>[9-12]</sup>和属性值变化的增量属性约简算法<sup>[13-15]</sup>. 我们发现许多增量属性约简算法主要解决单个决策信息系统对象、属性及属性值动态变化如何快速更新属性约简的问题, 但是对于分布决策信息系统中数据动态变化的增量属性约简算法研究甚少. 因此, 本文提出了决策分布信息系统对象变化后的增量属性约简去解决分布决策信息系统对象变化后快速更新约简的问题.

## 2 分布决策信息系统属性约简算法

### 2.1 分布决策信息系统的基础知识

分布决策信息系统中的基本概念<sup>[8,17]</sup>.

**定义 1:**  $DS = (U, A, V, F) = \bigcup_{i=1}^m S_i$  是分布决策信息系统, 分布信息系统相互独立的子信息系统为  $S_i = (U_i, A_i, V_i, f_i)$  是  $DS$  相互独立的子分布决策信息系统,  $DS$  必须满足如下条件:

- (1)  $U = \bigcup_{i=1}^m U_i$ ;
- (2)  $U_j \cap U_k \neq \emptyset, \forall j, k \in \{1, 2, \dots, m\}$  且  $j \neq k$ ;
- (3)  $A = A_i$ ;
- (4)  $V = \bigcup_{i=1}^m V_i$ ;
- (5)  $f = f_i$ ;

**定义 2:**  $DS = (U, A, V, F) = \bigcup_{i=1}^m S_i$  是分布决策信息系统,  $S_i = (U_i, A_i, V_i, f_i)$  和  $S_j = (U_j, A_j, V_j, f_j)$  是分布决策信息系统任意两个独立的子决策信息系统,  $n_i = |U_i|, n_j = |U_j|, x_l \in |U_i| (l = 1, 2, \dots, n_i)$  及  $x_k \in |U_j| (l = 1, 2, \dots, n_j)$ ,  $S_i$  和  $S_j$  的等价关系矩阵是  $M_{(n_i \times n_j)R_C}^{ij} = (m_{lk}^{ij})_{n_i \times n_j}, M_{(n_i \times n_j)R_C}^{ij}$  的元素被定义为:

$$m_{lk}^{ij} = \begin{cases} 1, & V_C(x_l) = V_{jC}(x_k) \\ 0, & V_C(x_l) \neq V_{jC}(x_k) \end{cases} \quad (1)$$

用  $M_{n_i \times n_j}^{ij}$  表示  $M_{(n_i \times n_j)R_C}^{ij}$ .

**定义 3:**  $DS = (U, A, V, F) = \bigcup_{i=1}^m S_i$  是分布决策信息系统,  $S_i = (U_i, A_i, V_i, f_i)$  和  $S_j = (U_j, A_j, V_j, f_j)$  是分布决策信息系统任意两个独立的子决策信息系统,  $M_{n_i \times n_j}^{ij}$  是  $S_i$  和  $S_j$  的等价关系矩阵, 则分布决策信息系统的知识粒度被定义为:

$$GP_{\bigcup_{i=1}^m U_i}^m = \frac{\sum_{i=1}^m \sum_{j=1}^m \text{Sum}(M_{n_i \times n_j}^{ij})}{(\sum_{i=1}^m |U_i|)^2} \quad (2)$$

则  $\text{Sum}(M_{n_i \times n_j}^{ij})$  表示  $M_{n_i \times n_j}^{ij}$  中所有元素相加.

**定义 4:**  $DS = (U, A, V, F) = \bigcup_{i=1}^m S_i$  是分布决策信息系统,  $S_i = (U_i, A_i, V_i, f_i)$  和  $S_j = (U_j, A_j, V_j, f_j)$  是分布决策信息系统任意两个独立的子决策信息系统,  $S_j$  和  $S'$  在  $C$  和  $C \cup D$  下的等价关系为  $M_{n_i \times n_j}^{ij}$  及  $M'_{n_i \times n_j}{}^{ij}$ , 则分布决策信息的相对知识粒度被定义为:

$$GP_{\bigcup_{i=1}^m U_i}^m(D) = \frac{\sum_{i=1}^m \sum_{j=1}^m \text{Sum}(M_{n_i \times n_j}^{ij})}{(\sum_{i=1}^m |U_i|)^2} - \frac{\sum_{i=1}^m \sum_{j=1}^m \text{Sum}(M'_{n_i \times n_j}{}^{ij})}{(\sum_{i=1}^m |U_i|)^2} \quad (3)$$

**定义 5:**  $DS = (U, A, V, F) = \bigcup_{i=1}^m S_i$  是分布决策信息系统,  $\forall a \in C, S_i = (U_i, A_i, V_i, f_i)$  和  $S_j = (U_j, A_j, V_j, f_j)$  是分布决策信息系统任意两个独立的子决策信息系统,  $M_{K_{C-(a)}}^{ij}, M_{K_{C-(a)} \cup D}^{ij}, M_{K_C}^{ij}$  及  $M_{K_{C \cup D}}^{ij}$  分别是  $S_i$  和  $S_j$  在不同关系下的等价关系矩阵. 分布决策信息系统  $a$  关于  $C$  相对于  $D$  的内重要性被定义为:

$$\begin{aligned} \text{Sig}_{\bigcup_{i=1}^m U_i}^{\text{inter}}(a, C, D) &= \frac{\sum_{i=1}^m \sum_{j=1}^m \text{Sum}(M_{K_{C-(a)}}^{ij})}{(\sum_{i=1}^m |U_i|)^2} \\ &- \frac{\sum_{i=1}^m \sum_{j=1}^m \text{Sum}(M_{K_{C-(a)} \cup D}^{ij})}{(\sum_{i=1}^m |U_i|)^2} - \frac{\sum_{i=1}^m \sum_{j=1}^m \text{Sum}(M_{K_C}^{ij})}{(\sum_{i=1}^m |U_i|)^2} \\ &+ \frac{\sum_{i=1}^m \sum_{j=1}^m \text{Sum}(M_{K_{C \cup D}}^{ij})}{(\sum_{i=1}^m |U_i|)^2} \end{aligned} \quad (4)$$

**定义 6:**  $DS = (U, A, V, F) = \bigcup_{i=1}^m S_i$  是分布决策信息系统, 分布决策信息系统的核被定义为:

$$\text{Core}_{\bigcup_{i=1}^m U_i}^m(D) = \{a \in C \mid \text{Sig}_{\bigcup_{i=1}^m U_i}^{\text{inter}}(a, C, D) > 0\}$$

(5)

**定义 7:**  $DS = (U, A, V, F) = \bigcup_{i=1}^m S_i$  是分布决策信息系统,  $B \subseteq C, \forall a \in (C - B), S_i = (U_i, A_i, V_i, f_i)$  和  $S_j = (U_j, A_j, V_j, f_j)$  是分布决策信息系统任意两个独立的子决策信息系统,  $M_{R_B}^{ij}, M_{R_{B \cup D}}^{ij}, M_{R_{B \cup \{a\}}}^{ij}$  及  $M_{R_{B \cup \{a\}} \cup D}^{ij}$  分别是  $S_i$  和  $S_j$  在不同关系下的等价关系矩阵. 分布决策信息系统属性  $a$  关于  $C$  相对于  $D$  的外重要性被定义为:

$$\begin{aligned} \text{Sig}_{\bigcup_{i=1}^m U_i}^{\text{inter}}(a, C, D) = & \frac{\sum_{i=1}^m \sum_{j=1}^m \text{Sum}(M_{R_B}^{ij})}{(\sum_{i=1}^m |U_i|)^2} \\ & - \frac{\sum_{i=1}^m \sum_{j=1}^m \text{Sum}(M_{R_{B \cup D}}^{ij})}{(\sum_{i=1}^m |U_i|)^2} - \frac{\sum_{i=1}^m \sum_{j=1}^m \text{Sum}(M_{R_{B \cup \{a\}}}^{ij})}{(\sum_{i=1}^m |U_i|)^2} \\ & + \frac{\sum_{i=1}^m \sum_{j=1}^m \text{Sum}(M_{R_{B \cup \{a\}} \cup D}^{ij})}{(\sum_{i=1}^m |U_i|)^2} \end{aligned} \quad (6)$$

**定义 8:**  $DS = (U, A, V, F) = \bigcup_{i=1}^m S_i$  是分布决策信息系统,  $B \subseteq C, B$  是分布决策信息系统的约简必须满足如下条件:

- (1)  $GP_{\bigcup_{i=1}^m U_i}^m(D | C) = GP_{\bigcup_{i=1}^m U_i}^m(D | B)$
- (2)  $\forall a \in B, GP_{\bigcup_{i=1}^m U_i}^m(D | B - \{a\}) \neq GP_{\bigcup_{i=1}^m U_i}^m(D | B)$

## 2.2 分布决策信息系统静态约简算法

分布决策信息系统静态约简算法具体步骤如下:

输入: 分布决策信息系统  $DS = (U, A, V, F) = \bigcup_{i=1}^m S_i$ .

输出: 分布决策信息系统的约简  $\text{RED}_{\bigcup_{i=1}^m U_i}^m$ .

Setp1 计算分布决策信息系统的核  $\text{Core}_{\bigcup_{i=1}^m U_i}^m(D)$ ;

Setp2  $B \leftarrow \text{Core}_{\bigcup_{i=1}^m U_i}^m(D)$ ;

Setp3 计算分布决策信息系统属性集  $B$  和  $C$  相对知识粒度  $GP_{\bigcup_{i=1}^m U_i}^m(D)$  和  $GP_c^{\bigcup_{i=1}^m U_i}(D)$ ;

Setp4 如果  $GP_{\bigcup_{i=1}^m U_i}^m(D) = GP_c^{\bigcup_{i=1}^m U_i}(D)$ , 则执行 Setp4, 否则执行 Setp6;

Setp5 当  $GP_{\bigcup_{i=1}^m U_i}^m(D) \neq GP_c^{\bigcup_{i=1}^m U_i}(D), \forall a \in (C - B)$ , 计算属性  $a$  在分布决策信息系统属性  $B$  相

对决策属性  $D$  的外重要性  $\text{Sig}_{\bigcup_{i=1}^m U_i}^{\text{outer}}(a, B, D)$ , 依次选取  $a_0 = \max(\text{Sig}_{\bigcup_{i=1}^m U_i}^{\text{outer}}(a, B, D))$  ( $\max$  为最大值) 加入到属性集  $B$  中, 计算分布决策信息系统属性集  $B$  的  $GP_{\bigcup_{i=1}^m U_i}^m(D)$ , 当  $GP_{\bigcup_{i=1}^m U_i}^m(D) = GP_c^{\bigcup_{i=1}^m U_i}(D)$ , 则停止计算;

Setp6  $\forall a \in B$ , 计算分布决策信息系统  $B - \{a\}$  属性的  $GP_{\bigcup_{i=1}^m U_i}^m(D)$ , 如果  $GP_{B - \{a\}}^{\bigcup_{i=1}^m U_i}(D) = GP_c^{\bigcup_{i=1}^m U_i}(D)$ , 则  $B \leftarrow B - \{a\}$ ;

Setp7  $\text{RED}_{\bigcup_{i=1}^m U_i}^m \leftarrow B$ , 的属性约简  $\text{RED}_{\bigcup_{i=1}^m U_i}^m$ .

## 3 分布决策信息系统增量属性约简算法

### 3.1 分布决策信息系统增量机制

**定义 9:**  $DS = (U, A, V, F) = \bigcup_{i=1}^m S_i$  是分布决策信息系统, 如果决策信息系统  $S' = (U', A, V, f)$  被增加到子分布决策信息系统  $S_j = (U_j, A_j, V_j, f_j)$  中,  $t = |U'|, x'_l \in |U_i| (l = 1, 2, \dots, t)$ , 决策信息系统  $S'$  在条件属性  $C$  下的等价关系矩阵为  $W_{t \times t}^{ij} = (w_{lk}^{ij})_{t \times t}, W_{t \times t}^{ij}$  的元素被定义为:

$$w_{lk}^{ij} = \begin{cases} 1, & V_C(x'_l) = V_C(x'_k) \\ 0, & V_C(x'_l) \neq V_C(x'_k) \end{cases} \quad (7)$$

**定义 10:** 给出一个分布信息系统  $DS = (U, A, V, F) = \bigcup_{i=1}^m S_i$ , 如果决策信息系统  $S' = (U', A, V, f)$  被增加到子分布决策信息系统  $S_j = (U_j, A_j, V_j, f_j)$  中,  $n_j = |U_j|, t = |U'|, x'_l \in |U_i| (l = 1, 2, \dots, t)$  及  $x_k \in |U_j| (l = 1, 2, \dots, n_j)$ , 决策信息系统  $S'$  和  $S_j$  在条件属性  $C$  下的等价关系为  $Q_{t \times n_j}^{ij} = (q_{lk}^{ij})_{t \times n_j}, Q_{t \times n_j}^{ij}$  的元素被定义为:

$$q_{lk}^{ij} = \begin{cases} 1, & V_C(x'_l) = V_{jC}(x_k) \\ 0, & V_C(x'_l) \neq V_{jC}(x_k) \end{cases} \quad (8)$$

**定义 11:** 给出一个分布信息系统  $DS = (U, A, V, F) = \bigcup_{i=1}^m S_i$ , 如果决策信息系统  $S' = (U', A, V, f)$  被增加到子分布决策信息系统  $S_j = (U_j, A_j, V_j, f_j)$  中,  $n_j = |U_j|, t = |U'|, x'_l \in |U_i| (l = 1, 2, \dots, t)$  及  $x_k \in |U_j| (l = 1, 2, \dots, n_j)$ ,  $S_j$  和  $S'$  在条件属性  $C$  下的等价关系为  $Z_{n_j \times t}^{ij} = (z_{kl}^{ij})_{n_j \times t}, Z_{n_j \times t}^{ij}$  的元素被定义为:

$$z_{kl}^{ij} = \begin{cases} 1, & V_{jC}(x_k) = V_C(x'_l) \\ 0, & V_{jC}(x_k) \neq V_C(x'_l) \end{cases} \quad (9)$$

**定理 1:**  $DS = (U, A, V, F) = \bigcup_{i=1}^m S_i$  是分布决策信息系统, 分布决策信息系统在条件属性  $C$  下的知识粒度是  $GP_c^{\bigcup_{i=1}^m U_i \cup U'}$ . 当分布决策信息系统增加

了对象集  $U'$ ,  $W_{i \times t}^{\bar{u}}$ ,  $Q_{i \times n_j}^{\bar{u}}$  及  $Z_{n_j \times t}^{\bar{u}}$  为在条件属性  $C$  下的等价关系矩阵, 则增加对象分布决策信息系统的知识粒度为:

$$GP_{\bigcup_{i=1}^m U_i \cup U'}^{\bar{u}} = GP_{\bigcup_{i=1}^m U_i}^{\bar{u}} + \frac{1}{|U|^2} \left( \sum_{j=1}^m \text{Sum}(Q_{i \times n_j}^{\bar{u}}) + \text{Sum}(W_{i \times t}^{\bar{u}}) + \sum_{j=1}^m \text{Sum}(Z_{n_j \times t}^{\bar{u}}) \right) \quad (10)$$

**定理 2:**  $DS = (U, A, V, F) = \bigcup_{i=1}^m S_i$  是分布决策信息系统, 分布决策信息系统在条件属性和决策属性  $C \cup D$  下的知识粒度是  $GP_{\bigcup_{i=1}^m U_i}^{\bar{u}}$ . 当分布决策信息系统增加了对象集  $U'$ ,  $W'_{i \times t}^{\bar{u}}$ ,  $Q'_{i \times n_j}^{\bar{u}}$  及  $Z'_{n_j \times t}^{\bar{u}}$  为在条件属性和决策属性  $C \cup D$  下的等价关系矩阵, 则增加对象分布决策信息系统的知识粒度为:

$$GP_{\bigcup_{i=1}^m U_i}^{\bar{u}} = GP_{\bigcup_{i=1}^m U_i}^{\bar{u}} + \frac{1}{|U|^2} \left( \sum_{j=1}^m \text{Sum}(Q'_{i \times n_j}^{\bar{u}}) + \text{Sum}(W'_{i \times t}^{\bar{u}}) + \sum_{j=1}^m \text{Sum}(Z'_{n_j \times t}^{\bar{u}}) \right) \quad (11)$$

**定理 3:**  $DS = (U, A, V, F) = \bigcup_{i=1}^m S_i$  是分布决策信息系统, 分布决策信息系统在条件属性和决策属性  $C$  下的相对知识粒度是  $GP_{\bigcup_{i=1}^m U_i}^{\bar{u}}(D)$ . 当分布决策信息系统增加了对象集  $U'$ ,  $W_{i \times t}^{\bar{u}}$ ,  $Q_{i \times n_j}^{\bar{u}}$  及  $Z_{n_j \times t}^{\bar{u}}$  为在条件属性  $C$  下的等价关系矩阵,  $W'_{i \times t}^{\bar{u}}$ ,  $Q'_{i \times n_j}^{\bar{u}}$  及  $Z'_{n_j \times t}^{\bar{u}}$  为在条件属性和决策属性  $C \cup D$  下的等价关系矩阵, 则增加对象分布决策信息系统的相对知识粒度为:

$$GP_{\bigcup_{i=1}^m U_i \cup U'}^{\bar{u}}(D) = GP_{\bigcup_{i=1}^m U_i}^{\bar{u}}(D) + \frac{1}{|U|^2} \left( \sum_{j=1}^m \text{Sum}(Q_{i \times n_j}^{\bar{u}}) + \text{Sum}(W_{i \times t}^{\bar{u}}) + \text{Sum}(W_{i \times t}^{\bar{u}}) + \sum_{j=1}^m \text{Sum}(Z_{n_j \times t}^{\bar{u}}) - \sum_{j=1}^m \text{Sum}(Q'_{i \times n_j}^{\bar{u}}) \text{Sum}(W'_{i \times t}^{\bar{u}}) - \sum_{j=1}^m \text{Sum}(Z'_{n_j \times t}^{\bar{u}}) \right) \quad (12)$$

### 3.2 分布决策信息系统增量约简算法

根据上述分布决策信息系统的增量机制, 设计了分布决策信息系统增量属性约简算法, 算法具体步骤如下:

输入: 分布决策信息系统  $DS = (U, A, V, F) = \bigcup_{i=1}^m S_i$  分布决策信息系统的约简为  $\text{RED}_{\bigcup_{i=1}^m U_i}^{\bar{u}}$ , 增

量对象集  $U'$ .

输出: 增加对象后  $DS$  的属性约简  $\text{RED}_{\bigcup_{i=1}^m U_i \cup U'}^{\bar{u}}$ .

Setp1  $B \leftarrow \text{RED}_{\bigcup_{i=1}^m U_i}^{\bar{u}}$ ;

Setp2 计算分布决策信息系统增加对象后的属性集  $B$  和  $C$  相对知识粒度  $GP_{\bigcup_{i=1}^m U_i \cup U'}^{\bar{u}}(D)$  和  $GP_{\bigcup_{i=1}^m U_i}^{\bar{u}}(D)$ ;

Setp3 如果  $GP_{\bigcup_{i=1}^m U_i \cup U'}^{\bar{u}}(D) = GP_{\bigcup_{i=1}^m U_i}^{\bar{u}}(D)$ , 则执行 Setp4, 否则执行 Setp6;

Setp4 当  $GP_{\bigcup_{i=1}^m U_i \cup U'}^{\bar{u}}(D) \neq GP_{\bigcup_{i=1}^m U_i}^{\bar{u}}(D)$ ,  $\forall a \in (C - B)$ , 计算属性  $a$  在分布决策信息系统属性  $B$  相对决策属性  $D$  的外重要性  $\text{Sig}_{\bigcup_{i=1}^m U_i \cup U'}^{\text{outer}}(a, B, D)$ , 依次选取  $a_0 = \max(\text{Sig}_{\bigcup_{i=1}^m U_i \cup U'}^{\text{outer}}(a, B, D))$  (max 为最大值) 加入到属性集  $B$  中, 计算分布决策信息系统属性集  $B$  的  $GP_{\bigcup_{i=1}^m U_i \cup U'}^{\bar{u}}(D)$ , 当  $GP_{\bigcup_{i=1}^m U_i \cup U'}^{\bar{u}}(D) = GP_{\bigcup_{i=1}^m U_i}^{\bar{u}}(D)$ , 则停止计算;

Setp5  $\forall a \in B$ , 计算分布决策信息系统  $B - \{a\}$  属性的  $GP_{\bigcup_{i=1}^m U_i \cup U'}^{\bar{u}}(D)$ , 如果  $GP_{\bigcup_{i=1}^m U_i \cup U'}^{\bar{u}}(D) = GP_{\bigcup_{i=1}^m U_i}^{\bar{u}}(D)$ , 则  $B \leftarrow B - \{a\}$ ;

Setp6  $\text{RED}_{\bigcup_{i=1}^m U_i \cup U'}^{\bar{u}} \leftarrow B$ , 增加对象后  $DS$  的属性约简  $\text{RED}_{\bigcup_{i=1}^m U_i \cup U'}^{\bar{u}}$ .

## 4 仿真实验分析

### 4.1 数据集及软硬件环境

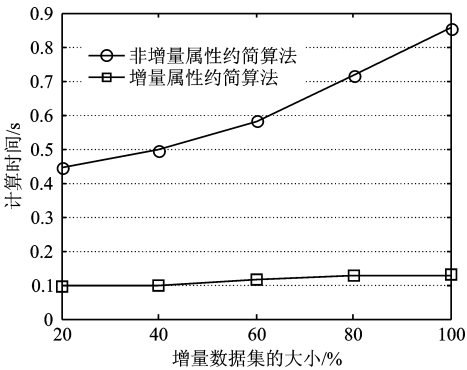
为了验证了分布决策信息系统增量约简算法能够有效处理动态分布数据属性约简的问题, 我们用分布决策信息系统静态约简算法和分布决策信息系统增量约简算法做了一些对比, 实验分析所用的数据集从 UCI 网站下载且其具体描述如表 1 所示. 另外, 实验分析操作系统为旗舰版 Windows 7.0. 实验分析编写代码的软件环境为 Visual Studio 2015. 实验分析的硬件环境为: 处理器: Intel 酷睿 i5 6400, 内存: 4.0 GB.

表 1 UCI 数据集

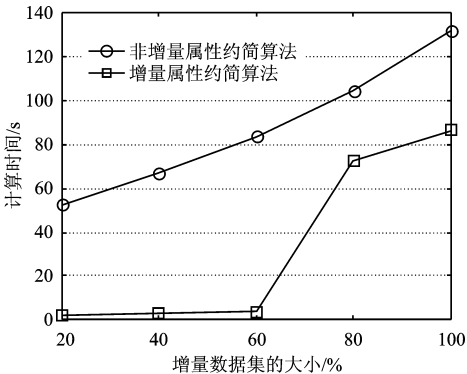
| 序号 | 数据集名称    | 对象数目  | 条件属性 | 决策属性 |
|----|----------|-------|------|------|
| 1  | Cancer   | 683   | 9    | 2    |
| 2  | Mushroom | 5 644 | 22   | 2    |
| 3  | Kr-vs-kp | 3 196 | 36   | 2    |

4.2 属性约简结果

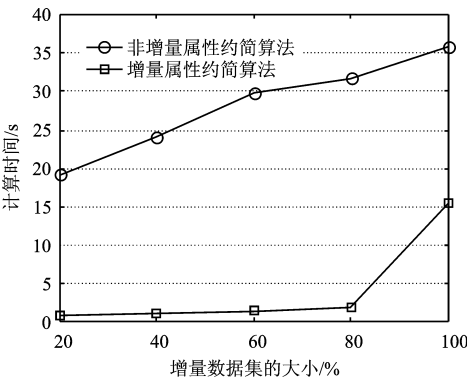
在对比实验分析过程中,表 1 中所有数据集按照对象集分成两个相等的子数据集,把一个子数据集均匀分成 3 部分,作为分布决策信息系统,另外,把另一个子数据集按照对象集的 20%,40%,60,80%,100%分成 5 部分,作为增量数据集,把这些增量数据集依次添加到分布决策信息系统中的任意一个子决策信息系统中,分别用分布决策信息系统增量属性约简算法和分布决策信息系统非增量属性约简算法运行增加对象前后的数据集.对比实验分析结果如图 1.



(a)Cancer



(b)Mushroom



(c)Kr-vs-kp

图 1 比较非增量算法和增量算法计算时间

4.3 属性约简分类精确度分析

把表 1 中所有数据集按照对象集分成两个相等的子数据集,一个子数据集均匀分成 3 部分,作为分布决策信息系统,另一个子数据集作为增量数据集,把增量数据集添加到分布决策信息系统中的任意一个子决策信息系统中,分别用静态约简算法和分布决策信息系统增量约简算法计算决策分布系统的约简,然后用贝叶斯分类方法和 1 折交叉验证方法计算分布决策信息系统静态约简算法和分布决策信息系统增量约简算法所得约简的分类精确度,分类结果如表 2 所示.

表 2 静态、动态增量属性约简算法分类精度结果比较

| 数据集      | 静态属性约简算法 | 增量属性约简算法 |
|----------|----------|----------|
| Cancer   | 0.747 2  | 0.747 2  |
| Mushroom | 0.989 2  | 0.997 1  |
| Kr-vs-kp | 0.901 4  | 0.912 1  |

从表 2 容易看出分布决策信息系统静态约简算法和分布决策信息系统增量约简算法的分类精确度是相近的,说明了本文提出的增量约简算法的分类能力与非增量算法的分类能力相近的情况下,能节省大量的计算时间.

5 结束语

本节给出了基于矩阵方法的分布决策信息系统属性约简算法主要特点和将来研究的方向如下:

(1)提出了分布决策信息系统知识粒度的增量机制原理,并给出严格的数据表达式,为分布决策信息系统增量属性约简算法设计鉴定了基础.

(2)基于上述增量更新机制的基础上,设计了分布决策信息系统增量属性约简算法,该算法能够在原有结果的基础上快速更新决策信息系统变化的约简.

(3)实验分析结果验证了分布决策信息系统属性约简算法在运行时间上比非增量算法具有较明显的优势.

(4)由于分布决策信息系统属性也随着时间动态发生变化,未来将研究属性动态变化后分布决策信息系统增量属性约简方法.

参考文献:

[1] PAWLAK Z, SKOWRON A. Rough sets: some extensions[J]. Information Sciences, 2007, 177 (1): 28-40.

[2] ZHU L, PANG S N, BAO T, et al. Incremental and

- decremental max-flow for online semi-supervised learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2016, 28(8): 2115-2127.
- [3] 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.
- [4] CHEN H M, LI T R, LUO C, et al. A decision-theoretic rough set approach for dynamic data mining[J]. IEEE Transactions on Fuzzy Systems, 2015, 23(6): 1958-1970.
- [5] LIANG J Y, WANG F, DANG C Y, et al. A group incremental approach to feature selection applying rough set technique[J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(2): 294-308.
- [6] XU Y T, WANG L S, ZHANG R Y. A dynamic attribute reduction algorithm based on 0-1 integer programming[J]. Knowledge-Based Systems, 2011, 24(8): 1341-1347.
- [7] JIANG F, SUI Y F, CAO C G. An incremental decision tree algorithm based on rough sets and its application in intrusion detection[J]. Artificial Intelligence Review, 2013, 40(4): 517-530.
- [8] JING Y G, LI T R, LUO C, et al. An incremental approach for attribute reduction based on knowledge granularity [J]. Knowledge-Based Systems, 2016(104): 24-38.
- [9] WANG F, LIANG J Y, DANG C Y. Attribute reduction: A dimension incremental strategy[J]. Knowledge-Based Systems, 2013(39): 95-108.
- [10] ZENG A P, LI T R, LIU D, et al. A fuzzy rough set approach for incremental feature selection on hybrid information systems [J]. Fuzzy Sets and Systems, 2015(258): 39-60.
- [11] CHEN D G, YANG Y Y, DONG Z. An incremental algorithm for attribute reduction with variable precision rough sets[J]. Applied Soft Computing, 2016(45): 129-149.
- [12] SHU W H, SHEN H. Updating attribute reduct in incomplete decision systems with the variation of attribute set[J]. International Journal of Approximate Reasoning, 2014(55): 867-884.
- [13] WANG F, LIANG J Y, QIAN Y H. Attribute reduction for dynamic data sets[J]. Applied Soft Computing, 2012(18): 1-18.
- [14] CHEN H M, LI T R, LUO C, et al. A Rough Set-based Method for Updating Decision Rules on Attribute Values' Coarsening and Refining[J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(12): 2886-2899.
- [15] JING Y G, LI T R, HUANG J F, et al. A group incremental reduction approach with varying data values [J]. International Journal of Intelligent Systems, 2017, 32(9): 500-525.
- [16] ZHANG J B, WONG J S, PAN Y, et al. A parallel matrix-based method for computing approximations in incomplete information systems[J]. IEEE Transactions on Knowledge and Data Engineering, 2015, 27(2): 326-339.
- [17] JING Y G, LI T R, FUJITA H, et al. An incremental attribute reduction approach based on knowledge granularity with a multi-granulation[J]. Information Sciences, 2017(411): 23-38.

#### 作者简介:

**徐岩柏** 女, (1982-), 硕士, 讲师. 研究方向为粗糙集理论与粒计算.

**景运革(通讯作者)** 男, (1970-), 教授. 研究方向为粗糙集理论与粒计算. E-mail: jyg701022@163.com.