An incremental attribute reduction algorithm based on the matrix dimension reduction method

Yunyun Ma*, **Ting Quan**
Xi'an Innovation College of Yan'an University, Xi'an 710000, Shaanxi, China

Abstract: In view of the dynamic change of decision system data in practical problems, the concept of attribute importance function is proposed, which effectively avoids the deficiency of considering only each attribute based on importance reduction algorithm. In the calculation process of matrix reduction algorithm, the sample combination explosion between big data causes huge time consumption. The matrix dimension reduction processing method is proposed, which greatly improves the efficiency and accuracy of calculation. Finally, an incremental reduction algorithm based on matrix dimensionality reduction is presented, and the correctness of the algorithm is tested. Meanwhile, the results of the four algorithms are compared to verify the effectiveness of the algorithm.

Keywords: Attribute reduction; Matrix dimension reduction; Attribute importance function; Incremental learning

一种基于矩阵降维方法的增量式属性约简算法

马云云* 权婷
延安大学西安创新学院 中国·陕西西安 710000

摘要：针对实际问题中决策系统数据动态变化的情况，提出属性重要度函数的概念，有效避免了基于重要度约简算法仅考虑每一个属性；基于矩阵约简算法在计算过程中，大数据间样本组合爆炸造成时间消耗巨大的问题，提出矩阵降维处理的方法，大大提高了计算的效率和准确性；最后，通过改进算法给出了基于矩阵降维的增量式约简算法，对算法进行了正确性检验，同时通过四种算法的结果比较验证了算法的有效性。

关键词：属性约简；矩阵降维；属性重要度函数；增量式学习

1 引言

粗糙集理论[1]是由波兰学者Z. Pawlak在1982年提出的，是一种处理不精确、不相容和不完整数据的数学工具，目前已经广泛应用于知识发现、机器学习、模式识别、数据挖掘、决策分析等领域。属性约简在粗糙集理论中占据核心地位，而属性约简就是指在保持系统分类能力不变的 情况下，删除系统中无关的或不重要的属性。由于属性组合爆炸的原因，求信息系统（或决策表）的约简或最优约简已经被证明是一个NP问题。目前，在粗糙集属性约简方面已有许多学者做出了成果，比如基于相对粒度的约简算法、基于属性重要度的约简算法、基于矩阵的约简算法[2–3]等常用的经典算法；但由于大多数的属性约简算法都是以静态信息系统（或决策表）为基础设计的，针对动态变化的信息系统（或决策表）的约简算法研究较少，而我们的实际生活中，几乎所有的数据库都是不断更新变化的，因此，在面对海量的动态变化的数据库时，原有就约简算法在此时就体现出了不足，比如算法时间、空间复杂度高，算法稳定性较低等，所以，在处理动态变化的数据库时需要提出更高效的约简算法。

增量变化的数据库有分为几种情况：样本动态变化、属性动态增加、样本与属性同时变化[4–6]。目前，针对前两种情况已经有许多学者进行了约简算法的研究，但是，针对第三种情况的算法研究相对较少，还需要更进一步的对算法进行改进。本文中主要针对对象与属性同时变化的数据库，提出了一种基于矩阵降维方法的增量式属性约简算法，并对算法的约简结果和正确性进行了验证。

2 粗糙集基础知识

定义1[7]信息系统(S)通常在形式上描述为一个四元组 $IS=(U,A,V,f)$，其中 $U=\{x_1,x_2,\cdots,x_n\}$表示所有样本（对象）集合，$A=\{a_1,a_2,\cdots,a_n\}$表示属性（特征）集合，将属性集合A划分为两部分，条件属性C，决策属性D，即$A=C\cup D$，且$C\cap D=\emptyset$，$D\neq\emptyset$，此

项目：1.基于矩阵降维方法的增量式属性约简算法（2024XJKY20）；2.“因需施教，分类教学”的实效性研究—以数学与应用数学为例（SGH23Y2953）

作者简介：马云云（1995.10—），女，汉，陕西，硕士研究生，助教，研究方向：粗糙集，教育统计。

时，信息系统IS就称为一个决策系统。

定义2[7]给定一个知识库K = (U, S)和知识库中的一个等价关系族P ⊆ S，∀R ∈ P，若 $IND(P) = IND(P-\{R\})$成立，则称知识R为P中不必要的，否则称R为P中不必要的。

如果对于每一个R ∈ P，R都为P中必要的，则称P为独立的，否则称P是依赖的或者不独立的。

定义3[7]给定一个知识库K = (U, S)和知识库上的一族等价关系P ⊆ S，对任意的G ⊆ P，若G满足以下两条：

（1）G是独立的，

（2）IND(G) = IND(P)。

则称G是P的一个约简，记为G ∈ RED(P)，其中，RED(P)表示P的全体约简组成的集合。

定义4[7]给定一个知识库K = (U, S)和知识库上的一族等价关系P ⊆ S，对任意的R ∈ P，若R满足 $IND(P-\{R\}) \neq IND(P)$，则称R为P中必要的，P中所有必要的知识组成的集合称为P的核，记为CORE(P)。

定理1 $CORE(P) = \cap RED(P)$。

定义5[7]（属性重要度）设给定信息系统的表示 $S=(U, C\cup D, V, f)$，其中α，β为条件属性，B为条件属性集，D为决策属性，∀B ⊆ C，∀β ∈ C以及∀α ∈ C − B，定义

$$SIG(\alpha,B;D)=\delta_{IND(B\cup\{\alpha\})}(D)-\delta_{IND(B)}(D)=\frac{|pos_{B\cup\{\alpha\}}(D)|-|pos_B(D)|}{|U|}$$

为属性α对B相对于D的重要度。

3 二进制矩阵与属性重要度函数

3.1 二进制矩阵

在属性约简的算法过程中，数据系统都包含较大的数据量，利用传统方法计算时，求取等价类的过程会造成很大的时间消耗，本节提出二进制矩阵的概念，可以提高计算等价类过程中的时间效率。

定义6决策系统 $DT=(U,A,V,f)$，其中 $U=\{x_1,x_2,\cdots,x_n\}$ 表示所有样本（对象）集合，$C=\{c_1,c_2,\cdots,c_m\}$表示条件属性，$D=\{d\}$表示决策属性，i，j表示决策表中第i和第j个样本x_i与x_j，则决策表DT对应的二进制可分辨矩阵BM中元素为：

$$m(i,j)=\begin{cases}1, & c_p(x_i)\neq c_p(x_j)\wedge D(x_i)\neq D(x_j) \quad i,j=1,\cdots,m, p=1,\cdots,m\\ 0, & else\end{cases}$$

由 $m(i,j)$ 组成的矩阵即为决策表DT对应的二进制可分辨矩阵BM。

定理2 二进制可分辨矩阵BM中，对于$\forall\alpha_i\in C$，都满足$m_{\alpha_i}(i,j)=1$，则$\alpha_i\subset CORE$，若$\forall\alpha_j\in C-\alpha_i, i\neq j$，则$\alpha_j$为必要属性，即$\alpha_j\subset NEC$，若$\forall\alpha_i\in C$，都满足$m_{\alpha_i}(i,j)=0$，则$\alpha_i$为不必要属性，即$\alpha_i\subset DIS$。

证明：假设属性$\alpha(\alpha\in C)$，使得每一个$m_\alpha(i,j)=1$，则属性α能够区分所有对象，分类能力为$IND(\alpha)\neq IND(C)$，且$SIG(\alpha,C;D)=SIG(\alpha,C;D)=1$，该属性为核属性；若存在属性$\alpha(\alpha\in C)$，使得$m_\alpha(i,j)=1$，且存在某个$m_\alpha(p,q)=0$, $p\neq i, q\neq j$，$IND(\alpha)\neq IND(C)$，且$0<SIG(\alpha,C;D)\neq SIG(\alpha,C;D)<1$，则该属性为必要属性；若存在属性$\alpha(\alpha\in C)$，使得$m_\alpha(i,j)=1$，则$SIG(\alpha,C;D)=0$，该属性为不必要属性，证毕。

3.2 属性重要度函数

在属性约简过程中，往往会出现多个矩阵重要性相同的情况，传统算法在解决这种状况时，一般会从中任意选择一个属性将其加入到约简结果中，但这种处理方法存在一定的不足，选择的随机性使得约简结果不唯一，且在多个约简结果中也并未指出最优约简，为了得到最优的约简结果，提出了一个新的定义属性重要性函数，达到进一步有效提取分类更优更好属性的效果，最终得到最优约简。

定义7给定决策系统$DT=(U,C\cup D,V,f)$，其中C为条件属性集，D为决策属性集，对条件属性∀α ∈ C，在二进制可分辨矩阵BM中，当有多个重要性相同的属性时，利用新定义的属性重要度函数来筛选，定义属性α的属性重要度函数$Fun(\alpha,C;D)$为：

$$Fun(\alpha,C;D)=\frac{f(\alpha_i)}{f(C_j^{'})}$$

其中$f(\alpha_i)$的定义如下：

$$f(\alpha_i)=\sum_{s=1}^{N}\left((m(i,j)k)\circ\left(\sum_{t=1}^{C^{'}}m(i,j)t\right)^{-1}\right),\ \alpha_k\in C^{'}$$

其中，N为二进制可分辨矩阵BM中元素项的个数$\left(0\le N\le|U^{'}|(|U^{'}|-1/2)\right)$，$|C^{'}|$为BM中条件属性的个数。

4 增量式属性约简算法

4.1 样本与属性同时变化的增量式属性约简算法

实际决策系统中，由样本两两比较得到的二进制矩阵显然是一个很多维的矩阵，利用该矩阵计算约简结果时，由于多维度造成的时间浪费和各种人为因素导致计算准确性降低，因而，需要对高维矩阵进行降维处理，使得计算的效率和准确率得到提升。提出矩阵降维方法，并证明了方法的正确性。

定义8决策系统$DT=(U,C\cup D,V,f)$，其中$C=\{\alpha_1,\alpha_2,\cdots,\alpha_m\}$，$B_k(k\le n)$表示所有划分约

简集合，若$\alpha_i \in \mathrm{DIS}_{U_i}$，则$\alpha_i$属于不必要属性集$\mathrm{DIS}_{U_i} = \bigcup_{k\le n} B_k$；若$\alpha_i \in \mathrm{CORE}_{U_i}$，则$\alpha_i$属于核属性集$\mathrm{CORE}_{U_i} = \bigcap_{k\le n} B_k$；若$\alpha_i \in \mathrm{CORE}_{U_i}$，则$\alpha_i$属于必要属性集$\mathrm{NEC}_{U_i} = \bigcup_{k\le n} B_k - \mathrm{CORE}_{U_i}$。

算法 基于矩阵降维的增量式约简算法

输入：决策系统$\mathrm{DT} = (U, C\cup D, V, f)$.

输出：属性约简$\mathrm{RED}_C(D)$.

步骤1：删除样本集U中重复或不相容的样本后，得到新的决策系统的样本集$U^{'}$，根据定义8与性质1对决策系统新的样本集进行分组，分组规则如下：

设总样本个数为γ，每组样本数为μ，若γ可以整除μ，将$U^{'}$分为$[\gamma|\mu] = n$个子系统，其中$U^{'} = U^{'}_1 \cup U^{'}_2 \cup \cdots \cup U^{'}_n$，若$\gamma$不能整除μ，将$U^{'}$分为$[\gamma|\mu]+1 = n+1$个子系统，其中$U^{'} = U^{'}_1 \cup U^{'}_2 \cup \cdots \cup U^{'}_n \cup U^{'}_{n+1}$；

步骤2：根据定义6求出$U^{'}$每一个子系统$U^{'}_l$对应的二进制可分辨矩阵BM_l，其中BM_l中元素$m(i,j)$表示为：

$$m(i,j) = \begin{cases} 1, & c_p(x_i) \ne c_p(x_j) \wedge D(x_i) \ne D(x_j) \quad i,j = 1,\cdots,m, p = 1,\cdots,m \\ 0, & else \end{cases}$$

步骤3：根据定理2对步骤1中得到的二进制可分辨矩阵BM_l进行简化删除处理，求出核属性集$\mathrm{CORE}_{U^{'}_i} = \bigcap_{k\le n} B_k$，必要属性集$\mathrm{NEC}_{U^{'}_i} = \bigcup_{k\le n} B_k - \mathrm{CORE}_{U^{'}_i}$，不必要属性集$\mathrm{DIS}_{U^{'}_i} = \bigcup_{k\le n} B_k$，$i = 1,2,\cdots,n,n+1$，在处理二进制可分辨矩阵$\mathrm{BM}_l$时，当有多个重要性相同的属性时，根据定义8提出的属性重要度函数$\mathrm{Fun}(\alpha, C; D)$，计算出属性重要度值相同的属性的$\mathrm{Fun}(\alpha, C; D)$值，选出其中属性重要度函数值最大的$\mathrm{Fun}(\alpha_{max}, C; D)$对应的属性$\alpha_{max}$，其中：

$$\mathrm{Fun}(\alpha, C; D) = \frac{f(\alpha_i)}{f(C_j)};$$

步骤4：设增加属性集S的属性个数为$\theta(\theta \ge 0)$，若θ=0，则转向步骤5，否则转向步骤2；

步骤5：设增加样本集Y的样本数为$\varphi(\varphi \ge 0)$，若φ=0，则转向步骤7，否则转向步骤6；

步骤6：判断步骤2中子系统$U^{'}_{n+1}$中含有的样本数δ与每组样本数μ的大小，

（1）若$\delta = \mu$，则直接对新增样本集Δ中的样本根据定义8进行分组，对新的子系统$U^{'}_{n+1}$，$U^{'}_{n+2}$，…，$U^{'}_{n+p}$进行步骤3的计算出相应的属性集；

（2）若$\delta < \mu$，则将新增样本集Δ中$\mu-\delta$差值个数的样本加入到子系统$U^{'}_{n+1}$中，然后把剩余样本$\omega = \varphi - (\mu - \delta)$按照定义8进行分组处理，对新的子系统$U^{'}_{n+1}$，$U^{'}_{n+2}$，…，$U^{'}_{n+p}$进行步骤3的计算出相应的属性集；

步骤7：令$\mathrm{CORE}_{U^{'}} = \bigcup_{i=1}^{n+1} C^{'}_{U^{'}_i}$，$\mathrm{NEC}_{U^{'}} = \bigcup_{i=1}^{n+1} \mathrm{NEC}^{'}_{U^{'}_i} - \mathrm{CORE}_{U^{'}_i}$，$\mathrm{DIS}_{U^{'}} = C^{'} - \mathrm{CORE}_{U^{'}} - \mathrm{NEC}_{U^{'}}$；

步骤8：若$\mathrm{CORE}_C(D) = \mathrm{CORE}_{U^{'}}$为的约简集，则$\mathrm{RED}_C(D) = \mathrm{CORE}$，转向步骤9，否则$\mathrm{RED}_C(D) = \mathrm{CORE}_{U^{'}} \cup \mathrm{NEC}_{U^{'}}$，转向步骤9；

步骤9：输出$\mathrm{RED}_C(D)$，算法结束。

4.2 算法复杂度分析

算法处理增量属性约简，若增加属性集S的属性个数为$\theta(\theta = 0)$，增加样本集Y的样本数为$\varphi(\varphi = 0)$时，时间复杂对为$O(\sum_{i=1}^{n} |U^{'}_i||C^{'}|)$；当$\theta(\theta > 0)$，$\varphi(\varphi > 0)$时，将$U^{'}$论域分为$[\gamma|\mu] = n$个子系统，步骤3–8求解约简结果的时间复杂度为$O(\sum_{i=1}^{n} |U^{'}_i||C^{'}|) + O(\sum_{i=1}^{n} |U^{''}_i - U^{'}_i||C^{''} - C^{'}|)$；$U^{'}$分为$[\gamma|\mu]+1 = n+1$个子系统时，时间复杂度为$O(\sum_{i=1}^{n+1} |U^{'}_i||C^{'}|) + O(\sum_{i=1}^{n+1} |U^{''}_i - U^{'}_i||C^{''} - C^{'}|)$，求解中如果出现多个(N)属性的重要性相同时，需要利用步骤3中属性重要度函数进行最优属性的筛选，此时时间复杂度为$O(\sum_{i=1}^{n} |U^{'}_i||C^{'}|) + O(\sum_{i=1}^{n} |U^{''}_i - U^{'}_i||C^{''} - C^{'}|) + O(|N|)$或者$O(\sum_{i=1}^{n+1} |U^{'}_i||C^{'}|) + O(\sum_{i=1}^{n+1} |U^{''}_i - U^{'}_i||C^{''} - C^{'}|) + O(|N|)$。

4.3 非增量实例分析

例表4.1为删除掉重复与不相容数据后的决策表，其中论域$U = \{x_1, x_2, x_3, x_4, x_5, x_6\}$，条件属性集$C = \{\alpha_1, \alpha_2, \alpha_3\}$，决策属性$D = \{d\}$，利用本文算法求解约简结果如下：

表4.1 决策表

$U^{'}$	α_1	α_2	α_3	d
x_1	1	1	0	0
x_2	1	2	0	1
x_3	0	0	2	1
x_4	0	1	0	0
x_5	0	0	1	0
x_6	0	1	2	1

根据算法步骤1利用定义8与性质1对样本集进行分组，实现对二进制矩阵的降维。由表可知总样本个数为$\gamma = 6$，令每组样本数为$\mu = 3$，则将$U^{'}$分为$[\gamma|\mu] = 2$个子系统$U^{'}_1$与$U^{'}_2$，然后根据定义6生成子系统的二级制可分辨矩阵为BM_1，BM_2。

表4.2 子系统$U^{'}_1$的二进制可分辨矩阵BM_1

$m(i,j)$	α_1	α_2	α_3
(1,2)	0	1	0
(1,3)	1	1	1
(2,3)	1	1	1

根据算法步骤2与步骤3由定义7属性重要度函数求得每个属性的函数值为：$\mathrm{Fun}(\alpha_1, C^{'}; D) = \frac{2}{3}$，$\mathrm{Fun}(\alpha_2, C^{'}; D) = 1$，$\mathrm{Fun}(\alpha_3, C^{'}; D) = \frac{2}{3}$。根据表4.2的$\mathrm{BM}_1$与定理2，可以得到

的属性集为：核属性集$\mathrm{CORE}_{U'_1}=\{\alpha_2\}$，必要属性集$\mathrm{NEC}_{U'_1}=\{\alpha_1\}$，或$\mathrm{NEC}_{U'_1}=\{\alpha_3\}$，不必要属性集$\mathrm{DIS}_{U'_1}=\emptyset$。

表4.3 子系统U'_2的二进制可分辨矩阵BM_2

$\mathrm{m}(i,j)$	α_1	α_2	α_3
(4,5)	0	1	1
(4,6)	0	1	1
(5,6)	0	0	1

由定义7属性重要度函数求得每个属性的函数值为：$\mathrm{Fun}(\alpha_1,C';D)=0$，$\mathrm{Fun}(\alpha_2,C';D)=\frac{2}{3}$，$\mathrm{Fun}(\alpha_3,C';D)=1$。根据表4.3的$\mathrm{BM}_2$与定理2，可以得到的属性集为：核属性集$\mathrm{CORE}_{U'_2}=\{\alpha_3\}$，必要属性集$\mathrm{NEC}_{U'_2}=\{\alpha_2\}$，不必要属性集$\mathrm{DIS}_{U'_2}=\{\alpha_1\}$。

根据性质1对以上计算结果取以上计算结果的合取范式即为原决策系统的结果，核属性集$\mathrm{CORE}_{U'}=\mathrm{CORE}_{U'_1}\vee\mathrm{CORE}_{U'_2}=\{\alpha_2,\alpha_3\}$，必要属性集$\mathrm{NEC}_{U'}=\mathrm{NEC}_{U'_1}\vee\mathrm{NEC}_{U'_2}=\{\alpha_2,\alpha_3\}$，不必要属性集$\mathrm{DIS}_{U'}=\mathrm{DIS}_{U'_1}\vee\mathrm{DIS}_{U'_2}=\{\alpha_1\}$，所以，约简结果为$\mathrm{RED}_C(D)=\mathrm{CORE}_{U'}=\{\alpha_2,\alpha_3\}$。

4.4 增量实例分析

表4.4 增加样本集

Y	α_1	α_2	α_3	d
x_7	1	2	1	1
x_8	0	1	0	1
x_9	1	1	2	0

（1）在增量算法中，对4.1节实例进行增量变化处理，当样本增加时（表4.4），已知，样本数据$[\gamma|\mu]$是整除的情况，那么可直接进行算法中步骤1对新增样本进行分组。由于每组样本数为$\mu=3$，则增加的原本组成子系统U'_3。

表4.4 增加属性集

S	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
α_4	0	1	2	2	1	2	0	2	1
α_5	0	2	0	0	1	1	1	0	1
α_6	2	0	2	0	0	0	1	2	2

（2）属性与样本同时增加时，删除掉重复与不相容数据后的决策表得到论域U''，表4.2中新增元素为，$\mathrm{m}(1,2)=\{(x_{1,\alpha_4},x_{2,\alpha_4}),(x_{1,\alpha_5},x_{2,\alpha_5}),(x_{1,\alpha_6},x_{2,\alpha_6})\}$，$\mathrm{m}(1,3)=\{(x_{1,\alpha_4},x_{3,\alpha_4}),(x_{1,\alpha_5},x_{3,\alpha_5}),(x_{1,\alpha_6},x_{3,\alpha_6})\}$，$\mathrm{m}(2,3)=\{(x_{2,\alpha_4},x_{3,\alpha_4}),(x_{2,\alpha_5},x_{3,\alpha_5}),(x_{2,\alpha_6},x_{3,\alpha_6})\}$。同理，对子系统$U'_2$与$U'_3$进行增量处理，得出子系统$U''_1$、$U''_2$与$U''_3$、相对应的二进制可分辨矩阵$\mathrm{BM}_{12}$、$\mathrm{BM}_{22}$和$\mathrm{BM}_{32}$。

通过算法2对增加属性与样本后的论域U''进行降维与约简求解，计算得到核属性集$\mathrm{CORE}_{U''}=\mathrm{CORE}_{U''_1}\vee\mathrm{CORE}_{U''_2}\vee\mathrm{CORE}_{U''_3}=\{\alpha_2,\alpha_3,\alpha_4\}$，必要属性集$\mathrm{NEC}_{U''}=\mathrm{NEC}_{U''_1}\vee\mathrm{NEC}_{U''_2}\vee\mathrm{NEC}_{U''_3}=\{\alpha_5\}$，不必要属性集$\mathrm{DIS}_{U''}=\mathrm{DIS}_{U''_1}\vee\mathrm{DIS}_{U''_2}\vee\mathrm{DIS}_{U''_3}=\{\alpha_1,\alpha_6\}$，所以，约简结果为$\mathrm{RED}_C(D)=\mathrm{CORE}_{U''}\cup\mathrm{NEC}_{U''}=\{\alpha_2,\alpha_3,\alpha_4,\alpha_5\}$。

为进一步验证本文算法的正确性与有效性，将本文基于矩阵降增量式约简算法与文献（DRICA）算法进行了约简结果与时间复杂度的比较，结果如下表4.5：

表4.5 约简结果比较

算法	本文算法	文献（DRICA）算法
算法类型	基于矩阵降维的增量式约简算法	一种差别矩阵的增量式算法
约简结果 $\mathrm{RED}_C(D)$	$\{\alpha_2,\alpha_3,\alpha_4,\alpha_5\}$	$\{\alpha_2,\alpha_3,\alpha_4,\alpha_5\}$
时间复杂度	$O\left(\sum_{i=1}^{n}\lvert U'_i\rvert\lvert C'\rvert\right)+O\left(\sum_{i=1}^{n}\lvert U''_i-U'_i\rvert\lvert C''-C'\rvert\right)+O(\lvert N\rvert)$	$O\left(\sum_{i=1}^{n}\lvert U'_i\rvert\lvert C'\rvert\right)+O\left(\lvert U''\rvert^2\lvert C''\rvert^2\right)$

由上表可知，约简算法得到的约简结果是一致的，且都得到了最优的唯一约简结果，验证了算法的可行性。通过时间复杂度计算，已知U'为初始论域，C'为初始属性集，U''增量后的论域，C''增量后的属性集，因为对矩阵降维处理后矩阵的元素会成倍减少，计算量也大大减小，显然，$O(\sum_{i=1}^{n}|U'_i||C'|)$时间相等，但$U''\gg\sum_{i=1}^{n}U''_i-U'_i$，$C'>C''-C'$，可以得到，$O(\sum_{i=1}^{n}|U''_i-U'_i||C''-C'|)<O\left(|U''|^2|C''|^2\right)$，所以本文算法相比DRICA算法的计算效率较高，且DRICA算法在如何调节粗糙集中的参数，使得上近似和下近似更好地满足要求，方面做的不够完善，另外须用簇中所有点来代表一个，对于大数据集在内存有限的情况下对动态的数据集进行增量式聚类难以适用。

5 结束语

提出了基于矩阵降维的增量式约简算法，本文算法能够动态的处理数据动态更新的问题，在一些方面克服了传统静态属性约简算法存在的问题。改进了筛选重要属性存在的随机性，导致不能得到最优约简的情况，通过矩阵降维的方法将高维度矩阵求解的过程化繁为简，在数据更新过程中，可以利用原由矩阵的计算结果，有效避免了原有矩阵元素的二次计算，提高里计算的准确性，也降低了过程中不必要时间的消耗，提高里算法的计算效率。但在处理实际问题时，数据可能存在缺失或遗漏的情况，该算法对这类数据是否适用，如何改进都将成为下一步继续解决和处理的研究内容。

参考文献

[1] Z. Pawlak. Rough Sets-Theoretical Aspects of Reasoning about Data [M].Netherlands: Kluwer Acadamic

Publisher , 1991.

[2]闫鑫,景运革.矩阵增量属性约简算法[J].小型微型计算机系统,2018,(06):1245-1249.

[3]孙士保,秦克云,王育辉.基于区分矩阵和区分函数进行属性约简的数据分类[J].河南科技大学学报(自然科学版),2005,(04): 37-40.

[4]申雪峰,谢珺,刘海冯,等.一种改进的基于相对正域的增量式属性约简算法[J].广西师范大学学报(自然科学版),2013,31(03):45-50.

[5]李成,赵海琳.基于粗糙集的不完备信息系统增量式属性约简[J].测控技术,2018,37(11):50-54.

[6]景运革,景罗希,王宝丽等.属性值和属性变化的增量属性约简算法[J].山东大学学报(理学院),2020,(01):62-68.

[7]苗夺谦,李道国.粗糙集理论、算法与应用[M].北京:清华大学出版社,2008,4.