2008年 2月 沈阳理工大学学报 Vol. 27 No. 1
第 27卷 第 1期 TRANSACTIONS OF SHENYANG LIGONG UNIVERSITY Feb. 2008

文章编号：1003－1251(2008)01－0018－03

基于差别对象对的属性约简算法

景永霞，王治和，苟和平

(西北师范大学 数学与信息科学学院，甘肃 兰州 730070)

摘　要：属性约简是粗糙集理论的核心问题之一，当问题的规模比较大时，基于差别矩阵的属性约简算法存放差别矩阵的空间过大，相应地，其时间复杂度也比较高．针对这一问题，提出了基于差别对象对的改进属性约简算法，由于该算法不再需要存储差别矩阵，因而降低了存储量和计算量，从而提高了算法的效率．

关 键 词：粗糙集；属性约简；差别对象对

中图分类号：TP18　　**文献标识码**：A

A Attribute Reduction Algorithm Based on Discernibility Object Pair

JING Yong-xia, WANG Zhi-he, Gou He-ping

(College of Mathematics and Information Science, Northwest Normal University, Lanzhou 730070, China)

Abstract: Attributes reduction is one of the core issues of rough set theory. When the scale of the problem is larger, the attribute reduction algorithm based on discernibility matrix needs a large storage space, and its complexity in time is high accordingly. An improved algorithm of attribution reduction based on the discernibility object pair is proposed, it can cut down the computing and storing capacity greatly, thus improving the efficiency of the algorithm.

Key words: rough set; attribute reduction; discernibility object pair

Pawlak[1]于 1982年首次提出了粗糙集(RS)，它是一种研究不精确、不确定性知识的数学工具．粗糙集的主要优势在于它无需提供问题所需处理的数据集合之外的任何先验知识，粗糙集理论已广泛应用于数据挖掘、决策支持系统、知识发现等领域．

粗糙集知识发现的一个很重要的方面涉及到搜索特定的条件属性的子集[2]，这些子集能够提供与条件属性集合相同的分类，即保持知识库的分类能力不变．众所周知，信息系统或者决策系统通常会有一些不相关或冗余的信息存在，而这些信息会干扰我们提取有用信息、做出明智决策．为了能从包含冗余或不相关信息的决策系统中获取简洁的决策规则，属性约简是必不可少的．

本文将提出一种基于差别对象对的改进算法来提高属性约简的效率，基于差别矩阵的属性约简算法先要求出差别矩阵，当问题的规模比较大时，基于差别矩阵的属性约简算法存放差别矩阵的空间过大，相应地，其时间复杂度也比较高．

收稿日期：2007－09－04

作者简介：景永霞(1984－)，女，硕士研究生；通讯作者：王治和(1965－)，男，教授，硕士．研究方向：软件工程，数据库技术及其应用．

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

1 粗糙集的相关概念

定义 1　一个决策信息系统 S可以表示为 $S=(U, A, V, f)$ 其中：U是对象的集合；A是属性集合；$V=\bigcup_{a\in A} V_a$, V_a 表示属性 a的值域；$f: U\times A\to V$ 是信息函数，它指定 U中每个对象 x的属性值，即对 $x\in U$, $a\in A$, 有 $f(x, a)\in V_a$. C, D 分别为条件属性集和决策属性集，即 $C\cup D=A$, $C\cap D=\Phi$.

定义 2　C是条件属性集合，D是决策属性集合，其中 $C, D\subseteq A$, 若 $C'\subseteq C$ 是满足 $\gamma(C', D)=\gamma(C, D)$ 的最小集合，则 C'为 C相对于 D的约简，其中

$$\gamma(C, D)=\frac{|POS_C(U, D)|}{|U|}$$

$$B_*(X)=\bigcup\{y\in U/ind(B) | Y\subseteq X\}$$

$$POS_C(U, D)=\bigcup_{X\in U/ind(D)} C_*(X)$$

定义 3[3]　在决策表 $S=(U, C, D, V, f, d)$中，差别矩阵 $M=(m_{ij})$, 其元素 m_{ij}定义如下

$$\begin{cases} \{c_k | c_k\in C, f(x_i, c_k)\neq f(x_j, c_k), d(x_i, D)\neq d(x_j, D)\} \\ \Phi \quad 否则 \\ 其中\ k=1, 2, \cdots, r \end{cases}$$

定义 4[4]　在决策表 $S=(U, C, D, V, f, d)$中，$\forall a\in C$, 记

$R_a=\{<x_i, x_j> | x_i, x_j\in U(i<j)$,

$f(x_i, a)\neq f(x_j, a)\wedge(\exists s\in D, d(x_i, s)\neq d(x_j, s)\}$,

称 R_a 为属性 a的差别对象对集. 记

$R_C=\{<x_i, x_j> | x_i, x_j\in U(i<j)$,

$\exists a\in C\Rightarrow f(x_i, a)\neq f(x_j, a)\wedge$

$\exists s\in D, d(x_i, s)\neq d(x_j, s)\}$,

称 R_C 为条件属性集的差别对象对集.

2 基于差别矩阵的属性约简算法

决策表核的确定和属性约简算法是粗糙集理论研究的焦点问题，人们已提出了若干个属性求核和属性约简算法，其中应用较多的是基于差别矩阵以及在此基础上的一些改进算法[5]，这些算法的主要缺点是空间复杂度高，计算繁琐，不适用大规模数据.

在基于差别矩阵的属性约简算法中，通常是先求出差别矩阵，然后再根据所设置的启发信息选取一个属性 a放入属性约简中，再在差别矩阵的元素中删除所有包含该属性 a的元素，直至差别矩阵为空.

算法思想[4, 6]：

(1) 生成差别矩阵 $M_{n\times n}$, 属性约简集合 B初始化为空；

(2) 不断进行 (3)操作，直到差别矩阵 $M_{n\times n}$ 为空；

(3) 根据启发信息在 C－B中选取属性 a, 从 $M_{n\times n}$中去掉所有包含属性 a的元素，然后 $B=B\cup\{a\}$.

通过差别矩阵可以方便地求取核属性，以核属性为出发点，再求取差别函数 P的最小析取范式，最后得到的每个析取分量对应一个约简. 因此，一定可以得到最小约简.

但基于差别矩阵的属性约简算法需要先生成差别矩阵，差别矩阵将占用大量的存储空间，尤其是当对象较多，属性的规模较大时；其次，问题规模比较大时，差别函数本身的化简就是一个 NP－hard问题.

3 基于差别对象对的属性约简算法

算法 DOP:

输入：决策表 $S=(U, C, D, V, f)$, 其中，$U=\{x_1, x_2, \cdots, x_m\}$, $C=\{c_1, c_2, \cdots, c_m\}$

输出：属性约简集合 B

```
RC =Φ;
B=Φ;
for( i=1; i<= |U |-1; i++) { //生成每个属性的 Rci
    for( j=i+1; j<= |U |; j++) {
        for( k=1; k< |C |; k++) {
    if( d( xi, ck)! =d( xj, ck) )
        if( f( xi, ck)! =f( xj, ck) )
        Rck =Rck ∪ {<xi, xj>} } }
            temp[ i] = |Rck |; }
k=1;
for( i=1; i<= |C |; i++)
if temp[ i]>temp[ k]
    k=i
```

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.   http://www.cnki.net

$R_C = R_C - R_{c_k}$;

$B = B \cup R_{c_k}$;

Do while $R_C != \Phi$ {

对于 $\forall b \in C - B$

计算 $R_C - R_b$

选取一个使得 $|R_C - R_b|$最小的属性 b(假如有不止一个,则任选其一)

$R_C = R_C - R_b$

$B = B \cup \{b\}$ }

Output B

上述 DOP算法在生成每个条件属性的差别对象对集时,同时求出它的差别对象对集中元素的总个数(数值越大,表明该属性越重要),这样找出其差别对象对集中元素的总个数最多的属性 c_i 就可以直接将其加入到 B中,避免了生成 R_C,而是直接生成 $R_C = R_C - R_{c_i}$,节省了存储空间;然后不断去寻找使得 $|R_C - R_b|$最小的属性 b将它加入到属性约简中,并在 R_C 中减去其差别对象对集 R_b,直到 $R_C = \Phi$.

4 实例分析

对于文献[7]中的一个决策表,利用算法 DOP对其进行属性约简.

(1)首先计算各个属性所对应的差别对象对集 Rs, Re和 Rc,数组 temp用于计数 Rs={<u1, u2>, <u1, u3>, <u1, u6>, <u1, u7>, <u1, u8>, <u2, u4>, <u2, u5>, <u3, u4>, <u3, u5>, <u3, u8>, <u4, u6>, <u4, u7>, <u4, u8>, <u5, u6>, <u5, u7>, <u6, u8>, <u7, u8>} temp[1]=17

Re={<u1, u2>, <u1, u6>, <u1, u7>, <u1, u8>, <u2, u5>, <u3, u4>, <u3, u8>, <u4, u5>, <u4, u6>, <u5, u6>, <u5, u7>, <u6, u8>} temp[2]=12

Rc={<u1, u2>, <u1, u3>, <u1, u6>, <u1, u7>, <u1, u8>, <u2, u4>, <u2, u5>, <u3, u5>, <u3, u8>, <u4, u5>, <u4, u7>, <u4, u8>, <u5, u6>, <u5, u7>, <u6, u8>} temp[3]=15

(2)产生初始条件属性集的差别对象对集 R_C,为了节省空间,我们通过第一步各个属性的差别对象对集的元素个数计数,可以确定 Temp[1]最大,因而直接产生 $R_C = R_C - R_s$,并将 Temp[1]对应的属性大小(s)添加到属性约简集合 B中,于是

$R_C = R_C - Rs = \{<u4, u5>\}$

B={大小(s)}

(3)$R_C != \Phi$,从剩余条件属性中选择属性添加到属性约集合中,由于 $|R_C - Re|$和 $|R_C - Rc|$都为零,选择任意一个添加到属性约简集合即可,至此,$R_C = \Phi$,算法结束.于是得到该决策表的属性约简有两个

B_1 = {大小(s),发动机(e)}

B_2 = {大小(s),颜色(c)}.

实验结果表明,在基于差别矩阵的算法中,存放差别矩阵的空间可能很大,仍以文献[7]中的决策表为例,其中对象的个数为 8,条件属性的个数为 3时,存放差别矩阵的空间为 $1+9\times3+8\times2=44$.而本文提出的算法由于在第一步计算时就将 Temp[i]最大的属性的差别对象对集从 R_C 中除去,因而显著减少了它的存储量,就本例而言,由于属性大小(s)的差别对象对集的元素个数最多(即 Temp[i]最大),$R_C - Rs = \{<u4, u5>\}$,因而它所占的存储空间仅为 $2\times1=2$,由这个例子可以看出本算法对存储空间复杂度的显著降低.由于差别对象对集中元素个数的减少,其比较次数也将显著较少,因而其计算复杂度也随之得以改善.

5 结束语

基于差别对象对的属性约简算法没有生成差别矩阵,因此所需的辅助空间比基于差别矩阵的算法要小得多.在基于差别矩阵的算法中,特别当决策表中的对象很多且属性也比较多时,生成的差别矩阵太大,以至占空间太多,不利于算法的运行,而本文所提出的基于差别对象对的改进算法能明显降低存储量和计算量,从而提高了算法的效率.

(下转第 69页)

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

3.3 温度的影响

输入图 3中给定的 D/A值，改变刀架的环境温度，其磁滞回线如图 8所示．其中系列 1为常温（25℃）时的磁滞回线图，系列 2为高温（50℃）时的磁滞回线图．

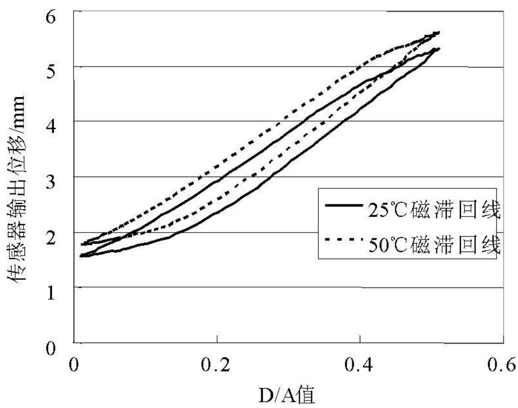

图 8　温度影响的磁滞回线

由图 8可以看出，温度高的磁滞回线相对于温度低的磁滞回线整体上移，温度对该刀架的输出位移有直接的影响．磁性棒周围的线圈在通入电流时，会产生大量的热，并将部分热量传递给磁性棒，因此，在加工时为了避免由于温度的原因产生的影响，可以将磁性棒周围通入冷却液来保持恒温．

4 结束语

通过对沈阳贝特数控机械有限公司的超磁致伸缩微位移刀架的实验研究，得出了以下结论：

（1）在标准 D/A值输入时，提高主轴转速可以减小磁滞回线对刀架的影响．

（2）在非标准 D/A值输入时，D/A值变化越大，磁滞回线对磁性棒的影响大，因此，为了减小磁滞回线带来的影响，尽量减小输入的 D/A值．

（3）磁性棒存在热胀冷缩现象，为了避免由于温度带来的误差，可以采取在磁性棒周围通入冷却液的方法来保持磁性棒的恒温．

参考文献：

[1] Jenner A G, Greenough R D, Wilkinson A J, et al. Performance of magnetostrictive rare earth iron compounds for device [J]. IEEE Trans On Magn, 1990, 26(5): 2016-2022.

[2] 袁惠群，周卓，李成英．稀土超磁致伸缩材料应力与电磁耦合特性的实验研究［J］．力学与实践，2000, 22(1)：27-30.

[3] 翁玲，王博文，孙英，等．超磁致伸缩致动器动态特性分析与实验研究［J］．河北工业大学学报，2005, 34(5)：14-17.

[4] 杨李色，周卓，李成英，等．稀土超磁致伸缩材料电磁参数的实验研究［J］．辽宁工学院学报，1999, 19(1)：14-19.

[5] 曹淑英，王博文，闫容格，等．超磁致伸缩致动器的磁滞的非线性动态模型［J］．中国电机工程学报，2003, 11(23)：145-149.

（上接第 20页）

参考文献：

[1] Pawlak Z. Rough sets[J]. International Journal of Computer and Information Science, 1982, 11(5): 341-356.

[2] Pei Xiaobing, Wang Yuanzhen. An Approximate Approach to Attribute Reduction International[J]. Journal of Information Technology, 2006, 12(4): 128-130.

[3] 徐章艳，杨炳儒，宋威．一个基于差别矩阵的快速求核算法［J］．计算机工程与应用，2006,（6）：4-6.

[4] 丁军，李凡，冯嘉礼．一种快速属性约简算法［J］．华中科技大学学报（自然科学版），2006,（8）：40-42.

[5] 芦晓红，陈世权，吴今培．基于可辨识矩阵的启发式属性约简方法及其应用［J］．计算机工程，2003, 29(1)：56-58.

[6] 陈淑珍．基于粗集的几种属性约简算法分析［J］．武汉工业学院学报，2005, 24(3)：118-119.

[7] 徐一新，叶东毅．知识约简的差别矩阵启发式算法［J］．福州大学学报（自然科学版），2000, 28(3)：121-123.

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net