# 基于信息熵的二进制差别矩阵属性约简算法

钱文彬[1],徐章艳[1],黄丽宇[1],杨炳儒[2]
QIAN Wen-bin[1],XU Zhang-yan[1],HUANG Li-yu[1],YANG Bing-ru[2]

1.广西师范大学 计算机科学与信息工程学院,广西 桂林 541004
2.北京科技大学 信息工程学院,北京 100083
1.School of Computer Science and Information Engineering,Guangxi Normal University,Guilin,Guangxi 541004,China
2.School of Information Engineering,University of Science and Technology Beijing,Beijing 100083,China
E-mail:qianwenbin1027@126.com

**QIAN Wen-bin,XU Zhang-yan,HUANG Li-yu,et al.Attribution reduction algorithm based on binary discernibility matrix of information entropy.Computer Engineering and Applications,2010,46(6):120-123.**

**Abstract:** The definition of attribution reduction of simplified binary discernibility matrix is provided.At the same time,it is proved that the above definition of attribution reduction is the same as the definition of attribution reduction based on information entropy.In order to compute simplified binary discernibility matrix,a quick algorithm for simplified decision table is designed,its time complexity is $O(|C||U|)$.In this condition,a quick attribution reduction algorithm based on simplified binary discernibility matrix of information entropy is designed,the time complexity and space complexity of the new algorithm are $\max\{O(|C||U|),O(|C|^2|U/C|^2)\}$ and $\max\{O(|C||U/C|^2),O(|U|)\}$ respectively.At last,an example is used to illustrate the efficiency of the new algorithm.

**Key words:** rough set;information entropy;simplified binary discernibility matrix;attribution reduction;algorithm complexity

**摘 要:**给出一个简化的二进制差别矩阵的属性约简定义,并证明该属性约简的定义与基于信息熵的属性约简的定义是等价的。为求出简化的二进制差别矩阵,设计了一个快速求简化决策表的算法,其时间复杂度为 $O(|C||U|)$。在此基础上,设计了基于信息熵的简化二进制差别矩阵的快速属性约简算法,其时间复杂度和空间复杂度分别为 $\max\{O(|C||U|),O(|C|^2|U/C|^2)\}$和 $\max\{O(|C||U/C|^2),O(|U|)\}$,最后用一个实例说明了新算法的高效性。

**关键词:**粗糙集;信息熵;简化的二进制差别矩阵;属性约简;算法复杂度
**DOI:**10.3778/j.issn.1002-8331.2010.06.034 **文章编号:**1002-8331(2010)06-0120-04 **文献标识码:**A **中图分类号:**TP301.6

## 1 引言

在粗糙集理论[1-2]中,属性约简是非常重要的研究内容之一,很多学者针对不同的需求提出了不同的属性约简。目前,一些学者从信息论的角度提出基于信息熵的属性约简[3-4],并以信息熵作为启发信息设计了一些属性约简算法,其中文献[3]中算法的时间复杂度为 $\max\{O(|C||U|^2),O(|U|^3)\}$,文献[4]中设计了两个算法,它们的时间复杂度分别为 $\max\{O(|C||U|^2),O(|U|^3)\}$和 $\max\{O(|C|^2|U|),O(|C||U|^3)\}$。近来,一些学者利用二进制差别矩阵设计了基于正区域的属性约简算法[5-6],其时间复杂度为 $O(|C||U|^4)$,空间复杂度为 $O(|C||U|^2)$。由于二进制差别矩阵便于计算机实现,且占用的存储空间较小,故该种方法是一种较好的设计属性约简的方法。经过深入研究,在文献[5-6]的基础上,用简化二进制差别矩阵设计了一个新的基于正区域的属性约简算法[7],使得其时间复杂度和空间度得到了改善。文献[8]指出基于信息熵的属性约简与基于正区域的属性约简是不等价的。到目前为止,尚没有学者利用简化的二进制差别矩阵的思想设计基于信息熵的属性约简算法。经研究,将其二进制差别矩阵的思想引入信息熵的属性约简中,设计出了基于信息熵的简化二进制差别矩阵的快速属性约简算法。

为了降低算法的复杂度,首先给出简化决策表,然后给出了一个基于简化的二进制差别矩阵的属性约简的定义,并证明了该定义与基于信息熵的属性约简的定义是等价的。在此基础上设计了基于信息熵的简化的二进制差别矩阵的快速属性约简算法,其时间复杂度和空间复杂度分别为:$\max\{O(|C||U|),O(|C|^2|U/C|^2)\}$和 $\max\{O(|C||U/C|^2),O(|U|)\}$。

## 2 基本概念

**定义** 1 五元组 $S=(U,C,D,V,f)$是一个决策表,其中 $U=$

**基金项目:**广西研究生科研创新基金项目(No.200910602M61);广西教育厅科研基金项目(No.200807MS015)。
**作者简介:**钱文彬(1984-),男,硕士研究生,主要研究方向:粗糙集理论及应用与数据挖掘;徐章艳(1972-),男,博士,副教授,主要研究方向:粗糙集,模糊集,数据挖掘;黄丽宇(1980-),女,硕士研究生,主要研究方向:粗糙集理论及应用与数据挖掘;杨炳儒(1943-),男,教授,博士生导师,主要研究方向:人工智能,数据挖掘和柔性建模。
**收稿日期:**2008-09-02 **修回日期:**2008-11-18

$\{x_1,x_2,\cdots,x_n\}$表示对象的非空有限集合,称为论域;$C$ 表示条件属性的非空有限集;$D$ 表示决策属性的非空有限集且 $C\cap D=\varnothing$;$V=\bigcup_{a\in C\cup D}V_a$,$V_a$ 是属性 $a$ 的值域,$f:U\times C\cup D\to V$ 是一个信息函数,它对一个对象的每一个属性赋予一个信息值,即 $\forall a\in C\cup D,x\in U$,有 $f(x,a)\in V_a$;每一个属性子集 $B\subseteq(C\cup D)$决定了一个二元不可区分关系 $IND(B)$:

$$IND(B)=\{(x,y)\in U\times U|\forall a\in B,f(x,a)=f(y,a)\}$$

关系 $IND(B)$构成了 $U$ 的一个划分,用 $U/IND(B)$,简记为 $U/B$,$U/B$ 中的任何元素$[x]_B=\{y|\forall a\in B,f(x,a)=f(y,a)\}$称为等价类。

**定义 2** 在决策表 $S=(U,C,D,V,f)$中,可以认为 $U$ 上任一属性集合 $B\subseteq C\cup D$(知识,等价关系簇,$U/B=\{B_1,B_2,\cdots,B_t\}$)是定义在 $U$ 上的子集组成的代数上的一个随机变量,其概率分布通过如下方法来确定(把该公式中的 $S$ 改成了 $B$):

$$[B:p]=\begin{bmatrix}B_1 & B_2 & \cdots & B_t\\ p(B_1) & p(B_2) & & p(B_t)\end{bmatrix}$$

其中 $p(B_i)=|B_i|/|U|$,$i=1,2,\cdots,t$。

**定义 3**[3] 在决策表 $S=(U,C,D,V,f)$中,决策属性集 $D(U/D=\{D_1,D_2,\cdots,D_k\})$相对于条件属性集 $C(U/C=\{C_1,C_2,\cdots,C_m\})$的条件熵 $H(D|C)$定义为:

$$H(D|C)=-\sum_{i=1}^{m}p(C_i)\sum_{j=1}^{k}p(D_j|C_i)\log p(D_j|C_i)$$

其中 $p(D_j|C_i)=|D_j\cap C_i|/|C_i|$。

**定义 4**[3] 在决策表 $S=(U,C,D,V,f)$中,对 $\forall b\in B\subseteq C$,若 $H(D|B)=H(D|(B-\{b\}))$,则称 $b$ 为 $B$ 中相对于 $D$ 是可省的(不必要的);否则 $b$ 为称 $B$ 中相对于 $D$ 是不可省的(必要的)。对 $\forall B\subseteq C$,若其任一元素相对于 $D$ 都是必要的,则称 $B$ 相对于 $D$ 是独立的。

**定义 5**[3] 在决策表 $S=(U,C,D,V,f)$中,若 $\forall B\subseteq C$,$H(D|B)=H(D|C)$且 $B$ 相对于 $D$ 是独立的,则称 $B$ 是 $C$ 的相对于 $D$ 的属性约简(基于信息熵的属性约简定义)。

**定义 6** 在决策表 $S=(U,C,D,V,f)$中,$\forall B\subseteq C$,$\forall x\in U$,定义

$$\mu_B(x)=\left(\frac{|D_1\cap[x]_B|}{|[x]_B|},\frac{|D_2\cap[x]_B|}{|[x]_B|},\cdots,\frac{|D_k\cap[x]_B|}{|[x]_B|}\right)$$

称 $\mu_B(x)$为 $B$ 关于 $D$ 的概率分配函数。

**定义 7** 设在一个决策表中 $S=(U,C,D,V,f)$ 中,记 $U/C=\{[x_1']_C,[x_2']_C,\cdots,[x_m']_C\}$,记 $U'=\{x_1',x_2',\cdots,x_m'\}$,称决策表 $S'=(U',C,D,V,f)$为简化决策表。

**定义 8** 在决策表 $S=(U,C,D,V,f)$中,$S'=(U',C,D,V,f)$为简化的决策表,定义简化的二进制差别矩阵为:$\boldsymbol{M}'=(m((i',j'),k))$,其元素定义为如下:

$$m((i',j'),k))=\begin{cases}1,c_k\in C,f(x_i',c_k)\neq f(x_j',c_k)\wedge\mu_C(x_i')\neq\mu_C(x_j')\\0,\text{else}\end{cases}$$

**定义 9** 设 $\boldsymbol{M}'=(m(i',j'),k)$为决策表 $S=(U,C,D,V,f)$的简化二进制差别矩阵,$B\subseteq C$,若 $B$ 满足:(1)由 $B$ 中所有属性对应的各列所构成的 $\boldsymbol{M}'$的子阵中,不全为 0 的行数等于 $\boldsymbol{M}'$中不全为 0 的行数;(2)$\forall B'\subset B$ 均不满足(1);则称 $B$ 是 $C$ 关于 $D$ 的一个属性约简(基于简化的二进制差别矩阵的属性约简)。

## 3 两种属性约简的等价性

**定理 1**[8-9] 在决策表 $S=(U,C,D,V,f)$中,$\forall P\subseteq C$,记为 $U/P=\{P_1,P_2,\cdots,P_t\}$,$U/C=\{C_1,C_2,\cdots,C_m\}$,$U/D=\{D_1,D_2,\cdots,D_k\}$,则 $H(D|P)=H(D|C)$的充分必要条件是:$\forall x\in U\Rightarrow\mu_P(x)=\mu_C(x)$。

**定理 2** 设 $\boldsymbol{M}'=(m((i',j'),k))$为决策表 $S=(U,C,D,V,f)$的简化二进制差别矩阵,$\forall P\subseteq C$,若 $H(D|P)=H(D|C)$,则 $P$ 中所有属性对应的各列所构成的 $\boldsymbol{M}'$的子阵中,不全为 0 的行数等于 $\boldsymbol{M}'$中不全为 0 的行数。

**证明** 假设结论不成立,即有由 $P=\{c_{i_1},c_{i_2},\cdots,c_{i_t}\}$中所有属性对应的各列所构成的 $\boldsymbol{M}'$的子阵 $M'(P)$中,不全为 0 的行数不等于 $\boldsymbol{M}'$中不全为 0 的行数;则在 $M'(P)$中一定存在一行 $(m((i,j),i_1),m((i,j),i_2),\cdots,m((i,j),i_t))=0$,而在 $\boldsymbol{M}'$中对应的这一行为$(m((i,j),1),m((i,j),2),\cdots,m((i,j),r))\neq 0$(这里 $C=\{c_1,c_2,\cdots,c_r\}$),即存在属性 $c_h\in C\wedge c_h\notin P$,使得 $m((i,j),h)=1$。由定义 8(简化的二进制差别矩阵的定义)可知,存在属性 $c_h\in C$ 且 $x_i'\in[x_i']_C\wedge x_j'\in[x_j']_C$,使得 $f(x_i',c_h)\neq f(x_j',c_h)\wedge\mu_C(x_i')\neq\mu_C(x_j')$,故$[x_i']_C\neq[x_j']_C$ 和$[x_i']_C\cup[x_j']_C\subseteq[x_i']_P$;另一方面,由于 $H(D|P)=H(D|C)$,故由定理 1 可知有:$\mu_P(x_i')=\mu_C(x_i')\wedge\mu_P(x_j')=\mu_C(x_j')$;由于$[x_i']_C\cup[x_j']_C\subseteq[x_i']_P$,由定义 6(概率分配函数的定义)可知 $\mu_P(x_i')=\mu_P(x_j')$,从而有 $\mu_C(x_i')=\mu_C(x_j')$。这与 $\mu_C(x_i')\neq\mu_C(x_j')$矛盾,故假设不成立。

综上所述,命题成立。

**定理 3** 设 $\boldsymbol{M}'=(m((i',j'),k))$为决策表 $S=(U,C,D,V,f)$的简化二进制差别矩阵,$\forall P\subseteq C$,若由 $P$ 中所有属性对应的各列所构成的 $\boldsymbol{M}'$的子阵中,不全为 0 的行数等于 $\boldsymbol{M}'$中不全为 0 的行数,则有 $H(D|P)=H(D|C)$。

**证明** 假设结论不成立,即有 $H(D|P)\neq H(D|C)$,由定理 1 可知有:$\exists x_i'\in U\Rightarrow\mu_P(x_i')\neq\mu_C(x_i')$,故有$[x_i']_P\neq[x_j']_C$;设$[x_i']_P=[x^0]_C\cup[x^1]_C\cup\cdots\cup[x^l]_C$(其中$[x_i']_C\cap[x_j']_C=\varnothing(i,j=1,2,\cdots,l;i\neq j)$,$x_i'=x^0,l\geqslant 1$),故一定存在 $x_j\in\{x^0,x^1,\cdots,x^l\}$使得 $\mu_P(x_j)\neq\mu_C(x_i')$,否则有 $\mu_C(x_i')=\mu_C(x^1)=\mu_C(x^2)=\cdots=\mu_C(x^l)$,由定理 1 的证明过程可知:$\mu_P(x_j)=\mu_C(x_i')$,这与 $\mu_P(x_j)\neq\mu_C(x_i')$矛盾;另一方面,由$[x_j]_C\cap[x_i']_C=\varnothing$ 且$[x_j]_C\cup[x_i']_C\subseteq[x_i']_P$,故一定 $\exists c_h\in C-P$ 使得 $f(x_j,c_h)\neq f(x_i',c_h)\wedge\mu_C(x_j)\neq\mu_C(x_i')$,根据由定义 8(简化的二进制差别矩阵的定义)可知,有 $m((i,j),h)=1$;而$[x_j]_C\cup[x_i']_C\subseteq[x_i']_P$,故 $\forall c_h\in P$,有 $m((i,j),h)=0$;即在 $\boldsymbol{M}'$中存在元素不全为 0 的一行而该行对应在由 $P$ 中所有属性对应的各列,所构成的 $\boldsymbol{M}'$子阵 $M'(P)$中的那一行的元素全为 0,这与条件矛盾,故假设不成立。

综上所述,命题成立。即有 $H(D|P)=H(D|C)$成立。

**定理 4** 基于信息熵的属性约简定义与基于简化的二进制差别矩阵的属性约简定义是等价的。

**证明** 由定理 2 和定理 3 即得命题成立。

定理 4 说明:基于信息熵的属性约简定义与基于简化的二进制差别矩阵的属性约简定义是等价的,这样就可以将基于信息熵的属性约简建立在简化的二进制差别矩阵上。该定理为设计新的快速属性约简算法提供了理论依据。

## 4 快速属性约简算法

### 4.1 求简化决策表的算法

求简化决策表,其实质就是求出不可区分关系 $IND(C)$。利用文献[10]的思想设计了一个求简化二进制差别矩阵的算法,其时间复杂度为 $O(|C||U|)$,空间复杂度为 $O(|U|)$[10]。

**算法 1** 求简化决策表

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

$\cdots,c_s\}$。

输出:$U'$,$m_i$,$M_i(i=1,2,\cdots,s)$,$\forall x\in U'$,$\mu_C(x)$。

(1)对每一个 $c_i(i=1,2,\cdots,s)$统计 $f(x_j,c_i)(j=1,2,\cdots,n)$的最大值和最小值,分别记为 $M_i$ 和 $m_i$;

(2)以静态链表依次存储对象 $x_1,x_2,\cdots,x_n$;令表头指针指向 $x_1$;

(3)for($i$=1,$i$<$s$+1,$i$++)

(3.1)第 $i$ 趟"分配":建立 $M_i-m_i+1$ 空队列,令 $front_k$ 和 $end_k(k=0,1,2,\cdots,M_i-m_i)$分别为第 $k$ 个队列的头指针和尾指针。将链表中的对象 $x\in U$ 按链表中的次序分配到第 $f(x,c_i)-m_i$ 个队列中去;

(3.2)第 $i$ 趟"收集":表头指针指向第一个非空队列的头指针,修改每一个非空队列的尾指针,令其指向下一个非空队列的队头对象,将 $M_i-m_i+1$ 个队列重新组成一个链表;

(4)设由第(3)步得到链表中的对象序列为 $x_1',x_2',\cdots,x_n'$;

$t$=1;$B_t$={$x_1'$};

for($j$=2,$j$<$n$+1,$j$++)

若任一 $c_i\in C(i=1,2,\cdots,s)$均有 $f(x_j',c_i)=f(x'_{j-1},c_i)$,则 $B_t=B_t\cup\{x_j'\}$;

否则{$t$=$t$+1;$B_t$={$x_j'$}};

(5)$U'$=Ø;

for($i$=1,$i$=$t$+1,$i$++)

取出 $B_i$ 中的第一个对象并入 $U'$并计算出 $\mu_C(y)$。

## 4.2 计算简化的二进制差别矩阵的算法

**算法 2** 求简化的二进制差别矩阵

输入:决策表 $S=(U,C,D,V,f)$,$U=\{x_1,x_2,\cdots,x_n\}$,$C=\{c_1,c_2,\cdots,c_s\}$。

输出:简化的二进制差别矩阵 $\boldsymbol{M}'=(m((i',j'),k))$;

(1)由算法 1,求出:$U'=(x_1',x_2',\cdots,x_m')$和 $\forall x\in U'$,$\mu_C(x)$;

(2)$\boldsymbol{M}'$=Ø;

(3)for($i$=1;$i$<$m$;$i$++)

for($j$=$i$+1;$j$<$m$+1;$j$++)

for($k$=1;$k$<$s$+1;$k$++) //把 for 循环中的 $r$ 改成了 $s$

if($\mu(x_i')\neq\mu(x_j')\wedge f(x_i',c_k)\neq f(x_j',c_k)$)

$m((i,j),k)$=1;

else

$m((i,j),k)$=0;

(4)对简化的二进制差别矩阵 $\boldsymbol{M}'=(m((i',j'),k))$的每一行做如下处理:

若该行的元素全为 0 或全为 1,则删除该行;

算法 2 的复杂度分析:算法 2 的第(1)步的时间复杂度为 $O(|C||U|)$;第(3)步的最坏时间复杂度为 $O(|C||U/C|^2)$;第(4)步的最坏时间复杂度为 $O(|C||U/C|^2)$;因此该算法的最坏时间复杂度为 $\max\{O(|C||U/C|^2),O(|C||U|)\}$。算法 2 的第(1)步的最坏空间复杂度为 $O(|U|)$;第(3),(4)步的最坏空间复杂度为 $O(|C||U/C|^2)$,故算法的最坏空间复杂度为 $\max\{O(|C||U/C|^2),O(|U|)\}$。

## 4.3 基于信息熵的快速属性约简算法

**算法 3** 求基于信息熵的属性约简

输入:简化的二进制差别矩阵 $\boldsymbol{M}'=(m(i',j'),k))$,$(1\leqslant k\leqslant s)$; //把括号中的 $r$ 改成了 $s$

输出:属性约简 $R$;

(1)$R$=Ø;

(2)在 $\boldsymbol{M}'$中,对每一行,若用该行只有一个元素为 1,则将该元素所对应的属性 $c_h$ 并入 $R$,并将属性 $c_h$ 对应的列上为 1 的元素所在的行和所对应的列去掉;$C=C-R$;

(3)将简化二进制差别矩阵的各行纵向相加,并将结果存入相应 $col[c_i]$中,$(c_i\in C)$;

(4)将 $col_{\max}[c_i]=\max\limits_{c_i\in C}\{col[c_i]\}$,若这样的属性不只一个,则任取一个;并在 $\boldsymbol{M}'$中将属性 $c$ 对应列上值为 1 的元素所在的行和所对应的列去掉;

(5)$R=R\cup\{c_i\}$;$C=C-R$;

(6)若 $\boldsymbol{M}'\neq\varnothing$,则返回步骤(3);否则算法终止,并输出属性约简 $R$。

算法 3 的复杂度分析:算法 3 的第(2)步的最坏时间复杂度为 $O(|C||U/C|^2)$;第(3),(4)步的最坏时间复杂度 $O(|C||U/C|^2)$;从第(2)步到第(6)步,最多执行$|C|$次,故其最坏时间复杂度为 $O(|C|^2|U/C|^2)$。结合算法 2 和算法 3 的时间复杂度,则基于信息熵的简化二进制差别矩阵快速属性约简算法的最坏时间复杂度为 $\max\{O(|C||U|),O(|C|^2|U/C|^2)\}$;而易知该快速属性约简算法的最坏空间复杂度为 $\max\{O(|C||U/C|^2),O(|U|)\}$。由于文献[5-6]中属性约简算法的时间复杂度和空间复杂度分别为 $O(|C||U|^4)$ 和 $O(|C||U|^2)$。因此该快速属性约简算法无论在时间复杂度上还是在空间复杂度上都较以前的算法好。

## 5 实例分析

用决策表 1 说明新属性约简算法的快速性。

对决策表 1 中的 15 个对象 $U$,利用算法 1 得到 $U'$:={$X1$,$X2$,$X8$,$X4$,$X6$,$X7$,$X13$}(其计算过程,请参看文献[10])。

$U/D$={{$X1$,$X3$,$X4$},{$X2$,$X5$,$X8$,$X12$,$X13$},{$X6$,$X7$,$X10$,$X11$,$X15$},{$X9$,$X14$}}={$D_1,D_2,D_3,D_4$}

**表 1 决策表 1**

| $U$ | $a$ | $b$ | $c$ | $d$ | $D$ | $U$ | $a$ | $b$ | $c$ | $d$ | $D$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $X1$ | 1 | 1 | 1 | 1 | 0 | $X9$ | 3 | 1 | 2 | 1 | 3 |
| $X2$ | 2 | 2 | 2 | 1 | 1 | $X10$ | 1 | 2 | 3 | 2 | 2 |
| $X3$ | 1 | 1 | 1 | 1 | 0 | $X11$ | 3 | 1 | 2 | 1 | 2 |
| $X4$ | 2 | 3 | 2 | 3 | 0 | $X12$ | 2 | 3 | 1 | 2 | 1 |
| $X5$ | 2 | 2 | 2 | 1 | 1 | $X13$ | 4 | 3 | 4 | 2 | 1 |
| $X6$ | 3 | 1 | 2 | 1 | 2 | $X14$ | 1 | 2 | 3 | 2 | 3 |
| $X7$ | 1 | 2 | 3 | 2 | 2 | $X15$ | 4 | 3 | 4 | 2 | 2 |
| $X8$ | 2 | 3 | 1 | 2 | 1 | | | | | | |

由概率分配函数可得 $\mu_c(*)$:$\mu_C(X1)=(1,0,0,0)$;$\mu_C(X2)=(0,1,0,0)$;$\mu_C(X4)=(1,0,0,0)$;$\mu_C(X6)=(0,0,2/3,1/3)$;$\mu_C(X7)=(0,0,2/3,1/3)$;$\mu_C(X8)=(0,1,0,0)$;$\mu_C(X13)=(0,1/2,1/2,0)$。

得到简化的决策表如表 2 所示。

**表 2 表 1 的简化决策表**

| $U$ | $a$ | $b$ | $c$ | $d$ | $\mu_c(*)$ |
|---|---|---|---|---|---|
| $X1$ | 1 | 1 | 1 | 1 | (1,0,0,0) |
| $X2$ | 2 | 2 | 2 | 1 | (0,1,0,0) |
| $X4$ | 2 | 3 | 2 | 3 | (1,0,0,0) |
| X6 | 3 | 1 | 2 | 1 | (0,0,2/3,1/3) |
| $X7$ | 1 | 2 | 3 | 2 | (0,0,2/3,1/3) |
| $X8$ | 2 | 3 | 1 | 2 | (0,1,0,0) |
| $X13$ | 4 | 3 | 4 | 2 | (0,1/2,1/2,0) |

对于简化决策表 2,利用算法 2 计算出简化的二进制差别矩阵如表 3 所示。

表 3 简化的二进制差别矩阵

| $m(i,j)$ | $a$ | $b$ | $c$ | $d$ | $m(i,j)$ | $a$ | $b$ | $c$ | $d$ |
|---|---|---|---|---|---|---|---|---|---|
| $m(1,2)$ | 1 | 1 | 1 | 0 | $m(4,6)$ | 1 | 1 | 0 | 1 |
| $m(1,6)$ | 1 | 0 | 1 | 0 | $m(4,8)$ | 0 | 0 | 1 | 1 |
| $m(1,7)$ | 0 | 1 | 1 | 1 | $m(4,13)$ | 1 | 0 | 1 | 1 |
| $m(1,8)$ | 1 | 1 | 0 | 1 | $m(7,8)$ | 1 | 1 | 1 | 0 |
| $m(2,4)$ | 0 | 1 | 0 | 1 | $m(7,13)$ | 1 | 1 | 1 | 0 |
| $m(2,6)$ | 1 | 1 | 0 | 0 | $m(8,13)$ | 1 | 0 | 1 | 0 |
| $m(2,7)$ | 1 | 0 | 1 | 1 | | | | | |

对于简化的二进制差别矩阵,利用算法 3 得出的属性约简的计算过程如下:

由算法 3 的第(4)步可知:第一次循环得到的 $col_{max}[c_i]=10$,所对应的属性为$\{a\}$,取出属性 $a$,则有 $R=\{a\}$,简化的二进制差别矩阵变为表 4 所示。

表 4 第一次循环后的决策表

| $m(i,j)$ | $a$ | $b$ | $c$ | $d$ |
|---|---|---|---|---|
| $m(1,7)$ | 0 | 1 | 1 | 1 |
| $m(2,4)$ | 0 | 1 | 0 | 1 |
| $m(4,8)$ | 0 | 0 | 1 | 1 |

由算法 3 的第(4)步可知:第二次循环得到的 $col_{max}[c_i]=3$,所对应的属性为$\{d\}$,取出属性 $d$,则有 $R=\{a,d\}$,简化的二进制差别矩阵变为空,算法结束,得到的属性约简为 $R=\{a,d\}$。

若用文献[5]中的算法所求得的二进制差别矩阵是 105 行,再用其算法中的变换,使计算量大大增加。而该算法所求得简化的二进制差别矩阵仅为 13 行。可见该算法不仅空间存储量小,而且计算量也大大减少了很多,这是因为在生成简化的二进制差别矩阵时可以减少大量的计算,而且在后续的计算中又能快速地缩小空间,从而能节省大量的计算时间。因此该新算法是一个快速的属性约简算法。

## 6 结束语

首先引入简化决策表从而有效地降低属性约简算法的复杂度,然后给出了基于简化的二进制差别矩阵的属性约简的定义,并证明了该定义与基于信息熵的属性约简的定义是等价的。在此基础上,利用简化的二进制差别矩阵的思想设计了一个高效的基于信息熵的属性约简算法,并分析了该算法的时间复杂度和空间复杂度,它们分别为:$\max\{O(|C||U|),O(|C|^2|U/C|^2)\}$和 $\max\{O(|C||U/C|^2),O(|U|)\}$。和文献[5-6]中的算法相比该算法的时间复杂度和空间复杂度都较低。由于该文的空间复杂度还不是很理想,因此利用差别矩阵的本身性质设计低空间复杂度的高效算法,将是后续工作。

## 参考文献:


[1] Pawlak Z.Rough set[J].International Journal of Computer and Information Science,1982,11(5):314-356.

[2] Pawlak Z,Wong S K M,Ziarko W.Rough sets:Probabilistic versus deterministic approach[J].Computational Intelligence,1988,29:81-95.

[3] 苗夺谦,胡桂荣.知识约简的一种启发式算法[J].计算机研究与发展,1999,36(6):681-684.

[4] 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):759-766.

[5] 支天去,苗夺谦.二进制可辨别矩阵的变换及高效属性约简算法的构造[J].计算机科学,2002,29(2):140-142.

[6] 周海岩,杨汀.基于二进制可辨识矩阵属性约简算法的改进[J].计算机工程与设计,2003,24(12):35-42.

[7] 徐章艳,杨炳儒,宋威.基于简化的二进制差别矩阵的快速属性约简算法[J].计算机科学,2006,33(4):65-68.

[8] Xu Zhang-yan,Yang bing-ru,Song Wei.Comparative study of different attribute reduction based on decision table[J].Chinese Journal of Electronics,2006,15(4A):953-956.

[9] 徐章艳,杨炳儒,郭燕萍,等.基于信息熵的快速求核算法[J].小型微型计算机系统,2007,28(2):279-282.

[10] 徐章艳,刘作鹏,杨炳儒,等.一个复杂度为 $\max\{O(|U||C|),O(|C|^2|U/C|)\}$的快速属性约简算法[J].计算机学报,2006,29(3):391-399.


(上接 88 页)

并公布 $g^{\frac{x_1x_2}{y_1y_2}} \bmod N$。

(3)利用协议 7,使 $p_i$ 各自得到一个 $z_i$($i$=1,2)。

(4)对 $i$=1,2,$p_i$ 计算 $g^{z_i} \bmod N$,并公布 $g^{z_i} \bmod N$。

(5)对 $i$=1,2,$p_i$ 计算 $g^{z_1}g^{z_2} \bmod N=g^{z_1+z_2} \bmod N$。

(6)对 $i$=1,2,$p_i$ 比较 $g^{z_1}g^{z_2} \bmod N$ 与 $g^{\frac{x_1x_2}{y_1y_2}} \bmod N$ 是否相等,若相等,表示输出是正确的,否则终止协议。

## 4 结束语

近年来,很多学者在半诚实模型下给设计出了一些基础协议,但关于恶意模型的协议的研究相对较少,给出了有关除法的几个协议,并分别对半诚实模型和恶意模型都给出了具体的安全双方计算协议。其中,协议 4~7,协议 9,协议 11 很容易可以推广到多方,由于篇幅的关系,这里不再叙述。协议 8,协议 10 如何推广到多方情形是值得进一步研究的问题。

## 参考文献:


[1] Yao A C.Protocols for secure computation[C]//Proc of the 23rd Annual IEEE Symposium on Foundations of Computer Science,Chicago,1982:160-164.

[2] Goldreich O,Micali S,Wigderson A.How to play any mental game[C]//Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing,1987:218-229.

[3] Naor M,Pinkas B.Oblivious transfer with adaptive queries[C]//Proceedings of Cryto'99,Santa Barbara,California,1999:573-590.

[4] Goldreich O,Goldwasser S,Linial N.Fault-tolerant computation in the full information model[C]//32nd FOCS,1991:447-457.

[5] Ostrovsky R,Yung M.How to withstand mobile virus attacks[C]//Proceedings of the 10th Annual ACM Symposium on Principles of Distributed Computing,1991:51-59.

[6] 罗文俊,李祥.多方安全矩阵乘积协议及应用[J].计算机学报,2005,28(7):1230-1235.

[7] 张禾瑞,郝鈉新.高等代数[M].4 版.北京:高等教育出版社,1999.

[8] 罗文俊.安全多方计算理论及其应用研究[D].贵阳:贵州大学,2005.

[9] 柯召,孙琦.数论讲义[M].北京:高等教育出版社,1986.