第37卷 第14期 计 算 机 工 程 2011年7月
Vol.37 No.14 Computer Engineering July 2011

·软件技术与数据库· 文章编号：1000—3428(2011)14—0036—03 文献标识码：A 中图分类号：TP312

# 不完备信息系统属性约简的矩阵算法

王 炜[1]，徐章艳[1,2]，王 帅[1]，杨炳儒[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)

**摘 要：**通过分析不完备信息系统中基于相容矩阵的属性约简算法，定义一个在单个条件属性下的新二进制相容矩阵，给出新矩阵间的吸收运算规则，利用该规则得到不完备信息系统条件属性集的二进制相容矩阵。通过判断新二进制相容矩阵间运算后得到的矩阵是否与条件属性集下的矩阵等价，找出不完备信息系统的属性约简。证明该属性约简与不完备信息系统中基于正区域的属性约简是等价的，并给出属性约简算法。实例证明该算法是合理的。

**关键词：**不完备信息系统；相容矩阵；二进制相容矩阵；属性约简；正区域

# Matrix Algorithm for Attribute Reduction in Incomplete Information System

WANG Wei[1], XU Zhang-yan[1,2], WANG Shuai[1], YANG Bing-ru[2]

(1. College of Computer Science and Information Engineering, Guangxi Normal University, Guilin 541004, China;
2. School of Information Engineering, University of Science and Technology Beijing, Beijing 100083, China)

**【Abstract】**By analyzing attribute reduction based on tolerance matrix in incomplete information system, a new binary tolerance matrix under single condition attribute is defined. At the same time, absorption of operation rule between two of them is also defined. The binary tolerance matrix under the whole condition attributes is gained by that rule. If the binary tolerance matrix based on some of condition attributes is equal to the binary tolerance matrix under the whole condition attributes, those condition attributes make up of an attribute reduction. It is proved that the attribute reduction acquired from this new method is equivalent to it based on positive region in incomplete information system, and an algorithm for attribute reduction is presented. An example is used to illustrate the rationality of this new algorithm.

**【Key words】**incomplete information system; tolerance matrix; binary tolerance matrix; attribute reduction; positive region

**DOI:** 10.3969/j.issn.1000-3428.2011.14.010

## 1 概述

粗糙集理论是由Pawlak教授在1982年提出的。属性约简是粗糙集理论研究的重要内容。然而在现实生活中，多种原因导致不完备信息系统大量存在。文献[1]给出了5种不同的知识约简形式。文献[2]利用论域中对象到决策属性值域的幂集上的映射进行属性约简和系统中的对象约简。文献[3-4]通过定义相容矩阵和分配决策矩阵，利用矩阵间的比较提取所有分配规则，并同时得到所有约简。文献[5]通过构造约束相似关系的二进制可辨矩阵得到下近似的属性核和属性约简。文献[6]讨论集合容差关系、条件属性的可辨识矩阵以及决策表的属性重要性3种属性约简算法。

本文提出一个在单个条件属性下的新二进制相容矩阵的定义，给出新矩阵间的吸收运算规则。同时，证明了利用该规则得到的属性约简与基于正区域的属性约简是一致的，并给出新二进制相容矩阵的属性约简算法及实例说明。

## 2 不完备信息系统

一个信息系统可以表示为四元组：$S=(U,A=C\cup D,V,f)$，其中，$U$表示对象集合，一般称为论域；$A=C\cup D$表示属性的非空有限集；$C$、$D$分别表示条件属性集和决策属性集，且$C\cap D=\varnothing$；$V=\cup V_a$，$V_a$表示属性$a\in A$的值域；$f$表示$U\times A\to V$的一个信息函数，它为每个对象在每个属性上赋予一个信息值，即$\forall a\in A, x\in U$有$f(x,a)\in V_a$。如果$D=\varnothing$，则称信息系统为信息表，否则称为决策表。若存在一个$x\in U$有$f(x,a)$未知(记作$f(x,a)=*$)，则称信息系统为不完备的，否则，称为完备的。

**定义1** 在不完备信息系统$S=(U,A,V,f)$中，设$P\subseteq A$，定义由$P$决定的容差关系$T$为：

$$T(P)=\{(x,y)\in U\times U\mid \forall a\in P, f(x,a)=f(y,a)\vee f(x,a)=*\vee f(y,a)=*\}$$

显然容差关系$T$是自反且对称的，但不一定是传递的。令：$T_P(x)=\{y\mid y\in U,\ (x,y)\in T(P)\}$，则$T_P(x)$表示在属性集合$P$下与对象$x$满足容差关系$T$的个体对象集合，称$T_P(x)$是对象$x$在属性集合$P$下的容差类，并且$U$上所有的容差类$T_P(x)$构成$U$的一个覆盖，即$\cup T_P(x)=U$。

**定义2** 不完备信息系统$S=(U,A,V,f)$中设$Q\subseteq U$，$P\subseteq A$，正区域定义为：

$$POS_P(Q)=\bigcup_{Y\in U/IND(Q)}\{x\mid T_P(x)\subseteq Y\}$$

如果取$P=C$，$Q=U$，则$POS_C(D)=\cup\{x\mid T(x)\subseteq D_i\}$，$D_i\in U/D$表示决策属性集$D$对$U$的划分，一般令$U_{pos}=$

**基金项目：**国家自然科学基金资助项目(60963008)

**作者简介：**王 炜(1985－)，女，硕士研究生，主研方向：粗糙集理论，数据挖掘；徐章艳，教授、博士；王 帅，硕士研究生；杨炳儒，教授、博士生导师

**收稿日期：**2010-12-16 **E-mail：**wangweihdne@126.com

$POS_C(D)$，$U_{neg}=U-POS_C(D)$。

对表1所示的不完备决策表，根据以上定义容易得到该决策表的正区域为 $U_{pos}=\{u_1,u_2,u_3\}$。

**表1 决策表**

| Car | $a_1$ | $a_2$ | $a_3$ | $a_4$ | $d$ |
|---|---|---|---|---|---|
| $U_1$ | High | Low | Full | Low | Good |
| $U_2$ | Low | * | Full | Low | Good |
| $U_3$ | * | * | Compact | High | Poor |
| $U_4$ | High | * | Full | High | Good |
| $U_5$ | * | * | Full | High | Excellent |
| $U_6$ | Low | High | Full | * | Good |

**定义3** 给定不完备决策表 $S=(U,A,V,f)$，对于 $\forall b\in B\subseteq C$，若 $POS_B(D)=POS_{B-\{b\}}(D)$，称 $b$ 为 $B$ 中相对于 $D$ 是不必要的，否则，称 $b$ 为 $B$ 中相对于 $D$ 是必要的；对于 $B\subseteq C$，如果 $B$ 中任意属性相对于 $D$ 都是必要的，称 $B$ 相对于 $D$ 是独立的。

**定义4** 给定不完备决策表 $S=(U,A,V,f)$，对于 $B\subseteq C$，如果 $POS_B(D)=POS_C(D)$ 且 $B$ 相对于 $D$ 是独立的，则称 $B$ 是 $C$ 相对于 $D$ 的属性约简。

对于表1,可以计算出 $POS_{\{a_3,a_4\}}(D)=POS_C(D)$,即 $\{a_3,a_4\}$ 是其基于正区域的属性约简。

## 3 相容矩阵及其属性约简的定义

**定义5** 在不完备决策表 $S=(U,A,V,f)$ 中,属性 $a$ ( $a\in C$ ) 的二进制相容矩阵为 $\boldsymbol{M}_a=(m_{ij})_{U_{pos}\times U}$，其元素定义如下:

$$m_{ij}=\begin{cases}1 & f(x_i,a)\neq f(x_j,a)\wedge f(x_i,a)\neq *\wedge f(x_j,a)\neq *,\\ & \text{当}x_i\in U_{pos}\wedge f(x_i,D)\neq f(x_j,D);\quad \text{者}f(x_i,a)\neq \\ & f(x_j,a)\vee f(x_i,a)=*\vee f(x_j,a)=*,\quad x_i\in U_{pos}\wedge \\ & x_j\notin U_{pos}\wedge f(x_i,D)=f(x_j,D)\wedge x_j\notin T_C(x_i)\\ 0 & \text{否则}\end{cases}$$

由上述定义可以看出，$m_{ij}=1$ 表示 $x_i$、$x_j$ 可以由条件属性 $a$ 区分，如果两者至少有一个在正区域且决策属性不同；或者 $x_i$、$x_j$ 可能由条件属性 $a$ 区分，如果两者只有一个在正区域中，决策属性相同且不满足相容关系。

**定义6** 在不完备决策表 $S=(U,A,V,f)$ 中,矩阵 $\boldsymbol{M}_a$、$\boldsymbol{M}_b$ 分别是属性 $a\in C$、$b\in C$ 的二进制相容矩阵，称两矩阵中对应元素进行或运算为 $\boldsymbol{M}_a$、$\boldsymbol{M}_b$ 的吸收运算，记为:

$$\boldsymbol{M}_{\{a,b\}}=\boldsymbol{M}_a\vee\boldsymbol{M}_b=(m_{ij})_{U_{pos}\times U}$$

其中，$\boldsymbol{M}_{\{a,b\}}$ 中的 $m_{ij}=\begin{cases}1 & \boldsymbol{M}_a、\boldsymbol{M}_b\text{中至少有一个}m_{ij}=1\\ 0 & \boldsymbol{M}_a、\boldsymbol{M}_b\text{中的}m_{ij}=0\end{cases}$。

**定义7** 在不完备决策表 $S=(U,A,V,f)$ 中，$P\subseteq C$，令 $\boldsymbol{M}_P=\vee\boldsymbol{M}_{a_i}$，$a_i\in P$，称$\boldsymbol{M}_P$为属性集 $P$ 下的相容矩阵。

那么条件属性 $C$ 中所有单个属性的相容矩阵进行吸收运算的结果为条件属性集的相容矩阵 $\boldsymbol{M}_C$，即 $\boldsymbol{M}_C=\underset{a\in C}{\vee}\boldsymbol{M}_a$。

**定义8** 在不完备决策表 $S=(U,A,V,f)$ 中，$P\subseteq C$，若 $\boldsymbol{M}_P=\boldsymbol{M}_C$ 且 $\forall a\in P,P'=P-a$，$\boldsymbol{M}_{P'}\neq\boldsymbol{M}_C$，则称 $P$ 是 $C$ 关于 $D$ 的一个属性约简。

**性质1** 若 $P\subseteq C$ 且在 $\boldsymbol{M}_P$ 中有 $m_{ij}=1$,则在 $\boldsymbol{M}_C$ 中 $m_{ij}=1$。

**性质2** 若 $P\subseteq C$，则有 $|\boldsymbol{M}_P|\leqslant|\boldsymbol{M}_C|$($|\boldsymbol{M}_P|$ 表示二进制相容矩阵 $\boldsymbol{M}_P$ 中元素为1的个数)。

这说明条件属性子集的二进制相容矩阵中元素为1的个数小于等于 $\boldsymbol{M}_C$ 中元素为1的个数。

**定理1** 在不完备决策表 $S=(U,A,V,f)$ 中，$P$ 是 $C$ 的一个属性约简，那么 $|\boldsymbol{M}_P|=|\boldsymbol{M}_C|$。

证明：根据定义8及其性质直接可证。

**定理2** 在不完备决策表 $S=(U,A,V,f)$ 中，$P\subseteq C$，若 $\boldsymbol{M}_P=\boldsymbol{M}_C$，则 $POS_P(D)=POS_C(D)$。

证明：假设有 $POS_P(D)\neq POS_C(D)$，则至少存在 $x_i\in U_{pos}$ 使得 $T_C(x_i)$ 在决策属性 $D$ 上是一致的，且 $T_P(x_i)$ 在 $D$ 上是不一致的。那么一定存在 $x_j\in T_P(x_i)$ 使得 $f(x_i,D)\neq f(x_J,D)$，因此，$T_P(x_i)\neq T_C(x_i)$，即一定存在 $c_k\in C-P$ 使得 $f(x_i,c_k)\neq f(x_j,c_k)$。由于 $x_i\in U_{pos}$，根据定义5，在相容矩阵 $\boldsymbol{M}_{c_k}$ 中有 $m_{ij}=1$，因此在条件属性集的相容矩阵 $\boldsymbol{M}_C$ 中有 $m_{ij}=1$。另一方面，由于 $x_j\in T_P(x_i)$，$f(x_i,D)\neq f(x_J,D)$，根据定义5，$\forall p_k\in P$ 在相容矩阵 $\boldsymbol{M}_{p_k}$ 中有 $m_{ij}=0$，在相容矩阵 $\boldsymbol{M}_P$ 中 $m_{ij}=0$，因此 $\boldsymbol{M}_P\neq\boldsymbol{M}_C$，这与条件矛盾，所以，假设不成立。综上所述，命题成立。

**定理3** 在不完备决策表 $S=(U,A,V,f)$ 中，$P\subseteq C$，若 $POS_P(D)=POS_C(D)$，则 $\boldsymbol{M}_P=\boldsymbol{M}_C$。

证明：假设 $\boldsymbol{M}_P\neq\boldsymbol{M}_C$，由 $P\subseteq C$，则在 $\boldsymbol{M}_P$ 中一定有 $m_{ij}=0$,在 $\boldsymbol{M}_C$ 中有 $m_{ij}=1$。即至少存在一个属性 $c_k\in C-P$ 使 $\boldsymbol{M}_{c_k}$ 中有 $m_{ij}=1$，而不存在 $p_k\in P$ 使 $\boldsymbol{M}_{p_k}$ 中有 $m_{ij}=1$。(1)当 $x_j\in U_{pos}$ 时，由于 $\boldsymbol{M}_{c_k}$ 中有 $m_{ij}=1$，因此有 $f(x_i,c_k)\neq f(x_j,c_k)$，$f(x_i,D)\neq f(x_j,D)$，因此，$x_j\notin T_C(x_i)$。因为在 $P$ 中不存在属性 $p_k$ 使 $f(x_i,p_k)\neq f(x_j,p_k)$，所以有 $x_j\in T_P(x_i)$。又因为 $f(x_i,D)\neq f(x_j,D)$，所以 $x_i\notin POS_P(D)$。然而，$x_i\in POS_C(D)$，因此，$POS_P(D)\neq POS_C(D)$。这与条件矛盾，假设不成立。(2)当 $x_j\notin U_{pos}$ 时，根据定义5可知，$m_{c_k}$ 中 $m_{ij}=1$ 的情况有2种，分别是:

1) $f(x_i,a)\neq f(x_j,a)\wedge f(x_i,a)\neq *\wedge f(x_j,a)\neq *\wedge f(x_iD)\neq f(x_j,D)$

2) $(f(x_i,a)\neq f(x_j,a)\vee f(x_i,a)=*\vee f(x_j,a)=*\wedge f(x_i,D)=f(x_j,D)\wedge x_j\notin T_C(x_i)$

在情况1)下，由于 $P$ 中不存在属性 $p_k$ 使得 $f(x_i,p_k)\neq f(x_J,p_k)$，因此 $x_j\in T_P(x_i)$。又由于 $f(x_i,D)\neq f(x_J,D)$，因此 $x_i\notin POS_P(D)$，而 $x_i\in POS_C(D)$，因此 $POS_P(D)\neq POS_C(D)$。这与条件矛盾,所以假设不成立。在情况2)下,不存在属性 $p_k$ 满足 $f(x_i,p_k)\neq f(x_j,p_k)\vee f(x_i,p_k)=*\vee f(x_j,p_k)=*$，即 $\forall p_k\in P$，都有 $f(x_i,p_k)=f(x_J,p_k)\neq *$，则有 $T_P(x_i)=T_P(x_j)$，所以，$x_i$ 和 $x_j$ 必同时属于 $POS_P(D)$ 或同时不属于 $POS_P(D)$。而 $x_i\in POS_C(D)$，$x_j\notin U_{pos}$，即 $x_j\notin POS_C(D)$，因此，$POS_P(D)\neq POS_C(D)$。这与条件矛盾，所以，假设不成立。

综上所述，命题成立。

## 4 新二进制相容矩阵的属性约简算法

在新的二进制相容矩阵属性约简中，首先要求出不完备决策表中的正区域 $POS_C(D)$，然后根据定义5得到每个条件属性下的相容矩阵，再进行属性约简。所以，先给出求单个属性下的相容矩阵算法

**算法1** 求属性 $a$ 的相容矩阵

**输入** 不完备决策表 $S=(U,A,V,f)$

**输出** 条件属性 $a$ 下的新的二进制相容矩阵 $\boldsymbol{M}_a=\left(m_{ij}\right)_{U_{pos}\times U}$

**Step1** 求出不完备决策表 S=(U,A,V,f)中每一个对象的相容类 $T_C(x_i)$和正区域 $POS_C(D)$，记 $POS_C(D)=\{x'_1,x'_2,\cdots,x'_n\}$;

**Step2** $U_{pos}=\{x'_1,x'_2,\cdots,x'_s\}$,$U_{neg}=\{x'_{s+1},x'_{s+2},\cdots,x'_n\}$;

**Step3** for(i=1;i<=s； i++)

for(j=1； j<=s;j++){

if(f(x'i,a)≠f(x'j,a)&& f(x'i,a)≠*

&& f(x'j,a)≠*&& f(x'i,D)≠f(x'j,D))

$m_{ij}$=1;

else $m_{ij}$=0;

}

**Step4** for(i=1;i≤ s;i++)

for(j=s+1;j≤ n;j++){

if(f(x'i,a)≠f(x'j,a)&&f(x'i,a)≠*&& f(x'j,a)≠*&& f(x'i,D)≠f(x'j,D))

$m_{ij}$=1;

if(f(x'i,D)==f(x'j,D)&& x'j 不属于 $T_C(x'_i)$){

if(f(x'i,a)≠f(x'j,a)||f(x'i,a)==*|| f(x'j,a)==*)

$m_{ij}$=1;

}

else $m_{ij}$=0;

}

**Step5** 输出 $M_a=(m_{ij})_{Upos\times U}$

**算法 2** 新二进制相容矩阵的属性约简算法

**输入** 不完备决策表 $S=(U,A,V,f)$

**输出** 属性约简 $P$

**Step1** 对条件属性集 $C=\{a_1,a_2,\mathrm{L}\ ,a_m\}$ 中的任一条件属性调用算法 1 求出相容矩阵，记为 $\boldsymbol{M}_{a_1}$，$\boldsymbol{M}_{a_2}$，$\cdots$，$\boldsymbol{M}_{a_{|C|}}$；$P=\varnothing$；$\boldsymbol{M}_P=0$。

**Step2** 根据定义 6 得到条件属性集 $C$ 下的二进制相容矩阵 $\boldsymbol{M}_C$。

**Step3** 将 $\boldsymbol{M}_{a_1}$，$\boldsymbol{M}_{a_2}$，$\cdots$，$\boldsymbol{M}_{a_{|C|}}$ 与 $\boldsymbol{M}_P$ 进行吸收运算分别得到 $\boldsymbol{M}_{a_i}=\boldsymbol{M}_P\vee\boldsymbol{M}_{a_i}$ $(i=1,2,\mathrm{L}\ ,|C|)$，并统计 $\boldsymbol{M}_{a_i}$ 中 1 的个数，对应的记为 $count_i$。

**Step4** 将 $count_i$ 值最大的对应的条件属性 $a'$ 放入 $P$ 中，即 $P=P\cup\{a'\}$，$\boldsymbol{M}_P=\boldsymbol{M}_{a_i}$，删除 $count_i$ 及其对应的相容矩阵 $\boldsymbol{M}_{a_i}$。

**Step5** 判断 $\boldsymbol{M}_P$ 是否等于 $\boldsymbol{M}_C$，相等，则输出 $P$；否则，重复步骤 3。

在算法 1 中，Step1 需要比较论域中的任意 2 个对象在条件属性集下是否满足相容关系，所以，时间复杂度为 $O(|C||U|^2)$；Step3 和 Step4 进行 $U_{pos}$ 与 $U$ 中的对象在每个条件属性下比较，时间复杂度为 $O(|C||U_{pos}||U|)$。算法 2 由于要调用算法 1 且要在单个条件属性下的相容矩阵之间进行吸收运算，因此其时间复杂度为：

$$\max(O(|C||U|^2),O(|C|^2|U_{pos}||U|))$$

下面以表 1 为例说明该算法的合理性。首先得到 $U_{pos}=\{u_1,u_2,u_3\}$，对 $C=\{a_1,a_2,a_3,a_4\}$ 中每个条件属性，根据算法 1 计算出二进制相容矩阵如下：

$$\boldsymbol{M}_{a_1}=\begin{bmatrix}0&0&0&0&0&1\\&0&0&1&0&0\\&&0&0&0&0\end{bmatrix}\quad\boldsymbol{M}_{a_2}=\begin{bmatrix}0&0&0&1&0&1\\&0&0&1&0&0\\&&0&0&0&0\end{bmatrix}$$

$$\boldsymbol{M}_{a_3}=\begin{bmatrix}0&0&1&0&0&0\\&0&1&0&0&0\\&&0&1&1&1\end{bmatrix}\quad\boldsymbol{M}_{a_4}=\begin{bmatrix}0&0&1&1&1&1\\&0&1&1&1&0\\&&0&0&0&0\end{bmatrix}$$

$$\boldsymbol{M}_C=\bigvee_{a\in C}\boldsymbol{M}_a=\begin{bmatrix}0&0&1&1&1&1\\&0&1&1&1&0\\&&0&1&1&1\end{bmatrix}$$

根据算法 2 的 Step3、Step4 知 $P=\{a_4\}$，因为属性 $a_4$ 对应的二进制相容矩阵中 *count* 值最大，但是 $\boldsymbol{M}_P\neq\boldsymbol{M}_C$，所以重复 Step3，计算 $\boldsymbol{M}_{P\cup\{a_1\}}$、$\boldsymbol{M}_{P\cup\{a_2\}}$、$\boldsymbol{M}_{P\cup\{a_3\}}$，限于篇幅，本文省略了具体的矩阵。由于

$$\boldsymbol{M}_{P\cup\{a_3\}}=\begin{bmatrix}0&0&1&1&1&1\\&0&1&1&1&0\\&&0&1&1&1\end{bmatrix}=\boldsymbol{M}_C$$

因此算法结束。所以，$\{a_3,a_4\}$ 是其属性约简，显然与基于正区域的属性约简是一致的。

## 5 结束语

本文为不完备决策表属性约简提供了一种新的算法。新算法在时间复杂度上有所提高。同时，如何设计出高效的求相容类的算法是降低算法时间复杂度的关键，是需要进一步研究的问题。

## 参考文献


[1] Kryszkiewicz M. Comparative Study of Alternative Types of Knowledge Reduction in Inconsistent System[J]. International Journal of Intelligent Systems, 2000, 16(1): 105-120.

[2] 蒙祖强, 史忠植. 不完备信息系统中基于相容粒度计算的知识获取方法[J]. 计算机研究与发展, 2008, 45(增刊): 264-267.

[3] 黄 兵, 周献中. 不完备信息系统分配约简与规则提取的矩阵算法[J]. 计算机工程, 2005, 31(17): 20-22.

[4] 黄 兵, 何 新, 周献中. 基于相容矩阵的粗计算[J]. 自动化学报, 2004, 30(3): 364-370.

[5] 赵亚鹏, 丁以中. 一种不完备信息系统的属性约简算法研究[J]. 计算机系统应用, 2008, 17(7): 46-48.

[6] 曾晓辉, 文 展. 不完备信息系统的属性约简算法[J]. 计算机工程, 2009, 35(24): 185-187.


编辑 张正兴

(上接第 35 页)

## 参考文献


[1] 陈志敏, 沈 洁, 赵 耀. 基于相关均值的协同过滤推荐算法[J]. 计算机工程, 2009, 35(22): 53-55.

[2] 邓爱林, 朱扬勇, 施伯乐. 基于项目评分预测的协同过滤推荐算法[J]. 软件学报, 2003, 14(9): 1621-1628.

[3] Lekakos G, Giaglis G. Improving the Prediction Accuracy of Recommendation Algorithms: Approaches Anchored on Human Factors[J]. Interacting with Computers, 2006, 18(3): 410-431.

[4] Pazzani M J. A Framework for Collaborative, Content-based and Demographic Filtering[J]. Artificial Intelligence Review, 1999, 13(5/6): 393-408.

[5] Oard D, Kim J. Implicit Feedback for Recommender Systems[C]// Proc. of AAAI'98 Workshop on Recommender Systems. [S. l.]: AAAI Press, 1998.

[6] 徐雁斐, 张 亮, 刘 炜. 基于协同标记的个性化推荐[J]. 计算机应用与软件, 2008, 25(1): 9-13.


编辑 金胡考