计算机科学 2004Vol.31№.2

知识表达系统的简化与集族的极小子集（Ⅱ）*）

李小霞　陈绵云
（华中科技大学控制科学与工程系　武汉 430074）

摘　要　论文定义知识表达系统的不可分辨矩阵，并且揭示如何通过不可分辨矩阵解决与知识表达系统三大属性简化相关的问题，包括判断一个属性是否在核中，判断一个对象是否在正区中，判断决策表是否协调，如何将知识表达系统属性简化、决策表条件属性简化、协调决策表条件属性简化问题转化为生成集族的Ⅱ型极小子集问题。

关键词　Ⅱ型极小子集，知识表达系统属性简化，决策表条件属性简化

Reduct in a Knowledge Representation System and Teeny Subset Related to a Subset Family(Ⅱ)

LI Xiao-Xia　CHEN Mian-Yun
(Dept. of Control Sci. & Eng,Huazhong Univ. of Sci. & Tech. ,Wuhan 430074)

Abstract　The paper defines indiscernibility matrix of a Knowledge Representation System(KRS). Then it reveals how to use indiscernibility matrix to solve the related problems with the three attribute reduct problems of a KRS, including determining whether an attribute is in the core, whether an object is included in positive region, whether a decision table is consistent, and how to convert the attribute reduct problem of a KRS, the condition attribute reduct problem of a decision table, and the condition attribute reduct problem of a consistent decision table to the problem of producing type Ⅱ teeny subset related to a subset family.

Keywords　Type Ⅱ teeny subset, Attribute reduct, Condition attribute reduct, Decision table, KRS

知识表达系统的简化、决策表条件属性简化、协调决策表条件属性简化是基本粗集理论[1,2]中很重要的问题。论文旨在研究这三大简化问题与集族的极小子集问题之间的转化关系。《知识表达系统的简化与集族的极小子集（Ⅰ）》已经论述了与集族的极小子集相关的一些问题。在此基础上，本文将定义不可分辨矩阵，研究如何用它将知识表达系统的三大简化问题转化为集族的Ⅱ型极小子集问题。事实上，可以用类似的方法导出如何用 Skowron[3]定义的分辨矩阵将知识表达系统三大简化问题转化为生成集族的Ⅰ型极小子集问题。

1. 基本概念

定义 3（知识表达系统的不可分辨矩阵）　S=(U,A,V)是一个知识表达系统，记 $U=\{x_1,x_2,\cdots,x_m\}$，定义 S 的不可分辨矩阵为 $\tilde{M}(S)$，简记为 $\tilde{M}$，其元素 $\tilde{M}_{ij}$定义为：$\tilde{M}_{ij}=\{a\mid a\in A$，并且 $a(x_i)=a(x_j)\}$，$i\neq j$，$i,j=1,2,\cdots,m$，称 $\tilde{M}_{ij}$为以 x_i 为行标，以 x_j 为列标的矩阵元素，记为 $\tilde{M}(S)(x_i,x_j)$，简记为 $\tilde{M}(x_i,x_j)$，定义 $\tilde{M}_{ii}=A$，$i=1,2,\cdots,m$。

容易知道知识表达系统 S 的不可分辨矩阵 $\tilde{M}(S)$为对称阵，并且它的主对角线上的元素为 A，因此在下文中，我们用 $\tilde{M}(S)$主对角线下面的部分来表示它。

定义 4（知识表达系统的属性简化[4]）　(U,A,V)是一个知识表达系统，C 是任意的 A 的非空子集，如果 C 满足下面的两个条件：a) $\bigcap \mathrm{Ind}(A)=\bigcap \mathrm{Ind}(C)$；b)不存在 C 的非空真子集 C^*，使得 $\bigcap \mathrm{Ind}(C^*)=\bigcap \mathrm{Ind}(A)$，则称 A 是 S 的属性简化。

定义 5（决策规则[4]）　(U,A,V)是一个知识表达系统，x 是 S 的对象，F、G 是 S 的属性，$F\cap G=\varnothing$，定义 $R(x,F,G)=(\{(a,a(x))\mid a\in F\},\{(b,b(x))\mid b\in G\})$，称为 S 的以 x 为标识，以 F 为条件属性，以 G 为决策属性的规则，称$\{(a,a(x))\mid a\in F\}$为规则的条件，记为 Con(x,F,G)，称$\{(b,b(x))\mid b\in G\}$为 R(x,F,G)的决策，记为 Dec(x,F,G)。如果 R(x,F,G)满足对任意的 $y\in U$，Con(y,F,G)=Con(x,F,G)蕴含 Dec(y,F,G)=Dec(x,F,G)，则称 R(x,F,G)协调。

定义 6（决策表条件属性的简化[4]）　(U,C,D)是一个决策表。定义 $\mathrm{Pos}(C,D)=\{x\mid x\in U, R(x,C,D)$协调$\}$，称为(U,C,D)的正区。$a^*$是 C 中的元素，如果 $C-\{a^*\}\neq\varnothing$，并且 $\mathrm{Pos}(C-\{a^*\},D)=\mathrm{Pos}(C,D)$，则称 a^*可以从 C 中去掉。如果 C 中任意一个元素都不能去掉，则称(U,C,D)独立。C^*是 C 的任意的非空子集，如果满足下面的两个条件：(1)$\mathrm{Pos}(C^*,D)=\mathrm{Pos}(C,D)$，(2)$\mathrm{Pos}(C^{**},D)\neq\mathrm{Pos}(C,D)$，$C^{**}$是 C^*的任意的非空真子集，则称 C^*是(U,C,D)条件属性的简化。

2. 知识表达系统的简化与集族的Ⅱ型极小子集

定理 5　S=(U,A,V)是一个知识表达系统，则有下面的结论成立：(1)a^*是 A 中任意一个元素，a^*在 S 的属性简化中是不可去掉的当且仅当 $A-\{a^*\}$是 $\tilde{M}(S)$的一个元素。(2)定义 $\tilde{F}(S)=\{A^*\mid\varnothing\neq A^*\subset A, A^*$是 $\tilde{M}(S)$的一个元素$\}$，对 A 的任意非空子集 C 而言，C 是 S 的属性简化的充要条件是 C 是$(A,\tilde{F}(S))$的Ⅱ型极小子集。

证明：(1)a^*是 A 中任意一个元素，a^*在 S 的属性简化中是不可去掉的，等价于$\bigcap \mathrm{Ind}(A-\{a^*\})\neq\bigcap \mathrm{Ind}(A)$，等价于 $\exists\, x,y\in U$，$(x,y)\in\bigcap \mathrm{Ind}(A-\{a^*\})$，并且$(x,y)\notin\bigcap \mathrm{Ind}$

*）基金项目：自然科学基金(79970025)。**李小霞**　博士生，研究方向包括知识发现、粗集理论及其应用；**陈绵云**　教授，博导，研究方向包括灰色动态建模与控制、模糊控制与决策、一般系统论及应用。

(A)，等价于∃ x，y∈U，$\tilde{M}(x,y)\supseteq A-\{a^*\}$并且$\tilde{M}(x,y)$不包含A，等价于∃ x，y∈U，$\tilde{M}(x,y)=A-\{a^*\}$。(2)C是$(A,\tilde{F}(S))$的Ⅰ型极小子集，等价于$\varnothing\neq C\subseteq A$；$\tilde{F}(S)$中任意一个元素都不包含C；C的任意非空真子集都不具有这样的性质，等价于$\varnothing\neq C\subseteq A$；∀ x，y∈U，如果$\varnothing\neq\tilde{M}(x,y)\subset A$，则$\tilde{M}(x,y)$不包含C；C的任意非空真子集都不具有这样的性质，等价于$\varnothing\neq C\subseteq A$；∀ x，y∈U，如果$(x,y)\notin\bigcap \mathrm{Ind}(A)$，则$(x,y)\notin\bigcap \mathrm{Ind}(C)$；C的任意非空真子集都不具有这样的性质，等价于$\varnothing\neq C\subseteq A$；∀ x，y∈U，如果$(x,y)\in\bigcap \mathrm{Ind}(C)$，则$(x,y)\bigcap \mathrm{Ind}(A)$；C的任意非空真子集都不具有这样的性质，等价于$\varnothing\neq C\subseteq A$；$\bigcap \mathrm{Ind}(C)\subseteq\bigcap \mathrm{Ind}(A)$；C的任意非空真子集都不具有这样的性质，因为$C\subseteq A$中，使得$\bigcap \mathrm{Ind}(A)\subseteq\bigcap \mathrm{Ind}(C)$，所以前面的命题等价于$\varnothing\neq C\subseteq A$；$\bigcap \mathrm{Ind}(C)=\bigcap \mathrm{Ind}(A)$；C的任意非空真子集都不具有这样的性质，等价于C是(U，A，V)的属性简化。证毕。

表1

U	*a*	*c*	*d*	*e*	*o*
1	0	1	1	1	1
2	1	1	0	1	0
3	1	0	0	1	1
4	1	0	0	1	0
5	1	0	0	0	0
6	1	1	0	1	1

表1可以表示为知识表达系统S=({1，2，3，4，5，6}，{a，c，d，e，o}，{0，1})，根据定义，它的不可分辨矩阵

$$\tilde{M}(S)=\left[\begin{array}{cccccc} & 1 & 2 & 3 & 4 & 5 \\ 2 & ce & & & & \\ 3 & eo & ade & & & \\ 4 & e & adeo & acde & & \\ 5 & \varnothing & ado & acd & acdo & \\ 6 & ceo & acde & adeo & ade & ad \end{array}\right]$$

因为{a，d，e，o}、{a，c，d，e}、{a，c，d，o}在$\tilde{M}(S)$中，所以根据定理5的第一个结论，可知c、e、o在S的属性简化中不能去掉。根据相关定义，可得$\tilde{F}(S)$={{c，e}{c，o}{a，d}{e}{c，e，o}{a，d，e}{a，d，o}{a，c，d}{a，d，e，o}{a，c，d，e}{a，c，d，o}}，因为({a，c，d，e，o}，$\tilde{F}(S)$)的所有Ⅰ型极小子集为{c，e，o，a}{c，e，o，d}，所以根据定理5的第二个结论，S=(U，{a，c，d，e，o}，{0，1})的所有的属性简化为{c，e，o，a}{c，e，o，d}。

定理6 S=(U，A，V)是一个知识表达系统，|A|>1，d∈A，有下面的结论成立：(1)x是U中任意一个元素，C是A－{d}的任意的非空子集，x∈Pos(C，{d})的充要条件是$C\subseteq\tilde{M}(x,y)$蕴含$d\in\tilde{M}(x,y)$，其中y是U中任意一个元素。(2)x是U中任意一个元素，x∈Pos(A－{d}，{d})的充要条件是不存在y∈U，使得$\tilde{M}(x,y)=A-\{d\}$。(3)a^*是A－{d}中任意一个元素，a^*在决策表(U，A－{d}，{d})的条件属性简化中不可去掉，等价于U中存在元素x，y，使得x∈Pos(A－{d}，{d})并且$\tilde{M}(x,y)=A-\{a^*,d\}$。(4)定义$\tilde{F}(DT)=\{A^*\mid\varnothing\neq A^*\subset A-\{d\}$，并且存在x，y∈U，满足x∈Pos(A－{d}，{d})，$\tilde{M}(x,y)=A^*\}$，则对A－{d}的任意非空子集C而言，C是决策表(U，A－{d}，{d})的条件属性简化的充要条件是C是(A－{d}，$\tilde{F}(DT)$)的Ⅰ型极小子集。

证明：(1)∀ x∈U，∀ C⊆A－{d}，C≠∅，x∈Pos(C，{d})，等价于规则R(x，C，{d})协调，等价于∀ y∈U，Con(x，C，{d})=Con(y，C，{d})蕴含Dec(x，C，{d})=Dec(y，C，{d})，等价于∀ y∈U，$C\subseteq\tilde{M}(x,y)$蕴含$d\in\tilde{M}(x,y)$。(2)根据此定理的第一个结论，∀ x∈U，x∈Pos(A－{d}，{d})，等价于∀ y∈U，$A-\{d\}\subseteq\tilde{M}(x,y)$蕴含$d\in\tilde{M}(x,y)$，等价于不存在$y\in U$，使得$\tilde{M}(x,y)=A-\{d\}$。(3)$a^*$是A－{d}中任意一个元素，$a^*$在(U，A－{d}，{d})的条件属性简化中不可去，等价于Pos(A－{a^*，d}，{d}≠(Pos(A－{d}，{d})，等价于∃ x∈U，x∈Pos(A－{d}，{d})并且x∉Pos(A－{a^*，d}，{d})，等价于∃ x，y∈U，x∈Pos(A－{d}，{d})，Con(x，A－{a^*，d}，{d})=Con(y，A－{a^*，d}，{d})，并且Dec(x，A－{a^*，d}，{d})≠Dec(y，A－{a^*，d}，{d})，等价于∃ x，y∈U，x∈Pos(A－{d}，{d})，$A-\{a^*,d\}\subseteq\tilde{M}(x,y)$，$d\notin\tilde{M}(x,y)$，等价于∃ x，y∈U，x∈Pos(A－{d}，{d})，$\tilde{M}(x,y)=A-\{a^*,d\}$或者$\tilde{M}(x,y)=A-\{d\}$，因为x∈Pos(A－{d}，{d})，根据此定理的第二个结论，可知前面的命题等价于∃ x，y∈U，x∈Pos(A－{d}，{d})并且$\tilde{M}(x,y)=A-\{a^*,d\}$。(4)C是(U，A－{d}，{d})条件属性的简化，等价于∅≠C⊆A－{d}；Pos(A－{d}，{d})=Pos(C，{d})；C的任意非空真子集都不具有这样的性质，等价于∅≠C⊆A－{d}；∀ x∈U，x∈Pos(A－{d}，{d})蕴含x∈Pos(C，{d})；C的任意非空真子集都不具有这样的性质，根据此定理的第一个结论，可知前面的命题等价于∅≠C⊆A－{d}；∀ x∈U，x∈Pos(A－{d}，{d})蕴含∀ y∈U，若$C\subseteq\tilde{M}(x,y)$，则$d\in\tilde{M}(x,y)$；C的任意非空真子集都不具有这样的性质，等价于∅≠C⊆A－{d}；∀ x∈U，x∈Pos(A－{d}，{d})蕴含∀ y∈U，若$d\notin\tilde{M}(x,y)$，则C不包含在$\tilde{M}(x,y)$中；C的任意非空真子集都不具有这样的性质，等价于∅≠C⊆A－{d}；∀ x∈U，x∈Pos(A－{d}，{d})蕴含∀ y∈U，若$\tilde{M}(x,y)\subseteq A-\{d\}$，则C不包含在$\tilde{M}(x,y)$中；C的任意非空真子集都不具有这样的性质，根据此定理的第二个结论，可知前面的命题等价于∅≠C⊆A－{d}；∀ x∈U，x∈Pos(A－{d}，{d})蕴含∀ y∈U，若$\tilde{M}(x,y)\subset(A-\{d\}$，则C不包含在$\tilde{M}(x,y)$中；C的任意非空真子集都不具有这样的性质，等价于∅≠C⊆A－{d}；∀ x∈U，x∈Pos(A－{d}，{d})蕴含∀ y∈U，若$\varnothing\neq\tilde{M}(x,y)\subset A-\{d\}$，则C不包含在$\tilde{M}(x,y)$中；C的任意非空真子集都不具有这样的性质，等价于∅≠C⊆A－{d}；∀ x，y∈U，如果x∈Pos(A－{d}，{d})并且$\varnothing\neq\tilde{M}(x,y)\subset(A-\{d\}$，则C不包含在$\tilde{M}(x,y)$中；C的任意非空真子集都不具有这样的性质，等价于∅≠C⊆A－{d}；C不被$\tilde{F}(DT)$中任何元素所包含；C的任意非空真子集都不具有这样的性质，等价于C是(A－{d}，$\tilde{F}(DT)$)的Ⅰ型极小子集。证毕。

表2

U	*a*	*b*	*d*
1	1	0	0
2	1	2	0
3	2	1	1
4	2	1	2
5	1	2	3

表2中的知识表达系统S=({1，2，3，4，5}，{a，b，d}，{0，1，2，3})的不可分辨矩阵

$$\tilde{M}(S)=\left[\begin{array}{ccccc} & 1 & 2 & 3 & 4 \\ 2 & ad & & & \\ 3 & \varnothing & \varnothing & & \\ 4 & \varnothing & \varnothing & ab & \\ 5 & a & ab & \varnothing & \varnothing \end{array}\right]$$

(下转第25页)

件学报,2000,11(11):1460~1466
2 Debar H,Dacier M,Wespi A. Towards a taxonomy of intrusion-detection systems. Computer Networks,1999,31(8):805~822
3 Paxson V. Bro:A system for detecting network intruders in real-time. Computer Networks,1999,31(23):2435~2463
4 Manganaris S, Christensen M, Zerkle D, et al. Data mining analysis of RTID alarms. Computer Networks,2000,34(4):571~577
5 Puketza N,Zhang K,Chung M,et al. A methodology for testing intrusion detection systems. IEEE Transactions on Software Engineering,1996,22(10):719~729
6 蔡忠闽,孙国基,卫军胡,等. 入侵检测系统评估环境的设计与实现. 系统仿真学报,2002,14(3):377~380
7 Puketza N,Chung M,Olsson R A,et al. A Software Platform for Testing Intrusion Detection Systems. IEEE Software,1997,14(5):43~51
8 Mchugh J. Testing Intrusion Detection Systems: A Critique of the 1998 and 1999 DARPA Intrusion Detection System Evaluations as Performed by Lincoln Laboratory. ACM Transactions on Information and System Security,2000,3(4):262~294
9 Lippmann R, Fried D, Graf I, et al. Evaluating Intrusion Detection Systems: the 1998 DARPA Off-Line Intrusion Detection Evaluation. In: Proc. of the 2000 DARPA Information Survivability Conf. and Exposition (DISCEX), Hilton Head: IEEE. 2000,2:1012~1035
10 Lippmann R,Haines J,Fried D,et al. The 1999 DARPA Off-Line Intrusion Detection Evaluation. Computer Networks, 2000, 34(4):579~595
11 Haines J,Rossey L,Lippmann R,et al. Extending the DARPA Off-Line Intrusion Detection Evaluations. http://www.cs.rpi.edu/~brancj/publications/discex01_paper.pdf
12 Lippmann R,Haines J,Fried D,et al. Analysis and Results of the 1999 DARPA Off-Line Intrusion Detection Evaluation. http://www.cs.fit.edu/~pkc/id/related/lippmann-raid00.pdf.
13 Durst R,Champion T,Witten B,et al. Testing And Evaluating Computer Intrusion Detection Systems. Communications Of The ACM,1999,42(7):53~61
14 Pickering K. Evaluating The Viability of Intrusion Detection System Benchmarking. http://www.cs.virginia.edu/~evans/theses/pickering.pdf.
15 Korba J. Windows NT Attacks for the Evaluation of Intrusion Detection Systems. http://www.ll.mit.edu/IST/ideval/pubs/2000/jkorba_thesis.pdf
16 Kendall K. A Database of Computer Attacks for the Evaluation of Intrusion Detection Systems, http://www.ll.mit.edu/IST/ideval/pubs/1998/kkendall_thesis.pdf.
17 Das K. Attack Development for Intrusion Detection Evaluation. http://www.cc.gatech.edu/~wenke/ids-readings/attack_development_thesis.pdf

(上接第10页)

根据定理6的第二个结论,可知Pos({a,b},{d})={1}。DT=({1,2,3,4,5},{a,b},{d}),根据定理6的第三个结论,可知b在DT的条件属性简化中不能去掉。根据相关的定义,可知$\tilde{F}$(DT)={{a}},({a,b},$\tilde{F}$(DT))所有的I型极小子集为{b},因此根据定理6的第四个结论,可知决策表({1,2,3,4,5},{a,b}{d})所有的条件属性简化为{b}。

定理7 S=(U,A,V)是一个知识表达系统,|A|>1,d∈A,有下面的结论成立:(1)∅≠C⊆A-{d},决策表(U,C,{d})协调,当且仅当C⊆$\tilde{M}$(x,y)蕴含d∈$\tilde{M}$(x,y),x,y是U中任意的元素。(2)决策表(U,A-{d},{d})协调,当且仅当A-{d}不是$\tilde{M}$(S)的元素。(3)a^*是A-{d}中任意一个元素,a^*在(U,A-{d},{d})的条件属性简化中不可去,当且仅当A-{a^*,d}是$\tilde{M}$(S)中的元素。(4)定义$\tilde{F}$(DT)={A^*|∅≠A^*⊂A-{d},A^*是$\tilde{M}$(S)的元素},其中DT=(U,A-{d},{d}),如果DT协调,则对A-{d}的任意非空子集C而言,C是决策表DT条件属性的简化,当且仅当C是(A-{d},$\tilde{F}$(DT))的I型极小子集。

证明:(1)(U,C,{d})协调,等价于∀x∈U,x∈Pos(C,{d}),根据定理6的第一个结论,可知前面的命题等价于∀x,y∈U,C⊆$\tilde{M}$(x,y)蕴含d∈$\tilde{M}$(x,y)。(2)(U,A-{d},{d})协调,等价于∀x∈U,x∈Pos(A-{d},{d}),根据定理6的第二个结论,可知前面的命题等价于∀x∈U,不存在y∈U,使得$\tilde{M}$(x,y)=A-{d},等价于A-{d}不是$\tilde{M}$(S)的元素。(3)a^*是A-{d}中任意的一个元素,根据定理6的第三个结论,可知a^*在(U,A-{d},{d})的条件属性简化中不可去,等价于∃x,y∈U,x∈Pos(A-{d},{d})并且$\tilde{M}$(x,y)=A-{a^*,d},因为(U,A-{d},{d})协调,所以前面的命题等价于A-{a^*,d}是$\tilde{M}$(S)的元素。(4)定义F'(DT)={A^*|∅≠A^*⊂A-{d},并且存在x,y∈U,满足x∈Pos(A-{d},{d}),$\tilde{M}$(x,y)=A^*},因为(U,A-{d},{d})协调,所以F'(DT)={A^*|∅≠A^*⊂A-{d},A^*是$\tilde{M}$(S)的元素},即F'(DT)等于这里定义的$\tilde{F}$(DT),根据定理6的第四个结论,可知C是(U,A-{d},{d})条件属性的简化,等价于C是(A-{d},F'(DT))的I型极小子集,进而等价于C是(A-{d},$\tilde{F}$(DT))的I型极小子集。证毕。

表3

U	*a*	*b*	*c*	*d*
1	1	0	2	0
2	2	1	0	1
3	2	1	2	2
4	1	2	2	0
5	1	2	0	2

表3中的知识表达系统S=({1,2,3,4,5},{a,b,c,d},{0,1,2})的不可分辨矩阵

$$\tilde{M}(S)=\begin{bmatrix} & 1 & 2 & 3 & 4 \\ 2 & \varnothing & & & \\ 3 & c & ab & & \\ 4 & acd & \varnothing & c & \\ 5 & a & c & d & ab \end{bmatrix}$$

因为{a,b,c}不是$\tilde{M}$(S)的元素,所以根据定理7的第二个结论,决策表DT=({1,2,3,4,5},{a,b,c},{d})协调。因为{a,b}是$\tilde{M}$(S)的元素,所以根据定理7的第三个结论可知c在DT条件属性的简化中是不可去的。根据定义,$\tilde{F}$(DT)={{c}{a}{a,b}},({a,b,c},$\tilde{F}$(DT))所有的I型极小子集为{c,a}{c,b},因此,根据定理7的第四个结论,可知决策表({1,2,3,4,5},{a,b,c},{d})所有的条件属性简化为{c,a}{c,b}。

至此,论文使用不可分辨矩阵将知识表达系统的三大简化问题转化为生成集族的I型极小子集问题,可以使用类似的方法导出用Skowron[3]定义的分辨矩阵将知识表达系统三大简化问题转化为生成集族的I型极小子集问题。

参考文献

1 Pawlak Z. Rough Sets-Theoretical Aspects of Reasoning about Data. Dordrecht: Kluwer Academic ,1991
2 曾黄麟. 粗集理论及其应用—关于数据推理的新方法(修订版). 重庆大学出版社,1998
3 Skowron A,Rauszer C. The discernibility matrices and functions in information systems. In: Slowinski R,ed. Intelligent Decision Support-Handbook of Applications and Advances of the Rough Sets Theory. Dordrecht,Kluwer: Academic Publishers,1992. 331~362
4 李小霞,陈绵云. 决策表条件属性的简化. 华中科技大学学报(已录用)