维普资讯 http://www.cqvip.com

集值决策信息系统的知识约简与规则提取*)

宋笑雪[1,2] 解争龙[2] 张文修[1]
(西安交通大学理学院信息与系统科学研究所 西安 710049)[1] (咸阳师范学院计算机系 咸阳 712000)[2]

摘 要 本文用集值信息系统描述不完备信息系统。在集值信息系统中定义了两种不同的关系:相容关系和优势关系,给出了两种不同关系下集合的上下近似概念及其性质。研究了集值决策信息系统在两种不同关系下的广义决策约简,得到了广义决策约简的判定定理和辨识矩阵,从而得到了约简的具体操作方法。最后,基于两种关系从集值决策信息系统提取了最优广义决策规则,并进行了讨论。

关键词 集值信息系统,知识约简,辨识矩阵,决策规则

Knowledge Reduction and Rule Extraction in Set-valued Decision Information System

SONG Xiao-Xue[1,2] XIE Zheng-Long[2] ZHANG Wen-Xiu[1]
(Institute of Information and System Science, Xi'an Jiaotong University, Xi'an 710049)[1] (Xianyang Normal College, Xianyang 712000)[2]

Abstract In this paper we describe the incomplete information system by the set-valued information system. The two relations are defined in set-valued decision information system: the tolerance relation and the dominance relation. The rough approximation on these relations is defined, and the different properties of a set-valued information system based on these two relations are studied. Then we give the judgment theorems and the discernibility matrices of the generalized decision reduction in set-valued decision information system based on these relations, from which we provide a material approach to knowledge reduction. Finally, the optimal generalized decision rules are discussed based on these two relations.

Keywords Set-valued information system, Knowledge reduction, Discernibility matrices, Decision rule

1 引言

粗糙集理论是由波兰数学家 Pawlak 于 20 世纪 80 年代初提出的[1]。最初,粗糙集理论主要用于完备数据集的研究(即每个对象的属性值均为已知)。近年来,粗糙集已扩展处理不完备数据集(即属性值有缺失的情形)[2~7]。

具有缺失属性的不完备信息系统可以用集值信息系统来处理。本文将讨论属性值是属性域子集的不完备信息系统,对于缺失的属性值可以用该属性的所有可能取值表示,这样的信息系统称为集值信息系统。在集值信息系统上,定义了两种不同的关系:相容关系和优势关系。同时,在这两种关系上给出了集值信息系统上下近似的定义及其性质。讨论了集值决策信息系统基于两种不同关系的广义决策约简,给出了广义决策约简的判定定理及辨识矩阵,进而从集值决策信息系统中提取了最优广义决策规则。

2 集值信息系统

定义 2.1[6] 称(U,A,F)是集值信息系统,其中U是有限非空的对象集,A是有限非空属性集,$F=\{f_l : l\leqslant m\}$是U与A的关系集,$f_l : U\to\mathscr{P}(V_l)(l\leqslant m)$,$V_l$是属性$a_l$的值域,$\mathscr{P}(V_l)$表示$V_l$的非空子集全体。

例 1 表 1 是一个不完备信息系统,其属性域为:$V_{Price}=\{high,low\}$,$V_{Mileage}=\{high,low\}$,$V_{Size}=\{full,compact\}$,$V_{MaX\text{-}Speed}=\{high,low\}$。

表 1 不完备信息系统

Car	Price	Mileage	Size	Max-Speed
1	high	low	full	low
2	low	*	full	low
3	*	*	compact	low
4	high	*	full	high
5	*	*	full	high
6	low	high	full	*

表 1 所示的不完备信息系统可以转化为表 2 所示的集值信息系统。

表 2 集值信息系统

Car	Price	Mileage	Size	Max-Speed
1	{high}	{low}	{full}	{low}
2	{low}	{high,low}	{full}	{low}
3	{high, low}	{high,low}	{compact}	{low}
4	{high}	{high,low}	{full}	{high}
5	{high, low}	{high,low}	{full}	{high}
6	{low}	{high}	{full}	{high,low}

设(U,A,F)是一个集值信息系统,任意属性子集$B\subseteq A$,定义二元关系:

*)本文得到国家 973 计划项目基金(项目号:No. 2002CB312200)和咸阳师范学院专项科研基金项目(04XSYK225)资助。**宋笑雪** 副教授,博士生,主要研究方向为人工智能的数学基础;**解争龙** 副教授,主要研究方向为软件工程;**张文修** 教授,博导,主要研究方向为粗糙集、遗传算法、模糊集、人工智能等。

维普资讯 http://www.cqvip.com

$R_B^{\cap}=\{(x,y)\in U\times U\mid f_a(x)\cap f_a(y)\neq\phi(\forall a\in B)\}$

并记$[x]_B^{\cap}=\{y\in U\mid (x,y)\in R_B^{\cap}\}=\{y\in U\mid f_a(x)\cap f_a(y)\neq\phi(\forall a\in B)\}$。显然，$R_B^{\cap}$ 是自反和对称的，未必是等价关系，称之为相似关系或相容关系，$[x]_B^{\cap}$ 称为相容类。

容易证明：如果对象 x 和 y 具有 Kryszkiewicz 所定义的相容关系 $SIM(B)$[3]：$SIM(B)=\{(x,y)\in U\times U:\forall a\in B, f_a(x)=f_a(y) or f_a(x)=* or f_a(y)=*\}$，则 x 和 y 也具有关系 $R_B^{\cap}$。

集合之间的基本关系除了"相交"，还有"包含"，所以在集值信息系统中除了定义相容关系外，还可以定义对象之间的序关系。

设(U,A,F)是一个集值信息系统，任意属性子集 $B\subseteq A$，定义二元关系：

$R_B^{\supseteq}=\{(x,y)\in U\times U\mid f_a(x)\supseteq f_a(y)(\forall a\in B)\}$

并记$[x]_B^{\supseteq}=\{y\in U:(x,y)\in R_B^{\supseteq}\}=\{y\in U\mid f_a(x)\supseteq f_a(y)(\forall a\in B)\}$。显然，$R_B^{\supseteq}$ 是自反和传递的，未必是等价关系，称之为优势关系，$[x]_B^{\supseteq}$ 称为优势类。

容易证明：如果对象 x 和 y 具有 Stefanowski 所定义的相容关系 $S(x,y)$[2]：$S(x,y)=\{(x,y)\in U\times U:\forall a\in B, f_a(x)\neq *, f_a(x)=f_a(y)\}$，则 x 和 y 也具有关系 $R_B^{\supseteq}$。

定理 2.1 设(U,A,F,d)是一个集值信息系统，$B\subseteq A$，则以下性质成立：

(1)当 $B_1\subseteq B_2\subseteq A$ 时，$R_{B_1}^{\cap}\supseteq R_{B_2}^{\cap}\supseteq R_A^{\cap}$；$R_{B_1}^{\supseteq}\supseteq R_{B_2}^{\supseteq}\supseteq R_A^{\supseteq}$。

(2)当 $B_1\subseteq B_2\subseteq A$ 时，$[x]_{B_1}^{\cap}\supseteq[x]_{B_2}^{\cap}\supseteq[x]_A^{\cap}$；$[x]_{B_1}^{\supseteq}\supseteq[x]_{B_2}^{\supseteq}\supseteq[x]_A^{\supseteq}$。

(3)$I=\{[x]_B^{\cap}\mid x\in U\}$构成 U 的一个覆盖；$\mathscr{J}=\{[x]_B^{\supseteq}\mid x\in U\}$也构成 U 的一个覆盖。

(4)当 $y\in[x]_B^{\supseteq}$ 时，$[y]_B^{\supseteq}\subseteq[x]_B^{\supseteq}$；但对于$[x]_B^{\cap}$，此式不成立。

(5)$\forall(x,y)\in R_B^{\supseteq}$，有$(x,y)\in R_B^{\cap}$。

设 $B\subseteq A$，集合 $X\subseteq U$ 在相容关系 $R_B^{\cap}$ 下关于属性集 B 的上下近似定义如下：

$\underline{R_B^{\cap}}(X)=\{x\in U\mid [x]_B^{\cap}\subseteq X\}$

$\overline{R_B^{\cap}}(X)=\{x\in U\mid [x]_B^{\cap}\cap X\neq\phi\}$

集合 $X\subseteq U$ 在优势关系 $R_B^{\supseteq}$ 下关于属性集 B 的上下近似定义如下：

$\underline{R_B^{\supseteq}}(X)=\{x\in U\mid [x]_B^{\supseteq}\subseteq X\}$

$\overline{R_B^{\supseteq}}(X)=\{x\in U\mid [x]_B^{\supseteq}\cap X\neq\phi\}$

定理 2.2 设(U,A,F)是一个集值信息系统，若 $B\subseteq A$，则对于任意 $X,Y\subseteq U$，以下性质对于关系 $R_B^{\cap}$ 和 $R_B^{\supseteq}$ 都成立：

(1)$\underline{R_B}(X)\subseteq X\subseteq\overline{R_B}(X)$。

(2)$\overline{R_B}(X)=\sim\overline{R_B}(\sim X)$，$\overline{R_B}(X)=\sim\underline{R_B}(\sim X)$。

(3)$\overline{R_B}(\phi)=\phi$，$\underline{R_B}(U)=U$。

(4)$\underline{R_B}(X\cap Y)=\underline{R_B}(X)\cap\underline{R_B}(Y)$，$\overline{R_B}(X\cup Y)=\overline{R_B}(X)\cup\overline{R_B}(Y)$。

(5)$\overline{R_B}(X\cup Y)\supseteq\underline{R_B}(X)\cup\underline{R_B}(Y)$，$\overline{R_B}(X\cap Y)\subseteq\overline{R_B}(X)\cap\overline{R_B}(Y)$。

(6)$X\subseteq Y\Rightarrow\underline{R_B}(X)\subseteq\underline{R_B}(Y)$，$\overline{R_B}(X)\subseteq\overline{R_B}(Y)$。

基于关系 $R_B^{\cap}$ 和 $R_B^{\supseteq}$ 的上下近似分别满足以下性质：

(7)$\overline{R_B^{\cap}}\underline{R_B^{\cap}}(X)\subseteq X\subseteq\underline{R_B^{\cap}}\overline{R_B^{\cap}}(X)$。

(8)$\underline{R_B^{\supseteq}}(X)=\underline{R_B^{\supseteq}}(\underline{R_B^{\supseteq}}(X))$，$\overline{R_B^{\supseteq}}(X)=\overline{R_B^{\supseteq}}(\overline{R_B^{\supseteq}}(X))$。

(9)$\underline{R_B^{\cap}}(X)\subseteq\underline{R_B^{\supseteq}}$，$\overline{R_B^{\supseteq}}(X)\subseteq\overline{R_B^{\cap}}(X)$。

性质(9)说明，在关系 $R_B^{\supseteq}$ 下，关于属性集 B 的下近似和上近似优于在关系 $R_B^{\cap}$ 下关于属性集 B 的下近似和上近似。

3 集值决策信息系统和决策规则

定义 3.1[6] 称$(U,A,F,d,\{g_d\})$是集值决策信息系统(简记为(U,A,F,d))，其中 U 是对象集，A 是条件属性集，$d\notin A$ 是决策属性。$F=\{f_l:l\leqslant m\}$是 U 与 A 的关系集，$f_l:U\to\mathscr{P}_0(V_l)(l\leqslant m)$，$V_l$ 是属性 a_l 的值域，$\mathscr{P}_0(V_l)$表示的非空子集全体。$g_d:U\to V_d$ 是 U 与 d 的关系集，$V_d=\{1,2,\cdots,r\}$是 g_d 的有限值域。

定义 3.2 设(U,A,F,d)是一个集值决策信息系统，$B\subseteq A$，关于相似关系 $R_A^{\cap}$ 的广义决策函数$\partial_B^{\cap}:U\to P(V_d)$定义如下：

$\partial_B^{\cap}(x)=\{i\mid i=d(y)\quad 并且\ y\in[x]_B^{\cap}\}$

关于优势关系 $R_B^{\supseteq}$ 的广义决策函数$\partial_B^{\supseteq}:U\to P(V_d)$定义如下：

$\partial_B^{\supseteq}(x)=\{i\mid i=d(y)\quad 并且\ y\in[x]_B^{\supseteq}\}$

如果 $\mathrm{card}(\partial_B(x))=1$(即$\partial_B(x)$的基等于 1)，那么该集值决策信息系统是协调的，否则称为不协调的。

隐藏在决策信息系统中的命题知识可以表示为决策规则的形式：

$t\to s$，其中 $t=\wedge(c,v),c\in A,v\in V_c$；$s=\vee(d,w),w\in V_d$。

其中 t 和 s 分别称为规则的条件部分和决策部分。任何一个集值决策信息系统都可以看作是广义决策规则的集合。

定义 3.3 称 $B\subseteq A$ 是集值决策信息系统(U,A,F,d)关于相似关系 $R_A^{\cap}$(或优势关系 $R_A^{\supseteq}$)的广义决策约简，若满足$\partial_B^{\cap}=\partial_A^{\cap}$(或$\partial_B^{\supseteq}=\partial_A^{\supseteq}$)$(\forall x\in U)$，且对任意 $B'\subset B$，$\partial_{B'}^{\cap}\neq\partial_A^{\cap}$(或$\partial_{B'}^{\supseteq}\neq\partial_A^{\supseteq}$)。

称 $B\subseteq A$ 是集值决策信息系统(U,A,F,d)在相似关系 $R_A^{\cap}$(或优势关系 $R_A^{\supseteq}$)下关于对象 x 的广义决策约简，若满足$\partial_B^{\cap}(x)=\partial_A^{\cap}(x)$(或$\partial_B^{\supseteq}(x)=\partial_A^{\supseteq}(x)$)，且对任意 $B'\subset B$，$\partial_{B'}^{\cap}(x)\neq\partial_A^{\cap}(x)$(或$\partial_{B'}^{\supseteq}(x)\neq\partial_A^{\supseteq}(x)$)。

记 $\mathscr{D}^{\cap}=\{D^{\cap}(x_i,x_j),(x_i,x_j\in U)\}$，其中

$$D^{\cap}(x_i,x_j)=\begin{cases}\{a\in A\mid f_a(x_i)\cap f_a(x_j)=\phi\}, & d(x_j)\notin\partial_A^{\cap}(x_i),\\ \phi, & d(x_j)\in\partial_A^{\cap}(x_i)。\end{cases}$$

称 $\mathscr{D}^{\cap}$ 是集值决策信息系统(U,A,F,d)关于 $R_A^{\cap}$ 的广义决策辨识矩阵。

记 $\mathscr{D}^{\supseteq}=\{D^{\supseteq}(x_i,x_j)(x_i,x_j\in U)\}$，其中

$$D^{\supseteq}(x_i,x_j)=\begin{cases}\{a\in A\mid f_a(x_i)\supseteq f_a(x_j)\}, & d(x_j)\notin\partial_A^{\supseteq}(x_i),\\ \phi, & d(x_j)\in\partial_A^{\supseteq}(x_i)。\end{cases}$$

称 $\mathscr{D}^{\supseteq}$ 是集值决策信息系统(U,A,F,d)关于 $R_A^{\supseteq}$ 的广义决策辨识矩阵。

定理 3.1 设(U,A,F,d)是集值决策信息系统，$B\subseteq A$，则

$B\cap D^{\cap}(x_i,x_j)\neq\phi(D^{\cap}(x_i,x_j)\neq\phi)$当且仅当$\partial_B^{\cap}=\partial_A^{\cap}$。

证明：因为对任意 $B\subseteq A$，$[x]_B^{\cap}\supseteq[x]_A^{\cap}$，所以$\partial_B^{\cap}(x)\supseteq\partial_A^{\cap}(x)$。

因此我们只需要证明

$B\cap D^{\cap}(x_i,x_j)\neq\phi(D^{\cap}(x_i,x_j)\neq\varphi)\Leftrightarrow\partial_B^{\cap}\subseteq\partial_{\cap A}$。

$(\Rightarrow)$设 $B\cap D^{\cap}(x_i,x_j)\neq\phi(D^{\cap}(x_i,x_j)\neq\phi)$。

假设对任意 $x'_i\in[x_i]_B^{\cap}$ 且 $x'_i\notin[x_i]_A^{\cap}$，有 $d(x'_i)\notin\partial_A^{\cap}(x_i)$，则 $D^{\cap}(x_i,x'_i)\neq\phi$，所以 $B\cap D^{\cap}(x_i,x'_i)\neq\phi$。然而，由

维普资讯 http://www.cqvip.com

$x'_i \in [x_i]_B^{\cap}$，可以得到$(x_i, x'_i) \in R_B^{\cap}$，也即 $f_b(x_i) \cap f_b(x'_i) \neq \phi(\forall b \in B)$，则 $B \cap D^{\cap}(x_i, x'_i) = \phi$。矛盾。

所以当 $B \cap D^{\cap}(x_i, x_j) \neq \phi(D^{\cap}(x_i, x_j) \neq \phi$，如果 $x'_i \in [x_i]_B^{\cap}$ 且 $x'_i \notin [x_i]_A^{\cap}$，则 $d(x'_i) \in \partial_{\bar{A}}^{\cap}(x_i)$。因此$\partial_B^{\cap}(x_i) \subseteq \partial_A^{\cap}(x_i)$。

($\Leftarrow$)设$\partial_B^{\cap} \subseteq \partial_A^{\cap}$，如果 $d(x_j) \notin \partial_A^{\cap}(x_i)$，则 $d(x_j) \notin \partial_B^{\cap}(x_i)$，即 $x_j \notin [x_i]_B^{\cap}$。因此存在 $b \in B$ 使得$(x_i, x_j) \notin R_b^{\cap}$，则 $b \in D^{\cap}(x_i, x_j)$。所以 $B \cap D^{\cap}(x_i, x'_i) = \phi(D^{\cap}(x_i, x_j) \neq \phi)$。

定理 3.2 设(U, A, F, d)是集值决策信息系统，$B \subseteq A$，则 $B \cap D^{\supseteq}(x_i, x_j) \neq \phi(D^{\supseteq}(x_i, x_j) \neq \phi)$当且仅当$\partial_B^{\supseteq} = \partial_{\bar{A}}^{\supseteq}$。

证明：类似于定理 3.1 的证明。

定理 3.3 设(U, A, F, d)是集值决策信息系统，如果 $D^{\cap}(x_i, x_j) \neq \phi$，则 $D^{\cap}(x_i, x_j) \subseteq D^{\supseteq}(x_i, x_j)$。

证明：由$\partial_{\bar{A}}^{\supseteq}(x) \subseteq \partial_A^{\cap}(x)(\forall x \in U)$即得。

这说明基于 $R_A^{\cap}$ 的广义决策辨识矩阵中的非空元素是基于$R_{\bar{A}}^{\supseteq}$ 的广义决策辨识矩阵中的相应元素的子集。但是由关系 $R_A^{\cap}$ 得到的最优广义决策规则与由关系 $R_{\bar{A}}^{\supseteq}$ 得到的最优广义决策规则并无关联。

例 2 表 3 所示是由表 1 扩充的集值决策信息系统。

表 3 集值决策信息系统

Car	Price	Mileage	Size	Max-Speed	d
1	{high}	{low}	{full}	{low}	good
2	{low}	{high, low}	{full}	{low}	good
3	{high, low}	{high, low}	{compact}	{low}	poor
4	{high}	{high, low}	{full}	{high}	good
5	{high, low}	{high, low}	{full}	{high}	excel
6	{low}	{high}	{full}	{high, low}	excel

由定义可以得到

$[1]_A^{\cap} = \{1\}, [2]_A^{\cap} = \{2,6\}, [3]_A^{\cap} = \{3\}, [4]_A^{\cap} = \{4,5\}, [5]_A^{\cap} = \{4,5,6\}, [6]_A^{\cap} = \{2,5,6\}$；

$[1]_{\bar{A}}^{\supseteq} = \{1\}, [2]_{\bar{A}}^{\supseteq} = \{2\}, [3]_{\bar{A}}^{\supseteq} = \{3\}, [4]_{\bar{A}}^{\supseteq} = \{4\}, [5]_{\bar{A}}^{\supseteq} = \{4,5\}, [6]_{\bar{A}}^{\supseteq} = \{6\}$。

表 3 基于关系 $R_A^{\cap}$ 和关系 $R_{\bar{A}}^{\supseteq}$ 的广义决策函数值如表 4 所示。

表 4 集值决策信息系统的广义决策函数值

Car	d	$\partial_A^{\cap}$	$\partial_{\bar{A}}^{\supseteq}$
1	good	{good}	{good}
2	good	{good, excel}	{good}
3	poor	{poor}	{poor}
4	good	{good, excel}	{good}
5	excel	{good, excel}	{good, excel}
6	excel	{good, excel}	{good, excel}

该集值决策信息系统在关系 $R_A^{\cap}$ 下的广义决策辨识矩阵为(其中 P, M, S, X 分别表示 Price, Mileage, Size, Max-Speed)：

$$\mathscr{D}^{\cap} = \begin{pmatrix} \phi & \phi & S & \phi & X & P,M \\ \phi & \phi & S & \phi & \phi & \phi \\ S & S & \phi & S,X & S,X & S \\ \phi & \phi & S,X & \phi & \phi & \phi \\ \phi & \phi & S,X & \phi & \phi & \phi \\ \phi & \phi & S & \phi & \phi & \phi \end{pmatrix}$$

因此 $B_1 = \{\text{Pr } ice, Size, MaX\text{-}Speed\}$，$B_2 = \{Mileage, Seze, MaX\text{-}Speed\}$是表 3 所示集值决策信息系统关于相似关系 $R_A^{\cap}$ 的广义决策约简。$\{\text{Pr } ice, Size, MaX\text{-}Speed\}$与$\{Mileage, Size, MaX\text{-}Speed\}$是对象 1 的广义决策约简，$\{Size\}$是对象 2 和对象 3 的广义决策约简等等。

由表 3 可得，对象 1 支持的最优决策规则有两个：

$(\text{Pr } ice, high) \wedge (Size, full) \wedge (MaX\text{-}Speed, low) \rightarrow (d, good)$，

$(Mileage, low) \wedge (Size, full) \wedge (MaX\text{-}Speed, low) \rightarrow (d, good)$；

对象 2 支持的最优决策规则有一个：

$(Size, full) \rightarrow (d, good) \vee (d, excel)$；

对象 3 支持的最优决策规则有一个：

$(Size, compact) \rightarrow (d, poor)$；

对象 4 和对象 5 支持的最优决策规则有两个：

$(Size, full) \rightarrow (d, good) \vee (d, excel)$

$(MaX\text{-}Speed, high) \rightarrow (d, good) \vee (d, excel)$；

对象 6 支持的最优决策规则有一个：

$(Size, full) \rightarrow (d, good) \vee (d, excel)$；

综合起来，表 3 所示集值决策信息系统关于相似关系 $R_A^{\cap}$ 可推出如下最优决策规则：

$r1$：$(\text{Pr } ice, hgih) \wedge (Size, full) \wedge (MaX\text{-}Speed, low) \rightarrow (d, good)$，//由对象 1 生成

$r2$：$(Mileage, low) \wedge (Size, full) \wedge (MaX\text{-}Speed, low) \rightarrow (d, good)$，//由对象 1 生成

$r3$：$(Size, full) \rightarrow (d, good) \vee (d, excel)$，//由对象 2,4,5,6 生成

$r4$：$(Size, compact) \rightarrow (d, poor)$，//由对象 3 生成

$r5$，$(MaX\text{-}Speed, high) \rightarrow (d, good) \vee (d, excel)$，//由对象 4,5 生成

该集值决策信息系统在关系 $R_{\bar{A}}^{\supseteq}$ 下的广义决策辨识矩阵为

$$\mathscr{D}^{\supseteq} = \begin{pmatrix} \phi & \phi & P,M,S & \phi & P,M,X & P,M,X \\ \phi & \phi & P,S & \phi & P,X & X \\ S & S & \phi & S,X & S,X & S,X \\ \phi & \phi & P,S,X & \phi & P & P,X \\ \phi & \phi & S,X & \phi & \phi & \phi \\ P,M & M & P,M,S & P,M & \phi & \phi \end{pmatrix}$$

因此 $B' = \{\text{Pr } ice, Mileage, Size, MaX\text{-}Speed\}$是表 3 所示集值决策信息系统关于优势关系 $R_{\bar{A}}^{\supseteq}$ 的广义决策约简。$\{\text{Pr } ice\}$、$\{Mileage\}$和$\{Size, MaX\text{-}Speed\}$是对象 1 的广义决策约简，$\{MaX\text{-}Speed\}$是对象 2 的广义决策约简等等。

表 3 所示集值决策信息系统关于优势关系 $R_{\bar{A}}^{\supseteq}$ 可推出如下最优决策规则：

$r1$，$(\text{Pr } ice, high) \rightarrow (d, good)$，//由对象 1 和对象 4 生成

$r2$，$(Mileage, low) \rightarrow (d, good)$，//由对象 1 生成

$r3$，$(Size, full) \wedge (MaX\text{-}Speed, low) \rightarrow (d, good)$，//由对象 1 生成

$r4$，$(MaX\text{-}Speed, low) \rightarrow (d, good)$，//由对象 2 生成

$r5$，$(Size, compact) \rightarrow (d, poor)$，//由对象 3 生成

$r6$，$(Size, full) \rightarrow (d, good) \vee (d, excel)$，//由对象 5 生成

$r7$，$(MaX\text{-}Speed, high) \rightarrow (d, good) \vee (d, excel)$，//由对象 5 生成

$r8$，$(Mileage, high) \rightarrow (d, excel)$，//由对象 6 生成

(下转第 191 页)

维普资讯 http://www.cqvip.com

确识别率。可见，选用了新的柔性可信度函数后，对提取出的规则集而言，最大识别率要优于使用阶梯函数的效果。

5.3 lenses 数据集

lenses 数据集是完备、无冲突、无噪声的，其中有 24 个样本，每个样本包含 4 个条件属性和 1 个决策属性。决策属性有 3 个决策类，且这 3 个决策类是不等量划分的，分别占样本总数的 16.67%，20.83%和 62.5%。在可信度阈值 $Tr=0.95$ 和陡度因子 $a=25$ 的条件下，得到 9 条规则，使得样本的识别率为 100%，与标准结果相同。这说明，本文的规则度量函数可以很好地处理决策类不等量划分的情况。

结论 本文使用免疫算法作为分类规则提取的工具，结合小生境技术，提出了柔性可信度的概念和新的规则重要性度量方法，将其作为免疫算法中抗体亲和力的度量公式进行了仿真实验。实验结果表明，无论是处理无噪声的数据，还是有噪声或决策类不均衡的数据，本文的方法都可以取得较好的效果。

参 考 文 献

1 Khoo L P, Zhai L Y. A prototype genetic algorithm-enhanced rough set-based rule induction system. Computers in Industry, 2001, 46(1): 95～106

2 He M, et al. Genetic algorithm based multi-knowledge extraction method for rough set. Mini-Micro System, 2005, 26(4): 651～654

3 Freitas A A. On rule interestingness measure. Knowledge-Based System, 1999, 12(4): 309～315

4 黄席樾，等. 现代智能算法理论及应用. 北京：科学出版社，2005

5 Ghosh A, Nath B. Multi-objective rule mining using genetic algorithms. Information Sciences, 2004, 163(1-3): 123～133

6 Wu Z J, et al. A Chromosome-based Evaluation Model for Computer Defense Immune Systems. In: Proceedings of 2003 IEEE Congress on Evolutionary Computation, 2003. 1363～1369

7 Zheng J G, Wang L, Jiao L C. An immune algorithm for generalized rule induction. In: Proceedings of CIE International Conference of Radar Proceedings, 2001. 1023～1026

8 Goldberg D E, Richardson J J. Genetic algorithms with sharing for multimodal function optimization. Genetic algorithms and their application. In: Proceedings of the Second International Conference of Genetic Algorithms, 1987. 41～49

（上接第 184 页）

由表 3 所示集值决策信息系统在关系 $R_A^{\cap}$ 和 $R_A^{\supseteq}$ 下的广义决策辨识矩阵，可以得到基于 $R_A^{\cap}$ 的广义决策辨识矩阵中的非空元素是基于 $R_A^{\supseteq}$ 的广义决策辨识矩阵中的相应元素的子集。但是由关系 $R_A^{\cap}$ 得到的最优广义决策规则与由关系 $R_A^{\supseteq}$ 得到的最优广义决策规则并无关联。

结论 通常我们所处理的信息表中属性的值是单一的并且是确定的，但这只是一种理想状态。在实际情况中，我们所得到的数据往往既不单一也不确定，可能缺失部分数据，可能是数据集的一个子集。本文用集值信息系统来处理数据不完全的信息系统。集值信息系统是处理不完备信息系统的一种方法。本文在集值信息系统上定义了两种关系：相容关系和优势关系。同时，给出了集值信息系统在这两种关系下的上下近似的定义及其性质。在关系 $R_A^{\supseteq}$ 下关于属性集 B 的下近似和上近似优于在关系 $R_A^{\cap}$ 下关于属性集 B 的下近似和上近似。讨论了集值决策信息系统基于两种不同关系的广义决策约简，给出了广义决策约简的判定定理及辨识矩阵，进而从集值决策信息系统中提取了最优广义决策规则。

参 考 文 献

1 Pawlak Z. Rough Sets: Theoretical Aspects of Reasoning About Data. Kluwer Academic Publishers, Boston, 1991

2 Stefanowski J, Tsoukias A. On the Extension of Rough sets under Incomplete Information. In: Zhong N, et al. eds. RSFDGrC 1999, LNAI 1711, 1999. 73～82

3 Kryszkiewicz M. Rule in incomplete information systems. Information Science, 1999, 113: 271～292

4 Grzymala-Busse J W. Incomplete Data and Generalization of Indiscernibility Relation, Definability, and Approximations. In: Slezak, et al. eds. RSFDGrC 2005, LNAI 3641, 2005. 244～253

5 Slowinski R, Vanderpooten D. A Generalized Definition of Rough Approximations Based on Similarity. IEEE Transactions on Knowledge and Data Engineering, 2000, 12(2): 331～336

6 张文修，梁怡，吴伟志. 信息系统与知识发现. 北京：科学出版社，2003

7 王国胤. Rough 集理论与知识获取. 西安：西安交通大学出版社，2001

（上接第 187 页）

小结 本文探讨了粗糙集理论中最小决策规则的求取方法，提出决策依赖度的定义，尝试从最短的条件属性组合中提取尽可能多的决策规则。只有现有长度的决策规则无法完全覆盖所有样本时，才会考虑增加决策规则的长度。由于条件属性的所有组合是条件属性数的幂集，本文提出了 3 种减少计算复杂性的方案：1）引入跳跃系数 λ，先计算较长的条件属性组合的决策依赖度 β。如果 β 为 0，则表明该属性组合的所有子集均不必再计算；如果 β 不为 0，再回溯到其较短的子集进行计算。如此跳跃式计算，在属性较多的情况下可以大大减少计算复杂性。而回溯功能则可以避免决策树学习方法经常会遇到的收敛到局部最优答案的情况。2）在计算中只对具有相同决策值的样本进行等价类划分，这样处理的好处是避免了对含有不同决策值的等价类的无用划分。3）设计 Remain 集合，只针对 Remain 中的样本进行等价类的划分。随着 Remain 中样本数的减少，计算量会大幅下降。实验表明，本文所提出的基于决策依赖度的跳跃式决策规则求取方法可以直接应用于不完备信息系统，因此具有良好的实用价值。

参 考 文 献

1 曾华军，张银奎译. 机器学习. 北京：机械工业出版社，2003

2 文硕频，乔胜勇，陈彩云，等. 基于决策树的不完全决策表的数据及规则提取. 计算机应用，2003，23(11)：17～22

3 孙长嵩，董西国，张健沛. 一个基于粗糙集和决策树的最简分类规则集生成算法. 哈尔滨工程大学学报，2002，23(5)：87～91

4 Yang Jing, Wang Hao, Hu Xue-Gang, et al. A new classification algorithm based on rough set and entropy. International Conference on Machine Learning and Cybernetics, 2003, 1: 364～367

5 Quinlan J R. C4. 5: Programs for machine learning. San Mateo, CA: Morgan Kaufmann

6 Witten I H, MacDonald B A. Using concept learning for knowledge acquisition. International Journal of Man-Machine Studies, 1988, 27: 349～370

7 Mitchell T M, Utgoff P E, Banerji R. Learning by experimentation: acquiring and modifying problem-solving heuristics. Machine Learning, 1: 163～190

8 王东锴，梁梁. 不完全决策表中的决策规则发现. 计算机科学，2001，28(5. 专刊)：83～85

9 文硕频，乔胜勇，陈彩云，等. 基于决策树的不完全决策表的数据及规则提取. 计算机应用，2003，23(11)：17～22