第40卷 第6期 华侨大学学报(自然科学版) Vol. 40 No. 6
2019年11月 Journal of Huaqiao University (Natural Science) Nov. 2019

DOI:10.11830/ISSN.1000-5013.201810106

决策信息系统协调性的关系矩阵表示

陈应生[1],李进金[1,2]

(1. 华侨大学 数学科学学院,福建 泉州 362021;
2. 闽南师范大学 数学科学与统计学院,福建 漳州 363000)

摘要: 从矩阵视角研究不完备信息系统的知识表示和属性约简.首先,引入关系矩阵,基于容差关系提出不完备决策信息系统协调性判定方法.其次,利用关系矩阵求正域,并提出属性协调集的矩阵判定定理,从而给出一种保持正域不变的约简的新方法.最后,基于上述框架,给出属性重要性度,进一步提出一种属性约简的启发式算法,并通过分析和实例证明该方法的有效性.

关键词: 不完备决策信息系统;关系矩阵;容差关系;正域

中图分类号: TP 18 **文献标志码:** A **文章编号:** 1000-5013(2019)06-0823-07

Relation Matrix Representation for Consistent of Decision Information System

CHEN Yingsheng[1], LI Jinjin[1,2]

(1. School of Mathematical Sciences, Huaqiao University, Quanzhou 362021, China;
2. School of Mathematics and Statistics, Minnan Normal University, Zhangzhou 363000, China)

Abstract: This paper studies the knowledge representation and attribute reduction of incomplete information systems from the perspective of matrix. Firstly, the relation matrix is introduced, and based on tolerance relation, a method for judging the consistent of incomplete decision information system is proposed. Secondly, the positive domain is obtained by using the relation matrix, and the matrix decision theorem of the attribute consistent set is proposed, and a new reduction method keeping the positive domain invariant is given. Finally, based on the above framework, an attribute importance significance is given, a heuristic algorithm for attribute reduction is proposed, and the effectiveness of the method is proved by analysis.

Keywords: incomplete decision information system; matrix of relation; tolerance relation; positive domain

1982年,波兰数学家Pawlak提出经典粗糙集理论[1],为处理信息系统中模糊、不确定的知识提供一种有效的数学工具,该理论被广泛应用于人工智能、模式识别、机器学习、数据挖掘、管理决策等领域[2-5].经典粗糙集以等价关系诱导的等价类为基本粒,这限制了粗糙集的应用范围.因此,为了适应现实中不同数据类型、不同知识类型的研究需要,文献[6-9]提出多种拓展模型.为了处理知识的模糊性,文献[7-8]提出模糊粗糙集,文献[9]提出覆盖粗糙集.在实际问题中,数据往往具有不完全、缺失等特点,因此,将信息系统分为完备信息系统与不完备信息系统.由于信息系统每个属性对应一个二元关系,通过在二元关系上定义一个关系矩阵,论域中对象的关系情况便会通过一个布尔矩阵进行直观的描述.因此,布尔矩阵就成为研究信息系统的一种有效工具[10-14].在粗糙集领域,王磊等[10]基于矩阵视角揭示知识粒度与等价关系矩阵间的关系,进一步探讨属性增删时知识粒度的变化规律及属性约简方法.林艺

收稿日期: 2018-10-31
通信作者: 陈应生(1976-),男,讲师,主要从事粗糙集和粒计算的研究.E-mail:cyssheng@hqu.edu.cn.
基金项目: 国家自然科学基金资助项目(11871259,61379021,11701258)

东等[11]研究覆盖族动态变化时知识近似的矩阵计算方法. Tan 等[12]定义新的矩阵运算，提出一种基于覆盖信息系统的知识提取的快速方法. Ma[14]研究 32 对覆盖近似算子的矩阵表示. 本文将对关系矩阵在不完备信息系统中的应用进行研究.

1 基础知识

定义 1[2] 定义信息系统为 $S=(U,A,V)$，其中，$U=\{x_1,x_2,\cdots,x_n\}$是一个非空有限集合，$A=\{a_1,a_2,\cdots,a_m\}$是非空有限属性集. 对于任意 $a\in A$，满足 $a:U\to V_a$，即 $a(x)\in V_a$，其中，$V_a=\{a(x)\mid x\in U\}$是属性 a 的值域，$V=\bigcup_{a\in A}V_a$，通常把信息系统简记为 $S=(U,A)$.

若 $\forall a\in A$，$\forall x\in U$，$a(x)$是一个单点集，称 S 是一个完备的信息系统；若 $\exists a\in A$，$\exists x\in U$，$a(x)$不是一个单点集，称 S 是一个不完备的信息系统. 在不完备的信息系统中，若 x 对应属性 a 的属性值缺失或者不唯一，常将其属性值看成 V_l，也有把属性值缺失或者唯一的情况给定一个空值，用 $*$ 表示.

定义 2[2] 如果 $S=(U,A)$是不完备信息系统，对于 $a\in A$，$x\in U$，U 上的容差关系 R_a 和 x 的容差类分别定义为

$$R_a=\{(x,y)\in U\times U\mid a(x)=a(y)\vee a(x)=*\vee a(y)=*\},$$

$$R_a(x)=\{y\in U\mid (x,y)\in R_a\}.$$

对于 $B\subseteq A$，$R_B=\bigcap_{a\in B}R_a$，显然，$\{R_B(x)\mid x\in U\}$是 U 的一个覆盖. R_B 具有自反性和对称性，但不具有传递性. 对于 $B\subseteq A$，$X\subseteq U$，定义 X 的上下近似集分别为

$$\overline{R}_B(X)=\{x\in U\mid R_B(x)\cap X\neq\varnothing\},$$

$$\underline{R}_B(X)=\{x\in U\mid R_B(x)\subseteq X\}.$$

定义 3[12] 对于不完备的信息系统 $S=(U,A)$，$B\subseteq A$，定义容差关系 R_B 对应的关系矩阵为 $\boldsymbol{M}_B=(r_{i,j})_{n\times n}$，其中，$r_{i,j}=\begin{cases}1,(x_i,x_j)\in R_B,\\0,(x_i,x_j)\notin R_B.\end{cases}$

定义 4[12] 设 $U=\{x_1,x_2,\cdots,x_n\}$，定义 $X\subseteq U$ 的特征向量为 $\boldsymbol{f}(X)=(f_1,f_2,\cdots,f_n)$，其中，$f_i=\begin{cases}1,x_i\in X,\\0,x_i\notin X.\end{cases}$

对于不完备信息系统 $S=(U,A)$，$B\subseteq A$，关系矩阵 $\boldsymbol{M}_B$ 体现了 B 上的所有关系类，即若 $r_{i,j}=1$，则 $y_j\in R_B(x_i)$，所有 x_i 的关系类 $R_B(x_i)$就是关系矩阵 $\boldsymbol{M}_B$ 第 i 行上所有 1 对应列的元素全体，所以由 $\boldsymbol{M}_B$ 的每一行，即可得出每个元素的关系类. 由于容差关系 R_B 是对称的，所以 $\boldsymbol{M}_B$ 也是对称的，由 $\boldsymbol{M}_B$ 的每一列也可得出每个元素的关系类.

$X\subseteq U$ 的特征向量 $\boldsymbol{f}(X)$上 1 对应的元素就属于 X，所以由 $\boldsymbol{f}(X)$可以得出集合 X.

例 1 不完备信息系统，如表 1 所示. 不完备信息系统所确定的容差关系的关系矩阵为

表 1 不完备信息系统

Tab. 1 Incomplete information system

U	A			
	a_1	a_2	a_3	a_4
x_1	0	0	1	*
x_2	1	2	*	0
x_3	0	*	1	*
x_4	*	2	2	2
x_5	2	1	1	1
x_6	2	1	2	*

$$\boldsymbol{M}_A=\begin{pmatrix}1&0&1&0&0&0\\0&1&0&1&0&0\\1&0&1&0&0&0\\0&1&0&1&0&0\\0&0&0&0&1&0\\0&0&0&0&0&1\end{pmatrix}.$$

定义 5 设 $\boldsymbol{M}_1=(r_{i,j})_{n\times n}$，$\boldsymbol{M}_2=(r'_{i,j})_{n\times n}$均为关系矩阵，定义如下运算：

1) $\boldsymbol{M}_1\wedge\boldsymbol{M}_2=(r_{i,j}\times r'_{i,j})_{n\times n}$；

2) $\boldsymbol{M}_1^{\mathrm{C}}=(1-r_{i,j})_{n\times n}$；

3) 若 $\boldsymbol{M}_1\wedge\boldsymbol{M}_2=\boldsymbol{M}_1$，则称 $\boldsymbol{M}_1\preceq\boldsymbol{M}_2$.

定义 6　设 $\boldsymbol{M}=(m_{i,j})_{n\times n}, m_{i,j}\geqslant 0, i,j=1,2,\cdots,n$，定义 $\boldsymbol{M}=(r_{i,j})_{n\times n}, r_{i,j}=\begin{cases}1, m_{i,j}\geqslant 1,\\ 0, m_{i,j}<1.\end{cases}$

例 2　如 $\boldsymbol{P}=\begin{pmatrix}1 & 2\\ 0 & 3\end{pmatrix}$，则 $(\approx\boldsymbol{P})=\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}$.

定义 7　对于不完备信息系统 $S=(U,A)$，有

1) 若 $B\subseteq A$，使得 $R_B(x)=R_A(x), \forall x\in U$，称 B 是信息系统 S 的一个协调集；

2) 若 B 是 A 的一个协调集，但对 B 的任何真子集都不是 A 的协调集，称 B 是 S 的一个约简集；

3) 若 $a\in A$，满足 $R_{A-\{a\}}(x)=R_A(x), \forall x\in U$，则称 a 是不必要属性，否则，称为必要属性；

4) 系统 S 的所有必要属性组成的集合称为 S 的核，记为 $\mathrm{Core}(A)$.

由定义 4，5 直接可得定理 8.

命题 1　设论域 $U=\{x_1,x_2,\cdots,x_n\}$，$X,Y\subseteq U$，则有 $Y\subseteq X\Leftrightarrow \boldsymbol{f}(Y)\preceq\boldsymbol{f}(X)\Leftrightarrow\boldsymbol{f}(Y)\wedge(\boldsymbol{f}(X))^C=0$.

命题 2　对于不完备信息系统 $S=(U,A)$，$B\subseteq A$，其容差关系 R_B 所对应的关系矩阵为 $\boldsymbol{M}_B$，因此有 $\boldsymbol{f}(R_B(x_i))=(r_{i,1},r_{i,2},\cdots,r_{i,n})$.

例 3　由例 1 对应的信息系统，有 $\boldsymbol{f}(R_A(x_1))=(1,0,1,0,0,0)$，$\boldsymbol{f}(R_A(x_6))=(0,0,0,0,0,1)$.

命题 3　对于不完备信息系统 $S=(U,A)$，$B,C\subseteq A$，则有以下 2 个结论成立.

1) $R_B\subseteq R_C$ 的充要条件是 $\boldsymbol{M}_B\preceq\boldsymbol{M}_C$.

2) $R_B=R_C$ 的充要条件是 $\boldsymbol{M}_B=\boldsymbol{M}_C$.

证明：1) 因为 $R_B\subseteq R_C\Leftrightarrow\forall x_i\in U$，所以有 $R_B(x_i)\subseteq R_C(x_i)\Leftrightarrow\forall x_i\in U, \boldsymbol{f}(R_B(x_i))\preceq\boldsymbol{f}(R_C(x_i))\Leftrightarrow \boldsymbol{M}_B\preceq\boldsymbol{M}_C$.

命题 4　对于不完备信息系统 $S=(U,A)$，$B,C\subseteq A$，则有

1) $\boldsymbol{M}_{B\cup C}=\boldsymbol{M}_B\wedge\boldsymbol{M}_C$；

2) $B\subseteq A, B=(a_1,a_2,\cdots,a_s)\boldsymbol{M}_B=\boldsymbol{M}_{a_1}\wedge\boldsymbol{M}_{a_2}\wedge\cdots\wedge\boldsymbol{M}_{a_s}$；

3) $C\subseteq B\Rightarrow\boldsymbol{M}_B\preceq\boldsymbol{M}_C$；

4) $a\in A$ 是不必要属性 $\Leftrightarrow\boldsymbol{M}_A=\boldsymbol{M}_{A-\{a\}}$；

5) $B\subseteq A$，则 B 是 A 的协调集 $\Leftrightarrow\boldsymbol{M}_B=\boldsymbol{M}_A$；

6) 若 $B\subseteq A, \boldsymbol{M}_B=\boldsymbol{M}_A$，而对任意 $C\subset B, \boldsymbol{M}_C\neq\boldsymbol{M}_A$，则 B 是 A 的约简集.

2　利用关系矩阵讨论决策信息系统的协调性

定义 8[2]　设 $S=(U,A\cup\{d\})$ 为不完备决策信息系统，其中，(U,A) 是不完备信息系统，这里 $U=\{x_1,x_2,\cdots,x_n\}$ 是一个非空有限集合，称为论域，$A=\{a_1,a_2,\cdots,a_m\}$ 是非空有限属性集，称为条件属性集，$d\notin A$ 称为决策属性，$d:U\to V_d$ 是一个单值映射，对任意的 $x\in U$，$d(x)\in V_d$，$V_d=\{d(x)\mid x\in U\}$ 是属性 d 的值域.

由 d 诱导的等价关系为 $R_d=\{(x,y)\in U\times U\mid d(x)=d(y)\}$. $[x]_d=\{y\in U\mid(x,y)\in R_d\}$ 为 x 的等价类，$U/R_d=\{R_d(x)\mid x\in U\}$ 为 R_d 的商集. 对不完备决策信息系统 $S=(U,A\cup\{d\})$，有

1) 若 $R_A\subseteq R_d$，称系统 S 是协调的，否则，称 S 为不协调的；

2) 若系统 S 是协调的，$B\subseteq A$，满足 $R_B\subseteq R_d$，称 B 是决策协调集，若 B 是决策协调集，但对任意 B 的真子集都不是决策协调集，称 B 是决策约简集；

3) 对于 $B\subseteq A$，$\mathrm{Pos}_B(d)=\bigcup_{U/R_d}\{\underline{R}_B(D)\}$ 称为条件属性子集 B 在决策属性 d 上的正域，假如存在 $B\subseteq A$，使得 $\mathrm{Pos}_B(d)=\mathrm{Pos}_A(d)$，称 B 为系统 S 的保正域决策协调集，若 B 为保正域决策协调集，而 B 的任意真子集都不是系统 S 的保正域决策协调集，称 B 为系统 S 的保正域决策约简集.

显然，$S=(U,A\cup\{d\})$ 协调当且仅当 $\mathrm{Pos}_A(d)=U$.

2.1　协调的决策信息系统

定理 1　设 $S=(U,A\cup\{d\})$ 为不完备决策信息系统，则 S 是协调的 $\Leftrightarrow R_A\subseteq R_d\Leftrightarrow\boldsymbol{M}_A\preceq\boldsymbol{M}_d\Leftrightarrow\boldsymbol{M}_A\wedge(\boldsymbol{M}_d)^C=0$.

定理 2 设 $S=(U,A\cup\{d\})$是不完备决策信息系统,有

1) 若 S 是协调的,B 是决策协调集$\Leftrightarrow R_B\subseteq R_d\Leftrightarrow \boldsymbol{M}_B\preceq\boldsymbol{M}_d$;

2) 若 S 是协调的,$B\subseteq A$ 是决策约简集$\Leftrightarrow\boldsymbol{M}_B\preceq\boldsymbol{M}_d$ 成立,但对 B 的任意真子集 C,$\boldsymbol{M}_C\preceq\boldsymbol{M}_d$ 不成立;

3) $a\in A$ 是不必要属性$\Leftrightarrow\boldsymbol{M}_{A-\{a\}}\preceq\boldsymbol{M}_d$,$a$ 是核心属性$\Leftrightarrow$ $\boldsymbol{M}_{A-\{a\}}\preceq\boldsymbol{M}_d$ 不成立.

定理 3 设 $S=(U,A\cup\{d\})$是不完备决策信息系统,A 的所有核心属性构成的集合称为 A 的核,记为 $\mathrm{Core}(A)$,对于 $a\in A$,若有 $\boldsymbol{M}_{\mathrm{Core}(A)}\preceq\boldsymbol{M}_a$,则 a 是绝对不必要属性.

证明:由于 $\boldsymbol{M}_{\mathrm{Core}(A)}\preceq\boldsymbol{M}_a$,所以 $\forall x\in U$,有 $R_{\mathrm{Core}(A)}(x)\subseteq R_a(x)$. 又对 $\forall A$ 的约简集 B,有 $\mathrm{Core}(A)\subseteq B\subseteq A$,$R_B(x)\subseteq R_{\mathrm{Core}(A)}(x)\subseteq R_a(x)$,故 a 为绝对不必要属性.

例 4 利用关系矩阵判定不完备决策信息系统(表 2)的协调性,有

表 2 例 4 的不完备决策信息系统

Tab. 2 Incomplete decision information system of example 4

U	A				
	a_1	a_2	a_3	a_4	d
x_1	0	0	1	*	*
x_2	1	2	*	*	1
x_3	0	*	1	*	0
x_4	*	2	2	2	1
x_5	2	1	1	1	2
x_6	2	1	2	*	2

$$\boldsymbol{M}_A=\begin{pmatrix}1&0&1&0&0&0\\0&1&0&0&0&0\\1&0&1&0&0&0\\0&0&0&1&0&0\\0&0&0&0&1&0\\0&0&0&0&0&1\end{pmatrix},\quad \boldsymbol{M}_d=\begin{pmatrix}1&0&1&0&0&0\\0&1&0&1&0&0\\1&0&1&0&0&0\\0&1&0&1&0&0\\0&0&0&0&1&1\\0&0&0&0&1&1\end{pmatrix},\quad \boldsymbol{M}_B=\begin{pmatrix}1&0&1&0&0&0\\0&1&0&1&0&0\\1&0&1&0&0&0\\0&1&0&1&0&0\\0&0&0&0&1&0\\0&0&0&0&0&1\end{pmatrix}.$$

因 $\boldsymbol{M}_A\preceq\boldsymbol{M}_d$,故不完备决策信息系统是协调的,又 $\boldsymbol{M}_B\preceq\boldsymbol{M}_d$,则 $B=\{a_1,a_2,a_3\}$是一个决策协调集. 经计算,$\boldsymbol{M}_{a_1,a_2}\preceq\boldsymbol{M}_d$,$\boldsymbol{M}_{a_1,a_3}\preceq\boldsymbol{M}_d$,$\boldsymbol{M}_{a_1,a_4}\preceq\boldsymbol{M}_d$,$\boldsymbol{M}_{a_2,a_4}\preceq\boldsymbol{M}_d$,$\boldsymbol{M}_{a_3,a_4}\preceq\boldsymbol{M}_d$,$\boldsymbol{M}_{a_2,a_3}\preceq\boldsymbol{M}_d$ 皆不成立,且 $\boldsymbol{M}_{a_1,a_2,a_4}\preceq\boldsymbol{M}_d$,$\boldsymbol{M}_{a_1,a_3,a_4}\preceq\boldsymbol{M}_d$,$\boldsymbol{M}_{a_2,a_3,a_4}\preceq\boldsymbol{M}_d$ 也都不成立,故 $B=\{a_1,a_2,a_3\}$是一个决策约简集,a_4 是不必要属性.

2.2 不协调决策信息系统

定理 4 设 $S=(U,A\cup\{d\})$是不完备决策信息系统,$B\subseteq A$,记 $\boldsymbol{K}_B=\boldsymbol{M}_B\wedge(\boldsymbol{M}_d)^C$,$\boldsymbol{K}_B$ 的第 i 行的和记为 $K_B(i)(i=1,2,\cdots,n)$,则有

$$\mathrm{Pos}_B(d)=\bigcup\{x_i\in U\mid K_B(i)=0\}.$$

证明:因 $\mathrm{Pos}_B(d)=\bigcup_{D\in U/R_d}\{\underline{R}_B(D)\}=\bigcup\{x_i\in U\mid R_B(x_i)\subseteq R_d(x_i)\}$,若 $R_B(x_i)\subseteq D_j$,因 $R_B(x_i)\cap R_d(x_i)\neq\varnothing$,得 $D_j\cap R_d(x_i)\neq\varnothing$,所以 $D_j=R_d(x_i)$. 由命题 1,2,3 及 $K_B(i)$的定义,有

$$R_B(x_i)\subseteq R_d(x_i)\Leftrightarrow\boldsymbol{f}(R_B(x_i))\leqslant\boldsymbol{f}(R_d(x_i))\Leftrightarrow\boldsymbol{f}(R_B(x_i))\wedge[\boldsymbol{f}(R_d(x_i))]^C=0\Leftrightarrow K_B(i)=0.$$

所以,$\mathrm{Pos}_B(d)=\bigcup\{x_i\in U\mid K_B(i)=0\}$.

定理 5 设 $S=(U,A\cup\{d\})$是不完备决策信息系统,$B\subseteq A$,记 $\boldsymbol{H}_B=\boldsymbol{M}_B\wedge\boldsymbol{M}_d$,$\boldsymbol{H}_B$ 的第 i 行的和记为 $H_B(i)$,$\boldsymbol{M}_B$ 的第 i 行的和,记为 $M_B(i)(i=1,2,\cdots,n)$,则

$$\mathrm{Pos}_B(d)=\bigcup\{x_i\in U\mid H_B(i)=M_B(i)\}.$$

证明:由于 $\mathrm{Pos}_B(d)=\bigcup_{D\in U/R_d}\{\underline{R}_B(D)\}=\bigcup\{x_i\in U\mid R_B(x_i)\subseteq R_d(x_i)\}$,由命题 1,2,3,以及 $K_B(i)$的定义,可得

$$R_B(x_i)\subseteq R_d(x_i)\Leftrightarrow$$
$$\boldsymbol{f}(R_B(x_i))\leqslant\boldsymbol{f}(R_d(x_i))\Leftrightarrow$$
$$\boldsymbol{f}(R_B(x_i))\wedge\boldsymbol{f}(R_d(x_i))=\boldsymbol{f}(R_B(x_i))\Leftrightarrow$$
$$H_B(i)=M_B(i),$$

所以 $\mathrm{Pos}_B(d)=\bigcup\{x_i\in U\mid H_B(i)=M_B(i)\}$.

例 5 利用关系矩阵求不完备信息系统(表 3)的正域,有

表 3 例 5 的不完备决策信息系统

Tab. 3 Incomplete decision information system of example 5

U	A						
	a_1	a_2	a_3	a_4	a_5	a_6	d
x_1	1	1	2	0	*	0	1
x_2	0	1	0	1	*	0	1
x_3	*	0	0	2	1	1	1
x_4	1	*	2	0	*	*	2
x_5	*	0	1	1	*	0	2
x_6	0	*	0	1	0	*	0
x_7	1	2	*	*	*	1	2
x_8	0	2	*	2	*	0	0

$$\boldsymbol{M}_A=\begin{pmatrix}1&0&0&1&0&0&0&0\\0&1&0&0&0&1&0&0\\0&0&1&0&0&0&0&0\\1&0&0&1&0&0&1&0\\0&0&0&0&1&0&0&0\\0&1&0&0&0&1&0&0\\0&0&0&1&0&0&1&0\\0&0&0&0&0&0&0&1\end{pmatrix},\quad \boldsymbol{M}_d=\begin{pmatrix}1&1&1&0&0&0&0&0\\1&1&1&0&0&0&0&0\\1&1&1&0&0&0&0&0\\0&0&0&1&1&0&1&0\\0&0&0&1&1&0&1&0\\0&0&0&0&0&1&0&1\\0&0&0&1&1&0&1&0\\0&0&0&0&0&1&0&1\end{pmatrix},$$

$$\boldsymbol{K}_A=\begin{pmatrix}0&0&0&1&0&0&0&0\\0&0&0&0&0&1&0&0\\0&0&0&0&0&0&0&0\\1&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0\\0&1&0&0&0&0&0&0\\0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0\end{pmatrix},\quad \boldsymbol{H}_A=\begin{pmatrix}1&0&0&0&0&0&0&0\\0&1&0&0&0&0&0&0\\0&0&1&0&0&0&0&0\\0&0&0&1&0&0&1&0\\0&0&0&0&1&0&0&0\\0&0&0&0&0&1&0&0\\0&0&0&1&0&0&1&0\\0&0&0&0&0&0&0&1\end{pmatrix}.$$

$(K_A(1),K_A(2),\cdots,K_A(n))=(1,1,0,2,0,1,0,0)$，由定理 5，得 $\mathrm{Pos}_B(d)=\{x_3,x_5,x_7,x_8\}$，则有

$$\begin{pmatrix}M_A(1)&M_A(2)&\cdots&M_A(8)\\H_A(1)&H_A(2)&\cdots&H_A(8)\end{pmatrix}=\begin{pmatrix}2&2&1&3&1&2&2&1\\1&1&1&2&1&1&2&1\end{pmatrix}.$$

定理 6　设 $S=(U,A\cup\{d\})$ 是不完备决策信息系统，$B\subseteq A$，记 $\boldsymbol{K}_B=\boldsymbol{M}_B\wedge(\boldsymbol{M}_d)^C$，$\boldsymbol{K}_B$ 的第 i 行的和记为 $K_B(i)(i=1,2,\cdots,n)$，$\alpha_{\boldsymbol{K}_B}=(K_B(1),K_B(2),\cdots,K_B(n))$，则有 $\boldsymbol{f}(\mathrm{Pos}_B(d))=(\approx\alpha_{\boldsymbol{K}_B})^C$.

证明：由定理 4，得 $\mathrm{Pos}_B(d)=\bigcup\{x_i\in U\mid K_B(i)=0\}$，且 $f_i=1\Leftrightarrow x_i\in\mathrm{Pos}_B(d)\Leftrightarrow K_B(i)=0$，$f_i=0\Leftrightarrow x_i\notin\mathrm{Pos}_B(d)\Leftrightarrow K_B(i)=1$. 结合定义 6，可得 $\boldsymbol{f}(\mathrm{Pos}_B(d))=(\approx\alpha_{\boldsymbol{K}_B})^C$.

定理 7　假设 $S=(U,A\cup\{d\})$ 是不完备决策信息系统，$B\subseteq A$，记 $\boldsymbol{H}_B=\boldsymbol{M}_B\wedge\boldsymbol{M}_d$，$\boldsymbol{H}_B$ 的第 i 行的和记为 $K_B(i)$，$\boldsymbol{M}_B$ 的第 i 行的和记为 $M_B(i)(i=1,2,\cdots,n)$. 记 $\beta_{\boldsymbol{H}_B}=(\dfrac{H_B(1)}{M_B(1)},\dfrac{H_B(2)}{M_B(2)},\cdots,\dfrac{H_B(n)}{M_B(n)})$，则可得 $\boldsymbol{f}(\mathrm{Pos}_B(d))=(\approx\beta_{\boldsymbol{H}_B})$.

证明：由定理 5，得 $\mathrm{Pos}_B(d)=\bigcup\{x_i\in U\mid H_B(i)=M_B(i)\}$，并且有 $f_i=1\Leftrightarrow x_i\in\mathrm{Pos}_B(d)\Leftrightarrow K_B(i)=M_B(i)$；$f_i=0\Leftrightarrow x_i\notin\mathrm{Pos}_B(d)\Leftrightarrow K_B(i)<M_B(i)$. 结合定义 6 可得 $\boldsymbol{f}(\mathrm{Pos}_B(d))=(\approx\beta_{\boldsymbol{H}_B})$.

例 6　对例 5 给定的不完决策信息系统，可得 $\alpha_{\boldsymbol{K}_A}=(1,1,0,2,0,1,0,0)$，由定义 6，得 $(\approx\alpha_{\boldsymbol{K}_A})=(1,1,0,1,0,1,0,0)$，$\boldsymbol{f}(x_3,x_5,x_7,x_8)=(0,0,1,0,1,0,1,1)=(\approx\alpha_{\boldsymbol{K}_A})^C$，$\beta_{\boldsymbol{H}_A}=(\dfrac{1}{2},\dfrac{1}{2},1,\dfrac{1}{2},1,0,1,1)$，由定义 6，得 $\beta_{\boldsymbol{H}_A}=(0,0,1,0,1,0,1,1)$，$\boldsymbol{f}(x_3,x_5,x_7,x_8)=(0,0,1,0,1,0,1,1)=\beta_{\boldsymbol{H}_A}$.

定理 8　对不完备的决策信息系统 $S=(U,A\cup\{d\})$，$B\subseteq A$，则

1) $\mathrm{Pos}_B(d)=\mathrm{Pos}_A(d)\Leftrightarrow\boldsymbol{f}(\mathrm{Pos}_B(d))=\boldsymbol{f}(\mathrm{Pos}_A(d))\Leftrightarrow(\approx\alpha_{\boldsymbol{K}_B})=(\approx\alpha_{\boldsymbol{K}_A})\Leftrightarrow(\approx\beta_{\boldsymbol{H}_B})=(\approx\beta_{\boldsymbol{H}_A})$；

2) a 是不必要属性 $\Leftrightarrow(\approx\alpha_{\boldsymbol{K}_{A-\{a\}}})=(\approx\alpha_{\boldsymbol{K}_A})\Leftrightarrow\beta_{\boldsymbol{H}_{A-\{a\}}}=\beta_{\boldsymbol{H}_A}$.

例 7　求例 5 对应的不完备信息系统的一个保正域的决策协调集. 取 $B=A-\{a_5,a_6\}$，有

$$\boldsymbol{H}_B=\begin{pmatrix}1&0&0&0&0&0&0&0\\0&1&0&0&0&0&0&0\\0&0&1&0&0&0&0&0\\0&0&0&1&0&0&1&0\\0&0&0&0&1&0&0&0\\0&1&0&0&0&1&0&0\\0&0&0&1&0&0&1&0\\0&0&0&0&0&1&0&1\end{pmatrix}(M_B(i),H_B(i))=\begin{pmatrix}2&1\\2&1\\1&1\\3&2\\1&1\\3&1\\2&2\\2&2\end{pmatrix}.$$

由定理 7 可知，$(\approx\beta_{\boldsymbol{H}_B})=(0,0,1,0,1,0,1,1)$，$(\approx\alpha_{\boldsymbol{K}_B})=(\approx\alpha_{\boldsymbol{K}_A})$，$\mathrm{Pos}_B(d)=(x_3,x_5,x_7,x_8)$，$\mathrm{Pos}_B(d)=\mathrm{Pos}_A(d)$,$B$ 是一个保正域的决策协调集.

定理 9 对不完备的决策信息系统 $S=(U,A\cup\{d\})$,$C\subseteq B\subseteq A$，则下列结论成立:1) $(\approx\alpha_{\boldsymbol{K}_C})\leqslant(\approx\alpha_{\boldsymbol{K}_B})$;2) $\beta_{\boldsymbol{H}_B}\leqslant\beta_{\boldsymbol{H}_C}$.

证明:1) $C\subseteq B\subseteq A\Rightarrow R_B\subseteq R_C\Rightarrow\mathrm{Pos}_B(d)\subseteq\mathrm{Pos}_C(d)\Rightarrow(\approx\alpha_{\boldsymbol{K}_B})^C\leqslant(\approx\alpha_{\boldsymbol{K}_C})^C\Rightarrow(\approx\alpha_{\boldsymbol{K}_C})\leqslant(\approx\alpha_{\boldsymbol{K}_B})$;

2) $C\subseteq B\subseteq A\Rightarrow R_B\subseteq R_C\Rightarrow\mathrm{Pos}_B(d)\subseteq\mathrm{Pos}_C(d)\Rightarrow(\approx\beta_{\boldsymbol{H}_B})\leqslant(\approx\beta_{\boldsymbol{H}_C})$.

定理 10 对不完备决策信息系统 $S=(U,A\cup\{d\})$,则有 1) B 是 A 的决策约简集;2) $(\approx\alpha_{\boldsymbol{K}_B})=(\approx\alpha_{\boldsymbol{K}_A})$,但对任意 B 的真子集 C,$(\approx\alpha_{\boldsymbol{K}_C})\neq(\approx\alpha_{\boldsymbol{K}_A})$;3) $\beta_{\boldsymbol{H}_B}=\beta_{\boldsymbol{H}_A}$,但对任意 B 的真子集 C,$\beta_{\boldsymbol{H}_C}\neq\beta_{\boldsymbol{H}_A}$.

最后,利用关系矩阵定义一个属性重要度,利用启发式算法来求不完备不协调决策信息系统保正域的一个约简集.

定义 9 设不完备决策信息系统 $S=(U,A\cup\{d\})$是不完备决策信息系统和 $B\subseteq A$. 对任意 $a\in A-B$,称 $I(B|a)=\|(\approx\beta_{\boldsymbol{H}_{B\cup\{a\}}})\wedge(\approx\beta_{\boldsymbol{H}_B})^C\|$为属性 a 相对于 B 的重要度.

定义 10 对不完备的决策信息系统 $S=(U,A\cup\{d\})$,$B\subseteq A$ $\forall a\in B$,定义属性 a 的重要度为 $I(a|B)=\|(\approx\beta_{\boldsymbol{H}_B})\wedge(\approx\beta_{\boldsymbol{H}_{B-\{a\}}})^C\|$. 这里 $\|\boldsymbol{M}\|$ 表示布尔矩阵 $\boldsymbol{M}$ 中 1 的个数.

推论 1 设 $S=(U,A\cup\{d\})$是不完备决策信息系统和 $a\in A$. 若 $I(a|A)>0$,则 a 是核心属性.

证明:由 $I(a|A)=\|(\approx\beta_{\boldsymbol{H}_A})\wedge(\approx\beta_{\boldsymbol{H}_{A-\{a\}}})^C\|>0$,则有$(\approx\beta_{\boldsymbol{H}_{A-\{a\}}})<(\approx\beta_{\boldsymbol{H}_A})$,所以 $R_{A-\{a\}}\subset R_A$,a 是核心属性.

$I(B|a)$和 $I(a|B)$均是属性 a 的重要性度量,前者反映 a 加入 B 时,a 相对于 B 的重要度;而后者反映的是将 a 从 B 中删除时,a 相对于 B 的重要度.

如果 $I(a|B)=0$,则有$(\approx\beta_{\boldsymbol{H}_B})\wedge(\approx\beta_{\boldsymbol{H}_{B-\{a\}}})^C=0$,则有$(\approx\beta_{\boldsymbol{H}_{B-\{a\}}})=(\approx\beta_{\boldsymbol{H}_B})$,即 $R_{B-\{a\}}=R_B$,由定理 8 的结论 2)可得,a 相对 B 不必要,$I(a|B)$越大,说明属性 a 对 B 保正域分类的影响越大. 其算法为

输入:不完备决策信息系统 $S=(U,A,d)$,其中,$|U|=n$ 且$|A|=m$;

输出:约简 red.

步骤 1 计算 $\boldsymbol{M}_A$,$\boldsymbol{M}_d$ 和 $\boldsymbol{H}_A\leftarrow\boldsymbol{M}_A\wedge\boldsymbol{M}_d$;

步骤 2 $\beta_{\boldsymbol{H}_A}\leftarrow\left(\dfrac{H_A(1)}{M_A(1)},\cdots,\dfrac{H_A(n)}{M_A(n)}\right)$且令 red$\leftarrow\varnothing$;

步骤 3 对 $a\in A$,计算 $I(a|A)$. 如果 $I(a|A)>0$,则 red$\leftarrow$red$\cup\{a\}$;

步骤 4 $A\leftarrow A-$red;

步骤 5 While $A\neq\Phi$

步骤 6 计算 $I(\mathrm{red}|a)=\max_{b\in A}\{I(\mathrm{red}|b)\}$;

步骤 7 如果 $I(\mathrm{red}|a)>0$,则 red$\leftarrow$red$\cup\{a\}$且 $A\leftarrow A-a$;如果 $I(\mathrm{red}|a)=0$,结束循环;

步骤 8 End While.

步骤 9 For $a\in$ red

步骤 10 若 $I(a|\mathrm{red})=0$,则 red$\leftarrow$red$-\{a\}$;

步骤 11 End For

根据定义,计算 $\boldsymbol{M}_A$ 需要耗费 $O(|U|^2|A|)$,因此,步骤 1~2 的复杂度为 $O(|U|^2|A|)$. 步骤 3 中,由于需要计算每个 $I(a|A)$,所以其复杂度为 $O(|U|^2|A|^2)$. 步骤 5~8 中,由于最坏的情形是 $B=\varnothing$ 及 red$=A$,所以复杂度为 $O(|U|^2|A|^2)$. 由此可知,此算法的时间复杂度为 $O(|U|^2|A|^2)$.

例 8 求如下不完备决策信息系统(表 4)的一个决策约简集.

表 4 例 7 的不完备决策信息系统

Tab. 4 Incomplete decision information system of example 7

U	A					
	a_1	a_2	a_3	a_4	a_5	d
x_1	0	0	2	1	1	1
x_2	0	0	2	0	1	1
x_3	1	0	*	1	2	2
x_4	1	0	1	*	*	2
x_5	1	*	*	2	1	3
x_6	1	1	*	*	0	3

1）$\beta_{H_A}=(1,1,1,0,0,1)$，由于 $I_A(a)=\|(\approx\beta_{H_A})\wedge(\approx\beta_{H_{A-\{a\}}})^C\|$，经计算，得 $I(a_1|A)=0$，$I(a_2|A)=1$，$I(a_3|A)=I(a_4|A)=I(a_5|A)=0$，因此，$a_2$ 为核心属性，则由步骤 3，得 red$=\{a_2\}$.

2）令 $A=A-$red$=\{a_1,a_3,a_4,a_5\}$，因为 $I(\mathrm{red}|a_1)=0$，$I(\mathrm{red}|a_3)=0$，$I(\mathrm{red}|a_4)=0$，$I(\mathrm{red}|a_5)=1$，由步骤 6～7，得 red$=\{a_2,a_5\}$，于是 $A=\{a_1,a_3,a_4\}$. 同样地，重复步骤 5～7，可得 red$=\{a_1,a_2,a_5\}$.

3）由步骤 9～10，得 $I(a_1|\mathrm{red})=2$，$I(a_2|\mathrm{red})=1$，$I(a_3|\mathrm{red})=1$，故 red$=\{a_1,a_2,a_5\}$是一个约简.

3　结束语

基于不完备信息决策系统，在容差关系中引入关系矩阵，研究系统的协调性，给出了一个用关系矩阵判定协调性的方法. 对于不协调的情况，利用关系矩阵，提出一个求正域的新方法，进一步研究了判定协调集的方法. 基于此理论框架，定义了属性重要性度量，设计一个属性约简的启发式算法并分析了其时间复杂度. 在后续的研究中，将进一步进行数值实验，以验证本算法的运算效率及分类性能等.

参考文献：

[1] PAWLAK Z. Rough sets[J]. International Journal of Computer and Information Sciences，1982，11(5)：341-354. DOI：10.1007/BF01001956.

[2] WU Weizhi，QIAN Yuhua，LI Tongjun. On rule acquisition in incomplete multi-scale decision tables[J]. Information Sciences，2017，378(C)：282-302. DOI：10.1016/j.ins.2016.03.041.

[3] LI Feng，HU Baoqing，WANG Jun. Stepwise optimal scale selection for multi-scale decision tables via attribute significance[J]. Knowledge-Based Systems，2017，129：4-16. DOI：10.1016/j.knosys.2017.04.005.

[4] CHEN Degang，ZHANG Xiaoxia，LI Wanlu. On measurements of covering rough sets based on granules and evidence theory[J]. Information Sciences，2015，317(C)：329-348. DOI：10.1016/j.ins.2015.04.051.

[5] 石慧，魏玲. 面向对象(属性)概念格的布尔表达[J]. 南京大学学报(自然科学版)，2015，51(2)：415-420. DOI：10.13232/j.cnki.jnju.2015.02.028.

[6] 李招文，文国秋，林福宁. 变精度不协调 SFD 信息系统中的拓扑学方法[J]. 控制与决策，2014，11：2066-2070. DOI：10.13195/j.kzyjc.2013.0936.

[7] SUN Bingzhen，MA Weimin，QIAN Yuhua. Multigranulation fuzzy rough set over two universes and its application to decision making[J]. Knowledge-Based Systems，2017，123：61-74. DOI：10.1016/j.knosys.2017.01.036.

[8] WANG Chongzhong. Topological structures of L-fuzzy rough sets and similarity sets of L-fuzzy relations[J]. International Journal of Approximate Reasoning，2017，83：160-175. DOI：10.1016/j.ijar.2017.01.002.

[9] WANG Changzhong，HE Qiang，CHEN Degang，*et al*. A novel method for attribute reduction of covering decision systems[J]. Information Sciences，2014，254：181-196. DOI：10.1016/j.ins.2013.08.057.

[10] 王磊，李天瑞. 一种基于矩阵的知识粒度计算方法[J]. 模式识别与人工智能，2013，26(5)：448-453. DOI：10.3969/j.issn.1003-6059.2013.05.005.

[11] 林艺东，张燕兰，林梦雷. 覆盖族动态变化时粗集计算的矩阵方法[J]. 计算机应用，2015，35(11)：3208-3212. DOI：10.11772/j.issn.1001-9081.2015.11.3208.

[12] TAN Anhui，LI Jinjin，LIN Guoping，*et al*. Fast approach to knowledge acquisition in covering information systems using matrix operations[J]. Knowledge-Based Systems，2015，79：90-98. DOI：10.1016/j.knosys.2015.02.003.

[13] 陈恒新. 可正定化矩阵的判别定理[J]. 华侨大学学报(自然科学版)，2011，32(5)：356-360. DOI：10.11830/issn.1000-5013.2011.03.0356.

[14] MA Liwen. The investigation of covering rough sets by Boolean matrices[J]. International Journal of Approximate-eReasoning，2018，100：69-84. DOI：10.1016/j.ijar.2018.05.008.

（**责任编辑：**陈志贤　　**英文审校：**黄心中）