第 17 卷第 1 期 空 军 工 程 大 学 学 报(自然科学版) Vol.17 No.1
2016 年 2 月 JOURNAL OF AIR FORCE ENGINEERING UNIVERSITY(NATURAL SCIENCE EDITION) Jan.2016

非一致决策表的一种新的水平分解方法

周 炜， 王晓丹， 路艳丽
(空军工程大学防空反导学院,西安,710051)

摘要 在提出三异等价类和三异组这 2 个新概念的基础上,给出了基于粗糙集的一种新的非一致决策表水平分解方法。该方法也可以像现有分解方法那样将非一致决策表分解为 1 个一致决策表和 1 个完全不一致决策表,但分解得到的一致决策表可能比用现有方法分解得到的一致决策表更大一些,造成的信息损失更小一些。同时,新方法还可以用来分解一部分完全不一致的决策表,而现有方法无法分解这些完全不一致的决策表。

关键词 粗糙集;决策表;水平分解;三异等价类;三异组
DOI 10.3969/j.issn.1009-3516.2016.01.018
中图分类号 TP18 **文献标志码** A **文章编号** 1009-3516(2016)01-0096-04

A New Horizontal Decomposition Approach of Inconsistent Decision Tables

ZHOU Wei, WANG Xiaodan, LU Yanli
(Air and Missile Defense College, Air force Engineering University, Xi′an 710051, China)

Abstract: On the basis of the presented two new concepts of tri-singular equivalence class and tri-singular set a new approach based on the rough set for horizontal decomposition of inconsistent decision tables is proposed. The new approach, just as the existing approach does, can decompose an inconsistent decision table into two decision tables, i.e., a consistent decision table and a complete inconsistent decision table. But the consistent decision table decomposed may be greater than that obtained by using the existing approach, thus causing less information loss. It is noteworthy that the new approach may be used to decompose parts of complete inconsistent decision tables, whereas complete decision tables can not be done by using the existing approach.

Key words: Rough set; decision table; horizontal decomposition; tri-singular equivalence class; tri-singular set

波兰学者 Pawlak 在 1982 年[1]提出的粗糙集理论已得到了国内外广泛深入的研究。粗糙集理论的重要应用之一是决策表的约简和分解[2-4],国内外学者在这方面做了大量工作[5-13]。决策表有一致和非一致之别。一致决策表无疑比非一致决策表更有价值,然而在实践中却常常须要处理非一致决策表。为了从非一致决策表中提取有用的信息,人们给出了将其分解为一致决策表和一个完全不一致决策表的水平分解方法[2-5]。本文在提出三异等价类和三异组概念的基础上以定理形式提出非一致决策表的一种新的水平分解方法,按该方法分解决策表得到的一致决策表可能比按现有方法分解得到的一致决

收稿日期:2015-03-24
基金项目:国家自然科学基金(61273275;60975026)
作者简介:周 炜(1962—),男,陕西凤翔人,副教授,主要从事智能信息处理与信息安全研究.E-mail: phoenixshaanxi@163.com

引用格式:*周炜,王晓丹,路艳丽. 非一致决策表的一种新的水平分解方法[J]. 空军工程大学学报:自然科学版,2016,17(1):96-99. ZHOU Wei, WANG Xiaodan, LU Yanli. A New Horizontal Decomposition Approach of Inconsistent Decision Tables[J]. Journal of Air Force Engineering University: Natural Science Edition,2016,17(1):96-99.*

策表更大一些，保留的样本更多一些，因而丢失的信息更少一些。

本文假设 U 是一个非空有限集合，称为论域，并以 $\bar{X}$ 表示 $X \subseteq U$ 的补 $U-X$ 。

1　上、下近似和正域

定义 1[1]　设 R 是 U 上的等价关系，$X \subseteq U$ 。定义 X 的上、下 R 近似如下：

$R^*(X)=\{x \mid x \in U \wedge [x]_R \cap X \neq \varnothing\}$ ，

$R_*(X)=\{x \mid x \in X \wedge [x]_R \subseteq X\}$ 。

对于等价关系族 Ω ，定义 $\Omega_*(X) = \text{ind}(\Omega)_*(X)$ 。

定理 1　设 P 、Q 是 U 上的关系。$P \subseteq Q$ 的充要条件是：对于任意的 $x \in U$ ，成立 $[x]_P \subseteq [x]_Q$ 。

定理 2　设 Ω 是 U 上的一族等价关系，则 $\text{ind}(\Omega)=\bigcap_{R\in\Omega} R$ 也是 U 上的一个等价关系，且任一 $x \in U$ 在 $\text{ind}(\Omega)$ 下的等价类是 $[x]_\Omega=\bigcap_{R\in\Omega}[x]_R$ 。

定理 1 和**定理 2** 常用来计算等价类。

定理 3　设 Π 、Ω 是 U 上的两族等价关系。$\text{ind}(\Pi) \subseteq \text{ind}(\Omega)$ 当且仅当对于任意的 $x \in U$ ，有 $[x]_\Pi \subseteq [x]_\Omega$ 。

定理 4[14]　设 R 是 U 上的一个等价关系，$X \subseteq U$ ，则：$R_*(X)=\bigcup_{x\in X\wedge[x]_R\subseteq X}[x]_R$ ，$R^*(X)=\bigcup_{x\in X}[x]_R$ 。

定义 2[1]　设 R 是 U 上的等价关系，$X \subseteq U$ ，则称 $\text{pos}_R(X)=R_*(X)$ 为 X 的 R 正域。

定义 3[1]　设 P 和 Q 是 U 上的等价关系。称 $\text{pos}_P(Q)=\bigcup_{X\in U/Q}\text{pos}_P(X)$ 为 Q 的 P 正域。

定理 5[14]　设 P 和 Q 是 U 上的等价关系，则：

$\text{pos}_P(Q)=\{x \mid x \in U \wedge [x]_P \subseteq [x]_Q\}$ 。

2　信息系统和决策表的概念

定义 4[2-4]　信息系统是一个四元组 $S=(U,A,V,f)$ ，其中：U 是一个有限论域；$A \neq \varnothing$ 是非空有限的属性集合；$V=\bigcup_{a\in A}V_a$ ，V_a 是属性 a 的值域；$f:U\times A \to V$ 是一个信息函数，它为每一个对象 $x \in U$ 在每一个属性 $a \in A$ 上指定唯一的属性值 $f(x,a) \in V_a$ 。

在信息系统 S 中，对于每一个 $a \in A$ ，如此定义 U 上的一个二元关系 $E(a)$ ：$xE(a)y$ 当且仅当 $f(x,a)=f(y,a)$ 。则 $E(a)$ 是 U 上的一个等价关系。于是，对于每一个 $B \subseteq A$ ，$\{E(a) \mid a \in B\}$ 是 U 上的一族等价关系，并记 $\text{ind}(B)=\bigcap_{b\in B}E(b)$ ，$U/B=U/\text{ind}(B)$ ，$[x]_B=[x]_{\text{ind}(B)}$ 。

若 $B \subseteq B' \subseteq A$ ，则 $\text{ind}(B') \subseteq \text{ind}(B)$ 。于是 $\text{ind}(B')=\text{ind}(B)$ 当且仅当 $\text{ind}(B) \subseteq \text{ind}(B')$ 。

定义 5[2-4]　设 $S=(U,A,V,f)$ 是一个信息系统，$C,D \subseteq A$ 。下面的 2 个函数：

$\text{pos}_C(D)=\text{pos}_{\text{ind}(C)}(\text{ind}(D))$

$\gamma_C(D)=|\text{pos}_C(D)|/|U|$

分别称为在信息系统 S 中 D 的 C 正域和 D 关于 C 的相对依赖度。

信息系统可以用二维表来表示。二维表的行对应着 U 中的元素，列对应着属性。每一对象在各属性上的取值形成的元组表达了该对象的信息。

定义 6[2-4]　设 $S=(U,A,V,f)$ 是一个信息系统。如果能将属性集合 A 划分为一个条件属性集合 $C \neq \varnothing$ 和一个决策属性集合 $D \neq \varnothing$ 的并 $A=C \cup D$ ，且 $C \cap D=\varnothing$ ，则称此信息系统是一个决策表，记作 $T=(U,A,C,D)$ 。若 $\gamma_C(D)=1$，则称此决策表是一致的；若 $\gamma_C(D)<1$，则称此决策表是非一致的；特别地，若 $\gamma_C(D)=0$，则称此决策表是完全不一致的。

定理 6　设 $T=(U,A,C,D)$ 是一个决策表，$\varnothing \subset B \subseteq C$ 。$\text{pos}_B(D)=\text{pos}_C(D)$ 当且仅当 $\text{pos}_C(D) \subseteq \text{pos}_B(D)$ 。

证明：因 $B \subseteq C$ ，故 $\text{ind}(C) \subseteq \text{ind}(B)$ 。对于每一个 $x \in U$ ，根据**定理 3** 有 $[x]_C \subseteq [x]_B$ 。若 $[x]_B \subseteq [x]_D$ ，则也有 $[x]_C \subseteq [x]_D$ 。因此由**定理 5** 得到 $\text{pos}_B(D) \subseteq \text{pos}_C(D)$ 。

由此可见，$\text{pos}_B(D)=\text{pos}_C(D)$ 当且仅当 $\text{pos}_C(D) \subseteq \text{pos}_B(D)$ 。

3　决策表的一个新分解定理

定理 7[2-4]　设 $T=(U,A,C,D)$ 是一个决策表，则 T 可以分解为 2 个决策表 T_1 和 T_2，记作 $T=\{T_1,T_2\}$ 。这里 $T_1=(U_1,A,C,D)$ ，$T_2=(U_2,A,C,D)$ ，$U_1=\text{pos}_C(D)$ 和 $U_2=\bigcup_{X\in U/D}bn_{\text{ind}(C)}(X)$ 分别在原决策表 T 中计算，$U_1 \cap U_2=\varnothing$ ，$U_1 \cup U_2=U$ 。决策表 T_1 是一致的，而决策表 T_2 是完全不一致的。

定理 7 是决策表的现有分解定理，它只是给出了分解出一个一致决策表和一个完全不一致决策表的方法，但不是唯一的方法。

例 1　表 1 给出了一个决策表 $T=(U,A,C,D)$ ，其中 $C=\{a,b,c\}$ ，$D=\{d,e\}$ 。若

此决策表是非一致的,对其进行分解。

表 1 一个非一致决策表

Tab.1 Aninconsistent decision table

U	a	b	c	d	e
1	2	2	0	1	1
2	1	2	0	0	1
3	1	2	0	1	0
4	0	0	0	0	1
5	1	0	1	0	1
6	2	0	1	1	0
7	1	2	0	1	1

解:先确定 U/C 和 U/D ,并计算 $\mathrm{pos}_C(D)$:

$U/C=\{\{1\},\{2,3,7\},\{4\},\{5\},\{6\}\}$,

$U/D=\{\{1,7\},\{2,4,5\},\{3,6\}\}$,

$\mathrm{pos}_C(D)=\{1,4,5,6\}$, $\gamma_C(D)=\dfrac{4}{7}<1$。

可见,此决策表是非一致的。

令 $U_1=\mathrm{pos}_C(D)\cup\{7\}=\{1,4,5,6,7\}$, $U_2=\overline{\mathrm{pos}_C(D)}=\{2,3\}$,得到原决策表的一个分解 $T=\{T_1,T_2\}$ 。

在决策表 $T_1=(U_1,A,C,D)$ 中,计算得到:

$U/C=\{\{1\},\{4\},\{5\},\{6\},\{7\}\}$,

$U/D=\{\{1,7\},\{4,5\},\{6\}\}$,

$\mathrm{pos}_C(D)=U_1$, $\gamma_C(D)=1$。

在决策表 $T_2=(U_2,A,C,D)$ 中,计算得到:

$U/C=\{\{2,3\}\}$, $U/D=\{\{2\},\{3\}\}$,

$\mathrm{pos}_C(D)=\varnothing$, $\gamma_C(D)=0$。

此例并没有用**定理 7** 的方法去分解所给的决策表,分解结果完全符合**定理 7** 的要求但优于按**定理 7** 分解的结果。因为如果按定理 7 分解,会得到 $U_1=\mathrm{pos}_C(D)=\{1,4,5,6\}$,信息损失更大一些。下面以此为背景给出一种新的分解方法。

定义 7 设 $T=(U,A,C,D)$ 是一个决策表, $X\in U/C$ 。如果存在 3 个不同元素 $x,y,z\in X$,使 $X\cap[x]_D$ 、$X\cap[y]_D$ 、$X\cap[z]_D$ 两两不同,则称 X 是一个三异的 C 等价类(Tri-singular C-equivalence Class),而三元素集合 $\{x,y,z\}$ 称为 X 的一个三异组(Tri-singular Set)。决策表 T 的所有三异 C 等价类的族记作 $T(C)$ 。对于每一个 $X\in T(C)$,任意选择一个三异组,记作 $\mathrm{Trs}(X)$;同时在 $\mathrm{Trs}(X)$ 中选择一个元素,记作 $\mathrm{trs}(X)$ 。

设 $X\in T(C)$, $\mathrm{Trs}(X)=\{x,y,z\}$ 。显然 $[x]_C=[y]_C=[z]_C=X$, $X\cap[x]_D$ 、$X\cap[y]_D$ 、$X\cap[z]_D$ 两两不相交,从而 $[x]_D$ 、$[y]_D$ 、$[z]_D$ 也两两不相交。更进一步,还有:

$X\subseteq[x]_D$, $X\subseteq[y]_D$, $X\subseteq[z]_D$;

$X-[x]_D\subseteq[y]_D$, $X-[x]_D\subseteq[z]_D$;

$X-[y]_D\subseteq[x]_D$, $X-[y]_D\subseteq[z]_D$;

$X-[z]_D\subseteq[x]_D$, $X-[z]_D\subseteq[y]_D$ 。

定理 8 设 $T=(U,A,C,D)$ 是一个决策表,则 T 可以分解为两个决策表 T_1 和 T_2,记作 $T=\{T_1,T_2\}$ 。这里 $T_1=(U_1,A,C,D)$, $T_2=(U_2,A,C,D)$, $U_1=\mathrm{pos}_C(D)\cup\bigcup\limits_{X\in T(C)}[trs(X)]_C\cap[trs(X)]_D$ 和 $U_2=\overline{U_1}$ 分别在原决策表 T 中计算, $U_1\cap U_2=\varnothing$, $U_1\cup U_2=U$ 。决策表 T_1 是一致的,而决策表 T_2 是完全不一致的。

证明:当 $T(C)=\varnothing$ 时,由**定理 7** 直接得出结论。以下设 $T(C)\neq\varnothing$ 。

对于每一个 $x\in U_1$,若 $x\in\mathrm{pos}_C(D)$,则 $[x]_C\subseteq[x]_D$,故 $[x]_C\cap U_1=[x]_C\subseteq[x]_D\cap U_1$;否则,若 $x\notin\mathrm{pos}_C(D)$,则存在 $X\in T(C)$,使 $x\in[\mathrm{trs}(X)]_C\cap[\mathrm{trs}(X)]_D$, $[x]_C=[\mathrm{trs}(X)]_C$, $[x]_D=[\mathrm{trs}(X)]_D$,因此 $[x]_C\cap U_1=[x]_C\cap[x]_D\subseteq[x]_D\cap U_1$。所以,当 $U_1\neq\varnothing$ 时,对于决策表 T_1,成立 $\gamma_C(D)=1$, T_1 是一致的。

设 $x\in U_2$,记 $X=[x]_C$ 。若 $X\in T(C)$,则 $[x]_C\cap[x]_D\subseteq U_1$,而 $X-[x]_D\subseteq U_2$。因此 $[x]_C\cap U_2=X-[x]_D\subseteq U_2$。这时候,定理前面已分析得知 $[x]_C\cap U_2\subseteq[x]_D\cap U_2$。若 $X\notin T(C)$,则 $[x]_C\subseteq U_2$。因 $[x]_C\subseteq[x]_D$,存在 $y\in[x]_C$,使 $y\notin[x]_D$,即 $[x]_D\cap[y]_D=\varnothing$ 。显然 $X\cap([x]_D\cup[y]_D)\subseteq[x]_C$ 。可以断言: $[x]_C=X\cap([x]_D\cup[y]_D)$ 。否则存在 $z\in[x]_C$,使 $z\notin X\cap([x]_D\cup[y]_D)$ 。故 $z\notin[x]_D\cup[y]_D$ 。从而 $[x]_D\cap[z]_D=\varnothing$, $[y]_D\cap[z]_D=\varnothing$ 。从而 $X\cap[x]_D$ 、$X\cap[y]_D$ 和 $X\cap[z]_D$ 两两不同,与 $X\notin T(C)$ 矛盾。所以 $[_x=X\cap([_x\cup[_y)=(X\cap[_x)\cup(X\cap[_y)$。因 $(X\cap[y]_D)\subseteq[x]_D\cap U_2$,故 $[x]_C\cap U_2=[x]_C\subseteq[x]_D\cap U_2$。

可见在决策表 T_2 中,没有 $x\in U_2$ 满足 $[x]_C\cap U_2\subseteq[x]_D\cap U_2$。因此,当 $U_2\neq\varnothing$ 时, T_2 是完全不一致的。

例 2 表 2 给出了一个决策表 $T=(U,A,C,D)$,其中 $C=\{a,b,c\}$, $D=\{d,e\}$ 。若此决策表是非一致的,对其进行分解。

解:先确定 U/C 和 U/D ,并计算 $\mathrm{pos}_C(D)$:

$U/C=\{\{1,3,5,7\},\{2,4,6,8\}\}$,

$U/D=\{\{1,4,7\},\{2,5,8\},\{3,6\}\}$,

$\mathrm{pos}_C(D)=\varnothing$, $\gamma_C(D)=0$。

可见,此决策表是完全不一致的。其中 $X_1=\{1,3,5,7\}$ 和 $X_2=\{2,4,6,8\}$ 都是三异的 C 等价类, X_1 有三异组 $\mathrm{trs}(X_1)=\{1,3,5\}$, X_2 有三异组 $\mathrm{Trs}(X_2)=\{2,4,6\}$ 。

取 $trs(X_1)=1$, $trs(X_2)=2$。根据**定理 8**,令 $U_1=pos_C(D)\cup(\{1,3,5,7\}\cap\{1,4,7\})\cup(\{2,4,6,8\}\cap\{2,5,8\})=\{1,2,7,8\}$, $U_2=\overline{U_1}=\{3,4,5,6\}$,得到原决策表的一个分解 $T=\{T_1,T_2\}$。在决策表 $T_1=(U_1,A,C,D)$ 中,计算得到 $U/C=\{\{1,7\},\{2,8\}\}$, $U/D=\{\{1,7\},\{2,8\}\}$, $pos_C(D)=U_1$, $\gamma_C(D)=1$。

在决策表 $T_2=(U_2,A,C,D)$ 中,计算得到 $U/C=\{\{3,5\},\{4,6\}\}$, $U/D=\{\{3,6\},\{4\},\{5\}\}$, $pos_C(D)=\varnothing$, $\gamma_C(D)=0$。

表 2　一个完全不一致决策表

Tab.2　A completely inconsistent decision table

U	a	b	c	d	e
1	1	1	0	1	1
2	0	2	1	0	1
3	1	1	0	1	0
4	0	2	1	1	1
5	1	1	0	0	1
6	0	2	1	1	0
7	1	1	0	1	1
8	0	2	1	0	1

4　结语

将一个非一致决策表分解成一个一致决策表和一个完全不一致决策表的方法不是唯一的,定理 7 给出的只是使分解出包含 $pos_c(D)$ 的一致决策表最小的分解方法。按**定理 8** 的方法分解决策表得到的一致决策表可能比按**定理 7** 的方法分解得到的一致决策表更大一些,含有的原始样本更多一些,因而信息损失更小一些。值得注意的是,用**定理 8** 的方法还可以从那些具有三异 C 等价类且完全不一致的决策表中分解出一致的子决策表。

本文提出的非一致决策表分解方法并不是最优的,分解的结果仍然有信息损失,更好的分解方法需进一步进行研究。

参考文献(References):

[1]　Pawlak Z. Rough Sets[J]. International Journal of Computer and Information Sciences, 1982, 11(5): 341-356.

[2]　Pawlak Z. Rough Sets and Decision Tables[J]. Lecture Notes and Computer Sciences, 1985, 208: 187-196.

[3]　Pawlak Z. On Decision Tables[J]. Bulletin of The Polish Academy of Sciences: Technical Sciences, 1986, 34: 563-589.

[4]　史忠植. 知识发现[M]. 北京:清华大学出版社, 2001.
SHI Zhongzhi. Knowledge Discovery[M]. Beijing: Tsinghua University Press, 2001.(in Chinese)

[5]　Pawlak Z, Sowinski R. Rough Set Approach to Multi-attribute Decision Analysis[J]. European Journal of Operational Research-EJOR, 1994, 72(3): 443-459.

[6]　Pawlak Z. Rough Set Theory and Its Applications to Data Analysis[J]. Cybernetics and Systems, 1998, 29(7): 661-688.

[7]　王国胤,于洪,扬大春. 基于条件信息熵的决策表约简[J]. 计算机学报, 2002, 25(7): 759-766.
WANG Guoyin, YU Hong, YANG Dachun. Decision Table Reduction Based on Conditional Information Entropy[J]. Chinese Journal of Computers, 2002, 25(7): 759-766.(in Chinese)

[8]　王国胤. 决策表核属性的计算方法[J]. 计算机学报, 2003, 26(5): 611-615.
WANG Guoyin. Calculation Methods for Core Attributes of Decision Table[J]. Chinese Journal of Computers, 2003, 26(5): 611-615.(in Chinese)

[9]　刘旭东,秦亮曦. 基于决策表分解的属性约简算法[J]. 计算机工程与设计, 2014, 35(8): 2872-2875.
LIU Xudong, QIN Liangxi. Decision Table Decomposition Based Attributes Reduction Algorithm[J]. Computer Engineering and Design, 2014, 35(8): 2872-2875.(in Chinese)

[10]　叶东毅. 不相容决策表分解的若干性质[J]. 小微型计算机系统, 2006, 27(4): 695-697.
YE Dongyi. Some Properties of Decomposition of Inconsistent Decision Tables[J]. Journal of Chinese Computer Systems, 2006, 27(4): 695-697. (in Chinese)

[11]　王加阳,刘柳明,罗安. 大型决策表分解方法研究[J]. 计算机科学, 2007, 34(8): 211-214.
WANG Jiayang, LIU Liuming, LUO An. Research on Decomposition of Large Decision Table[J]. Computer Science, 2007, 34(8): 211-214. (in Chinese)

[12]　翟翠红,秦克云,不协调决策表中基于对象的近似约简[J]. 计算机科学, 2014, 41(3): 245-248.
ZHAI Cuihong, QIN Keyun, Approximate Reduction for Objects in Inconsistent Decision Tables[J]. Computer Science, 2014, 41(3): 245-248. (in Chinese)

[13]　安涛,何宇廷,舒文军,等,基于粗糙集的老龄飞机金属结构安全性评估方法[J]. 空军工程大学学报:自然科学版, 2014, 15(2): 91-94.
AN Tao, HE Yuting, SHU Wenjun, et al. A Research on Metal Structure Safety Evaluation of Aging Aircraft[J]. Journal of Air Force Engineering University: Natural Science Edition, 2014, 15(2): 91-94. (in Chinese)

[14]　周炜,雷英杰. Pawlak 粗糙集的若干重要性质[J]. 空军工程大学学报:自然科学版, 2008, 9(3): 76-78.
ZHOU Wei, LEI Yingjie. A Few Important Properties of Pawlak Rough Sets[J]. Journal of Air force Engineering University: Natural Science Edition, 2008, 9(3): 76-78. (in Chinese)

(编辑:徐楠楠)