

## 协调多源决策表的规则提取

万 青<sup>1,3\*</sup>, 魏 玲<sup>2,3</sup>, 任睿思<sup>2,3</sup>

(1. 西安工程大学理学院, 西安, 710048; 2. 西北大学数学学院, 西安, 710127;

3. 西北大学概念、认知与智能研究中心, 西安, 710127)

**摘 要:** 规则提取是知识发现的一个重要研究方向. 多源数据是一类重要的数据集, 从不同的角度对多源数据进行规则提取可为决策行为提供更加可信的依据. 以多源决策表为数据基础, 首先以数据源和决策规则的结论为切入点, 提出两类多源决策规则的概念及其支持度和覆盖度的刻画方式; 其次, 分析这两类多源决策规则之间的联系; 最后通过引入单源化决策表的定义, 研究这两类多源决策规则的获取方法, 并基于此讨论了两类多源决策规则的简化方法.

**关键词:** 多源决策表, 多源决策规则, 支持度, 覆盖度, 规则简化

**中图分类号:** TP18

**文献标识码:** A

## Rule acquisition of consistent multi-source decision tables

Wan Qing<sup>1,3\*</sup>, Wei Ling<sup>2,3</sup>, Ren Ruisi<sup>2,3</sup>

(1. School of Science, Xi'an Polytechnic University, Xi'an, 710048, China;

2. School of Mathematics, Northwest University, Xi'an, 710127, China;

3. Institute of Concepts, Cognition and Intelligence, Northwest University, Xi'an, 710127, China)

**Abstract:** Rule acquisition is one of important research fields of knowledge discovery. Multi-source data is an important data set, and obtaining rules in multi-source data from different perspectives can provide a more reliable basis for decision making. In this paper, based on multi-source decision tables, the definitions of two types of multi-source decision rules are presented from the perspectives of the data source and the conclusion of decision rule. Then, the support degree and the coverage degree of two types of rules are proposed. After that, the relationships between two types of multi-source decision rules are discussed. Finally, the approach to rule acquisition and rule simplification of the first and second types of multi-source decision rules are investigated by introducing the unisource decision table of multi-source decision table.

**Key words:** multi-source decision table, multi-source decision rule, support degree, coverage degree, rule simplification

粗糙集理论(Rough Set Theory)<sup>[1]</sup>是知识发现的重要工具, 而规则提取是粗糙集理论的研究热点. 在粗糙集理论中, 数据的具体表现形式为信息系统. 信息系统是一个二维数据表, 记为  $(U, A, F)$ , 其中  $U$  是对象集,  $A$  是属性集,  $F$  是  $U$  和  $A$  之间的关系集. 进一步, 若将属性分为条件属性  $A$  和决策属性  $d$ , 则可得到决策表  $(U, A, F,$

$d)$ . 在决策表中, 通过研究两类属性的关系可以获得决策知识, 即决策规则.

目前, 针对决策表, 已有许多学者研究了其决策规则提取问题. 例如, 常犁云等<sup>[2]</sup>从属性约简和值约简两方面同时出发简化了决策表的决策规则; Kryszkiewicz<sup>[3]</sup>针对不完备决策表提出了获取所有最优确定决策规则的方法; 黄兵和周献中<sup>[4]</sup>

基金项目: 国家自然科学基金(61772021, 61976130), 陕西省教育厅专项基金(19JK0380), 西安工程大学博士科研启动基金(BS1528)

收稿日期: 2020-06-20

\* 通讯联系人, E-mail: wqysbe@163.com

基于矩阵的方法,利用条件属性矩阵和决策矩阵研究不协调决策表的决策规则提取方法;钱宇华等<sup>[5]</sup>探讨了决策表中的粒度思想,并建立了知识粒度与决策规则置信度之间的动态关系;Hao et al<sup>[6]</sup>将多粒度粗糙集中的决策规则与形式概念分析中的决策规则进行了比较研究;Chen et al<sup>[7]</sup>在不协调决策表中研究了由属性值的粗化和细化所引起的决策规则更新问题;Wu et al<sup>[8]</sup>针对不完备多粒度标记决策表在协调和不协调两种情况下研究了决策规则提取问题;Zhang et al<sup>[9]</sup>在区间集决策表中引入一个评估决策规则覆盖能力的度量,通过设置阈值得到了具有较高置信度且具有较强覆盖能力的决策规则。

规则提取是知识发现的一个重要研究方向,通过提取潜在的有效规则可为决策行为提供可靠依据.上述关于决策规则的研究方法仅限于针对单一来源数据构成的数据表,对于由多个源头的数据构成的数据表,若直接使用上述方法则提取的决策规则不能完全体现此类数据源头多的特点.该类多源数据在实际生活中普遍存在,是复杂数据类型之一,而且是一类重要的数据集.从不同的角度对多源数据进行知识获取研究,可进一步丰富复杂数据分析的方法.

现今,针对多源数据的研究主要集中在信息融合和约简两个方面<sup>[10-19]</sup>,而关于多源决策表的决策规则,目前的研究相对较少.其中,林国平等<sup>[18]</sup>基于多粒度粗糙集研究了决策规则评价函数的融合方法,给出了决策规则的整体评价指标.万青等<sup>[19]</sup>基于新提出的多源决策表,从乐观策略的角度给出了多源决策规则的定义.而对于多源决策表,从多个层次、多个粒度对其展开分析,可以同时从不同角度挖掘数据中潜在的有用知识,进而对其进行有效融合,可为决策行为提供更多有效的依据.

为了达到从多个角度、多个层次出发分析多源数据的目的,本文在万青等<sup>[19]</sup>研究的基础上,进一步研究多源决策表的决策规则获取问题.首先针对协调多源决策表,分别以数据源和决策规则的结论为切入点,定义第一类多源决策规则和第二类多源决策规则的概念;其次,为了刻画这两类多源决策规则的数字特征,给出了其支持度和

覆盖度的定义,并讨论了这两类多源决策规则之间的联系;最后,提出了单源化决策表的定义,基于此研究了两类多源决策规则的获取方法,继而提出保持两类多源决策规则不变的约简.

## 1 预备知识

本节主要回顾多源信息系统与多源决策表的相关概念.

**定义 1**<sup>[19]</sup> 称  $MIS = \{IS_k | IS_k = (U, A, F_k), k = 1, 2, \dots, s\}$  是多源信息系统. 其中  $IS_k = (U, A, F_k)$  为信息系统,  $U = \{x_1, x_2, \dots, x_m\}$  为对象集, 每个  $x_i (i \leq m)$  称为一个对象;  $A = \{a_1, a_2, \dots, a_n\}$  为属性集, 每个  $a_j (j \leq n)$  称为一个属性;  $F_k = \{f_{kl}: U \rightarrow V_l (l \leq |A|, a_l \in A)\}$  为关系集,  $V_l$  是  $a_l$  的值域.

称  $IS_k = (U, A, F_k)$  为多源信息系统  $MIS$  的第  $k$  个信息系统. 记:

$$R_{kA} = \{(x_i, x_j) \in U \times U | f_{kl}(x_i) = f_{kl}(x_j), a_l \in A\}$$

称  $R_{kA}$  为  $U$  在第  $k$  个信息系统  $IS_k$  上关于  $A$  的一个等价关系. 记:

$$[x_i]_{kA} = \{x_j | (x_i, x_j) \in R_{kA}\}$$

则称  $U/R_{kA} = \{[x_i]_{kA} | x_i \in U\}$  为  $U$  在第  $k$  个信息系统  $IS_k$  上关于  $A$  的一个划分.

**定义 2**<sup>[19]</sup> 称  $MDT = \{DT_k | DT_k = (U, A \cup d, F_k, d_k), k = 1, 2, \dots, s\}$  是多源决策表. 其中  $DT_k = (U, A \cup d, F_k, d_k)$  为决策表,  $A$  为条件属性集,  $d$  为决策属性;  $d_k: U \rightarrow V_d$  为关系集,  $V_d$  为  $d$  的值域.

称  $DT_k = (U, A \cup d, F_k, d_k)$  为多源决策表  $MDT$  的第  $k$  个决策表. 记:

$$R_{kd} = \{(x_i, x_j) \in U \times U | f_{kd}(x_i) = f_{kd}(x_j)\}$$

$$U/R_{kd} = \{[x_i]_{kd} | x_i \in U\}$$

称  $R_{kd}$  为  $U$  在第  $k$  个决策表  $DT_k$  上关于  $d$  的一个等价关系,  $U/R_{kd}$  为  $U$  在第  $k$  个决策表  $DT_k$  上关于  $d$  的一个划分.

记  $\mathcal{F} = \{F_k | k = 1, 2, \dots, s\}$ , 称  $\mathcal{F}$  为  $MDT$  的数

据源集合,从而也称  $DT_k = (U, A \cup d, F_k, d_k)$  为  $MDT$  的第  $k$  个数据源.

**定义 3**<sup>[19]</sup> 设  $MDT = \{DT_k | DT_k = (U, A \cup d, F_k, d_k), k = 1, 2, \dots, s\}$  是多源决策表.  $\forall k \in \{1, 2, \dots, s\}$ , 若  $R_{kA} \subseteq R_{kd}$ , 则称  $MDT$  为协调多源决策表. 否则, 称  $MDT$  为不协调多源决策表.

需要指出, 本文约定对于  $MDT$  中任意两个不同的决策表, 若一个对象在条件属性下的取值完全相等, 则它们在决策属性下的取值也一定相等.

针对上述相关概念, 通过一个例子来解释.

**例 1**<sup>[19]</sup> 表 1 是一个中学教师课堂教学评议表. 其中  $U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9\}$  是对象集, 分别代表语文、数学、英语、历史、地理、政治、生物、化学和物理共九门课程的代课教师;  $A = \{a_1, a_2, a_3, a_4\}$  是条件属性集, 分别代表教学态度、教学内容、教学方法和教学效果四个指标, 其值域均为  $\{1, 2, 3, 4\}$ , 分别代表  $[60, 70]$   $(70, 80]$   $(80, 90]$   $(90, 100]$  四个分数段;  $\mathcal{F} = \{F_1, F_2, F_3\}$  是数据源的集合, 分别代表三个教学班级;  $d$  是决策属性, 表示评价结果, 其值域为  $V_d = \{1, 2, 3\}$ , 分别表示中、良和优.

表 1 多源决策表  $MDT$

Table 1 A multi-source decision table  $MDT$

$U$	$F_1$					$F_2$					$F_3$				
	$a_1$	$a_2$	$a_3$	$a_4$	$d$	$a_1$	$a_2$	$a_3$	$a_4$	$d$	$a_1$	$a_2$	$a_3$	$a_4$	$d$
$x_1$	4	3	3	3	3	4	3	3	2	3	4	3	3	4	3
$x_2$	4	3	3	3	3	4	3	3	2	3	4	3	3	3	3
$x_3$	4	3	3	3	3	3	3	3	2	2	4	3	3	3	3
$x_4$	4	3	3	2	3	3	2	3	2	2	3	2	3	2	2
$x_5$	3	2	3	2	2	3	3	3	2	2	4	3	3	4	3
$x_6$	3	2	2	1	1	2	2	2	1	1	2	2	2	1	1
$x_7$	4	3	3	2	3	3	3	3	2	2	3	2	3	2	2
$x_8$	3	2	3	2	2	2	2	2	1	1	2	2	2	1	1
$x_9$	3	2	2	1	1	2	2	2	1	1	2	2	2	1	1

由表 1 可得下述结果:

$$U/R_{1A} = \{\{x_1, x_2, x_3\}, \{x_4, x_7\}, \{x_5, x_8\}, \{x_6, x_9\}\} = \{E_{11}, E_{12}, E_{13}, E_{14}\}$$

$$U/R_{2A} = \{\{x_1, x_2\}, \{x_3, x_5, x_7\}, \{x_4\}, \{x_6, x_8, x_9\}\} = \{E_{21}, E_{22}, E_{23}, E_{24}\}$$

$$U/R_{3A} = \{\{x_1, x_5\}, \{x_2, x_3\}, \{x_4, x_7\}, \{x_6, x_8, x_9\}\} = \{E_{31}, E_{32}, E_{33}, E_{34}\},$$

$$U/R_{1d} = \{\{x_1, x_2, x_3, x_4, x_7\}, \{x_5, x_8\}, \{x_6, x_9\}\}$$

$$U/R_{2d} = \{\{x_1, x_2, x_4\}, \{x_3, x_5, x_7\}, \{x_6, x_8, x_9\}\}$$

$$U/R_{3d} = \{\{x_1, x_2, x_3, x_5\}, \{x_4, x_7\}, \{x_6, x_8, x_9\}\}$$

于是可得  $R_{1A} \subseteq R_{1d}$ ,  $R_{2A} \subseteq R_{2d}$  和  $R_{3A} \subseteq R_{3d}$ . 因此, 根据定义 3 可知表 1 是协调多源决策表.

## 2 两类多源决策规则及其支持度、覆盖度

在协调决策表中, 通过对比分析条件属性和决策属性之间的关系, 可以得到决策规则如下.

**定义 4**<sup>[20]</sup> 设  $DT = (U, A, F, d)$  为协调决策表. 对于任意  $x \in U$ , 必有  $[x]_d$  存在, 使得  $[x]_A \subseteq [x]_d$ , 于是得到决策规则  $r$ , 即:

$$\text{If } \bigwedge_{a_i \in A} (a_i, f_i(x)), \text{ then } d = f_d(x)$$

简记为:

$$(f_1(x), f_2(x), \dots, f_{|A|}(x)) \Rightarrow d = f_d(x)$$

对于决策规则  $r$ , 记其支持度为  $\mu(r)$ , 即  $\mu(r) = \frac{|[x]_A \cap [x]_d|}{|U|}$ , 其前提记作  $\text{pre}(r)$ , 结论

记作  $\text{con}(r)$ , 所有决策规则构成的集合记作  $\Omega$ .

此外, 由定义 4 可知, 在协调决策表中, 决策规则的个数等于条件等价类的个数, 即  $|\Omega| = |U/R_A|$ . 由此可以得出结论: 协调决策表中信息系统的划分约简不仅可以简化决策规则的前件, 还可以保持该决策表的决策规则个数不变.

记  $DT_k$  的所有决策规则构成的集合为  $\Omega_k$ , 称  $\Omega_M = \bigcup_{k=1}^s \Omega_k$  为  $MDT$  的决策规则集. 也就是说, 称所有数据源中的决策规则构成的集合为多源决策表的决策规则集.

**例 2** (续例 1) 针对表 1 的协调多源决策表, 每一个决策表  $DT_k (k = 1, 2, 3)$  的决策规则集及相应的支持度如下所示.

$DT_1$  中的决策规则及相应的支持度为:

$$r_{11}:(3,2,2,1) \Rightarrow d=1, \mu(r_{11}) = \frac{2}{9}$$

$$r_{12}:(3,2,3,2) \Rightarrow d=2, \mu(r_{12}) = \frac{2}{9}$$

$$r_{13}:(4,3,3,2) \Rightarrow d=3, \mu(r_{13}) = \frac{2}{9}$$

$$r_{14}:(4,3,3,3) \Rightarrow d=3, \mu(r_{14}) = \frac{3}{9}$$

$DT_2$ 中的决策规则及相应的支持度为:

$$r_{21}:(2,2,2,1) \Rightarrow d=1, \mu(r_{21}) = \frac{3}{9}$$

$$r_{22}:(3,2,3,2) \Rightarrow d=2, \mu(r_{22}) = \frac{1}{9}$$

$$r_{23}:(3,3,3,2) \Rightarrow d=2, \mu(r_{23}) = \frac{3}{9}$$

$$r_{24}:(4,3,3,2) \Rightarrow d=3, \mu(r_{24}) = \frac{2}{9}$$

$DT_3$ 中的决策规则及相应的支持度为:

$$r_{31}:(2,2,2,1) \Rightarrow d=1, \mu(r_{31}) = \frac{3}{9}$$

$$r_{32}:(3,2,3,2) \Rightarrow d=2, \mu(r_{32}) = \frac{2}{9}$$

$$r_{33}:(4,3,3,3) \Rightarrow d=3, \mu(r_{33}) = \frac{2}{9}$$

$$r_{34}:(4,3,3,4) \Rightarrow d=3, \mu(r_{34}) = \frac{2}{9}$$

于是,MDT的决策规则集 $\Omega_M$ 为:

$$r_1:(2,2,2,1) \Rightarrow d=1$$

$$r_2:(3,2,2,1) \Rightarrow d=1$$

$$r_3:(3,2,3,2) \Rightarrow d=2$$

$$r_4:(3,3,3,2) \Rightarrow d=2$$

$$r_5:(4,3,3,2) \Rightarrow d=3$$

$$r_6:(4,3,3,3) \Rightarrow d=3$$

$$r_7:(4,3,3,4) \Rightarrow d=3$$

由例2的结果可以看出,多源决策表的决策规则与决策表的决策规则在结构上没有任何区别,这使得多源数据源头多的这一特点并未充分的体现在获取的决策规则中.

为解决这一问题,针对协调多源决策表,分别从数据源和决策规则的结论出发,对其决策规则集进行融合,给出第一类多源决策规则和第二类多源决策规则的概念.

## 2.1 第一类多源决策规则

**定义5** 设MDT是协调多源决策表, $\Omega_M$ 是其决策规则集, $\mathcal{F} = \{F_k | k=1, 2, \dots, s\}$ 是其数据源集合, $N \subseteq \mathcal{F}, N \neq \emptyset$ . 若对任意的 $F_h \in N$ ,都有

$\Omega_M$ 中的决策规则 $(f_1(x), f_2(x), \dots, f_{|A|}(x)) \Rightarrow d = f_d(x)$ 成立,则记:

$$(f_1(x), f_2(x), \dots, f_{|A|}(x)) \stackrel{N}{\Rightarrow} d = f_d(x)$$

进一步,若对任意的 $F_z \in \mathcal{F} - N$ ,有

$$(f_1(x), f_2(x), \dots, f_{|A|}(x)) \stackrel{N \cup F_z}{\Rightarrow} d = f_d(x)$$

不成立,则称 $(f_1(x), f_2(x), \dots, f_{|A|}(x)) \stackrel{N}{\Rightarrow} d = f_d(x)$ 为第一类多源决策规则,记作 $r_{\text{pre}}^M$ .

第一类多源决策规则实质上是在多源决策表的决策规则中增加了数据源的信息,该类多源决策规则 $(f_1(x), f_2(x), \dots, f_{|A|}(x)) \stackrel{N}{\Rightarrow} d = f_d(x)$ 可解释为:决策规则 $(f_1(x), f_2(x), \dots, f_{|A|}(x)) \Rightarrow d = f_d(x)$ 在 $N \subseteq \mathcal{F}$ 中的每一个数据源下都成立,且 $N$ 是最大的满足此条件的数据源子集.

称MDT的所有第一类多源决策规则构成的集合为第一类多源决策规则集,记作 $\Omega_M^I$ . 由定义5可知MDT的第一多源决策规则的个数与其决策规则的个数相同,即 $|\Omega_M^I| = |\Omega_M|$ . 于是, $\Omega_M^I$ 可表示为:

$$\Omega_M^I = \left\{ \text{pre}(r) \stackrel{N}{\Rightarrow} \text{con}(r) \mid r \in \Omega_M, N \subseteq \mathcal{F} \right\}$$

由于第一类多源决策规则融入了数据源的信息,比决策规则包含的知识更为丰富,从而此类多源决策规则的度量方式也应当加入数据源这个因素. 根据第一类多源决策规则和决策规则之间的关系,将决策规则 $(f_1(x), f_2(x), \dots, f_{|A|}(x)) \Rightarrow d = f_d(x)$ 在 $N \subseteq \mathcal{F}$ 的各个数据源中支持度的最大值定义为第一类多源决策规则的支持度,将决策规则在所有数据源 $\mathcal{F}$ 中出现的频率定义为其覆盖度. 具体的定义形式如下.

**定义6** 设MDT是协调多源决策表, $\Omega_M^I$ 是其第一类多源决策规则集, $N \subseteq \mathcal{F}$ . 则:

(1)  $\forall \text{pre}(r) \stackrel{N}{\Rightarrow} \text{con}(r) \in \Omega_M^I$ , 定义其支持度为:

$$\mu(\text{pre}(r) \stackrel{N}{\Rightarrow} \text{con}(r)) = \max_{i \leq |N|} \{\mu_i(r) : r \in \Omega_M\}$$

其中 $\mu_i(r)$ 是决策规则 $r$ 在第 $i$ 个数据源中的支持度;

(2)  $\forall \text{pre}(r) \Rightarrow \text{con}(r) \in \Omega_M^1$ , 定义其覆盖度为:

$$\psi(\text{pre}(r) \Rightarrow \text{con}(r)) = \frac{|N|}{|\mathcal{F}|}$$

**例 3**(续例 2) 结合例 2 的结果, 根据定义 5 可得表 1 所示协调多源决策表的第一类多源决策规则为:

$$r_{1 \text{ pre}}^M: (2, 2, 2, 1) \xrightarrow{(F_2, F_3)} d = 1$$

$$r_{2 \text{ pre}}^M: (3, 2, 2, 1) \xrightarrow{(F_1)} d = 1$$

$$r_{3 \text{ pre}}^M: (3, 2, 3, 2) \xrightarrow{\mathcal{F}} d = 2$$

$$r_{4 \text{ pre}}^M: (3, 3, 3, 2) \xrightarrow{(F_2)} d = 2$$

$$r_{5 \text{ pre}}^M: (4, 3, 3, 2) \xrightarrow{(F_1, F_2)} d = 3$$

$$r_{6 \text{ pre}}^M: (4, 3, 3, 3) \xrightarrow{(F_1, F_3)} d = 3$$

$$r_{7 \text{ pre}}^M: (4, 3, 3, 4) \xrightarrow{(F_3)} d = 3$$

进一步, 根据定义 6, 可得每一个第一类多源决策规则的支持度和覆盖度分别为:

$$\mu(r_{1 \text{ pre}}^M) = \frac{3}{9}, \psi(r_{1 \text{ pre}}^M) = \frac{2}{3}$$

$$\mu(r_{2 \text{ pre}}^M) = \frac{2}{9}, \psi(r_{2 \text{ pre}}^M) = \frac{1}{3}$$

$$\mu(r_{3 \text{ pre}}^M) = \frac{2}{9}, \psi(r_{3 \text{ pre}}^M) = 1$$

$$\mu(r_{4 \text{ pre}}^M) = \frac{3}{9}, \psi(r_{4 \text{ pre}}^M) = \frac{1}{3}$$

$$\mu(r_{5 \text{ pre}}^M) = \frac{2}{9}, \psi(r_{5 \text{ pre}}^M) = \frac{2}{3}$$

$$\mu(r_{6 \text{ pre}}^M) = \frac{3}{9}, \psi(r_{6 \text{ pre}}^M) = \frac{2}{3}$$

$$\mu(r_{7 \text{ pre}}^M) = \frac{2}{9}, \psi(r_{7 \text{ pre}}^M) = \frac{1}{3}$$

**2.2 第二类多源决策规则** 接下来, 以决策规则的结论为切入点, 给出第二类多源决策规则的定义. 为了叙述方便, 在此先给出以下记号.

设  $\Omega_M$  是协调多源决策表  $MDT$  的决策规则集, 记:

$$R_{d(v)} = \{(r_p, r_q) \mid \text{con}(r_p) = \text{con}(r_q), r_p, r_q \in \Omega_M\}$$

$$[r_p]_{d(v)} = \{r_q \mid (r_p, r_q) \in R_{d(v)}\}$$

$$\Omega_M / R_{d(v)} = \{[r_p]_{d(v)} \mid r_p \in \Omega_M\}$$

则  $R_{d(v)}$  是  $\Omega_M$  的一个等价关系,  $[r_p]_{d(v)}$  是其等价

类,  $\Omega_M / R_{d(v)}$  是  $\Omega_M$  的一个划分. 其中  $d(v)$  表示决策属性  $d$  的取值为  $v$ .

**定义 7**<sup>[19]</sup> 设  $MDT$  是协调多源决策表,  $\Omega_M$  为其决策规则集.  $\forall r_p \in \Omega_M$ , 称:

$$\left( \bigvee_{r_q \in [r_p]_{d(v)}} \text{pre}(r_q) \right) \Rightarrow d = v$$

为协调多源决策表的第二类多源决策规则, 记作  $r_{d(v)}^M$ .

由定义 7 可知, 第二类多源决策规则是依据决策属性的取值将多源决策表的部分决策规则的前件通过逻辑“或”运算“ $\vee$ ”进行融合而得到的.

第二类多源决策规则  $\left( \bigvee_{r_q \in [r_p]_{d(v)}} \text{pre}(r_q) \right) \Rightarrow d = v$  可

解释为:  $[r_p]_{d(v)}$  中所有决策规则的前提, 其结论都为  $d = v$ .

称  $MDT$  的所有第二类多源决策规则构成的集合为第二类多源决策规则集, 记作  $\Omega_M^{\text{II}}$ . 则由定义 7 可知  $|\Omega_M^{\text{II}}| = |V_d|$ . 将  $\Omega_M^{\text{II}}$  表示为:

$$\Omega_M^{\text{II}} = \left\{ \left( \bigvee_{r_q \in [r_p]_{d(v)}} \text{pre}(r_q) \right) \Rightarrow d = v \mid r_p, r_q \in \Omega_M, v \in V_d \right\}$$

第二类多源决策规则反映了多源决策表中决策规则的共性, 结合该特点, 将  $[r_p]_{d(v)}$  中决策规则支持度的最大值定义为第二类多源决策规则  $r_{d(v)}^M$  的支持度, 将  $[r_p]_{d(v)}$  中决策规则在全体数据源中出现频率的最大值定义为  $r_{d(v)}^M$  的覆盖度. 下面给出第二类多源决策规则集的支持度和覆盖度的概念.

**定义 8** 设  $MDT$  是协调多源决策表,  $\Omega_M^{\text{II}}$  是其第二类多源决策规则集. 则:

$$(1) \forall \left( \bigvee_{r_q \in [r_p]_{d(v)}} \text{pre}(r_q) \right) \Rightarrow d = v \in \Omega_M^{\text{II}}$$

定义其支持度为:

$$\mu \left( \left( \bigvee_{r_q \in [r_p]_{d(v)}} \text{pre}(r_q) \right) \Rightarrow d = v \right) = \max_{j \leq |[r_p]_{d(v)}|} \left\{ \mu^j(r_q) : r_q \in [r_p]_{d(v)} \right\}$$



其中,  $\mu^j(r_q)$  是  $[r_p]_{d(v)}$  中第  $j$  个决策规则  $r_q$  的支持度.

$$(2) \forall \left( \bigvee_{r_q \in [r_p]_{d(v)}} \text{pre}(r_q) \right) \Rightarrow d = v \in \Omega_M^{\Pi}$$

定义其覆盖度为:

$$\psi \left( \left( \bigvee_{r_q \in [r_p]_{d(v)}} \text{pre}(r_q) \right) \Rightarrow d = v \right) = \max \left\{ \frac{|H_q|}{|\mathcal{F}|} : r_q \in [r_p]_{d(v)} \right\}$$

其中  $H_q = \{k \mid r_q \in \Omega_k\}$  是具有决策规则  $r_q$  的数据源集合.

**例4(续例2)** 针对表1的协调多源决策表, 由例2和定义7可得第二类多源决策规则集  $\Omega_M^{\Pi}$  为:

$$r_{d(1)}^M: (2, 2, 2, 1) \vee (3, 2, 2, 1) \Rightarrow d=1$$

$$r_{d(2)}^M: (3, 2, 3, 2) \vee (3, 3, 3, 2) \Rightarrow d=2$$

$$r_{d(3)}^M: (4, 3, 3, 3) \vee (4, 3, 3, 2) \vee (4, 3, 3, 4) \Rightarrow d=3$$

根据定义8可得第二类多源决策规则的支持度和覆盖度如下所示:

$$\mu(r_{d(1)}^M) = \frac{3}{9}, \psi(r_{d(1)}^M) = \frac{2}{3}$$

$$\mu(r_{d(2)}^M) = \frac{3}{9}, \psi(r_{d(2)}^M) = 1$$

$$\mu(r_{d(3)}^M) = \frac{3}{9}, \psi(r_{d(3)}^M) = \frac{2}{3}$$

**2.3 两类多源决策规则之间的联系** 第一类多源决策规则与决策规则相比增加了数据源的信息, 而第二类多源决策规则与决策规则相比前件包含的信息量更多. 除此之外, 这两类多源决策规则之间具有下述的关系.

**定理1** 设  $MDT$  是协调多源决策表,  $\Omega_M^I$  是其第一类多源决策规则集.  $\forall r_{\text{pre}}^M \in \Omega_M^I$ , 记:

$$Y_{d(v)} = \{ \text{pre}(r_{\text{pre}}^M) \mid \text{con}(r_{\text{pre}}^M) = (d=v) \}$$

则一定存在  $r_{d(v)}^M \in \Omega_M^{\Pi}$ , 有  $(\bigvee Y_{d(v)}) = \text{pre}(r_{d(v)}^M)$ .

**证明**  $\forall r_{\text{pre}}^M \in \Omega_M^I$ ,  $\exists r_q \in \Omega_M$ , 由定义5知  $\text{pre}(r_{\text{pre}}^M) = \text{pre}(r_q)$ ,  $\text{con}(r_{\text{pre}}^M) = \text{con}(r_q)$ . 因此可得:

$$Y_{d(v)} = \{ \text{pre}(r_q) \mid \text{con}(r_q) = (d=v) \}$$

从而有:

$$Y_{d(v)} = \{ \text{pre}(r_q) \mid r_q \in [r_p]_{d(v)} \}$$

进一步, 根据定义7可得:

$$(\bigvee Y_{d(v)}) \Rightarrow d = v$$

为第二类多源决策规则, 即:

$$(\bigvee Y_{d(v)}) = \text{pre}(r_{d(v)}^M)$$

定理1表明可以通过第一类多源决策规则得到第二类多源决策规则.

根据多源决策表的两类多源决策规则之间的联系以及它们与决策规则之间的关系, 下面给出两类多源决策规则的另一种获取方法.

### 3 两类多源决策规则的获取及其简化

**3.1 两类多源决策规则的获取** 从多源信息系统  $MIS = \{IS_k \mid IS_k = (U, A, F_k), k=1, 2, \dots, s\}$  的结构上分析, 其可以看作是将构成  $MIS$  的  $s$  个信息系统  $IS_k = (U, A, F_k)$  按水平方向放置得到的数据表. 如果将这  $s$  个信息系统按竖直方向放置, 并将  $IS_k$  中的对象集记为  $\{(k, x) \mid x \in U\}$ , 则  $IS_k$  可表示为  $(\{(k, x) \mid x \in U\}, A, F)$ . 继而, 该多源信息系统可表示为:

$$\left( \bigcup_{k=1}^s \{(k, x) \mid x \in U\}, A, F \right)$$

其中,

$$F = \left\{ f_i: \bigcup_{k=1}^s \{(k, x) \mid x \in U\} \rightarrow V_i \mid i \leq |A|, a_i \in A \right\}$$

该数据表的结构与信息系统的一致, 称该数据表为  $MIS$  的单源化信息系统, 记为  $M_{IS}$ , 并称这一过程为多源信息系统的单源化.

对于任意一个多源信息系统, 都可以转化为一个与其对应的信息系统. 类似地, 对于任意一个多源决策表, 也都可以转化为一个与其对应的决策表. 为了便于叙述, 下面给出单源化决策表的形式化定义.

**定义9** 设  $MDT = \{DT_k \mid DT_k = (U, A \cup d, F_k, d_k), k=1, 2, \dots, s\}$  是多源决策表,  $(\dot{U}, A, F, d)$  是决策表, 其中,

$$\dot{U} = \bigcup_{k=1}^s \{(k, x) \mid x \in U\}, \dot{U} \rightarrow V_d$$

为关系集. 若  $\forall x \in U, \exists k \in \{1, 2, \dots, s\}$ , 使  $f_{kl}(x) = f_l((k, x))(a_l \in A)$ ,  $d_k(x) = d((k, x))$ , 则称  $(\dot{U}, A, F, d)$  为  $MDT$  的单源化决策表, 记为:

$$M_{DT} = \left( \bigcup_{k=1}^s \{(k, x) | x \in U\}, A, F, d \right)$$

由定义 2 和定义 9 可得, 对任意  $k \in \{1, 2, \dots, s\}$ , 有:

$$\left\{ f_{kl}: U \rightarrow V_l (l \leq |A|, a_l \in A) \right\} = \left\{ f_l: \bigcup_{k=1}^s \{(k, x) | x \in U\} \rightarrow V_l (l \leq |A|, a_l \in A) \right\}$$

即:  $\mathcal{F} = F$ .

记单源化决策表  $M_{DT}$  的条件属性集对应的划分为  $\dot{U}/R_A = \{E_i | i \in \tau\}$ , 决策属性对应的划分为  $\dot{U}/R_d = \{H_w | w \in \sigma\}$ , 其中  $\tau$  和  $\sigma$  为指标集.

由单源化决策表的定义可知: 若  $MDT$  是协调多源决策表, 则  $M_{DT}$  为协调决策表; 反之亦然. 因此, 可以通过判断单源化决策表  $M_{DT}$  的协调性得到多源决策表  $MDT$  的协调性.

**例 5**(续例 1) 考查表 1 的多源决策表, 其单源化决策表如表 2 所示.

由表 2 可得,  $\dot{U}/R_A$  中的元素分别为:

$$E_1 = \{(1, x_6), (1, x_9)\}$$

$$E_2 = \{(2, x_6), (2, x_8), (2, x_9), (3, x_6), (3, x_8), (3, x_9)\}$$

$$E_3 = \{(1, x_5), (1, x_8), (2, x_4), (3, x_4), (3, x_7)\},$$

$$E_4 = \{(2, x_3), (2, x_5), (2, x_7)\},$$

$$E_5 = \{(1, x_1), (1, x_2), (1, x_3), (3, x_2), (3, x_3)\}$$

$$E_6 = \{(1, x_4), (1, x_7), (2, x_1), (2, x_2)\}$$

$$E_7 = \{(3, x_1), (3, x_5)\}.$$

$\dot{U}/R_d$  中的元素分别为:

$$H_1 = \{(1, x_6), (1, x_9), (2, x_6), (2, x_8), (2, x_9), (3, x_6), (3, x_8), (3, x_9)\}$$

$$H_2 = \{(1, x_5), (1, x_8), (2, x_3), (2, x_4), (2, x_5), (2, x_7), (3, x_4), (3, x_7)\}$$

$$H_3 = \{(1, x_1), (1, x_2), (1, x_3), (1, x_4), (1, x_7), (2, x_1), (2, x_2), (3, x_1), (3, x_2), (3, x_3), (3, x_5)\}$$

由此可以判定表 2 所示的决策表是协调的.

**表 2** 表 1 的单源化决策表  $M_{DT}$

**Table 2** A unisource decision table  $M_{DT}$  of Table 1

	$a_1$	$a_2$	$a_3$	$a_4$	$d$
$(1, x_1)$	4	3	3	3	3
$(1, x_2)$	4	3	3	3	3
$(1, x_3)$	4	3	3	3	3
$(1, x_4)$	4	3	3	2	3
$(1, x_5)$	3	2	3	2	2
$(1, x_6)$	3	2	2	1	1
$(1, x_7)$	4	3	3	2	3
$(1, x_8)$	3	2	3	2	2
$(1, x_9)$	3	2	2	1	1
$(2, x_1)$	4	3	3	2	3
$(2, x_2)$	4	3	3	2	3
$(2, x_3)$	3	3	3	2	2
$(2, x_4)$	3	2	3	2	2
$(2, x_5)$	3	3	3	2	2
$(2, x_6)$	2	2	2	1	1
$(2, x_7)$	3	3	3	2	2
$(2, x_8)$	2	2	2	1	1
$(2, x_9)$	2	2	2	1	1
$(3, x_1)$	4	3	3	4	3
$(3, x_2)$	4	3	3	3	3
$(3, x_3)$	4	3	3	3	3
$(3, x_4)$	3	2	3	2	2
$(3, x_5)$	4	3	3	4	3
$(3, x_6)$	2	2	2	1	1
$(3, x_7)$	3	2	3	2	2
$(3, x_8)$	2	2	2	1	1
$(3, x_9)$	2	2	2	1	1

对于协调单源化决策表, 由于  $\dot{U}/R_A$  中每一个元素对应一条决策规则. 因此, 若记  $r_E^i$  为等价类  $E_i$  对应的决策规则, 则  $E_i$  与  $r_E^i$  之间是一一对应关系. 于是, 记  $M_{DT}$  的决策规则集为  $\Omega_E$ , 且将  $\Omega_E$  可表示为:

$$\Omega_E = \left\{ r_E^i \mid E_i \in \dot{U}/R_A, i \leq \left| \dot{U}/R_A \right| \right\}$$

下面给出协调多源决策表的决策规则与其单源化决策表的决策规则之间的联系.

**定理 2** 设  $M_{DT}$  是协调多源决策表  $MDT$  的单源化决策表. 则有  $\Omega_E = \Omega_M$ .

**证明** 根据定义 4 可知, 协调决策表中每一个条件等价类对应一条决策规则. 因此, 对于协调多源决策表  $MDT$  中的每一个决策表  $DT_k = (U, A \cup d, F_k, d_k)$ , 均有:

$$\Omega_k = \left\{ r_{kj}^i \mid E_{kj} \in U/R_{kA}, j \leq |U/R_{kA}| \right\}$$

其中,  $r_{kj}^i$  是由第  $j$  个条件等价类  $E_{kj}$  确定的决策规则, 继而可得:

$$\Omega_M = \bigcup_{k=1}^s \Omega_k = \bigcup_{k=1}^s \left\{ r_{kj}^i \mid E_{kj} \in U/R_{kA}, j \leq |U/R_{kA}| \right\}$$

又由定义2和定义9可得  $\mathcal{F} = F$ , 即:

$$\left\{ f_{kl}: U \rightarrow V_l \mid l \leq |A|, a_l \in A \right\} = \left\{ f_i: \bigcup_{k=1}^s \{(k, x) \mid x \in U\} \rightarrow V_l \mid l \leq |A|, a_l \in A \right\}$$

且有:

$$\Omega_E = \left\{ r_E^i \mid E_i \in \dot{U}/R_A, i \leq |\dot{U}/R_A| \right\}$$

因此可知:

$$\bigcup \left\{ \text{pre}(r_{kj}^i) \mid r_{kj}^i \in \Omega_M \right\} = \bigcup \left\{ \text{pre}(r_E^i) \mid r_E^i \in \Omega_E \right\}$$

从而可证  $\Omega_E = \Omega_M$ .

定理2表明可由协调多源决策表  $MDT$  的单源化决策表  $M_{DT}$  获取其决策规则集.

结合定理2以及第一类多源决策规则的定义, 可得下述由协调多源决策表的单源化决策表获取第一类多源决策规则的方法.

**定理3** 设:

$$M_{DT} = \left( \bigcup_{k=1}^s \{(k, x) \mid x \in U\}, A, F, d \right)$$

是协调多源决策表  $MDT$  的单源化决策表,  $\dot{U}/R_A$  是其条件属性集对应的划分,  $\Omega_E$  是其决策规则集. 定义映射  $g: \dot{U}/R_A \rightarrow \mathcal{F}$  如下:

$$g(E_i) = \left\{ F_k \mid (k, x) \in E_i, \forall E_i \in \dot{U}/R_A \right\}$$

则有:

$$\Omega_M^1 = \left\{ \text{pre}(r_E^i) \Rightarrow \text{con}(r_E^i) \mid E_i \in \dot{U}/R_A, i \leq |\dot{U}/R_A| \right\}$$

其中,  $\mathcal{F} = \{F_k \mid k = 1, 2, \dots, s\}$  为  $MDT$  的数据源集合.

**证明** 由定理2知  $\Omega_E = \Omega_M$ , 即对任意  $r_E^i \in \Omega_E$ , 存在  $r \in \Omega_M$ , 使得  $\text{pre}(r) = \text{pre}(r_E^i)$ ,  $\text{con}(r) = \text{con}(r_E^i)$ , 进而由  $|\Omega_M| = |\Omega_M^1|$ , 可得:

$$|\Omega_M^1| = |\Omega_E| = |\dot{U}/R_A|$$

进一步, 由映射  $g$  的定义以及等价类与决策规则的关系可知,  $g(E_i)$  是最大的使得

$\text{pre}(r_E^i) \Rightarrow \text{con}(r_E^i)$  成立的数据源子集, 因此, 根据定义5可知  $\text{pre}(r_E^i) \Rightarrow \text{con}(r_E^i)$  是第一类多源决策规则. 从而可证:

$$\Omega_M^1 = \left\{ \text{pre}(r_E^i) \Rightarrow \text{con}(r_E^i) \mid E_i \in \dot{U}/R_A, i \leq |\dot{U}/R_A| \right\}$$

此外, 通过结合定理1和定理2, 可得由协调多源决策表的单源化决策表获取第二类多源决策规则的方法如下.

**定理4** 设:

$$M_{DT} = \left( \bigcup_{k=1}^s \{(k, x) \mid x \in U\}, A, F, d \right)$$

是协调多源决策表  $MDT$  的单源化决策表,  $\Omega_E$  是  $M_{DT}$  的决策规则集.  $\forall v \in V_d$ , 定义:

$$Y_{d(v)} = \left\{ \text{pre}(r_E^i) \mid \text{con}(r_E^i) = d(v) \right\}$$

则有:

$$\Omega_M^{\text{II}} = \left\{ \left( \bigvee Y_{d(v)} \right) \Rightarrow d = v \mid v \in V_d \right\}$$

**证明** 由定理1和定理2易证.

**例6** (续例5) 对于表2的单源化决策表, 其等价类所对应的数据表如表3所示.

表3 表2的简化表

Table 3 A simplified table of Table 2

$\dot{U}/R_A$	$a_1$	$a_2$	$a_3$	$a_4$	$d$
$E_1$	2	2	2	1	1
$E_2$	3	2	2	1	1
$E_3$	3	2	3	2	2
$E_4$	3	3	3	2	2
$E_5$	4	3	3	2	3
$E_6$	4	3	3	3	3
$E_7$	4	3	3	4	3

根据表3, 由条件等价类与决策规则之间的联系, 易得表2单源化决策表的决策规则为:

$$r_E^1: (2, 2, 2, 1) \Rightarrow d=1$$

$$r_E^2: (3, 2, 2, 1) \Rightarrow d=1$$

$$r_E^3: (3, 2, 3, 2) \Rightarrow d=2$$

$$r_E^4: (3, 3, 3, 2) \Rightarrow d=2$$

$$r_E^5: (4, 3, 3, 2) \Rightarrow d=3$$

$$r_E^6: (4, 3, 3, 3) \Rightarrow d=3$$

$$r_E^7: (4, 3, 3, 4) \Rightarrow d=3$$



该结果与例 2 中多源决策表的决策规则集是完全相同的. 由此也验证了定理 2 的结论.

此外, 由映射  $g$  可得:

$$g(E_1) = \{F_2, F_3\}$$

$$g(E_2) = \{F_1\}$$

$$g(E_3) = \{F_1, F_2, F_3\}$$

$$g(E_4) = \{F_2\}$$

$$g(E_5) = \{F_1, F_2\}$$

$$g(E_6) = \{F_1, F_3\}$$

$$g(E_7) = \{F_3\}$$

从而根据定理 3 便可得到第一类多源决策规则集, 其结果与例 3 一致.

因  $V_d = \{1, 2, 3\}$ , 故根据定理 4 可得:

$$Y_{d(1)} = \{(2, 2, 2, 1), (3, 2, 2, 1)\}$$

$$Y_{d(2)} = \{(3, 2, 3, 2), (3, 3, 3, 2)\}$$

$$Y_{d(3)} = \{(4, 3, 3, 2), (4, 3, 3, 3), (4, 3, 3, 4)\}$$

进而可得到第二类多源决策规则集, 其结果与例 4 一致.

**3.2 两类多源决策规则的简化** 在决策表中, 通过删除冗余的条件属性可以获取简洁、紧凑的决策规则. 因此, 针对多源决策表, 讨论两类多源决策规则的简化问题也非常必要.

下面以保持多源决策表决策规则的恒真性及其个数不变为目标, 对其前件进行简化. 在本文中称其为保持决策规则不变的约简.

**定义 10** 设:

$$MDT =$$

$$\{DT_k | DT_k = (U, A \cup d, F_k, d_k), k = 1, 2, \dots, s\}$$

是多源决策表.  $\forall k \in \{1, 2, \dots, s\}$ , 若存在  $B \subseteq A$ , 使得  $R_{kB} \subseteq R_{kA}$  且  $|\Omega_{kB}| = |\Omega_{kA}|$ , 则称  $B$  为  $MDT$  的保持决策规则不变的协调集. 进一步, 若  $B$  为保持决策规则不变的协调集, 且  $B$  的任何真子集都不是保持决策规则不变的协调集, 则称  $B$  为  $MDT$  的保持决策规则不变的约简. 其中  $\Omega_{kB}$  是  $(U, B \cup d, F_k, d_k)$  中决策规则的集合.

对于多源决策表:

$$MDT =$$

$$\{DT_k | DT_k = (U, A \cup d, F_k, d_k), k = 1, 2, \dots, s\}$$

为方便描述与分析, 给出下述一些记号.

$\forall B \subseteq A$ , 记  $\Omega_{M_B}$  为由属性子集  $B$  得到的决策规则集. 类似地, 记  $\Omega_{M_B}^I$  和  $\Omega_{M_B}^{II}$  分别为由属性子集  $B$  得到的第一类和第二类多源决策规则集. 对于  $\Omega_{M_B}$  中的任意一条决策规则:

$$\text{If } \bigwedge_{a_l \in B} (a_l, f_l(x)), \text{ then } d = f_d(x)$$

简记为:

$$(f_{l_1}(x), f_{l_2}(x), \dots, f_{l_{|B|}}(x))_{l_1 l_2 \dots l_{|B|}} \Rightarrow d = f_d(x)$$

其中, 序列  $l_1 l_2 \dots l_{|B|}$  是  $B$  中属性的下角标对应的数值. 例如, 取  $B = \{a_1, a_2, a_4\}$ , 则决策规则  $(a_1, 2) \wedge (a_2, 2) \wedge (a_4, 1) \Rightarrow d = 1$  简记为  $(2, 2, 1)_{124} \Rightarrow d = 1$ . 类似地, 对于  $\Omega_{M_B}^I$  和  $\Omega_{M_B}^{II}$  中的多源决策规则, 也有相同的解释.

事实上, 根据协调多源决策表  $MDT$  的定义和多源信息系统  $MIS$  划分约简的定义可知,  $MDT$  的保持决策规则不变的约简就是其  $MIS$  的保持划分不变的约简. 进一步, 根据  $MDT$  的两类多源决策规则与其决策规则之间的关系, 给出这样的定义: 若  $B \subseteq A$  为  $MDT$  保持决策规则不变的约简, 则称  $B$  为  $MDT$  的保持第一类和第二类多源决策规则不变的约简.

由于这两类多源决策规则是在多源决策表的决策规则集的基础上得到的, 而多源决策表与其单源化决策表具有相同的决策规则集, 所以借助单源化决策表中信息系统的划分约简, 易得这两类多源决策规则的简化方法如下.

**定理 5** 设:

$$M_{DT} = \left( \bigcup_{k=1}^s \{(k, x) | x \in U\}, A, F, d \right)$$

是协调多源决策表  $MDT$  的单源化决策表,  $B \subseteq A$  是

$$M_{IS} = \left( \bigcup_{k=1}^s \{(k, x) | x \in U\}, A, F \right)$$

的划分约简. 则  $B$  为  $MDT$  的保持第一类和第二类多源决策规则不变的约简.

**例 7** (续例 6) 对于表 2,  $B = \{a_1, a_2, a_4\}$  是其信息系统  $M_{IS}$  的划分约简. 因此,  $B = \{a_1, a_2, a_4\}$  也是保持第一类和第二类多源决策规则不变的约简. 根据表 3, 属性子集  $B$  所对应的简化表格如表 4 所示.

表4 由 $B$ 产生的 $U$ 的划分Table 4 A partition of  $U$  based on  $B$ 

$\dot{U}/R_B$	$a_1$	$a_2$	$a_4$	$d$
$E_1$	2	2	1	1
$E_2$	3	2	1	1
$E_3$	3	2	2	2
$E_4$	3	3	2	2
$E_5$	4	3	2	3
$E_6$	4	3	3	3
$E_7$	4	3	4	3

简化的决策规则为:

$$(2, 2, 1)_{124} \Rightarrow d = 1, (3, 2, 1)_{124} \Rightarrow d = 1,$$

$$(3, 2, 2)_{124} \Rightarrow d = 2, (3, 3, 2)_{124} \Rightarrow d = 2,$$

$$(4, 3, 3)_{124} \Rightarrow d = 3, (4, 3, 2)_{124} \Rightarrow d = 3,$$

$$(4, 3, 4)_{124} \Rightarrow d = 3,$$

进而可得简化的第一类多源决策规则为:

$$r_{1\text{pre}}^{M_B}: (2, 2, 1)_{124} \xrightarrow{(F_2, F_3)} d = 1$$

$$r_{2\text{pre}}^{M_B}: (3, 2, 1)_{124} \xrightarrow{(F_1)} d = 1$$

$$r_{3\text{pre}}^{M_B}: (3, 2, 2)_{124} \xrightarrow{F} d = 2$$

$$r_{4\text{pre}}^{M_B}: (3, 3, 2)_{124} \xrightarrow{(F_2)} d = 2$$

$$r_{5\text{pre}}^{M_B}: (4, 3, 2)_{124} \xrightarrow{(F_1, F_2)} d = 3$$

$$r_{6\text{pre}}^{M_B}: (4, 3, 3)_{124} \xrightarrow{(F_1, F_3)} d = 3$$

$$r_{7\text{pre}}^{M_B}: (4, 3, 4)_{124} \xrightarrow{(F_3)} d = 3$$

简化的第二类多源决策规则为:

$$r_{d(1)}^{M_B}: (2, 2, 1)_{124} \vee (3, 2, 1)_{124} \Rightarrow d = 1$$

$$r_{d(2)}^{M_B}: (3, 2, 2)_{124} \vee (3, 3, 2)_{124} \Rightarrow d = 2$$

$$r_{d(3)}^{M_B}: (4, 3, 3)_{124} \vee (4, 3, 2)_{124} \vee (4, 3, 4)_{124} \Rightarrow d = 3$$

需要说明的是,一般情况下信息系统的划分约简不唯一,因此通过删除两类多源决策规则前件中的冗余属性,不仅可以简化规则,还可以增加规则的个数,为决策行为提供更多的有效依据.

## 4 结 论

本文借助多源决策表数据源头多的特点,分别以数据源和决策规则的结论为切入点,提出了第一类和第二类多源决策规则的概念,给出了这

两类多源决策规则支持度和覆盖度的度量方法以及两者之间的联系.最后,通过单源化多源决策表,研究了获取两类多源决策规则的方法和保持决策规则不变的约简.后续将进一步基于单源化多源决策表研究不协调多源决策表中的知识获取以及知识融合问题.

## 参考文献

- [1] Pawlak Z. Rough sets. International Journal of Computer & Information Science, 1982, 11(5): 341—356.
- [2] 常犁云,王国胤,吴淦.一种基于 Rough Set 理论的属性约简及规则提取方法.软件学报,1999,10(11): 1206—1211. (Chang L Y, Wang G Y, Wu Y. An approach for attribute reduction and rule generation based on rough set theory. Journal of Software, 1999, 10(11): 1206—1211.)
- [3] Kryszkiewicz M. Rules in incomplete information systems. Information Sciences, 1999, 113(3—4): 271—292.
- [4] 黄兵,周献中.不一致决策表中规则提取的矩阵算法.系统工程与电子技术,2005,27(3): 441—445. (Huang B, Zhou X Z. Matrix computation for rule extraction in inconsistent decision tables. Systems Engineering and Electronics, 2005, 27(3): 441—445.)
- [5] 钱宇华,梁吉业,曹付元等.决策表决策规则与知识粒度.电脑开发与应用,2006,19(3): 27—29. (Qian Y H, Liang J Y, Cao F Y, et al. Decision rule in decision table and knowledge granulation. Computer Development & Applications, 2006, 19(3): 27—29.)
- [6] Hao C, Li J H, Fan M, et al. Optimal scale selection in dynamic multi-scale decision tables based on sequential three-way decisions. Information Sciences, 2017, 415—416: 213—232.
- [7] Chen H M, Li T R, Luo C, et al. A rough set-based method for updating decision rules on attribute values' coarsening and refining. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(12): 2886—2899.
- [8] Wu W Z, Qian Y H, Li T J, et al. On rule acquisition in incomplete multi-scale decision tables. Information Sciences, 2017, 378: 282—302.
- [9] Zhang X, Mei C L, Chen D G, et al. Multi-confidence rule acquisition and confidence-preserved attribute

- reduction in interval - valued decision systems. *International Journal of Approximate Reasoning*, 2014, 55(8):1787—1804.
- [10] Khan M A, Banerjee M. Formal reasoning with rough sets in multiple-source approximation systems. *International Journal of Approximate Reasoning*, 2008, 49(2):466—477.
- [11] Qian Y H, Liang J Y, Yao Y Y, et al. MGRS: a multi-granulation rough set. *Information Sciences*, 2010, 180(6):949—970.
- [12] Qian Y H, Li S Y, Liang J Y, et al. Pessimistic rough set based decisions: a multigranulation fusion strategy. *Information Sciences*, 2014, 264:196—210.
- [13] Lin G P, Liang J Y, Qian Y H. An information fusion approach by combining multigranulation rough sets and evidence theory. *Information Sciences*, 2015, 314:184—199.
- [14] Che X Y, Mi J S, Chen D G. Information fusion and numerical characterization of a multi - source information system. *Knowledge - Based Systems*, 2018, 145:121—133.
- [15] Che X Y, Mi J S. Attributes set reduction in multigranulation approximation space of a multi - source decision information system. *International Journal of Machine Learning and Cybernetics*, 2019, 10(9):2297—2311.
- [16] Xu W H, Yu J H. A novel approach to information fusion in multi-source datasets: a granular computing viewpoint. *Information Sciences*, 2017, 378:410—423.
- [17] Guo Y T, Xu W H. Attribute reduction in multi-source decision systems//Flores V. *Rough Sets*. Springer Berlin Heidelberg, 2016:558—568.
- [18] 林国平, 梁吉业, 李进金. 多源决策信息系统的决策规则性能评价. *模式识别与人工智能*, 2015, 28(7):657—664. (Lin G P, Liang J Y, Li J J. Evaluation of decision rules performance for multi-source decision information systems. *Pattern Recognition and Artificial Intelligence*, 2015, 28(7):657—664.)
- [19] 万青, 马盈仓, 魏玲. 基于多粒度的多源数据知识获取. *山东大学学报(理学版)*, 2020, 55(1):41—50. (Wan Q, Ma Y C, Wei L. Knowledge acquisition of multi-source data based on multigranularity. *Journal of Shandong University (Natural Science)*, 2020, 55(1):41—50.)
- [20] 张文修, 仇国芳. 基于粗糙集的不确定决策. 北京: 清华大学出版社, 2005:66—67.

(责任编辑 杨可盛)