

## 置信规则库规则约简的粗糙集方法

王应明<sup>1a</sup>, 杨隆浩<sup>1b</sup>, 常雷雷<sup>2</sup>, 傅仰耿<sup>1b</sup>

(1. 福州大学 a. 决策科学研究所, b. 数学与计算机科学学院, 福州 350116; 2. 国防科技大学 信息系统与管理学院, 长沙 410073)

**摘要:** 针对置信规则中规则数的“组合爆炸”问题, 目前的解决方法主要是基于特征提取的规则约简方法, 有效性依赖于专家知识. 鉴于此, 提出基于粗糙集理论的无需依赖规则库以外知识的客观方法, 按照等价类划分思想逐条分析置信规则, 进而消除冗余的候选值. 最后, 以装甲装备能力评估作为实例进行分析, 分别从规则约简数、决策准确性方面与具有代表性的主观方法进行对比, 结果表明, 所提出方法是有效可行的, 且优于现有规则约简主观方法.

**关键词:** 置信规则库; 粗糙集; 规则约简; 主观方法; 客观方法

**中图分类号:** TP28

**文献标志码:** A

### Rough set method for rule reduction in belief rule base

WANG Ying-ming<sup>1a</sup>, YANG Long-hao<sup>1b</sup>, CHANG Lei-lei<sup>2</sup>, FU Yang-geng<sup>1b</sup>

(1a. Institute of Decision Sciences, 1b. School of Mathematics and Computer Science, Fuzhou University, Fuzhou 350116, China; 2. College of Information System and Management, National University of Defense Technology, Changsha 410073, China. Correspondent: FU Yang-geng, E-mail: ygfu@qq.com)

**Abstract:** The number of rules in belief rule base(BRB) may induce the problem of combinatorial explosion. However, most previous works on rule reduction are based on feature extraction, whose effectiveness depends on the expert knowledge. Therefore, an objective method based on rough set theory is proposed, which does not depends on any knowledge in addition to belief rule base. The method of rule reduction analyzes each belief rule according to equivalence class division thought, and then eliminates the redundancy of referential values. Finally, a numerical case study to evaluate armored system is analyzed and compared with the typical subjective method in the number of reduced rule and the accuracy of decision-making. The results show that the proposed method is feasible and effective, and superior to the existing subjective method of rule reduction.

**Key words:** belief rule base; rough set; rule reduction; subjective method; objective method

## 0 引言

置信规则库(BRB)系统由传统的If-Then规则库系统<sup>[1]</sup>扩展而来, 由于在If-Then规则中加入了置信框架, BRB系统不仅具有处理定性知识和定量信息的能力, 还能够在信息处理过程中保留信息的不确定性部分. BRB系统的核心是基于证据推理的置信规则库推理方法<sup>[2]</sup>(RIMER), 该方法在D-S证据理论<sup>[3-4]</sup>、决策理论<sup>[5]</sup>、模糊理论<sup>[6]</sup>和传统If-Then规则库<sup>[7]</sup>的基础上发展起来, 具有对带有含糊或模糊不确定性、不完整性或概率不确定性以及非线性特征的数据进行建模的能力. 目前, BRB系统已成功应用于工程系统安全评估<sup>[7]</sup>、石墨成分检测<sup>[8]</sup>、管线检漏<sup>[9]</sup>等领域.

RIMER包括知识表示和知识推理两部分, 其中知识表示主要依靠置信规则库中的置信规则, 因此构建一个健壮的置信规则库是搭建BRB系统过程中不可或缺的步骤. 然而, 置信规则库的构建需要覆盖所有前提属性和各属性的候选值, 导致置信规则库中的置信规则数成指数量级的趋势增长, 当实际问题过于复杂时, 易引发“组合爆炸”问题<sup>[10]</sup>. 目前, 解决该问题主要采用基于特征提取的规则约简方法, 如: 周志杰等<sup>[11-12]</sup>提出了“统计效用”的概念进行规则约简, 首先使用“统计效用”评价每一条规则, 然后根据评价的值判断哪些规则需要保留, 最后依此分析各前提属性的关键候选值, 最终达到规则约简的目的.

收稿日期: 2013-10-15; 修回日期: 2014-03-29.

基金项目: 国家杰出青年科学基金项目(70925004); 国家自然科学基金面上项目(71371053); 国家自然科学基金青年基金项目(61300026); 福建省教育厅科技项目(JA13036); 福州大学科技发展基金项目(2014-XQ-26).

作者简介: 王应明(1964—), 男, 教授, 博士生导师, 从事决策理论与方法、数据包络分析等研究; 杨隆浩(1990—), 男, 硕士生, 从事智能决策技术的研究.

常雷雷等<sup>[13]</sup>引入灰靶 (GT)、多维尺度 (MDS)、等距映射 (Isomap)、主成分分析 (PCA) 等特征提取的方法进行规则约简, 首先邀请专家对各前提属性作出评价, 然后使用上述特征提取的方法对评价信息进行处理, 从而筛选出关键前提属性, 最后由关键前提属性重构规模缩小的置信规则库. 上述约简方法均有效地缩减了置信规则库中规则数, 但呈现出共同的局限性, 即方法的有效性依赖于额外的信息或专家的知识. 由于主观因素占主导地位, 本文将该类方法称为规则约简主观方法.

鉴于此, 本文基于粗糙集 (RS) 理论<sup>[14-15]</sup>提出新的规则约简方法, 该方法在规则约简中相对上述规则约简主观方法的主要区别在于, 无需提供规则库以外的任何先验知识, 规则约简时主要利用规则库上的等价关系对规则的不确定程度进行度量, 在保持分类能力不变的前提下, 消除规则库中不必要的规则, 导出最终的决策或分类规则. 由于规则约简是依据置信规则中已有的规则间等价关系, 与规则约简主观方法形成鲜明的对比, 将此类方法称为规则约简客观方法. 在实验分析中, 引入装甲装备体系综合能力评估<sup>[13]</sup>的置信规则库作为分析实例, 并将基于 PCA 的规则约简方法作为比较对象, 根据约简后置信规则库中的规则数和参与激活规则合成的结果集数验证所提出规则约简方法的有效性, 分别以特殊方案和一般方案作为实例输入, 结果表明, 所提出的规则约简方法能够保证 BRB 系统推理的准确性.

## 1 RIMER 方法和问题描述

RIMER 方法主要包括构建置信规则库、计算激活权重和利用证据推理 (ER) 算法合成激活规则 3 部分. 在置信规则库的构建中, 由于置信规则数量与前提属性数量和候选值数量成正相关, 当实际问题过于复杂时, 容易出现“组合爆炸”问题, 现有的解决方法主要是基于特征提取的规则约简主观方法.

### 1.1 置信规则库的表示

规则库是专家系统的一个重要组成部分. 设一个规则库模型表示为

$$R = \langle U, A, D, F \rangle. \quad (1)$$

其中:  $U = \{U_i; i = 1, 2, \dots, T\}$  为前提属性集合;  $A = \{A_1, A_2, \dots, A_T\}$  为前提属性的候选值集合, 候选值集合中每个元素具体表示为  $A_i = \{A_{i,j}; j = 1, 2, \dots, J_i = |A_i|\}$ ,  $A_{i,j}$  为前提属性  $U_i$  的第  $j$  个候选值;  $D = \{D_n; n = 1, 2, \dots, N\}$  为规则库中的结果集;  $F$  为逻辑函数, 反映前提属性与结果集的逻辑关系.

一般地, 传统规则库中的第  $k$  条规则表示为

$$R_k : \text{If } A_1^k \wedge A_2^k \wedge \dots \wedge A_{T_k}^k, \\ \text{Then } D_n, k = 1, 2, \dots, L. \quad (2)$$

在传统规则的基础上, 杨剑波等<sup>[2]</sup>引入置信框架, 提出置信规则的表示形式, 允许置信规则中保留不确定性的信息, 使规则在表示知识时更加贴近实际. 置信规则表示为

$$R_k : \text{If } A_1^k \wedge A_2^k \wedge \dots \wedge A_{T_k}^k, \\ \text{Then } \{(D_1, \bar{\beta}_{1,k}), (D_2, \bar{\beta}_{2,k}), \dots, (D_N, \bar{\beta}_{N,k})\}. \quad (3)$$

其中: 规则权重为  $\theta_k$ , 前提属性权重集合为

$$\delta = \{\delta_{k,1}, \delta_{k,2}, \dots, \delta_{k,T_k}; k = 1, 2, \dots, L\}.$$

在置信规则中, 当表示的知识完整时,  $\sum_{i=1}^N \bar{\beta}_{i,k} = 1$ ; 当

表示的知识不完整时,  $\sum_{i=1}^N \bar{\beta}_{i,k} < 1$ .

### 1.2 RIMER 算法的主要公式

RIMER 算法主要包含计算激活权重和利用 ER 算法合成激活规则两步. 激活权重的计算取决于输入数据、前提属性权重和规则权重, 因此, 置信规则库中第  $k$  条规则的激活权重计算为

$$\omega_k = \frac{\theta_k \prod_{i=1}^{T_k} (\alpha_i^k)^{\delta_{k,i}}}{\sum_{l=1}^L \left( \theta_l \prod_{i=1}^{T_l} (\alpha_i^l)^{\delta_{l,i}} \right)}, \\ \bar{\delta}_{k,i} = \frac{\delta_{k,i}}{\max_{j=1,2,\dots,T_k} \{\delta_{k,j}\}}, \quad (4)$$

其中  $\alpha_i^k$  为第  $k$  条规则中第  $i$  个前提属性的输入数据. 由于输入数据可能不完整, 被激活规则结果集的置信度需要进一步修正, 第  $k$  条规则结果集中第  $i$  个等级的置信度修正公式为

$$\beta_{i,k} = \bar{\beta}_{i,k} \frac{\sum_{t=1}^{T_k} \left( \tau(t, k) \sum_{j=1}^{J_t} \alpha_{t,j} \right)}{\sum_{t=1}^{T_k} \tau(t, k)}. \\ \tau(t, k) = \begin{cases} 1, & U_t \in R_k, t = 1, 2, \dots, T_k; \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

利用上述公式可以求得全部激活规则的激活权重和修正的结果集置信度, 每条被激活的规则可视为 ER 方法中的一个基本属性, 因此可用 ER 算法<sup>[16]</sup>将所有激活规则进行合成. 以下简要介绍 ER 算法.

在 ER 算法中, 由每个基本属性的置信度和属性权重可以求得基本属性的基本可信值为

$$m_{n,i} = \omega_i \beta_{n,i}, \quad n = 1, 2, \dots, N, \quad i = 1, 2, \dots, L;$$

$$m_{H,i} = 1 - \omega_i \sum_{n=1}^N \beta_{n,i} = \bar{m}_{H,i} + \tilde{m}_{H,i};$$

$$\tilde{m}_{H,i} = \omega_i \left( 1 - \sum_{n=1}^N \beta_{n,i} \right);$$

$$\bar{m}_{H,i} = 1 - \omega_i. \quad (6)$$

在基本可信值的基础上, 利用解析公式<sup>[17]</sup>将所有的基本属性一次合成, 合成公式为

$$C_n = k \left[ \prod_{j=1}^L (m_{n,j} + \bar{m}_{H,j} + \tilde{m}_{H,j}) - \prod_{j=1}^L (\bar{m}_{H,j} + \tilde{m}_{H,j}) \right],$$

$$\tilde{C}_H = k \left[ \prod_{j=1}^L (\bar{m}_{H,j} + \tilde{m}_{H,j}) - \prod_{j=1}^L \bar{m}_{H,j} \right],$$

$$\bar{C}_H = k \prod_{j=1}^L \bar{m}_{H,j},$$

$$k^{-1} = \sum_{n=1}^N \prod_{j=1}^L (m_{n,j} + \bar{m}_{H,j} + \tilde{m}_{H,j}) - (N-1) \prod_{j=1}^L (\bar{m}_{H,j} + \tilde{m}_{H,j}). \quad (7)$$

求得综合属性的置信度为

$$\beta_n = \frac{C_n}{1 - \bar{C}_H}, \quad n = 1, 2, \dots, N;$$

$$\beta_H = \frac{\tilde{C}_H}{1 - \bar{C}_H}. \quad (8)$$

经过 ER 算法合成激活规则后得到的置信度即可作为 BRB 系统的输出, 其中第  $i$  个置信规则库相关的置信度分布输出为

$$\text{BRB}_i : \{(D_1, \beta_{1,i}), (D_2, \beta_{2,i}), \dots, (D_N, \beta_{N,i})\}. \quad (9)$$

为了直观地分析同一输入数据在不同置信规则库中推理所得置信度分布的差异, 将置信度分布转化为平均效用值和相似度. 对于平均效用值, 假设  $N$  个评价等级的等级效用值满足  $\mu_1 \leq \mu_2 \leq \dots \leq \mu_N$ , 则第  $i$  个置信规则库的平均效用值表示为

$$\mu_{avg,i} = \sum_{n=1}^N (\mu_n \beta_{n,i}) + \frac{(\mu_1 + \mu_N)}{2} \left( 1 - \sum_{n=1}^N \beta_{n,i} \right). \quad (10)$$

对于相似度, 假设有两个置信规则库  $\text{BRB}_i$  和  $\text{BRB}_j$ , 在两个置信规则库上推理所得的置信度分布的相似程度为

$$\phi_{i,j} = 1 - \sum_{k=1}^N (\beta_{k,j} - \beta_{k,i})^2 - \left( \sum_{k=1}^N (\beta_{k,i} - \beta_{k,j}) \right)^2. \quad (11)$$

### 1.3 BRB 的“组合爆炸”问题和规则约简主观方法

置信规则库在构建时需要覆盖所有的前提属性和各属性的候选值, 因此置信规则库中的规则数可达  $\prod_{i=1}^T J_i$ . 例如, 当前提属性的个数为 5、各前提属性候选值的个数均为 3 时, 所构建置信规则库的规则数为  $3^5 = 243$  条. 可见, 置信规则库中的规则数是一个指数量级的规模, 因此, 当问题过于复杂时容易引起“组合爆炸”问题.

解决“组合爆炸”问题最直接、有效的方式是减少前提属性或候选值的个数, 依据文献[18]对关键性能参数的介绍可知, 置信规则库中的每个前提属性或候选值并非都是必不可少的, 因此, 通过筛选关键的前提属性或候选值重建置信规则库, 可以有效地缩减置信规则库中的规则数. 受此启发, 人们提出了基于特征提取的规则约简方法, 首先利用专家的知识确定前提属性的评价矩阵, 然后利用特征提取方法对评价矩阵进行处理, 从而得到关键前提属性, 最后由关键前提属性重构规模缩小的置信规则库. 该方法在缩减规则数和保证系统推理的准确性方面均取得了理想的约简效果, 但实际上方法的有效性依赖于额外信息或专家知识, 由于主观因素占主导地位, 本文将其归为规则约简主观方法.

规则约简主观方法与多属性决策问题中的主观赋权方法<sup>[19]</sup>类似, 其共同点是需要邀请专家根据自己的经验和对实际的主观判断给出对属性的评价信息, 若选取的专家不同, 则得出的评价信息也不相同, 因此该方法的最大缺点在于主观随意性大. 此外, 规则约简主观方法在约简置信规则库时是相对于整体约简某个前提属性, 因此规则约简的过程中容易产生不一致的置信规则, 例如, 在置信规则库中存在如下两条置信规则:

$$R_k : \text{If } (C_3 = A) \wedge (C_4 = A) \wedge (C_5 = A),$$

$$\text{Then } \{(A, 1.0), (B, 0.0), (C, 0.0)\}; \quad (12a)$$

$$R_t : \text{If } (C_3 = A) \wedge (C_4 = A) \wedge (C_5 = B),$$

$$\text{Then } \{(A, 0.8), (B, 0.2), (C, 0.0)\}. \quad (12b)$$

假设由规则约简主观方法可确定  $C_5$  为非关键属性, 则约简前提属性  $C_5$  后可使置信规则库中出现两条不一致的置信规则, 即两条置信规则的前提属性候选值相同, 但结果集中的置信度分布不相同, 该类型置信规则所表示的知识彼此相互矛盾, 对 BRB 系统决策结果的影响是不可预估的. 在规则约简主观方法中, 对于不一致置信规则的处理方式仅是合并为如下形式的一致性置信规则:

$$R_k : \text{If } (C_3 = A) \wedge (C_4 = A), \\ \text{Then } \{(A, 1.0), (B, 0.0), (C, 0.0)\} \vee \\ \{(A, 0.8), (B, 0.2), (C, 0.0)\}. \quad (12c)$$

规则的合并虽然有效地缩减了置信规则库中的规则数,但由于结果集中包含两种可能,在使用该置信规则与其他置信规则作比较时,需要依次比较置信规则中不同的结果集,只有结果集均完全相同,两条置信规则才是相同的.同样地,当该置信规则被激活时,需要将两个结果集参与到置信规则的合成中.由此可见,基于特征提取的规则约简主观方法实质上并未有效约简置信规则库.

## 2 规则约简的粗糙集方法

现有的规则约简方法中,主要是基于特征提取的规则约简主观方法,目前该方法用于解决置信规则库的“组合爆炸”问题略显成效,但在实际使用中存在诸多局限性和缺点.鉴于此,本文基于粗糙集理论提出新的规则约简方法.

### 2.1 基于粗糙集理论规则约简方法

粗糙集理论是 Pawlak<sup>[14-15]</sup>提出的一种新的数学工具,由于它能够使用严谨的数学公式分析模糊和不确定问题,粗糙集理论在机器学习和模式识别等领域得到了广泛应用.在粗糙集理论中,知识约简是其研究的核心内容之一,主要包括判断近似空间中的每个等价关系是否都是必要的和如何删除不必要的知识.基于此功能本文提出新的规则约简方法,并以粗糙集理论中的属性值约简算法作为新规则约简方法的算法核心.为了方便叙述,以下简称基于粗糙集理论的约简方法为 RS 约简方法.

RS 约简方法与规则约简主观方法的主要区别在于前者无需提供规则库以外的任何先验知识,规则约简时主要利用规则库上的等价关系对规则的不确定程度进行度量,在保持分类能力不变的前提下,约简每条置信规则中冗余的候选值,以达到精简置信规则库的目的.由于 RS 约简方法在约简置信规则库时受人为因素的影响较小,可将 RS 约简方法归为规则约简的客观方法.此外,RS 约简方法与规则约简主观方法在处理置信规则库中冗余规则的方式上也大相径庭.RS 约简方法在规则约简过程中相对于置信规则库,未约简任一前提属性,仅是约简置信规则中非关键候选值.非关键候选值被约简后会致置信规则库中产生重复的置信规则,进而可通过合并这些重复置信规则达到缩减置信规则库中规则数的目的,其中合并重复置信规则是 RS 约简方法能够约简置信规则库的主要原因.

### 2.2 RS 约简方法的步骤

RS 约简方法的核心是粗糙集理论中的属性值约简算法<sup>[20]</sup>,对置信规则库进行约简时的具体算法步骤如下.

**Step 1** 检查置信规则库中是否有不一致置信规则,若存在,则合并为如式(12c)所示的一致性置信规则.

**Step 2** 对置信规则库中每条置信规则的候选值进行考察.由于置信规则库的每条置信规则中均包含同一前提属性的候选值或空值,在置信规则库中可由候选值对应的前提属性确定属性列.考察候选值时假定删除相关属性列的情况如下:

- 1) 若产生不一致置信规则,则保留不一致置信规则的原候选值,该候选值表示不可约简,同时该候选值对应的前提属性为当前置信规则的核值属性,为方便叙述,将  $R_i$  规则的核值属性集记为  $\text{CoreValue}(R_i)$ ;
- 2) 若未产生不一致置信规则并且含有重复置信规则,则将重复置信规则的候选值标为“null”,该候选值表示可以约简;
- 3) 若未产生不一致置信规则并且不含重复置信规则,则将候选值标为“unknown”,表示该候选值暂时无法判定是否可以约简.

**Step 3** 首先遍历置信规则库删除重复的置信规则,然后依次考察每条置信规则中标为“unknown”的候选值,考察时分如下 3 种处理情况:

- 1) 若当前置信规则中未被标记的候选值可判别出唯一结果集,则将置信规则中“unknown”改为“null”;
- 2) 在条件 1) 不成立且只有一个“unknown”的情况下,将“unknown”改为原候选值;
- 3) 在条件 1) 和条件 2) 均不成立的情况下,作如下处理:

① 初始化集合变量  $Y$  和  $Z$ . 有

$$Y = \{U_k \in U | f(R_i, U_k) = \text{“unknown”}\}, \\ Z = \text{CoreValue}(R_i) \cup Y. \quad (13)$$

计算  $Y$  中各前提属性的粗糙隶属度  $\mu$ 、关键系数  $\lambda_1$  和  $\lambda_2$ , 有

$$\mu_D^{\{a\}}(R) = \frac{|[R]_{\{a\}} \cap [R]_D|}{|[R]_{\{a\}}|}, a \in Y; \\ \lambda_1^{\{a\}} = |[R]_{U-\{a\}} - [R]_D|, a \in Y; \\ \lambda_2^{\{a\}} = |[R]_{\{a\}} - [R]_D|, a \in Y. \quad (14)$$

将  $Y$  中的前提属性按照参数的优先级从高到低进行排序,其中参数间优先级从高到低依次为粗糙隶属度  $\mu$ 、关键系数  $\lambda_1$ 、关键系数  $\lambda_2$ . 在参数值的优先级中,粗糙隶属度  $\mu$  和关键系数  $\lambda_1$  属于数值大的优先

级高, 关键系数  $\lambda_2$  属于数值小的优先级高, 若依旧无法判别前提属性间的先后顺序, 则采取随机排序的方式.

②从  $Y$  中逐个取出前提属性, 记为  $a$ . 若  $Z - \{a\}$  与  $Z$  两个集合的粗糙隶属度相同, 则将置信规则中隶属于前提属性  $a$  的候选值改为“null”, 并从  $Y$  中取出下一个前提属性赋予  $a$ , 继续执行上述步骤. 若  $Z - \{a\}$  与  $Z$  两个集合的粗糙隶属度不相同, 则将置信规则中剩余的标有“unknown”的候选值改为原候选值. 完成对  $Y$  中属性的操作后, 检查置信规则库, 删除候选值均被标为“null”和重复的置信规则.

**Step 4** 当两条置信规则仅有一处候选值不同, 且其中一条置信规则的候选值被标为“null”时, 作如下处理:

1) 对于候选值被标为“null”的置信规则, 如果可由未被标为“null”的候选值判断出唯一的结果集, 则删除候选值未被标为“null”的置信规则;

2) 否则, 删除候选值被标为“null”的置信规则.

经上述算法步骤后, 可以约简每条置信规则中冗余的候选值, 进而建立新的置信规则库. 属性值约简算法的流程如图 1 所示.

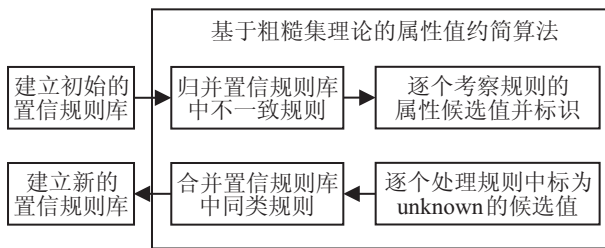


图 1 属性值约简算法流程

### 2.3 算法复杂度分析

下面对所提出算法的计算复杂度进行分析, 以此表明 RS 约简方法良好的计算性能. Step 1 的对比置信规则库中任意两条置信规则的前提属性候选值和结果集置信度, 时间复杂度为  $O(LL(T + N))$ . Step 2 逐列预处理置信规则库内的每条置信规则, 时间复杂度为  $O(TLL(T + N))$ . Step 3 可分为 3 部分: 第 1 部分为约简重复置信规则, 时间复杂度为  $O(LL(T + N))$ ; 第 2 部分为计算粗糙隶属度, 时间复杂度为  $O(TLL(T + N))$ ; 第 3 部分为比较前提属性间的优先级, 时间

复杂度为  $O((LL(T + N) + T + T \log T))$ . Step 4 处理每条置信规则内标为“null”的候选值, 计算复杂度为  $O(LL(T + N))$ . 因此, 综合上述各步骤的时间复杂度可知, RS 约简方法的总时间复杂度为  $O(TLL(T + N))$ .

## 3 实验分析和性能比较

为了检验 RS 约简方法在实际应用中的适用性, 本文引入一个装甲装备体系综合能力评估的置信规则库<sup>[13]</sup>, 目前该置信规则库主要用于验证规则约简主观方法的约简效果. 装甲装备体系综合能力评估的置信规则库含有 5 个前提属性, 分别为“指令与控制”、“机动性”、“火力”、“防御力”和“通讯”, 各前提属性简记为  $\{C_1, C_2, C_3, C_4, C_5\}$ . 假设每个前提属性的识别框架均为  $\{A, B, C\}$ , 由此可知装甲装备体系综合能力评估的置信规则库中共有规则数 243 条, 具体见文献 [13].

### 3.1 置信规则库的约简

利用 RS 约简方法对装甲装备体系综合能力评估的置信规则库作约简时, 可将被约简的置信规则依据结果集中置信度的分布分成两个类别, 第 1 类是置信度分布为非  $\{(A, 0.0)(B, 0.0)(C, 1.0)\}$  的第 1~72 条置信规则, 在约简时依照算法逐列分析每条置信规则的候选值, 判断其是否可以约简, 按照 Step 2 约简后还需作处理的置信规则如表 1 所示. 表 1 中, 置信规则编号为 21 与 59、24 与 44 的置信规则出现了重复的情况, 将每对置信规则合并为同一条置信规则; 编号为 21 和 23 的两条置信规则仅一处不相同且不相同属性列中的候选值为“null”, 可按照属性值约简算法的 Step 4 约简编号为 23 的置信规则.

第 73~243 条置信规则是被约简的第 2 类置信规则, 其置信度分布为  $\{(A, 0.0)(B, 0.0)(C, 1.0)\}$ , 同时这 171 条置信规则中存在一个隐性特点, 即仅有上述置信规则的前提属性  $C_1, C_2$  和  $C_3$  中包含候选值  $C$ . 因此, 在置信规则库中, 当结果集为  $\{(A, 0.0)(B, 0.0)(C, 1.0)\}$  且前提属性  $C_1, C_2$  和  $C_3$  中任意一个前提属性的候选值为  $C$  时, 可将置信规则中其他前提属性的候选值改为“null”, 完成属性值约简算法的 Step 2 后, 再去除重复的置信规则即可得到 3 条只有一个前提属性且前提属性各不相同的置信规则. 这 3

表 1 属性值约简后部分置信规则

编号	前提属性	结果集
21	$(C_1 = B) \wedge (C_2 = A) \wedge (C_3 = B) \wedge (C_4 = \text{null}) \wedge (C_5 = A)$	$\{(A, 0.4)(B, 0.6)(C, 0.0)\}$
23	$(C_1 = B) \wedge (C_2 = A) \wedge (C_3 = B) \wedge (C_4 = \text{null}) \wedge (C_5 = \text{null})$	$\{(A, 0.4)(B, 0.6)(C, 0.0)\}$
24	$(C_1 = B) \wedge (C_2 = A) \wedge (C_3 = B) \wedge (C_4 = B) \wedge (C_5 = \text{null})$	$\{(A, 0.4)(B, 0.6)(C, 0.0)\}$
44	$(C_1 = B) \wedge (C_2 = A) \wedge (C_3 = B) \wedge (C_4 = B) \wedge (C_5 = \text{null})$	$\{(A, 0.4)(B, 0.6)(C, 0.0)\}$
59	$(C_1 = B) \wedge (C_2 = A) \wedge (C_3 = B) \wedge (C_4 = \text{null}) \wedge (C_5 = A)$	$\{(A, 0.4)(B, 0.6)(C, 0.0)\}$

条置信规则可以简写为

$$\text{If } (C_1 = C) \vee (C_2 = C) \vee (C_3 = C), \\ \text{Then } \{(A, 0.0), (B, 0.0), (C, 1.0)\}. \quad (15)$$

式(15)与文献[13]中的第73条置信规则相同,进而验证了当前提属性“指挥与控制”、“机动性”、“火力”的候选值为 $C$ 时,可以直接判定结果集为 $\{(A, 0.0), (B, 0.0), (C, 1.0)\}$ 的规则特点<sup>[13]</sup>.经RS约简算法约简后,置信规则库中的规则数可以缩减至72条.

### 3.2 有效性对比分析

置信规则库的规则约简中,重构置信规则库中的规则数和参与激活规则合成的结果集数是衡量约简方法有效性的重要指标.本文为了直观地比较RS约简方法的有效性和相对于主观方法的优势,在对比初始置信规则库的同时,比较文献[13]中具有代表性的主观方法.由文献[13]中的分析可知,基于PCA的规则约简主观方法在将原置信规则库约简至4个前提属性时的推理准确性最为理想,因此,本文主要对比经PCA约简方法约简后由4个前提属性构成的置信规则库.为了叙述方便,以下简称经PCA约简方法约简后由4个前提属性构成的置信规则库为PCA-BRB,称经RS约简方法约简后的置信规则库为RS-BRB,称原置信规则库为O-BRB.

首先比较O-BRB、PCA-BRB和RS-BRB三者的规则数,如图2所示.PCA-BRB和RS-BRB与O-BRB相比,置信规则库中的规则数均明显被缩减,RS-BRB中的规则数比PCA-BRB少9条.由此可知,RS约简方法能够有效降低置信规则库中规则数的规模,且约简效果优于基于PCA的规则约简主观方法.

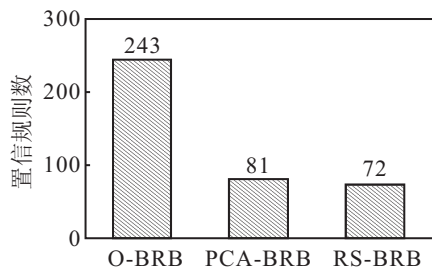


图2 不同置信规则库中规则数的对比

进一步分析RS-BRB中置信规则的组成成分,如图3所示.图3中,缺省规则数代表经RS约简方法约简后未发生任何改变的规则数.RS-BRB中缺省规则数为67条,占原有规则数的27.5%,由此可见,RS约简方法能够对置信规则库中绝大部分置信规则进行约简处理;在非缺省的置信规则中,有171条置信规则因被约简非关键候选值后与其他置信规则重复而被删除,由此可见,RS约简方法缩减置信规则库中规则数的效果明显.综上所述,粗糙集理论的引入能够

对置信规则库中绝大多数置信规则起到约简作用,进而有效地缩减了置信规则库中置信规则的数量,同时相比于主观约简方法,RS约简方法约简的效果更为理想.

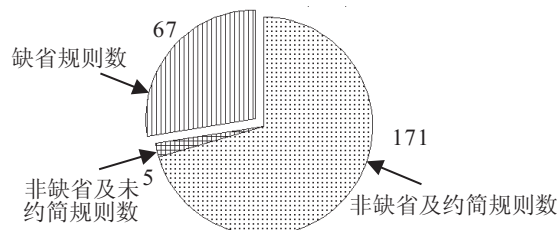


图3 RS-BRB中规则的组成成分

比较在O-BRB、PCA-BRB和RS-BRB中参与激活规则合成的结果集个数.首先假设5个前提属性中仅有A、B等级上有置信度,按照“指令与控制”、“机动性”、“火力”、“防御力”和“通讯”的顺序,依次改变输入数据的分布情况,即将原来只有A、B等级上有置信度转变为A、B、C等级上均有置信度,并在此基础上比较激活规则合成中参与合成的结果集个数,实验结果如图4所示.

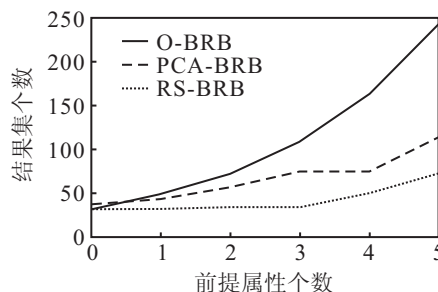


图4 激活规则合成中结果集个数

图4中,当前提属性个数为0,即输入数据未发生改进,仅在评价等级A、B上有置信度时,在O-BRB、PCA-BRB、RS-BRB上进行推理参与合成的结果集个数基本相同.随着评价等级C上置信度不为0的前提属性个数的增加,3种置信规则库中参与合成的结果集个数也增加,但始终保持O-BRB多于PCA-BRB、PCA-BRB多于RS-BRB,其中在RS-BRB中进行激活规则合成时,参与合成结果集个数增加的幅度最不明显.由此可见,使用RS约简方法不仅能够有效地约简置信规则数量,还能有效减少RIMER算法中的合成次数,降低ER算法的时间复杂度.

### 3.3 准确性对比分析

在BRB系统的置信规则库约简中,除了缩减置信规则的数量外,还需要保证BRB系统推理的准确性,本文主要从置信度分布、平均效用值和相似度等方面对比分析.对于平均效用值的计算,假定识别框架中各评价等级的等级效用值为

$$u = \{(A, 1.0), (B, 0.6), (C, 0.3)\}. \quad (16)$$

首先, 以特殊方案  $P$  作为实例输入, 有

$$P = \{(\text{指令与控制} = (A, 0.8), (B, 0.2)),$$

$$(\text{机动性} = (A, 0.7), (B, 0.3)),$$

$$(\text{火力} = (A, 0.9), (B, 0.1)),$$

$$(\text{防御力} = (A, 0.6), (B, 0.4)),$$

$$(\text{通讯} = (A, 0.7), (B, 0.3))\}.$$

依据方案  $P$ , 分别在 O-BRB、PCA-BRB 和 RS-BRB 上进行推理, 推理的结果如图 5 和图 6 所示.

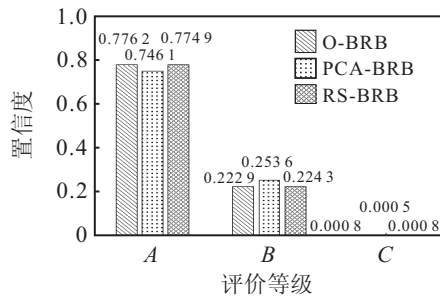


图 5 RS约简前后置信度分布的比较

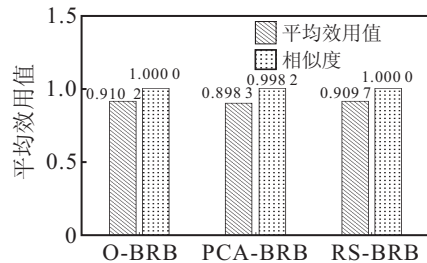


图 6 特例方案中效用值与相似度的比较

从置信度分布的角度分析, 置信规则库约简前后由 RIMER 推理所得的置信度分布基本相同, 且主要集中在评价等级  $A$  上, 且在该等级上, RS-BRB 与 O-BRB 相差的置信度为 0.0013, PCA-BRB 与 O-BRB 相差的置信度更加明显, 差值为 0.0301, 在其他等级上也同样表现出 RS-BRB 比 PCA-BRB 更加接近 O-BRB. 由此可知, 使用 RS 约简方法对置信规则库约简后能够使推理所得置信度保持较高的一致性.

从效用值和相似度的角度分析, 由图 6 可知, 置信规则库约简前平均效用值为 0.9102, 经过 RS 方法约简后平均效用值为 0.9097, 相比于 PCA-BRB 的 0.8983 具有更高的准确性. 在相似度的对比分析中, 由于仅保留小数点后 4 位有效数值, RS-BRB 计算所得的相似度近似为 1.0, PCA-BRB 只有 0.9982. 由相似度计算公式可知, 相似度反应置信规则库约简前后 BRB 系统推理所得置信度分布的相似程度, 因此, 进一步表明 RS 约简方法对 BRB 系统推理的准确性几乎未产生影响. 综上所述, RS 约简方法能够有效地保证 BRB 系统推理的准确性, 且相对于主观约简方法,

RS 约简方法能够确保被约简的置信规则库具有更高的准确性.

为了进一步验证 RS 约简方法能够保证 BRB 系统决策的准确性, 用一般方案作为实例输入, 分析时借鉴灵敏度分析方法<sup>[21]</sup>. 首先选取前提属性“指挥与控制”作为分析属性, 然后按照灵敏度分析中控制单一变量的方式, 将评价等级  $A$ 、 $B$  上的置信度按等比例减少, 并将减少的置信度转移至评价等级  $C$  上, 建立如下等式:

$$\hat{\beta}_A = \beta_A(1 - 0.1k),$$

$$\hat{\beta}_B = \beta_B(1 - 0.1k),$$

$$\hat{\beta}_C = (\beta_A + \beta_B)0.1k, \quad (17)$$

其中灵敏度因子  $k = 1, 2, \dots, 10$ . 分析时, 可在灵敏度因子  $k$  的取值范围内依次增大灵敏度因子  $k$  的值, 并依次比较 BRB 系统输出的平均效用值和相似度, 实验结果如图 7 所示. 由图 7 可见, 随着  $k$  值的增大, 在 3 个置信规则库中推理所得的平均效用值都呈下降趋势, 同时 RS-BRB 中推理所得的平均效用值比 PCA-BRB 更接近 O-BRB. 由图 7 中 3 条曲线可知, RS-BRB 中推理所得的置信度分布与 O-BRB 基本相同, 当  $\beta_C$  中的置信度接近 0.5 时, RS-BRB 与 O-BRB 的相差程度最大, 相比于规则约简主观方法, PCA-BRB 始终都表现出比 RS-BRB 更大的相异性. 由此可见, 对于一般方案, RS 约简方法同样能够有效地保证 BRB 系统推理的准确性, 且优于基于 PCA 的规则约简主观方法.

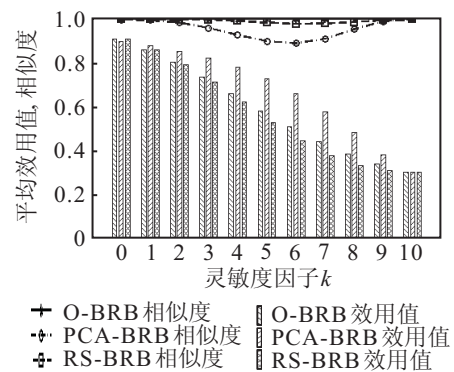


图 7 一般方案中效用值及相似度的比较

综上所述, 使用 RS 约简方法对置信规则库进行约简, 能够有效减少置信规则库中规则的数量, 保证 BRB 系统推理的准确性, 且相对于基于 PCA 的规则约简主观方法, 在规则的约简和准确性上都能取得更好的效果.

## 4 结 论

本文针对置信规则库中的“组合爆炸”问题和规则约简主观方法存在的不足, 提出基于粗糙集理论的规则约简客观方法. 该方法在约简过程中无需依赖规

则以外的任何先验知识,与现有规则约简主观方法形成鲜明对比.在实验中,将基于粗糙集理论的规则约简方法的有效性和准确性与具有代表性的主观方法进行比较分析,结果表明,所提出的约简方法无论从约简的规则数量还是推理的准确性,均优于现有的规则约简主观方法,同时能够减低 RIMER 方法推理的复杂度.下一步的研究方向是结合粗糙集理论,提出更具一般性的规则约简客观方法.

### 参考文献(References)

- [1] Sun Ron. Robust reasoning: Integration rule-based and similarity-based reasoning[J]. *Artificial Intelligence*, 1995, 75(2): 241-295.
- [2] Yang Jian-bo, Liu Jun, Wang Jin, et al. Belief rule-base inference methodology using the evidential reasoning approach-RIMER[J]. *IEEE Trans on Systems, Man and Cybernetics-Part A: Systems and Humans*, 2006, 36(2): 266-285.
- [3] Dempster A P. A generalization of bayesian inference[J]. *J of the Royal Statistical Society*, 1968, 30(2): 205-247.
- [4] Shafer Glenn. A mathematical theory of evidence[M]. Princeton: Princeton University Press, 1976: 1-314.
- [5] Hwang Ching-lai, Yoon Kwang-sun. Methods for multiple attribute decision making[M]. Berlin: Springer Berlin Heidelberg, 1981: 58-191.
- [6] Zadeh L Z. Fuzzy sets[J]. *Information and Control*, 1965, 8(3): 338-353.
- [7] Liu Jun, Yang Jian-bo, Ruan Da, et al. Self-tuning of fuzzy belief rule bases for engineering system safety analysis[J]. *Annals of Operations Research*, 2008, 163(1): 143-168.
- [8] Yang Jian-bo, Liu Jun, Xu Dong-ling, et al. Optimization model for training belief-rule-based systems[J]. *IEEE Trans on Systems, Man and Cybernetics, Part A: Systems and Humans*, 2007, 37(4): 569-585.
- [9] Zhou Zhi-jie, Hu Chang-hua, Yang Jian-bo, et al. Online updating belief-rule-based system for pipeline leak detection under expert intervention[J]. *Expert Systems with Applications*, 2009, 36(4): 7700-7709.
- [10] 姜江. 证据网络建模、推理及学习方法研究[D]. 长沙: 国防科技大学信息系统与管理学院, 2011. (Jiang J. Modeling, reasoning and learning approach to evidential network[D]. Changsha: Information System And Management College, National University of Defense Technology, 2011.)
- [11] 周志杰. 置信规则库在线建模方法与故障预测[D]. 西安: 西安科技大学高新学院, 2010. (Zhou Z J. Online modeling methods of belief rule base and failure prognosis[D]. Xi'an: Institute of High-Tech, Xi'an University of Science and Technology, 2010.)
- [12] 周志杰, 杨剑波, 胡昌华, 等. 置信规则库专家系统与复杂系统建模[M]. 北京: 科学出版社, 2011: 1-119. (Zhou Z J, Yang J B, Hu C H, et al. Belief rule base expert system and complex system modeling[M]. Beijing: Science Press, 2011: 1-119.)
- [13] Chang Lei-lei, Zhou Yu, Jiang Jiang, et al. Structure learning for belief rule base expert system: A comparative study[J]. *Knowledge-Based Systems*, 2013, 39(1): 159-172.
- [14] Pawlak Z. Rough sets[J]. *International J of Information and Computer Science*, 1982, 11(5): 341-356.
- [15] Pawlak Z. Rough set-theoretical aspects of reasoning about data[M]. Boston: Kluwer Academic Publishers, 1991: 1-248.
- [16] Yang Jian-bo, Xu Dong-ling. On the evidential reasoning algorithm for multiple attribute decision analysis under uncertainty[J]. *IEEE Trans on Systems, Man and Cybernetics, Part A: Systems and Humans*, 2002, 32(2): 289-304.
- [17] Wang Ying-ming, Yang Jian-bo, Xu Dong-ling, et al. The evidential reasoning approach for multiple attribute decision analysis using interval belief degrees[J]. *European J of Operational Research*, 2006, 175(1): 35-66.
- [18] Mcchrysal S A. Joint capabilities integration and development system[M]. Washington DC: Chairman of the Joint Chiefs of Staff Instruction, 2009: 1-30.
- [19] 金菊良, 魏一鸣. 复杂系统广义智能评价方法与应用[M]. 北京: 科学出版社, 2008: 1-246. (Jin J L, Wei Y M. Generalized intelligent assessment methods for complex systems and applications[M]. Beijing: Science Press, 2008: 1-246.)
- [20] 林嘉宜, 彭宏, 郑启伦. 一种新的基于粗糙集的值约简算法[J]. *计算机工程*, 2003, 29(4): 70-71. (Lin J Y, Peng H, Zheng Q L. A new algorithm for value reduction based on rough set[J]. *Computer Engineering*, 2003, 29(4): 70-71.)
- [21] Yang Jian-bo, Xu Dong-ling. Nolinear information aggregation via evidential reasoning in multiattribute decision analysis under uncertainty[J]. *IEEE Trans on Systems, Man and Cybernetics, Part A: Systems and Humans*, 2002, 32(3): 376-393.

(责任编辑: 郑晓蕾)