

文章编号: 1001-0920(2007)06-0657-06

# 基于粗集的多知识库决策融合

王加阳<sup>1</sup>, 王国仁<sup>2</sup>

(1. 中南大学 信息科学与工程学院, 长沙 410083; 2. 东北大学 信息科学与工程学院, 沈阳 110004)

**摘要:** 粗糙集理论研究的重要内容是分类与约简, 其目的在于获取优良的规则知识, 实现准确的决策. 通过提出规则集合的决策度量思想, 从整体上体现了对一个规则集合的衡量, 为基于多知识库的决策奠定了基础. 基于模型集成的基本思想, 将规则知识库作为一个决策模型, 根据规则集度量选择模型, 通过模型集成实现决策融合.

**关键词:** 粗糙集; 多知识库; 决策融合

**中图分类号:** TP18      **文献标识码:** A

## Decision fusion of redundant knowledge based on rough set

WANG Jia-yang<sup>1</sup>, WANG Guo-ren<sup>2</sup>

(1. College of Information Science and Engineering, Central South University, Changsha 410083, China; 2. College of Information Science and Engineering, Northeastern University, Shenyang 110004, China. Correspondent: WANG Jia-yang, E-mail: csuwjy@163.com)

**Abstract:** The important research contents of rough set are classification and reduction, the purpose of which is to get rule knowledge and realize accuracy decision. The metrics for rule set is presented and the properties of a rule set are showed in general, which play an important role for decision of redundant knowledge. Based on the basic theory of model integration rough set, a formal model for multi-repositories is provided. One repository is regarded as a decision model. By building a model integration procedure decision fusion is realized.

**Key words:** Rough set; Redundant knowledge; Decision fusion

### 1 引言

从数据集中对对象进行归纳学习和分类是智能计算<sup>[1]</sup>中的重要领域,旨在发现数据中隐藏的、未知的、潜在有用的知识. 对给定的决策信息系统,规则获取在于发现属性间的依赖特性,不同级别的信息泛化可以有不同的依赖,高级别知识粒度中的强函数依赖蕴涵着低级别的函数依赖. 属性依赖的问题已有相关研究<sup>[2,3]</sup>,并成功应用于机器学习、知识获取和决策支持等领域<sup>[4,6]</sup>.

### 2 规则集决策分析

给定决策信息系统  $S = (U, Q = C \cup D, V, F)$ . 其中:  $U$  为论域;  $C$  为条件属性集合;  $D$  为决策属性集合;  $V$  是所有属性值的并集且  $V = \bigcup_{q \in Q} V_q$ ,  $V_q$  是属性  $q$  的值域;  $F$  是一个信息函数  $F: U \times Q \rightarrow V$ , 给每个对象的每个属性赋值,对所有  $u \in U$ , 每个  $q \in Q$ , 均有  $F(u, q) \in V_q$ . 根据条件属性  $C$  产生的条件分类为  $U/C = \{X_1, X_2, \dots, X_{|U/C|}\}$ , 根据决策属性  $D$

产生的决策分类为  $U/D = \{Y_1, Y_2, \dots, Y_{|U/D|}\}$ , 则根据决策信息系统的条件分类和决策分类表示的规则为  $R_{ij}: X_i \rightarrow Y_j \text{ iff } X_i \cap Y_j \neq \emptyset$ .

在复杂的决策分析中,有必要从整体上讨论一个决策信息系统规则集合的整体性能,以此来衡量从样本集合得到的规则集知识库的决策性能,并可在规则集间进行比较.

#### 2.1 平均粒度

从决策信息系统的整体来看,平均粒度应该是各个块出现的概率与各块粒度的乘积之和,而各块出现的概率就是其块大小占整个数据集的比例.

**定义 1** 给定决策信息系统  $S = (U, Q = C \cup D, V, F)$ ,  $X_i \in U/\text{IND}(C)$ ,  $Y_j \in U/\text{IND}(D)$ , 则有

$$P(X_i) = |X_i| / |U|, 0 \leq P(X_i) \leq 1, \quad i = 1, 2, \dots, n; \quad (1)$$

$$P(Y_j) = |Y_j| / |U|, 0 \leq P(Y_j) \leq 1, \quad j = 1, 2, \dots, m. \quad (2)$$

收稿日期: 2006-03-20; 修回日期: 2006-05-24.

基金项目: 国家十五攻关计划项目(2002BA218C); 湖南省自然科学基金项目(06JJ20075).

作者简介: 王加阳(1966—),男,长沙人,教授,博士,从事计算智能、决策支持等研究; 王国仁(1966—),男,湖北崇阳人,教授,博士,从事数据挖掘、先进数据管理技术等研究.

条件类的平均粒度为

$$(C) = \sum_{i=1}^n (p(x_i) \cdot |x_i|) = \sum_{i=1}^n \left( \frac{|X_i|}{|U|} \cdot \frac{|X_i|}{|U|} \right) = \sum_{i=1}^n \frac{|X_i|^2}{|U|^2}. \quad (3)$$

决策类的平均粒度为

$$(D) = \sum_{j=1}^m (p(y_j) \cdot |y_j|) = \sum_{j=1}^m \frac{|Y_j|^2}{|U|^2}. \quad (4)$$

当所有的条件(或决策)类分得过细,单元素为一个类时,平均粒度最小,为 $1/|U|$ ;当整个论域构成一个类时,平均粒度为最大1.这与文献[7]的观点是一致的,但需要着重考虑决策系统的规则描述.

从规则的一致性角度来看,平均粒度对规则的不确定性程度有很大影响.条件分类 $U/\text{IND}(C)$ 粒度过大,对论域的划分过粗,容易产生很多不确定性规则,分类率也较低;粒度过小,对论域的划分过细,分类率可能很高,但使得每条规则支持的平均对象数较少,已知数据对规则的支撑度不够,使规则的随机性增大,缺乏对数据的代表性,对新的对象的分类预测可靠性下降.

## 2.2 规则集支持度

**定义2** 给定决策信息系统 $S = (U, Q = C, D, V, F)$ ,  $X_i \in U/\text{IND}(C)$ ,  $Y_j \in U/\text{IND}(D)$ ,则有

$$\text{Dsupp}(C, D) = \sum_{i=1}^n \sum_{j=1}^m \left( \frac{|X_i \cap Y_j|}{|U|} \cdot \frac{|X_i \cap Y_j|}{|U|} \right) = \sum_{i=1}^n \sum_{j=1}^m \frac{|X_i \cap Y_j|^2}{|U|^2}, \quad (5)$$

其中 $\text{Dsupp}(C)$ 为规则集合的平均支持度.

平均支持度描述了规则集的整体强度.从决策信息系统的整体来看,平均支持度应该是各个规则出现的概率与各规则支持度的乘积之和,而各规则出现的概率就是其对象数占整个论域的比例,也恰好与支持度从形式和值大小上是一样的.

从约简的角度看,不失一般性,假设一个属性的约简使条件等价类中的 $X_{n-1}$ 与 $X_n$ 合并,得到一个新条件类 $X_{n-1} + X_n$ .平均支持度的变化体现在

$$\text{Dsupp}(C, D) = \sum_{j=1}^m \left( \frac{|(X_{n-1} + X_n) \cap Y_j|^2}{|U|^2} - \frac{|X_{n-1} \cap Y_j|^2 + |X_n \cap Y_j|^2}{|U|^2} \right). \quad (6)$$

由于 $\text{Dsupp}(C, D) > 0$ ,即约简使得规则集合平均支持度增大.

## 2.3 规则集置信度

置信度定义了一条规则发生时,所得结论的可信程度.实际上,当条件类 $X_i$ 发生时,某一决策 $Y_j$ 的置信度就是条件概率的概念,即

$$P(Y_j | X_i) = \text{cer}(X_i, Y_j) \quad (7)$$

表示了 $X_i$ 发生的情况下,决策 $Y_j$ 发生的概率.

**定义3** 给定决策信息系统 $S = (U, Q = C, D, V, F)$ ,  $X_i \in U/\text{IND}(C)$ ,  $Y_j \in U/\text{IND}(D)$ ,则有

$$\text{Dcer}(C, D) = \sum_{i=1}^n \sum_{j=1}^m p(X_i, Y_j) \text{cer}(X_i, Y_j). \quad (8)$$

称 $\text{Dcer}(C, D)$ 为规则集合的置信度,由所有规则的规则强度与置信度共同决定,体现了规则集合决策充分性判断的整体程度,也可理解为规则集的整体置信度.

不失一般性,假设一个属性的约简,使条件等价类中的 $X_{n-1}$ 与 $X_n$ 合并,得到新条件类 $X_{n-1} + X_n$ ,记为 $X = X_{n-1} + X_n$ ,此时的规则集置信度为

$$\text{Dcer}(C, D) = \sum_{i=1}^{n-1} \sum_{j=1}^m p(X_i, Y_j) \text{cer}(X_i, Y_j). \quad (9)$$

约简前后的规则集置信度相减,有

$$\text{Dcer}(C, D) = \text{Dcer}(C, D) - \text{Dcer}(C, D) > 0. \quad (10)$$

由此证明,一个决策系统规则集合的整体置信度随着约简的进行呈现非单调递减趋势,实际也体现了不确定性的增大,二者具有一致性.

## 2.4 规则集覆盖度

**定义4** 给定决策信息系统 $S = (U, Q = C, D, V, F)$ ,  $X_i \in U/\text{IND}(C)$ ,  $Y_j \in U/\text{IND}(D)$ ,则有

$$\text{Dcov}(D, C) = \sum_{j=1}^m \sum_{i=1}^n p(X_i, Y_j) \text{cov}(X_i, Y_j). \quad (11)$$

称 $\text{Dcov}(C, D)$ 为规则集合的覆盖度,由所有规则的规则强度与覆盖度共同决定,体现了规则集合决策必要性判断的整体程度,也可理解为规则集的整体覆盖度.

不失一般性,假设一个属性的约简,使条件等价类中的 $X_{n-1}$ 与 $X_n$ 合并,得到新条件类 $X_{n-1} + X_n$ ,记 $X = X_{n-1} + X_n$ ,此时的规则集覆盖度为

$$\text{Dcov}(D, C) = \sum_{j=1}^m \sum_{i=1}^{n-1} p(X_i, Y_j) \text{cov}(X_i, Y_j). \quad (12)$$

约简前后的规则集覆盖度相减,有

$$\text{Dcov}(D, C) = \text{Dcov}(D, C) - \text{Dcov}(D, C) > 0. \quad (13)$$

可知规则集覆盖度的变化趋势与规则集置信度正好相反,随着约简进行呈现非单调递增趋势.这也可直观理解为,约简使条件粒度增大,从而当一项决策发生时,判断其发生条件的可信度增加.

### 3 多知识库决策融合

粗糙集方法能够从信息表中导出决策规则集合,并用于分类新的对象.根据粗集属性约简的概念和方法,一个决策系统的属性集合可能存在多个约简,每个约简对应的属性和规则互不相同,各个约简都可形成一个知识库,从而形成多个知识库.在一般情况下,很难把握利用哪一个约简得出的规则集合进行推理最佳.根据信息融合原理,可进行多知识库的信息融合,提高决策的准确性与综合性.

#### 3.1 模型定义

把知识库作为一个模型看待,从而将决策问题以模型与模型集成的形式表现,实现更为复杂的决策综合.模型集成已有相关的研究<sup>[8,9]</sup>,最初的研究工作主要集中在模型表示上,提出了一些经典的模型表示法,如结构化建模、逻辑建模、基于图的表示法等.模型表示的研究是模型管理研究工作的起点<sup>[10]</sup>,但是仅对模型表示研究还不足以解决模型管理中的许多问题,如多个模型组合运算问题,因此随着模型研究的深入,逐步形成发展了对模型集成的研究,以实现决策的综合.

模型集成目前尚无统一的概念,直观的理解是通过多个模型的连接组合构成一个更大的模型.模型集成主要可以分为模型模式上的集成和模型操纵上的集成<sup>[11]</sup>两类.模式上的集成是指在模型的定义部分的综合,即在模型定义部分完成集成,它与模型的表示紧密相关,操纵上的集成是指模型在运行时,通过模型间参数的传递来实现集成.

一个完备的模型管理至少必须做到模型的选择、连接和执行,这就需要模型的形式化框架,并通过模型输入输出的组合关系,构造复杂的模型.这也使得异构模型的组合成为可能,从而支持更高级别的模型集成.

定义 5 模型可表示为  $m = (IN, OUT)$ , 其中:

1) IN 是输入参数集合,  $m(IN) = \{in_1, \dots, in_p\}$  或  $m(IN) = \{m(in_1), \dots, m(in_p)\}$ ,  $p$  为输入参数的个数, 记为  $p = |m(IN)|$ .

2) OUT 是输出参数集合, 不失一般性, 通常可假定 OUT 只含有一个元素, 记为  $m(OUT) = \{out\}$  或者  $m(OUT) = \{m(out)\}$ . 例如,  $m = (IN, OUT)$ ,  $m(OUT) = \{out\}$ ,  $m(IN) = \{in_1, in_2, in_3\}$  表示模型  $m$  有 1 个输出参数, 3 个输入参数.

一般地, 外部输入数据、从数据库中取得数据

等也是模型, 把这些模型的输入参数集合设定为空集合. 在一个系统中, 所有单个模型构成一个模型集合, 由  $n$  个模型组成的模型集合记为  $M = \{m_i | 1 \leq i \leq n\}$ .

定义 6 给定模型集合  $M = \{m_1, \dots, m_n\}$ , 对  $M$  中任意二个模型  $m_i \in M$  和  $m_j \in M, i \neq j$ ,  $m_i(out), m_j(in)$  表示模型  $m_i$  的输出参数为模型  $m_j$  提供了唯一输入参数,  $m_i(out), m_j(in)$  称为模型有序对, 所有模型有序对可构成一个有限集合  $R$ , 则  $MR = (M, R)$  称为模型  $M$  的模型组合关系.

由定义 6, 模型集合  $M$  中任意模型的输入参数个数记为  $|m(IN)|$ , 则该模型的  $|m(IN)|$  个输入参数必然来自  $|m(IN)|$  个不同模型的输出参数. 可知, 一个模型只能为一个模型提供一个输入, 而且还可为多个模型提供一个输入.

由于任意模型输出参数的惟一性, 有序模型对  $m_i(out), m_j(in)$  可以简记为  $m_i, m_j(in)$ . 如果不必关注模型  $m_j$  的输入参数中的哪一个参数对应  $m_j(in)$ , 则模型有序对可以进一步简记为  $m_i, m_j$ , 它简明地表示了二个模型之间具有输入输出关系.

模型组合关系包含了模型集合  $M$  中所有模型间的可能输入输出关系, 而一个有序对  $m_i(out), m_j(in)$  是否属于  $R$  取决于所讨论问题的特性.

定义 7 给定模型集合  $M = \{m_1, \dots, m_n\}$ ,  $MR = (M, R)$  为模型组合关系, 有

$$\forall m_j \forall in_j (m_j \in M \wedge in_j \in m_j(IN) \wedge \exists m_i (m_i \in M \wedge out = m_i(OUT) \wedge m_i, m_j(in) \in R)), \quad (14)$$

则称  $MR$  是完备的.

$MR$  的完备性表明对  $M$  中任意模型  $m_j$  的任意输入  $m_j(in)$  必然存在至少一个模型  $m_i \in M$ , 其输出  $m_i(out)$  使得  $m_i(out), m_j(in)$  存在. 因为  $M$  是有限集合, 任意模型的输入输出也是有限的, 所以完备性是可判定的.

决策融合思想在于模型定义与集成, 通过集成实现决策融合. 给定一模型集与所期望的输出, 在模型的运行中组合多个模型运算得到结果的过程称为基于模型操纵的模型集成<sup>[12]</sup>. 这意味着只能用给定的模型集中的模型进行模型集成, 并且这些模型组合运算的最终结果应是所期望的输出结果.

#### 3.2 模型集成

定义 8 给定模型组合关系  $MM = (M, R)$ , OUTPUT 为期望输出, 模型集成就是要构造集合  $MM = (MI, RI)$ , 且满足下列条件:

1)  $MI \subseteq M, RI \subseteq R, MI$  是模型集成后的最终模型集, 集合  $RI$  描述了  $MI$  中模型的输入输出关

系.

2)  $\exists !mt [ mt \in MI \mid OUTPUT = mt(OUT) ] \exists m(m \in MI \mid mt, m \in RI) ]$ ,  $MI$  中模型  $mt$  的输出是集成后的唯一期望输出,称为终端模型.

3)  $\forall m_i \exists m_j [ (m_i \in MI \mid m_i \in mt) \wedge (m_j \in MI \mid m_i, m_j \in RI) ]$ ,  $MI$  中非终端模型的输出必是  $MI$  中其他模型的输入.

4)  $\forall m_j \exists !m_i [ m_j \in MI \mid (m_i \in MI \mid m_i, m_j \in RI) ]$ .  $MI$  中任意模型的一个输入只能由  $MI$  中的唯一模型提供.

5) 不存在有序集合  $\{m_1, \dots, m_L\} \subseteq MI$ , 满足  $m_1, m_2 \in RI, \dots, m_k, m_{k+1} \in RI, \dots, m_{L-1}, m_L \in RI, m_L, m_1 \in RI$ . 这意味着不存在参数循环传递的有序模型集合  $\{m_1, \dots, m_L\} \subseteq MI$ .

满足条件 1) ~ 4) 的  $MM = (MI, RI)$  称为复合模型,而满足条件 1) ~ 5) 的  $MM = (MI, RI)$  称为无环复合模型.

一旦模型组合关系  $MR$  和期望输出集合  $OUTPUT$  给定,在模型的运行中选择组合多个模型构成一个复合模型称为基于模型操纵的模型集成.这意味着在  $M$  中选择模型进行集成,集成结果或复合模型的输出是期望输出.

给定模型集合  $M$ ,  $MR$  可根据  $M$  的特性计算出来,再给定期望输出集合  $OUTPUT$ ,则可构造模型集成中所调用的模型集合  $MI$ .由定义的思想,对  $MI$  中的任意二个模型,根据输入输出参数的传递关系,二个相关模型间的输入输出关系可形成一个有序对,所有有序对构成了关联模型集合  $RI$ ,这时就得出一个复合模型.

根据  $MI$  和  $RI$ ,可知道在复合模型中所有模型之间的参数传递关系.在给定  $MR$  和期望输出集合的情况下,模型集成就是要找到  $MI$  和  $RI$ ,这样复合模型存在.复合模型是否存在,如何构造和实现复合模型是模型集成的基本问题.

### 3.3 决策融合

复合模型即模型集成后构成的综合模型,体现了对问题的解,表达了决策融合的思想.给定模型组合关系与预定输出集合,模型集成的主要问题在于是否可以完成期望的集成,实现决策融合.

**定理 1** 给定模型组合关系  $MR = (M, R)$  和期望输出集合  $OUTPUT$ ,终端模型为  $mt \in M$ ,  $OUTPUT = mt(OUT)$ ,则  $MR$  的完备性是模型集成或构造一个复合模型的充要条件.

**证明** 若模型集成是可完成的,由定义 7,  $MR$  必须是完备的.换言之,若  $MR$  不是完备的,则当构造复合模型时,至少有一个模型的输入参数之一不

能被提供,从而不能生成一个完整的复合模型.相反地,若  $MR$  是完备的,通过算法构造  $MM = (MI, RI)$ ,可知  $MM$  满足定义 8 的条件 1) ~ 4),从而得到复合模型.

在模型组合关系  $MR$  中,首先必须找到一个模型的输出是期望输出,若该模型的输入非空,则在模型组合关系  $MR$  中继续寻找其他模型,其输出提供了所需的输入参数.不断进行此过程,直到所有的输入参数已被提供.最终,所构造的符合模型的输出即是期望输出.下面提出复合模型构造算法.

首先在模型集中找出一模型,其输出参数为期望输出,若该模型有输入参数,则调用其他模型,使其输出参数能提供给前一模型的输入参数;其次,对于每个被调用的模型,若其有输入参数,则继续调用别的模型来提供该输入参数,这样一直进行下去,直到所有模型的输入参数均可被提供为止,通过模型的层层调用计算出期望输出.

模型选择的标准根据规则集决策分析提供的度量来权衡.模型的选择是组合动态过程,复合模型最终表示了对一个问题的综合决策,实现对复杂问题的建模.

**算法 1** 复合模型构造.输入为模型集合  $M$ ,模型组合关系  $MR$ ,期望输出  $OUTPUT$ .输出为复合模型.

```
(1) 令  $MI = \emptyset, RI = \emptyset$ .
// 假定  $MI$  中的模型按加入先后排序,用指针 CurrentMI 指向  $MI$  中目前被处理的模型.
(2) 找出  $mt \in M, mt(OUT) = OUTPUT$ ,且使  $MI = \{mt\}, RI = \{mt, NIL\}$ , CurrentMI =  $mt$ .
(3)
REPEAT
  IF CurrentMI(IN)  $\neq \emptyset$ 
    Because of the completeness of  $MR$ 
    Finding out  $m_1, m_2, \dots, m_P$  in model set  $M, P = |CurrentMI(IN)|$ 
    Satisfying  $m_1, CurrentMI \in R, \dots, m_P, CurrentMI \in R$ 
    LET  $RI = RI \cup \{m_1, CurrentMI, m_2, CurrentMI, \dots, m_P, CurrentMI\}$ 
    FOR  $i = 1$  TO  $P$ 
      IF  $m_i \notin MI$  THEN
         $MI = MI \cup \{m_i\}$ 
      ENDIF
    ENDFOR
  ENDIF
```

Current  $MI$  moves to next model in ordered-set  $MI$

UNTIL Current  $MI = NIL$ .

(4) 最终构造出复合模型  $MM = (MI, RI)$ .

$MR$  的完备性保证了(3)的合理性. 由于  $|MI| / |M|$ , 且  $M$  是有限集合, 则(3)将在有限步终止. 显然, 对于  $MM = (MI, RI)$ , 定义 8 中的条件 1) 是正确的, 即  $MI \subseteq M, RI \subseteq R$ . 步骤(2)保证了定义 8 中的条件 2). 根据构造思想方法, 步骤(3)使得定义 8 中的条件 3) 和 4) 成立.

由  $mt \ M, OUTPUT = mu(OUT)$  和  $MR$  的完备性, 复合模型  $MM$  可被构造. 由构造方法,  $MM$  满足定义 8 中条件 1) ~ 4). 若条件 5) 也能满足, 则存在无环组合模型, 而实际上只有无环组合模型才可在模型计算中实现.

**定理 2** 给定模型组合关系  $MR = (M, R)$  和期望输出集合  $OUTPUT$ , 且有  $mt \ M$   $OUTPUT = mt(OUT)$ . 假定  $MR$  是完备的, 且不存在有序集合  $\{m_1, \dots, m_L\} \subseteq M$  满足  $m_k, m_{k+1} \ R$ , 当  $k = L$ , 使  $k + 1$  为 1, 则满足定义 8 条件 1) ~ 5) 的无环组合模型  $MM = (MI, RI)$  存在.

**证明** 因  $MR$  是完备的, 由定理 1, 可构造  $MM = (MI, RI)$  满足定义 8 的条件 1) ~ 4). 由  $MR$  的特性, 且  $MM$  由  $MR$  构造,  $MI \subseteq M, RI \subseteq R$ , 根据已知, 不可能存在有序集合  $\{m_1, \dots, m_L\} \subseteq MI$  满足

$$m_1, m_2 \ RI, \dots, m_k, m_{k+1} \ RI, \dots, m_{L-1}, m_L \ RI, m_L, m_1 \ RI.$$

则  $MM = (MI, RI)$  满足定义 8 的条件 5), 因而  $MM = (MI, RI)$  是一个无环复合模型.

定理 2 意味着, 若模型组合关系  $MR = (M, R)$  是无环的, 那么由它构造复合模型是存在的, 并且必定是无环复合模型.

**定理 3**  $MR = (M, R)$  是模型组合关系,  $OUTPUT$  是期望输出集合,  $\exists mt \ M$ , 且  $OUTPUT = mt(OUT)$ , 若  $R$  是完备的,  $\exists ML = \{m_1, \dots, m_L\} \subseteq M$ , 且有: 1)  $\forall m_i \exists m_j [m_i \ ML \ m_j \ ML \ (in \ m_i(IN) \ out \ m_j(OUT) \ m_j, m_i \ R)]$ ; 2)  $\forall m_i \exists m_j [m_i \ ML \ m_j \ ML \ (out \ m_i(OUT) \ in \ m_j(IN) \ m_i, m_j \ R)]$ . 则满足定义 8 的无环复合模型存在.

**证明** 假定集合  $ML = \{m_1, \dots, m_L\} \subseteq M$  存在, 满足  $m_1, m_2 \ R, \dots, m_L, m_1 \ R$ , 则将有:

1)  $\forall m_k \ ML, \exists m_{k-1} \ ML, in \ m_k(IN)$

$out \ m_{k-1}(OUT)$ , 有  $m_{k-1}, m_k \ R$ , when  $k = 1$ , let  $k - 1$  be  $L$ ,

2)  $\forall m_k \ ML, \exists m_{k+1} \ ML, out \ m_k(OUT)$

in  $m_{k+1}(IN)$ , 有  $m_k, m_{k+1} \ R$ , when  $k = L$ , let  $k + 1$  be 1.

这与定理所给定的条件矛盾, 因而假设不成立, 由定理 2 可证定理 3 成立.

**定理 4**  $MR = (M, R)$  为模型组合关系,  $OUTPUT$  为期望输出集合,  $mt \ M$  ( $OUTPUT = mt(OUT)$ ). 假定  $MR$  是完备的,  $MM = (MI, RI)$  按定理 1 描述的方式构成. 若  $(\exists !ML = \{m_1, \dots, m_L\} \subseteq M) \ (MI \ ML) = \emptyset$ , 且有  $m_1, m_2 \ R, \dots, m_{L-1}, m_L \ R, m_L, m_1 \ R$ , 则  $MM = (MI, RI)$  是无环复合模型.

**证明** 因  $MM = (MI, RI)$  可由  $MR$  构造, 由构造方法知,  $MM$  满足定义 8 的条件 1) ~ 4).

如果不能满足定义 8 的条件 5), 则有集合  $M_0 = \{m_1, \dots, m_L\} \subseteq MI$ , 且

$$m_1, m_2 \ RI, \dots, m_L, m_1 \ RI.$$

由已知,  $ML$  是具有给定特性集合  $M$  的唯一子集合, 则有  $M_0 \subseteq MI \subseteq M, RI \subseteq R$ , 可推得  $M_0 = ML$ , 因而与给定条件  $MI \ ML = \emptyset$  不一致, 由此可知  $MM$  满足定义 8 的条件 5).

实际上, 可能存在许多与  $ML$  具有相同特性的子集合, 但只要这些集合与  $MI$  不相交, 按定理 1 思想构造的复合模型  $MM = (MI, RI)$  就是无环复合模型.

定理 4 意味着如果在模型组合关系  $MR$  中存在循环传递, 去除所有的环, 复合模型  $MM = (MI, RI)$  仍然存在, 则  $MM$  必是无环复合模型, 这描述了一个比较理想的状态.

根据粗糙集理论, 一个规则集可由决策信息系统的约简导出, 并用来分类新对象. 在很多情况下, 使用最小规则集分类新对象是不够充分的, 因为一个最小规则集仅代表一个知识库的分类能力, 只是由一个决策信息系统的约简产生的, 对噪声数据较敏感.

在形式体系上, 多知识库与粗糙集合理论中的属性约简一致. 从一个约简表可以导出一个知识库, 使用不同的约简, 可以得到不同的知识库. 一个决策信息系统可能存在多个约简, 从而形成多个知识库. 因此, 粗糙集理论能够为生成多知识库提供一种很好的理论模型和形式化方法, 根据知识模型进行信息推理和决策层数据融合.

### 4 结 论

粗糙集理论研究的重要内容是分类与约简, 目的在于获取优良的规则集合, 实现准确的决策. 通过提出规则集合的决策度量, 从整体上体现了一个规则集合的性能. 规则集合的度量体现了整个规则集



合的性质,而约简分析则正是体现了规则集合度量的变化,从多角度综合描述了规则集合的性质,特别是对于多知识库的决策融合,提供了具有整体性的决策模型选择.基于模型集成的基本理论,将规则知识库作为一个知识库模型,通过模型的集成实现决策的融合,在模型一级实现了决策融合.

### 参考文献(References)

- [1] Wong S K M, Ziarko W, Li Y R. Comparison of rough set and statistical methods in inductive learning[J]. Int J of Man-machine Studies, 1986, 25(1): 53-72.
- [2] Agrawal R, Imielinski T, Swami A. Database mining: A performance perspective [J]. IEEE Trans on Knowledge and Data Engineering, 1993, 5(6): 169-173.
- [3] Hu X, Cercone N, Han J. Fuzzy sets and knowledge discovery[M]. Berlin: Springer, 1994: 90-99.
- [4] 李剑, 范小军, 黄沛. 基于粗糙集的知识理论及其应用[J]. 系统工程理论方法应用, 2001, 10(3): 184-188. (Li J, Fan X J, Huang P. Theory of knowledge based on rough sets and its application [J]. Systems Engineering — Theory Methodology Applications, 2001, 10(3): 184-188.)
- [5] 印勇, 曹长修, 张邦礼. 基于粗糙集理论的分类规则发现[J]. 重庆大学学报, 2000, 23(1): 63-65. (Yin Y, Cao C X, Zhang B L. Classification rule discovery based on rough set theory[J]. J of Chongqing University, 2000, 23(1): 63-65.)
- [6] Pawlak Z. Rough classification [J]. Int J of Man-machine Studies, 1984, 20(5): 469-483.
- [7] 苗夺谦, 范世栋. 知识的粒度计算及其应用[J]. 系统工程理论与实践, 2002, 22(1): 48-56. (Miao D Q, Fan S D. The calculation of knowledge granulation and its application[J]. Systems Engineering Theory and Practice, 2002, 22(1): 48-56.)
- [8] Watson H J. Decision support in the data warehouse [M]. New Jersey: Prentice-Hall, 1998.
- [9] Kimball R, Strehlo K. Why decision support fails and how to fix it[J]. Acm Sigmod Record, 1995, 24(3): 65-77.
- [10] Saifur Rahman, Rahul Bhatnagar. An expert system based algorithm for short term load forecast[J]. IEEE Trans on PWRs, 1998, 3(2): 886-892.
- [11] Huang T Y. Intelligent decision support systems[M]. Beijing: Publishing House of Electronics Industry, 2001: 73-102.
- [12] Kottemann J E, Dolk D R. Process-oriented model integration[C]. Proc of the 21st Hawaii Int Conf on System Sciences. Hawaii: IEEE Computer Society Press, 1998: 656-668.

(上接第 656 页)

### 参考文献(References)

- [1] Pawlak Z. Rough sets [J]. Int J of Information and Computer Science, 1982, 11(5): 341-356.
- [2] Pawlak Z. Rough set approach to multi-attribute decision analysis [J]. European J of Operational Research, 1994, 72(3): 443-459.
- [3] 苗夺谦, 王珏. 粗糙集理论中概念与运算的信息表示[J]. 软件学报, 1999, 10(2): 113-116. (Miao D Q, Wang J. A information representation of the concepts and operations in rough set theory[J]. J of Software, 1999, 10(2): 113-116.)
- [4] Wang J, Wang J. Reduction algorithm based on discernibility matrix the ordered attributes method[J]. J of Computer Science and Technology, 2001, 16(6): 489-504.
- [5] Hu X H, Cercone N. Learning in relational databases: A rough set approach[J]. Computational Int: An Int J, 1995, 11(2): 323-338.
- [6] 叶东毅, 陈昭炯. 一个新的差别矩阵及其求核方法[J]. 电子学报, 2002, 30(7): 1086-1088. (Ye D Y, Chen Z J. A new discernibility matrix and the computation of a core[J]. Chinese J of Electrics, 2002, 30(7): 1086-1088.)
- [7] Zheng Z, Wang G Y, Wu Y. Objects' combination based simple computation of attribute core[C]. Proc of the 2002 IEEE Int Symposium on Intelligent Control. Vancouver, 2002: 514-519.
- [8] 杨明, 孙志挥. 改进的差别矩阵及其求核方法[J]. 复旦大学学报, 2004, 43(5): 865-868. (Yang M, Sun Z H. Improvement of discernibility matrix and the computation of a core[J]. J of Fudan University, 2004, 43(5): 865-868.)
- [9] 杨明. 一种基于改进差别矩阵的核增量式更新算法[J]. 计算机学报, 2006, 29(3): 407-413. (Yang M. An incremental updating algorithm of the computation of a core based on the improved discernibility matrix[J]. Chinese J of Computers, 2006, 29(3): 407-413.)
- [10] Wang G Y, Zhao J, An J J, et al. Theoretical study on attribute reduction of rough set theory: Comparison of algebra and information views [C]. Proc of the 3rd IEEE Int Conf on Cognitive Informatics. Washington DC: IEEE Computer Society, 2004: 148-155.