

粗糙集理论在故障诊断中的问题分析

赵熙临, 刘 辉

(湖北工业大学 电气与电子工程学院, 湖北 武汉 430068)

摘 要:粗糙集理论能对系统中的冗余信息进行约简,但其处理过程完全基于样本集,样本集的完备性对其处理结果有直接影响。对粗糙集理论及其在故障诊断中属性约简存在的问题进行了分析,通过实例证明了在故障样本集不完备的情况下,利用粗糙集进行的属性约简会由于新故障样本的引入而导致前后约简结果的不一致,从而影响诊断的准确性;指出了该问题产生的关键原因及解决的办法,并给出了相关的实现算法,以提高系统的故障诊断自适应性。

关键词:粗糙集;故障诊断;区分函数;约简

中图分类号:TP301.6;TP277

文献标识码:A

文章编号:1673-629X(2008)01-0132-04

An Analysis of Problem about Rough Set Theory in Fault Diagnosis

ZHAO Xi-lin, LIU Hui

(School of Electrical and Electronic Engineering, Hubei University of Technology, Wuhan 430068, China)

Abstract: Rough set theory can reduce the redundant information in the system. But the process of the reduction only based on the sample set. The integrality of the sample set will influence the result directly. Rough set theory and the problem consist in the reduction of the attribute in fault diagnosis are analyzed in this paper. It proved that when the sample set of the fault is incomplete, the result of the attribute reduction will be variant when a new sample is joined into the sample set, and this will influence the correctness of the diagnosis. Pointed out the pivotal reason and the method that can resolve the problem, and the resolved algorithm is also put forward. It will enhance the ability of the self-adaptation of the system.

Key words: rough set; fault diagnosis; discernibility matrix; reduction

0 引言

粗糙集理论是波兰学者 Z. Pawlak 于 1982 年提出的,作为处理不完整不精确知识的新型数学方法。该理论不需要任何初始或附加信息,直接对数据进行分析处理,发现其中隐含的规律,提取有用特征,得到简明扼要的知识表达形式。从而在模式识别、机器学习、故障诊断、决策分析等领域得到了广泛应用。

在故障诊断领域,由于故障发生的复杂性、不确定性及人们对故障及故障原因认识的局限性,对相关故障信息的分析与处理就存在一定的困难。因而粗糙集理论在该领域得到了越来越多的关注。目前,粗糙集理论在故障诊断中的应用主要有两种:一种是直接从故障征兆信息中约简掉冗余信息,利用其不可约特征信息生成决策树,再利用决策树进行故障的诊断^[1,2];

另一种是将粗糙集与其它理论相结合,利用粗糙集的信息处理能力对数据进行预处理,约简掉其中冗余的信息以简化信息规模,再利用其它方法对处理过的信息进行故障的判断,以提高系统的诊断效率^[3]。

但不论是何种应用方式,由于粗糙集对数据的处理完全基于样本集,样本的完整性直接影响其处理结果。而实际应用中,故障样本来源于实验和专家的经验,故障的判断是根据多个条件的状态取值确定的,全部的条件组合状态不可能都得到验证。因此,故障样本通常是不完整的,这就会使粗糙集在故障诊断中的应用面临一定的困难。文中通过实例对此问题进行了分析,得出了问题存在的原因,并提出了解决的算法。

1 粗糙集理论基本知识

粗糙集理论的主要思想是在保持分类能力不变的前提下,通过知识约简,导出问题的决策分类规则,它把知识看成是关于论域的划分。设 $U \neq \emptyset$ 是人们感兴趣的对象组成的有限集合,称为论域。 U 中的任何概念族称为关于 U 的知识。在 U 上对某一知识的一个

收稿日期:2007-04-04

基金项目:湖北省教育重点项目(D200614013)

作者简介:赵熙临(1969-),男,山东郓城人,讲师,博士,研究方向为故障诊断;刘 辉,副教授,博士,研究方向为网络控制系统、智能控制。

划分 $F = \{X_1, X_2, \dots, X_n\}$; 其中 $X_i \subseteq U, X_i \neq \emptyset, X_i \cap X_j = \emptyset (i \neq j), \bigcup_{i=1}^n X_i = U$ 。 U 上的一族划分称为关于 U 的一个知识库。一个知识库可用一个关系 $K = (U, R)$ 表示, 其中 U 为对象集合, R 是 U 上的一族等价关系。若 $P \subseteq R$, 且 $P \neq \emptyset$, 则 $\cap P$ 也是一个等价关系, 称为 P 上的不可区分关系, 记为 $\text{ind}(P)$ 。 $U/\text{ind}(P)$ 表示与等价关系族 P 相关的知识, 称为 K 中关于 U 的 P 基本知识。

粗糙集理论在实际应用中的一个重要特点就是可不需要预先给定某些特征或属性的数量描述, 而直接从给定问题的描述集合出发, 通过不可分辨关系和不可分辨类确定给定问题的近似域, 从而找出该问题的内在规律, 对数据集进行约简。

定义1: 在一个知识库中, R 为一族等价关系, $C \in R$, 如果 $\text{ind}(R) = \text{ind}(R - \{C\})$, 则称 C 为 R 中不必要的; 否则 C 为 R 中必要的^[4,5]。如果对于知识库中的知识划分, 某一知识是不必要的, 则可以将其约简掉。设 $Q \subseteq P$, 如果 Q 是独立的, 且 $\text{ind}(Q) = \text{ind}(P)$, 则称 Q 是 P 的一个约简。 P 中所有约简的交集称为 P 的核, 记为 $\text{core}(P)$ 。

把知识库中的一类关系称作条件属性, 另一类关系称作决策属性。两者的交集为空, 并集是整个关系集合。具有条件属性和决策属性的知识表达系统称为决策表。

定义2: 利用区分矩阵能容易地计算约简和核, 决策表 S 的区分矩阵是一个 $n \times n$ 矩阵, 其任一元素为:

$$a * (x, y) = \{a \in C \mid f(x, a) \neq f(y, a) \text{ 且 } \omega(x, y)\}$$

其中 $f(x, a)$ 是数据对象 x 在条件 a 下的值。对于 $x, y \in U, \omega(x, y)$ 满足 $x \in \text{POS}_C(D)$ 且 $y \notin \text{POS}_C(D)$, 或者 $x \notin \text{POS}_C(D)$ 且 $y \in \text{POS}_C(D)$, 或者 $x, y \in \text{POS}_C(D)$ 且 $(x, y) \in \text{ind}(D)$ ^[5]。

决策表 S 的区分函数 Δ^* 定义为: $\Delta^* = \prod_{(x, y) \in U \times U} a * (x, y)$ 。区分函数 Δ^* 的极小析取范式中的所有合取式是 C 的所有 D 约简。

如果 $C' \subseteq C$ 是满足条件 $C' \cap a * (x, y) \neq \emptyset, \forall a * (x, y) \neq \emptyset$ 的极小子集, 则 C' 是 C 的 D 约简。

2 粗糙集故障诊断问题的描述

在故障诊断过程中, 通常是通过故障信息的收集形成故障样本数据库, 然后通过适当的方法对该数据库进行分析处理再应用于故障的诊断。如专家系统是通过故障样本的分析提取出诊断规则, 然后利用该诊断规则对出现的故障进行判断; 神经网络诊断系统

是利用故障样本集对设计好的神经网络进行训练, 调整网络中的相关参数, 然后通过训练好的神经网络对诊断对象进行运行模式识别以判断是否有故障发生^[1,3,6]。因此, 作为系统进行分析处理的故障样本数据, 其准确性直接影响故障诊断的效率及正确率。

随着控制规模的扩大, 要求检测的对象也急剧增加, 如果按传统的方法进行故障诊断, 必然会使诊断系统过于复杂而导致诸如在线诊断速度下降、诊断正确率降低等问题的发生。而很多实际情况中, 对于一个故障点的判断存在大量的信息冗余。因此, 如果有方法能够对检测的数据进行处理, 在不影响原来数据的表达效果的基础上约简掉其中的冗余信息, 无疑会提高整个系统的诊断效率。粗糙集理论正好适合对这方面数据的处理。利用粗糙集理论可对检测的数据进行约简, 消除其冗余信息。不失一般性, 以下表对旋转机械进行故障诊断的一个样本集^[2]进行分析, 检测数据都经过了离散化处理。其中 U 为对象集, $a \sim g$ 为条件属性, D 为决策属性, 表示在所列 15 种取值条件下会发生已知的 5 种故障。

U	a	b	c	d	e	f	g	D
1	1	1	1	5	1	1	1	1
2	1	1	1	4	1	2	1	1
3	1	1	1	3	1	2	1	1
4	1	1	1	3	1	1	2	1
5	1	1	1	2	3	1	1	2
6	1	1	1	3	2	1	1	2
7	1	1	1	3	2	2	1	2
8	1	1	1	4	1	1	1	3
9	1	1	1	2	1	2	2	3
10	2	2	1	2	1	1	1	4
11	1	3	1	2	1	1	1	4
12	1	2	1	3	1	1	1	5
13	1	2	1	2	1	1	2	5
14	2	2	1	2	1	1	2	5
15	1	2	2	1	1	1	1	5

根据前述粗糙集理论, 可得到该样本集的区分矩阵(因该矩阵为对称矩阵, 现只写出其下半部; 为书写方便, 矩阵中省去了析取符号, 如 $d \vee e$ 简写为 de):

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1															
2															
3															
4															
5	de	def	def	deg											
6	de	def	ef	eg											
7	def	de	e	efg											
8	d	f	df	dg	de	de	def								
9	dfig	dg	dg	df	efg	defg	deg								
10	abd	abdf	abdf	abdg	abe	abde	abdef	abd	abfg						
11	bd	bdf	bdf	bdg	be	bde	bdef	bd	bfg						
12	bd	bdf	bf	bg	bde	be	bef	bd	bdfg	ad	bd				
13	bdg	bdfg	bdfg	bd	beg	bdeg	bdefg	bdg	bf	eg	fg				
14	abdg	abdfg	abdfg	abd	abeg	abdeg	abdefg	abdg	abf	g	abg				
15	bcd	bcdi	bcdi	bdg	bode	bode	bocdef	bcd	bcdi	acd	bcd				

其区分函数为矩阵中所有元素的合取式, 即 Δ^*

$$= (d \vee e) \wedge (d \vee e) \wedge (d \vee e \vee f) \wedge d \wedge (d \vee f \vee g) \wedge (a \vee b \vee d) \cdots (b \vee g) \wedge (a \vee b \vee g) \wedge (b \vee c \vee d)$$

化简得 $\Delta^* = d \wedge e \wedge f \wedge g$ 。因此,该样本集条件属性的约简为 d, e, f, g ,而 a, b 和 c 是冗余的。消去原样本集中冗余的条件属性 a, b, c 后,得到简化的样本集如下:

U	d	e	f	g	D
1	5	1	1	1	1
2	4	1	2	1	1
3	3	1	2	1	1
4	3	1	1	2	1
5	2	3	1	1	2
6	3	2	1	1	2
7	3	2	2	1	2
8	4	1	1	1	3
9	2	1	2	2	3
10	2	1	1	1	4
11	2	1	1	1	4
12	3	1	1	1	5
13	2	1	1	2	5
14	2	1	1	2	5
15	1	1	1	1	5

说明根据提供的故障情况,在 7 个条件属性中,只需根据约简得到的 4 个属性就可以获得与原来 7 个条件属性相同的信息,进行相同程度的判断。在系统的故障诊断过程中,只需对这 4 个条件属性进行分析处理即可,这样可大大地提高故障诊断的效率^[7]。比如,当直接利用条件属性构造决策树进行故障诊断时,如果得到检测数据 $a = 1, b = 3, c = 1, d = 2, e = 1, f = 1, g = 1$,约简后只需根据约简得到的 4 个条件属性值 d, e, f, g 对照上述决策表即可判断故障类型为 4,与约简前根据全部 7 个条件值得到的判断结果相同。

3 粗糙集应用的局限性及其示例分析

利用粗糙集理论进行约简是一种定量、客观的简化,约简计算以严密的定理为基础。虽然利用粗糙集理论可以约简故障样本集,减少条件属性的数目,提高故障诊断的效率,但粗糙集在实际应用中还是面临一定的问题。首先,粗糙集模型由于完全基于样本,有限的样本中只能获得有限的知识,不完备的样本中的知识同样可能不完备,因此存在一个知识有限外推性的问题,即不能保证对每一个新样本都能给出判断。其次,有一些关键信号,即属于粗糙集信息系统核属性的信号源不能缺失,否则粗糙集将会失效。对于后一种情况,需要采用一些方法对缺失数据进行估计,但对于前一种情况则需更深入的探讨。

考虑如下情况,当检测到条件属性 $a = 1, b = 1, c = 1, d = 3, e = 1, f = 1, g = 1$ 时发生故障类型 3,该故障在原故障样本集中是不存在的,是一个新的故障

样本。而如果按照由原样本集约简得到的决策表进行判断 $d = 3, e = 1, f = 1, g = 1$ 对应故障类型 5,显然发生了错判。这说明约简的条件属性无法正确反应新故障的状态,现考虑是否由于新故障的发生导致了原样本集约简的变化。将该新故障加入到原样本集,重新计算区分函数分析其中的变化。

如上得到样本集的区分矩阵,其区分函数为矩阵中所有元素的合取式,即 $\Delta^* = (d \vee e) \wedge (d \vee e) \wedge (d \vee e \vee f) \wedge d \wedge d \wedge (d \vee f \vee g) \cdots (d \vee g) \wedge (a \vee d \vee g) \wedge (c \vee d) = b \wedge d \wedge e \wedge f \wedge g$,该样本集条件属性的约简为 b, d, e, f, g 。因此,对照前后两个区分函数可见约简发生了变化,故障的判断情况自然也发生变化。新的决策表如下表所示,根据第 9 行可判断故障类型为 3。

U	b	d	e	f	g	D
1	1	5	1	1	1	1
2	1	4	1	2	1	1
3	1	3	1	2	1	1
4	1	3	1	1	2	1
5	1	2	3	1	1	2
6	1	3	2	1	1	2
7	1	3	2	2	1	2
8	1	4	1	1	1	3
9	1	3	1	1	1	3
10	1	2	1	2	2	3
11	2	2	1	1	1	4
12	3	2	1	1	1	4
13	2	3	1	1	1	5
14	2	2	1	1	2	5
15	2	1	1	1	1	5

这样,当将其它方法与粗糙集结合进行故障诊断,在出现新故障时可能有两种情况发生:一种如前分析会改变粗糙集的约简,另一种情况对约简没有影响。当后者发生时,由于约简能正确反映系统的故障状态,因此对故障的判断正确率没有影响。但当前者发生时,由于条件属性的约简发生了改变,一些需要的条件属性被系统约简掉,根据变化前的条件设计的诊断系统显然无法适应变化后的情况。

因此,利用粗糙集进行故障诊断,就需解决好故障样本不全面的情况下条件属性的约简问题。在有新故障出现的情况下,如何使系统自动地对新的样本集进行处理,使其约简集能根据实际情况进行自适应调整是提高故障诊断效率及正确率的关键。

4 粗糙集应用局限性解决算法研究与实现

通过上述分析,考虑利用粗糙集对故障样本进行约简后,系统出现故障时有三种可能情况发生:

①检测到的状态值与约简后的条件属性值相符,发生的故障状态也与决策属性判断相符,故障会被正常诊断。

②检测到的状态值与约简后的条件属性值相符,但实际发生的故障状态与根据条件属性判断得到的故障不相符,出现错误判断。

③检测到的状态值与约简后的条件属性值不相符,由于没有相应的故障类型,无法进行正确诊断。

对于第二种情况说明原样本集中,因为样本的不全面造成条件属性的约简不能完全反映故障类型,还有其它条件属性应该添加。第三种情况说明约简样本集不完整,还有新的故障状态无法进行判断,这时虽然可利用其它方法进行推断,但因为此时已说明原样本集是不完备的,也就说明在此基础上进行的约简也可能是不完全的,根据此约简进行的推断是否正确就值得怀疑。因此,对后两种情况的出现,都需要系统对新的样本集进行相关处理,以适应新的情况。

针对后两种情况,可根据下列步骤实现相关的操作以重新进行样本的约简。

Step1:新故障发生后,将新故障及各状态信息加入到样本集。

Step2:针对各条件属性对新样本集中对象进行分类,分别得出 U/C_i 及 U/D , 其中, C_i 为各条件属性值, D 为决策属性;并计算出 U/C , U/C 为根据全部条件属性进行的分类。

Step3:根据 U/C 和 U/D 求出 $POS_C(D)$ 。

Step4:列出区分矩阵,求出区分函数 Δ^* 。

Step5:求出区分函数 Δ^* 的极小析取式,其所有合取式是 C 的所有 D 约简,根据新的约简可得到新的决策表。

其中第二步的核心就是分类及求 U/C_i 的交集并计算出 U/C , 可利用相关算法^[4]进行程序的设计。

求 $POS_C(D)$ 的算法如算法1所示。该算法用于计算 D 的 C 正域,即算出 U 中所有根据分类 U/C 的信息可以准确划分到关系 D 的等价类中去的对象集合。设 $U/C = \{V_{1n}, V_{2n}, \dots, V_{sn}\}$, $U/D = \{V_{1m}, V_{2m}, \dots, V_{tm}\}$ 。

算法1:P1:初始化 $x = 1, y = 1, r = 0$

P2: V_{xn} 中取元素 u_y , 查找 u_y 属于何 U/D 子集,不妨设为 V_{im}

P3: if $u_y \in V_{im}$ go to P4, otherwise go to P6

P4: if $y > |V_{sn}|$ $V_r = V_{im}$ $x = x + 1$ go to P5, otherwise go to P3

P5: if $x > s$ 算法结束, otherwise $r = r + 1$ go to P2

P6: $x = x + 1$ if $x > s$ 算法结束, otherwise go to P2

第四步求区分函数的算法如算法2所示,该算法

结束时,可得到该信息系统的区分函数。设该系统条件属性是 c_1, c_2, \dots, c_m , $f(x_i, a)$ 表示对象 x_i 在条件属性 a 时的取值。

算法2:P1:初始化 $i = 1, j = 1, n = |U|, g_{ij} = F, g = T$

P2: $i = i + 1$ if $i > n - 1$ 算法结束, otherwise go to P3

P3: $j = i + 1$ go to P4

P4: if $j > n$ go to P8 otherwise go to P5

P5: $j = j + 1$ $a = c_j$ go to P6

P6: if $f(x_i, a) \neq f(x_j, a)$ $g_{ij} = g_{ij} \vee a$ go to P7

P7: if $a = c_m$ go to P4, otherwise

$a = \text{next } c_i$ go to P6

P8: $g = g_{ij} \wedge g$, go to P2

5 结论

粗糙集理论是一种知识处理的数学方法,将其应用于故障诊断领域可以对冗余信息进行约简,提高诊断的效率。但粗糙集对数据的处理完全来源于样本,样本的完备性会影响到约简的结果,从而会影响到诊断的准确率。文中对此进行了研究,指出故障样本的变化会影响属性的约简,提出了在故障样本集发生变化时粗糙集对约简重新处理的自适应算法。从而在必要时重新计算约简集,以提高系统的故障诊断自适应性。

参考文献:

- [1] 冯志鹏,宋希庚,薛冬新. 基于广义粗糙集理论的旋转机械故障诊断[J]. 振动与冲击, 2004, 23(1): 47-51.
- [2] 孙海军,蒋东翔,钱立军,等. 基于粗糙集理论的旋转机械故障诊断方法[J]. 动力工程, 2004, 24(1): 73-77.
- [3] 李千目,戚湧,张宏,等. 基于粗糙集神经网络的网络故障诊断新方法[J]. 计算机研究与发展, 2004, 41(10): 1696-1702.
- [4] 张文修,吴伟志,梁吉业,等. 粗糙集理论与方法[M]. 北京:科学出版社, 2001.
- [5] 王国胤. Rough集理论与知识获取[M]. 西安:西安交通大学出版社, 2001.
- [6] Liu Mei, Quan Taifan, Luan Shaohua. An Attribute Recognition System Based on Rough Set Theory - Fuzzy Neural Network and Fuzzy Expert System[C]// Proceedings of the 5th World Congress on Intelligent Control and Automation. Hangzhou, China: [s. n.], 2004: 2355-2359.
- [7] Zhao Xilin, Zhou Jianzhong, Yang Junjie, et al. An Analysis of Rule Generation in Fault Diagnosis Based on Rough Set Theory[C]// Proceedings of 2006 International Conference on Artificial Intelligence. Beijing, China: [s. n.], 2006: 694-697.