

# 基于粗糙集的装备故障诊断系统的算法研究

张凤琴<sup>1</sup>, 张水平<sup>1</sup>, 白钢<sup>2</sup>, 严晓梅<sup>1</sup>, 张月玲<sup>1</sup>  
(1. 空军工程大学 电讯工程学院, 陕西 西安 710077; 2. 95880 部队, 北京 100843)

**摘要:**通过设计基于粗糙集的装备故障诊断系统框架,以大屏幕投影系统的设备故障记录作为样本集,研究了粗糙集在装备故障诊断系统中的应用。针对数据预处理中缺失数据的填补问题,采用粗糙集理论和统计学理论相结合的方式,最大限度的保留了原始信息、产生规则的数量和可靠性;针对知识约简中规则生成速度慢等问题,采用了基于粒计算的粗糙集知识发现算法,将决策表中的属性集转换为二进制的矩阵对象,把复杂的知识约简转化为矩阵运算,降低了知识约简的时间复杂度,提高了生成规则运算效率,实现了故障诊断知识的快速自动获取。

**关键词:**粗糙集;故障诊断;缺失值处理;知识约简

**DOI:**10.3969/j.issn.1009-3516.2010.06.010

**中图分类号:** TP182 **文献标识码:** A **文章编号:** 1009-3516(2010)06-0047-05

随着装备种类增多,复杂程度、新技术含量提高,如何在短时间内全面掌握故障诊断技术等问题日益凸现。因此,研究一种技术新、操作性强的自动化装备故障诊断系统尤为重要。有关故障诊断的处理方法各有优势,但都不同程度地存在一些问题。例如,利用粗糙集理论在缺失数据填补时会导致数据的准确度降低;知识约简环节中又存在效率较低、实现复杂等缺点<sup>[1]</sup>。本文以大屏幕投影的故障数据为样本集,在研究基于粒计算的粗糙集知识发现方法的基础上,设计了装备故障诊断技术框架,提出缺失数据的填补和知识约简环节中的改进算法,有效地提高了知识规则的生成效率,实现了快速诊断装备故障的目的。

## 1 装备故障诊断系统的框架

基于粗糙集的装备故障诊断系统主要包括装备故障诊断和知识规则生成 2 部分<sup>[2-3]</sup>,见图 1。

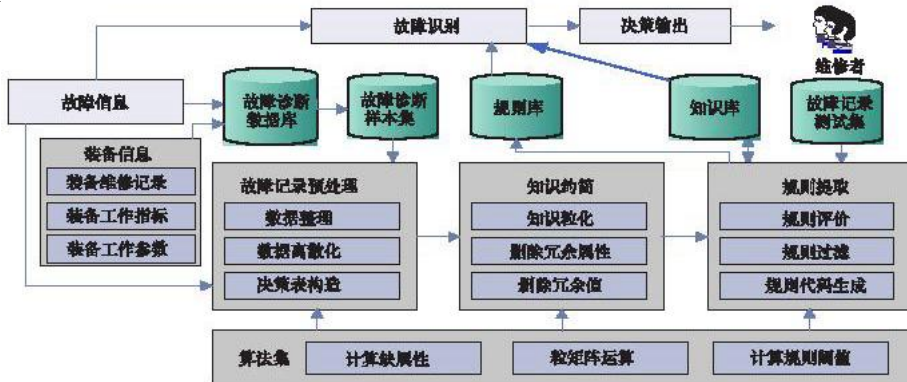


图 1 基于粗糙集的装备故障诊断系统的框架

Fig. 1 Framework of fault diagnosis system based on rough set

\* 收稿日期:2010-05-04

基金项目:陕西省自然科学基金资助项目(2010JM8035)

作者简介:张凤琴(1964-),女,山西芮城人,副教授,主要从事数据挖掘、信息安全研究;

E-mail:fengqin\_zhang@126.com

张水平(1956-),女,山西侯马人,教授,主要从事数据挖掘、信息分发研究。

装备故障诊断是根据专家的经验和本系统生成的知识规则,对故障进行自动推理和诊断,确定故障的原因和部位,给出装备诊断结果并提出维修或保养建议。知识规则生成则是根据设备信息、故障信息和专家的经验,将长期积累的设备故障记录作为样本训练集进行预处理,形成了二维故障决策表,利用知识约简算法和规则提取标准,形成规则库和知识库的内容。本文将对系统涉及到的数据缺失值处理算法和知识约简算法进行研究和分析。

## 2 故障数据缺失值处理算法

粗糙集的数据预处理是将故障记录信息转化为二维决策表,从而提高数据处理的质量<sup>[4-5]</sup>。故障记录预处理主要解决数据整理与检测、缺失值的处理以及数据的离散化等问题<sup>[6]</sup>。常用的缺失值的处理算法有:基于较高支持度的数据值填补算法、基于特殊的属性值的数据填补算法和基于统计学的数据填补算法等。采用统计学方法实现数据值填补,其算法思想简单,实现便利,相关方法主要有删除法、Mean Completer和 Combinatorial Completer 算法<sup>[7]</sup>。根据目前算法存在的缺陷,考虑将概率统计方法与粗糙集方法进行结合,本文提出了缺失数据补齐的改进算法,算法步骤为:

1) 在决策表中加入“概率”列,用于记录该条记录实际出现的次数,该列既不属于条件属性,也不属于决策属性,其值仅为填补缺失值提供统计信息。处理后该信息表  $S = \{C, P\}$ ,  $C = \{c_1, c_2, \dots, c_m\}$  为信息表的属性集合,  $m$  为属性的个数,  $P$  为概率列。

2) 若某记录中有一个属性值缺失,其它记录中该属性有  $L$  个取值,则  $V_1, V_2, \dots, V_{m-1}$  分别为与该记录有 1 到  $m-1$  个属性值相同的记录集合,  $V_i = \{v, n\}$ ,  $v = \{v_{i1}, v_{i2}, \dots, v_{in}\}$ ,  $1 \leq i \leq m-1$ ,  $v_{ip}$  为有  $i$  个属性值相同的情况下有  $L$  个缺失属性的值的集合,  $n$  为该值对应的记录数,  $\sum_{k=1}^n P_k$  为与该值对应原始记录的累加和。

3) 计算不同的  $v_{in}$  的概率  $p_{in}$ 。当缺失属性的值为  $v_{in}$  时,分别计算记录集合  $V_i$  中取该值的  $\sum_{k=1}^n P_k$ ,  $p_{in1} = \frac{i}{m-1} \sum_{k=1}^n P_k$  为该记录集合取该值的概率,最后把所有集合取该值的概率累加,即  $p_{in} = p_{in1} + p_{in2} + \dots + p_{inl} = \sum_{i=1}^{m-1} \left( \frac{i}{m-1} \sum_{k=1}^n P_k \right)$ , 就是  $v_{in}$  最终的概率  $p_{in}$ 。

4) 选取概率最大的  $v_{in}$  作为填补缺失属性的值。

5) 当某记录中缺失的属性值不止一个,例如有 2 个缺失值时,则步骤 2) 中与该记录有相同属性值的记录集合应为 1 到  $m-2$  个,以此类推。

本文以大屏幕投影故障数据的 967 条故障信息作为样本集,其中训练集数据占 2/3 共 645 条,测试集数据占 1/3 共 322 条。利用 Rosetta 软件对基于统计知识的删除法、Mean Completer、Combinatorial Completer 和本系统使用的缺失数据填补算法进行了测试和分析。经检测,该数据集中有空属性值的记录共 145 条。分别运用以上 4 种算法进行数据空缺值填补处理,得到的记录数比较结果见图 2。经过约简(使用遗传算法)后得到的规则数量见图 3。

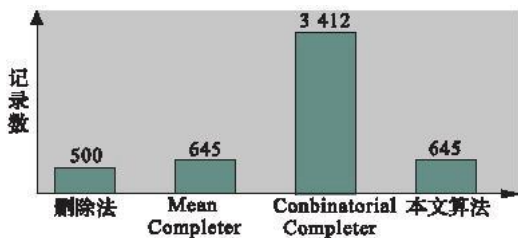


图2 空缺值填补算法得到的数据集记录数比较结果

Fig. 2 Result of comparing of data set records acquired from algorithm of blank filling

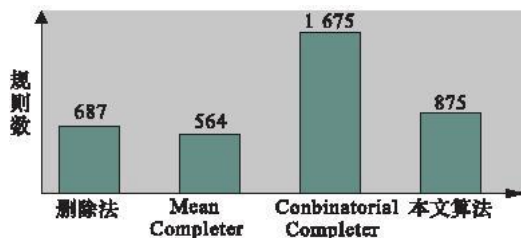


图3 空缺值填补算法得到的规则集规则数比较结果

Fig. 3 Result of comparing of rules' number in rule set acquired from algorithm of blank filling

从图 2-3 可以看出,删除法由于删除了包含空缺值的记录,遗漏的信息较多,因此得出的规则较少。Mean Completer 算法不改变记录集数量,由于选取出现最多的属性值填补空缺值,因此填补结果的错误率较高。Combinatorial Completer 把所有可能取值的组合都添加进数据集,记录数量会随着空缺值数量的增加呈指数增长,极大地增加了后续知识约简环节的计算量,虽然故障的查全率较高,但会导致无效规则数量的增加,因此查准率较低,也降低了产生规则的可靠性。相比较而言,本文算法填补空缺值后没有改变记录集的数量,最大限度地保留了原始信息,提高填补数据的质量和准确性,保证了产生规则的数量和可靠性。

### 3 故障数据知识约简算法

知识约简是在保持信息系统分类能力不变的条件下,使用知识发现算法,删除其中不相关或不重要的冗余属性和冗余值,以简化判断规则<sup>[4-5]</sup>。粗糙集中进行知识约简常用算法有<sup>[8]</sup>:基于差别函数和差别矩阵的算法、基于属性重要性的算法、遗传算法、动态约简、Johnson 贪婪算法等。本文从实际应用需求出发,提出了基于粒计算的知识约简算法。

#### 3.1 基于粒计算的知识约简算法的基本思想

二进制粒矩阵定义<sup>[6]</sup>:设集合  $X$  是论域  $U$  中的另一种划分  $X = \{X_j \mid (1 \leq j \leq q)\}$ , 则一定有:  $\cup X_j = U, \cap X_j \neq \emptyset$ 。令  $Y_{p \times m} \triangleq \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1l} \\ a_{21} & a_{22} & \cdots & a_{2l} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{ml} \end{bmatrix}, X_{q \times m} \triangleq \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1l} \\ b_{21} & b_{22} & \cdots & b_{2l} \\ \cdots & \cdots & \cdots & \cdots \\ b_{n1} & b_{n2} & \cdots & b_{nl} \end{bmatrix}, C_{p \times q} = C_{YX} \triangleq YX^T$ , 定义二进制粒矩阵为  $\{Y_{p \times m}, X_{q \times m}, C_{p \times q}\}$ 。  $C_{ij} = \sum_{k=1}^l (a_{ik} b_{kj})$ ,  $i=1, 2, \dots, p, j=1, 2, \dots, q$ 。  $C_{ij}$  清晰地反映了  $Y_i$  和  $X_j$  中元素的从属关系,显示了  $Y_i$  包含于  $X_j$  的元素的个数。用  $N(i)$  表示  $C_i$  中非零元素的个数,  $C_i$  是矩阵  $C_{p \times q}$  中的第  $i$  行元素。当  $N(i) = 1$  且  $C_{ij} \neq 0$  时,有  $Y_i \subseteq X_j$ ; 当  $C_{ij} \neq 0$  时,有  $Y_i \cap X_j \neq \emptyset$ , 进而有:

$$R_* X = \{ \cup Y_i \mid N(i) = 1 \}, \quad \text{card} R_* X = \sum_{N(i)=1} C_{ij} \quad (1)$$

$$R^* X = \{ \cup Y_i \mid C_{ij} \neq 1 \}, \quad \text{card} R^* X = \sum C_{ij} \quad (2)$$

本算法将决策表进行二进制的粒矩阵处理,通过计算决策属性在条件属性下的正区域和依赖度,判断该属性或属性值是否为冗余。若冗余,则可删除。该算法的计算单位不再是单个属性值,而是二进制的矩阵对象,所以使约简操作变换为矩阵的运算,大大地提高了算法的执行速度。

#### 3.2 基于粒计算的知识约简算法的流程

将数据预处理后的决策表按照二进制数进行知识粒化,并按照二进制粒矩阵进行计算。对于一个属性

$C_i$ , 通过计算其依赖度  $k = \frac{\sum_{N(i)=1} C_{ij}}{\text{card}(U)}$  ( $k$  表示属性对分类能力的影响) 来判定该属性是否被删除。当  $k = 1$  说明该属性对决策表无分类能力,则可删去属性  $C_i$ 。当  $k < 1$  说明删除该属性后分类能力会下降,则不能删去属性  $C_i$ 。对于属性值  $C_j$ , 通过  $U \setminus \text{IND}(C - C_j) = \{Y_i\}$ ,  $U \setminus \text{IND}(D) = \{X_i\}$ , 当  $Y_i \subseteq X_i$ , 则该属性值  $C_j$  可以被省略,否则不能被省略。算法步骤见图 4。

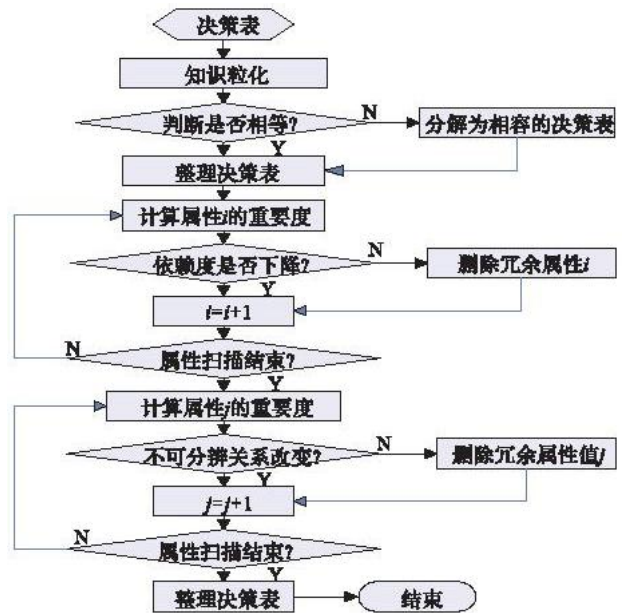


图 4 基于粒计算的知识约简算法流程  
Fig. 4 Flow of knowledge simplification based on granular computing

#### 3.3 知识约简算法比较

将大屏幕投影的 645 条故障记录集做为数据源,使用遗传算法、动态约简(使用穷尽算法作为约简计算

方法)、Johnson 贪婪算法和本文提出的算法,对故障信息进行知识约简,它们约简后的规则数、耗费时间、准确率的比较结果见表 1。

表 1 各种算法的规则数、耗费时间、准确率比较

Tab. 1 Comparing of rules number, time consuming, precision of different algorithm

比较项目	遗传算法	动态约简 (穷尽算法)	Johnson 贪婪算法	粒计算知识 约简算法
规则数	357	332	92	327
耗费时间/ms	1 200	1 600	4 000	1 000
准确率(%)	80.5	84.6	78.2	84.1

可看出,基于粒计算的知识约简算法在耗费时间较少的情况下,得出了较少的规则数量,且准确率较高,保持了较好的知识分类能力,并且由于矩阵运算已经有很成熟的工具和算法,编程实现容易。与经典的粗糙集处理方法相比较降低了运算的时间复杂度,提高了规则的生成效率。

约简算法生成的知识规则并非全部作为规则库的内容,而是需要在保证分类效果的前提下,根据实际需要,确定提取有效规则的标准<sup>[9]</sup>,并将选择出的有效规则加入到规则库中,为故障诊断提供支持。

## 4 结束语

针对我军目前装备普遍存在维修难度大的问题,提出了基于粗糙集的装备故障诊断系统的框架,以大屏幕投影系统的设备故障数据为例,研究数据预处理中对缺失数据的处理方法,提出了基于统计原理的缺失值处理的改进算法,通过对比实验,说明该算法在没有改变记录集数量的情况下,最大限度地保留了原始信息,保证了产生规则的数量和可靠性。在知识约简算法中提出了基于粒计算的粗糙集知识发现约简算法,保持了较好的知识分类能力,并且提高了规则生成速度。本文是以大屏幕投影的设备故障数据为测试对象,但研究的算法可用于各种装备故障诊断系统中。

## 参考文献:

- [1] 侯青剑,王宏力. 基于计算智能的故障诊断方法综述[J]. 战术导弹技术, 2009, 30(1): 12-15.  
HOU Qingjian, WANG Hongli. Survey of Fault Diagnosis Methods Based on Computational Intelligence[J]. Tactical Missile Technology, 2009, 30(1): 12-15. (in Chinese)
- [2] 张凤琴,白钢. 基于粗糙集的大屏幕系统故障诊断研究[J]. 微计算机信息, 2009, 25(19): 108-109.  
ZHANG Fengqin, BAI Gang. Research of Equipment Fault Diagnosis System Based on Rough Set[J]. Micro-electronics Information, 2009, 25(19): 108-109. (in Chinese)
- [3] 肖蕾,李颖晖,赵鸥,等. 基于小波变换的飞机电力作动系统故障诊断[J]. 空军工程大学学报:自然科学版, 2009, 10(5): 55-58.  
XIAO Lei, LI Yinghui, ZHAO Kun, et al. Fault Diagnosis System of Plane Power Based on Small Waves[J]. Journal of Air Force Engineering University: Natural Science Edition, 2009, 10(5): 55-58. (in Chinese)
- [4] Yeung D S, Chen Degang, Tsang E C, et al. On the Generalization of Fuzzy Rough Sets Fuzzy Systems[J]. IEEE Transactions, 2005, 38(5): 89-92.
- [5] Raymond R T. Rule-based Life Cycle Impact Assessment Using Modified Rough Set Induction Methodology[J]. Environmental Modelling & Software, 2005, 24(3): 120-122.
- [6] 张腾飞,王锡淮. 粗糙集理论在故障诊断中的应用综述[J]. 上海海事大学学报, 2005, 26(4): 20-25.  
ZHANG Tengfei, WANG Xihuai. Survey on Application of Rough Set Theory in Fault Diagnosis[J]. Journal of Shanghai Maritime University, 2005, 26(4): 20-25. (in Chinese)
- [7] 苗夺谦,李道国. 粗糙集理论、算法与应用[M]. 北京:清华大学出版社, 2008: 215-220.  
MIAO Duoqian, LI Daoguo. Rough Set theory, Algorithm and Application[M]. Beijing: Tsinghua University Press, 2008: 215-220. (in Chinese)
- [8] Fan Tuan-Fang, Liu Duen-Ren, Tzeng Gwo-Hshiang. Rough Set-based Logics Formulticriteria Decision Analysis[J]. European Journal of Operational Research, 2007, 182: 340-355.
- [9] HAN Zhixin, Swiniarski R W, Skowron A. Rough Set Method in Feature Selection and Recognition[J]. Pattern Recognition Let-

ter,2003,24(6):833-849.

(编辑:徐敏)

## Algorithm Research on Equipment Fault Diagnosis System Based On Rough Set

ZHANG Feng-qin<sup>1</sup>, ZHANG Shui-ping<sup>1</sup>, BAI Gang<sup>2</sup>, YAN Xiao-mei<sup>1</sup>, ZHANG Yue-ling<sup>1</sup>

(1. Telecommunication Engineering Institute, Air Force Engineering University, Xi'an 710077, China; 2. Unit 95880, Beijing 100843, China)

**Abstract:** By designing the prime frame of equipment fault diagnosis system and using equipment fault records of large screen system as a sample data, this article demonstrates and analyzes the application of Rough Set in the equipment fault diagnosis system. The following work has been done; by the method of combining the Rough Set theory with the Statistical theory, the datasets with missing values of records can be recovered; the algorithm of knowledge acquisition of Rough Set based on Granular Computing can be used to transform the attribute in decision sheet into matrix objects, then the complex knowledge simplification is transformed into matrix operation, thus the time consuming can be deducted, and the knowledge of fault diagnosis can be acquired automatically and more effectively.

**Key words:** rough set; faults diagnosis; datasets with missing values; knowledge simplification

(上接第41页)

## Research on Data Merging Together with Optimization and Reliable Transmission of Cooperation Spectrum Sensing

WANG Yin-he, LI Yan, NIU De-zhi

(Science Institute, Air Force Engineering University, Xi'an 710051, China)

**Abstract:** In order to improve spectrum utilization ratio, it is very necessary to enhance the practicability and veracity of the cooperation spectrum sensing. This paper analyzes and discusses the key technology of cooperation spectrum sensing and data transmission reliability by using the way of multipath data to increase the reliability and practicability during the data transmission process. By this means the data in every perceptive node can be transmitted to the central node more exactly, reliably and practically to make a quick and accurate judgment. And using this method the influences caused by interference and noises during the data transmission process can be overcome. Simulation results show that the method is feasible and lays a foundation for practical application of cooperation spectrum detection technology.

**Key words:** cognitive radio; cooperation spectrum sensing; reliable transmission; multipath