

文章编号: 1000-8055(2008)10-1842-05

基于变精度粗糙集的航空发动机故障诊断

索中英, 朱林户, 吴 华, 苏 强
(空军工程大学 工程学院, 西安 710038)

摘 要: 针对航空发动机故障诊断的实际情况, 在分析粗糙集理论缺陷的基础上, 引入变精度粗糙集模型对专家给出的样本集进行处理, 并将所得到的极小化决策算法用于历史故障样本集的分析, 得到了高的识别率, 从实践的角度证明了该方法的有效性以及小样本情况下所得决策算法的普适性。

关 键 词: 故障诊断; 航空发动机; 粗糙集; 变精度粗糙集

中图分类号: V233

文献标识码: A

Aeroengine fault diagnosis based on variable precision rough sets

SUO Zhong-ying, ZHU Lin-hu, WU Hua, SU Qiang

(The Engineering Institute, Air Force Engineering University, Xi'an 710038, China)

Abstract: According to the practical conditions of aeroengine fault diagnosis, the paper firstly analyzed limitation of rough sets, then processed the samples set from experts experience by introducing variable precision rough sets model. And the samples sets of history faults were analyzed by the received decision-rules. The rate of identification was high. The result proves the effect of this method and the applicability of the decision-rules received from the small samples set.

Key words: fault diagnosis; aeroengine; rough sets; variable precision rough sets (VPRS)

航空发动机是非常昂贵且复杂的设备, 因其复杂的结构及高速的恶劣工作环境, 可靠性很难保证, 发生故障后可能引发飞行事故, 后果严重。因此, 对于它进行及时的状态监测和故障诊断具有非常重大的实际意义^[1]。航空发动机故障诊断是保证发动机安全工作, 延长发动机使用寿命的一种重要的技术措施。

故障诊断技术经过近几十年的研究, 取得了长足的发展^[2]。概括地讲, 诊断方法主要有以下三类: ①基于解析模型的方法, 如参数估计方法、状态估计方法和等价空间方法等。②基于信号处理的方法, 如相关分析、频谱分析、小波分析等。③基于知识的方法, 如智能诊断、模糊推理、神经

网络等。

毋庸置疑, 基于知识的诊断方法应该是一种很有前途的方法, 尤其是在非线性系统领域。它的智能化技术和丰富的专家知识给用户提供了一个简单易用而又可靠的系统。然而, 一个众所周知的原因, 制约着这项技术向实用化的推广, 那就是故障样本数的不足^[1]。因此, 尝试对原有方法的改进就显得很有实际意义。

航空发动机故障诊断需要在发动机运行过程中连续监测多个参数, 参数本身的值及相对之间的变化关系, 分别对应着发动机的诸多状态。其对应关系非常复杂, 很难用函数关系描述, 可以用基于知识的方法来解决。而且对于航空发动机的多

收稿日期: 2007-09-28; 修订日期: 2008-02-27

基金项目: 总装备部科研资助项目

作者简介: 索中英(1981 -), 女, 山西忻州人, 博士生, 主要从事智能信息处理与决策的研究。E-mail: csy_316 @163. com

数故障而言,故障样本的数目极其有限,属于典型的小样本集问题^[1].

鉴于此,本文运用变精度粗糙集理论挖掘不确定数据间的关系,发现潜在知识,力求将由存在大量冗余信息和矛盾信息的小规模的对象集中得到的结论,应用到大规模的对象中去^[3]. 提高小样本情况下,所得结论的普适性.

1 粗糙集理论介绍

粗糙集(Rough set, RS)理论是由波兰的 Pawlak Z 教授于 20 世纪 80 年代初提出的,被认为是处理具有不精确和不完全信息问题的新的数学工具.

1.1 粗糙集的基本概念^[4-7]

1) 信息系统与决策表

RS 理论主要研究了由对象集和属性集构成的二元组数据结构 $S = (U, A)$, 称之为知识表达系统或信息系统. 信息系统 S 可方便地表示为数据表的形式,称为决策表,因此信息系统也常被称为决策系统.

2) 不可分辨性

若 $P \subseteq R$, 且 $P \neq \emptyset$, 则 P 中所有等价关系的交集 $\cap P$ 也是一个等价关系,称为 P 上的不可分辨关系,记为 $\text{ind}(P)$, 且有 $[x]_{\text{ind}(P)} = \bigcap_{r \in P} [x]_r$.

3) 条件属性的约简

对于系统 $S = (U, A)$, 任何最小集 $P \in A$, 且 $\text{ind}(A) = \text{ind}(P)$ 定义为系统 S 的一个约简. 记 $\text{red}(A)$ 表示所有的约简集, 它们的交集称为系统 S 的核.

4) 决策规则的约简

决策规则定义为

$$r_{i,j}: \text{des}(X_i) \rightarrow \text{des}(Y_j), Y_j \cap X_i \neq \emptyset$$

规则的确定性因子

$$\mu(X_i, Y_j) = |Y_j \cap X_i| / |X_i|,$$

$$0 < \mu(X_i, Y_j) \leq 1 \quad (1)$$

其中 $\text{des}(X_i)$ 和 $\text{des}(Y_j)$ 分别表示等价类 X_i 和 Y_j 对于各条件属性值和各决策属性值的特定取值, $X \neq \emptyset, |X|$ 表示 X 的基数.

决策规则的约简是利用决策逻辑分别消去决策算法中每一个决策规则的不必要条件,它不是整体上的简化属性,而是针对每一个决策规则,去掉表达该规则的冗余属性值,以便进一步简化决策算法.

5) 决策算法的极小化

一个基本算法 F 称为是极小的,如果其中的每一个规则是既约的且对每一个决策规则 $\phi \rightarrow \Psi$, $F\Psi$ 是既约的.

其中,一个决策规则集 $F\Psi$ 称为是既约的,如果它是自己的一个约简.

限于篇幅,关于属性与规则的获取和简化算法可参见相关文献^[6].

1.2 基本粗糙集的缺陷

基本 RS 理论是严格按照等价类进行分类的,它的一个局限性是所处理的分类必须是完全正确或肯定的,而没有某种程度上的“包含”或“属于”. 它的另一个局限性是所处理的对象是已知的,且所得的决策规则仅对这些对象包含的集合能得到正确的结论,当应用到更广泛的对象数据集的诊断时无法保证较高的正确诊断率,即容错性不强^[6].

因此在数据集存在噪声等干扰以及数据小样本情况下,基本 RS 模型会由于对数据的过拟合使其对新数据的预测或分类能力大为降低. 为增强 RS 模型的抗干扰能力, Ziarko 提出了一种变精度粗糙集模型,通过引入一个误差精度,使其具有一定的容错性^[3].

1.3 变精度粗糙集模型^[3]

VPRS (Variable precision rough sets) 模型是基本 RS 模型的扩充,它在基本 RS 模型中引入分类误差参数 β ($0 \leq \beta \leq 0.5$),即允许一定程度的错误分类率存在,从而产生具有一定不确定性的近似分类,实现在整体上较优的信息处理能力. 一方面完善了近似空间的概念,另一方面有利于用粗糙集理论从认为不相关的数据中发现相关信息. 当 $\beta = 0$ 时, VPRS 模型就是基本 RS 模型了.

定义 1: 设 X, Y 是有限论域 U 的两个非空子集. 如对于任意 $x \in X$, 有 $x \in Y$, 则称 Y 包含 X , 记为 $Y \supseteq X$. 令

$$c(X, Y) = \begin{cases} 1 - \text{card}(X \cap Y) / \text{card}(X), & (\text{card}(X) > 0) \\ 0, & (\text{card}(X) = 0) \end{cases} \quad (2)$$

称 $c(X, Y)$ 为集合 X 关于集合 Y 的相对错误率.

定义 2: 令 $0 \leq \beta \leq 0.5$, 多数包含关系定义为 $Y \supseteq_{\beta} X \Leftrightarrow c(X, Y) \leq \beta$.

定义 3: 对于 $X \subseteq U$, 定义 X 的 β 下近似为 $r_{-\beta}(X) = \bigcup \{E \in U / r \mid c(E, X) \leq \beta\}$, X 的 β 上近似为 $r_{\beta}(X) = \bigcup \{E \in U / r \mid c(E, X) < 1 - \beta\}$.

定义 4: 设信息系统 $S = (U, A)$, $P, Q \subseteq A$ 分别为条件属性集和决策属性集. 定义 β 依赖性为

$$\gamma(P, Q, \beta) = \frac{\text{card}(\bigcup_{Y_j \in U/Q} (\text{ind}(P) - \beta(Y_j)))}{\text{card}(U)} \quad (3)$$

它描述了系统对执行具有分类误差 β 的对象分类质量性能的好坏.

在 VPRS 模型中, 知识的约简是指选择和决策属性 Q 的 β 依赖性相同的最小条件属性子集, 且最小条件属性子集不改变原系统的相容性. 从而有下列 β 近似约简: 设 $\text{red}(P, Q, \beta)$ 是条件属性 P 的一个子集, 且满足

- ① $\gamma(P, Q, \beta) = \gamma(\text{red}(P, Q, \beta), Q, \beta)$;
- ② 从 $\text{red}(P, Q, \beta)$ 中去掉任何一个属性, 都将使 ① 不成立.

由上述讨论可知, VPRS 模型扩充了基本 RS 粗糙集理论, 更好地体现了数据分析中数据的相关性, 从而为获取近似决策规则奠定了理论基础.

2 实例分析

随着航空发动机性能的不断提高, 使得发动机各个部件的工作状况变得更加恶劣. 滑油的消耗量及滑油中金属含量值能比较准确反映发动机的轴承、附件和齿轮的工作情况. 当滑油消耗量或滑油中某些金属含量偏高, 则说明轴承、机匣或齿轮磨损厉害. 通过对滑油消耗量及滑油中镁、铝、铁和铜含量值及其趋势监控, 可达到对发动机轴承、机匣或齿轮工作情况的监控, 由此判断发动机相关部件的磨损状况^[1].

依据 24 组专家诊断样本建立基于粗集的某航空发动机的油液故障特征模型库, 如表 1 所示.

表 1 中将该航空发动机油液分析指标作为粗集的属性集 $\{a, b, c, d, e, f, g, h\}$, 分别为铁、铝、镁、铜在正常滑动和严重滑动下的磨粒以及油渣微粒的铁谱数量等级. 为便于粗集处理, 已采用 Naive Scaler 算法将其离散化为三种状态: 1(少量)、2(中量)、3(大量), 离散化标准如表 2 所示.

诊断结果作为目标属性 J 取值 J_1, J_2, J_3 分别对应发动机正常工作、轻微故障、严重故障.

从表 1 中看出, 24 位专家给出的诊断样本中存在着大量的冗余信息和矛盾信息. 并且对该故障特征模型库处理的最终目的是将其结果应用于大规模的故障对象的诊断. 鉴于这两点, 本文将采用变精度粗糙集模型来对其进行处理.

2.1 约 简

采用基本粗集中区分矩阵和区分函数的方法

对模型库进行条件属性的约简.

表 1 某航空发动机故障决策表

Table 1 Decision-making sheet of a certain aeroengine fault

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>	<i>i</i>	<i>J</i>
1	2	1	1	1	1	1	1	2	1	J_0
2	1	1	1	1	1	1	1	1	1	J_0
3	1	1	1	1	1	1	1	1	1	J_0
4	1	1	1	1	2	1	1	1	1	J_0
5	1	1	1	1	2	1	1	1	1	J_0
6	1	1	1	1	2	1	1	1	1	J_0
7	1	1	1	1	2	1	1	1	1	J_0
8	1	1	1	1	2	1	1	1	1	J_1
9	1	1	2	1	1	2	2	1	1	J_1
10	2	2	3	2	2	3	3	2	2	J_1
11	1	1	2	1	1	2	2	1	1	J_1
12	1	1	2	1	1	2	2	1	1	J_1
13	1	1	2	1	1	2	2	1	1	J_1
14	1	1	2	1	1	2	2	1	1	J_1
15	1	1	2	1	1	2	2	1	1	J_1
16	1	1	2	1	1	2	2	1	1	J_1
17	1	1	2	1	1	2	2	1	1	J_1
18	1	1	2	1	1	2	2	1	1	J_1
19	1	1	2	1	1	2	2	1	1	J_2
20	3	3	2	2	1	1	2	2	1	J_2
21	2	1	1	1	1	1	1	2	1	J_2
22	2	1	1	1	1	1	1	2	1	J_2
23	2	1	1	1	1	1	1	2	1	J_2
24	2	1	1	1	1	1	1	2	1	J_2

表 2 铁谱数量等级的离散化标准

Table 2 Discrete standard of quantity rank of metal spectrum

量化值	1	2	3
铁谱数量等级	[0.0, 3.0]	[3.0, 7.0]	[7.0, 10.0]

所得的区分函数

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

由上式可知: 模型库的核为 $\{e\}$, 从多种约简中选择属性最少的子集之一 $\{a, c, e\}$ 进行接下来的处理.

当 β 取不同值时系统对执行具有分类误差 β 的对象分类质量性能的好坏对比如表 3 所示.

表 3 对象分类质量对比表

Table 3 Comparison table of classified quality in samples

β	0	0.1	0.2
$\gamma(P, Q, \beta)$	0.17	0.58	1.0

因此本文在 $\beta=0.2$ 情况下,进行知识约简处理.对属性子集 $\{a,c,e\}$ 进行进一步的近似约简

$$\mathcal{V}(P,Q,0.2) \neq \mathcal{V}(P-\{a\},Q,0.2);$$

$$\mathcal{V}(P,Q,0.2) \neq \mathcal{V}(P-\{c\},Q,0.2);$$

$$\mathcal{V}(P,Q,0.2) = \mathcal{V}(P-\{e\},Q,0.2).$$

由上式可知:在精度 $\beta=0.2$ 的条件下,条件属性可进一步约简为子集 $\{a,c\}$,可得决策算法规则以及规则的确定性因子

$a2c1 \rightarrow J_0$ 确定性因子为 0.2

$a2c1 \rightarrow J_2$ 确定性因子为 0.8

$a1c1 \rightarrow J_0$ 确定性因子为 0.86

$a1c1 \rightarrow J_1$ 确定性因子为 0.14

$a1c2 \rightarrow J_1$ 确定性因子为 0.9

$a1c2 \rightarrow J_2$ 确定性因子为 0.1

$a2c3 \rightarrow J_1$ 确定性因子为 1.0

$a3c2 \rightarrow J_2$ 确定性因子为 1.0

进一步,对决策算法规则进行决策规则约简和决策算法极小化得

$a1c1 \rightarrow J_0$

$a1c2 \vee c3 \rightarrow J_1$

$a2c1 \vee a3 \rightarrow J_2$

2.2 验证

表 4 中所示为 13 组前述发动机的历史故障记录,采用表 2 离散化标准进行处理后的结果.

表 4 某航空发动机历史故障类型知识库
Table 4 Knowledge base of History faults
of a certain aeroengine

	a	b	c	d	e	f	g	h	i	J
1	1	1	1	1	2	1	1	1	1	J_0
2	1	1	1	1	1	1	1	1	1	J_0
3	1	1	1	1	1	1	1	1	1	J_0
4	1	1	2	1	1	2	2	1	1	J_1
5	1	2	3	2	3	3	3	2	2	J_1
6	2	1	2	1	1	1	1	2	1	J_2
7	3	3	2	2	1	1	2	2	1	J_2
8	2	1	1	1	1	1	1	2	1	J_2
9	2	2	3	2	2	3	3	2	2	J_1
10	2	1	1	1	1	1	1	3	1	J_2
11	1	1	1	1	1	1	1	1	1	J_0
12	3	3	2	2	3	1	2	3	2	J_2
13	1	1	2	1	2	3	3	2	1	J_1

依据 2.1 节得到的极小算法,对其进行诊断与实际故障类型对比如下:

由表 5 可得,极小化决策规则对该 13 组历史故障类型的识别率为 $(12/13) \times 100\% = 92.3\%$.

表 5 故障类型对比表

Table 5 Comparison table of fault types

样本	故障类型	诊断结果
1	J_0	J_0
2	J_0	J_0
3	J_0	J_0
4	J_1	J_1
5	J_1	J_1
6	J_2	不能识别
7	J_2	J_2
8	J_2	J_2
9	J_1	J_1
10	J_2	J_2
11	J_0	J_0
12	J_2	J_2
13	J_1	J_1

3 结论

VPRS 模型通过在基本 RS 模型中引入分类误差参数 $\beta(0 \leq \beta \leq 0.5)$,即允许一定程度的错误分类率存在,从而克服了基本 RS 模型由于对数据的过拟合使其对新数据的预测或分类能力大为降低的缺陷.

VPRS 模型产生具有一定不确定性的近似分类,实现在整体上较优的信息处理能力,有利于用粗糙集理论从认为不相关的数据中发现相关信息,从而可以将一些小规模的对象集中得到的结论应用到大规模的对象集中去.

由于种种原因,对于航空发动机的多数故障而言,故障样本的数目及其有限,属于典型的小样本集问题.

由此分析,将 VPRS 模型应用于航空发动机的故障诊断不失为一种有意义的尝试,而且通过具体的实例分析结果:92.3%的识别率,也从实践的角度验证了方法的有效性.

然而,在应用的过程中也发现了 VPRS 模型的不足:对具有矛盾的诊断规则的处理过于粗糙简单化,缺乏深入的理论刻画.还有待理论方面的进一步完善以期获得更好的结果.

参考文献:

[1] 李应红,尉询楷,刘建勋.支持向量机的工程应用[M].北京:兵器工业出版社,2004.
LI Yinghong, WEI Xunkai, LIU Jianxun. Application of support vector machine to engineering[M]. Beijing: Enginery Press,2004. (in Chinese)

- [2] 李凌均,张周锁,何正嘉. 支持向量机在机械故障诊断中的应用研究[J]. 计算机工程与应用,2002,19:19-21.
LI Lingjun, ZHANG Zhousuo, HE Zhengjia. Application research on support vector machines in mechanical fault diagnosis [J]. Computer Engineering and Applications, 2002,19:19-21. (in Chinese)
- [3] Ziarko W. Variable precision rough set model[J]. Journal of Computer and System Science, 1993, 46(1) : 39-59.
- [4] Pawlak Z. Rough sets[J]. International Journal of Information and Computer Sciences,1982,11:341-356.
- [5] Trunq N T. Rough set approach to domain knowledge approximation [R]. Fundamenta Informaticae, 2004, 59 (2-3) :261-270.
- [6] 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.
ZHANG Wenxiu, WU Weizhi, LIANG Jiye, et al. Rough sets theory and methods [M]. Beijing: Science Press, 2001. (in Chinese)
- [7] 邢敬华,赵新泽,严新华. 基于粗集理论的柴油机典型故障的磨粒识别[J]. 内燃机学报,2005,23(1) :92-95.
XING Jinghua, ZHAO Xinze, YAN Xinhua. Wear particles recognition of typical fault in diesel engine based on rough sets[J]. Transactions of CSICE, 2005,23(1) :92-95. (in Chinese)