

粗糙集约简在飞行器故障识别中的应用

赵方¹ 王静¹ 杨海龙²

(北京林业大学信息学院 北京 100083)¹ (航天东方红卫星有限公司 北京 100094)²

摘要 故障识别是飞行器结构健康监测的重要内容,不同类型的结构测量数据之间存在互补和矛盾的成分。粗糙集约简具有良好的数据挖掘和归类能力,为处理多传感器冗余信息提供了有效的工具。首先从飞行器结构的位移、加速度、应变响应以及模态参数中提取综合的故障评价指标,然后利用粗糙集约简技术对特征属性进行约简,求得系统的最简特征集合,最后通过概率神经网络进行了结构故障的识别。直升机结构的损伤仿真算例表明粗糙集约简方法不仅可以显著减低特征属性的维度,而且能提高故障识别的精度。

关键词 粗糙集,飞行器,故障识别

中图分类号 TP18 **文献标识码** A

Application of Rough Set Reduction in Spacecraft Fault Diagnosis

ZHAO Fang¹ WANG Jing¹ YANG Hai-long²

(School of Information Science, Beijing Forestry University, Beijing 100083, China)¹

(DFH SATELLITE CO. LTD, Beijing 100094, China)²

Abstract Damage detection of the aerospace structures is a critical issue, during which the data from multiple types of sensors are redundant and inconsistency to each other. Rough set reduction is an effective tool for data mining and classification, which can be used for the above problems. Firstly, the damage features are extracted from the multi-sensor information (e. g. displacement, acceleration or strain sensor) and modal parameters. Then, the rough set reduction technique is employed to obtain the core set of all features. Finally, the structural damage conditions are identified through probabilistic neural network. The numerical simulation of a helicopter demonstrates that the rough set reduction technique can not only decrease the spatial dimension of feature attributions, but also improve the identification accuracy.

Keywords Rough set reduction, Spacecraft, Fault diagnosis

1 引言

粗糙集(Rough sets)理论是 1982 年由 Z. Pawlak 教授提出的一种用于处理不确定性和含糊性知识的数学工具。该理论已经在数据挖掘、机器学习、过程控制、决策分析和模式识别等领域得到了广泛的应用,并取得了良好的效果^[1]。飞行器结构部件多、系统复杂,在健康监测过程中,通常会同时使用多个、不同类型的传感器。各种不同类型的传感器信息提供的数据不可避免地存在冗余和矛盾的部分。如何充分利用这些测试数据提取结构异常或故障信息,是能否完成结构故障识别及健康监测的关键。

基于粗糙集的信息约简是处理不精确、不确定、不一致和冗余信息的有效工具,在航空航天故障识别领域得到了极大的重视。通过粗糙集数据约简,可以简单、直观地删除信息系统的多余属性,得到系统的最小条件属性集合。

2 粗糙集约简理论

设 $X, Y \in U, R$ 是一个定义在 U 上的等价关系族,集合 X 关于 R 的下近似集定义为

$$\underline{R}(X) = \cup \{Y \in U / IND(R); Y \subseteq X\} \quad (1)$$

式中, $\underline{R}(X)$ 是根据现有知识判断肯定属于 X 的对象组成的最大集合,称为正域,记为 $POS(X)$;类似地, X 关于 R 的上近似集定义为:

$$\overline{R}(X) = \cup \{Y \in U / IND(R); Y \cap X \neq \emptyset\} \quad (2)$$

$\overline{R}(X)$ 是所有与集合 X 相交非空的等效类的并集,是有可能属于 X 的对象组成的最小集合。

集合边界区域定位为:

$$Bnd(X) = \overline{R}(X) - \underline{R}(X) \quad (3)$$

如果 $Bnd(X)$ 是空集,则称 X 关于 R 是清晰的;反之,则称 X 为 R 的粗糙集^[2]。

一个知识系统 S 可以表示为: $S = \langle U, A, V, f \rangle$, 其中 U 是对象集,也称为论域; A 是属性集,属性 A 可以表示为: $A = C \cup D$ 并且 $C \cap D = \emptyset$, 其中 C 为条件属性集合, D 为决策属性集合,则该信息系统称为信息决策系统或决策表。

在信息系统中,每个属性子集 $R \subseteq A$,不可分辨关系 $IND(R)$ 为:

$$IND(R) = \{(x, y) \in U^2 : \forall r \in R, r(x) = r(y)\} \quad (4)$$

显然 $IND(R)$ 是一个等价关系,对象 x 在属性集 R 上的

赵方(1958-),男,博士,副教授,主要研究方向为软件开发方法、软件开发模式、软件测试、企业信息技术等, E-mail: zhaofang1958@163.com; 王静(1983-),女,研究生,主要研究方向为计算机软件设计、仿真技术。

等价类 $[x]_R$ 定义为:

$$[x]_R = \{y \in U \mid (x, y) \in IND(R)\} \quad (5)$$

属性约简是在保持信息系统的分类或决策能力不变的情况下,删除其中的冗余属性。如果 $IND(R) = IND(R - r)$,那么称 r 在 R 中是可省略的。决策表约简的目的是用最少的条件辨别每一个决策表,即完全确定决策属性的值,对决策表进行条件属性约简。

信息系统 R 中所有不可约简属性构成的集合称为属性集的核集,记作 $CORE(R)$ 。核集是各约简集合的交集,去掉核集中的任意属性都将导致整个知识系统分类能力的下降。

3 算法实现

利用飞行器结构的多传感器测量数据(位移、应变和加速度数据)建立系统的信息决策表,通过粗糙集分析技术确定各属性的重要程度和约简集合,具体包括下面两部分内容:

(1) 利用K均值聚类算法对特征参数集合进行预处理,计算归一化的信息决策表。

(2) 计算属性集合的核集和每个条件属性的重要程度,求取系统的最简条件属性集合。

3.1 提取故障评价指标

对于任意一个信号 $f(t)$,以 $\psi(t)$ 为小波基函数,经过 j 阶小波包分解,可以得到:

$$f_j(t) = \sum_{j,k=-\infty}^{\infty} c_{j,k} \psi_{j,k}(t) \quad (6)$$

小波包系数:

$$c_{j,k} = \int_{-\infty}^{\infty} f(t) \psi_{j,k}(t) dt \quad (7)$$

对于第 j 个不同的小波包频段,各组分信号的能量为:

$$E(j, n) = \sum_{j=1}^m \sum_{k=0}^{2^j-1} c_{j,k}^2 \quad (m < 2^j) \quad (8)$$

设 $E_i = E(i, n)$,则归一化的小波包能量指标 $NWEI$ 定义如下:

$$NWEI_i = \frac{E_{\bar{a}} - E_{\underline{a}}}{E_{\bar{a}}} \quad (9)$$

式中, $E_{\bar{a}}$ 和 $E_{\underline{a}}$ 分别是结构在健康和故障时第 i 频段的小波包能量。

参考姜绍飞等^[3]关于结构故障识别指标的研究,定义归一化的应变能变化率指标($NSEC$)、归一化的频率变化率指标($NFCI$)以及归一化的混合故障评价指标($NMDC$)如下:

$$NSEC_i = \frac{SEC_i(k)}{\sum_{j=1}^n |SEC_j(k)|} \quad (10)$$

$$NFCI_i = \frac{FCI_i}{\sum_{j=1}^n |FCI_j|} \quad (11)$$

$$NMDC_i = \frac{MDC_i(k)}{\sum_{j=1}^n |MDC_j(k)|} \quad (12)$$

式中, $SEC_i = \frac{\{SE_{\bar{a}}\} - \{SE_{\underline{a}}\}}{\{SE_{\bar{a}}\}}$, $FCI_i = \frac{f_{\bar{a}} - f_{\underline{a}}}{f_{\bar{a}}}$, $MDC_i =$

$\frac{\{SE_{\bar{a}}\} - \{SE_{\underline{a}}\}}{f_{\bar{a}}^2 - f_{\underline{a}}^2}$ 分别是第 i 阶模态对应的应变能量变化率、模态频率变化率和混合故障指标; n 为测量节点总数; m 为模态总数; $f_{\bar{a}}$, $f_{\underline{a}}$ 是结构故障前后的第 i 阶模态频率; $\{SE_{\bar{a}}\}$ 和 $\{SE_{\underline{a}}\}$ 分别是第 i 阶模态时,结构故障前后的应变能量向量。

特征指标 $NWEC$ 、 $NSEC$ 、 $NFCI$ 和 $NMDC$ 充分利用位移、应变和加速度数据,实现了原始测量信号的数据层融合,能从不同的侧面反映出结构的故障水平,从而对其健康状况进行更加全面的评估。图1为结构故障诊断路线图。

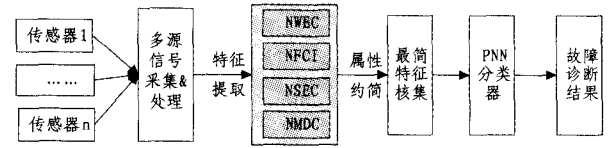


图1 结构故障诊断路线图

3.2 属性约简算法

属性约简是在保持原有信息分类能力不变的前提下,通过知识约简找出系统的重要信息和分类规则。它无需提供相关数据集合以外的任何先验信息,适合于发现数据库中隐含的、潜在有用的规律,找出其内部数据的关联关系和特征。

运用粗糙集理论进行属性约简时,要求信息决策表中的值是统一的离散数据。对于不同传感器的测量数据,需要首先对条件属性或决策属性的值域进行离散化处理,然后将离散值进行合并和归一化处理,得到统一的离散数据分布。利用归一化后的条件属性和决策属性集合构造系统信息决策表,才能够计算其核集,并按照每个条件属性的重要程度,求出最简属性集合。

(1) K-means 聚类

K-means算法是一种聚类分析算法^[4],该算法把 n 个样本点分为 k 个簇($k < n$),各簇内的样本点具有较高的相似性,而各簇间的样本点相似程度较低,相似度的计算是依据一个簇中样本点的平均值来进行的。

K-means离散化的实质就是在保持决策表分类能力不变,即条件属性和决策属性相对关系不变的条件下,寻找合适的分割点集,并对条件属性构成的空间进行划分。而连续属性离散化就是在特定的连续属性值域范围内设定若干个离散化划分点,将属性的值域范围划分成一些离散的区间,再用不同的符号或整数值代表属于每个区间的属性值。

(2) 提取约简集合

粗糙集属性约简的思想为:利用相对的正域求出相容与不相容信息系统的核集。然后,以属性的依赖程度定义属性的重要度,设定属性的重要度阈值,选择重要度大的属性加入约简集,直到满足终止条件为止。具体运算步骤如下:

输入 决策表系统 $S = \langle U, A, V, f \rangle$

输出 属性约简 Red

Step1 取 $CORE(A) = \emptyset$;

Step2 对于每个条件属性 $q \in A$

计算相对正域 $POS_{A-\{q\}}(d)$

如果 $POS_{A-\{q\}}(d) \neq POS_A(d)$

$CORE(A) = CORE(A) + \{a\}$;

结束

Step3 $Red = CORE(A)$;

Step4 计算约简集合的依赖程度 $k_R(d)$,设 ϵ 为设定的条件属性的重要程度阈值;

Step5 如果 $k_R(d) \geq \epsilon$,则计算 Step10,否则继续向下计算;

Step6 对于每一个 $q \in A - Red$,计算属性的重要程度 $SIG(q, R, d)$;

Step7 选择重要性最大的属性 q , $Red = Red \cup \{q\}$;

Step8 计算约简属性集 Red 的依赖度 $k_R(d)$;

Step9 如果 $k_R(d) \geq \epsilon$, 继续结果输出, 否则转向 Step6, 进行循环计算;

Step10 输出条件属性的约简集合 Red.

3.3 PNN 故障识别

运用概率神经网络进行结构故障模式的识别。PNN 分类计算要求有大量的训练和测试数据, 用下式添加测量噪声以模拟测量数据。

$$S_i = S_i^0 \times (1 + \epsilon R) \quad (13)$$

式中, S_i 是加噪后 $NWEC$, $NFCI$, $NSEC$ 和 $NMDC$ 的第 i 个分量; S_i^0 是加噪前的故障识别指标; R 是按标准正态分布的随机序列; ϵ 是相对噪声水平。

4 仿真算例

本文采用北航 FH-1 共轴式无人直升机(见图 2)模型验证多源信息融合在结构故障识别中的应用。直升机机身是一种典型的框架结构, 通过降低直升机机身斜撑刚度的方法来模拟结构的故障, 共包括 6 种结构故障状态与健康状态。故障为发生在 3 个不同位置的两种不同程度的故障(20%和 40%), 如表 1 所列。

表 1 结构故障形式

故障类型	故障程度	故障位置
Pattern 1	弹性模量降低 20%	斜撑 6423-8323 与 6579-8323 故障
Pattern 2	弹性模量降低 40%	斜撑 6423-8323 与 6579-8323 故障
Pattern 3	弹性模量降低 20%	斜撑 6423-8323 与 6660-8593 故障
Pattern 4	弹性模量降低 40%	斜撑 6423-8323 与 6660-8593 故障
Pattern 5	弹性模量降低 20%	斜撑 4791-6423, 3203-5133 及 6660-8593 发生故障
Pattern 6	弹性模量降低 40%	斜撑 4791-6423, 3203-5133 及 6660-8593 发生故障

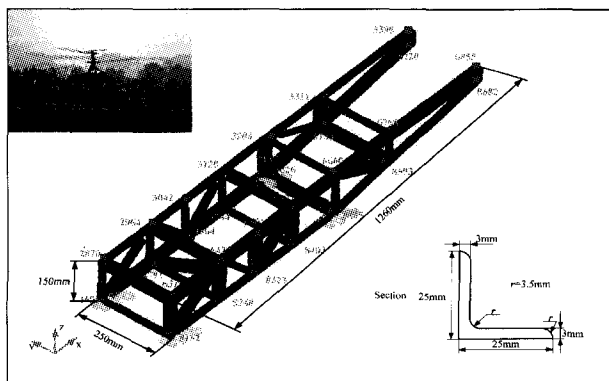


图 2 直升机机身结构图

表 2 不同噪声水平下, 属性约简前后识别结果

识别精度 (IA)/%	噪声水平与输入矩阵 IN									
	0.06		0.07		0.08		0.09		0.1	
	1-40	1-17	1-40	1-17	1-40	1-17	1-40	1-17	1-40	1-17
Pattern 1	99	99	97	97	94	91	91	88	89	87
Pattern 2	100	99	92	97	92	94	90	95	89	90
Pattern 3	95	100	94	99	96	100	86	95	81	96
Pattern 4	93	100	88	97	85	97	81	97	82	93
Pattern 5	100	100	100	100	100	100	100	100	100	100
Pattern 6	100	100	100	100	100	100	100	100	100	100
总精度(IA)	97.8	99.7	95.2	98.3	94.5	97.0	91.3	95.8	90.3	94.3

算例仿真结果表 2 和图 3 表明, 对直升机机身结构的故障进行约简属性集输入时, PNN 的故障识别精度比输入之前

4.1 特征参数提取

依据文献[9,10], 对直升机机身的有限元模型, 在 MSC/NASTRAN 中进行正交模态分析。分别提取机身的前十阶模态频率和每阶模态的应变能, 利用上述仿真数据计算特征参数 $NSEC$, $NFCI$ 和 $NMDC$ 。

按照直升机的实际结构形式, 对机身与起落架的连接处的 4 个节点 4685、4686、4693 和 4694 施加六自由度完全约束。在直升机机身支架 3120-6579 的中点施加 $100\sin(2\pi \cdot 40 \cdot t)N$ 的载荷作用, 对直升机机身进行瞬态响应分析。载荷作用时间为 2s, 时间间隔为 0.001s, 模拟旋翼额定转速时 (1200RPM) 作用在机身的周期性激励。以 1000Hz 的采样频率, 记录机身各节点的位移响应。以此, 提取小波包能量特征指标 $NWEI$, 连同上述模态相关指标一起作为 PNN 融合计算和属性约简的初始输入数据。

4.2 粗糙集约简

为了对结构的故障状况进行全面的了解, 采用大量不同类型的传感器, 因而数据冗余现象严重, 这给 PNN 网络训练[5]带来了难度, 同时不利于清晰地对识别影响因素进行分析。粗糙集约简是一种处理大量、不精确及不确定数据的方法, 能够从各种不完备的数据信息中发掘隐含的规律, 去除冗余的信息。

本文采用 K-means 聚类算法和求取特征属性的核集来对属性决策表进行约简。首先, 使用 MATLAB 函数库中的 KMEANS 函数对样本的条件属性集进行聚类分析, 选取聚类中心为 5 个, 通过迭代计算, 使每个簇内所有点到中心点的距离最小, 将数据聚为 5 簇, 用每个样本数据所属的簇来表示其属性值。

利用聚类后的特征因子集合构造结构故障识别决策表。其中, 条件属性由 4 个特征指标的 40 个分量组成, 决策属性为 6 种故障模式, 分别考虑噪声水平为 0.06, 0.07, 0.08, 0.09 及 0.1 时的情况。得到特征因子的核集为: 1~17, 即前 17 个条件属性集合 $NWEC_1, NWEC_2, \dots, NWEC_{11}, NFCI_1, \dots, NFCI_6$, 而其余的条件属性是冗余的。这表明属性约简能大大降低 PNN 的空间维数。此外, 结构 11 个位置处的小波能量指标结合六阶模态频率变化率指标能够对结构的故障状态进行识别, 而不影响辨识精度。这也从一个侧面说明小波包能量变化率指标可以更灵敏地指示结构局部故障情况。以约简后的属性集为输入, 重新构造一个 17-600-6 的网络结构来进行分类计算。实际测量过程中, 传感器数据经常会受到噪声的干扰。因此, 分别利用式(16)向特征向量数据中添加 0.06~0.1 的噪声, 识别结果如表 2 和图 3 所示。

好, 噪声从 0.06—0.1 的识别结果分别为 99.7%, 98.3%, (下转第 456 页)

重新选择合适的资源。其他方法没有考虑这个问题。因此,该方法对资源的不确定性不加控制。本文方法使用实时监控方式,如果在规定次数 λ 内发送的申请无响应,则返回重新刷新列表,重新选择一个合适的资源执行任务。那么此时本方法对资源的不确定性进行控制,保证了任务总能找到一个合适的服务资源来完成计算。

结束语 本文针对动态资源的几个典型特性,设计并实现了一种基于动态服务资源选择的方法。该方法弥补了基于历史资源信息方法的不足,能够动态可控地选择一个合理服务资源来执行特定任务,加强了服务发现技术的可靠性。但是一些问题仍然存在,典型的情况是历史表的大小按照何种方式来排序才能在查找资源的时候达到最佳的效果;当历史表记录不能作为参考时,窗口方式的较优资源选择对资源选择有很大影响,那么窗口应该选择多少个合适,窗口的合适大小为多少最佳等问题还需要进一步研究。

参考文献

[1] Cheung W K, Liu Ji-ming, Tsang K H, et al. Dynamic Resource Selection For Service Composition in The Grid[C] // IEEE/WIC/ACM International Conference on Web Intelligence (WI'04). IEEE Press, 2004; 412-418

[2] Foste I, Kesselmn C, Nick J. Grid services for distributed system integration[J]. IEEE Press, 2002, 35(6): 37-46

[3] Maximilien E M, Singh M P. Multiagent System for Dynamic Web Services Selection[C] // Workshop on Service-Oriented Computing and Agent-Based Engineering (SOCABE 2005). Utrecht, The Netherlands, July 2005; 25-29

[4] Roman M, Hess C K, Cerqueira R. et al. Gaia: A Middleware In-

frastructure to Enable Actives Paces[J]. IEEE Pervasive Computing Magazine, 2002, 1(1)

[5] Zeng L Z, Benatallah B, Dumas M. Quality driven Web service composition. In: Proc. of the WWW 2003[C] // Budapest; ACM, 2003; 411-421

[6] Benatallah B, Dumas M, Sheng Q Z, et al. Declarative composition and peer-to-peer provisioning of dynamic Web services[C] // Proc. of the 18th International Conference on Data Engineering. San Jose; IEEE Computer Society, 2002; 297-308

[7] Casati F, Ilnicki S, Jin L J, et al. eFlow: A platform for developing and managing composition e-services[R]. HPL-2000-36. HP Laboratories Palo Alto, 2000

[8] Liu Y T, Anne H H, Zeng L Z. QoS computation and policing in dynamic Web service selection[C] // Proc. of the WWW 2004. New York; ACM, 2004; 66-73

[9] Zhao J F, Xie B, Zhang L, et al. A Web services composition method supporting domain feature[J]. Chinese Journal of Computers, 2005, 28(4): 731-738

[10] 许兰,朱巧明,李培峰,等.一种基于历史信息反馈的动态服务资源选择模型[J]. 计算机应用, 2007, 27(9): 2283-2286

[11] 刘书雷,刘云翔,张帆,等.一种服务聚合中 Qos 全局最优服务动态选择算法[J]. 软件学报, 2007, 18(3): 646-656

[12] 王勇,代桂平,侯亚荣.信任感知的组合服务动态选择方法[J]. 计算机学报, 2009, 32(8): 132-138

[13] 李清,李志蜀,朱明放,等.基于自适应算法的动态网络服务选择方法[J]. 计算机工程, 2007, 22(13): 23-25

[14] 丁永生. 智能计算(理论与应用)[M]. 北京: 科学出版社, 2004

(上接第 431 页)

97%, 95.8% 与 94.3%。并且在相同噪声水平下,约简属性集输入时, PNN 结构故障识别结果明显好于所有 40 个条件属性都输入的情况。

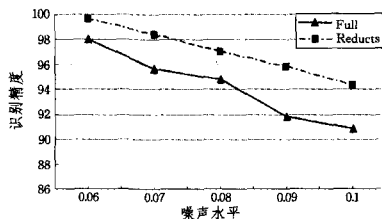


图3 不同噪声水平下粗糙集识别结果

结束语 本文利用粗糙集约简技术对特征属性进行约简,求得系统的最简特征集合,最后通过概率神经网络进行了结构故障的识别,通过仿真算例分析并对比属性约简前后直升机结构故障识别结果,可以得出下列结论:

- 1) 粗糙集属性约简可以显著减少 PNN 输入层神经元的数目,从而降低网络结构的复杂性,减少样本训练时间。
- 2) 经过粗糙集属性约简,可以获得比全部特征属性输入时更高的识别精度。
- 3) 粗糙集约简技术用于结构故障识别具有很好的抗噪声能力,即便噪声水平为 0.1 时,属性约减后的总识别精度仍为 94.3%,识别技术抗干扰能力增强。

参考文献

[1] Pawlak Z. Rough Sets, Decision Algorithms and Bayes's Theorem[J]. European Journal of Operational Research, 2002, 136: 181-189

[2] 薛定宇,陈阳泉. 高等应用数学问题的 MATLAB 求解[M]. 北京:清华大学出版社, 2008

[3] 姜绍飞,姚娟. 基于粗集和数据融合的结构故障识别方法[J]. 工程力学, 2009, 26(4): 207-213

[4] 韩晓红,胡玉. K-means 聚类算法的研究[J]. 太原理工大学学报, 2009, 40(3): 236-239

[5] Specht D F. Probabilistic neural networks [J]. International Journal of Neural Networks, 1990, 3: 109-118

[6] 贾清泉,杨奇逊,杨以涵. 基于故障测度概念与证据理论的配电网单相接地故障多判据融合[J]. 中国电机工程学报, 2003, 23(12): 6-11

[7] Hao L N, Xu X H. Applications of rough set theory in intelligent fault diagnosis[J]. China machine engineering, 2002, 13(21): 1856-1858

[8] 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, 2003

[9] 王前震,蔡瑞英,等. 概率粗糙集模型在机械故障诊断中的应用[J]. 计算机工程与应用, 2009, 23(4): 146-151

[10] 何明,冯博琴,马兆丰,等. 一种基于粗糙集的粗糙神经网络构造方法[J]. 西安交通大学学报, 2004, 18(12): 426-430