

文章编号:1004-9037(2012)01-0080-05

基于邻域粗糙集的航空发电机健康诊断方法

崔建国¹ 宋博翰¹ 董世良² 吕 瑞² 蒋丽英¹ 李忠海¹

(1. 沈阳航空航天大学自动化学院, 沈阳, 110136; 2. 沈阳飞机设计研究所, 沈阳, 110035)

摘要:提出一种基于邻域粗糙集和支持向量机相结合的航空发电机智能健康诊断方法。采用专业健康试验平台对某型战斗机的真实航空发电机进行试验,得到转速、负载、油压等大量表征发电机健康状态的监测数据。引入数据挖掘思想,采用邻域粗糙集理论对监测数据进行属性约简,将约简后的属性集输入给所设计的支持向量机健康诊断器,对航空发电机的健康状态进行了诊断研究。研究表明,该方法能够很好实现对某真实航空发电机的健康诊断,具有较高的推广应用价值。

关键词:邻域粗糙集;属性约简;支持向量机;航空发电机;健康诊断

中图分类号:TP206

文献标识码:A

Health Diagnosis of Aero-Generator Based on Neighborhood Rough Sets Theory

Cui Jianguo¹, Song Bohan¹, Dong Shiliang², Lü Rui², Jiang Liying¹, Li Zhonghai¹

(1. School of Automation, Shenyang Aerospace University, Shenyang, 110136;

2. Shenyang Aircraft Design & Research Institute, Shenyang, 110035)

Abstract: Health diagnosis of an aero-generator is important to flight safety. One of requirements is availability of extracting useful information from raw data. A health diagnosis method is presented based on neighborhood rough sets theory and support vector machine (SVM). Raw data are obtained from a specific aero-generator test platform. An approach of attribute reduction using neighborhood rough sets theory is outlined and a diagnosis classifier is designed based on SVM for further carrying out health diagnosis of aero-generator. The effectiveness of the proposed method is demonstrated through an experimental research. The result shows a better performance of the classifier that uses attribute reduction subsets as inputs, and also indicates that the method can have wide popularization and application potential.

Key words: neighborhood rough sets; attribute reduction; support vector machine; aero-generator; health diagnosis

引 言

航空发电机作为飞机供电系统的核心部件,其健康与否,直接影响着供电系统工作效能的发挥。因此,对其进行健康诊断技术研究,具有重要的军事意义与价值。目前,国内外学者对航空发电机健康诊断问题已进行了一些研究,采用的主要方法有数学建模法^[1]、频谱分析法^[2]、小波变换法^[3]、模

糊推理法^[4]以及神经网络法^[5]等。但随着航空发电机系统复杂度的不断提高,数学建模方法便受到一定限制。频谱分析和小波变换方法虽可对航空发电机进行诊断,但其主要是针对某类具体故障。神经网络方法在航空发电机诊断过程中具有自学习和较强的非线性拟合能力,但是航空发电机在实际中存在着大量的多特征、多过程和突发性故障,而神经网络方法不能解释其推理过程和依据。同时,军用飞机智能化发展不可避免地造成了监测数据量

基金项目:航空科学基金(2010ZD54012)资助项目;辽宁省教育厅科研基金(2008544)资助项目;国防基础科研计划(A0520110023)资助项目。

收稿日期:2011-03-22;修订日期:2011-06-22

的激增^[6],健康监测与诊断效率也随之降低,因此有必要引入数据挖掘思想,对原始数据信息进行精确约简,从海量数据中快速挖掘潜在知识,提取有用信息,剔除冗余数据,实现对航空发电机健康状态的快速有效诊断。

本文以某型军用航空发电机为具体研究对象,提出一种新的健康诊断方法。该方法首先利用邻域粗糙集属性约简算法对由多个传感器采集得到的状态信息进行约简,以降低信息维数,然后设计并采用支持向量机(Support vector machine, SVM)健康诊断器对约简后数据进行识别,从而可快速并准确的得到健康诊断结果。

1 基于邻域粗糙集的属性约简

粗糙集理论是一种建立在离散空间等价关系之上的专门研究不精确性和不确定性知识的数学工具。该理论的特性是不需要任何先验信息,仅使用数据本身的内部信息便能从中发现隐含知识,揭示潜在规律,对不完整不精确数据进行有效处理。然而,在实际应用中,对象的取值往往是实数中的连续值,若使用经典粗糙集理论,一般先将连续数据离散化,这样会导致原始信息的丢失^[7-8]。因此,为保证健康诊断的准确性和原始信息的完整性,本文提出首先运用邻域粗糙集理论^[9]进行属性约简,在此基础上对航空发电机健康信息样本空间进行邻域粒化,直接计算样本距离,确定样本之间相邻关系。

定义邻域决策系统 $NDT = \langle U, A, D \rangle$, 其中, A 是描述论域 U 的实数特征集合, 决策属性 D 将论域 $U = \{x_1, x_2, \dots, x_n\}$ 划分为 N 个等价类 $X_j (j=1, 2, \dots, N)$ 。 $\forall a \subseteq A$, 定义决策 D 关于属性 a 的上近似和下近似为

$$\overline{\text{apr}}_a(D) = \bigcup_{j=1}^N \{x_i | \delta_a(x_i) \cap X_j \neq \emptyset, x_i \in U\} \quad (1)$$

$$\underline{\text{apr}}_a(D) = \bigcup_{j=1}^N \{x_i | \delta_a(x_i) \subseteq X_j, x_i \in U\} \quad (2)$$

式中: $\delta_a(x)$ 是依据属性 a 生成的邻域粒子。决策 D 的下近似(决策正域)越大,说明依据属性分类效果越强,各类重叠程度越小。因为正域的大小是相对论域整体计算的,所以定义决策属性对条件属性 a 的依赖性为

$$\gamma_a(D) = \text{Card}(\underline{\text{apr}}_a(D)) / \text{Card}(U) \quad (3)$$

式(3)表示在样本集中依据条件属性 a 的描述,能够被某一决策包含的样本数占全体样本数的

比率,即决策 D 的下近似集中元素个数所占论域中元素总数的比率。有了属性的依赖性函数,接下来可以给定一个邻域系统 $NDT = \langle U, A, D \rangle, B \subseteq A, \forall a \in A - B$, 定义 a 相对于 B 的重要度为

$$SIG(a, B, D) = \gamma_{B \cup \{a\}}(D) - \gamma_B(D) \quad (4)$$

根据属性重要度指标,可以构造贪心式属性约简算法。初始化属性约简子集为空集,循环计算剩余属性的重要度,选择重要度最大的属性加入约简子集当中,直至所有剩余属性重要度为0,约简算法终止。算法的输入为 $NDT = \langle U, A, D \rangle$, 输出为属性约简子集 B , 步骤如下:

(1) 初始化属性约简子集 $B = \emptyset$;

(2) 对于任意剩余属性 $a_i \in A - B$, 计算该属性相对于约简子集 B 的重要度

$$SIG(a_i, B, D) = \gamma_{B \cup \{a_i\}}(D) - \gamma_B(D)$$

(3) 选择重要度最大的属性 $SIG(a_k, B, D) = \max_i(SIG(a_i, B, D))$, 如果 $SIG(a_k, B, D) > 0$, 则将属性 a_k 加入属性约简子集 $B, B = B \cup \{a_k\}$, 并跳至第(2)步继续计算,否则程序终止,输出最终计算结果。

2 支持向量机

SVM 是在统计学基础上产生的一种新型的机器学习机器,它遵循结构风险最小化原则,能够很好解决高维问题和局部极值问题,泛化能力强^[10],同时还具有结构简单、训练速度快、结果精度高等众多优异特性^[11],特别适用于航空发电机健康诊断问题。

设有两类样本集 $(x_i, x_j), x_i \in \mathbf{R}^d, y_i \in \{+1, -1\}, i=1, 2, \dots, N$, 其中 y 为样本类别标号, d 为样本空间维数, N 为样本总数。通过引入拉格朗日函数等一系列方法,将 SVM 最小风险界问题转化为在条件 $\alpha_i \geq 0$ 和 $\sum_{i=1}^n y_i \alpha_i = 0$ 的约束下,对拉格朗日系数 α_i 求解函数最大值的问题

$$\max Q(\alpha, K(x_i, x_j)) = -\frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) + \sum_{i=1}^n \alpha_i \quad (5)$$

对应判别函数式为

$$f(x) = \text{sgn} \left[\sum_{i=1}^n \alpha_i y_i K(x_i, x_j) + b^* \right] \quad (6)$$

在以上式中, sgn 为符号函数, n 为支持向量数目, b^* 为分类阈值, $K(x_i, x_j)$ 为满足 Mercer 条件的核函数。因此,只要设计适当的核函数 $K(x_i, x_j)$, 便可由判别函数 $f(x)$ 进行诊断与识别。

3 航空发电机健康监测数据预处理

针对某型真实航空发电机,采用其健康监测专用试验平台对其进行长期大量试验,得到转速、负载、电流、电压等大量表征发电机健康状态的监测数据。选取含有不同健康状态的120组原始数据作为研究样本,这些样本共包括11个条件属性参数与3种故障状态类别标号。部分原始数据如表1所示。

表1 某型航空发电机健康状态部分表征参数监测数据

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	D
8 702	26.94	401.4	79.68	114.4	116	130	0.24	0.42	1.8	47.7	1
8 701	30.54	401.1	90.3	114.1	133	151	0.22	0.38	1.73	45.9	1
8 693	35.82	401.1	105.7	114.1	123	137	0.24	0.41	1.77	48	1
8 698	53.22	395.6	157.6	113.9	132	153	0.23	0.37	1.12	46.2	2
8 701	54.06	397	157	114	146	165	0.25	0.42	0.88	49.3	3
8 698	26.82	401.4	79.02	114.3	120	138	0.23	0.4	1.93	47.3	1
8 689	53.46	400	157.5	114.1	120	144	0.26	0.4	1.05	48	2

在表1中,A1~A11为样本监测数据的11个属性参数,分别为:输入转速、负载、频率、电流、电压、进油温度、出油温度、进口压力、出口压力、注油压力、回油流量。将这11个参数作为邻域粗糙集约简算法的条件属性输入。表中最后一列D是对各样本健康状态类别的标号,将该列作为邻域粗糙集约简算法的决策属性输入。

由于数据各属性参数数量纲不同,数量级上存在差异,而在邻域粗糙集属性约简算法中需要对属性参数统一设置邻域直径,因此,为得出更加准确的属性约简结果,在进行邻域粗糙集属性约简之前,有必要对各属性参数进行归一化处理。采用最大最小值法,按如式(7)所示映射,将数据归一化至[0,1]区间。式中 x_{\max} , x_{\min} 分别表示某项属性参数中数据的最大与最小值。

$$f: x \rightarrow y = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (7)$$

经过归一化处理后,监测数据值全部位于[0,1]区间内,此时将120个样本数据的条件属性和决策属性输入属性约简算法。邻域直径以步长为0.01从0到1进行取值。将属性约简结果输入SVM,对数据进行多次分类试验,对比分类结果发现当邻域直径取值为0.05时,试验效果最佳,此时属性约简算法输出结果为属性约简子集 $B = \{A1, A2, A5, A10\}$,即约简得到输入转速、负载、电压和注油压力共4个属性参数。依据粗糙集属性约简后不降低系统分类性能的原则,可以推测,仅需此4个属性参数便可对发电机健康状态进行有效识别与诊断。

4 航空发电机健康诊断试验

4.1 支持向量机健康诊断器的构建

研究表明,SVM常采用的核函数有多项式核函数、Sigmoid核函数和径向基核函数^[12]。为使设计的SVM健康诊断器得到的诊断效果最佳,本文分别对此三种核函数进行了试验研究,结果表明,采用径向基核函数取得的效果最佳,运用该核函数设计健康诊断器的主要过程如下:

(1)确定SVM核函数。选用高斯径向基核函数

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (8)$$

(2)确定SVM相关参数。对式(8)中的核函数宽度 σ 和SVM惩罚因子 C 进行优化。本文采用交叉验证方法对参数进行寻优,以交叉验证准确率作为性能指标。当同时出现多组最佳参数时,选择惩罚因子 C 最小的一组,以避免出现SVM过学习状态。

(3)创建SVM健康诊断器。依据以上参数构建SVM模型,设计健康诊断算法,创建SVM健康诊断器,实现其健康诊断效能。

4.2 健康诊断试验

将120个样本分成两组,每组60个样本,分别用于SVM健康诊断器的训练和测试。样本划分如表2所示。

表2 样本划分信息

数据样本	状态类别		
	状态1	状态2	状态3
训练样本数	24	24	12
测试样本数	24	24	12
各类样本总数	48	48	24

状态1表示航空发电机处于正常工作状态,状态2表示航空发电机接近非正常工作状态,主要表现为注油压力不足,但此时发电机尚能基本满足供电需求,状态3表示航空发电机处于故障状态,不能正常工作。

为对比属性约简前后健康诊断效果,首先将训练样本的全部条件属性A1~A11作为SVM的输入,将样本对应的决策属性D作为期望输出,对SVM健康诊断器进行训练。训练过程中对参数 σ , C 进行寻优,可得到18组在交叉验证准确率为95.08%时的最佳参数 (σ, C) 。为避免SVM出现过学习状态,选取惩罚因子最小的一组参数 $\sigma = 0.25$,

$C=4$ 作为最终的 SVM 模型参数。利用训练好的 SVM 健康诊断器对 60 个测试样本(每个样本包括全部 11 个属性参数)进行健康诊断,将诊断器诊断结果与原始样本进行对比,可得 SVM 诊断正确率为 93.33%。

依据以上属性约简结果,将训练样本的属性 A_1, A_2, A_5, A_{10} 作为输入,决策属性 D 作为期望输出,训练新的 SVM 健康诊断器。训练过程中对参数 σ, C 进行寻优,可得到 65 组在交叉验证准确率为 100% 时的最佳参数 (σ, C) , 结果如图 1 所示。选取惩罚因子最小的一组参数 $\sigma=16, C=2$ 作为属性约简后最终的 SVM 模型参数。利用属性约简子集训练得到的 SVM 健康诊断器对属性约简后的 60 个测试样本进行健康诊断试验,将诊断结果与原始样本实际状态进行对比,如图 2 所示。

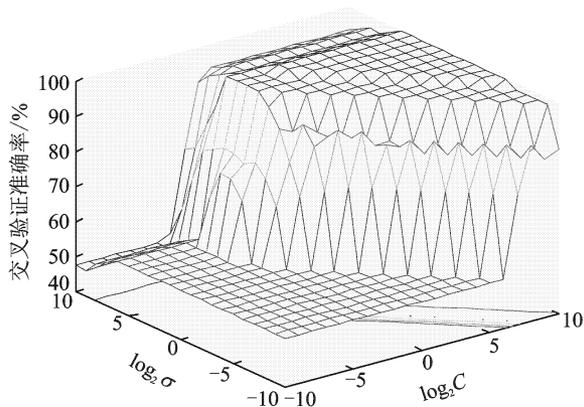


图1 属性约简子集输入时参数寻优结果

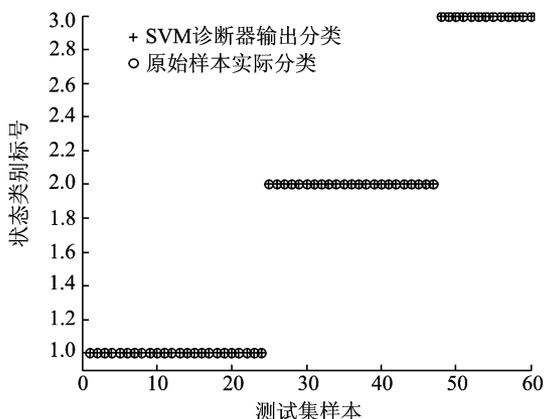


图2 属性约简子集输入时的分类效果图

由图 2 可知,采用属性约简子集作为 SVM 诊断器的输入,其诊断正确率为 100%。与全属性输入时的诊断效果相比,其诊断正确率与其他相关试验信息见表 3。

表 3 试验结果信息

数据源	未约简数据	约简后数据
输入属性	A_1, A_2, \dots, A_{11} (全属性)	A_1, A_2, A_5, A_{10}
决策属性	D	D
训练样本数	60	60
模型训练时间/s	0.001 041	0.000 870
测试样本数	60	60
诊断所需时间/s	0.006 740	0.001 097
诊断正确率	93.33%	100%

试验表明,进行属性约简后,利用 4 个属性参数即可有效对航空发电机健康状态进行诊断,在不降低诊断正确率的同时, SVM 输入维数得到大幅度降低,诊断所需数据量明显减少, SVM 健康诊断器训练和诊断所需时间均显著缩短,提高了健康诊断的时效性。

5 结束语

本文提出一种基于邻域粗糙集和 SVM 的航空发电机健康诊断方法。通过引入数据挖掘思想,采用邻域粗糙集策略对大量原始数据进行属性约简,得到属性约简子集,以此属性约简子集作为 SVM 诊断器的输入,对某真实航空发电机健康状态进行诊断研究,结果表明,其诊断正确率为 100%,且提高了健康状态诊断的时效性。此外,由于采用了邻域粗糙集理论,无需先验知识和数据离散化,约简了信息属性,减少输入数据量,因此该方法对于具有多属性数据的复杂系统进行健康诊断有很好的适用性,并具有很强的泛化能力,在航空航天等领域具有很好的应用前景。

参考文献:

- [1] 刘晓琳. 飞机电源系统的故障诊断数学模型[J]. 科技创新导报, 2008(4): 191.
Liu Xiaolin. Mathematical model of fault diagnosis technique of airplane electrical power system [J]. Science and Technology Innovation Herald, 2008, 4 (4): 191.
- [2] 刘向群, 仇越, 张洪钺. 基于频谱法与神经网络的航空起动发电机的故障检测与诊断[J]. 航空学报, 2005, 25(2): 158-161.
Liu Xiangqun, Qiu Yue, Zhang Hongyu. Fault detection and diagnosis of aero-starter-generator based on spectrum analysis and neural network method [J]. Acta Aeronautica et Astronautica Sinica, 2004, 25(2): 158-161.

- [3] 杨传道, 马建卫, 韩建定. 小波变换在航空发电机异常检测中的应用研究[J]. 电气应用, 2006, 27(12): 105-108.
Yang Chuandao, Ma Jianwei, Han Jianding. Research on application of wavelet transform to anomaly detection of aviation generator [J]. Electrical Application, 2006, 27(12): 105-108.
- [4] Ping Xu, Qi Shuangma. The fault diagnosis of aircraft generator using fuzzy petri nets[C]//2nd Asia-Pacific Power and Energy Engineering Conference. Chengdu: IEEE Power and Energy Society Press, 2010: 1-4.
- [5] 曹振锋, 高喜银, 程光明. 基于模糊推理的陆航飞机机电故障诊断专家系统设计研究[J]. 中国科技信息, 2006(12): 111-112.
Cao Zhenfeng, Gao Xiyin, Cheng Guangming. Design and research by fuzzy inference of LUHANG helicopter machine fault diagnosis expert system [J]. China Science and Technology Information, 2006 (12): 111-112.
- [6] 刘同明. 数据挖掘技术及其应用[M]. 北京: 国防工业出版社, 2001: 55-100.
Liu Tongming. Data mining techniques and its application [M]. Beijing: National Defence Industry Press, 2001: 55-100.
- [7] Lech P. Data-mining and knowledge discovery: case-based reasoning, nearest neighbor and rough sets, encyclopedia of complexity and systems science [M]. Germany: Springer Reference, 2009: 1789-1810.
- [8] Choo Y, Bakar A A, Muda A K. Capturing uncertainty in associative classification model [C]//2nd Conference on Data Mining and Optimization. Kajakang: Data Mining and Optimization Group, 2009: 84-89.
- [9] 胡清华, 于达仁, 谢宗霞. 基于邻域粒化和粗糙集逼近的数值属性约简[J]. 软件学报, 2008, 19(3): 640-649.
Hu Qinghua, Yu Daren, Xie Zongxia. Numerical attribute reduction based on neighborhood granulation and rough approximation [J]. Journal of Software, 2008, 19(3): 640-649.
- [10] 张国新, 汤青波, 许德昌. 基于支持向量机的液压泵健康诊断[J]. 煤矿机械, 2007, 28(8): 193-195.
Zhang Guoxin, Tang Qingbo, Xu Dechang. Hydraulic pump health diagnosis based on support vector machine [J]. Coal Mine Machinery, 2007, 28(8): 193-195.
- [11] 边肇祺, 张学工. 模式识别[M]. 北京: 清华大学出版社, 2000: 296-302.
Bian Zhaoqi, Zhang Xuegong. Pattern recognition [M]. Beijing: Tsinghua University Press, 2000: 296-302.
- [12] 刘祥楼, 贾东旭, 李辉姜, 等. 说话人识别中支持向量机核函数参数优化研究[J]. 科学技术与工程, 2010, 7(10): 1670-1673.
Liu Xianglou, Jia Dongxu, Li Huijiang, et al. Research on kernel parameter optimization of support vector machine in speaker recognition [J]. Science Technology and Engineering, 2010, 7(10): 1670-1673.

作者简介: 崔建国(1963-), 男, 博士后, 教授, 研究方向: 飞行器健康诊断、预测与综合健康管理、信号检测与控制、可视化仿真技术与应用, E-mail: gordon.cjg@163.com; 宋博翰(1987-), 男, 硕士研究生, 研究方向: 飞行器健康诊断、预测与综合健康管理; 董世良(1963-), 男, 研究员, 研究方向: 飞机健康诊断; 吕瑞(1968-), 男, 研究员, 研究方向: 飞机综合后勤保障; 蒋丽英(1976-), 女, 副教授, 研究方向: 飞机寿命预测; 李忠海(1962-), 男, 教授, 研究方向: 图像处理与飞机综合健康管理。