

基于 PCA 和 SVM 的铬层裂纹声发射识别方法

李 力, 李 骥

(三峡大学 水电机械设备设计与维护湖北省重点实验室, 宜昌 443002)

摘 要: 针对采用幅值、能量等统计特征参数分类识别声发射(AE)信号时存在的信息冗余问题, 提出利用主成分分析(PCA)方法减少信息冗余, 提取 AE 信号统计特征。设计了钢板表面铬层裂纹试验, 对统计特征参数进行主成分分析, 提取了两个主成分。设计了支持向量机(SVM)分类器, 以主成分为输入向量, 分类识别铬层裂纹 AE 信号。验证了主成分可以有效表征 AE 信号统计特征, 减少了信息冗余, 提高了分类效率及准确率。

关键词: 声发射; 统计特征参数; 主成分分析(PCA); 支持向量机(SVM); 识别

中图分类号: TG115. 28; TB559

文献标志码: A

文章编号: 1000-6656(2014)02-0001-04

Cracking Acoustic Emission Identification Method based on PCA and SVM

LI Li, LI Ji

(Hubei Key Laboratory of Hydroelectric Machinery Design & Maintenance,
China Three Gorges University, Yichang 443002, China)

Abstract: Information redundancy is a big problem in acoustic emission(AE) signal identification based on statistical feature parameters such as amplitude, energy counts, etc. Here, principle component analysis(PCA) was employed to reduce information redundancy and extract statistical feature of AE signals. The AE data were collected in the AE test for Cr-coating cracking on the surface of a steel plate, AE statistical feature parameters were analyzed using PCA, and two principle components were extracted. The principle components were employed as the input vector of a SVM classifier, and the AE signals caused by Cr-coating cracking were identified. It demonstrated that principle components could represent statistical feature of AE signals, reduce information redundancy, and effectively raise identification efficiency and accuracy.

Keywords: Acoustic emission; Statistical feature parameters; PCA; SVM; Identification

铬由于具有熔点高、耐腐蚀等优点而被广泛用于刀具、枪管等机械零件基体涂层防护。但铬层质地硬脆, 易开裂, 从而导致铬层与基体分离、剥落, 严重影响其防护性能, 所以检测铬层裂纹十分必要^[1]。铬层开裂时会释放应变能并发出弹性波, 即出现声发射(Acoustic Emission, AE)现象。因而, 可以利用 AE 技术检测铬层裂纹。但 AE 检测中通常存在大量的干扰信号, 需要对 AE 信号分类, 识别出铬层裂纹 AE 信号。而在分类识别中, 提取 AE 信号特征参数是关键^[2]。目前, 最常用的为 AE 信号统

计特征参数^[3], 主要包括上升时间、持续时间、幅值、能量等 9 个参数。然而, 基于统计特征参数的 AE 信号分类识别存在如下问题: ① 统计特征参数间存在相关性, 信息冗余较多。② 部分参数的分类能力较差, 影响分类效率。③ 参数选择多依赖经验^[4-5], 且只采用部分参数, 可能丢失许多有用信息。

针对上述问题, 笔者对 9 个统计特征参数进行主成分分析(Principle Component Analysis, PCA), 提取表征质量较高, 分类能力强的特征参数。为验证方法的可行性, 设计了钢板表面铬层裂纹试验, 采集试验中的 AE 信号。采用 PCA 提取了两个主成分特征参数作为支持向量机(Support Vector Machine, SVM)分类器输入向量, 对铬层裂

收稿日期: 2013-04-01

基金项目: 国家自然科学基金资助项目(51175401)

作者简介: 李 力(1964—), 女, 教授, 主要从事无损检测研究。

纹和干扰 AE 信号分类,证明了主成分特征参数可以有效表征 AE 信号,减少统计特征参数的信息冗余,提高分类效率和准确率。

1 PCA 与 SVM 原理

1.1 PCA 原理

主成分分析 PCA 主要用于提取数据关键信息,减少信息冗余^[6],其基本原理如下:设 p 个原始变量 x_1, x_2, \dots, x_p ,通过式(1)所示的线性变换将原始变量转化为 p 个主成分 z_1, z_2, \dots, z_p ,具体变换步骤参考文献^[6]。

$$\begin{cases} z_1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1p}x_p \\ z_2 = a_{21}x_1 + a_{22}x_2 + \dots + a_{2p}x_p \\ \vdots \\ z_p = a_{p1}x_1 + a_{p2}x_2 + \dots + a_{pp}x_p \end{cases} \quad (1)$$

式中: z_i 为第 i 主成分; a_{ij} 为主成分系数; $i, j=1, 2, \dots, p$ 。主成分满足以下条件:

(1) z_i, z_j 互不相关 ($i \neq j, i, j=1, 2, \dots, p$);

(2) 主成分方差 $\text{Var}(z_1) > \text{Var}(z_2) > \dots > \text{Var}(z_p)$

其中:

$$\begin{aligned} \text{Var}(z) &= \frac{1}{p-1} \sum_{i=1}^p (z_i - \bar{z})^2 z_i \\ \bar{z} &= \frac{1}{p} \sum_{i=1}^p z_i \end{aligned} \quad (2)$$

由式(1)可知,主成分包含了原始变量的所有信息。由条件(1)知各主成分包含的信息相互独立。由条件(2)知各主成分包含的信息量互不相等。为了定量描述各主成分包含的信息,常采用以下统计量^[6],即特征值 λ_i 、方差贡献率 C_{r_i} 和累计方差贡献率 AC_{r_k} 。

$$\lambda_i = \text{Var}(z_i)$$

$$C_{r_i} = \frac{\lambda_i}{\sum_{i=1}^p \lambda_i}$$

$$AC_{r_k} = \sum_{i=1}^k C_{r_i} \quad i, k = 1, 2, \dots, p \quad (3)$$

其中,特征值 λ_i 越大,则该主成分影响力越大,该主成分对数据信息的解释能力越强;方差贡献率 C_{r_i} 值越大,则该主成分所包含的信息量越大;如果累积贡献率 AC_{r_k} 达到 85% 以上,则可保证选取的 k 主成分能包括数据的绝大部分信息。

1.2 SVM 原理

SVM 是从线性可分情况下的最优分类面发展

而来的,基本思想是^[7]:设在给定的内积空间 H 中存在一个超平面 $\{(w \cdot x) + b = 0 | x \in H, w \in H, b \in R\}$ 将样本集 T 分离。其中, w 是一个垂直于超平面的向量,设 w 为单位长度,则 $(w \cdot x)$ 是向量 x 沿 w 方向的长度。通过内积(核)函数 ϕ 将输入空间变换到一个高维特征空间中,然后在特征空间中寻找最优分类超平面(线)将不同类别分开,可用图 1 所示的二维平面来说明。图 1 中小方框和圆圈代表两类样本,特征空间中间的实线为分类线,其附近的两条虚线分别为过各类中离分类线最近的样本且平行于分类线的直线,它们之间的距离就是分类间隔。所谓最优分类线就是要求分类线不但能将两类正确分开,即训练错误率为零,而且使分类间隔最大。

对分类线 $(w \cdot x) + b = 0$ 进行标准化处理,使得分类间隔等于 $2/\|w\|$,使间隔最大等价于使 $\|w\|/2$ 最小。训练样本正确可分,且使 $\|w\|/2$ 最小的分类面就是最优分类面,位于两虚线上的训练样本点就称作支持向量。构建最优超平面的问题可以转化为最优化(二次规划)问题: $\min(\|w\|^2/2)$ 。

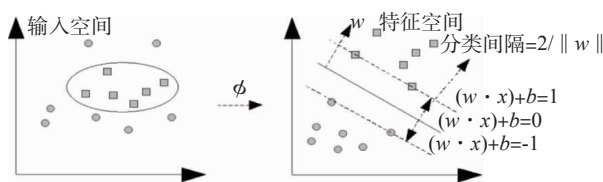


图 1 SVM 原理示意

2 试验与 AE 信号采集

试验试样为表面附铬层的钢板(图 2)。钢板厚 1 mm,其表面通过电解法附着一层 0.4 mm 厚的铬层。采用声发射试验台对钢板加载。试验台如图 2 所示,主要由支架、千斤顶和压力传感器组成。采用千斤顶向钢板施加向上的载荷,让钢板弯曲,即可使钢板表面的铬层因弯曲而开裂。



图 2 AE 信号采集试验系统

采用美国 PAC 公司的 PCI-2 声发射信号采集系统采集试验数据。试验传感器布置如图 2 所示,其中 1,2 位置为宽频带 WD 传感器,3,4 位置为谐振式 R15a 传感器。前置放大器增益 40 dB,采样频率 1 MHz,门槛值为 40 dB,带通滤波 100 ~ 400 kHz。峰值鉴别时间 PDT=200 ms,撞击鉴别时间 HDT=600 ms,撞击闭锁时间 HLT=800 ms。按照图 3 所示循环施加载荷,使用声发射信号采集系统采集 AE 信号,整个试验过程 AE 能量的时间历程如图 4 所示。试验过程中,在 $t=448$ s 时出现了明显的铬层裂纹,如图 5 所示,由此可以确定此刻的 AE 信号由铬层裂纹产生。在 $t=472$ s 时已经停止对钢板加载,但图 4 显示 $t=472$ s 以后仍有 AE 信号,由此可确定该 AE 信号完全由干扰引起。

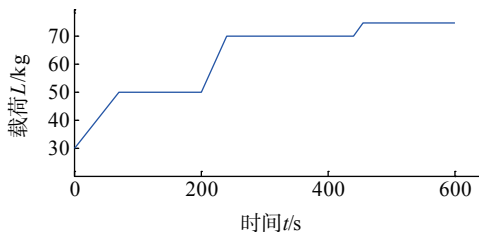


图3 加载循环

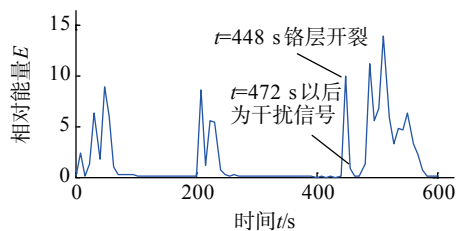


图4 AE能量-时间历程

3 分类识别应用

选取 $t=448$ s 时铬层裂纹 AE 信号 100 个, $t=$

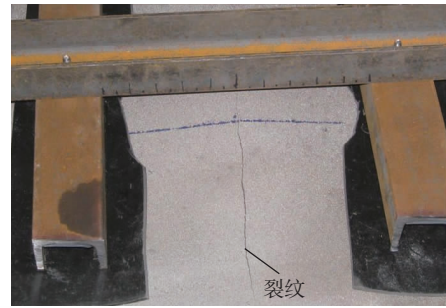


图5 铬层裂纹

472 s 时刻以后干扰 AE 信号 100 个。设定 AE 信号的统计特征参数为上升时间 x_1 、振铃计数 x_2 、能量 x_3 、持续时间 x_4 、幅值 x_5 、平均频率 x_6 、RMS x_7 、ASL x_8 、峰值频率 x_9 , 令 $X=(x_1, x_2, \dots, x_9)$ 。

3.1 AE 统计参数的相关性分析

相关分析主要是揭示变量之间的线性相关的密切程度^[6]。相关系数 r_p 定义如式(4)所示,它反映变量 x_i 和 x_j 间的相关关系。

$$r_p = \frac{\text{cov}(x_i, x_j)}{\sigma_{x_i} \sigma_{x_j}} \quad (4)$$

式中: $\text{cov}(x_i, x_j)$ 为变量 x_i 和 x_j 的协方差; σ_{x_i} 和 σ_{x_j} 分别为变量 x_i 和 x_j 的标准方差。

表 1 为 200 个 AE 信号的统计特征参数间的相关系数。由表 1 知,部分统计特征参数间存在较强的相关性,如上升时间 x_1 与持续时间 x_4 的相关系数达到 0.908,上升时间 x_1 与峰值频率 x_9 的相关系数达到 0.954,振铃计数 x_2 与持续时间 x_4 的相关系数达到 0.916,这表明统计特征参数间存在着大量的信息冗余。

3.2 基于 PCA 的特征提取

采用 PCA 对 200 个 AE 信号的 9 个统计特征参数进行分析得 9 个主成分。各主成分的特征值、方差贡献率和累计方差贡献率如表 2 所示。

表1 AE信号间的相关系数

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
x_1	1.000	0.809	0.540	0.908	0.497	-0.045	0.040	0.067	0.954
x_2	0.809	1.000	0.695	0.916	0.843	-0.054	-0.131	-0.126	0.908
x_3	0.540	0.695	1.000	0.746	0.766	-0.044	-0.003	0.032	0.552
x_4	0.908	0.916	0.746	1.000	0.772	-0.086	-0.014	0.019	0.901
x_5	0.497	0.843	0.766	0.772	1.000	-0.109	-0.131	-0.127	0.600
x_6	-0.045	-0.054	-0.044	-0.086	-0.109	1.000	-0.036	-0.058	-0.038
x_7	0.040	-0.131	-0.003	-0.014	-0.131	-0.036	1.000	0.941	-0.039
x_8	0.067	-0.126	0.032	0.019	-0.127	-0.058	0.941	1.000	-0.019
x_9	0.954	0.908	0.552	0.901	0.600	-0.038	-0.039	-0.019	1.000

表 2 AE 统计参数的主成分方差贡献率

主成分	特征值 λ_i	方差贡献率 $C_{r_i}/\%$	累计方差贡献 率 $AC_{r_k}/\%$
z_1	4.840	53.778	53.778
z_2	1.977	21.964	75.742
z_3	1.002	11.132	86.873
z_4	0.752	8.357	95.230
z_5	0.266	2.957	98.187
z_6	0.078	0.864	99.051
z_7	0.056	0.621	99.672
z_8	0.022	0.245	99.917
z_9	0.007	0.083	100.000

由表 2 可知, $\lambda_1, \lambda_2, \lambda_3$ 大于 1, 即主成分 z_1, z_2, z_3 对数据信息的解释能力较强; 主成分 z_1, z_2, z_3 的方差贡献率大于 10%, 表明它们包含的数据信息较大; 主成分 z_1, z_2, z_3 的累计方差贡献率为 86.873%, 表明采用 z_1, z_2, z_3 可以包含数据的绝大部分信息。而 z_1, z_2 的累计方差贡献率已达到 75.742%, 为了尽量减少数据量, 提高分类效率, 这里只采用 z_1, z_2 作为 AE 信号的特征参数, 其中:

$$\begin{aligned} z_1 = & 0.881x_1 + 0.967x_2 + 0.787x_3 + 0.977x_4 \\ & + 0.830x_5 - 0.086x_6 - 0.078x_7 - 0.055x_8 \\ & + 0.921x_9 \\ z_2 = & 0.141x_1 - 0.075x_2 \\ & + 0.057x_3 + 0.075x_4 - 0.109x_5 - 0.096x_6 \\ & + 0.977x_7 + 0.982x_8 + 0.047x_9 \end{aligned} \quad (5)$$

3.3 基于 SVM 的分类识别

以径向基函数为核函数构造 SVM 分类器, 分别以统计特征参数 $X = (x_1, x_2, \dots, x_9)$ 和 z_1, z_2 作为 SVM 分类器输入向量, 采用粒子群优化算法优化核函数参数^[8], 从铬层裂纹和干扰 AE 信号两类数据中各选取 50 组数据训练, 剩余 100 组数据预测, 对铬层裂纹 AE 信号和干扰 AE 信号分类。分类准确率及分类时间如表 3 所示。由表 3 可知, 以

表 3 SVM 分类准确率及分类时间

输入向量	准确率/%	分类时间/s
x_1, x_2, \dots, x_9	88	43.495
z_1, z_2	96	28.595

主成分 z_1, z_2 为输入向量时, 分类准确率 96%, 与统计特征参数 X 相比, 分类准确率提高了 8%, 参数个数减少了 7 个, 且分类时间减少了 14.9 s。

4 结语

AE 信号的统计参数相关性研究表明, 部分参数间存在较强的相关性。部分参数的相关性在 0.90 以上, 表明统计参数间存在大量信息冗余。

提取了 2 个主成分 z_1, z_2 , 对铬层裂纹和干扰 AE 信号分类, 分类准确率为 96%。与统计特征参数相比, 有效提高了分类准确率, 大大减少了计算数据量, 减少了分类时间, 提高了分类效率。

参考文献:

- [1] 杨班权. 涂层断裂韧性的声发射辅助拉伸测量方法[J]. 兵工学报, 2008, 29(4): 420-424.
- [2] 金文, 陈长征, 金志浩, 等. 声发射源识别中的三比值特征提取方法研究[J]. 仪器仪表学报, 2008, 29(3): 530-534.
- [3] 杨明伟, 耿荣生. 声发射检测[M]. 北京: 机械工业出版社, 2005.
- [4] 邵永波, 裴珍, 朱祖铭, 等. 基于声发射技术的钢丝绳断丝模式识别[J]. 应用声学, 1999, 18(1): 1-5.
- [5] 钟建强, 柳颖, 杨娟, 等. 利用声发射技术检测储罐的腐蚀损伤状态[J]. 无损检测, 2011, 33(11): 24-28.
- [6] 巴克豪斯. 多元统计方法[M]. 上海: 格致出版社, 2009.
- [7] LI Li, LI Ji, CHEN Bao-jia. Wavelet packet and support vector machine for engine fault diagnosis[J]. Advanced Materials Research, 2011(230-232): 1-6.
- [8] 王东, 吴湘滨. 利用粒子群算法优化 SVM 分类器的超参数[J]. 计算机应用, 2008, 28(1): 134-139.

摘 要 撰 写

摘要是以提供文献内容梗概为目的, 不加评论和补充解释, 简明、确切地论述文献重要内容的短文, 因此, 必须符合拥有与论文同等的主要信息的原则。摘要可包括研究工作的主要对象和范围, 以及具有情报价

值的其它重要信息, 内容不应空洞、过简, 应具备四个要素, 即研究目的、方法、结果和结论, 其中, 后两个是最重要的。中文摘要一般为 200 字左右, 如需要可以略多; 英文摘要内容应与中文摘要一致。