

# 基于自适应 SVM 决策树的焊缝缺陷类型识别

李 坤<sup>1</sup>,文 斌<sup>2</sup>,任清安<sup>3</sup>,罗爱民<sup>4</sup>

(1. 四川大学电子信息学院,成都 610064;2. 成都信息工程学院,成都 610225;  
3. 四川大学数学学院,成都 610065;4. 四川大学轻纺与食品学院,成都 610065)

**摘要:**针对传统 X 射线焊缝缺陷检测方法普遍存在分类识别精度不高的问题,提出了一种基于分离程度的自适应 SVM 决策树算法。首先对滤波后的 X-Ray 焊缝缺陷图像进行数学形态学重建,然后根据分离程度,每次将分离程度最大的缺陷类别首先分离出来,构造自适应二叉树的 SVM 分类器,从而达到了减小二叉树的累积误差,得到了分类性能优良的 SVM 决策树,并用其对 X-Ray 焊缝缺陷图像进行分类识别。实验结果表明,该算法取得了好的分类精度和识别效果。

**关键词:**决策二叉树;支持向量机;分离程度;数学形态学;缺陷识别

中图分类号: TG115.28 文献标志码:A 文章编号:1000-6656(2010)03-0171-04

## Welding Defects Classification Based on Adaptive SVM Decision Tree

LI Kun<sup>1</sup>, WEN Bin<sup>2</sup>, REN Qing-An<sup>3</sup>, LUO Ai-Min<sup>4</sup>

(1. School of Electronics and Information Engineering, Sichuan University, Chengdu 610064, China; 2. Chengdu University of Information Technology, Chengdu 610225, China; 3. School of Mathematics, Sichuan University, Chengdu 610065, China; 4. School of Light Industry and Food, Sichuan University, Chengdu 610065, China)

**Abstract:** An adaptive SVM(Support Vector Machines) based on binary tree using the degree of separation is proposed in this paper, aiming at the problem that it's difficult for traditional detection methods to accurately identify the welding defects of X-Ray images. Firstly, mathematical morphological reconstruction is applied to the filtered X-Ray images of welding defects. It is proposed to separate category of defects with the largest degree of separation as a priority, and to construct adaptive SVM classifiers based on binary tree, thus decreasing the accumulated error. Finally, a SVM decision tree of good classification performance can be obtained, which is used to classify and identify the X-Ray images of welding defects, and it shows that the algorithm has made a good classification and recognition accuracy results.

**Keywords:** Binary decision tree; Support vector machine; Degree of separation; Mathematical morphology; Welding defects classification

焊缝缺陷是造成锅炉、压力容器失效和事故的主要原因。长期对锅炉 X-ray 焊缝检测,发现气孔、夹钨、未焊透、夹渣,裂纹或未熔合是其主要的缺陷<sup>[1]</sup>,这类缺陷严重危害锅炉的安全运行。每种缺陷都有各自的产生原因、主要形态和影像特征,准确

识别焊缝缺陷类别,对于改进焊接工艺,提高焊接的质量和电站安全保障具有一定的意义。

国内在对焊缝 X 射线底片的自动评判进行了一些深入研究<sup>[2-4]</sup>,国外 Nacereddine, Nafa 和 Drai Redouane 等人利用根据边缘的几何不变矩,利用 BP 神经网络技术实现了焊缝缺陷的分类<sup>[5]</sup>,俄罗斯飞行器焊接及表面涂层技术研究所的 A E Kapustin 和 I I Bardusova 等人也研究了焊缝底片的计算机辅助识别技术<sup>[6-7]</sup>。但都存在缺陷定位不准、对比度不高及微小缺陷丢失现象,对裂纹和未熔合等

收稿日期:2009-04-03

基金项目:成都信息工程学院自然科学与技术发展基金资助  
(csrf200805)

作者简介:李 坤(1987—),男,在读本科,主要研究方向是光信息科学与技术。

缺陷的分类和识别仍存在分类精度不高等问题。

SVM 是近年发展很快的用于焊接缺陷的分类和识别的方法<sup>[8]</sup>。SVM 本质上是一个二分类器, 目前对于多分类 SVM 的算法有① multi-class 算法<sup>[9]</sup>。② 一对一的多分类算法<sup>[10]</sup>。③ 一对多的多分类算法<sup>[11]</sup>; 但这些算法存在计算复杂程度高、泛化能力差等问题。

基于二叉决策树的 SVM, 由于采用分类器和决策树的结构, 使得 SVM 决策树有很多的优点, 特别是识别的精度比较高。基于此考虑, 笔者设计了一种自适应二叉树的 SVM, 用来解决 X-Ray 图像中的焊接缺陷的分类和识别, 提高识别焊缝缺陷的准确率, 以达到进一步改进焊接工艺的目的。

## 1 SVM 理论基础

对于给定样本集  $(x_i, y_i), i=1, \dots, n, x_i \in R^n, y_i \in \{-1, +1\}$ , 若是线性可分, SVM 通过在特征空间中构造最优分类面  $(w \cdot x) + b = 0$  来实现结构风险最小化, 最优分类面的求解为:

$$\text{Max } \Phi(w) = \frac{1}{2} \|w\|^2 \quad (1)$$

$$\begin{aligned} \text{Subject to } & \begin{cases} (w \cdot x_i) + b \geq 1 & y_i = 1 \\ (w \cdot x_i) + b \leq -1 & y_i = -1 \end{cases} \\ & \Leftrightarrow y_i[(w \cdot x_i) + b] \geq 1 \end{aligned}$$

若是线性不可分, 则引入松弛变量  $\xi_i$ , 把式(1)的求解改写成下式:

$$\text{Min } \left( \frac{1}{2} \|w\|^2 + C \sum_i \xi_i \right) \quad (2)$$

$$\text{Subject to } y_i(w \cdot x_i + b) \geq 1 - \xi_i, \xi_i \geq 0$$

其原理为将输入向量映射到一个高维的特征向量空间, 并构造最优分类面。通过定义核函数转化到输入空间进行计算, 且不同的核函数可以构造不同的支持向量机, 目前, SVM 普遍采用的内积核函数主要有多项式、径向基 RBF 和多层次感知器三类<sup>[12]</sup>。

## 2 自适应多类 SVM 决策树算法

### 2.1 分离程度的定义

目前测量向量类与类间距离时有基于计算各类重心间距离的欧氏距离。但对于在映射空间有类交叉情况(如图 1 所示), 当仍用欧氏距离进行计算时, 得到图 1(a)类间的欧氏距离与图 1(b)一样, 从而无法判别图 1(a)与(b)中各类的分离难易程度。但明显图 1(b)比图 1(a)中的类更容易分离出来, 由此可知类间分离难易程度, 不仅与类与类的重心间距离

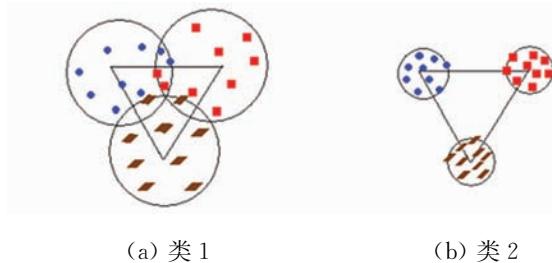


图 1 类分布图

有关, 也与各类所包围圆的半径有关。故而定义分离程度为:  $g_{ij} = \frac{d_{ij}}{r_i^2 + r_j^2}$ ,  $d_{ij}$  为类  $i$  与类  $j$  中心间的距离;  $r_i$  为在类  $i$  中以  $c_i$  为圆心, 以距离  $c_i$  最远的元素为半径。分离程度  $g_{ij}$  越大, 则类  $i$  与类  $j$  也越容易分离开。

### 2.2 基于分离程度的 SVM 决策树算法

针对传统二叉树的 SVM 存在的问题诸如分类器性能很大程度取决于树的结构、依赖于先验知识、易产生大的累积误差和推广性能不好等, 笔者设计了基于分离程度的 SVM 决策树算法。

算法思想是通过计算  $k$  类样本空间中各类的分离程度, 每次只先将分离程度最大的类分离出来, 构成一个结点的 SVM 分类器, 计算出来的最优分类面则对应决策树中相应根结点, 直至得到  $k-1$  个 SVM 分类器, 并利用其检测识别缺陷类型。这种基于分离程度的思想, 每次都先把最易分出的类先分离出来, 可使错分点远离决策树根结点, 从而有效地减小了累积误差, 大大提高了分类器的性能。具体分类算法步骤如下:

(1) 根据训练样本在属性空间的几何分布位置计算出每个类各自的中心:  $c_i = \frac{1}{N_i} \sum_{x_i \in X_i} x_i$  ( $i=1, 2, \dots, k$ ),  $N_i$  为类  $X_i$  的样本个数。

(2) 求类  $i$  与类  $j$  的分离程度  $g_{ij} = \frac{d_{ij}}{r_i^2 + r_j^2}$ ,  $d_{ij} = \|c_i - c_j\|$ ,  $g_{ij}$  越大, 表示类  $i$  与类  $j$  的分离程度越高, 且易知  $d_{ij} = d_{ji}$ 。

(3) 对于  $k$  类样本, 计算类  $i$  与其它各类的分离程度  $g_{ij}$  ( $j=1, 2, \dots, k$  且  $j \neq i$ ), 以类  $i$  与其他所有类的最小分离程度作为类  $i$  的分离程度, 即  $g_i = \min_{j \neq i} g_{ij}$ 。

(4) 用训练数据集选择样本  $x$  中分离程度最大的类  $i$ , 作为决策树根结点的左子结点, 剩下的类为右子结点, 利用 SVM 训练算法构造根节点处的二值 SVM 子分类器。

(5) 重新调整样本数据集  $x = x - i$ , 转步骤(4), 直至得到  $k-1$  个 SVM 子分类器。

(6) 用训练得到的  $k-1$  个二值 SVM 子类器对样本缺陷进行测试识别, 判断焊缝的具体缺陷类型。

(7) 程序结束。

### 3 试验

#### 3.1 X 射线焊缝图像的特征提取

采用东方汽轮机厂的电站锅炉的 X-Ray 图像进行分析, 先对滤波后的二值图像进行形态学开运算  $f \circ b = (f \Theta b) \oplus b$ <sup>[13]</sup>, 选取结构元素  $b = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ , 开运算可以去除较小的散点和毛刺(相

对结构元素而言), 即对图像进行平滑处理, 同时保留所有的灰度和较大的亮区特征不变。然后进行形态学开运算重建, 其中形态学开和闭重建运算分别定义为:  $O_r^{(rec)}(f) = D_r^{(rec)}(f \circ b, r)$ ,  $C_r^{(rec)}(f) = E_r^{(rec)}(f \cdot b, r)$ <sup>[14]</sup>,  $f$  是给定的图像;  $r$  为参考图像。其缺陷原始图像及经过数学形态学重建后的图像分别如图 2 所示, 形态学重建后共得到 763 个缺陷样本数据, 包括气孔缺陷 146 个, 夹钨 124 个, 夹渣 126 个, 未焊透 122 个, 裂纹 122 个, 未熔合 123 个。



图 2 X-Ray 缺陷原始图像和数学形态学重建后图像

对重建后的焊接图像特征提取是区分焊接缺陷的关键点, 其中几何形状是判断焊接缺陷的最重要的依据。基于此将图像特征分为以下几类<sup>[15-16]</sup>: ① 区域的密集度  $C$ ; ② 边缘平滑度  $R_v$ ; ③ 端部形态变化  $T$ ; ④ 长宽比  $L_1/L_2$ ; ⑤ 重心坐标相对焊缝中心的位置  $d$ 。

区域的密集度  $C$  越高, 围成的面积就越大; 邻接码值差的平均值若越大, 则目标区域边界越不平滑,  $R_v$  越小, 根据  $R_v$  可区分夹渣与气孔缺陷; 端部形态变化  $T$  越大表示端部形态越尖; 根据长宽比  $L_1/L_2$  可区分长形缺陷和圆形缺陷; 重心坐标相对焊缝中心的位置  $d$  也是反映缺陷类别的重要指标。表 1 为气孔和夹渣缺陷的特征参数测量值。

表 1 样本缺陷的特征参数

缺陷类型	$c$	$R_v$	$T$	$L_1/L_2$	$d$
气孔 1	0.058 4	2.051 3	0.974 4	1.111 8	0.670 3
气孔 2	0.055 9	2.666 7	1.239 5	1.150 1	0.628 6
夹渣 1	0.003 8	6.796 4	0.997 1	22.634 6	0.288 2
夹渣 2	0.007 9	5.574 3	1.021 1	23.910 3	0.277 9

针对以上 5 个特征信息对得到的缺陷样本进行 SVM 训练测试, 由于不同的特征参数具有不同的物



(a) 气孔



(b) 形态学重建的气孔二值图像



(c) 金属夹钨



(d) 形态学重建后的金属夹钨二值图像



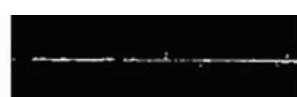
(e) 夹渣



(f) 形态学重建后的夹渣缺陷的二值图像



(g) 未焊透



(h) 形态学重建后的未焊透缺陷的二值图像



(i) 裂纹



(j) 形态学重建后的裂纹缺陷的二值图像



(k) 未熔合



(l) 形态学重建后的未熔合缺陷的二值图像

理意义,数值区间差别很大,因此对实际输入的特征参数进行归一化处理,将输入数据变换为[0,1]区间的数值,这样可大大提高 SVM 学习和测试精度。一组学习样本归一化处理后如表 2 所示。

表 2 归一化后样本缺陷的特征参数

缺陷类型	$c$	$R_v$	$T$	$L_1/L_2$	$d$
气孔 1	0.032 5	0.403 7	0.835 5	0.041 3	0.936 4
气孔 2	0.052 7	0.403 7	0.204 9	0.021 5	0.972 4
气孔 3	0.037 3	0.016 1	0.157 6	0.461 8	0.613 3
气孔 4	0.030 5	0.145 3	0.058 9	0.096 2	0.947 5
夹渣 1	0.012 8	0.744 1	0.483 5	0.687 2	0.359 1
夹渣 2	0.373 1	0.749 5	0.449 4	0.697 1	0.253 9
夹渣 3	0.043 3	0.695 3	0.479 9	0.278 9	0.309 5
夹渣 4	0.047 5	0.620 3	0.216 8	0.057 3	0.253 9
夹渣 5	0.087 1	0.369 5	0.514 7	0.688 6	0.371 1
未熔合 1	0.780 7	0.166 8	0.466 6	0.077 3	0.266 5
未熔合 2	0.328 9	0.949 9	0.449 6	0.256 5	0.244 7

### 3.2 SVM 建模及试验结果

对已得的缺陷样本参数,采用本文的自适应的 SVM 决策树算法,其中分类器采用 SVM,其核函数采用 RBF。通过测试不同的参数对( $C, \sigma$ )来获取推广精度高的  $C$  和  $\sigma$  参数,以气孔和夹渣为例来进行交叉验证,计算得到的最优惩罚系数  $C$  和 RBF 核参数  $\sigma$  为: $C=4\ 348.010\ 8, \sigma=0.001\ 649$ 。

通过 SVM 决策树算法,先对第一个根结点中 6 种缺陷类型计算类与类间的分离程度,将分离程度最大的类提出来作为左子结点,利用二值 SVM 训练算法构造二叉树第一个根结点的最优超平面,得到分类器 SVM-1,依此类推 SVM-2,SVM-3 等分类器分别表示对二叉树第二个、第三个内结点进行训练得到的分类器。由本文算法可知对于 6 种缺陷类型会在 5 个结点处得到 5 个 SVM 分类器。

经过 SVM 训练后,得到的最后的二叉树结构如图 3。

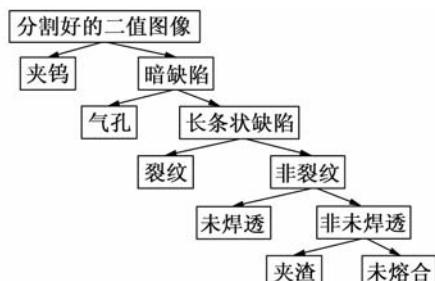


图 3 基于分离程度的 SVM 决策树

最后计算出来的气孔、夹钨、未焊透、夹渣、裂纹和未熔合的正确识别率如表 3 所示,并与其它的多类 SVM 算法进行了比较如表 4 所示。

表 3 最后的识别结果

类别	气孔	金属夹钨	夹渣	未熔合	未焊透	裂纹
训练数据	85	73	75	71	71	71
测试数据	61	51	51	52	51	51
训练数据识别率/%	96.3	100	96.7	91.0	91.6	92.1
测试数据识别率/%	95.5	100	91.4	87.8	88.5	87.0

表 4 与其它多分类 SVM 算法结果比较 (%)

类别	气孔	夹钨	夹渣	未熔合	未焊透	裂纹
一对一	93.2	98.7	88.6	80.4	81.9	81.0
一对多	90.4	94.6	91.2	78.5	83.8	80.8
自适应 SVM 决策树	95.5	100	91.4	87.8	88.5	87.0

### 4 结论

(1) 在构造决策树的过程中,通过计算分离程度来逐层优先选择最容易分的类出来,最大程度地减小了累积误差,从而提高了缺陷的分类识别精度。

(2) 在与其它多分类 SVM 精度比较后,表明在不需要先验知识的情况下,改进后的 SVM 决策树比其它 SVM 多分类算法具有更高的正确分类率。

(3) 良好的识别分类特性表明,支持向量机在 X-Ray 焊接图像中的缺陷识别是很有发展前途的技术。

### 参考文献:

- [1] 中国机械工程学会. 射线检测[M]. 北京: 机械工业出版社, 1997: 4—177.
- [2] 李德元, 邵成吉, 徐鲁宁. 焊接缺陷自动检测中区分典型条形缺陷判据的建立[J]. 焊接技术, 1998, 28(1): 7—8.
- [3] 傅德胜. 焊接缺陷计算机自动识别模式的研究[J]. 控制与决策, 1998, 13(增刊): 469—474.
- [4] 孙忠诚, 李鹤歧, 陶维道. 焊缝 X 射线实时探伤数字图像处理方法研究[J]. 无损检测, 1992, 14(2): 37—40.
- [5] Nacereddine Nafa, Drai Redouane. Weld defect extraction and classification in radiographic testing based artificial neural networks[C]// Proceeding of 15th WCNDT. Roma: [s. n. ], 2000.

(下转第 178 页)

因此,如果根据反射信号的声程信息来判断缺陷的深度,误差较大。现场焊缝和试块上的波形图见图 13 和图 14。

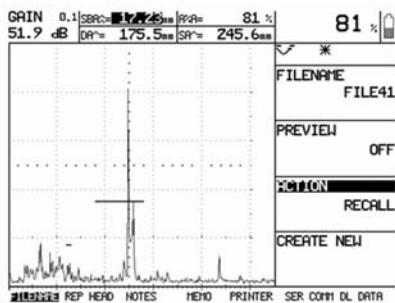


图 13 现场焊缝检测波形图



图 14 试块上的波形图

### 3.4 检测结果分析

从上述检测结果来看,LLT 探头能发现明显的回波,信噪比 $>20$  dB,而无缺陷区域则没有任何明显的回波显示,这说明焊缝中确实存在有利于 LLT 探头声束传播的反射体;2Z10×10A45探头发现信

(上接第 174 页)

- [6] Kapustin A E, Bardusova I I. Computer technologies and X-ray flaw detection of welds[C]// Proceeding of 15th WCNDT. Roma:[s. n.],2000.
- [7] Nacereddine N, Tridi M. Weld defects in industrial radiography based invariant attributes and neural networks[J]. Image and Signal Processing and Analysis, 2005(7):15—17.
- [8] Burges C. A tutorial on support vector machines for pattern recognition[J]. Data mining and knowledge discovery,1998,2(2):121—167.
- [9] Hus C W, Lin C J. A comparison of methods for multi-class support vector machines[J]. IEEE Transaction on Neural Networks,2002,26(13):414—425.
- [10] Brown M P, Grundy W N, Lin D, et al. Knowledge-based analysis of microarray gene expression data by using support vector machines[C]// Proceeding of

号时信噪比也比较高,同时移动探头时能观察到明显的端点反射信号,而这是裂纹的典型特征;直探头检测时需要将仪器增益提高到很高的数值才能发现微弱的反射信号,据此可以辅助判断该缺陷是面积性缺陷,而不是体积性缺陷。

根据以上分析,检测人员最终确定该缺陷为裂纹,并对裂纹的自身高度进行了测量。

### 4 结语

从现场检测效果来看,LLT 探头对垂直于检测面的区域性缺陷有着很高的检测灵敏度和信噪比。因此,对厚壁窄间隙焊缝进行超声检测时,除采用传统斜探头检测外,在现场条件允许的情况下,还应采用 LLT 探头尽量扫查到整个焊缝截面,发现信号后再采用端点反射法对缺陷高度进行测量。但是,肯定 LLT 技术检测能力的同时,也应该注意该技术在工件上、下表面存在检测盲区,且当缺陷方向倾斜角度较大时,采用 LLT 技术也可能得不到很好的效果。因此,LLT 法必须结合其他超声检测方法一起使用,才能保证检测结果的准确性。

致谢:本文得到了国核电站运行服务技术公司毕炳荣老师的指导和提出的宝贵意见。

### 参考文献:

- [1] Komura Ichiro. Nondestructive inspection technologies for the inservice inspection of NPP[J]. Journal of the Japan Institute of Energy,2004,83(7):485—492.
- the National Academy of Sciences of the USA. USA: [s. n. ],2000:262—267.
- [11] 孙林,付静.基于二叉树的支持向量机旋转机械故障诊断方法[J].设计与研究,2008,35(3):1—4.
- [12] Trevor Hastie, Robert Tibshirani, Jerome Friedman. The Elements of Statistical Learning Data Mining, Inference, and Prediction[M]. American: Springer-verlag,2004. 263—272.
- [13] 冈萨雷斯.数字图像处理[M].北京:电子工业出版社,2003:427—431.
- [14] Crespo J, Serra J. Theoretical aspects of morphological filters by reconstruction[J]. Signal Processing, 1995,47(2):201—225.
- [15] Katoh Y, O kumura T, Itoga K, et al Development of the automatic system for radiographic film interpretation[J]. NDT of Japan,1992,41(4):186—195.
- [16] 张晓光,林家骏.X 射线检测焊缝的图像处理与缺陷识别[J].华东理工大学学报,2004,30(2):199—202.