

DOI: 10.15923/j.cnki.cn22-1382/t.2019.1.13

改进自适应遗传 SVM 滚动轴承故障诊断

冯 新

(福州大学至诚学院 计算机工程系, 福建 福州 350002)

摘 要: 采用自适应遗传算法优化 SVM 相关参数, 建立了 IAGA-SVM 滚动轴承故障诊断模型。通过实验验证了该模型的精确性和有效性。

关键词: 滚动轴承; 支持向量机; 遗传算法; 故障诊断

中图分类号: TP 18 **文献标志码:** A **文章编号:** 1674-1374(2019)01-0072-05

Fault diagnosis of rolling bearing based on the improved Adaptive Genetic Algorithm optimization SVM

FENG Xin

(Department of Computer Engineering, Zhicheng College, Fuzhou University, Fuzhou 350002, China)

Abstract: Adaptive Genetic Algorithm is used to optimize parameters of SVM, and an IAGA-SVM rolling bearing fault diagnosis model is established. Experiments verify the feasibility of the model.

Key words: rolling bearing; support vector machine; genetic algorithm; fault diagnosis.

0 引 言

滚动轴承是轨道工程机械设备的关键组成元件, 如果可以在滚动轴承性能衰退的过程中及时检测出故障部位及故障类型, 就可以变被动维修到主动维修^[1]。一方面提高了滚动轴承的使用寿命, 另一方面也大大降低了因故障发现不及时而造成的各种不良后果。所以, 对轨道工程机械设备滚动轴承进行故障分类和故障预测具有重大的工程价值。

文中提出一种基于小波变化和改进的自适应遗传算法优化 SVM 分类的滚动轴承故障诊断模型。首先对采集到的数据进行预处理并做去噪处

理, 其次利用小波变换对去噪后的信号进行特征提取, 继而对提取获得的特征向量做进一步的归一化处理, 并使用改进的自适应遗传算法对 SVM 的相关参数进行优化处理, 建立了 IAGA-SVM 故障诊断模型, 最后采用仿真实验来说明笔者所提方法的优越性。

1 模型的描述与建立

1.1 支持向量机原理分析

支持向量机是 Vapnik 等^[2]在统计学理论的结构风险最小化和 VC 维的基础上提出的一种新的机器学习方法^[3]。SVM 能够在有限特征信息情况下较好地解决小样本集的回归和分类问题,

收稿日期: 2018-10-10

基金项目: 福建省本科高校教育教学改革研究项目(FBJG20180294)

作者简介: 冯 新(1983—), 男, 汉族, 福建福州人, 福州大学至诚学院讲师, 硕士, 主要从事数据挖掘方向研究, E-mail: fxfengxin@qq.com.

它比神经网络具有更强的泛化能力^[4]。鉴于SVM自身的优越性,现已成为目前最常用、最有效的分类器之一。

假设 n 个训练样本的样本集表示为:

$$D = \{(x_i, y_i)\}_{i=1}^n$$

$$x \in R^n, y \in \{-1, 1\}$$

当样本集 D 为线性可分时,则式(1)为样本的某一超平面方程:

$$\omega \cdot x + b = 0 \quad (1)$$

其最优分类超平面必须满足如下方程:

$$y_i(\omega \cdot x_i + b) \geq 1 \quad (2)$$

$$i = 1, 2, \dots, n$$

此时可采用下式对训练样本数据进行分类:

$$\min_{\omega, b} \frac{1}{2} \|\omega\|^2 = \min_{\omega, b} \frac{1}{2} \omega^T \omega \quad (3)$$

利用拉格朗日函数来对上述问题进行求解:

$$L(\omega, b, \alpha) = \frac{1}{2} \|\omega\|^2 - \sum_{i=1}^n \alpha_i [y_i(\omega x_i + b) - 1] \quad (4)$$

式中: α_i —— Lagrange 乘子, $\alpha_i \geq 0, i = 1, 2, \dots, n$ 。

在 $\sum_{i=1}^n \alpha_i y_i = 0$ 和 $\alpha_i \geq 0, i = 1, 2, \dots, n$ 的约束条件下,对 Lagrange 乘子 α_i 求最大值:

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \quad (5)$$

式(5)由于约束条件的存在,故其解仅有一个。假设其最优解是 α_i^* ,则可求出其最优分类函数为:

$$f(x) = \text{sgn}(\omega^* \cdot x_i + b^*) = \text{sgn}\left\{\sum_{i=1}^n \alpha_i^* y_i (x_i \cdot x) + b^*\right\} \quad (6)$$

当训练样本为线性不可分时,则需要寻求一个合适的映射函数对数据进行非线性映射。用字母表达为:

$$\begin{aligned} \varphi: R^n &\rightarrow H \\ x &\rightarrow \varphi(x) \end{aligned}$$

由 Mercer 条件可知,若核函数 $K(x_i, x_j)$ 可以写成

$$K(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j)$$

那么问题就转化为在约束条件

$$\sum_{i=1}^n \alpha_i y_i = 0$$

$$\alpha_i \geq 0, i = 1, 2, \dots, n$$

求解二次规划问题:

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i \cdot x_j) \quad (7)$$

求解所得的最优分类函数式为:

$$f(x) = \text{sgn}\left\{\sum_{i=1}^n \alpha_i^* y_i K(x_i \cdot x) + b^*\right\} \quad (8)$$

从上述表达式可以看出,核函数的选择对SVM性能的优劣起关键作用。文中选择径向基核函数 RBF,即

$$K(x_i, x_j) = \exp(-\gamma^* \|x_i, x_j\|^2)$$

$$\gamma = \frac{1}{2\sigma^2} > 0$$

γ 是描述核函数的作用范围,决定了模型的复杂程度。引入松弛变量 $\xi_i \geq 0$,则此时的约束条件由

$$\begin{aligned} y_i(\omega x_i + b) &\geq 1 \\ i &= 1, 2, \dots, n \end{aligned}$$

变为

$$\begin{aligned} y_i(\omega \cdot x_i + b) &\geq 1 - \xi_i \\ i &= 1, 2, \dots, n \end{aligned} \quad (9)$$

相对应的二次凸规划问题变为

$$\min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n \xi_i \quad (10)$$

式中: C —— 大于零的惩罚因子,用来权衡对错样本的惩罚力度。

SVM 分类模型最初被提出来的时候仅仅是用来实现二值分类问题,但实际工程应用中二值分类是非常少有的,往往是多分类问题更加常见。目前使用较多的则为“一对一”、“一对多”和二叉树结构的多分类方法^[5],文中选取基于二叉树结构的支持向量机多分类方法(DT-SVM)。

1.2 自适应遗传算法

自适应遗传算法(Adaptive Genetic Algorithm, AGA)的改进之处是将遗传算法中固定的交叉概率和变异概率进行线性的自适应调整^[6]。交叉概率 P_c 决定了遗传算法的全局搜索能力,变异概率 P_m 决定了遗传算法的局部搜索能力,所以 P_c 和 P_m 对遗传算法的寻优能力起决定性作用。两者的适应度调整公式为:

$$P_c = \begin{cases} \frac{\lambda_1 (f_{\max} - f_{\text{bigger}})}{f_{\max} - f_{\text{arg}}}, & f_{\text{bigger}} \geq f_{\text{arg}} \\ \lambda_2, & f_{\text{bigger}} < f_{\text{arg}} \end{cases} \quad (11)$$

$$P_m = \begin{cases} \frac{\lambda_3 (f_{\max} - f)}{f_{\max} - f_{\text{arg}}}, & f \geq f_{\text{arg}} \\ \lambda_4, & f < f_{\text{arg}} \end{cases} \quad (12)$$

式中： $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ ——在 $[0, 1]$ 内随机取值的常数；

f_{\max} ——群体中最大的适应度值；

f_{bigger} ——参与交叉运算的两个个体中较大的适应度值；

f_{arg} ——群体的平均适应度值；

f ——当前变异个体的适应度值。

由式(11)和式(12)可知, 如果 $f_{\max} = f_{\text{arg}}$, 则

$$P_c = \begin{cases} P_{\min 1} + \frac{P_{\max 1} - P_{\min 1}}{1 + \exp\left(A \left(\frac{2(f_{\text{bigger}} - f_{\text{arg}})}{f_{\max} - f_{\text{arg}}}\right)\right)}, & f_{\text{bigger}} \geq f_{\text{arg}} \\ P_{\max 1}, & f_{\text{bigger}} < f_{\text{arg}} \end{cases} \quad (13)$$

$$P_m = \begin{cases} P_{\min 2} + \frac{P_{\max 2} - P_{\min 2}}{1 + \exp\left(A \left(\frac{2(f - f_{\text{arg}})}{f_{\max} - f_{\text{arg}}}\right)\right)}, & f \geq f_{\text{arg}} \\ P_{\max 2}, & f < f_{\text{arg}} \end{cases} \quad (14)$$

式中： $P_{\max 1}, P_{\min 1}$ ——分别表示交叉概率的上限和下限；

$P_{\max 2}, P_{\min 2}$ ——分别表示变异概率的上限和下限。

这样的设置使得 P_c 和 P_m 的适应度值在 f_{\max} 和 f_{arg} 之间进行非线性的调整, 在很大程度上规避了局部收敛的弊端。

2 IAGA-SVM 故障诊断模型

文中考虑到支持向量机和遗传算法各自的优

式(11)和式(12)的分母都变为零, 这样的表达在数学中是不成立的。此外 $f_{\max} = f_{\text{arg}}$ 表示群体中所有个体的基因组成都是一样的, 这时候的优良个体有很大可能是局部最优解而不是全局最优解, 从而导致进化无法继续进行下去。

1.3 改进的自适应遗传算法

为了尽可能地避免群体陷入停滞不前的境况, 文中对上述自适应遗传算法的公式进行改进:

缺点, 提出一种利用改进的自适应遗传算法优化支持向量机核函数参数和惩罚因子的方法, 以此来提高支持向量机的分类能力和泛化能力。采用二进制和实数混合编码的方式对 SVM 中的相关参数进行遗传编码, 以此来提高算法的求解精度, 将最终搜索得到的最优结果作为 SVM 模型的参数进行建模。基于改进的自适应遗传算法优化 SVM 的故障诊断流程如图 1 所示。

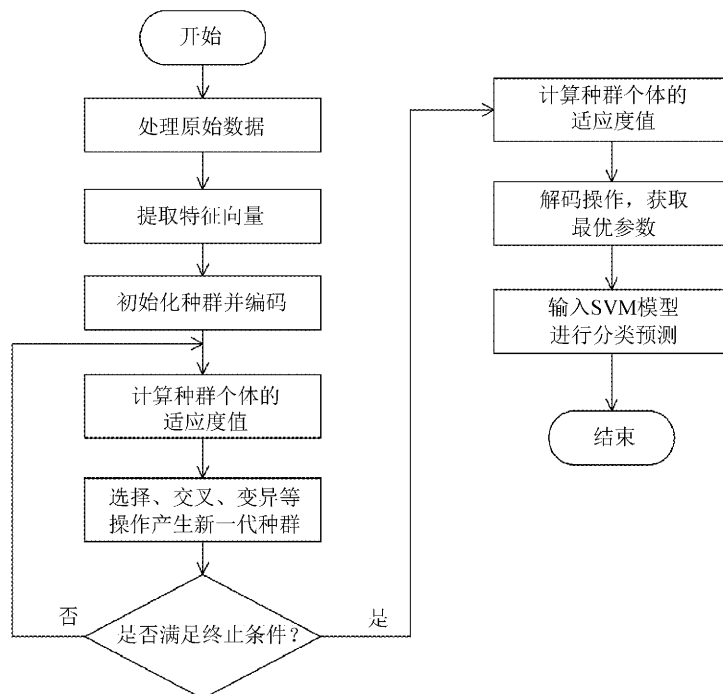


图1 IAGA-SVM 故障诊断模型流程图

3 实验结果与分析

3.1 实验系统的确立

为了证明文中所提 IAGA-SVM 故障诊断模型的有效性,设计如下实验进行验证。实验数据源自美国凯斯西储大学(Case Western Reserve University)^[7]所提供的滚动轴承数据。传感器数据采集频率为 12 000 Hz,转速为 1 730 r/min,故障损伤直径选用 0.053 cm。实验中将滚动轴承故障分为 4 种状态,分别是正常状态、轴承内圈故障、轴承外圈故障和轴承球体故障^[8]。

3.2 IAGA-SVM 模型实验结果分析

文中通过小波对信号进行去噪后提取特征向量,再选取各状态下 80 组样本作为训练样本(共 320 组),用于训练模型,选取各状态下 20 组特征向量为测试样本(共 80 组)作为测试所用。根据经验设置惩罚系数为 $[0.005, 100]$,RBF 径向基函数的参数为 $[0.05, 100]$,最大迭代次数设置为 200 次。

为了证明文中所提模型的高效性,特设计对比模型来进行实验。遗传算法对支持向量机参数寻优(GA-SVM)、自适应遗传算法对支持向量机参数寻优(AGA-SVM)、改进自适应遗传算法对支持向量机参数寻优(IAGA-SVM)的适应度曲线如图 2 所示。

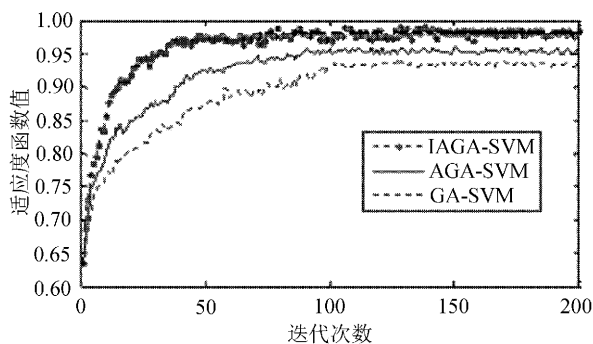


图2 GA-SVM、AGA-SVM 和 IAGA-SVM 的适应度曲线

三种模型的适应度函数均取支持向量机最终分类准确率。由图 2 可以看出,GA-SVM 于 102 代左右时适应度变化趋于平稳,但容易造成最终收敛速度慢,甚至存在早熟现象;AGA 能够自适应调整交叉率以及变异率,使得其收敛速度加快,因此,可以看到 AGA-SVM 在 63 代左右已趋于收敛,相对于 GA-SVM 而言,它的收敛速度已经加快不少,但 AGA 算法在种群进化初期容

易导致进化停滞不前,陷入局部最优且不利于增强算法的鲁棒性;IAGA-SVM 模型仅迭代 45 次便趋于收敛,明显优于前两种寻优模型,从而使得改进的自适应遗传算法能够更高效的完成收敛,验证了文中所提模型的高效性。

为了测试 IAGA-SVM 模型对滚动轴承 4 种状态的辨识度,现对每种故障状态分别选取 20 个测试样本,使用经过训练之后的 IAGA-SVM 分类模型对滚动轴承进行测试,测试结果如图 3 所示。

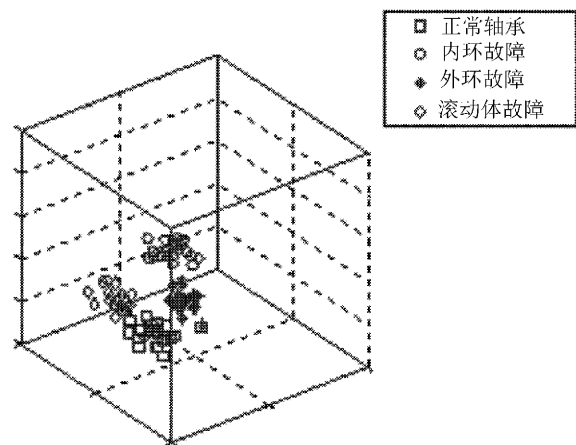


图3 IAGA-SVM 模型分类结果图

由图 3 可以看出,正常状态、内环故障、外环故障 20 个测试样本均确诊,其分类准确率分别达到 100%,但滚动体故障 20 个测试样本只有 19 个测试样本确诊,一个测试样本误诊为正常轴承,分类准确率达到 95%。由此可以得到该模型的平均准确率约高达 98.75%。可见 IAGA-SVM 模型作用于滚动轴承故障诊断识别率高,效果明显,具有较大的工程意义。

4 结语

在对滚动轴承的原始数据进行去噪处理之后,进一步对处理后的信号进行特征提取。针对 SVM 方法的弊端,特选用自适应遗传算法对 SVM 的相关参数进行优化处理。但由于自适应遗传算法自身所存在的缺陷,并不能达到预期的分类结果,故而对自适应遗传算法进行改进。

实验结果可以看出,所提出的 IAGA-SVM 模型对滚动轴承进行故障分类识别,无论在收敛速度上还是在分类的准确率上,优于其他两种实验模型,从而验证了文中所提的故障诊断模型对

轴承故障分类的优越性,具有可观的工程价值。

参考文献:

- [1] 秦喜文,郭宇,郭佳静,等.基于 LMD 与随机森林的滚动轴承故障诊断[J].长春工业大学学报,2018,39(5):417-423.
- [2] Cortes C, Vapnik V. Support-vector networks[J]. Machine Learning,1995,20(3):273-297.
- [3] 崔清亮.多核学习方法在分类问题中的应用研究[D].兰州:兰州交通大学,2014.
- [4] 黄剑锋.基于振动信号 SVM 的管壳式换热器堵塞故障诊断方法研究[D].广州:华南理工大学,2016.
- [5] 李志农,刘卫兵,肖尧先,等.基于局域均值分解包络谱和 SVM 的滚动轴承故障诊断方法研究[J].机械设计与制造,2011(11):170-172.
- [6] Srinivas M, Patnaik L M. Adaptive probabilities of crossover and mutation in genetic algorithms[J]. IEEE Transactions on Systems Man & Cybernetics,2002,24(4):656-667.
- [7] 美国西储大学轴承数据[EB/OL]. (2018-04-15) [2018-09-18]. <http://csegroups.case.edu/bearing-datecenter/pages/download-data-file>.
- [8] 贾艳秋,张兵,陈雪梅.滚动轴承的故障机理及诊断[J].化工装备技术,2011,32(4):55-57.