

DOI:10.3969/j.issn.1001-4551.2021.05.012

基于 CEEMDAN 多尺度熵和 SSA-SVM 的 滚动轴承故障诊断研究*

李 怡, 李焕锋*, 刘自然

(河南工业大学 机电工程学院, 河南 郑州 450001)

摘要:针对支持向量机(SVM)应用在轴承故障分类时,传统的智能算法优化 SVM 的参数容易存在寻优速度慢、调节参数多,以及容易陷入局部最优值等问题,提出了一种基于 CEEMDAN 多尺度熵与 SSA-SVM 相结合的故障诊断方法。对滚动轴承的故障特征提取和 SVM 参数优化进行了研究,引入了一种新的群智能优化算法,用麻雀搜索算法(SSA)对 SVM 参数进行了优化,提高了寻优速度以及轴承的故障分类准确率;该方法先采用自适应白噪声完整经验模态分解(CEEMDAN)算法分解信号,获得了若干个固有模态函数(IMF);再采用相关系数方法选择有用 IMF 分量,并进行了重新组合;最后,计算重构信号的多尺度熵作为特征向量,输入 SSA 优化的 SVM 进行了故障分类。研究表明:采用该方法能够准确地获得故障信息,且识别准确率高;与 PSO、GA 优化的 SVM 相比,该方法的故障诊断分类性能更好。

关键词:自适应白噪声完整经验模态分解;多尺度熵;麻雀搜索算法;支持向量机;故障诊断

中图分类号:TH133.33

文献标识码:A

文章编号:1001-4551(2021)05-0599-06

Fault diagnosis of rolling bearing based on CEEMDAN multi-scale entropy and SSA-SVM

LI Yi, LI Huan-feng, LIU Zi-ran

(School of Mechanical and Electrical Engineering, Henan University of Technology, Zhengzhou 450001, China)

Abstract: Aiming at the problem that when support vector machine (SVM) was applied to bearing fault classification, the traditional intelligent algorithm optimization of SVM parameters had problems such as slow optimization speed, more adjustment parameters, and easy to fall into local optimal values, a fault diagnosis method based on CEEMDAN and SSA-SVM was proposed. The fault feature extraction and SVM parameter optimization of rolling bearings were studied. A new swarm intelligence optimization algorithm sparrow search algorithm (SSA) was introduced to optimize the parameters of the SVM to improve the optimization speed and the accuracy of bearing fault classification. Firstly, the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) algorithm was used to decompose the signal to obtain several intrinsic mode functions (IMF). Then the correlation coefficient method was used to select the useful IMF components and recombine them. Finally, the multi-scale entropy of the reconstructed signal was calculated as the feature vector and inputted into the SVM optimized by SSA for fault classification. The results indicate that this method can accurately obtain fault information and has high recognition accuracy. Comparing with SVM optimized by PSO and GA, this method has better fault diagnosis and classification performance.

Key words: complete ensemble empirical mode decomposition with adaptive noise(CEEMDAN); multi-scale entropy; sparrow search algorithm(SSA); support vector machine(SVM); fault diagnosis

收稿日期:2020-10-07

基金项目:河南省自然科学基金资助项目(182300410234)

作者简介:李怡(1996-),女,黑龙江佳木斯人,硕士研究生,主要从事故障诊断技术方面的研究。E-mail:1325013527@qq.com。

通信联系人:李焕锋,男,教授,硕士生导师。E-mail:synglclhf@haut.edu.cn。

0 引 言

滚动轴承作为旋转机械的重要组成部分,一旦发生故障时,旋转机械就会受到很大的影响。因此,对滚动轴承故障诊断研究是十分必要的^[1]。

在故障特征提取方面,自适应白噪声完整经验模态分解(CEEMDAN)是 Torres 等人在集合经验模态分解(EEMD)算法的基础上,提出出来的一种信号处理算法^[2]。该算法对 EEMD 算法进行了优化处理。与 EEMD 算法相比,CEEMDAN 算法改善了分解的完整性,降低了重构的误差,还提高了分解的效率。

在故障分类方面,支持向量机(SVM)被广泛应用于机械故障诊断方面^[3]。它在解决非线性、小样本以及高维模式识别的问题中有很好的表现^[4]。但在 SVM 中,惩罚因子 C 以及核函数参数 σ 的取值对故障诊断的准确性起决定性的作用^[5]。为了提高 SVM 故障识别的准确率,张小龙等人^[6]提出了一种基于 ITD 复杂度和 PSO-SVM 的滚动轴承故障诊断。赵蕾等人^[7]提出了一种基于 FOA-WPT 降噪和 PSO-SVM 的滚动轴承故障诊断方法。时培明等人^[8]提出了一种基于分形维数和 GA-SVM 的轴承故障诊断方法。然而传统的智能算法在优化 SVM 参数时,存在寻优速度慢、调节参数多,以及容易陷入局部最优值等问题,从而导致其故障分类的准确率较低。

本文引入一种麻雀搜索算法优化支持向量机(SSA-SVM),提出一种基于 CEEMDAN 多尺度熵与 SSA-SVM 相结合的故障诊断方法,用于轴承故障诊断,并通过实验来证明该方法能够准确地获得故障信息,且故障识别效果更好。

1 故障特征提取

1.1 CEEMDAN 算法

设 $x(t)$ 为实验数据,则 CEEMDAN 算法对信号分解过程具体如下:

(1) IMF_1 分量与 EEMD 算法分解得到的 IMF_1 分量相同;

$$IMF_1(t) = \frac{1}{N} \sum_{i=1}^N IMF_{i1} \quad (1)$$

从 $x(t)$ 中去除 $IMF_1(t)$ 得残余分量 $r_1(t)$:

$$r_1(t) = x(t) - IMF_1(t) \quad (2)$$

(2) 采用 EMD 算法将 N 次高斯白噪声分解,获得 IMF 和残余分量 $r_1(t)$ 相加,再通过 EMD 算法,取均值得到的第一个 IMF ,将其作为 $IMF_2(t)$:

$$IMF_2(t) = \frac{1}{N} \sum_{i=1}^N E_1(r_1(t) + \varepsilon_1 E_1(v_i(t))) \quad (3)$$

式中: $E_1(\cdot)$ —通过 EMD 算法提取的第 1 个 IMF ; v_i —单位方差均值为 0 的高斯白噪声; ε_1 —幅值。

则第二个残余分量为:

$$r_2(t) = r_1(t) - IMF_2(t) \quad (4)$$

(3) 重复上面步骤,可求得第 j 个 $IMF_j(t)$ 和残余分量 $r_j(t)$;

$$IMF_j(t) = \frac{1}{N} \sum_{i=1}^N E_{j-1}(r_{j-1}(t) + \varepsilon_{j-1} E_{j-1}(v_i(t))) \quad (5)$$

$$r_j(t) = r_{j-1}(t) - IMF_{j-1}(t) \quad (6)$$

若残余分量不能在分解,则中断上述过程,此时信号 $x(t)$ 为:

$$x(t) = \sum_{j=1}^j IMF_j(t) + r(t) \quad (7)$$

1.2 多尺度熵

多尺度熵是计算不同时间尺度下信号的复杂程度^[9]。它具有良好的抗干扰和抗噪效果^[10]。当一个信号越复杂,波动越大,多尺度熵就越大;同样,信号越规则,波动越小,多尺度熵就越小。

该算法详细步骤如下所示:

(1) 粗粒化处理。粗粒化处理时间序列就表示为对不同数量的连续点取平均值,来创建不同尺度的信号,即:

$$y_i(\tau) = \frac{1}{\tau} \sum_{j=(i-1)\tau}^{i\tau} x_j \quad (8)$$

式中: x_i —时间序列; τ —时间尺度。

(2) 求各时间尺度下,经过粗粒化处理过的信号的样本熵记 MSE,即:

$$MSE(X, \tau, m, r) = \text{SampEn}(y(\tau), m, r) \quad (9)$$

式中: m —嵌入维数; r —相似容量。

2 麻雀搜索算法优化支持向量机

2.1 支持向量机

SVM 自身理论比较复杂,难于理解,所以研究人员自己实现该算法有难度。但是通过多年的摸索与研究,现在已经有许多研究人员研究出了相关的软件包,故直接安装运行即可。该实验采用的 SVM 是 Libsvm 软件包中的一个部分。

在采用 SVM 进行故障分类时,确定适合的核函数、惩罚因子 C 以及核函数参数 σ 是 SVM 的核心^[11]。

基于径向基核函数(RBF)有很强的局部性能以及抗干扰等特点,在轴承故障诊断中 SVM 常选择 RBF

核函数。因此,惩罚因子 C 以及核函数参数 σ 的取值对 SVM 故障识别的准确性至关重要。

2.2 麻雀搜索算法

麻雀搜索算法 (sparrow search algorithm, SSA) 是薛建凯等人^[12]于 2020 年基于麻雀的觅食行为和反捕食行为提出的一种新的群智能优化算法。该算法具有全局搜索能力强、寻优速度快、收敛速度快等优点。

SSA 的社会行为可以通过以下列数学模型来描述:

假设麻雀种群表示为:

$$\mathbf{X} = \begin{bmatrix} x_1^1 & x_1^2 & \cdots & x_1^d \\ x_2^1 & x_2^2 & \cdots & x_2^d \\ \cdots & \cdots & \cdots & \cdots \\ x_n^1 & x_n^2 & \cdots & x_n^d \end{bmatrix} \quad (10)$$

式中: n —麻雀的数量; d —要优化的变量的维数。

则所有麻雀的适应度值可以用以下向量来表示:

$$\mathbf{F}_x = \begin{bmatrix} f([x_{1,1} & x_{1,2} & \cdots & x_{1,d}]) \\ f([x_{2,1} & x_{2,2} & \cdots & x_{2,d}]) \\ \vdots & \vdots & \vdots & \vdots \\ f([x_{n,1} & x_{n,2} & \cdots & x_{n,d}]) \end{bmatrix} \quad (11)$$

式中: \mathbf{F}_x 中每一行的值一个体的适应度值。

在搜索过程中,适应度较高的麻雀会先得到食物。探索者作为麻雀种群的领导者负责搜索食物,且提供搜索食物的方向,故探索者的搜索食物范围最大。

当麻雀种群周围存在没有捕食者时,探索者可以随意进行食物搜索;一旦发现了周围存在捕食者,探索者会带着追随者向安全地方移动。

探索者的位置如下:

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j} \cdot \exp\left(-\frac{i}{\alpha \cdot \text{ite} r_{\max}}\right), R_2 < ST \\ X_{i,j} + Q \cdot L, R_2 \geq ST \end{cases} \quad (12)$$

式中: t —当前迭代次数; $\text{ite} r_{\max}$ —最大迭代次数; X_{ij} —第 i 个麻雀在第 j 维的位置信息; α —随机数, $\alpha \in (0, 1]$; R_2 —预警值, $R_2 \in (0, 1]$; ST —安全值, $ST \in (0.5, 1]$; Q —服从正态分布的随机数; L — $1 \times d$ 的矩阵,其中该矩阵内每个元素全部为 1。

当 $R_2 < ST$ 时,表示周围没有捕食者;反之,发现周围存在捕食者。

在整个麻雀种群中,探索者与追随者的比重是不变的,只要能够找到好的食物,就能成为探索者,反之就是追随者。麻雀种群中追随者的觅食环境与范围很差,因此,它们就会随时注意探索者的情况并与探索者

争抢食物,若抢夺成功,它们就会获得探索者的食物代替去更远的地方搜索食物。

则追随者的位置为:

$$X_{i,j}^{t+1} = \begin{cases} Q \cdot \exp\left(\frac{X_{\text{worst}} - X_{i,j}^t}{t^2}\right), i > n/2 \\ X_p^{t+1} + |X_{i,j} - X_p^{t+1}| \cdot A^+, i \leq n/2 \end{cases} \quad (13)$$

式中: X_{worst} —全局最差位置; X_p —当前发现者中最优位置; A —代表一个 $1 \times d$ 的矩阵,其中该矩阵内每个元素为 1 或 -1,并且 $A^+ = A^T(AA^T)^{-1}$ 。

当 $i > 2/n$ 时,表示此时追随者没有获得食物,需要去更远的地方搜索食物。

在整个麻雀种群中,假设警惕者占总数量的 10% ~ 20%,且警惕者的初始位置是在种群中随机分布的,当警惕者意识到周围存在捕食者时,外围的麻雀将快速地向安全的地方飞行,来获取优越的搜索环境;内部的麻雀将在安全区域内一直走动,减少被捕食概率。

警惕者的数学模型为:

$$X_{i,j}^{t+1} = \begin{cases} X_{\text{best}}^t + \beta \cdot |X_{i,j}^t - X_{\text{best}}^t|, f_i > f_g \\ X_{i,j}^t + K \cdot \left(\frac{|X_{i,j}^t - X_{\text{worst}}^t|}{(f_i - f_w) + \varepsilon}\right), f_i = f_g \end{cases} \quad (14)$$

式中: X_{best} —当前的全局最优位置; β —步长且服从标准正态分布; K —代表一个随机数, $K \in [-1, 1]$; f_i —当前麻雀个体的适应度值; f_g —当前全局最佳适应度; f_w —当前全局最差的适应度值; ε —常数,以避免分母出现 0。

当 $f_i > f_g$ 时,表示外围麻雀发现捕食者;反之,中间麻雀发现捕食者。

2.3 SSA-SVM 的优化步骤

SSA-SVM 的具体优化步骤如下:

(1) 设置 SSA 算法中的麻雀群总数 n , 最大迭代次数 $\text{ite} r_{\max}$, 发现者的比例、追随者的比例以及 C 、 σ 的取值范围,随机初始化麻雀种群;

(2) 计算每个麻雀的适应度,并进行排序,定义每只麻雀所属的种群;

(3) 根据式(12~14)更新每种麻雀种群的位置;

(4) 对更新位置后的每只麻雀重新进行适应度计算,对更新前后的适应度进行对比,保留更优的适应度继续进行更新;

(5) 判断迭代次数是否为 $\text{ite} r_{\max}$ 。若不是 $\text{ite} r_{\max}$,跳到(2)继续进行,直到为 $\text{ite} r_{\max}$ 为止,终止运行;

(6) 得到的最优的适应度 X_{best} 的位置就为 SVM 的参数 C 和 σ 。

3 基于CEEMDAN 多尺度熵-SSA-SVM 故障诊断流程

基于 CEEMDAN 多尺度熵-SSA-SVM 故障诊断流程如下：

- (1) 将振动信号随机分为训练样本和测试样本，利用 CEEMDAN 算法对其分解得到若干 IMF；
- (2) 提取敏感 IMF 并进行信号重组和多尺度熵的计算，然后进行归一化处理作为提取的特征向量；
- (3) 将得到的故障特征向量经 SSA-SVM 进行故障诊断。

其总体流程示意图如图 1 所示。

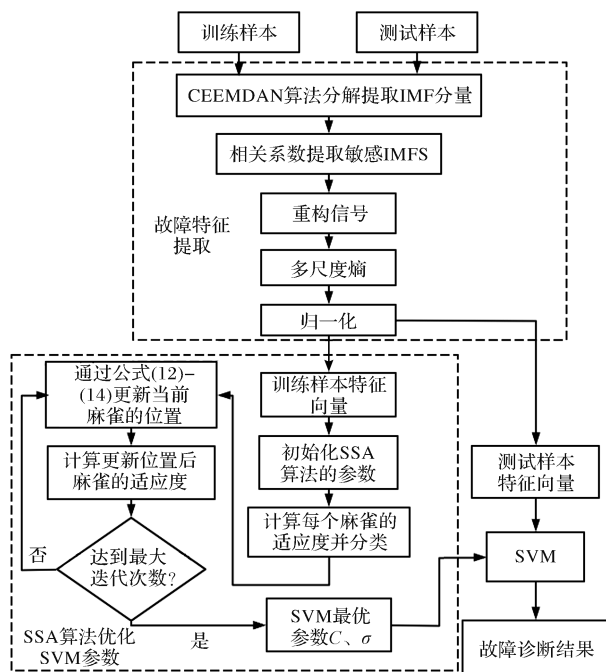


图 1 流程示意图

4 实验及结果分析

4.1 实验数据准备

为了检验上述过程的真实性，笔者采用美国西储大学公开的电机轴承数据进行研究^[13-15]。

被测轴承为 6205-2RSJEMSK，采用的数据为驱动端数据；其采样频率是 12 kHz，转速为 1 797 r/min。

此处分别取轴承正常状态、内圈故障、外圈故障和滚动体故障 4 种状态下的振动信号各 60 组，训练与测试样本各 30 组，每组样本的采样长度为 2 048 个点。

其中，多尺度熵中 $\tau = 10, r = 0.15S$ (S —样本的标准差)， $m = 2$ 。滚动轴承 4 种状态的标签为 1~4。

实验所采用的 4 种状态的时域图如图 2 所示。

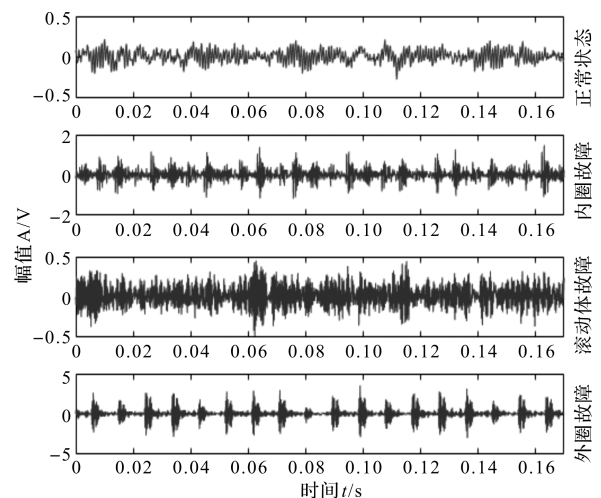


图 2 4 种状态的时域图

4.2 实验结果分析

该实验将 4 种状态下的数据通过 CEEMDAN 算法进行分解，共获得 12 个 IMF。

因为篇幅关系，此处只给出内圈故障状态下通过 CEEMDAN 算法提取的前 5 个 IMF，如图 3 所示。

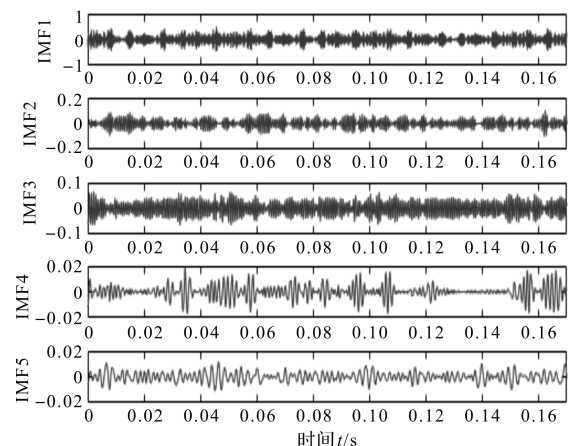


图 3 内圈故障信号 CEEMDAN 处理结果

由于若干 IMF 中存在一些无用信息，则选取相关系数超过 0.1 的 IMF 作为有用信息保留。

各个 IMF 相关系数大小如表 1 所示。

表 1 各个 IMFS 的相关系数值

IMF ₁	IMF ₂	IMF ₃	IMF ₄	IMF ₅	IMF ₆
0.947 5	0.327 7	0.303 0	0.031 6	0.004 0	0.001 6
IMF ₇	IMF ₈	IMF ₉	IMF ₁₀	IMF ₁₁	IMF ₁₂
0.002 3	0.003 0	0.001 1	0.000 7	0.000 1	0.002 0

根据表 1 可得，超过 0.1 的为 IMF₁、IMF₂、IMF₃，并对 IMF₁、IMF₂、IMF₃ 进行信号重组。

对重组信号求取尺度熵，并作为特征向量分别输入到 PSO-SVM、GA-SVM 以及 SSA-SVM 中，进行对比试验。

其中:每种 SVM 模型迭代次数为 100 次,种群数量为 20。

SSA-SVM 模型的单次分类准确率如图 4 所示。

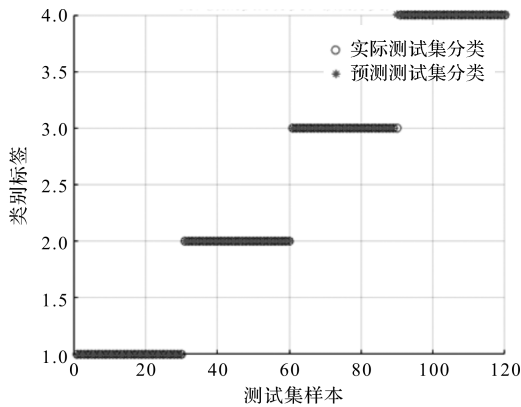


图 4 SSA-SVM 模型的单次分类准确率

PSO-SVM 模型的单次分类准确率如图 5 所示。

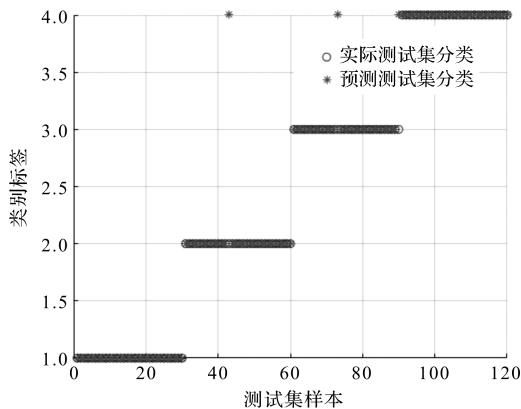


图 5 PSO-SVM 模型的单次分类准确率

GA-SVM 模型的单次分类准确率如图 6 所示。

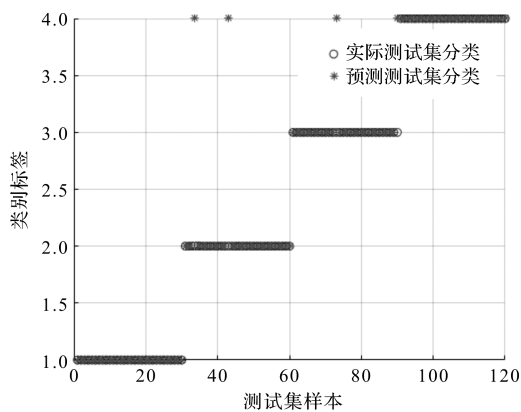


图 6 GA-SVM 模型的单次分类准确率

通过图(4~6)可知:

GA-SVM 模型与 PSO-SVM 模型对正常状态分类与外圈故障准确,但将内圈故障与滚动体故障不同程度的识别成了外圈故障。而 SSA-SVM 模型对正常状态、内圈故障与外圈故障均识别正确,仅在滚动体故障识别出现略微错误。

为了使该实验的实验结果更有说服力,笔者将 GA-SVM、PSO-SVM 以及 SSA-SVM 模型均进行 50 次实验。

实验环境为在 2.40 Hz 的英特尔 i5-9300H 处理器和 8 GB 内存电脑上的 Matlab2016a 中进行的。

此处笔者选用平均准确率和 3 种优化算法对 SVM 参数 C 与 σ 寻优的 50 次平均时间作为指标,对 3 种 SVM 模型的性能进行分析比较,可以很清楚地得出 3 种模型的差异。

GA-SVM、PSO-SVM 以及 SSA-SVM 模型的 50 次寻优时间如图 7 所示。

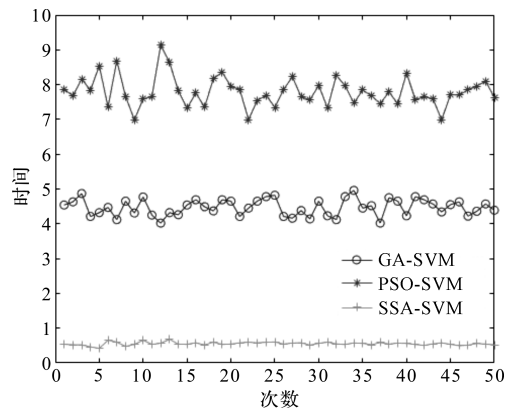


图 7 3 种模型 50 次寻优时间

3 种模型比较结果如表 2 所示。

表 2 3 种模型比较结果

分类方法	平均分类准确率/(%)	50 次平均寻优时间/s
GA-SVM	96.67	4.47
PSO-SVM	97.5	7.79
SSA-SVM	99.17	0.554

通过对图 7 与表 2 分析可得:

GA-SVM 模型的平均寻优时间 4.47 s 且寻优时间略不稳定,平均分类准确率为 96.67%,该模型的故障诊断效果最差;PSO-SVM 模型的寻优时间高于 GA-SVM 且寻优时间极差较大,但平均分类准确率略提高了 0.83%。SSA-SVM 模型与其他两种模型相比,其平均寻优时间最短,每次的寻优时间较为稳定以及平均分类准确率最高。

综上所述,SSA-SVM 模型分类准确度高,寻优速度快,在轴承故障诊断中表现出了更好的故障诊断性能。

5 结束语

针对支持向量机(SVM)应用在轴承故障分类时,传统的智能算法优化 SVM 的参数容易存在寻优速度慢、调节参数多,以及容易陷入局部最优值等问题,本

文提出了一种基于 CEEMDAN 多尺度熵与 SSA-SVM 相结合的故障诊断方法。

该故障诊断方法利用 CEEMDAN 多尺度熵对数据进行了故障特征提取,再利用 SSA-SVM 对提取的故障特征进行了分类;并对美国西储大学公开轴承故障数据进行了 PSO-SVM、GA-SVM 以及 SSA-SVM 模型对比实验。

实验结果表明:相比于 PSO-SVM 与 GA-SVM 模型,SSA-SVM 模型具有更快的收敛速度、更好的分类效果以及更短的寻优时间等优点,对研究其他方向有一定的参考价值。

参考文献 (References) :

- [1] 许 凯,许黎明,周大朝,等. 基于 Stockwell 变换的滚动轴承故障诊断方法[J]. 机械制造,2019,57(4):98-102.
- [2] 邢 昀,荣 剑. 常见不同模态信号分解方法探讨[J]. 现代计算机(专业版),2018(36):7-11.
- [3] 何庆飞,陈小虎,姚春江,等. 基于最小二乘支持向量分类机的齿轮泵故障诊断研究[J]. 流体机械,2019,47(8):32-36.
- [4] 张 鑫,赵建民,李海平,等. 基于 NIC-DWT-WOASVM 的齿轮箱混合故障诊断[J]. 振动与冲击,2020,39(11):146-151.
- [5] 刘 冲. 基于优化支持向量机的采煤机故障诊断技术[J]. 山东农业大学学报:自然科学版,2015,46(1):132-135.
- [6] 张小龙,张 氢,秦仙蓉,等. 基于 ITD 复杂度和 PSO-SVM 的滚动轴承故障诊断[J]. 振动与冲击,2016,35(24):102-107.
- [7] 赵 蕾,傅 攀,胡龙飞,等. FOA-WPT 降噪和 PSO-SVM 在滚动轴承故障诊断中的应用[J]. 机械与电子,2018,36(12):3-8.
- [8] 时培明,梁 凯,赵 娜,等. 基于分形维数和 GA-SVM 的风电机组齿轮箱轴承故障诊断[J]. 计量学报,2018,39(1):61-65.
- [9] 付 岩. 果蝇算法优化的广义回归神经网络滚动轴承故障预测[D]. 哈尔滨:哈尔滨理工大学机械动力工程,2018.
- [10] ZHUANG C C, YI H, ZHANG J. Application of EEMD multi-scale entropy and LSSV Minanalog circuit fault diagnosis[J]. *Microelectronics & Computer*, 2019, 36(10):78-82.
- [11] 林茂六,陈春雨. 基于傅立叶核与径向基核的支持向量机性能之比较[J]. 重庆邮电学院学报:自然科学版,2005(6):647-650.
- [12] 薛建凯. 一种新型的群智能优化技术的研究与应用[D]. 上海:东华大学信息科学与技术学院,2020.
- [13] 钟小倩,马文科,宋萌萌. 基于 GA 和 LM 组合优化 BP 神经网络的滚动轴承故障诊断方法[J]. 组合机床与自动化加工技术,2014(12):91-95.
- [14] 高立龙,井祥鹤,陈欣鹏,等. 基于改进 EEMD 方法的工程机械液压系统冲击振动信号研究[J]. 液压气动与密封,2020(3):6-9.
- [15] 梁治华,曹江涛,姬晓飞. 基于 EEMD 和 CS-SVM 的滚动轴承故障诊断研究[J]. 机电工程,2019,36(6):622-627.

[编辑:李 辉]

本文引用格式:

李 怡,李焕锋,刘自然. 基于 CEEMDAN 多尺度熵和 SSA-SVM 的滚动轴承故障诊断研究[J]. 机电工程,2021,38(5):599-604.

LI Yi, LI Huan-feng, LIU Zi-ran. Fault diagnosis of rolling bearing based on CEEMDAN multi-scale entropy and SSA-SVM[J]. *Journal of Mechanical & Electrical Engineering*, 2021, 38(5):599-604.

《机电工程》杂志: <http://www.meem.com.cn>