

基于 CEEMDAN-SVM 的高铁轴承故障诊断研究

杨 帅, 郝如江

(石家庄铁道大学 机械工程学院, 河北 石家庄 050043)

摘要: 高铁轴承在高铁运行中起着重要作用, 对其进行状态检测和故障诊断有着十分重要的作用和意义。总结出一种基于自适应辅助噪声的完备集合经验模态分解(CEEMDAN)和样本熵(SampEn)相结合的高铁轴承故障诊断方法。振动信号经过分解获得诸多的本征模态分量(IMF), 计算其样本熵特征参数来表征不同故障状态下的轴承信号的相关特征, 并构造相应的训练和测试样本数据, 而后将样本数据录入支持向量机(SVM)并配合灰狼优化算法(GWO)进行训练和测试, 完成轴承故障的分类和识别。实验结果表明, 此方法能够有效区分不同故障状态下的轴承振动信号。

关键词: 自适应噪声的完备集合经验模态分解; 样本熵; 灰狼优化算法; 支持向量机

中图分类号: TP277 **文献标志码:** A **文章编号:** 2095-0373(2021)02-0116-07

0 引言

高铁轴承是高铁的重要部件之一, 对于高铁安全稳定运行起着决定性作用^[1]。由于轴承运行状态的很多相关信息都可以在轴承振动信号中得以体现, 因此在轴承故障诊断方法中, 振动信号分析法是最常用也是最有效的方法。CEEMDAN 算法^[2]是一种信号的自适应分解算法, 能够有效改善其他分解算法中存在的模态混叠现象^[3-4], 并且重构误差较小, 完备性较好。在信号处理和分析方面, 特征提取是一种常用的技术手段, 近些年来, 排列熵、样本熵、奇异值等无量纲参数被引入机械设备振动信号的特征提取之中, 给机械故障诊断提供了更多的方法和思路。刘珍珍等^[5]利用多尺度排列熵算法, 成功地对轴承故障样本进行了特征分类; 别峰峰等^[6]利用奇异值作为特征提取方式, 对往复泵的振动信号样本进行了特征提取; 张宁等^[7]则利用样本熵作为特征提取手段, 进行了行星齿轮箱故障诊断研究。支持向量机^[8]是一种分类算法, 可以用于不同类型信号的分类和识别, 且其对小样本数据的支持性较好, 运行效率也比较高。狼群优化算法是一种新型的启发式参数寻优算法, 因为能够进行全局寻优, 且收敛速度快、寻优精度高而被广泛应用于各种领域, 周同乐等^[9]利用狼群算法来辅助生成无人机攻防策略数学模型, 成功提高攻防策略的作战效能并降低了行动代价; 蒋一锄^[10]则利用狼群算法来优化图书馆借阅数据聚类模型的初始参数值, 并成功提高了聚类精度。本次实验以高铁轴承不同故障状态下的振动信号作为实验对象, 并将振动信号分解后提取出的各阶分量的样本熵特征值作为样本和测试数据, 利用 SVM 进行高铁轴承故障的诊断。结果表明, 此方法能有效进行高铁轴承的故障分类, 准确率也较高。

1 相关原理

1.1 CEEMDAN 原理

CEEMDAN 算法是对 EMD 算法的一种改进, 通过添加自适应噪声来辅助分解, 使得分解结果能够在一定程度上改善模态混叠现象, 并且信号重构的完备性较好, 误差也比较小^[11], 下面简述 CEEMDAN 分解算法的原理。

(1) 对于任意的原始信号 $S(t)$, 对其添加噪声, 构造新的信号 $S(t) = S(t) + a_0 n^i(t)$, 其中, a_0 为所添加辅

收稿日期: 2020-12-16 责任编辑: 车轩玉 DOI: 10.13319/j.cnki.sjztdxxbzb.20200183

作者简介: 杨帅(1994—), 男, 硕士研究生, 研究方向为机械设备状态检测与故障诊断。E-mail: 1490124911@qq.com

杨帅, 郝如江. 基于 CEEMDAN-SVM 的高铁轴承故障诊断研究[J]. 石家庄铁道大学学报(自然科学版), 2021, 34(2): 116-122.

助噪声的幅值; i 为添加噪声构造新信号的次数,其取值为 $1, 2, 3, \dots, M$, 为正整数, M 是集合平均的总次数。而后将这些新得到的信号进行一阶段的 EMD 分解, 获得多个第一阶段的本征模态分量 $IMF_1^i(t)$, 对这些模态分量进行集合平均, 得到最终的第一阶本征模态分量 $IMF_1(t)$, 计算方法如式(1)所示, 将第一阶本征模态分量从原始信号中减去, 得到剩余信号 $S_1(t)$, 计算方法如式(2)所示。

$$IMF_1(t) = \frac{1}{M} \sum_{i=1}^M IMF_1^i(t) \quad (1)$$

$$S_1(t) = S(t) - IMF_1(t) \quad (2)$$

(2) 定义一种新的计算方式 $E_k(\cdot)$, 其含义为任意信号经过 EMD 分解后得到的第 k 个分量。在剩余信号 $S_1(t)$ 中加入噪声再次构造新的信号 $S_1(t) = S_1(t) + a_1 E_1[n^i(t)]$, 再对其进行一阶段的 EMD 分解, 将获得的多个第一阶段的本征模态分量再次集合平均, 得到最终的第二阶本征模态分量 $IMF_2(t)$, 计算方法如式(3)所示, 得到剩余信号 $S_2(t)$, 计算方法如式(4)所示。

$$IMF_2(t) = \frac{1}{M} \sum_{i=1}^M IMF_1^i(t) \quad (3)$$

$$S_2(t) = S_1(t) - IMF_2(t) \quad (4)$$

(3) 按照上文中的规则可以一直计算下去, 直到最后的剩余信号不满足再次分解所需要的条件, 分解结束, 信号经过分解后得到多个本征模态分量和剩余信号, 其结果的表达形式为

$$S(t) = \sum_{k=1}^K IMF_k(t) + R(t) \quad (5)$$

1.2 样本熵原理

样本熵^[12]是评价信号混乱程度的一种方式, 计算简单, 不依赖于数据长度, 且对于噪声等干扰条件的抵抗能力较强, 能有效表征不同信号在时间序列上的复杂性。下面简述其计算方式。

(1) 首先, 对于任意一个时间序列 $x = \{x_1, x_2, x_3, \dots, x_n\}$, 构造一个形式如式(6)的新的 M 维的 $X(i)$, 其中 i 的取值范围是 $\{1, 2, n-M+1\}$, n 是原始信号的数据长度, M 是预先设定的模式维数, 新数据的含义是原始信号从第 i 个点开始截取的 M 个连续的信号数值。

$$X_M(i) = \{x(i), x(i+1), \dots, x(i+M-1)\} \quad (6)$$

(2) 定义向量 $X_M(i)$ 和 $X_M(j)$ 之间的距离为 D , 其计算方式为 2 个不同向量之间相对应的元素差值的绝对值的最大者, 即 $D\{X_M(i), X_M(j)\} = \text{Max}\{|x(i+q) - x(j+q)|\}$, 其中, q 的取值范围为 $\{1, 2, \dots, M-1\}$ 。

(3) 根据事先确定的阈值 R , 统计距离 D 小于 R 的个体数目, 记为 B_i , 其中, $1 \leq i \leq N-M$, 计算出其在总个数 $N-M$ 中所占的比值, 记为 $B_i(R) = B_i / (N-M-1)$, 对于所有的 $B_i(R)$ 按照式(7)取其算术平均值, 记为 $B(R)$ 。

$$B(R) = \frac{1}{N-M} \sum_{i=1}^{N-M} B_i(R) \quad (7)$$

(4) 将模式维数设定为 $M+1$ 维, 再次重复以上步骤, 可以计算出 $M+1$ 维度下新的 $B(R)$ 值, 将其记为 $A(R)$, 那么对于含有有限个点的数据 x 而言, 其样本熵便定义为 $\text{SampEn}(x) = -\ln[A(R)/B(R)]$ 。样本熵的计算需要设定模式维度 M 和阈值 R 2 个参数, 一般情况下, M 的取值为 1 或者 2, 而阈值的取值可以落在 0.1~0.2 倍原始信号标准差这个区间范围内, 在这样的前提下, 计算结果比较准确合理。

1.3 狼群优化算法

灰狼优化算法是模拟自然界中狼群狩猎捕食行为的一种启发式优化算法, 它按照自然界中狼群等级将种群分为 $\alpha, \beta, \delta, \omega$ 4 种等级, 其中 α 狼的等级最高, 又被称为头狼, 它负责统领并协调其他狼, 为种群的领导者, 其他 3 种狼受头狼的指挥, 整个寻优计算过程就是狼群之间相互配合不断接近直到最终捕捉猎物的自然行为。它的主要过程如下:

(1) 包围。这一步代表的是狼群搜索猎物并接近猎物的行为, 数据的更新规则如式(8)所示, 其中的 t 表示当前的迭代次数, A 和 C 是协同系数向量, X_p 代表猎物位置, $X(t)$ 表示灰狼位置, 2 个 r 是 $[0, 1]$ 之间

的随机向量。

$$\begin{cases} D=C \circ X_p(t)-X(t) \\ X(t+1)=X_p-A \circ D \\ A=2\alpha \circ r_1-a \\ C=2r_2 \end{cases} \quad (8)$$

(2) 狩猎。这一步模型假定 α, β, δ 3 匹狼识别潜在猎物的能力较强, 因此可以保留最优的 3 匹狼的位置信息来更新其他搜索代理的位置, 更新规则如式(9)所示, 其中 $X_\alpha, X_\beta, X_\delta$ 表示当前 3 匹狼的位置信息, X 表示灰狼的位置信息, D 表示当前候选灰狼和最优 3 匹狼之间的距离。

$$\begin{cases} D_\alpha=C_1 \circ X_\alpha-X & X_1=X_\alpha-A_1 \circ D_\alpha \\ D_\beta=C_2 \circ X_\beta-X & X_2=X_\beta-A_2 \circ D_\beta \\ D_\delta=C_3 \circ X_\delta-X & X_3=X_\delta-A_3 \circ D_\delta \\ X(t+1)=(X_1+X_2+X_3)/3 \end{cases} \quad (9)$$

(3) 攻击。这一步便是狼群捕食的最后一步, 也就是算法得到最优解的过程, 其中 α 头狼所代表的的结果便是寻找的最优解结果。

2 实验方案设计

本次实验采用东方所振动信号采集仪(INV3018A)以及 DASP 软件来采集铁路轴承综合性能实验台的数据, 其中采样频率设定为 51 200 Hz, 驱动电机转速为 1 200 r/min, 不添加外部荷载。实验所用的轴承为实验室的高铁轮对轴承, 其外径长度为 240 mm, 内径长度为 130 mm, 节径长度为 185 mm, 采用人为加工缺陷来模拟不同部位的轴承故障, 故障尺寸长宽深分别为 5、1、0.7 mm。分别采集正常轴承、内圈故障轴承和外圈故障轴承 3 种状态下的高铁轴承振动信号, 传感器为振动加速度传感器, 型号为 PCB35A26。本次实验一共采用了 360 组数据, 其中的 150 组用作训练数据, 其余 210 组用作测试数据。具体的实验方案及流程设计如下:

(1) 采集不同类型的振动信号, 并将每一个信号样本作为原始数据输入 CEEMDAN 分解算法进行自适应分解, 分解过程中辅助白噪声的幅值设定为原始信号标准差的 0.2 倍, 保存分解结果。

(2) 计算每一组分解后获得的本征模态分量的样本熵特征值, 其中模式维度设定为 2, 并将结果作为 SVM 模型的训练样本数据和测试样本数据。

(3) 将训练数据和测试数据进行预处理, 例如归一化操作等, 利用训练数据创建 SVM 的模型, 其中的参数惩罚因子 C 和核函数参数 G 采用狼群算法进行优化, 创建完模型后利用测试数据来测试模型, 利用训练数据和测试数据进行 SVM 模型的故障诊断实验, 记录分类效果, 统计分类的正确率, 评价故障诊断模型的效果。模型的创建、训练和识别的流程如图 1 所示。

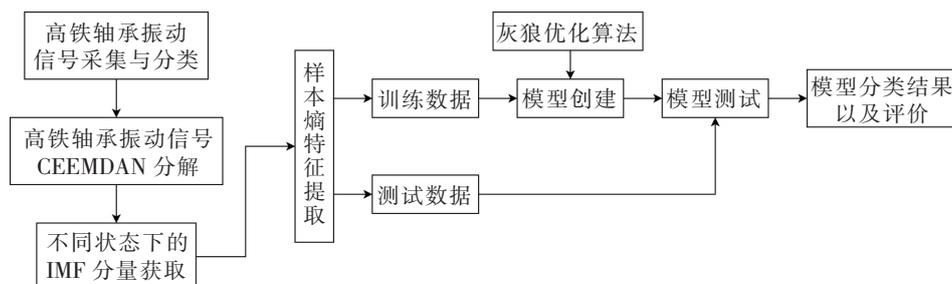


图 1 故障诊断实验流程图

3 高铁轴承故障诊断实验

(1) 启动铁路轴承测试实验台的设备和 DASP 采集系统, 从实验台中分别采集正常轴承、内圈故障轴

承、外圈故障轴承 3 种不同状态下的轴承振动信号,典型的 3 种不同状态的振动信号如图 2 所示。从 3 种信号的时域波形图中无法有效分辨 3 种信号的差异,需要提取更加深层次的信息。

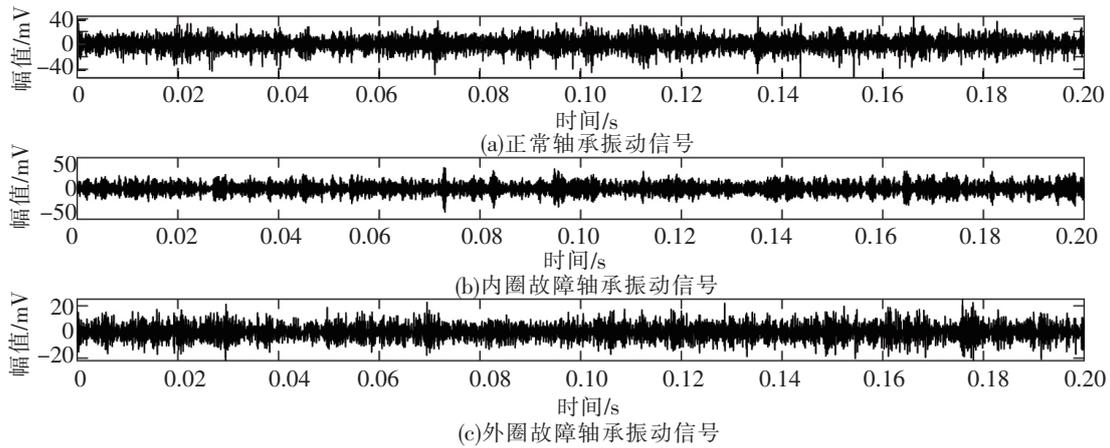


图 2 轴承振动信号时域波形图

(2)将不同状态的振动信号进行 CEEMDAN 自适应分解,从中挑选任意 3 种不同信号分解后获得的分量的时域波形图进行展示,限于篇幅,只展示了其前 5 阶段的本征模态分量,其分解结果如图 3~图 5 所示,观察信号模态分量的时域图仍然难以明确分辨 3 种信号分量之间的具体差异。

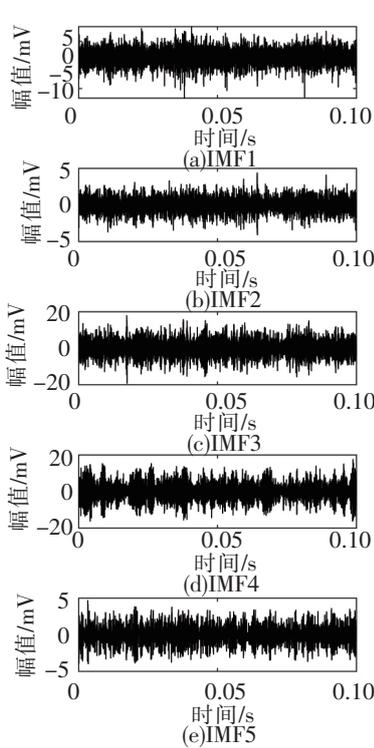


图 3 正常轴承信号 CEEMDAN 分解图

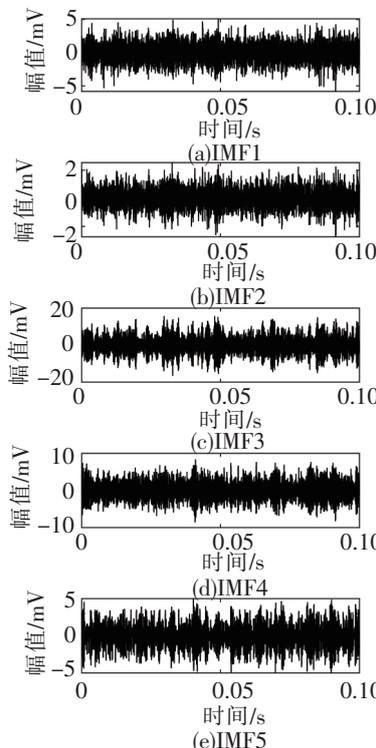


图 4 内圈故障信号 CEEMDAN 分解图

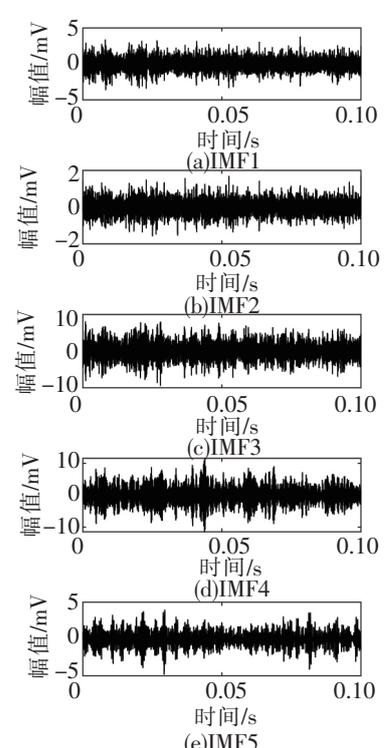


图 5 外圈故障信号 CEEMDAN 分解图

(3)分别计算任意 3 组数据的样本熵特征参数值,其结果如图 6 所示。从图 6 中不难发现,在图中黑色竖线标记的左侧,三者的特征值的数值都在 0.5 以上,并且三者有着较为明显的差距,而在竖线右侧,特征值小于 0.5,且三者的特征值差别不大,无法有效代表不同类型信号之间的差异,因此,为使样本的特征值数据有最好的识别和区分度,本次实验选择 3 种信号分解后提取出的前 5 阶样本熵特征参数值数据作为 SVM 模型的训练和识别数据。

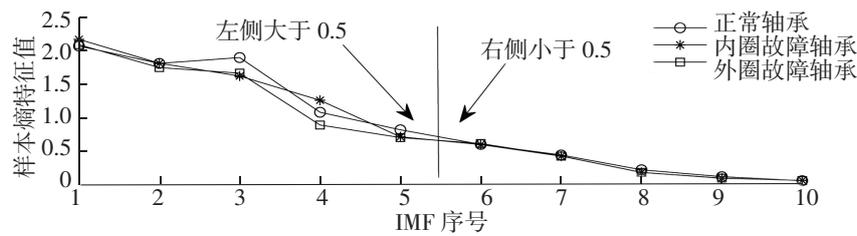


图 6 样本熵特征参数值对比图

(4)对所有采集到的不同故障状态下的轴承振动信号分别进行以上的 CEEMDAN 分解和计算特征值的操作,其部分特征值如表 1 的内容所示。将得到的 150 组训练样本和 210 组测试样本先后输入支持向量机 SVM 中,训练模型时,正常信号的标签被设定为 1,内圈故障轴承信号的标签设定为 2,外圈故障轴承信号的标签设定为 3。训练所采用的参数 C 和 G 利用灰狼优化算法来确定,在灰狼算法的计算下,可以得到优化后的参数值。在实验中,狼群的种群规模设定为 20,最大的进化迭代次数设定为 25,优化参数数量设定为 2 个,取值范围限定于 $[0.01, 100]$ 区间范围内,经过计算,优化后的惩罚因子 C 的值为 70.93,核函数参量 G 的值为 9.91。将以上训练数据和优化结果录入 SVM 进行模型的创建,在创建完模型后,采用相同的标签设定规则对测试数据进行标签设定,而后再将其输入 SVM 进行分类,SVM 的识别结果如图 7 所示。采用相同的实验数据,利用未优化参数的 SVM 模型进行了对比实验,标签设定与 SVM 模型设定方式相同,其中 2 个参数仍在 $[0, 100]$ 范围内取值。本次实验进行了 2 次人为取值,取值分别为 $[80, 60]$ 和 $[90, 80]$,采用与上文相同的实验方法进行故障识别,其识别结果如图 8 所示。

表 1 部分实验数据表

样本类型	训练数据					测试数据				
	IMF1	IMF2	IMF3	IMF4	IMF5	IMF1	IMF2	IMF3	IMF4	IMF5
正常信号	2.013 9	1.823 4	1.913 0	1.103 8	0.786 1	1.995 6	1.869 8	1.938 3	1.125 9	0.802 1
正常信号	2.004 5	1.882 3	1.978 1	1.099 1	0.797 5	2.034 6	1.925 9	1.995 9	1.129 7	0.792 5
正常信号	2.034 1	1.864 3	1.908 0	1.149 9	0.794 8	2.029 2	1.848 7	1.896 8	1.173 1	0.764 5
内圈故障信号	2.079 2	1.812 4	1.629 7	1.269 0	0.679 8	2.139 3	1.857 8	1.747 8	1.310 3	0.714 4
内圈故障信号	2.130 1	1.834 8	1.685 7	1.261 8	0.703 3	2.138 6	1.840 7	1.714 5	1.270 2	0.692 8
内圈故障信号	2.107 7	1.816 5	1.644 3	1.315 4	0.697 8	2.128 2	1.866 8	1.674 3	1.266 4	0.685 6
外圈故障信号	1.915 0	1.750 7	1.572 6	0.861 0	0.714 9	2.082 7	1.797 9	1.658 9	0.881 9	0.715 9
外圈故障信号	2.096 3	1.832 6	1.682 0	0.878 2	0.723 3	1.925 5	1.751 1	1.695 7	0.907 2	0.699 5
外圈故障信号	2.051 4	1.758 6	1.622 5	0.858 3	0.698 5	1.958 4	1.740 8	1.585 8	0.883 2	0.693 8

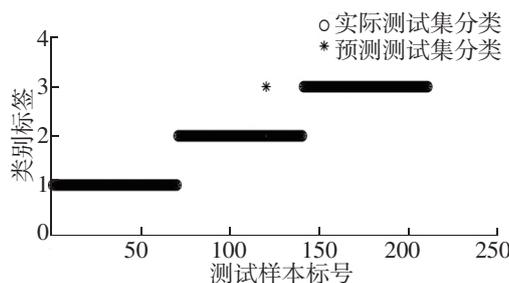


图 7 优化后的 GWOSVM 计算结果图

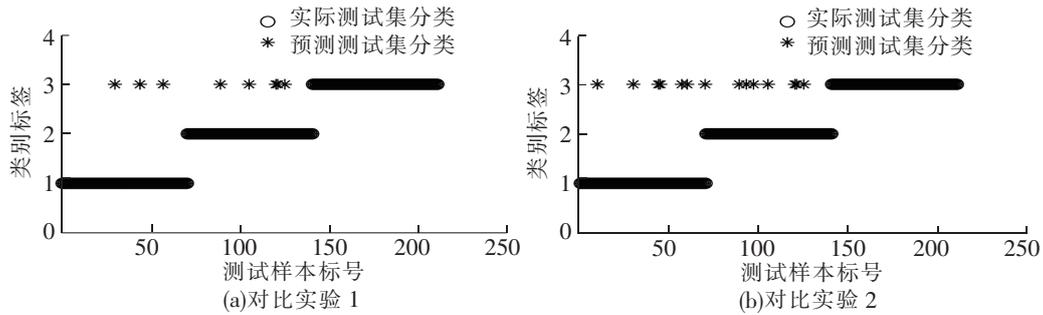


图 8 未优化的 SVM 计算结果图

(5)具体的计算结果数据汇总于表格 2 中。由表 2 中数据可以看出,在 GWO 优化的 SVM 模型下,其中正常信号和外圈故障信号全部识别正确,内圈故障信号只有一个样本被错误识别成了外圈故障样本,总体的正确率可以到达比较高的 99.52%。在第一次未优化参数的实验中,正常信号数据中有 3 个被误判为外圈故障信号数据,内圈故障信号数据中有 5 个被误判为外圈故障信号数据,整体的正确率为 96.19%;在第二次未优化参数的实验中,正常信号数据中有 7 个被误判为外圈故障信号数据,内圈故障信号数据中有 7 个被误判为外圈故障信号数据,整体的正确率为 93.33%,2 次模型分类的准确率均比参数优化后的模型分类的准确率要低。由以上数据可以得出,GWO 优化后的 SVM 取得了较高的识别正确率,并且由于 GWO 算法可以根据实际的数据来自动确定参数,无需人工设定,让实验流程更加科学合理;未优化参数的 SVM 模型的识别正确率较低,说明人为选定参数可能会对识别结果产生影响,也增大了故障诊断过程中潜在的人为经验误差。

表 2 分类识别结果

模型类别	正常	内圈故障	外圈故障	正确率/%
未优化的 SVM 实验 1	67/70	65/70	70/70	96.19
未优化的 SVM 实验 2	63/70	63/70	70/70	93.33
优化后 GWOSVM 实验	70/70	69/70	70/70	99.52

4 结论

总结出一种基于 CEEMDAN 和 SampEn 特征参数和 GWOSVM 的高铁轴承故障诊断方法,并利用实际采集的高铁轴承振动信号进行了模型可行性测试实验。实验结果表明,对于高铁轴承不同故障类型振动信号的识别和分类,此模型是有效的,且能取得较好的分类结果,GWO 算法的加入让参数的设定更加科学,避免了人为操作带来的误差影响,是一种科学合理的高铁轴承故障诊断方案。

参 考 文 献

[1]曹国晋. 高铁轴承疲劳特性分析及可靠性寿命预测[D]. 上海:上海应用技术大学,2018.
 [2]韩宏志,唐振浩. 基于 CEEMDAN 与回声状态网络的风速预测方法[J]. 电力系统保护与控制,2020,48(12):90-96.
 [3]朱敏,段志善,郭宝良,等. CEEMDAN 结合 LMS 算法在轴承信号降噪中的应用[J]. 噪声与振动控制,2018,38(2):144-149.
 [4]张建文,刘洋,张大朋,等. 基于 CEEMDAN 与小波自适应阈值的去噪新方法[J]. 电测与仪表,2018,55(10):14-18,33.
 [5]刘珍珍,陈志雄,陈进. 基于 CEEMDAN-MPE 的滚动轴承故障识别[J]. 组合机床与自动化加工技术,2018(11):57-61.
 [6]别锋锋,徐鹏青,裴峻峰,等. 基于 CEEMDAN 与奇异值分解的往复机械故障诊断方法研究[J]. 噪声与振动控制,2018,38(4):180-185,207.
 [7]张宁,魏秀业,徐晋宏. 基于 LMD 样本熵与 ELM 的行星齿轮箱故障诊断[J]. 机械传动,2020,44(4):152-157.
 [8]陈祖成,王硕禾,赵超策,等. 基于 GA-BP 和小波-SVM 算法的风电场短期功率预测[J]. 石家庄铁道大学学报(自然科学版),2020,33(1):104-109.

- [9]周同乐,陈谋,朱荣刚,等.基于狼群算法的多无人机协同多目标攻防满意决策方法[J].指挥与控制学报,2020,6(3):251-256.
- [10]蒋一锄.基于灰狼优化聚类算法的读者行为分析[J].吉首大学学报(自然科学版),2020,41(4):38-43.
- [11]金成功.基于 CEEMDAN 能量熵和马氏距离的齿轮箱轴承故障诊断方法[J].机床与液压,2020,48(16):218-223.
- [12]陈鹏,赵小强,朱奇先.基于多尺度排列熵和改进多分类相关向量机的滚动轴承故障诊断方法[J].电子测量与仪器学报,2020,34(2):20-28.

Research on Bearing Fault Diagnosis of High-speed Railway Based on CEEMDAN -SVM

Yang Shuai, Hao Rujiang

(School of Mechanical Engineering, Shijiazhuang Tiedao University, Shijiazhuang 050043, China)

Abstract: High speed rail bearing plays an important part in the operation of high-speed railway, so it is very important to carry out state detection and fault diagnosis. In this paper, a bearing fault diagnosis method based on complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) and Sample Entropy (SampEn) has been proposed. The vibration signal is decomposed into many intrinsic mode functions (IMF), the Sample Entropy parameters are calculated to represent the relevant characteristics of different signals, and the corresponding train and test data samples are constructed. After that, the data samples are input into support vector machine (SVM) and trained and tested with gray wolf optimization algorithm (GWO) to complete the classification and identification of bearing faults. The results show that this method can effectively distinguish the bearing vibration signals under different fault states.

Key words: CEEMDAN; sample entropy; GWO algorithm; SVM