

一种基于改进极端梯度提升决策树的轴承故障诊断方法*

张天瑞, 赵伟博, 周福强, 朱芷仪, 宋雨儒, 贾泽轩
(沈阳大学 机械工程学院, 沈阳 110044)

摘要:针对传统故障诊断模型面向海量故障数据时诊断准确度低的问题,首先,提出了一种局部均值分解与固定点算法联合降噪方法,以消除轴承振动信号中的噪声;其次,为了避免原始信号中敏感特征难以提取的问题,提出了一种基于核主成分分析的降维方法;再次,构建了一种基于改进极端梯度提升决策树的故障诊断模型,采用GS-PSO算法优化SVM性能,进而运用改进极端梯度提升决策树思想修正分类模型的残差以提升模型分类精度,应用Spark-大数据平台,通过并行处理技术进行科学计算;最后,采用CWRU提供的滚动轴承数据进行训练与仿真,证明构建的模型能实现对不同类型滚动轴承的识别诊断,并保证诊断结果的准确率。通过对4种不同故障诊断模型的对比分析,表明本文模型具有可行性和优越性。

关键词:滚动轴承;故障诊断;大数据;特征提取;XGBoost

中图分类号:O221.2;TP391.9

文献标志码:A

文章编号:1672-6693(2023)05-0030-10

随着人工智能、大数据技术的开发和应用,机器故障诊断作为一个综合性、交叉性研究领域也得到了快速发展。滚动轴承在机器零部件中的应用十分广泛,它的安全状态关系到机组整体是否能稳定运行。由于滚动轴承实际使用环境恶劣,故障频发,导致它成为现代工业中最易受损的元器件之一^[1]。因此,对滚动轴承的工作运行状态进行预先诊断非常有必要。通过建立准确、高效、适应性强的故障诊断模型,以应对不同生产实际中的各种复杂环境,对生产实践具有十分重要的指导作用。

近些年来,国内外学者对滚动轴承故障诊断方面的研究成果层出不穷。运用有效的方法、算法与诊断模型相结合对轴承故障诊断具有重要意义。Han等人^[2]考虑多尺度符号动态信息熵,并利用局部均值分解(local mean decomposition, LMD)方法处理轴承信号。车昱娇等人^[3]考虑Dropout策略的长短时记忆神经网络,并基于核主成分分析(key principal component analysis, KPCA)降维方法进行轴承特征提取。刘甜甜等人^[4-5]基于改进的PSO算法和LSSVM预测磁轴承转子的位置,取得了良好效果。

在各类大数据处理技术中,深度学习和数据挖掘等是其中比较重要的技术^[6]。目前在面对大数据中的各类重要数据集进行处理时,存在分类时间长、构造模型复杂等问题。Paper等人^[7]考虑了Spark平台的基本架构和生态,从数据预处理和模块计算2个方面进行了优化,并基于Spark平台实现了数据并行计算。李鑫灵等人^[8]构建了PSO-SVM突出预警模型,在此基础上融合Spark大数据平台开发了掘进工作面突出预警系统。总的来看,基于Spark计算框架进行SVM参数算法的并行性研究和实现,更适合大数据环境下机器学习方面和故障诊断的研究。

极端梯度提升决策树(extreme gradient boosting decision tree, XGBoost)是Chen等人^[9-10]首先提出。XGBoost的应用十分广泛,尤其在面向故障诊断的问题,XGBoost的优势主要体现在多个分类器集成上,在针对特征集合的选取、参数优化、故障识别和模型评分等方面具有优势。

1 基本理论

1.1 LMD+FPA联合降噪过程

LMD方法是基于时域-频域的自适应信号处理分析方法^[11],本质是将原始信号分解为若干个分量和残余

* 收稿日期:2022-05-07 修回日期:2022-11-11 网络出版时间:2023-06-26T10:49

资助项目:中央引导地方科技发展资金计划项目(No. 2021JH6-10500149);辽宁省自然科学基金项目(No. 20180551001)

第一作者简介:张天瑞,男,副教授,博士,研究方向为运筹与管理,E-mail:trzhang@syu.edu.cn;通信作者:赵伟博,男,E-mail:332828953@qq.com

网络出版地址:https://link.cnki.net/urlid/50.1165.n.20230621.1723.002

值。盲源欠定是指原始信号中的信号来源不同,而混合的不同源信号数量超过观测信号的数量,固定点(fixed-point algorithm, FPA)算法是目前比较流行的一种盲源分离算法^[12],在应对数据迭代处理过程中计算数据量大的情况,具有更快的处理速度。为了克服盲源欠定问题,本文采用 LMD 方法和 FPA 算法联合,以实现盲源分离的过程。FPA 算法流程如图 1 所示,其中 X 表示观测数据, Z 表示白化处理数据, W_i 表示初始权向量。

本文使用基于负熵最大的联合降噪方法,即 LMD+FPA 降噪方法,基本流程如图 2 所示。主要过程为:1) 借助 LMD 信号处理过程获得多个分量;2) 计算实际信号同分量之间的关系,以相关系数确定评定标准关系程度最大的信号进行重构;3) 以虚拟观测信号和原始信号构建二维输入矩阵,运用 FPA 算法予以分析分解达成滤波效果。

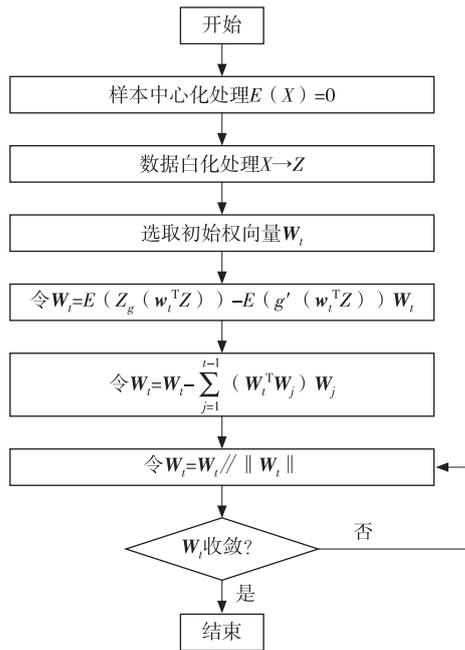


图 1 FPA 算法流程

Fig. 1 FPA algorithm flow

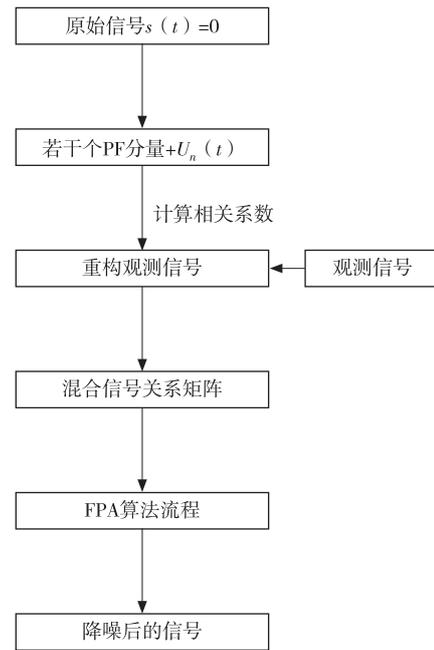


图 2 LMD+FPA 联合降噪流程

Fig. 2 LMD+FPA combined noise reduction flow

1.2 主成分计算过程与核函数的选取

主成分计算(kernel principal component analysis, KPCA)方法又称为核主分量分析^[13],采用非线性的方式进行样本集的空间变换,通过核函数的映射过程达到数据降维的目的。由于 KPCA 具有非线性分类方式的特点,因此对非线性问题的解决效果更好。KPCA 方法分析的主要计算过程如下。

首先, KPCA 方法涉及空间变换,数据由低维空间变换到高维空间的同时,计算高维空间 F 的协方差矩阵 $C = \frac{1}{m} \sum_{j=1}^m \Phi(x_j) \Phi(x_j)^T$, 其中 $\Phi(x_j)$ 为非线性函数。 C 的特征向量 V 和特征值 λ 满足:

$$\lambda V - CV = 0. \tag{1}$$

再将 $\Phi(x_j)$ 代入式(1),得到 $\lambda \Phi(x_j) V - \Phi(x_j) CV = 0$ 。用高维特征空间 F 的样本点 $\Phi(x_i)$ 来线性描述特征向量 V , 即 $V = \sum_{i=1}^m a_i \Phi(x_i)$ 。其中: x_i 表示低维空间样本, m 表示低维输入空间样本个数。

最后,输入核函数 $K_{ij} = K(x_i, x_j) = \Phi(x_i) \Phi(x_j)$, 计算可得 $m\lambda \alpha - K\alpha = 0$ 。其中: K 是核矩阵, α 是 K 的特征向量。

综上,数据样本在高维特征空间 F 中的主元成分 $\Phi(x)$ 上的投影为:

$$V\Phi(x) = \sum_{i=1}^m \alpha_i \Phi(x_i) \Phi(x) = \sum_{i=1}^m \alpha_i K(x_i, x).$$

本文采用 KPCA 方法将轴承振动信号的时域和频域特征进行融合处理,目的是降低数据中特征向量的维度。在此过程中,主成分 N 的个数和原始信息之间密切相关,本文选择 N 的原则为:计算累计的方差贡献率

r_{CPV} , 再通过 r_{CPV} 的大小判断主元的数量。其中, 协方差矩阵 C 的特征值的排序方式以从大到小的原则进行排列, 以此最终确定主元。第 k 个主元的方差贡献率为: $\xi_k = \frac{\lambda_k}{\sum_{k=1}^m \lambda_k}$, λ_k 表示第 k 个特征值。则累计前 s 个主元的

$$r_{CPV} = \left(\frac{\sum_{k=1}^s \lambda_k}{\sum_{k=1}^m \lambda_k} \right) \times 100\% (s < m)。$$

通常 r_{CPV} 的大小和结果准确率的大小相关, r_{CPV} 的值越大则计算结果越好。

KPCA 方法中的核函数的选取也十分重要。目前, 在工程领域应用较为广泛的核函数主要类型有: 1) 线性核函数 $K(x_i, x) = \langle x_i, x \rangle$; 2) d 阶多项式核函数 $K(x_i, x) = [\langle x_i, x \rangle + 1]^d$; 3) RBF 核函数 $K(x_i, x) = e^{-\frac{\|x_i - x\|^2}{2\sigma^2}}$; 4) 多层感知核函数 $K(x_i, x) = \tanh[v \langle x_i, x \rangle + c]$ 。

由于径向基(radial basis function, RBF)核函数在工程实际应用中较多^[14], 在处理数据样本时的精度、速度均优于其他核函数, 所以本文使用 RBF 核函数。

2 滚动轴承故障诊断模型

为了提高滚动轴承故障诊断模型的诊断准确率, 构建基于支持向量机(support vector machine, SVM)改进极端梯度提升决策树的滚动轴承故障诊断模型。

2.1 基于 GS-PSO 的 SVM 分类模型

误差惩罚参数和核函数参数对 SVM 分类效果至关重要, 为了保证 SVM 模型的良好性能, 本文通过网格搜索算法(grid search, GS)和粒子群算法(partical swarm algorithm, PSO)联合寻优确定 SVM 的输入参数, 提升诊断模型的效果。联合寻优过程既能解决 PSO 单独寻优易陷入局部最优的问题, 又能提高寻优的准确率和分类速度。图 3 显示了 GS-PSO 对 SVM 输入参数进行寻优的流程。

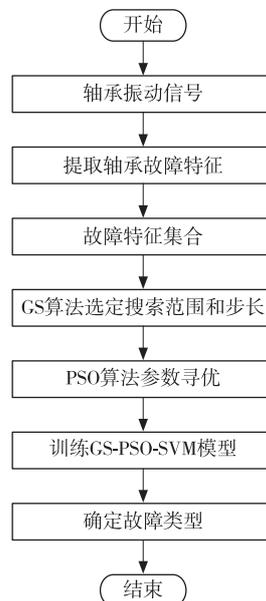


图 3 GS-PSO 流程

Fig. 3 GS-PSO flow

2.2 基于改进 XGBoost 的轴承故障诊断模型

XGBoost 作为提升准确度的依据, 极大地提升了模型分类精度, 这体现在对泰勒二阶展开的计算方面, 并在目标函数之外加入惩罚项以控制分类模型的性能^[15]。XGBoost 中一部分为求解损失函数, 用以计算每种特征集合最终的得分; 另一部分为正则惩罚项, 包括叶子节点的数量和叶子节点的分数。针对滚动轴承特征集合中存在特征类别不均衡的问题, 借助设计带有学习敏感度的新损失函数以修正特征不平衡的现象, 其中: XGBoost 的

目标函数和正则项函数分别为:

$$O_{bj} = \sum_{i=1}^n (y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k), \Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2.$$

式中: y_i 为 i 轮真实值; \hat{y}_i 为 i 轮预测值; f_k 为第 k 棵 CART 树预测输出; Ω 为正则惩罚项; $i = 1, 2, \dots, n$, n 为求解次数; $k = 1, 2, \dots, K$, K 为 CART 树数量; λ 为叶子节点数; γ 为正则项系数; ω 为叶子向量(叶子节点分数); $j = 1, 2, \dots, T$, T 为叶子节点数量。

针对 XGBoost 做出如下改进:

$$L(Y, (P(Y | X))) = -\ln P(Y | X) = -\frac{1}{n} \left[\sum_{i=1}^n y_i (1 + e^{\bar{y}_i}) \right],$$

$$L_{NEW}(Y, P(Y | X)) = -\frac{1}{n} \left(\sum_{i=1}^n y_i \ln(1 + e^{-\bar{y}_i}) \times S^{\frac{1}{y_i+1}} \right). \quad (2)$$

式(2)为改进的目标函数。上述公式中: X, Y 表示不同的样本集合, $P(Y|X)$ 表示概率值, \bar{y}_i 表示 i 轮真实平均值, S 表示学习权重。

为解决特征集中包含多维特征和残余信息而导致的故障模型识别准确率低的问题,本文基于改进 XGBoost 设计了 2 种优化特征集选取的流程。通过每次计算得分来不断修正模型预测误差,最终达到提升精度的目的。利用改进 XGBoost 计算每个特征集合的分数值,并以增益值、覆盖率和特征相关程度作为评估指标对特征集合的重要性进行衡量,具体如下。

1) 重要性指数为 $\alpha = \frac{c+g}{2}$, 其中: c 表示增益值, g 表示覆盖率;

2) 各特征间的相关程度为 $R = \frac{\text{cov}(x_i, x_j)}{\sqrt{\text{var}(x_i)\text{var}(x_j)}}$ 。

基于改进 XGBoost,以优化特征集合的流程采用 2 种特征选取过程,序列向前和浮动向后原则通过计算得分进行集合的筛选,如图 4 和图 5 所示。

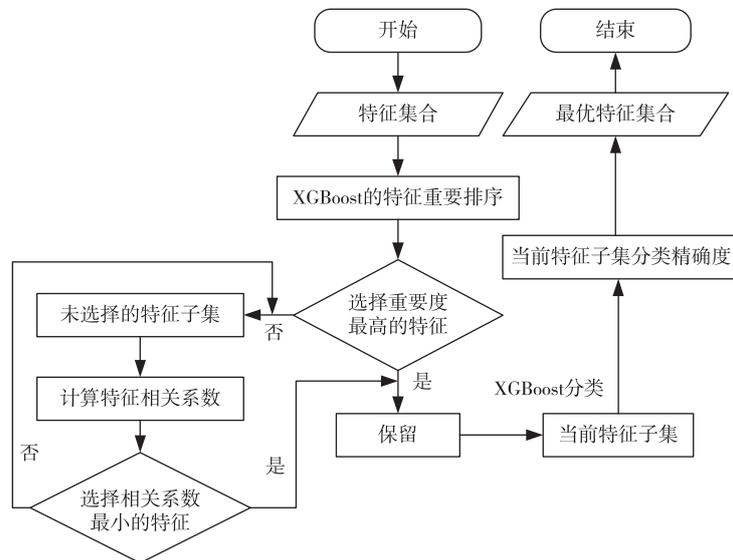


图 4 改进 XGBoost-序列向前分类流程

Fig. 4 Improved XGBoost sequence forward classification flow

基于改进 XGBoost 的分类流程中,2 种不同原则的操作步骤为:

1) 序列向前添加。a) 根据之前获取的轴承特征集合,利用改进 XGboost 对所有特征集合进行分类;b) 借助数据集合对改进 XGBoost 进行训练,训练之后的改进 XGBoost 故障诊断模型需要计算特征值的重要性,根据特征重要性的大小进行特征集合的排序。

2) 浮动向后剔除。a) 改进 XGBoost 选定最优特征集之后,再通过指标评估的过程剔除最差的特征集合,直到未选择的特征集中无法选择下一个最优特征;b) 计算重要性排序最差的特征和未选特征的相关系数,删除相

关系数绝对值最大的特征;c) 借助改进 XGBoost 分类器计算所有子集的特征得分,得分最高的即为最优特征集合,若有 2 个或 2 个以上的特征子集分类精度相同且最高,则选择特征数量较少的作为最优特征集合。

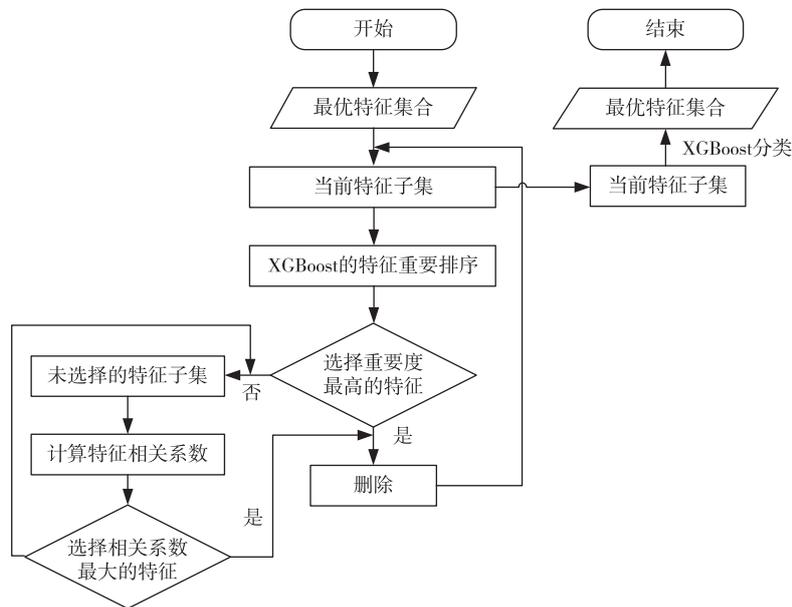


图 5 改进 XGBoost-浮动向后分类流程

Fig. 5 Improved XGBoost floating backward classification flow

3 仿真分析

为了证实本文模型的有效性和在实际生产中发挥的作用,对轴承故障诊断模型进行了仿真试验以及对比分析。针对滚动轴承的一般性故障问题,采用美国凯斯西储大学(Case western reserve university, CWRU)电气工程实验室提供的轴承数据集^[16],该实验平台通过改变滚动轴承直径、检测的位置、马达的负载和转速、取样频率等方式获得多种有效数据。CWRU 的实验平台如图 6 所示。

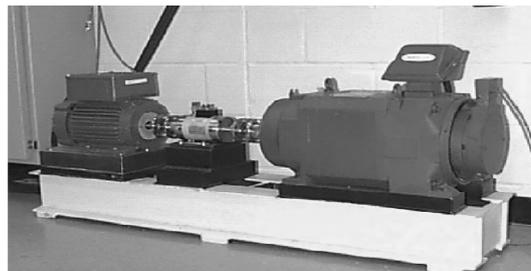


图 6 CWRU 轴承故障试验台

Fig. 6 CWRU bearing fault test-bed

图 6 中的 CWRU 的平台由 1 个三相感应电动机、轴和测力计组成,测试轴承为 SKF-6205-2R5 深沟球轴承。在滚动轴承的故障数据中,故障直径包括 177.8、355.6 和 533.4 μm ,数据长度 2^{10} ;轴承状态包含正常状态和 3 种不同的异常状态,涉及到除保持架外的全部故障问题。由于保持架的损坏问题较为少见,故本文不进行分析和验证。实验数据集合分为 1 和 2,如表 1 所示。

3.1 基于改进 XGBoost 的轴承故障诊断模型的仿真实验

基于 Python 和 Matlab 对改进 XGBoost 的滚动轴承故障诊断模型仿真。其中,Python 版本为 3.10,Spark 大数据平台用于处理数据,Anaconda 中的 Jupyter 部分用于表现可视化数据的处理结果。改进 XGBoost 模型中,学习权重 S 的大小是影响模型最终分类准确度的关键,为了找到最优值,首先对 S 值进行寻优计算,如图 7 所示。以最佳适应度和平均适应度来观察改进 XGBoost 诊断的结果,从图 7 可以看出,经过参数寻优之后的 S 最优值为 2.7。

GBDT 和 XGBoost 在本文中以准确率和 AUC 值作为评价标准,表 2 为二者结果的对比分析。

表 1 两组实验数据

Tab. 1 Two groups experimental data

数据	训练样本个数	测试样本个数	轴承状态	故障尺寸/ μm	标签
1	40	20	正常	0	0
	60	30	内圈故障	177.8、355.6、533.4	1
	48	24	滚动体故障	177.8	2
	52	26	外圈故障	177.8、355.6、533.4	3
2	20	20	正常	0	0
	20	20	内圈故障	177.8、355.6、533.4	1
	20	20	滚动体故障	177.8	2
	20	20	外圈故障	177.8、355.6、533.4	3

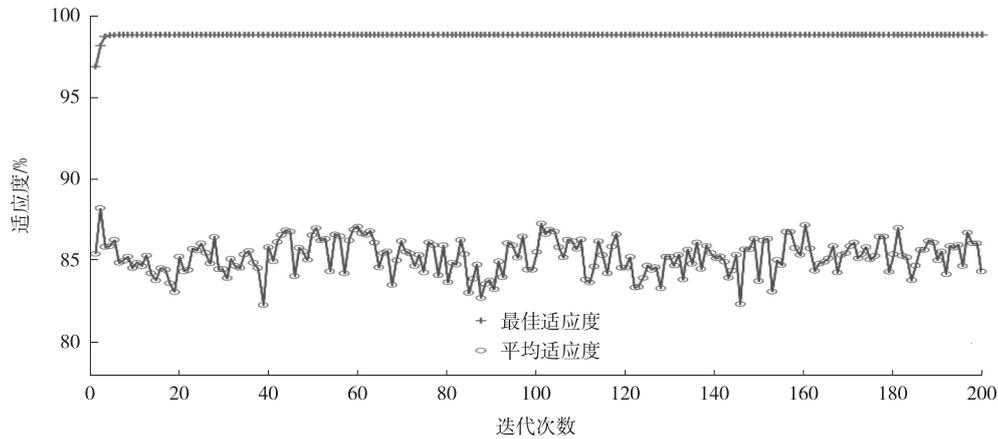


图 7 S 值寻优

Fig. 7 Optimizing for S value

表 2 改进 XGBoost 和 GBDT 诊断结果

Tab. 2 Improved XGBoost 和 GBDT diagnostic results

算法	AUC 值	诊断准确率/%
GBDT	0.541 8	95.51
改进 XGBoost	0.708 5	98.11

GBDT 是 XGBoost 的基础,因此作为对比参考项,并以 AUC 和诊断的准确率作为评判标准。众所周知,AUC 的值越接近 1 结果越好。从表 2 可以看出,XGBoost 的 AUC 值为 0.708 5,GBDT 的 AUC 值为 0.541 8,因此,XGBoost 优于 GBDT。就准确率而言,XGBoost 的值为 98.11%,GBDT 的值为 95.51%,仍是 XGBoost 优于 GBDT。因此,可以判断 XGBoost 的效果更好。

改进 XGBoost 可应用于数据集之间特征分布不均的问题,而数据集 3 为基于 K-means 方法采用均值填充缺失的数值特征集合,如表 3 所示。

在选取改进 XGBoost 重要参数 S 并计算最佳适应度、平均适应度和数值填充后,本文设计了序列向前选择、浮动向后 2 种优化特征集合的筛选流程,选择重要程度较高的特征作为分类依据,并将最终的寻优集合输入到最终的故障诊断模型中。设计寻优流程最终目的是提升诊断模型的准确度。因此,本文以准确度作为评定标准,基于改进 XGBoost 进行数据仿真。改进 XGBoost 的诊断模型以数据集 1 进行最终测试,同样包含 4 种轴

承状态,最终的诊断结果如图 8 所示。

表 3 填充实验数据及特征

Tab. 3 Filling experimental data and its characteristics

数据集	特征名称	特征个数	故障尺寸/ μm	标签
3	正常状态	40	0	0
	内圈故障	5	177.8	1
	滚动体故障	5	177.8	2
	外圈故障	5	177.8	3

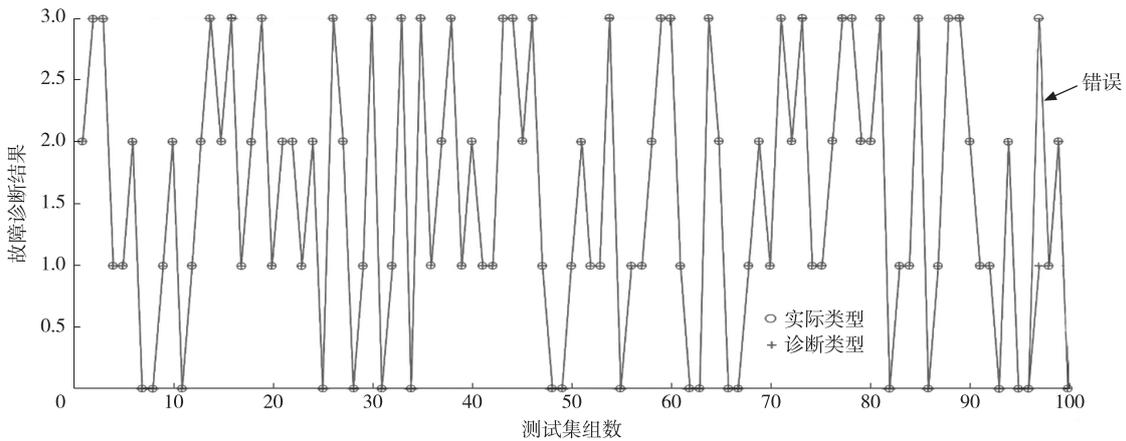


图 8 改进 XGBoost 模型故障诊断

Fig. 8 Fault diagnosis of improved XGBoost model

从图 8 的最终诊断结果可以看出:诊断结果错误数为 1,发生在第 97 组,正确率为 99%,诊断的结果和实际结果如表 4 所示。

表 4 改进 XGBoost 诊断结果

Tab. 4 Improved XGBoost diagnostic results

测试集样本组	模型诊断结果	实际结果	故障尺寸/ μm	标签
97	内圈故障	外圈故障	177.8	3

3.2 诊断模型仿真对比分析

本文构建 SVM、GS-PSO-SVM 和改进 XGBoost 这 3 种滚动轴承故障诊断模型,并使用不同类型的滚动轴承状态数据作为训练和测试参数。为了进一步验证改进 XGBoost 模型的有效性,对构建的 3 种模型进行仿真对比分析,故障诊断过程以数据集 1 为例、参数寻优使用数据集 2、数值填充使用数据集 3,最终以准确度作为模型评定标准。为了进一步说明模型的有效性,文中加入 BP-NN 诊断模型又进行了 20 次实验,诊断结果的准确率如表 5 所示。

为了对比不同模型的诊断结果,将表 5 中的数据生成改进 XGBoost 和 BP-NN 诊断模型的准确率对比结果,如图 9 所示。改进 XGBoost 和 SVM、GS-PSO-SVM 准确率对比结果则如图 10 所示。

从图 9、图 10 不难看出,改进 XGBoost 的平均准确率为 97.52%,BP-NN 平均准确率为 95.63%,SVM 平均准确率为 89.00%;GS-PSO-SVM 的平均准确率为 94.16%。即其余 3 个模型的诊断平均准确率均低于改进 XGBoost 模型,由此证明该诊断模型诊断效果最佳。

3.3 案例分析

经过 20 次滚动轴承实验验证了本文所提方法在确定轴承的故障类型中具有一定良好的效果。为了进一步证实该方法的有效性,以某企业提供的 6 205 轴承数据为例进行仿真分析,轴承故障包括内圈故障、滚动体故障及外圈故障 3 类,如表 6 所示。

表 5 4 种模型准确率对比
Tab. 5 Model accuracy comparison

实验次数	1	2	3	4	5	6	7
改进 XGBoost	99.02	99.14	100	99.58	99.12	98.78	100
SVM	91.08	90.12	89.58	89.26	90.40	90.56	90.88
GS-PSO-SVM	97.36	96.14	96.12	95.58	96.50	95.52	97.88
BP-NN	95.60	97.44	98.24	97.28	97.58	96.06	96.70
实验次数	8	9	10	11	12	13	14
改进 XGBoost	99.66	98.20	99.24	97.36	96.28	96.54	96.02
SVM	89.42	90.26	89.08	87.86	88.90	88.32	89.70
GS-PSO-SVM	95.82	95.02	96.68	94.22	93.28	91.52	92.66
BP-NN	98.36	97.42	95.22	95.18	96.28	94.64	94.02
实验次数	15	16	17	18	19	20	
改进 XGBoost	95.78	95.66	97.62	95.04	94.08	93.20	
SVM	88.12	88.24	88.12	87.16	86.52	86.46	
GS-PSO-SVM	91.98	92.46	91.58	91.62	90.12	91.16	
BP-NN	93.86	95.54	96.50	93.38	92.28	91.06	

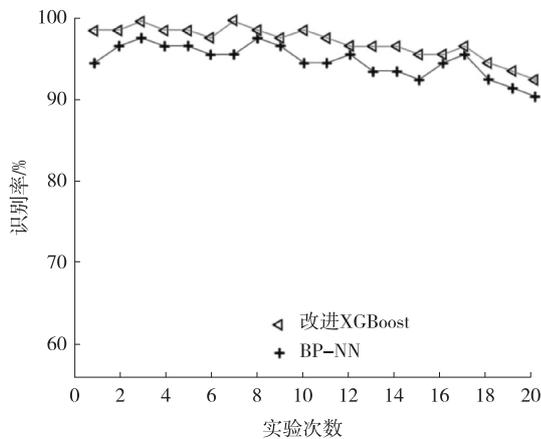


图 9 改进 XGBoost 和 BP-NN 的准确率对比
Fig. 9 Accuracy comparison of improved XGBoost and BP-NN model

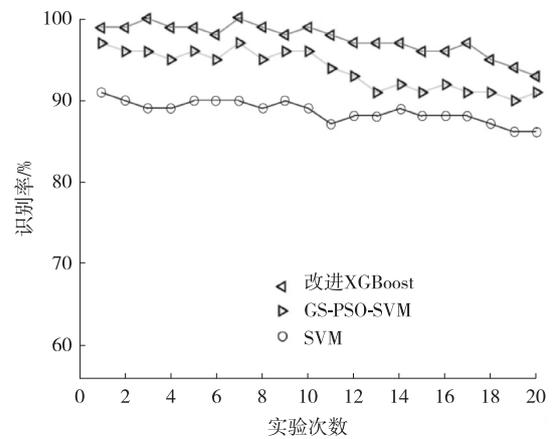


图 10 改进 XGBoost、SVM 和 GS-PSO-SVM 的准确率对比
Fig. 10 Accuracy comparison of improved XGBoost, SVM and GS-PSO-SVM model

表 6 实际轴承数据
Tab. 6 Actual bearing data

数据	故障分类	组数	分类标签
4	内圈故障	10	1
	滚动体故障	10	2
	外圈故障	10	3

由表 6 可知,实例验证故障轴承共 30 组,经诊断模型计算后,得出每一组诊断类型和轴承故障实际类型,诊断结果如图 11 所示。由图可知,模型分类结果错误出现在第 23 组,将外圈故障诊断成内圈故障,准确率为 96.67%,与模型仿真结果接近。因此,进一步证明了改进模型的可行性和有效性。

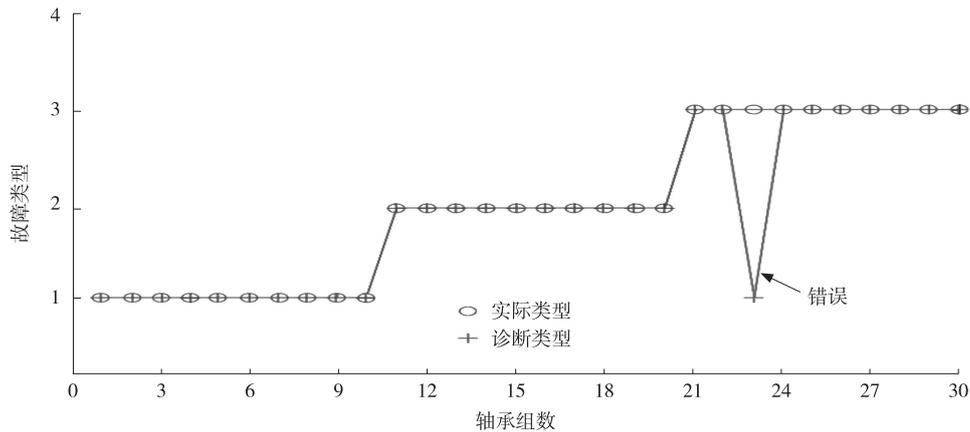


图 11 30 组实际数据

Fig. 11 30 groups actual data

4 结论

1) 本文结合大数据技术和轴承故障诊断基本理论提出了基于 LMD+FPA 的振动信号联合降噪方法,该方法可以实现将噪声信号从振动信号中剥离的效果,同时解决了信号盲源欠定的问题。

2) 针对传统故障诊断中单一指标难以准确反映故障类型的问题,本文采用 KPCA 方法选取多维指标以构建多维故障特征集合,解决了滚动轴承故障特征难以选取的问题。

3) 为了进一步提高故障诊断的准确率,本文通过 GS-PSO-SVM 对 XGBoost 进行了改进,构建了适用于轴承一般性问题的基于改进 XGBoost 的轴承故障诊断模型。通过模型仿真对比分析和实例应用验证了模型能够实现滚动轴承的故障分类,且在准确度方面具有较好的稳定性等优势。

参考文献:

- [1] SHIZA M, ISLAM M, MUHAMMAD S. Deep learning aided data-driven fault diagnosis of rotatory machine: a Comprehensive Review[J]. *Energies*, 2021, 14(16): 113-125.
- [2] HAN M, WU Y, WANG Y. Roller bearing fault diagnosis based on LMD and multi-scale symbolic dynamic information entropy [J]. *Journal of Mechanical Science and Technology*, 2021, 35(5): 1993-2005.
- [3] 郑恒, 姜宏, 章翔峰. 聚 KPCA 在高维轴承故障诊断中的应用[J]. *机床与液压*, 2021, 49(11): 179-180.
ZHENG H, JIANG H, ZHANG X F. Application of poly-KPCA in fault diagnosis of high-dimensional bearings[J]. *Machine tools and hydraulics*, 2021, 49(11): 179-180.
- [4] NI G, CHEN J. Degradation assessment of rolling bearing towards safety based on random matrix single ring machine learning [J]. *Safety Science*, 2019, 118: 403-408.
- [5] 张亚靓, 纪俊卿, 孟祥川, 等. 基于指数小波阈值与 PSO-DP-LSSVM 的发动机轴承故障诊断[J]. *机床与液压*, 2020, 48(19): 196-200.
ZHANG Y L, JI J Q, MENG X C, et al. Engine bearing fault diagnosis based on exponential wavelet threshold and PSO-DP-LSSVM[J]. *Machine tools and hydraulics*, 2020, 48(19): 196-200.
- [6] OPRESNIK D, TAISCH M. The value of big data in servitization[J]. *International Journal of Production Economics*. 2015, 165: 174-184.
- [7] 王秋秋. Spark 环境下半参数支持向量机的研究与优化[D]. 西安: 西安科技大学, 2019.
WANG Q Q. Research and optimization of semi-parametric support vector machine in Spark environment[D]. Xi'an: Xi'an University of Science and Technology, 2019.
- [8] 李鑫灵, 袁梅, 董洪, 等. PSO-SVM 模型在掘进工作面突出预警系统中的应用[J]. *煤矿安全*, 2021, 52(9): 90-95.
LI X L, YUAN M, DONG H, et al. Application of PSO-SVM model in excavation face outburst warning system[J]. *Coal Mine Safety*, 2021, 52(9): 90-95.
- [9] PAPER D. Hands-on Scikit-learn for machine learning applications[M]. Apress, Berkeley, CA: 2020.
- [10] CHEN W B, KUN F, ZUO J W. Radar emitter classification for large data set based on weighted-XGBoost[J]. *IET Radar*,

- Sonar&Navigation, 2017, 11(8):1203-1207.
- [11] 者娜. 基于大数据技术的滚动轴承故障诊断及剩余寿命预测方法研究[D]. 北京:北京化工大学, 2018.
ZHE N. Research on fault diagnosis and remaining life prediction method of rolling bearing based on big data technology[D]. Beijing:Beijing University of Chemical Technology, 2018.
- [12] 李祥. 基于改进 LMD 和多尺度散布熵的滚动轴承损伤识别研究[D]. 昆明:昆明理工大学, 2020.
LI X. Research on damage identification of rolling bearing based on improved LMD and multi-scale scatter entropy[D]. Kunming:Kunming University of Science and Technology, 2020.
- [13] 车昱娇, 陈云霞, 崔宇轩. KPCA 和改进 LSTM 在滚动轴承剩余寿命预测中的应用研究[J]. 电子测量与仪器学报, 2021, 35(2):109-114.
CHE Y J, CHEN Y X, CUI Y X. Research on the application of KPCA and improved LSTM in the prediction of residual life of rolling bearings[J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(2):109-114.
- [14] MAO D Y, LI C F, MIKULOVICH V I, et al. Method of assessing the state of a rolling bearing based on the relative compensation distance of multiple-domain features and locally linear embedding[J]. Mechanical Systems and Signal Processing, 2017, 86:40-57.
- [15] CHEN W B, KUN F, ZUO J W. Radar emitter classification for large data set based on weighted-XGBoost[J]. IET Radar, Sonar&Navigation, 2017, 11(8):1203-1207.
- [16] 张钰. XGboost 在滚动轴承故障诊断中的应用[J]. 噪声与振动控制, 2017, 37(4):166-170.
ZHANG Y. Application of XGboost in fault diagnosis of rolling bearing[J]. Noise and Vibration Control, 2017, 37(4):166-170.

Operations Research and Cybernetics

A Bearing Fault Diagnosis Method Based on Improved XGBoost

ZHANG Tianrui, ZHAO Weibo, ZHOU Fuqiang, ZHU Zhiyi, SONG Yuru, JIA Zexuan

(School of Mechanical Engineering, Shenyang University, Shenyang 110044, China)

Abstract: Aiming at the problem of low accuracy of traditional fault diagnosis models facing massive fault data, a joint denoising method of local mean decomposition and fixed point algorithm was proposed to eliminate the noise in bearing vibration signals. Secondly, in order to solve the problem of difficult extraction of sensitive features from original signals, a dimension reduction method based on kernel principal component analysis is proposed. Thirdly, a fault diagnosis model based on the improved eXtreme Gradient Boosting decision tree was constructed. The GS-PSO algorithm was used to optimize the performance of SVM, and then the residual error of classification model was modified by using the improved XGBoost idea to improve the classification accuracy of the model. Spark- Big data platform was used to carry out scientific calculation by parallel processing technology. Finally, the rolling bearing data provided by CWRU is used for training and simulation, which proves that the established model can realize the identification and diagnosis of different types of rolling bearings and ensure the accuracy of diagnosis results. Through comparative analysis of four different fault diagnosis models, the results show that the model is feasible and advantageous.

Keywords: rolling bearings; fault diagnosis; big data; feature extraction; XGBoost

(责任编辑 黄颖)