

基于 BO-DKELM 的滚动轴承故障诊断

聂新华, 秦玉峰, 李尚懿

(海军航空大学 岸防兵学院, 山东 烟台 264001)

摘要: 滚动轴承作为旋转机械中的必需元件, 其任何故障都可能导致机器乃至整个系统发生故障, 从而导致巨大的经济损失和时间的浪费, 因此必须及时准确地诊断滚动轴承故障; 针对传统极限学习机中模型参数对滚动轴承故障诊断精度影响较大的问题, 提出了一种基于贝叶斯优化的深度核极限学习机的滚动轴承故障诊断方法; 首先, 将自动编码器与核极限学习机相结合, 构建了深度核极限学习机 (DKELM) 模型; 其次, 利用贝叶斯优化 (BO) 算法对 DKELM 中的超参数进行寻优, 使得训练数据集和验证数据集在 DKELM 模型中的分类错误率之和最低; 然后, 将测试数据集输入到训练好的 BO-DKELM 中进行故障诊断; 最后, 采用凯斯西储大学轴承故障数据集对所提方法进行验证, 最终故障诊断精度为 99.6%, 与深度置信网络和卷积神经网络等传统智能算法进行对比, 所提方法具有更高的故障诊断精度。

关键词: 滚动轴承; 故障诊断; 深度核极限学习机; 贝叶斯优化; 深度学习

Rolling Bearing Fault Diagnosis Based on BO-DKELM

NIE Xinhua, QIN Yufeng, LI Shangcong

(College of Coastal Defense Force, Naval Aviation University, Yantai 264001, China)

Abstract: As a necessary component in rotating machinery, any fault of rolling bearings may lead to mechanical or even system failures, resulting in huge economic loss and time wastage, therefore, it is necessary to promptly and accurately diagnose the rolling bearing fault. In response to the problem that there is the large influence of the model parameters on the fault diagnosis accuracy of rolling bearings, a rolling bearing fault diagnosis method based on deep kernel extreme learning machine with Bayesian optimization is proposed. Firstly, the deep kernel extreme learning machine (DKELM) model is constructed by combining the auto encoder (AE) with the kernel extreme learning machine (KELM). Secondly, the Bayesian optimization algorithm is used to search the optimal hyperparameters in the DKELM, and minimize the classification error rates of the training and validation datasets in the DKELM model. Then, the test dataset is then input to the trained BO-DKELM for the fault diagnosis. Finally, the proposed method is validated on the bearing fault dataset of the Case Western Reserve University, the result shows that the final fault diagnosis accuracy is 99.6%, compared with traditional intelligent algorithms such as deep belief networks and convolutional neural networks, the proposed method has a higher fault diagnosis accuracy.

Keywords: rolling bearing; fault diagnosis; deep kernel extreme learning machine; Bayesian optimization; deep learning

0 引言

旋转机械设备目前正广泛应用于电力、制造和航空等各种工业部门中^[1-2]。作为故障诊断领域的一个主要问题, 滚动轴承的故障诊断已引起研究人员的广泛关注^[3-7]。随着机械设备的复杂化和智能化, 设备中普遍安装了大量传感器, 这些传感器能够采集到丰富的运行数据。智能故障诊断算法能够挖掘这些数据的深层特征并将其应用于设备的故障诊断中, 因此基于数据驱动的故障诊断方法受到了广泛关注并取得了丰硕成果^[8-10]。基于深度学习的诊断方法是一种典型的数据驱动方法, 该方法通过训练神经网络直接建立样本到故障类别的映射关系, 而无需人工提取故障特征, 具有更强的适用性。

分类器作为智能故障诊断过程中的关键因素, 直接影响着故障诊断的精度。极限学习机 (ELM, extreme learning

machine) 是一种单隐层前馈神经网络 (SLFN, single hidden layer forward neural network) 学习方法。在网络训练过程中, ELM 随机初始化输入权重和偏置并得到相应的输出权重, 且不需要更新输入权重和隐层阈值, 因此相比于传统的神经网络, ELM 具有学习速度快的优点^[11-13]。大量研究成果表明: 相比于 BP 神经网络和支持向量机, ELM 具有更快的训练速度和更好的泛化能力^[14-16], 因此 ELM 在智能故障诊断中具有广泛的应用前景。Gao 等人^[17]提出了一种集成 ELM 故障诊断模型; Lei 等人^[18]将本征时间尺度分解技术与核极限学习机 (KELM, Kernel extreme learning machine) 相结合对铣削刀具磨损情况进行分类; Chen 等人^[19]提出了一种求和高斯核极限学习机, 并成功地应用于输电线路故障诊断; Li 等人^[20]将 ELM 与自动编码器 (AE, auto encoder) 相结合, 提出了一种基于稀疏保邻域深度极限学习机的滚动轴承故障诊断新算法; Rodriguez 等人^[21]提出了一种结合平稳小波变换

收稿日期: 2023-04-30; 修回日期: 2023-06-19。

作者简介: 聂新华(1985-), 男, 博士, 讲师。

通讯作者: 秦玉峰(1995-), 男, 博士, 讲师。

引用格式: 聂新华, 秦玉峰, 李尚懿. 基于 BO-DKELM 的滚动轴承故障诊断[J]. 计算机测量与控制, 2024, 32(4): 8-14.

和奇异值分解的 ELM 算法对滚动轴承故障进行故障诊断; Ye 等人^[22]提出了一种基于稀疏贝叶斯极限学习机的故障诊断方法。ELM 或 KELM 等故障诊断模型具有学习速度快、鲁棒性强等优点,然而上述文献未考虑故障诊断模型中的超参数对故障诊断结果的影响,如果超参数设置的不合适可能导致最终的故障诊断结果准确率较低。

针对上述问题,本文提出了一种基于 BO-DKELM (Bayesian Optimization-deep Kernel Extreme Learning Machine) 的滚动轴承故障诊断方法。由于 KELM 具有学习速度快、泛化能力强等优点,同时考虑到 AE 能够充分学习输入数据隐含特征的特点,将 AE 与 KELM 进行结合,构建了 DKELM 模型。DKELM 模型中超参数较多,若通过手动随机设置超参数的取值往往无法得到最优的结果。因此进一步利用贝叶斯优化算法对 DKELM 的超参数进行寻优,构建 BO-DKELM 故障诊断模型,旨在提高 DKELM 的故障诊断性能,提高系统的故障诊断能力。

1 DKELM 故障诊断模型构建

1.1 KELM 基本理论

ELM 模型的网络结构与 SLFN 一致,其区别在于 ELM 在训练阶段不采用传统 BP 神经网络中的梯度下降的算法,而采用随机的输入层权重和偏置,输出层权重则通过广义逆矩阵理论计算得到,在计算得到所有节点的权值和偏差后便完成了 ELM 的训练。区别于传统 BP 神经网络中的梯度下降算法,ELM 以任意连续的概率分布随机对隐含层中权值 \mathbf{W} 和偏置 \mathbf{b} 进行初始化,通过非线性激活函数将输入数据映射到新的高维空间,然后基于最小逼近平方差方法求解输出层权重 $\boldsymbol{\beta}$ 。ELM 算法具体分析如下:

假设存在 N 个训练样本 $\{(\mathbf{x}_i, \mathbf{t}_i)\}_{i=1}^N$,其中 $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{im}]^T \in \mathbb{R}^n$ 为输入样本,目标输出值 $\mathbf{t}_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T \in \mathbb{R}^m$ 。对于含有 L 个隐含层节点的 ELM 数学模型可以用如下公式进行表示:

$$y_j = \sum_{i=1}^L \beta_i g(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) \quad j = 1, 2, \dots, N \quad (1)$$

式中, β_i 为连接隐含层与输出层的输出权重; g 为激活函数,表示隐含层对于输入样本的响应方式。 $\mathbf{w}_i = [w_{i1}, w_{i2}, \dots, w_{im}]^T$ 为连接第 i 个隐含层节点和输入节点之间的权重; b_i 为偏置。ELM 的网络结构如图 1 所示。

若激活函数 g 能够使得网络输出任意逼近全部 N 个输入样本,即: $\sum_{i=1}^N \|y_i - t_i\| = 0$,则有:

$$t_j = \sum_{i=1}^L \beta_i g(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) \quad j = 1, 2, \dots, N \quad (2)$$

上述方程可以用矩阵形式进行表示:

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{T} \quad (3)$$

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \\ \vdots \\ \mathbf{h}_N \end{bmatrix} = \begin{bmatrix} g(\mathbf{w}_1 \cdot \mathbf{x}_1 + b_1) & \cdots & g(\mathbf{w}_L \cdot \mathbf{x}_1 + b_L) \\ g(\mathbf{w}_1 \cdot \mathbf{x}_2 + b_1) & \cdots & g(\mathbf{w}_L \cdot \mathbf{x}_2 + b_L) \\ \vdots & \cdots & \vdots \\ g(\mathbf{w}_1 \cdot \mathbf{x}_N + b_1) & \cdots & g(\mathbf{w}_L \cdot \mathbf{x}_N + b_L) \end{bmatrix} \quad (4)$$

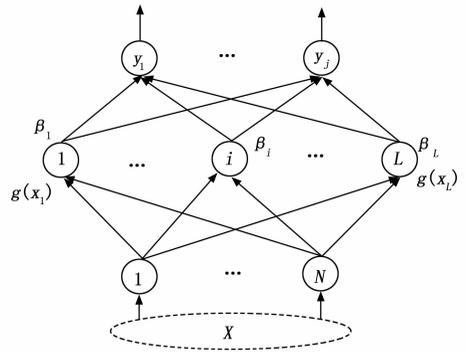


图 1 ELM 网络结构

式中, \mathbf{H} 为极限学习机的隐含层输出矩阵; $\boldsymbol{\beta}$ 为输出权重向量; \mathbf{T} 为期望输出向量。通过求解线性方程组的最小二乘解,即寻找最优的权值 $\boldsymbol{\beta}^*$ 使得 $\|\mathbf{H}\boldsymbol{\beta} - \mathbf{T}\|^2$ 取最小值,进而求取输出权重向量。根据广义逆的理论,其解为:

$$\boldsymbol{\beta}^* = \mathbf{H}^+ \mathbf{T} \quad (5)$$

式中, \mathbf{H}^+ 表示隐含层输出矩阵 \mathbf{H} 的 Moore-Penrose 广义逆。

上述方法是基于训练误差最小情况下进行计算的,容易产生过拟合的现象,从而无法获得最优的分类效果。因此对问题进行折中考虑,使得训练误差和输出权值同时达到最小化:

$$\begin{cases} \min \|\boldsymbol{\beta}\|^2 \\ \min \sum_{i=1}^N \|h_i \boldsymbol{\beta} - t_i\| \end{cases} \quad (6)$$

式 (6) 可以改写为:

$$\begin{cases} \min L_{ELM} = \frac{1}{2} \|\boldsymbol{\beta}\|^2 + \frac{1}{2} C_1 \sum_{i=1}^N \|\xi_i\|^2 \\ s. t. \quad h_i \boldsymbol{\beta} = t_i - \xi_i \quad i = 1, \dots, N \end{cases} \quad (7)$$

其中: ξ_i 是第 i 个样本 \mathbf{x}_i 与隐含层输出之间的误差; C_1 为正则化系数,用来控制误差在最小化问题中的占比。根据 KKT 条件,引入拉格朗日乘子求解公式:

$$\min L_{ELM} = \frac{1}{2} \|\boldsymbol{\beta}\|^2 + \frac{1}{2} C_1 \sum_{i=1}^N \|\xi_i\|^2 - \sum_{i=1}^N \sum_{j=1}^m \alpha_{ij} [h_j \boldsymbol{\beta}_i - t_{ij} + \xi_{ij}] \quad (8)$$

其中: α_{ij} 为拉格朗日乘子。求解公式的优化条件可得:

$$\begin{cases} \frac{\partial L_{ELM}}{\partial \boldsymbol{\beta}_j} = 0 \Rightarrow \boldsymbol{\beta}_j = \sum_{i=1}^N \boldsymbol{\alpha}_i h_i^T \Rightarrow \boldsymbol{\beta} = \mathbf{H}^T \boldsymbol{\alpha} \\ \frac{\partial L_{ELM}}{\partial \xi_i} = 0 \Rightarrow \boldsymbol{\alpha}_i = C_1 \xi_i \quad i = 1, \dots, N \\ \frac{\partial L_{ELM}}{\partial \boldsymbol{\alpha}_i} = 0 \Rightarrow h_i \boldsymbol{\beta} - t_i^T + \xi_i^T = 0 \quad i = 1, \dots, N \end{cases} \quad (9)$$

根据公式 (9) 可得如下方程组:

$$\begin{cases} h_1 \mathbf{H}^T C_1 \xi_1 - t_1^T + \xi_1^T = 0 \\ \vdots \\ h_N \mathbf{H}^T C_1 \xi_N - t_N^T + \xi_N^T = 0 \end{cases} \quad (10)$$

公式 (10) 等价于:

$$T = \left(\frac{I}{C_1} + HH^T \right) \alpha \tag{11}$$

其中： I 为单位矩阵。结合公式 (5) 可以得到 ELM 网络的输出权值的最小二乘解：

$$\hat{\beta} = H^T \left(\frac{I}{C_1} + HH^T \right)^{-1} T \tag{12}$$

传统 ELM 模型采用单隐层结构，在隐含层中设置一定数量的神经元节点，通过非线性激活函数将输入数据映射到高维特征空间，并且数据样本在被映射到高空间后能够线性分离，这与前文中核方法的原理是一致的。所以可以根据核方法基本定理，将 ELM 对输入矩阵的非线性映射过程利用核函数进行替换，从而提高 ELM 的非线性逼近能力和稳定性。因此，将核函数引入到 ELM 中，提出如下 KELM 算法：

首先应用 Mercer's 条件定义核矩阵^[23]：

$$K_{ij} = \mathbf{h}_i \cdot \mathbf{h}_j = k(\mathbf{x}_i, \mathbf{x}_j) \tag{13}$$

其中： K_{ij} 为核矩阵 \mathbf{K} 中第 i 行第 j 列元素。通过核函数 $k(\mathbf{x}_i, \mathbf{x}_j)$ 将输入样本从 n 维输入空间映射到高维特征空间。核函数包括 RBF 核函数、线性核函数和多项式核函数等，通常设定为 RBF 核：

$$k(\mathbf{x}_i, \mathbf{x}_j) = e^{(-\|\mathbf{x}_i - \mathbf{x}_j\|^2)/\delta} \tag{14}$$

式中， δ 为核宽度。核参数选择完毕后，则核矩阵 \mathbf{K} 是确定的，因此通过定义核函数则可以将 ELM 中的随机矩阵 HH^T 替换为确定的核矩阵 \mathbf{K} 。由以上分析可得，KELM 模型的求解公式可以表示为：

$$f(\mathbf{x}) = \mathbf{h}_i H^T \left(\frac{I}{C_1} + HH^T \right)^{-1} T = \begin{bmatrix} k(\mathbf{x}, \mathbf{x}_1) \\ \vdots \\ k(\mathbf{x}, \mathbf{x}_N) \end{bmatrix} \left(\frac{I}{C_1} + \mathbf{K} \right)^{-1} T \tag{15}$$

因此可得 KELM 模型输出权值为：

$$\beta = (I/C_1 + \mathbf{K})^{-1} T$$

在 KELM 算法中，由于利用核矩阵对随机矩阵 HH^T 进行了替换，因此只需要定义核矩阵 $k(\mathbf{x}_i, \mathbf{x}_j)$ 即可计算得到输出函数的值。同时，因为核函数直接采用内积的形式，在求解输出函数值时不需要主观设定网络隐含层节点数目，也不需要设定隐含层初始权重和偏置。

1.2 自动编码器

AE 是一种前馈全连接的对称神经网络，其网络结构如图 2 所示。

AE 主要由编码层和解码层构成，其中：编码层由输入层和隐含层构成，解码层由隐含层和输出层组成，解码层与编码层的结构类似，且输出与输入的维度相同^[24]。AE 通过接收、压缩和编码数据，然后以损失最小化为目标，通过无监督训练直到输出尽可能接近地输入数据。其目的在于通过训练学习数据的特征。AE 的编码过程为通过激活函数将数据的高维特征降维压缩为隐含层的低维特征表示，实现对输入数据的编码；AE 的解码过程为通过激活函数将隐含层的低维特征表示以输入数据为目标进行重构，将输

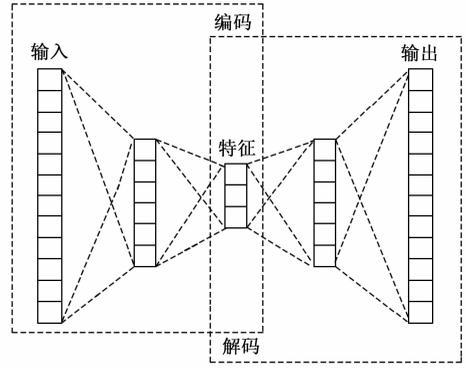


图 2 AE 网络结构

入从编码中重建为原始维度，通过对数据进行编码和解码可以学习到输入数据的隐含特征。AE 的数学模型如下：

假设存在 N 个训练样本 $\{(\mathbf{x}_i, \mathbf{t}_i)\}_{i=1}^N$ ，其中 $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in \mathbb{R}^n$ 为输入样本，则原始输入样本的编码过程可以表示为：

$$\mathbf{H} = g(\mathbf{w} \cdot \mathbf{x} + \mathbf{b}) \tag{16}$$

其中： \mathbf{H} 为隐含层的输出特征矩阵； \mathbf{w} 为编码层的权值矩阵； \mathbf{b} 为编码层的偏置向量； g 为激活函数。解码过程为输入数据的重构过程，隐含层输出特征矩阵 \mathbf{H} 的解码过程可以表示为：

$$\hat{\mathbf{x}} = g^*(\mathbf{w}^* \cdot \mathbf{H} + \mathbf{b}^*) \tag{17}$$

其中： $\hat{\mathbf{x}}$ 为输入 \mathbf{x} 的重构数据； \mathbf{w}^* 为解码层的权值矩阵； \mathbf{b}^* 表示解码层的偏置向量。

1.3 DKELM 模型构建

可以发现若 AE 仅含有一个隐含层，则 ELM 与 AE 在结构上非常相似，因此将 AE 的思想与 ELM 相结合中，构造作为自编码器的极限学习机 ELM-AE，使 ELM 的输出重构为输入。ELM-AE 的网络结构如图 3 所示。

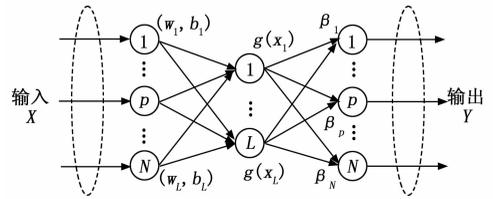


图 3 作为自动编码器的极限学习机结构

与 ELM 类似，ELM-AE 隐含层输出权重可以由公式转化为：

$$\hat{\beta} = H^T \left(\frac{I}{C_2} + HH^T \right)^{-1} X \tag{18}$$

其中： C_2 为 ELM-AE 的正则化系数。根据 ELM-AE 的特征学习能力，本文构造了一种深度极限学习机模型，将 ELM-AE 作为 DKELM 的基本单元，利用逐层贪婪的训练方法对 DKELM 进行训练。DKELM 的基本思想为：利用 AE 的特征学习能力，令 ELM-AE 的输出在最小重构误差的情况下无限接近原始输入，通过多个 ELM-AE 单元的分

层无监督训练确定 DKELM 中每个隐含层的权重。经过多层 ELM-AE 单元的训练后, DKELM 模型能够深层次地提取输入数据的隐含特征, 从而提高最终分类效果。DKELM 模型的训练过程如图 4 所示。

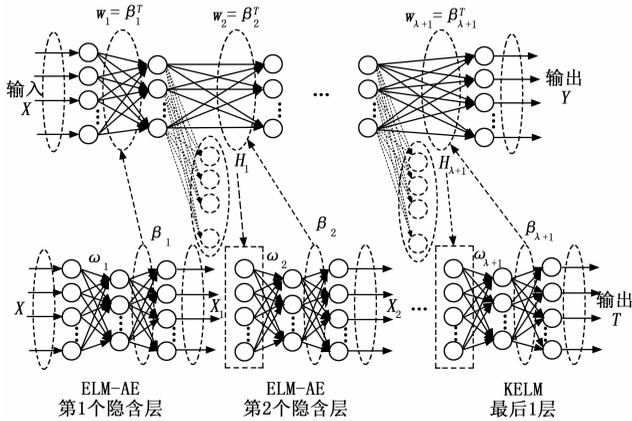


图 4 DKELM 模型训练过程

其中: $H_{\lambda+1}$ 为 KELM 层的输出矩阵, T 为样本标签, KELM 的输入权重矩阵为 $w_{\lambda+1} = \beta_{\lambda+1}^T$ 。假设存在 λ 个 ELM-AE 模块, 将输入数据样本 X 作为第 1 个 ELM-AE 的目标输出: $X_1 = X$, 求取第 1 个 ELM-AE 单元的输出权值 β_1 ; 然后将 DKELM 第 1 个隐含层的输出矩阵 H_1 作为第 2 个 ELM-AE 的目标输出: $X_2 = H_1$, 求取第 2 个 ELM-AE 单元的输出权值 β_2 , 并依此类推。模型的输出层利用 KELM 来训练, 并根据公式 (15) 对 KELM 模型进行求解。

根据分析, 本文所构建的 DKELM 模型中包含 λ 个 ELM-AE 模型和一个 KELM 模型。虽然 DKELM 模型能够用于系统的故障诊断, 但是存在以下问题: 由于模型中超参数过多, 主要包括: DKELM 中各隐含层的节点个数 n_i , KELM 中的正则化系数 C_1 与核参数 δ , ELM-AE 中的正则化系数 C_2 。各超参数对最终的故障诊断精度都存在着较大影响, 若通过手动随机设置各超参数往往无法得到最优的结果。因此本文提出一种 BO-DKELM 模型, 利用贝叶斯优化算法对 DKELM 的超参数进行寻优, 旨在提高 DKELM 的性能, 更好地应用于实际故障诊断。

2 BO-DKELM 模型

目前, 贝叶斯优化在求解黑盒函数问题中应用越来越广泛, 已经成为超参数估计的主流方法^[25-26]。区别于网格搜索方法和随机搜索方法, 贝叶斯优化采用序贯的算法模型, 通过选取参数点不断更新概率模型, 在过去搜索结果的基础上获得当前的最优值。贝叶斯优化方法中主要包括两个组成部分: 1) 概率代理模型: 概率代理模型描述了目标函数的历史结果, 每一次迭代过程概率代理模型根据当前数据点的目标函数结果进行改进, 直至找到符合要求的结果或者达到最大迭代次数。概率代理模型用来代理描述目标函数, 对数据点集合和目标函数值进行统计分析; 2) 采集函数: BO 通过将概率代理模型代入采集函数中计

算极值, 估计当前情况下最优数据点最有可能出现的一些位置, 并将其作为选择下一组需要计算的数据点。下面从概率代理模型和采集函数两个方面介绍 BO 的基本原理。

2.1 概率代理模型

由于优化目标函数可能不存在显式的函数表达式, 因此 BO 采用概率代理模型对目标函数进行近似表示。基于初始采样数据点开始构建初步的概率代理模型, 随着数据点信息量的增加, 逐步对概率代理模型进行修正和更新。概率代理模型一般分为参数模型和非参数模型两种。目前高斯过程 (GP, gaussian process) 回归作为概率代理模型取得了较好的应用。高斯过程回归属于非参数模型, 其每个有限子集都服从多元正态分布, 通过使用高斯过程的先验信息进行回归分析。高斯过程是多元高斯分布向无穷维的扩展, 由均值函数和协方差函数组成:

$$f(x) \sim GP(m(x), k(x, x')) \quad (19)$$

其中: $f(x)$ 为目标函数; $m(x)$ 为均值函数; $k(x, x')$ 为协方差函数矩阵。在第 t 次试验后, 可以获得数据点集合 $\{x_{1:t}, f_{1:t}\}$, 由于高斯过程上任意点 f_{t+1} 与之前的观测数据服从联合高斯分布, 进一步可以得到预测分布:

$$P(f_{t+1} | D_{1:t}, x_{t+1}) = N(\mu_t(x_{t+1}), \sigma_t^2(x_{t+1})) \quad (20)$$

对于均值和协方差函数的推导和具体形式本文不再赘述。根据式 (20), 则可以根据高斯过程的后验分布对未知函数 $f(x)$ 在任意位置的值和所在位置的均值、方差做出估计。

2.2 采集函数

直接从概率代理模型中随机选择数据点寻找最优解通常需要耗费大量的时间。BO 采集函数在每次迭代过程中获得下一个采样数据点。采集函数根据已经观测到的数据更新的后验概率分布进行构造, 下一个评估的采样数据点通过最大化后验概率分布进行选择, 即:

$$x_{t+1} = \operatorname{argmax}_x \alpha_t(x; D) \quad (21)$$

常见的采集函数包括概率提升函数 (PI, probability of improvement)、期望提升函数 (EI, expected improvement) 以及上置信边界函数 (UCB, upper confidence bound) 等。由于本文使用高斯过程作为概率代理模型, 其后验概率分布预测中包括均值和方差, 因此通过使用代理模型计算得到预测分布中均值和方差的关系, 使采集函数最大化, 以此来估计最优的下一个采样结构点 x_{t+1} :

$$\alpha_{\text{UCB}}(x) = \mu(x) + \gamma\sigma(x) \quad (22)$$

其中: γ 为调节参数, 采集函数通过 γ 控制对探索未知和已知数据信息之间的平衡。

2.3 BO-DKELM 模型训练过程

BO-DKELM 模型的训练过程如图 5 所示。

在图 5 中, 贝叶斯优化的目标函数设计为分类错误率, 优化目标为在超参数空间内寻找到一组超参数组合 $x = \{n_i, C_1, C_2, \delta\}$, 使得训练集和验证集的错误率之和最低:

$$\xi = \operatorname{argmin}_x 2 - Acc_t - Acc_v \quad (23)$$

其中: Acc_t 为训练集的准确率; Acc_v 为验证集的准确率。

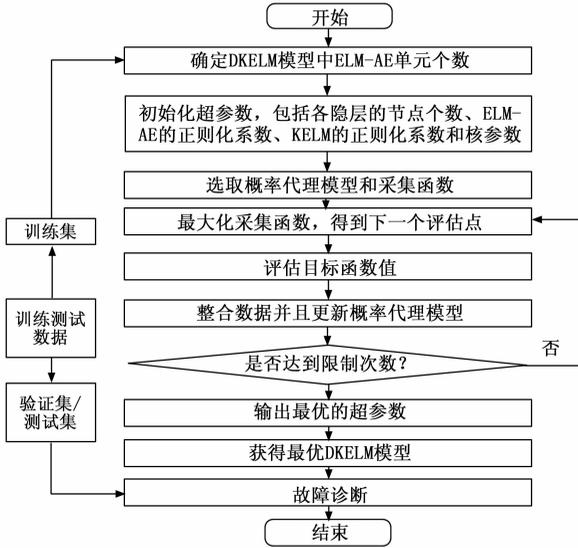


图 5 BO-DKELM 故障诊断模型训练过程

此时可确定概率代理模型的输入为数据点集合 $\{x_{1:t}, \xi_{1:t}\}$, 输出为超参数与目标函数之间的函数关系 $f(x)$ 。基于 BO-DKELM 模型的故障诊断流程可以分为 4 部分:

- 1) 将采样数据划分为训练集、验证集和测试集;
- 2) 构建 DKELM 模型, 确定网络层数、ELM-AE 采用的激活函数、KELM 采用的核函数;
- 3) 训练 BO-DKELM 模型, 输出 DKELM 模型中各隐层的节点个数、ELM-AE 的正则化系数、KELM 的正则化系数和核参数, 构建 BO-DKELM 模型;
- 4) 通过测试数据集验证所构建模型的故障分类性能。

3 案例验证

为了验证所提方法的可行性和有效性, 本文采用凯斯西储大学实验室的实验数据, 实验设施如图 6 所示^[27]。

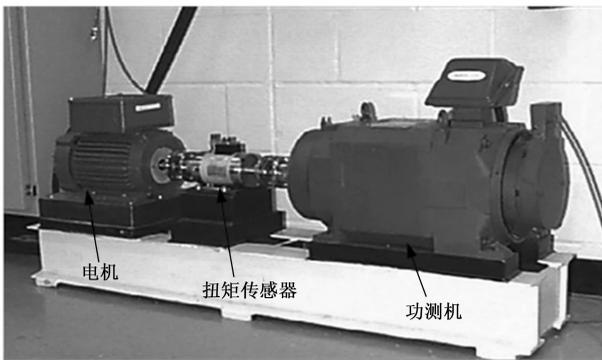


图 6 故障实验设施

电机作为整套系统的动力源, 驱动滚动轴承旋转带动负载工作; 功测机用于模拟测试平台的负载, 其实际输出扭矩可根据测试需求进行调节, 并通过扭矩传感器进行测量; 利用加速度作为振动传感器, 通过测量加速度对振动强度进行估计; 各种不同故障类型的滚动轴承用于模拟实

际生产中的滚动轴承故障。故障位置主要包括滚珠缺陷、外圈缺陷和内圈缺陷。轴承的故障由电加工产生, 故障尺寸为 7、14 和 21 mil。每一种故障都有上述的 3 种损伤情况, 代表着不同严重程度的故障, 因此一共有 10 种不同的轴承健康状况数据, 包括正常情况的轴承和不同严重程度的故障情况。具体信息列于表 1 中。

表 1 故障信息表

故障编号	故障部位	故障幅度/mil	故障方位	标签
f_1	滚珠	7	—	1
f_2	滚珠	14	—	2
f_3	滚珠	21	—	3
f_4	内圈	7	—	4
f_5	内圈	14	—	5
f_6	内圈	21	—	6
f_7	外圈	7	@6:00	7
f_8	外圈	14	@6:00	8
f_9	外圈	21	@6:00	9
f_{10}	无	—	—	10

注:符号 '@6:00' 表示故障方向在 6 点钟方向

将样本的数据点设置为 1 024, 将测试信号作为 BO-DKELM 的输入。对于每个类别的样本, 从原始振动信号中随机构建 250 个样本, 因此构建了含有 2 500 个样本的样本集, 设置训练集: 验证集: 测试集的比例为 7 : 2 : 1。设置 ELM-AE 单元个数为 2 个, 其中激活函数选用 sigmoid 函数。其中: ELM-AE 单元的节点数 n_1 和 n_2 的搜索范围设置为 $[0, 200]$, ELM-AE 的正则化系数 C_2 、KELM 中正则化系数 C_1 和核参数 δ 的搜索范围统一设置为 $[0.001, 1\ 000]$, 最大搜索次数设置为 20 次, 最小目标值搜索过程和故障诊断结果如图 7 和图 8 所示。

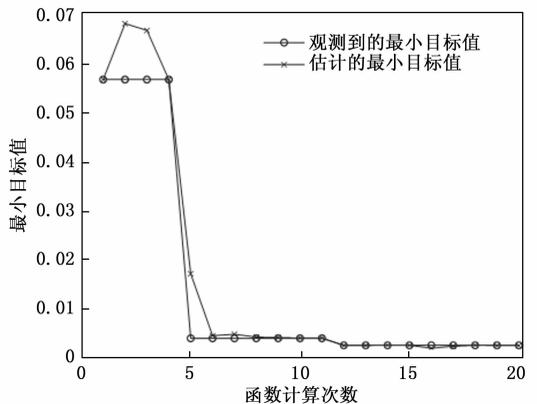


图 7 最小目标值搜索过程

从仿真结果可以看出: 1) BO-DKELM 算法在第 12 次迭代时结果收敛, 训练集与测试集的错误率之和最低为 0.25%; 2) 经过 20 次搜索后, 贝叶斯优化算法给出了模型的最优超参数分别为: $n_1 = 199, n_2 = 159, C_1 = 539.205\ 0, C_2 = 622.172\ 6, \delta = 38.807\ 9$; 3) BO-DKELM 算法的测试集分类正确率为 99.60%, F_1 得分为 0.996 4。

输出类别	1	28 11.2%	1 0.4%	0 0.0%	96.6% 3.4%										
	2	0 0.0%	26 10.4%	0 0.0%	100% 0.0%										
	3	0 0.0%	0 0.0%	24 9.6%	0 0.0%	100% 0.0%									
	4	0 0.0%	0 0.0%	0 0.0%	21 8.4%	0 0.0%	100% 0.0%								
	5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	28 11.2%	0 0.0%	100% 0.0%							
	6	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	20 8.0%	0 0.0%	100% 0.0%						
	7	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	27 10.8%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	8	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	23 9.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	9	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	28 11.2%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	10	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	24 9.6%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
		100% 0.0%	96.3% 3.7%	100% 0.0%	99.6% 0.4%										
		1	2	3	4	5	6	7	8	9	10				
		实际类别													

图 8 基于 BO-DKELM 的故障诊断结果

为说明本文方法的优越性, 将本文所提 BO-DKELM 故障诊断模型与 DKELM、KELM 和 ELM 等模型进行对比, 各模型的激活函数都选择为 Sigmoid 函数, 表 2 给出了算法其余超参数具体数值。

表 2 模型超参数设置

模型	n_1	n_2	C_1	C_2	δ
BO-DKELM	199	159	539.205 0	622.172 6	38.807 9
DKELM	100	50	10	10	1
KELM	—	50	10	—	1
ELM	—	50	—	—	—

所有模型所用的故障数据集相同。一共进行了 20 次实验, 取 20 次实验结果的平均值, 最终故障诊断结果对比如表 3 所示。

表 3 故障诊断结果对比

模型	训练集准确率/%	验证集准确率/%	测试集准确率/%	F_1 得分
BO-DKELM	99.90	99.78	99.60	0.994 4
DKELM	93.17	92.50	94.44	0.943 9
KELM	84.69	82.37	83.64	0.871 4
ELM	81.83	80.80	79.48	0.801 2

可见, 基于 BO-DKELM 的诊断模型相比于 DKELM、KELM 和 ELM, 测试集准确率分别提高了 5.06%、15.86% 和 20.02%, F_1 得分分别提高了 0.050 5、0.123 0 和 0.193 2。进一步将本文所提方法与 CNN、2D-CNN^[28]、DBN^[29]、CNN+2D-CNN 进行比较, 这些方法的详细描述如下:

1) CNN 包括一个输入层, 两个一维卷积层 (卷积规模分别为 1×6 和 6×16), 两个池化层, 一个全连接层, 一个 softmax 分类器和一个输出层, 卷积层和池化层的核尺寸分别为 5、6、3、6。

2) 由于原始信号数据的维度为 1 024, 将原始信号数据转化为 32×32 维的矩阵作为 2D-CNN 的输入。2D-CNN 包括一个输入层, 两个一维卷积层 (卷积规模分别为 3×6 和 6×16), 两个池化层, 一个全连接层, 一个 softmax 分类器和一个输出层, 卷积层和池化层的核尺寸分别为 3、2、2、2。

3) DBN 中包含一个输入层, 两个隐含层和一个输出层, 网络结构为 [1 024, 50, 20, 10], 学习率为 0.05, 小批量尺寸为 8, 迭代次数为 500。

4) CNN+2D-CNN 通过特征融合层将 CNN 和 2D-CNN 两个通道全连接层的特征进行拼接。设置 SVM 为分类器, 选取核函数为高斯核, 核参数为 0.01, 惩罚因子为 1。学习率设置为 0.005, 小批量尺寸设置为 8, 总迭代次数为 500, 损失函数为交叉熵。这些方法在样本集上的故障诊断准确率和 F_1 得分如表 4 所示、

表 4 与现有故障诊断结果方法对比

方法编号	故障诊断准确率/%			F_1 得分
	训练集	验证集	测试集	
BO-DKELM	99.90	99.78	99.60	0.99
CNN	100	80.25	83.50	0.84
2D-CNN ^[28]	100	96.00	96.00	0.96
DBN ^[29]	100	72.50	71.50	0.73
CNN+2D-CNN	100	90.00	88.50	0.89

可以看出, 本文所提出的 BO-DKELM 故障诊断能力更强, 在这些模型中, DBN 获得的故障诊断准确率和 F_1 得分的值最低, 与其余方法结果相比, 由于本文所提出的 BO-DKELM 故障诊断模型通过贝叶斯优化使得模型的超参数达到最优, 能够有效提高系统的故障诊断能力, 上述结果验证了本文所提方法的优越性。

4 结束语

本文提出了一种基于 BO-DKELM 的故障诊断方法。通过贝叶斯优化算法对 DKELM 中的隐含层节点个数、正则化参数和核宽度进行优化, 构建 BO-DKELM 模型, 然后利用数据集对优化后的 DKELM 模型进行训练。通过故障诊断结果可以看出, 本文所提方法能够得到最优的超参数, 相比于未优化的 DKELM 算法和传统的 ELM、KELM 等算法, 本文所提方法可以有效提高滚动轴承故障诊断的准确性。然而本文所提方法需要利用优化算法对超参数进行寻优, 因此诊断时间有所增加。如何在保证故障诊断具有较高准确性的前提下进一步缩短诊断时间, 是下一步研究的主要内容。

参考文献:

[1] ZHENG Z, FU J, LU C, et al. Research on rolling bearing fault diagnosis of small dataset based on a new optimal transfer learning network [J]. Measurement: Journal of the International Measurement Confederation, 2021, 177: 109285.

[2] RABAH A, ABDELHAFID K. Rolling bearing fault diagnosis based on improved complete ensemble empirical mode of decom-

- position with adaptive noise combined with minimum entropy deconvolution [J]. *Journal of Vibroengineering*, 2018, 20 (1): 240 - 257.
- [3] ANDREAS K, ROBBERSMYR K G. Cross-correlation of whitened vibration signals for low-speed bearing diagnostics [J]. *Mechanical Systems and Signal Processing*, 2019, 118: 226 - 244.
- [4] CHEN Z, CEN J, XIONG J. Rolling bearing fault diagnosis using time-frequency analysis and deep transfer convolutional neural network [J]. *IEEE Access*, 2020, 8: 150248 - 150261.
- [5] UDMALE S, SINGH S. Application of spectral kurtosis and improved extreme learning machine for bearing fault classification [J]. *IEEE Transactions on Instrumentation and Measurement*, 2019, 68 (11): 4222 - 4233.
- [6] KANNAN V, LI H, DAO D V. Demodulation band optimization in envelope analysis for fault diagnosis of rolling element bearings using a real-coded genetic algorithm [J]. *IEEE Access*, 2019, 7: 168828 - 168838.
- [7] ROY S S, DEY S, CHATTERJEE S. Autocorrelation aided random forest classifier based bearing fault detection framework [J]. *IEEE Sensors Journal*, 2020, 20 (18): 10792 - 10800.
- [8] SAMANTA A, CHOWDHURI S, WILLIAMSON S S. Machine learning-based data-driven fault detection/diagnosis of lithium-ion battery: a critical review [J]. *Electronics*, 2021, 10 (11): 1309.
- [9] XIA Y, XU Y. A transferrable data-driven method for IGBT open-circuit fault diagnosis in three-phase inverters [J]. *IEEE Transactions on Power Electronics*, 2021, 70 (12): 13474 - 13488.
- [10] MA B, CAI W, HAN Y, et al. A novel probability confidence CNN model and its application in mechanical fault diagnosis [J]. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70: 3517111.
- [11] HUANG G B, ZHOU H, DING X, et al. Extreme learning machine for regression and multiclass classification [J]. *IEEE Transactions on Systems Man & Cybernetics Part B*, 2012, 42 (2): 513 - 529.
- [12] LENDASSE A, HE Q, MICHE Y, et al. Advances in extreme learning machines (ELM2013) [J]. *Neurocomputing*, 2014, 149: 158 - 159.
- [13] CAMBRIA E, HUANG G B, KASUN L, et al. Extreme learning machines [trends & controversies] [J]. *IEEE Intelligent Systems*, 2013, 28 (6): 30 - 59.
- [14] HAN F, HUANG D S. Improved extreme learning machine for function approximation by encoding a priori information [J]. *Neurocomputing*, 2006, 69 (16/17/18): 2369 - 2373.
- [15] YAN L, WANG H, WANG H, et al. An improved OS-ELM based Real-time prognostic method towards singularity perturbation phenomenon [J]. *Measurement*, 2021 (2): 109673.
- [16] SUN Y, XU B, WANG X. Pseudo fourth-order moment based bearing fault feature reconstruction and diagnosis [J]. *ISA Transactions*, 2021, 118: 238 - 246.
- [17] GAO W, WAI R J, QIAO S P, et al. Mechanical faults diagnosis of high-voltage circuit breaker via hybrid features and integrated extreme learning machine [J]. *IEEE Access*, 2019, 7: 60091 - 60103.
- [18] LEI Z, ZHOU Y, SUN B, et al. An intrinsic timescale decomposition-based kernel extreme learning machine method to detect tool wear conditions in the milling process [J]. *The International Journal of Advanced Manufacturing Technology*, 2020, 106 (3): 1203 - 1212.
- [19] CHEN Y Q, FINK O, SANSAVINI G. Combined fault location and classification for power transmission lines fault diagnosis with integrated feature extraction [J]. *IEEE Transactions on Industrial Electronics*, 2018, 65: 561 - 569.
- [20] LI K, XIONG M, LI F, et al. A novel fault diagnosis algorithm for rotating machinery based on a sparsity and neighborhood preserving deep extreme learning machine [J]. *Neurocomputing*, 2019, 350: 261 - 270.
- [21] RODRIGUEZ N, LAGOS C, CABRERA E, et al. Extreme learning machine based on stationary wavelet singular values for bearing failure diagnosis [J]. *Studies in Informatics & Control*, 2017, 26 (3): 287 - 294.
- [22] YE Q, PAN H, LIU C. A framework for final drive simultaneous failure diagnosis based on fuzzy entropy and sparse bayesian extreme learning machine [J]. *Computational Intelligence & Neuroscience*, 2015: 427965.
- [23] YAO G, WANG Y, BENBOUZID M, et al. A hybrid gearbox fault diagnosis method based on GWO-VMD and DE-KELM [J]. *Applied Sciences-Basel*, 2021, 11 (11): 4996.
- [24] VINCENT P, LAROCHELLE H, LAJOIE I, et al. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion [J]. *Journal of Machine Learning Research*. 2010, 11 (12): 3371 - 3408.
- [25] LI Z, WANG Y, MA J. Fault diagnosis of motor bearings based on a convolutional long short-term memory network of bayesian optimization [J]. *IEEE Access*, 2021, 9: 97546 - 97556.
- [26] SNOEK J, RIPPEL O, SWERSKY K, et al. Scalable bayesian optimization using deep neural networks [J]. *Statistics*, 2015: 1861 - 1869.
- [27] ZHAO B, ZHANG X, HAI L, et al. Intelligent fault diagnosis of rolling bearings based on normalized CNN considering data imbalance and variable working conditions [J]. *Knowledge-Based Systems*, 2020, 199: 105971.
- [28] PENG X J, ZHANG B K, GAO D. Research on fault diagnosis method of rolling bearing based on 2DCNN [C] // *Proceeding of 2020 Chinese Control And Decision Conference (CCDC)*. Hefei: IEEE, 2020: 693 - 697.
- [29] MA L, YANG Y, WANG H. DBN based automatic modulation recognition for ultra-low SNR RFID signals [C] // *Proceeding of 35th Chinese Control Conference*, Chengdu, 2016: 7054 - 7057.