

DOI:10.19651/j.cnki.emt.2209263

# 基于改进 AAKR 的风电机组齿轮箱状态监测<sup>\*</sup>

田雯雯 吕丽霞 冯雪凯 王梓齐

(华北电力大学控制与计算机工程学院 河北 保定 071000)

**摘要:** 针对自联想核回归(AAKR)算法在计算相似度时未考虑状态向量中各元素对欧氏距离贡献程度不一、模型参数常依据主观经验进行标定而导致模型精度较低的问题,提出基于旗鱼优化(SFO)的改进 AAKR 算法建立齿轮箱正常行为模型的非参数建模方法。首先,以全参数等间隔划分方法构建记忆矩阵;其次,在 AAKR 模型中引入距离权重系数并通过 SFO 算法对 AAKR 模型中的宽度系数和距离权重系数进行优化;最后基于滑动窗口和残差数据构造健康指数实现风电机组齿轮箱的状态监测。以某台 2 MW 风电机组实测数据为例进行验证,结果表明,相比于传统 AAKR、加权 AAKR 和稳健状态估计模型,所提算法平均精度分别提高了 1.55%、0.6%、0.76%,在故障预警时通过所构造的健康指数能够更灵敏、准确的反映齿轮箱的早期故障及其发展趋势。

**关键词:** 旗鱼优化算法;自联想核回归算法;加权欧氏距离;齿轮箱状态监测

**中图分类号:** TM315 **文献标识码:** A **国家标准学科分类代码:** 470.4047

## Condition monitoring of wind turbine gearbox based on improved AAKR

Tian Wenwen Lyu Lixia Feng Xuekai Wang Ziqi

(School of Control and Computer Engineering, North China Electric Power University, Baoding 071000, China)

**Abstract:** The auto associative kernel regression (AAKR) algorithm does not consider the contribution of each element in the state vector to the Euclidean distance when calculating the similarity, and the model parameter is often calibrated based on subjective experience. As a result, the accuracy of the model is relatively low. A non parametric modeling method for establishing the normal behavior model of gearbox is proposed based on the SFO algorithm and the modified AAKR algorithm. Firstly, the memory matrix is constructed by full parameter equal interval partition method. Secondly, the distance weight coefficient is introduced into the AAKR model, and the width coefficient and distance weight coefficient in the AAKR model are optimized by SFO algorithm. Finally, the health index is constructed based on sliding window and residual data to realize the condition monitoring of wind turbine gearbox. Taking the measured data of a 2 MW wind turbine as an example, the results show that compared with the traditional AAKR, weighted AAKR and robust state estimation model, the average accuracy of the proposed algorithm is improved by 1.55%, 0.6% and 0.76% respectively. In fault early warning, the constructed health index can more sensitively and accurately reflect the early fault and development trend of gearbox.

**Keywords:** sailfish optimization algorithm; auto associative kernel regression algorithm; weighted euclidean distance; gearbox condition monitoring

## 0 引言

风电机组工作环境较为恶劣,很多环境因素都会导致机组发生故障,如风速波动频繁、地域温差大等。调查研究表明,齿轮箱的故障频率在所有主要设备中排在第4位,仅次于变频器、发电机和变桨系统,在维修过程中消耗了大量人力物力<sup>[1]</sup>。因此通过行之有效的方法对齿轮箱状态进行

监测,进而完成故障预警是十分有必要的。

研究齿轮箱故障的数据主要来源于两种:1)监控和数据采集(supervisory control and data acquisition, SCADA)系统;2)振动数据。由于当前风电场均配备有 SCADA 系统及相关软硬件,极易获得 SCADA 数据<sup>[2]</sup>,因此基于 SCADA 数据分析的风电机组故障预警方法成为了研究热点。文献[3]通过最小二乘及离散度分析并以功率特性曲

收稿日期:2022-03-14

\* 基金项目:中央高校基本科研业务费(2020JG006,2020MS117)项目资助

线为标准,对 SCADA 数据进行预处理,用风电机组预处理后剩余的正常工况数据训练神经网络模型,一定程度上实现了模型精度的提升。但神经网络易出现“过拟合”现象,预测精度相对较低;文献[4]在 SCADA 数据软聚类分析时使用了 GMM 算法,完成了风电机组的运行工况辨识,并且在每个子运行工况中分别建立了 DAE 模型,完成了对输入数据的重构,但该模型处理时间相对较长,不适合工程实际应用;文献[5]通过粒子群算法对双支持向量机的核参数和惩罚因子进行寻优,提高了模型预测精度,但 PSO 算法寻优精度较低、收敛速度较慢;文献[6]利用人工鱼群算法对径向基神经网络建模参数进行优化,克服了参数随机性影响,但人工鱼群算法只适用于一些精度要求不高的场合,适用范围受限。

以上状态监测方法中,使用的建模方法大都为参数建模方法<sup>[7]</sup>。参数建模方法主要有支持向量机<sup>[8]</sup>、神经网络<sup>[9]</sup>、深度学习<sup>[10]</sup>等,这类方法具有训练速度快、方法简便等优点,但是存在训练过程随机性以及模型鲁棒性较差等问题,使得运用于风电机组故障预警较为困难。非参数建模方法主要有 AAKR<sup>[11]</sup>、非线性状态估计(nonlinear state estimation technique, NSET)<sup>[12]</sup>等。其中 AAKR 通过计算观测向量和记忆矩阵状态向量之间的相似度对当前观测向量中出现的偏差进行修正,在建模时能够只依托于正常状态下的检测参数而无需故障数据即可实现设备的状态监测,快捷高效。文献[13]提出一种加权 AAKR 算法建立了设备的状态估计模型,有较好的准确性,但加权 AAKR 的权重乘法因子向量仅凭借经验人为规定,随机性较大;文献[14]对 AAKR 状态估计模型进行改进,用稳健距离代替传统的马氏距离,提高了自回归模型的稳健性和抗污染能力。但该方法将距离算子中误差最大的参数直接剔除,难免会对模型造成一定影响。

以上研究在一定程度上增强了 AAKR 的实用性,但在计算观测向量和历史向量间的相似度时,并未有效解决状态向量中各元素对欧氏距离贡献程度不一的问题,且宽度系数取值也会很大程度影响建模精度,因此选择合适的寻优算法确定参数值很有必要。

综上,本文将 AAKR 算法中传统欧氏距离加入权重系数以此克服不良参数的偏差污染,并选用收敛速度快的旗鱼优化算法对宽度系数及距离权重系数进行优化。以全参数等间隔划分方法构建记忆矩阵,采用 SFO 算法对 AAKR 算法中核函数宽度系数以及欧氏距离系数进行参数优化。最后以某 2 MW 风电机组 SCADA 采集数据为例,对机组故障停机前工况进行实验。依据验证集数据确定预警阈值,以测试集残差为依据构造健康指数指标,实现了风电机组齿轮箱状态监测。

## 1 AAKR 模型

### 1.1 AAKR 算法原理

作为多变量状态估计的 AAKR 算法,其原理是对大量

的正常历史数据进行分析,通过筛选出的典型工况数据构造设备正常运行工况下的记忆矩阵。它使用历史的无故障观测数据重构当前观测数据来校正其中出现的误差,通过比较当前观测向量与存储的历史正常状态向量之间的差异对当前状态进行判断<sup>[15-16]</sup>,当目标参数预测值和实际值之差越过报警限并持续越限时即可对设备进行预警。

由大量历史数据筛选出的典型工况向量存储在记忆矩阵  $\mathbf{X}$  中,对于有  $n$  个观测点的  $m$  个观测值数据,该记忆矩阵可写为:

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,m} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,m} \end{bmatrix} \quad (1)$$

式中:  $x_{i,j}$  是第  $i$  个观测数据中的第  $j$  个观测值。

因此,当前时刻的观测向量可以表示为  $1 \times m$  的向量:

$$\mathbf{x} = [x_1 \quad x_2 \quad \cdots \quad x_m] \quad (2)$$

观测向量通过记忆矩阵完成重构得到经验模型,其具体实现分为以下 3 步:

1) 在实时监测过程中,对于数据集中每一个新的观测状态向量  $\mathbf{x} = [x_1 \quad x_2 \quad \cdots \quad x_m]$ ,需要计算此状态向量与记忆矩阵  $\mathbf{X}$  中每个状态向量间的距离,距离测度采用欧氏距离衡量,计算公式如下:

$$d_i(\mathbf{X}_i, \mathbf{x}) = \sqrt{(X_{i,1} - x_1)^2 + (X_{i,2} - x_2)^2 + \cdots + (X_{i,m} - x_m)^2} \quad (3)$$

以上计算出的  $\mathbf{d}$  为  $n \times 1$  的距离向量:

$$\mathbf{d} = [d_1 \quad d_2 \quad \cdots \quad d_n]^T \quad (4)$$

2) 将距离向量  $\mathbf{d}$  中的距离参数代入高斯核函数计算出相应权值,并得出权值向量:

$$w_i = \frac{1}{\sqrt{2ph}} e^{-\frac{d_i^2}{2h^2}} \quad (5)$$

式中:  $h$  是核函数的宽度系数,  $h$  大小与其对应的权值成反比关系。

3) 通过式(5)得到的权值向量  $w_i$  与记忆矩阵计算出预测向量,公式为:

$$\hat{\mathbf{x}} = \frac{\sum_{i=1}^n (w_i \cdot \mathbf{X}_i)}{\sum_{i=1}^n w_i} \quad (6)$$

### 1.2 改进 AAKR 算法

AAKR 在计算观测向量与记忆矩阵各状态向量之间相似性时用了欧氏距离。但传统欧氏距离中各元素的权重相同,为去除因扩大权重对监测效果带来的不利影响,使欧氏距离能够更加精确的反映观测向量与历史向量之间的贴近程度,本文在计算距离时对各元素加入权重系数  $b_i$ , 即:

$$d_i(\mathbf{X}_i, \mathbf{x}) = \sqrt{\sum_{k=1}^m b_k (X_{i,k} - x_k)^2} \quad (7)$$

此外,模型的连续性受到高斯核函数中宽度系数  $h$  的较大影响,模型在某些带宽下可以正常产生重构向量,相反,在某些带宽下则可能因无法重构而导致预测值不连续,因此选择 SFO 算法对宽度系数  $h$  和距离权重系数进行优化。

## 2 改进 AAKR 状态监测方法

### 2.1 SFO 基本原理

2019 年提出的旗鱼优化算法是一种收敛速度快、寻优能力强的新型寻优算法<sup>[17]</sup>,能够有效避免局部寻优问题。SFO 通过模仿旗鱼捕食沙丁鱼这一过程实现参数寻优,首先定义两个种群分别模拟捕猎者和狩猎者群体;其次捕猎者通过轮换攻击的方式破坏分组猎物的集体抵御;最后更新猎物在搜索空间中的运动,并且为使结果变得更优整个寻优过程中允许捕猎者捕捉合适的猎物。具体流程如图 1 所示。

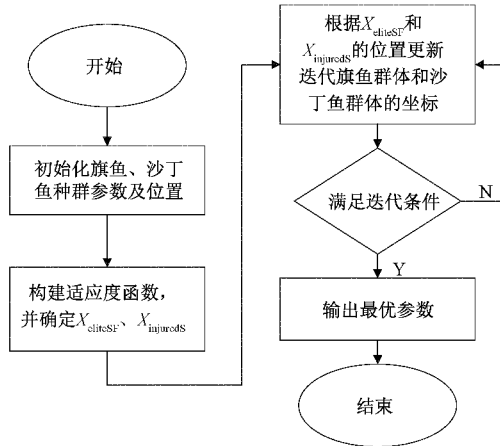


图 1 旗鱼优化算法流程图

1) 旗鱼和沙丁鱼种群初始化。用矩阵  $X_{SF}$  记录所有旗鱼的位置,矩阵  $X_s$  记录所有沙丁鱼的位置。并定义矩阵  $X_{eliteSF}$  和  $X_{injuredF}$  分别用来记录旗鱼适应度值最好的种群和沙丁鱼适应度值最好的种群。

2) 分别计算旗鱼和沙丁鱼的适应度值,将每次迭代后所得的最优适应度值以及对应位置记录下来。为保证预警的时效性和精度,同时确保模型重构向量的连续性,在优化系数时本文采用如下适应度函数:

$$f = RMSE + \sqrt{\sum_{i=1}^m b_i^2} \quad (8)$$

式中:  $RMSE$  是基于测试集所得的均方根误差,  $b_i$  是距离权重系数。

适应度函数中对权重系数采用了二范数约束,其结果越小表示预测结果越逼近实际值,建模精度越高。

3) 更新旗鱼和沙丁鱼的位置信息。其中沙丁鱼位置更新时需要依据攻击力度的大小,若攻击力度值大于 0.5,需对其全部位置进行更新;若攻击力度小于 0.5,则计算沙丁

鱼数量  $\alpha$ 、维度  $\beta$ , 只对部分位置进行更新。

4) 比较旗鱼和沙丁鱼位置,若沙丁鱼位置比旗鱼更优,则将此时沙丁鱼位置赋给旗鱼,并去除沙丁鱼位置信息。

$$X_{SF}^i = X_s^i f(S_i) < f(SF_i) \quad (9)$$

式中:  $X_{SF}^i$ 、 $X_s^i$  分别是第  $i$  次迭代中旗鱼、沙丁鱼的位置;  $f(SF_i)$ 、 $f(S_i)$  分别是旗鱼、沙丁鱼的适应度函数值。

5) 计算所有适应度值,并更新记录最优适应度值和位置。

6) 判断是否符合迭代终止条件,即所有沙丁鱼的适应度值均优于所有旗鱼的适应度值,如果符合,则  $X_{injuredF}$  所对应的适应度值为结果,对应最优旗鱼坐标即为模型参数最优值,否则重复步骤 2)~5)。

### 2.2 改进 AAKR 状态监测流程

基于改进 AAKR 的风电机组齿轮箱状态监测方法由数据预处理、模型建立、模型改进及参数优化、实时状态监测四部分组成,其流程如图 2 所示。

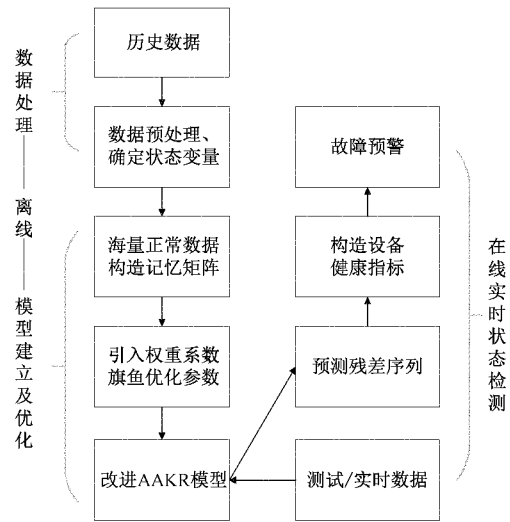


图 2 状态预警流程

1) 数据预处理:取机组正常运行工况下的一段历史数据,剔除风速在切入风速和切出风速以外、有功功率非正以及数据缺失的异常点,并依据皮尔逊相关系数选取模型所需的状态变量;由于各状态变量都具有不同的单位,为防止模型预测精度受到量纲影响,对所有数据做归一化处理。

2) 模型建立:从正常运行工况历史数据中选取一段作为训练集,对其以全参数等间隔划分方法构建记忆矩阵完成模型建立,具体步骤见 2.3 节。

3) 模型改进及参数优化:设置旗鱼优化算法参数,以验证集数据为依据对核函数宽度系数和欧氏距离权重系数寻优,完成模型优化。

4) 实时状态监测:基于 SCADA 系统的测试数据或现场采集实时数据,计算观测向量中齿轮箱轴承温度预测值,进而得到预测值和实际值间的残差,并对残差序列作进一步处理,将其构造成现场人员能够直观感受且具有一定物

理意义的健康指数以实现风电机组齿轮箱状态监测。

### 2.3 构建记忆矩阵

记忆矩阵  $X$  直接影响  $\Delta$ AKR 模型的实用性,而记忆矩阵包含的历史状态向量(测点)的数量决定了 AAKR 模型能够观测的向量范围。历史状态向量过少会导致 AAKR 模型不精确,无法反映一些细小的设备故障,进而 AAKR 模型的可信度下降。历史状态向量过多会导致记忆矩阵过大,重构观测向量时计算量剧增,模型的实时性与实用性得不到充分发挥。

就单一状态包含的参数个数而言,风电机组参数繁多,为保证模型运算效率,在齿轮箱监测中常结合相关性及设备运行原理,选取与目标参数即齿轮箱轴承温度关系较大的参数构成状态向量。

对于 AAKR 模型,不同的历史向量矩阵和带宽  $h$  决定了模型的准确性。并且作为观测向量的重构依据,记忆矩阵必须尽可能的包含齿轮箱的所有工况,否则将出现较大偏差甚至模型失效,此时应当对记忆矩阵进行更新,记忆矩阵构建过程如图 3 所示。

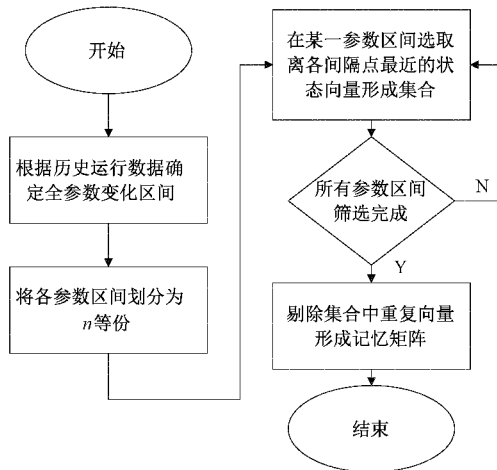


图 3 记忆矩阵构建过程

- 1) 基于训练集数据,分别求得归一化后各输入变量和输出变量的最大值、最小值,确定各特征变量的变化区间。
- 2) 将各特征变量的变化区间划分为 200 等份。
- 3) 针对每个特征变量,选取离各间隔点最近的状态向量作为初始集合。
- 4) 剔除初始集合中的重复状态向量,剩余状态向量即构成最终的记忆矩阵。

## 3 算例实验

### 3.1 研究对象

为了验证本文所提方法的有效性和实用性,以福建省某风电场机组齿轮箱故障之前的一段运行数据为例进行研究。该风电机组的基本信息如下:2 MW 双馈式风力发电机,切入风速 4 m/s,切出风速 25 m/s,齿轮箱为一级行星齿轮和二级螺旋齿轮结构,SCADA 系统记录数据间隔时

间为 10 min。

该风电场机组因齿轮箱故障于 2016 年 7 月 13 日 10:20 停运,经维修于 7 月 18 日 9:30 机组正常运行。导出 SCADA 系统中 2016 年 1 月 1 日 0:00~7 月 13 日 10:20 之间的所有运行数据,其中有参考价值的运行参数共有 8 个,分别为风速、叶轮转速、发电机转速、环境温度、风向角、无功功率、有功功率、齿轮箱轴承温度。

### 3.2 数据预处理及构建记忆矩阵

SCADA 系统记录的风电机组历史运行数据中往往包含着数量巨大的非正常数据,当基于数据进行建模时,这些异常数据的存在会给建模过程及建模精度带来较大的影响,因此在建模前我们需要对电厂原始数据进行预处理,剔除风速小于 4 m/s、大于 25 m/s、有功功率非正以及数据缺失的异常点,再按照拉依达准则对数据进行筛选,最终得到 14 000 组数据进行实验。

取剩余数据中第 1~5 000 号样本作为训练集,用于建立记忆矩阵;取 5 001~7 000 号样本为验证集,用于评价状态变量模型性能及求取阈值;取 7 001~14 000 号数据为测试集,用于检验模型故障预警能力。本实验基于 MATLAB 2019(运行于 AMD Ryzen 7 4800H 16.0 GB RAM 的 PC 机)进行。

为了有效减少模型输入的冗余信息,使得  $\Delta$ AKR 模型的泛化能力实际增强,在提高模型运算效率的同时更好地反映风电机组的运行状态,需要对  $\Delta$ AKR 模型的记忆矩阵  $X$  单一状态包含的变量进行筛选。为防止量纲不同对建模精度产生影响,先对各参数进行归一化处理,继而分别计算每个参数与齿轮箱轴承温度的皮尔逊相关系数,并列各参数的变化范围如表 1 所示。

表 1 运行参数变化范围和皮尔逊系数

运行参数	变化范围	皮尔逊系数
风速/( $m \cdot s^{-1}$ )	[4.00,14.09]	0.853
叶轮转速/( $r \cdot min^{-1}$ )	[9.00,15.03]	0.931
发电机转速/( $r \cdot min^{-1}$ )	[1 001.3,1 687.2]	0.924
环境温度/ $^{\circ}C$	[2.0,30.0]	-0.214
风向角/ $^{\circ}$	[6.9,353.7]	-0.144
无功功率/V <sub>ar</sub>	[-0.04,0.05]	-0.110
有功功率/kW	[50.7,2 007.3]	0.834
齿轮箱轴承温度/ $^{\circ}C$	[43.88,63.01]	1.000

由表 1 可知,风速、叶轮转速、发电机转速以及有功功率 4 个参数与齿轮箱轴承温度之间的皮尔逊相关系数均较大,表明这 4 个变量均与齿轮箱轴承温度存在较强的相关性。但是发电机转速和叶轮转速之间冗余性较大,二者只选其一即可;虽然环境温度与齿轮箱轴承温度间的皮尔逊相关系数较小,但是齿轮箱工作易受到环境温度变化的影响,因此该变量也作为状态变量。综上,本文选用风速、发



电机转速、环境温度、有功功率和齿轮箱轴承温度 5 个变量作为记忆矩阵的状态变量。

在选定模型状态变量后,根据图 3 所示流程对处理后的 1~5 000 组数据进行训练构建记忆矩阵,最终确定齿轮箱状态监测参数记忆矩阵包含的状态向量有 777 组,至此完成 AAKR 模型的建立。

### 3.3 模型改进及参数寻优

由于传统 AAKR 模型并未考虑到状态向量中各元素对欧氏距离贡献程度不一的问题,导致模型精度较低。为更加精确的反映观测向量与历史向量之间的贴近程度,去除向量中元素因扩大权重对齿轮箱监测效果带来的不利影响,本文在计算距离时对各元素加入权重系数,且采用旗鱼优化算法对各距离系数和宽度系数  $h$  进行优化。

依据经验和仿真测试,当核函数宽度系数  $h$  取 0.1 时 AAKR 模型初步显示出良好的预测效果,因此取其初值为 0.1。旗鱼优化算法参数如下:迭代次数 100 次,种群规模 50(其中旗鱼数量占 30%,沙丁鱼数量占 70%),攻击力度  $A$ 、 $e$  分别为 4、0.001。优化分为单参数优化和联合优化两部分,其中单参数优化分为仅优化宽度系数  $h$  和仅优化各距离系数。

保持 3.2 节中的记忆矩阵不变,利用验证集对各类参数进行优化,模型优化后参数如表 2 所示。

表 2 改进 AAKR 模型优化后参数

参数	仅优化 $h$	仅优化距离系数 $b_i$	联合优化
$h$	0.027 3	0.1	0.011 4
$b_1$	1	0.166 2	0.114 2
$b_2$	1	0.193 8	0.202 1
$b_3$	1	0.378 6	0.276 2
$b_4$	1	0.053 6	0.074 8
$b_5$	1	0.519 1	0.493 9

以联合优化为例,绘制其收敛过程如图 4 所示。

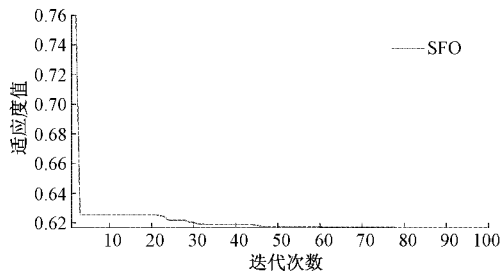


图 4 联合优化适应度收敛曲线

由图 4 可知,旗鱼优化收敛速度较快,在第 3 次迭代时已经基本达到最优解,最终得到最优适应度值为 0.618。

## 4 结果及分析

为了验证改进 AAKR 模型精度及实时性,从模型预测

精度和状态监测效果两方面进行对比分析。

模型精度对比分为以下两部分:

- 1) 优化前后  $\Delta$ AKR 模型精度对比分析;
- 2) 多模型(加权 AAKR 模型、稳健状态估计、改进 AAKR 模型)精度对比分析。

状态监测效果对比分为以下两部分:

- 1) 基于残差曲线的齿轮箱状态监测;
- 2) 基于健康指数的齿轮箱状态监测。

### 4.1 模型预测精度对比分析

1) 优化前后 AAKR 模型精度对比分析

基于训练样本和验证样本,将表 2 中 3 种情况下的优化结果依次作为 AAKR 模型的参数,分别计算模型训练误差和验证误差(二者均为均方根误差)结果如表 3 所示。其中均方根误差计算公式如下:

$$RMSE = \sqrt{\frac{1}{L} \sum_{i=1}^L (\hat{y}_i - y_i)^2} \quad (10)$$

式中:  $\hat{y}_i$  是齿轮箱轴承温度预测值,  $y_i$  是齿轮箱轴承温度实际值,  $L$  是观测次数。

表 3 优化前后 AAKR 模型的训练及验证结果

参数	训练误差	验证误差
优化前	0.016 8	0.018 1
仅优化带宽 $h$	0.014 3	0.017 5
仅优化距离系数 $b_i$	0.009 7	0.010 3
联合优化	0.001 8	0.002 2

由表 3 可知,优化前模型的训练和验证误差均较大,模型精度较低;对模型进行单参数优化后精度相对有所提高,且仅优化距离权重系数比仅优化带宽  $h$  的精度更高,说明在模型中引入距离权重系数可以切实克服不良参数的影响,对于改善模型精度有较大帮助;联合优化的精度则明显高于优化前及单参数优化的精度,进一步说明引入距离权重系数并对其和  $h$  同时优化能够有效改进 AAKR 模型,相比传统 AAKR 模型有着较大优势。

2) 多模型精度对比分析

选取加权  $\Delta$ AKR 模型及稳健状态估计两种模型与本文所提改进算法进行对比。为保证每种情况下模型均为最优,使用旗鱼优化算法分别对每个模型的带宽  $h$  进行优化,得到加权 AAKR 模型对应带宽系数为 0.101 8,稳健状态估计算法对应带宽系数为 0.072 2,改进 AAKR 模型对应参数即为 3.3 节中联合优化对应参数值。

基于训练样本和验证样本,对各状态变量模型进行训练和验证,分别计算模型训练误差和验证误差(二者均为均方根误差)结果如表 4 所示。

由表 4 可知,相比于加权 AAKR 和稳健状态估计模型,改进 AAKR 模型精度更高,可能是因为加权 AAKR 模型精度受到权重乘法因子向量取值的限制,稳健状态估计

表 4 不同模型训练及验证结果

模型	训练误差	验证误差
加权 AAKR	0.007 1	0.008 9
稳健状态估计	0.008 6	0.010 6
改进 AAKR	0.001 8	0.002 2

模型在计算距离时直接剔除了误差最大的参数,对模型预测效果产生影响。进一步说明改进 AAKR 算法有效解决了计算距离时各参数贡献程度不一的问题,更准确地描述了观测向量和状态向量间的相似程度,大幅度提升了模型精度。

#### 4.2 齿轮箱状态监测效果对比分析

残差是一个随机变量,一般服从正态分布。根据  $3\sigma$  准则,在正态分布中,数据点有 99.74% 分布在  $\mu \pm 3\sigma$  范围内,本文假定分布在  $\mu \pm 3\sigma$  范围外的点为异常点,通过验证集求取报警上下阈值,其计算公式为<sup>[18]</sup>:

$$UCL = \mu + 3\sigma \quad (11)$$

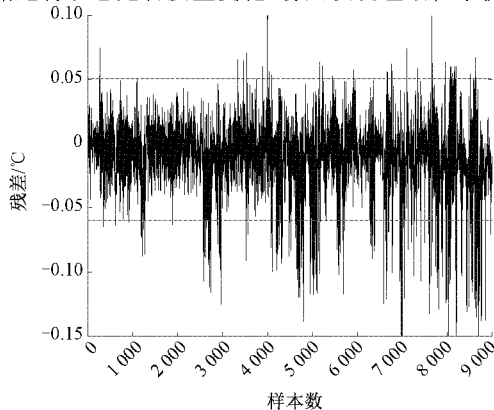
$$LCL = \mu - 3\sigma \quad (12)$$

式中:  $UCL$  为上界阈值,  $LCL$  为下界阈值,  $\mu$  和  $\sigma$  为基于验证集得到的齿轮箱轴承温度预测残差的均值与标准差。

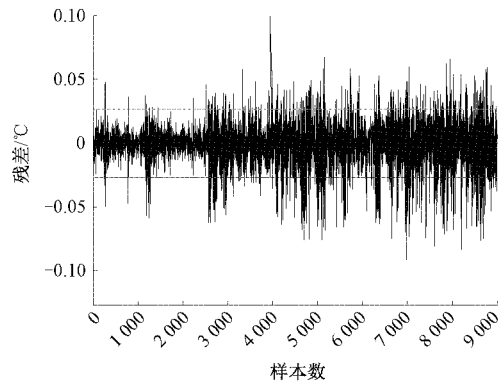
##### 1) 基于残差曲线的齿轮箱状态监测

取传统 AAKR、加权 AAKR、稳健状态估计以及改进 AAKR 4 个模型,基于验证和测试数据集,绘制齿轮箱轴承温度预测值与实际值残差曲线如图 5 所示。根据式(11)和(12)计算得到传统 AAKR 模型的报警阈值约为 0.050 1;加权 AAKR 模型的报警阈值约为 0.026 5;稳健状态估计模型的报警阈值约为 0.031 8;改进 AAKR 模型的报警阈值约为 0.006 3。

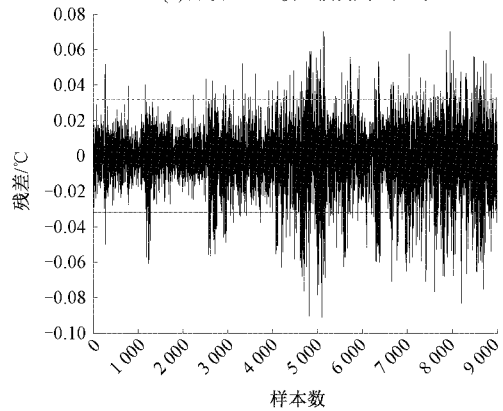
由图 5(a)可知,传统 AAKR 模型在 2 500 个样本点之后残差曲线开始越过阈值线进行预警,但模型精度较低导致残差波动较大,超出阈值次数较少,误报率较高;由图 5(b)和(c)可知,加权 AAKR 和稳健状态估计模型的预测残差在 2 500 个点之后较为频繁地越过阈值线,但与前 2 500 个点之间差距仍不是很明显,现场人员依旧较难判断齿轮箱运行状态是否发生变化,易延误机组最佳维护时机;



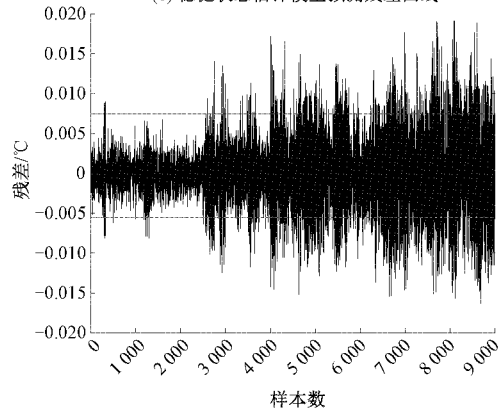
(a) 传统 AAKR 模型预测残差曲线



(b) 加权 AAKR 模型预测残差曲线



(c) 稳健状态估计模型预测残差曲线



(d) 改进 AAKR 模型预测残差曲线

图 5 模型预测残差曲线

由图 5(d)可知,对于改进 AAKR 模型,其残差曲线在 2 500 个点之后频繁超出阈值线,且残差幅值越来越大,相比其他 3 个模型 AAKR 模型,改进 AAKR 预测残差的变化更能反映早期故障持续且不可逆的劣化趋势。

由图 5 还可以看到,改进 AAKR 模型报警阈值明显低于其余 3 个模型,在故障预警中表现为对齿轮箱轴承温度变化更为敏感,故障预警能力增强。

##### 2) 基于健康指数的齿轮箱状态监测

尽管图 5 中残差曲线前后期对比在一定程度上能够说明齿轮箱运行状态发生变化,但是现场人员仅凭借观察残差曲线很难直观判断风电机组是否出现异常。因此,对残

差序列作进一步处理,将其与阈值相结合构造出能够更加直观体现机组运行状态的健康指标是很有必要的。

本文选用健康指数这一指标来进行预警,其计算公式如下:

$$I = \left(1 - \frac{N}{M}\right) \times 100\% \quad (13)$$

式中:  $N$  为滑动窗口中超过预警阈值的残差个数;  $M$  为残差序列的总个数;  $I$  的单位为百分比且值域为  $[0, 100\%]$ 。

当设备处于正常运行状态时,齿轮箱轴承温度预测值和实际值相差较小,残差绝大多数位于预警阈值以内,健康指数  $I$  相对较高;当设备发生故障时,模型预测结果将会显著偏离实际值,窗口中超过阈值的残差个数增多,  $I$  会显著下降。

基于验证和测试数据,分别计算 4.1 节中 4 个模型在各自预警阈值下的健康指数  $I$ , 结果如图 6 所示。(取滑动窗口的长度  $M=1\ 500$ )。

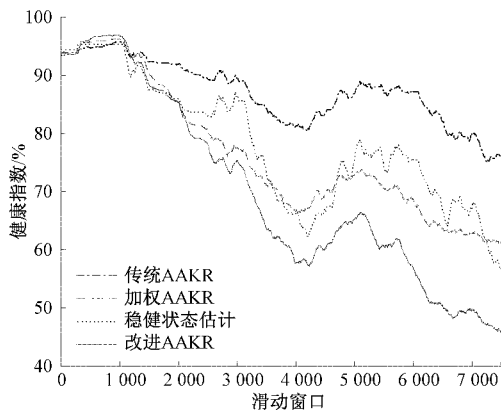


图 6 模型预测健康指数  $I$

由图 6 可知,在构造阈值和健康指数方法相同的前提下,4 种模型健康指数曲线变化差别较大:初始时刻健康指数基本相同,在前 1 500 左右个窗口内改进 AAKR 的健康指数略高于其他 3 个模型的健康指数,这是因为此时机组尚未开始预警,运行状态良好;约 1 500 个窗口起,模型健康指数开始下降,可以看出:稳健状态估计算法健康曲线波动幅度较为明显,单调性差,故障预警效果较差;加权 AAKR 模型单调性略好,健康指数下降趋势更为平稳,但最终健康指数只下降到 63% 左右,不足以说明故障情况;改进 AAKR 模型曲线的下降幅度明显大于其余 3 种模型且单调性较好,从最初的约 94% 最终降至约 47%,下降速度也明显高于其他算法,说明改进 AAKR 能够更准确的反映齿轮箱故障的发展趋势、故障预警能力相对更优。

## 5 结 论

本文针对风电机组齿轮箱状态监测问题,提出了一种基于旗鱼优化和改进 AAKR 算法的风电机组齿轮箱故障预警策略。在传统 AAKR 算法的欧氏距离中引入距离权

重系数,并利用旗鱼优化算法对核函数宽度系数以及欧氏距离权重系数进行优化,实现了风电机组齿轮箱状态监测并得到如下结论:

1) 相比于传统 AAKR、加权 AAKR、稳健状态估计模型,改进 AAKR 模型可以有效解决欧式距离中各参数贡献程度不一的问题,大幅度提升预测精度及故障预警能力。

2) 使用旗鱼优化算法可以有效标定模型参数,优化前、单参数优化、参数联合优化效果依次提升,优化后的模型均能使齿轮箱的早期故障得到更加直观、灵敏的反映。

3) 改进后的 AAKR 模型具有更加良好的残差污染抑制能力,对异常参数诊断具有促进作用,更加适合于工程实际应用。

## 参 考 文 献

- [1] SU C, HU Z. Reliability assessment for Chinese domestic wind turbines based on data mining techniques[J]. Wind Energy, 2017, 21(3): 198-208.
- [2] 金晓航,许壮伟,孙毅,等. 基于生成对抗网络的风电机组在线状态监测[J]. 仪器仪表学报, 2020, 41(4): 68-76.
- [3] 乔福宇,马良玉,马永光. 基于功率曲线分析与神经网络的风电机组故障预警方法[J]. 中国测试, 2020, 46(8): 44-50.
- [4] 王东林,吕丽霞,王梓齐,等. 基于 GMM 工况识别和 DAE 的风机齿轮箱状态监测[J]. 中国测试, 2021, 47(4): 89-95.
- [5] 方贤,铁治欣,崔仕文,等. 基于粒子群优化双支持向量机的 SCR 烟气脱硝效率预测模型[J]. 热力发电, 2018, 47(1): 53-58.
- [6] 金秀章,于静,刘岳. 基于人工鱼群-径向神经网络 NO<sub>x</sub> 预测模型[J]. 动力工程学报, 2021, 41(7): 551-557.
- [7] 刘长良,张书瑶,王梓齐. 基于改进 KNN 回归算法的风电机组齿轮箱状态监测[J]. 中国测试, 2021, 47(1): 153-159.
- [8] 王一鹏,陈学振,李连玉. 基于小波包混合特征和支持向量机的机床主轴轴承故障诊断研究[J]. 电子测量与仪器学报, 2021, 35(2): 59-64.
- [9] 赵宁. 基于神经网络的三相整流电路故障诊断策略研究[J]. 国外电子测量技术, 2021, 40(3): 88-93.
- [10] 徐永财. 深度学习算法的机电设备工作状态检测[J]. 电子测量技术, 2020, 43(11): 34-38.
- [11] 蒋佳炜,胡以怀,李方玉,等. 基于 AAKR 模型的船用低速柴油机状态监测方法[J]. 舰船科学技术, 2021, 43(3): 158-162.
- [12] 郭鹏, DAVID, INFELD, 等. 风电机组齿轮箱温度趋势状态监测及分析方法[J]. 中国电机工程学报, 2011, 31(32): 129-136.
- [13] 刘双白,朱龙飞,仇晓智,等. 基于加权 AAKR 算法的

- 发电设备状态预警技术研究[J]. 热能动力工程, 2020, 35(7):235-241.
- [14] 李刚, 仇晨光, 曹帅, 等. 基于稳健状态估计的设备状态预警研究[J]. 热力发电, 2020, 49(11):1-7.
- [15] AN S H, HEO G, CHANG S H. Detection of process anomalies using an improved statistical leaning framework[J]. Expert System with Application, 2011, 38(3): 1356-1363.
- [16] HU Y, PALME T, FINK O. Fault detection based on signal reconstruction with auto-associative extreme learning machines [J]. Engineering Application of Artificial Intelligence, 2017(57): 105-117.
- [17] 易正元, 苏怀智, 杨立夫. 混凝土坝变形监控模型的随机森林与旗鱼优化组合建模方法[J]. 水电能源科学, 2021, 39(10):106-109, 143.
- [18] 黄荣舟, 汤宝平, 杨燕妮, 等. 基于长短时记忆网络融合 SCADA 数据的风电齿轮箱状态监测[J]. 太阳能学报, 2021, 42(1):235-239.

#### 作者简介

田雯雯, 硕士研究生, 主要研究方向为风电机组故障预警。

E-mail: 1173947627@qq.com