

基于改进 HAC 和 MSET 的给水泵故障预警方法

蔡吉磊¹, 茅大钧¹, 李玉珍²

(1. 上海电力大学 自动化工程学院, 上海 200090; 2. 上海长庚信息技术股份有限公司, 上海 201209)

摘要: 汽动给水泵是火电厂汽水系统重要的辅机设备, 但给水泵运行环境恶劣, 且故障种类多, 导致给水泵故障频发, 对电厂的经济性和安全性造成了影响。对此, 提出了一种基于改进的层次凝聚聚类 (Hierarchical Agglomerative Clustering, HAC) 和多元状态估计 (Multivariate State Estimation, MSET) 的给水泵故障预警方法, 首先选取与给水泵故障相关测点的历史数据, 使这些数据可以涵盖给水泵正常运行时所有动态变化情况; 通过主元分析法 (Principal Component Analysis, PCA) 和改进的小波去噪对数据进行预处理, 实现数据的降噪和降维, 再采用层次凝聚聚类算法构建记忆矩阵 D , 并引入了距离检测的方法对 HAC 进行改进, 以此构建 MSET 预警模型, 最后通过滑动窗口法分析残差, 实现故障预警, 并和最小二乘支持向量机 (LS-SVM) 预警模型比较分析, 经验证该模型可以准确高效地实现给水泵的早期预警。

关键词: 小波变换; 主元分析; 层次凝聚聚类; 多元状态估计; 记忆矩阵

中图分类号: TM621 文献标识码: A DOI: 10.16146/j.cnki.rndlgc.2023.08.022

[引用本文格式] 蔡吉磊, 茅大钧, 李玉珍. 基于改进 HAC 和 MSET 的给水泵故障预警方法[J]. 热能动力工程, 2023, 38(8): 176-184. CAI Ji-lei, MAO Da-jun, LI Yu-zhen. Feed water pump fault early warning method based on improved HAC and MSET[J]. Journal of Engineering for Thermal Energy and Power, 2023, 38(8): 176-184.

Feed Water Pump Fault Early Warning Method based on Improved HAC and MSET

CAI Ji-lei¹, MAO Da-jun¹, LI Yu-zhen²

(1. College of Automation Engineering, Shanghai University of Electric Power, Shanghai, China, Post Code: 200090;
2. Shanghai Changgeng Information Technology Co., Ltd., Shanghai, China, Post Code: 201209)

Abstract: The steam-driven feed water pump is an important auxiliary equipment for the steam-water system of the thermal power plant, but the operation environment of the feed water pump is harsh, and there are many types of failures, resulting in frequent failure of the feed water pump, which has an impact on the economy and safety of the power plant. Therefore, this paper proposes a feed water pump fault early warning method based on improved hierarchical agglomerative clustering (HAC) and multivariate state estimation technique (MSET). Firstly, the historical data of relative measuring point of feed water pump fault is selected to cover all dynamic changes in the normal operation of the feed water pump; through principal component analysis (PCA) and improved wavelet denoising to preprocess the data to achieve data noise reduction and dimensionality reduction; then, the memory matrix D is constructed by the HAC algorithm, and the distance detection method is introduced to improve the HAC, so as to construct the MSET early warning model; finally, the residual error is analyzed by the sliding window method to realize the fault early warning, and compared with the least squares support vector machine (LS-SVM) early

收稿日期: 2022-05-25; 修订日期: 2022-06-29

基金项目: 上海市“科技创新行动计划”地方院校能力建设专项项目(19020500700)

Fund-supported Project: Shanghai University Science and Technology Innovation Action Plan Local University Capacity Building Project (19020500700)

作者简介: 蔡吉磊(1997-), 男, 上海电力大学硕士研究生.

通讯作者: 茅大钧(1966-), 男, 上海电力大学教授.

warning model. It is verified that the model can accurately and efficiently realize the early warning of the feed water pump.

Key words: wavelet transform, principal component analysis (PCA), hierarchical agglomerative clustering (HAC), multivariate state estimation technique (MSET), memory matrix

引 言

据国家能源局数据显示,2021 年全国火力发电装机容量达 12.96 亿千瓦,火力发电在我国仍占有很大比重。但随着大型火电厂的辅机设备单机容量不断增大,其复杂度也在不断的提升,导致火电厂辅机设备的故障频繁发生。

给水泵是火电厂重要的辅机设备,作用是提升给水压力并将水送入锅炉,锅炉给水泵组主要由小汽轮机驱动的给水泵组成,保障其安全稳定运行是保障机组安全运行的关键一环。给水泵的工作状态会影响整个汽水系统,一旦发生故障,就会导致机组负荷降低,不仅减少了电厂的发电量,还增加了电厂的维修费用。当发生严重故障时,电厂将不得不停机,给火电厂造成巨大的经济损失。

西安热工研究院的徐红伟等人^[1]基于分类算法对给水泵进行故障预测,分别通过 5 种分类算法建立预警模型,其中 BP 神经网络、支持向量机和组合分类算法分类效果更优。吴昕等人^[2]结合了主成分分析和 K 邻近算法的优点,基于某燃气-蒸汽联合循环机组高压给水泵,研究论证了一种高效的给水泵振动预警方法。华北电力大学董玉亮等人^[3]提出基于正交局部保持投影(OPLL)和自组织映射(SOM)的给水泵健康衰退评价方法,可以提前发现给水泵出现的潜在隐患,实现给水泵的早期预警。

目前,国内对于给水泵的故障预警研究方面还较少,还有很多好的预警方法没有被应用到给水泵。由于给水泵工作环境比较复杂,状态变量较多,采用单一的方法很难准确地捕捉到给水泵潜在的隐患,如复杂的环境因素造成模型失效、专家的实际操作经验难以合理化描述等。且给水泵内部的工作机制会随着时间的变化而变化,采用机理来构建预警模型有很多不足之处。因此,本文采用数据挖掘的故障预警方法,通过采集历史数据建立故障预警模型,

采用此方法具有较高的实时性、适应性和准确性。

1 数据预处理

1.1 状态参数的选择

大型机组的给水泵多采用变转速小汽轮机直接驱动给水泵,小汽轮机和给水泵通过联轴器联接在一起的两个设备,在运行时互相影响,故将其作为一个整体,以提高分析结果的准确性和实用性。

根据每台给水泵都有自己规定的扬程和泵级数,给水泵在正常工作时每个测点都必须在一定的范围内,而且某些测点之间有一定的联系。例如,给水泵出口流量过小,无法达到机组负荷的要求,给水泵的叶轮可能堵塞或进水口发生堵塞;给水泵入口压力或入口温度过大时,会导致给水泵发生气蚀现象,若泵长期在汽蚀状态下运行,泵的某些部位会遭到腐蚀破坏;如果给水泵出口母管压力和入口压力的压差过小,无法达到给水泵的扬程,会导致泵不出水;给水泵的轴承温度过高可能是因为轴承进油管滤网堵塞,导致油量减少,使轴承不能得到较好冷却;轴承振动过大可能是润滑失效或推力轴承和油环套之间的间隙不当,导致泵体剧烈振动或产生噪音;小汽轮机的主蒸汽压力和主蒸汽流量大致呈正相关,否则可能存在漏气的现象;小汽轮机的转速与主蒸汽流量的比值应在一个范围之间,若低于这个范围,小汽轮机的叶片发生了磨损、裂纹或腐蚀变形。

针对上述故障特征,本文选取华能集团浙江某电厂的一台汽动给水泵作为研究对象来验证故障预警模型,并从电厂厂级监控系统(Safety Instrumented System, SIS)中选取 24 个与故障特征相关的测点参数作为状态变量^[4],具体测点包括机组负荷、给水泵入口压力、给水泵入口温度、给水泵出口母管压力、再循环阀阀位、给水泵出口流量、驱动端回水温度、自由端回水温度、驱动端轴承 X 向和 Y 向振动、自由端轴承 X 向和 Y 向振动、驱动端轴承温度、自

由端轴承温度、小汽轮机调节阀阀位、小汽轮机转速、小汽轮主蒸汽压力、小汽轮机主蒸汽流量、小汽轮机主蒸汽温度、小汽轮机排汽室温度、小汽轮机排汽压力、小汽轮机排汽温度、泵上部壳体温度及泵下部壳体温度。

1.2 改进的小波变换去噪

电厂在采集各种信号时,会受到外界环境因素的影响,采集的给水泵数据信号会掺杂一些噪声信号。噪声信号大多都是高频信号,直接采用含噪信号会造成较大误差,影响后续预警模型的准确性。因此,必须对初始数据进行降噪。

小波变换(WT)可以有效地对数据进行去噪,其基本思想为:将信号经过小波分解后,得到小波系数,真实信号的小波系数较大,噪声信号的小波系数较小,通过选取一个合适的阈值和阈值函数,小于阈值的小波系数为噪声信号,将其置0,大于阈值的信号保留,从而达到去噪的目的^[5-6]。小波降噪流程如图1所示。

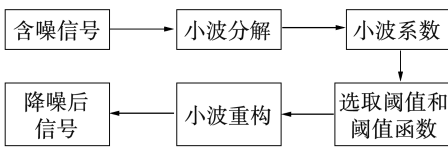


图1 小波降噪流程图

Fig.1 Wavelet denoising flow chart

其中,不同的阈值函数对信号的降噪效果不同,常用的阈值函数公式如下:

硬阈值函数:

$$\hat{d}_{i,j} = \begin{cases} d_{i,j}, & |d_{i,j}| \geq \lambda \\ 0, & |d_{i,j}| < \lambda \end{cases} \quad (1)$$

软阈值函数:

$$\hat{d}_{i,j} = \begin{cases} d_{i,j} - \lambda \operatorname{sgn}(d_{i,j}), & |d_{i,j}| \geq \lambda \\ 0, & |d_{i,j}| < \lambda \end{cases} \quad (2)$$

Garrote 阈值函数:

$$\hat{d}_{i,j} = \begin{cases} d_{i,j} - \frac{\lambda^2}{d_{i,j}}, & |d_{i,j}| \geq \lambda \\ 0, & |d_{i,j}| < \lambda \end{cases} \quad (3)$$

式中: $\hat{d}_{i,j}$ —估计的小波系数; $d_{i,j}$ —原始的小波系数; λ —选定的阈值; i —分解的尺度; j —正整数。

但上述的3种阈值函数去噪时存在一定的不足,硬阈值函数在原始小波系数等于阈值处不是连续的,处理信号时会造成震荡现象;软阈值函数连续性较好,但是会造成固定的偏差,从而使误差较大;Garrote 阈值函数克服了这两种阈值函数的不足,但在处理突变的信号,会使重构信号存在一定的失真,降噪效果仍有不足之处。

为了解决上述问题,选择一种新的阈值函数来改进传统的小波去噪,具体表达式如公式(4)所示。

$$\hat{d}_{i,j} = \begin{cases} d_{i,j} - \frac{\lambda^2}{d_{i,j} \cdot e^{\frac{|d_{i,j}-\lambda|}{N}}}, & |d_{i,j}| \geq \lambda \\ 0, & |d_{i,j}| < \lambda \end{cases} \quad (4)$$

式中: $e^{\frac{|d_{i,j}-\lambda|}{N}}$ —调整量; N —调节因子。

在 Garrote 阈值函数的基础上,新的阈值函数加入了一个调整量,可以通过改变调节因子有效的改变去噪效果。

新构造的阈值函数在 Garrote 阈值函数的基础上加入调整量,控制调节因子 N 来改变函数逼近速度,当降噪后的信号有很多尖峰震荡,说明还存在部分噪声信号,可以适当增大 N 值;当信号特别平滑,可能会造成一部分原始信号丢失,导致误差增大,可以适当减小 N 值。

与传统的 Garrote 阈值函数相比,改进后的阈值函数在处理突变信号和高频信号时,有更高的信噪比和较小的误差,且处理起来灵活方便^[7]。

1.3 PCA 数据降维

因给水泵的状态变量较多,如果采用所有的状态变量建模会使矩阵的规模太大,将会加大后续建模的复杂度,因此必须对数据进行降维,提高预警模型的速度和精度。

主成分分析(PCA)是一种降低数据维度的有效方法。对一个 n 维数据,通过对坐标轴进行旋转,并平移,使数据可以用较少的维度表示。具体是通过对方差矩阵进行特征分解,得到特征值与特征向量,每个特征值代表一个主元,特征向量为该主元的主成分,特征值越大,说明该主元的方差贡献率越高,保留方差贡献率高的主成分,即可实现数据的降维。PCA 算法流程如下:

假设样本数据 X 是一个 $m \times n$ 的矩阵,

$$\mathbf{X}_{m \times n} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \quad (5)$$

式中: m —有 m 个样本点; n —有 n 个状态变量。

由于各个变量间存在量级关系,降维前先将样本数据进行 Z-score 标准化处理:

$$\mathbf{Z}_{ij} = \frac{a_{ij} - \bar{a}_j}{S_j} \quad (6)$$

式中: \bar{a}_j, S_j —第 j 个状态变量所有样本的平均值和标准差,具体计算如下:

$$\bar{a}_j = \frac{1}{m} \sum_{i=1}^m a_{ij} \quad (7)$$

$$S_j = \sqrt{\frac{1}{m-1} \sum_{i=1}^m (a_{ij} - \bar{a}_j)^2} \quad (8)$$

得到标准化矩阵 \mathbf{Z} 为:

$$\mathbf{Z}_{m \times n} = \begin{pmatrix} z_{11} & z_{12} & \cdots & z_{1n} \\ z_{21} & z_{22} & \cdots & z_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ z_{m1} & z_{m2} & \cdots & z_{mn} \end{pmatrix} \quad (9)$$

计算矩阵 \mathbf{Z} 的相关系数矩阵 \mathbf{R} ,可以反应各变量之间的相关性系数。

$$r_{ij} = \frac{\sum_{k=1}^n (z_{ij} - \bar{z}_k)(z_{ij} - \bar{z}_l)}{\sqrt{\sum_{k=1}^n (z_{ij} - \bar{z}_k)^2 \sum_{l=1}^n (z_{ij} - \bar{z}_l)^2}} \quad (10)$$

式中: \bar{z}_k 和 \bar{z}_l —标准化矩阵 \mathbf{Z} 第 k 列和第 l 列的平均值。

得到相关系数矩阵 \mathbf{R} :

$$\mathbf{R}_{n \times n} = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{pmatrix} \quad (11)$$

计算矩阵 \mathbf{R} 的特征值 λ 和特征向量 ν ,按照特征值从大到小进行排序 $\lambda_1 > \lambda_2 > \cdots > \lambda_n$ 。

计算给水泵各监测参数变量的贡献率 c ,公式如下:

$$c_j = \lambda_j / \sum_{j=1}^n \lambda_j \quad (12)$$

计算给水泵各监测参数变量的累积贡献率:

$$C_k = \sum_{j=1}^k \lambda_j / \sum_{j=1}^n \lambda_j \quad (13)$$

在工业上,要求累计贡献率大于 85%,说明这 k 个主元包含了原始数据的绝大部分信息,可以用这 k 个主元来代替原来 n 维数据,从而达到降维的目的^[8]。

2 基于改进 HAC 和 MSET 预警模型构建

2.1 MSET 技术原理

多元状态估计技术(MSET)^[9]是一种比较先进的模式识别技术,最早是由美国阿尔贡国立实验室旨在检测核电厂中传感器、设备以及运行参数的初始劣化点而开发的一种非参数建模方法。目前在很多领域都有被广泛应用。

MSET 通过选取对象正常运行时的历史数据,记为记忆矩阵 \mathbf{D} ,每一行代表一个状态变量,每一列代表某一时刻设备的状态,并挖掘各个状态参数之间的潜在关系,每当输入新的观测值 \mathbf{X}_{obs} ,MSET 通过将观测值和上述构建的记忆矩阵进行比较,并经过计算得到估计值 \mathbf{X}_{est} ,估计向量是记忆矩阵和权值向量的积:

$$\mathbf{X}_{\text{est}} = \mathbf{D} \cdot \mathbf{W} = \mathbf{D} \cdot [w_1, w_2, \dots, w_m]^T \quad (14)$$

权值向量 \mathbf{W} 是观测值与记忆矩阵之间的一种相似性测度,权值向量可通过最小化残差 ε 来获得。残差可由观测向量和估计向量做差得到:

$$\varepsilon = \mathbf{X}_{\text{est}} - \mathbf{X}_{\text{obs}} \quad (15)$$

ε 越小,说明预测结果更加准确,对 ε 求最小平方误差:

$$\begin{aligned} \min \varepsilon &= \min(|\mathbf{X}_{\text{est}} - \mathbf{X}_{\text{obs}}|) \\ &= \min(|\mathbf{X}_{\text{obs}} - \mathbf{D} \cdot \mathbf{W}|) \end{aligned} \quad (16)$$

即:

$$\begin{aligned} \min \varepsilon^2 &= [(\mathbf{X}_{\text{obs}} - \mathbf{X}_{\text{est}})^T \cdot (\mathbf{X}_{\text{obs}} - \mathbf{X}_{\text{est}})] \\ &= \min[(\mathbf{X}_{\text{obs}} - \mathbf{D} \cdot \mathbf{W})^T \cdot (\mathbf{X}_{\text{obs}} - \mathbf{D} \cdot \mathbf{W})] \end{aligned} \quad (17)$$

令 $\frac{\partial \varepsilon^2}{\partial \mathbf{w}} = 0$, 便可得到:

$$\mathbf{W} = (\mathbf{D}^T \cdot \mathbf{D})^{-1} \cdot (\mathbf{D}^T \cdot \mathbf{X}_{\text{obs}}) \quad (18)$$

但是 $(\mathbf{D}^T \cdot \mathbf{D})$ 存在不可逆的情况,因为给水泵各测点之间可能存在一定的相关性,可以通过非线性算子运算来代替上式的点乘,非线性算子可用欧氏距离来计算。

$$\otimes (X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (19)$$

式中: x_i 和 y_i 是矩阵 X 、 Y 中的列向量。

将公式(18)代入公式(14)可得给水泵当天状态对应的估计向量为:

$$X_{\text{est}} = D \cdot (D^T \otimes D)^{-1} \cdot (D^T \otimes X_{\text{obs}}) \quad (20)$$

设备运行时的状态可由输入的状态向量和输出的估计向量之间的残差来反映,当残差超过规定的阈值,给水泵可能发生了故障^[10-11]。

2.2 改进的 HAC 构建记忆矩阵

构建记忆矩阵 D 是 MSET 技术最重要的一环,对计算效率和估计准确性都有重大的影响。最理想的方式是选取所有的历史数据来构建记忆矩阵 D ,但是大量给水泵的历史数据会使计算机耗时很长,导致实时性较差,且部分相似的数据会增加数据的冗余度。因此,要选取比较典型的数据来反映给水泵各正常工作状态。构建记忆矩阵可以采用等间隔抽样方法和聚类算法,但抽样法可能会遗漏部分正常运行状态的数据。因此,本文采用聚类算法来构建记忆矩阵。

凝聚层次算法(HAC)是根据某种距离度量准则逐步合并,将每个数据当作一个聚类,计算各个聚类之间的距离,每次聚类将距离最近的聚类进行合并,设定一个最终聚类数目,直到聚类个数为设定值,完成聚类,用每一类的聚类中心代表这一类中的数据^[12-14]。

但是传统的凝聚层次算法在大量处理高维密度不均匀数据时,存在一部分数据偏离其所在类的聚类中心,显然用这个聚类中心代表这一类所有的数据会产生较大的误差,因此本文提出一种基于距离检测的改进凝聚层次算法。改进算法的具流程如下:

(1) 分别对降维后的每个主元数据进行归一化:

$$x_i^* = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (21)$$

式中: x_i^* 和 x_i —归一化前后的值; x_{\min} 和 x_{\max} —该数据的最小值和最大值。

(2) 由于历史数据量比较多,每次聚类需要计算 $(k-1)!$ 次距离,计算量太庞大了,因此先对历

史数据按照时间对数据进行分区,每个分区是给水泵一天的历史数据,分别对每个分区进行聚类,这样可以大大减少计算量,降低时间和空间复杂度^[15]。

(3) 数据间距离度量的方法选择欧式距离:

$$d_{j,k} = \sqrt{\sum_{i=1}^n (x_{ij} - y_{ik})^2} \quad (22)$$

式中: j 和 k —第 j 和第 k 个时刻的数据; n —数据的维度:

(4) 类间距离度量方法的选择质心距离法,设该聚类中有 n 个点,质心的坐标为:

$$M = \left(\frac{1}{n} \sum_{i=1}^n x_i, \frac{1}{n} \sum_{i=1}^n y_i, \dots \right) \quad (23)$$

将每个聚类中所有点的质心作为聚类中心。

(5) 设定最终聚类数为原始数据量的 10%; 并设定一个距离阈值。

(6) 分别计算每个聚类中心之间的欧式距离,合并距离最近的两个聚类,为一个新的聚类,并更新聚类中心,直到聚类数为设定的最终聚类数。

(7) 计算各个聚类中的每个数据距离质心的距离,得到距离质心最大的点:

$$d_{\max} = \max(d_1, d_2, \dots, d_l) \quad (24)$$

若质心与最远点的距离大于设定的阈值,则这个聚类为异常类,对这个聚类中的所有数据继续进行聚类,设定最终聚类数为这个异常类的 50%,重复步骤(6),否则转至步骤(8)。

(8) 聚类结束,用各个聚类中心代表它们所在的聚类中的数据,并将每个分区的所有聚类中心构成多元状态估计的记忆矩阵 D 。

与传统的聚类方法相比,采用改进后的层次凝聚聚类解决了处理大量高维密度不均匀数据时数据偏离聚类中心的问题,通过对数据分区和距离检测法提高了预警模型运行的速度和准确性。

2.3 设定预警阈值

给水泵工作环境复杂,经常会受到外界随机因素的影响,因此选择滑动窗口法来设定预警阈值,选择合适的窗口宽度,可以及时、迅速地获取残差统计特性的连续变化,设残差序列共有 n 个^[16]:

$$\varepsilon = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n] \quad (25)$$

设定一个合适的窗口宽度 N , 计算窗口内连续 N 个残差的平均值:

$$\bar{\varepsilon} = \frac{1}{N} \sum_{i=1}^n \varepsilon_i \quad (26)$$

设滑动窗口均值序列的最大值为 E_{\max} , 则预警阈值为:

$$E_w = k \times E_{\max} \quad (27)$$

式中, k —预警阈值系数。

预警系数 k 的值一般由现场运行经验确定, k 的取值如果取过大, 阈值设定的较大, 就会导致故障预警的灵敏性降低, 无法充分提前预警; k 的取值如果取过小, 阈值设定的较小, 就会导致故障预警的准确性降低^[17]。

3 应用实例

3.1 数据的降噪和降维

采集了 2020/6/1/5:00 至 2020/6/24/16:10 期间共 6 759 组数据, 采样周期为 5 min, 得到 $6\ 760 \times 24$ 的样本矩阵, 其中该电厂给水泵在 6 月 23 日发生了故障, 其它时间都是正常工作状态。

采集的数据含有大量高频噪声信号, 采用改进的小波变换对这 24 个测点的数据进行去噪处理, 降噪阈值设定为启发式阈值, 小波基采用 db4, 分解层数设为 4 层, 改进的阈值函数 N 分别设定为 10 和 50, 并和 3 种常见阈值函数比较, 表 1 为不同阈值函数对给水泵出口母管压力信号去噪效果对比。

表 1 去噪结果定量比较

Fig. 1 Quantitative comparison of denoising results

阈值函数	信噪比	均方误差
硬阈值法	22.329 2	0.681 4
软阈值法	19.156 5	1.256 5
Garrote 阈值法	20.590 8	0.928 4
改进阈值法 ($N=10$)	26.739 2	0.426 8
改进阈值法 ($N=50$)	24.875 5	0.585 2

由表 1 可知, 相比于常见的 3 种阈值函数法, 改进后的阈值函数有较高的信噪比和较小的均方误差, 当调节因子 $N=10$ 时, 去噪效果更好, 图 2 为改进阈值法降噪前后的对比图, 可以看出, 降噪后的信号变得平滑。

数据降噪完后, 通过 PCA 对给水泵的 24 个状态变量进行降维, 表 2 为各个主元的特征值和方差贡献率。

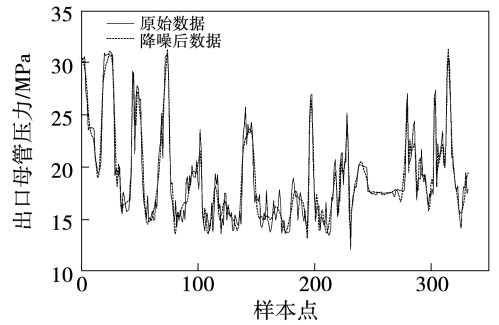


图 2 小波降噪结果图

Fig. 2 Wavelet noise reduction result graph

表 2 主成分方差贡献率

Fig. 2 Principal component variance contribution rate

主元个数	特征值	方差贡献率/%	累计方差贡献率/%
1	13.816 7	57.57	57.57
2	3.328 6	13.87	71.44
3	1.778 8	0.074 1	78.85
4	1.476 2	6.15	85.00
5	1.078 1	4.49	89.49
6	0.494 8	2.06	91.55
7	0.427 4	1.78	93.34

工业上要求累计方差贡献率要大于 85%, 为了使预警模型有更高的准确性, 选择主元个数为 6, 此时累计方差贡献率为 91.55%。图 3 为各个主元特征值的陡坡图, 当主元个数大于 6 时, 特征值趋于平缓而且接近于 0。由此可说明, 前 6 个主元保留了给水泵 24 个测点的绝大部分信息, 实现了数据的降维, 降低了后续建模的复杂度。

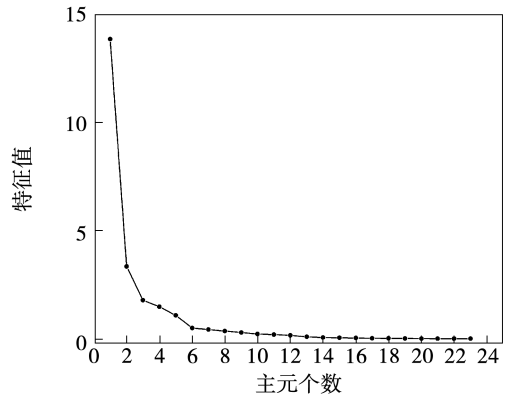


图 3 陡坡图

Fig. 3 Steep slope map

3.2 预警模型的建立

经过对原始数据降噪和降维处理后,得到 $6 \times 6\ 759$ 的矩阵,选取 2020/6/1/5:00 到 2020/6/21/24:00 共 5 989 个数据来构建记忆矩阵 D ,2020/6/22/0:00 到 2020/6/24/16:00 共 770 个数据来验证预警模型,预警模型流程如图 4 所示。

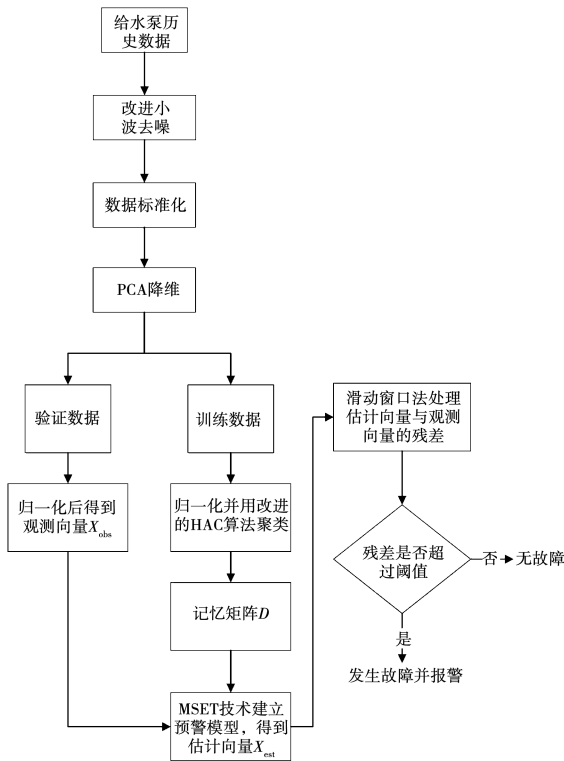


图 4 故障预警流程图

Fig. 4 Fault warning flow chart

通过改进的层次凝聚聚类算法 (HAC) 对 5989 组历史数据进行聚类,并以每一类的聚类中心来代替这一类的数据,最终得到 6×984 的记忆矩阵,并通过轮廓系数 (SC),归一化互信息 (NMI),调整兰德指数 (ARI) 来评价传统 HAC 和改进后的 HAC 的聚类效果,轮廓系数的取值在 $[-1, 1]$ 之间,值越大说明同类样本相距越近,聚类效果越好。归一化互信息指的是两个随机变量之间的关联程度,取值区间 $[0, 1]$,越接近 1 聚类效果越好,调整兰德指数的取值区间也是 $[0, 1]$,ARI 越接近 1 表示聚类效果准确性越高,同时每个类内的纯度越高^[18]。传统 HAC 和改进后的 HAC 聚类性能比较如表 3 所示。

表 3 传统 HAC 和改进后的 HAC 聚类性能对比

Fig. 3 Traditional HAC and improved HAC clustering

performance comparison

聚类算法	SC	NMI	ARI
传统的 HAC	0.935 6	0.982 4	0.978 2
改进的 HAC	0.884 5	0.961 8	0.943 9

由表 3 可知,改进后的层次凝聚聚类算法有更好的聚类性能,说明改进后的 HAC 算法更适合记忆矩阵的构建。

3.3 预警模型的验证

为了验证预警模型的准确性,采用 6 月 22 日到 6 月 24 日的数据来测试预警模型,并采用最小二乘支持向量机 (LS-SVM) 预警模型与此模型作为对比,分析两种模型的预测性能。图 5 为这两种预警模型对给水泵出口流量的预测对比图,并计算真实值与预测值的均方根误差和平均绝对误差来评价两种预测模型。表 4 为两种模型对给水泵出口流量和母管出口压力的均方根误差和平均绝对误差。

表 4 MSET 和 LS-SVM 模型预测指标评价

Fig. 4 HAC and improved HAC clustering

performance comparison

给水泵参数	模型	均方根误差	平均绝对误差
出口流量	MSET	0.459 1	0.176 4
	LS-SVM	0.695 7	0.331 2
母管出口压力	MSET	0.341 3	0.243 9
	LS-SVM	0.516 7	0.393 4

由表 4 可知,MSET 模型的预测值有更小的均方根误差和平均绝对误差,说明该预警模型比 LS-SVM 预警模型具有更高的精确度。

图 6 为浙江某电厂 6 月 22 日到 6 月 24 日 (共 770 个样本点) 的预警模型滑动残差图。选取的预警阈值为 0.38,在第 382 个样本点 (6 月 23 日 01:00) 滑动残差超过预警阈值,说明此时给水泵发生了故障,在第 418 个样本点 (6 月 23 日 04:00) 运检人员发现了水泵有进水口堵塞现象,导致流量不足,并进行了检修,消除了这次故障,采用此预警模型比人工巡检可以提前 3 h 发现故障,可以实现

给水泵早期的故障预警,减少故障发生时电厂的经济损失。

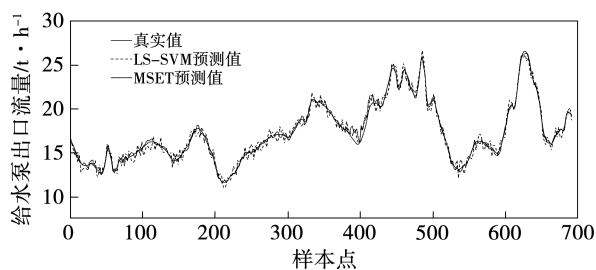


图 5 给水泵出口流量预测对比

Fig. 5 Prediction and comparison of outlet flow of feed water pump

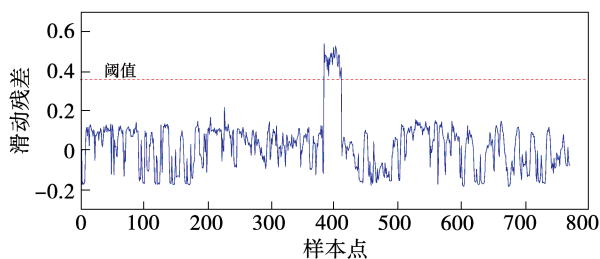


图 6 MSET 模型滑动残差

Fig. 6 MSET model sliding residual

4 结 论

选取能反映汽动给水泵运行状态的 24 个监测参数,利用 MSET 技术建立预警模型,并利用滑动窗口法设定预警阈值,当残差超过设定的阈值,预警模型就会发出预警信息,提醒电厂人员对给水泵进行检修。通过仿真可以得到以下结论:

(1) 采用改进的小波去噪法对数据进行降噪,除去原始数据中的高频信号,与传统的小波去噪相比,改进后的方法信噪比更高,且均方误差小。再通过 PCA 对大量状态参数进行降维,6 个主元以 91.55% 的精度代表原 24 个变量,实现数据的降维。

(2) 通过改进的 HAC 算法构建记忆矩阵 D ,引入了距离检测的方法对层次凝聚聚类算法进行改进,减少了冗余的历史数据,且提高了建模的精度,使记忆矩阵可以包含给水泵各种正常运行状态。

(3) 将 MSET 模型与 LS-SVM 模型对比,MSET 预警模型具有更高的准确性。并利用浙江某电厂汽

动给水泵对预警模型进行验证,可以提前 3 h 对给水泵进行故障预警,及时准确地发现给水泵的故障。

参考文献:

- [1] 徐红伟,刘振宇,李崇晟.基于分类算法的汽动给水泵组故障预测[J].热力发电,2019,48(4):128-134.
XU Hong-wei, LIU Zhen-yu, LI Chong-sheng. Fault prediction for turbine driven boiler feed water pump set based on classification algorithm [J]. Thermal Power Generation, 2019, 48(4): 128-134.
- [2] 吴昕,陈侃,谢昌亚,等.基于 PCA-KNN 的给水泵振动预警方法研究[J].化工机械,2022,49(1):137-142.
WU Xin, CHEN Kan, XIE Chang-ya, et al. Research on early warning method of feed water pump vibration based on PCA-KNN [J]. Chemical Engineering & Machinery, 2022, 49(1): 137-142.
- [3] 董玉亮,苏焯,王浩,等.基于正交局部保持投影与自组织映射的汽动给水泵组健康衰退评价[J].动力工程学报,2015,35(8):639-645.
DONG Yu-liang, SU Ye, WANG Hao, et al. Health degradation assessment for turbine-driven feed water pump based on orthogonal locality preserving projection and self-organizing map [J]. Journal of Chinese Society of Power Engineering, 2015, 35(8): 639-645.
- [4] 侯全放.汽动给水泵性能参数研究及状态在线监测的实现[D].哈尔滨:哈尔滨工业大学,2016.
HOU Quan-fang. Study on steam-driven feed water pump performance parameters and realize on-line monitoring of state [D]. Harbin: Harbin Institute of Technology, 2016.
- [5] 张阳峰,韦仕鸿,邓娜娜,等.基于小波降噪的振动传感器数据分析[J].计算机科学,2019,46(6A):537-539,565.
ZHANG Yang-feng, WEI Shi-hong, DENG Na-na, et al. Vibration sensor data analysis based on wavelet denoising [J]. Computer Science, 2019, 46(6A): 537-539, 565.
- [6] 刘剑,秦飞龙.改进的小波阈值法及其在地震数据降噪处理中的应用[J].物探与化探,2020,44(4):784-789.
LIU Jian, QIN Fei-long. The application of the improved wavelet threshold method to seismic data de-noising [J]. Geophysical and Geochemical Exploration, 2020, 44(4): 784-789.
- [7] 李维松,许俊杰,张涛.基于小波变换阈值去噪算法的改进[J].计算机仿真,2021,38(6):348-351,356.
LI Wei-song, XU Wei-jie, ZHANG Tao. Improvement of threshold denoising method based on wavelet transform [J]. Computer Simulation, 2021, 38(6): 348-351, 356.
- [8] 韩万里,茅大钧,印琪民.基于 PCA 和多元状态估计的引风机故障预警[J].热能动力工程,2020,35(1):91-97.
HAN Wan-li, MAO Da-jun, YIN Qi-min. Induced draft fan fault warning based on PCA and multivariate state estimation technique

- [J]. Journal of Engineering for Thermal Energy and Power, 2020, 35(1): 91-97.
- [9] 胡杰, 唐静, 谢仕义. 基于实时动态基线的运行设备多元状态估计方法[J]. 热力发电, 2021, 50(2): 125-131.
- HU Jie, TANG Jing, XIE Shi-yi. Multivariate state estimation technique for equipment running condition using real-time dynamic baseline[J]. Thermal Power Generation, 2021, 50(2): 125-131.
- [10] 刘涛. 基于多元状态估计的电站风机故障预警研究及系统开发[D]. 北京: 华北电力大学, 2016.
- LIU Tao. Research and development of early fault warning of fan in power plant based on multivariate state estimation technique [D]. Beijing: North China Electric Power University, 2016.
- [11] 牛玉广, 李晓彬, 张佳辉. 基于多元状态估计与自适应阈值的电站辅机故障预警[J]. 动力工程学报, 2019, 39(9): 717-724.
- NIU Yu-guang, LI Xiao-bin, ZHANG Jia-hui. Early fault warning for auxiliary equipment of power plants based on multivariate state estimation and adaptive threshold [J]. Journal of Chinese Society of Power Engineering, 2019, 39(9): 717-724.
- [12] 赵春, 潘建平, 陈璟. 一种新的基于层次凝聚的高效复合聚类算法[J]. 科技通报, 2012, 28(4): 125-127.
- ZHAO Chun, PAN Jian-ping, CHEN Jing. A new efficient combined clustering algorithm based on agglomerative hierarchical [J]. Bulletin of Science and Technology, 2012, 28(4): 125-127.
- [13] 肖娜. 基于密度和方向的聚类算法研究[D]. 长沙: 湖南大学, 2019.
- XIAO Na. Research on clustering algorithm based on density and direction [D]. Changsha: Hunan University, 2019.
- [14] 李敏, 陈果, 沈大千, 等. 基于改进凝聚层次聚类算法的变压器绕组及铁心故障诊断研究[J]. 高压电器, 2018, 54(1): 236-242.
- LI Min, CHEN Guo, SHEN Da-qian, et al. Research on fault diagnosis of transformer winding and core based on the improved agglomerative hierarchical clustering algorithm [J]. High Voltage Apparatus, 2018, 54(1): 236-242.
- [15] 杨金花. 基于 Web 挖掘的层次凝聚聚类算法研究[J]. 电子设计工程, 2012, 20(12): 30-32.
- YANG Jin-hua. Algorithm-based web mining the level of cohesion class [J]. Electronic Design Engineering, 2012, 20(12): 30-32.
- [16] 徐霖琳, 徐健锋, 刘龙, 等. 面向滑动窗口法概念格漂移计算研究[J]. 计算机科学与探索, 2021, 15(6): 1145-1154.
- XU Ji-lin, XU Jian-feng, LIU Long, et al. Research on drift calculation of concept lattice for sliding window method [J]. Journal of Frontiers of Computer Science and Technology, 2021, 15(6): 1145-1154.
- [17] 刘耘彰. 电厂风机故障诊断与预警研究[D]. 杭州: 浙江大学, 2019.
- LIU Yun-zhang. Research on fault diagnosis and early warning of power plants fans [D]. Hangzhou: Zhejiang University, 2019.
- [18] 赵娜娜. 聚类算法及聚类有效性指标的研究[D]. 无锡: 江南大学, 2016.
- ZHAO Na-na. Research on fuzzy clustering algorithm and cluster validity index [D]. Wuxi: Jiangnan University, 2016.

(金圣迪 编辑)