

# 控制与决策

Control and Decision

基于SCADA参量关联互信息自编码的风电机组故障检测方法

刘小峰, 史长振, 晏锐, 柏林

引用本文:

刘小峰,史长振,晏锐,柏林. 基于SCADA参量关联互信息自编码的风电机组故障检测方法[J]. *控制与决策*, 2023, 38(10): 2953–2961.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2021.1755>

---

您可能感兴趣的其他文章

Articles you may be interested in

[基于广义主成分分析的重构故障子空间建模方法](#)

Reconstructed fault subspace modelling method based on generalized principal component analysis

*控制与决策*. 2021, 36(4): 808–814 <https://doi.org/10.13195/j.kzyjc.2019.0818>

[改进集成深层自编码器在轴承故障诊断中的应用](#)

Application of improved ensemble deep auto-encoder in bearing fault diagnosis

*控制与决策*. 2021, 36(1): 135–142 <https://doi.org/10.13195/j.kzyjc.2019.0270>

[基于Fisher Score与最大信息系数的齿轮箱故障特征选择方法](#)

Fault feature selection method of gearbox based on Fisher Score and maximum information coefficient

*控制与决策*. 2021, 36(9): 2234–2240 <https://doi.org/10.13195/j.kzyjc.2019.1770>

[基于批次图像化的卷积自编码故障监测方法](#)

Fault detection of batch image-based convolutional autoencoder

*控制与决策*. 2021, 36(6): 1361–1367 <https://doi.org/10.13195/j.kzyjc.2019.1342>

[基于核典型相关性-熵成分分析的工业过程质量监测方法](#)

A quality monitoring method for industrial process based on kernel canonical correlation-entropy component analysis

*控制与决策*. 2021, 36(12): 2999–3006 <https://doi.org/10.13195/j.kzyjc.2020.0548>

# 基于 SCADA 参量关联互信息自编码的风电机组 故障检测方法

刘小峰<sup>†</sup>, 史长振, 晏锐, 柏林

(重庆大学 机械与运载工程学院, 重庆 400044)

**摘要:** 针对风力发电机组数据采集与监视控制系统 (supervisory control and data acquisition, SCADA) 监测参量间的耦合关联性, 提出基于多参数耦合关联互信息编码的风电机组故障检测方法. 该方法构建了 SCADA 数据的耦合关联矩阵, 采用互信息变分自编码器对关联矩阵进行编码重构; 将 SCADA 参量关联矩阵的重构误差作为机组健康评估指标, 结合指数加权移动平均模型的迭代更新, 对机组实时故障阈值进行自适应设置. 两个风场的风电机组 SCADA 数据分析结果表明, 所提方法充分利用了 SCADA 数据的耦合关联结构信息, 能有效提高风电机组故障检测的准确性及对环境工况的鲁棒性.

**关键词:** 多参数关联耦合; 变分自编码器; 最大互信息; 健康指数构建; 风电机组故障检测

中图分类号: TM761

文献标志码: A

DOI: 10.13195/j.kzyjc.2021.1755

引用格式: 刘小峰, 史长振, 晏锐, 等. 基于 SCADA 参量关联互信息自编码的风电机组故障检测方法[J]. 控制与决策, 2023, 38(10): 2953-2961.

## Wind turbine fault detection based on mutual information auto-encoding of SCADA data correlation

LIU Xiao-feng<sup>†</sup>, SHI Chang-zhen, YAN Rui, BO Lin

(College of Mechanical and Vehicle Engineering, Chongqing University, Chongqing 400044, China)

**Abstract:** Owing that the monitoring variables of the wind turbine supervisory control and data acquisition (SCADA) system are highly correlated and coupled, a wind turbine fault detection method is proposed based on multi-parameter correlation coupling and mutual information auto-encoder. The correlation matrix of multidimensional time series is established to describe the coupled relationship of SCADA data. The mutual information based variational auto-encoder is proposed for the encoding-decoding reconstruction of the correlation matrix. The wind turbine health evaluation index is constructed based on the reconstruction error of the correlation matrix, and then the early fault threshold is set through the update iteration of the exponential weighted moving average model. The proposed method is verified using the SCADA monitoring data of wind turbines in two wind farms. The results show that this method can effectively mine the internal correlation coupling information of SCADA multivariable time series, which can effectively improve the accuracy of wind turbine anomaly detection and the robustness to environmental interferences.

**Keywords:** multiple parameters correlation coupling; variational auto-encoder; maximum mutual information; health index construction; fault detection of wind turbine

## 0 引言

对于长期处于恶劣环境中持续运行的风电机组 (wind turbine, WT), 准确可靠的故障异常预警与视情维修, 确保机组可靠高效运行, 降低运维成本, 是风电系统优化运行亟待解决的关键问题<sup>[1]</sup>. 数据采集与监视控制系统 (supervisory control and data acquisition, SCADA) 数据驱动的故障检测方法, 无需

安装额外的传感器即可实现风电机系统的全方位检测, 大大降低了状态监测的投入成本, 已成为近年来 WT 状态监测的研究热点<sup>[2]</sup>. 基于 SCADA 数据的 WT 故障检测常见解决方案是对 WT 异常状态进行统计处理后, 将偏离正常样本的状态设定为故障异常, 涉及到的技术方法可大致分为聚类分析法、模型重构法及参量关系法等. 聚类分析法的主旨是建立 WT 正

收稿日期: 2021-10-13; 录用日期: 2022-06-15.

基金项目: 国家自然科学基金项目 (51975067, 52175077).

<sup>†</sup>通讯作者. E-mail: liuxfeng0080@126.com.

常状态下的监测样本聚类中心,远离聚类中心的测试样本即为故障异常.此类方法对工况稳定性的要求较高,对监测数据的规整性依赖较强,因此其在强时变工况、高随机运行环境下的故障检测准确性有待提高.

鉴于SCADA数据的海量、多源、异常数据缺乏和无标签等特点,基于深度学习的检测模型得到长足的发展.林涛等<sup>[3]</sup>采用堆栈式自编码器对SCADA数据进行建模,通过重构误差的分布差异性进行WT的早期故障检测.黄荣舟等<sup>[4]</sup>使用正常状态下SCADA数据训练长短时记忆网络预测模型,以预测值残差为基础,结合 $3\sigma$ 准则进行风电齿轮箱状态检测与故障预警. Renström等<sup>[5]</sup>针对风电机组SCADA系统监测数据,提出了一种基于自动编码器模型,通过重构所有输入信号实现了对WT各组件的异常检测. Yang等<sup>[6]</sup>基于风电机SCADA数据,应用时空模式网络来提取系统中变量的时空特征,使用基于能量的受限玻尔兹曼机(RBM)进行WT异常检测.赵洪山等<sup>[7]</sup>采用逐层编码网络模型对WT主轴承正常状态数据进行训练,采用模型重构误差对主轴承进行故障预测.上述模型重构法中,输入数据与输出数据间的重构关系是建立在累积网络模型对每个输入样本的复杂非线性响应基础上的,当输入不同于正常状态样本但却与WT部件故障异常无关的奇异数据时,网络模型的非线性响应输出也不同,因此易将未出现的正常样本误判为异常,增加了虚警风险.

监测参量关系法以多参量间的关系模型为基础,根据关系模型参数的变化或者偏差来辨识异常状态.魏书荣等<sup>[8]</sup>采用灰色关联分析筛选出与WT运行温度相关的状态变量,通过长短记忆网络对WT运行温度进行预测,根据温度预测残差进行了风机早期故障的预警. Yang等<sup>[9]</sup>利用多项式最小二乘法建立模型,实现了独立于工况的WT状态监测.李霸军等<sup>[10]</sup>采用多项式回归法建立SCADA监控数据之间的拟合关系,根据正常状态与异常状态下回归模型系数的差异性进行WT异常状态的检测.张帆等<sup>[11]</sup>采用多项式回归拟合方法,构建WT运行状态输入/输出参数关系数学模型,对WT运行状态出现异常进行预警. Gill等<sup>[12]</sup>采用Copula函数对WT功率曲线的概率进行了拟合,利用SCADA系统运行参数对WT的变桨系统的早期故障进行了预测.上述这些参量关系法大多通过多维数据的加权融合来构建某一个时间序列的预测模型,缺乏对各个参量变化的相互影响关系的综合考虑,因此容易受变量间冗余作用、误差积

累和关联信息缺乏等的影响.

由上述分析可知,基于SCADA数据的WT故障检测方法,其检测精度的高低一方面取决于SCADA数据蕴含的状态信息的有效提取,另一方面取决于WT正常状态描述模型的准确建立. WT作为一个关联多个变量的复杂动力系统,其演化运行受到大量内部因素与外部环境的综合制约和影响,通过分析其变量间的关系,可以在消除信息冗余的情况下,揭露系统内部的结构信息.因此,针对WT各部件关联性强、监测点位多、监测参量耦合关系复杂的特点,本文提出一种基于SCADA参量关联互信息编码的WT故障检测方法.该方法充分利用SCADA参量间的关联耦合信息,采用互信息变分自编码器(mutual information variational auto-encoders, MIVAE)建立WT正常状态下参量耦合关系的描述空间,设定WT故障预警动态阈值,能够有效提高WT早期故障检测准确率与工况鲁棒性.

## 1 多维时间序列关联矩阵建立

### 1.1 去趋势波动分析

去趋势波动分析(detrended fluctuation analysis, DFA)通过计算去趋势波动函数来达到最小化外部趋势对自相关性的影响,能够排除外部噪声影响从而获取数据本征特征,适合于非平稳时间序列的长程相关程度分析<sup>[13]</sup>.在时间序列 $\{x(t) : t = 1, 2, \dots, N\}$ 的DFA处理中,首先需计算 $x(t)$ 的累计差

$$X(i) = \sum_{t=1}^i (x(t) - \bar{x}), \quad i = 1, 2, \dots, N, \quad (1)$$

其中 $\bar{x} = \frac{1}{N} \sum_{t=1}^N x(t)$ .采用长度为 $n$ 的时间滑动窗将累积序列 $X(i)$ 分割成 $N - n$ 个子区间 $D_v = (X(v), X(v+1), \dots, X(v+n-1))^T, v = 1, 2, \dots, N - n$ .设置窗口滑动步长为1,以减少相邻分割内数据的跳跃性.对每个子区间内的 $n$ 个点,使用最小二乘法进行多项式拟合,计算子区间内累积序列的局部趋势

$$P_{X,v}(i) = a_0 + a_1 i + a_2 i^2 + \dots + a_k i^k. \quad (2)$$

这里取 $k = 3$ .计算子区间内的波动局部趋势

$$F_X(v, n) = \frac{1}{n} \sum_{i=1}^n [X[(v-1)n + i] - P_{X,v}(i)]^2. \quad (3)$$

求取 $F_X(v, n)$ 的平均值,得到波动函数

$$F_x(n) = \left\{ \frac{1}{N-n} \sum_{v=1}^{N-n} F_X(v, n) \right\}^{1/2}. \quad (4)$$

若时间序列 $x(t)$ 具有长程自相关性,则 $F_x(n)$ 与 $n$ 满足幂律关系: $F_x(n) \sim n^\alpha$ .其中 $\alpha$ 为DFA的标度指数,

其反映了序列的长程自相关性。

## 1.2 消除趋势互相关分析

消除趋势互相关分析(detrended cross-correlation analysis, DCCA)可有效避免由于数据非平稳性所导致的序列之间的伪相关现象,系统地滤去各阶趋势成分,从而检测出含有噪声且叠加有多项式趋势信号的长程相关性<sup>[14]</sup>。对于两个时间序列 $\{x(t) : t = 1, 2, \dots, N\}$ 和 $\{y(t) : t = 1, 2, \dots, N\}$ ,按照式(1)分别得到两个序列的累加信号 $X(i)$ 和 $Y(i)$ ,并对两个累积序列进行区域划分,然后按照式(2)对 $X(i)$ 和 $Y(i)$ 每个区域中的 $n$ 个数据点进行多项式最小二乘拟合,得到每个区域内 $X(i)$ 和 $Y(i)$ 的局部趋势 $P_{X,v}(i)$ 和 $P_{Y,v}(i)$ ,继而计算两者的互协方差函数

$$F_{XY}(v, n) = \frac{1}{n} \sum_{i=1}^n \{(X[(v-1)n+i] - P_{X,v}(i))(Y[(v-1)n+i] - P_{Y,v}(i))\}. \quad (5)$$

对于各个区间内的累积序列,求取 $F_{XY}(v, n)$ 的平均值,得到波动函数

$$F_{xy}(n) = \left\{ \frac{1}{N-n} \sum_{v=1}^{N-n} |F_{XY}(v, n)|^2 \right\}^{1/2}. \quad (6)$$

若时间序列 $x(t)$ 具有长程相关性,则 $F_{xy}(n)$ 与 $n$ 满足幂律关系: $F_{xy}(n) \sim n^h$ 。其中 $h$ 为自相似参量,即DCCA的标度指数,定量表征了两组非平稳序列间的相关性。

## 1.3 关联矩阵的建立

由于DCCA标度参数 $h$ 无法严格量化两个序列之间的相关程度,引入DCCA互相关系数<sup>[15]</sup>以量化序列 $x(t)$ 与 $y(t)$ 之间的相关程度,即

$$\rho_{xy}(n) = \frac{F_{xy}^2(n)}{F_x(n)F_y(n)}. \quad (7)$$

$\rho_{xy}$ 的大小表明两组时间序列之间的互相关性。若 $\rho_{xy} = 0$ ,则表示两组序列毫无关联;若 $0 < \rho_{xy} < 1$ ,则表明在过去的某一段时间内某一组序列变量存在增加(或减少)的趋势,将会导致另一组变量在未来一定时间尺度上也存在相同的趋势,且 $\rho_{xy}$ 越大,两组时序变量的正相关性越强;若 $-1 < \rho_{xy} < 0$ ,则表明两组序列变量之间存在持续的反相关性。

WT故障异常是指在整个机组系统中由于部分零部件出现故障,导致采集的SCADA数据中存在单个或多个偏离整体的参量数据,由于这种偏离程度较小,且受到工况环境的影响,往往不容易通过直接数据趋势拟合或模型描述方法检测到,需要通过对整体数据隐含的机组状态信息进行深度挖掘。对于WT的

SCADA系统记录的按时间顺序排列的参数序列,通过剔除数据中的空值以及停机阶段的数据,可得到数据集

$$D = \left\{ \begin{matrix} s_{1,1} & s_{1,2} & \dots & s_{1,i} & \dots \\ s_{2,1} & s_{2,2} & \dots & s_{2,i} & \dots \\ \vdots & \vdots & \ddots & \vdots & \ddots \\ s_{M,1} & s_{M,2} & \dots & s_{M,i} & \dots \end{matrix} \right\}, \quad (8)$$

其中 $s_{m,i}$ ( $m = 1, 2, \dots, M$ )为 $t_i$ 时刻WT的第 $m$ 个状态参数。采用数据长度为 $w$ 的等步长移动的时间滑动窗口对 $D$ 进行截断处理,滑动窗口随着时间递增顺序逐步推移,即可构建 $K$ 个 $M$ 维等长时间序列 $D_k = [S_{1k}, S_{2k}, \dots, S_{Mk}]^T$ 。其中: $S_{mk}$ 表示SCADA系统各个传感器记录的第 $k$ 个时间段窗口内的第 $m$ 个参量时间序列, $S_{mk} = (s_{m,k-w}, s_{m,k-w+1}, \dots, s_{m,k})$ , $k = 1, 2, \dots, K$ 。

WT的状态变化会引起监测参数的数值变化,也会引起参量间的耦合关系变化,为了对SCADA参量时间序列间的相互关联性进行量化描述,根据式(7)计算 $D_k$ 中任意两个参量时间序列 $S_{mk}$ 与 $S_{nk}$ 之间的DCCA相关系数 $\rho_{(s_{mk}, s_{nk})}$ ,构建WT的多参数耦合关联矩阵

$$G_k = \begin{bmatrix} \rho_{s_{1k}, s_{1k}} & \rho_{s_{1k}, s_{2k}} & \dots & \rho_{s_{1k}, s_{Mk}} \\ \rho_{s_{2k}, s_{1k}} & \rho_{s_{2k}, s_{2k}} & \dots & \rho_{s_{2k}, s_{Mk}} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{s_{Mk}, s_{1k}} & \rho_{s_{Mk}, s_{2k}} & \dots & \rho_{s_{Mk}, s_{Mk}} \end{bmatrix}. \quad (9)$$

时间滑动窗内关联矩阵 $G_k$ 发生显著变化时,预示着WT的运行状态可能出现了故障异常。以 $G_k$ 为数据驱动的WT故障检测方法,摒弃了传统的受工况环境影响较大的监测参量驱动理念,充分利用了SCADA系统中非平稳时间序列间的相互依赖、影响变化等结构信息,可在一定程度上提高WT故障检测对环境噪声的抗干扰性以及按时变工况的鲁棒性,弥补SCADA监测数据对WT状态信息表征的不足。

## 2 基于MIVAE的故障检测

### 2.1 MIVAE网络架构

SCADA参量间的耦合关系是复杂多变的,其蕴含的机组健康状态信息的潜隐性高,以关联矩阵为驱动的故障检测模型,不仅需要考虑每个关系结点的本征特征,还需要学习表征结点间关联性结构信息。另者,在关联矩阵建立过程中,不明显的参量相关性会被当作结点处理,生成的多个零值项,导致关联矩阵的稀疏性,不利于模型的重构。因此,本文在传统变分自编码器(variational auto-encoders, VAE)的基础上进行深度扩展,构建如图1所示的MIVAE网络模型。

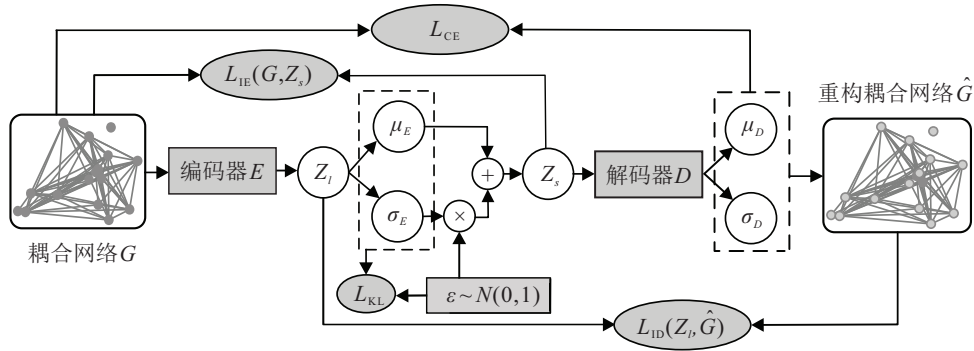


图1 MIVAE结构

图1中MIVAE的编码器 $E$ 将输入的关联矩阵 $G$ 转换为低维编码 $Z_l$ ,然后通过连接层与BN层,得到 $Z_l$ 的均值向量 $\mu_E$ 与方差向量 $\sigma_E$ ,再对 $Z_l$ 添加噪声并进行随机采样,构造隐变量 $Z_s$ 。同理,在解码阶段,先采用解码器从隐变量 $Z_s$ 中生成均值为 $\mu_D$ 、标准差为 $\sigma_D$ 的多元正态分布解码数据,从该分布中抽样生成重构矩阵 $\hat{G}$ 。MIVAE一方面在传统VAE的基础上增加了低维编码层,扩展了网络深度,以便建立 $G$ 与 $Z_s$ 更合理的非线性映射关系;另一方面在潜在特征提取与隐层变量空间构造的基础上,采用特征数值与其概率分布来表征输入数据,更能体现数据的本质信息,在面对变化的工况时也能精确地捕捉其中蕴含的设备健康信息。

## 2.2 MIVAE损失函数设置

由于隐变量 $Z_s$ 的真实分布 $P(Z_s)$ 与编码器学习的分布 $Q_\varphi(Z_s|G)$ ( $\varphi$ 为编码器参数)存在着偏差,通常希望这种偏差越小越好,KL散度可用于量化这个偏差,对应损失函数为

$$L_{KL} = -\text{KL}(Q_\varphi(Z_s|G) \| P_\theta(Z_s)). \quad (10)$$

其中:先验分布 $P_\theta(Z_s) = N(Z_s|0, I)$ , $\theta$ 是解码器的参数。MIVAE通过“再参数化”采样得到潜在向量,令 $Z_s = \mu_E + \sigma_E \odot \varepsilon$ ,将从正态分布 $N(\mu_E, \sigma_E^2)$ 采样转换为从正态分布 $N(0, I)$ 采样,从而保证了模型的训练效率。

由于解码器 $D$ 输入的是经白噪声处理过的具有高随机性的向量 $Z_s$ ,解码器易将不同的 $Z_s$ 映射到相同的重构矩阵 $\hat{G}$ ,导致 $\hat{G}$ 仅仅依赖于 $Z_s$ 的少数维度,使得 $P_\theta(G|Z_s)$ 只是真实数据分布的“子分布”,导致解码重构中会丢失一些真实的分布特征。为此,MIVAE采用 $G$ 与 $Z_s$ 间的最大互信息来约束编码器,使得 $G$ 的每个维度都尽可能地与隐变量 $Z$ 产生特定对应关系,迫使预测分布能更加逼近真实分布,对应损失函数设置为

$$L_{IE} = -I(G, Z_s) = \sum P_{G, Z_s} \log \frac{P_{G, Z_s}}{\tilde{P}_{G, Z_s}}. \quad (11)$$

其中: $P$ 表示两个变量的联合分布, $\tilde{P}$ 表示两个变量的边缘分布乘积, $P_{G, Z_s} = P(G, Z_s)$ , $\tilde{P}_{G, Z_s} = P(G)(Z_s)$ 。

与传统VAE一样,MIVAE采用式(12)中的重构误差衡量原始关联矩阵 $G$ 与重构关联矩阵 $\hat{G}$ 之间的差距,提高解码器恢复输入样本的能力,损失函数为

$$L_{CE} = \|G - \mu_D\|^2. \quad (12)$$

由于关联矩阵的稀疏性,计算交叉熵时会倾向于重构零项而非矩阵中表征重要关联性的非零项,降低了重构效果。因此,为了抑制重构的零值倾向性,采用直接编码特征 $Z_l$ 与 $\hat{G}$ 之间的互信息控制解码器 $D$ 的解码过程,使 $\hat{G}$ 的每个维度都尽可能地与 $Z_l$ 产生特定的关系,从而减少解码信息损失,对应损失函数设定为

$$L_{ID} = -I(Z_l, \hat{G}) = -\sum P(Z_l, \hat{G}) \log \frac{P(Z_l, \hat{G})}{P(Z_l)P(\hat{G})}. \quad (13)$$

综上,MIVAE总的损失函数为

$$L = \alpha L_{KL} + \beta L_{IE} + \gamma L_{CE} + \delta L_{ID}. \quad (14)$$

其中: $\alpha$ 、 $\beta$ 、 $\gamma$ 、 $\delta$ 为损失之间的权重,本文取 $\alpha = 0.01$ , $\beta = 0.5$ , $\gamma = 1$ , $\delta = 0.5$ 。可以看出,总损失函数的第1部分KL项用来表征关联矩阵的一阶相似度;第2部分从编码特征角度来消除预测分布与真实分布差异性带来的细节模糊性;第3部分重构项用来表征关联矩阵的二阶相似度;第4部分从解码特征角度来消除关联矩阵稀疏性带来的重构困难问题。

选取JS散度来表示互信息的最大化<sup>[16]</sup>,并使用负采样的方法对JS散度进行计算,最终的目标函数可以表示为

$$\begin{aligned} & \min\{\alpha \text{KL}[N(\mu_E, \sigma_E^2 I) \| N(0, I)] - \\ & \beta(E_{P_{G, Z_s}}[\log(T(G, Z_s))] + E_{\tilde{P}_{G, Z_s}}[\log(1 - \\ & (T(G, Z_s))]) - \gamma \|G - \mu_D\|^2 - \\ & \delta(E_{P_{Z_l, \hat{G}}}[\log(T(Z_l, \hat{G}))] + E_{\tilde{P}_{Z_l, \hat{G}}}[\log(1 - \end{aligned}$$

$$(T(Z_l, \hat{G}))\} \}, \quad (15)$$

其中 $T(\cdot)$ 表示鉴别网络 $T$ 的输出。

### 2.3 基于MIVAE的故障异常检测

在健康评估指数( $H$ )构建阶段,利用已训练完毕的MIVAE中的编码器 $E$ 与解码器 $D$ 对测试样本对应WT的 $H$ 指数进行构建。当待测关联矩阵 $G$ 来自健康状态样本时, $G$ 与重构矩阵 $\hat{G}$ 之间的差异极小;当 $G$ 来自异常状态样本时,正常状态空间模型不能正常提取异常状态下的关联矩阵潜隐特征,亦不能正确还原异常状态关系矩阵,此时 $G$ 与 $\hat{G}$ 之间的差异变大。故障异常检测的具体实现步骤如下:

step 1: 构建健康评估指数 $H$ 。对于WT正常状态数据, MIVAE模型中的输入关联矩阵与解码输出的重构矩阵之间的差异值可表示为

$$R = \|G - \hat{G}\|^2. \quad (16)$$

采用长为5个时间点的平滑窗对 $R$ 序列进行平滑处理,得到WT健康评估指数 $H$ ,即

$$H_t = \frac{1}{5} \sum_{i=1}^5 R(t-i+1), \quad (17)$$

则长度为 $l$ 的正常状态下的 $H$ 序列可表示为 $H_{\text{normal}} = \{H_1, H_2, \dots, H_l\}$ 。

step 2: 鉴于 $H$ 数值随工况环境的动态波动性较强,在WT运行过程中, $H$ 指数会随时间动态变化,“正常 $H$ 指数”并非一个特定的状态,而是会随着工况环境以及相关因素的变化不断更新状态。因此,对于WT故障异常的判断应基于当前状态的正常基准值。由于 $H$ 序列的变动具有非线性趋势性,本文采用指数加权移动平均(exponentially weighted moving-average, EWMA)模型<sup>[17]</sup>对 $H_{\text{normal}}$ 中的 $H$ 序列进行建模,有

$$\hat{H}_t = \lambda \sum_{j=0}^{t-1} (1-\lambda)^j H_{t-j} + (1-\lambda)^t H_0, \quad (18)$$

$$t = 1, 2, \dots, l.$$

其中 $\lambda$ 为随样本平均数的增加而呈几何下降趋势的权重参数。

step 3: 为了充分描述 $H$ 序列的上下波动特性,以EWMA模型预测值 $\hat{H}_t$ 为分界线,设定向上波动序列为

$$H_{t_1}^{\text{up}} = \begin{cases} H_t - \hat{H}_t, & H_t > \hat{H}_t; \\ 0, & \text{otherwise.} \end{cases} \quad (19)$$

计算 $H_{t_1}^{\text{up}}$ 序列的均值为

$$\alpha_{t_1} = \sum_{t=1}^L H_t^{\text{up}} / l,$$

其变异系数

$$\text{CV}_{t_1} = \sqrt{\sum_{t=1}^L (H_t^{\text{up}} - \alpha_{t_1})^2 / L} / \alpha_{t_1}.$$

同理,可以得到向下波动序列的均值 $\alpha_{t_2}$ 和其变异系数 $\text{CV}_{t_2}$ 。结合 $H$ 序列的波动规律与 $t$ 时刻的预测值, $H$ 序列在 $t$ 时刻的故障阈值为

$$F_t = \hat{H}_t + \alpha_{t_1} + \text{CV}_{t_1} + \alpha_{t_2} + \text{CV}_{t_2}. \quad (20)$$

step 4: 将测试数据的耦合关联矩阵 $G_{\text{test}}$ 输入训练好的MIVAE模型,得到原始耦合网络 $G_{\text{test}}$ 与重构耦合网络 $\hat{G}_{\text{test}}$ 间的差异值

$$R_{\text{test}} = \|G_{\text{test}} - \hat{G}_{\text{test}}\|_F^2. \quad (21)$$

根据式(17)进行滑动平均后得到 $G_{\text{test}}$ 对应的WT健康评价指数 $H_{\text{test}}$ 。

step 5: 当 $H_{\text{test}} < F_t$ 时,判断为正常,将该数据点 $H_{\text{test}}$ 作为第 $l+1$ 个正常 $H$ 值加入集合 $H_{\text{normal}}$ ,进行 $H_{\text{normal}}$ 数据集更新,即 $H_{\text{normal}} = \{H_1, H_2, \dots, H_l, H_{\text{test}}\}$ 。反之, $H_y > F_t$ ,则判断为异常。将预测值 $\hat{H}_{\text{test}}$ 作为该测试时间的正常 $H$ 值计入集合 $H_{\text{normal}}$ ,得到更新数据集 $H'_{\text{normal}} = \{H_1, H_2, \dots, H_l, \hat{H}_{\text{test}}\}$ 。

step 6: 采用 $H'_{\text{normal}}$ 建立新的EWMA模型,计算每个测试时间点对应的故障阈值,进而确立WT总的健康基线,  $\text{Baseline} = \{F_t\}_{t=1}^{T_{\text{total}}}$ ,  $T_{\text{total}}$ 为训练样本与测试样本总长度。

## 3 实例验证

### 3.1 数据描述

为了验证本文所提方法在WT故障检测中的有效性,选用来自某风场一台2 MW双馈式WT(编号为#1)的SCADA系统监测数据进行分析,并与传统故障异常检测方法进行结果比较。WT的切入风速和切出风速分别为4 m/s和25 m/s, SCADA数据采样率为1 sample/s。为降低噪声数据和缺失数据影响,取SCADA数据的10 min平均值进行分析。风电机组SCADA系统多源监测数据包含多组时间序列,包括表1中的16个状态参数以及偏航位置和桨距角等共计64类状态参数数据。为了避免过多参数带来的模型复杂化及冗余信息问题,选择表1中的监测参量进行关联耦合分析。

选取#1WT在2014年11月10日~2014年12月25日期间风速介于4 m/s至25 m/s的SCADA数据,并对其进行归一化处理,得到6630组样本数据。SCADA系统显示#1WT在2014年11月10日~2014年12月17日运行正常,期间未出现参数超限报警情

表1 #1WT用于构建关联矩阵的SCADA数据表

序号	SCADA 参数	序号	SCADA 参数
0	平均风速	8	转子轴承温度
1	风轮转速	9	齿轮箱轴承温度
2	平均有功功率	10	齿轮箱进口油温
3	平均叶片角度	11	齿轮箱油底壳温度
4	环境温度	12	发电机转速
5	机舱温度	13	发电机定子温度
6	机舱柜温度	14	发电机轴承1温度
7	转子温度	15	发电机轴承2温度

况,且限功率运行情况较少,因此判定#1WT在此期间处于正常状态.#1WT于2014年12月1日进行了一次定期维修,在2014年12月23日凌晨出现故障。

### 3.2 关联矩阵建立

根据1.3节中关联矩阵的建立方法,采用数据长度  $w = 720$  的等步长移动窗对#1WT的SCADA数据集进行处理,即每次截取5天的参量时间序列,滑动窗的时移步长为1,构建5911个16维等长时间序列  $D_k = [S_{1k}, S_{2k}, \dots, S_{Mk}]^T (M = 1, 2, \dots, 16, k = 1, 2, \dots, 5911)$ . 对于每个  $D_k$  多维时间序列,按照DCCA方法建立关联矩阵(9),进而得到5911个矩阵样本  $G_k (k = 1, 2, \dots, 5911)$ .

图2给出了12月18日~25日故障异常前后风电机组SCADA监测数据的耦合关系的可视化结果.图2中数字节点表示表1中SCADA参量编号,节点间的连线表示消除趋势相关矩阵中各元素的值,节点间无连线说明两监测变量间无明显关联性.由图2可知,当#1WT处于正常状态时,从正常状态A到正常状态B的过程中,监测变量之间的耦合关系发生了微妙的变化,变量3由正常状态A时与变量0、2、11和12耦合,变化到了正常状态B时与变量0、4、5、11、12和15耦合.可见,参量间的耦合关系不仅与WT的健康状态有关,也会受到环境工况的影响而产生变化.当环境工况引起监测参量幅值大幅波动时,参量间的耦合关系并没有产生很大的变化,也就是说,关联矩阵对环境干扰具有更好的鲁棒性.

#1WT从正常状态B到异常状态C的过程中,变量3与其他变量的耦合关系发生了显著变化,更明显的关系变化表现在变量2和变量12,由正常状态B时与其他15组变量都具有耦合关系,变化到异常状态C时完全脱离耦合,变量9和变量11也由正常状态B时与其他15组变量都具有耦合关系,分别变化到仅与变量1、3、14和仅与变量0、6、7、8、10、13、15耦合.可见,风电机组正常状态下与异常状态下的参量耦合关系存在明显不同.对比风电机组异常状态C到异常状态D下的耦合关系图可知,变量11与其

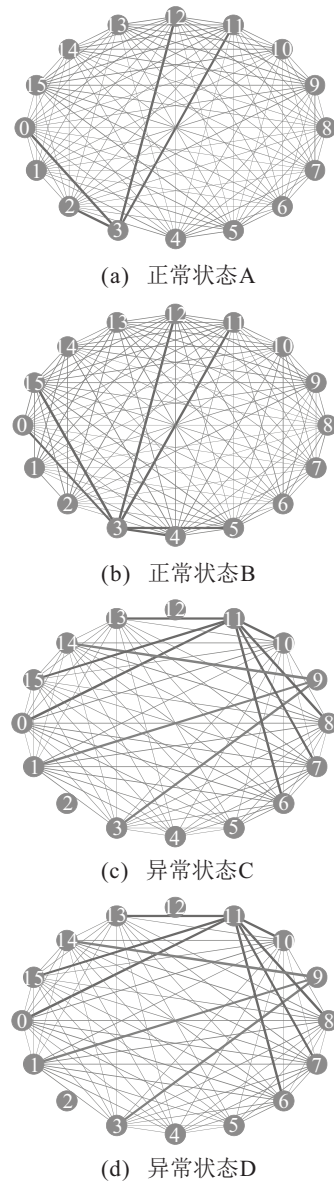


图2 关联耦合图

他变量之间的耦合关系产生了微小变化,但两者的耦合关系结构基本一致.这说明风电机组一旦发生故障异常,各监测变量之间的耦合关系便会较正常状态发生不可逆的显著变化,这也为本文方法对WT的故障检测的稳定性提供了理论基础.

### 3.3 模型训练

选取#1WT在2014年11月10日~2014年12月17日正常运行状态下的4759个关联矩阵作为MIVAE的训练集,12月18~25日的1152个关联矩阵作为测试集.MIVAE模型参数设置如下:编码器E为2层卷积和1层Dense结构,卷积核的大小为 $3 \times 3$ ,采用Relu函数进行激活.采用全连接层与BN层来拟合低维编码特征的均值和方差,神经元个数为32.解码器D为1层Dense和3层反卷积结构,卷积核的大小为 $3 \times 3$ ,前2层反卷积层采用Relu函数激活,后1层采用Sigmoid函数激活.使用Adam优化器训练200

次迭代,学习率设为0.01.采用式(15)中的目标函数进行模型迭代训练,经Batch内平均后的结果如图3所示.可见,随着迭代次数的增加,模型损失逐渐减小,当迭代次数超过150次时,训练误差曲线趋于平缓,达到收敛.

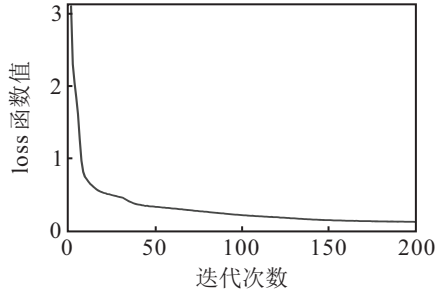


图3 损失函数收敛曲线

为了验证MIVAE模型在构建 $H$ 指数方面的有效性与优越性,图4给出了在同等输入支撑条件下采用MIVAE模型、孤立森林(IFOREST)<sup>[18]</sup>、离群点检测(LOF)<sup>[19]</sup>、深度卷积自编码器(DCAE)<sup>[5]</sup>及对抗自编码器(AAE)<sup>[20]</sup>构建的2014年12月10日~17日期间的正常运行的#1WT的 $H$ 曲线.由图4可知,对比其他4种传统模型所构建的健康评估指数,本文方法所构建的WT正常状态下 $H$ 指数波动幅度与波动频率都较小,对机组正常状态表征的一致性较强.这说明本文模型所建立的正常状态空间模型具有更好的紧致性,对机组健康状态的描述更稳定,受工况波动性影响更小,具有更好的噪声鲁棒性.

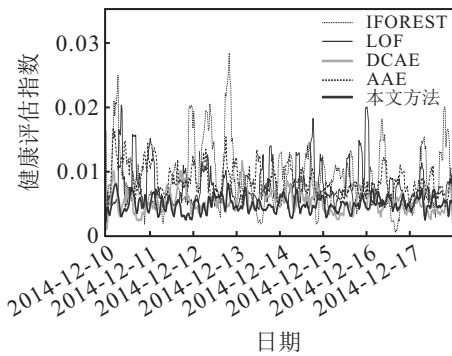


图4 不同模型构建的健康评价指数

为了量化各个模型提取 $H$ 指数的性能,采用表2的参数指标对各个模型提取的正常状态下 $H$ 指数进行评价.从表2可以看出,本文模型提取的正常状态下的 $H$ 指数的方差和变异系数都明显小于其他模型建立的 $H$ 指数,具有更好的状态表征一致性.

表2 健康评估指数的比较

方法	IFOREST	LOF	DCAE	AAE	本文方法
均值	0.0090	0.0077	0.0055	0.0086	0.0048
方差/ $10^{-6}$	23.5810	9.6713	2.5551	5.2547	1.2773
变异系数	0.5390	0.4064	0.2932	0.2680	0.2336

### 3.4 WT故障检测结果

采用正常状态下健康评估曲线建立EWMA模型,得到正常状态下 $H$ 序列的预测值,结合式(19)和(20)计算得到正常状态下的健康基线.图5给出了2014年12月10日~17日期间#1WT健康基线,健康基线对 $H$ 指数有较好的跟随性,两者的波动趋势基本一致.可见,本文设定故障阈值能够随着对#1WT的环境工况的变化而自适应调节,充分考虑了WT SCADA数据的时序性与相关性.

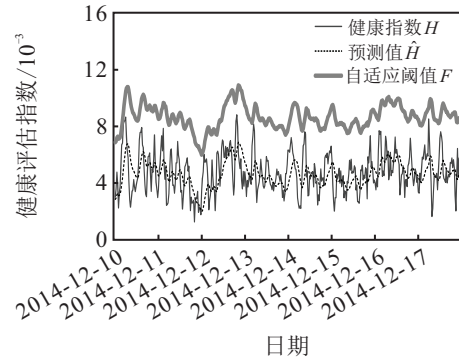


图5 #1WT正常状态下的 $H$ 指数

将2014年12月18日~25日(故障前后8天)的#1WT SCADA数据的1152个关联矩阵作为测试样本,输入至训练好的IMVAE网络,得到这8天的风电机组 $H$ 指数,如图6所示.结合测试样本 $H$ 指数对图5中的正常状态下的EWMA预测模型进行更新,得到每个测试时间点上的指数预测值 $\hat{H}_t$ ,继而计算对应故障阈值 $F_t$ .通过比较测试数据的 $H$ 值与其对应的故障阈值 $F_t$ ,进行故障异常检测,结果如图6所示.

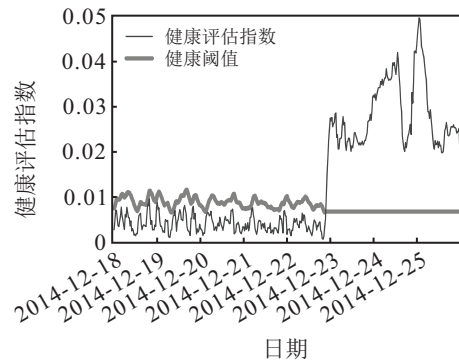


图6 #1WT测试结果

从图6可知,在2014年12月22日之前 $H$ 指数均小于故障阈值,#1WT运行正常.在12月22日23:00之后,模型的健康评估指数值超出对应的健康阈值,且呈上升趋势后产生大幅波动,但始终高于健康阈值,据此判定风电机组运行状态异常,并进行故障预警.实际情况为该风机在23日凌晨出现异常,3天后进行维护检修时,发现齿轮箱温控阀损坏,进行停机更换后恢复正常,预警结果符合实际情况.

### 3.5 验证实例

采用来自于爱尔兰南部海岸附近3 MW直驱式WT(编号为#2)的SCADA监测系统数据<sup>[21]</sup>进一步验证本文所提出方法在WT运行状态异常检测中的有效性与可行性. 选取该机组在2014年11月14日~2015年1月4日期间的SCADA数据进行分析, 并对其清洗过滤和归一化处理, 得到7494组样本数据. 状态数据记录显示机组2014年11月14日~2014年12月27日期间整体运行正常, 在2015年1月1日18时左右出现故障, 于2015年1月5日中午进行停机维修. 针对直驱式WT, 选择表3中的监测参量计算关联耦合矩阵.

表3 #2WT用于构建关联矩阵的SCADA数据表

序号	SCADA 参数	序号	SCADA 参数
0	平均风速	8	转子温度
1	平均旋转角	9	定子温度
2	平均功率	10	机舱温度
3	平均桨距角	11	主轴温度
4	旋转轴承温度	12	整流器柜温度
5	前轴承温度	13	环境温度
6	后轴承温度	14	控制柜温度
7	变桨柜叶片端温度	15	变压器温度

将该风机于2014年11月14日~2014年12月27日正常运行期间SCADA数据样本计算得到的5623个关联矩阵作为训练集建立MIVAE模型. 2014年12月28日~2015年1月4日的1152个测试样本矩阵得到的检测结果如图7所示.

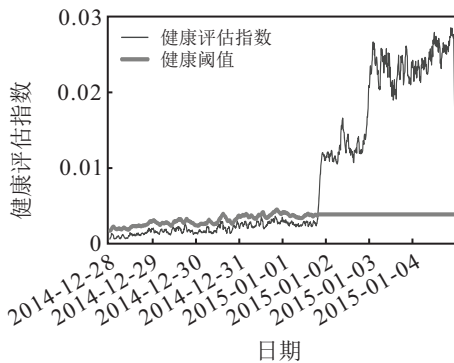


图7 #2WT测试结果

2014年12月28日~2015年1月4日期间的H指数在1月1日之前均小于健康阈值, 风电机组运行正常; 在1月1日18时之后, 模型的健康评估指数超出对应的健康阈值, 且呈现先后两次较大幅度上升后在一定范围波动, 据此判定#2WT运行状态异常, 并进行故障预警. 实际情况为, 该风机在1月1日18时50分出现馈电故障, 1月5日中午机组停止运行, 维修后于1月7日恢复正常, 预警结果符合实际情况.

为了验证本文方法在#2WT故障检测中的优

势性, 在同等数据支撑条件下采用IFOREST、LOF、DCAE及AAE四种常见的异常检测模型, 对实例中的SCADA数据进行分析, 并采用正确率(accuracy), 精确率(precision)、召回率(recall)及 $F_1$ 值来度量检测的准确性, 结果如图8所示. 同时, 图8还给出了以#2WT SCADA原始参量序列 $D_k$ 作为IMVAE模型的输入时的检测结果. 本文方法的正确率与精确率分别为99.31%与98.30%, 召回率为1,  $F_1$ 值为0.98, 均高于其他方法; 与采用 $D_k$ 作为模型输入相比, 采用多参量关联矩阵作为模型输入的方法运算效率更高, 更具优越性, 能够更有效地提高IMVAE在风电机组故障检测中的准确度.

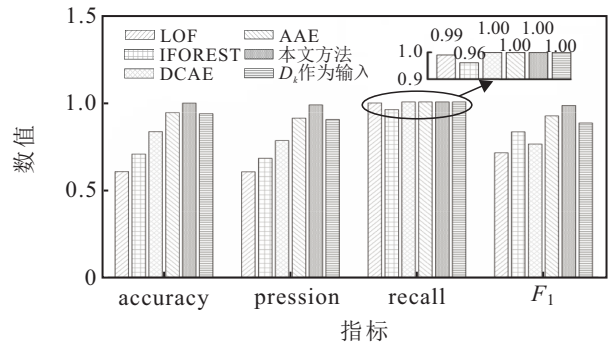


图8 检测结果比较

## 4 结论

1) 采用基于DCCA的耦合关联矩阵对SCADA数据多维参数间的关联耦合性进行量化表征, 能在消除信息冗余的情况下, 揭示WT健康状态变化引起的内部结构信息变化.

2) 构建的MIVAE模型能减少关联矩阵在编码解码中的信息损失, 其采用隐变量分布估计与特征逼近方法, 能够提高对多参量关联矩阵的重构复现能力.

3) 在充分考虑WT运行状态变化的时序性与相关性的情况下, 采用EWMA的更新迭代法对故障阈值进行自适应设置, 能够克服传统故障阈值通过人为设置的主观性, 降低复杂工况环境下WT运行状态的误警率.

### 参考文献(References)

[1] 张晓红, 曾建潮, 石冠男, 等. 基于退化状态空间划分的风电机组视情维修决策[J]. 控制与决策, 2019, 34(9): 1909-1916.  
(Zhang X H, Zeng J C, Shi G N, et al. Optimal decision of condition-based maintenance of wind turbines based on deterioration state-space partition[J]. Control and Decision, 2019, 34(9): 1909-1916.)

[2] 姜娜, 严蜜, 李柠. 基于时序上采样卷积神经网络的风机叶片结冰检测[J]. 控制与决策, 2022, 37(8): 2017-2025.  
(Jiang N, Yan M, Li N. Icing detection of wind

- turbine blade based on the time-dimensional upsampling convolutional neural network[J]. *Control and Decision*, 2022, 37(8): 2017-2025.)
- [3] 林涛, 赵成林, 刘航鹏, 等. 基于改进栈式自编码器的风电机组发电机健康评估[J]. *计算机工程与科学*, 2020, 42(3): 517-522.  
(Lin T, Zhao C L, Liu H P, et al. Health assessment of wind turbine generator based on improved stacked auto-encoder[J]. *Computer Engineering & Science*, 2020, 42(3): 517-522.)
- [4] 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.  
(Huang R Z, Tang B P, Yang Y N, et al. Condition monitoring of wind turbine gearbox based on lstm neural network fusing scada data[J]. *Acta Energiæ Solaris Sinica*, 2021, 42(1): 235-239.)
- [5] Renström N, Bangalore P, Highcock E. System-wide anomaly detection in wind turbines using deep autoencoders[J]. *Renewable Energy*, 2020, 157: 647-659.
- [6] Yang W G, Liu C, Jiang D X. An unsupervised spatiotemporal graphical modeling approach for wind turbine condition monitoring[J]. *Renewable Energy*, 2018, 127: 230-241.
- [7] 赵洪山, 刘辉海. 基于深度学习网络的风电机组主轴轴承故障检测[J]. *太阳能学报*, 2018, 39(3): 588-595.  
(Zhao H S, Liu H H. Fault detection of wind turbine main bear based on deep learning network[J]. *Acta Energiæ Solaris Sinica*, 2018, 39(3): 588-595.)
- [8] 魏书荣, 张鑫, 符杨, 等. 基于GRA-LSTM-Stacking模型的海上双馈风力发电机早期故障预警与诊断[J]. *中国电机工程学报*, 2021, 41(7): 2373-2383.  
(Wei S R, Zhang X, Fu Y, et al. Early fault warning and diagnosis of offshore wind DFIG based on GRA-LSTM-stacking model[J]. *Proceedings of the CSEE*, 2021, 41(7): 2373-2383.)
- [9] Yang W X, Court R, Jiang J S. Wind turbine condition monitoring by the approach of SCADA data analysis[J]. *Renewable Energy*, 2013, 53: 365-376.
- [10] 李霸军, 杨茜芝. 基于SCADA数据的风电机组传动系统故障状态监测[J]. *西北水电*, 2018(1): 116-119.  
(Li B J, Yang Q Z. SCADA-data-based monitoring of failures in WTG transmission system[J]. *Northwest Hydropower*, 2018(1): 116-119.)
- [11] 张帆, 刘德顺, 戴巨川, 等. 一种基于SCADA参数关系的风电机组运行状态识别方法[J]. *机械工程学报*, 2019, 55(4): 1-9.  
(Zhang F, Liu D S, Dai J C, et al. An operating condition recognition method of wind turbine based on SCADA parameter relations[J]. *Journal of Mechanical Engineering*, 2019, 55(4): 1-9.)
- [12] Gill S, Stephen B, Galloway S. Wind turbine condition assessment through power curve copula modeling[J]. *IEEE Transactions on Sustainable Energy*, 2012, 3(1): 94-101.
- [13] 凌点熠. 非平稳时间序列的多重分形时间加权去趋势偏互相关分析[D]. 湘潭: 湘潭大学, 2018.  
(Ling D Y. Multifractal temporally weighted detrended partial cross-correlation analysis of nonstationary time series[D]. Xiangtan: Xiangtan University, 2018.)
- [14] 刘春琼, 刘萍, 吴生虎, 等. 基于DCCA方法分析气候变化对四川省粮食产量的影响[J]. *中国农业气象*, 2016, 37(1): 43-50.  
(Liu C Q, Liu P, Wu S H, et al. Impacts of climate change on grain output in Sichuan based on the detrended cross-correlation analysis method[J]. *Chinese Journal of Agrometeorology*, 2016, 37(1): 43-50.)
- [15] 张聪, 朱永生, 杨敏燕, 等. 采用多变量耦合网络与变分图自编码器的机械设备异常检测方法[J]. *西安交通大学学报*, 2021, 55(4): 20-28.  
(Zhang C, Zhu Y S, Yang M Y, et al. Anomaly detection method with multivariable coupling network and variational graph autoencoder[J]. *Journal of Xi'an Jiaotong University*, 2021, 55(4): 20-28.)
- [16] Krizhevsky A, Hinton G. Learning multiple layers of features from tiny images[R]. Toronto: University of Toronto, 2009.
- [17] Haq A, Khoo M B C, Ming H, et al. Enhanced adaptive multivariate EWMA and CUSUM charts for process mean[J]. *Journal of Statistical Computation and Simulation*, 2021, 91(12): 2361-2382.
- [18] Domingues R, Filippone M, Michiardi P, et al. A comparative evaluation of outlier detection algorithms: Experiments and analyses[J]. *Pattern Recognition*, 2018, 74: 406-421.
- [19] Ma H H, Hu Y, Shi H B. Fault detection and identification based on the neighborhood standardized local outlier factor method[J]. *Industrial & Engineering Chemistry Research*, 2013, 52(6): 2389-2402.
- [20] Mao J X, Wang H, Spencer B F Jr. Toward data anomaly detection for automated structural health monitoring: Exploiting generative adversarial nets and autoencoders[J]. *Structural Health Monitoring*, 2021, 20(4): 1609-1626.
- [21] Leahy K, Hu R L, Konstantakopoulos I C, et al. Diagnosing wind turbine faults using machine learning techniques applied to operational data[C]. *IEEE International Conference on Prognostics and Health Management*. Ottawa, 2016: 1-8.

## 作者简介

刘小峰(1980—), 女, 教授, 博士生导师, 从事设备状态监测与故障诊断、智能测试与仪器等研究, E-mail: liuxfeng0080@126.com;

史长振(1997—), 男, 硕士生, 从事机械设备状态监测与故障诊断、剩余寿命预测等研究, E-mail: changzhenshi@163.com;

晏锐(1996—), 女, 硕士生, 从事机械设备故障异常检测、剩余寿命预测等研究, E-mail: 2452131005@qq.com;

柏林(1972—), 男, 教授, 博士生导师, 从事机械设备故障诊断与寿命预测、非平稳信号处理等研究, E-mail: bolin0001@aliyun.com.