

基于多尺度全卷积神经网络的核电主泵状态异常检测方法

龚安, 魏金铭

中国石油大学(华东)青岛软件学院, 计算机科学与技术学院, 青岛 266580

摘要 为解决目前主泵状态异常检测复杂数据建模难、早期异常检出难和异常程度评估难的问题,提出一种基于多尺度全卷积神经网络的编解码器网络结构异常检测方法 MSFCNAD (multi-scale FCN-based anomaly detection)。在考虑主泵状态多变量时序特征的基础上,利用全卷积神经网络编解码进行像素级训练,精准定位核电主泵状态数据的异常范围;同时兼顾主泵状态异常时段特征,提取核电主泵状态的多尺度特征矩阵,通过不同尺度的特征矩阵所检测到的异常范围分级判断其异常程度。在此基础上,利用真实核电主泵数据进行实验,对比自回归滑动平均模型(ARMA)、双向长短期记忆网络(BiLSTM)、全卷积神经网络(FCN)、自动编码器(AEs)等多个模型的分类效果。结果显示,MSFCNAD模型的召回率、 F_1 分数均优于文中列举的模型,分别为78.44%、80.30%,优于其他模型中最高的77.53%、69.74%,表明MSFCNAD模型比其他异常检测方法的性能更好,同时能够通过异常程度判断异常的严重性,在故障发生前优先维护,保障主泵正常运行。

关键词 异常检测;多变量时间序列;全卷积神经网络;多尺度分析;深度学习

核电作为一种清洁高效能源符合绿色转型的要求,已经成为世界能源发展的大趋势。然而核电的发展机遇与挑战并存。纵观历史上几次大的核电事故,无一不造成不可估量的损失,甚至至今仍受到核辐射的影响,因此确保核电安全十分重要。核电安全受多方面影响,保证核电设备完全稳定运行至关重要。作为一回路系统能量传递的唯一动

力源,核电主泵的安全运行尤其重要。主泵传感器能否准确收集和处理状态信号进行反馈不仅对核电站的安全运行至关重要,而且还影响设备维护成本、设备使用寿命以及整体发电效率。因此,针对核电主泵运行状态有效监测和异常检测,及时检测出早期的异常并进行维护处理,保障核电站安全运转是不可或缺的。

收稿日期:2022-10-09;修回日期:2022-11-11

基金项目:中石油重大科技项目(ZD2019-183-004);中央高校基本科研业务费专项(20CX05019A)

作者简介:龚安,副教授,研究方向为大数据智能处理,电子信箱:414625329@qq.com

引用格式:龚安,魏金铭.基于多尺度全卷积神经网络的核电主泵状态异常检测方法[J].科技导报,42(16):114-125;

doi: 10.3981/j.issn.1000-7857.2022.10.01513

目前,国内外关于核电主泵状态异常检测方法主要包括:基于固定阈值的方法、基于相似性的方法、基于专家系统的方法、基于统计的方法、基于偏差的方法等。目前针对核电主泵异常检测常用的方法仍以阈值判断和人工巡查为主,导致设备早期异常无法被发现;另外专家系统也在主泵上有所应用,但专家经验无法处理未知异常。由于核电主泵运行过程中涉及的传感器种类数量较多,高精度传感器传出的数据关系复杂,所以目前针对核电主泵状态异常检测的研究仍很难保证同时实现高检出率和低误检率。Maurya等^[1]提出了一种基于支持向量数据描述的鲁棒性异常检测算法,对核电站真实数据进行了大量实验证明算法的有效性,同时讨论算法在高维和非线性可分离数据中寻找异常的扩展,但仍局限于单维数据的异常检测技术。Caliva等^[2]提出了一种结合卷积神经网络(CNN)、去噪自编码器(DAE)及K均值聚类(K-means)的深度学习方法来检测核电反应堆信号异常,将核反应器的26层展开成二维图像,通过归一化的交叉关联和度量来定位异常来源,异常检出率和可解释性显著提升。You等^[3]分析核电机机械仪器信号,针对核电信号的异常情况及时随机性影响,提出了一种基于贝叶斯推论的信号异常识别新决策来确定信号状态并量化故障概率,能够提前发出警报并提供核电机机械仪器故障概率。

近年来,深度学习在复杂多变量时序数据上的研究越发深入,同各个领域交叉结合取得了不错的效果,使其在处理复杂动态系统故障诊断和异常检测中兼具应用价值和理论意义。Cheng等^[4]提出了一种多尺度长短期记忆网络(long short-term memory, LSTM)模型以完全表征不同尺度的时间依赖性。Sabokrou等^[5]最先将全卷积神经网络(full convolutional network, FCN)应用于异常检测,用级联的方式来检测和定位异常区域。陈磊等^[6]针对异常检测阶段重建误差易在阈值附近的序列无法正确判断异常这一问题,提出了长短期记忆自编码器集成框架,提高了异常检测精度。由于核电数据保密性、核能设备安全性、核电建设减缓发展等原因,目前核电主泵异常检测的相关研究还没有很好地与

深度学习领域进行结合,亟待研究出合适的深度学习核电主泵异常检测方法。

针对核电站主泵状态传感器所产生的多变量时间序列数据,提出一种基于多尺度FCN的核电主泵状态异常检测方法MSFCNAD(multi-scale FCN-based anomaly detection),通过分析主泵状态设备的真实测点监测数据及时发现异常行为并进行评估,在状态异常出现早期及时检测上报,以减少运维危险。

1 相关理论

1.1 全卷积神经网络

针对CNN下采样会丢失细节信息^[7]的缺点,用卷积层代替全连接层,Long等^[8]提出了全卷积神经网络。通过像素级的训练方式,FCN能获得更高更好的精确度、鲁棒性和运算效率。

FCN通过交替进行卷积与池化,对输入特征进行提取,如图1^[8]所示。卷积层采用卷积核与输入特征进行卷积运算,运用激活函数实现非线性变换,缩放指数线性单元(SELU)^[9]激活函数能够使样本分布自动归一化,保证训练过程中梯度不会爆炸或者消失,卷积过程及SELU激活函数非线性变换过程的数学表达式为

$$g(i) = \sum_{x=1}^m \sum_{y=1}^n \sum_{z=1}^p u_{x,y,z} * w_{x,y,z}^i + v^i, \quad i = 1, 2, \dots, q \quad (1)$$

$$y(i) = f(g(i)) = \max\{\alpha e^{g(i)} - \alpha, g(i)\}, \quad i = 1, 2, \dots, q \quad (2)$$

式中, $g(i)$ 为第*i*个卷积核所构造的特征张量; u 、 v 、 w 分别为输入数组、卷积核偏置、卷积核宽度; x 、 y 、 z 为输入特征维度; $y(i)$ 为第*i*个经SELU激活函数实现的非线性变换。

池化层是对上层数据的缩放与映射,池化过程^[10]包括最大值和均值池化,其数学表达式为

$$P_{\max}^{l(i,j)} = \max_{(j-1)w < t < jw} \{u^{l(i,t)}\}, \quad i = 1, 2, \dots, q \quad (3)$$

$$P_{\text{average}}^{l(i,j)} = \text{average}_{(j-1)w < t < jw} \{u^{l(i,t)}\}, \quad i = 1, 2, \dots, q \quad (4)$$

式中, $u^{l(i,t)}$ 为第*l*层中第*i*个特征张量的第*t*个神经

元; j 为第 j 个池化层。

反卷积层就是相对于卷积在神经网络结构的正向和反向传播^[11]中做相反的运算,也称作逆卷积、转置卷积,其数学表达式为

$$g(i) = \sum_{x=1}^m \sum_{y=1}^n \sum_{z=1}^p u_{x,y,z} \otimes w_{x,y,z}^i + v^i, i = 1, 2, \dots, q \quad (5)$$

式中, \otimes 为反卷积操作。

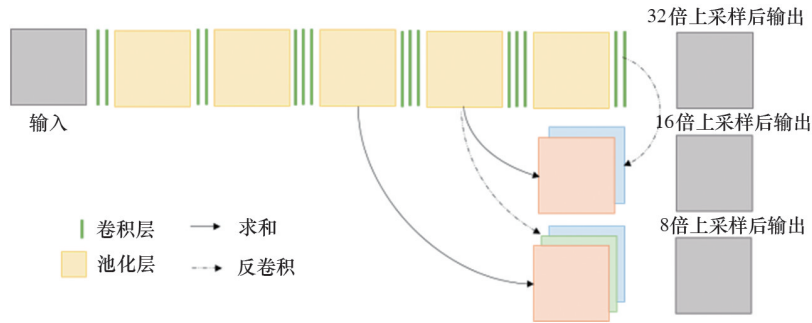


图1 FCN逻辑结构

1.2 双向长短期记忆网络

循环神经网络(recurrent neural network, RNN)是一种针对序列能够考虑时间先后顺序问题的神经网络结构。由于RNN反向传播存在梯度消失问题,无法处理较长序列,因此引入“门”的内部机制,向隐藏层神经单元中加入记忆单元^[12],提出长短期记忆网络来解决长距离依赖问题,但LSTM只能单向处理语义依赖。将前向LSTM和后向LSTM结合,形成双向长短期记忆网络(bi-directional long short-term memory, BiLSTM),用于处理双向语义依赖^[13]。

LSTM通过遗忘门、输入门、输出门这3个门控状态实现对有效信息的记忆和对无效信息的遗忘。当 t 时刻的 x_t 输入,同 $t-1$ 时刻的细胞状态 C_{t-1} 、输出 h_{t-1} ,参与运算得到 t 时刻的 h_t 输出,LSTM具体更新公式为

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (6)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (7)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (8)$$

$$C_t = f_t \otimes C_{t-1} + i_t \otimes \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (9)$$

$$h_t = o_t \otimes \tanh C_t \quad (10)$$

式中, f 、 i 、 o 分别表示遗忘门、输入门、输出门,分别储存旧信息的权重向量、获取新信息的权重向量、输出待输出的候选向量; W 、 b 分别为3个门控状态对应权重和偏置向量; $\sigma(\cdot)$ 为sigmoid激活函数;

$\tanh(\cdot)$ 为双曲正切激活函数。

BiLSTM的内部单元逻辑结构同LSTM相同,在其基础上增加了双向输入。BiLSTM不仅解决了RNN长序列依赖缺失^[14]的问题,还能同时捕获序列前后方向的特征信息^[15],从更多角度学习序列特征信息。如图2所示,BiLSTM在序列两端各构建不同方向的单向LSTM且连接于同一层,同时捕捉双向特征信息。BiLSTM神经网络层的输出为

$$h_t = W_{mh} \vec{m}_t + W_{\bar{m}h} \bar{m}_t + b_h \quad (11)$$

式中, m 为对应上标方向的记忆神经网络更新状态向量,即前一时间戳的LSTM状态。

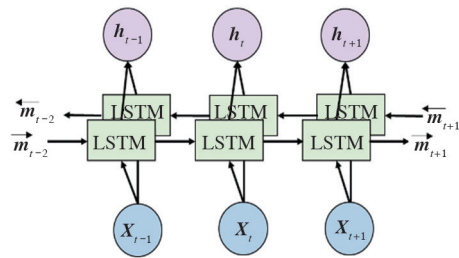


图2 BiLSTM特征提取过程

2 模型构建

2.1 问题描述

原始主泵状态时间序列数据由含有 n 个时间步长为 T 的单维时间序列组成:

$$\mathbf{X} = (x_1, x_2, \dots, x_n)^T \in \mathbf{R}_N^T \quad (12)$$

主泵状态“异常”为:与其余大多数状态数据有着明显区别的数据,或偏离其大部分正常历史状态^[16]的数据。由于主泵关系着核电安全,本着不能漏检的原则,根据检测难易程度、异常复杂程度和

异常持续时间,对主泵状态数据进行分析,争取在异常早期尽快检测出异常并量化异常程度,优先处理影响范围较大的异常。图3给出了本文异常检测模型的整体流程框架。

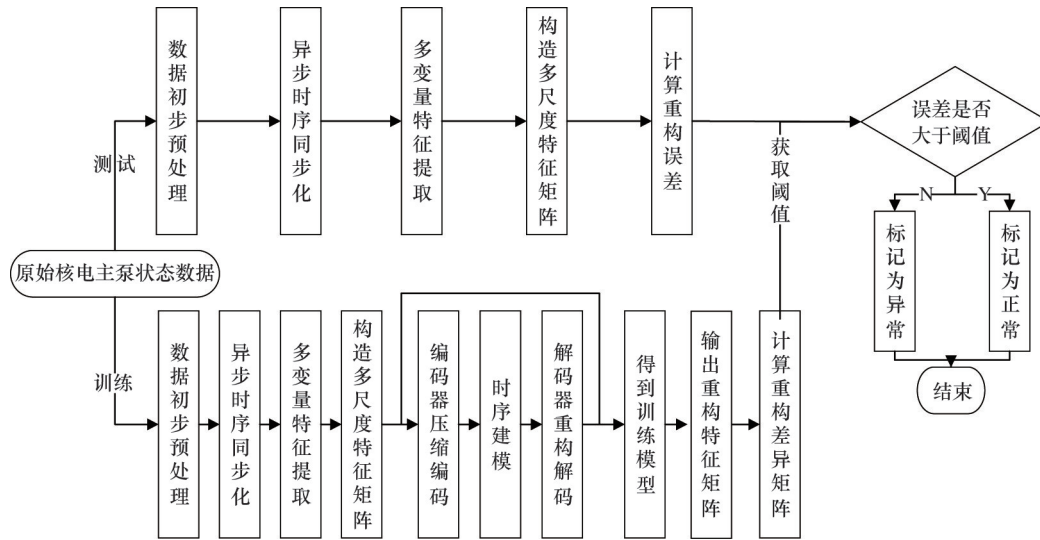


图3 异常检测模型框架

2.2 数据预处理

由于主泵状态的传感器数据存在一定的数据缺失和数据噪声,各个传感器测点^[17]之间的数据存在相关性,为了保证异常检测模型的准确性,需要对数据进行预处理。

主泵状态数据分布均匀,数据缺失较少,可能是由传感器信号无法解析或者传输信号数据出现损失引起,缺失数据占比低于 10^{-5} ,数据缺失量较少,因此采用加权KNN(k -nearest-neighbors)^[18]近似结果补充缺失数据是可行的。找到缺失点附近数据最相似的若干个历史数据点,对缺失值进行填补。同时,通过同组数据集建立判定树模型,对近似结果进行判定,判定数据填充是否对整体数据造成影响。

将原始主泵状态时间序列进行归一化处理后,其单维主泵状态时间序列表示为

$$x_i^t = \left[\sum_{s=0}^S \alpha_{i,s} \sin(\omega_s(t + \beta_{i,s})) \right] + \gamma_i + \varepsilon_i^t \quad (13)$$

式中, α 、 ω 、 β 、 γ 分别为时间序列的振幅、角频率、相位、基; ε 为噪声,约等于 $\alpha_{\varepsilon, i, t}$ 。

利用快速傅里叶变换将时域信号转化为频域信号,提取时域信号的频率信息,找到异步时序的对齐参考系原点 o' ,利用正弦变换对时间序列的相位进行变换,得到具有同步性^[19]的主泵状态时间序列。

$$x_i^t = \left[\sum_{s=0}^S \alpha_{i,s} \sin(\omega_s(t + \beta_{i,s} - \beta_i')) \right] + \gamma_i + \varepsilon_i^t \quad (14)$$

其中,变换相位和变换噪声分别为

$$\beta_i' = \beta_{0,i} + (t_i - o') \quad (15)$$

$$\varepsilon_i^t = \varepsilon_i' \sin \left[\omega_0 \left(\frac{v_0}{4} - \beta_i' \right) \right] \quad (16)$$

通过傅里叶变换和sin变换将其同步化后,主泵状态数据的部分异步特征及其同步表示如图4所示。

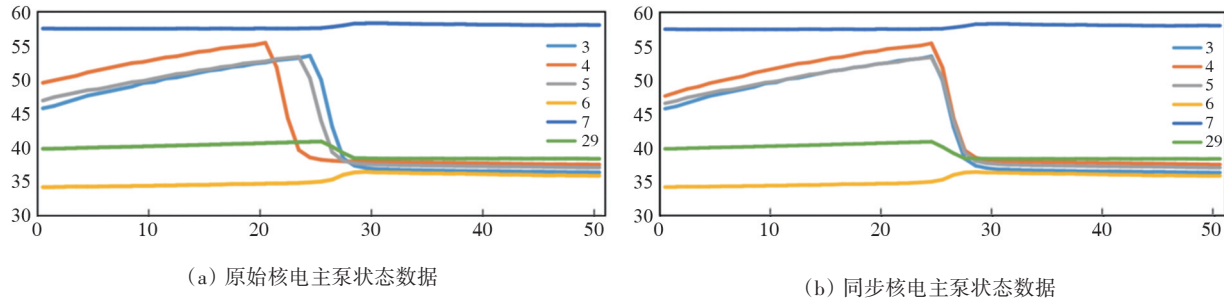


图4 异步时序的同步化

2.3 多尺度特征矩阵提取

在主泵实际运行过程中,由于设备属性、运行工况、周围环境的差异,主泵的状态监测参量、多维参量之间的相关关系比较复杂。针对多个变量之间的相关关系,通过两两参量之间的相关性建立特征矩阵^[20]获得。序列之间的特征矩阵不仅能够获取两两序列间的形状相关性和值尺度相关性,而且时序数据的不稳定性对其影响较小,导致特征矩阵对输入噪声有一定的鲁棒性。

对于时间窗大小为 a 的两个短序列 x_i, x_j 的相关性可以表示为

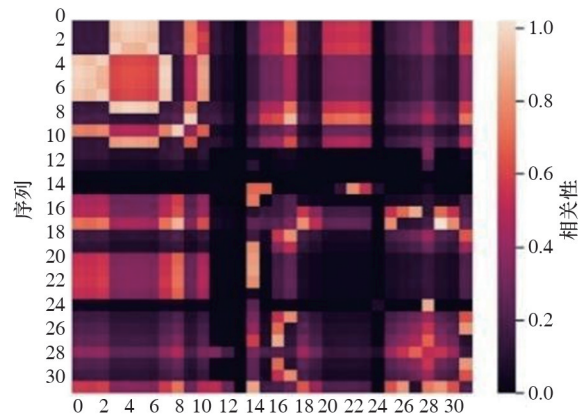
$$m_{i,j}^t = \frac{1}{a^2} \sum_{k=0}^a \langle x_i^{t-k} x_j^{t-k} \rangle \quad (17)$$

n 个序列间的两两相关性依次随着时间窗向后移动计算,最终得到 n 个序列的相关性矩阵,即特征矩阵 $\mathbf{M}_{n \times n}^a$ 。对于不同尺度下的主泵状态,采用不同的时间窗大小 a 依次向后提取得到多尺度^[21]特征矩阵,用于描述不同时间步长所对应的主泵状态。

根据主泵异常状态的时间长短,设定尺度 $m=3$,得到分别为 10、60、240 大小的时间窗,即多尺度特征矩阵 $\mathbf{M}_{n \times n}^a$ 的 a 分别为 10、60、240。特征矩阵如图 5 所示,横纵坐标为 32 个序列,所交叉点为两序列间的相关性,从 0 到 1 数值越大,相关性越强。同一时间序列不同尺度的特征矩阵根据所提取的特征信息进行比较,所检测出的异常可能不同,因此根据异常检测的难易程度对异常进行评估。

2.4 MSFCNAD 模型

主泵状态数据是具有时间相关性的多变量时间序列数据,为了实现对主泵状态异常的有效检测

图5 $a=60$ 时所提取的特征矩阵

和程度评估,在编码器-解码器结构^[22]的基础上,设计了一种新的异常检测模型 MSFCNAD,如图 6 所示。

编码器:采用 FCN 对含有序列间相关性的特征矩阵 $\mathbf{M}_{n \times n}^a$ 进行编码,将多个不同尺度的特征矩阵连接为初始张量 \mathbf{X}_0^t 输入多个卷积层中。编码器部分主要包含 4 个 FCN 卷积层,对于第 h 个卷积层,输入为上一层的输出张量 \mathbf{X}_{h-1}^t ,对其进行卷积操作得到输出张量 \mathbf{X}_h^t 。

BiLSTM 建模:结合多头注意力机制 (multi-head attention)^[23]和堆叠 BiLSTM^[24]对编码器的输出进行建模并作为解码器的输入,作为连接编码器和解码器的中介。利用多层堆叠 BiLSTM^[25]捕获编码器输出张量的时间依赖性,通过多头注意力机制对隐含特征分配权重,根据不同时间步长的多尺度特征矩阵自适应地选择相应的隐藏状态 α ,最终与建模得到的特征映射拼接得到建模张量 \mathbf{X}_h^t 。

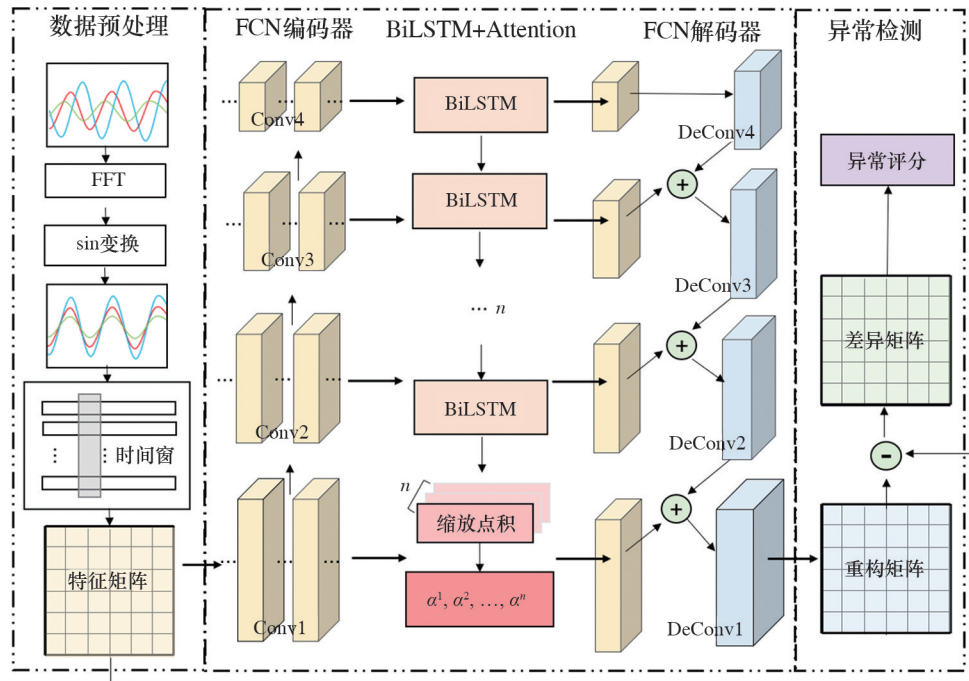


图6 MSFCNAD模型

解码器:在BiLSTM建模特征的基础上,利用FCN的反卷积层进行反卷积和连接合并操作。解码器部分主要包含4个FCN反卷积层,输入为BiLSTM建模张量 X'_h ,对于第4个反卷积层,对其输入直接进行反卷积操作得到输出张量 Y'_4 ;对于其他第 h 个反卷积层,输入与其上一反卷积层的输出张量 Y'_{h+1} 进行连接后再进行反卷积操作得到输出张量 Y'_h 。通过解码器得到相应重构矩阵 $M_{n \times n}^{a,m}$,最终输出多个不同尺度的重构特征矩阵。

2.5 误差阈值选取

将多尺度特征矩阵 $M_{n \times n}^{a,m}$ 与经过MSFCNAD模型重构后的重构矩阵 $M_{n \times n}^{a,m'}$ 相减得到差异矩阵,利用小批量随机梯度下降法^[26](mini-batch stochastic gradient descent)和Adam优化器最小化重构损失,差异重构损失定义为

$$\text{Loss}(t) = \sum_t \sum_{k=1}^m \|X_0^{t,k} - Y_0^{t,k}\|_F^2 \quad (18)$$

在训练期,通过足量的训练学习得到合适的神经网络参数来推理测试集和验证集的重构矩阵,最终得到相应的差异矩阵。利用第二极值理论中的DSPOT^[27](streaming peaks over threshold with drift)

对差异矩阵选取阈值 τ ,相对值超过阈值判定为异常。

$$\tau \approx \tau_0 + \frac{\hat{\beta}}{\hat{\gamma}} \left[\left(\frac{tN}{N_\tau} \right)^{-\hat{\gamma}} - 1 \right] \quad (19)$$

式中, τ_0 为初始阈值; $\hat{\beta}$ 和 $\hat{\gamma}$ 分别为运用最大似然估计(MLE)^[28]得到的参数估计值,表示广义帕累托分布中形状参数和尺度参数的任一值; N 为观测点个数; N_τ 为超过阈值范围的观测点个数。

2.6 异常检测模型评价指标

模型的评估少不了借助一些策略和评价指标。异常类似于分类,假设分类目标只有2类:正例(positive)和负例(negative)。目前用于评价一个机器学习模型好坏的指标都直接或间接与混淆矩阵有关,如表1所示。

表1 样本混淆矩阵

实际	预测为正常	预测为异常
正常	TP(true positive)	FN(false negative)
异常	FP(false positive)	TN(true negative)

对于分类不平衡的情况,查准率(precision)和

查全率(recall)这2个评价指标非常重要。查准率 P 又称为精确率,指正确预测占有所有预测的比例。查全率 R 又称为召回率,指正确预测占有所有正例的比例。另外, F_1 为精确率和召回率的调和平均值,其假定精确率和召回率同样重要。上述3个指标的值越大,异常检测模型的性能越佳。具体计算公式为

$$P = \frac{TP}{TP + FP} \quad (20)$$

$$R = \frac{TP}{TP + FN} \quad (21)$$

$$F_1 = \frac{2 \times P \times R}{P + R} \quad (22)$$

由于在时间序列的异常检测中,实验评判标准更侧重于异常样本的查全率 R 和综合评价指标 F_1 。因此本文主要以查全率 R 和 F_1 作为实验评价指标,以查准率 P 作为辅助标准。

3 实验与分析

3.1 实验环境

实验所使用的计算机配置如下:处理器为(英特尔)Intel(R) Core(TM) i7-6700K @ 4.00 GHz 四核;内存为16.00 GB(威刚 DDR4 2400 MHz);GPU为GeForce GTX TITAN X和RTX 3090;主硬盘为西数 WDC WD20EFRX-68EUZN0;主板为技嘉 Z170X-Gaming 3;操作系统为Ubuntu 16.04;集成开发环境为PyCharm Community Edition 2020.2,所用的Python版本为3.7.8,所用的框架为TensorFlow 2.1。

3.2 实验数据集

实验所使用的数据集以核电主泵状态数据为主。该数据集是核电站主泵状态传感器所采集的真实历史测点监测数据,共计2组机组8台主泵,采样频率1次/min,每个主泵有32个测点,主泵的测点含温度、轴承壳振、轴向位移、主泵转速、压力和压差等几种类型。数据时间范围包含7年的数据,每个主泵拥有360多万条数据,总数据量有近3000万条数据。

考虑到整体数据量较大,实验过程十分耗时,

数据可能受外界环境影响,因此选取机组1主泵4一年正常运行无异常的主泵状态数据作为训练集,选取训练集后机组1主泵4的3个月正常运行无异常的主泵状态数据作为验证集。考虑到核电主泵异常状态较为罕见,如若测试集不存在异常,最终只能分析出噪声。因此选取多个测试集,这样既能分析出异常,证明有效性,又能通过与未分析出异常进行对照证明实用性。通过对机组1主泵4历史数据进行初步分析与核电主泵现场监测维护记录对比,检测出机组1主泵4在历史数据中发生过3次异常,选取异常所在月份的主泵状态数据作为A测试集和B测试集,另选取正常运行2月的主泵状态数据集(reactor coolant pump status dataset, RCP)。另选取公共数据集Secure Water Treatment (SWaT)^[29]和Server Machine Dataset (SMD)^[30]。SWaT是由新加坡科技设计大学SUTD的iTrust机构收集的包含连续11 d数据,7 d正常操作,4 d阶段性袭击(41次)的安全用水处理数据集,数据来自51个传感器和执行器,时间粒度为1 s。SMD是连续5周采集28个不同服务器,每个服务器38个维度获得的服务器数据集,时间粒度为1 min。最终所使用的实验数据集如表2所示。

表2 实验数据集

数据集	测点	异常率/%	数据量	训练量	测试量
RCP	32	2.64	922320	527040	395280
SWaT	51	11.98	946719	496800	449919
SMD	28×38	4.16	1416825	708405	708420

3.3 多尺度FCN编解码器(MSFCNAD)构建

主要使用TensorFlow 2深度学习库构建多尺度FCN编解码器模型。利用最小-最大值标准化(min-max scaler)方法对数据进行归一化处理,采用函数式API(functional application programming interface)定义构建编码器-解码器结构的网络模型,将每层网络看作一个函数进行调用。

编码器定义为FCN全卷积层,共Conv1~Conv4这4层,其卷积核数量分别为32、64、128、256,卷积核大小分别为3×3×3、3×3×32、2×2×64、2×2×128,

Conv1层的滑动步长为 1×1 ,其余3层的滑动步长为 2×2 ,激活函数为SELU,Conv1~Conv4均返回输入张量。中间建模模块定义为基于多头注意力机制的多层堆叠BiLSTM,共5层,根据当前时间步向前注意5个时间步,隐藏层神经元个数分别为64、48、24、48、64,Dropout比率为0.2,激活函数为sigmoid激活函数和tanh激活函数,根据编码器的输入获取时序性并返回对应建模张量。解码器定义为FCN反卷积层,共Deconv1~Deconv4 4层,其卷积核数量分别为3、32、64、128,卷积核大小分别为 $3 \times 3 \times 64$ 、 $3 \times 3 \times 64$ 、 $2 \times 2 \times 128$ 、 $2 \times 2 \times 256$,Deconv1层的滑动步长为 1×1 ,其余三层的滑动步长为 2×2 ,激活函数为SELU,Deconv4的输出为建模张量反卷积得到输出张量,Deconv1~Deconv3的输出均为建模张量同上一反卷积层连接后反卷积得到的输出张量,最终通过Deconv1输出重构结果。

模型编译时,使用小批量随机梯度下降法和Adam优化器作为梯度下降优化算法(Optimizer),学习率(Learning_rate)设为0.001。模型训练时,设置批次训练样本数(Batch_size)为32,训练轮次(Epoch)为50。

3.4 数据预处理与特征矩阵

核电主泵状态数据是一个具有32个变量的多变量异步时间序列,通过傅里叶变换和sin变换将其同步化处理后,根据其序列对间的形状相关性和值尺度相关性构造特征矩阵,截取不同时间窗大小的特征矩阵,分别为10、60、240大小的时间窗。由于异常的持续时间不同,在时间窗较大的特征矩阵中,异常可能占比较低而检测不到异常。通过对3种大小的特征矩阵进行检测,判断其异常严重程度,对异常进行评估。

为了进一步确定合适的时间窗大小,对核电主泵状态序列进行分析:如果截取时间段过短,编码器在编码压缩过程无法获取时序重要特征,特征提取效果较差;如果截取时间段过长,编码器在编码压缩过程可能忽略部分时序特征,特征提取不够完整。因此通过2次实验对比调整时间窗大小,确定时间窗 $a=10, 60, 40$ 时,主泵特征提取较为完整且能够获取时序重要特征,来针对主泵状态时序不同

长度的异常情况进行检测评估。MSFCNAD在时间窗 $a=10, 60, 240$ 的情况下,平均查全率的异常检测情况如图7所示。

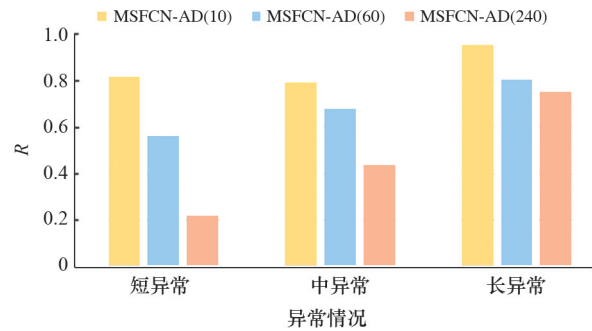


图7 MSFCNAD的异常检测情况

通过MSFCNAD的平均查全率可以看出,针对不同持续时间的异常,MSFCNAD(10)几乎都能检测出来,MSFCNAD(60)只能检测出部分时长短的异常,MSFCNAD(240)几乎不能检测出时长短的异常而且只能检测出部分中等时长的异常。因此将检测结果所对应的异常分为3类,根据时间持续时间长短判断异常的严重程度,在维修和检修时优先处理,避免危险的发生。

3.5 实验结果

为了验证本文模型MSFCNAD在复杂时序下的异常检测性能,将MSFCNAD和其他基线方法进行实验比较,比较的模型包括ARMA^[31]、BiLSTM^[32]、FCN^[33]、AEs^[34]、VAEs (variational autoencoders)^[35]、MAD-GAN (medical anomaly detection generative adversarial networks)^[36]和LSTM-ED (long short term memory encoder-decoder)^[37]。

实验结果如表3所示,表中加粗部分为评价指标中最优结果。采用核电主泵状态3个测试集进行实验验证,取平均查全率、平均 F_1 及平均查准率作为最终评价指标。其中A测试集存在2个异常且具有明显的噪声,B测试集存在1个异常,C测试集不存在异常但存在缺失值。与基线方法相比,MSFCNAD在各评价指标都取得了不错的效果,查全率和 F_1 均超过其他基线方法,查准率的评价效果也比较优异。下面对比分析MSFCNAD与其他基线方法的实验表现。

表3 不同方法异常检测结果

模型	RCP			SWaT			SMD		
	P	R	F_1	P	R	F_1	P	R	F_1
ARMA	0.7632	0.3956	0.5211	0.5271	0.4683	0.4960	0.3467	0.3915	0.3677
BiLSTM	0.5667	0.7753	0.6548	0.5849	0.2157	0.3152	0.5412	0.4723	0.5044
FCN	0.3164	0.6682	0.4295	0.5711	0.1843	0.2787	0.3614	0.2035	0.2604
AEs	0.6955	0.6048	0.6469	0.6363	0.5846	0.6094	0.4416	0.3447	0.3872
VAEs	0.6814	0.7142	0.6974	0.6667	0.7237	0.6940	0.8042	0.7741	0.7889
MAD-GAN	0.7137	0.5344	0.6112	0.9135	0.8264	0.8677	0.7654	0.8198	0.7916
LSTM-ED	0.8361	0.5137	0.6364	0.8824	0.7991	0.8387	0.7846	0.6924	0.7356
MSFCNAD	0.8226	0.7844	0.8030	0.9041	0.9768	0.9391	0.8125	0.8418	0.8269

ARMA 自回归滑动平均模型是针对非稳定序列进行异常检测的时间序列模型。从实验结果看 ARMA 模型的查全率表现较差,数据维数较多导致 ARMA 模型预测效果差无法处理这些复杂的数据场景。BiLSTM 是常用于解决时序问题的神经网络模型。BiLSTM 模型的查全率表现较好,能够较好地处理时序的长期依赖问题,但查准率表现较差,整体精确率不够。FCN 是一个主要用于提取特征学习特征的全卷积神经网络,利用卷积核在输入端对数据进行扁平化处理,能够更好的保留数据的原始信息。FCN 模型的查准率表现较差但查全率表现良好,没有处理好时序间的相关性关系。AEs 和 VAEs 是基于自编码器的深层网络,通过编码器和解码器完成数据重构,根据重构误差大小来进行异常检测。由于编解码阶段缺少额外约束,处理复杂数据时常常难以捕获有效数据特征。MAD-GAN 是基于 GAN 网络的多变量数据异常检测方法,通过调整实验设置来检查主泵状态的异常。但由于 MAD-GAN 中使用的 LSTM-RNN 建模时间相关性时间较长,无法很好的处理数据量较大的多变量时序。LSTM-ED 是基于记忆神经网络的编解码器模型,用于处理高维非线性的复杂时序。LSTM-ED 可以很好的模拟数据的相关性特征,预测较为准确,但处理数据较为缓慢。

在 3 个数据集 RCP、SWaT、SMD 上,MSFCNAD 的各评价指标都比较优异,其模型性能较其他基线方法也比较优异。3 个数据集均是多维时间序列数据,是数据量较大的复杂时序。MSFCNAD 利用

全卷积层获得了较好的精准性、运算效率和鲁棒性,提高了建模性能;同时 BiLSTM 和编解码器结构结合,保证了时序数据的相关性和长期依赖,能够很好的处理大量复杂时序数据。经实验对比分析可知,MSFCNAD 在 3 个数据集上的表现证明模型在复杂时序数据上的抗噪声干扰能力和数据异常检测能力。另外,MSFCNAD 较其他模型还具有异常评估能力,能够通过不同维度的异常先后顺序和持续时间,对异常进行分级评估。

图 8 展示的是数据异常时 10000 min 的情况,红线是异常出现的时刻。抽取了 32 维特征中具有代表性的特征,3、4、7 分别为定子绕组温度、推力轴承温度和导向轴承温度,12 为轴承壳振,22、24 分别为密封腔室压力和密封压差,28 为冷却水温差。第 1 次异常可以看出特征 24 先出现了明显的尖峰,明显区别于之前正常情况下的平稳状态,随后除 28 外其他特征发生激增。第 2 次异常可以看出特征 28 出现骤升后,特征 3、7、24 出现尖峰,所有特征均在异常发生后数值下降。通过以上分析,得知出现异常后,各个特征的异常程度不同,甚至可能在异常持续时间内某特征无法检测出异常。另外由于异常的持续时间不同,各测点出现的异常强烈程度不同,在维修时需要分清轻重缓急,优先处理较严重的异常测点。因此,针对核电主泵状态这一复杂时序,相比于现有核电异常检测方法,MSFCNAD 既能检测多变量时序的异常,又能对异常进行评估,得到的结果更加全面有效。

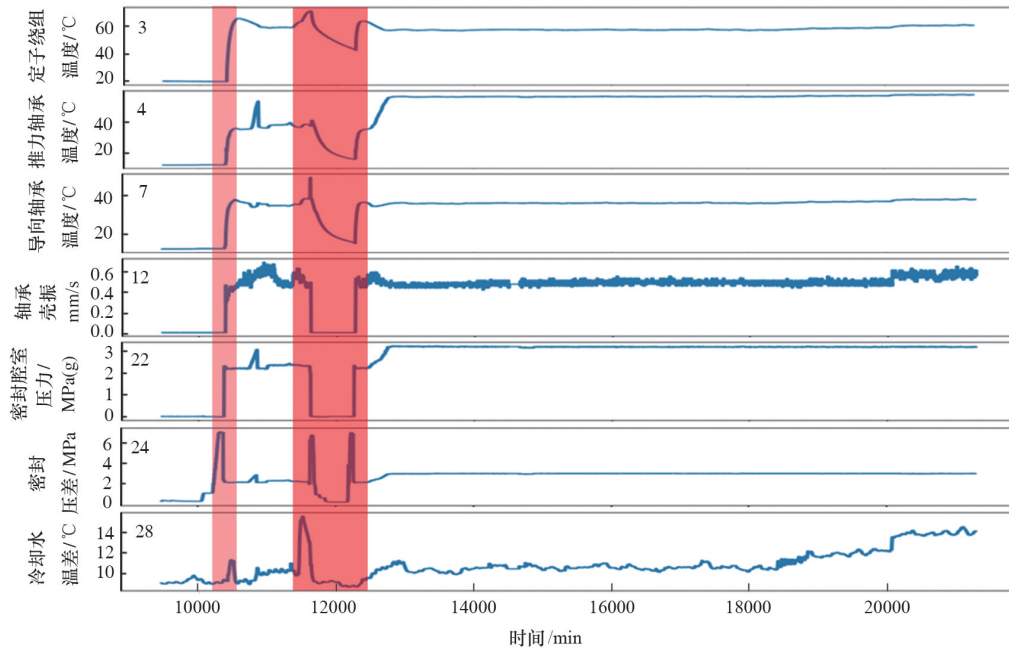


图8 案例异常检测结果分析

4 结论

针对核电主泵状态数据复杂、高维以及早期异常难检测的特点,提出了一种基于多尺度FCN的核电主泵异常检测方法MSFCNAD,引入多尺度分析的思想,在FCN、BiLSTM的基础上,利用编解码器框架构造模型。通过对比实验和性能评估证明了MSFCNAD的有效性,在核电主泵状态数据集上MSFCNAD比其他基线方法具有更好的异常检测性能,能够更好地建模复杂状态数据,检测早期异常并进行异常评估。但MSFCNAD在性能上仍有很大的改进空间,寻找更有效的特征提取方法,用Conv-BiLSTM或Transformer替代BiLSTM的模型组合等,从而进一步提升检测性能是未来研究工作的一个主要方向。

参考文献(References)

- [1] Maurya C K, Toshniwal D. Anomaly detection in nuclear power plant data using support vector data description [C]//Proceedings of the 2014 IEEE Students' Technology Symposium. Piscataway: IEEE Press, 2014: 82-86.
- [2] Calivá F, De Ribeiro F S, Mylonakis A, et al. A deep learning approach to anomaly detection in nuclear reactors[C]//2018 International Joint Conference on Neural Networks (IJCNN). Piscataway: IEEE Press, 2018: 1-8.
- [3] You D D, Shen X C, Liu G J, et al. Signal anomaly identification strategy based on Bayesian inference for nuclear power machinery[J]. Mechanical Systems and Signal Processing, 2021, 161: 107967.
- [4] Cheng M, Li Q, Lv J M, et al. Multi-scale LSTM model for BGP anomaly classification[J]. IEEE Transactions on Services Computing, 2021, 14(3): 765-778.
- [5] Sabokrou M, Fayyaz M, Fathy M, et al. Deep-anomaly: fully convolutional neural network for fast anomaly detection in crowded scenes[J]. Computer Vision and Image Understanding, 2018, 172: 88-97.
- [6] 陈磊, 秦凯, 郝矿荣. 基于集成LSTM-AE的时间序列异常检测方法[J]. 华中科技大学学报(自然科学版), 2021, 49(11): 35-40.
- [7] Szczotka A B, Shakir D I, Clarkson M J, et al. Zero-shot super-resolution with a physically-motivated downsampling kernel for endomicroscopy[J]. IEEE Transactions on Medical Imaging, 2021, 40(7): 1863-1874.
- [8] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39

- (4): 640–651.
- [9] Jian J, Xia W, Zhang R, et al. Multiple instance convolutional neural network with modality-based attention and contextual multi-instance learning pooling layer for effective differentiation between borderline and malignant epithelial ovarian tumors[J]. *Artificial Intelligence in Medicine*, 2021, 121: 102194.
- [10] Labonne M, Olivereau A, Polve B, et al. Unsupervised protocol-based intrusion detection for real-world networks[C]// 2020 International Conference on Computing, Networking and Communications (ICNC). Piscataway: IEEE Press, 2020: 299–303
- [11] Mao W D, Lin J, Wang Z F. F-DNA: Fast convolution architecture for deconvolutional network acceleration[J]. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 2020, 28(8): 1867–1880.
- [12] Ergen T, Kozat S S. Unsupervised anomaly detection with LSTM neural networks[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, 31(8): 3127–3141.
- [13] Malhotra P, Ramakrishnan A, Anand G, et al. LSTM-based encoder-decoder for multi-sensor anomaly detection[DB/OL]. arXiv preprint: 1607.00148, 2016.
- [14] Tejasri M, Lakshmi K S, Narayan K G R, et al. Unsupervised distance-based anomaly disclosure in RNN[J]. *International Journal of Computer Sciences and Engineering*, 2018, 6(3): 439–441.
- [15] 田园, 马文. 基于 Attention-BiLSTM 的电网设备故障文本分类[J]. *计算机应用*, 2020, 40(增刊2): 24–29.
- [16] 侯修群, 蒋庆磊, 包彬彬, 等. 基于相关系数的核电主泵振动异常定位方法研究[J]. *核科学与工程*, 2021, 41(5): 920–928.
- [17] 周勇, 朱鹏树, 陈星, 等. 核电厂应急柴油发电机组监测数据异常检测综述[J]. *仪器仪表用户*, 2020, 27(10): 52–55, 78.
- [18] 向玲, 朱浩伟, 丁显, 等. 基于 CAE 与 BiLSTM 结合的风电机组齿轮箱故障预警方法研究[J]. *动力工程学报*, 2022, 42(06): 514–521.
- [19] Galante L, Banisch R. A comparative evaluation of anomaly detection techniques on multivariate time series data[J]. *International Journal of Computational Science and Engineering*, 2019, 18: 17–29.
- [20] 夏英, 韩星雨. 融合统计方法和双向卷积 LSTM 的多维时序数据异常检测[J]. *计算机应用研究*, 2022, 39(5): 1362–1367, 1409.
- [21] 于旭, 何亚东, 杜军威, 等. 一种结合显式特征和隐式特征的开发者混合推荐算法[J]. *软件学报*, 2022, 33(5): 1635–1651.
- [22] Hou J, Ni K, Hawari A. An artificial neural network based anomaly detection algorithm for nuclear power plants[J]. *Transactions*, 2019, 120(1): 219–222.
- [23] Bhattacharya G, Ghosh K, Chowdhury A S. Granger causality driven AHP for feature weighted kNN[J]. *Pattern Recognition*, 2017, 66: 425–436.
- [24] Abdulaal A, Liu Z H, Lancewicki T. Practical approach to asynchronous multivariate time series anomaly detection and localization[C]// Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. New York: ACM, 2021: 2485–2494.
- [25] Zhang C X, Song D J, Chen Y C, et al. A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, 33(1): 1409–1416.
- [26] Jin M, Liu Y, Zheng Y, et al. Anemone: Graph anomaly detection with multi-scale contrastive learning[C]// Proceedings of the 30th ACM International Conference on Information & Knowledge Management. New York: ACM, 2021: 3122–3126.
- [27] Cai R C, Xu B Y, Yang X Y, et al. An encoder-decoder framework translating natural language to database queries[DB/OL]. arXiv preprint: 1711.06061, 2017.
- [28] Chen Y R, Zhang H, Wang Y N, et al. MAMA net: multi-scale attention memory autoencoder network for anomaly detection[J]. *IEEE Transactions on Medical Imaging*, 2021, 40(3): 1032–1041.
- [29] Shalyga D, Filonov P, Lavrentyev A. Anomaly detection for water treatment system based on neural network with automatic architecture optimization[DB/OL]. arXiv preprint arXiv: 1807.07282, 2018.
- [30] Kumar A, Narapareddy V T, Aditya Srikanth V, et al. Sarcasm detection using multi-head attention based bidirectional LSTM[J]. *IEEE Access*, 2020, 8: 6388–6397.
- [31] Kadri F, Harrou F, Chaabane S, et al. Seasonal ARMA-based SPC charts for anomaly detection: Application to emergency department systems[J]. *Neurocomputing*, 2016, 173: 2102–2114.
- [32] Chen X, Feng F, Wu J, et al. Anomaly detection for drinking water quality via deep biLSTM ensemble[C]// Proceedings of the Genetic And Evolutionary Computation Conference Companion. New York: ACM, 2018: 3–4.

- [33] Ye H, Ma X P, Pan Q F, et al. An adaptive approach for anomaly detector selection and fine-tuning in time series [DB/OL]. arXiv preprint: 1907.07843, 2019.
- [34] Xiao Q F, Shao S K, Wang J. Memory-augmented adversarial autoencoders for multivariate time-series anomaly detection with deep reconstruction and prediction[DB/OL]. arXiv preprint: 2110.08306, 2021.
- [35] Liu Y X, Lin Y F, Xiao Q F, et al. Self-adversarial variational autoencoder with spectral residual for time series anomaly detection[J]. Neurocomputing, 2021, 458: 349–363.
- [36] Han C, Rundo L, Murao K, et al. MADGAN: Unsupervised medical anomaly detection GAN using multiple adjacent brain MRI slice reconstruction[J]. BMC Bioinformatics, 2021, 22(2): 1–20.
- [37] Kao I F, Zhou Y L, Chang L C, et al. Exploring a Long Short-Term Memory based Encoder-Decoder framework for multi-step-ahead flood forecasting[J]. Journal of Hydrology, 2020, 583: 124631.

Reactor coolant pump status anomaly detection method based on multi-scale fully convolutional networks

GONG An, WEI Jinming

School of Qingdao Software, School of Computer Science and Technology, China University of Petroleum(East China), Qingdao 266580, China

Abstract The large number of types of reactor coolant pump condition sensors leads to three difficulties in current main pump condition anomaly detection: difficulty in modeling complex condition data, difficulty in detecting early abnormalities, and difficulty in assessing the degree of abnormality. The study of rapidly developing deep learning techniques such as neural networks can provide new ideas to solve these problems. To this end, a codec network structure anomaly detection method MSFCNAD (multi-scale FCN-based anomaly detection) based on multi-scale fully convolutional neural networks is proposed. Based on the multivariate temporal characteristics of the main pump state, this method uses the full convolutional neural network codec for pixel-level training to precisely locate the abnormal range of reactor coolant pump state data. At the same time, taking into account the main pump state abnormal time characteristics, the multi-scale feature matrix of reactor coolant pump state is extracted, and the abnormal extent is judged by the graded abnormal range detected by the feature matrix of different scales. On this basis, experiments are conducted using real nuclear main pump data to compare the classification effects of several models such as ARMA, BiLSTM, FCN and AEs. The results show that the MSFCNAD model outperforms the models listed in the paper in terms of recall and F1 score, which are 78.44% and 80.30%, respectively, better than the highest 77.53% and 69.74% of the other models. The experimental results show that this method has better performance compared with other anomaly detection methods, and it can also judge the severity of anomalies by the degree of anomalies and prioritize maintenance processing.

Keywords anomaly detection; multivariate time series; fully convolutional neural networks; multiscale analysis; deep learning



(责任编辑 刘志远)