

基于深度学习 LSTM 的风电机组故障趋势分析技术

王桂兰, 赵洪山, 米增强
华北电力大学, 河北省保定市 071003

Wind Turbine Fault Trend Analysis Based on Deep-learning LSTM Model

WANG Gui-lan, ZHAO Hong-shan, MI Zeng-qiang
North China Electric Power University, Baoding 071003

摘要: 风电机组多数故障发生前有较长时间的潜伏期, 利用实测 SCADA 数据预测故障发展趋势, 有利于早期定位故障, 对减少故障停机时间、降低机组维护代价具有重要意义。SCADA 数据属于多变量时间序列数据, 深度学习 LSTM 模型能够预测时间序列中间隔和延迟相对较长的重要事件, 本文提出基于 SCADA 数据、利用 LSTM 模型来预测风机故障。首先对 SCADA 数据根据时间进行聚合、并归一化处理, 转换为 LSTM 的 3D 格式[样本、时间步长、特征]; 然后构建状态堆叠 LSTM 时序分类神经网络模型并实现; 最后利用两组实测 SCADA 数据集分别对模型训练和测试, 实验证明此方法可有效预测风电机组故障。

关键词: 风电机组; SCADA; 人工智能; 长短期记忆网络; 故障预测

ABSTRACT: There is always a long latency period of time before most of the wind turbine failures, so predicting the development trend of faults based on the measured SCADA data is of great significance to reduce the failure time and maintenance cost, and also help to early fault localization. LSTM, a deep learning model, can predict time series data important events with long time interval and delay. For wind turbine SCADA data is multivariate time series data, so we propose to use LSTM model to predict wind turbine fault based on SCADA data. Firstly, through aggregating according to time and normalizing process, the wind turbine SCADA data is converted to the LSTM 3D format [sample, time step, feature]; then, a state stacked LSTM sequential classification neural network model is constructed and implemented; finally, two sets of measured SCADA data sets are used to train and test the model respectively. The results show that this method can effectively predict the fault of wind turbine.

KEY WORD: Wind turbines; SCADA; Artificial Intelligence; LSTM; Fault Prediction.

1 引言

近年来, 随着我国清洁能源战略的实施, 陆续投入建设了一批大规模风电场, 安装部署了大量风电机组; 然而由于实际安装需要, 风电场多位于偏远且气候条件恶劣的地区, 造成风电机组的运行工况极为复杂。为保障其安全运行, 及时掌握风电机组运行状态, 减少由于设备故障造成的停机现象, 降低维护代价, 需要对风电机组进行实时远程在线状态监测, 因此大多数风电场均安装了 SCADA(Supervisory Control And Data Acquisition, 数据采集与监视控制)系统^[1]。SCADA 系统以较高的频率实时采集风电机组运行数据, 包括各重要部件的振动、温度及功率等方面的数据。SCADA 数据具有大数据量、多类

型等特点, 蕴涵了反映风机运行状态的丰富而有价值的信息, 因此基于 SCADA 数据分析可评估、预测风机运行状态, 且相关研究已在实践中取得了良好的效果。如在文献[2]中利用风电机组温度变化趋势预测齿轮箱工作状态, 文献[3]利用数据相关性实现了风电机组健康状态评估等。

另一方面, 随着数据集规模的扩大和算法精准确度的提高, 人工智能技术在电力行业获得了广泛关注和应用^[4, 5]。利用人工智能中的深度学习算法, 基于实测 SCADA 运行数据预测风电机组故障是当前重要研究内容之一。如在文献[6]中, 基于深度自编码网络模型实现了风电机组齿轮箱故障检测; 文献[7]提出一种将去噪编码融入稀疏自动编码器的深度神经网络, 实现非监督学习的

特征提取并用于感应电动机的故障诊断；文献[8]利用深度神经网络实现了变压器故障诊断。

风电机组 SCADA 数据属于多变量时间序列数据^[9]，且多数故障在发生前有较长一段时间的潜伏期，存在一个渐变的过程。因此可借助时间序列中数据前后关系及发展趋势来预测故障，也就是算法需要具有长期记忆功能。但普通神经网络算法均较少考虑 SCADA 的多变量时间序列数据特征及数据之间的前后关联关系，同时，经典线性方法也难以适应多变量或多输入预测问题。而深度学习中的长短期记忆网络(LSTM, Long Short-Term Memory)^[10]模型适合处理前后有关联的多变量时间序列数据，因此本文提出利用 LSTM 模型，实现基于 SCADA 数据的风电机组故障预测。

LSTM 是递归神经网络(Recurrent Neural Network, RNN)的一种，可记忆长期信息，适于处理和预测时间序列中间隔和延迟相对较长的重要事件。LSTM 模型是解决长时间序列依赖问题的有效技术，该技术具有高普适性，适用于多种应用场景，已应用在电力短期负荷预测^[11]、风电场发电功率超短期预测^[12]、机械状态预测^[13]等多个领域。依据风电机组 SCADA 数据特点，利用 LSTM 记忆重要故障事件，可期待发现故障发展趋势，有效预测风电机组设备故障。

本文利用 LSTM 模型的多变量时间序列分析功能处理风电机组 SCADA 数据，构建状态堆叠时序分类模型，有效实现了风电机组设备故障的预测。

2 LSTM 模型简介

2.1 递归神经网络 RNN

递归神经网络 RNN 是处理时间序列模型中常用且功能强大的工具之一，属于深度神经网络，适于分析相互之间有关联的数据。与普通神经网络各计算结果之间相互独立不同，RNN 每一次隐含层的计算结果都与当前输入和上一次隐含层结果相关。通过这种方式，RNN 的计算结果便具备了记忆前几次输入数据特征的功能。常用的 RNN 模型的结构如图 1 所示。

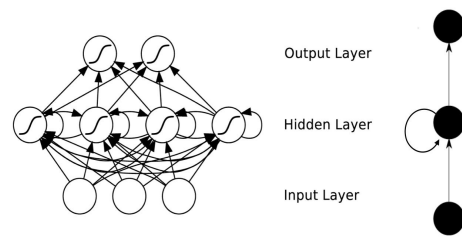


图 1 RNN 模型的结构

Fig. 1 Structure of RNN model

如图 1 所示，与一般神经网络不同，RNN 在隐藏层的节点之间建立了关联关系。RNN 模型隐藏层中每一个具体节点的展开结构如图 2 所示。

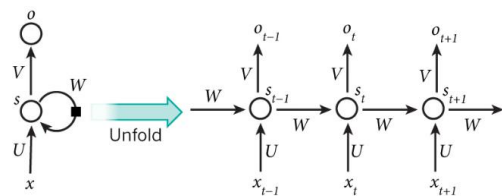


图 2 RNN 节点的展开结构

Fig. 2 The unfolded structure of RNN node

如图 2 所示， U 、 W 、 V 分别为输入到隐含层、隐含层到隐含层、隐含层到输出的权重； x_t 为神经网络输入的带有时间戳的时序数据，表示输入序列在时刻 t 的值； o_t 为神经网络在 t 时刻的输出； s_t 为隐含层当前 t 时刻的状态。RNN 模型为实现长期记忆，需要将当前隐含态的计算与前 n 次的计算结合，如公式 1 所示。

$$s_t = f(s_{t-1}, x_t) \quad (1)$$

其中， f 表示 Sigmoid 函数， x_t 表示输入序列在时刻 t 的值。但这会造成计算量随着层数增加呈指数级增长，且模型训练时间大幅增加；另一方面，存在梯度弥散或梯度爆炸问题，因此 RNN 模型一般用来进行长期记忆计算。

2.2 LSTM 模型

LSTM 模型是长短期记忆网络，是对 RNN 模型的一种改进，最早由 RNN 之父 Juergen Schmidhuber 提出。LSTM 的特点是在 RNN 结构以外为各层添加了门节点，门的状态可以打开或关闭，当门打开时，前面模型的训练结果就会关联到当前的模型计算，而当门关闭时之前的计算结果就不再影响当前的计算。因此，通过调节门的开关可以实现早期序列对最终结果的影响。如

果不希望之前结果对后续产生影响,比如自然语言处理中的开始分析新段落或新章节,那么将门的状态改为关闭即可;同样,在风机故障修复后,门的状态也应为关闭。

LSTM 节点展开后如图 3 所示,包括 3 类门:遗忘门(Forget Gate) f_t ,它决定了上一时刻的单元状态 S_{t-1} 对当前时刻状态 S_t 的影响程度;输入门(Input Gate) i_t ,它决定了当前时刻网络的输入 x_t 对 S_t 的影响程度;输出门(Output Gate) o_t ,控制单元状态 S_t 到当前输出值 o_t 的量。

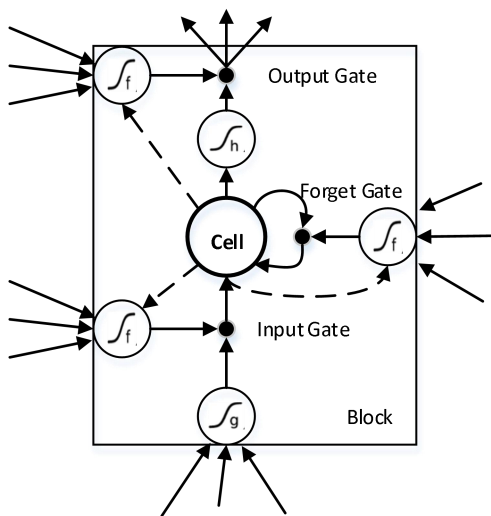


图 3 LSTM 节点的展开结构

Fig. 3 The unfolded structure of LSTM node

Cell 是核心记忆模块,是一个递归的节点,与自己的上一个状态相关,可以记住或者遗忘上一个状态,遗忘门输出 $f_t(0\sim 1.0)$ 来控制遗忘程度,类似梯度下降法中的冲量。因此得到三个门公式如公式(2)-(4)所示。

$$f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f) \quad (2)$$

$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \quad (3)$$

$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \quad (4)$$

三个公式分别对应 f_t 、 i_t 、 o_t 三个门,其输入均为 $[x_t, h_{t-1}]$,但参数不同,经过一个激活函数后,将值缩放到 $[0-1]$ 区间。

LSTM 适合于处理和预测时间序列中间隔和延迟相对较长的事件,已经在众多领域获得了

多种应用。如利用基于 LSTM 的系统来实现语言翻译、手写识别、聊天机器人的控制、音乐合成等任务。风电机组故障一般会间隔较长时间发生,且故障存在从轻微到明显的发展期,因此利用 LSTM 可以较好的处理此类数据,发现变化趋势,从而预测可能故障的发生。

3 模型构建

3.1 数据预处理

通常 SCADA 数据具有采集频率高、数据规模大等特点,需要经过预处理和特征工程来删除无效数据、修正错误数据,从而达到精简数据、提高数据质量的目的。依据 LSTM 对数据格式要求,需要将 SCADA 数据转换为 LSTM 可以处理的多变量时序数据。风电机组 SCADA 系统采集数据的主要特征及说明如表 1 所示。

表 1 SCADA 数据特征表

Tab. 1 Data feature of SCADA

序号	名称	中文含义
1	wind_speed	风速
2	gearbox_temperature_input_shaft1	齿轮箱输入轴 1 温度
3	gearbox_temperature_output_shaft2	齿轮箱输入轴 2 温度
4	gearbox_oil_temperature_gearbox	齿轮箱油温
5	generator_bearing_temperature_b	发电机轴承 b 温度
6	nacelle_temperature	机舱内部温度
7	grid_power	并网功率
8	main_bearing_rotor_side_temperature	主轴承转子侧温度
9	main_bearing_gearbox_side_temperature	齿轮箱侧主轴承温度
10	converter_motor_speed	减速比

大多数 SCADA 数据采集频率以秒为单位,数据规模庞大、数据粒度偏细,不易发现数据变化趋势,也不方便 LSTM 算法处理,因此为方便应用算法,应将数据根据不同日期粒度进行聚合,形成以分钟、小时、天为单位的聚合数据,对各个测量值取平均值。由于不同特征数据的量纲不同,通过归一化处理,转化为无量纲表达式,将数据映射到 $[0-1]$ 区间,降低不同量纲数据对结果的影响。采用的公式如下:

$$\bar{x}_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (5)$$

其中 x_i 表示采集数据, x_{\max} 表示最大值, x_{\min} 表示最小值。选取以天为单位聚合后的数据, 以某风机三个月数据(90条记录)为例, 绘制归一化后的数据展示如图4所示。

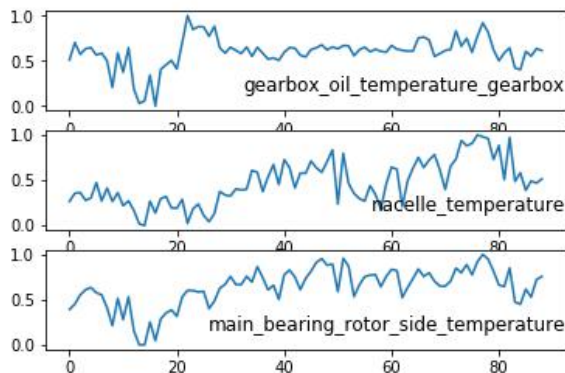


图4 归一化后的数据图

Fig. 4 Normalized data graph

LSTM 是基于时间序列的预测, 需要确定窗口长度, 即根据过去若干长度的时间来预测下一个时间周期的故障可能性。比如采用窗口长度为3, 即用 $t-2$ 、 $t-1$ 、 t 次的时间间隔进行模型训练, 然后用 $t+1$ 次时间数据对结果进行验证。考虑 SCADA 数据的特征及分析需求, 必须提供超过包含一个完整故障周期的输入时间步长; 且由于在解决序列预测问题时, LSTM 是通过时间进行反向传播的, 因此, 最终形成的数据格式如表2所示。其中采用的时间间隔为1, 即利用 $t-1$ 时刻的数据预测 t 时刻风机的状态。

表2 LSTM 数据格式

Tab. 2 LSTM data format

编号	风速($t-1$)	温度($t-1$)	功率($t-1$)	...	状态(t)
1	0.111214	0.518830	0.487279	...	0
2	0.392770	0.684834	0.635220	...	0
3	0.280142	0.624866	0.605011	...	1
4	0.440743	0.717604	0.719489	...	1
5	0.623932	0.744774	0.753545	...	1

3.2 状态堆叠 LSTM 时序分类

数据准备完成后, 即可开始搭建状态堆叠

LSTM 神经网络, 实现对 SCADA 时序数据的分类预测。一般来说, LSTM 模块的层数越多, 对高级别时间序列的学习能力就越强。但随着 LSTM 模块的层数增多, 训练时模型不容易收敛, 实践证明一般不超过3层是较好的选择。

在模型构建时, 第一步需要确定 LSTM 模块的激活函数, 每层使用不同的激活函数, 以及接收 LSTM 输出的全连接人工神经网络的激活函数, 默认为 linear; 第二步, 为了防止过拟合 Overfitting, 需确定每一层网络节点的舍弃率 Dropout, 默认值设定为 0.5; 第三步, 确定误差的计算方式, 根据应用需求的特点, 这里采用对数损失(binary_crossentropy)方式进行计算; 第四步, 确定权重参数的迭代更新方式, 这里采用常用于 RNN 网络的 RMSprop 算法, RMSProp 通过增加衰减系数可控制历史信息的获取多少; 第五步, 依据数据规模确定模型训练迭代次数 epochs 和批次数据大小 batch size; 最后, 在最外层增加一层普通的神经网络层, 用于输出结果的降维。

具体操作时, 可通过堆叠3个 LSTM 层, 使模型能够学习高级时序特征: 前两层 LSTM 返回全部输出序列, 而最后一层只返回输出序列最后一步, 舍弃时序维度, 将输入序列转化为单一向量。其内部记忆状态由处理一批次样本后重复利用, 作为下一批次的初始状态输入, 这使得处理更长的序列并保持卷积复杂度可控成为可能。

4 实验验证

4.1 实验环境及工具

为简化开发过程, 实验工具采用开源 Keras 程序包, Keras 是基于 Google 机器学习平台 TensorFlow^[14] 的深度学习框架, 其设计参考了 Torch, 使用 Python 语言编写, 是高度模块化的神经网络库, 可以帮助用户方便快捷的构建深度神经网络模型。Keras 提供一致而简洁的开放 API, 可大大减少一般应用下多数用户的工作量。Keras 将网络层、损失函数、优化器、初始化策略、激活函数、正则化方法均定义为独立的模块, 用户可以使用这些模块来构建自己的模型。Keras 也提供了 LSTM 层构造和训练函数, 方便搭建 LSTM 训练模型^[15]。

4.2 实验模型及结果

首先需要准备训练和测试两个数据集,依据已收集的包含故障信息的两组风机 SCADA 数据,将一组(#45 风机)作为训练集,另一组(#46 风机)作为测试集;即使用#45 风机的数据来拟合模型,然后用#46 风机的数据进行评估。可能的故障有:风机轴承滚动体、保持架及滚道空间没有油脂,滚动体出现轻微倾斜并有 3-4 mm 的晃动,滚动体外侧有环形摩擦痕迹。

每个数据集均包含正常和异常数据。将训练集和测试集分为输入和输出变量。最后,将输入(X)重构为 LSTM 预期的 3D 格式,即[样本,时间步长,特征]。构建四层的 LSTM 模型,其中前三层为 LSTM 层,第四层是用于预测状态的输出层,包含 1 个神经元。LSTM 层返回全部输出序列,但最后一层只返回输出序列最后一步,舍弃时序维度。输入形状是 1 个时间步长,具有 8 个特征。对应的 Keras 模型如图 5 所示。

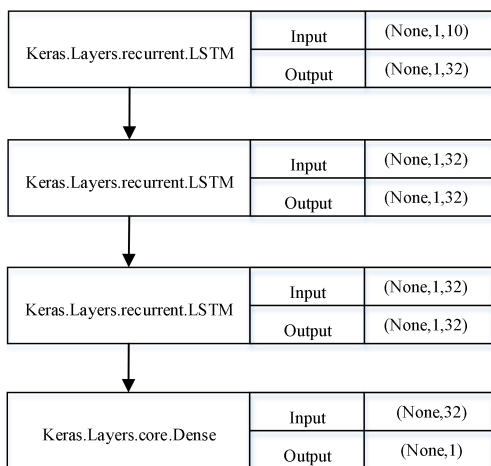


图 5 LSTM 对应 Keras 模型

Fig. 5 LSTM corresponding to Keras model

模型中使用对数损失函数计算误差,该模型将拟合 50 个批量大小为 8 的训练数据。通过在 fit() 函数中设置 validation_data 参数来跟踪训练过程中的训练和测试失败。在运行结束时,绘制训练和测试损失对应图,结果如图 6 所示。

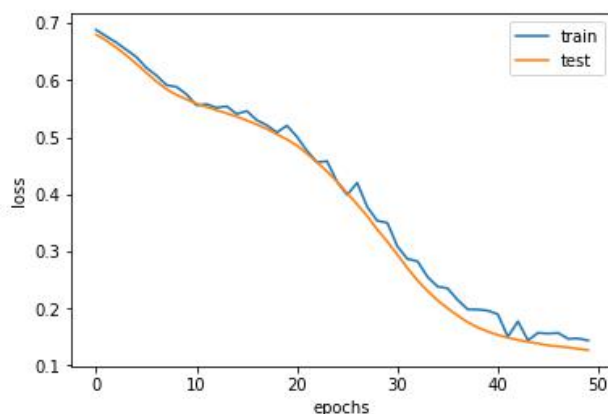


图 6 训练和测试损失对应图

Fig. 6 Loss corresponding graph of train and test

从图 6 中可以看出,随着训练次数增加,算法在训练和测试数据集上的损失逐渐降低,在 50 次左右,算法损失率为 0.1657,准确率达到 0.9775,算法趋于稳定。针对测试数据集,绘制预测值与实际值的对应关系,如图 7 所示。从图中可看出预测值和实际值基本吻合,说明了利用 LSTM 模型分析 SCADA 数据的有效性。

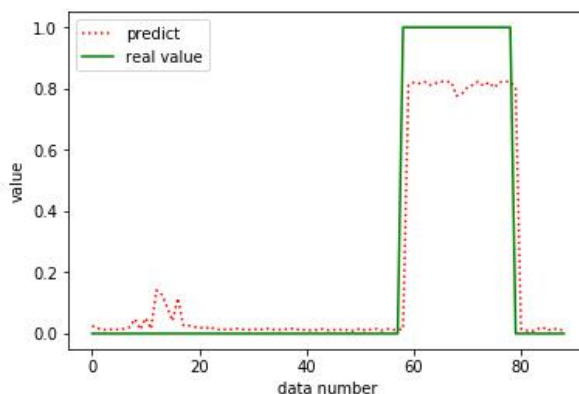


图 7 预测值与实际值对比

Fig. 7 Comparison between the predicted value and the actual value

另一组实验采用某风机一个月内采集的数据,时间间隔为 10 分钟,总数据规模为 4288 条,使用前 3904 条数据训练,后 384 条数据测试,实验结果如图 8 所示,测试准确率为 0.88。

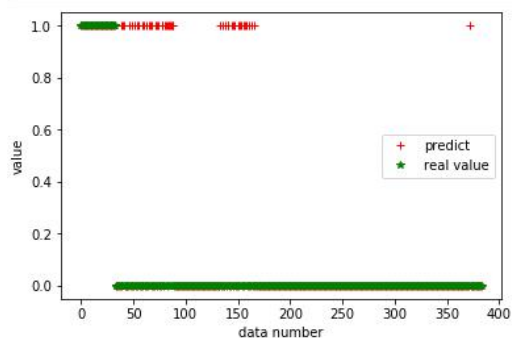


图 8 预测值与实际值对比

Fig. 8 Comparison between the predicted value and the actual value

5 结论

随着近年来人工智能技术在多个行业领域应用中取得的突破性进展,人工智能在电力行业中的应用获得广泛关注,将是智慧能源的核心支撑技术之一。本文从风电机组 SCADA 运行数据出发,分析数据特点,利用 LSTM 神经网络实现多变量时间序列数据的分析预测,并通过实际运行数据检验了模型的有效性和准确性。随着收集 SCADA 数据的进一步丰富和多样化,以及更精确人工智能算法的推出,后续将针对具体故障种类开展预测和诊断工作,提高风机故障诊断的针对性和实际应用价值。

致谢

本论文承蒙国家自然科学基金面上项目(51677072)及中央高校基本科研业务费理工类面上项目(2019MS112)支持,特此致谢。

参考文献

- [1]梁颖,方瑞明. 基于 SCADA 和支持向量回归的风电机组状态在线评估方法[J]. 电力系统自动化, 2013, 37(14):7-12. LIANG Ying, FANG Ruiming. An Online Wind Turbine Condition Assessment Method Based on SCADA and Support Vector Regression[J]. Automation of Electric Power Systems, 2013, 37(14):7-12.
- [2]郭鹏, David, Infield, 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. GUO Peng, DAVID Infield, YANG Xiyun. Wind Turbine Gearbox Condition Monitoring Using Temperature Trend Analysis. Proceeding of the CSEE, 2011, 31(32): 129-136.
- [3]Yang Wenxian, Richard Court, Jiang Jiasheng. Wind Turbine Condition Monitoring by the Approach of SCADA Data Analysis[J]. Renewable Energy, 2013: 365-376.
- [4]韩祯祥,文福拴,张琦. 人工智能在电力系统中的应用[J]. 电力系统自动化, 2000, 24(2):2-10. Han Zhenxiang, Wen Fushuang, Zhang Qi. Artificial Intelligence Applications in Power Systems [J]. Automation of Electric Power Systems, 2000, 24(2):2-10.
- [5]王守鹏,赵冬梅. 电网故障诊断的研究综述与前景展望[J]. 电力系统自动化, 2017, 41(19):164-175. WANG Shoupeng, ZHAO Dongmei. Research Review and Prospects for Power Grid Fault Diagnosis[J]. Automation of Electric Power Systems, 2017, 41(19):164-175.
- [6]刘辉海,赵星宇,赵洪山,宋鹏,邓春. 基于高度自编码网络模型的风电机组齿轮箱故障检测[J]. 电工技术学报, 2017, 32(17): 156-163. Liu Huihai, Zhao Xingyu, Zhao Hongshan, Song Peng, Deng Chun. Fault Detection of Wind Turbine Gearbox Based on Deep Autoencoder Network[J]. Transactions of China Electrotechnical Society, 2017, 32(17): 156-163.
- [7]孙文珺,邵思羽,严如强. 基于稀疏自动编码深度神经网络的感应电动机故障诊断[J]. 机械工程学报, 2016, 52(9):65-71. SUN Wenjun, SHAO Siyu, YAN Ruqiang. Induction Motor Fault Diagnosis Based on Deep Neural Network of Sparse Auto-encoder. Journal of Mechanical Engineering, 2016, 52(9): 65-71.
- [8]石鑫,朱永利. 深度学习神经网络在电力变压器故障诊断中的应用[J]. 电力建设, 2015, 36(12): 116-122. SHI Xin, ZHU Yongli. Application of Deep Learning Neural Network in Fault Diagnosis of Power Transformer[J]. ELECTRIC POWER CONSTRUCTION, 2015, 36(12): 116-122.
- [9]张艳霞. 风电机组状态监测与故障预测方法研究[D]. 华北电力大学, 2016.
- [10]Zheng S, Hao Y, Lu D, et al. Joint Entity and Relation Extraction Based on A Hybrid Neural Network[J]. Neurocomputing, 2017, 257(000):1-8.
- [11]张宇航,邱才明,贺兴,等. 一种基于 LSTM 神经网络的短期用电负荷预测方法[J]. 电力信息与通信技术, 2017(9):19-25. ZHANG Yu-hang, QIU Cai-ming, HE Xing, LING Ze-nan, SHI Xin. A Short-Term Load Forecasting Based on LSTM Neural Network[J]. Electric Power Information and Communication Technology, 2017(9):19-25.
- [12]朱乔木,李弘毅,王子琪,等. 基于长短期记忆网络的风电场发电功率超短期预测[J]. 电网技术, 2017(12): 3797-3802.
- [13]陈再发,刘彦呈,刘么源. 长短期记忆神经网络在机械状态预测中的应用[J]. 大连海事大学学报, 2017. CHEN Zai-fa, LIU Yan-cheng, LIU Si-yuan. Application of long-short term memory neural network in prediction of mechanical state[J]. Journal of Dalian Maritime University, 2017.
- [14]Abadi M, Barham P, Chen J, et al. TensorFlow: a system for large-scale machine learning[J]. 2016.
- [15]Abadi M. TensorFlow: learning functions at scale[J]. Acm Sigplan Notices, 2016, 51(9):1-1.

作者简介:

王桂兰(1979-),女,河北保定,博士,讲师,研究方向:风电机组故障检测

赵洪山(1965-),男,河北保定,博士,教授,研究方向:电力系统混杂建模、动态分析与控制以及电力系统优化计算等

米增强(1979-),男,河北保定,博士,教授,研究方向:电力系统自动化与混杂电力系统动态分析与控制