

基于计量数据挖掘分析的分布式光伏系统故障感知

刘锦宁¹, 吴裕宙¹, 苏盛², 王晓倩²

(1. 广东电网有限责任公司东莞供电局, 广东 东莞 523008; 2. 长沙理工大学电气与信息工程学院, 长沙 410014)

摘要: 分布式光伏系统点多面广, 加之可用数据严重匮乏, 难以及时感知设备故障, 容易长期带故障运行, 降低生命周期发电量。利用光伏系统故障异常最终会影响发电出力的特点提出了一种基于计量数据的分布式光伏故障感知方法。首先分析晴天的太阳辐照度特性, 提出晴空日筛选机制, 对不同电站进行相关性分析, 获取出力相关性高的光伏电站作为横向参考; 再选择待测电站不同晴空日的出力曲线进行纵向对比, 以排除异常检测中的各类干扰因素; 将排除干扰的出力数据输入时间卷积网络分位数回归(quantile regression temporal convolutional network, QRTCN)模型, 拟合出光伏正常出力区间后, 即可根据正常出力区间识别故障异常的分布式光伏电站。运用实际光伏系统数据的仿真分析表明, 所提方法能准确识别出存在故障异常的分布式光伏系统, 推动分布式光伏的精细化运维。

关键词: 分布式光伏; 故障感知; 晴空日; 分位数回归; 时间卷积网络; 相关性分析

Distributed Photovoltaic System Anomaly Detection Based on Metering Data Mining Analysis

LIU Jinning¹, WU Yuzhou¹, SU Sheng², WANG Xiaoqian²

(1. Dongguan Power Supply Bureau of Guangdong Power Grid Co., Ltd., Dongguan, Guangdong 523008, China; 2. College of Electrical and Information Engineering, Changsha University of Science and Technology, Changsha 410014, China)

Abstract: Distributed photovoltaic system has wide range of points, coupled with a severe lack of available data, making it difficult to detect equipment failures in a timely manner, and is prone to long-term operation with faults, reducing the life cycle power generation. An anomaly detection method of distributed photovoltaic system based on measurement data is proposed, taking advantage of the characteristic that abnormal faults in photovoltaic system will ultimately affect power generation output. Firstly, the characteristics of solar irradiance on sunny days are analyzed, and a clear sky day screening mechanism is proposed, conducting correlation analysis on different power stations, and obtaining photovoltaic power stations with high output correlation as horizontal references. Then the output curves of the tested power station on different clear days for longitudinal comparison are selected to eliminate various interference factors in abnormal detection. The output data excluding the above interference is input into the quantile regression temporal convolutional network model for training to obtain the fitting range of photovoltaic normal output, and then abnormal output of distributed photovoltaic power station is detected according to the normal output range. The simulation analysis using actual photovoltaic system data shows that the proposed method can accurately identify distributed photovoltaic systems with faults and anomalies, promoting the refined operation and maintenance of distributed photovoltaics.

Key words: distributed photovoltaic; anomaly detection; clear sky day; quantile regression; temporal convolutional network; correlation analysis

基金项目: 国家自然科学基金资助项目(51777015); 中国南方电网有限责任公司科技项目(031900KK52220039)。

Foundation item: Supported by the National Natural Science Foundation of China (51777015); the Science and Technology Project of China Southern Power Grid Co., Ltd. (031900KK52220039).

0 引言

作为当前技术成熟度高、成本低廉且推广潜力巨大的可再生能源发电技术,光伏发电是实现“双碳”战略目标的重要手段。光伏发电系统点多面广,散布于室外,故障概率明显高于其它发电形式,加之没有旋转部件,故障隐蔽性较强,如不能有效感知故障异常将导致大量分布式光伏长期带故障运行,明显降低全寿命周期发电量。

光伏系统故障形式多样,既可能表现为阵列开路/短路和逆变器过流/过压,也可能表现为功率管开路及热斑等形式。近年来,国内外学者提出了一些针对性的故障诊断方法,利用逆变器的电气量及光伏组件红外热像等状态量识别乃至准确定位故障,为光伏系统精细化运维检修奠定了坚实基础。既有光伏系统故障检测主要围绕光伏组件 I-V 特性曲线展开,文献[1-2]根据 I-V 曲线的形态特征建立光伏故障诊断模型,文献[3]采集动态电压电流数据估计面板内在参数。Oozeki T 等人选取光伏发电系统电气参数中的电流、电压测点数据,对比电流与电压的测量值曲线和理论曲线来确定是否发生故障^[4]。为了提高对复杂故障的检测能力,文献[5]选取短路电流和开路电压训练神经网络模型进行故障诊断,文献[6]提出了使用少量标记的时序电压电流样本的半监督梯形网络故障诊断模型。需要指出的是,前述故障诊断方法是以掌握有完备测数据为基础的,但分布式光伏与集中式光伏在多方面存在显著差异,这些方法并不适用,原因如下。

1) 分布式光伏仅通过计量系统报送上网电表计量数据,不报送其它可反应系统故障的状态量指标,系统可观性显著劣于集中式光伏。

2) 分布式光伏点多面广,主要矛盾是如何识别出需要进行现场运维的异常系统,诊断和定位故障点退化为次要矛盾,可由运维人员携带工器具现场采集状态量指标后诊断识别。

综上,分布式光伏故障诊断的目标是基于计量数据识别出异常系统,靶向性提示进行现场巡视以排查修复故障,与集中式光伏存在明显差异。尽管光伏系统故障有多种表现形式,但最终总会表现为发电量和出力的下降。根据分布式光伏可用状态监测数据有限的特点提出基于计量数据挖掘分析的分布式光伏故障感知方法。首先分析了光伏系统不同

类型故障异常的外在表征和感知光伏故障状态的潜在干扰因素;然后提出了可排除气象因素干扰的晴朗天气识别方法;再利用相关性强的电站互相作为参考排除持续性故障等干扰;之后可采用时间卷积网络分位数回归(quantile regression temporal convolutional network, QRTCN)检测超出正常区间的异常光伏,最后结合现场数据验证了所提方法的有效性。

1 光伏系统故障异常的外在表征

并网光伏发电系统主要由光伏组件、控制器、逆变器及其他配件组成,常见接线方式示意图 1 所示,其中光伏阵列先通过串联接线方式形成组,组与组之间以并联方式连接。

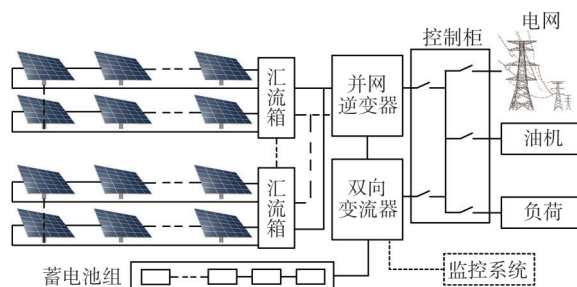


图 1 光伏系统接线示意图

Fig. 1 Schematic diagram of PV system

1.1 光伏系统故障的外在表征

光伏系统故障可分为可逆故障和不可逆故障,详细分类列如表 1 所示。可逆故障若不及时解决,多数将发展成不可逆故障。如杂物覆盖及阴影遮挡容易使光伏组件出现热斑效应,如不能及时发现和清除,不但将导致出力降低,甚至还可引发火灾造成重大损失。此外,恶劣环境会导致某些组件加速老化,造成组件间短路/开路、降低发电效率。表 1 所列故障除汇流箱烧毁较直观外都是工程实际中难以发现、可能长期潜伏恶化发展的。

科研人员围绕光伏系统故障检测提出了一系列行之有效的方法。利用光伏阵列热斑存在温度差的特点,可根据红外热像识别异常阵列^[7],根据光伏组件输出特性曲线,也可判断组件是否发生故障^[8]。文献[9]基于对不同组串电流差值的快速过采样主成分分析,识别阵列电流有异的故障组串。需要指出的是,分布式光伏缺乏必要监测数据,有的逆变器甚至还会故意虚报上网功率,很难直观判

表 1 光伏系统故障类型
Tab. 1 Type of PV system faults

可逆故障	电站设计失误	组件:失配、失效、滑脱掉落、倾斜位置不当
		逆变器:效率不实、厂家虚报
		汇流箱:接线不当、接触电阻大小错误
	恶劣环境因素	环境:灰尘、阴影遮挡
气象:高温、极端天气、云层等因素		
不可逆故障	组件	热斑、开裂、内部短路、开路
		接线老化、阵列间短路、开路
	逆变器	IGBT故障、内部过压、过流致功率管损坏,效率降低
	汇流箱	电缆老化、接线端子灼烧、汇流箱烧毁

断故障异常。

光伏系统部件故障表现各有不同,最终总会体现为出力曲线变化。图2中光伏系统完全正常时出力曲线为最上方绿色曲线;组件出现热斑时系统出力略微下降,可能表现为蓝色曲线;而逆变器内部过流、过压、功率管开路和线路老化造成效率降低时,系统出力可能大幅下降为黑色曲线;而汇流箱故障时则可能导致系统出力为0^[10-11]。

部分瞬时性故障(可逆故障)如漏电流过大可能导致出力突发性波动后会自愈或处理后恢复。面向集中式光伏的故障诊断方法利用完备的高实时状态监测数据在线检出瞬时性故障,基于计量数据的故障感知采用分钟级计量数据无法实时感知瞬时性故障,需要故障表现为稳定的出力异常下降时才可有效感知。

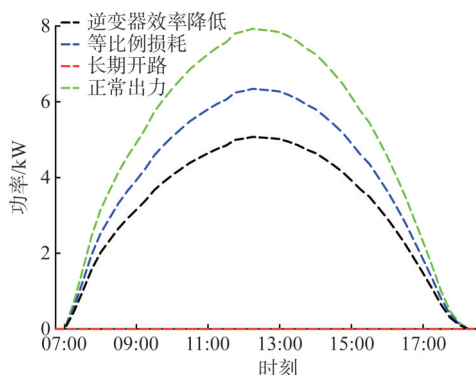


图2 光伏阵列正常出力与不同故障时出力曲线

Fig. 2 Output curves of PV array under diversified scenarios

1.2 基于计量数据的光伏故障异常诊断分析

与集中式光伏不同,分布式光伏一般没有完备的状态监测系统,可用的主要是电表报送到供电公

司的电流、电压、有功功率、无功功率和电量等计量数据。从计量系统提供的实际数据来看,电流数据还经常缺项,真正稳定可用的主要是后几项。利用故障异常最终会表现为电量和出力下降的特点有可能识别故障光伏系统。分布式光伏异常诊断中存在一些影响光伏出力的干扰因素,容易造成混淆误判,可能存在的干扰因素消除方法如下。

1)气象的变换导致光伏发电出力波动,其中阴天天气占主要影响因素。晴朗天气下无云层遮蔽,光伏出力曲线平滑无波动,日出力曲线呈拱门形。选择晴朗天气可排除云层遮蔽的影响。

2)除云层遮蔽外气温和纬度是最大影响光伏出力的因素。可利用地理位置接近时气象条件相似的特点选择同一地区的分布式光伏,排除气温和纬度因素影响。

3)光伏阵列转换效率、仰角和朝向对发电出力也有很大的影响。但此类因素在投产后保持不变,在光伏出力的横向对比中不影响诊断结果。

4)光伏系统投产后可能出现部分元件故障异常,此后发电出力会持续降额运行。可通过光伏出力的横向对比识别异常的光伏系统。

因此选择同一地区光伏系统在晴朗天气下的发电出力数据进行横向对比,可排除气象因素及光伏系统自身对发电出力异常诊断的干扰,识别出存在故障异常的光伏系统。

2 晴空日判别方法

为了排除云层对光伏系统故障感知的影响,需定义晴空日。气象领域根据天空云量定义晴空日,将云量小于25%时定义为晴天^[12]。但光伏系统部件异常时可能造成90%降额运行,出力降幅有限,而天空中有25%云量时发电出力的波动幅度足以掩盖光伏系统异常,因此需要根据需要定义晴空日。

云层遮蔽阳光引起太阳辐照度下降,光伏发电出力下降。因云层会带来辐照度在波动性上的显著差异,可根据波动性判别是否为晴空日。

图3为晴空日太阳辐照度曲线。由图3可见晴空条件下在峰值左侧各点辐照度递增,而右侧递减。因不受云层的影响辐照度曲线顺滑无突变。利用这一特性可定义识别光伏异常所需的晴空日。具体规则和实现步骤如下。

1)步骤1:太阳辐照度在灰霾天气时也可能和

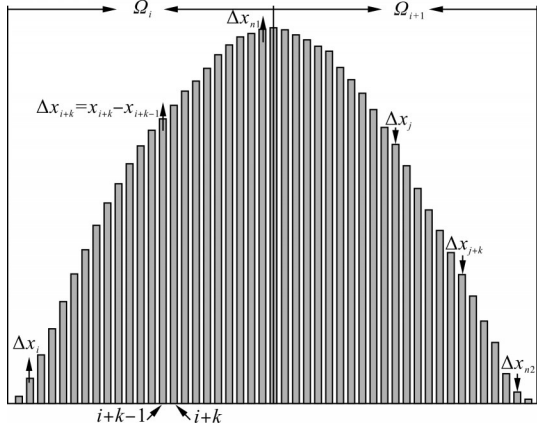


图3 晴空日太阳辐照度变化曲线

Fig. 3 Curves of solar irradiance variation on clear sky days

晴空一样顺滑增减,仅依照曲线形态难以准确地判断晴空。首先由太阳辐照度历史数据得到每天的最大辐照度数值,然后按最大值排序并选择大的前25%

四分位数作为阈值。待检测日最大辐照度小于该阈值时,说明是存在环境干扰的非晴空日,不能用于异常检测。

2)步骤2:将所筛选出晴空日辐照度数据归一化,可得15 min间隔标么化太阳辐照度数据序列 $X_j = \{x_{1,j}, x_{2,j}, \dots, x_{96,j}\}$ 。

3)步骤3:用式(1)计算日区间有效数据单位时段的差分值,构建单位时段波动量序列 $F_i = \{f_1, f_{2,j}, \dots, f_{35,j}\}$ 。 $x_{i,j}$ 为第 j 个有效数据时段内第 i 个单位时刻的辐照度大小; $f_{i,j}$ 为第 j 个有效数据时段内第 i 个单位时刻相比上一单位时刻的辐照度大小差值; F_i 为当前波动区间内单位时段辐照度波动的集合。

$$f_{i,j} = x_{i,j} - x_{i-1,j} \quad (1)$$

4)步骤4:将波动量序列逐个与0比较定义不同的波动区间 $\Omega^{[13]}$,将前后波动量正负相反的时刻标记为一个波动点;由于完全无云的天气极少,为丰富后期异常检测模型可用天数,若波动点的波动量绝对值小于阈值 β ,认为没有新的波动区间生成;否则有新的波动区间生成。将该点的上一时段作为此波动区间的结束,该点下一时段为下一个波动区间的开始,以日为单位统计波动区间数。波动区间定义如图3所示。

5)步骤5:若步骤4中统计的日波动区间数大于2,则该日为非晴空日;若等于2,则将每个波

动区间的波动量做二次差分;若二次差分绝对值都小于阈值 γ ,则认为该日为晴空日。

晴空日筛选流程如图4所示,该方法的输出结果依赖于参数二次差分阈值 γ 设置,阈值越小时筛选出的辐照度曲线越光滑,但可用天数也越少。为满足多样化场景应用需要,需结合当地天气情况和专家经验设置合理的阈值。

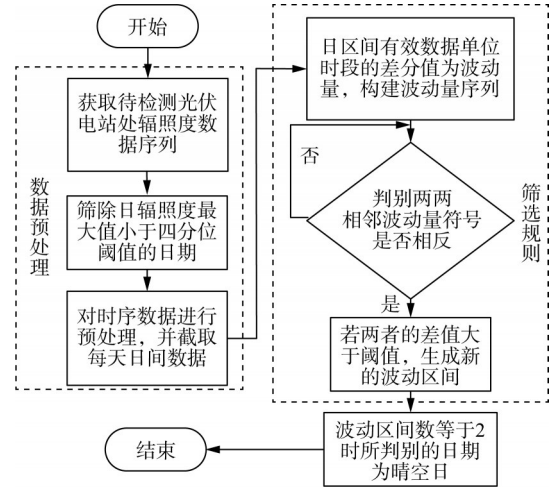


图4 晴空日筛选流程图

Fig. 4 Flow chart of clear sky day screening

3 光伏发电的出力相关性分析

3.1 邻近光伏发电系统出力相关性分析

以下分析某省两地区的40个光伏系统(分别为24个和16个),两地相隔196 km,两地区间气象条件存在一定差异,而同一地区电站气象基本一致。各地区涵盖多类型和不同倾角的光伏阵列,两地区之间及地区内部的光伏系统出力曲线存在一定差异,需计算不同电站的皮尔逊相关系数选择可互相作为参考的光伏系统。

光伏系统电表会每15 min记录和报送一次计量数据,取 N 个不同光伏系统出力数据,其皮尔逊相关系数矩阵如下。

$$\mathbf{R} = [\rho]_{N \times N} \quad (2)$$

$$\rho = r(P^j, P^k) = \frac{\sum_{i=1}^n (P_i^j - \bar{P}^j)(P_i^k - \bar{P}^k)}{\sqrt{\sum_{i=1}^n (P_i^j - \bar{P}^j)^2 \sum_{i=1}^n (P_i^k - \bar{P}^k)^2}} \quad (3)$$

式中: $\mathbf{R}$ 为同一时段 N 个光伏系统出力的相关系数矩阵; ρ 为同一时段第 j 号光伏系统和第 k 号光伏系

统出力的相关系数; P 为发电功率; n 为采样点序号; P_i^j 和 P_i^k 分别为 i 时刻第 j 号光伏系统和第 k 号光伏系统出力大小; $\bar{P}^j$ 和 $\bar{P}^k$ 分别为计量时间段内第 j 号和第 k 号光伏系统出力的期望值。

3.2 不同光伏电站出力相关性分析

根据光伏电站出力的相关性可以定量评估不同光伏电站之间的关联性。不同光伏电站之间出力相关性高说明所处的气象环境条件相似^[15], 可能是两者之间的地理位置相近, 可以互相作为横向参考, 以此来排除当某一电站具有持续性故障和气温对异常检测的干扰, 并作为同一个光伏系统进行异常检测。

对这 40 个光伏系统 1 个月 15 min 间隔的光伏出力数据进行相关性分析, 利用式(3)计算两地区不同光伏系统出力相关性, 计算结果前一地区的 24 个光伏电站相互间相关性大于 0.97, 而后一地区 16 个光伏电站间相关性也在 0.95 以上, 但两地区之间光伏出力的相关性都小于 0.86, 明显低于地区内部的相关性, 不适合用来进行异常检测。利用同一地区光伏电站出力基本一致的特点, 可相互校验进行异常检测。

3.3 同一光伏电站不同晴空日期的出力相关性分析

使用皮尔逊系数计算同一光伏系统不同晴空日下的出力相关性, 可分析如电缆损耗及触点损耗、线路老化和设备陈旧等设备固有损耗以及太阳入射角度季节性变化的影响。选取某一光伏发电系统 2022 年全年的 33 个晴空日数据计算相关性系数可见, 所有晴空日之间的相关系数都大于 0.95, 说明光伏系统设备在一年内未明显老化损耗; 另外, 相近月份的辐照度强度和气温接近, 相关性也会更高; 虽然不同月份气象条件变化较大, 晴空日下光伏出力相关性都大于 0.95, 保持了很高的相似性。因此, 同一光伏系统在不同季节晴空日的出力数据可以用来进行异常检测, 对不同晴空日下的光伏电站进行相关性分析可排除时间差异的干扰。

综上所述, 可将横、纵向相关性都大于 0.95 的光伏系统作为同一标准下的待检测系统进行下一步异常检测, 以排除光伏故障与晴空日波动引起的功率下降混淆以及同一地区光伏同时出现故障导致误排除等问题。

4 基于 QRTCN 的检测流程

光伏发电受气象条件影响的主要因素有辐照度、温度、空气质量(臭氧含量、水汽含量、林顿浑浊度)等^[16]。首先阴云和雾霾天气的干扰可以通过筛选晴空日排除; 再对不同电站进行相关性分析, 在检测电站持续性故障时排除气温等因素的干扰; 纵向对比不同日期的出力可排除同一电站阵列倾角不同和设备随时间产生老化损耗的干扰。将相关性高的晴空日出力数据输入 QRTCN 模型, 提高估计的光伏出力正常区间精准率。

分位数回归可以拟合光伏出力的条件分位数, 正常区间阈值设置为能够包含真实光伏出力曲线的分位数上下界, 进而形成晴空日下光伏正常出力区间。运维人员将待检测序列与光伏出力拟合的正常区间对比, 可判别光伏系统是否异常。与长短记忆神经网络分位数回归(long short-term memory neural network quantile regression, QRLSTM)^[17-18]相比, 在不筛选晴空日的情况下, 光伏出力被不规则的云层影响, 不具有明显的日周期性, QRLSTM 可能表现更好, 但相对于筛选晴空日下运用 QRTCN 拟合模型的性能有所降低。

4.1 线性分位数回归模型

线性回归时光伏出力拟合样本 $Y = [Y_1, Y_2, \dots, Y_i, \dots, Y_m]$ 是光伏出力真实值 $X = [X_1, X_2, \dots, X_i, \dots, X_m]$ 的响应变量, 其中, X 为解释变量, X_i 和 Y_i 分别为 i 时刻的光伏出力的真实值和拟合值, m 为样本量。则线性回归的模型参数可求解式(4)的目标函数获得, ξ 为回归系数^[19]。

$$\min \sum_{i=1}^m (Y_i - X_i \xi) \quad (4)$$

式(4)中估计参数向量 ξ 可转化为求解式(5)的优化问题。

$$\min_{\xi} \sum_{i=1}^m \rho_{\tau}(Y_i - X_i' \xi) = \min_{\xi} \sum_{i|Y_i \geq X_i' \xi} \tau |Y_i - X_i' \xi| + \sum_{i|Y_i < X_i' \xi} (1 - \tau) |Y_i - X_i' \xi| \quad (5)$$

式中: $X_i' (i = 1, 2, \dots, m)$ 为分位数条件下的解释变量; $i|Y_i \geq X_i' \xi$ 为第 i 个响应变量实际值大于等于线性回归估计值; 分位数 $\tau \in (0, 1)$; $\rho_{\tau}(u) = u[\tau - I(u)]$, $I(\cdot)$ 为示性函数, u 为自变量, 如式(6)所示。

$$I(u) = \begin{cases} 1, & u < 0 \\ 0, & u \geq 0 \end{cases} \quad (6)$$

由式(6)得出,不同分位数 τ 可得到不同的参数估计 $\xi(\tau)$,即可测算出在不同分位数 τ 处解释变量对响应变量条件分位数的影响。 τ 在 $(0, 1)$ 连续取值时,即可得到响应变量的条件分位数。

4.2 QRTCN 模型

在式(5)所示线性回归模型中限定了解释变量和响应变量之间只能是线性关系,但光伏出力拟合中之间存在非线性关系。TCN可以在残差连接的帮助下使用非常深的网络,因此可以更好地学习到长期的简单的晴天光伏出力规律,并且可以在空洞卷积的帮助下查看很远的过去进行预测,QRTCN实现过程如下。

$$p(x|\theta) = \prod_{i=0}^{N-1} P(\hat{x}_{i+1}|x_0, x_1, \dots, x_i, \theta) \quad (7)$$

式中: $p(x|\theta)$ 为参数 θ 下 x_i 的概率; N 为样本数量; x_i 为光伏出力历史值,结构采用堆叠空洞卷积层,输出层的特征映射为:

$$(w_h^l * f^{l-1}) = \sum \sum w_{h,(j,k)}^l f^{l-1}(i-dj, k) \quad (8)$$

式中: d 为空洞因子; $*$ 为空洞因子的卷积运算; l 为卷积层数量; $f^{l-1}(\cdot)$ 为第 $l-1$ 层卷积的输出函数; w_h^l 为 l 层第 h 个卷积核的值; (j, k) 为第 j 个和第 k 个输入; j 为卷积核序号; k 为卷积核的步数。

时间卷积网络分位数回归可以参考式(5),用Adam随机梯度下降法求解式(9),得到权重参数 W 和偏置项 b 。权重参数决定了网络中每个连接的强度,偏置项决定了神经元的激活阈值。

$$f_{\text{cost}} = \sum_{i=1}^N \rho_{\tau}[Y_i - f(X_i, W, b)] = \sum_{i|Y_i \geq f(X_i, W, b)} \tau |Y_i - f(X_i, W, b)| + \sum_{i|Y_i < f(X_i, W, b)} (1 - \tau) |Y_i - f(X_i, W, b)| \quad (9)$$

式中: f_{cost} 为损失函数,用于衡量模型预测值与真实值之间的不一致程度; $W = \{W_f, W_{ic}, W_o\}$ 为TCN的权重参数; $b = \{b_f, b_{ic}, b_o\}$ 为网络结构的偏置项权重; $f(\cdot)$ 为卷积神经网络模型预测值;下标 f 、 ic 和 o 分别表示卷积核的属性、输入通道和输出通道。

根据QRTCN回归模型的结构和式(9)所示的时间卷积网络分位数回归方法可建立一种QRTCN模

型,将其对网络结构参数估计的代价函数转化为式(10)所示的分位数回归目标函数,并将参数估计看作式(11)所示的优化问题,即可用Adma随机梯度下降法求解该优化问题。

$$\min_{W, b} f_{\text{cost}} + \frac{\lambda}{2} \|(\hat{W}(\tau), \hat{b}(\tau))\|^2 \quad (10)$$

$\hat{W}(\tau)$ 、 $\hat{b}(\tau)$ 是带分位数条件的 W 、 b ,求解参数 $\hat{W}(\tau)$ 、 $\hat{b}(\tau)$ 后代入式(10)可得到 Y 的条件分位数估计。

$$\hat{Q}_Y(\tau|X) = f(X, \hat{W}(\tau), \hat{b}(\tau)) \quad (11)$$

当 τ 在 $(0, 1)$ 上连续取值时,条件分位数曲线 $\hat{Q}_Y(\tau|X)$ 就被称为条件分布,文献[16]从分布函数 $F(F^{-1}(\tau)) = \tau$ 出发推导条件密度拟合如式(12)所示。

$$P(\hat{Q}_Y(\tau|X)) = \frac{d\tau}{d\hat{Q}_Y(\tau|X)} \quad (12)$$

对式(12)执行 X 条件化、 τ 离散化操作,可用密度估计得到 Y 的条件密度拟合 $P(\hat{Q}_Y(\tau|X))$,即光伏出力的概率区间。

在不同分位数下进行时间卷积网络回归得到多点拟合结果,进而采用高斯核进行概率密度估计得到概率密度函数,图5为QRTCN算法的流程图。

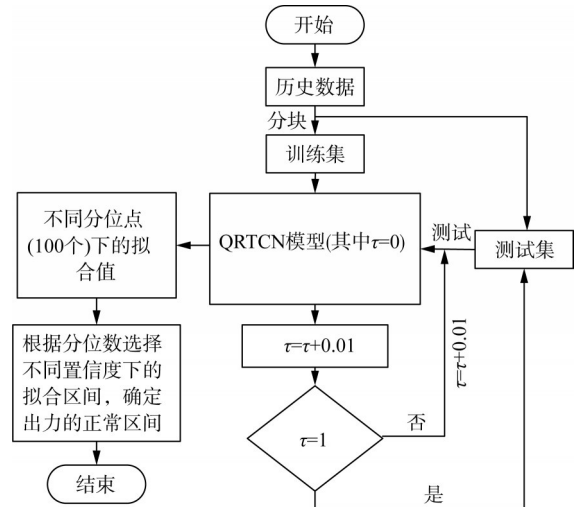


图5 时间卷积网络分位数回归模型实现流程

Fig. 5 Flow chart of QRTCN model

4.3 异常检测流程设计

应用所提模型进行光伏出力异常检测流程如下。

1)数据预处理:获取光伏出力发电功率数据及太阳辐照度数据,根据相似日对应时刻数据查漏

补缺。

2) 确定待分析的电站以及晴空日: 根据第 2 章步骤获取晴空日期, 再通过相关性分析对电站进行横、纵向对比, 排除各类干扰因素影响。

3) 计算光伏电站正常出力区间: 将晴空条件下且出力数据在同一标准下的电站输入 QRTCEN 模型训练获得光伏电站各时刻正常出力区间。

4) 故障感知: 将获取的各时刻正常出力最小值构成的出力序列作为正常最低出力的阈值, 若待检测序列任意时刻值小于对应时刻的最低正常值, 则认为存在异常。

5 仿真

5.1 算例数据

选取 2022 年全年某地 30 个光伏电站 15 min 间隔数据进行验证分析, 验证集的异常样本将光伏出力正常数据按 8% 的比例降低出力构造产生。为验证 QRTCEN 模型拟合的准确性, 首先将它与 QRLSTM 考虑确定晴空日与不确定晴空日作为异常识别场景进行对比。仿真实验计算机配置为 CPU 酷睿 i7-12700KF、内存 32 GB、GPU 3070Ti 8 GB。通过 Keras 深度学习框架将每个分位数下的 TCN 迭代 50 个轮次。图 6 为 5 d 晴空日的 90% 置信度下 QRTCEN 模型区间拟合结果, 由图可知真实值全部落在拟合区间内, 且接近于区间中部, 说明 QRTCEN 拟合结果性能较好, 具体对比分析如后文分析。

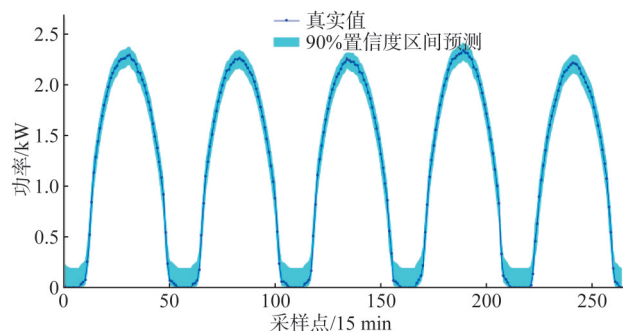


图 6 晴空日下 90% 置信度的 QRTCEN 区间拟合结果

Fig. 6 QRTCEN interval fitting results of 90% confidence in clear sky days

5.2 多场景下拟合精度分析

拟合模型评价常用的指标分为点拟合效果指标和概率拟合效果指标。其中前者的拟合效果指标主

要有平均绝对百分比误差 (mean absolute percent error, MAPE)、均方误差 (root mean square error, MSE) 和衡量回归模型拟合优度的决定系数 R^2 等统计指标; 后者常用可靠性指标 (拟合区间覆盖率) 和敏锐性指标 (拟合区间平均宽度)^[20] 等。

由于部分真实数据数值为 0, 不能采用 MAPE 评价, 以下选用 MSE 和 R^2 指标定量评价分位数中的中位数回归效果, 对不同算法在未筛选晴空日和晴空日下的拟合误差统计如表 2 所示。

表 2 不同算法在不同天气条件下拟合均方误差和 R^2

Tab. 2 MSE and R^2 of different algorithm fitting in different conditions

测试场景	MSE 值	$R^2/\%$
未筛选晴空日下 QRTCEN	0.013 1	79.60
未筛选晴空日下 QRLSTM	0.024 0	78.71
晴空日下 QRTCEN	0.001 8	98.15
晴空日下 QRLSTM	0.003 5	98.11

由表 2 误差指标可见:

1) 不筛选晴空日时, 受气象因素干扰影响, MSE 较大而 R^2 较小, 拟合效果存在较明显偏差。筛选晴空日后可排除云层干扰, 拟合效果明显提升。由于晴空日下光伏出力曲线具有明确的规律特征, 各种算法能更好地学习和利用其特征。

2) 相较于深度学习 LSTM 模型, TCN 模型对本文非线性数据的分位数回归拟合有更高的点拟合精度。主要原因是 LSTM 模型学习复杂时间序列的效果优于 TCN 算法。而光伏电站出力时序数据具有确定的日周期, 时间序列特征简单明确。

未筛选晴空日时, QRLSTM 与 QRTCEN 的性能差距大于晴空日下的性能差距。这是因为 LSTM 模型具有较强的记忆属性, 能记住非晴空日时出力曲线更复杂的特征, 从而导致晴空日条件下时间序列的过拟合。TCN 是因果关系, 并且可以将任何序列映射到相同长度的输出序列。以下通过概率区间拟合评价指标可靠性和敏锐性进行不同维度的定量评价, 各算法在不同场景的可靠性和敏锐性指标列如表 3 所示。可靠性、敏锐性是在置信度 $1-\alpha$ 条件下衡量真实值落在拟合区间、拟合区间平均宽度的性能指标, 可综合反映拟合模型的优劣。更高的可靠性和更低的敏锐性表明拟合结果更贴近真实值, 可将尽可能多的真实值包含在内, 并使拟合区间平均

宽度尽可能小。由表3可见:

表3 不同算法在不同气象条件下的拟合可靠性和敏锐性

Tab. 3 Fitting reliability and sensitivity of different algorithms in different conditions

置信度	不同算法	可靠性/%	敏锐性/kW
90%	晴空日下 QRTC�	100	0.067 3
	晴空日下 QRLSTM	74.40	0.184 9
	未筛选晴空日下 QRTC�	73.78	0.204 5
	未筛选晴空日下 QRLSTM	80.36	0.315 8
80%	晴空日下 QRTC�	92.36	0.024 7
	晴空日下 QRLSTM	51.60	0.121 1
	未筛选晴空日下 QRTC�	63.31	0.085 0
	未筛选晴空日下 QRLSTM	57.73	0.161 6

1)置信区间为90%时晴空日的QRTC�模型可靠性为100%,敏锐性为0.067 3 kW,性能远超QRLSTM。

2)未筛选晴空日时QRLSTM可靠性为80.36%,略优于QRTC�;但QRTC�的敏锐性优于QRLSTM,对规律性强的简单时间序列拟合效果更优。

3)光伏出力时间序列在晴空日下规律性强,QRTC�模型能很好拟合晴空日条件下光伏出力,为异常识别提供可靠的参考区间。

4)置信度区间为80%时各拟合模型的可靠性都有所下降,敏锐性能有所提升,但QRLSTM可靠性下降幅度大于QRTC�模型。

5.3 基于拟合正常区间的光伏系统异常识别

本文所用QRTC�模型可拟合出时间序列的短期正常区间。TCN学习正常时以日为单位的光伏出力特性并获得光伏正常出力拟合模型,再结合分位数回归可将某时刻的正常值在学习光伏正常出力的基础上拓展为区间值。本文隐式地将异常定义为不在拟合正常区间中的观测值^[21],因而可根据拟合正常区间进行异常检测,具体步骤如下。

1)运用QRTC�模型获得所选日期在置信度90%条件下的光伏出力拟合正常区间。

2)以日为单位将正常区间每日对应时刻的最小值进行比较,获得一天中每个时刻的全局最小值作为判别异常的阈值。

3)若待检测光伏电站对应时刻出力小于阈值则认为此时存在异常,可能由飘挂物遮挡所致,但若持续小于阈值则可能出现不可逆故障,需人工

巡视。

表4 不同气象环境下QRTC�模型检测结果

Tab. 4 QRTC� model detection results in different climatic environments

测试场景	误报率	检测率
晴空日下	2.25	95.15
未筛选晴空日下	10.26	89.31

根据4.2节分析,QRTC�模型拟合区间时表现较好,能为异常检测提供可靠的参考区间。分别获取晴空日与未筛选晴空日下QRTC�模型拟合正常区间,按上述检测步骤获得一日最小光伏有效出力序列作为阈值标杆。获取待检测的光伏电站数据,分别与一日最小光伏有效出力序列按照是否筛选晴空日分别进行对比判别,其检测率和误报率如表4。由表4可见,筛选晴空日可明显提高QRTC�模型拟合区间异常检测的检测率、降低误报率;如不筛选晴空日,则易被气象因素干扰将正常光伏判别为异常,牺牲人力进行不必要的巡视。

5.4 对比实验

近年来涌现了一批光伏系统故障检测与诊断方法^[22-25],可根据状态量数据的聚类与分类分析识别和定位故障。基于映射距离比离群因子的离群点检测算法(mapping distance ratio outlier factor, MDROF)是基于密度的局部离群点检测算法,在异常检测领域中有广泛应用^[26]。以下将本文所提方法与基于MDROF的异常检测进行对比分析,检测结果列如表5所示。由表4—5对比可见:

1)晴空日下MDROF异常检测的误报率偏高,检测率也明显低于本文所提方法。MDROF算法要求异常点处于低密度区域,且对参数的选取很敏感,对光伏出力异常检测适用性较弱。

2)未筛选晴空日时,受阴云天气干扰影响,误报率和检测率均劣于晴空日条件。

综上,本文所提方法采用概率性判断,在所选光伏电站出力异常检测中有更高的检测率和更低的误报率,能较好满足应用需求。

6 结语

针对分布式光伏系统难以及时感知设备故障异常、将会降低生命周期发电量的问题,提出了基于电表报送计量有功功率数据挖掘分析的光伏系统故

表 5 不同气象环境下 MDROF 算法检测结果

Tab. 5 MDROF model detection results in different climatic

environments		%
测试场景	误报率	检测率
晴空日下	31.40	92.61
未筛选晴空日下	41.29	87.31

障感知方法。提出了晴空日筛选机制,可剔除阴天天气对异常检测的干扰;利用时空关联分析,进一步排除持续性故障、大面积区域光伏系统同时损坏等干扰因素的影响;最后对化归至同一标准的光伏出力曲线进行 QRTCN 拟合,再基于拟合正常区间,可将超出的分布式光伏判断为故障异常,提高运维巡视的靶向性。并采用实际系统数据验证了所提方法的有效性。

需要指出的是,本文所提方法依赖于晴空日筛选来提高检测效果,但相当部分地区并不一定有足够频密的晴空日,为提高基于计量数据感知分布式光伏系统故障异常的实用性,未来还需要进一步研究非完美晴空日下的故障感知方法。

参考文献

- [1] 高伟,黄俊铭.基于 SSELm 的光伏组件故障智能诊断方法[J].太阳能学报,2021,42(12):465-470.
- [2] 李光辉,段晨东,武珊.基于半监督机器学习法的光伏阵列故障诊断[J].电网技术,2020,44(5):1908-1913.
- [3] WANG Wenguang, LIU Chunfor, CHUNG Shuhung, et al. Fault diagnosis of photovoltaic panels using dynamic current-voltage characteristics [J]. IEEE Transactions on Power Electronics, 2015, 31(2): 1588-1599.
- [4] TAKASHI Oozeki, TAKAO Yamada, KAZUHIKO Kato. On-site measurements of photovoltaic systems for detection of failure modules [C]//2008 33rd IEEE Photovoltaic Specialists Conference. May 11-16, 2008, San Diego, USA. New York: IEEE, 2008: 1-6.
- [5] 王元章,吴春华,周笛青,等.基于 BP 神经网络的光伏阵列故障诊断研究[J].电力系统保护与控制,2013,41(16):108-114.
- [6] CHEN Shiqun, YANG Gengjie, GAO Wei, et al. Photovoltaic fault diagnosis via semisupervised ladder network with string voltage and current measures [J]. IEEE Journal of Photovoltaics, 2021, 11(1): 219-231.
- [7] 蒋琳,苏建徽,施永,等.基于红外热图像处理的光伏阵列热斑检测方法[J].太阳能学报,2020,41(8):180-184.
- [8] JIANG Lin, SU Jianhui, SHI Yong, et al. Hot spots detection of operating PV arrays through IR thermal image [J]. Acta Energiae Solaris Sinica, 2020 41(8):180-184.
- [9] WANG Hongxia, WAN Bo, ZHANG Jiaxin, et al. Free probability theory based event detection for power grids using IoT-enabled measurements [J]. Journal of Modern Power Systems and Clean Energy, 2024, 12(5): 1396-1407.
- [10] 姜斌,龚雁峰,李岩.模块化多电平换流器子模块短路故障分析、检测与定位方法[J].南方电网技术,2019,13(3):73-78,88.
- [11] JIANG Bin, GONG Yanfeng, LI Yan. Short-Circuit fault analysis, detection and location method for sub-modules of modular multilevel converters [J]. Southern Power System Technology, 2019, 13(3): 73-78,88.
- [12] 葛磊蛟,杜天硕,孙冰.分布式光伏功率数据的 IMOWOA 和 LightGBM 混合虚拟采集方法[J].中国电机工程学报,2024,44(3):1035-1047.
- [13] GE Leijiao, DU Tianshuo, SUN Bing. Hybrid virtual collection method of IMOWOA and LightGBM for distributed PV power data [J]. Proc. of the CSEE, 2024, 44(3): 1035-1047.
- [14] 邓序之,刘淇,叶傲霜,等.基于气象耦合特征分析及改进 XGBoost 算法的用户分布式光伏短期出力预测模型[J].南方电网技术,2023,17(12):80-89.
- [15] DENG Xuzhi, LIU Qi, YE Aoshuang, et al. Short-term output forecasting model of user side distributed photovoltaic based on meteorological coupling characteristics analysis and improved XGBoost algorithm [J]. Southern Power System Technology, 2023, 17(12): 80-89.
- [16] 郭元喜,龚道溢,汪文珊,等.中国中东部秋季 PM10 时空变化及其与日气温的关系[J].地理学报,2012,67(9):1155-1164.
- [17] GUO Yuanxi, GONG Daoyi, WANG Wenshan, et al. Spatiotemporal variation of PM10 concentration and its relationship with autumn daily temperature over central and eastern China [J]. Acta Geographica Sinica, 2012, 67(9): 1155-1164.
- [18] 王兴媛,王春雨,李志,等.基于 EAST 的二次回路端子编号检测方法[J].电力科学与技术学报,2022,37(5):215-221.
- [19] WANG Xingyuan, LI Zhi, et al. Detection method of terminal number of secondary circuit based on EAST [J]. Journal of Electric Power Science and Technology, 2022, 37(5): 215-221.
- [20] DONG Xiaohong, SUN Yingyun, LI Ye, et al. Spatio-temporal convolutional network based power forecasting of multiple wind farms [J]. Journal of Modern Power Systems and Clean Energy, 2022, 10(2): 388-398.
- [21] 陆双,彭曙蓉,杨云皓,等.基于平均影响值-启发式前向搜索的异常光伏用户识别方法[J].电力自动化设备,2022,42(2):106-111.
- [22] LU Shuang, PENG Shurong, YANG Yunhao, et al. Identification method of abnormal photovoltaic users based on mean impact value and heuristic forward searching [J]. Electric Power Automation Equipment, 2022, 42(2): 106-111.
- [23] 张翔颖,杨永标,徐青山,等.基于多时段相似日理论的光伏功率组合预测方法[J].南方电网技术,2023,17(2):57-65.
- [24] ZHANG Xiangying, YANG Yongbiao, XU Qingshan, et al.

- Photovoltaic power combination prediction method based on multitemporal similarity day theory [J]. Southern Power System Technology, 2023, 17(2): 57 – 65.
- [17] MA Xiaolei, TAO Zhimin, WANG Yinhai, et al. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data [J]. Transportation Research Part C Emerging Technologies, 2015(54):187 – 197.
- [18] 曾昭法, 游悦. 基于神经网络分位数回归的金融风险定位[J]. 统计与决策, 2020, 36(14): 137 – 140.
- ZENG Zhaofa, YOU Yue. Financial risk pre-warning based on neural network quantile regression [J]. Statistics & Decision, 2020, 36(14): 137 – 140.
- [19] 李彬, 彭曙蓉, 彭君哲, 等. 基于深度学习分位数回归模型的风电功率概率密度预测[J]. 电力自动化设备, 2018, 38(9): 15 – 20.
- LI Bin, PENG Shurong, PENG Junzhe, et al. Wind power probability density forecasting based on deep learning quantile regression mode [J]. Electric Power Automation Equipment, 2018 38(9): 15 – 20.
- [20] 彭曙蓉, 李彬, 彭君哲, 等. 基于CEEMDAN分解的深度学习分位数回归电价预测[J]. 电力系统及其自动化学报, 2020, 32(4): 138 – 144.
- PENG Shurong, LI Bin, PENG Junzhe, et al. Electricity price forecasting based on CEEMDAN decomposition and deep learning quantile regression [J]. Proceedings of the CSU-EPSA, 2020, 32(4): 138 – 144.
- [21] 杨秀, 陈斌超, 朱兰, 等. 基于相关性分析和长短期记忆网络分位数回归的短期公共楼宇负荷概率密度预测[J]. 电网技术, 2019, 43(9): 3061 – 3071.
- YANG Xiu, CHEN Binchao, ZHU Lan, et al. Short-term public building load probability density prediction based on correlation analysis and long- and short-term memory network quantile regression [J]. Power System Technology, 2019, 43(9): 3061 – 3071.
- [22] 李训嘉. 数据驱动的航天器遥测数据异常检测方法研究[D]. 长沙: 国防科技大学, 2019.
- [23] LIU Fei, TING Kaiming, ZHOU Zhihua. Isolation-based anomaly detection [J]. ACM Transactions on Knowledge Discovery from Data, 2012, 6(1): 1 – 39.
- [24] 王阳, 马伟东, 刘泊溟, 等. 基于关联规则与重构误差的二次系统故障检测方法[J]. 中国电力, 2024, 57(8): 159 – 167.
- WANG Yang, MA Weidong, LIU Jiming, et al. Secondary system fault detection method based on association rules and reconstruction error [J]. Electric Power, 2024, 57(8): 159 – 167.
- [25] 林建熙, 刘必成, 肖晃庆, 等. 大规模海上风电接入的受端电网频率特性[J]. 广东电力, 2023, 36(3): 23 – 31.
- LIN Jianxi, LIU Bicheng, XIAO Huangqing, et al. Frequency characteristics of the receiving grid for large-scale offshore wind power integration [J]. Guangdong Electric Power, 2023, 36(3): 23 – 31.
- [26] 郭丽娟, 张玉波, 尹立群, 等. 基于离群点检测的变电主设备异常辨识与规律分析[J]. 南方电网技术, 2018, 12(9): 14 – 21.
- GUO Lijuan, ZHANG Yubo, YIN Liqun, et al. Identification and analysis of main substation equipment abnormal data based on outlier detection method [J]. Southern Power System Technology, 2018, 12(9): 14 – 21.

收稿日期: 2023-06-08; 网络首发日期: 2024-04-20

作者简介:

刘锦宁(1979), 男, 高级工程师, 硕士, 研究方向为智能用电和用电异常检测, liujn@csg.cn;

吴裕宙(1983), 男, 高级工程师, 硕士, 研究方向为电力营销数据分析与高级应用, aeou11983@163.com;

苏盛(1975), 男, 通信作者, 教授, 博士, 研究方向为配用电大数据应用、电力气象灾害和电力系统网络安全防护, eessheng@163.com。