

基于时间序列的分布式光伏电站发电数据采集分析

黄金华

(中交光伏科技有限公司, 上海 200125)

摘要: 由于社会发展水平的显著提升, 我国各个行业的发展速度也在不断加快, 其中, 分布式光伏电站的发电数据采集技术也得到了迅速的革新。但是, 因为当前阶段所采用的发电数据采集技术, 并没有及时修复发电数值的缺失值, 因而造成了多种问题, 所测得的数据精准度也亟待完善。基于此, 本文将针对基于时间序列的分布式光伏电站发电数据采集方法, 进行详细的分析和阐述, 以期对于发电站数据剔除与优化及相关发电量曲线的分析提供帮助, 推动行业进步, 为相关领域深度研究提供借鉴。

关键词: 时间序列; 分布式光伏电站; 发电数据采集

中图分类号: TM615

文献标识码: A

DOI: 10.12230/j.issn.2095-6657.2022.27.053

当前分布式光伏电站的建设与应用是十分广泛的, 原因在于分布式光伏电站本身具备安装便捷、成本投资少、效用发挥速度快的特点。但是, 在实际应用中, 相关部门需要重点关注的就是分布式光伏电站还存在着信息数据采集功能不完善的问题, 一旦出现数据采集量过大的情况, 就极易导致相应故障问题, 最终, 就会使已经采集的数据缺乏准确性与参考性。

1 时间序列的分析

时间序列指把同一统计指标的数值, 按照发生时间, 进行排列的数据。它的作用就在于可使用已有的历史数据, 进行未来数据的预测, 其中, “时间”可以是任意一种时间形式, 包括年、月、日等。在具体应用的过程中, 相关部门以及工作人员还需要重点关注其构成要素。时间序列的构成要素包含四点: 首先就是不规则变动, 这种变动毫无规律可循, 其中包含不规则突发性影响很大的变动以及严格的随机变动两种; 其次是循环变动, 整体现象的呈现是以“若干年”为一个周期, 所表达的是在这若干年中出现的、有规律的波浪起伏的变动; 再次是季节变动, 在一年时间内, 因季节变化而产生的周期性变化; 最后是长期趋势, 相应的现象一直受到根本性因素的影响而产生的总体变动趋势^[1]。

2 发电站异常数据剔除

据相关调查显示, 分布式发电站经常因为受到多种因素的影响, 而产生相应的故障问题, 这时, 就会有异常数据产生。那么, 相关工作人员需要重点关注的就是如何将这“异常数据”剔除, 而在负荷异常值中, 进行时间序列模型构建的公式为:

$$y(t) = x(t) + \sum_{i=1}^n \omega_i \alpha(j) I_t^{(T)} \quad (1)$$

其中, $y(t)$ 代表复合序列模型; $x(t)$ 代表无异常数据复合序列; ω_i 代表其中存在的影响因子; n 代表异常值数量; $\alpha(j)$ 代表异常数据复合序列; $I_t^{(T)}$ 代表特征量值^[2]。

基于发电站的异常数据, 需要计算目前时刻负荷均值、相同时间点负荷均值, 除此之外, 还需要进行时间序列至中心负荷密度区域的确定, 点与点之间的距离确定为 d , 密度半径设置为 r , 而 s 代表的则是到达中心点的距离, 最后是以阈值的设定结果来进行异常负荷点的获取, 具体的表达式为:

$$\delta = \frac{\sum_{i=1}^m \frac{d}{r_d}}{|d_s|} \quad (2)$$

其中, δ 代表异常负荷点; r_d 代表可达距离; d_s 代表全部距离之间均值; m 代表的则是常数。

上述计算完成后, 就可以进行所获取异常数据的剔除了, 以时间序列模型作用的发挥来剔除, 具体的公式为:

$$x_j = x(t) \left(1 + \frac{1}{m} \sum_{i=1}^m \frac{x(t-j)}{x'(t-j)} \right) \quad (3)$$

其中, $x(t)$ 代表不存在异常负荷; $x(t-j)$ 代表第 j 个负荷数据。在上述计算完成后, 异常数据的负荷点剔除工作完成。这时就能够发现电站数据中的缺失数据, 后续需要进行的工作就是对上述缺失数据的修复、插补^[3]。

3 关于周发电量的趋势曲线分析以及所需修复值的计算

3.1 关于周发电量的趋势曲线分析

分布式光伏电站中的发电量曲线, 其波动规律基本相同, 将周发电量曲线设置为 C_t , 整体关系的公式为:

$$C_t = [e_{t1}, e_{t2}, e_{t3} \dots e_{t7}] \quad (4)$$

其中, e_{t1} 代表 C_t 中的第一个元素, 以此类推。基于此, 为了及时有效地进行的获取, 需要将周发电量曲线系数和集合 S 进行有机结合, 并用 S_b 来进行二者结合后, 表达 (S_b (集合)) > 阈值 P_{set} (集合)。在计算时, 将 S_b 中第 i 个分布式光伏归一曲线进行设定, 表示为 C_{bi} , 计算其数值的公式为:

$$C_{bi} = [e_{i1}, e_{i2}, e_{i3} \dots e_{i7}] \quad (5)$$

公式中各部分所代表含义与上一个公式基本一致, 不同的是, 在 C_{bi} 之中, e_{i1} 可以表达为:

$$e_{i1} = \frac{E_{i1}}{E_{iB}} \quad (6)$$

在该公式中, E_{iB} 代表 S_b 中第 i 个周发电量的基准数值, E_{i1} 代表的则是周发电量的趋势曲线^[4]。其中, 需要重点关注的是 E_{iB} 的计算方式, 其计算公式为:

$$E_{iB} = \max(E_{i1}, E_{i2}, E_{i3} \dots E_{i7}) \quad (7)$$

其中, E_{iB} 中的第一个元素表达为 E_{i1} , 以此类推。

随后, 相关工作人员经过对于 C_{bi} 的分析显示, 其与周发电量曲线之间的波动规律是相同的, 但是, 二者之间的幅值却有着明显的不同。由此得出, 二者之间关联程度较高, 除此之外, 需要关注的就是 S_b 的周发电量归一化曲线与上述两者有一定的联系。依照如下表达式得出 C_t 中的元素值。

$$e_{t1} = \frac{1}{n} \sum_{i=1}^n e_{i1} \quad (8)$$

其中, n 代表光伏数量。公式计算完毕后, 所获得的结果和 C_{bi} 公式意义是一致的。由此看出, C_t 和 C_i 、 C_{bi} 之间都有较强的联系。因此, 关于相应数据插补、修复工作的开展, 可以按照 C_t 的变化趋势, 来进行实操。

3.2 关于所需修复值的计算

需要重点关注的是, 第 i 个发电量的缺失, 会造成 C_i 有空缺值产生, 而在进行所需修复值计算的过程中, 工作人员可以以 M 表示空缺值 C_i 对应序号的集合, 再加上 C_t 、 C_i 之间波动性相同、关联性较强, 所以, 计算时也需要将二者的关系进行表达, 公式为:

$$\frac{E_{ij}}{\sum_{k=1}^7 E_{ik}} = \frac{e_{tj}}{\sum_{k=1}^7 e_{tk}} \quad (9)$$

其中 k 是不属于 M 的, 一旦 E_{ij} (第 j 天分布式光伏发电量) 产生缺失情况, 就可以进行其修复值 R_{ij} 的计算, 公式为:

$$R_{ij} = \sum_{k=1}^7 E_{ik} \frac{e_{tj}}{\sum_{k=1}^7 e_{tk}} \quad (10)$$

随后还需要进行修复数据的核验, 由于篇幅原因, 在这里不过多赘述。

4 利用 Back Propagation 神经网络进行的发电数据采集分析

完成上述基本流程后, 需要相关工作人员利用 Back

Propagation 神经网络, 采集发电数据。Back Propagation 神经网络, 简称 BP 神经网络, 指依照误差逆向传播算法, 进行训练的多层前馈神经网络, 在相关行业中的应用十分广泛, Back Propagation 神经网络是在输入、输出层之间, 增加一层或者多层的神经元, 这些神经元又被称为“隐单元”, 这些单元与外界无直接联系, 但是, 如果这些隐单元状态发生改变, 输入与输出之间的关系将会受到相应的影响, 同时需要注意的是, 在每一层之间, 都会有多个节点的存在^[5]。整体计算过程由反向计算过程、正向计算过程两部分构成, 前者是把误差信号沿原连接通路, 进行返回, 以修改相应神经元权值, 达成误差信号最小的目的; 后者是自输入层, 经过隐藏单元层, 并逐级处理, 然后转向输出层, 仅有上一层神经元影响下一层神经元状态的存在。需要注意的是, Back Propagation 神经网络有三个层次。接下来, 笔者将针对其训练过程, 进行详细阐述、分析与探究。

训练过程中, 将输入量设置为 x_i , 并将隐含层神经元设定为 y_i , 其表达式为:

$$y_i = \sum_{j=1}^n w_{ji} x_j - \theta_j \quad (11)$$

其中, 第 i 个连接权值由 w_{ji} 表示, 阈值则由 θ_j 来表示。而输出层神经元表达式与以最小二乘法进行误差检定的表达式如下:

$$y_k = \sum_{j=1}^n w_{kj} b_j - \theta_k \quad (12)$$

$$E = \frac{1}{2} \sum_{j=1}^k (y - y_j)^2 \quad (13)$$

在上述两个公式中, 输出层连接权值—— w_{kj} ; 因子—— b_j ; 阈值—— θ_k ; 而在后一个公式中, 误差——E; 固有系数—— $\frac{1}{2}$; 期望输出值——y; 第 j 个期望输出值—— y_j 。

4.1 基于时间序列的分布式光伏电站发电数据采集模型构建

随后就是基于时间序列的分布式光伏电站发电数据采集模型的构建, 采用上述 Back Propagation 神经网络, 进行数据采集模型构建实操。在实际构建过程中, 相关工作人员需要对 BP 神经网络每一层的神经元数目, 进行有效确定。这种确定, 主要有两层的划分。

(1) 隐含层的节点选取, 在隐含层中, 对于其节点的选取是与后者的节点选取息息相关的。简单来说, 其数量庞大, 网络承载压力就会更大, 模型的运行效率就难以得到有效保证, 而如果出现数量过少现象, 所采集的数据就难以满足相应需求^[6]。基于此, 在隐含层节点选取中, 需要依照下列四种方式来进行, 表达式如下:

$$m = 2n + 1 \quad (14)$$

$$m = \log_2 n \quad (15)$$

$$m = \sqrt{n+l} + \alpha \quad (16)$$

$$m = \sqrt{nl} \quad (17)$$

在四种表达式中, 隐含层节点数量—— m ; 输入层、输出层节点数量—— n 、 l ; 常数—— α 。

(2) 输出、输入层的节点选取。输出为待采集日时间段的输出功率, 表达式为:

$$Y = (e_1, e_2, e_3 \dots e_{14}) \quad (18)$$

其中, 输出节点—— e 。

而 BP 神经网络的输入, 由分布式光伏电站辐照度、带采集辐照度代表, 输入层节点的公式为:

$$D(t) = (d_1, d_2, d_3 \dots d_{14}) \quad (19)$$

该公式中, 相似日—— t ; 输入节点—— d 。

4.2 基于时间序列分布式光伏电站发电数据采集结果的相关分析

根据上述模型构建, 进行数据采集结果分析, 公式为:

$$P_i = \frac{P_n}{P_D} \times P_f \quad (20)$$

其中, 待采集数据—— P_i ; 电站容量—— P_n ; 采集装置电站 D 装机容量—— P_D ; 已采集数据—— P_f 。

5 结语

综上所述, 本文对基于时间序列的分布式光伏电站发电数据采集进行了三点阐述, 分别是缺失数据的修复、采集模型的构建以及数据采集结果的分析。这种数据采集方式, 能够有效保障数据采集结果的精准度。同时, 本文从时间序列的分析、

发电站异常数据剔除、关于周发电量的趋势曲线分析以及所需修复值的计算, 还有就是利用 Back Propagation 神经网络进行的发电数据采集分析四个角度出发, 具体分析了整个信息数据采集的流程, 希望能够为我国相关领域工作的开展, 提供借鉴。

参考文献:

- [1] 王婧骅, 张娟, 赵婉茹, 等. 基于时间序列的分布式光伏电站发电数据采集方法 [J]. 电网与清洁能源, 2022, 38 (06): 137-142.
- [2] 张锦元, 贾松达, 董亚旭, 等. ZigBee 在分布式光伏电站监控系统中的应用研究 [J]. 光学与光电技术, 2022, 20 (01): 97-101, 114.
- [3] 吴保华, 翟志成, 韩诗地, 等. 分布式光伏电站监控平台的实现原理及开发技术分析 [J]. 电子世界, 2021, (02): 33-34.
- [4] 张凌浩, 张明, 嵇文路, 等. 基于灰色关联理论和 BP 神经网络的分布式光伏电站运维数据虚拟采集方法 [J]. 电力建设, 2021, 42 (01): 125-131.
- [5] 张冲标. 小容量分布式光伏电站信息采集系统 [J]. 农村电工, 2020, 28 (10): 35-36.
- [6] 马大燕, 谢祥颖, 那峙雄, 等. 基于调度自动化系统的低压分布式光伏电站接入估算模型 [J]. 电信科学, 2020, 36 (02): 90-94.

作者简介: 黄金华 (1986-), 男, 上海人, 大学本科, 工程师, 主要从事新能源光伏研究。