

基于离群点分析的分布式光伏发电量异常检测算法

王立斌¹, 赵佩¹, 王杉杉²

(1. 国网河北省电力有限公司电力科学研究院, 河北 石家庄 050021;
2. 国网石家庄供电公司, 河北 石家庄 050051)

摘 要:在实际运行中,部分分布式光伏存在单位容量日发电量明显偏低,即带病运行的现象,另有部分分布式光伏单位容量日发电量明显过高,存在骗取国家补贴的嫌疑。针对分布式光伏带病运行及骗取国家补贴的现象,基于离群点分析中的格拉布斯准则和分布式光伏的特点,提出了一种分布式光伏发电量异常检测算法。该算法可以筛选出单位容量日发电量过高与过低的分布式光伏用户,以供相关工作人员进行现场核实,进而避免分布式光伏长期带病运行,以及遏制骗取国家补贴的行为,从而保证分布式光伏高效规范运行。最后以某县分布式光伏数据进行了案例分析,验证了该算法的有效性。

关键词:分布式光伏;离群点分析;格拉布斯准则;发电量;异常检测

文章编号:2096-4633(2018)02-0013-05 **中图分类号:**TM7 **文献标志码:**B

随着社会经济的发展,能源短缺以及环境污染等问题日益突出,开发与利用可再生能源已经成为发展的必然趋势^[1-3]。在可再生能源发电中,光伏发电具有安全可靠、地域限制少,建设周期短等优势,目前已经有较大的产业规模。且近几年随着光伏组件成本的日益降低,分布式光伏发展迅速。另外,国家还制定了相应的补贴政策^[4-5],以支持光伏发电的发展。

在实际工作中通过用电信息采集系统可以发现有部分分布式光伏的单位容量日发电量明显过低,可能是因分布式光伏带病运行(即部分元器件故障老化、太阳能电池板安装角度不合适以及太阳能电池板积尘过厚等)引起;还有部分分布式光伏的单位容量日发电量明显过高,存在骗取国家补贴的嫌疑。因此,有必要对分布式光伏的发电量进行检测,以发现发电量异常的分布式光伏,进而保证分布式光伏合规高效的运行。

文章基于离群点分析提出了一种分布式光伏发电量异常检测算法,并根据用电信息采集系统中的数据对算法的有效性进行了验证。

1 分布式光伏发电系统

1.1 分布式光伏原理结构

分布式光伏发电系统一般布置在居民的屋顶,

无需额外的安装空间,其结构如图 1 所示^[6],

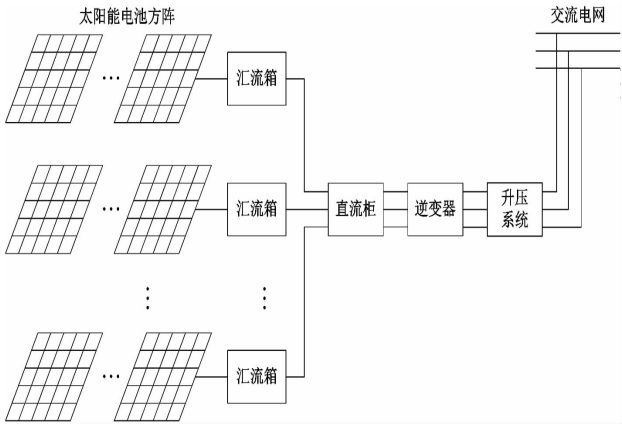


图 1 分布式光伏发电系统结构示意图
Fig. 1 The structural of distributed photovoltaic system

太阳能电池方阵将太阳能转换为电能,再由汇流箱汇集到直流柜内。然后由逆变器将直流电转换为交流电,经升压系统升压后注入交流电网。

1.2 分布式光伏带病运行

若太阳能电池板的安装倾角不当、积尘过厚或被阴影遮挡,逆变器效率低下以及部分元器件损坏等,均会导致分布式光伏发电系统的发电量明显降低,本文将上述现象称为分布式光伏发电

系统带病运行。

1.3 分布式光伏用户骗补原理

在实际当中,有部分不法分子在供电公司的配电线路上挂接单相电力变压器、单相整流桥、逆变器等设备,在电网取电后返送到光伏发电计量表计,使表计多计量发电量,虚发电量以骗取国家补贴,其原理图如图 2 所示。

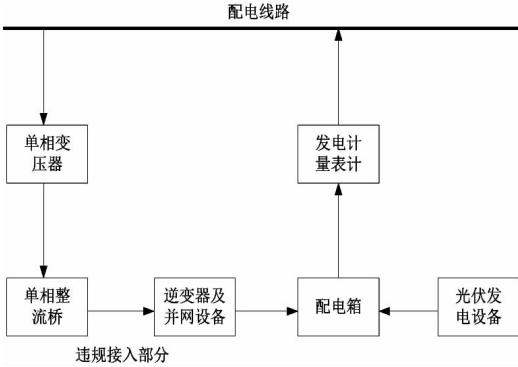


图 2 分布式光伏用户骗取补贴原理图

Fig. 2 Schematic diagram of distributed photovoltaic users defrauding subsidies

2 离群点分析

离群点是指一个数据点偏离其他数据点非常之大,从而导致怀疑该数据点是由特殊原因产生的。离群点分析的目的在于找出这些令人怀疑的数据点^[7]。

当前,离群点分析已经成功应用在信用卡异常数据检测、结构缺陷检测、时间序列监视、网络性能诊断、贷款受理、行为监控、新颖文本检测等领域中^[8-10]。

经典的离群点分析算法主要有:基于统计的离群点分析、基于距离的离群点分析、基于密度的离群点分析、基于深度的离群点分析以及基于聚类的离群点分析。其中基于统计的离群点分析适合处理数值型单维数据集^[11],其他几种离群点分析适合处理多维数据集。

格拉布斯准则是一种常见的基于统计的离群点分析方法,其具有较强的异常数据筛选能力^[12-14]。

3 发电量异常检测算法

在某一地理区域内(如一个县或一个镇),同一天的气象条件是相同的,所以该地理区域内所有分

布式光伏的单位容量日发电量相差不大。因此,对某一地理区域内所有分布式光伏的单位容量日发电量进行离群点分析,即可筛选出发电量异常的分布式光伏。

由于分布式光伏的单位容量日发电量为单维数据,选用基于统计的离群点分析中的格拉布斯准则对分布式光伏的单位容量日发电量进行分析。

分布式光伏的单位容量日发电量 e 的计算方法如(1)式所示,

$$e = \frac{E}{S} \quad (1)$$

式中, E 为分布式光伏的日发电量, S 为分布式光伏的装机容量。

某一地理区域内所有分布式光伏的单位容量日发电量的平均值 $\bar{e}$ 以及标准差 s 的计算方法分别如(2)式与(3)式所示,

$$\bar{e} = \frac{1}{n} \sum_{i=1}^n e_i \quad (2)$$

式中, n 为该地理区域内分布式光伏的数量, e_i 为第 i 个分布式光伏的单位容量日发电量。

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (e_i - \bar{e})^2} \quad (3)$$

式中, $n, e_i, \bar{e}$ 定义同(2)式。

找出 e_k , 使得满足(4)式,

$$|e_k - \bar{e}| = \max_{1 \leq i \leq n} |e_i - \bar{e}| \quad (4)$$

(4)式中, k 的取值范围为 $[1, n]$ 。

若 e_k 满足(5)式,则判定 e_k 为异常数据,将其从数据集中剔除。

$$|e_k - \bar{e}| \geq g_0 s \quad (5)$$

式中, s 为标准差, g_0 为临界值,其可以根据显著水平 α 与数据量 n 查表得出。 α 越小,则对数据要求越严格,非异常数据的范围越窄^[15], α 一般取 0.01 或 0.05。表 1 给出了 n 在 $[15, 25]$ 范围内时对应的 g_0 值。

若 e_k 满足(6)式,则该分布式光伏有可能处于带病运行状态。

$$e_k < \bar{e} \quad (6)$$

若 e_k 满足(7)式,则该分布式光伏存在骗取国家补贴的嫌疑。

$$e_k > \bar{e} \quad (7)$$

对剔除 e_k 后的新数据集重复用(2) - (7)式进

行计算分析,直至数据集中无异常数据为止。

表 1 格拉布斯临界值		
Tab. 1 Critical value of Grubbs		
	0. 05	0. 01
15	2. 409	2. 705
16	2. 443	2. 747
17	2. 475	2. 785
18	2. 504	2. 821
19	2. 532	2. 854
20	2. 557	2. 884
21	2. 580	2. 912
22	2. 603	2. 939
23	2. 624	2. 963
24	2. 644	2. 987
25	2. 663	3. 009

4 算例分析

4.1 算例

从用电信息采集系统中选择某县全部分布式光伏于 2017 年 3 月 12 日的发电量进行算例分析^[16],待分析分布式光伏的相关数据如表 2 所示,

利用第 3 节的算法对表 2 中数据进行分析,显著水平 α 取 0. 05,得出发电量异常的分布式光伏如表 3 所示,

由表 3 可知,用户 C2、C11 的单位容量日发电量明显低于其他用户,其有可能处于带病运行状态^[17];C25 的单位容量日发电量则明显高于其他用户,其存在骗取国家补贴的嫌疑。

4.2 算法有效性验证

用户 C1 ~ C25 的单位容量日发电量的柱状图如图 3 所示,

由图 3 可知,用户 C2、C11 的单位容量日发电量明显低于其他用户,而 C25 的单位容量日发电量则明显高于其他用户。

由此可见,利用该检测算法所筛选出的发电量异常的分布式光伏用户与由图 3 分析出的发电量异常的分布式光伏用户一致,说明该算法有效可行。

表 2 分布式光伏相关数据			
Tab. 2 The data of distributed photovoltaic systems			
用户	装机容量 /kW	发电量 /kW · h	单位容量日发电量 /(kW · h/kW)
C1	9	52. 02	5. 78
C2	10	39. 22	3. 92
C3	10	58. 24	5. 82
C4	40	215. 05	5. 38
C5	30	181. 8	6. 06
C6	5	30. 5	6. 10
C7	20	117. 97	5. 90
C8	15	87. 6	5. 84
C9	5	31. 14	6. 23
C10	3	17. 23	5. 74
C11	15	57. 96	3. 86
C12	15	85. 29	5. 69
C13	5	32. 72	6. 54
C14	10	62. 67	6. 27
C15	10	62. 95	6. 30
C16	5	27. 98	5. 60
C17	10	62. 41	6. 24
C18	11	58. 35	5. 30
C19	12	72. 26	6. 02
C20	12	73. 81	6. 15
C21	7	37. 96	5. 42
C22	18	114. 28	6. 35
C23	28	162. 9	5. 82
C24	20	132. 14	6. 61
C25	6	48. 4	8. 07

表 3 发电量异常的分布式光伏	
Tab. 3 Distributed photovoltaic systems with abnormal power generation	
用户	单位容量日发电量/(kW · h/kW)
C2	3. 92
C11	3. 86
C25	8. 07

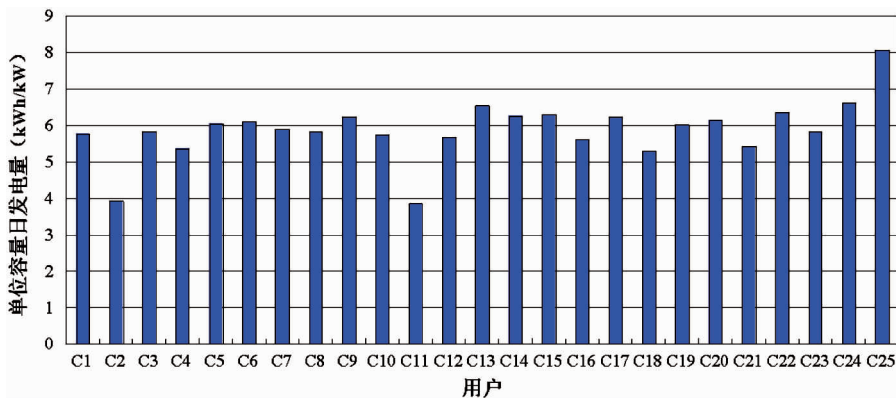


图 3 分布式光伏单位容量日发电量柱状图

Fig. 3 The histogram of distributed photovoltaic systems generation of the unit capacity per day

5 结束语

本文基于离群点分析中的格拉布斯准则以及分布式光伏的特点,提出了一种分布式光伏发电量异常的检测算法。该算法可以筛选出单位容量日发电量过高与过低的分布式光伏用户,能遏制骗取国家补贴的行为,还能避免分布式光伏长期带病运行。以用电信息采集系统中某县的分布式光伏数据进行了算例分析,验证了算法的有效性。

参考文献:

[1] 常健,陈峦,董立淑,等. 光伏并网有功功率平滑措施研究[J]. 贵州电力技术,2017,20(3):17-22.
CHANG Jian, CHEN Luan, DONG Lishu, et al. Research on smoothing measures of photovoltaic grid-connected active power [J]. Guizhou Electric Power Technology, 2017, 20(3): 17-22.

[2] 张程熠,唐雅杰,李永杰,等. 适用于小样本的神经网络光伏预测方法[J]. 电力自动化设备,2017,37(1):101-106.
ZHANG Chengyi, TANG Yajie, LI Yongjie, et al. Photovoltaic power forecast based on neural network with a small number of samples[J]. Electric Power Automation Equipment, 2017, 37(1): 101-106.

[3] 韩富佳,王淳. 基于 Matlab 的分布式光伏并网发电系统对配电网电能质量的影响[J]. 电测与仪表,2015,52(14):16-21.
HAN Fujia, WANG Chun. Influences of the distributed photovoltaic grid-connected generation system on the distribution network power quality based on Matlab[J]. Electrical Measurement & Instrumentation, 2015, 52(14): 16-21.

[4] 林焱. 分布式光伏电站引起功率因数下降的分析[J]. 贵州电力技术,2017,20(2):43-48.
LIN Yan. Investigation of power factor drop events caused by distributed photovoltaic power plant [J]. Guizhou Electric Power

Technology, 2017, 20(2): 43-48.

[5] 林小进,吴蓓蓓,李红涛,等. 分布式光伏发电系统现场检测技术研究[J]. 电测与仪表,2014,51(23):79-82.
LIN Xiaojin, WU Beibei, LI Hongtao, et al. On-site testing technology for distributed photovoltaic power plant [J]. Electrical Measurement & Instrumentation, 2014, 51(23): 79-82.

[6] 沈金荣,惠杰,董炜. 基于地理位置相关性的光伏故障监控方法研究[J]. 可再生能源,2016,12(12):1754-1761.
SHEN Jinrong, HUI Jie, DONG Wei. Research on the method of photovoltaic fault monitoring based on the geographical position correlation[J]. Renewable Energy Resources, 2016, 12(12): 1754-1761.

[7] 程超,张汉敬,景志敏,等. 基于离群点算法和用电信息采集系统的反窃电研究[J]. 电力系统保护与控制,2015,43(17):69-74.
CHENG Chao, ZHANG Hanjing, JING Zhimin, et al. Study on the anti-electricity stealing based on outlier algorithm and the electricity information acquisition system [J]. Power System Protection and Control, 2015, 43(17): 69-74.

[8] 王震. 基于距离的离群点检测算法分析与研究[D]. 重庆:重庆大学,2011.

[9] 胡婷婷. 数据挖掘中的离群点检测算法研究[D]. 厦门:厦门大学,2014.

[10] 黄洪宇,林甲祥,陈崇成,等. 离群数据挖掘综述[J]. 计算机应用研究,2006,(8):8-13.
HUANG Hongyu, LIN Jiaxiang, CHEN Chongcheng, et al. Review of outlier detection [J]. Application Research of Computers, 2006, (8): 8-13.

[11] 周英,卓金武,卞月青. 大数据挖掘:系统方法与实例分析[M]. 北京:机械工业出版社. 2016.

[12] 侯瑞东,夏宏森,李可心. 基于 Grubbs 的雷达靶场试验数据处理[J]. 现代雷达,2009,31(3):11-13.
HOU Ruidong, XIA Hongsen, LI Kexin. Grey processing of radar range test data based on Grubbs [J]. Modern Rada, 2009, 31(3): 11-13.

[13] 赵庆权. 高速公路补充预算定额原始数据异常值的改进格拉布斯法应用研究[J]. 价值工程,2013(12):82-83.
ZHAO Qingquan. Application of improved Grubbs Law in raw data outliers of highway supplementary budget quota[J]. Value Engineering,2013(12):82-83.

[14] 齐敏芳. 大数据技术及其在电站机组分析中的应用[D]. 北京:华北电力大学,2016.

[15] 林丽芬,肖化,吴先球. 肖维勒准则和格拉布斯准则的比较[J]. 大学物理实验,2012,25(6):86-88.
LIN Lifen, XIAO Hua, WU Xianqiu. Comparison between Chauvenet's Criterion and Grubbs Criterion[J]. Physical Experiment of College,2012,25(6):86-88.

[16] 程向文,詹雄铿,李钙. 不同运营模式下光伏发电经济效益[J]. 广东电力,2017,30(8):63-67.
CHENG Xiangwen,ZHANG Xiongkeng,LI Gai. Economic benefit

of photovoltaic generation under different operation patterns[J]. Guangdong Electric Power,2017,30(8):63-67.

[17] 曾江,凌毓畅. 光伏电站接入电网的电能质量分析及其应用[J]. 广东电力,2017,30(8):38-46.
ZENG Jiang, LING Yuchang. Analysis and application of power quality of grid-connected photovoltaic power station[J]. Guangdong Electric Power,2017,30(8):38-46.

收稿日期:2017-11-22

作者简介:



王立斌(1985),男,硕士,高级工程师,主要从事电能计量与用电信息采集工作。

(本文责任编辑:王燕)

Anomaly detection algorithm for distributed photovoltaic generation
based on outlier analysis

WANG Libin¹, ZHAO Pei¹, WANG Shanshan²

(1. State Grid Hebei Electric Power Research Institute, Shijiazhuang 050021 Hebei, China;
2. State Grid Shijiazhuang Power Supply Company, Shijiazhuang 050051 Hebei, China)

Abstract:In practice, the power generation of the unit capacity per day of some distributed photovoltaic is too low, that is the phenomenon of the sick running. And the power generation of the unit capacity per day of some distributed photovoltaic is too high, that is suspected of fraudulent state subsidies. According to distributed photovoltaic system sick running and defraud state subsidies behavior, an anomaly detection algorithm for distributed photovoltaic generation is proposed, based on the outlier analysis of Grubbs criterion and the characteristics of the distributed photovoltaic system. The algorithm can filter out the users, whose power generation of the unit capacity per day is too high or too low, so that distributed photovoltaic system long-term sick running can be avoided and defraud state subsidies behavior can be curbed. It ensures the efficient and legitimate operation of distributed photovoltaic. Finally analyzed the data of the distributed photovoltaic systems in a county, and the result shows the effectiveness of the algorithm.

Key words:distributed photovoltaic system; outlier analysis; Grubbs criterion; power generation; anomaly detection