

DOI: 10.13382/j.jemi.B1902804

基于孤立森林算法的电涌保护器异常参数监测研究

马俊彦^{1,2} 杨仲江^{1,2} 王昊^{1,2}

(1. 南京信息工程大学 中国气象局气溶胶与云降水重点开放实验室 南京 210044;

2. 南京信息工程大学 大气物理学院 南京 210044)

摘要:应用孤立森林算法构建电涌保护器(SPD)的异常参数监测模型,通过令模型学习正常SPD参数数据的分布,来对无法拟合至该分布内的SPD异常参数数据进行识别,进而对器件实际状态及劣化程度进行评价预警。研究先对SPD进行多参数采样,将得到的390组实验数据进行预处理,结合算法模型计算每组采样数据对应的异常值,来识别对应异常的采样数据,最后根据每组数据对应的标签,对算法的准确度进行验证。检验结果表明,基于孤立森林算法构建的监测模型可以对SPD采样数据的异常程度进行评价,锁定SPD在系统中可能出现不良或劣化状态时所对应的参数值。对于实验数据集,该算法模型在选择不同参数下的性能指标AUC值不低于96%,为SPD性能的监测研究提供了新的思路。

关键词:电涌保护器;多参数采样;异常值;孤立森林算法

中图分类号: TM862; TN303 **文献标识码:** A **国家标准学科分类代码:** 510.3030

Research on abnormal parameter monitoring of surge protective devices based on isolation forest algorithm

Ma Junyan^{1,2} Yang Zhongjiang^{1,2} Wang hao^{1,2}

(1. Key Laboratory for Aerosol-Cloud-Precipitation of China Meteorological Administration, Nanjing University of Information Science & Technology, Nanjing 210044, China; 2. School of Atmospheric Physics, Nanjing University of Information Science and Technology, Nanjing 210044, China)

Abstract: In this paper, isolation forest algorithm is used to construct an abnormal parameter monitoring model for surge protectors. By making the model learn the distribution of normal SPD parameter data, the abnormal parameter data of SPD that cannot fit into the distribution will be identified, then the actual state and deterioration degree of the device can be evaluated and warned. The study first performs multi-parameter sampling on the SPD, pre-processes the obtained 390 sets of experimental data, and combines the algorithm model to calculate the abnormal value corresponding to each set of sampling data to identify the corresponding abnormal sampling data. Finally, according to the label corresponding to each set of data, Verify the accuracy of the algorithm. The test results show that the monitoring model based on the isolation forest algorithm can evaluate the abnormality of the SPD sampling data, and lock the parameter values corresponding to the SPD that may appear bad or degraded in the system. For the experimental data set, the performance index AUC value of the algorithm model under different parameters is not less than 96%, which provides a new idea for the monitoring research of SPD performance.

Keywords: surge protective devices; multi-parameter sampling; outliers; isolation forest

0 引言

电涌保护器(SPD)是为设备提供雷电防护能力的必

备装置^[1],被广泛的应用于各类电气系统中。SPD在运行时除遭受雷电过电压、操作过电压等威胁外,还受到温湿、化学污染等外部环境因素的影响,导致其电气特性和物理状态产生衰减^[2]。这种衰减使发生SPD持续性劣

化,为了能够对系统中各 SPD 的实际性能进行实时的监测评价,找出潜在物理状态不佳的器件,研究人员以此做出了许多的研究工作^[3-7]。

在 SPD 监测技术的研究中,杨仲江等^[8]通过研究压敏电阻的非线性系数 α ,发现随着 SPD 不断受到雷电流的冲击,其非线性系数 α 会随之规律的变化。因此提出了基于非线性系数 α 的 SPD 劣化监测评价方法,但避雷器本身的劣化状况需要通过多个相关参数进行综合评价,仅使用非线性系数 α 一个参数会对评价结果产生一定的误差。龚敏等^[9]研究了氧化锌阀片老化与雷电冲击幅值及冲击的次数间的定量关系,提出一种带电监测方法。徐志钮等^[10]提出一种组容性电流补偿法,利用阻容性电流的正交性提取阻性电流,以此监测 SPD 实际性能的衰减程度。但此类方法易受系统中谐波电压影响而产生较大误差。近年来,将数据与算法相结合作为新的研究方向,杨仲江^[11]提出了基于遗传算法的 MOV 在线监测,通过最优化的求解等效模型中反映器件劣化的参数,来对期间进行性能的监测,为后续在线监测技术的研究提供了新的方向,但该算法中需要人工调试的参数较多,各参数的选择组合对输出值产生的变化较大。

本文针对 SPD 传统监测方式的不足之处进行补充,通过数据分析的思路,提出一种简单高效具有可行性的方法,基于孤立森林(isolation forest, IForest)算法构建 SPD 的异常参数监测模型,在时间序列上对电气系统内各 SPD 的多项参数数据进行采样记录,根据相应的计算策略,对每组采样数据进行判定,得出是否具有异常性的判断结果。将结果与对应的原始采样数据和被采样的器件信息进行匹配。这样就可以对某个器件的某一次多参数采样数据是否开始具有异常的分布情况,进行监测识别。

1 监测方案设计

为验证本文提出的模型用于监测任务的准确性及可行性,设计采用对多个同型号的全新 SPD 进行多次冲击的冲击试验,记录每次冲击后的采样数据,以模拟 SPD 在实际电气系统中持续长期运行的情景。由于实验条件及设备的限制,仅记录 SPD 物理状态的原始参数,包括冲击前后的压敏电压、冲击前后的泄漏电流(全电流)、 $C_p(50\text{ Hz})$ 、 $C_p(10\text{ kHz})$ 、 $L_s(10\text{ kHz})$ 和 $R_s(10\text{ kHz})$ 。

方案中冲击试品采用 6 个同型号的全新 SPD,其最大持续运行电压 U_c 为 385 V,电压保护水平 U_p 为 1.5 kV,标称放电电流 I_n 为 10 kA,最大放电电流 I_{\max} 为 20 kA,将其编号为 A1~A6。实验器械选用北京某测试中心的多脉冲冲击平台,根据冲击实验规定,采用 5 个等幅值、时间间隔为 20 ms 的 8/20 μs 雷电脉冲对 6 个全新试品 SPD 各进行 65 次冲击,脉冲的幅值为 10 kA。每模

拟冲击一次,待器件完全冷却后对所测参数进行记录。将所有采样数据进行整理,最后通过算法模型计算每组数据对应的异常值,实现 SPD 异常参数的监测。

2 监测数据的分析与处理

2.1 数据分析

在实际操作中对编号为 A1~A6 的 SPD 进行了 65 次冲击试验,得到共计 390 组实验数据,将其统计作为数据集。对每次冲击试验前后测得的压敏电压和泄漏电流取均值,使得每一组数据包括 6 个参数特征。

数据集中各参数随冲击而产生的变化趋势如图 1 所示。由此可知,随着冲击试验的次数增加,所测 SPD 的压敏电压不断下降,器件的压敏电压最终都分布在阈值 575 V 处,而 A1 压敏电压的下降趋势呈现出一个更大的梯度。同样,对于 SPD 的泄漏电流(全电流)随着冲击试验次数的变化关系,可以观察到 A4 的泄漏电流出现明显的大梯度上升,A6 泄漏电流的梯度也呈现出陡峭的都先升高后下降的趋势。所测器件的泄漏电流在 10 μA 内分布较为密集,10 μA 以上的分布较为稀疏。另外,各测试器件的电容、电感随冲击次数的变化趋势都较为相似,数据分布都较为密集。但可以发现,在时间序列上出现了“峰值”起伏的电容或电感变化现象。

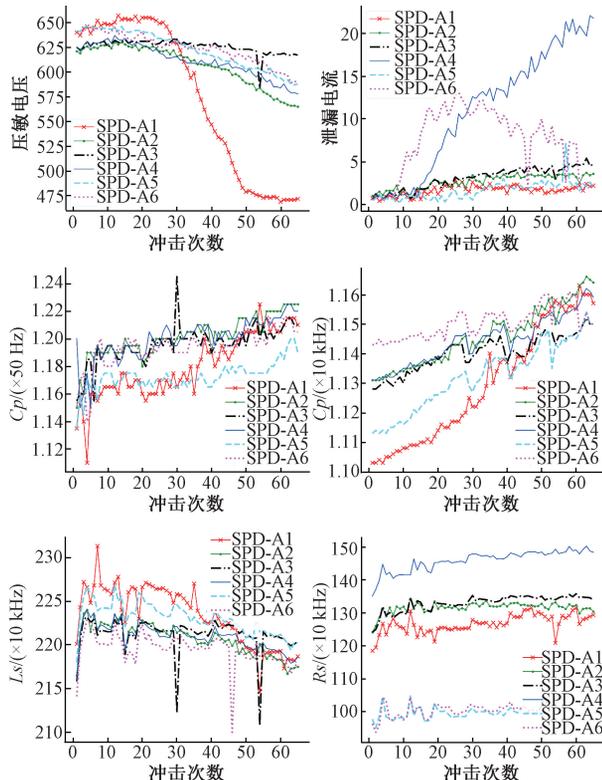


图1 实验数据特征的变化趋势

Fig. 1 Changes in experimental data characteristics

通过对比分析发现,对于同一个 SPD 器件,一些参数随时间的变化趋势与其他器件相似,而另一些参数的变化趋势则具有很大差异。目前还没有一个统一的方式,使得完全仅通过一个采样参数的变化,来对 SPD 实时的性能状态进行追踪判断。所以在对 SPD 的实际状态进行监测时,需要结合所有的测量参数,找到合理的统计方法,考虑每个参量的变化情况后,进行综合评价。

2.2 数据预处理

对数据集进行主成分分析(PCA),提取前两个成分的累积解释方差为 72%,前 3 个成分的累积解释方差为 85%,使用 2 个或 3 个主成分可以较好的描述实验数据。图 2、3 所示分别为提取主成分后,原始数据集分别降维至 2 维和 3 维数据空间上的可视化。

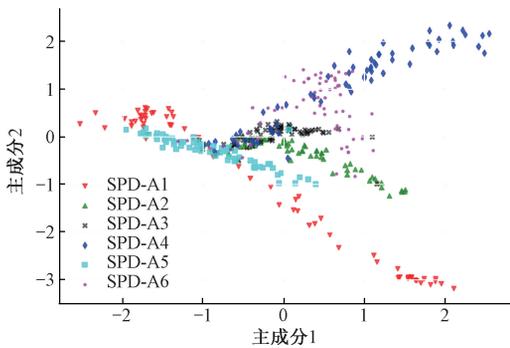


图 2 采样数据在二维中的分布

Fig. 2 Distribution of sampled data in 2-D space

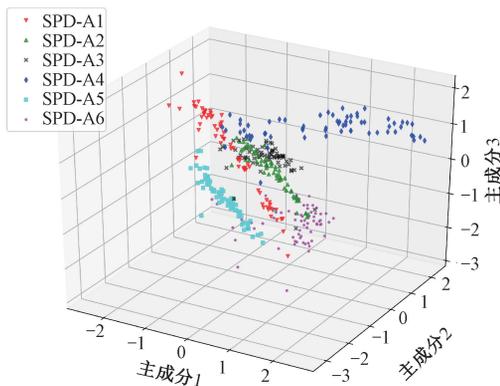


图 3 采样数据在三维中的分布

Fig. 3 Distribution of sampled data in 3-D space

通过 PCA 组合特征后,对数据集标准化处理,在图中表现为以坐标轴 0 点为中心分布。可以发现,大部分的数据点密集的围绕圆心分布,对应着图 1 随时序正常变化的 SPD 采样数据特征。而在空间中稀疏分布在外侧的数据点则对应着图 1 某个时刻开始异常变化的数据。这说明了正常数据在数据空间的分布是密集的,而异常数据在数据空间的分布是稀疏的,两者的分布之间具有明显的差异。

3 异常监测模型

3.1 IForest 算法

IForest 算法^[12]是一种基于树模型的异常值监测算法。它的作用在于将数据样本分布中的异常数据利用“孤立”的策略检测出来,能够有效的应用在电气安全领域之中。

在 IForest 中,异常数据被定义为易被“孤立”的离群点,具体表现为自身分布稀疏,且与密集分布的点的区域存在相应距离的点。对于图 2、3 降维后的 SPD 采样数据分布,可以看作性能良好的数据点具有较高的概率密集分布,而分稀疏分布的数据点则是低概率且异常的。因此 IForest 对异常点的决策如图 4 所示,通过递归的随机数次的分割数据集,直到所有的样本不能够再被分割下去。在这种方式下,“孤立”密集的簇需要分割多次,而那些分布稀疏的点只需要分割几次就能被“孤立”出来。

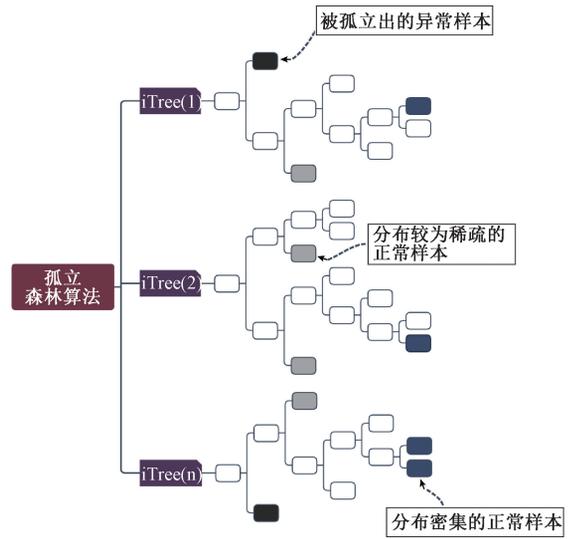


图 4 运用孤立森林进行异常监测

Fig. 4 Using isolation forest for anomaly monitoring

3.2 计算步骤

IForest 具有集成学习的思想,通过生成多个“树结构”的学习器,根据每个学习器的输出进行综合决策,从而得到精确度更高的最终结果。每个学习器 (ITree) 的构建基于完全随机生成的过程。

1) 初始构建 ITree 时,从 n 组采样数据中均匀抽样出 m 个样本作为子样本。

2) 在子样本中,随机选取一个特征维度,在这个特征的所有值范围内(最小值与最大值之间)随机选一个值 k 作为节点,对样本进行二叉划分,将样本中小于 k 值的划分至其左侧,大于等于 k 值的划分至右侧。

3) 分别在新产生的左右两边的数据子集上重复步骤 2) 的过程, 当数据不能继续分割或者分割次数到达 $\log_2 k$ 时, 停止分割操作。

在构建完成 t 个 ITree 后, 令样本数据 x 遍历每个 ITree, 并计算每次的路径长度 $h(x)$ 。 $h(x)$ 为从根节点至叶子节点, 走过的边的数量。最后, 将 $h(x)$ 代入式 (1), 计算每条待测数据的异常值, x 的异常值定义为:

$$S(x, n) = 2^{-\frac{E(h(x))}{c(n)}} \quad (1)$$

$S(x, n)$ 为每个采样数据样本所具有的异常值, $E(h(x))$ 为样本 x 在所有 ITree 中的具有的路径长度的期望。

$$c(n) = 2H(n-1) - \frac{2(n-1)}{n} \quad (2)$$

$c(n)$ 为给定样本数 n 时, 路径长度的平均值, 用于标准化样本的路径长度 $h(x)$ 。 $H(i)$ 为调和数, 该值可以被估计为 $\ln(i) + \gamma$, γ 为欧拉常数。

S 的取值范围在 0~1, 当 S 越接近 1 时, 数据被判断为异常的几率越高; 当 $S < 0.5$ 时, 数据可以判断为非异常的; 当 S 处于 0.5 附近时, 数据认为不具备明显的异常特征, 有待观察。

3.3 SPD 异常参数监测流程与思想

对 SPD 的异常参数进行监测的主要流程如图 5 所示, 包括对原始采集数据进行预处理, 删除冗余、错误记录, 对特征数量降维。之后基于预处理后的数据构建多个 ITree, 最后通过 IForest 确定每组采样数据的异常值, 找出高异常值对应的原始数据。

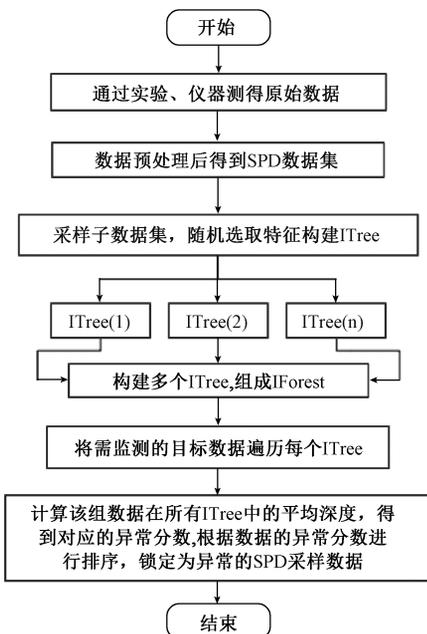


图 5 SPD 异常参数监测流程

Fig. 5 SPD abnormal parameter monitoring process

上述步骤具有可行性理由如下: 与 SPD 异常采样参数分布的多样性相对应的, 是正常采样参数服从固定分布的有限性。这里的有限性指, 正常的 SPD 采样参数特征, 都具有着一定规律的变化趋势。从统计学的角度, 可以假设正常的的数据都服从于一个有限的分布, 而对于异常数据, 尽管具有各种各样的分布, 但是它们都不能被拟合到正常数据的分布之中。

这意味着可以结合算法与大量的 SPD 多参数采样数据, 来训练出一个模型去学习正常参数数据的分布, 并用这个学习到的分布, 去检测出无法被拟合到这个分布中的异常参数数据。当数据的异常值越大, 表示 SPD 的实际性能越糟糕, 通过对 SPD 的异常参数进行监测, 实现了对其实状态及劣化程度进行评价预警。

4 监测性能分析

结合图 1 分析, 根据相关的电气、SPD 规范标准及专业人员的判断, 对数据集中对应器件实际性能完全损耗的数据条目做相应的标签。令正常数据添加标签为 1 类, 异常数据添加标签为 0 类, 在 390 组数据中标记了 38 个异常点, 异常值占比 0.097 4。模型的准确性及可靠性通过 ROC 曲线下面积 (area under curve, AUC) 量化。AUC 分数取值在 0.5~1, 分数越接近 1。表示模型结合查准率和查全率的综合性能越强。

表 1 是选取不同数量下的 ITree 和每个 ITree 随机选择不同样本数量的 AUC 值。在不同的参数的组合下, 模型介具有很高的 AUC 值。当选取集成 ITree 的数量为 60, 样本采样数量为 256 时, 模型的 AUC 值最高。

表 1 选取不同参数下的 AUC 值

ITree 数量	样本数量			
	64	128	256	390
20	0.967 02	0.964 75	0.970 3	0.966 94
40	0.971 91	0.979 87	0.972 09	0.980 93
60	0.978 71	0.976 22	0.981 15	0.973 19
80	0.974 29	0.978 38	0.979 87	0.977 43
100	0.976 58	0.979 58	0.976 92	0.975 60

图 6 所示为 IForest 在低维下 (以 PCA 选取 2 个主成分为例) 对本文 SPD 采样数据集进行异常监测的可视化结果。此时集成 ITree 数量的设置为 100, 每个 ITree 中样本采样数量为 256。图 6 中红色点为模型计算出的对应数据集中的异常点, 蓝色点为对应的正常点, 黑色点为模型计算出来异常值最小的数据点 (设置显示为 10 个)。对于分布稀疏的异常数据, IForest 能较好的将其检测出来, 即使异常数据的簇较大。这是因为 IForest 通过对数据集进行子采样建立局部模型, 从而赋予每个 ITree 的不同数据, 使其专门用来识别特定的子数据样本。

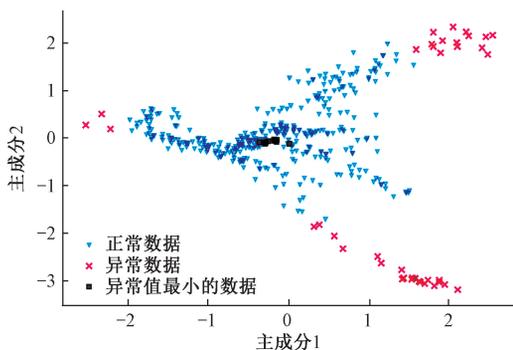


图 6 数据集在二维中的异常监测结果

Fig. 6 Detection results of the data set in 2-D space

表 2 为基于数据集实际标签,模型对数据集中每组数据异常性判别的精度统计。计算得到查准率为 98.01%,查全率为 98.29%,整体的准确度为 96.67%。表明模型经过了本文数据集的检验,能够准确的对一组 SPD 采样数据的异常程度进行判断。

表 2 模型识别结果

Table 2 Results of model recognition

类别	预测为正常	预测为异常
正常/1 类	346	6
异常/0 类	7	31

在本次 390 组数据中,IForest 还存在着对 13 组数据的错误判断,主要的原因在于由于数据的不足,模型所拟合的分布的“泛化”能力还有所欠缺。基于 3.3 节的模型用于监测任务的思想,其目标在于去学习到代表正常与异常数据的各自的分布。对于一组采样数据,如果它无法被拟合至正常数据的密集分布之中,那么就会被判断为是稀疏的异常的。

随着采样数据的增多,用于区分正常分布和异常分布的决策边界(图 7 的红色界线)会愈之具有鲁棒性。而正常数据在空间中的分布变化也会愈之固定(图 7 黄色区域,为模型学习到的正常数据的分布)。因此,符合大数据的思想,随着在实际应用中海量数据的支撑,基于 IForest 的 SPD 异常参数监测模型的性能会随着越来越精确的方向迭代。

5 结 论

为对电涌保护器件的实际状态及劣化程度的评价预警,本文基于孤立森林算法的构建电涌保护器的异常参数监测模型,对同时刻每组 SPD 多参数采样数据进行异常性的判别。根据模型在实验数据集上的监测结果,验证了该模型的方法思想对器件性能监测应用的可行性。当然,模型的计算策略也存在一些局限性,主要在于 IForest 不适合监测具有大量特征的数据的异常性。由于 IForest 对数据空间的分割为随机选取一个维度,所以在

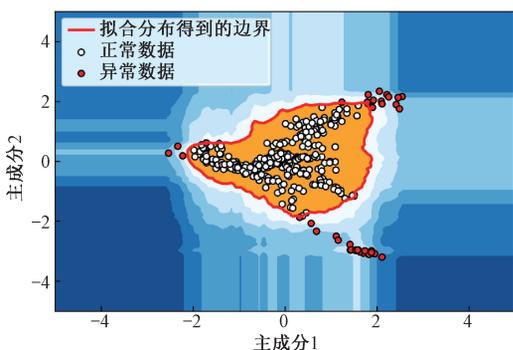


图 7 数据的异常性分布决策边界

构建完 ITree 后使得余下的维度信息被冗余,导致模型的鲁棒性降低。在未来的研究中,可以保持异常监测的思想不变,更换其他的技术模型作为改进工作。

参考文献

[1] 顿伟娟. 雷电电涌防护在低压交流电源系统的应用研究[J]. 通信电源技术,2017,34(3):147-148.
DUN W J. The application of lightning surge protection in low voltage AC power supply system [J]. Telecom Power Technology, 2017, 34(3): 147-148.

[2] 翁慧颖,陆海波,钱江,等. 一种电涌保护器智能监测系统的研制[J]. 电力科学与技术学报,2012,27(1): 59-63.
WENG H Y, LU H B, QIAN J, et al. Development of intelligent monitoring system for surge protective devices[J]. Journal of Electric Power Science and Technology, 2012, 27(1): 59-63.

[3] 陈登义,周雪会,李鹏. 电涌保护器老化监测指标对比研究[J]. 电瓷避雷器,2016,139(5):90-93.
CHEN D Y, ZHOU X H, LI P. Comparative study about index of SPD aging monitoring[J]. Insulators and Surge Arresters, 2016, 139(5): 90-93.

[4] 韩晗,潘学萍. 氧化锌避雷器在线监测方法的现状与发展[J]. 河海大学学报(自然科学版),2017, 45(3):282.
HAN H, PAN X P. Present state and future of online monitoring methods of ZnO arresters [J]. Journal of Hohai University (Natural Sciences), 2017, 45(3): 282.

[5] 盛泓,徐伟,冯晓宇,等. 电涌保护器老化研究的回顾与进展[J]. 电子测量技术,2017,40(11):16-20.
SHENG F, XU W, FENG X Y, et al. Review and progress on the aging of surge protective device [J]. Electronic Measurement Technology, 2017,40(11): 16-20.

[6] 徐伟,盛泓,行鸿彦. 电涌保护器中漏电阻性分量的监测及提取方法[J]. 电子测量与仪器学报,2018, 32(9): 133-141.

- XU W, SHENG F, XING H Y. Monitoring and extraction of resistive components of leakage current in surge protectors [J]. Journal of Electronic Measurement and Instrumentation, 2018, 32(9): 133-141.
- [7] 黄海宏, 颜碧琛, 赵常威, 等. 氧化锌避雷器泄漏电流检测的优化 FFT 分析 [J]. 电子测量与仪器学报, 2019, 33(4): 87-94.
- HUANG H H, YAN B SH, ZHAO CH W, et al. Optimized FFT analysis for leakage current detection of zinc oxide arrester [J]. Journal of Electronic Measurement and Instrumentation, 2019, 33(4): 87-94.
- [8] 王梧熠, 杨仲江, 于忠江, 等. 多脉冲下 MOA 老化劣化在线监测指标研究 [J]. 电瓷避雷器, 2019(1): 37-43.
- WANG W Y, YANG ZH J, YU ZH J, et al. Research on online monitoring index of aging deterioration of MOA under multi-pulse [J]. Insulators and Surge Arresters, 2019(1): 37-43.
- [9] 李祥超, 马骁骐. ZnO 压敏电阻在直流电压作用下老化劣化的分析 [J]. 电瓷避雷器, 2018(2): 1-7.
- LI X CH, MA X Q. Study of aging and deterioration of Zinc Oxide varistor under DC voltage [J]. Insulators and Surge Arresters, 2018(2): 1-7.
- [10] 王森, 牛博, 郭洁, 等. 采用避雷器动作电流实现过电压在线监测的新方法 [J]. 电网技术, 2013, 37(1): 201-205.
- WANG S, NIU B, GUO J, et al. A new approach for online overvoltage monitoring based on action current of MOA [J]. Power System Technology, 2013, 37(1): 201-205.
- [11] 黄新波, 罗兵, 王勇, 等. 用灰关联法分析气象因素对 MOA 在线监测的影响 [J]. 高电压技术, 2010, 36(6): 1468-1474.
- HUANG X B, LUO B, WANG Y, et al. Influence of climatic factors on on-line monitoring of MOA using grey relational analysis [J]. High Voltage Engineering, 2010, 36(6): 1468-1474.
- [12] 张红梅, 李佑光. 基于灰色模型的电涌保护器老化预测研究 [J]. 电瓷避雷器, 2016, 139(6): 73-76.
- ZHANG H M, LI Y G. Aging prediction of SPD based on gray model [J]. Insulators and Surge Arresters, 2016, 139(6): 73-76.
- [13] 颜旭, 陈绍东, 江润志, 等. 自然雷电下氧化锌避雷器残压特征分析 [J]. 中国电力, 2013, 46(7): 72-76.
- YAN X, CHEN SH D, JIANG R ZH, et al. Analysis on characteristics of residual voltage in ZnO SPD based on natural lightning [J]. Electric Power, 2013, 46(7): 72-76.
- [14] 李祥超, 陈则煌, 周中山, 等. 限压型电涌保护器在不同条件下劣化性能的分析 [J]. 低压电器, 2014(14): 12-17.
- LI X CH, CHEN Z H, ZHOU ZH SH, et al. Analysis on aging and deterioration of voltage limiting type surge protective device [J]. Low Voltage Apparatus, 2014(14): 12-17.
- [15] 杨仲江, 张彬, 柴健, 等. 氧化锌压敏电阻老化过程中非线性系数变化的研究 [J]. 电子元件与材料, 2011, 30(9): 27-30.
- YANG ZH J, ZHANG CH, CHAI J, et al. Research on the varying of nonlinear coefficient during the degradation of ZnO varistor [J]. Electronic Components & Materials, 2011, 30(9): 27-30.
- [16] 龚敏, 何智强, 饶玉凡, 等. 基于雷电冲击特性的带间隙氧化锌避雷器在线监测方法 [J]. 电瓷避雷器, 2018(5): 127-131.
- GONG M, HE ZH Q, RAO Y F, et al. On line monitoring method of Zinc Oxide surge arrester with gap based on lightning impulse characteristic [J]. Insulators and Surge Arresters, 2018(5): 121-125.
- [17] 徐志钮, 赵雨娟, 丁傲, 等. 一种新的 MOA 阻性电流提取算法 [J]. 电力自动化设备, 2010, 30(12): 47-51.
- XU ZH N, ZHAO L J, DING AO, et al. Calculation of MOA resistive current [J]. Electric Power Automation Equipment, 2010, 30(12): 47-51.
- [18] 杨仲江, 曹洪亮, 李鹏飞, 等. 基于遗传算法的金属氧化物避雷器在线监测 [J]. 高电压技术, 2015, 41(9): 3104-3109.
- YANG ZH J, CAO H L, LI P F, et al. On-line monitoring of metal oxide arrester using genetic algorithm [J]. High Voltage Engineering, 2015, 41(9): 3104-3109.
- [19] LIU F T, TING K M, ZHOU Z H. Isolation-based anomaly detection [J]. ACM Transactions on Knowledge Discovery from Data, 2012, 6(1): 1-39.
- [20] 李新鹏, 高欣, 阎博, 等. 基于孤立森林算法的电力调度流数据异常检测方法 [J]. 电网技术, 2019, 43(4): 1447-1456.
- LI X P, GAO X, YAN B, et al. An approach of data anomaly detection in power dispatching streaming data based on isolation forest algorithm [J]. Power System Technology, 2019, 43(4): 1447-1456.

作者简介



马俊彦, 现为南京信息工程大学硕士研究生, 主要研究方向为机器学习结合雷电预警技术。

E-mail: 112074898@qq.com

Ma Junyan currently a M. Sc. candidate at Nanjing University of Information Science and Technology. His main research direction is machine learning combined with lightning warning technology.