

球隙最短路径电场特征集与特征选择方法

金 颀^{1,2}, 邱志斌³, 阮江军¹, 王学宗¹, 全 好⁴, 杜志叶¹, 邓永清¹

(1. 武汉大学电气与自动化学院, 湖北 武汉 430072; 2. 中国舰船研究设计中心, 湖北 武汉 430064; 3. 南昌大学信息工程学院, 江西 南昌 330031; 4. 国网湖北省电力有限公司武汉供电公司, 湖北 武汉 430051)

摘要:球隙作为典型空气间隙是一种研究空气击穿特性的重要电极结构,依据球隙最短路径上的电场分布定义一组电场特征集。以 IEC 60052 标准中的 271 个球隙作为样本集,采用相关性系数、敏感性系数、遗传算法和随机森林 4 种方法进行特征选择,通过小样本数据对支持向量回归机进行训练,对其他测试样本的工频击穿电压进行预测。结果表明,在不进行特征选择时,所有样本的预测值与试验基准值相比平均绝对百分比误差为 4.21%,与试验值基本吻合,验证了所定义的最短路径电场特征集的有效性,采用上述 4 种特征选择方法进行特征选择后,预测结果平均绝对百分比误差分别为 1.92%、2.00%、3.86% 和 2.04%,其中相关性系数法具有最高的预测精度,对比结果发现, E_{\min} 、 L_{w90} 这 2 个特征量在模型中起到了至关重要的作用。该方法对复杂间隙乃至工程间隙的放电特性研究具有指导意义。

关键词:球隙;击穿电压;特征选择;支持向量回归;最短路径;电场特征集

DOI:10.19781/j.issn.1673-9140.2020.06.002 中图分类号:TM855 文章编号:1673-9140(2020)06-0012-09

Study on electric field features for the shortest path of sphere gap and the feature selection methods

JIN Qi^{1,2}, QIU Zhibin³, RUAN Jiangjun¹, WANG Xuezhong¹,
QUAN Yu⁴, DU Zhiye¹, DENG Yongqing¹

(1. School of Electrical Engineering and Automation, Wuhan University, Wuhan 430072, China; 2. China Ship Development and Design Center, Wuhan 430064, China; 3. Information Engineering School of Nanchang University, Nanchang 330031, China; 4. Wuhan Power Supply Company, State Grid Hubei Electric Power Co., Ltd., Wuhan 430051, China)

Abstract: As a typical air gap, the sphere gap is a kind of important electrode structure for studying air gap breakdown characteristics. Firstly, an electric field feature set is defined on the basis of electric field distribution on the shortest path. 271 power frequency breakdown voltages collected from IEC 60052: 2002 are considered as samples set. Then, four feature selection methods including the Pearson correlation coefficient, sensitivity coefficient, genetic algorithm and random forest are adopted to select features. In the end, the support vector regression (SVR) is trained by small sample data to predict power frequency breakdown voltages of other samples. The prediction results show that with-

收稿日期:2018-06-12;修回日期:2018-10-17

基金项目:国家自然科学基金(51477120);中国博士后科学基金(2016M602354)

通信作者:邱志斌(1991-),男,特聘教授,主要从事空气绝缘预测、输变电设备外绝缘、电磁场数值计算等研究;E-mail:qiuzyb@ncu.edu.cn

out the feature selection, the mean absolute percentage error is 4.21% for breakdown voltages of all samples with test results as the reference. Thus, the effectiveness of the feature set is verified. The mean absolute percentage errors of the results by adopting four feature selection methods are 1.92%, 2.00%, 3.86% and 2.04% respectively. Pearson correlation coefficient corresponds to the highest prediction accuracy. By comparing the results, it is concluded that the features of E_{\min} and L_{w90} play crucial roles in the model. The proposed method has a guiding significance to the study on discharge characteristics of complex gaps and engineering gaps in the future.

Key words: sphere gap; breakdown voltage; feature selection; support vector regression; the shortest path; electric field feature set

空气间隙是交、直流输电线路及变电站、换流站的主要外绝缘形式,其放电特性是输变电工程外绝缘设计的重要依据^[1-3]。由放电理论可知,空气间隙击穿电压主要与间隙结构、大气条件和加载电压类型有关,且放电过程极其复杂。为有效实现空气间隙击穿电压计算或预测,可以借助适合处理多维非线性问题的人工智能算法直接建立上述因素与击穿电压的灰关联性^[4]。

针对大气条件对击穿电压的影响研究,文献[5]研究了温度、相对湿度、气压、风速和辐照度与空气间隙放电电压的灰色关联度大小关系,并结合神经网络准确预测了空气间隙的放电电压;文献[6]应用了人工神经网络预测棒—板短空气间隙在淋雨条件下的交流放电电压,在训练范围内预测精度较高;文献[7]以球隙为研究对象,使用 BP 神经网络对无凝露情况下的空气间隙放电电压进行预测,预测准确率达 80.7%。针对间隙结构对击穿电压的影响研究;文献[8]提出了一种通过电场分布特征来表征间隙结构的方法,并提取多个区域(整个区域、放电通道、电极表面和最短路径)的电场特征量近似反映电场分布情况,借助支持向量机建立起了电极在空气中的击穿电压预测模型;文献[9-10]利用这一预测方法,分别预测了典型电极间隙如球隙、棒—板间隙、球—板间隙和异形电极间隙如环—板间隙、绞线—板间隙、绞线—球间隙的击穿电压,预测值与试验值基本相符。实际上,该方法的关键点在于如何通过更加有效的电场特征集来表征电极结构,目前的方法对放电通道的定义并不统一,难以推广到复杂的间隙结构中,且对电场特征量的定义方式存在较大改进空间,对是否存在通用且有效的电场特征集值得探索。

常见的空气间隙包含球隙、棒—棒间隙、棒—板

间隙乃至复杂的工程间隙等,当间隙结构存在差异、间隙距离不同时,间隙击穿特性也会不同。球隙作为一种典型电极空气间隙,其试验数据易于获取,试验平台易于搭建,是研究空气间隙放电理论的常见电极结构,对研究复杂间隙结构的击穿特性具有指导意义。为此,该文以球隙作为研究对象,提出一组最短路径电场特征集,将复杂的三维空间电场分布用最短路径电场分布来表征,解决了空间区域定义不统一的问题,利用支持向量回归机(support vector regression, SVR)对球隙击穿电压进行预测,以验证该文提出特征集的有效性,并研究多种特征选择方式对预测结果的影响,找到具有显著效果的最短路径电场特征,为今后对复杂的工程间隙放电特性研究奠定基础。

1 球隙击穿电压预测基本思想

1.1 最短路径电场特征集定义

从宏观角度来讲,空气间隙发生放电的影响因素可归纳为电极结构、电压波形以及大气参数 3 类。电极结构包括电极的几何形状、电极之间的距离、电极对地高度与周围物体之间的距离等,这些因素可统一通过单位电压下的静电场来表征。由于放电随机性的存在,当进行无数次球隙放电之后,放电发展路径必然满足某一统计学规律,即球隙的放电过程沿着轴线从高压电极向低压电极发展的可能性最大。且电场是无穷维分布,不同球径不同间距的最短路径电场分布有着显著的不同,因此,可将三维空间电场分布映射到选取最短路径上的电场分布来表征电极结构,简化了对空间静电场的描述过程。

当两球电极正对、一球施加单位电位、一球施加零电位时,最短路径上呈现“U”形的电场强度分布

曲线,由于无穷远处电位为零,一球电位为零,两球中心连线的中点不是一个完全对称点,因此该 U 形曲线并不会对称。为尽可能地描述最短路径上的电场分布,可定义主要包括电场强度、电场梯度、电场平方量、电场积分(电位)、路径长度以及与场强、梯度、平方量相关的比例系数等最短路径电场的特征量,各特征量示意如图 1 所示。

球隙最短路径上电场可由计算机仿真获取,将最短路径上的电场强度离散化, n 为路径取点总数, E_i 为最短路径上第 i 个点的场强,可定义 34 个最短路径电场特征量,如表 1 所示。

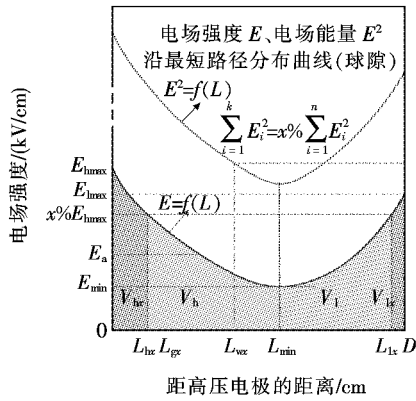


图 1 球隙最短路径特征量示意

Figure 1 Schematic diagram of features in the shortest path of sphere gap

表 1 最短路径电场特征量

Table 1 The shortest path electric field features

类别	电场特征量	个数
电场强度	$E_{hmax}, E_{lmax}, E_{min}, E_a, E_{std2}, E_{std}$	6
电场梯度	$E_{ghmax}, E_{glmax}, E_{ga}$	3
电场平方量	W, W_a	2
路径长度	$L_{E90}, L_{E75}, L_{g90}, L_{g75}, L_{w90}, L_{w75}, L_{min}$	7
电场分布不均匀程度	$E_d, L_{Er90}, L_{Er75}, L_{gr90}, L_{gr75}, E_{lh}, E_{rm}, L_{rm}, L_{wr90}, L_{wr75}, W_{r90}, W_{r75}$	12
电场积分	V_{90}, V_{75}, V_h, V_l	4

具体各个特征量的定义与计算公式如下:

1) 电场强度。包括最短路径上的高压电极场强最大值 E_{hmax} 、低压电极场强最大值 E_{lmax} 、最小值 E_{min} 和平均值 E_a 以及电场强度方差 E_{std2} 和标准差 E_{std} , 用来表征电场强度的离散程度。

$$E_{min} = \min E_i, i = 1, 2, \dots, n \quad (1)$$

$$E_{hmax} = \max E_i, i = 1, 2, \dots, m \quad (2)$$

$$E_{lmax} = \max E_i, i = m + 1, m + 2, \dots, n \quad (3)$$

$$E_a = \sum_{i=1}^n E_i / n \quad (4)$$

$$E_{std2} = \frac{1}{n} \sum_{i=1}^n (E_i - E_a)^2 = \frac{1}{n} \sum_{i=1}^n E_i^2 - E_a^2 \quad (5)$$

$$E_{std} = \sqrt{\frac{1}{n} \sum_{i=1}^n (E_i - E_a)^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n E_i^2 - E_a^2} \quad (6)$$

其中, m 为路径上场强最小值对应点的数目。

2) 电场梯度。包括从高压电极场强最大值 E_{hmax} 到场强最小值 E_{min} 这段路径上的电场梯度最大值 E_{ghmax} 、从场强最小值 E_{min} 到低压电极场强最大值 E_{lmax} 这段路径上的电场梯度最大值 E_{glmax} 和整条路径上的梯度平均值 E_{ga} 。

$$E_{ghmax} = \max(|-\text{grad } E_i|), i = 1, 2, \dots, m \quad (7)$$

$$E_{glmax} = \max(|-\text{grad } E_i|), i = m, m + 1, \dots, n \quad (8)$$

$$E_{ga} = \sum_{i=1}^n (|-\text{grad } E_i|) / n \quad (9)$$

3) 电场平方量。采用路径上场强平方的积分 W 和路径上场强平方的期望值 W_a 来表征。

$$W = \int_0^d E^2 dl \approx \sum_{i=1}^n E_i^2 d_i \quad (10)$$

$$W_a = \frac{W}{d} \quad (11)$$

其中, d 为间隙距离, d_i 为在路径上的 n 个点将间隙划分成每一段的长度, $d_i = d / (n - 1)$ 。

4) 路径长度。包括最短路径上场强值超过 $x\% \cdot E_{hmax}$ 的路径长度 L_{Ex} 、电场梯度超过 $x\% E_{ghmax}$ 的路径长度 L_{gx} 、累计电场平方量为 $x\% W$ 的路径长度 L_{wx} 以及场强最小值点到高压电极的距离 L_{min} 。 $x\%$ 取为 90%、75%, 因此, L_{Ex} 包含 L_{E90} 、 L_{E75} , L_{gx} 包含 L_{g90} 、 L_{g75} , L_{wx} 包含 L_{w90} 、 L_{w75} 。其中, L_{Ex} 和 L_{gx} 由如图 1 所示两段距离相加得到。路径长度类特征量共 7 个,路径长度类特征量的计算公式都具有形式:

$$L_x = \sum_{i=1}^p d_i \quad (12)$$

式中 p 为最短路径上满足对应场强、梯度、平方量相关条件的取点个数; d_i 为分段长度。

5) 电场分布不均匀程度。采用最短路径上与电场强度、电场梯度相关的比例系数表征电场分布的不均匀程度,包括场强畸变率 E_d 、场强值超过 $x\% \cdot E_{hmax}$ 的路径长度所占间距的比例 L_{Er90} 、 L_{Er75} ; 电场梯度值超过 $x\% E_{ghmax}$ 的路径长度所占间距的比例 L_{gr90} 、 L_{gr75} 以及两侧场强比例系数 E_{lh} 、最小场强比例系数 E_{rm} 和最小场强路径比例系数 L_{rm} 。

$$E_d = E_{hmax} / E_a \quad (13)$$

$$L_{rx} = L_x / d \quad (14)$$

$$E_{lh} = E_{lmax} / E_{hmax} \quad (15)$$

$$E_{rm} = E_{min} / E_{hmax} \quad (16)$$

L_{Erx} 、 L_{grx} 、 L_{rm} 的计算公式均为式(14)的形式。采用最短路径上与平方量相关的比例系数表征电场分布不均匀程度,累计电场平方量为 $x\% W$ 的路径长度所占间距的比例 L_{Wr90} 、 L_{Wr75} ; 场强值超过 $x\% \cdot E_{hmax}$ 的路径上电场平方量所占比例 W_{r90} 、 W_{r75} 。

$$W_{rx} = W_x / W = \left(\sum_{E_i \geq E_x} E_i^2 d_i \right) / W \quad (17)$$

式中 $E_x = x\% E_{hmax}$ 。此外, L_{Wrx} 的计算公式与式(14)具有相同的形式。

6) 电场强度的积分,即电位分布。包括场强值超过 $x\% E_{hmax}$ 的路径上的电场积分 V_{90} 、 V_{75} 以及高、低压电极到场强最小值点间的电位 V_h 、 V_l 。

$$V_x = \int_{E_i \geq E_x} E_i dl \approx \sum_{E_i \geq E_x} E_i d_i \quad (18)$$

$$V_h = \int_{E_{min} \leq E_i \leq E_{hmax}} E_i dl \approx \sum_{E_{min} \leq E_i \leq E_{hmax}} E_i d_i \quad (19)$$

$$V_l = \int_{E_{min} \leq E_i \leq E_{lmax}} E_i dl \approx \sum_{E_{min} \leq E_i \leq E_{lmax}} E_i d_i \quad (20)$$

1.2 预测流程图

支持向量回归机是在研究二分类问题的基础上提出的^[11],可用于解决回归问题,该文的 SVR 模型选择泛化能力较好的径向基核函数。基于 SVR 和最短路径特征集的球隙击穿电压预测流程如图 2 所示。

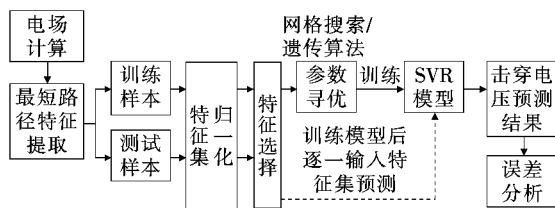


图 2 球隙击穿电压预测流程

Figure 2 Flow chart of breakdown voltage prediction in sphere-gap

惩罚系数 C 和核函数参数 g 决定了 SVR 的回归性能。为了获得一个较优的参数组,该文基于 K 折交叉验证的思想,通过网格搜索法或遗传算法找到较优的 C 和 g 。

误差分析是评价预测模型优劣的重要步骤,误差指标包括均方根误差 RMSE、平均绝对百分比误差 MAPE 和均方百分比误差 MSPE^[8]。

2 特征选择方法的基本原理

特征选择是机器学习领域的重要问题,进行特征选择的优点在于一方面能够揭示引起结果变化的重要特征,另一方面能够降低空间维数,缩小求解的规模,降低计算量^[12-13]。特征选择方法可分为过滤式^[14-16]、封装式^[17-18]和两者的组合^[19]。

2.1 相关性系数

皮尔逊(Pearson)相关性系数是用来衡量定距变量间的线性关系,是一种过滤式特征选择方法的评价准则^[19]。相关性系数的绝对值越大,相关性越强。通常情况下,相关性系数的绝对值在 0.8~1.0 之间时为极强相关,在 0.6~0.8 之间时强相关,在 0.4~0.6 之间时中等程度相关,在 0.2~0.4 之间时弱相关,0~0.2 之间时极弱相关或无相关。

假设有 N 个样本,需考察 x 和 y 之间的相关性,皮尔逊相关性系数的计算公式为

$$r_j = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (21)$$

式中 N 为球隙训练样本数; r_j 为第 j 个最短路径特征量与击穿电压的相关性系数; x_i 为第 i 个样本第 j 个最短路径特征量的数值; y_i 为第 i 个样本击穿电压的数值。

2.2 敏感性系数

敏感性系数反映最短路径特征量随间隙距离和球径变化的敏感程度,是一种过滤式特征选择方法的评价准则。在概率统计中,标准差用来反映组内个体间的离散程度,变异系数(coefficient of variation, CV)是一组数据的标准差与平均值的比,变异系数能消除测量尺度和量纲的影响,进行客观比较,是衡量结构特征量对影响因素敏感性的较好依据。

假设有 N 个训练样本,那么最短路径特征量在 N 个训练样本内的敏感性系数为

$$CV_j = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu_j)^2}}{\mu_j} \quad (22)$$

式中 CV_j 为第 j 个最短路径特征量的敏感性系数; x_i 为该最短路径特征量的数值; μ_j 为第 j 个最短路径特征量在样本空间中的平均值。

2.3 遗传算法

遗传算法是模拟达尔文生物进化论的自然选择和遗传学机理的生物进化过程的计算模型,是一种模拟自然进化过程搜索最优解的方法^[20]。在遗传算法中,每个种群包含一定数量的个体,每个个体具有一定位数的基因,当使用遗传算法进行特征选择时,基因的位数表征特征集合的总特征数,基因内的取值 0 或 1,表示去掉或保留该特征量,那么每一个个体就对应一种特征选择方式。选取训练 SVR 模型时交叉验证后均方误差 MSE 平均值的倒数作为适应度函数,适应度函数越大表明对于验证集来说,特征组合越优。这样通过一代代的交叉变异,当满足迭代终止条件时,输出的末代种群中适应度最高的个体便是筛选出的最具有代表性的特征组合,特征选择过程完成。基于遗传算法进行特征选择的流程如图 3 所示。

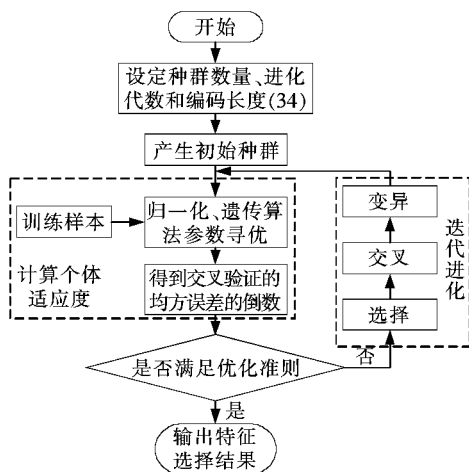


图 3 基于遗传算法进行特征选择的流程

Figure 3 Schematic diagram of feature selection based on genetic algorithm

2.4 随机森林

随机森林(random forest, RF)是一种集成机器学习方法,它利用随机重采样技术和节点随机分裂

技术构建多棵决策树,通过投票得到最终结果。其变量重要性度量可以作为高维数据特征选择的工具,是一种封装式特征选择方法^[21-23]。当使用随机森林算法进行特征选择时,各特征量的值即成为决策树的节点,当随机改变某一特征量的值时,基于随机森林的最终结果也会改变。因此,可以用 2 个指标来衡量特征量的重要性,分别是平均减少的准确度(mean decrease in accuracy, MDA)和平均减少的基尼指数(mean decrease in gini index, MDGI)。MDA 是指当把一个特征量变成随机数时,随机森林预测准确度的降低程度,该值越大表示该特征量的重要性越大。MDGI 是通过基尼指数计算每个特征量对分类树上每个节点的观测值的异质性影响,该值越大表示该变量的重要性越大。

3 预测及分析

3.1 训练样本集的选取

球隙研究对象取自 IEC 60052—2002^[24] 国际标准,共 271 个不同直径、不同间距的球隙在标准大气条件下的工频击穿电压值,大气参数一致,不考虑间距极小($d < 1.0$ cm)和球径极小($D < 5$ cm)以及准确度较低($d > 0.5D$)的数据。共有 11 种球直径的数据,分别是 5、6.25、10、12.5、15、25、50、75、100、150、200 cm,间隙距离的范围为 1~100 cm。

训练样本数过少,则特征选择不具有说服力和普遍性,SVR 模型极易出现“欠学习”,预测精度较差;训练样本数过多,则又淡化了所定义的特征在 SVR 模型中的作用。故训练样本共选取 20 个,训练样本集,如表 2 所示。

测试样本共 251 个,测试样本个数是训练样本个数的 12 倍多,并且,训练样本分布多集中在小球径短间隙处,故 SVR 模型在大球径长间隙时的预测

表 2 训练样本集

		cm					
球径	间距	球径	间距	球径	间距	球径	间距
5	1.4	12.5	2.2	15	7.0	50	6.5
5	2.2	12.5	2.5	15	7.5	50	22.0
6.25	2.4	12.5	4.5	25	5.5	150	5.5
10	4.5	12.5	5.0	25	6.5	150	11.0
12.5	1.2	15	1.5	25	7.0	150	22.0

精度能较大程度上反映最短路径特征集表征电场分布的能力。

3.2 特征选择结果

在特征选择过程中,所有信息仅从训练样本中提取,不包含任何测试样本的信息。

按相关性系数来划分,相关性系数的绝对值大于 0.6 为与击穿电压强相关或极强相关的特征量,按其值大于 0.6 进行特征选择。若按敏感性系数来划分,该组训练样本集中,34 维最短路径特征量的敏感性系数最大为 1.531,最小为 0.004 39,按敏感性系数大于 0.8 进行特征选择。

基于遗传算法进行特征选择时,种群数量为 30,进化代数为 600,特征选择的适应度函数进化曲线如图 4 所示。基于随机森林算法进行特征选择时,该训练样本集下平均减少的准确度和基尼指数指标如图 5 所示,2 项指标具有一定的相似性,选择大于 50% 平均减少的准确度的特征量作为特征选择结果。

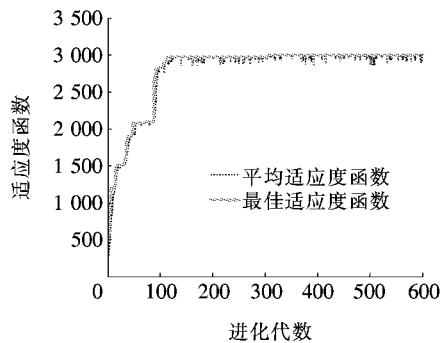


图 4 遗传算法的适应度变化曲线

Figure 4 Fitness change curve of genetic algorithm

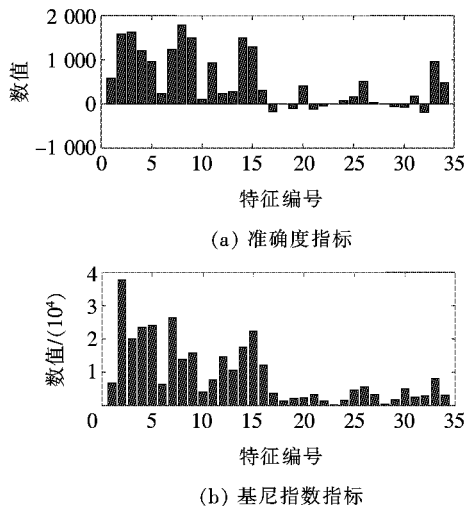


图 5 随机森林算法的评价指标

Figure 5 Evaluation index of random forest algorithm

4 种特征选择方式选择结果如表 3 所示(排序不分先后),其中,4 种特征选择方法都保留了 E_{\min} 和 L_{w90} 这 2 个最短路径特征,说明球隙最短路径上的最小电场强度和累计电场平方量为 90% 总平方量的路径长度与其击穿电压有极为重要的关联,而 L_{w75} 、 L_{E75} 、 E_{lmax} 、 E_a 、 L_{E90} 这 5 个特征量被 3 种特征选择方法选取,也与球隙击穿电压有着较为重要的关联。遗传算法和随机森林算法具有随机性,并不能保证每次都取得相同的结果。

表 3 特征选择结果

Table 3 The results of feature selection

特征选择方式	特征选择结果	个数
相关性系数	L_{w75} 、 L_{w90} 、 L_{\min} 、 L_{g75} 、 L_{g90} 、 L_{E75} 、 E_{hmax} 、 E_{lmax} 、 W 、 E_a 、 E_{\min} 、 L_{E90}	12
敏感性系数	L_{w75} 、 L_{w90} 、 L_{\min} 、 L_{g75} 、 L_{g90} 、 L_{E75} 、 E_{\min} 、 W_d 、 E_{ga} 、 E_{glmax} 、 E_{ghmax} 、 L_{E90} 、 W_{r90} 、 V_{90} 、 V_{r90} 、 L_{r90} 、 V_{r75}	17
遗传算法	L_{w90} 、 E_{lmax} 、 E_a 、 E_{\min} 、 L_{gr75} 、 E_{ghmax} 、 V_h 、 V_1 、 L_{E90} 、 V_{75}	10
随机森林	L_{w75} 、 L_{w90} 、 L_{E75} 、 E_{lmax} 、 W 、 E_a 、 E_{\min} 、 W_d 、 E_{ga} 、 E_{ghmax} 、 V_h	11

3.3 击穿电压预测结果与分析

在参数寻优过程中,均采用网格搜索法进行参数寻优,惩罚系数 C 的区间为 $[2^3, 2^{20}]$,核函数参数 g 的区间为 $[2^{-12}, 2^{-3}]$,寻优步长为 $2^{0.1}$,采用五折交叉验证,在搜索空间内选取使得交叉验证后均方误差最小的 C 和 g ,通过该参数对 SVR 模型进行训练,再将测试样本的信息一一输入至 SVR 模型中,得到预测结果。不同特征选择方法的寻优参数和误差指标如表 4 所示。11 种球径基于不同特征选择方法预测结果如图 6 所示。

由预测结果可知,不进行特征选择时,271 个样

表 4 寻优参数和误差指标

Table 4 Indexes of optimization parameters and errors

特征选择方式	C	g	RMSE	MAPE	MSPE
不进行特征选择	8 779.97	3.45×10^{-4}	35.856	0.042 1	0.003 66
相关性系数	17 559.94	2.06×10^{-2}	23.078	0.019 2	0.001 98
敏感性系数	70 239.75	4.25×10^{-4}	30.234	0.020 0	0.002 20
遗传算法	7 131.55	2.40×10^{-3}	29.945	0.038 6	0.003 19
随机森林算法	49 667.00	2.62×10^{-4}	29.488	0.020 4	0.002 05

本平均绝对百分比误差为 4.21%。在球径为 5 和 6.25 cm、间隙距离小于 2 cm 时,预测结果较试验值相比有较大的偏差。当球径和间隙距离同时最小时绝对误差可达 6 kV 左右,相对误差达到 21%。当球径为 10~25 cm 时,不论是否进行特征选择,预测结果和试验值有极高的吻合度,说明在极其有限的训练样本中,SVR 能够通过所定义或选择的特征对击穿电压进行预测,证明了所定义的最短路径电场特征的有效性。当球径大于 50 cm、间隙距离大于 30 cm、不进行特征选择时,其预测结果同试验值

有较大偏差,在 $D=150$ 、 $d=75$ cm 时,绝对误差达 151 kV,相对误差为 -10.5%。

经过 4 种特征选择方法进行特征选择后,271 个样本平均绝对百分比误差有不同程度的降低,其中,按相关性系数进行特征选择的平均绝对百分比误差最低,其值为 1.92%,基于遗传算法进行特征选择后对原结果有轻微的优化,平均绝对百分比误差为 3.86%,说明所定义的 34 维最短路径电场特征能够一定程度上反映球隙的空间电场分布情况,但是存在部分具有冗余信息的最短路径特征。由图 6

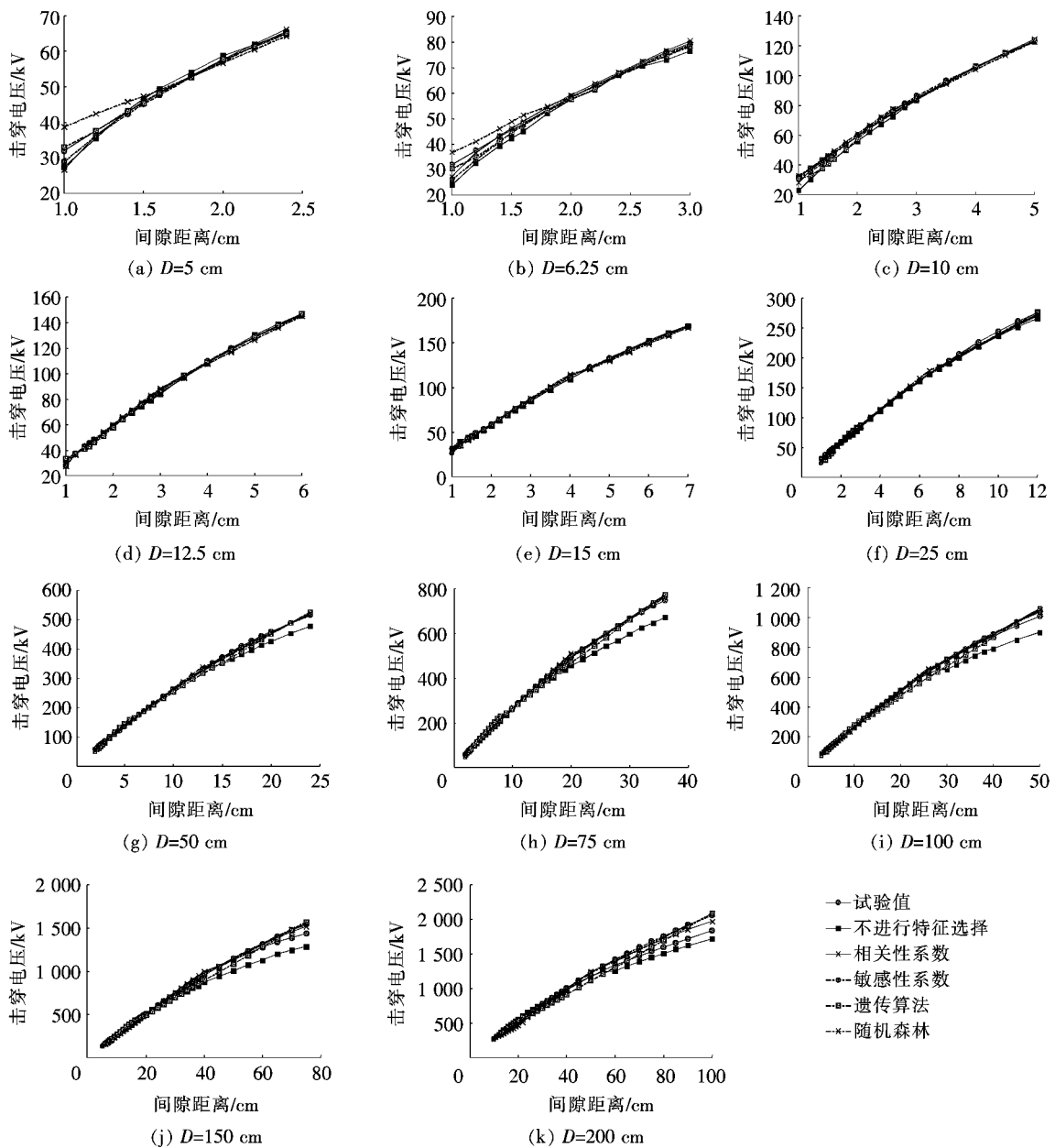


图 6 预测结果

Figure 6 Prediction results

可知,经过特征选择之后,在大球径长间隙的预测精度明显提高,按相关性系数进行特征选择时有最好的预测结果, $D=150$ 、 $d=75$ cm 时绝对误差减小为 82 kV,相对误差仅为 5.7%。虽然部分样本出现大于 5%的相对误差,但根据最短路径电场特征绝大多数样本都能很准确的预测出球隙击穿电压,进一步验证了所定义的最短路径特征集的有效性。

4 结语

1)定义了 34 维球隙最短路径电场特征集,该特征集能一定程度上反映球隙的三维空间电场分布情况,可看作空间电场分布在最短路径上的映射,并且经过 SVR 模型对球隙击穿电压进行预测,预测结果验证了该特征集的有效性。

2)采用相关性系数、敏感性系数、遗传算法和随机森林四种特征选择方法都能够降低 271 个球隙样本预测结果的平均绝对百分比误差,其中,采用相关性系数进行特征选择后预测精度最高,平均绝对百分比误差可从 4.21%降至 1.92%。

3)该方法初步证实了最短路径电场特征集能一定程度上反映间隙空间电场分布,对空气间隙的击穿特性研究具有指导意义,可将该方法推广到复杂间隙乃至工程间隙的放电特性研究中。

参考文献:

- [1] 黄道春,卢威,姚涛,等. 植被火条件下导线一板短空气间隙泄漏电流特性研究[J]. 电工技术学报,2019,34(16): 3487-3493.
HUANG Daochun, LU Wei, YAO Tao, et al. Study on leakage current characteristics of conductor-plane short air gap under vegetation fire condition[J]. Transactions of China Electrotechnical Society, 2019, 34(16): 3487-3493.
- [2] 霍锋,胡伟,徐涛,等. 1 000 kV 交流紧凑型输电线路杆塔空气间隙放电特性[J]. 高电压技术,2011,37(8): 1875-1881.
HUO Feng, HU Wei, XU Tao, et al. Air-gaps flashover characteristics for 1 000 kV AC compact tower[J]. High Voltage Engineering, 2011, 37(8): 1875-1881.
- [3] 林浩然,牛海清,张耿斌,等. 高湿条件下空气间隙放电规律的试验研究[J]. 高压电器,2014,50(7): 87-91.
LIN Haoran, NIU Haiqing, ZHANG Gengbin, et al. Experimental study on air-gap discharge characteristics in heavy humidity[J]. High Voltage Apparatus, 2014, 50(7): 87-91.
- [4] 舒胜文,刘畅,阮江军. 空气间隙击穿电压预测方法研究综述[J]. 高压电器,2016,52(7): 19-26.
SHU Shengwen, LIU Chang, RUAN Jiangjun. Review on prediction methods for breakdown voltage of air gap[J]. High Voltage Apparatus, 2016, 52(7): 19-26.
- [5] 牛海清,许佳,吴炬卓,等. 气隙放电电压的大气条件灰联度分析及预测[J]. 华南理工大学学报:自然科学版,2017,45(7): 48-54.
NIU Haiqing, XU Jia, WU Juzhuo, et al. Gray correlation analysis of atmospheric conditions and prediction of air gap discharge voltage[J]. Journal of South China University of Technology: Natural Science Edition, 2017, 45(7): 48-54.
- [6] 袁耀,蒋兴良,杜勇,等. 应用人工神经网络预测棒-板短空气间隙在淋雨条件下的交流放电电压[J]. 高电压技术,2012,38(1): 102-108.
YUAN Yao, JIANG Xingliang, DU Yong, et al. Predictions of the AC discharge voltage of short rod-plane air gap under rain conditions with the application of ANN[J]. High Voltage Engineering, 2012, 38(1): 102-108.
- [7] 张耿斌,罗新,沈杨杨,等. 大气条件对气隙放电电压的影响及神经网络在放电电压预测中的应用[J]. 高电压技术,2014,40(2): 564-571.
ZHANG Gengbin, LUO Xin, SHEN Yangyang, et al. Effect of atmosphere condition on discharge characteristics of air gap and the application of neural network[J]. High Voltage Engineering, 2014, 40(2): 564-571.
- [8] 舒胜文,阮江军,黄道春,等. 稍不均匀电场空气间隙击穿电压计算的新方法[J]. 高电压技术,2015,41(2): 662-668.
SHU Shengwen, RUAN Jiangjun, HUANG Daochun, et al. A new calculation method for the breakdown voltage of air gap in slightly uneven electric field[J]. High Voltage Engineering, 2015, 41(2): 662-668.
- [9] Qiu Z B, Ruan J J, Huang D C, et al. A prediction method for breakdown voltage of typical air gaps based on electric field features and support vector machine[J]. IEEE Transactions on Dielectrics & Electrical Insulation, 2015, 22(4): 2125-2135.
- [10] Qiu Z B, Ruan J J, Huang D C, et al. A method for

- breakdown voltage prediction of short air gaps with a-typical electrodes[J]. IEEE Transactions on Dielectrics & Electrical Insulation, 2016, 23(5): 2685-2694.
- [11] Vapnik V N. The nature of statistical learning theory [M]. New York, USA: Springer, 2000.
- [12] 姚旭, 王晓丹, 张玉玺, 等. 特征选择方法综述[J]. 控制与决策, 2012, 27(2): 161-166.
YAO Xu, WANG Xiaodan, ZHANG Yuxi, et al. Summary of feature selection algorithms[J]. Control and Decision, 2012, 27(2): 161-166.
- [13] 李邦彦, 齐伟强, 杨兆静, 等. 基于多特征选择方法的高压断路器故障诊断[J]. 高压电器, 2020, 56(6): 218-224.
LI Bangyan, QI Weiqiang, YANG Zhaojing, et al. Fault diagnosis of high-voltage circuit breaker based on multi-feature selection approaches[J]. High Voltage Apparatus, 2020, 56(6): 218-224.
- [14] Jain A K, Duin R P W, Mao J. Statistical pattern recognition: A review[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2000, 22(1): 4-37.
- [15] 李扬, 顾雪平. 基于改进最大相关最小冗余判据的暂态稳定评估特征选择[J]. 中国电机工程学报, 2013, 33(34): 179-186.
LI Yang, GU Xueping. Feature selection for transient stability assessment based on improved maximal relevance and minimal redundancy criterion[J]. Proceedings of the CSEE, 2013, 33(34): 179-186.
- [16] 刘文慧, 徐遵义, 张旭冉, 等. 基于互信息和 PCA 理论的湿法烟气脱硫工况特征提取方法[J]. 中国电力, 2020, 53(8): 158-163.
LIU Wenhui, XU Zunyi, ZHANG Xuran, et al. Feature selection method of wet flue gas desulfurization based on MI-PCA theory[J]. Electric Power, 2020, 53(8): 158-163.
- [17] Hsu W H. Genetic wrappers for feature selection in decision tree induction and variable ordering in Bayesian network structure learning[J]. Information Sciences, 2004, 163(17): 103-122.
- [18] Shima K, Todoriki M, Suzuki A. SVM-based feature selection of latent semantic features[J]. Pattern Recognition Letters, 2004, 25(9): 1051-1057.
- [19] 计智伟, 胡珉, 尹建新. 特征选择算法综述[J]. 电子设计工程, 2011, 19(9): 46-51.
JI Zhiwei, HU Min, YIN Jianxin. A survey of feature selection algorithm[J]. Electronic Design Engineering, 2011, 19(9): 46-51.
- [20] 马永杰, 云文霞. 遗传算法研究进展[J]. 计算机应用研究, 2012, 29(4): 1201-1206.
MA Yongjie, YUN Wenxia. Research progress of genetic algorithm[J]. Application Research of Computers, 2012, 29(4): 1201-1206.
- [21] Chan C W, Paelinckx D. Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery[J]. Remote Sensing of Environment, 2008, 112(6): 2999-3011.
- [22] 乔黎伟, 王静怡, 郭炜, 等. 基于随机森林算法的中短期用电量预测[J]. 电力科学与技术学报, 2020, 35(2): 150-156.
QIAO Liwei, WANG Jingyi, GUO Wei, et al. Medium and short-term electricity demand prediction based on random forests algorithm[J]. Journal of Electric Power Science and Technology, 2020, 35(2): 150-156.
- [23] 刘芹, 彭在兴, 王颂, 等. 基于随机森林算法的断路器分合闸线圈故障电流曲线识别[J]. 高压电器, 2019, 55(7): 93-100.
LIU Qin, PENG Zaixing, WANG Song, et al. Fault current curves identification of circuit breaker opening/closing coil based on random forest algorithm[J]. High Voltage Apparatus, 2019, 55(7): 93-100.
- [24] IEC 60052—2002. Voltage measurement by means of standard air gaps[S].