

DOI: 10.12158/j.2096-3203.2019.06.025

# 数据驱动的绝缘子积污特征量识别与污秽度预测

吴胜磊<sup>1</sup>, 滕松<sup>2</sup>, 刘振华<sup>1</sup>, 王新宽<sup>1</sup>, 迟鹏<sup>3</sup>

(1. 国网重庆市电力公司经济技术研究院, 重庆 401121; 2. 国网江苏省电力有限公司徐州供电公司, 江苏 徐州 221100; 3. 中国矿业大学电气与动力工程学院, 江苏 徐州 221116)

**摘要:**为准确预测复杂环境下输电线路绝缘子污秽度, 实现污闪的有效预警, 提出数据驱动的绝缘子积污特征量识别与污秽度预测方法。该方法结合改进粗糙集与样本加权方法, 基于粒子群算法优化的误差反向传播神经网络, 得到绝缘子积污特征量综合量化模型, 对影响积污程度的因素进行量化识别。在识别的基础上, 构建基于改进粗糙集的特征加权支持向量机, 来预测绝缘子污秽程度, 识别污闪风险。实例分析表明, 该方法完全基于数据驱动, 避免人为干预, 能实现不同运行环境下绝缘子积污特征量的准确识别。相较于其他方法, 所提污秽度预测和风险评估方法更精确, 误差更小, 具有良好的应用前景。

**关键词:**数据驱动; 绝缘子; 粗糙集; 支持向量机; 污秽度预测

中图分类号: TM726

文献标志码: A

文章编号: 2096-3203(2019)06-0179-08

## 0 引言

绝缘子是架空输电线路的一个重要组成部件, 承担导线和杆塔之间的机械连接和电气隔离责任<sup>[1-3]</sup>。绝缘子性能受到雾、小雨、小雪等恶劣天气及厂矿粉尘、煤烟的实时考验<sup>[4-7]</sup>, 绝缘子脏污潮湿导致的污秽闪络将造成输电网大面积、长时间停电<sup>[8-9]</sup>。输电线路污闪跳闸事故威胁着电网的安全稳定运行, 因此有必要对线路绝缘子的污闪防治进行深入研究。

绝缘子的积污不仅受型号、电压等级等影响, 更与外界环境有关。国内外学者对不同的绝缘子类型、气象条件、环境条件、大气污染等因素与等值盐密之间的定性关系进行了大量研究, 结果表明这些因素与绝缘子污秽度之间都存在一定的相关性<sup>[10-14]</sup>。因此对众多因素进行识别, 是实现绝缘子污秽度预测的前提。绝缘子等值盐密是确定污秽等级和绘制电网污区分布图的重要依据, 电力部门通过测量年度盐密划分污区, 根据污区图开展绝缘子的定期清扫工作<sup>[15]</sup>。这种方法虽然能在一定程度上防止污闪事故发生, 但由于污区划分基于往年数据而不能及时反映当前的真实情况<sup>[16]</sup>, 因此有必要寻求一种绝缘子表面等值盐密的实时预测方法, 来直接、有效地指导户外绝缘子的防污闪工作。目前已有相关研究利用污秽影响因素进行盐密预测, 周龙武等人以降雨量、暴雨次数、酸雨 PH 值、大气质量等气象数据为输入量, 以一定区域内绝缘子的

盐密均值为输出量, 最终得到盐密的概率密度函数<sup>[17]</sup>。王思华等人选取月降水量、月均风速、月均相对湿度、月降尘量 4 个特征量, 提出利用人工鱼群算法优化误差反向传播(back propagation, BP)神经网络, 建立绝缘子污秽等级预测模型<sup>[18-21]</sup>。这些研究不需要带电测试, 仅利用气象等数据实现了盐密的实时预测, 但由于没有明确影响因素对积污过程的贡献程度, 无法区分积污特征量的重要性, 仅通过人为规定来选择特征量, 选取的特征量是否具有代表性、预测方法是否准确还有待进一步研究。

文中结合信息熵改进的粗糙集和样本加权方法, 提出基于改进 BP 神经网络的绝缘子积污特征量识别模型, 得到关键的积污特征量, 去除低关联度数据。以关键特征量为输入, 提出基于改进粗糙集的特征加权支持向量机(rough set-weight-support vector machine, RS-W-SVM), 在降低计算复杂度的同时提高预测的准确性。实例分析表明, 所提方法能够实现绝缘子污秽度的准确预测和污闪风险的有效识别。

## 1 数据驱动的绝缘子积污特征量识别

数据驱动即完全通过数据本身得到所需的信息和知识。影响绝缘子积污的因素多、数据量大, 因此需要突出有效数据, 剔除低关联度数据。利用数据分析理论明确各个积污特征量与污秽度的关联程度, 再通过 BP 神经网络训练识别模型, 实现输入样本的积污特征量识别。

### 1.1 信息熵改进的粗糙集理论

粗糙集理论是一种处理模糊和不确定数据的

收稿日期: 2019-05-13; 修回日期: 2019-07-09

基金项目: 国家重点研发计划资助项目(2017YFC0804400); 国家电网有限公司总部科技项目(J2018078)

新型数学工具<sup>[22]</sup>。粗糙集理论把分类理解为在特定空间上的等价关系,等价关系则构成对数据的划分。给定一个决策表  $S=(U, A, V, f)$ ,其中  $U$  为论域, $A$  为属性集合, $V$  为  $A$  的值域, $f$  为映射关系。假设  $M$  是论域  $U$  上具有属性划分特性的随机变量, $M=\{M_1, M_2, \dots, M_n\}$  有如下概率分布:

$$[M: P] = \begin{bmatrix} M_1 & M_2 & \dots & M_n \\ p(M_1) & p(M_2) & \dots & p(M_n) \end{bmatrix} \quad (1)$$

其中  $p(M_i) = |M_i| / |U|$ ,  $|M_i|$  为集合  $M_i$  的基数。信息熵用来表征某个特征所有可能的平均不确定性。若某个特征有  $z_1, z_2, \dots, z_n$  种取值,对应概率为  $p_1, p_2, \dots, p_n$ ,且各信息之间出现概率彼此独立。信息熵即为单个不确定性的统计平均值,即:

$$H(U_i) = - \sum_{i=1}^n (p_i \ln p_i) \quad (2)$$

从概率论的角度,可将信息熵的熵函数引入到粗糙集理论中来度量论域  $U$  的不确定性。在信息论中用条件信息熵来度量已知随机变量  $X$  对随机变量  $Y$  的信息熵,记为  $H(Y|X)$ 。在决策表  $S$  中,决策集  $D$ (污秽度)对特征集  $C$  的条件信息熵可表示为:

$$H(D|C) = \sum_{j=1}^n \frac{|c_j|^2}{|U|^2} \sum_{i=1}^m \frac{|d_i \cap c_j|}{|c_j|} \left( 1 - \frac{|d_i \cap c_j|}{|c_j|} \right) \quad (3)$$

融合信息熵和粗糙集理论后,特征的重要度为:

$$K(c_j) = H(D|C) - H(D|C - \{c_j\}) + H(D|\{c_j\}) \quad (4)$$

在式(4)中,前两项作差考察的是在特征  $c_j$  剔除后的变化情况,若变化越大,则该属性越重要。 $H(D|\{c_j\})$  为特征  $c_j$  相对决策集的信息熵,是传统意义上的加权。因此任意特征  $c_j$  的相对权重可表示为:

$$w(c_j) = \frac{K(c_j)}{\sum_{k=1}^n K(c_k)} \quad (5)$$

## 1.2 样本加权与特征量综合量化

绝缘子积污特征数据可能缺失或存在误差,不同的样本对污秽度的预测贡献是不同的。因此某些与其他样本存在明显区别的样本会对特征量的量化识别产生更大的影响,通过样本加权方法进行处理。

假设样本集合  $X=(x_{ij} | i=1, 2, \dots, n; j=1, 2, \dots, p)$ ,  $X$  为  $n \times p$  矩阵,  $x_i$  为第  $i$  个样本,  $x_{ij}$  为第  $i$  个样本中第  $j$  个特征的值。通过样本间距实现样本加

权,首要步骤是将原始空间当中的所有样本映射到间距空间当中:

$$x'_{ij} = \sum_{l=1}^{|S|} |x_{ij} - x_{lj}(S)| - \sum_{l=1}^{|O|} |x_{ij} - x_{lj}(O)| \quad (6)$$

式中: $S$  为样本  $x_i$  同类样本的集合; $x_l(S)$  为  $S$  中的第  $l$  个样本; $|x_{ij} - x_{lj}(S)|$  为样本  $x_i$  与同类样本  $x_l(S)$  第  $j$  个特征上的偏差; $O$  为  $x_i$  不同类样本的集合; $x_l(O)$  为  $O$  中的第  $l$  个样本; $|S|$ ,  $|O|$  分别为 2 个集合中的样本个数; $|x_{ij} - x_{lj}(O)|$  为样本  $x_i$  与不同类样本  $x_l(O)$  第  $j$  个特征上的偏差。在原始特征空间中,同一种类别的样本在空间中聚集为一族。直接用数学公式识别显著样本较为复杂。映射到间距空间以后,无论是同类还是非同类,只要通过样本间距的计算便能识别出显著样本,即距离其他所有样本越远的样本越明显。

样本越“显著”即与其他样本的距离越大,对识别的影响越大。所以需要赋予这些样本较小的权值来消除样本的影响,任意样本  $x_i$  的权重  $\theta_i$  计算公式可以表达为:

$$\theta_i = \frac{1/\overline{dis(x'_i)}}{\sum_{i=1}^n (1/\overline{dis(x'_i)})} \quad (7)$$

其中,  $dis(x'_i)$  是通过欧氏距离计算得到的第  $i$  个样本与其他所有样本的平均间隔。

前面通过式(5)的计算,得到了各个特征的相对加权重  $w_1, w_2, \dots, w_p$  代表着第  $p$  个特征与污秽度的关联程度,可以看作是对样本的纵向加权,而式(7)所得的样本加权  $\theta_1, \theta_2, \dots, \theta_n$  是为了解决样本的不稳定性,可以看作是横向加权,由此可以得到每一个特征数据的综合量化值:

$$H_{np} = \alpha_1 \theta_n + \alpha_2 w_p \quad (8)$$

即原始数据与综合量化值有以下映射关系:

$$\begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \longrightarrow \begin{bmatrix} H_{11} & H_{12} & \dots & H_{1p} \\ H_{21} & H_{22} & \dots & H_{2p} \\ \vdots & \vdots & & \vdots \\ H_{n1} & H_{n2} & \dots & H_{np} \end{bmatrix} \quad (9)$$

随着输入数据量的增大,计算复杂度也将增大,而算法的效率将变低。采用 BP 神经网络对数据库进行学习训练,得到数据驱动的智能量化模型,实现给定一组新的测试数据,无需重复计算就能够快速准确地得到量化数值。

## 1.3 基于改进 BP 神经网络的绝缘子积污特征量量化识别模型

BP 神经网络是采用误差反向传播算法的有监

督学习方法,在电力领域中的应用已有大量的相关研究。为了加快 BP 网络的收敛速度、增强全局寻优能力,采用粒子群算法优化 BP 神经网络<sup>[23]</sup>。

基于数据驱动的绝缘子积污特征量量化识别的分析流程如图 1 所示。

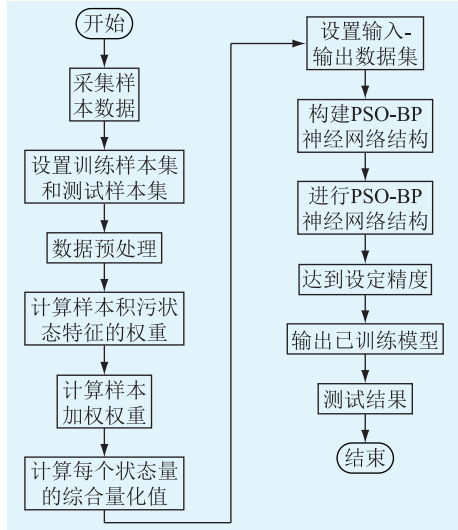


图 1 数据驱动的绝缘子积污特征量量化识别模型

Fig.1 Quantitative recognition model of insulator pollution state variable based on data driven

(1) 采集  $n$  个绝缘子的  $p$  个特征量的历史数据作为样本数据。将数据库分类,把前  $q$  组样本数据作为训练集,把后  $n \times q$  组样本数据作为测试集,并对数据进行预先处理。

(2) 计算样本集绝缘子影响因素的特征权重  $w_1, w_2, \dots, w_p$ , 计算样本集在间距空间中每个样本的加权权重  $\theta_1, \theta_2, \dots, \theta_n$ , 由 2 种权重得到每个样本的特征值的综合量化值  $H_{np} = \alpha_1 \theta_n + \alpha_2 w_p$ 。

(3) 根据原始样本和综合量化值的映射关系,得到综合量化样本数据库。将新的数据库输入到粒子群算法优化的误差反向传播 (particle swarm optimization-back propagation, PSO-BP) 神经网络中进行学习和训练,判断模型的精度是否达到预期,若达到预期则输出训练好的模型进行测试,若未达到预期则返回继续学习和训练。

(4) 将测试样本带入训练完成的模型中进行测试,输出测试样本的综合量化值,确定对测试样本影响程度最高的关键特征量,准备下一步的预测。

## 2 基于 RS-W-SVM 的绝缘子污秽度预测与污闪预警

通过积污特征量的量化识别筛除非关键数据,并利用关键特征量与污秽度的高关联性,采用 RS-W-SVM 方法,通过输入积污样本数据计算得到污秽

度的预测值,在减少了数据运算量的同时也提高了预测的准确性。

### 2.1 特征加权算法与核函数构造

绝缘子的各积污特征量对分类或回归的贡献程度是不同的。在使用支持向量机进行污秽度预测时,核函数的运算会受到特征相关程度的影响。为了区别每个特征的重要性,根据每个特征与目标函数之间的关联程度,对不同特征进行加权处理赋予  $[0, 1]$  的值,特征越重要权值越高。RS-W-SVM 的特点是选择合适的特征加权方法,构造特征加权的核函数。式(4)和式(5)中的属性相对权重  $w$  可用于特征加权核函数的构造,求取的权重  $w$  越大,表明信息增益越大,特征包涵信息量越多,特征相对决策属性越重要,对回归的贡献也就越大。

设已知训练样本集  $D = \{(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2), \dots, (\mathbf{x}_n, \mathbf{y}_n)\}$ , 其中  $\mathbf{x}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{ip})^T$  是  $p$  维的向量,由式(10)得到  $p$  维特征向量加权函数  $w_p$ 。下面构造特征加权的核函数, $p$  维样本  $\mathbf{x}_i$  对应的核矩阵  $k_p = \langle \varphi(\mathbf{x}_i), \varphi(\mathbf{x}_i)^T \rangle$ , 由线性配准方法对核矩阵加权:

$$w_p k_p = \langle w_p \varphi(\mathbf{x}_i), w_p \varphi(\mathbf{x}_i)^T \rangle \quad (10)$$

由核函数构造原理可知构造的特征加权核函数为:

$$K(\mathbf{x}_i, \mathbf{x}_j) = w_1(\varphi(\mathbf{x}_{i1}), \varphi(\mathbf{x}_{j1})) + w_2(\varphi(\mathbf{x}_{i2}), \varphi(\mathbf{x}_{j2})) + \dots + w_p(\varphi(\mathbf{x}_{ip}), \varphi(\mathbf{x}_{jp})) \quad (11)$$

其中,  $\mathbf{x}_{ip}$  为  $\mathbf{x}_i$  第  $p$  维;  $\mathbf{x}_{jp}$  为  $\mathbf{x}_j$  第  $p$  维。采用径向基函数(RBF) 通过内积关系来解析样本集合在高维空间中的映射关系。对 RBF 函数进行扩展得到适用于绝缘子污秽度预测的特征加权径向基函数  $w$ -RBF。

$$K_{\text{RBF}}(\mathbf{x}_i, \mathbf{x}_j) = e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}} \quad (12)$$

$$K_{w\text{-RBF}}(\mathbf{x}_i, \mathbf{x}_j) = \sum_{p=1}^n w_p e^{-\frac{\|\mathbf{x}_{ip} - \mathbf{x}_{jp}\|^2}{2\sigma^2}} \quad (13)$$

核函数反映了数据间的映射关系,在构建的  $w$ -RBF 函数的各项中,对各维特征进行特征加权计算,分配相应的权重,达到优化核函数的目的。用构造的  $w$ -RBF 函数代替支持向量机对偶条件中的核函数  $K(\mathbf{x}_i, \mathbf{x}_j)$ <sup>[24]</sup>:

$$\min_{w, b} \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i - \beta_i)(\alpha_j - \beta_j) K_{w\text{-RBF}}(\mathbf{x}_i, \mathbf{x}_j) + \varepsilon \sum_{i=1}^n (\alpha_i + \beta_i) - \sum_{i=1}^n y_i (\alpha_i - \beta_i) \quad (14)$$

$$\text{s.t.} \sum_{i=1}^n (\alpha_i - \beta_i) = 0 \quad (15)$$

$$\alpha_i \geq 0, \beta_i \leq C \quad i = 1, 2, \dots, n \quad (16)$$

得到基于改进粗糙集的 RS-W-SVM 回归的标准模型。

## 2.2 算法流程

结合综合量化识别结果,可以得到基于改进粗糙集的 RS-W-SVM 机污秽度预测算法,流程如图 2 所示,步骤如下:

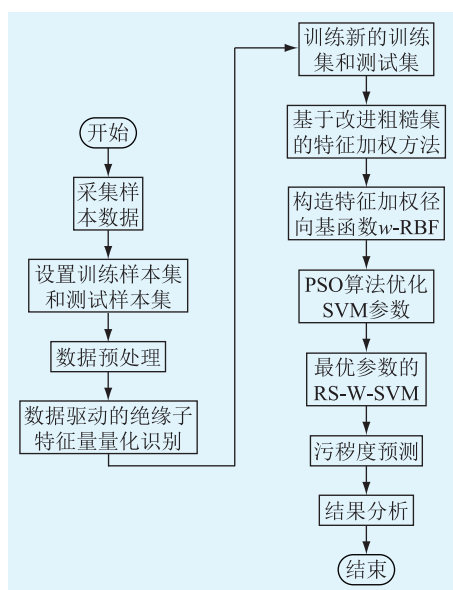


图 2 基于 RS-W-SVM 的绝缘子污秽度预测流程

Fig.2 Flowchart of pollution prediction of insulator based on RS-W-SVM

(1) 数据预处理。包括缺失数据、异常数据的处理以及数据的归一化和离散化。

(2) 基于数据驱动的积污特征量量化识别,确定关键影响因素作为输入特征,根据量化结果简化样本数据,划分训练集和测试集。

(3) 基于信息熵改进的粗糙集计算各特征的权重,并对特征向量相应的核矩阵进行加权。

(4) 选择基础核函数 RBF,引入加权的核矩阵构建再生核函数  $w$ -RBF。

(5) 将加权核函数  $w$ -RBF 带入到 SVM 求解框架中,应用 PSO 算法以预测值和实际值误差作为适应度,优化 SVM 的惩罚因子和核函数参数。

(6) 通过 RS-W-SVM 回归模型对测试绝缘子的污秽程度进行预测。

## 3 实例分析

### 3.1 积污特征量识别

对已公布研究中的绝缘子污秽数据进行了分析<sup>[25-28]</sup>。其中绝缘子环境和电气数据来自于运维数据库,气象数据来源于国家气象局。选取相同电压等级下的 80 组绝缘子污秽数据进行模型的仿真

验证,对特征量的综合量化系数进行分析。每组绝缘子污秽数据包含了特征量 12 个,特征量值为月测量平均值,如表 1 所示,AQI 为空气质量指数。

表 1 样本原始数据

对象	1	2	...	79	80
降水量/mm	3.3	1.4	...	21.7	39.7
温度/°C	-4	-5	...	11	21
相对湿度/%	55	57	...	41	65
降尘量/(g·m <sup>-2</sup> )	28.8	26.1	...	38.5	24.4
风速/(m·s <sup>-1</sup> )	5.5	4.6	...	4.8	3.6
风向	东风	东风	...	东风	西风
AQI	77	94	...	77	78
NO <sub>2</sub> /(ug·m <sup>-3</sup> )	37	31	...	74	68
SO <sub>2</sub> /(ug·m <sup>-3</sup> )	50	43	...	64	29
海拔高度/m	1 685	1 720	...	1 928	1 776
形状系数	0.74	0.74	...	4.90	3.65
运行时间/月	65	64	...	59	63
盐密/(mg·m <sup>-2</sup> )	0.085	0.093	...	0.105	0.068
污秽等级	3	3	...	4	3

设置 1—65 组数据为训练样本,66—80 组数据为测试样本。根据所提出的积污特征量量化方法,首先计算所有样本数据的综合量化值,权重系数均取 0.5。将更新后的训练样本带入 PSO 优化的 BP 神经网络中进行训练和学习,所有算法程序在 Matlab 中实现。经过多次仿真实验,设置 PSO 算法的种群的规模为 60,加速度系数为 2,粒子最大速度为 0.05,最大迭代次数设置为 500 次,BP 网络的收敛精度为  $10^{-3}$ ,输入-隐含-输出结构层数为 12-10-12。

将 15 组测试样本输入训练好的模型,得到 15 组样本 12 个综合量化值如表 2 所示。对于测试样本来说,12 个特征中降水量、相对湿度、降尘量、风速、AQI 指数、运行时间的综合量化值整体处于较高水平(>0.5)。综合量化值高即该 6 个特征量对测试绝缘子的污秽度影响最大,与污秽度的关联性最高。因此这 6 个特征量应作为测试绝缘子的关键特征量,在进行污秽度预测和污闪预警的时候,应以这 6 个特征量的数据作为主体,其他特征量作为辅助来进行计算和分析。

选取 SVM 和多元线性回归建立综合量化识别模型。以降水量为例,对比各个模型的测试结果和原始计算值,如图 3 所示。

由图 3 可知,所提方法测试得到的综合量化值更贴近于原始计算值曲线,测试值与原始值之间的距离最小,波动也最小。说明所提方法相较于其他

表2 综合量化值测试结果

Table 2 Comprehensive quantitative value results

特征量	1	2	...	7	8	...	14	15
降水量/mm	0.60	0.51	...	0.55	0.54	...	0.62	0.66
温度/°C	0.30	0.25	...	0.30	0.44	...	0.34	0.35
相对湿度/%	0.68	0.55	...	0.60	0.62	...	0.63	0.52
降尘量/(g·m <sup>-2</sup> )	0.55	0.53	...	0.52	0.52	...	0.60	0.54
风速/(m·s <sup>-1</sup> )	0.59	0.57	...	0.62	0.62	...	0.46	0.58
风向	0.20	0.20	...	0.18	0.17	...	0.19	0.24
AQI	0.60	0.60	...	0.52	0.64	...	0.53	0.56
NO <sub>2</sub> /(ug·m <sup>-3</sup> )	0.31	0.28	...	0.37	0.30	...	0.38	0.31
SO <sub>2</sub> /(ug·m <sup>-3</sup> )	0.33	0.38	...	0.35	0.34	...	0.35	0.30
海拔高度/m	0.22	0.22	...	0.19	0.20	...	0.22	0.20
形状系数	0.32	0.29	...	0.33	0.31	...	0.30	0.35
运行时间/月	0.65	0.60	...	0.52	0.58	...	0.63	0.65

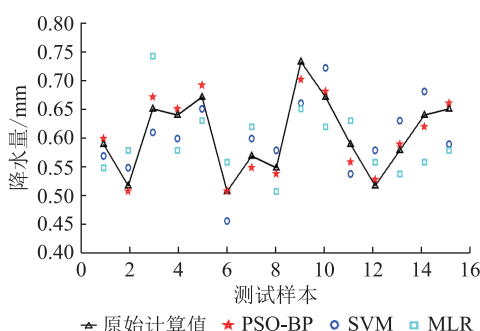


图3 不同方法量化值对比(降水量)

Fig.3 Comparison of comprehensive measurement value of precipitation

方法,对于特征量综合量化的拟合能力和非线性映射能力最强,应用文中所提方法能够得到准确的量化值。

### 3.2 污秽度预测与污闪风险预警

通过积污特征量识别,得到降水量、相对湿度、降尘量、风速、AQI 指数、运行时间 6 个关键特征量。由改进的粗糙集计算 6 个特征量的加权权重如表 3 所示,并带入 RS-W-SVM 中进行回归预测。根据 PSO 优化得到支持向量机的模型参数取  $C = 15.3, \sigma = 6.5$  时达到最优。

表3 特征加权重

Table 3 Weight of characteristics

特征	权重	特征	权重
降水量/mm	0.232	风速/(m·s <sup>-1</sup> )	0.127
相对湿度/%	0.209	AQI 指数	0.150
降尘量/(g·m <sup>-2</sup> )	0.177	运行时间/月	0.105

为证明所提出的绝缘子污秽度预测模型的预测性能,选取传统 SVM 模型、BP 神经网络模型对案例进行预测,并对 3 种模型的预测结果进行分析和

评价。以相对误差和平均相对误差作为预测模型的评价指标:

$$E_{rel} = \frac{y(i) - y^*(i)}{y(i)} \times 100\% \quad (18)$$

$$E_{ave} = \frac{1}{N} \sum_{i=1}^N \left| \frac{y(i) - y^*(i)}{y(i)} \right| \times 100\% \quad (19)$$

式中: $y(i)$ 为实际的污秽度; $y^*(i)$ 为预测的污秽度。

结合国网公司标准 Q/GDW 1152.1—2014,可以根据盐密值将绝缘子污闪风险等级划为非常轻、轻、中等、严重和非常严重 5 个等级,如表 4 所示。风险等级越高,表示绝缘子的污秽程度越严重,越容易发生污闪事故。当污闪风险等级较高时,应提前布置人力和清污装置,最大程度地降低污闪事故发生的可能性。

表4 绝缘子污闪风险等级

Table 4 Risk rank of pollution flashover of insulator

盐密/(mg·cm <sup>-2</sup> )	污闪风险等级	污闪风险程度
<0.025	1	非常轻
0.025~0.05	2	轻
0.05~0.10	3	中等
0.10~0.25	4	严重
>0.25	5	非常严重

通过 3 种预测模型得到测试绝缘子的污秽度预测结果,预测值、实际值、误差以及污闪风险预警结果如表 5 所示。

由表 5 可知,3 种方法中 BP 神经网络的相对误差最大值和最小值分别达到了 8.37% 和 -7.56%,并且误差的波动最大。SVM 相对误差最大值和最小值要小于 BP 神经网络,分别为 6.18% 和 -5.73%。这表明了对于样本规模较小的问题来说,SVM 的非线性拟合能力和模型的回归性能要优于 BP 神经网络。事实上,BP 神经网络在处理小样本问题时拟合能力较差,在数据量不断扩充以后学习能力能得到一定的提高。

由图 4 可知,文中所提出的 RS-W-SVM 相对误差的最大值和最小值仅为 2.80% 和 -2.96%,并且相对误差的波动和预测结果的离散程度最小。相较于前 2 种方法,所提方法预测的精确度最高且稳定性最好。

结果表明,RS-W-SVM 具有较强的学习能力和非线性拟合能力。进行污秽度预测时,首先在 SVM 的前端进行了积污特征量的量化识别,将与污秽度预测不相关和弱相关的数据进行了约简,保留了可读性和相关性最好的信息,加强了模型的计算性能。然后利用改进的粗糙集对特征进行加权处理,

表 5 不同模型的污闪风险预警结果

Table 5 Risk assessment of different model pollution flashover

测试 样本	实际值/ ( $\text{mg}\cdot\text{cm}^{-2}$ )	实际 风险	RS-W-SVM		SVM		BP 神经网络	
			预测值/ ( $\text{mg}\cdot\text{cm}^{-2}$ )	风险 程度	预测值/ ( $\text{mg}\cdot\text{cm}^{-2}$ )	风险 程度	预测值/ ( $\text{mg}\cdot\text{cm}^{-2}$ )	风险 程度
1	0.043	2	0.043 5	2	0.041 8	2	0.045 5	2
2	0.024	1	0.024 1	1	0.024 4	1	0.025 4	2
3	0.092	3	0.092 8	3	0.096 3	3	0.095 3	3
4	0.098	3	0.098 5	3	0.094 1	3	0.089 8	3
5	0.050	2	0.048 6	2	0.047 6	2	0.046 8	2
6	0.222	4	0.216 2	4	0.230 2	4	0.233 8	4
7	0.192	4	0.190 4	4	0.200 7	4	0.181 9	4
8	0.096	3	0.097 8	3	0.101 5	4	0.102 0	4
9	0.258	5	0.265 4	5	0.248 2	4	0.263 7	5
10	0.213	4	0.216 9	4	0.221 8	4	0.229 1	4
11	0.047	2	0.046 2	2	0.045 0	2	0.043 9	2
12	0.021	1	0.020 6	1	0.020 0	1	0.020 1	1
13	0.189	4	0.194 6	4	0.182 9	4	0.199 7	4
14	0.105	4	0.102 4	4	0.108 5	4	0.097 9	3
15	0.068	3	0.069 4	3	0.063 8	3	0.071 6	3
平均相对误差/%			1.79		4.10		5.68	
识别率/%			100		86.67		80	

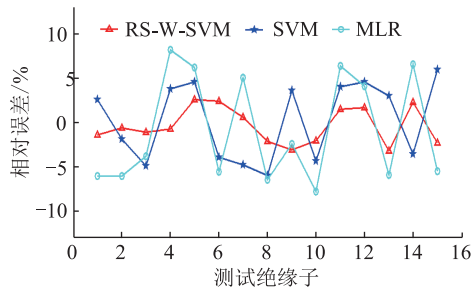


图 4 不同模型的相对误差对比

Fig.4 Comparison of different model relative error

对重要的特征赋予较大的权重,进一步增强了模型的拟合能力和学习能力。

同时,RS-W-SVM 对污闪风险的识别率达到了 100%,而 SVM 和 BP 神经网络分别有 2、3 个样本识别错误。相比之下,RS-W-SVM 能准确实现污闪风险的识别,使用所提出的方法能够为绝缘子污闪事故的预警和绝缘子的维护提供合理的指导。

#### 4 结语

文中建立了数据驱动的特征量量化和污秽度预测模型,并进行了实例分析验证。所建模型具有以下特点:

(1) 数据驱动的绝缘子积污特征量量化识别模型通过数据的综合量化,能够筛选出重要特征量,减少低关联度的数据。

(2) 基于 RS-W-SVM 的绝缘子污秽度预测模型利用经特征加权的关键特征量,在减少了数据运算量的同时也提高了预测的准确性。

(3) 算例分析表明数据驱动方法利用数据本身特征,避免了人为主观影响。相较于传统方法更准确、误差波动更小。该模型具有较强的实用性,适用于绝缘子污秽度的预测和污闪预警,也可用于其他电力设备的评估预测。

#### 参考文献:

- [1] 关志成. 绝缘子及输变电设备外绝缘[M]. 北京: 清华大学出版社, 2006.  
GUAN Zhicheng. Insulators and external insulation of transmission and distribution equipment[M]. Beijing: Tsinghua University Press, 2006.
- [2] 闫克平,李树然,冯卫强,等. 高电压环境工程应用研究关键技术问题分析及展望[J]. 高电压技术, 2015, 41(8): 2528-2544.  
YAN Keping, Li Shuran, FENG Weiqiang, et al. Analysis and prospect on key technology of high-voltage discharge for environmental engineering study and application[J]. High Voltage Engineering, 2015, 41(8): 2528-2544.
- [3] GUBANSKI S M, DERNFALK A, ANDERSSON J, et al. Diagnostic methods for outdoor polymeric insulators[J]. IEEE Transactions on Dielectrics and Electrical Insulation, 2007, 15(5): 1065-1080.
- [4] 李恒真,叶晓君,刘刚,等. 广州地区输电线路沿线绝缘子自

- 然污秽化学成分的来源分析[J]. 高电压技术, 2011, 37(8): 1937-1943.
- LI Hengzhen, YE Xiaojun, LIU Gang, et al. Source analysis on the chemical composition of natural contamination on the line insulator in Guangzhou area [J]. High Voltage Engineering, 2011, 37(8): 1937-1943.
- [5] 王康, 王建国, 姚森敬, 等. 深圳地区支柱绝缘子自然积污特性[J]. 高电压技术, 2014, 40(3): 780-786.
- WANG Kang, WANG Jianguo, YAO Senjing, et al. Natural contamination characteristics of post insulator in shenzhen area [J]. High Voltage Engineering, 2014, 40(3): 780-786.
- [6] 李永刚, 黄成才, 汪佛池, 等. 雾霾天气对复合绝缘子运行性能的影响[J]. 电网技术, 2014, 38(7): 1792-1797.
- LI Yonggang, HUANG Chengcai, WANG Fochi, et al. Influences of haze weather on operating performance of composite insulators [J]. Power System Technology, 2014, 38(7): 1792-1797.
- [7] 王少华, 胡文堂, 龚坚刚, 等. 浙江电网架空输电线路绝缘子串的自然积污特性[J]. 高电压技术, 2014, 40(4): 1002-1009.
- WANG Shaohua, HU Wentang, GONG Jianguo, et al. Natural pollution accumulation characteristics of overhead transmission line insulator strings of Zhejiang electric power grid [J]. High Voltage Engineering, 2014, 40(4): 1002-1009.
- [8] 宿志一. 防止大面积污闪的根本出路是提高电网的基本外绝缘水平[J]. 中国电力, 2003, 36(12): 57-61.
- SU Zhiyi. To intensify basic external insulation level of power system-fundamental way for prevention of large-scale pollution flashover [J]. Electric Power, 2003, 36(12): 57-61.
- [9] 宿志一, 李庆峰. 我国电网防污闪措施的回顾和总结[J]. 电网技术, 2010, 34(12): 124-130.
- SU Zhiyi, LI Qingfeng. Historical review and summary on measures against pollution flashover occurred in power grids in China [J]. Power System Technology. 2010, 34(12): 124-130.
- [10] 胡霖, 董彦武, 陈怡, 等. 大气环境参数与电网污秽等级划分间经验算式的修正[J]. 高电压技术, 2012, 38(3): 632-638.
- HU Ji, DONG Yanwu, CHEN Yi, et al. Revise of empirical formula of atmosphere environment parameter and dividing grid pollution level [J]. High Voltage Engineering, 2012, 38(3): 632-638.
- [11] 张楚岩, 王黎明, 周军, 等. 多串并联绝缘子交直流污闪特性[J]. 高电压技术, 2014, 40(11): 3357-3364.
- ZHANG Chuyan, WANG Liming, ZHOU Jun, et al. AC and DC pollution flashover performance of multiple parallel suspension insulators [J]. High Voltage Engineering, 2014, 40(11): 3357-3364.
- [12] 李清, 胡文, 李黎, 等. 绝缘子自然积污饱和趋势及气象因素影响分析[J]. 高电压技术, 2014, 40(11): 3336-3342.
- LI Qing, HU Wen, LI Li, et al. Analysis of natural contamination saturation tendency of insulators and influence of meteorological factors [J]. High Voltage Engineering, 2014, 40(11): 3336-3342.
- [13] BAKER A C, FARZANEH M, GORUR R S, et al. Insulator selection for AC overhead lines with respect to contamination [J]. IEEE Transactions on Power Delivery, 2009, 24(3): 1633-1641.
- [14] 黄俊杰, 阮羚, 熊宇, 等. 一种动态参数优化的绝缘子等值盐密预测模型[J]. 武汉大学学报, 2016, 49(3): 423-428.
- HUANG Junjie, RUAN Ling, XIONG Yu, et al. A forecast model of ESDD for insulator based on dynamic parameters optimization [J]. Engineering Journal of Wuhan University, 2016, 49(3): 423-428.
- [15] 宿志一. 用饱和盐密确定污秽等级及绘制污区分布图的探讨[J]. 电网技术, 2004, 28(8): 16-19.
- SU Zhiyi. Study on defining pollution classes and mapping pollution areas by saturated equal salt deposit density [J]. Power System Technology, 2004, 28(8): 16-19.
- [16] 彭功茂, 关志成, 张福增, 等. 复合绝缘子的直流人工污秽试验方法[J]. 高电压技术, 2011, 37(3): 570-576.
- PENG Gongmao, GUAN Zhicheng, ZHANG Fuzeng, et al. Test method of DC artificial contaminated composite insulators [J]. High Voltage Engineering, 2011, 37(3): 570-576.
- [17] 周龙武, 龚泽, 上官帖, 等. 基于概率统计与神经网络相结合的绝缘子盐密预测[J]. 水电能源科学, 2015, 33(12): 172-175.
- ZHOU Longwu, GONG Ze, SHANGGUAN Tie, et al. Prediction of insulator's ESDD based on probability statistics and neural network [J]. Water Resources and Power, 2015, 33(12): 172-175.
- [18] 王思华, 景弘. 一种接触网绝缘子污闪预测方法的研究[J]. 铁道学报, 2018, 40(3): 58-67.
- WANG Sihua, JING Hong. Study on method for predicting pollution flashover of insulators in contact network [J]. Journal of the China Railway Society, 2018, 40(3): 58-67.
- [19] HONG J, WANG S H. Research on pollution flashover probability of contact network insulator based on meteorological factors [J]. International Journal of Electrical Engineering, 2016, 23(4): 133-142.
- [20] 王曦冉, 何英静, 来聪, 等. 基于加权熵的电网清洁性评估[J]. 浙江电力, 2018, 37(1): 37-41.
- WANG Xiran, HE Yingjing, LAI Cong, et al. Power system cleanliness evaluation based on weighted entropy [J]. Zhejiang Electric Power, 2018, 37(1): 37-41.
- [21] 李德智, 张文文, 嵇建飞. 基于果蝇优化模糊神经网络模型的售电市场评估[J]. 电力需求侧管理, 2019, 21(1): 27-31.
- LI Dezhi, ZHANG Wenwen, JI Jianfei. Electricity retail market evaluation based on fruit fly optimized fuzzy neural network model [J]. Power Demand Side Management, 2019, 21(1): 27-31.
- [22] PAWLAK Z, SKOWRON A. Rudiments of rough sets [J]. Information Sciences, 2007, 177(1): 3-27.
- [23] 罗新, 牛海清, 来立永, 等. 粒子群优化自适应小波神经网络在带电局放信号识别中的应用[J]. 电工技术学报, 2014, 39(11): 3336-3342.

- 2014, 29(10): 326-333.
- LUO Xin, NIU Haiqing, LAI Liyong, et al. Application of adaptive wavelet neural network based on particle swarm optimization algorithm in online PD pattern recognition [J]. Transactions of China Electrotechnical Society, 2014, 29(10): 326-333.
- [24] BI J, BENNETT K P. Duality, geometry, and support vector regression [J]. Advances in Neural Information Processing Systems, 2002, 4(9): 593-600.
- [25] 王健, 杨志超, 葛乐, 等. 基于 BP 神经网络和模糊逻辑的绝缘子污秽等级预测 [J]. 南京工程学院学报(自然科学版), 2013, 11(4): 17-22.
- WANG Jian, YANG Zhichao, GE Le, et al. Prediction of insulator pollution severity class based on BP neural network and fuzzy logic [J]. Journal of Nanjing Institute of Technology (Natural Science Edition), 2013, 11(4): 17-22.
- [26] 岳良顺, 刘念, 梁杉, 等. 小波模糊神经网络在绝缘子污秽在线监测中的应用 [J]. 高电压技术, 2010, 36(10): 2483-2487.
- YUE Liangshun, LIU Nian, LIANG Shan, et al. Application of wavelet fuzzy neural network to contamination on-line monitoring of insulators [J]. High Voltage Engineering, 2010, 36(10): 2483-2487.
- [27] 景弘. 接触网绝缘子污闪预测方法的研究 [D]. 兰州: 兰州交通大学, 2017.
- JING Hong. Research on contact line insulator pollution flash-over prediction method [D]. Lanzhou: Lanzhou Jiaotong University, 2017.
- [28] 何相佑, 向凤红, 忽建蕊. 基于 BP 神经网络的绝缘子等值附盐密度预测 [J]. 绝缘材料, 2008, 41(4): 58-61.
- HE Xiangyou, XIANG Fenghong, HU Jianrui. Prediction of ESDD of outdoor insulators based on BP artificial network with fuzzy output [J]. Insulating Materials, 2008, 41(4): 58-61.

作者简介:



吴胜磊

吴胜磊(1994),男,硕士,从事输电线路状态评估工作(E-mail:p5411578@163.com);

滕松(1986),男,硕士,工程师,从事电力系统自动化,柔性直流输电工作;

刘振华(1985),男,硕士,工程师,从事输电线路规划与设计工作。

## Identification of pollution characteristics of transmission line insulator and pollution prediction based on data driven

WU Shenglei<sup>1</sup>, TENG Song<sup>2</sup>, LIU Zhenhua<sup>1</sup>, WANG Xinkuan<sup>1</sup>, CHI Peng<sup>3</sup>

(1. State Grid Chongqing Electric Power Company Economic Research Institute, Chongqing 401121, China;

2. Xuzhou Power Supply Company of State Grid Jiangsu Electric Power Co., Ltd., Xuzhou 221100, China;

3. School of Electrical and Power Engineering, China University of Mining and Technology, Xuzhou 221116, China)

**Abstract:** In order to accurately predict the pollution of transmission line insulators in complex environment and to achieve warning of pollution flashover, a method for identification of pollution characteristics of transmission line insulator and pollution prediction based on data driven is proposed. Combined with improved rough set and sample weighting method, a comprehensive quantitative model of insulator pollution characteristics based on PSO-BP neural network, and important characteristics affecting pollution can be quantified. On the basis of identification, a weighted support vector machine based on improved rough set is constructed to predict the pollution of the insulator and identify the risk of pollution flashover. The results show that the method is completely based on data driving and the characteristics of insulators with different operating environments and complex data types can be accurately identified. Compared with other methods, the proposed pollution prediction and risk identification method is more accurate and has smaller error due to the importance of the characteristics. The method proposed has good application prospects.

**Keywords:** data driven; insulator; rough set; support vector machine; pollution prediction

(编辑 钱悦)