

# 基于支持向量机的复杂环境条件下 绝缘子闪络电压的预测

舒立春<sup>1</sup>, 白困利<sup>2</sup>, 胡琴<sup>1</sup>, 蒋兴良<sup>1</sup>, 杨庆<sup>1</sup>

(1. 重庆大学高电压与电工新技术教育部重点实验室, 重庆市沙坪坝区 400044;

2. 重庆电力公司江北供电局, 重庆市江北区 401147)

## Insulator Flashover Voltage Forecasting Under Complex Circumstance Based on Support Vector Machine

SHU Li-chun<sup>1</sup>, BAI Kun-li<sup>2</sup>, HU Qin<sup>1</sup>, JIANG Xing-liang<sup>1</sup>, YANG Qing<sup>1</sup>

(1. Key Laboratory of High Voltage and Electrical New Technology of Ministry of Education, Chongqing University, Shapingba District, Chongqing 400044, China; 2. Chongqing Electric Power Corporation Jiangbei Power Supply Bureau, Jiangbei District, Chongqing 401147, China)

**ABSTRACT:** According to the test result on XZP-160 insulator in a large artificial climate chamber, a flashover voltage forecasting method based on support vector machine (SVM) is put forward. It is a new statistical study method in which the traditional empirical risk minimization principle is replaced by structural risk minimization principle. Using environmental conditions (atmosphere pressure, ice weight and pollution degree) as inputs, insulator flashover voltage as outputs, the relation between environmental conditions and flashover voltage is trained and the flashover voltage forecasting model is built. The forecasting result is in concordance with test result. The method provides a new way to select external insulation under complex circumstance conditions.

**KEY WORDS:** support vector machine; complex circumstance; insulator; forecasting of flashover voltage

**摘要:** 在大型人工气候实验室对 XZP-160 绝缘子试验数据的基础上, 提出了一种基于支持向量机的绝缘子闪络电压预测方法。支持向量机是以统计学习理论为基础的, 采用结构风险最小化原则代替传统经验风险最小化原则的新型统计学习方法。该文以气压、覆冰、污秽程度等环境条件作为输入, 绝缘子的闪络电压作为输出, 对环境条件和闪络电压的关系进行训练, 建立绝缘子闪络电压的预测模型。结果表明预测的闪络电压与实测结果基本一致。该方法为复杂环境条件下外绝缘的选择提供了一种新的途径。

基金项目: 国家自然科学基金项目(90210026)。

Project Supported by National Science Foundation of China (90210026).

**关键词:** 支持向量机; 复杂环境; 绝缘子; 闪络电压预测

## 0 引言

我国一次能源和经济发展在区域分布上极不平衡。丰富的水力资源主要集中在西部高海拔、高寒山区, 而经济发达的地区主要集中在东南沿海省市<sup>[1]</sup>。这种能源资源分布和经济发展区域不平衡的特点, 宏观上决定要将强大的电力从西部地区输送到东南经济发达地区, 输电线路必然要跨越一些复杂环境地区, 这些地区的气候条件与平原地区有很大的差异, 高海拔(低气压)、覆冰、污秽等环境条件对输电线路绝缘子的外绝缘性能有显著的影响<sup>[1-3]</sup>。而闪络电压是绝缘子电气性能的基本特征<sup>[4]</sup>, 绝缘子的闪络电压与污秽、覆冰、气压等因素有关, 它综合地反映了复杂环境中的各种影响因素对绝缘子外绝缘特性的影响。而且, 绝缘子闪络电压也是输电线路外绝缘选择和设计的重要参考依据。

目前, 复杂环境条件下外绝缘选择主要依赖于实验室的试验研究结果, 对试验数据的分析并进行外绝缘选择一般采用传统的惯用法、统计法和简化统计法确定结果<sup>[5]</sup>。但是由于环境信息具有多维性, 环境信息与绝缘子的闪络电压的关系复杂, 可能不符合现有的某个统计模型或统计分布假设, 或者说根本就不知道应该符合哪一种统计模型或统计分布, 这种关系具有非线性的特性, 且一般这些试验数据样本容量有限, 属于小样本的情况。而传统的

惯用法、统计法和简化统计法属于大样本的方法，而且经过一定的统计假设<sup>[5]</sup>。所以，如果仍用这些方法来分析和指导外绝缘的选择，效果会不理想。

此外，国内外学者对外绝缘特性也提出了一些预测方法，对闪络电压、等值附盐密度(ESDD)等都进行过预测，采用的方法包括传统的最小二乘法、神经网络等<sup>[6-10]</sup>。文献[6-7]采用传统的最小二乘法的多元统计线性回归分析来预测多种绝缘介质的绝缘强度和闪络电压，文献[8-9]也采用这种方法来预测输电线路绝缘子在多种因素影响下的 ESDD。但是基于传统的最小二乘法的多元统计回归分析，对误差做了正态假设，属于大样本统计方法，而且该方法处理线性关系容易，但是处理非线性关系很困难，另外该方法不具有稳健性。文献[10]采用 BP 神经网络的方法来预测输电线路绝缘子在多种因素影响下的 ESDD，虽然人工神经网络具有很强的自学习能力，能够实现复杂的非线性关系，但是由于神经网络的学习算法采用经验风险最小化原理(empirical risk minimization principle, ERM)<sup>[11]</sup>，仅仅使经验风险最小化，并没有使期望风险最小化，所以该方法存在过学习和低泛化问题，特别是在小样本的情况下，问题尤为严重；而且该方法得到的最优解可能只是局部最优解，不能保证是全局最优解。人工神经网络与传统的最小二乘法相比，在原理上缺乏实质性的突破，同时缺乏理论依据<sup>[11]</sup>。

而支持向量机(support vector machine, SVM)具备了以上方法无法具备的优点<sup>[12-15]</sup>。首先，支持向量机采用结构风险最小化原理(structural risk minimization principle, SRM)<sup>[13]</sup>，同时最小化训练样本的经验风险和置信范围(也称 VC 界)，这样就保证了学习的推广性，对未来样本有较好的泛化性能，特别是对小样本的情况。而且，支持向量机引入了特征空间和核函数的概念，把非线性问题映射到高维空间当作线性问题解决，同时解决了“维灾”问题。另外，核函数的 Mercer 条件保证了支持向量机训练的二次规划问题是凸优化问题，即得到的最优解为全局最优解。最后，支持向量机的结果只与支持向量有关，通过参数的选择，保证结果具有稳健性。所以本文采用支持向量机来预测绝缘子的闪络电压。

## 1 支持向量机回归估计的方法和原理

### 1.1 支持向量机的训练

支持向量机是从线性可分的最优分类超平面(optimal hyperplane)提出的，最初是解决模式识

别的分类问题<sup>[15]</sup>。设  $n$  个训练样本集  $D=\{(x_i, y_i) | i=1, 2, \dots, n\}$ ， $x \in R^n$ ， $y \in [1, -1]$ ，能被一个超平面  $w \cdot x + b = 0$  没有错误地分开，则最优分类超平面可以由下面的最优化问题表示

$$\min f(w) = \frac{1}{2} \|w\|^2 \quad (1)$$

$$\text{s.t. } y_i[(w \cdot x) + b] - 1 \geq 0, i=1, 2, \dots, n$$

式中参数  $b$  是阈值。

在分类问题的基础上引入一种新的损失函数— $e$  不敏感损失函数，这样支持向量机就能解决回归估计问题。 $e$  不敏感损失函数<sup>[14]</sup>：如果预测值和实际值之间的差别小于  $e$ ，则损失等于 0。记为

$$|y - f(x)| = \begin{cases} 0, & |y - f(x)| \leq 0 \\ |y - f(x)| - e, & \text{其它} \end{cases} \quad (2)$$

式中： $y$  为实际值； $f(x)$  为预测值。

这样，最优分类超平面的优化问题可以转化为

$$\min \frac{1}{2} \|w\|^2 \quad (3)$$

$$\text{s.t. } \begin{cases} y_i - w \cdot x_i - b \leq e \\ w \cdot x_i + b - y_i \leq e, i=1, \dots, n \end{cases}$$

针对限制条件，引入相应的松弛变量  $x$ ， $x^*$  和惩罚参数  $C$ ，上式的最优化问题转化为

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (x_i + x_i^*) \quad (4)$$

$$\text{s.t. } \begin{cases} y_i - w \cdot x_i - b \leq e + x_i \\ w \cdot x_i + b - y_i \leq e + x_i^* \\ x_i, x_i^* \geq 0, i=1, \dots, n \end{cases}$$

把式(4)的最优化问题转化为对应的对偶问题

$$\max W(a, a^*) = \sum_{i=1}^n (a_i - a_i^*) y_i - e \sum_{i=1}^n (a_i + a_i^*) - \frac{1}{2} \sum_{i,j=1}^n (a_i - a_i^*) (a_j - a_j^*) K(x_i, x_j) \quad (5)$$

$$\text{s.t. } 0 \leq a_i, a_i^* \leq C, \sum_{i=1}^n (a_i - a_i^*) = 0, i=1, \dots, n$$

式中： $K(x_i, x_j)$  为核函数； $a, a^*$  为拉格朗日乘子。

从式(5)可以看出支持向量机的训练问题就是一个二次规划问题。

### 1.2 支持向量机回归函数

由式(5)训练的结果就可以知道哪些是支持向量，以此确定回归函数。根据 KKT 条件<sup>[14]</sup>： $a_i a_i^* = 0$ ，而且限制条件要保证拉格朗日乘子大于 0。这样就可以分 5 种情况来判断是否为支持向量<sup>[16]</sup>：①  $a_i = 0$ ， $a_i^* = 0$  所对应的  $x_i$  不是支持向量；②  $0 < a_i < C$ ，

$a_i^* = 0$  所对应的  $x_i$  是标准支持向量；③  $a_i = 0$ ,  $0 < a_i^* < C$  所对应的  $x_i$  是标准支持向量；④  $a_i = C$ ,  $a_i^* = 0$  所对应的  $x_i$  是边界支持向量；⑤  $a_i = 0$ ,  $a_i^* = C$  所对应的  $x_i$  是边界支持向量。除了标准支持向量(normal support vector, NSV), 其它情况对结果没有影响。

对支持向量机的结构可以归结为图 1。

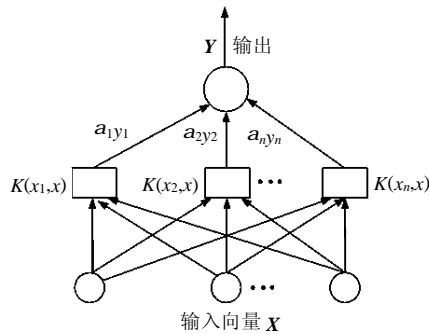


图 1 支持向量机结构示意图

Fig.1 Sketch map of support vector machines

在已知支持向量(support vector, SV)的情况下, 可以求出参数  $b$  :

$$b = \frac{1}{N_{NSV}} \left\{ \sum_{0 < a_i < C} [y_i - \sum_{x_j \in SV} (a_j - a_j^*) K(x_i, x_j) - e] + \sum_{0 < a_j^* < C} [y_i - \sum_{x_j \in SV} (a_j - a_j^*) K(x_i, x_j) - e] \right\} \quad (6)$$

式中  $N_{NSV}$  为标准支持向量机的个数。

得到了参数  $b$ , 则回归函数:

$$f(x) = \sum_{SV} (a_i - a_i^*) K(x_i, x_j) + b \quad (7)$$

### 1.3 支持向量机的参数选择

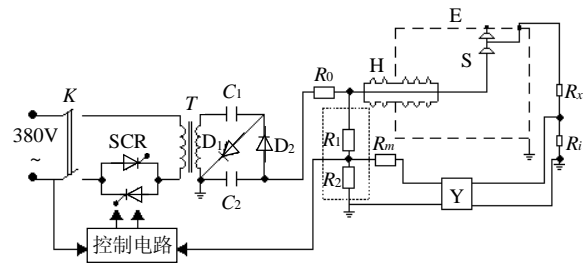
支持向量机的优点之一就是可以通过参数的选择达到控制解的目的。这些参数包括: 核函数、核参数、 $e$  不敏感损失函数、惩罚参数  $C$ 。核函数实现了支持向量机将输入数据映射到高维空间, 把非线性问题转化为线性问题, 提高了学习机器的计算能力<sup>[13]</sup>, 它的选择可以精确确定高维空间的结构, 以此来控制解的复杂性。惩罚参数  $C$  的引入是为了防止最优化问题无解的情况, 它的选择可以约束和限制奇异数据对结果的影响, 一般情况根据训练数据的大小而定, 不宜选得过小。对于  $e$  的选择, 一般说来,  $e$  值越大, 支持向量数目就越少, 因而解的表达就越稀疏。然而, 过大的  $e$  值也能降低数据点的逼近精度, 从这一意义上讲,  $e$  也是解的稀疏程度与数据点精度之间的平衡因子<sup>[11]</sup>。

## 2 基于支持向量机的绝缘子闪络电压预测

### 2.1 试验方法

试验是在低温低气压人工气候室内完成的。人工气候室内径 2.0m, 内长 3.8m, 室内最低气温可达  $-36^\circ\text{C}$ , 最低气压可达 34.6kPa, 满足绝缘子人工覆冰的要求。本试验采用可控硅双反馈直流电源, 额定电压 100kV, 在泄漏电流为 0.5A 时, 动态压降小于 3%, 符合 IEC 相关标准对试验电源的要求。

试品为 XZP-160 直流绝缘子。首先用蒸馏水洗净试品, 用一定电导率的覆冰水喷雾, 同时对 5 片绝缘子覆冰, 直到表面冰重达到预定值后停止喷雾并冷冻约 10min。分别对 5 片绝缘子进行一次正极性直流闪络试验, 闪络电压取 5 次试验的平均值。试验过程中保持室内温度在  $0^\circ\text{C}$  左右。通过改变气压、覆冰水电导率、表面冰重即可得到闪络电压与气压、覆冰水电导率、表面冰重之间的一组数据。试验原理如图 2 所示。



K—空气开关; T—变压器;  $C_1$ 、 $C_2$ —倍压电容;  $D_1$ 、 $D_2$ —高压硅堆; SCR—可控硅元件; F—分压器;  $R_1$ —分压器高压臂;  $R_2$ —分压器低压臂; H—穿墙套管;  $R_0$ 、 $R_m$ 、 $R_x$ —保护电阻;  $R_f$ —分流电阻; Y—示波器; S—试品; E—人工气候室

图 2 试验原理示意图

Fig.2 Test circuit

### 2.2 数据样本的归一化处理

数据样本来源于人工气候室对单片 XZP-160 绝缘子进行正极性闪络试验的试验数据, 通过改变绝缘子表面冰重、污秽程度(覆冰水电导率)、气压等参数模拟复杂环境进行直流闪络试验。对 20 个试验数据样本随机分为 15 个训练样本, 5 个验证样本。

由于作为训练样本输入的环境信息(气压、覆冰水电导率、冰重等参数)和训练的目标值(闪络电压)应该选取无量纲的向量, 而且为了进一步反应任意输入向量变化引起输出向量变化的大小, 所以把输入环境信息归一到  $[0, 1]$  区间<sup>[17]</sup>。

设  $f$  为某一环境信息,  $f_{\max}$  为此环境信息中最大的量值(可以是根据现场情况而得的估计值),  $f_{\min}$  为此环境信息中最小的量值, 那么对于任意的环境信息  $f(x)$ , 按以下方式进行归一化处理:

如果此环境信息与绝缘子的闪络电压正相关, 则归一化后的环境信息  $f_n(x)$  为

$$f_n(x) = \frac{f(x) - f_{\min}}{f_{\max} - f_{\min}} \quad (8)$$

如果此环境信息与绝缘子的闪络电压负相关, 则归一化后的环境信息  $f_n(x)$  为

$$f_n(x) = \frac{f_{\max} - f(x)}{f_{\max} - f_{\min}} \quad (9)$$

作为输入的环境信息中冰重、覆冰水电导率与绝缘子闪络电压负相关, 按式(8)进行归一化处理; 气压、闪络电压与绝缘子闪络电压属于正相关, 按式(9)进行归一化处理。根据实验室试验情况, 各环境信息具体的变化区间按  $([f_{\min}, f_{\max}])$  划分, 各参量的变化区间见表1。

表1 各参量的变化区间

Tab.1 The range of the each parameters

参量	闪络电压/kV	气压/kPa	冰重/kg	覆冰水电导率/( $\mu\text{S}/\text{cm}$ )
变化区间	[43, 93]	[70, 101.3]	[0, 2]	[200, 2 500]

## 2.3 数据样本的训练

对于数据的训练, 如果采用式(5), 由于该二次规划问题是双参数的, 为了更好利用 MATLAB 软件里的相关函数, 通过代换把双参数变为单参数。设:  $b_i = a_i - a_i^*$ , 代入式(5)和式(6), 由 KKT 条件<sup>[14]</sup>, 可以知道  $a_i a_i^* = 0$ , 则前面的训练问题、相应的参数和回归函数, 即式(5)~(7)相应变换为:

$$\max W(b) = \sum_{i=1}^n b_i y_i - e \sum_{i=1}^n (|b_i|) - \frac{1}{2} \sum_{i,j=1}^n b_i b_j K(x_i, x_j) \quad (10)$$

$$\text{s.t. } -C \leq b_i \leq C, \sum_{i=1}^n b_i = 0, i=1, \dots, n$$

$$b = \frac{1}{N_{\text{NSV}}} \left\{ \sum_{0 < b_i} [y_i - \sum_{x_j \in \text{SV}} b_j K(x_j, x_i) - e] + \sum_{0 > b_i} [y_i - \sum_{x_j \in \text{SV}} b_j K(x_j, x_i) + e] \right\} \quad (11)$$

$$f(x) = \sum_{\text{SV}} b_i K(x_i, x_j) + b \quad (12)$$

由于  $b_i = a_i - a_i^*$  的代入, 相应的支持向量的判断也要做一定的调整: ①  $b_i = 0$  所对应的  $x_i$  不是支持向量; ②  $0 < b_i < C$  所对应的  $x_i$  是标准支持向量; ③  $-C < b_i < 0$  所对应的  $x_i$  是标准支持向量; ④  $b_i = C$  所对应的  $x_i$  是边界支持向量; ⑤  $b_i = -C$  所对应的  $x_i$  是边界支持向量。

通过这样的代换后, 就可以利用 MATLAB 的 `fmincon` 函数编制相应的程序, 对数据进行训练后得到回归函数。

## 2.4 闪络电压预测

对于参数的选择, 核函数选择径向基核函数:

$K(x, x_i) = \exp\{-s \|x - x_i\|^2\}$ , 因为该核函数的应用范围是最广的, 它直观反映了两个数据的距离, 即分类问题, 而广义上讲, 回归问题就是一个多分类问题。经过参数选择不同值的对比分析, 核参数  $s$  选取 10, 惩罚参数  $C$  选取 100 比较合适, 由于训练样本的目标值(闪络电压的归一值)的取值是保留小数点后两位, 所以选择  $e = 0.01$  比较合适。这样通过对 15 个样本训练后, 得到 13 个支持向量, 相应的回归函数为:

$$f(x) = \sum_{\text{SV}} b_i K(x_i, x_j) + 0.4869 \quad (13)$$

这样在回归函数中输入相应的环境信息, 就可以得到预测的闪络电压。

表2给出5个验证样本预测结果与试验结果的比较。

表2 试验结果与预测结果对照

Tab.2 Test results and simulating results

样本号	试验结果/kV	SVM 预测结果/kV	相对误差/%
1	65	67.02	3.107
2	61	63.54	4.172
3	73	71.91	1.493
4	70	69.72	0.407
5	65	64.56	0.676

由表2可见, 试验结果与支持向量机计算出来的结果最大误差不超过 5%, 从对比的结果来看, 预测结果基本准确, 能满足工程的需要。

## 3 结论

本文应用支持向量机的回归模型, 用气压、污秽、覆冰3个变量作为预测模型的输入参数, 建立了预测绝缘子闪络电压的预测模型。结果表明, 预测的闪络电压与实际闪络电压基本吻合, 能满足工程的需要。

作为一种新兴技术, 支持向量机具有良好的学习性能和泛化能力, 不受维数、样本容量和非线性的限制。而复杂环境中外绝缘的选择也存在着影响因素多、关系复杂、试验数据有限等问题, 采用支持向量机研究复杂环境条件下的绝缘子闪络电压, 为输电线路外绝缘选择开辟了一条新的途径。

## 参考文献

- [1] 孙才新, 司马文霞, 舒立春. 大气环境与电气外绝缘[M]. 北京: 中国电力出版社, 2002.
- [2] 蒋兴良, 易辉. 输电线路覆冰及防护[M]. 北京: 中国电力出版社,

- 2002.
- [3] 孙才新, 舒立春, 蒋兴良. 高海拔、污秽、覆冰环境条件下超高压线路绝缘子交直流放电特性及闪络电压校正研究[J]. 中国电机工程学报, 2002, 22(11): 115-120.
- Sun Caixin, Shu Lichun, Jiang Xingliang. AC/DC flashover performance and its voltage correction of UHV insulators in high altitude and icing and pollution environments[J]. Proceedings of the CSEE, 2002, 22(11): 115-120(in Chinese).
- [4] 顾乐观, 孙才新. 电力系统的污秽绝缘[M]. 重庆: 重庆大学出版社, 1990.
- [5] 张纬钺, 何金良, 高玉明. 过电压防护及绝缘配合[M]. 北京: 清华大学出版社, 2002.
- [6] Shwehdi M H, Farag A S, Izzularab M A. Estimating the insulation strength of two series non-ceramic dielectrics on distribution systems: a statistical approach[C]. Electrical Insulation Conference, Rosemont, IL, 1997, 22(7): 799-804.
- [7] Shwehdi M H, Shahzad F. A novel method for predicting critical flashover (CFO) voltages insulation strength of multiple dielectrics on distribution overhead lines[C]. Conference Record of the 1996 IEEE International Symposium on Electrical Insulation, Montreal Que, 1996, 1(6): 316-319.
- [8] Almad A S, Ahmad H, Salam M A, et al. Regression technique for prediction of salt contamination severity on high voltage insulators [C]. Annual Report Conference on Electrical Insulation and Dielectric Phenomena, Victoria BC, 2000, 1(10): 218-221.
- [9] Ahmad A S, Ahmad H, Salam M A, et al. Prediction of salt contamination on high voltage insulators in rainy season using regression technique[C]. Proceedings of TENCON 2000, Kuala Lumpur, 2000, 3(9): 24-27.
- [10] Ahmad A S, Ghosh P S, Aljunid S A K, et al. Modeling of various meteorological effects on contamination level for suspension type of high voltage insulators using ANN[C]. IEEE/PES Transmission and Distribution Conference and Exhibition 2002: Asia Pacific, Yokohama Japan, 2002, 2(10): 1030-1035.
- [11] 李元诚, 方廷健, 于尔铿. 短期负荷预测的支持向量机方法研究[J]. 中国电机工程学报, 2003, 23(6): 55-59.
- Li Yuancheng, Fang Tingjian, Yu Erkeng. Study of Support vector machines for short-term load forecasting[J]. Proceedings of the CSEE, 2003, 23(6): 55-59.
- [12] 赵登福, 王蒙, 张讲社. 基于支撑向量机方法的短期负荷预测[J]. 中国电机工程学报, 2002, 22(4): 26-30.
- Zhao Dengfu, Wang Meng, Zhang Jianshe, et al. A support vector machine approach for short term load forecasting[J]. Proceedings of the CSEE, 2002, 22(4): 26-30(in Chinese).
- [13] Vladimir N. Vapnik. 统计学习理论的本质[M]. 北京: 清华大学出版社, 2004.
- [14] Nello C, Shawe-Taylor J. An introduction to support vector machines and other kernel-based learning methods[M]. 北京: 电子工业出版社, 2004.
- [15] 张恒喜, 郭基联, 朱家元, 等. 小样本多元数据分析方法和应用[M]. 西安: 西北工业大学出版社, 2002.
- [16] 杜树新, 吴铁军. 用于回归估计的支持向量机方法[J]. 系统仿真学报, 2003, 15(11): 1580-1585,1633.
- Du Shuxin, Wu Tiejun. Support vector machines for regression [J]. Journal of System Simulation, 2003, 15(11): 1580-1585,1633(in Chinese).
- [17] 杨庆, 司马文霞, 蒋兴良, 等. 复杂环境条件下绝缘子闪络电压预测神经网络模型的建立和应用[J]. 中国电机工程学报, 2005, 25(13): 155-159.
- Yang Qing, Sima Wenxia, Jiang Xingliang, et al. The buliding and application of a neural network model for forecasting the flashover voltage of the insulator in complex ambient conditions [J]. Proceedings of the CSEE, 2005, 25(13): 155-159(in Chinese).

收稿日期: 2006-04-05。

作者简介:

舒立春(1964—), 男, 博士, 教授, 博士生导师, 从事高电压与绝缘技术的研究工作, [lcshu@cqu.edu.cn](mailto:lcshu@cqu.edu.cn);

白困利(1974—), 男, 硕士研究生, 从事高电压与绝缘技术的研究工作;

胡 琴(1981—), 男, 博士研究生, 助教, 从事高电压与绝缘技术的研究工作。

(编辑 车德竞)