

文章编号:1005-0523(2018)02-0136-08

基于支持向量机的弓网间电弧诊断策略

刘仕兵,曾聿田,刘欢,马志方

(华东交通大学电气与自动化工程学院,江西 南昌 330013)

摘要:电力机车在运行的过程中,通过受电弓与接触网间的电接触而获取电能,在机车受流过程中弓网间电弧是较为常见的一个现象。弓网间的电弧不仅会对接触网与受电弓造成不同程度上的损伤,同时也会对电力机车上的电力设备造成一定的干扰。弓网间的电弧具有随机性,发生的原因也具有多样性,因此对电弧进行诊断具有一定难度。针对目前对弓网间电弧诊断存在的技术上的问题,文章基于支持向量机(support vector machine, SVM)来实现对弓网间电弧的诊断。在获取弓网间电流的原始数据后,通过计算出原始数据的功率谱熵,构造出电弧诊断所需的特征向量,应用支持向量机对这些特征向量进行分类,能实现对弓网间电弧电流的正常电流的正确区分。仿真结果表明,应用本文所建立起的诊断模型对样本进行诊断,结果正确率能达到90%以上,为弓网间电弧诊断提供了一个实用的方法和研究思路。

关键词:弓网系统;功率谱熵;支持向量机;弓网间电弧

中图分类号:U225

文献标志码:A

截止2015年底,我国铁路运营总里程突破12万km,其中高速铁路超过了2.3万km,电气化铁路更是达到了7万km以上,供电在铁路运输安全生产中的重要性进一步提高^[1]。在如今电气化铁路的牵引供电系统中,主要采用的是受电弓与接触网相接触的受流方式为电力机车供电。在铁路的牵引供电系统中,最容易出现故障的设备就是接触网。接触网一般没有备用,一旦弓网间出现故障,则会影响整个供电区间的正常运行。因此加强对弓网间的状态检测与故障诊断是保证电力机车正常运行的关键。随着我国电力机车朝着高速化、重载化发展,对弓网间电弧的检测与诊断是评价高速铁路弓网受流质量的重要方式。

电弧现象是一个电场、磁场、气流场及热场等多物理场之间相互耦合变化的复杂过程。电力机车在运行的过程中,一旦弓网间出现接触性能不稳定,或者发生分离,便会引起弓网间的燃弧现象。随着我国电力机车运行速度的不断提高,接触网振动、轨道不平顺等因素造成受电弓和接触网发生不良接触而引起的弓网间燃弧现象愈发频繁。弓网间电弧不仅会使车载电力设备承受高频振荡过电压外,轻者会烧蚀接触线,缩短接触线与受电弓的使用寿命,影响车载电力设备的正常运行,重者烧断接触线,造成列车停运、线路损坏等重大事故。在实际中,受现场运营条件、设备安装空间、技术手段不成熟等因素的限制,很难在现场对弓网电弧进行直接检测与诊断^[2]。

国内外对弓网间电弧检测的研究取得了一定的进展,检测方法主要为以下几种^[3]:在接触网附近安装高帧率相机,对弓网间电弧进行图像处理以达到实时检测,但这种方法的缺点在于需要处理的图像数据过于庞大,造成计算机负担过于沉重而影响效率;根据弓网离线时其电阻会变大的原理实现对电弧电流的检测,但这种方法的准确性较差;利用传感器对弓网间的接触压力与温度进行测量,而实现对电弧的诊断,但是这种方法容易受到外界其他因素的干扰。这些方法具有一定局限性,对电弧诊断的准确度也难以保证。本文将基于支持向量机建立起一个弓网间电弧诊断模型,通过对线路中电流信号的检测,以实现对弓网间电弧的诊断,并为弓网间电弧故障诊断提供一个新思路。

收稿日期:2017-09-20

基金项目:国家自然科学基金项目(11162006);江西省教育厅科技项目(GJJ150530);江西省教育厅科技项目(GJJ160488)

作者简介:刘仕兵(1970—),男,教授,主要研究方向为电气化铁路接触网技术。

1 原始数据获取与分析

1.1 弓网电流采集系统

采用如图 1 所示的电弧试验系统来对电流数据进行采集^[4]。电动机带动转盘转动,转盘带动接触线运动,使接触线与受电弓滑板之间有相对运动,对机车运行时弓网间作用进行了模拟;电流传感器能够测量出受电弓与接触线中的运行电流,通过数据采集卡采集后传送到计算机;同时高速摄像机能够记录下发生电弧时的图像与时间点;R、L 为用于模拟机车负载的电阻与电感。

本次实验模拟了 500 个弓网间电弧电流信号,再任取 500 个弓网正常状态下的电流信号,构成一份数量为 1000 的原始数据集。

1.2 原始数据分析

对采集到的线路电流进行初步分析,任取某一弓网正常状态的电流,其波形如图 2 所示。同时,任取某一弓网间发生电弧时的电流,其波形如图 3 所示。

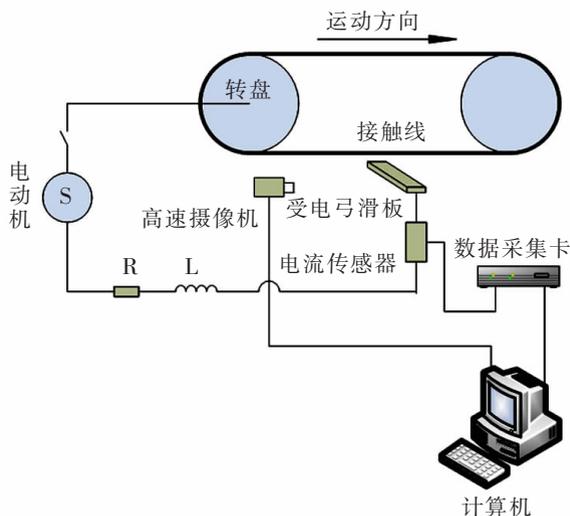


图 1 电流采集系统结构图

Fig.1 Structure graph of current acquisition system

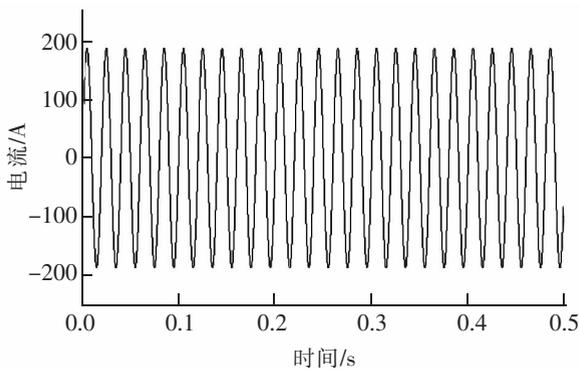


图 2 正常状态电流波形

Fig.2 Current waveform in normal state

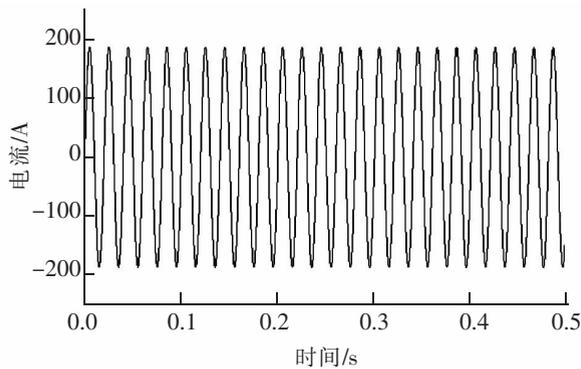


图 3 电弧状态电流波形

Fig.3 Current waveform in arc state

对比图 2 与图 3 可以看出,弓网间发生电弧时的电弧电流与正常运行状态下的电流在波形上并无明显的畸变,其峰值也没有明显的增大。因此仅通过电流波形无法判断线路中是否出现了电弧。

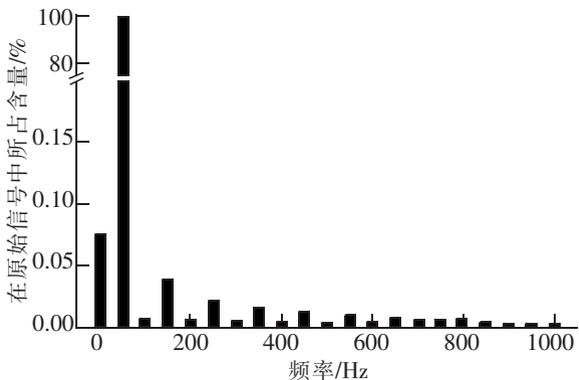


图 4 正常状态电流谐波含量

Fig.4 THD of current in normal state

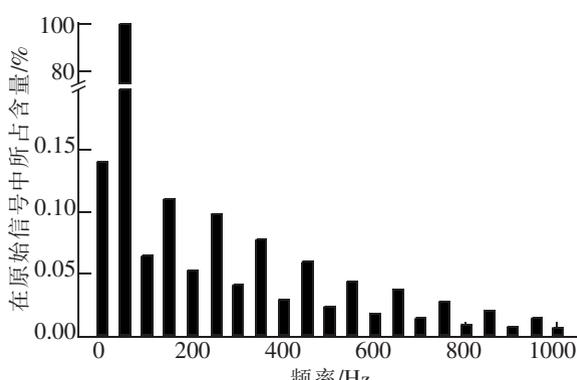


图 5 电弧状态电流谐波含量

Fig.5 THD of current in arc state

对这两种状态下的电流进行总谐波失真(THD)分析,结果如图4、图5所示。图中给出了频率在0~1 000 Hz范围内的整数次谐波含量,通过对比可以看出,弓网间产生电弧后电流中谐波含量相对于正常电流会略有增加,其总体显现出一定的混沌特征。

2 电流信号的特征提取

弓网间出现电弧现象有不同的原因,非线性因素对电流信号有不同的影响,当无法直观地对故障类型进行判断时,就需要提取出不同故障类型下的特征值,提高对故障识别、分类的准确度。

2.1 电流信号的混沌特性判断

从图4,图5中看出,机车运行时弓网间电流都存在一定的谐波,总体表现出一定的混沌特性。最大李雅普诺夫指数表示相空间相邻轨迹的平均指数发散率的数值特征,用于识别混沌运动的主要特征。计算一个信号最大李雅普诺夫指数,指数为正时表明该信号具备混沌特性^[5]。

现采用 wolf 算法对正常状态与电弧电流信号的最大李雅普诺夫指数^[6]进行计算,下表1为计算所得结果。从表中可以看出,这两种状态下的电流信号的指数值均为正,表面它们都具备一定混沌特性。

表1 两种状态下电流的最大李雅普诺夫指数

Tab.1 The maximum Lyapunov exponent of current in two states

电流状态	指数
正常状态	0.0143
电弧电流	0.0416

2.2 功率谱熵

信息熵最早是由美国数学家 Shannon 于 1948 年提出,其目的是为了解决对信息的量化度量问题。对于非线性的随机变量,信息熵可以给出某特征的定量表述。其基本定义为:假设 X 为一个可测集合 U 生成的 δ 代数 and 具有 μ 测度的勒贝格空间,且空间 X 可表示有界划分 $A=A_i$ 中互不相容集合的形式,对于该划分 A 的信息熵 $H(A)$ 可用下式表示:

$$H(A) = - \sum_{i=1}^n \mu(A_i) \lg \mu(A_i) \quad (1)$$

对于某个系统来说,信息熵是用来表示该系统内各个因素的混乱程度。

功率谱熵是通过对信号的频域特征进行提取而获得的。给定一个长度为 $N(N>0)$ 的离散信号 $x(i)$, 其中 $i=0,1,2,\dots,N-1$, 对其进行 Fourier 变换后得到频谱序列 $X(k)$:

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j \frac{2\pi}{N} nk} \quad (k=0,1,2,\dots,N-1) \quad (2)$$

该频谱序列在频率 k 处的功率谱值 $S(k)$ 如下:

$$S(k) = \frac{1}{N} |X(k)|^2 \quad (k=0,1,2,\dots,N-1) \quad (3)$$

用 p_i 表示各频率分量归一化后的功率谱密度 p_i 函数:

$$p_i = \frac{|S(k)|^2}{\sum_{k=0}^{N-1} |S(k)|^2} \quad (4)$$

结合信息熵的定义,可以写出相应功率谱熵 H 的表达式:

$$H = - \sum_{i=1}^n p_i \lg p_i \quad (5)$$

上式中, n 为对集合 $S(k)$ 的划分参数。

信息熵能够体现出一个系统中不确定因素的混乱程度^[7], 而功率谱熵作为信息熵在频域上的表征, 能够

反映出系统中谱型结构情况。电弧电流与正常电流在频谱上存在差异,因此功率谱熵值能在一定程度上将这种差异表现出来。

3 基于 SVM 的弓网间电弧诊断

3.1 SVM 简介

在机器学习领域中,支持向量机(SVM)是与学习算法有关的监督学习模型,它基于结构风险最小化原则和 VC 维理论等统计学原理,能应用于数据识别、非回归分析和监督分类等方面。支持向量机能有效解决样本小、非线性和高维识别等实际问题,其具备通用性广、鲁棒性好、相关理论完善、计算简单和有效性高等优点,并能有效地克服维数灾难、局部过优和过度学习等传统分类器具有的缺点^[8]。

支持向量机是通过对少量训练样本的特征子集,即支持向量进行训练,从而实现对整个样本进行分类。对于某样本集合: $(x_i, y_i), i=1, 2, 3, \dots, n$,其中 $x_i \in R^n, y_i \in \{-1, +1\}$,取该 n 维空间中的判别函数 $\varphi(x) = (\omega \cdot x) + b$,则该样本集中有一最优超平面 $(\omega \cdot x) + b = 0$ 能将这这些样本全部划分开来,分类原理如图 6 所示。

根据最优超平面的定义,支持向量与该超平面之间的距离为 $1/|\omega|$,因此寻求该超平面的问题可以转化为如下形式的二次规划问题:

$$\min \frac{1}{2} \omega^T \omega + C \sum_{i=1}^N \xi_i \quad \text{s.t.} \quad y_i (\omega^T \varphi(x_i) + b) \geq 1 - \xi_i \quad (6)$$

式中: ω 为权重向量; C 为惩罚因子且 $C \geq 0$; ξ_i 为松弛因子; b 为偏置项; N 为样本个数; $\varphi(x_i)$ 为非线性转换函数,其作用是将 x_i 映射到高位特征空间之中; y_i 用来表示分类结果,对分类结果 y_i 定义如下一个分类函数:

$$y_i = \text{sgn}(\omega^T \varphi(x_i) + b) \quad (7)$$

式(6)中目标函数与约束条件都是凸函数,根据最优化理论可知该函数能找到一个全局最优解。为解决该约束优化问题,现引入 Lagrange 算子 α_i ,将式(6)转化为无约束的目标函数:

$$L(\omega, \xi_i, \alpha_i, b) = \frac{1}{2} \omega^T \omega + C \sum_{i=1}^N \xi_i - \sum_{i=1}^N \alpha_i [y_i (\omega^T \varphi(x_i) + b) - 1 + \xi_i] \quad (8)$$

令式(8)中 ω, ξ_i, α_i 和 b 的偏导值等于 0,可得:

$$\omega = \sum_{i=1}^N \alpha_i y_i \varphi(x_i) \quad (9)$$

式(6)中的偏置项 b 可用下式计算:

$$b = \frac{1}{N_{SV}} \sum_{i \in SV} \omega^T \varphi(x_i) - y_i \quad (10)$$

其中 $SV = \{i: \alpha_i > 0\}$, N_{SV} 为样本中支持向量的个数,即 $\alpha_i > 0$ 时所对应的 x_i 个数。

观察式(7)与式(10),可以看出目标函数与偏置项中都存在数量积 $\omega^T \varphi(x_i)$,引入核函数 k 对其进行计算:

$$\omega^T \varphi(x_j) = \sum_{i=1}^N y_i \alpha_i \varphi(x_i)^T \varphi(x_j) = \sum_{i=1}^N y_i \alpha_i k(x_i, x_j) \quad (11)$$

通过引入核函数 k ,将该数量积进行非线性变换映射到高维空间中。因此,决策函数用下式表示:

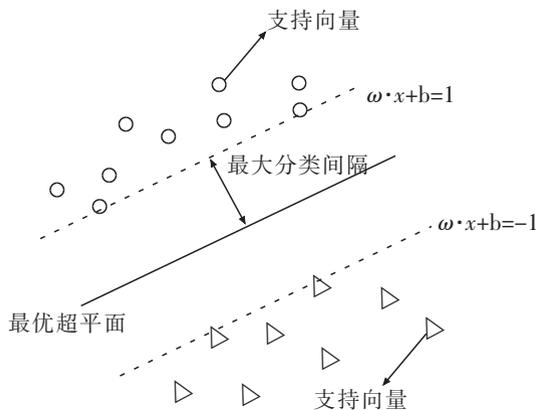


图 6 SVM 分类原理

Fig.6 Classification principle of SVM

$$y_i^* = \text{sgn}\left(\sum_{i=1}^N y_i \alpha_i k(x_i, x_j) + b\right) \quad (12)$$

支持向量机的核心就在于通过非线性映射将变量 x 映射到高维空间中,在这个空间中找到一个最优分类的超平面将变量进行完美划分。而核函数 $k(x_i, x_j)$ 能够避免在这个高维空间中进行复杂的计算,常用的核函数有线性内核,多项式内核,径向基内核(RBF),sigmoidal 核。由于 RBF 核适用于样本线性不可分且样本数量适中的情形,因此本次实验选取 RBF 核函数,其表达式如下:

$$k(x_i, x_j) = \exp(-g\|x_i - x_j\|^2) \quad (13)$$

式中: g 为核参数。

3.2 弓网间电弧诊断过程

获取的原始数据是连续时间信号,对原始数据进行采样后(采样频率为 5 kHz,采样点数为 800),构建一份数量为 1 000 的样本集合。对这 1 000 个样本进行特征提取,构造出弓网间电弧电流的特征向量 x_i (x_i 为计算所得的功率谱熵)。并用 y_i 来对样本进行相应的类别标记,属于电弧状态的样本标记 $y_i = -1$,正常状态的标记 $y_i = 1$ 。样本集合如表 2 所示。

表 2 样本集合
Tab.2 The sample database

	样本序号	功率谱熵值 x_i	类别划分 y_i
训练集	1	1.4723	1
	2	2.5416	-1
	3	1.1471	1

	600	2.6357	-1
测试集	601	1.4312	1

	1000	2.7129	-1

电弧电流频带比正常电流的频带要宽,因此其功率谱熵值大于正常电流。

在这 1 000 份样本种,前 600 份样本作为支持向量机的训练集,用于建立诊断模型;剩下 400 份样本作为测试集,用以测试本次实验所搭建起的诊断模型的诊断效果。

构建好训练集与测试集后,通过 MATLAB 软件中 LIB-SVM 工具箱,采用径向基(RBF)核函数对训练集进行建模,初步建立起隔离开关故障诊断模型。建模后,将测试集的样本导入到模型中进行结果测试,将诊断结果与真实值进行比较,计算出该模型的诊断准确率。若效果不理想,则需要对模型参数进行修改以达到满意结果。

影响模型诊断准确率的关键的因素在于式(13)中核参数 g 与惩罚因子 C 的选取。目前在国际上对 SVM 的参数优化选取并没有一个公认统一的办法。在实际中常用的方法就是通过网格搜索法^[9-10](Grid Search),使 g 与 C 的取值在一定范围内划分网格并对网格内所有点进行取值,对于所选定的 g 与 C 的值,通过交验证法(CV, cross validation)验证在此组 g 与 C 值下模型诊断的准确率,取能使对训练集分类准确度最高的一组 g 与 C 作为最终的模型参数。参数寻优的步骤如下:

1) 建立起网格坐标,设定网格搜索变量(g, C)的取值范围与搜索步长。其中 C 的范围设置为 $[2^{-10}, 2^{10}]$, g 的范围设置为 $[2^{-10}, 2^3]$,步长设置为 0.1。

2) 使用 K-fold 交叉验证法对训练集进行验证。将训练集分为 K 个子集($K \geq 3$ 以保证训练集数量大于

测试集),以其中任意一个子集作为测试集,剩下 $K-1$ 个子集作为训练集,对网格中的每一组 (g, C) 进行验证。使用训练集建立起诊断模型,然后用测试集对模型诊断效果进行测试,并计算出诊断结果的均方误差 (MSE, mean-squared-error),计算公式如下:

$$\sigma = \frac{\sum_{i=1}^n (y_i^* - y_i)^2}{n} \quad (14)$$

其中 n 为测试集中样本个数; y_i^* 为诊断结果; y_i 为实际值。

3) 将测试集更换为 K 个子集中的另一个子集,并将剩余的 $K-1$ 个子集做为训练集,直到将 K 个子集都用做为训练集进行效果测试后,取这 K 个均方误差的平均值作为该组参数 (g, C) 下的诊断准确率。

4) 重复以上步骤,找出使均方误差的平均值最小的一组 (g, C) ,即为该诊断模型下的最优参数组合。

基于网格划分法对参数进行交叉验证,能大大提高对参数寻优的效率与准确率,也能有效避免训练集中样本的随机性对模型性能的影响。

3.4 诊断结果分析

应用训练集中 600 个样本建立起诊断模型后,对测试集中 400 个样本进行分类诊断,将诊断结果 y_i^* 与测试集中的真实值 y_i 进行比较,验证该模型的诊断效果。分别进行了 3、5、7 折的交叉验证并对比 RBF 核函数与 Sigmoidal 核函数的诊断效果,结果如表 3 所示。模型诊断结果的准确率用式(15)计算。

$$P = 1 - \frac{\sum_{i=1}^n |y_i^* - y_i|}{2n} \times 100\% \quad (15)$$

式中, n 为测试集中样本个数。

表 3 诊断结果
Tab.3 Diagnosis results

核函数	CV 折数	准确率/%
RBF	3-fold CV	92.75
RBF	5-fold CV	96.75
RBF	7-fold CV	97.25
Sigmoidal	3-fold CV	85.75
Sigmoidal	5-fold CV	89.25
Sigmoidal	7-fold CV	90.50

从表 3 中可以看出,在诊断准确率上 RBF 核函数明显优于 Sigmoidal 核函数;同时交叉验证的折数越大其诊断准确度也越高。但交叉验证的折数越多诊断所耗费的时间也更多,7 折交叉验证的正确率相比 5 折略有提高,但在实验过程中耗时却大大增加,对于样本数量更巨大的测试集来说耗时也会更加可观。

弓网间出现电弧时,持续时间很短,因此对原始信号进行采样时,采样频率与采样长度的不同会在一定程度上影响对电弧信号表述。现对原始信号进行不同长度与频率的采样,对重新构造出的样本以 5-fold CV 进行诊断,诊断准确率如表 4,表 5 所示。

表4 不同采样点数所得诊断结果(采样频率=5 kHz)

Tab.4 The diagnosis results at different sampling number (sample frequency=5 KHz)

采样点数/m	准确率/%
600	94.00
800	96.75
1000	95.00
1 200	94.75

表5 RBF 不同采样频率所得诊断结果(采样点数=800)

Tab.5 The diagnosis results at different sample frequencies (sampling number=800)

采样频率/kHz	准确率/%
5	96.75
10	96.25
20	96.50

从表4,表5中可以看出,采样点数对诊断结果有一定影响,但由于样本集合不大,对断准确率影响不是很大;而不同的采样频率对诊断结果的准确率影响非常小。

4 结论

本文基于支持向量机建立起数学模型,通过对线路中的电流信号的检测来判断弓网间是否出现电弧现象,为弓网间电弧诊断提供一个新思路。通过实验,可以得到以下几点结论:

1) 弓网间出现电弧时,电弧电流与正常电流波形相比并无明显的畸变,无法直观对它们进行分辨。但电弧电流中混有很多高频噪声,具有一定的混沌特性。针对电弧电流的这种混沌特性可以通过计算其功率谱熵而进行特征提取,在本次实验中这种方法能较好地地区分电弧状态电流与正常状态电流。

2) 本文所建立的模型准确率较高,但仍存在一定的误差。误差原因主要来自于模型参数、样本数量,以及某些电弧电流与正常状态电流在频域与时域上的区别较小,样本的随机性会对模型的诊断性能造成一定的影响。此外,本文仅采集了弓网间电流,数据维度较为单一,但诊断结果较为理想;对原始数据进行处理时,可考虑采用电流加电压的方式构建特征向量提高数据维度,以实现更为准确的诊断结果。

3) 应用支持向量机建立起的电弧诊断模型对本次实验所收集到的数据诊断正确率比较理想,总体能正确率能达到90%以上。证明了该方法对电弧诊断的可行性与有效性。当给定一个适当的阈值对普通电弧电流与故障电弧电流进行区分,使用该模型就能有效实现电弧故障的诊断,为高速铁路弓网间的故障诊断提供了一个新的思路。

参考文献:

- [1] 刘仕兵,朱雪龙,武磊,等. 基于灰色聚类和组合赋权法的高速铁路接触网健康状态评估研究[J]. 铁道学报,2016,38(7):57-63.
- [2] 王万岗,吴广宁,高国. 高速铁路弓网电弧试验系统[J]. 铁道学报,2012,34(4):22-27.
- [3] 王英,刘志刚,范福强,等. 弓网电弧模型及其电气特性的研究进展[J]. 铁道学报,2013,35(8):35-43.
- [4] 王万岗,吴广宁,高国强,等. 高速铁路弓网电弧试验系统[J]. 铁道学报,2012,34(4):22-27.
- [5] 刘仕兵,葛俊祥. 基于 K-means 聚类法的牵引供电隔离开关故障状态监测[J]. 华东交通大学学报,2017,34(3):109-117.
- [6] 蔡聪波. 混沌系统最大 Lyapunov 指数的计算及其在脑电数据分析中的应用[D]. 厦门:厦门大学,2001.
- [7] QU LIANGSHENG,LI LIANGMING,LEE JAY. Enhanced diagnostic certainty using information entropy theory[J]. Advanced Engineering Informatics,2003,17(3-4):141-150.
- [8] 白鹏,张喜斌. 支持向量机理论及工程实例[M]. 西安:西安电子科技大学出版社,2008.
- [9] 张剑,王波. Grid-Search 和 PSO 优化的 SVM 在 Shibor 回归预测中的应用研究[J]. 经济数学,2017,34(2):84-88.
- [10] 王兴玲,李占斌. 基于网格搜索的支持向量机核函数参数的确定[J]. 中国海洋大学学报,2005,35(5):859-862.

Diagnosis Strategy for Arc State Between Catenary and Pantograph Based on Support Vector Machine

Liu Shibing, Zeng Yutian, Liu Huan, Ma Zhifang

(School of Electrical and Automation Engineering, East China Jiaotong University, Nanchang 330013, China)

Abstract: The electric locomotive obtains the electric energy through the electrical contact between pantograph and catenary. And the arc fault between the pantograph and catenary is a common phenomenon when the electric locomotive is running. The arc fault in pantograph–catenary system will not only damage the catenary and the pantograph, but also disturb the operation of the electric equipment in electric locomotive. The arc fault has the characteristic of randomness and irregularity, so it is difficult to diagnose the fault timely and accurately. According to the existing problems in arc fault diagnosis, this paper introduced a new diagnosis method based on the support vector machine(SVM). After obtaining the raw data of the current in pantograph–catenary system, the power spectrum entropy was adopted to extract the feature vectors needed by arc fault diagnosis and the SVM was used to classify these feature vectors. Then, the normal current from arc fault can be recognized. The research results show that the diagnosis model established in this paper has a high accuracy in classifying the fault state and normal state, which provides an useful method and research approach for arc fault diagnosis.

Key words: pantograph–catenary system; power spectrum entropy; SVM; electric arc