

DOI: 10.19783/j.cnki.pspc.231036

# 基于 IF-AD-ELM 的特高压直流输电系统故障辨识

杨新宇, 赵庆生, 韩肖清, 梁定康, 王旭平

(太原理工大学电力系统运行与控制山西省重点实验室, 山西 太原 030024)

**摘要:** 针对现有的特高压直流(ultra high voltage direct current, UHVDC)输电系统故障检测方法灵敏度低、难以识别高阻接地故障的问题, 提出了一种基于整数因子(integer factor, IF)-近似导数(approximate derivative, AD)和极限学习机(extreme learning machine, ELM)的特高压直流输电系统故障辨识方法。其中整数因子用于分析不同采样频率下的信号, 近似导数法用于获得信号不同程度的细节系数。首先, 基于不同的整数因子对信号进行下采样, 并利用近似导数法对所得信号求一阶、二阶和三阶近似导数。其次, 分别计算各个子信号的熵特征。然后, 用基于交叉验证的递归特征消除(recursive feature elimination with cross validation, RFECV)算法对得到的一系列特征进行特征筛选, 并结合 ELM 对特高压直流输电系统进行故障辨识。最后, 在 Matlab/Simulink 环境中搭建了 $\pm 800$  kV 的 UHVDC 系统模型, 模拟不同故障类型。实验结果表明, 所提方法在识别特高压直流输电系统不同类型故障时有更高的准确率, 且耐受过渡电阻能力强。

**关键词:** 特高压直流; 下采样; 特征选择; 极限学习机; 故障辨识

## Fault identification of a UHVDC transmission system based on IF-AD-ELM

YANG Xinyu, ZHAO Qingsheng, HAN Xiaoqing, LIANG Dingkan, WANG Xuping

(Shanxi Key Laboratory of Power System Operation and Control, Taiyuan University of Technology, Taiyuan 030024, China)

**Abstract:** There is a problem of low sensitivity and difficulty in identifying high-resistance ground faults in existing fault detection methods for ultra high voltage direct current (UHVDC) transmission system. Thus a fault identification method for a UHVDC transmission system based on the integer factor (IF)-approximate derivative (AD) and an extreme learning machine (ELM) is proposed. The IF is used to analyze the signals at different sampling frequencies, and the AD method is used to obtain different degrees of detail coefficients for the signals. First, the signal is down-sampled based on different IFs, and the AD method is used to calculate the first, second and third order approximate derivatives of the obtained signal. Secondly, the entropy characteristics of each sub-signal are calculated. Then, recursive feature elimination with a cross validation (RFECV) algorithm is used to screen the features of the obtained series of features, and the ELM is used to identify the UHVDC transmission system fault types. Finally, the UHVDC system model of  $\pm 800$  kV is built in the Matlab/Simulink environment to simulate different fault types. The experimental results show that the proposed method has higher accuracy and strong tolerance to transition resistance when identifying different types of faults in a UHVDC transmission system.

This work is supported by the National Natural Science Foundation of China (No. 51777132).

**Key words:** UHVDC; down-sampling; feature selection; ELM; fault identification

## 0 引言

特高压直流输电技术有着传输容量大、输送距离远、控制灵活等特点, 在远距离输电领域应用广

泛<sup>[1-3]</sup>。近年来, 由于电力电子技术的不断发展, 特高压直流输电的稳定性得到进一步提升。但由于输电线路较长、地理环境复杂和运行环境恶劣, 电力传输系统容易发生故障, 而这些故障会严重影响电力传输的效率。因此, 为了确保整个输电系统的稳定运行, 快速准确地实现故障识别显得尤为重要。目前, 故障辨识主要基于故障时电气量的差异, 而人工智能方法应用较少。

**基金项目:** 国家自然科学基金项目资助(51777132); 国家自然科学基金青年基金项目资助(51907138); 国网山西省电力公司科技项目资助(520510220002)

基于边界元件的阻抗幅频特性,单端特定频率信号可用于建立区内外故障的判断标准,但该方法会受到控制系统触发角和换相重叠角的影响<sup>[4]</sup>。在文献[5]中,当输电线路发生故障时,观察两端无功功率的极性差异可以判断区内外故障。文献[6]考虑了高压直流输电系统的谐波等效电路、控制策略和直流滤波器的幅频特性,利用区内外故障时两端特征谐波电流的极性差异实现故障辨识。

随着人工智能的发展,许多学者开始将机器学习应用于电力系统故障检测<sup>[7-8]</sup>。文献[9]利用主成分分析法提取直流电压和电流的特征,然后利用支持向量机(support vector machine, SVM)对这些特征进行训练和测试,以实现多端高压直流输电系统区内外故障识别。文献[10]提出了一种基于多分辨奇异谱熵和 SVM 相结合的方法,该方法能够利用少量数据来识别线路故障,并准确识别区内外故障,但不能进行故障选极。文献[11]提出了一种基于 K-means 的故障识别方法,利用高压直流输电系统逆变器的电压和电流信息来识别故障。然而,该方法可能不适用于其他输电系统,且难以识别高阻接地故障。虽然上述机器学习方法在特定情况下可能表现良好,但当系统结构发生变化时,其效果可能会受到限制。

近年来,深度学习正在快速发展,并在图像识别、语音识别以及自然语言处理等领域取得了显著的进展<sup>[12]</sup>,大量学者将其应用于故障识别领域。文献[13]采用经验模态分解(empirical mode decomposition, EMD)分析整流侧和逆变侧的故障信号,然后将这些处理过的信号用于卷积神经网络(convolutional neural network, CNN)训练,利用 CNN 对故障进行精确定位。文献[14]提出了一种改进的自适应噪声完备集合经验模态分解(modified complete ensemble empirical mode decomposition with adaptive noise, MCEEMDAN)算法,可以将信号分解为一系列本征模态函数。通过对本征模态函数进行伪彩色编码,将原始信号转换为二维图像,并利用条件生成对抗网络产生大量新样本实现数据增广,最后利用卷积神经网络实现故障识别。文献[15]在高压直流输电系统中利用离散小波变换(discrete wavelet transformation, DWT)处理直流电流,最后用人工神经网络(artificial neural network, ANN)实现故障辨识。但上述信号处理算法较为复杂,不利于快速识别故障。相比于机器学习算法,深度学习可以快速、准确地提取隐藏在海量数据中的深层抽象特征,但辨识结果比较依赖神经网络的结构参数。针对不同的电力系统场景,神经网络的调优和训练可能耗时较长。

为了解决上述问题,本文提出了一种基于整数因子(integer factor, IF)-近似导数(approximate derivative, AD)和极限学习机(extreme learning machine, ELM)的特高压直流输电系统故障辨识方法。其中, IF 用于分析不同采样频率下的信号,而 AD 法用于获得信号不同程度的细节系数。然后计算所得信号的 9 个熵特征,使用基于交叉验证的递归特征消除(recursive feature elimination with cross validation, RFECV)算法进行特征选择,最后利用 ELM 实现故障辨识。结果表明,本文所提方法在高阻接地时能准确识别 UHVDC 故障类型。

## 1 数据处理

### 1.1 整数因子

信号的采样频率可以表示为

$$f_s = N/T \quad (1)$$

式中:  $f_s$  为信号的采样频率;  $N$  为数据长度;  $T$  为信号的采样时间。

基于整数因子的信号下采样是信号处理中的基本方法,其作用是降低信号的采样频率。当 IF 为  $n$  时,通过保留信号中第一个样本以及第一个样本后的每隔  $n$  个本来降低原信号的采样率。

假设有信号  $L = [L(1), L(2), L(3), \dots, L(12)]$ , 信号持续时长为 0.01 s。

当  $n=1$  时,有

$$L_1 = [L(1), L(2), L(3), \dots, L(12)] \quad (2)$$

此时信号的长度为 12。

当  $n=2$  时,有

$$L_2 = [L(1), L(3), \dots, L(9), L(11)] \quad (3)$$

此时信号的长度为 6。

当  $n=4$  时,有

$$L_4 = [L(1), L(5), L(9)] \quad (4)$$

此时信号长度为 3。

由此可知,不同的 IF 值,如“1, 2, 4, ...,  $m$ ”,可以得到不同采样频率的信号。本文中不同 IF 值下对应的采样频率如表 1 所示。

表 1 基于 IF 的采样频率

IF	数据长度	采样频率/kHz
1	1280	12.8
2	640	6.4
4	320	3.2
8	160	1.6
16	80	0.8
32	40	0.4
64	20	0.2
128	10	0.1

由上文分析可知, 当 IF 为 1 时, 所得信号与原始信号相同。本文以特高压直流(ultra high voltage direct current, UHVDC)输电系统正极经  $200\ \Omega$  过渡电阻接地时在整流侧所测的电压信号为例进行分析。如图 1(a)所示, 电压信号的持续时间为 0.1 s,

信号采样频率为 12.8 kHz, 包含 1280 个数据点。如图 1 所示, 基于 IF 的变化, 信号的数据长度和采样频率各不相同。每种情况下的采样频率按式(1)计算。本文中 IF 值分别取 1、2、4、8、16、32、64、128。因此, 故障电压信号的采样频率在 0.1~12.8 kHz 之间。

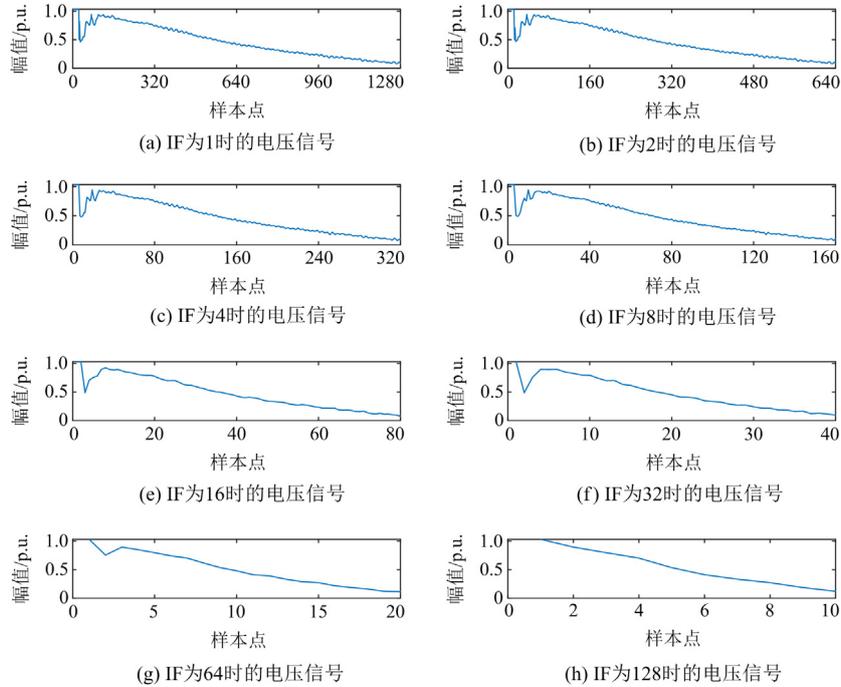


图 1 不同 IF 值下的电压信号

Fig. 1 Voltage signals under different IF values

## 1.2 近似导数

近似导数是一种简单的信号处理算法<sup>[16]</sup>。设向量  $\mathbf{F}$  的长度为  $k$ ,  $\mathbf{F} = [F(1), F(2), \dots, F(k)]$ 。用近似导数法计算  $\mathbf{F}$  的相邻元素之差。

设  $\mathbf{F}'$  是  $\mathbf{F}$  的一阶近似导数, 可以将  $\mathbf{F}'$  写为

$$\mathbf{F}' = d(\mathbf{F}) \quad (5)$$

若令  $F(2) - F(1) = T(1)$ ,  $F(3) - F(2) = T(2)$ ,  $\dots$ ,  $F(k) - F(k-1) = T(k-1)$ , 则

$$\mathbf{F}' = [T(1), T(2), T(3), \dots, T(k-1)] \quad (6)$$

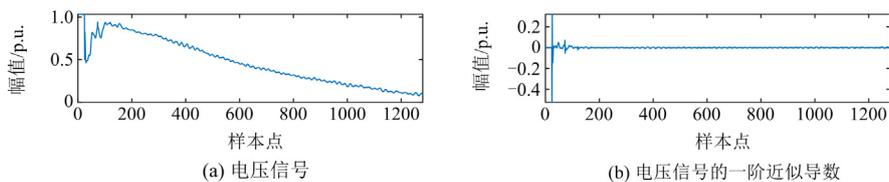
$$\mathbf{F}'' = [T(2) - T(1), \dots, T(k-1) - T(k-2)] \quad (7)$$

$$\mathbf{F}''' = (\mathbf{F}'')' \quad (8)$$

式中:  $\mathbf{F}$  的维度为  $k$ ;  $\mathbf{F}'$  的维度为  $k-1$ ;  $\mathbf{F}''$  的维度为  $k-2$ 。以此类推, 向量的维度随着近似导数阶数增加而减小。

在 AD 法中, 可以确定信号不同程度的细节系数。以图 1(a)中的电压信号为例, 采用 AD 法处理故障电压信号。如图 2 所示, 可以获得电压信号的一阶、二阶、三阶、四阶和五阶近似导数。在求得三阶近似导数后, 得到了类似的图形。对于不同的 UHVDC 系统故障类型可以得到类似的结果。因此, 本文选取 UHVDC 系统故障期间的电压信号以及电压信号的一阶、二阶、三阶近似导数为研究对象, 确定故障辨识特征。

将基于 IF 的下采样法与 AD 法相结合, 可以将一个信号分解为多个信号进行分析, 从而得到更多的分类信息。如图 3 所示, 根据不同的 IF 值(1、2、4、8、16、32、64、128), 将电压信号分解为 8 个信号。并分别对每个 IF 值下的信号求一阶、二阶和三阶近似导数, 共得到 32 个信号 ( $8 \times 4$ )。



(a) 电压信号

(b) 电压信号的一阶近似导数

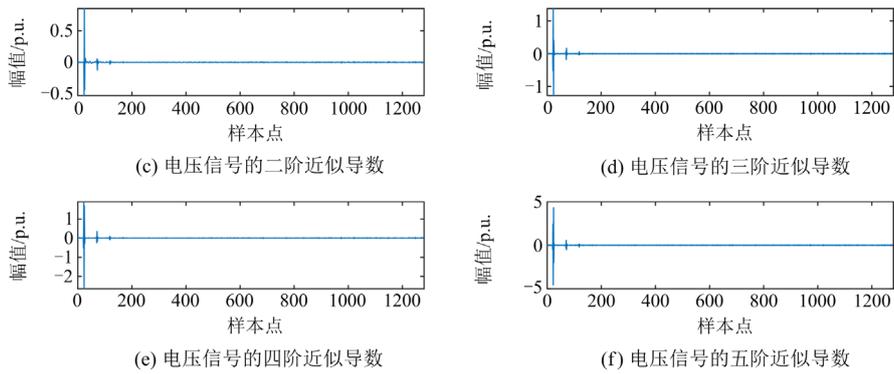


图 2 电压信号的各阶近似导数

Fig. 2 Each order approximate derivative of voltage signal

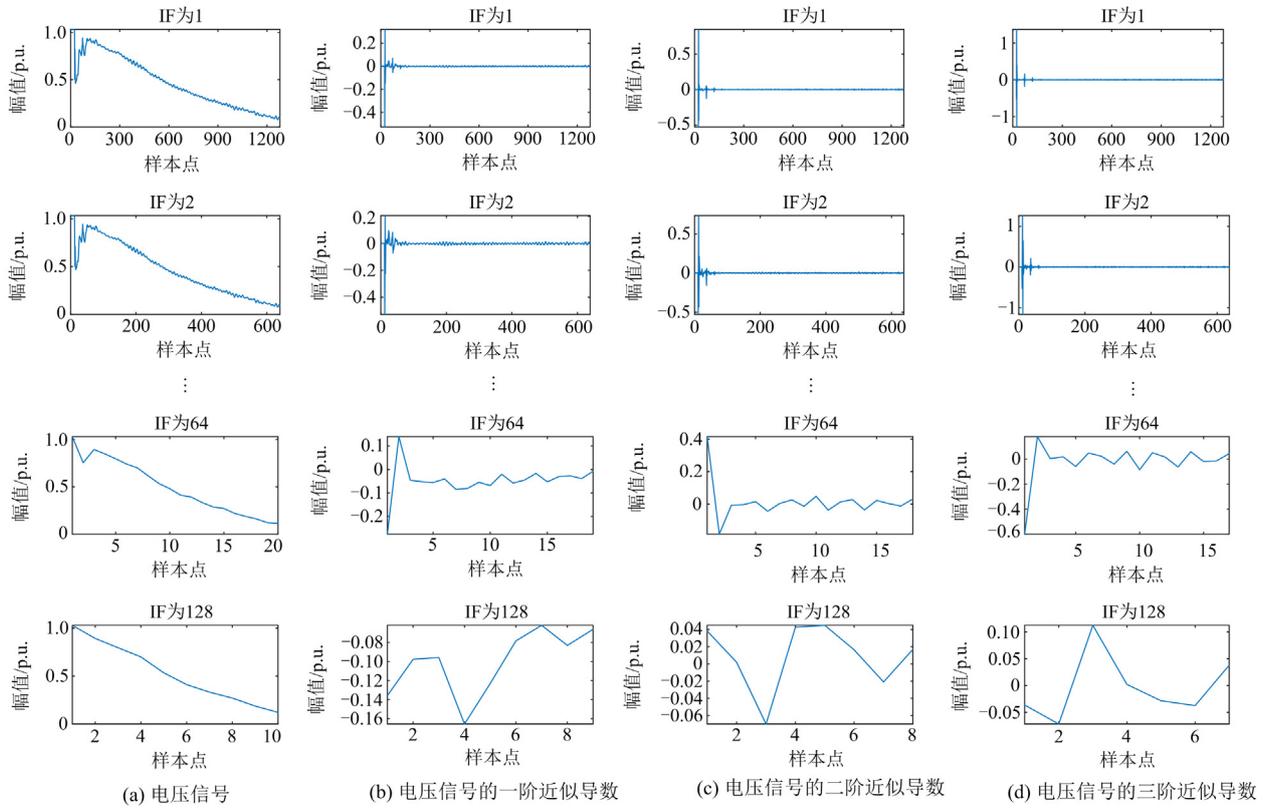


图 3 电压信号经 IF-AD 处理得到的多个信号

Fig. 3 Multiple signals obtained by IF-AD processing of voltage signal

## 2 辨识方法

### 2.1 特征提取

原始信号中含有大量干扰信息,无法为分类器提供优质的辨识特征。而特征提取的主要目的是通过计算不同的特征为分类器提供更多的分类信息。根据不同 IF 值和近似导数阶数可将 1 个故障电压信号分解为 32 个子信号。基于 32 个子信号,考虑 9 种熵特征,包括能量、香农熵、对数能量、标准差、平均幅

值、平均值、偏度、峰值和最小值,可得到 288(32×9) 个分类特征。表 2 给出了本文使用的不同熵特征公式,其中  $\mu$  为信号平均值,  $\sigma$  为信号的标准差。

设有信号  $Q=[Q_1, Q_2, \dots, Q_m]$ , 其中  $Q_i$  为信号的第  $i$  个样本点,  $1 \leq i \leq m$ ,  $m$  表示样本点总数。利用表 2 中的特征公式,提取了 288 个特征用于特高压直流输电系统故障类型的识别。然而,某些特征可能为冗余特征。这些特征将降低模型的性能,采用 RFECV 特征选择算法可以获得优质的分类特征。

表 2 熵特征  
Table 2 Entropy feature

序号	特征名称	公式
1	能量	$E = \sum_{i=1}^m  Q_i ^2$
2	香农熵	$S_E = -\sum_{i=1}^m Q_i^2 \log(Q_i^2 \log(Q_i^2))$
3	对数能量	$L_E = \log \sum_{i=1}^m Q_i^2$
4	标准差	$\sigma = \left(\frac{1}{m} \sum_{i=1}^m (Q_i - \mu)^2\right)^{0.5}$
5	平均幅值	$N_P = \frac{1}{m} \sum_{i=1}^m  Q_i $
6	平均值	$\mu = \frac{1}{m} \sum_{i=1}^m Q_i$
7	偏度	$S_k = \frac{1}{m} \sum_{i=1}^m [(Q_i - \mu) / \sigma]^3$
8	峰值	$\text{Max}( Q_i )$
9	最小值	$\text{Min}(Q_i)$

## 2.2 特征选择

输入数据的维度过高会导致故障识别的时间成本过大。因此,有必要减少输入数据的维度,以提高识别速度。

RFECV 是一种特征选择算法,它可用于从高维数据中选择最佳的特征子集,从而提高模型的预测能力。RFECV 算法是递归特征消除(recursive feature elimination, RFE)算法和交叉验证(cross validation, CV)算法的结合。其基本思想是先使用 RFE 算法对特征进行逐步选择,然后使用 CV 算法验证,以选出最优特征子集<sup>[17]</sup>。具体地, RFE 算法需要一个外部估计器来为每个特征分配权重。本文选取 SVM 为估计器,并计算每个特征子集的准确率。RFECV 算法首先使用 RFE 算法对所有特征进行排序,然后从排名最高的特征开始,逐步剔除一定数量的特征,直到特征子集中特征数量为零。在每个剔除特征的步骤中, RFE 算法对特征子集进行评估,并记录每个子集的得分。本文选择交叉验证的次数为 5,根据得分排名,可以筛选出最优的特征子集<sup>[18]</sup>。

RFECV 算法的步骤如下。

步骤 1: 将数据集随机等分成 5 份, 4 份用于模型训练, 1 份用于模型测试。

步骤 2: 将 RFE 算法和 SVM 结合计算特征重要性并排序。

步骤 3: 开始阶段包含所有特征, 依次删除一个最不重要的特征, 组成新的特征子集。将其作为 SVM 的输入, 并计算新特征子集的准确率, 直到特征子集中特征数量为零。

步骤 4: 转至步骤 1, 更换训练集和测试集进行交叉验证。最后选择准确率最高的特征子集为最优特征子集。

## 2.3 极限学习机

极限学习机是一种单隐层前馈神经网络, 它可以随机初始化模型的权重与偏置<sup>[19]</sup>。这种随机初始化权重和偏置的方法使得 ELM 的训练速度非常快, 因为它不需要像传统的神经网络一样通过迭代训练来更新权重和偏置。并且在大多数情况下, 它的泛化性能比传统的神经网络更好。

ELM 的另一个重要特点是它可以处理高维数据, 因为它只需要初始化输入层的权重和偏置, 而不需要迭代训练整个网络。这使得 ELM 非常适合处理大规模数据集和高维数据, 其结构图如图 4 所示。

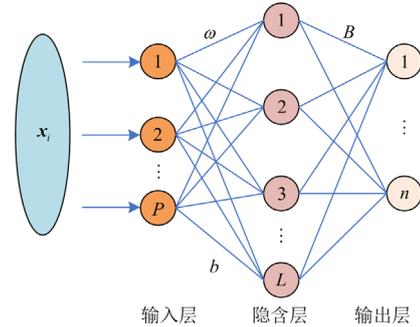


图 4 ELM 结构图

Fig. 4 ELM structure diagram

假设共有  $m$  个样本  $(x_i, y_i)$ ,  $1 \leq i \leq m$ , 其中  $x_i$  为第  $i$  个样本的输入向量,  $y_i$  为  $x_i$  的期望输出。设输入层维度为  $P$ , 隐含层维度为  $L$ , 输出层维度为  $n$ 。令  $\omega_j$  为输入层连接隐含层的权重,  $\omega_j = [\omega_{j1}, \omega_{j2}, \dots, \omega_{jP}]^T$ ,  $1 \leq j \leq L$ ;  $B_j$  为隐含层连接输出层的权重,  $B_j = [B_{j1}, B_{j2}, \dots, B_{jn}]^T$ ;  $b$  为隐含层阈值,  $b = [b_1, b_2, \dots, b_L]^T$ ;  $f(\omega, x, b)$  为激活函数。则输出函数为

$$G(x_i) = \sum_{j=1}^L [B_j f(\omega_j \cdot x_i + b_j)] \quad (9)$$

记  $H$  为网络隐含层输出矩阵,  $Y$  为给定数据的期望输出矩阵, 则可将式(9)简化为式(10)。

$$HB = Y, Y = [y_1, y_2, \dots, y_m] \quad (10)$$

$$H = \begin{pmatrix} f(\omega_1 \cdot x_1 + b_1) & \cdots & f(\omega_L \cdot x_1 + b_L) \\ \vdots & & \vdots \\ f(\omega_1 \cdot x_m + b_1) & \cdots & f(\omega_L \cdot x_m + b_L) \end{pmatrix}_{m \times L} \quad (11)$$

记  $H$  的 Moore-Penrose 广义逆为  $H^+$ , 则  $B$  的解为:  $B = H^+ Y$ 。

## 3 基于 IF-AD-ELM 的特高压直流输电故障辨识

本文的 UHVDC 故障辨识流程如图 5 所示。

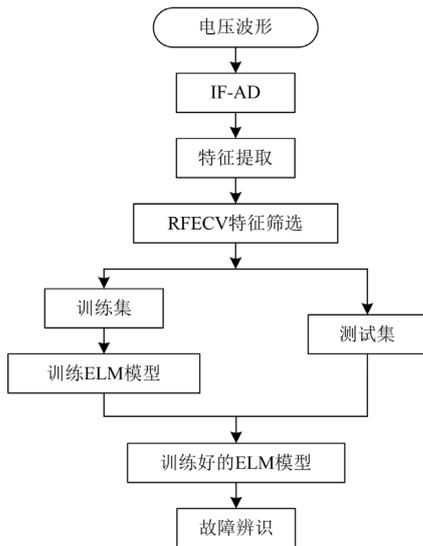


图 5 故障辨识流程图

Fig. 5 Fault identification flow chart

步骤 1: 将故障电压波形用 IF-AD 算法进行处理, 利用不同 IF 值获得 8 个不同采样频率的信号。对不同 IF 值下的信号分别求一阶、二阶、三阶近似导数, 共得到 32 个子信号。

步骤 2: 对 32 个子信号分别求表 2 中的 9 种熵特征, 则一个原始电压信号可获得 288 个特征。

步骤 3: 应用 RFECV 算法对获得的数据集求最佳特征子集。

步骤 4: 将获得的最佳特征子集按 7:3 随机分成训练集和测试集, 分别用于训练和测试 ELM 模型。

步骤 5: 输出故障辨识结果。

## 4 仿真实验

### 4.1 模型参数

在 Matlab/Simulink 中搭建了±800 kV 的 UHVDC 模型进行仿真分析, 如图 6 所示。其额定功率为 8000 MVA, 额定电压为 800 kV, 额定频率为 50 Hz, 输电线路全长为 1095.7 km。换流器采用的是 12 脉波桥式电路, 整流侧和逆变侧的交流滤波器均分为 4 大组, 每组由 4 小组滤波器组成。

### 4.2 故障实验

本文共设计了 7 种故障。F1 为整流侧交流系统单相接地故障, F2 为逆变侧交流系统单相接地故障, F3 为正极整流侧出口接地故障, F4 为正极逆变侧出口接地故障, F5 为正极直流输电线路区内接地故障, F6 为负极直流输电线路区内接地故障, F7 为两极短路故障。本文的故障仿真时间为 0.7~0.8 s, 信号采样频率为 12.8 kHz, 故障信号持续时长为 0.1 s。

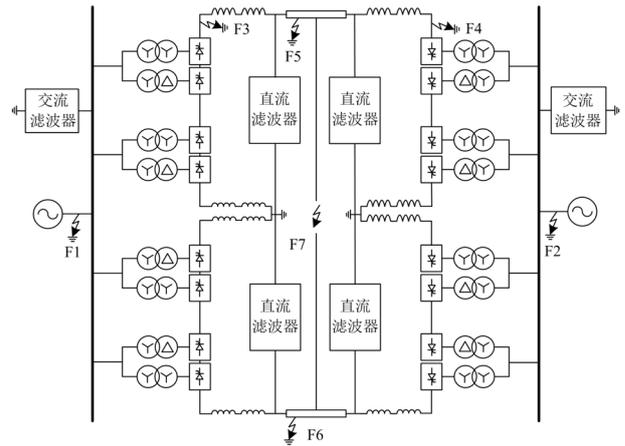


图 6 UHVDC 模型示意图

Fig. 6 UHVDC schematic diagram

对于 F1、F2、F3 和 F4 这 4 种故障, 通过改变过渡电阻的大小来获得不同的样本。对于这 4 种故障本文选取的过渡电阻大小为 5~1100 Ω。前 100 个样本: 过渡电阻为 5~500 Ω, 步长为 5 Ω; 后 50 个样本: 过渡电阻为 512~1100 Ω, 步长为 12 Ω。对于 F5、F6 和 F7, 选择的过渡电阻为 0.01 Ω、50 Ω、100 Ω、200 Ω、350 Ω、500 Ω、650 Ω、800 Ω、950 Ω、1100 Ω; 距离整流侧的距离为 70~1050 km, 步长为 70 km。本文共设计了 7 种故障实验, 每种实验 150 个样本, 共 150×7=1050 个样本。每种故障模拟方法如表 3 所示。

表 3 仿真设置

Table 3 Simulation settings

类型	数量	模型参数设置
F1	150	过渡电阻 前 100 个样本: 过渡电阻为 5~500 Ω, 步长为 5 Ω; 后 50 个样本: 过渡电阻为 512~1100 Ω, 步长为 12 Ω。
F2	150	
F3	150	
F4	150	
F5	150	过渡电阻、故障位置 过渡电阻: 0.01 Ω、50 Ω、100 Ω、200 Ω、350 Ω、500 Ω、650 Ω、800 Ω、950 Ω、1100 Ω; 距离整流侧: 70~1050 km, 步长为 70 km。
F6	150	
F7	150	

## 5 结果与讨论

### 5.1 IF-AD 方法分析

以图 1(a)中的电压信号为例, 不同的 IF 和 AD 值将得到不同的子信号。由图 3 可知, 一个电压信号经 IF-AD 处理后可得到 32 个不同的子信号。以熵特征中的能量特征为例, 不同故障情况下的正极整流侧直流电压信号经 IF-AD 处理后, 有不同的能量, 如图 7 所示。

与常用的信号处理算法对比, 如离散小波变换 (discrete wavelet transform, DWT)。本文将不同故障类型的电压信号经 DWT 处理, 得到 8 个细节系数, 计算各细节系数的能量如图 8 所示, 也能为分类器提供故障辨识信息。

本文还研究了 IF-AD 和 DWT 处理一个电压信号所需时长。同时在不同信噪比<sup>[16]</sup>下测试了两种算法的耗时, 结果如表 4 所示。

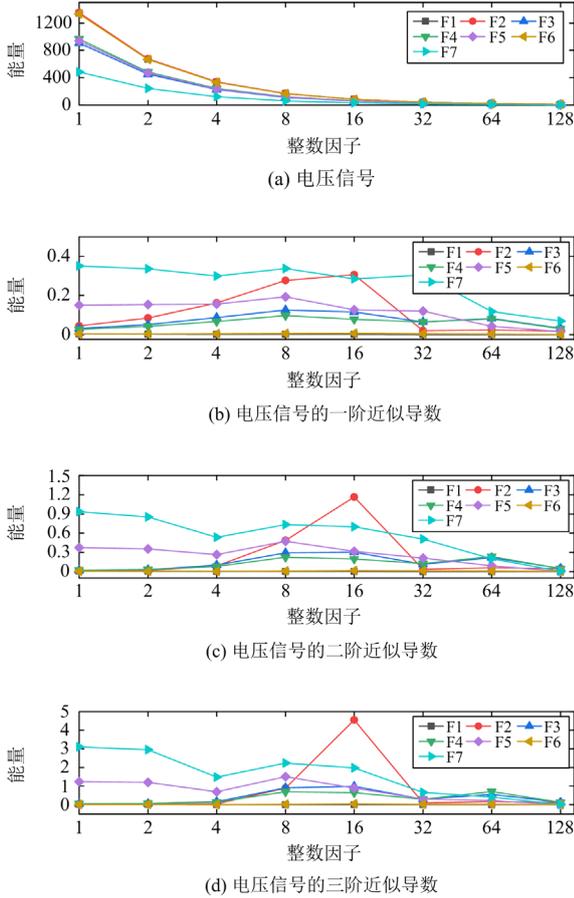


图 7 不同故障类型下电压信号的能量

Fig. 7 Energy of voltage signals for different fault types

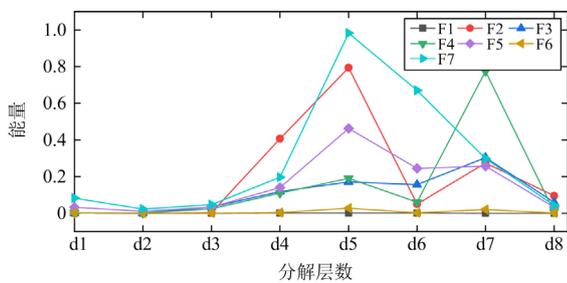


图 8 不同故障类型的电压信号经 DWT 处理后的能量

Fig. 8 Energy of voltage signals for different fault types after DWT processing

表 4 IF-AD 和 DWT 耗时

Table 4 Running time of IF-AD and DWT

算法	无噪声	20 dB	30 dB
DWT	0.11	0.12	0.13
IF-AD	0.022	0.021	0.022

由图 7 和图 8 可知, 相比于 DWT, IF-AD 算法能获得更多特性曲线和子信号, 这将为 ELM 模型提供更多的分类信息, 从而提高故障辨识的准确率。同时, 由表 4 可知, IF-AD 算法耗时更短。

### 5.2 RFECV 特征选择结果

一个样本数据可以得到 288 个特征, 进行特征选择前数据集的大小为  $1050 \times 288$ 。利用 RFECV 算法进行特征选择, 得到 14 个优质特征, 结果如表 5 所示。其中 AD 值表示近似导数阶数, 表中“信号”代表仅经过 IF 下采样得到的信号。

表 5 特征筛选结果

Table 5 Feature screening results

序号	IF 值	AD 值	特征
1	1	2	最小值
2	2	1	香农熵
3	2	1	最小值
4	2	3	香农熵
5	4	信号	对数能量
6	4	1	对数能量
7	4	2	平均值
8	4	2	峰值
9	4	2	最小值
10	4	3	香农熵
11	16	3	能量
12	16	3	偏度
13	32	1	香农熵
14	64	信号	对数能量

### 5.3 ELM 模型辨识结果

将大小为  $1050 \times 14$  的特征数据集按 7:3 的比例随机分为训练集和测试集, 分别用于训练和测试 ELM 模型。表 6 给出了 ELM 模型的辨识结果, 同时计算了每种故障的辨识准确率。为了验证 RFECV 算法的有效性, 在未进行特征筛选的数据集上用 ELM 进行故障辨识, 结果如表 7 所示。

本文以召回率  $R$ 、 $F_{score}$  和准确率  $A$  衡量模型性能, 这些指标越高, 则模型性能越好。假设  $n_{TP}$  表示实际为正被预测为正的样本数量,  $n_{FN}$  表示实际为正但被预测为负的样本数量,  $n_{FP}$  表示实际为负但被预测为正的样本数量,  $n_{TN}$  表示实际为负被预测为负的样本数量。召回率反映实际为正的样本中预测结果为正的比例<sup>[20]</sup>, 即

$$R = \frac{n_{TP}}{n_{TP} + n_{FN}} \quad (12)$$

$F_{score}$  用于衡量模型精确度, 可表示为<sup>[21]</sup>

$$F_{score} = \frac{2n_{TP}}{2n_{TP} + n_{FN} + n_{FP}} \quad (13)$$

准确率为所有正确预测的样本占总样本的比例, 即

$$A = \frac{n_{TP} + n_{TN}}{n_{TP} + n_{TN} + n_{FP} + n_{FN}} \quad (14)$$

表 6 ELM 辨识结果

Table 6 ELM identification results

类型	识别正确	识别错误	样本总数	准确率/%
F1	45	0	45	100
F2	45	0	45	100
F3	45	0	45	100
F4	44	1	45	97.78
F5	43	2	45	95.56
F6	45	0	45	100
F7	42	3	45	93.33
总体	309	6	315	98.10

表 7 未经 RFECV 筛选的辨识结果

Table 7 Identification results without RFECV screening

类型	识别正确	识别错误	样本总数	准确率/%
F1	45	0	45	100
F2	45	0	45	100
F3	45	0	45	100
F4	45	0	45	100
F5	40	5	45	88.89
F6	43	2	45	95.56
F7	39	6	45	86.67
总体	302	13	315	95.87

将 ELM 与其他分类算法对比, 如 SVM、ANN、随机森林(random forest, RF)算法等<sup>[22-23]</sup>, 结果如表 8 所示。为了验证所提模型的鲁棒性, 在不同分类算法中, 用不同信噪比的电压数据进行测试, 结果如表 9 所示。

最后将常用的信号处理算法 DWT 和 EMD 与 IF-AD 算法对比, 结果如表 10 所示。

表 8 不同分类算法比较

Table 8 Comparison with other classification algorithms

算法	召回率	$F_{score}$	准确率/%
IF-AD-SVM	0.9268	0.9238	92.70
IF-AD-RF	0.9110	0.9114	90.16
IF-AD-CNN	0.9720	0.9714	96.83
IF-AD-ANN	0.8390	0.8398	82.86
IF-AD-ELM	0.9810	0.9810	98.10

表 9 不同信噪比下模型性能

Table 9 Model performance under different signal-to-noise ratios

算法	信噪比	
	30 dB	40 dB
SVM	84.76%	89.21%
RF	83.49%	87.30%
CNN	87.62%	92.70%
ANN	74.60%	77.14%
ELM	96.83%	97.78%

表 10 不同信号处理算法对比

Table 10 Comparison of different signal processing algorithms

算法	召回率	$F_{score}$	准确率/%
DWT-ELM	0.9619	0.9599	95.56
EMD-ELM	0.8908	0.8858	87.94
IF-AD-ELM	0.9810	0.9810	98.10

由表 6 和表 7 可知, 经过 RFECV 算法筛选后模型辨识准确率为 98.10%。而在未经筛选的数据集上, 模型准确率为 95.87%。说明 RFECV 算法能有效筛选优质特征子集, 提高辨识准确率。由表 8 可知, 本文采用的 ELM 算法召回率、 $F_{score}$  和准确率均高于其他分类算法, 辨识效果更好。同时, 由表 9 和表 10 可知, 本文所提的故障辨识模型有较强的鲁棒性, 相比于其他辨识模型, 准确率更高。

## 6 结论

本文提出了一种 IF-AD-ELM 特高压直流输电系统故障辨识模型。将不同故障情况下的电压信号经 IF-AD 算法处理后提取熵特征, 通过 RFECV 算法进行特征筛选, 最后利用 ELM 算法进行故障辨识。实验结果表明本文提出的特高压直流输电系统故障辨识方法在高阻接地情况下相比于其他辨识模型有更高的准确率, 且抗干扰能力强。

## 参考文献

- [1] 王渝红, 陈立维, 曾琦, 等. 特高压直流分层接入方式下预防换相失败的协调控制策略[J]. 电力系统保护与控制, 2022, 50(8): 160-170.  
WANG Yuhong, CHEN Liwei, ZENG Qi, et al. Coordinated control strategy for preventing commutation failure in a UHVDC system hierarchically connected to an AC grid[J]. Power System Protection and Control, 2022, 50(8): 160-170.
- [2] 李生虎, 赵慧洁, 周慧敏, 等. 基于可靠性灵敏度的 UHVDC 系统预防性检修优化[J]. 电力系统保护与控制, 2022, 50(3): 68-75.  
LI Shenghu, ZHAO Huijie, ZHOU Huimin, et al. Reliability sensitivity-based optimization for a UHVDC transmission system with preventive maintenance[J]. Power System Protection and Control, 2022, 50(3): 68-75.

- [3] 王顺亮, 谢洋, 马俊鹏, 等. 基于直流电流控制的特高压直流分层接入系统协调控制策略[J]. 电力系统保护与控制, 2022, 50(19): 167-178.  
WANG Shunliang, XIE Yang, MA Junpeng, et al. Coordinated control strategy for a UHVDC hierarchical connection system based on DC current control[J]. Power System Protection and Control, 2022, 50(19): 167-178.
- [4] SONG G B, CHU X, GAO S P, et al. A new whole-line quick-action protection principle for HVDC transmission lines using one-end current[J]. IEEE Transactions on Power Delivery, 2014, 30(2): 599-607.
- [5] LUO S X, DONG X Z, SHI S X, et al. A directional protection scheme for HVDC transmission lines based on reactive energy[J]. IEEE Transactions on Power Delivery, 2015, 31(2): 559-567.
- [6] ZHANG Y K, LI Y L, SONG J Z, et al. A new protection scheme for HVDC transmission lines based on the specific frequency current of DC filter[J]. IEEE Transactions on Power Delivery, 2018, 34(2): 420-429.
- [7] MUZZAMMEL R, RAZA A. Low impedance fault identification and classification based on Boltzmann machine learning for HVDC transmission systems[J]. Journal of Modern Power Systems and Clean Energy, 2022, 10(2): 440-449.
- [8] WU J Y, LAN S, XIAO S J, et al. Single pole-to-ground fault location system for MMC-HVDC transmission lines based on active pulse and CEEMDAN[J]. IEEE Access, 2021, 9: 42226-42235.
- [9] MUZZAMMEL R, RAZA A. A support vector machine learning-based protection technique for MT-HVDC systems[J]. Energies, 2020, 13.
- [10] 陈仕龙, 曹蕊蕊, 毕贵红, 等. 利用多分辨奇异谱熵和支持向量机的特高压直流输电线路区内外故障识别方法[J]. 电网技术, 2015, 39(4): 989-994.  
CHEN Shilong, CAO Ruihui, BI Guihong, et al. Distinguish faults located inside /outside protection zone of UHVDC transmission line by multi-resolution singular spectrum entropy and support vector machine[J]. Power System Technology, 2015, 39(4): 989-994.
- [11] FARSHAD M. Detection and classification of internal faults in bipolar HVDC transmission lines based on K-means data description method[J]. International Journal of Electrical Power & Energy Systems, 2019, 104: 615-625.
- [12] 罗潇, 於锋, 彭勇. 基于深度学习的无人机电网巡检缺陷检测研究[J]. 电力系统保护与控制, 2022, 50(10): 132-139.  
LUO Xiao, YU Feng, PENG Yong. UAV power grid inspection defect detection based on deep learning[J]. Power System Protection and Control, 2022, 50(10): 132-139.
- [13] LAN S, CHEN M J, CHEN D Y. A novel HVDC double-terminal non-synchronous fault location method based on convolutional neural network[J]. IEEE Transactions on Power Delivery, 2019, 34(3): 848-857.
- [14] HOU S Z, GUO W, WANG Z Q, et al. Deep-learning-based fault type identification using modified CEEMDAN and image augmentation in distribution power grid[J]. IEEE Sensors Journal, 2021, 22(2): 1583-1596.
- [15] TORRES J A, DOS SANTOS R C, YANG Q Q, et al. Analyses of different approaches for detecting, classifying and locating faults in a three-terminal VSC-HVDC system[J]. International Journal of Electrical Power & Energy Systems, 2022, 135.
- [16] AKMAZ D. Approximate-derivative-based signal-processing method to segment power-quality disturbances[J]. IET Generation, Transmission & Distribution, 2020, 14(21): 4835-4846.
- [17] FANG G, XU P, LIU W B. Automated ischemic stroke subtyping based on machine learning approach[J]. IEEE Access, 2020, 8: 118426-118432.
- [18] ZHOU D, ZHUANG X, ZUO H F, et al. Deep learning-based approach for civil aircraft hazard identification and prediction[J]. IEEE Access, 2020, 8: 103665-103683.
- [19] ZENG W L, CAO Y F, FENG L T, et al. Hybrid CEEMDAN-DBN-ELM for online DGA serials and transformer status forecasting[J]. Electric Power Systems Research, 2023, 217.
- [20] 侯慧, 陈希, 李敏, 等. 一种 Stacking 集成结构的台风灾害下停电空间预测方法[J]. 电力系统保护与控制, 2022, 50(3): 76-84.  
HOU Hui, CHEN Xi, LI Min, et al. A space prediction method for power outage in a typhoon disaster based on a stacking integrated structure[J]. Power System Protection and Control, 2022, 50(3): 76-84.
- [21] VINAYAGAM A, VEERASAMY V, RADHAKRISHNAN P, et al. An ensemble approach of classification model for detection and classification of power quality disturbances in PV integrated microgrid network[J]. Applied Soft Computing, 2021, 106.
- [22] 王渝红, 杜婷, 廖建权, 等. 基于故障电流预测的柔性直流电网自适应限流策略[J]. 电力系统自动化, 2023, 47(20): 134-142.  
WANG Yuhong, DU Ting, LIAO Jianquan, et al. Adaptive current limiting strategy for flexible DC power grid based on fault current prediction[J]. Automation of Electric Power Systems, 2023, 47(20): 134-142.
- [23] RAO P S, PARIDA P, SAHU G, et al. A multi-view human gait recognition using hybrid whale and gray wolf optimization algorithm with a random forest classifier[J]. Image and Vision Computing, 2023, 136.

收稿日期: 2023-08-12; 修回日期: 2024-01-24

作者简介:

杨新宇(1998—), 男, 硕士研究生, 主要研究方向为高压直流输电; E-mail: yxy18481910593@163.com

赵庆生(1969—), 男, 通信作者, 博士, 副教授, 硕士生导师, 主要研究方向为电力系统运行与控制、高压直流输电; E-mail: zhaoqs1996@163.com

韩肖清(1964—), 女, 博士, 教授, 博士生导师, 主要研究方向为电力系统运行与控制、新能源发电与微电网。E-mail: hanxiaqing@tyut.edu.cn

(编辑 魏小丽)