

# 基于 Hoeffding Tree 的电能质量在线扰动分类

丁建光,张沛超

(上海交通大学 电气工程系 电力传输与功率变换控制教育部重点实验室,上海 200240)

**摘要:** 为满足电能质量扰动事件的在线分类需求,提出了一种基于 Hoeffding Tree 的电能质量扰动在线分类方法。对电能质量在线扰动分类中的关键技术进行了研究,提出用小波变换和离散傅里叶变换相结合的判别方法检测电能质量扰动,该算法采用自适应滑动数据窗算法,能够根据扰动持续时间提取完整的扰动事件。以小波信号能量以及基波有效值构成特征向量,利用 Hoeffding Tree 算法构建增量式分类训练模型。仿真结果表明,所提方法的准确度和效率均满足电能质量扰动事件在线检测和分类的要求。

**关键词:** 电能质量; 扰动; 小波变换; 自适应滑动数据窗; 数据挖掘; Hoeffding Tree; 噪声; 数据流

**中图分类号:** TM 77

**文献标识码:** A

**DOI:** 10.3969/j.issn.1006-6047.2014.09.014

## 0 引言

随着电网中接纳越来越多的可再生能源,电能质量扰动 PQD(Power Quality Disturbance)事件的监视与识别成为智能电网的一项重要研究内容。PQD 识别属于典型的分类问题,已有研究大多基于数据挖掘 DM(Data Mining)技术,并结合现代数字信号处理方法完成电能质量数据的去噪、特征量提取、分类等处理。如文献[1-3]利用小波变换(WT)、dq 变换、快速傅里叶变换(FFT)、Hilbert 变换、数学形态学和短窗功率算法等方法进行 PQD 检测;文献[4-5]研究了检测过程中信号去噪的问题;文献[6]通过小波变换和决策树方法进行扰动分类;文献[7]分析了决策树、贝叶斯分类、支持向量机、人工神经网络等数据挖掘算法。但上述研究均基于传统数据挖掘技术,数据挖掘的实时性不高,甚至仅适用于离线分析。由于电能质量数据具有高速连续到达、潜在无限输入等海量数据的特征,数据必须得到及时处理而不能累积。同时,未来的智能电网将存在大量在线决策环节,而实时数据处理是在线决策的基本前提。所以有必要研究海量电能质量数据的实时扰动识别和在线分类挖掘问题。

数据流(data stream)<sup>[8-9]</sup>技术是近年来数据库和数据挖掘领域的一个研究热点。数据流是实时、连续、有序、时变、无限长的带时标的元组序列,目前数据流系统仍处于研发的原型阶段,针对电力系统的应用研究则更为鲜见。文献[10-11]从电网和配电自动化的角度,利用数据流技术,提出集传感测量、实时

数据分析和在线决策于一体的统一解决方案。文献[12]利用数据流方法分析 PQD 事件,但其采用固定长度的滑动数据窗 SW(Sliding Window)且未研究数据流分类算法。由于 PQD 事件具有很大的随机性,这种固定长度的时间窗难以满足在线分类的需要。

由于 PQD 数据具有典型的数据流特征,使得 PQD 的实时识别和在线分类需解决几个关键问题:首先是实时监测和识别数据流的变化,并能有效区分变化和噪声;其次,对于连续数据流不能采用阻塞方式处理,而必须采用变长滑动数据窗<sup>[13]</sup>技术,根据扰动特点将无限长的数据流转化为一系列可控的数据单元;最后,需要建立高效的数据流分类模型,即分类算法应能利用有限的时间、空间处理无限的流数据。

本文围绕上述 3 个关键问题展开研究。首先将 WT 和离散傅里叶变换(DFT)相结合,对采样数据进行信号去噪和扰动检测;然后提出自适应滑动数据窗 ASW(Adaptive Sliding Window)的构建方法,以适应不同长度的扰动,最大限度地保留扰动信息的能量分布特征;最后,引入具有低存储空间需求的 Hoeffding Tree 算法,对 PQD 事件进行分类。

## 1 技术路线

常见 PQD 信号可分为 6 类<sup>[6]</sup>,分别是电压凸起、电压凹陷、电压间断、谐波、脉冲暂态和振荡暂态,其数学模型见表 1。表中,  $A_m$  为正常情况下电压的幅值;  $A_f$  为电压凸起、凹陷或间断的幅度比例;  $\mu(t)$  为阶跃函数,即当  $t < 0$  时  $\mu(t) = 0$ , 当  $t \geq 0$  时  $\mu(t) = 1$ ;  $t_1$  为扰动的起始时间,  $t_2$  为扰动的结束时间;  $\alpha_k$  为  $k$  次谐波的幅值;  $b$ 、 $c$  分别为振荡的频率和衰减速度;  $\alpha$  为脉冲的幅值;  $\delta(t)$  为  $t$  时刻的单位脉冲函数。

脉冲暂态持续时间在 0.5 ~ 2 ms (即 4 ~ 12 个采样点)之内随机确定,其余扰动类型持续时间在 0.02 ~ 0.5 s 范围内随机确定;电压凸起的幅值大于  $1.2 A_m$ ,

收稿日期:2013-10-13;修回日期:2014-07-10

基金项目:国家高技术研究发展计划(863 计划)资助项目(2012-AA050503);上海市科委资助项目(11dz1210402)

Project supported by the National High Technology Research and Development Program of China (863 Program) (2012-AA050503) and Shanghai Science & Technology Commission (11dz1210402)

表 1 电能质量扰动数学模型  
Tab.1 Mathematical models of PQD

PQD 类型	公式	参数
正常	$u(t)=A_m \sin(\omega t)$	—
电压凸起	$u(t)=A_m\{1+A_r[\mu(t_2)-\mu(t_1)]\} \sin(\omega t)$	$A_r>0.2$
电压凹陷	$u(t)=A_m\{1-A_r[\mu(t_2)-\mu(t_1)]\} \sin(\omega t)$	$0.2<A_r<0.8$
电压间断	$u(t)=A_m\{1-A_r[\mu(t_2)-\mu(t_1)]\} \sin(\omega t)$	$0.8<A_r<1.0$
谐波	$u(t)=A_m[\alpha_1 \sin(\omega t)+\alpha_3 \sin(3\omega t)+\alpha_5 \sin(5\omega t)+\alpha_7 \sin(7\omega t)]$	$0.05 A_m<\alpha_3, \alpha_5, \alpha_7<0.15 A_m, \sum_{k=1,3,5,7} \alpha_k^2=1$
脉冲暂态	$u(t)=A_m[\sin(\omega t)+\alpha \delta(t-t_0)]$	$20<\alpha<500$
振荡暂态	$u(t)=A_m\{\sin(\omega t)+\alpha e^{-\beta(t-t_0)} \sin(b\omega t)[\mu(t_2)-\mu(t_1)]\}$	—

电压凹陷的幅值在  $0.2A_m \sim 0.8A_m$  之间,电压间断的幅值小于  $0.2A_m$ ,脉冲暂态的幅值大于  $20A_m$  且小于  $500A_m$ ,这 4 种扰动电压幅值在指定范围内随机产生;电压谐波的各谐波幅值  $\alpha_3, \alpha_5, \alpha_7$  在  $0.05A_m \sim 0.15A_m$  范围,且满足  $\sum \alpha_k^2=1$ ,其中  $k=1, 3, 5, 7$ 。

扰动信号可以抽象为数据流  $S$ ,  $S$  可以形式化地表示为一个无限的时间序列:

$$S = \{\langle s_1, t_1 \rangle, \langle s_2, t_2 \rangle, \dots, \langle s_i, t_i \rangle, \dots\} \quad (1)$$

其中,  $\langle s_i, t_i \rangle$  表示  $t_i$  时刻产生的  $s_i$  元组,该元组可以是数值,也可以是向量,本文以变量形式表示。

滑动数据窗技术使得 PQD 分析可以针对最新的一个有限数据子集,而非所有可见的数据流。设当前时刻为  $t_n$ ,则当前的滑动数据窗可以表示为:

$$W_n(K) = (\langle s_{n-K+1}, t_{n-K+1} \rangle, \langle s_{n-K+2}, t_{n-K+2} \rangle, \dots, \langle s_n, t_n \rangle) \quad (2)$$

其中,  $K$  为数据窗长度;  $s_{n-K+1}, s_{n-K+2}, \dots, s_n$  为采样值。根据  $K$  变化与否,滑动数据窗可分为定长数据窗和变长数据窗。

通过对电能质量采样数据进行扰动检测和特征提取,可以得到扰动事件的特征向量时间序列,该序列同样可以利用数据流表示为:

$$V = (\langle v_1, t_1 \rangle, \langle v_2, t_2 \rangle, \dots, \langle v_i, t_i \rangle, \dots) \quad (3)$$

其中,特征向量  $v_i = [a_1, a_2, \dots, a_j, \dots, a_m, C]$  是一个分类实例,  $a_j$  为实例的属性,  $C$  为实例的类标签(class label)。当  $C$  非空时,该实例是用于训练的样本。

综合以上分析,形成图 1 所示的总体技术路线。

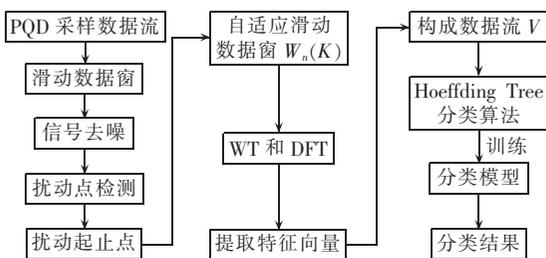


图 1 PQD 在线检测和分类

Fig.1 Online detection and classification of PQD

## 2 基于数据流的 PQD 检测

### 2.1 扰动点检测

扰动点检测是实现自适应滑动数据窗的关键步骤,本文提出 WT 和 DFT 相结合的扰动起止类型综

合判别算法的流程。本文选取 db4 小波作为基函数,并进行 3 层分解。

电力系统中的噪声以高斯白噪声为主,本文采用以 Stein 无偏风险阈值为基础的软阈值去噪方法<sup>[5]</sup>,在去噪的同时能准确保留扰动特征。对滑动数据窗中的采样数据进行 WT,通过第 1 层细节分量系数可以分析是否存在扰动。以电压凸起信号为例,正常电压幅值为 1 p.u.,凸起时为 1.2 p.u.,扰动发生在第 640 个采样点处,持续至第 1280 个采样点处,如图 2(a)所示。以 db4 小波进行 3 层分解,得到第 1 层细节分量系数,扰动起止点对应分量系数明显高于其他点,如图 2(b)所示。

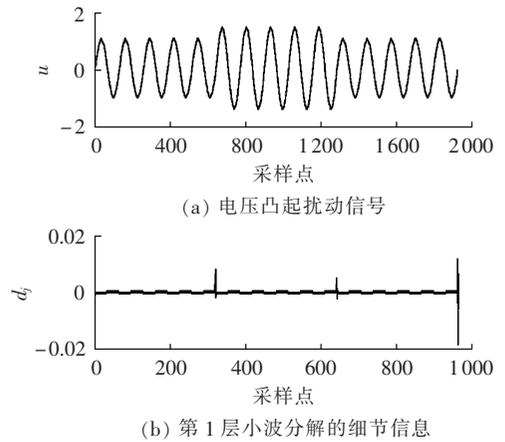


图 2 电压凸起的扰动检测

Fig.2 Disturbance detection for voltage swell

检测到扰动后,以扰动点为中心点,提取一个完整周期的采样值,通过 DFT 计算基波有效值  $K_1$ ,并与正常情况下基波有效值  $K_0$  作比较。若满足  $|K_1 - K_0| < 0.02K_0$ ,则判定该点由残余噪声引起,重新提取下一个滑动数据窗;如果扰动点不是由残余噪声引起,则提取扰动点下一个周期的采样值并计算基波有效值  $K_2$ ,并和  $K_0$  作比较。如果满足  $|K_2 - K_0| < 0.02K_0$ ,则判定该扰动点为 PQD 的终止点,提取事件后,重新提取下一滑动数据窗;否则判定为起始点,缓存后重新提取下一滑动数据窗。

经过上述处理可得扰动点序列:

$$D = (\langle D_1, t_1 \rangle, \langle D_2, t_2 \rangle, \dots, \langle D_i, t_i \rangle, \dots) \quad (4)$$

其中,  $D_i$  为扰动点的起止类型,  $i$  取 0、1、2 分别表示噪声点、扰动的起始点和终止点。如果用向上箭头表示扰动起始点, 向下箭头表示终止点, 无向竖线表示由残余噪声引起的扰动, 则上述处理可检测出图 3 所示的 2 类扰动点序列。其中, 图 3(a) 所示序列由 1 个起始点和 1 个终止点组成, 对应电压凸起、电压凹陷、电压间断、谐波和振荡暂态等扰动; 图 3(b) 所示序列仅包含 1 个终止点, 对应脉冲暂态扰动。图 3 中还包含了一些残余噪声扰动点, 这些扰动点既可能距起始点和终止点超过 1 个周期 (如  $N_1$ ), 也可能紧邻起始点和终止点 (如  $N_2$ )。

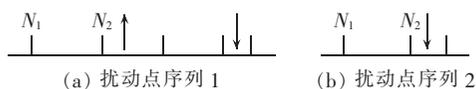


图 3 扰动点序列

Fig.3 Sequence of disturbance points

## 2.2 自适应滑动数据窗

本文提出自适应滑动数据窗, 能够根据式(4)所形成的扰动点序列, 自适应地调整滑动数据窗的长度, 算法描述如下。

设在式(4)中,  $D_i$  为第 1 个非 0 扰动点; 如果  $D_i=1$  (起始点), 则从  $D_i$  开始找到第 1 个终止点, 设为  $D_j$ , 于是记本 PQD 事件的数据窗为  $[t_i, t_j]$ ; 如果  $D_i=2$  (终止点), 则数据窗记为  $[t_i, t_i]$ , 如此类推。然后, 将上述数据窗的起始时刻向前、终止时刻向后各延伸 1 个周期, 形成一个自适应滑动数据窗, 所截取出的采样点能够完整包含一个 PQD 事件。如果在起始时刻之前或终止时刻之后 1 个周期内存在噪声, 则以噪声扰动点为基点, 延展数据窗范围。

## 2.3 特征向量提取

基波有效值能直接反映电压幅值偏离的大小与方向, 且能有效区分残余噪声, 但其计算数据窗较长, 难以有效识别短暂时态事件。文献[14]提出了小波能量特征, 利用小波的多分辨率原理将信号分解到不同频段。根据 Parseval 定理, 信号能量可通过小波概要系数  $c_j(k)$  和细节系数  $d_j(k)$  进行重构:

$$\int |f(t)|^2 dt = \sum_k |c_l(k)|^2 + \sum_{j=1}^l \sum_k |d_j(k)|^2 \quad (5)$$

其中,  $l$  为小波分解层数。综上, 本文采用 3 层小波 ( $l=3$ ) 分解能量和基波有效值作为特征向量:

$$v_i = [c_3, d_1, d_2, d_3, a, C] \quad (6)$$

其中,  $c_3, d_1, d_2, d_3$  分别为小波概要系数和细节系数的能量;  $a$  为基波有效值特征量;  $C$  为类标签。

## 3 基于数据流的 PQD 分类算法

传统的决策树、支持向量机等分类算法采用批处理算法, 训练过程中需要一次性加载所有样本实例, 时间和空间开销很大。针对数据流的实例数目巨大

且容易发生概念漂移的特点, 本文选择 Hoeffding Tree 分类算法<sup>[15]</sup>。该算法能增量式地建立决策树, 每次只完成一个实例的处理, 并且在处理完成后不保存实例信息, 仅保存决策树信息, 这样在保证分类准确性的同时极大降低了算法的时间和空间复杂度。

### 3.1 Hoeffding Tree 算法的分类原则

该算法引入熵的概念衡量一段数据所载有的信息量, 通过计算信息增益来判别何时进行分枝。如果类标签  $C$  有  $n$  个不同的值, 第  $i$  个值在整体样本中对应的概率为  $p_i$ , 则熵的计算为<sup>[17]</sup>:

$$e(p_1, p_2, \dots, p_n) = -\sum_{i=1}^n p_i \log_2 p_i \quad (7)$$

信息增益通过使用类分布的熵减去以某属性进行分枝的加权平均熵而计算获得, 即如果使用属性  $a_j$  进行分枝, 属性  $a_j$  可以将整个数据样本划分为  $S_1, S_2, \dots, S_v$ , 每一份有一个对应的熵。将其加权平均, 可以获得以  $a_j$  划分子集的信息增益为:

$$G(a_j) = e(p_1, p_2, \dots, p_n) - E(a_j) \quad (8)$$

$$E(a_j) = \sum_{i=1}^v w_i e_i(p_{i1}, p_{i2}, \dots, p_{in}) \quad (9)$$

其中, 权重  $w_i$  为第  $i$  份样本在整个样本中的比重;  $e_i$  为第  $i$  份样本的熵;  $p_{i1}, p_{i2}, \dots, p_{in}$  分别为各标签值在第  $i$  份样本中对应的概率。在传统的批处理分类方式下, 很容易使用信息增益最高的属性进行分枝, 但是在增量处理方式下, 进行分枝需要引入 Hoeffding 边界 HB (Hoeffding Bound) 的概念。

$$HB = \sqrt{R^2 \ln(1/\delta) / (2m)} \quad (10)$$

其中,  $R = \log_2 n$ ,  $n$  为类标签数;  $\delta$  为设定的置信因子,  $1-\delta$  描述了在树的每个节点上, 正确属性被选中的预期概率, 显然希望这个概率尽量接近 1, 根据文献[15], 可将  $\delta$  设为  $10^{-7}$ ;  $m$  为已处理的实例个数。

HB 衡量了信息增益最高的属性  $a'_j$  和次高的属性  $a''_j$  之间信息增益的差值的最小值, 如果该差值最小值超过 HB, 则将信息增益最高的属性  $a'_j$  确定为分类属性。

### 3.2 Hoeffding Tree 算法逻辑图和参数设置

首先生成仅有 1 个根节点的树结构  $T$ , 然后依次对每个实例进行评估, 即增量式地完成学习任务。算法的流程如图 4 所示。

在评估的过程中, 如果存在 1 个叶子节点, 该节点的实例数目  $n_i$  是设定值  $n_{\min}$  的整数倍, 且所有节点不属于同一个类标签, 即满足图 4 中的判别条件 1, 则在该节点进行分枝条件判别。设置参数  $n_{\min}$  是因为单一实例对信息增益影响很小, 当某个叶子节点累计增加  $n_{\min}$  个实例后, 才考虑对该节点进行后续的分枝条件判断。

分枝条件判别需计算每个属性的信息增益  $G(a_j)$ ,

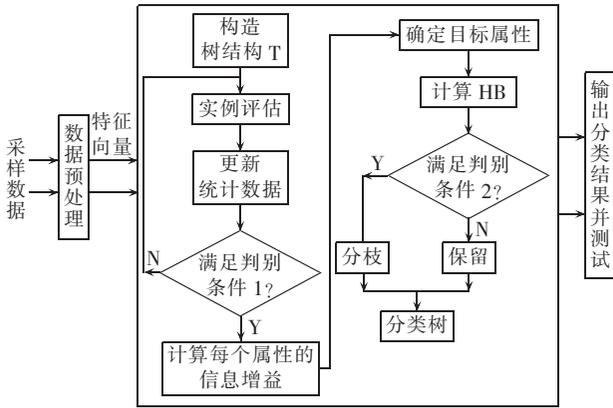


图 4 Hoeffding Tree 算法示意图

Fig.4 Schematic diagram of Hoeffding Tree algorithm

确定增益最高的 2 个属性。若  $G(a'_i) - G(a''_i) > HB$  或者  $HB < \tau$  ( $\tau$  是预设的 HB 的最小值), 即满足图 4 中的判别条件 2, 则进行分枝, 否则继续进行下一个实例的评估。通过预设  $\tau$  可避免随着分配到该节点的实例越来越多, HB 越来越小, 进而造成在该节点始终无法分枝。

## 4 算例与分析

### 4.1 实验数据

本文采用电压凸起、电压凹陷、电压间断、谐波、脉冲暂态和振荡暂态这 6 种 PQD 模型进行算例分析, 采样频率为 6.4 kHz。为模拟实际情况, 所有扰动事件的起始和持续时间皆随机产生, 模型参数设置见表 1。扰动数据中添加 40 dB 高斯白噪声, 每种扰动采用 20000 个实例训练和 10000 个实例测试。

### 4.2 PQD 检测算法测试

在 MATLAB 仿真环境下, 随机产生每种扰动事件各 20000 个, 并记录各扰动的起始时刻, 然后测试 PQD 检测的准确度, 结果见表 2。

表 2 PQD 检测结果

Tab.2 Results of PQD detection

PQD 类型	检测到的 PQD 数量	检测错误率/%
电压凸起	19832	0.84
电压凹陷	19994	0.03
电压间断	20000	0
谐波	19985	0.08
脉冲暂态	19897	0.52
振荡暂态	18831	5.85

### 4.3 PQD 分类算法测试

在数据流挖掘分类过程中, 采用 Hoeffding Tree 分类算法, 以信息增益作为分枝标准, 信息增益评估周期采用 200 个实例, 即  $n_{min} = 200$ ; 置信因子  $\delta = 10^{-7}$ ;  $\tau = 0.05$ 。分类结果见表 3。

由表 3 可知, 以小波能量和电压基波有效值为特征向量, 通过本文的分类算法可以很好地识别 PQD 类型, 每种扰动分类的准确率均在 99% 以上。

表 3 PQD 分类测试结果  
Tab.3 Results of PQD classification

PQD 类型	分类结果						
	正弦电压	电压凸起	电压凹陷	电压间断	谐波	脉冲暂态	振荡暂态
正弦电压	10000	0	0	0	0	0	0
电压凸起	0	9999	0	0	1	0	0
电压凹陷	0	0	9995	4	0	0	1
电压间断	0	0	13	9987	0	0	0
谐波	1	0	0	0	9999	0	0
脉冲暂态	0	0	0	0	0	10000	0
振荡暂态	0	0	1	0	0	0	9999

针对相同类型和参数的扰动, 本文将 Hoeffding Tree 算法与典型的 C4.5 分类算法<sup>[6]</sup>做了对比, 结果如表 4 所示。Hoeffding Tree 算法本身在分类准确率方面与 C4.5 相近<sup>[15]</sup>, 本文算法准确率的提高主要是因为采用了自适应变长滑动数据窗技术。

表 4 Hoeffding Tree 和 C4.5 算法准确率对比

Tab.4 Comparison of correctness rate between Hoeffding Tree and C4.5 algorithms

PQD 类型	算法准确率/%	
	C4.5 决策树算法	Hoeffding Tree 算法
电压凸起	87.5	99.99
电压凹陷	93.0	99.95
电压间断	96.5	99.87
谐波	100	99.99
脉冲暂态	97.0	100
振荡暂态	95.0	99.99

目前, 针对电能质量复合扰动分类的研究还不多, 已有方法主要采用传统单一扰动识别方法<sup>[17]</sup>, 区分效果较差。本文针对所提算法进行了复合扰动的测试。按照文献<sup>[1]</sup>中暂态扰动和稳态扰动相复合的原则, 并参考文献<sup>[18]</sup>的复合扰动类型, 设定存在各种单一扰动和“谐波+电压凸起”、“谐波+电压凹陷”、“谐波+电压间断”3 种复合扰动, 每种扰动采用 20000 个实例训练、10000 个实例测试, 本文算法对复合扰动的分类效果如表 5 所示。其中, 识别率指正确识别出复合扰动类型或识别出其中一种单一扰动类型的比率; 准确率指准确识别出复合扰动类型的比率; 漏分指仅检测出其中一种单一扰动; 错分指未检测出复合扰动的任意一种。由表 5 可见, 在复合扰动类型识别方面, 本文算法仍需做进一步改进。

表 5 复合扰动分类效果

Tab.5 Effect of complex PQD classification

复合扰动类型	识别率/%	准确率/%	漏分率/%	错分率/%
谐波+电压凸起	90.2	79.0	11.2	9.8
谐波+电压凹陷	94.7	83.8	10.9	5.3
谐波+电压间断	96.0	87.2	8.8	4.0

### 4.4 算法的时间与空间性能测试

基于数据流挖掘的 PQD 在线分类算法应具有良好的时间和空间性能, 能够高效处理实时、高速和无限的采样数据流而不造成数据堵塞。算法仿真用

计算机的 CPU 主频为 3.10 GHz、内存为 3.24 GB、操作系统为 32 位。分别进行 PQD 检测和特征量提取,以及分类算法训练和测试,算法的时间和空间性能测试结果如表 6 所示。

表 6 算法时间和空间复杂度分析

Tab.6 Analysis of temporal and spatial complexities of algorithm

测试项目	运行时间/s	占用内存/MB
PQD 检测和特征量提取	12.74	85.33
数据流分类	0.81	68.00

连续采样时间为 2240 s,每种扰动包含 1000 个事件。算法用于扰动检测和特征量提取耗时 12.74 s,用于数据流分类耗时 0.81 s,算法合计用时 13.55 s。若以算法执行时间与样本持续采样时间之比衡量算法实时性<sup>[6]</sup>,则实时性指标为  $13.55/2240=0.00605$ ,表明算法具有良好的实时性。

目前,绝大部分研究仅关注分类准确度,而未对实时性进行研究,仅有文献[12]给出了实时性测试结果,但该文没有给出测试用计算机的具体配置。文献[12]中的实时性指标为  $9.464/160=0.059$ ,而本文实时性指标为 0.00605。故本文算法的实时性优于文献[12]中的算法。

## 5 结论

本文将电能质量数据视为数据流,对基于数据挖掘的 PQD 在线识别与分类的关键问题进行了研究。将 WT 和 DFT 相结合,实现采样数据去噪和扰动检测,进而提出自适应变长滑动数据窗的构建方法,从而将无限长的数据流转变为有限长的可控处理单元,并最大限度地保留扰动信息的能量分布特征;以小波概要系数、细节系数能量和基波有效值构成特征向量,并引入能够进行增量式学习的 Hoeffding Tree 算法,对 PQD 进行分类。

本文的研究重点是单一扰动的在线识别与分类。针对复合扰动,本文认为可以采用在线和离线相结合的方法,以兼顾实时性和复合扰动识别的准确性,这将是下一步的研究重点。

## 参考文献:

[1] 张秀娟,徐永海,肖湘宁. 基于  $dq$  变换与小波变换的电能质量扰动检测与识别方法[J]. 电力自动化设备,2005,25(7):1-5.  
ZHANG Xiujian,XU Yonghai,XIAO Xiangning. Power quality disturbance detection and identification based on  $dq$  conversion and wavelet transform[J]. Electric Power Automation Equipment, 2005,25(7):1-5.

[2] 李天云,陈晓东,赵为红,等. 几种短时电能质量扰动分类和监测的双小波分析法[J]. 电力系统自动化,2003,27(22):26-30.  
LI Tianyun,CHEN Xiaodong,ZHAO Weihong,et al. Double wavelets measurements and classification of short duration power quality

disturbances[J]. Automation of Electric Power Systems,2003,27(22):26-30.

[3] 李天云,郭跃霞,王静,等. 基于数学形态学和短窗功率算法的电能质量扰动检测方法[J]. 电力自动化设备,2008,28(7):37-40.  
LI Tianyun,GUO Yuexia,WANG Jing,et al. Power quality disturbance detection based on mathematical morphology and power algorithm of short data window[J]. Electric Power Automation Equipment,2008,28(7):37-40.

[4] 张全明,刘会金,兰泉妮. 基于小波混合阈值方法的电能质量信号去噪[J]. 电力自动化设备,2008,28(8):28-30,35.  
ZHANG Quanming,LIU Huijin,LAN Quanni. Consumer importance classification based on VOLL[J]. Electric Power Automation Equipment,2008,28(8):28-30,35.

[5] 欧阳森,宋政湘,陈德桂,等. 小波软阈值去噪技术在电能质量检测中的应用[J]. 电力系统自动化,2002,26(19):56-60.  
OUYANG Sen,SONG Zhengxiang,CHEN Degui,et al. Application of wavelet soft-threshold de-noising technique to power quality detection[J]. Automation of Electric Power Systems, 2002,26(19):56-60.

[6] 孔英会,车麟麟,苑津莎,等. 基于小波分解和数据挖掘中决策树算法的电能质量扰动识别方法[J]. 电网技术,2007,31(23):78-82.  
KONG Yinghui,CHE Linlin,YUAN Jinsha,et al. A power quality disturbance identification method based on wavelet decomposition and decision tree algorithm in data mining[J]. Power System Technology,2007,31(23):78-82.

[7] 唐良瑞,杨雪. 基于三角模融合算子的电能质量去噪算法[J]. 电工技术学报,2007,22(9):154-158.  
TANG Liangrui,YANG Xue. A de-noising method of power quality based on triangle module operator[J]. Transactions of China Electrotechnical Society,2007,22(9):154-158.

[8] 黄树成,曲亚辉. 数据流分类技术研究综述[J]. 计算机应用研究,2009,26(10):3604-3609.  
HUANG Shucheng,QU Yahui. Survey on data stream classification technologies[J]. Application Research of Computers,2009,26(10):3604-3609.

[9] BABCOCK B,BABU S,DATAR M,et al. Models and issues in data stream system[C]//Proceedings of the 21st ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems. New York,USA:ACM,2002:1-30.

[10] 赵加奎,杨国凤,沐连顺,等. 数据流技术在电网自动化中的应用研究[J]. 电网技术,2011,35(8):6-11.  
ZHAO Jiakui,YANG Guofeng,MU Lianshun,et al. Research on the application of the data stream technology in grid automation[J]. Power System Technology,2011,35(8):6-11.

[11] 王永利,徐宏炳,董逸生,等. 配电自动化的数据流管理系统设计[J]. 电力系统自动化,2004,28(13):85-89.  
WANG Yongli,XU Hongbing,DONG Yisheng,et al. Design of distribution automation system supporting DSMs[J]. Automation of Electric Power Systems,2004,28(13):85-89.

[12] 车麟麟,孙景芳. 基于数据流技术的电能质量扰动在线识别[J]. 中国科技信息,2010(9):16-17.  
CHE Linlin,SUN Jingfang. Online power quality disturbances identification based on data streams technology [J]. China Science and Technology Information,2010(9):16-17.

[13] 钱峰,胡光岷. 基于滑动时窗的小波变换实时算法[J]. 信号处理,2007,23(3):361-364.  
QIAN Feng,HU Guangmin. A realtime wavelet algorithm based

- on sliding window technology[J]. Signal Processing, 2007, 23(3):361-364.
- [14] 孔英会,蔡维,何伟. 基于特征组合的 SVM 电能质量扰动信号分类[J]. 华北电力大学学报, 2010, 37(4):72-77.  
KONG Yinghui, CAI Wei, HE Wei. Power quality disturbance signal classification using support vector machine based on feature combination[J]. Journal of North China Electric Power University, 2010, 37(4):72-77.
- [15] BIFET A, KIRKBY R. Data stream mining—a practical approach [M]. [S.l.]: COSI, 2009: 48-63.
- [16] ACKENHUSEN J G. Real-time signal processing; design and implementation of signal processing systems[M]. [S.l.]: Prentice Hall PTR, 1999.
- [17] 刘志刚, 张巧革, 张杨. 电能质量复合扰动分类的研究进展[J]. 电力系统保护与控制, 2013, 41(13):146-153.  
LIU Zhigang, ZHANG Qiaoge, ZHANG Yang. Review of power quality mixed disturbances identification[J]. Power System Pro-

- tection and Control, 2013, 41(13):146-153.
- [18] 周维维, 管春, 卢伟国. 多标签分类法在电能质量复合扰动分类中的应用[J]. 中国电机工程学报, 2011, 31(4):45-50.  
ZHOU Luwei, GUAN Chun, LU Weiguo. Application of multi-label classification method to categorization of multiple power quality disturbances[J]. Proceedings of the CSEE, 2011, 31(3): 45-50.

#### 作者简介:



丁建光

丁建光(1989-),男,河北石家庄人,硕士研究生,主要研究方向为智能变电站、电力系统数据挖掘(**E-mail**:jgding@163.com);

张沛超(1970-),男,江苏连云港人,副教授,博士,主要研究方向为电力系统保护与控制(**E-mail**:pczhang@sjtu.edu.cn)。

## Online power quality disturbance classification based on Hoeffding Tree

DING Jianguang, ZHANG Peichao

(Key Laboratory of Control of Power Transmission and Conversion, Ministry of Education, Department of Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China)

**Abstract:** An online classification method based on Hoeffding Tree is proposed for the online classification of PQD (Power Quality Disturbance). The key technologies used in the online PQD classification based on power quality data stream are researched and a PQD detection method combining the wavelet transform and the DFT (Discrete Fourier Transform) is proposed, which adopts an adaptive sliding window to extract a complete PQD event according to its duration. The characteristic vector is composed of the wavelet energy and the fundamental RMS and the Hoeffding Tree algorithm is applied to build the incremental classification training model. Simulative results show that, the accuracy and efficiency of the proposed method meet the requirements of online PQD detection and classification.

**Key words:** power quality; disturbance; wavelet transforms; adaptive sliding window; data mining; Hoeffding Tree; noises; data stream

(上接第 83 页 continued from page 83)

## State estimation for power system of seafloor observatory network

FENG Yingbin<sup>1,2</sup>, LI Zhigang<sup>2</sup>, WANG Xiaohui<sup>2</sup>

(1. University of Chinese Academy of Sciences, Beijing 100049, China; 2. State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China)

**Abstract:** Because the traditional WLS (Weighted Least Squares) method has low robustness and accuracy when it is applied in the state estimation for the power system of seafloor observatory network with low measurement redundancy, a method combining the wavelet analysis with WLS is proposed, which adopts the wavelet de-noising theory to improve the accuracy of WLS state estimation and the wavelet singularity detection theory to improve the WLS robustness by identifying the sensor faults. The simulative results based on the power system model for seafloor observatory network verify the superiority of the proposed method.

**Key words:** seafloor observatory network; state estimation; electric power systems; WLS method; wavelet analysis; de-noising; singularity detection