

应用组合分类法对暂态电能质量扰动分类

张文涛, 潘瑾, 饶群, 薛俊范

(河北省电力研究院, 河北 石家庄 050021)

摘要: 有机地结合贝叶斯分类法、感知准则函数分类法和最小错分样本数准则分类法的理论优点, 提出了一种新的组合分类方法。该方法保留贝叶斯法的最优点, 去除其使用的必备条件, 保留了最小错分样本数准则分类法可对交叉样本分类, 去除其计算复杂的缺点。与常用的最小错分样本数准则函数分类法以及贝叶斯分类法、感知准则函数分类法相比, 该组合分类法在理论上具有最优分类、可对交叉样本分类、计算简便的优点。在对暂态电能质量扰动分类的仿真分析过程中, 充分显示了新方法的上述优点。

关键词: 分类; 交叉样本; 电能质量; 暂态扰动

Classifying the transient power quality disturbances using combined methods

ZHANG Wen-tao, PAN Jin, RAO Qun, XUE Jun-fan

(Hebei Electric Power Research Institute, Shijiazhuang 050021, China)

Abstract: A new classification method is put forward in this paper which combines the advantages of the Bayes method and the rule of perception function and the rule of least wrong. This new method keeps the advantages of the traditional methods of classification but removes their shortcomings. Comparing to the classification methods commonly used, the new method has several advantages including optimal classification and simple computation and classifying the cross samples. When classifying the transient power quality disturbances, the new method shows all these advantages.

Key words: classification; cross sample; power quality; transient power quality disturbance

中图分类号: TM 711 文献标识码: A 文章编号: 1674-3415(2010)05-0001-04

0 引言

随着国家科技水平的提升, 用户对电能质量的要求越来越高, 对各种扰动进行治理也成为迫切的要求。产生不同种类电能质量暂态扰动的原因不同, 为了有目的的治理, 对扰动进行分类非常重要。有人研究了多种相关方法^[1-3]。这些方法大都理论复杂, 使用计算机实现困难。

一般在对暂态电能质量扰动分类时, 只提取几种特征值, 如扰动持续时间和扰动电压幅值, 这时多个样本间会产生交叉样本。感知准则函数分类法计算简单, 但无法应对样本交叉带来的线性不可分问题, 即计算不收敛。贝叶斯分类方法具有“最优分类”的优点, 但在很多情况下一些必要参数难以确定, 因而应用受到限制。最小错分样本数准则分类法不受是否有交叉样本的限制, 但计算方法过于复杂^[4-5]。

本文提出新分类法将贝叶斯分类法和感知准则

函数分类法的优点集中在一起, 有计算方法简单、准确的优点, 又不受是否有交叉样本限制。对比新方法和最小错分样本数准则分类法, 仿真和分析结果表明新方法有较好的分类效果。

1 特征向量的提取

当发生电能质量暂态扰动时, 总有扰动开始和结束的时间。有多种方法可以确定这两个时间, 并得到暂态扰动的持续时间:

$$t_d = |t_e - t_s| \quad (1)$$

式中: t_d 为扰动持续时间; t_e 为扰动结束时间; t_s 为扰动开始时间。

扰动期间, 测得的电压有多个值, 作为分类标准的扰动电压只能是一个值:

$$U_d = \frac{1}{|t_s - t_e|} \cdot \sum_{i=t_s}^{t_e} (|U_{fi}| - |U_{si}|) \quad (2)$$

式(2)中: U_d 是扰动电压; U_{fi} 是扰动时的测量电压; U_{si} 是无扰动时应有的电压, 此值要在扰动前对应的时刻找到。取绝对值是为了避免负值的影响。

2 几种对样本分类的方法

2.1 贝叶斯最优分类法

基于最小错误率的贝叶斯分类法判定准则为:

$$p(x \in R_2 | \omega_1)P(\omega_1) = p(x \in R_1 | \omega_2)P(\omega_2) \quad (3)$$

此时最小的分类错误率为:

$$p(e) = p(x \in R_2 | \omega_1)p(\omega_1) + p(x \in R_1 | \omega_2)p(\omega_2) \quad (4)$$

式(3)和式(4)中: ω_1 和 ω_2 分别为第1、2类的样本, x 是两类中的某一个样本, R_1 和 R_2 是由分类线确定的样本分类范围。以下的符号类同。

在现实应用中, 很多样本的概率和条件概率难以得到确切的值, 使贝叶斯最优分类法无法应用。

2.2 感知准则函数分类法

设有一组样本集 y_1, y_2, \dots, y_N , 其中 y_n 是规范化增广样本向量, 我们要找一个解向量 a , 使得:

$$a^T y_n > 0, \quad n=1, 2, \dots, N \quad (5)$$

先构造这样一个准则函数:

$$J_p(a) = \sum_{y \in Y^h} (-a^T y) \quad (6)$$

式(6)中 Y^h 是被权向量 a 错分类的样本集合, 当 y 被错分时, 就有 $a^T y \leq 0$ 或

$$-a^T y \geq 0 \quad (7)$$

因此式(6)总是大于等于零, 仅当 a 在解区边界上时, $J_p(a)$ 才为零。

它在20世纪50年代由 Rosenblatt 提出来, 一般称为感知准则函数。感知准则函数是一种简单且常用的一种分类方法。但它使用的梯度下降算法只适用于线性可分情况, 对于线性不可分情况, 迭代过程永远不会终结, 即算法不收敛。

2.3 最小错分样本数准则分类法

最小错分样本数准则分类法是一种既适用于无交叉样本的线性可分情况又适用于有交叉样本的线性不可分情况的算法。

该算法对于线性可分问题, 可以得到一个如感知函数那样的解向量; 对于线性不可分问题则可以得到一个使两类样本集错分数目最少的权向量。具体实现过程为:

式(5)可推广写为:

$$Ya \geq b > 0 \quad (8)$$

其中: Y 是 $N \times \hat{d}$ 规范化增广样本矩阵; b 为引入余量, 可定义为:

$$b = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \left. \vphantom{\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}} \right\} N \text{ 个 } 1 \quad (9)$$

$$Y = \begin{bmatrix} y_1^T \\ y_2^T \\ \vdots \\ y_N^T \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1\hat{d}} \\ y_{21} & y_{22} & \cdots & y_{2\hat{d}} \\ \vdots & \vdots & & \vdots \\ y_{N1} & y_{N2} & \cdots & y_{N\hat{d}} \end{bmatrix} \quad (10)$$

对于式(8)可以定义准则函数

$$J_q(a) = \left\| (Ya - b) - |Ya - b| \right\|^2 \quad (11)$$

如果 $Ya > b$, 则 $(Ya - b)$ 和 $|Ya - b|$ 同号, 因此, $J_q(a) = 0$, 反之, 如果有某些 y_i 不满足 $a^T y_i > b_i$, 则 $(Ya - b)$ 和 $|Ya - b|$ 异号, 因此 $J_q(a) > 0$ 。不满足的 y_i 越多, $J_q(a)$ 越大。显然, $J_q(a)$ 取极小值时的 a 为最优解 a^* , 并且在不定式组一致的情况下, $J_q(a^*) = 0$, 在不一致情况下, $J_q(a^*) > 0$ 。我们称 $J_q(a)$ 为最小错分样本数准则。

对此, Nagaraja 和 Krishna 提出了在 Fletcher-Reeves 共轭梯度算法的计算方法和计算步骤。

3 改进分类法

3.1 对贝叶斯分类法的改进^[6-7]

贝叶斯样本分类如图1所示。图中: 曲线表示按某种特征向量分类的两类样本的分布率, T 左边的样本归于第一类; T 右边的样本归于第二类。假定样本数无穷大, 则可以认为上述样本的分布率曲线是光滑的。样本按照分布率曲线分为如图所示的区域1~4。在以下的各式中:

$\sum S(1,1)$ 代表区域1中第一类的样本总数, 分类归于第一类, 属正确分类。

$\sum S(2,2)$ 代表区域2中第二类的样本总数, 分类归于第二类, 属正确分类。

区域3、4都是由2类样本混和构成, 即线形相交的样本, 各分为2种来表示样本数。

$\sum S(3,1)$ 代表区域3中第1类的样本总数, 分类归于第一类, 属正确分类。

$\sum S(3,2)$ 代表区域 3 中第 2 类的样本总数, 分类归于第一类, 属错误分类。

$\sum S(4,1)$ 代表区域 4 中第 1 类的样本总数, 分类归于第二类, 属错误分类。

$\sum S(4,2)$ 代表区域 4 中第 2 类中的样本总数, 分类归于第二类, 属正确分类。

由以上分类知两类样本的总数为:

$$\sum S = \sum S(1,1) + \sum S(2,2) + \sum S(3,1) + \sum S(3,2) + \sum S(4,1) + \sum S(4,2) \quad (12)$$

由贝叶斯最优分类的定义, 有:

$$p(\omega_1) = \frac{\sum S(1,1) + \sum S(3,1) + \sum S(4,1)}{\sum S} \quad (13)$$

$$p(\omega_2) = \frac{\sum S(2,2) + \sum S(3,2) + \sum S(4,2)}{\sum S} \quad (14)$$

则在贝叶斯最优分类后, 在归于第一类的条件下, 分类错误的概率为:

$$p(\mathbf{x} \in R_1 | \omega_2) = \frac{\sum S(3,2)}{p(\omega_2)} \quad (15)$$

在归于第二类的条件下, 分类错误的概率为:

$$p(\mathbf{x} \in R_2 | \omega_1) = \frac{\sum S(4,1)}{p(\omega_1)} \quad (16)$$

贝叶斯最优分类准则是使错误率最小的分类规则, 即使两类样本的平均错误率 $P(e)$ 达到最小, 在分类时要求两类样本的错误率相同。

$$\frac{\sum S(1,1) + \sum S(3,1) + \sum S(4,1)}{\sum S} \cdot \frac{\sum S(4,1)}{p(\omega_1)} = \frac{\sum S(2,2) + \sum S(4,2) + \sum S(3,2)}{\sum S} \cdot \frac{\sum S(3,2)}{p(\omega_2)} \quad (17)$$

由此可得:

$$\sum S(3,2) = \sum S(4,1) \quad (18)$$

由式(18)可知: 在最优分类时, 两类交叉样本的错误分类样本数相同, 也就是使分类错误的样本数相同的分类标准就是最优的分类标准。由此也可得出结论: 去除 $\sum S(3,2)$ 和 $\sum S(4,1)$ 的样本对最优分类并无影响, 但经过修正的两类样本却由线性不可分变为线性可分, 改善了两类样本间的特性。

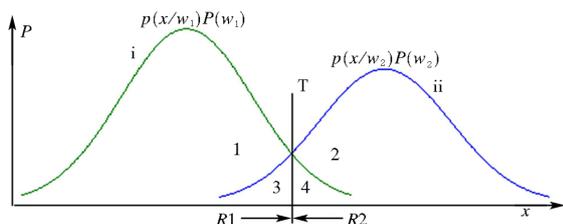


图1 贝叶斯样本分类
Fig.1 Classification using Bayes method

3.2 两类方法优点的结合

贝叶斯分类法是“最优”分类法, 具有优良的分类特性。但使用它要先知道样本的分布概率和条件概率分布, 这在很多情况下难以做到, 使它无法应用。

使用感知准则函数分类简便易行, 但无法用于有交叉的线性不可分样本。

使用最小错分样本数准则分类可用于线性不可分的样本, 但它的计算复杂程度也是显而易见的。实际工程中, 面对动辄数万个节点的分类, 又要求快速、实时, 计算机也难以承受。

改进方法的使用过程为: 先使用贝叶斯最优分类法对样本进行处理, 去掉交叉样本, 使有交叉样本变为无交叉样本, 再使用感知准则函数分类。这样感知准则函数分类的都是无交叉样本, 无需分类有交叉样本。改进的方法在计算的复杂程度上有明显的不同。两种方法的分类过程如图 2、3 所示。

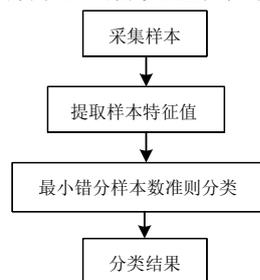


图2 使用最小错分样本数准则分类

Fig.2 Classification using rule of least wrong samples

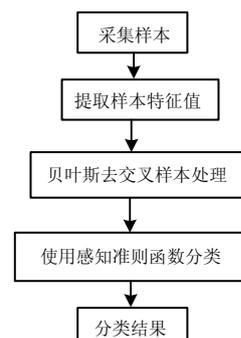


图3 使用新方法分类

Fig.3 Classification using new method

4 仿真分析

使用 Matlab 仿真, 如图 4 所示: 有 6 种电能质量扰动待分类: ①电压脉冲; ②电压缺口; ③电压膨胀; ④电压暂降; ⑤电压中断; ⑥无扰动。由图 4 可见, 有相邻边界的两类扰动之间易产生样本交叉现象, 若两类样本的特征值不相邻, 则难以发生样本交叉现象。

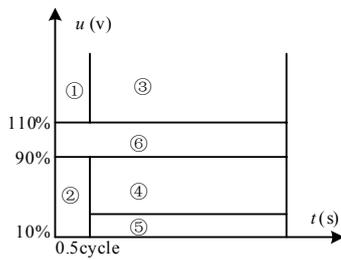


图 4 暂态扰动的分类

Fig.4 The classification of transient disturbances

上述的 6 类电能质量扰动类型共取 1 000 个样本，使用上节提出的两种方法分类。分类的结果见表 1 所示。

表 1 两种分类方法的结果对比

Tab.1 The classification result of two methods

	分类准确率/(%)	计算使用时间/s
原方法	98.8	12
新方法	99.2	1

表 1 中的计算使用时间是在启动计算时开始计时，计算结束时完成的人工计时，误差在 0.5 s 内。

5 结论

由两种分类方法的比较结果来看，在分类的准确率上，改进法的准确率较高，但差别不是很大。改进法使用“最优”分类后，再使用较为简单的方法分类，整个过程没有复杂运算，对提高分类的准确率有帮助。最小错分样本数准则分类法的复杂计算会降低计算的精度。

在计算使用时间的比较上，改进法有较大的优势，这是因为最小错分样本数准则分类法的计算复杂，需要大量时间，而改进法在“最优”的基础上极大地简化了计算，缩短了计算时间，这在大规模结点计算并有快速实时的要求时，改进法有很大的优势。

在计算存储量的比较上，从改进法的计算方法上可看出，占用的总空间相对于待分类样本数据并不多，而最小错分样本数准则分类法则要占用较大的计算存储空间，这在大规模数据运算时，是必须考虑的一个因素。本文没有对这方面做实际比较。

参考文献

[1] 朱桂萍, 王树民. 电能质量控制技术综述[J]. 电力系统

自动化, 2002, 26 (19): 28-31.

ZHU Gui-ping, WANG Shu-min. A Survey on Power Quality Control Technology[J]. Automation of Electric Power Systems, 2002, 26 (19): 28-31.

[2] 文继锋, 刘沛. 一种电能质量扰动的检测和分类方法[J]. 电力系统自动化, 2002, 26 (1): 42-44.

WEN Ji-feng, LIU Pei. A Method for Detection and Classification of Power Quality Disturbance[J]. Automation of Electric Power Systems, 2002, 26 (1): 42-44.

[3] 王成山, 王继东. 基于小波包分解的电能质量扰动分类方法[J]. 电网技术, 2004, 28 (15): 78-82.

WANG Cheng-shan, WANG Ji-dong. Classification Method of Power Quality Disturbance Based on Wavelet Packet Decomposition[J]. Power System Technology, 2004, 28 (15): 78-82.

[4] 杨光正, 吴岷, 张晓莉. 模式识别[M]. 合肥: 中国科学技术大学出版社, 2001.

[5] 黄德双. 神经网络模式识别系统理论[M]. 北京: 电子工业出版社, 1996.

HUANG De-shuang. System Theory of Neural Networks Pattern Recognition[M]. Beijing: Publishing House of Electronics Industry, 1996.

[6] 张文涛, 王成山. 基于改进贝叶斯分类法的电能质量扰动分类[J]. 电网技术, 2007, 31 (7): 22-25.

ZHANG Wen-tao, WANG Cheng-shan. Classification of Power Quality Disturbances Based on Improved Bayes Method[J]. Power System Technology, 2007, 31 (7): 22-25.

[7] 张文涛, 王成山. 基于概率神经网络的电压暂态扰动分类[J]. 电力系统自动化, 2007, 31 (18): 34-38.

ZHANG Wen-tao, WANG Cheng-shan. Classification of the Voltage Transient Disturbances Based on Probabilistic Neural Networks[J]. Automation of Electric Power Systems, 2007, 31 (18): 34-38.

收稿日期: 2009-03-19; 修回日期: 2009-04-05

作者简介:

张文涛(1972-), 男, 博士研究生, 工程师, 主要从事电能质量方面的研究; E-mail: wtzhang@yahoo.com.cn

潘瑾(1966-), 男, 高级工程师, 主要从事电器技术方面的研究;

饶群(1968-), 女, 华北院A级专家, 主要从事电能质量方面的研究。