

基于支持向量机的动态电能质量扰动分类方法

宋晓芳¹, 陈劲操²

(1. 国电自动化研究院, 江苏 南京 210003;
2. 南京理工大学 动力工程学院, 江苏 南京 210094)

摘要: 将支持向量机 SVM(Support Vector Machine)引入到动态电能质量分类问题中。在 Matlab 中编程建立了谐波、电压暂升、电压跌落、瞬时中断、电压波动、瞬变 6 种常见动态电能质量扰动数学模型,利用傅里叶变换和小波变换对产生的样本波形进行特征提取,产生训练和测试样本。给出了利用 LIBSVM 解决电能质量扰动分类问题的步骤,并根据分类结果对影响分类效果的参数进行了分析。对训练好的支持向量分类器进行测试,效果良好,当采用 C-SVC,RBF 核时调整参数可以得到最优分类效果,最高分类率可达到 96.67%。

关键词: 动态电能质量; 支持向量机; 分类方法; 多类分类

中图分类号: TM 71;TP 181 **文献标识码:** A **文章编号:** 1006-6047(2006)04-0039-04

常见的动态电能质量扰动有电压暂升、电压跌落、谐波及电压瞬变等,动态电能质量扰动问题已成为电力系统中的重要研究课题。到目前为止,人工神经网络^[1]、小波变换^[2-4]、基于尺度的变换方法(如傅里叶变换)等已被用于扰动类型分类。由于电能质量扰动涉及的特征量太多,分类判据复杂易错,因此大多采用数学变换与人工智能^[5]相结合的方法对扰动进行分类。

支持向量机 SVM(Support Vector Machine)是近年发展的一个统计学习算法,在解决小样本、非线性模式识别问题中表现出许多优势,并在许多实际问题中取得了很好效果。本文将 SVM 引入动态电能质量扰动分类,对谐波、电压跌落和波动等电能质量进行分类。

收稿日期:2005-06-14;修回日期:2005-10-13

1 支持向量机分类原理

SVM 是从线性可分情况下的最优超平面发展而来的,基本思想可用图 1 的二维情况说明。

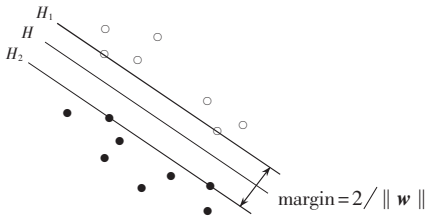


图 1 最优超平面

Fig.1 Optimal hyper-plane for classification

在二维分类的情况下,最优超平面简化成最优分类直线。

图 1 中,“○”和“●”分别表示 2 类可分的训练样

本, H 为分类面, 将 2 类样本无错误地分开, H_1, H_2 分别是过 2 类样本中离 H 最近的点且平行于 H 的直线, 推广到高维空间, 最优分类线就成为最优分类面。设分类线方程 $\mathbf{w} \cdot \mathbf{x} + \mathbf{b} = 0$, 将其归一化, 使得对线性可分样本集 $(x_i, y_i), i=1, \dots, n; \mathbf{x} \in R^d; y \in \{+1, -1\}$, 满足 $y_i[(x_i \cdot \mathbf{w}) + \mathbf{b}] - 1 \geq 0$ 。

此时分类间隔 (margin) 为 $2 / \|\mathbf{w}\|$, 使间隔最大的分类面就称为最优超平面, H_1, H_2 上的训练样本点就称作支持向量。

在线性可分的情况下, 构造最优超平面的问题被转化为优化问题, 引入拉格朗日函数可将优化问题转化为相应的对偶问题, 可求得优化问题的唯一解。

SVM 由训练样本集和核函数完全描述, 常用的核函数有

- 线性核函数: $K(\mathbf{x}, x_i) = (\mathbf{x} \cdot x_i)$;
- 多项式核函数: $K(\mathbf{x}, x_i) = [(\mathbf{x} \cdot x_i) + c]^q$, 其中 c, q 为常数, $q=1, 2, \dots$;
- 高斯核函数: $K(\mathbf{x}, x_i) = \exp[1 - (\mathbf{x} - x_i)^2] / \sigma^2$;
- Sigmoid 核函数: $K(\mathbf{x}, x_i) = \tanh[v(\mathbf{x} \cdot x_i) + c]$ 。

2 多类分类问题

支持向量机是针对 2 类分类问题提出的, 因此存在一个如何将其推广到多类分类的问题, 动态电能质量扰动分类就是一个典型的多类分类问题。对于多类分类问题, 可以进行如下描述。

根据给定的训练集如式 (3) 所示:

$$T = \{(x_1, y_1), \dots, (x_l, y_l)\} \in (X \times Y)^l \quad (1)$$

其中 $x_i \in X = R^n, y_i \in Y = \{1, 2, \dots, M\}, i=1, 2, \dots, l$, 寻找一个决策函数 $f(\mathbf{x}): X = R^n \rightarrow Y$ 。

由此可见, 求解多类分类问题, 实质上就是找到一个把 R^n 上的点分成 M 部分的规则。

目前, 对于多类分类问题的常见算法^[6]有一类对余类 (one against all), SVM 决策树, 确定多类目标函数方法 (multi-class SVM) 及成对分类 (one against one)。在本文中用的是成对分类方法, 下面简单介绍该算法的原理。

设有 k 类数据, 选取第 i 类数据和第 j 类数据构造一个分类器, 其中 $i < j$ (一般设 i 为正例, j 为负例), 这样需要训练 $k(k-1)/2$ 个分类器。对于第 i 和第 j 类数据, 需解决如式 (4) 表示的 2 分类问题。

$$\min_{\mathbf{w}^{ij}, \mathbf{b}^{ij}, \xi^{ij}} \frac{1}{2} (\mathbf{w}^{ij})^T \mathbf{w}^{ij} + C \sum_l \xi_l^{ij} \quad \xi_l^{ij} \geq 0 \quad (2)$$

$$(\mathbf{w}^{ij})^T \boldsymbol{\phi}(x_l) + \mathbf{b}^{ij} \geq 1 - \xi_l^{ij} \quad y_l = i$$

$$(\mathbf{w}^{ij})^T \boldsymbol{\phi}(x_l) + \mathbf{b}^{ij} \leq -1 + \xi_l^{ij} \quad y_l \neq i$$

用投票法解决上述问题, 若判断 $\mathbf{x}$ 属于第 i 类, 第 i 类的票数加 1; 反之, 第 j 类票数加 1。 $\mathbf{x}$ 最后归为拥有最多票数的类。这种投票法称为“Max Wins”法。但若 2 类具有同样的票数, 该方法就不太适用了, 这时可以选择索引值较小的类。

在实际问题中, 通常解式 (2) 的对偶问题, 因为它的变量个数和 2 类的数据个数相同。若平均每类有 $1/k$ 个数据点, 可以解决 $k(k-1)/2$ 个二次规划问题, 其中每个有大约 $2/k$ 个变量。

3 仿真算例

SVM 分类可以通过编程实现, 也可以通过已有的 SVM 软件实现。本文选用工具箱 LIBSVM^[7] 实现分类。LIBSVM 是 SVM 最常用的软件之一, 该软件可以解决 C -SVC 分类、 ε -SVR 回归等问题以及基于成对分类算法的多类模式识别问题。LIBSVM 主要用到 2 个程序, 训练建模 (svmtrain) 和使用已有的模型进行预测 (svmpredict), 此外还有数据归一化 (svmscale) 和交叉验证 (cross-validation)。

本文采用 LIBSVM 对谐波、电压暂升、电压跌落、瞬时中断、电压波动、瞬变 6 种电能质量扰动进行分类。各种电能质量扰动的数学模型可见文献[8], 这里只给出瞬变和跌落的数学模型。

a. 瞬变 (呈阻尼衰减的突变、暂态高频等信号):

$$u(t) = \begin{cases} U_0 \sin(\omega t + \alpha_0) & t < t_1, t > t_2 \\ U_0 \sin(\omega t + \alpha_0) + \sum_{i=2}^N A_i \sin(i\omega t + \alpha_i) + A e^{-\lambda t} & t_1 < t < t_2 \end{cases}$$

式中 $\sum_{i=2}^N A_i \sin(i\omega t + \alpha_i)$ 为高次谐波分量; $A e^{-\lambda t}$ 为直流衰减分量。

b. 电压跌落:

$$u(t) = \begin{cases} U_0 \sin(\omega t + \alpha) & t < t_1, t > t_2 \\ a U_0 \sin(\omega t + \alpha_1) & t_1 < t < t_2 \end{cases}$$

其中, $0.1 < a < 0.9$ 为典型跌落幅度; $0.01 \text{ s} < t_2 - t_1 < 1 \text{ min}$ 为典型跌落持续时间。

在实际的电力系统中, 电压信号总是受到不同程度的噪声影响, 所以在以上模型中需再叠加一定的噪声。

在 Matlab 中根据电能质量扰动的数学模型模拟产生所研究的 6 种扰动波形, 共产生 300 个样本波形, 每个样本波形包括 10 个工频周期。波形产生后对采样后的波形进行傅里叶变换、小波变换产生所需特征。

3.1 特征提取

选择如表 1 所示特征量用来对扰动波形分类。

表 1 电能质量特征量计算公式

Tab.1 Calculation formulas for unique features of power quality

特征量	计算公式
基波分量	$V_n = \sqrt{2} \mathbf{V}^n[1] / N$
相角偏移	$\alpha_n = \text{angle}(\mathbf{V}^n[1]) - \text{angle}(\mathbf{V}^1[1])$
小波系统能量	$EW_n = \sum_{k=1}^n WC^n[k] $
低频谐波畸变率	$TS_n = \frac{\sum_{k=2}^{10} \mathbf{V}^n[k] ^2}{ \mathbf{V}^1[1] }$
总谐波畸变率	$THD_n = \frac{\sqrt{\sum_{k=2}^{\text{int}(N/2)} \mathbf{V}^n[k] ^2}}{ \mathbf{V}^1[1] }$

表中, $V^n[k]$ 是样本信号的傅里叶变换, $V^n[k] = \sum_{i=0}^{N-1} v[i + (n-1)N] e^{-j(2\pi k i / N)}$, 其中 $v[i]$ 是样本信号采样值, $i=0, 1, \dots, L-1$, L 为信号长度, N 为每周期的采样点数, 取为 64, j 为虚数单位, $n=1, 2, \dots, 10$; WC^n 为采样信号每周期的小波系数, $k=1, 2, \dots, \eta$, η 为 WC^n 的长度; $\text{angle}(\cdot)$ 用于求输入量角度。

3.2 LIBSVM 解决电能质量扰动分类问题的步骤

利用 LIBSVM 实现电能质量扰动分类的具体流程如下:

- a. 将生成的特征量按 LIBSVM 软件使用的数据格式输入;
- b. 数据归一化, 即将训练和测试的特征输入量归一化到一个小的范围, 一般将特征量数据归一化到 $[-1 \ 1]$ 或者 $[0 \ 1]$, 本文将数据归一化到 $[0 \ 1]$;
- c. 参数寻优, 目前 LIBSVM 只能对 C -SVC, RBF 核函数的参数 C 和 γ 进行寻优, 对于其他核函数参数只能依靠经验或尝试法设定合适的值;
- d. 在 `svmtrain` 命令中设置 SVM 类型、核函数类型及惩罚系数等参数, 进行训练, 训练结束后产生模型文件, 文件中包括支持向量样本数、支持向量样本等必须的参数;
- e. 输入测试数据, 利用训练好的模型进行分类, 得到分类结果, 分类后得到分类正确率和均方根差值等。

在分类过程中, 扰动类型由数据的标识符确定, 跌落、暂升、谐波、瞬态中断、瞬变和波动的标识符类型依次定义为 1~6, 每个扰动样本数据由提取出的特征代替, 训练时输入样本标识符和特征。

3.3 分类结果分析

对于多类分类问题, 优先采用 C -SVC, RBF 核^[7] 进行分类。首先, 利用交叉验证对 C 和 γ 的值进行寻优, 然后利用搜索到得 C 和 γ 值进行训练, 训练结束后产生的模型文件, 利用模型文件对跌落、瞬变等测试数据进行测试, 表 2, 3 分别为电压瞬变、电压跌落的 10 组测试数据分类后的情况, 篇幅所限, 对于其他 4 种电能质量指标的分类情况不再给出。

表 2 C -SVC, RBF 核, $C=128, \gamma=2$ 时瞬变的 10 组测试数据结果
Tab.2 The classification results of transients by C -SVC, RBF kernel, $C=128$ and $\gamma=2$

样本组数	特征量					测试标识	输出标识
	V	α	EW	TS	THD		
1	0.2479	0.8316	0.0011	0.0192	0.0610	5	5
2	0.3117	0.6006	0.0003	0.0177	0.0610	5	1
3	0.2042	0.8956	0.0035	0.0216	0.0528	5	5
4	0.2482	0.8956	0.0038	0.0216	0.0528	5	5
5	0.2620	0.8985	0.0056	0.0226	0.0447	5	5
6	0.2095	0.5990	0	0.0237	0.0610	5	1
7	0.2805	0.8172	0.0046	0.0220	0.0488	5	5
8	0.2048	0.8331	0.0099	0.0255	0.0056	5	5
9	0.1859	1.0000	0.0082	0.0263	0.0244	5	5
10	0.2119	0.7791	0.0060	0.0237	0.0386	5	5

表 3 C -SVC, RBF 核, $C=128, \gamma=2$ 时对电压跌落的 10 组测试数据结果
Tab.3 The classification results of sags by C -SVC, RBF kernel, $C=128$ and $\gamma=2$

样本组数	特征量					测试标识	输出标识
	V	α	EW	TS	THD		
1	0.4555	0.5744	0.0825	0.0023	0.0023	1	1
2	0.4586	0.5739	0.0788	0.0021	0.0012	1	1
3	0.4652	0.5772	0.0700	0.0020	0.0014	1	1
4	0.4677	0.5749	0.0735	0.0018	0.0045	1	1
5	0.4737	0.5805	0.0846	0.0018	0.0017	1	1
6	0.4683	0.6271	0.1353	0.0040	0.0027	1	3
7	0.4798	0.5753	0.0703	0.0017	0.0018	1	1
8	0.4859	0.5783	0.0960	0.0019	0.0002	1	1
9	0.4938	0.5755	0.0753	0.0015	0.0004	1	1
10	0.4996	0.5669	0.1180	0.0063	0.0021	1	1

由表 2 可知, 对于所选择的电压瞬变数据, 有 2 组电压瞬变波形数据被误判为电压跌落; 由表 3 可知, 有 1 个电压跌落波形数据被误判为谐波。对大量测试数据统计得到分类正确率为 96.67%。

为分析 SVM 类型和核函数及其参数对分类的影响, 改变 SVM 类型、核函数类型等进行训练和测试, 对于 C -SVC, 考察其参数惩罚系数 $C(C>0)$, 对于 ν -SVC, 考察参数 $\nu(\nu \in (0, 1])$ 。分类结果见表 4, 5。

表 4 $\gamma=2.0$ 时惩罚系数 C 对不同核函数的分类结果的影响
Tab.4 The influence of penalty parameter on different kernel functions when $\gamma=2.0$

C	分类正确率/%			
	线性核	多项式核	RBF 核	Sigmoid 核
0.1	65.33	55.33	68.67	63.33
1	76.67	55.33	68.67	63.33
10	82.67	55.33	90.67	81.33
100	91.33	71.33	93.67	85.33
1000	92.67	79.33	92.67	92.00
10 000	92.00	92.67	92.00	92.67
100 000	92.00	92.00	87.33	92.67
最大分类率/%	92.67	92.67	93.67	92.67

表 5 ν 值改变对不同核函数分类结果的影响
Tab.5 The influence of ν on different kernel functions

ν	分类正确率/%			
	线性核	多项式核	RBF 核	Sigmoid 核
0.1	87.33	62.67	89.33	92.67
0.2	90.67	65.33	81.33	90.67
0.3	90.67	74.67	90.00	92.00
0.4	88.67	72.67	88.67	90.00
0.5	88.67	89.33	90.00	88.67
0.6	88.67	80.00	88.67	87.33
0.7	82.67	76.00	82.67	83.33
0.8	82.00	70.67	82.00	82.00
0.9	74.00	59.33	76.00	74.00
最大分类率/%	90.67	89.33	90.00	92.67

由表 4 可知, 惩罚系数不同对分类结果有较大影响, 适于一套样本的惩罚系数在某一范围内存在, 但当惩罚系数很大时, 分类结果趋于稳定。文献[9]研究表明当惩罚系数大于某一极限值时, 决策函数相同, 分类效果也相同。

事实上 LIBSVM 的交叉验证寻优时,也是在某一范围内寻求最适合的参数^[10],而寻求的范围也是要使用者确定。

由表 5 可知, ν 值改变对分类结果有影响,总之, ν 取 0.5 时,分类效果比较好。由 ν -SVC 定义^[6]可知,参数 ν 影响到“支持向量个数”和“间隔错误样本数点”,这 2 个参数都会影响分类结果。

由上可知,使用 SVM 分类器可对电能质量扰动指标进行有效分类,采用不同核函数的支持向量机,影响分类结果。对于本文的数据采用 C-SVC, RBF 核可以得到比较理想的分类效果。

4 结语

本文将支持向量机引入到动态电能质量分类问题中,利用傅里叶变换和小波变换对各种扰动样本波形进行特征提取,产生训练和测试样本。由分类结果可知:支持向量机可以用于电能质量扰动的分类,当采用 C-SVC, RBF 核时,调整参数可以得到最优分类效果,最高分类率可达到 96.67%。

若在一段时间内对电网中电压信号等间隔提取样本,提取特征量后,用支持向量机进行分类,对分类结果进行处理就可以统计出电网在这段时间内各种扰动发生的次数,进而得出电网的电能质量情况。这可以作为电能质量分析的后处理工具。

参考文献:

- [1] HUANG Jian-sheng, THONG N D. A neural fuzzy classifier for recognition of power quality disturbances [J]. **IEEE Transactions on Power Delivery**, 2002, 17 (2): 609-616.
- [2] SANTOSO S, POWERS J E, GRADY M W. Power quality disturbance data compression using wavelet transform methods [J]. **IEEE Transactions on Power Delivery**,

1997, 12(3):1250-1257.

- [3] WILKINSON W A, COX M D. Discrete wavelet analysis of power quality transient [J]. **IEEE Transactions on Power Systems**, 1996, 11(4):2038-2044.
- [4] SANTOSO S, GRADY M W, POWERS J E, et al. Power quality disturbance wavelet recognition using wavelet based neural classifier. Part 2: application [J]. **IEEE Transactions on Power Delivery**, 2000, 15(1):229-235.
- [5] LIAO Yuan, LEE Jong-beom. A fuzzy-expert system for classifying power quality disturbance [J]. **Electrical Power and Energy System**, 2004(26):199-205.
- [6] 邓乃扬, 田英杰. 数据挖掘中的新方法:支持向量机 [M]. 北京:科技出版社, 2004.
- [7] HSU Chih-wei, CHANG Chih-chung, LIN Chih-jen. A practical guide to support vector classification [EB/OL]. (2004-11-17) [2005-01-20]. <http://www.csie.ntu.edu.tw/~cjlin/papers/guide.pdf>.
- [8] MEI Xue, WU Wen-lin, CHEN Xiao-dong. Analysis of power quality variations [C] // **Proceeding of 2003 IEEE PES T&D Conference**. [S.l.]:[s.n.], 2003:41-44.
- [9] KEERTHI S S, LIN C J. Asymptotic behaviors of support vector machines with Gaussian kernel [J]. **Neural Computation**, 2003, 15(7):1667-1689.
- [10] CHANG Chih-chung, LIN Chih-jen. LIBSVM: a library for support vector machines (Version 2.7) [EB/OL]. [2005-01-20]. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.

(责任编辑:康鲁豫)



宋晓芳

作者简介:

宋晓芳(1980-),女,河北沧州人,硕士,研究方向为电能质量分析和电力系统控制 (E-mail:songxiaofang@126.com);

陈劲操(1956-),男,江苏南京人,教授,研究方向为电能质量控制、电力电子在电力系统中的应用。

Classification method of dynamic power quality disturbances based on SVM

SONG Xiao-fang¹, CHEN Jin-cai²

(1.Nanjing Automation Research Institute, Nanjing 210003, China;

2.Nanjing University of Science and Technology, Nanjing 210094, China)

Abstract: The SVM(Support Vector Machine) method is introduced to classification of power quality disturbances. The concerned disturbances, including voltage sags, swells, interruptions, switching transients, flickers and harmonics, are modeled with Matlab. The features of sample waves are extracted by Fourier transform and wavelet transform to form the training and testing samples. The steps of disturbance classification using LIBSVM are described. The influence factors are analyzed according to the classification result. The trained support vector classifier is tested and validated effective. When using C-SVC and RBF kernel, parameters can be adjusted to achieve the optimal effect, while the maximal classification ratio reaches 96.67%.

Key words: dynamic power quality; support vector machine; classification method; multi-class classification