

基于改进 SVM 模型的电能质量扰动分类

俞晓冬¹, 周栾爱²

(1. 山东轻工职业学院, 山东 济南 250353; 2. 青岛港湾职业技术学院, 山东 青岛 266404)

摘要: 提出一种改进的支持向量机模型, 对电能质量扰动进行分类。支持向量机(SVM)在对大规模样本集的训练和分类时, 需要占用大量内存, 耗时过高, 运算速度缓慢。针对这种情况提出一种改进的 SVM 模型: 将原始训练样本集应用粗糙集理论(RS)去除冗余信息, 然后在 SVM 中引入概率分布函数, 用一个小规模的样本集训练得到一个初始的分类器, 用这个初始分类器对大规模训练集进行修剪, 修剪后得到一个规模很小的约减集, 再用这个约减集进行训练得到最终的分类器。实验表明: 这种改进的 SVM 模型有效降低了训练样本集的规模, 提高了分类能力。

关键词: 电能质量; 支持向量机; 粗糙集; 缩减训练集; 小波变换

Classification method of power quality disturbances based on improved SVM model

YU Xiao-dong¹, ZHOU Luan-ai²

(1. Shandong Institute of Light Industry, Jinan 250353, China;

2. Qingdao Harbour Vocational and Technical College, Qingdao 266404, China)

Abstract: A new support vector machine(SVM) model is proposed to classify power quality disturbances. The improved SVM method improves the speed of classification when SVM treats the large training set. Firstly, using rough set(RS) theory to eliminate redundant information of the large initial training set. Secondly, utilizing a probabilities function in SVM, training an initial classifier with a small training set and pruning the large training set with the initial classifier to obtain a small reduction set. Then, training with the reduction set, final classifier is obtained. Experiments show that this method effectively reduces the training set and improves the classification ability.

Key words: power quality; SVM; RS; reduced training set; wavelet transform

中图分类号: TM714 文献标识码: A 文章编号: 1674-3415(2010)03-0015-05

0 引言

近几年来, 电能质量问题引起的矛盾越来越凸现^[1-2]。随着电网中非线性负载的大量增加和敏感电子设备的推广应用, 电能质量问题日益严重, 如以电力电子装置为代表的非线性负荷的使用、冲击性负荷的大量使用以及各种大型用电设备的起停等, 使电网中的谐波污染、三相电压的不对称性以及电压波动和闪变日趋严重, 而各种复杂的、精密的、对电能质量敏感的用电设备的不断普及, 对电能质量的要求也比以前高得多。近年来, 基于各种人工智能技术进行电能质量的分析研究^[3-4], 取得了一定的效果。其中支持向量机(Support Vector Machines, SVM)是以统计学习理论为基础发展起来的一种新的通用学习方法, 有效地解决了小样本、高维数、非线性等实际问题, 并克服了人工神经网络学习结构不确定和存在局部最优等缺点, 大大地提高了学习

方法的泛化能力。

传统SVM方法对大规模样本集的训练比较困难, 会导致系统资源占用过高、训练时间迅速增长、分类处理速度下降, 该文针对该问题提出了一种改进的SVM方法进行电能质量的分类, 有效删减训练集中的非有效样本, 取得了更优的分类效果。

1 支持向量机

支持向量机是近年来刚刚发展起来的一种新的机器学习方法^[5-9], 支持向量机算法最终将转化为一个二次型寻优问题, 从理论上讲得到的将是全局最优点, SVM的最优求解基于结构风险最小化思想, 因此比其他非线性函数逼近方法具有更强的泛化能力, 基于此, 使得SVM迅速发展和完善。目前, 支持向量机已经广泛用于解决分类、回归、异常值检测以及时间序列预测等问题, 表现出独特的优势和良好的应用前景。

1.1 支持向量机的数学原理

支持向量机主要是基于如下思想:首先选择一非线性映射 $\Psi(\cdot)$, 把 n 维样本向量从原空间 $(x_1, y_1), \dots, (x_n, y_n) \in R^n, y_i \in \{+1, -1\}$ 映射到特征空间, 在此高维特征空间中构造最优线性决策函数 $y(x) = \text{sgn}[\Psi(x)w + b]$ 。我们称 $y(x) = \text{sgn}[\Psi(x)w + b]$ 为支持向量机。在构造最优决策函数时, 利用了结构风险最小化原则, 同时引入了间隔的概念, 并巧妙地利用原空间的核函数取代了高维特征空间的点积运算, 避免了复杂计算。

对标准支持向量机, 优化问题为:

$$\min J(\omega, \xi) = \frac{1}{2} \omega \cdot \omega + c \sum_{i=1}^n \xi_i$$

$$\text{s.t. } y_i = [\Psi(x_i)\omega + b] \geq 1 - \xi_i$$

$$\xi_i \geq 0, i = 1, 2, \dots, n$$

其中: ξ_i 为误差, 即允许错分的松弛变量; C 为惩罚参数。用拉格朗日法求解上述优化问题, 支持向量机优化问题转化为下面的二次规划:

$$\max W(\alpha) = -\frac{1}{2} \sum_{i,j=1}^n \alpha_i y_i y_j K(x_i, x_j) \alpha_j + \sum_{i=1}^n \alpha_i$$

$$\text{s.t. } \sum_{i=1}^n \alpha_i y_i = 0$$

$$0 \leq \alpha_i \leq c, i = 1, 2, \dots, n$$

其中: $\alpha_i \leq 0$, 为Lagrange函数。这是一个不等式约束下的凸二次规划问题, 存在唯一解。

支持向量机理论提供了一种避开高维空间的复杂性, 直接用此空间的内积函数(既是核函数), 再利用在线性可分的情况下的求解方法直接求解对应的高维空间的决策问题。当核函数已知, 可以简化高维空间问题的求解难度。同时SVM是基于小样本统计理论的基础上的, 这符合机器学习的目的。而且支持向量机比神经网络具有较好的泛化推广能力。

1.2 存在的问题

虽然支持向量机得到了广泛的应用, 但仍存在一些问题制约着SVM的运用及发展, 比如: 对于样本容量比较大的情况, 支持向量机的运算速度会明显变慢, 不利于在线应用。

由于SVM是借助二次规划来求解支持向量, 而求解二次规划将涉及 m 阶矩阵的计算 (m 为样本的个数), 当 m 数目很大时该矩阵的存储和计算将耗费大量的机器内存和运算时间。对大规模样本集的训

练导致系统资源占用过高、训练时间迅速增长、分类处理速度下降。

针对此问题, 多种快速改进算法都曾被提出, 如文献[10-11], 这些方法在一定程度上发挥了作用。其中文献[10]的LSVM方法通过删减训练样本集中的无效样本, 缩小数据集规模, 从而降低训练代价。该方法首先从大规模样本集 L 中随机抽取一个小规模样本集 N , 然后用小规模样本集 N 训练得到初始的分类器, 用初始分类器修剪大规模样本集 L , 得到规模远小于原样本集的缩减集。修剪的具体方法是: 设初始分类器的分类超平面为 H , 对于 L 的任意一个样本 s , 设 s 与 H 的距离 (大于等于零) 为 d 。若 $-\varepsilon < d - 1 < \varepsilon$, 则保留此样本, 否则就删除它 (其中 $0 < \varepsilon < 1$ 为可以调整的阈值)。最后利用这个缩减集进行训练, 得到最终SVM分类器。它极大地减轻了学习负担, 但也存在着严重的问题: 初始随机小规模样本集的选取有很大的随机性, 难以保证分类的准确度。在初始集上建立的分类超平面很可能出现与实际最优分类面差距较大的情况, 此时上面的裁剪方法将导致部分本应该作为支持向量的样本被淘汰, 从而使最终分类结果的精度大幅下降。

2 改进的SVM方法

为了克服文献[10]的LSVM方法中初始随机小规模样本集的选取存在很大随机性的问题, 本文提出一种新的改进的支持向量机方法。本方法主要包括两部分: (1) 应用粗糙集理论对原始样本集进行预处理, 在保留重要信息的前提下消除多余的数据, 去除冗余信息, 形成新的样本集 L ; (2) 修改LSVM原有的裁剪方法^[11], 引入概率函数, 进一步删减新样本集 L 中的无效样本, 缩减数据集规模。最后利用这个缩减集进行训练, 得到最终SVM分类器。

(1) 对原始训练样本集用粗糙集理论进行预处理^[12-13]

粗糙集理论是一门新兴的智能信息处理技术, 它对各种不完整数据进行分析、推理, 发现数据间的关系, 提取有用特征, 简化信息处理。其最大的特点就是在保持分类能力不变的前提下, 能对大量数据进行信息约简。其优势就在于处理大数据集和去除冗余信息方面的强大能力, 有助于克服SVM在处理大数据集时的缺点。因为粗糙集理论无需对知识或数据的局部给与主观评价, 因而利用粗糙集理论进行约简是一种定量、客观的简化, 约简计算以严密的定理为基础。它显然要优于仿真验证而不是定理保证的其它一些方法。这样可以优势互补, 避免初始训练集选取的随机性, 提高分类的准确度。粗

糙集化简数据的详细方法见文献[15]。

(2) 对文献[10]的LSVM原先的裁剪方法进行修改^[11]

对LSVM原先的裁剪方法, 引入概率分布函数 $P(x)$ 。以距离值 $|d-1|$ 为参数, 按照概率分布函数 $P(x)$ 给出样本点的分类归属概率, 其中 $P(x)$ 应满足: 对任意非负实数 x_1, x_2 , 若 $x_1 < x_2$, 则 $P(x_1) > P(x_2) > 0$ 。在区间 $[0, U] (U > 0)$ 内产生一个随机值 ε ; 如 $\varepsilon < P(|d-1|)$, 则将样本 s 加入缩减训练集 S 中; 否则, 舍弃 s 。这样得到的缩减集包含了更广泛的有效信息。最后利用缩减训练集 S 进行 SVM 训练, 得到最终分类器, 提高分类的精确度。一般, $P(x)$ 可采用中心点为 0 的正态分布函数。

3 基于改进SVM模型的电能质量扰动分类

本文利用改进的SVM模型对6种电能质量扰动进行分类。这6种电能质量扰动的数学模型如下所示:

(1) 电压暂降

$$u(t) = \begin{cases} (1-\delta)\sin(\omega t) & t_1 \leq t \leq t_2, 0.1 < \delta < 0.9 \\ \sin(\omega t) & t < t_1, t > t_2 \end{cases}$$

$0.01s < t_2 - t_1 < 60s$

(2) 电压暂升

$$u(t) = \begin{cases} (1+\delta)\sin(\omega t) & t_1 \leq t \leq t_2, 0.1 < \delta < 0.3 \\ \sin(\omega t) & t < t_1, t > t_2 \end{cases}$$

$0.01s < t_2 - t_1 < 60s$

(3) 谐波

$$u(t) = \sin(\omega t) + a_3 \sin(3\omega t) + a_5 \sin(5\omega t) + a_7 \sin(7\omega t)$$

其中: a_3, a_5, a_7 分别是 3 次, 5 次, 7 次和 11 次谐波的幅值。忽略其它谐波。

(4) 电压波动

$$u(t) = (1 + a \sin(b\omega t)) \sin(\omega t)$$

$0.1 \leq a \leq 0.2, 0.1 \leq b \leq 0.4$

(5) 电压瞬变

$$u(t) = \begin{cases} \sin(\omega t + \delta) & t < t_1, t > t_2 \\ \sin(\omega t + \delta) + \sum_{i=2}^N A_i \sin(i\omega t + \delta_i) + Ae^{-\lambda t} & t_1 \leq t \leq t_2 \end{cases}$$

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

(6) 电压短时中断

$$u(t) = \begin{cases} 0.1 \sin(\omega t) & t_1 \leq t \leq t_2 \\ \sin(\omega t) & t < t_1, t > t_2 \end{cases} \quad t_2 - t_1 > 0.01s$$

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200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。

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^[14], 产生训练和测试样本。选择如表1所示特征量用来对扰动波形进行分类。

表1中, $V^n[k]$ 是样本信号的傅里叶变换, $V^n[k] = \sum_{n=0}^{N-1} V[i + (n-1)N] e^{-j(2k\pi)n/N}$, 其中: $V[i]$ 是样本信号采样值; N 为每周期的采样点数; j 为虚数单位; $n=1, 2, \dots, 10$; WC^n 为采样信号每周期的的小波系数, WC^n 由 $WC^n[k]$, ($k=1, 2, \dots, \eta$) 组成, η 为 WC^n 的长度。

表1 电能质量特征量计算公式

Tab.1 Calculation formulae for unique feature of power quality

特征量	计算公式
基波分量	$V_n = \sqrt{2} (V^n[1])/N $
低频谐波畸变率	$LHD_n = \frac{\sqrt{\sum_{k=2}^7 V^n[k] ^2}}{V^n[1]}$
总谐波畸变率	$THD_n = \frac{\sqrt{\sum_{k=2}^{m(N/2)} V^n[k] ^2}}{V^n[1]}$
小波系统能量	$EW_n = \sum_{k=1}^n WC^n[k] $
能量偏差	$EWD_n = \frac{(EW_n - EW'_n)}{\sum_{n=1}^{10} EW_n - EW'_n }$

在分类过程中, 扰动类型由数据的标识符确定, 电压暂降、电压暂升、谐波、电压短时中断、电压瞬变、电压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。

基于改进SVM模型的电能质量扰动分类的步

骤:

(1) 利用傅里叶变换和小波变换对样本波形进行特征提取。

(2) 将生成的特征量进行数据归一化, 形成原始训练样本集。

(3) 利用粗糙集理论对原始训练样本集进行约简, 去除冗余信息, 形成新样本集L, 见文献[15]。

(4) 在新样本集L中随机选取一小规模样本集N。

(5) 利用N进行SVM训练, 得到初始分类超平面H。

(6) 对样本集L中的每个样本s。

a) 计算出它到H的距离d及P(|d-1|), 其中, P(x)满足: $\forall x_2 > x_1 \geq 0, P(x_1) > P(x_2) > 0$;

b) 如 $\varepsilon < P(|d-1|)$, 则将样本s加入缩减训练集S中, ε 为大于0的随机值;

c) 否则, 舍弃s。

(7) 利用缩减的训练集S进行SVM训练, 得到最终分类器。

一般, P(x)可采用中心点为0的正态分布函数,

即 $P(x) = (\sqrt{2\pi}\sigma)^{-1} e^{-x^2/2\sigma^2}, -\infty < x < \infty$ 它能较好地

反映现实中误差的分布情况, 其中, 参数 σ 代表了对初始分类超平面的信任程度。 σ 越小, 那些偏离由初始分类面所确定的支持向量位置的样本点能保留到缩减训练集中的机会越小, 而处于这些位置的样本点被淘汰的概率也越低; 相反, σ 越大, 偏离的样本点进入缩减集的概率越高, 处于预测位置的样本点也越有可能被淘汰。

用缩减的训练样本集S训练的最终SVM分类器, 其分类时间更短, 分类精度更高。

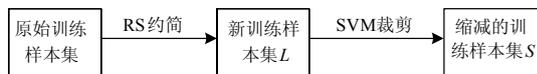


图1 训练样本集的约简步骤

Fig.1 Reduction steps of the training set

4 仿真结果

特征提取后产生的原始训练样本集共有 200 条规则, 利用 RS 软件 Rosetta 来进行约简。约简后的规则由原始样本集中的 200 条减少为现在的 125 条, 形成新样本集 L。可见新样本集 L 的规模大大减小。

首先从新样本集 L 中随机抽取 50 个样本作为小规模训练集(N), 训练得到初始 SVM 分类器, 然后利用初始分类器修剪样本集 L 得到缩减训练集 S, 最后利用缩减集 S 训练得到最终分类器。

本文采用径向基 (RBF) 核函数。最小二乘支

持向量机算法主要参数是惩罚参数 c 和径向基核函数参数 δ , c 值过小易对训练数据造成欠学习现象, 过大易对训练数据造成过学习现象而导致泛化能力下降; δ 值过小易对训练数据造成过学习现象, 过大则造成欠学习现象。这 2 个参数在很大程度上决定了 SVM 的学习能力。可以通过交叉验证等方法来优化参数。也可以采用网格搜索的方法: 如 $C=0:0.2:1, \delta=0:0.2:1$, 那么 C 与 δ 的组合就有 $6 \times 6=36$ 种, 对这 36 种组合训练支持向量机, 然后选择正确识别率最大的一组参数作为最优的 C 与 δ , 如果结果均不理想, 就需要重新考虑 C 与 δ 的范围。

采用 Matlab 工具箱 LS-SVMLAB 训练 SVM, 训练建模用函数 trainlssvm, 测试用函数 simlssvm, 交叉验证优化参数用函数 tunelssvm。

取 100 个样本作为测试集, 分别检测两种 SVM 模型的分类效果。由表 2 可以看出: 改进 SVM 方法与直接 SVM 相比, 提高了分类的准确率。

表 2 分类结果比较

Tab.2 Comparison of classification results

扰动类型	测试数据	正确分类数	
		改进 SVM	LSVM
谐波	100	100	97
电压暂降	100	95	92
电压暂升	100	89	86
电压短时中断	100	92	90
电压波动	100	85	82
电压瞬变	100	82	79
总计		543	526
(正确率)	600	(90.5%)	(87.6%)

5 结论

本文提出了一种改进的电能质量扰动分类的方法。传统 SVM 在处理大规模样本集时内存占用大, 分类速度慢。针对这种情况, 提出一种改进的 SVM 模型。首先应用粗糙集理论去除原始样本集的冗余信息, 然后在概率选择基础上对 SVM 训练集进行进一步的缩减, 实验表明, 这种方法有效降低了训练样本集的规模, 改善了 SVM 的分类效果。

参考文献

[1] Sabin D D, Sundaram A. Quality Enhances Reliability[J]. IEEE Spectrum, 1996.

[2] Olivier P, Pascal R, Michel M. Detection and Measurement of Power Quality Disturbances Using Wavelet Transform[J]. IEEE Trans on Power Delivery, 2000, 15(3): 1039-1044.

- [3] Kanitpanyacharoen W, Premrudeepreechacharn S. Power Quality Problem Classification Using Wavelet Transformation and Artificial Neural Networks[J]. IEEE Power Systems Conference and Exposition, 2004(3):1496-1501.
- [4] HUANG Jian-sheng, Thong N D. A Neural Fuzzy Classifier for Recognition of Power Quality Disturbances [J]. IEEE Trans on Power Delivery, 2002, 17 (2) : 609-616.
- [5] 胡国胜, 钱玲, 张国红. 支持向量机的多分类算法[J]. 系统工程与电子技术, 2006, 28 (1) :127-132.
HU Guo-sheng , QIAN Ling, ZHANG Guo-hong. Survey of Multi-classification Algorithms Based on Support Vector Machines[J]. Aerospace Electronics Information Engineering and Control, 2006, 28(1): 127- 132.
- [6] Weston J, Watkins C. Multi-class Support Vector Machines[A]. in: Proceedings of ESANN99[C]. Brussels(Belgium): 1999.
- [7] Hu G S, Chen Y T. Power System Reliability Evaluation Using Support Vector Machine[A]. in: Inter Symposium on Computational Intelligence and Industrial Applications[C]. Hainan(China): 2004.
- [8] Lee J, Lee D. An Improved Cluster Labeling Method for Support Vector Clustering[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2005, 27: 461-464.
- [9] Hsu Chih-wei, Lin Chih-jen. A Comparison of Methods for Multi-class Support Vector Machines[J]. IEEE Trans on Neural Networks, 2002, 13(2).
- [10] 李红莲, 王春花. 针对大规模的支持向量机的学习策略[J]. 计算机学报, 2004, 27(5): 716-718.
LI Hong-lian, WANG Chun-hua. A Learning Strategy of SVM Used to Large Training Set[J]. Chinese Journal of Computers, 2004, 27(5): 716-718.
- [11] 林杨, 刘贵全, 杨立身. 基于改进SVM方法的入侵检测[J]. 计算机工程, 2007, 33(14): 151-153.
LIN Yang, LIU Gui-quan, YANG Li-shen. Intrusion Detection Based on Improved SVM Algorithm[J]. Computer Engineering, 2007, 33(14): 151-153.
- [12] Jelonek J, et al. Rough Set Reduction of Attributes and Their Domains for Neural Networks[J]. Computational Intelligence, 1995, 11(2): 339-347.
- [13] 曾黄麟. 粗糙集理论及其应用[M]. 重庆: 重庆大学出版社, 1998.
ZENG Huang-lin. Rough Set Theory and Its Application[M]. Chongqing: Publishing House of Chongqing University, 1998.
- [14] Santoso S, Powers J E, Grady M W. Power Quality Disturbance Data Compression Using Wavelet Transform Methods[J]. IEEE Trans on Power Delivery.
- [15] YU Xiao-dong, ZANG Hong-zhi. Transformer Fault Diagnosis Based on Rough Sets Theory and Artificial Neural Networks[A]. in: International Conference on Condition Monitoring and Diagnosis[C]. Beijing: 2008.

收稿日期: 2009-02-25; 修回日期: 2009-06-14

作者简介:

俞晓冬(1974-), 女, 副教授, 研究方向为人工智能在电力系统中的应用等; E-mail:xiaodongyu2001@163.com
周荣爱(1973-), 女, 研究方向为电力系统及其自动化等。

(上接第 14 页 continued from page 14)

- [10] 祁才君, 王小海. 基于插值 FFT 算法的间谐波参数估计 [J]. 电工技术学报, 2003, 18(1) :92-95.
QI Cai-jun, WANG Xiao-hai. Interharmonics Estimation Based on Interpolation FFT Algorithm[J]. Transactions of China Electrotechnical Society, 2003, 18(1): 92-95.
- [11] 邹江锋, 刘涤尘, 谭子求, 等. 基于 FFT 算法的分次谐波测量与分析[J]. 高电压技术, 2003, 29(9): 29-31.
ZOU Jiang-feng, LIU Di-chen, TAN Zi-qiu, et al. The Analysis and Measurement of Power System Sub-harmonic[J]. High Voltage Engineering, 2003, 29 (9): 29-31.
- [12] 俞卞章. 数字信号处理 [M]. 西安:西北工业大学出版社, 2002.
YU Bian-zhang. Digital Signal Processing[M]. Xi'an: Northwestern Polytechnical University Press, 2002.

收稿日期: 2009-02-01; 修回日期: 2009-05-13

作者简介:

蒋春芳(1978-), 女, 硕士, 工程师, 从事电力系统可视化分析与计算、电力系统暂态稳定分析等工作;E-mail:jianglaoli@yahoo.cn
刘敏(1973-), 女, 硕士, 工程师, 从事电力系统保护控制及自动化方向研究。