基于遗传最小二乘小波支持向量机的分类研究

代龙泉　黄嘉亮　沈民奋
（汕头大学工学院电子工程系，广东　汕头　515063）
（07lqdai@stu.edu.cn）

摘要：最小二乘小波支持向量机（LS-WSVM）具有速度快和精度高的优点，但也存在模型参数过多，难以确定的问题。基于遗传算法和 LS-WSVM，本文提出了遗传最小二乘小波支持向量机（GA-LS-WSVM）。利用遗传算法优化 LS-WSVM 的参数，提高了辨识效果。两个实验仿真结果表明，在同等条件下，GA-LS-WSVM 的精度比神经网络和 LS-SVM 都更高，因而更适合与工程应用。

关键词：最小二乘　小波　支持向量机　遗传算法

Classification Research On Genetic Least Square Wavelets SVM

Longquan Dai, Jialiang Huang, Minfen Shen
(Electronic department　Shantou University　ShanTou　Guangdong　515063)
(07lqdai@stu.edu.cn)

Abstract：Although LS-WSVM has a lot of advantages, there are some stubborn problems that are open to discuss. In this paper, we proposed genetic least square svm based on genetic algorithm and LS-WSVM. Genetic algorithm is used to optimize the coefficients of LS-WSVM. Final two experiences show that GA-LS-WSVM is better than LS-SVM.

Key Words：least square wavelets；svm；genetic algorithm

支持向量机（Support Vector Machine，SVM）是一种新的通用的机器学习方法，它的理论基础是 Vapnik 等提出的统计学习理论，其基本思想是通过非线性变换将输入空间变换到一个高维空间，并在这个新空间中求取最优线性分类面，而这个非线性变换是通过事先选择的内积核函数实现的。与传统的学习方法（前馈型神经网络）相比，SVM 不存在局部极小、隐层节点数的选择和维数灾难等问题。最小二乘小波支持向量机（Least Squares Support Wavelet Vector Machine，LS-WSVM）结合了最小二乘法和小波和函数，是传统 SVM 的一种扩展。LS-WSVM 采用最小二乘线性系统作为优化函数，只受等式约束，求解过程变成了解一组等式方程，求解速度较快；另一方面，小波核函数不仅是近似正交的，而且适用于信号的局部分析、信噪分离和突变信号的检测，从而提高了支持向量机的泛化能力。

但是在应用上，LS-WSVM 仍然存在着选择模型参数的问题。LS-WSVM 的参数要比标准支持向量机多，而其参数对其性能影响很大。目前，也无统一的参数选取标准和理论。本文利用遗传算法优化 LS-WSVM 的参数，构成了遗传最小二乘小波支持向量机（Genetic Algorithm Least Squares Support Wavelet Vector Machine，GA-LS-WSVM）分类器。再通过两个实验来验证其有效性。

1. 支持向量机

设训练样本集为$[x_i(n)\quad y_i(n)]$，$xi(t)\in R^m$，$i=1,2,\cdots,l$。支持向量机的算法归结为一个二次规划问题：

$$\min\frac{1}{2}\langle w,w\rangle+C\sum_{i=1}^{l}\xi_i \quad s.t.\quad y_i(\langle w,w\rangle+b)-1+\xi_i\geqslant 0 \tag{1}$$

其中$\langle w,w\rangle$表示两向量之间的内积；$\xi_i\geqslant 0$为松弛项，表示错分样本的惩罚程度；C为惩罚因子，用于控制对错分样本惩罚的程度，实现错分样本数与模型复杂度之间的折衷；w 和 b 为决策函数$f(x)=\langle w,x\rangle+b$中的权向量和阀值。当无错分样本时，最小化目标函数的第一项等价于最大化两类间的间隔，可降低分类器的 VC 维，实现结构风险最小化。

引入 Lagrange 函数，上述二次规划问题的对偶形式为：

$$\max\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j K(x_i,x_j)\quad s.t.\quad \begin{array}{l}\sum_{i=1}^{n}y_i\alpha_i=0,C\geqslant\alpha_i^*,\\ \alpha_i\geqslant 0,i=1,\cdots,n\end{array} \tag{2}$$

式中α_i^*和α_i为 Lagrange 乘子，$K(x_i,x_j)$为核函数。依据支持向量机的稀疏性特征，优化问题（2）中只有少量的系数α_i不为零，α_i不为零所对应的样本(x_i,x_j)就是支持向量（Support ector）。由支持向量得到支持向量机的决策函数：

$$f(x)=\sum_{i=1}^{n}\alpha_i y_i K(x_i,x)+b \tag{3}$$

2. 最小二乘小波支持向量机

小波分析属于时频分析的一种，是当前数学中一个迅速发展的新领域。它具有多分辨分析的特点，通过伸缩和平移等运算功能对函数或信号进行多尺度细化分析，被誉为“数学显微镜”。在信号处理、图像压缩、语音编码、模式识别、地震勘探以及许多非线性科学领域内获得了广泛的应用。

若$h\in L^2\cap L^1$，且$\hat{h}(0)=0$，则按如下方式生成的函数族$\{h_{a,m}\}$是连续小波。

$$h_{(a,m)}(x)=|a|^{-1/2}h\left(\frac{x-m}{a}\right)\quad m\in R,a\in R,a\neq 0 \tag{4}$$

a为尺度因子或频率因子，m为平移尺度因子；$h(x)$为基本小波或母小波。对于能量有限信号或时间序列$f(x)\in L^2(R)$，其连续小波变换定义为：

$$W_f(a,b)=|a^{-1/2}\int_R f(x)\bar{h}\left(\frac{x-m}{a}\right)dx \tag{5}$$

式中，$\bar{h}(x)$是$h(x)$的复共轭函数。连续小波逆变换定义为：

$$f(x)=C_h^{-1}\iint a^{-2}W_f(a,m)h\left(\frac{x-m}{a}\right)dadm \tag{6}$$

式中，C_h为与$h(\bullet)$相关的常数。

小波变换的基本思想是利用一簇小波的叠加实现任意函数$f(x)$的表示。

利用张量积理论，可得到d维的分离小波函数的乘积：

$$h_d(x)=h_d(x_1,x_2,\cdots,x_d)=\prod_{i=1}^{d}h(x_i) \tag{7}$$

设$h(x)$是一个母小波，a、m分别为伸缩、平移尺度因子，那么小波核函数为：

$$K(x,x')=\prod_{i=1}^{d}h\left(\frac{x_i-m_i}{a_i}\right)h\left(\frac{yi-m_i'}{a_i'}\right) \text{其中} \begin{array}{l} x,x'\in R^d;a_i,a_i'\neq 0 \\ x_i,x_i',a_i,a_i',m_i,m_i'\in R \end{array} \quad (8)$$

选择常用的墨西哥草帽小波函数

$$h(x)=(1-x^2)\exp\left(-\frac{x^2}{2}\right) \quad (9)$$

来构造小波核函数，如下：

$$K(x,x')=K(x-x')=\prod_{i=1}^{d}h\left(\frac{x-x'}{a_i}\right)=\prod_{i=1}^{d}\left(1-\frac{\|x_i-x_i'|}{a_i^2}\right)\exp\left(\frac{\|x_i-x_i'\|^2}{2a_i^2}\right) \quad (10)$$

可以证明，墨西哥草帽小波核函数满足 Mercer 条件。因此，把式（10）代入式（3），可得小波支持向量机的决策函数：

$$f(x)=\sum_{i=1}^{l}\alpha_i\prod_{j=1}^{d}\left(1-\frac{\|x_{i,j}-x_j|^2}{a_{i,j}^2}\right)\bullet\exp\left(-\frac{|x_{i,j}-x_j\|^2}{2a_{i,j}^2}\right)+b \quad (11)$$

式中，$x_{i,j}$ 表示第 i 个训练数据的第 j 个分量，x_j 表示 x 的第 j 个分量，d 代表输入向量的维数。

式（11）便是基于小波核函数和 LS-SVM 的最小二乘小波支持向量机（Least Squares Wavelet Support Vector Machine，LS-WSVM）。本质上，LS-WSVM 是一种最小二乘支持向量机，它的核函数采用了小波核函数。

由于 LS-SVM 不能优化核函数的参数，选择 $l\times d$ 个参数 $a_{i,j}, i=1,\cdots,l; j=1,\cdots$，$d$ 是困难的。因而必须借助其他方法来确定参数。

3. 遗传最小二乘小波支持向量机分类器

支持向量机是建立在统计学习坚实的理论基础之上的，具有理论的完备性，但是在应用上，仍然存在一些问题，典型的问题就是模型参数的选择，目前，也无统一的模型选取标准和理论。支持向量机的参数对核空间特征提取算法及其性能影响很大。应用支持向量机之前，必须慎重考虑如何选取向量机的最优参数这个十分重要的问题。

对于由式（2-30）所表示的最小二乘小波支持向量机分类器，需要确定的参数有小波核函数参数 $a_{i,j}, i=1,\cdots,l; j=1,\cdots$，$d$，以及 LS-SVM 的惩罚因子 C，共 $l\times d+1$ 个。为了达到 LS-WSVM 的最佳性能，需要得到 $a_{i,j}$ 和 C 的最优组合。显然这是一个优化问题，若采取穷举的方式搜索最优值，效率低下，计算量将十分巨大以至于无法实现。由于遗传算法具有隐含的并行性和强大全局搜索能力，可以在很短的时间内搜索到全局最优点，因此利用遗传算法来优化 LS-WSVM 分类器的参数。

要应用遗传算法，对于特定问题需要设计特定的方案，以得到最佳性能。针对 LS-WSVM 的特点，本文采用的遗传算法的运行参数、编码方法、适应度函数、遗传算子等各组成部分如下：

（1）运行参数。

群体大小：20 个个体；

终止条件：运行 20 代，或分类正确率达到 100%；

交叉概率：0.4；

变异概率：0.1。

（2）编码。

C 和 $a_{i,j}$ 的取值范围都较大，$C\in[0.1,1000]$，$a_{i,j}\in[0.01,100]$。为了能保持较高精度，采

用浮点数编码。

在实验中发现，C 或 $a_{i,j}$ 值同样的变化，在其值小时比值大时对性能的影响要大。因此较好的方法应该使得搜索步长在值小时较小，而在值大时较大。作如下非线性变换：

$$C' = \lg C, a'_{i,j} = \lg a_{i,j} \tag{12}$$

以 C' 和 $a'_{i,j}$ 为优化对象，这样就能满足上述要求。

（3）适应度函数。

利用训练样本集对分类器进行学习训练，而用测试样本集来对训练后的分类器进行验证，测试结果的错误率越小，显然分类器的分类效果越理想。理想情况是对测试样本集完全正确分类。衡量分类器的分类精度采用识别率 RR (recognition rate)，其定义为：

$$RR = \frac{\text{分类正确的项目}}{\text{样本总数}} \tag{13}$$

如直接使用识别率 RR 作为适应度函数，由于识别率相对地一般都差别不大，导致生存压力不够。假设总样本数为 800 个，两个人口分别能正确识别 600 个样本和 610 个样本，RR 分别为 0.75 和 0.7625，十分接近，算法就近似于随机搜索而不是适者生存了。为增加生存压力，作如下线性变换：

$$\begin{aligned} &F(i) = RR(i) - \frac{\min(RR)}{3} \\ &i = 1, \cdots, 20 \end{aligned} \tag{14}$$

其中 $\min(RR)$ 为当代所有人口的 RR 最小值。

（4）选择算子。

采用比例选择算子，为保证算法的收敛性，必须采用最优保存策略。但传统的最优保存策略容易使得某个局部最优个体不易被淘汰掉反而快速扩散，从而使得算法的全局搜索能力不强。因此改进如下：

进化出新群体后找出当前群体的最佳个体 G2 与上一代的最佳个体 G1 比较。只有在 G1 比 G2 的适应度高时，才用 G1 替换当前群体中的最差个体。

与传统的保优策略有所不同的是，改进后的保优策略并不一定在每一代都用最佳个体换掉最差的个体，因此在保证算法收敛的同时有效地抑制了局部最优解的扩散。

（5）交叉算子。

采用算术交叉算子。

假设在两个体 X_1^t，X_2^t 之间进行算术交叉，则交叉运算后所产生出的两个新个体是：

$$\begin{cases} X_1^{t+1} = aX_1^t + (1-a)X_2^t \\ X_2^{t+1} = aX_2^t + (1-a)X_1^t \end{cases} \quad 0 < a < 1 \tag{15}$$

本文取 $a = 0.5$。

（6）变异算子。

采用非均匀变异算子。

在对第 n 代的基因 C'_n 进行非均匀变异时，取值范围为［-2，4］，则新的基因值 C'_{n+1} 由下式确定

$$C'_{n+1} = \begin{cases} C'_n + (4 - C'_n)(1 - r^{(1-t/T)b}) & if \quad random(0,1) = 0 \\ C'_n + (C'_n + 2)(1 - r^{(1-t/T)b}) & if \quad random(0,1) = 1 \end{cases} \tag{16}$$

式中，r 为［0，1］范围内符合均匀概率分布的一个随机数，t 是当前代数，T 是最大进化代数，b 是一个系统参数，它决定了随机扰动对 t 的依赖程度。

至此就设定好遗传算法的各个组成部分了。用遗传算法来优化最小二乘小波支持向量机

的参数，就构造了遗传最小二乘小波支持向量机分类器。

4. 遗传最小二乘小波支持向量机的性能

为验证最小二乘小波支持向量机的有效性，以下设计方形分类问题和双螺旋分类问题两个实验来测试它的性能，并把它和神经网络与最小二乘支持向量机做横向比较。

最小二乘支持向量机的核函数使用最常用的径向基核。由式（2-3）和（2-21）可得最小二乘支持向量机分类器的决策函数：

$$f(x)=\sum_{i=1}^{l}\alpha_i\bullet\exp\left(-\frac{1}{2\sigma^2}\|x_i-x\|^2\right)+b,\sigma>0 \tag{17}$$

其中需要确定的参数有惩罚因子 C 和径向基核参数 σ。对于只有两个变量的优化问题，可以使用网格法，步骤如下：

（1）确定参数 C 和 σ 的取值范围。依据最小二乘支持向量机的原理，$C\in[10^{-1},10^{3}]$，$\sigma\in[10^{-1},10^{3}]$。

（2）在最大取值范围内选取参数值，构建参数对 (C,σ)。把取值范围用对数表示，即 $C\in[10^{-1},10^{3}]$，$\sigma\in[10^{-1},10^{3}]$，$C$和$\sigma$各均匀取9个值，$\{10^{-1},10^{-0.5},10^{0},10^{0.5},10^{1},10^{1.5},10^{2},10^{2.5},10^{3}\}$，构成 9×9 网格和 81 个参数对。

（3）分别输出每个参数对到 LS-SVM，取识别率 RR 最高的参数对 (C,σ) 为最优参数对。

4.1 方形分类问题

（1）构造数据集。

在直角平面 xoy 上，在正方形 $x^2\leqslant 0.5,y^2\leqslant 0.5$ 内定义为一类，标记为“+1”；在正方形外定义为一类，标记为“-1”。

训练样本 80 个。在方形内外按均匀分布随机各选取 40 个点。

测试样本 400 个。在区域 $x^2\leqslant 1,y^2\leqslant 1$ 内按均匀分布随机选取 400 个点。

（2）实验结果。

图 4-1 是用于训练向量机的数据点集，共 80 个点。理想分类面是一个正方形。分别用“。”和“+”表示方形里外两类样本。

图 4-2 是 2-4-1 BP 神经网络的分类结果。其分类效果一般，在图中具体表现为其分类面内，上侧包含了较多的“。”点，属于分类错误的样本。

图 4-3 是 LS-SVM 的分类结果。其分类效果要优于 BP 神经网络，但其分类面接近圆形，与理想分类面差别较大，泛化能力不足。

图 4-4 是 GA-LS-WSVM 的分类结果。其错分样本最少，分类效果较好。由于小波核有很强的对信号的局部分析和突变检测的能力，因此其分类面最接近理想分类面方形。

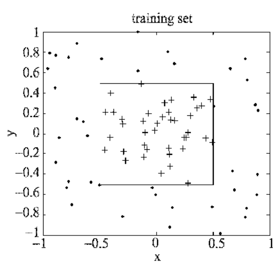


图 4-1　方形分类问题：训练集

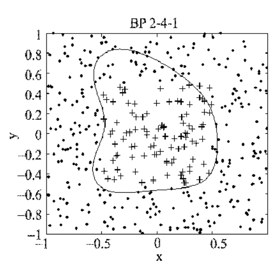
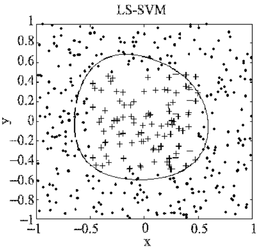


图 4-2　方形分类问题：2-4-1 BP 神经网络分类结果



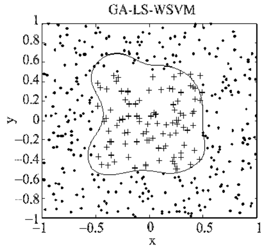


图 4-3 方形分类问题：LS-SVM 分类结果　　图 4-4 方形分类问题：GA-LS-WSVM 分类结果

表 4-1 是在方形分类问题上的结果。可以看到三种方法的识别率都达到了 90%以上，其中 GA-LS-WSVM 的性能最佳，达到 97.25%。

用网格法找到 LS-SVM 的参数是 $C=0.32,\sigma=3.16$。GA-LS-WSVM 的惩罚因子 $C=0.29$，而由于小波核函数的参数 $a_{i,j}$ 共有 $80*2*400=64000$ 个，因此不在这里列出。

表 4-1　方形分类问题

方形分类问题	正确样本数	识别率	参数
BP 2-4-1	364 个	91%	—
LS-SVM	377 个	94.25%	$C=0.32,\sigma=3.16$
GA-LS-WSVM	389 个	97.25%	$C=0.29$

4.2　双螺旋分类问题

（1）构造数据集。

在极坐标系中，把曲线 K_1：$r=\theta$ 定义为一类，标记为“+1”；上的子集；把曲线 K_2：$r=-\theta$ 定义为另一类，标记为“-1”。

训练样本 156 个。令 $\theta=i\pi/2, i=1,\cdots,12$，以及区间 $[i\pi/2,(i+1)\pi/2]$ 的 $i+1$ 等分点。这样在 K_1、K_2 上各有 78 点，共 156 点训练数据。

测试样本 400 个。随机选取 400 个 θ 值，两类各 200 点。

（2）实验结果。

图 4-5 画出了双螺旋分类问题的训练数据集。实线代表曲线 K_1：$r=\theta$，实线上的点是标记为“+1”类的训练数据；虚线代表曲线 K_2：$r=-\theta$，虚线上的点是标记为“-1”类的训练数据。

图 4-6 是 2-50-1 BP 神经网络的分类结果。分别用“。”和“x”来表示被正确分类和错误分类的样本。2-50-1 BP 神经网络不适用于这个分类问题，大量样本被错误划分，其分类面是一些自圆心往外的放射线，完全不能识别两类样本。

图 4-7 是 LS-SVM 的分类结果。其分类效果较为准确。图中的一条螺旋线是 LS-SVM 的分类面，基本上能把两类样本划分开来。但在原点附近，两类样本的距离很近，LS-SVM 出现了划分错误，图中表现为原点附件有少量的“x”点。这也说明了 LS-SVM 的分辨能力还不够高。

图 4-8 是 GA-LS-WSVM 的分类结果。GA-LS-WSVM 的分类而同样是一条螺旋线。得益于小波核的微观分析能力，GA-LS-WSVM 对于原点附近的两类样本都能识别，完美的把两类样本 K_1、K_2 分离开来。没有一个样本被错误划分，分类效果极佳。

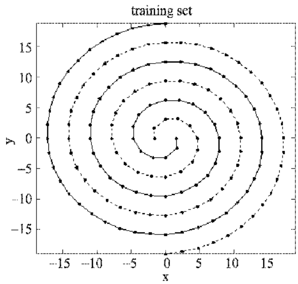


图 4-5 双螺旋分类问题：训练集

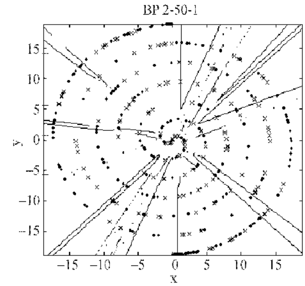


图 4-6 双螺旋分类问题：2-50-1 BP 神经网络分类结果

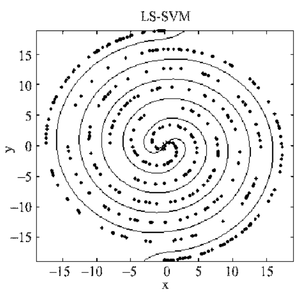


图 4-7 双螺旋分类问题：LS-SVM 分类结果

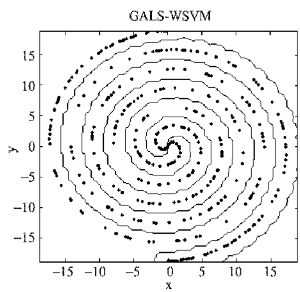


图 4-8 双螺旋分类问题：GA-LS-WSVM 分类结果

从表 4-2 可以看出，BP 神经网络在双螺旋分类这个问题上性能不好，识别率只刚刚过 50%。而 GA-LS-WSVM 的识别率达到了 100%！用网格法找到 LS-SVM 的参数是 $C=100, \sigma=3.16$。GA-LS-WSVM 的惩罚因子 $C=93.59$，此处同样没列出 GA-LS-WSVM 的 $156*2*400=124800$ 个小波核函数参数 $a_{i,j}$。

表 4-2 双螺旋分类问题

双螺旋分类问题	正确样本数	识别率	参数
BP 2-50-1	216 个	54%	—
LS-SVM	385 个	96.25%	$C=100, \sigma=3.16$
GA-LS-WSVM	400 个	100%	$C=93.59$

5. 结论

由以上两个实验的结果，可以得出以下结论：

对实验 1 和 2，本文算法均获得了非常好的仿真结果。GA-LS-WSVM 对两个实验数据集的识别率均达到了 90%以上，整个学习过程不需要人工干扰，大大提高了模式识别的自动化程度和精度。

对于双螺旋这样的经典分类难题，许多文献均进行了详细的研究，通常认为用常规的 BP 网络很难获得好的泛化特性。而通过本文算法，得到最小二乘小波支持向量机分类器的最优参数$(C, a_{i,j})$，识别率达到 100%。整个学习过程也自动完成。由此可见，与 BP 神经网络和

LS-SVM 相比较，GA-LS-WSVM 具有明显的优势，更适合于工程应用。

本文受国家自然科学基金（60571066，60271023），广东省自然科学基金重点项目资助。

参 考 文 献

[1] Schǒlkopf B, Smola A. Learning with kernels: support vector machines, regularization, and beyond [R] . Cambridge , MA:MIT Press, 2002.

[2] Goldberg D. Genetic Algorithms in Search, Optimization and Machine Learning [M] . Addison-Wesley, Reading, MA, 1989.

[3] Suykens J A K, Vandewalle J. Least squares support vector machine classfiers [J] . Neural Processing Letters, 1999, 9:293 -300.

[4] Smits G F, Jordan E M. Improved SVM Regression using Mixtures of Kernels [J] . IEEE, 2002, 3:2785 -2790.

[5] Li Y, Ren Y, Shan XM. Radar HRRP Classification with Support Vector Machines [C] . Proceedings of the 2001 International Conference on Info-tech and Info-net, 2001, 1: 218 - 222.

[6] Nello Cristianini, John Shawe-Taylor. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods [R] .电子工业出版社，2004.

[7] 周明，孙树栋. 遗传算法原理及应用. 国防工业出版社，1999.

[8] Suykens J A K. Nonlinear Modelling and Support Vector Machine [A] .IEEE Instrumentation and Measurement Technology Conference [C] . Budapest ,Hungary :2001. 21–23.

[9] ZHU J Y, REN B , ZHAN G H X , et al .Time series prediction via new support vector machine [A] . Proceedings of the First International Conference on Machine Learning and Cybernetics [C] . 2002 : 364 -366.