

子波核函数网络*

张莉 周伟达 焦李成

(西安电子科技大学雷达信号处理重点实验室, 陕西, 西安, 710071)

摘要 提出一种子波核函数网络作为支撑向量机的一种替代学习机, 仿真实验验证了子波核函数网络的逼近性能和识别性能都可以与相应的支撑向量机相媲美, 并优于子波神经网络。

关键词 支撑向量机, 子波核函数网络, 函数逼近, 模式识别, 子波神经网络。

WAVELET KERNEL FUNCTION NETWORK*

ZHANG Li ZHOU Wei-Da JIAO Li-Cheng

(National Key Laboratory for Radar Signal Processing, Xidian University, Xi'an, Shaanxi 710071, China)

Abstract A new wavelet kernel function network (WKFN) was proposed as an alternative of support vector machine (SVM). Simulation experiments show that the performance of approximation and recognition of WKFN is equal to that of the corresponding SVM, but gets an advantage over that of a general neural network.

Key words support vector machine, wavelet kernel function network, function approach, pattern recognition, wavelet neural network.

引言

支撑向量机是一种新的学习机, 在函数逼近和模式识别中已得到成功应用^[1,2]。但支撑向量机的学习性能很大程度上依赖于不易求解的有约束的二次凸规划, 对大规模问题更是如此, 不但精度难于保证, 而且计算复杂, 计算量非常庞大。另外, 不管是模式识别支撑向量机还是回归估计支撑向量机, 核函数自带的参数没有办法得到优化, 只能凭经验选择。

神经网络^[3]由于采用非线性节点函数, 具有很强的非线性处理能力和各种优化学习算法(如梯度下降法、共轭梯度法、序贯训练方法等), 但是神经网络对高维样本的处理能力缺乏有效的方法, 训练结果容易陷入局部极小。

能不能把支撑向量机和神经网络的优点结合起来, 消除它们的一些不足, 这是本文工作的出发点。子波神经网络在模式识别和函数逼近方面已经得到广泛的应用并取得了良好的效果^[4,5]。在支撑向量机中径向基函数也得到了很好的应用, 这主要得益于其良好的非线性映射能力, 我们知道许多子波函数

同样具有良好的映射能力, 其傅里叶维数为无穷维。本文中选用的子波就是这样的一种子波函数, 同时可选作核函数的子波函数需满足 Mercer 条件, 下文将作具体分析。

1 子波核函数网络

1.1 支撑向量机及支撑向量核

支撑向量核是用于支撑向量机上的一种核函数, 通过该核函数的映射, 可以把低维的输入模式空间映射到高维的特征空间。在这个特征空间中, 所要解决的问题由非线性可分(可逼近)变成了线性可分(可逼近), 比如对分类问题, 在输入空间分界面是曲线或非曲面, 而在特征空间分界面是超平面。

对模式识别问题, 支撑向量机需要求解下面的有约束对偶二次规划^[2]:

$$\text{最大化 } W(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j K(x_i \cdot x_j), \quad (1)$$

约束 $\sum_{i=1}^l \alpha_i y_i = 0$, 其中 $\alpha_i \in [0, C], i = 1, \dots, l$ 。输入空间对分类问题可构造如下形式的非线性决策函数

* 国家“863”计划(编号 863-306-ZT06-06-1)和高等学校博士点专项基金资助项目

稿件收到日期 2000-05-17, 修改稿收到日期 2001-02-19

* Supported by the National “863” Project (No. 863-306-ZT06-06-1) and the National Education Ministry of Doctor Fund

Received 2000-05-17, revised 2001-02-19

$$f(x) = \text{sgn}\left(\sum_{i=1}^l a_i y_i K(x, x_i) - b\right). \quad (2)$$

其中 $\{(x_i, y_i) | x_i \in R^N, y_i \in \{1, -1\}, i=1, \dots, l\}$ 表示训练样本对, $K(x, x_i)$ 正是先验选择的支撑矢量核, x 是检验样本, $x_i \geq 0$ 是权系数, b 是阈值.

相应地,对逼近问题的优化问题则有^[1]

$$W(a^{**}) = -\varepsilon \sum_{i=1}^l (a_i^* + a_i) - \sum_{i=1}^l (a_i^* - a_i) y_i - \frac{1}{2} \sum_{i,j=1}^l (a_i^* - a_i)(a_j^* - a_j) K(x_i, x_j), \quad (3)$$

约束 $\sum_{i=1}^l (a_i^* - a_i) = 0$, 其中 $a_i^{**} \in [0, \frac{\varepsilon}{\gamma}]$, a_i^{**} 表示 a_i 和 a_i^* . 决策函数为

$$f(x) = \sum_{i=1}^l (a_i - a_i^*) K(x, x_i) + b. \quad (4)$$

这里训练样本为 $\{(x_i, y_i) | x_i \in R^N, y_i \in R, i=1, \dots, l\}$, 权系数变成 $(a_i - a_i^*)$. 在这两种情况中, 只要权系数不为零, 则相应的训练样本就被称为是支撑矢量.

通常的方法是利用数学规划的方法来求解大规模二次规划问题, 很明显, 随着训练样本的增多, 二次凸规划的求解复杂度也大大增加. 这里, 我们利用神经网络来优化权值, 可以避免求解这一大规模二次规划问题.

由于核函数的引入使支撑矢量机有了非线性处理的能力, 但这种核函数需要满足一定的条件, 即 $K(x, x') = (\Phi(x) \cdot \Phi(x'))$. 一般支撑矢量核有两种形式: 点积形式核函数 $K(x, x') = K(\langle x, x' \rangle)$ 和平移不变形式核函数 $K(x, x') = K(x - x')$. 这两种核函数形式都必须满足上述核函数条件的等价充要条件 (Mercer 条件)^[1,2]. 由于本文只涉及平移不变形式的核函数, 在此我们以定理的形式给出容许条件, 它与相应的 Mercer 条件是等价的.

定理 1^[1] 一种平移不变核 $K(x, x') = K(x - x')$ 是一种可容许支撑矢量核, 当且仅当其傅里叶变换

$$F[k](\omega) = (2\pi)^{-N/2} \int_{R^N} \exp(-i(\omega \cdot x)) K(x) dx \quad (5)$$

的值是非负的, 其中 $x \in R^N, i = \sqrt{-1}$.

定理 1 无论是在检验一个核是否是可容许支撑矢量核时, 还是在构造新的可容许支撑矢量核时, 都是很有用的工具.

1.2 子波核函数

记母子波函数为 $h(x)$, 伸缩为 a , 且 $x, a \in R$. 则

$x, x' \in R^N$ 满足平移不变核定理子波核为

$$K(x, x') = \prod_{i=1}^N \left| h\left(\frac{x_i - x'_i}{a_i}\right) \right|. \quad (6)$$

由此我们能够得到子波支撑矢量机的估计函数 (用于函数逼近)

$$f(x) = \sum_{i=1}^l (a_i - a_i^*) \prod_{j=1}^N h\left(\frac{x'_j - x_j}{a_j}\right) + b \quad (7)$$

和决策函数 (用于分类)

$$f(x) = \text{sgn}\left(\sum_{i=1}^l a_i y_i \prod_{j=1}^N h\left(\frac{x'_j - x_j}{a_j}\right) + b\right), \quad (8)$$

其中 x'_j 表示第 i 个训练样本的第 j 个分量.

当然, 由于子波函数要满足严格的定理 1 的允许条件, 才能作为核函数使用, 所以并不是所有的子波函数都可作为核函数. 幸运的是, 通过神经网络得到广泛使用的子波函数为 $h(x) = \cos\left(\frac{1.75x}{a}\right) \exp\left(-\frac{x^2}{2a^2}\right)$. 我们验证了它的平移不变形式是满足容许条件的, 下面我们以定理的形式给出.

定理 2 若母波函数为 $h(x) = \cos(1.75x) \exp\left(-\frac{x^2}{2}\right)$, 伸缩因子为 a , 且 $a, x \in R$, 变量 $x, x' \in R^N$. 则子波核

$$K(x, x') = \prod_{i=1}^N \left| h\left(\frac{x_i - x'_i}{a_i}\right) \right| = \prod_{i=1}^N \left(\cos\left(1.75 \times \frac{x_i - x'_i}{a_i}\right) \exp\left(-\frac{\|x_i - x'_i\|^2}{2a_i^2}\right) \right) \quad (9)$$

是一种可容许的支撑矢量核.

定理 2 的证明见附录, 在此需指出: 可以把子波核看作是一种多维子波函数. 我们的子波支撑矢量机也就是在多维子波基上张成的空间中, 用优化的方法来找到最佳的子波系数, 从而得到最佳的估计函数或分类决策函数.

1.3 子波核函数网络模型

如图 1 所示, 子波核函数网络是一类前向网络. 假设训练样本为 $\{(x_i, y_i) | x_i \in R^N, y_i \in R, i=1, \dots, l\}$, 则可用如下方法构造相应的网络模型: 对输入层, 输入变量 $x = (x_1, \dots, x_N) \in R^N$, 我们把输入变量直接馈入子波核函数中, 通过函数作用, 把输入空间映射到特征空间, 使得能够以线性的方式来解决; 对隐层, 隐层的单元数为 l , 选用隐层单元的非线性函数为子波函数式 (9). 输入层到隐层之间没有权值联结, 隐层到输出层之间的权值为 W ; 对输出层, 单元数为 1, 分类问题的输出函数表达式为

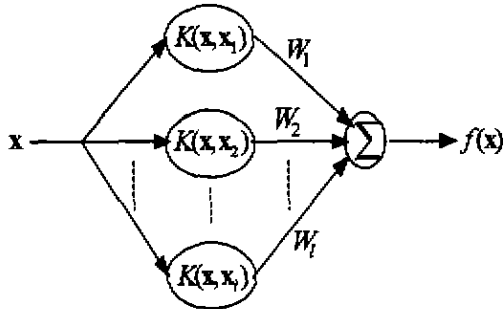


图 1 子波核函数网络

Fig. 1 Wavelet kernel function network

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^l W_i K(x, x_i) + b\right), \quad (10)$$

而对逼近问题,其输出函数可描述为

$$f(x) = \sum_{i=1}^l W_i K(x, x_i) - b, \quad (11)$$

上式中, $K(x, x_i) = \prod_{j=1}^N \cos\left[\frac{1.75(x_j - x'_j)}{a_i}\right] \exp\left[-\frac{\|x_j - x'_j\|^2}{2a_i^2}\right]$, 这里 x_i 表示第 i 个训练样本的第 j 个分量, b 表示阈值。

我们提出的子波核函数网络的数学模型式(10)和(11)及支撑矢量机决策函数式(2)和(4)是等价的,它们都是把输入空间的变量通过核函数映射到特征空间,然后在特征空间进行线性的识别或逼近。子波核函数网络与支撑矢量机优化算法有所不同,支撑矢量机采用凸二次规划优化,子波核函数网络则可以采用有效神经网络学习方法^[3],如梯度下降法,共轭梯度法等。在本文中,为了试验子波核函数的性能,我们采用最基础的神经网络学习算法,对于网络权值 W_i ,分类问题我们采用感知器学习算法,函数逼近问题我们采用 BP 算法,对于伸缩因子 a_i ,我们可以根据先验知识来设置,也可以用训练权值的方法来优化,这些算法是最简单和最通用的神经网络学习算法,由于构造了新非线性函数,因而收到了很好的效果,如果采用一些改进措施和新的学习算法,可取得更为优良的结果。另外,由于子波核函数网络在输入层与隐层之间权值固定为 1,整个网络可明确分为映射过程和分类过程两部分,而其分类过程只相当于单层感知器,可以证明采用均方误差最小学习算法的子波核函数网络没有局部极小点。

子波核函数网络优越于一般神经网络使网络的规模不会随着输入模式维数的增长而呈指数增长(即维数灾难问题),网络隐层的单元数只与训练样本的个数有关,而与输入模式的维数无关,在训练好

网络后,我们可以根据网络的权值 W_i 来删减或修剪网络的结构,理论上,如果 $W_i = 0$, 则就可以把第 i 个单元删掉,实际学习中,或多或少存在着一些误差,不可能精确地为 0,我们可以给出一个小的阈值来减小网络的复杂度,并使网络具有良好的间隔。

2 仿真实验

这里,我们给出两个仿真例子,一个是一维函数逼近的例子,另一个是多类飞机一维图像识别的例子,同时给出子波核函数网络和支撑矢量机的性能比较。

2.1 我们对下式的一维非线性函数进行逼近:

$$f(x) = \sin x + \frac{1}{3} \sin 3x - 2 \sin \frac{x}{2}, \quad x \in [0, 2\pi] \quad (12)$$

对式(12)的函数曲线进行间隔采样,取 63 个样本,其中用 32 个作为训练样本,31 个作为检验样本,在相同条件下,分别用子波核函数网络和支撑矢量机对式(12)的非线性函数进行逼近,先设定伸缩因子和迭代次数,结果见表 1,表中的逼近误差^[1]由下面的表达式计算得到:

$$\delta = \sqrt{\frac{\sum_{i=1}^l (y_i - \bar{y}_i)^2}{\sum_{i=1}^l (y_i - \bar{y})^2}}, \quad \text{且 } \bar{y} = \frac{1}{l} \sum_{i=1}^l y_i, \quad (13)$$

图 2 和图 3 分别是两种方法估计曲线的结果,由表 1 可知,这两种方法的逼近误差有相同的数量级,从而验证了子波核函数网络的逼近性能和支撑矢量机是近似等价的。

2.2 飞机一维像识别

我们分别用子波核函数网络和支撑矢量机对 3 类飞机的一维像进行识别,3 类飞机模型分别是 B-52、歼-6 和歼-7,这些数据都是在微波暗室中测得的,测量了这 3 种飞机从 $0^\circ \sim 160^\circ$ 方位角的一维距离像,表 2 列出了用子波-核函数网络和支撑矢量机对 3 种飞机的分类结果,可以看出,子波核函数网络的分类性能和支撑矢量机是差不多的,两种方法都能够完成对不同飞机的识别。

在用网络方法对高维问题进行分类时,通常都要进行预处理,即对样本数据进行降维处理,而这里

表 1 一维非线性函数逼近结果

Table 1 Approximation of one-dimensional function

算法	伸缩因子	迭代次数	逼近误差
子波核函数网络	$a=1$	1000 次	0.0104
支撑矢量机	$a=1$	1000 次	0.0237

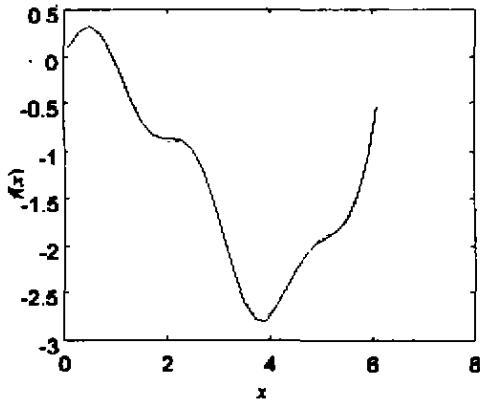


图2 支撑向量机估计函数
Fig. 2 Approximation result of SVM

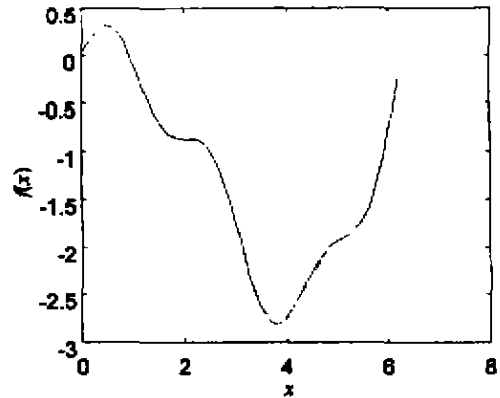


图3 子波核函数网络估计函数
Fig. 3 Approximation result of WKFN

的子波核函数网络是不需要进行降维处理的,也就是说,高维输入对该网络的性能没有影响.这里,一维像识别问题的输入空间是64维的,文献[6]中提出了一种局域进化子波网络来对之进行识别,为了与文献[6]的方法进行比较,我们将3种飞机的一维像分成两组来进行识别,识别结果见表3.文献[6]得到的识别率分别为94.3%、84.6%和93.2%,而这里得到的识别率则分别为98.87%、92.92%和94.23%,都高于进化子波网络的识别结果.子波核函数网络中尺度参数 α 的选取,可以在网络训练过程中用算法进行优化^[6],但本文提出的子波核函数

网络中,从实验结果看, α 的选取可以在很大一个范围内,均能使训练误差达到最小,所以没有必要进行优化,只是作经验性的选取.一般 α 只要取一个能使子波函数主瓣与样本分布的区域直径的1/6同等数量的值即可.

3 结论

本文提出了一种子波核函数网络,它结合了支撑向量机和神经网络的一些优点.这种方法一方面沿袭了核映射的特点,把模式输入空间映射到高维的特征空间,在此空间中对样本进行非线性的逼近

表2 两种算法对3种飞机的分类结果
Table 2 Classification results of three-class planes

算法	伸缩因子	飞机种类	训练样本		检验样本	
			识别率	误识率	识别率	误识率
子波核函数网络	$\alpha=0.1$	B-52	100%	0%	94.17%	5.83%
		歼-6	100%	0%	88.33%	11.67%
		歼-7	100%	0%	83.33%	16.67%
支撑向量机	$\alpha=0.1$	B-52	100%	0%	98.75%	1.25%
		歼-6	100%	0%	88.75%	11.25%
		歼-7	100%	0%	80.83%	19.17%

表3 子波核函数网络的识别结果
Table 3 Recognition results of WKFN

方位角	参数	飞机种类	训练样本		检验样本	
			个数	识别率	个数	识别率
$0^\circ \sim 100^\circ$	$\alpha=2$	B-52	51	100%	200	99%
		歼-6	41	100%	160	92.5%
		歼-7	56	100%	224	94.64%
$80^\circ \sim 160^\circ$	$\alpha=5.5$	B-52	30	100%	118	98.73%
		歼-6	31	100%	120	93.33%
		歼-7	47	100%	186	93.82%

或识别,同时它的网络结构可根据支撑矢量的概念加以调整,该网络具有前向网络的结构,其中的非线性函数采用了子波核函数,可以用子波核函数网络在核映射后的特征空间中进行逼近或识别,避免了求解有约束的二次凸规划问题,并且该网络不存在局部极小点,该网络最大的优点是网络规模不会随着输入模式维数的增长而呈指数增长(即维数灾难问题),网络隐层的单元数只与训练样本的个数有关,而与输入模式的维数无关,文中仿真实验验证了子波核函数网络的逼近性能和分类性能都可以和相应的支撑向量机相媲美,而且要优于子波函数网络。

附录: 定理 2 的证明

证明: 根据定理 2,我们只要证明

$$F[K](\omega) = (2\pi)^{-N} \int_{R^N} e^{-j\omega x} K(x) dx \geq 0, \quad (14)$$

这里用 j 来表示虚数单位, $K(x) = \prod_{i=1}^N h(\frac{x_i}{a_i}) = \prod_{i=1}^N \cos$

$(\frac{1.75x_i}{a_i}) e^{-\frac{1}{2a_i^2} x_i^2}$, 先计算积分项

$$\begin{aligned} & \int_{R^N} \exp(-j\omega x) K(x) dx \\ &= \prod_{i=1}^N \frac{|a_i|}{2} \sqrt{2\pi} \left(\exp\left(-\frac{(1.75 - \omega_i a_i)^2}{2}\right) + \exp\left(-\frac{(1.75 + \omega_i a_i)^2}{2}\right) \right). \end{aligned} \quad (15)$$

把式(15)代入式(14),即可得到子波核的 Fourier 变换为

$$F[K](\omega) = \left(\frac{|a_i|}{2} \right)^N \prod_{i=1}^N \left(\exp\left(-\frac{(1.75 - \omega_i a_i)^2}{2}\right) + \exp\left(-\frac{(1.75 + \omega_i a_i)^2}{2}\right) \right), \quad (16)$$

由式(16),有

$$F[K](\omega) \geq 0, \quad (17)$$

从而定理 2 得证。

REFERENCES

- [1] Smola A, Sholkopf B. A tutorial on support vector regression. *Neuro COLT, Rep.*, 1998, **19**. Available <http://svm.first.gmd.de>
- [2] Burges C. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 1998, **2**(2): 1-47
- [3] JIAO Li-Cheng. *Theory of Neural Networks System*. Xi'an: Xidian University Press (焦李成, 神经网络系统理论. 西安: 西安电子科技大学出版社), 1990.
- [4] Zhang Q, Benveniste A. Wavelet Networks. *IEEE Trans. Networks*, 1992, **3**(6): 889-898
- [5] Szu H, Telfer B, Kadambe S. Neural network adaptive wavelets for signal representational and classification. *Opt. Eng.*, 1992, **31**: 1907-1916
- [6] ZHANG Yan-Ning. Study of the intelligent object recognition method. Xidian University, Post-doctor thesis (张艳宁, 智能目标识别方法研究. 西安电子科技大学博士后研究报告), 1999, 61-63