

# 基于粒子群支持向量机的海杂波序列回归预测

苏昭斌, 朱人杰, 陈红卫

(江苏科技大学 电子信息学院, 江苏 镇江 212003)

**摘要:** 在雷达数据处理中, 为更好地抑制海杂波, 预测海杂波是必要的; 海杂波具有混沌特性, 而支持向量机算法能够有效地对混沌序列进行回归预测, 文章提出了一种改进的支持向量机海杂波序列回归预测算法; 文中给出了算法的框架结构, 采用了互信息法和改进的伪邻近点法提取海杂波混沌特性的延迟时间和嵌入维数, 利用相空间重构求取 SVM 训练样本, 应用改进的 PSO 算法优化 SVM 的核函数参数以及惩罚系数, 并仿真了预测模型; 仿真实验结果表明: 海杂波回归预测能达到满意的精度, 而 PSO-SVM 方法比 SVM 方法的预测精度更高。

**关键词:** 海杂波; 混沌; 支持向量机; 粒子群; 回归预测

## Sea Clutter Sequences Regression Prediction Based on PSO-SVM Method

Su Zhaobin, Zhu Renjie, Chen Hongwei

(School of Electronics and Information, Jiangsu University of Science and Technology, Zhenjiang 212003, China)

**Abstract:** In the radar data processing, in order to better suppress sea clutter, sea clutter prediction is necessary. Sea clutter has chaotic features, and support vector machine algorithm can effectively carry out regression prediction of chaotic sequences, this paper proposes an improved support vector machine regression sea clutter sequence prediction algorithm. The paper gives the algorithm framework, using a mutual information method and improved pseudo-point method to extract neighboring sea clutter chaotic time delay and embedding dimension, using phase space reconstruction to strike SVM training sample, applying improved PSO algorithm to optimize kernel function parameter and penalty coefficient of SVM and simulate a prediction model. Simulation results show that: regression model to predict sea clutter is feasible and PSO-SVM method can further improve the prediction of sea clutter prediction accuracy.

**Key words:** sea clutter; chaos; SVM; PSO; regression model

### 0 引言

1995 年 Simmon Haykin 在 IEEE 上首次详细介绍了对海杂波混沌特性的研究情况, 并断言海杂波具有混沌特性, 而不是随机的一个全新论断<sup>[1]</sup>。混沌的短期可预测性为海杂波序列的回归预测提供了可能性。

支持向量机 (support vector machines, SVM) 是 Corinna Cortes 和 Vapnik 等在 1995 年首先提出的统计学习理论<sup>[2]</sup>。统计学习理论采用结构风险最小原则, 在最小化样本点误差的同时, 缩小模型泛化误差的上界, 从而提高了模型的泛化能力, SVM 理论正是在这一基础上发展起来的, 经过几十年的发展, 该理论已经开始逐步应用于模式识别、函数拟合、时间序列预测等领域。另外, 由于 SVM 法不依赖于系统的数学模型, 同时具有自学习自调整模型的特点, 因此能对混沌时间序列产生较好的预测效果<sup>[3]</sup>。

但 SVM 法中的参数影响预测效果, 一般采用实验法确定 SVM 参数。但实验法费事费力, 而且不易找到最优的参数。粒子群优化算法能够比较容易找到群体最优解, 且简单实用, 目前已经广泛用于参数优化, 模式分类, 神经网络训练等领域。本文从雷达实测数据中提取海杂波混沌特性, 采用相空间重构得到 SVM 训练源, 应用改进的 PSO 算法优化 SVM 的参数, 最终建立回归模型, 最后利用回归模型预测海杂波并与实

测海杂波进行比较。

### 1 海杂波序列回归预测框架

图 1 是本文改进的粒子群算法优化支持向量机 (PSO-SVM) 参数进行海杂波序列回归预测方法的整体框架结构。以下围绕此框架进行。

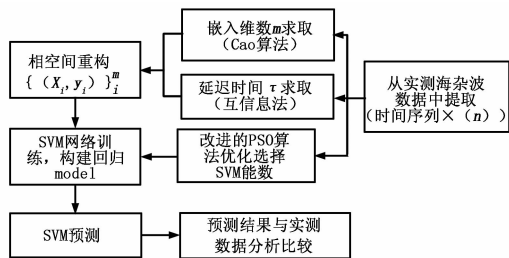


图 1 PSO-SVM 算法框架

### 2 相空间重构

相空间重构是一种研究非线性时间序列的有效方法, 其基本思想是: 在高维空间中, 为重构系统的状态空间, 可以利用一个分量的时间序列, 将它在某些时间的延迟点上的观测值作为新的坐标, 由它们共同确定多维状态空间的点, 重构相空间与原相空间拓扑等价。海杂波序列  $\{x(n), n = 1, 2, \dots, N\}$ , 相空间重构数学表达式如式 (1) 所示。

$$X(n) = (x(n), x(n+\tau), \dots, x(n+(m-1)\tau))$$

$$n = 1, 2, \dots, M \tag{1}$$

式中,  $x(t)$  是单变量的测量值;  $\tau$  为延迟时间;  $m$  为嵌入位数;  $M$  为相点数  $M = N - (m - 1)\tau$ 。

收稿日期: 2013-12-27; 修回日期: 2014-02-24。

基金项目: 船舶预研支撑技术基金项目 (12J3.3.4)。

作者简介: 苏昭斌 (1988-), 男, 硕士研究生, 主要从事计算机应用方向的研究。

对于单变量构成的海杂波时间序列  $\{x(n), n = 1, 2, \dots, N\}$ , 如果能适当选定嵌入维数和延迟时间, 就能进行相空间重构, 重构后变为  $G = \{(X_i, y_i)\}_{i=1}^m$  形式, 作为 SVM 的训练源。

### 2.1 延迟时间 $\tau$ 的确定

延迟时间  $\tau$  若太小, 将造成相空间矢量的相邻分量过于接近无法区分, 从而无法提供独立的坐标分量; 若取值太大, 则两坐标在统计意义上完全独立, 导致混沌吸引子失去相关性。本文采用互信息法计算延迟时间  $\tau$ , 其表达式为式 (2)<sup>[4]</sup>:

$$M(x_i, x_{i-\tau}) = \sum_{i,j} p_{i,j}(\tau) \ln \frac{p_{i,j}(\tau)}{p_i p_j} \quad (2)$$

式中,  $p_i$  为  $x_i$  在第  $i$  个时间间隔中的概率;  $p_{i,j}(\tau)$  为点  $x_i$  在  $t$  时刻落入第  $i$  个时间间隔和  $t + \tau$  时刻落入第  $j$  个时间间隔的联合概率。 $M$  的大小描述了已知  $x(t)$  的情况下,  $x(t + \tau)$  的确定性。为了简化计算, 重构时选用  $M$  的第一个极小值作为延迟时间参数  $\tau$ 。

### 2.2 嵌入维数 $m$ 的确定

计算嵌入维数较常使用的方法是伪邻近点法 (False neighbors method, FNM)。FNM 法的主要思想: 在高维相空间中并不相邻的两点可能投影在一维空间轴上, 即伪邻近点, 这也是混沌时间序列呈现出无规律的原因。而随着嵌入维数的增大, 混沌吸引子逐渐打开, 伪邻近点也会被剔除, 从而将混沌运动的轨迹得到恢复。但是 FNM 算法存在对噪声敏感、主观选择性强等缺点, 因此在此采用 Cao Liangyue 提出改进的 FNN 算法 (Cao 法)<sup>[5]</sup>, 此方法具有: 1) 除延迟时间外不包含其它主观参数; 2) 对数据长度依赖性不强, 使用较小数据量就可以求得的优点。

Cao 算法主要引入了  $E_1(m)$  和  $E_2(m)$  两个变量。对于时间序列  $x_1, x_2, \dots, x_N$ , 当延迟时间为  $\tau$ , 嵌入维数为  $m$  时, 空间第  $i$  个向量  $Y_i(m)$  及其最邻近点  $Y_i^{NN}(m)$  分别为:

$$Y_i(m) = \{x_i, x_{i+\tau}, \dots, x_{i+(m-1)\tau}\} \quad (3)$$

$$Y_i^{NN}(m) = \{x_i^{NN}, x_{i+\tau}^{NN}, \dots, x_{i+(m-1)\tau}^{NN}\} \quad (4)$$

定义  $a_2(i, m)$ 、 $E(m)$  和  $E_1(m)$  3 个变量如式 (5) ~ (7):

$$a_2(i, m) = \frac{\|Y_i(m+1) - Y_i^{NN}(m+1)\|}{\|Y_i(m) - Y_i^{NN}(m)\|} \quad (5)$$

$$E(m) = \frac{1}{N - m\tau} \sum_{i=1}^{N-m\tau} a_2(i, m) \quad (6)$$

$$E_1(m) = E(m+1)/E(m) \quad (7)$$

如果时间序列是确定的, 则嵌入维是存在的, 即  $E_1(m)$  将在  $m$  大于某一特定值  $m_0$  后将不再变化。若时间序列是随机信号, 则  $E_1(m)$  应逐渐增大, 但在实际应用中序列  $E_1(m)$  是否缓慢变化已不容易, 因此补充第二判断准则为:

$$E^*(m) = \frac{1}{N - m\tau} \sum_{i=1}^{N-m\tau} |x(i + m\tau) - x^{NN}(i + m\tau)| \quad (8)$$

$$E_2(m) = E^*(m+1)/E^*(m) \quad (9)$$

由于随机数据间没有相关性,  $E_2(m)$  将恒等于 1; 而对于海杂波序列, 数据之间的相关性是依赖于嵌入维数的, 因此总存在一些  $m$  使得  $E_2(m) \neq 1$ 。因此, 可以通过  $E_2(m)$  判定信号数据是随机还是确定序列。

## 3 粒子群优化支持向量机模型

### 3.1 支持向量机<sup>[6]</sup>

支持向量回归算法的基本思想是通过一个非线性映射  $\Phi$ , 将数据  $x$  映射到高维的特征空间  $F$ , 并在该空间进行线性回

归, 因此, 对经过相空间重构后的数据, 可以用式 (10) 进行回归估计:

$$f(X) = \langle \omega, \Phi(X) \rangle + b \quad (10)$$

式中,  $\omega$  和  $b$  表示回归因子, 可以通过最小化式 (11) 的正则化风险泛函获得:

$$R(\omega) = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^m L(f(X_i), y_i) \quad (11)$$

式中, 系数  $C$  称为惩罚系数,  $L(\cdot)$  表示损失函数, 通常取一次不敏感损失函数  $L_\epsilon(f(x), y)$  如式 (12) 所示。

$$L_\epsilon(f(x), y) = \max(|f(x) - y| - \epsilon, 0) \quad (12)$$

最小化  $R(\omega)$  可得  $\omega = \sum_{i=1}^m (\alpha_i + \alpha_i^*) \Phi(X_i)$ , 其中  $\alpha_i, \alpha_i^*$  是最小化  $R(\omega)$  对偶问题的解, 将  $\omega$  代入式 (10) 得

$$f(x) = \sum_{i=1}^m (\alpha_i + \alpha_i^*) \langle \Phi(X_i), \Phi(X) \rangle + b = \sum_{i=1}^m (\alpha_i + \alpha_i^*) k(X_i, X) + b \quad (13)$$

式中,  $k(X_i, X) = \langle \Phi(X_i), \Phi(X) \rangle$  称为核函数, 它是对称的正实数函数, 同时满足 Mercer 条件。常用的核函数有径向基核函数  $k(x_i, x) = \exp\{-|x - x_i|^2 / 2\sigma^2\}$  和多项式核函数  $k(x_i, x) = [(x, x_i) + c]^q, q \in \mathbf{N}, c \geq 0$  等, 文中采用径向基核函数法。

当  $b$  取在边界上的一点, 便可根据 KKT 最优条件定理进行计算, 但出于稳定性考虑, 一般取边界上的平均值

$$b = \text{average}_k \{\delta_k + y_k - \sum_{i=1}^m (\alpha_i + \alpha_i^*) k(X_i, X)\} \quad (14)$$

式中,  $\delta_k$  为预测误差。

通过对 SVM 模型的分析可知, 在支持向量回归估计算法中核函数的类型、核函数的参数以及惩罚系数都是很重要的参数, 为获得更好的推广能力有必要对这些参数进行优化调整。

### 3.2 粒子群优化支持向量机

#### 3.2.1 改进的粒子群算法

标准粒子群算法优化过程中, 每个种群粒子都会根据适应度在每一代进化中更新速度  $V$  和粒子当前值  $X$ , 为了避免优化过程早熟, 提前陷入局部搜索, 扩大算法的搜索空间, 平衡算法在全局和局部的搜索能力, 文献 (7) 提出迭代次数线性变化的惯性权重, 如式 (16) 所示。惯性权重保持微粒运行的惯性, 从而使算法具有所搜新区域的能力。

$$V(k+1) = \omega(k) \cdot V(k) + c_1 \cdot r_1 \cdot (Pbest - X(k)) + c_2 \cdot r_2 \cdot (Gbest - X(k)) \quad (15)$$

$$\omega(k) = (\omega_s - \omega_e) \cdot \frac{k_{\max} - k}{k_{\max}} + \omega_e \quad (16)$$

式中,  $c_1, c_2$  为粒子加速系数;  $r_1, r_2$  为范围内的随机数;  $pbest$  为单个粒子最好适应度时的值;  $Gbest$  为所有种群最好适应度的粒子值;  $k$  为当前种群迭代次数;  $\omega$  为惯性权重系数;  $\omega_s$  为惯性权重初始值;  $\omega_e$  为惯性权重结束值;  $k_{\max}$  为种群最大迭代次数。

线性递减的惯性权重结构简单, 易于计算, 但是实际的优化过程往往是高度复杂的, 随代数线性递减的策略不能正确反映实际问题。同时在一次迭代过程中, 所有种群粒子都是用同一惯性权重, 忽略了粒子差异性。基于这种不足, 本文提出一种新的粒子群改进算法。

定义  $D_k = |x_k - Gbest| \cdot |H(x_k) - H(Gbest)|$  为粒子与全

局最优的合成相对距离，由两部分组成， $|x_k - Gbest|$  为粒子位置与全局最优的相对距离， $|H(x_k) - H(Gbest)|$  为粒子适应度与全局最优适应度的相对值。

$$D_a = \frac{\sum_{k=1}^N D_k}{N}$$

定义为合成相对距离的平局值， $D_{max}$  为合成相对距离的最大值。

则  $w(k) = (w_s - w_e) \cdot \frac{D_k - D_a}{D_{max} - D_a} + w_e$  为新的惯性权重表达式，合成相对距离比较大的粒子，其惯性权重比较大，可以使其迅速的向全局最优优点靠近，合成相对距离比较小的粒子，其惯性权重就小，可以再精细的距离内向全局最优优点靠近。

3.2.2 改进的 PSO 算法用于优选 SVM 参数  
改进的 PSO 算法对核函数参数  $\sigma$  以及惩罚系数  $C$  进行参数优选<sup>[8]</sup>的具体步骤如下：

- (1) 导入学习样本数据，进行归一化处理；
- (2) 初始化设置，包括种群规模、迭代次数，随机给出粒子的初始位置  $z_i^0$  和初始速度  $v_i^0$ 。粒子个体对应  $\sigma$  和  $C$ ；
- (3) 用粒子对应的  $\sigma$  和  $C$ ，建立支持向量机的预测模型。构造适应度函数  $f(z_i)$ ：

$$f(z_i) = \min(\max(\{\frac{|x_k - x'_k|}{x'_k}\})) \quad (17)$$

式中， $x_k$  为  $z_k$  粒子对应的第  $k$  个测试样本的位移预测值； $x'_k$  为第  $k$  个样本的实测值。

(4) 根据改进的粒子群算法，若  $f(z_i) < f(pbest)$ ，则用新的适应度值代替粒子原有的适应度值，用新的粒子取代原来粒子。

(5) 将各个粒子的自身最好适应值  $f(pbest)$  与全局最优适应度值  $f(gbest)$  进行比较。如果  $f(pbest) < f(gbest)$ ，则用每个粒子的最好适应度值取代原有粒子的最好适应度值，同时保持粒子的当前状态。

(6) 判断适应度是否满足要求或者迭代次数已到，不满足则按式 (15)、(16) 进行移动，产生新的粒子，返回 (3)，直到运行结束，找出最优的  $\sigma$  和  $C$ 。

(7) 利用最优的  $\sigma$  和  $C$  建立支持向量机回归预测模型，进行时间序列的预测。

## 4 仿真实验

### 4.1 海杂波实测数据

本文的数据研究对象来源于加拿大 MCMaster 大学 tivec-systems 实验室公开的 IPIX (Intelligent Pixel-processing) 雷达数据，文件名为 19931107\_135603\_starea。该时间序列是复数形式，对实测数据进行实部与虚部平方和开根号 ( $\sqrt{Imag^2 + Real^2}$ ) 处理。相空间重构以及回归预测时都采用该文件名的数据。

### 4.2 仿真实验及结果分析

实验 1：直接采用 SVM 回归预测海杂波

支持向量回归 (SVR) 中选择出的核函数参数  $\sigma = 0.125$  以及惩罚系数  $C = 0.5$ ，均方差选择  $MSE = 0.075464$ 。

图 (2) 给出了原始数据与回归预测数据的对比效果，预测数据为训练样本数据以后的 100 个点。

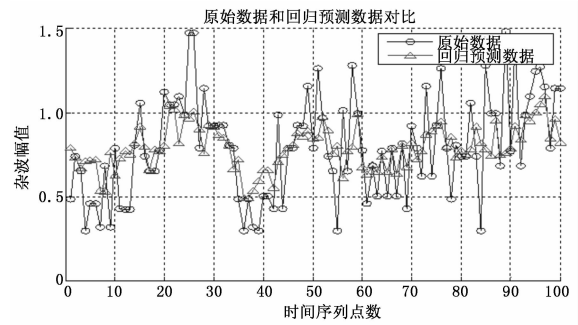


图 2 海杂波原始数据与回归预测数据对比 (SVM)

图 3 表示了预测数据相对于实际测量数据的相对误差。

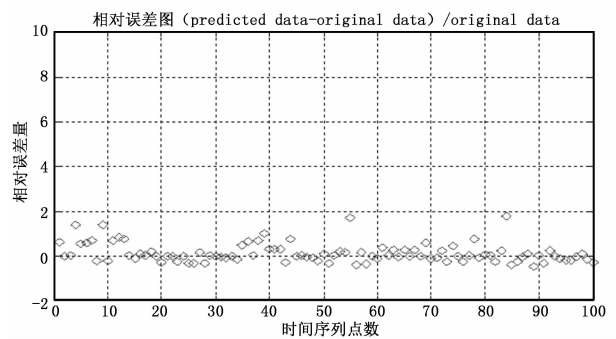


图 3 海杂波预测数据相对误差图 (SVM)

实验 2：采用 PSO-SVM 回归预测海杂波

利用粒子群优化的 SVM 计算出的核函数参数  $\sigma = 0.01$ ，惩罚系数  $C = 19.5$ ，均方差选择为  $MSE = 0.067541$ 。

图 (4) 是原始数据与 PSO-SVM 回归预测数据的对比效果，预测数据为训练样本数据以后的 100 个点。

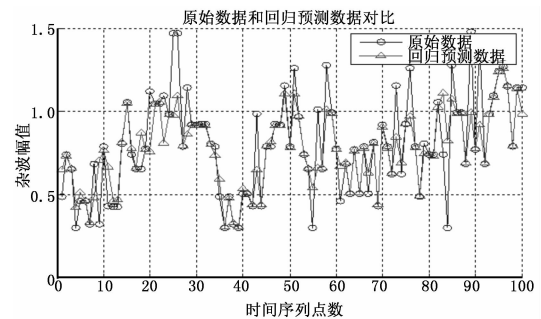


图 4 海杂波原始数据与回归预测数据对比 (PSO-SVM)

图 5 是预测数据相对于实际测量数据的相对误差。比较图 5 与图 3，采用 PSO-SVM 回归预测的相对误差的明显减小。计算 SVM 与 PSO-SVM 回归海杂波的均方差与相关系数如表 1 所示，由表 1 可看出：本文 PSO-SVM 回归预测海杂波性能较好。

表 1 SVM 与 PSO-SVM 方法预测效果

方法	SVM	PSO-SVM
均方差 MSE	0.900 91	0.709 52
相关系数 R(%)	58	67

对应的峰值信噪比、余量误差 2 范数和运行时间进行比较，结果如表 1 所示。

表 1 不同算法重构结果比较

算法名称	峰值信噪比(dB)	余量误差 2 范数	运行时间(s)
OMP	25.653	0.142 1	31.24
ROMP	26.784	0.094 3	13.535
文献[7]	27.674	0.063 4	10.564
文献[8]	29.454	0.056 3	24.344
文中方法	30.124	0.043 2	14.241

从表 1 中可以看出，文中方法的重构性能最好，这表现在文中方法具有最大的峰值信噪比和最小的余量误差 2 范数，由于文中方法求取了两次候选子集，因此执行时间略长于 ROMP 算法和文献 [7]，综合上述，文中算法在具有较少运行时间的同时能有最小的余量误差和最大的峰值信噪比，因此，文中方法更优。

## 6 结论

为了有效和快速地对规模巨大的数据量进行处理，需要在对信号进行压缩的基础上实现信号重构，为此，设计了一种基于压缩感知理论和改进的自适应正交匹配算法的信号重构方法。通过在信号重构阶段对候选原子集进行两次筛选，即选择具有最大能量的候选原子集和具有最大相关性的原子集，并取两者的交集作为最终的支配原子集，在此基础上通过重复迭代更新余量实现信号的精确重构，仿真实验证明了文中方法能实现信号的精确重构，具有较小的余量误差，较其他方法具有较大优越性。

### 参考文献：

[1] Baraniuk R G. Compressive sensing [J]. IEEE Signal Processing Magazine, 2007, 24 (4): 118 - 121.

~~~~~

(上接第 1542 页)

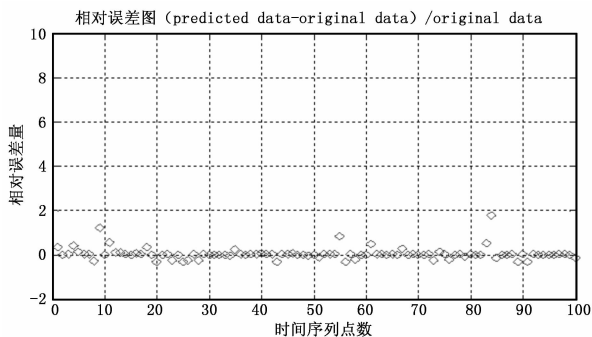


图 5 海杂波预测数据相对误差图 (PSO-SVM)

## 5 结束语

海杂波具有混沌特性，而支持向量机算法能够有效地对混沌序列进行回归预测。本文采用互信息法和改进的伪邻近点法从雷达实测数据中提取海杂波混沌特性的延迟时间和嵌入维数，再利用相空间重构得到 SVM 训练样本，应用改进的 PSO 算法优化 SVM 的参数，最终建立回归模型，并进行仿真实验。仿真实验结果表明：回归模型预测海杂波是可行的，而 PSO-SVM 法预测海杂波能进一步提高预测精度。

[2] Donoho D L, Tsaig Y, Drori I, etc. Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit [J]. IEEE Transactions on Information Theory, 2012, 58 (2): 1094 - 1121.

[3] 杨真真, 杨震, 孙林慧. 信号压缩重构的正交匹配追踪类算法综述 [J]. 2013, 29 (4): 486 - 489.

[4] Karahanoglu N B, Erdogan H. A orthogonal matching pursuit: Best-first search for compressed sensing signal recovery [J]. Digital signal processing: A review journal, 2012, 22 (4): 555 - 568.

[5] 路锦正, 张启衡, 徐智勇, 等. 超完备稀疏表示的图像超分辨率重构方法 [J]. 系统工程与电子技术, 2012, 34 (2): 403 - 408.

[6] 马春晖. 压缩感知重构算法研究 [D]. 杭州: 杭州电子科技大学, 2012.

[7] 陈艳, 王彪. 基于压缩感知的稀疏信道估计方法 [J]. 计算机工程与设计, 2013, 34 (7): 2338 - 2341.

[8] 白凌云, 梁志毅, 徐志军. 基于压缩感知信号重建的自适应正交多匹配追踪算法 [J]. 计算机应用研究, 2011, 28 (11): 4061 - 4063.

[9] 刘寅, 吴顺君, 张怀根, 等. 一种快速的基于压缩感知的多普勒高分辨方法 [J]. 西安电子科技大学学报 (自然科学版), 2011, 38 (2): 82 - 87.

[10] Mallat S, Zhang Z. Matching pursuits with time-frequency dictionaries [J]. IEEE Transactions on Signal processing, 1993, 41 (12): 3397 - 3415.

[11] Tropp J. Greed is good: algorithmic results for sparse approximation [J]. IEEE Trans. inf. Theory, 2004, 50 (10): 2231 - 2242.

[12] Donoho D L, Tsaig Y, Drori I, et al. Sparse solution of underdetermined linear equations by stage orthogonal matching pursuit [R]. Technical Report, 2006.

[13] 左来. 基于小波包分析和 LS-SVM 的柴油机故障诊断方法 [J]. 计算机测量与控制, 2009, 17 (11): 2150 - 2152.

### 参考文献：

[1] 张云展. 混沌学与海杂波 [J]. 火控雷达技术, 2008, 37 (4): 1 - 6.

[2] 张天文, 袁山增. 基于分形理论的目标航路预测 [J]. 指挥控制与仿真, 2008, 30 (2): 21 - 23.

[3] 马晓光, 胡非. 利用支撑向量机预报大气污染物浓度 [J]. 自然科学进展, 2004, (3): 349 - 353.

[4] 张小花, 陈玮. 基于相空间重构的制造系统混沌研究 [J]. 机械科学与技术, 2013, 32 (2): 309 - 312.

[5] 李中群, 裴承鸣. 一种新的多嵌入延迟相空间重构方法 [J]. 计算机工程与应用, 2009, 45 (33): 120 - 123.

[6] 混沌-支持向量机回归在流量预测中的应用研究 [J]. 计算机科学, 2009, 36 (7): 243 - 245.

[7] 刘楠, 黄金全. 应用改进粒子群算法的涡轮增压器性能寻优 [J]. 南京航空航天大学学报, 2013, 45 (3): 303 - 308.

[8] 邵信光, 杨慧中, 陈刚. 基于粒子群优化算法的支持向量参数选择及其应用 [J]. 控制理论与应用, 2006, 23 (5): 740 - 748.

[9] 尹波, 夏静波, 付凯, 等. 基于 IPSO 混沌支持向量机的网络流量预测研究 [J]. 计算机应用研究, 2012, 29 (11): 4293 - 4299.

[10] 田海累, 李洪儒, 许葆华. 基于改进人工鱼群算法的支持向量机预测 [J]. 计算机工程, 2013, 39 (4): 222 - 225.

[11] 姜斌, 王宏强, 黎湘. S 波段雷达实测数据海杂波混沌分型特性分析 [J]. 电子与信息学报, 2007, 29 (8): 1809 - 1812.

[12] 宋绍楼, 陈龙虎, 陈晓菊, 等. 基于粒子群多峰值 MPPT 算法的光伏系统研究 [J]. 计算机测量与控制, 2012, 20 (5): 1354 - 1356.