

## 基于鲁棒小波 $\nu$ -支持向量机的 产品销售预测模型

吴奇<sup>1,2</sup> 严洪森<sup>1</sup> 王斌<sup>1</sup>

**摘要** 针对产品销售时序具有正态高斯分布、幅值较大、奇异点等混合噪音,设计一种鲁棒损失函数,并采用小波核函数,由此得到一种新的小波  $\nu$ -支持向量机,即鲁棒小波  $\nu$ -支持向量机 (Robust wavelet  $\nu$ -support vector machine, RW $\nu$ -SVM). 它可以有效地压制销售时序的多种噪音和奇异点,具有很强的鲁棒性,而且它比标准小波  $\nu$ -支持向量机 (W $\nu$ -SVM) 具有更简洁的对偶优化问题. 最后进行了汽车销售预测的实例分析,结果表明基于 RW $\nu$ -SVM 的预测模型是有效可行的.

**关键词** 支持向量机, 小波核函数, 鲁棒损失函数, 预测  
**中图分类号** TP391

## Product Sales Forecasting Model Based on Robust Wavelet $\nu$ -Support Vector Machine

WU Qi<sup>1,2</sup> YAN Hong-Sen<sup>1</sup> WANG Bin<sup>1</sup>

**Abstract** Aiming at the normal Gaussian distributional noise, greater breadth noise and oddity point noise of product sales series and combing a designed robust loss function with wavelet kernel function, we propose a new wavelet  $\nu$ -support vector machine, named as robust wavelet  $\nu$ -support vector machine (RW $\nu$ -SVM). The RW $\nu$ -SVM, which has a stronger robustness and simpler dual optimization problem than standard wavelet-support vector machine (W $\nu$ -SVM), can inhibit some types of noise and disturbing oddity point noise of product sales series effectively. Finally, the RW $\nu$ -SVM is applied to the forecasts of car sales, and the results show that the forecasting model based on the proposed RW $\nu$ -SVM is effective and feasible.

**Key words** Support vector machine (SVM), wavelet kernel function, robust loss function, forecasting

产品销售预测过程是一个复杂的动态过程,其预测行为受很多因素的影响,这些因素大多具有随机性、非线性等特点,这些因素之间复杂的非线性关系,在低维空间很难用一种确定的方法来表达.而且实际的销售时序常受到一些随机误差的影响而退化,通常称这个退化为噪音 (Noise).在数据的收集,传输或处理过程中都可能出现各种噪音,噪音可能依赖于销售时序,也可能与其无关.理想的噪音,称为白噪音 (White noise),它是退化的最坏估计.而白噪音的一个特例是高斯噪音 (Gaussian noise).针对样本数据本身的特点,对于线性时间序列,主要的预测模型有自

收稿日期 2008-06-05 收修改稿日期 2008-11-14

Received June 5, 2008; in revised form November 14, 2008

国家高技术研究发展计划 (863 计划) (2007AA04Z112), 国家自然科学基金 (50875046, 60574062) 资助

Supported by National High Technology Research and Development Program of China (863 Program) (2007AA04Z112) and National Natural Science Foundation of China (50875046, 60574062)

1. 复杂工程系统测量与控制教育部重点实验室 (东南大学自动化学院) 南京 210096 2. 东南大学机械工程学院 南京 210096

1. Key Laboratory of Measurement and Control of Complex Systems of Engineering (School of Automation, Southeast University), Ministry of Education, Nanjing 210096 2. School of Mechanical Engineering, Southeast University, Nanjing 210096

DOI: 10.3724/SP.J.1004.2009.01027

回归求和滑动平均模型 (Autoregressive integrated moving average, ARIMA)、自回归条件异方差模型 (Autoregressive conditional heteroskedasticity model, ARCH) 等模型<sup>[1-2]</sup>. 对于非线性时序, 主要的预测模型有神经网络、模糊神经网络等模型<sup>[3]</sup>. 神经网络主要用于大量样本数据, 但其泛化能力较差, 涉及的输入变量较多而使得神经网络模型面临“维数灾难”问题; 对于受季节性因素影响的时序, 主要方法有 Holt-Winter<sup>[4]</sup> 模型. 这些模型各有自己的优缺点, 线性回归模型可以较好地处理线性问题, 但它需要大量的历史数据, 神经网络模型是基于大样本的模型, 其泛化能力较差, 模糊神经网络可以对一些模糊性的问题进行处理和预测, 但对小样本预测问题效果不明显. 理论上, 上述模型的建立都没有考虑对依赖于时序中的噪音如何进行惩罚 (降噪). 对较少的样本数据, 样本数据间的不确定性关系, 无法在样本数据维空间用一种确定的方式来表达这个不确定关系. 但采用结构风险最小原则的支持向量机, 将样本空间的数据关系映射到一个高维空间, 在高维空间里, 数据间的关系可以用一种确定的表达方式来表示. 支持向量机<sup>[5]</sup> 是由 Vapnik 等人提出的一种全新的机器学习方法, 由于其具有很好的泛化能力, 得到的解是全局最优解, 而且能较好地解决小样本、非线性、高维数等实际问题, 在国内外学术界已受到广泛重视, 并已在模式识别、函数估计、产品开发时间估计、预测建模等方面取得越来越多的应用<sup>[6-11]</sup>. 值得一提的是 Yan 和 Xu<sup>[11]</sup> 建立了一种新的模糊支持向量机, 并运用于产品开发时间估计方面研究. 由于标准支持向量机的不敏感损失函数的缺陷性, 不能够有效处理一些正态分布, 幅值较大和奇异点的噪音, 本文设计了一种新的损失函数, 将松弛变量的作用区间分为三个, 靠近最优分界的区间为管道区, 中间区间为高斯区, 最外层的区间为线性 and 奇异点区. 这样落在各个区间内的噪音会被各自区间的损失函数有效压制. 基于这个思想启发, 本文设计一种鲁棒性损失函数, 采用小波核函数, 提出一种新的小波支持向量机, 即鲁棒小波  $\nu$ -支持向量机 (Robust wavelet  $\nu$ -support vector machine, RW $\nu$ -SVM). 并将这种模型与产品销售预测的特点相结合, 从而得到一种产品销售智能短期预测方法.

## 1 鲁棒小波 $\nu$ -支持向量机

目前大多学者仍然使用不敏感损失函数来处理时间序列, 使用方法不具有针对性. Rossi 和 Villab 等<sup>[8]</sup> 用支持向量机对数据进行分类, 使用的是不敏感损失函数, 没有针对数据特征设计对应的损失函数. Shim 和 Hwang 等<sup>[12]</sup> 使用一个平方损失函数的支持向量机作为鲁棒支持向量机, 这种损失函数没有从理论上说明其优点, 而且无法处理幅度较大和奇异点的噪音.

一些学者对标准向量机的提法做了改进<sup>[13]</sup>, 以便支持向量机所形成的对偶问题更容易求解. 这种研究是值得提倡的. 本文从结构风险最小化原则出发, 把小波支持向量机最终决策回归函数的偏置量参数的影响直接纳入最优化问题的原问题, 建立起参数与其对偶问题最优解之间的关系, 获取一种新的小波支持向量机. 它减少了对偶问题的约束条件, 大大简化了最优化问题, 提高了计算效率.

根据结构风险最小化原则, 由于采用不同的损失函数, 支持向量机的表现形式具有多样性, 而标准支持向量机只是其中的一种形式<sup>[5]</sup>. 在回归估计中, 支持向量机的最优损失函数仅和样本数据中的噪音相关<sup>[14]</sup>, 由此可见, 松弛变量的表达式是由损失函数来决定. 目前常见的损失函数有  $\varepsilon$ -不敏感

损失函数, 高斯损失函数, Laplace 损失函数, 多项式损失函数等. 对于每种损失函数, 都有不同的物理意义. 例如, 高斯损失函数和 Laplace 损失函数侧重于对噪音进行降噪, 而多项式损失函数则侧重于松弛变量的数学表达式.

由此, 本文考虑一种由  $\varepsilon$ -不敏感损失函数, 高斯损失函数, Laplace 损失函数组成的混合损失函数, 兼顾数据噪音分布特征, 得到一种分段式损失函数, 即鲁棒损失函数, 作为小波支持向量机的损失函数. 因此, 对于分段式损失函数 (鲁棒损失函数), 必然对应于一种分段式支持向量机, 即在不同的松弛变量区域, 有不同的最优化问题.

基于以上考虑, 本文建立一种新的最优化问题, 设计一种鲁棒损失函数, 采用小波核函数, 提出一种新的小波支持向量机, 即鲁棒小波  $\nu$ -支持向量机 (RW $\nu$ -SVM).

### 1.1 鲁棒损失函数

样本集的最优损失函数与样本集数据内在特性之间存在某种关系, 对于正态分布噪音, 选择高斯函数作为损失函数, 可以得到最好的降噪效果. 对于幅值较大的噪音和奇异点, Laplace 损失函数 (线性损失函数) 表现得最好<sup>[14]</sup>. 本文综合各种损失函数的特点, 考虑到样本数据的未知性、随机性、不平稳性等特征, 有必要在设计损失函数的时把体现上述特征的函数纳入新的损失函数内, 即鲁棒损失函数, 其数学表达式为

$$L(\xi) = \begin{cases} 0, & |\xi| \leq \varepsilon \\ \frac{1}{2}(|\xi| - \varepsilon)^2, & \varepsilon < |\xi| \leq \varepsilon_\mu \\ \mu(|\xi| - \varepsilon) - \frac{1}{2}\mu^2, & |\xi| > \varepsilon_\mu \end{cases} \quad (1)$$

其中,  $\varepsilon + \mu = \varepsilon_\mu$ ,  $\varepsilon \geq 0$ ,  $\mu \geq 0$ . 鲁棒损失函数分为三部分:

- 1)  $|\xi| \leq \varepsilon$  部分, 即  $\varepsilon$  不灵敏区, 不惩罚小于  $\varepsilon$  的偏差, 使学习机的解具有稀疏性;
- 2)  $\varepsilon < |\xi| \leq \varepsilon_\mu$  部分, 选用高斯函数损失函数, 可以抑制符合高斯特征的噪音;
- 3)  $|\xi| > \varepsilon_\mu$  部分, 选用 Laplace 损失函数, 可以较好地抑制幅值较大的噪音和一些奇异点.

### 1.2 小波核函数

根据广义线性判别函数的思路, 要解决一个非线性问题, 可通过非线性变换将输入空间变换到一个高维特征空间, 在此特征空间中寻求最优或广义最优分类面, 且只需进行内积运算, 而没有必要知道采用的是非线性变换的形式. 只要在变换后的特征空间的内积可以用原空间中的变量直接计算得到, 则即使特征空间中的维数增加很多, 在其中求解最优分类面的问题并没有增加多少计算复杂度.

统计学习理论指出, 根据 Hilbert-Schmidt 原理, 只要一种运算满足 Mercer 条件, 它就即可作内积使用.

**引理 1 (Mercer 条件)**<sup>[15-17]</sup>.

$L^2(\mathbf{R}^n)$  下的对称函数  $\mathbf{K}(\mathbf{x}, \mathbf{x}')$  为特征空间中的内积的充分必要条件是, 对于任意的  $\varphi(\mathbf{x}) \neq 0$  且  $\int \varphi^2(\mathbf{x})d\mathbf{x} < \infty$ , 满足下列条件

$$\iint \mathbf{K}(\mathbf{x}, \mathbf{x}')\varphi(\mathbf{x})\varphi(\mathbf{x}')d\mathbf{x}d\mathbf{x}' \geq 0, \quad \mathbf{x}, \mathbf{x}' \in \mathbf{R}^n \quad (2)$$

则满足式 (1) 的点积核  $\mathbf{K}(\mathbf{x}, \mathbf{x}')$  是一个允许支持向量核.

**引理 2 (平移不变核)**<sup>[15-17]</sup>.

$L^2(\mathbf{R}^n)$  下的对称函数  $\mathbf{K}(\mathbf{x}, \mathbf{x}')$ ,  $\mathbf{x}, \mathbf{x}' \in \mathbf{R}^n$ , 平移函数满足  $\mathbf{K}(\mathbf{x}, \mathbf{x}') = \mathbf{K}(\mathbf{x} - \mathbf{x}')$ , 若满足 Mercer 条件, 那么它就

是一个允许支持向量核。

引理 1 和引理 2 提供了判断和构建核函数的方法。但是实际上很难将平移不变核分解成两个函数的内积, 引理 3 给出了平移不变核函数是一个支持向量核的充分必要条件。

引理 3<sup>[15-17]</sup>, 平移不变核函数  $K(\mathbf{x}, \mathbf{x}') = K(\mathbf{x} - \mathbf{x}')$  是一个允许支持向量核, 当且仅当  $K(\mathbf{x})$  的傅里叶变换成立。

$$F[\mathbf{x}](\boldsymbol{\omega}) = (2\pi)^{-\frac{n}{2}} \int_{\mathbf{R}^n} \exp(-j(\boldsymbol{\omega} \cdot \mathbf{x})) K(\mathbf{x}) d\mathbf{x} \geq 0, \quad \boldsymbol{\omega} \in \mathbf{R}^n, \mathbf{x} \in \mathbf{R}^n \quad (3)$$

将母小波函数替换成某些具体形式, 就可以得到相应的小波核函数, 下面是构造小波核函数的定理。

引理 4<sup>[15-17]</sup>, 设母小波函数为  $\psi(\mathbf{x}_i)$ ,  $\mathbf{x}_i, \mathbf{x}'_i \in \mathbf{R}^d$ ,  $\mathbf{x}'$  和  $\mathbf{x}'_i$  都是  $d$  维列向量,  $\mathbf{a}$  为  $l$  维尺度因子列向量,  $\mathbf{m}$  和  $\mathbf{m}'$  是  $d$  维位移因子列向量, 其中  $\mathbf{a} \in \mathbf{R}^l$ ,  $\mathbf{m}, \mathbf{m}' \in \mathbf{R}^d$ , 则满足 Mercer 条件的点积形式的小波核函数

$$K(\mathbf{x}, \mathbf{x}') = \prod_{i=1}^l \psi\left(\frac{\mathbf{x}_i - \mathbf{m}}{a_i}\right) \psi\left(\frac{\mathbf{x}'_i - \mathbf{m}'}{a_i}\right) \quad (4)$$

和满足平移不变核定理的小波核函数

$$K(\mathbf{x}, \mathbf{x}') = \prod_{i=1}^l \psi\left(\frac{\mathbf{x}_i - \mathbf{x}'_i}{a_i}\right) \quad (5)$$

均是可容许的支持向量核。

式 (5) 中没有出现位移因子  $\mathbf{m}$ , 但有一个变量  $\mathbf{x}'_i$ , 该变量的作用等价于位移因子。因此, 可以用小波函数来构造具有平移不变性的小波核函数。

定理 1<sup>[17]</sup>, Morlet 母小波函数为

$$\psi(\mathbf{x}) = \cos(\omega_0 \mathbf{x}_i) \exp(-\|\mathbf{x}_i\|^2/2), \quad \omega_0 \in \mathbf{R}$$

则小波核函数

$$K(\mathbf{x}, \mathbf{x}') = \prod_{i=1}^l \cos\left(\omega_0 \times \frac{\mathbf{x}_i - \mathbf{x}'_i}{a_i}\right) \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}'_i\|^2}{2a_i^2}\right), \quad \mathbf{x}' \text{ 和 } \mathbf{x}'_i \text{ 都是 } d \text{ 维列向量} \quad (6)$$

是一种可容许的支持向量核。

那么, 对于 Mexican 小波

$$\psi(\mathbf{x}) = (1 - \|\mathbf{x}_i\|^2) \exp\left(-\frac{\|\mathbf{x}_i\|^2}{2}\right) \quad (7)$$

根据引理 4 和定理 1, Mexican 小波核函数能够表示为

$$K(\mathbf{x}, \mathbf{x}') = \prod_{i=1}^l \left(1 - \frac{\|\mathbf{x}_i - \mathbf{x}'_i\|^2}{a_i^2}\right) \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}'_i\|^2}{2a_i^2}\right), \quad \mathbf{x}' \text{ 和 } \mathbf{x}'_i \text{ 都是 } d \text{ 维列向量} \quad (8)$$

### 1.3 鲁棒小波 $\nu$ -支持向量机

在对标准小波支持向量机的求解过程中, 参数  $b$  的精确确定是困难的, 导致在模型参数寻优过程中消耗迭代时间。为此本文从结构风险最小化原则出发, 在建立模型最初时, 参数  $b$  被考虑进最优化问题的原问题中, 获取一种新的小波

支持向量机原问题。它减少了一个对偶问题的约束条件, 参数  $b$  将不出现在最终决策函数中, 整个求解过程无需对参数  $b$  进行辨识, 从而提高了计算效率。而参数  $b$  放入最优化问题原问题中的研究, 作者在这方面已经做了前期工作<sup>[18]</sup>, 具体推导过程如下:

设鲁棒小波  $\nu$ -支持向量机的回归函数具有  $\bar{f}(\mathbf{x}) = \bar{\mathbf{w}} \cdot \bar{\mathbf{x}}$  形式, 标准小波  $\nu$ -支持向量机的回归函数有标准形式  $f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b$ 。对于样本集  $T = \{\mathbf{x}_i, y_i\}_{i=1}^l$ , 其中  $\mathbf{x}_i \in \mathbf{R}^d$ ,  $y_i \in \mathbf{R}$ ,  $\mathbf{x}_i$  为  $d$  维列向量。考虑由  $\bar{\mathbf{x}} = (\mathbf{x}_i^T, \eta)^T$  构成的 Hilbert 空间, 并定义  $\bar{\mathbf{x}}_1 = (\mathbf{x}_1^T, \eta)^T$  和  $\bar{\mathbf{x}}_2 = (\mathbf{x}_2^T, \eta)^T$  的内积  $\bar{\mathbf{x}}_1 \cdot \bar{\mathbf{x}}_2 = \mathbf{x}_1 \cdot \mathbf{x}_2 + \eta^2$ , 令  $\bar{\mathbf{w}} = (\mathbf{w}^T, b/\eta)^T$ , 其中  $\mathbf{w}$  为  $d$  维列向量,  $\bar{\mathbf{w}}$  为  $d+1$  维列向量,  $\eta$  为不等于零的实数, 于是有  $\bar{f}(\mathbf{x}) = \bar{\mathbf{w}} \cdot \bar{\mathbf{x}} = \mathbf{w} \cdot \mathbf{x} + (b/\eta)\eta = \mathbf{w} \cdot \mathbf{x} + b = f(\mathbf{x})$  成立。由此可见, 本文所建立的最终决策回归函数中不出现参数  $b$  的鲁棒小波  $\nu$ -支持向量机的最优化问题  $\bar{f}(\mathbf{x})$ , 等价于最终决策回归函数中出现参数的标准小波  $\nu$ -支持向量机的最优化问题  $f(\mathbf{x})$ 。

于是, 对于样本集  $T = \{\mathbf{x}_i, y_i\}_{i=1}^l$ , 其中  $\mathbf{x}_i \in \mathbf{R}^d$ ,  $y_i \in \mathbf{R}$ ,  $\mathbf{x}_i$  为  $d$  维列向量, 将参数  $b$  的影响直接纳入最优化问题的原问题, 由此产生一种新的小波  $\nu$ -支持向量机, 并采用鲁棒损失函数 (1), 得到鲁棒小波  $\nu$ -支持向量机 (RW $\nu$ -SVM), 其数学表达式为

$$\begin{aligned} \min_{\mathbf{w}, \boldsymbol{\xi}^{(*)}, b, \varepsilon} & \frac{1}{2} (\|\mathbf{w}\|^2 + b^2) + \\ & C \cdot \left( \nu \cdot \varepsilon + \frac{1}{l} \sum_{i \in I_1} \frac{1}{2} (\xi_i^2 + \xi_i^{*2}) + \frac{1}{l} \sum_{i \in I_2} \mu(\xi_i + \xi_i^*) \right) \\ \text{s.t.} & \begin{cases} y_i - \mathbf{w} \cdot \mathbf{x}_i - b \leq \varepsilon + \xi_i \\ \mathbf{w} \cdot \mathbf{x}_i + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, \nu \in (0, 1], \varepsilon \geq 0, i = 1, \dots, l \end{cases} \quad (9) \end{aligned}$$

其中,  $\mathbf{w}$  为  $d$  维列向量,  $C > 0$  为惩罚系数,  $\nu \in (0, 1]$  用来控制模型复杂性与训练误差之间的平衡; 为控制支持向量的参数,  $\varepsilon$  为控制管道大小的参数;  $\xi_i, \xi_i^* (i = 1, \dots, l)$  为松弛变量, 保证了约束的满足;  $I_1$  是表示松弛变量落在  $0 < |\xi_i|, |\xi_i^*| \leq \varepsilon_\mu$  区间内的样本集,  $I_2$  是表示松弛变量落在  $\varepsilon_\mu < |\xi_i|, |\xi_i^*|$  区间内的样本集。

为求上述 RW $\nu$ -SVM 的对偶问题, 定义 Lagrange 函数如下

$$\begin{aligned} L(\mathbf{w}, b, \boldsymbol{\alpha}^{(*)}, \beta, \boldsymbol{\xi}_*, \varepsilon, \boldsymbol{\eta}^*) &= \frac{1}{2} (\|\mathbf{w}\|^2 + b^2) + \\ & C \cdot \left( \nu \cdot \varepsilon + \frac{1}{l} \sum_{i \in I_1} \frac{1}{2} (\xi_i^2 + \xi_i^{*2}) + \frac{1}{l} \sum_{i \in I_2} \mu(\xi_i + \xi_i^*) \right) - \\ & \beta \cdot \varepsilon - \sum_{i=1}^l (\eta_i \xi_i + \eta_i^* \xi_i^*) - \\ & \sum_{i=1}^l \alpha_i (\varepsilon + \xi_i + y_i - \mathbf{w} \cdot \mathbf{x}_i - b) - \\ & \sum_{i=1}^l \alpha_i^* (\varepsilon + \xi_i^* - y_i + \mathbf{w} \cdot \mathbf{x}_i + b) \quad (10) \end{aligned}$$

其中, 乘子  $\alpha_i^*, \eta_i^*, \beta \geq 0$ 。分别对  $\mathbf{w}, b, \varepsilon, \boldsymbol{\xi}_*$  计算偏导, 即  $\nabla_{\mathbf{w}} L = 0, \nabla_b L = 0, \nabla_{\varepsilon} L = 0, \nabla_{\boldsymbol{\xi}_*} L = 0$ 。其中  $\nabla_{\mathbf{w}} L = 0$  和  $\nabla_b L = 0$  时能够得到

$$\begin{cases} \mathbf{w} = \sum_{i=1}^l (\alpha_i^* - \alpha_i) \mathbf{x}_i \\ b = \sum_{i=1}^l (\alpha_i^* - \alpha_i) \end{cases} \quad (11)$$

然后利用对偶原理、Karush-Kuhn-Tucker (KKT) 条件和小波核函数技巧,可以得到优化问题 (9) 的对偶问题如下

$$\begin{aligned} \min_{\alpha, \alpha^*} & \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) (\mathbf{K}(\mathbf{x}, \mathbf{x}') + 1) - \\ & \sum_{i=1}^l y_i (\alpha_i^* - \alpha_i) + \frac{l}{2C} \sum_{i=1}^l (\alpha_i^2 - \alpha_i^{*2}) \\ \text{s.t.} & \begin{cases} \mathbf{e}^T (\boldsymbol{\alpha} + \boldsymbol{\alpha}^*) \leq C \cdot \nu, \\ 0 \leq \alpha_i, \alpha_i^* \leq \min \left( \frac{C}{l}, \frac{C \cdot \mu}{l} \right) \end{cases} \end{aligned} \quad (12)$$

其中,对于二次损失函数部分的约束条件为:  $0 \leq \alpha_i, \alpha_i^* \leq C/l$ ; 对于一次损失函数部分的约束条件为:  $0 \leq \alpha_i, \alpha_i^* \leq C\mu/l$ . 为同时满足上面二个条件,取  $\min \left( \frac{C}{l}, \frac{C \cdot \mu}{l} \right)$  作为  $\alpha_i, \alpha_i^*$  的最大值.

将上式改写成矩阵形式,可得

$$\begin{aligned} \min_{\alpha, \alpha^*} & \frac{1}{2} [(\boldsymbol{\alpha}^*)^T, \boldsymbol{\alpha}^T] \begin{bmatrix} Q + \frac{El}{C} & -Q \\ -Q & Q + \frac{El}{C} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\alpha}^* \end{bmatrix} + \\ & [-\mathbf{y}^T, \mathbf{y}^T] \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\alpha}^* \end{bmatrix} \\ \text{s.t.} & \begin{cases} \mathbf{e}^T (\boldsymbol{\alpha} + \boldsymbol{\alpha}^*) \leq C \cdot \nu, \\ 0 \leq \alpha_i, \alpha_i^* \leq \min \left( \frac{C}{l}, \frac{C \cdot \mu}{l} \right) \end{cases} \end{aligned} \quad (13)$$

其中,  $Q_{ij} = K(\mathbf{x}_i, \mathbf{x}_j) + 1, i = 1, \dots, l, \mathbf{e} = [1, \dots, 1]^T, \mathbf{e}$  为  $l$  维列向量;  $E$  为  $l$  阶单位矩阵;  $\boldsymbol{\alpha}, \boldsymbol{\alpha}^*$  为 Lagrange 乘子组成的向量,它们都是  $l$  维一个非负列向量.

进一步将式 (13) 改写成简洁形式

$$\begin{aligned} \min & \frac{1}{2} \bar{\boldsymbol{\alpha}}^T H \bar{\boldsymbol{\alpha}} + \bar{\mathbf{y}}^T \bar{\boldsymbol{\alpha}} \\ \text{s.t.} & \begin{cases} \mathbf{e}^T (\boldsymbol{\alpha} + \boldsymbol{\alpha}^*) \leq C \cdot \nu, \\ 0 \leq \alpha_i, \alpha_i^* \leq \min \left( \frac{C}{l}, \frac{C \cdot \mu}{l} \right) \end{cases} \end{aligned} \quad (14)$$

其中,  $\bar{\boldsymbol{\alpha}} = \begin{bmatrix} \boldsymbol{\alpha}^* \\ \boldsymbol{\alpha} \end{bmatrix}, H = \begin{bmatrix} Q + \frac{El}{C} & -Q \\ -Q & Q + \frac{El}{C} \end{bmatrix}, \bar{\mathbf{y}} = \begin{bmatrix} -\mathbf{y} \\ \mathbf{y} \end{bmatrix}.$

将式 (5) 和 (11) 代入 RW $\nu$ -SVM 的回归估计函数  $f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b$ , 得到 RW $\nu$ -SVM 的输出

$$f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b = \sum_{i=1}^l (\alpha_i^* - \alpha_i) \prod_{j=1}^l \psi \left( \frac{x_j - x_{ij}}{a_i} + 1 \right) \quad (15)$$

其中,  $x_j$  表示输入向量  $\mathbf{x}$  的第  $j$  个分量,  $x_{ij}$  表示第  $i$  个训练样本的第  $j$  个分量.

与标准小波  $\nu$ -支持向量机 (W $\nu$ -SVM) 相比较, RW $\nu$ -SVM 的对偶问题减少了一个约束条件,简化了计算. 而且

RW $\nu$ -SVM 的输出表达式 (15) 不含有参数  $b$ , 因此避免了参数  $b$  的辩识.

用获得的最佳回归模型,就可以对样本进行回归分析,对样本的变化趋势进行预测.

**定理 2.** 设  $\boldsymbol{\alpha}^{(*)} = (\alpha_1, \dots, \alpha_l, \alpha_1^*, \dots, \alpha_l^*)^T$  是最优化问题 (10) 的解, 则

1) 若  $\alpha_i = \alpha_i^*$ , 则相应的样本点  $(x_{ij}, y_i)$  一定在  $\varepsilon$ -带的内部或边界上;

2) 若  $\alpha_i \in \left( 0, \min \left( \frac{C}{l}, \frac{C}{l} \cdot \mu \right) \right), \alpha_i^* = 0$  或  $\alpha_i = 0, \alpha_i^* \in \left( 0, \min \left( \frac{C}{l}, \frac{C}{l} \cdot \mu \right) \right)$ , 则相应的样本点  $(x_{ij}, y_i)$  一定在  $\varepsilon$ -带的边界上;

3) 若  $\alpha_i = \min \left( \frac{C}{l}, \frac{C}{l} \cdot \mu \right), \alpha_i^* = 0$  或  $\alpha_i = 0, \alpha_i^* = \min \left( \frac{C}{l}, \frac{C}{l} \cdot \mu \right)$ , 则相应的样本点  $(x_{ij}, y_i)$  一定在  $\varepsilon$ -带的边界上.

**证明.** 若  $\boldsymbol{\alpha}^{(*)} = (\alpha_1, \dots, \alpha_l, \alpha_1^*, \dots, \alpha_l^*)^T$  是最优化问题 (10) 的解, 根据 KKT 条件, 在最优解处, 有

$$\alpha_i (\varepsilon + \xi_i - y_i + \mathbf{w} \cdot \mathbf{x}_i + b) = 0 \quad (16)$$

$$\alpha_i^* (\varepsilon + \xi_i^* + y_i - \mathbf{w} \cdot \mathbf{x}_i - b) = 0 \quad (17)$$

$$\left( \min \left( \frac{C}{l}, \frac{C}{l} \cdot \mu \right) - \alpha_i \right) \xi_i = 0 \quad (18)$$

$$\left( \min \left( \frac{C}{l}, \frac{C}{l} \cdot \mu \right) - \alpha_i^* \right) \xi_i^* = 0 \quad (19)$$

以下分别证明 1)~3).

1) 若  $\alpha_i = \alpha_i^* = 0$ , 由式 (18) 和 (19) 可以看出,  $\xi_i = \xi_i^* = 0$ , 那么根据鲁棒损失函数式 (1) 可知, 相应的样本点  $(x_{ij}, y_i)$  一定在  $\varepsilon$ -带的内部或边界上.

2) 若  $\alpha_i \in \left( 0, \min \left( \frac{C}{l}, \frac{C}{l} \cdot \mu \right) \right), \alpha_i^* = 0$  或  $\alpha_i = 0, \alpha_i^* \in \left( 0, \min \left( \frac{C}{l}, \frac{C}{l} \cdot \mu \right) \right)$ , 由式 (18) 和 (19) 可以看出,  $\xi_i = \xi_i^* = 0$ , 结合式 (16) 和 (17), 可知  $y_i = \mathbf{w} \cdot \mathbf{x}_i + b + \varepsilon$  一定成立; 因而相应的样本点  $(x_{ij}, y_i)$  一定在  $\varepsilon$ -带的边界上.

3) 若  $\alpha_i = \min \left( \frac{C}{l}, \frac{C}{l} \cdot \mu \right), \alpha_i^* = 0$ , 那么由式 (18) 可以看出

$$\xi_i \geq 0, \quad \xi_i^* = 0 \quad (20)$$

再由式 (16) 得到

$$y_i = \varepsilon + \xi_i + \mathbf{w} \cdot \mathbf{x}_i + b \geq \varepsilon + \mathbf{w} \cdot \mathbf{x}_i + b \quad (21)$$

若  $\alpha_i = 0, \alpha_i^* = \min \left( \frac{C}{l}, \frac{C}{l} \cdot \mu \right)$ , 那么由式 (19) 可以看出

$$\xi_i^* \geq 0, \quad \xi_i = 0 \quad (22)$$

再由式 (17) 可以得到

$$y_i = \mathbf{w} \cdot \mathbf{x}_i + b - \varepsilon - \xi_i^* \leq \mathbf{w} \cdot \mathbf{x}_i + b - \varepsilon \quad (23)$$

通过分析式 (20)~(23), 可以得到相应的样本点  $(x_{ij}, y_i)$  一定在  $\varepsilon$ -带的边界上.  $\square$

### 1.4 基于 RW $\nu$ -SVM 的回归估计

RW $\nu$ -SVM 的训练涉及到  $C$ ,  $\nu$ ,  $\mu$  和核函数常量系数  $\alpha$  这些参数, 本文中只考虑  $\mu = 1$  的情况. 利用粒子群算法对 RW $\nu$ -SVM 模型参数进行辨识, 获取最优参数组合  $(C, \nu, a)$ . 这样, 基于 RW $\nu$ -SVM 模型的回归估计步骤如下:

步骤 1. 对多维样本数据进行归一化处理, 形成样本集;

步骤 2. 对样本时序  $y_i$  ( $i = 1, 2, \dots, l$ ) 进行多尺度小波变换, 选择与原始时序相吻合的小波函数, 粗估尺度变化范围;

步骤 3. 按照式 (5) 构造小波函数  $K(\mathbf{x}, \mathbf{x}')$ , 建立形式如式 (9) 的最优化目标函数;

步骤 4. 利用粒子群算法对参数  $(C, \nu, a)$  进行辨识, 利用凸二次规划求解优化目标函数问题并判断 KKT 条件, 得到参数  $\alpha^{(*)}$ ;

步骤 5. 对新的输入  $\mathbf{x}$ , 根据式 (15) 计算预测值.

## 2 应用实例

通过向若干个相关汽车企业的销售工程师、部门负责人、项目管理和资料管理人员作问卷调查, 综合分析后获得主要的汽车销售因素. 以汽车销售量的预测为例, 考虑三个产品内在特征因素和三个外在经济形势因素: 汽车价格波动因素 PW、汽车性能参数 PF、汽车油价影响数 OF、季节性因素 SC、销售工程师经验数 SE 和居民可支付购买力 PP, 其中 PW, PF, OF, SC, SE 和 PP 为六个语言型数据. 利用模糊综合评价法获取语言型因素的影响权重, 作为影响因素输入集. 具体如下: 建立因素集  $U$ , 评语集  $V$ , 根据因素集中六个评价因素对汽车销售影响程度的不同, 分别赋予不同的权重, 产生权重集  $A$ , 权重集是  $U$  上的一个模糊子集, 各因素权重的分配采用层次分析法、调查分析法确定. 这样获取的权重集  $A$  中某因素对汽车销售影响程度的数据就是数值型了, 该数据可以直接作为支持向量机训练集  $T$  中的训练数据. 选取 60 个汽车销售数据加上由模糊综合评价法获取的  $60 \times 6$  个影响因素集的数值型数据组成小波支持向量机的样本集进行研究.

利用 Matlab 7.1 编制相应的模型训练和预测程序, 运行环境为 Core (TM) 2 CPU 1.80 GHz、内存 1.00 G 的微机. 利用小波理论获取在不同尺度范围内汽车销售时序的小波变换. 本文选择的小波函数有 Morlet, Mexh 和 Gaussian. 粗估计在尺度  $a \in (0, 2]$  范围内, Mexican 母小波要优于 Morlet 母小波, 更能与原始销售时序相吻合. 由此, RW $\nu$ -SVM 的核函数选用 Mexican 小波核, 三个参数的变化范围设置如下:  $C \in [0.0001, 10000]$ ,  $\nu \in [0.001, 1]$ ,  $a \in (0, 2]$ . 粒子群参数设定为:  $w = 0.9$ ,  $c_1 = c_2 = 2$ . 利用粒子群算法获取最优参数组合为:  $C = 10000$ ,  $\nu = 0.92$  和  $a = 1$ . RW $\nu$ -SVM 给出最近 12 个月的归一化样本预测均方差为  $3.6064 \times 10^{-4}$ , 而用标准 W $\nu$ -SVM 获取的均方差为  $6.7968 \times 10^{-4}$ , 可见 RW $\nu$ -SVM 预测能力要好于标准 W $\nu$ -SVM.

从图 1 (见下页) 可以发现, 靠近样本均值的时序, RW $\nu$ -SVM 给出的预测精度高, 而远离均值的时序, RW $\nu$ -SVM 给出的预测精度较低. 最近 12 个月的预测结果如表 1 (见下页) 所示. 下文中第 1 组对应于表 1 中的 No.1, 第 2 组对应于表 1 中的 No.2, 以此类推, 第 12 组对应于表 1 中的 No.12. 从自回归滑动平均模型 (Autoregressive moving average, ARMA) 预测的结果来看, 可以发现, 其最大预测误差为 54.07% (第 1 组), 最小预测误差为 12.7% (第 2 组), 这说明 ARMA 对于具有明显非线性、随机扰动等特性的样本的预测能力差, 而对线性、平稳样本的预测能力好. 而用小波支持向量机模型的预测结果明显好于 ARMA 的预测结果, 其中 W $\nu$ -SVM 模型给出的最大预测误差为 4.59% (第 6 组), 最小预测误差为 0.27% (第 2 组), RW $\nu$ -SVM 模型给出的最大预测误差为 3.95% (第 6 组), 最小预测误差为 0 (第 2 组和第 11 组). 求解 RW $\nu$ -SVM 的过程, 实际上和 W $\nu$ -SVM 一样都是求解一个凸二次规划问题, 模型求解的复杂度是一样的, 运行效率也是一样的. 考虑到企业销售产品的现实状况, 数据的收集和估计不可避免地存在着偏差, 因此预测所得结果还是非常满意的. 由此可见, 小波支持向量机对于具有小样本、非线性、随机扰动、不平稳等多维特征的时序具有强大的高维映射能力. 从表 1 也能看出距离样本均值较近的样本点, RW $\nu$ -SVM 比标准 W $\nu$ -SVM 提供的预测精度更高, 而距离样本均值较远的样本点, RW $\nu$ -SVM 提供的预测精度也比标准 W $\nu$ -SVM 高. 从预测结果来看, RW $\nu$ -SVM 将落在  $[-\varepsilon_\mu, -\varepsilon] \cup (\varepsilon, \varepsilon_\mu]$  范围内的噪音, 采用高斯损失函数进行惩罚, 可以有效压制正态分布噪音; 将落在  $(-\infty, -\varepsilon_\mu) \cup (\varepsilon_\mu, +\infty)$  范围内的噪音, 采用 Laplace 损失函数进行惩罚, 可以有效压制幅值较大的噪音和一些奇异点. 因此, 相比于  $\varepsilon$ -不敏感损失函数的标准 W $\nu$ -SVM, 具有鲁棒损失函数的 RW $\nu$ -SVM 有较好的泛化能力. 为了便于对三种模型的预测值进行分析, 表 2 给出了基于平均绝对误差 (Mean absolute error, MAE), 平均绝对百分误差 (Mean absolute percentage error, MAPE) 和均方误差 (Mean square error, MSE) 的误差统计分析. 从表 2 可以明显看出, 对于本算例的销售时序, ARMA 给出的误差精度远远低于小波支持向量机给出的误差精度. 而且, 本文给出的具有鲁棒性的 RW $\nu$ -SVM 的误差指标值均好于标准 W $\nu$ -SVM 模型的误差指标值, RW $\nu$ -SVM 的预测精度接近于 W $\nu$ -SVM 预测精度的二倍. 这说明了基于 RW $\nu$ -SVM 模型的预测方法是行之有效的.

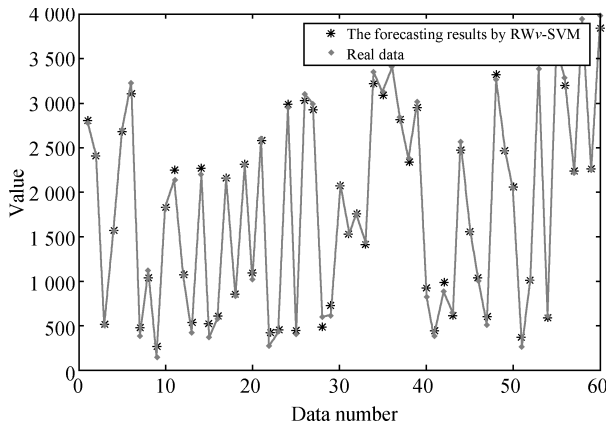
表 2 三种预测模型的误差统计

Model	MAE	MAPE	MSE
ARMA	1209.000	0.965	1970100
W $\nu$ -SVM	88.167	0.0759	11664
RW $\nu$ -SVM	59.583	0.0491	6466

表 1 三种不同模型预测比较

Table 1 Comparison of forecasting results from three different models

No.	1	2	3	4	5	6	7	8	9	10	11	12
Real data	3982	2252	3948	2205	3382	3595	607	3379	1011	259	2056	2457
ARMA	1829	1966	1771	1825	1723	1613	1831	1586	1536	1732	1764	1795
W $\nu$ -SVM	3812	2258	3773	2254	3173	3430	544	3261	1034	399	2042	2431
RW $\nu$ -SVM	3834	2252	3859	2231	3196	3453	585	3475	1007	358	2056	2460

图1 RW $\nu$ -SVM 预测效果图Fig.1 The forecasting results by RW $\nu$ -SVM

### 3 结论

产品销售量的预测是产品生产规划和控制的基础, 随着市场竞争的加剧, 面对日益增加的产品复杂性和多样性, 企业迫切需要精确的产品销售量预测方法, 以面向销售的“零库存”管理来控制生产规模。

企业产品销售量的预测受很多因素影响, 观测得到的因素数据不仅具有模糊性和不确定性, 而且存在小样本问题, 利用传统方法很难建立确切的非线性销售预测模型. 本文提出一种适合于处理小样本, 不确定信息的 RW $\nu$ -SVM 模型可以有效压制各种混合噪音, 提高回归预测精度, 且它具有简洁的对偶形式, 不需要求参数, 函数预测的期望值部分直接由样本数据和小波核函数给出. 实例分析表明 RW $\nu$ -SVM 具有良好的性能, 易于在工程实际中应用。

### References

- Box G, Jenkins G M, Reinsel G. *Time Series Analysis: Forecasting and Control (Third Edition)*. New Jersey: Prentice-Hall, 1994
- Engle R F, Granger C W J. Combining competing forecasts of inflation using a bivariate ARCH model. *Journal of Economic Dynamics and Control*, 1984, **8**(2): 151–165
- Kuo R J. A sales forecasting system based on fuzzy neural network with initial weights generated by genetic algorithm. *European Journal of Operational Research*, 2001, **129**(3): 496–517
- Chatfield C, Yar M. Holt-winters forecasting: some practical issues. *The Statistician*, 1988, **37**: 129–140
- Vapnik V N. *The Nature of Statistical Learning*. New York: Springer, 1995
- Wu Q, Yan H S, Yang H B. A forecasting model based support vector machine and particle swarm optimization. In: *Proceedings on the 2008 Workshop on Power Electronics and Intelligent Transportation System*. Guangzhou, China: IEEE, 2008. 218–222
- Wu Q, Yan H S, Yang H B. A hybrid forecasting model based on chaotic mapping and improved support vector machine. In: *Proceedings on the 9th International Conference for Young Computer Scientists*. Washington D. C., USA: IEEE, 2008. 2701–2706
- Rossi F, Villab N. Support vector machine for functional data classification. *Neurocomputing*, 2006, **69**(7-9): 730–742

- Zhang Ri-Dong, Wang Shu-Qing, Li Ping. Support vector machine based predictive control for nonlinear systems. *Acta Automatica Sinica*, 2007, **33**(10): 1066–1073 (张日东, 王树青, 李平. 基于支持向量机的非线性系统预测控制. *自动化学报*, 2007, **33**(10): 1066–1073)
- Cao Yuan-Yuan, Yang Bo, Xu Guang-You. Netlike texture detection method using fractal texture features and wavelet transform. *Acta Automatica Sinica*, 2007, **33**(7): 688–692 (曹媛媛, 杨波, 徐光祐. 基于分形纹理特征和小波变换的网状纹理检测方法. *自动化学报*, 2007, **33**(7): 688–692)
- Yan H S, Xu D. An approach to estimating product design time based on fuzzy  $\nu$ -support vector machine. *IEEE Transactions on Neural Networks*, 2007, **18**(3): 721–731
- Shim J, Hwang C, Nau S. Robust LS-SVM regression using fuzzy C-means clustering. In: *Proceedings of the 2nd International Conference on Advances in Natural Computation*. Xi'an, China: Springer, 2006. 157–166
- Francis E H T, Cao L J. Modified support vector machines in financial time series forecasting. *Neurocomputing*, 2002, **48**(1-4): 847–861
- Vapnik V N. *The Nature of Statistical Learning*. New York: Springer, 2000
- Khandoker A H, Lai D T H, Begg R K, Palaniswami M. Wavelet-based feature extraction for support vector machines for screening balance impairments in the elderly. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2007, **15**(4): 587–597
- Widodo A, Yang B S. Wavelet support vector machine for induction machine fault diagnosis based on transient current signal. *Expert Systems with Applications: An International Journal*, 2008, **35**(1-2): 307–316
- Fernandez V. Wavelet and SVM based forecasts: an analysis of the U.S. metal and materials manufacturing industry. *Resources Policy*, 2007, **32**(1-2): 80–89
- Wu Q. The forecasting model based on wavelet  $\nu$ -support vector machine. *Expert Systems with Applications: An International Journal*, 2009, **36**(4): 7604–7610

吴 奇 东南大学博士后. 主要研究方向为计算机视觉, 生物计算, 模式识别和计算机集成制造系统. 本文通信作者.

E-mail: wuqi7812@163.com

(WU Qi Postdoctor at Southeast University. His research interest covers computer vision, biologic computing, pattern recognition, and computer integrated manufacturing systems. Corresponding author of this paper.)

严洪森 东南大学自动化学院教授. 主要研究方向为柔性制造系统, 计算机集成制造系统, 敏捷制造, 并行工程和知识化制造系统.

E-mail: Hsyan@seu.edu.cn

(YAN Hong-Sen Professor at the School of Automation, Southeast University. His research interest covers flexible manufacturing systems, computer integrated manufacturing systems, agile manufacturing, concurrent engineering, and knowledgeable manufacturing systems.)

王 斌 东南大学自动化学院博士研究生. 主要研究方向为并行工程, 智能控制和计算机仿真. E-mail: wbsky@163.com

(WANG Bin Ph.D. candidate at the School of Automation, Southeast University. His research interest covers concurrent engineering, intelligent control, and computer emulation.)