

# Product Sale Forecasting Model Based on Iterative Support Vector Machine

Xin Tu<sup>1,2</sup>, Hongsen Yan<sup>1,3</sup>

<sup>1</sup>School of Automation, Southeast University, Nanjing Jiangsu

<sup>2</sup>College of Big Data and Information Engineering, Guizhou University, Guiyang Guizhou

<sup>3</sup>Key Laboratory of Measurement and Control of Complex Systems of Engineering, Ministry of Education, Southeast University, Nanjing Jiangsu

Email: 5654157\_cn@sina.com

Received: Dec. 22<sup>nd</sup>, 2019; accepted: Jan. 6<sup>th</sup>, 2020; published: Jan. 13<sup>th</sup>, 2020

---

## Abstract

Aiming at data characteristics of small samples and noise existing in the product sale series, an iterative support vector machine ( $I\epsilon$ -SVM) is proposed in this paper. During the gradually reducing process of  $I\epsilon$ -SVM's parameter  $\epsilon$ , the samples greatly affected by noise are iteratively amended to reduce their influence on the final forecasting model generated.  $I\epsilon$ -SVM is applied to a numerical value example and the automobile sales forecasting in contrast with the  $\epsilon$  support vector machine ( $\epsilon$ -SVM). The experiment results indicate that  $I\epsilon$ -SVM is effective and feasible, by which more accurate forecasting results are obtained over the  $\epsilon$ -SVM.

## Keywords

Sale Series, Noise, Iteration, Support Vector Machine, Forecasting

---

# 基于迭代的支持向量机的产品销售预测模型

涂 歆<sup>1,2</sup>, 严洪森<sup>1,3</sup>

<sup>1</sup>东南大学自动化学院, 江苏 南京

<sup>2</sup>贵州大学大数据与信息工程学院, 贵州 贵阳

<sup>3</sup>东南大学复杂工程系统测量与控制教育部重点实验室, 江苏 南京

Email: 5654157\_cn@sina.com

收稿日期: 2019年12月22日; 录用日期: 2020年1月6日; 发布日期: 2020年1月13日

---

## 摘 要

针对产品销售时序具有小样本、含噪声的数据特征, 本文设计了一种迭代的支持向量机模型( $I\epsilon$ -SVM)。

**I $\epsilon$ -SVM**采用迭代的方式,在SVM中参数 $\epsilon$ 逐步减小的过程中,一步步修正那些可能受噪声影响较大的样本点信息,降低这些样本点对最终生成的预测模型的影响。**I $\epsilon$ -SVM**与 $\epsilon$ -SVM一起被应用于处理一个数值算例和一个汽车销售预测实例中,仿真实验结果表明**I $\epsilon$ -SVM**是有效可行的,可获得比 $\epsilon$ -SVM更精确的预测结果。

## 关键词

销售时序, 噪声, 迭代, 支持向量机, 预测

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

产品销售预测是一个复杂的动态过程,其预测行为受多重因素影响,这些因素与产品销售量之间通常是非线性关系,目前难以用一种确定性模型来对这种关系进行建模。实际的销售时序常受到一些噪声(Noise)的干扰,在数据的收集,传输或处理过程中都可能出现噪声。如果不能在建模时有效抑制噪声的干扰,将会直接影响到预测模型的预测精度。同时,随着生产力水平的提高,现在的产品生命周期逐渐变短,这导致建立模型时很难收集到足够多的样本数据,所以我们需要面对的是对一个小样本、含噪声的数据集进行非线性建模的问题。

支持向量机(SVM)是 Vapnik 等提出的一种机器学习方法[1],它在解决小样本、非线性及高维模式识别中表现出许多特有的优势[2][3],在处理分类问题与回归问题均取得了很好的效果[3]-[9]。同时,针对样本集数据含有噪声的情况,目前也出现了一些基于支持向量机的针对性方法,取得了不少成果[10][11][12]。文献[11][12][13][14]都是通过在支持向量机方法中引入模糊数来对含噪声的数据进行建模;文献[15]采用重采样的方式来降低噪声的干扰,以期获得较为精确的训练样本集;文献[16]设计了高斯损失函数来抑制样本数据中含有的高斯噪声;文献[17]设计了概率支持向量机来对受噪声影响的数据集进行分类,取得了较好的效果。

尽管支持向量机方法在对小样本的数据集进行非线性建模时具有一定的优势,并已取得了许多的成果;但由于支持向量机方法的特性,其模型建立后仅受支持向量的影响,而支持向量均由偏离回归曲线较远的样本点即通常是受噪声影响较大的样本点构成,已有的噪声抑制方法并不能改变所建立的支持向量机模型对它们的依赖;因此本文提出了迭代的支持向量机方法,文献[18][19][20][21]也研究了迭代的支持向量机方法,但它们是以提高模型的训练效率为目的,通过设计迭代算法来加快寻找支持向量机模型的最优解;而本文方法主要以抑制噪声样本对模型的干扰为目标,依据每次支持向量机的回归结果对训练样本集进行迭代更新,在样本集迭代更新的过程中逐步修正那些可能受噪声影响较大的样本点信息,以此来降低它们对最后生成的回归模型的影响,仿真结果表明该方法在面对小样本、含噪声的数据集时比一般的支持向量机模型预测效果更好。

## 2. 支持向量机

支持向量机(SVM)是一种基于结构风险最小化思想的小样本机器学习方法。设样本集  $T = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_l, y_l)\}$ , 其中  $\mathbf{x}_i \in \mathbf{R}^d$  为  $d$  维输入向量,  $y_i \in \mathbf{R}$  为系统输出,  $i = 1, 2, \dots, l$ . 考虑用线性函数  $f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b$  拟合数据,采用  $\epsilon$  误差不敏感函数,并允许拟合误差  $\xi_i^{(*)}$  存在的软间隔支持向量回归( $\epsilon$ -SVR)可描述如下:

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

其中  $\varepsilon > 0$  为控制拟合精度的参数,  $C > 0$  为惩罚系数,  $\boldsymbol{\xi}^{(*)} = (\xi_1, \xi_1^*, \dots, \xi_l, \xi_l^*)^\top$  为松弛变量。利用 Lagrange 优化方法, 可求得问题(1)的对偶优化问题:

$$\begin{aligned} \min_{\boldsymbol{\alpha}, \boldsymbol{\alpha}^*} \quad & W(\boldsymbol{\alpha}, \boldsymbol{\alpha}^*) = \frac{1}{2} \sum_{i,j=1}^l (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) k(\mathbf{x}_i, \mathbf{x}_j) \\ & - \sum_{i=1}^l (\alpha_i^* - \alpha_i) y_i + \sum_{i=1}^l (\alpha_i^* + \alpha_i) \varepsilon \\ \text{s.t.} \quad & \begin{cases} \sum_{i=1}^l (\alpha_i^* - \alpha_i) = 0 \\ \alpha_i, \alpha_i^* \in [0, C], i=1, \dots, l \end{cases} \end{aligned} \quad (2)$$

其中  $K(\mathbf{x}_i, \mathbf{x}_j) = (F(\mathbf{x}_i) \cdot F(\mathbf{x}_j))$  是核函数, 它等价于将向量  $\mathbf{x}$  映射到空间  $F(\mathbf{x})$  后的内积。目前比较常用的核函数包括: 多项式核函数  $K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j + c)^d$  (其中  $c \geq 0, d$  为任意正整数); 高斯径向基核函数(RBF核)  $K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma^2}\right)$ ; Sigmoid 核函数  $K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\kappa(\mathbf{x}_i \cdot \mathbf{x}_j) + \nu)$  (其中  $\kappa > 0, \nu < 0$ ) 等。

通过求解问题(2)得解向量  $(\boldsymbol{\alpha}, \boldsymbol{\alpha}^*)$ , 对新输入的  $\mathbf{x}$ , 构造回归预测函数

$$f(\mathbf{x}) = \sum_{i=1}^l (\alpha_i^* - \alpha_i) k(\mathbf{x}_i, \mathbf{x}) + b \quad (3)$$

### 3. 迭代支持向量机(I $\varepsilon$ -SVM)

$\varepsilon$ -SVM 采用  $\varepsilon$  不敏感函数, 夹在曲线  $f(\mathbf{x}) = \sum_{i=1}^l (\alpha_i^* - \alpha_i) k(\mathbf{x}_i, \mathbf{x}) + b + \varepsilon$  和

$f(\mathbf{x}) = \sum_{i=1}^l (\alpha_i^* - \alpha_i) k(\mathbf{x}_i, \mathbf{x}) + b - \varepsilon$  中间的区域被称为  $\varepsilon$ -带, 当样本点位于  $\varepsilon$ -带中时则认为模型在该点没有损失, 只有当样本点位于  $\varepsilon$ -带之外时, 才有损失出现。针对样本点  $(\mathbf{x}_i, y_i)$  和  $\varepsilon$ -带之间的位置关系, 我们有下面的定理。

**定理 1 [2]:** 设  $\boldsymbol{\alpha}^{(*)} = (\alpha_1, \alpha_1^*, \dots, \alpha_l, \alpha_l^*)^\top$  是最优化问题(2)的解, 则

- i) 若  $\alpha_i = \alpha_i^* = 0$ , 则相应的样本点  $(\mathbf{x}_i, y_i)$  一定在  $\varepsilon$ -带的内部或边界上。
- ii) 若  $\alpha_i \in (0, C), \alpha_i^* = 0$  或  $\alpha_i = 0, \alpha_i^* \in (0, C)$ , 则相应的样本点  $(\mathbf{x}_i, y_i)$  一定在  $\varepsilon$ -带的边界上。
- iii) 若  $\alpha_i = C, \alpha_i^* = 0$  或  $\alpha_i = 0, \alpha_i^* = C$ , 则相应的样本点  $(\mathbf{x}_i, y_i)$  一定在  $\varepsilon$ -带的外部或边界上。

**证明:** 若  $\boldsymbol{\alpha}^{(*)} = (\alpha_1, \alpha_1^*, \dots, \alpha_l, \alpha_l^*)^\top$  是最优化问题(2)的解。则 KKT 互补条件为:

$$\alpha_i (\varepsilon + \xi_i - y_i + \mathbf{w} \cdot \mathbf{x}_i + b) = 0 \quad (4)$$

$$\alpha_i^* (\varepsilon + \xi_i + y_i - \mathbf{w} \cdot \mathbf{x}_i - b) = 0 \quad (5)$$

$$(C - \alpha_i) \xi_i = 0 \quad (6)$$

$$(C - \alpha_i^*) \xi_i^* = 0 \quad (7)$$

以下分别证明结论 i)~iii):

i) 由式(6)和(7)可以看出  $\xi = \xi_i^* = 0$ 。

ii) 由式(6)和(7)可以看出  $\xi = \xi_i^* = 0$ , 同时结合式(4)和(5)可断定, 当  $\alpha_i \neq 0$  时,  $\varepsilon + \xi_i - y_i + \mathbf{w} \cdot \mathbf{x}_i + b = 0$ ; 当  $\alpha_i^* \neq 0$  时,  $\varepsilon + \xi_i + y_i - \mathbf{w} \cdot \mathbf{x}_i - b = 0$ 。

iii) 由式(4)~(7)可以看出。

定理 1 得证。

由定理 1 可知, 若样本点  $(\mathbf{x}_i, y_i)$  在  $\varepsilon$ -带外部, 则  $\alpha_i = C, \alpha_i^* = 0$  或  $\alpha_i = 0, \alpha_i^* = C$ 。

从定理 1 可以看出, 离最终拟合曲线距离越远的样本点  $(\mathbf{x}_i, y_i)$  对应着更大的  $\alpha_i$  或  $\alpha_i^*$ 。 $\varepsilon$ -SVM 具有很强的建模能力, 但在实际问题当中, 样本点数据  $(\mathbf{x}_i, y_i)$  通常还包含了未知的噪声信息,  $\varepsilon$ -SVM 的建模机理使得最终生成的模型更依赖于那些可能受噪声影响较大的样本点。为避免这一点, 我们提出了迭代的  $\varepsilon$ -SVM(I $\varepsilon$ -SVM)。I $\varepsilon$ -SVM 采用迭代的方式, 通过一步步修正那些可能受噪声影响较大的样本点信息来使得 I $\varepsilon$ -SVM 获得更好的泛化性能。

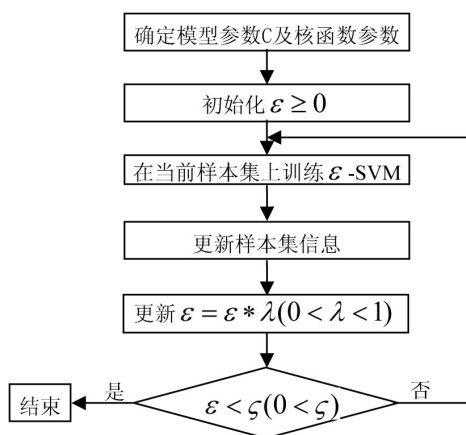


Figure 1. The flow chart of I $\varepsilon$ -SVM

图 1. I $\varepsilon$ -SVM 流程图

I $\varepsilon$ -SVM 的运行机制如图 1 所示, 我们首先确定模型参数  $C$  及可能存在的核函数参数, 为  $\varepsilon$  选取一个较大的初始值, 训练  $\varepsilon$ -SVM, 更新落入  $\varepsilon$ -带外部的样本点  $(\mathbf{x}_i, y_i)$  的值,  $\lambda$  是个预先给定的比例系数, 每次按比例  $\lambda$  缩小  $\varepsilon$ ;  $\zeta$  是一个预先给定的阈值, 如果  $\varepsilon < \zeta$  则终止算法, 否则更新  $\varepsilon$  的值, 并在更新后的样本集上训练  $\varepsilon$ -SVM。

### 更新样本集信息

当选定一个  $\varepsilon$ , 并训练好  $\varepsilon$ -SVM, 更新此时落入  $\varepsilon$ -带外部的样本点  $(\mathbf{x}_i, y_i)$  的值, 让它们向对应的  $\varepsilon$ -带边界靠近。根据定理 1, 有三种可能的情况。

1) 若  $\alpha_i \neq C$  且  $\alpha_i^* \neq C$ , 此时样本点  $(\mathbf{x}_i, y_i)$  在  $\varepsilon$ -带的内部或边界上, 则令  $\tilde{y}_i = y_i$ 。

2) 若  $\alpha_i = C, \alpha_i^* = 0$ , 此时  $y_i \geq \mathbf{w} \cdot \mathbf{x}_i + b + \varepsilon$ , 我们令  $\tilde{y}_i = \mathbf{w} \cdot \mathbf{x}_i + b + \varepsilon$ 。

3) 若  $\alpha_i = 0, \alpha_i^* = C$ , 此时  $y_i \leq \mathbf{w} \cdot \mathbf{x}_i + b - \varepsilon$ , 我们令  $\tilde{y}_i = \mathbf{w} \cdot \mathbf{x}_i + b - \varepsilon$ 。

按上面方法对每个样本点进行更新, 获得更新后的样本集  $\tilde{T} = \{(\mathbf{x}_1, \tilde{y}_1), \dots, (\mathbf{x}_l, \tilde{y}_l)\}$ , 下面定理说明了当  $\varepsilon$ -SVM 作用于样本集  $\tilde{T}$  时可以获得比作用于原样本集  $T$  时更小的  $\|\mathbf{w}\|^2$ 。

定理 2: 对于给定的  $\varepsilon$ -SVM 参数  $\varepsilon, C$ ,  $T = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_l, y_l)\}$  为原样本集,  $\tilde{T} = \{(\mathbf{x}_1, \tilde{y}_1), \dots, (\mathbf{x}_l, \tilde{y}_l)\}$ , 为更新样本点信息后获得的样本集, 将  $\varepsilon$ -SVM 分别作用于样本集  $T$

和  $\tilde{T}$  上获得最优解  $(\bar{\mathbf{w}}, \bar{b}, \bar{\xi}^{(*)})$  和  $(\tilde{\mathbf{w}}, \tilde{b}, \tilde{\xi}^{(*)})$ , 则  $\|\tilde{\mathbf{w}}\|^2 \leq \|\bar{\mathbf{w}}\|^2$ 。

证明: 将  $\varepsilon$ -SVM 作用于样本集  $\tilde{T}$ , 获得如下的最优化问题:

$$\begin{aligned} \min_{\mathbf{w}, b, \xi^{(*)}} \quad & \tau(\mathbf{w}, b, \xi^{(*)}) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\ \text{s.t.} \quad & \begin{cases} \tilde{y}_i - (\mathbf{w} \cdot \mathbf{x}_i + b) \leq \varepsilon + \xi_i \\ (\mathbf{w} \cdot \mathbf{x}_i + b) - \tilde{y}_i \leq \varepsilon + \xi_i^* \\ \xi^{(*)} \geq 0 \end{cases} \end{aligned} \quad (8)$$

样本集  $\tilde{T}$  是由样本集  $T$  更新后获得, 根据更新规则,  $\tilde{y}_i$  只有三种可能取值。

i)  $\tilde{y}_i = y_i$ , 此时  $(\mathbf{x}_i, y_i)$  落入曲线  $y = \bar{\mathbf{w}} \cdot \mathbf{x} + \bar{b}$  的  $\varepsilon$ -带内部或边界上, 有:

$$\begin{aligned} \tilde{y}_i - (\bar{\mathbf{w}} \cdot \mathbf{x}_i + \bar{b}) &\leq \varepsilon, \\ (\bar{\mathbf{w}} \cdot \mathbf{x}_i + \bar{b}) - \tilde{y}_i &\leq \varepsilon. \end{aligned}$$

ii)  $\tilde{y}_i = \bar{\mathbf{w}} \cdot \mathbf{x}_i + \bar{b} + \varepsilon$ , 此时  $(\mathbf{x}_i, y_i)$  落入曲线  $y = \bar{\mathbf{w}} \cdot \mathbf{x} + \bar{b}$  的  $\varepsilon$ -带外部或边界上, 有:

$$\begin{aligned} \tilde{y}_i - (\bar{\mathbf{w}} \cdot \mathbf{x}_i + \bar{b}) &\leq \varepsilon, \\ (\bar{\mathbf{w}} \cdot \mathbf{x}_i + \bar{b}) - \tilde{y}_i &\leq \varepsilon. \end{aligned}$$

iii)  $\tilde{y}_i = \bar{\mathbf{w}} \cdot \mathbf{x}_i + \bar{b} - \varepsilon$ , 此时  $(\mathbf{x}_i, y_i)$  同样落入曲线  $y = \bar{\mathbf{w}} \cdot \mathbf{x} + \bar{b}$  的  $\varepsilon$ -带外部或边界上, 有:

$$\begin{aligned} \tilde{y}_i - (\bar{\mathbf{w}} \cdot \mathbf{x}_i + \bar{b}) &\leq \varepsilon, \\ (\bar{\mathbf{w}} \cdot \mathbf{x}_i + \bar{b}) - \tilde{y}_i &\leq \varepsilon. \end{aligned}$$

由(i)~(iii)可知  $(\bar{\mathbf{w}}, \bar{b}, 0)$  是最优化问题(8)的可行解, 所以有:

$$\tau(\bar{\mathbf{w}}, \bar{b}, 0) \geq \tau(\tilde{\mathbf{w}}, \tilde{b}, \tilde{\xi}^{(*)}),$$

即  $\frac{1}{2} \|\bar{\mathbf{w}}\|^2 + C \sum_{i=1}^l (\bar{\xi}_i + \bar{\xi}_i^*) \leq \frac{1}{2} \|\tilde{\mathbf{w}}\|^2$ , 而  $\bar{\xi}^{(*)} \geq 0$ , 所以有:

$$\|\tilde{\mathbf{w}}\|^2 \leq \|\bar{\mathbf{w}}\|^2.$$

## 4. 仿真实验

为验证  $\text{I}\varepsilon$ -SVM 的有效性, 我们将  $\text{I}\varepsilon$ -SVM 应用于处理一个数值算例和一个应用实例中, 并选取  $\varepsilon$ -SVM 与本文提出的  $\text{I}\varepsilon$ -SVM 进行比较, 通过网格搜索选取各自最优的参数组合。

### 4.1. 比较准则

假设预测模型在某个测试样本点误差为  $e_i$ :

$$e_i = |y_i - y'_i|, \quad (9)$$

其中  $y_i$  是第  $i$  个样本的真实输出值,  $y'_i$  是预测模型在该点的预测值。为了综合考察所建立模型的预测效果, 通过计算  $\text{I}\varepsilon$ -SVM 与  $\varepsilon$ -SVM 在测试样本集  $T_2$  上预测误差的统计特性来比较它们所建模型的优劣。

1) 最大预测误差( $\max(e)$ )

$$\max(e) = \max_{1 \leq i \leq l_2} (e_i), \quad (10)$$

其中  $l_2$  为测试样本集  $T_2$  的规模。

2) 最小预测误差( $\min(e)$ )

$$\min(e) = \min_{1 \leq i \leq l_2} (e_i) \quad (11)$$

3) 平均预测误差( $\text{mean}(e)$ )

$$\text{mean}(e) = \frac{1}{l_2} \left( \sum_{i=1}^{l_2} e_i \right) \quad (12)$$

4) 预测误差方差( $\text{var}(e)$ )

$$\text{var}(e) = \frac{1}{l_2 - 1} \sum_{i=1}^{l_2} (e_i - \text{mean}(e))^2 \quad (13)$$

## 4.2. 数值算例

为了更好的考察  $\varepsilon$ -SVM 与  $\text{I}\varepsilon$ -SVM 在噪声环境下的建模能力, 我们建立一个只包含一个自变量的回归模型,

$$y = x + \xi, \quad (14)$$

$\xi$  是服从  $N(0, 0.05)$  的随机数,  $x$  分别取区间  $[1, 2]$  中平均分布的 36 个节点, 并通过式 14 计算相应的  $y$  值, 产生 36 组数据  $(x, y)$  构成样本集  $T$ 。将样本集  $T$  分为训练样本集  $T_1$  和测试样本集  $T_2$  两个部分, 训练样本集  $T_1$  取样本集  $T$  得前  $l_1$  个样本, 余下样本作为测试样本集  $T_2$ 。 $l_1$  分别取 6、12、24 三种不同情况, 以便比较两种模型在不同规模训练样本集下的表现。分别将  $\varepsilon$ -SVM 与  $\text{I}\varepsilon$ -SVM 作用于训练样本集  $T_1$  建立模型,  $\varepsilon$ -SVM 与  $\text{I}\varepsilon$ -SVM 采用相同的核函数:

$$k(x_i, x_j) = x_i \cdot x_j$$

通过网络搜索的方式寻找到  $\varepsilon$ -SVM 的最优参数组合  $\varepsilon = 0.001, C = 38.4$ ,  $\text{I}\varepsilon$ -SVM 的参数  $C$  同样取为 38.4,  $\varepsilon$  的初始取值为 0.2, 比例缩小系数  $\lambda = 0.8$ , 阈值  $\zeta = 10^{-4}$ 。

表 1 列出了两种模型在不同规模训练样本集下的预测误差比较, 当训练样本集规模  $l_1 = 24$  时, 两种模型的预测效果非常接近,  $\text{I}\varepsilon$ -SVM 的各项指标稍优于  $\varepsilon$ -SVM; 当  $l_1 = 12$  时,  $\varepsilon$ -SVM 具有更小的最大和最小误差, 但  $\text{I}\varepsilon$ -SVM 具有更小的平均误差; 当  $l_1 = 6$  时,  $\text{I}\varepsilon$ -SVM 的预测误差的各项统计指标均明显优于  $\varepsilon$ -SVM。综合来看, 两种模型的预测误差都随着训练样本集规模  $l_1$  的减小而增大, 但在三种情况下  $\text{I}\varepsilon$ -SVM 所获的平均预测误差均小于  $\varepsilon$ -SVM 的平均预测误差, 且它们之间的差距随着  $l_1$  的减小而增大, 由此可见,  $\text{I}\varepsilon$ -SVM 更适合于对具有小样本、含噪声特点的数据集进行建模。

**Table 1.** The forecasting errors of two models in the numerical experiment

**表 1.** 数值算例中两种模型的预测误差比较

		Max( $e$ )	Min( $e$ )	Mean( $e$ )	Var( $e$ )
$l_1 = 24$	$\varepsilon$ -SVM	0.1063	4.68e-04	0.0298	0.0012
	$\text{I}\varepsilon$ -SVM	0.1059	5.69e-05	0.0296	0.0012
$l_1 = 12$	$\varepsilon$ -SVM	0.1024	3.39e-04	0.0388	0.0010
	$\text{I}\varepsilon$ -SVM	0.1123	0.0012	0.0363	0.0012
$l_1 = 6$	$\varepsilon$ -SVM	0.1500	0.0036	0.0627	0.0015
	$\text{I}\varepsilon$ -SVM	0.1408	0.0017	0.0582	0.0013

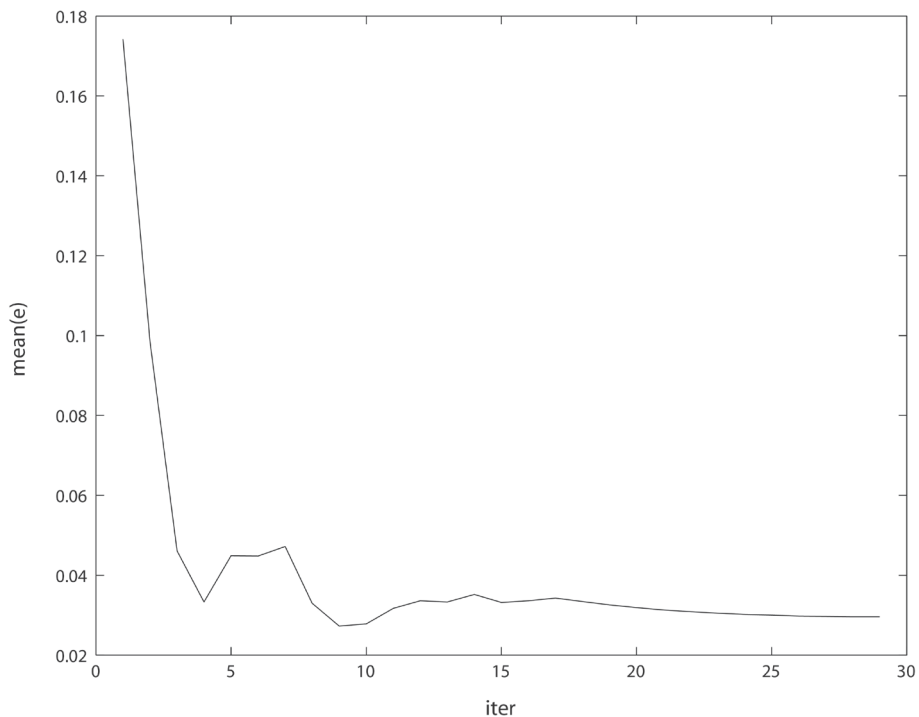


Figure 2. The decline curve of the average forecasting error of  $l_{\varepsilon}$ -SVM in iteration when  $l_1 = 6$

图 2.  $l_1 = 6$  时,  $l_{\varepsilon}$ -SVM 在迭代过程中的平均预测误差下降曲线

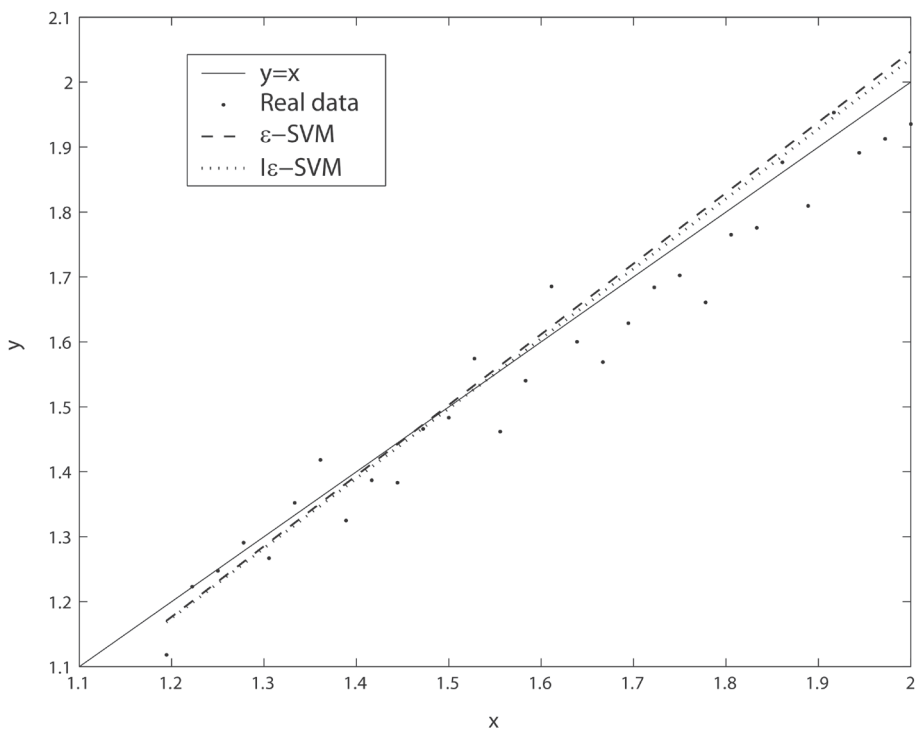


Figure 3. The forecasting results of two models when  $l_1 = 6$

图 3.  $l_1 = 6$  时, 两种模型的预测效果比较图

图 2 是  $l_{\varepsilon}$ -SVM 在  $l_1 = 6$  的训练样本集上的平均预测误差下降曲线, 在开始阶段, 随着  $\varepsilon$  的减小,  $\text{mean}(e)$

有一个快速下降的过程, 而在迭代的最后阶段,  $\text{mean}(e)$  处于一个非常平缓的下降过程。所以在迭代过程中,  $I\epsilon$ -SVM 的预测误差可以近似看成是收敛的。图 3 是在  $l_1 = 6$  的训练样本集上, 两种模型的预测效果比较图。从图中可以看出,  $\epsilon$ -SVM 与  $I\epsilon$ -SVM 所建立的模型非常接近, 与直线  $y = x$  均有一定偏差, 但从图的右上角部分仍可明显看出  $I\epsilon$ -SVM 所建立的模型更接近直线  $y = x$ 。

### 4.3. 应用实例

以汽车销售量的预测为例, 考虑销售量的影响因子集由 6 个元素组成, 分别是居民可支付购买力 (Purchases of People) PP、市场竞争程度 (Market Competition) MC、季节性因素 (Season Component) SC、汽车油价影响因素 (Oil Factor) OF、汽车性能参数 (Performance Factor) PF 和汽车价格波动因素 (Price Wave) PW。其中 PP, MC, SC, OF, PF 和 PW 为六个语言型数据, 利用综合评价法获取其数值型数据。选取 48 个月度的汽车销售数据加上由综合评价法获取的  $48 \times 6$  个影响因素集的数值型数据组成样本集  $T$  进行实例研究。同样将样本集  $T$  分为训练样本集  $T_1$  和测试样本集  $T_2$  两个部分, 前 36 组数据作为  $T_1$ , 后 12 组数据作为  $T_2$ 。分别将  $\epsilon$ -SVM 与  $I\epsilon$ -SVM 作用于训练样本集  $T_1$  建立模型,  $\epsilon$ -SVM 与  $I\epsilon$ -SVM 采用相同的核函数:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma^2}\right)$$

通过网格搜索的方式寻找到  $\epsilon$ -SVM 的最优参数组合  $\epsilon = 0.0059, C = 2155, \sigma = 6.354$ ,  $I\epsilon$ -SVM 采用与  $\epsilon$ -SVM 相同的  $C$  和  $\sigma$ ,  $\epsilon$  的初始取值为 0.2, 比例缩小系数  $\lambda = 0.8$ , 阈值  $\zeta = 10^{-4}$ 。

表 2 列出了  $\epsilon$ -SVM 与  $I\epsilon$ -SVM 建立的汽车销量模型的预测误差比较,  $I\epsilon$ -SVM 明显获得了更好的预测效果,  $\epsilon$ -SVM 仅在  $\min(e)$  一项稍优于  $I\epsilon$ -SVM, 其它各项指标均劣于  $I\epsilon$ -SVM。

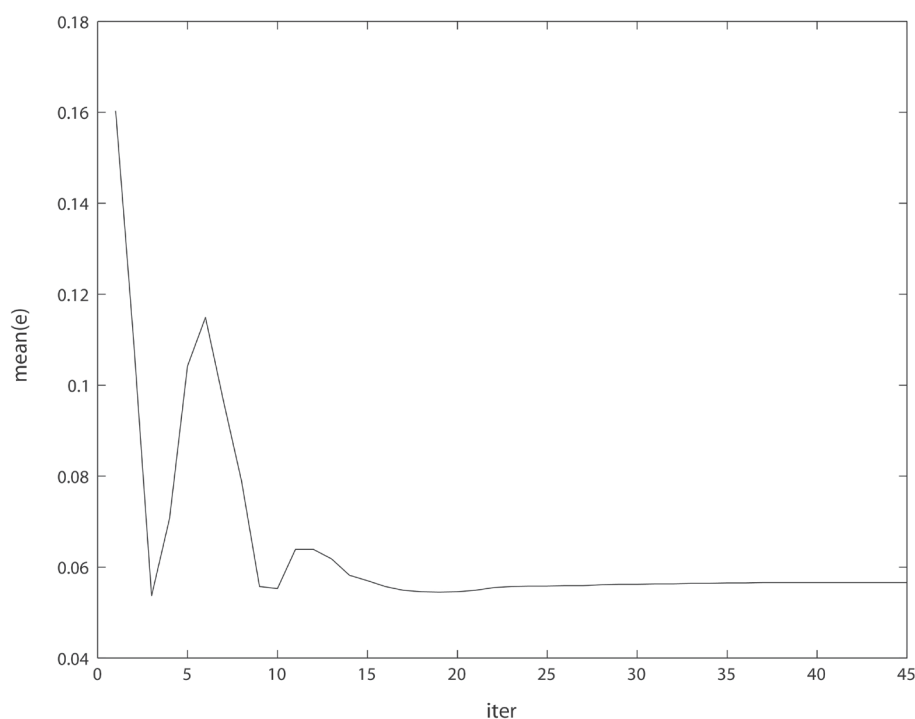


Figure 4. The decline curve of the average forecasting error of  $I\epsilon$ -SVM in iteration

图 4.  $I\epsilon$ -SVM 在迭代过程中的平均预测误差下降曲线



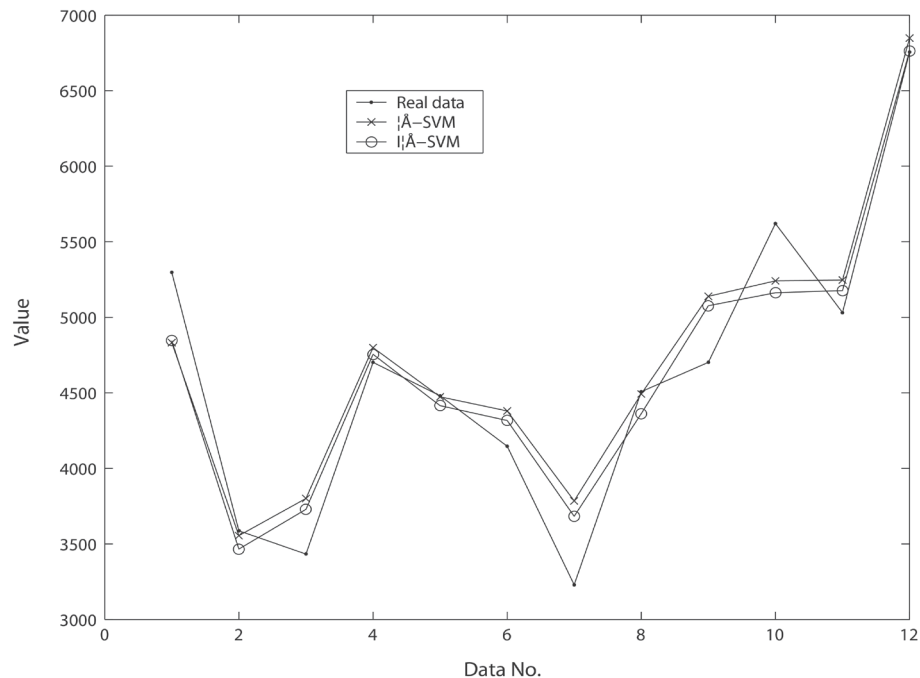


Figure 5. The forecasting results of two models  
图 5. 两种模型的预测效果比较图

Table 2. The forecasting errors of two models in the practical experiment  
表 2. 应用实例中两种模型的预测误差比较

	Max( $e$ )	Min( $e$ )	Mean( $e$ )	Var( $e$ )
$\varepsilon$ -SVM	0.1330	0.0017	0.0578	0.0022
$l\varepsilon$ -SVM	0.1101	0.0018	0.0549	0.0017

图 4 是  $l\varepsilon$ -SVM 在迭代过程中的平均预测误差下降曲线图, 同样在迭代的最后阶段, Mean( $e$ ) 处于一个非常平缓的下降过程, 可以看成在迭代过程中  $l\varepsilon$ -SVM 的预测误差是近似收敛的。图 5 是两种模型的预测效果比较图。从图中可以看出,  $\varepsilon$ -SVM 与  $l\varepsilon$ -SVM 所建立的模型在 12 个预测点中互有优劣, 在第 2、5、8、10 号预测点,  $\varepsilon$ -SVM 的预测效果更好, 在第 1、4、12 号预测点, 两种模型的预测效果接近, 而在剩下的 5 个点, 则是  $l\varepsilon$ -SVM 所获的预测值更接近于真实值。

## 5. 结论

产品销售预测是企业生产规划和库存控制的基础。在激烈的市场竞争环境中, 随着产品的复杂性和多样性的增加, 企业迫切希望能对未来产品的需求量有更精确的预测, 据此制定生产计划, 达到在一段时间内供应链成本最优的目的。

针对产品销售预测时序具有小样本、含噪声, 非线性的数据特征, 本文设计了迭代的支持向量机方法, 通过在迭代过程中逐步修正那些可能受噪声影响较大的样本点信息来降低这些样本点对最后生成的支持向量机模型的影响, 所获方法与  $\varepsilon$ -SVM 进行了数值仿真比较实验, 结果表明所获方法在预测精度上具有一定的优越性。

## 基金项目

国家自然科学基金(50875046); 国家自然科学基金(60934008); 国家自然科学基金(61065010)。

## 参考文献

- [1] Vapnik, V.N. (2000) *The Nature of Statistical Learning Theory*. Springer, New York, 138-167. <https://doi.org/10.1007/978-1-4757-3264-1>
- [2] 邓乃扬, 田英杰. 数据挖掘中的新方法—支持向量机[M]. 北京: 科学出版社, 2006.
- [3] 周辉仁, 郑丕谔, 任仙玲. 最小二乘支持向量机的参数优选方法及应用[J]. 系统工程学报, 2009, 24(2): 248-252.
- [4] 姚宝珍, 杨成永, 于滨. 动态公交车辆运行时间预测模型[J]. 系统工程学报, 2010, 25(3): 365-370.
- [5] 李大铭, 于滨. 公交运营的协控准点滞站调度模型[J]. 系统工程学报, 2012, 27(2): 248-255.
- [6] 商志根, 严洪森. 基于模糊支持向量机的产品设计时间预测[J]. 控制与决策, 2012, 27(4): 531-534.
- [7] 时培明, 梁凯, 赵娜, 等. 基于深度学习特征提取和粒子群支持向量机状态识别的齿轮智能故障诊断[J]. 中国机械工程, 2017, 28(9): 1056-1061.
- [8] 张燕君, 王会敏, 付兴虎, 等. 基于粒子群支持向量机的钢板损伤位置识别[J]. 中国激光, 2017, 44(10): 197-203.
- [9] 肖鹏飞, 张超勇, 罗敏, 等. 基于自适应动态无偏最小二乘支持向量机的刀具磨损预测建模[J]. 中国机械工程, 2018, 29(7): 842-849.
- [10] 张英, 苏宏业, 褚健. 基于模糊最小二乘支持向量机的软测量建模[J]. 控制与决策, 2005, 20(6): 621-624.
- [11] 姚潇, 余乐安. 模糊近似支持向量机模型及其在信用风险评估中的应用[J]. 系统工程理论与实践, 2012, 32(3): 549-554.
- [12] Lin, C.F. and Wang, S.D. (2004) Training Algorithms for Fuzzy Support Vector Machines with Noisy Data. *Pattern Recognition Letters*, **25**, 1647-1656. <https://doi.org/10.1016/j.patrec.2004.06.009>
- [13] Li, D., Hu, W.C., Xiong, W. and Yang, J.B. (2008) Fuzzy Relevance Vector Machine for Learning from Unbalanced Data and Noise. *Pattern Recognition Letters*, **29**, 1175-1181. <https://doi.org/10.1016/j.patrec.2008.01.009>
- [14] 张桂香, 费岚, 杜喆, 刘三阳. 非均衡数据的去噪模糊支持向量机新方法[J]. 计算机工程与应用, 2008, 44(16): 142-144.
- [15] 蒋蔚, 伊国兴, 曾庆双. 一种基于 SVM 重采样的似然粒子滤波算法[J]. 控制与决策, 2011, 26(2): 243-247.
- [16] 吴奇, 严洪森. 基于具有高斯损失函数支持向量机的预测模型[J]. 计算机集成制造系统, 2009, 15(2): 306-312.
- [17] Li, H.X., Yang, J.L., Zhang, G. and Fan, B. (2013) Probabilistic Support Vector Machines for Classification of Noise Affected Data. *Information Sciences*, **221**, 60-71. <https://doi.org/10.1016/j.ins.2012.09.041>
- [18] Keerthi, S.S., Shevade, S.K., Bhattacharyya, C. and Murthy, K.R.K. (2000) A Fast Iterative Nearest Point Algorithm for Support Vector Machine Classifier Design. *IEEE Transactions on Neural Networks*, **11**, 124-136. <https://doi.org/10.1109/72.822516>
- [19] Zhou, S.S., Liu, H.W., Ye, F. and Zhou, L.H. (2009) A New Iterative Algorithm Training SVM. *Optimization Methods and Software*, **24**, 913-932. <https://doi.org/10.1080/10556780902867906>
- [20] Ye, Q.L., Zhao, C.X., Ye, N. and Chen, Y.N. (2010) Iterative Support Vector Machine with Guaranteed Accuracy and Run Time. *Expert Systems*, **27**, 338-348. <https://doi.org/10.1111/j.1468-0394.2010.00550.x>
- [21] Liu, D.H., Qian, H., Dai, G. and Zhang, Z. (2013) An Iterative SVM Approach to Feature Selection and Classification in High-Dimensional Datasets. *Pattern Recognition*, **46**, 2531-2537. <https://doi.org/10.1016/j.patcog.2013.02.007>