

基于具有自适应分段损失函数支持向量机的产品销售预测模型

涂 歆^{1,2}, 严洪森¹

(1. 东南大学 a. 自动化学院, b. 复杂工程系统测量与控制教育部重点实验室, 南京 210096; 2. 贵州大学 大数据与信息工程学院, 贵阳 550025)

摘要: 针对产品销售时序包含噪声的数据特征, 提出一种基于自适应分段损失函数的支持向量机模型(AS ϵ -SVM). AS ϵ -SVM 为每个样本点赋一个单独的不敏感损失值, 以此降低模型对包含较大噪声的样本点的依赖性, 并从理论上证明了该方法可增强模型部分的泛化性能. 将 AS ϵ -SVM 与 ϵ -SVM 共同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, AS ϵ -SVM 是有效可行的, 可获得比 ϵ -SVM 更精确的预测结果.

关键词: 销售时序; 噪声; 自适应分段损失函数; 支持向量机; 预测

中图分类号: TP391

文献标志码: A

Product sale forecasting method based on support vector machine with adaptive segmented loss function

TU Xin^{1,2}, YAN Hong-sen¹

(1a. School of Automation, 1b. MOE Key Laboratory of Measurement and Control of Complex Systems of Engineering, Southeast University, Nanjing 210096, China; 2. College of Big Data and Information Engineering, Guizhou University, Guiyang 550025, China. Correspondent: TU Xin, E-mail: 5654157.cn@sina.com)

Abstract: Aiming at data characteristics of noise existing in the product sale series, a support vector machine based on the adaptive segmented loss function(AS ϵ -SVM) is proposed. In the AS ϵ -SVM, a separate insensitive loss value is assigned to each sample point adaptively, which can reduce the influence of inaccurate samples on the final model. It is proved in theory that the method can enhance partial generalization performance of the model. The AS ϵ -SVM is applied to a numerical value example and the automobile sales forecasting in contrast with the ϵ support vector machine (ϵ -SVM). The experiment results show that the AS ϵ -SVM is effective and feasible, by which more accurate forecasting results are obtained over the ϵ -SVM.

Keywords: sale series; noise; adaptive segmented loss function; support vector machine; forecasting

0 引 言

为达到在一段时间内供应链成本最优的目的, 企业迫切希望能对未来产品的需求量有精确的预测, 据此制定生产计划, 以适应激烈的市场竞争环境. 由于产品自身特点的不同, 产品销量通常受多个内外部因素的影响, 它们之间的关系通常难以用一种确定性模型来刻画. 实际的销售时序通常包含一些随机噪声(Noise), 在数据的收集、传输或处理过程中均可能出现噪声. 如果不能在建立产品销售预测模型时有效抑制噪声的干扰, 则会直接影响到所建立模型的预测精度.

支持向量机(SVM)是基于统计学习理论的一种

机器学习方法^[1], 它克服了传统分类器与神经网络的过学习、局部极值点和维数灾难等缺点, 具有较强的泛化能力^[2]. 通过引入核函数, 可以对小样本的数据集进行非线性建模, 已广泛应用于处理分类问题和回归问题^[3-9]. 针对样本集数据含有噪声的情况, 一些针对性的方法也取得了不少成果^[10-12]. 文献[10-14]运用模糊支持向量机的方法来减少噪声在 SVM 建模过程造成的负面影响. 文献[15-16]运用小波变换对样本数据进行预处理, 以达到降噪的目的. 文献[17]利用重采样的方式获得较为精确的训练样本集, 以减少噪声的干扰. 文献[18]设计了高斯损失函数抑制样本数据中含有的高斯噪声. 文献[19]设计了概率支持向

收稿日期: 2013-11-05; 修回日期: 2015-01-10.

基金项目: 国家自然科学基金项目(50875046, 60934008, 61065010); 中央高校基本科研业务费专项资金项目(2242014K10031).

作者简介: 涂歆(1982-), 男, 博士生, 从事智能算法、人工智能的研究; 严洪森(1957-), 男, 教授, 博士生导师, 从事预测、生产计划与调度、知识化制造等研究.

量机对受噪声影响的数据集进行分类,并取得了较好的效果.文献[20-22]将损失函数设计为分段函数的形式,预先给定统一的阈值,各样本点依据阈值判定具体使用何种形式的损失函数.尽管这样可以在建模时减少含噪声样本点的影响力,但它们对模型的影响仍大于那些较精确的样本点.文献[23]在对噪声数据集进行分类时考虑了类内部的分散度,以抑制部分噪声数据的干扰.文献[24]提出了一种加权的最小二乘支持向量机来处理含噪声的样本数据集,但如何为每个样本点赋予一个合理的权值是该方法应用中的一个难点.文献[25]基于贝叶斯方法提出了一种新的损失函数来抑制含有一般噪声的样本数据集,取得了较好的效果.

尽管支持向量机在含噪声的数据集进行非线性建模领域取得了许多成果,但是在支持向量机回归方法中,不敏感损失函数值被用于控制回归估计的精度,普通的支持向量机模型对各样本点通常采用相同的不敏感损失值计算方法,这导致了在假设回归曲线足够精确的前提下,回归模型会在那些可能受噪声影响较大的样本点处产生较大的损失,以至于回归模型会在尽力拟合这些样本点的过程中丧失部分的泛化性能,现有的抑制噪声的方法并不能减小支持向量机对这些含噪声较大的样本点的依赖.基于此,本文提出了基于自适应分段损失函数的支持向量机方法(AS ε -SVM),通过对可能受噪声影响较大的样本点自动赋予一个较大的不敏感损失值 ε_i 来降低这些样本点对最后生成的支持向量机模型的影响,以此提升模型的泛化性能.仿真结果表明,所提出的方法在面对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.

1 AS ε -SVM 模型

1.1 支持向量机

设样本集 $T = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_l, y_l)\}$,其中 $\mathbf{x}_i \in R^d$ 为 d 维输入向量, $y_i \in R$ 为系统输出.用线性函数 $f(\mathbf{x}) = \mathbf{w}\mathbf{x} + b$ 拟合数据,采用 ε 误差不敏感函数的 ε -SVM模型可描述如下:

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

其中: $C > 0$ 为惩罚系数, $\varepsilon > 0$ 为控制拟合精度的参数, $\xi^{(*)}$ 为松弛变量组成的向量.利用Lagrange优化方法,可求得问题(1)的对偶优化问题

$$\begin{aligned} \min_{\mathbf{a}, \mathbf{a}^{(*)}} & \frac{1}{2} \sum_{i,j=1}^l (a_i - a_i^*)(a_j - a_j^*) \mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) - \\ & \sum_{i=1}^l (a_i - a_i^*) y_i + \sum_{i=1}^l (a_i + a_i^*) \varepsilon; \\ \text{s.t. } & \sum_{i=1}^l (a_i - a_i^*) = 0, \\ & a_i, a_i^* \in [0, C], i = 1, 2, \dots, l. \end{aligned} \quad (2)$$

其中 $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{F}(\mathbf{x}_i)\mathbf{F}(\mathbf{x}_j))$ 为核函数,它等价于将向量 \mathbf{x} 映射到空间 $\mathbf{F}(\mathbf{x})$ 后的内积.通过求解问题(2)得到解向量 $(\mathbf{a}, \mathbf{a}^*)$,对新输入的 \mathbf{x} ,构造回归预测函数

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

1.2 自适应分段 ε 支持向量机(AS ε -SVM)

ε -SVM采用 ε 不敏感损失函数,当样本点位于曲线 $f(\mathbf{x}) = \sum_{i=1}^l (a_i^* - a_i) \mathbf{K}(\mathbf{x}_i, \mathbf{x}) + b + \varepsilon$ 和 $f(\mathbf{x}) = \sum_{i=1}^l (a_i^* - a_i) \mathbf{K}(\mathbf{x}_i, \mathbf{x}) + b - \varepsilon$ 之间的带子里时,认为在该点没有损失,称两条曲线构成的带子为 ε -带.只有当样本点位于 ε -带之外时,才有损失出现,且这些样本点如果离 ε -带越远,则产生的损失越大.实际问题中收集到的样本数据通常包含了未知的噪声信息,这些离最终拟合曲线距离很远的样本点 (\mathbf{x}_i, y_i) 很可能是受噪声影响较大的样本点, ε -SVM的建模机理导致了在建模过程中, ε -SVM为了拟合这些受噪声影响较大的样本点而损失模型的部分泛化性能.

为了避免这种情况,采用对 ε 进行分段的方法,为每一个样本点 (\mathbf{x}_i, y_i) 赋予不同的不敏感损失值 ε_i ,对那些离拟合曲线较远的样本点赋较大的 ε_i ,以此提升模型的泛化性能.采用分段 ε 的支持向量机可描绘如下:

$$\begin{aligned} \min_{\mathbf{w}, b, \xi^{(*)}} & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*); \\ \text{s.t. } & y_i - (\mathbf{w}\mathbf{x}_i + b) \leq \varepsilon_i + \xi_i, \\ & (\mathbf{w}\mathbf{x}_i + b) - y_i \leq \varepsilon_i + \xi_i^*, \\ & \xi^{(*)} \geq 0. \end{aligned} \quad (4)$$

为求问题(4)的对偶优化问题,定义Lagrange函数如下:

$$\begin{aligned} L(\mathbf{w}, b, \mathbf{a}^{(*)}, \xi^{(*)}, \eta^{(*)}) = \\ \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) - \sum_{i=1}^l (\eta_i \xi_i + \eta_i^* \xi_i^*) - \end{aligned}$$

$$\sum_{i=1}^l a_i(\varepsilon_i + \xi_i - y_i + \mathbf{w}\mathbf{x}_i + b) - \sum_{i=1}^l a_i^*(\varepsilon_i + \xi_i^* + y_i - \mathbf{w}\mathbf{x}_i - b).$$

分别对 $\mathbf{w}, b, \xi^{(*)}$ 计算偏导, 由 $\nabla_{\mathbf{w}}L = 0, \nabla_b L = 0, \nabla_{\xi^{(*)}}L = 0$, 得到

$$\mathbf{w} = \sum_{i=1}^l (a_i - a_i^*)\mathbf{x}_i, \sum_{i=1}^l (a_i - a_i^*) = 0, C - a_i^{(*)} - \eta_i^{(*)} = 0.$$

将上述极值条件代入 Lagrange 函数, 并对 $\mathbf{a}^{(*)}$ 求极大, 得到问题 (4) 的对偶问题. 为了与问题 (2) 相对应, 将目标函数乘一负号, 转化为极小化问题, 即

$$\begin{aligned} \min_{\mathbf{a}, \mathbf{a}^{(*)}} & \frac{1}{2} \sum_{i,j=1}^l (a_i - a_i^*)(a_j - a_j^*)\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) - \sum_{i=1}^l (a_i - a_i^*)y_i + \sum_{i=1}^l (a_i + a_i^*)\varepsilon_i; \\ \text{s.t.} & \sum_{i=1}^l (a_i - a_i^*) = 0, \\ & a_i, a_i^* \in [0, C], i = 1, 2, \dots, l. \end{aligned} \quad (5)$$

求解问题 (5) 得到解向量 $(\mathbf{a}, \mathbf{a}^*)$, 代入式 (3) 获得回归预测模型.

2 分段 ε 的计算方法

AS ε -SVM 需要为每个样本 (\mathbf{x}_i, y_i) 赋予一个不同的不敏感损失值 ε_i , 假设 $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_l)$. 为了获得 ε , 选择一个较小的初值 ε_0 , 令 $\varepsilon_i = \varepsilon_0, i = 1, 2, \dots, l$, 通过迭代的方式逐步更新 ε_i 的值以确定最后的 ε . 计算分段 ε 的迭代算法描述如下.

Step 1: 确定 SVM 模型参数 C 和核函数参数, 确定 ε 的迭代初值 ε_0 和迭代步长 $\varepsilon_s > 0$, 初始化 $\varepsilon_i = \varepsilon_0, i = 1, 2, \dots, l, \bar{\varepsilon} = \varepsilon_0$.

Step 2: 更新 $\bar{\varepsilon}$, 令 $\bar{\varepsilon} = \varepsilon_0 + \varepsilon_s$.

Step 3: 将参数集 $(C, \bar{\varepsilon})$ 代入模型 (2) 中求得 $(\mathbf{a}, \mathbf{a}^*)$, 将 $(\mathbf{a}, \mathbf{a}^*)$ 代入式 (3) 中求得每个样本点的预测值 $f(\mathbf{x}_i)$.

Step 4: 计算此时落入回归曲线 ε -带之外的样本点指标集 I , 若 $|y_i - f(\mathbf{x}_i)| > \bar{\varepsilon}$, 则 $i \in I$.

Step 5: 若指标集 I 不为空, 则对于每个 $i \in I$, 更新 $\varepsilon_i = \bar{\varepsilon}$, 转至 Step 2, 否则结束算法.

将由上述算法确定的 ε 代入模型 (5), 求解获得最终的回归预测模型. 从确定 ε 的算法步骤中可以看出, 本文为距离回归曲线较远的样本点 (\mathbf{x}_i, y_i) 赋予了一个较大的不敏感损失值 ε_i , 该值不会大到使样本点 (\mathbf{x}_i, y_i) 在建立 AS ε -SVM 模型时造成损失, 以至于

浪费该样本点的信息, 同时又能降低 AS ε -SVM 在该点的损失, 避免在建立回归模型时需拟合该样本点而丧失模型的部分泛化性能.

引理 1^[2] 最优化问题 (1) 关于 \mathbf{w} 的解唯一.

引理 2 对于给定的 SVM 参数 (ε_0, C) , 按上述算法计算 $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_l)$, 将参数集 (ε_0, C) 与 (ε, C) 分别代入模型 (1)、(4), 得到最优解 $(\bar{\mathbf{w}}, \bar{b}, \bar{\xi}^{(*)})$ 和 $(\tilde{\mathbf{w}}, \tilde{b}, \tilde{\xi}^{(*)})$. 将 $(\bar{\mathbf{w}}, \bar{b})$ 和 $(\tilde{\mathbf{w}}, \tilde{b})$ 分别代入模型 (4)、(1), 求最小的 $\xi^{(*)}$, 获得可行解 $(\bar{\mathbf{w}}, \bar{b}, \bar{\gamma}^{(*)})$ 和 $(\tilde{\mathbf{w}}, \tilde{b}, \tilde{\gamma}^{(*)})$, 则有

$$\begin{aligned} 0 & \leq C \sum_{i=1}^l (\bar{\xi}_i + \bar{\xi}_i^*) - C \sum_{i=1}^l (\bar{\gamma}_i + \bar{\gamma}_i^*) \leq \\ & C \sum_{i=1}^l (\tilde{\gamma}_i + \tilde{\gamma}_i^*) - C \sum_{i=1}^l (\tilde{\xi}_i + \tilde{\xi}_i^*) \leq \\ & C \sum_{i=1}^l (\varepsilon_i - \varepsilon_0). \end{aligned}$$

若

$$\begin{aligned} C \sum_{i=1}^l (\bar{\xi}_i + \bar{\xi}_i^*) - C \sum_{i=1}^l (\bar{\gamma}_i + \bar{\gamma}_i^*) & = \\ C \sum_{i=1}^l (\tilde{\gamma}_i + \tilde{\gamma}_i^*) - C \sum_{i=1}^l (\tilde{\xi}_i + \tilde{\xi}_i^*), \end{aligned}$$

则有

$$\begin{aligned} \frac{1}{2} \|\bar{\mathbf{w}}\|^2 + C \sum_{i=1}^l (\bar{\xi}_i + \bar{\xi}_i^*) & = \\ \frac{1}{2} \|\tilde{\mathbf{w}}\|^2 + C \sum_{i=1}^l (\tilde{\gamma}_i + \tilde{\gamma}_i^*). \end{aligned}$$

证明 因为 $(\bar{\mathbf{w}}, \bar{b}, \bar{\xi}^{(*)})$ 是问题 (1) 的最优解, $(\tilde{\mathbf{w}}, \tilde{b}, \tilde{\gamma}^{(*)})$ 是问题 (1) 的可行解, 有

$$\begin{aligned} \frac{1}{2} \|\bar{\mathbf{w}}\|^2 + C \sum_{i=1}^l (\bar{\xi}_i + \bar{\xi}_i^*) & \leq \\ \frac{1}{2} \|\tilde{\mathbf{w}}\|^2 + C \sum_{i=1}^l (\tilde{\gamma}_i + \tilde{\gamma}_i^*). \end{aligned} \quad (6)$$

$(\tilde{\mathbf{w}}, \tilde{b}, \tilde{\xi}^{(*)})$ 是问题 (4) 的最优解, $(\bar{\mathbf{w}}, \bar{b}, \bar{\gamma}^{(*)})$ 是问题 (4) 的可行解, 有

$$\begin{aligned} \frac{1}{2} \|\tilde{\mathbf{w}}\|^2 + C \sum_{i=1}^l (\tilde{\xi}_i + \tilde{\xi}_i^*) & \leq \\ \frac{1}{2} \|\bar{\mathbf{w}}\|^2 + C \sum_{i=1}^l (\bar{\gamma}_i + \bar{\gamma}_i^*). \end{aligned} \quad (7)$$

同时, 因为 $\varepsilon_i \geq \varepsilon_0 (i = 1, 2, \dots, l)$, 所以有

$$\bar{\gamma}_i + \bar{\gamma}_i^* \leq \bar{\xi}_i + \bar{\xi}_i^*, i = 1, 2, \dots, l,$$

再结合式 (6) 和 (7), 有

$$\frac{1}{2} \|\tilde{\mathbf{w}}\|^2 + C \sum_{i=1}^l (\tilde{\xi}_i + \tilde{\xi}_i^*) \leq$$

$$\begin{aligned} \frac{1}{2}\|\bar{\mathbf{w}}\|^2 + C \sum_{i=1}^l (\bar{\gamma}_i + \bar{\gamma}_i^*) &\leq \\ \frac{1}{2}\|\bar{\mathbf{w}}\|^2 + C \sum_{i=1}^l (\bar{\xi}_i + \bar{\xi}_i^*) &\leq \\ \frac{1}{2}\|\tilde{\mathbf{w}}\|^2 + C \sum_{i=1}^l (\tilde{\gamma}_i + \tilde{\gamma}_i^*). &\quad (8) \end{aligned}$$

进而有

$$\begin{aligned} 0 \leq C \sum_{i=1}^l (\bar{\xi}_i + \bar{\xi}_i^*) - C \sum_{i=1}^l (\bar{\gamma}_i + \bar{\gamma}_i^*) &\leq \\ C \sum_{i=1}^l (\tilde{\gamma}_i + \tilde{\gamma}_i^*) - C \sum_{i=1}^l (\tilde{\xi}_i + \tilde{\xi}_i^*). &\quad (9) \end{aligned}$$

根据模型(1)和(4)的构造形式,对于 $\forall i \in \{1, 2, \dots, l\}$,若 $|y_i - (\tilde{\mathbf{w}}\mathbf{x}_i + \tilde{b})| < \varepsilon_i$,则 $\tilde{\xi}_i + \tilde{\xi}_i^* = 0$.又因为 $\varepsilon_i \geq \varepsilon_0$,所以 $\tilde{\gamma}_i + \tilde{\gamma}_i^* + \varepsilon_0 \leq \varepsilon_i$,即

$$\tilde{\gamma}_i + \tilde{\gamma}_i^* - (\tilde{\xi}_i + \tilde{\xi}_i^*) \leq \varepsilon_i - \varepsilon_0.$$

若 $y_i - (\tilde{\mathbf{w}}\mathbf{x}_i + \tilde{b}) \geq \varepsilon_i$,则

$$|y_i - (\tilde{\mathbf{w}}\mathbf{x}_i + \tilde{b})| = \varepsilon_0 + \tilde{\gamma}_i + \tilde{\gamma}_i^* = \varepsilon_i + \tilde{\xi}_i + \tilde{\xi}_i^*,$$

即 $\tilde{\gamma}_i + \tilde{\gamma}_i^* - (\tilde{\xi}_i + \tilde{\xi}_i^*) = \varepsilon_i - \varepsilon_0$.综合可得

$$C \sum_{i=1}^l (\tilde{\gamma}_i + \tilde{\gamma}_i^*) - C \sum_{i=1}^l (\tilde{\xi}_i + \tilde{\xi}_i^*) \leq C \sum_{i=1}^l (\varepsilon_i - \varepsilon_0).$$

若

$$\begin{aligned} C \sum_{i=1}^l (\bar{\xi}_i + \bar{\xi}_i^*) - C \sum_{i=1}^l (\bar{\gamma}_i + \bar{\gamma}_i^*) &= \\ C \sum_{i=1}^l (\tilde{\gamma}_i + \tilde{\gamma}_i^*) - C \sum_{i=1}^l (\tilde{\xi}_i + \tilde{\xi}_i^*), & \end{aligned}$$

则有

$$\begin{aligned} \frac{1}{2}\|\bar{\mathbf{w}}\|^2 + C \sum_{i=1}^l (\bar{\xi}_i + \bar{\xi}_i^*) - \frac{1}{2}\|\bar{\mathbf{w}}\|^2 + C \sum_{i=1}^l (\bar{\gamma}_i + \bar{\gamma}_i^*) &= \\ \frac{1}{2}\|\tilde{\mathbf{w}}\|^2 + C \sum_{i=1}^l (\tilde{\gamma}_i + \tilde{\gamma}_i^*) - \frac{1}{2}\|\tilde{\mathbf{w}}\|^2 + C \sum_{i=1}^l (\tilde{\xi}_i + \tilde{\xi}_i^*). & \end{aligned}$$

结合式(8)即得到

$$\begin{aligned} \frac{1}{2}\|\bar{\mathbf{w}}\|^2 + C \sum_{i=1}^l (\bar{\xi}_i + \bar{\xi}_i^*) &= \\ \frac{1}{2}\|\tilde{\mathbf{w}}\|^2 + C \sum_{i=1}^l (\tilde{\gamma}_i + \tilde{\gamma}_i^*). & \quad \square \end{aligned}$$

推论 1 假设 $\varepsilon' = (\varepsilon'_1, \varepsilon'_2, \dots, \varepsilon'_l)$, $\varepsilon'_i \leq \varepsilon_i$, $i = 1, 2, \dots, l$.将参数集 (ε', C) 和 (ε, C) 分别代入模型(4)中得到最优解 $(\hat{\mathbf{w}}, \hat{b}, \hat{\xi}^{(*)})$ 和 $(\tilde{\mathbf{w}}, \tilde{b}, \tilde{\xi}^{(*)})$.将 $(\hat{\mathbf{w}}, \hat{b})$ 和 $(\tilde{\mathbf{w}}, \tilde{b})$ 分别代入以 (ε, C) 和 (ε', C) 为参数集的模型(4)中求得最小的 $\xi^{(*)}$,获得可行解 $(\hat{\mathbf{w}}, \hat{b}, \hat{\gamma}^{(*)})$ 和 $(\tilde{\mathbf{w}}, \tilde{b}, \tilde{\gamma}^{(*)})$.设指标集 $J \subseteq \{1, 2, \dots, l\}$,假设当 $i \notin J$ 时,有 $\varepsilon'_i = \varepsilon_i$,则有

$$0 \leq C \sum_{i \in J} (\hat{\xi}_i + \hat{\xi}_i^*) - C \sum_{i \in J} (\hat{\gamma}_i + \hat{\gamma}_i^*) \leq$$

$$C \sum_{i \in J} (\tilde{\gamma}_i + \tilde{\gamma}_i^*) - C \sum_{i \in J} (\tilde{\xi}_i + \tilde{\xi}_i^*) \leq C \sum_{i \in J} (\varepsilon_i - \varepsilon'_i).$$

证明 根据模型(1)、(4)的相似性和 $\varepsilon'_i \leq \varepsilon_i$ ($i = 1, 2, \dots, l$),由类似引理2的证明可得

$$0 \leq C \sum_{i=1}^l (\hat{\xi}_i + \hat{\xi}_i^*) - C \sum_{i=1}^l (\hat{\gamma}_i + \hat{\gamma}_i^*) \leq$$

$$C \sum_{i=1}^l (\tilde{\gamma}_i + \tilde{\gamma}_i^*) - C \sum_{i=1}^l (\tilde{\xi}_i + \tilde{\xi}_i^*) \leq C \sum_{i=1}^l (\varepsilon_i - \varepsilon'_i).$$

因为对于任意 $i \in \{1, 2, \dots, l\}$,若 $i \notin J$,则 $\varepsilon'_i = \varepsilon_i$,此时有

$$\begin{aligned} 0 &= (\hat{\xi}_i + \hat{\xi}_i^*) - (\hat{\gamma}_i + \hat{\gamma}_i^*) = \\ (\tilde{\gamma}_i + \tilde{\gamma}_i^*) - (\tilde{\xi}_i + \tilde{\xi}_i^*) &= (\varepsilon_i - \varepsilon'_i). \end{aligned}$$

综合可得

$$0 \leq C \sum_{i \in J} (\hat{\xi}_i + \hat{\xi}_i^*) - C \sum_{i \in J} (\hat{\gamma}_i + \hat{\gamma}_i^*) \leq$$

$$C \sum_{i \in J} (\tilde{\gamma}_i + \tilde{\gamma}_i^*) - C \sum_{i \in J} (\tilde{\xi}_i + \tilde{\xi}_i^*) \leq C \sum_{i \in J} (\varepsilon_i - \varepsilon'_i). \quad \square$$

定理 1 假设指标集 $J = \{j | \varepsilon_0 < \varepsilon_j, j = 1, 2, \dots, l\}$,当 $i \notin J$ 时必有 $\varepsilon_i = \varepsilon_0$.若引理2中模型(1)最优解满足 $|y_i - (\bar{\mathbf{w}} \cdot \mathbf{x}_i + \bar{b})| \geq \varepsilon_i$, $i \in J$,则有 $\bar{\mathbf{w}} = \tilde{\mathbf{w}}$.

证明 对于 $i \in J$,因为 $|y_i - (\bar{\mathbf{w}}\mathbf{x}_i + \bar{b})| \geq \varepsilon_i$,所以有

$$C \sum_{i \in J} (\bar{\xi}_i + \bar{\xi}_i^*) - C \sum_{i \in J} (\bar{\gamma}_i + \bar{\gamma}_i^*) = C \sum_{i \in J} (\varepsilon_i - \varepsilon_0).$$

由于 $\varepsilon_i \geq \varepsilon_0$, $i = 1, 2, \dots, l$,对于 $i \notin J$,必有 $\varepsilon_i = \varepsilon_0$,此时有 $(\bar{\xi}_i + \bar{\xi}_i^*) - (\bar{\gamma}_i + \bar{\gamma}_i^*) = \varepsilon_i - \varepsilon_0 = 0$,综合可得

$$C \sum_{i=1}^l (\bar{\xi}_i + \bar{\xi}_i^*) - C \sum_{i=1}^l (\bar{\gamma}_i + \bar{\gamma}_i^*) = C \sum_{i=1}^l (\varepsilon_i - \varepsilon_0). \quad (10)$$

进而有

$$\begin{aligned} C \sum_{i=1}^l (\bar{\xi}_i + \bar{\xi}_i^*) - C \sum_{i=1}^l (\bar{\gamma}_i + \bar{\gamma}_i^*) &= \\ C \sum_{i=1}^l (\tilde{\gamma}_i + \tilde{\gamma}_i^*) - C \sum_{i=1}^l (\tilde{\xi}_i + \tilde{\xi}_i^*). &\quad (11) \end{aligned}$$

再根据引理2可得

$$\frac{1}{2}\|\bar{\mathbf{w}}\|^2 + C \sum_{i \in J} (\bar{\xi}_i + \bar{\xi}_i^*) = \frac{1}{2}\|\tilde{\mathbf{w}}\|^2 + C \sum_{i=1}^l (\tilde{\gamma}_i + \tilde{\gamma}_i^*). \quad (12)$$

所以 $(\tilde{\mathbf{w}}, \tilde{b}, \tilde{\gamma}^{(*)})$ 也是问题(1)的最优解,根据引理1关于解的唯一性得到 $\bar{\mathbf{w}} = \tilde{\mathbf{w}}$. \square

定理 2 在引理2中, $(\bar{\mathbf{w}}, \bar{b})$ 为模型(1)的最优解,假设指标集 $K = \{k | |y_k - (\bar{\mathbf{w}}\mathbf{x}_k + \bar{b})| < \varepsilon_k, k = 1, 2, \dots, l\}$,则有 $C \sum_{i \in K} (\bar{\xi}_i + \bar{\xi}_i^*) \leq C \sum_{i \in K} (\tilde{\gamma}_i + \tilde{\gamma}_i^*)$.

证明 设指标集 $K = \{1, 2, \dots, k\}$,构造 $\varepsilon' = (\varepsilon_0, \dots, \varepsilon_k, \varepsilon_{k+1}, \dots, \varepsilon_l)$,比较 ε' 与 ε 的形式可知, ε'_i

$\leq \varepsilon_i, i = 1, 2, \dots, l$, 当 $i \notin K$ 时, 有 $\varepsilon'_i = \varepsilon_i$, 根据推论 1 可得

$$0 \leq C \sum_{i \in K} (\hat{\xi}_i + \hat{\xi}_i^*) - C \sum_{i \in K} (\hat{\gamma}_i + \hat{\gamma}_i^*) \leq C \sum_{i \in K} (\tilde{\gamma}_i + \tilde{\gamma}_i^*) - C \sum_{i \in K} (\tilde{\xi}_i + \tilde{\xi}_i^*). \quad (13)$$

由于 $\hat{\xi}_i^{(*)}$ 是使模型 (4) 满足约束条件最小的 $\xi_i^{(*)}$, $\hat{\xi}_i$ 与 $\hat{\xi}_i^*$ 中最多只有一个不为零, 且 $\hat{\xi}_i + \hat{\xi}_i^* = \max(|y_i - (\hat{w}x_k + \hat{b})| - \varepsilon'_i, 0)$. 因为当 $i \in K$ 时, 有 $\varepsilon'_i = \varepsilon_0$, 当 $i \notin K$ 时, 有 $|y_i - (\bar{w}x_i + \bar{b})| \geq \varepsilon'_i$, 所以根据定理 1, 将参数集 (ε_0, C) 和 (ε', C) 分别代入模型 (1)、(4) 中可得到除 $\xi^{(*)}$ 之外相同的最优解 (\bar{w}, \bar{b}) . 当 $i \in K$ 时, 有 $\varepsilon'_i = \varepsilon_0$, 所以有

$$\hat{\xi}_i + \hat{\xi}_i^* = \max(|y_i - (\hat{w}x_k + \hat{b})| - \varepsilon'_i, 0) = \max(|y_i - (\bar{w}x_k + \bar{b})| - \varepsilon_0, 0) = \bar{\xi}_i + \bar{\xi}_i^*, i \in K.$$

同理可得 $\hat{\gamma}_i + \hat{\gamma}_i^* = \bar{\gamma}_i + \bar{\gamma}_i^*, i \in K$. 再代入到式 (13) 中即得到

$$0 \leq C \sum_{i \in K} (\bar{\xi}_i + \bar{\xi}_i^*) - C \sum_{i \in K} (\bar{\gamma}_i + \bar{\gamma}_i^*) \leq C \sum_{i \in K} (\tilde{\gamma}_i + \tilde{\gamma}_i^*) - C \sum_{i \in K} (\tilde{\xi}_i + \tilde{\xi}_i^*). \quad (14)$$

因为 $K = \{k | |y_k - (\bar{w}x_k + \bar{b})| < \varepsilon_k, k = 1, 2, \dots, l\}$, 所以有

$$C \sum_{i \in K} (\bar{\gamma}_i + \bar{\gamma}_i^*) = 0 \leq C \sum_{i \in K} (\tilde{\xi}_i + \tilde{\xi}_i^*). \quad (15)$$

结合式 (14) 得到

$$C \sum_{i \in K} (\bar{\xi}_i + \bar{\xi}_i^*) \leq C \sum_{i \in K} (\tilde{\gamma}_i + \tilde{\gamma}_i^*). \quad \square$$

定理 1 表明, 当 ε -SVM 预测曲线在各样本点的误差均大于自适应分配给各样本点的不敏感损失值 ε_i 时, 假设在 ε -SVM 中不存在过拟合的噪声样本点, 最终 $AS\varepsilon$ -SVM 获得与 ε -SVM 相同的回归曲线. 定理 2 表明, 当 ε -SVM 预测曲线在某些样本点的误差小于自适应分配给各样本点的不敏感损失值 ε_i 时, 假设 ε -SVM 在建模时过拟合了这些样本点, 最后 $AS\varepsilon$ -SVM 所获得的回归曲线适当疏远了与这些样本点间距离, 以避免模型因过拟合这些样本点而损失部分泛化性能.

3 仿真实验

为验证 $AS\varepsilon$ -SVM 的有效性, 将 $AS\varepsilon$ -SVM 应用于处理一个数值算例和一个应用实例, 并与 ε -SVM 进行比较, 选取最大预测误差 ($\max(e)$)、最小预测误差 ($\min(e)$)、平均预测误差 ($\text{mean}(e)$) 和预测误差方差 ($\text{var}(e)$) 作为比较两种模型优劣的评价指标. 假设预测模型在某个测试样本点误差为

$$e_i = |y_i - y'_i|. \quad (16)$$

其中: y_i 为第 i 个样本的真实输出值, y'_i 为预测模型在该点的预测值. 有

$$\max(e) = \max_{1 \leq i \leq l_2} (e_i), \quad (17)$$

$$\min(e) = \min_{1 \leq i \leq l_2} (e_i), \quad (18)$$

$$\text{mean}(e) = \frac{1}{l_2} \left(\sum_{i=1}^{l_2} e_i \right), \quad (19)$$

$$\text{var}(e) = \frac{1}{l_2 - 1} \left(\sum_{i=1}^{l_2} (e_i - \text{mean}(e))^2 \right). \quad (20)$$

其中 l_2 为测试样本集的规模.

3.1 数值算例

为了更好地考察 ε -SVM 和 $AS\varepsilon$ -SVM 在噪声环境下的建模能力, 建立一个只包含一个自变量的回归模型

$$y = x + \zeta.$$

ζ 是服从 $N(0, 0.02)$ 的随机数, x 分别取区间 $[1, 2]$ 中平均分布的 36 个节点, 并计算对应的 y 值, 产生 36 组数据 (x, y) 构成样本集. 将样本集分为训练样本集和测试样本集, 取样本集的前 24 个样本作为训练样本集, 余下样本作为测试样本集. 分别将 ε -SVM 和 $AS\varepsilon$ -SVM 作用于训练样本集建立模型, ε -SVM 和 $AS\varepsilon$ -SVM 采用相同的核函数, 有

$$K(x_i, x_j) = x_i x_j.$$

通过网格搜索的方式找到 ε -SVM 的最优参数组合 $\varepsilon = 0.001, C = 38$, $AS\varepsilon$ -SVM 的参数 C 同样取为 38, ε_0 的初始取值为 0.001, 迭代步长 $\varepsilon_s = 0.01$. 为了尽量避免随机因素造成的影响, 将上述实验独立重复执行 60 遍, 每次产生全新的训练样本集和测试样本集. 在 60 次独立实验中, 对两种模型获得的平均预测误差进行比较, 其中 ε -SVM 有 26 次获得了更小的平均预测误差 $\text{mean}(e)$, $AS\varepsilon$ -SVM 在另外的 34 次实验中获得了更小的 $\text{mean}(e)$. 由于篇幅限制, 表 1 仅列出两种模型在前 8 次实验中预测误差的统计性能比较. 由于每次实验中的训练样本集和测试样本集都是全新产生的, 由表 1 可见, ε -SVM 和 $AS\varepsilon$ -SVM 在不同训练样本集和测试样本集上的预测结果会有一个较大的波动; 在第 1 次和第 2 次实验中, 两种模型的预测结果非常接近, 出现这种情况可能是由于 ε -SVM 预测曲线在各样本点的误差都大于自适应分配给各样本点的不敏感损失值 ε_i , 此时 $AS\varepsilon$ -SVM 获得与 ε -SVM 相同的回归曲线, 印证了定理 1 中的结论; 在剩下的 6 次实验中, 两种模型的预测效果互有优劣, 这是由于数据中含有噪声信息. 综合来看, 在第 5 次和第 6 次实验中, ε -SVM 的预测效果更好, 而在剩下的 4 次实验中则是 AS -SVM 的预测精度更高. 这表明

AS ϵ -SVM在面对噪声数据时泛化能力更强.

表1 两种模型在数值算例中的预测误差比较

		max(e)	min(e)	mean(e)	var(e)
第1次	ϵ -SVM	0.0511	0.0011	0.0246	0.2318e-03
	AS ϵ -SVM	0.0510	0.0012	0.0249	0.2295e-03
第2次	ϵ -SVM	0.0330	0.0012	0.0152	0.1048e-03
	AS ϵ -SVM	0.0330	0.0012	0.0151	0.1059e-03
第3次	ϵ -SVM	0.0332	0.0003	0.0145	0.1409e-03
	AS ϵ -SVM	0.0332	0.0003	0.0121	0.1132e-03
第4次	ϵ -SVM	0.0352	0.0046	0.0177	0.0823e-03
	AS ϵ -SVM	0.0331	0.0017	0.0162	0.0725e-03
第5次	ϵ -SVM	0.0499	0.0034	0.0171	0.1857e-03
	AS ϵ -SVM	0.0499	0.0050	0.0201	0.2002e-03
第6次	ϵ -SVM	0.0452	0.0001	0.0186	0.1994e-03
	AS ϵ -SVM	0.0535	0.0001	0.0206	0.2947e-03
第7次	ϵ -SVM	0.0476	0.0002	0.0181	0.1873e-03
	AS ϵ -SVM	0.0442	0.0028	0.0168	0.1514e-03
第8次	ϵ -SVM	0.0408	0.0007	0.0195	0.1924e-03
	AS ϵ -SVM	0.0399	0.0024	0.0141	0.1887e-03

3.2 应用实例

以预测某种型号的汽车销售量为例,考虑影响销售量的因子集包含6个元素,分别是市场竞争程度MC、汽车性能参数PF、汽车油价影响因素OF、汽车价格波动因素PW、居民可支付购买力PP和季节性因素SC.其中MC、PF、OF、PW、PP和SC为6个语言型数据,利用综合评价法获取其数值型数据.选取48个该型号汽车的历史销售数据,加上由综合评价法获取的486个影响因素集的数值型数据组成支持向量机的样本集 (x, y) 进行实例研究.将样本集分为两个部分,前36组数据作为训练样本集,后12组数据作为测试样本集.分别将 ϵ -SVM和AS ϵ -SVM作用于训练样本集建立模型, ϵ -SVM和AS ϵ -SVM采用相同的核函数,有

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|}{\sigma^2}\right).$$

通过网格搜索的方式找到 ϵ -SVM的最优参数组合 $\epsilon = 0.0059$, $C = 2155$, $\sigma = 6.354$, AS ϵ -SVM采用与 ϵ -SVM相同的 C 和 σ , ϵ_0 的初始取值为0.0059,迭代步长 $\epsilon_s = 0.01$.表2列出了 ϵ -SVM与AS ϵ -SVM建立的汽车销量模型的预测误差比较,AS ϵ -SVM明显获得了更好的预测效果, ϵ -SVM仅在min(e)一项稍优于AS ϵ -SVM,其他各项指标均劣于AS ϵ -SVM.

表2 两种模型在应用实例中的预测误差比较

	max(e)	min(e)	mean(e)	var(e)
ϵ -SVM	0.1328	0.0018	0.0576	0.0021
AS ϵ -SVM	0.1245	0.0023	0.0522	0.0018

图1是两种模型的预测误差比较.由图1可见, ϵ -SVM和AS ϵ -SVM所建立的模型在12个预测点中互有优劣,这是受销售数据中含有噪声信息的影响.

在第2、8、10号预测点, ϵ -SVM的预测效果更好,在第1、5号预测点,两种模型的预测效果接近,在剩下的7个点,AS ϵ -SVM所获得的预测值更接近于真实值.所以综合来看,AS ϵ -SVM的整体预测效果优于 ϵ -SVM.

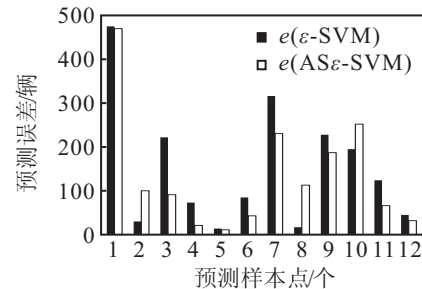


图1 两种模型的预测误差比较

4 结论

产品销售预测时序具有典型的非线性、含噪声等数据特征,为了抑制噪声对预测模型的影响,本文设计了基于自适应分段损失函数的支持向量机模型.通过对距离拟合曲线较远的样本点赋予较大的不敏感损失函数值 ϵ_i 来降低预测模型对样本点的依赖,并从理论上证明了该方法在 ϵ -SVM没有过拟合噪声时将获得与 ϵ -SVM相同的拟合曲线,否则会疏远噪声样本点,以提高模型的泛化性能.仿真结果表明,所提出方法比一般支持向量机方法在预测精度上具有一定的优越性.

参考文献(References)

- [1] Vapnik V N. The nature of statistical learning theory[M]. New York: Springer, 2000: 138-167.
- [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: 224-270. (Deng N Y, Tian Y J. A new method of data mining—Support vector machine[M]. Beijing: Science Press, 2006: 224-270.)
- [3] 陶新民, 郝思媛, 张冬雪, 等. 基于样本特性欠取样的不均衡支持向量机[J]. 控制与决策, 2013, 28(7): 978-984. (Tao X M, Hao S Y, Zhang D X, et al. Support vector machine for unbalanced data based on sample properties under-sampling approaches[J]. Control and Decision, 2013, 28(7): 978-984.)
- [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. (Yao B Z, Yang C Y, Yu B. Dynamic bus travel-time prediction model[J]. J of System Engineering, 2010, 25(3): 365-370.)
- [5] 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. (Zhou H R, Zheng P E, Ren X L. Method for selecting

- parameters of least squares support vector machines and application[J]. J of System Engineering, 2009, 24(2): 248-252.)
- [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.
(Shang Z G, Yan H S. Forecasting product design time based on fuzzy support vector machine[J]. Control and Decision, 2012, 27(4): 531-534.)
- [7] Lee Y J, Hsieh W F, Huang C M. SSVR: A smooth support vector machine for insensitive regression[J]. IEEE Trans on Knowledge and Data Engineering, 2005, 17(5): 678-685.
- [8] Gu J R, Zhu M C, Jiang L G Y. Housing price forecasting based on genetic algorithm and support vector machine[J]. Expert Systems with Applications, 2011, 38(4): 3383-3386.
- [9] Yanga B S, Hwanga W W, Kima D J, et al. Condition classification of small reciprocating compressor for refrigerators using artificial neural networks and support vector machines[J]. Mechanical Systems and Signal Processing, 2005, 19(2): 371-390.
- [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.
(Zhang Y, Su H Y, Chu J. Soft sensor modeling based on fuzzy least squares support vector Machines[J]. Control and Decision, 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.
(Yao X, Yu L A. A fuzzy proximal support vector machine model and its application to credit risk analysis[J]. Systems Engineering-Theory & Practice, 2012, 32(3): 549-554.)
- [12] Lin C F, Wang S D. Training algorithms for fuzzy support vector machines with noisy data[J]. Pattern Recognition Letters, 2004, 25(14): 1647-1656.
- [13] Li D F, Hu W C, Xiong W, et al. Fuzzy relevance vector machine for learning from unbalanced data and noise[J]. Pattern Recognition Letters, 2008, 29(9): 1175-1181.
- [14] 张桂香, 费岚, 杜喆, 等. 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.
(Zhang G X, Fei L, Du Z, et al. New noise-immune fuzzy SVM algorithm for unbalanced data[J]. Computer Engineering and Applications, 2008, 44(16): 142-144.)
- [15] Yang Z Z, Jin L J, Wang M H. Forecasting baltic panamax index with support vector machine[J]. J of Transportation Systems Engineering & Information Technology, 2011, 11(3): 50-57.
- [16] Guo J, Zhou J Z, Qin H, et al. Monthly stream flow forecasting based on improved support vector machine model[J]. Expert Systems with Applications, 2011, 38(10): 13073-13081.
- [17] 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.
(Jiang W, Yi G X, Zeng Q S. Likelihood particle filter based on support vector machines resampling[J]. Control and Decision, 2011, 26(2): 243-247.)
- [18] 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.
(Wu Q, Yan H S. Forecasting method based On support Vector m achine with Gaussian loss function[J]. Computer Integrated Manufacturing System, 2009, 15(2): 306-312.)
- [19] Li H X, Yang J L, Zhang G, et al. Probabilistic support vector machines for classification of noise affected data[J]. Information Sciences, 2013, 221(1): 60-71.
- [20] 吴奇, 严洪森. 基于鲁棒 ν -支持向量机的产品销售预测模型[J]. 计算机集成制造系统, 2009, 15(6): 1081-1087.
(Wu Q, Yan H S. Product sales forecasting model based on robust ν -support vector machine[J]. Computer Integrated Manufacturing System, 2009, 15(6): 1081-1087.)
- [21] 胡根生, 邓飞其. 具分段损失函数的支持向量机回归及在投资决策中的应用[J]. 控制理论与应用, 2006, 23(2): 315-318.
(Hu G S, Deng F Q. Support vector regression with piecewise loss function and its application in investment decision[J]. Control Theory & Applications, 2006, 23(2): 315-318.)
- [22] 胡根生, 邓飞其. 具有多分段损失函数的多输出支持向量机回归[J]. 控制理论与应用, 2007, 24(5): 711-714.
(Hu G S, Deng F Q. Multi-output support vector regression with piecewise loss function[J]. Control Theory & Applications, 2007, 24(5): 711-714.)
- [23] An W J, Liang M G. Fuzzy support vector machine based on within-class scatter for classification problems with outliers or noises[J]. Neurocomputing, 2013, 110(6): 101-110.
- [24] Yang X W, Tan L J, He L F. A robust least squares support vector machine for regression and classification with noise[J]. Neurocomputing, 2014, 140(9): 41-52.
- [25] Hu Q H, Zhang S G, Xie Z X, et al. Noise model based ν -support vector regression with its application to short-term wind speed forecasting[J]. Neural Networks, 2014, 57(9): 1-11.