

基于模糊聚类的稳健支撑向量回归机及火焰图像处理

陈晓峰¹⁾ 王士同¹⁾ 曹苏群^{1),2)} 崔运静³⁾ 马培勇⁴⁾ 仇性启³⁾

¹⁾ (江南大学信息工程学院, 无锡 214122) ²⁾ (淮阴工学院机械系, 淮安 223001) ³⁾ (中国石油大学(华东)机电工程学院, 东营 257061) ⁴⁾ (中国科学技术大学工程科学学院, 合肥 230026)

摘 要 由于离群点会降低支撑向量回归机的性能, 因此为了提高支撑向量回归机的图像处理性能, 提出了一种具有抗离群点性能的模糊稳健支撑向量回归机(FRSVR), 并首先给出了在任意代价函数下支撑向量回归机的求解方法; 然后讨论了构建稳健支撑向量机的代价函数所需的性质, 并在此基础上, 引入了损失代价函数族; 接着根据支撑向量回归机的训练误差, 用模糊 C 均值聚类(FCM)查找离群点; 最后通过迭代的方法实现了模糊稳健支撑向量回归机。为了对火焰图像进行有效处理, 还将 FRSVR 算法应用于乳化油燃烧火焰图像处理, 以去除火焰图像上的离群点。实验结果表明, FRSVR 算法处理图像的性能优于 ε -SVR 算法和自适应 SVR 滤镜(ASBF), 不仅能有效地查找离群点, 而且可去除较大的离群点区域, 还能显著的降低离群点的影响, 并具有良好的泛化性能。

关键词 离群点 支撑向量回归 模糊聚类

中图法分类号: TP391.41 **文献标识码:** A **文章编号:** 1006-8961(2009)03-0463-08

Fuzzy Clustering Based Robust SVR and Flame Image Processing

CHEN Xiao-feng¹⁾, WANG Shi-tong¹⁾, CAO Su-qun^{1),2)}

CUI Yun-jing³⁾, MA Pei-yong⁴⁾, QIU Xing-qi³⁾

¹⁾ (School of Information Technology, Jiangnan University, Wuxi 214122)

²⁾ (School of Mechanical Engineering, Huaiyin Institute of Technology, Huaian 223001)

³⁾ (School of Process Equipment and Control, China University of Petroleum, Dongying 257061)

⁴⁾ (School of Engineering Science, University of Science and Technology of China, Hefei 230026)

Abstract A novel support vector regression method—FRSVR(fuzzy robust support vector regression) based on traditional ε -SVR is proposed in this paper. First, a solution for support vector regression with arbitrary cost function is given. Second the properties which a cost function should have in order to construct a robust support vector regression are discussed. Then a family of cost functions is introduced. In the training procedure of FRSVR, outliers can also be detected in terms of different training error ranges between normal examples and outliers using fuzzy c-means algorithm (FCM). Through iteration, FRSVR is obtained. Since it is based on ε -SVR, various optimization methods for epsilon support vector regression can be used to solve FRSVR. In the experimental part of the paper, FRSVR is applied to process emulsified oil combustion flame images such that outliers therein can be detected and removed, then flame shapes are accordingly segmented. Experimental results show that FRSVR performs better than ε -SVR and ASBF filter in the sense of removing outliers and enhancing the generalization ability.

Keywords outlier, support vector regression(SVR), fuzzy clustering

基金项目:国家高技术研究发展计划(863)项目(2006AA10Z313);国家自然科学基金项目(60773206/F020106,60704047/F030304);国防应用基础研究基金项目(A1420461266);2004 年教育部跨世纪优秀人才支持计划基金项目(NCET-04-0496);2005 年教育部科学研究重点基金项目(105087)

收稿日期:2007-03-23;**改回日期:**2007-09-26

第一作者简介:陈晓峰(1977 ~),男。现为江南大学信息工程学院博士研究生。从事人工智能、模式识别的研究。E-mail: brianchen1977@

1 引言

支撑向量回归机(SVR)是基于统计学习理论的函数拟合与回归的方法^[1],如今研究人员已提出了各种 SVR 算法,如 ε -SVR 算法、 ν -SVR 算法^[2]。支撑向量回归机的拟合性能要受到训练数据中离群点的影响。由于离群点在训练中对应较大的训练误差,拟合曲线的平滑性受到它们的影响远比受到的其他训练样本大,因此导致拟合结果与没有离群点时相比,会有很大变化。因此,在 SVR 训练中,应尽量减小离群点的影响,因为它对构造稳健的支撑向量回归机具有重要的意义。

支撑向量机主要解决模式识别中的分类问题。在构造稳健支撑向量机分类算法方面,去除离群点的影响有以下几种方法:限制样本与其所属类的质心之间的距离^[3]、最小化“留一法”交叉验证的训练误差上界^[4]、用 Hinge 损失函数检测并移除离群点^[5]、自适应非线性地调整正则化参数^[6]等。在构造稳健的支撑向量回归机方面有:加权稳健支撑向量回归方法^[7]、加权最小平方损失支撑向量方法^[8]、基于稳健损失函数的径向基函数网方法^[9]、基于分段代价函数的稳健 SVR^[10]、基于证据框架的稳健支撑向量参数优选方法^[11-12],以及自适应误差损失函数支撑向量机^[13]等方法。在上述方法中,第 1 类方法,如文献[8]、[9]的方法,其是以传统的 SVR 算法为基础来设计变体结构的支撑向量机和支撑向量回归机;第 2 类方法,如文献[7]、[10]、[13]的方法,是先设计新的代价函数,然后通过迭代的方式,根据训练误差来调整与样本对应的正则化参数,由此减小离群点的影响。第 2 类方法的优点是,由于算法的本质仍然是支撑向量算法,因此可以将解决 SVR 问题的优化算法应用到新算法中。

由于支撑向量算法有良好的性能,因此被用于设计图像噪声的过滤算法,以消减图像的噪声,如 SVC (support vector classifier) 滤镜^[14]、ASBF (adaptive SVR based filter) 滤镜^[15]。其中 SVC 滤镜由支撑向量噪声判别器和两相均值滤镜组成,如果像素点被判定为噪声数据,则用两相均值滤镜计算一个新值来替换原值,它的优点是可以同时处理固定值噪声和随机噪声。ASBF 滤镜是用支撑向量回归机检测脉冲噪声,其方法是,先用 SVR 拟合待检测像素点所在的局部图像的像素值,然后用它预

测待检测像素点的像素值,如果待检测像素点的预测值与实际值的差值超过预设阈值,则视该像素点为噪声点。

本文首先给了一种在任意代价函数下支撑向量回归机的求解方法,同时讨论了构建稳健支撑向量机的代价函数所需的性质,然后在此基础上,引入一类稳健代价函数族,实现了模糊稳健支撑向量回归机(FRSVR)。FRSVR 算法是根据支撑向量回归机的训练误差,先用模糊聚类方法判别离群点,然后通过迭代的方法得到稳健的支撑向量回归机。本文用 FRSVR 算法处理乳化油燃烧火焰图像,通过去除图像上的离群点,分割出了较好的乳化油燃烧火焰图像。

2 模糊稳健支撑向量回归机

2.1 ε -SVR 算法

设训练样本集为 $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_l, y_l)\} \subset \mathbf{R}^N \times \mathbf{R}$, 其中 $\mathbf{x}_i$ 为输入向量, y_i 为输出值, N 为输入向量的维数, l 为训练样本数。在 ε -SVR 算法中,先估计函数 $f(\mathbf{x})$, 并使 $|f(\mathbf{x}_i) - y_i| \leq \varepsilon$ ($i = 1, \dots, l$), 且 $f(\mathbf{x})$ 要尽可能平滑。假设 $f(\mathbf{x})$ 为线性函数(可以通过核方法扩展到非线性情况), 即 $f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b$, 其中 $\mathbf{w} \in \mathbf{R}^N, b \in \mathbf{R}, \langle \cdot, \cdot \rangle$ 代表内积, 则求 $f(\mathbf{x})$ 即可表述为如下求极小值问题:

$$\text{minimize } \frac{1}{2} \mathbf{w} \mathbf{w}^T + C \sum_{i=1}^l (\xi_i + \xi_i^*) \quad (1)$$

$$\text{且 } \begin{cases} y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle - b \leq \varepsilon + \xi_i \\ \langle \mathbf{w}, \mathbf{x}_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad (2)$$

其中, C 为正则化参数。 $\xi_i^{(*)}$ (表示 ξ_i 或 ξ_i^* , 下文类同)。可由下述 ε -不敏感损失函数计算:

$$|\xi|_\varepsilon = \begin{cases} 0 & |\xi| \leq \varepsilon \\ |\xi| - \varepsilon & \text{其他} \end{cases} \quad (3)$$

上述问题可以通过 Lagrange 求极值方法转为如下的 Wolfe 对偶的二次规划问题^[1]:

$$\begin{aligned} \text{minimize } & -\frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle \mathbf{x}_i, \mathbf{x}_j \rangle \\ & - \varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) + \sum_{i=1}^l y_i (\alpha_i - \alpha_i^*) \end{aligned} \quad (4)$$

$$\text{且} \quad \begin{cases} \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, C] \end{cases} \quad (5)$$

其中, α_i^* ($i=1, \dots, l$) 为 Lagrange 系数。待求量 $\mathbf{w}, b$ 可表示为

$$\mathbf{w} = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \mathbf{x}_i \quad (6)$$

$$\begin{aligned} & \max \{ -\varepsilon + y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle \\ & \text{且} |\alpha_i < C \text{ 或 } \alpha_i^* > 0\} \leq b \\ & \leq \min \{ -\varepsilon + y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle \mid \alpha_i > 0 \text{ 或 } \alpha_i^* < C \} \end{aligned} \quad (7)$$

2.2 FRSVR 算法

设任意代价函数的一般形式为 $L(x)$, 令 $L'(x)$ 表示 $L(x)$ 的一阶导数。则可根据 $L(x)$ 求解 SVR, 建立的规划问题如下(下面的公式推导, 假定 $f(\mathbf{x})$ 为线性函数, 并可以通过核方法扩展到非线性情况):

$$\text{minimize} \quad \frac{1}{2} \mathbf{w} \mathbf{w}^T + C \sum_{i=1}^l (L_i(\xi_i) + L_i(\xi_i^*)) \quad (8)$$

$$\text{且} \quad \begin{cases} y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle - b \leq \varepsilon_i + \xi_i \\ \langle \mathbf{w}, \mathbf{x}_i \rangle + b - y_i \leq \varepsilon_i^* + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad (9)$$

$$L_i(\xi_i^{(*)}) = L(x_i, y_i, \varepsilon_i^{(*)}, \xi_i^{(*)}) \quad (10)$$

其中, $\xi_i^{(*)}$ (表示 ξ_i 或 ξ_i^*) 为训练误差, $L_i(\xi_i^{(*)})$ 表示与第 i 个样本对应的代价函数。

若用 Lagrange 求极值方法求解上述问题, 则原问题就转化为如下问题:

$$\begin{aligned} L_{\text{Lag}} = & \frac{1}{2} \mathbf{w} \mathbf{w}^T + C \sum_{i=1}^l (L_i(\xi_i) + L_i(\xi_i^*)) - \\ & \sum_{i=1}^l \alpha_i (\langle \mathbf{w}, \mathbf{x}_i \rangle + b - y_i + \varepsilon_i + \xi_i) - \\ & \sum_{i=1}^l \alpha_i^* (y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle - b + \varepsilon_i^* + \xi_i^*) - \\ & \sum_{i=1}^l u_i \xi_i - \sum_{i=1}^l u_i^* \xi_i^* \end{aligned} \quad (11)$$

其中, L_{Lag} 表示 Lagrange 等式, $\alpha_i^{(*)}$ 和 $u_i^{(*)}$ ($i=1, \dots, l$) 为 Lagrange 系数。分别求 L_{Lag} 对 $\mathbf{w}, b, \xi_i^{(*)}$ 的偏导, 并令其为零:

$$\frac{\partial L_{\text{Lag}}}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^l \alpha_i \mathbf{x}_i + \sum_{i=1}^l \alpha_i^* \mathbf{x}_i = 0, \text{ 即}$$

$$\mathbf{w} = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \mathbf{x}_i \quad (12)$$

$$\frac{\partial L_{\text{Lag}}}{\partial b} = - \sum_{i=1}^l \alpha_i + \sum_{i=1}^l \alpha_i^* = 0, \text{ 即}$$

$$\sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \quad (13)$$

$$\frac{\partial L_{\text{Lag}}}{\partial \xi_i^{(*)}} = CL'_i(\xi_i^{(*)}) - \alpha_i^{(*)} - u_i^{(*)} = 0, \text{ 即}$$

$$\alpha_i^{(*)} + u_i^{(*)} = CL'_i(\xi_i^{(*)}) \quad (14)$$

若将式(12)~(14)代入式(11), 则上述规划问题转化为如下对偶规划问题:

$$\begin{aligned} \text{maximize} \quad & -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \mathbf{x}_i \mathbf{x}_j^T + \\ & \sum_{i=1}^l (\alpha_i - \alpha_i^*) y_i + C \sum_{i=1}^l (L_i(\xi_i) - L'_i(\xi_i) \xi_i) + \\ & C \sum_{i=1}^l (L_i(\xi_i^*) - L'_i(\xi_i^*) \xi_i^*) - \\ & \sum_{i=1}^l \alpha_i \varepsilon_i - \sum_{i=1}^l \alpha_i^* \varepsilon_i^* \end{aligned} \quad (15)$$

$$\text{且} \quad \begin{cases} \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \\ 0 \leq \alpha_i^{(*)} \leq CL'_i(\xi_i^{(*)}) \end{cases} \quad (16)$$

构建稳健的支撑向量回归机的关键在于构造合适的代价函数。设代价函数的一般形式为如下所示的 $L(x)$ 的分段函数:

$$L(x) = \begin{cases} 0 & x \leq \varepsilon \\ x - \varepsilon & \varepsilon < x \leq \tau \\ b^{\text{up}}(x) & x > \tau \end{cases} \quad (17)$$

其中, $b^{\text{up}}(x)$ 为误差上界约束函数。

为了减小离群点对回归的影响, 代价函数必须是有界的, 即无论训练误差有多大, 代价函数不能无限的大。此外, 代价函数应具有适当的结构, 以方便求解二次规划对偶问题, 并使正则化参数能根据训练误差 $\xi_i^{(*)}$ ($i=1, \dots, l$) 进行自动的调整。而 $b^{\text{up}}(x)$ 应当具有如下性质: (1) $\lim_{x \rightarrow +\infty} b^{\text{up}}(x)$ 存在, 即有上界, 该性质使代价函数是有界的; (2) $b^{\text{up}}(x)$ 有一阶导数 $(b^{\text{up}}(x))'$, $(b^{\text{up}}(x))'$ 在区间 $[0, +\infty]$ 内有定义, 值域为正数, 且函数单调减, 该性质使在求解对偶规划问题时, $\alpha_i^{(*)}$ 的取值范围随误差的增大而减小。满足上述两个条件的函数有多种, 称之为限上界函数族。在 FRSVR 算法中有如下 3 种函数:

$$b_1^{\text{up}}(x,\sigma) = \text{erf}(x,\sigma) = \frac{2}{\sqrt{\pi}\sigma} \int_0^x e^{-\frac{t^2}{\sigma^2}} dt \quad (18)$$

$$b_2^{\text{up}}(x,\sigma) = \int_0^x \frac{1}{1 + \left(\frac{t}{\sigma}\right)^{2n}} dt \quad (19)$$

$$b_3^{\text{up}}(x,\sigma) = \tanh\left(\frac{x}{\sigma}\right) = \frac{e^{\frac{x}{\sigma}} - e^{-\frac{x}{\sigma}}}{e^{\frac{x}{\sigma}} + e^{-\frac{x}{\sigma}}} \quad (20)$$

上述 3 种函数的一阶导数分别为

$$(b_1^{\text{up}})'(x,\sigma) = \text{erf}'(x,\sigma) = \frac{2}{\sqrt{\pi}\sigma} e^{-\left(\frac{x}{\sigma}\right)^2} \quad (21)$$

$$(b_2^{\text{up}})'(x,\sigma) = \frac{1}{1 + \left(\frac{x}{\sigma}\right)^{2n}} \quad (22)$$

$$(b_3^{\text{up}})'(x,\sigma) = \tanh'(x/\sigma) = \frac{4}{\sigma(e^{\frac{x}{\sigma}} + e^{-\frac{x}{\sigma}})^2} \quad (23)$$

因 3 种限上界函数的图形是相似的,故仅绘出 $\tanh(x,\sigma)$ 和 $\tanh'(x,\sigma)$ 函数,以作为例子说明(如图 1 和图 2 所示)。3 种限上界的函数均满足性质 (1) 和性质 (2),其证明过程较容易,仅需微积分知识,在此省略。

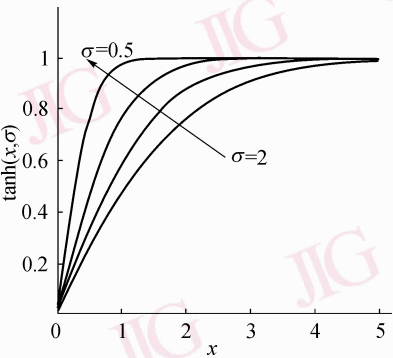


图 1 $\tanh(x,\sigma)$ 函数
Fig. 1 Function $\tanh(x,\sigma)$

在每轮迭代中,FRSVR 算法用基于距离的方法查找离群点。其方法是,在支撑向量回归机的训练中,与离群点对应的训练误差比与正常样本对应的训练误差要大得多。可将训练误差视为样本点与支撑向量回归机之间的距离,并可将它分为以下两类:第 1 类为与正常样本对应的距离,第 2 类为与离群点对应的距离。第 2 类距离的均值要比第 1 类均值

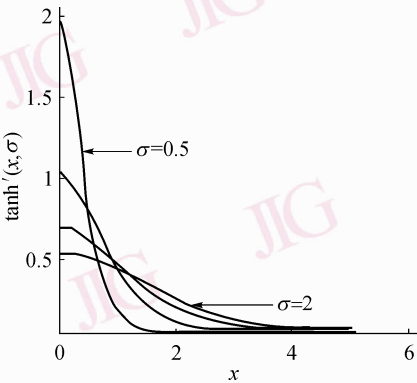


图 2 $\tanh'(x,\sigma)$ 函数
Fig. 2 Function $\tanh'(x,\sigma)$

大得多。图 3 示意了离群点和正常样本之间的区别,在图中,“+”表示训练样本,它们是根据函数 $f(x) = 0.1x^3 + 0.2x^2 + 0.1x + 1$ 产生,并混入 3 个离群点,在用 ε -SVR 算法拟合样本曲线时,核函数为高斯核,参数 C 取 100, σ 取 0.01, ε 取 0.005,拟合结果如图 3 所示,样本中与 3 个离群点对应的曲线部分扭曲明显,离群点和拟合的曲线之间的距离较大,也就是训练误差比较大,由于离群点与拟合曲线的距离不同,因此曲线扭曲的程度是有差别的。在 3 个离群点中,与中间的离群点对应的训练误差相对较小,与其余两个离群点对应的训练误差较大。

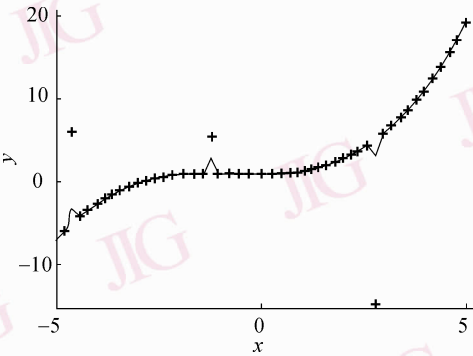


图 3 离群点与正常样本的区别
Fig. 3 Difference between normal examples and outliers

在实际问题中,由于离群点的先验信息,如概率分布,或者偏移正常样本的程度等是未知的,因此,仅知道离群点的训练误差比正常样本的训练误差大这一定性信息,并不能检测出离群点。与训练样本对应的误差有多大时,才可将样本视为

离群点,带有一定的模糊性。由于模糊聚类算法在解决模糊性问题中具有良好的性能,因此,在FRSVR算法中,考虑到查找离群点的模糊性,本文用模糊聚类方法,根据训练误差的大小来查找离群点。

FRSVR算法是用模糊C均值聚类算法(FCM)^[16]将距离分为两类,并将其中均值较大的一类视为与离群点对应的距离,FCM算法如下:

设 $\xi_i^{(*)}$ ($i=1, \dots, l$)为训练误差,则可用FCM算法将 $\xi_i^{(*)}$ 分为两类,设 v 为聚类中心, μ 为隶属度,则FCM算法聚类的目标函数为

$$J_m(\mu, v) = \sum_{k=1}^l \sum_{i=1}^2 (\mu_{i,k})^m d(v_i, \xi_k^{(*)}) \quad (24)$$

$$d(v_i, \xi_k^{(*)}) = \|\xi_k^{(*)} - v_i\|^2 \quad (25)$$

$$\sum_{i=1}^2 \mu_{i,k} = 1 \quad (26)$$

$$\mu_{i,k} \in (0, 1) \quad (27)$$

其中, $d(v_i, \xi_k^{(*)})$ 表示 $\xi_k^{(*)}$ 和聚类中心 v_i 之间的距离, $1 \leq m < \infty$,通常 m 值取2。

聚类中心 v 和隶属度 μ 按如下公式更新:

$$v_i = \frac{\sum_{k=1}^l \mu_{i,k}^m \xi_k^{(*)}}{\sum_{k=1}^l \mu_{i,k}^m} \quad i = 1, 2 \quad (28)$$

$$\mu_{i,k} = \frac{1}{\sum_{j=1}^2 \left(\frac{d(v_i, \xi_k^{(*)})}{d(v_j, \xi_k^{(*)})} \right)^{\frac{2}{m-1}}} \quad i = 1, 2; k = 1, \dots, l \quad (29)$$

将距离聚类为两类后,再引入参数 τ , τ 为第1类距离的所有样本与第1类距离聚类中心的最大距离,其计算公式如下:

$$\tau = \max \|x_{1,i} - v_1\| \quad i = 1, \dots, l_1 \quad (30)$$

其中, l_1 为聚类后的属于第1类距离的样本总数, $x_{1,i}$ 是第1类距离聚类的样本。设 $\xi_i^{(*)}$ 为与训练集的第 i 个样本对应的训练误差,如果有 $\|\xi_i^{(*)} - v_1\| \geq \tau$,则认为该样本为离群点。在式(16)中, $\alpha_i^{(*)}$ 的取值范围可根据 $CL'(\xi_i^{(*)}, \sigma)$ 进行调整。由于 $L'(\xi, \sigma)$ 随着误差变大而迅速减小,而且与离群点对应的训练误差远比非离群点大,因此,与离群点对应的

$\alpha_i^{(*)}$ 的取值范围迅速减小,而且离群点在训练中的影响也随之降低。

FRSVR算法步骤如下:

(1)用常规的 ε -SVR算法训练出初始的支持向量回归机;

(2)选择代价函数的形式;

(3)设置代价函数的参数 σ ,以及 σ 的下降系数参数 s ($s > 1$)和 σ 的下限 σ^b ;

(4)计算 $\sigma = \sigma/s$,若更新后的 $\sigma \geq \sigma^b$,且当前迭代次数小于最大迭代次数,则转到第5步,否则,转到第8步;

(5)计算 $\xi_i^{(*)}$,并计算对应的 $C_i^{(*)} = L'(\xi_i^{(*)}, \sigma)C$;

(6)根据 $\xi_i^{(*)}$,用FCM算法对样本进行分类,并识别离群点;

(7)根据更新后的 $C_i^{(*)}$,训练新的 ε -SVR;

(8)返回第4步;

(9)停止运算,返回 w, b ,离群点。

3 乳化油燃烧火焰图像处理

委内瑞拉的奥里努考地区有3 000亿吨的超重粘油,但这种粘油很难用常规的炼油工艺提取更高使用价值的烃类产品。我国进口了奥里油,并将其乳化作为燃料油^[17]。在炼油过程中即采用以数字图像技术为核心的燃烧控制系统对乳化油燃烧进行自动控制,可见优化燃烧具有重要的现实意义。数字图像技术,即从火焰图像中通过获取锅炉燃烧过程的信息来提取相关特征值的技术。其技术的关键是首先需要对火焰图像进行处理和分割,以获得二值化图像;然后计算火焰的面积、中心、燃烧剧烈程度、长度、直径等参数。其中,火焰的面积、燃烧剧烈程度同乳化油的燃烧效率相关,而火焰的中心、长度和直径则要受到锅炉炉膛的限制,否则会发生喷雾碰壁现象和引起炉膛结焦。由此可见,对以上参数进行测量和控制是保障锅炉高效稳定运行的关键技术。然而在湍流燃烧中,由于火焰轮廓不明显,形状也不规则,火焰不稳定,因此测量火焰的长度和宽度很不容易^[18],并且由于燃烧本身是个剧烈的化学反映过程,在采集图像的过程中,瞬时拍摄的图像会受到烟气、油品雾滴、雾化不均等不利因素的影响,因此会导致火焰图像有不规则的亮点、暗点、带状亮

区等。由于这些区域是燃烧过程中随机出现的,不仅会影响对火焰图像的分割,还会影响火焰的多个性能参数的计算。

图 4 为一幅采集到的典型乳化油火焰燃烧图像,图像上部有条状亮区,同时火焰边缘也较为模糊。图 5 为由专家手工辨识的图 4 火焰图像,其可以认为是火焰的实际图像。图 6 为用 OTSU 算法^[19]对图 4 进行二值化分割后的图像,由该图上可明显看到存在若干不连续的亮区,这些亮区影响到计算火焰的面积、中心和直径。



图 4 火焰图像
Fig. 4 Image of flame



图 5 人工辨识的火焰图像
Fig. 5 Manually identified flame image



图 6 图 4 的二值化分割结果
Fig. 6 Binary thresholding result of Fig. 4

实验时,先用 FRSVR 算法拟合火焰燃烧图像,然后再用 OTSU 算法进行二值化分割来提取火焰图像,并利用 FRSVR 算法的稳健性降低离群区对火焰图像的影响。同时,将 FRSVR 算法的分割结果与 ASBF 滤镜和 ε -SVR 算法拟合的分割结果作对比。FRSVR 算法、 ε -SVR 算法和 ASBF 滤镜的核函数均为 rbf 核,核函数形式如下:

$$K(\boldsymbol{x},\hat{\boldsymbol{x}})=\mathrm{e}^{-r\|\boldsymbol{x}-\hat{\boldsymbol{x}}\|^2}\tag{31}$$

实验参数取 $\sigma=10$,最大迭代次数为 10, $C=40$, $\varepsilon=1$, $\gamma=2$ 。ASBF 滤镜的过滤窗参数取 3。FRSVR 算法中,误差限上界函数为 erf 函数, $\sigma^b=0.5$, $s=2.1$ 。

对火焰图像先分别用 ε -SVR 算法、ASBF 滤镜和 FRSVR 算法进行处理,然后将处理后的图像分别用 OTSU 算法进行分割,其处理结果如图 7~图 12 所示。

图 8 是图 4 用 ε -SVR 算法处理分割后的火焰图像,图 10 是图 4 用 ASBF 滤镜处理分割后的火焰图像,图 12 是图 4 用 FRSVR 算法处理分割后的火焰图像。从上述图中可以看出,图 8、图 10 中均存在



图 7 ε -SVR 算法的拟合结果
Fig. 7 Regression result using ε -SVR



图 8 图 7 的二值化分割结果
Fig. 8 Binary thresholding result of Fig. 7



图 9 ASBF 滤镜的过滤结果
Fig. 9 Filtering result using ASBF



图 10 图 9 的二值化分割结果
Fig. 10 Binary thresholding result of Fig. 9



图 11 FRSVR 算法拟合结果
Fig. 11 Regression result using FRSVR



图 12 图 11 的二值化分割结果
Fig. 12 Binary thresholding result of Fig. 11

若干不连续的亮区,与图 6 相比,亮区的范围虽有所减小,但仍然对火焰区域的面积、中心和直径的计算产生不利影响,而在图 12 中,则亮区不存在,分割后的火焰图像较准确。与图 4、图 7、图 9 相比,图 11 有明显的区别,即火焰上方的条状亮区被弱化,亮度变小,且分散成若干不连续的区块,不仅整幅图像亮度更为均匀,而且对比度有一定的提高。

本文以图 5 为基准图像,将图 6、图 8、图 10、图 12 与图 5 进行对比,以准确率为指标来衡量火焰图像的分割性能。准确率的计算公式^[20]为

$$\eta = 100\% \times \frac{|B_o \cap B_r| + |F_o \cap F_r|}{|B_o| + |F_o|} \quad (32)$$

其中, B_o 和 B_r 分别表示图 5 和待对比的火焰图像的背景像素集,也就是图中的黑色区域, F_o 和 F_r 分别表示图 6 和该火焰图像的前景像素集,即火焰图中的白色区域, $|\bullet|$ 表示像素集的元素个数,脚标“o”和“r”分别表示原图和参考图。

与图 5 对比,图 6、图 8、图 10 和图 12 的分割准确率如表 1 所示:

表 1 各图的分割准确率
Tab. 1 Accuracy rates of different images

图号	单位: %			
	图 6	图 8	图 10	图 12
精度	89. 62	93. 21	90. 36	95. 46

由表 1 的准确率数据可知,在图像分割准确程度上,如果直接用未处理的原图来分割火焰图像,其分割准确率是最低的,为 89. 62%。FRSVR 算法的图像分割准确率是最高的,为 95. 46%,其次为 ε -SVR 算法,分割准确率为 93. 21%。ASBF 滤镜的分割准确率为 90. 36%。对于用 FRSVR 算法处理的乳化油的燃烧火焰图像,通过对处理后图像的分割结果(图 12)和原图的分割结果(图 6)进行对比可见,火焰图像上面的亮区被准确地去除了。FRSVR 算法处理后的火焰图像(图 11)也较原火焰图像(图 4)平缓,亮区部分被明显地暗化,因此,在用 OTSU 算法分割火焰图像时,被暗化的亮区更有可能被归入到背景区域里,这样就提高了火焰图像提取的准确率。 ε -SVR 算法能在一定程度上弱化离群区域,其分割结果(图 8)与原图分割结果(图 6)相比,亮区有一定的减少。而 ASBF 滤镜只适用于处理像素的高斯噪声(即图像中的“椒盐”类型的噪声),并不适合处理图像中的包含较多像素点的亮区和暗区。由于离群点的区域比较大,ASBF 滤镜无法去除其影响,因此分割准确率仅略高于原图直接分割的结果。在计算时间上,由于 FRSVR 算法是以 ε -SVR 算法为基础,通过迭代的方式进行计算,因此,它的时间代价高于 ε -SVR 算法,本试验根据参数设置,FRSVR 算法的计算时间为 ε -SVR 算法的 4 倍。ASBF 滤镜是通过平滑图像窗口的方式去除噪音,其时间代价主要取决于图像窗口的宽度,本试验中,ASBF 滤镜的计算时间高于 ε -SVR 算法,低于 FRSVR 算法。

4 结 论

本文给出了一种任意代价函数下支撑向量回归机的求解方法,同时讨论了构建稳健支撑向量机的代价函数所需的性质,并在此基础上,通过引入一类稳健代价函数族,实现了模糊稳健支撑向量回归机。FRSVR 算法可以很好地改进支撑向量回归算法的稳健性,其不但可以查找样本中的离群点,而且可以将离群点在训练中的影响大幅度降低,得到泛化性能更好的拟合函数。将 FRSVR 算法应用于乳化油燃烧火焰图像处理的实验表明,它能有效去除火焰图像的离群亮区和能分割出准确率高的火焰图像。迭代方式的稳健支撑向量机的缺点是训练时间较

长,因为其需要多次求解二次规划问题。这在训练样本数较少的情况下可以承受,但当处理较大规律数据集时,会消耗较多的训练时间。为了提高FRSVR算法的可用性,进一步的工作将是研究如何提高其求解速度。

参考文献 (References)

- 1 Smola A J, Schölkopf B. A tutorial on support vector regression [J]. *Statistics and Computing*, 2004, **14**(3): 199-222.
- 2 Schölkopf B, Smola A J, Williamson R C, *et al.* New support vector algorithm [J]. *Neural Computation*, 2000, **12**(12): 1207-1245.
- 3 Song Q, Hu W, Xie W. Robust support vector machine with bullet hole image classification [J]. *IEEE Transactions on Systems, Man and Cybernetics*, 2002, **32**(4): 440-448.
- 4 Weston J, Herbrich R. Adaptive margin support vector machines [A]. In: Smola A J, Bartlett P, Schölkopf B, *et al.* eds; *Advances in Large Margin Classifiers* [C], Cambridge, MA, USA: MIT Press, 2000: 281-295.
- 5 Xu Lin-li, Crammer K, Schuurmans D. Robust support vector machine training via convex outlier ablation [A]. In: *Proceedings of the 21st National Conference on Artificial Intelligence* [C], Boston, Massachusetts, USA, 2006: 536-546.
- 6 Zhan Yi-qiang, Shen Ding-gang. An adaptive error penalization method for training an efficient and generalized SVM [J]. *Pattern Recognition*, 2006, **39**(3): 342-350.
- 7 Zhang Jiang-she, Guo Gao. Reweighted robust support vector regression method [J]. *Chinese Journal of Computer*, 2005, **28**(7): 1171-1177. [张讲社, 郭高. 加权稳健支撑向量回归方法[J]. *计算机学报*, 2005, **28**(7): 1171-1177.]
- 8 Suykens J A K, De Brahanter J, Lukas L, *et al.* Weighted least squares support vector machine: robustness and sparse approximation [J]. *Neurocomputing*, 2002, **48**(1-4): 85-105.
- 9 Chuang C C, Su F F, Jeng J T, *et al.* Robust support regression networks for function approximation with outliers [J]. *IEEE Transactions on Neural Networks*, 2002, **13**(6): 1322-1330.
- 10 Zhan Yong, Cheng Hao-zhong. A robust support vector algorithm for harmonic and interharmonic analysis of electric power system [J]. *Electric Power Systems Research*, 2005, **73**(3): 393-400.
- 11 Wang Shi-tong, Zhu Jia-gang, Chung Fu-Lai, *et al.* Experimental study on parameter choices in norm-r support vector regression machines with noisy input [J]. *Soft Computing*, 2006, **10**(3): 219-223.
- 12 Wang Shi-tong, Zhu Jia-gang, Chung Fu-lai, *et al.* Theoretically optimal parameter choices for support vector regression machines with noisy input [J]. *Soft Computing*, 2005, **9**(10): 732-741.
- 13 Chen Xiao-feng, Wang Shi-tong, Cao Su-qun. SVR with adaptive error penalization [J]. *Journal of Electronics & Information Technology*, 2008, **30**(2): 367-370. [陈晓峰, 王士同, 曹苏群. 自适应误差惩罚支撑向量回归机[J]. *电子与信息学报*, 2008, **30**(2): 367-370.]
- 14 Lin Tau-chao, Yu Pao-ta. Adaptive two-pass median filter on support vector machines for image restoration [J]. *Neural Computation*, 2004, **16**(2): 333-354.
- 15 Zhu Jia-gang, Wang Shi-tong, Wu Xi-sheng, *et al.* A novel adaptive SVR based filter ASBF for image restoration [J]. *Soft Computing*, 2006, **10**(8): 665-672.
- 16 Dunn J C. A fuzzy relative of the ISODATA process and its use in detecting compact, well-separated clusters [J]. *Journal of Cybernetics*, **3**(3): 32-57, 1973.
- 17 Duan Xi-li, Wang Zong-ming, Wang Li-juan, *et al.* Experimental study of the turbulent/diffusion flame dimensions of a multi-jet gas burner [J]. *Journal of Engineering for Thermal Energy and Power*, 2004, **19**(2): 144-147. [段希利, 王宗明, 王丽娟等. 多股射流瓦斯燃烧器湍流扩散火焰尺度的实验研究[J]. *热能动力工程*, 2004, **19**(2): 144-147.]
- 18 Wang Zong-ming, Qiu Xing-qi. Design and experiment on the Orinoco emulsified oil burner [J]. *Petrol-chemical Equipment*, 2002, **31**(1): 23-25. [王宗明, 仇性启. 奥里乳化油燃烧器设计与实验[J]. *石油化工设备*, 2002, **31**(1): 23-25.]
- 19 Otsu N. A threshold selection method from gray-level histograms [J]. *IEEE Transactions on Systems, Man and Cybernetics*, 1979, **9**(1): 62-66.
- 20 Tizhoosh H R. Image thresholding using type II fuzzy sets [J]. *Pattern Recognition*, 2005, **38**(12): 2363-2372.