

融合概率分布和单调性的支持向量回归算法

张青, 颜学峰[†]

(华东理工大学 化工过程先进控制和优化技术教育部重点实验室 上海 200237)

摘要: 传统支持向量回归是单纯基于样本数据的输入输出值建模, 仅使用样本数据信息, 未充分利用其他已知信息, 模型泛化能力不强. 为了进一步提高其性能, 提出一种融合概率分布和单调性先验知识的支持向量回归算法. 首先将对偶二次规划问题简化为线性规划问题, 在求解时, 加入与拉格朗日乘子相关的单调性约束条件; 通过粒子群算法优化惩罚参数和核参数, 优化目标包括四阶矩估计表示的输出样本概率分布特性. 实验结果表明, 融合这两部分信息的模型, 能使预测值较好地满足训练样本隐含的概率分布特性及已知的单调性, 既提高了预测精度, 又增加了模型的可解释性.

关键词: 支持向量回归; 概率分布; 单调性; 粒子群优化

中图分类号: TP273 文献标识码: A

Improved support vector regression algorithm combining with probability distribution and monotone property

ZHANG Qing, YAN Xue-feng[†]

(Key Laboratory of Advanced Control and Optimization for Chemical Processes of Ministry of Education, East China University of Science and Technology, Shanghai 200237, China)

Abstract: The traditional support vector regression (SVR) is only based on data information, and a great deal of prior knowledge is neglected. In order to improve its performance, a new SVR algorithm combining with probability distribution and monotone property is proposed. Firstly, the dual quadratic programming problem is simplified as a linear one. Secondly, the monotonic constraints associated with Lagrange multiplier are added. Thirdly, the particle swarm optimization (PSO) is employed to optimize the penalty and kernel parameters. And the fitness function of PSO is the deviation of the probability distribution estimated by four-order moments. The experiment results show that the performance of the proposed SVR model is improved and the developed model satisfies probability distribution and monotone property.

Key words: support vector regression; probability distribution; monotone property; particle swarm optimization

1 引言(Introduction)

支持向量回归(support vector regression, SVR)是 Vapnik 在统计学习理论的基础上提出的一种机器学习算法^[1], 标准支持向量回归算法的基本思路是通过有限个样本数据, 建立起反映输出和输入之间的函数关系. 它通过结构风险最小化原理来提高泛化能力, 较好地解决了小样本、非线性、高维度等实际问题. 从本质上说, SVR 属于黑箱建模, 在学习过程中, 没有考虑已知的先验知识. 在机器学习领域, 有效利用先验知识有助于提高学习算法的泛化能力. 文献[2]考虑了几种不同类型的先验知识, 包括已知函数在特殊点的值、导数值, 在给定区域的界以及在给定区域的任何导数信息. 文献[3]将梯度信息作为第2类变量融入到

核矩阵中直接构建优化模型. 通常, 先验知识会以约束的形式结合到支持向量回归算法中^[2,4-5]. 对于一个化工系统, 由于宏观和微观上的复杂性, 虽然它的全部行为还是无法准确地用数学语言去描述, 但根据系统的能量和质量守恒定律, 及动态平衡原则等可以得出部分的信息. 比如某个反应过程, 系统的熵永远是只增不减, 反应速率随着反应温度的升高保持单调递增, 而单调性正是经常可以获得的化工系统特性. 文献[6]中采用单调性改进神经网络, 文献[7]将单调性约束支持向量分类算法, 都使模型性能提升. 因此本文首先考虑单调性先验知识, 采用不等式约束的形式将单调性有效地融入 SVR 建模. 随后, 考虑到工业过程采集的样本数据中, 只有一小部分有对应输入输出

收稿日期: 2015-11-09; 录用日期: 2017-03-15.

[†]通信作者. E-mail: xfyang@ecust.edu.cn; Tel.: +86 21-64253254.

本文责任编辑: 李少远.

国家自然科学基金项目(21176073), 国家“973”计划(2013CB733605)资助.

Supported by National Natural Science Foundation of China (21176073) and National “973” Plan (2013CB733605).

对的样本数据. 这是由于工业过程分析测试比较耗时, 主要分析指标通常间隔若干小时测试一次, 造成大量输入数据没有对应的输出数据. 但这些缺失对应输出数据的样本也包含大量有用信息, 主要体现在概率分布(probability distribution, PD)信息上. 这些信息在现有SVR建模中很少被利用. 文献[8]假设了风速的季节性分布规律, 并用直方图估计出风速概率, 给样本扩充一个概率维度, 得到较好的时间序列预测结果. 本文从这两个角度出发, 分别在模型推导阶段以差分方程约束的形式, 以及在参数优化阶段以调整适应度函数的方式, 首次同时将单调性和样本概率分布一致性这两种先验知识融合到SVR中, 并结合线性规划方法给出了一种新的融合概率分布和单调性的线性规划支持向量机回归方法, 以下简称MonLPSVR(w/PD).

2 支持向量回归(Support vector regression)

2.1 标准支持向量回归(Standard SVR)

给定一个有标记的训练样本集

$$D = \{(\mathbf{x}_i, y_i) \mid \mathbf{x}_i \in \mathbb{R}^d, y_i \in \mathbb{R}\},$$

其中: $\mathbf{x}_i \in \mathbb{R}^d$ 为 d 维输入变量的值, $y_i \in \mathbb{R}$ 为相应的输出值, n 为训练样本数. 定义 ε 不敏感损失函数为

$$|y - f(\mathbf{x})|_\varepsilon = \begin{cases} 0, & |y - f(\mathbf{x})| \leq \varepsilon, \\ y - f(\mathbf{x}) - \varepsilon, & \text{其他,} \end{cases} \quad (1)$$

其中: $f(\mathbf{x})$ 为通过对样本集的学习而构造的回归估计函数, y 为与 \mathbf{x} 对应的目标输出值, $\varepsilon > 0$ 为误差参数, 规定了估计函数在样本数据上的误差要求. 在采用支持向量机研究非线性样本集时, 通过非线性函数 $\varphi(\mathbf{x})$ 将训练集数据映射到一个具有高维线性的特征空间, 在这个维数可能为无穷大的特征空间中构造估计函数. 假定 $f(\mathbf{x})$ 为如下形式:

$$f(\mathbf{x}) = \boldsymbol{\omega}^T \cdot \varphi(\mathbf{x}) + b, \quad (2)$$

其中: $\boldsymbol{\omega}^T \cdot \varphi(\mathbf{x})$ 表示向量 $\boldsymbol{\omega}^T$ 与 $\varphi(\mathbf{x})$ 的内积, $\boldsymbol{\omega}^T$ 的维数为特征空间维数, $b \in \mathbb{R}$.

由对偶问题的最优性条件得

$$\boldsymbol{\omega}^T = \sum_{i=1}^N \alpha_i \cdot \varphi^T(\mathbf{x}_i),$$

再引入核函数:

$$k(\mathbf{x}, \mathbf{x}_i) = \langle \varphi(\mathbf{x}_i), \varphi(\mathbf{x}) \rangle = \varphi^T(\mathbf{x}_i) \cdot \varphi(\mathbf{x}). \quad (3)$$

则式(2)可改写为最终回归模型

$$f(\mathbf{x}) = \sum_{i=1}^N \alpha_i \cdot k(\mathbf{x}, \mathbf{x}_i) + b. \quad (4)$$

核函数的引入解决了非线性、高维数据的回归问题, 通过映射后特征内积的等价, 避免了直接寻找非

线性映射的困难. 目前针对支持向量机核函数的选择及构造的研究比较多. 文献[9]构造了一个新的混合核函数用于精矿品位预测, 同时提高了算法的学习能力与泛化能力. 但是目前最通用的核函数还是高斯核.

为方便比较研究, 并突出融合先验知识的优势, 本文采用高斯核函数

$$k_{\text{rbf}}(\mathbf{x}, \mathbf{x}_i) = \exp[-\lambda |\mathbf{x} - \mathbf{x}_i|^2], \quad (5)$$

其中 λ 是核半径的倒数, 以下称为核尺度.

设 $\mathbf{A} \in \mathbb{R}^{N \times d}$ 是样本输入集, $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_N]^T$,

$$K(\mathbf{x}, \mathbf{A}) = [k(\mathbf{x}, \mathbf{x}_1), \dots, k(\mathbf{x}, \mathbf{x}_N)]^T;$$

用矩阵形式描述式(4)得

$$f(\mathbf{x}) = \boldsymbol{\alpha}^T \cdot K(\mathbf{x}, \mathbf{A}) + b. \quad (6)$$

根据统计学理论^[1], 回归函数的估计转换成如下最优化问题:

$$\min_{\boldsymbol{\alpha}, b} \frac{1}{2} \|\boldsymbol{\alpha}\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*), \quad (7)$$

$$\text{s.t.} \quad \begin{cases} y_i - \boldsymbol{\alpha}^T K(\mathbf{x}, \mathbf{A}) + b \leq \varepsilon + \xi_i, \\ \boldsymbol{\alpha}^T K(\mathbf{x}, \mathbf{A}) + b - y_i \leq \varepsilon + \xi_i^*, \\ \xi_i, \xi_i^* \geq 0, \quad i = 1, \dots, N, \end{cases}$$

式中: C 表示训练误差最小化与回归函数泛化能力最大化的折衷常数, 向量 $\boldsymbol{\alpha}$ 决定了回归模型的复杂度, 向量 $\boldsymbol{\alpha}$ 中的不等于零的单元 α_i 所对应的样本输入 \mathbf{x}_i 即为支持向量.

2.2 线性规划支持向量回归(Linear-process SVR)

考虑到之后的SVR改进工作需要增加新的约束条件, 并要通过反复迭代来搜寻最优参数, 如果维持使用二次约束下的二次规划算法, 其求解过程会比较复杂. 在这里采用线性规划支持向量回归(LPSVR)来简化标准的二次规划支持向量回归(QPSVR). 文献[10]将LPSVR与多核支持向量结合, 取得较好结果. 研究表明: LPSVR比QPSVR具有更好的模型稀疏性以及能够使用更一般性的核函数^[11].

如果使用LPSVR, 那么对偶的最优化算式就变为以下形式:

$$\min_{\boldsymbol{\alpha}, b} \|\boldsymbol{\alpha}\|_1 + C \sum_{i=1}^N (\xi_i), \quad (8)$$

$$\text{s.t.} \quad \begin{cases} y_i - \boldsymbol{\alpha}^T K(\mathbf{x}, \mathbf{A}) - b \leq \varepsilon + \xi_i, \\ \boldsymbol{\alpha}^T K(\mathbf{x}, \mathbf{A}) + b - y_i \leq \varepsilon + \xi_i, \\ \xi_i \geq 0, \quad i = 1, \dots, N. \end{cases}$$

在求解上述算式时, 一范数 $\|\boldsymbol{\alpha}\|_1 = |\alpha_1| + \dots + |\alpha_n|$, 需要把 α_i 和 $|\alpha_i|$ 分解为

$$\alpha_i = \alpha_i^+ - \alpha_i^-, \quad |\alpha_i| = \alpha_i^+ + \alpha_i^-,$$

式中 $\alpha_i^+, \alpha_i^- \geq 0$. 对于一个已知 α_i , 上述两个方程只

有一个满足, 而且两个变量不能同时大于零, 即需要满足 $\alpha_i^+ \cdot \alpha_i^- = 0$. 用线性规划优化算法求解, 得到回归函数

$$f(\mathbf{x}) = \sum_{i=1}^N (\alpha_i^+ - \alpha_i^-) k(\mathbf{x}, \mathbf{x}_i) + b. \quad (9)$$

3 先验知识支持向量回归(Priori knowledge SVR)

3.1 单调性先验知识描述及融合(Description and combination of monotonic prior knowledge)

如果一个方程可导, 单调递增意味着非负的导数. 如果不假设方程可导, 单调递增可描述为非负的一阶差分. 对于一维单变量回归, 用矩阵形式的核回归函数 $f(x) = \boldsymbol{\alpha}^T K(x, \mathbf{A}) + b$ 直接来构造一阶非负差分方程可得^[12](以单调递减为例)

$$\begin{aligned} f(\mathbf{v}_m + \Delta h) - f(\mathbf{v}_m) &\leq \mathbf{0} \Rightarrow \\ \boldsymbol{\alpha}^T [K(\mathbf{v}_m + \Delta h, \mathbf{A}) - K(\mathbf{v}_m, \mathbf{A})] &\leq \mathbf{0}, \end{aligned} \quad (10)$$

其中: f 为核回归函数, $\mathbf{v}_m, \mathbf{v}_m + \Delta h$ 为含有 m 个虚拟离散变量点的向量, Δh 是相对于点 \mathbf{v}_m 增加的步长, $\mathbf{0}$ 为 m 维零向量. 由此得到保证单调性的 m 个不等式约束条件.

所以, 优化问题(7)修正如下:

$$\begin{aligned} \min_{\boldsymbol{\alpha}, b} \|\boldsymbol{\alpha}\|_1 + C \sum_{i=1}^N (\xi_i), \quad (11) \\ \text{s.t.} \quad \begin{cases} y_i - \boldsymbol{\alpha}^T K(x, \mathbf{A}) + b \leq \varepsilon + \xi_i, \\ \boldsymbol{\alpha}^T K(x, \mathbf{A}) + b - y_i \leq \varepsilon + \xi_i, \\ \boldsymbol{\alpha}^T [K(\mathbf{v}_m + \Delta h, \mathbf{A}) - K(\mathbf{v}_m, \mathbf{A})] \leq \mathbf{0}, \\ \xi_i \geq 0, i = 1, \dots, N. \end{cases} \end{aligned}$$

3.2 样本概率分布先验知识融合 (Sample PD knowledge integration)

3.2.1 样本概率分布差异度量 (Sample PD difference metric)

已知工业样本有一部分有对应输入输出数据, 这在实验中就相当于测试样本集, 但是还有大部分工业数据是只有输入, 没有对应分析结果的数据. 若单以这部分输入通过模型进行预测, 其获得输出结果分布应该与已有输出样本分布相一致. 这部分信息需要进一步提取, 加以有效利用.

文献[13]针对领域适应学习任务, 从缩小领域间数据分布距离的角度, 提出一种新颖的核分布一致局部领域适应学习机, 达到很好的分类效果. 本文的回归建模是在相同领域, 并且考虑到之前得到的参考模型的预测精度已经较高, 值得信任, 所以可以认为真实分布与参考分布的偏差不会太大. 从而通过设定两

个分布的矩差, 可以近似刻画真实分布与参考分布的偏离程度.

本文借助文献[14]中的4阶概率分布差异矩度量方法:

定义 1 对于随机变量 x 与 y , 若

$E(x) = \mu_x, D(x) = \sigma_x^2, E(y) = \mu_y, D(y) = \sigma_y^2$, x 的偏度和峰度分别为 ν_3^x 和 ν_4^x , y 的偏度和峰度分别为 ν_3^y 和 ν_4^y , 称

$$\begin{aligned} \text{MD}_4(F_y \| F_x) &= \frac{(\mu_x - \mu_y)^2}{\sigma_x^2} + \frac{1}{2} \left(1 - \frac{\sigma_y^2}{\sigma_x^2}\right)^2 + \\ &\quad \frac{1}{6} (\nu_3^x - \nu_3^y)^2 + \frac{1}{24} (\nu_4^x - \nu_4^y)^2 \end{aligned} \quad (12)$$

为随机变量 x 与 y 的4阶概率分布差异矩度量, 记作 $\text{MD}_4(F_y \| F_x)$.

接下来的问题是如何应用概率分布的先验知识来改进支持向量回归, 本文主要是将它融入基于智能优化算法的模型参数的选择上, 包括惩罚系数 C 和核尺度 λ , 因为这两个参数对SVR模型的性能会产生很大的影响. 已有大量智能优化算法及其改进工作^[15]力求提高参数寻优的收敛速度与收敛精度, 但它们都是从解空间的结构分析和搜索策略出发, 没有对适应度函数的改进. 本文选取最常用的粒子群优化, 用样本的概率分布特征改进其适应度函数, 并进行改进前后的对比.

3.2.2 粒子群优化(Particle swarm optimization)

粒子群优化^[16](Particle swarm optimization, PSO) 是于1995年由Eberhart和Kennedy提出的一种智能优化算法. PSO算法是通过不断调整粒子的位置来搜索解的. 假设 D 维搜索空间中, 由 n 个粒子构成种群 $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$, 则第 i 个粒子的当前位置为 $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{id})$, 粒子当前的飞行速度为 $\mathbf{v}_i = (v_{i1}, v_{i2}, \dots, v_{id})$, 粒子 i 所经过的最好位置为 $\mathbf{p}_i = (p_{i1}, p_{i2}, \dots, p_{id})$, 所有粒子经过的最好位置为 $\mathbf{p}_g = (p_{g1}, p_{g2}, \dots, p_{gd})$. 则第 i 个粒子在 $t + 1$ 时刻为

$$\mathbf{v}_i^{t+1} = \mathbf{v}_i^t + c_1 r_1 (\mathbf{p}_i^t - \mathbf{x}_i^t) + c_2 r_2 (\mathbf{p}_g^t - \mathbf{x}_i^t), \quad (13)$$

$$\mathbf{x}_i^{t+1} = \mathbf{x}_i^t + \mathbf{v}_i^{t+1}, \quad (14)$$

其中: $1 \leq d \leq D; 1 \leq i \leq n; r_1$ 和 r_2 为均匀分布在 $(0, 1)$ 区间上的随机数; c_1 和 c_2 称为学习因子, 通常取 $c_1 = c_2 = 2$.

3.2.3 基于概率分布特性的适应度函数设计 (Fitness function based on PD)

设从工业过程采集包含输入输出的 N_1 个样本数据 $(\mathbf{x}_i, \mathbf{y}_i) (i = 1, 2, \dots, N_1)$ 称为完整样本. 采集仅包含输入没有对应输出的 N_2 个样本数据 $\mathbf{x}_i^s (i = 1, 2, \dots,$

N_2)称为缺失样本. 对完整数据集再采用交叉验证的分割思想, 随机分出训练集(training set)和验证集(validation set). 分别记为:

$$(\mathbf{x}_i^t, y_i^t), i = 1, 2, \dots, N_1^t; (\mathbf{x}_i^v, y_i^v), i = 1, 2, \dots, N_1^v.$$

在应用PSO选择最优参数对 (C, λ) 时, 通常是先取每次迭代时的一组值, 用训练样本 (\mathbf{x}_i^t, y_i^t) 代入, 训练得到一个回归模型, 再将验证集的输入 \mathbf{x}_i^v 代入回归模型得到预测值 $f(\mathbf{x}_i^v)$, 把它们与验证集真实输出值 y_i^v 的均方根误差(RMSE)作为粒子的适应值. 然而, 本文在这里把缺失样本 \mathbf{x}_i^s 也代入回归模型得到预测输出 $f(\mathbf{x}_i^s)$, 这部分输出和训练样本真实输出 y_i^t 的概率分布应该保持一致, 把这个隐含条件加入粒子的适应值, 作为最优参数的评判标准. 这样就多了一重回归模型预测精度的保障.

结合定义1给出新的适应度函数为

$$\text{Fitness} = \text{RMSE} + \text{MD}_4 = \sqrt{\frac{1}{N_1^v} \sum_{i=1}^{N_1^v} (y_i^v - f(\mathbf{x}_i^v))^2 + \text{MD}_4(y^t \| f(\mathbf{x}^s))}, \quad (15)$$

其中: N_1^v 为验证集中的样本数目, $f(\mathbf{x}_i^v)$ 为验证集样本的预测值, y_i^v 为验证集样本的实际值, y^t 为测试集输出值, \mathbf{x}^s 为缺失样本输入值. 由上式可知, 粒子适应值越小目标参数越优.

3.2.4 算法步骤(Algorithm steps)

1) 初始化粒子群 (C, λ) , 确定群体规模、算法最大迭代次数、算法权重因子 ω (以下实验中均设置群体规模为24, 最大迭代次数为2000, 权重因子为0.5), 以及粒子群飞行速度范围 $[-v_{\max}, v_{\max}]$, 给定SVR损失函数的参数 ε ;

2) 采用结合单调性的线性支持向量回归(参数即各粒子的坐标位置值), 对训练样本进行学习并建立各自对应的回归模型.

3) 用验证集代入得到的支持向量回归模型, 并按照式(15)计算粒子的适应度值, 根据所获得的粒子适应度函数值调整粒子的个体最优位置 \mathbf{p}_i 和全局最优位置 \mathbf{p}_g ;

4) 按照式(13)–(14)更新粒子群算法, 获得一组新的组合参数;

5) 判断是否满足循环停止条件, 当适应值小于 ξ (默认 $1e-25$)或达到最大迭代次数时则结束, 输出结果, 否则转2);

6) 此时的 (C, λ) 值, 就是能使回归模型既满足单调性先验知识, 又融合了概率分布特征的最优组合参数.

4 仿真实验(Experiment)

为了验证上述算法的有效性, 设计或使用下面的案例, 包括单调递增的S函数, 单调递减的高原函数, 和一个实际工业过程原油蒸馏实沸点的离散采样数据, 以及6维的燃烧副反应数据. 以下实验中损失函数(1)的误差容许宽度都取为 $\varepsilon = 0.001$;

1) S函数

$$f(x) = \frac{1}{1 + e^{-x}}, x \in (-\infty, \infty). \quad (16)$$

2) 高原函数

$$f(x) = \frac{3}{1 + e^{(10(x-1))}} + \frac{2}{1 + e^{(5(x-4))}}, 0 \leq x \leq 5. \quad (17)$$

为模拟工业过程采样, 以上两个函数的采样方法均遵照3.2.3节, 在定义域中随机均匀采取完整样本200对, 缺失样本200对. 则 $N_1 = N_2 = 200$, 且 x_i, x_i^s 均满足区间在定义域上的随机均匀分布, 记为 $\mathbf{x}, \mathbf{x}^s \sim U(0, 5)$ 或 $U(-10, 10)$.

为了更好地区别改进SVR算法与标准SVR算法的预测效果, 也同时仿真工业过程的噪声, 本文给完整样本中分割出的训练样本的函数值 y_i^t 加入额外的高斯噪声, 即 $y_i^{t*} = y_i^t + e_i$, 验证集则无需加噪声. 其中 e_i 满足均值为0, 方差为 δ^2 的正态分布, 分别表示为: $\mathbf{e} \sim N(0, \delta^2)$. 还需指出的是, 在加入单调约束时, 函数(16)是在 $(-10, 10)$ 区间上每隔0.1取一个点, 由取出的201个点组成200个约束不等式; 函数(17)是在 $(0, 5)$ 区间上每隔0.05取一个点, 组成200个不等式约束.

经过3.2.4的PSO算法步骤, 待回归函数式(16)的SVR建模过程中的两个最优参数由原来的 $C = 8079, \lambda = 1.6$ 修正为 $C = 7432, \lambda = 1.4$. 待回归函数式(17)的SVR的两个最优参数由原来的 $C = 4630, \lambda = 3.57$ 修正为 $C = 5637, \lambda = 3.73$. 不同算法的效果比较如图1(a)和图1(b).

3) 原油蒸馏实沸点.

原油蒸馏过程^[17]中, 不同组分会对应不同的沸点. 常以馏出组分的质量百分比 p (这里直接转为小数)横坐标, 以馏出温度 T 为纵坐标. 本文采用的样本为30组实测值. 随机分出10组训练集和10组缺失样本集, 和10组验证集. 由于这些值里已包含测量误差, 无需重复添加, 且在原油沸点数据中, 已明确非单调区间位于0.3至0.5, 所以在此区间每隔0.01取21个点组成20组不等式约束. 最终建模过程最优参数由原来的 $C = 44600, \lambda = 20$ 修正为 $C = 500000, \lambda = 28.57$. 不同算法的效果比较如图1(c).

最后实验1, 2得到的模型误差由10折交叉验证后的MSE形式给出, 如式(18), 而实验3的值域范围较大, 所以由留一法验证后的平均相对误差的绝对值给

出, 如式(19).

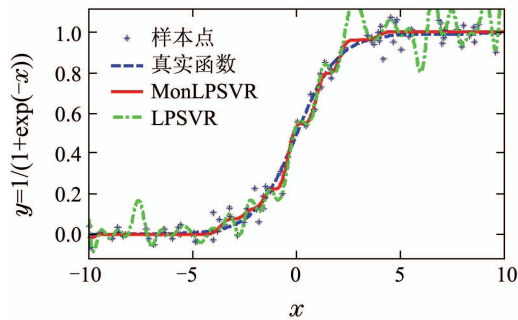
$$MSE = \frac{1}{k * m_v} \sum_{j=1}^k \sum_{i=1}^{m_v} (y_i^v - f(x_i^v))^2, \quad (18)$$

交叉验证折数 k 取 10, m_v 为验证集的样本数.

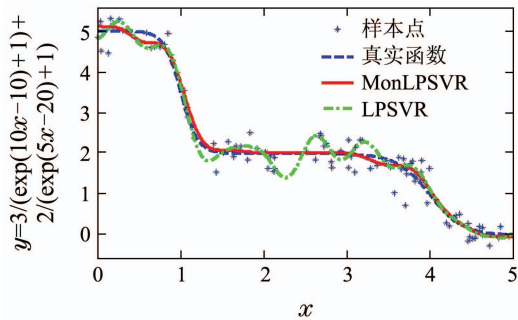
$$r = \frac{1}{n} \sum_{j=1}^n \left| \frac{1}{m_v} \sum_{i=1}^{m_v} \frac{(y_i - f_i(x))}{y_i} \right|. \quad (19)$$

由于原油沸点曲线的非单调出现在第 22, 23 个样本点, 所以不抽取它们作为验证点, 模型数 n 取 28.

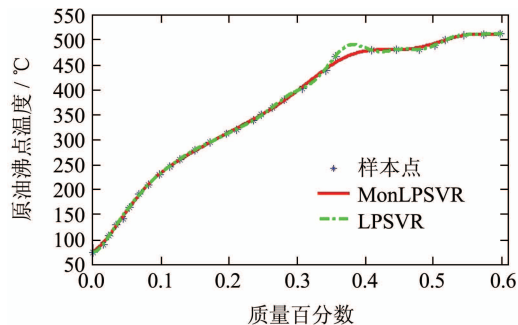
根据以上描述, 结合前一节的算法步骤操作, 绘制比较图如下:



(a) S 函数



(b) 高斯函数



(c) 原油实沸点曲线

图 1 改进SVR与标准SVR回归效果对比

Fig. 1 Improved SVR and standard SVR regression effect

图1(a)和图1(b)的噪声方差只取为0.1, 已经可以明显看出未加单调的回归曲线有较大的起伏, 不能满足模型的单调特性, 而经过本文改进的MonLPSVR算法很好地控制住了单调特性, 使回归曲线更加精确合理. 图1(c)中的实测值回归也体现出了良好效果.

由于PSO算法和线性规划算法得出的结果不唯一, 每次都有一定波动, 为保严谨, 独立重复进行上述实验50次, 将平均误差记于表1.

表 1 改进SVR算法与标准SVR算法精度对比

Table 1 Comparison between improved SVR algorithm and standard SVR algorithm

	噪声方差	LPSVR	MonLPSVR (w/o PD)	MonLPSVR (w/ PD)
S函数	0.1	0.0124	0.0012	0.0008
	0.2	0.1125	0.0035	0.0023
	0.5	1.0865	0.0211	0.0147
高原函数	0.1	0.0037	0.0020	0.0016
	0.2	0.0131	0.0052	0.0042
	0.5	0.1594	0.0166	0.0082
原油实沸点	—	0.0046	0.0023	0.0019

MonLPSVR(w/o PD)和MonLPSVR(w/ PD)分别表示未融合样本概率分布和有融合概率分布的单调线性规划支持向量回归算法, 下面简称为方法2和方法3.

接着, 分别选取方差为0.5的两组实验, 比较SVR模型得到的预测值与样本输出值的概率分布偏差, 偏差指标依然选取式(12)的4阶矩形式, 记于表2.

表 2 PSO改进前后SVR预测值概率分布对比

Table 2 Comparison of probability distribution before and after PSO-SVR improvement

	噪声方差	MonLPSVR (w/o PD)	MonLPSVR (w/ PD)
S函数	0.5	0.0121	0.0091
高原函数	0.5	0.0085	0.0044
原油实沸点	—	0.0016	0.0008

从表2可以看出加入样本概率分布先验知识后得到的SVR模型, 其预测值与真实值之间不仅在“微观上”的各自偏差较小, 而且整体上的概率分布特征依然保持高度一致. 使得预测结果蕴含了更多的信息, 更加可信.

4) 燃烧副反应数据.

接着采用6维的燃烧副反应数据^[11]进一步验证本文改进算法在多变量输入单变量输出情况下的有效性, 数据的输入和输出都归一化到[0, 1]之间, 将144组数据随机取出30, 50, 70作为缺失样本, 余下的再随机平分作为训练样本和测试样本. 单调约束的步长取0.01. 建模过程最优参数由原来的 $C = 460$, $\lambda = 5.9$ 修正为 $C = 400$, $\lambda = 3.3$. 每组实验进行50次, 取平均MSE作为误差评价, 记于表3.

表3 改进SVR算法与标准SVR算法精度对比(多维)
Table 3 Comparison between improved SVR algorithm and standard SVR algorithm(multidimensional)

	缺失 样本数	LPSVR	MonLPSVR (w/o PD)	MonLPSVR (w/ PD)
	30	0.0133	0.0102	0.0098
燃烧副反应	50	0.0157	0.0147	0.0142
	70	0.0185	0.0173	0.0159

从图1和表1-3可以看出本文提出的改进方法不仅使回归模型很好地满足了原函数所具备的单调性特征,在加入样本概率分布的修正后,使预测精度也有了明显提高,方法3较方法2和标准LPSVR都表现出其优越性.尤其是当噪声水平较高时,差距更大.

5 结语(Conclusions)

本文针对传统支持向量回归中,样本先验知识未被有效利用,而导致训练模型泛化能力较弱的问题,提出了结合样本概率分布特性和单调性先验知识的支持向量回归算法.为了使加入单调不等式约束后的算法求解简便,本文还将标准支持向量回归采用的二次规划改为线性规划.最后在优化模型重要参数(惩罚系数 C 和高斯核尺度 λ)的PSO算法中,加入完整样本的输出与缺失样本的预测输出之间的概率分布矩差异,用来修正粒子适应值,使两个概率分布尽量趋同.最终的回归模型既满足了单调性的先验知识,又融合了样本的概率分布特征.数值仿真实验很好地证明了该算法的有效性和优越性.进一步的工作可以考虑能否将概率分布特征直接融合在标准支持向量机的优化目标中.

参考文献(References):

- [1] VAPNIK V N. *The Nature of Statistical Learning Theory* [M]. The 2nd Edition. New York: Springer, 2000.
- [2] LAUER F, BLOCH G. Incorporating prior knowledge in support vector regression [J]. *Machine Learning*, 2008, 70(1): 89 – 118.
- [3] ZHOU Xiaojian. Enhancing ε -support vector regression with gradient information [J]. *Acta Automatica Sinica*, 2014, 40(12): 2908 – 2915 (周晓剑. 考虑梯度信息的 ε -支持向量回归机 [J]. 自动化学报, 2014, 40(12): 2908 – 2915.)
- [4] FUNG G M, MANGASARIAN O L, SHAFLIK J W. Knowledge-based support vector machine classifiers [C] // *Advances in Neural Information Processing Systems*, Vancouver: MIT Press, 2002, 15: 1 – 9.
- [5] MANGASARIAN O L, WILD E W. Nonlinear knowledge in kernel approximation [J]. *IEEE Transactions on Neural Networks*, 2007, 18(1): 300 – 306.
- [6] XU Yong, CHEN Zengqiang, YUAN Zhuzhi. Monotone approximation of RBF neural network [J]. *Systems Engineering*, 2004, 22(8): 5 – 9.

- (徐勇, 陈增强, 袁著祉. 单调RBF神经网络的逼近性分析 [J]. 系统工程, 2004, 22(8): 5 – 9.)
- [7] LAUER F, BLOCH G. Incorporating prior knowledge in support vector machines for classification: a review [J]. *Neurocomputing*, 2008, 71(7): 1578 – 1594.
- [8] CHEN J, XUE X, HA M, et al. Support vector regression method for wind speed prediction incorporating probability prior knowledge [J]. *Mathematical Problems in Engineering*, 2014, 2014(1): 149 – 160.
- [9] LIU Changxin, DING Jinliang, JIANG Bo, et al. Adaptive online support vector prediction method for concentrate grade of ore dressing process [J]. *Control Theory & Applications*, 2014, 31(3): 386 – 391.
(刘长鑫, 丁进良, 姜波, 等. 选矿过程精矿品位自适应在线支持向量预测方法 [J]. 控制理论与应用, 2014, 31(3): 386 – 391.)
- [10] ZHOU Jinzhu, HUANG Jin. Support vector regression based on multi kernel linear programming with integrated prior knowledge [J]. *Acta Automatica Sinica*, 2011, 37(3): 360 – 370.
(周金柱, 黄进. 集成先验知识的多核线性规划支持向量回归 [J]. 自动化学报, 2011, 37(3): 360 – 370.)
- [11] MANGASARIAN O L, MUSICANT D R. Large scale kernel regression via linear programming [J]. *Machine Learning*, 2002, 46(1 – 3): 255 – 269.
- [12] PAN C, DONG Y, YAN X, et al. Modified kernel regression integrated with monotonicity knowledge and its application to chemical engineering [J]. *Journal of Chemical Engineering of Japan*, 2014, 47(1): 52 – 59.
- [13] TAO Jianwen, WANG Shitong. Consistent local domain adaptive learning with kernel distribution [J]. *Acta Automatica Sinica*, 2013, 39(8): 1295 – 1309.
(陶剑文, 王士同. 核分布一致局部领域适应学习 [J]. 自动化学报, 2013, 39(8): 1295 – 1309.)
- [14] TAO Guiping, HAN Liyan, SONG Jie. Probability distribution difference measurement based on moments [J]. *Systems Science and Mathematics*, 2013, 33(9): 1071 – 1082.
(陶桂平, 韩立岩, 宋捷. 基于矩的概率分布差异度量 [J]. 系统科学与数学, 2013, 33(9): 1071 – 1082.)
- [15] JI Gang, WANG Jingcheng, GE Yang, et al. Gravitational search algorithm-least squares support vector machine model forecasting on hourly urban water demand [J]. *Control Theory & Applications*, 2014, 31(10): 1377 – 1382.
(耿钢, 王景成, 葛阳, 等. 城市小时级需水量的改进型引力搜索算法—最小二乘支持向量机模型预测 [J]. 控制理论与应用, 2014, 31(10): 1377 – 1382.)
- [16] KENNEDY J, EBERHART R. Particle Swarm Optimization [C] // *IEEE International Conference on Neural Networks*, Perth: IEEE, 1995: 1942 – 1948.
- [17] CHEN Chongwei, CHEN Dezhaohao. Simulation of crude boiling point distillation curve based on prior knowledge of feedforward network [J]. *Journal of Chemical Engineering*, 2001, 15(4): 351 – 356.
(陈翀伟, 陈德钊. 基于先验知识的前馈网络对原油实沸点蒸馏曲线的仿真 [J]. 高校化学工程学报, 2001, 15(4): 351 – 356.)

作者简介:

张青 (1991–), 男, 硕士研究生, 主要从事机器学习的研究, E-mail: zqonly@qq.com;

颜学峰 (1972–), 男, 博士生导师, 从事系统工程研究, E-mail: xfyfan@ecust.edu.cn.