

支持向量回归机元参数优化方法

宋彦坡¹, 彭小奇^{1,2}, 胡志坤¹

(1. 中南大学能源科学与工程学院, 湖南长沙 410083;
2. 湖南第一师范学院信息科学与工程系, 湖南长沙 410205)

摘要: 为了优化 ϵ 不敏感支持向量回归机(ϵ -support vector regression, ϵ -SVR)的三类元参数, 根据其耦合程度将其优化问题分解为核参数优化和结构参数(即不敏感参数和正则化参数)优化两个子问题, 并提出了相应的优化方法。首先, 提出了一种新的核校准系数以优化核参数; 其次, 提出了一种基于期望训练误差的结构参数优化方法; 最后, 为准确估算 ϵ -SVR 的期望训练误差, 还提出了一种根据实际训练误差分布特征评估和校正期望误差的方法。仿真结果表明, 该文方法具有与交叉检验法近似的优化效果, 且时间效率更高。

关键词: 支持向量回归机; 参数优化; 核校准; 期望误差

中图分类号: TP 274

文献标志码: A

DOI: 10.3969/j.issn.1001-506X.2010.10.46

Meta-parameters optimization method for support vector regression

SONG Yan-po¹, PENG Xiao-qi^{1,2}, HU Zhi-kun¹

(1. School of Energy Science and Engineering, Central South Univ., Changsha 410083, China;
2. Dept. of Information Science and Engineering, Hunan First Normal Univ., Changsha 410205, China)

Abstract: To optimize the meta-parameters of ϵ insensitive support vector regression (ϵ -SVR), the meta-parameters optimization problem is divided into two sub-problems named as kernel parameter optimization and structure parameters (including insensitive parameter and regularization parameter) optimization according to the coupling degrees among them, and corresponding optimization methods are proposed. First, a new kernel alignment coefficient is proposed for the former. Second, a method based on expectation training error is proposed for the latter. Finally, to estimate accurately the expectation error of the ϵ -SVR, a method to evaluate and adjust expectation error according to the distribution characteristics of the real training errors is proposed. Simulation results show that the proposed method is nearly as accurate as the cross validation method, and much more rapid.

Keywords: support vector regression (SVR); parameter optimization; kernel alignment; expectation error

0 引言

ϵ 不敏感支持向量回归机(ϵ -support vector regression, ϵ -SVR)具有 3 类元参数: 核参数 p (本文仅考虑高斯核函数)、不敏感参数 ϵ 和正则化参数 C 。这 3 类元参数对 ϵ -SVR 的建模效果具有较大影响。因此, 元参数的优化问题一直备受关注, 但至今仍未得到很好解决。

交叉检验法^[1-2]的优化效果好, 但需要多次读取数据和训练模型, 时间复杂度极高, 即使采用混沌优化^[3]、微粒群算法^[4]、遗传算法^[5]等智能优化方法也难以从根本上解决该问题。文献[6-7]认为, ϵ 与噪声方差成正例, 但其对比例系数的计算方法基于“大数定理”, 不完全适用于小样本

数据集。文献[8]考虑到样本集大小对 ϵ 的影响, 对 Smola 的方法进行了修正, 但该修正缺乏理论基础, 带有较大主观性。文献[9]认为, C 与样本应变量的取值范围有关, 并提出了根据应变量取值设置 C 的方法。文献[8]考虑到异常点的影响, 对该方法进行了修正, 但从文献[8-9]的相关推导可知, 这两种方法仅利用了 C 的取值下限信息, 完全忽略了其他因素的影响。文献[8]还认为核参数 p 与自变量的取值范围有关, 并提出了 p 的计算方法, 但该方法完全忽略了应变量信息, 显然不够合理。文献[10-11]提出了根据特征变量空间和目标变量(应变量)空间内积关系的相似性判断核函数的合理性的核校准系数, 但该系数对特征空间和目标空间的坐标平移敏感, 因此不能真正体现核函数与

收稿日期: 2009-08-31; 修回日期: 2010-01-28。

基金项目: 国家自然科学基金重点项目(60634020); 国家自然科学基金青年项目(60904077); 湖南省科技计划项目(2010FJ4132)资助课题

作者简介: 宋彦坡(1979-), 男, 讲师, 博士, 主要研究方向为数据挖掘、智能建模。E-mail: songyanpo@126.com

数据的适配度。而且,除交叉检验之外,上述方法均未考虑三个元参数之间的协调。

为此,本文以同时提高优化效果和时间效率为目标,为 ϵ -SVR 提出一种新的元参数优化方法。

1 SVR 优化问题分析

ϵ -SVR 的回归过程本质上可理解为:首先运用非线性映射 $\phi(\cdot)$ 把原始空间中的数据 $\{\mathbf{X}_i, y_i\}_1^n$ 映射到高维特征空间 $\{\phi(\mathbf{X}_i), y_i\}_1^n$, 然后在特征空间线性回归,即构造最优线性函数

$$\hat{y} = \{\phi(\mathbf{X}), \mathbf{w}\} + b \tag{1}$$

式中, \mathbf{w} 和 b 分别为待定的权重向量和偏置项。确定 \mathbf{w} 和 b , 需求解如下优化问题

$$\begin{aligned} \min & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \hat{\xi}_i) \\ \text{s. t. } & \{\phi(\mathbf{X}_i), \mathbf{w}\} + b - y_i \leq \epsilon + \xi_i \\ & y_i - \{\phi(\mathbf{X}_i), \mathbf{w}\} - b \leq \epsilon + \hat{\xi}_i \\ & \xi_i, \hat{\xi}_i \geq 0, i = 1, 2, \dots, n \end{aligned} \tag{2}$$

因为, $\phi(\mathbf{X})$ 维数往往很高甚至为无穷,故该优化问题难以直接求解,为此需要通过求解其对偶问题来间接求解,即

$$\begin{aligned} \max & \sum_{i=1}^n (\hat{\alpha}_i - \alpha_i) y_i - \epsilon \sum_{i=1}^n (\hat{\alpha}_i + \alpha_i) - \\ & \frac{1}{2} \sum_{i,j=1}^n (\hat{\alpha}_i - \alpha_i) (\hat{\alpha}_j - \alpha_j) k(\mathbf{X}_i, \mathbf{X}_j) \\ \text{s. t. } & 0 \leq \alpha_i, \hat{\alpha}_i \leq C, i = 1, 2, \dots, n \\ & \sum_{i=1}^n (\hat{\alpha}_i - \alpha_i) = 0, i = 1, 2, \dots, n \end{aligned} \tag{3}$$

式中, $k(\mathbf{X}_i, \mathbf{X}_j) = \{\phi(\mathbf{X}_i), \phi(\mathbf{X}_j)\}$ 为核函数(本文取高斯核函数 $k(\mathbf{x}, \mathbf{z}) = \exp(-\|\mathbf{x} - \mathbf{z}\|^2 / p^2)$)。解式(3)可得

$$\hat{y} = \sum_{i=1}^n \alpha_i^* k(\mathbf{X}_i, \mathbf{X}) + b \tag{4}$$

式中, $\alpha_i^* = \hat{\alpha}_i - \alpha_i \neq 0$ 只对部分样本成立,这部分样本被称为支持向量。

显然,正则化参数 C 、不敏感系数 ϵ 、核参数 p 均能影响支持向量机建模效果,且三者之间存在一定的耦合,但耦合程度各有不同。原始空间到特征空间的映射由 p 单独决定,而回归问题的损失函数由三者共同决定。因此, ϵ 和 C 之间存在强烈耦合,而 p 与 ϵ 和 C 的耦合程度相对较小。根据其耦合程度不同,本文将 ϵ -SVR 的参数优化问题近似解耦为超结构参数(核参数)优化和结构参数(不敏感参数和正则化参数)优化两个相对独立的问题,并分别提出解决方案。

2 SVR 参数优化方法

2.1 核参数的优化

核函数的作用在于将数据从原始变量空间映射到特征空间,为后续的线性回归创造条件。因此,核函数的合理性在一定程度上可以根据线性回归的可行性反映。由式(1)可知,线性回归可行的前提条件是样本在特征空间沿某个

方向的相对距离关系和样本在应变量空间的相对距离关系一致,最理想的情况是满足前提条件且样本在特征空间成一条直线。因此,根据样本在特征空间的实际分布和理想分布的差异可以评价核函数的合理性。

文献[10-11]提出的核校准方法与上述思想有近似之处,他们定义的核校准系数(本文简称其为 Nello 系数)为

$$A(S, k) = \frac{\{\mathbf{K}, \mathbf{Y}\}}{\sqrt{\{\mathbf{K}, \mathbf{K}\} \{\mathbf{Y}, \mathbf{Y}\}}} \tag{5}$$

式中, \mathbf{K} 为核矩阵, $\mathbf{K}_{i,j} = k(\mathbf{X}_i, \mathbf{X}_j)$; \mathbf{Y} 为应变量内积矩阵, $\mathbf{Y}_{i,j} = \{y_i, y_j\}$; $\{\mathbf{X}, \mathbf{Y}\} = \sum_{i=1}^R \sum_{j=1}^L \mathbf{X}_{i,j} \mathbf{Y}_{i,j}$ 为两等维矩阵的内积,其中 R 和 L 分别为矩阵的行数和列数。

Nello 系数可以理解成两个二维向量 \mathbf{K} 和 \mathbf{Y} 夹角的余弦值,它反映了样本在特征空间和应变量空间内积关系的近似度。由于样本的位置关系由内积关系确定,因此 Nello 系数间接反映了样本在特征空间的实际分布和理想分布的差异,并反映了核函数对实际样本的适配性,即核函数的合理性。但是,由于内积关系和距离关系并非一一对应,故 Nello 系数对特征空间或应变量空间的坐标平移敏感,两个对样本的适配度相同的核函数,可能具有不同的 Nello 系数,这说明该系数作为核函数的优化标准不尽合理。为此,本文利用二次距离矩阵替换式(5)中的内积矩阵构造一种修正的核校准系数(本文简称其为本文系数)

$$RA(S, k) = \frac{\{\mathbf{D}_k, \mathbf{D}_y\}}{\sqrt{\{\mathbf{D}_k, \mathbf{D}_k\} \{\mathbf{D}_y, \mathbf{D}_y\}}} \tag{6}$$

式中, \mathbf{D}_k 、 \mathbf{D}_y 分别为特征空间、应变量空间的二次距离矩阵

$$\mathbf{D}_{k(i,j)} = k(x_i, x_i) + k(x_j, x_j) - 2k(x_i, x_j) \tag{7}$$

$$\mathbf{D}_{y(i,j)} = (y_i - y_j)^2 \tag{8}$$

显然,本文系数直接体现了特征空间中样本的实际分布与理想分布的差异,避免了内积关系的不确定性,因此更适用于优化核参数。

需要指出,上述的两种核校准系数反映的均是整个特征空间和应变量空间的适配性,核校准系数越高,相应的核函数为最优的概率越大,但两者都不能绝对对准确的描述核函数对数据的适配度。

2.2 不敏感参数与正则化参数的优化

不敏感参数 ϵ 和正则化参数 C 之间存在较强的耦合,因此文献[6-9]采用的对 ϵ 和 C 分别优化的策略存在理论上的缺陷,而优化效果较好的交叉检验法时间效率过低。为此,本文提出了一种基于期望训练误差估算的 ϵ 和 C 的优化方法。

利用数据建立的模型,其训练误差存在最优值(称其为“期望训练误差”),此值过大则导致“欠拟合”,过小则导致“过拟合”。模型的期望训练误差可根据建模用数据特征估算,实际训练误差和期望训练误差之间的偏差可作为模型优化的一种新判据。当模型训练误差近似等于期望训练误差,并具有很好的泛化能力时,可认为模型是最优的。对于 ϵ -SVR,可认为模型最优的一个近似充要条件为

$$MSE_{tr} \approx MSE_{t0}, \|\mathbf{w}\|^2 \text{ 最小} \tag{9}$$

式中, MSE_{tr} 为模型的实际训练误差; MSE_{t0} 为模型的期望训练误差。

在核参数确定的情况下,可以按式(9)选择 ϵ 和 C 。尽管该问题是一个非线性规划问题,不易直接求解,但因 MSE_{tr} 和 $\|w\|^2$ 均可计算,故可以利用文献[8]的方法设定两类参数的初始值。然后,调整两类参数,直至式(9)满足。

受式(9)的约束,两个参数的优化搜索方向是相对确定的。例如,假定两组不同的参数 (ϵ_1, C_1) 、 (ϵ_2, C_2) 均满足条件 $MSE_{tr} = MSE_{tr0}$, 则

$$(\epsilon_1 - \epsilon_2)(C_1 - C_2) > 0 \tag{10}$$

从搜索策略看,该优化问题的计算量相对于单参数优化问题不成指数增长关系,求解相对简单。

2.3 期望训练误差的估算与校正

若建模样本 $\{X_i, y_i\}_i^n$ 的应变变量数据中存在噪声污染,即 $y_i = y_i^* + \delta_i$ (其中, y_i^* 为理想应变变量值; δ_i 为随机噪声),则理想模型的训练误差为^[12]

$$MSE_{tr} = \sum_{i=1}^n \delta_i^2 / K = \sigma^2 \tag{11}$$

式中, σ^2 为数据的噪声方差。几乎所有的数据建模方法的训练误差一般都小于检验误差,因此模型的期望训练误差满足以下条件

$$MSE_{tr0} < \sigma^2, MSE_{tr0} \propto \sigma^2 \tag{12}$$

数据的噪声方差 σ^2 可以根据数据特点估算^[13-14], 本文选用文献[13]的方法,即

$$\hat{\sigma}^2 = \frac{1}{2n} \sum_{i=1}^n (y_{i,n} - y_i)^2 \tag{13}$$

式中, n 为样本个数; $y_{i,n}$ 为 y_i 的最近邻估计,即在自变量空间 X 中距离 X_i 最近的样本应变变量值。

因噪声方差估算存在误差,并且期望训练误差和噪声方差之间的比例关系难以确定,故期望训练误差的估算存在一定误差。本文针对 ϵ -SVR 的特点对期望训练误差估算值做适当的判别和校正。

ϵ -SVR 本质上是特征空间里的线性回归,因此当训练误差过大时,样本误差与其应变变量在数值上存在较大的线性负相关性;当训练误差适中时,样本误差分布特点和样本中的噪声分布特点类似,样本误差和应变变量之间不存在线性相关性;当训练误差过小时,样本误差和应变变量之间一般也不存在线性相关性。针对 sinc 函数回归问题,图 1、图 2 示意了训练误差过大(SVR1,error1)、适中(SVR2,error2)和过小(SVR3,error3)三种情况下支持向量机模型的回归效果和误差分布。

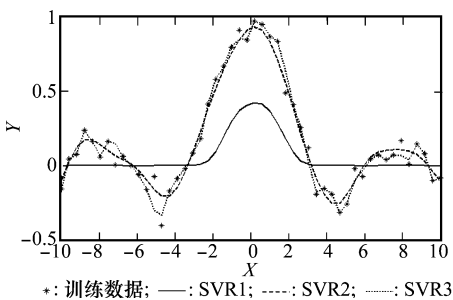


图 1 不同训练误差时 SVR 模型的效果

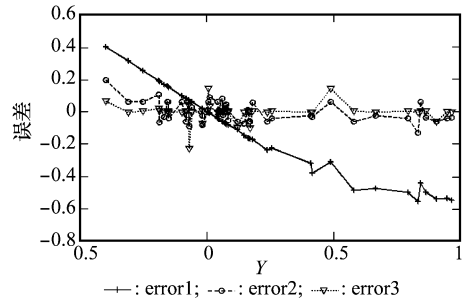


图 2 样本预报误差与应变变量之间的关系

结果如图 1 和图 2 所示,与理论分析一致,说明期望训练误差估值的准确性可以根据训练误差序列和应变变量序列的相关性进行评估和校正。

两数据序列间的线性相关性可根据其线性相关系数及其 t 检验量确定

$$R(x, y) = \frac{\{x, y\}}{\sqrt{\{x, x\}\{y, y\}}} \tag{14}$$

$$T(R_\alpha) = \frac{R_\alpha \sqrt{n-2}}{\sqrt{1-R_\alpha^2}} \sim t(n-2) \tag{15}$$

式中, α 为相关显著性水平,可根据 $T(R_\alpha)$ 值并查 t 分布表获得。

$T(R_\alpha)$ 的计算考虑了样本容量的影响,相关显著性水平 α 比相关系数更能反映变量间的线性相关程度,本文主要根据该参数判断误差序列和应变变量序列的相关性。

仅根据其相关性无法区分模型的“合理”和“过拟合”,为防止模型“过拟合”,误差序列和应变变量序列间应存在较弱的负相关性。为此,本文将 α 的范围设为 $0.1 \sim 0.25$, 相应的相关系数范围可利用式(15)和 t 分布表求取。当样本个数分别为 50, 100, 150, 400 时,相应的相关系数范围分别为: $-0.19 \sim -0.10$, $-0.13 \sim -0.07$, $-0.11 \sim -0.06$, $-0.07 \sim -0.03$ 。

需要指出,实际应用中,样本对应的理想应变变量值是未知的,但可以以 ϵ -SVR 的输出作为应变变量,当模型训练误差 (MSE_{tr}) 显著小于应变变量方差时,这样使用是合理的。

2.4 支持向量机参数优化算法

本文提出的支持向量机元参数优化方法流程如图 3 所示,该方法包含四个基本环节,部分环节间存在嵌套关系:

(1) 初始化。依据本文提出的校准系数优化核参数,利用式(13)估计期望训练误差。

(2) 核参数的优化。核参数决定了支持向量模型训练误差的下限,本环节检验该误差下限是否小于期望训练误差。若是,则不必调整核参数,否则应调整核参数(对高斯核函数而言,即减小 p)。

(3) 期望训练误差的检验和校正。根据误差序列和应变变量序列的相关系数 R_α , 判断期望训练误差的设置是否合理,若不合理则进行适当调节。

(4) 支持向量结构参数 ϵ 、 C 的优化。调节 ϵ 、 C , 使

SVM 模型满足受限最优模型的近似充要条件(即 $MSE_{tr} \approx MSE_{tr0}$, $\|w\|^2$ 最小),优化 ϵ 是为了使 $\|w\|^2$ 最小,优化 C 是为了使 $MSE_{tr} \approx MSE_{tr0}$ 。

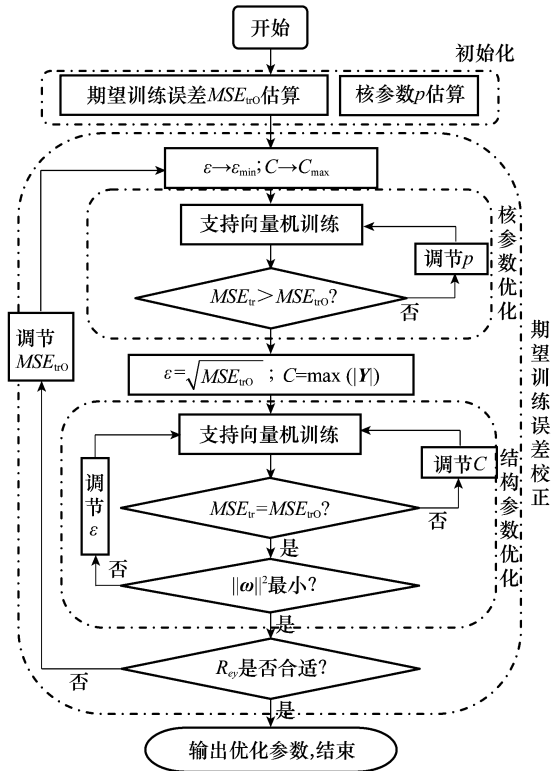


图 3 支持向量机参数优化算法流程

其中,第(3)、第(4)个环节涉及需要多次对支持向量机模型进行训练(或多次求解二次规划问题),因此算法的执行时间主要消耗在这两个环节。每改变一次支持向量参数,本文方法只需训练支持向量机模型一次,而 k 折交叉检验方法需要 k 次、留一法交叉检验需要 n 次(n 为训练样本个数),因此本文方法的时间消耗远小于交叉检验类方法。

3 仿真分析

利用 Matlab6.5 对本文方法进行仿真测试,所有实验均在一台 PentiumIV 2.8 GHz 的 PC 机上进行,测试数据选用机器学习领域广为应用的两类测试数据 sinc 数据(测试一元回归方法)和 Friedman 数据(测试多元回归方法)。sinc 函数定义如下

$$f(x) = \sin |x| / |x| \quad (16)$$

式中, $x \in [-10, 10]$ 。在论域范围内随机采集一定数量的样本,并在样本应变量数据上添加方差不同的高斯噪音,构成 6 组样本规模、噪音方差不等的数据集。

Friedman-1 函数定义如下

$$f(x) = 10\sin(\pi x_1 x_2) + 20(x_3 - 0.5)^2 + 10x_4 + 5x_5 \quad (17)$$

式中, $x_i \in [0, 1]$ 。在论域范围内随机采集 400 组数据,并添加不同强度的高斯噪音,构成 2 组训练数据样本。

模型效果主要采用误差平均功率 EAP 和误差相对功率 EVS 两类指标描述,分别定义如下

$$EAP = \frac{\int \dots \int (f(\mathbf{X}) - \hat{f}(\mathbf{X}))^2 \prod_{j=1}^m dx_j}{\int \dots \int \prod_{j=1}^m dx_j} \quad (18)$$

$$EVS = \frac{\int \dots \int (f(\mathbf{X}) - \hat{f}(\mathbf{X}))^2 \prod_{j=1}^m dx_j}{\int \dots \int (f(\mathbf{X}) - \bar{f}(\mathbf{X}))^2 \prod_{j=1}^m dx_j} \quad (19)$$

式中, $f(\mathbf{X})$ 、 $\hat{f}(\mathbf{X})$ 和 $\bar{f}(\mathbf{X})$ 分别为目标函数、支持向量机模型和目标函数均值; $\mathbf{X} = (x_1, x_2, \dots, x_m)$ 为自变量矢量。

针对上述 8 组数据,利用本文方法和十折交叉检验方法进行仿真实验,测试结果见表 1。表中, Sinc、Fried 分别表示 sinc、Friedman 函数数据; size 为数据集中样本个数; $\hat{\sigma}^2$ 为的噪声方差估值^[13],并作为第一次优化的期望误差; $t_{\text{opt}1}$ 为本文方法第一次优化(包括初始化环节及支持向量机结构参数优化环节)所消耗的计算机时间; $t_{\text{opt}2}$ 为整个优化过程消耗的计算机时间; t_{opt} 为采用十折交叉检验法消耗的计算机时间。

表 1 相关测试结果

数据集	第一次优化结果		最终优化结果		十折交叉验证法	
	参数设置	模型效果	参数设置	模型效果	参数设置	模型效果
Sinc1 size=50 $\hat{\sigma}^2=0.2^2$ $\hat{\sigma}^2=0.0421$	$p=1.1700$ $\epsilon=0.2258$ $C=0.1802$	$EAP=0.0276$ $EVS=0.2219$ $R_{cv}=-0.4719$ $t_{\text{opt}1}=18.53$	$MSE_{tr0}=0.0295$ $p=1.1700$ $\epsilon=0.0172$ $C=0.2953$	$EAP=0.0094$ $EVS=0.0760$ $R_{cv}=-0.1625$ $t_{\text{opt}2}=42.97$	$p=3.1125$ $\epsilon=1.5404e-009$ $C=1.7911$	$EAP=0.0057$ $EVS=0.0460$ $t_{\text{opt}}=3.36e+4$
Sinc2 size=50 $\hat{\sigma}^2=0.1^2$ $\hat{\sigma}^2=0.0089$	$p=1.3600$ $\epsilon=0.0847$ $C=0.1517$	$EAP=0.0057$ $EVS=0.0462$ $R_{cv}=-0.6036$ $t_{\text{opt}1}=15.42$	$MSE_{tr0}=0.004$ $p=1.3600$ $\epsilon=0.0631$ $C=2.7985$	$EAP=0.0019$ $EVS=0.0150$ $R_{cv}=-0.2018$ $t_{\text{opt}2}=71.40$	$p=2.5723$ $\epsilon=0.0486$ $C=230.1773$	$EAP=0.0020$ $EVS=0.0162$ $t_{\text{opt}}=5.61e+3$
Sinc3 size=100 $\hat{\sigma}^2=0.2^2$ $\hat{\sigma}^2=0.0387$	$p=1.1700$ $\epsilon=0.2558$ $C=0.1677$	$EAP=0.0091$ $EVS=0.0731$ $R_{cv}=-0.4432$ $t_{\text{opt}1}=145.00$	$MSE_{tr0}=0.0290$ $p=1.1700$ $\epsilon=0.0909$ $C=1.1206$	$EAP=0.0042$ $EVS=0.0339$ $R_{cv}=-0.0986$ $t_{\text{opt}2}=322.83$	$p=1.4399$ $\epsilon=0.0853$ $C=0.6006$	$EAP=0.0032$ $EVS=0.0259$ $t_{\text{opt}}=1.29e+4$

续表

数据集	第一次优化结果		最终优化结果		十折交叉验证法	
	参数设置	模型效果	参数设置	模型效果	参数设置	模型效果
Sinc4 size=100 $\sigma^2=0.1^2$ $\hat{\sigma}^2=0.0075$	$p=1.3100$	EAP=0.0012	MSE _{tr0} =0.0071	EAP=0.0010	$p=0.8941$	EAP=0.0024
	$\epsilon=0.0865$	EVS=0.0100	$p=1.3100$	EVS=0.0082	$\epsilon=0.0758$	EVS=0.0193
	$C=0.2009$	$R_{cv}=-0.2157$	$\epsilon=0.0956$	$R_{cv}=-0.1078$	$C=1.8800$	$t_{\alpha p}=2.04e+4$
		$t_{\alpha p1}=109.9690$	$C=0.5625$	$t_{\alpha p2}=213.66$		
Sinc5 size=150 $\sigma^2=0.2^2$ $\hat{\sigma}^2=0.0297$	$p=1.1900$	EAP=0.0015	MSE _{tr0} =0.0267	EAP=0.0038	$p=3.7661$	EAP=0.0029
	$\epsilon=0.1665$	EVS=0.0124	$p=0.7500$	EVS=0.0303	$\epsilon=0.1012$	EVS=0.0234
	$C=0.2448$	$R_{cv}=-0.1732$	$\epsilon=0.1743$	$R_{cv}=-0.0933$	$C=19.2945$	$t_{\alpha p}=7.19e+4$
		$t_{\alpha p1}=299.50$	$C=2.1544$	$t_{\alpha p2}=680.06$		
Sinc6 size=150 $\sigma^2=0.1^2$ $\hat{\sigma}^2=0.0092$	$p=1.3200$	EAP=0.0030	MSE _{tr0} =0.0083	EAP=0.0020	$p=3.7661$	EAP=8.4e-4
	$\epsilon=0.0767$	EVS=0.0240	$p=1.32$	EVS=0.0134	$\epsilon=0.0767$	EVS=0.0068
	$C=0.0837$	$R_{cv}=-0.3399$	$\epsilon=0.0909$	$R_{cv}=-0.1034$	$C=97.1151$	$t_{\alpha p}=9.92e+4$
		$t_{\alpha p1}=406.52$	$C=1.018$	$t_{\alpha p2}=499.55$		
Fried1 size=400 $\sigma^2=1^2$ $\hat{\sigma}^2=3.45$	$p=1$	EAP=2.6863	MSE _{tr0} =0.701	EAP=0.3468	$p=1.2418$	EAP=0.3499
	$\epsilon=1.8686$	EVS=0.1149	$p=1$	EVS=0.0146	$\epsilon=0.2348$	EVS=0.0147
	$C=6.1405$	$R_{cv}=-0.1443$	$\epsilon=0.8589$	$R_{cv}=-0.0472$	$C=773.9671$	$t_{\alpha p}=5.67+5$
		$t_{\alpha p1}=2.10e+3$	$C=1.4536e+3$	$t_{\alpha p2}=9.3e+3$		
Fried2 size=400 $\sigma^2=3^2$ $\hat{\sigma}^2=10.19$	$p=0.8$	EAP=3.6267	MSE _{tr0} =6.5209	EAP=1.6690	$p=0.8879$	EAP=1.2804
	$\epsilon=3.1920$	EVS=0.1498	$p=0.8000$	EVS=0.0699	$\epsilon=1.0788$	EVS=0.0529
	$C=4.7721$	$R_{cv}=-0.0962$	$\epsilon=2.0429$	$R_{cv}=-0.059$	$C=50.2347$	$t_{\alpha p}=6.36e+5$
		$t_{\alpha p1}=3.78e+3$	$C=138.2395$	$t_{\alpha p2}=5.99e+3$		

由表 1 可知:① 大多数情况下,文献[13]的方法能较为准确地估计样本的噪声方差,经过适当优化后的期望训练误差小于噪声方差(事实上,多数基于数据的建模方法的期望训练误差均小于噪声方差);② 本文方法和交叉检验法的优化效果差异不大,但交叉检验法消耗的计算机时间远大于本文方法。

需要指出的是,在对“交叉检验”进行算法设计时,采用了相对保守的策略(搜索步长较小,并采取了一些避免局部极小的措施),这在保证结果近似全局最优的同时也增加了其时间消耗。尽管采用一些相对“冒险”的优化策略会一定程度上减少其时间消耗,但由于其在基本原理上的不足,其时间效率几乎不可能高于本文方法。

4 结 论

本文针对 ϵ -SVR 元参数优化提出了一种新方法,能取得和交叉检验法相当的优化效果,同时在时间效率上具有较大优势。此外,对比本文方法和交叉检验法的优化结果不难看出,尽管在多数数据集上两种优化方法的准确性近似,但其参数的具体取值却有较大不同。这说明 ϵ -SVR 的效果主要取决于三类参数的组合匹配程度而非单个参数的具体设置。这也说明了,只针对单个参数而不考虑参数间的匹配关系的优化策略难以取得最好的优化效果。

参考文献:

[1] Lee M M S, Keerthi S S, Ong C J, et al. An efficient method for computing leave-one-out error in support vector machines with Gaussian kernels[J]. *IEEE Trans. on Neural Networks*, 2004, 15(3):750-757.

[2] Duan K, Keerthi S S, Poo A N. Evaluation of simple performance measures for tuning SVM hyper-parameters[J]. *Neurocomputing*, 2003,51(4):41-59.

[3] Yuan X F, Wang Y N. Parameter selection of support vector machine

for function approximation based on chaos optimization[J]. *Journal of Systems Engineering and Electronics*, 2008,19(1):191-197.

[4] Huang C L, Dun J F. A distributed PSO-SVM hybrid system with feature selection and parameter optimization[J]. *Applied Soft Computing*, 2008,8(4):1381-1391.

[5] Huang C L, Wang C J. A GA-based feature selection and parameters optimization for support vector machines[J]. *Expert Systems with Applications*, 2006,31(2):231-240.

[6] Smola A J, Murata N, Scholkopf B, et al. Asymptotically optimal choice of epsilon-loss for support vector machines[C] // *Proc. of the 8th International Conference on Artificial Neural Networks*, 1998:105-110.

[7] Kwok J T, Tsang I W. Linear dependency between ϵ and the input noise in ϵ -support vector regression[J]. *IEEE Trans. on Neural Networks*, 2003,14(3):544-553.

[8] Cherkassky V, Ma Y Q. Practical selection of SVM parameters and noise estimation for SVM regression[J]. *Neural Networks*, 2004,17(1):113-126.

[9] Matterna D, Haykin S. Support vector machines for dynamic reconstruction of a chaotic system[M] // Schölkopf B, Burges J, Smola A. *Kernel Methods: Support Vector Machine*, USA: MIT Press,1999:211-241.

[10] Cristianini N, Kandola J, Elisseeff A, et al. *On kernel target alignment*[M]. USA: MIT Press, 2002:367-373.

[11] Kandola J, Shawe-Taylor J, Cristianini N. On the extensions of kernel alignment[EB/OL]. <http://www.neurocolt.org>.

[12] Peng X Q, Song Y P, Tang Y. Restricted optimal modeling method supervised by expectation error[C] // *Proc. of WRI Global Congress on Intelligent Systems*, 2009:151-154.

[13] Devroye L, Györfi L, Schaffer D, et al. The estimation problem of minimum mean squared error[J]. *Statistics and Decisions*, 2003,21(1):15-28.

[14] Radfar M H, Dansereau R M, Sayadiyan A. Nonlinear minimum mean square error estimator for mixture-maximisation approximation[J]. *Electronics Letters*, 2006,42(12):724-725.