

带结构风险最小化的最优区间回归模型辨识

刘小雍^{1†}, 方华京², 陈孝玉¹

(1. 遵义师范学院 工学院, 贵州 遵义 563006; 2. 华中科技大学 自动化学院, 湖北 武汉 430074)

摘要: 针对来自模型结构、参数以及测量数据的不确定性等因素, 传统的辨识方法获取的是确定性数学模型的点输出, 其鲁棒性差, 易受外界干扰. 因此, 采用区间输出比点输出更易于实际问题的研究. 基于复杂系统的不确定性测量数据以及系统参数的不确定性, 提出了最优区间回归模型辨识的一种新方法, 该方法将逼近误差的 L_∞ 范数思想与结构风险最小化理论相结合, 建立求解区间模型的最优化问题, 应用线性规划独立求解区间模型的上界和下界模型. 该方法在保证模型辨识精度的同时, 其泛化性能得到进一步提高. 实验分析表明, 提出的方法对来自噪声以及参数不确定性的数据, 可以从区间模型的辨识精度和泛化性能之间取其平衡.

关键词: 结构风险最小化; 不确定性分析; 逼近误差的 L_∞ 范数优化; 最优区间回归模型; 线性规划

引用格式: 刘小雍, 方华京, 陈孝玉. 带结构风险最小化的最优区间回归模型辨识. 控制理论与应用, 2020, 37(3): 560 – 573

DOI: 10.7641/CTA.2019.80786

Identification of optimal interval regression model with structural risk minimization

LIU Xiao-yong^{1†}, FANG Hua-jing², CHEN Xiao-yu¹

(1. College of Engineering, Zunyi Normal University, Zunyi Guizhou 563006, China;

2. School of Automation, Huazhong University of Science and Technology, Wuhan Hubei 430074, China)

Abstract: Aiming at the characteristics from a family of uncertain nonlinear functions or the systems with uncertain physical parameters, the problem of the conventional nonlinear system modeling, referred to as the deterministic modeling method whose output is a single value (or a point output), is prone to produce a poor robustness and is subject to external disturbance. This paper proposes a novel method for identifying optimal interval regression model (OIRM) with sparsity only based on the uncertain measurements of complex system. The OIRM, differently from standard deterministic models, is composed of upper regression model (URM) and lower regression model (LRM), and returns an interval output as opposed to a point output. The method combines sparsity stemming from the idea of structural risk minimization (SRM) principle, and optimality using L_∞ -norm of approximation errors with some notions from linear programming (LP) problem. The optimization problems corresponding to URM and LRM with constraints in a form of convex inequality and linear equality are independently solved by LP. Finally, the equilibrium between modeling accuracy and generalization performance of the proposed OIRM are demonstrated by the experimental cases using the two indices, the fractions of utilised support vectors (SVs) and root mean square error (RMSE).

Key words: structural risk minimization; uncertain analysis; L_∞ -norm optimization on approximation errors; optimal interval regression model; linear programming

Citation: LIU Xiaoyong, FANG Huajing, CHEN Xiaoyu, et al. Identification of optimal interval regression model with structural risk minimization. *Control Theory & Applications*, 2020, 37(3): 560 – 573

收稿日期: 2018–10–12; 录用日期: 2019–07–10.

[†]通信作者. E-mail: liuxy204@163.com; Tel.: +86 851-28939072.

本文责任编辑: 周杰.

国家自然科学基金项目(61473127), 贵州省科技计划项目(黔科合基础[2018]1179, 黔科合LH字[2016]7002号, 黔科合LH字[2016]7004号), 贵州省教育厅科技人才成长项目(黔教合KY字[2016]254), 贵州省高层次创新人才项目([2017]19)资助.

Supported by the National Natural Science Foundation of China (61473127), the Science and Technology Foundation of Guizhou Province ([2018] 1179, [2016]7002, [2016]7004), the Youth Foundation of Department of Education of Guizhou Province ([2016]254) and the Program for High-Level Innovative Talents of Guizhou ([2017]19).

1 引言

工业过程中的大多数控制系统结构主要是基于传统控制理论方法进行设计的,其适用范围受到了一定的限制,包括假定该过程为线性变化且整个过程的数学模型为准确已知^[1].然而,大多数物理系统包含复杂的非线性及耦合关系等因素,导致很难建立准确的数学模型;此外,对来自控制系统中的过程参数变化、外部干扰或传感器失效也对控制系统的设计增加了诸多不确定性^[2].这些现象的存在更需要探索一种更有效的建模方法,能够实现复杂的输入-输出映射及非线性函数逼近.因此,出现了基于数据或数据驱动的建模方法^[3-4],即通过传感器或其他数据获取设备获取被控系统的输入-输出数据,采用神经网络^[5]、Takagi-Sugeno(T-S)模糊模型^[6]、支持向量机^[7]等方法建立复杂系统的数学模型,这些方法的共同特点是建立的数学模型一般是确定的,不受系统参数变化、测量噪声或其他不确定性等因素的影响,具有一定的局限性:1)确定的数学模型不能自适应系统的变化;2)基于训练数据建立的数学模型结构较复杂、泛化性能差等.

然而,在众多的实际应用中,获取的信息往往呈现出不确定、不准确以及不完整等特征^[8],对于这一类特征的信息描述,若仍采用传统的确定性数学模型建模,显然不能更好的去捕捉这一类不确定性复杂系统的特征变化^[4].因此,针对不确定性复杂系统建模的研究,开始了从确定性数学模型到不确定性数学模型的扩展,包括模糊理论结合其他参数求解方法^[9-10]、概率统计中的分位方法^[11]、神经网络的上下界建模方法^[12-13]等.文献[10]最早将possibility与necessity回归模型相结合,采用模糊输出对区间回归进行分析,并应用二次规划对possibility与necessity回归模型进行参数求解.基于文献[10]的探索,又出现了模糊数据的另一种重要建模方法,即基于概率 α 分位和神经网络(neural network, NN)的区间回归分析方法.针对含噪声工业时间序列数据,文献[14]再次将NN与概率分位数方法相结合,提出了工业数据的区间模型预测方法.对于前一种方法^[11],首先从已知的不确定性数据中建立数据的主要变化趋势,再引入 α 分位分别获取区间模型的上、下界模型.后一种的区间模型分别用上界 $g^*(x)$ 和下界 $g_*(x)$ 模型来描述,求解过程仍然采用传统的反向传播(back propagation, BP)方法.众所周知,NN在模型结构选择上存在主观性较大、复杂,以及参数求解主要是基于最小二乘的目标优化,导致较差的模型泛化性能.近年来,基于结构风险最小化的支持向量机(support vector machine, SVM)成了重点关注,在模式识别和函数逼近问题上得到了成功的应用,文献[4, 15-16]提出了基于支持向量学习方

法的模糊回归分析,该方法较传统神经网络方法做了较好的改进.

进一步,考虑到基于SVM的回归模型求解存在两个问题:1)从SVM中非线性映射的大规模数据中求解有一定的困难;2)不同样本数据集间的大小不平衡,往往出现一个类的样本集比另一样本集大,引起分割间隔的偏移.因此,针对上述问题,文献[17]提出了简化SVM的区间回归分析方法来减小支持向量(support vector, SV)的个数,然而在提出的方法中是通过将 L_1 范数的结构风险与逼近误差的 L_∞ 范数相结合来实现.此外,文献[18]从应用的角度出发,通过将获取到的确定性测量数据(单值)转换为区间描述,建立了基于非参数加性模型的区间回归分析方法,正确实现区间测量数据可能的非线性模式识别.针对区间数据现有的回归分析使用确定的数据参考点再加上半径范围来建立区间的上、下界方法存在一定的限制,例如不同的数据集应该考虑采用不同的数据参考点描述区间,文献[19]提出了如何更好地从回归变量中提取最佳参考点的参数化方法.此外,从区间数据的另一个角度出发,当区间数据出现异常值的情况时,文献[20]对区间值变量的指数核鲁棒回归模型进行了研究,通过指数型核函数的使用,对区间值观测值中的异常值引入权值进行惩罚,进而保证了参数估计算法的收敛速度以及较低的计算成本.从提高模型建模精度出发,文献[21]将概率分位与群体智能化方法相结合对参数进行估计,并论证了方法的有效性与优越性.

在本文中,建立了一种新的区间回归模型辨识方法:1)基于文献[22]提出了最小化最大逼近误差的 L_∞ 范数引理,作为分别建立区间模型的上、下边回归模型优化问题的基础;2)提出了基于 L_1 范数框架下的结构风险最小化的代价函数,在保证辨识区间回归模型精度的同时,尽可能对模型结构复杂性进行有效控制,进而提高模型的泛化性能;3)基于引理1,提出了独立求解上、下边回归模型的两个定理,并应用简单线性规划对其求解.

2 凸二次规划到线性规划的支持向量回归

2.1 凸二次规划的支持向量回归

由Vapnik及其团队所建立的SVM^[23-24]很好的克服了传统过拟合的分类问题.因此,在SVM中引入不敏感损失函数可发展为非线性的回归问题,即支持向量回归(support vector regression, SVR),已在最优控制^[25]、时间序列预测^[26]、区间回归分析^[27]等方面得到了广泛应用.假设通过传感器或其他数据获取设备获取到一组带有噪声的测量数据 $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_N, y_N)\}$,SVR方法是对这组数据所对应的未知数学模型进行逼近,其中 $\mathbf{x}_k = (x_k^1, x_k^2, \dots, x_k^d)$ 表示 d 维的输入模式.假设一组线性无关的基函

数 $\{\omega_s(\mathbf{x}), s = 1, \dots, m\}$, 可用基函数的线性展开描述任意的非线性系统. 因此, 函数的逼近问题可转化为寻求如下基函数线性展开的最优参数求解^[24]:

$$f(\mathbf{x}, \boldsymbol{\theta}) = \sum_{s=1}^m \theta_s \omega_s(\mathbf{x}) + b, \quad (1)$$

其中: $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_m)$ 为需要被寻优的参数向量; b 是一个常量. 进一步, 该问题的参数寻优即为寻找满足如下优化问题的非线性函数 f :

$$\min: R(f) = \sum_{k=1}^N L_\varepsilon(y_k - f(\mathbf{x}_k)) + \gamma \|\mathbf{w}\|_2^2, \quad (2)$$

$R(f)$ 为结构风险; γ 表示规则化常量, 体现对模型结构与建模精度重视程度的权衡; $\|\mathbf{w}\|_2^2$ 的引入在于控制模型的复杂度; $L_\varepsilon(\cdot)$ 为描述 ε 不敏感损失函数, 定义为

$$L_\varepsilon(y_k - f(\mathbf{x}_k)) = \begin{cases} 0, & |y_k - f(\mathbf{x}_k)| \leq \varepsilon, \\ |y_k - f(\mathbf{x}_k)| - \varepsilon, & \text{其他}. \end{cases} \quad (3)$$

从上述 ε 域定义可知, 如果 $|y_k - f(\mathbf{x}_k)|$ 的值在该 ε 区域内, 损失为0; 否则为 $|y_k - f(\mathbf{x}_k)|$ 与 ε 的差值.

通过应用拉格朗日乘子方法, 对式(2)的最小化可转化为它的对偶优化问题^[28]:

$$\begin{aligned} \min: W(\boldsymbol{\alpha}^+, \boldsymbol{\alpha}^-) = & \varepsilon \sum_{k=1}^N L_\varepsilon(\alpha_k^- + \alpha_k^+) - \sum_{k=1}^N y_k(\alpha_k^+ - \alpha_k^-) + \\ & \frac{1}{2} \sum_{k,i=1}^N ((\alpha_k^+ - \alpha_k^-)(\alpha_i^+ - \alpha_i^-) \cdot \\ & \sum_{s=1}^N g_s(\mathbf{x}_k)g_s(\mathbf{x}_i)), \\ \text{s.t. } & \sum_{k=1}^N \alpha_k^+ = \sum_{k=1}^N \alpha_k^-, \quad 0 \leq \alpha_k^+, \alpha_k^- \leq \gamma, \\ & k = 1, 2, \dots, N, \end{aligned} \quad (4)$$

α_k^+, α_k^- 表示拉格朗日乘子. 式(4)中 $g_s(\mathbf{x})$ 的内积可用如下核函数代替:

$$K(\mathbf{x}_k, \mathbf{x}_i) = \sum_{s=1}^N g_s(\mathbf{x}_k)g_s(\mathbf{x}_i). \quad (5)$$

核函数确定了解的平滑特性, 选取时应该更好的反映数据的先验知识. 式(4)的优化问题可重写为

$$\begin{aligned} \min: W(\boldsymbol{\alpha}^+, \boldsymbol{\alpha}^-) = & \varepsilon \sum_{k=1}^N L_\varepsilon(\alpha_k^- + \alpha_k^+) - \sum_{k=1}^N y_k(\alpha_k^+ - \alpha_k^-) + \\ & \frac{1}{2} \sum_{k,i=1}^N ((\alpha_k^+ - \alpha_k^-)(\alpha_i^+ - \alpha_i^-) \sum_{s=1}^N K(\mathbf{x}_k, \mathbf{x}_i)), \end{aligned} \quad (6)$$

其中: $\boldsymbol{\alpha}^+ = (\alpha_1^+, \dots, \alpha_N^+)$, $\boldsymbol{\alpha}^- = (\alpha_1^-, \dots, \alpha_N^-)$. 基

于Vapnik的研究, SVR方法的解以核函数的线性展开描述如下:

$$f(\mathbf{x}, \boldsymbol{\alpha}^+, \boldsymbol{\alpha}^-) = \sum_{k=1}^N (\alpha_k^+ - \alpha_k^-) K(\mathbf{x}_k, \mathbf{x}_i) + b, \quad (7)$$

其中常量 b 的计算为

$$b = y_k - \sum_{k=1}^N (\alpha_k^+ - \alpha_k^-) K(\mathbf{x}_k, \mathbf{x}_i) + \varepsilon \cdot \text{sgn}(\alpha_k^+ - \alpha_k^-). \quad (8)$$

显然, 仅当 $(\alpha_k^+ - \alpha_k^-) \neq 0$ 时, 对应的样本 $\mathbf{x}_k$ 称为SV. 由于本文使用高斯核函数, 可将式(7)写成

$$f(\mathbf{x}, \boldsymbol{\alpha}^+, \boldsymbol{\alpha}^-) = \sum_{k=1}^N (\alpha_k^+ - \alpha_k^-) \exp\left\{\frac{\|\mathbf{x} - \mathbf{x}_k\|^2}{\sigma^2}\right\} + b, \quad (9)$$

σ 为高斯核参数.

2.2 SVR的优化问题转化

众所周知, SVR是将基于二范数的结构风险最小化作为优化目标, 实现对回归模型结构复杂性的控制, 从而提高模型的泛化性能; 然而, 在文献[29]中指出, SVR中的优化目标求解采用的是二次规划, 会产生模型的冗余描述以及昂贵的计算成本; 同时, 二范数的优化问题不能直接用于本文提出的最优区间回归模型辨识的优化问题, 需要进行二范数到一范数的转化. 对于二次规划(quadratic programming-SVR, QP-SVR)^[30], 基于式(2)的优化问题如下:

$$\begin{aligned} \min: R(f) = & \sum_{k=1}^N L_\varepsilon(y_k - f(\mathbf{x}_k)) + \gamma \|\mathbf{w}\|_2^2, \\ \text{s.t. } & \begin{cases} y_k - \langle \mathbf{w}, \psi(\mathbf{x}_k) \rangle - b \leq \varepsilon + \xi_k, \\ \langle \mathbf{w}, \psi(\mathbf{x}_k) \rangle - b - y_k \leq \varepsilon + \xi_k^*, \\ \xi_k, \xi_k^* \geq 0, \end{cases} \end{aligned} \quad (10)$$

其中: $\psi(\cdot)$ 表示从输入空间到高维空间的非线性特征映射, 即 $\psi: \mathbb{R}^n \rightarrow \mathbb{R}^m (m > n)$; ξ_k, ξ_k^* 为松弛变量, 分别对应超出正、负方向偏差值时的大小; 常量 γ 大于0, 反应非线性 f 与偏差大于 ε 时两者之间的平衡. 对于式(9), 令 $\beta_k = \alpha_k^+ - \alpha_k^-$, 则有

$$f(\mathbf{x}, \boldsymbol{\beta}) = \sum_{k=1}^N \beta_k \exp\left(\frac{-\|\mathbf{x} - \mathbf{x}_k\|}{2\sigma^2}\right) + b, \quad (11)$$

$\boldsymbol{\beta} = (\beta_1 \ \beta_2 \ \dots \ \beta_N)^T$. 考虑到式(2)的优化问题, $\|\mathbf{w}\|_2^2$ 范数的引入是为了控制模型的复杂度, 根据范数的等价性可知, 在结构风险中引入其他范数也可以同样对模型复杂性进行控制^[31]. 接下来, 将QP-SVR的优化问题(10)变成

$$\min: R(f) = \sum_{k=1}^N L_\varepsilon(y_k - f(\mathbf{x}_k)) + \gamma \|\boldsymbol{\beta}\|_1, \quad (12)$$

其中: $f(\mathbf{x}_k)$ 以式(9)形式描述; $\|\boldsymbol{\beta}\|_1$ 表示系数空间的1-范数. 因此, 新的约束优化问题为^[32]

$$\begin{aligned} \min: R(f) &= \gamma \cdot \sum_{k=1}^N (\xi_k + \xi_k^*) + \|\beta\|_1, \\ \text{s.t.} \quad &\begin{cases} y_k - \sum_{k=1}^N \beta_k \exp\left(\frac{-\|\mathbf{x} - \mathbf{x}_k\|}{2\sigma^2}\right)^2 - \\ b \leq \varepsilon + \xi_k, \\ \sum_{k=1}^N \beta_k \exp\left(\frac{-\|\mathbf{x} - \mathbf{x}_k\|}{2\sigma^2}\right)^2 + \\ b - y_k \leq \varepsilon + \xi_k^*, \\ \xi_k, \xi_k^* \geq 0. \end{cases} \end{aligned} \quad (13)$$

从几何的角度来看, ξ_k 和 ξ_k^* 之间的关系在SVR中满足 $\xi_k \cdot \xi_k^* = 0$. 因此, 在优化问题(13)中仅引入一个松弛变量 ξ_k 即可^[29], 即

$$\begin{aligned} \min: R(f) &= \gamma \cdot \sum_{k=1}^N \xi_k + \|\beta\|_1, \\ \text{s.t.} \quad &\begin{cases} y_k - \sum_{k=1}^N \beta_k \exp\left(\frac{-\|\mathbf{x} - \mathbf{x}_k\|}{2\sigma^2}\right)^2 - \\ b \leq \varepsilon + \xi_k, \\ \sum_{k=1}^N \beta_k \exp\left(\frac{-\|\mathbf{x} - \mathbf{x}_k\|}{2\sigma^2}\right)^2 + \\ b - y_k \leq \varepsilon + \xi_k, \\ \xi_k \geq 0. \end{cases} \end{aligned} \quad (14)$$

为了转化上述优化问题为线性规划问题, 将 β_k 和 $|\beta_k|$ 进行如下分解^[32]:

$$\begin{aligned} \beta_k &= \alpha_k^+ - \alpha_k^-, \\ |\beta_k| &= \alpha_k^+ + \alpha_k^-, \end{aligned} \quad (15)$$

其中 $\alpha_k^+, \alpha_k^- \geq 0$. 考虑到 α_k^+ 和 α_k^- 不能同时大于0, 对应3种不同情况的解, 即 $(\alpha_k^+, 0)$, $(0, \alpha_k^-)$, $(0, 0)$, 因此式(15)的分解是唯一的. 下面对线性规划问题(14)在式(15)的分解下, α_k^+ 和 α_k^- 不能同时大于0做简单证明. 采用反证法, 假设在问题(14)的最优解中存在某一个正整数 k , 使得 α_k^+ 和 α_k^- 同时大于零, 现取 $\tau = \min(\alpha_k^+, \alpha_k^-)$, 将 α_k^+ 和 α_k^- 分别换成 $\alpha_k^+ - \tau$ 和 $\alpha_k^- - \tau$, 仍然满足 $\alpha_k^+ - \tau \geq 0$ 和 $\alpha_k^- - \tau \geq 0$, 但是目标函数中的 $\alpha_k^+ + \alpha_k^-$ 将减少 $\alpha_k^+ + \alpha_k^- - 2\tau$, 这显然与最优解的假设相矛盾. 因此, α_k^+ 和 α_k^- 不能同时大于0, 即 $\alpha_k^+ \cdot \alpha_k^- = 0$.

基于式(15), 优化问题(14)进一步变成

$$\begin{aligned} \min: R(f) &= \gamma \cdot \sum_{k=1}^N \xi_k + \sum_{k=1}^N (\alpha_k^+ + \alpha_k^-), \\ \text{s.t.} \quad &\begin{cases} y_k - \sum_{k=1}^N (\alpha_k^+ - \alpha_k^-) \exp\left(\frac{-\|\mathbf{x} - \mathbf{x}_k\|}{2\sigma^2}\right)^2 - \\ b \leq \varepsilon + \xi_k, \\ \sum_{k=1}^N (\alpha_k^+ - \alpha_k^-) \exp\left(\frac{-\|\mathbf{x} - \mathbf{x}_k\|}{2\sigma^2}\right)^2 + \\ b - y_k \leq \varepsilon + \xi_k, \\ \xi_k \geq 0. \end{cases} \end{aligned} \quad (16)$$

现定义向量

$$\mathbf{c} = \underbrace{(1 \ 1 \ \cdots \ 1)}_N \underbrace{(1 \ 1 \ \cdots \ 1 \ 1 \ 1 \ \cdots \ 1)}_N \underbrace{(\gamma \ \gamma \ \cdots \ \gamma)}_N \quad (17)$$

和向量 β 的 L_1 范数

$$\|\beta\|_1 = \underbrace{(1 \ 1 \ \cdots \ 1)}_N \underbrace{(1 \ 1 \ \cdots \ 1)}_N \begin{pmatrix} \alpha^+ \\ \alpha^- \end{pmatrix}. \quad (18)$$

以向量形式将优化问题(16)构造为线性规划问题

$$\begin{aligned} \min \quad &\mathbf{c}^T \begin{pmatrix} \alpha^+ \\ \alpha^- \\ \xi \end{pmatrix}, \\ \text{s.t.} \quad &\begin{cases} \begin{pmatrix} \mathbf{K} & -\mathbf{K} & -\mathbf{I} \\ -\mathbf{K} & \mathbf{K} & -\mathbf{I} \end{pmatrix} \begin{pmatrix} \alpha^+ \\ \alpha^- \\ \xi \end{pmatrix} \leq \\ \begin{pmatrix} \mathbf{y} + \varepsilon \\ \varepsilon - \mathbf{y} \end{pmatrix}, \\ \alpha^+, \alpha^- \geq 0, \xi \geq 0, \end{cases} \end{aligned} \quad (19)$$

其中: $\xi = (\xi_1 \ \xi_2 \ \cdots \ \xi_N)^T$; $\mathbf{I}$ 为 $N \times N$ 单位矩阵;

$$\mathbf{y} = (y_1 \ y_2 \ \cdots \ y_N)^T,$$

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

线性规划问题(19)可通过单纯型算法或内点算法进行求解^[33]. 对于QP-SVR, 在 ε 域之外的所有数据点将被选择为支持向量(SV); 而对于LP-SVR, 即便 ε 域选择为0时, 由于软约束在优化问题中的使用, LP-SVR 仍然能够获取稀疏解. 通常情况下, 稀疏解往往通过设定非零的 ε 域来获取.

3 基于逼近误差的 L_∞ 范数的回归模型辨识

回归模型的辨识离不开参数的求解, 为了能够控制模型结构复杂性控制的优化问题应用到回归模型辨识, 下面将讨论将基于逼近误差的 L_∞ 范数作为评判标准, 建立回归模型辨识的优化问题. 假设通过传感器或数据获取设备获取一组测量数据 $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_N, y_N)\}$, 其中 $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ 描述输入测量数据, 对应的输出定义为 $\{y_1, y_2, \dots, y_N\}$. 设测量满足如下非线性系统模型:

$$y_k = g(\mathbf{x}_k), \quad k = 1, 2, \dots, N. \quad (20)$$

根据统计学理论可知^[34], 存在以式(11)描述的非线性回归模型 f 对测量模型 g 的任意逼近, 当逼近精度越小时, 需要的支持向量越少; 反之, 逼近精度越高, 则支持向量越多. 因此, 对任意给定的实连续函数 g 及 $\eta >$

0, 存在如下回归模型 f 满足:

$$\sup_{\mathbf{x}_k \in \mathcal{S}} |f(\mathbf{x}_k) - g(\mathbf{x}_k)| < \eta, \forall k. \quad (21)$$

值得指出的是, 较小的 η 值, 对应式(11)较多的支持向量. 现讨论回归模型, 式(11)的另一种参数求解方法. 在非线性系统模型的逼近情况下, 定义实际输出与由式(11)定义的SVR模型输出之间的偏差 e_k :

$$e_k = y_k - f(\mathbf{x}_k), \forall k. \quad (22)$$

为了估计SVR模型的最优参数, 考虑如下最大建模误差的最小化:

$$\max_{\mathbf{x}_k \in \mathcal{Z}} |y_k - f(\mathbf{x}_k)|, \forall k, \quad (23)$$

$\mathcal{Z}$ 表示整个输入数据集. 显然, 这是一个最小-最大(min-max)优化问题. 在式(11)描述的回归模型情况下, 式(23)的最小化可通过如下两个阶段完成: 1) 核函数中的核宽度 σ 的参数寻优, 通常采用经典的 k 阶交叉验证或其他优化方法来实现; 2) 式(11)的参数确定可通过min-max优化问题求解, 即

$$\begin{aligned} \beta = \\ \arg \min_{\beta} \max_{\mathbf{x}_k \in \mathcal{Z}} |y_k - \sum_{k=1}^N \beta_k \exp(\frac{-\|\mathbf{x} - \mathbf{x}_k\|^2}{2\sigma^2}) - b|. \end{aligned} \quad (24)$$

基于文献[22], 优化问题(24)可通过如下引理1求解, 该引理也是本文提出方法用于辨识最优区间回归模型的基础.

引理1 min-max优化问题中的参数可通过最小化 λ , 且满足如下不等式约束的线性规划求解, 即

$$\begin{aligned} \min : \lambda, \\ \text{s.t.} \begin{cases} y_k - \sum_{k=1}^N \beta_k \cdot \\ \exp(\frac{-\|\mathbf{x} - \mathbf{x}_k\|^2}{2\sigma^2}) - b \leq \lambda, \\ -y_k + \sum_{k=1}^N \beta_k \cdot \\ \exp(\frac{-\|\mathbf{x} - \mathbf{x}_k\|^2}{2\sigma^2}) + b \leq \lambda, \\ \lambda \geq 0, k = 1, 2, \dots, N, \end{cases} \end{aligned} \quad (25)$$

其中: $\beta = (\beta_1 \ \beta_2 \ \dots \ \beta_N)^T$ 和 b 为被求解参数, λ 表示最大逼近误差.

证 定义 λ 如下:

$$\lambda = \max_{\mathbf{x}_k} |y_k - \sum_{k=1}^N \beta_k \exp(\frac{-\|\mathbf{x} - \mathbf{x}_k\|^2}{2\sigma^2})|, \quad (26)$$

可直接得出如下不等式成立:

$$\begin{aligned} |y_k - \sum_{k=1}^N \beta_k \exp(\frac{-\|\mathbf{x} - \mathbf{x}_k\|^2}{2\sigma^2})| \leq \lambda, \\ k = 1, 2, \dots, N. \end{aligned} \quad (27)$$

根据去绝对值运算, 可知式(1)的约束条件成立.

证毕.

4 基于 L_∞ 范数区间回归模型辨识

有了基于 L_∞ 范数的回归模型辨识的基础之后, 接下来将讨论区间回归模型的辨识的一种新方法, 为了给出辨识方法的清晰思路, 整个辨识过程如图1所示.

假设不确定非线性函数或非线性系统属于某函数簇 Γ , 由名义函数 g 和不确定性 $\Delta g(\mathbf{x})$ 两部分组成, 即

$$\begin{aligned} \Gamma = \{g : \mathcal{S} \rightarrow \mathbb{R}^1 | g(\mathbf{x}) = \\ g_{\text{nom}}(\mathbf{x}) + \Delta g(\mathbf{x}), \mathbf{x} \in \mathcal{S}\}, \end{aligned} \quad (28)$$

g_{nom} 为标称函数, 不确定性 $\Delta g(\mathbf{x})$ 满足

$$\sup_{\mathbf{x} \in \mathcal{S}} |\Delta g(\mathbf{x})| \leq \delta, \delta \in \mathbb{R}. \quad (29)$$

现考虑来自函数簇 Γ 的成员函数 g , $\mathbf{x} \in \mathbb{R}^d$, 对应输入 $\mathbf{x}$ 上的测量输出 $\mathbf{Y} = \{y_1, y_2, \dots, y_N\}$, 即 $y_k = g(\mathbf{x}_k)$, $g \in \Gamma$, $\mathbf{x}_k \in \mathcal{S}$, $k = 1, 2, \dots, N$. 最优区间回归模型辨识的思想是, 在满足如下约束条件(30)的情况下, 辨识下边回归模型 $f_L(\mathbf{x}_k)$ 和上边回归模型 $f_U(\mathbf{x}_k)$:

$$f_L(\mathbf{x}_k) \leq g(\mathbf{x}_k) \leq f_U(\mathbf{x}_k), \forall \mathbf{x}_k \in \mathcal{S}. \quad (30)$$

在上式约束的意义下, 来自函数簇的任一成员函数总能在区间带 $[f_L(\mathbf{x}_k), f_U(\mathbf{x}_k)]$ 中找到. 这样的区间带有无穷多个, 本文的目的就是根据提出的约束, 确定尽可能窄的区间带. 该问题的研究, 文献[35]采用了连续分段线性函数的逼近方法来实现. 在本文, 是通过最小化最大逼近误差来给出回归模型更好的逼近, 或寻找一个更紧凑的逼近带. 基于式(11), 可给出上、下边回归模型的表达式如下:

$$f_L(\mathbf{x}, \beta_L, b_L) = \sum_{k=1}^N \beta_{L,k} \exp(\frac{-\|\mathbf{x} - \mathbf{x}_k\|^2}{2\sigma^2}) + b_L, \quad (31)$$

$$f_U(\mathbf{x}, \beta_U, b_U) = \sum_{k=1}^N \beta_{U,k} \exp(\frac{-\|\mathbf{x} - \mathbf{x}_k\|^2}{2\sigma^2}) + b_U. \quad (32)$$

基于引理1, 式(32)和式(31)所对应的上、下边回归模型 $f_U(\mathbf{x})$, $f_L(\mathbf{x})$ 可通过线性规划对如下优化问题进行求解.

$$\begin{aligned} \min_{f_L} \max_{\mathbf{x}_k \in \mathcal{S}} |y_k - f_L(\mathbf{x}_k)|, \\ \text{s.t. } y_k - f_L(\mathbf{x}_k) \geq 0, k = 1, 2, \dots, N, \end{aligned} \quad (33)$$

$$\begin{aligned} \min_{f_U} \max_{\mathbf{x}_k \in S} |y_k - f_U(\mathbf{x}_k)|, \\ \text{s.t. } y_k - f_U(\mathbf{x}_k) \leq 0, k = 1, 2, \dots, N. \end{aligned} \quad (34)$$

因此, 进一步可得如下定理.

定理 1 对于下边回归模型 $f_L(\mathbf{x})$ 的参数 β_L, b_L 的求解, 对应min-max优化问题(34)可通过最小化 λ_1 , 且满足如下不等式约束的线性规划求解, 即

$$\begin{aligned} \min: \lambda_1, \\ \text{s.t. } \begin{cases} y_k - \sum_{k=1}^N \beta_k^L \exp\left(\frac{-\|\mathbf{x} - \mathbf{x}_k\|}{2\sigma^2}\right)^2 - b_L \leq \lambda_1, \\ y_k - \sum_{k=1}^N \beta_k^L \exp\left(\frac{-\|\mathbf{x} - \mathbf{x}_k\|}{2\sigma^2}\right)^2 - b_L \geq 0, \\ \lambda_1 \geq 0, k = 1, 2, \dots, N, \end{cases} \end{aligned} \quad (35)$$

λ_1 表示对应下边模型 f_L 最大逼近误差.

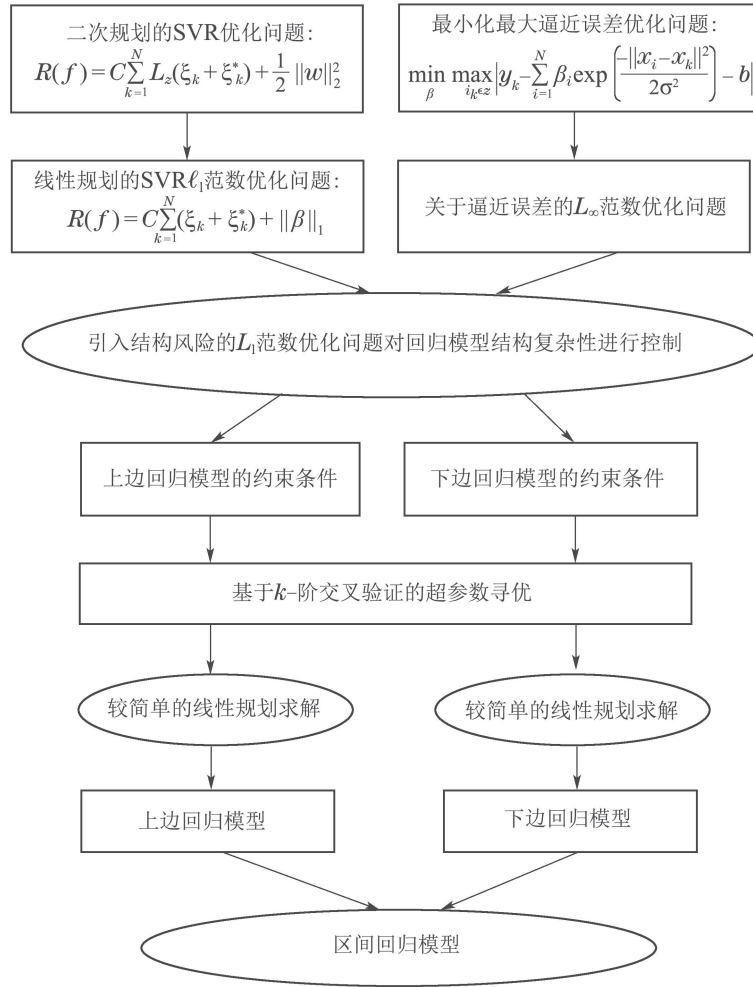


图 1 一种新的区间回归模型辨识方法流程图

Fig. 1 Flow chart of identifying interval regression model using proposed method

定理 2 对于上边回归模型 $f_U(\mathbf{x})$ 的参数 β_U, b_U 的求解, 对应min-max优化问题(34)可通过最小化 λ_2 , 且满足如下不等式约束的线性规划求解, 即

$$\begin{aligned} \min: \lambda_2, \\ \text{s.t. } \begin{cases} -y_k + \sum_{k=1}^N \beta_k^U \exp\left(\frac{-\|\mathbf{x} - \mathbf{x}_k\|}{2\sigma^2}\right)^2 + b_U \leq \lambda_2, \\ y_k - \sum_{k=1}^N \beta_k^U \exp\left(\frac{-\|\mathbf{x} - \mathbf{x}_k\|}{2\sigma^2}\right)^2 - b_U \leq 0, \\ \lambda_2 \geq 0, k = 1, 2, \dots, N, \end{cases} \end{aligned} \quad (36)$$

λ_2 表示对应上边模型 f_U 最大逼近误差.

证 上述定理1和定理2可直接通过引理1推出. 证毕.

从上述区间回归模型辨识的思想来看, 仅考虑如何让上、下边模型输出与实际输出之间的逼近误差最小, 而对回归模型本身的结构复杂性控制却没有被考虑, 这样一来, 通过上述优化问题获取的参数解有可能出现不全为零的情况, 不具有稀疏性, 对应 N 个样本数据可能对区间模型的建立都在起作用, 从而导致较复杂的模型结构. 为了解决模型稀疏解的问题, 在求解区间回归模型的优化问题中, 有必要将结构风险

最小化的思想融合其中, 在保证回归模型逼近精度的同时, 尽可能让模型结构复杂性得到有效控制. 基于此, 将区间回归模型所对应的上、下边回归模型优化

问题(35)和(36), 融合到基于结构风险最小化的优化问题(16). 因此, 基于式(15), 对应下边回归模型新的优化问题为

$$\begin{aligned} \min: R(f) &= \gamma \cdot \sum_{k=1}^N \xi_k + \sum_{k=1}^N (\alpha_{L,k}^+ + \alpha_{L,k}^-) + \lambda_1 + b_L, \\ \text{s.t.} \quad &\begin{cases} y_k - \sum_{k=1}^N (\alpha_{L,k}^+ - \alpha_{L,k}^-) \exp\left(\frac{-\|\mathbf{x} - \mathbf{x}_k\|}{2\sigma^2}\right)^2 - b_L \leq \varepsilon + \xi_k, \\ \sum_{k=1}^N (\alpha_{L,k}^+ - \alpha_{L,k}^-) \exp\left(\frac{-\|\mathbf{x} - \mathbf{x}_k\|}{2\sigma^2}\right)^2 + b_L - y_k \leq \varepsilon + \xi_k, \\ y_k - \sum_{k=1}^N (\alpha_{L,k}^+ - \alpha_{L,k}^-) \exp\left(\frac{-\|\mathbf{x} - \mathbf{x}_k\|}{2\sigma^2}\right)^2 - b_L \leq \lambda_1, \\ y_k - \sum_{k=1}^N (\alpha_{L,k}^+ - \alpha_{L,k}^-) \exp\left(\frac{-\|\mathbf{x} - \mathbf{x}_k\|}{2\sigma^2}\right)^2 - b_L \geq 0, \\ \xi_k \geq 0, \lambda_1 \geq 0, k = 1, 2, \dots, N, \end{cases} \end{aligned} \quad (37)$$

其中 λ_1 表示最大逼近误差, 参数 $\alpha_{L,k}^+, \alpha_{L,k}^-, b_L, \varepsilon, \xi_k$ 与第2节的定义一样.

对于上边回归模型 $f_U(\mathbf{x})$ 的优化问题, 有

$$\begin{aligned} \min: R(f) &= \gamma \cdot \sum_{k=1}^N \xi_k + \sum_{k=1}^N (\alpha_{U,k}^+ + \alpha_{U,k}^-) + \lambda_2 + b_U, \\ \text{s.t.} \quad &\begin{cases} y_k - \sum_{k=1}^N (\alpha_{U,k}^+ - \alpha_{U,k}^-) \exp\left(\frac{-\|\mathbf{x} - \mathbf{x}_k\|}{2\sigma^2}\right)^2 - b_U \leq \varepsilon + \xi_k, \\ \sum_{k=1}^N (\alpha_{U,k}^+ - \alpha_{U,k}^-) \exp\left(\frac{-\|\mathbf{x} - \mathbf{x}_k\|}{2\sigma^2}\right)^2 + b_U - y_k \leq \varepsilon + \xi_k, \\ -y_k + \sum_{k=1}^N (\alpha_{U,k}^+ - \alpha_{U,k}^-) \exp\left(\frac{-\|\mathbf{x} - \mathbf{x}_k\|}{2\sigma^2}\right)^2 + b_U \leq \lambda_2, \\ y_k - \sum_{k=1}^N (\alpha_{U,k}^+ - \alpha_{U,k}^-) \exp\left(\frac{-\|\mathbf{x} - \mathbf{x}_k\|}{2\sigma^2}\right)^2 - b_U \leq 0, \\ \xi_k \geq 0, \lambda_2 \geq 0, k = 1, 2, \dots, N. \end{cases} \end{aligned} \quad (38)$$

相似地, λ_2 表示最大逼近误差, 参数 $\alpha_{U,k}^+, \alpha_{U,k}^-, b_U, \varepsilon, \xi_k$ 与第2节的定义一样. 从优化问题(37)和(38)可知, 为典型的线性规划问题, 可用向量及矩阵形式表述如下:

$$\begin{aligned} \min \quad &\mathbf{c}^T(\boldsymbol{\alpha}_L^+, \boldsymbol{\alpha}_L^-, \boldsymbol{\xi}, \lambda_2, b_L)^T, \\ \text{s.t.} \quad &\begin{cases} \begin{pmatrix} -\mathbf{K} & \mathbf{K} & -\mathbf{I} & \mathbf{0}_{N \times 1} & -\mathbf{E}_{N \times 1} \\ \mathbf{K} & -\mathbf{K} & -\mathbf{I} & \mathbf{0}_{N \times 1} & \mathbf{E}_{N \times 1} \\ -\mathbf{K} & \mathbf{K} & \mathbf{Z} & -\mathbf{E}_{N \times 1} & -\mathbf{E}_{N \times 1} \\ \mathbf{K} & -\mathbf{K} & \mathbf{Z} & \mathbf{0}_{N \times 1} & \mathbf{E}_{N \times 1} \end{pmatrix} \\ \begin{pmatrix} \alpha_L^+ \\ \alpha_L^- \\ \boldsymbol{\xi} \\ \lambda_2 \\ b_L \end{pmatrix} \leq \begin{pmatrix} -\mathbf{y} + \varepsilon \\ \varepsilon + \mathbf{y} \\ -\mathbf{y} \\ \mathbf{y} \end{pmatrix}, \\ \alpha_L^+, \alpha_L^- \geq 0, \boldsymbol{\xi} \geq 0, 0 \leq \lambda_2 \leq 1, \end{cases} \end{aligned} \quad (39)$$

以及

$$\begin{aligned} \min \quad &\mathbf{c}^T(\boldsymbol{\alpha}_U^+, \boldsymbol{\alpha}_U^-, \boldsymbol{\xi}, \lambda_1, b_U)^T, \\ \text{s.t.} \quad &\begin{cases} \begin{pmatrix} -\mathbf{K} & \mathbf{K} & -\mathbf{I} & \mathbf{0}_{N \times 1} & -\mathbf{E}_{N \times 1} \\ \mathbf{K} & -\mathbf{K} & -\mathbf{I} & \mathbf{0}_{N \times 1} & \mathbf{E}_{N \times 1} \\ \mathbf{K} & -\mathbf{K} & \mathbf{Z} & -\mathbf{E}_{N \times 1} & \mathbf{E}_{N \times 1} \\ -\mathbf{K} & \mathbf{K} & \mathbf{Z} & \mathbf{0}_{N \times 1} & -\mathbf{E}_{N \times 1} \end{pmatrix} \\ \begin{pmatrix} \alpha_U^+ \\ \alpha_U^- \\ \boldsymbol{\xi} \\ \lambda_1 \\ b_U \end{pmatrix} \leq \begin{pmatrix} -\mathbf{y} + \varepsilon \\ \varepsilon + \mathbf{y} \\ \mathbf{y} \\ -\mathbf{y} \end{pmatrix}, \\ \alpha_U^+, \alpha_U^- \geq 0, \boldsymbol{\xi} \geq 0, 0 \leq \lambda_1 \leq 1, \end{cases} \end{aligned} \quad (40)$$

其中: $\mathbf{I}$ 为 $N \times N$ 单位矩阵, $\mathbf{Z}$ 表示 N 阶零矩阵, $\mathbf{E}$ 和 $\mathbf{0}$ 是元素分别为1和0的列向量,

$$\mathbf{c} = (\underbrace{1 \cdots 1}_N \underbrace{1 \cdots 1 \ 1 \ 1 \cdots 1}_N \underbrace{\gamma \cdots \gamma}_N \ 1 \ 1)^T,$$

$$\mathbf{y} = (y_1 \ y_2 \ \cdots \ y_N),$$

$$\alpha_U^+ = (\alpha_{U,1}^+ \ \alpha_{U,2}^+ \ \cdots \ \alpha_{U,N}^+)^T,$$

$$\alpha_U^- = (\alpha_{U,1}^- \ \alpha_{U,2}^- \ \cdots \ \alpha_{U,N}^-)^T,$$

$$\xi = (\xi_1 \ \xi_2 \ \cdots \ \xi_N)^T, \mathbf{Z} = 0_{N \times N},$$

核矩阵 $\mathbf{K}$ 的元素定义为

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

σ 为可调核参数. 显然, 应用内点法或单纯性方法可以求解优化问题(40)和(39), 进而得到区间回归模型所对应的上、下边回归模型 $f_U(\mathbf{x})$ 和 $f_L(\mathbf{x})$:

$$f_U(\mathbf{x}) = \sum_{k=1}^N (\alpha_{U,k}^+ - \alpha_{U,k}^-) \exp\left(\frac{-\|\mathbf{x} - \mathbf{x}_k\|}{2\sigma^2}\right)^2 + b_U, \quad (41)$$

$$f_L(\mathbf{x}) = \sum_{k=1}^N (\alpha_{L,k}^+ - \alpha_{L,k}^-) \exp\left(\frac{-\|\mathbf{x} - \mathbf{x}_k\|}{2\sigma^2}\right)^2 + b_L. \quad (42)$$

从提出的优化问题中来建立 $f_U(\mathbf{x})$ 和 $f_L(\mathbf{x})$ 的整个过程来看, 优化问题既包括了对模型结构复杂性控制的目标函数, 又包括了如何获取较好的模型精度所对应的逼近误差作为目标函数, 而且模型结构复杂性和模型精度之间的权衡可以通过超参数(包括核宽度、正则化参数以及不敏感域)进行调整. 如果超参数选取不当, 会极大影响所建立区间模型的泛化性能. 当前, 关于超参数寻优的方法包括元启发式算法^[36]、 k -阶交叉验证法^[37]、正弦-余弦算法(sine cosine algorithm SCA)^[38]、遗传算法(genetic algorithms)^[39]、粒子群优化算法(partical swarm optimization, PSO)^[40]等. 然而, 在众多参数寻优方法中, k -阶交叉验证法可以较好的避免模型过拟合问题^[37], 在理论研究和实际应用中得到了广泛应用. 总而言之, 提出方法在保证获取区间模型尽可能窄的区间带的同时, 而且还对模型结构复杂性进行有效控制, 从而提高区间回归模型的泛化性能^[41]. 因此, k -阶交叉验证法将用于本文的参数寻优, 其原理为: 1) 首先将整个训练集划分为 k 个互斥且大小相同的子集, 其中一个作为测试集, 剩下的 $k-1$ 个作为训练集; 2) 将被寻优的参数对(核宽度 σ 、正则化参数 γ)以递增方式设定, 例如 $\sigma = 2^{-15}, 2^{-13}, \dots, 2^3$, $\gamma = 2^{-5}, 2^{-3}, \dots, 2^{15}$; 3) 再用提出的方法以及参数对分别建立上、下边回归模型 $f_U(\mathbf{x})$ 和 $f_L(\mathbf{x})$. 这样一来, 训练集中的每个样本都将被用于模型的一次预测得到相应的预测值, 并与真实值比较形成交叉验证精度; 4) 将最好的交叉验证精度所对应的参数对作为区间模型的最优参数选择.

所建立的区间回归模型可以用于系统的故障检测, 例如当系统运行在健康状态时, 获取所对应的无故障数据, 应用提出的方法对无故障数据进行建模, 得到相应的区间回归模型, 从而获取系统健康状态下的区间带. 在系统运行期间, 检测某个测量参数输出是否

包含于区间带, 若实际输出在对应的区间带内, 则判断系统运行正常, 否则故障发生; 此外提出的区间回归模型也可用于鲁棒控制设计、信息压缩等方面.

5 实验研究

这部分内容将通过如下实验分析, 论证所提出方法的最优性与稀疏性; 同时为了更直观的去评判提出的方法, 将考虑如下两个性能指标, 即均方根误差(root mean square error, RMSE)和支持向量占整个样本数据的百分比SVs%. 对于RMSE 指标定义如下:

$$\text{RMSE} = \sqrt{\frac{\sum_{k=1}^N (y_k - \hat{y}_k)^2}{N}}, \quad (43)$$

N 表示测试数据的总数, y_k 为实际输出, $\hat{y}_k$ 是模型的被估输出. RMSE反映了用提出方法所建立区间回归模型在满足各自约束条件下, 即 $f_U(\mathbf{x}) - y \geq 0$ 以及 $f_L(\mathbf{x}) - y \leq 0$, 模型输出与实际测量数据之间的逼近程度; RMSE越小, 逼近程度越好, 反之越差. 此外, 对应优化问题(39)–(40)的求解, 若有 $\alpha_k^+ - \alpha_k^- \neq 0$ 或 $\alpha_k^+ - \alpha_k^- \geq 1 \times 10^{-8}$, 则对应第 k 个样本数据为支持向量, 假设通过该条件判断获取 N_k 个支持向量, 则支持向量(SV)占整个样本数据的百分比SVs%可定义如下:

$$\text{SVs\%} = \frac{N_k}{N} \times 100\%. \quad (44)$$

显然, 在保证模型精度的同时, 指标SVs%越小越好, 越小则表示求解的区间模型有稀疏解, 模型结构简单, 具有较好的泛化性能.

下面将对提出的方法从区间模型的辨识精度以及稀疏特性展开实验分析, 论证其合理性与优越性, 包括不受噪声干扰、受噪声干扰的不确定性测量数据以及参数的不确定性. 其中 $\text{sinc}(x) = \frac{\sin x}{x}$ 是支持向量回归理论^[30]产生以来, 以及用于论证其他方法^[42]最常采用的仿真.

首先考虑 $\text{sinc}(x)$ 无噪声情况下的分析, 在区间 $[-10, 10]$ 上获取100个等间隔无噪声样本. 采用5-阶交叉验证法^[37]获取参数集 $(\epsilon, \gamma, \sigma)$ 为 $(0.0001, 1000, 2.8)$, 再通过提出的方法建立最优区间回归模型如图2所示, 图3给出了URM $f_U(\mathbf{x})$ 和LRM $f_L(\mathbf{x})$ 的逼近误差. 从图2可知, 区间模型与无噪声的实际输出基本上重合, $f_U(\mathbf{x})$ 与 $f_L(\mathbf{x})$ 无限逼近实际输出. 与其他方法相比较, 提出的方法分别需要12个支持向量(SV)建立区间模型的 $f_U(\mathbf{x})$ 和 $f_L(\mathbf{x})$, 模型预测输出与实际输出之间的最大误差为0.0057和0.0108; 文献[34]用了39个SV, 最大误差0.01; 文献[42]用了9个SV, 最大误差为0.0087. 因此, 无论从模型稀疏特性还是建模精度来看, 提出的方法具有一定的优越性. 此外, 当被建模对象无噪声干扰时, 提出的最优区间回归模型变成了传统的确定性建模.

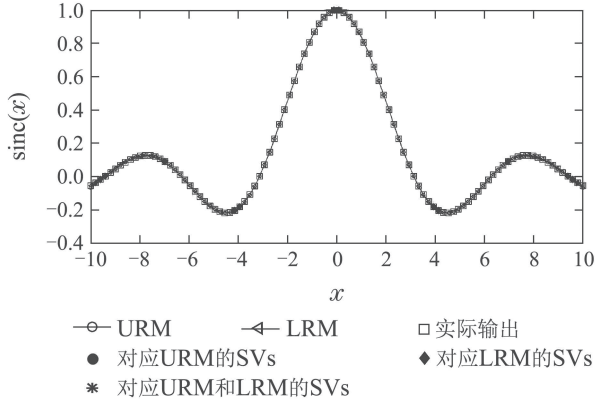


图2 提出方法的区间输出

Fig. 2 The output of IRM

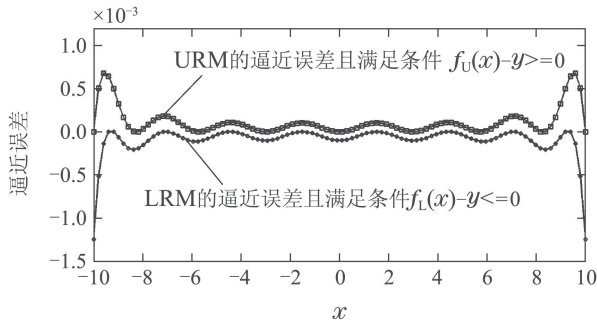


图3 URM和LRM的逼近误差

Fig. 3 Approximation error with noise-free for URM and LRM

接下来,考虑 $\text{sinc}(x)$ 受噪声干扰的情况.类似地,在区间 $[-10, 10]$ 上获取100个带均匀分布噪声为 $[-0.2, 0.2]$ 的等间隔噪声样本.应用5-阶交叉验证法^[37]获取设超参数集 $(\epsilon, \gamma, \sigma)$ 为 $(0.01, 100, 2.5)$,则 $\text{sinc}(x)$ 的区间输出逼近如图4所示,URM和LRM分别从100个数据中,仅仅用了其中的10和9个数据建立回归模型,这些数据被称SV,对应的SVs%则为10%和9%,反映了提出方法所建立区间模型的稀疏特性;RMSE为0.1291和0.1238反映了建立区间模型的精度.

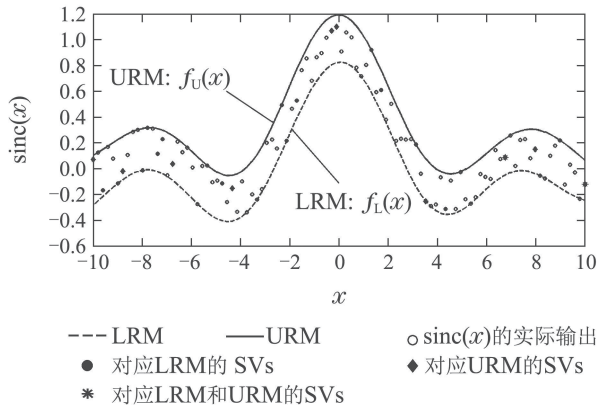


图4 提出方法对100个噪声样本建立的区间输出

Fig. 4 The output of IRM corresponding to 100 noise samples using the proposed method

从建模精度指标RMSE的数值来看,尽管所反映的区间带变宽,紧凑性变差,但模型具有较好的稀疏特性.因此,建模精度与模型稀疏特性是一对矛盾体,根据建模的需要,从两者之间取其平衡.继续对 $\text{sinc}(x)$ 在相同噪声作用下,产生1000个样本建立区间模型,仍然采用超参数集 $(\epsilon, \gamma, \sigma)$ 为 $(0.01, 100, 2.5)$,图5给出了提出方法的区间输出,在保证模型精度条件下,稀疏特性也进一步得到保证,例如URM和LRM分别从1000个数据中,仅用了13和12个数据,对应的SVs%为1.3%和1.2%,RMSE为0.2264和0.2301.图6给出了IRM的逼近误差,其中 $f_U(x_k) - y_k \geq 0$ 以及 $f_L(x_k) - y_k \leq 0$.

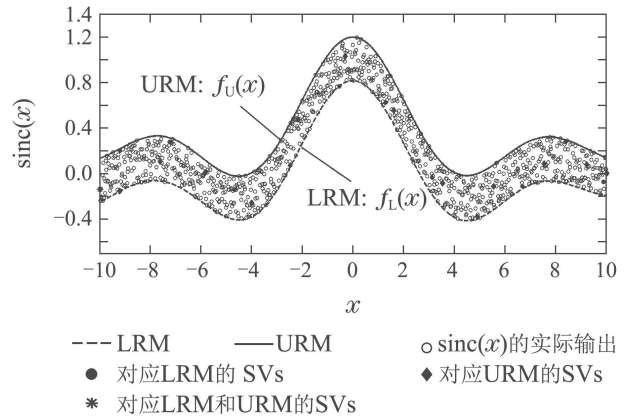


图5 提出方法对1000个噪声样本建立的区间输出

Fig. 5 The output of IRM corresponding to 1000 noise samples using the proposed method

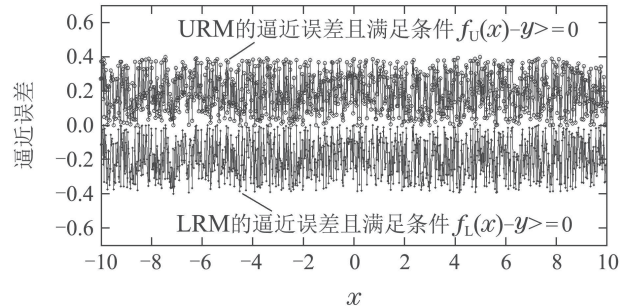


图6 区间模型的逼近误差

Fig. 6 Approximation error with noise-free for URM and LRM

为了进一步论证提出方法的稀疏特性,在实验仿真过程中,事先设定大于0的较小值 $\eta = 1.0 \times 10^8$,判断回归模型的展开项中系数 $(\alpha_k^+ - \alpha_k^-) \geq \eta$, $k = 1, 2, \dots, N$ 是否成立,如果满足该条件,说明展开项中的第 k 项对回归模型起作用,对应第 k 个样本 x_k 为支持向量,如图7-8所示,分别对应URM和LRM中 $(\alpha_k^+ - \alpha_k^-) \geq \eta$.

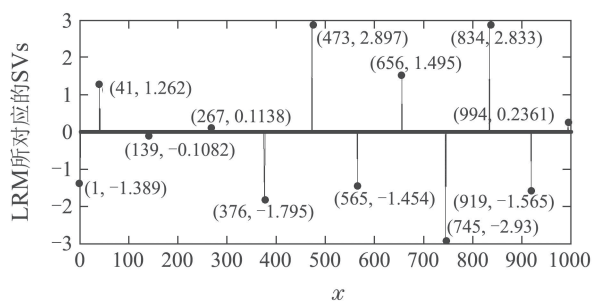


图7 对应LRM展开项系数 $\{|\alpha_k^+ - \alpha_k^-| \geq \delta\}_{k=1}^N$

Fig. 7 The coefficient of expansion term for LRM corresponding to $\{|\alpha_k^+ - \alpha_k^-| \geq \delta\}_{k=1}^N$ (where δ is selected as $1.0E-08$ in computational simulation)

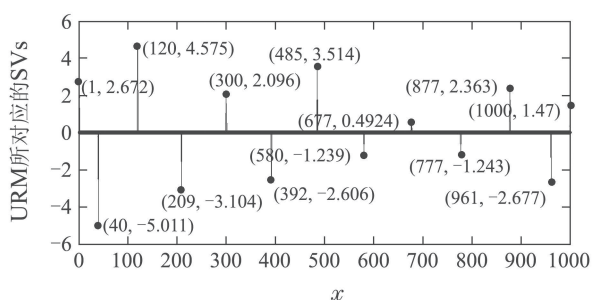


图8 对应URM展开项系数 $\{|\alpha_k^+ - \alpha_k^-| \geq \delta\}_{k=1}^N$

Fig. 8 The coefficient of expansion term for URM corresponding to $\{|\alpha_k^+ - \alpha_k^-| \geq \delta\}_{k=1}^N$ (where δ is selected as $1.0E-08$ in computational simulation)

继上述来自测量数据的不确定性实验分析后, 下面考虑来自模型参数不确定性的最优区间模型辨识. 假设带有不确定性参数的非线性函数类 Γ , 其成员函数为 $f(x)$, 由名义函数 $f_{\text{nom}}(x)$ 和不确定性 $\Delta f(x)$ 构成, 即

$$f(x) = f_{\text{nom}}(x) + \Delta f(x), \quad (45)$$

$$f_{\text{nom}}(x) = \cos x \cdot \sin x, \quad (46)$$

$$\Delta f(x) = \tau \cos(8x), \quad 0 \leq \tau \leq 1.$$

设该函数类的定义域为 $-1 \leq x \leq 1$, 为获取建立模型所需要的样本数据, 不妨取 $x_k = 0.021k$, $k = -47, -46, \dots, 46, 47$, 图9给出了由不确定性参数 τ 引起的不确定性输出. 接下来用提出的区间回归模型包络由不确定性参数 τ 值所引起的测量输出. 应用5-阶交叉验证获取超参数集 $(\varepsilon, \gamma, \sigma)$ 为 $(0.001, 100, 3.5)$, 由上边、下边模型构成的区间回归模型如图10所示, 区间模型的曲线平坦度变化相对较大, 分别对应URM与LRM的SVs%都为45.26%, RMSE为0.0037和0.0036, 辨识精度进一步提高, 但区间模型趋于复杂, 其中点线属于 Γ ; 若选取 $(\varepsilon, \gamma, \sigma)$ 为 $(0.001, 100,$

4.5), 如图11所示, $f_U(x)$ 和 $f_L(x)$ 的RMSE分别为0.0073和0.0074, 反映了提出方法的辨识精度较图10模型结构相对简单, 此时URM与LRM从96个数据中, 仅用了21个数据(支持向量)建立区间模型的上、下边界(对应SVs%都为22.11%), 稀疏特性较好, 曲线较平坦, 实线分别对应上边回归模型逼近 $f_U(x)$ 以及下边回归模型逼近 $f_L(x)$. 换言之, 图11所对应区间模型的稀疏特性比图10好, 但辨识精度却比图10差. 此外, 图12和图13给出了在满足各自约束条件的 $f_U(x)$ 和 $f_L(x)$ 的逼近误差曲线. 为了更清晰的论证提出方法在辨识精度和稀疏特性之间的平衡, 表1和图14在不同核参数情况下的两个指标RMSE和SVs%的变化, 分别反映了模型辨识精度及模型的稀疏特性.

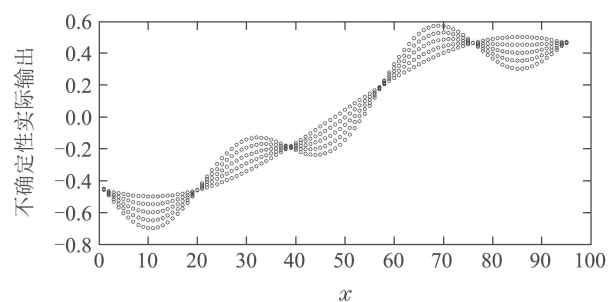


图9 由参数不确定性引起的非线性函数簇输出

Fig. 9 Actual of the family of nonlinear function caused by the uncertain parameter τ

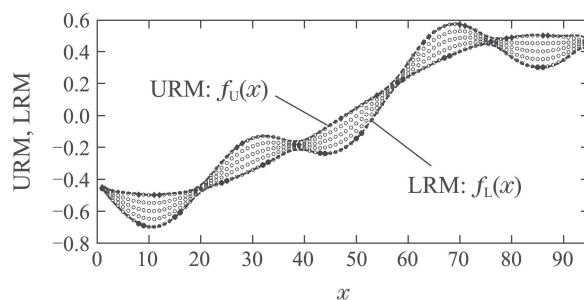


图10 当核宽度选择为3.5时所对应的的区间输出

Fig. 10 The output of IRM when the kernel width is selected as 3.5

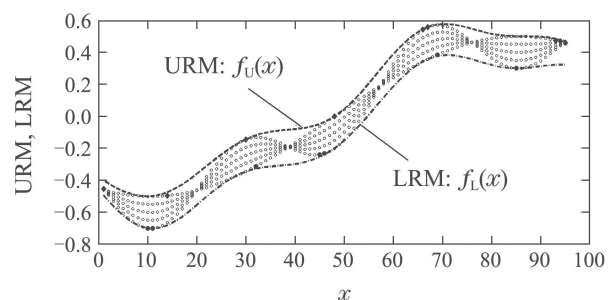


图11 当核宽度选择为4.5时所对应的的区间输出

Fig. 11 The output of IRM when the kernel width is selected as 4.5

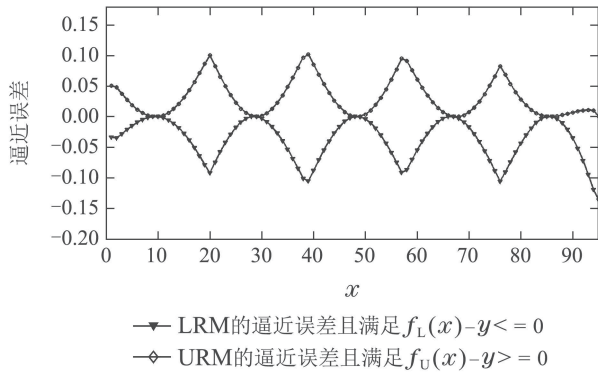


图12 区间模型的逼近误差

Fig. 12 Approximation error for URM and LRM

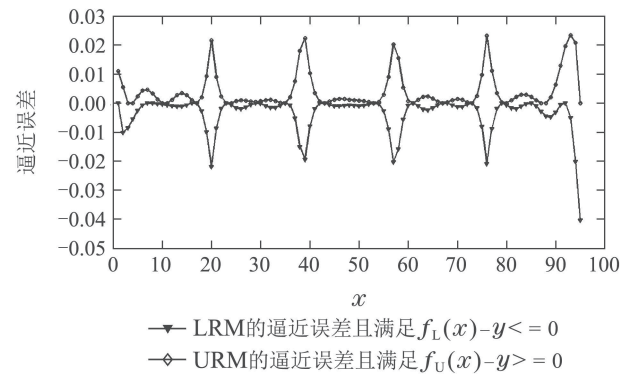


图13 区间模型的逼近误差

Fig. 13 Approximation error for URM and LRM

表1 不同核宽度 σ 的比较结果Table 1 Comparison results of the selected different kernel width σ for URM and LRM

核宽度 σ	RMSE		SVs/%	
	URM	LRM	URM	LRM
0.5	0.7035E-03	0.6649E-03	0.9895	1.000
1	0.7043E-03	0.6597E-03	0.9895	0.9895
1.5	0.0026	0.0027	0.7368	0.7368
2	0.0034	0.0036	0.5474	0.5684
2.5	0.0037	0.0036	0.4526	0.4526
3	0.0046	0.0046	0.3684	0.3684
3.5	0.0055	0.0059	0.3053	0.3053
4	0.0060	0.0063	0.2737	0.2842
4.5	0.0073	0.0074	0.2211	0.2211
5	0.0090	0.0108	0.2105	0.1895
5.5	0.0100	0.0117	0.1895	0.1895
6	0.0104	0.0118	0.1684	0.1684

核宽度 σ	RMSE		SVs/%	
	URM	LRM	URM	LRM
6.5	0.0111	0.0118	0.1579	0.1474
7	0.0123	0.0128	0.1368	0.1474
7.5	0.0140	0.0132	0.1474	0.1368
8	0.0158	0.0143	0.1263	0.1263
8.5	0.0207	0.0180	0.1263	0.1474
9	0.0274	0.0242	0.1263	0.1158
9.5	0.0342	0.0359	0.1053	0.1053
10	0.0376	0.0435	0.0947	0.0947
10.5	0.0401	0.0473	0.0842	0.1053
11	0.0420	0.0498	0.0947	0.0842
11.5	0.0430	0.0501	0.0842	0.0842

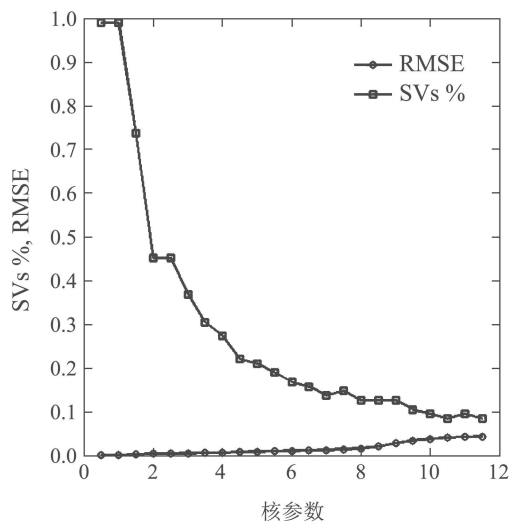
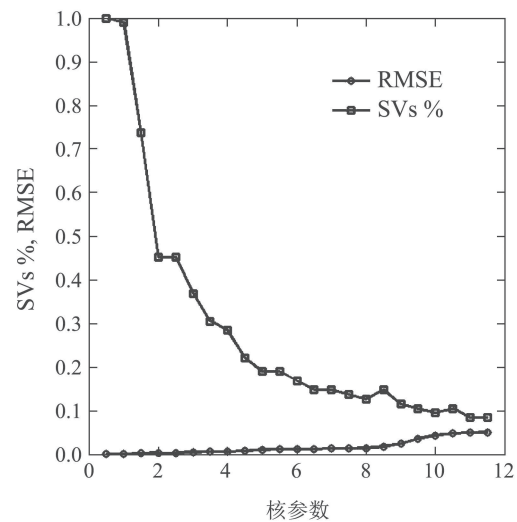
(a) URM: $f_U(x)$ (b) LRM: $f_L(x)$

图14 RMSE与SVs%在不同核宽度下的比较结果

Fig. 14 Comparison results both RMSE and SVs% with the different σ is shown for URM (left) and LRM (right)

进一步,将提出的方法与文献[15]从两个指标RMSE, SVs%进行分析比较. 根据文献[15]定义如下回归估计量 $\hat{f}(x)$:

$$\hat{f}(x) = \frac{1}{2}(f_U(x) + f_L(x)), \quad (47)$$

则有回归估计的均方根误差为

$$\text{RMSE} = \sqrt{\frac{\sum_{k=1}^N (f(x) - \hat{f}(x))^2}{N}}, \quad (48)$$

N 表示测试数据的总数, $f(x)$ 为实际输出, $\hat{f}(x)$ 是模型的被估输出. 值得注意的是, 该式的定义与式(43)是不一样的, 式(43)描述的是 $f_U(x)$ 和 $f_L(x)$ 与不确定性数据之间的RMSE, 而式(48)描述的是回归估计量 $\hat{f}(x)$ 的RMSE. 在文献[15]中, 下面的非线性系统仿真被考虑:

$$\begin{cases} y_k = 0.2 \cdot \sin(2\pi x_k) + 0.2 \cdot x_k^2 + 0.3 + \\ (0.1 \cdot x_k^2 + 0.05) \cdot \text{noise}, \\ x_k = 0.02 \cdot (k-1), k = 1, 2, \dots, 51, \end{cases} \quad (49)$$

其中noise表示在区间 $[-1, 1]$ 上随机产生的实数. 从式(48)分别获取训练数据和测试数据, 当应用5-阶交叉验证获取超参数 $(\varepsilon, \gamma, \sigma)$ 为 $(0.001, 10, 0.5)$ 时, 应用提出的方法所建立的区间模型以及对应的上、下边回归模型的逼近误差分别如图15-16所示, 上、下边回归模型从51个数据中都用了4个SV. 对于 $f_U(x)$ (上边回归模型), 对应的第 k 个SV以及 $\alpha_{U,k}^+ - \alpha_{U,k}^-$ 值如表2所示, 其表达式为

$$\begin{aligned} f_U(x) = & -3.4183 \times \exp\left(\frac{-\|x - 0\|}{2 \times 0.5^2}\right)^2 + \\ & 6.1808 \times \exp\left(\frac{-\|x - 0.26\|}{2 \times 0.5^2}\right)^2 - \\ & 5.7676 \times \exp\left(\frac{-\|x - 0.7\|}{2 \times 0.5^2}\right)^2 + \\ & 3.8100 \times \exp\left(\frac{-\|x - 1\|}{2 \times 0.5^2}\right)^2. \end{aligned} \quad (50)$$

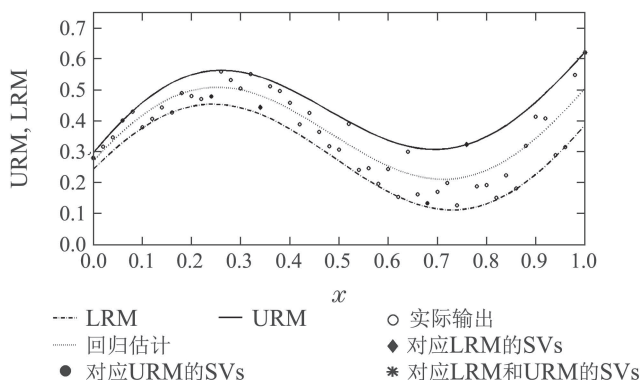


图 15 当核宽度选择为0.5时所对应的区间输出

Fig. 15 The output of IRM when the kernel width is selected as 0.5

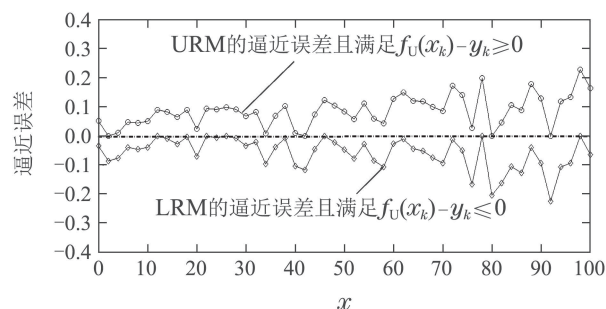


图 16 当核宽度选择为0.5时所对应的区间输出逼近误差
Fig. 16 The approximation error of IRM when the kernel width is selected as 0.5

表 2 $f_U(x)$ 所对应的第 k 个SV以及 $\alpha_{U,k}^+ - \alpha_{U,k}^-$ 值
Table 2 The k th SV with $\alpha_{U,k}^+ - \alpha_{U,k}^-$ 值

k	1	14	36	51
x_k	0	0.26	0.7	1
$\alpha_{U,k}^+ - \alpha_{U,k}^-$	-3.4183	6.1808	-5.7676	3.8100

同理, 对于 $f_L(x)$ (下边回归模型), 对应的第 k 个SV以及 $\alpha_{L,k}^+ - \alpha_{L,k}^-$ 值如表3所示, 其表达式为

$$\begin{aligned} f_L(x) = & -3.2228 \times \exp\left(\frac{-\|x - 0\|}{2 \times 0.5^2}\right)^2 + \\ & 6.3270 \times \exp\left(\frac{-\|x - 0.30\|}{2 \times 0.5^2}\right)^2 - \\ & 7.4278 \times \exp\left(\frac{-\|x - 0.74\|}{2 \times 0.5^2}\right)^2 + \\ & 4.9534 \times \exp\left(\frac{-\|x - 1\|}{2 \times 0.5^2}\right)^2. \end{aligned} \quad (51)$$

表 3 $f_L(x)$ 所对应的第 k 个SV以及 $\alpha_{L,k}^+ - \alpha_{L,k}^-$ 值
Table 3 The k th SV with $\alpha_{L,k}^+ - \alpha_{L,k}^-$ 值

k	1	16	38	51
x_k	0	0.30	0.7400	1
$\alpha_{L,k}^+ - \alpha_{L,k}^-$	-3.2228	6.3270	-7.4278	4.9534

当超参数 $(\varepsilon, \gamma, \sigma)$ 的选取与文献[15]一样时, 即 $(100, 0.03, 0.18)$, 所对应的RMSE和SVs%如表4所示, 无论是从训练还是测试RMSE, 提出方法的建模精度较文献[15]的好, 而且模型结构也较稀疏, 对应SVs%为15.68%.

表 4 本文提出方法与文献[15]的比较

Table 4 Comparison results of the proposed method and [15]

方法	RMSE		SVs/%
	训练	测试	
ν -SVIRN ^[15]	0.1016	0.0116	29.41
SVIRM ^[15]	0.1032	0.0128	98.04
本文提出的方法	0.0574	0.0081	15.68

因此,通过上述实验对提出的两个指标,即SVs%以及RMSE进行分析可知,提出的方法可以:1)很好的建立来自不确定性测量数据或系统模型参数不确定性的区间模型,采用区间输出代替传统的点输出建模方法,更易于实际问题的研究;2)很好的解决了模型结构复杂性与辨识精度之间的平衡,在提高模型泛化性能的同时,模型辨识精度也能得到保证。

6 结论

当获取一组由测量或参数不确定性等因素导致的测量数据集时,提出了一种新的区间回归模型辨识方法。该方法将结构风险最小化理论与逼近误差的 L_∞ 范数最优化思想相结合,给出了在保证被辨识模型精度条件下,对模型结构复杂性进行有效控制的目标优化问题,同时将复杂的凸二次规划求解转化为较简单的线性规划问题求解。克服了传统方法在辨识非线性系统过程中,目标优化问题仅考虑模型输出与实际输出之间的偏差达到最小,致使模型结构复杂。提出的区间回归模型由包含所有测量数据的上、下边回归模型构成,可应用于被观测系统的参数在某个区间变化时的函数簇建模、也可用于数据挖掘中的信息压缩、鲁棒控制系统辨识以及故障检测。

参考文献:

- [1] KUKOLJ D, LEVI E. Identification of complex systems based on neural and Takagi-Sugeno fuzzy model. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2004, 34(1): 272 – 282.
- [2] LI Jinna, GAO Xize, CHAI Tianyou, et al. Data-driven operational optimization control of industrial processes. *Control Theory & Applications*, 2016, 33(12): 1584 – 1592.
(李金娜, 高溪泽, 柴天佑, 等. 数据驱动的工业过程运行优化控制. 控制理论与应用, 2016, 33(12): 1584 – 1592.)
- [3] HABBI H, KIDOUCE M, ZELMAT M. Data-driven fuzzy models for nonlinear identification of a complex heat exchanger. *Applied Mathematical Modelling*, 2011, 35(3): 1470 – 1482.
- [4] YANG Lingling, ZHANG Yun, CHEN Zhenfeng. Bilinear control based on model bias separation for uncertain nonlinear systems. *Acta Automatica Sinica*, 2010, 36(10): 1432 – 1442.
(杨玲玲, 章云, 陈贞丰. 不确定非线性系统基于偏差分离的双线性控制. 自动化学报, 2010, 36(10): 1432 – 1442.)
- [5] MASSIMILIANO L, MAURIZIO P, ANTONELLO R, et al. A novel neural networks ensemble approach for modeling electrochemical cells. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, 30(2): 1 – 12.
- [6] LÜ Z, ZHAO J, ZHAI Y W, et al. Non-iterative T-S fuzzy modeling with random hidden-layer structure for BFG pipeline pressure prediction. *Control Engineering Practice*, 2018, 76: 96 – 103.
- [7] ZHOU X J, JIANG T. An effective way to integrate ε -support vector regression with gradients. *Expert Systems with Applications*, 2018, 99: 126 – 140.
- [8] LIU Qin. Operator-based robust nonlinear perfect tracking control to a manipulator with unknown uncertainties. *Journal of Xinyang Normal University (Natural Science Edition)*, 2014, 27(4): 577 – 580.
(刘琴. 基于演算子理论的含有未知不确定性机械臂的鲁棒非线性精确跟踪控制. 信阳师范学院学报(自然科学版), 2014, 27(4): 577 – 580.)
- [9] HONG D H, HWANG C. Interval regression analysis using quadratic loss support vector machine. *IEEE Transactions on Fuzzy Systems*, 2005, 13(2): 229 – 237.
- [10] TANAKA H, KOYAMA K, LEE H. Interval regression analysis based on quadratic programming. *Proceedings of the Fifth IEEE International Conference on Fuzzy Systems*. New Orleans, USA: IEEE, 1996: 325 – 329.
- [11] TANAKA H, LEE H. Interval regression analysis by quadratic programming approach. *IEEE Transactions on Fuzzy Systems*, 1998, 6(4): 473 – 481.
- [12] LEE H, TANAKA H. Upper and lower approximation models in interval regression using regression quantile techniques. *European Journal of Operational Research*, 1999, 116(3): 653 – 666.
- [13] HUANG L, ZHANG B L, HUANG Q. Robust interval regression analysis using neural networks. *Fuzzy Sets and Systems*, 1998, 97(3): 337 – 347.
- [14] LIU Ying, LIU Ruoyu, ZHAO Jun, et al. Variational inference-based echo state network ensemble model for intervals prediction of industrial data. *Control Theory & Applications*, 2018, 35(8): 1066 – 1073.
(刘颖, 刘若愚, 赵珺, 等. 基于变分推理回声状态网络集成模型的工业数据区间预测. 控制理论与应用, 2018, 35(8): 1066 – 1073.)
- [15] HAO P Y. Interval regression analysis using support vector networks. *Fuzzy Sets and Systems*, 2009, 160(17): 2466 – 2485.
- [16] SHI Lei, HOU Liping. Parameter optimization of fuzzy support vector classifiers based on the improved PSO. *Journal of Xinyang Normal University (Natural Science Edition)*, 2013, 26(2): 288 – 291.
(石磊, 侯丽萍. 基于改进PSO算法参数优化的模糊支持向量分类机. 信阳师范学院学报(自然科学版), 2013, 26(2): 288 – 291.)
- [17] HUANG C H. A reduced support vector machine approach for interval regression analysis. *Information Sciences*, 2012, 217: 56 – 64.
- [18] LIM C. Interval-valued data regression using nonparametric additive models. *Journal of the Korean Statistical Society*, 2016, 45(3): 358 – 370.
- [19] SOUZA L C, SOUZA R M C R, AMARAL G J A, et al. A parametrized approach for linear regression of interval data. *Knowledge-Based Systems*, 2017, 131: 149 – 159.
- [20] NETO E A L, DE CARVALHO F A T. An exponential-type kernel robust regression model for interval-valued variables. *Information Sciences*, 2018, 454: 419 – 442.
- [21] SOARES Y M G, FAGUNDES R A A. Interval quantile regression models based on swarm intelligence. *Applied Soft Computing*, 2018, 72: 474 – 485.
- [22] PEDROSO J P, MURATA N. Support vector machines with different norms: motivation, formulations and results. *Pattern Recognition Letters*, 2001, 22(12): 1263 – 1272.
- [23] BASAK D, PAL S, PATRANABIS D C. Support vector regression. *Neural Information Processing-Letters and Reviews*, 2007, 11(10): 203 – 224.
- [24] SMOLA A J, SCHÖLKOPF B. A tutorial on support vector regression. *Statistics and Computing*, 2004, 14(3): 199 – 222.
- [25] KEMAL U, GULAY O G. Model free adaptive support vector regressor controller for nonlinear systems. *Engineering Applications of Artificial Intelligence*, 2019, 81: 47 – 67.
- [26] ZHANG L, ZHOU W D, CHANG P C, et al. Iterated time series prediction with multiple support vector regression models. *Neurocomputing*, 2014, 99: 411 – 422.
- [27] HONG D H, HWANG C. Interval regression analysis using quadratic loss support vector machine. *IEEE Transactions on Fuzzy System*, 2005, 13(2): 229 – 237.
- [28] CHUANG C C. Fuzzy weighted support vector regression with a fuzzy partition. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2007, 37(3): 630 – 640.

- [29] LU Z, SUN J, BUTTS K. Linear programming SVM-ARMA_{2K} with application in engine system identification. *IEEE Transactions on Automation Science and Engineering*, 2011, 8(4): 846 – 854.
- [30] VAPNIK V, GOLOWICH S E, SMOLA A J. Support vector method for function approximation, regression estimation and signal processing. *Advances in Neural Information Processing Systems*, 1997, 9: 281 – 287.
- [31] RIVAS-PEREA P, COTA-RUIZ J. An algorithm for training a large scale support vector machine for regression based on linear programming and decomposition methods. *Pattern Recognition Letters*, 2013, 34(4): 439 – 451.
- [32] ZHAO Y P, SUN J G. Multikernel semiparametric linear programming support vector regression. *Expert Systems with Applications*, 2011, 38(3): 1611 – 1618.
- [33] SMOLA A, SCHOLKOPF B, RATSCH G. Linear programs for automatic accuracy control in regression. *Proceedings of the ninth International Conference on Artificial Neural Networks*. Edinburgh, UK: IET, 1999: 575 – 580.
- [34] VAPNIK V N. An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, 1999, 10(5): 988 – 999.
- [35] JULIÁN P, JORDÁN M, DESAGES A. Canonical piecewise-linear approximation of smooth functions. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 1998, 45(5): 567 – 571.
- [36] GOGNA A, TAYAL A. Metaheuristics: review and application. *Journal of Experimental & Theoretical Artificial Intelligence*, 2013, 25(4): 503 – 526.
- [37] JIANG P, CHEN J J. Displacement prediction of landslide based on generalized regression neural networks with K-fold cross-validation. *Neurocomputing*, 2016, 198: 40 – 47.
- [38] MIRJALILI S. SCA: A sine cosine algorithm for solving optimization problems. *Knowledge-Based Systems*, 2016, 96: 120 – 133.
- [39] LIU Y, WANG R. Study on network traffic forecast model of svr optimized by gafsa. *Chaos, Solitons & Fractals*, 2016, 89: 153 – 159.
- [40] BARATI M, SHARIFIAN S. A hybrid heuristic-based tuned support vector regression model for cloud load prediction. *The Journal of Supercomputing*, 2015, 71(11): 4235 – 4259.
- [41] CHANG C C, LIN C J. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2011, 2: 1 – 27. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [42] TIPPING M E. Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, 2001, 1: 211 – 244.

作者简介:

刘小雍 博士, 副教授, 主要研究方向为支持向量机、数据-模型融合建模及故障检测, E-mail: liuxy204@163.com;

方华京 教授, 博士生导师, 主要研究方向为复杂系统的故障诊断、鲁棒及故障容错控制, E-mail: hjfang@mail.hust.edu.cn;

陈孝玉 硕士, 讲师, 主要研究方向为动态系统的状态估计及故障检测, E-mail: 66070056@qq.com.