

最优实验设计与 Laplacian 正则化的 WNN 的非线性预测控制

任世锦^{1,2} 王高峰¹ 李新玉³ 杨茂云^{1,3} 徐桂云³

(1. 江苏师范大学计算机科学与技术学院, 徐州, 221116; 2. 浙江大学工业控制国家重点实验室, 杭州, 310027; 3. 中国矿业大学机电工程学院, 徐州, 221114)

摘要: 提出了基于最优实验设计与 Laplacian 正则化的自适应小波神经网络(Wavelet neural network, WNN)的非线性预测控制算法。该方法迭代地从 WNN 隐含节点候选集选取隐含小波神经元, 并使用扩展卡尔曼滤波(Extended Kalman filter, EKF)方法调整该节点参数。为了控制 WNN 的复杂度, 提出采用 Laplacian 正则化和最优实验设计选择重要的 WNN 隐含节点, 使用最小描述长度(Minimum description length, MDL)准则确定节点数量。使用在线基于 Gustafson-kesscl(GK)的模糊满意聚类算法确定 WNN 初始参数值和权重更新策略, 该策略具有直观性和物理意义。最后给出基于 WNN 线性化模型的预测函数控制方法。对工业焦化装置温度控制进行仿真, 结果说明了算法的有效性。

关键词: 小波神经网络; 扩展卡尔曼滤波; 预测控制; Laplacian 正则化; 满意模糊聚类

中图分类号: TP273 文献标志码: A

Nonlinear Predictive Control of WNN Using Optimal Experimental Design and Laplacian Regularization

Ren Shijin^{1,2}, Wang Gaofeng¹, Li Xinyu³, Yang Maoyun^{1,3}, Xu Guiyun³

(1. College of Computer Science & Technology, Jiangsu Normal University, Xuzhou, 221116, China; 2. National Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou, 310027, China; 3. School of Mechatronic Engineering, China University of Mining and Technology, Xuzhou, 221114, China)

Abstract: A nonlinear predictive control algorithm based on wavelet neural network (WNN) integrating optimal experimental design with manifold regularization is presented for the complex processes. Firstly, the wavelet hidden nodes are recursively selected from candidate node set to be added into WNN and the optimal parameters of selected nodes are obtained through extended Kalman filter (EKF). The optimum experimental design and Laplacian regularization are then integrated to select salient WNN hidden nodes, and minimum description length (MDL) is utilized to determine the number of hidden nodes. Initial WNN parameters and associated weight updating scheme are provided via an online Gustafson-kesscl (GK) based fuzzy satisfactory clustering algorithm with intuitive interpretation and physic meaning. Finally, a predictive functional control law is given by linearizing WNN. The simulation of industrial coking equipment shows the efficiency of the proposed algorithm.

Key words: wavelet neural networks (WNN); extended Kalman filter; predictive control; Laplacian reg-

ularization; satisfactory fuzzy clustering

引 言

预测控制自 20 世纪 70 年代以来,已经发展成为理论丰富、实践应用广泛的新的学科分支,在复杂过程中的成功应用显示其巨大的发展潜力^[1-3]。预测函数控制(Predictive function control, PFC)具有在线计算量小、跟踪能力强和鲁棒性好的优点,能够确定控制输入的结构,控制输入具有直观的物理意义^[2,3]。现代工业过程正朝着大型化、产品多批次趋势发展,过程对象表现出多工况、非线性和动态时变特性,很难建立或辨识准确的过程模型^[3-6]。预测控制是基于模型的先进控制算法,不管采取何种非线性模型,预测控制输出均是基于工作点附近增量线性化基础之上^[7]。准确辨识复杂过程数学模型是提高产品质量平稳性和企业经济效益的前提条件。目前线性预测控制理论已经成熟,过程对象的自身不确定性因素和外部干扰会严重影响预测控制性能,针对多工况和强非线性的控制对象的非线性预测控制成为人们研究的重要内容。

目前非线性预测控制算法大致分为如下 3 个方面:(1) 对非线性模型在工作点线性化的预测控制方法。该类方法首先使用非线性建模方法辨识控制对象,然后对非线性模型进行线性化并使用预测控制方法实现非线性系统的控制。小波神经网络(Wavelet neural network, WNN)^[8,9]、递归神经网络、模糊神经网络^[10-12]、支持向量机及其他模型(如 Winer 模型)^[13]常用于非线性系统的辨识。(2) 基于多局部模型的预测控制算法。该类方法主要使用聚类分析算法或者先验知识把非线性空间划分为多个局部线性子空间,每个局部线性模型对应一个运行状态。采用一定的控制切换策略、T-S 模糊建模等方法实现非线性系统的预测控制,降低控制算法的计算量和复杂度^[3,14-16]。(3) 基于非线性系统约束优化求解的非线性预测控制方法。该类算法针对非线性机理或者辨识模型,通过求解非线性约束优化问题实现预测控制^[17-19],且成功应用的关键是准确辨识非线性工业过程系统以及降低在线非线性辨识、非凸约束优化问题求解的计算量。一些学者还提出多层预测控制算法、参数预测控制等其他非线性预测控制算法^[1,5]。但是该类方法距离实际应用还有相当大的差距,有关非线性预测控制综述参见文献^[1,5,17,20,21]。

现有的多模型预测控制方法往往假定工况数量已知,需要较为完备的过程先验知识,在实际应用中很难满足该假设。另外,需要在线调整辨识模型克服工业过程运行中存在未知干扰和设备参数时变问题以保证控制对象的平稳运行。考虑到非线性预测控制在理论和实际应用存在的问题,鉴于 WNN 良好的局部表示以及以任意精度逼近任意函数能力^[21,22],本文提出一种最优实验设计(Optimal experimental design)A-optimum 和 Laplacian 正则化结合的在线 WNN 的 PFC 算法,提高多工况、动态时变工业过程对象的控制性能。该方法的主要贡献如下:(1) 在线调整 WNN 权重。该方法在样本数据滑动窗口内使用 QR 分解算法,通过增加最新样本及移去最老样本的方法调整 WNN 权重,更好地逼近当前时刻实际对象,有效地解决了动态性、不确定因素对控制性能的影响;(2) WNN 隐含节点增加和参数学习。该方法使用误差下降最大准则选择新的 WNN 隐含节点,然后基于扩展卡尔曼滤波算法(Extended Kalman filter, EKF)快速调整新增隐含节点参数方法,解决了控制对象工况变化导致辨识模型精度降低的问题,克服了节点候选集中小波神经元参数选取困难的问题;(3) WNN 模型选择。基于流形学习和最小二乘支持向量机(Least square support vector machine, LSSVM)理论,提出 Laplacian 正则化 LSSVM 与 A-optimum 实验设计结合的选择 WNN 隐含节点方法,并使用最小描述长度(Minimum description length, MDL)准则确定 WNN 隐含节点数量,保证 WNN 的推广性能;(4) 初始小波神经元参数值的确定以及更新策略判定。本文使用在线 Gustafson-kessel (GK) 模糊满意聚类确定小波神经元参数值以及判定 WNN 模型权重更新或结构更新,适合更广泛应用场合。通过工业焦化装置温度控制

的仿真实验证明了本文方法的有效性。

1 基于集成学习的 WNN 构造算法

集成学习就是对多个基分类器或者回归模型输出结果进行线性组合,每个基分类器可以看作对应复杂空间的局部线性子空间。如果把 3 层神经网络的每个隐含节点看作基本分类器或回归模型,那么神经网络的输出则可以看作多个基本模型的线性组合。基于上述分析,提出基于集成学习理论构造 WNN,并使用 EKF 方法调整小波神经元参数。假设给出 N 个数据样本组成的集合 $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$, WNN 表示为

$$y(\mathbf{x}) = \sum_{i=1}^k w_i g_i(\mathbf{x}) + e(\mathbf{x}) \quad (1)$$

式中: $g_i(\mathbf{x})$ 为径向基小波函数, $g_i(\mathbf{x}) = (d - \|\mathbf{x} - \mathbf{c}_i\|_i^2 / \sigma^2) \exp(-\mathbf{x} - \mathbf{c}_i^2 / 2\sigma^2)$; k 为 WNN 的隐含节点数目。

对于输入数据 \mathbf{x} , 式(1)所示 WNN 模型相应的输出记为 $y^{(k)}(\mathbf{x}) = \sum_{i=1}^k w_i g_i(\mathbf{x})$ 。令 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]^T$, $\mathbf{y}^{(k)} = [y^{(k)}(\mathbf{x}_1), \dots, y^{(k)}(\mathbf{x}_N)]^T$, $\tilde{\mathbf{y}}^{(k)} = \mathbf{y} - \mathbf{y}^{(k)}$ 为 WNN 对样本 (\mathbf{x}_i, y_i) 的逼近误差, $i=1, 2, \dots, N$, $\tilde{\mathbf{y}}^{(k)} = [\tilde{y}_1^{(k)}, \tilde{y}_2^{(k)}, \dots, \tilde{y}_N^{(k)}]^T$ 。为选择合适的小波函数作为新增小波隐含节点, 首先估计出候选节点集中每个节点 $p_i(\cdot)$ 的重要性, 其重要性根据 $\mathbf{p}_i = \mathbf{p}_i(\mathbf{X}) \triangleq [p_i(\mathbf{x}_1), p_i(\mathbf{x}_2), \dots, p_i(\mathbf{x}_N)]^T$ 在 $\mathbf{y}^{(k)}$ 上投影产生的误差下降程度进行估计。 \mathbf{p}_i 在 $\tilde{\mathbf{y}}^{(k)}$ 上的投影计算为

$$\text{Proj}_{\mathbf{p}_i} \tilde{\mathbf{y}}^{(k)} = \mathbf{p}_i \frac{\mathbf{p}_i^T \tilde{\mathbf{y}}^{(k)}}{\mathbf{p}_i^T \mathbf{p}_i} = \tilde{\mathbf{p}}_i \tilde{\mathbf{p}}_i^T \tilde{\mathbf{y}}^{(k)} \quad (2)$$

式中: $\tilde{\mathbf{p}}_i = \mathbf{p}_i / \|\mathbf{p}_i\|$ 为规格化向量。式(2)可以看作 $\tilde{\mathbf{y}}^{(k)}$ 在向量 \mathbf{p}_i 方向上的逼近, 其相应的投影误差为 $\tilde{\mathbf{y}}^{(k+1)} = \tilde{\mathbf{y}}^{(k)} - \text{Proj}_{\mathbf{p}_i} \tilde{\mathbf{y}}^{(k)}$ 。很显然, 使得误差减少最大的节点作为新的 WNN 隐含节点, 即

$$\mathbf{p}_{k^*} = \min_{\mathbf{p}_i \in \text{candiatorset}} (\|\tilde{\mathbf{y}}^{(k)}\|^2 + C_r / \|\mathbf{p}_i\|^2) \quad (3)$$

式中: C_r 为很小的正数, 可以看作是正则化因子, 目的是防止因 $\|\mathbf{p}_i\|$ 太小而导致过拟合问题。增加小波神经元会减小 WNN 对样本集的逼近误差。因此有必要分析 WNN 构造算法的收敛性以及收敛速度。

假设求解式(3)优化问题后选取小波节点函数 $p_i(\cdot)$, 那么 $p_i(\cdot)$ 与样本集 $\{(\mathbf{x}_i, \tilde{y}_i^{(k)})\}_{i=1}^N$ 之间的相关程度为

$$\lambda_i = \frac{|p_i(\mathbf{X})^T \tilde{\mathbf{y}}^{(k)}|}{\|\mathbf{p}_i(\mathbf{X})\| \|\tilde{\mathbf{y}}^{(k)}\|} = \frac{\|\tilde{\mathbf{p}}_i(\mathbf{X})^T \tilde{\mathbf{y}}^{(k)}\|}{\|\tilde{\mathbf{y}}^{(k)}\|} \quad (4)$$

式中: $\lambda_i \in [0, 1]$, λ_i 越大表示函数 $p_i(\mathbf{X})$ 与 $\tilde{\mathbf{y}}^{(k)}$ 之间关联程度越大。根据式(2, 5), 投影 $\text{Proj}_{\mathbf{p}_i} \tilde{\mathbf{y}}^{(k)}$ 与相关程度 λ_i 满足 $\text{Proj}_{\mathbf{p}_i} \tilde{\mathbf{y}}^{(k)} = \tilde{\mathbf{p}}_i(\mathbf{X}) \|\tilde{\mathbf{y}}^{(k)}\| \lambda_i$ 。一般 WNN 的拟合误差均方被视为经验风险函数, 即

$$\text{err}_k = \frac{1}{2} \sum_{i=1}^N (y_i - y^{(k)}(\mathbf{x}))^2 = \frac{1}{2} \|\mathbf{y} - \mathbf{y}^{(k)}\|^2 = \frac{1}{2} \|\tilde{\mathbf{y}}^{(k)}\|^2 \quad (5)$$

那么投影误差可写为

$$\|\tilde{\mathbf{y}}^{(k)} - \text{Proj}_{\mathbf{p}_i} \tilde{\mathbf{y}}^{(k)}\|^2 = \|\tilde{\mathbf{y}}^{(k)}\|^2 - \|\tilde{\mathbf{p}}_i^T \tilde{\mathbf{y}}^{(k)}\|^2 + \|\tilde{\mathbf{p}}_i^T \tilde{\mathbf{y}}^{(k)}\|^2 = \|\tilde{\mathbf{y}}^{(k)}\|^2 (1 - \lambda_i^2) \quad (6)$$

选取最优节点函数 $p_i(\cdot)$ 后, 相应的拟合误差可以记为 $\tilde{\mathbf{y}}^{(k+1)} = \tilde{\mathbf{y}}^{(k)} - \text{Proj}_{\mathbf{p}_i} \tilde{\mathbf{y}}^{(k)}$, 有 $\|\tilde{\mathbf{y}}^{(k+1)}\|^2 = \|\tilde{\mathbf{y}}^{(k)}\|^2 - \lambda_i^2 \|\tilde{\mathbf{y}}^{(k)}\|^2$ 。根据式(4, 5)可以写成以下等价形式

$$\text{err}_{k+1} = \frac{1}{2} \|\tilde{\mathbf{y}}^{(k+1)}\|^2 = \text{err}_k (1 - \lambda_i^2) \quad (7)$$

根据式(7)可知, WNN 加入隐含节点函数 $p_i(\mathbf{x})$ 后的拟合误差以 λ_i^2 速率减少。因此, 选择与样本集相关性大的小波函数作为 WNN 新增隐含节点有利于加快算法的收敛速度。根据投影理论^[23], WNN

的新增隐含节点 $p_l(\mathbf{x})$ 与逼近误差之间的最佳权重 ω_l^* 就是投影系数 θ_l , 即

$$\omega_l^* = \theta_l = \frac{\mathbf{p}_l(\mathbf{X})^T \bar{\mathbf{y}}^{(k)}}{\|\mathbf{p}_l(\mathbf{X})\|^2} = \sum_{i=1}^N p_l(\mathbf{x}_i) \bar{y}_i^{(k)} / \sum_{i=1}^N p_l(\mathbf{x}_i)^2 \quad (8)$$

式中: ω_l^* 也为使得误差最小的 WNN 新增隐含神经元 $p_l(\mathbf{x})$ 对应的最佳权重。WNN 构造算法收敛的结论如下。

定理 1 设样本集 $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$ 是满足某种有界未知的有界函数关系 $y = p(\mathbf{x}) \in [-B, B]$, 若存在正数 $\lambda_{\min} > 0$, 使得对所有的 $\lambda_i > \lambda_{\min}$ 成立, 则对于给定的拟合误差阈值 $\epsilon > 0$, 本文 WNN 构造算法至多经过 $T = \lfloor \ln(2\epsilon/NB^2) / \ln(1 - \lambda_{\min}^2) \rfloor$ 次迭代增加小波神经元后, 拟合误差不超过 ϵ , 即 $\text{err}_T =$

$$\frac{1}{2} \sum_{i=1}^N |\bar{y}_i^{(T)}|^2 \leq \epsilon. \quad \text{其中, } \lambda_{\min} = \operatorname{argmin}_{i=1,2,\dots,T} \lambda_i; \lfloor \cdot \rfloor \text{ 为取整算子。}$$

证明: 设算法 T 次迭代增加隐含小波神经元, 每一步由式(4)求得的相关性为 $\lambda_i, i = 1, 2, \dots, T$ 。根据式(7)有

$$\begin{aligned} \|\bar{\mathbf{y}}^{(1)}\|^2 &= (1 - \lambda_1^2) \|\bar{\mathbf{y}}^{(0)}\|^2 \\ \|\bar{\mathbf{y}}^{(2)}\|^2 &= \|\bar{\mathbf{y}}^{(1)}\|^2 (1 - \lambda_2^2) = \bar{y}^{(0)2} (1 - \lambda_1^2) (1 - \lambda_2^2) \\ &\vdots \\ \|\bar{\mathbf{y}}^{(T)}\|^2 &= \|\bar{\mathbf{y}}^{(0)}\|^2 \prod_{i=1}^T (1 - \lambda_i^2) \end{aligned}$$

其中, $\bar{\mathbf{y}}^{(0)} = \mathbf{y}$ 。令 $\lambda_{\min} = \operatorname{argmin}_{i=1,2,\dots,T} \lambda_i$, 根据相关条件有

$$\frac{1}{2} \|\mathbf{y}\|^2 \prod_{i=1}^T (1 - \lambda_i^2) \leq \frac{1}{2} NB^2 \prod_{i=1}^T (1 - \lambda_i^2) \leq \frac{1}{2} NB^2 (1 - \lambda_{\min}^2)^T \quad (9)$$

如果给定拟合误差阈值 $\epsilon > 0$ 使得

$$\frac{1}{2} NB^2 (1 - \lambda_{\min}^2)^T \leq \epsilon \quad (10)$$

那么根据式(9)一定有式(11)成立, 即

$$\text{err}_T = \frac{1}{2} \|\mathbf{y}\|^2 \prod_{i=1}^T (1 - \lambda_i^2) \leq \epsilon \quad (11)$$

对式(10)两边取对数、整理并取整得

$$T \geq \lfloor \ln(2\epsilon/NB^2) / \ln(1 - \lambda_{\min}^2) \rfloor \quad (12)$$

很显然, 当 $T = \lfloor \ln(2\epsilon/NB^2) / \ln(1 - \lambda_{\min}^2) \rfloor$ 时拟合误差不大于 ϵ 。

定理 1 表明, 对现有 WNN 增加有限数量的隐含节点可以一致逼近任意的非线性系统, 而构造合适的节点候选集是算法成功的重要环节。在实际中, 人们根据经验方法选择小波函数参数, 不可避免地会导致小波节点数量过多问题。过多冗余隐含节点不仅会增加算法的计算量, 而且会影响 WNN 模型的泛化性能。为克服上述问题, 本文在 WNN 构造过程中增加对已选小波神经元参数优化步骤, 降低样本拟合误差, 从而使用更少隐含节点, 以提高 WNN 的推广性能。

考虑到基于 EKF 的神经网络训练方法能够显著地加快神经网络收敛速度, 且在理论上能够保证算法的收敛性^[23, 24], 本文采用 EKF 方法求取新增小波神经元最优参数。常规训练方法的学习速率固定, 虽然较小的学习速率以保证算法收敛性, 但算法的收敛时间会大大增加, 不利于在线实施。为加快训练算法的收敛速度, 本文引入参数学习率因子 η 。令 $\boldsymbol{\theta} = [\sigma_k, \mathbf{c}_k, \omega_k]^T$, 类似于文献[23], 该方法迭代学习式为: $\boldsymbol{\theta}(n+1) = \boldsymbol{\theta}(n) + \eta \mathbf{K}(n) e(n)$, $\mathbf{K}(n) = \mathbf{P}(n) \mathbf{H}(n) (\mathbf{H}^T(n) \mathbf{P}(n) \mathbf{H}(n) + R(n))^{-1}$, $e(n) = \bar{y}(n) - \hat{y}^{(k)}(n)$, $\mathbf{P}(n+1) = (\mathbf{I} - \mathbf{K}(n) \mathbf{H}(n)) \mathbf{P}(n)$ 。其中, $\bar{y}(n)$ 与 $\hat{y}^{(k)}(n)$ 分别为局部小波神经元 k 输出和期望输出; $e(n)$ 为训练误差; $\mathbf{H}(n) = \frac{\partial \hat{y}^{(k)}(n)}{\partial \boldsymbol{\theta}(n)}$; $\boldsymbol{\eta} = \text{diag}(\eta_1, \eta_2, \eta_3) \triangleq \text{diag}(\eta_{\sigma}, \eta_{\mathbf{c}}, \eta_{\omega})$ 为学习速率对角矩阵; \mathbf{K}

(n)为 Kalman 增益向量; $R(n)$ 为估计的噪声协方差,由 $R(n) = R(n-1) + \frac{[e_i(n)^2 - R(n-1)]}{n}$ 进行递归求取; $\mathbf{P}(n)$ 为状态估计误差协方差矩阵; n 为 EKF 算法的迭代次数。上述算法的收敛速度依赖于学习速率 η 。为加快该学习算法的收敛性,学习速率自适应调整方法。

定理 2 令 $\theta = [\sigma_{k^*}, c_{k^*}, w_{k^*}]$ 为隐含节点 k^* 的参数; $\hat{y}(n)$ 为隐含节点的预测输出; $y(n)$ 为隐含节点 k^* 的对样本 $\mathbf{x}(n)$ 的期望输出; $\boldsymbol{\eta}$ 为相应参数 $\theta = [\sigma_{k^*}, c_{k^*}, w_{k^*}]$ 的学习速率, $\mathbf{K}(n)$ 为 EKF 增益向量。如果学习速率 η_i 满足

$$0 < \eta_i < \frac{2}{\mathbf{H}(n)^\top \mathbf{K}(n)} \quad i=1,2,3 \quad (13)$$

那么基于 EKF 的 WNN 训练算法一致收敛。

证明:令 WNN 在加入隐含节点 k^* 后样本 $(\mathbf{x}(n), y(n))$ 的预测输出为 $\hat{y}(n)$,则 WNN 的预测误差为 $e(n) = y(n) - \hat{y}(n)$ 。取 Lyapunov 函数为

$$V(n) = \frac{1}{2} e(n)^2$$

那么离散 Lyapunov 函数的微分为

$$\Delta V(n) = V(n+1) - V(n) = \frac{1}{2} \Delta e(n) (2e(n) + \Delta e(n)) \quad (14)$$

式中: $\Delta e(n)$ 为误差的微分。根据式(13),有

$$\Delta e(n) = \left[\frac{\partial e(n)}{\partial \theta(n)} \right]^\top \Delta \theta(n) = - \left[\frac{\partial \hat{y}(n)}{\partial \theta(n)} \right]^\top \boldsymbol{\eta} \mathbf{K}(n) e(n) \quad (15)$$

如果令 $\mathbf{H}(n) = \frac{\partial \hat{y}(n)}{\partial \theta(n)}$,那么把式(15)代入式(14)表示的 Lyapunov 的微分 $\Delta V(n)$,可得

$$\Delta V(n) = -\frac{1}{2} \mathbf{H}(n)^\top \boldsymbol{\eta} \mathbf{K}(n) (2 - \mathbf{H}(n)^\top \boldsymbol{\eta} \mathbf{K}(n)) e(n)^2 \quad (16)$$

根据 Lyapunov 理论,如果满足下面条件

$$-\frac{1}{2} \mathbf{H}(n)^\top \boldsymbol{\eta} \mathbf{K}(n) (2 - \mathbf{H}(n)^\top \boldsymbol{\eta} \mathbf{K}(n)) < 0 \quad (17)$$

那么 $\Delta V(n) < 0$,则训练算法一致收敛。对式(17)进一步整理可知,其等价于 $0 < \eta_j < 2 / \mathbf{H}(n)^\top \mathbf{K}(n)$, $j=1,2,3$ 。

定理 2 表明,EKF 算法可以自适应确定算法速率,能够保证 WNN 训练算法一致收敛。使用 EKF 算法求取隐含节点 k^* 最优参数,然后计算拟合误差。拟合误差如果满足条件 $\|\hat{\mathbf{y}}^{(k)}\| \leq \epsilon$ 表明 WNN 构建结束,否则 $k=k+1$,重复上述方法从候选集中选取合适的小波函数作为新增节点加入到 WNN 中,然后使用 EKF 求取最优 WNN 新增隐含神经元参数。

在实际中需要 WNN 随着控制对象工况变化调整 WNN 模型。因此 WNN 的规模随着时间逐渐加大,其复杂度也随之提高。因此有必要采取措施选择合适 WNN 模型,从而提高 WNN 泛化性能。

2 WNN 模型选择

本文提出 WNN 隐含节点选择算法以及 MDL 准则控制模型复杂度,前者基于最优实验设计原理^[25,26]。

2.1 最优实验设计

对给定的数据集 $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$,假设它们存在下面的线性关系

$$\mathbf{y} = \mathbf{w}^\top \mathbf{x} + \epsilon \quad (18)$$

式中: $\varepsilon \in N(0, \sigma^2)$ 为高斯噪声。最优实验设计思想就是选择包含最有代表性的实验数据作为训练数据, 使得预测函数 $f(\mathbf{x}) = \mathbf{w}^T \mathbf{x}$ 对数据样本集的预测误差最小。令 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbf{R}^{d \times N}$, $\mathbf{y} = [y_1, y_2, \dots, y_N]^T$, 则 $\mathbf{y} = \mathbf{X}^T \mathbf{w} + \varepsilon$ 。最优实验设计估计方法就是使用最小均方误差作为代价函数, 其最优解为

$$\hat{\mathbf{w}} = (\mathbf{X}\mathbf{X}^T)^{-1} \mathbf{X}\mathbf{y} \quad (19)$$

为了保证回归模型的泛化性, 还期望模型参数的方差最小, 即保证隐含节点权重的一致性。注意到

$$E(\hat{\mathbf{w}} - \mathbf{w}) = E((\mathbf{X}\mathbf{X}^T)^{-1} \mathbf{X}\varepsilon) - E(\mathbf{w}) = 0$$

因此, 式(19)所示的模型参数估计为无偏估计, 那么模型权重协方差可以表示为

$$\text{Cov}(\hat{\mathbf{w}}) = \sigma^2 E((\mathbf{X}\mathbf{X}^T)^{-1} \mathbf{X}\mathbf{X}^T (\mathbf{X}\mathbf{X}^T)^{-1}) = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \quad (20)$$

由预测模型为 $\hat{y} = \hat{\mathbf{w}}^T \mathbf{x}$, 预测输出方差表示为

$$\text{var}(\hat{y}) = \mathbf{x}^T \text{Cov}(\hat{\mathbf{w}}) \mathbf{x} = \sigma^2 \mathbf{x}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x} \quad (21)$$

由式(21)可见, 模型参数预测方差最小等价于模型估计参数方差最小。目前已经出现了多个优化准则度量模型参数方差, 其中 A-optimum 和 D-optimum 已经引起人们广泛的关注^[28], 本文使用 A-optimum 准则。

2.2 基于 A-optimum 设计和 MDL 的 WNN 模型选择

为利用训练数据和测试数据的局部几何信息, 本文提出使用 Laplacian 正则化最小二乘支持向量机 (Least square support vector machine with Laplacian regularization, LR-LSSVM) 估计 WNN 的权值, 然后使用 A-optimum 准则序贯地选择 WNN 隐含节点, 并使用 MDL 准则确定隐含节点数量。

假设 WNN 存在 M 个隐含节点, 设输入样本 $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$, $\mathbf{x}_i \in \mathbf{R}^{d \times 1}$, $y_i \in \mathbf{R}$, WNN 输出及其隐含节点输出记为 \hat{y}_i 和 $\mathbf{z}_i \in \mathbf{R}^{M \times 1}$ 。记 $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N] \in \mathbf{R}^{M \times N}$, 令 $\mathbf{f}_i \in \mathbf{R}^{N \times 1}$ ($i=1, 2, \dots, d$) 为数据集 \mathbf{Z} 的 i -th 特征组成的向量, 记特征集为 $\mathbf{F} = \{f_1, f_2, \dots, f_M\}$ 。很显然, 数据集 \mathbf{Z} 的 i -th 特征对应 WNN 的 i -th 隐含节点。因此选择 p ($p < M$) 个重要的 WNN 节点就等价于从数据集 \mathbf{Z} 选择 p 个重要特征。假设选择 p 个特征记为 $\mathbf{S}_f = \{\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_p\} \subset \mathbf{F}$ 。令 $\mathbf{Z}^s = [\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_p]^T \triangleq [Z_1^s, Z_2^s, \dots, Z_N^s]$ 为由选择特征组成的样本集, 其中, $Z_i^s \in \mathbf{R}^{p \times 1}$, $i=1, 2, \dots, N$ 。令 $\mathbf{\Lambda} = \text{diag}\{\Lambda_{11}, \Lambda_{22}, \dots, \Lambda_{MM}\}$, $\Lambda_{ii} \in \{0, 1\}$, $i=1, 2, \dots, M$, 其定义为, 如果 $f_i \in \mathbf{S}_f$ 则 $\Lambda_{ii} = 1$, 否则为 0。这样有, $\mathbf{Z}^s = \mathbf{\Lambda} \mathbf{Z}$, $\mathbf{z}_j^s = \mathbf{\Lambda} \mathbf{z}_j$, $j=1, 2, \dots, N$ 。那么在选择的特征空间的回归模型具有如下形式

$$\mathbf{y} = \mathbf{w}^T \mathbf{z}^s + \varepsilon$$

为充分利用数据的局部几何结构信息, 更好地选择有效 WNN 隐含节点, 本文使用 LR-LSSVM 求取 \mathbf{w} 的最优值。基于流形学习理论和最小 LSSVM 理论^[27, 28], LR-LSSVM 优化问题的形式如下

$$J(\mathbf{w}) = \min \frac{1}{2} \|\mathbf{w}\|^2 + \frac{\nu}{2} \sum_{i,j=1}^N \|\mathbf{w}^T \mathbf{z}_i^s - \mathbf{w}^T \mathbf{z}_j^s\| S_{ij} + \frac{\gamma}{2} \sum_{i=1}^N y_i - \mathbf{w}^T \mathbf{z}_i^s \|^2 = \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{\nu}{2} \mathbf{w}^T \mathbf{Z} \mathbf{L} (\mathbf{Z}^s)^T \mathbf{w} + \frac{\gamma}{2} (\mathbf{y} - (\mathbf{Z}^s)^T \mathbf{w})^T (\mathbf{y} - (\mathbf{Z}^s)^T \mathbf{w}) \quad (22)$$

式中: $\nu, \gamma > 0$, 矩阵 \mathbf{L} 和权重 S_{ij} 定义见文献[28]。类似 LSSVM 求解方法可得 \mathbf{w} 最优解, 即有

$$\hat{\mathbf{w}} = (\mathbf{I} + \nu \mathbf{Z} \mathbf{L} (\mathbf{Z}^s)^T + \gamma \mathbf{Z}^s (\mathbf{Z}^s)^T)^{-1} \mathbf{Z} \mathbf{y} \quad (23)$$

式中: \mathbf{I} 为 $p \times p$ 的单位矩阵。令 $\mathbf{H} = \mathbf{I} + \nu \mathbf{Z} \mathbf{L} (\mathbf{Z}^s)^T + \gamma \mathbf{Z}^s (\mathbf{Z}^s)^T$, 那么

$$\hat{\mathbf{w}} = \mathbf{H}^{-1} \mathbf{Z} \mathbf{y} \quad (24)$$

注意到 $\mathbf{y} = (\mathbf{Z}^s)^T \mathbf{w} + \varepsilon$, \mathbf{H} 正定对称性以及 $\text{Cov}(\mathbf{y}) = \text{Cov}(\varepsilon) = \sigma^2 \mathbf{I}$, 令 $\mathbf{\Xi} = \nu \mathbf{Z} \mathbf{L} (\mathbf{Z}^s)^T + \mathbf{I}$, 估计参数 $\hat{\mathbf{w}}$ 的有偏性和方差为

$$E(\hat{\mathbf{w}} - \mathbf{w}) = \frac{1}{\gamma} (\mathbf{H}^{-1} (\mathbf{H} - \mathbf{\Xi}) \mathbf{w} - \mathbf{w}) = -\frac{1}{\gamma} \mathbf{H}^{-1} \mathbf{\Xi} \mathbf{w}$$

$$\text{Cov}(\hat{\mathbf{w}}) = \sigma^2 \mathbf{H}^{-1} \frac{1}{\gamma} (\mathbf{H} - \mathbf{E}) \mathbf{H}^{-1} = \sigma^2 \frac{1}{\gamma} (\mathbf{H}^{-1} - \mathbf{H}^{-1} \mathbf{E} \mathbf{H}^{-1}) \quad (25)$$

根据式(24)可知预测值为 $\hat{\mathbf{y}} = \hat{\mathbf{w}}^T \mathbf{Z} = (\mathbf{Z}')^T \mathbf{H}^{-1} \mathbf{Z}$, 那么 WNN 预测误差的方差为

$$E(y - \hat{y})^2 = \sigma^2 + (\mathbf{Z}')^T E[(\mathbf{w} - \hat{\mathbf{w}})(\mathbf{w} - \hat{\mathbf{w}})^T] \mathbf{Z}' \quad (26)$$

注意到 $E(\hat{\mathbf{w}}) = \mathbf{H}^{-1} \mathbf{X}' \mathbf{x}'^T \mathbf{w}$ 和 $E((\mathbf{w} - \hat{\mathbf{w}})E(\mathbf{w} - \hat{\mathbf{w}})^T) = E(\mathbf{w} - \hat{\mathbf{w}})(\mathbf{w} - \hat{\mathbf{w}})^T$, $\text{Cov}(\mathbf{w}) = 0$, 根据式(25), 参数估计误差的方差为

$$\text{Cov}(\mathbf{w} - \hat{\mathbf{w}}) = \sigma^2 \frac{1}{\gamma} (\mathbf{H}^{-1} - \mathbf{H}^{-1} \mathbf{E} \mathbf{H}^{-1}) + \frac{1}{\gamma^2} \mathbf{H}^{-1} \mathbf{E} \mathbf{w} \mathbf{w}^T \mathbf{E} \mathbf{H}^{-1} \quad (27)$$

把式(27)代入式(26)得到预测方差得, 即

$$E(y - \hat{y})^2 = \sigma^2 + (\mathbf{Z}')^T \left(\sigma^2 \frac{1}{\gamma} (\mathbf{H}^{-1} - \mathbf{H}^{-1} \mathbf{E} \mathbf{H}^{-1}) + \frac{1}{\gamma^2} \mathbf{H}^{-1} \mathbf{E} \mathbf{w} \mathbf{w}^T \mathbf{E} \mathbf{H}^{-1} \right) \mathbf{Z}'$$

一般来说, 正则化系数 ν 设置比较小, 而误差惩罚系数 γ 设置较大。注意到 \mathbf{L} 和 $\mathbf{Z}'(\mathbf{Z}')^T$ 为正定矩阵, 且 $\det(\mathbf{H}^{-1}) \leq 1$, 有

$$\text{Cov}(\hat{\mathbf{w}}) = \sigma^2 \left(\frac{1}{\gamma} \mathbf{H}^{-1} - \mathbf{H}^{-1} \left(\frac{\nu}{\gamma} \mathbf{Z}' \mathbf{L} (\mathbf{Z}')^T + \frac{1}{\gamma} \mathbf{I} \right) \mathbf{H}^{-1} \right) \approx \frac{\sigma^2}{\gamma} (\mathbf{H}^{-1} - \mathbf{H}^{-1} \mathbf{H}^{-1}) \approx \frac{\sigma^2}{\gamma} \mathbf{H}^{-1}$$

类似地, 有

$$E(y - \hat{y})^2 = \sigma^2 + (\mathbf{Z}')^T \left(\sigma^2 \frac{1}{\gamma} \mathbf{H}^{-1} + \frac{1}{\gamma^2} \mathbf{H}^{-1} \mathbf{w} \mathbf{w}^T \mathbf{H}^{-1} \right) \mathbf{Z}' \approx \sigma^2 + (\mathbf{Z}')^T \left(\sigma^2 \frac{1}{\gamma} \mathbf{H}^{-1} \right) \mathbf{Z}' = \sigma^2 + (\mathbf{Z}')^T \frac{1}{\gamma^2} \mathbf{H}^{-1} (\sigma^2 \gamma)$$

基于最优实验设计原理^[26,28], 本文期望选择的特征子集能够使得估计参数的协方差矩阵最小。而最小化 \mathbf{H}^{-1} 也能够使得新样本的预测误差最小, 该问题可以等价于最优实验设计准则 A-optimum, 即

$$\min_{\mathbf{H}} \text{tr}(\mathbf{H}^{-1})$$

对矩阵 $\mathbf{H} = \mathbf{I} + \nu \mathbf{Z}' \mathbf{L} (\mathbf{Z}')^T + \gamma \mathbf{Z}' (\mathbf{Z}')^T = \mathbf{I} + \mathbf{Z}' (\nu \mathbf{L} + \gamma \mathbf{I}) (\mathbf{Z}')^T$, \mathbf{L} 为正半定矩阵, $\nu \mathbf{L} + \gamma \mathbf{I}$ 为正定、可逆矩阵。令 $\mathbf{G} = (\nu \mathbf{L} + \gamma \mathbf{I})^{-1}$, 根据 Woodbury 公式^[29], 有

$$\mathbf{H}^{-1} = \mathbf{I} - \mathbf{Z}' (\mathbf{G} + (\mathbf{Z}')^T (\mathbf{Z}'))^{-1} (\mathbf{Z}')^T$$

注意到迹性质 $\text{tr}(\mathbf{A}\mathbf{B}) = \text{tr}(\mathbf{B}\mathbf{A})$, A-optimum 准则的目标函数可以写为

$$\text{tr}(\mathbf{H}^{-1}) = p - \text{tr}((\mathbf{G} + (\mathbf{Z}')^T \mathbf{Z}')^{-1} (\mathbf{Z}')^T \mathbf{Z}') = p - m + \text{tr}((\mathbf{G} + (\mathbf{Z}')^T \mathbf{Z}')^{-1} \mathbf{G}) \quad (28)$$

考虑到 p, m 为参数, 因此选择 WNN 节点问题就转化为如下优化问题

$$\min_{S \subset F} \text{tr}((\mathbf{G} + (\mathbf{Z}')^T \mathbf{Z}')^{-1} \mathbf{G})$$

注意到 \mathbf{Z}' 只包含 p 个选择特征: $\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_p$, 因此 $(\mathbf{Z}')^T \mathbf{Z}'$ 可以写为 $(\mathbf{Z}')^T \mathbf{Z}' = \sum_{i=1}^p \mathbf{g}_i \mathbf{g}_i^T$ 。由于 $\mathbf{Z}' = \mathbf{A}\mathbf{Z}$, 则 $(\mathbf{Z}')^T \mathbf{Z}' = \mathbf{Z}^T \mathbf{A}\mathbf{Z}$ 。这样式(28)所示优化问题转化为

$$\min_{S \subset F} \text{tr}((\mathbf{G} + \sum_{i=1}^p \mathbf{g}_i \mathbf{g}_i^T)^{-1} \mathbf{G}) = \text{tr}((\mathbf{G} + \sum_{i=1}^n \Lambda_i \mathbf{f}_i \mathbf{f}_i^T)^{-1} \mathbf{G}) \quad \Lambda_i \in \{0, 1\}; \sum_{i=1}^d \Lambda_i = p \quad (29)$$

使用序贯优化方法对式(29)所示优化问题进行求解^[26]。首先假设已经选择 $k(k < p)$ 个节点 $\mathbf{f}_{s_1}, \mathbf{f}_{s_2}, \dots, \mathbf{f}_{s_k}$, 可以通过下面的优化问题选取第 $(k+1)$ 个节点 $\mathbf{f}_{s_{k+1}}$, 即

$$\mathbf{f}_{s_{k+1}} = \min_{i \in \{s_1, s_2, \dots, s_k\}} \text{tr}((\mathbf{G} + \sum_{j=1}^k \mathbf{g}_j \mathbf{g}_j^T + \mathbf{X}^T \mathbf{e}_i \mathbf{e}_i^T \mathbf{X})^{-1} \mathbf{G}) = \text{tr}((\mathbf{G} + \mathbf{f}_i \mathbf{f}_i^T + \sum_{j=1}^k \mathbf{f}_j \mathbf{f}_j^T)^{-1} \mathbf{G}) \quad (30)$$

令 $\mathbf{D}_k = \mathbf{G} + \sum_{j=1}^k \mathbf{f}_j \mathbf{f}_j^T$, 根据 Woodbury 和 Sherman-Morrison 公式^[29] 可得

$$\text{tr}((\mathbf{D}_k + \mathbf{f}_i \mathbf{f}_i^T)^{-1} \mathbf{G}) = \text{tr}(\mathbf{D}_k^{-1} \mathbf{G}) - \frac{1}{1 + \mathbf{f}_i^T \mathbf{D}_k^{-1} \mathbf{f}_i} \text{tr}(\mathbf{D}_k^{-1} \mathbf{f}_i \mathbf{f}_i^T \mathbf{D}_k^{-1} \mathbf{G}) \quad (31)$$

由于 \mathbf{D}_k^{-1} 和 \mathbf{G} 为常矩阵, 式(30)所示优化问题的序贯求解方法为

$$f_{k+1} = \max_{i \in \{s_1, s_2, \dots, s_k\}} \frac{1}{1 + \mathbf{f}_i^T \mathbf{D}_k^{-1} \mathbf{f}_i} \text{tr}(\mathbf{D}_k^{-1} \mathbf{f}_i \mathbf{f}_i^T \mathbf{D}_k^{-1} \mathbf{G}) \quad (32)$$

这样,通过求解上面的优化问题可以依次选取 WNN 隐含节点。隐含节点的数量与数据拟合精度密切相关,直接影响 WNN 的泛化性能。使用最小描述长度准则确定小波神经网络隐含节点数量。MDL 准则的形式为^[30]

$$\text{MDL} = -\ln f + \frac{k}{2} \ln N \quad (33)$$

式中: $f = \max_{\theta} q(\mathbf{x}^N | \theta)$ 为最大似然函数, $\theta = (\mathbf{w}, \sigma)$ 为模型参数。为了使用 MDL 找到最优的模型参数 N_h , 本文引入了如下的 WNN 统计模型

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{h}(\mathbf{x}) + \varepsilon \quad (34)$$

式中: \mathbf{w} 为权重向量; $\mathbf{h}(\mathbf{x})$ 为小波网络隐含节点的输出; ε 为均值为 0、方差为 δ^2 的高斯白噪声。那么上述模型的概率密度函数为高斯分布, 即

$$p(f(\mathbf{x}) | \theta = (\mathbf{w}, \sigma)) = \frac{1}{\sqrt{2\pi\delta^2}} \exp\left(-\frac{(f(\mathbf{x}) - \mathbf{w}^T \mathbf{h}(\mathbf{x}))^2}{2\delta^2}\right) \quad (35)$$

则 WNN 对应的 MDL 形式转换为

$$\text{MDL}_{\text{WNN}} = -\ln f + \frac{(N_h)^2}{2} \ln \frac{N}{2\pi} \quad (36)$$

这里,最大对数似然函数

$$\ln f = -\frac{N}{2} \ln 2\pi\delta^2 - \frac{1}{2\delta^2} \sum_{i=1}^N (y_i - \hat{\mathbf{w}}^T \hat{\mathbf{h}}(\mathbf{x}_i))^2 \quad (37)$$

式中: $\hat{\mathbf{w}}, \hat{\mathbf{h}}(\mathbf{x}_i)$ 分别为 WNN 估计权重和隐含节点输出; $\delta^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{\mathbf{w}}^T \hat{\mathbf{h}}(\mathbf{x}_i))^2$ 。最小化式(33)可以得到最优的 N_h 。在给定训练数据集和构造出 WNN 后,使用 MDL 准则选择相应最优的 WNN 模型,且无需人为设定参数。

在线调整 WNN 权重是克服工业过程控制对象的动态不确定特性导致已有 WNN 与时间控制对象之间的误差问题的有效方法,该方法可以参考文献[31,32]推导出 WNN 在线更新算法。

3 WNN 更新策略判定方法

本文需要根据过程样本自适应调整 WNN 权重或模型结构,解决被控对象因工况变化、外界环境及设备状态变化等因素导致 WNN 预测精度下降的问题。在实际中通过准则判定固定长度滑动窗口内数据变化决定 WNN 更新策略^[33]。该类方法往往需要根据先验知识确定合适阈值,且容易受到噪声的影响。从直观上讲,每个工况对应一个聚类,不同工况的过渡过程对应聚类边界重叠样本。因此可以根据新来的过程数据样本在聚类空间的位置确定过程对象运行状态,从而采取合适更新策略。为此,提出在线基于 GK 的满意模糊聚类方法^[33]确定 WNN 权重参数或 WNN 结构更新策略,在线 GK 模糊满意聚类方法由文献[33]提出,它能够根据用户关心的性能指标自适应地确定模糊聚类的数量,算法收敛的速度很快。

基于上述聚类算法,模型更新策略准则为:(1)WNN 模型保持不变。对于新来的样本,首先计算样本属于当前聚类的隶属度大于 0.5 时,则认为系统当前参数没有变化;(2)WNN 结构调整。当需要增加一个新的聚类时或者属于另外一个聚类时,则认为系统的工况发生了变化,需要改变 WNN 的结构;(3)WNN 权重更新。当样本在多个聚类的交叉重叠处,认为系统的参数发生了变化。该方法直观简

单、物理意义明显以及计算量小、适合在线计算。

4 基于 WNN 线性化的预测函数控制

4.1 非线性模型线性化处理

类似文献[34,35],对由 K 个隐含神经元构成的 WNN 进行瞬时线性化,即在中心点 $\mathbf{x}_0 = \mathbf{x}(t-1)$ 利用 Taylor 公式对 WNN 数学表达式展开,并将 $y(\mathbf{x}(t))$ 写为线性离散差分方程形式为

$$y(k) = a_1 y(k-1) + \dots + a_{n_y} y(k-n_y) + b_1 u(k-d-1) + b_2 u(k-d) + \dots + b_{n_u} u(k-d-n_u) + \sum_{i=1}^N \beta_i e_i(\mathbf{x}(t)) + \sum_{i=1}^N \beta_i G_i + b_i \quad (38)$$

式中:系数 a_i, b_j 求取方法与文献[35]类似。

经过线性化处理后的离散化模型分为 3 部分:(1) 线性部分,其相应系数为 $a_1, \dots, a_{n_y}, b_1, \dots, b_{n_u}$;(2) 误差,由 $e(k) = \sum_{i=1}^N \beta_i e_i(\mathbf{x}(k))$ 表示系统的非线性部分及其误差部分;(3) 常数,由 $G = \sum_{i=1}^N \beta_i G_i$ 表示。

那么系统输出可表示为

$$y(k) = a_1 y(k-1) + \dots + a_{n_y} y(k-n_y) + b_1 u(k-d-1) + \dots + b_{n_u} u(k-d-n_u) + G + e(k) + b_i \quad (39)$$

4.2 预测函数控制律设计

如果 WNN 在工作点附近辨识精度足够高,非线性部分 G_i 可以忽略。对式(39)两边乘以差分算子 $\Delta = 1 - z^{-1}$ 可得预测模型

$$y(k) = A_1 y(k-1) + \dots + A_{n_y+1} y(k-n_y-1) + B_1 u(k-d-1) + B_2 u(k-d) + \dots + B_{n_u+1} u(k-d-n_u-1) + \Delta e(k) \quad (40)$$

式中: $B_{1,i} = b_i, i = 1, 3, \dots, n_u$; $A_1 = 1 + a_1, A_i = a_i - a_{i-1}, i = 2, 3, \dots, n_y, A_{n_y+1} = a_{n_y}$ 。

最优预测输出 $y(k)$ 由两部分组成:模型自由响应 $y_{\text{past}}(k)$,它是只依赖于过去的控制量和输出量;另外一部分为模型的受迫响应 $y_{F_0}(k)$,它是 $u(k)$ 引起的系统响应^[35]。 $\Delta e(k)$ 可看作外界干扰、模型失配等引起的预测模型和实际过程之间的误差。为推导方便,可假设 $\Delta e(k) = 0$ 。使用文献[33,34]方法,递推计算 $k+h$ 时刻的模型响应和受迫响应分别为

$$y_{\text{past}}(k+h) = \sum_{i=1}^{n_y+1} A_i y'_M(k+h-i) + \sum_{i=1}^{n_u+1} B_i u'(k+h-i)$$

$$y_{F_0}(k+h) = \sum_{i=1}^{n_u+1} C_i u(k+h-j) \quad h = 1, \dots, H \quad (41)$$

式中: H 为预测时域;如果 $h < i+1$,那么 $y'(k+h-i) = y(k+h-i)$,否则 $y'(k+h-i) = y_{\text{past}}(k+h-i)$;如果 $h < i$,那么 $u'(k+h-i) = u(k+h-i)$,否则为 0; C_i 根据以下形式递推计算得到: $C_1 = B_1, C_i =$

$$\sum_{t=1}^{\min(i-1, n_y+1)} A_t C_{i-t} + B_i (2 \leq j \leq m), C_i = \sum_{t=1}^{\min(i-1, n_u+1)} A_t C_{i-t} (m < j \leq H)。$$

预测函数控制被认为基于过程特性和跟踪设定有关基函数的组合。设参考轨迹 $y_r(k+H)$ 为如下形式

$$y_r(k+H) = c(k+H) - \gamma^H (c(k) - y(k)) \quad (42)$$

式中: $\gamma = e^{-\frac{\tau}{T_s}}$ 为柔化因子; T_s 为采样周期; T_r 为参考轨迹的响应时间。取 PFC 的二次优化问题为

$$J = \min(\|y_r - \hat{y}\|^2 + \tau \|\Delta u\|^2)$$

式中: $\tau > 0$ 为加权因子。为适应多步预测控制算法的要求, 取预测时域为 H , 设预测控制域内控制输出相等: $u(t) = u(t+1) = \dots = u(t+H-1)$ 。考虑到实际过程噪声和模型失配等误差, 对预测输出进行校正, 取预测输出

$$e(k+h) = y_p(k) - y_M(k) \quad h=1, \dots, H \quad (43)$$

基于单值预测控制思想, 那么优化问题变为

$$J_p = \min [y_M(k+H) + e(k+H) - y_r(k+H)]^2$$

即 $y_M(k+H) + e(k+H) = y_r(k+H)$ 。注意到 $y_M(k) = y_{past}(k) + y_{Fo}(k)$, 将式(38, 42, 43)代入式(39)得到最优控制律为

$$u(k) = [c(k+H) - \gamma^H c(k) - (1 - \gamma^H) y_p(k) + y_M(k) - y_{past}(k+H)] / \sum_{i=1}^H C_i \quad (44)$$

这样由式(44)很容易计算出 k 时刻最优输入 $u(k)$ 作用于系统。下一时刻重新进行采样、预测和优化, 实现了系统的最优控制。

5 实验仿真

工业焦化装置是炼油化工企业最普遍的生产单元。其执行部分为带有非线性死区的蝴蝶阀, 在运行中往往会受到负荷、锅炉燃烧及输入燃料品质等变化因素的影响, 传统的 PID 控制方法很难取得满意的效果^[35]。本文使用控制方法实现对焦化锅炉的平稳控制。

本文算法步骤为: 首先使用在线 GK 模糊满意聚类算法^[33]对过程数据进行聚类, 基于聚类结果构建候选节点集; 根据样本集以及候选节点集利用本文第 1, 2 节方法构建 WNN 模型; 根据新数据样本在聚类数据空间的分布情况确定 WNN 更新策略。然后根据更新后的 WNN 模型计算 PFC 控制器的输出。在实际应用时, 首先对采集的过程输入输出数据进行低通滤波预处理, 把噪声的干扰降低到最小; 然后对数据进行归一化处理。

焦化装置结构如图 1 所示。该装置由 1 个分馏塔(T102)、1 个过程锅炉(T101/3)和 2 个焦化塔(T101/5, 6)组成。过程的工作流程描述如下: 原料渣油从焦化炉两侧分别送入加热炉对流室预热, 然后合并进入到分馏塔底, 与焦炭塔顶的油气进行热传递, 实现原料中轻组分和重组分的分离; 分馏塔底的重组分经加热炉辐射进料泵(102/1, 2, 3)分两路(FRC8107, FRC8108)进入加热炉辐射室(101/3)加热至 495 °C, 然后进入焦化

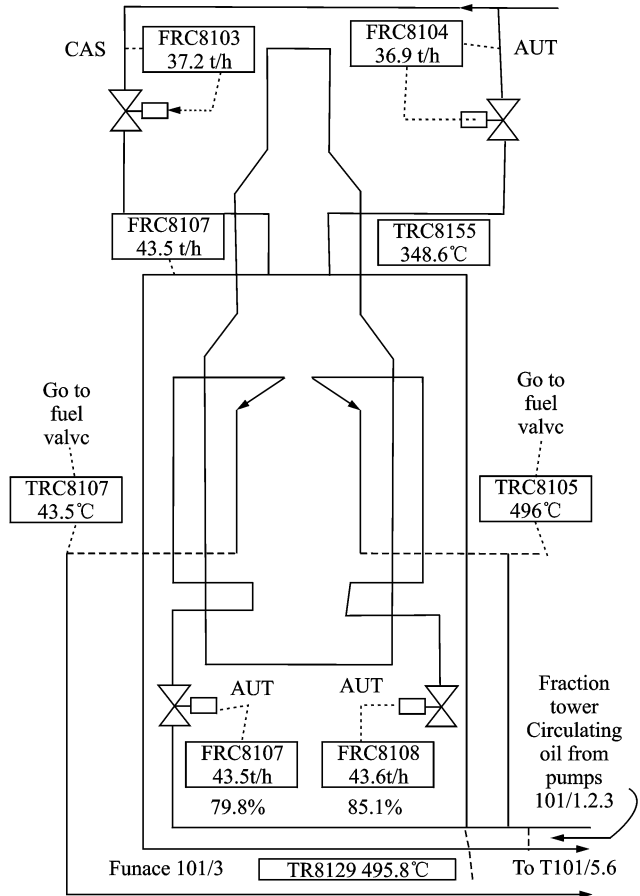


图 1 焦化炉流程图

Fig. 1 Flowchart of coking furnace

塔(T101/5)进行裂解反应。该装置一种复杂的间歇过程,在满足温度控制的前提下获得经济效益最大化。

本文控制目标是保持加热炉出口期望温度尽可能接近 495 °C。因此温度设定值为 495 °C 时,由实测过程的加热炉出口温度数据求取实际响应曲线图 2。从图 2 可以看出,该过程是一个大滞后、非线性、具有反向特性的对象。

对于局部在线自适应 WNN 算法,影响该算法速度的关键是候选隐含节点的参数确定,直接影响模型结构大小和模型参数调整时间。小波网络隐含节点的平移参数和尺度参数具有明确的物理意义。虽然径向基小波不是紧支撑,但是该函数可以快速衰减到 0。基于此性质,选择合适的小波参数使得聚类区域恰好分布在小波函数的紧半支撑域。

对 m 维输入的径向基小波函数 $\varphi_{\sigma,c}(\mathbf{x}, \mathbf{c}) = (n - \|\mathbf{x} - \mathbf{c}\|^2 / \sigma^2) \exp(-\|\mathbf{x} - \mathbf{c}\|^2 / (2\sigma^2))$, 其中 $\mathbf{x} = [x_1, x_2, \dots, x_m]^T, 0 \leq x_i \leq 1; \mathbf{c} = [c_1, c_2, \dots, c_m]^T$ 。假定伸缩因子 $0 \leq \sigma \leq 1$,小波函数半紧支撑界内数据满足 $|x_i - c_i| / \sigma \leq 3$,即 $-3\sigma + x_i \leq c_i \leq 3\sigma + x_i, i = 1, 2, \dots, m$ 。对中心 \mathbf{c}' 和半径 r 超球形的聚类,由 $|x_i - c_i| / |\sigma| \leq 3$ 可知尺度参数应满足 $\sigma \geq r/3$ 。因此对给定任意的聚类,可以根据聚类信息确定候选小波神经元数目以及伸缩、平移因子参数。

基于以上的讨论,设时滞为 50,本文使用以上 300 组数据,使用上面参数初始化方法和本文 WNN 学习方法辨识过程系统。在辨识过程中,LR-LSSVM 模型参数 γ 和 ν 分别在区间 $[1, 10]$ 及 $[1, 3]$ 内使用交叉验证方法确定;相似度矩阵由热核函数计算,即 $A_{i,j} = \exp(-\|(\mathbf{x}_i - \mathbf{x}_j)\|^2 / (\sigma_i \sigma_j))$,这里 σ_i 表示样本 \mathbf{x}_i 近邻的局部尺度,其值由 $\sigma_i = \|\mathbf{x}_i - \mathbf{x}_i^{(K)}\|$ 计算得到,其中 $\mathbf{x}_i^{(K)}$ 为 \mathbf{x}_i 的第 K 个近邻样本,一般情况下取 $K=8$ 就能达到很好的效果。基于 GK 的满意模糊聚类算法性能指标采用均方根误差 $RMSE =$

$$\sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2}$$

算法阈值参数设定为 0.23,初始化聚类数量设为 2。

为说明本文算法的有效性,本文给出在模型匹配和模型失配两种情况下的模型随动跟踪仿真结果。首先给出模型匹配时的 PFC 控制器的预测时域 $H=10$,遗忘因子 $\gamma=0.95$,图 3 给出本文算法控制在随动跟踪矩形波性能结果,这里矩形波为设定温度值。为说明本算法的有效性,图 4 给出广义预测控制 (Generalized predictive control, GPC) 控制效果,图 4 中方波为设定值,曲线为控制输出。从图 3,4 可以看出,本文算法所提出的算法明显优于 GPC 控制效果,说明本文所提出的新的 WNN 算法能够对复杂过程准确的建模。

下面讨论模型变化时的控制效果。对 WNN 进行线性化,使用 Pade 近似方法^[33] 计算得到线性模型为 $y_m(k) = a_1 y(k-1) + a_2 y(k-2) + b_1 u(k-51) + b_2 u(k-52)$ 。在 0 时刻,模型参数 a_1 改变为 $a_1 - 0.0075, b_1$ 变化为 $b_1 + 0.007$,其余参数不变;在 1200 时刻时,模型参数 a_1 改变为 $a_1 - 0.0075, b_1$ 变化为 $b_1 + 0.007, a_2$ 变为 $a_2/10$,系统模型结构和参数均发生变化。在上述两种情况下,本文算法和 GPC 的控制仿真效果如图 5,6 所示,图 6 中方波为设定值,曲线为控制输出。从仿真效果可以看出,本文算法

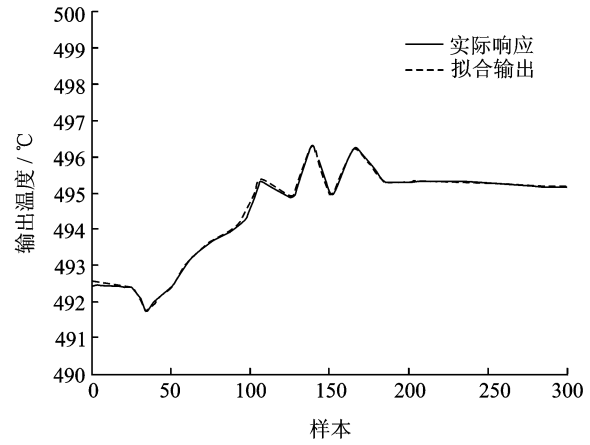


图 2 系统响应曲线和辨识结果
Fig.2 System response and the corresponding identification result

在模型参数和结构变化时均能自适应调整 WNN, 取得较为满意的控制效果。相比较而言, GPC 的控制效果不能令人满意。

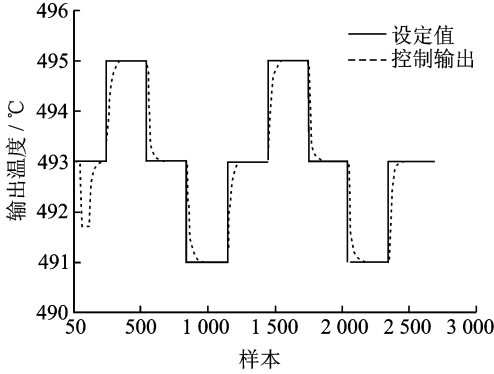


图3 本文算法矩形波设定值跟踪控制结果

Fig. 3 Tracking performance of proposed algorithm with rectangular setting temperature

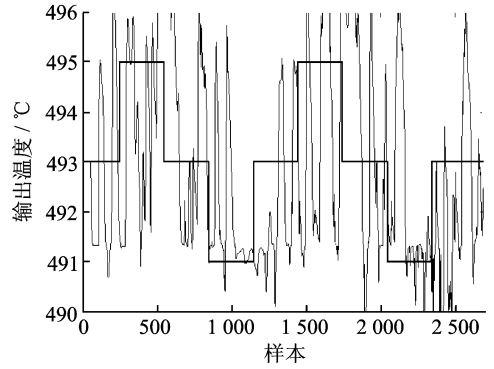


图4 GPC 矩形波设定值跟踪控制结果

Fig. 4 Tracking performance of GPC with rectangular setting temperature

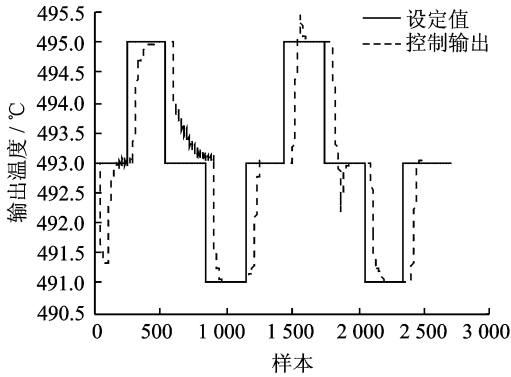


图5 模型参数和结构变化时本文算法控制效果

Fig. 5 Controlling performance of the proposed algorithm with model parameters and structure changing

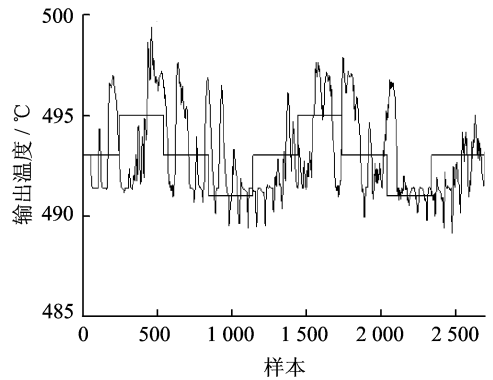


图6 模型参数和结构变化时 GPC 控制效果

Fig. 6 Controlling performance of GPC with model parameters and structure changing

6 结束语

本文基于集成学习、流形学习理论和 MDL 原理研究了在线 WNN 权重和结构自适应调整算法,并用于非线性预测控制中。该方法迭代地从 WNN 隐含节点候选集选取小波神经元并使用 EKF 方法自适应调整该节点的参数。基于 LR-LSSVM、最优实验原则 A-optimum 以及 MDL 准则实现 WNN 模型选择,较好地解决了时变动态、多工况非线性系统的辨识问题。基于在线 GK 满意模糊聚类算法,根据新样本在聚类空间的位置自动确定 WNN 的更新方法,克服传统在线更新判定方法需要人为确定额外参数的问题,且可以根据聚类结果构建候选节点参数值。实验结果表明,本文算法适合多工况非线性系统控制问题。虽然本文方法能够根据过程对象变化保证 WNN 模型的逼近精度,但是本文算法的鲁棒性以及推广到更多应用场合还需要进一步研究。

参考文献:

- [1] 席裕庚, 李德伟, 林姝. 模型预测控制——现状与挑战[J]. 自动化学报, 2013, 39(3): 222-236.
Xi Yugeng, Li Dewei, Lin Shu. Model predictive control—Status and challenges[J]. Acta Automatica Sinica, 2013, 39(3): 222-236.
- [2] Zhou Lei, Fei Shumin, Huang Jiakai, et al. A predictive functional control algorithm for multivariable systems with time delay [J]. Journal of Control Theory Application, 2013, 11(4): 656-660.
- [3] Wen S H, Yang J W, Rad A B, et al. Weighted multimodel predictive function control for automatic train operation system [EB/OL]. <http://dx.doi.org/10.1155/2014/520627>, 2014-05-04.
- [4] Heidarnejad M, Liu J, Christofides P D. Distributed model predictive control of switched nonlinear systems with scheduled mode transitions [J]. American Institute of Chemical Engineering Journal, 2013, 59: 860-871.
- [5] Song Yiming, Wang Xiaoxiao, Teng Haipeng, et al. Nonlinear parametric predictive control for the temperature control of bench-scale batch reactor[J]. Applied Thermal Engineering, 2016, 102: 134-143.
- [6] Tulsyan A, Huang B, Gopaluni B R, et al. On simultaneous on-line state and parameter estimation in non-linear state-space models [J]. Journal of Process Control, 2013, 23(4): 516-526.
- [7] 罗雄麟, 于洋, 许巍. 化工过程预测控制的在线优化实现机制[J]. 化工学报, 2014, 65(10): 3984-3992.
Luo Xionglin, Yu Yang, Xu Jun. Online optimization implementation on model predictive control in chemical process [J]. CIESC Journal, 2014, 65(10): 3984-3992.
- [8] Meng Yuebo, Zou Jianhua, Gan Xusheng, et al. Research on WNN aerodynamic modeling from flight data based on improved PSO algorithm[J]. Neurocomputing, 2012, 83: 212-221.
- [9] Mohsen Farahani. Intelligent control of SVC using wavelet neural network to enhance transient stability[J]. Engineering Applications of Artificial Intelligence, 2013, 26: 273-280.
- [10] Seyab R K A, Cao Y. Nonlinear system identification for predictive control using continuous time recurrent neural networks and automatic differentiation [J]. Journal of Process Control, 2008, 18(6): 568-581.
- [11] Maciej Lawryńczuk. Explicit nonlinear predictive control algorithms with neural approximation [J]. Neurocomputing, 2014, 129: 570-584.
- [12] Sevil A, Michail P, Alexandar I. Fuzzy model-based predictive control applied to multi-variable level control of multi-tank system[C]//IEEE International Conference on Intelligent Systems. London: IEEE, 2010: 456-461.
- [13] Maciej Lawryńczuk. Modelling and predictive control of a neutralisation reactor using sparse support vector machine Wiener models[J]. Neurocomputing, 2016, 205: 311-328.
- [14] Siam A, Brandon C, Prashant M, et al. Data-based modeling and control of Nylon-6, 6 batch polymerization [J]. IEEE Transactions on Control Systems Technology, 2013, 21(1): 94-106.
- [15] Na M G, Hwang I J, Lee Y J. Design of a fuzzy model predictive power controller for pressurized water reactors [J]. IEEE Transactions on Nuclear Science, 2006, 53(3): 1504-1514.
- [16] Peng H, Nakano K, Shioya H. Nonlinear predictive control using neural nets-based local linearization ARX model—Stability and industrial application[J]. IEEE Transactions on Control Systems Technology, 2007, 15(1): 130-143.
- [17] Xi Yugeng, Li Dewei, Lin Shu. Model predictive control—Status and challenges [J]. Acta Automatica Sinica, 2013, 39(3): 222-236.
- [18] Rodrigo L N, Fernando J D, Lorenz T B, et al. Fast nonlinear model predictive control: Formulation and industrial process applications [J]. Computers and Chemical Engineering, 2013, 51: 55-64.
- [19] Craven S, Whelan J, Glennon B. Glucose concentration control of a fed-batch mammalian cell bioprocess using a nonlinear model predictive controller [J]. Journal of Process Control, 2014, 24: 344-357.
- [20] Findeisen R, Allgöwer F, Biegler L T. Assessment and future directions of NMPC [M]. Berlin: Springer, 2007.
- [21] Yang X, Biegler L T. Advanced-multi-step nonlinear model predictive control [J]. Journal of Process Control, 2013, 23: 1116-1128.
- [22] Wei H L, Billings S A, Zhao Y, et al. Lattice dynamical wavelet neural networks implemented using particle swarm optimization for spatio-temporal system identification [J]. IEEE Transactions on Neural Networks, 2009, 20(1): 181-185.
- [23] Lary D J, Mussa H Y. Using an extended Kalman filter learning algorithm for feed-forward neural networks to describe tracer correlations [J]. Atmos Chemical Physics Discussion, 2004, 4: 3653-3667.
- [24] Alessandri A, Cuneo M, Pagnan S, et al. On the convergence of EKF based parameters optimization for neural networks[C]//Proc of 42nd IEEE Conference on Decision and Control. [S.l.]: IEEE, 2003, 6: 6181-6186.
- [25] He Xiaofei, Li Ming, Zhang Chiyuan, et al. A variance minimization criterion to feature selection using Laplacian regularization

- [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(10):2013-2025.
- [26] Atkinson A C, Donev A N. *Optimum experimental designs* [M]. Oxford: Oxford University Press, 2007.
- [27] 孙卫红, 童晓, 李强. 改进 PSO 优化参数的 LSSVM 燃煤锅炉 NO_x 排放预测[J]. *数据采集与处理*, 2015, 30(1):231-238.
Sun Weihong, Tong Xiao, Li Qiang. NO_x emission concentration of coal-fired boiler prediction based on improved PSO parameter optimized LSSVM[J]. *Journal of Data Acquisition and Processing*, 2015, 30(1):231-238.
- [28] He Xiaofei, Ji Ming, Zhang Chiyuan, et al. A variance minimization criterion to feature selection using Laplacian regularization[J]. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2011, 33(10):2013-2025.
- [29] 张贤达. *矩阵分析与应用*[M]. 北京: 清华大学出版社, 2004.
Zhang Xianda. *Matrix analysis and application*[M]. Beijing: Tsinghua University Press, 2004.
- [30] 黄桂根, 孟兵, 李品. 一种基于 MDL 原理的信号分选新算法[J]. *数据采集与处理*, 2015, 30(3):656-668.
Huang Guigen, Meng Bing, Li Pin. New radar signal deinterleaving algorithm based on MDL criterion[J]. *Journal of Data Acquisition and Processing*, 2015, 30(3):656-668.
- [31] 谢磊, 梁武星, 张泉灵. 基于快速滑窗 QR 分解的自适应子空间辨识[J]. *化工学报*, 2009, 59(6):1448-1453.
Xie Lei, Liang Wuxing, Zhang Quanling. Adaptive subspace identification based fast moving window QR decomposition [J]. *CIESC Journal*, 2009, 59(6):1448-1453.
- [32] 李丽娟, 苏宏业, 褚健. 基于在线最小二乘支持向量机的广义预测控制[J]. *自动化学报*, 2007, 33(11):1182-1188.
Li Lijuan, Su Hongye, Cu Jian. Online least square support vector machine-based generalized predictive control [J]. *Acta Automatica Sinica*, 2007, 33(11):1182-1188.
- [33] 李少远. *全局工况系统的预测控制及其应用*[M]. 北京: 科学出版社, 2007.
Li Shaoyuan. *Predictive control for global operating conditions system and its applications* [M]. Beijing: Science Press, 2007.
- [34] 张日东, 王树青. 基于神经网络的非线性系统预测函数控制 [J]. *控制理论与应用*, 2006, 24(6):949-953.
Zhang Ridong, Wang Shuqing. Neural network-based predictive control for nonlinear system [J]. *Control Theory & Application*, 2006, 24(6):949-953.
- [35] 苏成利, 李平, 邓淑贤, 等. 一种基于 T-S 模糊模型的自适应预测函数控制[J]. *控制工程*, 2007, 14(6):610-613.
Su Chengli, Li Ping, Deng Shuxian, et al. A T-S fuzzy model-based adaptive predictive control[J]. *Control Engineering*, 2007, 14(6):610-613.

作者简介:



任世锦 (1972-), 男, 博士, 副教授, 研究方向: 复杂工业过程建模、监测及机器学习, E-mail: sjren_phd@163.com。



王高峰 (1976-), 男, 博士研究生, 讲师, 研究方向: 机器学习。



李新玉 (1968-), 男, 博士研究生, 高级工程师, 研究方向: 机械故障检测。



杨茂云 (1968-), 男, 博士研究生, 讲师, 研究方向: 机械信号处理、概率图理论。



徐桂云 (1961-), 女, 教授, 研究方向: 机器人机构学、机器人动力学。

