

面向建模误差 PDF 形状与趋势拟合优度的动态过程优化建模

周平¹ 赵向志¹

摘要 本文提出一种新的数据驱动建模思路及方法,即面向建模误差概率密度函数(Probability density function, PDF)形状与趋势拟合优度(相似度)的动态过程多目标优化建模方法.首先,针对均方根误差(Root mean square error, RMSE)等常规一维性能指标不能完全刻画建模误差在时间和空间二维随机动态特性的问题,引入 PDF 指标来对动态过程的建模误差在时间和空间进行二维尺度的全面刻画和评价,并采用核密度估计技术对关于时间的建模误差序列的 PDF 进行估计;其次,为了刻画动态过程数据建模的拟合趋势,进一步引入趋势拟合优度指标,从而构造综合描述数据建模误差 PDF 形状与趋势拟合相似性的多目标性能指标;在此基础上,采用 NSGA-II 算法优化数据模型的参数集,获取一大类满足上述多目标性能优化的智能模型参数解.数值仿真及工业数据验证表明,所提方法的建模误差 PDF 逼近设定的期望 PDF,并且模型输出与样本数据拟合趋势接近,好于常规最小化一维 RMSE 指标的数据建模方法.

关键词 建模误差 PDF, 拟合优度, 数据建模, 核密度估计, 多目标优化, 污水处理

引用格式 周平, 赵向志. 面向建模误差 PDF 形状与趋势拟合优度的动态过程优化建模. 自动化学报, 2021, 47(10): 2402-2411

DOI 10.16383/j.aas.c200001

Optimized Modeling of Dynamic Process Oriented Towards Modeling Error PDF Shape and Goodness of Fit

ZHOU Ping¹ ZHAO Xiang-Zhi¹

Abstract This paper proposes a novel data-driven modeling method, which is a multi-objective optimized modeling method for dynamic process oriented towards modeling error probability density function (PDF) shape and goodness of fit (similarity). First, aiming at the problem that the conventional modeling performance indicators such as the root mean square error (RMSE) cannot fully characterize the two-dimensional stochastic dynamic characteristics of modeling errors. The PDF index is introduced to comprehensively characterize and evaluate the modeling errors of dynamic systems in two dimensions on time and space, while the kernel density estimation technology is used to estimate the PDF of modeling error sequence. Second, in order to characterize the fitting trend of dynamic process data modeling, the goodness of fit is further introduced to construct a multi-objective performance indicator that comprehensively describes the data modeling error PDF shape and trend fitting similarity. Based on this, the parameter set of the data model is optimized using the NSGA-II algorithm to obtain the optimized parameter solutions for a large class of intelligent models. Finally, numerical simulation and industrial data verification show that the modeling error PDF of the proposed method approximates the set target PDF, and the model output is close to the actual data fitting trend, which is better than the conventional data modeling methods of minimizing the one-dimensional RMSE index.

Key words Modeling error PDF, goodness of fit, data modeling, kernel density estimation (KDE), multi-objective optimization, sewage treatment

Citation Zhou Ping, Zhao Xiang-Zhi. Optimized modeling of dynamic process oriented towards modeling error PDF shape and goodness of fit. *Acta Automatica Sinica*, 2021, 47(10): 2402-2411

收稿日期 2020-01-01 录用日期 2020-04-10

Manuscript received January 1, 2020; accepted April 10, 2020

国家自然科学基金项目(61890934, 61790572), 辽宁省“兴辽英才计划”项目(XLYC1907132), 央高校基本科研业务费项目(N180802003)

Supported by National Natural Science Foundation of China (61890934, 61790572), Liaoning Revitalization Talents Program (XLYC1907132), Fundamental Research Funds for the Central Universities (N180802003)

本文责任编辑 郭戈

Recommended by Associate Editor GUO Ge

1. 东北大学流程工业综合自动化国家重点实验室 沈阳 110819

1. State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110819

实际复杂工业过程如高炉炼铁过程、磨矿过程、造纸制浆过程、污水处理过程等,通常涉及复杂的物理化学反应,具有多变量、强耦合、非线性、大滞后等综合复杂动态特性,利用传统的物理化学等机理方法时,难以建立精确的数学模型^[1-4].近年来,随着大数据和人工智能技术的发展,对于难以进行机理建模的工业过程,甚至缺乏机理模型的复杂工业系统,数据驱动建模常被看作是一种非常有效的建模方法^[1,4-5].目前,数据驱动建模主要采用人工智能

技术或多元统计分析技术来描述过程输入与输出之间的复杂未知动态关系, 在此基础上建立具有一定结构和适当模型参数的过程数据模型. 由于数据模型输出与实际过程输出之间存在一定的偏差, 为了最优化模型性能, 通常需要采用相关算法来优化关于建模误差的性能指标, 如均方根误差 (Root mean square error, RMSE)、均方差 (MSE) 及平均绝对误差 (Mean absolute error, MAE) 等, 以此获得满意的数据模型参数^[6].

在实际工业过程建模时, 单纯的 RMSE、MSE、MAE 等性能指标均以最小化一维的统计建模误差均值为目标, 并不能完全描述和刻画动态系统建模过程的随机性和不确定性^[6]. 在复杂工业过程中, 外部不确定因素和随机动态干扰往往具有非高斯特性, 将其假设为高斯分布特征时, 难以获得满意的建模效果^[7]. 另一方面, 复杂工业过程中, 其建模误差通常是一个未知的随机变量, 因而建模误差在时空尺度上的二维概率密度函数 (Probability density function, PDF) 形状分布能够包含动态系统建模误差的所有分布和统计信息. 因此, 误差 PDF 形状优化的思想受到了越来越广泛的关注, 并逐渐用于复杂工业过程的数据建模与控制中, 如以非高斯动态系统 PDF 形状为目标的随机分布控制已成为解决随机动态系统控制的非常有效的方法, 得到了广泛的应用^[7-8]. 近年, 面向建模误差 PDF 形状优化的数据建模方法也逐渐引起重视. 文献 [6] 中, 作为有界随机分布系统建模与控制方法的扩展, Zhou 等将输出 PDF 控制或随机分布控制思想引入到过程建模中. 通过优化建模误差 PDF 形状, 进而对模型参数求解, 使得实际建模误差 PDF 形状接近设定的期望 PDF 形状^[6]. 该方法不仅可以获得较为满意的建模效果, 而且一定程度降低建模过程中的随机性和不确定性. 此外, 文献 [9] 通过优化建模误差 PDF 形状, 建立了选矿过程精矿品位的最小支持向量机模型. 而文献 [10] 通过优化建模误差 PDF 形状, 间接对模型参数进行调节, 建立了间歇过程的模糊神经网络模型.

上述方法均是期望实际模型的建模误差 PDF 形状更好地跟踪期望的高斯分布形状, 以此建立具有最优参数的过程数据模型. 然而, 不管是常规建模方法的误差 RMSE 指标, 还是上述改进方法提到的建模误差 PDF 指标, 均仅仅体现过程模型输出与实际输出之间的误差大小情况, 难以衡量模型输出与实际动态过程输出之间拟合趋势是否一致. 实际上, 实际工业动态系统中, 过程输出变化趋势的估计和预测, 对于基于模型的预测控制、生产过程运行态势的把握与调控等诸多工程应用, 都具有十

分重要的作用. 因此, 在动态系统建模时, 除了需要优化建模误差的 PDF 形状, 同时也需要考虑建模输出与样本数据之间拟合趋势最接近, 即曲线拟合动态变化趋势的相似度最大^[11].

针对上述动态系统建模的实际需求和现有方法的不足, 本文以小波神经网络 (Wavelet neural network, WNN)^[12-13] 数据建模为例, 提出一种新型的面向建模误差 PDF 形状与趋势拟合优度 (相似度) 多目标优化的动态系统数据建模方法. 所提方法不仅引入二维尺度的 PDF 指标来对动态建模误差在时间和空间进行全面刻画, 同时引入拟合优度 (相似度) 指标^[11, 14] 刻画动态系统数据建模的拟合趋势. 通过采用核密度估计 (Kernel density estimation, KDE)^[15-17] 技术对实际建模误差 PDF 形状进行估计, 以及采用 NSGA-II 算法^[18] 对建模误差 PDF 形状的偏差以及拟合优度指标进行多目标优化, 从而建立具有最优模型参数的 WNN 模型. 数值仿真以及污水处理过程^[19-20] 数据验证表明所提方法的实际建模误差 PDF 能够更好地逼近设定的期望 PDF, 并且模型输出与样本数据拟合趋势接近.

1 建模思路与策略

1.1 WNN 简介

小波神经网络是结合小波分析与神经网络的一种前馈型网络. WNN 用小波函数代替传统 Sigmoid 函数作为激励函数, 通过仿射变换建立起小波变换与网络参数之间的连接, 能以任意精度对函数进行逼近^[12-13]. 如图 1 所示, WNN 通常采用三层网络结构, 其中: x_1, x_2, \dots, x_M 为 WNN 的输入变量, y_1, y_2, \dots, y_N 为输出变量, $\omega_{I,j}$ 和 $\omega_{H,l}$ 分别是输入层到隐含层的连接权值以及隐含层到输出层的连接权值, $\theta_{H,j}, \theta_{O,l}$ 分别为隐含层和输出层节点阈值, a_j 和 b_j 分别为小波基函数的伸缩因子和平移因子.

WNN 隐层激励函数 $\phi(t)$ 通常采用如下的 Morlet 母小波函数^[12]:

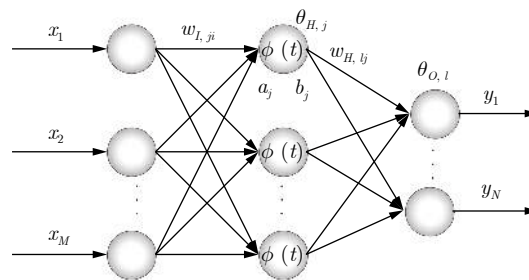


图 1 WNN 结构图

Fig. 1 Structure diagram of WNN

$$\phi(t) = \cos(1.75t) \exp\left(-\frac{t^2}{2}\right) \quad (1)$$

输出层激励函数 $f(t)$ 则采用 Sigmoid 函数, 即:

$$f(t) = (1 + e^{-t})^{-1} \quad (2)$$

此外, 定义三层 WNN 的输入层、隐含层、输出层节点数分别为 M, n 和 N , 则隐含层第 j 个节点的输入 net_j 和输出 z_j 分别为:

$$net_j = \sum_i \omega_{I,j,i} x_i - \theta_{H,j}, i = 1, \dots, M, \quad j = 1, 2, \dots, n \quad (3)$$

$$z_j = \varphi\left(\frac{net_j - b_j}{a_j}\right), \quad j = 1, \dots, n \quad (4)$$

式中, b_j 为小波基函数 $\varphi(t)$ 的平移因子, a_j 为小波基函数的伸缩因子, $\varphi(t)$ 为 Morlet 母小波函数, $\theta_{H,j}$ 为隐含层节点的阈值, 则 WNN 的最终输出为:

$$\hat{y}_l = f\left(\sum_j \omega_{H,l,j} z_j - \theta_{O,l}\right), \quad j = 1, \dots, n, l = 1, \dots, N \quad (5)$$

式中, $\theta_{O,l}$ 为输出层节点的阈值, 函数 $f(\cdot)$ 为输出层的激励函数.

1.2 建模策略

对于常规 WNN 等现有多数建模方法, 通常采

用如下均方根误差 (RMSE)、均方差 (MSE)、平均绝对误差 (MAE) 等单一的误差性能指标, 通过性能指标数值大小来评价建模精度.

$$\begin{cases} J_{\text{RMSE}} = \sqrt{\frac{1}{m} \sum_{l=1}^m (y_l - \hat{y}_l)^2} \\ J_{\text{MSE}} = \frac{1}{m} \sum_{l=1}^m (y_l - \hat{y}_l)^2 \\ J_{\text{MAE}} = \frac{1}{m} \sum_{l=1}^m |y_l - \hat{y}_l| \end{cases} \quad (6)$$

然而, 式 (6) 所示传统性能指标是从建模误差的均值角度评价模型精度, 并不能全面描述动态系统建模误差在时空尺度上的随机特性. 此外, 对于时序相关动态系统建模, 运行数据拟合趋势的估计对于建模效果有很大影响, 并且更有实际意义. 而式 (6) 所示常规建模性能评价指标仅希望建模输出与实际数据之间偏差最小, 却难以描述动态系统的拟合趋势好坏.

为了解决上述问题, 本文以 WNN 智能建模为基础, 通过引入建模误差概率密度函数 (PDF) 指标从时空二维角度对建模误差进行全面刻画, 以及引入拟合优度指标对动态系统数据建模的拟合趋势进行相似性评估, 从而提出图 2 所示的面向建模误差 PDF 形状与趋势拟合优度的动态系统优化建模方

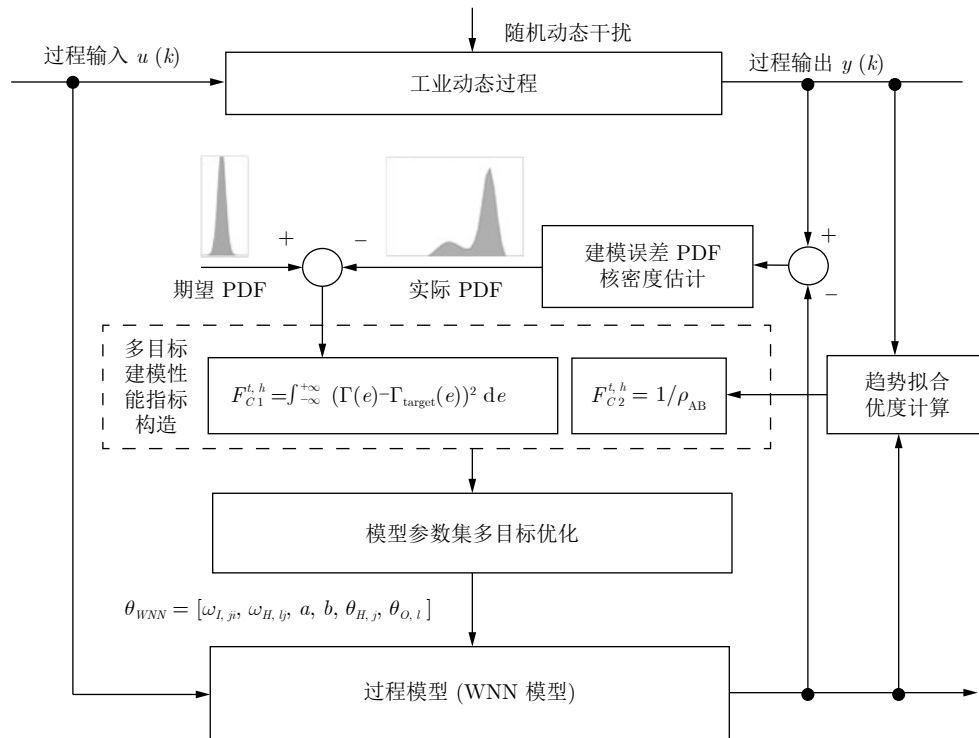


图 2 面向建模误差 PDF 形状与趋势拟合优度的优化建模策略

Fig. 2 Optimized modeling strategy towards modeling error PDF shape and goodness of fit

法, 具体如下:

1) 首先, 构建动态系统数据建模的实际建模误差 PDF 与期望建模误差 PDF 的偏差平方积分作为多目标优化计算的第一个评价指标, 如下所示:

$$J_1 = \int_{-\infty}^{+\infty} (\Gamma(e) - \Gamma_{\text{target}}(e))^2 de \quad (7)$$

式中, $\Gamma(e)$ 和 $\Gamma_{\text{target}}(e)$ 分别为实际建模误差 PDF 和期望建模误差 PDF. 本文实际建模误差 PDF 是采用核密度估计技术对所建立数据模型的建模误差序列进行求解获得, 而期望建模误差 PDF 是设置的一个较为理想的 (即均值为 0、方差尽量小) 高斯分布形状的 PDF, e 为建模误差 PDF 的自变量.

2) 其次, 引入式 (8) 所示的拟合优度指标 ρ_{AB} ^[11,14] 对动态系统数据建模的动态拟合趋势进行相似性评估, 然后构建式 (9) 所示关于拟合优度的性能指标作为第二个评价指标.

$$\rho_{AB} = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{\left(\sum_m \sum_n (A_{mn} - \bar{A})^2\right) \left(\sum_m \sum_n (B_{mn} - \bar{B})^2\right)}} \quad (8)$$

$$J_2 = \frac{1}{\rho_{AB}} \quad (9)$$

式中, A, B 为两个数据矩阵, \bar{A}, \bar{B} 分别为数据矩阵 A, B 的均值. 事实上, ρ_{AB} 是衡量数据矩阵 A 和 B 之间近似程度的量, $|\rho_{AB}| \rightarrow 1$ 表示数据矩阵 A 和 B 之间相关性很强, 而 $|\rho_{AB}| \rightarrow 0$ 意味着数据矩阵 A 和 B 之间相关性较弱. 由于本文要衡量建模输出与实际输出的时序相关数据之间的动态拟合趋势, 所以本文式 (8) 中 A 和 B 分别表示小波神经网络模型输出和实际过程输出所构成的时序相关数据矩阵.

3) 最后, 分别将式 (8) 和式 (9) 作为数据建模的综合性能评价指标的适应度函数, 采用运算速度快、解集收敛性好的 NSGA-II 算法^[11] 来获得 WNN 模型的最优参数集 $[\omega_{I,j_i}, \omega_{H,l_j}, a_j, b_j, \theta_{H,j}, \theta_{O,l}]$.

2 建模算法

所提方法的具体建模算法包括如下几个过程: 首先, 采用式 (1) ~ (5) 所示算法构建初始的 WNN 数据模型, 通过比较 WNN 数据模型输出与过程输出或者相应的实际值, 可以得到特定时间内的建模误差序列. 然后, 采用第 3.1 节的核密度估计技术对实际建模误差 PDF 进行计算. 最后, 采用 NSGA-

II 算法优化式 (8) 和式 (9) 所示的多目标性能指标, 获得同时具有较好建模误差 PDF 形状与拟合优度值的多组 WNN 模型参数集解.

2.1 建模误差 PDF 核密度估计

核密度估计 (Kernel density estimation, KDE) 是由 Parzen 提出的一种非参数估计方法^[15-17], 用于求解给定随机变量数据集分布的概率密度函数. 假设 $x_i \in R, i = 1, \dots, n$ 为独立同分布的随机变量数据集, 其所服从的分布密度函数为 $f(x), x \in R$, 则 $f(x)$ 的核密度估计 $\hat{f}_h(x)$ 定义如下:

$$\hat{f}_h(x) = \frac{1}{nh_p} \sum_{i=1}^n \phi\left(\frac{x_i - x}{h_p}\right), \quad x \in R \quad (10)$$

式中, 窗宽 h_p 是一个给定的正数, $\phi(x)$ 为核函数, n 为样本数.

对于所提建模方法, 采用 KDE 对建模误差 PDF 进行估计, 可以得到估计的建模误差概率密度函数 Γ_e 为:

$$\Gamma_e = \frac{1}{K} \sum_{k=1}^K \frac{1}{h_p} \phi\left(-\frac{e - e(k)}{h_p}\right) \quad (11)$$

式中, K 为建模误差样本数目, 通过设置期望的建模误差 PDF, 可以构造式 (7) 所示的性能指标. 式 (11) 所示基于 KDE 的 WNN 实际建模误差 PDF 估计求解步骤如下:

1) 选择核函数: 在估计随机变量未知概率密度函数时, 常用的核函数有高斯核函数、矩形窗核函数、Epanechnikov 核函数等. 核函数的不同选择在 KDE 中不敏感, 当样本数据很大时, 对核函数密度估计的结果影响不大. 本文选取高斯核函数, 其表达式如下:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\|x\|^2}{2}} \quad (12)$$

2) 选择窗宽: 窗宽 h_p 的选择对核函数的密度估计起着局部光滑的作用, 如果 h_p 过大会使模型误差 PDF 形状很光滑, 使其主要部分的某些特征 (如多峰性) 被掩盖起来, 从而增加估计量的偏差; 而若 h_p 过小, 则整个密度函数表现粗糙. 本文基于正态参照规则方法^[17] 进行窗宽选择, 假设建模误差服从正态分布, 则窗宽 h_p 设置为 $h_p = 1.06\sigma K^{-1/5}$ 其中 σ 由 $\min\{s, Q/1.34\}$ 估计, s 表示样本标准差, Q 为四分位数间距.

3) 求解模型误差 PDF: 根据步骤 1) 和 2) 选择合适窗函数和窗宽参数, 然后代入式 (11), 可以得到 WNN 建模误差 PDF 函数的估计值为:

$$\Gamma_e = (x(k), y(k), \theta_{WNN}, \varphi(e), h_p) = \frac{1}{K} \sum_{k=1}^K \frac{1}{h_p} \varphi \left(-\frac{e - \Theta(x(k), y(k), \theta_{WNN})}{h_p} \right) \quad (13)$$

2.2 模型参数多目标优化求解

式(1)~(5)所示基本WNN数据模型的参数主要包括:输入层连接权值 $\omega_{I,ji}$ 、隐含层连接权值 $\omega_{H,lj}$ 、隐含层阈值 $\theta_{H,j}$ 、输出层阈值 $\theta_{O,l}$ 、小波基函数的伸缩因子 a_j 以及平移因子 b_j .这些参数的取值直接决定了WNN数据模型的性能,因而基于前述构建的多目标建模性能指标,采用NSGA-II算法对模型参数进行优化,步骤如下:

1) 网络参数的编码.将WNN模型参数集 $\theta_{WNN} = [\omega_{I,ji}, \omega_{H,lj}, a_j, b_j, \theta_{H,j}, \theta_{O,l}]$ 与每条染色体相对应,即对WNN模型参数进行如下形式的编码:

$$R = [\omega_{I,11}, \dots, \omega_{I,1n}, \dots, \omega_{I,Mn}, \omega_{H,11}, \dots, \omega_{H,1m}, \dots, \omega_{H,nN}, a_1, \dots, a_n, b_1, \dots, b_n, \theta_{H,1}, \dots, \theta_{H,n}, \theta_{O,1}, \dots, \theta_{O,n}]$$

式中,染色体基因数为 $S = (M + 3)n + (n + 1)N$, $P = [S_1, S_2, \dots, S_i, \dots, S_Q]^T$ 表示包含 Q 条染色体的初始种群.

2) 个体适应度计算.每条染色体中的各个基因分别代表WNN的各个参数,将第 t 代种群中第 h 条染色体上的各个基因代入下式的第 t 代、第 h 个个体多适应度函数中:

$$F_{C1}^{t,h} = \int_{-\infty}^{+\infty} (\Gamma(e) - \Gamma_{\text{target}}(e))^2 de \quad (14)$$

$$F_{C2}^{t,h} = \frac{1}{\rho_{AB}} \quad (15)$$

3) 选择算子.根据非支配排序结果,选择非支配排序中支配层较低的个体.如果有多个个体在同一支配层,从种群多样性角度考虑,选择拥挤度距离较大的个体.

4) 模拟二进制交叉.基于实数编码,交叉后代为父代的线性组合,即

$$G_{1,i}^{t+1} = 0.5 [(1 - \beta_k(\varepsilon))G_{1,i}^t + (1 + \beta_k(\varepsilon))G_{2,i}^t] \\ G_{2,i}^{t+1} = 0.5 [(1 + \beta_k(\varepsilon))G_{1,i}^t + (1 - \beta_k(\varepsilon))G_{2,i}^t]$$

式中, ε 为在(0,1)内服从均匀分布的随机数.当 $\varepsilon > 0.5$ 时 $\beta_k(\varepsilon) = ([2(1 - \varepsilon)]^{(\eta_c+1)^{-1}})^{-1}$,当 $\varepsilon \leq 0.5$ 时, $\beta_k(\varepsilon) = (2\varepsilon)^{(\eta_c+1)^{-1}}$, η_c 为交叉分布指数, $i = 1, 2$ 为目标函数的个数.

5) 多项式变异.二进制交叉后,进行多项式变

异,变异后的个体为:

$$G_i^{t+1} = G_i^{t+1} + (B^U - B^L)\delta_k$$

式中, B^U, B^L 分别为优化变量的上、下限, δ_k 为变异参数.当 $r_k > 0.5$ 时, $\delta_k = (2r_k)^{(\eta_m+1)^{-1}}$,而当 $r_k \leq 0.5$ 时, $\delta_k = (1 - [2(1 - r_k)])^{(\eta_m+1)^{-1}}$, r_k 为在(0,1)服从均匀分布的随机数, η_m 为变异分布指数.

采用NSGA-II算法优化WNN模型参数时,每个待优化的参数对应染色体上的一个基因.在遗传算法中,适应度函数的选择决定着遗传优化的精度和收敛速度.描述个体性能的指标主要通过适应度函数值体现,依据适应度值的大小对个体进行优胜劣汰.本文多目标适应度函数为实际建模误差PDF与期望建模误差PDF之间的二维偏差平方和以及趋势拟合优度的倒数,并通过基因之间的选择、二进制交叉、变异产生最优个体即最优模型参数.

3 数值仿真及工业数据验证

3.1 数值仿真

为了验证所提方法的有效性和优越性,首先使用下述两输入一输出非线性动态系统进行数值验证:

$$y(k+1) = u^3(k) + \frac{y(k)}{1 + y^2(k)} + \omega(k) \quad (16)$$

式中, $y(0) = 0.1, u(k)$ 为在区间(0, 1)内服从均匀分布的随机序列, $\omega(k)$ 为通过参数 σ 描述的、服从瑞利分布的非高斯随机干扰序列.针对以上非线性系统,利用提出的建模方法进行建模,所要建立的WNN数据模型可以表示为:

$$\tilde{y}(k+1) = f_{WNN}(y(k), u(k), \omega(k), \theta_{WNN})$$

假设 ω 为随机产生、服从瑞利分布且参数为0.2的非高斯干扰.WNN隐层节点数选择为6,迭代优化步长 r 为0.003.采用NSGA-II算法对WNN模型参数进行寻优时,交叉分布指数 $\eta_c = 20$,变异分布指数 $\eta_m = 20$,优化变量的上限与下限分别设定为1和-1,交叉率和突变率分别设为0.9和0.1.

建模后,得到60组Pareto前沿解进化过程如图3所示,而图4为60组多目标优化解对应的拟合优度变化曲线,这里将所提方法与常规WNN方法以及近年文献[6]中提出的面向建模误差PDF优化的WNN方法进行比较.由于文献[6]是采用梯度下降方法来优化WNN模型的建模误差PDF,因而本文将其称为GD-WNN.从图4可以看出,采用所提建模方法可以获得具有较大动态变化趋势拟合优度的一组解集,这些解对应的拟合优度均远好

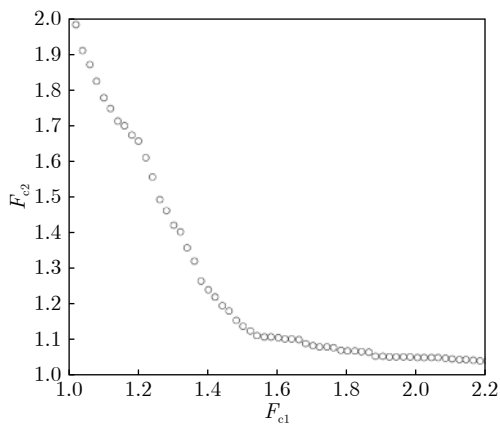


图 3 Pareto 前沿解进化过程

Fig.3 Evolution process of Pareto front

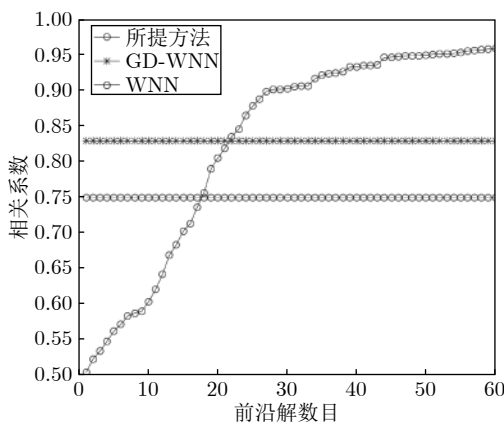


图 4 不同优化解对应的拟合优度值变化曲线

Fig.4 Change curve of goodness of fit corresponding to different optimization solutions

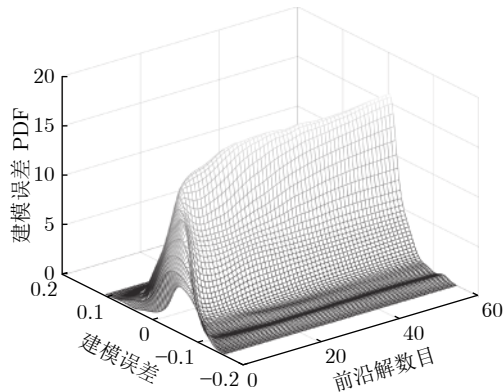


图 5 不同优化解对应的建模误差 PDF 变化曲面

Fig.5 PDF changing surface corresponding to different optimization solutions

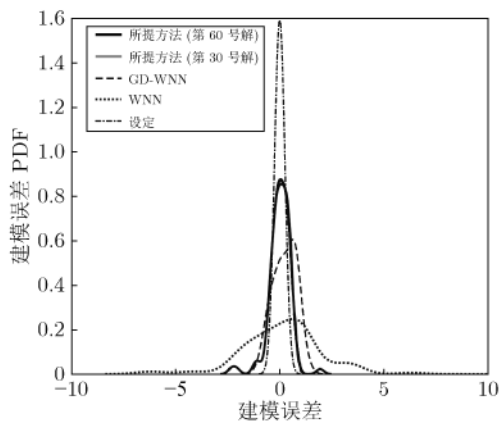


图 6 不同方法建模误差 PDF 比较

Fig.6 Comparison of modeling error PDF with different methods

于常规 WNN 方法以及 GD-WNN 方法. 所提方法得到的解对应的拟合优度指标最高达到 0.96, 而文献 [6] 中方法的拟合优度指标仅为 0.83, 以及常规 WNN 方法的拟合优度指标甚至仅为 0.75. 并且图 4 还可以看出所提方法有 39 组解的拟合优化度指标好于文献 [6] 中方法得到的拟合优度值. 图 5 是图 3 中 1 到 60 号解对应的建模误差 PDF 曲线变化图, 可以看出 60 号解对应模型的建模误差 PDF 最好. 图 6 为多目标优化后 30 号解与 60 号解的建模误差 PDF 曲线与其他两种现有方法的建模误差 PDF 曲线的对比图, 本文设置的期望 PDF 为均值为 0 方差为 0.25 的高斯型概率密度函数.

可以看出, 本文方法获得的非最优 30 号解对应的建模误差 PDF 曲线也要远好于常规 WNN 方法和 GD-WNN 方法. 所提方法得到的建模误差 PDF 较高且较窄, 与其他方法相比方差更小, 即模

型的随机性和不确定性更小, 这也表明所提方法的有效性和优越性. 图 7 和图 8 分别是本文方法中非最优的 30 号解对应的建模效果和新样本测试效果, 可以看出所提方法得到非最优解对应的模型不论是建模和新样本测试均好于其他两种方法.

3.2 活性污泥污水处理过程生化反应池出水 COD 含量建模

目前, 城市污水处理广泛采用活性污泥法^[3, 19-22]. 图 9 为典型活性污泥污水处理的工艺流程图, 主要包括三个级别的处理过程. 一级处理主要进行物理反应, 除去原生污水中的悬浮固体. 二级处理过程包括曝气池及二沉池处理两个部分. 曝气池是污水处理的核心部分, 主要进行微生物自身的代谢活动, 从而达到对污水中有机污染物如氮、磷的去除以及有氧生物的降解.

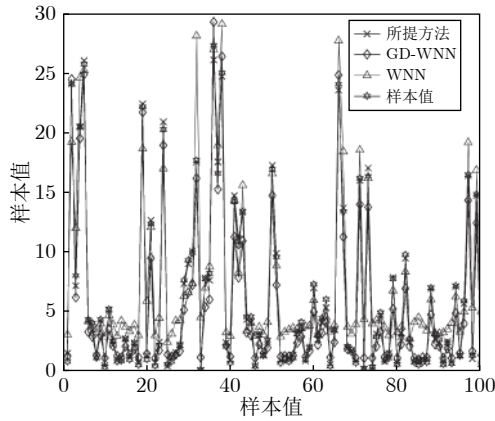


图 7 所提方法 30 号优化解对应的建模效果

Fig. 7 Modeling result corresponding to the 30th optimization solution of the proposed method

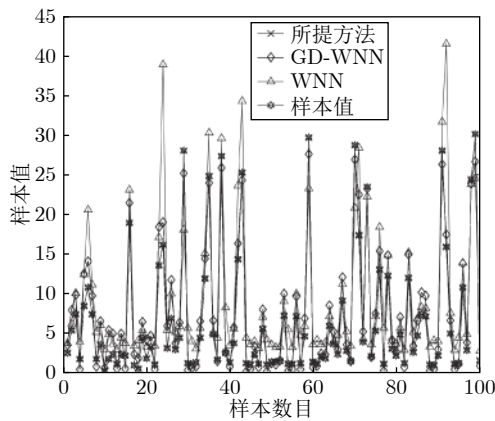


图 8 所提方法 30 号优化解对应的测试效果

Fig. 8 Testing result corresponding to the 30th optimization solution of the proposed method

经过曝气池处理后的污水流入二沉池进行固液分离,上层是澄清的液体,下层的污泥一部分回流至曝气池,以维持曝气池内的污泥浓度,另一部分污泥排出系统. 经过二级处理后的污水进入三级处理过程,通过加入药剂进而得到达标的出水. 判断水质是否达标主要通过水质参数进行衡量. 在众多水质参数指标中,出水 COD 含量不仅代表污水中含有的有机物的量,同时还包括污水中还原性无机物被氧化时所消耗的氧气量. COD 数值越小,在氧化过程中氧气的消耗量就越少,即水体中有机物的量越少. 因此,该指标能够反映有机污染物受纳的程度,是非常重要的出水水质指标. 虽然目前有许多 COD 含量的在线检测仪,但是都存在检测周期长、价格昂贵的问题. 所以,通过基于数据的智能建模技术来预测曝气池出水 COD 含量对于判断出水水质是否达标具有重要意义.

影响污水出水 COD 含量的参数较多,包括:入水流量 (Q)、入水化学需氧量 (COD)、溶解氧浓度 (DO)、污泥浓度 (MLSS)、悬浮固体浓度 (SS)、出水 PH 值等. 为此,根据过程机理分析,确定出水 COD 预测建模的输入变量为:入水 COD、入水流量 (Q) 和污泥浓度 (MLSS). 为消除变量间的量纲影响,建模所用训练与测试数据都归一化处理. 设定 WNN 隐层节点数为 6,迭代优化步长为 0.003,得到的 100 组多目标优化 Pareto 前沿解如图 10 所示. 图 11 和图 12 分别为 100 组多目标优化解对应的拟合优度变化曲线和建模误差 PDF 变化曲线. 同样,这里也将所提方法与常规 WNN 方法以及文献 [6] 中的 GD-WNN 方法进行比较. 可以看出,采用所提建模方法可以获得具有较大动态趋势拟合优

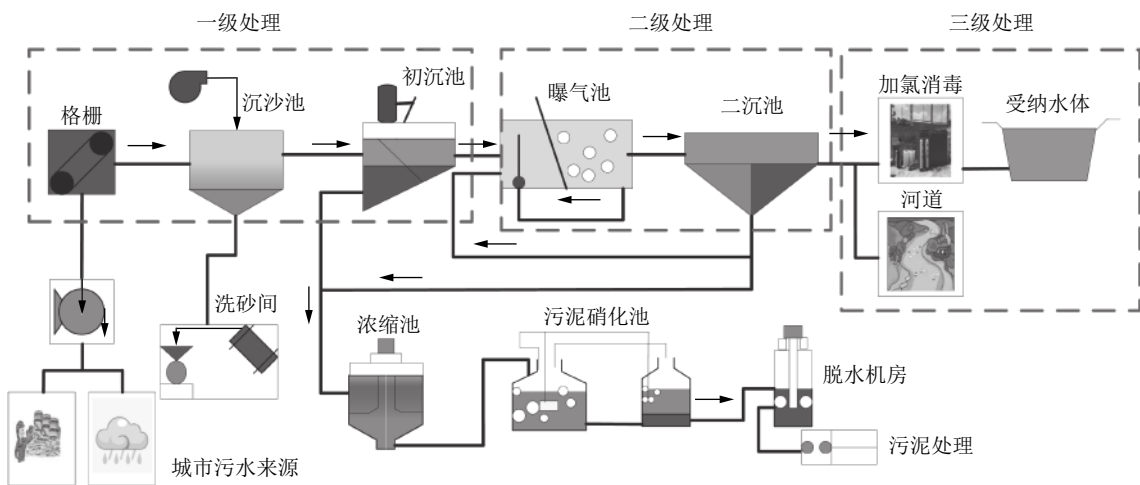


图 9 典型活性污泥法污水处理过程工艺流程图

Fig. 9 Flow chart of a typical activated sludge wastewater treatment process

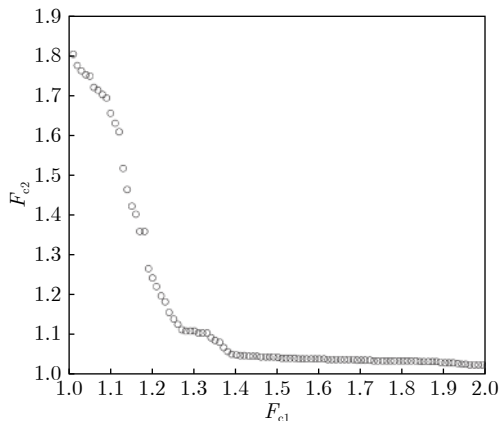


图 10 COD 含量建模 Pareto 前沿进化过程
Fig.10 Pareto front evolution process of COD content modeling

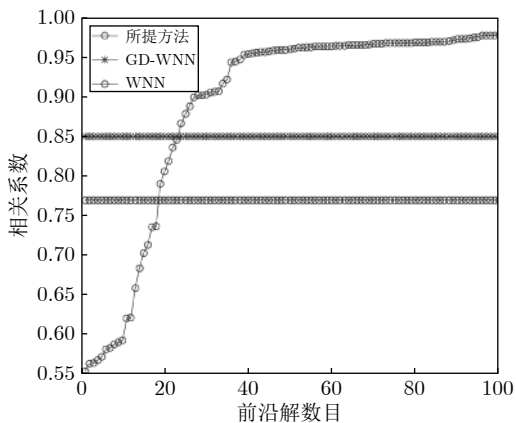


图 11 不同优化解对应的拟合优度值变化曲线
Fig.11 Change curve of goodness of fit corresponding to different optimization solutions

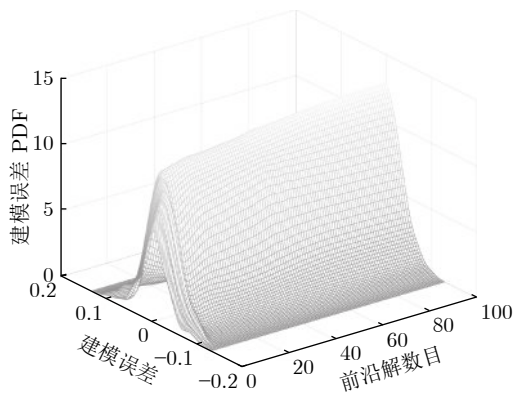


图 12 不同优化解对应的 COD 含量建模误差 PDF 变化曲面
Fig.12 PDF changing surface of COD content modeling error corresponding to different optimization solutions

度的 70 余组解, 这些解对应的拟合优度均远好于常规 WNN 方法以及文献 [6] 中的 GD-WNN 方法, 并且所提方法所得解的最优拟合优度已非常接近 1. 从图 12 所有多目标优化解的建模误差 PDF 变化曲线可以看出, 从第 1 号解到第 100 号解, 建模误差 PDF 的形状越来越窄而尖, 并且越来越接近设定的理想 PDF 形状, 即模型的随机性和不确定性很小.

图 13 为 50 号解与 100 号解的建模误差 PDF 曲线与其它方法的建模误差 PDF 曲线对比图, 这里设置的期望 PDF 为均值为 0 方差为 0.3 的高斯型概率密度函数. 图中可以看出, 本文方法获得的非最优 50 号解对应的建模误差 PDF 形状为窄而高的形状, 仍然远好于常规 WNN 方法和 GD-WNN 方法. 图 14 和图 15 分别是本文方法所得非最优 50 号解的建模效果以及对新样本的测试效果, 虽然是选取的本文方法的非最优解, 但是从图中可

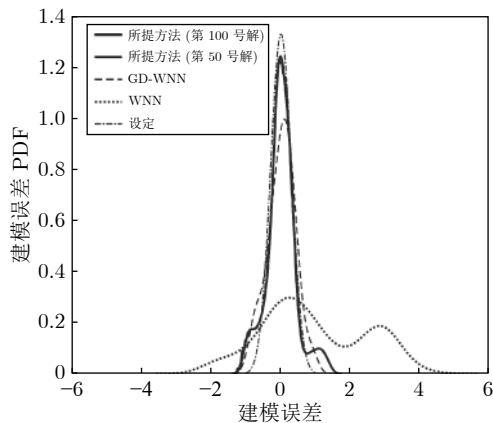


图 13 不同方法 COD 含量建模误差 PDF 比较
Fig.13 PDF comparison of COD content modeling error with different methods

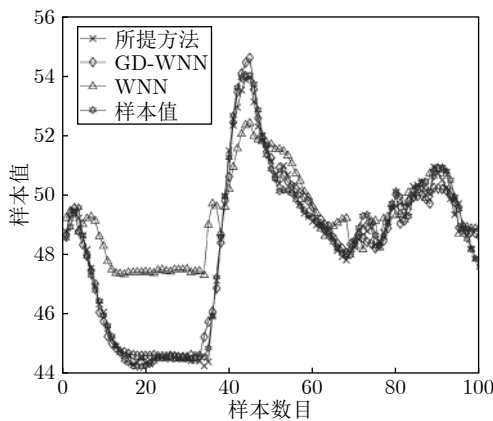


图 14 所提方法 50 号优化解对应的 COD 含量建模效果
Fig.14 Modeling result of COD content corresponding to the 50th optimization solution of the proposed method

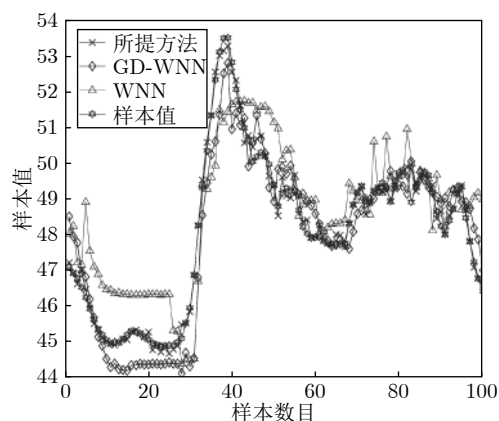


图 15 所提方法 50 号优化解对应的 COD 含量测试效果
Fig.15 Testing result of COD content corresponding to the 50th optimization solution of the proposed method

以看出非最优解的建模和新样本测试效果均好于其他两种对比方法。

4 结语

基于误差最小的数据驱动工业系统建模时,通常基于单一的 RMSE 等一维性能指标.但是 RMSE 等时间维度的一维性能指标并不能充分体现动态系统建模的随机性和不确定性.同时,对于传统动态系统建模方法,并没有考虑模型输出和动态系统实际输出之间的拟合趋势.为此,本文基于数据驱动小波神经网络智能建模、多目标参数优化以及核密度估计技术,提出综合考虑建模误差 PDF 形状与趋势拟合优度的动态系统优化建模方法.其中多目标参数优化的性能指标分别为实际建模误差 PDF 与期望建模误差 PDF 之间二维偏差平方、趋势拟合优度.仿真实验以及污水处理过程数据验证表明:相比于对比的两种现有建模方法,所提方法不仅具有更好的建模精度和泛化能力,还可控制建模误差的空间分布状态,使得所提方法的建模误差 PDF 比传统建模方法的建模误差 PDF 更高、更窄,即模型中含有的随机性和不确定性更小.此外,所提方法还可以获得一大类具有建模误差 PDF 形状接近期望分布形状,且模型输出与实际输出的趋势拟合优度值较大的数据模型参数解,因而具有更好的实用性.

References

- 1 Song He-Da, Zhou Ping, Wang Hong, Chai Tian-You. Nonlinear subspace modeling of multivariate molten iron quality in blast furnace ironmaking and its application. *Acta Automatica Sinica*, 2016, **42**(11): 1664–1679
(宋贺达,周平,王宏,柴天佑.高炉炼铁过程多元铁水质量非线性

子空间建模及应用. *自动化学报*, 2016, **42**(11): 1664–1679)

- 2 Zhou P, Lv Y B, Wang H, Chai T Y. Data-driven robust RVFLNs modeling of blast furnace ironmaking process using Cauchy distribution weighted M-estimation. *IEEE Transactions on Industrial Electronics*, 2017, **64**(9): 7141–7151
- 3 Han H G, Qiao J F. Nonlinear model-predictive control for industrial processes: an application to wastewater treatment process. *IEEE Transactions on Industrial Electronics*, 2014, **61**(4): 1970–1982
- 4 Zhou P, Lu S W, Chai T Y. Data-driven soft-sensor modeling for product quality estimation using case-based reasoning and fuzzy-similarity rough sets. *IEEE Transactions on Automation Science and Engineering*, 2014, **11**(4): 992–1003
- 5 Janssens P, Pipeleers G, Swevers J. A data-driven constrained norm-optimal iterative learning control framework for LTI systems. *IEEE Transactions on Control Systems Technology*, 2013, **21**(2): 546–551
- 6 Zhou P, Wang C Y, Li M J, et al. Modeling error PDF optimization based wavelet neural network modeling of dynamic system and its application in blast furnace ironmaking. *IEEE Neurocomputing*, 2018, **285**: 167–175
- 7 Wang H. Bounded Dynamic Stochastic Systems: Modeling and control. London: Springer-Verlag Ltd, 2000.
- 8 Zhou Y Y, Wang A P, Zhou P, Wang H, and Chai T Y. Dynamic performance enhancement for nonlinear stochastic systems using RBF driven nonlinear compensation with extended Kalman filter. *Automatica*, DOI: [10.1016/j.automatica.2019.108693](https://doi.org/10.1016/j.automatica.2019.108693)
- 9 Ding J, Chai T, Wang H. Offline modeling for product quality prediction of mineral processing using modeling error PDF shaping and entropy minimization. *IEEE Transactions on Neural Networks*, 2011, **22**(3): 408–419
- 10 Jia L, Cao L M, Chui M S. Modeling error PDF shape based data-driven model for batch processes. *Chin. J. Sci. Instrum*, 2012, **33**(7): 1505–1512
- 11 Zhou P, Guo D W, Wang H, Chai T Y. Data-driven robust M-LS-SVR-based NARX modeling for estimation and control of molten iron quality indices in blast furnace ironmaking. *IEEE Trans. Neural Netw. Learn. Syst.*, 2018, **29**(9): 4007–4021
- 12 Li Huan-Qin, Wan Bai-Wu. A wavelet neural network model for industrial product quality based on filled function algorithm. *Acta Automatica Sinica*, 2004, **30**(2): 283–287
(李换琴,万百五.基于填充函数算法的工业产品小波网络质量模型. *自动化学报*, 2004, **30**(2): 283–287)
- 13 Rana M, Koprinska I. Forecasting electricity load with advanced wavelet neural networks. *Neurocomputing*, 2016, **182**(3): 118–132
- 14 Wu M, Xu C H, She J H, Yokoyama R. Intelligent integrated optimization and control system for lead-zinc sintering process. *Control Engineering Practice*, 2009, **17**(2): 280–290
- 15 Parzen E. On estimation of a probability density function and mode. *Ann. Math. Stat.*, 1962, **33**(3): 1065–1076

- 16 Jones M C, Marron J S, Sheather S J. A brief survey of bandwidth selection for density estimation. *J. Am. Stat. Assoc.*, 1996, **91**(433): 401–407
- 17 Buch-larsen T, Nielsen J P, Guillén M. Kernel density estimation for heavy-tailed distributions using the champernowne transformation. *Social Science Electronic Publishing*, 2005, **39**(6): 503–516
- 18 Deb K, Pratap A, Agarwal S and Meyarivan T. A fast and elitist multi-objective genetic algorithm: NSGA II. *IEEE Trans. Evol. Comput.*, 2002, **6**(2): 182–197
- 19 Zhang Shuai, Zhou Ping. Recursive bilinear subspace modeling and model-free adaptive control of wastewater treatment. *Acta Automatica Sinica*, DOI: [10.16383/j.aas.c190514](https://doi.org/10.16383/j.aas.c190514)
(张帅, 周平. 污水处理过程递推双线性子空间建模及无模型自适应控制. 自动化学报, DOI: [10.16383/j.aas.c190514](https://doi.org/10.16383/j.aas.c190514))
- 20 Han H G, Zhang L, Qiao J F. Data-based predictive control for wastewater treatment process. *IEEE Access*, 2018, **6**: 1498–1512
- 21 Qiao J F, Zhou H B. Modeling of energy consumption and effluent quality using density peaks-based adaptive fuzzy neural network. *IEEE/CAA Journal of Automatica Sinica*, 2018, **5**(5): 968–976
- 22 Qiao Jun-Fei, Bo Ying-Chun, Han Guang. Application of ESN-based multi indices dual heuristic dynamic programming on wastewater treatment process. *Acta Automatica Sinica*, 2013, **39**(7): 146–1151
(乔俊飞, 薄迎春, 韩广. 基于ESN的多指标DHP控制策略在污水处理过程中的应用. 自动化学报, 2013, **39**(7): 146–1151)



周平 东北大学教授. 分别于 2003 年, 2006 年, 2013 年获得东北大学学士学位、硕士学位和博士学位. 主要研究方向为工业过程运行反馈控制, 数据驱动建模与控制. 本文通信作者.
E-mail: zhouping@mail.neu.edu.cn
(**ZHOU Ping** Ph.D., Professor at

Northeastern University. He received his bachelor degree, master degree, and Ph. D. degree from Northeastern University in 2003, 2006, and 2013, respectively. His research interest covers operation feedback control of industrial process, data-driven modeling and control. Corresponding author of this paper.)



赵向志 东北大学硕士研究生. 2018 年获得东北石油大学学士学位. 主要研究方向为数据驱动建模与控制, 机器学习算法.
E-mail: 2092371322@qq.com

(**ZHAO Xiang-Zhi** Master student at Northeastern University. She received his bachelor degree from Northeast Petroleum University in 2018. Her research interest covers data-driven modeling and control, and machine learning algorithm.)