

控制与决策

Control and Decision

基于深度学习特征提取与多目标优化集成修剪的选择性集成学习软测量方法

金怀平, 王建军, 董守龙, 钱斌, 杨彪, 陈祥光

引用本文:

金怀平, 王建军, 董守龙, 钱斌, 杨彪, 陈祥光. 基于深度学习特征提取与多目标优化集成修剪的选择性集成学习软测量方法[J]. 控制与决策, 2023, 38(3): 738–750.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2021.1220>

您可能感兴趣的其他文章

Articles you may be interested in

[一种时空协同的图卷积长短期记忆网络及其工业软测量应用](#)

A spatio-temporal synergistic graph convolution long short-term memory network and its application for industrial soft sensors

控制与决策. 2022, 37(1): 77–86 <https://doi.org/10.13195/j.kzyjc.2020.0901>

[基于改进集成学习分类的代理辅助进化算法](#)

Improved ensemble learning classification based surrogate-assisted evolutionary algorithm

控制与决策. 2022, 37(10): 2456–2466 <https://doi.org/10.13195/j.kzyjc.2021.0388>

[低质量渲染图像的目标物体6D姿态估计](#)

6D object pose estimation for low-quality rendering images

控制与决策. 2022, 37(1): 135–141 <https://doi.org/10.13195/j.kzyjc.2020.1057>

[改进集成深层自编码器在轴承故障诊断中的应用](#)

Application of improved ensemble deep auto-encoder in bearing fault diagnosis

控制与决策. 2021, 36(1): 135–142 <https://doi.org/10.13195/j.kzyjc.2019.0270>

[基于批次图像化的卷积自编码故障监测方法](#)

Fault detection of batch image-based convolutional autoencoder

控制与决策. 2021, 36(6): 1361–1367 <https://doi.org/10.13195/j.kzyjc.2019.1342>

基于深度学习特征提取与多目标优化集成修剪的 选择性集成学习软测量方法

金怀平^{1†}, 王建军¹, 董守龙², 钱斌¹, 杨彪¹, 陈祥光²

(1. 昆明理工大学 信息工程与自动化学院, 昆明 650500; 2. 北京理工大学 化学与化工学院, 北京 100081)

摘要: 集成学习已成为一种广泛使用的软测量建模框架, 但是建立高性能的集成学习软测量模型依然面临特征选择不当、基模型多样性不足、基模型估计性能不佳等诸多挑战. 为此, 提出一种基于堆栈自编码器多样性生成机制的选择性集成学习高斯过程回归 (selective ensemble of stacked autoencoder based Gaussian process regression, SESAEGR) 软测量建模方法. 该方法充分发挥深度学习在特征提取方面的优势, 通过构建多样性的堆栈自编码器 (stacked autoencoder, SAE) 网络, 建立基于隐特征的高斯过程回归 (Gaussian process regression, GPR) 基模型. 基于模型性能提升率和进化多目标优化对 SAEGR 基模型实施两次集成修剪, 以降低集成模型复杂度、保持甚至进一步提升模型估计性能, 最后, 引入 PLS Stacking 集成策略实现基模型融合. 所提出方法显著优于传统全局和全集成软测量建模方法, 其有效性和优越性通过青霉素发酵过程和 Tennessee Eastman 化工过程得到验证.

关键词: 软测量; 深度学习; 集成学习; 堆栈自编码器; 隐特征; 多目标优化; 高斯过程回归

中图分类号: TP273 文献标志码: A

DOI: 10.13195/j.kzyj.2021.1220

引用格式: 金怀平, 王建军, 董守龙, 等. 基于深度学习特征提取与多目标优化集成修剪的选择性集成学习软测量方法[J]. 控制与决策, 2023, 38(3): 738-750.

Selective ensemble learning for soft sensor development based on deep learning for feature extraction and multi-objective optimization for ensemble pruning

JIN Huai-ping^{1†}, WANG Jian-jun¹, DONG Shou-long², QIAN Bin¹, YANG Biao¹, CHEN Xiang-guang²

(1. Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China; 2. School of Chemistry and Chemical Engineering, Beijing Institute of Technology, Beijing 100081, China)

Abstract: Ensemble learning has become a widely used soft sensor modeling framework. However, the establishment of high-performance ensemble learning soft sensor models still encounters many challenges such as improper feature selection, insufficient diversity of base models, and poor base model estimation performance. Therefore, a selective ensemble of stacked autoencoder based Gaussian process regression (SESAEGPR) is proposed for soft sensor modeling. By fully utilizing the advantages of deep learning in feature extraction, the SESAEGPR method first builds a set of diverse stacked autoencoder (SAE) networks and then establishes a set of Gaussian process regression (GPR) models based on the already extracted latent features. Then, a two-stage ensemble pruning is performed. The first is achieved based on the model performance improvement, and the evolutionary multi-objective optimization approach is used for the second. Ensemble pruning enables the reduction of ensemble model complexity while maintaining or even further improving the ensemble estimation performance. Finally, a PLS Stacking ensemble mechanism is employed to achieve the combination of the selected base models. The proposed method performs significantly better than the traditional global and fully integrated soft sensing methods, and its effectiveness and superiority have been verified through the penicillin fermentation process and the Tennessee Eastman chemical process.

Keywords: soft sensor; deep learning; ensemble learning; stacked autoencoder; latent variables; multi-objective optimization; Gaussian process regression

收稿日期: 2021-07-13; 录用日期: 2021-11-26.

基金项目: 国家自然科学基金项目 (62163019, 61763020, 61863020); 云南省应用基础研究计划项目 (202101AT070096).

责任编辑: 阳春华.

[†]通讯作者. E-mail: jinhuaping@126.com.

0 引言

在化工、冶金、生物、制药、污水处理等流程工业中,过程参数的实时在线检测是实现智能优化制造的基础条件^[1-3],但一些关键质量参数由于缺乏在线检测设备或者因为高昂的成本而难以实现实时测量,数据驱动软测量技术为解决此类问题提供了有效途径^[4-5].该方法通过构建主导变量与辅助变量之间的数据驱动模型实现对难测变量的实时估计,其代表性方法有:主成分回归(PCR)^[6]、偏最小二乘(PLS)^[7]、人工神经网络(ANN)^[8]、支持向量回归(SVR)^[9]、高斯过程回归(GPR)^[10]、深度学习^[11]等.

传统的软测量方法通常采用单一模型结构,但此类建模方式往往存在模型泛化能力较差、鲁棒性不强的问题.因此,集成学习^[12]作为一种有效的机器学习性能提升策略,被广泛引入软测量应用领域^[13-17].

集成学习软测量建模方法成功的关键往往在于构建一组多样性充分且估计精度高的基模型,其中多样性的生成机制尤为关键.常规集成学习软测量建模方法的多样性生成方式主要有样本扰动、特征扰动和模型参数扰动等^[18].例如,文献[19]利用异质相似度选择多样性的建模样本子集,进而构建集成即时学习软测量模型.文献[20]结合随机重采样和偏互信息相关分析(PMI)构建多样性特征子空间,从而建立集成学习高斯过程回归软测量模型.文献[21]通过使用多样性的模型超参数组合,建立集成学习在线支持向量回归软测量模型.尽管多样性扰动机制的类型很多,但本文聚焦于输入特征扰动多样性生成机制及集成学习软测量建模方法的研究.

在软测量建模中,选择合适的输入特征对于确保模型的估计性能至关重要^[22].一般而言,特征选择方法分为两大类:一类是通过某种优化准则直接选取部分原始特征用于建模,此类方法的典型代表有过滤式选择、包裹式选择和嵌入式选择;另一类则通过某种线性或非线性变换将原始特征映射到一个新的特征空间,称为隐特征提取,其典型代表为投影映射法,例如主成分分析.目前,选择合适的建模特征仍然面临一些困难:1)特征筛选法会因为抛弃一些变量而造成信息损失;2)常规的隐特征提取方法无法有效地挖掘复杂的过程数据特征;3)由于过程数据、模型参数、优化参数选择的差异性会导致特征选择结果存在一定的不确定性,难以获取一组全局最优的建模特征.

近年来,深度学习作为一种新兴技术,在图像处

理、计算机视觉、自然语言处理等领域均取得了巨大的成功^[23].该方法采用深层网络结构,通过多层非线性映射,能够从数据中提取更深层的抽象特征,在处理复杂过程数据建模问题上具有不可替代的优势,目前已被广泛应用到软测量建模领域.深度学习取得成功的关键因素之一便是其强大的特征学习能力,通过将原始信号进行逐层非线性特征变换,从过程数据中学习获得层次化的特征表示,从而更有利于解决分类或回归等特定问题.堆栈自编码器(stacked autoencoder, SAE)^[24]作为一种代表性的深度学习方法,已被大量应用于软测量建模领域.文献[25]设计了一种变量加权自编码器,用于提取与关键质量参数相关的潜在特征.文献[26]提出了一种基于半监督预训练策略的SAE并应用于软测量.文献[27]基于门控循环单元,设计了一种目标相关的堆栈自编码器,以提升质量参数估计精度.鉴于深度学习在隐特征提取上的明显优势,本文重点研究基于SAE深度学习隐特征提取的多样性生成机制,结合集成学习构建高性能的软测量模型.此外,构建高性能集成学习软测量模型的另一个关键步骤在于基模型的融合.根据Zhou等^[28]的研究,在模型融合阶段经常存在“多胜于全(many could be better than all)”的现象,即选择部分基模型进行集成能够获得比融合全部基模型更佳的泛化性能.随之而来的一个问题是,选择哪些基模型参与集成,这便是集成修剪问题^[29-31].究其根本,集成修剪的目的在于确保基模型更好地满足准确性和多样性的要求,这本质上是一个多目标优化问题.但是,常规的集成修剪方法往往仅考虑单一的优化目标,无法有效确保准确性和多样性目标的平衡.

本文提出一种基于堆栈自编码器多样性生成机制的选择性集成学习高斯过程回归(selective ensemble of stacked autoencoder based Gaussian process regression, SESAE GPR)软测量建模方法.该方法充分发挥深度学习在特征提取方面的优势,通过构建多样性的堆栈自编码器(stacked autoencoder, SAE)网络,建立基于隐特征的高斯过程回归(Gaussian process regression, GPR)基模型.基于模型性能提升率和进化多目标优化对SAEGPR基模型实施两次集成修剪,以降低集成模型复杂度、保持甚至进一步提升模型估计性能,最后,引入PLS Stacking集成策略实现基模型融合.所提出方法显著优于传统全局和全集成软测量建模方法,其有效性和优越性通过青霉素发酵过程和Tennessee Eastman化工过程得到验证.

1 算法介绍

1.1 高斯过程回归

高斯过程回归 (Gaussian process regression, GPR)^[32] 是一种基于贝叶斯理论和统计学习理论发展而来的概率建模技术, 具有容易实现、模型参数少、能够提供估计不确定度等优点, 适用于处理小样本、高维数和非线性等复杂建模问题。

高斯过程 (Gaussian process, GP) 是任意有限个具有联合高斯分布的随机变量的集合. 为描述函数分布, 需直接在函数 f 空间上定义一个完全由均值函数 $m(\mathbf{x})$ 和协方差函数 $C(\mathbf{x}, \mathbf{x}')$ 确定的高斯过程 (GP), 即

$$m(\mathbf{x}) = E[f(\mathbf{x})], \quad (1)$$

$$C(\mathbf{x}, \mathbf{x}') = E[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))]. \quad (2)$$

其中: \mathbf{x} 为任意输入向量, $f(\mathbf{x}) \sim \text{GP}(m(\mathbf{x}), C(\mathbf{x}, \mathbf{x}'))$ 为所定义的高斯过程。

对于有标签的样本集 $\mathbf{D} = \{\mathbf{X}, \mathbf{y}\} = \{\mathbf{x}_i, y_i\}_{i=1}^n$ 和所针对的回归问题, 考虑如下模型:

$$y = f(\mathbf{x}) + \varepsilon, \varepsilon \sim N(0, \sigma_n^2). \quad (3)$$

其中: $f(\cdot)$ 为待拟合函数, n 为总的输入样本数, σ 是符合均值为 0 方差为 σ_n^2 分布的高斯噪声, y 为受到 ε 干扰的观测值。

假设对数据进行适当的预处理, 使其满足训练样本服从均值为 0 的高斯过程, 即输出观测值服从如下分布:

$$y \sim N(0, \mathbf{C}), \quad (4)$$

其中 \mathbf{C} 为 $n \times n$ 阶对称正定的协方差矩阵。

给定一个测试样本点, 其测试输入向量为 \mathbf{x}_* , 则对应的训练输出 \mathbf{y} 与测试输出 y_* 的联合先验分布为

$$\begin{bmatrix} \mathbf{y} \\ y_* \end{bmatrix} \sim \left(\mathbf{0}, \begin{bmatrix} \mathbf{C} & \mathbf{Z}_* \\ \mathbf{Z}_*^T & C(\mathbf{x}_*, \mathbf{x}_*) \end{bmatrix} \right). \quad (5)$$

其中: $\mathbf{Z}_* = [C(\mathbf{x}_*, \mathbf{x}_1), \dots, C(\mathbf{x}_*, \mathbf{x}_n)]$ 为 $n \times 1$ 阶矩阵, 表示测试样本点 \mathbf{x}_* 与训练集输入 $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ 之间的协方差; $C(\mathbf{x}_*, \mathbf{x}_*)$ 为测试样本点 \mathbf{x}_* 自身的协方差. 给定协方差函数, 测试输出 y_* 的后验分布为

$$y_* | \mathbf{X}, \mathbf{y}, \mathbf{x}_* \sim N(\hat{y}_*, \hat{\sigma}_*^2). \quad (6)$$

其中

$$\hat{y}_* = E(y_*) = \mathbf{Z}_*^T \mathbf{C}^{-1} \mathbf{y}, \quad (7)$$

$$\hat{\sigma}_*^2 = \text{Var}(y_*) = C(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{Z}_*^T \mathbf{C}^{-1} \mathbf{Z}_*. \quad (8)$$

\hat{y}_* 和 $\hat{\sigma}_*^2$ 分别为 GPR 模型在测试样本点 \mathbf{x}_* 上的估计均值和方差, $E(\cdot)$ 和 $\text{Var}(\cdot)$ 分别为均值和协方差算

子. 选择带噪声项的 Matern 协方差函数

$$C(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 \left(1 + \frac{\sqrt{3} \|\mathbf{x}_i - \mathbf{x}_j\|}{l} \right) \exp \left(- \frac{\sqrt{3} \|\mathbf{x}_i - \mathbf{x}_j\|}{l} \right) + \sigma_n^2 \delta_{ij}. \quad (9)$$

其中: σ_f^2 为输出尺度; l 为输入尺度; σ_n^2 为噪声方差; 当 $i = j$ 时 $\delta_{ij} = \sigma_f^2$, 否则 $\delta_{ij} = 0$. 模型超参数集合 $\theta = \{\sigma_f^2, l, \sigma_n^2\}$ 一般通过最大似然法获得。

1.2 自编码器和堆栈自编码器

1.2.1 自编码器

自编码器 (autoencoder, AE)^[33] 是一种单隐层深度学习方法, 也是构建深度堆栈自编码器的基本单元. 自编码器试图通过参数学习和最小化损失函数重构原始输入, 从而在隐藏层学习到特征的潜在表示, 通常称其为隐特征或潜在特征. AE 结构如图 1 所示, 包含输入层、隐藏层和输出层 3 个网络层次, 其中输入层和隐藏层构成编码器部分, 隐藏层和输出层构成解码器部分。

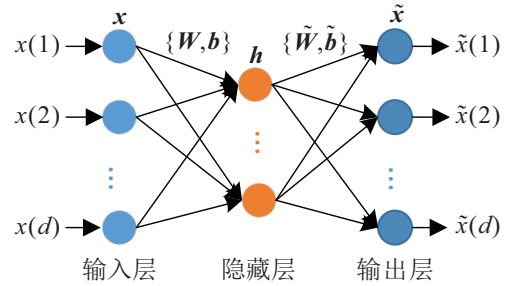


图 1 AE 结构

假设 $\mathbf{x} = [x(1), x(2), \dots, x(d)]^T \in R^{d \times 1}$ 表示 AE 的原始输入样本, d 为样本维度, 通过编码层将原始输入映射到隐藏层, 获得隐特征 $\mathbf{h} = [h(1), h(2), \dots, h(v)]^T \in R^v$, v 表示隐层节点数. 通过解码层将 \mathbf{h} 映射到输出层获得 d 维输出 $\tilde{\mathbf{x}} = [\tilde{x}(1), \tilde{x}(2), \dots, \tilde{x}(d)]^T \in R^d$. 在编码阶段, \mathbf{x} 通过非线性激活函数 f 转换为隐层特征 \mathbf{h} , 有

$$\mathbf{h} = f(\mathbf{W}\mathbf{x} + \mathbf{b}), \quad (10)$$

其中 $\mathbf{W} \in R^{v \times d}$ 和 $\mathbf{b} \in R^v$ 分别为编码层 $v \times d$ 阶的权重矩阵和 v 维的偏置向量. 在解码阶段, 将隐特征 \mathbf{h} 从隐层映射到输出层获得重构之后的输出向量 $\tilde{\mathbf{x}}$, 有

$$\tilde{\mathbf{x}} = f(\tilde{\mathbf{W}}\mathbf{h} + \tilde{\mathbf{b}}), \quad (11)$$

其中 $\tilde{\mathbf{W}} \in R^{d \times v}$ 和 $\tilde{\mathbf{b}} \in R^d$ 分别为解码层 $d \times v$ 阶的权重矩阵和 d 维偏置向量. 不失一般性, 采用非线性激活函数 Sigmoid 函数 $f(\mathbf{x}) = 1/(1 + \exp^{-\mathbf{x}})$.

给定 n 个训练样本, 样本矩阵 $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$, $\mathbf{H} = \{\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_n\}$ 为其对应的潜在特征矩阵, 重构样本矩阵为 $\tilde{\mathbf{X}} = \{\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_n\}$. 为确保

输入数据的重构误差足够小,需通过梯度下降法和反向传递算法不断更新优化模型参数 $\Theta_{AE} = \{\mathbf{W}, \mathbf{b}; \tilde{\mathbf{W}}, \tilde{\mathbf{b}}\}$, 直至达到迭代终止条件. 重构误差函数和参数更新公式分别如下所示:

$$\begin{cases} J(\mathbf{W}, \mathbf{b}; \tilde{\mathbf{W}}, \tilde{\mathbf{b}}) = \frac{1}{2n} \sum_{i=1}^n \|\tilde{\mathbf{x}}_i - \mathbf{x}_i\|^2, \\ J(\Theta_{AE}) = \frac{1}{2n} \sum_{i=1}^n \|\tilde{\mathbf{x}}_i - \mathbf{x}_i\|^2; \end{cases} \quad (12)$$

$$\begin{cases} \mathbf{W} \leftarrow \mathbf{W} - L_{rate} \frac{\partial J(\Theta_{AE})}{\partial \mathbf{W}}, \\ \mathbf{b} \leftarrow \mathbf{b} - L_{rate} \frac{\partial J(\Theta_{AE})}{\partial \mathbf{b}}. \end{cases} \quad (13)$$

其中: L_{rate} 为学习率, $\partial J(\Theta_{AE})/\partial \mathbf{W}$ 为 $J(\Theta_{AE})$ 对权重 \mathbf{W} 求偏导, $\partial J(\Theta_{AE})/\partial \mathbf{b}$ 为 $J(\Theta_{AE})$ 对偏置 \mathbf{b} 求偏导.

1.2.2 堆栈自编码器

传统 AE 结构简单、变换单一,难以获得更为复杂和抽象的隐特征,因此可通过逐层堆叠自编码器的方式,构建出结构灵活多样的堆栈自编码器 (stacked autoencoder, SAE)^[24,34], 由此提取出更复杂抽象的隐含层特征. SAE 的网络结构如图 2 所示.

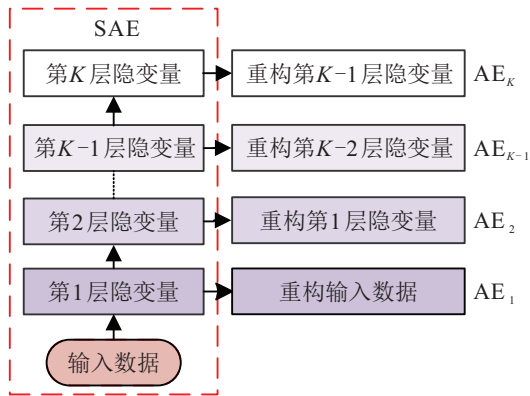


图 2 SAE 结构

SAE 网络训练过程分为两个阶段,如图 3 所示. 一个是无监督逐层贪婪训练阶段,也称为“预训练”,在该阶段参与叠加的每个 AE 将以逐层独立的方式进行训练,一层训练结束,下一层训练开始,并将上一层 AE 的隐特征作为本层 AE 的输入. 对于具有 K 个隐层的 SAE 网络,假设 $\mathbf{x} \in R^{d \times 1}$ 是第 1 个自编码器的输入样本, AE_1 通过最小化 \mathbf{x} 的重构误差学习和更新编码层的参数集 $\Theta_1 = \{\mathbf{W}_1, \mathbf{b}_1\}$, 其隐层特征向量为 \mathbf{h}_1 ; 将 \mathbf{h}_1 作为第 2 个自编码器的输入继续训练模型,学习参数集 $\Theta_2 = \{\mathbf{W}_2, \mathbf{b}_2\}$ 以及 AE_2 的隐层特征向量 \mathbf{h}_2 . 依次类推,逐层训练至 AE_K , 即完成整个 SAE 网络的预训练,获得 SAE 网络的初始参数 $\Theta_{SAE} = \{\mathbf{W}_1, \mathbf{b}_1; \mathbf{W}_2, \mathbf{b}_2; \dots; \mathbf{W}_K, \mathbf{b}_K\}$ 和 K 个隐层分别对应的潜在特征向量 $\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_K$. 另一

个是网络有监督微调阶段. 在预训练结束后,在网络的输出端加一个输出层,以标记样本的主导变量 y 与其映射 \tilde{y} 之间均方误差作为 SAE 的损失函数,通过批量梯度下降法和反向传播算法微调整个网络的参数 Θ_{SAE} , 且有

$$\tilde{y} = f_*(\mathbf{W}_* \mathbf{h}_k + \mathbf{b}_*). \quad (14)$$

其中: $\mathbf{W}_*, \mathbf{b}_*$ 分别为输出层的权重矩阵和偏置向量, f_* 为输出层激活函数. 微调回归网络的损失函数如下:

$$J(\Theta_{SAE}) = \frac{1}{2n_L} \sum_{i=1}^{n_L} \|\tilde{y}_i - y_i\|^2. \quad (15)$$

其中: n_L 为标记样本数, y_i 和 \tilde{y}_i 分别为主导变量的真实值和估计值.

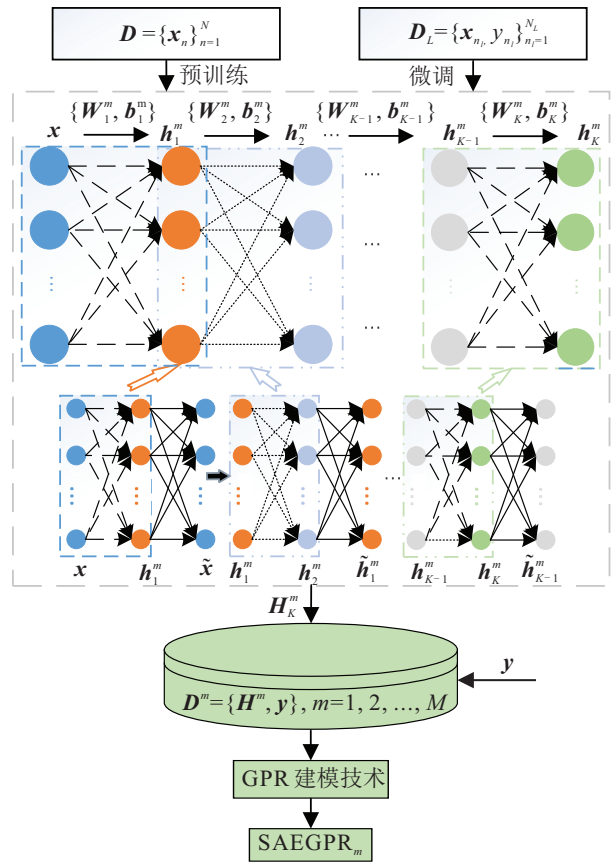


图 3 SAEGPR 基模型构建

2 SESAEGPR 软测量建模方法

2.1 多样性隐特征提取及基模型构建

本文提出使用 SAE 深度学习网络生成多样性的隐特征,进而生成多样性的 GPR 基模型,简称为 SAEGPR 基模型. 影响 SAE 隐特征提取结果的一种主要因素是训练样本的选择. 鉴于实际生产过程往往存在标记样本有限、非标记样本丰富的普遍现象,考虑将标记样本输入数据与非标记样本数据混合后再对隐特征进行提取. 为了确保隐特征提取的多样性,

考虑将有限标记样本的输入数据与多样性的非标记数据子集进行混合后再实施隐特征提取,同时采用标记样本对SAE网络进行微调.

此外,隐特征提取结果还极易受到网络结构和训练参数设置^[35]等因素的影响.一般而言,SAE训练主要涉及参数如表1所示.目前,一般通过经验法、网格搜索法、试错法、交叉验证法等确定模型参数,但确定一组最优的SAE网络配置参数在现实情况下很难实现.实际上,建模样本和SAE网络参数的微小变化都可能导致隐特征提取结果的较大差异,而且选择单一的参数设置容易造成较大的建模风险.因此退而求其次,一种更为现实的选择是采用多样性的参数配置提取多样性的隐特征,从而构建多样性的SAEGPR模型.

表1 SAE网络结构和训练参数符号及描述

符号	描述
Hnode1	第1层节点数
Hnode2	第2层节点数
Lrate1	预训练阶段学习率
Nepoch1	预训练阶段 epoch 数
Bsize1	预训练阶段样本批次大小
Lrate2	微调阶段学习率
Nepoch2	微调阶段 epoch 数
Bsize2	微调阶段样本批次大小

综合上述两方面因素,本文采用一种融合训练样本扰动和模型参数扰动的多样性基模型生成机制.令标记数据集 $D_L = \{\mathbf{X}_{N_l}, \mathbf{y}\} = \{\mathbf{x}_{n_l}, y_{n_l}\}_{n_l=1}^{N_l}$ 和非标记数据集 $D_U = \{\mathbf{X}_U\} = \{\mathbf{x}_{n_u}\}_{n_u=1}^{N_u}$ 分别表示第 n_l 个标记样本的输入向量和对应的标签, \mathbf{x}_{n_u} 为第 n_u 个非标记样本, N_l 和 N_u 分别为标记样本总数和非标记样本总数.多样性的SAEGPR基模型构建原理如图3所示.

构建基模型的关键步骤如下:

step 1: 确定模型训练数据.不失一般性,当需要训练第 m 个SAE深度学习模型时,通过Bootstrapping随机重采样方法从 N_u 个非标记样本中选择 $N_{u'}$ 个非标记样本 $D_{U_m} = \mathbf{X}_{N_{u'}} = \{\mathbf{x}_{n_{u'}}\}_{n_{u'}=1}^{N_{u'}}$; 然后获得扩充训练集 $D = \{\mathbf{X}_{N_l} \cup \mathbf{X}_{N_{u'}}\} = \{\mathbf{x}_n\}_{n=1}^N$, 其中 \mathbf{x}_n 为第 n 个输入数据, N 为输入样本总数,且 $N = N_l + N_{u'}$.此外,从表1给出的参数候选范围内选出一组模型结构和学习参数.

step 2: 根据 step 1 的SAE模型参数配置,将数据集 D 用于SAE预训练,标记数据集 D_L 用于模型微

调,从而获得原始输入数据 \mathbf{X}_{N_l} 对应的隐特征数据 \mathbf{H}^m , 并构建新的建模样本集 $D^m = \{\mathbf{H}^m, \mathbf{y}\}$, 将 D^m 用于GPR建模,构建得到第 m 个SAEGPR基模型.

step 3: 返回 step 1, 直到生成 M 个SAEGPR基模型为止.

根据上述步骤获得的SAEGPR_m基模型可描述为

$$\begin{cases} \hat{y}_{\text{new}}^m = E(y_{\text{new}}^m) = (\mathbf{z}_{\text{new}}^m)^T (\mathbf{C}^m)^{-1} \mathbf{y}^m, \\ (\sigma_{\text{new}}^m)^2 = \text{Var}(y_{\text{new}}^m) = \\ C(\mathbf{h}_{k,\text{new}}^m, \mathbf{h}_{k,\text{new}}^m) - (\mathbf{z}_{\text{new}}^m)^T (\mathbf{C}^m)^{-1} \mathbf{z}_{\text{new}}^m, \\ m = 1, 2, \dots, M. \end{cases} \quad (16)$$

其中: $\mathbf{z}_{\text{new}}^m = [C(\mathbf{h}_{k,\text{new}}^m, \mathbf{h}_{k,1}^m), \dots, C(\mathbf{h}_{k,\text{new}}^m, \mathbf{h}_{k,n_l}^m)]^T$, \hat{y}_{new}^m 和 $(\sigma_{\text{new}}^m)^2$ 分别为该基模型的估计均值和方差.

综上,多样性隐特征提取是构建多样性SAEGPR基模型的关键所在,其对于提升软测量模型估计性能的重要意义主要可以归纳为3方面:

1) 流程工业过程中普遍存在数据维度高、特征之间存在冗余等问题,而基于深度学习的隐特征提取技术可有效实现复杂过程数据关键信息的提取、特征降维及冗余信息的剔除.

2) 流程工业普遍存在多时段、多模式等复杂过程特征,而全局建模由于局部过程特征处理能力不足而导致模型估计性能受限.相比而言,多样性的隐特征提取有助于充分发挥“分而治之”的优势,通过构建在局部区域具有较高预测能力的SAEGPR基模型,然后引入集成学习以提升软测量模型的全局估计能力.

3) 基于深度学习提取隐特征需要设置大量的参数,在建模样本和先验知识受限的情况下,寻求单一最优参数组合的努力是不切实际的.为此,采用多样性的隐特征提取信息有助于降低参数选择不当带来的建模风险.

2.2 双重集成修剪

2.2.1 基于模型性能提升率的一次集成修剪

初始建模阶段,为了最大程度地激发SAEGPR基模型的多样性,需要构建尽可能多的基模型.其中不乏估计性能差的基模型,这不但会影响最终的建模精度,还会造成极大的计算负担.因此,将基模型在隐特征提取前后的估计性能变化作为集成修剪的一项重要依据.算法具体实施步骤如下.

step 1: 基于原始建模样本构建GPR模型.

step 2: 评价GPR和 $\{\text{SAEGPR}_m\}_{m=1}^M$ 在验证样本集 $D_{\text{val}} = \{\mathbf{X}_{\text{val}}, \mathbf{y}_{\text{val}}\}$ 上的估计均方根误差

RMSE_{val}和 $\{\text{RMSE}_{\text{val}}^m\}_{m=1}^M$, 定义为

$$\begin{cases} \text{RMSE}_{\text{val}} = \sqrt{\frac{\sum_{i=1}^{n_{\text{val}}} (\hat{y}_{\text{val},i} - y_{\text{val},i})^2}{n_{\text{val}}}}, \\ \text{RMSE}_{\text{val}}^m = \sqrt{\frac{\sum_{i=1}^{n_{\text{val}}} (\hat{y}_{\text{val},i}^m - y_{\text{val},i})^2}{n_{\text{val}}}}, \end{cases} \quad m = 1, 2, \dots, M. \quad (17)$$

其中: n_{val} 为验证样本数, $y_{\text{val},i}$ 为第 i 个验证样本的真实值, $\hat{y}_{\text{val},i}$ 和 $\hat{y}_{\text{val},i}^m$ 分别为第 i 个验证样本在 GPR 和第 m 个 SAEGPR 基模型上的估计值.

step 3: 计算每个 SAEGPR 基模型相对于 GPR 模型的性能提升率

$$I_{\text{rate}}^m = (\text{RMSE}_{\text{val}} - \text{RMSE}_{\text{val}}^m) / \text{RMSE}_{\text{val}}. \quad (18)$$

若 $I_{\text{rate}}^m > I_{\text{rate}}$, 则保留第 m 个 SAEGPR 基模型, 其中 I_{rate} 表示筛选阈值.

经过上述步骤后, 假设保留了 M' 个多样性的 SAEGPR 基模型.

2.2.2 基于进化多目标优化的二次集成修剪

虽然第 2.2.1 节剔除了部分性能较差的基模型, 但所保留的部分基模型之间可能仍然面临冗余度较高、差异性较小等问题, 这在模型融合阶段势必会造成集成模型泛化性能不佳、鲁棒性不强等问题, 因此需要考虑实施二次修剪. 本文以精度指标和多样性指标作为优化目标, 利用进化多目标优化算法进行求解, 以求获得一组具有较高准确性和多样性的 SAEGPR 基模型.

首先确定集成修剪多目标优化问题的数学模型, 关键在于确定决策变量、目标函数和约束条件.

1) 决策变量: 基模型的筛选问题实质是一个 0-1 整数优化问题, 0 表示未选中模型, 1 表示选中模型, 因此考虑用一组二进制变量表示决策变量, 即 $\mathbf{s} = [s_1, s_2, \dots, s_{M'}] \in \{0, 1\}^{M'}$.

2) 目标函数: 大量研究表明, 确保基模型的精度与它们之间的多样性是集成建模取得成功的关键所在, 因此考虑估计精度优化目标函数 f_1 和多样性优化目标函数 f_2 .

3) 约束条件: 以最终基模型选择数 M'' 作为限制条件.

根据上述分析, 可将集成修剪多目标优化问题描述为

$$\begin{aligned} & \max [f_1(\mathbf{s}), f_2(\mathbf{s})]; \\ & \text{s.t.} \quad \sum_{i=1}^{m'} S_i = M''. \end{aligned} \quad (19)$$

下面给出估计精度和多样性目标函数的具体定义.

定义 1 (估计精度指标) 利用验证集 D_{val} 估计所有待选择基模型的均方根误差 $\{\text{RMSE}_{\text{val}}^m\}_{m=1}^{M''}$, 并据此定义整体的估计精度为

$$\text{RMSE}_{\text{avg, val}} = \frac{\sum_{m=1}^{M''} \text{RMSE}_{\text{val}}^m}{M''}, \quad (20)$$

其中 $\text{RMSE}_{\text{val}}^m$ 为第 m 个基模型的验证均方根误差.

定义 2 (多样性指标) 采用相关系数衡量基模型之间的多样性, 基模型之间差异越大, 估计误差序列之间的相关系数值便越小. 任意两个基模型之间的相关性系数可估计为

$$r(e_i, e_j) = \frac{\text{Cov}(e_i, e_j)}{\sqrt{\text{Var}(e_i)\text{Var}(e_j)}}. \quad (21)$$

其中: e_i, e_j 分别为任意两个 SAEGPR 基模型的估计误差序列, $\text{Cov}(\cdot, \cdot)$ 和 $\text{Var}(\cdot)$ 分别为协方差和方差算子. 获得任意两个基模型的相关系数值后, 将其平均值作为整体多样性指标, 有

$$r_{\text{avg, val}} = \frac{\sum_{\beta=\alpha+1}^{M''} \sum_{\alpha=1}^{M''-1} r(e_i, e_j)}{((M'')^2 - M'')/2}. \quad (22)$$

由式 (20) 和 (22) 可知: 估计误差越小, 模型精度越高; 模型估计误差序列之间的相关性越小, 多样性越强. 于是式 (19) 的最大化优化问题转化为最小化多目标优化问题

$$\min [\text{RMSE}_{\text{avg, val}}, r_{\text{avg, val}}]. \quad (23)$$

为求解上述多目标优化问题, 选择基于非支配排序的多目标遗传算法 (nondominated sorting genetic algorithm II, NSGA-II). 该方法是 Deb 等^[36] 通过引入快速非支配排序算法、精英保留策略以及拥挤度分配策略对 NSGA 的改进版本, 有效改善了 NSGA 因目标问题个数与种群大小带来的高计算复杂度问题和优化过程中优异解丢失的问题.

假设初始种群大小、迭代次数和模型选择数分别为 P, T 和 M'' , 则采用 NSGA-II 算法实现集成修剪的步骤如下.

step 1: 根据问题定义对待二次选择的 M' 个基模型进行二进制编码, 并随机初始化个体的染色体, 生成规模为 P 的初始种群 $P_t, t = 0$.

step 2: 从染色体上解码决策变量, 确定被选中 SAEGPR 基模型, 利用验证样本集 $D_{\text{val}} = \{\mathbf{X}_{\text{val}}, \mathbf{y}_{\text{val}}\}$ 对种群 P_t 中的每个个体进行适应度评价.

step 3: 重复以下步骤, 直至达到预期迭代次数 T :

step 3.1: 对种群 P_t 进行非支配排序, 并计算全部非支配解的拥挤距离;

step 3.2: 通过锦标赛法从种群 P_t 中选择个体进行交叉、变异, 生成子代种群 Q_t , 合并父代和子代种群, 使 $R_t = P_t \cup Q_t$, 转至 step 2 对 R_t 进行适应度评估;

step 3.3: 根据精英保留策略和拥挤距离, 从 R_t 选出最佳的 P 个个体组成新种群 P_{t+1} .

step 4: 迭代终止, 根据非支配排序从最后一代个体中选出合适的 Pareto 最优解并对其解码, 即可获得 M'' 个 SAEGPR 基模型用于集成建模.

2.3 基模型融合

完成双重集成修剪后, 采用 PLS Stacking 集成策略对 M'' 个 SAEGPR 基模型进行融合. 为进一步有效处理基模型之间可能存在的冗余问题, 采用 PLS 方法构建融合模型, 其基本思路为: 估计每个基模型在验证集 $D_{val} = \{X_{val}, y_{val}\}$ 上的估计输出 $\hat{y}_{val} = \{\hat{y}_{val,1}, \hat{y}_{val,2}, \dots, \hat{y}_{val,M''}\}$, 并以 \hat{y}_{val} 为输入、 y_{val} 为输出构建 PLS 回归模型. 该过程需要离线完成.

在线估计阶段, 给定一个测试样本 x_{new} , 可从基模型获得 M'' 个相应的估计均值 $\hat{y}_{new} = \{\hat{y}_{new,1}, \hat{y}_{new,2}, \dots, \hat{y}_{new,M''}\}$ 和方差 $\sigma_{new}^2 = \{\sigma_{new,1}^2, \sigma_{new,2}^2, \dots, \sigma_{new,M''}^2\}$. 基于 PLS Stacking 的集成输出可以表示为

$$\hat{y}_{new} = \omega_0 + \omega_1 \hat{y}_1 + \omega_2 \hat{y}_2 + \dots + \omega_{M''} \hat{y}_{M''}, \quad (24)$$

其中 $\omega_0, \omega_1, \dots, \omega_{M''}$ 为 PLS 模型回归系数.

根据不确定度理论^[37], 集成估计方差 σ_{new}^2 可以估计为

$$\sigma_{new}^2 = \sum_{i=1}^{M''} \left(\frac{\partial y}{\partial y_i} \right)^2 \sigma_i^2 + 2 \sum_{1 \leq i < j} \frac{\partial y}{\partial y_i} \frac{\partial y}{\partial y_j} \rho_{ij} \sigma_i \sigma_j. \quad (25)$$

其中: σ_i 和 σ_j 为任意两个基模型的估计输出不确定度, ρ_{ij} 为任意二者的相关系数.

由式(26)可知, $w_i = \partial y / \partial y_i$, 则式(25)可重写为

$$\sigma_{new}^2 = \sum_{i=1}^{M''} w_i^2 \sigma_i^2 + 2 \sum_{1 \leq i < j} w_i^2 w_j^2 \rho_{ij} \sigma_i \sigma_j. \quad (26)$$

由式(26)可以看出, 基模型自身的估计方差水平与估计方差之间的相关程度决定了集成估计方差. 若 $\rho_{ij} = 0$, 则表明基模型完全独立, 集成估计方差 $\sigma_{new}^2 = \sum_{i=1}^{M''} w_i^2 \sigma_i^2$. 但在实际应用中, 基模型之间难免存在某种程度的相关性, 即 $\rho_{ij} \neq 0$. 由于真实的 ρ_{ij}

值无从得知, 采用下式对其进行估计:

$$\rho_{ij} = \frac{\text{Cov}(\sigma_{i, \text{val}}, \sigma_{j, \text{val}})}{\sqrt{\text{Var}(\sigma_{i, \text{val}}) \text{Var}(\sigma_{j, \text{val}})}}, \quad (27)$$

其中 $\{\hat{y}_{i, \text{val}}, \hat{\sigma}_{i, \text{val}}, \hat{y}_{j, \text{val}}, \hat{\sigma}_{j, \text{val}}\}$ 为任意两个基模型在验证样本集上的估计均值向量和标准差向量.

2.4 实施原理

SESAEGPR 软测量方法的原理如图4所示. 具体实施步骤如下.

step 1: 采集建模样本, 并实施样本划分、异常值处理和归一化等预处理操作.

step 2: 生成 M 组多样性的 SAE 深度学习参数组合和训练数据集, 训练相应的 SAE 网络, 提取原始输入对应的隐特征, 并据此构建 $\{SAEGPR_m\}_{m=1}^M$ 基模型.

step 3: 评价 GPR 和 SAEGPR 基模型在给定验证样本集上的估计性能并计算使用隐特征后的性能提

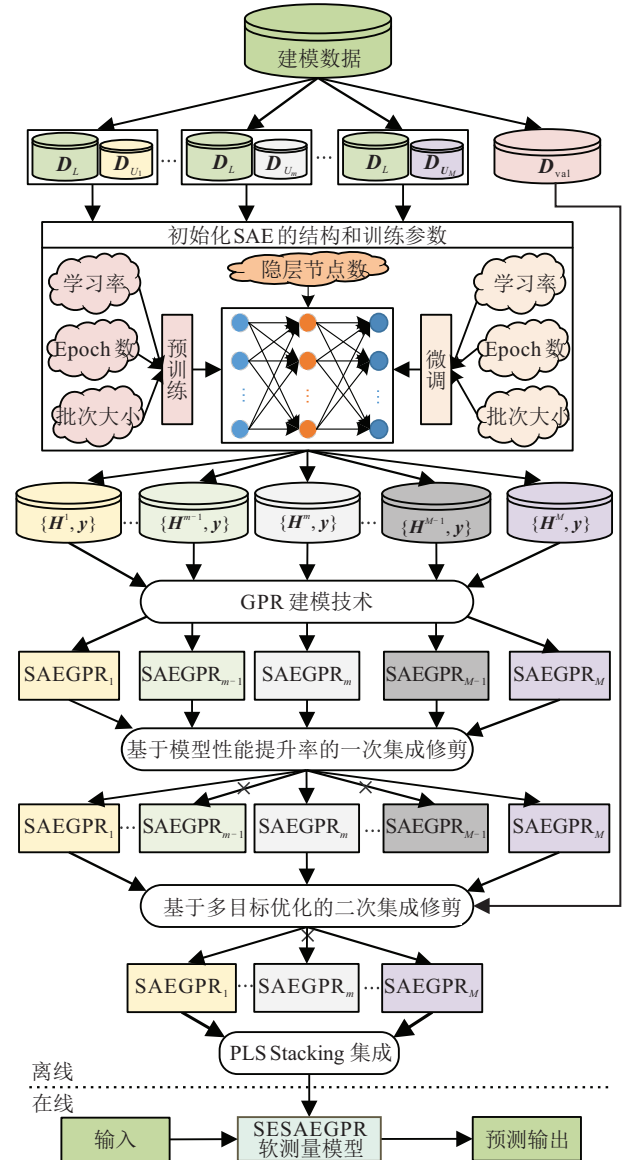


图4 SESAEGPR 软测量方法原理

升幅度,基于性能提升率实现一次集成修剪,剔除泛化性能较差的SAEGPR基模型。

step 4: 以 step 3 获得的 SAEGPR 基模型为候选对象,构建集成修剪多目标优化问题,采用 NSGA-II 算法实现问题求解。

step 5: 选择一个满意的 Pareto 解,将其解码后获得二次集成修剪后的 SAEGPR 基模型。

step 6: 构建 PLS Stacking 集成模型。

step 7: 在线实施阶段,将测试样本提供给每一个选中的 SAEGPR 基模型,计算获得相应的估计均值和方差,并采用 PLS Stacking 集成模型将其融合为最终的估计结果。

3 应用研究

为了验证所提出方法 SESAEGPR 的有效性和优越性,在青霉素发酵过程和 Tennessee Eastman 化工过程中开展应用研究,参与对比的方法有:

- 1) GPR: 高斯过程回归软测量模型;
- 2) SAE: 堆栈自编码回归软测量模型;
- 3) BPNN: 前馈线性回归软测量模型;
- 4) SAEGPR: 基于隐特征提取的高斯过程回归软测量模型;
- 5) ESAEGPR: 融合所有 SAEGPR 基模型的集成软测量模型,未考虑集成修剪;
- 6) SESAEGPR: 本文所提出方法。

采用均方根误差 (root-mean-square error, RMSE) 和决定系数 (R^2) 作为模型的性能评价指标,有

$$RMSE = \sqrt{\frac{1}{N_{test}} \sum_{i=1}^{N_{test}} (\hat{y}_i - y_i)^2}, \quad (28)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N_{test}} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{N_{test}} (\hat{y}_i - \bar{y})^2}. \quad (29)$$

其中: N_{test} 为测试样本数, \hat{y}_i 和 y_i 分别为第 i 个测试样本的估计值、实际值, \bar{y} 为实际值的均值。

3.1 青霉素发酵过程

3.1.1 过程描述

青霉素是一种二次代谢产物,其生产制备是一个典型的多时段间歇生产过程,广泛用于间歇过程的建模、监测和控制的研究。本文以青霉素浓度作为主导变量,以表 2 所示辅助变量作为输入变量,开展软测量建模研究。

建模数据来源于 PenSim 2.0^[38] 仿真实验平台。将仿真周期设为 400 h, 采样间隔设为 2 h, 共采集 20

表2 用于青霉素发酵过程软测量建模的辅助变量

序号	变量/单位	序号	变量/单位
1	培养时间/h	8	底物进料温度/K
2	曝气率/(L·h ⁻¹)	9	溶解氧浓度/(g·L ⁻¹)
3	搅拌功率/W	10	培养体积/L
4	底物进料速度/(L·h ⁻¹)	11	二氧化碳浓度/(g·L ⁻¹)
5	产生热量/kcal	12	发酵温度/K
6	酸流量/(L·h ⁻¹)	13	pH 值
7	底物流量/(L·h ⁻¹)	14	冷却水流量/(L·h ⁻¹)

批青霉素发酵过程数据,将其划分为标记训练集(8批 1 600 个样本)、验证集(4批 800 个样本)、测试集(4批 800 个样本),并从训练批次中挑选出 1 600 个非标记样本。为了激发隐特征提取结果的多样性,选用 1 600 个标记样本和 400 个随机采样的非标记样本用于每一个 SAE 模型训练。

为了确保软测量模型的泛化性能,需要对一些关键参数进行精心挑选。对比方法通过模型在验证样本上的估计误差评估参数的优劣性,不同方法的参数选择结果如下:

- 1) GPR: 确定协方差函数后无需额外的超参数优化。
- 2) SAE: 结构为 14-6-3-1, 预训练学习率、epoch 数和批次大小分别为 0.2、300、20, 微调学习率、epoch 数和样本批次大小分别为 0.5、500、20。
- 3) BPNN: 结构为 14-7-4-1, 学习率、epoch 数和样本批次大小分别为 0.5、300、20。
- 4) SAEGPR: 用于提取隐特征的 SAE 结构为 14-6-3-1, 预训练学习率、epoch 数和样本批次大小分别为 0.2、300、20, 微调学习率、epoch 数和样本批次大小分别为 0.5、500、20。
- 5) SESAEGPR: 方法不寻求一组最优的参数设置,而是在如表 3 所示的参数范围内产生多样性 SAE 网络结构和训练配置参数,进而构建多样性 SAEGPR 基模型,然后引入 PLS Stacking 集成实现基模型的融合。

根据表 3 给定的候选参数,共获得 600 组多样性的参数组合,以此构建 600 个 SAEGPR 基模型,并采用两次集成修剪策略实现模型精简。首先,执行一次集成修剪,将性能提升率阈值 I_{rate} 设为 0, 剔除无性能提升的基模型,最终保留 192 个基模型;然后,在第 2 次集成修剪操作中,对 192 个基模型选择进行二进制编码,然后采用 NSGA-II 多目标优化算法实现问题求解。将初始种群数、迭代数和基模型选择数分别设为 400、100 和 25。

表3 SESAEGPR方法中的相关参数及其候选范围

参数	范围
Hnode1	[1:2:20]
Hnode2	[1:2:20]
Lrate1	[0.1:0.1:0.3]
Nepoch1	300
Bsize1	20
Lrate2	[0.2:0.1:0.3]
Nepoch2	600
Bsize2	20

3.1.2 实验结果及分析

表4为不同软测量模型在测试集上的估计性能. 总体来看, 由于集成学习策略的引入, ESAEGPR和SESAEGPR的性能大幅优于全局模型GPR、SAE、BPNN和SAEGPR. 而在全局模型中, SAEGPR表现最佳, 这主要得益于GPR建模技术和SAE隐特征提取的综合优势. 但是, SAEGPR难免会因为建模样本的选择及模型参数设置的多样性导致模型性能呈现出一定程度的不确定性. 为此, ESAEGPR通过融合集成学习有效提升了估计性能, 但该方法依然存在基模型数过多、模型复杂度较高的问题. 相比而言, SESAEGPR采用两次集成修剪策略, 不仅大幅削减了基模型规模, 而且进一步改善了模型估计性能. 这些结果充分表明所提出方法相较于传统的全局模型和全集成模型具有较大的优越性. 图5给出了GPR、ESAEGPR和SESAEGPR三种方法的估计结果趋势曲线. 如图5可见, GPR估计结果趋势曲线平滑度较差, 偏离真实值幅度较大, 部分估计置信区间较宽, 反映出GPR模型估计结果不佳. 相比而言, ESAEGPR的估计结果明显改善, 置信区间较窄但边缘呈锯齿状, 表明并非参与集成的所有基模型都有优异的估计性能. SESAEGPR的估计结果与真实值高度吻合, 估计置信区间较窄, 表明该方法具有良好的估计性

表4 不同软测量方法对青霉素浓度的估计结果比较

参数	范围	RMSE	R^2
GPR	单模型	0.0287	0.9961
SAE	单模型	0.0265	0.9967
BPNN	单模型	0.0284	0.9961
SAEGPR	单模型	0.0246	0.9971
ESAEGPR	全集成	0.0217	0.9978
SESAEGPR	选择性集成	0.0195	0.9982

能. 总体上看, ESAEGPR和SESAEGPR相较于GPR而言均有大幅的性能提升. 但从模型实施效率上而言, SESAEGPR由于采用了两次集成修剪策略, 不仅大幅降低了集成模型复杂度, 而且进一步显著提升了估计精度, 在实际运用中更具实用性和吸引力.

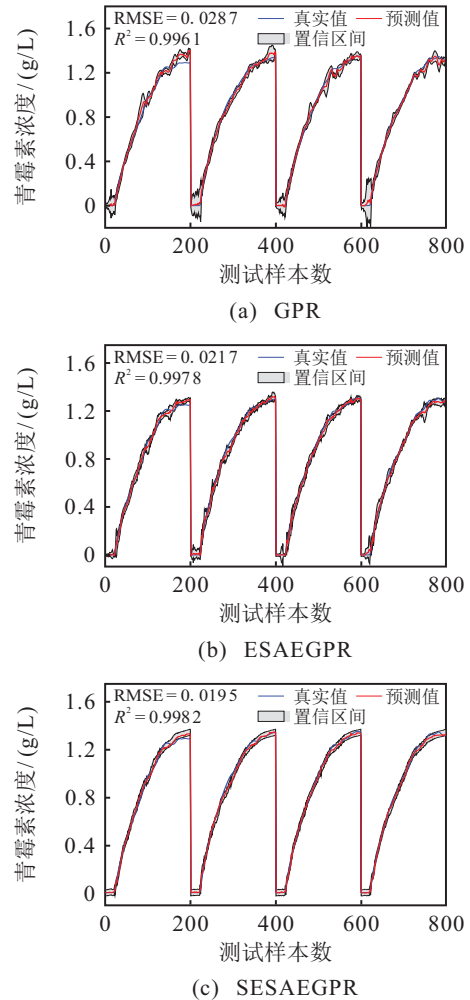


图5 青霉素浓度估计趋势曲线

综上, 由于综合了深度学习隐特征提取、集成学习和双重集成修剪的优势, 所提出SESAEGPR软测量方法在青霉素浓度估计任务中表现出良好的估计性能, 与常规全局模型和全集成模型相比, 该方法优势明显.

3.2 TE化工过程

3.2.1 过程描述

TE(tennessee eastman process)化工过程^[39]是一个模拟实际工业过程的仿真案例. 该生产过程主要有A、C、D和E四种气态物料参与反应, 生产产品G和H, 而F为生产过程中的副产品. TE过程总计有53个过程变量, 包括41个检测变量和12个操纵变量. 将其中23个连续测量变量和9个操纵变量作为模型输入(见表5), 以Stream6中的C成分浓度作为模型输出, 开展软测量建模研究. 为了验证所提出方法在多

模式过程中的有效性,如表6所示,根据G/H质量比和生产率的变化设置3种操作模式,每个模式采集721个标记样本,共计2136个标记样本,将其进一步划分

为:1083个训练样本,540个验证样本,540个测试样本.此外,采集2160个非标记样本.隐特征提取过程中随机选择800个非标记样本加入SAE模型训练.

表5 TE过程用于构建SESAEGPR模型的辅助变量

序号	变量/单位	序号	变量/单位
1	时间/h	17	汽提塔压力/kPa
2	A进料(流1)/(km ³ ·h ⁻¹)	18	汽提塔塔底流量(流11)/(km ³ ·h ⁻¹)
3	D进料(流2)/(kg·h ⁻¹)	19	汽提塔温度/°C
4	E进料(流3)/(kg·h ⁻¹)	20	汽提塔流量/(kg·h ⁻¹)
5	A和C总进料(流4)/(km ³ ·h ⁻¹)	21	压缩机功率/kW
6	再循环流量(流8)/(km ³ ·h ⁻¹)	22	反应器冷却水出口温度/°C
7	反应器进料速度(流6)/(km ³ ·h ⁻¹)	23	分离器冷却水出口温度/°C
8	反应器压力/kPa	24	D进料量(流2)/(kg·h ⁻¹)
9	反应器液位/%	25	E进料量(流3)/(kg·h ⁻¹)
10	反应器温度/°C	26	A进料量(流1)/(km ³ ·h ⁻¹)
11	排放速度/(km ³ ·h ⁻¹)	27	A和C进料量(流4)/(km ³ ·h ⁻¹)
12	产品分离器温度/°C	28	排放阀(流9)/%
13	产品分离区液位/%	29	分离罐液体流量(流10)/(km ³ ·h ⁻¹)
14	产品分离器压力/kPa	30	汽提塔液体产品流量(流11)/(m ³ ·h ⁻¹)
15	产品分离器塔底低流量(流10)/(km ³ ·h ⁻¹)	31	反应器冷却水流量/(m ³ ·h ⁻¹)
16	汽提塔液位/%	32	冷凝器冷却水流量/(m ³ ·h ⁻¹)

表6 TE过程的操作模式

操作模式	G/H质量比	生产率/(m ³ /h)
1	50/50	18.04
2	10/90	18.04
3	90/10	18.04

不同软测量方法的参数寻优结果如下:

1) GPR: 无需确定额外的超参数.

2) SAE: 结构为32-22-10-1,预训练学习率、epoch数和样本批次大小分别为0.3、600、20,微调学习率、epoch数和样本批次大小分别为0.3、600、20.

3) BPNN: 结构为32-20-5-1,学习率、epoch数和样本批次大小分别为0.1、800、20.

4) SAEGPR: 用于提取隐特征的SAE结构为32-22-10-1,预训练学习率、epoch数和样本批次大小分别为0.3、600、20,微调学习率、epoch数和样本批次大小分别为0.3、600、20.

5) SESAEGPR: 方法选定的参数候选范围如表7所示.

在ESAEGPR和SESAEGPR建模中,根据多样性隐特征提取生成扰动机制,先产生180组SAE初始化参数,从而构建180个SAEGPR基模型.利用一次集

表7 SESAEGPR方法中的相关参数及其候选范围

参数	范围
Hnode1	[20: 4: 32]
Hnode2	[1: 4: 20]
Lrate1	[0.1: 0.1: 0.3]
Nepoch1	600
Bsize1	20
Lrate2	[0.1: 0.1: 0.3]
Nepoch2	600
Bsize2	20

成修剪,初步保留80个基模型,随后采用NSGA-II多目标优化算法对80个基模型进行二次集成修剪.将初始种群数、迭代数和基模型选择数分别设置为300、100和35.

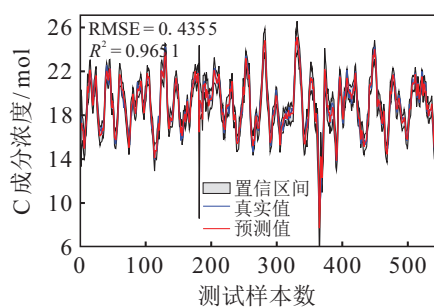
3.2.2 实验结果及分析

基于不同软测量方法的C成分浓度(Stream 6)估计结果如表8所示.不难看出,类似于青霉素浓度估计,在本案例研究中,全局模型估计性能明显不如集成模型.通过比较不同的全局模型估计性能也可以发现,SAEGPR显著优于GPR、SAE和BPNN.进一步分析两种集成模型估计性能可以看出,尽管SESAEGPR大幅削减了基模型数目,但其估计性能

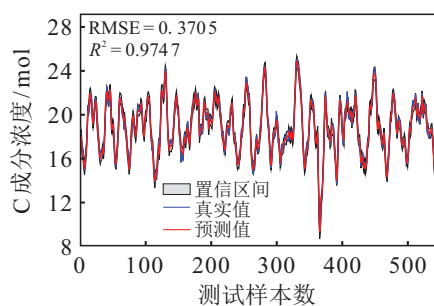
仍然比全局集成模型ESAEGPR稍好,再一次凸显了选择性集成策略在降低集成模型复杂度和保持甚至改善估计性能方面的显著优势.此外,图6比较了GPR、ESAEGPR和SESAEGPR方法的估计结果趋势曲线.显然,就估计偏差和估计不确定度性能指标而言,两种集成模型显著优于全局模型.这些实验结果再一次表明,相较于全局建模和全集成建模,所提出方法由于融合了多样性隐特征提取和双重集成修剪策略而展现出更为优异的估计性能.

表8 不同软测量方法对TE过程的估计结果比较

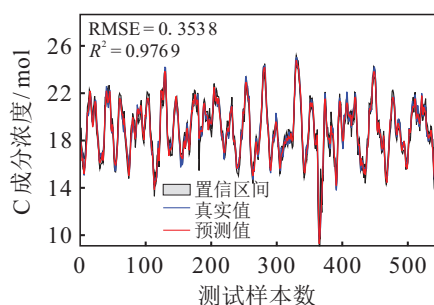
参数	范围	RMSE	R^2
GPR	单模型	0.4355	0.9651
SAE	单模型	0.4821	0.9572
BPNN	单模型	0.4826	0.9570
SAEGPR	单模型	0.4108	0.9689
ESAEGPR	全集成	0.3705	0.9747
SESAEGPR	选择性集成	0.3538	0.9769



(a) GPR



(b) ESAEGPR



(c) SESAEGPR

图6 TE过程中C成分浓度(Stream 6)的估计趋势曲线

4 结论

本文针对单一模型估计性能不佳以及传统特征提取方法难以挖掘复杂过程特征的问题,提出了一种基于深度学习多样性特征提取与多目标优化的选择性集成学习软测量建模方法.主要贡献为:1)充分发挥深度学习在挖掘复杂过程数据特征方面的优越性,提出了一种基于堆栈自编码器的多样性隐特征提取方法,为构建多样性的基模型提供了一种新的思路,为构建高性能的集成学习软测量模型奠定了基础.2)提出了一种双重集成修剪策略,首先根据模型性能提升率完成一次集成修剪,然后将基模型选择问题转化为多目标优化问题,并通过进化多目标优化实现问题求解.该策略的实施将有助于挑选合适的基模型,降低集成模型复杂度,同时保持甚至进一步改进模型估计性能.3)通过青霉素发酵过程和TE化工过程对所提出方法进行了应用验证,结果充分表明所提出方法SESAEGPR相较于传统全局建模和全集成建模的显著优越性.

本研究是综合集成学习和深度学习优势以构建高性能软测量模型的一次初步尝试,但还有如下问题值得进一步研究:

1) 基模型的多样性是确保集成建模成功的关键,然而产生模型多样性的途径并不唯一.因此,将深度学习多样性隐特征提取与其他多样性扰动机制相结合将为构造性能更强的集成软测量模型提供可能.

2) 大量研究表明,在集成学习中普遍存在“多胜于全(many could be better than all)”的现象,表明集成修剪的不可或缺性.尽管所提出的双重集成修剪策略表现良好,但是如何显性平衡估计精度和多样性指标并通过多目标优化方法实现集成修剪仍存有很大的探索空间.

3) 深度学习研究成果日新月异,新型隐特征提取技术正在大量涌现,如何将该领域最新研究成果与所提出软测量建模框架相结合也是值得探索的课题.

参考文献(References)

- [1] Yuan X F, Ge Z Q, Huang B, et al. A probabilistic just-in-time learning framework for soft sensor development with missing data[J]. IEEE Transactions on Control Systems Technology, 2017, 25(3): 1124-1132.
- [2] Alford J S. Bioprocess control: Advances and challenges[J]. Computers & Chemical Engineering, 2006, 30(10/11/12): 1464-1475.
- [3] Liu Y, Yang C, Liu K X, et al. Domain adaptation transfer learning soft sensor for product quality prediction[J]. Chemometrics and Intelligent Laboratory Systems, 2019,

- 192: 103813.
- [4] Fortuna L, Graziani S, Rizzo A, et al. Soft sensors for monitoring and control of industrial processes[M]. London: Springer, 2007: 27-51.
- [5] 熊伟丽, 孙文心, 马君霞. 带自相关约束的NARX动态软测量模型[J]. 控制与决策, 2020, 35(4): 816-822. (Xiong W L, Sun W X, Ma J X. Autocorrelation constrained NARX dynamic soft sensing model[J]. Control and Decision, 2020, 35(4): 816-822.)
- [6] Yuan X F, Ge Z Q, Huang B, et al. Semisupervised JITL framework for nonlinear industrial soft sensing based on locally semisupervised weighted PCR[J]. IEEE Transactions on Industrial Informatics, 2017, 13(2): 532-541.
- [7] Shao W M, Tian X M, Wang P. Soft sensor development for nonlinear and time-varying processes based on supervised ensemble learning with improved process state partition[J]. Asia-Pacific Journal of Chemical Engineering, 2015, 10(2): 282-296.
- [8] Wang Y L, Wu D Z, Yuan X F. A two-layer ensemble learning framework for data-driven soft sensor of the diesel attributes in an industrial hydrocracking process[J]. Journal of Chemometrics, 2019, 33(12): e3185.
- [9] 马建, 邓晓刚, 王磊. 基于深度集成支持向量机的工业过程软测量方法[J]. 化工学报, 2018, 69(3): 1121-1128. (Ma J, Deng X G, Wang L. Industrial process soft sensor method based on deep learning ensemble support vector machine[J]. CIESC Journal, 2018, 69(3): 1121-1128.)
- [10] Jin H P, Shi L X, Chen X G, et al. Probabilistic wind power forecasting using selective ensemble of finite mixture Gaussian process regression models[J]. Renewable Energy, 2021, 174: 1-18.
- [11] Sun Q Q, Ge Z Q. A survey on deep learning for data-driven soft sensors[J]. IEEE Transactions on Industrial Informatics, 2021, 17(9): 5853-5866.
- [12] Zhou Z H. Ensemble methods[M]. New York: Chapman and Hall/CRC, 2012: 186-229.
- [13] Shao W M, Tian X M. Adaptive soft sensor for quality prediction of chemical processes based on selective ensemble of local partial least squares models[J]. Chemical Engineering Research and Design, 2015, 95: 113-132.
- [14] 邵伟明, 田学民, 宋执环. 基于集成学习的多产品化工过程软测量建模方法[J]. 化工学报, 2018, 69(6): 2551-2559. (Shao W M, Tian X M, Song Z H. Ensemble learning-based soft sensor method for multi-product chemical processes[J]. CIESC Journal, 2018, 69(6): 2551-2559.)
- [15] 阎高伟, 贺敏, 汤健, 等. 基于最大均值差异多源域迁移学习的湿式球磨机负荷参数软测量[J]. 控制与决策, 2018, 33(10): 1795-1800. (Yan G W, He M, Tang J, et al. Soft sensor of wet ball mill load based on maximum mean discrepancy multi-source domain transfer learning[J]. Control and Decision, 2018, 33(10): 1795-1800.)
- [16] Sheng X C, Ma J X, Xiong W L. Smart soft sensor design with hierarchical sampling strategy of ensemble Gaussian process regression for fermentation processes[J]. Sensors: Basel, Switzerland, 2020, 20(7): 1957.
- [17] Sun Q Q, Ge Z Q. Deep learning for industrial KPI prediction: When ensemble learning meets semi-supervised data[J]. IEEE Transactions on Industrial Informatics, 2021, 17(1): 260-269.
- [18] Brown G, Wyatt J, Harris R, et al. Diversity creation methods: A survey and categorisation[J]. Information Fusion, 2005, 6(1): 5-20.
- [19] Pan B, Jin H P, Yang B, et al. Soft sensor development for nonlinear industrial processes based on ensemble just-in-time extreme learning machine through triple-modal perturbation and evolutionary multiobjective optimization[J]. Industrial & Engineering Chemistry Research, 2019, 58(38): 17991-18006.
- [20] 金怀平, 黄思, 王莉, 等. 基于进化多目标优化的选择性集成学习软测量建模[J]. 高校化学工程学报, 2019, 33(3): 680-691. (Jin H P, Huang S, Wang L, et al. Selective ensemble learning based on evolutionary multi-objective optimization for soft sensor development[J]. Journal of Chemical Engineering of Chinese Universities, 2019, 33(3): 680-691.)
- [21] Kaneko H, Funatsu K. Adaptive soft sensor based on online support vector regression and Bayesian ensemble learning for various states in chemical plants[J]. Chemometrics and Intelligent Laboratory Systems, 2014, 137: 57-66.
- [22] Guyon I, Elisseeff A. An introduction to variable and feature selection[J]. Journal of Machine Learning Research, 2003, 3: 1157-1182.
- [23] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554.
- [24] 袁小锋, 王雅琳, 阳春华, 等. 深度学习在流程工业过程数据建模中的应用[J]. 智能科学与技术学报, 2020, 2(2): 107-115. (Yuan X F, Wang Y L, Yang C H, et al. The application of deep learning in data-driven modeling of process industries[J]. Chinese Journal of Intelligent Science and Technology, 2020, 2(2): 107-115.)

- [25] Yuan X F, Huang B, Wang Y L, et al. Deep learning-based feature representation and its application for soft sensor modeling with variable-wise weighted SAE[J]. IEEE Transactions on Industrial Informatics, 2018, 14(7): 3235-3243.
- [26] Yuan X F, Ou C, Wang Y L, et al. A novel semi-supervised pre-training strategy for deep networks and its application for quality variable prediction in industrial processes[J]. Chemical Engineering Science, 2020, 217: 115509.
- [27] Sun Q Q, Ge Z Q. Gated stacked target-related autoencoder: A novel deep feature extraction and layerwise ensemble method for industrial soft sensor application[J]. IEEE Transactions on Cybernetics, 2020, 16(7): 4456-4467.
- [28] Zhou Z H, Wu J X, Tang W. Ensembling neural networks: Many could be better than all[J]. Artificial Intelligence, 2002, 137(1/2): 239-263.
- [29] Tsoumakas G, Partalas I, Vlahavas I. An ensemble pruning primer[M]. Berlin, Heidelberg: Springer, 2009: 155-165.
- [30] 刘卓, 柴天佑, 汤健. 基于多尺度振动和振声频谱特征自适应提取与选择的磨机负荷参数软测量[J]. 控制与决策, 2019, 34(12): 2603-2610.
(Liu Z, Chai T Y, Tang J. Soft sensing modeling of mill load parameter based on adaptive feature selection and extraction of multi-scale vibration and acoustic frequency spectrum[J]. Control and Decision, 2019, 34(12): 2603-2610.)
- [31] Liu T, Chen S, Liang S, et al. Selective ensemble of multiple local model learning for nonlinear and nonstationary systems[J]. Neurocomputing, 2020, 378: 98-111.
- [32] Williams C K, Rasmussen C E. Gaussian processes for machine learning[M]. Cambridge: MIT Press, 2006: 15-28.
- [33] Goodfellow I, Bengio Y, Courville A. Deep learning[M]. Cambridge: MIT Press, 2016, 22(4): 351-354.
- [34] Yuan X F, Gu Y J, Wang Y L, et al. A deep supervised learning framework for data-driven soft sensor modeling of industrial processes[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 31(11): 4737-4746.
- [35] Yu L, Wu Y, Tang L, et al. Investigation of diversity strategies in RVFL network ensemble learning for crude oil price forecasting[J]. Soft Computing, 2021, 25(5): 3609-3622.
- [36] Deb K, Agrawal S, Pratap A, et al. A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II[C]. Parallel Problem Solving from Nature. Berlin, Heidelberg: Springer, 2000: 849-858.
- [37] Mei C L, Yin L, Yang M, et al. A multi-model soft sensor development using Gaussian process regression for fermentation processes[J]. Computers and Applied Chemistry, 2016, 33(12): 1279-1285.
- [38] Birol G, Ündey C, Çinar A. A modular simulation package for fed-batch fermentation: Penicillin production[J]. Computers & Chemical Engineering, 2002, 26(11): 1553-1565.
- [39] Downs J J, Vogel E F. A plant-wide industrial process control problem[J]. Computers & Chemical Engineering, 1993, 17(3): 245-255.

作者简介

金怀平(1987—), 男, 副教授, 博士, 从事复杂工业过程建模、软测量技术与应用等研究, E-mail: jinhuaping@126.com;

王建军(1995—), 男, 硕士生, 从事复杂工业过程建模的研究, E-mail: wangjianjun195@126.com;

董守龙(1979—), 男, 讲师, 博士, 从事在线检测与控制技术、智能仪表及嵌入式系统开发等研究, E-mail: sldong@bit.edu.cn;

钱斌(1976—), 男, 教授, 博士生导师, 从事智能优化调度理论与方法等研究, E-mail: bin.qian@vip.163.com;

杨彪(1974—), 男, 教授, 博士, 从事复杂过程建模、微波能等新能源的优化与评估等研究, E-mail: yb_chenggong@163.com;

陈祥光(1953—), 男, 教授, 博士生导师, 从事智能测控技术研究、智能仪器仪表研发技术等研究, E-mail: xgc1@bit.edu.cn.

(责任编辑: 郑晓蕾)