

DOI: 10.11949/0438-1157.20190635

优化增量型随机权神经网络及应用

姜乐, 周平

(东北大学流程工业综合自动化国家重点实验室, 辽宁 沈阳 110819)

摘要: 针对传统增量型随机权神经网络(I-RVFLNs)存在网络参数难以优化确定、模型收敛速度慢和结构复杂的问题, 提出一种优化增量型随机权神经网络算法, 即O-I-RVFLNs。与传统I-RVFLNs不同, 所提O-I-RVFLNs算法首先设定了一个期望的建模残差向量, 然后在每次新增隐层节点时, 选择可以达到或小于此节点期望残差的输入权值和偏置作为该节点的输入参数, 进而提高网络的收敛速度。除此之外, 考虑到算法在不断迭代更新过程中建模误差越来越小, 下降趋势越来越不明显的问题, 将各指标参数相邻两次迭代均方根误差的差值考虑在算法终止条件内, 并借鉴统计过程控制中的西电规则制定了相应的算法收敛判定准则。最后, 基于UCI能效数据和实际高炉工业数据, 对所提O-I-RVFLNs算法进行了验证和应用。结果表明, 相对于其他RVFLNs算法, 所提算法建立的数据模型能够获得更紧凑的网络结构以及更好的泛化性能和预测精度。

关键词: 神经网络; 过拟合; 高炉炼铁; 优化; 动态建模

中图分类号: TQ 028.8

文献标志码: A

文章编号: 0438-1157 (2019) 12-4710-12

Optimized incremental random vector functional-link networks and its application

JIANG Yue, ZHOU Ping

(State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110819, Liaoning, China)

Abstract: Aiming at the problem that network parameters are difficult to be optimally determined, the model convergence speed is slow and the structure is complex in the traditional incremental random vector functional-link networks (I-RVFLNs), an optimized incremental random vector functional-link networks algorithm, namely O-I-RVFLNs, is proposed. Different from the traditional I-RVFLNs, the proposed O-I-RVFLNs algorithm sets a desired residual error vector, and then selects the input weights and biases that can reach or less than the expected residual error as the input parameters of the node each time a hidden node is added, thereby improving the convergence rate of the network. In addition, considering that the modeling error of the algorithm is smaller and smaller and the downward trend is less obvious in the process of continuous iteration updating, the RMSE difference between adjacent iterations of each index parameter is considered in the termination condition of the algorithm, and the corresponding convergence criteria are formulated by referring to the Western Electricity Rules in statistical

收稿日期: 2019-06-10 修回日期: 2019-08-12

通信作者: 周平(1980—), 男, 博士, 教授, 博士生导师, zhouping@mail.neu.edu.cn

第一作者: 姜乐(1994—), 女, 硕士研究生, 1649577095@qq.com

基金项目: 国家自然科学基金项目(61890934, 61473064, 61790572, 61890930); 中央高校基本科研业务费专项资金(N180802003); 矿冶过程自动控制技术国家(北京市)重点实验室开放课题(BGRIMM-KZSKL-2017-04)

引用本文: 姜乐, 周平. 优化增量型随机权神经网络及应用[J]. 化工学报, 2019, 70(12): 4710–4721

Citation: JIANG Yue, ZHOU Ping. Optimized incremental random vector functional-link networks and its application[J]. CIESC Journal, 2019, 70 (12): 4710–4721

process control. Finally, based on UCI energy efficiency data and actual blast furnace industrial data, the proposed O-I-RVFLNs algorithm is verified and applied. The results show that compared with other RVFLNs algorithms, the data model built by the proposed algorithm can obtain more compact network structure, better generalization performance and prediction accuracy.

Key words: neural networks; overfitting; blast furnace ironmaking; optimization; dynamic modeling

引 言

作为一种简单而特殊的前馈型神经网络(FNN),单隐层前馈网络(SLFN)在实际应用中起到了非常重要的作用,其中随机权神经网络(RVFLNs)作为其典型的代表,在过去几十年中,进行了理论和应用上的广泛研究^[1-4]。随机权神经网络(RVFLNs)是由Pao等^[1-2]首次提出,由于其训练速度快,易于实现且泛化能力强等特点,受到了学术界和工程界的广泛关注。目前,RVFLNs已被广泛应用于处理分类、回归、预测和控制等实际问题^[5-9]。RVFLNs算法的最大特点是网络的输入权值和偏置在特定范围内随机选取,仅需根据最小二乘思想调整网络输出权值,以最小化训练误差,因而学习速度极快,计算过程极为简化^[5]。研究表明:若RVFLNs的输入权值和偏置在满足一定条件的概率空间选取,则网络的计算输出可以以任意小的误差逼近期望输出^[1]。Huang等^[10]证明了如果隐层激活函数是无限可微的,则具有随机输入权值和偏置的单隐层前馈神经网络可以以任意精度逼近任意连续函数,进而在此基础上提出了极限学习机(ELM)。在ELM中,网络的输入权值与偏置无关,隐层参数完全独立于给定的目标函数和建模数据,通过随机分配隐层节点的输入权值和偏置,然后采用最小二乘法即可获得输出权值^[10-12]。此方法提高了网络的训练速度,有效避免网络训练时陷入局部极小值的问题。

基本RVFLNs中,网络中的隐层节点数需要预先设定,设定数值的大小直接决定了网络的学习和泛化性能以及计算效率^[5,7,13]。然而,目前还没有一种系统、有效的方法来确定RVFLNs最佳隐层节点数目。众所周知,任何神经网络若隐层节点数太少,则网络不能很好地学习给定问题;而若隐层节点数太多,则会导致网络学习速度过慢,且易出现“过拟合”问题。因此,研究能够自动找到神经网络适当数量隐层节点的算法非常必要且意义重大。Huang等^[14]于2006年提出了一种增量型结构的随机权神经网络,即增量RVFLNs(I-RVFLNs),并将其命

名为增量极限学习机(I-ELM)。I-RVFLNs在训练过程中,隐层节点逐一地增加到网络中,而且对于每个新增的隐层节点,均在给定范围内随机生成其对应的输入权值和偏置,然后调整输出权值来不断逼近网络输出,因此模型计算效率较高^[14-15]。虽然I-RVFLNs算法有效地解决了RVFLNs的过拟合问题,并且可以自动计算隐层节点的数目。但是,当算法满足终止条件时,网络中的隐层节点数过多且很多隐层节点对网络残差的减小贡献很小,因而导致网络结构过于复杂,模型训练时间较长^[16-18]。

为解决传统I-RVFLNs存在的上述问题,文献[17]提出了一种增强增量型随机权神经网络(EI-RVFLNs)。该算法提高了I-RVFLNs的收敛速度,但由于对网络中各隐层节点的输入参数进行了筛选,因此模型训练时间过长。文献[18]将Barron的凸优化方法^[19]引入到I-RVFLNs中,提出一种凸增量随机权神经网络(CI-RVFLNs)。传统I-RVFLNs算法在训练过程中,每当新增加一个隐层节点时,仅计算该新增隐层节点的输出权值,所有已有隐层节点的输出权值均被固定,因此算法收敛速度较慢。而CI-RVFLNs中,每当新增隐层节点时,所有已有隐层节点的输出权值均被重新计算。实验结果表明,该算法在保持传统I-RVFLNs的简单性和高效性的同时可以实现更快的收敛速度,但是该算法只适用于多输入单输出问题。

针对上述问题,本文提出一种优化增量型随机权神经网络算法,即O-I-RVFLNs算法。与传统I-RVFLNs不同的是,所提算法在模型训练的过程中,通过设定一个期望的残差向量,并在每次新增隐层节点时,选择可以达到或小于此节点期望残差的输入权值和偏置作为该节点的输入参数,以此类推,直到满足算法终止条件。同时考虑到算法在不断迭代更新过程中各指标参数的误差值越来越小,下降趋势越来越不明显,因此借鉴统计过程监测中的西电规则思想,将各指标参数相邻两次迭代的均方根误差差值考虑在终止条件内。最后,基于标准数据集集中的能效数据以及我国华南某大型高炉的实

际数据建立基于 O-I-RVFLNs 的多指标预测模型,并与基本 RVFLNs、I-RVFLNs 和 EI-RVFLNs 算法进行对比实验。

1 O-I-RVFLNs 算法

1.1 传统 I-RVFLNs 算法简介

2006 年, Huang 等^[14]在固定型随机权神经网络的基础上提出了增量型结构的随机权神经网络(I-RVFLNs)。针对 N 个任意样本集合 $(\mathbf{x}_i, \mathbf{t}_i)$, $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{in}] \in \mathbf{R}^n$, $\mathbf{t}_i = [t_{i1}, t_{i2}, \dots, t_{im}] \in \mathbf{R}^m$, 具有 L 个隐含层节点的单隐层前馈网络为

$$\sum_{j=1}^L \beta_j \phi_j(\mathbf{w}_j \mathbf{x}_i + b_j) = \mathbf{t}_i, \mathbf{w}_j \in \mathbf{R}^n, b_j \in \mathbf{R}, i = 1, \dots, N \quad (1)$$

式中, $\mathbf{w}_j$ 表示输入层神经元与第 j 个隐含层神经元之间的随机权值, β_j 表示第 j 个隐含层节点与输出层节点之间的输出权值, b_j 表示隐含层上第 j 个节点的偏置, ϕ_j 是隐含层节点的输出函数。

根据文献[14]的理论, 针对线性节点给定任意一个有界非常数分段连续函数 $\phi: \mathbf{R} \rightarrow \mathbf{R}$ 且 $\int_{\mathbf{R}} \phi(\mathbf{x}) d\mathbf{x} \neq 0$, 并设隐层随机输出矩阵为 H_L , 则对于任意线性目标函数 f 有 $\lim_{n \rightarrow \infty} \|f - f_L\| = 0$ 以概率 1 存在, 只要输出权值 β 为

$$\beta_j = \frac{E_{j-1} H_j^T}{H_j H_j^T} = \frac{\sum_{i=1}^N \mathbf{e}(i) h(i)}{\sum_{i=1}^N h^2(i)} \quad (2)$$

式中, $h(i)$ 是第 i 个输入样本对应的第 j 个隐含层节点的激活函数值, $\mathbf{e}(i)$ 是未加入新增隐层节点时对应的剩余残差, $H_j = [h(1), \dots, h(N)]$ 是 N 个训练样本对应的第 j 个隐层节点的激活相量, $E_{j-1} = [\mathbf{e}(1), \dots, \mathbf{e}(N)]$ 是未加入新增隐层节点时对应的残差矩阵。

加入新增隐层节点 j 后的剩余残差 $\mathbf{E}$ 可计算为

$$\mathbf{E}_j = \mathbf{E}_{j-1} - \beta_j H_j \quad (3)$$

在传统的 I-RVFLNs 中, 隐层节点逐一增加到网络中, 并且在每个新的隐层节点加入时, 所有现存隐层节点的输出权值均固定不变。因此, 模型的计算复杂度较高, 且最终得到的模型结构也比较复杂。

1.2 O-I-RVFLNs 算法描述

常规 I-RVFLNs 中, 各隐层节点的输入权值 $\mathbf{w}$ 和偏置 b 是随机生成的, 但从实验中发现, 不同的输入权值和偏置使得残差减小的程度不同, 进而影响

模型的收敛速度和模型结构的复杂度。针对此问题, 本文提出一种优化增量型随机权神经网络, 即 O-I-RVFLNs, 算法策略如图 1 所示。在每次新增加一个隐层节点到网络中时, 均对其输入权值和偏置进行筛选, 即选择影响网络残差减小程度较大的参数作为新增隐层节点的输入参数, 进而提高模型的收敛速度。

算法训练过程中, 首先设定一组期望的残差向量 $\mathbf{E}_r = [\mathbf{E}_{r1}, \mathbf{E}_{r2}, \dots, \mathbf{E}_{rj_{\max}}]$, 并设定输入权值 $\mathbf{w}$ 及隐层偏置 b 的取值范围, 然后逐一增加网络中的隐层节点数 j , 并随机生成新增隐层节点 j 的输入权值 $\mathbf{w}_j^k, 1 \leq k \leq k_{\max}$ 及偏置 b_j^k , 进而得到隐层节点 j 的输出权值 β_j^k 为

$$\begin{cases} \beta_j^k = \mathbf{E}_{j-1}^k (H_j^k)^T / (H_j^k (H_j^k)^T) \\ H_j^k = [h_j^k(1), \dots, h_j^k(N)] \\ \mathbf{E}_{j-1}^k = [\mathbf{e}_{j-1}^k(1), \dots, \mathbf{e}_{j-1}^k(N)] \end{cases} \quad (4)$$

式中, H_j^k 是新增节点 j 对应的 N 个训练样本的激活向量, $\mathbf{E}_{j-1}^k$ 是未加入新增隐层节点 j 时对应的残差矩阵, 加入新增隐层节点 j 后的网络剩余残差 $\mathbf{E}_j^k$ 为

$$\mathbf{E}_j^k = \mathbf{E}_{j-1}^k - \beta_j^k H_j^k \quad (5)$$

当对应隐层节点 j 的第 k 组输入参数满足

$$\|\mathbf{E}_j^k\| \leq \mathbf{E}_{\eta}, k \leq k_{\max}$$

则所获得的各参数为

$$\begin{cases} \mathbf{w}'_j = \mathbf{w}_j^k \\ b'_j = b_j^k \\ \beta'_j = \beta_j^k \\ \mathbf{E}'_j = \mathbf{E}_j^k \end{cases} \quad (6)$$

若对应隐层节点 j 的 $k_{\max}$ 组输入参数均满足 $\|\mathbf{E}_j^k\| > \mathbf{E}_{\eta}, k = 1, 2, \dots, k_{\max}$, 则隐层节点 j 的各参数为

$$\begin{cases} \mathbf{w}'_j = \mathbf{w}_j^{k'} \\ b'_j = b_j^{k'} \\ \beta'_j = \beta_j^{k'} \\ \mathbf{E}'_j = \mathbf{E}_j^{k'} \end{cases} \quad (7)$$

其中 k' 为

$$k' = \left\{ k \mid \min \|\mathbf{E}_j^k\|, 1 \leq k \leq k_{\max} \right\}$$

期望残差向量 $\mathbf{E}_r = [\mathbf{E}_{r1}, \mathbf{E}_{r2}, \dots, \mathbf{E}_{rj_{\max}}]$ 的设定规则为: 针对传统的 I-RVFLNs, 在模型训练过程中, 对每个隐层节点的输入参数均随机选择 c 组, 然后分别计算每组参数对应当前隐层节点 j 的输出权值 β_j 以及残差 $\mathbf{E}_j$, 选择 c 组 $\|\mathbf{E}_j\|$ 中最小的一组作为所提 O-I-RVFLNs 算法中隐层节点 j 的期望残差 $\mathbf{E}_{\eta}$ 。

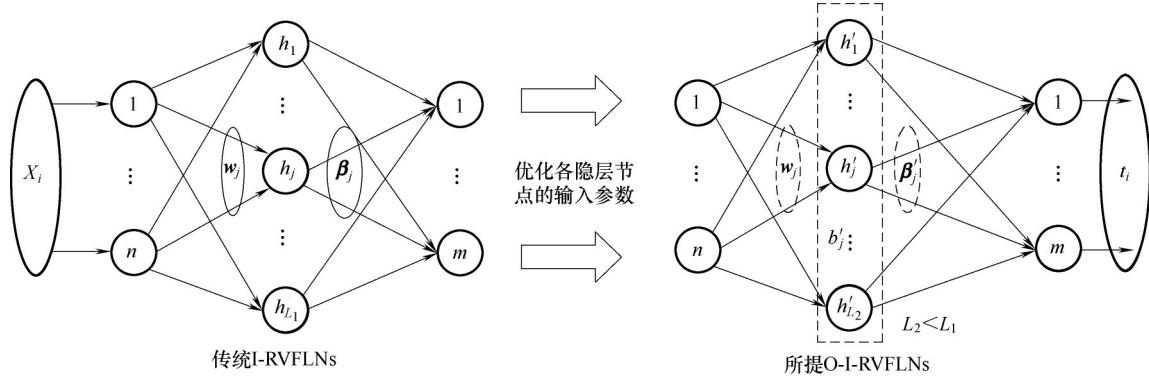


图1 优化增量型随机神经网络策略图

Fig.1 Strategic diagram of optimized incremental random vector functional-link networks

1.3 迭代终止条件改进

针对输入维度为 n 、输出维度为 m 的多输入多输出数据建模问题,从实验中发现,随着算法迭代次数的不断增加,各指标参数的误差值逐渐减小。但当迭代次数大于一定值时,各输出指标拟合误差值的下降趋势越来越不明显,近乎趋于定值,从而严重影响算法训练时间。对此,为缩短模型的训练时间,简化数据模型结构,本文除了将传统 I-RVFLNs 中的最大隐层节点数和期望误差作为模型训练的终止条件外,同时将各估计指标相邻两次迭代的均方根误差差值的大小也考虑在终止条件内,提出如下多条件综合约束的算法迭代终止判定

$$\begin{cases} \varepsilon_{1j} = \max(|e_{1j} - e_{1j-1}|, |e_{1j-1} - e_{1j-2}|, \dots, |e_{1j-7} - e_{1j-8}|) \\ \varepsilon_{2j} = \max(|e_{2j} - e_{2j-1}|, |e_{2j-1} - e_{2j-2}|, \dots, |e_{2j-7} - e_{2j-8}|) \\ \vdots \\ \varepsilon_{mj} = \max(|e_{mj} - e_{mj-1}|, |e_{mj-1} - e_{mj-2}|, \dots, |e_{mj-7} - e_{mj-8}|) \end{cases} \quad (j > 8) \quad (8)$$

1.4 算法实现步骤

优化增量型随机神经网络算法实现步骤概括如下,包括初始化阶段和学习阶段两个部分。

初始化阶段: 给定数据集 $Z = \{(\mathbf{x}_i, \mathbf{y}_i) | \mathbf{x}_i \in \mathbf{R}^n, \mathbf{y}_i \in \mathbf{R}^m, i = 1, \dots, N\}$, 输入权值 $\mathbf{w}$ 及隐层偏置 b 的取值范围为 $[-1, 1]$, 激活函数 $\phi(x)$, 最大隐层节点数 $j_{\max}$, 期望残差向量 $\mathbf{E}_r = [\mathbf{E}_{r1}, \mathbf{E}_{r2}, \dots, \mathbf{E}_{rj_{\max}}]$, 期望训练误差 $\mathbf{E}_0 = [e_{10}, e_{20}, \dots, e_{m0}]$ 及相邻两次 RMSE 差值的设定值 $\varepsilon_0 = [\varepsilon_{10}, \varepsilon_{20}, \dots, \varepsilon_{m0}]$, 且设定 $\varepsilon_0 = 10^{-3} \times \mathbf{E}_0$, 当前隐层节点数 $j = 0$, 当前隐层节点 j 对应的输入参数组数 $k, 1 \leq k \leq k_{\max}, k_{\max} = 10$ 。

学习阶段:

While $j < j_{\max}$ and $(e_{1j} > e_{10}, e_{2j} > e_{20}, \dots, e_{mj} > e_{m0})$ and $(\varepsilon_{1j} >$

条件

$$j > j_{\max} \text{ or } (e_{1j} < e_{10}, e_{2j} < e_{20}, \dots, e_{mj} < e_{m0}) \text{ or } (\varepsilon_{1j} < \varepsilon_{10}, \varepsilon_{2j} < \varepsilon_{20}, \dots, \varepsilon_{mj} < \varepsilon_{m0})$$

式中, j 为当前隐层节点数, $j_{\max}$ 为设定的最大隐层节点数, $e_{ij}, e_{i0}, i = 1, \dots, m$, 分别为 m 个输出的实际均方根误差和期望均方根误差, $\varepsilon_{ij}, \varepsilon_{i0}$ 分别为 m 个输出相邻两次迭代均方根误差差值的实际值和设定值。

为防止相邻估计指标的误差差值波动较大对算法训练结果造成干扰,本文借鉴统计过程控制 (SPC) 领域西电规则^[14]思想建立 m 维输出的相邻两次迭代均方根误差差值 $\varepsilon_{1j}, \varepsilon_{2j}, \dots, \varepsilon_{mj}$ 为

$$\varepsilon_{10}, \varepsilon_{20} > \varepsilon_{20}, \dots, \varepsilon_{mj} > \varepsilon_{m0})$$

逐一增加隐层节点 $j: j = j + 1$;

For $k = 1: k_{\max}$

随机生成隐层节点 j 的输入权值 $\mathbf{w}_j^k$ 和偏置 b_j^k ;

根据式(4)计算隐层节点 j 的输出权值;

根据式(5)计算隐层节点 j 加入后的剩余残差;

If $\|\mathbf{E}_j^k\| \leq \mathbf{E}_j$

Break;

End if

End for

If $(k \leq k_{\max} \& \|\mathbf{E}_j^k\| \leq \mathbf{E}_j)$

计算隐层节点 j 对应的各参数: $\mathbf{w}_j = \mathbf{w}_j^k, b_j = b_j^k, \beta_j = \beta_j^k$,

$\mathbf{E}_j = \mathbf{E}_j^k$;

根据式(8)计算各指标相邻两次迭代的 RMSE 差值。

Else

计算 $k' = \{k | \min\|\mathbf{E}_j^k\|, 1 \leq k \leq k_{\max}\}$;

求网络中各参数: $w_j = w_j^{k'}$, $b_j = b_j^{k'}$, $\beta_j = \beta_j^{k'}$, $E_j = E_j^{k'}$;

根据式(8)计算各指标相邻两次迭代的 RMSE 差值。

End if

End while

2 算法性能评估及应用

2.1 UCI能效数据测试

为了验证算法的有效性,将提出的 O-I-RVFLNs 算法利用 UCI 标准数据集中的能效数据 (energy efficiency data set) 进行数据测试。UCI 能效数据集一共有 767 组数据,是一个 8 输入(即相对密度、表面积、墙体面积、顶部面积、总高度、方向、玻璃面积以及玻璃面积分布)、2 输出(即制热负荷和制冷负荷)的多输入多输出样本。数据测试时,采用留出法将 500 组数据作为训练集,剩下的 267 组数据作为测试集。在建立模型时,将输入、输出数据都归一化到 $[-1, 1]$, 输入权重和隐层偏置取值范围为 $[-1, 1]$, 最大隐层节点数设为 1000, 期望误差为 $[1, 1]$, 相邻两次迭代的均方根误差差值为 $[0.001, 0.001]$ 。为防止各 RVFLNs 算法中输入权值和隐层偏置的随机性对结果产生影响,以下对比实验均采用 30 次实验的平均值作为最终的结果。

图 2 为采用所提 O-I-RVFLNs 算法的训练集制冷负荷和制热负荷的均方根误差(RMSE)收敛曲线。可以看出,随着隐层节点数(迭代次数)的增加,制冷负荷与制热负荷的 RMSE 均不断减小并逐渐收敛到定值。图 3 为制冷负荷和制热负荷相邻两次迭代的

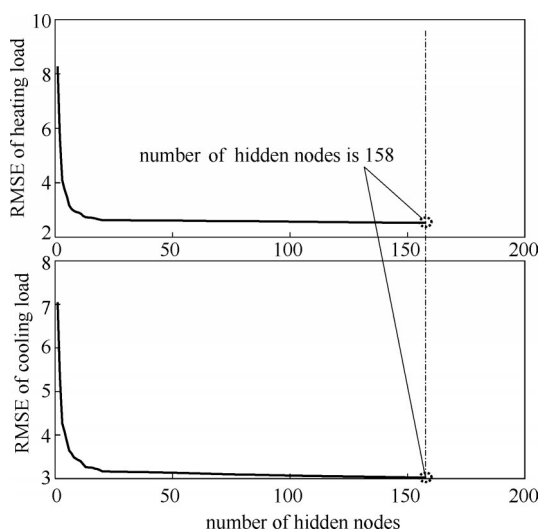


图 2 各指标 RMSE 与隐层节点数的关系

Fig.2 Relationship between RMSE of each index and number of hidden nodes

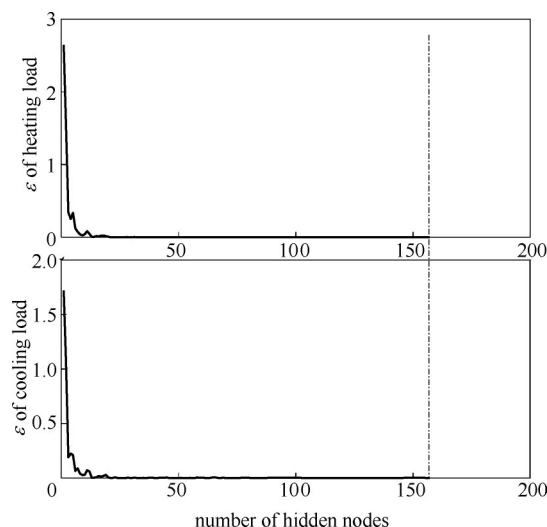


图 3 各指标 ε 与隐层节点数的关系

Fig.3 Relationship between ε of each index and number of hidden nodes

表 1 相同隐层节点下采用不同算法的建模结果对比

Table 1 Comparison of modeling results using different algorithms under the same hidden layer node

算法	隐层节点数目	RMSE	
		Heating load	Cooling load
RVFLNs	40	3.8066	4.5477
RVFLNs	50	4.5690	5.8099
RVFLNs	60	4.8659	6.3166
I-RVFLNs	40	3.6951	3.8338
I-RVFLNs	50	3.5300	3.5697
I-RVFLNs	60	3.3696	3.4109
O-I-RVFLNs	40	3.0929	3.3198
O-I-RVFLNs	50	2.9963	3.1709
O-I-RVFLNs	60	2.8696	3.0851

RMSE 差值与隐层节点数(迭代次数)的关系。由图可以看出,随着隐层节点数(迭代次数)的增加,各参数相邻两次迭代的 RMSE 差值均逐渐下降,当隐层节点数大于 20 时,下降趋势越来越不明显,且当隐层节点数为 158 时迭代停止。

图 4 为相同条件下传统 I-RVFLNs 算法与所提 O-I-RVFLNs 算法的建模误差变化对比曲线。可以看出,相对于传统 I-RVFLNs 算法,所提 O-I-RVFLNs 算法能够获得更快的收敛速度。进一步由表 1 发现,在相同隐层节点数的条件下,所提算法得到的模型预测精度明显高于基本 RVFLNs 和传统 I-RVFLNs 算法,而且随着隐层节点数的增加,所提算法不会出现类似 RVFLNs 的过拟合现象。

为了说明所提算法的优越性,将其与 RVFLNs、

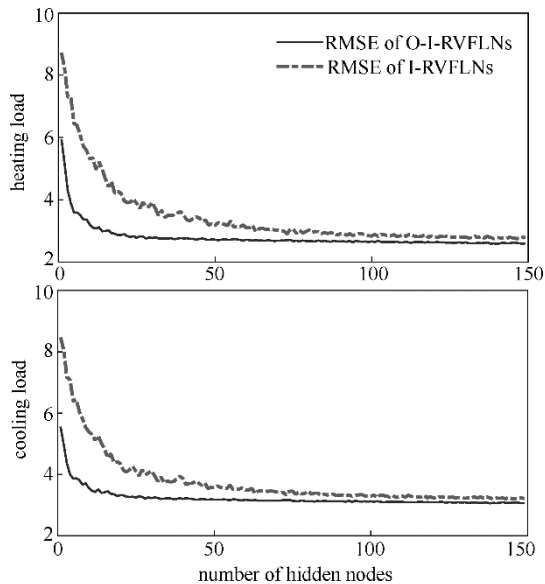


图4 不同方法RMSE随隐层节点数增加的变化曲线

Fig.4 RMSE variation curves of different methods as number of hidden layer nodes increases

I-RVFLNs 和 EI-RVFLNs 在相同数据集上进行对比实验,如图5所示。通过对比发现,相对于其他三种已有 RVFLNs 算法,采用所提 O-I-RVFLNs 方法建立的模型能够获得更好的建模效果,其模型估计输

出与实际输出最为接近。表2对比了不同增量型 RVFLNs 的建模结果,可以看出,在模型到达终止条件时,所提 O-I-RVFLNs 的隐层节点数最少,模型结构最简单,因此其运算效率高于传统 I-RVFLNs 和 EI-RVFLNs 算法。

2.2 应用于高炉多元铁水质量预测

2.2.1 高炉炼铁过程描述及铁水质量模型建立 高炉炼铁是钢铁生产中最重要的一环,其作用就是用焦炭等还原剂在高温下将含铁原料还原成液态生铁的过程。如图6所示,高炉炼铁是借助高炉本体及其附属系统来完成的,而高炉本体是冶炼生铁的主体设备,它自下而上由炉喉、炉身、炉腰、炉腹、炉缸五部分组成。高炉上部不断装入原矿石、焦炭、石灰石等炉料和导出炉顶煤气,下部不断鼓入煤粉和高温热风。全过程是在炉料自上而下、煤气自下而上的相互紧密接触过程中完成的。低温的矿石在下降过程中被煤气由外向内逐渐夺取氧而还原,同时又自高温煤气得到热量。矿石升到一定的温度时先软化,后熔融滴落,实现渣铁分离^[20-25]。生产出的铁水对后续转炉炼钢有着很大的影响,因此有必要获得实时准确的炉况信息,来保

表2 不同增量型随机神经网络算法的建模结果对比

Table 2 Comparison of modeling results of different incremental random vector functional-link networks algorithms

算法	隐层节点数	迭代次数	训练时间	测试时间	RMSE			
					Training		Testing	
					Heating load	Cooling load	Heating load	Cooling load
I-RVFLNs	1000	1000	5.7600	0.9840	2.6126	3.0687	3.4773	3.7051
EI-RVFLNs	842	842	8.8090	0.8340	2.5144	3.0250	3.1070	3.4349
O-I-RVFLNs	158	158	3.2600	0.1540	2.3604	2.9174	2.6709	2.9803

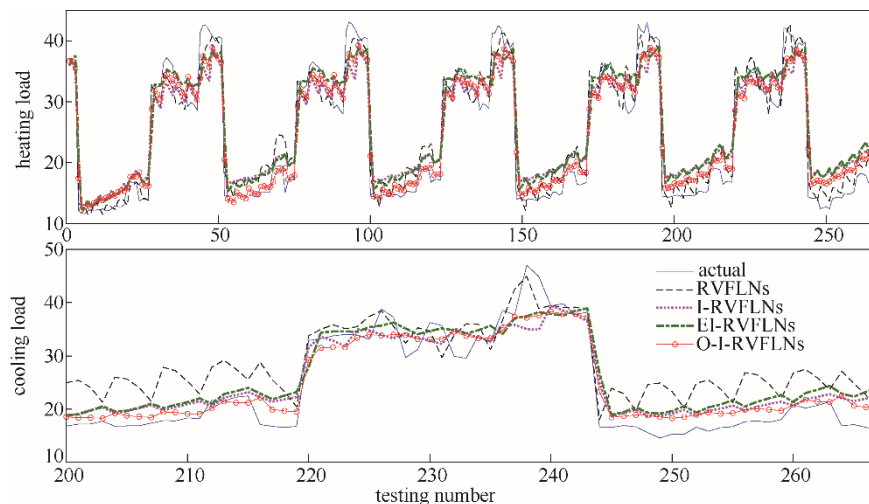


图5 基于不同算法的能效预测建模效果比较

Fig.5 Comparison of modeling effects of energy efficiency prediction based on different algorithms

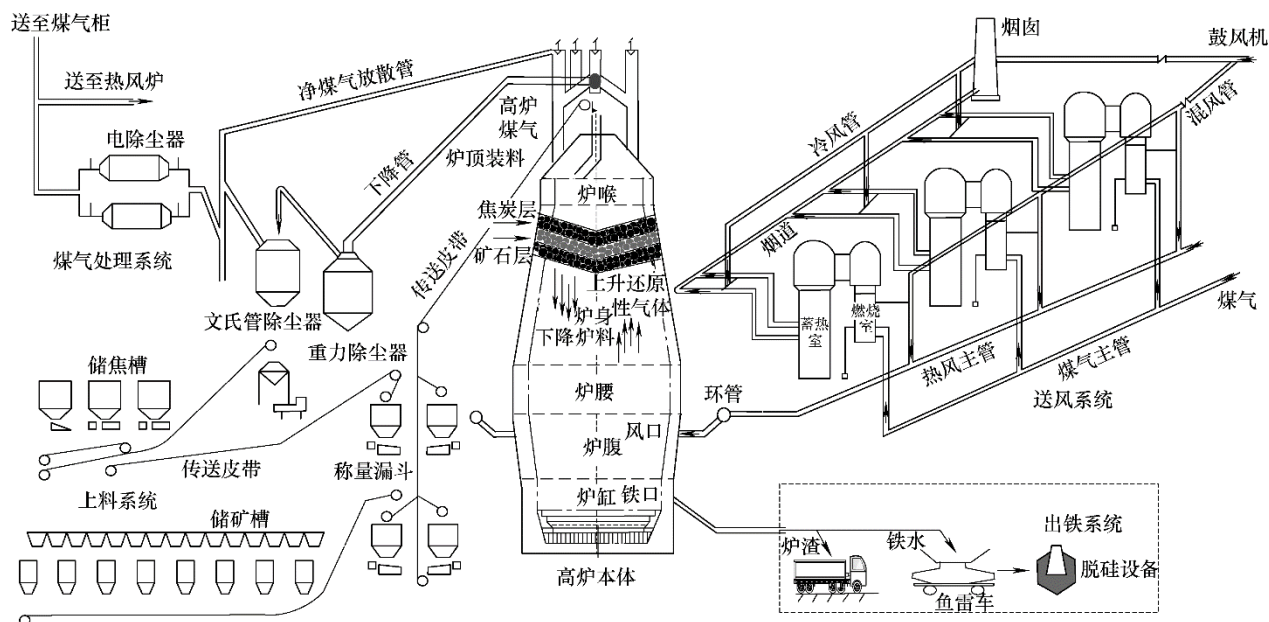


图6 高炉炼铁工艺示意图

Fig.6 Schematic diagram of blast furnace ironmaking process

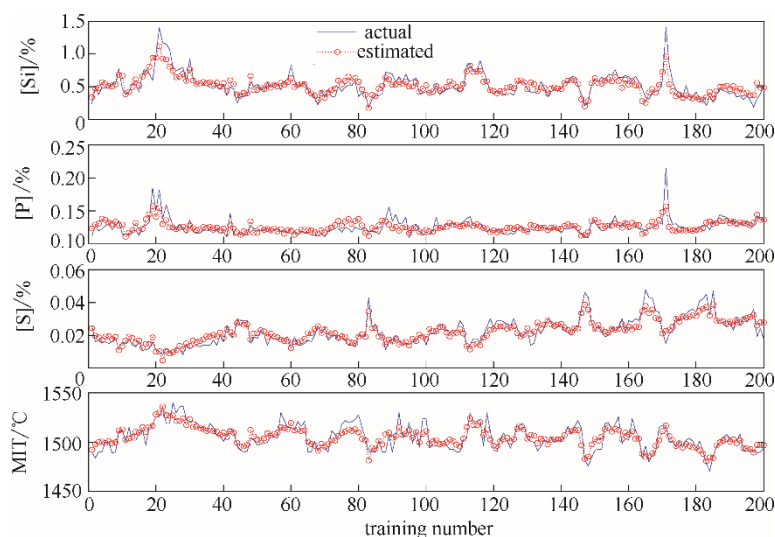


图7 所提O-I-RVFLNs算法的铁水质量建模结果

Fig.7 Modeling results of molten iron quality based on proposed O-I-RVFLNs algorithm

障高炉的平稳运行。然而高炉炼铁是一个发生在密闭容器中包含气、固、液三相混合,高度复杂的物理和化学过程,且内部环境十分恶劣,导致操作人员难以对内部运行状态的变化进行实时监测。因此有必要建立准确可靠的铁水质量数据模型来反映高炉当前和预期的内部温度和指标参数变化,为现场操作人员提供实时铁水质量信息^[20-22]。

为了实现多元铁水质量指标的在线估计,以我国华南某容积为 2650 m³ 的大型高炉为应用研究背景,采用提出的 O-I-RVFLNs 算法建立多元铁水质量指标的预测模型。根据该高炉炼铁工艺及相关

仪器仪表设置,确定影响铁水质量指标绝大部分信息的关键过程变量为:炉腹煤气量 x_1 (kg/t)、热风温度 x_2 (°C)、热风压力 x_3 (kPa)、富氧率 x_4 (%)、鼓风湿度 x_5 (g/m³) 和喷煤量 x_6 (m³/min)^[26]。

高炉炼铁系统是一个非线性、强耦合、大时滞的复杂动态系统,为了更好地反映高炉非线性动态特性和输入输出变量的时序及时滞关系,将当前时刻的输入、上一时刻的输入和上一时刻的模型(系统)输出作为数据建模的综合输入,而 [Si]、[P]、[S] 和 MIT 作为数据建模的输出变量,从而建立基于所提 O-I-RVFLNs 的铁水质量非线性自回归(NARX)动

态模型。共采用了270组实际高炉工业数据,采用留出法将其中200组用于训练,剩余70组用于测试。建模过程中,所有数据都归一化到 $[-1,1]$,输入权重和隐层偏置的取值范围为 $[-1,1]$,最大隐层节点数设为1000,期望误差为 $[0.1, 0.005, 0.005, 9]$,相邻两次迭代的RMSE差值为 $[1 \times 10^{-4}, 5 \times 10^{-6}, 5 \times 10^{-6}, 9 \times 10^{-3}]$ 。

2.2.2 多元铁水质量的建模和预测效果 图7为所提方法对多元铁水质量(即[Si],[P],[S]和MIT)在训练集上的建模效果。通过与实际曲线进行对比可知,各个铁水质量指标的建模输出值与其实际值基本吻合,变化趋势也比较接近。图8为采用RVFLNs、I-RVFLNs、EI-RVFLNs算法和所提O-I-RVFLNs算法在相同测试数据集上进行的预测对比实验,各算法均采用Sigmoid函数作为激励函数。通过比较可知,基于所提O-I-RVFLNs建立的模型对各铁水质量指标的预测结果最准确。而且相对于其他三种建模算法,采用所提方法建立的模型的预测输出曲线形状与实际曲线匹配最好,即所提算法的泛化性能最高。

图9为采用不同增量型随机神经网络算法所获得的各铁水质量指标RMSE与隐层节点数(迭代

次数)的关系变化曲线。可以看出,随着隐层节点数的增加,不同算法中各铁水质量指标的RMSE均逐渐减小,说明各算法的建模误差均是收敛的。但是通过进一步比较,可知所提O-I-RVFLNs算法的收敛速度明显快于传统I-RVFLNs算法,而且在获得相同建模RMSE的情况下,所提O-I-RVFLNs算法所需的隐层节点数最少,因而采用所提算法建立的模型能够获得最紧凑的网络结构。图10为各铁水质量指标相邻两次迭代的均方根误差差值与隐层节点数(迭代次数)的关系。可以看出,随着隐层节点数(迭代次数)的增加,各铁水质量指标相邻两次迭代的均方根误差差值均呈逐渐减小的趋势并最终趋于设定值,而且当隐层节点数为173时迭代停止。

随后,研究所提O-I-RVFLNs算法在训练过程隐层节点数与训练集、测试集误差的关系,如图11所示。可以看出,随隐层节点数的增加,训练集和测试集中各铁水质量指标的RMSE均逐渐减少,即没有“过拟合”现象。作为对比,图12所示为基本RVFLNs模型在训练过程中隐层节点数与训练集、测试集的RMSE关系曲线。可以看出,随着隐层节点数的增加,训练集各铁水质量指标的RMSE整体

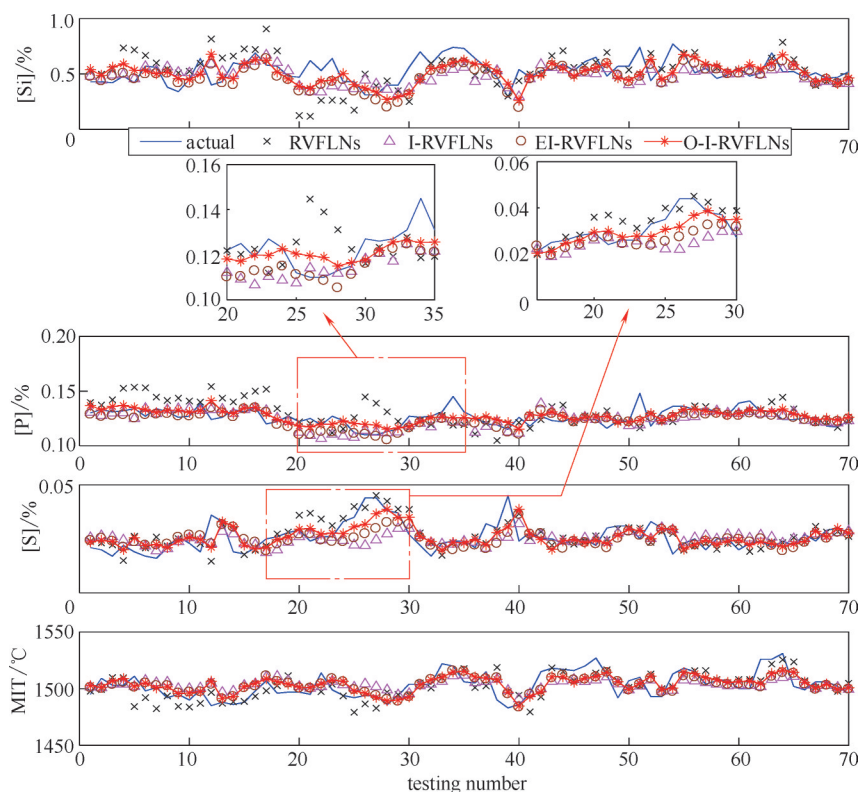


图8 基于不同算法的铁水质量在线预测结果

Fig.8 On-line molten iron quality prediction based on different algorithms

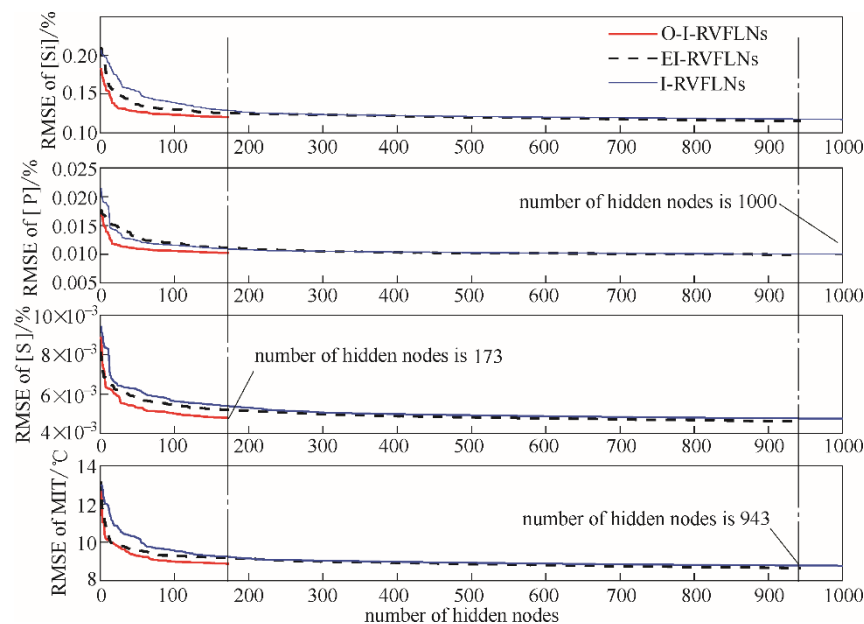
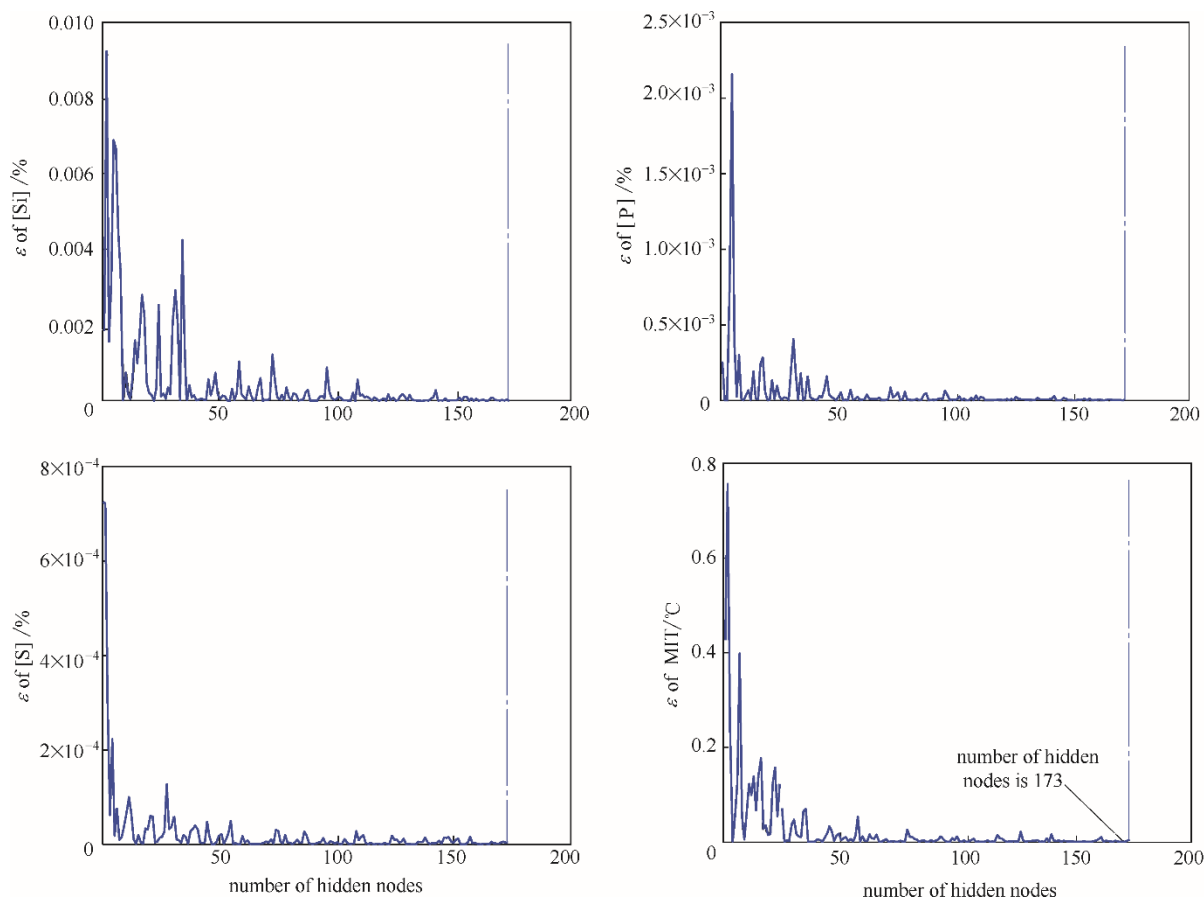


图9 各铁水质量指标的RMSE与隐层节点数的关系

Fig.9 Relationship between RMSE of each molten iron quality index and number of hidden layer nodes

图10 各铁水质量指标的 ε 与隐层节点数的关系Fig.10 Relationship between ε of each molten iron quality index and number of hidden layer nodes

呈现逐渐下降的趋势,而测试集各铁水质量指标的RMSE整体呈明显上升趋势。当隐层节点数大于30时,基本RVFLNs开始出现明显的“过拟合”现象。

表3为采用不同增量型随机神经网络算法在相同数据集上做的对比实验,由于各增量型RVFLNs隐层节点的输入权值和偏置是在给定范围

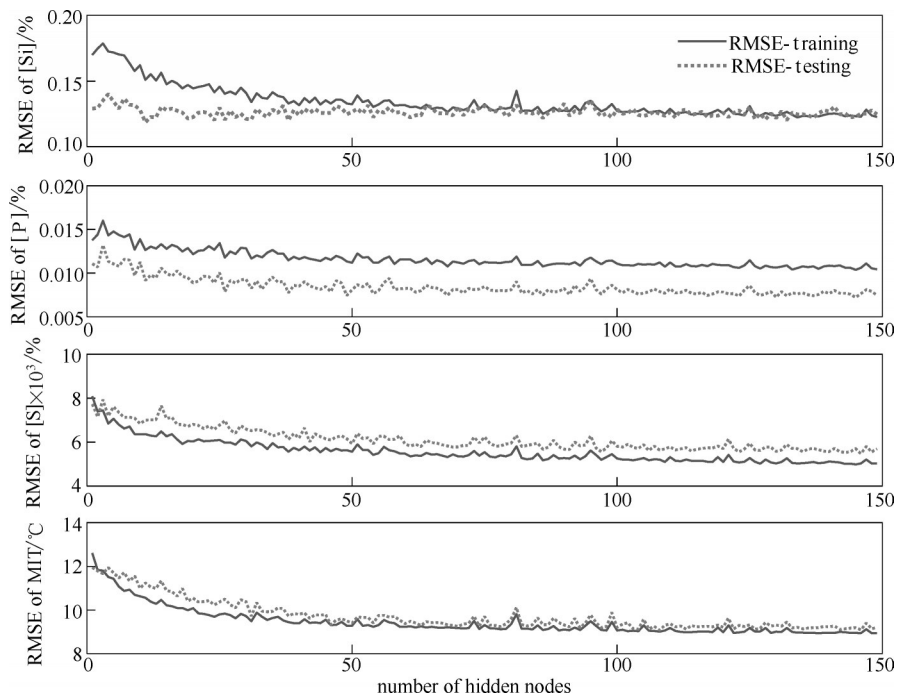


图 11 逐一增加隐层节点数时训练集和测试集的 RMSE 变化(O-I-RVFLNs)

Fig.11 RMSE changes of training set and testing set as number of hidden layer nodes increases(O-I-RVFLNs)

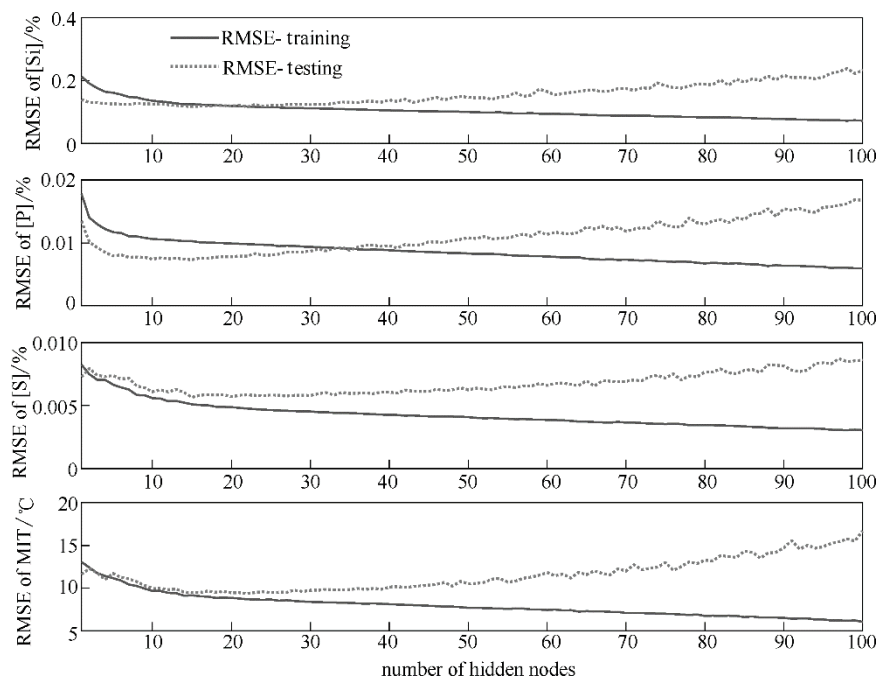


图 12 逐一增加隐层节点数时训练集和测试集的 RMSE 变化(RVFLNs)

Fig.12 RMSE changes of training set and testing set as number of hidden layer nodes increases(RVFLNs)

内随机选取的,为了克服随机性对结果造成的不稳定,对比实验取 30 次实验的平均值作为最终结果。通过对比发现,传统 I-RVFLNs 算法在到达终止条件时,网络的隐层节点数已经达到了算法的最大允许值 1000。因此,传统 I-RVFLNs 建立的模型结构过于复杂,测试时间较长。EI-RVFLNs 算法在到达

终止条件时,网络中的隐层节点数与 I-RVFLNs 相差不大,由于该算法在训练过程中对每个隐层节点的输入参数进行了筛选,因此模型预测精度要高于 I-RVFLNs,但模型训练时间较长。相比之下,所提 O-I-RVFLNs 算法在满足停止条件时,网络中的隐层节点数仅为 173,因此采用所提方法建立的模型

表 3 不同增量型随机神经网络算法的铁水质量建模结果对比

Table 3 Comparison of modeling results of molten iron quality based on different incremental random vector functional-link networks algorithms

算法	隐层节点数目	训练时间	测试时间	RMSE			
				[Si]	[P]	[S]	MIT
I-RVFLNs	1000	2.6090	0.2630	0.1177	0.0083	0.0057	9.1580
EI-RVFLNs	943	11.1340	0.2410	0.1172	0.0080	0.0055	9.1565
O-I-RVFLNs	173	2.1330	0.0490	0.1156	0.0078	0.0054	8.9221

结构最简单,测试时间最短,而且模型预测的精度

也较高。

3 结 论

本文针对传统 I-RVFLNs 存在的网络参数难以优化确定、模型收敛速度慢和模型结构复杂的问题,提出一种优化增量型随机神经网络(O-I-RVFLNs)算法。与传统 I-RVFLNs 采用随机生成隐层节点参数不同的是,所提 O-I-RVFLNs 算法通过提前设定一个期望的残差向量,然后在每次新增隐层节点时,选择可以达到或小于此节点期望残差的输入权值和偏置作为该节点的输入参数。此外,算法在迭代计算的过程中,将各指标参数相邻两次迭代的均方根误差差值考虑在终止条件内。最后,基于标准 UCI 数据集中的能效数据和我国华南某大型高炉的实际数据进行数据验证研究。数据实验结果表明:相对于 RVFLNs、I-RVFLNs 和 EI-RVFLNs,所提 O-I-RVFLNs 算法具有更好的泛化性能、更紧凑的模型结构以及更快的收敛速度,因此能够较较好地实际工业应用。

参考文献

- [1] Igel'nik B, Pao Y H. Stochastic choice of basis functions in adaptive function approximation and the functional-link net[J]. IEEE Trans. Neural Networks, 1995, 6(6): 1320-1329.
- [2] Pao Y H, Takefuji Y. Functional-link net computing: theory, system, architecture, and functionalities[J]. Computer, 1992, 25(2): 76-79.
- [3] Suganthan P N. On non-iterative learning algorithms with closed-form solution[J]. Applied Soft Computing, 2018, 70: 1078-1082.
- [4] Ren Y, Suganthan P N, Srikanth N, et al. Random vector functional link network for short-term electricity load demand forecasting[J]. Information Sciences, 2016, 367: 1078-1093.
- [5] Zhou P, Lv Y B, Wang H, et al. Data-driven robust RVFLNs modeling of blast furnace ironmaking process using Cauchy distribution weighted M-estimation[J]. IEEE Trans. Industrial Electronics, 2017, 64(9): 7141-7151.
- [6] Dai W, Chen Q X, Chu F. Robust regularized random vector functional link network and its industrial application [J]. IEEE Access, 2017, 5: 16162-16172.

- [7] Pratama M, Angelov P P, Lughofer E, et al. Parsimonious random vector functional link network for data streams [J]. Information Science, 2018, 430/431: 519-537.
- [8] Zhang L, Suganthan P N. Visual tracking with convolutional random vector functional link network [J]. IEEE Trans. Cybernetics, 2017, 47(10): 3243-3253.
- [9] Dai W, Liu Q, Chai T Y. Particle size estimate of grinding processes using random vector functional link networks with improved robustness[J]. Neurocomputing, 2015, 169: 361-372.
- [10] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: theory and applications[J]. Neurocomputing, 2006, 70(1): 489-501.
- [11] Huang G B, Zhou H, Ding X, et al. Extreme learning machine for regression and multiclass classification[J]. IEEE Trans. Syst., Man, Cybern. Syst., 2012, 42(2): 513-529.
- [12] Oneto L, Fumeo E, Clerico G. Dynamic delay predictions for large-scale railway networks: deep and shallow extreme learning machines tuned *via* thresholdout[J]. IEEE Trans. Syst., Man, Cybern. Syst., 2017, 47(10): 2754-2767.
- [13] 贺敏, 汤健, 郭旭琦, 等. 基于流形正则化域适应随机神经网络的湿式球磨机负荷参数软测量[J]. 自动化学报, 2019, 45(2): 398-406.
- [14] He M, Tang J, Guo X Q, et al. Soft sensor for ball mill load using DAMRRWNN model[J]. Acta Automatica Sinica, 2019, 45(2): 398-406.
- [15] Huang G B, Lei C, Siew C K. Universal approximation using incremental constructive feedforward networks with random hidden node[J]. IEEE Trans. Neural Networks, 2006, 17(4): 879-892.
- [16] Zhang L, Zhou P, Song H D, et al. Multivariable dynamic modeling for molten iron quality using incremental random vector functional-link networks[J]. Journal of Iron and Steel Research International, 2016, 23(11): 1151-1159.
- [17] Ying L. Orthogonal incremental extreme learning machine for regression and multiclass classification[J]. Neural Computing and Applications, 2016, 27(1): 111-120.
- [18] Huang G, Chen L. Enhanced random search based incremental extreme learning machine[J]. Neurocomputing, 2008, 71(16): 3460-3468.
- [19] Huang G, Chen L. Convex incremental extreme learning machine [J]. Neurocomputing, 2007, 70(16): 16-18.
- [20] Barron A R. Universal approximation bounds for superpositions of a sigmoidal function[J]. IEEE Trans. Information Theory, 1993, 39(3): 930-945.

- [20] Jian L, Gao C H, Xia Z H. Constructing multiple kernel learning framework for blast furnace automation[J]. IEEE Trans. Automation Science and Engineering, 2012, **9**(4): 763-777.
- [21] 宋贺达, 周平, 王宏, 等. 高炉炼铁过程多元铁水质量非线性子空间建模及应用[J]. 自动化学报, 2016, **42**(21): 1664-1679.
- Song H D, Zhou P, Wang H, et al. Nonlinear subspace modeling of multivariate molten iron quality in blast furnace ironmaking and its application[J]. Acta Automatica Sinica, 2016, **42**(21): 1664-1679.
- [22] Zhou P, Guo D, Wang H, et al. Data-driven robust M-LS-SVR-based NARX modeling for estimation and control of molten iron quality indices in blast furnace ironmaking[J]. IEEE Trans. Neural Networks & Learning Systems, 2018, **29**(9): 4007-4021.
- [23] 宋菁华, 杨春节, 周哲, 等. 改进型EMD-Elman神经网络在铁水硅含量预测中的应用[J]. 化工学报, 2016, **67**(3): 729-735.
- Song J H, Yang C J, Zhou Z, et al. Application of improved EMD-Elman neural network to predict silicon content in hot metal[J]. CIESC Journal, 2016, **67**(3): 729-735.
- [24] 李泽龙, 杨春节, 刘文辉, 等. 基于LSTM-RNN模型的铁水硅含量预测[J]. 化工学报, 2018, **69**(3): 992-997.
- Li Z L, Yang C J, Liu W H, et al. Research on hot metal Si-content prediction based on LSTM-RNN[J]. CIESC Journal, 2018, **69**(3): 992-997.
- [25] 蒋朝辉, 董梦林, 桂卫华, 等. 基于Bootstrap的高炉铁水硅含量二维预报[J]. 自动化学报, 2016, **42**(5): 715-723.
- Jiang Z H, Dong M L, Gui W H, et al. Two-dimensional prediction for silicon content of hot metal of blast furnace based on bootstrap[J]. Acta Automatica Sinica, 2016, **42**(5): 715-723.
- [26] Zhou P, Yuan M, Wang H, et al. Multivariable dynamic modeling for molten iron quality using online sequential random vector functional-link networks with self-feedback connections[J]. Information Sciences, 2015, **325**: 237-255.