

# 基于小波采样理论的新型准则函数



毛伟伟<sup>1</sup>, 张治国<sup>2\*</sup>, 金晓宇<sup>1</sup>

(1. 海军士官学校 5 系, 蚌埠 233012; 2. 电子科技大学 自动化工程学院, 成都 610054)

**摘要** 为解决在噪声环境下建模的过拟合问题, 基于小波采样理论, 提出一种适用于小波神经网络的新型准则函数, 并设计了相应的训练算法。这种算法能够利用样本分布和误差训练输入和输出层权值, 因此可以大大提高小波神经网络的学习效率。理论和试验表明, 新型准则函数有力地保证了小波神经网络的泛化能力, 其相应的算法具有全局收敛性, 并对噪声变化具有良好的鲁棒性。

**关键词** 广义采样; 神经网络; 过拟合; 小波采样

**中图分类号** TN911.97

**文献标志码** A

**DOI** 10.12178/1001-0548.2022128

## A Novel Cost Function Based on Wavelet Sampling Theory

MAO Weiwei<sup>1</sup>, ZHANG Zhiguo<sup>2\*</sup>, and JIN Xiaoyu<sup>1</sup>

(1. Fifth Department, Naval Petty Officer Academy, Bengbu 233012, China;

2. School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China)

**Abstract** In order to solve overfitting of modeling in noisy circumstance, a novel cost function with corresponding training algorithm is proposed for wavelet networks based on sampling theory. Since such an algorithm can use sample distributions and errors respectively to train input and output weights, learning efficiencies of wavelet networks are improved greatly. The theories and experiments show that this novel cost function can ensure generalizations of wavelet networks. Simultaneously, the new algorithm can converge globally and is robust to noise varying.

**Key words** generalized sampling; neural networks; overfitting; wavelet sampling

复杂环境下数据采集条件恶劣, 训练样本往往带有大量噪声, 且分布密度低。这导致神经网络建模极易发生过拟合。通常认为, 复杂模型和样本噪声是导致神经网络过拟合的主要原因, 因此在经典算法中, 限制模型复杂度及滤除噪声, 是防止过拟合的主要手段。

从限制模型复杂度的观点出发, 及早停止法是一种常用的防止过拟合方法。它常和交叉验证法结合, 广泛应用于河流和大气等复杂系统的建模<sup>[1-3]</sup>。然而该方法训练效率较低, 且很难适用于在线建模。模型选择法是另一种基于限制神经网络复杂程度来克服过拟合的有效方法。通过设定模型选择标准, 如 AIC (Akaike Information Criterion)<sup>[4-5]</sup> 和 BIC (Bayesian Information Criteria)<sup>[6-7]</sup> 标准, 这类方法可以有效地抑制训练误差的减小程度, 进而保

证神经网络的泛化能力。但研究表明, 只有当训练样本密度足够高时, 选择法才能同时保证模型的泛化能力和精度。这成为模型选择法在很多实际应用中效果不佳的主要原因。

从噪声引起过拟合的观点出发, 去除或限制样本噪声成为防止过拟合的另一种重要手段, 如主成分分析 (Principle Component Analysis)<sup>[8-9]</sup>、再生核算子<sup>[10-11]</sup> 等。然而这类方法在去除噪声时, 常容易滤除信号自身的细节, 而导致模型精度下降。

正则化技术是一种通过保证神经网络输出光滑性, 防止过拟合的方法。在该项技术中, 准则函数由正则项和误差项组成。由于正则项的存在, 神经网络训练可以缓和插值误差的减小程度, 提高神经网络的泛化能力<sup>[12-13]</sup>。正则化技术便于操作, 样本数据利用效率高、防止过拟合效果显著, 因此得到

收稿日期: 2022-05-15; 修回日期: 2023-11-21

作者简介: 毛伟伟, 主要从事信号处理、小波神经网络、伺服控制方面的研究。

\*通信作者 E-mail: zhiguo Zhang@uestc.edu.cn

广泛的应用<sup>[14]</sup>。但正则项的选取依赖于噪声特性, 而该特性往往会随着时间发生变化, 这大大降低了正则化技术对噪声的鲁棒性。

从上面的分析可知, 尽管经典防止过拟合技术可以提高神经网络建模的泛化能力, 但常存在计算复杂、样本密度要求高、数据利用效率低、对噪声变化缺乏鲁棒性等问题, 难以适用于复杂系统建模。为此本文将小波神经网络作为基础, 引入一种新的准则函数, 并提出一种输入和输出层权值相分离的训练算法。理论和实验表明该算法具有良好的泛化能力和全局收敛性。

### 1 小波神经网络结构

本算法采用跟文献 [15] 中相同的小波神经网络结构。为便于论述本文算法, 本节将简单描述该神经网络的特性。关于其更为详细内容, 可以参见文献 [15]。

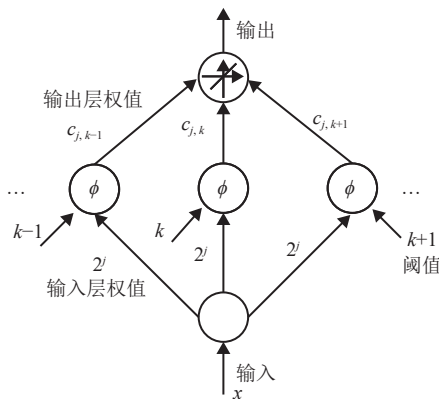


图 1 小波神经网络结构

本算法针对三层结构的小波神经网络, 其以小波父函数 (尺度函数)  $\phi(x)$  作为激励函数, 结构如图 1 所示, 其中  $\{c_k\}_{k \in \mathbb{Z}}$  表示输出层权值,  $2^j$  表示输入层权值,  $\{k\}_{k \in \mathbb{Z}}$  表示阈值, 这里  $j \in \mathbb{Z}$  (见文献 [15])。神经网络解析表达可以写为:

$$f_{ne}(x) = \sum_{k=N}^M c_k \phi(2^j x - k) = \sum_{k=N}^M c_k \phi_{j,k}(x) \quad (1)$$

将单维尺度函数替换成多维尺度函数, 式 (1) 所示单维小波网络可以拓展为多维小波网络<sup>[15-16]</sup>。如果令  $\{c_k\} = \{c_{k_1, \dots, k_d}\}_{k_1, \dots, k_d \in \mathbb{Z}}$ ,  $\mathbf{M} = [M_1, \dots, M_d]$ ,  $\mathbf{N} = [N_1, \dots, N_d]$ ,  $\mathbf{k} = [k_1, \dots, k_d]$ ,  $\mathbf{x} = [x_1, \dots, x_d]$ ,  $\mathbf{j} = [j_1, \dots, j_d]$ ,  $2^{\mathbf{j}} = [2^{j_1}, \dots, 2^{j_d}]$ , 那么多维小波网络可以表述为:

$$f_{ne}(\mathbf{x}) = \sum_{k=N}^M c_k \phi(2^{\mathbf{j}} \circ \mathbf{x} - \mathbf{k}) =$$

$$\sum_{k_d=N_d}^{M_d} \cdots \sum_{k_1=N_1}^{M_1} c_{k_1, \dots, k_d} \phi(2^{j_1} x_1 - k_1, \dots, 2^{j_d} x_d - k_d) =$$

$$\sum_{k_1=N_d}^{M_d} \cdots \sum_{k_d=N_1}^{M_1} c_{k_1, \dots, k_d} \phi(2^{j_1} x_1 - k_1) \times \cdots \times \phi(2^{j_d} x_d - k_d) \quad (2)$$

式中,  $2^{\mathbf{j}} \circ \mathbf{x} = [2^{j_1} x_1, \dots, 2^{j_d} x_d]$  表示向量  $2^{\mathbf{j}}$  和向量  $\mathbf{x}$  之间的点积;  $2^{j_d}$  表示神经网络在  $x_d$  方向上的输入层权值。此外  $\phi(2^{\mathbf{j}} \circ \mathbf{x} - \mathbf{k}) = \phi(2^{j_1} x_1 - k_1, \dots, 2^{j_d} x_d - k_d) = \phi(2^{j_1} x_1 - k_1) \times \cdots \times \phi(2^{j_d} x_d - k_d)$  表示由张量积生成的多维尺度函数<sup>[15-16]</sup>。

### 2 一维学习算法

#### 2.1 基于小波采样理论的准则函数

为简便起见, 首先讨论一维 (单输入) 情况下的算法, 然后再扩展到高维。由小波采样理论可知, 插值尺度函数  $\{S_{j,k}^\phi(x)\}_{k \in \mathbb{Z}}$  和单位正交尺度函数  $\{\theta_{j,k}(x)\}$  之间存在一一映射, 即:

$$S^\phi(2^j x) = \sum_{k=-\infty}^{+\infty} \lambda_k \theta(2^j x - k) \quad (3)$$

式中,  $\{\lambda_k\}_{k \in \mathbb{Z}} \in l^2(\mathbb{Z})$ ;  $\theta_{j,k}(x) = \theta(2^j x - k)$ 。

设  $f_s(x) \in V_j$  为建模对象函数,  $V_j$  表示  $\{\phi_{j,k}(x)\}_{k \in \mathbb{Z}}$  张成的尺度函数空间。令  $\mathbf{Q} = [\lambda_{k-n}]_{k \times n}$  表示在第  $k$  行、第  $n$  列、以  $\lambda_{k-n}$  为元素的矩阵, 那么基于式 (3), 构建准则函数:

$$F_C(f_s, f_{ne}) = (\mathbf{F}_s - \mathbf{F}_{ne})^T \mathbf{Q}^T \mathbf{Q} (\mathbf{F}_s - \mathbf{F}_{ne}) =$$

$$(\mathbf{F}_s - \mathbf{F}_{ne})^T \mathbf{A} (\mathbf{F}_s - \mathbf{F}_{ne}) \quad (4)$$

式中,  $\mathbf{F}_s$ ,  $\mathbf{F}_{ne}$  分别表示以  $\{f_s(k \cdot T)\}_{k \in \mathbb{Z}}$ ,  $\{f_{ne}(k \cdot T)\}_{k \in \mathbb{Z}}$  为元素的向量, 而矩阵  $\mathbf{A}$  表示为:

$$\mathbf{A} = \mathbf{Q}^T \mathbf{Q} \quad (5)$$

显然式 (3) 表明序列  $\{\lambda_k\}_{k \in \mathbb{Z}}$  可以构成 Riesz 基  $\{S_{j,k}^\phi(x)\}_{k \in \mathbb{Z}}$  和  $\{\theta_{j,k}(x)\}$  的可逆性。基于此, 这里引入以下定理。

**定理 1** 式 (5) 中  $\mathbf{A}$  为正定阵

证明: 令:

$$\mathbf{F}_{s,\lambda} = \mathbf{Q} \mathbf{F}_s \quad (6)$$

显然对于任何  $\mathbf{F}_s \in l^2(\mathbb{Z})$ , 总存在唯一  $\mathbf{F}_{s,\lambda} \in l^2(\mathbb{Z})$ 。下面说明任何  $\mathbf{F}_{s,\lambda} \in l^2(\mathbb{Z})$  对应的  $\mathbf{F}_s \in l^2(\mathbb{Z})$  也是唯一。假设  $f_{s,\lambda}(n \cdot T)$  表示  $\mathbf{F}_{s,\lambda}$  中第  $n$  个元素 (这里  $T$  表示对  $f_s(x)$  的采样间隔), 那么由式 (3) 和式 (6)

可得:

$$\sum_{n=-\infty}^{+\infty} f_{s,\lambda}(n \cdot T) \theta(x-n) = \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} \lambda_{n-m} f_s(m \cdot T) \theta(x-n) = \sum_{m=-\infty}^{+\infty} f_s(m \cdot T) S^\phi(x-m) \quad (7)$$

由于 $\{\theta(x-n)\}_{n \in \mathbb{Z}}$ 和 $\{S^\phi(x-m)\}_{m \in \mathbb{Z}}$ 都构成 $V_0$ 的 Riesz 基 (这里 $V_0$ 是 $\{\phi(x-k)\}_{k \in \mathbb{Z}}$ 张成的空间), 因此对于任何 $\{f_s(m \cdot T)\}_{m \in \mathbb{Z}} \in l^2(\mathbb{Z})$ , 使得式 (8) 成立的 $\{f_{s,\lambda}(n \cdot T)\}_{n \in \mathbb{Z}} \in l^2(\mathbb{Z})$ 都是唯一的。这反过来意味着对于任何 $\{f_{s,\lambda}(n \cdot T)\}_{n \in \mathbb{Z}} \in l^2(\mathbb{Z})$ , 只存在唯一的 $\{f_s(m \cdot T)\}_{m \in \mathbb{Z}} \in l^2(\mathbb{Z})$ 使得式 (7) 成立, 因而从 $\{f_s(m \cdot T)\}_{m \in \mathbb{Z}}$ 到 $\{f_{s,\lambda}(n \cdot T)\}_{n \in \mathbb{Z}}$ 的映射, 具有可逆性。

因为式 (6) 是可逆映射, 所以矩阵 $Q$ 必为可逆矩阵。式 (6) 意味着:

$$F_{s,\lambda} = \mathbf{0} \text{ 当且仅当 } F_s = \mathbf{0} \quad (8)$$

由式 (6) 和 (8) 可知, 当 $F_s \neq \mathbf{0}$ 时:

$$F_s^T \Lambda F_s = F_s^T Q^T Q F_s = F_{s,\lambda}^T F_{s,\lambda} > 0 \quad (9)$$

即 $\Lambda$ 为正定阵。证毕。

显然对于 $F_s$ ,  $F_{ne} \in l^2(\mathbb{R})$ , 当 $\Lambda$ 为正定阵时, 式 (4) 可以有效衡量 $f_s(x)$ 和 $f_{ne}(x)$ 之间的距离平方。为此这里提出用式 (4) 作为算法的准则函数。

## 2.2 由样本分布确定输入层权值

由小波理论及式 (3) 可知, 小波网络输入层的权值 $2^j$ 决定了逼近空间, 而输出层权值 $\{c_{j,k}\}_{k \in \mathbb{Z}}$ 对应函数 (被建模对象) 在逼近空间中的坐标。基于这种特点及式 (4) 所示准则函数, 这里提出一种新的神经网络训练算法。在该算法中, 将依据训练样本分布确定输入权值 $2^J$ , 而基于样本误差确定输出层权值。

考虑一维均匀分布的样本集合 $\{k \cdot T, f_s(k \cdot T)\}_{k \in \mathbb{Z}}$ , 这里 $T$ 表示输入值之间的间隔,  $k \cdot T$ 表示第 $k$ 个输入值,  $f_s(k \cdot T)$ 表示第 $k$ 个输出值。同时基于归一技术, 这里假设存在 $J \in \mathbb{Z}$ , 使得:

$$2^J = \frac{1}{T} \quad (10)$$

此时由小波采样理论可知, 如果 $f_s(x) \in V_J$  (这里 $V_J$ 表示 $\{\phi_{J,k}(x)\}_{k \in \mathbb{Z}}$ 张成的函数空间), 总有下式成立:

$$f_s(x) = \sum_{k=-\infty}^{+\infty} f_s(k \cdot T) S^\phi(2^J x - k) \quad (11)$$

这里 $f_s(x)$ 表示被建模对象的函数形式,  $S^\phi(x)$ 表示插值尺度函数。

因为 $\{S^\phi(2^J x - k)\}_{k \in \mathbb{Z}}$ 和 $\{\phi(2^J x - k)\}_{k \in \mathbb{Z}}$ 同时构成 Riesz 基, 因此式 (12) 意味着存在唯一的系数 $\{c_{J,k}^0\}_{k \in \mathbb{Z}}$ , 使得:

$$f_s(x) = \sum_{n=-\infty}^{+\infty} c_{J,n}^0 \phi(2^J x - n) \quad (12)$$

式 (12) 表明如果利用 $\phi(x)$ 作为激励函数, 那么输入层权值为 $2^j = 2^J$ 的小波网络 $f_{ne}(x)$ , 具有足够的逼近能力, 建立 $f_s(x) \in V_J$ 的数学模型, 这里 $J$ 由式 (10) 确定。为此, 本算法的输入层权值 $2^j$ 将由采样间隔 $T$ 确定, 即输入层权值 $2^j$ 满足条件:

$$2^j = 2^{J-q} \leq 2^J = 1/T \quad (13)$$

式中,  $q = J - j$ 。

## 2.3 基于样本误差训练输出层权值

当输入层权值 $2^j$ 确定后, 由式 (1) 可知, 在样本位置上的神经网络输出, 可以表示为:

$$F_{ne} = \Phi C \quad (14)$$

式中,  $\Phi = [\phi_{j,m}(n \cdot T)]_{n,m}$ ;  $T$ 表示样本间隔。基于式 (14), 神经网络训练可以表述为以下迭代过程:

$$\begin{cases} E^{(k+1)} = F_s - \Phi C^{(k+1)} \\ C^{(k+1)} = C^{(k)} + A_f E^{(k)} \end{cases} \quad (15)$$

式中, 向量 $E^{(k)}$ 和 $C^{(k)}$ 分别表示第 $k$ 步的插值误差和输出层权值; 矩阵 $A_f$ 表示误差反馈系数。事实上, 当式 (13) 成立时,  $\Phi^T \Lambda \Phi$ 为可逆矩阵<sup>[17]</sup>。因此这里可以得到以下收敛算法。

**定理 2** 设 $\lambda_A$ 为满足 $|1 - \lambda_A| < 1$ 的常数, 那么选取反馈矩阵为 $A_f = \lambda_A (\Phi^T \Lambda \Phi)^{-1} \Phi^T \Lambda$ 时, 输出层权值 $C$ 总可以通过式 (15), 收敛到:

$$C^0 = (\Phi^T \Lambda \Phi)^{-1} \Phi^T \Lambda F_s \quad (16)$$

并且在收敛点 $C^0$ 上, 准则函数式 (4) 减为最小。

证明: 将 $E^{(k+1)} = F_s - \Phi C^{(k+1)}$ 代入 $C^{(k+1)} = C^{(k)} + A_f E^{(k)}$ , 可得:

$$C^{(k+1)} = (I - A_f \Phi) C^{(k)} + A_f F_s \quad (17)$$

简化可得:

$$C^{(k+1)} = (I - \lambda_A) C^{(k)} + \lambda_A (\Phi^T \Lambda \Phi)^{-1} \Phi^T \Lambda F_s \quad (18)$$

设定 $B^{(k)} \in l^2(\mathbb{Z})$ 为另一迭代过程 (跟 $C^{(k)}$ 迭代过程的初始值不同), 在第 $k$ 步的输出层权值为:

$$B^{(k+1)} = (I - \lambda_A) B^{(k)} + \lambda_A (\Phi^T \Lambda \Phi)^{-1} \Phi^T \Lambda F_s \quad (19)$$

那么由式 (18) 和式 (19) 可得:

$$\|C^{(k+1)} - B^{(k+1)}\|_{\rho} = |1 - \lambda_A| \|C^{(k)} - B^{(k)}\|_{\rho} \quad (20)$$

式中,  $\|\cdot\|_{\rho}$  表示  $L^2(\mathbb{Z})$  空间的范数。式 (20) 意味着当  $|1 - \lambda_A| < 1$  时, 有:

$$\|C^{(k+1)} - B^{(k+1)}\|_{\rho} < \|C^{(k)} - B^{(k)}\|_{\rho} \quad (21)$$

这意味着式 (18) 为压缩映射。因此根据压缩映射定理, 式 (18) 将收敛于唯一的固定点  $C^0$ 。固定点  $C^0$  满足:

$$C^0 = (I - A_f \Phi) C^0 + A_f F_s \quad (22)$$

可解得式 (16)。

另一方面, 由式 (4) 可知:

$$\begin{cases} \frac{\partial F_C}{\partial C} = 4(\Phi^T \Lambda \Phi C - \Phi^T \Lambda F_s) \\ \frac{\partial^2 F_C}{\partial C^2} = 4\Phi^T \Lambda \Phi \end{cases} \quad (23)$$

将  $C^0 = (\Phi^T \Lambda \Phi)^{-1} \Phi^T \Lambda F_s$  代入式 (23), 可得:

$$\frac{\partial F_C}{\partial C} = 0 \quad (24)$$

因为  $\Lambda$  为正定阵, 因此矩阵  $\Phi^T \Lambda \Phi$  主对角线上的元素为正, 因此式 (24) 意味着  $C^0$  为准则函数  $F_C(f_s, f_{ne})$  的最小值点。证毕。

定理 2 意味着无论如何选择输出层权值的初始值, 只要按照式 (13) 确定输入层权值, 并依据式 (16) 确定反馈系数, 那么式 (15) 总可以收敛到固定点  $C^0$ , 在该点上插值误差减为最小。这表明本文提出的算法具有全局收敛性。

### 3 学习算法多维扩展

#### 3.1 多维算法的准则函数和输入层权值确定

因为多维尺度函数由单维尺度函数通过张量积生成, 所以单维小波神经网络的性质, 很容易拓展到多维的情况。设定  $F_s, F_{ne}$  分别表示多维样本以及多维小波网络在样本位置的输出:

$$F_s = [f_s(T_1, T_2, \dots, T_d), f_s(2T_1, T_2, \dots, T_d), \dots, f_s(nT_1, T_2, \dots, T_d), f_s(nT_1, 2T_2, \dots, T_d), \dots, f_s(nT_1, nT_2, \dots, T_d), \dots, f_s(nT_1, \dots, nT_{d-1}, 2T_d), \dots, f_s(nT_1, \dots, nT_{d-1}, nT_d)]^T \quad (25)$$

$$F_{ne} = [f_{ne}(T_1, T_2, \dots, T_d), f_{ne}(2T_1, T_2, \dots, T_d), \dots, f_{ne}(nT_1, T_2, \dots, T_d), f_{ne}(nT_1, 2T_2, \dots, T_d), \dots, f_{ne}(nT_1, nT_2, \dots, T_d), \dots, f_{ne}(nT_1, \dots, nT_{d-1}, 2T_d), \dots, f_{ne}(nT_1, \dots, nT_{d-1}, nT_d)]^T \quad (26)$$

由张量积性质可知, 式 (2) 所示多维小波神经网络的输出  $F_{ne}$  可以表示成:

$$F_{ne} = \Phi_1 \otimes \dots \otimes \Phi_i \otimes \dots \otimes \Phi_d C = \Phi C \quad (27)$$

式中,  $\Phi_i$  表示在  $x_i$  方向上的矩阵  $\Phi$ ;  $\otimes$  表示矩阵 Kronecker 积, 而:

$$\Phi = \Phi_1 \otimes \dots \otimes \Phi_i \otimes \dots \otimes \Phi_d \quad (28)$$

跟式 (3) 相似, 在多维情况下, 有:

$$S^\phi(2^j \circ x) = \sum_{k=-\infty}^{+\infty} \lambda_k \theta(2^j \circ x - k) \quad (29)$$

式中,  $\theta(2^j \circ x - k) = \theta(2^{j_1} x_1 - k_1) \times \dots \times \theta(2^{j_d} x_d - k_d)$ ;  $S^\phi(2^j \circ x) = S^\phi(2^{j_1} x_1) \times \dots \times S^\phi(2^{j_d} x_d)$ 。由  $L^2(\mathbb{R}^d)$  空间内积算子性质可得式 (29) 中  $\lambda_k = \lambda_{k_1} \times \dots \times \lambda_{k_d}$ , 这里  $\{\lambda_{k_i}\}_{k_i \in \mathbb{Z}}$  表示在  $x_i$  方向上的系数  $\{\lambda_k\}_{k \in \mathbb{Z}}$ , 即当  $k_i = k$  时,  $\lambda_{k_i} = \lambda_k$ 。

令  $Q_i = [\lambda_{n_i - k_i}]_{n_i, k_i}$  表示在  $(k_i, n_i)$  位置上的矩阵, 这里  $k_i, n_i \in \mathbb{Z}$ 。显然  $Q_i$  对应于在  $x_i$  方向上的  $Q$  矩阵。基于  $Q_i$  构建矩阵:

$$Q = Q_1 \otimes \dots \otimes Q_i \otimes \dots \otimes Q_d \quad (30)$$

那么由矩阵 Kronecker 积的性质可得:

$$\Lambda = Q^T Q = \Lambda_1 \otimes \dots \otimes \Lambda_i \otimes \dots \otimes \Lambda_d \quad (31)$$

$$\Lambda_i = Q_i^T Q_i \quad i \in \mathbb{Z} \quad (32)$$

令  $2^j = [2^{j_1}, \dots, 2^{j_i}, \dots, 2^{j_d}]$  表示多维情况下的输入层权值, 其中  $2^{j_i}$  表示在  $x_i$  方向上输入层权值的分量。由定理 1 可知, 当  $2^{j_i}$  满足式 (13) 时, 即:

$$2^{j_i} \leq 1/T_i \quad (33)$$

式 (31) 中  $\Lambda_i$  为正定阵。如果对应于每个  $x_i$  方向上的  $\Lambda_i$  都为正定阵, 那么根据 Kronecker 积的性质, 意味着  $\Lambda$  也为正定阵。

与式 (4) 相似, 这里设定多维准则函数为:

$$F_C(f_s, f_{ne}) = (F_s - F_{ne})_s^T \Lambda (F_s - F_{ne}_s) = \|F_s - F_{ne}\|_{\Lambda}^2 \quad (34)$$

#### 3.2 多维输出层权值算法

与式 (16) 相类似, 在多维情况下, 当输入层权值确定后, 输出层权值训练过程可以表示为:

$$\begin{cases} E^{(k+1)} = F_s - \Phi C^{(k+1)} \\ C^{(k+1)} = C^{(k)} + A_f E^{(k)} \end{cases} \quad (35)$$

式中,  $E^{(k)}$ ,  $C^{(k)}$  和  $A_f$  分别表示神经网络在第  $k$  步的插值误差、输出层权值和反馈系数矩阵。

对比式 (15) 和 (35), 以及式 (4) 和式 (34), 可

以发现, 单维神经网络和多维神经网络的迭代过程是相似的。这意味只要  $\Phi^T \Lambda \Phi$  具有可逆性, 那么在定理 2 的证明中, 只要简单地将矩阵  $\Phi$ 、 $F_s$ 、 $A_f$  替换成  $\Phi$ 、 $F_s$ 、 $A_f$ , 那么单维迭代收敛算法同样适用于多维情况。

事实上, 根据张量积的性质, 可以获得以下推论。

**推论 1** 矩阵  $\Phi^T \Lambda \Phi$  可逆的充要条件是输入层权值满足式 (33)。

证明: 由式 (27) 和式 (30), 及矩阵性质可得:

$$\begin{aligned} Q\Phi &= (Q_1 \otimes \cdots \otimes Q_d)(\Phi_1 \otimes \cdots \otimes \Phi_d) = \\ &Q_1 \Phi_1 \otimes \cdots \otimes Q_d \Phi_d \end{aligned} \quad (36)$$

将式 (31) 和式 (36) 代入  $\Phi^T \Lambda \Phi$ , 可得:

$$\begin{aligned} \Phi^T \Lambda \Phi &= \Phi^T Q^T Q \Phi = (Q\Phi)^T Q\Phi = \\ &(\Phi_1^T Q_1^T \otimes \cdots \otimes \Phi_d^T Q_d^T)(Q_1 \Phi_1 \otimes \cdots \otimes Q_d \Phi_d) = \\ &\Phi_1^T \Lambda_1 \Phi_1 \otimes \cdots \otimes \Phi_d^T \Lambda_d \Phi_d \end{aligned} \quad (37)$$

式 (37) 和矩阵 Kronecker 积的性质意味着  $\Phi^T \Lambda \Phi$  可逆, 当且仅当  $\Phi_i^T \Lambda_i \Phi_i$  可逆。由文献 [16] 可知, 当  $\Lambda_i$  为正定阵时,  $\Phi_i^T \Lambda_i \Phi_i$  可逆的充要条件为式 (33) 成立, 因此推论 1 成立。证毕。

推论 1 意味着如果按式 (33) 选择输入层权值, 那么  $\Phi^T \Lambda \Phi$  为可逆矩阵。跟单维训练算法相似, 如果构建矩阵  $A_f = \lambda_A (\Phi^T \Lambda \Phi)^{-1} \Phi^T \Lambda$  那么式 (35) 收敛, 且在收敛值  $C^0 = (\Phi^T \Lambda \Phi)^{-1} \Phi^T \Lambda F_s$  上, 使得准则函数式 (34) 减为最小。

## 4 算法泛化能力

根据小波理论, 当  $j \rightarrow +\infty$  时, 式 (2) 中的神经网络  $f_{ne}(\mathbf{x})$  可以逼近  $L^2(\mathbb{R}^d)$  空间中的任何元素<sup>[15]</sup>。由于  $\{\phi(2^j \circ \mathbf{x} - \mathbf{k})\}_{\mathbf{k} \in \mathbb{Z}^d}$  可以构成  $V_j$  空间中的 Riesz 基, 因此这里总假设对象函数  $f_s(\mathbf{x}) \in V_j$ , 这里  $V_j$  表示由  $\{\phi(2^j \circ \mathbf{x} - \mathbf{k})\}_{\mathbf{k} \in \mathbb{Z}^d}$  张成的尺度函数空间。

由于  $V_j \subset L^2(\mathbb{R}^d)$ , 因此这里采用  $L^2$  范数来表示逼近误差:

$$e(f_s, f_{ne}) = \|f_s - f_{ne}\|_{L^2}^2 = \int_{-\infty}^{+\infty} |f_s(\mathbf{x}) - f_{ne}(\mathbf{x})|^2 d\mathbf{x} \quad (38)$$

式中,  $\int_{-\infty}^{+\infty} |f(\mathbf{x})|^2 d\mathbf{x} = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} |f(x_1, \cdots, x_d)|^2 dx_1 \cdots dx_d$ ;  $\|\cdot\|_{L^2}$  表示  $L^2$  范数范数。基于  $e(f_s, f_{ne})$  及准则式 (34), 有以下定理。

**定理 3** 当式 (33) 成立时, 逼近误差式 (38) 可以表示为:

$$e(f_s, f_{ne}) = \int_{-\infty}^{+\infty} |f_{ne}(\mathbf{x}) - f_s(\mathbf{x})|^2 d\mathbf{x} = F_C(f_s, f_{ne}) \quad (39)$$

证明: 当式 (33) 成立时, 由插值基的性质可得:

$$\begin{aligned} f_s(\mathbf{x}) - f_{ne}(\mathbf{x}) &= \sum_{\mathbf{k}} f_s(\mathbf{k} \circ \mathbf{T}) S^\phi(2^J \circ \mathbf{x} - \mathbf{k}) - \\ &\sum_{\mathbf{k}} f_{ne}(\mathbf{k} \circ \mathbf{T}) S^\phi(2^J \circ \mathbf{x} - \mathbf{k}) \end{aligned} \quad (40)$$

将式 (29) 代入式 (40) 可得:

$$\begin{aligned} f_s(\mathbf{x}) - f_{ne}(\mathbf{x}) &= \left( \sum_{\mathbf{k}} f_s(\mathbf{k} \circ \mathbf{T}) - \right. \\ &\sum_{\mathbf{k}} f_{ne}(\mathbf{k} \circ \mathbf{T}) \left. \right) \sum_{\mathbf{m}} \lambda_{\mathbf{m}} \theta(2^J \circ \mathbf{x} - \mathbf{m} - \mathbf{k}) = \\ &\sum_{\mathbf{n}} \sum_{\mathbf{k}} f_s(\mathbf{k} \circ \mathbf{T}) \lambda_{\mathbf{n}-\mathbf{k}} \theta(2^J \circ \mathbf{x} - \mathbf{n}) - \\ &\sum_{\mathbf{n}} \sum_{\mathbf{k}} f_{ne}(\mathbf{k} \circ \mathbf{T}) \lambda_{\mathbf{n}-\mathbf{k}} \theta(2^J \circ \mathbf{x} - \mathbf{n}) \end{aligned} \quad (41)$$

因为  $\{\theta(2^J \circ \mathbf{x} - \mathbf{n})\}_{\mathbf{n} \in \mathbb{Z}^d}$  为正交基, 将式 (41) 代入式 (38) 可得:

$$\begin{aligned} e(f_s, f_{ne}) &= \|f_s - f_{ne}\|_{L^2}^2 = \\ &\sum_{\mathbf{n}} \left| \sum_{\mathbf{k}} f_s(\mathbf{k} \circ \mathbf{T}) \lambda_{\mathbf{n}-\mathbf{k}} - \sum_{\mathbf{k}} f_{ne}(\mathbf{k} \circ \mathbf{T}) \lambda_{\mathbf{n}-\mathbf{k}} \right|^2 \end{aligned} \quad (42)$$

另一方面, 将式 (34) 的矩阵展开, 可得:

$$F_C(f_s, f_{ne}) = \sum_{\mathbf{k}} \left| \sum_{\mathbf{n}} \lambda_{\mathbf{k}-\mathbf{n}} (f_s(\mathbf{n} \circ \mathbf{T}) - f_{ne}(\mathbf{n} \circ \mathbf{T})) \right|^2 \quad (43)$$

由式 (42) 和式 (43) 可得式 (39)。证毕。

所谓的过拟合, 就是准则函数减小, 而逼近误差增大的情况。然而式 (39) 意味着, 如果采用本文设计的准则函数, 那么准则函数的减小, 就等于逼近误差的减小。因此本文提出的准则函数可以有效抑制过拟合的发生, 继而保证本文算法的泛化能力。

## 5 仿真

正则化是一种有效的抗过拟合技术, 且被广泛应用于系统建模。为验证本文算法的有效性, 本文算法将应用于建立多维模型 (四维模型), 并跟正则化技术进行比较。

### 5.1 仿真中的建模对象

#### 5.1.1 建模对象的函数

设定仿真中的建模对象函数为:

$$\begin{aligned} g_s(x_1, x_2, x_3, x_4) &= \sum_{k_4=1}^4 \sum_{k_3=1}^4 \sum_{k_2=1}^4 \\ &\sum_{k_1=1}^4 a_{k_1, k_2, k_3, k_4} e^{(x_1-5k_1)^2 + (x_2-5k_2)^2 + (x_3-5k_3)^2 + (x_4-5k_4)^2} \end{aligned} \quad (44)$$

这里  $\{a_{k_1, k_2, k_3, k_4}\}_{k_1 \in [1, \dots, 4], \dots, k_4 \in [1, \dots, 4]}$  为在区间  $[0.5, 1.5]$  上均匀分布的随机数。仿真在区域  $[-20, 20] \times [-20, 20] \times [-20, 20] \times [-20, 20]$  上使用间隔为  $1/8$  的样本训练神经网络。用间隔为  $1/16$  的样本验证模型精度和泛化能力。

### 5.1.2 训练中的噪声

在工程实践中, 噪声常常服从正态和均匀分布。为检验算法的抗噪声能力, 这里在训练样本中, 分别加入以下 4 种工程实践中常见的噪声。

1) 均值为 0, 方差为  $\sigma^2 = 10^{-3}$  且满足正态分布的平稳随机噪声。

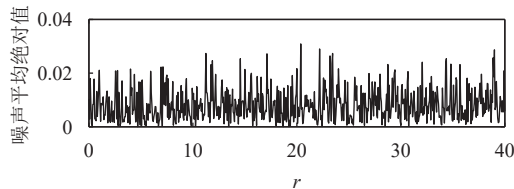
2) 均值为 0, 方差为  $\sigma^2 = 3.0 \times 10^{-4}$  且满足均匀分布的平稳随机噪声。

3) 均值为 0, 方差为  $\sigma^2(r) = 10^{-5} + 10^{-5}r^3$  且满足正态分布的非平稳随机噪声。

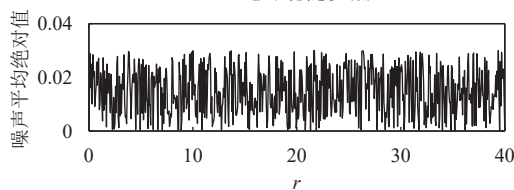
4) 均值为 0, 方差为  $\sigma^2(r) = 0.5 \times 10^{-4} + 10^{-4}r^3$  且满足均匀分布的非平稳随机噪声。

其中,  $r = \sqrt{x_1^2 + x_2^2 + x_3^2 + x_4^2}$  表示原本点到原点的距离,  $-20 \leq x_i \leq 20 (i = 1, \dots, 4)$ ,  $0 \leq r \leq 40$ 。令  $n_s(x_1, x_2, x_3, x_4)$  表示在坐标  $(x_1, x_2, x_3, x_4)$  处的噪声幅值。对于某次随机噪声实现, 将具有相同距离  $r$  的  $n_s(x_1, x_2, x_3, x_4)$  分为一组, 并求取相应的绝对平均值  $|\widetilde{n_s(r)}|$ 。以  $r$  为横坐标, 噪声平均绝对值为纵坐标, 绘制  $|\widetilde{n_s(r)}|$ , 分布如图 2 所示。

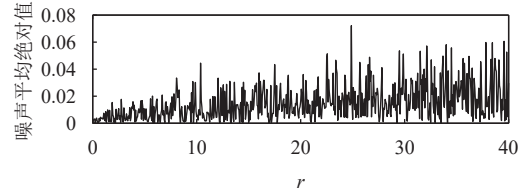
图中, 第 1 和第 2 种噪声, 属于较为常见的平稳随机过程, 因此可以较好地检验算法对于普通噪声的抑制能力。第 3 和第 4 种噪声的方差随位置的变化而变化, 即在靠近原点 ( $r=0$ ) 的地方能量较小, 而在远离原点时能量较大。因此第 3 和第 4 种噪声属于非平稳随机过程。这类噪声可以有效地检验本算法对噪声变化的鲁棒性。



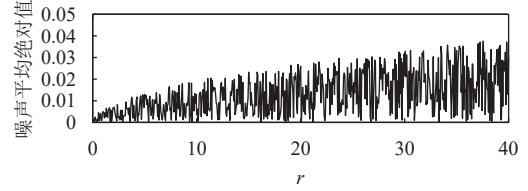
a. 正态平稳随机噪声



b. 均匀平稳随机噪声



c. 正态非平稳随机噪声



d. 均匀非平稳随机噪声

图 2 随机噪声实现的  $|\widetilde{n_s(r)}|$

## 5.2 小波神经网络参数

### 5.2.1 小波网络激励函数选取

由经典小波采样理论可知, Daubechies 多分辨率的尺度函数空间存在插值基。这满足本文算法激励函数的选取要求。为此这里选用四阶 Daubechies 尺度函数  $\phi(x)$  作为激活函数。

### 5.2.2 小波网络准则函数

小波网络准则函数, 由式 (3) 中的参数  $\{\lambda_k\}_{k \in \mathbb{Z}}$  决定。由经典小波采样理论可知, 对于 Daubechies 小波, 有:

$$\sum_{k=-\infty}^{+\infty} \lambda_k e^{-i\omega k} = \frac{1}{\sum_{k=-\infty}^{+\infty} \hat{\phi}(\omega + 2k\pi)} \quad (45)$$

式中,  $\hat{\phi}(\omega)$  表示  $\phi(x)$  的傅里叶变换。另一方面, 四阶 Daubechies 尺度函数  $\phi(x)$  的傅里叶变换满足:

$$\begin{aligned} \sum_{k=-\infty}^{+\infty} \hat{\phi}(\omega + 2k\pi) = & 1.0072e^{-i\omega} - 0.0338e^{-2i\omega} + \\ & 0.0396e^{-3i\omega} - 0.0118e^{-4i\omega} - \\ & 1.1982 \times 10^{-3}e^{-5i\omega} + 1.8838 \times 10^{-5}e^{-6i\omega} \end{aligned} \quad (46)$$

将 (46) 代入 (45) 可得:

$$\begin{aligned} \lambda_{-1} = 0.9929, \lambda_0 = 0.0333, \lambda_1 = -0.0379, \lambda_2 = 0.0091, \\ \lambda_3 = 0.0034, \lambda_4 = -0.0007, \lambda_5 = -0.0001, \lambda_6 = 0.0001 \end{aligned} \quad (47)$$

且  $\lambda_k = 0 (k \neq -1, \dots, 6)$ 。这样基于式 (47) 可以构建矩阵  $\mathbf{Q}_i = [\lambda_{k-n}]$  和  $\mathbf{\Lambda}_i = \mathbf{Q}_i^T \mathbf{Q}_i (i = 1, \dots, 4)$ , 进而应用式 (31) 和 (34) 构建对应的准则函数:

$$\mathbf{\Lambda} = \mathbf{\Lambda}_1 \otimes \dots \otimes \mathbf{\Lambda}_4 \quad (48)$$

## 5.3 正则化神经网络参数

在本仿真中, 将采用文献 [12] 中的经典正则

化技术。正则化神经网络具有与 RBF 网络相似的结构，其准则函数可表示为：

$$\sum_{n_4=-\infty}^{+\infty} \sum_{n_3=-\infty}^{+\infty} \sum_{n_2=-\infty}^{+\infty} \sum_{n_1=-\infty}^{+\infty} (f_{ne}(x_{n_1}, x_{n_2}, x_{n_3}, x_{n_4}) - f_s(x_{n_1}, x_{n_2}, x_{n_3}, x_{n_4}))^2 + \mu \|\bar{D}f_{ne}(x_1, x_2, x_3, x_4)\| \quad (49)$$

式中， $\mu \|\bar{D}f_{ne}(x,y)\|$ 表示正则化项； $\bar{D}$ 为微分类算子； $\mu$ 为正则化参数，其决定了神经网络输出平滑度和样本误差之间的折中程度。

微分算子 $\bar{D}$ 决定了正则化神经网络激励函数的选取。注意到式(44)中，对象函数是高斯函数的线性组合，为此仿真中通过选取微分算子 $\bar{D}$ ，使得正则化神经网络的激励函数为高斯函数。此外通过尝试，发现对于图 2 的 4 种噪声，当 $\mu = 0.02$ 时正则化神经网络可以获得最好的逼近结果，因此仿真中，本文算法将跟 $\mu = 0.02$ 时的正则化神经网络的学习结果进行比较。

5.4 仿真结果

令 $e_{ap}(x_1, x_2, x_3, x_4) = f_s(x_1, x_2, x_3, x_4) - f_{ne}(x_1, x_2, x_3, x_4)$ 表示神经网络 $f_{ne}(x_1, \dots, x_4)$ 和建模对象 $f_s(x_1, \dots, x_4)$ 之间的误差。仿真重复 500 次试验，每次试验随机产生参数 $\{a_{k_1, k_2, k_3, k_4} | k_1 \in [1, \dots, 4], \dots, k_4 \in [1, \dots, 4]\}$ 和噪声 $n_s(x_1, x_2, x_3, x_4)$ 。将两种神经网络 500 次试验误差取绝对值，其相应的统计值如表 1 所示。

表 1 绝对误差统计

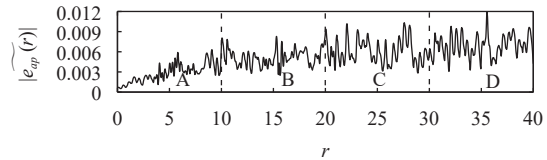
噪声	最大值	平均值
小波网络(正态平稳噪声)	0.017 5	0.003 14
正则化网络(正态平稳噪声)	0.039 50	0.006 92
小波网络(均匀平稳噪声)	0.026 22	0.005 09
正则化网络(均匀平稳噪声)	0.043 32	0.011 76
小波网络(正态非平稳噪声)	0.036 32	0.005 39
正则化网络(正态非平稳噪声)	0.088 58	0.012 40
小波网络(均匀非平稳噪声)	0.028 09	0.004 37
正则化网络(均匀非平稳噪声)	0.047 20	0.010 20

从表 1 可以看出，在平稳正态和均匀噪声条件下，小波网络的最大绝对误差和平均绝对误差分别小于正则化网络的最大绝对误差和平均绝对误差。另一方面，在非平稳正态和均匀噪声条件下，正则化网络的最大绝对误差和平均绝对误差也大于小波网络的平均绝对误差和平均绝对误差。

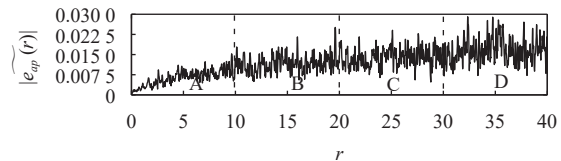
上述情况说明无论是在平稳还是非平稳噪声条件下，小波神经网络的抗噪声能力都要远好于正则

化技术。从表 1 的数据还可以看出，相比于正则化技术，小波网络在克服非平稳噪声上拥有更大的优势。

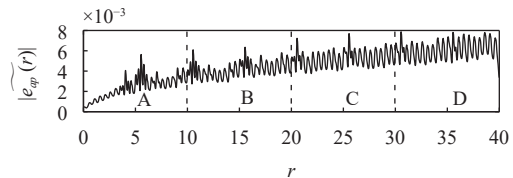
满足正态分布和均匀分布的两种非平稳随机噪声，其统计特性（方差）随距离 $r$ 的变化而变化，为阐明上述现象，这里将在相同距离 $r$ 上的误差 $e_{ap}(x_1, x_2, x_3, x_4)$ 分成一组，并求取绝对平均值 $|\widetilde{e_{ap}}(r)|$ 。以 $r$ 为横坐标、 $|\widetilde{e_{ap}}(r)|$ 为纵坐标绘制误差分布图，如图 3 所示。将横坐标分成 4 个区间，即 $[0,10)$ （A 区间）、 $[10,20)$ （B 区间）、 $[20,30)$ （C 区间）、 $[30,40)$ （D 区间）。然后统计各个区间上逼近误差的平均值，如表 2 所示。



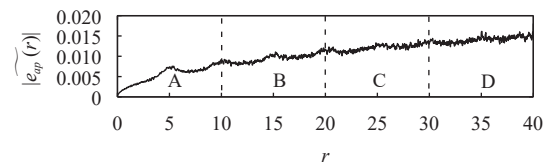
a. 波网络在正态非平稳噪声下的误差



b. 正则化网络在正态非平稳噪声下的误差



c. 小波网络在均匀非平稳噪声下的误差



d. 正则化网络在均匀非平稳噪声下的误差

图 3 神经网络逼近误差 $|\widetilde{e_{ap}}(r)|$

表 2 非平稳噪声条件下逼近误差绝对平均值

网络	区间A	区间B	区间C	区间D
小波网络(正态)	0.002 81	0.005 14	0.006 02	0.006 88
正则化网络(正态)	0.006 18	0.011 68	0.014 07	0.016 69
小波网络(均匀)	0.002 37	0.004 15	0.005 20	0.006 07
正则化网络(均匀)	0.005 37	0.009 49	0.012 11	0.014 10

由图 2c 和图 2d 可知，随着距离 $r$ 逐渐增大，无论是小波网络还是正则化网络的逼近误差都随之

变大。然而由表 2 可知, 对于非平稳正态噪声, 在 A 区间, 正则化网络误差是小波网络误差的 2.20 倍, 但是在 D 区间, 正则化网络误差扩大为小波网络误差的 2.42 倍。类似地, 对于非平稳均匀噪声, 在 A 区间, 正则化网络为小波网络的 2.27 倍, 在 D 区间, 逐渐扩大为 2.33 倍。这主要是因为正则化技术抑制噪声的能力, 主要取决于式 (49) 中参数  $\mu$  的选取。 $\mu$  较大时, 虽然可以在 D 区间较好地抑制强噪声, 但容易导致噪声较小的 A 区间丢失信号的细节。反之较小的  $\mu$  参数虽然有助于保持 A 区间的信号细节, 却难以抑制 D 区间的强噪声。因此当噪声能量变化时, 很难选取合理的参数  $\mu$ 。这意味着在噪声统计特性发生变化时, 小波网络可以获得更高的逼近精度, 因此小波网络相比正则化技术对噪声更具鲁棒性。

## 6 结束语

应用插值基和正交基之间的映射关系, 本文构建了新型准则函数。同时利用该准则函数, 证明了当输入层权值满足条件时, 总可以获得反馈矩阵, 使得训练迭代具有全局收敛性, 进而提出了相应的训练算法, 并将该算法推广至多维的情况。此外, 本文还证明了该算法能够使准则函数的减小等价于逼近误差的减小, 因而可以有效避免过拟合。最后仿真结果验证了本文算法的有效性。

## 参考文献

- [1] TAWFIK A M. River flood routing using artificial neural networks[J]. Ain Shams Engineering Journal, 2023, 14(3): 1-10.
- [2] ANNA R, ALEXANDER S, ANDREY S, et al. Three-day forecasting of greenhouse gas CH<sub>4</sub> in the atmosphere of the Arctic Belyy Island using discrete wavelet transform and artificial neural networks[J]. Neural Computing and Applications, 2021, 33(16): 0311-10322.
- [3] RANA M A, ANDREA P, SALIM H. Comparison of different methodologies for rainfall-runoff modeling: Machine learning vs conceptual approach[J]. Natural Hazards, 2021, 105(3): 2987-3011.
- [4] CAVANAUGH J E, NEATH A A. The Akaike information criterion: background, derivation, properties, application, interpretation, and refinements[J]. Wiley Interdisciplinary Reviews: Computational Statistics, 2019, 11(3): E1460.
- [5] ASADULLAH K, SARWAT A I. Unified univariate-neural network models for lithium-ion battery state-of-charge forecasting using minimized Akaike Information Criterion algorithm[J]. IEEE Access, 2021, 9: 39154-39170.
- [6] LIU J Q. Sensor fault analysis of aero-engine using ensemble SCNN and Bayesian interval estimation[J]. Engineering Applications of Artificial Intelligence, 2023, 125: 1-14.
- [7] QIAN G Q, WU Y H, XU M. Multiple change-points detection by empirical Bayesian information criteria and gibbs sampling induced stochastic search[J]. Applied Mathematical Modelling, 2019, 72: 202-216.
- [8] GAO Y G, MA C L, AN S. Compound fault diagnosis for cooling dehumidifier based on RBF neural network improved by kernel principle component analysis and adaptive genetic algorithm[J]. Soft Computing, 2023, 27(3): 1599-1613.
- [9] SANDHYA S, HEMANT P. Hyperspectral image classification using principle component analysis and deep convolutional neural network[J]. Journal of Ambient Intelligence and Humanized Computing, 2022, DOI: 10.1007/s12652-022-03876-z.
- [10] FRANCESCA B, ERNESTO De V, LORENZO R, et al. Understanding neural networks with reproducing kernel Banach spaces[J]. Applied and Computational Harmonic Analysis, 2023, 62: 194-236.
- [11] BAEK J H, CHEN J S, KRISTEN S. A neural network-enhanced reproducing kernel particle method for modeling strain localization[J]. International Journal for Numerical Methods in Engineering, 2022, 123(18): 4422-4454.
- [12] TOMASO P, FEDERICO G. Networks for approximation and learning[J]. Proceedings of the IEEE, 1990, 78(9): 1481-1497.
- [13] LI G, GU Y T, DING J.  $L^1$  regularization in two-layer neural networks[J]. IEEE Signal Processing Letters, 2022, 29: 135-139.
- [14] GUI S, CHEN Z D, LU B Z. Molecular sparse representation by a 3D ellipsoid radial basis function neural network via  $L^1$  regularization[J]. Journal of Chemical Information and Modeling, 2020, 60(12): 6054-6064.
- [15] JUN Z, WALTER G G, MIAO Y. Wavelet neural networks for function learning[J]. IEEE Transactions on Signal Processing, 1995, 43(6): 1485-1496.
- [16] MALLAT S. A wavelet tour of signal processing[M]. 2ed. Beijing: China Machine Press, 2003.
- [17] ZHANG Z G, LI Y. Recovery of the optimal approximation from samples in wavelet subspace[J]. Digital Signal Processing, 2012, 22(5): 795-807.

编辑 叶芳