

多稀疏回声状态网络预测模型

沈力华¹, 陈吉红^{1†}, 曾志刚², 杜宝瑞³, 金 健¹

(1. 华中科技大学 机械科学与工程学院, 湖北 武汉 430074;

2. 华中科技大学 自动化学院, 湖北 武汉 430074; 3. 沈阳飞机工业(集团)有限公司, 辽宁 沈阳 110034)

摘要: 针对单回声状态网络难以充分描述数据信息的问题, 提出多稀疏回声状态网络预测模型. 通过对相关回声状态网络的组合权值及由相关样本得到的基函数的权值同时进行学习, 获得优化的多个稀疏回声状态网络组合模型. 所提模型不同于双稀疏相关向量机等多核学习模型, 它不需要选择特定的核函数及相应的核参数. 因此, 该模型不但能更好的描述数据信息, 避免了双稀疏相关向量机及其他多核学习中核函数及其参数不易选择的问题. 同时, 所提模型不需要采用交叉验证的方式确定回声状态网络的谱半径和稀疏度, 只需确定相应的区间. 本文通过两组标杆数据和一组实际数据仿真实验, 与传统回声状态网络方法相比, 验证了所提模型具有更好的预测性能.

关键词: 回声状态网络; 稀疏; 预测模型; 相关向量机

引用格式: 沈力华, 陈吉红, 曾志刚, 等. 多稀疏回声状态网络预测模型. 控制理论与应用, 2018, 35(4): 421–428

中图分类号: TP183 **文献标识码:** A

Prediction model with multiple sparse echo state network

SHEN Li-hua¹, CHEN Ji-hong^{1†}, ZENG Zhi-gang², DU Bao-rui³, JIN Jian¹

(1. School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan Hubei 430074, China;

2. School of Automation, Huazhong University of Science and Technology, Wuhan Hubei 430074, China;

3. Shenyang Aircraft Industry Group Co., LTD, Shenyang Liaoning 110034, China)

Abstract: Considering the problem that using a single echo state network (ESN) is difficult to describe the data information adequately, we propose a multiple sparse echo state network prediction model. The optimized combination model of echo state network is achieved by learning the sparse weights of the related ESN and the sparse weights of related basis functions determined by related sample simultaneously. And the proposed model is achieved with no need of determining the kernel functions and the related kernel parameters, which is different from the double sparse relevance vector machine and the other multiple kernel learning models. So the proposed model not only can describe the information of the datasets better but also can avoid the selection procedure of kernel functions and kernel parameters. There is no need of selecting the spectral radius and sparsity of ESN by cross validation in the proposed model and only the interval of spectral radius and sparsity are needed to be determined. The experimental results of two groups of benchmarking data and a group of real-world dataset demonstrate that the proposed model has better prediction performance.

Key words: echo state network; sparse; prediction model; relevance vector machine

Citation: SHEN Lihua, CHEN Jihong, ZENG Zhigang, et al. Prediction model with multiple sparse echo state network. *Control Theory & Applications*, 2018, 35(4): 421–428

1 引言(Introduction)

在现实应用中, 数据常常表现为非线性特性. 由于神经网络具有较强的非线性逼近能力, 因此, 基于神经网络的预测模型得到了越来越快的发展, 并广泛应用于工业领域中^[1–3]. 与前向神经网络相比, 递归神经网络能够描述出动态系统内部特性, 因此, 其作为时间序列及回归预测模型, 往往能够得到更好的预测结

果. 特别是在时间序列预测中, 前一时刻与下一时刻的序列值往往具有较强的相关性, 递归神经网络通过递归结构描述这一相关特性, 从而获得了更高的预测精度^[4–8].

回声状态网络(echo state network, ESN)^[9]作为一种新型的递归神经网络, 其随机产生大规模稀疏连接的储备池, 将模型输入映射到高维空间中, 使数据在

收稿日期: 2017–05–12; 录用日期: 2017–11–22.

†通信作者. E-mail: d201477195@hust.edu.cn; Tel.: +86 15242484949.

本文责任编辑: 阳春华.

国家自然科学基金项目(51575210), 国家科技重大专项项目(2014ZX04001051)资助.

Supported by the National Natural Science Foundation of China (51575210) and the National Science and Technology Major Project of China (2014ZX04001051).

高维空间中具有线性特性. ESN首先随机初始化输入权值和内部权值, 然后只对输出权值进行学习, 因此, 其能够获得全局最优解, 且具有学习速度快的特点. 由于ESN的上述优点, 基于ESN及其改进算法, 如: 支持向量回声机 (support vector echo-state machine, SVESM)^[10]、基于模糊方法的ESN^[11]、基于优化算法的ESN^[12]等已被大量提出, 取得了快速的发展并广泛应用于时间序列及回归模型预测中.

然而, 由于现实应用中, 数据往往呈现多源、异构性, 采用单个回声状态网络或单个核函数的映射方式往往不能全面的描述系统的特性和信息, 针对上述问题, 多核学习模型获得了较多的应用^[13-16]. 特别是文献[17]提出一种双稀疏相关向量机 (doubly sparse relevance vector machine, DSRVM), 以下简称DSRVM. DSRVM作为多核学习模型的一种, 它通过学习得到了两组稀疏解, 提高了模型的泛化能力, 并被应用于脸部表情预测中. 然而, DSRVM与其他多核学习模型类似, 都面临着核函数及其相应参数不易选择的问题, 而不同的核函数和核参数, 将很大程度上影响预测结果.

针对上述问题, 基于DSRVM和回声状态网络, 本文提出一种多稀疏回声状态网络 (multiple sparse echo state network, MSESN), 将多个回声状态网络进行组合, 充分表达系统动态特性, 并采用变分推理及期望最大化同时学习多个ESN组合权值及由相关样本获得的基函数权值. 另外, 所提方法避免了核函数和核参数的选择, 且不需要交叉验证获得ESN的稀疏度和谱半径, 只需给定稀疏度和谱半径的区间, 模型即会自动选择最适合数据特性的参数, 需要确定的主要参数即是储备池的规模, 因此, 大大简化了模型训练过程. 两组标杆数据和一组实际数据的仿真实验验证了所提模型能够更好的刻画系统特性, 具有更高的预测精度.

2 回声状态网络和双稀疏相关向量机 (Echo state network and double sparse relevance vector machine)

2.1 回声状态网络 (Echo state network)

ESN由输入层、隐层和输出层组成, 其状态方程和输出方程如下所示:

$$\begin{aligned} \mathbf{x}(k+1) &= \\ f(\mathbf{W}_{\text{in}} \cdot \mathbf{u}(k+1) + \mathbf{W}_{\text{x}} \cdot \mathbf{x}(k)), & (1) \\ y(k+1) &= \mathbf{w}^T \cdot \mathbf{x}(k+1), & (2) \end{aligned}$$

其中: $\mathbf{u}(k+1)$ 为输入向量, $\mathbf{x}(k) \in \mathbb{R}^r$ 为储备池内部状态向量, r 为储备池规模, $y(k+1) \in \mathbb{R}$ 为模型输出. $\mathbf{W}_{\text{in}} \in \mathbb{R}^{r \times d}$, $\mathbf{W}_{\text{x}} \in \mathbb{R}^{r \times r}$ 和 $\mathbf{w} \in \mathbb{R}^r$ 分别为输入权值、内部连接权值和输出权值, f 为激活函数. ESN通常舍

去部分初始暂态过程以提高模型预测精度. 设 $\mathbf{A} \in \mathbb{R}^{N \times r}$ 和 \mathbf{y} 如下所示且 \mathbf{a}_k 为 \mathbf{A} 的第 k 个行向量:

$$\mathbf{A} = [x(l_z) \ x(l_z + 1) \ \cdots \ x(l_z + N - 1)]^T, \quad (3)$$

$$\mathbf{y} = [y(l_z) \ y(l_z + 1) \ \cdots \ y(l_z + N - 1)]^T, \quad (4)$$

其中 l_z 为训练样本的开始索引. 根据ESN的结构特性, 可得到下式:

$$\mathbf{A}\mathbf{w} = \mathbf{y}, \quad (5)$$

伪逆方法为ESN学习的基本方法之一, 通过伪逆方法求得的 \mathbf{w} 如下式所示:

$$\mathbf{w} = \mathbf{A}^\dagger \mathbf{y} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y}, \quad (6)$$

其中 \mathbf{A}^\dagger 指 \mathbf{A} 的伪逆, 可通过奇异值分解 (singular value decomposition, SVD) 求得.

由于伪逆方法容易产生病态问题, 降低模型泛化能力. 岭回归回声状态网络 (ridge regression echo state network, RESN) 通过在目标函数中引入正则项避免了伪逆方法病态解的问题. 加入正则项的目标函数如下所示:

$$J(\mathbf{w}) = \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{w}\|^2 + \frac{1}{2} C \|\mathbf{w}\|^2, \quad (7)$$

由此得到的输出权值 \mathbf{w} 如下所示:

$$\mathbf{w} = (\mathbf{A}^T \mathbf{A} + C\mathbf{I})^{-1} \mathbf{A}^T \mathbf{y}, \quad (8)$$

其中: C 为正则化参数, $\mathbf{I} \in \mathbb{R}^{r \times r}$ 为单位矩阵.

2.2 双稀疏相关向量机 (Double sparse relevance vector machine)

DSRVM^[17] 是基于相关向量机提出的, 它定义回归函数 y 具有如下形式:

$$y(\mathbf{x}; \mathbf{w}, \mathbf{v}) = \sum_{m=1}^M \sum_{k=1}^K w_m v_k \kappa_k(\mathbf{x}, \mathbf{x}_m), \quad (9)$$

上式中: $\{\kappa_k\}_{k=1}^K$ 为一组预先定义的核, v_k 为对应核的权值, 在DSRVM中选择的是一组具有不同核参数的高斯核. w_m 为第 m 个基函数的权值, 第 m 个基函数可以表示为 $\varphi_d = v_k \kappa_k(\mathbf{x}, \mathbf{x}_m)$. 假设有 N 个样本对 (\mathbf{x}, t) , 则上式中 $M = N$. DSRVM基于贝叶斯框架解决上式中的回归问题. 假设 ϵ 为服从均值为0, 方差为 σ^2 的高斯分布的误差, 如 $t = y(\mathbf{x}; \mathbf{w}, \mathbf{v}) + \epsilon$, t 为目标输出. 由此可得到下式:

$$p(t|\mathbf{w}, \mathbf{v}, \sigma^2, \mathbf{x}) \sim N(t; y(\mathbf{x}; \mathbf{w}, \mathbf{v}), \sigma^2), \quad (10)$$

其中: $N(t; y, \sigma^2)$ 指关于变量 t 以 y 为均值; σ^2 为方差的高斯分布. 上式适用于所有训练数据和测试数据. 进一步假设所有观察值之间相互独立, 同时DSRVM定义了两个高斯先验分布:

$$p(\mathbf{w}|\boldsymbol{\alpha}) \sim N(\mathbf{w}; 0, \mathbf{A}_0^{-1}), \quad (11)$$

$$p(\mathbf{v}|\boldsymbol{\beta}) \sim N(\mathbf{v}; 0, \mathbf{B}^{-1}), \quad (12)$$

其中 \mathbf{A}_0 和 \mathbf{B} 分别是由向量 $\boldsymbol{\alpha}$ 和向量 $\boldsymbol{\beta}$ 组成的对角矩阵. 假设 $\Omega = (\mathbf{w}, \mathbf{v}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2)$ 为所有超参数的集合, 则对于一个给定的样本输入 \mathbf{x}_{new} , 其待预测的目标值 t_{new} 的概率分布如下式所示:

$$p(t_{\text{new}}|\mathbf{t}, \mathbf{X}, \mathbf{x}_{\text{new}}) = \int p(t_{\text{new}}|\Omega, \mathbf{x}_{\text{new}})p(\Omega|\mathbf{t}, \mathbf{X})d\Omega, \quad (13)$$

模型训练过程需要找到超参数的后验概率分布, 而 (\mathbf{w}, \mathbf{v}) 的联合后验概率无法解析获得, 因此采用了变分近似方法进行求解, \log 似然函数 $p(\mathbf{t}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2)$ 的变分下界为

$$L = \int q(\mathbf{w})q(\mathbf{v}) \log\left(\frac{p(\mathbf{w}, \mathbf{v}, \mathbf{t}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2)}{q(\mathbf{w})q(\mathbf{v})}\right)d\mathbf{w}d\mathbf{v}, \quad (14)$$

$q(\mathbf{w})$ 和 $q(\mathbf{v})$ 通过变分方法获得, 其他3个超参通过给定 $q(\mathbf{w})$ 和 $q(\mathbf{v})$, 最大化上式中的变分下限得到.

DSRVM在变分和相关向量机框架下提出了一种多核组合优化方法, 该方法在表情识别回归数据中得到了较好的结果, 但DSRVM仍然存在以下问题: 该模型需要事先确定核函数的类型, 而核函数的选择则直接影响模型的预测性能, 当选择了不适合待预测数据集的核函数时, 往往不能获得较好的预测结果, 同时多个核函数的核参数选择也是一个非常耗时的任务. 因此, 针对上述问题, 本文基于DSRVM, 提出一种不需要确定核函数和核参数的方法, 采用多个回声状态网络对数据进行映射, 多个回声状态网络根据数据特性自适应的学习模型参数, 无需事先设定核类型, 避免了由于核函数和核参数选择不当对数据预测结果产生的影响. 此外, 所提模型不需要确定回声状态网络具体的稀疏度和谱半径, 只需给定一个范围, 模型即会自动获取更适合数据集的参数, 因此, 对回声状态网络的参数选择具有更宽松的要求.

3 多稀疏回声状态网络(Multiple sparse echo state network)

2004年, Jaeger发现采用ESN对典型时间序列进行预测^[9], 大大提高了时间序列预测的准确性, 表明通过ESN对数据进行映射, 能够更好的挖掘数据内部信息特性. 与核方法类似, ESN也是将数据映射到高维空间, 不同的是, 核方法为隐性映射, ESN为显性映射. 文献[10]采用ESN替代支持向量机中的核而得到一种无核支持向量机, 取得了更好的预测结果. 因此, 本文尝试采用多个ESN组合优化代替DSRVM中的多个核函数, 得到一种多稀疏回声状态网络模型. 实验结果表明, 所提方法与DSRVM相比, 具有更高的预测精度.

3.1 模型推理(Model inference)

为解决单回声状态网络表达信息单一的问题, 并基于DSRVM提出一种多稀疏回声状态网络预测模型. MSESN通过将多个回声状态网络进行优化组合, 充分表达系统动态信息, 并采用变分推理及期望最大化的方法同时学习不同回声状态网络的组合权值及由相关样本确定的基函数的权值, 且得到的两组权值均具有稀疏性, 进一步提高了模型的泛化能力. 对于给定的 N 个 d 维训练样本: $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N$, 其中 $\mathbf{X}_i = [x_1(i) \ x_2(i) \ \dots \ x_d(i)]^T \in \mathbb{R}^d$, $i = 1, 2, \dots, N$, 定义MSESN输出方程如下所示:

$$y(\mathbf{X}) = \sum_{i=1}^N w_i \psi_i(\mathbf{X}), \quad (15)$$

上式中: w_i 为第 i 个基函数的权值, $\psi_i(\mathbf{X})$ 为经过第 i 个样本和多个ESN映射再加权后的第 i 个基函数. 假设采用 K 个ESN对输入向量进行组合映射, 第 k 个ESN的权值设为 E_k , 则 $\psi_i(\mathbf{X})$ 可表示为

$$\psi_i(\mathbf{X}) = \sum_{k=1}^K E_k \varphi_k(\mathbf{X}, \mathbf{X}_i), \quad (16)$$

假设输入向量 \mathbf{X}_i 经第 k 个ESN通过式(1)映射到 r 维空间得到的向量为

$$\boldsymbol{\phi}_i = [\phi_{i1} \ \phi_{i2} \ \dots \ \phi_{ir}], \quad (17)$$

则所有训练样本经第 k 个ESN通过式(1)映射到 r 维空间得到的矩阵设为

$$\Phi_k = \begin{bmatrix} \phi_{11} & \dots & \phi_{1r} \\ \vdots & \ddots & \vdots \\ \phi_{N1} & \dots & \phi_{Nr} \end{bmatrix}. \quad (18)$$

在 r 维空间中两个样本间的欧几里得距离矩阵设为 $\mathbf{D}_k \in \mathbb{R}^{N \times N}$, \mathbf{D}_k 的第 i 行第 j 列个元素为 $\mathbf{D}_k(i, j) = (\boldsymbol{\phi}_i - \boldsymbol{\phi}_j)(\boldsymbol{\phi}_i - \boldsymbol{\phi}_j)^T$, \mathbf{D}_k 对角线上的元素为0.

$$\mathbf{D}_k = \begin{bmatrix} 0 & \dots & D_k(1, N) \\ \vdots & 0 & \vdots \\ D_k(N, 1) & \dots & 0 \end{bmatrix}, \quad (19)$$

则经过 K 个ESN的映射, 将得到 K 个距离矩阵 \mathbf{D}_k . 设三维矩阵 $\boldsymbol{\Psi} \in \mathbb{R}^{N \times M \times K}$, 其中 $N = M$, 在此是为了区分第1维和第2维, $\boldsymbol{\Psi}$ 的第 k 个矩阵分量设为 \mathbf{D}_k .

$$\varphi_k(\mathbf{X}_n, \mathbf{X}_m) = \boldsymbol{\Psi}(n, m, k). \quad (20)$$

从以上各式可以看出, 不同于DSRVM及其他多核学习, MSESN不需要选择核函数, 而是采用多个ESN映射替代多个核的映射, 且不需要确定核参数, 大大简化了模型学习过程, 而且避免了核函数选择不当影响预测结果的问题. 当有测试样本需要预测时, 可通过以上各式, 确定输出方程(9)的具体形式. 需要估计的参数为 $\mathbf{w} = [w_1 \ \dots \ w_i \ \dots \ w_N]$ 和 $\mathbf{E} = [E_1 \ \dots \ E_k \ \dots \ E_K]$.

在 \mathbf{w} 和 \mathbf{E} 的优化过程中,通过采用变分推理方法获得 \mathbf{w} 和 \mathbf{E} 的后验分布 $q(\mathbf{w})$ 和 $q(\mathbf{E})$,已知 $q(\mathbf{w})$ 和 $q(\mathbf{E})$ 后,将其带入式(14),再令式(14)对其他3个超参数求导,令导数为零获得3个超参数的值,如此迭代直至算法收敛.采用DSRVM中的变分推理和期望最大化学习算法进行参数估计^[17],根据文献[17],假设 \mathbf{w} 和 \mathbf{E} 的先验分布为0均值高斯分布,通过变分推理与期望最大化的混合方法迭代近似求解 \mathbf{w} 和 \mathbf{E} 的后验分布 $q(\mathbf{w})$ 和 $q(\mathbf{E})$,最终确定两个后验分布的参数值,并将后验分布的均值作为最终学习得到的 \mathbf{w} 和 \mathbf{E} .

ESN模型通过对状态矩阵求伪逆获得输出权值,其计算复杂度为 $O(r^2N + r^3)$,其中: r 为储备池规模, N 为训练样本数.与DSRVM类似,由于所提模型中的相关向量和相关ESN的个数都远小于训练样本数,MSESN的复杂度为 $O(N^3)$.相比于ESN,MSESN虽增加了计算复杂度,但在实际应用中,随着计算能力的提升,预测精度仍是很多实际问题重点考虑的内容,因此,具有更高预测精度的MSESN是具有实际应用价值和意义的.

3.2 模型预测(Prediction)

通过学习得到 \mathbf{w} 和 \mathbf{E} 后,由于DSRVM和所提模型本质上是基于相关向量机的学习方法,因此得到的 \mathbf{w} 和 \mathbf{E} 具有稀疏性,可进一步提高模型的泛化性能,并提高模型的预测效率.在模型预测过程中,首先从 \mathbf{w} 和 \mathbf{E} 中,提取出非零元素,假设 \mathbf{w} 中的非零元素有 s 个,且非零元素组成的向量为 $\tilde{\mathbf{w}} = [\tilde{w}_1 \ \tilde{w}_2 \ \cdots \ \tilde{w}_s]$, \mathbf{E} 中的非零元素有 e 个,非零元素组成的向量为 $\tilde{\mathbf{E}} = [\tilde{E}_1 \ \tilde{E}_2 \ \cdots \ \tilde{E}_e]$.此时,输出方程可表示成如下形式:

$$y(\mathbf{X}) = \sum_{i=1}^s \tilde{w}_i \tilde{\psi}_i(\mathbf{X}). \quad (21)$$

对于每一个ESN,从 $\tilde{\Phi}_k$ 中选取 \mathbf{w} 中非零元素对应索引的行向量,组成新的矩阵 $\tilde{\Phi}_k \in \mathbb{R}^{s \times r}$,同时去除掉 \mathbf{E} 中零元素对应索引的 $\tilde{\Phi}_k$.此时得到 e 个矩阵 $\tilde{\Phi}_k$,此时 $k = 1, 2, \dots, e$.对于三维矩阵 $\mathbf{H} \in \mathbb{R}^{e \times s \times r}$, \mathbf{H} 为由 e 个矩阵分量 $\tilde{\Phi}_k$ 组成的. \mathbf{H} 其实为 s 个相关训练样本通过 e 个相关ESN映射到 r 维空间上的一个三维矩阵.假设有 l 个测试样本 $\mathbf{X}t_1, \mathbf{X}t_2, \dots, \mathbf{X}t_l$,其中 $\mathbf{X}t_i = [xt_1(i), xt_2(i), \dots, xt_d(i)]^T \in \mathbb{R}^d$, $i = 1, 2, \dots, l$,首先采用模型训练时产生的 e 个相关ESN,将测试集映射到 r 维空间得到一个三维矩阵 $\tilde{\mathbf{H}} \in \mathbb{R}^{e \times l \times r}$, $\tilde{\mathbf{H}}$ 由 e 个矩阵 $\tilde{\mathbf{H}}_k \in \mathbb{R}^{l \times r}$ 组成,其中 $k = 1, 2, \dots, e$.将 $\tilde{\mathbf{H}}_k$ 与 $\tilde{\Phi}_k$ 按照式(13)计算得到欧几里得距离矩阵 $\tilde{\mathbf{D}}_k \in \mathbb{R}^{l \times s}$,设 $\tilde{\mathbf{H}}_k$ 的第 i 行向量为 $\tilde{\mathbf{h}}_{ki}$, $\tilde{\Phi}_k$ 的第 j 个行向量为 $\tilde{\phi}_{kj}$, $\tilde{\mathbf{D}}_k$ 的第 i 行第 j 列元素 $\tilde{D}_k(i, j)$ 可通过如下公式计算:

$$\tilde{D}_k(i, j) = (\tilde{\mathbf{h}}_{ki} - \tilde{\phi}_{kj})(\tilde{\mathbf{h}}_{ki} - \tilde{\phi}_{kj})^T. \quad (22)$$

式(15)中

$$\tilde{\psi}_i(\mathbf{X}) = \sum_{k=1}^e E_k \varphi_k(\mathbf{X}, \mathbf{X}_i), \quad (23)$$

其中

$$\varphi_k(\mathbf{X}_i, \mathbf{X}_j) = \tilde{D}_k(i, j). \quad (24)$$

将式(22)–(24)代入式(21),即可得到测试样本的预测值.由此得到一种多稀疏回声状态网络模型(MSESN),得到的MSESN由 e 个回声状态网络加权得到,且预测值由 e 个回声状态网络及 s 个相关向量共同确定,每个回声状态网络的权值及由相关向量确定的基函数的权值通过同时优化、学习获得.MSESN避免了核函数的选择及核参数的确定问题,它通过多个回声状态网络将系统信息映射到高维空间,对系统信息进行了更充分的表达和描述,从而获得了更高的预测精度.MSESN不需要采用交叉验证的方式确定ESN的谱半径和稀疏度,只需确定相应的区间即可,两组标杆数据和一组实际数据的仿真实验证明了所提模型的有效性.

4 仿真实例(Simulation examples)

为证明所建立模型的泛化性能,在训练数据集上完成模型的训练后,在测试数据集上,将MSESN与ESN^[9],SVESM^[10],DSRVM^[17]等方法进行比较,通过均方误差(mean square error, MSE)、归一化根均方误差(normalized root mean square error, NRMSE)、平均绝对偏差(mean absolute deviation, MAD)来评价模型预测性能和泛化性能,MSE, NRMSE, MAD的计算公式如下所示:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (p_i - o_i)^2, \quad (25)$$

$$\text{NRMSE} = \left(\sum_{i=1}^N (p_i - o_i)^2 / \sum_{i=1}^N (o_i - \bar{o})^2 \right)^{1/2}, \quad (26)$$

$$\text{MAD} = \frac{1}{N} \sum_{i=1}^N |o_i - p_i|. \quad (27)$$

设所提模型的储备池规模为 r ,在3组数据仿真分析中,所提模型的参数设置如表1所示,从表1可以看出,不同于常规ESN训练过程,采用交叉验证得到稀疏度和谱半径的值,而是给定相应的最小值、每个值之间的间隔以及最大值,[0.01:0.01:0.04]表示设定的区间及间隔,最小值为方括号中的第1个值0.01,最大值为方括号中的第3个值0.04,每个值距离下一个值的间隔为方括号中的第2个值0.01.给多个ESN分配不同的稀疏度和谱半径,再通过优化学习,选择最适合训练数据的储备池.所建立模型采用多个回声状态网络加权组合优化,代替多个核函数.采用多个回声状态网络将原始数据映射到高维空间的方法来表达多源异构信息.文章对SinC函数和函数 $f(x) = \sin(ax) + \sin(bx) + cx$ 进行预测,第1个函数由一个sin函数和

一个线性函数相除获得, 第2个函数由不同频率的两个sin函数和一个线性函数组成, 上述两组函数是由周期函数和线性函数进行组合运算获得的, 其具有多源异构信息的特性. 文章最后通过一组UCI数据集中的机翼自噪声数据集(airfoil self-noise data set)^[18]进行仿真实验, 进一步验证了所提方法的有效性. MSES N采用多个ESN进行组合优化获得, 它通过随机初始化获得输入权值, 因此, 在每次实验中, 会得到不同的预测精度. 鉴于以上, 文章在评价 MSES N, ESN 和 SVESM等模型性能时, 均是采用50次预测结果的平均值, 来获得均方误差、归一化根均方误差、平均绝对偏差的值. 文章在仿真实验过程中, 将训练集中20%的数据作为验证集, 用来分别确定 MSES N, ESN, SVESM以及DSRVM等模型的全局最优参数值.

表 1 MSES N模型参数设置

Table 1 Parameters setting of MSES N

参数	r	稀疏度	谱半径	ESN个数
SinC函数	45	[0.01:0.01:0.04]	[0.59:0.01:0.69]	10
组合函数	45	[0.01:0.01:0.04]	[0.59:0.01:0.99]	40
UCI数据	80	[0.005:0.005:0.02]	[0.59:0.01:0.99]	40

4.1 SinC函数(SinC function)

SinC函数如下式所示:

$$y(x) = \begin{cases} \frac{\sin(\pi x)}{\pi x}, & x \neq 0, \\ 1, & x = 0, \end{cases} \quad (28)$$

在(-10, 10)上随机生成1000组数据作为训练样本, 同时向训练目标值加入0.2倍的高斯噪声作为干扰. 同时采用式(28)产生1000组测试样本, 测试样本不加入噪声干扰, 实验结果如表2所示.

表 2 不同模型比较结果(SinC函数)

Table 2 SinC function simulation prediction results

模型	ESN	SVESM	DSRVM	MSES N
MSE	0.0071	0.0046	0.0020	0.0010
NRMSE	0.4174	0.3347	0.2121	0.1842
MAD	0.0590	0.0514	0.0364	0.0276

从表中可以看出, 相比于其他几种方法, 所提模型 MSES N在MSE, NRMSE, MAD3个性能指标上均优于其他3种方法, DSRVM虽然采用多个核函数对输入向量进行映射, 预测精度仍低于MSES N, 因此, 所提模型通过多个回声状态网络更好的表达了系统内部信息, 获得了更好的预测结果.

图1为所提模型的拟合曲线, 从图中可以看出, 即使在训练样本中加入较多噪声干扰, MSES N仍能较好的拟合SinC函数, 图2为DSRVM的拟合曲线, 从图中可以看出, 当自变量大于4时, DSRVM已经不能很

好的捕捉SinC函数特性, 因此预测精度相对较低. 图3为所提模型选择的ESN个数随着迭代的变化情况, 从图中可以看出, 当迭代步数小于60时, ESN的个数即已确定, 最终选择了最能表达数据信息的4个ESN进行优化组合. 图4为DSRVM核函数选择的收敛情况, DSRVM最终选择了3个核函数.

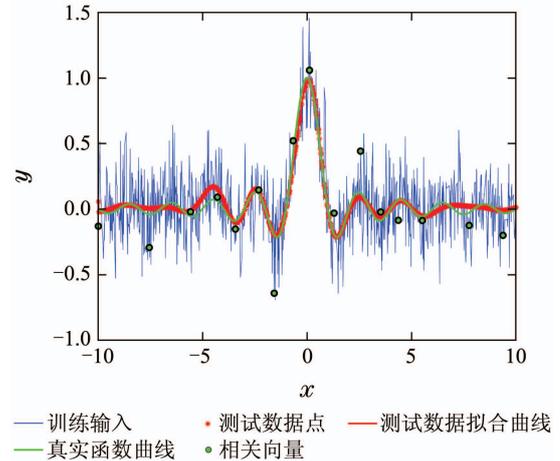


图 1 MSES N拟合曲线(SinC函数)

Fig. 1 MSES N fitted curve (SinC function)

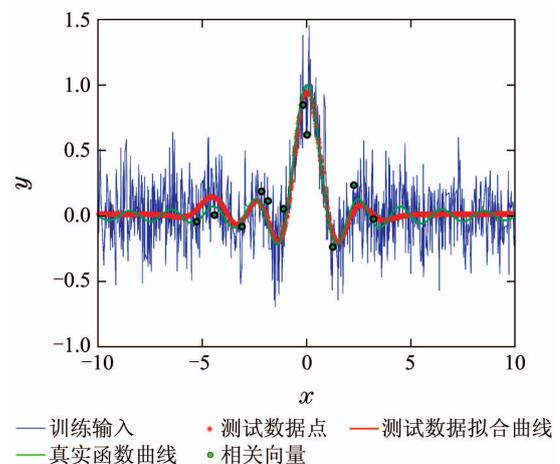


图 2 DSRVM拟合曲线(SinC函数)

Fig. 2 DSRVM fitted curve (SinC function)

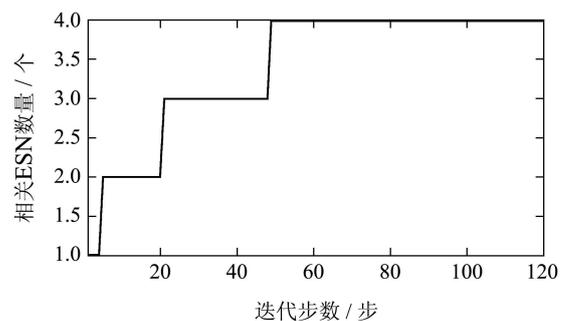


图 3 MSES N模型ESN个数收敛曲线(SinC函数)

Fig. 3 Convergence curve of ESN number of MSES N (SinC function)

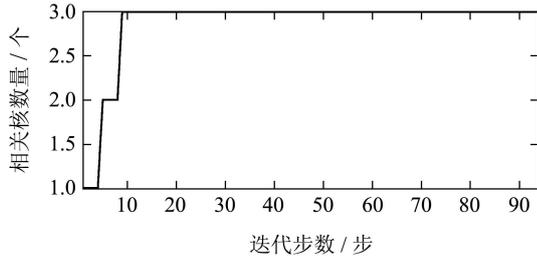


图4 DSRVM模型核函数个数收敛曲线(SinC函数)

Fig. 4 Convergence curve of kernel number of DSRVM (SinC function)

4.2 组合函数(Combined functions)

在另一组实验中,采用文献[16]所使用的标杆函数进行仿真分析,该函数由一个线性函数和两个sin函数组成,该函数更能反映出模型对于多源异构信息的捕捉能力,函数具体形式如下式所示:

$$f(x) = \sin(ax) + \sin(bx) + cx, \quad (29)$$

其中:

$$a = 3, b = 1, c = 0.1.$$

在(0, 20)上随机生成500组数据作为训练样本,同时为进一步验证所提模型的抗干扰能力,向训练目标值加入0.3倍的高斯噪声作为干扰.同时采用式(29)产生500组测试样本,测试样本不加入噪声干扰.将MSESN与ESN, SVESM, DSRVM的预测结果进行比较分析,同样采用RMSE, NRMSE, MAD来评价回归模型性能.实验结果如表3所示.

表3 不同模型比较结果(组合函数)

Table 3 Combined functions simulation prediction results

模型	ESN	SVESM	DSRVM	MSESN
MSE	0.9542	0.5212	0.0110	0.0090
NRMSE	0.8826	0.6524	0.0942	0.0859
MAD	0.8204	0.6043	0.0830	0.0742

从表3可以看出,当函数为两个sin函数加上一个线性函数且加入较大噪声时,RESN和SVESM已不能得到较好的预测结果,虽然DSRVM仍具有较好的预测精度,而所提MSESN能更全面的描述函数中的多源信息,具有更好的预测性能.

将图5和图6进行比较分析,可以看出,MSESN对函数进行了更好的拟合,特别是自变量在14到16区间的函数峰值处,可以明显的看出DSRVM没有得到准确的拟合曲线.图7和图8分别为MSESN的ESN个数收敛曲线和DSRVM的核函数个数收敛曲线.图7显示MSESN最终选择了7个相关的ESN进行组合优化,而DSRVM选择了6个相关核.

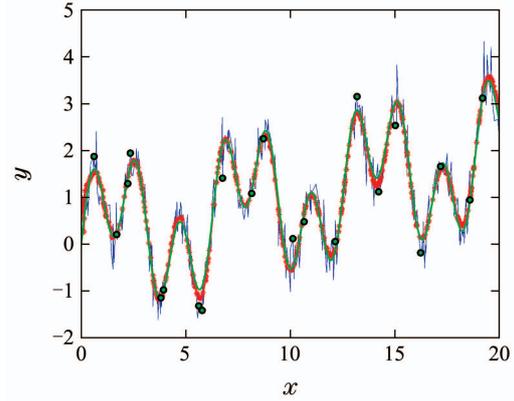


图5 MSESN拟合曲线(组合函数)

Fig. 5 MSESN fitted curve (combined functions)

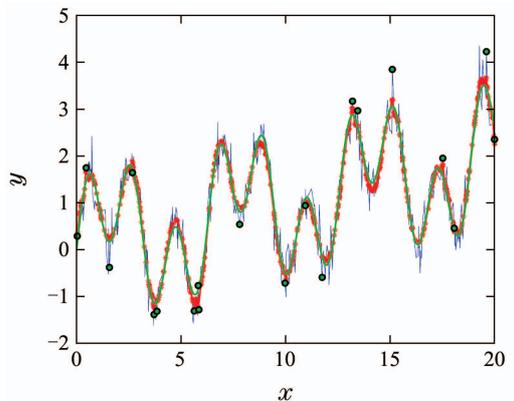


图6 DSRVM拟合曲线(组合函数)

Fig. 6 DSRVM fitted curve (combined functions)

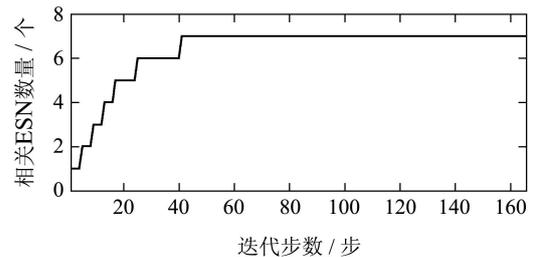


图7 MSESN模型ESN个数收敛曲线(组合函数)

Fig. 7 Convergence curve of ESN number of MSESN (combined functions)

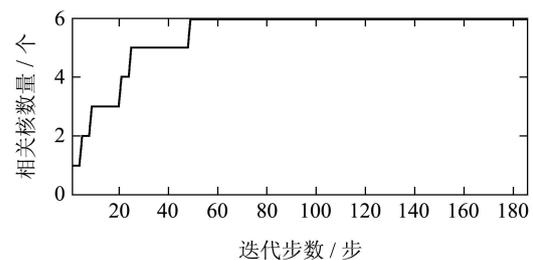


图8 DSRVM模型核函数个数收敛曲线(组合函数)

Fig. 8 Convergence curve of kernel number of DSRVM (combined functions)

从两组仿真实验中也可以看出,当拟合函数从SinC函数变为组合函数时,MSESN和DSRVM都分别选择了更多的ESN和更多的核,分析原因,主要是由于函数变为了多个sin函数和线性函数的组合,函数具有了更多特性和信息,因此两种模型根据数据特点自动选择了更适合数据的更多的ESN和核函数,以更全面的表达数据信息,从而也可以看出所提模型具有自适应的特点。

4.3 实际数据(Real-world data)

为进一步验证所提MSESN模型的有效性,将其应用于机翼自噪声数据集中,该数据集是由NASA提供的,该数据集通过无回声风洞中实施一系列空气动力学和声学测试实验获得,该数据含有1503个样本,6个属性,其中频率、迎角、弦长、自由气流速度和吸力面位移厚度作为输入,需要预测的目标输出是第6个属性,按比例缩小的声压级。在本仿真实验中,随机选择50%的数据作为训练数据,另外50%的数据作为测试数据,超参数的选择通过在训练数据集中交叉验证获得。将MSEN, ESN, SVESM, DRRVM分别应用于该数据集目标输出的预测中。预测结果如表4所示。

表4 不同模型比较结果(机翼自噪声数据)

Table 4 Airfoil self-noise dataSet prediction results

模型	ESN	SVESM	DSRVM	MSESN
MSE	15.2165	13.9412	17.8431	12.4795
NRMSE	0.5573	0.5315	0.6091	0.5141
MAD	2.9901	2.7476	3.2778	2.7278

从表4中可以看出,在3个性能指标上,所提模型相比于其他模型,都具有更好的预测结果,在该数据仿真中也可以看出,不同于前两组数据,SVESM的预测结果要好于DSRVM,分析其原因,主要是由于DSRVM所设定的核函数是有限的,一些特定的核函数更适于特定的数据集,而采用回声状态网络的预测模型则不需要确定特定的核函数,因此更具有自适应性,而MSESN通过将多个ESN优化组合,更好的描述了数据内部特性和信息,因此获得了更好的预测结果,该实际数据的仿真实验同样证明了所提模型的有效性。

5 结论(Conclusions)

文章通过将多个ESN进行组合优化,同时学习相关ESN的加权重值和由相关向量确定的相关基函数的权值,获得优化的多个稀疏回声状态网络组合模型,所提模型不但能充分表达复杂多源异构信息,且不需要选择特定的核函数及相应的核参数,解决了双稀疏相关向量机及其他多核学习方法中核函数及其参数不易选择的问题,同时MSESN只需确定谱半径和稀疏度的区间值以及相邻值的间隔即可,模型即会自动

选取适合训练数据集的ESN. 避免了交叉验证选择稀疏度和谱半径耗时的问题. 文章最后通过两组标杆数据和一组实际数据仿真实验证明其具有较好的预测性能和抗干扰能力。

参考文献(References):

- [1] QUAN H, SRINIVASAN D, KHOSRAVI A. Short-term load and wind power forecasting using neural network-based prediction intervals [J]. *IEEE transactions on neural networks and learning systems*, 2014, 25(2): 303 – 315.
- [2] JIANG Zhaohui, LI Xiyue, GUI Weihua. Blast furnace stockline prediction by segmented linear-regression and dynamic weighting neural network [J]. *Control Theory & Applications*, 2015, 32(6): 801 – 809. (蒋朝辉, 李晞月, 桂卫华. 分段线性回归和动态加权神经网络融合的高炉料位预测 [J]. 控制理论与应用, 2015, 32(6): 801 – 809.)
- [3] ZHANG Dongqing, NING Xuanxi, LIU Xueni. On-line prediction of nonlinear time series using RBF neural networks [J]. *Control Theory & Applications*, 2009, 26(2): 151 – 155. (张冬青, 宁宣熙, 刘雪妮. 基于RBF神经网络的非线性时间序列在线预测 [J]. 控制理论与应用, 2009, 26(2): 151 – 155.)
- [4] KUREMOTO T, KIMURA S, KOBAYASHI K. Time series forecasting using a deep belief network with restricted Boltzmann machines [J]. *Neurocomputing*, 2014, 137(1): 47 – 56.
- [5] RATHER A M, AGARWAL A, SASTRY V N. Recurrent neural network and a hybrid model for prediction of stock returns [J]. *Expert Systems with Applications*, 2015, 42(6): 3234 – 3241.
- [6] LI Zhanying, WANG Kejun, XU Liang. Approach of prediction of ship rolling based on chaotic diagonal recurrent neural networks [J]. *Control & Decision*, 2012, 27(11): 1681 – 1684. (李占英, 王科俊, 徐亮. 混沌对角递归神经网络的船舶横摇预报方法 [J]. 控制与决策, 2012, 27(11): 1681 – 1684.)
- [7] LI Huijun, XIAO Bing. Multistep recurrent neural network model predictive controller without constraints [J]. *Control Theory & Applications*, 2012, 29(5): 642 – 648. (李会军, 肖兵. 一种无约束多步递归神经网络预测控制器 [J]. 控制理论与应用, 2012, 29(5): 642 – 648.)
- [8] HAN Min, SHI Zhiwei, XI Jianhui. Learning the trajectories of periodic attractor using recurrent neural network [J]. *Control Theory & Applications*, 2006, 23(4): 497 – 502. (韩敏, 史志伟, 席剑辉. 应用递归神经网络学习周期运动吸引子轨迹 [J]. 控制理论与应用, 2006, 23(4): 497 – 502.)
- [9] JAEGER H, HAAS H. Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication [J]. *Science*, 2004, 304(5667): 78 – 80.
- [10] SHI Z, HAN M. Support vector echo-state machine for chaotic time-series prediction [J]. *IEEE Transactions on Neural Networks*, 2007, 18(2): 359 – 372.
- [11] HAN S I, LEE J M. Fuzzy echo state neural networks and funnel dynamic surface control for prescribed performance of a nonlinear dynamic system [J]. *IEEE Transactions on Industrial Electronics*, 2014, 61(2): 1099 – 1112.
- [12] CHOUIKHI N, AMMAR B, ROKBANI N. PSO-based analysis of Echo state network parameters for time series forecasting [J]. *Applied Soft Computing*, 2017, 55(1): 211 – 225.
- [13] WANG Xinying, HAN Min. Multivariate chaotic time series prediction using multiple kernel extreme learning machine [J]. *Acta Physica Sinica*, 2015, 64(7): 070504-1 – 070504-7. (王新迎, 韩敏. 多元混沌时间序列的多核极端学习机建模预测 [J]. 物理学报, 2015, 64(7): 070504-1 – 070504-7.)

- [14] RAKOTOMAMONJY A, BACH F R, CANU S, et al. SimpleMKL [J]. *Journal of Machine Learning Research*, 2008, 9(3): 2491 – 2521.
- [15] YANG H, XU Z, YE J, et al. Efficient sparse generalized multiple kernel learning [J]. *IEEE Transactions on Neural Networks*, 2011, 22(3): 433 – 446.
- [16] SONNENBURG S. Large scale multiple kernel learning [J]. *Journal of Machine Learning Research*, 2006, 7(1): 1531 – 1565.
- [17] KALTWANG S, TODOROVIC S, PANTIC M. Doubly sparse relevance vector machine for continuous facial behavior estimation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 38(9): 1748 – 1761.
- [18] BACHE K, LICHMAN M. *UCI machine learning repository* [EB/OL]. <http://archive.ics.uci.edu/ml>, 2013.

作者简介:

沈力华 (1984–), 女, 博士研究生, 研究方向为神经网络理论及应用, E-mail: sacslh@126.com;

陈吉红 (1965–), 男, 教授, 研究方向为智能数控系统、工业大数据建模与分析, E-mail: d201477195@hust.edu.cn;

曾志刚 (1971–), 男, 教授, 研究方向为神经网络理论与应用、切换系统控制理论与应用, E-mail: zgzung@hust.edu.cn;

杜宝瑞 (1970–), 男, 研究员, 研究方向为智能制造技术、工业大数据建模与分析, E-mail: sacdbr@163.com;

金健 (1964–), 女, 副教授, 研究方向为机床可靠性、工业大数据建模与分析, E-mail: 315901928@qq.com.