

基于深度去噪核映射的长期预测模型

王强, 吕政[†], 王霖青, 王伟

(大连理工大学控制科学与工程学院, 辽宁大连 116024)

摘要: 针对最小二乘支持向量机核函数结构较浅对其长期预测模型精度的限制, 采用深度学习中逐层特征提取的思想, 提出基于深度去噪核映射的最小二乘支持向量机长期预测模型. 该模型通过深度核网络的逐层变换, 将样本数据映射到深度特征空间, 从而有效提高其长期预测的精度. 此外, 为了提高模型对含高噪声数据的拟合性能, 将去噪算法融入深度核网络的训练过程中, 并通过反向传播算法对核网络参数进行整体微调. 标准数据集及实际工业数据的仿真实验结果表明, 所提方法能够有效提取数据中蕴含的特征信息, 提高预测模型的精度.

关键词: 特征提取; 深度学习; 最小二乘支持向量机; 长期预测; 去噪

中图分类号: TP18

文献标志码: A

Deep denoising kernel mapping-based long-term prediction model

WANG Qiang, LYU Zheng[†], WANG Lin-qing, WANG Wei

(School of Control Science and Engineering, Dalian University of Technology, Dalian 116024, China)

Abstract: In this study, employing the idea of layer by layer feature extraction in deep learning, a deep denoising kernel mapping-based least square support vector machine long-term prediction model is proposed. The proposed model can deal with the poor long-term prediction ability problem of the least square support vector machine with shallow kernel structure. By transforming through the deep kernel network layer by layer, the sample data are mapped to the deep feature space to improve the prediction accuracy. In addition, the denoising algorithm is incorporated into the training process of the deep kernel network to improve the fitting performance for the data with high level noises. Furthermore, the whole network is fine-tuned to further improve the modeling ability by using the back propagation algorithm. The results of the standard dataset and the actual industrial data simulation experiments show that the proposed method can extract the feature information contained in the data, and effectively improve the prediction accuracy.

Keywords: feature extraction; deep learning; least square support vector machine; long-term prediction; denoising

0 引言

数据预测在当今的工业生产中起着极为重要的作用, 对生产过程进行及时准确的预测可以极大地提高生产效率, 为工业生产调度和控制决策提供有力的支撑^[1].

针对数据预测问题, 学者进行了大量的研究. 文献 [2] 对输入和输出之间采用线性映射的自回归滑动平均模型 (Auto-regressive moving average model, ARMA)、人工神经网络模型 (Artificial neural network, ANN) 和采用非线性映射的自适应模糊神经推理系统模型 (Adaptive neuro-fuzzy inference systems, ANFIS) 进行了对比, 并对风力数据进行了预测, 实验结果表明基于非线性模型的预测精度较

高; 文献 [3] 提出了一种基于增强粒子群优化的混合预测模型, 其主要是将反向传播神经网络 (Back propagation neural network, BPNN) 和径向基函数神经网络相结合对风力数据进行预测; 文献 [4] 提出了一种基于反向传播的小波神经网络, 通过扩展的卡尔曼滤波来训练这个神经网络以建立输入和输出之间的关系, 最后对实际的负荷数据进行了短期预测; 文献 [5] 建立了一种贝叶斯神经网络 (Bayesian neural network, BNN) 的电力系统短期负荷预测模型, 采用贝叶斯自相关分析方法来确定影响电力负荷变化的因素, 通过这些因素对系统进行预测; 文献 [6] 提出了一种基于高效极限学习机和差分进化算法的混合模型来预测混沌时间序列, 其中高效极限学习机用了一

收稿日期: 2017-10-30; 修回日期: 2018-03-12.

基金项目: 国家自然科学基金项目 (61533005, 61703070, 61603069); 中央大学基础研究基金项目 (DUT16RC(3)031).

作者简介: 王强 (1985—), 男, 博士生, 从事流程工业过程建模与优化的研究; 王伟 (1955—), 男, 教授, 博士生导师, 从事复杂系统建模与控制、流程工业综合自动化技术等研究.

[†]通讯作者. E-mail: lvzheng@dlut.edu.cn.

个简化的完全正交分解替换普通的奇异值分解来计算输出权重,进而提高了极限学习机的学习速度.然而,以上这些预测方法均存在一定的缺点,如模型易过拟合、训练时需要较大的计算资源、模型结构不易确定等.此外,上述方法均针对数据进行短期预测,当预测长度过大时,模型的性能明显下降,预测精度无法达到使用需求^[7].

支持向量机(Support vector machine, SVM)是基于统计学习理论的小样本学习机,具有严格的数学理论基础,其泛化能力强、计算速度快,不存在局部极值问题^[8],已经在许多预测领域得到应用.文献[9]分别应用最小二乘支持向量机(Least square support vector machine, LSSVM)、多变量自适应回归样条函数(Multivariate adaptive regression splines, MARS)和M5模型树(M5 model tree, M5Tree)对河流水污染数据进行了长期预测建模,实验结果表明,LSSVM和MARS在长期预测精度上基本接近,均好于M5Tree;文献[10]提出了一种基于平滑支持向量机回归的长期预测模型,并对飞机发动机振动数据进行了长期预测;文献[11]提出了一种基于遗传算法的SVM模型,并对水力发电站的水流量数据进行了长期预测,该模型通过采用快速收敛率的遗传算法对参数进行优化,提高了计算速度和预测精度.在基于LSSVM的长期预测模型中,核函数的选取是其中的重要一环,核函数选取的结果对最终的分类或预测结果有着重要影响^[12].然而,上述文献的LSSVM核函数基本都是浅层结构的核函数,对含高噪声的复杂问题建模能力具有一定的局限性.

本文提出一种基于深度去噪核映射的最小二乘支持向量机长期预测模型,用深度去噪核函数替代浅层核函数.根据深度学习的思想,对深度核函数网络进行逐层预训练后对网络整体进行微调,进而优化网络参数.为了使深度核函数网络能够应对高噪声数据,提出将去噪算法融入到深度核函数网络,进而形成深度去噪核函数网络.该深度核结构能够提取数据中的深度鲁棒特征,进而提高最小二乘支持向量机的预测精度.实际数据仿真实验表明,本文所提方法能够有效提取数据中蕴含的特征信息,从而提高预测模型的精度.

1 基于深度去噪核映射的最小二乘支持向量机长期预测模型

1.1 最小二乘支持向量机

LSSVM是一种基于结构风险最小化原理的机器学习算法^[13].对于给定的训练样本集, $D = \{(x_i,$

$y_i) | i = 1, 2, \dots, n\}$.其中: $x_i \in R^m$ 是 m 维的输入, $y_i \in R$ 是一维输出.LSSVM方法相当于求解如下的优化问题:

$$\min_{w,b,e} J(w, e) = \frac{1}{2}w^T w + \frac{\gamma}{2} \sum_{i=1}^n e_i^2; \quad (1)$$

$$\text{s.t. } y_i = w^T \varphi(x_i) + b + e_i. \quad (2)$$

其中: w 和 b 为待求的参数, $\varphi(x_i)$ 为映射函数, γ 为惩罚系数, n 为训练样本个数, e_i 为每一个训练样本的拟合误差.引入拉格朗日乘子 α ,将约束优化问题变为无约束优化问题

$$L(w, b, e, \alpha) = J(w, e) - \sum_{i=1}^n \alpha_i (w^T \varphi(x_i) + b + e_i - y_i). \quad (3)$$

根据KKT条件,对式(3)求导可得下列方程组:

$$\begin{bmatrix} 0 & 1_n^T \\ 1_n & K + \frac{1}{\gamma} E_n \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}. \quad (4)$$

其中: E_n 为 n 阶单位矩阵, $1_n = [1, \dots, 1]^T$, $K(x_i, x_j) = \varphi^T(x_i)\varphi(x_j)$ 为核函数.

由式(4)可以求解出参数 b 和 α ,进而可以得到最小二乘支持向量机的预测模型如下:

$$f(x) = \sum_{i=1}^n \alpha_i K(x, x_i) + b. \quad (5)$$

1.2 深度去噪核网络

针对浅层结构核函数存在提取样本数据特征不充分、对复杂问题建模能力有限及数据中的高噪声会导致模型泛化能力不足等问题^[14],提出深度去噪核函数,它是一种深层结构的核函数网络,如图1所示.

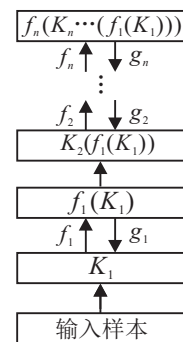


图1 深度核结构

由图1可知,所提的深度去噪核网络由多个浅层去噪核经过非线性映射叠加而成.其训练过程分为两个步骤,第1步是预训练每一层的浅层去噪核网络,第2步是微调整个深度去噪核网络,进而得出所提的深度去噪核函数.

为了使网络能够更好地应对高噪声数据,通过对训练数据加入噪声,然后强迫网络拟合真正的没有被噪声污染过的输入,构成去噪网络,以学习去除噪声的能力来获得输入数据更加鲁棒的表达,提高其泛化能力. 具体过程如下:通过一个随机的映射变换 q_D 对原始输入数据 x 加入噪声,从而得到一个含有噪声的训练样本数据 \tilde{x} . 将含有噪声的训练样本数据 \tilde{x} 输入到网络中,其隐含层输出为 $y = f_\theta(\tilde{x})$,输出层输出为 $z = g_{\theta'}(y)$. 其中: f_θ 和 $g_{\theta'}$ 为相应的映射函数, θ 和 θ' 为相应的参数. 通过最小化网络输出 z 与不含噪声的原始数据 x 之间的误差,即最小化 z 与 x 之间的代价函数 $L_H(x, z)$ 来求取网络参数 θ 和 θ' . 去噪过程如图2所示.

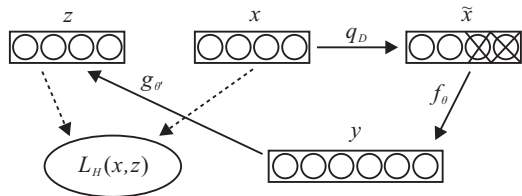


图2 去噪过程

1.2.1 深度核网络的预训练

为了降低深度结构的训练复杂度,对深度结构的每一层参数依次进行预训练,其训练过程如下:

- 1) 将训练样本数据 U 代入核函数中,得到其核函数值 $K(U)$.
- 2) 提取核函数值 $K(U)$ 的稀疏特征. 采用稀疏自编码神经网络^[15]对如图3所示的网络结构进行训练. 其中: $\{x_1, x_2, \dots, x_h\} = K(U), \{x'_1, x'_2, \dots, x'_h\} = \{x_1, x_2, \dots, x_h\}$.

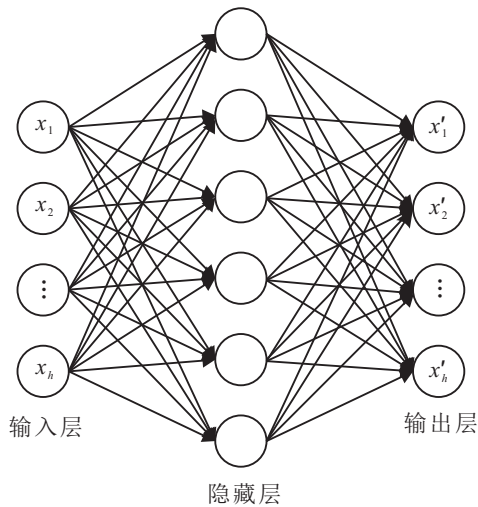


图3 稀疏自编码神经网络结构

- 3) 训练好网络参数后,将样本数据 $\{x_1, x_2, \dots, x_h\}$ 进行一次前向传播,计算隐藏层的激活值作为下

一层网络的输入样本.

- 4) 从底层至顶层,依次对如图1所示的深度核结构进行上述计算,得到深度核网络的预训练参数.

1.2.2 深度去噪核网络的建立及微调

在经过自下而上的逐层预训练后,所建立的深度去噪核网络能够提取数据的深度特征,对含高噪声的复杂数据具有较好的建模效果. 为了进一步提升网络的性能,对整个网络参数进行进一步微调,在调整过程中不断减小原始数据和网络重构数据之间的重构误差. 具体过程是将重构误差经过逐层的反向传播,利用梯度下降算法计算更新各层网络权值及偏置,直至满足最小误差要求. 因为网络在微调之前已经通过逐层预训练得到了较好的参数初始值,所以有效克服了传统神经网络因随机初始化参数而容易陷入局部最小和训练时间过长的缺点^[16],大大缩减了参数寻优的收敛时间. 其具体步骤为: 1) 前向传播. 将训练数据样本输入到网络的输入层,并由下向上逐层计算每层的输出直到最顶层. 2) 反向传播. 将网络最顶层的实际输出值与相对应的理想输出值之间的误差函数作为代价函数,如下式所示:

$$J = \frac{1}{m} \sum \left(\frac{1}{2} \|Y - Z\|^2 \right). \quad (6)$$

其中: m 为样本个数, Z 为网络的输出值, Y 为真实值. 运用梯度下降算法,通过调整权值以减小代价函数的值直至满足最小误差要求,以达到网络微调的目的.

在调整好网络的所有参数后,深度去噪核函数可以表示为

$$K_{\text{deep}} = f_n(K_n \cdots (f_1(K_1))) = \sigma(W_n(K_n \cdots (\sigma(W_1 K_1 + b_1))) + b_n). \quad (7)$$

其中: f 为映射函数, σ 为激活函数, W 为权值矩阵, b 为偏置项.

1.3 算法流程

为了使所提的基于深度去噪核映射的最小二乘支持向量机的长期预测模型结构更易于理解,将其算法结构表示如下.

算法1 基于深度去噪核映射的最小二乘支持向量机的长期预测算法.

输入: 训练集 $D = \{(x_i, y_i) | i = 1, \dots, n\}$, 测试集 $S = \{z_j | j = 1, \dots, p\}$, 深度去噪核映射最小二乘支持向量机, 网络层数 L ;

输出: 测试集的预测结果 $Q = \{c_u | u = 1, \dots, p\}$.

- 1) 初始化深度去噪核网络的权值 W 和偏置 b , 选

定核函数 K ;

- 2) 根据训练集输入样本 x_i 计算核函数 K 的值;
- 3) For $l = 1$ to L ;
- 4) 应用随机映射在 K^l 中添加随机噪声产生样本 K_t^l ;

5) For epoch = 1 to maxepoch do;

6) 前向传播, 计算激活值

$$a^l = f(K_t^l) = \sigma(W_1^l K_t^l + b_1^l),$$

$$c^l = g(a^l) = \sigma(W_2^l a^l + b_2^l);$$

7) 反向传播, 计算误差

$$\delta_3^l = -(K^l - c^l) \cdot g'(W_2^l a^l + b_2^l),$$

$$\delta_2^l = \left(\left(\sum_{j=1}^{s_2} W_{j2}^l \delta_3^l \right) + \beta \left(-\frac{\rho}{\hat{\rho}_2} + \frac{1-\rho}{1-\hat{\rho}_2} \right) \right) f'(W_1^l K_t^l + b_1^l);$$

8) 计算权值和偏置的导数

$$\frac{\partial J}{\partial W_1^l} = \delta_2^l (a^l)^T, \quad \frac{\partial J}{\partial b_1^l} = \delta_2^l,$$

$$\frac{\partial J}{\partial W_2^l} = \delta_3^l (c^l)^T, \quad \frac{\partial J}{\partial b_2^l} = \delta_3^l;$$

9) 计算

$$\Delta W_1^l = \Delta W_1^l + \frac{\partial J}{\partial W_1^l},$$

$$\Delta b_1^l = \Delta b_1^l + \frac{\partial J}{\partial b_1^l},$$

$$\Delta W_2^l = \Delta W_2^l + \frac{\partial J}{\partial W_2^l},$$

$$\Delta b_2^l = \Delta b_2^l + \frac{\partial J}{\partial b_2^l};$$

10) 更新参数

$$W_1^l = W_1^l - \varepsilon \left[\left(\frac{1}{n} \Delta W_1^l \right) + \lambda W_1^l \right],$$

$$b_1^l = b_1^l - \varepsilon \left[\frac{1}{n} \Delta b_1^l \right],$$

$$W_2^l = W_2^l - \varepsilon \left[\left(\frac{1}{n} \Delta W_2^l \right) + \lambda W_2^l \right],$$

$$b_2^l = b_2^l - \varepsilon \left[\frac{1}{n} \Delta b_2^l \right];$$

11) End;

12) End;

13) 根据训练好的参数 W_1^l 和 b_1^l 求取深度去噪核函数

$$K = f^l(K^l \cdots (f^1(K^1))) = \sigma(W_1^l (K^l \cdots (\sigma(W_1^l K^l + b_1^l))) + b_1^l);$$

14) 根据求取的核函数 K , 用训练样本训练

LSSVM, 求取参数 α 和 b ;

15) 将测试样本代入训练好的 LSSVM 预测公式

$$f(x) = \sum_{i=1}^n \alpha_i K(x, x_i) + b$$

求取预测值.

2 实际数据仿真验证及分析

为了验证所提方法的有效性, 分别选取波士顿住房数据集和某钢铁厂煤气管网系统中4个转炉煤气(Linz Donawitz converter gas, LDG)产消流量与柜位数据进行建模和验证.

采用多隐含层反向传播神经网络(Multi-hidden-layer back propagation neural network, MHLBPNN)、LSSVM、回声状态网络(Echo state network, ESN)分别建立相应的预测模型与本文所提方法进行比较. 同时选取均方根误差(Root mean square error, RMSE)和平均绝对百分比误差(Mean absolute percentage error, MAPE)作为判断预测效果的依据, 其计算公式分别为

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^T (y(t) - y_d(t))^2}, \quad (8)$$

$$\text{MAPE} = \frac{100}{T} \sum_{t=1}^T \frac{|y(t) - y_d(t)|}{y_d(t)}. \quad (9)$$

其中: T 为预测长度, $y(t)$ 为预测值, $y_d(t)$ 为真实值.

2.1 模型参数选取原则

当训练 MHLBPNN 模型时, 由于采用将输入数据归一化到 $[0, 1]$ 区间的方法, 激活函数也选用值域为 $[0, 1]$ 的 sigmoid 函数; 隐层神经元数一般选取比嵌入维数大, 以便更好地拟合数据, 针对各个具体情况采用试验法进行选取; 学习率采用经验值法进行选取, 一般选取值在 $0.6 \sim 0.99$ 之间. 由于是多隐含层网络, 隐含层层数要大于 1, 其隐含层层数采用构造法进行选取. 当训练 LSSVM 模型时, 其核函数根据传统选取方法一般选取高斯核函数; 高斯核宽度根据各个数据的具体情况采用试验法进行选取; 惩罚系数采用经验值法加试验法进行选取. 当训练 ESN 模型时, 储备池稀疏度采用经验值法进行选取, 一般选取在 $0.5 \sim 1$ 之间, 其值越大, 非线性逼近能力越强; 储备池的神经元数根据样本个数选取, 储备池规模越大, ESN 对给定动态系统的描述越准确, 但是会带来过拟合问题, 根据具体的样本个数, 采用经验值法和实验法确定各个神经元的具体个数; 储备池内部连

接权谱半径为连接权矩阵的绝对值最大的特征值,采用经验值法加试凑法进行选取. 当训练本文所提模型时,核函数、高斯核宽度、惩罚系数的选取方法同LSSVM模型. 由于是深度网络结构,深度核层数要大于1层,具体采用构造法进行选取.

2.2 波士顿住房数据集

所选取的波士顿住房数据集^[17]是一个测试模型性能的标准数据集. 这个数据集主要是用来评价波士顿郊区的住房价值,样本个数为506,包含与房屋价值有关的13个连续属性和1个二值属性,具体如下:CRIM:城镇人均犯罪率;ZN:居住用地超过25 000平方英尺的比例;INDUS:每个城镇非零售商务所占比例;CHAS:查尔斯河虚拟变量,判断房屋土地是否临河,临河设为1,否则为0;NOX:氮氧化物浓度;RM:每个公寓房屋数目的平均值;AGE:1940年之前建造的业主自住单位楼所占的比例;DIS:到波士顿5个就业中心的加权距离;RAD:通往高速公路的通行指数;TAX:每10 000美金的全部价值财产税税率;PTRATIO:城镇中教师的比率;B: $1\ 000(B_k - 0.63)^2$,这里 B_k 表示城镇中黑人所占的比例;LSTAT:低地位人群所占百分比;MEDV:1 000美金以内的业主自住房价中位数. 问题可描述为通过房屋的属性信息,决策者对其房屋价值进行评价. 选取其中的300个样本作为训练样本,200个样本作为测试样本.

2.2.1 各个模型参数的具体选取

根据模型参数选取原则,结合具体的波士顿住房数据集集中的数据,其各个模型的参数选取如表1所示.

表1 模型参数选取

预测方法	参数	值
MHLBPNN	激活函数	sigmoid
	隐层神经元数	40
	学习率	0.85
	隐含层层数	3
LSSVM	核函数	高斯核
	高斯核宽度	75
	惩罚系数	27
ESN	储备池稀疏度	0.83
	储备池神经元数	87
	谱半径	0.73
本文方法	核函数	高斯核
	高斯核宽度	78
	惩罚系数	30
	深度核层数	3

2.2.2 预测曲线对比

为了将本文所提方法与其他3种方法的预测结果作直观对比,图4所示为4种预测方法的预测结果曲线.

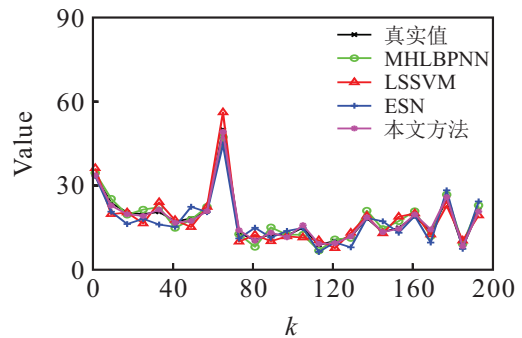


图4 4种方法预测结果曲线

从图4可以看出:LSSVM和ESN方法的预测曲线在某些区间段上发生了偏离真实值较大的情况;MHLBPNN预测方法的预测结果曲线总体上能够跟踪真实值曲线的上升、下降等大趋势,然而在某些点上,其预测误差仍然较大. 本文所提方法的预测曲线较其他3种预测方法最接近真实值曲线,预测误差较大的点也最少.

为了对4种预测方法的预测值和真实值作完整的对比统计分析,给出预测值与真实值的散点图如图5所示.

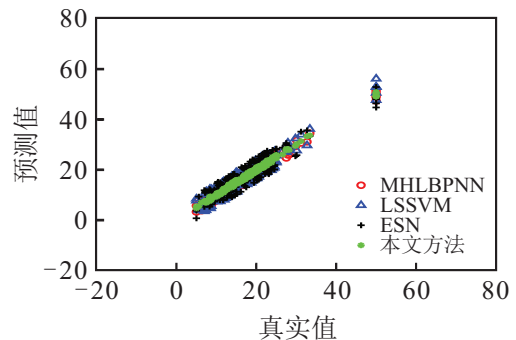


图5 4种方法的预测值与真实值的散点图

从图5可以看出:本文所提方法的预测值较其他3种预测方法更接近直线 $y = x$,基本没有偏离程度较大的点,说明预测效果最好;其他3种对比方法的预测值都有不同程度偏离对角线的情况,其中尤以LSSVM和ESN方法最为突出,有相当一部分点离对角线距离较大,而MHLBPNN方法也有部分点远离了直线 $y = x$.

2.2.3 预测误差对比

为了更进一步地量化各个方法的预测精度,将4种方法预测结果的RMSE和MAPE列于表2中.

表2 4种预测方法的预测误差对比

误差	MHLBPNN	LSSVM	ESN	本文方法
RMSE	1.657	2.444	2.893	0.577
MAPE	0.099	0.146	0.166	0.032

由表2中的对比数据可知,LSSVM和ESN方法的RMSE和MAPE都较大,MHLBPNN方法的预测误差虽然小于这2种方法,但是也大于本文所提方法.因此,所提方法的预测精度最高.

2.3 钢铁厂煤气管网系统数据

所选取的某钢铁厂煤气管网系统中的4个转炉煤气柜位受到系统中很多用户因素的影响,可选取其中影响最大的几个因素作为引起柜位变化的主要因素,如转炉LDG回收量、煤气柜位、热轧、冷轧、发电厂、低压锅炉.这些主要因素中总共包含18个具体影响因素,将其作为模型的输入.选取数据的个数为10000个,以min为单位,将数据中的前8000个数据作为训练样本,剩下的2000个数据作为测试样本,预测长度为200个点.之所以选取这些数据,是因为LDG系统数据在煤气管网系统中比较有代表性,可以验证所提方法在工业现场应用的有效性.

2.3.1 各个模型参数的具体选取

根据模型参数选取原则,结合具体的转炉煤气系统数据,其各个模型针对不同柜的参数选取值如表3所示.

表3 模型参数选取

预测方法	参数	1号柜	2号柜	3号柜	4号柜
MHLBPNN	激活函数	sigmoid	sigmoid	sigmoid	sigmoid
	隐层神经元数	43	40	46	45
	学习率	0.87	0.8	0.84	0.83
	隐含层数	4	4	4	4
LSSVM	核函数	高斯核	高斯核	高斯核	高斯核
	高斯核宽度	79	82	80	77
	惩罚系数	30	24	28	29
ESN	储备池稀疏度	0.8	0.85	0.83	0.88
	储备池神经元数	90	95	83	100
	谱半径	0.75	0.75	0.75	0.75
本文方法	核函数	高斯核	高斯核	高斯核	高斯核
	高斯核宽度	90	88	95	87
	惩罚系数	32	38	27	30
	深度核层数	4	4	4	4

2.3.2 预测曲线对比

为了将本文所提方法与其他3种方法的预测结果作一直观对比,4种预测方法对4个LDG柜位的预

测结果如图6所示.

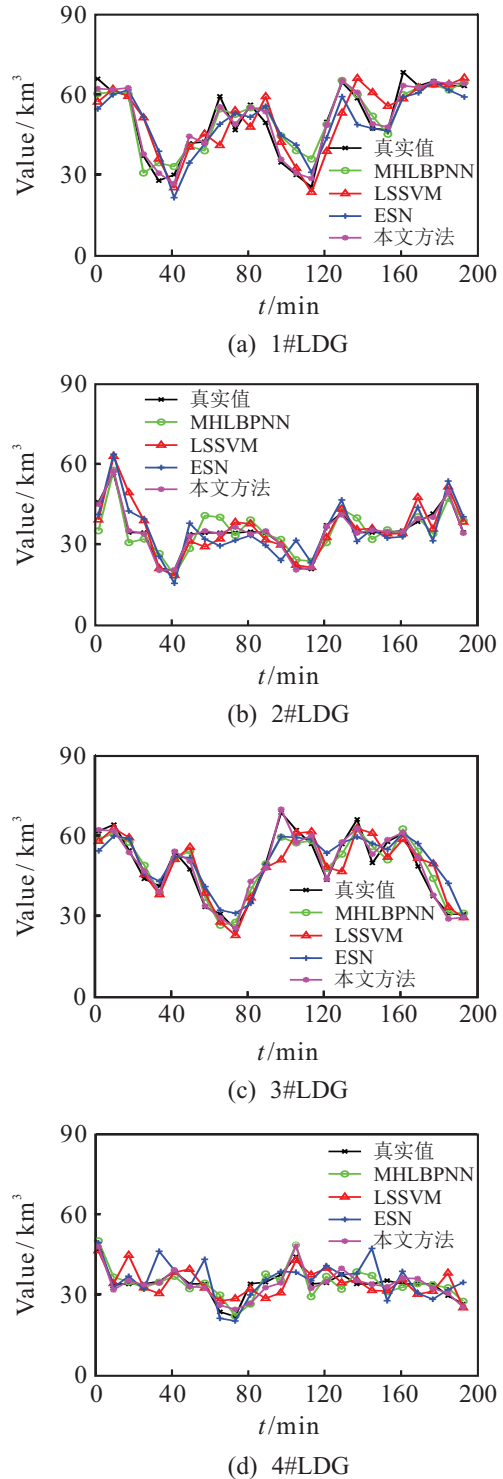


图6 4种方法对4个LDG柜位预测图

由图6中的预测对比曲线可知,对于4个LDG柜位值预测,本文所提方法的预测效果都是最好的,能够准确地跟踪真实值的上升、下降等趋势,且从整体而言,偏离真实值曲线的程度最低,较大偏离真实值的点也最少.在图6(a)中,LSSVM的预测曲线波动较大,在某些时间段上偏离真实值曲线程度较高,而MHLBPNN和ESN的预测曲线基本能够跟踪真实值曲线的趋势,其中MHLBPNN的预测曲线较

ESN的预测曲线更接近真实值曲线,但准确程度不及本文方法. 在图6(b)中,LSSVM的预测曲线要比ESN的预测曲线偏离真实值曲线的波动程度小,但不及MHLBPNN方法,而本文方法好于MHLBPNN方法. 在图6(c)中,MHLBPNN的预测曲线接近真实值曲线的程度紧次于本文方法,好于LSSVM和ESN方法. 在图6(d)中,ESN方法的预测曲线上偏离真实值的点最多,个别点偏离程度较大,LSSVM预测曲线虽然偏离真实值点的个数较ESN少,但多于MHLBPNN方法,而本文方法的预测曲线上偏离真实值的点最少.

为了对本文方法对于4个LDG柜位值的预测结果作完整的对比统计分析,给出预测值与真实值的散点图如图7所示.

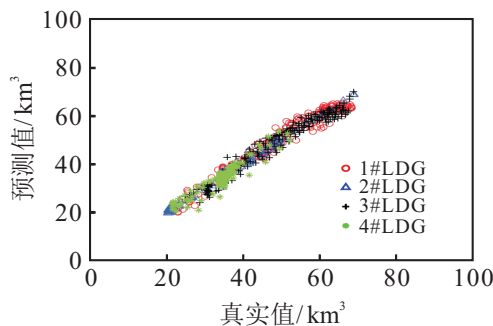


图7 预测值与真实值的散点图

从图7可以看出,本文方法对4个LDG柜位的预测值基本都分布在直线 $y = x$ 两侧,说明预测精度较高. 其中:2#LDG的预测值与对角线接近的程度最高,且基本很少有偏离程度较大的点;对于1#LDG的预测值有个别的点发生了较大程度偏离对角线的情况,而大多数的点还是紧密分布在对角线两侧;虽然3#LDG和4#LDG也有极个别的预测值偏离了真实值,但总体而言预测效果较好.

2.3.3 预测误差对比

为了更进一步地量化各个方法的预测精度,将4种方法预测结果的RMSE和MAPE列于表4中.

表4 4种预测方法的预测误差对比

柜位	误差	MHLBPNN	LSSVM	ESN	本文方法
1#LDG	RMSE	5.894	8.035	7.370	2.614
	MAPE	0.101	0.156	0.139	0.046
2#LDG	RMSE	4.450	5.242	5.656	1.015
	MAPE	0.098	0.106	0.124	0.024
3#LDG	RMSE	4.640	6.453	5.720	2.577
	MAPE	0.087	0.147	0.106	0.041
4#LDG	RMSE	3.437	5.214	6.047	1.945
	MAPE	0.081	0.125	0.159	0.038

从表4中的对比数据可知:LSSVM、ESN这2种方法对4个LDG柜位值预测效果并不相同,对于1#LDG数据,ESN预测误差小于LSSVM预测误差;对于2#LDG数据,LSSVM预测误差小于ESN预测误差;对于3#LDG数据,LSSVM预测误差大于ESN预测误差;对于4#LDG数据,LSSVM预测误差小于ESN预测误差. LSSVM和ESML这2种方法对不同数据的建模能力有较大差别,说明它们的普适性较差. 虽然MHLBPNN方法对4个LDG柜位值的预测精度均高于LSSVM和ESN方法,但是本文方法在这4种数据下预测结果的RMSE和MAPE都明显低于其他3种方法,这说明了本文方法的有效性更好.

2.4 结果分析

根据波士顿住房数据集和4个转炉煤气柜位数据的预测曲线和预测误差对比结果,通过分析可知,ESN是利用储备池来存储信息,其内部神经元连接权值是固定的,通过计算从储备池状态空间到输出节点线性回归所需要的输出权值使训练误差达到最小,克服了局部最优问题,然而ESN模型对于含噪声较高的数据建模能力较差,预测精度较低. LSSVM模型虽然能够克服局部最优值、训练时间长、计算复杂度高、训练结果存在随机性以及过学习等问题,然而它和ESN一样,都是仅含单层非线性变换的浅层学习结构,其共性是仅含单个将原始输入信号转换到特定问题空间特征的简单结构. 浅层结构的局限性在于有限的样本和计算单元情况下对复杂函数的表示能力有限,对初始值选取敏感,在处理高噪声多复杂性的数据问题时其泛化能力受到一定制约,模型效果有限,因而它们的预测结果较差. 而深层网络结构是将数据经过多个隐含层的特征变换,提取出最能够刻画数据本质的特征,因而能够拟合高度非线性的数据,进而提高数据预测的准确性. MHLBPNN模型属于多隐含层的深层网络结构,因而预测精度高于LSSVM和ESN这样的浅层网络模型. 然而,MHLBPNN模型的基本组件是BPNN模型,由于BPNN模型神经元之间的连接形式固定,整体网络性能受隐藏层神经元的影响较大,容易过拟合,限制了网络的泛化能力,最终导致预测精度降低. 而本文方法是将深度学习思想融入到LSSVM预测模型中的深层网络结构,既保留了LSSVM预测模型的优点,又通过深度核结构发挥了深度网络的优势,摒弃了多隐含层网络的缺点,同时通过对网络模型加入去噪处理,提高了网络对含高噪声数据的拟合能力,进

一步提高了预测精度. 因而, 本文方法的预测精度明显好于MHLBPNN、LSSVM和ESN方法.

3 结论

针对最小二乘支持向量机对含高噪声数据的预测能力有限的问题, 本文根据深度学习的思想, 从提高核函数映射能力的角度, 提出了基于深度去噪核映射的最小二乘支持向量机长期预测模型. 所提模型采用深度核网络代替传统最小二乘支持向量机中的浅层核结构, 通过深度核网络的逐层特征映射, 提取数据样本的深度特征, 进而提高模型的建模能力, 解决了基于浅层核结构模型预测精度不高的问题. 为了能够更好地处理高噪声数据, 提出将数据加入随机噪声并让模型拟合未加入噪声的原始数据来提高模型的去噪能力. 在通过逐层预训练网络后, 提出采用反向传播算法来对网络整体进行微调, 进一步提高了整个网络的建模能力. 最后采用本文方法对标准数据集和工业系统中的实际数据进行预测, 通过与其他几种预测方法的对比实验结果表明了本文方法对于提高预测精度的有效性.

参考文献(References)

- [1] Wang K, Zhuo L C, Shao Y, et al. Toward distributed data processing on intelligent leak-points prediction in petrochemical industries[J]. *IEEE Trans on Industrial Informatics*, 2016, 12(6): 2091-2102.
- [2] Giorgi M G D, Ficarella A, Tarantino M. Error analysis of short term wind power prediction models[J]. *Applied Energy*, 2011, 88(4): 1298-1311.
- [3] Chang W Y. Short-term wind power forecasting using the enhanced particle swarm optimization based hybrid method[J]. *Energies*, 2013, 6(9): 4879-4896.
- [4] Guan C, Luh P B, Michel L D, et al. Hybrid Kalman filters for very short-term load forecasting and prediction interval estimation[J]. *IEEE Trans on Power Systems*, 2013, 28(4): 3806-3817.
- [5] Lauret P, Fock E, Randrianarivony R N, et al. Bayesian neural network approach to short time load forecasting[J]. *Energy Conversion and Management*, 2008, 49(5): 1156-1166.
- [6] Guo W, Xu T, Lu Z L. An integrated chaotic time series prediction model based on efficient extreme learning machine and differential evolution[J]. *Neural Computing & Applications*, 2016, 27(4): 883-898.
- [7] Zhou J, Li X B, Shi X Z. Long-term prediction model of rockburst in underground openings using heuristic algorithms and support vector machines[J]. *Safety Science*, 2012, 50(4): 629-644.
- [8] Suykens J A K, Gestel T V, Brabanter J D, et al. *Least squares support vector machines*[M]. Singapore: World Scientific Publishing Co, 2002: 1-27.
- [9] Kisi O, Parmar K S. Application of least square support vector machine and multivariate adaptive regression spline models in long term prediction of river water pollution[J]. *J of Hydrology*, 2016, 534: 104-112.
- [10] Jin X, Sun Z, Wang H, et al. Application of improved support vector machine regression analysis for medium-and long-term vibration trend prediction[J]. *J of Vibroengineering*, 2013, 15(2): 942-950.
- [11] Xu Y, Wang J Y, Su H Y. Application of support vector machine based on genetic algorithm to mid-term and long-term run-off prediction[J]. *J of Water Resources & Architectural Engineering*, 2014, 12(5): 42-45.
- [12] Jebarani E S, Kumar S S, Jayakumar J. Torque modeling of switched reluctance motor using LSSVM-DE[J]. *Neurocomputing*, 2016, 211: 117-128.
- [13] Mall R, Suykens J A K. Very sparse LSSVM reductions for large-scale data[J]. *IEEE Trans on Neural Networks & Learning Systems*, 2015, 26(5): 1086-1097.
- [14] Zhang X. White noise testing and model diagnostic checking for functional time series[J]. *J of Econometrics*, 2016, 194(1): 76-95.
- [15] Xu J, Xiang L, Liu Q S, et al. Stacked sparse autoencoder(SSAE) for nuclei detection on breast cancer histopathology images[J]. *IEEE Trans on Medical Imaging*, 2016, 35(1): 119-130.
- [16] Zhang L, Wang F, Sun T, et al. A constrained optimization method based on BP neural network[J]. *Neural Computing & Applications*, 2016, 7(11): 1-9.
- [17] Harrison D, Rubinfeld D L. Hedonic housing prices and the demand for clean air[J]. *J of Environmental Economics and Management*, 1978, 5(1): 81-102.

(责任编辑: 齐 霁)