计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.200700164

基于时空优化的多尺度卷积神经网络空气质量预测模型

周　杰[1]　罗云芳[1]　雷耀建[2]　李文敬[3]　封　宇[1]
1 广西职业技术学院机电与信息工程学院　南宁 530226
2 广西幼儿师范高等专科学校　南宁 530022
3 南宁师范大学物流管理与工程学院　南宁 530001
(779489658@qq.com)

摘　要　目前针对空气质量数值预测多采用单一站点时间序列特征进行浓度预测,没有考虑空气质量数值的变化受空间特征的影响。针对该问题提出一种基于时空优化的多尺度神经网络(MSCNN-GALSTM)模型用于空气质量预测,利用一维多尺度卷积核(MSCNN)提取空气质量数据中的局部时间关系和空间特征关系,并进行线性拼接融合,得出多站点多特征的相互时空特征关系,结合长短记忆网络(LSTM)处理时间序列的优势,并引入遗传算法(GA)对LSTM网络的参数集进行全局寻优,把多站点多特征的相互时空关系输入至LSTM网络中,进而输出多站点多特征的长期特征依赖关系。最后将MSCNN-GALSTM模型与单一LSTM基准模型和单尺度卷积神经网络模型进行对比,均方根误差(RMSE)下降约11%,平均预测准确率提升约20%。实验结果表明,MSCNN-GALSTM预测模型在空气质量数据预测中特征提取更加全面、层次更深,预测精度更高,并且表现出了更好的泛化能力。

关键词:多尺度卷积;空间特征;时空优化;LSTM;预测模型
中图法分类号　TP183

Multi-scale Convolutional Neural Network Air Quality Prediction Model Based on Spatio-Temporal Optimization

ZHOU Jie[1],LUO Yun-fang[1],LEI Yao-jian[2],LI Wen-jing[3] and FENG Yu[1]
1 College of Mechanical,Electrical and Information Engineering,Guangxi Vocational & Technical College,Nanning 530226,China
2 Guangxi College for Preshcool Education,Nanning 530022,China
3 School of Logistics Management and Engineering,Nanning Normal University,Nanning 530001,China

Abstract　At present,the air quality prediction is mainly based on the time series of a single station,without considering the influence of the spatial characteristics on the air quality. To solve this problem,a multi-scale neural network (MSCNN-GALSTM) model based on spatiotemporal optimization is proposed for air quality prediction. one-dimensional multi-scale convolution kernel (MSCNN) is used to extract the local temporal and spatial characteristic relations in air quality data,the LINEAR SPLICING and fusion are carried out to obtain the space-time characteristic relation of multi-sites and multi-features,combine the advantage of long-short memory network (LSTM) to process time series,and introduce genetic algorithm (Ga) to optimize the parameter set of LSTM network globally,the time-space relationship of multi-site and multi-feature is input into the LSTM network,and then the long-term feature dependence of multi-site and multi-feature is output. Finally,the MSCNN-GALSTM model was compared with the single LSTM reference model and the single scale convolutional neural network model. The root mean square error (RMSE) decreased by about 11% and the average prediction accuracy increased by about 20%. The results show that the MSCNN-GALSTM model has more comprehensive feature extraction,deeper level,higher prediction accuracy and better generalization ability.

Keywords　Multiscale convolution,Spatial features,Spatio-temporal optimization,LSTM,Prediction model

1　引言

空气污染对个人的生活、工作、学习等造成了负面影响,我国在该问题上投入了大量的精力,以加强对空气质量数据的监控与预测。人们若能预知未来空气质量的情况,将有利于人们对空气污染进行防控。在当前的空气质量预测领域研

基金项目:国家自然科学基金项目(61866006);广西教育厅自然科学基金项目(2019ITA01002,2018KY0951,2019KY1220,2017KY0980);广西职业技术学院自然科学类课题(桂职院(2019)176号 191202)
This work was supported by the National Natural Science Foundation of China(61866006),Natural Science Foundation of Guangxi Education Department(2019ITA01002,2018KY0951,2019KY1220,2017KY0980) and Natural Science Project of Guangxi Vocational & Technical College (2019-176-191202).
通信作者:罗云芳(123377307@qq.com)

究中，众多的研究者采用人工特征选择的方法对空气中所存在的污染物进行预测，虽然在此领域中取得了一定的成绩，但目前预测模型存在的不足是工作效率低、研究人员对空气质量数据处理不当等问题，从而会影响到最终的预测结果。

一些研究表明，空气质量数据存在时间序列问题，国内外学者从这一角度出发，对空气质量进行预测分析。Gamze等[1]采用季节性自回归移动平均模型(SARIMA)用于预测SO2空气质量指数。实验结果显示，SARIMA模型为空气质量参数提供了可靠且令人满意的预测精度。Rahman等[2]提出了人工神经网络时间序列模型ANNs，用于预测2000—2009年3个不同站点的空气污染指数(API)。实验结果显示，ANNS预测模型比季节自回归移动平均模型(SARIMA)、传统人工神经网络(ANN)的预测效果更好，在性能指标上也表现得更优异。Biancofiore等[3]采用RNN循环网络与无循环网络模型来预测PM2.5浓度。通过对比实验，RNN循环网络的预测精度较高，证明了循环网络在PM2.5浓度预测上有着较大的发展空间。Song等[4]提出一种基于LSTM循环网络和Kalman滤波的空气质量预测模型(LSTM-Kalman)，用于预测具有长期和短期特征的时间序列数据。其中，LSTM循环网络用特有的记忆特性来“存储”时间序列信息，然后通过Kalman滤波动态调整LSTM循环网络处理得到的时间序列信息，最终对空气质量进行预测。实验结果显示，LSTM-Kalman预测模型比单一LSTM预测模型的性能更好。Chiou-Jye等[5]采用卷积神经网络(CNN)提取空气质量数据的时间特征关系，并结合LSTM神经网络(CNN-LSTM模型)，用于预测未来PM2.5浓度，最终该模型的预测效果优于单一的CNN、LSTM预测模型，但它们均没有考虑到空间特征的关联性，可能还有一定的提升空间。

2 问题描述

针对以上问题，采用卷积神经网络提取空气质量数据中的局部时间和空间两个方面的特征，但通常卷积操作采用的是大小相同的卷积核对数据进行特征提取，这样的方法会在一定程度上使网络达到瓶颈状态。本文在此基础上进行改进，采用多尺度(Multi-scale)的卷积核(MSCNN)进行特征提取，再将多尺度卷积核所提取的特征进行拼接融合，使网络具有更强的泛化能力和鲁棒性。

将MSCNN提取的时空特征关系输入至LSTM网络中，进而提取空气质量数据的长期依赖关系，即MSCNN-LSTM预测模型，但LSTM网络中存在大量的初始参数需要调整，且连接权值及偏置很难有效确定，在此引入了遗传算法(GA)对LSTM网络的参数集进行全局寻优，最终得到了MSCNN-GALSTM空气质量预测模型[23]。

3 理论与模型

3.1 MSCNN空气质量数据时空特征提取

卷积神经网络已成功应用于图像识别方向，验证了该网络对特征图的特征提取能力具有强大的效果，而本文的应用场景是对城市未来1天PM2.5的预测，需要对空气质量数据及气象因素进行时空特征的提取。本文对数据集进行分析发现，数据集中包含多站点多特征，且均是以数值的形式表现，而不是特征图的形式，因此本文首先对数据进行预处理，将数据的特征合并成一张特征图，最终输入至卷积神经网络进行时空特征的提取。另外，特征图通常以二维结构为输入，而我们将多站点多个特征做简单变换，构成并列多站点单特征的一维结构特征图[6-8]。

如图1所示，以构建多站点单特征(PM10)的一维结构特征图为例，某时刻多站点的数据组成PM10一维时间序列向量，不同时刻的时间序列从上到下组成多站点单特征(PM10)的一维结构特征图，如仅采用一维单尺度卷积核对特征图进行时空特征的提取，这会造成站点与站点之间的时空特征关系过于单一，且深度不够，最终导致网络陷入一个瓶颈状态，因此本文采用一维多尺度卷积核(MSCNN)对特征图进行时空特征的提取，将多尺度时空特征进行拼接融合，使得站点与站点之间的时空特征关系考虑得更加全面，这将有利于提高网络的泛化能力[23-24]。

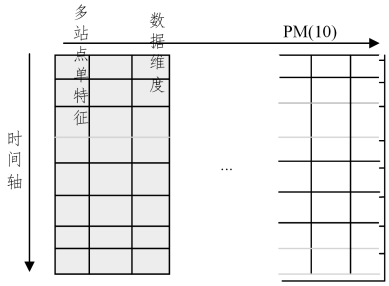

图1 多站点单特征(PM10)特征图

Fig. 1 Multi-site single feature diagram(PM10)

以单特征PM10为例，多站点PM10的时空特征提取如图2所示。其中，特征图通过一维多尺度卷积核(1*3，1*5，1*7)在数据特征轴上从左到右进行遍历完成卷积操作，步数为1，不同卷积核输出的特征向量进行拼接融合，为此得到多站点的空间特征关系。在时间轴上，随着卷积核从上到下遍历完成卷积操作，步数为1，可得到多站点随时间变化的局部趋势。最终把拼接融合的特征向量往数据特征方向合并，输出多站点PM10的时空特征。

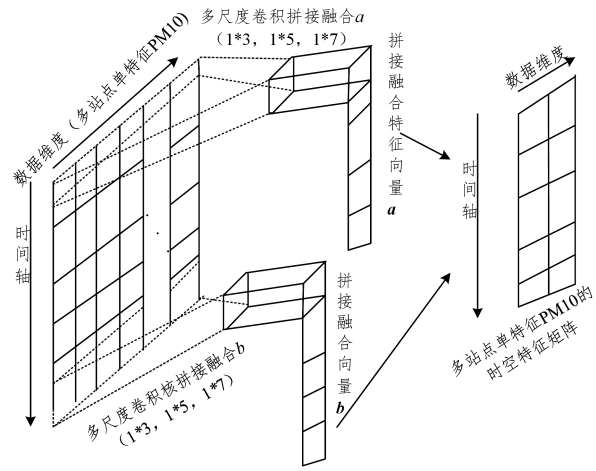

图2 一维多尺度卷积核特征提取过程图

Fig. 2 Feature extraction process of one-dimensional multi-scale convolution kernel

下文是MSCNN对特征图进行卷积操作的公式推理步骤。特征图包含P个站点提供的N个样本数据，并包含M种污染物浓度特征，那么多站点单特征n的特征图公式如下：

$$X_n=[x_n^1,x_n^2,\cdots,x_n^N]^{\mathrm{T}} \tag{1}$$

$$\boldsymbol{X}_n^{t:t+T-1}=[\boldsymbol{x}_n^t,\boldsymbol{x}_n^{t+1},\cdots,\boldsymbol{X}_n^{t+T-1}]^{\mathrm{T}} \tag{2}$$

其中，$\boldsymbol{x}_n^t=[x_{n,1}^t,x_{n,2}^t,\cdots,x_{n,p}^t]\in R^{1\times P}$ 表示 P 个站点单特征 n 在 t 时刻的向量，$\boldsymbol{X}_n^{t:t+T-1}$ 则表示 X_n 在$[t,t+T-1]$时间范围的 T 组向量，T 表示矩阵转置。

卷积操作通过权值矩阵 $\boldsymbol{W}_i$ 与 $\boldsymbol{X}_n^{t:t+T-1}$ 相乘：

1）$\boldsymbol{W}_i$ 在数据特征轴上与 $X_n^{t:t+T-1}$ 相乘，得到多站点单特征的空间特征关系。

2）$\boldsymbol{W}_i$ 在时间轴上与 $X_n^{t:t+T-1}$ 相乘，得到多站点单特征随时间变化的局部趋势。

当第 1 个卷积核在时间轴上遍历整个特征图，步数为 1 时，得到特征向量 $\boldsymbol{a}_n^i$，其大小为$[N-T+1]\times 1$，多个卷积核 V 得到的特征向量往数据特征方向合并为$[N-T+1]\times V$ 大小的 $\boldsymbol{A}_n$，$\boldsymbol{A}_n$ 代表多站点单特征的时空特征矩阵。

$$\boldsymbol{a}_n^i=[a_{t+T-1,n}^i,a_{t+T,n}^i,\cdots,a_N^i] \tag{3}$$

$$\boldsymbol{A}_n=[a_n^1,a_n^2,\cdots,a_n^V] \tag{4}$$

至此完成了多站点单特征的时空特征的提取，但数据集中还包含了其他的特征，如 PM10、CO 、O3 等共 M 种，因此，我们把 M 种特征通过以上相同的操作，即可提取并列多站点单特征的时空特征矩阵，然后把它们进行线性拼接融合，最终形成多站点多特征融合的时空特征矩阵 $\boldsymbol{A}$，如式(5)所示：

$$\boldsymbol{A}=[A_1,A_2,\cdots,A_M] \tag{5}$$

基于 MSCNN 卷积神经网络对空气质量数据时空特征进行提取，该方法对二维特征图进行简单变换，形成并列的一维特征图，这样对特征图的灵活设计，使得网络训练表现出了较好的泛化能力。同时，卷积神经网络自动提取特征的方法代替了传统的人工特征选择方法，这使得特征提取有了更全面、更深层次的效果。

3.2　MSCNN-LSTM 空气质量预测模型

MSCNN-LSTM 预测模型流程图如图 3 所示。

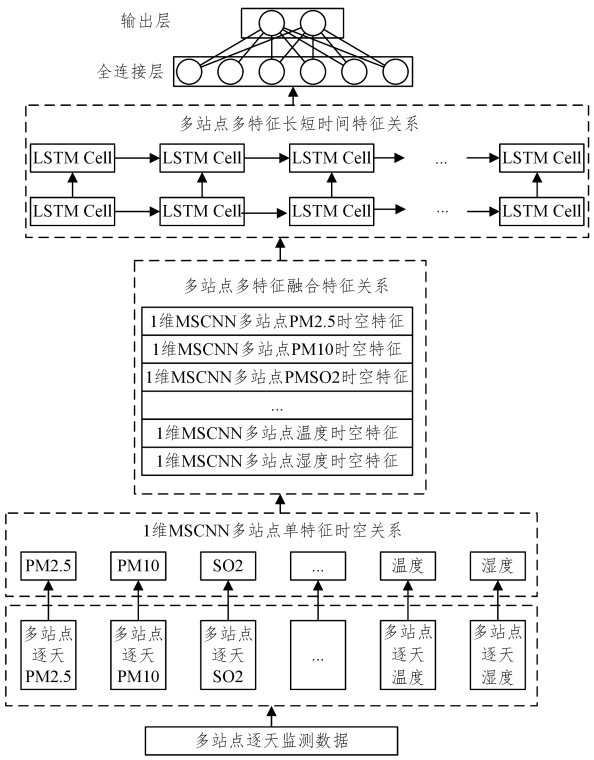

图 3　MSCNN-LSTM 预测模型流程图

Fig. 3　MSCNN-LSTM prediction model flow diagram

图 3 中，多站点逐天监测数据包括 AQI，PM2.5，PM10，SO_2，CO，NO_2，O_3 等 7 项；气象数据包括最高、平均、最低温度、最高、平均、最低湿度、最大、平均、最小风速等 9 项。根据空气质量监测数据和气象数据，选取这 15 种影响因子的数据作为整个网络预测模型的输入变量。

MSCNN-LSTM 模型由 1 维 MSCNN 多站点单特征时空关系、多站点多特征融合时空关系、多站点多特征长短时间特征关系、全连接层 4 个部分组成，每个部分在整个模型中有着关键性的作用，其作用如下：

1）多站点单特征的 1 维 MSCNN。每个站点监测数据包含多种特征，本文把它们的每一种特征分别输入至卷积神经网络中，并通过 1 维 MSCNN 提取它们每一种特征在每个站点之间的时空特征关系。

2）多站点多特征融合。将 1 维 MSCNN 提取的多站点单特征的时空特征关系进行简单的线性拼接融合，得出多站点多特征的相互时空特征关系。

3）多站点多特征长短时间特征关系。利用 LSTM 处理长短时间序列的优势，把多站点多特征的相互时空关系输入至 LSTM 网络中，进而输出多站点多特征的长期特征依赖关系。

4）全连接层。将 LSTM 网络输出的多站点多特征的长期特征依赖关系输入至全连接层，最终输出城市未来 1 天的空气质量污染物浓度。

3.3　MSCNN-GALSTM 预测模型流程图

首先对 MSCNN-GALSTM 预测模型的参数进行随机初始化，对 LSTM 网络中的参数集进行实数化编码并定义适应度函数，接着对种群进行个体适应度的计算并对个体进行选择、交叉、变异操作，如果未达到目标优化要求，遗传算法继续迭代训练，否则输出网络最优参数，进而对预测模型进行训练，计算误差值，调整最优权值及阈值，如未达到结束条件，则网络继续训练，否则输出最优模型，得出空气质量数据中多站点多特征的长期依赖的时空特征关系，最终将得到的时空特征关系输入到全连接层进行空气质量预测。MSCNN-GALSTM 预测模型的流程图如图 4 所示。

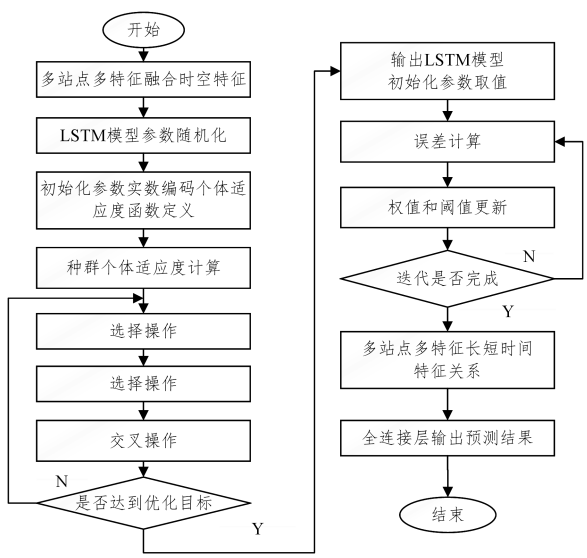

图 4　MSCNN-GALSTM 预测模型的流程图

Fig. 4　MSCNN-GALSTM prediction model flow diagram

4 仿真实验及结果分析

4.1 性能评估指标

通过评价指标来判断网络模型性能的好坏，我们采用以一致性指数(IA)、均方根误差(RMSE)以及平均绝对误差(MAE)作为性能评价指标。其中，一致性指数公式如下：

$$IA=1-\frac{\sum_{i=1}^{n}(y_i'-y_i)^2}{\sum_{i=1}^{n}(|y_i'-\bar{y}|+|y_i-\bar{y}|)^2} \tag{6}$$

均方根误差计算公式如下：

$$RMSE=\sqrt{\frac{\sum_{i=1}^{n}(y_i'-y_i)^2}{n}} \tag{7}$$

平均绝对误差计算公式如下：

$$MAE=\frac{1}{n}\sum_{i=1}^{n}|y_i-y_i'| \tag{8}$$

以上 3 个公式中，n 为数据个数，y_i 为第 i 天空气污染指数真实值，x_i'为第 i 天空气污染指数的预测值，y 为真实值的平均值。3 个指标均能反映模型的整体预测性能，*RMSE* 和 *MAE* 值越小表示预测误差越小，而 *IA* 值的取值范围在 0～1 之间，计算结果越接近 1 说明拟合效果越好。

4.2 超参数调整实验

以预测南宁市未来 1 天的 PM2.5 浓度为例，对 MSCNN-LSTM 预测模型进行超参数调整，最终确定最优的超参数。

MSCNN-LSTM 预测模型中，LSTM 网络隐藏层层数及节点数的合理设定，有利于提升网络的性能。节点数少时，网络可能无法学习到样本的关联特征；节点数多时，造成节点出现冗余，计算资源浪费，网络参数增加，从而使得网络训练时间长。我们根据经验初步确定隐藏层的层数和节点数，选择参数值较好的情况进行训练。实验采用隐藏层的层数分别为 1～4 层，且隐藏层神经元数目每层为 40，分别进行实验。实验通过一致性指标(IA)进行对比，由表 1 可知，随着隐藏层层数增加，网络模型性能并没有得到很大的提升空间，隐藏层为 2 时网络性能最优。最终我们确定 MSCNN-LSTM 网络模型中 LSTM 网络隐藏层的层数为 2。

表 1 不同隐藏层相同神经元的误差对比

Table 1 Error comparison of the same neuron in different hidden layers

层数	神经元	*IA*
1	40	0.8856
2	40	0.8988
3	40	0.8936
4	40	0.8917

由上文可知，网络结构的节点数也会对网络性能产生影响，节点数太少会造成较大的系统误差，节点数太多，容易出现过拟合，因此需要反复进行实验来确定。为了使预测模型达到较优的性能，预测精度存在的误差较低，以下实验将隐藏层设置为 2，隐藏层每层节点数设置为 [1,100]，迭代次数均设为 200，最终确定最优的节点数。

MSCNN-LSTM 预测模型的性能如图 5 所示，横坐标为隐藏层神经元个数，纵坐标为评价指标 IA 值。IA 值随着隐藏层节点数的变化而变化，其中 IA 的值越接近 1 说明网络模型的性能越好。对图 5 进行分析，隐藏层节点数在[50,80]区间内，IA 值在 0.92～0.93 范围内，隐藏层节点个数太少，导致网络训练能力不足，对样本的关联性学习欠佳，使得 IA 值较低；隐藏层节点个数太多，导致网络模型出现过拟合问题，使得 IA 值有下降的趋势。

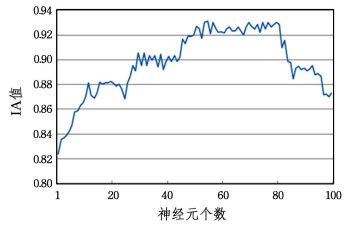

图 5 随着节点数增加 IA 值的变化图

Fig. 5 IA value change as the node increases

LSTM 神经网络 batch_size 超参实验中，batch_size 值越小，一定层度上精度得以提高，但网络训练时间被拉长，且收敛速度慢，影响训练效率。batch_size 值越大，网络训练收敛过快，很快就会陷入局部最优值，且 GPU 内存占用率高。因此选择合适的 batch size 在整个网络训练中至关重要，根据经验设定 batch size 为 10，20，40，60，100，epoch 为 200，它们的网络训练效果如图 6 所示，当 batch size 取值为 20 时，均方根误差(RMSE)最小，说明此时网络性能表现得比较优异。

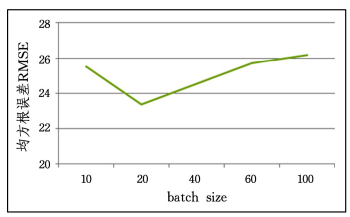

图 6 不同 batch size 均方根误差对比图

Fig. 6 Comparison of the root-mean-square error of different bitch size

MSCNN-LSTM 预测模型通过以上实验显示，当 LSTM 网络结构中隐藏层的层数为 2、每层隐藏层的节点数为 70、batch size 为 20 时，MSCNN-LSTM 预测模型在评价指标上表现得更优异。

在 MSCNN-GALSTM 预测模型中，引入了遗传算法对 LSTM 网络的初始参数进行全局寻优。采用传统的经验设定参数值设计，在算法迭代后期，算法收敛容易陷入局部最优。针对该问题，我们对遗传算法的初始化参数进行动态设定。使用较大的扰动概率，并且随着迭代的次数逐渐增大，使其跳出局部最优。反复进行实验，最终参数设置如表 2 所列，取得了较好的收敛效果。

表 2 遗传算法初始参数设置

Table 2 Genetic algorithm initializes parameter setting

参数	值
遗传代数	20
种群规模	100
交叉率	0.3
变异率	0.1

再者，遗传算法优化后的 LSTM 网络具有良好的预测效果，模型拟合能力更强。导入不同地区的空气质量数据集均表现出了较好的拟合预测能力，说明其普适性更好。相比传统的经验设定参数，利用遗传算法进行 LSTM 网络初始参数寻优，能将该模型应用到更多的需求场景中。

4.3 仿真实验结果及对比分析

采用广西区 9 个城市的空气质量数据及气象数据来预测南宁市未来 1 天的 PM2.5 浓度。实验结果展现了 100 个样

本在不同预测模型下的预测效果。我们对实验结果进行对比分析，发现每个预测模型的真实值与预测值都有着相同的趋势，但 LSTM 预测模型的预测精度明显低于 CNN-LSTM，MSCNN-LSTM，MSCNN-GALSTM 预测模型，原因在于后三者对空气质量数据及气象数据的空间特征进行了提取，并融合了时间特征，从而得到了更全面、更深层次的时空特征，最终促使网络输出层得到更好的预测效果，由此可见，在空气质量预测问题中，时空特征对预测结果有着明显的影响。

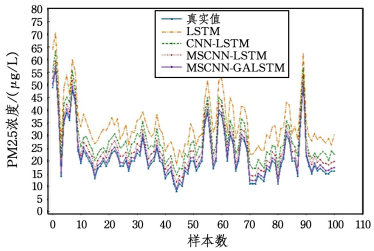

图 7　4 种预测模型拟合结果与真实值对比图

Fig. 7　Comparison of fitting results and real values of 4 prediction models

LSTM，CNN-LSTM，MSCNN-LSTM，MSCNN-GALSTM 预测模型最终的预测精度分别为 88.1%，93.0%，94.6%，95.5%。另外，当某个样本的 PM2.5 浓度出现峰值时，LSTM，CNN-LSTM，MSCNN-LSTM，MSCNN-GALSTM 预测模型的平均准确率分别为 75.1%，87.3%，89.5%，90.8%。由此可见，LSTM 收敛速度较慢，CNN 网络在收敛过程中震荡严重。MSCNN-LSTM 相比 LSTM 网络收敛速度更快，相比 CNN 网络在收敛过程中精度更高。本文改进的 MSCNN-LASTM，MSCNN-GALSTM 预测模型表现出了更好的泛化能力，说明遗传算法动态参数的设定在跳出局部最优方面具有明显的效果，具有更强的全局寻优的能力。

为了更直观地对每个模型的性能进行评价，表 3 列出了南宁 PM2.5 浓度预测在不同模型下 3 个评价指标的表现。由表 3 可知，MSCNN-LSTM，MSCNN-GALSTM 预测模型在各性能指标上均优于其他预测模型。

表 3　不同模型的性能指标

Table 3　Performance metrics for different models

模型	评价指标		
	RMSE	MAE	IA
LSTM	23.1822	17.9653	0.8839
CNN-LSTM	16.6671	12.4720	0.9302
MSCNN-LSTM	15.5675	11.2646	0.9466
MSCNN-GALSTM	12.8315	9.6134	0.9551

结束语　对实验结果进行分析发现，单一的 LSTM 预测模型仅考虑了空气质量数据及气象数据的时间序列问题，未把空间特征加以考虑，使得预测模型的精度较低，而 CNN-LSTM，MSCNN-LSTM，MSCNN-GALSTM 预测模型预测考虑了时空特征的关联性，使得预测精度得到了较大的提升。另外，MSCNN-LSTM 预测模型优于 CNN-LSTM 预测模型，原因在于前者采用了多尺度卷积来提取空气质量数据及气象数据的时空特征，使得特征提取更加全面、多样化、层次更深，进而提升了预测精度；MSCNN-GALSTM 预测模型优于 MSCNN-LSTM 网络模型，原因在于前者引入了遗传算法对 LSTM 网络进行参数寻优，使得网络对最优参数集有更强的捕抓能力，得出空气质量数据隐藏的长期依赖关系更加准确，使得预测精度也有进一步的提升。

参 考 文 献

[1] ÖZELKADILAR G，KADILAR C. Assessing air quality in Aksaray with time series analysis[C]// American Institute of Physics Conference Series. AIP Publishing LLC，2017.

[2] RAHMAN N H A，LEE M H，SUHARTONO，et al. Artificial neural networks and fuzzy time series forecasting：an application to air quality[J]. Quality & Quantity，2015，49(6)：2633-2647.

[3] BIANCOFIORE F，BUSILACCHIO M，VERDECCHIA M，et al. Recursive neural network model for analysis and forecast of PM10 and PM2.5[J]. Atmospheric Pollution Research，2017，8(4)：652-659.

[4] SONG X，HUANG J，SONG D. Air Quality Prediction based on LSTM-Kalman Model[C]// 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC). IEEE，2019：695-699.

[5] CHIOU-JYE H，PING-HUAN K. A Deep CNN-LSTM Model for Particulate Matter (PM2.5) Forecasting in Smart Cities[J]. Sensors，2018，18(7)：2220-2241.

[6] VERMA I，AHUJA R，MEISHERI H，et al. Air Pollutant Severity Prediction Using Bi-Directional LSTM Network[C]// 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI). ACM，2018：651-654.

[7] BECKERMAN B S，JERRETT M，SERRE M，et al. A hybrid approach to estimating national scale spatiotemporal variability of PM2.5 in the contiguous United States[J]. Environmental Science & Technology，2013，47(13)：7233-7241.

[8] HWA-LUNG Y，CHIH-HSIN W. Retrospective prediction of intraurban spatiotemporal distribution of PM2.5 in Taipei[J]. Atmospheric Environment，2010，44(25)：3053-3065.

[9] BARI M A，KINDZIERSKI W B. Characteristics of air quality and sources affecting fine particulate matter (PM2.5) levels in the City of Red Deer，Canada[J]. Environmental Pollution，2016，221：367-376.

[10] LIAO Z H，SUN J R，FAN S J，et al. Variation characteristics and influencing factors of air pollution in Pearl River Delta area from 2006 to 2012[J]. ZhongguoHuanjingKexue/china Environmental Science，2015，35(2)：329-336.

[11] LI X，PENG L，YAO X，et al. Long short-term memory neural network for air pollutant concentration predictions：Method development and evaluation[J]. Environmental Pollution，2017，231(1)：997-1004.

[12] ZHU Y J，LI Q，HOU J X，et al. Spatio-temporal modelingand prediction of PM2.5concentration based on Bayesian method [J]. Science of Surveying and Mapping，2016，41(2)：44-48.

[13] HAN X G，LI B Y，GUAN Z Y. Atmospheric Quality Prediction Model Based on RBF Neural Network-Markov Chain of Grey Correlational Analysis Filter Indexes[J]. Acta ScientiarumNaturalium Universitatis Nankaiensis，2013，46(2)：22-27.

[14] GAO S，HU H P，LI Y，et al. AQI Prediction Based on Improved Mind EvolutionaryAlgorithm and BP Neural Network [J]. Mathematics in Practice and RACTICE Theory，2018，48(19)：151-157.

[15] YUAN G，YANG W. Evaluating China's Air Pollution Control Policy with Extended AQI Indicator System：Example of the Beijing-Tianjin-Hebei Region[J]. Sustainability，2019，11(3)：939.

[16] HU Z，LI W，QIAO J. Prediction of PM2.5 based on Elman neu-

ral network with chaos theory[C]// 2016 35th Chinese Control Conference (CCC). IEEE, 2016: 3573-3578.
[17] SAYEGH A S, MUNIR S, HABEEBULLAH T M. Comparing the Performance of Statistical Models for Predicting PM10 Concentrations[J]. Aerosol & Air Quality Research, 2014, 14(3): 653-665.
[18] GLOROT X, BENGIO Y. Understanding the difficulty of training deep feedforward neural networks[J]. Journal of Machine Learning Research, 2010(9): 249-256.
[19] YE X Z, TAO F F, QI R Z, et al. Improvement on Activation Functions of Recurrent Neural Network Architectures[J]. Jisuanji Yu Xiandaihua, 2016(12): 29-33.
[20] LI H D. Portforlio selection based on recurrent neural network [D]. Zhengzhou: Zhengzhou University, 2018.
[21] DING L, FANG W, LUO H, et al. A deep hybrid learning model to detect unsafe behavior: Integrating convolution neural networks and long short-term memory[J]. Automation in Construction, 2018, 86: 118-124.
[22] LEI Y J. Research on Urban Air Quality Prediction Based on Temporal and Spatial Optimization Neural Network[D]. Nanning: Nanning Normal University, 2020.
[23] JI L. Research and Implementation ofPM2. 5 Prediction Based on CNNs-GRU Deep Learning[D]. Chongqing: Chongqing University of Posts and Telecommunications, 2019.

ZHOU Jie, born in 1992, postgraduate. His main research interests include intelligent computing and parallel computing.

LUO Yun-fang, born in 1981, associate professor. His main research interests include big data, computer application technology and computer teaching.

(上接第 507 页)

结束语 论文针对工业生产现场海量流数据的实时频谱观察需求，提出一种基于 Apache Storm 的增量式 FFT 方法，并基于 Apache Storm 和清华 DWF 将其进行了实现，同时采用 Bently 转子实验台的不对中故障流数据进行了验证。可以得出如下结论：1)增量式的流数据处理算法必须是非递归算法；2)流数据计算过程中加入队列结构缓冲数据，可以解决数据传输过程中的丢失和乱序问题；3)复制和上锁队列可以解决线程安全问题。

参 考 文 献

[1] LI G J, CHENG X Q. Research Status and Scientific Thinking of Big Data[J]. Bulletin of Chinese Academy of Sciences, 2012, 27(6): 647-657.
[2] YAN X, SUN Y, WAN J, et al. Industrial Big Data for Fault Diagnosis: Taxonomy, Review, and Applications[J]. IEEE Access, 2017, PP(99): 1-1.
[3] LEI Y G, JIA F, KONG D T, et al. Opportunity and challenge of mechanical intelligent fault diagnosis under big data[J]. Journal of Mechanical Engineering, 2018, 54(5): 94-104.
[4] LYONS RG. Understanding digital signal processing[M]. Englewood Cliffs, New Jersy: Prentice Hall, 2010.
[5] CHEN JL, LI ZP, PAN J, et al. Wavelet transform based on inner product in fault diagnosis of rotating machinery: A review [J]. Mechanical Systems and Signal Processing, 2016, 70: 1-35.
[6] QIN SJ. Process Data Analytics in the Era of Big Data[J]. Aiche Journal, 2014, 60(9): 3092-3100.
[7] JIANG X C, SHENG G G. Research and application of big data analysis of power equipment state[J]. High Votage Engineering, 2018, 44(4): 1041-1050.
[8] QIAO X, LIU F, YU B H. Design of distributed digital signal processing algorithm library based on spark [J]. Computer Systems & Applications, 2018, 27(8): 214-218.
[9] YANG C, BAO W, ZHU X, et al. A Parallel Fast Fourier Transform Algorithm for Large-Scale Signal Data Using Apache Spark in Cloud[C]// International Conference onAlgorithms and Architectures for ParallelProcessing. Cham: Springer International Publishing, 2018: 293-310.
[10] JI P, LI H, CHEN M, et al. Dofft: a fast Fourier transform method based on Distributed Database[J]. Computer and Modernization, 2018(6): 19-24, 29.
[11] ZHANG S M, MAO D, WANG B Y. Application of big data processing technology in gearbox fault diagnosis and early warning of wind turbine[J]. Automation of Electric Power Systems, 2016, 40(14): 129-134.
[12] HU H, BO T, GONG XJ, et al. Intelligent fault diagnosis of the high-speed train with big data based on deep neural networks [J]. IEEE Transactions on Industrial Informatics, 2017, 13(4): 2106-2116.
[13] LIU Y, YANG Q, AN D, et al. An improved fault diagnosis method based on deep wavelet neural network[C]//2018 Chinese Control And Decision Conference. 2018: 1048-1053.
[14] LI H, ZHANG Q, QIN X R, et al. Bearing fault diagnosis method based on STFT and convolution neural network[J]. Journal of Vibration and Shock, 2018, 37(19): 124-131.
[15] COOLEY JW, TUKEY JW. An algorithm for the machine calculation of complex Fourier series[J]. Mathematics of Computation, 1965, 19(90): 297-301.
[16] JAIN A, NALYA A. Learning Apache Storm[M]. New York: Packt Publishing, 2014.

ZHAO Xin, born in 1994, Ph. D candidate. His main research interests includes schema evolution and data integration.