Journal of Computer Applications
计算机应用,2020,40(11):3385 - 3392
ISSN 1001-9081
CODEN JYIIDU
2020-11-10
http://www.joca.cn

文章编号:1001-9081(2020)11-3385-08
DOI:10.11772/j.issn.1001-9081.2020040471

面向空气质量的时空混合预测模型

黄伟建,李丹阳,黄　远*
(河北工程大学 信息与电气工程学院,河北 邯郸 056038)
(*通信作者电子邮箱757918272@qq.com)

摘　要:由于城市中各区域空气质量同时存在时间与空间维度上的相关性,而传统深度学习模型结构比较单一,并且难以从时空角度进行建模。针对该问题提出一种可以同时提取空气质量间复杂时空关系的STAQI模型用于空气质量预测。该模型由局部组件和全局组件构成,分别用于描述本地污染物浓度和邻近站点空气质量状况对目标站点空气质量预测产生的影响,并利用加权融合组件输出获得预测结果。在全局组件中,利用图卷积网络改进门控循环单元网络的输入部分,从而提取出输入数据中的空间特征。最后将STAQI模型与多种基准模型和变体模型进行对比。其中,STAQI模型与门控循环单元模型和全局组件变体模型相比,均方根误差(RMSE)分别下降约19%和16%。结果表明STAQI模型对于任意时间窗口都具有最佳预测性能,并且对不同目标站点的预测结果验证了该模型具有较强的泛化能力。

关键词:空气质量预测;时空数据;图卷积网络;长短期记忆;门控循环单元;深度学习

中图分类号:TP391　　**文献标志码**:A

Spatio-temporal hybrid prediction model for air quality

HUANG Weijian, LI Danyang, HUANG Yuan*
(*School of Information and Electrical Engineering, Hebei University of Engineering, Handan Hebei* 056038, *China*)

Abstract: Because the air quality in different regions of the city are correlated with each other in both time and space, the traditional deep learning model structure is relatively simple, and it is difficult to model from the perspectives of time and space. Aiming at this problem, a Spatio Temporal Air Quality Index (STAQI) model that can simultaneously extract the complex spatial and temporal relationships between air qualities was proposed for air quality prediction. The model was composed of local components and global components, which were used to describe the influences of local pollutant concentration and air quality states of adjacent sites on the air quality prediction of target site, and the prediction results were obtained by using the weighted fusion component output. In the global component, the graph convolutional network was used to improve the input part of the gated recurrent unit network, so as to extract the spatial characteristics of the input data. Finally, STAQI model was compared with various baseline models and variant models. Among them, the Root Mean Square Error (RMSE) of STAQI model is decreased by about 19% and 16% respectively compared with those of the gated recurrent unit model and the global component variant model. The results show that STAQI model has the best prediction performance for any time window, and the prediction results of different target sites verify the strong generalization ability of the model.

Key words: air quality prediction; spatio-temporal data; Graph Convolutional Network (GCN); Long Short-Term Memory (LSTM); Gated Recurrent Unit (GRU); deep learning

0　引言

由于气候变化、工业生产和人口聚集等原因,我国多地的空气质量状况不容乐观。持续恶化的空气质量状况已严重影响了中国的经济发展和公众的身体健康。随着我国空气质量监测体系的逐步完善,我国已建成多层次的空气质量监测网络。空气质量指数(Air Quality Index, AQI)根据大气环境中二氧化硫(SO_2)、二氧化氮(NO_2)、一氧化碳(CO)、臭氧(O_3)、可吸入颗粒物($PM_{2.5}$、PM_{10})浓度值计算得出用于表示空气清洁或污染程度。提高空气质量分析和预测的科学性与准确性可以帮助政府提升对空气质量评价、管理和决策的可靠性,及时采取有效的防控措施规避大气污染造成的损害,保护公众健康。

早期的空气质量预测方法主要包括数值预测和回归统计[1]。数值预测方法通过利用多个环境监测站的监测信息建立气象排放和化学模型来模拟污染物的迁移、交换、扩散和排放过程[2]。该方法受复杂的先验知识、不可靠和有限的数据以及各种使用约束[3],因此用来模拟真实的大气环境具有一定理论难度。回归统计方法避免了复杂的理论模型,通过使

收稿日期:2020-04-15;**修回日期**:2020-06-27;**录用日期**:2020-06-28。

基金项目:河北省自然科学基金资助项目(F2015402077);河北省高等学校科学技术研究项目(QN2018073)。

作者简介:黄伟建(1964—),男,山西吕梁人,教授,博士,主要研究方向:信息系统、云计算、时间序列预测评估;　李丹阳(1995—),女,河北邯郸人,硕士研究生,主要研究方向:深度学习、时间序列预测评估;　黄远(1987—),男,山西吕梁人,讲师,博士,主要研究方向:数据挖掘、信息处理、预测算法。

用基于统计的模型来预测空气质量。然而影响空气质量的各项因子与大气污染物浓度之间复杂的线性或非线性关系很难用确定的数学模型进行描述[4]。

随着物联网与传感器技术的快速发展和应用,城市中各种传感器及相关数据采集设备采集的大气数据为空气质量预测提供了必要的数据来源。由于传统的浅层学习模型在处理空气质量这类大数据方面仍存在瓶颈,因此新的空气质量预测方法需要数据驱动模型的支持[5]。文献[6]利用长短期记忆(Long Short-Term Memory, LSTM)网络来挖掘空气质量数据中存在的深层次时间依赖特征,通过提取输入序列中的长时间依赖关系对空气质量进行预测。由于LSTM网络无法对空气质量间复杂的空间相关性进行建模分析,因此预测精度较低。文献[7]提出基于历史空气污染物浓度数据、气象数据和时间戳数据的空气污染物浓度预测LSTME(Long Short-Term Memory Extended)模型。该模型能够对具有长时间依赖性的时间序列进行建模,并能自动确定最优滞后时间。然而LSTME模型随着预测步长的增加,其预测性能有所下降。文献[8]提出一个通用且有效的DAL(Deep Air Learning)模型来解决细粒度空气质量中的插值、预测和特征分析。该模型主要思想在于嵌入式特征选择以及对深度学习网络中不同层的半监督学习,利用未标记空气质量数据的相关信息来提高插值和预测性能。

除了对空气质量数据间的时间相关性进行建模外,非线性空间依赖性也是影响空气质量预测性能的重要因素之一。由于地面环境监测站呈现非均匀分布,因此如何将空间特征处理嵌入到各种基于神经网络的方法中一直是以往工作中的难题。一些折中的方法尝试将卷积神经网络(Convolutional Neural Network, CNN)应用于欧氏空间,通过人工将监测站点重新排列为二维阵列来挖掘其中的空间依赖性。如文献[9]采用融合CNN和LSTM结构的ConvLSTM(Convolutional Long Short-Term Memory)模型来模拟站点测量数据间的时空关系。但该模型是基于CNN构建的网络,因此最适用于欧氏数据中的空间关系。为了使城市中多个环境监测站点间的空间特征处理能够嵌入到深度学习方法中以进一步提高空气质量预测精度,本文利用图卷积网络(Graph Convolutional Network, GCN)使原先只适用于处理欧氏数据的卷积操作扩展至可处理任意图数据的特点[10],提出一种基于GCN、门控循环单元(Gated Recurrent Unit, GRU)网络、LSTM网络构成的时空混合STAQI(Spatio Temporal Air Quality Index)模型用于空气质量预测。

本文主要贡献如下:

1)对门控循环单元网络进行改进。通过将图卷积网络嵌入门控循环单元,使该网络对输入数据具有空间特征提取的能力。这使原本适用于处理时间序列的门控循环单元网络可同时提取数据间时空依赖特征,同时与长短期记忆网络结合,提出一种新型的空气质量预测模型。

2)为防止过拟合问题的发生,通过在原有损失函数的基础上加入L2正则化项来降低模型复杂度和不稳定性。

3)在真实的数据集中从不同角度评估STAQI模型和其他多种模型的预测性能,验证了该模型具有最佳预测效果以及较强的泛化能力。

1 问题分析

空气质量监测站通过在站内安装多参数自动传感器来连续自动对周围空气质量进行监测。城市中部署的环境监测站分布于不同的地理位置空间并以一定时间间隔进行自动监测。因此监测站点产生大量具有空间坐标和时间戳的数据,这些数据被称为时空数据[11]。

从空间维度来看,某一区域的环境污染排放会受其他空间单元环境污染排放空间溢出效应的影响[12]。地理层次较高的位置具有较粗的粒度信息,父节点粒度由其子节点粒度组成。例如,一个监测站位于某一地区,该监测数据能准确反映周围区域细粒度空气质量状况。而此地区又位于某一城市,整个城市粗粒度空气质量信息由该城市各监测站点的监测数据共同决定。此外,两个站点间的地理距离与两地空气质量相关性成正比。

从时间维度来看,大气污染物排放影响因素的变化过程往往是缓慢的,当前污染排放积累值会受到前期污染排放的影响。对监测数据中每个实例的时间戳按时间顺序进行排序可生成顺序属性,其中相邻的时间戳通常比远处的时间戳具有更高的相似性。

综上所述,空气质量间的动态时空相关性可表示为两部分:第一是各站点传感器间相关性。从图1中可以看出不同传感器时间序列之间的空间相关性是高度动态的,并随时间不断变化。第二是传感器内相关性。地理感知时间序列通常遵循周期性变化模式(如图1中站点4产生的时间序列),并随时间与地理位置的不同而变化[13]。因此,空气质量预测是一个典型的时空序列预测问题,时间与空间依赖关系应该被明确建模。

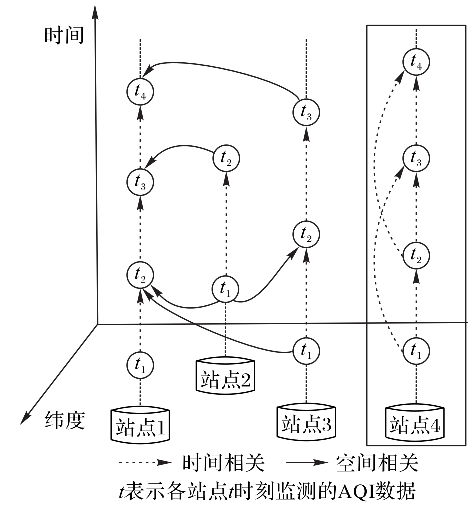

图1 多站点间的动态时空相关性

Fig. 1 Dynamic spatio-temporal correlation between multiple sites

2 时空混合STAQI模型

2.1 STAQI模型结构

STAQI模型由全局组件和局部组件构成,对目标区域污染物以及周围邻近区域内空气质量动态转化对目标站点空气质量预测产生的影响进行分析研究。在不用人为处理数据提取特征的情况下,STAQI模型通过反复训练可从大量有监督数据集中自动学习空气质量数据中动态时空依赖特征进而对

目标站点进行细粒度空气质量预测。该模型结构如图2所示，下面对STAQI模型实现过程进行主要介绍。

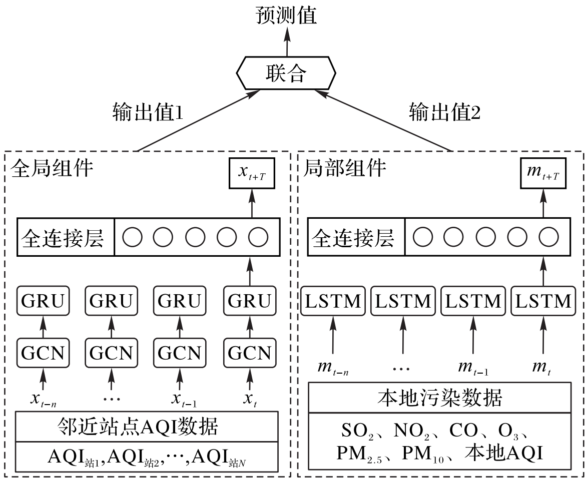

图2　STAQI模型结构

Fig. 2　Structure of STAQI model

2.1.1　全局组件

城市各监测站点间的空气质量并不是相互孤立而是存在一定相关性。通过选取北京市东四、天坛与官园站点同一时段内空气质量数据进行对比，结果如图3所示，可以发现各站点空气质量随时间具有相似的变化趋势。因此全局组件从时空依赖的角度分别进行建模并融合以提取邻近站点空气质量对预测产生的影响。

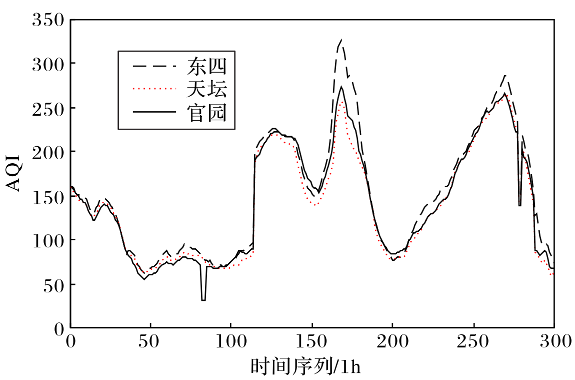

图3　东四、天坛、官园站点空气质量对比

Fig. 3　Comparison of air quality at Dongsi, Tiantan and Guanyuan sites

1)空间依赖性建模。

城市中分布的各监测站点间距并不是等同的，因此构成的拓扑网络为图结构，这意味着无法使用CNN处理这类非欧氏数据以提取空间特征。GCN是一种对图数据进行深度学习的方法[14]，其核心思想是学习一个函数映射使得图中的节点可以聚合自身节点和邻居节点的特征来生成节点新表示。

定义1　拓扑图G。利用未加权图$G = (V, E)$来描述城市各环境监测站间的地理位置拓扑结构。每个站点作为图中的一个节点，V代表城市中所有站点的集合，即$V = \{V_1, V_2, \cdots, V_N\}$，$N$为监测站点总个数。$E$代表各监测站点间边的集合。由于大气环境是一个实时动态变化的系统，根据地理第一定律，各站点间存在不同程度的影响关系。因此利用式(1)来计算两两站点间的距离并取其倒数作为连边权重值存储在邻接矩阵$\boldsymbol{A}$中，$\boldsymbol{A} \in \mathbf{R}^{N \times N}$。

$$L = 2r * \arcsin\left(\sqrt{\left(\sin\frac{a-c}{2}\right)^2 + \cos a * \cos c * \left(\sin\frac{b-d}{2}\right)^2}\right) \tag{1}$$

其中：$X(a, b)$、$Y(c, d)$代表两个站点位置，a、c为纬度，b、d为经度，r为地球半径。

定义2　特征矩阵$\boldsymbol{X}^{N \times P}$。将目标站点所在城市各环境监测站历史AQI值作为网中节点的属性特征，表示为$\boldsymbol{X} \in \mathbf{R}^{N \times P}$。$P$代表节点属性的数量特征，即历史时间序列长度。$\boldsymbol{X}_i \in \mathbf{R}^{N \times i}$表示$i$时刻城市所有环境监测站的AQI当前值。

GCN通过利用目标站点所在城市各环境监测站地理位置拓扑结构G和特征矩阵$\boldsymbol{X}$的信息学习复杂映射函数以提取空间特征，主要步骤如下：

①根据目标站点所在城市各环境监测站地理位置构建拓扑图G和邻接矩阵$\boldsymbol{A}$。

②利用各站点监测数据的AQI信息构建特征矩阵$\boldsymbol{X}$。

③基于邻接矩阵$\boldsymbol{A}$进行拉普拉斯变换生成矩阵$\hat{\boldsymbol{A}}$，可表示为$\hat{\boldsymbol{A}} = \tilde{\boldsymbol{D}}^{-\frac{1}{2}}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{D}}^{-\frac{1}{2}}$，其中$\tilde{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{I}_N$为自连接邻接矩阵，$\boldsymbol{I}_N$为单位矩阵，$\tilde{\boldsymbol{D}}$为度矩阵。

④利用式(2)提取特征矩阵中的空间信息：

$$\boldsymbol{H}^{(i+1)} = \sigma\left(\hat{\boldsymbol{A}}\boldsymbol{H}^{(i)}\boldsymbol{W}^{(i)}\right) \tag{2}$$

其中：$\sigma(\cdot)$表示激活函数，$\boldsymbol{W}^{(i)}$表示第i层权值矩阵，$\boldsymbol{H}^{(i)}$表示第i层的激活值，且$\boldsymbol{H}^{(0)} = \boldsymbol{X}$。

2)时间依赖性建模。

循环神经网络(Recurrent Neural Network, RNN)是一个可用于处理时间序列的网络[15]。然而，在实际的运用中会存在梯度消失、梯度爆炸等问题[16]，因此传统的RNN在时序预测方面存在一定的局限性。GRU是RNN的一种变体结构，该网络具有结构相对简单、参数较少、训练能力较快等优势。因此在全局组件中利用GRU网络提取邻近站点空气质量数据间的时间依赖特征。

全局组件原理如图4所示。首先根据目标站点所在城市各环境监测站地理位置信息构成拓扑图，根据拓扑图计算生成相应的邻接矩阵、度矩阵和拉普拉斯矩阵。利用GCN根据拉普拉斯矩阵来动态捕获各环境监测站的不同影响权重，获得空间特征。其次将提取空间特征后的时间序列作为GRU的输入，通过单元之间的信息传递获取时间特征。由于GRU整体输出为最后一个单元时刻的隐藏状态，它的维度不同于标签向量，因此最后通过全连接层做维度转换。图5左侧显示了全局组件预测过程，其中每一个循环单元结构简称为GG(GCN GRU)，右侧显示了GG单元细胞具体结构，$\boldsymbol{h}_{t-1}$表示$t-1$时刻的输出，GCN是图卷积操作，输入数据$\boldsymbol{x}_t$经过GCN生成$\boldsymbol{x}_t'$，$\boldsymbol{u}_t$、$\boldsymbol{r}_t$分别是t时刻的更新门和重置门，$\boldsymbol{c}_t$为t时刻细胞状态，$\boldsymbol{h}_t$表示t时刻的输出。

全局组件计算过程如下所示：

$$\boldsymbol{u}_t = \sigma(\boldsymbol{W}_u[f(\boldsymbol{A}, \boldsymbol{X}_t), \boldsymbol{h}_{t-1}] + \boldsymbol{b}_u) \tag{3}$$

$$\boldsymbol{r}_t = \sigma(\boldsymbol{W}_r[f(\boldsymbol{A}, \boldsymbol{X}_t), \boldsymbol{h}_{t-1}] + \boldsymbol{b}_r) \tag{4}$$

$$\boldsymbol{c}_t = \tanh(\boldsymbol{W}_c[f(\boldsymbol{A}, \boldsymbol{X}_t), (\boldsymbol{r}_t * \boldsymbol{h}_{t-1})] + \boldsymbol{b}_c) \tag{5}$$

$$\boldsymbol{h}_t = \boldsymbol{u}_t * \boldsymbol{h}_{t-1} + (1 - \boldsymbol{u}_t) * \boldsymbol{c}_t \tag{6}$$

其中：$f(\boldsymbol{A}, \boldsymbol{X}_t)$表示图卷积过程，$\boldsymbol{W}$和$\boldsymbol{b}$代表训练过程中的权重和偏置。

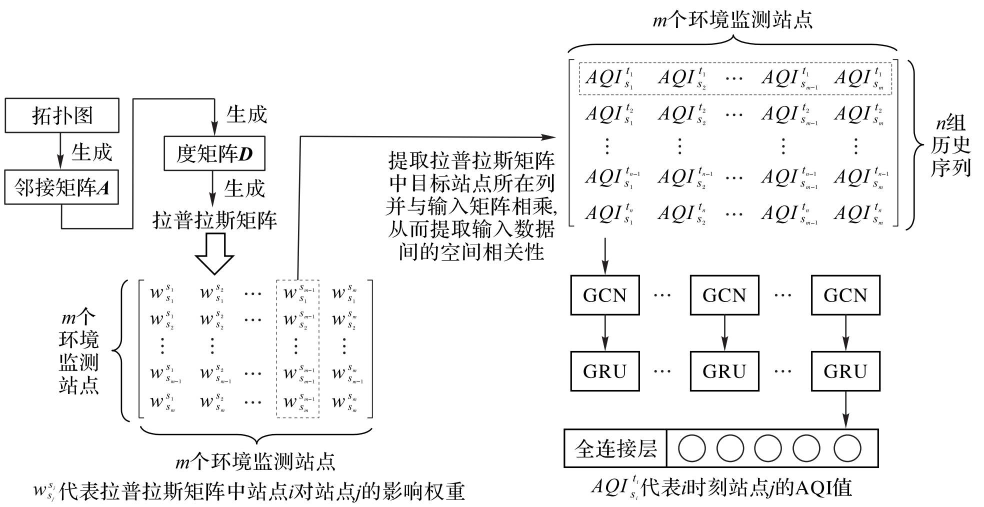

图4　全局组件原理

Fig. 4　Principle of global component

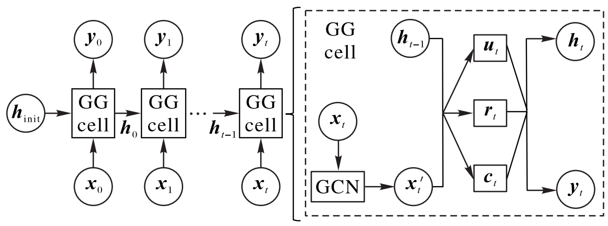

图5　全局组件预测过程

Fig. 5　Prediction process of global component

$\boldsymbol{h}_t$接入输出层，经过Softmax函数后取得预测值，进而得到残差$\boldsymbol{E}$：

$$\begin{cases} \hat{\boldsymbol{y}}_t = \mathrm{Softmax}(\boldsymbol{V}\boldsymbol{h}_t + \boldsymbol{b}_v) \\ \boldsymbol{E} = -\sum_{i=1}^{\mathrm{len}(\boldsymbol{x}_t)} \boldsymbol{y}_i \log(\hat{\boldsymbol{y}}_t) \end{cases} \tag{13}$$

2）反向计算每个神经元的误差项$\boldsymbol{\delta}$值。误差项的反向传播包括两个方面：一个是沿时间轴的反向传播，即根据t时刻的误差项$\boldsymbol{\delta}_t$来计算$t-1$时刻的误差项$\boldsymbol{\delta}_{t-1}$。其中$\boldsymbol{\delta}_t$定义为$\boldsymbol{\delta}_t = \frac{\partial \boldsymbol{E}}{\partial \boldsymbol{h}_t}$，则$t-1$时刻的误差项$\boldsymbol{\delta}_{t-1}$为：

$$\boldsymbol{\delta}_{t-1} = \frac{\partial \boldsymbol{E}_t}{\partial \boldsymbol{h}_{t-1}} = \frac{\partial \boldsymbol{E}_t}{\partial \boldsymbol{h}_t}\frac{\partial \boldsymbol{h}_t}{\partial \boldsymbol{h}_{t-1}} = \boldsymbol{\delta}_t \frac{\partial \boldsymbol{h}_t}{\partial \boldsymbol{h}_{t-1}} \tag{14}$$

另一个是将误差向上一层传播。即假设当前为第l层来计算$l-1$层的误差项。

3）根据相应的误差项来计算各权重矩阵和偏置向量的梯度（共计12个），从而更新各参数。

2.1.3　组件融合

STAQI模型采用加权融合的方式来联合局部组件输出o_v与全局组件输出o_c以形成目标站点AQI预测值Y，其计算过程如下所示：

$$Y = \alpha o_v + (1-\alpha) o_c \tag{15}$$

其中α为组合权值（$0 \leqslant \alpha \leqslant 1$）。在实验部分，将对组件分配不同权重从而进行预测性能对比并根据实验结果选取α最佳取值。

2.2　损失函数

模型训练过程以最小化损失函数为目标，使模型更精确地拟合训练数据。但若模型参数w太多，会导致模型复杂度大大提高，模型在训练集上具有出色的拟合效果，但在测试集上性能欠佳。因此本文在原有的损失函数$L_0(w)$基础上加入L2正则化项来对权重系数加以约束限制，使得模型尽量简单，整体的损失函数$L(w)$可表示为：

$$L(w) = L_0(w) + \frac{\lambda}{2n}\sum_{i=1}^{n} {w_i}^2 \tag{16}$$

对$L(w)$进行求导：

2.1.2　局部组件

一方面，当前空气质量状况会受到过去一段时间内污染物累积的影响；另一方面，站点监测到的空气质量数据是一组具有一定周期性的序列。比如一个工厂周期性地排放污染气体，那么附近站点的空气质量可能会随着时间的推移而重复。因此空气质量在时间维度上既具有短期性质，也具有长期性质。

LSTM在传统RNN基础上通过在隐藏层各神经单元中增加记忆单元从而使时间序列上的记忆信息可控[17]。LSTM单元细胞由遗忘门$\boldsymbol{f}_t$、输入门$\boldsymbol{i}_t$和输出门$\boldsymbol{o}_t$构成[18]，各部分更新公式如下所示：

$$\boldsymbol{f}_t = \sigma(\boldsymbol{W}_f \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_f) \tag{7}$$

$$\boldsymbol{i}_t = \sigma(\boldsymbol{W}_i \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_i) \tag{8}$$

$$\tilde{\boldsymbol{c}}_t = \tanh(\boldsymbol{W}_c \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_c) \tag{9}$$

$$\boldsymbol{c}_t = \boldsymbol{f}_t \circ \boldsymbol{c}_{t-1} + \boldsymbol{i}_t \circ \tilde{\boldsymbol{c}}_t \tag{10}$$

$$\boldsymbol{o}_t = \sigma(\boldsymbol{W}_o \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_o) \tag{11}$$

$$\boldsymbol{h}_t = \boldsymbol{o}_t \circ \tanh(\boldsymbol{c}_t) \tag{12}$$

其中：$\circ$表示按元素乘，$\boldsymbol{W}$和$\boldsymbol{b}$代表训练过程中的权重和偏置，$\boldsymbol{h}_t$、$\boldsymbol{c}_t$分别表示t时刻的隐藏状态和细胞状态。

局部组件以目标站点过去n个时间步长的本地污染物浓度与目标站点AQI值作为输入，通过LSTM网络提取输入数据中的时间相关特性，最后将LSTM的输出经过全连接层转换成与标签向量一致的维度，从而得到局部组件的输出。局部组件的训练算法主要包含以下3个步骤：

1）前向计算LSTM每个神经元的输出值，即分别根据式（7）~式（12）计算$\boldsymbol{f}_t$、$\boldsymbol{i}_t$、$\boldsymbol{c}_t$、$\boldsymbol{o}_t$、$\boldsymbol{h}_t$向量值。每个时刻的隐藏状态

$$\frac{\partial L(w)}{\partial w}=\frac{\partial L_0(w)}{\partial w}+\lambda w \tag{17}$$

对参数 w 更新如下所示:

$$w'=w-\eta\frac{\partial L(w)}{\partial w}=w-\frac{\eta\lambda w}{n}-\frac{\partial L_0(w)}{\partial w}=\left(1-\frac{\eta\lambda}{n}\right)w-\frac{\partial L_0(w)}{\partial w} \tag{18}$$

由此可以看出当 w 趋于0时,加入L2正则化项可以使得参数减小到很小范围内,从而降低模型复杂度。综上所述,STAQI模型中使用的损失函数如式(19)所示,其中用 Y_t 和 $\hat{Y}_t$ 来表示目标站点实际AQI值和预测AQI值。损失函数中第一项用于最小化站点实际AQI值与预测AQI值之间的误差。第二项 L_{reg} 是L2正则化项,有助于避免过拟合问题,λ 为超参数。

$$loss=\left\|Y_t-\hat{Y}_t\right\|+\lambda L_{\mathrm{reg}} \tag{19}$$

2.3　STAQI模型算法流程

输入　测试数据;

输出　预测值。

1)初始化模型多组超参数值。

2)加载邻接矩阵 $\boldsymbol{A}$ 和数据集。

3)使用Min-Max方法将数据归一化到[-1, 1]。

4)根据不同时间窗口,利用滑动窗口机制构建有监督数据集和测试集。

5)根据批尺寸大小计算训练总批次。

6)搭建全局组件。首先实现GG单元细胞结构,通过继承RNNCell类改写_init_和_call_方法来实现。_init_方法中需要指定激活函数、节点数量、隐藏单元数量并获取根据邻接矩阵计算得到的拉普拉斯矩阵。_call_方法中对输入数据进行变化,通过提取拉普拉斯矩阵中目标站点所在列与原始输入数据相乘,利用提取空间特征后的输入数据新表示来计算GRU单元细胞状态。调用自定义GG细胞类,利用最后单元状态输出值通过全连接层转化成与标签向量相同的维度。

7)搭建局部组件。设置LSTM节点数量和隐藏单元数量,利用LSTM网络最后单元输出值作为全连接层的输入,作维度转化。

8)融合全局组件与局部组件输出。

9)定义损失函数和均方根误差。

10)使用Adam优化器以最小化损失值为目标。

11)定义以训练总批次为循环次数的循环体,在循环体中首先实现对训练集中输入数据的划分,即划分为全局组件与局部组件输入。其次调用已定义的全局和局部组件。最后将当前批次局部组件输入、全局组件输入、标签数据作为填充数据来动态执行优化器、损失函数值以及预测值的计算。

12)对训练好的模型输入测试数据进行测试,获得模型输出值。

13)将模型输出值进行反归一化处理,得到模型实际预测值并输出。

3　实验与结果

3.1　数据获取与处理

本文选取北京地区35个环境监测站从2014年4月1日至2019年12月31日的历史监测信息作为实验数据集。数据从http://beijingair.sinaapp.com网站中获得。空气质量监测数据为小时粒度信息,包括 SO_2、NO_2、CO、O_3、$PM_{2.5}$、PM_{10} 六种大气污染物浓度值和AQI值。通过在Pycharm开发环境中安装配置Tensorflow框架并使用Python编程语言来完成有关实验。

收集的数据必须经过预处理才可作为STAQI模型的输入,主要由于以下两方面原因:第一,外界多种不确定性因素的干扰导致收集的数据中存在某些时间序列或属性值缺失,因此需要遍历整个数据集对缺失的时间序列进行插入并利用空值填充方法填充其他特征数据项。第二,空气质量受多重因素影响,各影响因子量纲不同。为保证模型预测性能,本文采用Min-Max方法将数据归一化到[-1, 1],最终在评价模型的预测效果时,再将预测值重新还原到原来的大小并与真实值进行比较。整个数据集按照8∶2比例划分为训练集和测试集用于模型训练及测试。训练集中的每条数据由同一时刻目标站点6种污染物浓度值和目标站点所在城市各环境监测站点AQI值组成。

3.2　实验

3.2.1　实验流程

本实验首先需要收集相关实验数据,对收集的数据进行预处理使数据变得完整以符合模型的输入要求。其次将整个数据集划分为训练集与测试集。利用训练集中的数据,使用滑动窗口机制来建立模型输入和输出结果之间的关系从而构成有监督数据集。每条有监督数据集由特征窗口和标签窗口构成。通过将大量有监督数据集作为模型输入并在一定迭代次数内不断训练调整模型参数,使得模型可以学习特征和标签之间复杂的非线性关系。最后使用测试集对训练好的模型性能进行测试,在面对只有特征没有标签的数据时,训练好的STAQI模型可以推断出与之对应的标签进而完成空气质量预测任务。STAQI模型的整个实验流程如图6所示。

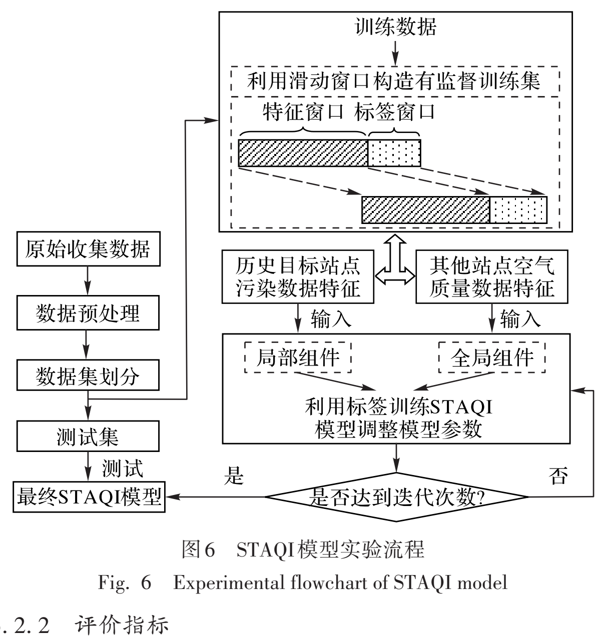

图6　STAQI模型实验流程

Fig. 6　Experimental flowchart of STAQI model

3.2.2　评价指标

为了评估STAQI模型训练后的预测性能,本文使用均方根误差(Root Mean Square Error, RMSE)来评估站点空气质量真实值与预测值之间的差异,RMSE数值越低表示模型的预测精度越高,其计算公式如下所示:

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_t-\hat{Y}_t)^2} \tag{20}$$

3.2.3 模型超参数设置

模型超参数的不同取值对模型预测性能将产生一定影响，主要包括迭代次数、网络深度、学习率、输入层的向量大小、隐藏层单元数量、批尺寸等超参数[19]。本实验采用Adam优化器，手动设置学习率为0.001，批尺寸为64，迭代次数为300。为了最小化隐藏单元数量的不同取值对STAQI模型预测性能产生的客观影响，将从[8,16,32,64,100,128]中选择不同的隐藏单元数量进行实验，结果如表1所示。在固定α取值为0.5时，当增加隐藏单元数量时，RMSE值先降低后增加。这主要是因为当隐藏单元数量大于一定阈值时，模型复杂度和计算难度大大增加，从而增加了预测误差。当隐藏单元数量为64时，RMSE误差值最小，因此在实验中将LSTM模型与GRU模型中的隐藏单元数量设置为64。

表1 隐藏单元数量对RMSE的影响

Tab. 1 Influence of number of hidden cells on RMSE

隐藏单元数量	RMSE	隐藏单元数量	RMSE
8	5.5	64	3.9
16	4.2	100	4.1
32	4.0	128	4.3

此外，由于STAQI模型输出是一个局部组件和全局组件的加权参数值α($0 \leqslant \alpha \leqslant 1$)。为设置$\alpha$的最佳取值，本文在0.0和1.0之间改变$\alpha$大小来比较模型RMSE值，实验结果如表2所示，选取RMSE最低值对应的0.6作为本实验中α的取值。通过对α参数寻优的结果可以发现局部组件具有更大的权重，这意味着本地污染物浓度对目标站点空气质量预测产生的影响要高于邻近站点空气质量对目标站点空气质量预测产生的影响。

表2 α取值对RMSE的影响

Tab. 2 Effect of α value on RMSE

α	RMSE	α	RMSE
0.0	6.4	0.6	4.8
0.1	5.6	0.7	5.3
0.2	5.9	0.8	5.9
0.3	5.5	0.9	5.7
0.4	5.7	1.0	6.7
0.5	5.3		

3.3 结果与分析

3.3.1 基准模型比较

本节从不同时间窗口和预测不同目标站点两个方面将STAQI模型与自回归综合移动平均模型(Auto Regressive Integrated Moving Average model，ARIMA)、支持向量回归(Support Vector Regression，SVR)[20]、反向传播(Back Propagation，BP)神经网络[21]、GRU进行对比。其中SVR使用线性核函数，惩罚项为0.001。

1)基于不同时间窗口的空气质量预测。

利用时间窗口描述模型使用过去N小时的历史数据来预测未来T小时后的目标站点空气质量，简记为(N,T)。通过使用不同基准模型对北京市东四监测站点进行预测，实验结果如表3所示。从表3可以得到以下四方面结论：第一，GRU模型强调了时间特征建模的重要性，通常比其他基线模型(如ARIMA、SVR、BP模型)具有更高的预测精度。第二，基于时空特性的STAQI模型与只考虑时间特征的GRU模型相比能获得更好的预测效果。第三，对于包含时序处理模块的模型(如GRU、STAQI模型)来说，在相同预测步长的情况下，模型输入的历史数据越多，可进一步提高其预测能力。第四，无论时间窗口如何变化，STAQI模型都可以通过训练获得最佳预测性能，相较于基线模型中性能较优的GRU模型，STAQI模型RMSE值大约下降19%。因此STAQI模型不仅可以用于短期预测，还适用于中长期预测。

表3 不同时间窗口的预测结果

Tab. 3 Prediction results of different time windows

时间窗口	ARIMA	SVR	BP	GRU	STAQI
(6,5)	11.31	9.65	8.75	7.86	5.70
(6,10)	11.49	9.78	8.45	7.91	5.81
(6,15)	13.25	10.25	9.31	8.15	5.96
(6,20)	12.25	9.61	8.41	8.23	6.58
(12,5)	11.14	9.94	8.52	7.66	5.65
(12,10)	11.25	9.71	8.21	7.87	5.70
(12,15)	11.86	8.75	8.62	7.94	5.88
(12,20)	11.94	9.78	9.04	8.10	6.40
(18,5)	10.85	9.68	8.65	7.31	5.53
(18,10)	11.28	8.84	8.89	7.66	5.62
(18,15)	11.32	8.81	8.62	7.82	5.79
(18,20)	11.84	9.75	8.55	7.95	6.22
(24,5)	11.32	9.84	8.23	6.92	5.50
(24,10)	11.88	9.65	8.32	7.36	5.60
(24,15)	11.26	8.98	8.54	7.52	5.63
(24,20)	11.27	8.71	8.62	7.60	6.01

2)基于不同目标站点的空气质量预测。

为检验STAQI模型对不同目标站点的预测能力是否具有显著差异，本文使用不同基线模型对北京市天坛站点未来一定时间步长的AQI进行预测，实验结果如表4所示。从表4可以发现STAQI模型相比其他基线模型对于任意AQI预测步长仍能取得最佳预测性能。其中与性能较优的GRU模型相比，STAQI模型RMSE值大约下降6%，因此STAQI模型对地理位置不同、具有地势差异的不同站点的空气质量预测能力具有较强的泛化性。

表4 不同模型对天坛站点AQI预测的RMSE值

Tab. 4 RMSE values predicted by different models for AQI of Tiantan site

模型	预测时间步长/h			
	5	10	15	20
ARIMA	28.36	30.58	29.21	30.24
SVR	20.87	23.02	26.83	24.25
BP	18.65	19.25	20.18	23.16
GRU	14.59	16.84	18.91	20.01
STAQI	10.71	12.02	12.75	13.52

此外，对于北京市35个环境监测站点，本文使用不同基准模型对各环境站点未来一定时间步长的空气质量进行预测并统计各模型最优预测站点数量，实验结果如表5所示。从表5可以看出在对未来1 h的短期预测中各模型的最优预测站点数量大致相同，但随着预测窗口逐渐增大，ARIMA、SVR、BP模型对空气质量这类时空数据预测能力越来越低。由于GRU模型是一种专门用于处理时间序列的网络，随着预测步长增加，其最优站点统计数量要高于ARIMA、SVR和BP模型，但低于STAQI模型。基于时空混合的STAQI模型随预测

步长的增加相比其他基线模型显示出明显的预测优势。

表 5　不同模型最优预测站点数统计

Tab. 5　Statistics on the number of optimal prediction sites by different model

模型	预测时间步长/h			
	1	3	5	7
ARIMA	6	2	0	0
SVR	6	3	0	0
BP	4	2	0	0
GRU	9	5	4	0
STAQI	10	23	31	35

3.3.2　变体模型比较

为了检验STAQI模型中各个组件对模型预测性能的有效性,本文将STAQI模型与由GCN和GRU构成的全局组件模型(LOCAL)和由LSTM构成的局部组件模型(GLOBAL)进行比较,实验结果如表6所示。从表6可以看出对于不同预测步长STAQI模型的RMSE都取得最低值。相较于变体模型中性能较优的局部组件模型,STAQI模型RMSE值大约下降16%。这是由于局部组件模型只考虑了空气质量数据中的时间特性,忽略了空间相关性。而全局组件模型把对目标站点的空气质量预测完全依赖于邻居节点的空气质量状况,忽略了本地污染物浓度对目标站点空气质量预测的影响。STAQI模型将时间与空间、局部与全局因素进行结合,使得该模型可以更好地拟合空气质量变化趋势,提高预测精度。

表 6　变体模型预测的 RMSE 值比较

Tab. 6　Comparison of RMSE values predicted by variant models

预测时间步长/h	LOCAL	GLOBAL	STAQI
5	8.10	8.62	6.05
10	8.22	8.74	6.23
15	8.41	8.65	6.34
20	8.53	8.73	6.81

此外本文对北京市35个环境监测站点未来6 h的空气质量使用变体模型和STAQI模型分别进行预测,各模型的RMSE值如图7所示。从图7可以发现对于任一站点STAQI模型的RMSE都取得最低值,因此融合了全局组件和局部组件的STAQI模型可以提高对空气质量的预测能力。

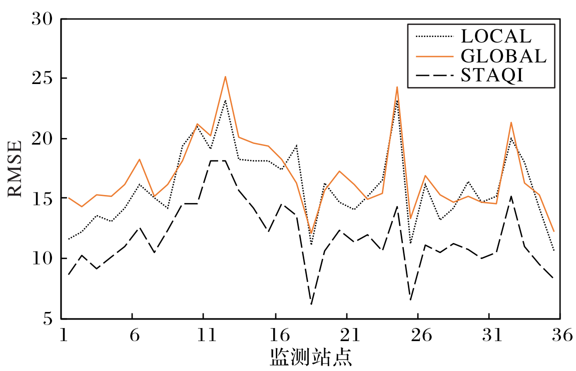

图7　变体模型对北京市35个站点的预测结果比较

Fig. 7　Comparison of prediction results of variant models for 35 sites in Beijing

3.3.3　模型预测及改进分析

为了直观显示STAQI模型在真实数据集中的预测效果,本文使用STAQI模型对东四站点进行预测,实验结果如图8所示。从图8可以看出STAQI模型通过提取空气质量中存在的时空特征可以很好地拟合空气质量变化趋势;但是STAQI模型在拐点处的预测误差要大于对平稳序列预测的误差,这是由于受极端天气条件如暴风、暴雨等影响,导致空气质量在极短的时间内发生突变。因此在未来工作中,可以在模型中融合气象因素以提高模型在空气质量拐点处的预测能力。

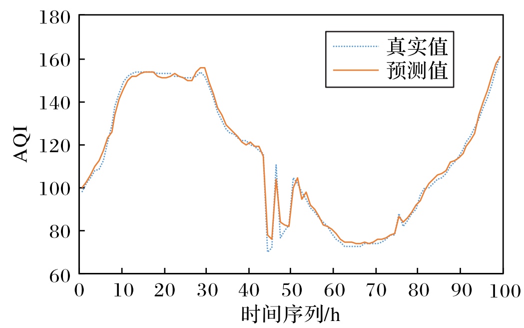

图8　STAQI模型对东四站点的预测结果

Fig. 8　Predicted results of STAQI model for Dongsi site

4　结语

针对空气质量间存在的复杂时空特征,本文提出一种面向空气质量的新型STAQI时空混合预测模型,它由全局组件和局部组件构成。该模型对地面多个环境监测站之间的时空依赖性进行明确建模,可以自动学习多变量空气质量相关时间序列数据局部趋势的相关特征和时空依赖模式。与传统仅考虑时间相关性的方法相比,STAQI模型在保留原有监测站点位置分布的基础上将空间特征处理嵌入深度学习方法中以同时提取城市中多站点间的复杂时空相关性。通过在真实的数据集上进行评估,并与不同基准模型和变体模型在不同预测水平下进行比较,结果表明STAQI模型对空气质量预测具有更精准的预测效果。

为了进一步提高模型预测性能,未来研究工作中将以以下两方面作为研究重点。首先,空气质量受多种复杂因素影响,仅仅利用污染数据和空气质量数据作为影响因素具有一定片面性,因此可以在模型中融合气象因素、交通流量等其他影响因子对问题进行建模。其次,可在LSTM网络中引入注意力机制,使得该网络对不同时期的历史数据可分配不同权重的注意力,从而获得更精准的预测。

参考文献 (References)

[1] FENG X, LI Q, ZHU Y, et al. Artificial neural networks forecasting of $PM_{2.5}$ pollution using air mass trajectory based geographic model and wavelet transformation [J]. Atmospheric Environment, 2015, 107: 118-128.

[2] BAKLANOV A, MESTAYER P G, CLAPPIER A, et al. Towards improving the simulation of meteorological fields in urban areas through updated/advanced surface fluxes description [J]. Atmospheric Chemistry and Physics, 2008, 8(3): 523-543.

[3] STERN R, BUILTJES P, SCHAAP M, et al. A model inter-comparison study focussing on episodes with elevated PM10 concentrations [J]. Atmospheric Environment, 2008, 42(19): 4567-4588.

[4] 董婷,赵俭辉,胡勇. 基于时空优化深度神经网络的AQI等级预测[J]. 计算机工程与应用, 2017, 53(21):17-23, 41. (DONG T, ZHAO J H, HU Y. AQI levels prediction based on deep neural network with spatial and temporal optimizations [J]. Computer

Engineering and Applications, 2017, 53(21): 17-23, 41.)

[5] HSIEH H P, LIN S D, ZHENG Y, et al. Inferring air quality for station location recommendation based on urban big data [C]// Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2015: 437-446.

[6] 张春露,白艳萍. 基于TensorFlow的LSTM模型在太原空气质量AQI指数预测中的应用[J]. 重庆理工大学学报(自然科学版), 2018, 32(8):137-141. (ZHANG C L, BAI Y P. Application of LSTM prediction model based on TensorFlow in Taiyuan air quality AQI index [J]. Journal of Chongqing University of Technology (Natural Science), 2018, 32(8): 137-141.)

[7] 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(Pt 1): 997-1004.

[8] QI Z, WANG T, SONG G, et al. Deep air learning: interpolation, prediction, and feature analysis of fine-grained air quality [J]. IEEE Transactions on Knowledge and Data Engineering, 2018, 30 (12): 2285-2297.

[9] HUANG C J, KUO P H. A deep CNN-LSTM model for particulate matter ($PM_{2.5}$) forecasting in smart cities [J]. Sensors, 2018, 18 (7): No. 2220.

[10] COLEY C W, JIN W, ROGERS L, et al. A graph-convolutional neural network model for the prediction of chemical reactivity [J]. Chemical Science, 2019, 10(2): 370-377.

[11] ZHANG J, ZHENG Y, QI D, et al. DNN-based prediction model for spatio-temporal data [C]// Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. New York: ACM, 2016: No. 92.

[12] 刘华军,杨骞. 环境污染、时空依赖与经济增长[J]. 产业经济研究, 2014(1):81-91. (LIU H J, YANG Q. Environmental pollution, temporal-spatial dependence, and economic growth [J]. Industrial Economics Research, 2014(1): 81-91.)

[13] ZHANG J, ZHENG Y, QI D, et al. Deep spatio-temporal residual networks for citywide crowd flows prediction [C]// Proceedings of the 31th AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2017: 1655-1661.

[14] QI Y, LI Q, KARIMIAN H, et al. A hybrid model for spatiotemporal forecasting of $PM_{2.5}$ based on graph convolutional neural network and long short-term memory [J]. Science of the Total Environment, 2019, 664: 1-10.

[15] 杨丽,吴雨茜,王俊丽,等. 循环神经网络研究综述[J]. 计算机应用, 2018, 38(S2):1-6, 26. (YANG L, WU Y X, WANG J L, et al. Research on recurrent neural network [J]. Journal of Computer Applications, 2018, 38(S2): 1-6, 26.)

[16] 于家斌,尚方方,王小艺,等. 基于遗传算法改进的一阶滞后滤波和长短期记忆网络的蓝藻水华预测方法[J]. 计算机应用, 2018, 38(7):2119-2123, 2135. (YU J B, SHANG F F, WANG X Y, et al. Cyanobacterial bloom forecast method based on genetic algorithm-first order lag filter and long short-term memory network [J]. Journal of Computer Applications, 2018, 38(7): 2119-2123, 2135.)

[17] ESPOSITO E, DE VITO S, SALVATO M, et al. Dynamic neural network architectures for on field stochastic calibration of indicative low cost air quality sensing systems [J]. Sensors and Actuators, B: Chemical, 2016, 231: 701-713.

[18] GREFF K, SRIVASTAVA R K, KOUTNÍK J, et al. LSTM: a search space odyssey [J]. IEEE Transactions on Neural Networks and Learning Systems, 2017, 28(10): 2222-2232.

[19] VINCENT P, LAROCHELLE H, LAJOIE I, et al. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion [J]. Journal of Machine Learning Research, 2010, 11:3371-3408.

[20] 胡世前,姜倩雯,凌冰,等. 基于改进支持向量机的空气质量监测预警模型[J]. 江苏大学学报(自然科学版), 2016, 37(4): 491-496. (HU S Q, JIANG Q W, LING B, et al. Monitoring and early warning of air quality based on improved support vector machines [J]. Journal of Jiangsu University (Natural Science Edition), 2016, 37(4): 491-496.)

[21] KANG Z, QU Z. Application of BP neural network optimized by genetic simulated annealing algorithm to prediction of air quality index in Lanzhou [C]// Proceedings of the 2nd IEEE International Conference on Computational Intelligence and Applications. Piscataway: IEEE, 2017: 155-160.

This work is partially supported by the Natural Science Foundation of Hebei Province (F2015402077), the Scientific Research Foundation of the Higher Education Institutions of Hebei Province (QN2018073).

HUANG Weijian, born in 1964, Ph. D., professor. His research interests include information system, cloud computing, time series prediction and evaluation.

LI Danyang, born in 1995, M. S. candidate. Her research interests include deep learning, time series prediction and evaluation.

HUANG Yuan, born in 1987, Ph. D., lecturer. His research interests include data mining, information processing, prediction algorithm.