

DOI:10.19651/j.cnki.emt.2209553

基于CRQA-DBN-ELM空气质量数据预测模型*

李志刚^{1,2} 秦林林^{1,2} 付多民^{1,2} 孙晓川^{1,2}

(1.华北理工大学人工智能学院 唐山 063210; 2.河北省工业智能感知重点实验室 唐山 063210)

摘要: 对空气质量影响因素进行分析与预测是确定关键因素和追溯污染源的简单而有效的途径。针对当前空气质量预测方法预测精度不高,且极易陷入局部最优值的问题,本文提出了一种基于交叉递归定量分析(CRQA)与深度置信网络-极限学习机(DBN-ELM)的空气质量数据预测方法。首先,采用CRQA对多种空气质量影响因素间的关联度进行分析,筛选出影响空气质量的重要因素。然后,将获取的空气质量主要影响因素输入到DBN-ELM模型中进行预测。其中,DBN用于提取空气质量主要因素的关键特征,ELM用于最终空气质量时序数据的非线性逼近。实验结果表明,在北京奥体中心站点空气质量数据集上,该模型的RMSE值为1.7759, R^2 为0.9833,优于其他对比模型。进一步,采用散点图与分位比较图方法验证了所提出模型的有效性。

关键词: 空气质量预测;CRQA;DBN-ELM

中图分类号: TP181 文献标识码: A 国家标准学科分类代码: 520.2060

Air quality data prediction method based on CRQA-DBN-ELM

Li Zhigang^{1,2} Qin Linlin^{1,2} Fu Duomin^{1,2} Sun Xiaochuan^{1,2}

(1. College of Artificial Intelligence, North China University of Science and Technology, Tangshan 063210, China;

2. Hebei Key Laboratory of Industrial Intelligent Perception, Tangshan 063210, China)

Abstract: It is a simple and effective way to determine the key factors and trace the causes of pollution by analyzing and predicting the influencing factors of air quality. Aiming at the prediction accuracy of current air quality prediction methods is not high, and it is easy to fall into the local optimal value problem, a novel model is proposed, which is based on Cross Recurrence Quantification Analysis (CRQA) and Deep Belief Network-Extreme Learning Machine (DBN-ELM) air quality data prediction method. Firstly, CRQA is used to analyze the correlation degree among various factors affecting air quality and screen out the critical factors affecting air quality. Then, the main influencing factors of air quality obtained are input into the DBN-ELM model for prediction. Concretely, DBN is used to extract key features of main air quality factors, and ELM is used for nonlinear approximation of final air quality time series data. The experimental results show that in the air quality data set of the Beijing Olympic Sports Center, the RMSE value and R^2 value of this model are 1.7759 and 0.9833 respectively, which are better than other models. Furthermore, the effectiveness of the proposed model is verified by scatterplot and quantile-quantile plot.

Keywords: air quality prediction;CRQA;DBN-ELM

0 引言

空气质量一直是世界各国关注的重要问题^[1]。随着人类社会经济的发展,酸雨、雾霾和全球变暖等空气污染引发的问题不断显现^[2-3]。空气污染不仅阻碍了人与自然的协调发展,还给公众健康和市民日常活动带来负面影响。众所周知,空气质量受到物理、化学、气象等多种复杂因素的影响。空气质量数据的分析与预测对制定大气管理决策、

控制空气污染具有重要作用,是促进可持续发展的重要组成部分。

空气质量时序数据具有非线性、多元化、动态性等特征,这使得空气质量数据的分析与建模相当困难。近年来,大量研究致力于空气质量影响因素分析与评价。例如,孙瑞玲等采用主成分分析研究PM_{2.5}中氯代类污染物与各种气象因素之间的关联性^[4]。Wang等^[5]采用空间杜宾模型量化二氧化氮等多种因素的空间差异性。此外,皮尔逊

收稿日期:2022-04-10

* 基金项目:河北省高等学校科学技术研究项目(ZD2021088)、国家重点研发计划项目(2017FE0135700)资助

相关系数分析与灰度关联分析也是量化空气质量数据相关性的主流方法^[6-8]。然而上述方法易受空气质量时序数据分布的影响,并且只关注于一维空间下空气质量影响因素间的关联度量化,忽略了空气质量数据的潜在特征。交叉递归定量分析(cross recurrence quantification analysis, crqa)是分析非线性时序数据的有力工具,为时间序列在高维空间中的潜在动态行为提供了丰富的度量指标^[9]。crqa为时序数据提供了高维特征视角,能够从多角度度量两组时序数据间的关联性,它已经被广泛应用于医学、环境、智慧海洋等领域的多种因素相似性度量分析^[10-12]。鉴于CRQA的优势,本文采用CRQA量化空气质量影响因素间的相似性,筛选影响空气质量的关键因素,这对后续空气质量数据预测和应急政策制定具有重要作用。

近年来,机器学习算法迅猛发展,支持向量机(support vector machine, SVM)、反向传播神经网络(back propagation, BP)、长短期记忆网络(long short-term memory, LSTM)等算法在空气质量预测任务中的性能已经被验证^[13-15]。特别地,作为机器学习算法的优秀代表之一,深度置信网络(deep belief network, DBN)具备卓越的特征提取性能,已经广泛用于空气质量数据预测任务等多个领域^[16-18]。众所周知,DBN的梯度下降反向传播算法导致其在训练过程中极易陷入局部最优值,并且大量的迭代计算也导致其收敛速度过慢,因此精度仍有待提升。为此,我们引入了极限学习机算法(extreme learning machine, ELM)^[19-20]。ELM具有学习速度快、泛化能力强,能够有效弥补反向传播算法的缺陷。此外,在DBN-ELM模型中,ELM可以利用DBN提供的深度特征达到相对稳定的预测性能,避免了传统ELM中数据特征提取不充分和随机初始化造成的不适定问题。

综上,本文构建了一种基于CRQA-DBN-ELM的空气质量数据预测模型。众所周知,PM2.5是影响空气质量的关键指标。因此,本文探究了多种空气质量影响因素间的高维动态特性,挖掘它们与PM2.5之间的关联性,筛选出与PM2.5相似性较高的因素,为后续PM2.5预测提供先验数据。经过CRQA筛选后的因素被输入到DBN-ELM中,完成最终的PM2.5多变量预测任务。

1 模型理论

本文构建了一种基于CRQA-DBN-ELM的空气质量数据预测模型。该模型主要包含相空间重构、CRQA分析、DBN特征提取、ELM回归预测四部分内容,描述如下。

1.1 相空间重构

相空间重构能够将一维的空气质量影响因素时间序列转换至高维相空间中。对比其原始序列,重构后的空气质量序列能够表现出更多非线性动态特性。对于一个一维的空气质量影响因素时间序列,其重构至 h 维空间中可表示如下:

$$F_w(h) = \{x_w, x_{w+\tau}, \dots, x_{w+(h-1)\tau}\} \quad (1)$$

其中, h 表示嵌入维度, τ 表示延迟时间,分别采用虚假临近点法和互信息法确定。为计算交叉递归定量分析指标,需选择相同重构参数,本文选择较大嵌入维度与较小延迟时间作为PM2.5与其他因素的最佳重构参数^[12]。

1.2 CRQA分析

CRQA是评估时间序列数据间相似性的量化方法。本文主要考虑递归率(recurrence rate, R)、确定性(determinism, D)、层流性(laminarity, L)、平均对角线长度(mean diagonal length, M)4个指标度量PM2.5与其他因素的动态相似性。

R量化了两组空气质量影响因素序列回到相空间中相同区域的概率大小,表示如下:

$$R(\epsilon) = \frac{1}{N^2} \sum_{i,j=1}^N \mathbf{M}_{i,j}(\epsilon) \quad (2)$$

其中, N 表示空气质量影响因素序列的长度, ϵ 表示阈值,设为最大相空间直径的0.7%。 \mathbf{M} 表示交叉递归矩阵,计算如下:

$$\mathbf{M}_{i,j}(\epsilon) = \Theta(\epsilon - \|x_i - y_j\|) \quad (3)$$

其中, x 和 y 分别表示重构后的空气质量影响因素。 $\Theta(\cdot)$ 为赫维赛德函数。由此可知, \mathbf{M} 为一个只包含0和1的矩阵。

D量化了两组空气质量影响因素间相似状态的可预测性与规律性,表示如下:

$$D = \frac{\sum_{l=l_{\min}}^{N_l} lP(l)}{\sum_{l=1}^{N_l} lP(l)} \quad (4)$$

其中, l_{\min} 为2, N_l 表示 \mathbf{M} 中为1值所构成的对角线结构的个数, $P(l)$ 定义如下:

$$P(l) = \sum_{i,j=1}^{N_j} (1 - \mathbf{M}_{i-1,j-1})(1 - \mathbf{M}_{i+l,j+l}) \times \prod_{k=0}^{l-1} \mathbf{M}_{i+k,j+k} \quad (5)$$

L量化了两组空气质量影响因素序列的平稳状态时间,计算如下:

$$L = \frac{\sum_{v=v_{\min}}^{N_v} vP(v)}{\sum_{v=1}^{N_v} vP(v)} \quad (6)$$

其中, v_{\min} 为2, N_v 表示 \mathbf{M} 中为1值所构成的垂直线结构的个数, $P(v)$ 定义如下:

$$P(v) = \sum_{i,j=1}^{N_v} (1 - \mathbf{M}_{i-1,j-1})(1 - \mathbf{M}_{i,j+v}) \times \prod_{k=0}^{v-1} \mathbf{M}_{i,j+k} \quad (7)$$

M量化了两组空气质量影响因素保持相似的平均时间。它是度量相似性的主要指标,定义如下:

$$\mathbf{M} = \frac{\sum_{l=l_{\min}}^{N_l} lP(l)}{\sum_{l=1}^{N_l} lP(l)} \quad (8)$$

通过上述 4 个指标的量化,可获得各个因素对 PM2.5 的综合作用结果,排名前 50% 的因素与 PM2.5 共同作为后续预测模型的输入。

1.3 DBN 特征提取

DBN 是深度学习中的一个重要模型,它是由一系列受限玻尔兹曼机(restricted boltzmann machine, RBM)单元组成的概率生成模型。经过 CRQA 筛选处理后的空气质量多维数据被输入到第一个 RBM 中。RBM 是由可见层 $\mathbf{v} = \{v_1, v_2, \dots, v_i\}$ 与隐藏层 $\mathbf{h} = \{h_1, h_2, \dots, h_j\}$ 组成的二值神经网络。RBM 的联合概率分布 $P_\theta(\mathbf{v}, \mathbf{h})$ 由能量函数 $E_\theta(\mathbf{v}, \mathbf{h})$ 所得:

$$P_\theta(\mathbf{v}, \mathbf{h}) = \frac{1}{Z(\theta)} \exp(-E_\theta(\mathbf{v}, \mathbf{h})) \quad (9)$$

$$E_\theta(\mathbf{v}, \mathbf{h}) = -\sum_i a_i v_i - \sum_j b_j h_j - \sum_{i,j} v_i w_{i,j} h_j \quad (10)$$

其中, $\theta = \{a_i, w_{i,j}, b_j\}$, $w_{i,j}$ 表示可见层第 i 个神经元与隐藏层第 j 个神经元之间的连接权重。 a 与 b 分别表示可见层与隐藏层中神经元的偏置值。 $Z(\theta)$ 代表分布函数,定义如下:

$$Z(\theta) = \sum_{\mathbf{v}, \mathbf{h}} \exp(-E_\theta(\mathbf{v}, \mathbf{h})) \quad (11)$$

隐藏层与可见层神经元间的条件概率分别定义如下:

$$P_\theta(v_i = 1 | \mathbf{h}) = \text{Sigmoid}(a_i + \sum_j w_{i,j} h_j) \quad (12)$$

$$P_\theta(h_j = 1 | \mathbf{v}) = \text{Sigmoid}(b_j + \sum_i w_{j,i} v_i) \quad (13)$$

通常情况下,采用对比散度算法来完成权重更新:

$$\Delta w_{ij} = \alpha(\langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{recon}) \quad (14)$$

$$\Delta a_i = \alpha(\langle v_i \rangle_{data} - \langle v_i \rangle_{recon}) \quad (15)$$

$$\Delta b_j = \alpha(\langle h_j \rangle_{data} - \langle h_j \rangle_{recon}) \quad (16)$$

其中, α 表示学习率, $\langle \cdot \rangle_{data}$ 表示空气质量多维数据本身的期望, $\langle \cdot \rangle_{recon}$ 表示重构后的数据期望。通过重复式(14)~(16),直至所有 RBM 网络完成权重更新,最后一层 RBM 的输出向量即为空气质量数据中最具有代表性的特征。

1.4 ELM 非线性逼近

经过 DBN 提取后的空气质量特征被输入到 ELM 中,由 ELM 完成最终的 PM2.5 预测任务。假设 ELM 隐藏层神经元个数为 num , 对于一个给定包含 q 个样本 (\mathbf{X}_i, t_i) , $i = 1, 2, \dots, q$, ELM 可表示如下:

$$\sum_{i=1}^{num} \beta_i \text{Sigmoid}(\mathbf{W}_i \cdot \mathbf{X}_j + b_i) = y_j, j = 1, 2, \dots, q \quad (17)$$

其中, Sigmoid 表示激活函数, \mathbf{W} 表示输入层与隐藏层神经元间的连接权重, β 表示隐藏层与输出层神经元之间的连接权重, b_i 表示隐含层第 i 个神经元的偏置值。该单隐藏层的 ELM 用于近似输入采样,最小化输出 PM2.5 与实际 PM2.5 值之间的误差,即:

$$\sum_{i=1}^{num} \beta_i \text{Sigmoid}(\mathbf{W}_i \cdot \mathbf{X}_j + b_i) = t_j, j = 1, 2, \dots, q \quad (18)$$

若隐藏层神经元的输出为 \mathbf{U} , 期望输出为 \mathbf{H} , 则有:

$$\mathbf{U}\beta = \mathbf{H} \quad (19)$$

在 ELM 中,连接权重 \mathbf{W} 与偏置值 b 是随机初始化的。因此,只需确定隐藏层的神经元个数与激活函数,即 β :

$$\beta = \mathbf{U}^T \mathbf{H} \quad (20)$$

2 实 验

2.1 数据来源及参数设置

实验采用的数据来自北京奥体中心站点捕获的空气质量数据^[21]。该数据集中包含:PM2.5、PM10;化学因素:二氧化硫(sulfur dioxide, SO₂)、二氧化氮(nitrogen dioxide, NO₂)、一氧化碳(carbon monoxide, CO)与臭氧(ozone, O₃)、温度(temperature, T)、气压(pressure, P)、露点温度(dew point temperature, DT)、雨量(rain, R)、风向(wind direction, WD)与风速(wind speed, WS),共 12 种空气质量影响因素。本文选取了 2016 年 3 月 1 日~2016 年 11 月 5 日共计 6 000 条数据。其中,训练集 4 500 条,测试集 1 500 条,均采用移动均值法进行处理。本实验选取对人类健康危害最大的环境污染因素 PM2.5 作为空气质量的主要度量指标,量化其他影响因素对 PM2.5 的影响程度。筛选出前 50% 的因素,联合 PM2.5 作为模型输入数据。CRQA-DBN-ELM 主要参数设置如表 1。为避免随机性,所有预测结果为指定参数设置下运行 10 次的均值。

表 1 模型参数设置

参数	设置
RBM 个数	2
DBN 隐层神经元数	10~10
ELM 隐层神经元数	12
学习率	0.01
激活函数	<i>Sigmoid</i>

2.2 评估指标

为了评估模型的性能,采用均方根误差(root mean square error, RMSE)与决定系数(coefficient of determination, R²)两种评估指标,其公式如下:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_{pi} - y_{ai})^2} \quad (21)$$

$$R^2 = \frac{\sum_{i=1}^N (y_{pi} - \bar{y}_a)^2}{\sum_{i=1}^N (y_{ai} - \bar{y}_a)^2} \quad (22)$$

其中, N 表示数据长度, y_{pi} 表示预测值, y_{ai} 表示真实值, \bar{y}_a 表示真实值的平均值。

2.3 CRQA 分析

采用了 R 、 D 、 L 、 M 指标从多角度度量各种因素与 PM2.5 关联度,其结果如表 2 所示。 R 表示了两组空气质量影响因素在同一相空间中出现相似行为的概率,其值越

大,证明两组空气质量序列间出现相似状态的频率就越高。根据表2, R 与 $PM_{2.5}$ 间的 R 值最高。众所周知,大量雨水能够冲刷空气中的雾霾颗粒,雨后的空气质量会明显变好。作为两个空气质量序列的确定性度量, D 值越高表明空气质量序列间相似状态的规律性与可预测性越高。显然, SO_2 与 $PM_{2.5}$ 间确定性达到最高。通常情况下,多种燃料燃烧时,会产生 SO_2 和多种颗粒。 L 度量了其他因素与 $PM_{2.5}$ 间的同步性。在我们的实验中, SO_2 与 $PM_{2.5}$ 间的同步性最高。 M 是度量两组序列相似性的主要指标。在该指标下, P 与 $PM_{2.5}$ 表现出最强的相似性,然而这一现象并未在前三种指标中体现,我们发现单一指标无法综合度量影响因素间的相似性。因此,我们综合了四种指标结果,用于评估各个空气质量影响因素与 $PM_{2.5}$ 的相似性。显然,各个因素与 $PM_{2.5}$ 相似性的综合排名由高到低为: P 、 SO_2 、 R 、 T 、 DT 、 O_3 、 CO 、 NO_2 、 PM_{10} 、 WD 、 WS 。根据排序结果,选取前6种因素作为影响 $PM_{2.5}$ 的重要指标,融合 $PM_{2.5}$ 因素共同作为后续输入数据。

2.4 预测结果对比

图1给出了不同对比模型的 $PM_{2.5}$ 预测曲线与实际

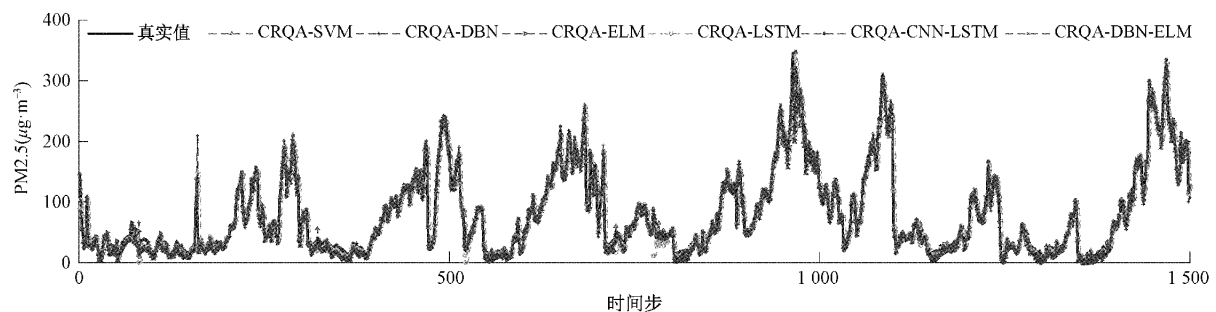


图1 $PM_{2.5}$ 预测任务中不同模型预测曲线对比

图2给出了不同模型预测值与实际值之间的绝对误差曲线对比。当曲线越接近 $y=0$ 时,表明预测结果与实

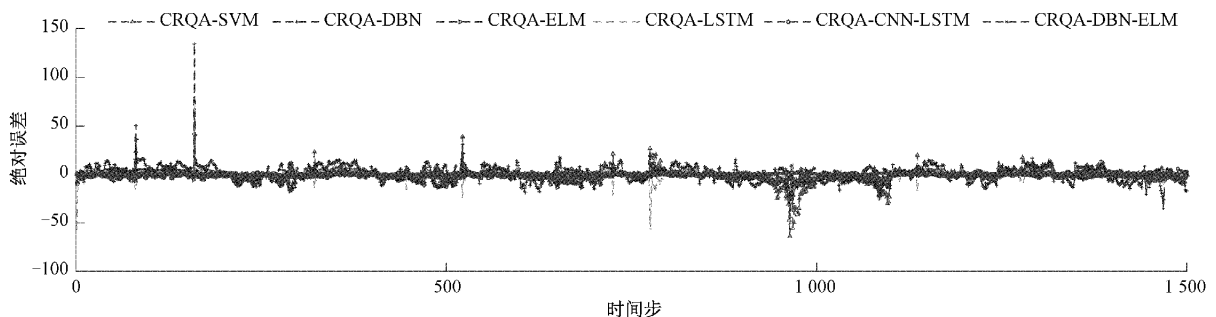


图2 $PM_{2.5}$ 预测任务中不同模型绝对误差对比

表3给出了6组模型在 $PM_{2.5}$ 预测任务中的性能评价指标。其中,RMSE值越小,表明模型预测值越接近实际值。 R^2 值越低,表明模型的预测值越偏离实际值。显然,CRQA-DBN-ELM的RMSE和 R^2 优于其它预测模型,即该模型可以更精确地建模 $PM_{2.5}$ 数据的变化趋势。

表2 CRQA度量结果

因素	R	D	L	M
PM_{10}	0.010 5	0.605 2	0.804 1	2.911 3
SO_2	0.065 7	0.861 7	0.888 8	4.281 8
NO_2	0.021 1	0.685 1	0.874 7	3.023 0
CO	0.026 2	0.723 0	0.843 6	3.232 0
O_3	0.011 8	0.716 2	0.858 2	3.512 3
T	0.022 6	0.815 4	0.870 3	3.680 3
P	0.021 9	0.859 8	0.870 9	4.371 8
DT	0.011 0	0.769 3	0.817 5	3.553 4
R	0.077 2	0.856 0	0.880 4	4.222 1
WD	0.009 1	0.229 7	0.836 3	2.281 0
WS	0.011 5	0.269 2	0.829 9	2.229 7

曲线对比图。横轴表示时间步,纵轴表示 $PM_{2.5}$ 值。由图1可知,CRQA-DBN模型的预测曲线与实际 $PM_{2.5}$ 曲线偏离程度较大,这说明该模型的预测性能较差。CRQA-CNN-LSTM与CRQA-DBN-ELM的预测曲线与实际曲线的更为接近,即它们的预测性能较好。

际 $PM_{2.5}$ 值之间的误差越小。由图2可知,CRQA-DBN-ELM的预测结果最接近实际的 $PM_{2.5}$ 数据。

2.5 统计分析

为了进一步验证CRQA-DBN-ELM在 $PM_{2.5}$ 预测任务中的优越性,本文从统计学角度出发,采用散点图与分位比较(quantile-quantile, QQ)图对模型的非线性逼近性能进行分析。

表 3 模型评估

模型	RMSE	R^2
CRQA-SVM	6.370 4	0.940 2
CRQA-DBN	8.490 4	0.920 3
CRQA-ELM	4.623 3	0.956 6
CRQA-LSTM	4.610 3	0.956 7
CRQA-CNN-LSTM	2.832 4	0.973 4
CRQA-DBN-ELM	1.775 9	0.983 3

图 3 给出了 CRQA-SVM、CRQA-DBN、CRQA-ELM、

CRQA-LSTM、CRQA-CNN-LSTM 与 CRQA-DBN-ELM 在 PM2.5 预测任务中的散点图对比。它可以直观反应回归分析中 PM2.5 数据在平面中的分布。图 3 中 $y=x$ 的黑色直线表示基准线,当模型的预测点越接近基准线,表明相应模型的拟合性能越好。由图 3(a)和(b)可知,CRQA-DBN、CRQA-SVM 的预测值偏离基准线较远,即这两种模型的预测性能最差。此外,根据图 3(e)和(f),CRQA-CNN-LSTM 与 CRQA-DBN-ELM 模型的预测数据集中在 $y=x$ 基准线的附近,这表明 CRQA-DBN-ELM 与 CRQA-CNN-LSTM 的预测数据点最接近真实的 PM2.5 值,即上述模型的拟合性能较好。

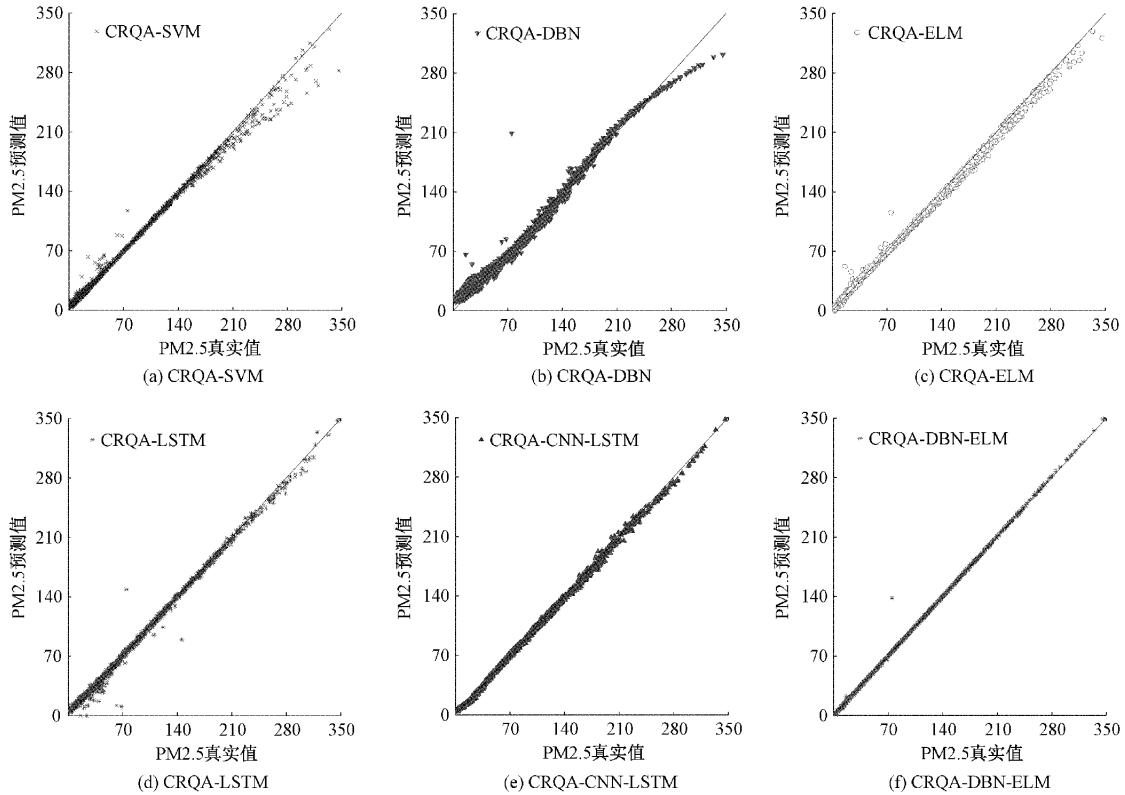
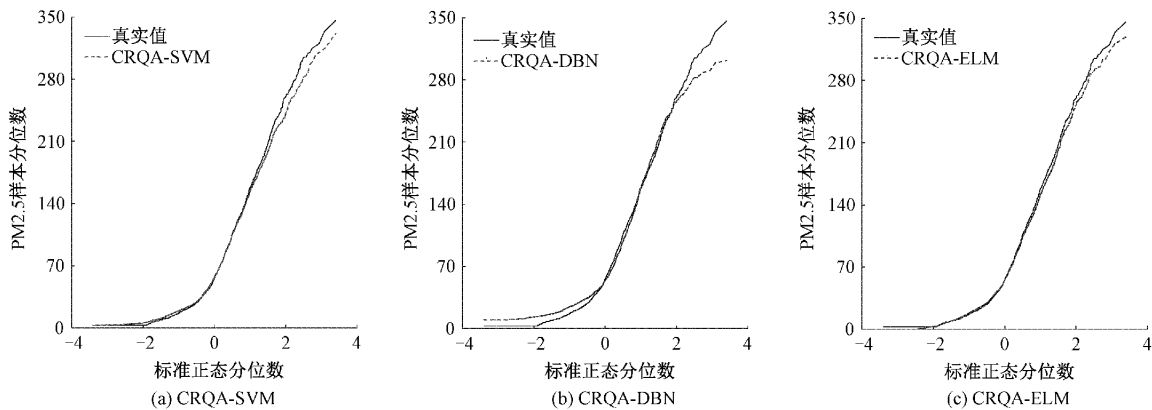


图 3 模型预测值与真实值散点图对比

图 4 给出了多个模型的预测数据与实际数据的 QQ 图对比。通过度量 PM2.5 的实际值与模型预测输出值的分位数分布,直观展示了预测数据与实际数据分布曲线之间

的差异。由图 4(f)可知,CRQA-DBN-ELM 仍是最优的预测模型。相比之下,CRQA-DBN 与 CRQA-SVM 预测结果的分位数更偏离真实数据分位数。



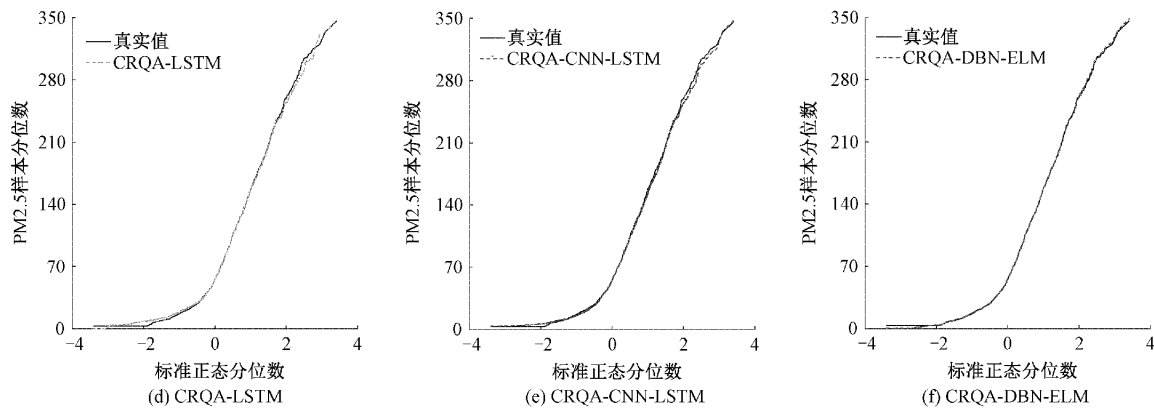


图4 模型预测值与真实值 QQ 图对比

3 结 论

本文提出了一种新的空气质量数据预测架构,该架构由相空间重构、CRQA 分析、DBN 特征提取与 ELM 非线性逼近四部分组成。其中,CRQA 能够探究空气质量影响因素间的动态相关性,挖掘对 PM_{2.5} 影响较大的因素,为后续预测提供优质先验数据。DBN-ELM 模型能够建模多维空气质量影响因素与 PM_{2.5} 的非线性关系。实验结果表明,对比 CRQA-SVM、CRQA-DBN、CRQA-ELM、CRQA-LSTM、CRQA-CNN-LSTM,该模型在空气质量数据预测中具有优越的拟合性能。未来的工作将关注于多站点空气质量影响因素关联性分析与预测等方法的构建。

参考文献

- [1] WANG J, SONG G. A deep spatial-temporal ensemble model for air quality prediction [J]. *Neurocomputing*, 2018, 314:198-206.
- [2] 高嵩,何卓骏,刘子岳,等.基于机器学习的差异融合分析在空气质量预测中的应用[J]. *电子测量技术*, 2021, 44(18):85-92.
- [3] GRENNFELT P, ENGLERYD A, FORSIUS M, et al. Acid rain and air pollution: 50 years of progress in environmental science and policy[J]. *Ambio*, 2020, 49(4):849-864.
- [4] 孙瑞玲,岳永丽,张国祯,等.主成分分析用于 PM_{2.5} 中氯代类污染物的研究[J]. *环境科学与技术*, 2020, 43(5):88-93.
- [5] WANG S, HUA G, ZHOU H. What are the key factors affecting air pollution? Research on Jiangsu, China from the perspective of spatial differentiation[J]. *Sustainability*, 2020, 12(6):2371-2386.
- [6] 侯素霞,张鉴达,李静.上海市大气污染物时空分布及其相关性因子分析[J]. *生态环境学报*, 2021, 30(6):1220-1228.
- [7] 张爱琳,白丽娜,韩液,等.基于多变量分数阶灰色

模型预测郑州市的空气质量[J/OL]. *安全与环境学报*: 1-14 [2021-12-02]. <https://doi.org/10.13637/j.issn.1009-6094.2021.0886>.

- [8] CHEN H, ZHANG L, ZOU W, et al. Regional differences of air pollution in China: Comparison of clustering analysis and systematic clustering methods of panel data based on gray relational analysis[J]. *Air Quality, Atmosphere & Health*, 2020, 13(10): 1257-1269.
- [9] YANG Y X, GAO Z K, WANG X M, et al. A recurrence quantification analysis-based channel-frequency convolutional neural network for emotion recognition from EEG[J]. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 2018, 28(8):1-9.
- [10] WANG P, WANG H, YAN R. Bearing degradation evaluation using improved cross recurrence quantification analysis and nonlinear auto-regressive neural network[J]. *IEEE Access*, 2019, 7:38937-38946.
- [11] BAI A, HIRA S, DESHPANDE P S. Recurrence based similarity identification of climate data [J]. *Discrete Dynamics in Nature and Society*, 2017, (2017): 1-21.
- [12] LI Z, CAI D, WANG J, et al. Machine learning based dynamic correlation on marine environmental data using cross-recurrence strategy [J]. *IEEE Access*, 2019, 7:185121-185130.
- [13] 左姣姣,倪志伟,朱旭辉,等.融合协同进化人工鱼群法和 SVM 的雾霾预测方法[J]. *模式识别与人工智能*, 2018, 31(8):725-739.
- [14] 李泽群,韦骏.利用人工智能神经网络预测广州市 PM_{2.5} 日浓度 [J]. *北京大学学报:自然科学版*, 2021, 57(4):645-652.
- [15] 武艳,杜景林,全力,等.一种基于 PCA-IPSO 的 BiLSTM 神经网络预测 PM_{2.5} [J]. *现代电子技术*, 2022, 45(1):119-124.

- [16] 李佩宜, 王鹏, 张羲海, 等. 基于深度信念网络的变频电机局部放电起始电压预测[J]. 仪器仪表学报, 2021, 42(4):121-130.
- [17] GUO R, MA Y, WANG S, et al. Establishment of air quality forecast model based on deep learning[C]. 2020 IEEE 6th International Conference on Computer and Communications(ICCC), IEEE, 2020:1500-1504.
- [18] 宋娟, 倪志伟, 李萍, 等. 基于参数优化深度置信网络的雾霾预测模型[J]. 系统科学与数学, 2020, 40(9):1644-1661.
- [19] 张杜娟, 王震. 基于 PSO-ELM 的疾病诊断方法研究[J]. 国外电子测量技术, 2021, 321(8):82-86.
- [20] 王毅, 李曙, 李松浓, 等. 瞬时特征下极限学习机在接地故障诊断中的应用[J]. 电子测量与仪器学报, 2022, 253(1):212-219.
- [21] ZHANG S, GUO B, DONG A, et al. Cautionary tales on air-quality improvement in Beijing[J]. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2017, 473:1-14.

作者简介

李志刚, 博士, 教授, 主要研究方向为人工智能、智能控制。

秦林林, 硕士研究生, 主要研究方向为递归图、机器学习。

付多民, 硕士研究生, 主要研究方向为深度学习、计算机视觉。

孙晓川, 博士, 副教授, 主要研究方向为储备池计算、深度学习。

E-mail: sunxiaochuan@ncst.edu.cn