

融合 2 维卷积与注意力以预测 $\text{PM}_{2.5}$ 浓度的 S-TCN 模型^{*}

李春辉¹ 张瑛琪² 孙 洁¹

(1. 华北理工大学电气工程学院 唐山 063210; 2. 河北医科大学第一医院(急诊科) 石家庄 050031)

摘 要:针对传统预测模型对 $\text{PM}_{2.5}$ 浓度预测精度较低、可解释性差的缺陷,提出一种融合 2 维卷积层(2D convolution)和注意力层的时空卷积网络预测模型(spatio-2D-temporal convolutional networks attention, S-2D-TCNA)。选取北京市 2014 年 5 月 1 日~2015 年 4 月 30 日的 36 个监测站点逐小时空气质量和气象数据,通过对多个站点时空相关性分析,将符合相关性阈值的监测站数据输入至卷积进行升维再降维的处理方式,得出具有时空序列的输入特征;将注意力融入时间卷积网络预测模型,用于预测未来 1 h 的中心监测站 $\text{PM}_{2.5}$ 浓度。在模型训练优化参数过程中,通过 Adam 来训练深度学习模型的参数,然后使用贝叶斯优化来调整模型的超参数,这种方法能找到模型的最佳参数,使其均方根误差、平均绝对误差分别减少 3.791% 和 5.576%,拟合优度增大 0.67%;在质量方面,所提出的 S-Conv2D-TCNA 模型均方根误差、平均绝对误差和拟合优度分别为 16.020 9、10.610 0 和 0.942 8,该预测模型在准确性和稳定性方面优于基线模型。结果表明,该预测模型空气污染的预警、区域预防和控制方面大有可为。

关键词:时空序列;注意力;时间卷积网络(TCN); $\text{PM}_{2.5}$ 浓度

中图分类号: X513 **文献标识码:** A **国家标准学科分类代码:** 510.4030

S-TCN model fusing 2D convolution and attention to predict $\text{PM}_{2.5}$ concentrations

Li Chunhui¹ Zhang Yingqi² Sun Jie¹

(1. North China University of Science and Technology, School of Electrical Engineering, TangShan 063210, China;

2. Emergency Department, The First Hospital of Hebei Medical University, Shijiazhuang 050031, China)

Abstract: Aiming at the shortcomings of the traditional prediction model for $\text{PM}_{2.5}$ concentration prediction with low accuracy and poor interpretability, a spatio-temporal convolutional network prediction model integrating 2D convolution and attention (S-2D-TCNA) layers is proposed. Hourly air quality and meteorological data from 36 monitoring stations in Beijing from May 1, 2014, to April 30, 2015, are selected. Through the analysis of spatio-temporal correlations among multiple stations, the monitoring station data that meet the correlation threshold are inputted into a convolutional network that adopts a dimensionality expansion and reduction approach to obtain input features with spatio-temporal sequences. Attention is incorporated into the Temporal convolutional network model for predicting the $\text{PM}_{2.5}$ concentrations at the central monitoring station for the next hour. In the process of optimizing the parameters for model training, the parameters of the deep learning model are trained by Adam, and then Bayesian optimization is used to adjust the hyper-parameters of the model, and this method finds the best parameters of the model, which reduces the root mean squared error and mean absolute error by 3.791% and 5.576%, respectively, and increases the goodness of fit by 0.67%. In terms of quality, the S-Conv2D-TCNA model has a mean root mean square error, mean absolute error and goodness of fit are 16.020 9, 10.610 0 and 0.942 8, respectively, and this prediction model is better than the baseline model in terms of accuracy and stability. The results show that the forecasting system is promising in the early warning, regional prevention, and control of air pollution.

Keywords: spatio sequence; attention; temporal convolutional network (TCN); $\text{PM}_{2.5}$ concentration

收稿日期:2023-08-01

^{*} 基金项目:2020 年河北省省级科技计划(20477703D)项目资助

0 引言

随着工业化和城市化的迅速发展,空气污染已成为特别重要的社会问题。中国北方的城市经常面临 $\text{PM}_{2.5}$ 高危预警,空气污染给人民的健康和环境造成了严重威胁。 $\text{PM}_{2.5}$ 是指空气动力学直径小于 $2.5\ \mu\text{m}$ 的粒子,是威胁健康和环境的主要污染物^[1-2]。这些微小的颗粒物可以在空气中悬浮很长时间,因此能够被远距离传播。 $\text{PM}_{2.5}$ 与呼吸道疾病、心血管疾病、癌症等多种健康问题之间存在密切关联,已被世界卫生组织列为致癌物质。因此,必须高度重视并采取有效的措施来应对 $\text{PM}_{2.5}$ 带来的健康风险。

$\text{PM}_{2.5}$ 的浓度预测大致分为两类,物理预测模型和数据驱动预测模型。在物理预测模型方面,这些模型主要依靠分析和挖掘大气的化学、物理成分与 $\text{PM}_{2.5}$ 浓度之间的关系。传统的物理模型在捕捉复杂的时空关系和环境变化方面可能存在一定的局限性。正因为如此,数据驱动的预测模型变得尤为重要。根据数据被构建的方式不同,以往的预测模型分为线性和非线性预测模型两大类。线性预测模型有 Guo 等^[3]使用地理和时间的加权回归模型具有更好的预测结果和 Moisan 等^[4]多元线性回归具有较好的预测效果和研究意义,同时多元线性回归被广泛应用。

基于非线性预测模型分为机器学习和深度学习。传统的机器学习算法是一种有效的解决方案,比如支持向量机已经得到了广泛的应用,Zimmerman 等^[5]使用随机森林以及 Yafouz 等^[6]使用人工神经网络,利用空气质量与其他因素之间的相关性,进而提升了模型的预测性能。

然而,随着大数据技术进入环境研究领域,逐渐积累了大量的监测数据,传统的机器学习模型对于如此巨量数据的预测性能有所下降。深度学习具有较强的泛化能力,可以处理大规模、高维度的数据,适用于复杂的空气污染预测问题。其中卷积神经网络(convolutional neural network, CNN)具有很强的数据特征提取能力,郭芳青等^[7]利用 CNN 优秀的空间信息的综合处理能力,实现多种情绪的高效准确识别对当前脑电信号情感计算。提高情绪识别准确率,且鲁棒性更强。循环神经网络(recurrent neural network, RNN)等结构对时空关系和时间序列数据进行建模,进一步提升了预测性能。Li 等^[8]通过使用长短时记忆模型(long short-term memory, LSTM)模型与支持向量回归(support vector machine, SVR)的模型对 1 h 的预测任务具有非常良好的预测表现,并且对多不同时间跨度的多尺度预测,也取得了令人满意的性能。

邢红涛等^[9]提出融合卷积和长短期记忆网络的预测模型,该模型在预测精度和泛化能力方面具有显著优势。但在处理大量数据时,LSTM 和 RNN 在训练期间需要大量内存。2019 年,Murali 等^[10]提出了时间卷积网络(temporal convolutional network, TCN),这是一种通用卷积架构,具有并行处理序列和预测的能力。TCN 能有效地从因果卷积的体系结构中获得关于所有感兴趣的历史观察

和变量的信息且仅基于过去时间段的信息进行预测,而不考虑未来时间段的信息。这种因果性约束确保了模型不存在未来信息泄漏到过去的情况。此外,为了应对深度学习网络中可能出现的梯度问题,TCN 巧妙地采用了残差网络,从而提高了网络的稳定性。总的来说,TCN 是一种具有广泛应用前景的卷积架构,为时间序列预测提供了强大的工具。

近年来,TCN 已经被应用在各种环境来预测时间序列且效果较好。高学金等^[11]通过使用堆栈降噪自编码网络提取非线性特征以及特征重构,并根据重构误差构造平方预测误差(squared prediction error, SPE)统计量作为故障状态特征。最后利用时间卷积网络实现状态趋势预测。Tang 等^[12]提出一种 TCN 加载气象条件来短期预测的模型。Samal 等^[13]提出了基于深度学习的 TCN 自动编码器结构构建时空污染预测模型。它捕捉气象因素和 $\text{PM}_{2.5}$ 污染物之间的时间相关性来估算 $\text{PM}_{2.5}$ 时间序列的缺失值并执行多步、多输出和多站点污染物预测,以提高训练效率。Samal 等^[14]提出使用 TCN 和去噪自编码器来解决非线性的大数据集。韩天翔等^[15]提出基于互补集合经验模态分解与时间卷积网络模型来预测受热面积灰的方法,都能够有效的提高模型预测精度。

然而,这些基于 TCN 的方法和研究不完善,仍有一些不足之处。1)当前研究的 TCN 模型通常专注于时间维度的建模,在同时捕捉时间和空间的多维特征方面仍有挑战。这意味着在某些应用场景下,模型可能无法充分考虑到空间上的相关性,从而导致预测结果的局限性。2)确定 TCN 模型内部的超参数值通常依赖于经验和试验,缺乏一种简便、科学的方法。这使得模型的调整和优化变得相对困难,而一种更系统的超参数优化方法可能有助于提高模型性能并加速研究进展。3)虽然 TCN 模型已经改进了多站点信息融合,但仍然存在一些限制。传统的模型通常只关注目标站点的历史数据,无法充分考虑到多站点之间的影响关系。这可能导致模型在处理复杂的多站点问题时受到限制,因为它们无法全面理解站点之间的相互作用,而这些相互作用可能对于准确的预测至关重要。

本文研究目标是克服上述研究的局限性,建立了融合 2 维卷积层(2D convolution)和注意力层的时空因果卷积网络预测模型(spatio-2D-temporal convolutional networks attention, S-2D-TCNA),用于空气污染的预测。结合了卷积神经网络和时序卷积网络的结构,二维卷积用于提取空间特征,而时序卷积网络用于处理时序数据。通过结合两者的优势,能够更好地捕捉数据中的空间和时序关系。在时序预测层引入注意力层,使模型能够更加关注重要的时间步或特征;为了更加高效地搜索和优化模型的内部参数组合,通过融合 Adam 优化器和贝叶斯优化对模型的内部参数调参,以此来提高模型的性能和准确性;S-2D-TCNA 预测模型不仅考虑了目标站点的历史特征信息,还考虑了其他站点的信息。通过科学的多站点信息融合,能够捕捉

到多个站点之间的相关性,使模型能够更好地理解站点之间的影响关系,提高模型预测性能。

1 实验材料与方法

1.1 数据来源

本文使用了微软研究院城市计算团队的空气质量预测数据集。这个数据集有许多输入变量,具有很高的质量和完整性。该数据集从2014年5月1日~2015年4月30日,北京市内36个监测站的278 023个空气质量观测实例和116 867个气象观测实例以及所有的数据都有经纬度的地理坐标(<https://www.microsoft.com/en-us/research/project/urban-air/?from=https://research.microsoft.com/en-us/projects/urbanair&type=exact>)。

本文研究对象是2014年5月1日~2015年4月30日北京市内36个监测站的空气质量检测数据(监测时间为1 h)其中包含 $\text{PM}_{2.5}$ 、 PM_{10} 、 NO_2 、 CO 、 O_3 、 SO_2 浓度值以及温湿度、压强和风速风向,并且数据都是当日某时刻的实时监测值。通过删除异常值,并通过线性插值填补缺失值,以减少对模型的干扰,使其更加稳定。依据36个监测站的经纬度,北京市36个监测站分布如图1所示。

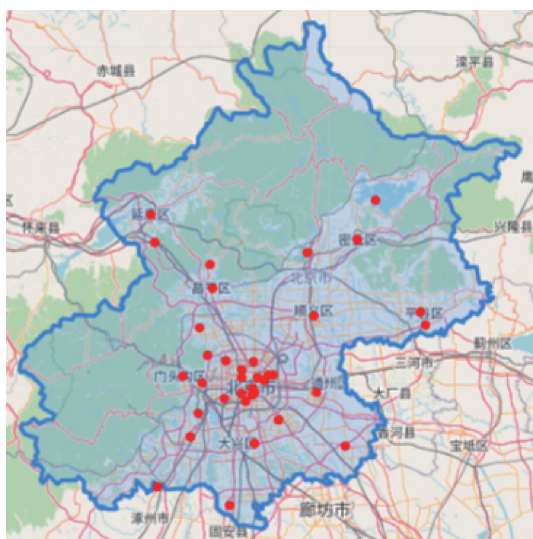


图1 北京市36个监测站分布

1.2 研究方法

1) 时空相关性分析

一个地区的空气污染在空间上取决于其他地区的污染。如果只考虑单个监测站的历史数据,模型的性能会受到限制。为了解决这个问题,必须对多个监测站点的数据进行时空相关性分析。可以通过计算 Pearson 相关系数来表示两个站之间的影响程度,将目标监测站和其他监测站之间的时间序列相关性,假设有 N 个监测站的历史数据集,输入的特征可以表示为一个三维向量。

$$\mathbf{X} \in \mathbf{R}^{N \times T \times M} \quad (1)$$

式中: N 表示有 N 个监测站的历史数据集; T 表示时间步

数和 M 表示与 $\text{PM}_{2.5}$ 浓度相关的环境特征。其中 \mathbf{X} 有 N 个特征矩阵的向量,分别为 $[\mathbf{X}_{S1}, \dots, \mathbf{X}_{Si}, \dots, \mathbf{X}_{SN}]$ 。其中 $\mathbf{X}_{S1} \in \mathbf{R}^{T \times M}$ 代表第1个监测站的特征矩阵。

首先用 Pearson 相关系数来计算目标站 \mathbf{Y}^* 与其他站 \mathbf{Y}_i 之间的 Pearson 相关系数,如下所示:

$$\rho(\mathbf{Y}^*, \mathbf{Y}_i) = \frac{\text{Cov}(\mathbf{Y}^*, \mathbf{Y}_i)}{(\sigma \mathbf{Y}^* \times \sigma \mathbf{Y}_i)} \quad (2)$$

式中: $\text{Cov}(\mathbf{Y}^*, \mathbf{Y}_i)$ 表示 \mathbf{Y}^* 和 \mathbf{Y}_i 协方差矩阵;其中 $\sigma \mathbf{Y}^*$ 和 $\sigma \mathbf{Y}_i$ 分别表示目标站 \mathbf{Y}^* 和其他站 \mathbf{Y}_i 的标准差。

在给定相关阈值的情况下,可以将所有其他监测站与目标监测站的相关性组织成一个相关向量:

$$\boldsymbol{\rho}^* = [\rho(\mathbf{Y}^*, \mathbf{Y}_1), \dots, \rho(\mathbf{Y}^*, \mathbf{Y}_i), \dots, \rho(\mathbf{Y}^*, \mathbf{Y}_N)] \quad (3)$$

其中, $\rho(\mathbf{Y}^*, \mathbf{Y}_i)$ 表示目标站 \mathbf{Y}^* 与第 i 个站 \mathbf{Y}_i 之间的 Pearson 相关系数。

然后,鉴于并非所有其他台站都对目标台站有明显的影响,设置一个相关度阈值来提取对目标监测站有相对较高影响的监测站数据是合理的,构成最终的输入特征向量。因此,最终的输入特征向量可以确定如下:

$$\mathbf{X}^* = \{\mathbf{X}_i \mid \rho(\mathbf{Y}^*, \mathbf{Y}_i) > \rho_{th}, i \in 1, \dots, N\} \quad (4)$$

这意味着只有当与目标站的相关性高于阈值 ρ_{th} 的监测站数据才会被保留在输入特征向量 \mathbf{X}^* 中,代表目标监测站的时空信息。确定了输入特征向量 \mathbf{X}^* 之后,它可以作为所提出的模型的输入,用于预测目标站的 $\text{PM}_{2.5}$ 浓度或进行其他相关任务。这样可以减少不具有明显影响的监测站数据对模型的干扰,提高模型的预测性能和解释能力。

2) 注意力机制

注意力机制起源于对人类视觉的研究。人类有选择地将有限的注意力集中在最重要的事情上看到的,忽略那些被认为无用的信息。因此,注意力机制的核心思想是搜索原始数据,从而忽略不相关的噪声并突出显示权重较高的信息。

该预测模型将注意力机制融入 TCN 中,如图2所示,将注意力机制融入 TCN 的最后一层,提高了模型的可解释性和数值预测准确性。

3) 因果卷积

卷积神经网络是一种常用于处理时间序列数据的模型,特别适用于具有时间依赖关系的序列数据。CNN 基于因果卷积的原理,每个输出时间步长仅依赖于当前和过去的输入时间步,而不受未来时间步的影响,从而保持因果性。这个概念在时间序列领域广泛应用,并在各种任务中展现出优异的性能,如时间序列预测、语音识别和自然语言处理等领域。

本文在该模型中加入了一个因果卷积结构(一种特殊的 CNN 结构),来处理序列建模问题。因果卷积网络使用多层全卷积网络结构。与 CNN 不同的是,在因果卷积层中,早于特定时间的输入特征被用于卷积。

$$F(s) = (f * x)(s) = \sum_i^{k-1} f(i) \cdot x_{s-d \cdot i} \quad (5)$$

式中: s 是序列的元素; d 是扩张参数; $f(i)$ 是滤波函数 f 在位置 i 处的权重; $x_{(s-i)}$ 表示输入序列 x 在过去时间步 $s-i$ 处的值; 以及 $s-d \cdot i$ 表示过去的方向。因果卷积通过在卷积操作中使用过去时间步的输入值来确保因果性, 不受未来时间步的影响。正常的卷积操作可以被视为因果卷积的特例, 即当扩张参数 $d=1$ 时, 因果卷积退化为正常的卷积操作。

多层全卷积网络通常具有多个卷积层, 通过堆叠这些层来增加网络的深度。每个卷积层的输出作为下一层的输入, 逐渐提取更高级别的特征表示。最后的卷积层输出被用作最终的预测结果或传递给其他任务。研究多次使用因果卷积网络进行数据传递和处理。

4) 基于时空序列分析的 TCNA 模型

将 36 个监测站的时空数据进行时空相关性处理, 将部分时空相关性高于 0.85 的监测站数据给定输入特征

\mathbf{X}^* , 是包含不同站点的时空数据的 3D 向量, 在多个站点的维度上集成 \mathbf{X}^* 的输入信息是直观的。本文首次提出了用于深度神经网络的 1×1 卷积, 增加和减少输入特征的维数。TCNA 模型也使用 1×1 卷积来实现增加和减少输入特征的维数。图 2 中, 以第 1 个时间步长的第 1 个特征值 (用蓝色标记) 为例。这组值可以看作是一个 1×1 的多通道信息, 其中通道的数量等于选取的监测站数。使用多个 1×1 的卷积核扩展这个多通道信息的维度, 再使用一个 1×1 的卷积核将维度降低为 1。通过这个方法, 不仅可以通过增加维度提高对多个监测站信息的特征提取能力, 还可以将输入数据聚合为 2D 特征向量, 与第 2 部分因果网络的输入兼容。其他研究也表明, 1×1 卷积可以看作是一个维度增加或减少的模块, 可以有效地提高模型容量 (增加非线性) 并消除计算瓶颈。接着被因果卷积处理后的输入特征再输入至 TCN 和注意力机制融合预测模块进一步处理。TCNA 模型的框架如图 3 所示。

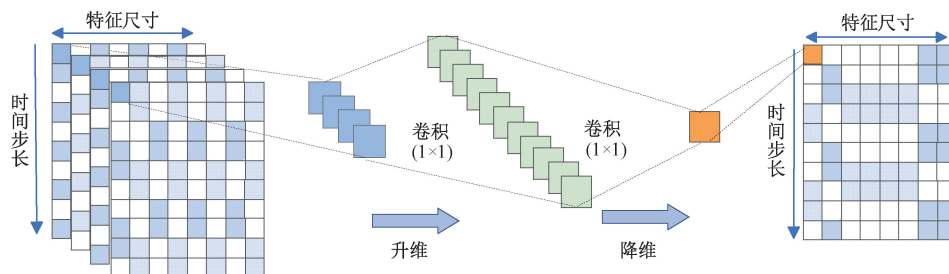


图2 基于时空序列的 36 个监测站数据特征整合

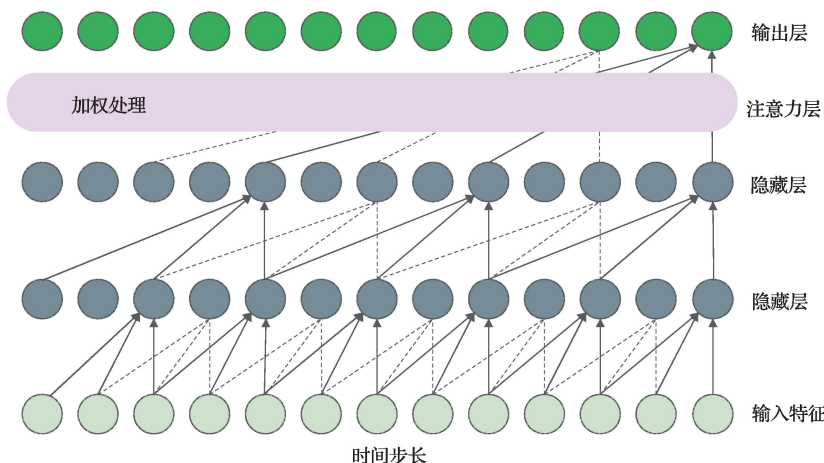


图3 TCNA 模型的框架

首先, 因果卷积将输入层特征进行二维卷积操作, 提取空间序列特征 (通过“先升维再降维”的方法)。经过二维卷积层的输出将作为下一层的输入, 并逐渐提取更高级别的特征表示。然后将注意力机制融入 TCN 模型: 融合模块会将 TCN 输出的特征与注意力权重进行逐元素的乘积操作。这样注意力层可以引导 TCN 模型更加关注对当

前任务有益的特征, 增强模型对重要信息的表达能力。输入特征处理流程如图 4 所示。

5) 实验设置

数据集分为 3 部分, 训练集、验证集和测试集。基于时间均匀分布的原则, 以确保验证的可靠性。训练集涵盖了每个月前 60% 的数据, 验证集涵盖了每个月后 20% 的



图4 S-2D-TCNA模型输入特征传递处理

数据。剩余的20%被保留为测试集。这种数据划分策略的目的在于保证验证方法的可靠性。最后选择均方误差作为训练过程的损失函数。

通过首先使用Adam优化器(训练时期和批量大小分别设置为200和32)对模型的随机目标函数进行梯度下降优化,确定关键的4个参数的值。随后扩大4个参数的优化范围,使用贝叶斯优化建立目标函数的概率模型来找到模型的最优超参数。它通过构建先验分布和不断更新后验分布来进行参数搜索。定义一个字典pbounds,用于指定超参数的搜索范围。

为了验证该实验模型性能的优越性,则选择了在PM_{2.5}浓度预测被广泛应用的LSTM、CNN、TCN、S-TCN和S-2D-TCNA模型。最后通过均方根误差(RMSE)、平均绝对误差(MAE)和决定系数 R^2 作为模型误差评估指标来评估PM_{2.5}预测的性能。当模型评估值RMSE和MAE越小, R^2 的值越大时,模型预测准确性越高。

2 实验结果与分析

2.1 时空相关性分析

空气质量监测是保护公众健康、改善城市环境的重要手段。时空分析是评估监测站数据之间关联性的有效方法,可用于预测和优化空气质量监测网络的布局和数据分析方法。研究旨在通过对北京市36个监测站的PM_{2.5}数据进行时空相关性分析,评估不同阈值设定对时空关系的影响,并找出最佳阈值以提高预测能力。

选择中心监测站点时,必须考虑其数据的完整性,确保中心点监测站的数据记录完整、准确。这可以提高时空相关性分析的可靠性和稳定性,因此选择1013作为中心点。中心站点1013在北京的分布如图5所示。

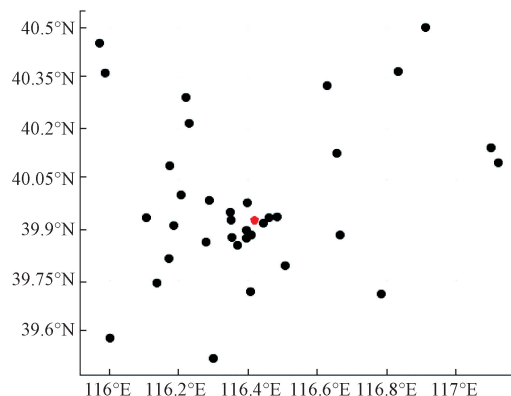


图5 北京监测站分布

根据不同的阈值分析,根据表1,不同的阈值对S-2D-TCNA模型的性能影响较大,显而易见当阈值为0.85时,相关监测站为19个站点时,S-2D-TCNA模型的性能最优。所以在以下训练中,将相关性阈值默认为0.85。

表1 不同的阈值对S-2D-TCNA模型的性能影响

阈值	相关监测站数量	RMSE	MAE	R^2
0.70	34	18.010	12.390	0.932
0.80	27	19.030	13.210	0.929
0.85	19	16.020	10.610	0.942
0.90	13	18.088	13.140	0.925
0.95	3	18.796	13.421	0.919

2.2 融合Adam优化和贝叶斯优化的参数优化与分析

首次提出融合Adam优化器及贝叶斯优化在训练模型中,首先使用Adam优化器优化参数,其内部参数随机种子为314、学习率 1×10^{-3} 。由于隐含层维数(hidden_size)、堆叠的层数(levels)、卷积核大小(kernel_size)和随机失活(dropout)这4个参数在模型训练中扮演着至关重要的角色。因此需要找出4个参数的合理范围,首先使用Adam优化器优化后,得到隐含层维数、堆叠的层数、卷积核大小和随机失活这4个参数的最佳参数组分别为32、4、4和0.25。所以贝叶斯优化的参数组中隐含层维数应设置在16~48,堆叠的层数应设置在2~5,卷积核的大小应设置在2~5和随机失活率应设置在0.1~0.5。

先通过Adam优化器优化,再通过构建贝叶斯优化器对模型参数进行自动寻优,以高效、精确地捕获最佳参数组合,优势在于其自适应地探索参数空间并充分利用历史信息。通过贝叶斯优化方法,将均方根误差作为损失函数,隐含层维数、堆叠的层数、卷积核大小和随机失活这4个参数经过多次迭代优化,看均方根误差的上下起伏的变化而得出最佳参数。如图6~9所示,每一个贝叶斯优化周期变化,4个参数也在变化,当均方根误差在最低点时,

此时的4个参数值就是最佳参数组,并且最佳参数组能够提高模型的预测精度。如图10所示,当贝叶斯优化周期(epoch)为15时,损失函数(RMSE)最低,此时隐含层维数、堆叠的层数、卷积核大小和随机失活这4个参数的最佳参数组的值分别25、2、2和0.1。

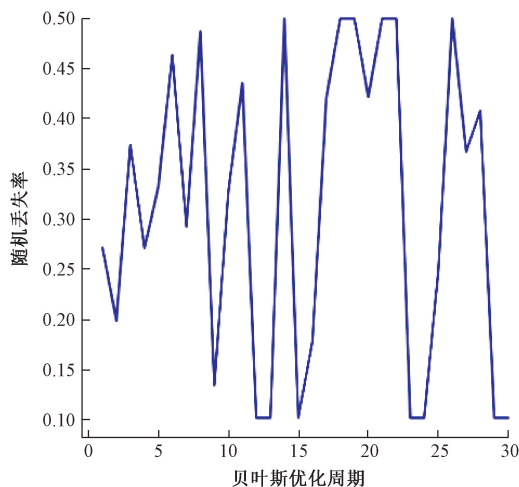


图6 参数(随机丢失率)曲线

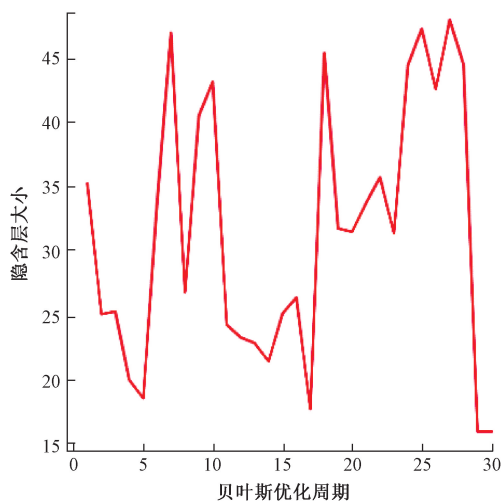


图7 参数(隐含层大小)曲线

Adam 优化器的隐含层维数、堆叠的层数、卷积核大小和随机失活4个参数的最佳参数组分别为32、4、4和0.25。融合Adam优化和贝叶斯优化的隐含层维数、堆叠的层数、卷积核大小和随机失活这4个参数最佳参数组分别为25、2、2和0.1。融合Adam优化和贝叶斯优化的参数组使其均方根误差、平均绝对误差分别减少3.791%和5.576%，拟合优度增大0.67%。表2是两种参数组的对比,可以看出RMSE、MAE和 R^2 得到明显的改进。

在该领域中,许多权威的研究和实践已经证明了贝叶斯优化在参数调优中的有效性。贝叶斯优化能够在有限的迭代次数内找到最佳解,并具备对高维参数空间的探索能力。然而若设置参数时将其取值范围设的太大,会占用

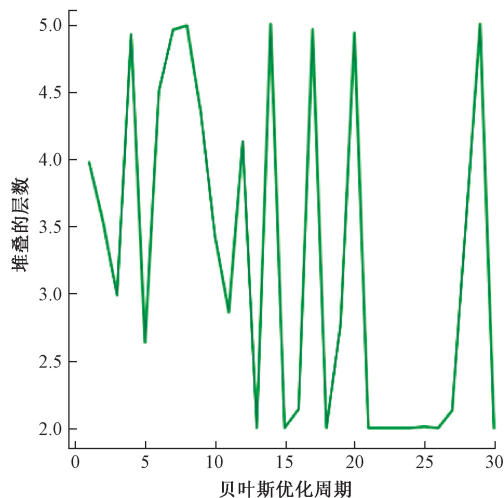


图8 参数(堆叠的层数)曲线

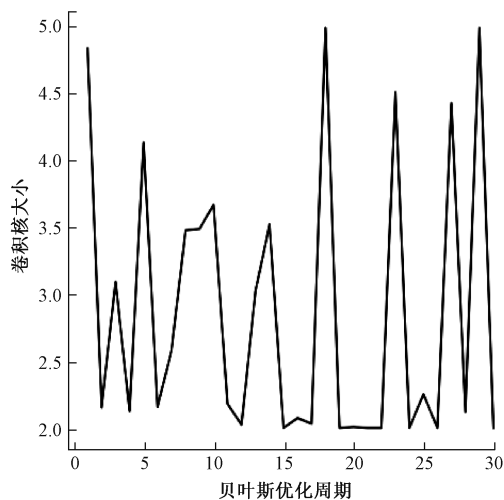


图9 参数(卷积核大小)曲线

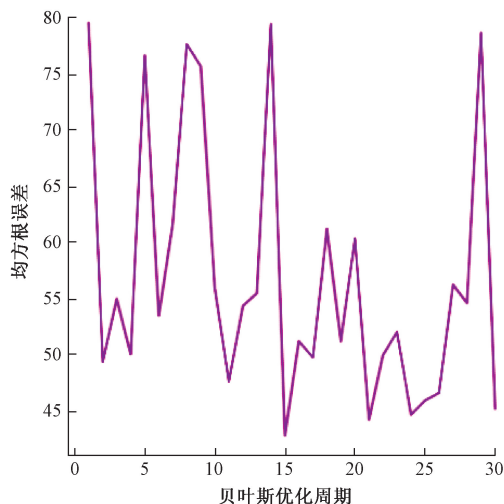


图10 损失函数(RMSE)曲线

计算机的运行内存且用时太长;若设置参数时将其取值范

表2 两种机器学习参数优化对比

评估指标	Adam 优化器 最佳参数组	贝叶斯优化 最佳参数组
RMSE	16.652 3	16.020 9
MAE	11.236 5	10.610 0
R^2	0.936 5	0.942 8

围设的太小,得到的参数组反而会降低模型的预测精度。因此通过融合 Adam 优化器和贝叶斯优化来调优参数,并结合将均方根误差作为评估指标,能够获得更准确、稳健的模型。这将有助于更好地理解 and 解释模型预测可靠性,并为进一步的研究和决策提供可靠的基础。

2.3 评价指标

3个评价指标,即 MAE、RMSE 和 R^2 用于评价 $PM_{2.5}$ 的预测性能,计算如下:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_{true}^i - y_{pred}^i| \quad (6)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_{true}^i - y_{pred}^i)^2} \quad (7)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_{true}^i - y_{pred}^i)^2}{\sum_{i=1}^N (y_{mean} - y_{pred}^i)^2} \quad (8)$$

MAE 和 RMSE 是评价绝对误差的评价指标,数值越小,模型性能越好。 R^2 的值越接近于 1,预测结果越接近实际值^[16]。

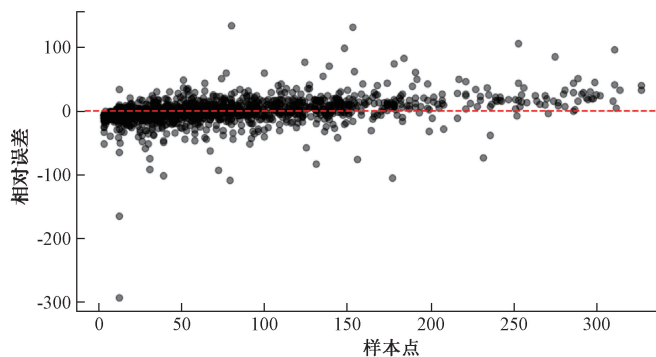
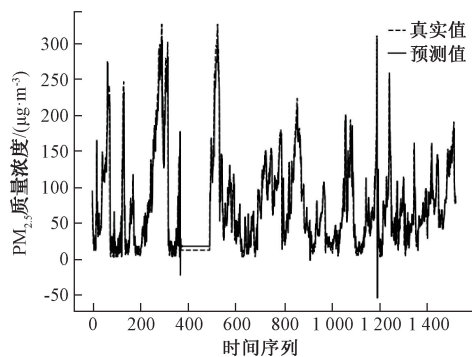


图11 CNN网络模型的预测结果

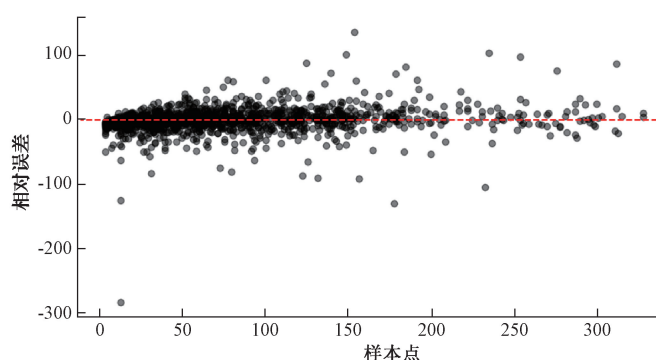
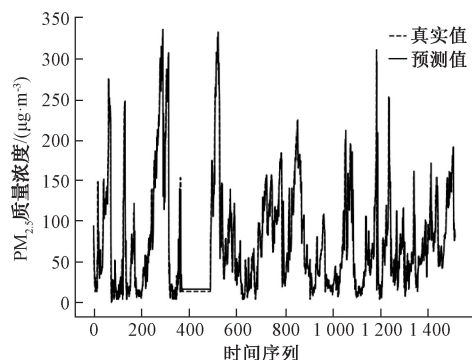


图12 LSTM网络模型的预测结果

2.4 消融实验

消融实验是一种可控变量方法。通过逐步消融某一变量或因素,可以评估每个变量对结果的贡献,从而确定哪些因素对研究问题最为重要。如改变深度学习内部层来研究改进的必要性。

为了获得更好的预测精度,S-2D-STCNA 模型融入了 2 维卷积层和注意力层,如表 3 所示。在没有注意力层的情况下,模型的 RMSE、MAE 和 R^2 都低于 S-2D-STCNA;在没有 2 维卷积层来处理时空输入特征的情况下,模型的 RMSE、MAE 和 R^2 都低于 S-2D-STCNA。因此融入注意力层和 2 维卷积层可以提高模型的平滑度和预测的准确性。

表3 S-2D-STCNA 模型消融实验

模型	RMSE	MAE	R^2
S-2D-TCNA	16.020 9	10.610 0	0.942 8
注意力层	16.860 1	11.657 9	0.936 5
2 维卷积层	17.615 1	12.317 2	0.926 1
TCN	20.926 6	14.129 9	0.913 2

2.4 模型预测与对比

将本文 S-2D-TCNA 与 STCN、TCN、LSTM、CNN 这 4 种预测模型进行预测结果对比。将 RMSE、MAE 和 R^2 作为评估 4 种不同的 $PM_{2.5}$ 浓度预测模型。

预测结果如图 11~15 所示,S-2D-TCNA 模型在预测目标站的 $PM_{2.5}$ 浓度中取得良好的表现,与 LSTM、CNN、

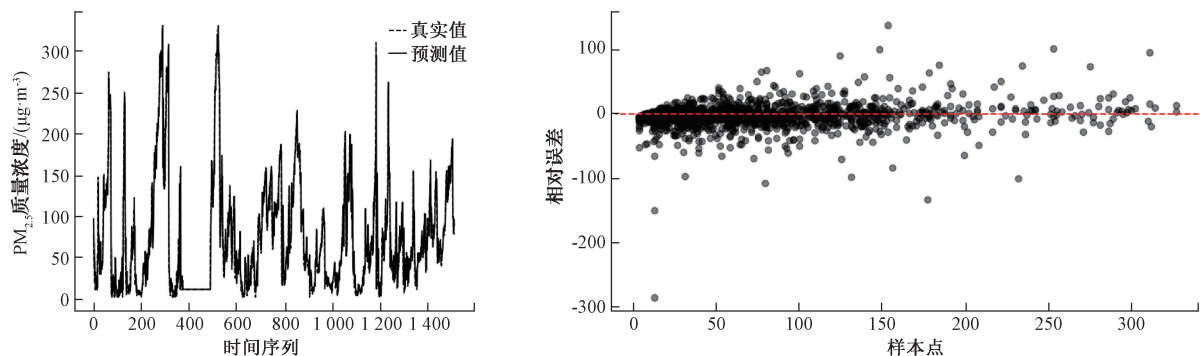


图 13 TCN 网络模型的预测结果

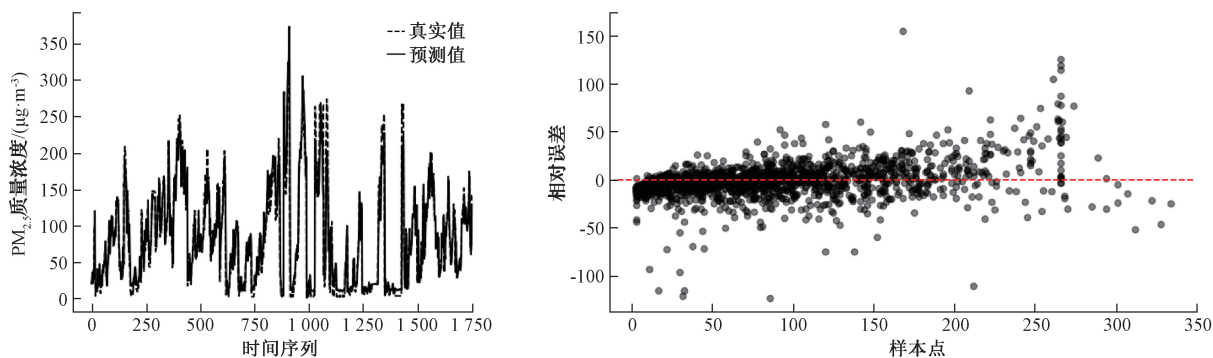


图 14 S-TCN 网络模型的预测结果

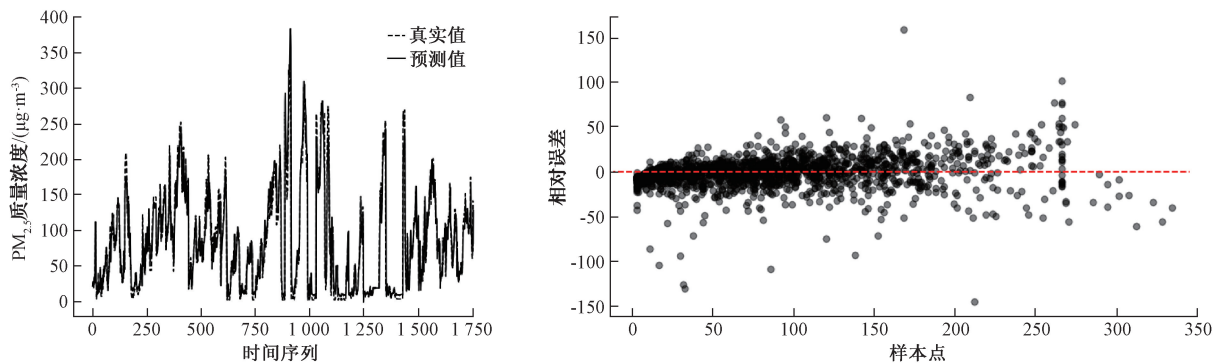


图 15 S-2D-TCNA 网络模型的预测结果

TCN 和 S-TCN 4 种模型比较, S-2D-TCNA 的 RMSE 分别降低了约 34.81%、33.26%、23.44% 和 13.44%, 在 MAE 上分别降低了约 37.43%、34.17%、24.91% 和 11.79%, 此外, S-2D-TCNA 模型相较于 LSTM、CNN、TCN 和 S-TCN 4 种模型在 R^2 值上分别提高了约 3.74%、4.42%、3.14% 和 1.59% 的拟合优度。

LSTM、CNN、TCN 和 S-TCN 4 种模型与 S-2D-TCNA 模型对比分析, 可以看出, S-2D-TCNA 模型具有更好的预测性能, 更容易捕捉非线性非平稳的时间序列, 并能够对实际 $PM_{2.5}$ 浓度值更精准的预测。如表 4 所示。

表 4 4 种模型的性能结果对比

预测模型	RMSE	MAE	R^2
LSTM ^[8]	24.574 2	16.957 3	0.907 5
CNN ^[17]	24.153 0	16.116 6	0.901 1
TCN ^[18]	20.926 6	14.129 9	0.913 2
S-TCN ^[19]	18.515 6	12.032 7	0.928 1
S-2D-TCNA	16.020 9	10.610 0	0.942 8

3 结论

TCN 模型是一个强大的时间序列预测模型,但在空气污染预测这一领域的研究数量较少;传统模型在捕捉时空关系、特征提取和数据融合方面的方法较为单一。为了补充这些空缺,本文设计了 S-2D-TCNA 模型。

1) TCN 模型内部超参数的取值通常依赖于经验和试验,缺乏一种简便、科学的方法。本文首次提出融合 Adam 优化器与贝叶斯优化的方法,共同优化神经网络内部参数。在模型训练参数优化过程,先通过 Adam 优化器得出模型的内部参数值,将 Adam 优化器的结果作为贝叶斯优化器的输入范围。贝叶斯优化器在每次迭代中评估不同参数组合下的模型性能,并将结果作为反馈信息,贝叶斯优化算法可以逐步搜索到更优的参数组合,提高模型 S-2D-TCNA 的预测精度。

2) 在某些应用场景下,模型可能无法充分考虑到空间上的相关性,从而导致预测结果的局限性。为了解决这个问题,该模型利用了多个卷积层和注意力层,二维卷积层将时空分析结果作为新的输入特征,采用“先降维再升维”的方法提取有用的特征,时间卷积再将输入特征层层提取和加权,以此捕捉输入数据的空间特征和时间序列的时空和时序关系。

3) 传统的模型通常只关注目标站点的历史数据,无法充分考虑到多站点之间的影响关系。而 S-2D-TCNA 预测模型不仅考虑目标站点的历史特征信息,还考虑其他站点的信息。通过科学的多站点信息融合,能够捕捉到多个站点之间的相关性。这使模型能够更好地理解站点之间的影响关系,识别和利用站点之间的共同特征(即使输入数据的某些站点的数据质量较差或有缺失),从而提高模型的精度和鲁棒性,提高模型预测的准确性。

综上所述,S-2D-TCNA 模型的核心思路在于充分采用 2 维卷积和结合 Adam 优化与贝叶斯优化的参数优化策略,以及高效的特征提取和多站点信息融合;同时利用注意力机制,提高模型对时间特征的关注度和表达能力。这些创新性技术和设计使得 S-2D-TCNA 模型在空气污染预测任务中具有先进性和必要性。通过对复杂的空气污染数据准确建模和预测,该模型为环境保护和公共健康领域的决策制定者提供了重要的工具和参考依据。

本文所提 S-2D-TCNA 模型,为相关领域的学者和研究人员提供了使用 TCN 模型预测空气污染的新思路和方法。在未来的研究中,学者和研究人员可以进一步探索 S-2D-TCNA 模型在其他环境数据预测任务的应用研究,进一步优化和改进该模型的性能。这将推动深度学习在环境科学和大气环境领域的应用研究,有助于更好地保护人类健康并提高生活质量。

参考文献

- [1] 朱菊香,谷卫,任明煜,等. 基于 SWT-ISSA-LSTM 的地铁空气质量预测建模[J]. 国外电子测量技术, 2023,42(7):164-174.
- [2] 刘恩海,付英健,张智,等. 联合 TRANSFORMER 注意力机制的 PM_{2.5} 浓度预测网络研究[J]. 安全与环境学报, 2023,23(10):3760-3768.
- [3] GUO B, WANG X, PEI L, et al. Identifying the spatiotemporal dynamic of PM_{2.5} concentrations at multiple scales using geographically and temporally weighted regression model across China during 2015-2018[J]. Science of the Total Environment, 2021, 751: 141765.
- [4] MOISAN S, HERRERA R, CLEMENTS A. A dynamic multiple equation approach for forecasting PM_{2.5} pollution in Santiago, Chile [J]. International Journal of Forecasting, 2018, DOI: 10.1016/j.ijforecast.2018.03.007.
- [5] ZIMMERMAN N, PRESTO A A, KUMAR S P N, et al. A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring [J]. Atmospheric Measurement Techniques, 2018,11(1):291-313.
- [6] YAFOUZ A, AHMED A N, SHERIF M, et al. Comprehensive comparison of various machine learning algorithms for short-term ozone concentration prediction[J]. Alexandria Engineering Journal, 2021, DOI: 10.1016/j.aej.2021.10.021.
- [7] 郭芳青,赵丽,张志雯,等. 基于交错组卷积神经网络的脑电信号情绪识别研究[J]. 国外电子测量技术, 2022,41(8):112-117.
- [8] LI X, PENG L, YAO X J, et al. Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation[J]. Environmental Pollution, 2017,231(P1):997-1004.
- [9] 邢红涛,郭江龙,刘书安,等. 基于 CNN-LSTM 混合神经网络模型的 NO_x 排放预测[J]. 电子测量技术, 2022,45(2):98-103.
- [10] MURALI S, SWAPNA T R. An empirical evaluation of temporal convolutional network for offensive text classification[J]. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 2019,8(8):2179-2184.
- [11] 高学金,马东阳,韩华云,等. 基于 DAE 和 TCN 的复杂工业过程故障预测[J]. 仪器仪表学报, 2021, 42(6):140-151.
- [12] TANG X L, CHEN H X, XIANG W H, et al. Short-term load forecasting using channel and temporal attention based temporal convolutional network [J]. Electric Power Systems Research, 2022,205: 107761.
- [13] SAMAL K K R, BABU K S, DAS S K. Multi-output

- TCN autoencoder for long-term pollution forecasting for multiple sites[J]. Urban Climate, 2021, DOI: 10.1016/j.uclim.2021.100943.
- [14] SAMAL K K R, BABU K S, DAS S K. Temporal convolutional denoising autoencoder network for air pollution prediction with missing values[J]. Urban Climate, 2021, DOI:10.1016/j.uclim.2021.100872.
- [15] 韩天翔,史元浩,李孟威,等. 融合 CEEMD 和 TCN 的受热面积灰预测研究[J]. 电子测量与仪器学报, 2022, 36(10):108-114.
- [16] 张文帅,王占刚. 基于改进麻雀算法优化深度极限学习机的缺失数据预测[J]. 电子测量技术, 2022, 45(15):63-67.
- [17] YAN R, LIAO J Q, YANG J, et al. Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering[J]. Expert Systems with Applications, 2020, 169(4):114513.
- [18] NI S, JIA P F, YANG X, et al. Prediction of CO concentration in different conditions based on Gaussian-TCN[J]. Sensors and Actuators: B Chemical, 2023, 376(12): 133010.
- [19] 林涛,吉萌萌,付崇阁,等. 基于改进时间卷积网络的空气质量预测研究[J]. 计算机仿真, 2022, 39(10): 451-456, 501.

作者简介

李春辉, 硕士研究生, 主要研究方向为大数据挖掘。

E-mail: 584576303@qq.com

张瑛琪(通信作者), 主任医师, 主要研究方向为急诊医学。

E-mail: zhangyingqi@hebm. edu. cn

孙洁, 博士, 教授, 主要研究方向为控制理论与控制工程、测控网络。

E-mail: 13931590668@139.com