

# 基于注意力机制和高斯概率估计的 PM<sub>2.5</sub> 浓度预测<sup>①</sup>



杨 柳<sup>1,2</sup>, 张 镒<sup>1,2</sup>, 李建熹<sup>3</sup>, 康丽荣<sup>3</sup>, 赵思彤<sup>4</sup>

<sup>1</sup>(中国科学院 沈阳计算技术研究所, 沈阳 110168)

<sup>2</sup>(中国科学院大学, 北京 100049)

<sup>3</sup>(辽宁省生态环境监测中心, 沈阳 110161)

<sup>4</sup>(沈阳市第二十二中学, 沈阳 110075)

通信作者: 张 镒, E-mail: zhangdi@sict.ac.cn

**摘 要:** 作为衡量空气污染物浓度的重要指标, 对 PM<sub>2.5</sub> 浓度进行监控预测, 能够有效地保护大气环境, 进一步地减少空气污染带来的危害. 随着空气质量自动监测站的大范围建立, 由传统的机器学习搭建的空气质量预测模型已经不能满足当今的需求. 本文提出了一种基于多头注意力机制和高斯概率估计的高斯-注意力预测模型, 并对沈阳市某监测站点的数据进行了训练和测试. 该模型考虑了 PM<sub>2.5</sub> 浓度受到其他空气质量数据的影响, 将空气质量数据的分层时间戳(周、日、小时)的信息对齐作为输入, 使用多头注意力机制对于不同子空间的时间序列关联特征进行提取, 能够获得更加完善有效的特征信息, 再经过高斯似然估计得到预测结果. 通过与多种基准模型进行对比, 相较于性能较优的 DeepAR, 高斯-注意力预测模型的 MSE、MAE 分别下降了 21%、15%, 有效地提高了预测准确率, 能够较准确地预测出 PM<sub>2.5</sub> 浓度.

**关键词:** PM<sub>2.5</sub>; 高斯分布; 注意力机制

引用格式: 杨柳, 张镒, 李建熹, 康丽荣, 赵思彤. 基于注意力机制和高斯概率估计的 PM<sub>2.5</sub> 浓度预测. 计算机系统应用, 2023, 32(7): 305-311. <http://www.c-s-a.org.cn/1003-3254/8808.html>

## PM<sub>2.5</sub> Concentration Prediction Based on Attention Mechanism and Gaussian Probability Estimation

YANG Liu<sup>1,2</sup>, ZHANG Di<sup>1,2</sup>, LI Jian-Xi<sup>3</sup>, KANG Li-Xing<sup>3</sup>, ZHAO Si-Tong<sup>4</sup>

<sup>1</sup>(Shenyang Institute of Computing Technology, Chinese Academy of Sciences, Shenyang 110168, China)

<sup>2</sup>(University of Chinese Academy of Sciences, Beijing 100049, China)

<sup>3</sup>(Liaoning Ecological and Environmental Monitoring Center, Shenyang 110161, China)

<sup>4</sup>(Shenyang No.22 Middle School, Shenyang 110075, China)

**Abstract:** PM<sub>2.5</sub> is an important indicator for measuring the concentration of air pollutants, and monitoring and predicting its concentration can effectively protect the atmospheric environment and further reduce the harm caused by air pollution. As automatic air quality monitoring stations are constructed on a large scale, the air quality prediction model built by traditional machine learning can no longer meet the current needs. This study proposes a Gaussian-attention prediction model based on the multi-head attention mechanism and Gaussian probability estimation and utilizes the data from a monitoring station in Shenyang for training and tests. Because PM<sub>2.5</sub> concentration is affected by other air quality data, this model uses the information alignment of hierarchical time stamps (week, day, and hour) of air quality data as input and extracts the time-series correlation features of different subspaces with the multi-head attention mechanism. More complete and effective feature information is thereby obtained, and prediction results are then acquired by Gaussian

① 收稿时间: 2022-01-14; 修改时间: 2022-02-15, 2022-02-24, 2022-03-21; 采用时间: 2022-04-20; csa 在线出版时间: 2022-09-01

CNKI 网络首发时间: 2022-11-17

likelihood estimation. A comparison with multiple benchmark models is conducted, and the mean squared error (MSE) and mean absolute error (MAE) of the proposed Gaussian-attention prediction model are respectively 21% and 15% lower than that of the DeepAR model. Effectively improving prediction accuracy, the proposed model can accurately predict  $PM_{2.5}$  concentration.

**Key words:**  $PM_{2.5}$ ; Gaussian distribution; attention mechanism

## 1 引言

随着经济和社会的快速发展,大量空气污染物排放到大气中,造成严重的大气污染,对人体造成严重的危害.近些年来,我国政府加大了生态文明建设的投资,加强了生态环境保护的力度,在全国各个地区搭建空气质量监测站,用来监控空气污染物( $PM_{2.5}$ ,  $PM_{10}$ ,  $NO_2$ 等)质量数据. $PM_{2.5}$ 作为衡量空气污染物浓度的重要指标,对其浓度进行准确的监控预测,能够减少对人们身健健康的侵害,进一步推进大气污染物监控治理工作,能够有效地保护大气环境<sup>[1,2]</sup>.

当前  $PM_{2.5}$  浓度预测的方法主要有传统机器学习和深度学习.传统机器学习方法主要包括移动平均自回归 (autoregressive integrated moving average model, ARIMA)、支持向量机 (support vector machine, SVM)、随机森林 (RF)、集成学习、智能优化算法等.严宙宁等<sup>[3]</sup>利用 ARIMA 模型对深圳日平均  $PM_{2.5}$  浓度进行预测,得到较好的预测结果.石峰等<sup>[4]</sup>提出群智能灰狼最优化算法,模型具有较好的泛化能力、预测能力和鲁棒性.尹建光等<sup>[5]</sup>采用小波分解与最小二乘支持向量回归相结合的方法对  $PM_{2.5}$  进行建模预测,提高预测精度,特别是对于浓度变化较大的点预测精度较高.然而,上述传统方法都采用  $PM_{2.5}$  日均浓度作为输入,现实监测站是以小时为单位检测记录  $PM_{2.5}$  浓度,经过加工的日均浓度数据损失了原有的数据信息,得到的预测结果准确度较低<sup>[6]</sup>.此外,ARIMA、SVM、智能优化算法等受到算法本身的限制,只能对  $PM_{2.5}$  浓度进行短期预测,缺乏了对时间序列数据周期性、趋势性等长期特征的捕获,预测效果也会受到影响<sup>[7]</sup>.

近些年来,随着数据量的激增、计算机算力的大幅提升,深度学习在处理复杂时间序列的问题上取得较大的进步<sup>[8]</sup>.深度学习主要的网络模型有卷积神经网络 (convolutional neural network, CNN)、循环神经网络 (recurrent neural network, RNN)、对抗生成网络 (generative adversarial network, GAN) 等.在解决长时依赖的问题上,许多研究者都使用 RNN 的变体结构

LSTM.例如,刘旭林等<sup>[9]</sup>考虑  $PM_{2.5}$  浓度与气象要素的时空相关性,建立 CNN-Seq2Seq 预测模型,有效提高  $PM_{2.5}$  浓度未来一小时预测精度.邓丽等<sup>[10]</sup>考虑时间因素、空间因素、气象因素和空气污染物因素等多种因素,提出了 SSAE 深度特征学习和 LSTM 网络的模型.黄伟建等<sup>[11]</sup>考虑到  $PM_{2.5}$  多元数据间的时空相关性,通过构建本地记忆组件和邻域组件对时间和空间依赖关系,提出一种 TSMN 预测模型.由此可见,LSTM 因其能够捕捉中长期时间依赖的信息在空气质量预测领域的应用取得了不错的效果.2017年,Google 机器翻译团队采用 attention 机制来进行机器翻译任务,取得显著成效<sup>[12]</sup>.此后,注意力机制在语音、自然语言处理、计算机视觉等领域取得重大突破.已有文献将注意力机制应用到  $PM_{2.5}$  浓度预测的研究较少.

综上所述,本文提出一种基于高斯概率估计和注意力机制的高斯-注意力预测模型,将历史  $PM_{2.5}$  数据和其他空气质量数据的分层时间戳信息对齐作为输入<sup>[13]</sup>,通过多头注意力机制捕捉影响  $PM_{2.5}$  空气质量浓度关键的信息,再经过高斯似然估计得到预测结果,并对预测结果进行比较和评价.

## 2 高斯-注意力模型预测 $PM_{2.5}$ 浓度

### 2.1 问题分析

$PM_{2.5}$  浓度预测的问题本质上是时间序列的问题,利用历史  $N$  个时间段的数据,去预测未来  $M$  个时间间隔的数据<sup>[14]</sup>.即利用历史数据  $X_{his}=[X_{t-N+1}, \dots, X_t]$  去预测未来数据  $X_{fu}=[X_{t+1}, \dots, X_{t+M}]$ ,  $PM_{2.5}$  浓度预测示意图如图 1 所示.

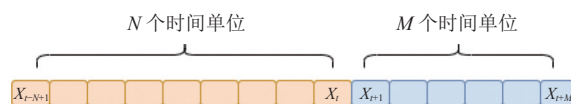


图 1  $PM_{2.5}$  浓度预测

$PM_{2.5}$  浓度不仅受到历史  $PM_{2.5}$  数据的影响,同时也受到大气污染物和气象要素等因素影响<sup>[15]</sup>.因此,应

当将同一时间点的空气质量监测站的其他监测数据如 NO<sub>2</sub>、SO<sub>2</sub> 等作为输入, 即第  $t$  个时间点的  $X_t=[PM_{2.5}, NO_2, \dots, SO_2]$ , 通过对多种历史数据特征进行建模处理, 实现对 PM<sub>2.5</sub> 浓度预测. 我们的目标是建模得到如式 (1) 所示的条件分布:

$$P(Y_{t+1:t+\tau}|X_{1:t};\theta) = \prod_{t+1}^{t+\tau} P(Y_t|X_{1:t-1};\theta) \quad (1)$$

其中,  $\theta$  表示模型中所有可学习的参数,  $X_{1:t}$  表示 1- $t$  时刻历史的监测数据,  $Y_t$  表示模型预测得到的 PM<sub>2.5</sub> 浓度数据.

### 2.2 高斯概率模型

在本文中, 由于空气监测站的检测数据都是实测数据, 这些长期、大量的观测数据受到多种因素的影响使得观测数据的分布呈现为高斯分布. 因此, 本文通过构建高斯概率密度函数直接预测下一个时间点的概率分布的所有参数  $\theta$  (均值和方差), 通过采样得到下一个时间节点的 PM<sub>2.5</sub> 预测值. 高斯概率密度函数如式 (2) 所示:

$$I_G(y|\mu, \sigma) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) \quad (2)$$

其中,  $\mu$  表示数学期望,  $\sigma$  表示标准差.

### 2.3 注意力机制

注意力机制的本质来自人类的视觉感知机制, 人们在进行视觉感知的时候, 会根据需求将注意力放在特定的一部分. 注意力机制通过对输入权重分配的关注, 可以快速提取稀疏数据的重要特征. 注意力函数是将一个查询 (query) 和键值对 (key-value) 映射到输出的方法<sup>[16]</sup>. 计算注意力得分如图 2 所示.

计算注意力得分主要分为 3 步: 第 1 步是将 query 和每个 key 进行相似度计算得到权重.

$$f(Q, K_i) = \begin{cases} Q^T K_i \text{dot} \\ Q^T W_\alpha K_i \text{general} \\ W_\alpha [Q; K_i] \text{concat} \end{cases} \quad (3)$$

其中,  $Q$  表示 query 矩阵,  $K_i$  表示第  $i$  个 key,  $W_\alpha$  表示权重矩阵,  $f(Q, K_i)$  表示 query 跟第  $i$  个 Key 的相似度,  $\text{dot}$  表示  $Q^T$  与  $K$  的点积,  $\text{general}$  表示  $Q^T$  与  $K$  的加权点积,  $\text{concat}$  表示  $Q^T$  与  $K$  的拼接.

第 2 步是使用一个 *Softmax* 函数对权重进行归一化处理.

$$a_i = \text{Softmax}(f(Q, K_i)) = \frac{\exp(f(Q, K_i))}{\sum_j \exp(f(Q, K_j))} \quad (4)$$

其中,  $a_i$  表示第  $i$  个权重归一化的值.

最后一步是将权重和相应的键值 value 进行加权求和得到最后的注意力得分.

$$\text{Attention}(Q, K, V) = \sum_i a_i V_i \quad (5)$$

其中,  $V_i$  表示第  $i$  个 value 的值,  $\text{Attention}(Q, K, V)$  表示最后的注意力得分.

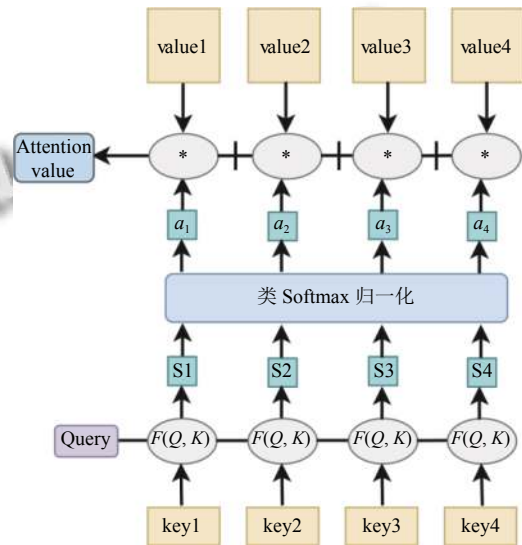


图 2 计算注意力得分图

### 2.4 高斯-注意力预测模型

本文通过注意力机制和高斯概率估计相结合实现. 高斯-注意力预测模型由输入、注意力层、高斯概率估计、输出构成. 高斯-注意力预测模型根据历史空气质量观测数据来预测未来 PM<sub>2.5</sub> 浓度值. 高斯-注意力预测模型如图 3 所示.

高斯-注意力预测模型的核心模块是多头注意力机制和高斯似然估计, 能够对长序列的空气质量数据进行建模, 捕获其复杂的数据特征从而进行预测. 其中多头注意力机制能够关注不同子空间的不同部分的信息, 能够将关联化的空气质量数据特征进行深层次挖掘, 得到更加有效特征. 高斯似然估计通过对拼接后的数据特征建模, 能够得到准确 PM<sub>2.5</sub> 浓度值的概率分布, 减低模型过拟合的风险. 本文使用的多头注意力机制如图 4 所示.

如图 4 所示,  $h$  表示多头注意力机制中的头数. 注意力集中机制中的  $Q, K, V$  通过不同的线性变换得到新的一组  $Q, K, V$ .  $h$  组不同的  $Q, K, V$  通过计算

注意力得分得到  $h$  组不同的注意力得分. 将  $h$  组的注意力得分拼接, 经过线性变换得到最后的输出.

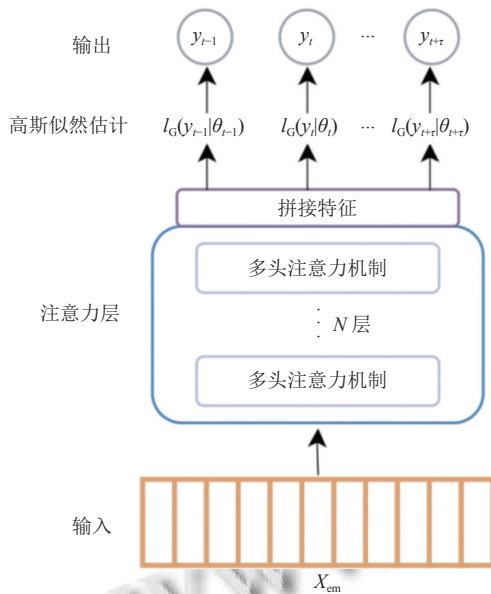


图3 高斯-注意力预测模型

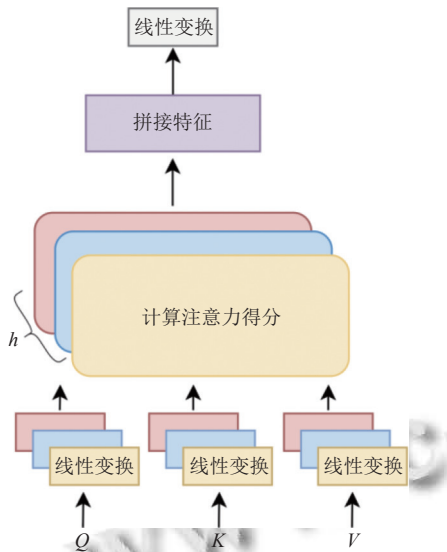


图4 多头注意力机制

为了捕获空气质量数据间的长期依赖和全局信息, 我们考虑空气质量数据的分层时间戳的信息(周, 日, 小时). 如图5所示, 本文提出一个统一、对齐的输入表示来解决这个问题.

模型的输入  $X_{em}$  包括原始信息、局部信息、全局信息3个部分, 长度为  $d$ . 局部信息是一个固定的位置嵌入如式(6), 式(7)所示:

$$PE_{(pos,2i)} = \sin(pos/(2L_x)^{(2i/d)}) \quad (6)$$

$$PE_{(pos,2i+1)} = \cos(pos/(2L_x)^{(2i/d)}) \quad (7)$$

其中,  $i=[1, \dots, [d/2]]$ . 全局信息由两个可学习的日嵌入和周嵌入组成. 为了对齐输入的维度, 本文用一个一维的卷积滤波器对原始小时数据进行卷积处理得到原始信息.

模型在训练阶段中, 多头注意力机制能到看到全局的信息, 能够一次性的更新  $t$  长度的所有参数  $\theta$ . 模型预测阶段中, 由于只能看到当前的时间步的信息, 因此在计算注意力得分的时候需要将未来的信息进行掩码处理. 模型的训练阶段采用均方误差 MSE 作为模型的损失函数, 具体计算过程如式(8)所示.

$$loss = 1/N \sum_1^N (y^p - x^t)^2 \quad (8)$$

其中,  $N$  表示样本总数,  $y^p$  表示模型预测的  $t$  时刻  $PM_{2.5}$  浓度值,  $y^t$  表示  $t$  时刻真实的  $PM_{2.5}$  浓度值.

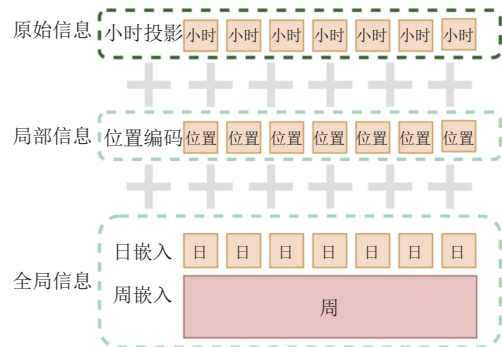


图5 输入对齐表示

### 3 实验结果与分析

#### 3.1 数据来源及分析

本文数据集选择 2019 年 6 月 1 日到 2021 年 3 月 11 日沈阳市某监测站点的 6 种空气质量监测数据 ( $PM_{2.5}$ 、 $PM_{10}$ 、 $SO_2$ 、 $NO_2$ 、 $NO$ 、 $CO$ ) 作为实验数据, 单位为  $\mu g/m^3$ . 原始数据周平均浓度如图6所示, 基本统计特征如表1所示.

由图6可知, 6 种空气质量监测数据浓度变化有相同的趋势,  $PM_{2.5}$ ,  $PM_{10}$ ,  $CO$  具有相同的数据特征变化较为剧烈.  $NO_2$ ,  $NO$ ,  $SO_2$  变化相对平缓.

由表1可知,  $CO$  的量级较大,  $NO$  和  $SO_2$  量级较小,  $PM_{2.5}$ 、 $PM_{10}$ 、 $NO_2$  属于同一量级. 综上所述, 6 种空气质量监测数据具有密不可分的相似性, 同时不同数据特征间有高度的复杂性.

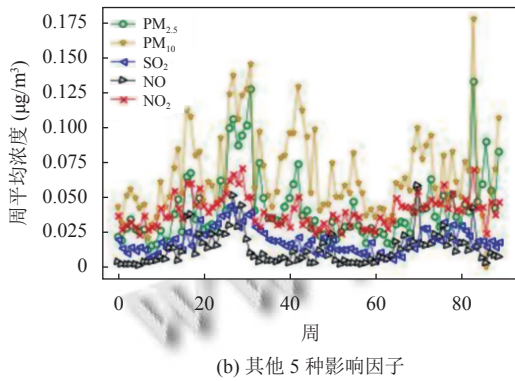
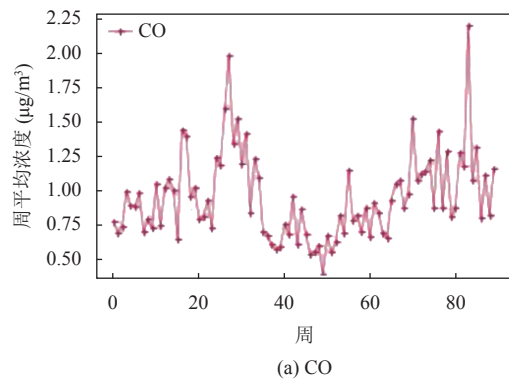


图6 2019-06-01至2021-03-11原始数据周平均浓度

表1 6种空气质量监测数据统计特征 (µg/m³)

影响因子	最大值	最小值	均值	标准差
PM <sub>2.5</sub>	1.109	0.001	0.042	0.043
PM <sub>10</sub>	1.302	0.002	0.068	0.064
SO <sub>2</sub>	1.049	0.001	0.019	0.023
NO <sub>2</sub>	1.173	0	0.040	0.023
CO	19.192	0.015	0.949	0.684
NO	0.525	0.004	0.013	0.028

由于一些不可控的因素例如网络信号丢失,采集器突然宕机等导致部分原始数据会出现异常情况,所以在此次研究中采用了24小时均值法对异常和缺失的原始数据进行数据预处理,使空气质量的监测数据尽可能地保存相应的特征.实验选择历史一周的小时空气质量数据作为输入,预测未来一天的PM<sub>2.5</sub>各小时浓度值.

### 3.2 实验参数

本实验的环境主要参数,平台: Google Colab, GPU显卡: Telsa P100, 显存为 16 GB, 操作系统为 Ubuntu 18.04. 本文提出高斯-注意力预测模型的超参数主要涉及多头注意力机制. 当高斯-注意力预测模型超参数设置如表2所示时,模型预测性能最优.

模型使用MSE作为损失函数,按照表2超参数组

合迭代完成15轮训练,模型在训练集和验证集loss值变化曲线如图7所示.

由图7可知,随着迭代次数的增加,训练集和验证集的loss值在训练初期快速下降,当迭代次数达到10次,训练集的loss趋于平稳,验证集的loss产生小幅度的波动.表明模型已经能很好表示数据间的复杂特征.为了避免模型发生过拟合现象,模型选择在15轮停止训练.

表2 最优超参数组合

参数	数值
批次大小 (batch)	32
学习率	0.001
激活函数	ReLU
优化策略	Adam
注意力头数 (h)	4
输入序列长度 (d)	168
隐藏层神经元	64
Dropout	0.2

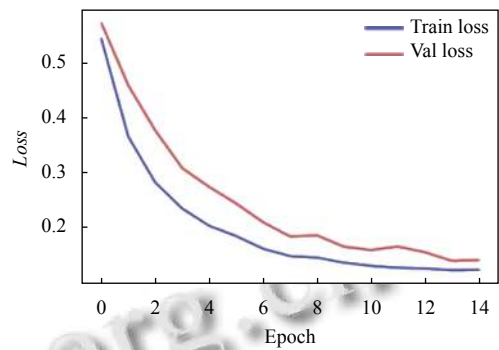


图7 模型loss变化曲线

### 3.3 实验结果

为了验证算法高斯-注意力预测模型的有效性,本文选择与当前主流空气质量预测方法ARIMA、LSTM、DeepAR<sup>[17]</sup>进行比较.对比实验在相同的实验环境中进行,上述3种模型的超参数设置均为最优超参数组合,本文模型与上述3种主流方法的部分预测值如表3.

表3 不同模型部分预测值 (µg/m³)

真实值	本文	ARIMA	LSTM	DeepAR
0.124 6	0.123 4	0.073 8	0.122 3	0.122 7
0.102 5	0.104 9	0.093 1	0.087 6	0.095 6
0.028 5	0.052 3	0.087 1	0.072 3	0.067 6
0.015 4	0.017 7	0.030 0	0.020 6	0.019 3
0.031 5	0.033 2	0.027 8	0.032 3	0.028 9
...	...	...	...	...
0.153 9	0.163 7	0.040 0	0.177 9	0.175 3

由表 3 可知, 高斯-注意力模型预测值相较于传统方法 ARIMA 及深度学习方法 LSTM、DeepAR 模型更接近真实的  $PM_{2.5}$  浓度值, 误差绝对值也有所降低.

为了更直观的体现各预测模型在测试集上的预测性能, 因此绘制 4 种模型在测试集上预测值和真实值的折线图, 如图 8 所示.

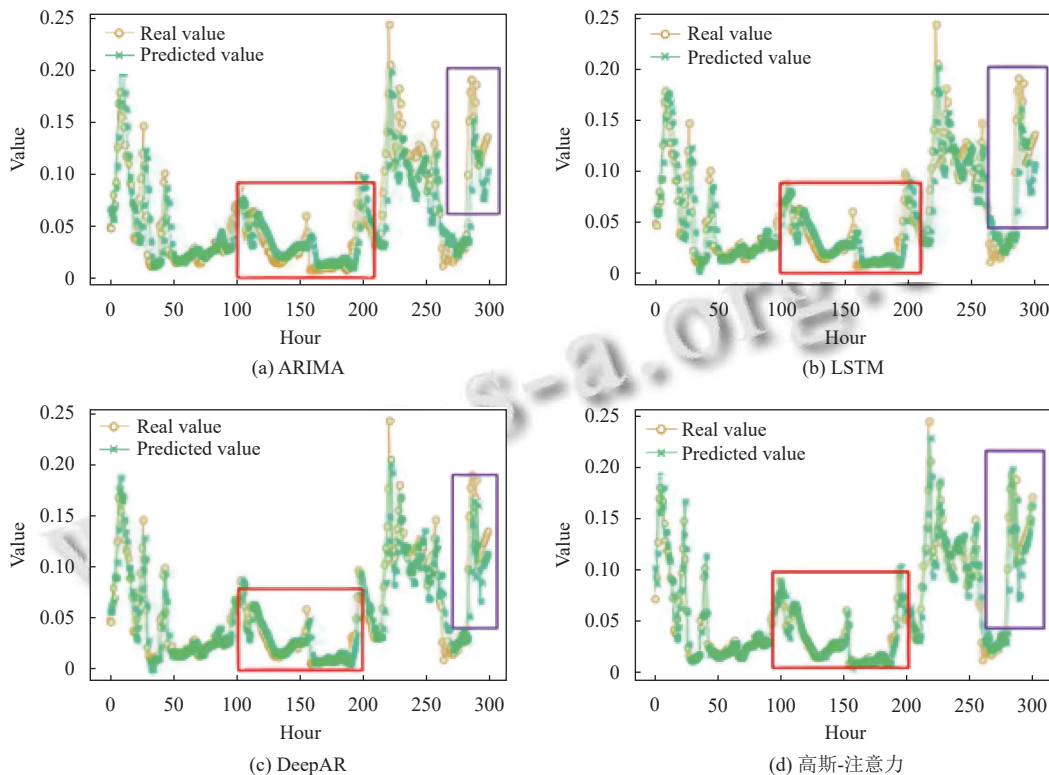


图 8 4 种方法  $PM_{2.5}$  浓度预测图

由图 8 可知, 高斯-注意力预测模型的整体预测结果好于 3 种基准模型 ARIMA、LSTM、DeepAR. 由图 8 的红框可知, 传统模型 ARIMA 预测结果有明显的滞后性, 整体的数据趋势出现了向后的偏移, 这表明相较于深度学习方法, 传统模型 ARIMA 对多维时序数据的处理能力不足, 不能表达深层次的特征信息. LSTM 模型能在一定程度上缓解这个问题, DeepAR 和高斯-注意力模型通过高斯似然估计对  $PM_{2.5}$  浓度进行采样预测, 解决了数据向后偏移的问题. 由图 8 的紫框可知, LSTM、DeepAR 模型在变化剧烈的序列上预测结果较差, 高斯-注意力预测模型得到更好的拟合结果. 由于 DeepAR 模型中提取特征的部分是 LSTM, 这表明相较于 LSTM, 本文提出的高斯-注意力预测模型中的多头注意力机制能够关注到全局信息, 通过不同子空间对高维空间的复杂特征进行抽取, 在变化剧烈的序列上有很好的预测性能.

标下, 高斯-注意力预测模型均取得最好的结果. 相较于基准模型中性能较优的 DeepAR, 高斯-注意力预测模型的 MSE、MAE 分别下降了 21%、15%. 该实验验证本文提出的高斯-注意力预测模型对解决多维时序问题的优越性, 能得到准确性和鲁棒性较强的空气质量预测结果.

由表 4 可知, 4 种模型在 MSE 和 MAE 的评价指

表 4 模型实验结果

算法模型	MSE	MAE
ARIMA	6.639	1.901
LSTM	4.401	1.317
DeepAR	3.762	1.219
高斯-注意力	2.963	1.031

不同的注意力层对输入数据有不同的注意力得分. 第 1、4 层注意力得分如图 9 所示, 注意力机制的第 1 层能够注意到每天的整体情况, 给予每天的输入数据相似的权重. 注意力机制的第 4 层进一步压缩了注意力机制的范围, 给予较长时间间隔的数据很少的关注, 反而更加在意最近 48 h 数据的影响.

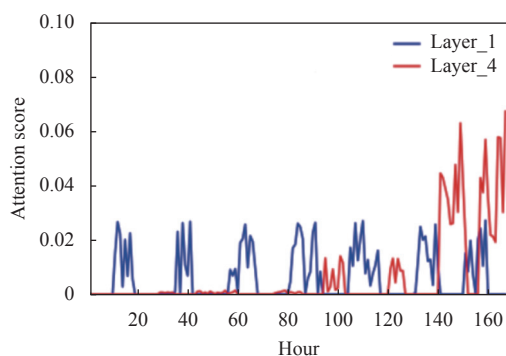


图9 第1、4层注意力得分

#### 4 结束语

本文提出了一种基于高斯概率估计和注意力机制的高斯-注意力预测模型. 为了捕获空气质量数据间的长期依赖和全局信息, 将不同时间戳的信息对齐得到原始信息、局部信息、全局信息作为输入. 引入多头注意力机制, 通过不同子空间对输入序列的全局信息进行学习, 得到深层次的数据特征, 再经过高斯似然估计对  $PM_{2.5}$  浓度质量数据进行概率估计. 最终通过与多种模型对比, 高斯-注意力预测模型在各个评价指标方面均取得最优结果. 另外, 当前模型只考虑了空气质量数据的时序特征, 未来将结合不同位置的空气质量监测站数据, 对空气质量数据的空间和时间特征进行挖掘, 得到更加有效的预测结果.

#### 参考文献

- Pui DYH, Chen SC, Zuo ZL.  $PM_{2.5}$  in China: Measurements, sources, visibility and health effects, and mitigation. *Particology*, 2014, 13: 1–26. [doi: 10.1016/j.partic.2013.11.001]
- Chen PF, Bi XH, Zhang JQ, *et al.* Assessment of heavy metal pollution characteristics and human health risk of exposure to ambient  $PM_{2.5}$  in Tianjin, China. *Particology*, 2015, 20: 104–109. [doi: 10.1016/j.partic.2014.04.020]
- 严宙宁, 牟敬锋, 赵星, 等. 基于 ARIMA 模型的深圳市大气  $PM_{2.5}$  浓度时间序列预测分析. *现代预防医学*, 2018, 45(2): 220–223, 242.
- 石峰, 楼文高, 张博. 基于灰狼群智能最优化的神经网络  $PM_{2.5}$  浓度预测. *计算机应用*, 2017, 37(10): 2854–2860. [doi: 10.11772/j.issn.1001-9081.2017.10.2854]
- 尹建光, 彭飞, 谢连科, 等. 基于小波分解与自适应多级残差修正的最小二乘支持向量回归预测模型的  $PM_{2.5}$  浓度预测. *环境科学学报*, 2018, 38(5): 2090–2098.
- 蒋洪迅, 石晓文, 孙彩虹, 等. 基于 DLNN 模型的沈阳地区  $PM_{2.5}$  浓度预测. *系统工程*, 2021, 39(2): 13–21.
- Mirzaei M, Amanollahi J, Tzani CG. Evaluation of linear, nonlinear, and hybrid models for predicting  $PM_{2.5}$  based on a GTWR model and MODIS AOD data. *Air Quality, Atmosphere & Health*, 2019, 12(10): 1215–1224.
- Lim B, Zohren S. Time-series forecasting with deep learning: A survey. *Philosophical Transactions of the Royal Society A*, 2021, 379(2194): 20200209. [doi: 10.1098/rsta.2020.0209]
- 刘旭林, 赵文芳, 唐伟. 应用 CNN-Seq2seq 的  $PM_{2.5}$  未来一小时浓度预测模型. *小型微型计算机系统*, 2020, 41(5): 1000–1006. [doi: 10.3969/j.issn.1000-1220.2020.05.017]
- 邓丽, 邬群勇, 杨水荣. 融合 SSAE 深度特征学习和 LSTM 网络的  $PM_{2.5}$  小时浓度预测. *环境科学学报*, 2020, 40(9): 3422–3434.
- 黄伟建, 李丹阳, 黄远. 基于深度学习的  $PM_{2.5}$  浓度长期预测. *计算机应用研究*, 2021, 38(6): 1809–1814.
- Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach: Curran Associates Inc., 2017. 6000–6010.
- Zhou HY, Zhang SH, Peng JQ, *et al.* Informer: Beyond efficient transformer for long sequence time-series forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, 35(12): 11106–11115.
- 傅颖颖, 张丰, 杜震洪, 等. 融合图卷积神经网络和注意力机制的  $PM_{2.5}$  小时浓度多步预测. *浙江大学学报(理学版)*, 2021, 48(1): 74–83.
- 张莹, 张婕, 王式功, 等. 成都市  $PM_{2.5}$  浓度变化的影响因素交互作用研究. *中国环境科学*, 2021, 41(10): 4518–4528. [doi: 10.3969/j.issn.1000-6923.2021.10.006]
- Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. *arXiv:1409.0473*, 2014.
- Salinas D, Flunkert V, Gasthaus J, *et al.* DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 2020, 36(3): 1181–1191. [doi: 10.1016/j.ijforecast.2019.07.001]

(校对责编: 孙君艳)