

基于二向注意力循环神经网络的PM_{2.5}浓度预测

杨亚莉¹, 李智伟¹, 钟卫军²

(1. 空军工程大学基础部, 西安, 710051; 2. 宇航动力学国家重点实验室, 西安, 710043)

摘要 针对PM_{2.5}浓度预测模型效果不稳定、泛化能力差的问题, 以循环神经网络和注意力机制为基础, 提出了二向注意力循环神经网络(TDA-RNN)。首先, TDA-RNN模型通过注意力机制获取输入数据的时序注意力和类别注意力, 并将其进行融合; 然后通过特征编码器对融合后的数据进行编码, 获得中间特征; 最后将中间特征与PM_{2.5}浓度的历史信息融合, 并通过特征解码器获取预测值。对北京地区的PM_{2.5}浓度进行了预测。结果表明, 相比前向型神经网络、长短期记忆神经网络、门控循环单元模型和滑动平均模型, TDA-RNN模型预测精度更高; 在抗干扰测试中, 当输入数据存在无关因素时, TDA-RNN模型的预测精度出现轻微下降, 但仍高于其他模型。该二向注意力循环神经网络特征提取能力强, 预测精度高, 同时可适用于其他场景的多变量时间序列预测。

关键词 PM_{2.5}; 时间序列预测; 深度学习; 循环神经网络; 注意力机制

DOI 10.3969/j.issn.1009-3516.2020.06.016

中图分类号 P456.8 文献标志码 A 文章编号 1009-3516(2020)06-0101-06

A Prediction of PM_{2.5} Concentration Based on a Two-Direction Attention-Based Recurrent Neural Network

YANG Yali¹, LI Zhiwei¹, ZHONG Weijun²

(1. Department of Basic Sciences, Air Force Engineering University, Xi'an 710051, China;

2. State Key Laboratory of Astronautic Dynamics, Xi'an 710043, China)

Abstract Aimed at the problems that PM_{2.5} concentration prediction model is unstable in efficiency and poor in generalization ability, a Two-Direction Attention-based Recurrent Neural Network (TDA-RNN) is proposed based on the cyclic neural network and attention mechanism. Firstly, the temporal attention and the category attention in inputting data through attention mechanism are obtained in making them fused by TDA-RNN model, and then the fused data are encoded to obtain intermediate features through feature encoder. Finally, the intermediate features are fused with historical information of PM_{2.5} concentration, and the predicted values are obtained by feature decoder. The PM_{2.5} concentration in Beijing is predicted by several models. The results show that the prediction accuracy of TDA-RNN is higher than that of the Back Propagation Neural Network, the Long Short-Term Memory, the Gate Recurrent Unit and the Moving Average model. In the anti-jamming test, while the input data having noise factors, the prediction accura-

收稿日期: 2020-08-06

基金项目: 国家自然科学基金(61801518); 空军工程大学基础部研究生创新基金

作者简介: 杨亚莉(1974—), 女, 陕西泾阳人, 教授, 硕士生导师, 主要从事微分方程及其应用研究。E-mail: yylhgr@126.com

通信作者: 李智伟(1995—), 男, 山西晋城人, 硕士生, 主要从事时间序列分析研究。E-mail: lzw116405@126.com

引用格式: 杨亚莉, 李智伟, 钟卫军. 基于二向注意力循环神经网络的PM_{2.5}浓度预测[J]. 空军工程大学学报(自然科学版), 2020, 21(6): 101-106. YANG Yali, LI Zhiwei, ZHONG Weijun. Prediction of PM_{2.5} Concentration Based on a Two-Direction Attention-Based Recurrent Neural Network[J]. Journal of Air Force Engineering University (Natural Science Edition), 2020, 21(6): 101-106.

cy of TDA-RNN decreases slightly, but still higher than that of other models. The TDA-RNN proposed in the paper is strong in feature extraction ability and high in prediction accuracy. And this can also be applied to multivariate time series prediction in other application scenarios.

Key words PM_{2.5}; time series prediction; deep learning; recurrent neural network; attention mechanism

随着我国工业化和城镇化的快速发展,大气污染逐渐成为我们在发展过程中不可忽视的问题。作为衡量空气质量的重要指标,细颗粒物(PM_{2.5})颗粒小、活性强、空中停留时间长,易携带有毒有害物质,对人体和环境的危害更严重,导致老人、儿童等弱体质人群易引发呼吸系统和心血管等疾病^[1]。对PM_{2.5}进行准确预测,有利于相关部门管理和治理空气问题,并可为居民提供空气污染预警,保障人民健康。

PM_{2.5}的形成原因及影响其浓度的因素仍未有统一的认识,同时各类影响因素具有一定的随机性,增加了预测PM_{2.5}浓度的难度^[2]。目前,预测方法主要有确定性模型和统计模型。确定性模型是一种理论模拟方法,充分描述大气污染物扩散和稀释过程中复杂的物理和化学变化。Shorshani等^[3]利用改进的高斯模型模拟了公路旁PM_{2.5}浓度的变化,Djalalova等^[4]利用卡尔曼滤波法预测了PM_{2.5}的浓度。统计模型是利用各类数值模拟技术对大量的空气质量数据与气象数据进行分析,进而挖掘潜在规律的方法。彭岩等^[5]提出了一种基于集成树-梯度提升决策树的PM_{2.5}浓度预测模型;罗宏远等^[6]开展了基于二层分解技术和改进极限学习机模型的PM_{2.5}浓度预测研究;张熙来等^[7]基于单时间序列数据的动态调整模型来进行预测。此外,还有大量基于机器学习的PM_{2.5}浓度预测模型,如支持向量机^[8]、随机森林^[9]、Seq2seq模型^[10]。

传统的统计模型方法是使用向量自回归模型^[11]、ARMAX^[12]、ARIMAX^[13]等模型,但这些方法只能提取到相对简单的线性特征,难以有效提取出较为复杂的非线性特征。随着机器学习研究的深入,深度学习模型以其强大的非线性特征提取和分析能力逐渐受到研究人员的青睐,其中,循环神经网络(Recurrent Neural Network, RNN)^[14]凭借其特殊的结构,在时间序列预测任务中展现了比其他模型更优越的性能。但标准的RNN网络在长序列训练过程中容易出现梯度消失和梯度爆炸的问题,为此,研究人员又提出了一类特殊的RNN网络——长短期记忆神经网络(Long Short-Term Memory, LSTM)^[15],进一步提升了模型对时序数据的特征提取能力。为了解决信息超载问题,研究人员通过

注意力机制,将计算资源分配给更重要的任务。基于注意力机制的深度学习网络往往能比其初始网络表现出更强的性能。

在以上研究的基础上,本文提出了二向注意力循环神经网络模型(Two-Direction Attention-Based Recurrent Neural Network, TDA-RNN)。

1 数据来源及预处理

本文建模和评估使用的是北京地区的空气质量数据集,来自于UCI机器学习仓库^[16]。该数据集记录了2010年1月1日到2014年12月31日每小时的北京空气质量,共计43 824组,包含8项不同的气象条件数据:PM_{2.5}浓度、露点、温度、大气压强、风向、风速、累积雪量和累积雨量。

由于不同变量采用了不同的单位来衡量,为使模型在训练过程中能收敛,在建立模型前,需对原始数据进行无量纲化处理:

$$\hat{p}_{i,j} = \frac{p_{i,j} - \min P_j}{\max P_j - \min P_j} \quad (1)$$

式中:P_j为数据集的第j类因素的集合;p_{i,j}为第j类因素的第i个原始数据;̂p_{i,j}为p_{i,j}经过数据标准化后的数据。

2 TDA-RNN 模型构建

为提高模型对多变量时序预测的精度,Yao Q等^[17]设计了基于双阶段注意力的循环神经网络。受该模型的启发,本文设计了二向注意力循环神经网络TDA-RNN。

TDA-RNN 主要处理流程分为4步:①首先通过类别注意力模块,对输入的多变量数据的不同变量进行分析,从而获取不同变量数据的类别注意力;②通过时序注意力模块,对输入的多变量数据的不同时间步进行分析,从而获取数据的时序注意力;③将时序注意力与类别注意力进行融合,获取每个输入数据的注意力,更新输入数据并对其进行特征编码;④将编码后的特征矩阵与PM_{2.5}浓度的历史数据进行融合,并将其输入特征解码器中进行解码,获取最终的预测值。TDA-RNN的网络结构见图1。

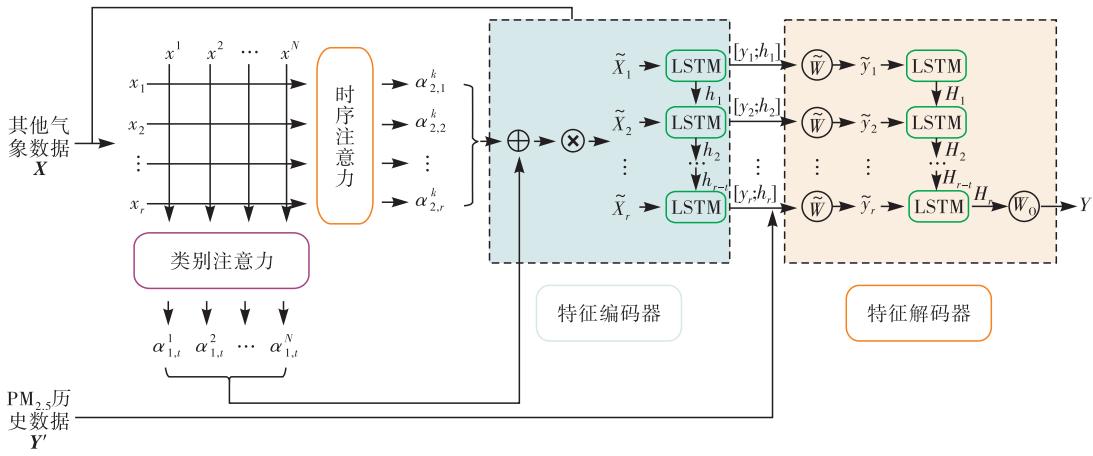


图1 TDA-RNN模型结构

TDA-RNN 是一个多输入单输出模型,需输入目标序列的历史数据 \mathbf{Y}' 和多变量数据 \mathbf{X} ,输出为预测变量值 \mathbf{Y} ,其中 $\mathbf{X}=(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T)^T=(x^1, x^2, \dots, x^N), \mathbf{Y}'=(y_1, y_2, \dots, y_T), \mathbf{x}_t \in R^N, x^n \in R^T, Y, y_t \in R, N$ 为影响目标序列的因素的数量, T 为模型输入的时间窗口长度。因此,模型 TDA-RNN 可表示为需要通过训练获得具体参数的非线性函数 F :

$$\mathbf{Y}=F(\mathbf{Y}', \mathbf{X})=F(y_1, y_2, \dots, y_T, x^1, x^2, \dots, x^N) \quad (2)$$

2.1 类别注意力

RNN 是一类以序列数据为输入的递归神经网络,其在序列的演进方向进行递归的特性使得 RNN 在语音识别、机器翻译、时间序列预测等方面性能显著^[18]。使用 RNN 进行时间序列预测,模型获取输入序列数据 $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T)$,其中 $\mathbf{x}_t \in R^n, n$ 为与目标序列相关的因素的数量,并将通过训练得到非线性函数 f :

$$\mathbf{h}_t=f(\mathbf{h}_{t-1}, \mathbf{x}_t) \quad (3)$$

式中: $\mathbf{h}_t \in R^m$ 是 RNN 在 t 时刻的隐藏状态; m 为隐藏状态的序列长度。非线性函数 f 可以是标准 RNN 层、LSTM 层或 GRU 层。

LSTM 层不仅保留了标准 RNN 层的允许信息持久化的特性,还改善了 RNN 长期依赖问题。因此,使用 LSTM 层来为模型提取序列数据中的长期和短期依赖特征。每个 LSTM 单元通过忘记门 f_t 、输入门 i_t 和输出门 o_t 这 3 个门结构来存储和更新模型在 t 时刻的细胞状态 s_t 和隐藏状态 \mathbf{h}_t :

$$\left\{\begin{array}{l} f_t=\sigma(\mathbf{W}_f[\mathbf{h}_{t-1}; \mathbf{x}_t]+\mathbf{b}_f) \\ i_t=\sigma(\mathbf{W}_i[\mathbf{h}_{t-1}; \mathbf{x}_t]+\mathbf{b}_i) \\ o_t=\sigma(\mathbf{W}_o[\mathbf{h}_{t-1}; \mathbf{x}_t]+\mathbf{b}_o) \\ s_t=f_t \odot s_{t-1}+i_t \odot \tanh (\mathbf{W}_s[\mathbf{h}_{t-1}; \mathbf{x}_t]+\mathbf{b}_s) \\ \mathbf{h}_t=o_t \odot \tanh (s_t) \end{array}\right. \quad (4)$$

式中: $[\mathbf{h}_{t-1}; \mathbf{x}_t] \in R^{m+n}$ 是由 $t-1$ 时刻的隐藏状态 \mathbf{h}_{t-1} 与 t 时刻输入的 \mathbf{x}_t 拼接而成; $\mathbf{W}_f, \mathbf{W}_i, \mathbf{W}_o, \mathbf{W}_s \in R^{m \times(m+n)}$ 和 $\mathbf{b}_f, \mathbf{b}_i, \mathbf{b}_o, \mathbf{b}_s \in R^m$ 是网络需要通过训练获

取的参数; σ 是 sigmoid 函数, \odot 是哈达玛积运算。

利用输入数据 \mathbf{X} 的第 k 类影响因素的序列数据 \mathbf{x}^k 以及 LSTM 单元的细胞状态 $s_{1,t}$ 和隐藏状态 $\mathbf{h}_{1,t-1}$,我们可以构造出输入数据 \mathbf{X} 不同类别的影响因素的不同注意力值 $e_{1,t}^k$:

$$e_{1,t}^k=\mathbf{v}_{1,e}^T \tanh (\mathbf{W}_{1,e}[\mathbf{h}_{1,t-1}; s_{1,t-1}]+\mathbf{U}_{1,e} \mathbf{x}^k) \quad (5)$$

式中: $\mathbf{v}_{1,e} \in R^T, \mathbf{w}_{1,e} \in R^{T \times 2m}, \mathbf{U}_{1,e} \in R^{T \times T}$, 是网络需要通过训练获取的参数。为了使注意力权重之和为 1,需利用 SoftMax 函数来获取最终的类别注意力值 $\alpha_{1,t}^k$:

$$\alpha_{1,t}^k=\frac{\exp (e_{1,t}^k)}{\sum_{i=1}^n \exp (e_{1,t}^i)} \quad (6)$$

2.2 时序注意力

时序注意力仍通过 LSTM 层进行计算。利用输入数据 \mathbf{X} 的时间 t 的不同影响因素的数据 \mathbf{x}_t 以及 LSTM 单元的细胞状态 s_2^k 和隐藏状态 \mathbf{h}_2^{k-1} ,我们可以构造出输入数据 \mathbf{X} 不同时间步的不同的注意力值 $e_{2,t}^k$:

$$e_{2,t}^k=\mathbf{v}_{2,e}^T \tanh (\mathbf{W}_{2,e}[\mathbf{h}_2^{k-1}; s_2^{k-1}]+\mathbf{U}_{2,e} \mathbf{x}_t) \quad (7)$$

式中: $\mathbf{v}_{2,e} \in R^n, \mathbf{w}_{2,e} \in R^{n \times 2T}, \mathbf{U}_{2,e} \in R^{n \times n}$ 是网络需要通过训练获取的参数。为了使注意力权重之和为 1,需利用 Softmax 函数来获取最终的类别注意力值 $\alpha_{2,t}^k$:

$$\alpha_{2,t}^k=\frac{\exp (e_{2,t}^k)}{\sum_{i=1}^n \exp (e_{2,t}^i)} \quad (8)$$

2.3 特征编码器

首先,将类别注意力与时序注意力进行融合,获得综合注意力 α_t^k :

$$\alpha_t^k=\alpha_{1,t}^k+\alpha_{2,t}^k \quad (9)$$

同时,根据综合注意力,可更新输入的多变量数据 $\mathbf{X}=(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T)$, 其中, $\mathbf{x}_t=(\alpha_t^1 x_t^1, \alpha_t^2 x_t^2, \dots, \alpha_t^n x_t^n)^T$ 。通过融合类别注意力和时序注意力,使得网络能更关注于输入的多变量数据中的重要信息,减弱干扰信息对网络的不利影响,从而提高网络的

特征提取能力。

利用新的 LSTM 层对更新后的多变量数据进行编码, 获取 LSTM 单元在不同时刻 t 的隐藏状态 \mathbf{h}_t , 并将其作为编码后的特征矩阵:

$$\mathbf{h}_t = f_1(\mathbf{h}_{t-1}, \mathbf{x}_t) \quad (10)$$

2.4 特征解码器

为了提高模型的预测精度, 将目标序列的历史数据与其他影响因素的历史数据分别输入模型中, 并将特征编码器获得的特征矩阵与目标序列的历史数据融合后进行初次解码:

$$\mathbf{y}_{t-1} = \mathbf{w}^T [\mathbf{y}_{t-1}; \mathbf{h}_{t-1}] + \tilde{b} \quad (11)$$

式中: $\tilde{\mathbf{w}} \in R^{1+m}$ 和 $\tilde{b} \in R$, 是网络需要通过训练获取的参数; $[\mathbf{y}_{t-1}; \mathbf{h}_{t-1}] \in R^{1+m}$ 为数据融合之后的层输入。

通过初次解码获得的 \mathbf{y}_{t-1} 可用于更新特征解码器在时刻 t 的隐藏状态 \mathbf{H}_t :

$$\mathbf{H}_t = f_2(\mathbf{H}_{t-1}, \mathbf{y}_{t-1}) \quad (12)$$

使用 LSTM 单元作为非线性函数 f_2 。 \mathbf{H}_t 通过方程组(13)进行更新:

$$\begin{cases} \mathbf{f}'_t = \sigma(\mathbf{W}_f' [\mathbf{H}_{t-1}; \mathbf{y}_{t-1}] + \mathbf{b}_f') \\ \mathbf{i}'_t = \sigma(\mathbf{W}_i' [\mathbf{H}_{t-1}; \mathbf{y}_{t-1}] + \mathbf{b}_i') \\ \mathbf{o}'_t = \sigma(\mathbf{W}_o' [\mathbf{H}_{t-1}; \mathbf{y}_{t-1}] + \mathbf{b}_o') \\ \mathbf{s}'_t = \mathbf{f}'_t \odot \mathbf{s}'_{t-1} + \mathbf{i}'_t \odot \tanh(\mathbf{W}_s' [\mathbf{H}_{t-1}; \mathbf{y}_{t-1}] + \mathbf{b}_s') \\ \mathbf{h}'_t = \mathbf{o}'_t \odot \tanh(\mathbf{s}'_t) \end{cases} \quad (13)$$

式中: $\mathbf{W}_f', \mathbf{W}_i', \mathbf{W}_o', \mathbf{W}_s' \in R^{m \times (m+1)}$; $\mathbf{b}_f', \mathbf{b}_i', \mathbf{b}_o', \mathbf{b}_s' \in R^m$ 为网络需要通过训练获得的参数。

对于获得更新后的 \mathbf{H}_t , 进行最终解码, 获得模型的预测值:

$$Y = \mathbf{w}_0^T \mathbf{H}_t + b_0 \quad (14)$$

式中: $\mathbf{w}_0 \in R^m$ 和 $b_0 \in R$ 为网络需要通过训练获得的参数; Y 为模型最终的预测值。

2.5 参数设置

将数据集按 0.7225 : 0.1275 : 0.15 的比例划分为训练集、验证集和测试集, 设置损失函数为 MSE, 并使用 Adam 作为模型训练的优化器。同时, 为了减少不必要的训练时间, 采用“早停法”, 即当模型在验证集上的损失函数在 20 个轮次的训练中未发生下降时, 模型停止训练。

3 实验结果及分析

3.1 模型评估标准

为了衡量模型的预测性能, 本文选择拟合优度 R^2 、均方根误差(RMSE)、平均绝对误差(MAE)和平均绝对百分比误差(MAPE)作为评价指标。

R^2 衡量的是自变量与因变量之间的整体拟合度。RMSE 能较好地反映出测量的精密度。MAE 可有效解决偏差相互抵消的问题, 并描述数据的离散程度。MAPE 可衡量预测值与观测值之间的相对误差。 R^2 越接近 1, 说明最终的预测值相对观测值的拟合程度越高, 模型性能越好。RMSE、MAE 和 MAPE 越小说明模型的预测性能越好。规定 y_i 为观测值, \hat{y}_i 为预测值, 各评价指标定义如下:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (15)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (16)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (17)$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (18)$$

3.2 不同模型性能对比

在 TDA-RNN 网络中, 有 2 个主要的参数, 分别是时间窗口长度 T 和隐藏状态的序列长度 m 。参数选择的不同, 会导致模型性能的差异。分别取 $T \in \{3, 6, 12, 18, 24\}$ 和 $m \in \{8, 16, 32, 64\}$ 并进行训练, 实验结果表明, 当 $m=64, T=12$ 时, TDA-RNN 模型的预测精度最高。

为了充分评估模型的预测性能, 使用前向型神经网络(Back Propagation Neural Network, BPNN)、门控循环单元(Gate Recurrent Unit, GRU)模型、LSTM 和滑动平均模型(Moving Average Model, MA)4 种模型作为对照模型。BPNN 模型是目前应用最广泛、构造过程最规范的一类神经网络; MA 是模型参量法谱分析方法之一, 也是现代谱估计中常用的模型; GRU 和 LSTM 均为标准循环神经网络的改进结构, 是目前性能较好的时间序列预测模型。

表 1 为不同模型的评估结果, 从中可以看出, 文中提出的 TDA-RNN 模型不仅拟合优度 R^2 最低, 而且 RMSE、MAE 和 MAPE 均远低于 4 个对照模型, 说明其整体预测性能最优。同时, 由于该模型是在 LSTM 模型的基础上构建且性能明显提高, 说明本文提出的二向注意力机制对模型产生了正向作用; BPNN、GRU 和 LSTM 的评估结果接近, 其中 GRU 的性能略强于 BPNN 和 LSTM; MA 模型除了在 MAPE 这一评估参数上优于 LSTM 模型外, 在 4 项评估参数上均表现最差, 说明线性模型由于其非线性特征提取能力较差, 已难以满足 $PM_{2.5}$ 浓度预测的需要。

表1 不同模型评估结果

指标	BPNN	GRU	LSTM	MA	TDA_RNN
R^2	0.9324	0.9355	0.9324	0.9097	0.9402
RMSE	20.54	20.06	20.54	23.74	19.31
MAE	11.89	11.34	11.95	13.64	10.89
MAPE	0.2625	0.2498	0.3026	0.2662	0.2285

图2为不同模型在测试集上预测结果的APE箱线图。根据图2可知,TDA-RNN模型在测试集上预测值的相对百分比误差的平均值和上边缘值均低于对照模型,且预测值误差的分布更集中。图2从误差分布方面进一步说明了TDA-RNN在该数据集上有较高的预测性能。

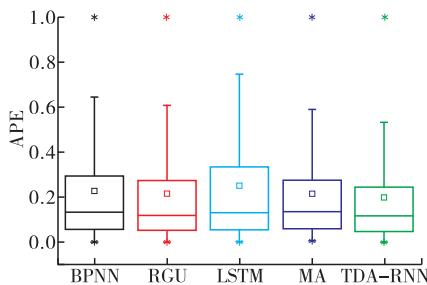


图2 不同模型预测结果的APE箱线图

3.3 参数敏感性

不同的参数,可能导致深度学习模型的性能有较大的差异。本文对TDA-RNN模型的主要参数 T 和 m 进行了参数敏感性实验。改变 T 或 m 时,保持其他参数不变并对模型进行训练,结果见图3。

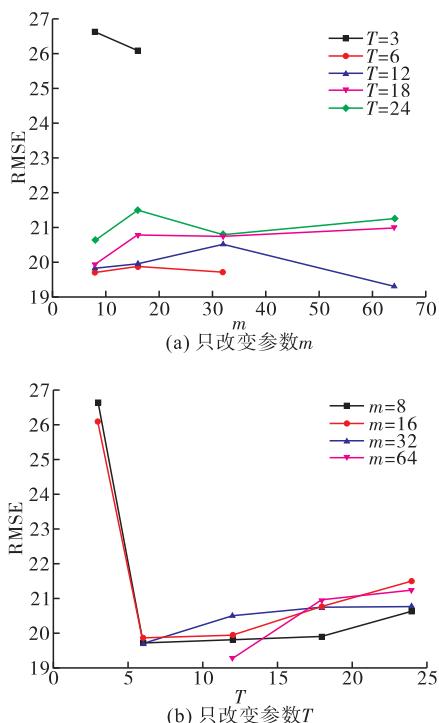


图3 不同参数的模型在测试集上的RMSE

当 $(T,m)=(3,32),(3,64),(6,64)$ 时,TDA-RNN模型对任何输入均有相同的输出,此时模型的

训练出现梯度消失现象,导致模型无法对数据集进行有效的训练。

从图3(a)可知,TDA-RNN在测试集上的RMSE随参数 m 的变化是没有规律的,甚至当 $T=12$ 和 $T=24$ 时的规律几乎是相反的。因此, m 的选择需要通过实验来确定。

从图3(b)可知, T 的取值较大或较小,均会导致模型预测误差的增大。这是因为当 T 取值较小时,模型获取的信息不充分,导致模型对缺乏足够的信息来进行预测;取值较大时,模型虽然获取了更多的信息,但是过量的信息增加了模型对重要信息提取和分析的难度。

3.4 抗干扰能力检测

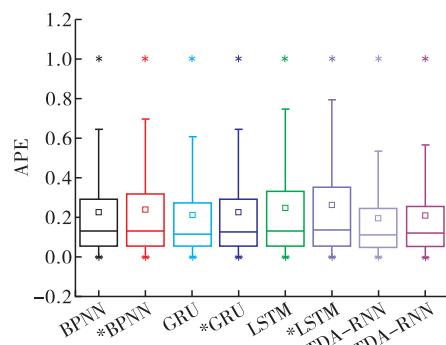
由于PM_{2.5}形成原因复杂,在进行PM_{2.5}浓度预测时使用的其他气象数据不一定与PM_{2.5}浓度变化有直接关系,此时,对网络来说相当于引入了干扰因素。因此,检测网络的抗干扰性能是非常必要的。

首先随机生成一列满足正态分布的数据,并将其作为一类气象数据添加至模型的输入 \mathbf{X} ;然后在不调整模型参数和结构的情况下,对TDA-RNN和其他对照模型进行训练。

表2为不同模型在含有干扰因素的数据集上的评估结果。当 $m=64,T=12$ 时,TDA-RNN的4项评估指标 R^2 、RMSE、MAE和MAPE分别为0.9349、20.15、11.43和0.2463,均优于对照模型。图4为不同模型在测试集上预测结果的APE箱线图。

表2 不同模型在含有干扰因素的数据集上的评估结果

指标	* BPNN	* GRU	* LSTM	* TDA-RNN
R^2	0.9317	0.9315	0.9308	0.9349
RMSE	20.64	20.67	20.79	20.15
MAE	12.32	11.7	12.4	11.43
MAPE	0.2826	0.2656	0.3262	0.2463



注: * 表示该模型是在含有干扰因素的数据集上训练。

图4 不同模型预测结果的APE箱线图

根据图4可知,在不调整模型参数和结构的情况下,各类模型在含有干扰因素的数据集上的预测

精度均有了不同程度的降低。其中, TDA-RNN 和 GRU 在测试集上的预测误差的上四分位数降低程度最低, 说明模型抗干扰能力较强。通过对比 TDA-RNN 与 LSTM 模型的评估结果可知本文提出的二向注意力机制增强了模型的抗干扰能力。

4 结语

针对现有 $PM_{2.5}$ 浓度预测模型效果不稳定、泛化能力差的问题, 本文提出了二向注意力循环神经网络 TDA-RNN, 实现了对 $PM_{2.5}$ 浓度的准确预测。模型分别在时序和类别这两个维度上使用注意力机制, 实现了对计算资源的优化分配, 使 TDA-RNN 模型拥有更高的预测精度。注意力机制的合理使用, 使模型对干扰因素有较强的识别能力, 增强了模型的抗干扰能力。同时, TDA-RNN 也适用于其他应用场景的多变量时间序列预测, 具有广阔的工程应用前景。

参考文献

- [1] 汤宇磊, 杨复沫, 詹宇. 四川盆地 $PM_{2.5}$ 与 PM_{10} 高分辨率时空分布及关联分析[J]. 中国环境科学, 2019, 39(12): 4950-4958.
- [2] 刘杰, 杨鹏, 吕文生, 等. 模糊时序与支持向量机建模相结合的 $PM_{2.5}$ 质量浓度预测[J]. 北京科技大学学报, 2014, 36(12): 1694-1702.
- [3] MASOUDI S, CHRISTIAN S, LUCIE P R, et al. Atmospheric Dispersion Modeling Near a Roadway Under Calm Meteorological Conditions[J]. Transportation Research (Part D): Transport and Environment, 2015, 34, 137-154.
- [4] IRINAD, LUCA D M, JAMES W. $PM_{2.5}$ Analog Forecast and Kalman Filter Post-processing for the Community Multiscale Air Quality (CMAQ) Model [J]. Atmospheric Environment, 2015, 108, 76-87.
- [5] 彭岩, 赵梓如, 吴婷娴, 等. $PM_{2.5}$ 浓度预测与影响因素分析[J]. 北京邮电大学学报, 2019, 42(6): 162-169.
- [6] 罗宏远, 王德运, 刘艳玲, 等. 基于二层分解技术和改进极限学习机模型的 $PM_{2.5}$ 浓度预测研究[J]. 系统工程理论与实践, 2018, 38(5): 1321-1330.
- [7] 张熙来, 赵俭辉, 蔡波. 针对 $PM_{2.5}$ 单时间序列数据的动态调整预测模型[J]. 自动化学报, 2018, 44(10): 1790-1798.
- [8] 张长江, 戴李杰, 马雷鸣. 应用 SVM 的 $PM_{2.5}$ 未来一小时浓度动态预报模型[J]. 红外与激光工程, 2017, 46(2): 252-259.
- [9] 任才溶, 谢刚. 基于随机森林和气象参数的 $PM_{2.5}$ 浓度等级预测[J]. 计算机工程与应用, 2019, 55(2): 213-220.
- [10] 马井会, 曹钰, 余钟奇, 等. 深度学习方法在上海市 $PM_{2.5}$ 浓度预报中的应用[J]. 中国环境科学, 2020, 40(2): 530-538.
- [11] ZOU J J, JIANG G F, XIE X X, et al. Application of a Combined Model with Seasonal Autoregressive Integrated Moving Average and Support Vector Regression in Forecasting Hand-Foot-Mouth Disease Incidence in Wuhan, China [J]. Medicine (Baltimore), 2019, 98, e14195.
- [12] CHOI S B, AHN I. Forecasting Seasonal Influenza-like Illness in South Korea after 2 and 30 weeks using Google Trends and Influenza Data from Argentina[J]. PLoS ONE, 2020, 15-7: e0233855.
- [13] SABHARWAL A, GROVER G, KAUSHIK S, et al. Modelling and Forecasting Positive and Negative Syndrome Scale Scores to Achieve Remission Using Time Series Analysis[J]. Int J Methods Psychiatr Res, 2019, 28-1, e1763.
- [14] CUI R, LIU M. RNN-Based Longitudinal Analysis for Diagnosis of Alzheimer's Disease[J]. Comput Med Imaging Graph, 2019, 73: 1-10.
- [15] MARAGATHAM G, DEVI S. LSTM Model for Prediction of Heart Failure in Big Data[J]. J Med Syst, 2019, 43-5, 111.
- [16] CHEN S X. Beijing $PM_{2.5}$ Data Data Set[EB/OL]. (2017-01-19)/[2020-08-06]. <https://archive.ics.uci.edu/ml/datasets/Beijing+PM2.5+Data#>.
- [17] QIN Y, SONG D, CHEN H, et al. A Dual-Stage Attention-Based Recurrent Neural Network for Time Series Prediction[C]//In Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne, Australia. 2017, 2627-2633.
- [18] GOODFELLOW I, BENGIO Y, COURVILLE A. Deep Learning (Vol. 1) [M]. Cambridge: MIT Press, 2016: 367-415.

(编辑:徐敏)