文章类型：论文|刊号（ISSN）：2630-4740 /（中图刊号）：715GL012

基于 LSTM 与 Transformer 的空气质量预测

闫凯
广东邮电职业技术学院
DOI:10.12238/eep.v7i9.2262

[摘　要] 空气质量预测对于公共健康具有重要意义,本文针对空气质量表征值的特点设计了LSTM模型与Transformer结合进行空气质量预测的方法。模型具有长期历史数据记忆能力与特征关注的泛化能力。文中将过去9年某地空气质量的原始观测数据作为输入,而后输出未来24小时的空气质量指数。实验结果显示,在测试集上,该模型的MAE和RMSE等评判指标均优于LSTM、GRU、CNN-LSTM和单独的Transformer模型,可以较好地完成空气质量预测。
[关键词] 空气质量指数；LSTM；Transformer；多头注意力
中图分类号：P465　**文献标识码**：A

Air Quality Prediction Based on LSTM and Transformer

Kai Yan
Guangdong Vocational College of Post and Telecom

[Abstract] The prediction of air quality is of great significance to public health. This article designs a method for air quality prediction by combining LSTM model and Transformer based on the characteristics of air quality characterization values. The model has the ability to remember long-term historical data and generalize feature attention. The article takes the raw observation data of air quality in a certain area over the past 9 years as input, and then outputs the air quality index for the next 24 hours. The experimental results show that on the test set, the evaluation indicators such as MAE and RMSE of the model are superior to LSTM, GRU, CNN-LSTM, and the standalone Transformer model, and can achieve better air quality prediction.
[Key words] AQI; LSTM; Transformer; Multi-Head Attention

引言

空气质量受到诸多复杂因素影响,预测存在诸多难点,原因之一在于模型精度受限,随着机器学习技术的发展,出现越来越多的组合模型用于空气质量预测。

余长慧[1]等使用seq2seq模型对北京PM2.5浓度进行了预测。刘媛媛[2]等基于CNN-LSTM模型,结合时空特征对山西运城的空气质量指数进行了预测,获得了较好效果。王克丽[3]等使用双层LSTM模型获得了较高的分类精度。周聪[4]等提出基于带有信息增益的LSTM模型对广州空气质量进行综合分析,也获得了较好效果。本文设计了一种LSTM结合Transformer的多头注意力模型,在关注数据泛化特征的同时,保持长时间记忆,从而提升预测效果。

1 模型构造

本提出一种LSTM与Transformer相结合的空气质量预测模型,该模型由LSTM层、Transformer层和预测层组成。

1.1 LSTM。LSTM(长短期记忆网络,Long Short-Term Memory)是一种递归结构的特殊神经网络(RNN),常用来处理和预测与时间相关的有序数据。LSTM的核心是其记忆单元与门控机制,包括遗忘门、输入门和输出门。它们共同决定了如何处理和更新网络的状态。其中遗忘门决定当前时刻需要丢弃的信息[5]。并根据当前输入x_t,和上一时刻的隐藏状态h_{t-1},而后通过sigmoid激活函数计算出一个介于0到1之间的值,公式如下：

$$f_t = \sigma\left(W_f \cdot \left[h_{t-1}, x_t\right] + b_f\right)$$

输出门决定当前时刻网络的输出。通过sigmoid激活函数计算输出门的值,然后将当前时刻的记忆单元状态C_t通过tanh处理,再结合输出门的结果生成最终的隐藏状态。

1.2多头注意力机制。自注意力机制可捕捉序列中任意2个位置之间的依赖关系,而不受距离的限制。自注意力机制通过对序列中元素之间的关系进行建模,计算序列中各元素的权重。与传统的固定权重的注意力机制不同,自注意力允许模型在不同位置之间动态地分配注意力。这使得模型可以更加灵活地捕捉输入序列内部各个元素之间的长程依赖关系,从而提升模型的表示能力[6]。

自注意力机制将输入序列中的每个元素表示转换为三种不同的向量：查询向量Q(query)、键向量K(key)、值向量V(value)。通过计算查询向量与所有键向量之间的相关性得到注意力分数，然后将这些注意力分数与对应的值向量相乘并加权求和，从而得到输出向量。如公式所示：

$$\text{Attention}(Q,K,V)=\text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

式中，Attention()为注意力函数，Softmax为归一化函数，d_k为向量维度。

多头注意力机制(Multi-head Attention)是对自注意力机制的一种扩展，能够增强模型对不同表示空间的关注能力，提升其表征能力和学习能力。多头注意力的公式为：

$$\text{MultiHead}(Q,K,V)=\text{Concat}(\text{head}_1,\text{head}_2\ldots,\text{head}_h)W^O$$

式中，MultHead为多头注意力函数，Concat为拼接函数，head_h为第h个头的注意力权权重，W^O多头自注意力函数的线性映射矩阵。

1.3 Transformer架构。Tansformer模型主要由由编码器(Encoder)和解码器(Decoder)两部分组成，它们之间通过注意力机制进行交互，每部分包含多个相同的层，每层又包含多头自注意力子层、层规范化、残差链接和前馈神经网络子层。

序列数据因其时序性不能直接输入模型需经过编码处理后才能输入，编码分为数据编码和位置编码，它可增强Transformer模型在时间序列预测任务上的表现。位置编码公式为：

$$PE_{(pos,2i)}=\sin\left(\frac{pos}{10000^{\frac{2i}{d_{\text{model}}}}}\right)$$

$$PE_{(pos,2i+1)}=\cos\left(\frac{pos}{10000^{\frac{2i}{d_{\text{model}}}}}\right)$$

式中，dmodel为预设的数据维度，本文取值为512；s为数据点所在序列的绝对位置信息。

编码器的主要作用是将输入序列映射到一个连续的表示空间。它由多个相同结构的层堆叠而成，每层都包含以下两个子层，多头自注意力层，这是Transformer的核心组件之一。它允许编码器在输入序列内部建立长距离的依赖关系，同时能够并行处理不同位置的信息，提升了模型的并行计算能力和学习效率。通过多层堆叠，输入序列的信息被逐渐丰富和抽象化，最终形成了编码器的输出表示。编码器中第I层的多头注意力子层的输出为：

$$\begin{cases}\mathbf{M}=\mathbf{C}(\text{head}_1,,\text{ead}_2,\cdots,\text{head}_H)\mathbf{W}^o\\ \text{head}_I=A_{\text{Mask}}\left(\mathbf{QW}_I^Q,\mathbf{KW}_I^K,\mathbf{VW}_I^V\right)\end{cases}$$

解码器根据编码器的输出和已生成的目标序列来预测下一个目标序列。解码器由N个相同的层堆叠而成，每层包括多头自注意力子层、编码器–解码器注意力子层和前馈神经网络子层。解码器的多头自注意力子层使用一个掩码来遮盖未生成的位置。带掩码的自注意力为：

$$A(\mathbf{Q},\mathbf{K},\mathbf{V})=\text{softmax}\left(\frac{\mathbf{QK}^T}{\sqrt{d_K}}\right)\mathbf{V}\ \text{Mask}$$

式中Mask为掩码操作。

层规范化和残差链接对矩阵进行处理，层规范化和残差连接能够在一定程度上解决深度神经网络因层数过多而导致的梯度消失和网络退化问题，加速模型预测结果的收敛。进入前馈神经网络，前馈神经网络的计算公式为：

$$F(k)=\max\left(0,k\mathbf{W}_1+\mathbf{b}_1\right)\mathbf{W}_2+\mathbf{b}_2$$

式中：F(k)为前馈神经网络函数；k为规范化层的输出值；W_1、b_1分别为第1个线性层的权重矩阵和偏置向量；W_2、b_2分别为2个线性层的权重矩阵和偏置向量。

2 实验分析和评估

2.1数据描述。本文使用广州2013年4月到2022年10月之间的空气质量指数原始数据进行实验分析，同时数据还包括PM2.5,PM10,SO2,NO2,CO,O3等辅助预测特征。采样间隔为24小时，共有3642个数据点。按照8-1-1的比例进行训练集，验证集，测试集划分。

2.2缺失值处理。针对数据中的缺失值问题，目前主要对处理数据缺失值的方法有：删除记录或变量、使用平均值、中位数或众数进行简单填充、利用线性或多项式插值法。在空气质量检查数据处理中，填充缺失值的一种常见方法是使用数据的平均值进行填充。

2.3异常值处理。空气监测传感器时常因为故障而导致出现异常值，为保证数据的合理性，需进行异常值检测处理。通常，将明显过高或过低且不符合测量曲线变化规律的数值称为异常值。判断异常值的方法有很多，其中一种常用的方法是采用3σ准则作为判断标准。基于此假设来计算标准偏差σ。通过确定一个概率区间，来判断数据的合理性。位于这个区间内的数据被认为是合理值，而超出该范围的数据则被视为异常值。

$$\sigma=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2}$$

式中n表示总体样本数量，xi是每个样本的观测值，μ是总体样本的平均值。当数据值位于区间(μ-3σ，μ+3σ)内时，认为该数据是正常的。若数据落在该区间之外，则判定其为异常值并予以舍弃。舍弃异常值后，再按照处理缺失值的方法进行处理。

2.4归一化处理。经过预处理后还需对数据进行归一化处理，以确保模型能够更好地处理和理解数据。归一化处理能够将不同量纲的数据转换到相同的尺度范围内，减少特征间的差异对模型的影响本文采用min-max标准化将数据缩放到[0,1]。

经过数据预处理的缺失值、和异常值的处理后累计获取到数据3542条，进一步构造数据集并按照8:2的比例切分训练集和测试集。

2.5模型评估指标。为对建立空气质量预测模型进行训练评估，本研究选用R2(决定系数)、MAE(平均绝对误差)和RMSE(均方根误差)作为模型的性能评估指标，其中R^2通常用来衡量模型拟合数据的程度，R2越接近1表示模型对数据的拟合越好。其计算如下：

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

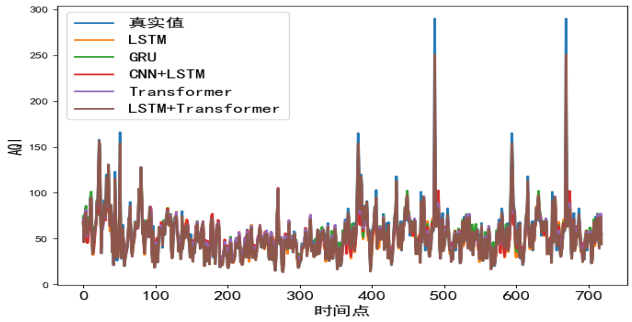
MAE用来衡量模型预测误差的平均绝对值，MAE越小表示模型的预测误差越小，其计算公式如下：

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

RMSE综合考虑了预测误差的平方，RMSE越小表示模型的预测精度越高，其计算公式如下：

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

其中，y_i代表真实值，$\hat{y}_i$ 是预测值，$\bar{y}$ 是真实值的均值。



模型	MAE	RMSE	R2
LSTM	3.7875	5.6938	0.9462
GRU	3.9383	5.9038	0.9428
CNN-LSTM	3.1684	5.0484	0.95668
Transformer	2.7555	4.4166	0.96591
本文模型	**2.3355**	**3.5618**	**0.9756**

2.6模型预测结果对比分析。为了验证LSTM-Transformer在空气质量监测预测问题上的优越性，我们对多种模型进行了对比分析，包括LSTM、GRU、CNN-LSTM和Transformer模型。图为各模型预测结果对比，从实验结果来看，LSTM-Transformer模型能够及时响应环境变化，准确预测风光功率的整体变化趋势，无论是在峰值还是谷值时，都能与真实值良好贴合。相较于LSTM，GRU模型在处理长序列时可能会遭遇梯度消失的问题，这使得它在捕捉长距离依赖关系时的表现不够理想。相反Transformer的自注意力机制具有自动选择和关注输入序列中最重要部分的能力，能够有效捕捉长距离依赖关系，从而提升预测的准确性。

在此基础上，LSTM-Transformer模型结合了LSTM在时间序列建模方面的优势和Transformer在特征表达方面的能力，使得模型能够生成更为丰富的特征表示。这种融合不仅提升了预测精度，还增强了模型对复杂动态变化的适应能力，从而在空气质量监测预测中展现出更为优越的性能。通过这次对比分析，我们可以清晰地看到不同模型在处理空气质量预测任务时的表现差异，进一步验证了LSTM-Transformer模型的有效性与实用性。

为了更清晰的对比不同方法，分别计算各模型的MAE、RMSE、R2指标，从而衡量各个模型的预测效果。预测效果如图表所示。

相比于LSTM，GRU，CNN-LSTM和Transformer，LSTM-Transformer预测模型MAE分别降低了38.35%，40.72%，26.29%，15.25%。而RMSE分别降低37.45%，39.68%，29.46%，19.38%。本文的预测模型在空气质量检查预测结果评判中取得最优，表明该模型能够有效的提高预测精度。

3 结束语

本文针对空气质量预测任务提出了一种基于LSTM+Transformer模型的预测方法，提高了预测的精度和模型的泛化能力。实验结果表明，所提出的模型在MAE和RMSE等评判指标上，均优于LSTM、GRU、CNN-LSTM和单独的Transformer模型。这表明该模型能够更好地捕捉复杂时序数据中的长距离依赖关系和关键特征变化，提升了空气质量预测的准确性。

[参考文献]

[1]余长慧,刘良.基于注意力机制的Seq2seq模型在PM2.5浓度预测中的研究[J].测绘地理信息,2023,48(04):126-131.

[2]刘媛媛.融合CNN-LSTM和注意力机制的空气质量指数预测[J].计算机时代,2022(01):58-60.

[3]王克丽,卢照.基于双层LSTM模型的空气质量预测[J].电脑与信息技术,2024,32(01):51-55.

[4]周聪,卢杰.基于深度学习的空气质量综合分析系统[J].科学技术创新,2024(21):67-70.

[5]Felix Gers,Jürgen Schmidhuber,Fred Cummins.Learning to Forget:Continual Prediction with LSTM[J].Neural Computation, 2000.

[6]Karim Ahmed,Nitish Shirish Keskar,Richard Socher.Weighted Transformer Network for Machine Translation[J/OL].https://arxiv.org/pdf/1711.02132.

作者简介：

闫凯(1988--),硕士,助教,研究方向：嵌入式系统应用、边缘计算、音频处理等。