

机器学习-时间序列分析

韩琳
hanlin3309@163.com

Abstract

本次报告基于模式识别课程内容学习的LSTM等时间序列模型，通过Kaggle上天气和污染水平的有关数据，建立空气污染时间预测模型。除了使用LSTM之外，该报告使用GRU、Transformer方法对空气污染数据进行预测，最终能基本地预测空气污染地趋势，实现了课堂内容的应用。

Introduction

LSTM（长短期记忆网络）是一种循环神经网络（RNN），专门设计用于捕捉长期依赖关系，有效处理序列数据。GRU（门控循环单元）是LSTM的简化版本，具有较少的参数，使其在某些情况下更快速且高效。Transformer模型则基于自注意力机制，通过并行处理序列数据，大幅提升了处理速度和性能，适合于长序列建模。以上三种模型在时间序列预测中表现出色，广泛应用于气象、金融和空气质量等领域。

Methodology

一、LSTM

LSTM（长短期记忆网络）是一种特殊的循环神经网络（RNN），旨在捕捉长期依赖关系。其组成包括单元状态和三种门控机制：遗忘门（控制保留信息），输入门（决定新信息输入），以及输出门（决定输出内容）。LSTM广泛应用于时间序列预测（如气象和金融）、自然语言处理（如文本生成和机器翻译）、语音识别及视频分析等领域，因其卓越的序列建模能力而备受青睐。

核心概念

- 单元状态 (Cell State)**：LSTM网络的主要功能是维护一个单元状态，它类似于传送带，可以保持信息相对不变，这使得信息可以有效地长时间存在于网络中。

2. 门控机制：

- 遗忘门 (Forget Gate)：决定丢弃哪些旧信息。
- 输入门 (Input Gate)：控制新信息的输入。
- 输出门 (Output Gate)：确定下一个隐藏状态的输出。

公式

对于每个时间步 (t)，LSTM的更新过程包括以下公式：

1. 遗忘门：

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

其中，(f_t) 是遗忘门的激活值，(W_f) 是权重矩阵，(b_f) 是偏置，(h_{t-1}) 是上一个时间步的隐藏状态，(x_t) 是当前输入。

2. 输入门：

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

其中，(i_t) 是输入门的激活值，(tilde{C}_t) 是新记忆内容。

3. 更新单元状态：

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

这里，(C_t) 是当前时间步的单元状态，(ast) 表示逐元素相乘。

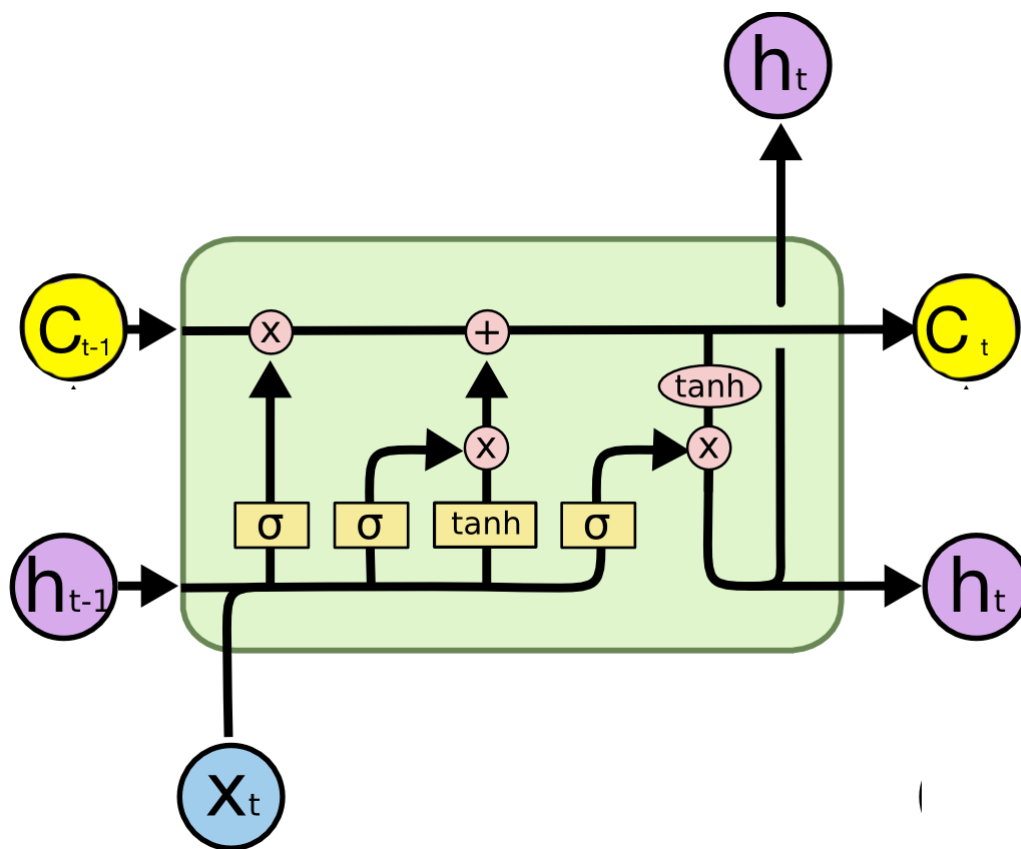
4. 输出门：

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t * \tanh(C_t)$$

其中，(o_t) 是输出门的激活值，(h_t) 是当前时间步的隐藏状态。

这些公式共同工作，使得LSTM能够在处理序列数据时有效地选择性保留和遗忘信息，从而优化学习和预测性能。



2.算法流程

- 数据处理
- 滑动窗口法构建LSTM序列
- 构建LSTM模型
- 训练模型
- 评估模型

3.实验结果

1. 训练损失 (Training Loss) 与测试损失 (Test Loss) :

- **训练损失下降趋势良好:** 从图中可以看到，训练损失在训练过程中逐渐降低，表明模型正在学习数据的规律并逐渐适应训练集。
- **测试损失波动较大:** 测试损失在训练过程中出现了一定的波动，尤其在训练后期，损失开始上升。这可能意味着模型在测试集上的表现不如训练集，可能存在过拟合问题。

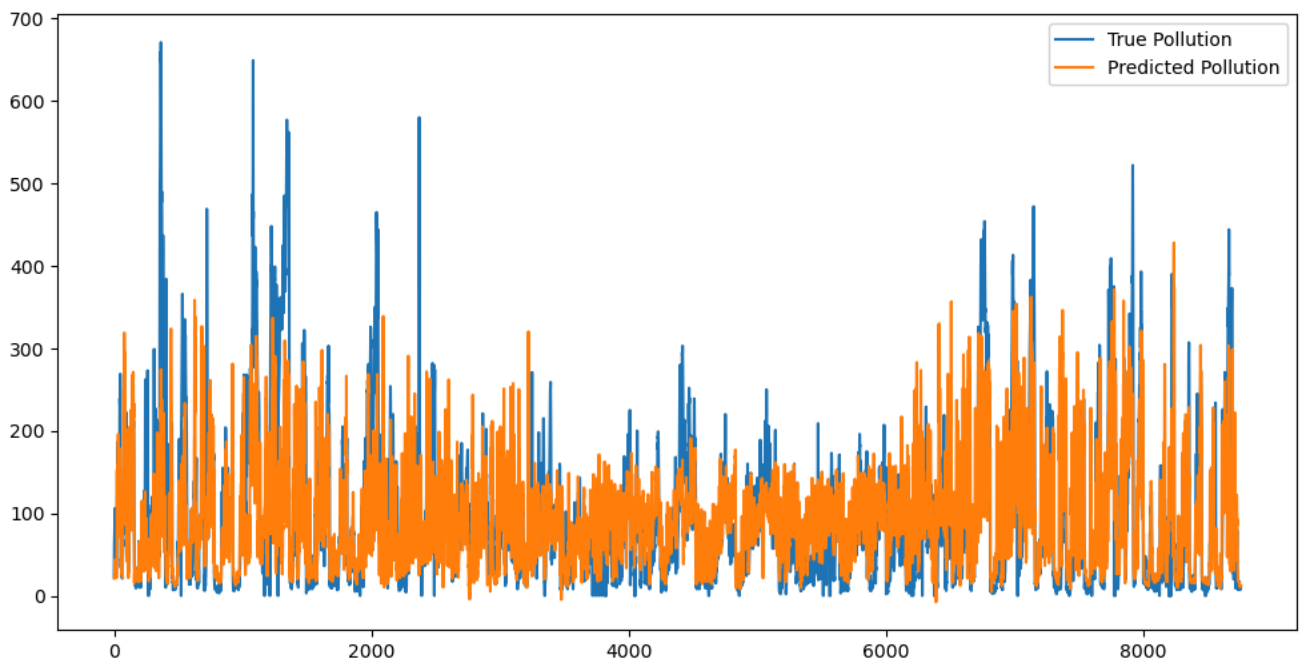
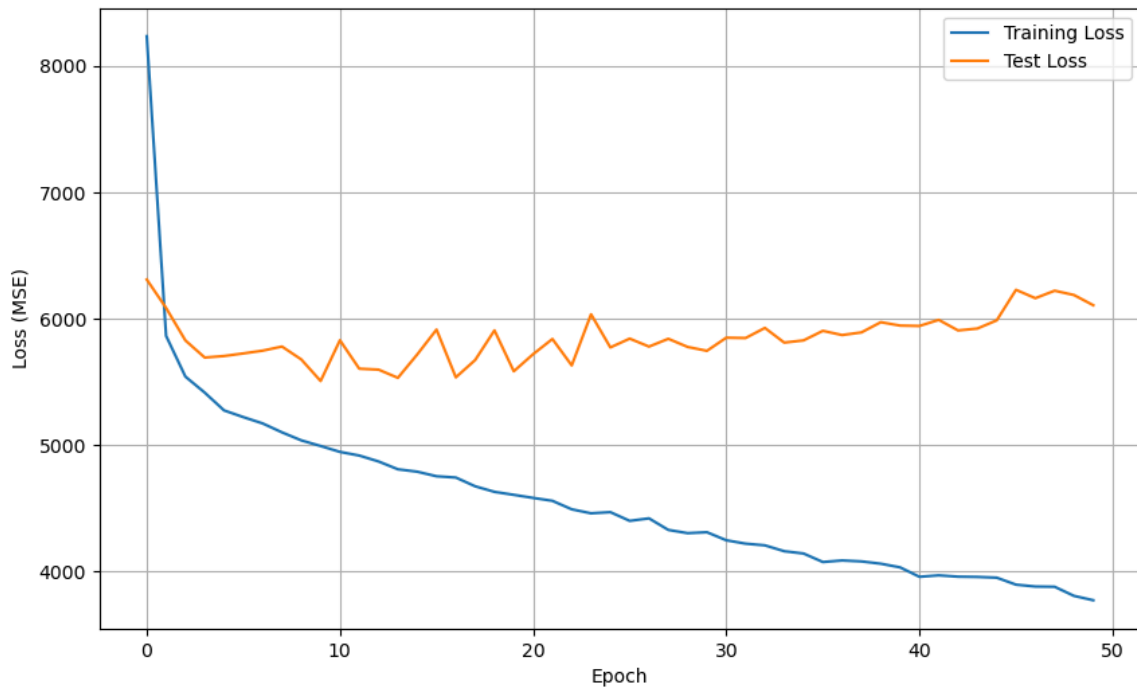
2. R^2 值:

- 模型的 R^2 值为 **0.2485**，这个值相对较低，表明模型对目标变量的解释能力有限。理想情况下， R^2 值接近1意味着模型能够较好地拟合数据，而较低的 R^2 值表示模型在预测上的误差较大。

3. 真实值与预测值的对比:

- 从下图的曲线对比来看，**预测值与真实值之间的差异较大**。虽然两者趋势相似，但预测值的波动性较强，特别是在高污染值的情况下，预测值出现了较大的偏差。
- 这种差异可能是由于模型没有完全捕捉到数据中的所有复杂性，或是训练数据的某些特征未能充分表示。

LSTM模型 R^2 : 0.2485



二.GRU预测

GRU（门控循环单元，Gated Recurrent Unit）是一种改进的循环神经网络（RNN），旨在解决传统RNN在处理长序列时的梯度消失问题。GRU通过引入门控机制，能够有效捕捉序列中的长期依赖关系，同时相较于LSTM（长短期记忆网络），GRU的结构更加简洁，参数更少，因此在一些任务中能够更快地训练,GRU广泛应用于时间序列预测、自然语言处理、语音识别、机器翻译等领域，尤其在需要处理长时间序列的任务中表现优秀。

1.GRU的核心概念

GRU主要由以下两个门组成：

- 1. **更新门（Update Gate）**： 决定当前时间步的隐藏状态有多少需要保留，多少需要更新。它通过对前一时刻的隐藏状态和当前输入进行加权平均来控制信息的传递。
- 2. **重置门（Reset Gate）**： 决定丢弃多少前一时刻的隐藏状态。当重置门的值接近0时，GRU会忽略前一时刻的隐藏状态，只关注当前输入的信息；当重置门的值接近1时，前一时刻的隐藏状态对当前状态的影响较大。

GRU公式

- 1. 重置门：

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r)$$

- 2. 更新门：

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z)$$

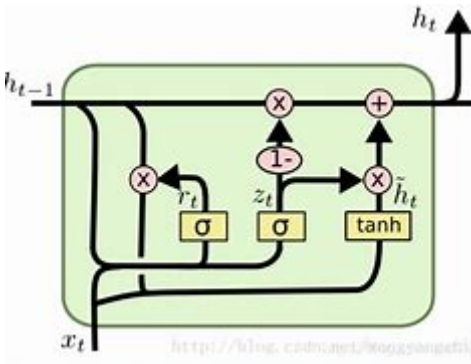
- 3. 候选隐藏状态：

$$\tilde{h}_t = \tanh(W_h \cdot [r_t * h_{t-1}, x_t] + b_h)$$

- 4. 最终隐藏状态：

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

其中， r_t 是重置门， z_t 是更新门， \tilde{h}_t 是候选隐藏状态， h_t 是当前时间步的隐藏状态， σ 是sigmoid激活函数， \tanh 是双曲正切激活函数。



2.算法流程

- 数据处理
- 滑动窗口法构建GRU序列
- 构建GRU模型
- 训练模型
- 评估模型

3.实验结果

1. 训练与测试损失:

- **训练损失**: 从图中可以看到, 训练损失(蓝线)在前几轮迅速下降, 表明模型在学习过程中有效地拟合了训练数据。训练损失保持平稳, 且没有出现过拟合的迹象。
- **测试损失**: 测试损失(橙线)波动较大, 尤其在训练后期, 测试损失有一定的上升趋势。模型对测试集的泛化能力较弱, 存在过拟合的风险。

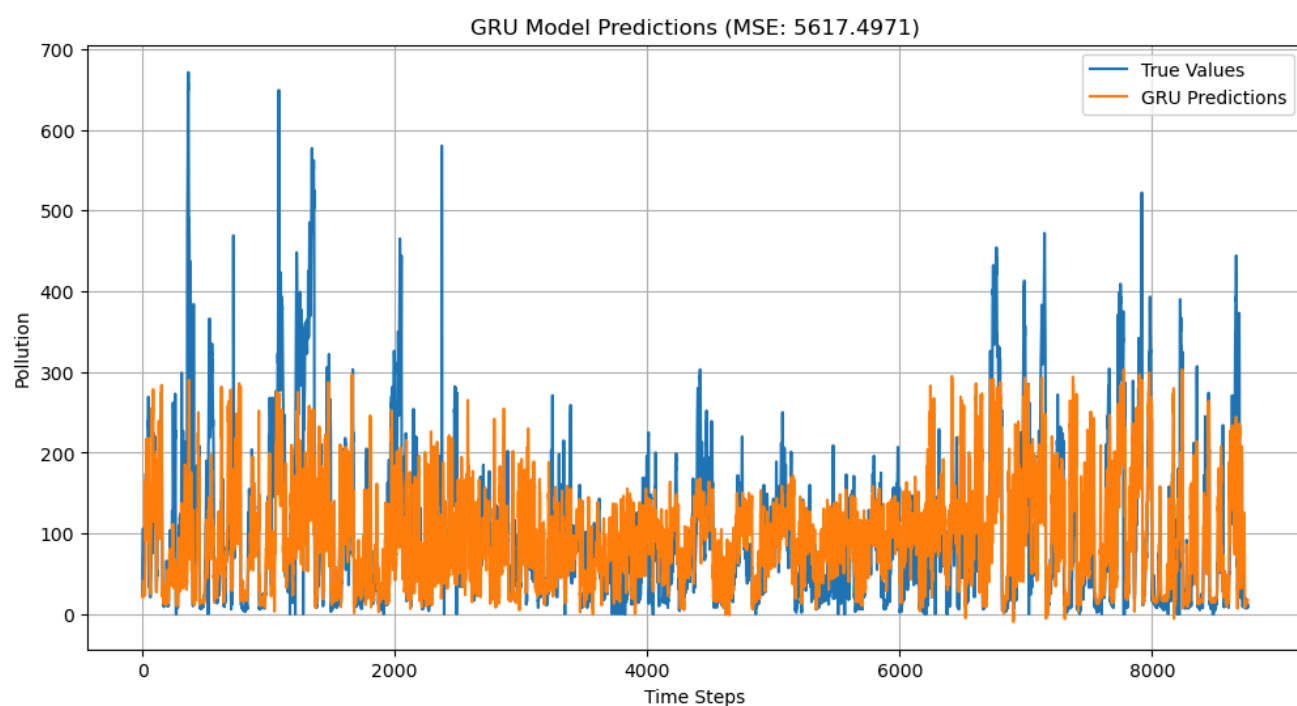
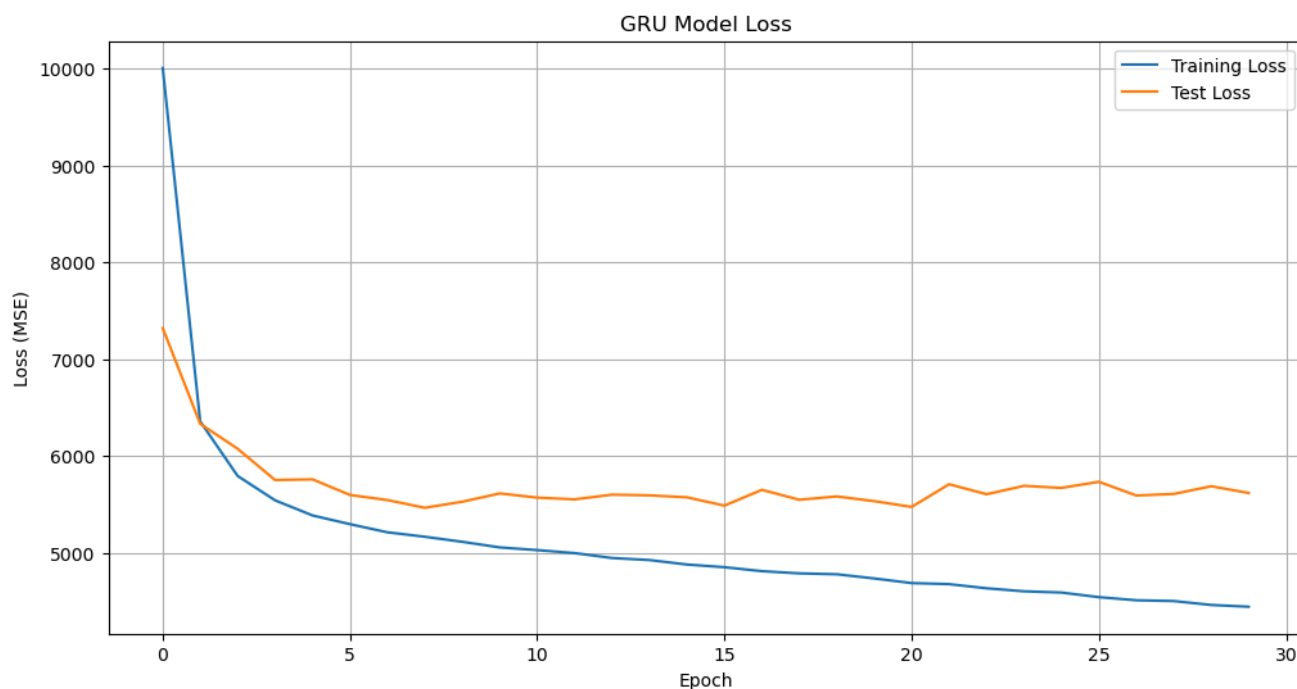
2. R^2 值:

- 模型的 **R^2 值为0.3584**, 这个值表明模型对数据的解释能力较弱。理想情况下, R^2 值应该接近1, 说明模型可以很好地解释数据中的方差。0.3584的 R^2 值表示模型的预测能力有限, 但是强于lstm模型, 仍然需要进一步优化

3. 预测结果与真实值对比:

- **真实值与预测值的差异**: 从第二张图中可以看到, GRU的预测值(橙线)与真实值(蓝线)之间存在明显的差异, 尤其是在一些高峰值部分, 预测值有较大的偏差。虽然整体趋势相似, 但在一些细节上, 模型预测的精度仍有待提高。
- 这种差异可能是由于模型在捕捉高波动性数据时的不足, 或者训练数据没有涵盖所有可能的模式, 导致模型在某些情况下预测不准确。

GRU模型 R^2 : 0.3584



三.Transformer预测

Transformer模型是一种基于自注意力机制的深度学习架构，最初由Vaswani等人在2017年提出，专门用于处理序列到序列（seq2seq）任务，尤其在自然语言处理（NLP）领域取得了巨大成功。与传统的循环神经网络（RNN）和长短期记忆网络（LSTM）不同，Transformer并不依赖于递归或时间步长顺序，而是通过自注意力机制来并行处理整个输入序列，极大地提高了计算效率和性能。

1.核心概念

1. 自注意力机制 (Self-Attention) :

- Transformer的核心创新是自注意力机制，允许模型在处理当前输入时，关注序列中所有其他位置的信息。通过计算每个词与其他词之间的关系权重，模型能够捕捉输入序列中的长距离依赖关系。
- 自注意力计算公式如下：

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

其中， Q (Query)、 K (Key) 和 V (Value) 是从输入序列中计算出的向量表示， d_k 是查询向量的维度。

2. 多头注意力机制 (Multi-Head Attention) :

- Transformer通过多头注意力机制并行计算多个自注意力头，从不同的子空间中学习信息。这使得模型能够更全面地捕捉输入序列中的信息。

3. 位置编码 (Positional Encoding) :

- 由于Transformer不依赖于递归结构，不能直接获取输入序列的顺序信息，因此使用位置编码 (Positional Encoding) 将位置信息注入到模型中，以帮助模型理解序列顺序。

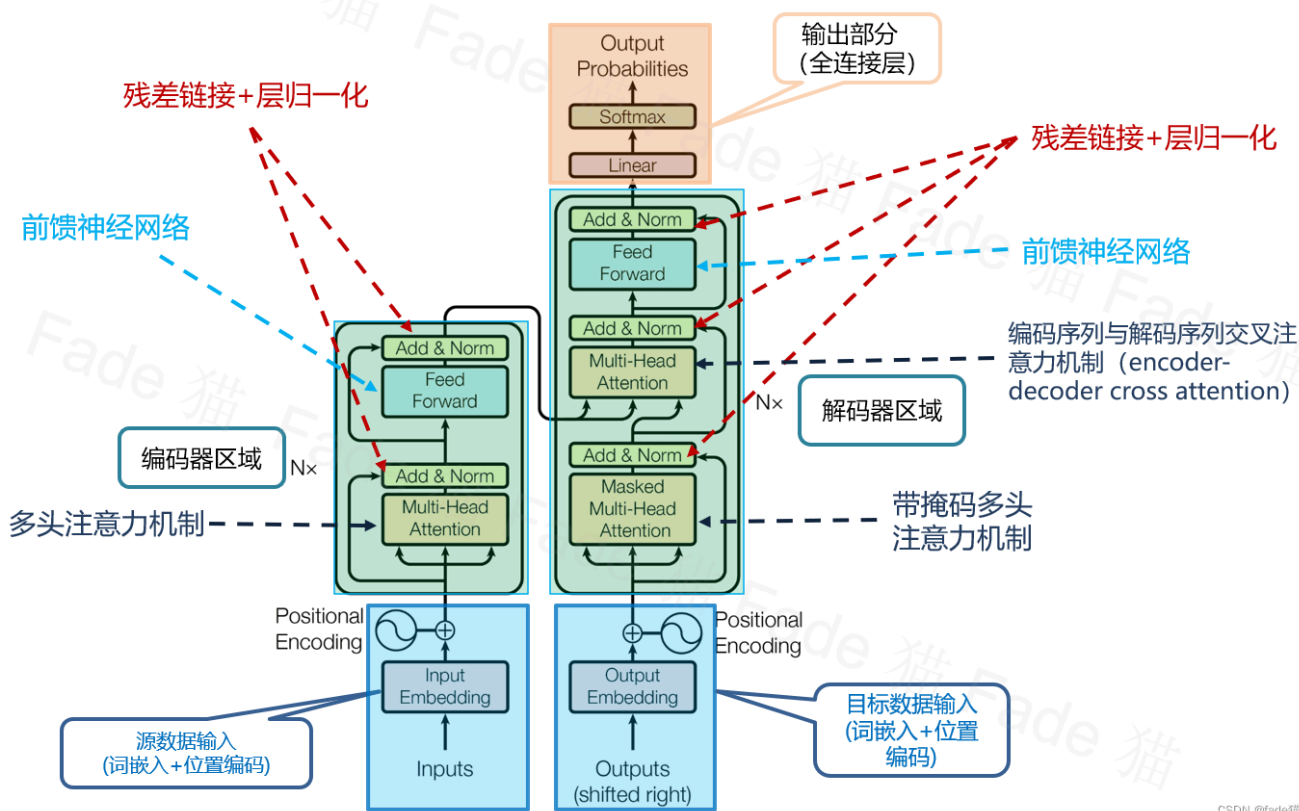
4. 前馈神经网络 (Feed-Forward Network) :

- Transformer的每一层还包括一个全连接的前馈神经网络，用于处理自注意力机制后的信息。该网络通常包含两个线性层和激活函数（如ReLU）。

2.Transformer架构

Transformer的整体架构分为编码器 (Encoder) 和解码器 (Decoder) 两部分，常用于序列到序列任务（例如机器翻译）：

- 1. 编码器：** 负责处理输入序列，将其转换为一个上下文相关的表示。每个编码器层由两个主要部分组成：
 - 多头自注意力机制
 - 前馈神经网络
- 2. 解码器：** 负责生成输出序列，逐步预测目标序列的每个词。解码器每一层包含三个主要部分：
 - 多头自注意力机制
 - 编码器-解码器注意力机制（用于关注编码器输出的相关信息）
 - 前馈神经网络



CSDN @fade猫

3.实验结果

Transformer 模型预测性能总结

1. R^2 值:

- 模型的 R^2 值为 **0.3276**, 表示模型解释了大约 **32.76%** 的方差。这是一个较低的 R^2 值, 表明模型的预测效果相对较差。

2. 损失函数 (Loss) :

- 训练损失 (Training Loss)** : 从训练损失曲线来看, 随着训练的进行, 损失逐渐下降, 表明模型在训练集上逐步拟合。
- 测试损失 (Test Loss)** : 测试损失在训练过程中波动较大, 并且在最后趋于稳定, 但依然较高。

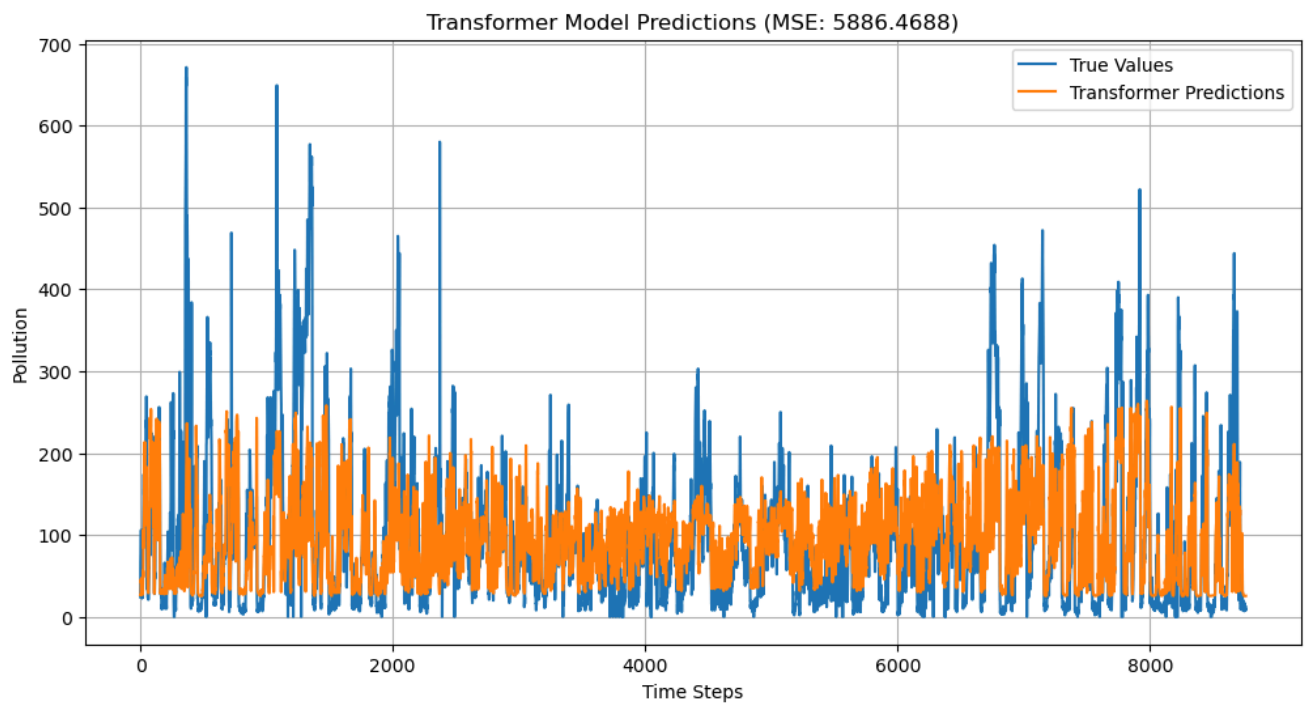
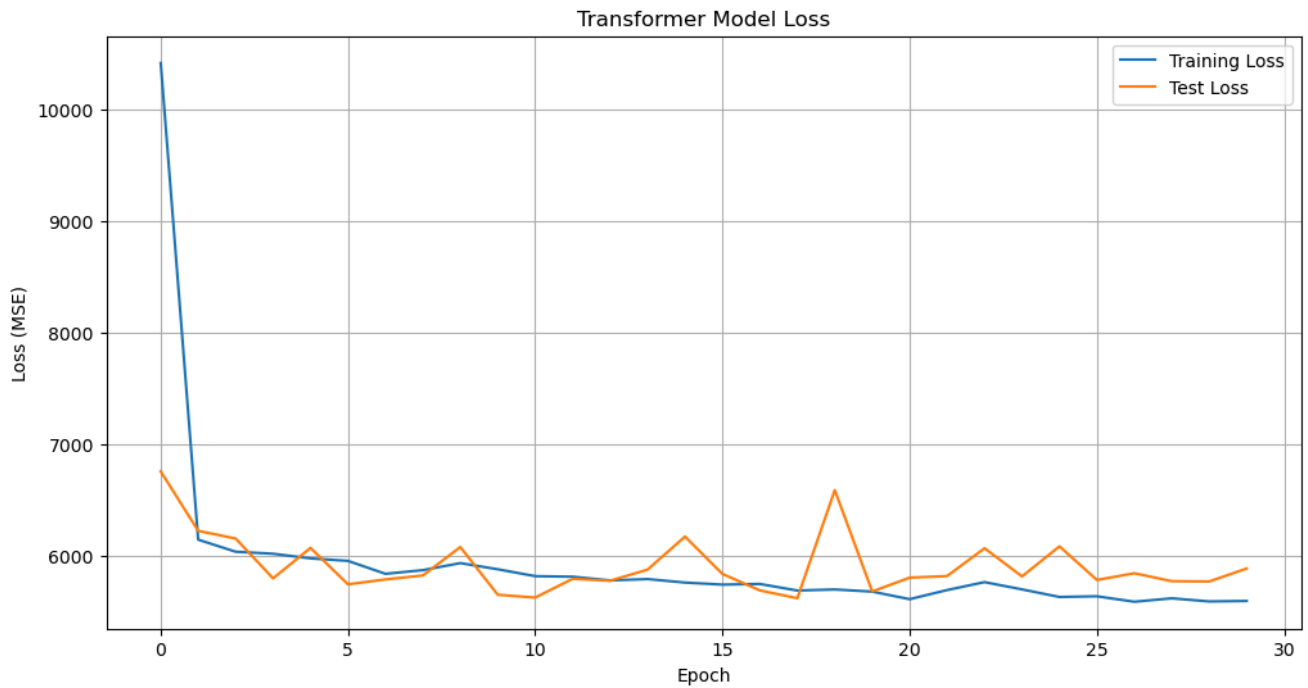
3. 预测值与真实值对比

- 在第二个图表中, 蓝色线代表真实值, 橙色线代表模型预测值。从图中可以看到, 模型的预测值与真实值有明显的偏差。预测值在某些时刻出现了较大的波动, 远离了真实值, 表现出模型未能很好地捕捉数据的真实规律。

总结:

尽管 Transformer 模型在训练过程中逐渐收敛, 但在测试集上的表现不理想, R^2 值和 MSE 都表明该模型的预测能力较弱, 可能需要更多的优化。

Transformer模型 R^2 : 0.3276



Conclusions

根据LSTM、GRU和Transformer模型在天气污染预测中的表现，较低的 R^2 值（均较小）表明模型的预测效果不佳。

- 数据质量问题：**数据可能存在缺失、噪声或不一致，影响模型的学习能力。如果特征选择不恰当或未充分反映污染的关键因素，模型将无法准确捕捉污染水平的变化。
- 模型结构与参数调整：**LSTM、GRU和Transformer模型的复杂性较高，且需要精确调整超参数。若没有合理的超参数调优或优化算法，可能导致过拟合或欠拟合，无法有效泛化到测试数据。

3. **未考虑外部因素：**污染水平不仅受到气象因素影响，还与交通流量、工业排放等外部因素密切相关。如果这些因素未被纳入模型，可能导致预测精度下降。因此，建议在数据预处理和特征工程方面进一步优化，考虑加入外部因素，并进行更细致的模型调优与实验，提升预测效果。
