



高效长序列水位预测模型的研究与实现

黄颖¹, 许剑¹, 周子祺¹, 陈树沛², 周帆^{2*}, 曹晟³

(1. 国能大渡河大数据服务有限公司技术管理中心 成都 610041; 2. 电子科技大学信息与软件工程学院 成都 610054;
3. 电子科技大学计算机科学与工程学院 成都 611731)

【摘要】 序列预测旨在利用历史序列模式信息预测未来长时间跨度的趋势, 在工业领域具有众多的实际应用需求。针对工业数据序列预测问题中时序长度较长的特点, 提出了一种高效的自注意力机制以适用于长序列数据建模与预测。该模型构建了新的嵌入表示, 增加了池化操作, 并且使用了生成式推断, 实现长距离依赖建模和时序信号预测。相比之前的自注意力模型, 该模型有效解决了现有方法在面对长序列预测时存在的预测精度不足、训练耗时过长等问题。在大规模水电站水轮机顶盖水位预测这一实际工业应用场景中, 相比其他基准模型, 该模型显著提高了长序列水位预测的精度和效率。

关键词 深度学习; 工业大数据; 长序列预测; 神经网络; 注意力机制
中图分类号 TP391.4 **文献标志码** A **doi**:10.12178/1001-0548.2022133

Research and Implementation of Efficient Long Sequence Model for Water Level Forecasting

HUANG Ying¹, XU Jian¹, ZHOU Ziqi¹, CHEN Shupe², ZHOU Fan^{2*}, and CAO Sheng³

(1. Technology Management Center, CHN Energy Dadu River Big Data Service CO., Ltd Chengdu 610041;
2. School of Information and Software Engineering, University of Electronic Science and Technology of China Chengdu 610054;
3. School of Computer Science and Engineering, University of Electronic Science and Technology of China Chengdu 611731)

Abstract Long-Sequence forecasting aims to model and predict future long-term time series trends by leveraging historical knowledge and patterns and has many practical applications in various industries. To fully utilize long-time series industrial data characteristics, this paper presents an improved self-attention mechanism suitable for modeling and forecasting long sequence industrial data. Our model builds a new embedding representation learning module, combined with the pooling operations, and uses the generative inference for long-range dependency modeling and time-series signal prediction. Compared with the previous self-attention-based method, the proposed model effectively solves the problems of insufficient prediction accuracy and high training cost in long sequence prediction. Our model significantly improves long-sequence water level prediction accuracy and efficiency compared with other benchmark methods. Experiments conducted on the real-world water level data from a large-scale hydropower station proved the superior performance of the proposed model in terms of both effectiveness and efficiency over existing state-of-the-art models.

Key words deep learning; industrial big data; long sequence time-series forecasting; neural network; self-attention mechanism

随着数据采集技术和大数据科学的发展, 水电工业相关领域汇聚了海量数据, 其中, 时间序列数据占据了较大比重。利用丰富的历史时序数据来预测未来一段时间的发展变化在电力特别是水电工业中具有重大的应用前景, 如智能电网调度^[1]、水电

负荷平衡^[2]、水库水位预测^[3]等。

现有的时间序列预测方法主要针对短期的时序预测问题, 而水电领域日益增长的长序列数据和长期趋势预测的需求, 更要求模型具有对长序列的建模与预测能力, 该问题被称为 LSTF (long sequence

收稿日期: 2022-05-09; 修回日期: 2022-10-16

基金项目: 国家自然科学基金(62072077, 62176043); 国家重点研发计划(2019YFB1406202); 四川省科技计划(2020GFW068, 2020ZHCG0058, 2021YFQ0007)

作者简介: 黄颖(1992-), 男, 主要从事时空数据挖掘和大数据分析方面的研究。

*通信作者: 周帆, E-mail: fan.zhou@uestc.edu.cn

time-series forecasting)。输入序列和预测序列的长度是影响时序模型预测性能的两大重要因素，在 LSTF 任务中，输入序列和预测序列的长度都大大增加，这对模型实现准确快速地时序预测提出了巨大挑战。由于 LSTF 任务难度巨大，现有的时间序列预测模型如循环神经网络模型和注意力模型，在处理 LSTF 任务时都存在着许多不足，需要进行更多的针对性改进。

水电站水轮机顶盖水位变化是一个典型的长序列预测问题。由于水轮机主轴密封装置漏水和水泵排水等因素，水轮机顶盖水位随时间不断变化。水轮机顶盖水位是影响水轮机运行安全的重要因素，顶盖水位过高会淹没水导轴承和漏油箱，引起机组用油变质，严重威胁设备的运行安全^[4]。水轮机顶盖积水主要有以下几个来源：主轴密封装置漏水、导叶套筒漏水、真空破坏阀无法复归、顶盖排水空心固定导叶阻塞或者空心固定导叶破坏、顶盖与止漏环冲蚀漏水等^[5]。对顶盖水位及时的监测和预警，可以监测顶盖水泵的运行效率，保证顶盖水位保持在安全可控的范围内，提升水电站运行的安全性。

本文根据 LSTF 问题的特点，对带自注意力机制的时间序列预测模型进行改进，设计了新的时间序列嵌入表示方法，增加池化操作，并结合生成式推断构建了一种高效的长序列预测模型。在大规模数据集上的实验表明，本文模型在水电站水轮机顶盖水位预测这一重要工业应用中，可以对长期水位的时间序列进行高效准确地预测。

1 相关工作

时间序列预测的现有方法大致可以分为两类：经典统计学习模型和深度学习方法。经典的时间序列预测模型基于统计学习方法，具有可解释性强和有完备理论支撑等优点^[6]，后续改进包括对丢失数据的容错^[7]和多种数据类型的支持等^[8]。在面对海量时序数据时，经典统计学习模型存在着性能上的瓶颈。深度学习方法主要利用循环神经网络 (recurrent neural networks, RNN) 及其各种变体如 GRU (gate recurrent unit)^[9]和长短期记忆网络 (long short-term memory, LSTM)^[10]等，实现了时间序列预测性能上的突破。

基于 RNN 的深度学习方法在实现长序列预测时，仍然存在性能问题，尤其是深度学习方法的顺序预测范式，主要存在以下不足。1) 模型的性能不

够稳定。尽管短距离预测结果准确，但它们在长序列预测中会因动态解码而累积误差，模型的精度会随着预测长度的增加而快速衰减，导致最终的预测结果出现较大误差^[11]。2) 由于梯度消失和内存约束问题，大多数现有方法无法从过去长时间的历史序列中学习有效的知识和模式。

为了捕获序列中的长期依赖关系，文献 [12] 提出了注意力机制，以改进自然语言翻译任务中编码器-解码器结构的词对齐。随后，文献 [13] 提出了使用位置注意和点积注意机制等多种变体，它们被广泛应用在各种序列分析预测问题中。采用注意力机制考虑到 RNN (包括 LSTM、GRU 等变体) 的计算机制是顺序的，即 RNN 相关算法只能从左向右或从右向左依次计算，这种机制带来了两个主要问题：1) 时间 t 时刻的计算依赖于 $t-1$ 时刻的计算结果，这种规则限制了模型的并行能力；2) 顺序计算的过程中会出现信息丢失现象，尽管 LSTM 等具有门控机制的模型一定程度上缓解了长期依赖的问题，但对于特别长期的依赖关系，RNN 模型的性能依然不佳。

基于自注意力机制的 Transformer 模型^[14]有效解决了前面两个问题。首先，它使用了自注意力机制，将序列中的任意两个位置之间的距离缩小为一个常量。其次，它避免了类似 RNN 中的顺序结构，具有更好的并行性，能够很好地适用于现有的 GPU 计算框架。但在使用原始的 Transformer 解决 LSTF 问题时，依然存在两个明显的缺陷：1) 模型计算量较大，时间复杂度和空间复杂度都达到了 $O(L^2)$ ；2) 在预测长距离输出时，推理速度大幅下降，因为 Transformer 的动态解码使得其预测时间和基于 RNN 的模型在推理时计算量是一致的。本文利用 Transformer 在解决时间序列预测问题上的优势，对模型进行针对性改进。

2 数据预处理

水轮机状态数据包括当前顶盖水位、机组流量、电压、电流、水轮机机组有功功率、主轴密封磨损量、密封水压力等多种不同类型的数据。由于传感器故障、水轮机工作异常等原因，原始的水轮机状态数据存在很多异常值和缺失值。异常值主要指由于传感器异常或设备运行异常等引发的严重偏离期望值的数据。异常值分为两种类型：设备运行本身的故障和状态异常；传感器示数错误引起的异常值。异常数据会降低模型拟合的精度，而且会导

致后续状态评估出现偏差, 引发误判, 异常数据应该通过数据清洗去除。水轮机状态数据中的缺失值主要是由传感器的短暂失效、通信端口异常、设备检修维护、记录失误等因素造成^[5]。通过剔除所有应被数据清洗的异常值和补全缺失值, 可以实现水轮机状态数据的数据预处理。

2.1 异常值检测

数据中由于传感器示数异常导致的异常值, 本文使用孤立森林算法^[6]进行检测并剔除。孤立森林是一种集成学习算法, 由多棵孤立树 (isolation tree, itree) 组成, 每棵孤立树都具有二叉树结构, 其核心思想是快速被划分到叶子节点的数据大概率是异常数据。构建一棵孤立树的具体步骤如下: 1) 从训练数据中采样 n 个数据作为孤立树的训练样本; 2) 在训练样本中随机选择一个特征, 并在该特征的最小值和最大值范围内随机选择一个值 P 作为当前孤立树的根节点; 3) 对样本进行划分, 将小于 P 的样本点划分到根节点的左边, 大于 P 值的样本点划分到根节点的右边; 4) 对划分产生的左右两部分重复步骤 2) 和步骤 3) 操作, 直到节点只包含一个样本或者达到树高度的最大限度。

重复上述操作构建 N 棵孤立树, 利用构建好的孤立森林模型检测出异常数据, 使每一个时刻数据遍历所有孤立树, 直到达到终止条件, 计算数据经过的路径长度。在这种随机分割策略下, 异常数据相较于正常数据经过的路径通常都较短, 会最先被分割出来, 进而在数据中被剔除。

2.2 缺失值插补

对于数据中的缺失值, 本文使用随机森林回归算法^[7]进行缺失值的插补。随机森林回归算法采用集成学习的思想, 在多个二叉决策树的构建上使用不同的参数选择, 通过组合多棵决策树进行最终预测。在训练阶段, 随机森林使用 bootstrap 采样方法从训练数据中采集多个不同的子训练数据集来依次训练多个不同的二叉决策树。在预测阶段, 随机森林将所有二叉决策树的预测结果取平均得到最终结果。二叉决策树的预测过程分为以下步骤: 1) 输入某一样本, 从二叉决策树的根节点起, 判断当前节点是否为叶子节点: 如果是, 则返回叶子节点的预测值 (即当前叶子中样本目标变量的平均值), 否则进入下一步。2) 根据当前节点的切分变量和切分值, 将样本中对应变量的值与节点的切分值作对比。如果样本变量值大于当前节点切分值, 则访问当前节点的右子节点; 如果样本变量值小于

等于当前节点切分值, 则访问当前节点的左子节点。3) 循环步骤 2), 直到访问到叶子节点, 并返回叶子节点的预测值。

2.3 数据归一化

利用孤立森林算法和随机森林回归算法实现了对数据的异常值处理和缺失值插补后, 由于原始数据中不同特征数据的量级及变化范围不同, 本文对清洗后的不同特征数据分别进行 MIN-MAX 归一化处理。归一化后的特征值 x' 的具体计算方式为: $x' = \frac{x - \min(X)}{\max(X) - \min(X)}$, 其中 x 表示原始值, $\min(X)$ 与 $\max(X)$ 分别表示该特征的最小值与最大值。

3 水轮机顶盖水位预测模型

本文模型摒弃了传统的卷积神经网络和循环神经网络, 基于自注意力机制构建整个网络结构。模型的主要框架由嵌入输入部分、基于自注意力机制的编码器、基于自注意力机制的解码器和预测输出模块这 4 部分组成。编码器主要由多头注意力模块和池化操作模块堆叠而成, 而解码器主要由掩盖多头注意力模块、多头注意力模块和前馈神经网络模块组成。模型的整体框架如图 1 所示。

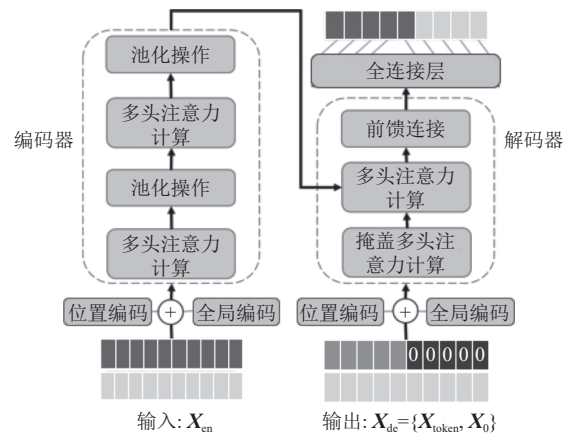


图 1 模型的整体框架示意图

3.1 数据嵌入表示

本文将预处理过的水轮机顶盖水位相关数据拼接起来, 并加入位置编码和全局编码, 作为自注意力模块的输入。 t 时刻的输入表示为 $x_t = \{x_1, x_2, \dots, x_{L_x} | x_i \in R^{d_x}\}$, 其中, L_x 指输入序列的长度, d_x 代表相关特征值的个数。由于自注意力机制将序列中任意两个位置之间的距离缩小为一个常量, 需要在嵌入表示中显式地增加位置编码, 使模型能够区分不同时刻的输入。位置编码可以分成相对位置编码和绝对位置编码两种类型, 本文使用绝对位置编码。

具体地, 模型使用了正、余弦函数位置编码。对于在输入序列中位置为 pos 的输入向量, 其位置编码 PE 的数值为:

$$\text{PE}_i(\text{pos}) = \begin{cases} \sin\left(\frac{\text{pos}}{(2L_x)^{2i/d_x}}\right) & \text{pos} = 2i \\ \cos\left(\frac{\text{pos}}{(2L_x)^{2i/d_x}}\right) & \text{pos} = 2i + 1 \end{cases} \quad (1)$$

式中, pos 表示当前输入在输入序列中的位置; i 表示该输入在对应的位置编码向量中的位置, 范围是 $i \in \{1, 2, \dots, \lfloor L_x/2 \rfloor\}$; L_x 表示输入序列的长度。位置编码 PE_i 即是用正弦余弦对表示原有序列中的位置。最终的嵌入编码 \mathbf{X}_i^{en} 定义为:

$$\mathbf{X}_i^{\text{en}}(j) = x_i(j) + \text{PE}_i(j) \quad j \in \{1, 2, \dots, L_x\} \quad (2)$$

整体数据嵌入 \mathbf{X}_i^{en} 是原始序列 x_i 与其位置编码信息 PE_i 按序列中的位置级联到一起构成的, 结合原始信息和全局位置上下文, 将作为自注意力模块的输入。

3.2 多头注意力计算

多头注意力模块是多个不同自注意力模块的集成, 是编码器和解码器的关键组成部分。自注意力模块的计算主要由 (query, key, value) 三元组构成:

$$\mathbf{Z} = \text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}}\right)\mathbf{V} \quad (3)$$

式中, \mathbf{Q} 表示查询 (query), $\mathbf{Q} \in R^{L_Q \times d}$; \mathbf{K} 表示键 (key), $\mathbf{K} \in R^{L_K \times d}$; \mathbf{V} 表示值 (value), $\mathbf{V} \in R^{L_V \times d}$ 。通过式 (3) 得到注意力特征向量 \mathbf{Z} , \mathbf{Z} 反应了不同位置输入之间的相互关系。 \mathbf{Q} 、 \mathbf{K} 、 \mathbf{V} 通过 3 个不同的权值矩阵与嵌入矩阵 \mathbf{X} 相乘得到: $\mathbf{Q} = \mathbf{W}^Q \mathbf{X}$, $\mathbf{K} = \mathbf{W}^K \mathbf{X}$, $\mathbf{V} = \mathbf{W}^V \mathbf{X}$, 其中, \mathbf{W} 为注意力头部的权值矩阵。利用多组不同的 \mathbf{W} 将嵌入向量进行多次不同的映射, 以捕捉映射到多个线性空间更加丰富的特征, 实现多头注意力计算。实际实验中, 一般取 $L_Q = L_K = L_V$, 使得 \mathbf{Q} 、 \mathbf{K} 、 \mathbf{V} 这 3 个矩阵的维度相同。将多个注意力头部的计算结果拼接并进行线性转换就得到了最后的特征矩阵 \mathbf{Z} 。在本文实验中, 模型使用了 6 个注意力头, 即 6 组不同的 \mathbf{W} 矩阵。另外, 在自注意力计算的过程中, 使用了残差网络中的 short-cut 结构, 以缓解深度学习中的梯度消失和梯度爆炸问题。

3.3 池化操作

得到特征矩阵 \mathbf{Z} 后, 原始的 Transformer 会使用一个双层前向反馈网络计算得到隐藏向量。为了减少模型的时间复杂度和空间复杂度, 且自注意力

计算中存在特征映射的冗余, 本文模型在编码器的计算中加入池化操作替换了原有的前向反馈网络。池化操作对具有主导特征的优势特征赋予更高的权重, 在下一层生成具有主导性的特征映射, 可以大幅削减输入的时间维度, 具体计算如下:

$$\mathbf{Z}_{s+1} = \text{MaxPool}(\text{SELU}(\text{Conv1d}([\mathbf{Z}_s]_{\text{AB}}))) \quad (4)$$

式中, $[\cdot]_{\text{AB}}$ 表示自注意力模块; s 表示池化操作的层数。通过卷积核为 3 的卷积、SELU 非线性变换和步长为 2 的 MaxPool, 经历两次下采样, 降低了计算量和内存消耗, 编码器的空间复杂度从 $O(L^2)$ 下降为 $O(L \log L)$ 。本文模型将多头注意力模块和池化模块堆叠使用 (如图 1 所示) 得到编码器最终的隐藏表示。

3.4 长序列预测

模型解码器使用标准 Transformer 解码器结构, 由两个多头注意力计算模块和一个前馈连接模块堆叠而成。首先, 解码器的输入是将预处理过的水轮机顶盖水位相关的起始数据和一段填充数据拼接在一起, 并添加位置编码和全局编码作为数据嵌入表示, 即解码器的输入。解码器的输入 \mathbf{X}_{de} 计算如下:

$$\mathbf{X}_{\text{de}} = \text{Concat}(\mathbf{X}_{\text{token}}, \mathbf{X}_0) \in R^{(L_{\text{token}} + L_y) \times d_{\text{model}}} \quad (5)$$

式中, $\mathbf{X}_{\text{token}} \in R^{L_{\text{token}} \times d_{\text{model}}}$ 是起始序列; $\mathbf{X}_0 \in R^{L_y \times d_{\text{model}}}$ 是目标序列的占位符, 设置为 0。

不同于编码器, 解码器会将数据嵌入表示输入到掩盖多头注意力计算模块, 从而学习解码器输入数据之间的相互作用关系。掩盖多头注意力计算模块需要遮盖输入序列的信息, 并在进行注意力计算时将输入序列中每个位置后面部分点积计算结果设置为负无穷。这样操作可防止每个位置关注未来的位置信息, 从而实现自回归学习和解码。接下来将数据传入另外一个多头注意力计算模块中, 该模块旨在学习编码器输入表示和解码器输入表示之间的关系, 使得解码器在预测时不但关注解码器的输入, 还能关注编码器的长序列输入, 使得模型能够高效地对长距离依赖进行建模。最后将多头注意力计算模块的输出输入到前馈连接模块中, 以丰富数据的表示信息, 增强模型的表达能力。

经过解码器的转换和学习之后, 模型获得输入数据的良好表示。在预测时, 使用一段目标序列之前的已知序列作为起始序列, 将 $\mathbf{X}_{\text{de}} = \{\mathbf{X}_{\text{token}}, \mathbf{X}_0\}$ 输入到生成式的解码器中。然后, 解码器通过一个全

连接网络, 将数据转换成和待预测的信息相同的维度, 预测输出全部预测序列。这种方式能准确地预测出未来一段时间水轮机顶盖水位相关的数据信息, 取代了传统的编解码器体系结构中耗时的“动态解码”, 极大地提高了解码速度。

在模型的训练过程中, 使用均方误差 (mean square errors, MSE) 作为最终的损失函数, 计算模型输出和真实序列值之间的误差, 进行梯度的反向传播, 实现模型参数的优化。由于注意力机制的存在, 模型可以并行运算, 即每个时刻都会去预测下一个时刻的结果而不依赖于其他数据, 不像循环神经网络一样每一步都需要依赖前一步的计算结果, 这种方式极大地提高了模型的训练速度。

4 实验

4.1 数据集

本文从国家能源集团大渡河公司大岗山水电站获取了 2018—2020 年的水电站运行数据, 选取了与水轮机顶盖水位相关的多个测点作为模型的输入特征。相关数据包括了: 水轮机机组流量、机组有功、主轴密封磨损量、密封水压力、密封水流量、顶盖压力脉动、顶盖 xyz 向振动、顶盖水位以及 4 台排水泵的启停信号。所有数据采样共 262 万余条, 其中包括模拟量 (传感器数值) 数据共 157 万余条, 其他开关量及统计量 (如机组启停记录和运行时长) 共 105 万余条。使用上述数据对本文模型进行实验验证。

4.2 基准模型

选取以下 4 种经典的时间序列预测模型和本文模型进行对比, 以更好地验证模型的性能。

ARIMA: 差分整合移动平均自回归模型, 一种最常见的基于历史数据进行时间序列预测的统计模型。

LSTM^[10]: 一种 RNN 的变种, 在 RNN 的基础上引入门控机制, 由输入门、遗忘门和输出门组成, 广泛应用于时间序列信息的提取和挖掘。

GRU^[9]: 在 LSTM 上加以改进, 将 LSTM 的遗忘门和输入门合并为一个更新门, 该模型只包含更新门和重置门, 减小了计算量和内存使用。

Transformer^[14]: 利用自注意力机制替换 RNN 网络结构, 实现时序信息的对齐和捕获, 利用层层叠加的自注意力机制更好地捕获多个输入序列中最重要的部分。

4.3 实验设置

本文实验模型使用 PyTorch 框架实现, 实验环

境基于 Ubuntu 20.04 操作系统, 使用了 Inter i7-6700 CPU 和 GTX3090 GPU。在模型训练方面, 运用 Adam 优化器, 学习率设置为 0.000 1; 编码器多头注意力模块的头数为 6; 多头注意力模块和池化模块的组合块数为 2; 解码器多头注意力模块的头数为 6; 训练最大迭代次数为 100; 批次样本数为 64。关于对比基线, 本文在预定义的参数中空间搜索, 选取最优配置的实验结果。具体地, ARIMA 中, 将 p 、 d 和 q 的值分别设为 64、1 和 32; 设定 LSTM 与 GRU 隐状态的维度为 256; Transformer 编码器的层数设为 3; 解码器层数设为 1; 多头注意力模块的头数设定为 8。数据集划分上, 将数据预处理后的初始数据集按年划分, 前 80% 作为训练集和验证集, 共计 42 万条数据; 将后 20% 作为测试集, 共计 10 万条数据。

4.4 评价指标

对于时间序列预测任务, 本文使用通用的评价指标来评估本文模型和其他基准模型的性能。采用均方误差 (MSE) 和平均绝对误差 (MAE) 作为最终的预测效果评价指标。MSE 和 MAE 计算如下:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (6)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (7)$$

4.5 模型性能比较

本文模型和基准模型的性能比较结果如表 1 所示, 考虑到编码器输入长度和解码器预测长度是影响本文模型性能的关键参数, 因此分别尝试了不同的输入长度和预测长度的组合来更加全面地对模型展开分析。

由表 1 和表 2 可知, 当预测长度固定时, 随着输入长度的增加, 所有模型的表现都在逐步提升, 这表明模型在更大的观察视野下通常可以捕获更多的周期等特征信息, 以实现更好地预测。与此同时, 本文发现在不同的输入长度和预测长度的组合情形下, 传统的基于统计学习的模型 ARIMA 表现最差, 采用了门控机制的 LSTM 和 GRU 表现次之, 这表明简单的基于统计学习的模型在短序列预测时表现较好, 但无法应对长序列水位的场景。标准 Transformer 模型性能较好, 仅次于本文所提出的模型, 这是由于本文模型在 Transformer 的基础上引入了池化操作, 使模型可以实现序列对齐, 寻找出更为关键的特征, 更好地应对长序列预测问题。

表 1 不同模型的预测结果比较

预测长度	编码器输入长度	ARIMA		LSTM		GRU		Transformer		本文模型	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
60	60	0.517	0.769	0.416	0.634	0.398	0.612	0.301	0.468	0.304	0.452
60	120	0.436	0.691	0.364	0.584	0.351	0.559	0.241	0.423	0.232	0.398
60	180	0.367	0.524	0.286	0.438	0.273	0.419	0.184	0.284	0.176	0.267
120	120	0.547	0.905	0.477	0.765	0.462	0.726	0.361	0.548	0.354	0.532
120	180	0.494	0.814	0.389	0.617	0.376	0.669	0.261	0.441	0.272	0.438
120	240	0.387	0.618	0.324	0.502	0.311	0.489	0.224	0.314	0.226	0.327

表 2 在长跨度设置下 (预测长度=720) 的结果比较

编码器输入长度	ARIMA		LSTM		GRU		Transformer		本文模型	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
120	0.865	0.812	0.650	0.755	0.654	0.754	0.425	0.633	0.412	0.624
180	0.813	0.790	0.573	0.578	0.564	0.654	0.375	0.568	0.358	0.556
240	0.778	0.767	0.547	0.641	0.545	0.647	0.345	0.524	0.322	0.503

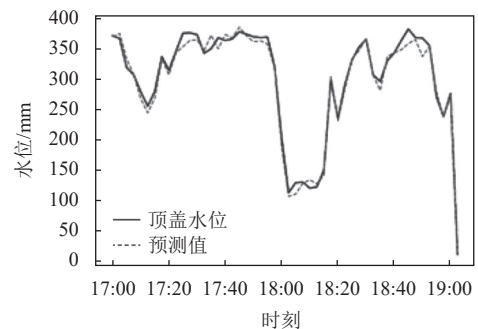
为了更好地呈现本文模型的实际预测性能,选取多组不同时间的顶盖水位变化数据和水轮机相关状态数据,使用历史 2 h 的数据预测后 2 h 的顶盖水位,即 120 步长的水位变化,实验结果如图 2 所示,实线表示真实值,虚线表示预测值。

从图中得到以下结论。1) 顶盖水位变化具有一定的周期性,这是由水轮机和水泵的工作特点和工作方式决定的。顶盖水泵的启停受到顶盖水位的控制,当水位上升到一定阈值,水泵开启,水位快速下降,下降到一定阈值,水泵关闭。本文模型能够捕获这种周期性,并在预测结果上体现了与真实数据相同的上升与下降的态势。2) 当顶盖水位在周期规律中出现波动时,本文模型所包含的多头注意力机制能够分析并捕捉历史观测中对未来水位变化的重要时刻,并据此实现良好的预测效果。多组案例可视化分析说明了,本文模型可以很好地预测顶盖水位的变化,预测的水位趋势及数值能够很好地拟合真实的水位变化情况。这一现象说明了本文模型可以有效地处理极长的输入序列,提取长序列信息并进行长距离的时序预测。

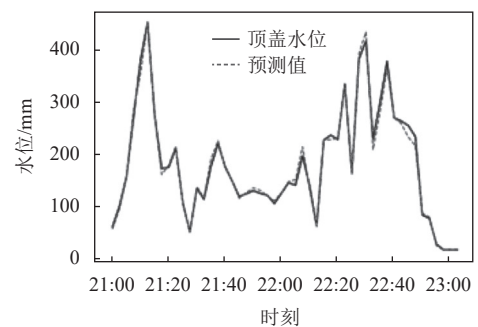
在本文模型的设计和实现中,针对长序列预测问题中预测长度长的特点,进行针对性设计以提高预测速度。图 3 中展示了不同模型预测时间的对比情况。

由图 3 可知,本文模型的预测效率高于其他基准模型,并随着预测长度的增加,速度优势越来越大。说明本文模型使用池化操作替换前馈神经网络

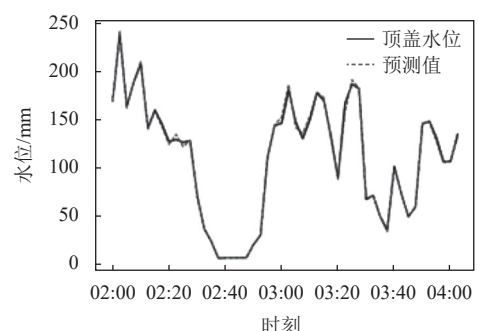
能有效减少了计算量。



a. 2018-06-06



b. 2019-03-24



c. 2019-12-16

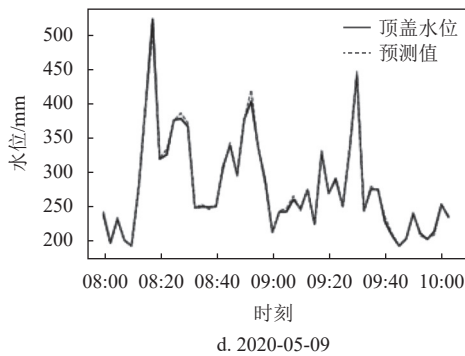


图2 水轮机顶盖水位变化预测

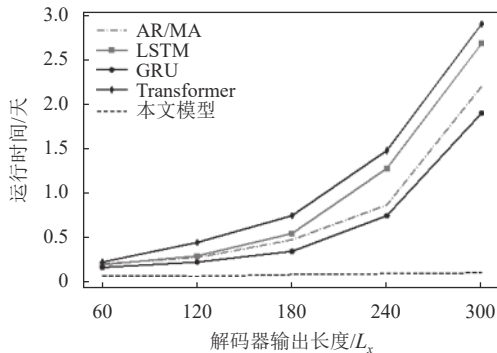


图3 不同模型的预测时间对比

5 结束语

本文针对长序列预测问题进行研究,提出了新模型来实现长序列预测。本文模型通过设计新的嵌入表示方法,保存长时间序列中的完整信息,增加了池化操作以解决原始 Transformer 内存空间占用较大的问题,有效降低了计算量和内存占用。同时,生成式解码器有效处理了传统编码器-解码器模型在预测速度上的局限性。针对水轮机顶盖水位预测这一典型的长序列预测问题,设计了较完善的数据预处理方案和模型实施方法。实验证明本文模型在预测精度和预测速率上取得了明显提升。在未来工作中,拟将本文模型应用在其他水电工业长序列预测问题中,以更好地拓展模型适应性和验证其可靠性。

参 考 文 献

- [1] CHEN P D, LIU S H, SHI C, et al. NeuCast: Seasonal neural forecast of power grid time series[C]//Proceedings of the 27th International Joint Conference on Artificial Intelligence. [S.l.]: ACM, 2018: 3315-3321.
- [2] ZHOU F, LI L, ZHANG K P, et al. Forecasting the evolution of hydropower generation[C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. [S.l.]: ACM, 2020: 2861-2870.
- [3] ZHOU F, LI L. Forecasting reservoir inflow via recurrent neural odes[C]//Proceedings of the 35th AAAI Conference on Artificial Intelligence. [S.l.]: AAAI Press, 2021: 15025-15032.
- [4] 新帅, 郑建民. 大型水轮发电机组水淹水车室预警决策模型研究[J]. 四川水力发电, 2020, 39(3): 137-140.
- JIN S, ZHENG J M. Study on the early warning decision model of flood in waterwheel room of large hydro-generator unit[J]. Sichuan Water Power, 2020, 39(3): 137-140.
- [5] 李冰颖, 耿清华, 王平德. 混流式水轮机顶盖水位上升原因分析及运行对策[J]. 水电与新能源, 2015, 138(12): 66-68.
- LI B Y, GENG Q H, WANG P D. Analysis and treatment of abnormal water level rising above the headcover of francis turbine[J]. Hydropower and New Energy, 2015, 138(12): 66-68.
- [6] BOX G E P, JENKINS G M, REINSEL G C, et al. Time series analysis: Forecasting and control[J]. Journal of Time, 2010, 31(3), DOI: 10.1111/j.1467-9892.2009.00643.x.
- [7] RAY W D. Time series: Theory and methods[J]. Journal of the Royal Statistical Society: Series A (Statistics in Society), 1990, 153(3): 400.
- [8] SEEGER M, RANGAPURAM S, WANG Y, et al. Approximate Bayesian inference in linear state space models for intermittent demand forecasting at scale [EB/OL]. (2017-09-22). <https://arxiv.org/abs/1709.07638>.
- [9] CHUNG J, GULCEHRE C, CHO K H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[EB/OL]. (2014-12-11). <https://arxiv.org/abs/1412.3555>.
- [10] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [11] KITAIEV N, KAISER L, LEVSKAYA A. Reformer: The efficient transformer[EB/OL]. (2020-02-18). <https://arxiv.org/abs/2001.04451v1>.
- [12] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[EB/OL]. (2020-02-18). <https://arxiv.org/abs/2001.04451v1>.
- [13] LUONG T, PHAM H, MANNING C D. Effective approaches to attention-based neural machine translation[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing. [S.l.]: The Association for Computational Linguistics, 2015: 1412-1421.
- [14] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. [S.l.]: ACM, 2017: 6000-6010.
- [15] 严英杰, 盛戈峰, 陈玉峰, 等. 基于时间序列分析的输变电设备状态大数据清洗方法[J]. 电力系统自动化, 2014, 39(7): 138-144.
- YAN Y J, SHENG G H, CHEN Y F, et al. Cleaning method for big data of power transmission and transformation equipment status based on time series analysis[J]. Automation of Electric Power Systems, 2014, 39(7): 138-144.
- [16] LIU F T, TING K M, ZHOU Z H. Isolation forest[C]//Proceedings of the 8th IEEE International Conference on Data Mining. [S.l.]: IEEE Computer Society, 2008: 413-422.
- [17] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.