

文章编号: 1674-5566(2023)03-0669-11

DOI:10.12024/jsou.20220903970

多要素局部-全局特征关联的有效波高预测模型

宋巍¹, 赵勐¹, 贺琪¹, 胡安铎², 张峰³

(1. 上海海洋大学 信息学院, 上海 201306; 2. 上海电力大学 电子与信息工程学院, 上海 201306; 3. 国家海洋局东海信息中心, 上海 200136)

摘要: 有效波高(Significant wave heights, SWH)是描述海浪的重要属性, SWH 的预测对于保障近岸工程设计以及海上作业安全具有重要意义。近年来, 深度学习的方法被用来对 SWH 进行预测, 但是目前存在的方法无法有效捕捉 SWH 的长时间相关性, 同时忽略了海洋多要素之间的局部关联。为此, 提出一种结合海洋多要素局部和全局特征的 SWH 预测模型(Multi-elements local and global correlation for wave height prediction, MLG-SWH)。以有效波高、风速、周期等多要素作为输入, 设计了局部-全局编码(Local-global embedding, LGE)模块, 提取海洋多要素的局部关联以及时间信息。采用编-解码器作为基础网络结构, 提取多要素海浪序列特征。在编-解码器中, 设计了空洞因果卷积自注意力模块来有效捕捉海洋多要素序列的全局长时间相关性, 并在解码器中利用生成推理方式避免单步迭代预测产生的误差累积。选取北大西洋海浪有效波高变化特点不同的两个站点数据进行实验。相较于经典时间序列预测模型以及主流深度学习的方法, 研究使用的 MLG-SWH 模型在 24 和 48 h 预测的均方根误差以及平均绝对误差均为最低, 并在长时序预测方面具有较大的优势。

关键词: 多要素; 长时间相关性; 有效波高预测; 空洞因果卷积自注意力模块

中图分类号: P 731.33 **文献标志码:** A

海浪是近岸海洋工程中的重要设计参数, 海浪条件影响着各种海洋活动, 对于保护近岸地区安全具有重要的意义^[1]。有效波高(Significant wave heights, SWH)是描述海浪的重要属性, 海浪在传播过程中会受到若干因素的影响, 例如风、海洋的深度以及温度等^[2], 这使得 SWH 的预测难度增加。

目前, 海浪运动的预测方法大致可以分为两类: 第一类是利用海浪传播的物理特性, 基于海浪数值模型模拟预测海浪的传播过程, 例如第三代波浪模型(Wave model, WAM)、基于 WAM 的近岸波浪模拟方法(Simulating wave nearshore, SWAN)^[3-5]。数值模型预测通常需要外海入射海浪条件等驱动因子, 且计算效率较低。第二类是基于传统时间序列模型对 SWH 进行预测, 例如自回归模型(Autoregressive, AR)、自回归移动平均模型(Autoregressive moving average, ARMA)以

及自回归综合移动平均模型(Autoregressive integrated moving average, ARIMA)等^[6-8]。AGRAWAL 等^[9]使用 ARIMA 模型在不同预测区间实现在线 SWH 预测。然而, 海浪在传播过程中会受到复杂的海洋要素影响, 而 ARIMA 只使用单要素的时间序列预测, 无法有效表征复杂多样的海洋要素带来的影响, 且 ARIMA 模型极其依赖平稳假设的条件。随着机器学习的发展, 人工神经网络(Artificial neural network, ANN)和支持向量机(Support vector machine, SVM)等开始被应用于 SWH 的预测, DEO 等^[10]提出基于 ANN 对 SWH 进行实时预测, 实验结果表明, ANN 在准确率和一致性方面都有较好的表现。CORNEJO-BUENO 等^[11]和 KUMAR 等^[12]利用前馈神经网络极限学习机(Extreme learning machine, ELM)对 SWH 进行预测, 获得较强泛化能力和较快求解速度。基于完善的数学理论的 SVM 应用十分广泛。

收稿日期: 2022-09-05 修回日期: 2023-01-01

基金项目: 国家重点研发计划(2021YFC3101602); 上海市科委部分地方高校能力建设基金项目(20050501900, 20020500700)

作者简介: 宋巍(1977—), 女, 博士, 教授, 研究方向为海洋大数据分析、计算机视觉等。E-mail: wsong@shou.edu.cn

通信作者: 贺琪, E-mail: qihe@shou.edu.cn

版权所有 ©《上海海洋大学学报》编辑部(CC BY-NC-ND 4.0)

Copyright © Editorial Office of Journal of Shanghai Ocean University(CC BY-NC-ND 4.0)

<http://www.shhydx.com>

MAHJOABI 等^[13]的研究表明,在 SWH 预测中 SVM 某些情况下比 ANN 有更好的性能。黄心裕等^[14]使用 Prophet 模型学习海浪序列的变化特性对海南地区海浪有效波高进行预测;夏天亮等^[15]提出了使用天牛搜素算法来优化 BP 神经网络并应用在南海北部地区海浪预测,提高了预测性能。随着自然语言处理(Natural language processing, NLP)的发展,研究人员发现循环神经网络(Recurrent neural network, RNN)的序列结构非常适用于自然语言处理、气象预报等序列预测任务。长短期记忆网络(Long short-term memory neural networks, LSTM)作为 RNN 的改进,被广泛地应用于预测风速、交通、太阳能和股票价格^[16-18]。李自立等^[19]结合 LSTM 与 ResNet 对北部湾的 SWH 进行短期预测,预测精度得到显著提高,但 LSTM 捕捉长时间相关性方面的能力不足,导致长期预测能力下降。在这种情况下,如何捕捉 SWH 的长时间相关性是提升模型预测性能的关键问题。

近年来,Transformer 作为一个新的体系架构被提出并应用于 NLP 领域^[20],它利用自注意力机制来处理序列数据。基于 RNN 的模型在当前时刻的计算时只依赖前一时刻的计算结果,然而,Transformer 的自注意力机制可以使用序列中任何时刻的值,更有利于捕捉时间序列的长时间相关性。WU 等^[21]在 Transformer 的基础上添加了生成对抗网络对模型的预测值进行修正,降低了模型预测的误差。LI 等^[22]使用因果卷积神经网络生成注意力机制中的查询(Query, Q)与键(Key, K),增强了捕捉局部趋势的能力,但由于其感受有限,因此对于历史信息使用的能力较低。ZHOU 等^[23]采用了卷积层与池化层对注意力矩阵进行操作,并在解码器中使用生成推理的预测方法进行多步预测。SHEN 等^[24]针对目前结合 CNN 与 Transformer 的模型中存在松耦合的问题,提出了 TCCT 模型,该模型在 Transformer 的原有基础上充分利用 CNN 的特性,加强了二者的耦合性。MADHUSUDHANAN 等^[25]提出的 Yformer 结合了 Transformer 与 U-Net 的优点,编码器部分对历史数据以及未来数据分别进行下采样并进行拼接,在解码器部分进行上采样后得到预测结果。

在海浪有效波高预测任务中,既要考虑海浪长期的周期性变化,又要应对各种环境和气象因

素引起的短期变化,因此,传统的 Transformer 应用于海浪有效波高预测任务存在着不足之处。为了解决上述问题,提出结合海洋多要素局部和全局特征的 SWH 预测模型(Multi-elements local and global correlation for wave height prediction, MLG-SWH),实现对 24 和 48 h 的 SWH 高精度预测。MLG-SWH 模型以有效波高及相关海洋多要素时间序列为输入,以编码器-解码器网络结构为基础,通过局部-全局编码(Local-global embedding, LGE)强调海洋多要素特征的时间信息嵌入能力,通过空洞因果卷积扩大自注意力模型的感受野,利用更多历史信息来预测未来。相比于 ARIMA、LSTM 以及改进的 Transformer 等深度学习网络,不仅可以更好地捕捉 SWH 的长时间相关性,而且可以有效地提取 SWH 与海洋多要素间的局部关联。此外,模型采用了生成式预测方式实现 SWH 的一次多步预测,有效降低了逐点迭代式多步预测中误差累积的影响。

1 数据来源及其预处理

1.1 数据来源

实验使用的数据来源于美国国家海洋和大气管理局网站,选取墨西哥湾 42019 号站点(27.910°N, 95.345°W)以及帕姆利科湾 41025 号站点(35.010°N, 75.454°W)作为研究对象且均为海面浮标站点。其中:位于墨西哥湾的 42019 号站点水深 83.5 m,地处热带与亚热带,高温多雨,降水量较多,且每年都有飓风季,因此海浪有效波高波动幅度较大;41025 号站点位于帕姆利科湾,水深 59.4 m,是美国东海岸最大的泻湖,海浪波动幅度相对平稳。两个站点数据采集的时间跨度均为 2018 年 1 月 1 日到 2020 年 12 月 31 日,数据采样间隔均为 1 h,数据集中包括风速(WSPD)、海浪平均周期(APD)、水温(WT)、气温(AT)等要素(表 1)。

表 1 所选站点信息

Tab. 1 Details of selected stations

特征 Variables	41025		42019	
	最大值 Max	最小值 Min	最大值 Max	最小值 Min
有效波高 SWH/m	5.31	0.39	6.88	0.13
风速 WSPD/(m/s)	19.10	0	21.40	0
海浪平均周期 APD/s	8.35	3.28	9.84	2.60
水温 WT/°C	5.20	31.40	32.70	18.50
气温 AT/°C	30.20	-4.40	33.50	3.40

1.2 数据预处理

海浪的形成受多种复杂地理与自然因素影响。从物理机制上,风浪的成长状态与风速、风时、风区三要素密切相关;涌浪一般具有周期长的特点,近岸浪还会受到海底地形和海岸线等的影响。模型基于数据驱动的方法研究海浪有效波高的预测,在现有数据范围内(地形数据不具备)尽可能选择与 SWH 相关的因素。因此,采用统计学的方法对 SWH 与各要素之间的相关性进行辅助验证,在保证显著性水平的情况下,通过计算各要素与 SWH 之间的皮尔逊相关系数(Pearson correlation coefficient),来验证各要素与 SWH 之间的相关性,计算公式如下:

$$\rho_{(A,B_m)} = \frac{\text{cov}(A,B_m)}{\sigma_A \sigma_{B_m}} \quad (1)$$

式中: A 为 SWH 序列; B_m 为第 m 个要素序列; $\text{cov}(A,B_m)$ 为 SWH 与第 m 个要素序列的协方差; σ_A 为 SWH 的方差; σ_{B_m} 为第 m 个要素序列的方差。

选取 42019 号站点数据集中连续 4 000 h 的各要素值进行计算,当皮尔逊相关系数 $|\rho_{(A,B_m)}| > 0.7$ 且显著性差异值 $P < 0.05$ 时,可认为该要素与 SWH 具有显著相关性。经过计算,风速(WSPD)、每小时最大风速(GST)以及海浪平均周期(APD)与 SWH 之间的相关系数分别为 0.726、0.774、0.789;显著性差异值均远小于 0.001,水温(WT)和气温(AT)与 SWH 之间的相关系数仅为 0.343 和 0.376。由于 WSPD 和 GST 之间存在较强相关性,而 APD 与 WSPD 和 GST 相关性较低(相关系数分别为 0.226 和 0.239),因此选择 WSPD 和 APD 与 SWH 一起作为多要素输入。

SWH 的预测为有监督的学习任务,为了生成模型训练所需的训练数据以及测试数据,采用定长滑动窗口的方式对时序数据进行处理,按 7:1:2 比例划分训练集、验证集以及测试集。此外,由于海洋不同要素的取值范围不同,使用 z-score 标准化对数据进行处理,使其服从于标准正态分布。

2 MLG-SWH 模型

图 1 展示了 MLG-SWH 模型,该模型由 LGE 模块、编码器(Encoder)模块以及解码器(Decoder)模块组成。编码器和解码器中的基本单元均由多头空洞因果卷积自注意力层(Multi-

head dconv self-attention)、残差连接及归一化层(Add & Norm)及前向反馈层(Feed forward)组成。LGE 模块用于对海洋多要素序列进行特征编码,编码器模块中的空洞卷积自注意力机制可以捕捉序列的长时间相关性,以及局部变化趋势。解码器模块用于对编码器模块输出的高维特征进行解码,最终连接全连接层得到预测结果。

传统的多步时间序列预测方法常采用迭代式预测即每次迭代只预测一个时间步长,下一次迭代时利用上一次预测的结果,随着预测长度的增加,不仅会出现误差累积的影响,而且会导致模型计算量增加。为此,研究采用生成式预测方法一次性输出海浪有效波高的多步预测结果,定义模型的输入为时间步长为 N 的多维时间序列 $Y = (y_1, y_2, \dots, y_N)$,其中 $y_i = (x_i, T_i)$ 为第 i 个时刻的输入,它由海洋多要素向量 $x_i = [x_i^1, x_i^2, \dots, x_i^a]^T$ 及其对应的时间 T_i 组成, a 为海洋要素的数量。模型预测输出是接下来 n 个时间步长的 SWH 序列,记作 $y = (x_{N+1}^1, x_{N+2}^1, \dots, x_{N+n}^1)$ 。

2.1 局部-全局编码

如图 2 所示,MLG-SWH 模型采用 LGE 模块对时间序列进行编码,使得其具备捕捉海洋多要素之间的局部特征以及有效利用时间信息的能力。

首先,采用一维卷积神经网络对海洋多要素序列 $X = (x_1, x_2, \dots, x_N) \in \mathbb{R}^{a \times N}$ 进行编码,捕捉海洋多要素之间的局部关联特征,编码后的输出序列为 $X \in \mathbb{R}^{a \times d_t}$ 。

由于注意力机制无法识别输入序列的位置顺序,因此,通常需要位置编码(Position embedding)来人为给序列添加位置关系。位置编码只能保留序列中元素的相对位置,且无法利用时间序列的时间信息,但对于时间序列问题,随着时间的变化,序列会呈现周期性变化的趋势,且在极端情况下,也会呈现非周期性的情况。因此,充分利用时间信息对于时间序列预测模型性能的提升有着重要的作用。

为此,MLG-SWH 模型采用 Time2Vec^[26] 将位置编码推广到连续的时间序列上,自适应地捕捉时间序列的周期性与非周期性模式。使用 Time2Vec 对采样时间序列 T 进行建模,定义为 $\xi(T)$,计算公式如下:

$$\xi_j(T) = \begin{cases} w_j T + \varphi_j, & j=0 \\ F(w_j T + \varphi_j), & 1 \leq j \leq d_r - 1 \end{cases} \quad (2)$$

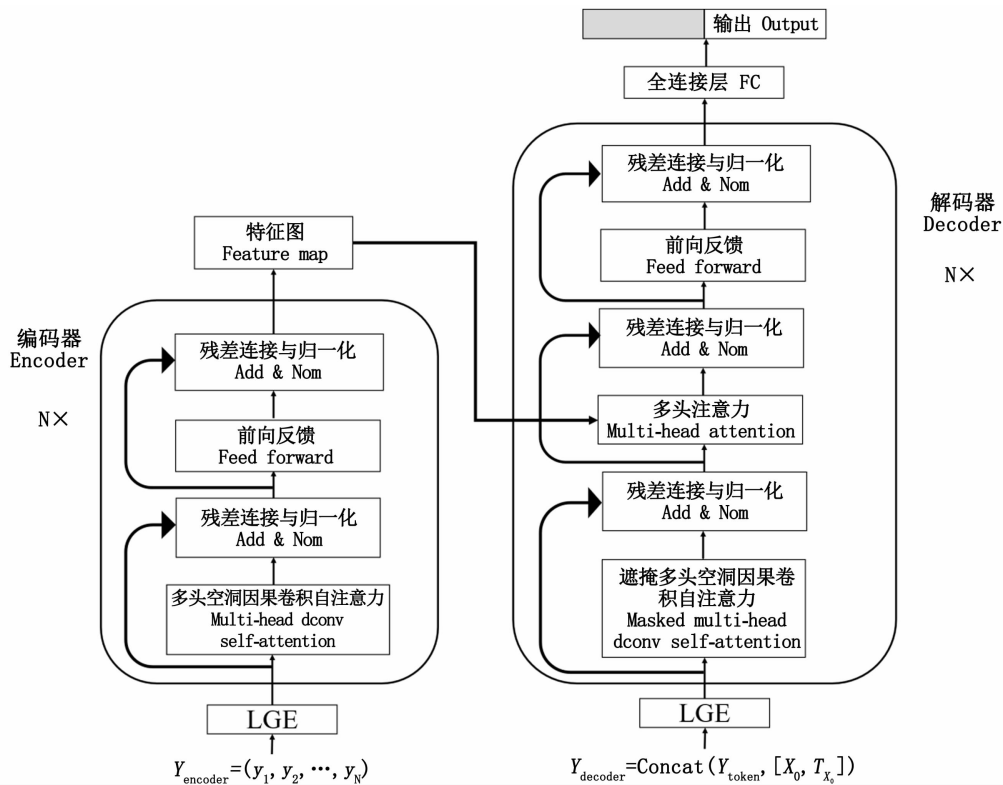


图 1 多要素局部-全局特征关联的有效波高预测模型

Fig. 1 Multi-elements local and global correlation for significant wave height prediction (MLG-SWH) model

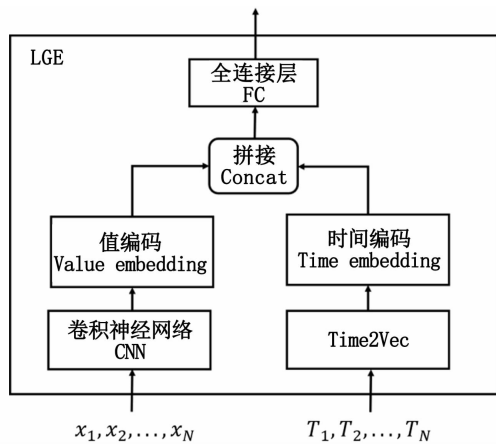


图 2 局部-全局编码模块

Fig. 2 Local-global encoding module

式中: $\xi_j(T)$ 为 $\xi(T)$ 中第 j 个元素; F 为 1 个带有周期性的激活函数, 实验中采用的是正弦函数; w_j 和 φ_j 为正弦函数的频率和偏移, 均为可学习的超参数。Time2Vec 中的正弦函数可用于捕获周期模式, 线性项可用于捕获非周期模式, 它相当于

使用正弦函数作为 1 个全连接层将采样时间序列 T 映射到了 d_r 维的空间中。

Transformer 和 LogTrans 均采用直接相加的方式对数据编码和位置编码进行处理, 但这样做不可避免地会导致信息损失^[20,22]。因此, LGE 模块将海洋要素序列的局部特征编码和时间编码进行拼接, 然后将其输入到全连接层中得到最终的编码结果 $D \in R^{d_m}$ 。

2.2 多头自注意力机制

自注意力机制起源于人类对视觉方面的研究, 是人类视觉所特有的大脑信号处理机制。人类视觉通过浏览全局的图像, 获得需要重点关注的目标位置, 然后对这一位置投入更多的注意力资源来获得更多所需要关注目标的细节信息。注意力机制很适用于时间序列模型, 其优点在于它可以很好地捕捉长时间序列的相关性。多头注意力机制 (Multi-head attention, MHA) 中不同的头通过学习可以捕捉序列中的不同时间模式。多头注意力机制的计算公式如下:

$$O_h = S \frac{(Q_h K_h^T)}{\sqrt{d_k}} V_h \quad (3)$$

式中: S 为 Sigmoid 函数; $Q_h = DW_h^Q$ 为查询矩阵; $K_h^T = DW_h^K$ 为键矩阵以及 $V_h = DW_h^V$ 为值矩阵; $h = \{1, 2, \dots, H\}$ 是头索引, 且 $W_h^Q, W_h^K \in R^{d_m \times d_k}, W_h^V \in R^{d_m \times d_v}$ 均是可以学习的参数。

经过这些线性变换之后, 通过缩放点积计算向量的注意力分数。再对 O_1, O_2, \dots, O_H 进行拼接操作后再次进行线性投影。在注意力分数输出时, 堆叠了两层全连接层以及一个 ReLU 激活函数层。

2.3 编码器模块

编码器模块由 4 组编码层堆叠而成。每组编码层分为 2 个子层: 多头空洞因果卷积自注意力层以及前向反馈层, 每个子层后均连接一个残差连接与归一化层。

传统自注意力机制计算查询和键之间的相关性是逐点计算的[图 3(a)], 因此, 自注意力模块无法学习到时间序列的局部信息。卷积神经网络通常被使用在图像领域, 用于捕捉图像的局部特征, 一些学者将卷积神经网络应用在时间序列预测任务中, 在捕捉局部特征方面取得了良好的效果^[27-29]。LI 等^[22]提出了因果卷积自注意力机制将卷积神经网络与自注意力机制相结合, 即用因果卷积神经网络代替全连接层来生成注意力机制所需的 Q 和 K [图 3(b)], 增强了模型捕

捉局部特性的能力。然而, 因果卷积神经网络只能回顾线性规模的历史数据, 由于感受野有限, 导致该模型利用历史信息的能力不足。虽然因果卷积神经网络可以通过不断叠加层数来增加感受野的大小, 但会导致参数量增大。

为了解决上述问题, 设计了空洞因果卷积神经网络来生成 Q 与 K [图 3(c)], 相较于因果卷积神经网络 [图 3(b)], 其感受野会随着卷积层数的增加以指数的形式增加。第 1 个空洞因果卷积层中的卷积核会对 $(2^{l-1} - 1)$ 位置上的元素进行卷积操作, 随着层数的增加, 感受野也会随之增加, 可以使得模型捕捉更长时间序列的信息, 充分利用历史数据的变化趋势。模型使用的空洞因果卷积维持了因果关系的性质: 时间序列在某时刻的值只与之前时刻的值进行卷积, 确保网络在构建和训练过程中不违反波高预测中时间顺序上的基本前后依赖关系; 同时, 空洞卷积可以更好地利用序列的历史信息, 在对某个时间步的数据计算相似度时, 可以利用该时间步的上下文关系 (例如局部趋势) 进行计算, 有助于提高预测的准确度。

在经过多头空洞因果卷积自注意力层后, 将输出结果与 LGE 模块的编码结果 D 输入到残差连接与归一化层, 之后通过前馈层对特征维度进行变化, 并再次输入到残差连接与归一化层得到最终的特征图 $F \in R^{d_m}$ 。

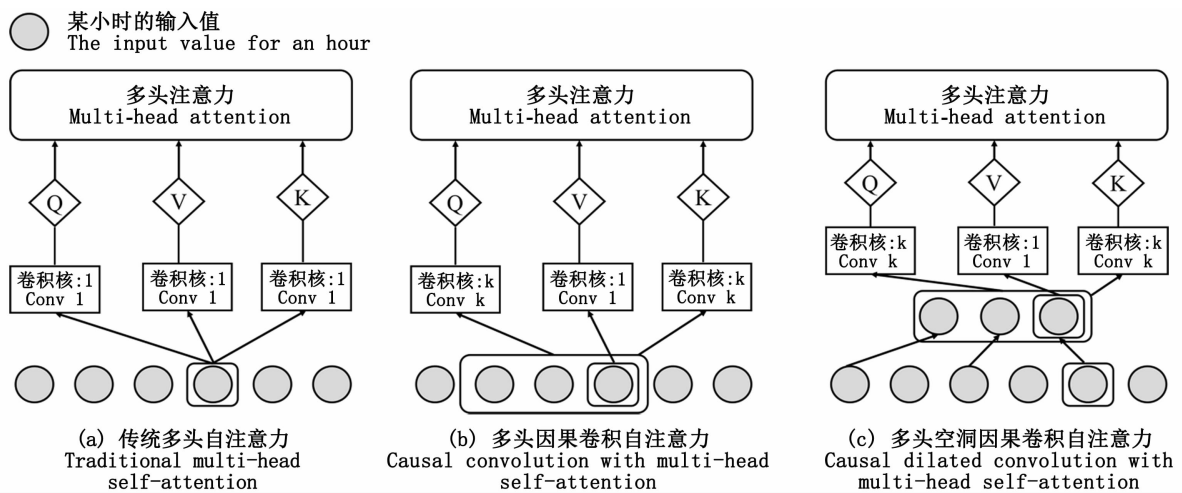


图 3 不同注意力之间的比较

Fig. 3 Comparison between different attention mechanisms

2.4 解码器模块

解码器模块采用与编码器模块类似的结构。在解码器的多头空洞因果卷积自注意力中添加的掩码矩阵 $M \in \mathbb{R}^{d_m \times d_m}$, 其所有上三角元素均设为 $-\infty$, 用来掩蔽 $Q_h K_h^T$ 矩阵中包含未来信息的部分, 当矩阵 M 通过 Softmax 函数计算后, 所有上三角元素的值均变为 0, 使得未来信息不参与计算注意力分数, 避免自回归。

传统多步预测方法采用迭代式预测, 会导致模型计算量增加, 且随着预测长度的增加会出现误差累计的情况。因此, 模型借鉴 Informer 的思想采用生成式预测方法, 截取编码器模块的部分输入作为起始序列段, 并与 1 个长度与目标预测序列相同的零矩阵进行拼接后添加时间信息作为解码器的输入^[23]。解码后, 取映射向量中后半部分(目标预测长度)作为预测结果, 该模型仅解码 1 次就可以得到多步预测结果, 解码器模块如图 4 所示。

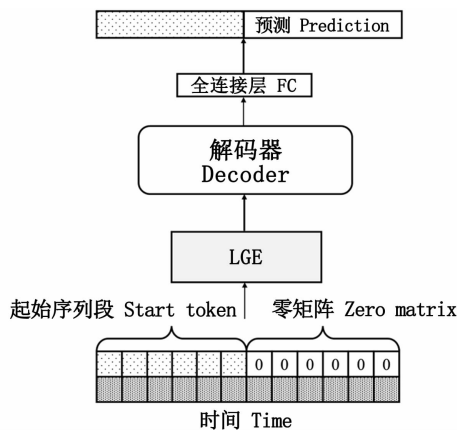


图 4 解码器模块

Fig. 4 Decoder module

MLG-SWH 模型采用编码器模块输入多要素序列的后半段作为起始序列段, 并与长度和目标序列相同的零矩阵进行拼接, 计算公式如下:

$$Y_{\text{decoder}} = \text{Concat}(Y_{\text{token}}, [X_0, T_{X_0}]) \quad (4)$$

式中: Y_{token} 为起始序列段, 在 24、48 h 的预测任务中, Y_{token} 采取的长度分别为 6、12; X_0 为长度和目标预测序列相同的零矩阵; T_{X_0} 为目标预测序列的时间。

将拼接后的结果输入到 LGE 模块中进行编码, 并将编码后的结果 D 输入到解码器模块进行解码。在解码时, 首先经过多头因果卷积自注意力层对 D 进行特征学习, 并添加残差连接与归一

化层得到矩阵 $Z \in \mathbb{R}^{d_m}$, 其次使用多头注意力计算编码器模块输出的高维特征与矩阵 Z 之间的注意力矩阵并输入到前向反馈层。最后, 在解码器模块后连接全连接层, 将编码器模块的高维特征在解码器模块进行解码后的结果进行线性映射。映射向量中后半部分(目标预测长度)即为预测结果。

3 实验

3.1 实验环境与参数设置

模型在搭建时, 采用了 Pytorch 深度学习框架, 并部署在 Tesla-P100 服务器上运行。采用平均绝对误差作为损失函数, 使用了 Adam 优化器, 学习率设置为 0.001, 数据编码维度 d_m 设置为 512, 时间编码维度 d_r 设置为 64, 多头注意力机制的头数 H 设置为 8。在模型训练时, 利用早停机机制避免模型迭代次数过多而出现过拟合。

3.2 实验评价指标

实验采用均方误差 (Mean square error, E_{MS}) 和平均绝对误差 (Mean absolute error, E_{MA}) 作为评价指标, 计算公式如下:

$$E_{MS} = \frac{1}{n} \sum_{i=1}^n (H_p^i - H_t^i)^2 \quad (5)$$

$$E_{MA} = \frac{1}{n} \sum_{i=1}^n |H_p^i - H_t^i| \quad (6)$$

式中: H_p^i 为 i 时刻 SWH 的预测值; H_t^i 为 SWH 的真实值; n 为预测的长度。模型预测越精准, 评价指标 E_{MS} 、 E_{MA} 的值越小。

3.3 实验结果与分析

为了体现 MLG-SWH 模型的预测效果, 将 MLG-SWH 模型与 ARIMA、LSTM、LogTrans 以及 Informer 进行比较。对于上述方法, 均使用 168 h 的历史数据预测未来 24 和 48 h 的 SWH。其中, ARIMA 是传统时间序列预测模型; LSTM 是 SWH 预测中主流的深度学习算法, 它在循环神经网络的基础上添加了遗忘门、输入门、输出门以及可以有效避免梯度消失以及梯度爆炸带来的影响; LogTrans 是在 Transformer 的基础上进行改进, 采用了因果卷积自注意力机制增强模型捕捉局部变化趋势的能力; Informer 在注意力机制后使用卷积层与池化层代替了残差连接与归一化层以及全连接层, 并采用了生成式预测方法。表 2 展示了每种方法的预测结果。

由表 2 可以看出, MLG-SWH 模型在 41025

与 42019 两个站点的 24 和 48 h 预测任务中均方差以及平均绝对误差均为最低。ARIMA 以及 LSTM 在 24 和 48 h 的预测误差较大,利用注意力机制的模型预测性能均明显优于 ARIMA 以及 LSTM。由于 42019 号站点所处地理位置具有高温多雨的特点,且每年都有飓风季,因此该站点的海浪变化更加复杂。除此之外,42019 号站点与 41025 号站点水深相差 20 多米,受到海底地形的影响更大。因此,42019 号站点的预测难度会

更大,预测误差略高于 41025 号站点。MLG-SWH 与 Informer 的预测结果较为接近,但在预测效率上,MLG-SWH 要高于 Informer。以 41025 号站点 48 h 预测为例,Informer 所耗的时间为 64 s,而 MLG-SWH 模型所耗的时间约为 43 s,节约 30% 以上。MLG-SWH 模型不仅提升了利用历史信息的能力,而且可以自适应地学习时间信息中的周期性与非周期性,模型预测性能相较于 LogTrans 以及 Informer 均有提升。

表 2 不同预测方法的性能对比

Tab.2 Performance comparison of different prediction methods

站点 Station	小时 Hour	ARIMA		LSTM		LogTrans		Informer		MLG-SWH	
		E_{MA}/m	E_{MS}/m	E_{MA}/m	E_{MS}/m	E_{MA}/m	E_{MS}/m	E_{MA}/m	E_{MS}/m	E_{MA}/m	E_{MS}/m
41025	24	0.482	0.463	0.469	0.429	0.385	0.338	0.369	0.318	0.341	0.287
	48	0.531	0.503	0.522	0.461	0.447	0.416	0.422	0.391	0.395	0.362
42019	24	0.533	0.476	0.478	0.447	0.412	0.391	0.395	0.364	0.382	0.343
	48	0.585	0.540	0.541	0.503	0.462	0.442	0.451	0.429	0.439	0.418

注:加粗表示最好的结果。
Notes: Bold means the best result.

为了更直观地体现模型在长期预测任务中的性能,通过在 41025 号站点的测试集数据中随机抽取了 168 h 的历史数据,输入到对比模型中

性能较好的 LogTrans、Informer 中进行预测,分别得到未来 48 h 的预测值。图 5 展示 3 个模型的 SWH 预测值与 SWH 真实值的对比。

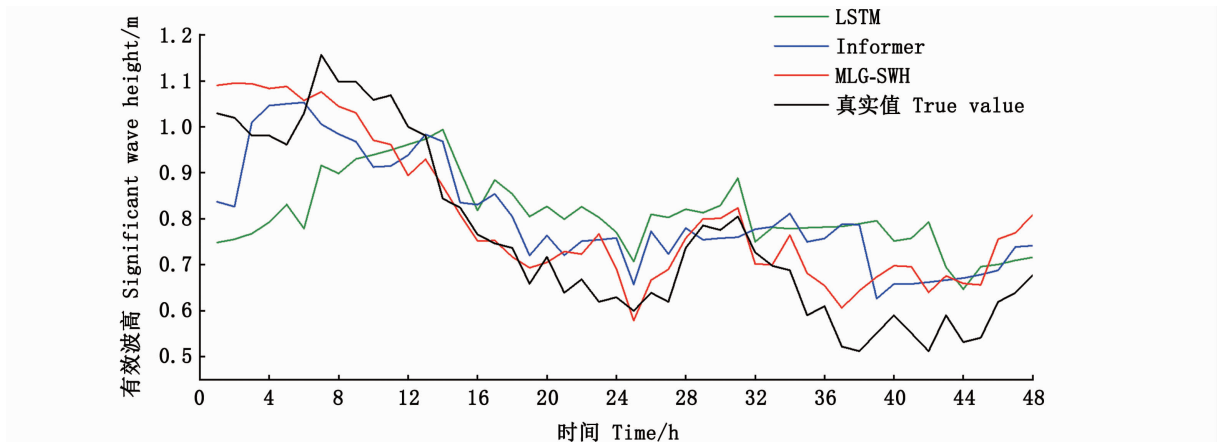


图 5 MLG-SWH 与 Informer、LSTM 对 41025 号站点 48 h 预测结果

Fig.5 48 hours prediction results of station 41025 by MLG-SWH, Informer, LSTM

由图 5 可以看出,相较于 MLG-SWH,Informer 和 LSTM 模型的预测值明显偏离真实值,这是因为 LSTM 的长时间序列建模能力不足,而 Informer 中卷积核大小固定,导致感受野有限,从而降低了模型捕捉序列变化趋势的能力。除此之外,在预测较长时间序列时,模型预测误差会随时间增加而增加。

为进一步显示 MLG-SWH 模型应用于长时间预测任务的效果,在 41025 号站点的测试集数据中抽取了 1 个月(720 h)的预测结果进行展示。为避免线条堆叠影响显示效果,图 6 展示了 3 个模型的 SWH 预测值与真实值的对比。

由图 6 可以看出,在预测较长时间序列时,模型预测误差会随时间增加而增加,Informer 和

LSTM 均存在不同程度的滞后性,原因是时间序列预测是 1 个回归问题,且序列本身存在着自相关性,导致模型在训练时为了使得损失函数最小,倾向于使用前面步长的真实值作为预测值,

而 MLG-SWH 模型的整体预测趋势与真实值更加接近,模型中的空洞因果卷积自注意力机制可以更好地捕捉海洋多要素序列局部变化趋势,使得模型对于局部趋势的预测更加准确。

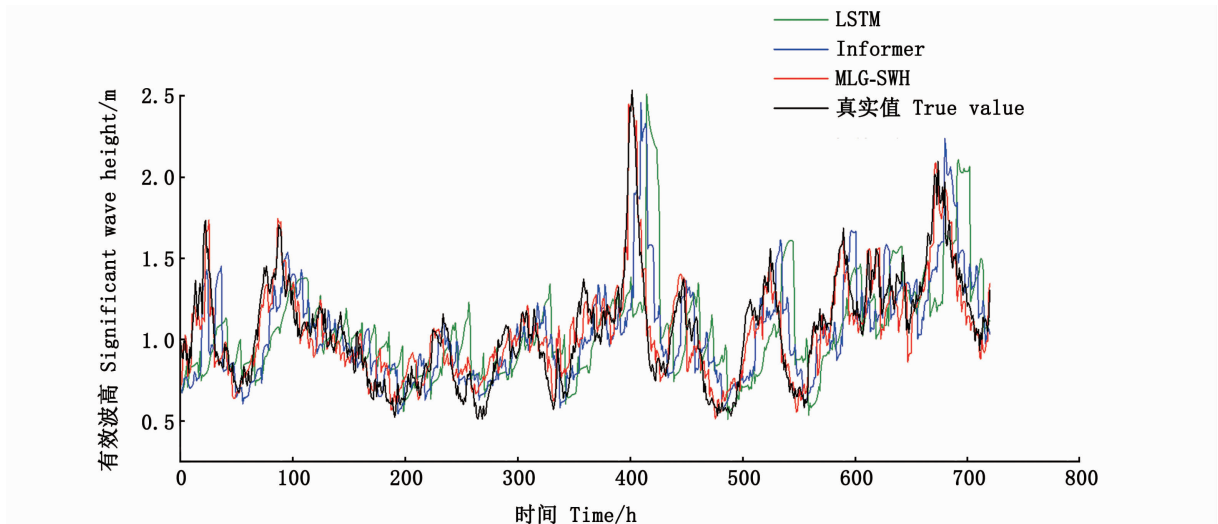


图 6 MLG-SWH 与 Informer、LSTM 对 41025 号站点 1 个月预测结果

Fig. 6 One month prediction results of station 41025 by MLG-SWH, Informer, LSTM

为进一步体现 MLG-SWH 模型的长时序预测性能,对比了 MLG-SWH 的 48 h 模型(MLG-SWH-48)和 24 h 模型(MLG-SWH-24)两个月的预测结果,其中 MLG-SWH-24 的平均绝对误差为 13.3%,MLG-SWH-48 为 17.2%,均符合海浪有效波高预测的业务化需求。图 7 展示了 MLG-SWH-24 分两次预测未来 48 h 与 MLG-SWH-48 一次性预测未来 48 h 的结果。可以看出,对未来

48 小时预测时,MLG-SWH-24 预测结果与真实值拟合程度更好,但在极端波高情况下,模型拟合程度较差。复杂的天气条件是极端波高出现的 1 个重要原因,且出现频率较低。因此,模型训练所需的信息以及训练样本不足会导致预测结果出现较大的偏差。在实际应用中,可以根据海浪预报的不同要求对模型进行选择。

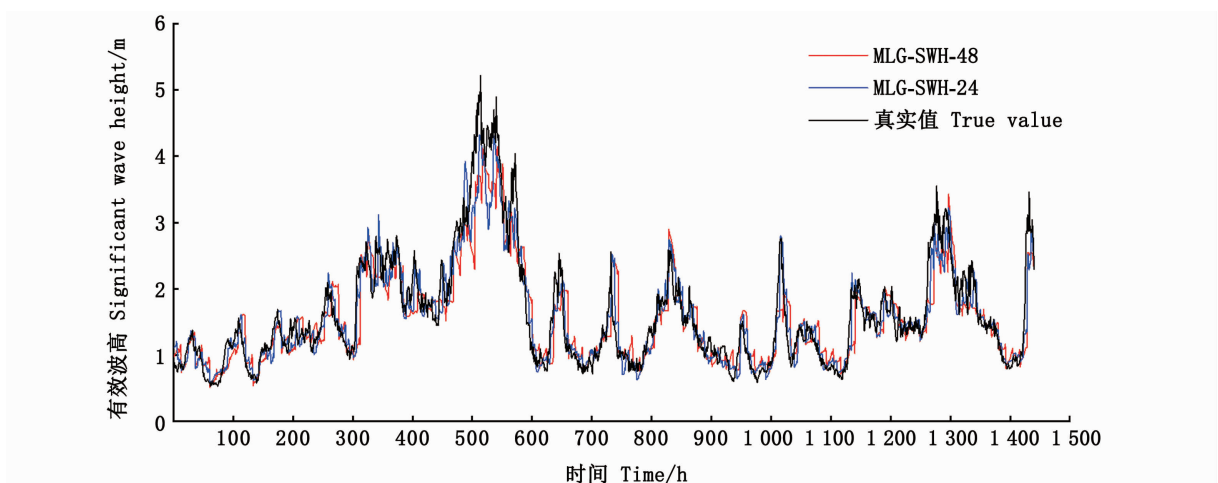


图 7 MLG-SWH-24、MLG-SWH-48 与真实值的比较

Fig. 7 Comparison of MLG-SWH-24, MLG-SWH-48 with the true value

3.4 消融实验

为了探究各个模块对模型性能带来的影响,对 LGE 模块、空洞因果卷积自注意力机制以及生成式预测方法等 3 部分进行了消融实验。

3.4.1 局部-全局编码模块的影响

为了研究 LGE 编码对于模型性能的影响,对未使用 LGE 的 MLG-SWH 使用全连接层对数据进行编码并添加位置编码作为模型的输入。41025 号站点的实验结果如表 3 所示。

表 3 Local-Global Embedding 的消融实验结果

Tab.3 Ablation experiment results of Local-Global Embedding

方法 Methods	24 h		48 h	
	E_{MA}/m	E_{MS}/m	E_{MA}/m	E_{MS}/m
MLG-SWH w/o LGE	0.366	0.314	0.420	0.387
MLG-SWH	0.341	0.287	0.395	0.362

注:加粗表示最好的结果。

Notes: Bold means the best result.

由表 3 可以看出,添加 LGE 模块对于模型预测能力有明显提升,在 41025 号站点的 48 h 任务中,未使用 LGE 模块与使用 LGE 模型的均方差分别为 0.387 与 0.362,表明 LGE 模块中的 Time2Vec 对时间进行编码后可以使得模型自适应的学习序列的时间信息,卷积神经网络可以提取 SWH 与海洋多要素之间的局部关联,使得模型可以充分利用多要素海浪序列的潜在信息,提高预测能力。

3.4.2 空洞因果卷积自注意力机制的影响

随着卷积核大小的增加,感受野会以指数级上升,使得模型可以捕捉更长时间的信息,从而提高模型捕捉局部变化趋势的能力。设计实验对不同卷积核大小对模型的影响进行讨论,其编码器中层数均为 4 层,解码器中层数为 1 层。当 $k=1$ 时,空洞因果卷积自注意力模块等同于传统自注意力模块。在 41025 号站点对未来 24 h 进行预测的结果如表 4 所示。

表 4 不同大小卷积核对于模型的影响

Tab.4 Influence of different size convolution kernels on the model

指标 Index	卷积核大小 Kernel size			
	1	2	3	6
E_{MA}/m	0.368	0.359	0.341	0.350
E_{MS}/m	0.314	0.310	0.287	0.299

注:加粗表示最好的结果。

Notes: Bold means the best result.

从表 4 可以看出,使用空洞因果卷积自注意力机制的模型性能普遍高于使用传统注意力机制的模型,其中卷积核大小为 3 时,模型性能达到最佳。然而,随着卷积核继续增大,模型的性能没有显著变化,这表明当卷积核大小为 3、卷积层数为 4 时,模型对于输入的序列中的局部信息已经可以充分提取,虽然更大的卷积核会使得模型的感觉野更大,但同时也会使模型训练的参数量增大,不利于模型深度的增加,同时影响模型性能。

3.4.3 生成式多步预测方法与迭代式多步预测方法的对比

在多步预测任务中,随着预测长度的增加,迭代式多步预测方法不仅会出现误差累积的情况,导致预测误差增加,而且会增加模型的计算量。因此,采用一步式生成预测结果的方法来降低误差累积的影响。实验结果如表 5 所示。

表 5 生成式多步预测方法与迭代式多步预测方法对比

Tab.5 Comparison between generative multi-step prediction method and iterative multi-step prediction method

站点 Station	小时 Hour	生成式预测方法 Generative prediction method		迭代式预测方法 Iterative prediction method	
		E_{MA}/m	E_{MS}/m	E_{MA}/m	E_{MS}/m
41025	24	0.341	0.287	0.375	0.325
	48	0.395	0.362	0.431	0.403
42019	24	0.382	0.343	0.401	0.357
	48	0.439	0.418	0.458	0.436

注:加粗表示最好的结果。

Notes: Bold means the best result.

由表 5 可以看出,生成式预测方法在两个站点预测未来 24、48 h 的误差均有所降低。迭代式预测方法进行多步预测时,需要进行迭代解码,因此迭代式预测方法的时间复杂度较高。生成式预测方法不仅有效降低了误差累计的影响,预测误差得到降低,而且在多步预测任务中仅解码 1 次就可得到预测结果,预测时间大幅度降低。

4 结论

为了解决现有模型无法有效捕捉海浪有效波高长时间性以及海洋多要素局部关联的问题,提出了 MLG-SWH 模型用于海浪有效波高预测,模型中的 LGE 模块可以有效地捕捉海浪有效波高与海洋多要素之间的局部关系以及充分利用

时间信息,多头空洞因果卷积自注意力通过增加感受野可以使模型更好地捕获长时间相关性以及局部变化趋势;模型使用的生成式预测方法有效地避免了长时间序列预测中误差累计的影响。通过使用 MLG-SWH 模型在北大西洋 41025 号站点以及墨西哥湾 42019 号站点进行实验,模型在预测未来 24、48 h 的任务中均取得最佳效果。然而模型还存在着不足之处:(1)相较于 LSTM 等基于循环神经网络的模型,MLG-SWH 模型的计算复杂度略高;(2)在现实情况中,有效波高与附近海域的海浪相关,在位置上具有空间相关性,未来研究可以引入附近站点数据进行协同预测;(3)引入更多的海洋要素对有效波高进行预测。同时,未来的研究可以更进一步地将海浪生成的原理与机器学习相结合,更加完善地实现有效波高预测任务。

参考文献:

- [1] JAIN P, DEO M C. Neural networks in ocean engineering[J]. *Ships and Offshore Structures*, 2006, 1(1): 25-35.
- [2] WANG H, FU D Y, LIAO S, et al. Significant wave height prediction based on MSFD neural network[C]//Proceedings of the 2019 Tenth International Conference on Intelligent Control and Information Processing. Marrakesh: IEEE, 2019: 1-5.
- [3] The Wamdi Group. The WAM Model—A third generation ocean wave prediction model [J]. *Journal of Physical Oceanography*, 1988, 18(12): 1775-1810.
- [4] BOOIJ N, RIS R C, HOLTHUIJSEN L H. A third-generation wave model for coastal regions: 1. Model description and validation[J]. *Journal of Geophysical Research*, 1999, 104(C4): 7649-7666.
- [5] TOLMAN H L. User manual and system documentation of WAVEWATCH III TM version 3. 14 [R]. Maryland: U. S. Department of Commerce, National Oceanic and Atmospheric Administration, National Weather Service, National Centers for Environmental Prediction, 2009: 220.
- [6] BOLLERSLEV T. Generalized autoregressive conditional heteroskedasticity[J]. *Journal of Econometrics*, 1986, 31(3): 307-327.
- [7] SAID S E, DICKEY D A. Testing for unit roots in autoregressive-moving average models of unknown order[J]. *Biometrika*, 1984, 71(3): 599-607.
- [8] BOX G E P, PIERCE D A. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models [J]. *Journal of the American Statistical Association*, 1970, 65(332): 1509-1526.
- [9] AGRAWAL J D, DEO M C. On-line wave prediction[J]. *Marine Structures*, 2002, 15(1): 57-74.
- [10] DEO M C, JHA A, CHAPHEKAR A S, et al. Neural networks for wave forecasting[J]. *Ocean Engineering*, 2001, 28(7): 889-898.
- [11] CORNEJO-BUENO L, NIETO-BORGE J C, GARCÍA-DÍAZ P, et al. Significant wave height and energy flux prediction for marine energy applications: a grouping genetic algorithm-Extreme Learning Machine approach [J]. *Renewable Energy*, 2016, 97: 380-389.
- [12] KUMAR N K, SAVITHA R, AL MAMUN A. Ocean wave height prediction using ensemble of extreme learning machine [J]. *Neurocomputing*, 2018, 277: 12-20.
- [13] MAHJOOBI J, ADELI MOSABBEH E. Prediction of significant wave height using regressive support vector machines[J]. *Ocean Engineering*, 2009, 36(5): 339-347.
- [14] 黄心裕,唐军,王晓宇. 基于 Prophet 算法的海南近海波浪长时段时序分析与预测[J]. *海洋学报*, 2022, 44(4): 114-121.
HUANG X Y, TANG J, WANG X Y. Long term time series analysis and prediction of waves at Hainan offshore zone based on Prophet algorithm[J]. *Haiyang Xuebao*, 2022, 44(4): 114-121.
- [15] 夏天亮,李醒飞,杨少波. 基于 BAS-BP 模型的南海北部波高预测[J]. *海洋湖沼通报*, 2021, 43(5): 9-16.
XIA T L, LI X F, YANG S B. Prediction of wave height based on BAS-BP model in the northern part of the South China Sea[J]. *Transactions of Oceanology and Limnology*, 2021, 43(5): 9-16.
- [16] ZHAO Z, CHEN W H, WU X M, et al. LSTM network: a deep learning approach for short-term traffic forecast[J]. *IET Intelligent Transport Systems*, 2017, 11(2): 68-75.
- [17] GENSLER A, HENZE J, SICK B, et al. Deep Learning for solar power forecasting—An approach using AutoEncoder and LSTM Neural Networks [C]//Proceedings of 2016 IEEE International Conference on Systems, Man, and Cybernetics. Budapest: IEEE, 2016: 2858-2865.
- [18] FAN S T, XIAO N H, DONG S. A novel model to predict significant wave height based on long short-term memory network[J]. *Ocean Engineering*, 2020, 205: 107298.
- [19] 李自立,蒙素素. 基于 LSTM-ResNet 模型的定点有效波高预测[J]. *海洋预报*, 2022, 39(2): 80-85.
LI Z L, MENG S S. Prediction of significant wave height of a fixed point based on LSTM-ResNet model [J]. *Marine Forecasts*, 2022, 39(2): 80-85.
- [20] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017: 6000-6010.
- [21] WU S F, XIAO X, DING Q G, et al. Adversarial sparse transformer for time series forecasting [C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, BC, Canada: Curran

- Associates Inc. , 2020: 1435.
- [22] LI S Y, JIN X Y, XUAN Y, et al. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting[C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver;Curran Associates Inc. , 2019: 471.
- [23] ZHOU H Y, ZHANG S H, PENG J Q, et al. Informer: beyond efficient transformer for long sequence time-series forecasting [C]//Proceedings of the 35th AAAI Conference on Artificial Intelligence. AAAI, 2020.
- [24] SHEN L, WANG Y Z. TCCT: tightly-coupled convolutional transformer on time series forecasting[J]. *Neurocomputing*, 2022, 480: 131-145.
- [25] MADHUSUDHANAN K, BURCHERT J, DUONG-TRUNG N, et al. Yformer: U-Net inspired transformer architecture for far horizon time series forecasting [J]. *arXiv*: 2110. 08255, 2021.
- [26] KAZEMI S M, GOEL R, EGHBALI S, et al. Time2Vec: learning a vector representation of time [J]. *arXiv*: 1907. 05321, 2019.
- [27] BAI S J, KOLTER J Z, KOLTUN V. Convolutional sequence modeling revisited[C]//Proceedings of the 6th International Conference on Learning Representations. Vancouver: OpenReview. net, 2018.
- [28] VAN DEN OORD A, DIELEMAN S, ZEN H G, et al. WaveNet: a generative model for raw audio[C]//Proceedings of the 9th ISCA Speech Synthesis Workshop. Sunnyvale: ISCA,2016: 125.
- [29] STOLLER D, TIAN M, EWERT S, et al. Seq-U-Net: a one-dimensional causal U-Net for efficient sequence modelling [C]//Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence. Yokohama: IJCAI.org, 2020: 2893-2900.

A prediction model of significant wave height based on local and global correlation of multi-elements

SONG Wei¹, ZHAO Meng¹, HE Qi¹, HU Anduo², ZHANG Feng³

(1. College of Information, Shanghai Ocean University, Shanghai 201306, China; 2. College of Electronics and Information, Shanghai University of Electric Power, Shanghai 201306, China; 3. East China Sea Information Center, State Oceanic Administration, Shanghai 200136, China)

Abstract: Significant Wave Heights (SWH) is an important attribute to describe ocean waves, and SWH prediction is of great significance for ensuring the design of offshore engineering and the safety of offshore operations. In recent years, deep learning methods have been used to predict SWH, but the existing methods cannot effectively capture the long-term correlation of SWH, thus ignoring the local associations between multiple elements of the ocean. To this end, this paper proposes a SWH prediction model (Multi-elements Local and Global Correlation for Wave height Prediction, MLG-SWH) that combines local and global features of marine multi-elements. First, using multiple factors such as significant wave height, wind speed and period as input, a Local-Global Embedding (LGE) module is designed to embed local correlation and time information of ocean multi-elements. Then, an encoder-decoder structure is used to extract the features of ocean wave height, where a casual dilated convolution self-attention module is designed to effectively capture the global long-term correlation of ocean multi-element sequences and the generative prediction method in the decoder is adopted to avoid errors accumulated in the single-step iterative prediction. Finally, the data of two stations with different characteristics of SWH variation in the North Atlantic are selected for experimental evaluations. Compared with classical time-series forecasting models and mainstream deep learning methods, the MLG-SWH model achieves the lowest mean square error and mean absolute error in 24 and 48 hours SWH forecasting, having a greater advantage in long-term time series prediction.

Key words: multi-element; long-term correlation; significant wave height prediction; casual dilated convolution self-attention module