

【电子与信息科学 / Electronics and Information Science】

基于双模型并联的复杂时序预测方法

郑洪英, 夏林中, 刘星

深圳信息职业技术学院中德机器人学院, 广东深圳 518172

摘要: 传统时序预测模型通常仅关注捕捉复杂时序中的趋势和模式, 而忽略了变量间的相互作用, 限制了该模型在复杂时序预测中的应用. 提出一种Dualformer双模型并联方案, 该模型并联iTransformer (inverted transformer)和PatchTST(patch time series transformer), 通过激活函数替代前馈神经网络, 并通过多层感知机计算输出结果. Dualformer利用注意力机制同时捕捉复杂时序中的时间维度和变量维度信息, 关注时间趋势与多变量交互. 实验结果显示, Dualformer在复杂时序预测效果上显著优于对比模型iTransformer、PatchTST和DLinear(decomposition linear), 在实际应用中可显著提高复杂时序预测的准确度, 具有广泛应用前景.

关键词: 人工智能; 深度学习; 复杂时序预测; 注意力机制; 多层感知机; Dualformer模型

中图分类号: TP183

文献标志码: A

DOI: 10.3724/SP.J.1249.2025.03334

Complex time series forecasting method based on dual-model parallelism

ZHENG Hongying, XIA Linzhong, and LIU Xing

Sino-German Robotics School, Shenzhen Institute of Information Technology, Shenzhen 518172, Guangdong Province, P. R. China

Abstract: Traditional forecasting models typically focus on capturing trends and patterns in time series data while often neglecting interactions among variables, which limits their effectiveness in complex time series scenarios. To address this, a dual-model parallel approach named Dualformer is proposed, which integrates the inverted transformer (iTransformer) and patch time series transformer (PatchTST) models in a parallel architecture. In this framework, the conventional feedforward neural network components are replaced with activation functions, and a multilayer perceptron is introduced to further process the outputs of the two parallel sub-models. The Dualformer model employs attention mechanisms to simultaneously capture both the temporal and variable dimensions in the complex time series, effectively capturing time-dependent patterns and multivariate interactions. Experimental results demonstrate that the Dualformer significantly outperforms the benchmark models, including iTransformer, PatchTST, and decomposition linear (DLinear) in complex time series forecasting tasks. These results highlight Dualformer's superior accuracy and its strong potential for real-world applications in forecasting complex time series data.

Key words: artificial intelligence; deep learning; complex time series forecasting; attention mechanism; multilayer perceptron; Dualformer model

复杂时序预测通过分析数据规律来预测未来发展趋势, 为行业决策提供重要指导^[1-3], 可广泛应

用于天气预报、金融交易及能源管理等多个领域. 随着时序数据量的不断增加及变量维度的提升, 复

Received: 2024-08-09; Accepted: 2024-11-05; Online (CNKI): 2025-01-03

Foundation: National Natural Science Foundation of China (62303327); Natural Science Foundation of Guangdong Province (KJ2023B001)

Corresponding author: Professor XIA Linzhong (xialz@szit.edu.cn)

Citation: ZHENG Hongying, XIA Linzhong, LIU Xing. Complex time series forecasting method based on dual-model parallelism [J]. Journal of Shenzhen University Science and Engineering, 2025, 42(3): 334-341. (in Chinese)

Open access: This article is licensed under the CC BY 4.0 License (<https://creativecommons.org/licenses/by/4.0/>).

<http://journal.szu.edu.cn>



杂时序的预测方法也在不断迭代. 传统方法主要依赖于领域专家知识构建的参数模型^[4], 但这些方法时常受到一定限制.

深度学习提供了一种纯粹由数据驱动的学习模型, 在自然语言处理中已获得巨大成功^[5-6]. 由于自然语言处理与复杂时序数据分析之间存在共通性, 深度学习技术在复杂时序分析领域也得到广泛应用, 循环神经网络(recurrent neural network, RNN)^[7]和卷积神经网络(convolutional neural network, CNN)^[8]等早期预测模型, 虽然对平稳数据有效, 但由于RNN基于马尔可夫假设进行建模, CNN通过时间卷积网络(temporal convolutional network, TCN)等技术提取时间维度变化信息, 它们在捕捉序列数据的长期依赖关系方面能力有限.

Transformer模型通过注意力机制和堆叠结构^[9], 能够从深层特征中捕捉时间相关性, 成为序列建模任务的理想选择, 并已在自然语言处理^[10]、计算机视觉^[11]、语音^[12]及复杂时序预测^[13]等多个领域取得显著成功. 当前的研究主要集中在对注意力算子和模型架构的改进方面.

针对注意力算子的研究主要进行了两类探索: 一类专注于时序依赖关系, 另一类同时考虑时序和变量依赖性. 在第1类研究中, 为降低Transformer的时间复杂度, ZHOU等^[14]在Informer中引入概率稀疏表示(probabilistic sparse representation, Prob-Sparse)的注意力机制, 利用注意力得分的稀疏性提高预测的准确性和效率; DU等^[15]通过数据段替代数据元素, 进一步降低了Preformer的复杂度; WU等^[16-17]分别在Autoformer和FEDformer(frequency enhanced decomposed transformer)中采用快速傅里叶变换(fast Fourier transform, FFT), 将注意力机制迁移至频域, 在保证高效计算性能的同时, 精细捕捉时序依赖性. 然而, 该研究路径面临来自线性模型挑战^[18-19], 对应模型具有较高的预测准确度和较低的时间复杂度.

第2类探索则在注意力机制中融合了时间维度和变量维度的依赖性. 通道独立(channel independent, CI)方法将多变量时序分解为独立的单变量序列, 虽简化了模型, 但忽略了变量间的相互作用, 典型方法如PatchTST^[20](patch time series transformer). 而通道依赖(channel dependent, CD)方法通过并行或串行使用时序和变量注意力模块, 捕捉跨时序和变量的依赖关系, 如Crossformer^[21]和

DSformer^[22](double sampling transformer). LIU等^[23]在iTransformer中通过调整输入数据的维度顺序, 转换注意力机制的焦点, 以提取时间维度上的非线性信息. 相比于CD, CI的数据拟合效果比较差, 但预测结果稳定性更好, 考虑到非平稳数据更依赖结果的稳定性, 采用CI会取得更好的预测效果^[24].

也有研究在架构层面进行改进^[24-25], 使Transformer模型适用于复杂时序建模, 如WU等^[16]通过减少Autoformer模型解码器的层数, 强化编码器的多层网络结构, 使模型更适应复杂时序数据的特点; CHEN等^[26-27]提出基于Pathformer的多尺度建模方法, 通过自适应路径和周期性解耦框架, 进一步增强模型对时间变化的捕捉能力.

通过多个编码器并行处理同一个输入, 并对编码器输出数据进行聚合, 可以有效提高模型的预测能力^[27-28]. ASHOK等^[29]根据不同时间周期对输入数据进行分解, 每个分量由独立的编码器处理, 最后进行聚合输出, 也已取得较好效果.

本研究基于PatchTST和iTransformer模型进行架构拓展, 提出PatchTST和iTransformer并联的架构方案, 称为Dualformer模型. 实验结果表明, 本模型在大多数情况下表现出了比单一维度优化模型更高的预测准确度, 基本解决了优化的瓶颈问题.

1 Dualformer模型

本研究模型主要作用于多变量的复杂时序预测任务, 假设输入时序数据 $\mathbf{X}$ 为

$$\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T\} \in \mathbf{R}^{T \times N} \quad (1)$$

其中, $\mathbf{x}_t$ 为由 t 时刻的所有变量组成的张量, T 为回溯窗口大小, $t = 1, 2, \dots, T$; N 为变量维度. 实际输入的时间序列数据 $\mathbf{X}$ 是3维张量, 第1维度通常是时序数据的批量大小. 由于批量主要用于计算加速, 不影响实际算法流程, 故本模型中省略了批量维度.

1.1 Dualformer架构

Dualformer模型架构如图1, 在时间和变量维度上以不同嵌入方式处理数据. 其中, 在时间维度上采用滑动窗口的形式将一段连续时间点的1个变量作为token, 方便学习时间变换的趋势信息; 在变量维度上采用1个变量的整段时间作为token, 方便学习变量间交互信息, 再分别通过Transformer中最核心的注意力模块分别提取两个维度的信息. 这

里去掉了 Transformer 架构中常用的前馈全连接层而改用简单的激活函数, 原因在于本模型已经通过并行维度的注意力模块提取双维度信息, 每一边缺失的 token 间信息表达由并行的另一注意力模块补充, 这也是本模型与其他单一维度优化模型的较大区别. 最后, 通过线性层进行信息的聚合, 综合两者贡献生成最终结果.

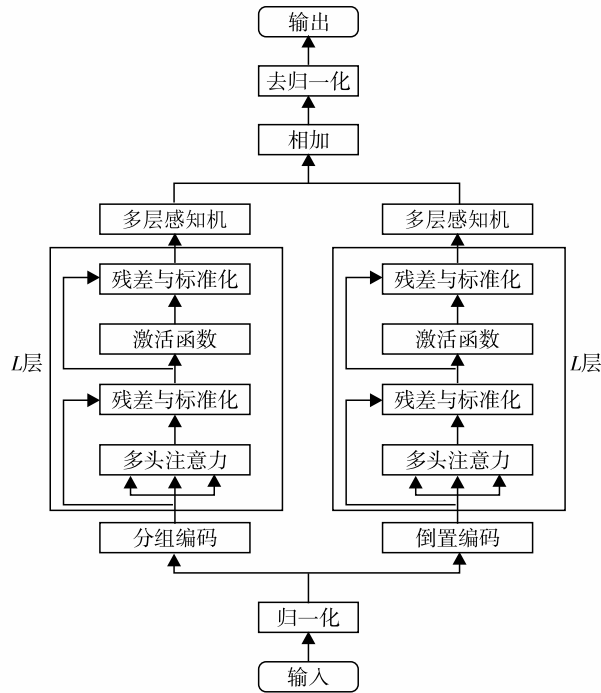


图 1 Dualformer 架构

Fig. 1 Diagram of Dualformer.

Dualformer 模型的流程与原理如下: ① 对输入数据进行归一化操作, 以消除复杂时序数据中的尺度差异性, 增加输入数据的分布稳定性, 提高模型的泛化能力. ② 对数据进行两种不同的嵌入处理: 一种是关注时间趋势的嵌入, 通过滑动窗口方式划分 Patch, 将注意力放在同一变量的多个 Patch 区间之间, 以使其更容易找到数据随时间的变换规律; 另一种是关注多变量交互的嵌入, 通过反转数据的最后两个维度, 将注意力放在多个变量 token 之间, 以使其更容易找到多变量间的影响规律. 嵌入结束后会分别进入 L 层的 Encoder 层, 与传统 Transformer 模型的 Encoder 层不同的是, 本研究采用 1 个简单的激活函数来代替原有的前馈全连接层 (feed forward network, FFN), 这样既保存了非线性语义, 也省去了计算开销, 且原有的前馈全连接层信息在并行的另一边 Encoder 的注意力层中计算, 也保证

了计算的准确度. ③ 两边分别通过 1 个多层感知器 (multilayer perceptron, MLP) 层转换到目标输出, 并进行加法运算, 最终得到输出数据.

1.2 归一化

归一化与去归一化模块的作用是为了解决序列的非平稳问题, 并使输入数据的分布范围处于激活函数作用期间, 改善学习效果. 本研究采用 Max-Min 归一化, 即

$$x'_t = \frac{x_t - x_{\min}}{x_{\max} - x_{\min}} \quad (2)$$

其中, x_t 是多变量 $\mathbf{x}_t$ 中的单维变量; $x_{\min}$ 和 $x_{\max}$ 分别代表 T 内的最小值和最大值; x'_t 是 x_t 归一化后的值. 在模型输出前进行去归一化, 为

$$y_t = y'_t \times (x_{\max} - x_{\min}) + x_{\min} \quad (3)$$

其中, y'_t 是 x'_t 经模型计算后输出的单维变量; y_t 是 y'_t 去归一化后的单维变量.

1.3 编码与嵌入

Dualformer 模型同时采用倒置编码和分组编码两种形式. 其中, 倒置编码只是将张量中的变量维度和时间维度对调, 即将 $\mathbf{X} \in \mathbf{R}^{T \times N}$ 倒置为 $\mathbf{X}^a \in \mathbf{R}^{N \times T}$; 分组编码则在倒置编码基础上, 对时间维进一步切分成若干块. 设 P_{num} 为切块数量, 每个切块具有相同长度 P_{len} , 若块与块之间没有重叠, 则 $T = P_{\text{num}} \times P_{\text{len}}$, 切块后的时间系列数据为 $\mathbf{X}^b \in \mathbf{R}^{N \times P_{\text{num}} \times P_{\text{len}}}$.

编码后还需要进行嵌入, 嵌入的作用是将最后 1 维数据映射到更大的空间, 从而有助于学习到更多有用的规律. 本研究用 Embedding 代表嵌入算子, 对于倒置编码, 其映射过程可表示为 $\mathbf{R}^T \mapsto \mathbf{R}^D$, 其中, D 是嵌入后的数据维度, 即嵌入后的变量数; 对于分组编码, 其映射过程可表示为 $\mathbf{R}^{P_{\text{len}}} \mapsto \mathbf{R}^D$. 该映射过程可由 MLP 算子实现, 倒置编码嵌入可表示为

$$\mathbf{H}^a = \text{Embedding}^a(\mathbf{X}^a) \quad (5)$$

其中, $\mathbf{H}^a \in \mathbf{R}^{N \times D}$. 对于分组编码采用数值嵌入加位置嵌入的方案, 即

$$\mathbf{H}^b = \text{Embedding}^b(\mathbf{X}^b) + \boldsymbol{\Psi} \quad (6)$$

$$\boldsymbol{\Psi}(\Phi, 2i) = \sin\left(\frac{\Phi}{10\,000^{\frac{2i}{D}}}\right) \quad (7)$$

$$\boldsymbol{\Psi}(\Phi, 2i + 1) = \cos\left(\frac{\Phi}{10\,000^{\frac{2i}{D}}}\right) \quad (8)$$

其中, $\mathbf{H}^b \in \mathbf{R}^{P_{\text{num}} \times D}$; $\boldsymbol{\Psi}$ 表示位置嵌入向量; $i =$

$0, \dots, D-1; \Phi = 0, 1, \dots, P_{\max} - 1; \Psi(\Phi, 2i)$ 和 $\Psi(\Phi, 2i+1)$ 分别表示在 $2i$ 和 $2i+1$ 索引的编码值.

1.4 计算层

Dualformer 由多层计算层串联而成, 计算层由函数 Computing() 表示, 该函数由多头自注意力机制、残差网络、标准化及激活函数构成.

多头自注意力机制为

$$H^{a:l+1} = \text{Computing}^a(H^{a:l}) \quad (9)$$

$$H^{b:l+1} = \text{Computing}^b(H^{b:l}) \quad (10)$$

$$H_{\text{out}} = \text{Multiheads}(Z_1, Z_2, \dots, Z_M) \quad (11)$$

$$Z_m = \text{Softmax}\left(\frac{Q_m K_m}{\sqrt{d_k}}\right) V_m \quad (12)$$

其中, $l = 0, 1, \dots, L-1$; L 为计算层的总层数; $m = 1, 2, \dots, M$, M 为自注意力的头数; $Q_m = H_{\text{in}} W_m^Q$; $K_m = H_{\text{in}} W_m^K$; $V_m = H_{\text{in}} W_m^V$; 初始值为随机赋值的权重矩阵 $W_m^Q, W_m^K \in \mathbf{R}^{D \times d_k}$, $W_m^V \in \mathbf{R}^{D \times d_v}$; $d_k = D/M$.

残差网络为

$$H = H_{\text{out}} + H_{\text{in}} \quad (13)$$

层归一化为

$$\text{LayerNorm}(H) = \left\{ \frac{h_n - \text{Mean}(h_n)}{\sqrt{\text{Var}(h_n)}}, n = 1, 2, \dots, N \right\} \quad (14)$$

其中, $h_n \in \mathbf{R}^{1 \times D}$; Mean() 为均值函数; Var() 为方差函数. Dualformer 模型的激活函数采用 GeLU().

2 实验

在多个复杂时序预测公开数据集对 Dualformer 模型进行评测, 并与 iTransformer、PatchTST 和 DLinear^[18] 模型的预测结果进行比较.

2.1 实验准备

实验选取的数据集覆盖了不同领域和类型的复杂时序数据, 能够更全面评估模型的性能和泛化能力^[14, 23]. 首先, 选取涵盖电力消费数据的 ECL (electricity) 数据集. 电力消费是具有明显季节性和周期性变化的复杂时序, 有助于测试模型对周期性变化的捕捉能力. 其次, 选择包含电力变压器操作数据的 ETTh2 (electricity transformer temperature hour 2) 数据集. 与 ECL 数据集不同, ETTh2 数据集中的复杂时序数据更具有变化的不确定性, 能够更

好地用来评估模型对于突发事件和不确定性因素的应对能力. 第3个是包含城市交通流量复杂时序数据的 Traffic 数据集. 交通流量数据具有明显的周期性和随机性, 有助于测试模型的稳定性和泛化能力. 第4个是包含多个城市天气情况复杂时序数据的 Weather 数据集. 天气数据受到气候和季节等因素影响, 有助于评估模型对于外部因素变化的适应能力. 第5个是包含不同货币汇率变化数据的 Exchange 数据集. 汇率数据受到宏观经济和国际政治等多方面因素的影响, 有助于评估模型对于复杂多变的外部因素的响应能力. 通过选取这些不同领域和类型的数据集, 可以更全面评估模型在不同情境下的表现, 从而准确了解本模型的优势和局限性.

在训练参数的选取方面, 本研究参考时序预测开源库 Tslib (time series library)^[30], 从中选取多个模型通用的训练参数, 这些参数已经过调优, 能较好地体现模型的预测效果.

2.2 预测结果

基于上述配置, 本研究测试4个模型分别在预测长度为96、192、336和720时的表现, 并通过平均绝对误差 (mean absolute error, MAE) 和均方误差 (mean squared error, MSE) 对结果进行评价. MAE 与 MSE 的比值越小, 表示预测准确度越高.

主要综合预测结果见表1. 其中, 加粗数据代表最好的预测结果; 下划线数据为次好预测结果. 可见, 在大部分数据集和预测长度上, Dualformer 均优于对比模型 PatchTST、iTransformer 和 DLinear. 由于 Dualformer 模型将前馈全连接层模块移除, 在保证模型正确性与效果的同时节约了计算时间, 一定程度降低了算法复杂程度, 能够兼顾预测效果和训练速度.

2.3 结果分析

从以上结果可见, Dualformer 模型在除 Traffic 数据集外都优于 iTransformer、PatchTST 及 DLinear 模型. 表2为 Dualformer 与 iTransformer 模型的平均 MSE 比值和平均 MAE 比值对比. 其中, 随着通道数 (变量数) 的增加, Dualformer 与 iTransformer 模型的平均 MSE 和 MAE 比值也随之增加. 比值 < 1 表示 Dualformer 优于 iTransformer; 比值 > 1 表示 iTransformer 优于 Dualformer. 可见, 在变量数较少的情况下, Dualformer 具有较大优势, 而在变量数较多时的结果正好相反. 结合模型的架构和特点, 本研

表1 四种模型在不同数据集上的预测结果

Table 1 The prediction results of the four models for the datasets

数据集	数据规模	Dualformer		PatchTST		iTransformer		DLinear	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ECL	96	<u>0.148</u>	<u>0.241</u>	0.196	0.287	0.147	0.239	0.210	0.301
	192	0.162	0.255	0.201	0.290	<u>0.166</u>	<u>0.257</u>	0.210	0.304
	336	0.176	0.271	0.217	0.306	<u>0.180</u>	<u>0.273</u>	0.222	0.319
	720	<u>0.216</u>	<u>0.306</u>	0.257	0.337	0.215	0.302	0.257	0.349
	A_{vg}	0.176	<u>0.268</u>	0.218	0.305	<u>0.177</u>	0.268	0.225	0.318
ETTh2	96	0.293	0.343	0.305	<u>0.349</u>	<u>0.302</u>	0.351	0.340	0.394
	192	0.373	0.392	0.388	0.401	<u>0.381</u>	<u>0.399</u>	0.482	0.479
	336	0.415	0.427	0.425	0.432	<u>0.423</u>	<u>0.433</u>	0.586	0.539
	720	0.421	0.439	0.431	0.446	<u>0.427</u>	<u>0.446</u>	0.838	0.660
	A_{vg}	0.375	0.400	0.387	0.407	<u>0.383</u>	<u>0.407</u>	0.562	0.518
Traffic	96	<u>0.407</u>	<u>0.272</u>	0.545	0.360	0.393	0.268	0.699	0.429
	192	<u>0.424</u>	<u>0.280</u>	0.540	0.354	0.413	0.277	0.647	0.407
	336	<u>0.436</u>	<u>0.285</u>	0.552	0.359	0.425	0.283	0.653	0.410
	720	<u>0.472</u>	<u>0.305</u>	0.586	0.377	0.459	0.300	0.694	0.429
	A_{vg}	<u>0.435</u>	<u>0.285</u>	0.556	0.363	0.422	0.282	0.673	0.419
Weather	96	0.171	0.211	<u>0.177</u>	0.218	0.178	<u>0.217</u>	0.196	0.255
	192	0.222	0.257	<u>0.225</u>	0.259	<u>0.225</u>	<u>0.257</u>	0.237	0.295
	336	0.280	0.299	0.283	0.302	<u>0.282</u>	<u>0.299</u>	0.282	0.331
	720	<u>0.355</u>	0.349	<u>0.356</u>	0.352	0.359	<u>0.349</u>	0.345	0.382
	A_{vg}	0.257	0.279	<u>0.260</u>	0.283	0.261	<u>0.281</u>	0.265	0.316
Exchange	96	0.084	0.204	<u>0.085</u>	<u>0.206</u>	0.086	0.206	0.098	0.233
	192	0.171	0.296	<u>0.173</u>	<u>0.299</u>	0.178	0.302	0.185	0.325
	336	0.312	0.404	0.329	0.417	0.324	<u>0.413</u>	<u>0.320</u>	0.431
	720	0.760	0.658	0.851	0.696	0.805	0.678	<u>0.769</u>	<u>0.664</u>
	A_{vg}	0.332	0.390	0.360	0.405	0.348	<u>0.400</u>	<u>0.343</u>	0.413

注: A_{vg} 表示所有预测长度结果的平均值;粗体表示最好结果;下划线表示次好结果。

究认为以上现象的原因在于,当变量数较少时,Dualformer模型能够综合考虑时间趋势信息和多变量交互信息,全面捕捉时序数据中的立体规律,因此,相对于以变量相关性为主的iTransformer和以时间相关性为主的PatchTST模型,具有较大优势。然而,在变量数大幅增加后,多变量之间的交互关系变得更加复杂且具有更深层次的关联,此时,关注多变量交互将变得更为重要且更容易发现规律,因此iTransformer更具优势。

Exchange数据集一般属于非平稳数据集,非平稳数据同时具有趋势性和季节性^[31],不同时间步长

表2 Dualformer与iTransformer模型的平均MSE比值和平均MAE比值

Table 2 Average MSE and MAE ratios between Dualformer and iTransformer models

数据集	通道数	平均MSE比值	平均MAE比值
ETTh2	7	0.979	0.983
Exchange	8	0.952	0.977
Weather	21	0.987	0.994
ECL	321	0.990	1.037
Traffic	862	1.031	1.011

上产生变化的协变量模式, 给预测带来了重大挑战. 本模型在 Exchange 数据集上也取得了优异结果, 表明 Dualformer 模型在金融预测领域具有较大应用优势.

为了更直观展示模型左右两个分支不同的特征捕获能力, 基于 Exchange 数据集, 在推理阶段左右两个多层感知机输出位置各自获取 1 组同一时刻数据, 通过热力图方式呈现其数据特征, 结果见图 2. 其中, 横坐标表示数据的 64 个时间步; 纵坐标表示数据的 8 个变量. 由图 2(a)可见, 大部分竖线跨越多个变量, 表明 iTransformer 模型能更多地捕获变量间的相关性. 由图 2(b)可见, 有 8 个界限清晰的横向带子, 且同一个带子在横坐标方向呈现较高的相似性, 这说明 PatchTST 模型能更多地捕获时

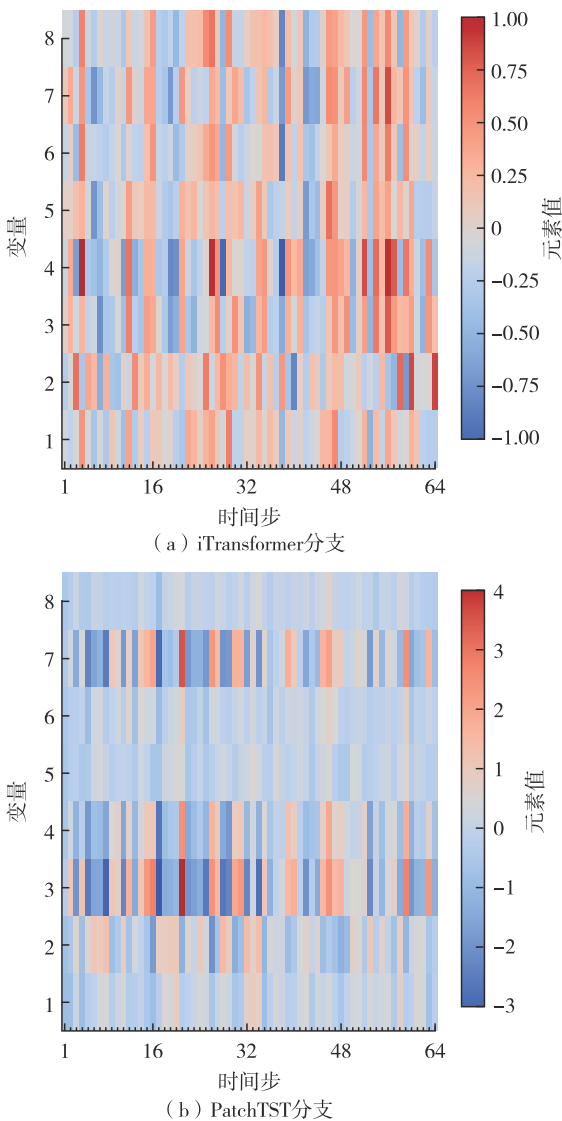


图2 本模型的输出数组数据热力图

Fig. 2 The output array data heatmap of this model.

间步之间的相关性. Dualformer 融合了 iTransformer 和 PatchTST 模型, 能起到较好的互补作用.

2.4 消融实验

为了验证采用激活函数代替 FFN 的合理性, 本研究设计消融实验对比 Dualformer(GeLU)和 Dualformer(FFN)模型的表现, 实验结果见表 3, 其中, 加粗数字表示效果最好. 可见, 使用 GeLU 激活函数的 Dualformer 模型在测试数据集和预测长度上总体优于使用 FFN 模块的 Dualformer 模型. 虽然原始 FFN 模块在提取垂直维度信息方面起到一定作用, 但相比只保留 GeLU 激活函数, FFN 模块仅增加了简单的线性变换. 在单视角模型中, 其在一定程度上能够弥补注意力机制缺失的另一维度信息. 然而, 在 Dualformer 模型中, 左右两边的注意力机制已涵盖复杂时序数据变量维度和时间维度的相关性计算, 因此, 保留线性变换操作反而会引入噪声,

表3 不同数据集下 FFN 模块存在与否的模型预测结果对比

Table 3 Model prediction results with and without FFN module on different datasets					
数据集	数据规模	Dualformer		Dualformer+FFN	
		MSE	MAE	MSE	MAE
ETTh2	96	0.293	0.343	0.296	0.345
	192	0.373	0.392	0.375	0.394
	336	0.415	0.427	0.418	0.429
	720	0.421	0.439	0.429	0.446
	A_{vg}	0.375	0.400	0.380	0.403
Exchange	96	0.084	0.204	0.091	0.212
	192	0.171	0.296	0.176	0.299
	336	0.312	0.404	0.316	0.409
	720	0.760	0.658	0.851	0.695
	A_{vg}	0.332	0.390	0.358	0.404
Weather	96	0.171	0.211	0.173	0.214
	192	0.222	0.257	0.225	0.258
	336	0.280	0.299	0.282	0.301
	720	0.355	0.349	0.357	0.351
	A_{vg}	0.257	0.279	0.259	0.281
Traffic	96	0.407	0.272	0.401	0.274
	192	0.424	0.280	0.426	0.285
	336	0.436	0.285	0.433	0.285
	720	0.472	0.305	0.465	0.306
	A_{vg}	0.435	0.285	0.426	0.283

注: A_{vg} 表示所有预测长度结果的平均值; 粗体表示最好结果.

导致预测效果变差。

结 语

为提高时序数据预测的准确度, 在 iTransformer 和 PatchTST 基础上, 提出并联的 Dualformer 模型, 其能较好地融合两个模型的优点. 该模型并联的两个注意力算子, 分别从时间维度和变量维度提取了复杂时序的相关性, 从而能够更全面地获取数据中的非线性特征. 此外, 在对并联结果相加之前, 各自增加 1 个 MLP 算子的计算, 有利于获取时间维度的全局相关性. 在 5 个公开数据集上的测试结果表明, 采用 Dualformer 模型的预测准确度总体明显优于 iTransformer 和 PatchTST 两个对比模型. 本文所提的 Dualformer 模型可应用于各领域的复杂时序数据的预测.

基金项目: 国家自然科学基金资助项目(62303327); 广东省自然科学基金资助项目(KJ2023B001)

作者简介: 郑洪英(hyzheng@126.com), 深圳信息职业技术学院副教授、博士. 研究方向: 人工智能、脑机接口及智能机器人.

引 文: 郑洪英, 夏林中, 刘星. 基于双模型并联的复杂时序预测方法[J]. 深圳大学学报理工版, 2025, 42(3): 334-341.

参考文献 / References:

- [1] LAM R, SANCHEZ-GONZALEZ A, WILLSON M, et al. Learning skillful medium-range global weather forecasting [J]. Science, 2023, 382(6677): 1416-1421.
- [2] 张建晋, 王韞博, 龙明盛, 等. 面向季节性时空数据的预测式循环网络及其在城市计算中的应用[J]. 计算机学报, 2020, 43(2): 286-302.
ZHANG Jianjin, WANG Yunbo, LONG Mingsheng, et al. Predictive recurrent networks for seasonal spatiotemporal data with applications to urban computing [J]. Chinese Journal of Computers, 2020, 43(2): 286-302. (in Chinese)
- [3] XU Zichao, ZHENG Hongying, CHEN Jianyong. Opemod: an optimal performance selection model for prediction of non-stationary financial time series [C]// Artificial Neural Networks and Machine Learning - ICANN 2022. Cham: Springer International Publishing, 2022: 304-315.
- [4] HARVEY A C. Forecasting, structural time series models and the Kalman filter [M]. Cambridge, UK: Cambridge University Press, 1990.
- [5] DEVLIN J, CHANG Mingwei, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [C]// Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, USA: ACL, 2019: 4171-4186.
- [6] 夏林中, 叶剑锋, 罗德安, 等. 基于 BERT-BiLSTM 模型的短文本自动评分系统[J]. 深圳大学学报理工版, 2022, 39(3): 349-354.
XIA Linzhong, YE Jianfeng, LUO Dean, et al. Short text automatic scoring system based on BERT-BiLSTM model [J]. Journal of Shenzhen University Science and Engineering, 2022, 39(3): 349-354. (in Chinese)
- [7] SALINAS D, FLUNKERT V, GASTHAUS J, et al. DeepAR: probabilistic forecasting with autoregressive recurrent networks [J]. International Journal of Forecasting, 2020, 36(3): 1181-1191.
- [8] LIU Minhao, ZENG Ailing, CHEN Muxi, et al. SCINet: time series modeling and forecasting with sample convolution and interaction [C]// Advances in Neural Information Processing Systems. Red Hook, USA: Curran Associates Inc., 2022: 5816-5828.
- [9] VASWANI A, SHAZEER N, PARMA N, et al. Attention is all you need [C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, USA: Curran Associates Inc., 2017: 6000-6010.
- [10] KALYAN K S, RAJASEKHARAN A, SANGEETHA S. AMMUS : a survey of transformer-based pretrained models in natural language processing [EB/OL]. (2021-08-28) [2024-06-05]. <https://arxiv.org/abs/2108.05542>.
- [11] 吴泱, 宋慧慧, 张开华, 等. 联合组间对抗数据混合与变换器学习的协同显著性检测[J]. 计算机学报, 2023, 46(9): 1838-1854.
WU Yang, SONG Huihui, ZHANG Kaihua, et al. Inter-group adversarial mixup and transformer learning for co-saliency detection [J]. Chinese Journal of Computers, 2023, 46(9): 1838-1854. (in Chinese)
- [12] KARITA S, CHEN Nanxin, HAYASHI T, et al. A comparative study on transformer vs RNN in speech applications [C]// IEEE Automatic Speech Recognition and Understanding Workshop (ASRU). Piscataway, USA: IEEE, 2019: 449-456.
- [13] WEN Qiusong, ZHOU Tian, ZHANG Chaoli, et al. Transformers in time series: a survey [C]// Proceedings of the 32nd International Joint Conference on Artificial Intelligence. Macao, China: IJCAI, 2023: 6778-6786.
- [14] ZHOU Haoyi, ZHANG Shanghang, PENG Jieqi, et al.

- Informer: beyond efficient transformer for long sequence time-series forecasting [C]// Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI Press, 2021: 11106-11115.
- [15] DU Dazhao, SU Bing, WEI Zhewei. Preformer: predictive transformer with multi-scale segment-wise correlations for long-term time series forecasting [C]// IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway, USA: IEEE, 2023: 1-5.
- [16] WU Haixu, XU Jiehui, WANG Jianmin, et al. Autoformer: decomposition transformers with auto-correlation for long-term series forecasting [C]// Proceedings of the 35th International Conference on Neural Information Processing Systems. Red Hook, USA: Curran Associates Inc., 2024: 22419-22430.
- [17] ZHOU Tian, MA Ziqing, WEN Qingsong, et al. FED-former: frequency enhanced decomposed transformer for long-term series forecasting [C]// Proceedings of the 39th International Conference on Machine Learning. Sardinia, Italy: PMLR, 2022: 27268-27286.
- [18] ZENG Ailing, CHEN Muxi, ZHANG Lei, et al. Are transformers effective for time series forecasting? [C]// Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI Press, 2023: 11121-11128.
- [19] DAS A, KONG Weihao, LEACH A, et al. Long-term forecasting with TiDE: time-series dense encoder [EB/OL]. (2024-04-04) [2024-06-05]. <https://arxiv.org/abs/2304.08424>.
- [20] NIE Yuqi, NGUYEN N H, SINTHONG P, et al. A time series is worth 64 words: long-term forecasting with transformers [C]// ICLR 2023. New York, USA: ICLR, 2023: 1-24.
- [21] ZHANG Yunhao, YAN Junchi. Crossformer: transformer utilizing cross-dimension dependency for multivariate time series forecasting [C]// ICLR 2023. New York, USA: ICLR, 2023: 1-21.
- [22] YU Chengqing, WANG Fei, SHAO Zezhi, et al. DSformer: a double sampling transformer for multivariate time series long-term prediction [C]// Proceedings of the 32nd ACM International Conference on Information and Knowledge Management. New York, USA: ACM, 2023: 3062-3072.
- [23] LIU Yong, HU Tengge, ZHANG Haoran, et al. iTransformer: inverted transformers are effective for time series forecasting [EB/OL]. (2024-03-14) [2024-06-05]. <https://arxiv.org/abs/2310.06625>.
- [24] HAN Lu, YE Hanjia, ZHAN Dechuan. The capacity and robustness trade-off: revisiting the channel independent strategy for multivariate time series forecasting [J]. IEEE Transactions on Knowledge and Data Engineering, 2024, 36(11): 7129-7142.
- [25] LIU Shizhan, YU Hang, LIAO Cong, et al. Pyraformer: low-complexity pyramidal attention for long-range time series modeling and forecasting [C]// ICLR 2022. New York, USA: ICLR, 2022: 1-20.
- [26] CHEN Peng, ZHANG Yingying, CHENG Yunyao, et al. Multi-scale transformers with adaptive pathways for time series forecasting [C]// ICLR 2024. New York, USA: ICLR, 2024: 1-18.
- [27] DAI Tao, WU Beiliang, LIU Peiyuan, et al. Periodicity decoupling framework for long-term series forecasting [C]// ICLR 2024. New York, USA: ICLR, 2024: 1-12.
- [28] ZHANG Yitian, MA Liheng, PAL S, et al. Multi-resolution time-series transformer for long-term forecasting [C]// Proceedings of the 27th International Conference on Artificial Intelligence and Statistics. Sardinia, Italy: PMLR, 2024: 4222-4230.
- [29] ASHOK A, MARCOTTE É, ZANTEDESCHI V, et al. TACTiS-2: better, faster, simpler attentional copulas for multivariate time series [EB/OL]. (2024-03-25) [2024-06-05]. <https://arxiv.org/abs/2310.01327>.
- [30] WANG Yuxuan, WU Haixu, DONG Jiaxiang, et al. Deep time series models: a comprehensive survey and Benchmark [EB/OL]. (2024-07-18) [2024-07-20]. <http://arxiv.org/abs/2407.13278?context=cs>.
- [31] HYNDMAN R J, FORECASTING G. Principles and practice [M]. [S. l.]: OTexts, 2018.

【中文责编: 方圆; 英文责编: 溯心】