

基于 Mamba 和 Time Machine 模型的数据预测方法

范亚茹¹,熊志豪²

(1.西南民族大学数学学院,四川成都610041;2.西南民族大学计算机与人工智能学院,四川成都610041)

摘要:提出了一个集成特征融合网络(Integrated Feature Fusion Net,IFFN),用于天气、交通等时间序列数据预测.该网络基于双向 Mamba 编码器捕获数据的全局与局部特征,利用 PS-Mixer (Polar-Vector and Strength-Vector Mixer) 模块提取时间序列的极性趋势和波动强度特征,从而实现对复杂时间序列的高效建模和精准预测.此外,IFFN 采用多尺度建模方法,融合时间序列的全局和局部特征,提升了模型的适应性与泛化能力.实验结果表明,IFFN 在多个公开数据集上均显著优于当前流行的数据预测模型,特别是在处理大规模和复杂模式的时间序列数据时表现卓越.

关键词:Mamba 模型;Time Machine 模型;神经网络;数据预测

中图分类号:O242

文献标志码:A

文章编号:2095-4271(2025)05-0567-06

Data prediction method based on Mamba and Time Machine models

FAN Yaru¹,XIONG Zhihao²

(1. School of Mathematics, Southwest Minzu University, Chengdu 610041, China;

2. School of Computer Science and Artificial Intelligence, Southwest Minzu University, Chengdu 610041, China)

Abstract:The Integrated Feature Fusion Network (IFFN) was proposed in this paper for the prediction of time series data such as weather and traffic. The network employed a bidirectional Mamba encoder to capture both global and local features, and utilized the PS-Mixer (Polar-Vector and Strength-Vector Mixer) module to extract polar trend and fluctuation features (polar and strength vectors) from the data, enabling efficient modeling and accurate prediction of complex time series. Furthermore, IFFN incorporated a multi-scale modeling approach, effectively integrating global and local features, thereby enhancing the model's adaptability and generalization ability. Experimental results demonstrated that IFFN significantly outperformed current state-of-the-art models across multiple publicly available datasets, especially in handling large-scale and complex-pattern time series data.

Keywords:Mamba model;Time Machine model;neural network;data prediction

1 介绍

Transformer^[1-3]能够有效预测时间数据,因其自注意力机制可以在一次计算中直接关注序列中任意时间点的信息,从而高效捕捉全局依赖关系,适合处理复杂的多变量时间序列.然而,Transformer 的计算复杂度为 $O(n^2)$,随着序列长度增加,计算和内存需求也呈二次增长,这使得处理长时间序列变得不现

实.为了克服这个缺点,PatchTST 模型^[4]通过将序列分割为子序列补丁作为 Transformer 输入,能够保留局部语义信息并减少计算和内存需求,适用于长时间跨度的历史数据.然而,对于金融、气象等数据,固定大小的补丁可能会导致丢失关键信息,特别是在处理复杂周期性、趋势性或突发事件的时间序列时,可能降低预测精度^[5-7].

收稿日期:2025-02-18

作者简介:范亚茹(1990-),女,副教授,博士,研究方向:信息与计算科学和图像处理.E-mail:yarufanfan@163.com

基金项目:中央高校基本科研业务费专项资金项目(ZYN2025013)

Abhimanyu Das^[8]等人提出的 TiDE 模型基于多层感知器,具有线性模型的简单性和速度优势,同时能处理协变量和非线性依赖关系.然而,该模型通常需要大量训练数据才能有效学习时间序列的长期依赖关系.在数据较少或标注困难的情况下,TiDE 模型可能会出现过拟合,且难以推广到未标注的数据. Maximilian Beck^[9]等人基于长短期记忆(Long Short-Term Memory, LSTM)神经网络引入指数门控机制、修改 LSTM 记忆单元并优化为 sLSTM 和 mLSTM 网络,进一步集成到残差模块中形成 XLSTM 架构.这些改进提升了网络的性能和扩展性,但在数据有限时, XLSTM 的复杂结构和超参数空间加重了调优困难.

Albert Gu^[10]等人研究出一种新的选择性结构状态空间模型名为 Mamba,通过引入动态加权机制,有效缓解了传统卷积神经网络在建模上的局限. Mamba 提供了类似于 Transformer 的高级建模能力,更重要的是 Mamba 在保证以上优点的同时避免了 Transformer 计算复杂的缺点.研究表明,采用了动态加权机制的 Mamba 非常适用于时间序列数据建模,它能够精准捕捉时间序列中的复杂模式和突发波动,避免了传统模型在处理长序列和复杂数据时的计算瓶颈.同时,它的非线性建模能力可很好地处理带噪声和有异常的数据.

Lin^[11]等人提出的 PS-Mixer 模型创新性地分离出情感分析中的极性和强度变量,将情感分析划分为两个维度,通过独立建模来保持两分量的特性.这种

分离建模方法提升了情感分析的准确性和多样性,更精确地捕捉了情感的正负倾向和强烈程度.本文首次提出采用 PS-Mixer 模块提取时间序列的极性趋势和波动强度特征,设计了一个集成特征融合网络(IFFN)用于复杂时间序列预测. IFFN 通过 Mamba 模型提取时间序列的多层次多尺度特征,捕捉各个特征之间的相关性和交互作用.将捕获的不同尺度的全局和局部上下文特征动态融合,从而准确预测时间序列数据.

2 集成特征融合网络

IFFN 网络以 Time Machine^[12]结构为基准架构,如图 1 所示.第一,通过可逆实例规范化(RevIN)将原始数据归一化,去除数据非平稳性,后将网络预测结果从归一化空间恢复到原始数据空间,确保预测与真实数据分布一致.第二,使用 DataEmbedding_inverted 方法对数据进行线性变换以及频率特征增强,高效获取时间序列的相关性和特征交互信息.第三,嵌入双向 Mamba^[13]编码层,有效地捕获上下文细节特征,并由多个全连接层(Fully Connected layers, FC)和一个 GELU 激活函数集成的生成器生成极性向量和强度向量.极性向量刻画序列的正数、负数状态趋势,强度向量表示序列的波动程度.采用自下而上的策略,充分利用时间特征.第四,通过四个 Mamba 状态空间模型进行特征融合,以获得最终预测结果.

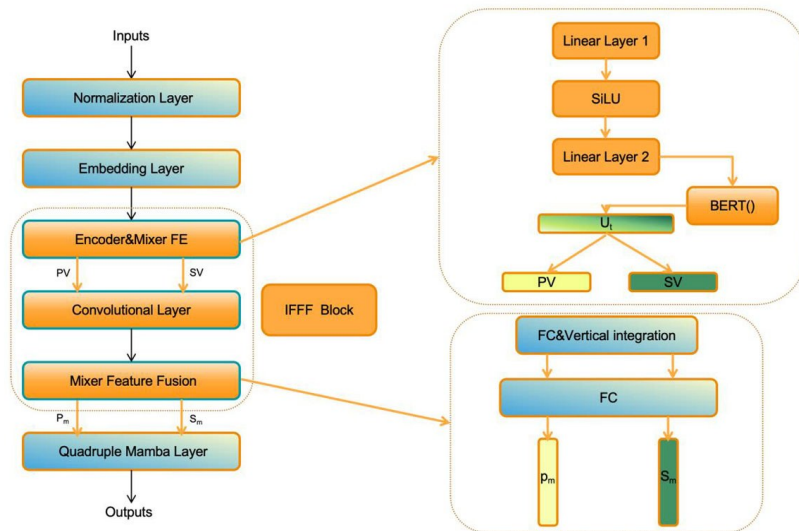


图 1 IFNN 模型

Fig. 1 The IFNN model

2.1 归一化 (Normalization)

首先采用可逆实例规范化 (RevIN) 将原始数据归一化^[14],公式如下:

$$\hat{u}_i = \frac{u_i - \mu}{\sigma} \cdot \gamma + \beta. \quad (1)$$

u_i 为时间序列的原始输入数据, μ 为每个输入序列的均值 (按实例计算, 单独对每个时间序列独立操作), σ 为每个输入序列的标准差, γ 和 β 分别为可学习的缩放和平移参数, 用于恢复重要的特征关系, 将 $U_{in} = [u_1, u_2 \dots u_L] \in \mathbb{R}^{L \times V}$ 去均值并归一化, 将时间序列的非平稳特性 (如趋势性变化) 去除后再通过下面公式进行反规范化, 将预测结果从归一化空间恢复到原始数据空间, 确保预测与真实数据的分布一致.

$$y_i = \frac{\hat{y}_i - \beta}{\gamma} \cdot \sigma + \mu. \quad (2)$$

其中 y_i 表示反归一化后的预测结果, 恢复到原始数据尺度的时间序列值. \hat{y}_i 表示模型输出的预测值 (归一化空间中的预测结果).

2.2 嵌入 (Embedding)

归一化后, 采用 DataEmbedding_inverted 方法将时间序列嵌入到模型所需的高维特征空间. 通过线性变换和频率特征增强, 生成固定维度的嵌入表示, 以适配后续模型的输入需求. 该嵌入方法适用于多通道时间序列任务, 通过时间频率特征编码和特征维度映射, 可以有效捕获时间序列的相关性和特征交互信息.

具体地, DataEmbedding_inverted 方法首先对输入的时间序列进行线性变换, 将数据从原始的低维空间映射到更高维的特征空间. 在这一过程中, 模型不仅依赖于原始数据的时间顺序, 还引入了频率特征的增强, 进一步丰富了数据的表示能力. 这种增强能够有效捕捉时间序列中时间步之间的相关性, 以及不同特征之间的交互信息. 通过这种方法, 时间序列不仅保持了时间依赖关系, 还通过特征映射捕获到不同维度上的更多信息, 为后续的模型训练提供了更加丰富的输入, 从而提升模型的预测能力^[15].

2.3 集成特征融合模块 (IFFF Block)

该模块主要运用 Mamba 编码器以及 PS-Mixer 模块进行特征提取. Mamba 的选择机制能够识别出各个变量的重要性, 类似于注意力机制, 但其计算量随着

变量数量的增加呈近似线性增长. Mamba 的单向性限制了其在全局变量上的关注能力, 其选择机制仅能基于前因变量进行合并, 因此本文中的 Mamba 编码层采用双向的 Mamba 块 (forward 和 backward), 用于捕获不同变量之间的相互关系, 公式如下:

$$\vec{Y} = \overrightarrow{\text{Mamba Block}}(U). \quad (3)$$

$$\overleftarrow{Y} = \overleftarrow{\text{Mamba Block}}(U). \quad (4)$$

其中, U 为 Embedding 层的输出结果, \vec{Y} 和 \overleftarrow{Y} 分别为两个 Mamba 块的特征提取结果. 通过 $Y = \vec{Y} + \overleftarrow{Y}$ 聚合成 Y , 输出特征 $U' = Y + U$. 采用提前训练好的 BERT 模型^[10] 将已提取出来的特征 U' 投射到低维度上, 以获得更高质量的信息和较低的计算复杂度. 将 U_i 数据分解为极性趋势向量 PV 和波动强度向量 SV, 分别刻画时间序列的极性趋势和波动强度, 公式如下:

$$PV = E_p(U_d; \theta^p). \quad (5)$$

$$SV = E_s(U_d; \theta^s). \quad (6)$$

E 为神经层, 下标 p 表示极性, s 表示强度, θ 是 PV 和 SV 的共享参数. 向量 PV 捕捉了数据的极性, 无论是正、负或是中性的. 向量 SV 捕捉了数据的具体的波动强度及成分. 如图 2 所示, 通过由多个全连接层和一个 GELU 激活函数集成的生成器 (MLP-Communicator) 可将 PV 和 SV 向量转化为刻画数据极性趋势的特征 P_m 和波动强度的特征 S_m , 公式如下:

$$P_m = BN(MLP - C(PV, \theta^p)). \quad (7)$$

$$S_m = BN(MLP - C(SV, \theta^s)). \quad (8)$$

其中 BN 是批处理规范化函数, 参数 θ^p 和 θ^s 是权值, 分别提取 P_m 的方向和 S_m 的尺度, 用于特征融合.

2.4 四个 Mamba 模型 (Quadruple Mamba)

该层通过集成四个 Mamba 状态空间模型, 同时处理时间序列中的通道混合和通道独立特性, 对不同尺度的上下文信息进行建模, 动态融合上述特征获取时间序列中的全局特征. 具体地, 该模块输入特征数据 P_m 、 S_m , 通过线性投影以确保输入数据在传递给卷积层之前进行线性变换. 通过卷积运算捕捉数据中的时序或空间依赖关系. 经过卷积层处理后, 数据进入四个 Mamba 模块. 第一个和第二个 Mamba 模块分别在不同的处理层次上提取数据特征. 第一个模块专注于捕捉全局时序依赖, 而第二个模块则更加注重局部时序依赖. 这两个 Mamba 模块帮助从输入序列中提取

全局上下文信息,并通过识别数据中跨时间点的长期依赖关系来提升模型性能.第三个和第四个 Mamba 模块处理更加局部的特征,提高模型对短期变化的捕捉

能力.这四个 Mamba 模块的输出最终通过融合机制进行合并,汇聚不同层次的特征用于最终预测,如图 3 所示.

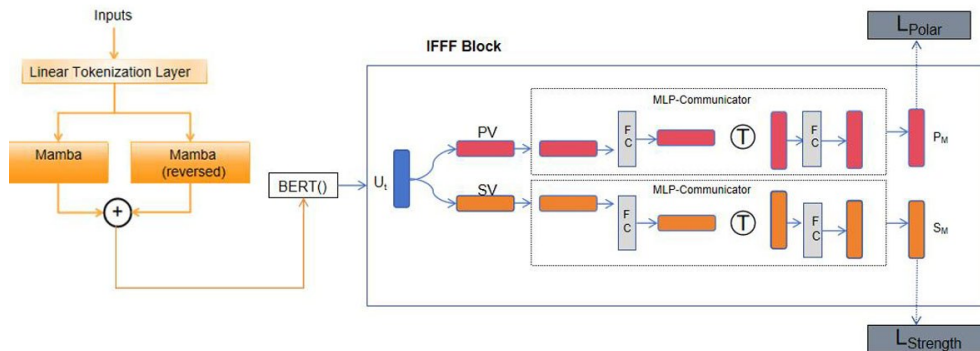


图 2 IFFF 模型架构

Fig. 2 The IFFF model Framework

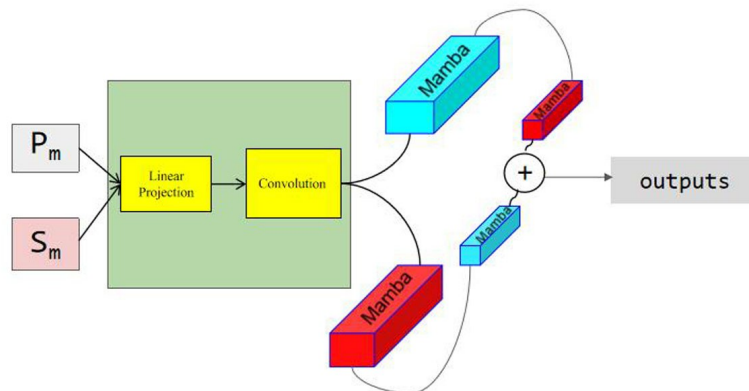


图 3 四个 Mamba 架构

Fig. 3 The Quadruple Mamba Framework

3 实验结果分析

3.1 实验数据集

我们采用 Weather、Ettm1、Ettm2、ILI、Solar、Electricity 等公开数据集作为实验数据. Weather 数据集来自气象监测站,记录了天气相关的指标; Ettm1 和 Ettm2 数据集 (Electricity Transformer Temperature) 来源于电力行业,记录了变压器的温度和其他运行参数;ILI (Influenza-Like Illness) 数据集由美国疾病控制与预防中心提供,包含了流感相关的病例统计数据;Solar 数据集来源于太阳能发电场,记录了光伏电站的发电情况;Electricity 数据集记录了来自电力公司的用电负荷数据,用于分析不同时间段的电力消耗模式.

在数据预处理方面采用如差分和滑动窗口等时间域方法,以及小波变换和傅里叶变换等频域方法,

对时间序列数据进行趋势去除、平稳化和去噪处理,减少噪声和异常波动对预测模型的负面影响,从而提高数据的稳定性和预测的准确性.

实验数据集配置信息如表 1 所示.实验环境配置如表 2 所示.

表 1 实验数据集配置表

数据集	特征数/个	观测值
Weather	21	52 696
Ettm1	7	69 680
Ettm2	7	69 680
ILI	7	966
Solar	137	52 560
Electricity	321	26 304

表 2 实验配置表

Table 2 Experimental configuration table

电脑配置	参数
CPU	14 vCPU Intel(R) Xeon(R) Platinum 8362
GPU	RTX 3090
操作系统	Windows 10
编程语言及版本	Python 3.11
深度学习框架	Pytorch

3.2 实验评价指标

在时间序列预测问题中,一般用均方误差(MSE)和平均绝对误差(MAE)作为性能评价指标.MSE 是预测值与真实值之间差异的平方的平均值,通过将误差进行平方处理,会显著放大误差对模型性能的影响.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (9)$$

其中, y_i 是真实值, \hat{y}_i 是预测值, n 是样本数量.

MAE 是预测值与真实值之间差异的绝对值的平

均值.MAE 使用误差的绝对值,而不对误差进行平方处理,这意味着它对异常值的敏感度相对较低.当数据中存在噪声或极端的异常值时,MAE 能够提供更稳健的性能评估.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|. \quad (10)$$

3.3 实验结果分析

表 3 展示了 IFFN 模型与当前流行的数据预测模型 iTransformer、crossformer、Dlinear、Rlinear、FEDformer、Autoformer 在长期时间序列预测任务中的预测结果,红色数据表示最优性能,蓝色数据为次优性能.在实验配置方面,将预测步长分为 96,192,336,720 四个等级分别进行预测,同时为 Mamba 编码器设置默认参数值:维度因子(D)=256,局部卷积宽度=2,状态扩展因子(N)=1.在六个数据集的预测任务中,图 4 展示出 IFFN 的预测精度超越了所有比较的模型.

表 3 实验结果

Table 3 Experimental results

数据集	步长	比较的模型													
		IFFN		iTransformer		crossformer		Dlinear		Rlinear		FEDformer		Autoformer	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Weather	96	0.159	0.204	0.174	0.214	0.158	0.230	0.196	0.255	0.192	0.232	0.217	0.296	0.266	0.336
	192	0.200	0.245	0.221	0.254	0.206	0.277	0.237	0.296	0.240	0.271	0.276	0.336	0.307	0.367
	336	0.266	0.297	0.278	0.296	0.272	0.335	0.283	0.335	0.292	0.307	0.339	0.380	0.359	0.395
	720	0.363	0.364	0.358	0.349	0.398	0.418	0.345	0.381	0.364	0.353	0.403	0.428	0.419	0.428
Ettm1	96	0.332	0.366	0.334	0.368	0.404	0.426	0.345	0.372	0.355	0.376	0.379	0.419	0.505	0.475
	192	0.346	0.377	0.377	0.391	0.450	0.451	0.380	0.389	0.391	0.392	0.426	0.441	0.553	0.496
	336	0.377	0.397	0.426	0.420	0.532	0.515	0.413	0.413	0.424	0.415	0.445	0.459	0.621	0.537
	720	0.439	0.436	0.491	0.459	0.666	0.589	0.474	0.453	0.487	0.450	0.534	0.490	0.671	0.561
Ettm2	96	0.178	0.261	0.180	0.264	0.287	0.366	0.193	0.292	0.182	0.265	0.203	0.287	0.255	0.339
	192	0.241	0.305	0.250	0.309	0.414	0.492	0.284	0.362	0.246	0.304	0.269	0.328	0.281	0.340
	336	0.278	0.335	0.311	0.348	0.597	0.542	0.369	0.427	0.307	0.342	0.325	0.366	0.339	0.372
	720	0.409	0.413	0.412	0.407	1.730	1.042	0.554	0.522	0.407	0.398	0.421	0.415	0.433	0.432
ILI	96	3.768	1.264	2.701	1.011	3.041	1.186	7.981	2.117	6.542	1.927	3.554	1.304	4.289	1.443
	192	2.243	0.939	1.994	0.900	3.460	1.232	6.546	1.873	4.034	1.442	2.963	1.182	3.462	1.278
	336	2.471	0.971	2.454	1.036	3.459	1.221	5.927	1.749	3.156	1.244	2.596	1.089	3.133	1.218
	720	1.882	0.915	1.942	0.931	3.640	1.305	5.774	1.739	3.116	1.225	2.595	1.086	3.114	1.195
Solar	96	0.189	0.230	0.203	0.237	0.310	0.331	0.290	0.378	0.322	0.339	0.242	0.342	0.884	0.711
	192	0.204	0.251	0.233	0.261	0.734	0.725	0.320	0.398	0.359	0.356	0.285	0.380	0.834	0.692
	336	0.219	0.266	0.248	0.273	0.750	0.735	0.353	0.415	0.397	0.369	0.282	0.376	0.941	0.723
	720	0.215	0.279	0.249	0.275	0.769	0.765	0.356	0.413	0.397	0.356	0.357	0.427	0.882	0.717
Electricity	96	0.137	0.234	0.148	0.240	0.219	0.314	0.197	0.282	0.201	0.281	0.193	0.308	0.201	0.317
	192	0.163	0.258	0.162	0.253	0.231	0.322	0.196	0.285	0.201	0.283	0.201	0.315	0.222	0.334
	336	0.176	0.275	0.178	0.269	0.246	0.337	0.209	0.301	0.215	0.298	0.214	0.329	0.231	0.338
	720	0.198	0.298	0.225	0.317	0.280	0.363	0.245	0.333	0.257	0.331	0.246	0.355	0.254	0.361

从表 3、图 4 可以看出,IFFN 模型 MSE 值均低于其他比较的模型.然而,基于 Transformer 的 Autoformer、iTransformer 和 crossformer 模型对于大的数据

集如 Solar、Ettm1 等进行预测时稍显不足.IFFN 模型在时间序列数据特征提取阶段加入 PS-Mixer 模块用于提取数据极性向量和强度向量,在增强局部特征刻

画的同时加入双向 Mamba 编码器,用于捕捉长距离依赖和全局模式,弥补局部特征的不足,生成兼顾细节与整体的综合特征,提高模型的适应性和泛化性,使模型可以适应更加复杂的时间序列预测任务。

4 总结与展望

本文提出一个用于复杂时间序列预测的集成特征融合网络 IFFN,通过嵌入双向 Mamba 编码器,捕捉

了各个特征之间的相关性和交互作用.该方法首次融合 PS-Mixer 模块,用于提取时间序列的极性趋势和波动强度特征.基于动态的特征融合,准确地预测时间序列数据.实验结果表明,IFFN 网络在长期时间序列预测任务中展现出卓越的性能,尤其在 Ettm1、Solar 两个公开数据集的预测任务中大幅领先当前流行的 iTransformer、Autoformer、Dlinear 模型.特别地,对于大规模数据集,IFFN 获得了更准确的预测结果。

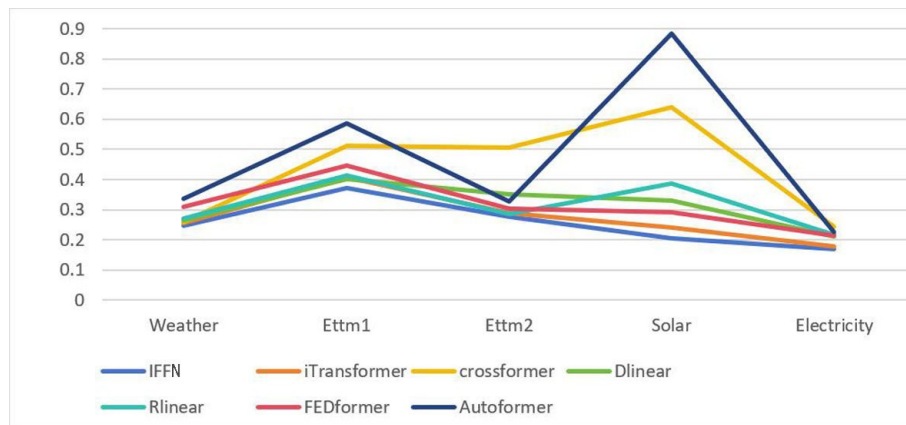


图 4 MSE 指标结果折线图

Fig. 4 Figure of MSE Results

参考文献

- [1] 李岚皓,严皓钧,周号益,等.基于神经网络的多尺度信息融合时间序列长期预测模型[J/OL].(2024-12-03)[2025-02-12].<https://link.cnki.net/urlid/51.1307.TP.20241202.1032.002>.
- [2] 陈嘉俊,刘波,林伟伟,等.基于 Transformer 的时间序列预测方法综述[J/OL].(2024-10-30)[2025-02-12].<https://link.cnki.net/urlid/50.1075.TP.20241030.0909.002>.
- [3] VASWANI A.Attention is all you need[J].Advances in Neural Information Processing Systems,2017;109-113.
- [4] NIE Y,NGUYEN N H,SINTHONG P,et al.A time series is worth 64 words;Long-term forecasting with transformers[J].arXiv preprint arXiv:2211.14730,2022.
- [5] 徐筠雯,陈宗镛,李天瑞,等.基于季节分解的混合神经网络的时间序列预测[J].计算机科学,2024,51(S2):553-559.
- [6] 范航舟,梅红岩,赵勤,等.融合双注意力机制的 GNN 多维时间序列预测[J/OL].(2024-08-28)[2025-02-12].<https://link.cnki.net/urlid/23.1538.TP.20240828.0932.006>.
- [7] 陈加毫,谢良,廖思灏,等.基于多粒度多尺度深度时空模型的长期时间序列预测方法[J].计算机科学,2025,52(2):91-98.
- [8] DAS A,KONG W,LEACH A,et al.Long-term forecasting with tide; Time-series dense encoder[J].arXiv preprint arXiv:2304.08424,2023.
- [9] BECK M,P6PPEL K,SPANRING M,et al.xLSTM; Extended Long Short-Term Memory[J].arXiv preprint arXiv:2405.04517,2024.
- [10] ALBERT G,DAO T.Mamba; Linear-time sequence modeling with selective state spaces[J].arXiv preprint arXiv:2312.00752,2023.
- [11] LIN H,ZHANG P L,LING J D,et al.PS-mixer; A polar-vector and strength-vector mixer model for multimodal sentiment analysis[J].Information Processing & Management,2023,60(2):103229.
- [12] AHAMED M A,CHENG Q.TimeMachine; A time series is worth 4 mambas for long-term forecasting [J]. arXiv preprint arXiv:2403.09898,2024.
- [13] WANG Z H,KONG F H,FENG S,et al.Is Mamba effective for time series forecasting? [J].Neurocomputing,2025,619:129178.
- [14] KIM T,KIM J,TAE Y,et al.Reversible instance normalization for accurate time-series forecasting against distribution shift[C].International Conference on Learning Representations,2021.
- [15] DAO T,GU A.Transformers are SSMS; Generalized models and efficient algorithms through structured state space duality[J].arXiv preprint arXiv:2405.21060,2024.

(责任编辑:张阳,殷锋,付强,和力新,肖丽;英文编辑:周序林,郑玉才)