華東理工大學学报 自然科学版 Natural Science Edition
JOURNAL OF EAST CHINA UNIVERSITY OF SCIENCE AND TECHNOLOGY

融合时序卷积网络与傅里叶分解的时序预测模型

李辰浩 郑红 张佳鑫 徐贤

Time Series Prediction Model Combining Temporal Convolutional Network and Fourier Decomposition

LI Chenhao, ZHENG Hong, ZHANG Jiaxin, XU Xian

在线阅读 View online: http://doi.org/10.14135/j.cnki.1006-3080.20251218001

您可能感兴趣的其他文章

Articles you may be interested in

基于Attention机制改进TimeGAN的小样本时间序列预测方法

Improved TimeGAN Based on Attention for Time Series Prediction Method with Few Shot

华东理工大学学报（自然科学版）. 2023, 49(6): 890-899 http://doi.org/10.14135/j.cnki.1006-3080.20220723001

基于时序分解和随机森林的时间序列多步预测算法

Time Series Multi-Step Prediction Algorithm Based on Time Series Decomposition and Random Forest

华东理工大学学报（自然科学版）. 2023, 49(6): 873-881 http://doi.org/10.14135/j.cnki.1006-3080.20220810001

基于误差分布补偿的混杂模糊神经网络时间序列预测模型

Hybrid Fuzzy Neural Network Based on Error Distribution Analysis for Time-Series Prediction Model

华东理工大学学报（自然科学版）. 2023, 49(4): 589-597 http://doi.org/10.14135/j.cnki.1006-3080.20220212001

基于时钟触发长短期记忆的多元时序预测

Multivariate Time Series Prediction Based on Clockwork Triggered Long Short Term Memory

华东理工大学学报（自然科学版）. 2023, 49(2): 255-268 http://doi.org/10.14135/j.cnki.1006-3080.20211102001

基于多通道图神经网络和CNN-BiLSTM的漏洞检测方法

Vulnerability Detection Method Based on Multi-Channel Graph Neural Network and CNN-BiLSTM

华东理工大学学报（自然科学版）. 2025, 51(6): 835-842 http://doi.org/10.14135/j.cnki.1006-3080.20250317001

基于宽深学习网络的音乐情感识别

Music Emotion Recognition Based on the Broad and Deep Learning Network

华东理工大学学报（自然科学版）. 2022, 48(3): 373-380 http://doi.org/10.14135/j.cnki.1006-3080.20210225007

扫码关注公众号，获取更多信息！

文章编号：1006-3080(2026)04-0553-08 DOI：10.14135/j.cnki.1006-3080.20251218001

融合时序卷积网络与傅里叶分解的时序预测模型

李辰浩， 郑 红， 张佳鑫， 徐 贤
（华东理工大学信息科学与工程学院，上海 200237）

摘要：针对时间序列预测任务中复杂非线性模型计算复杂度高且对噪声敏感的问题，本文提出了一种融合时序卷积网络(Temporal Convolutional Network, TCN)与傅里叶分解的FDTCNLinear模型。该方法先利用傅里叶变换分离时序信号的长期趋势与季节周期项，随后采用线性层高效外推长期趋势，并通过多尺度TCN挖掘季节项的长短期依赖。此外，本文还提出了一种自适应动态加权混合损失(Dynamic Weighted Hybrid Loss, DWHL)函数，结合均方误差与平均绝对误差动态调整权重以降低异常值干扰。在多个公开基准数据集上的实验表明，所提模型在不同预测长度下的精度均显著优于当前Transformer模型与线性模型。

关键词：长时间序列预测；傅里叶分解；时序卷积网络；深度学习；混合模型

中图分类号：TP391 **文献标志码：**A

在金融市场分析、能源负荷预测、交通流量调度、医疗健康监测以及气候变化建模等应用场景中，时间序列预测发挥着至关重要的作用。准确的预测结果不仅能为企业决策提供可靠依据，还能显著提升系统运行效率与社会资源利用率。然而，时间序列数据往往具有复杂的时序依赖性、高噪声水平以及长期趋势与短期波动并存的特点，这给建模与预测带来了巨大挑战。

长时间序列预测任务中，不同时期有不同的主流预测模型，但每个阶段的模型都有其特点以及局限性[1]。最早的长时间序列预测模型是自回归积分移动平均模型(ARIMA)[2]，其参数直观且易于理解，可以清楚显示时间序列的自回归部分和移动平均部分，便于解释模型结果。但对于具有复杂非线性模式的时间序列，该模型捕捉突发事件或异常值的影响时表现不佳。为解决上述问题，Hochreiter等[3]提出了长短时记忆网络(Long Short-Term Memory, LSTM)。LSTM能够捕捉长期依赖关系，特别适用于处理复杂的序列预测问题。但LSTM也存在一些明显的缺陷，如训练过程中的高计算资源需求和参数调优的复杂性，这些缺陷也经常给模型优化带来挑战。为了优化长序列数据处理的效率和效果，Transformer模型被提出并迅速成为主流[4]，已被广泛应用于各种序列到序列的任务[5]，如机器翻译、文本摘要和时间序列预测等。

目前，主流的长期时间序列预测模型大多采用基于Transformer的架构，例如Informer[6]、Pyraformer[7]、Autoformer[8]、FEDformer[9]和PatchTST[10]。其中，PatchTST和iTransformer[11]在长期时间序列预测任务上表现最佳，它们都主要依靠自注意力机制来捕捉时间序列特征。然而，自注意力机制对一些含有噪声的序列较为敏感。这是因为在自注意力机制的实现过程中，模型会对序列中的每个数据点进行特征提取和相关性计算，包括大量的噪声数据。由于模型无法有效区分噪声和有用信号，导致对噪声数据同样敏感，从而影响预测的准确性。近期研究重新审视了简单线性模型的潜力。Zeng等[12]提出了一种基于线性层的预测方法，通过分解趋势和季节性

收稿日期：2025-12-18
基金项目：上海市2024年度“科技创新行动计划”(24BC3200500, 24BC3200300)
作者简介：李辰浩(2000—)，男，安徽宣城人，硕士生，主要研究方向为时间序列预测。E-mail: 895175679@qq.com
通信联系人：郑 红，E-mail: zhenghong@ecust.edu.cn
引用本文：李辰浩，郑 红，张佳鑫，等. 融合时序卷积网络与傅里叶分解的时序预测模型 [J]. 华东理工大学学报(自然科学版)，2026，52(4)：553-560.
Citation: LI Chenhao, ZHENG Hong, ZHANG Jiaxin, *et al*. Time Series Prediction Model Combining Temporal Convolutional Network and Fourier Decomposition[J]. Journal of East China University of Science and Technology, 2026, 52(4): 553-560.

分量,在多个基准测试中表现出与 Transformer 媲美的性能,且计算复杂度仅为 $O(N)$。这表明复杂模型并非总是最佳选择,高效的简单模型在特定场景下具有竞争力。

时间序列分解作为剖析复杂时间序列数据的核心方法[13],旨在将其解构为若干关键构成要素。这些要素通常包括描绘长期演变方向的趋势项、揭示重复周期性模式的季节项,以及代表随机不可预测变动的余项。在选择分解模型时,需考量序列的特性:加法模型适用于时序数据的季节性或趋势周期波动幅度相对恒定的情形,即无论序列数值处于何种水平,此类波动的强度基本保持不变;乘法模型则更适用于波动幅度与序列整体水平成比例变化的场景,序列数值越高,季节性或周期性波动的幅度也随之增大。针对具有多个季节性周期的时间序列,多重季节趋势损失分解 (MSTL) 等方法采用加性原则进行分解[14-15];PROPHET 模型在分解趋势和季节性的基础上,还考虑到节假日等外部因素的影响[16];此外,Sen 等[17] 提出了利用矩阵分解将时间序列表达为基本成分的线性组合。

目前主流的长时间序列预测都使用均方误差(MSE)作为损失函数。MSE 在时间序列分析中得到了广泛应用,原因在于其二次性质可放大序列中的较大误差。这一特点使其特别适用于大误差比小误差更不可取的应用场景。如文献 [18] 中 MSE 在捕捉预测误差的方差方面表现出色,并提供了一个清晰的预测准确性度量标准。然而,Chai 等[19] 提出 MSE 对异常值十分敏感。在包含显著异常或噪声的数据集中,MSE 可能导致估计偏差,这强调了鲁棒预处理步骤的必要性。因此,如何构建一种能够根据预测误差实时大小、自适应调整误差惩罚权重的鲁棒损失函数,仍是当前长时间序列预测面临的关键挑战。

针对上述现有模型对高噪声敏感以及损失函数抗噪性不足的缺陷,本文提出了一种融合时序卷积网络与傅里叶分解的预测模型 FDTCNLinear。该模型利用傅里叶变换将时域信息转化为频域信息,将时间序列信号分解为趋势项与周期项,提取主要频率成分以降低噪声影响,从而提升模型的可解释性。在架构设计上,FDTCNLinear 模型采用双分支结构,其中线性层专注于建模长期趋势,而多尺度时序卷积网络(Temporal Convolutional Network, TCN)则用于捕捉季节性模式和局部时序特征。通过引入残差连接与层标准化(Layer Normalization),模型的训练稳定性与非线性表达能力得到了进一步提升。此外,本文还提出了一种自适应动态加权混合损失函数(Dynamic Weighted Hybrid Loss, DWHL)。该函数根据预测误差的实时大小,为每个样本动态生成 MSE 和平均绝对误差(MAE)的组合权重。

1 基于傅里叶分解改进的时序卷积网络线性模型

1.1 网络整体概述

首先,模型对输入的多元时间序列中的每个单变量时间序列进行移动平均分解,将每个序列分解为趋势项和季节项。其次,模型将每个时间序列的趋势项和季节项分别输入 FDTCNLinear 网络主干中,进行特征提取和预测。随后,在每个时间序列,将模型得到的预测趋势项 $\widehat{T}_t$ 和预测季节项 $\widehat{I}_t$ 进行相加,重构出时间序列的预测值 $\widehat{Y}_t$。最后,将所有单变量时间序列的预测结果进行拼接,形成最终的多元时间序列预测结果。FDTCNLinear 网络可以对时间序列数据进行高效的特征提取和预测。FDTCNLinear 的整体架构如图 1 所示。

1.2 傅里叶分解

傅里叶变换作为信号处理中的基础方法,它的原理在于将时域信号表示为一系列不同频率正弦波与余弦波的线性组合。通过这种转换,信号内在的周期性及趋势性特征得以在频域中清晰呈现。通常,时间序列中缓慢变化的部分对应低频分量,反映长期趋势;而快速波动部分对应高频分量,反映短期周期或噪声特征。快速傅里叶变换(Fast Fourier Transform, FFT)作为傅里叶变换的高效实现方法,其时间复杂度仅为 $O(N\log N)$,被广泛应用于时序信号分析与建模中,可在较低计算开销下实现高效的频谱分解。

本文设计了傅里叶分解模块,将输入的时间序列解耦为趋势项和季节性项,从而为后续的预测任务提供清晰的子成分。该模块的执行流程包含 4 个关键步骤:(1)频域变换,将时域序列映射至频域;(2)频率分量筛选,基于幅度大小选取主要周期成分;(3)信号重构,通过逆变换生成时域分量;(4)残差计算,提取未能被分解的随机扰动信息。

首先,对输入时间序列 $X\in\mathbb{R}^{B\times L\times C}$ 进行快速傅里叶变换,将时间域信号转换为频率域表示:

$$F(X)(k)=\sum_{n=0}^{L-1}X(n)\mathrm{e}^{-\frac{j2\mathrm{mathrm}\pi kn}{L}},\quad k=0,1,\cdots,L-1 \tag{1}$$

其中,$F(x)\in\mathbb{C}^{B\times L\times C}$;$k$ 表示频率分量的索引,对应的归一化频率 $f_k=k/L$;Δt 是采样间隔;L 是输入时间

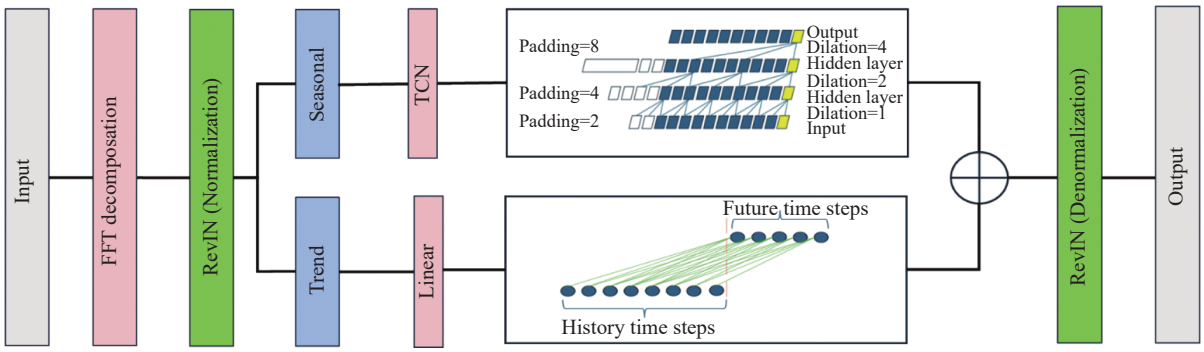


图 1　FDTCNLinear 网络整体架构

Fig. 1　Architecture of the FDTCNLinear network

序列的长度; B 是批次大小; C 是通道数。提取幅值最大的前 k 个主导频率分量(k 表现为保留的成分个数, 本文取 $k=5$), 用于构建季节性成分 $X_{\text{seasonality}}$, 其余低频分量用于重构趋势项 X_{trend} 。两类成分分别通过逆傅里叶变换(IFFT)重构如下:

$$X_{\text{seasonality}}(n)=\frac{1}{L}\sum_{k\in \text{top_}k}F(X)(k)\mathrm{e}^{\frac{j2\pi kn}{L}} \tag{2}$$

$$X_{\text{trend}}(n)=\frac{1}{L}\sum_{k\in \text{low_frequencies}}F(X)(k)\mathrm{e}^{\frac{j2\pi kn}{L}} \tag{3}$$

最后, 残差项 X_{residual} 通过从原始序列中减去趋势项与季节性项得到:

$$X_{\text{residual}}(n)=X(n)-X_{\text{trend}}(n)-X_{\text{seasonality}}(n) \tag{4}$$

将该模块输出的趋势项和季节性项分别输入线性层与多尺度 TCN, 进行特征提取与序列预测, 实现时间序列中长期趋势与周期模式的有效分离与建模。

1.3　线性模型

线性模型对傅里叶分解得到的趋势成分 X_{trend} 进行预测, 捕捉时间序列的长期变化模式。其中, X_{trend} 由傅里叶分解的低频分量构成, 反映时间序列的长期变化模式。线性模型通过线性变换拟合这些低频分量, 生成未来时间步 (P) 的趋势预测 Y_{trend} , 为后续预测提供长期变化的基础。线性模型通过全连接层将历史趋势映射到未来预测长度, 线性预测模型如图 2 所示。具体过程如下: 首先输入趋势成分 $X_{\text{trend}}\in\mathbb{R}^{B\times L\times C}$, 接着将输入展平为 $\mathbb{R}^{B\times L\times C}$, 通过线性层映射为 $\mathbb{R}^{B\times P\times C}$, 然后, 重塑为预测结果 $Y_{\text{trend}}\in\mathbb{R}^{B\times P\times C}$ 。线性变换公式如下:

$$Y_{\text{trend}}=X_{\text{trend}}\boldsymbol{W}+b \tag{5}$$

其中, $\boldsymbol{W}\in\mathbb{R}^{(L\times C)\times(P\times C)}$ 是权重矩阵, $b\in\mathbb{R}^{P\times C}$ 是偏置。

1.4　时序卷积网络

通过多层一维卷积堆叠与膨胀卷积机制[20],

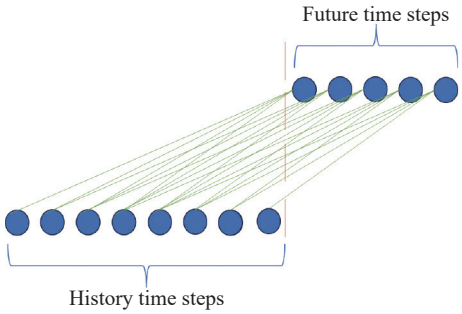


图 2　线性预测示意图

Fig. 2　Schematic diagram of linear prediction

TCN 能够在不同时间尺度上捕捉时序模式, 从而生成更加准确的季节性预测。因此, 本文用 TCN 处理傅里叶分解后的季节性分量。为了充分捕获季节性分量中多层次的时序动态变化, 本文在 TCN 模块中构建了多尺度特征提取机制。具体而言, 模型采用尺寸为 $k\in\{3,5,7\}$ 的一维卷积核进行并行处理: 小尺寸卷积核(k=3)专注于捕捉局部的短期波动模式, 中等尺寸(k=5)用于挖掘中期时序依赖, 而大尺寸(k=7)则致力于提取长期的周期性特征。为以更低的计算代价成倍扩张感受野, 模型引入了膨胀因果卷积, 并设置了多级递增的膨胀系数。这种多尺度卷积核与多级膨胀系数的结合, 使得网络在深层能够覆盖足够长的前置历史信息。该网络结构下的感受野的计算公式如下:

$$R=(\text{kernel_size}-1)\times\sum\text{dilation}+1 \tag{6}$$

每个 TCN 块均采用残差连接以缓解梯度消失问题, 并提升训练稳定性, 其计算公式如下:

$$\text{Output}=\text{ReLU}(x+\text{Conv}(x)) \tag{7}$$

在每次卷积操作后引入层标准化, 以进一步稳定训练过程并提高模型的收敛速度。通过调整最后一层卷积的步幅, TCN 能够灵活适配不同预测长度, 输出 $Y_{\text{seasonality}}\in\mathbb{R}^{B\times P\times C}$, 其中包含时间序列的周期性变化特征。通过多尺度卷积核与膨胀卷积的协同设

计, TCN 能够在不同时间尺度上同时捕捉短期波动与长期周期模式, 从而显著提升整体预测性能与模型鲁棒性。

1.5 损失函数

为了解决传统 MSE 对异常值敏感以及 MAE 在零点处梯度不连续的问题, 本文提出了 DWHL 函数。本文定义第 i 个样本的预测误差为 $e_i = \left|Y_i - \widehat{Y}_i\right|$, 引入一个基于误差衰减的动态权重系数 λ_i, 公式如下:

$$\lambda_i = \exp\left(-\frac{e_i}{\tau}\right) \tag{8}$$

其中, τ 为控制权重衰减速率的超参数(温度系数)。

总损失函数为 MSE 与 MAE 的动态加权和, 定义如下:

$$L_{\mathrm{DWHL}} = \frac{1}{n}\sum_{i=1}^{n}\left[\lambda_i \cdot \frac{1}{2}\left(Y_i - \widehat{Y}_i\right)^2 + (1-\lambda_i)\left|Y_i - \widehat{Y}_i\right|\right] \tag{9}$$

对 DWHL 函数的极端误差状态进行渐近性质分析: 当进入小误差状态($e_i \to 0$)时, $\lambda_i \to 1$, 总损失函数渐近退化为标准的 MSE 特性。当进入大误差状态($e_i \to \infty$)时, 针对噪声或突变导致的大误差, $\lambda_i \to 0$, 损失函数渐近退化为 MAE 特性。由于 MAE 施加恒定的线性惩罚, 避免了 MSE 对大误差的平方级放大, 从而自适应地抑制了异常值对整体梯度的破坏, 保障了模型的优化鲁棒性。此外, 超参数 τ 决定了损失函数从 MSE 向 MAE 过渡的敏感阈值。 τ 较小时, 模型对误差增大更敏感, 迅速切换为 MAE 模式以应对频繁的高幅噪声; τ 较大时, 过渡则更平缓。这种设计摆脱了传统 Huber 损失需要硬编码分段阈值的局限, 实现了全域可导与自适应平滑过渡。

2 实验结果与分析

2.1 数据集

本文使用的数据集为时间序列预测领域常用的 5 种数据集, 包含不同程度的噪声与异常样本(如交通流量突变、电力负载波动、汇率异常波动等), 客观反映了真实时间序列场景下的数据特性: (1) 交通(Traffic)数据集[21]: 记录了 2015 年至 2016 年期间, 旧金山高速公路每小时监测的道路占用率情况。(2) 电力变压器温度(ETT)系列数据集[22]: 涵盖小时粒度(ETTh1、ETTh2)与 15 min 粒度(ETTm1、ETTm2)两类, 收录了 2016 年 7 月至 2018 年 7 月期间电力变压器的油温与负载等关键运行指标。(3) 电力(Electricity)数据集[23]: 记录了 2012 年至 2014 年间 321 名居民的每小时用电量数据, 反映了不同用户的电力消耗模式。(4) 天气(Weather)数据集[22]: 采集了 2020 年德国的 21 个气象指标, 如温度、湿度等, 记录间隔为 10 min 一次。(5) 金融(Exchange-Rate)数据集[24]: 汇集 1990 年至 2016 年间包含中国、英国等 8 个国家的每日汇率记录。

2.2 评估指标

为确保与已有研究保持可比性, 选用 MSE与 MAE 作为核心评估指标。

2.3 对比基线与实验设计

本文选择多种基于 Transformer 的模型作为对比, 包括 PatchTST、iTransformer 和 TimesNet[25] 以及非 Transformer 模型 Dlinear、OLinear[26] 和 Softs[27]。为确保对比实验的公正性, 所有模型均采用相同的实验配置与数据划分方式。窗口预测长度分别考察 96、192、336、720 步这 4 种情况。所有模型均在完全相同的训练集与测试集划分下进行训练与性能评估, 以保证可信度和可重复性。

2.4 对比实验

表 1 示出了不同模型多元长时间序列预测结果对比。图 3 示出了 ETTh1 数据集上不同模型的预测结果对比。如表 1 与图 3 所示, 本文所提出的 FDTCNLinear 在多数实验设定下均优于现有的预测方法, 表现出了更优的预测性能。特别是对含有显著噪声及强非线性模式的时间序列进行预测时, 该方法展现了更佳的鲁棒性与泛化能力。

从表 1 可以看出, 本文提出的 FDTCNLinear 模型在长序列预测任务(如预测长度为 336 步和 720 步)中, 依靠多尺度 TCN 模块的广阔感受野有效抑制了长程误差累积: 以 ETTh1 数据集为例, 当预测长度从 96 步增至 720 步时, FDTCNLinear 的 MSE 仅从 0.370 缓慢上升至 0.431; 而同为线性架构的 OLinear 误差从 0.360 快速恶化至 0.463, 基于 Transformer 架构的 iTransformer 更是升至 0.495。同时, 采用精简结构化建模的 FDTCNLinear避免了过参数化自注意力机制对随机噪声的过拟合, 在 ETTh2(如预测长度为 336 步) 任务中以 0.400的 MSE 显著优于 Softs (0.410)等复杂深度学习模型。最后, 得益于 DWHL 函数对异常值惩罚权重的动态调节, 本模型在面对采样频率与特征分布差异较大的数据集时, 有效抵御了噪声干扰与分布漂移。

2.5 消融实验

为探究 TCN 模块与 DWHL 函数在 FDTCNLinear 模型中的作用, 实验选取了典型的 ETTh1 数据集进行消融实验, 实验分别构建了 3 种变体模型: 仅包含

表 1　不同模型多元长时间序列预测结果对比

Table 1　Comparison of multivariate long-term sequence prediction results of different models

Model metric	Prediction length	FDTCNLinear (Ours)		OLinear[26] (2025)		Softs[27] (2024)		iTransformer[11] (2024)		DLinear[12] (2023)		PatchTST[10] (2023)		TimesNet[25] (2023)	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.370	0.392	**0.360**	**0.382**	0.381	0.399	0.386	0.405	0.379	0.403	0.377	0.397	0.389	0.412
	192	**0.404**	**0.409**	0.416	0.414	0.435	0.431	0.424	0.440	0.408	0.419	0.409	0.425	0.440	0.443
	336	**0.429**	**0.433**	0.457	0.438	0.480	0.452	0.449	0.460	0.440	0.440	0.441	0.444	0.523	0.487
	720	**0.431**	0.463	0.463	**0.462**	0.499	0.488	0.495	0.487	0.471	0.493	0.467	0.477	0.521	0.495
ETTh2	96	**0.280**	**0.322**	0.284	0.329	0.297	0.347	0.297	0.348	0.300	0.364	0.292	0.342	0.334	0.370
	192	**0.357**	**0.370**	0.360	0.379	0.373	0.394	0.372	0.403	0.387	0.423	0.348	0.384	0.404	0.413
	336	**0.400**	**0.410**	0.409	0.415	0.410	0.426	0.388	0.417	0.490	0.487	0.377	0.416	0.389	0.435
	720	**0.410**	**0.424**	0.415	0.431	0.411	0.433	0.424	0.444	0.704	0.597	0.406	0.441	0.434	0.448
ETTm1	96	**0.296**	0.337	0.302	**0.334**	0.325	0.361	0.300	0.353	0.300	0.345	0.329	0.367	0.340	0.378
	192	**0.333**	**0.357**	0.357	0.363	0.375	0.389	0.341	0.380	0.336	0.366	0.367	0.385	0.392	0.404
	336	**0.367**	**0.379**	0.387	0.385	0.405	0.412	0.374	0.396	0.367	0.386	0.399	0.410	0.423	0.426
	720	**0.427**	**0.419**	0.452	0.426	0.466	0.447	0.429	0.430	0.419	0.416	0.454	0.439	0.475	0.453
ETTm2	96	**0.164**	0.250	0.169	**0.249**	0.180	0.261	0.175	0.266	0.164	0.255	0.175	0.259	0.189	0.265
	192	**0.221**	**0.289**	0.232	0.290	0.246	0.306	0.242	0.312	0.224	0.304	0.241	0.302	0.254	0.310
	336	**0.279**	0.333	0.291	**0.328**	0.319	0.352	0.282	0.337	0.277	0.337	0.305	0.343	0.313	0.345
	720	**0.388**	0.409	0.389	**0.387**	0.405	0.401	0.375	0.394	0.371	0.401	0.402	0.400	0.413	0.402
Exchange	96	0.082	**0.200**	0.082	0.200	0.089	0.209	0.086	0.205	**0.080**	0.202	0.087	0.204	0.112	0.242
	192	**0.167**	0.298	0.171	**0.293**	0.204	0.323	0.177	0.299	0.182	0.321	0.177	0.300	0.209	0.334
	336	0.300	**0.409**	0.331	0.414	0.358	0.434	0.331	0.417	0.327	0.434	**0.297**	0.399	0.358	0.435
	720	0.643	**0.601**	0.837	0.688	1.195	0.805	0.846	0.693	**0.578**	0.605	0.843	0.692	0.944	0.736
Weather	96	0.175	0.217	0.153	0.190	0.166	0.208	0.157	0.207	0.170	0.230	**0.150**	**0.200**	0.168	0.214
	192	0.215	0.257	0.200	0.235	0.217	0.253	0.200	0.248	0.216	0.273	**0.191**	**0.239**	0.219	0.262
	336	0.259	0.297	0.258	0.280	0.282	0.300	0.252	0.287	0.258	0.307	**0.242**	**0.279**	0.278	0.302
	720	0.330	0.363	0.337	0.333	0.356	0.351	0.320	0.336	0.323	0.362	**0.312**	**0.330**	0.353	0.351
Electricity	96	0.141	0.235	**0.131**	**0.221**	0.143	0.233	0.134	0.230	0.140	0.237	0.143	0.247	0.169	0.271
	192	0.154	0.247	**0.150**	**0.238**	0.158	0.248	0.154	0.250	0.154	0.251	0.158	0.260	0.180	0.280
	336	0.170	0.263	**0.165**	**0.254**	0.178	0.269	0.169	0.265	0.169	0.268	0.168	0.267	0.204	0.304
	720	0.204	0.294	**0.191**	**0.279**	0.218	0.305	0.194	0.288	0.204	0.301	0.214	0.307	0.205	0.304
Traffic	96	0.416	0.431	0.398	**0.226**	**0.376**	0.251	0.363	0.265	0.395	0.275	0.370	0.262	0.595	0.312
	192	0.438	0.272	0.439	**0.241**	**0.398**	0.261	0.384	0.273	0.407	0.280	0.386	0.269	0.613	0.322
	336	0.504	0.502	0.464	**0.250**	**0.415**	0.269	0.396	0.277	0.417	0.286	0.396	0.275	0.626	0.332
	720	0.621	0.581	0.502	**0.270**	**0.447**	0.287	0.445	0.308	0.454	0.308	0.435	0.295	0.635	0.340
Optimal result count		16	13	5	15	4	0	0	0	2	0	5	4	0	0

趋势项外推全连接层的“原始模型”、引入多尺度特征提取机制的“仅添加 TCN 变体”, 以及替换损失优化核心的“仅添加 DWHL 变体”。 FDTCNLinear 消融实验结果如表 2 所示。消融实验下不同模块在 ETTh1 数据集上的相对提升率如图 4所示。结合表 2 进行深度分析, 可以得出:

(1)在原始模型的线性架构基础上引入多尺度 TCN 模块后, 模型的预测精度在所有预测跨度上均实现了系统性提升。由于原始模型仅采用线性全连接层对信号进行宏观映射, 容易受到高频随机噪声的恶性干扰并发生过拟合。而 TCN 模块通过多尺度卷积核与多级递增膨胀因果卷积的有机结合, 能够有效分离并在频域重构后的季节周期项中高效捕捉长短期时序依赖关系。这种机制不仅显著减少了噪声信号对网络主干的干扰, 更增强了模型对局部动态波动的刻画能力。特别是在极端长序列预测任

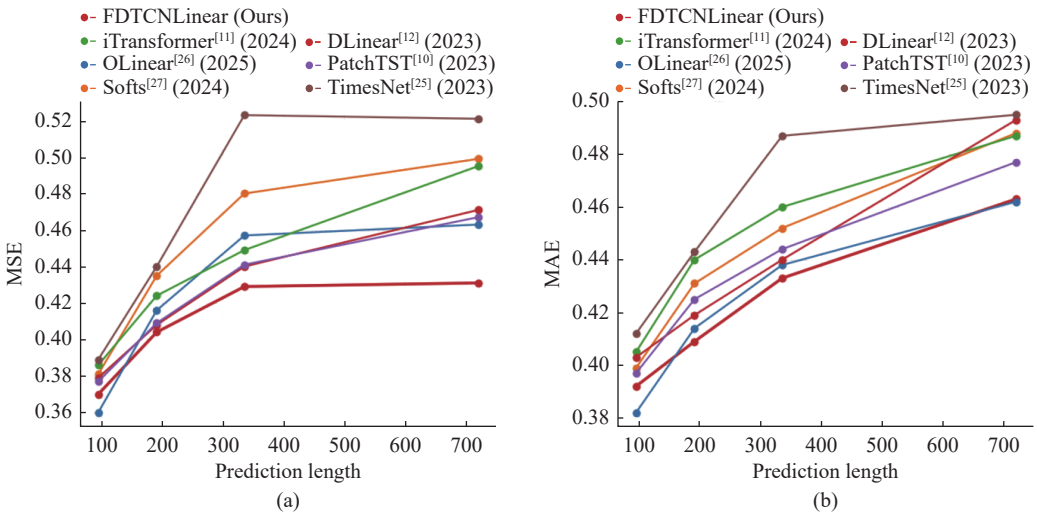


图 3 ETTh1 数据集上不同模型的预测结果对比

Fig. 3 Comparison of prediction results of different models on the ETTh1 dataset

表 2 FDTCNLinear 消融实验结果

Table 2 Results of FDTCNLinear ablation experiment

Model metric	Prediction length	Original model		Add TCN		Add DWHL		FDTCNLinear	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.375	0.400	0.375	0.399	0.372	0.397	**0.370**	**0.392**
	192	0.412	0.419	0.405	0.416	0.411	0.418	**0.404**	**0.409**
	336	0.439	0.443	0.430	0.435	0.431	0.434	**0.429**	**0.433**
	720	0.472	0.490	0.439	0.456	0.445	0.460	**0.431**	**0.433**

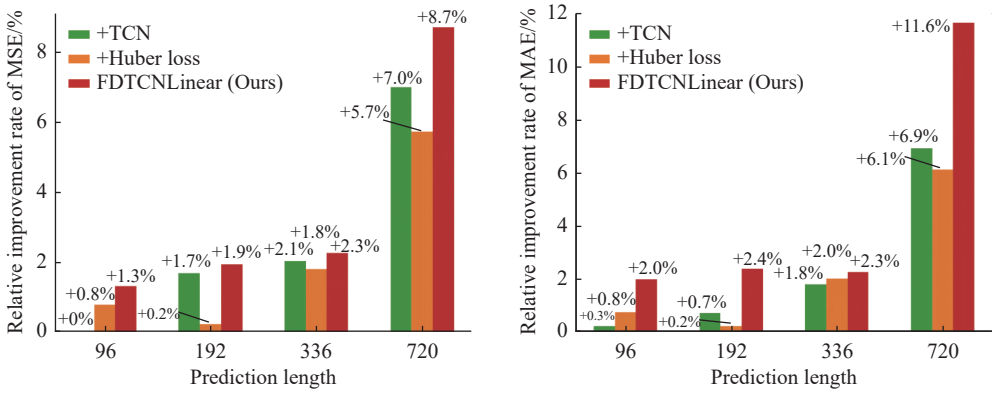


图 4 消融实验下不同模块在 ETTh1 数据集上的相对提升率

Fig. 4 Relative improvement rates of different modules under ablation experiments on the ETTh1 dataset

务(预测长度为 720 步)时,添加 TCN 模块使 MSE 从 0.472 大幅降低至 0.439,这充分证实了多尺度特征提取在抑制长程误差累积、稳固长时预测基准方面的优异性能。

(2) 消融结果表明,相较于传统的 MSE 损失函数,DWHL 损失函数在面对异常波动时,能显著提升模型在各预测周期的预测精度。由于 ETTh1 数据集蕴含大量由于电网负载随机波动导致的异常值,传统 MSE 会对这些突变大误差施加二次平方放大,使梯度在反向传播中过度偏向噪声点,导致估计偏差。而 DWHL 损失函数通过引入基于误差指数衰减的温度权重调节机制,实时预测误差自适应调节 MSE 与 MAE 的贡献权重:在面临大幅噪声和突变异常值时将损失平滑切换为线性惩罚特性,从而削弱了异常梯度对整体网络参数的底层破坏。在预测长度为 96 步的短周期任务中,仅添加 DWHL变体的

MAE 达到 0.397, 优于原始模型的 0.400, 这为其对复杂时序的平滑微调与自适应优化能力提供了实验数据支撑。

(3)当 TCN 模块与 DWHL 损失函数同时作用形成最终的 FDTCNLinear 模型时, 两者的优势表现出显著的协同叠加增益。在所有评测跨度下, 完整模型均取得了最优的预测性能。这一结果表明, 时域频域多尺度特征提取(TCN 分支)与参数底层鲁棒优化(DWHL 损失)从前向特征映射与反向梯度传播两个维度形成了互补闭环, 共同保障了模型在面临复杂、高噪声长时间序列场景下的卓越稳定性与极强的泛化表现。

3　结束语

针对基于注意力机制的类 Former 模型(如 iTransformer、PatchTST 等)在时序预测任务中易受噪声干扰、导致预测精度下降的问题, 本文提出了一种融合傅里叶分解与 TCN 的 FDTCNLinear 预测模型。实验结果表明, 与一系列基于 Linear 结构的时序预测模型相比, FDTCNLinear 在多个真实数据集上均取得了更优的性能表现。该模型能够有效抑制噪声影响, 捕捉复杂的非线性时序特征, 从而展现出更高的鲁棒性与泛化能力, 具有良好的实际应用前景。

同时, 实验结果也证实了 DWHL 函数相较于传统 MSE 在时序预测任务中具有更优异的优化性能。此外, 面对更复杂和更长时间跨度的预测任务, 本文模型仍有进一步的提升空间, 如探索更高效的注意力机制、改进序列分解方法、优化模型结构等。如何进一步结合焦点增强注意力机制, 提高输出预测的精度和模型的计算效率, 将是本文接下来的研究重点。

参考文献:

[1] 毛远宏, 孙琛琛, 徐鲁豫, 等. 基于深度学习的时间序列预测方法综述 [J]. 微电子学与计算机, 2023, 40(4): 8-17.

[2] BOX G E P, JENKINS G M. Some recent advances in forecasting and control[J]. Journal of the Royal Statistical Society: Series C (Applied Statistics), 1968, 17(2): 91-109.

[3] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.

[4] 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, CA, USA: Curran Associates Inc., 2017: 6000-6010.

[5] 郭津延, 郑红, 杜佳宇, 等. 基于双尺度自适应令牌注意力的交通流量预测 [J]. 华东理工大学学报 (自然科学版), 2025, 51(6): 817-826.

[6] ZHOU H, ZHANG S, PENG J, *et al*. Informer: Beyond efficient transformer for long sequence time-series forecasting[C]//Proceedings of the 35th AAAI Conference on Artificial Intelligence. [s.l.]: AAAI Press, 2021: 11106-11115.

[7] LIU S, YU H, LIAO C, *et al*. Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting[C/OL]//International Conference on Learning Representations. [s.l.]: ICLR, 2022. https://openreview.net/forum?id=0EXmFzUn5I.

[8] WU H, XU J, WANG J, *et al*. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting[C]//Advances in Neural Information Processing Systems. [s.l.]: Curran Associates Inc., 2021: 22419-22430.

[9] ZHOU T, MA Z, WEN Q, *et al*. FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting[C]// Proceedings of the 39th International Conference on Machine Learning. Baltimore, Maryland, USA: PMLR, 2022: 27268-27286.

[10] NIE Y, NGUYEN N H, SINTHONG P, *et al*. A time series is worth 64 words: Long-term forecasting with transformers[C/OL]// Proceedings of the 11th International Conference on Learning Representations. Kigali: ICLR, 2023. https://openreview.net/forum?id=8szSgdP0fD.

[11] LIU Y, HU T, ZHANG H, *et al*. iTransformer: Inverted transformers are effective for time series forecasting[C/OL]//Proceedings of the 12th International Conference on Learning Representations. Vienna, Austria: ICLR, 2024. https://openreview.net/forum?id=JePfAI8fah.

[12] ZENG A, CHEN M, ZHANG L, *et al*. Are transformers effective for time series forecasting[C] //Proceedings of the AAAI Conference on Artificial Intelligence. Washington, DC, USA: AAAI Press, 2023: 11121-11128.

[13] 文家璇, 王苗, 刘济. 基于时序分解和随机森林的时间序列多步预测算法 [J]. 华东理工大学学报 (自然科学版), 2023, 49(6): 873-881.

[14] XIONG T, LI C, BAO Y. Seasonal forecasting of agricultural commodity price using a hybrid STL and ELM method: evidence from the vegetable market in China[J]. Neurocomputing, 2018, 275: 2831-2844.

[15] WEN Q, ZHANG Z, LI Y, *et al*. Fast Robust STL: Efficient and robust seasonal-trend decomposition for time series with complex patterns[C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. [s.l.]: ACM, 2020: 2203-2213.

[16] GUO L, FANG W, ZHAO Q, *et al*. The hybrid PROPHET-SVR approach for forecasting product time series demand with seasonality[J]. Computers and Industrial Engineering,

2021, 161: 1075

[17] SEN R, YU F, DHILLON I. Think globally, act locally: A deep neural network approach to high-dimensional time series forecasting [C]// Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc., 2019: 14084-14095.

[18] HYNDMAN R J, KOEHLER A B. Another look at measures of forecast accuracy[J]. International Journal of Forecasting, 2006, 22(4): 679-688.

[19] CHAI T, DRAXLER R R. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature[J]. Geoscientific Model Development, 2014, 7(3): 1247-1250.

[20] BAI S, KOLTER Z, KOLTUN V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling [EB/OL]. (2018-03-04). https://doi.org/10.48550/arXiv.1803.01271.

[21] California Department of Transportation. Caltrans performance measurement system (PeMS)[DB/OL]. [2025-12-18]. https://pems.dot.ca.gov.

[22] Max Planck Institute for Biogeochemistry. Weather data of the max planck institute for biogeochemistry[DB/OL]. [2025-12-18]. https://www.bgc-jena.mpg.de/wetter.

[23] TRINDADE A. ElectricityLoadDiagrams20112014 data set [DS/OL]. (2015-03-13) [2025-12-18]. https://archive.ics.uci.edu/ml/datasets/ElectricityLoadDiagrams20112014.

[24] LAI G K. Multivariate time series data [DS/OL]. (2017-03-10) [2025-12-18]. https://github.com/laiguokun/multivariate-time-series-data.

[25] WU H, HU T, LIU Y, *et al*. TimesNet: Temporal 2D-variation modeling for general time series analysis[C/OL]//Proceedings of the 11th International Conference on Learning Representations. Kigali: ICLR, 2023. https://openreview.net/forum?id=ju_Uqw384Oq.

[26] YUE W, LIU Y, LI H, *et al*. OLinear: A linear model for time series forecasting in orthogonally transformed domain[EB/OL]. (2025-05-14). https://arxiv.org/abs/2505.08550.

[27] HAN L, CHEN X Y, YE H J, *et al*. SOFTS: Efficient multivariate time series forecasting with series-core fusion[C]//Advances in Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc., 2024: 64145-64175.

Time Series Prediction Model Combining Temporal Convolutional Network and Fourier Decomposition

LI Chenhao, ZHENG Hong, ZHANG Jiaxin, XU Xian

(School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China)

Abstract: To tackle the high computational complexity and noise sensitivity of complex nonlinear models for time series forecasting, this paper proposes FDTCNLinear, a model integrating Temporal Convolutional Network (TCN) and the Fourier decomposition. The model first leverages the Fourier transform to decompose time-series signals into long-term trends and seasonal-cyclic components. It then adopts a linear layer for efficient extrapolation of long-term trends, and employs a multi-scale TCN to mine long- and short-term temporal dependencies hidden within seasonal components. In addition, this study designs an adaptive Dynamic Weighted Hybrid Loss (DWHL) function. By dynamically balancing the weights of Mean Squared Error (MSE) and Mean Absolute Error (MAE), this loss alleviates disturbances caused by outliers. Experiments on multiple public benchmark datasets verify that the proposed model achieves remarkably higher prediction accuracy than state-of-the-art Transformer-based models and linear baselines across all prediction horizons.

Key words: long-term time series prediction; Fourier decomposition; temporal convolutional network; deep learning; hybrid model

（责任编辑：李娟）