

基于深度学习的时间序列预测方法综述

梁鑫¹, 黄荣华¹, 任旭²

1. 广西师范大学 数学与统计学院, 广西 桂林 541004

2. 国家统计局 普查中心, 北京 100826

DOI:10.61369/ASDS.2026030007

摘要: 时间序列预测作为解析历史数据动态演化模式并推断未来状态的核心技术, 在经济、交通、能源及气象等领域具有重要的决策支持价值。随着大数据技术的发展与计算能力的指数级提升, 深度学习凭借其卓越的非线性拟合与自动特征提取能力, 已克服传统统计模型在处理高维、非平稳及复杂时空相关性数据时的局限, 成为时间序列预测的主流范式。本文旨在系统综述基于深度学习的时间序列预测方法的研究进展。首先, 文章构建了时间序列预测的理论框架, 明确了任务定义与方法分类体系。其次, 深入剖析了四类主流深度学习架构的技术演进路径与核心机制: 基于循环神经网络 (RNN) 的方法经历了从经典 LSTM/GRU 到现代化 SegRNN、xLSTM 的演变, 证明了递归归纳偏置在序列依赖建模中的持续活力; 基于卷积神经网络 (CNN) 的方法利用扩张卷积与多尺度架构 (如 TimesNet), 在局部特征提取与并行计算效率方面展现出显著优势; 基于 Transformer 的方法凭借自注意力机制主导了长程依赖建模, 并对改造 Transformer 架构 (如 iTransformer 等) 突破计算瓶颈; 而基于多层感知机 (MLP) 的方法则以极简线性架构 (如 DLinear) 实现了精度与效率的平衡, 挑战了复杂模型的必要性。此外, 本文系统梳理了该领域面临的关键挑战, 包括不等长序列的鲁棒处理、黑箱模型的可解释性缺失、大规模数据的计算成本以及跨域泛化能力不足等问题。最后, 文章展望了未来的研究趋势, 指出时间序列基础模型的构建、多模态预训练范式及边缘设备上的轻量化部署将是下一阶段的创新重点。本文期望通过对现有方法的全面梳理与前瞻性分析, 为时间序列预测领域的理论研究与工程应用提供参考。

关键词: 时间序列预测; 深度学习; Transformer; 基础模型

A Review of Deep Learning-Based Time Series Forecasting Methods

Liang Xin¹, Huang Ronghua¹, Ren Xu²

1. School of Mathematics and Statistics, Guangxi Normal University, Guilin, Guangxi 541004

2. National Bureau of Statistics, Census Center, Beijing 100826

Abstract: Time series forecasting, as a core technology for analysing the dynamic evolution patterns of historical data and inferring future states, holds significant decision-support value across economic, transportation, energy, and meteorological domains. With the advancement of big data technologies and exponential growth in computational power, deep learning has overcome the limitations of traditional statistical models when handling high-dimensional, non-stationary, and complex spatiotemporal data through its exceptional nonlinear fitting and automatic feature extraction capabilities. It has consequently emerged as the mainstream paradigm for time series forecasting. This paper aims to provide a systematic review of research progress in deep learning-based time series forecasting methods. Firstly, the paper constructs a theoretical framework for time series forecasting, clarifying task definitions and establishing a classification system for methodologies. Secondly, it conducts an in-depth analysis of the technological evolution and core mechanisms of four mainstream deep learning architectures: methods based on recurrent neural networks (RNNs) have evolved from classical LSTM/GRU to modernised SegRNN and xLSTM, demonstrating the enduring vitality of recurrent induction bias in modelling sequential dependencies; Convolutional neural network (CNN)-based approaches, leveraging dilated convolutions and multi-scale architectures (e.g., TimesNet), demonstrate significant advantages in local feature extraction and parallel computational efficiency; Transformer-based approaches have dominated long-range dependency modelling through

基金项目: 国家自然科学基金资助项目 (12161009); 广西哲学社会科学课题 (23BTJ001)。

作者简介:

梁鑫 (1978-), 男, 副教授, 硕士生导师; E-mail: liangxin@gxnu.edu.cn;

黄荣华 (2002-), 男, 硕士研究生; E-mail: hhu14917@gmail.com;

任旭 (1990-), 女, 博士研究生; E-mail: 357249266@qq.com.

self-attention mechanisms, overcoming computational bottlenecks by modifying Transformer architectures (e.g., iTransformer); Meanwhile, Multi-Layer Perceptron (MLP)-based methods achieve a balance between accuracy and efficiency through minimalist linear architectures (e.g., DLinear), challenging the necessity of complex models. Moreover, this paper systematically reviews the key challenges confronting this field, including robust handling of sequences of varying lengths, the lack of interpretability in black-box models, computational costs associated with large-scale data, and insufficient cross-domain generalisation capabilities. Finally, the article outlines future research trends, indicating that constructing foundational time-series models, adopting multimodal pre-training paradigms, and enabling lightweight deployment on edge devices will constitute the primary focus for innovation in the next phase. This paper aims to provide a reference for theoretical research and engineering applications in the field of time series forecasting through a comprehensive review of existing methods and forward-looking analysis.

Keywords : time series forecasting; deep learning; Transformer; foundation mode

引言

在当今大数据快速发展的时代，时间序列数据广泛分布在各个领域，比如经济^[1,2]、交通^[3-5]、能源^[6-8]、气象^[9-11]以及医疗^[12,13]等。时间序列预测的核心目标在于通过解析历史观测数据的动态演化模式，实现对未来状态的精确推断。作为决策支持系统的核心环节，高精度预测对于优化资源分配效率、降低系统运营风险以及提升社会运行效能具有重要的理论意义与实践价值。

传统预测方法主要依赖于自回归移动平均模型与指数平滑法等统计学模型。此类方法基于严格的数学理论框架，具备良好的可解释性。然而，其建模能力受限于序列平稳性假设与线性依赖结构，在处理具有强非线性特征、高维特征空间及复杂时空相关性的大规模数据集时表现显著不足^[14]。后续发展的机器学习方法（比如支持向量回归与随机森林等）虽在非线性和关系建模方面取得一定进展，但其性能高度依赖人工特征工程，难以实现对深层次时间动力学特征的自动化表征^[15]。

近年来，随着计算能力的指数级提升与大数据技术的成熟，深度学习在特征提取与非线性拟合方面展现出显著优势，推动了时间序列预测范式的根本性变革。该领域的发展脉络可概括为：从循环神经网络（RNN）对序列依赖的建模，到卷积神经网络（CNN）对局部特征的提取，再到 Transformer 架构凭借自注意力机制在长程依赖建模中的突破性进展，深度学习模型已成为时间序列预测的主流范式。近期，基于多层感知机（MLP）的轻量化架构在预测精度与计算效率的权衡中展现出显著潜力^[16]。尽管该领域已取得显著进展，但随着预测窗口延长与数据复杂度提升，如何有效平衡模型的计算复杂度、鲁棒性及可解释性，仍是当前研究的核心挑战。此外，针对非平稳序列的动态建模、多变量间空间相关性的精准刻画以及大规模模型的泛化能力优化等问题，学术界已涌现出大量创新性研究成果^[17]。

为系统性梳理该领域的最新进展，本文对基于深度学习的时间序列预测方法进行全面综述。主要贡献如下：

- （1）系统性地构建时间序列预测的理论框架，明确方法分类体系；
- （2）深入剖析四类主流深度学习架构（RNN、CNN、Transformer、MLP）在预测任务中的发展路径，阐释代表性模型的核心思想；
- （3）概括当前时间序列预测面临的关键挑战，前瞻性地提出未来研究方向。

本文后续的结构安排如下：第2节阐述关于时间序列的基础理论与相关背景知识；第3节详细综述各类深度学习预测模型及其变体；第4节分析当前面临的挑战与技术难点；第5节总结全文并展望未来研究趋势。

一、背景知识

（一）时间序列

时间序列可定义为按时间顺序严格排列的随机变量序列，数学表示为 $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T\}$ ，其中 $\mathbf{x}_t \in \mathbb{R}^M$ 表示时刻 t 的 M 维观测向量。当 $M=1$ 时，序列称为单变量时间序列；当 $M>1$ 时，则称为多变量时间序列。

时间序列的核心特性在于其观测值间存在显著的时间依赖性，即当前观测值 \mathbf{x}_t 严格依赖于历史观测 $\{\mathbf{x}_{t-k}\}_{k=1}^t$ 。这种依赖性使时间序列区别于独立同分布数据，并在预测任务中引入关键理论概念：自相关性与平稳性。在时间序列分析中，平稳性假设具有基础性地位——一个严格平稳序列的联合分布不随时间平移而改变，其一阶矩（均值）、二阶矩（方差）及自协方差函数均与时间无关。非平稳序列通常需通过差分、对数变换等预处理操作转

化为平稳形式，以满足经典统计模型的理论前提^[14]。

时间序列的结构可分解为四个基本成分：

(1) 趋势 (Trend)：表征序列的长期非周期性变化方向（如经济指标的持续上升或下降）；

(2) 季节性 (Seasonality)：描述具有固定周期长度的重复模式（如月度销售数据的季度性波动）；

(3) 循环 (Cyclicality)：捕捉周期长度不固定的长期波动（如经济周期中的繁荣与衰退）；

(4) 随机性 (Randomness)：表征不可预测的随机波动，通常服从零均值高斯分布。

该分解框架 (STL 分解法^[18]) 为序列的特征提取与预测建模提供了理论基础。

(二) 时间序列预测任务

考虑一个包含 M 个观测变量的多变量时间序列系统。时间序列预测任务的目标是基于连续历史观测序列 $\mathbf{X}_{t-L:t} = (\mathbf{x}_{t-L+1}, \dots, \mathbf{x}_t)$ (其中 $\mathbf{x}_t \in \mathbb{R}^M$ 为时刻 t 的 M 维观测向量, $L \in \mathbb{Z}^+$ 为回溯窗口长度) 预测未来 $H \in \mathbb{Z}^+$ 个时间步的序列 $\mathbf{X}_{t+1:t+H} = (\mathbf{x}_{t+1}, \dots, \mathbf{x}_{t+H})$ 。该任务的数学表述可形式化为学习映射函数 $\mathcal{F}: \mathbb{R}^{L \times M} \rightarrow \mathbb{R}^{H \times M}$ 。

在深度学习框架下，该任务涉及以下关键概念：

(1) 回溯窗口 (Look-back Window) 与预测窗口 (Predicted Window)：回溯窗口指输入序列的长度 L ，即模型利用的连续历史观测跨度；预测窗口指输出序列的长度 H ，即目标预测的未来时间步数。

(2) 单步预测 (Single-step Forecasting) 与多步预测 (Multi-step Forecasting)：当 $H=1$ 时为单步预测；当 $H>1$ 时为多步预测。多步预测需同时估计连续 H 个未来时间点的状态。

(3) 迭代预测 (Iterative Forecasting) 与直接预测 (Direct Forecasting)：迭代预测采用自回归机制，将前一时刻的预测值作为预测下一时刻的输入，逐步生成预测序列。该方法容易导致误差累积 (error propagation)。直接预测通过构建端到端映射 $\mathcal{F}(\mathbf{X}_{t-L:t}) = \mathbf{X}_{t+1:t+H}$ ，一次性输出整个预测窗口的值，有效避免误差累积，已成为深度学习模型的主流范式。

(4) 内生变量 (Endogenous Variables) 与外生变量 (Exogenous Variables)：内生变量指目标预测序列的观测值 \mathbf{x}_t ；外生变量指影响目标序列的外部因素（如气象数据、节假日标记、政策指标等）。在预测任务中，有效整合外生变量（尤其是未来已知的外生变量）是提升预测精度的关键因素。

该任务框架为深度学习模型的设计提供了明确的输入-输出约束，其中回溯窗口 L 与预测窗口 H 的选择直接影响模型的计算复杂度与预测精度。

(三) 时间序列数据集

时间序列数据集构成了时间序列预测任务研究与评估的基石。如表所示，本文系统梳理了20个广泛应用于时间序列预测模型性能评估的公开真实数据集。该表详细列出了各数据集所属的应用领域、通道数（即变量数目）、样本总量及采样频率等关键元信息。这些数据集覆盖了能源、交通、环境、经济、医疗等多

个典型应用场景，具有较强的代表性和多样性。基于此类基准数据集开展实验，不仅有助于全面评估模型在不同场景下的预测性能，亦可有效检验其泛化能力，从而为模型在实际工程应用中的稳健性与可靠性提供实证支撑。

表：数据集统计信息

Table : Statistics of datasets

数据集名称	应用领域	通道数	样本总量	采样频率
ETTh1 ^[19]	能源	7	14,400	1小时
ETTh2 ^[19]	能源	7	14,400	1小时
ETTm1 ^[19]	能源	7	57,600	15分钟
ETTm2 ^[19]	能源	7	57,600	15分钟
Electricity ^[20]	能源	321	36,304	1小时
Solar ^[21]	能源	137	52560	10分钟
Wind ^[22]	能源	7	48673	15分钟
PEMS04 ^[23]	交通	307	16992	5分钟
PEMS07 ^[23]	交通	883	28224	5分钟
PEMS08 ^[23]	交通	170	17856	5分钟
Traffic ^[24]	交通	862	17544	1小时
Weather ^[24]	环境	21	52696	10分钟
AQWan ^[25]	环境	11	35064	1小时
AQShunyi ^[25]	环境	11	35064	1小时
Beijing Air Quality ^[26]	环境	12	35064	1小时
Exchange ^[21]	经济	8	7588	1天
NASDAQ ^[27]	经济	5	1244	1天
NYSE ^[27]	经济	5	1243	1天
ILI ^[24]	医疗	7	966	1周
Covid-19 ^[28]	医疗	848	1392	1天

(四) 时间序列预测方法概述

时间序列预测方法的发展历程可系统归纳为从经典统计学模型到机器学习模型，最终演进至深度学习模型的范式转变^[29]。基于建模机理与技术路径的差异，现有方法可划分为三类核心体系：

(1) 传统统计学方法：以自回归移动平均模型与指数平滑法为代表，此类方法建立在严格的概率论与时间序列理论基础之上。其核心假设为序列满足弱平稳性或可通过差分运算转化为平稳序列，侧重于建模线性动态关系。尽管该类模型具备良好的数学可解释性且在小规模数据集上表现稳健，但在处理非线性、高维特征空间及复杂时空相关性的大规模时序数据时，模型表达能力显著受限。

(2) 机器学习方法：随着计算资源的提升，支持向量回归、随机森林与梯度提升树等机器学习算法被引入时间序列预测领域。相较于统计模型，机器学习方法能够有效捕捉非线性特征，并通过特征工程集成多维外生变量（如滞后特征、滚动统计量）。然而，其性能高度依赖人工设计的特征提取过程，难以实现对原始序列深层时间依赖关系的自动化表征。

(3) 深度学习预测方法：近年来，深度学习凭借其卓越的非线性拟合能力与自动特征学习优势，已成为时间序列预测的主流范式。该类方法通过构建多层非线性映射函数（如深度神经网络），实现对大规模、高频次多变量时序数据的端到端学习。深度学习模型不仅能够有效捕获长程时间依赖，还可通过自注意力机制、频域分析等技术深度挖掘序列的周期性模式与趋势特征。

根据底层架构的不同，基于深度学习的时间序列预测方法

可进一步细分为基于 RNN、CNN、Transformer 以及 MLP 的方法。下文将针对这些具体的模型分类及其演进路线进行详细综述。

二、深度学习预测方法

随着大数据时代的到来与计算能力的飞跃，深度学习凭借其强大的自动特征提取能力与非线性映射机制，成为解决复杂时间序列预测问题的主流技术范式。本章节将系统梳理深度学习在时间序列预测领域的演进脉络与前沿进展。针对序列数据的不同特性与建模需求，研究界发展出了多种各具优势的模型架构。本章将重点围绕四类核心方法展开论述：一是基于 RNN 的预测方法，利用递归机制实现对历史信息的动态记忆与上下文建模；二是基于 CNN 的预测方法，通过卷积操作与感受野扩张，发挥局部特征提取与并行计算的优势；三是基于 Transformer 的预测方法，凭借自注意力机制突破长程依赖瓶颈，成为全局建模的强力工具；四是基于 MLP 的预测方法，近期通过结构重设计实现复兴，以简洁的线性架构挑战了复杂模型的必要性。通过对上述四类方法的深入分析，旨在揭示不同架构在面对长时序、多变量及不确定性预测挑战时的设计哲学、适用场景及性能差异^[30]。

（一）基于 RNN 的预测方法

基于 RNN 的预测方法是时间序列建模的经典范式。其核心优势在于通过隐状态的递归传递机制，天然适配变长序列数据，从而有效捕捉时间维度上的序列依赖关系。Jordan 于 1990 年最早提出了 RNN 的基本框架，通过引入循环连接结构，将前一时刻的隐藏状态映射为当前时刻的输入，实现了对历史信息的动态记忆与上下文建模^[31]。然而，标准 RNN 在反向传播过程中面临严重的梯度消失或爆炸问题，难以捕捉跨越长周期的深层依赖，严重制约了其在复杂长序列预测任务中的性能表现。

为突破这一瓶颈，Hochreiter 等人于 1997 年提出了长短期记忆网络（LSTM）^[32]。LSTM 引入了精巧的门控机制（遗忘门、输入门与输出门），通过可学习的参数动态调控信息流的保留与更新，显著缓解了梯度消失问题，奠定了其在电力负荷、金融市场等领域长时序预测中的基准地位。随后，Cho 等人于 2014 年提出了门控循环单元（GRU）^[33]。作为 LSTM 的轻量化变体，GRU 通过合并遗忘门与输入门为更新门，在降低参数冗余与计算开销的同时，保持了与 LSTM 相当的建模能力，成为资源受限场景下的首选架构。

随着预测任务向大规模、概率化方向演进，Salinas 等人于 2017 年提出了 DeepAR 模型^[34]。该模型基于自回归 RNN 架构，通过全局共享参数学习大量相关时间序列的共性模式，并输出预测值的概率分布（如高斯分布或负二项分布），有效支持了分位数预测与不确定性量化。针对多步分位数预测的精度问题，Wen 等人同年提出了 MQRNN（Multi-horizon Quantile Recurrent Neural Network）^[35]，通过结合序列到序列架构与分位数回归，并引入注意力机制对历史状态进行动态加权，有效规避了传统方法中的分位数交叉（Quantile Crossing）现象。

近年来，针对 Transformer 在超长序列建模中计算复杂度高的问题，RNN 架构迎来了新的复兴。Lin 等人于 2023 年提出的 SegRNN^[36] 采用了“分段循环”策略，将长序列划分为连续片段，利用底层 GRU 建模局部动力学，顶层 GRU 协调段间依赖，以线性复杂度实现了优异的预测性能。更具突破性的是 Alharthi 等人于 2024 年提出的 xLSTM^[37]。该模型对经典 LSTM 进行了现代化重构，通过引入指数型门控与矩阵内存，赋予了 RNN 类 Transformer 的并行训练能力与更强的长程建模潜力，标志着 RNN 范式在保持推理恒定内存优势的同时，重新具备了与大模型竞争的核心能力。

（二）基于 CNN 的预测方法

基于 CNN 的预测方法近年来因其高效的并行计算能力和局部特征提取优势而备受关注。通过一维卷积与扩张卷积的结合，CNN 能够在不显著增加网络深度的前提下指数级扩大感受野，从而有效捕捉长距离时间依赖^[31]。

Bai 等人于 2018 年提出的时域卷积网络（TCN）是该领域的奠基性工作之一^[38]。TCN 通过集成因果卷积、扩张卷积与残差连接，构建了一种非递归的纯卷积架构，既保证了预测过程不发生信息泄露，又实现了优于 RNN 的训练效率与梯度稳定性。在此基础上，Sen 等人于 2019 年提出的 DeepGLO^[39] 融合了全局矩阵分解与局部 TCN 网络，通过“全局正则化 + 局部时序建模”的混合策略，显著提升了高维时间序列预测的泛化能力。

为进一步挖掘序列的多尺度特征，Liu 等人于 2021 年提出了 SCINet^[40]。该模型设计了多层下采样卷积交互结构，通过在不同时间分辨率上递归提取并融合特征，实现了对复杂时序动态的精细刻画。针对时间序列中普遍存在的多周期性，Wu 等人于 2022 年提出了 TimesNet^[41]。该工作创新性地利用快速傅里叶变换（FFT）识别主要周期，将一维时间序列折叠为二维张量，进而利用 Inception 模块在二维空间中同时捕获周期内与周期间的变化规律。

随着轻量化模型需求的增加，Luo 等人于 2023 年提出了 ModernTCN^[42]。该模型借鉴现代卷积设计理念（如深度可分离卷积），通过通道混合与跨变量卷积显式建模多变量依赖，以极简的纯卷积架构在多项基准测试中超越了复杂的 Transformer 模型。最新进展方面，Stitsyuk 等人于 2024 年提出的 xPatch^[43] 则采用双流架构，结合指数平滑进行趋势 - 季节解耦，分别利用线性 MLP 与非线性 CNN 处理不同分量，并引入 Patching 机制与通道独立假设，进一步提升了模型的鲁棒性与计算效率。

（三）基于 Transformer 的预测方法

自 2017 年 Vaswani 等人提出 Transformer 架构以来，凭借其自注意力机制（Self-Attention）在全局依赖建模上的卓越表现，该架构迅速被引入时间序列预测领域^[44]。然而，原生 Transformer 存在时间复杂度 $O(L^2)$ 较高以及对时序位置信息不敏感等缺陷。为此，研究界提出了一系列针对性的改进方案。

早期的改进主要致力于降低计算复杂度。Li 等人提出的 LogTrans^[45] 通过稀疏注意力机制将复杂度降至 $O(L \log L^2)$ ，并结合因果卷积增强局部感知。Lim 等人提出的 TFT^[46]（Temporal

Fusion Transformer) 则专注于多源信息融合, 通过门控机制自适应整合静态协变量与动态输入, 并提供了良好的可解释性。Zhou 等人于2021年提出的 Informer^[19] 模型, 设计了 ProbSparse 自注意力机制与蒸馏操作, 大幅降低了内存占用, 并采用生成式解码器显著提升了长序列预测的推理速度。

随后的研究开始关注时间序列的结构特性, 特别是频域与分解思想的引入。Wu 等人提出的 Autoformer^[24] 基于经典的“趋势-季节”分解思想, 利用自相关机制代替点对点注意力, 在子序列级别进行信息聚合, 显著增强了对周期性模式的建模能力。Zhou 等人进一步提出的 FEDformer^[47] 结合傅里叶变换与小波变换, 在频域内进行低秩近似与特征提取, 实现了线性复杂度的同时也更精准地捕捉了全局分布特征。Liu 等人提出的 Pyraformer^[48] 则构建了金字塔注意力机制, 通过多尺度层级交互有效捕获不同时间分辨率下的依赖关系。

近期, Patching 机制与通道独立策略成为提升 Transformer 性能的关键。Nie 等人于2023年提出的 PatchTST^[49] 将时间序列分块 (Patch) 作为输入单元, 不仅保留了局部语义信息, 还将注意力机制的计算复杂度大幅降低, 结合通道独立策略, 在长时序预测任务上取得了显著的性能跃升。与其相对, Zhang 等人提出的 Crossformer^[50] 坚持对跨变量依赖进行显式建模, 通过两阶段注意力机制平衡了时间维度与变量维度的信息交互。

最新的研究趋势开始重新审视 Transformer 的建模范式。Liu 等人于2024年提出的 iTransformer^[51] 采用了“倒置”视角, 将整个时间序列视为一个 Token, 在变量维度上应用注意力机制, 有效解决了多变量预测中的泛化性问题。Wang 等人提出的 TimeXer^[52] 则针对外生变量场景, 设计了内生-外生解耦框架, 通过跨变量注意力机制精准量化外部因素对目标序列的因果干预。综上所述, 尽管面临线性模型与 CNN 的挑战, Transformer 架构凭借其强大的特征表示能力与灵活性, 仍是解决复杂大系统长期预测任务最具前景的技术路径之一。

(四) 基于 MLP 的预测方法

基于 MLP 的预测方法近年来因其结构简洁、计算高效以及在基准测试中的优异性能而受到广泛关注。这些方法挑战了复杂架构的必要性, 证明了线性映射结合适当的结构设计即可捕捉时间序列的复杂动态。相较于 RNN 的递归机制、CNN 的局部卷积以及 Transformer 的自注意力, MLP 架构通常实现线性或准线性时间复杂度, 同时在长时序预测任务中展现出与 SOTA (state-of-the-art) 模型相当或更优的精度。该类方法的核心思想在于通过分解、残差连接、多尺度混合或聚类机制引入时序领域的归纳偏差, 实现对趋势、季节性和非线性关系的有效建模。

Oreshkin 等人于2019年提出的 N-BEATS^[53] 是该领域的开创性工作。该模型采用深层堆叠的 MLP 块, 每块通过基函数扩展将序列分解为趋势和季节性组件, 利用残差链接逐步精炼预测。N-BEATS 有通用和可解释两种变体, 前者自由学习基函数, 后者显式使用多项式和傅里叶基。该设计提升了可解释性, 并在 M3 和 M4 竞赛数据集上超越传统统计方法。作为其扩展, Challu 等人于2022年提出的 N-HiTS^[54] 引入分层插值和多速率采样机制, 高

层块捕捉长期模式, 低层块处理短期细节, 通过插值融合多分辨率预测, 实现更低的参数量和计算成本, 在长视野任务中进一步提升准确性。

Zeng 等人于2022年提出的 DLinear^[55] 以极简线性架构挑战了 Transformer 的有效性。DLinear 将序列分解为趋势和季节组件, 通过独立的单层线性层处理并相加输出, 采用直接多步预测策略, 避免误差累积, 并在多个真实数据集上优于复杂 Transformer 模型。Das 等人于2023年提出的 TiDE^[56] (Time-series Dense Encoder) 是一种基于 MLP 的编码器-解码器模型, 专为多变量长时序预测设计。TiDE 通过特征投影压缩动态协变量, 使用密集 MLP 编码历史序列, 解码时结合未来协变量细化预测, 引入全局残差连接确保线性模型为其子类, 预测精度在基准数据集上优于 Transformer, 同时训练速度快 5-10 倍。

Wang 等人于2023年提出的 Koopa^[57], 将 Koopman 理论与 MLP 结合, 针对非平稳序列设计, 通过傅里叶滤波分解时不变和时变组件, 利用数据驱动的 Koopman 算子建模动态, 采用可堆叠块结构捕捉局部非平稳性。Wang 等人于2024年提出的 TimeMixer^[58] 聚焦于可分解的多尺度混合, 将历史序列下采样成多尺度表示, 通过 Past-Decomposable-Mixing 模块双向混合趋势和季节组件, Future-Multipredictor-Mixing 融合多预测器输出, 并通过消融实验验证了每个模块的重要性。

最新进展中, Liu 等人于2024年提出的 DUET^[59] 引入双聚类增强框架, 在时间维度使用分布路由器聚类异质模式, 在通道维度采用频率域软聚类建模相关性, 通过掩码注意力融合两者, 该通用框架在 25 个数据集上实现 SOTA 性能, 提高了多变量预测的泛化能力。综上, 基于 MLP 的方法正从简单线性映射向结构化、多模态方向演进, 其轻量化特性使其在边缘部署中极具潜力, 并已成为时间序列预测领域的关键创新路径。

三、面临的挑战

尽管基于深度学习的时间序列预测方法在多个基准数据集上取得了显著进展, 但在面对现实工业场景中复杂、动态且极具挑战性的数据环境时, 仍面临若干亟待解决的关键问题。

(一) 不等长序列的处理

在电力监测、医疗健康以及物联网传感器数据等领域, 由于采样频率不一致、设备故障导致的数据丢失及不同观测对象生命周期差异等因素的影响, 数据通常表现为变长、非对齐或极度稀疏。当前多数深度学习模型假定输入具有固定的序列长度和均匀采样间隔。然而, 对于不等长序列的有效处理, 目前尚缺乏系统而高效的方法。

(二) 缺乏可解释性

深度神经网络常被视为“黑箱”模型, 其内部复杂的非线性映射机制使得预测结果难以被直观理解。特别是在医疗诊断、金融风险等高敏感度领域, 模型的可解释性与预测精度同样重要。尽管注意力机制提供了一定程度的权重可视化, 但其本质上反映的是数据间的相关性而非因果关系, 这限制了对其逻辑进行

深入分析的可能性。因此，如何将领域知识（如物理定律、经济学原理）融入深度模型，或开发针对时间序列特征的因果归因技术，是提升模型可信度的重要方向。

（三）计算成本与效率

随着预测窗口的延长及变量维度的增长，模型的计算开销呈指数级上升。特别是对于采用自注意力机制的 Transformer 架构，其计算复杂度高达 $O(L^2)$ 。虽然近年来基于 MLP 的轻量化模型层出不穷，但在资源受限的边缘计算环境中，如何在有限的计算能力和能耗约束下实现高精度实时预测，依然是工业界关注的核心问题之一。探索更高效的模型压缩、量化策略以及开发线性复杂度的架构设计，依然具有广阔的前景。

（四）通用模型的构建与迁移

现有的时间序列预测模型大多遵循“特定任务、特定训练”的范式，即针对某一具体应用场景进行离线训练后，难以直接应用于其他领域。鉴于不同领域（例如交通与金融）中的时间序列在周期性、平稳性及噪声分布等方面的显著差异，构建一个能够跨越多个领域和尺度的通用时间序列预测模型极具挑战性。研究具备强大迁移学习能力和域适应能力的模型架构，是实现“一次训练，多处适用”目标的关键所在。

（五）大规模模型的应用

受大规模语言模型（Large Language Models, LLMs^[60]）成功的启发，时间序列基础模型（Time Series Foundation Models, TSFM）已成为新兴的研究前沿。然而，与文本数据相比，时间序列缺乏自然的“词汇表”结构，且数值范围广泛、语义信息密度较低。当前面临的挑战包括：如何设计有效的 Tokenization 方法来处理连续数值；如何构建大规模的多模态预训练时序数据集；以及如何利用 LLMs 的强大推理能力捕捉超长时间序列中的内在逻辑联系。大模型在零样本预测方面的潜力挖掘刚刚起步，未来的研究需要更加深入地探索这些方面。

四、总结与展望

本文系统综述了基于深度学习的时间序列预测方法的研究进展，从理论框架、主流架构演进及关键技术挑战三个维度进行了

深入剖析。随着大数据技术的普及与计算能力的提升，深度学习已成功超越传统统计模型与机器学习方法，成为处理复杂、高维、非线性时间序列数据的主流范式。通过对 RNN、CNN、Transformer 及 MLP 四类核心架构的梳理，可以发现技术演进呈现出鲜明的逻辑脉络：

（1）RNN 类方法经历了从标准循环结构到门控机制（LSTM/GRU），再到近期 SegRNN、xLSTM 等现代化重构的历程，证明了递归归纳偏置在处理序列依赖中仍具独特优势；

（2）CNN 类方法利用扩张卷积与多尺度架构（如 TimesNet、ModernTCN），在提取局部特征与并行计算效率方面展现出卓越性能；

（3）Transformer 类方法凭借自注意力机制在长程依赖建模中占据主导地位，并通过 Patching 机制、频域注意力及通道独立策略（如 PatchTST、iTransformer）克服计算瓶颈。

（4）MLP 类方法则以 N-BEATS、DLinear 为代表，通过极简的线性映射与结构重参数化，向复杂的深度架构发起了有力挑战，重新定义了预测精度与计算成本的平衡点。

尽管上述方法在公开基准数据集上取得了显著成果，面向真实工业场景的应用仍面临多重挑战：不等长与非对齐序列的鲁棒处理、模型可解释性与因果推理能力的缺失、高维长窗口下的计算效率瓶颈、跨领域泛化能力不足，以及大规模基础模型在时序语义建模中的适配难题。这些问题亟需通过跨学科融合（如结合信号处理、动力系统理论、因果推断与表示学习）加以突破。展望未来，时间序列预测研究将朝着以下方向深化发展：

- （1）构建兼具高效性、可解释性与物理一致性的混合架构；
- （2）发展面向非平稳、异步与稀疏观测的通用建模框架；
- （3）推动时间序列基础模型的预训练范式创新，探索连续数值 Tokenization、多模态对齐与零样本迁移机制；
- （4）强化模型在边缘设备上的部署能力，实现“精度 - 效率 - 能耗”三重优化。

随着算法、算力与数据生态的协同发展，深度学习驱动的时间序列预测有望在智慧城市、智能制造、精准医疗与气候模拟等关键领域发挥更深远的支撑作用。

参考文献

- [1]Ahmar A S, Boj Del Val E. Novel time series methods in economic forecasting: SutteARIMA evidence from Indonesian Consumer Price Index and currency exchange rates[J]. Cogent Economics & Finance, 2025, 13(1): 2566223.
- [2]Kahra D, Kahra V. VSF-Net: A variational encoder and Sharpness-Aware attention framework for robust forecasting of non-stationary financial time series[J]. International Journal of Information Technology, 2025.
- [3]曾俊斌, 杨建森, 孙璇. RD-RNN: 用于交通流量预测的残差分解循环神经网络 [J]. 控制工程, 2025, 32(12): 2207-2213.
- [4]Bielecki P, Hachaj T, Waś J. Sample and Aggregate Voronoi Neighborhood Weighted Graph Neural Network (SAGE-Voronoi) and Its Capability for City-Sized Vehicle Traffic Time Series Prediction[J]. Applied Sciences, 2025, 15(24): 12899.
- [5]Liu R, Shin S Y. A Review of Traffic Flow Prediction Methods in Intelligent Transportation System Construction[J]. Applied Sciences, 2025, 15(7): 3866.
- [6]Park Y J, Germain F, Liu J, et al. Probabilistic forecasting for building energy systems using time-series foundation models[J]. Energy and Buildings, 2025: 116446.

- [7]Zhang X, Song M, Xiao X, et al. EXP-Transformer time series prediction model for accident scenarios in high-reliability energy systems: Nuclear power plants case[J]. *Energy*, 2025, 322: 135481.
- [8]Zhang F, Leng Z, Chen L, et al. Joint Probabilistic Forecasting of Wind and Solar Power Exploiting Spatiotemporal Complementarity[J]. *Sustainability*, 2025, 17(8): 3584.
- [9]Sithole P P, Obagbuwa I C. Rainfall Prediction in South Africa Using Time - Series Analysis[J]. *Applied Computational Intelligence and Soft Computing*, 2025, 2025(1): 5599720.
- [10] 薛院红, 宋文广, 江琼琴, 等. VMD-LSTM-FEDformer 降水预测融合模型研究 [J]. *节水灌溉*, 2025.
- [11]Waqas M, Humphries U W, Chueasa B, et al. Artificial intelligence and numerical weather prediction models: A technical survey[J]. *Natural Hazards Research*, 2025, 5(2): 306-320.
- [12] 田剑, 刘全智, 张瑜, 等. 通用时间序列预测大模型在医学领域应用研究 [J]. *现代医院*, 2025, 25(12): 1923-1928.
- [13]Ahmed M, Hasan M, Kumar S, et al. Trends and Forecast of Heart Failure-Related Deaths in US Chronic Obstructive Pulmonary Disease Patients (1999 - 2030): Insights From Advanced Time-Series Modeling[J]. *Cardiology in Review*, 2025.
- [14]Box G E, Jenkins G M, Reinsel G C, et al. *Time series analysis: forecasting and control*[M]. John Wiley & Sons, 2015.
- [15]Zhou Z H. *Machine learning*[M]. Springer nature, 2021.
- [16]Torres J F, Hadjout D, Sebaa A, et al. Deep Learning for Time Series Forecasting: A Survey[J]. *Big Data*, 2021, 9(1): 3-21.
- [17]Wang Y, Wu H, Dong J, et al. Deep time series models: A comprehensive survey and benchmark[J]. *arXiv preprint arXiv:2407.13278*, 2024.
- [18]Cleveland R B, Cleveland W S, McRae J E, et al. STL: A seasonal-trend decomposition[J]. *J. off. Stat.*, 1990, 6(1): 3-73.
- [19]Zhou H, Zhang S, Peng J, et al. Informer: Beyond efficient transformer for long sequence time-series forecasting[C]//*Proceedings of the AAAI conference on artificial intelligence*: 35. 2021: 11106-11115.
- [20]Trindade A. *ElectricityLoadDiagrams20112014*[DS/OL]. UCI Machine Learning Repository, 2015. <https://archive.ics.uci.edu/dataset/321>.
- [21]Lai G, Chang W C, Yang Y, et al. Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks[C]//*The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. Ann Arbor MI USA: ACM, 2018: 95-104.
- [22]Li Y, Lu X, Wang Y, et al. Generative time series forecasting with diffusion, denoise, and disentanglement[J]. *Advances in Neural Information Processing Systems*, 2022, 35: 23009-23022.
- [23]Song C, Lin Y, Guo S, et al. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting[C]//*Proceedings of the AAAI conference on artificial intelligence*: 34. 2020: 914-921.
- [24]Wu H, Xu J, Wang J, et al. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting[J]. *Advances in neural information processing systems*, 2021, 34: 22419-22430.
- [25]Zhang S, Guo B, Dong A, et al. Cautionary tales on air-quality improvement in Beijing[J]. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2017, 473(2205): 20170457.
- [26]Chen S. *Beijing Multi-Site Air Quality*[DS/OL]. UCI Machine Learning Repository, 2017. <https://archive.ics.uci.edu/dataset/501>.
- [27]Feng F, He X, Wang X, et al. Temporal Relational Ranking for Stock Prediction[J]. *ACM Transactions on Information Systems*, 2019, 37(2): 1-30.
- [28]Panagopoulos G, Nikolentzos G, Vazirgiannis M. Transfer graph neural networks for pandemic forecasting[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*: 35. 2021: 4838-4845.
- [29]Cheng M, Liu Z, Tao X, et al. A comprehensive survey of time series forecasting: Concepts, challenges, and future directions[J]. *Authorea Preprints*, 2025.
- [30]LeCun Y, Bengio Y, Hinton G. Deep learning[J]. *nature*, 2015, 521(7553): 436-444.
- [31]Kim K G. Book Review: Deep Learning[J]. *Healthcare Informatics Research*, 2016, 22(4): 351.
- [32]Hochreiter S, Schmidhuber J. Long short-term memory[J]. *Neural computation*, 1997, 9(8): 1735-1780.
- [33]Cho K, Merriënboer B van, Gulcehre C, et al. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation[A]. *arXiv*, 2014.
- [34]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.
- [35]en R, Torkkola K, Narayanaswamy B, et al. A Multi-Horizon Quantile Recurrent Forecaster[A]. *arXiv*, 2018.
- [36]Lin S, Lin W, Wu W, et al. SegRNN: Segment Recurrent Neural Network for Long-Term Time Series Forecasting[A]. *arXiv*, 2023.
- [37]Alharthi M, Mahmood A. xlstmtime: Long-term time series forecasting with xlstm[J]. *AI*, 2024, 5(3): 1482-1495.
- [38]Bai S. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling[J]. *arXiv preprint arXiv:1803.01271*, 2018.
- [39]Sen R, Yu H F, Dhillon I S. Think globally, act locally: A deep neural network approach to high-dimensional time series forecasting[J]. *Advances in neural information processing systems*, 2019, 32.
- [40]Liu M, Zeng A, Chen M, et al. Scinet: Time series modeling and forecasting with sample convolution and interaction[J]. *Advances in Neural Information Processing Systems*, 2022, 35: 5816-5828.
- [41]Wu H, Hu T, Liu Y, et al. TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis[A]. *arXiv*, 2023.
- [42]Luo D, Wang X. ModernTCN: A modern pure convolution structure for general time series analysis[C]//*The twelfth international conference on learning representations*. 2024: 1-43.
- [43]Stitsyuk A, Choi J. xPatch: Dual-Stream Time Series Forecasting with Exponential Seasonal-Trend Decomposition[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*: 39. 2025: 20601-20609.
- [44]Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. *Advances in neural information processing systems*, 2017, 30.
- [45]Li S, Jin X, Xuan Y, et al. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting[J]. *Advances in neural information processing systems*, 2019, 32.
- [46]Lim B, Arik S O, Loeff N, et al. Temporal fusion transformers for interpretable multi-horizon time series forecasting[J]. *International journal of forecasting*, 2021, 37(4): 1748-1764.

- [47]Zhou T, Ma Z, Wen Q, et al. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting[C]//International conference on machine learning. PMLR, 2022: 27268–27286.
- [48]Liu S, Yu H, Liao C, et al. Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting[C]//# PLACEHOLDER_PARENT_METADATA_VALUE#. 2022.
- [49]Nie Y, Nguyen N H, Sinthong P, et al. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers[A]. arXiv, 2023.
- [50]Zhang Y, Yan J. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting[C]//The eleventh international conference on learning representations. 2023.
- [51]Liu Y, Hu T, Zhang H, et al. iTransformer: Inverted Transformers Are Effective for Time Series Forecasting[A]. arXiv, 2024.
- [52]Wang Y, Wu H, Dong J, et al. Timexer: Empowering transformers for time series forecasting with exogenous variables[J]. Advances in Neural Information Processing Systems, 2024, 37: 469–498.
- [53]Oreshkin B N, Carpov D, Chapados N, et al. N-BEATS: Neural basis expansion analysis for interpretable time series forecasting[A]. arXiv, 2020.
- [54]Challu C, Olivares K G, Oreshkin B N, et al. Nhits: Neural hierarchical interpolation for time series forecasting[C]//Proceedings of the AAAI conference on artificial intelligence: 37. 2023: 6989–6997.
- [55]Zeng A, Chen M, Zhang L, et al. Are transformers effective for time series forecasting?[C]//Proceedings of the AAAI conference on artificial intelligence: 37. 2023: 11121–11128.
- [56]Das A, Kong W, Leach A, et al. Long-term Forecasting with TiDE: Time-series Dense Encoder[A]. arXiv, 2024.
- [57]Liu Y, Li C, Wang J, et al. Koopa: Learning non-stationary time series dynamics with koopman predictors[J]. Advances in neural information processing systems, 2023, 36: 12271–12290.
- [58]Wang S, Wu H, Shi X, et al. TimeMixer: Decomposable Multiscale Mixing for Time Series Forecasting[A]. arXiv, 2024.
- [59]Qiu X, Wu X, Lin Y, et al. DUET: Dual Clustering Enhanced Multivariate Time Series Forecasting[C]//Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.1. Toronto ON Canada: ACM, 2025: 1185–1196.
- [60]Minaee S, Mikolov T, Nikzad N, et al. Large Language Models: A Survey[A]. arXiv, 2025.