

DPANET: 双金字塔注意力网络用于多变量时间序列预测

Qianyang Li^{*1}, Xingjun Zhang^{*1}, Shaoxun Wang¹, Wei Jia²

¹School of Computer Science and Technology, Xi'an Jiaotong University, Xi'an, China

²Department of Computer Science and Technology, Tsinghua University, Beijing, China

{liqianyang, shaoxunwang}@stu.xjtu.edu.cn, xjzhang@xjtu.edu.cn, weijia4473@mail.tsinghua.edu.cn

ABSTRACT

长期时间序列预测 (LTSF) 受制于建模跨越多个时间尺度和频率分辨率的复杂依赖关系的挑战。现有方法, 包括 Transformer 和基于 MLP 的模型, 通常难以以统一且结构化的方式捕捉这些交织的特点。我们提出了双金字塔注意力网络 (DPANet), 这是一种新型架构, 明确地解耦并同时建模时间多尺度动态和频谱多分辨率周期性。DPANet 构建了两个平行的金字塔: 一个基于逐步下采样的时间金字塔, 另一个基于带通滤波的频率金字塔。我们的模型核心是跨金字塔融合块 (Cross-Pyramid Fusion Block), 它通过交叉注意力促进对应金字塔层级之间的深度交互信息交换。这种融合按照从粗到细的层次进行, 使全局上下文能够指导局部表示学习。在公共基准上的广泛实验表明, DPANet 实现了最先进的性能, 显著优于先前模型。代码可在 <https://github.com/hit636/DPANet> 获得。

Index Terms— 时间序列预测, 注意力机制, 多尺度学习, 金字塔网络。

1. 介绍

时间序列预测在众多实际应用中是一项关键任务, 其广泛的实际效用和复杂的理论问题推动了这一领域的发展 [1]。该领域对多个关键部门产生了重要影响, 包括金融市场分析 [2]、能源需求管理 [3]、气象预报 [4] 以及交通系统优化 [5]。尽管深度学习的最新进

展, 特别是基于 Transformer 和 MLP 架构的发展 [6] 显著提高了性能, 但仍存在一个基本挑战: 有效建模时间序列数据中存在的层次结构。

真实世界的时间序列是由多尺度的时序模式 (例如, 日趋势与年趋势) 和多分辨率的频率成分 (例如, 高频波动与低频季节性) 组成的。这些属性是正交的但又紧密耦合; 例如, 长期的季节趋势会调制高频的日模式。然而, 现有的模型如 PatchTST [7] 虽然擅长处理长程依赖关系, 但却缺乏这种复杂多分辨率分析的明确结构。

本文认为, 解决长期时间序列预测 (LTSF) 的挑战需要一个能够反映时间序列数据固有多级结构的模型架构。借鉴信号和图像处理中的金字塔网络 [8], 并在多尺度融合模型的最新进展 [9] 的基础上, 我们提出了一种分解时域和频域层次并通过交互式融合集成它们的方法。我们引入了双金字塔注意力网络 (DPANet), 其采用双流架构, 包含两个金字塔: 一个通过下采样捕捉多尺度分层特征的时序金字塔, 以及一个使用带通滤波分离周期性模式的频率金字塔。核心创新是跨金字塔融合块, 通过交叉注意力实现对应金字塔层级之间的深度信息交换。

我们的贡献有三个方面: (1) 我们提出了一种新颖的双金字塔架构, 用于并行的时间和频率域分析。(2) 我们设计了一种跨金字塔融合机制, 以实现深层次的特征交互。(3) 我们实证建立了 DPANet 作为长期序列预测新强基线的优越性。

^{*}Corresponding author: Qianyang Li and Xingjun Zhang. This work is partially supported by National Natural Science Foundation of China 62372366.

2. 相关工作

2.1. 基于变压器的时序预测

基于 Transformer 的模型通过利用注意力机制来捕捉长距离依赖性，达到了最先进的水平。创新集中在提高效率和将该机制适应于时间数据上。值得注意的例子包括使用 ProbSparse 注意力的 Informer[10]、带有自相关机制的 Autoformer[11] 和在频域中操作的 FEDformer[12]。最近，受视觉 Transformer (ViT) 启发的 PatchTST[7] 通过分块处理时间序列展示了巨大的成功。尽管这些方法非常强大，但它们通常仅在一个领域内运行（无论是时间还是频域）。DPANet 通过并行显式建模这两个领域，并促进它们之间的深层次、层次化交互而区别开来。

2.2. 基于 MLP 的时间序列预测

近期的研究表明，轻量级的基于 MLP 的模型可以出奇地有效。N-BEATS[13] 使用了一种带有基函数扩展原理的深层 MLP 块堆栈。DLinear[6] 进一步简化了这一点，表明对分解的趋势和季节性应用单一线性层可以超越复杂的 Transformer。沿这一趋势，TSMixer[14] 等模型已成功将 MLP-Mixer 架构适应于高效的时间和通道特征提取。DPANet 从这一范式中学习，通过使用高效的计算块，但将其嵌入到结构化的分层框架内以明确捕捉简单的 MLPs 忽略的多尺度和多分辨率特性。

2.3. 多尺度与多分辨率方法

多尺度和多分辨率分析是时间序列的经典方法，已被整合进各种深度学习模型中。Pyraformer[15] 使用金字塔注意力结构来减少复杂性，而 TimesNet[16] 则根据通过 FFT 发现的周期性将 1D 系列重塑为 2D 张量。最近如 TimeMixer[17] 的混合模型利用了多尺度分解。DPANet 通过构建用于时间域和频谱域的两个不同且并行的金字塔，推进了这一领域的研究，并引入了一种新颖机制以实现它们的交互融合，从而实现了比先前工作更加全面和原则性的分层分析。

3. 方法

3.1. DPANet 架构

给定一个历史多元时间序列 $\mathbf{X} \in \mathbb{R}^{L_{in} \times C}$ ，其中 L_{in} 是回看窗口长度且 C 是变量数，我们的目标是预测未来序列 $\mathbf{Y} \in \mathbb{R}^{L_{pred} \times C}$ 的长度为 L_{pred} 。

如图 1 所示，DPANet 架构系统地解耦和融合多尺度时间和多分辨率信息。输入首先通过一个 RevIN 层 [18] 进行归一化。模型的核心是一系列新颖的**交叉金字塔融合块**操作，这些操作在并行的时间和频率金字塔上进行。融合遵循从粗到细的分层原则，其中高层次语义指导低层次特征学习。最终预测头将融合后的表示映射到预报，然后通过逆 RevIN 变换进行反归一化。

3.2. 双重金字塔表示法

为了同时捕捉不同时间粒度和频率带的动力学特性，我们从归一化的输入 $\mathbf{X}'$ 构建两个并行的特征金字塔。

时间金字塔: $\mathcal{P}_t = \{\mathbf{T}^{(0)}, \dots, \mathbf{T}^{(S-1)}\}$ ，是通过将输入序列 $\mathbf{X}'$ 在时间域内逐步降采样创建的。令 $\mathbf{T}^{(0)} = \mathbf{X}'$ 。每一后续级别通过以下方式获得：

$$\mathbf{T}^{(s)} = \text{AvgPool1d}(\mathbf{T}^{(s-1)}) \in \mathbb{R}^{(L_{in}/2^s) \times C} \quad (1)$$

其中 $s \in \{1, \dots, S-1\}$ 是尺度级别。该金字塔有效地捕捉了从细粒度细节到粗粒度趋势的序列结构。

频率金字塔: 频率金字塔 $\mathcal{P}_f = \{\mathbf{F}^{(0)}, \dots, \mathbf{F}^{(S-1)}\}$ 是通过将输入序列分解为不同的频率分辨率创建的。我们首先计算输入 $\mathcal{F} = \text{RFFT}(\mathbf{X}')$ 的实数傅里叶变换。然后将频谱划分成 S 个不相交的、对数间隔的带。这种对数分区至关重要：它为低频分量分配更精细的分辨率，这些分量通常对应于主导的长期季节性变化，而将高频噪声分组得更为粗糙。

对于每个频带 s ，我们创建一个二进制掩码 $M^{(s)}$ 来隔离其频率范围。对应的金字塔级别 $\mathbf{F}^{(s)}$ 通过应用此掩码并将结果逆 RFFT 变换回时域进行重建：

$$\mathbf{F}^{(s)} = \text{IRFFT}(\mathcal{F} \odot M^{(s)}) \in \mathbb{R}^{L_{in} \times C} \quad (2)$$

其中 $\odot$ 表示逐元素乘法。每个 $\mathbf{F}^{(s)}$ 代表原始序列的特

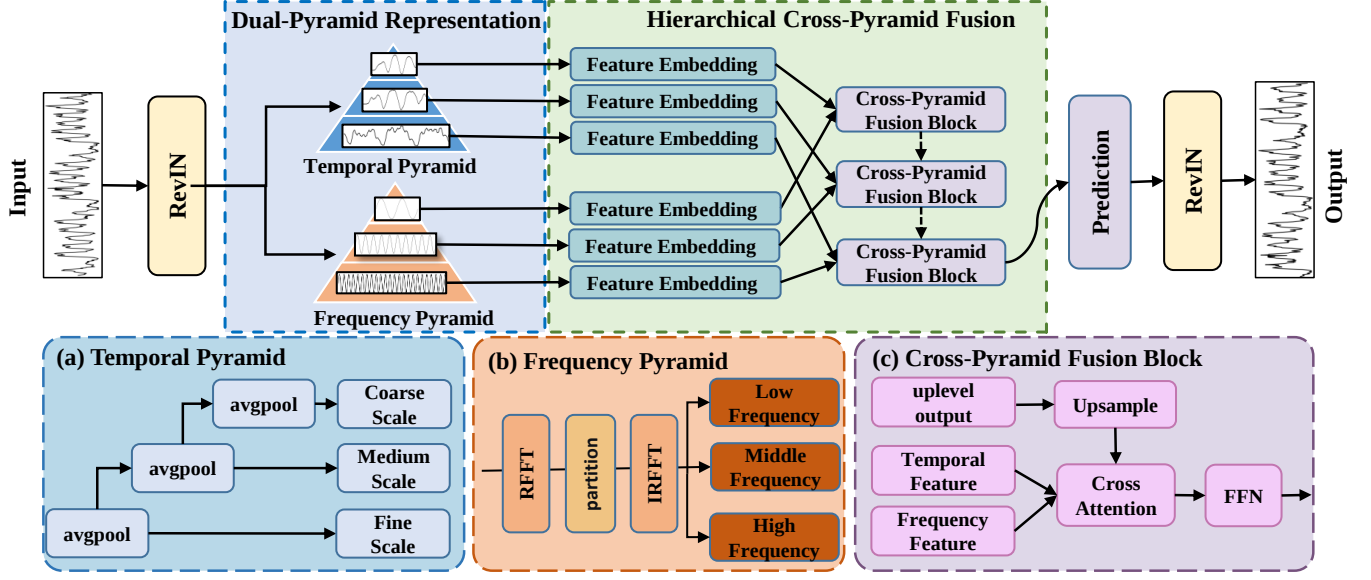


Fig. 1. DPANet 的整体架构。该模型将时间序列分解为并行的 (a) 时间金字塔和 (b) 频率金字塔。这些表示然后通过一组 (c) 跨金字塔融合块以从粗到细的层次结构进行融合。

定周期分量，有效地隔离不同的季节性模式以供融合阶段使用。

3.3. 分层交叉金字塔融合

DPANet 的核心是分层融合过程，该过程从最粗的尺度操作到最细的尺度。在每个尺度 s 上，采用交叉金字塔融合块来促进时间表示和频率表示之间的深层交互。

特征嵌入：在融合之前，每个金字塔的表示 $\mathbf{T}^{(s)}$ 和 $\mathbf{F}^{(s)}$ 首先被重塑以将每个通道视为一个独立实例，然后通过尺度特定的线性层嵌入到 d_{model} 维特征空间中，生成初始隐藏状态 $\mathbf{h}_t^{(s)}$ 和 $\mathbf{h}_f^{(s)}$ 。这种重塑将每个通道的时间序列视为初始嵌入中的独立实例，而它们之间的关系则由交叉注意力机制稍后捕获。

交叉金字塔融合块：这个模块实现了两个域之间的双向信息流动。令 $\mathbf{h}_t$ 和 $\mathbf{h}_f$ 为当前尺度的输入表示 $\mathbf{h}_t^{(s)}$ 和 $\mathbf{h}_f^{(s)}$ 。首先，两个并行的交叉注意力层丰富了这些表示：

$$\mathbf{h}_t' = \text{LayerNorm}(\mathbf{h}_t + \text{CrossAttn}(\mathbf{h}_t, \mathbf{h}_f, \mathbf{h}_f)) \quad (3)$$

$$\mathbf{h}_f' = \text{LayerNorm}(\mathbf{h}_f + \text{CrossAttn}(\mathbf{h}_f, \mathbf{h}_t, \mathbf{h}_t)) \quad (4)$$

其中 $\text{CrossAttn}(Q, K, V)$ 表示一个标准的多头注意力

模块。这两种丰富的表示随后被拼接，并通过前馈网络 (FFN) 进行深度融合。输出最终被分解回更新后的时域和频域表示， $\tilde{\mathbf{h}}_t^{(s)}$ 和 $\tilde{\mathbf{h}}_f^{(s)}$ 。

融合过程从最粗的尺度 ($s = S - 1$) 开始，逐步进行到最精细的 ($s = 0$)。对于每个更精细的尺度 s ，从前一个较粗糙尺度融合输出的 $\tilde{\mathbf{h}}_t^{(s+1)}$ 和 $\tilde{\mathbf{h}}_f^{(s+1)}$ 通过线性插值上采样以匹配当前序列长度。然后将它们作为残差连接添加到初始嵌入 $\mathbf{h}_t^{(s)}$ 和 $\mathbf{h}_f^{(s)}$ 中，形成当前融合块的输入：

$$\mathbf{h}_t^{(s)'} = \mathbf{h}_t^{(s)} + \text{UpSample}(\tilde{\mathbf{h}}_t^{(s+1)}) \quad (5)$$

$$\mathbf{h}_f^{(s)'} = \mathbf{h}_f^{(s)} + \text{UpSample}(\tilde{\mathbf{h}}_f^{(s+1)}) \quad (6)$$

其中 $\text{UpSample}(\cdot)$ 将序列长度翻倍。这种从粗到细的机制允许抽象的全局信息指导更精细层次上的表示学习。

3.4. 预测

最终预测是由最细粒度 ($s = 0$) 的时间表示 $\hat{\mathbf{h}}^{(0)}$ 生成的，其中包含了完全整合的多尺度和多分辨率信息。一个线性预测头首先将序列池化表示 $\hat{\mathbf{h}}^{(0)}$ 投影到所需的预测长度 L_{pred} ，得到归一化空间中的原始预

测。最后，通过应用存储的输入统计信息的逆 RevIN 变换，将该原始预测反归一化为最终输出。

4. 实验

4.1. 数据集和基线参考

我们在 8 个广泛使用的 LTSF 基准上进行了实验：ETT (ETTh1, ETTh2, ETTm1, ETTm2)、Weather、Electricity 和 Traffic。我们将 DPANet 与一组全面的 SOTA 基线进行了比较，包括基于 Transformer 的模型 (PatchTST, TimeXer)、基于 MLP 的模型 (DLinear, TSMixer)、多尺度模型 (TimeMixer, MSD-Mixer, TimesNet) 和其他基于金字塔的模型 (Pyraformer)。

4.2. 实验设置

我们遵循了标准的评估协议，所有模型均使用固定输入序列长度 $L = 96$ 在四个预测时域 $T \in \{96, 192, 336, 720\}$ 上进行了评估。所有模型都在相同的条件下进行训练，并公平地调整超参数。均方误差 (MSE) 和平均绝对误差 (MAE) 被用作评估指标。

4.3. 主要结果

如表 1 所示，我们的模型在七个基准测试中始终表现出优异的性能。DPANet 在总共 28 种设置中的 17 个 MSE 和 19 个 MAE 指标上取得了最佳结果。其优势尤其体现在具有复杂模式的数据集（如 Electricity 和 ETTh1）上，在这些数据集中，DPANet 在所有预测范围内的 MSE 和 MAE 都位居榜首。此外，在高度波动的 Traffic 数据集上，DPANet 通过在最具挑战性的长期预测（预测范围 720）中超越强大的竞争对手证明了其稳健性。这些结果验证了我们双金字塔架构的有效性。

4.4. 消融研究

我们进行了严格的消融研究以验证 DPANet 的关键组件（表 2）。完整模型始终优于所有变体。为了测试我们的双域假设，我们设计了两个专用版本：一个仅时间模型（融合两个相同的时间金字塔）和一个仅频率模型（融合两个频谱金字塔）。这两种变体表现

显著不佳，证实了异构时间和频率信息的融合至关重要。此外，将交叉注意力机制替换为更简单的方法（无 Cross-Fusion）导致最严重的性能下降。这一结果强调了我们的交互式融合块是最关键组件。

5. 结论

在本文中，我们介绍了双金字塔注意力网络 (DPANet)，以解决长时序预测 (LTSF) 中交织的多尺度和多分辨率模式建模挑战。通过构建并行的时间和频率金字塔，并通过一种原理上的、从粗到细的交叉注意力机制将它们融合，DPANet 实现了对复杂时间序列动态的更全面和结构化的表示。广泛的实验结果表明，DPANet 在多个基准测试中表现出优越性能。

Table 1. 长期预测任务的综合结果。最佳结果在**粗体**中突出显示，而第二好的结果是 下划线的。

模型	DPANet		时间混合器		MSD-混频器		时间X器		补丁TST		时间网		D线性		FEDformer		金字塔形器	
度量	均方误差	平均绝对误差	均方误差	平均绝对误差	均方误差	平均绝对误差	均方误差	平均绝对误差	均方误差	平均绝对误差	均方误差	平均绝对误差	均方误差	均方误差	均方误差	平均绝对误差	均方误差	均方误差
ETTm1	96	0.371	<u>0.398</u>	<u>0.375</u>	0.400	0.377	0.391	0.382	0.403	0.460	0.447	0.384	0.402	0.407	0.412	0.395	0.424	0.664
	192	0.420	0.419	0.429	<u>0.421</u>	<u>0.427</u>	0.422	0.429	0.453	0.477	0.429	0.436	0.429	0.446	0.441	0.469	0.470	0.790
	336	0.441	0.441	<u>0.458</u>	0.448	0.469	<u>0.443</u>	0.468	0.448	0.546	0.496	0.491	0.469	0.489	0.467	0.547	0.495	0.891
	720	0.461	0.458	0.498	0.482	0.485	0.475	<u>0.469</u>	<u>0.461</u>	0.544	0.517	0.521	0.500	0.513	0.510	0.598	0.544	0.963
ETTm2	96	0.292	0.340	<u>0.289</u>	<u>0.342</u>	0.284	0.345	0.308	0.355	0.745	0.584	0.340	0.374	0.340	0.394	0.358	0.397	0.645
	192	0.374	0.391	0.378	0.397	0.362	<u>0.392</u>	<u>0.363</u>	0.389	0.393	0.405	0.402	0.414	0.482	0.479	0.429	0.439	0.788
	336	0.383	0.411	<u>0.386</u>	<u>0.414</u>	0.399	0.428	0.414	0.423	0.427	0.436	0.452	0.452	0.591	0.541	0.496	0.487	0.907
	720	0.411	0.430	<u>0.412</u>	0.434	0.426	0.457	0.414	<u>0.432</u>	0.436	0.450	0.462	0.468	0.839	0.661	0.463	0.474	0.963
ETTh1	96	<u>0.319</u>	0.349	0.320	0.357	0.304	<u>0.351</u>	0.318	0.356	0.352	0.374	0.338	0.375	0.346	0.374	0.379	0.419	0.543
	192	0.343	<u>0.378</u>	0.361	0.381	<u>0.344</u>	0.375	0.362	0.383	0.374	0.387	0.374	0.387	0.382	0.391	0.389	0.387	0.557
	336	0.379	0.393	0.390	0.404	<u>0.380</u>	<u>0.395</u>	0.395	0.407	0.421	0.414	0.410	0.411	0.415	0.415	0.445	0.459	0.754
	720	<u>0.431</u>	0.422	0.454	0.441	0.427	<u>0.428</u>	0.452	0.441	0.462	0.449	0.478	0.450	0.473	0.451	0.543	0.490	0.908
ETTh2	96	<u>0.173</u>	0.255	0.175	0.258	0.169	0.259	0.171	<u>0.256</u>	0.183	0.270	0.187	0.267	0.193	0.293	0.203	0.287	0.435
	192	0.233	0.297	0.237	<u>0.299</u>	<u>0.232</u>	0.300	0.237	0.299	0.255	0.314	0.249	0.309	0.284	0.361	0.269	0.328	0.730
	336	0.291	0.335	0.298	0.340	<u>0.292</u>	<u>0.337</u>	0.296	0.338	0.309	0.347	0.321	0.351	0.382	0.429	0.325	0.366	1.201
	720	0.390	<u>0.395</u>	<u>0.391</u>	0.396	0.392	0.398	0.392	0.394	0.412	0.404	0.408	0.403	0.558	0.525	0.421	0.415	3.345
天气	96	0.168	<u>0.211</u>	0.202	0.261	0.148	0.212	<u>0.157</u>	0.205	0.186	0.227	0.172	0.220	0.195	0.252	0.217	0.296	0.622
	192	<u>0.207</u>	0.246	0.208	<u>0.250</u>	0.200	0.262	0.204	<u>0.247</u>	0.234	0.265	0.219	0.261	0.237	0.295	0.276	0.336	0.739
	336	0.250	<u>0.285</u>	<u>0.251</u>	0.278	0.256	0.310	0.261	0.290	0.284	0.301	0.280	0.306	0.282	0.331	0.339	0.380	1.004
	720	0.338	0.339	<u>0.339</u>	<u>0.341</u>	0.327	0.362	0.340	0.341	0.356	0.349	0.365	0.359	0.359	0.345	0.403	0.428	1.420
电力	96	0.152	0.245	<u>0.153</u>	<u>0.247</u>	0.161	0.253	0.182	0.278	0.190	0.296	0.168	0.272	0.210	0.305	0.169	0.273	0.386
	192	0.161	<u>0.257</u>	<u>0.166</u>	0.256	0.169	0.269	0.199	0.263	0.196	0.304	0.184	0.322	0.210	0.305	0.201	0.315	0.378
	336	0.183	0.273	<u>0.185</u>	<u>0.277</u>	0.188	0.286	0.193	0.312	0.217	0.319	0.198	0.300	0.223	0.319	0.200	0.304	0.376
	720	<u>0.227</u>	0.307	0.225	0.317	0.229	<u>0.311</u>	0.233	0.312	0.258	0.352	0.220	0.320	0.258	0.350	0.246	0.355	0.376
交通	96	<u>0.459</u>	<u>0.274</u>	0.462	0.285	0.500	0.324	0.428	0.271	0.526	0.347	0.593	0.321	0.650	0.396	0.587	0.366	0.867
	192	<u>0.471</u>	<u>0.291</u>	0.473	0.296	0.500	0.324	0.448	0.282	0.522	0.332	0.617	0.336	0.598	0.370	0.604	0.373	0.869
	336	<u>0.492</u>	<u>0.295</u>	0.498	0.296	0.528	0.341	0.473	0.289	0.517	0.334	0.629	0.336	0.605	0.373	0.621	0.383	0.881
	720	0.505	0.306	<u>0.506</u>	0.313	0.561	0.369	0.516	<u>0.307</u>	0.552	0.352	0.640	0.350	0.645	0.394	0.626	0.382	0.896
第1次计数	17		19	1	2	7	2	3	5	0	0	0	0	0	0	0	0	0

Table 2. DPANet 组件的消融研究在 ETTm2 和 Weather 数据集上的表现。

模型	DPANet		仅时间		仅频率		无交叉融合	
度量	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ettm2	96	0.173	0.255	0.182	0.276	0.186	0.259	0.215
	192	0.233	0.297	0.242	0.325	0.240	0.319	0.268
	336	0.291	0.335	0.316	0.362	0.309	0.355	0.342
	720	0.390	0.395	0.412	0.417	0.426	0.409	0.461
天气	96	0.168	0.211	0.179	0.223	0.176	0.218	0.192
	192	0.207	0.246	0.220	0.257	0.223	0.242	0.243
	336	0.250	0.285	0.268	0.312	0.261	0.304	0.268
	720	0.338	0.339	0.351	0.352	0.349	0.346	0.373

6. REFERENCES

- [1] Jongseon Kim, Hyungjoon Kim, HyunGi Kim, Dongjun Lee, and Sungroh Yoon, “A comprehensive survey of deep learning for time series forecasting: architectural diversity and open challenges,” Artificial Intelligence Review, vol. 58, no. 7, pp. 1–95, 2025.
- [2] Peng Zhu, Yuante Li, Yifan Hu, Qinyuan Liu,

Dawei Cheng, and Yuqi Liang, “Lsr-igru: Stock trend prediction based on long short-term relationships and improved gru,” in Proceedings of the 33rd ACM International Conference on Information and Knowledge Management, 2024, pp. 5135–5142.

- [3] Chirag Deb, Fan Zhang, Junjing Yang, Siew Eang Lee, and Kwok Wei Shah, “A review on time series forecasting techniques for building energy consumption,” Renewable and Sustainable Energy Reviews, vol. 74, pp. 902–924, 2017.
- [4] Xiangfei Qiu, Jilin Hu, Lekui Zhou, Xingjian Wu, Junyang Du, Buang Zhang, Chenjuan Guo, Aoying Zhou, Christian S Jensen, Zhenli Sheng, et al., “Tfb: Towards comprehensive and fair benchmarking of time series forecasting methods,” arXiv preprint arXiv:2403.20150, 2024.
- [5] Yuchen Fang, Yuxuan Liang, Bo Hui, Zezhi Shao, Liwei Deng, Xu Liu, Xinke Jiang, and Kai Zheng, “Efficient large-scale traffic forecast-

- ing with transformers: A spatial data management perspective,” in Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1, 2025, pp. 307–317.
- [6] Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu, “Are transformers effective for time series forecasting?,” in Proceedings of the AAAI conference on artificial intelligence, 2023, vol. 37, pp. 11121–11128.
- [7] Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam, “A time series is worth 64 words: Long-term forecasting with transformers,” arXiv preprint arXiv:2211.14730, 2022.
- [8] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie, “Feature pyramid networks for object detection,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2117–2125.
- [9] Shiyu Wang, Haixu Wu, Xiaoming Shi, Tengge Hu, Huakun Luo, Lintao Ma, James Y. Zhang, and JUN ZHOU, “Timemixer: Decomposable multiscale mixing for time series forecasting,” in The Twelfth International Conference on Learning Representations, 2024.
- [10] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang, “Informer: Beyond efficient transformer for long sequence time-series forecasting,” in Proceedings of the AAAI conference on artificial intelligence, 2021, vol. 35, pp. 11106–11115.
- [11] Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long, “Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting,” Advances in neural information processing systems, vol. 34, pp. 22419–22430, 2021.
- [12] Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin, “Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting,” in International conference on machine learning. PMLR, 2022, pp. 27268–27286.
- [13] Boris N Oreshkin, Dmitri Carpov, Nicolas Chapados, and Yoshua Bengio, “N-beats: Neural basis expansion analysis for interpretable time series forecasting,” arXiv preprint arXiv:1905.10437, 2019.
- [14] Vijay Ekambaram, Arindam Jati, Nam Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam, “Tsmixer: Lightweight mlp-mixer model for multivariate time series forecasting,” in Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining, 2023, pp. 459–469.
- [15] Shizhan Liu, Hang Yu, Cong Liao, Jianguo Li, Weiyao Lin, Alex X Liu, and Schahram Dustdar, “Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting,” in # PLACEHOLDER_PARENT_METADATA_VALUE#, 2022.
- [16] Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long, “Timesnet: Temporal 2d-variation modeling for general time series analysis,” in The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023, 2023, OpenReview.net.
- [17] Yuxuan Wang, Haixu Wu, Jiaxiang Dong, Guo Qin, Haoran Zhang, Yong Liu, Yunzhong Qiu, Jianmin Wang, and Mingsheng Long, “Timexer: Empowering transformers for time series forecasting with exogenous variables,” Advances in Neural Information Processing Systems, vol. 37, pp. 469–498, 2024.

- [18] Taesung Kim, Jinhee Kim, Yunwon Tae, Cheonbok Park, Jang-Ho Choi, and Jaegul Choo, “Reversible instance normalization for accurate time-series forecasting against distribution shift,” in International conference on learning representations, 2021.