

TimeMixer: Decomposable Multiscale Mixing For Time Series Forecasting

Shiyu Wang, Haixu Wu, Xiaoming Shi, Tengge Hu, Huakun Luo, Lintao Ma

摘要

时间序列预测广泛应用于交通规划和天气预报等领域。然而，现实世界中的时间序列数据通常具有复杂和非平稳的特性，展现出混合的多种时间模式，这些模式对预测任务构成了重大挑战。TimeMixer 的核心思想是用不同的尺度对原始的时间序列进行采样，得到的新的序列蕴含的时域信息是不同的。微观信息在细尺度上反映，宏观信息在粗尺度上体现，从而复杂的变动可以被内在地解耦。TimeMixer 是一种基于 MLP 架构的模型，主要包含三个模块，分别是序列分解模块、过去可分解混合模块 (PDM)、未来多预测器混合模块 (FMM)，从而充分利用解耦后的多尺度序列在过去信息提取和未来预测阶段的优势。具体来说，序列分解模块主要对输入序列进行不同程度的池化来下采样得到不同尺度的序列。PDM 模块将每个尺度的序列都分解为趋势成分和季节性成分，并分别在粗到细和细到粗的方向上对两个成分进行混合，从而聚合微观的季节性信息和宏观的趋势信息。FMM 模块使用多个预测器分别对每个尺度的序列进行预测，然后将所有尺度的预测结果相加得到最终的预测结果，充分利用了多尺度观测中的互补预测能力。因此，TimeMixer 能够在长期和短期预测任务中取得较好的结果，并具有较高的运行效率。

关键词：时间序列；MLP；多尺度预测

1 引言

时间序列预测在经济学、能源、交通规划和天气预测中的应用十分广泛，[19] 其目标是基于时间序列的过去观测值预测未来的时间变化。但是，由于现实世界或系统的复杂和非平稳特性 [29]，观测到的序列通常表现出复杂的时间模式，给预测任务带来了严峻的挑战 [5]。

近年来，深度学习模型在时间序列预测中取得了显著的进展 [13]。其中最具代表性的一些模型涵盖了多种基础的神经网络架构，包括卷积神经网络、循环神经网络和多层感知机 [7]。为了解决复杂的时间模式，基于这些模型架构，许多深度学习模型还融入了一些特殊的设计。最常用的设计方式包括序列分解和多周期性分析，序列分解作为最经典的时间序列分析技术之一 [6]，在很多深度学习模型中都被引入作为一个基础模块，它将复杂的时间模式分解成更易预测的组成部分，如季节性和趋势 [28]，从而有助于预测过程。此外，多周期性也被引入到时间序列预测中 [34]，它将混合的时间变化分解为具有不同周期长度的多个组成成分 [26]。通过这些特殊设计，深度学习模型能够从复杂的变化中提取特征并进一步提升预测性能。

基于上述的设计，作者进一步观察到时间序列在不同采样尺度上表现出不同的时间变化。例如，如果以小时频率对交通流量进行采样的话，时间序列会呈现出早高峰或者晚高峰的特

征 [37]; 对交通流量进行每天的采样的话, 这种细节的变化就会消失, 更关注的可能是节假日这种交通流量的变化 [21]。此外, 作者还观察到未来的变化显然是由多个尺度的变化共同决定的, 其中细尺度和粗尺度分别反映微观和宏观的信息 [9]。因此, 论文从多尺度混合的新视角出发, 尝试设计一种预测模型, 能够同时利用解耦的变化和来自多尺度序列的互补预测能力来提高预测性能 [10]。

在技术上, 论文提出了 *TimeMixer*, 这是一种具有多尺度混合架构的模型, 能够通过 *Past-Decomposable-Mixing(PDM)* 模块从过去的变化中提取关键信息, 然后通过 *Future-Multipredictor-Mixing(FMM)* 模块预测未来的时间序列。具体而言, *TimeMixer* 首先通过平均池化对原始的时间序列进行降采样生成多尺度的序列; 接着, *PDM* 模块采用可分解的设计, 通过在细到粗和粗到细的方向上分别混合分解后的多尺度季节性和趋势成分。借助这个新设计, *PDM* 能够从最细尺度的序列开始聚合详细的季节性信息, 并结合粗尺度的知识深入挖掘宏观趋势成分。在预测阶段, *FMM* 通过集成多个预测器, 利用来自多尺度序列中的互补预测能力, 使得多尺度信息在历史提取和未来预测阶段得以结合。凭借该架构, *TimeMixer* 在短期和长期预测任务中都取得了稳定的最先进的性能。

2 相关工作

2.1 深度时间序列中的时间建模

时间序列分析中的关键问题之一是时间建模, 这一问题目前已经得到了广泛的研究 [30]。根据深度学习模型的基础架构, 可以将其大致分为以下四种范式 [14]: 基于 RNN、基于 CNN、基于 Transformer、基于 MLP 的方法 [25]。通常, 基于 CNN 的模型沿时间维度使用卷积核来捕捉时间模式 [2]。而基于 RNN 的方法则采用递归结构来建模时间状态的转变 [18]。但是, RNN 和 CNN 都存在感受野有限的问题, 从而限制了它们在长期预测中的能力 [24]。近年来, 基于 Transformer 的模型在长期时间序列预测中得到了广泛认可 [29], 它能够通过注意力机制自适应地捕捉长期依赖关系 [11]。此外, 多层感知机在时间序列预测中也表现出优异的预测性能和效率 [17]。

此外, 还提出了若干特殊的设计以更好地捕捉复杂的时间模式, 包括序列分解和多周期性分析 [35]。首先, 关于序列分解, Autoformer 提出了基于移动平均的序列分解模块, 将复杂的时间变化分解为季节性和趋势成分 [22]。随后, FEDformer 通过多核移动平均增强了序列分解模块 [27]。DLinear 将序列分解作为线性回归前的预处理 [32]。MICN 也将输入序列分解为季节性和趋势项, 并结合全局和局部上下文进行预测 [20]。至于多周期性分析, N-BEATS 使用多个三角基函数来拟合时间序列 [23]。TimesNet 采用傅里叶变换将时间序列映射到多个不同周期长度的成分, 并提出了一个模块化架构来处理分解后的成分 [33]。

与上述设计不同, 作者团队探索了时间序列预测中的多尺度混合架构。尽管目前已经存在一些具有多时间尺度设计的模型, 比如采用金字塔注意力的 Pyraformer [4] 和具有分叉下采样树的 SCINet [1], 它们的未来预测并没有同时利用从过去观测中提取的不同尺度的信息。而在 *TimeMixer* 中, 作者提出了一种新的多尺度混合架构, 通过 *Past-Decomposable-Mixing* 利用分解后的序列进行多尺度学习, 并通过 *Future-Multipredictor-Mixing* 集成多尺度序列的互补预测能力, 以实现更好的预测效果。

2.2 混合网络

混合是一种有效的信息整合方式,已广泛应用于计算机视觉和自然语言处理。例如,MLP-Mixer 设计了一种用于图像识别的两阶段混合结构,依次通过线性层混合通道信息和块信息 [36]。FNet 将 Transformer 中的注意力层替换为简单的傅里叶变换 [31],实现了对句子的高效标记混合。与之前设计不同的是,TimeMixer 提出了一种可分解的多尺度混合架构,并区分了在过去信息提取和未来预测阶段的混合方法。

3 本文方法

3.1 本文方法概述

给定一个具有一个或多个观测变量的时间序列 \mathbf{x} , 时间序列预测的主要目标是 [8] 利用过去的观测值 (长度为 \mathbf{P}) 来获得最可能的未来预测 (长度为 \mathbf{F})。如上所述,准确预测的关键挑战在于处理复杂的时间变化 [16]。论文中提出了 *TimeMixer* 模型,采用多尺度混合方法,利用解耦的变化和来自多尺度序列的互补预测能力。从技术角度来看,TimeMixer 由一个多尺度混合架构组成,分别通过 *Past-Decomposable-Mixing* 和 *Future-Multipredictor-Mixing* 进行过去信息提取和未来预测。

如图 1 所示,为了解开时间序列复杂的时间变化,首先通过平均池化将过去的观测数据 $\mathbf{x} \in R^{P \times C}$ 下采样到 M 个尺度,并最终获得一组多尺度的时间序列 $\mathcal{X} = \{\mathbf{x}_0, \dots, \mathbf{x}_M\}$, 其中 $\mathbf{x}_m \in R^{\lfloor \frac{P}{2^m} \rfloor \times C}$, $m \in \{0, \dots, M\}$, C 表示变量的数量。最细尺度的序列 $\mathbf{x}_0 = \mathbf{x}$ 是原始的输入序列,包含最细致的时间变化,而最粗尺度的序列 \mathbf{x}_M 则用于表示宏观变化。然后,通过嵌入层将这些多尺度序列投影到深层特征 \mathcal{X}^0 , 可以表示为 $\mathcal{X}^0 = \text{Embed}(\mathcal{X})$ 。通过上述设计就能够得到输入序列的多尺度表示。

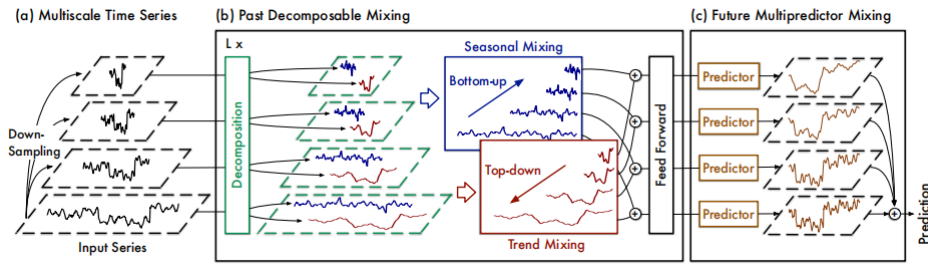


图 1. TimeMixer 的总体架构, 包括用于过去观测的过去可分解混合 (Past-Decomposable Mixing) 和用与未来预测的未来多预测混合 (Future-Multipredictor-Mixing)

然后,使用堆叠的 Past-Decomposable-Mixing (PDM) 块去混合不同尺度的历史信息。对于第 l 层, 输入为 \mathcal{X}^{l-1} , PDM 的过程可以表示为:

$$\mathcal{X}^l = PDM(\mathcal{X}^{l-1}), \quad l \in \{0, \dots, L\}, \quad (1)$$

其中, L 是总层数, $\{\mathbf{x}_0^l, \dots, \mathbf{x}_M^l\}$, $\mathbf{x}_m^l \in R^{\lfloor \frac{P}{2^m} \rfloor \times d_{model}}$ 表示具有 d_{model} 通道的混合过去表示。

对于未来预测阶段,采用 Future-Multipredictor-Mixing(FMM) 块来集成提取多尺度的过去信息 \mathcal{X}^L , 并生成未来预测, 公式表示为:

$$\hat{\mathbf{x}} = FMM(\mathcal{X}^L), \quad (2)$$

其中, $\hat{\mathbf{x}} \in R^{F \times C}$ 表示最终预测结果。通过上述设计, TimeMixer 能够成功地从解开复杂变化的多尺度观测中捕捉到重要的过去信息, 并利用多尺度过去信息预测未来。

3.2 过去可分解混合模块

我们观察到, 对于过去的观测值, 由于现实世界序列的复杂性, 即使是最粗尺度的序列也呈现出混合的变化 [3]。如图 1 所示, 最上层的序列仍然同时呈现出明显的季节性和趋势性。值得注意的是, 季节性和趋势成分在时间序列分析中具有不同的特性, 分别对应短期和长期变化, 或者平稳和非平稳动态。因此, 作者提出了 Past-Decomposable-Mixing (PDM) 模块, 用于分别混合多尺度中的季节性和趋势成分, 而不是直接将多尺度序列整体混合。

具体来说, 对于第 l 层 PDM 块, 首先通过来自 Autoformer 的序列分解块 [15], 将多尺度时间序列 \mathcal{X}^l 分解为季节性部分 $\mathcal{S}^l = \{s_0^l, \dots, s_M^l\}$ 和趋势部分 $\mathcal{T}^l = \{t_0^l, \dots, t_M^l\}$ 。如上所述, 考虑到季节性和趋势部分的不同特性, 分别对季节性和趋势项应用不同的混合操作, 以便在多个尺度间交互信息。总体而言, 第 l 层 PDM 块可以表示为:

$$\begin{aligned} s_m^l, t_m^l &= \text{SeriesDecomp}(x_m^l), m \in \{0, \dots, M\}, \\ \mathcal{X}^l &= \mathcal{X}^{l-1} + \text{FeedForward} \\ &\quad (S - \text{Mix}(\{s_m^l\}_{m=0}^M) + T - \text{Mix}(\{t_m^l\}_{m=0}^M)) \end{aligned} \quad (3)$$

其中, $\text{FeedForward}()$ 包含两个线性层, 中间使用 GELU 激活函数, 用于在通道之间进行信息交互, $S - \text{Mix}()$ 和 $T - \text{Mix}()$ 分别表示季节性和趋势项的混合操作。

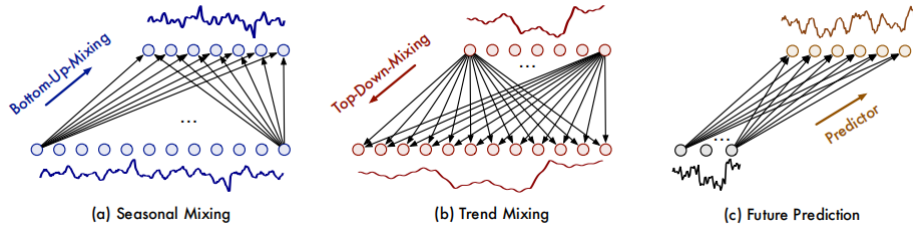


图 2. 季节性混合中的时间线性层 (a), 趋势混合中的时间线性层 (b) 和未来预测中的时间线性层 (c)

在季节性分析中, 较长的周期可以视为较短周期的聚合 [12], 例如由七天日变化组成的交通流量的每周周期, 强调了在预测未来季节性变化时, 细节信息的重要性。所以, 对于多尺度季节性部分 $\mathcal{S}^l = \{s_0^l, \dots, s_M^l\}$ 使用自下而上混合层 (Bottom-Up-Mixing) 以残差的方式作用于第 m 个尺度, 实现自下而上的季节性信息交互, 公式表示为:

$$\text{for } m : 1 \rightarrow M \text{ do: } s_m^l = s_m^l + \text{Bottom-Up-Mixing}(s_{m-1}^l). \quad (4)$$

其中, $\text{Bottom-Up-Mixing}()$ 是由两个线性层组成, 中间使用 GELU 激活函数, 沿时间维度操作, 输入维度为 $\lfloor \frac{P}{2^{m-1}} \rfloor$, 输出维度为 $\lfloor \frac{P}{2^m} \rfloor$, 如图 2 所示。

在趋势项分析中, 细尺度的趋势信息容易引入噪声, 从而影响宏观趋势的捕捉。需要注意的是, 较粗尺度的时间序列比较细尺度的时间序列更容易提供清晰的宏观信息。因此, 对

于多尺度趋势模块 $\mathcal{T}^l = \{t_0^l, \dots, t_M^l\}$ ，采用自上而下混合层以残差的方式作用于第 m 个尺度，实现自上而下的趋势信息交互。公式表示为：

$$\text{for } m : M - 1 \rightarrow 0 \quad \text{do: } t_m^l = t_m^l + \text{Top-Down-Mixing}(t_{m+1}^l). \quad (5)$$

其中， $\text{Top-Down-Mixing}()$ 由两个线性层组成，中间使用 GELU 激活函数，输入维度为 $\lfloor \frac{P}{2^{m+1}} \rfloor$ ，输出维度为 $\lfloor \frac{P}{2^m} \rfloor$ ，如图 2 所示。

3.3 未来多预测器混合模块

经过 L 个 PDM 块后，就能得到多尺度的过去信息 $\mathcal{X}^L = \{\mathbf{x}_0^L, \dots, \mathbf{x}_M^L\}$ ， $\mathbf{x}_m^L \in R^{\lfloor \frac{P}{2^m} \rfloor \times d_{model}}$ 。由于不同尺度的时间序列对应的预测能力不同，为了充分利用多尺度的信息，引入了未来多预测器混合块，每个预测器对应一个特定的尺度，其形式如下：

$$\hat{x}_m = \text{Predictor}_m(x_m^L), m \in \{0, \dots, M\}, \quad \hat{x} = \sum_{m=0}^M \hat{x}_m, \quad (6)$$

其中， $\hat{x}_m \in R^{F \times C}$ 表示来自第 m 个尺度序列的未来预测，最终输出为 $\hat{x} \in R^{F \times C}$ 。 $\text{Predictor}_m()$ 表示第 m 个尺度序列的预测器，首先通过一个单一的线性层直接回归到长度为 F 的未来值，基于从长度为 $\lfloor \frac{P}{2^m} \rfloor$ 提取的过去信息，然后将深度表示投影到 C 个变量上。需要注意的是，FMM 是多个预测器的集成，多个预测器基于来自不同尺度的过去信息，使得 FMM 能够整合多尺度序列的互补预测能力。

4 复现细节

4.1 与已有开源代码对比

时间序列是连续记录的，每个时间点只保存一些标量，单一的时间点通常不能提供足够的语义信息用于分析，所以我们通常集中在研究时间变化上，时间变化能够反映时间序列的内在特性，比如连续性、周期性、季节性、趋势等。由于现实世界的时间序列通常呈现多周期性，比如天气观测中的日变化和年变化，电力消耗中的周变化和季节变化。这些多个周期相互重叠并相互作用，导致对时间变化的建模变得十分困难。因此，在源代码的基础上采用傅里叶变换进一步提取关键的时间变化信息。在时序预测中，傅里叶变换是一种常用的工具，它将时域上的数据转换到频域上，从而揭示数据中的周期性模式和频率特征。此外，傅里叶变换有助于分离时间序列中的趋势成分和季节性成分。在很多时间序列中，趋势成分是低频的，而季节性成分是高频的。同时，傅里叶变换可以帮助识别并去除时间序列中的高频噪声，通过过滤掉不重要的频率成分，可以改善模型的预测性能。在本次研究中，我们致力于复现先前的工作，并在此过程中，我们参考了本论文的开源代码作为我们复现的基础，链接如下 <https://github.com/kwuking/TimeMixer>

4.2 实验环境搭建

整个实验在单块 NVIDIA A40 GPU 上跑，CUDA 版本是 12.1，pytorch 版本是 2.4.1，python 版本是 3.8.20，根据 github 上该团队提供的依赖文件，安装相关版本的依赖包。去

Google Drive 或者 Baidu Drive 上下载实验所需的数据集并存放至项目根目录的 dataset 文件夹下。

4.3 创新点

通过引入傅里叶变换技术，对时间序列进行处理。这旨在识别并去除时间序列中的高频噪声，同时进一步提取时间变化中的重要信息，这一改进有助于模型在长短期时间序列预测任务中更好地适应并提取关键信息。

5 实验结果分析

论文中主要进行了长期预测和短期预测。长期预测中主要使用了八种数据集：ETT 数据集（包括四个子集：ETTh1、ETTh2、ETTm1、ETTm2）、Weather、Solar-Energy、Electricity、Traffic。短期预测中主要使用了 PeMS 数据集（包括四个公共交通网络数据集）和 M4 数据集（包括 100000 条按照不同频率采集的时间序列）。

我主要复现了论文中的长期预测和短期预测的实验。下面是长期预测的复现结果，红色字体是论文中的实验数据。所有结果均取自 4 种不同预测长度（96、192、336、720）的平均值，并将输入长度固定为 96，通过引入傅里叶变换，能够较好地提取时间序列中的季节性信息和趋势信息，并且还去除了一些高频噪声，实验结果有略微提升。

prediction length (Electricity)	MSE	MAE		
96	0.153	0.245	0.153	0.247
192	0.166	0.257	0.166	0.256
336	0.185	0.275	0.185	0.277
720	0.224	0.312	0.225	0.31
Average	0.182	0.272	0.182	0.272
prediction length (Weather)	MSE	MAE		
96	0.163	0.208	0.163	0.209
192	0.210	0.254	0.208	0.25
336	0.263	0.293	0.251	0.287
720	0.344	0.347	0.339	0.341
Average	0.245	0.275	0.24	0.271
prediction length (Solar-Energy)	MSE	MAE		
96	0.216	0.308	0.189	0.259
192	0.243	0.316	0.222	0.283
336	0.246	0.303	0.231	0.292
720	0.227	0.289	0.223	0.285
Average	0.233	0.304	0.216	0.28

图 3. 长期预测实验结果

predicition length (Traffic)	MSE	MAE		
96	0.477	0.288	0.462	0.285
192	0.462	0.290	0.473	0.296
336	0.510	0.312	0.498	0.296
720	0.569	0.313	0.506	0.313
Average	0.504	0.301	0.484	0.297
prediction length (ETTh1)	MSE	MAE		
96	0.386	0.402	0.375	0.4
192	0.443	0.430	0.429	0.421
336	0.513	0.470	0.484	0.458
720	0.496	0.475	0.498	0.482
Average	0.459	0.444	0.447	0.44
prediction length (ETTh2)	MSE	MAE		
96	0.291	0.342	0.289	0.341
192	0.380	0.397	0.372	0.392
336	0.434	0.443	0.386	0.414
720	0.464	0.464	0.412	0.434
Average	0.392	0.411	0.364	0.395
prediction length (ETTm1)	MSE	MAE		
96	0.319	0.359	0.32	0.357
192	0.376	0.389	0.361	0.381
336	0.392	0.405	0.39	0.404
720	0.455	0.443	0.454	0.441
Average	0.385	0.399	0.381	0.395
prediction length (ETTm2)	MSE	MAE		
96	0.174	0.255	0.175	0.258
192	0.240	0.301	0.237	0.299
336	0.301	0.340	0.298	0.34
720	0.393	0.395	0.391	0.396
Average	0.277	0.323	0.275	0.323

图 4. 长期预测实验结果

下面是使用 PeMS 数据集做的多变量的短期预测结果，所有的输入长度都是 96。

短期预测	MAE	MAPE	RMSE
PEMS03	14.77 (14.63)	14.72 (14.54)	23.52 (23.28)
PEMS04	19.18 (19.21)	12.50 (12.53)	30.85 (30.92)
PEMS07	20.72 (20.57)	8.72 (8.62)	33.69 (33.59)
PEMS08	15.27 (15.22)	9.70 (9.67)	24.34 (24.26)

图 5. 多变量短期预测实验结果

下面是使用 M4 数据集做的单变量短期预测结果，所有的输入长度为 [6,48]。

短期预测	SMAPE	MASE	OWA
Yearly	13.21 (13.206)	2.955 (2.916)	0.776 (0.776)
Quarterly	9.996 (9.996)	1.171 (1.166)	0.881 (0.825)
Monthly	12.793 (12.605)	0.941 (0.919)	0.886 (0.869)
Others	4.535 (4.564)	3.132 (3.115)	0.971 (0.982)
Average	11.805 (11.723)	1.569 (1.559)	0.845 (0.840)

图 6. 单变量短期预测实验结果

6 总结与展望

在本文中,我们全面探讨了一种用于解决时序预测中复杂时间变化的模型 TimeMixer,并在此基础上进行了有益的扩展。主要的创新点包括引入傅里叶变换技术,以去除数据中的高频噪声,更好地提取周期性信息,以提高预测性能。

然而,在当前研究中仍然存在一些不足之处。例如,实验中固定了输入序列的长度,对于多样化的时间序列数据,可能会限制模型对不同长度时间序列的适应能力。此外,TimeMixer对含有缺失值或异常值的数据表现较差,需要额外的预处理步骤或在模型中加入鲁棒性模块。

未来的研究方向包括但不限于以下几点:首先,开发支持动态输入长度的机制,使模型能够根据任务或数据特性灵活调整滑动窗口,从而提高使用性。其次,可以通过改进损失函数或引入数据增强技术优化模型表现。此外,可以将 TimeMixer 与物理模型或领域知识融合,构建物理引导的混合模型,在提升预测精度的同时增强解释性。

参考文献

- [1] Francisco Massa A. Lerer James Bradbury Gregory Chanan-Trevor Killeen Z. Lin N. Gimelshein L. Antiga Alban Desmaison Andreas Köpf Edward Yang Zach DeVito Martin Raison Alykhan Tejani Sasank Chilamkurthy Benoit Steiner Lu Fang Junjie Bai Adam Paszke, S. Gross and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. 2019.
- [2] Lei Zhang Ailing Zeng, Muxi Chen and Qiang Xu. Are transformers effective for time series forecasting? In *Proc. AAAI Conf. on Artificial Intelligence*, 2023.
- [3] Niki Parmar Jakob Uszkoreit Llion Jones Aidan NGomez-Lukasz Kaiser Ashish Vaswani, Noam Shazeer and Illia Polosukhin. Attention is all you need. 2017.
- [4] Nicolas Chapados Boris N Oreshkin, Dmitri Carpov and Yoshua Bengio. N-beats: Neural basis expansion analysis for interpretable time series forecasting. In *Proc. Int. Conf. on Learning Representations*, 2019.
- [5] G. E. P. Box and Gwilym M. Jenkins. time series analysis, forecasting and control. In *Proc. Int. Conf. on Learning Representations*, 1970.
- [6] Alexander Skabardonis Pravin Pratap Varaiya Chao Chen, Karl F. Petty and Zhanfeng Jia. Freeway performance measurement system: Mining loop detector data. 2001.
- [7] Boris N Oreshkin Federico Garza Max Mergenthaler Cristian Challu, Kin G Olivares and Artur Dubrawski. N-hits: Neural hierarchical interpolation for time series forecasting. In *Proc. AAAI Conf. on Artificial Intelligence*, 2023.
- [8] Jan Gasthaus David Salinas, Valentin Flunkert and Tim Januschowski. Deepar: Probabilistic forecasting with autoregressive recurrent networks. 2020.

- [9] Georg Goerg. Forecastable component analysis. 2013.
- [10] Clive William John Granger and Paul Newbold. Forecasting economic time series. 2014.
- [11] Yiming Yang Guokun Lai, Wei-Cheng Chang and Hanxiao Liu. Modeling long-and short-term temporal patterns with deep neural networks. 2018.
- [12] Jianmin Wang Haixu Wu, Jiehui Xu and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. 2021.
- [13] Mingsheng Long Haixu Wu, Hang Zhou and Jianmin Wang. Interpretable weather forecasting for worldwide stations with a unified deep model. 2023b.
- [14] Jieqi Peng Shuai Zhang Jianxin Li Hui Xiong Haoyi Zhou, Shanghang Zhang and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proc. AAAI Conf. on Artificial Intelligence*, 2021.
- [15] Feihu Huang Jince Wang Junhui Chen Huiqiang Wang, Jian Peng and Yifei Xiao. Micn:multi-scale local and global context modeling for long-term series forecasting. In *Proc. Int. Conf. on Learning Representations*, 2023.
- [16] Alexander Kolesnikov Lucas Beyer Xiaohua Zhai Thomas Unterthiner Jessica Yung Andreas Steiner Daniel Keysers Jakob Uszkoreit Mario Lucic and Alexey Dosovitskiy Ilya Tolstikhin, Neil Houlsby. Mlp-mixer: An all-mlp architecture for vision. 2021.
- [17] Ilya Eckstein James Lee-Thorp, Joshua Ainslie and Santiago Ontanon. Fnet: Mixing tokens with fourier transforms. 2022.
- [18] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Proc. Int. Conf. on Learning Representations*, 2015.
- [19] Amir Abdi Lili Meng Amin Shabani and Tristan Sylvain. Scaleformer: iterative multi-scale refining transformers for time series forecasting. In *Proc. Int. Conf. on Learning Representations*, 2023.
- [20] Jesus Polo Ana Navarro Ruth Marchante Luis Martín, Luis F Zarzalejo and Marco Cony. Prediction of global solar irradiance based on time series analysis: Application to solar thermal power plants energy production planning. 2010.
- [21] Herbert K. H. Lee Marco A. R. Ferreira, Michael A. West and David M. Higdon. Multi-scale and hidden resolution time series models. 2006.
- [22] Muxi Chen Zhijian Xu Qiuxia Lai Lingna Ma Minhao Liu, Ailing Zeng and Qiang Xu. Scinet: time series modeling and forecasting with sample convolution and interaction. 2022a.
- [23] Michael C. Mozer. Induction of multiscale temporal structure. 1991.

- [24] Lukasz Kaiser Nikita Kitaev and Anselm Levskaya. Reformer: The efficient transformer. In *Proc. Int. Conf. on Learning Representations*, 2020.
- [25] Marcello Trovati Ella Pereira Morteza Ghahremani Francesco Palmieri Pradeep Hewage, Ardhendu Behera and Yonghuai Liu. Temporal convolutional neural (tcn) network for an effective weather forecasting using time-series data from the local weather station. 2020.
- [26] Jean E McRae Robert B Cleveland, William S Cleveland and Irma Terpenning. Stl: A seasonal-trend decomposition. 1990.
- [27] Cong Liao Jianguo Li Weiyao Lin Alex X Liu Shizhan Liu, Hang Yu and Schahram Dustdar. Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting. In *Proc. Int. Conf. on Learning Representations*, 2021.
- [28] Nate Yoder Sercan O. Arik Si-An Chen, Chun-Liang Li and Tomas Pfister. Tsmixer: An all-mlp architecture for time series forecasting. 2023.
- [29] Wei Cao Jiang Bian Xiaohan Yi Shun Zheng Tianping Zhang, Yizhuo Zhang and Jian Li. Less is more: Fast multivariate time series forecasting with light sampling-oriented mlp structures. 2022.
- [30] Jinze Wei Yanming Shen Heng Qi Xueyan Yin, Genze Wu and Baocai Yin. Deep learning on traffic prediction: Methods, analysis, and future directions. 2021.
- [31] Haifeng Chen Wei Cheng Guofei Jiang Yao Qin, Dongjin Song and Garrison Cottrell. A dual-stage attention-based recurrent neural network for time series prediction. In *Proc. Int. Joint Conf. on Artificial Intelligence*, 2017.
- [32] Jianmin Wang Yong Liu, Haixu Wu and Mingsheng Long. Non-stationary transformers: rethinking the stationarity in time series forecasting. 2022b.
- [33] Phanwadee Sinthong Yuqi Nie, Nam H Nguyen and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *Proc. Int. Conf. on Learning Representations*, 2023.
- [34] Yunhao Zhang and Junchi Yan. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In *Proc. Int. Conf. on Learning Representations*, 2023.
- [35] Lujia Pan Zhe Li, Zhongwen Rao and Zenglin Xu. Mts-mixers: Multivariate time series forecasting via factorized temporal and channel mixing. 2023.
- [36] Hamidreza Zareipour Zheng Qian, Yan Pei and Niya Chen. A review and discussion of decomposition-based hybrid models for wind energy forecasting applications. 2019.
- [37] Xingming Wu Peter CY Chen Zheng Zhao, Weihai Chen and Jingmeng Liu. Lstm network: a deep learning approach for short-term traffic forecast. 2017.