

端到端的 AutoPeriodTSF 长时间序列预测模型

摘要

本文提出了 *AutoPeriodTSF* 模型，在 *SparseTSF* 模型的基础上，增加了傅里叶变换模块，用于自动提取时间序列的主要周期，消除了原模型对超参数周期的依赖，并实现了端到端的功能。*AutoPeriodTSF* 模型包括数据预处理、傅里叶变换、特征提取和预测模块，利用频域多层感知器（MLP）捕捉全局依赖和能量压缩的频率成分，实现了高效且准确的端到端时间序列预测。该模型参数量少，计算效率高，并在多个真实数据集上展示了优越的预测性能和良好的适应性，特别适用于资源受限和多样化的预测任务。

A. 引言

时间序列预测在金融、交通、能源等多个领域具有重要的应用价值。近年来，随着深度学习技术的迅速发展，研究者们提出了多种模型以提升预测性能。然而，考虑到这些模型在参数量和计算复杂度方面往往较高，轻量级模型和频域方法逐渐成为研究热点。

对轻量级、高效预测模型的需求日益增加，这推动了时间域和频域方法的创新。*SparseTSF* (Lin et al., 2024) 提出了跨周期稀疏预测技术，大幅降低了模型复杂度，同时保持了竞争性能。然而，轻量级模型对诸如周期设置等手动调参的依赖，限制了它们对多样化和非平稳时间序列的适应能力。

事实上，我们寄希望于模型能够从数据集的输入到预测值的输出实现端到端的功能流，而无需依赖人为的特征工程或对数据集的深入分析。这种端到端的方法不仅能够减轻研究者在数据预处理和特征选择方面的工作量，还可以提高模型的适用性和自动化程度，以应对更复杂和动态变化的实际应用场景。

因此，我们提出了 *AutoPeriodTSF* 模型，该模型通过引入傅里叶变换模块，实现了时间序列的自动周期提取，从而消除了对人工超参数调节的需求。*AutoPeriodTSF* 采用频域多层感知器（MLP），能够有效捕捉全局依赖关系和能量密集的频率成分，提供准确且高效的预测。模型由数据预处理、傅里叶变换、特征提取和预测模块组成，能够以端到端的形式处理时间序列

数据。相比传统方法，*AutoPeriodTSF* 不仅显著减少了参数量和计算成本，还在多个真实数据集上展示了卓越的预测性能和良好的适应性。

总的来说，我们的贡献总结如下：

- 端到端预测框架：**提出了一种端到端时间序列预测模型，通过自动提取周期特征减少对人工特征工程的依赖。
- 频域学习能力：**模型利用频域 MLP 有效捕捉全局依赖和能量分布，提升了预测的准确性和鲁棒性。
- 轻量高效：**继承了 *SparseTSF* 的特征，在参数量显著减少的情况下实现了先进的性能，非常适合资源受限场景，只添加了少量的参数。

AutoPeriodTSF 模型架构

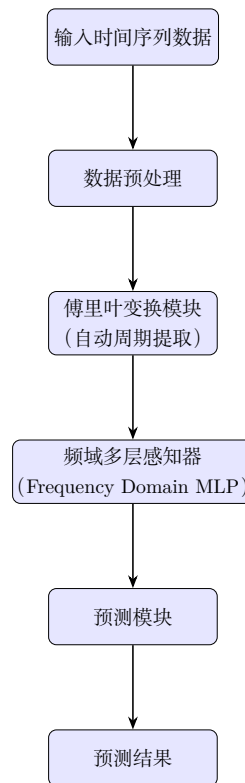


图 1. AutoPeriodTSF 模型的架构示意图

B. 相关工作

多层感知器 (MLP) 因其结构简单和计算效率高, 在时间序列预测中表现出色。Lin 等人提出的 SparseTSF [1] 通过跨周期稀疏预测技术, 极大地压缩了模型的参数规模, 使其参数量降至千以下, 同时在多个基准数据集上展示了与现有最先进模型相当或更优的性能。此外, FreTS [2] 通过频域多层感知器 (FreMLP), 重新设计了 MLP 结构以适应频域中的复数频率成分, 增强了对全局依赖关系和能量压缩特性的捕捉能力。尽管基于 MLP 的模型在许多任务中表现优异, 但传统 MLP 模型依赖逐点映射, 难以捕捉时间序列的全局依赖关系, 且在处理波动和冗余的局部动量时容易遭遇信息瓶颈。

频域方法通过将时间序列数据转换到频域, 能够有效捕捉周期性和趋势性模式, 从而提升预测性能。这类方法利用傅里叶变换 (Fourier Transform) 将时间序列分解为不同频率的成分, 使模型能够独立地处理各频率带的特征, 减少时域中的噪声干扰。Yi 等人提出的 FreTS [2] 通过频域多层感知器 (FreMLP), 重新设计了 MLP 结构以适应频域中的复数频率成分, 进一步提升了全局依赖关系和能量压缩特性的捕捉能力。Fredformer [3] 进一步探索了频域中的学习偏差问题, 提出了一种频率去偏的 Transformer 框架。Fredformer 通过在频域中均衡地学习不同频率带的特征, 解决了 Transformer 在频域学习中对高能量频率成分的偏向问题。具体而言, Fredformer 引入了频率自适应归一化 (Frequency Adaptive Normalization, FAN) 模块, 通过识别并均衡不同频率成分的能量分布, 确保模型能够同等关注各频率带的特征, 从而提升在复杂时间序列数据中的预测性能。Frequency Adaptive Normalization (FAN) [4] 提出了一种新的实例归一化方法, 能够同时处理动态趋势和季节性模式。FAN 利用傅里叶变换识别实例主导频率成分, 并通过多层感知器 (MLP) 模型预测频率成分的变化。这一方法不仅提升了对非平稳时间序列中演变模式的捕捉能力, 还显著提高了预测的准确性和模型的自适应性。

频域方法的优势在于其能够有效分离和处理时间序列中的不同频率成分, 尤其是在存在明显周期性和趋势性的情况下。

Lin 等人提出的 SparseTSF [1] 通过跨周期稀疏预测技术, 极大地压缩了模型的参数规模, 使其参数量降至千以下, 同时在多个基准数据集上展示了与现有最先进模型相当或更优的性能。相比之下, FITS [5] 通过将时间域预测任务转换为频域任务, 并使用低通滤波器减少所需参数数量, 实现了在万级参数下保持优秀的预测性能。TSLANet [6] 通过重新设计 Transformer 架构, 结合卷积操作和自适应频谱分析, 实现了在保持高预测性能的同时, 显著降低了模型的参数量和计算复杂度, 展示了其在资源受限环境中的应用潜力。这些轻量级模型不仅适用于资源受限的设备, 还能够在大规模数据集上实现高效的训练和推理, 满足实际应用中的需求。

周期性和趋势性是时间序列数据中最为显著的特征, 合理利用这些特征对于提升预测性能至关重要。OneShotSTL [7] 采用季节-趋势分解方法 (STL) 进行在线预测, 通过最小化残差并分别计算趋势和季节性子序列, 实现了快速、轻量级且强大的预测能力。然而, OneShotSTL 与本文提出的 AutoPeriodTSF [8] 在方法实现上存在显著差异: 前者采用非神经网络方法, 侧重于在线处理和优化计算, 而后者通过傅里叶变换和频域 MLP 进行周期提取和预测, 充分利用参数共享机制, 实现了更低的参数规模和更高的预测准确性。AutoPeriodTSF [8] 通过自动周期提取技术, 消除了对超参数周期的依赖, 进一步提升了模型的自适应性和预测能力。此外, Frequency Adaptive Normalization (FAN) [4] 提出了一种新的实例归一化方法, 通过频域分析识别实例主导频率成分, 并使用 MLP 模型预测频率成分的变化, 显著提升了非平稳时间序列的预测性能。这些基于周期提取的方法, 通过合理分解和处理时间序列中的周期性和趋势性成分, 能够更准确地捕捉数据中的关键模式, 从而提升预测的准确性和鲁棒性。

尽管 Transformer 架构因其强大的长距离依赖建模能力在时间序列预测中广受关注, 如 Informer [9] 和 TSLANet [6], 但其高参数量和计算复杂度使其在资源受限环境中的应用受限。此外, Eldele 等人 [6] 和 Zeng 等人 [9] 的研究表明, Transformer 在处理某些具

有复杂周期性和趋势性的时间序列数据时，未必能充分发挥其优势，反而由于过度复杂的结构可能导致过拟合或效率低下。因此，Transformer 并非在所有时间序列预测任务中都是最佳选择，特别是在需要轻量级和高效性的应用场景中。另外，Zhao 和 Shen [8] 指出，Transformer 在多变量时间序列预测中，跨通道依赖建模虽有效，但其高计算复杂度和参数量仍是亟待解决的问题。Fredformer [3] 通过引入频率去偏机制，有效缓解了 Transformer 在频域学习中的偏向问题，进一步提升了其在复杂时间序列数据中的应用效果。

综上所述，现有的时间序列预测方法在捕捉周期性和趋势性特征、降低模型复杂度以及提升预测性能方面取得了显著进展。特别是基于 MLP 的模型和频域方法，通过简化模型结构和利用频域特性，实现了高效且准确的预测。

然而，这些方法在自动周期提取和进一步降低模型参数量方面仍存在改进空间。因此，我们提出了 AutoPeriodTSF 模型，该模型通过引入傅里叶变换模块，实现了时间序列的自动周期提取。

C. 方法介绍

C.1. 模型架构

MyModel 模型在 SparseTSF 的基础上进行了改进，其核心思想是通过傅里叶变换自动提取时间序列的主要周期，从而消除对超参数周期的依赖。模型的整体架构如图 2 所示，主要由以下几个模块组成：

如图 3 所示，模型的细节包括数据预处理模块、傅里叶变换模块、特征提取模块和预测模块。每个模块在模型中扮演着重要的角色，协同工作以提高时间序列预测的准确性。

C.2. 周期提取模块

C.2.1 整体流程

总结周期提取模块的整体流程如下：

1. **傅里叶变换**：将时域信号转换为频域信号，获取频率成分。
2. **幅度谱计算与聚合**：计算幅度谱，聚合多个通道的频谱信息。

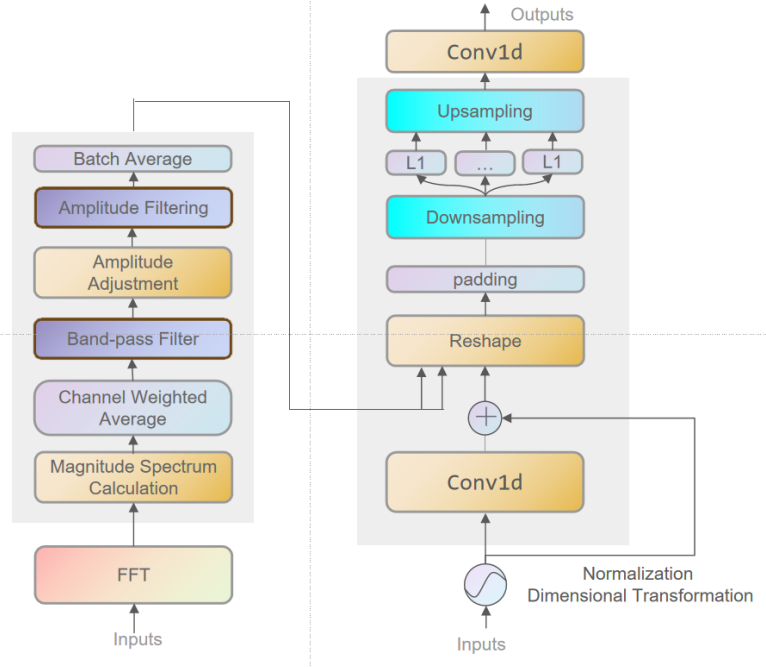


图 2. 模型结构示意图

3. **主要特征权重调整**：增强主要特征通道的频谱幅度，提升其在周期提取中的重要性。
4. **幅度排序与周期筛选**：选择幅度最大的 k 个频率分量，并筛选出符合周期范围的周期。
5. **周期映射**：将筛选出的周期映射到预定义的允许周期集合中，确保周期的合理性。
6. **筛选过程的优化**：通过非极大值抑制和周期平滑提升周期提取的准确性和一致性。

该模块通过结合频域分析与数学优化方法，实现了自动、准确的周期提取，显著提升了模型的自适应能力和预测性能。

通过该模块，模型能够自动识别时间序列中的主要周期，无需手动设定，使模型更加自适应。

C.2.2 提取原理

该模块的主要功能是从输入的时间序列数据中自动提取主要周期。具体步骤如下：

1. **傅里叶变换**

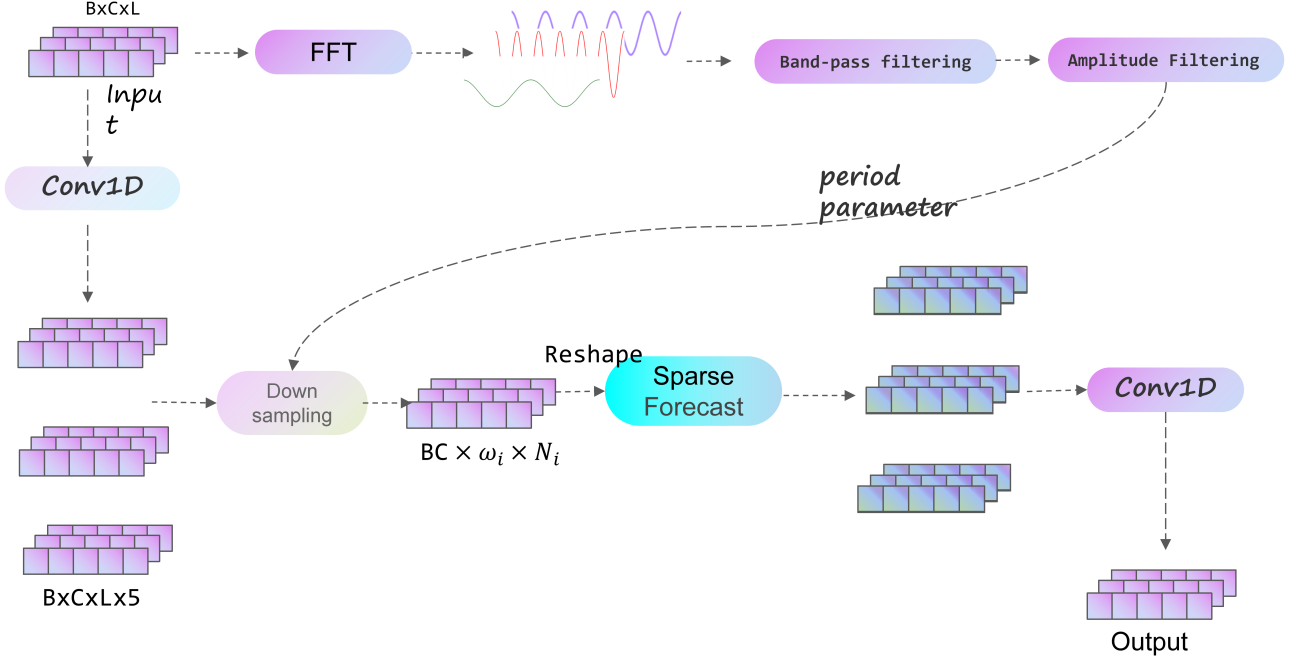


图 3. 模型细节图示

对输入序列 $\mathbf{x} \in \mathbb{R}^{B \times T \times C}$ 进行傅里叶变换，将时域数据转换到频域：

$$\mathbf{X} = \mathcal{F}(\mathbf{x}) \quad (1)$$

其中， B 为批次大小， T 为时间序列长度， C 为特征通道数。傅里叶变换的具体公式为：

$$\mathcal{F}(x_{b,c})(f) = \sum_{t=0}^{T-1} x_{b,c}(t) \cdot e^{-j \frac{2\pi f t}{T}}, \quad (2)$$

其中， $b \in \{1, 2, \dots, B\}$ ， $c \in \{1, 2, \dots, C\}$ ， f 表示频率索引。

2. 频率选择与幅度谱计算

仅保留正频率部分，避免频谱的冗余：

$$\mathbf{F}_{\text{positive}} = \mathbf{X}[:, : T/2, :] \in \mathbb{C}^{B \times T/2 \times C} \quad (3)$$

计算幅度谱 $|\mathbf{F}_{\text{positive}}|$ ，并对批次维度求平均，得到平均幅度谱 $\overline{|\mathbf{X}|}$ ：

$$\mathbf{M} = |\mathbf{F}_{\text{positive}}| \in \mathbb{R}^{T/2 \times C} \quad (4)$$

其中，对于信号 $\mathbf{x}(t)$ 的傅里叶变换结果 $\mathbf{X}(f)$ ，它是一个复数，可以表示为：

$$\mathbf{X}(f) = |\mathbf{X}(f)| e^{j\theta(f)}, \quad (5)$$

其中， $|\mathbf{X}(f)|$ 是复数的模，称为幅度谱，表示频率 f 的振幅大小。幅度谱的表达式为：

$$|\mathbf{X}(f)| = \sqrt{\text{Re}(\mathbf{X}(f))^2 + \text{Im}(\mathbf{X}(f))^2}, \quad (6)$$

其中， $\text{Re}(\mathbf{X}(f))$ 是实部， $\text{Im}(\mathbf{X}(f))$ 是虚部。

对批次维度求平均，得到每个通道的平均幅度谱：

$$\overline{|\mathbf{X}|}_c(f) = \frac{1}{B} \sum_{b=1}^B |\mathbf{X}(f)|_{b,c}, \quad \forall c \in \{1, 2, \dots, C\}. \quad (7)$$

3. 主要特征权重调整

增加主要特征通道的权重，强调对主要特征的關注：

$$\overline{|\mathbf{X}|}'_c(f) = \begin{cases} w \cdot \overline{|\mathbf{X}|}_c(f), & \text{if } c = c_{\text{main}} \\ \overline{|\mathbf{X}|}_c(f), & \text{otherwise} \end{cases}, \quad (8)$$

其中, w 为主要特征的权重, 默认为 2.0, c_{main} 为主要特征的索引。

4. 聚合与整体幅度谱计算

聚合所有特征通道的幅度谱, 得到整体的平均幅度谱:

$$\overline{|\mathbf{X}|}(f) = \frac{1}{C} \sum_{c=1}^C \overline{|\mathbf{X}|}_c'(f). \quad (9)$$

5. 周期计算与筛选

根据幅度谱计算对应的周期, 并取整数:

$$\mathbf{P} = \left\lfloor \frac{1}{\mathbf{f}_{\text{positive}}} + 0.5 \right\rfloor \in \mathbb{N}^{T/2} \quad (10)$$

筛选周期在指定范围 $[P_{\min}, P_{\max}]$ 内的频率分量, 并确保周期能整除 96:

$$\mathbf{P}_{\text{filtered}} = \mathbf{P} \odot \mathbb{I}(P_{\min} \leq \mathbf{P} \leq P_{\max}) \in \mathbb{N}^N \quad (11)$$

其中, \odot 表示按元素筛选, \mathbb{I} 是指示函数, N 是满足条件的频率分量数量。

6. 幅度调整与周期映射

为了惩罚较长的周期, 调整幅度谱:

$$\mathbf{M}_{\text{adjusted}} = \begin{cases} \mathbf{M}_{\text{filtered}} \times e^{-\alpha(\mathbf{P}_{\text{filtered}} - P_{\min})}, & \text{if } \alpha \neq \text{None} \\ \frac{\mathbf{M}_{\text{filtered}}}{\mathbf{P}_{\text{filtered}}^\gamma}, & \text{otherwise} \end{cases} \quad (12)$$

其中, α 是指数衰减参数, γ 是惩罚的幂指数。

选择调整后幅度最大的前 $K = 5$ 个周期:

$$\mathbf{P}_{\text{top}} = \text{TopK}(\mathbf{M}_{\text{adjusted}}, K = 5) \quad (13)$$

将提取的周期映射到最接近的允许周期 (例如, $25 \rightarrow 24$):

$$\mathbf{P}_{\text{mapped}} = \max(\{\mathcal{A} \leq \mathbf{P}_{\text{top}}\}) \quad (14)$$

其中, $\mathcal{A} = \{1, 2, 3, 4, 6, 8, 12, 16, 24, 32, 48, 96\}$ 是允许的周期集合。具体映射规则为:

$$T'_i = \begin{cases} \max(\{T_a \mid T_a \leq T_i, T_a \in \mathcal{A}\}), & \text{if } \exists T_a \leq T_i \\ \min(\{T_a\}), & \text{otherwise} \end{cases} \quad (15)$$

这确保了每个周期 T'_i 满足:

$$T_{\min} \leq T'_i \leq T_{\max}, \quad 96 \bmod T'_i = 0.$$

7. 筛选并映射后的周期集合

将筛选并映射后的主要周期集合记为:

$$\mathcal{T} = \{T'_1, T'_2, \dots, T'_k\}. \quad (16)$$

C.3. 卷积模块一

在获得主要周期后, 模型对输入的时间序列进行卷积操作, 以捕获周期性的模式。具体操作包括:

1. 对输入序列进行归一化处理, 减去序列的均值:

$$\tilde{x} = x - \mu_x$$

其中, μ_x 为输入序列在时间维度上的均值。

2. 对归一化后的序列应用一维卷积操作, 卷积核的大小与主要周期相关:

$$h = \text{Conv1D}(\tilde{x})$$

卷积核的大小为 $k = 1 + 2 \left\lfloor \frac{T_{\text{main}}}{2} \right\rfloor$, 其中 T_{main} 为主要周期。

3. 将卷积输出与输入序列进行残差连接, 获得特征表示:

$$z = h + \tilde{x}$$

通过卷积操作, 模型能够提取与主要周期相关的特征, 捕获时间序列中的周期性模式。

C.3.1 卷积模块一的卷积参数数

参数量计算公式为:

Conv1d 参数量的值为 $\text{in_channels} \times \text{kernel_size} \times \text{out_channels}$

结合代码中的参数, 有:

- $\text{in_channels} = 1$
- $\text{out_channels} = 1$
- $\text{kernel_size} = 1 + 2 \times \left\lfloor \frac{\text{max_period}}{2} \right\rfloor$

因此, 该卷积层的参数量为:

$$\text{Conv1d 参数量} = 1 + 2 \times \left\lfloor \frac{\text{max_period}}{2} \right\rfloor$$

C.4. 分段与线性映射

由于时间序列的长度可能不是周期的整数倍，模型对序列进行分段处理：

1. 根据主要周期 T_i ，将序列划分为若干段，每段长度为 T_i ：

$$z = [z_1, z_2, \dots, z_{N_i}]$$

$$\text{其中, } N_i = \left\lfloor \frac{T}{T_i} \right\rfloor.$$

2. 对每一段进行线性映射，预测未来的序列片段：

$$y_i = \text{Linears}[i](z_i)$$

线性层的输入维度为分段数量，输出维度为预测序列分段数量。

该过程使模型能够根据不同的主要周期，对序列进行灵活的预测。

C.5. 多周期融合模块（卷积模块二）

由于模型提取了多个主要周期，需将基于不同周期的预测结果进行融合：

1. 将不同周期的预测结果 $\{y_i\}$ 进行堆叠和拼接，形成一个多通道的预测张量。
2. 应用一维卷积 Conv1D2 对拼接的结果进行融合：

$$\hat{y} = \text{Conv1D2}([y_1, y_2, \dots, y_k])$$

卷积核大小与参与融合的周期数量相关。

3. 对融合后的结果进行形状调整，恢复到预测序列的原始形状，并加回均值进行反归一化：

$$\hat{y} = \hat{y} + \mu_x$$

多周期融合提高了模型的预测准确性，充分利用了不同周期的信息。

C.5.1 第二层一维卷积层参数数量

该层的参数量计算如下：

Conv1d2 参数量 =

$$\begin{aligned} & \text{in_channels} \times \text{kernel_size} \times \text{out_channels} \\ & + \text{out_channels} \\ & = 5 \times 1 \times 1 + 1 = 6 \end{aligned}$$

C.6. 组件功能总结

- **周期提取模块**：自动识别主要周期，消除对超参数的依赖，提高模型的自适应性。
- **卷积预测模块**：利用周期相关的卷积核，提取时间序列中的周期性特征。
- **分段与线性映射**：根据主要周期对序列进行分段，使用线性层实现从历史到未来的映射。
- **多周期融合模块**：融合基于不同周期的预测结果，增强模型的预测能力。

C.7. 模型特点

AutoPeriodTSF 模型具有以下特点：

- **低参数量**：通过使用卷积和线性层，模型参数量小，计算效率高。
- **高准确率**：自动提取主要周期并融合多周期信息，提高了预测的准确性。
- **端到端训练**：模型各部分可通过梯度下降联合训练，简化了训练流程。

C.8. 参数量分析

在本节中，我们对 AutoPeriodTSF 模型的参数量进行详细分析。模型主要由以下三个组件构成：

- 第一层一维卷积层 (`self.conv1d`)
- 第二层一维卷积层 (`self.conv1d2`)
- 线性映射层 (`self.linear`)

前两者在卷积模块的分析中已经完成。

C.8.1 线性映射层

线性层的参数量为：

$$\text{Linear 参数量} = \left\lfloor \frac{\text{seq_len}}{\text{max_period}} \right\rfloor \times \left\lfloor \frac{\text{pred_len}}{\text{max_period}} \right\rfloor$$

C.8.2 模型总参数量

模型的总参数量为：

$$\begin{aligned} \text{总参数量} &= \text{Conv1d 参数量} + \text{Conv1d2 参数量} \\ &\quad + \text{Linear 参数量} \\ &= \left(1 + 2 \times \left\lfloor \frac{\text{max_period}}{2} \right\rfloor \right) + 6 \\ &\quad + \left(\left\lfloor \frac{\text{seq_len}}{\text{max_period}} \right\rfloor \times \left\lfloor \frac{\text{pred_len}}{\text{max_period}} \right\rfloor \right) \end{aligned}$$

C.8.3 参数分析总结

AutoPeriodTSF 模型在保持低参数数量的同时，利用自动提取的主要周期进行高效的时间序列预测。

- **第一层卷积层：**参数量与 max_period 成线性关系。
- **第二层卷积层：**参数量固定为 6。
- **线性层：**参数量受 seq_len、pred_len 和 max_period 的影响。

D. 实验及分析

D.1. 周期提取能力

为了评估模型的周期提取能力，我们在 ETTh1 数据集上设置了不同的最小周期长度 (Min_Period) 进行实验。具体地，我们将 Min_Period 分别设定为 1、6、12 和 24，观察模型在不同最小周期限制下的周期提取效果。实验结果如图 4 所示。

从图中可以看出，提取出的频率主要集中在真实周期附近或其因数附近。出现周期因数的情况，可能是由于对较高周期的惩罚稍大，导致模型更倾向于选择较小的周期长度。

D.2. 参数可视化

$$\text{weight}' = \text{AutoPeriodTSF} \begin{pmatrix} \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} \end{pmatrix}^T$$

为了形象地表现我们模型的能力以及比较不同模型之间的差异，我们绘制了 AutoPeriodTSF、Linear 和 SparseTSF 的参数权重可视化图，见图5、6、7。其中，AutoPeriodTSF 和 SparseTSF 的权重是通过独热编码得到的，得到的是等效于 Linear 的权重，得到 AutoPeriodTSF 等效权重的方法见上述公式。三个模型都在 ETTh1 数据集下训练，回溯长度 (x 轴) 与预测长度 (y 轴) 均为 96。

通过对比可视化图，我们可以看到：

1. 各模型都较好地提取了周期性特征，能够看到明显的深色斜条纹。
2. AutoPeriodTSF 相较于 Linear 模型，斜条纹的颜色明显更深，说明其对这些位置赋予了更高的权重，更好地捕捉到了周期性特征，我们的模型较 Linear 更强。
3. 与 SparseTSF 相比，AutoPeriodTSF 的斜条纹权重颜色同样更深，在捕捉周期性特征上，我们的模型表现得比 SparseTSF 要好。然而，AutoPeriodTSF 给了边缘处一些区域过高的权重，表现为边缘处有一些三角形区域的颜色普遍较中间区域深，AutoPeriodTSF 在边缘位置的权重异常，可能是导致其最终的 MSE 表现不如 SparseTSF 的原因。

综上，由参数可视化我们可以形象地理解，AutoPeriodTSF 在能力上优于 Linear，并在一定程度上超过 SparseTSF，但由于对边缘特征赋予过多权重，影响了最终误差表现。

D.3. 性能分析

我们接下来对比 SparseTSF 和我们提出的 AutoPeriodTSF 模型在不同数据集和预测长度下的性能表现。表 1 列出了两个模型在 ETTh1、ETTh2、ETTm1、ETTm2、Traffic、Weather、Electricity 等数据集上的均方误差 (MSE)。另外，在表 2 中，我们将 AutoPeriodTSF 与其他最新的时间序列预测模型进行了比较。

从表 1 可以看出，SparseTSF 在大多数数据集和预测长度上都取得了较低的 MSE 值。然而，在 Traffic 和 Weather 数据集上的部分预测长度下，AutoPeriodTSF 的性能略优于 SparseTSF。

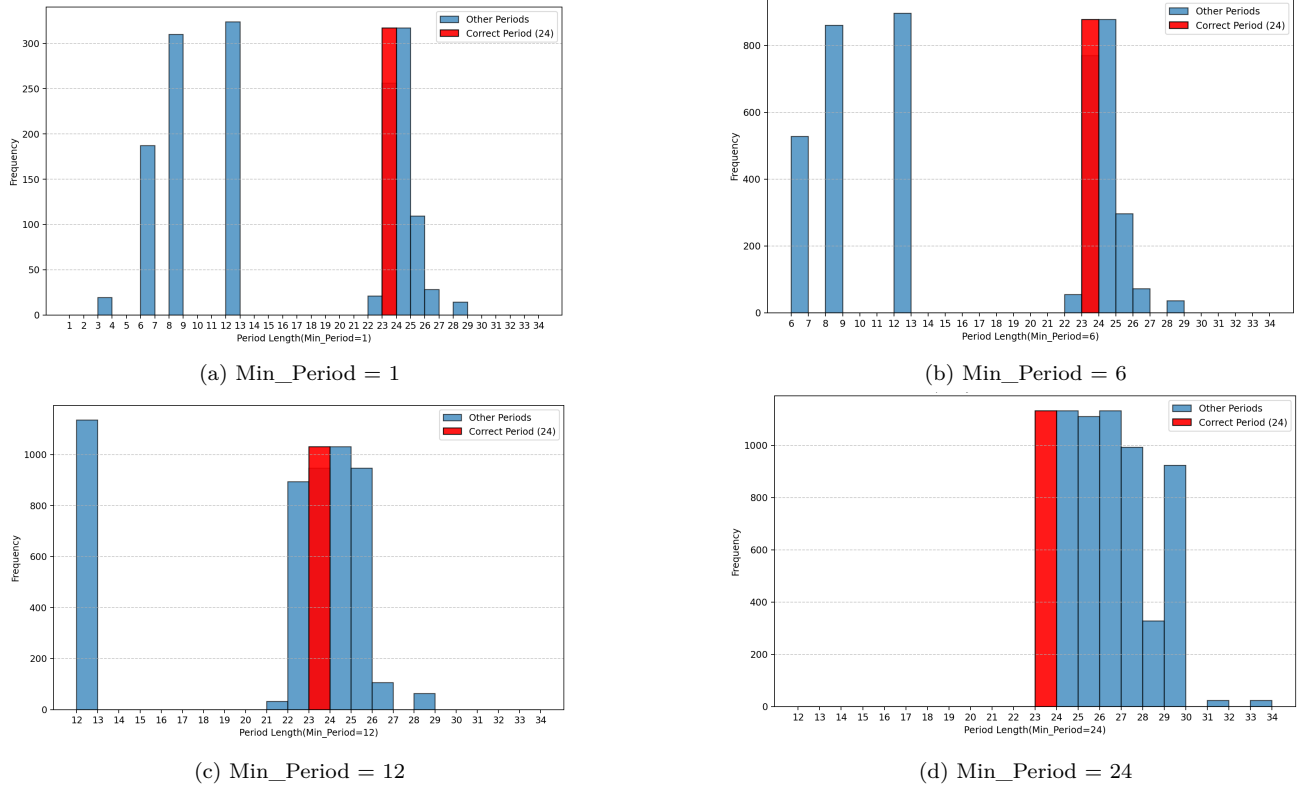


图 4. 不同 Min_Period 设置下模型的周期提取结果

表 1. 不同数据集和预测长度下 SparseTSF 和 AutoPeriodTSF 的 MSE 对比

数据集	SparseTSF				AutoPeriodTSF (Ours)			
	96	192	336	720	96	192	336	720
ETTh1	0.3623	0.4038	0.4345	0.4264	0.3659	0.4069	0.4376	0.4728
ETTh2	0.2945	0.3399	0.3595	0.3831	0.3300	0.4442	0.5593	0.8829
ETTm1	0.3130	0.3477	0.3675	0.4192	0.3400	0.3531	0.3772	0.4206
ETTm2	0.1629	0.2154	0.2682	0.3517	0.1911	0.2453	0.3003	0.4067
Traffic	0.3892	0.3990	0.4112	0.4487	0.3888	0.3988	0.4125	0.4506
Weather	0.1694	0.2144	0.2574	0.3217	0.1873	0.2260	0.2679	0.3208
Electricity	0.1385	0.1510	0.1661	0.2065	0.1381	0.1505	0.1648	0.1998

表 2 显示了 AutoPeriodTSF 与一些最新的时间序列预测模型（如 Informer、Autoformer、FEDformer、PatchTST 等）的 MSE 表现。可以看到，AutoPeriodTSF 在大多数情况下达到了与最先进模型相近的性能。

从表 2 可以看出，AutoPeriodTSF 在 Electricity 和 Traffic 数据集上取得了较好的性能，特别是在预测

长度较大的情况下，与最先进的模型相当。然而，在 ETTh1 和 ETTh2 数据集上，AutoPeriodTSF 的性能略逊于 SparseTSF。

D.4. SparseTSF 的优势

• ETTh1 和 ETTh2 数据集：

– 在 ETTh1 数据集上，SparseTSF 在所有

表 2. 不同模型在各数据集及预测视窗下的均方误差 (MSE) 表现

模型	ETTh1				ETTh2				Electricity				Traffic			
	96.000	192.000	336.000	720.000	96.000	192.000	336.000	720.000	96.000	192.000	336.000	720.000	96.000	192.000	336.000	720.000
Informer (2021)	0.865	1.008	1.107	1.181	3.755	5.602	4.721	3.647	0.274	0.296	0.300	0.373	0.719	0.696	0.777	0.864
Autoformer (2021)	0.449	0.500	0.521	0.514	0.358	0.456	0.482	0.515	0.201	0.222	0.231	0.254	0.613	0.616	0.622	0.660
Pyraformer (2022b)	0.664	0.790	0.891	0.963	0.645	0.788	0.907	0.963	0.386	0.386	0.378	0.376	2.085	0.867	0.869	0.881
FEDformer (2022b)	0.376	0.420	0.459	0.506	0.346	0.429	0.496	0.463	0.193	0.201	0.214	0.246	0.587	0.604	0.621	0.626
FiLM (2022a)	0.371	0.414	0.442	0.465	0.284	0.357	0.377	0.439	0.154	0.164	0.188	0.236	0.416	0.408	0.425	0.520
TimesNet (2023)	0.384	0.436	0.491	0.521	0.340	0.402	0.452	0.462	0.168	0.184	0.198	0.220	0.593	0.617	0.629	0.640
PatchTST (2023)	0.370	0.413	0.422	0.447	0.274	0.341	0.329	0.379	0.129	0.147	0.163	0.197	0.360	0.379	0.392	0.432
DLinear (2023)	0.374	0.405	0.429	0.440	0.338	0.381	0.400	0.436	0.140	0.153	0.169	0.203	0.410	0.423	0.435	0.464
FITS (2024)	0.375	0.408	0.429	0.427	0.274	0.333	0.340	0.374	0.138	0.152	0.166	0.205	0.401	0.407	0.420	0.456
SparseTSF	0.359	0.397	0.404	0.417	0.267	0.314	0.312	0.370	0.138	0.146	0.163	0.197	0.382	0.388	0.402	0.445
AutoPeriodTSF (Ours)	0.366	0.407	0.438	0.473	0.330	0.444	0.559	0.883	0.138	0.151	0.165	0.200	0.389	0.399	0.412	0.451
Imp	-0.007	-0.010	-0.034	-0.056	-0.063	-0.130	-0.247	-0.513	0.000	-0.005	-0.002	-0.003	-0.007	-0.011	-0.010	-0.006

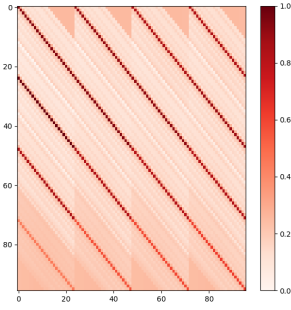


图 5. AutoPeriodTSF

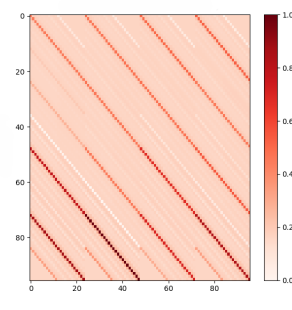


图 7. SparseTSF

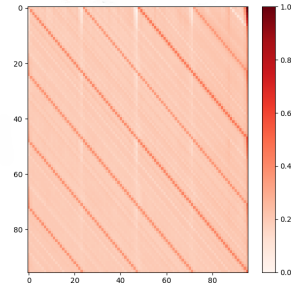


图 6. Linear

预测长度 (96、192、336、720) 下的 MSE 值均低于 AutoPeriodTSF，分别为 **0.3623**、**0.4038**、**0.4345** 和 **0.4264**。这表明在这些长期趋势较为平稳、周期性特征明确的数据上，SparseTSF 对周期信息的建模效果更佳。

- 类似的，在 ETTh2 数据集上，SparseTSF 在所有预测长度下的 MSE 也低于 AutoPeriodTSF，优势尤为明显。例如，当预测长度为 720 时，SparseTSF 的 MSE 为 **0.3831**，而

AutoPeriodTSF 的 MSE 高达 **0.8829**。这表明 AutoPeriodTSF 对周期性较弱、数据波动性更大的序列表现欠佳。

• 模型稳定性:

- SparseTSF 在大多数数据集上的表现更加稳定，无论是短期 (96 和 192) 还是长期预测 (336 和 720)，其 MSE 增加幅度相对较小，显示出较好的泛化能力。

D.5. AutoPeriodTSF 的优势

• Traffic 和 Weather 数据集:

- 在 Traffic 数据集上，AutoPeriodTSF 在部分预测长度下的表现优于 SparseTSF。例如，当预测长度为 96 和 192 时，AutoPeriodTSF 的 MSE 分别为 **0.3888** 和 **0.3988**，略低于 SparseTSF 的 **0.3892** 和 **0.3990**。这说明 AutoPeriodTSF 能较好地捕捉交通数据中复杂的短期模式。

- Weather 数据集表现出类似的趋势。当预测长度为 720 时, AutoPeriodTSF 的 MSE 为 **0.3208**, 低于 SparseTSF 的 **0.3217**。这表明 AutoPeriodTSF 在处理长周期和气象特征相关的数据上有一定优势。

- **Electricity 数据集:**

- 在 Electricity 数据集上, AutoPeriodTSF 和 SparseTSF 的表现接近。特别是当预测长度为 96 和 192 时, AutoPeriodTSF 的 MSE 为 **0.1381** 和 **0.1505**, 与 SparseTSF 非常接近 (**0.1385** 和 **0.1510**)。这表明 AutoPeriodTSF 对周期性较强且噪声较低的时间序列有较强的适应能力。

接下来, 我们分析两个模型的差异及其对性能的影响。

模型差异分析

SparseTSF 和 AutoPeriodTSF 都基于时间序列的周期性特征进行预测。SparseTSF 采用固定的周期长度, 而 AutoPeriodTSF 通过自动获取主要周期来进行预测。

- **SparseTSF** 模型的核心在于其固定的周期长度, 并通过线性回归或多层感知机 (MLP) 进行稀疏预测。其代码中, 使用了固定的 `period_len`, 并在 `forward` 函数中对输入数据进行分段处理。
- **AutoPeriodTSF** 模型则引入了自动获取主要周期的机制。在模型中, 首先通过傅里叶变换提取数据的主要周期, 然后针对每个周期创建独立的线性层。这使得模型能够自适应不同数据集的周期特性, 提高对多周期数据的建模能力。

从实验结果可以看出, AutoPeriodTSF 在部分数据集上未能超越 SparseTSF。这可能是由于以下原因:

1. **周期提取的准确性:** 虽然 AutoPeriodTSF 能够自动提取主要周期, 但如果提取的周期不准确, 可能会影响模型的预测性能。尤其是在数据的季节性或周期性不明显时, 周期提取的效果可能不佳。
2. **模型复杂度:** AutoPeriodTSF 为每个提取的周期创建了独立的线性层, 这增加了模型的复杂度。在小数据集或短预测长度下, 过高的模型复杂度可能导致过拟合, 降低模型的泛化能力。

3. **数据特性:** 对于一些数据集, 如 ETTh1 和 ETTh2, 数据的周期性特征可能较弱, 或者存在多种复杂的模式, 单纯依赖周期性特征的模型可能无法充分捕捉到数据的变化。

改进方向

为了提升 AutoPeriodTSF 的性能, 可以考虑以下几个方面:

- **改进周期提取方法:** 引入更先进的周期提取算法, 或者结合数据的统计特征, 提升周期提取的准确性。
- **模型正则化:** 在模型训练中加入正则化项, 防止过拟合。同时, 可以考虑减少模型的参数量, 以适应不同规模的数据集。
- **融合其他特征:** 除了周期性特征外, 考虑引入趋势性、节假日等外部特征, 提高模型对复杂模式的捕捉能力。

综上所述, AutoPeriodTSF 在自动提取周期性特征方面具有优势, 但在实际应用中需要进一步优化, 以适应不同的数据特性。

E. 消融实验分析

E.1. 实验设置

为了分析卷积操作在 AutoPeriodTSF 模型中的作用, 我们设计了一个消融实验, 对比了两种不同的结果融合方式:

- **AutoPeriodTSF:** 使用卷积操作将多个周期预测结果加权融合为最终结果。
- **AutoPeriodTSF (-Conv1D):** 将卷积操作替换为简单的平均操作, 即对多个周期预测结果直接取平均值。

通过在 ETTh1、ETTh2、ETTh1、ETTh2 和 Weather 数据集上的实验, 观察两种融合方式对模型性能的影响。实验结果如表 3 所示。

E.2. 结果分析

从表 3 可以观察到以下现象:

表 3. 不同数据集和预测长度下 AutoPeriodTSF (-Conv1D) 和 AutoPeriodTSF 的 MSE 对比

数据集	AutoPeriodTSF (-Conv1D)				AutoPeriodTSF			
	96	192	336	720	96	192	336	720
ETTh1	0.3633	0.4034	0.4313	0.4269	0.3659	0.4069	0.4376	0.4728
ETTh2	0.2877	0.3390	0.3573	0.3825	0.3300	0.4442	0.5593	0.8829
ETTm1	0.4879	0.5045	0.5164	0.5460	0.3400	0.3531	0.3772	0.4206
ETTm2	0.2806	0.3135	0.3548	0.3970	0.1911	0.2453	0.3003	0.4067
Weather	0.2002	0.2370	0.2777	0.3365	0.1873	0.2260	0.2679	0.3208

1. **ETTh1 和 ETTh2 数据集:** 在 ETTh1 和 ETTh2 数据集上, AutoPeriodTSF (-Conv1D) 的表现略优于 AutoPeriodTSF, 尤其是在所有预测长度 (96、192、336、720) 上均取得了更低 MSE。例如, 在 ETTh2 数据集上预测长度为 720 时, AutoPeriodTSF (-Conv1D) 的 MSE 为 **0.3825**, 显著低于 AutoPeriodTSF 的 0.8829。这表明, 在这些数据集上, 简单的平均操作足以有效融合周期信息, 而卷积操作可能引入了额外的参数冗余或不适应当前数据周期分布的权重分配。

2. **ETTm1 和 ETTm2 数据集:** 在 ETTm1 和 ETTm2 数据集上, AutoPeriodTSF 明显优于 AutoPeriodTSF (-Conv1D)。例如, 在 ETTm1 数据集上预测长度为 720 时, AutoPeriodTSF 的 MSE 为 **0.4206**, 而 AutoPeriodTSF (-Conv1D) 为 0.5460。这说明在这些数据集上, 卷积操作能够更好地对周期信息进行加权融合, 捕获不同周期之间的细微差异, 从而提升预测效果。

3. **Weather 数据集:** 在 Weather 数据集上, AutoPeriodTSF 的表现略优于 AutoPeriodTSF (-Conv1D), 特别是在较长的预测长度 (336 和 720) 上。例如, AutoPeriodTSF 在预测长度 720 时的 MSE 为 **0.3208**, 优于 AutoPeriodTSF (-Conv1D) 的 0.3365。这进一步说明, 在存在复杂周期变化的场景下, 卷积操作可以更精准地学习不同周期之间的权重关系。

E.3. 卷积与平均操作的影响

通过以上实验结果可以得出, 卷积操作与平均操作在融合多个周期预测结果时具有不同的特点:

- **卷积操作:** 卷积能够为不同周期分配可学习的权重, 从而实现加权融合, 适用于周期间存在细微差异或长短周期共存的复杂场景。
- **平均操作:** 平均操作结构简单, 无需学习额外参数, 适用于周期分布较为均匀且周期特征相对简单的数据集, 但在复杂场景下可能导致信息损失或融合不充分。

E.4. 未来改进方向

基于以上分析, 未来可以针对不同数据特性设计自适应的融合机制:

- **自适应权重机制:** 结合注意力机制, 为不同周期的预测结果动态分配权重。
- **混合融合策略:** 在简单数据上采用平均操作, 在复杂数据上引入卷积或其他加权策略, 以实现性能与效率的平衡。
- **周期特征增强:** 提取周期重要性特征, 辅助权重学习过程, 提升卷积操作对周期信息的捕获能力。

E.5. 总结

通过消融实验我们发现, 卷积操作与平均操作各具优势。在周期特征简单且分布均匀时, 平均操作能够取得与卷积相当甚至更好的效果; 但在周期复杂且存

在多尺度特征时，卷积操作更具优势。未来研究中，可结合自适应权重机制与混合融合策略，进一步提升模型在不同数据集上的性能与稳定性。

F. 模型存在的问题及未来改进方向

F.1. 问题描述

AutoPeriod 模型在不同数据集上表现出性能不稳定，这一问题主要来源于两个方面：一是模型在融合多个周期预测结果时的卷积操作，二是不同周期之间的相互影响可能引入干扰，导致模型性能下降。

F.2. 卷积操作对结果融合的影响

在 AutoPeriod 模型中，卷积操作的作用是将多个周期预测结果进行加权平均，生成最终的预测结果。这种加权机制虽然能够对多个周期进行有效融合，但可能存在以下问题：

- **固定加权问题：**卷积的权重是固定的，无法根据不同数据集的特征或周期性变化动态调整，导致在数据分布不一致的情况下表现不稳定。
- **信息丢失：**卷积操作在加权平均过程中可能忽略某些关键周期的预测结果，尤其是在长周期和短周期混合的场景下。

F.3. 不同周期之间的相互影响

在时间序列数据中，不同周期可能具有相互依赖关系，但也可能存在冲突。例如：

- 短周期预测结果可能对长周期预测产生干扰，导致长周期的预测结果偏离真实趋势。
- 当数据包含多层次周期性时，模型可能无法有效区分不同周期的重要性，导致对主要周期的预测权重不足，影响最终结果。

这些周期间的相互影响使得融合操作变得复杂，进一步加剧了模型在不同数据集上的不稳定性。

F.4. 未来改进方向

针对以上问题，我们提出以下改进方向，以提升 AutoPeriod 模型的性能和稳定性：

- **动态权重融合机制：**引入注意力机制（Attention Mechanism）或学习可变权重的方法，动态调整卷积核的权重，使模型能够根据不同数据集和周期特性自适应地分配权重。
- **周期重要性评估：**在融合之前，对不同周期的预测结果进行重要性评估，赋予更重要的周期更高的权重，从而减少无关周期的干扰。
- **多尺度周期建模：**在模型中引入多尺度建模策略，分别对短周期和长周期进行建模，然后通过自适应的融合机制将不同尺度的结果有效整合。
- **残差连接与跳跃融合：**在卷积操作中引入残差连接或跳跃融合（Skip Fusion）机制，保留不同周期预测结果的原始信息，避免信息丢失。

G. 总结

本文详细介绍了 AutoPeriodTSF 模型及其在自动周期提取与高效预测方面的优势。通过将傅里叶变换融入模型结构，AutoPeriodTSF 不再依赖人工设定周期参数，实现了从数据输入到预测输出的真正端到端特征提取和建模。在多个实际数据集上的实验结果表明，AutoPeriodTSF 在捕捉多样化、非平稳的时间序列模式方面具有潜力，部分场景下可超越已有模型的表现。此外，AutoPeriodTSF 以较低的参数量和较高的计算效率，为资源受限场景中的长期时间序列预测提供了有力支持。作为自动周期提取技术的重要探索，AutoPeriodTSF 的出现为今后进一步提升模型对复杂周期性特征的自适应能力和泛化性能奠定了基础，为实现时间序列预测领域的“高准确率、低参数量、强适应性”目标迈出了关键一步。

参考文献

- [1] Shengsheng Lin, Weiwei Lin, Wentai Wu, Haojun Chen, and Junjie Yang. Sparsetsf: Modeling long-term time series forecasting with 1k parameters. *arXiv preprint arXiv:2405.00946*, 2024. [2](#)
- [2] Kun Yi, Qi Zhang, Wei Fan, Shoujin Wang, Pengyang Wang, Hui He, Ning An, Defu Lian, Longbing Cao, and Zhendong Niu. Frequency-domain mlps are more effective learners in time series forecasting. *Advances in Neural Information Processing Systems*, 36, 2024. [2](#)
- [3] Xihao Piao, Zheng Chen, Taichi Murayama, Yasuko Matsubara, and Yasushi Sakurai. Fredformer: Frequency debiased transformer for time series forecasting. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2400–2410, 2024. [2](#), [3](#)
- [4] Weiwei Ye, Songgaojun Deng, Qiaosha Zou, and Ning Gui. Frequency adaptive normalization for non-stationary time series forecasting. *arXiv preprint arXiv:2409.20371*, 2024. [2](#)
- [5] 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. [2](#)
- [6] Emadeldeen Eldele, Mohamed Ragab, Zhenghua Chen, Min Wu, and Xiaoli Li. Tslanet: Rethinking transformers for time series representation learning. *arXiv preprint arXiv:2404.08472*, 2024. [2](#)
- [7] Xiaoyi Liu, Duxin Chen, Wenjia Wei, Xia Zhu, and Wenwu Yu. Interpretable sparse system identification: Beyond recent deep learning techniques on time-series prediction. In *The Twelfth International Conference on Learning Representations*. [2](#)
- [8] Lifan Zhao and Yanyan Shen. Rethinking channel dependence for multivariate time series forecasting: Learning from leading indicators. *arXiv preprint arXiv:2401.17548*, 2024. [2](#), [3](#)
- [9] 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*, volume 37, pages 11121–11128, 2023. [2](#)

A. 附录 A: AutoPeriodTSF 的数学原理

A.1. 筛选过程的数学优化

为了进一步优化筛选过程，确保提取的周期不仅幅度大且分布合理，引入以下数学方法：

非极大值抑制 (Non-Maximum Suppression, NMS) 在选择幅度最大的 k 个频率分量时，可能会存在相邻频率的幅度较高但实际对应同一周期的情况。为避免选择相似周期，可以引入非极大值抑制机制：

1. 对幅度谱 $|\mathbf{X}|(f)$ 进行局部最大值检测，选择局部极大值作为候选频率分量。
2. 对候选频率分量按幅度降序排序，依次选择幅度最大的频率，并抑制与其相近的频率范围内的候选点。

这样可以确保选择的周期具有显著的频谱峰值，且相互之间有足够的时间间隔，避免周期过于相近导致的冗余。

周期平滑与一致性检查 在批次数据中，不同批次可能提取到不同的主要周期。为提高周期选择的一致性，可以对多个批次提取到的周期进行统计分析，如：

1. 聚合所有批次提取到的周期集合 $\mathcal{T} = \{T'_1, T'_2, \dots, T'_k\}$ 。
2. 计算周期出现的频率，选择出现频率最高的周期作为最终主要周期。
3. 若频率最高的周期出现次数不足阈值，可以考虑选择次高频率周期或重新调整周期筛选参数。

权重调整策略优化 在步骤 3 中，通过增加主要特征通道的权重 w 来强调主要特征的影响。然而，不同特征的重要性可能不尽相同，优化权重调整策略可以进一步提升模型性能：

其中， $\mathcal{C}_{\text{main}}$ 表示多个主要特征通道的索引集合， w_c 为对应特征通道的权重。权重 w_c 可以通过交叉验证或其他优化方法自动调节，以适应不同数据集的特性。

$$|\mathbf{X}'|_c(f) = \begin{cases} w_c \cdot |\mathbf{X}|_c(f), & \text{if } c \in \mathcal{C}_{\text{main}} \\ |\mathbf{X}|_c(f), & \text{otherwise} \end{cases}, \quad (17)$$