

Beijing Forest Studio
北京理工大学信息系统及安全对抗实验中心



显式周期引导的长期时间序列预测研究

硕士研究生 邵羊飞

2025 年 11 月 02 日

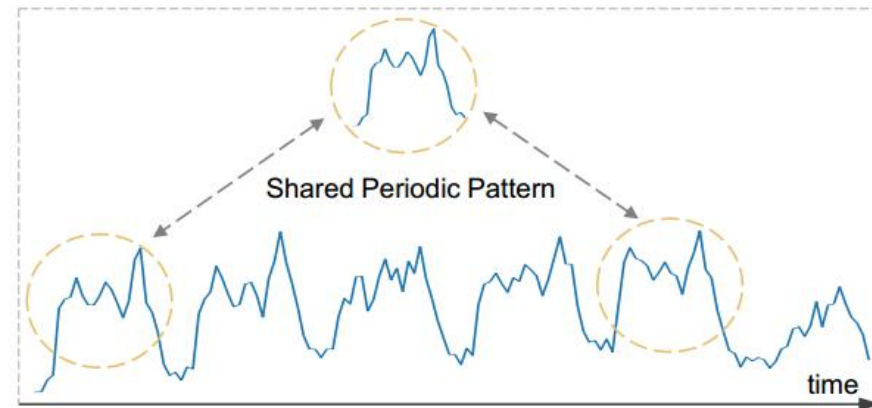
- 相关内容

- 2024.12.19 李新帅: 《多元时间序列异常检测方法研究》
- 2024.03.15 李新帅: 《不规则多元时间序列预测研究》
- 2023.06.11 李新帅: 《基于Transformer的时间序列分析》

- 预期收获
- 内涵解析与研究目标
- 研究背景与研究意义
- 研究历史
- 知识基础
- 算法原理
 - Periodformer
 - CycleNet
- 特点总结与工作展望
- 参考文献

- 预期收获
 - 了解长期时间序列预测（LTSF）的发展脉络与基本方法
 - 理解显式周期引导的LTSF的技术原理
 - 理解显式周期引导的LTSF的性能贡献和不足

- 研究目的
 - 面向**长期时间序列预测**任务
 - 结合**周期性建模**、深度学习和注意力机制等技术
 - 实现多元时间序列预测准确率和效率的提升
- 内涵解析
 - 时间序列数据（如电力消耗、交通流量）往往呈现出**稳定的周期性模式**，这种周期性为长期预测提供了可追溯的规律性基础
 - 显式周期指**直接**利用和建模时间序列的周期模式
 - 长期时间序列预测指基于**历史观测值**，预测未来长 horizon 步数的任务，horizon 通常**远超**一个季节性周期



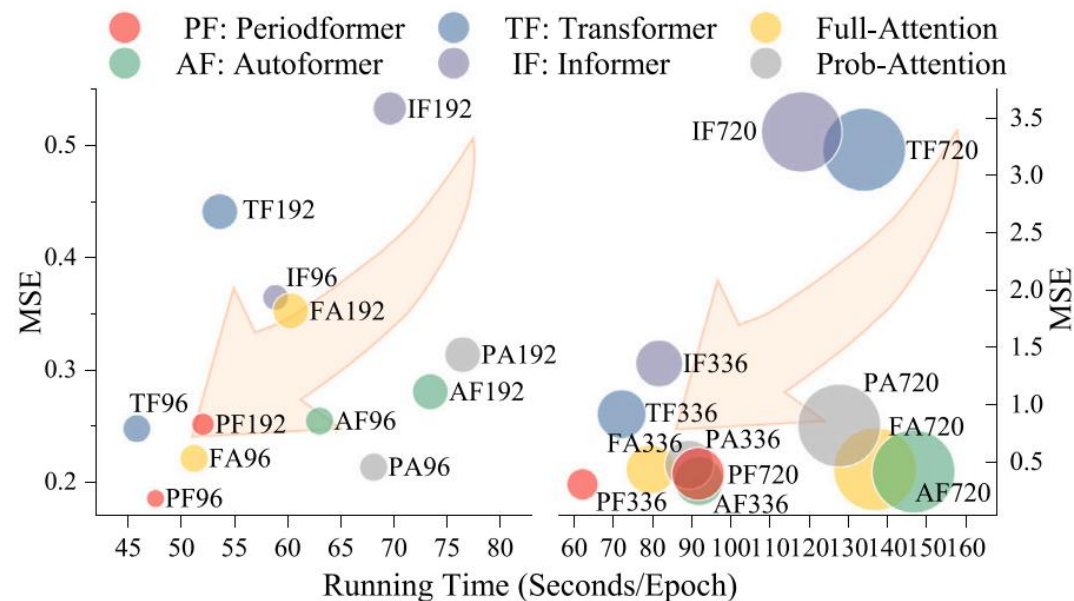
	短期预测	长期预测
高频数据 (交通流量)	预测未来1小时的流量， 时间跨度短	预测未来24小时的流量 趋势，时间 跨度长
低频数据 (季节GDP)	预测未来1个 季节周期， 预测步数小	预测未来3个 以上的季节 周期，预测 步数大

- 研究背景

- 长时预测应用于电力调度、交通管理、气象预报、经济规划等领域，能提供更早的预警和更长期的规划便利
- 主流深度学习模型（如Transformer）在处理长期预测时存在**计算效率低下、难以有效捕捉长期周期模式**

- 研究意义

- 范式创新：从“黑箱”学习转向基于**先验知识**的结构化建模
- 通过显式周期性建模，增强了模型的**可解释性和鲁棒性**
- 提升预测效率与**降低计算成本**



- 显式和隐式的区别
 - 显式周期性：内置模块去发现周期或直接利用先验周期结构
 - 隐式周期性：依赖模型的通用结构间接学习周期性
- 趋势发展

范式	隐式学习	显式发现	显式引导
核心思想	模型通用结构间接、被动地学习周期，周期是“副产品”	模型内置专用组件主动、从数据中发现主导周期，并显式基于该周期进行聚合	将先验周期知识作为超参数直接注入模型结构，进行强引导
如何利用周期	不保证。依赖注意力机制等全局学习	Autoformer (里程碑) 基于序列自相关度寻找周期，并对齐相似相位点	预设周期长度
优势	灵活，无先验要求	无需精确周期先验，能自适应不同数据集，可解释性强	效率极高，结构简单，在周期明确的数据上表现强大
劣势	黑箱、效率低、无法有效捕捉周期模式	发现机制的准确性依赖数据质量 内置模块引入不小的计算量	依赖准确的周期先验知识

研究历史 LSTF对周期性的利用

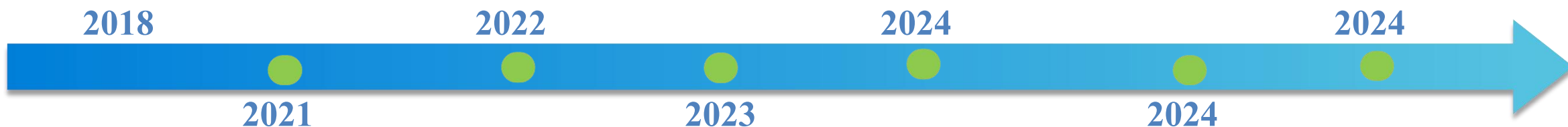


LSTNet结合**CNN与RNN和GRU**，分别捕捉短期局部依赖和长期趋势，同时通过Recurrent-skip机制显式建模周期性，并引入**时间注意力层自适应学习动态周期特征**，最后融合自回归模型增强对尺度变化的敏感性

FEDformer是在Autoformer**季节-趋势分解**的框架基础上进行的重要改进。并利用**频域增强模块**在傅里叶或小波变换后的频域中进行随机采样和**注意力计算**，有效捕获关键周期性模式。模型通过傅里叶变换将序列转换到频域，并采用随机选择固定数量频率分量的策略来构建紧凑表示

Leddham 首创**可学习序列分解技术**，结合通道与自回归双注意力机制，分别建模变量关联与时序变化。使用可学习分解核**替代**传统STD中的移动平均核，提升了分解的灵活性

CycleNet提出残差周期预测（RCF）技术，通过**可学习的循环单元**捕捉序列内在周期模式**并对残差进行预测**。该方法既可构建轻量且强大的独立模型，也能作为即插即用的模块有效提升现有模型的性能。

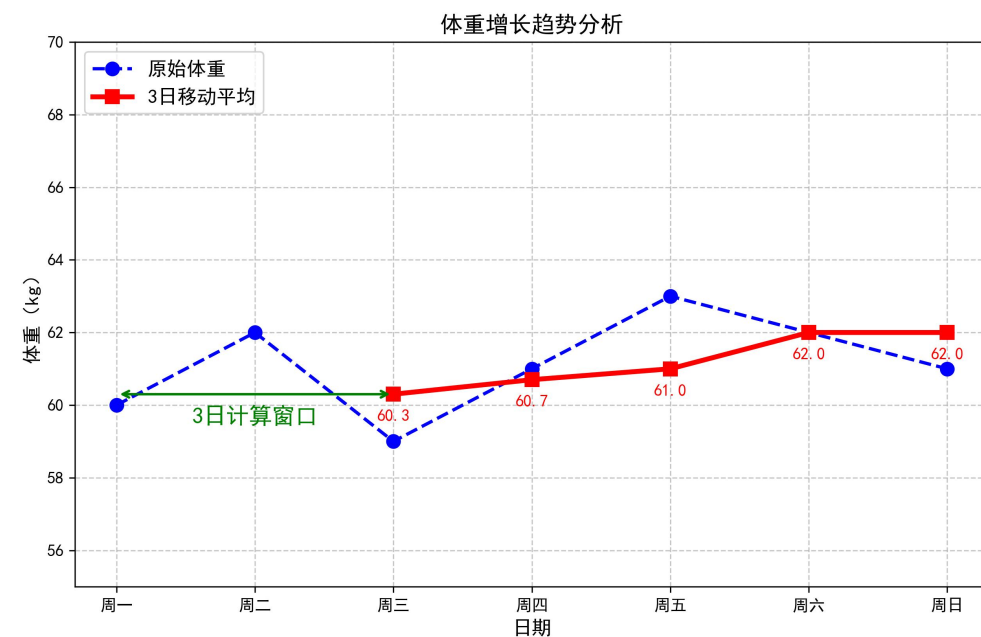


2018
Autoformer 是一项里程碑式的工作。其核心创新在于将**序列分解**作为内置算子集成到Transformer中，**利用移动平均（MA）提取趋势项**，不像传统Transformer那样隐式地学习周期，因为它专门设计了一个**自相关注意力机制**来主动地、显式地去发现序列中的周期模式

2021
DLinear 模型指出，一个简单的线性模型在将序列预先分解为季节项与趋势项后，即可取得卓越的性能。这一结果不仅凸显了序列分解技术本身的关键作用，更实证了：长时序列**经过恰当分解预处理后，简单的线性方法足以超越许多复杂的Transformer变体**

2022
Periodformer提出了轻量化的Period-Attention机制，其创新在于利用**显式周期性聚合长期子序列**，并基于内置邻近性聚合短期子序列，同时引入门控机制调节注意力影响，最终以线性计算复杂度实现强大序列建模

- 季节-趋势分解（STD）
 - 季节性（Seasonal）：代表固定周期的重复性波动（如日、周、年复一年）
 - 趋势-周期（Trend-Cyclic）：长期的方向性运动
 - 趋势：长期的上升或下降方向
 - 周期：周期长度不固定、间隔不固定的重复波动
 - 其他成分：噪声
- 分解方法：移动平均（Moving Average, MA）
 - 平均：用窗口平均值代替单日的波动值
 - 移动：时间窗口随着时间点向后移动
 - 通过平均来平滑短期内的随机波动，从而让长期、持续性的变化方向（即趋势）显现出来
 - 季节成分=原始序列-趋势成分



PERIODFORMER



Periodformer: An efficient long-term time series forecasting method based on periodic attention



T	目标	利用显式周期，降低LTSF的计算复杂度同时提升预测精度
I	输入	一个包含L个历史时间点、每个时间点具有D个特征维度的多元时间序列数据块X
P	处理	1. 数据预处理：MA-K对 时序分解 得到趋势分量和季节分量 2. 通过 周期性注意力 机制对长序列进行聚合，门控机制 调节 注意力 权重 ；前馈层聚合同一周期的短期子序列 3. 最终预测：季节性分量+ 累积 趋势分量
O	输出	模型预测的未来O个时间点、每个时间点同样具有D个特征维度的未来序列值

P	问题	1. 现有Transformer模型中的注意力机制在计算序列元素间相似度时计算 复杂度过高 2. 现有注意力机制在长序列预测中存在 过度拟合噪声 的情况
C	条件	多元长时序列
D	难点	1.如何降低注意力机制的计算复杂度 2.如何减低对噪声的过度拟合
L	水平	Knowledge-Based Systems 2024 中科院1区

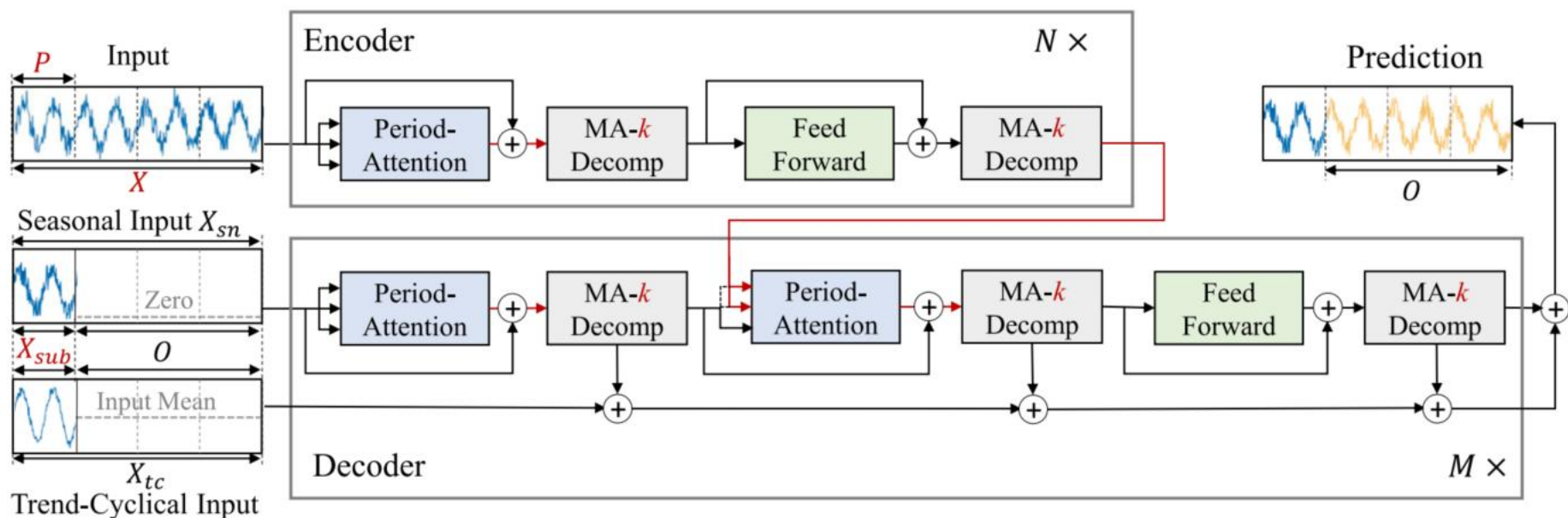
- Periodformer

- MA-K时序分解:

- 季节项: 利用编码器输出的历史周期性信息来修正和细化解码器对于未来周期的预测
 - 趋势项: 每一步都会剥离季节项中的趋势成分并累积

- 周期性注意力 (Period-Attention): 显式捕捉周期之间的依赖关系

- 前馈层 (Feed Forward): 基于邻近度聚合同一周期内的短期子序列



- 模型输入

- 编码器：输入 $X \in R^{D \times L}$ 包含 N_p 个周期数据
- 解码器： X_{sub} 即从 X 中取出最近几个周期子序列分解为
 - 趋势-周期成分 (trend-cyclical) : X_{tc}
 - 季节成分 (seasonal) : X_{sn}
- 占位符：交叉注意力 (Cross-Attention) 的格式需求

- MA-K

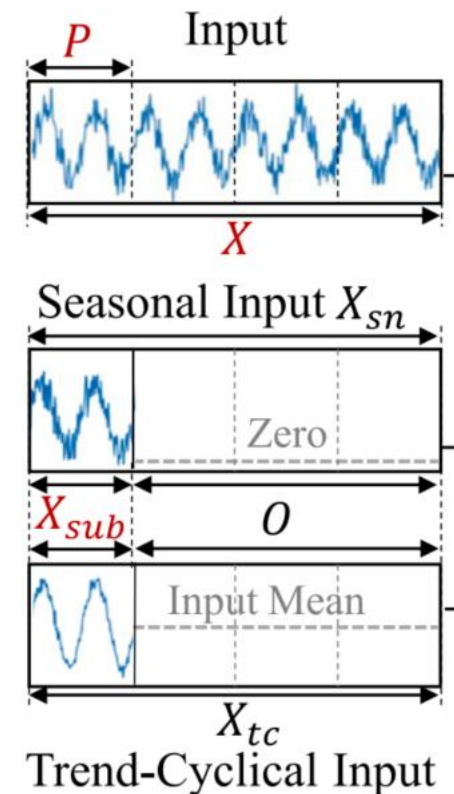
- 表示核大小为 k 的移动平均模块， k 是超参数
- 从长期序列中分离出趋势和季节部分
- 平滑输入数据，降低噪声过拟合风险

$$X_{tc}^l = X_{tc}^l + \text{MA}(\text{Pad}(X_{sn}^l), k)$$

$$X_{sn}^l = X_{sn}^l - X_{tc}^l$$

l 是模型层数

$$N_p = \left\lceil \frac{L}{P} \right\rceil (P \leq L)$$

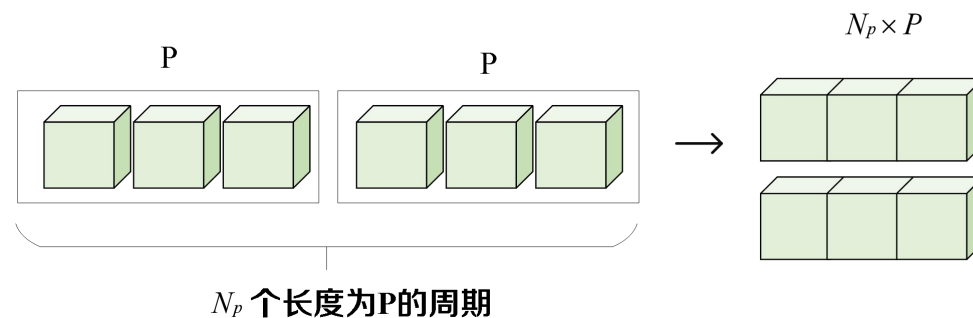
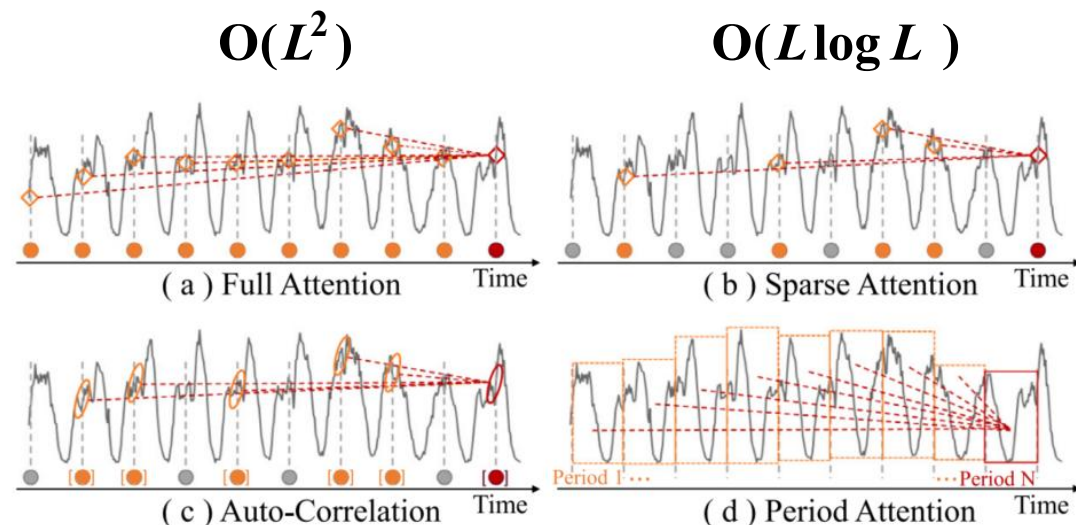


- 现有问题1

- 全注意力机制在输入序列 L 非常大时**计算量庞大**
- 其他方法虽降低了理论复杂度，但其注意力机制**实现复杂**
- 传统的注意力机制很难有效地捕捉序列长期的**周期性模式**

- Period-attention

- 引入显式**周期 P** ，是来自数据集的周期
- 对长序列基于**周期 P** 分段，仅计算周期之间的注意力
- 周期对齐，专注长序列中**重复出现**的模式



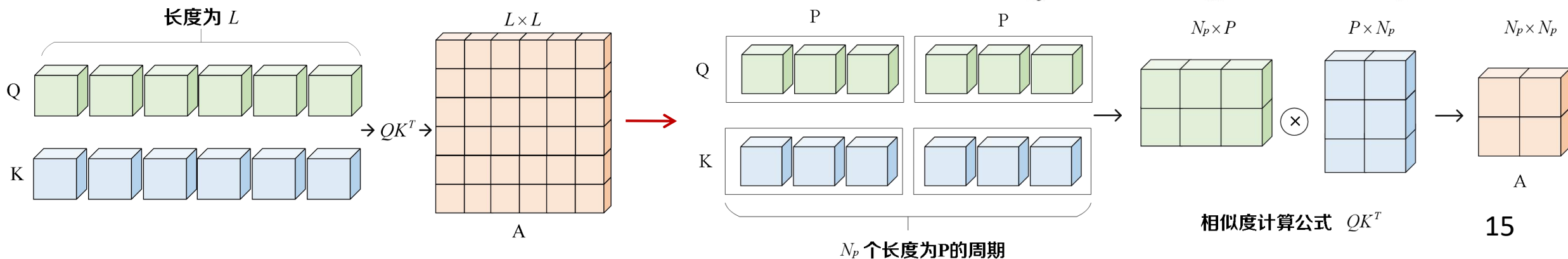
• Period-attention

– 相似度计算

- QKV矩阵被重塑，注意力矩阵变为 $N_p \times N_p$
- 对于每个注意力权重，聚合对应周期块内的 P 个时间点的信息

– 损失周期内点和点的相似度，乘法计算复杂度降为 $N_p^2 P = N_p L$

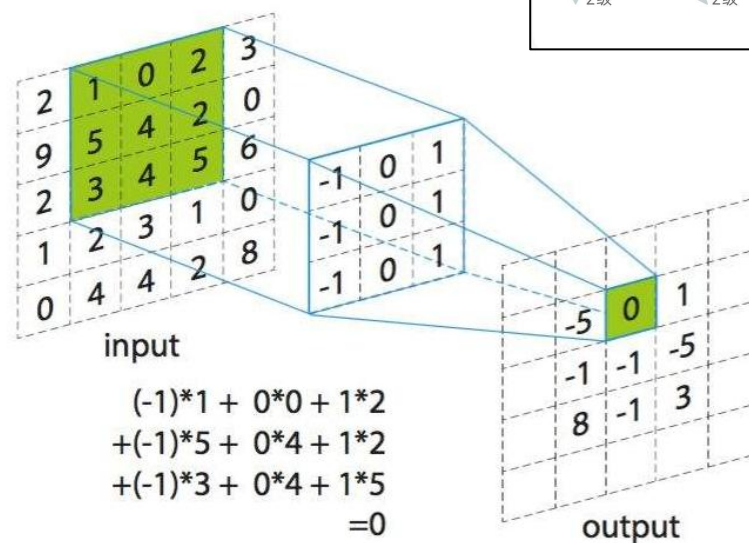
$$Att(Q, K, V) = \sigma\left(\frac{QK^T}{\sqrt{d}}\right)V \rightarrow Att(Q, K, V) = \sigma\left(\frac{QK^T}{\sqrt{P}}\right)V$$



- 前馈层

- 问题：Period-attention将序列划分为多个周期，**仅关注**周期之间的相关性，**损失**周期内点和点的相似度
- 同一周期内**距离较近的子序列**同样具有较高的相关性，即短期依赖
- 卷积运算（Conv）通过捕捉窗口内的**局部模式**，模拟临近子序列的**近似性**
- δ 是激活函数， κ 是卷积层的核大小

$$X^l = X^{l-1} + \text{Conv}(\delta(\text{Conv}((X^{l-1})^T, \kappa)), \kappa)^T$$



• 现有问题2

– 注意力机制的负面影响

- 对噪声过敏感：在低信噪比序列中，放大噪声导致过拟合

- 模型过于复杂：对于简单问题，引入不必要的参数与计算开销

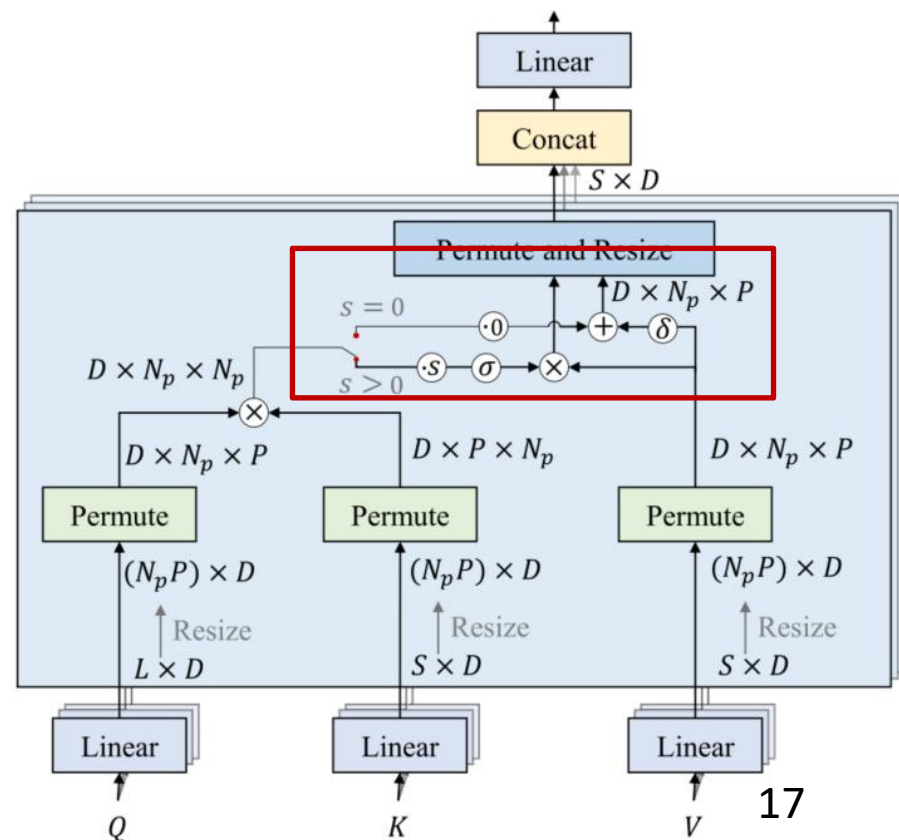
• 门控机制调节注意力权重

- 当比例因子 $s=0$ ，Attention 机制被取消，取而代之的是非线性激活操作 δ

$$Att(Q, K, V) = \sigma\left(\frac{QK^T}{\sqrt{P}}\right)V \rightarrow Att(Q, K, V) = \begin{cases} \sigma\left(\frac{QK^T \cdot s}{\sqrt{P}}\right)V, & s > 0 \\ \delta(V), & s = 0 \end{cases}$$

Negative impacts of attention mechanisms.

Datasets	ETTm2		Electricity	
Lengths	Transformer	w/o Attention	Transformer	w/o Attention
96	0.248	0.200	0.259	0.225
192	0.441	0.260	0.282	0.269
336	0.922	0.318	0.291	0.289
720	3.223	0.413	0.325	0.327



- 数据集

数据集	长度	特征数量	采样频率
ETTh1&ETTh2	17,420	7	1 hour
ETTm1&ETTm2	69,680	7	15 mins
Electricity	26,304	321	1小时
Exchange	7,588	8	1天
Traffic	17,544	862	1小时
Weather	52,696	21	10分钟
Illness	966	7	7天

- 评价指标: MSE, MAE
- 实验参数: 输入长度 $L=96$, 输出长度 $O=\{96, 192, 336, 720\}$
- 对比方法
 - 基于Transformer: FEDformer (2022)、Autoformer (2021)、Informer (2021)、LogTrans (2019)、Reformer (2020)
 - 基于大模型的方法: Time-LLM (2024)、UniTime (2024)、OneFitAll (2023)
 - 其他: LSTNet (2018)、LSTM (1997)、TCN (2018)

实验结果 多变量长时序序列预测实验



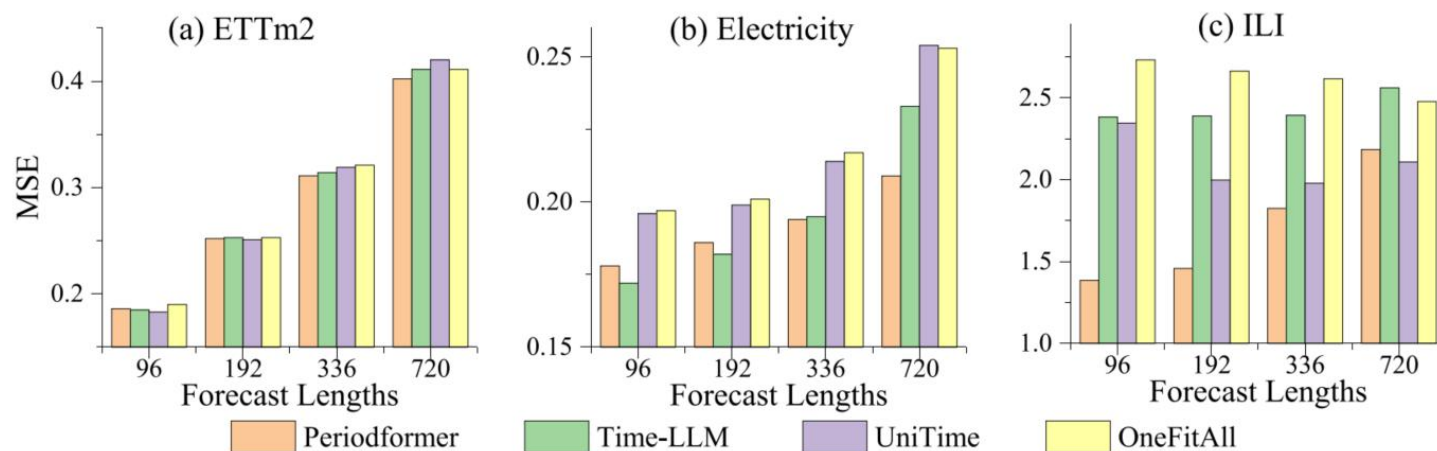
- 相比FEDformer和Autoformer，平均MSE分别降低13%和29%
- 在五项数据集中均取得最优性能，具有更强的泛化能力
- 单变量LTSF任务同样取得最优效果（图表略）

Model		Periodformer		FEDformer*		Autoformer		Informer		LogTrans		Reformer		LSTNet		LSTM		TCN	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Electricity	96	0.178	0.286	0.183	0.297	0.201	0.317	0.274	0.368	0.258	0.357	0.312	0.402	0.680	0.645	0.375	0.437	0.985	0.813
	192	0.186	0.297	0.195	0.308	0.222	0.334	0.296	0.386	0.266	0.368	0.348	0.433	0.725	0.676	0.442	0.473	0.996	0.821
	336	0.194	0.307	0.212	0.313	0.231	0.338	0.300	0.394	0.280	0.380	0.350	0.433	0.828	0.727	0.439	0.473	1.000	0.824
	720	0.209	0.320	0.231	0.343	0.254	0.361	0.373	0.439	0.283	0.376	0.340	0.420	0.957	0.811	0.980	0.814	1.438	0.784
Exchange	96	0.086	0.204	0.139	0.276	0.197	0.323	0.847	0.752	0.968	0.812	1.065	0.829	1.551	1.058	1.453	1.049	3.004	1.432
	192	0.175	0.296	0.256	0.369	0.300	0.369	1.204	0.895	1.040	0.851	1.188	0.906	1.477	1.028	1.846	1.179	3.048	1.444
	336	0.299	0.394	0.426	0.464	0.509	0.524	1.672	1.036	1.659	1.081	1.357	0.976	1.507	1.031	2.136	1.231	3.113	1.459
	720	0.829	0.690	1.090	0.800	1.447	0.941	2.478	1.310	1.941	1.127	1.510	1.016	2.285	1.243	2.984	1.427	3.150	1.458
Traffic	96	0.544	0.333	0.562	0.349	0.613	0.388	0.719	0.391	0.684	0.384	0.732	0.423	1.107	0.685	0.843	0.453	1.438	0.784
	192	0.559	0.338	0.562	0.346	0.616	0.382	0.696	0.379	0.685	0.390	0.733	0.420	1.157	0.706	0.847	0.453	1.463	0.794
	336	0.569	0.317	0.570	0.323	0.622	0.337	0.777	0.420	0.733	0.408	0.742	0.420	1.216	0.730	0.853	0.455	1.479	0.799
	720	0.594	0.354	0.596	0.368	0.660	0.408	0.864	0.472	0.717	0.396	0.755	0.423	1.481	0.805	1.500	0.805	1.499	0.804
Weather	96	0.176	0.231	0.217	0.296	0.266	0.336	0.300	0.384	0.458	0.490	0.689	0.596	0.594	0.587	0.369	0.406	0.615	0.589
	192	0.267	0.319	0.276	0.336	0.307	0.367	0.598	0.544	0.658	0.589	0.752	0.638	0.560	0.565	0.416	0.435	0.629	0.600
	336	0.316	0.353	0.339	0.380	0.359	0.395	0.578	0.523	0.797	0.652	0.639	0.596	0.597	0.587	0.455	0.454	0.639	0.608
	720	0.398	0.401	0.403	0.428	0.419	0.428	1.059	0.741	0.869	0.675	1.130	0.792	0.618	0.599	0.535	0.520	0.639	0.610
ILI	24	1.386	0.777	2.203	0.963	3.483	1.287	5.764	1.677	4.480	1.444	4.400	1.382	6.026	1.770	5.914	1.734	6.624	1.830
	36	1.458	0.813	2.272	0.976	3.103	1.148	4.755	1.467	4.799	1.467	4.783	1.448	5.340	1.668	6.631	1.845	6.858	1.879
	48	1.825	0.894	2.209	0.981	2.669	1.085	4.763	1.469	4.800	1.468	4.832	1.465	6.080	1.787	6.736	1.857	6.968	1.892
	60	2.185	0.961	2.545	1.061	2.770	1.125	5.264	1.564	5.278	1.560	4.882	1.483	5.548	1.720	6.870	1.879	7.127	1.918

- 在ETT数据集上的对比试验，实现最优预测效果
- 四个子集的区别：采样频率、变压器
- 在ETT_{h1}上，预测长度336时，MSE相比FEDformer-w降低4%
- 单变量LTSF任务同样取得最优效果（图表略）

Model		Periodformer		FEDformer-w		FEDformer-f		Autoformer		Informer		LogTrans		Reformer	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETT _{h1}	96	0.375	0.395	0.376	0.419	0.395	0.424	0.449	0.459	0.865	0.713	0.878	0.740	0.837	0.728
	192	0.413	0.421	0.420	0.448	0.469	0.470	0.500	0.482	1.008	0.792	1.037	0.824	0.923	0.766
	336	0.443	0.441	0.459	0.465	0.530	0.499	0.521	0.496	1.107	0.809	1.238	0.932	1.097	0.835
	720	0.467	0.469	0.506	0.507	0.598	0.544	0.514	0.512	1.181	0.865	1.135	0.852	1.257	0.889
ETT _{h2}	96	0.313	0.356	0.346	0.388	0.394	0.414	0.358	0.397	3.755	1.525	2.116	1.197	2.626	1.317
	192	0.389	0.405	0.429	0.439	0.439	0.445	0.456	0.452	5.602	1.931	4.315	1.635	11.120	2.979
	336	0.418	0.432	0.496	0.487	0.482	0.480	0.482	0.486	4.721	1.835	1.124	1.604	9.323	2.769
	720	0.427	0.444	0.463	0.474	0.500	0.509	0.515	0.511	3.647	1.625	3.188	1.540	3.874	1.697
ETT _{m1}	96	0.337	0.378	0.379	0.419	0.378	0.418	0.505	0.475	0.672	0.571	0.600	0.546	0.538	0.528
	192	0.413	0.431	0.426	0.441	0.464	0.463	0.553	0.496	0.795	0.669	0.837	0.700	0.658	0.592
	336	0.428	0.441	0.445	0.459	0.508	0.487	0.621	0.537	1.212	0.871	1.124	0.832	0.898	0.721
	720	0.483	0.483	0.543	0.490	0.561	0.515	0.671	0.561	1.166	0.823	1.153	0.820	1.102	0.841
ETT _{m2}	96	0.186	0.274	0.203	0.287	0.204	0.288	0.255	0.339	0.365	0.453	0.768	0.642	0.658	0.619
	192	0.252	0.317	0.269	0.328	0.316	0.363	0.281	0.340	0.533	0.563	0.989	0.757	1.078	0.827
	336	0.311	0.355	0.325	0.366	0.359	0.387	0.339	0.372	1.363	0.887	1.334	0.872	1.549	0.972
	720	0.402	0.405	0.421	0.415	0.433	0.432	0.422	0.419	3.379	1.338	3.048	1.328	2.631	1.242

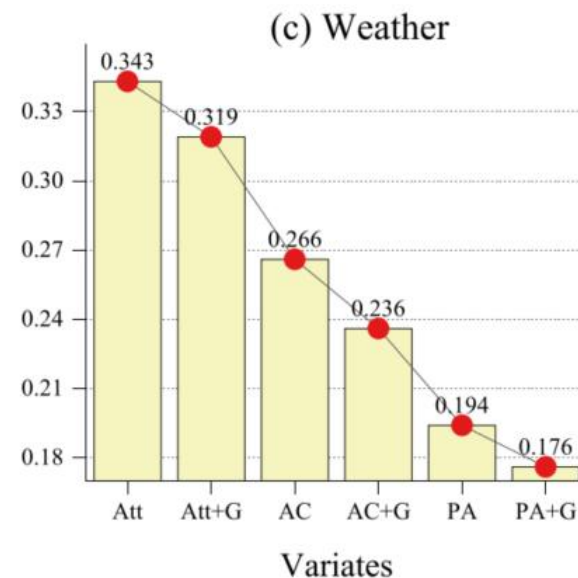
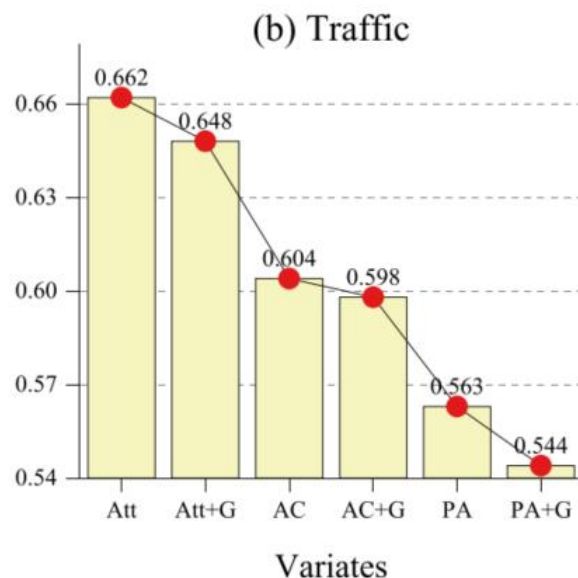
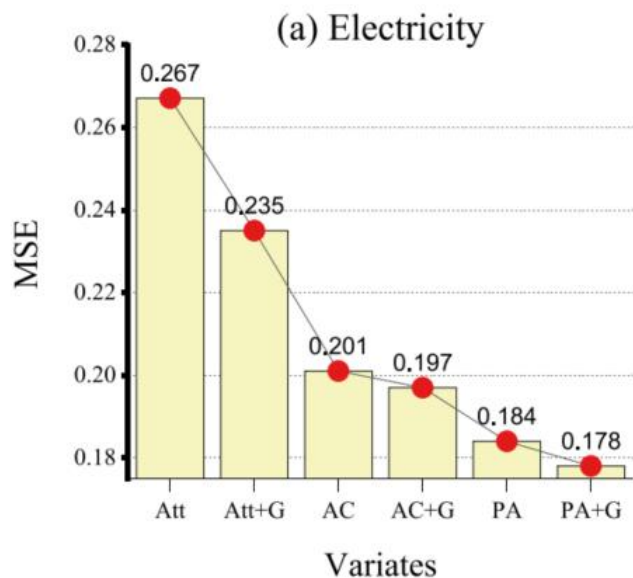
- Periodformer在平均预测性能上优于LLM-based方法



- 使用相同数据预处理条件（iTransformer）
 - Periodformer不依赖特殊预处理即可实现优越表现

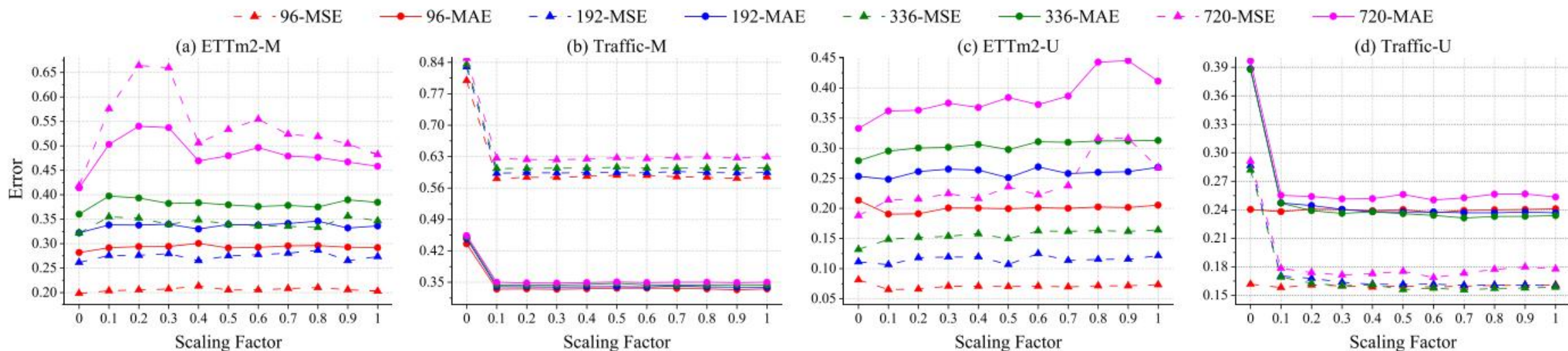
Models	Datasets	ETT				Electricity				Traffic				Weather			
		Lengths	96	192	336	720	96	192	336	720	96	192	336	720	96	192	336
Periodformer	MSE	0.295	0.378	0.42	0.423	0.144	0.164	0.173	0.223	0.391	0.412	0.431	0.463	0.169	0.223	0.275	0.353
	MAE	0.344	0.395	0.43	0.441	0.242	0.25	0.262	0.319	0.264	0.274	0.285	0.302	0.21	0.251	0.293	0.345
iTransformer	MSE	0.297	0.38	0.43	0.427	0.148	0.162	0.178	0.225	0.395	0.417	0.433	0.467	0.174	0.221	0.278	0.358
	MAE	0.349	0.4	0.43	0.445	0.24	0.253	0.269	0.317	0.268	0.276	0.283	0.302	0.214	0.254	0.296	0.349
TimesNet	MSE	0.34	0.402	0.45	0.462	0.168	0.184	0.198	0.22	0.593	0.617	0.629	0.64	0.172	0.219	0.28	0.365
	MAE	0.374	0.414	0.45	0.468	0.272	0.289	0.3	0.32	0.321	0.336	0.336	0.35	0.22	0.261	0.306	0.359
PatchTST	MSE	0.302	0.388	0.43	0.431	0.195	0.199	0.215	0.256	0.544	0.54	0.551	0.586	0.177	0.225	0.278	0.354
	MAE	0.348	0.4	0.43	0.446	0.28	0.289	0.305	0.337	0.359	0.354	0.358	0.375	0.218	0.259	0.297	0.348

- 验证不同Attention变体和门控机制的影响
 - Att: 朴素的Attention机制
 - **G**: 门控机制
 - AC: Auto-Correlation (来自Autoformer)
 - **PA**: Period-Attention



- 验证比例因子 s 的必要性

- 在Traffic数据集上, s 从0到0.1, 引入注意力后误差大幅减少
- 不同数据集和预测长度下, s 值的**大小**对性能有**调节**作用
- 数据特性: **噪声水平**直接影响时间序列的可预测性
- 当信噪比低时, 注意力模块可能会**过度拟合**噪声

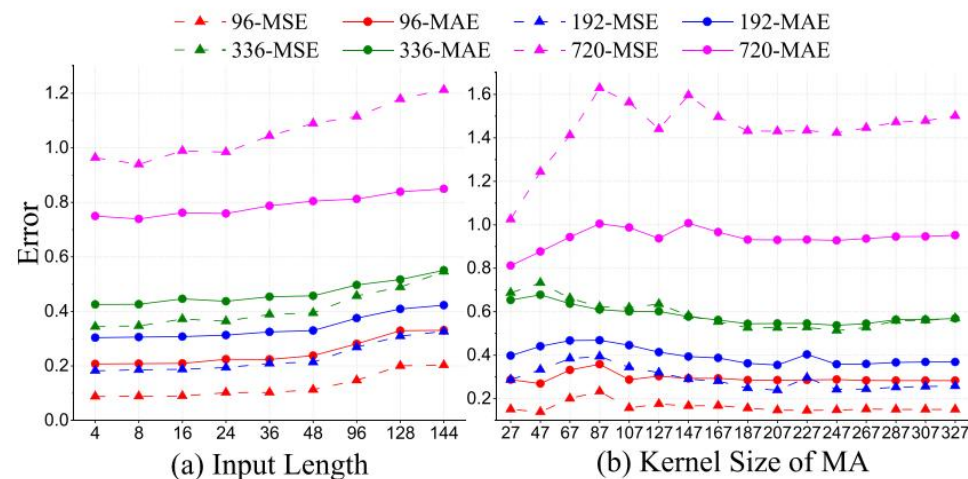


- Period-Attention有效性分析

- 使用朴素Transformer进行消融，测试输入长度和MA核大小对性能的影响
- 增加输入长度可能降低性能，而调节MA核大小能提升平均性能

- Period-Attention的优势

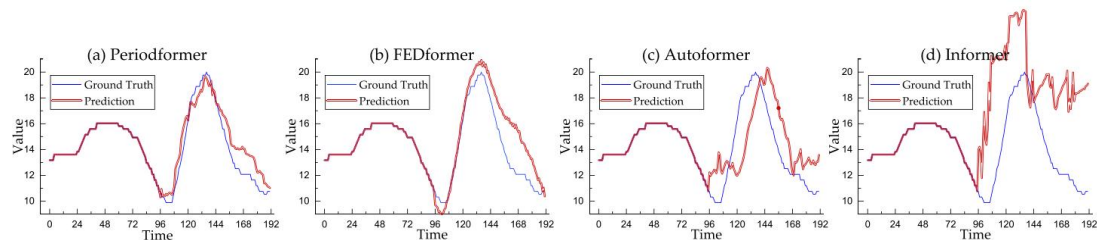
- 显式识别长期周期模式
- 聚焦周期中类似阶段相对应的相关时间点
- 减少时间衰减问题
 - 传统注意力机制在长序列中易受时间衰退影响，导致较早时间点的作用衰减
 - 周期性注意力通过对齐周期性模式，强化重复出现的关键信号，有效维持长期依赖关系



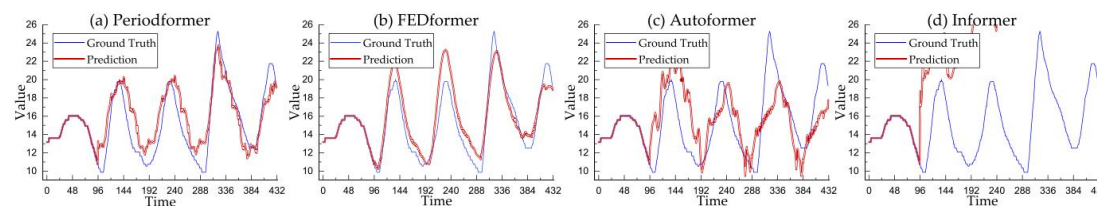
• 训练时间比较

Datasets	ETTm2			Electricity			Traffic		
Models	Training Time	Inference Time	FLOPS	Training Time	Inference Time	FLOPS	Training Time	Inference Time	FLOPS
Periodformer	47.6	13.1	24.6	29.8	12.5	27.6	24.7	12.9	35.2
iTransformer	55.8	15.3	32.0	96.7	34.6	31.9	73.4	28.6	38.3
FEDformer	242.22	60.9	182.6	349.1	68.8	79.4	238.9	47.1	90.8
Autoformer	63.0	21.3	38.4	50.5	19.3	44.7	40.6	18.3	44.7
Informer	58.87	16.1	35.1	34.1	14.1	39.4	27.9	13.6	39.4

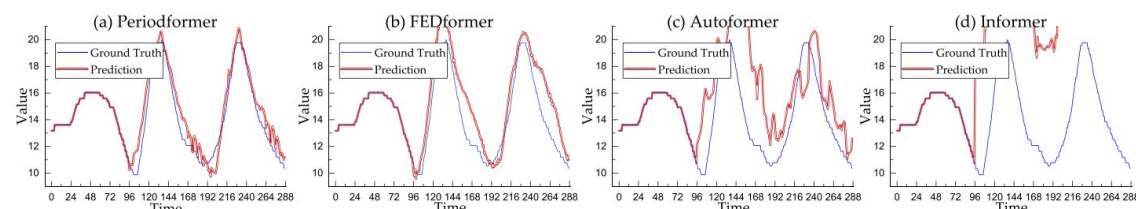
• 可视化 (ETTm2)



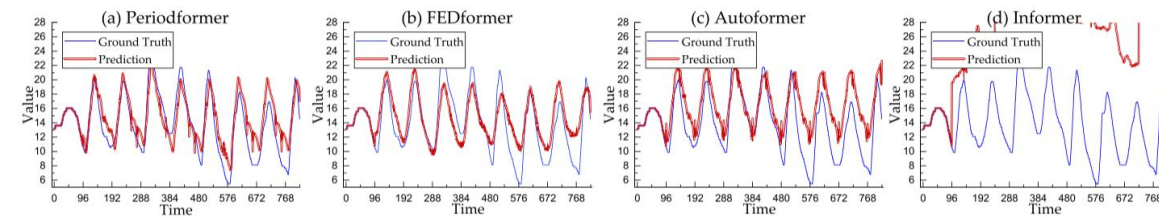
input-96-predict-96



input-96-predict-336



input-96-predict-192



input-96-predict-720

CAIOU/01

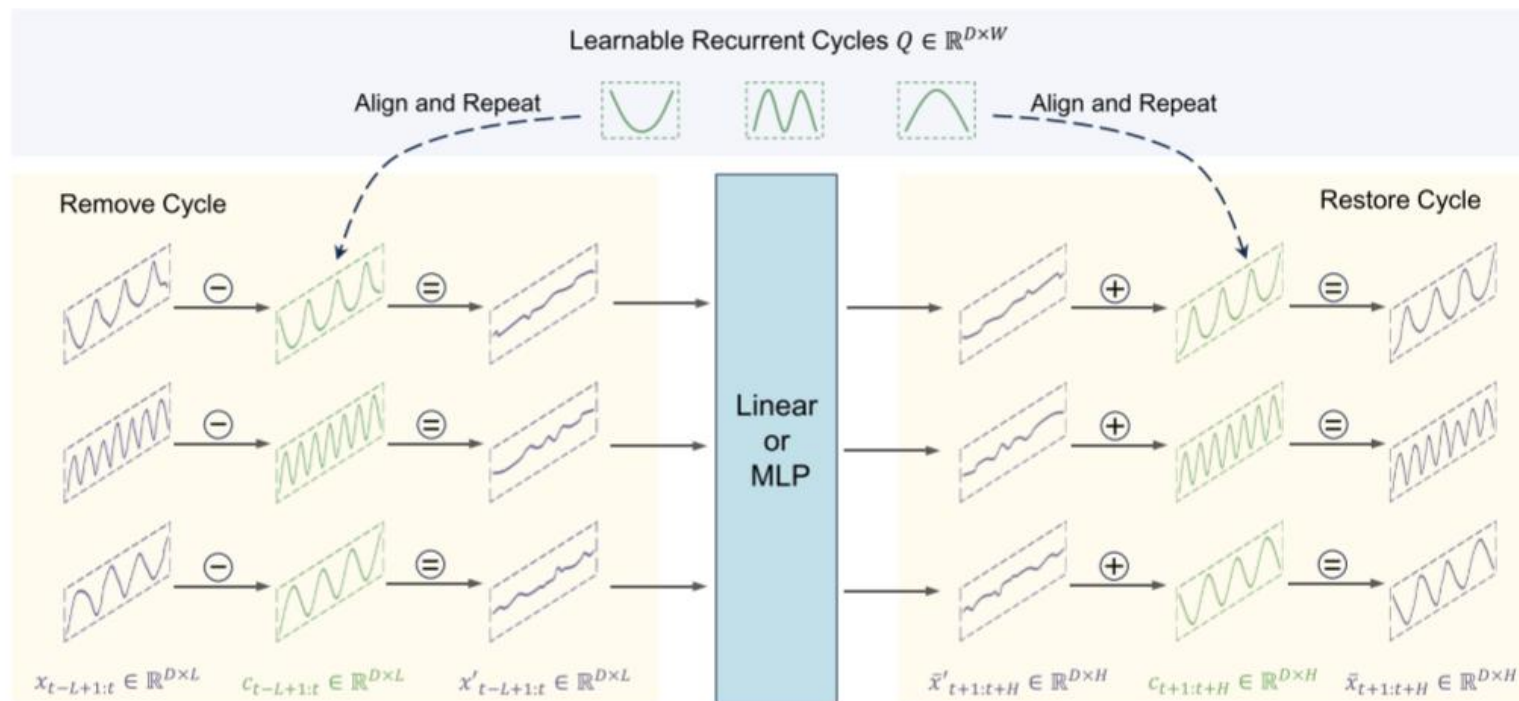


CycleNet: Enhancing Time Series Forecasting through Modeling Periodic Patterns

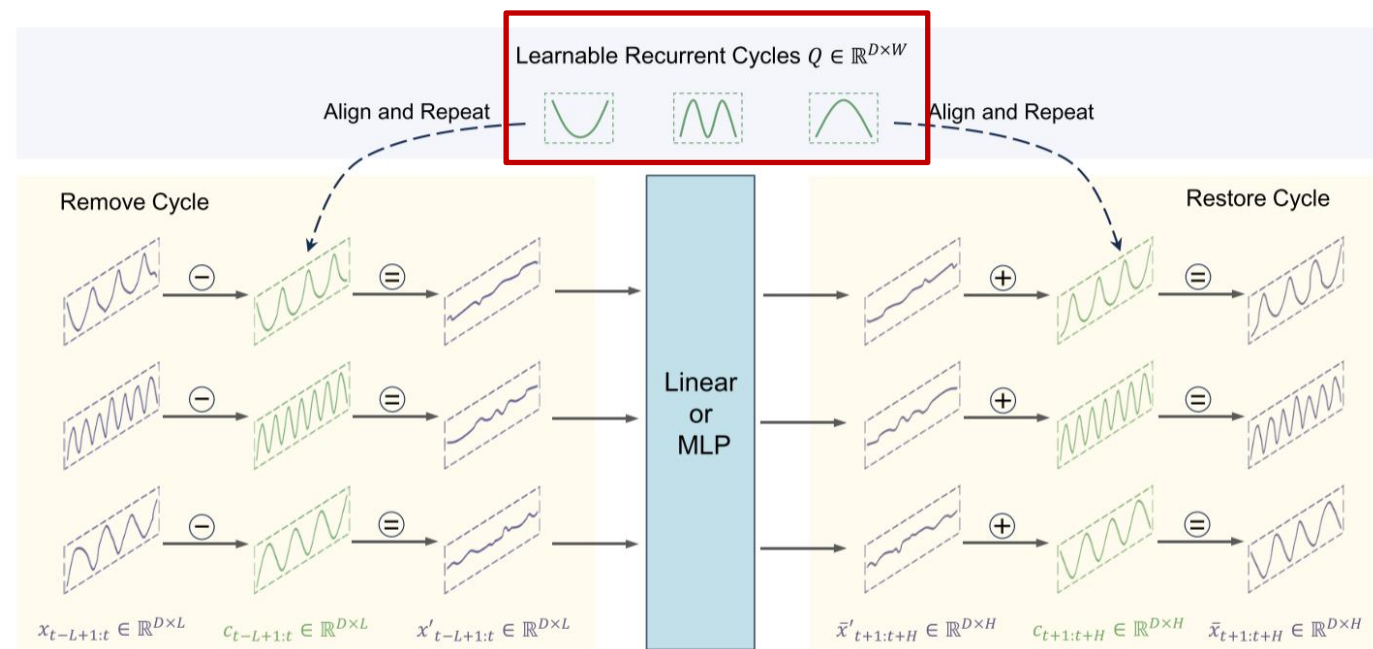
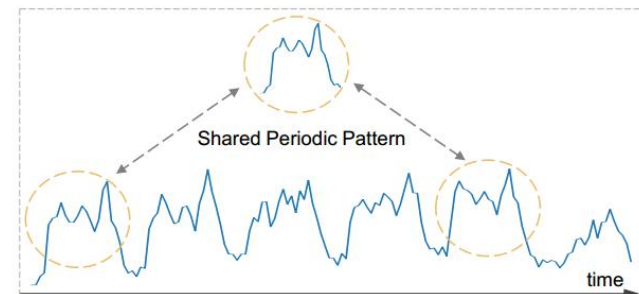
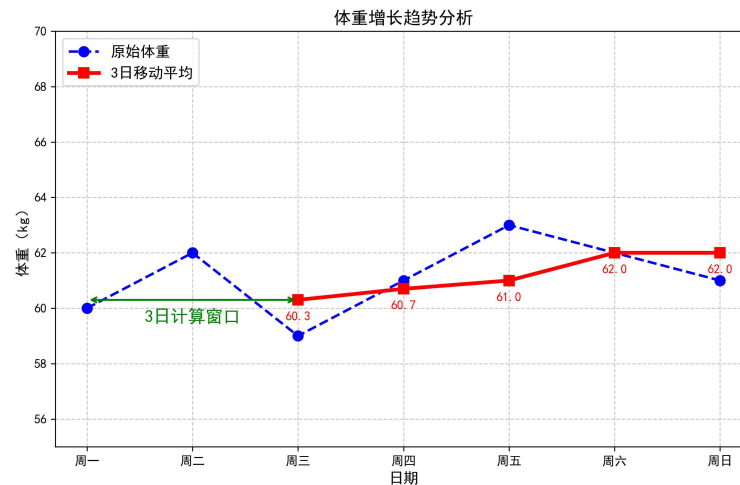
T	目标	通过显式周期建模，更好的从长期依赖关系中提取周期性特征
I	输入	一个包含L个历史时间点、每个时间点具有D个特征维度的多元时间序列数据块X
P	处理	1. 周期性模式建模：使用可学习的循环周期来显式建模序列中的固有周期性模式 2. 残差预测：对建模周期后的残差成分进行预测 3. 最终预测：周期成分+预测的残差成分
O	输出	模型预测的未来O个时间点、每个时间点同样具有D个特征维度的未来序列值

P	问题	周期性是时间序列预测的关键特征，复杂模型（如informer等）对周期性特征隐式学习效率低下，传统季节性分解方法建模周期性不够准确
C	条件	多元长时序列
D	难点	利用显式周期，高效地直接建模序列中的周期成分
L	水平	Neural Information Processing Systems 2024 CCF A

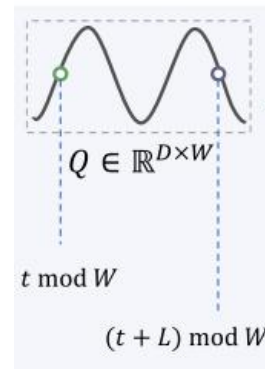
- CycleNet
 - 周期性模式建模
 - 可学习的循环队列
 - 复制生成周期组件
 - 残差预测
 - 去除周期成分
 - 主干预测残差
 - 重构预测结果
 - 数据预处理
 - 实例归一化



- MA分解：周期性建模**不准确**
 - 忽略窗口长度之外的周期性模式
 - 窗口过**小**则**去噪不足**，过**大**则**过度平滑**，掩盖短期重要波动
 - 端点信息丢失，边缘失真
- 周期性建模（RCF）
 - 延续**序列分解思想**，先建模周期，再得到趋势
 - **可学习**的循环队列 $Q \in \mathbb{R}^{D \times W}$ ，初始化为0，梯度反向传播训练
 - W 是数据集的**最大稳定周期**
 - Q 的每个通道是**全局共享**的



- 周期**对齐**、**复制**得到循环分量C
 - 索引计算：对于时间 t ，**偏移量** $i = t \bmod W$
 - 移位操作：将 Q 向左循环移位 i 位，得到 $Q^{(t_w)}$
 - 周期组件复制：将 $Q^{(t_w)}$ 重复 $\lfloor L/W \rfloor$ 次，并拼接剩余部分

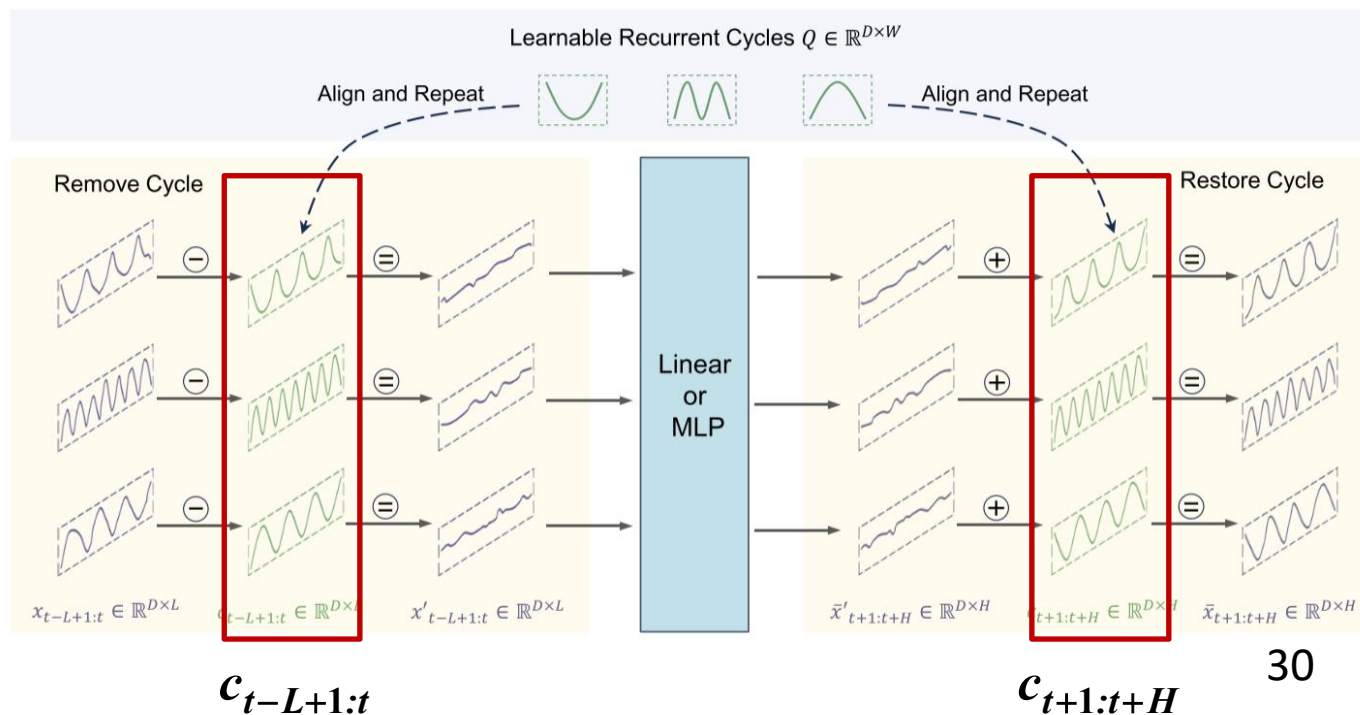


$$c_{t-L+1:t} = [\underbrace{Q^{(t_w)}, \dots, Q^{(t_w)}}_{\lfloor L/W \rfloor}, Q_{0:L \bmod W}^{(t_w)}]$$

$$c_{t+1:t+H} = [\underbrace{Q^{(t_w+L)}, \dots, Q^{(t_w+L)}}_{\lfloor H/W \rfloor}, Q_{0:H \bmod W}^{(t_w+L)}]$$

$$t_w = t - L + 1$$

L 是输入长度
 H 是输出长度



- 残差预测

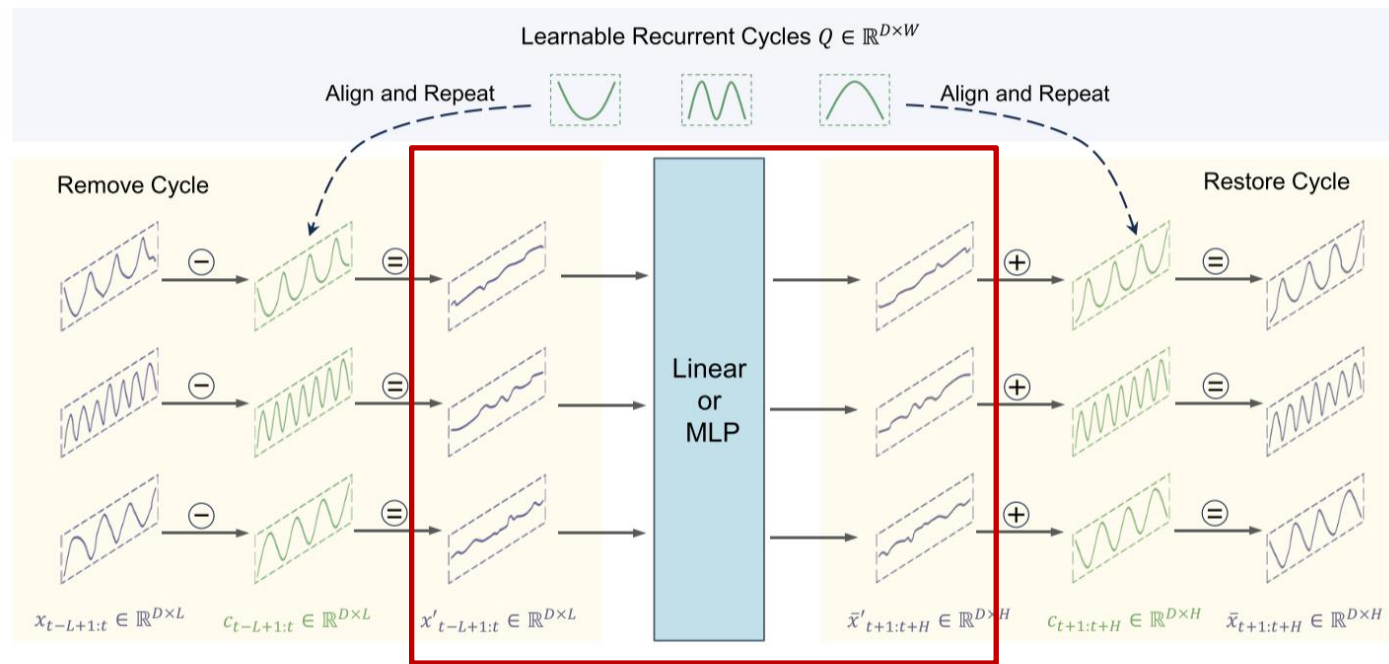
- 残差成分=原始序列-周期成分
- 通过主干获得对残差分量的预测
- 主干是单层Linear或双层MLP，可以换成任何其他时序预测模型
- 通道无关：不显式捕捉每个通道变量之间的关联性

- 数据预处理

- 实例归一化
- 缓解时序分布漂移问题

$$\mathbf{x}_{t-L+1:t} = \frac{\mathbf{x}_{t-L+1:t} - \mu}{\sqrt{\sigma + \varepsilon}}$$

$$\bar{\mathbf{x}}_{t+1:t+H} = \bar{\mathbf{x}}_{t+1:t+H} \times \sqrt{\sigma + \varepsilon} + \mu$$



- 数据集

数据集	长度	特征数量	采样频率	周期模式	周期长度
ETTh1&ETTh2	17,420	7	1 hour	Daily	24
ETTm1&ETTm2	69,680	7	15 mins	Daily	96
Electricity	26,304	321	1 hour	Daily & Weekly	168
Solar-Energy	52,560	137	10 mins	Daily	144
Traffic	17,544	862	1 hour	Daily & Weekly	168
Weather	52,696	21	10 mins	Daily	144

- 输入长度 $L=96$ ，输出长度 $H=\{96, 192, 336, 720\}$ ，预测结果取均值
- 评价指标：MSE，MAE
- 对比方法
 - 基于Transformer：iTransformer (2024)，PatchTST (2023)，Crossformer (2023)，**FEDformer** (2022)，**Autoformer** (2021)
 - 分解方法：TimesNet (2023)，DLinear (2023)
 - 传统深度学习：TimeMixer (2024)，SciNet (2022)，TiDE (2023)



- 在大多数数据集和预测范围上达到了最优或次优性能
- 在Electricity、Solar-Energy等周期性显著的数据集上优势明显
- 仅结合简单线性模型或浅层MLP，证明了RCF技术的强大
- 局限性分析：Traffic数据集表现较差，交通数据具有强烈的时空依赖性和时滞特性，通道无关策略无法有效捕捉变量间关系

Dataset	ETTh1		ETTh2		ETTm1		ETTm2		Electricity		Solar-Energy		Traffic		Weather	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Autoformer [2021]	0.496	0.487	0.450	0.459	0.588	0.517	0.327	0.371	0.227	0.338	0.885	0.711	0.628	0.379	0.338	0.382
FEDformer [2022]	0.440	0.460	0.437	0.449	0.448	0.452	0.305	0.349	0.214	0.327	0.291	0.381	0.610	0.376	0.309	0.360
SCINet [2022]	0.747	0.647	0.954	0.723	0.485	0.481	0.571	0.537	0.268	0.365	0.282	0.375	0.804	0.509	0.292	0.363
DLinear [2023]	0.456	0.452	0.559	0.515	0.403	0.407	0.350	0.401	0.212	0.300	0.330	0.401	0.625	0.383	0.265	0.317
TimesNet [2023]	0.458	0.450	0.414	0.427	0.400	0.406	0.291	0.333	0.192	0.295	0.301	0.319	0.620	0.336	0.259	0.287
TiDE [2023]	0.541	0.507	0.611	0.550	0.419	0.419	0.358	0.404	0.251	0.344	0.347	0.417	0.760	0.473	0.271	0.320
Crossformer [2023]	0.529	0.522	0.942	0.684	0.513	0.496	0.757	0.610	0.244	0.334	0.641	0.639	0.550	0.304	0.259	0.315
PatchTST [2023]	0.469	0.454	0.387	0.407	0.387	0.400	0.281	0.326	0.205	0.290	0.270	0.307	0.481	0.304	0.259	0.281
TimeMixer [2024]	0.447	0.440	0.364	0.395	0.381	0.395	0.275	0.323	0.182	0.272	0.216	0.280	0.484	0.297	0.240	0.271
iTransformer [2024]	0.454	0.447	0.383	0.407	0.407	0.410	0.288	0.332	0.178	0.270	0.233	0.262	0.428	0.282	0.258	0.278
CycleNet/Linear	0.432	0.427	0.383	0.404	0.386	0.395	0.272	0.315	0.170	0.260	0.235	0.270	0.485	0.313	0.254	0.279
CycleNet/MLP	0.457	0.441	0.388	0.409	0.379	0.396	0.266	0.314	0.168	0.259	0.210	0.261	0.472	0.301	0.243	0.271

- L=96, H=720, 每个epoch训练的平均值
- 比较参数量、MACs（乘加运算次数）和训练时间
- CycleNet/MLP的参数量比iTransformer少一个数量级（超过90%）
- 训练速度逊色于Dlneat, 考虑到带来的精度提升, CycleNet实现了精度和效率的平衡

Model	Parameters	MACs	Training Time(s)
Informer [2021]	12.53M	3.97G	70.1
Autoformer [2021]	12.22M	4.41G	107.7
FEDformer [2022]	17.98M	4.41G	238.7
DLinear [2023]	139.6K	44.91M	18.1
PatchTST [2023]	10.74M	25.87G	129.5
iTransformer [2024]	5.15M	1.65G	35.1
CycleNet/MLP	472.9K	134.84M	30.8
CycleNet/Linear	123.7K	22.42M	29.6
RCF part	53.9K	0	12.8



- RCF的有效性
 - 加入RCF技术后，简单的Linear/MLP骨干，预测精度提升约10%-20%
 - 复杂的PatchTST/iTransformer模型精度提升约5%-10%
 - 采用基于MOV的STD技术的DLinear，提供大约20%的改进，说明了区域合作框架的有效性和可移植性
 - 部分MAE降低，MSE提高：数据集含有极端值，干扰RCF学习历史平均周期

Dataset	Electricity								Traffic							
Horizon	96		192		336		720		96		192		336		720	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Linear	0.197	0.274	0.197	0.277	0.212	0.292	0.253	0.324	0.645	0.383	0.598	0.361	0.605	0.362	0.643	0.381
+ RCF	0.141	0.234	0.155	0.247	0.172	0.264	0.210	0.296	0.480	0.314	0.482	0.313	0.476	0.303	0.503	0.320
Improve	28.6%	14.6%	21.4%	10.8%	18.8%	9.5%	17.1%	8.7%	25.6%	18.0%	19.5%	13.2%	21.3%	16.2%	21.8%	16.1%
MLP	0.175	0.259	0.181	0.265	0.197	0.282	0.240	0.317	0.500	0.325	0.496	0.321	0.509	0.325	0.542	0.342
+ RCF	0.136	0.229	0.152	0.244	0.170	0.264	0.212	0.299	0.458	0.296	0.457	0.294	0.470	0.299	0.502	0.314
Improve	22.2%	11.6%	15.9%	8.0%	13.6%	6.3%	11.6%	5.7%	8.5%	8.9%	7.9%	8.3%	7.7%	8.0%	7.3%	8.1%
DLinear	0.195	0.278	0.194	0.281	0.207	0.297	0.243	0.331	0.649	0.398	0.599	0.372	0.606	0.375	0.646	0.396
+ RCF	0.143	0.240	0.156	0.253	0.171	0.270	0.204	0.302	0.506	0.317	0.499	0.317	0.512	0.325	0.545	0.343
Improve	26.6%	13.6%	19.7%	10.0%	17.4%	8.9%	16.3%	8.8%	22.1%	20.4%	16.6%	14.6%	15.4%	13.3%	15.6%	13.5%
PatchTST	0.168	0.260	0.176	0.266	0.193	0.282	0.233	0.317	0.436	0.281	0.449	0.285	0.464	0.293	0.499	0.310
+ RCF	0.136	0.231	0.153	0.246	0.170	0.264	0.211	0.299	0.438	0.264	0.457	0.270	0.469	0.275	0.509	0.292
Improve	19.0%	11.0%	13.0%	7.6%	11.7%	6.6%	9.4%	5.7%	-0.5%	6.1%	-1.8%	5.5%	-1.0%	6.3%	-2.0%	6.1%
iTransformer	0.148	0.240	0.162	0.253	0.178	0.269	0.225	0.317	0.395	0.268	0.417	0.276	0.433	0.283	0.467	0.302
+ RCF	0.136	0.231	0.153	0.247	0.168	0.263	0.194	0.287	0.415	0.263	0.440	0.271	0.456	0.278	0.491	0.294
Improve	8.1%	3.7%	5.6%	2.4%	5.8%	2.2%	13.8%	9.5%	-5.1%	1.9%	-5.5%	1.8%	-5.3%	1.8%	-5.1%	2.6%

- RCF序列分解的有效性

- 将Linear和三种序列分解方法结合，并比较性能
- 传统分解方法缺点：
 - 移动平均的滑动窗口需要大于季节性分量的**最大周期**；边缘失真
 - 等价于无约束或弱约束线性回归，在完全训练收敛后，在理论上等价于线性模型
- RCF是从整个训练集**全局学习周期模式**，而非局限于有限的历史窗口，从而能更准确地捕捉长期周期性

Setup	CLinear (RCF+Linear)		LDLinear (LD+Linear)		DLinear (MOV+Linear)		SLinear (Sparse+Linear)		Linear	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	0.418	0.434	0.427	0.439	0.425	0.437	<u>0.424</u>	<u>0.436</u>	0.427	0.439
ETTh2	0.451	0.456	<u>0.455</u>	<u>0.457</u>	0.471	0.467	0.460	0.460	0.460	0.462
ETTm1	0.349	0.382	0.365	0.387	0.367	0.390	<u>0.362</u>	<u>0.383</u>	0.362	0.384
ETTm2	0.266	0.330	0.273	0.336	0.280	0.341	0.290	0.352	<u>0.269</u>	<u>0.331</u>
Electricity	0.157	0.255	<u>0.167</u>	<u>0.264</u>	0.167	0.264	0.172	0.268	0.167	0.265
Solar-Energy	0.220	0.259	<u>0.253</u>	0.316	0.254	0.318	0.255	<u>0.315</u>	0.253	0.318
Traffic	0.423	0.289	0.434	0.296	<u>0.434</u>	0.296	0.435	<u>0.292</u>	0.434	0.296
Weather	0.245	0.300	<u>0.244</u>	<u>0.297</u>	0.244	0.296	0.246	0.298	0.245	0.297

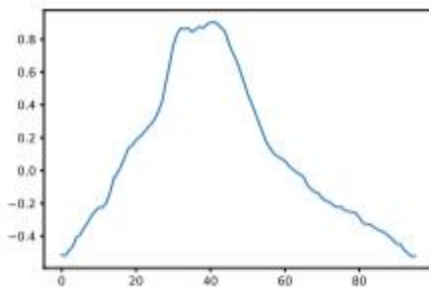
- 周期长度W

- 将超参数W正确设置为数据集的最大周期长度，取得最佳效果
- 进一步通过自相关函数（ACF）检验数据集周期
- 当模型失配时，模型的性能几乎与根本不使用RCF时相同
- 即使在最糟糕的情况下，RCF也不会带来重大的负面影响

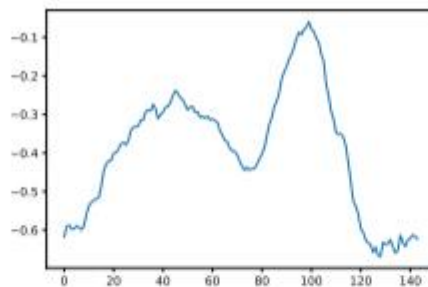
Setup	RCF/W=168		RCF/W=144		RCF/W=96		RCF/W=24		W/o. RCF	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Electricity	0.142	0.234	0.196	0.275	0.196	0.274	0.195	0.274	0.197	0.274
Traffic	0.480	0.314	0.617	0.386	0.617	0.385	0.618	0.385	0.645	0.383
Solar-Energy	0.289	0.376	0.208	0.256	0.276	0.365	0.287	0.375	0.286	0.375
ETTm1	0.350	0.369	0.340	0.366	0.325	0.363	0.348	0.367	0.351	0.372
ETTh1	0.395	0.402	0.384	0.395	0.383	0.393	0.377	0.391	0.384	0.392

- 周期模式可视化

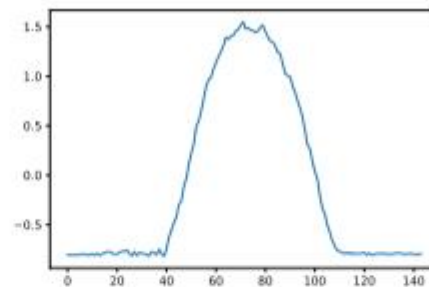
- (a-d)显示从不同数据集学习的不同周期模式
- (e-h)显示从同一数据集中的**不同通道**学习的不同周期模式



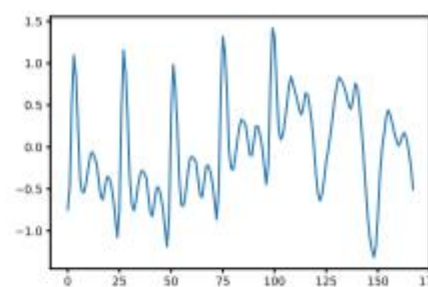
(a) ETTm1, 7th



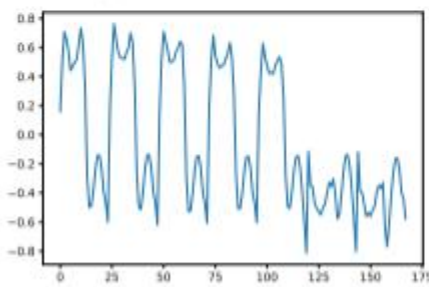
(b) Weather, 7th



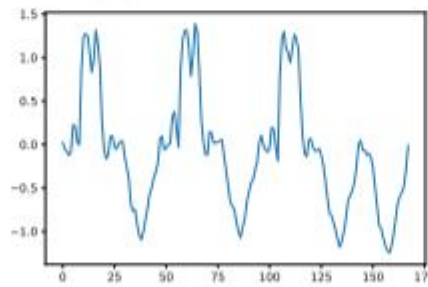
(c) Solar-Energy, 137th



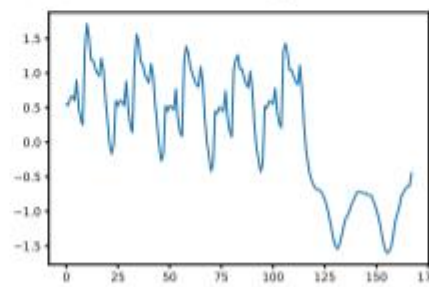
(d) Traffic, 607th



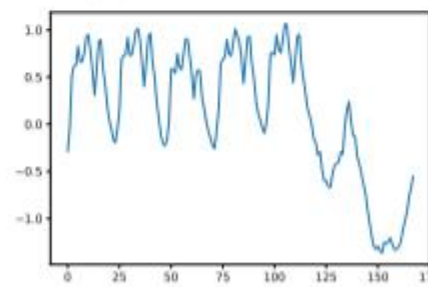
(e) Electricity, 311st



(f) Electricity, 318th



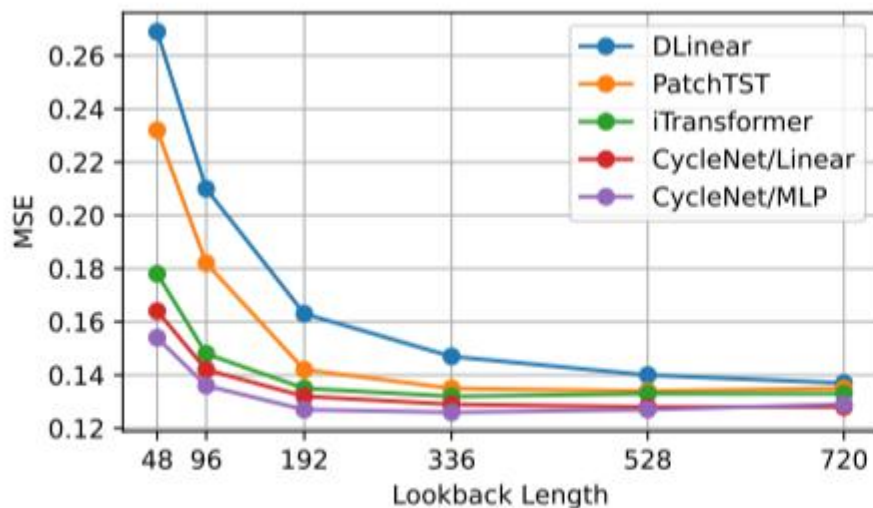
(g) Electricity, 320th



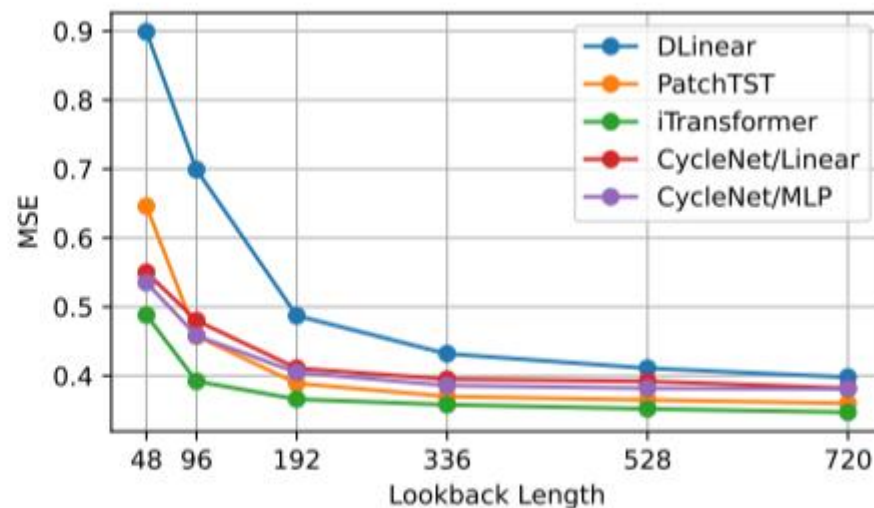
(h) Electricity, 321st

- 长时窗口回溯实验

- 通过更长的回顾长度实现更好的性能，表明模型具有较强的长期依赖建模能力
- 在电力数据集上，CycleNet在任何预测长度上都优于当前最先进的模型
- 在交通数据集上，CycleNet仍然落后于现有强大的多变量预测模型
- 直接将RCF技术应用于iTransformer并不会带来显著的改进



(a) Electricity



(b) Traffic

特点总结与未来展望



特点总结与未来展望

- 特点总结

- Periodformer

- 显式学习长期周期模式，**周期注意力**聚焦周期中类似阶段相对应的相关时间点
 - 实现线性复杂度，通过对齐周期模式强化重复出现的关键信号，有效维持长期依赖关系

- CycleNet

- 通过**可学习的循环单元**捕捉序列内在周期模式，并对残差进行预测
 - 方法简单高效，全局学习周期模式，更准确地捕捉长期周期性

- 不足分析

- 不适合周期**长度不稳定的**、每个**通道周期不同**数据集
 - 对**更长的依赖关系**（如一年）无能为力，数据可能需要数十年
 - 对于CycleNet，解决**极端离群值**问题和有效建模**多变量关系**是未来重要的发展方向

- [1]. Lin S, Lin W, Hu X, et al. Cyclenet: Enhancing time series forecasting through modeling periodic patterns[J]. Advances in Neural Information Processing Systems, 2024, 37: 106315-106345.**
- [2]. Liang D, Zhang H, Yuan D, et al. Periodformer: An efficient long-term time series forecasting method based on periodic attention[J]. Knowledge-Based Systems, 2024, 304: 112556.**
- [3]. 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.**

道可道，非常道。名可
名，非常名。无名天地
之始。有名万物之母。
故常无欲以观其妙。常
有欲以观其徼。此两者
同出而异名，同谓之玄。
玄之又玄，众妙之门。

谢谢！

