北京航空航天大學
BEIHANG UNIVERSITY

FEDformer_MindSpore

——用FEDformer求解时间序列预测问题

2025年06月16日

小组成员：谢浩志_ZY2406222 黄星阳_ZY2406437

北京航空航天大学
BEIHANG UNIVERSITY

目录

CONTENTS

德才兼备 知行合一

背景与问题定义

☐ 长时间序列预测的背景

- **数据难题：**长时间序列通常具有跨度大、特征维度高、噪声干扰强等特点，增加了建模难度与预测不确定性。
- **模型难题：**传统模型如RNN在长序列建模中易出现梯度消失或爆炸，难以捕捉长期依赖结构，预测精度受限。

☐ 问题形式化定义

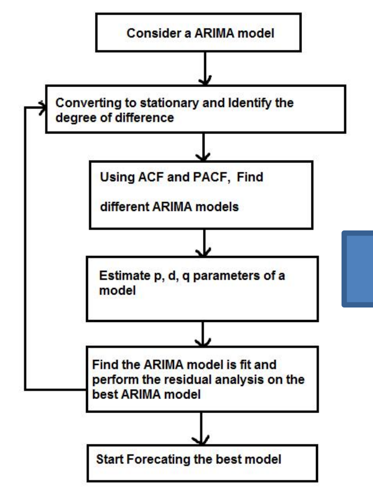
- 给定历史序列为 $X = x_1, x_2, ..., x_I \in \mathbb{R}^{I \times D}$，目标是预测未来序列 $Y = y_1, y_2, ..., y_O \in \mathbb{R}^{O \times D}$。
- 任务形式为Encoder-Decoder序列建模的过程，需要最小化预测值和真实值的误差，即

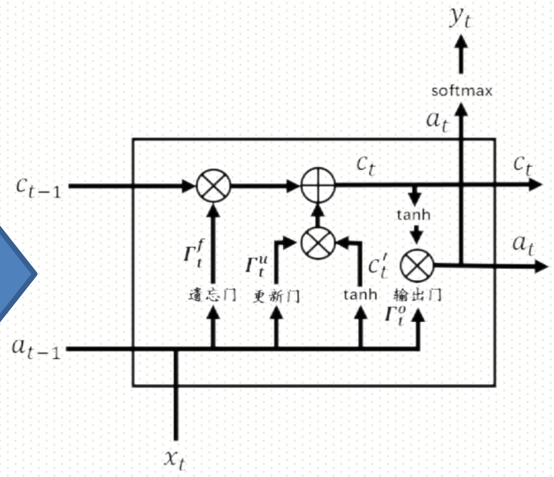
$$\min_\theta \mathcal{L}(\theta) = \frac{1}{O} \sum_{t=1}^{O} \|\hat{y}_{I+t} - y_{I+t}\|^2$$

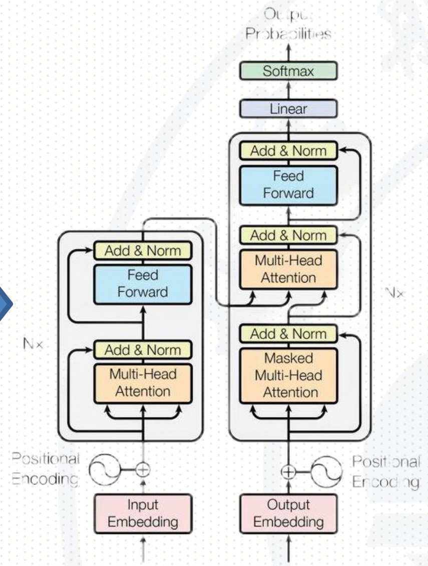
时间序列预测面临数据复杂与模型建模能力双重挑战

研究现状与主要挑战

□ 时间序列相关研究



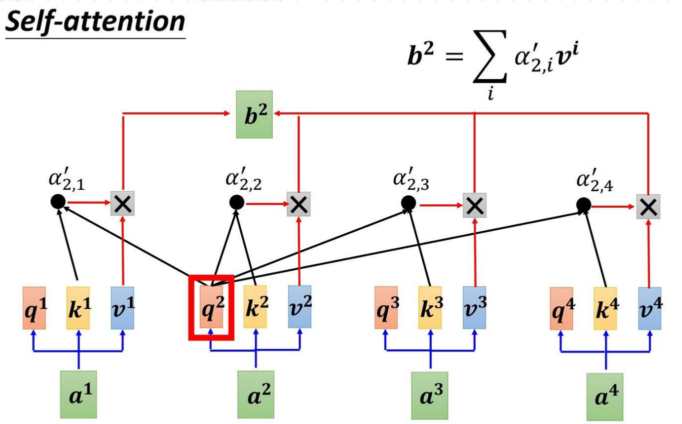




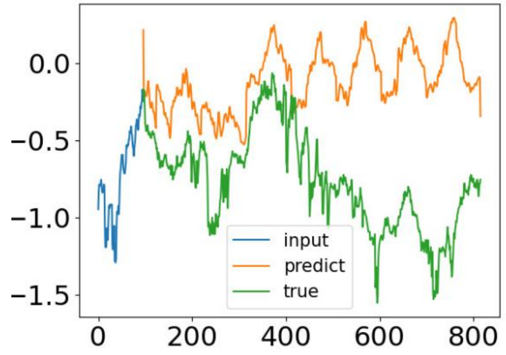
传统模型（ARIMA）：无法建模非线性关系

RNN-based模型：难以捕捉长期依赖

Transformer模型：使用较为广泛



挑战1：自注意力计算开销大



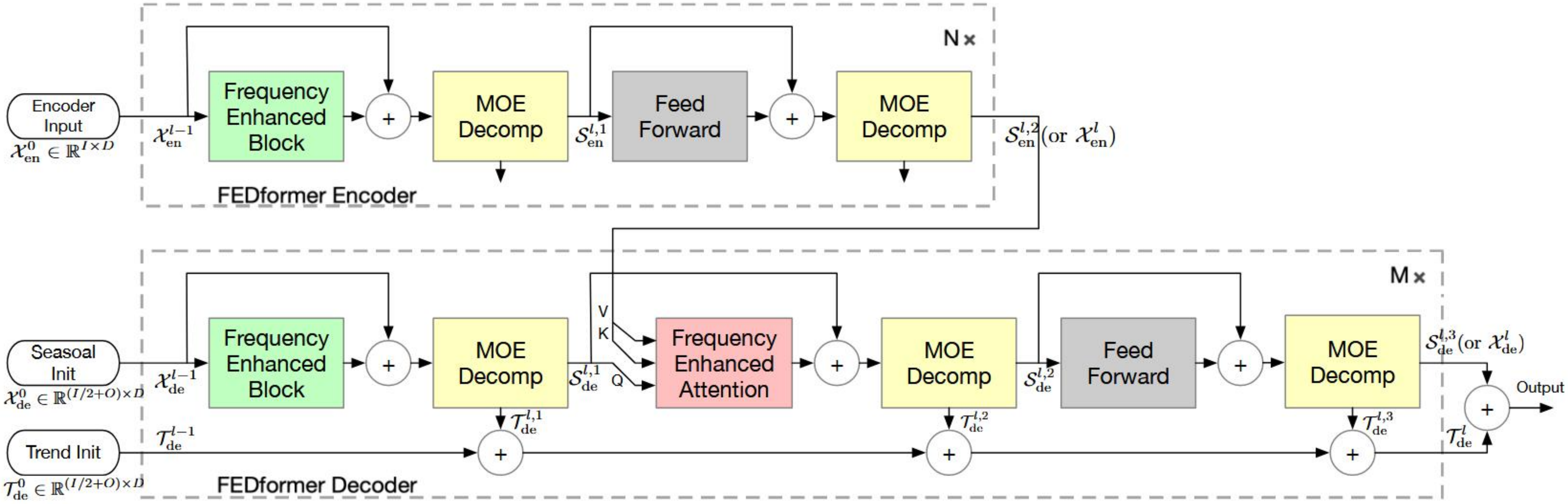
挑战2：难以捕捉季节/趋势特征

研究核心难点：降低计算成本、捕捉时序数据的全局结构特征

论文方法介绍

北京航空航天大學 BEIHANG UNIVERSITY

□ FEDformer模型结构与核心机制



- **核心结构1：**引入序列分解机制，将预测任务分解为趋势建模与季节建模【把握全局特征】；
- **核心结构2：**使用 Fourier / Wavelet 变换在频域执行注意力计算，减少复杂度至 $O(L)$【加速】。

FEDformer方案：融合分解建模与频域注意力，实现高效长期预测

复现挑战与解决方法

□ 迁移适配主要问题

- **问题一：** 昇腾平台不支持 FFTWithSize，频域模块无法运行
    - 迁移至 NVIDIA Tesla P100 显卡环境执行；
- **问题二：** Grad 算子不支持复数
    - 将复数计算模块封装为实部/虚部分离表示，规避反向传播问题；
- **问题三：** 图模式支持有限
    - 整体采用 PYNATIVE 模式运行，确保训练稳定性与模块兼容性。

```python
def complex_mul(a_real, a_imag, b_real, b_imag, mode):
    """使用实数表示执行复数乘法"""
    einsum = ops.Einsum(mode)
    real = einsum((a_real, b_real)) - einsum((a_imag, b_imag))
    imag = einsum((a_real, b_imag)) + einsum((a_imag, b_real))
    return real, imag  # 返回实部和虚部的实数张量
```

复数计算封装

```python
def _set_device(self):
    if self.args.use_gpu:
        context.set_context(mode=context.PYNATIVE_MODE, device_target="GPU", dev
        # context.set_context(mode=context.GRAPH_MODE, device_target="GPU", devi
        # context.set_context(mode=context.PYNATIVE_MODE, device_target="Ascend"
        print(f"Using GPU: {self.args.gpu}")
    else:
        context.set_context(mode=context.PYNATIVE_MODE, device_target="CPU")
        print("Using CPU")
```

运行模式设定

模型迁移：规避框架限制

实验结果与分析

□ 实验设置与结果分析

- ➢ **实验设置：**
  - ➢ 对比模型：Transformer、Informer、Autoformer；
  - ➢ 数据集：Electricity, Exchange, Traffic, Weather；
  - ➢ 设置：预测步长为96，多变量预测任务；
- ➢ **结果分析：**
  - ➢ FEDformer在Exchange与Weather数据集表现领先，针对**季节性趋势性数据预测效果良好**；
  - ➢ 与原文结果存在差距。原因包括**平台差异、复数运算近似**、动态图运行限制等；
  - ➢ 总体趋势一致，验证了FEDformer频域建模在长序列预测中的有效性与鲁棒性。

原始结果表

Methods	Metric	Electricity 96	Electricity 192	Electricity 336	Electricity 720	Exchange 96	Exchange 192	Exchange 336	Exchange 720	Traffic 96	Traffic 192	Traffic 336	Traffic 720	Weather 96	Weather 192	Weather 336	Weather 720
FEDformer-f	MSE	0.193	0.201	0.214	0.246	0.148	0.271	0.460	1.195	0.587	0.604	0.621	0.626	**0.217**	**0.276**	**0.339**	**0.403**
	MAE	0.308	0.315	0.329	0.355	0.278	0.380	0.500	0.841	0.366	0.373	0.383	0.382	**0.296**	**0.336**	**0.380**	**0.428**
FEDformer-w	MSE	**0.183**	**0.195**	**0.212**	**0.231**	**0.139**	**0.256**	**0.426**	**1.090**	**0.562**	**0.562**	**0.570**	**0.596**	0.227	0.295	0.381	0.424
	MAE	**0.297**	**0.308**	**0.313**	**0.343**	**0.276**	**0.369**	**0.464**	**0.800**	**0.349**	**0.346**	**0.323**	**0.368**	0.304	0.363	0.416	0.434
Autoformer	MSE	0.201	0.222	0.231	0.254	0.197	0.300	0.509	1.447	0.613	0.616	0.622	0.660	0.266	0.307	0.359	0.419
	MAE	0.317	0.334	0.338	0.361	0.323	0.369	0.524	0.941	0.388	0.382	0.337	0.408	0.336	0.367	0.395	0.428
Informer	MSE	0.274	0.296	0.300	0.373	0.847	1.204	1.672	2.478	0.719	0.696	0.777	0.864	0.300	0.598	0.578	1.059
	MAE	0.368	0.386	0.394	0.439	0.752	0.895	1.036	1.310	0.391	0.379	0.420	0.472	0.384	0.544	0.523	0.741
LogTrans	MSE	0.258	0.266	0.280	0.283	0.968	1.040	1.659	1.941	0.684	0.685	0.7337	0.717	0.458	0.658	0.797	0.869
	MAE	0.357	0.368	0.380	0.376	0.812	0.851	1.081	1.127	0.384	0.390	0.408	0.396	0.490	0.589	0.652	0.675
Reformer	MSE	0.312	0.348	0.350	0.340	1.065	1.188	1.357	1.510	0.732	0.733	0.742	0.755	0.689	0.752	0.639	1.130
	MAE	0.402	0.433	0.433	0.420	0.829	0.906	0.976	1.016	0.423	0.420	0.420	423	0.596	0.638	0.596	0.792

复现结果表

方法	指标\数据集	Electricity	Exchange	Traffic	Weather
Transformer	MAE	**0.830**	1.418	**0.911**	1.580
	RMSE	**1.014**	1.784	**1.315**	2.066
	MAPE	**3.238**	12.247	**5.847**	25.460
Informer	MAE	0.938	1.589	0.963	1.743
	RMSE	1.133	1.954	1.353	2.233
	MAPE	5.320	9.693	7.462	31.488
Autoformer	MAE	1.698	**0.278**	1.941	2.379
	RMSE	2.016	**0.385**	2.265	3.797
	MAPE	13.925	**1.691**	25.638	87.032
FEDformer	MAE	1.326	0.278	1.795	**0.319**
	RMSE	1.562	0.386	2.101	**0.501**
	MAPE	10.921	1.699	24.045	**18.864**

实验验证结论：特定数据性能领先，与原框架实现有一定差异

参考文献

北京航空航天大學 BEIHANG UNIVERSITY

[1]Box G E P, Jenkins G M. Some recent advances in forecasting and control[J]. Journal of the Royal Statistical Society. Series C (Applied Statistics), 1968, 17(2): 91-109.

[2]Box G E P, Pierce D A. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models[J]. Journal of the American statistical Association, 1970, 65(332): 1509-1526.

[3]Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735-1780.

[4]Chung J, Gulcehre C, Cho K H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[J]. arXiv preprint arXiv:1412.3555, 2014.

[5]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.

[6]Qin Y, Song D, Chen H, et al. A dual-stage attention-based recurrent neural network for time series prediction[J]. arXiv preprint arXiv:1704.02971, 2017.

[7]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.

[8]Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.

[9]Beltagy I, Peters M E, Cohan A. Longformer: The long-document transformer[J]. arXiv preprint arXiv:2004.05150, 2020.

[10]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.

[11]Zhu Z, Soricut R. H-transformer-1d: Fast one-dimensional hierarchical attention for sequences[J]. arXiv preprint arXiv:2107.11906, 2021.

[12]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. 2021, 35(12): 11106-11115.

[13]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.

[14]Wang S, Li B Z, Khabsa M, et al. Linformer: Self-attention with linear complexity[J]. arXiv preprint arXiv:2006.04768, 2020.

[15]Ma X, Kong X, Wang S, et al. Luna: Linear unified nested attention[J]. Advances in Neural Information Processing Systems, 2021, 34: 2441-2453.

[16]Xiong Y, Zeng Z, Chakraborty R, et al. Nyströmformer: A nyström-based algorithm for approximating self-attention[C]//Proceedings of the AAAI conference on artificial intelligence. 2021, 35(16): 14138-14148.

[17]Choromanski K, Likhosherstov V, Dohan D, et al. Rethinking attention with performers[J]. arXiv preprint arXiv:2009.14794, 2020.

[18]Mathieu M, Henaff M, LeCun Y. Fast training of convolutional networks through ffts[J]. arXiv preprint arXiv:1312.5851, 2013.

[19]Wen Q, He K, Sun L, et al. RobustPeriod: Robust time-frequency mining for multiple periodicity detection[C]//Proceedings of the 2021 international conference on management of data. 2021: 2328-2337.

[20]Li Z, Kovachki N, Azizzadenesheli K, et al. Fourier neural operator for parametric partial differential equations[J]. arXiv preprint arXiv:2010.08895, 2020.

[21]Gupta G, Xiao X, Bogdan P. Multiwavelet-based operator learning for differential equations[J]. Advances in neural information processing systems, 2021, 34: 24048-24062.

[22]Rahimi A, Recht B. Random features for large-scale kernel machines[J]. Advances in neural information processing systems, 2007, 20.

[23]Rawat A S, Chen J, Yu F X X, et al. Sampled softmax with random fourier features[J]. Advances in Neural Information Processing Systems, 2019, 32.

[24]Lee-Thorp J, Ainslie J, Eckstein I, et al. Fnet: Mixing tokens with fourier transforms[J]. arXiv preprint arXiv:2105.03824, 2021.

北京航空航天大學

BEIHANG UNIVERSITY

Thanks!

艰苦朴素 勤奋好学 全面发展 勇于创新

2025年06月16日

小组成员：谢浩志_ZY2406222 黄星阳_ZY2406437