



Deep learning for time series forecasting: a survey

用于时间序列预测的深度学习：一项调查

汇报人：薛彤

2025/6/8

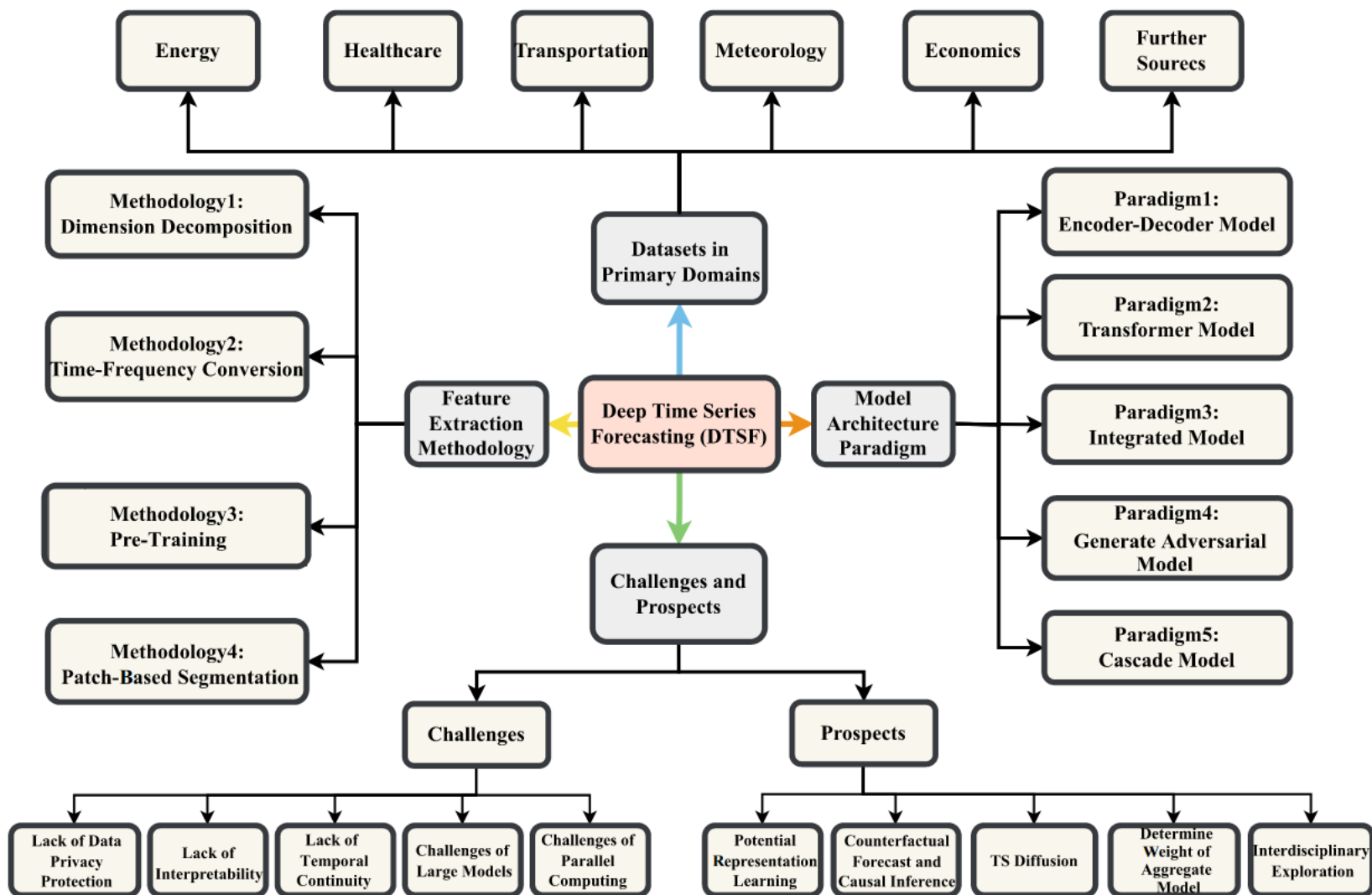
研究背景

■ 必要性

- 现有研究没有提供时序预测领域广泛模型架构的统一总结
- 也没有特征提取和数据集工作的详细摘要

■ 本文工作/贡献

- 动态和系统分类。从模型架构方面总结了深度时间序列预测（DTSF）的一般范式。
- 数据特征增强的综合回顾。总结了时间序列数据的特征增强方法，包括维度分解、时频变换、预训练和基于补丁的分割。
- 总结了近年来的主要 TSF 数据集。
- 强调了该领域面临的重大挑战和未来的研究方向。



时间序列预测

■ 定义

时间序列表示以规则或不规则时间间隔记录的一系列数据点，提供观察到的现象的时间顺序记录。这些观察的性质可以包含数值、标签等。

■ 特征

- ❑ 时间序列数据通常有噪声并且高维。降维、小波分析或滤波等技术可用于消除一些噪声和降低维数。
- ❑ 样本时间间隔对其有影响。不同采样频率下获得的时间序列分布不具有均匀的概率分布。
- ❑ 时间序列数据本质上是异构和动态的。

■ 预测任务

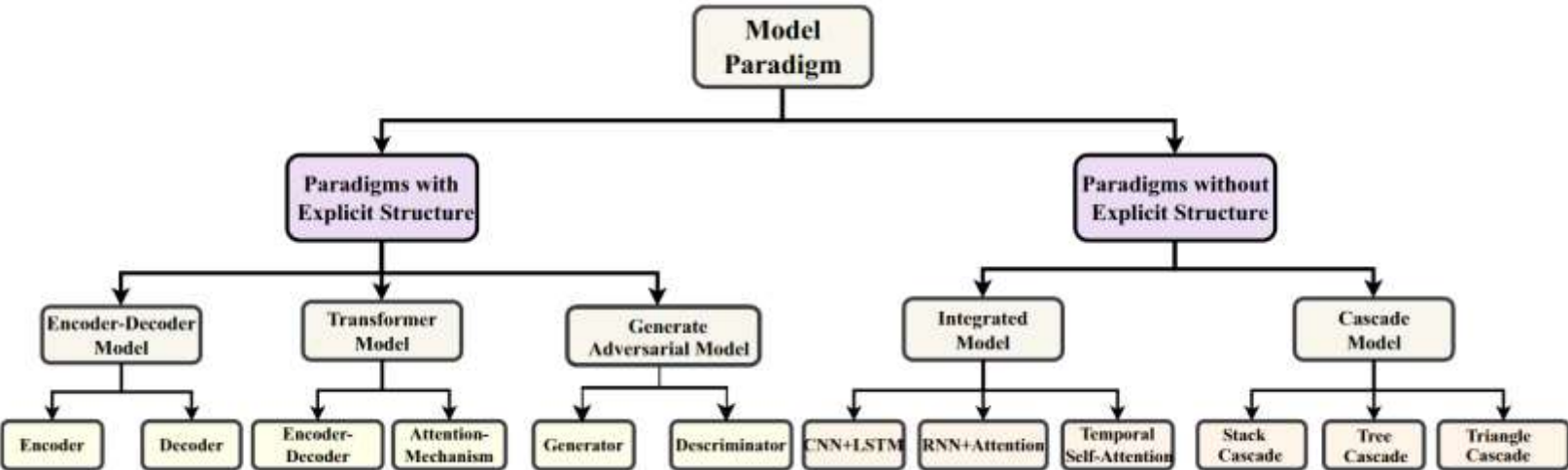
- ❑ 短期预测优先考虑精度，主要依赖于外推数据，适用于相对较短的时间内波动对预测结果至关重要的场景。
- ❑ 长期预测需要考虑长期趋势和季节影响，需要额外的因素，可能会影响准确性。
- ❑ 全局模型考虑整个时间序列数据集中的所有变量，而局部模型专注于数据的子集。

■ 统计预测模型

统计预测方法在处理非平稳性、非线性关系、噪声和复杂依赖关系方面存在较大的不足，对长期依赖关系和多特征预测任务的适应性也有限。

DTSF模型架构

■ 将五种模型架构分为显式结构范式和隐式结构范式。



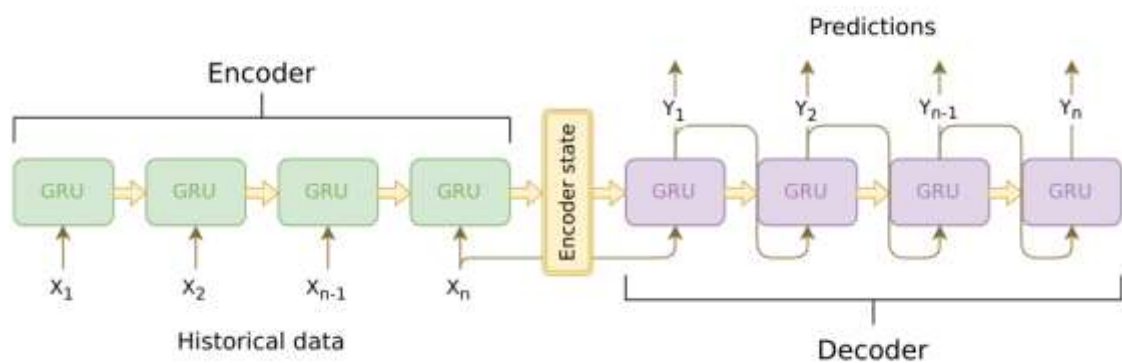
■ 几个经典的时间序列预测模型的比较分析：优点、缺点、应用和预测长度。

Model	Advantages	Disadvantages	Applications	Prediction horizon
Informer [292]	Efficient; strong representation; good generalization	Sensitive to data shifts	Energy; weather	Long
HANet [21]	Capture complex dependencies; flexible for multivariate data	High complexity	Weather; ecology	Long
Autoformer [251]	Efficient; good information	High complexity; depend on data periodicity	Finance; energy; electricity; traffic; weather; healthcare	Long
ETSformer [250]	Combine traditional methods with transformer; adaptive time window	High computational cost; requires large data	Finance; energy; electricity; traffic; weather; healthcare	Short
FEDformer [294]	Frequency enhancement; Better flexibility for long-term forecasts	High complexity; large data needed	Finance; energy; electricity; traffic; weather; healthcare	Long
TreeDRNet [295]	Capture time dynamics; efficient training with joint networks	High complexity; need large data	Finance; energy; electricity; traffic; weather; healthcare	Long
TATCN [242]	Capture temporal dependencies; extract local patterns	High computational cost; data dependence	Electricity; healthcare	Short

DTSF模型架构

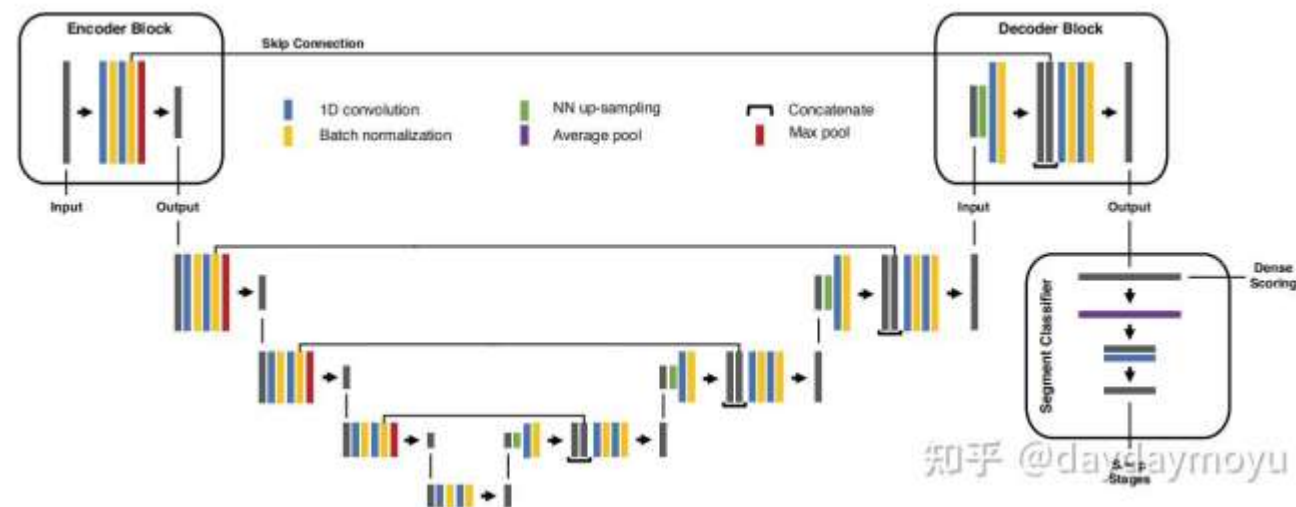
■ 显式结构模型：编码器-解码器模型

□ 经典模型：seq2seq，各种形式的Transformer，U-Time



seq2seq （2014年）模型概览

- 输入序列通过Encoder，得到语义向量C，语义向量C作为Decoder的初始状态h0，参与解码过程，生成输出序列。
- Seq2Seq使用的都是RNN单元，一般为LSTM和GRU。



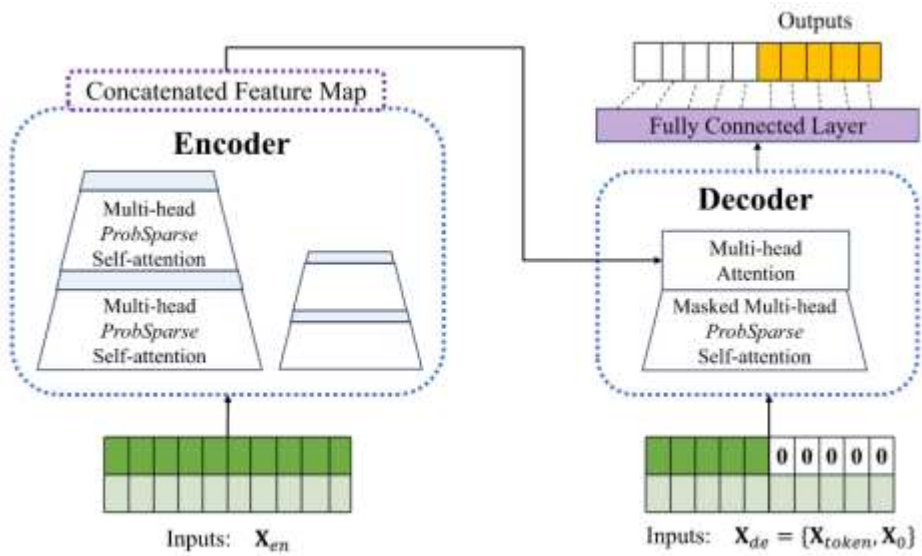
U-Time （NeurIPS 2019）模型概览

- 受 U-net 启发，含对称编解码器，区别在于，为了获得更大的Receptive field, 作者使用了更大的max-pooling window和Kernel size。
- 在大量睡眠脑电图 (EEG) 数据集上评估了用于睡眠阶段分类的 U-Time。

DTSF模型架构

- 显式结构模型：Transformer模型
- 经典模型：Transformer, Informer, Autoformer, TSMixer, ETSformer, FEDformer

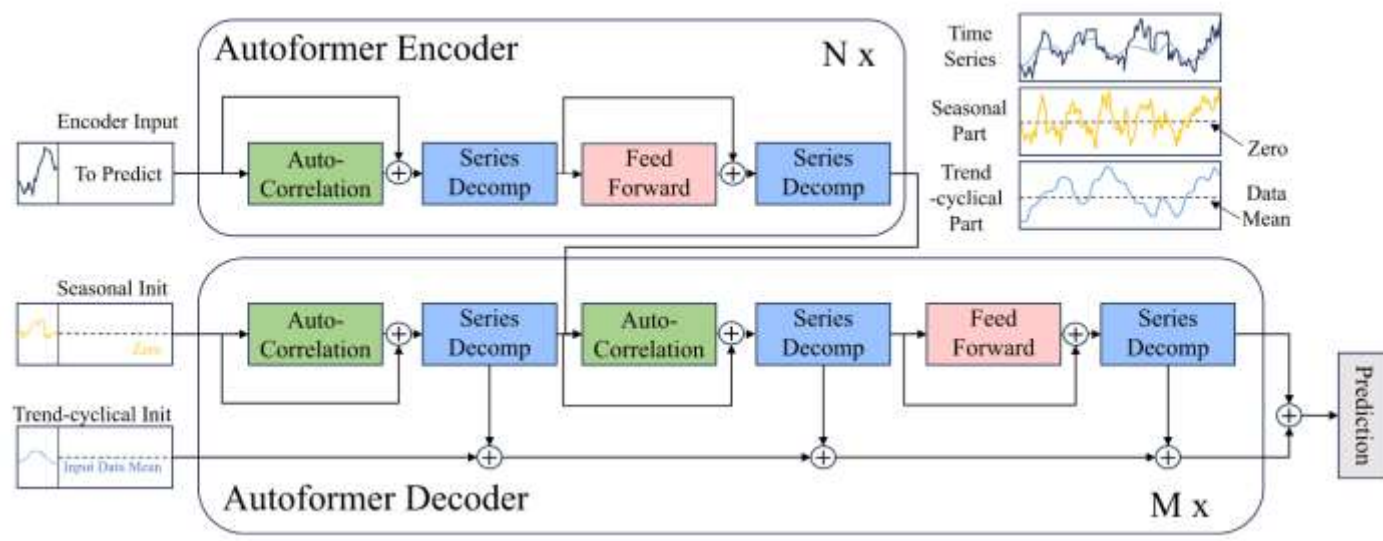
时间复杂度 $O(L^2)$ 时间复杂度 $O(L\log L)$, 自注意机制被ProbSparse自注意机制所取代 Decoder的输入是包含季节维度和趋势维度的子序列 将时间序列数据划分为多个patch, 使模型能够有效地捕获局部特征 设计了指平滑注意 (ESA) 和频率注意 (FA) 机制 结合了Transformer与季节趋势分解方法



Informer (AAAI'21 Best paper) 模型概览

- 较少的点积贡献了绝大部分的注意力得分。一个query的分布与均匀分布之间的差异可以用KL散度来度量, 选取 $M(q_{\{i\}}, K)$ 排名前u的若干query作为 \hat{Q} 。

$$A(Q, K, V) = \text{softmax} \left(\frac{\hat{Q}K^T}{\sqrt{d}} \right) V$$



Autoformer模型 (NeurIPS'21) 概览

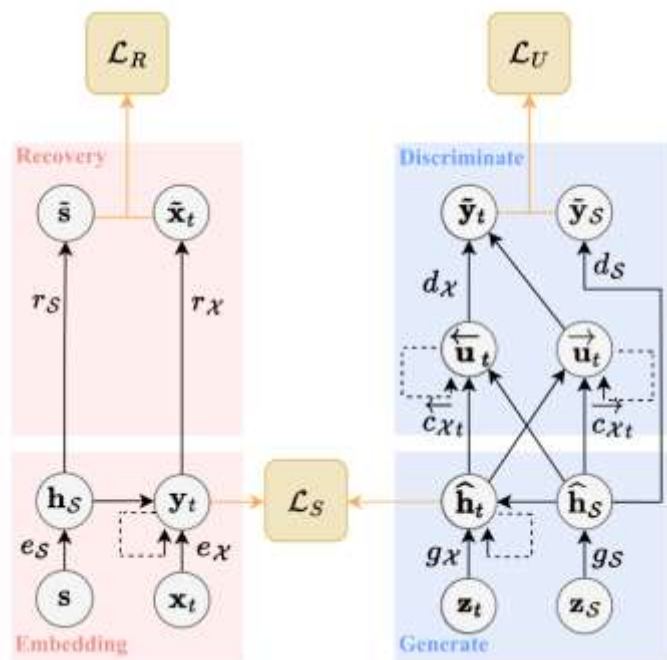
- 模型交替进行预测结果优化和序列分解, 即从隐变量中逐步分离趋势项与周期项, 实现渐进式分解。
- <https://zhuanlan.zhihu.com/p/385066440>

DTSF模型架构

■ 显式结构模型：生成对抗模型

□ 介绍：GAN（生成对抗网络）是一种生成模型，由包括鉴别器和生成器的显式结构组成。虽然 GAN 已广泛用于计算机视觉领域，但它们在 TSF 中的应用相对有限。

□ 经典模型：TimeGAN



TimeGAN（NeurIPS 2019）模型概览

■ 隐式结构模型：集成模型

□ 介绍：结合单个模型体系结构的优势，每个模型都专注于学习它擅长的特性。例如，卷积架构擅长学习局部特征模式，而循环架构擅长学习节点之间的时间依赖性。

□ 经典模型：TATCN 将卷积体系结构应用于序列领域

■ 隐式结构模型：级联模型

□ 介绍：级联网络通常由多个组件组成，每个组件都有不同的功能，共同形成更深入、更强大的网络模型。

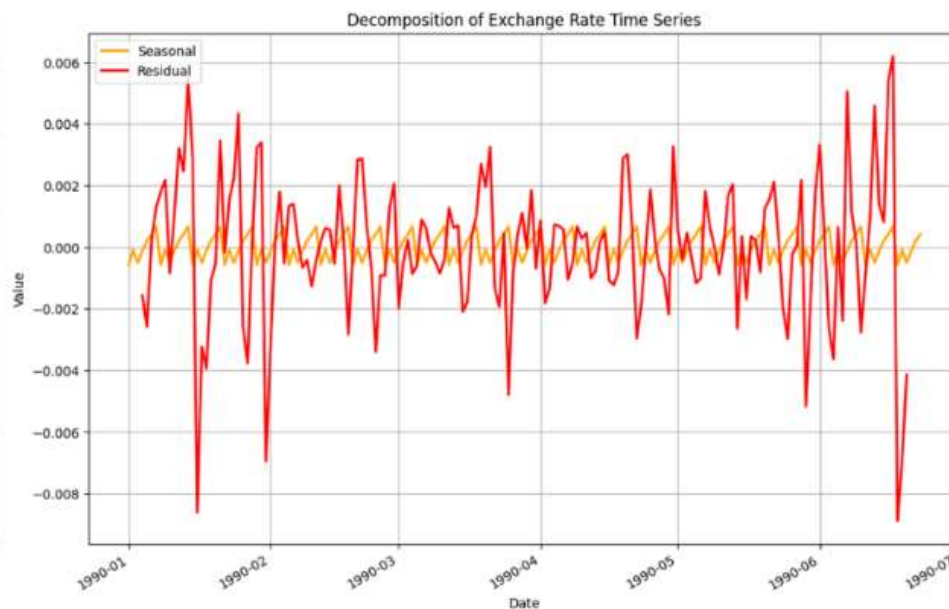
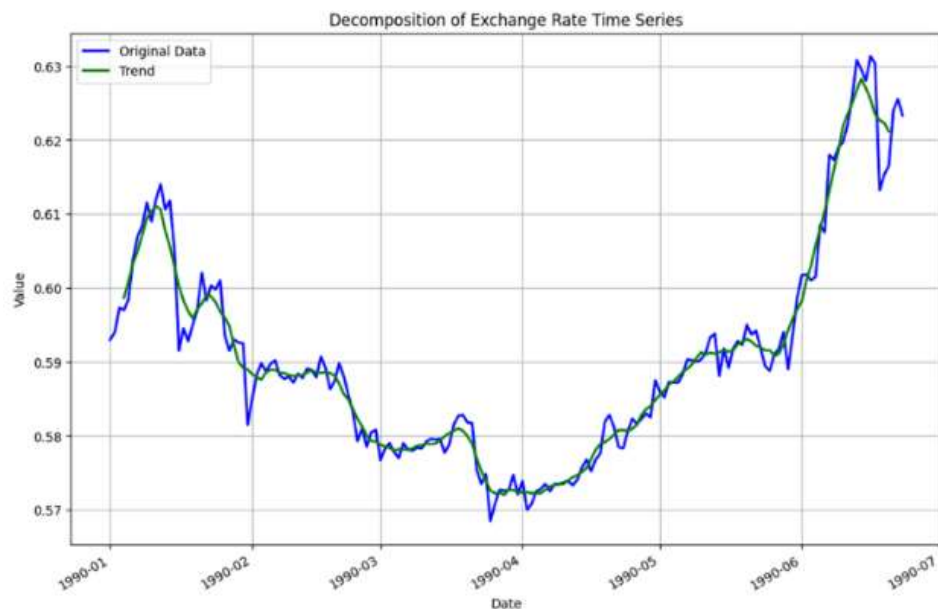
□ 经典模型：TreeDRNet 引入特征选择、模型集成和树结构

- 利用无监督模型的灵活性和有监督模型的控制性来生成真实的时间序列数据。
- 利用嵌入函数和恢复函数从时间序列数据中提取高维特征，然后将其输入序列生成器和鉴别器进行对抗训练。
- 鉴别器被训练来区分真实和生成的时间序列数据，目标是生成与真实数据无法区分的合成数据。
- <https://blog.csdn.net/Zcymatics/article/details/145011621>

时间序列的组成部分

时间序列数据可以分解为三个主要组成部分：趋势、季节性和残差或白噪声。

- 趋势：表示时间序列数据的长期变化。
- 季节性：参考时间序列数据中观察到的周期性变化，通常由季节性、每月、每周或其他时间单位影响引起。
- 残差：表示趋势和季节性分解后剩余的随机波动或噪声。残差反映了时间序列数据中没有建模的短期波动和不规则性。此外，残差表现出一些自相关，可以帮助我们识别和调整模型中潜在的缺陷。



时间序列的组件。数据来自1990年1月1日至1990年6月23日的Exchange-Rate数据集。蓝线表示原始数据，绿色表示趋势，黄色表示季节性，红色表示残差。

增强特征提取的方法

■ 方法一：维度分解

□ 经典模型：FEDformer, TDformer

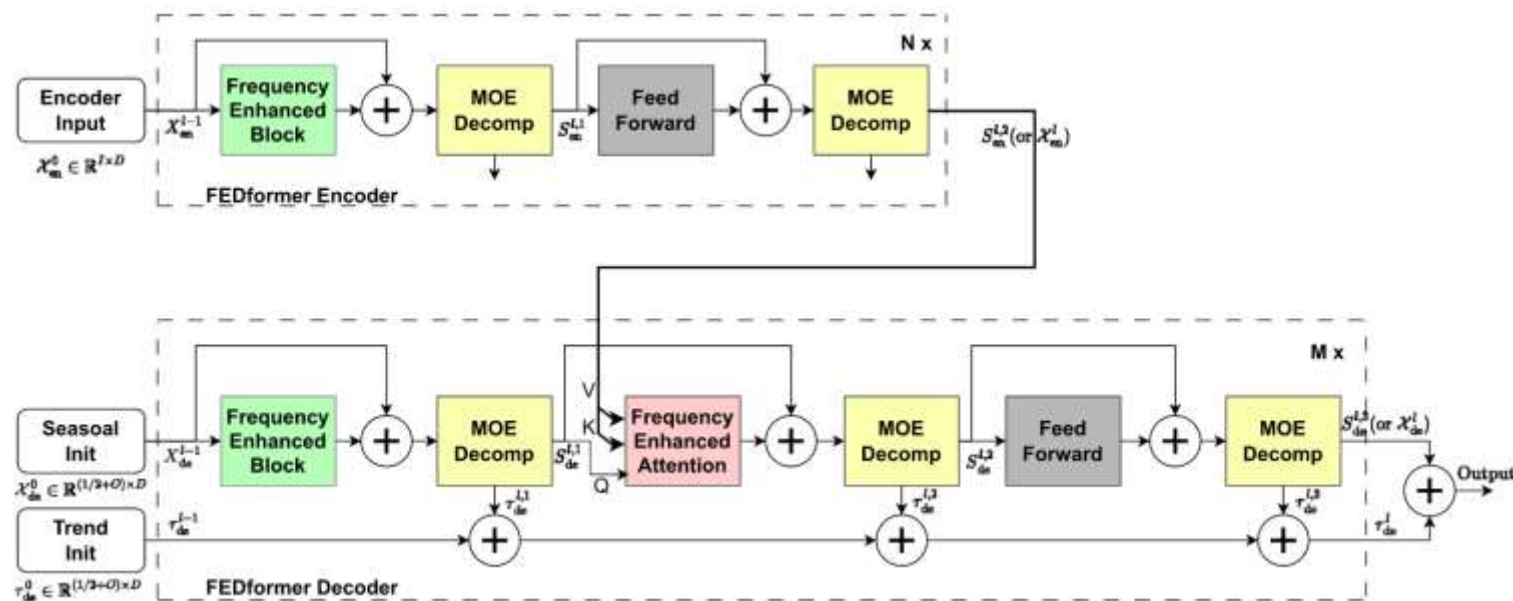
基于季节趋势的频率增强分解变压器体系结构
使用MLP对趋势进行建模，并使用傅里叶关注来模拟季节性

■ 方法二：时频转换

□ 介绍：将时域数据转换为其频域表示，从而能够更有效地分析时间序列数据内的频率、光谱特征和动态变化

□ 经典模型：FreTS, FiLM, FEDformer

由域转换和频率学习两个阶段组成
结合傅里叶分析和低秩矩阵近似实现了降噪



- 基于autoformer进行改进的。
- 使用了 low-rank approximation 而可以达到线性复杂度。
- <https://zhuanlan.zhihu.com/p/528131016>

FEDformer (ICML 2022) 模型概述

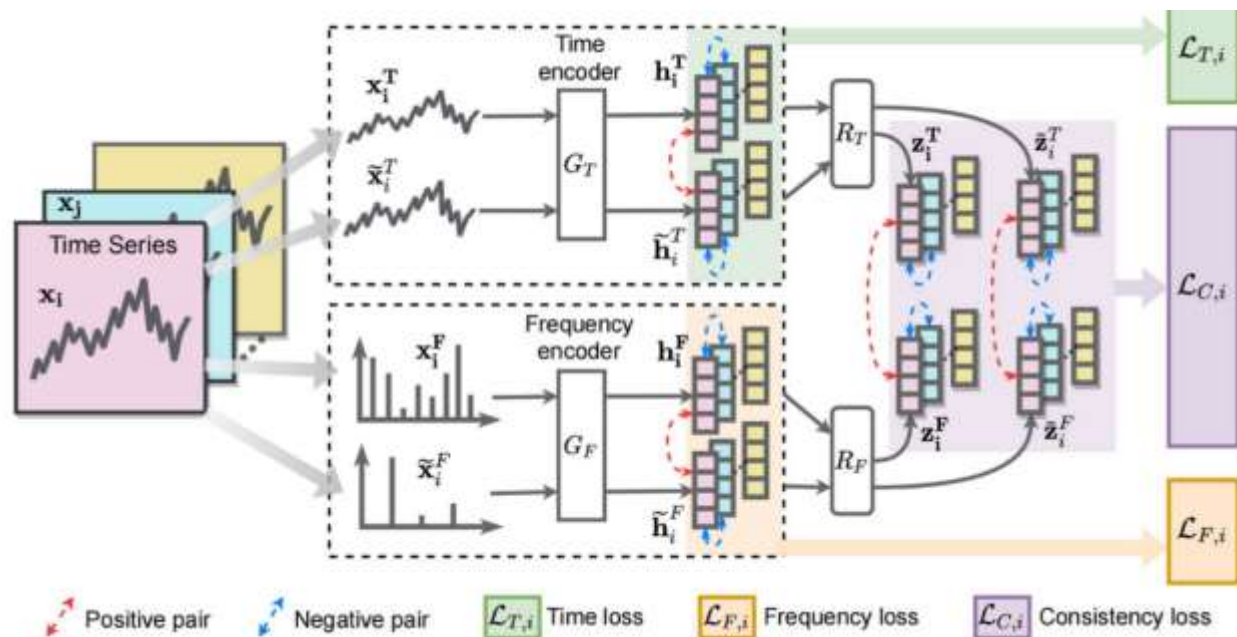
增强特征提取的方法

■ 方法三：预训练

□ 分类：

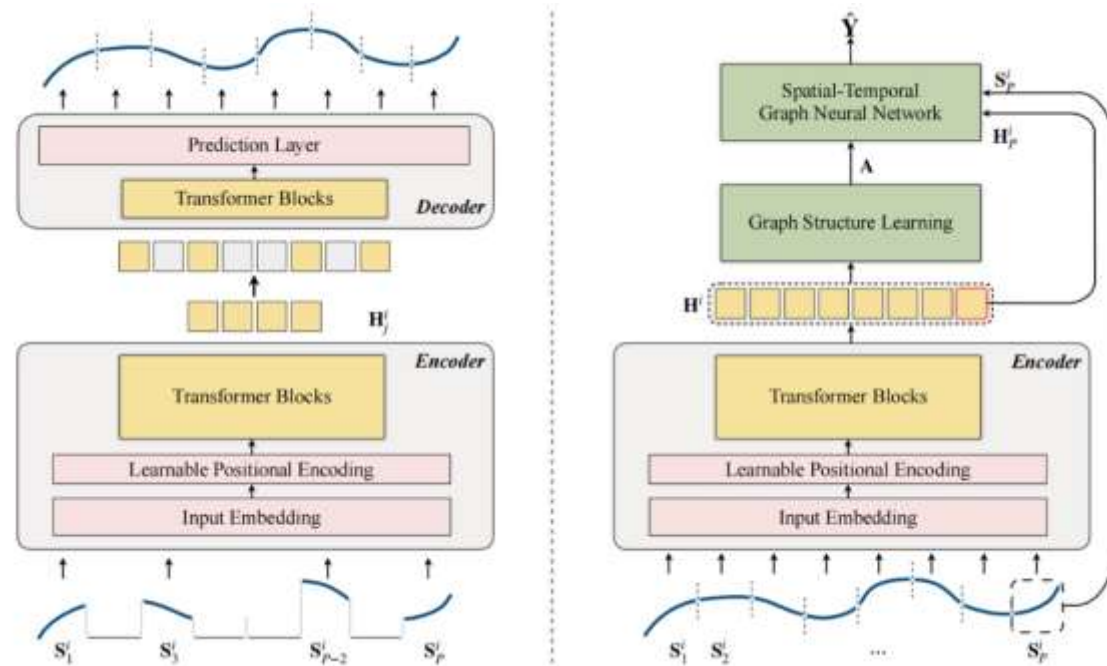
- 对比预训练：是一种自我监督学习形式，旨在训练输入编码器将正样本对映射得更近，将负样本对映射到更远。
- 屏蔽预训练：时间序列数据通常是连续的、有序的，但实际上表现出不完整性。此外，现实世界的时间序列数据通常包含噪声和不确定性，需要模型具有处理这种不确定性的稳健性。

□ 经典模型：TF-C, STEP



TF-C (NeurIPS 2022) 的概述

- 基于频率的对比增强来利用丰富的频谱信息并探索时间序列中的时频一致性。
- <https://zhuanlan.zhihu.com/p/650709265>



STEP (KDD 2022) 概述

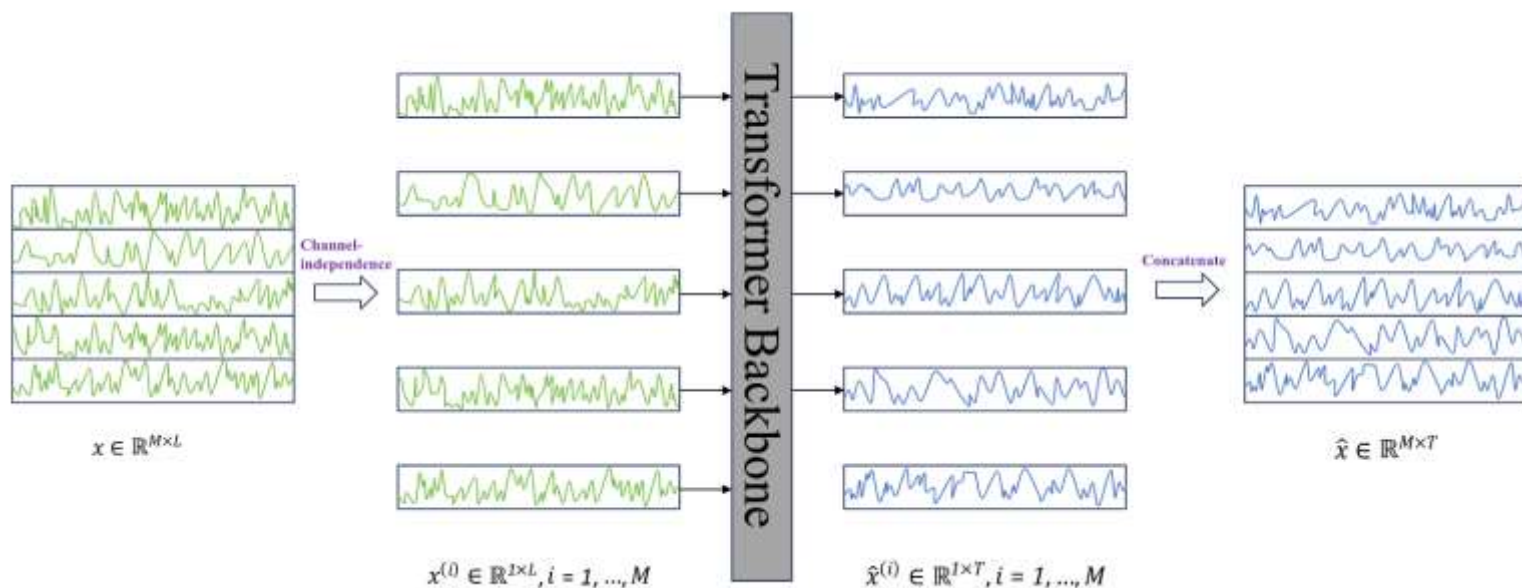
- 采用掩码自动编码策略进行训练。
- https://blog.csdn.net/weixin_45378275/article/details/128951452

增强特征提取的方法

■ 方法四：基于补丁的分割

□ 介绍：通过将长时间序列分割成更小的片段，该模型更善于捕捉序列中的短期和局部模式

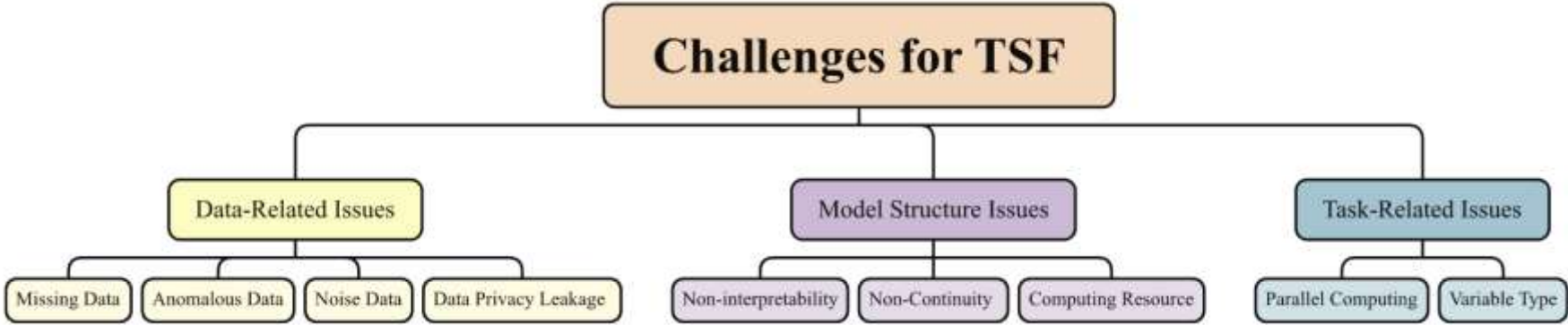
□ 经典模型：PatchTST



PatchTST (ICLR 2023) 的概述

- 将时间序列分割成子序列级别的补丁，作为Transformer的输入令牌。它们独立地为每个通道建模，以表示单个变量。
- <https://zhuanlan.zhihu.com/p/642866281>

挑战/前景



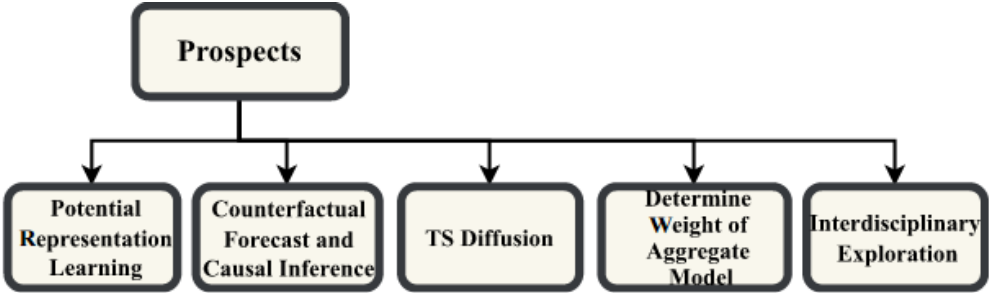
大型模型的挑战

数据隐私保护：使用联邦学习（FL），多个参与者可以协同训练模型，而无需共享敏感的原始数据。

大型模型的挑战：

- 2023 年 12 月 13 日，亚马逊发布了利用大型模型进行时间序列预测的工作。
- TimeGPT-1 模型将大型语言模型 (LLM) 背后的技术和架构应用于预测领域，成功地建立了第一个能够零样本推理的基础时间序列模型。

<https://zhuanlan.zhihu.com/p/683685850>



表征学习（RL）专注于获取低维和紧凑的数据表示，捕获数据的基本特征。

一些工作采用多模块架构或模型，另一些工作则使用具有去噪、平滑属性、孪生结构或 2D 变化建模的预训练。此外，对比学习帮助模型从未标记的数据中学习有意义的特征。

谢 谢！