

U-pre: U-Net是优秀的时序预测器

2024 年 11 月 23 日

摘要

时间序列预测是一项关键的回归任务。本文提出了一种基于 U-Net 架构的有效时间序列预测模型，命名为 U-pre。我们观察到 U-Net 架构与其固有的输入输出对称性和数学一致性使其非常适合时间序列预测问题，其中输入和输出序列通常具有相同的维度和统计特性。我们首先构建了 U-preV1，一个仅基于一维 U-Net (U-Net1D) 的模型，并在 ETT 数据集上进行了实验，结果表明其性能优于部分 2022 和 2023 年的基线模型。为了进一步提升模型的性能，我们提出了 U-preV2，该模型将 U-Net1D 与 BERT 编码器相结合，以捕捉时间序列中的局部和全局依赖关系。U-Net1D 负责提取局部时间特征，而 BERT 编码器则整合了 U-Net1D 的输出和原始时间序列，以捕捉全局上下文信息。实验结果验证了 U-preV2 的有效性。我们的代码已公开发布在：<https://github.com/Selen-Suyue/U-pre>

1 引言及相关工作

1.1 U-preV1: 基于 U-Net1D 的基线模型

本研究提出了 U-pre，一个用于时间序列预测的模型。我们注意到时间序列预测任务的输入和输出通常具有相同的维度和统计分布，这与 U-Net 架构的输入输出对称性高度契合。基于此观察，我们设计了 U-preV1，该模型直接利用 U-Net1D 对 ETT 数据集进行时间序列预测。虽然 U-preV1 的性能并未超越现有的最先进方法 (SOTA)，但其结果已优于部分 2022 年的基线模型，这表明 U-Net 架构在时间序列预测任务中具有显著的潜力。

1.2 U-preV2: 融合局部与全局注意力机制的改进模型

U-preV1 的主要局限在于一维卷积的局部感受野，限制了其对长期依赖关系的建模能力。为了克服这一缺陷，U-preV2 提出了一种混合方法，将 U-Net1D 与 BERT 编码器相结合，以同时捕获局部和全局信息。U-Net1D 作为局部注意力机制，提取局部时间特征；其输出与原始时间序列拼接后作为输入，送入单层 BERT 编码器，实现全局注意力机制，捕捉长程依赖关系。这种结合局部和全局注意力的策略，有效地提升了模型对时间序列复杂模式的建模能力。我们将 U-Net1D 的特征提取阶段定义为局部注意力，因为它隐式地为不同的时间步赋予了局部梯度权重；而 BERT 编码器则通过注意力机制实现全局注意力，捕捉全局上下文信息。

1.3 主要贡献

- 提出了 U-pre，一种基于 U-Net 架构的有效时间序列预测模型，并提出了其两个版本，U-preV1 和 U-preV2。
- U-preV2 巧妙地结合了 U-Net1D 和 BERT 编码器，有效地融合了局部和全局注意力，增强了模型对时间序列的建模能力。

- 在 ETT 数据集上，U-pre 取得了优于部分基线模型的性能。

1.4 相关工作: iTransformer

iTransformer[4] 是 2023 年提出的一个 SOTA 时间序列预测模型。其核心思想是重新组织输入数据，将 (B,L,N) 形式的数据转换为 (B,N,L) 形式，即以特征维度为序列维度，显著减少了输入 token 的数量，从而更有效地建模时间序列数据。这为我们使用 U-Net1D 提供了重要启示。在 U-pre 中，我们沿用 iTransformer 的数据组织方式，使其能够有效地捕捉时间序列的局部特征。在 U-preV2 中，我们也采用了 (B,N,L) 的模式输入 BERT 编码器。

2 方法

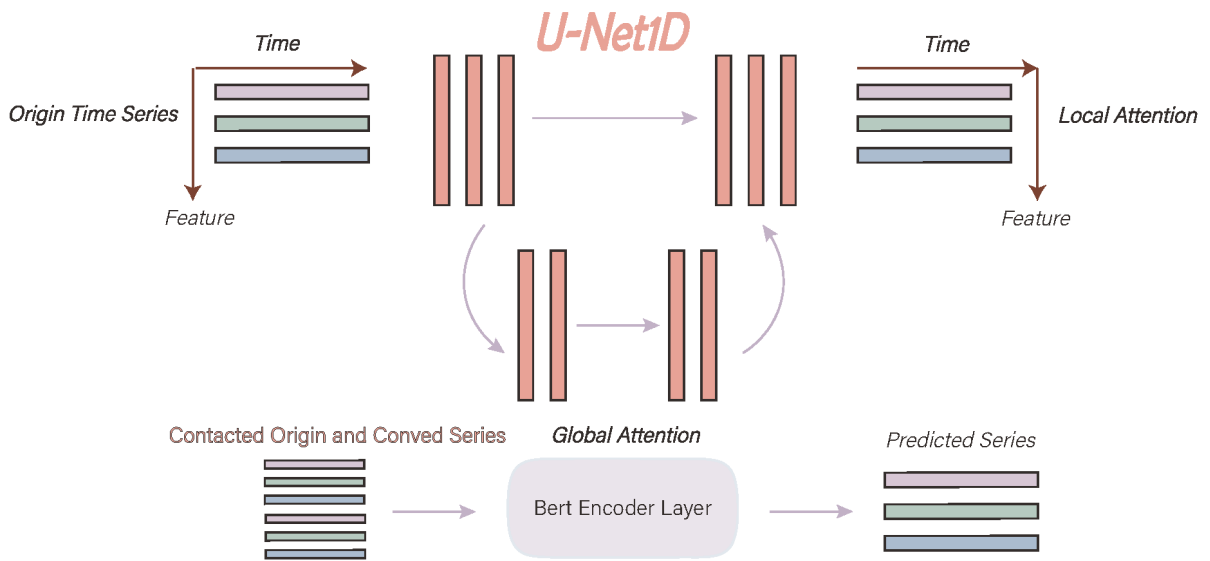


图 1: U-preV2 模型架构图

本节详细介绍 U-preV2 的模型架构，并简述 U-preV1 的流程。

2.1 U-Net1D

输入时间序列数据被组织为 (B, N, L) 的张量，其中 B 表示 batch size， N 表示特征数量， L 表示时间步长。U-Net1D 由多个卷积块组成，通过卷积操作进行下采样和上采样，最终输出与输入维度相同的张量。在编码阶段，卷积核大小设置为 8；在解码阶段，卷积核大小设置为 2。该设计保证了 U-Net1D 主要捕捉局部时间依赖关系。对于 U-preV1，U-Net1D 的输出直接通过一个线性层进行预测。

2.2 全局注意力机制 (Global Attention)

U-preV2 将 U-Net1D 的输出与原始输入在特征维度上拼接，形成 $(B, 2N, L)$ 的张量，然后输入到单层 BERT 编码器中。BERT 编码器对 $2N$ 个 token 进行处理，捕捉不同特征之间以及 U-Net1D 输出与原始输入之间的关系。U-Net1D 的输出隐式地编码了局部时间梯度信息，BERT 编码器通过注意力机制，关注这些梯度信息，从而捕捉时间序列的变化量，而非仅仅是原始数值。

3 实验

模型	ETTm1 (MSE/MAE)	ETTm2 (MSE/MAE)	ETTth1 (MSE/MAE)	ETTth2 (MSE/MAE)	平均 (MSE/MAE)
iTransformer [4]	0.334/0.368	0.180/0.264	0.386/0.405	0.297/0.349	0.299/0.347
RLinear [2]	0.355/0.376	0.182/0.265	0.386/0.395	0.288/0.338	0.303/0.344
PatchTST [6]	0.329/0.367	0.175/0.259	0.414/0.419	0.302/0.348	0.305/0.348
Crossformer [10]	0.404/0.426	0.287/0.366	0.423/0.448	0.340/0.374	0.364/0.404
TiDE [1]	0.364/0.387	0.207/0.305	0.384/0.402	0.340/0.374	0.324/0.367
TimesNet [7]	0.338/0.375	0.187/0.267	0.479/0.464	0.400/0.440	0.351/0.387
DLinear [9]	0.345/0.372	0.193/0.292	0.386/0.400	0.402/0.414	0.332/0.369
SCINet [3]	0.418/0.438	0.286/0.377	0.654/0.599	0.376/0.419	0.434/0.458
FEDformer [11]	0.379/0.419	0.203/0.287	0.513/0.491	0.449/0.459	0.386/0.414
Stationary [5]	0.386/0.398	0.192/0.274	0.449/0.459	0.526/0.516	0.388/0.412
Autoformer [8]	0.505/0.475	0.255/0.339	0.449/0.459	0.450/0.459	0.415/0.433
U-preV1	0.466/0.451	0.195/0.275	0.524/0.483	0.367/0.395	0.388/0.401
U-preV2	0.370/0.396	0.188/0.273	0.419/0.429	0.325/0.371	0.326/0.367

表 1: 不同模型在 ETT 数据集上的预测性能 (MSE/MAE) 比较 (96 步预测), 标红色的区域意味着两项指标都低于 U-preV2

我们的结果如表1所示, 选用的 Baseline: Autoformer (2021); SCINet, FEDformer, Stationary (2022), Else (2023, 2024)。通过比较实验我们发现, U-preV2 仅仅通过单层 EncoderLayer 就超过了许多基于 Transformer 的模型, 这说明 U-NET 的编码是有效的, 同时, V2 相比 V1 的提升也非常明显, 这一定程度上说明了一维卷积的局限性和长距注意力的必要性。图2展示了我们的部

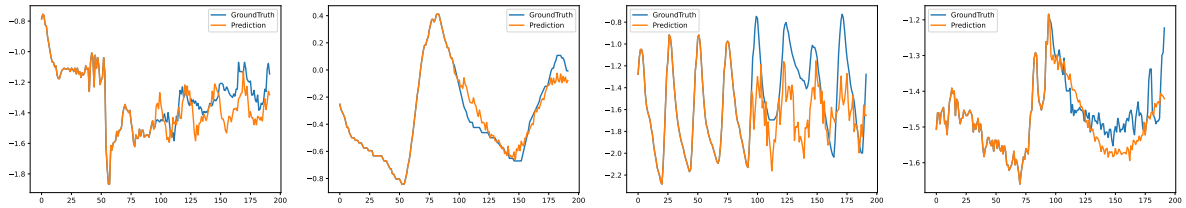


图 2: U-preV2 预测结果示例

分预测结果, 和大多数现有的时序预测模型一样, U-pre 无法克服在预测后期与 Ground Truth 不拟合的情况, 但是值得欣慰的是它能够基本判断未来时序在具体时刻的波动情况, 这是因为前期的 U-NET 处理捕捉了局部梯度, 也是 U-pre 相比现有模型的一个很大的优势。

4 结论

本文提出了 U-pre, 一种基于 U-Net 架构的有效时间序列预测模型。通过实验验证, U-pre 在 ETT 数据集上取得了不错的性能。尤其 U-preV2, 通过结合 U-Net1D 和 BERT 编码器, 有效地捕捉了局部和全局时间依赖关系, 进一步提升了模型的预测精度。未来工作将探索更复杂的 U-Net 结构以及更先进的注意力机制, 以进一步提升 U-pre 的性能。

参考文献

[1] Abhimanyu Das, Weihao Kong, Andrew Leach, Shaan K Mathur, Rajat Sen, and Rose Yu. Long-term forecasting with tiDE: Time-series dense encoder. *Transactions on Machine Learn-*

ing Research, 2023.

- [2] Zhe Li, Shiyi Qi, Yiduo Li, and Zenglin Xu. Revisiting long-term time series forecasting: An investigation on linear mapping, 2023.
- [3] Minhao LIU, Ailing Zeng, Muxi Chen, Zhijian Xu, Qiuxia LAI, Lingna Ma, and Qiang Xu. SCINet: Time series modeling and forecasting with sample convolution and interaction. In *Advances in Neural Information Processing Systems*, 2022.
- [4] Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting, 2024.
- [5] Yong Liu, Haixu Wu, Jianmin Wang, and Mingsheng Long. Non-stationary transformers: Exploring the stationarity in time series forecasting. In *Advances in Neural Information Processing Systems*, 2022.
- [6] Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *The Eleventh International Conference on Learning Representations*, 2023.
- [7] Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. In *The Eleventh International Conference on Learning Representations*, 2023.
- [8] Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with Auto-Correlation for long-term series forecasting. In *Advances in Neural Information Processing Systems*, 2021.
- [9] Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? 2023.
- [10] Yunhao Zhang and Junchi Yan. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In *The Eleventh International Conference on Learning Representations*, 2023.
- [11] Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *Proc. 39th International Conference on Machine Learning*, 2022.