

Transformer 在时间序列预测中的应用综述

夏雪¹ 闫恩来² 李喜武¹
XIA Xue YAN Enlai LI Xiwu

摘要

以 Transformer 为代表的深度学习理论在时间序列预测任务中表现优异。首先阐述了 Transformer 的基本原理和特点；然后着重分析了目前基于 Transformer 结构改进的相关预测模型，分别从性能优化和结构改进两方面进行了详细的论述，并列举了改进模型应用的场合和领域；最后总结了 Transformer 在时间序列预测上的优缺点，并对未来的研究工作进行了展望。

关键词

时间序列；预测；Transformer

doi: 10.3969/j.issn.1672-9528.2024.03.000

0 引言

时间序列预测技术是指根据时间序列所映射出来的发展趋势，推导出接下来一段时间的发展进程^[1]。该技术对人们日常生活的方方面面都产生了巨大影响，例如天气预测、种植物产量预测、交通流量预测，以及在城市未来的治理上发挥了极大的作用^[2]。

基于深度学习的时间序列预测模型发展脉络如图 1 所示，本文对 Transformer 在时间序列预测任务中的发展进行梳理、分析、总结。

传统方法，能更好地解决长距离依赖问题。Transformer 拥有编码器栈（Encoder stack）和解码器栈（Decoder stack），Encoder stack 和 Decoder stack 中分别为连续 N 个具有相同结构的 Encoder 和 Decoder。每一个 Encoder 和 Decoder 都是串联的组合，Encoder 和 Decoder 都包含 Self-attention 和前馈网络两个核心模块。Attention 机制本质是加权求和以获得对上下文的全局感知。Self-attention 计算每个部件的权重，标明互相间的关系，Self-attention 可以分解成 Multi-head attention，而前馈网络是根据这些权重进行一次变形。

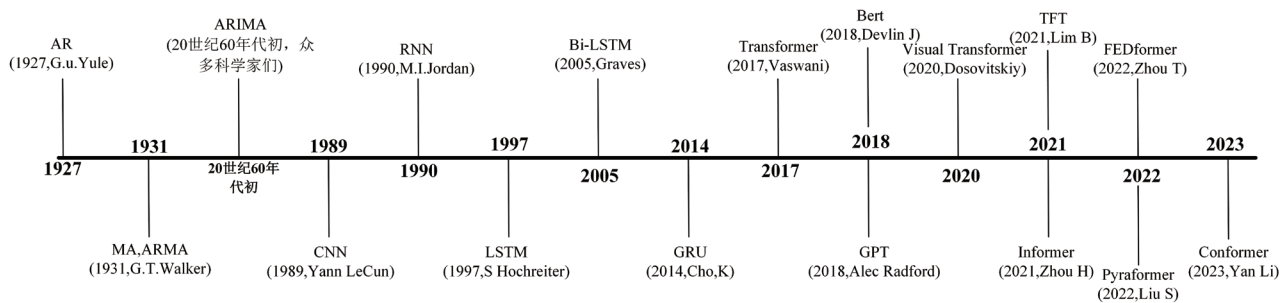


图 1 基于深度学习的时间序列预测模型发展图

1 Transformer 基础理论

Transformer 是一种深度学习架构，采用 Self-attention 机制而非传统的循环神经网络，适用于多种序列数据处理任务，并在各个领域取得了显著的成功。Transformer 的核心是 Self-attention 机制，在全局上下文信息的兼顾上表现极佳^[3]。Self-attention 机制在捕捉数据或特征的内部相关性方面优于

Transformer 的时间序列预测原理如图 2 所示，由嵌入部分、Encoder-Decoder 部分与逻辑回归部分组成。Transformer 输入模块首先通过一个嵌入层，对输入序列的每个数据进行词嵌入，将原本一维数据升维成二维矩阵，将新张量与位置编码相加，并通过 Multi-head attention 模块传递。Encoder-Decoder 的核心是 Self-attention 机制，Transformer 神经网络在 Self-attention 机制的基础上升级为 Multi-head attention 机制。Multi-head attention 模块是由单头注意力在输入序列中的通道维度上划分成很多个头形成的，每个头使用不同的可学习权重，对应生成不同的 Q 、 K 、 V 值。然后，Multi-head

1. 广东石油化工学院自动化学院 广东茂名 525000
2. 吉林化工学院信息与控制工程学院 吉林吉林 132022
[基金项目] 广东石油化工学院科研基金项目 (2018rc30)；
针对小分子污染物的在线分析方法及其应用研究 (U21A20290)

attention 模块的输出被传递到一个两层前馈网络, 该网络的输入和输出以残差方式与层归一化连接^[4]。逻辑回归部分由一个线性变换与 Softmax 映射组成, 其作用是将 Decoder 的输出回归到输出向量空间中并重新映射为下一时刻观察点的预测概率。

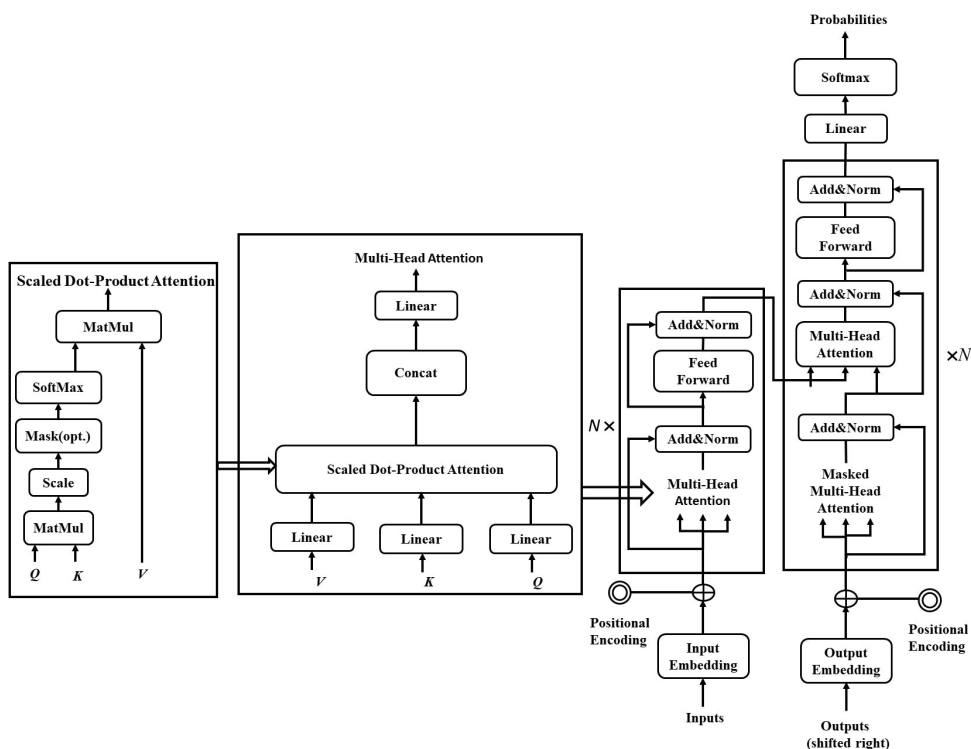


图2 基于 Transformer 的时间序列预测原理示意图^[5]

2 Transformer 在时间序列预测领域研究的进展

Transformer 的 Self-attention 机制和各种改进版本使其成为处理时间序列数据的有力工具。研究人员还在不断探索如何进一步优化和定制 Transformer 以满足不同时间序列预测任务的需求。本节主要介绍目前在时间序列预测领域已经推出的改进模型, 为后续的研究提供了方向和参考。

2.1 时域融合 Transformer (TFT)

TFT 是一种新的基于注意力的架构, 结合了高性能的多水平预测和可解释的时间动态洞察力。TFT 主要由 5 个部分组成, 分别是门控机制、变量选择网络、静态协变量编码器、时间处理以及多水平预测区间预测^[6]。

TFT 的运行过程是: 先输入静态元数据、时变的过去输入和时变的先验已知未来输入。静态元数据通过变量选择进入到静态协变量编码器中, 再输向门控残差网络 (GRN), 而过去输入和先验已知未来输入数据通过变量选择进入 LSTM 编码器, 经过门控机制 (Gate), 求和与归一化 (Add&Norm) 后输向门控残差网络。GRN 与遮挡 - 可解释多头

注意力模块, 以及门控机制、残差连接模块组成一个时间融合解码器, 它由序列到序列层、静态富集层、时间自我注意层、位置前馈层四个部分构成。TFT 也在点预测的基础上生成预测区间, 这是通过在每个时间步长同时预测各种百分位数来实现的。

TFT 已在多个时间序列预测任务中取得良好的效果, AMRIL 等人^[7]利用 TFT 对智能电网用户能需求进行预测, 根据客户需求生产能源, 以此节约资源。TFT 是一个非常具有前途的模型, 可以为时间序列预测领域带来更好的性能和效率。

2.2 Crossformer

Crossformer^[8]用以解决多元时间序列问题, 利用跨维依赖性进行多元时间序列预测 (MTS)。在 Crossformer 中, 输入 MTS 通过维度分段 (DSW) 嵌入到二维矢量数组中, 以保留时间和尺寸信息, 提出两阶段注意层 (TSA) 来捕获跨时间和跨维度的依赖关系。

利用 DSW 嵌入和 TSA 层, 建立了一个层次 Encoder-Decoder (HED), 通过不同尺度信息进行最终预测。DSW 的模型表达式为:

$$X_{1:T} = \left\{ X_{i,d}^{(s)} \mid 1 \leq i \leq \frac{T}{L_{seg}}, 1 \leq d \leq D \right\} \quad (1)$$

$$X_{i,d}^{(s)} = \{x_{i,d} \mid (i-1) \times L_{seg} < t \leq i \times L_{seg}\} \quad (2)$$

式中: $X_{1:T}$ 表示输入包含 T 个序列, 每个序列有 D 个维度; $X_{i,d}^{(s)}$ 维度 d 中的第 i 个段, 长度为 L_{seg} 。

Crossformer 是分层 Encoder-Decoder 的架构, 每个矢量的长度表示覆盖的时间范围。其主要包含 DSW Embedding、TSA 层和 Linear Projection。Encoder 使用 TSA 层和分段合并来捕捉不同尺度的依赖关系。同时, Decoder 在每个尺度上进行预测并将其结果相加进行最终预测。彭子彬^[9]构建了结合 Crossformer 模型的 Crossformer-CRF 模型, 对输出的字符特征进行注意力操作并通过条件随机场进行标签约束, 以获得最佳的命名实体标签序列。该模型能有效融合词汇信息以及相对位置信息, 提升了模型预测效果。

2.3 Informer

Informer^[10] 是基于 Transformer 改进的轻量化模型, 适合处理长时间序列任务。Informer 由多头概率稀疏自注意力层、Multi-head attention 机制、掩码多头概率稀疏自注意力层构成, 通过 Encoder 获取时间依赖, Decoder 生成预测结果。不同于 Transformer 的是, Informer 只有一个 Encoder 和 Decoder。

Informer 目前仍在逐步发展, 例如 Informer-DCR 在多云计算资源负载预测中发挥出最佳水平, 将 Encoder 中各注意力块之间的正则卷积替换为扩张因果卷积, 使深层网络中高层能够接收更大范围的输入信息来提高模型预测精度, 并保证了时序预测过程的因果性。苏燕等人^[11] 提出基于 K-means 聚类融合多元时空信息的 Informer-AD 大坝变形预测模型, 该模型对大坝变形预测有较高的适应性; 朱莉等人^[12] 使用二次分解策略组合 Informer 对短期电力负荷的预测, 对比模型得出该预测方法能有效降低数据的非平稳性并提高预测精度。

2.4 Autoformer

Autoformer^[13] 是一种基于 Transformer 的自动化神经网络架构设计模型, 通过自适应搜索空间和神经网络结构自动优化的方式来提高模型的性能和效率。Autoformer 的重点是序列分解, 分为季节性分解和趋势周期性分解, 分别反映了该序列的季节性和长期进展。Encoder 通过序列分解去除趋势项, Decoder 包含内部自相关和 Encoder-Decoder 自相关, 分别可以细化预测和利用过去的季节信息。

Autoformer 的模型表达式为:

$$\chi_t = \text{AvgPool}(\text{Padding}(\chi)) \quad (3)$$

$$\chi_s = \chi - \chi_t \quad (4)$$

式中: χ_s, χ_t 分别表示季节性部分和趋势周期性部分。

Autoformer 作为一种分解架构, 通过嵌入序列分解块, 从中间预测中逐步聚合长期趋势部分, 并在序列中发现相互之间依赖关系和对信息进行聚合。范杏蕊^[14] 提出基于 Autoformer 改进的 Sdformer 模型, 设计了深度模型的内部分解模块, 提取模型中隐藏状态的内在复杂时序趋势, 对短期电力负荷进行有效预测。

2.5 Pyraformer

Pyraformer^[15] 因其强大的数据推理能力和处理长序列的能力在各个行业中得到广泛应用, 重点在于引入了全局标记来降低时间复杂度。开发金字塔注意力机制以弥补远程依赖关系和实现低时间和空间复杂性之间的差距, 金字塔注意力机制可以分为尺度间连接和尺度内连接。尺度间连接构建了原始序列的多分辨率表示, 尺度内的边缘通过将相邻节点连接在一起来捕获每个分辨率的时间依赖性。

Pyraformer 在小麦重金属污染安全风险水平的预测上发挥重要作用, 为我国小麦重金属污染分级管理和预警提供了

技术手段, 也为农田小麦重金属污染动态监测和综合防治提供了科学依据。

2.6 FEDformer

FEDformer 在 Transformer 模型下开发了频率增强的功能, 进一步提高了 Transformer 的长期预测性能。FEDformer 的傅里叶分析是深入频域的关键部分, 使用少量选定的傅立叶分量保持时间序列的紧凑表示, 从而使 Transformer 达到高效计算, 这对于建模长序列尤为重要。基于傅立叶变换和小波变换的模块, 通过在频域进行固定数量的随机采样, 使得模型达到线性复杂度的同时提高精度。当傅立叶变换在频域中创建信号表示时, 小波变换也在频域和时域中创建表示, 使其能够有效地访问信号的局部信息。

孙亦皓等人^[16] 利用 GCN-FEDformer 对风速进行预测, 更全面地捕捉相邻风电场之间风速的时空相关性。FEDformer 被用于检测储藏环境对稻谷品质的影响, Zihan Li 等人^[17] 利用 FEDformer 模型与 k-medoids 算法结合, 构建了稻谷品质变化预测模型和分级评价模型, 在预测稻谷储藏过程中品质变化时体现出最高的准确性和最低的误差。

3 Transformer 在其他领域的发展状况

Transformer 本身具有对序列数据的处理能力, 适用于时域、频域和空间域之间的多模态特征提取和融合, 大幅提升对于 Speech 和 Video 相关任务的效果, 同样运用 BEV based 3DDetection 等前沿场景。Transformer 最初是为机器翻译而设计的, 其因灵活的架构已被广泛应用于 NLP 之外的各个领域, 包括 CV、音频处理和图像识别等。

(1) 自然语言处理。Transformer 及其变体已在 NLP 任务中得到广泛探索和应用, 例如机器翻译、语言建模和命名实体识别。DAI 等人^[18] 提出的 Transformer-XL 不仅能够捕捉长期依赖关系, 还解决了上下文碎片问题。

(2) 计算机视觉。Transformer 还适用于各种视觉任务, 例如图像分类、物体检测、图像生成和视频处理。朱哲峰^[19] 基于 Vision Transformer 提出了一种多尺度特征融合的图像识别框架 FFT (feature fusion transformer), 该框架能关注到图像中更加细致微小的特征, 为图像识别提供了更加丰富的特征信息, 有效提升图像识别的准确率。张艺超等人^[20] 提出一个基于 Self-attention 机制的层级融合高光谱图像分类框架, 增强各层级特征联动。Transformer 及其变体如雨后春笋般涌现, 给计算机视觉领域注入了新的活力, 引领了新的变革。

(3) 音频应用。Transformer 还可以扩展到音频应用上, 例如语音识别、语音合成、语音增强和音乐生成。YU 等人^[21] 提出一种基于认知计算的语音增强模型, 称为 SETransformer, 该模型可以在未知噪声环境中提高语音质量。

(4) 多模态应用。Transformer 还被应用于各种多模态

场景,例如视觉问答、视觉常识推理、字幕生成、语音到文本翻译和文本到图像生成。DING 等人^[22]提出 CogView,是一个具有强大生成能力和跨模态理解能力的模型,解决了文本到图像的生成问题。由于文本和音频输入表示不同,模态差距使得文本机器翻译数据及端到端模型与端到端语音翻译部分不兼容,对此,HAN 等人^[23]提出 Chimera 模型用以弥合这一差距。

4 优缺点以及未来工作

根据对 Transformer 应用领域的研究和分析,发现 Transformer 具有长序列处理能力强、并行计算能力强、不受位置信息限制且具有全局感受野的优点。但同时也存在模型解释性较差、训练和调参复杂、对于输入数据的噪声不敏感、预测速度慢,以及还需要更多内存和计算资源的缺陷。

未来 Transformer 在时间序列预测领域可以从以下几个方向开展工作。

- (1) 实施注意力的局部化,减少对整个序列的依赖。
- (2) 使用模型解释工具分析模型的预测结果。
- (3) 运用预训练模型进行微调,以此节省训练时间,或者尝试自动调参技术,更有效地搜索超参数空间。
- (4) 引入严格的输入数据预处理,以清理数据噪声。
- (5) 探索模型并行化和分布式计算,更好地利用计算资源。

5 结论

Transformer 作为深度学习领域的杰出代表,其全局建模能力和可扩展性使其在捕捉时间序列中复杂关系方面具有独特的优势,因此在时间序列预测任务中表现出色。本文总结了 Transformer 及其变体模型在时间序列预测领域中的相关工作,并分析了现有工作存在的不足,为今后开展利用 Transformer 进行时间序列预测方面的研究做好基础准备。未来将会继续推动 Transformer 在不同领域中的应用,并与其他模型进行有益的融合,这有助于拓展其在时间序列预测领域的潜力。通过这些努力,有望进一步深化对时间序列预测问题的理解,促进该领域的发展和创新。

参考文献:

- [1] 何亚磊,许乾坤.时间序列预测技术综述[J].信息通信,2018(11):35-36.
- [2] GAO J, SUN Y. The evolution of ecological and environmental governance attention allocation in J city based on big data analysis[J]. Discrete dynamics in nature and society, 2022,2022:1-8.
- [3] 李翔,张涛,张哲,等. Transformer 在计算机视觉领域的研究综述[J]. 计算机工程与应用,2023,59(1):1-14.

- [4] YI T, MOSTAFA D, DARA B, et al. Efficient transformers: a survey[J]. ACM cOMPUTING surveys, 2023,55(6):1-28.
- [5] ASHISH V, NOAM S, NIKI P, et al. Attention is all you need[J]. Conference on neural information processing systems, 2017, 30: 5998-6008.
- [6] LIM B, ARIK S O, LOEFF N, et al. Temporal fusion transformers for interpretable multi-horizon time series forecasting[J]. International journal of forecasting, 2021,37(4): 1748-1764.
- [7] AMRIL N, KHALIQUE A S, SALAM A S, et al. Forecasting energy consumption demand of customers in smart grid using temporal fusion transformer (TFT)[J]. Results in engineering,2023, 17: 1-14.
- [8] CHEN L, GE H, LI J. CrossFormer: multi-scale cross-attention for polyp segmentation[J]. IET image processing,2023,17(12): 3441-3452.
- [9] 彭子彬.面向司法领域的中文命名实体识别方法研究[D].南昌:江西财经大学,2023.
- [10] MINGJU G, CHANGCHENG Y, WEI X, et al. Short-term wind power forecasting model based on temporal convolutional network and informer[J]. Energy,2023,283:1-9.
- [11] 苏燕,黄姝璇,林川,等.融合多元时空信息的 Informer-AD 大坝变形预测模型[J].水力发电学报,2023(11):101-113.
- [12] 朱莉,韩凯萍,朱春强.二次分解策略组合 Informer 的短期电力负荷预测方法[J].国外电子测量技术,2023,42(6):23-32.
- [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] 范杏蕊,李元诚.基于改进 Autoformer 模型的短期电力负荷预测[J/OL].电力自动化设备,1-12[2024-03-16].<https://doi.org/10.16081/j.epae.202305011>.
- [15] DONG W, HU T, ZHANG Q, et al. Prediction of food safety risk level of wheat in China based on pyraformer neural network model for heavy metal contamination[J]. Foods, 2023, 12(9): 2-19.
- [16] 孙亦皓,刘浩,胡天宇,等.基于时空关联特征与 GCN-FEDformer 的风速短期预测方法[J/OL].中国电机工程学报,1-10[2024-03-16].<https://doi.org/10.13334/j.0258-8013.pcsee.231140>.
- [17] LI Z H, ZHANG Q C, DONG W, et al. FEDformer-Based paddy quality assessment model affected by toxin change in different storage environments[J]. Foods, 2023, 12(8): 1681.

板级高速互连线的物理特征优化方法

梁卓灏¹ 杨其宇¹

LIANG Zhuohao YANG Qiyu

摘要

针对板级高速互连线的物理特征优化问题,采用非支配排序遗传算法III(NSGA-III)、推拉搜索算法(PPS)和TIGE2算法对其进行优化,主要针对板级高速互连线信号完整性的反射、串扰和损耗的多目标优化问题,并结合实际工程设计时的生产工艺、性能要求和材料成本设定了多个约束条件。对比优化前起始结构的目标函数和三种遗传算法优化结构的目标函数,结果显示,三种遗传算法都可以用于处理板级高速互连线的物理特征优化问题,其中NSGA-III优化结构的目标函数总和最小,算法的运算时间较短,在实际的工程设计中更具备指导意义。

关键词

遗传算法; 信号完整性; 板级高速互连线; 非线性约束; 多目标优化

doi: 10.3969/j.issn.1672-9528.2024.03.031

0 引言

随着电子元器件不断地向着高性能、小体积和低功耗等方向发展,板级互连线的传输速度和布线密度在过去的几十年里不断提高,目前板级高速互连线单通道的传输速度设计要求高达56 GB/s^[1],同时,串扰、反射、损耗等信号完整性问题也随之出现^[2],国际半导体技术蓝图ITRS指出电子系统的未来发展将会受限于板级高速互连线的信号完整性性能^[3]。

随着机器学习在近年的快速发展,为了减少对硬件工程师设计经验的依赖,国内外的研究人员使用机器学习进行信号完整性的研究工作。Heegon Kim等人提出一种基于神经网络的板级高速互连的建模办法,采用W-Element RLGC矩阵与频率无关的特点,简化了互连线模型^[4]。黄中钊在Heegon Kim的基础上,考虑了生产工艺的影响,优化了人工神经网络对板级高速互连线的建模,代理模型预测得到的参数精度达到96.2%^[5]。

本文根据黄中钊提出的基于W-Element RLGC矩阵的板

1. 广东工业大学 广东广州 511400

[18]DAI Z H, YANG Z L, YANG Y M, et al. Transformer-XL: attentive language models beyond a fixed-length context[C]// Annual meeting of the association for computational linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2019: 2978-2988.

[19]朱哲峰. 基于Vision Transformer的图像分类算法研究[D]. 广州: 广州大学, 2023.

[20]张艺超, 郑向涛, 卢孝强. 基于层级Transformer的高光谱图像分类方法[J]. 测绘学报, 2023, 52(7): 1139-1147.

[21]YU W, ZHOU J, WANG H B, et al. SETransformer: speech enhancement transformer[J]. Cognitive computation, 2022, 14: 1-7.

[22]DING M, YANG Z Y, HONG W Y, et al. Cogview: Mastering text-to-image generation via transformers[J]. Advances in Neural Information Processing Systems, 2021, 34: 19822-19835.

[23]HAN C, WANG M X, JI H, et al. Learning shared semantic space for speech-to-text translation[C]// Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2021: 2214-2225.

【作者简介】

夏雪(1998—), 女, 安徽省合肥人, 硕士研究生, 研究方向: 润滑油监测与状态评估。

闫恩来(1998—), 男, 内蒙古包头人, 硕士研究生, 研究方向: 润滑油监测与状态评估。

李喜武(1971—), 男, 通信作者, 吉林长春人, 博士, 副教授, 研究方向: 设备智能监测与控制。

(收稿日期: 2023-12-12)