

基于门控多层感知机和Informer 的多通道电力负荷预测

孙卓远¹, 吕学文^{1,2}, 王继军³

¹上海电力大学计算机科学与技术学院, 上海

²全程上海智能科技有限公司, 上海

³赢科储能科技有限公司, 湖南 长沙

收稿日期: 2024年4月15日; 录用日期: 2024年5月24日; 发布日期: 2024年5月31日

摘要

在电力领域, 利用时间序列方法进行电力负荷预测已成为众多研究的热点。为了解决电力负荷预测准确率低的问题, 本文提出了融合门控多层感知器和增强因果卷积的多通道时间序列融合网络 GMEC-Informer, 提高了模型捕捉长短期时间序列信息依赖的能力。为了证明本文模型的优越性, 本文在广泛使用的数据集上与多个模型进行了比较, 实验结果表明本文提出的GMEC-Informer具有更高的预测精度, 可以为时间序列预测提供更好的研究方向。

关键词

负荷预测, 因果卷积, 深度学习, 注意力机制

Multi-Channel Power Load Forecasting Based on Gated Multilayer Perceptron and Informer

Zhuoyuan Sun¹, Xuewen Lv^{1,2}, Jijun Wang³

¹School of Computer Science and Technology, Shanghai University of Electric Power, Shanghai

²Intellqc Co., Ltd., Shanghai

³WINCLE Co., Ltd., Changsha Hunan

Received: Apr. 15th, 2024; accepted: May 24th, 2024; published: May 31st, 2024

文章引用: 孙卓远, 吕学文, 王继军. 基于门控多层感知机和 Informer 的多通道电力负荷预测[J]. 人工智能与机器人研究, 2024, 13(2): 375-387. DOI: 10.12677/airr.2024.132039

Abstract

In the field of electric power, power load forecasting using time series methods has become a hot spot in many researches. In order to solve the problem of low accuracy of power load prediction, this paper adds gated multilayer perceptual units into the Informer model, and proposes GMEC-Informer, a multichannel time-series fusion network that fuses gated multilayer perceptron and enhanced causal convolution, which improves the model's ability to capture long-short time series information dependence. In order to prove the superiority of the model in this paper, it is compared with several models on widely used datasets, and the experimental results show that GMEC-Informer has higher prediction accuracy and can provide a better research direction for time series prediction.

Keywords

Power Forecasting, Causal Convolution, Deep Learning, Attention Mechanism

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

时间序列数据在金融、气象、农业、工业和医疗等领域有着广泛应用[1] [2] [3]。近年来,随着传感器和网络技术的发展,各个领域生成和积累了大量时间序列数据。时间序列预测分为连续型预测和离散型预测,其核心是从历史数据中挖掘规律,估计未来趋势。随着时序数据量和维度的增长,时间序列预测方法不断演进,从数学统计到机器学习,再到深度学习。深度学习在自然语言处理中的有效性促进了其在时间序列研究中的应用[4]。

时间序列是一系列或多组随机变量,按照时间先后顺序生成,通常以固定的采样频率记录数据的变化过程。先前的时间点的数据可能包含对当前或未来时间点数据的某种规律或相关性[5]。电力分配问题涉及电网根据不同时段的需求,动态管理电力供应到各用户区域。然而,如何准确预测特定用户区域的未来电力需求受到多种因素的影响,如工作日与节假日、季节、天气、温度等的变化,准确的预测对未来的生产具有重要影响。传统的预测方法无法应用于当今社会长时间序列数据的高精度预测[6]。在电力负荷预测的场景下,当无法准确预测未来电力负荷水平或对未来预测水平预测偏差过大时,往往就会采用保守的发电量,导致电力过量储存和设备折旧的不必要浪费[7]。

近年来,国内外学者对电力领域发电量预测问题都进行了深入研究,在电力领域,利用时间序列方法预测电力负荷已成为许多研究的热点。Jakob 等人利用 EMD 分解原始负荷数据,然后建立 PSO-SVR 模型对每个分解成分进行预测,充分考虑了负荷数据的非线性、非稳态特点,提高了预测精度[8]。Yang 等人通过 CEEMD 和 VMD 分解了去噪原始电力负荷数据,结合时序卷积网络-双向长短时记忆神经网络建立了 CVMD-TCN-BiLSTM 模型进行分量预测,有效提高了短期负荷预测的准确性,并通过实验进行验证[9]。David 等人在文献[10]中发明了一种基于自动回归循环神经网络的概率预测模型,该模型能够为历史数据较少的项目生成准确的概率预测,并适应各种噪声分布,为解决概率预测问题提供了有效的基于深度学习的方法。Cui 等人提出了堆叠双向和单向 LSTM 网络架构,第一个特征学习层使用 BDLSTM,

后接 LSTM 层, 提高了预测性能[11]。Kim 等人提出将 LSTM、GRU 和 RNN 三种时间序列预测方法组合为集成模型, 以更准确地预测用电模式, 通过分类把工作日、假期的用电数据分开建模, 考虑周期性信息, 实验结果表明该集成模型相比单一模型能明显提高预测准确率[12]。田英杰等人设计了周期自动编码器, 将电力时间序列周期性地嵌入到向量空间中, 然后利用多种深度神经网络模型对个体水平的电力负荷进行预测[13]。Nathan 等人通过在 LSTM 的门内实施 GNN 并利用空间信息来进一步研究这两种方法的集成。并且引入了跳跃连接, 该连接对于共同捕获数据中的空间和时间模式也被证明了非常有效[14]。文章[15]提出了一项新的短期电力负荷预测方法, 该方法采用注意力机制来增强双向长短期记忆 (Bi-LSTM) 模型的性能, 并且结合了 XGBoost 模型, 以建立一种电力负荷组合预测模型, 该模型的权重是通过误差倒数法确定的。然而大部分时间序列问题中使用的递归神经网络普遍存在的问题是仅考虑了数据的时序特性, 因此只能递归遍历进行计算, 难以建模长期依赖关系, 反向传播时容易导致梯度消失的问题, 而且解释性不强[16]。

Transformer 是一种基于 Google 团队提出的自注意力机制的 Seq2Seq 结构模型, 该模型通常用于机器翻译[17]。这种基于注意力机制模型的优势之处在于捕捉序列不同位置之间的关联所需操作无需按照序列关系循环传递, 避免了由于传递过程导致的信息丢失, 可以进行并行计算。然而, 在 Transformer 中计算自注意权重的操作, 使得模型存在二次的计算复杂度[18]。

为了解决 Transformer 中并行计算带来的计算量大的问题, 文献[19]提出了局部敏感散列注意力来代替多焦点注意力, 发明了一种基于 Transformer 的新模型, 名为 Informer, 针对于长序列时间序列预测任务(LSTF)而设计, 有效降低了 Transformer 在处理长序列时的时间复杂度。

虽然基于 Transformer 改进的模型通过堆叠模型本身或与其他模型进行组合提高了预测性能, 但是 Informer 模型本身在时序预测方面仍存在以下问题: 其单一的稀疏自注意力机制无法有效提取长序列中对应元素之间的语义相关性的要求; 正则卷积只能回溯线性大小的历史信息, 造成模型感受野受限导致计算冗余, 并且不能保证未来信息对于时序预测结果的影响; 随着网络深度的增加, 在训练时可能导致网络退化、反向传播时的梯度的消失或爆炸, 造成模型训练收敛速度慢的缺点, 使得预测性能下降。基于以上问题, 为了提高 Informer 的数据特征提取能力, 同时防止网络结构造成未来信息与当前预测交互的问题, 本文提出了一种融合门控多层感知和增强因果卷积的多模态时序融合网络 GMEC-Informer。对于稀疏自注意层, 融合门控多层感知单元提取自注意机制中难以捕捉的局部序列相关信息; 对于稀疏自注意层间的一维 CNN 前馈网络, 采用不同大小的增强因果卷积进行串行处理, 以增强数据特征提取; 其次, 将残差网络(ResNet)添加到门控单元中, 以防止深度学习过程中遇到的降级。最后在广泛使用的数据集上将模型与传统方法进行了比较, 验证了其在时间序列预测中的有效性。

综上所述, 为进一步提升电力功率预测的精度, 本文的贡献可以总结如下: 1) 通过多层门控感知单元补充单个注意力机制, 并且可以扩展网络层的宽度, 避免了模型退化, 并可以检测各种尺度的特征信息; 2) 将增强因果前馈层引入特征提取映射以防止未来信息泄露, 提升了模型对时序数据处理的匹配度; 3) 使用本文模型与其它几种常用的电力负荷预测模型在广泛使用的数据集上进行了比较。实测数据的结果表明, 与过去常见的模型相比, 本文模型在预测任务上具有明显的优势, 并获得了更显著的提升。

2. 研究方法

2.1. Informer 模型

目前, 传统的 Transformer 在自注意力机制的实现上存在复杂度高、堆叠层内存瓶颈以及预测输出速度较慢等问题。为了应对这些问题, Informer 针对时序预测领域中 Transformer 存在的问题进行了改进。

该模型引入了稀疏概率自注意力(ProbSparse Self-attention)机制, 以代替传统的自注意力机制。通过精选最为关键的查询向量(query), Informer 实现了时间复杂度和内存使用量都为 $O(L\log L)$, 从而显著减少了网络的规模。此外, Informer 还引入了自注意力蒸馏(Self-attention Distilling)操作, 通过正则卷积和池化操作来降低维度和网络参数数量。在解码操作中, Informer 创造性地提出了生成式解码器, 能够一次性生成完整的预测序列, 避免了逐步预测所带来的误差累积, 从而有效缩短了预测时间。Informer 整体架构如图 1 所示。

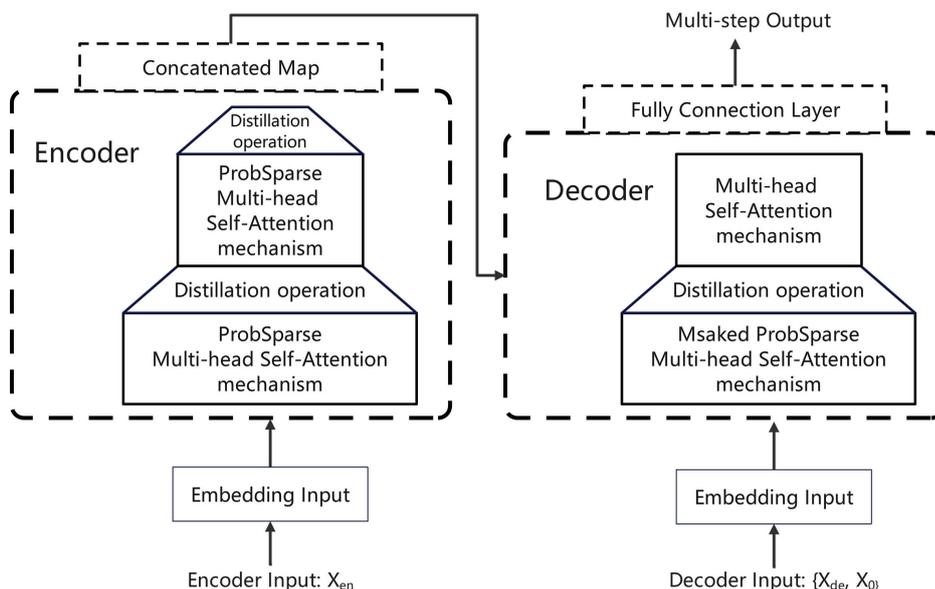


Figure 1. Overall architecture of Informer
图 1. Informer 模型整体架构

对于时间序列类型的预测问题, 数据遵循位置有序、与时间强相关的特点, 因此如何使模型学习到序列数据的位置信息对预测准确度至关重要。在 RNN 类模型中, 通过循环结构和时间戳来捕捉时间序列的规律性; 而 Transformer 则利用注意力机制和时间戳来捕捉当前位置的上下文。在处理时间序列预测问题时, 若需要捕捉数据的长期模式, 则必须考虑全局信息, 例如不同级别(周、月、年)的时间戳以及突发事件(节假日、事件)的时间戳。然而, 由于自关注点积计算中涉及多类型特征的融合, 这种信息可能导致潜在的精度下降[20], 因此, Informer 通过采用统一的输入表示来解决这一问题。在长时间序列电力负荷预测的稀疏自关注模型中, 输入部分由数据编码、位置编码和时间编码三个组成部分组成, 如图 2 所示。

首先, 将电力负荷数据映射到高维空间, 这是通过对原始数据进行一维卷积获得的。接着, 进行序列位置编码, 该编码方式对应 Transformer 模型中的位置编码。第三部分涉及数据的时间戳信息, 这一信息通过全连接层转换为高维表示。最终, 将上述三部分进行整合, 得到最终的模型输入[21]。

Informer 中编码器是由多头稀疏概率自注意力(Multi-head ProbSparse Self-attention)机制和蒸馏操作堆叠而成的。稀疏概率自注意力代替标准自注意力来接受长序列数据, 通过将输入序列映射到 Query、Key 和 Value 空间, 计算元素之间的相似度并分配注意力权重, 实现了元素之间的关联性建模。在 Informer 中, 通过稀疏概率注意力机制, 可以将每个 k 设置为只关注主查询, 如等式(1)所示。

$$\text{ProbAttention}(Q, K, V) = \text{softmax}\left(\frac{\bar{Q}K^T}{\sqrt{d}}\right) * V \tag{1}$$

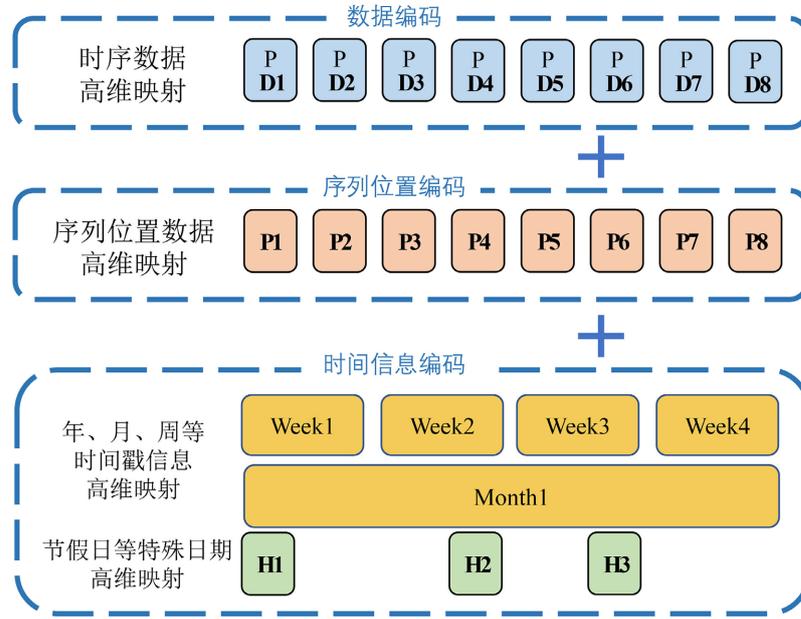


Figure 2. Input data information code
图 2. 输入数据信息编码

其中 $Q(Q \in R^{l_0 \times d})$ 表示查询矩阵, $K(K \in R^{l_k \times d})$ 表示键矩阵, V 表示值矩阵, 对 Q 稀疏化之后的矩阵, 具有与矩阵 q 相同的大小, 其仅包含稀疏度量 $M(q_i, K)$ 中自定义的极大相关查询, 其中 $M(q_i, K)$ 的公式如等式(2)所示。

$$M(q_i, K) = \ln \sum_j^{L_k} \frac{q_i k_j^T}{\sqrt{d}} - \frac{1}{L_k} \sum_j^{L_k} \frac{q_i k_j^T}{\sqrt{d}} \quad (2)$$

其中 $Q(Q \in R^{l_0 \times d})$ 表示查询矩阵, $K(K \in R^{l_k \times d})$ 表示键矩阵, V 表示值矩阵, 对 Q 稀疏化之后的矩阵, 具有与矩阵 q 相同的大小, 其仅包含稀疏度量 $M(q_i, K)$ 中自定义的极大相关查询, 其中 $M(q_i, K)$ 的公式如等式(2)所示。

Informer 模型在每个自注意块之间执行自注意蒸馏操作, 这一操作的目的是选择主要的自注意力信息, 以减小存储器和网络的占用大小, 从而降低后续计算复杂度和开销。从第 j 层到第 $(j + 1)$ 层进行蒸馏操作的过程如等式(3)所示。

$$X_i^t = \text{MaxPool} \left(\text{ELU} \left(\text{Convld} \left(\left[X_{i-1}^t \right]_{PB} \right) \right) \right) \quad (3)$$

其中 $\left[X_i^t \right]_{PB}$ 表示多头稀疏概率自注意块中的基本操作。Informer 采用了卷积核为 3 的 1D-CNN 提取负载特征, 但单一大小的卷积核难以捕捉不同尺度的时序信息, 影响了数据特征的评价。

2.2. 门控多层感知单元

门控多层感知单元 GMLP (Gated Multilayer Perceptron) 是一种用于分析时间序列的神经网络架构, 其特点在于能够有效地捕获局部空间尺度的特征[22], 通过引入门控机制和多层感知器结构, 用以处理序列数据和时序任务。如图 3 所示, GMLP 构建在 MLP (Multilayer Perceptron) 结构的基础上, 同时融合了注意力机制的一些思想, 使其在序列建模和时序预测方面表现卓越。这一模块的核心概念在于, 在 MLP 层之间巧妙地嵌入了门控单元, 以增强神经网络的表达能力, 并捕获序列数据中的依赖关系。

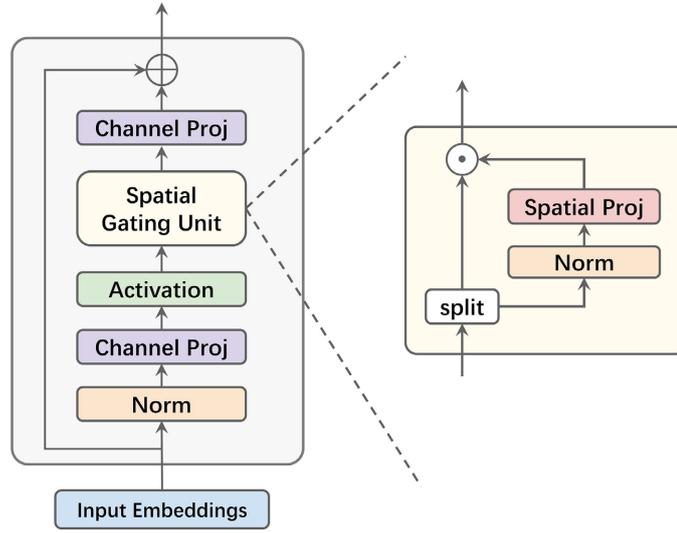


Figure 3. Inner structure of Gated Multilayer Perceptron
图 3. 门控多层感知单元内部结构

门控多层感知单元基本层由具有相同大小和结构的 L 个块的堆叠组成。设 $X \in R^{n \times d}$ 是序列长度 n 和维数 d 的符号表示。每个块定义为：

$$Z = \sigma(XU), \tag{4}$$

$$\tilde{Z} = s(Z), \tag{5}$$

$$Y = \tilde{Z}K \tag{6}$$

其中 σ 是激活函数，通常使用 GeLU。 U 和 K 限定长度等于通道尺寸的线性投影，并且 $s(\cdot)$ 是捕获空间交互的层。

其中空间投影单元是由以下定义的空间交互层：

$$s(Z) = Z \odot f_{w,b}(Z) \tag{7}$$

其中 \odot 表示逐元素乘法，并且 f 是定义为以下的线性投影：

$$f_{w,b}(Z) = WZ + b \tag{8}$$

2.3. 增强因果前馈单元

增强因果前馈(Enhanced Causal Feedforward Unit)是一种用于时间序列处理的多层卷积神经网络结构，适用于时间序列建模任务[23]。这一模块的设计旨在同时考虑两个关键方面：膨胀卷积(Dilated Convolution)和因果卷积(Causal Convolution)。膨胀卷积通过增加卷积核的感受野，能够有效地捕捉输入时间序列中的长期依赖关系，而不增加参数数量，从而提高了网络对序列内部特征的抽取能力。与此同时，因果卷积确保了模型的因果性，即模型只依赖过去的信息，不会使用未来的信息，这对于时间序列预测任务至关重要，以避免未来信息的泄漏[24]。

本文在增强因果前馈层中，首先使用膨胀卷积操作来提取处理后的输入序列中的长期依赖特征。膨胀卷积核的设置允许模块在不增加网络深度的情况下，显著扩展感受野，从而更好地理解时间序列中不同时间尺度上的特征。随后，在膨胀卷积的基础上引入因果卷积操作，确保模块满足因果关系的要求。因果卷积只考虑输入序列中当前时间步之前的信息，不会使用未来的信息，因此在时间序列建模中非常

有效。通过适当的填充设置，我们可以实现因果卷积，将卷积核限制在右侧，只考虑过去的信息，实现了模型避免未来信息的干扰。

最终，如图 4 所示的增强因果前馈模块将膨胀卷积和因果卷积的输出进行组合。这种组合允许模块既能够捕捉长期依赖特征，又具有因果性，确保了其在时间序列预测等任务中长期依赖捕捉的高效性和准确性。

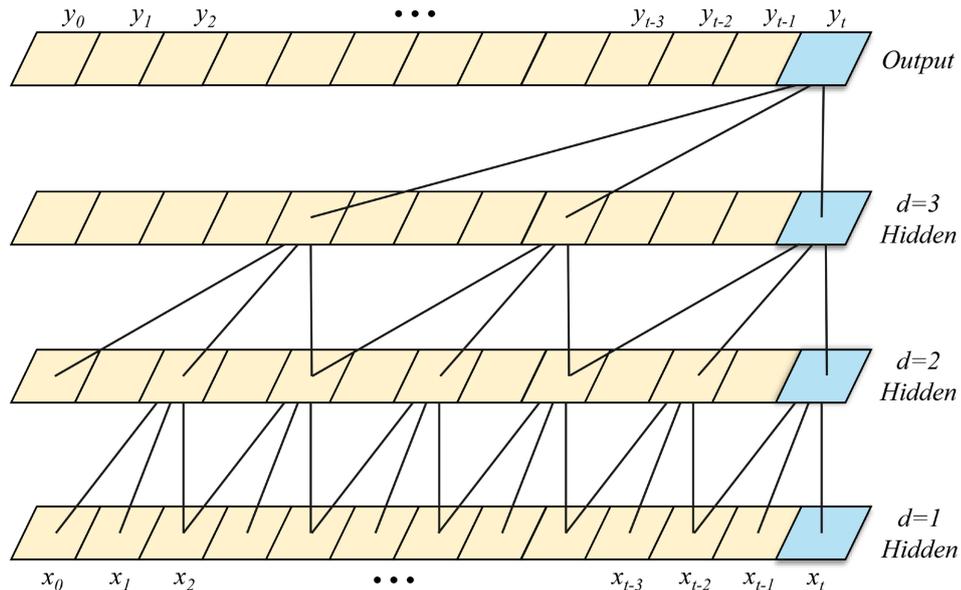


Figure 4. Enhanced causal convolutional structure

图 4. 增强因果卷积结构

2.4. GMEC-Informer

Informer 模型中蒸馏操作将自注意力块与正则卷积堆叠在一起，使感受野受限导致重复无意义的计算，又因为缩减蒸馏操作前，单纯的使用稀疏的自注意力机制会使得捕捉时间序列相关信息的能力下降，同时由于堆叠产生的深度增加引起了网络退化。因此 Informer 虽然在捕获长序列输入信息方面拥有较强的能力，预测性能也优于现有主流预测模型，然而该模型应用于时序预测时仍存在不足。

本文提出融合门控多层感知单元和增强因果的多模态时序融合网络 GMEC-Informer，并针对其内部结构进行优化，将原编码器中单一的注意力机制替换为残差的门控多层感知单元-注意力机制并联的多模态时序特征提取模块，并在编码器块之间增加增强因果卷积提取时间序列长期依赖关系，GMEC-Informer 模型整体架构如图 5 所示。

3. 实验结果与分析

3.1. 数据集选择

为了有效比较模型性能提升的有效性，GMEC-Informer 在电力变压器油温时间序列数据预测数据集上进行了实验，使用了 Informer 模型的基准 ETT 数据集进行评估。ETT (电力变压器温度)是由 Zhou 等人创建的一个独立数据集，旨在研究长短时序预测(LSTF)问题。该数据集收集了中国不同县市的两年数据，创建了小时级数据集(ETTh)。每个数据点包含目标变量“油温”以及高有用负荷(HUFL)、高无用负荷(HULL)等六个功率负荷特性。

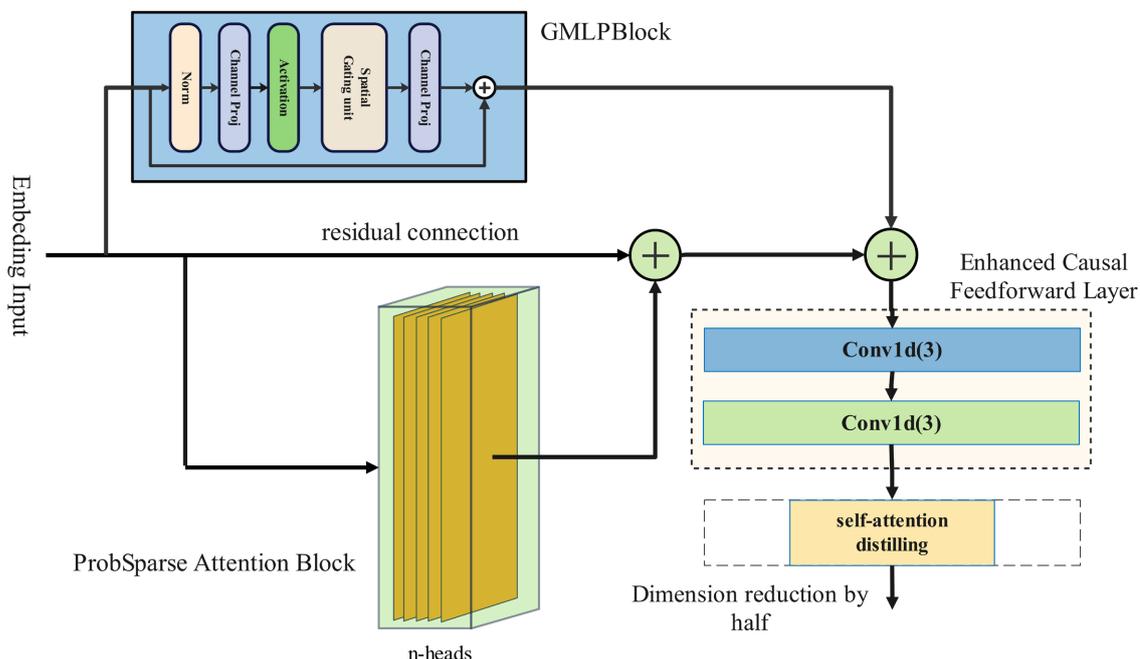


Figure 5. Encoder architecture of GMEC-Informer
 图 5. GMEC-Informer 编码器结构

在本文的研究中，采用 ETTh1 数据集进行了实验，并将其划分为三个部分，即训练集、验证集和测试集，划分比例分别为 7:2:1。这种数据集划分策略的选择旨在支持对模型性能的全面评估，并确保实验结果的可靠性。

3.2. 对照模型和评估指标

为了与本文方法进行比较，本文选择了六种不同的时间序列模型作为参照，分别是 RNN、ARIMA、GRU、LSTM、Informer 以及包含了标准自注意力机制的 Informer+，以评估本文方法在电力变压器油温时间序列数据预测任务中的性能。这六种模型代表了传统的统计方法，以及深度学习方法，提供了广泛的比较基准，有助于全面评测本文方法在该领域的效果。

本研究在 Python 3.7 平台上进行了模型开发，并使用了 PyTorch1.8.0 框架来构建模型。由于训练过程中涉及到大量的矩阵运算，研究选择了 GPU 来加速模型的训练速度，所使用的 GPU 型号是 GeForce RTX 2060，其显存容量为 6 GB。

本文模型处理前均选用了最大最小值归一化方法对数据集进行预处理，有助于将数据缩放到一个统一的范围内，以便模型更好地进行训练。考虑到内存和精度的平衡，批容量均采用为 32。初始学习率采用 $1e-4$ ，优化器采用 Adam。

本实验分别使用如下公式中定义的两个评价指标——平均绝对误差(MAE)和均方误差(MSE)。

$$MAE = \frac{1}{n} \sum_{i=1}^n |y - \hat{y}| \tag{9}$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y - \hat{y})^2 \tag{10}$$

其中， y 表示预测数据的大小， \hat{y} 表示实际值， n 表示预测值的个数。以上两个评价指标数值越低表示模型预测性能越好。

3.3. 实验参数对比分析

在进行实验分析之前，首先需要进行模型编码器端的输入序列长度的比较和分析。适当的输入序列长度可以帮助模型更好地捕捉时间序列数据之间的信息关联，从而优化模型性能。通过分析和优化模型的输入序列长度，可以进一步提高后续预测分析的精度。

本实验以 MSE 作为性能评估指标，分别采用不同天数时间窗口对应编码器输入序列长度为{24, 48, 96, 168, 366, 720}来评估模型的性能。如下图 6 显示了本文模型在不同输入序列长度下的性能评估结果对比。实验结果表明，在编码器的输入序列长度为 96 时，模型的预测性能表现最佳。因此，在后续实验分析中，适合采用输入序列长度为 96 的编码器配置。

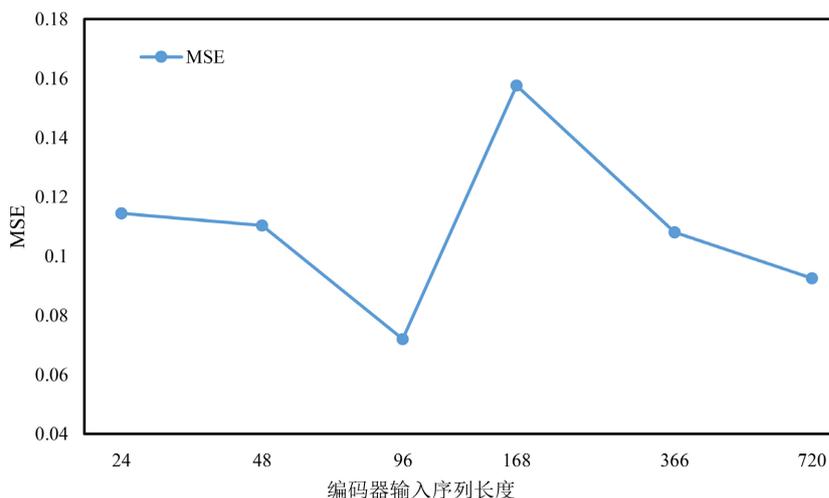


Figure 6. Performance comparison of Encoder input sequence length
图 6. 编码器输入序列长度性能对比

解码器的输入序列长度和编码器一样可以对模型的预测性能产生影响。在本实验中，同样以 MSE 作为性能评估指标，采用不同的解码器输入序列长度{24, 48, 96, 168, 288, 360}来评估模型的性能。如下图 7 展示了本文模型在不同解码器输入序列长度下的性能评估结果对比。

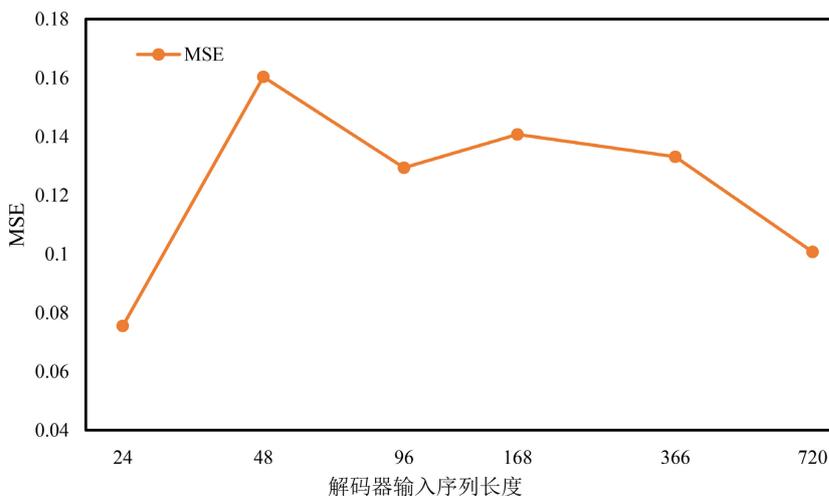


Figure 7. Performance comparison of Decoder input sequence length
图 7. 解码器输入序列长度性能对比

根据实验结果,可以得出结论,当解码器的输入序列长度为 24 时,模型的预测性能最为优秀。因此,在后续的实验分析中,将采用编码器输入长度为 96 和解码器输入长度为 24 的实验参数配置,以保证模型的预测性能。

3.4. 实验参数对比分析

为验证模型预测性能,本实验使用 RNN、ARIMA、LSTM、GRU、Informer、Informer+等 6 种预测模型在不同预测步长下与本文模型的预测性能对比。表 1 中呈现了所有参照模型在 ETT 数据集上的多变量对单变量评估结果,其中预测时间窗口逐渐扩大。为了更明显地突出显示不同预测期的最佳结果,表中采用了粗体的方式对模型在同一个评价指标上的最佳结果进行标注,并对该方法的有效性范围进行了全面检验。

Table 1. Performance of each model in terms of error at different prediction steps

表 1. 各个模型在不同预测步长上的误差表现

模型		预测步长				
		24	48	96	168	366
GMEC-Informer	MSE	0.0995	0.1782	0.1654	0.1603	0.2475
	MAE	0.2535	0.3608	0.3398	0.3350	0.4252
Informer	MSE	0.1244	0.2229	0.2607	0.3307	0.2901
	MAE	0.2879	0.3991	0.4313	0.5014	0.3989
Informer+	MSE	0.1126	0.2739	0.2415	0.3772	0.4058
	MAE	0.2725	0.4518	0.4133	0.5360	0.5632
LSTM	MSE	0.1214	0.1892	0.2034	0.2133	0.3235
	MAE	0.2638	0.1761	0.3472	0.3513	0.4473
GRU	MSE	0.1963	0.2048	0.2297	0.2426	0.2614
	MAE	0.3521	0.3592	0.3790	0.3902	0.4061
ARIMA	MSE	0.1310	0.2246	0.1680	0.3565	0.3678
	MAE	0.2816	0.3709	0.3475	0.3904	0.4037
RNN	MSE	0.1926	0.2194	0.3638	0.4331	0.6010
	MAE	0.3339	0.3498	0.4473	0.4709	0.5645

通过观察表格可以清晰地看出,通过引入 GMLP 和残差网络,本文模型能够联合使用自注意力机制,捕捉多尺度时间信息,从而获取不同时间尺度上的关键信息。另外,增强因果卷积网络的引入有助于保持时间序列信息在参照窗口内的稳定性,避免未来时间信息对当前预测产生不良影响。

因此,本文提出的 GMEC-Informer 模型具有以下显著优点:

1) GMEC-Informer 模型的卓越之处,与其他方法相比,在不同数据集上,GMEC-Informer 模型展现出更出色的预测性能,其误差较低的结果显著突出了其优势。这进一步突显了 GMEC-Informer 在时间序列预测领域的独特性和卓越性能,为提升预测准确性和可靠性提供了有力支持。

2) 对于 MAE 的评价指标,本文采用的 GMEC-Informer 模型在此评价指标上的最优性能个数均大于其他数理统计和深度学习模型,

3) 以编码器-解码器为基本架构的 GMEC-Informer 模型预测性能显著优于递归神经网络 LSTM 模型。该方法的 MSE 分别降低了 18.03% (24)、5.81% (48)、18.68% (96)、24.85% (366)和 23.49% (720)。结果表明, 基于编码器-解码器和自注意机制的模型比基于 RNN 的模型具有更好的预测能力。

相同的数据集下, GMEC-Informer 模型在短期负荷预测方面表现卓越, 相较于中长期负荷预测。这一趋势可归因于以下因素: 在短期负荷预测中, 影响负荷变化的因素相对较少, 而在中长期负荷预测中, 涉及的变量更加繁多, 包括但不限于环境温度和天气变化等, 这些因素对预测准确性具有显著影响。

3.5. 消融实验

要研究本文模型中各个组件部分对预测性能的影响, 以 GMEC-Informer 为基础, 进行了消融实验。在实验过程中, 移除多尺度空间信息提取模块的模型指代为 no GMLP, 同时移除增强因果卷积的模型指代为 no EC。此外, GMEC-Informer 与 no GMLP、no EC 以及 Informer 在不同条件下的性能也在表 2 中进行了对比分析。

Table 2. Comparison of ablation experiments of different modules

表 2. 不同模块消融实验对比

模型		预测步长				
		24	48	96	168	366
GMEC-Informer	MSE	0.0995	0.1783	0.1654	0.1603	0.2475
	MAE	0.2535	0.3608	0.3398	0.3350	0.4252
no GMLP	MSE	0.1187	0.1965	0.1984	0.3021	0.3946
	MAE	0.2860	0.3716	0.3750	0.4745	0.5578
no EC	MSE	0.1147	0.1952	0.2367	0.3012	0.2960
	MAE	0.2738	0.3688	0.4167	0.4775	0.4698
Informer	MSE	0.1244	0.2229	0.2607	0.3307	0.2901
	MAE	0.2879	0.3991	0.4313	0.5014	0.3989

在消融实验中, GMEC-Informer 模型相较于 Informer 模型在评价指标上只有一个指标稍有下降, 在其他所有指标上都表现出了最优水平。这表明 GMEC-Informer 模型在多个方面都取得了明显的改进, 特别是在那些被优化的指标上。以下是对这一结果的一些总结和分析:

GMEC-Informer 模型在多个指标上表现最优, 这表明该模型总体性能提升, 在处理时间序列预测任务时, 具有更好的综合性能。这对于应用于实际场景中的时间序列问题来说, 具有非常积极的意义。

尽管 GMEC-Informer 在大多数指标上超越了 Informer 模型, 但在预测步长为 366 时指标稍微差于 Informer。这种情况可能是由于预测时间步长的增加, 模型预测性能均出现明显下降和不稳定导致的。

在时间序列任务中, 不同的指标可能对不同的应用场景具有不同的重要性。因此, 在选择最佳模型时, 需要考虑到具体应用的需求。GMEC-Informer 模型在使用 MSE 作为评价指标的多数情况下表现优越, 对于不同的预测场景, 则可以权衡模型选择。

4. 结论

本研究创新性地提出了一种融合门控多层感知机和增强因果卷积的多模态时序融合网络——GMEC-Informer, 并用于电力负荷预测任务。该模型的优势体现在两方面: 相较于传统的全局注意力模

型, GMEC-Informer 采用门控多层感知单元, 能够更为精准地捕捉电力负荷随时间演变的局部时空模式, 即动态变化中蕴含的局部特征及其在时空上的相互作用; 通过引入增强因果前馈层, 模型能够在信息传递过程中避免与未来信息发生不必要的交互, 确保了预测仅基于过去已知数据, 有利于提升预测的稳健性和可信度。

实验结果证实了 GMEC-Informer 在处理电力变压器油温数据集时, 能够成功捕获电力负荷复杂的时空依赖关系, 展现出优异的预测性能, 优于现有的同类模型。这不仅验证了模型的有效性, 也为电力系统的高效运营与管理提供了强有力的支持工具。

未来研究将继续探索时间序列信息融合的先进机制, 旨在将这些方法与注意力机制深度整合, 以进一步提升模型在处理长预测步长任务时的精度。此外, 还计划将 GMEC-Informer 拓展至其他类型的时间序列预测任务, 以验证其在不同应用场景中的普适性和适用性, 为更广泛领域的预测问题提供高性能解决方案。

参考文献

- [1] 史建楠, 邹俊忠, 张见, 等. 基于 DMD-LSTM 模型的股票价格时间序列预测研究[J]. 计算机应用研究, 2020, 37(3): 662-666.
- [2] Song, W., Wang, C., Dong, T., *et al.* (2023) Hierarchical Extraction of Cropland Boundaries Using Sentinel-2 Time-Series Data in Fragmented Agricultural Landscapes. *Computers and Electronics in Agriculture*, **212**, Article 108097. <https://doi.org/10.1016/j.compag.2023.108097>
- [3] 张栗棕, 王谨平, 刘贵松, 等. 面向金融数据的神经网络时间序列预测模型[J]. 计算机应用研究, 2018, 35(9): 2632-2637.
- [4] 江洋洋, 金伯, 张宝昌. 深度学习在自然语言处理领域的研究进展[J]. 计算机工程与应用, 2021, 57(22): 1-14.
- [5] Ma, R., Angryk, R. and Scherer, R. (2022) Special Issue on Deep Learning for Time Series Data. *Neural Computing and Applications*, **34**, 13147-13148. <https://doi.org/10.1007/s00521-022-07536-3>
- [6] Singh, A., Srivastava, M.K. and Singh, N.K. (2019) AI-Based Short-Term Electric Time Series Forecasting. *International Journal of Innovative Technology and Exploring Engineering*, **8**, 3255-3261.
- [7] 谭风雷, 徐刚, 李义峰, 等. 基于相似日和相似时刻的变压器顶层油温预测方法[J]. 电力工程技术, 2022, 41(2): 193-200.
- [8] Wang, X. and Wang, Y. (2016) A Hybrid Model of EMD and PSO-SVR for Short-Term Load Forecasting in Residential Quarters. *Mathematical Problems in Engineering*, **2016**, Article ID: 9895639. <https://doi.org/10.1155/2016/9895639>
- [9] 杨汪洋, 魏云冰, 罗程浩. 基于 CVMD-TCN-BiLSTM 的短期电力负荷预测* [J/OL]. 电气工程学报, 1-10. <http://kns.cnki.net/kcms/detail/10.1289.TM.20230601.1229.002.html>, 2024-02-20.
- [10] Salinas, D., Flunkert, V., Gasthaus, J., *et al.* (2020) DeepAR: Probabilistic Forecasting with Autoregressive Recurrent Networks. *International Journal of Forecasting*, **36**, 1181-1191. <https://doi.org/10.1016/j.ijforecast.2019.07.001>
- [11] 崔星, 李晋国, 张照贝, 等. 基于改进粒子群算法优化 LSTM 的短期电力负荷预测[J]. 电测与仪表, 2024, 61(1): 131-136.
- [12] Kim, M., Lee, S. and Jeong, T. (2023) Time Series Prediction Methodology and Ensemble Model Using Real-World Data. *Electronics*, **12**, Article 2811. <https://doi.org/10.3390/electronics12132811>
- [13] 田英杰, 苏运, 郭乃网, 等. 基于时间序列嵌入的电力负荷预测方法[J]. 计算机应用与软件, 2018, 35(11): 55-60, 73.
- [14] Sesti, N., Garau-Luis, J.J., Crawley, E., *et al.* (2021) Integrating LSTMS and GNNS for COVID-19 Forecasting. arXiv preprint arXiv:2108.10052. <https://doi.org/10.48550/arXiv.2108.10052>
- [15] 代业明, 周琼. 基于改进 Bi-LSTM 和 XGBoost 的电力负荷组合预测方法[J]. 上海理工大学学报, 2022, 44(2): 138-147.
- [16] Knoll, G. and Lindner, B. (2022) Information Transmission in Recurrent Networks: Consequences of Network Noise for Synchronous and Asynchronous Signal Encoding. *Physical Review E*, **105**, Article 044411. <https://doi.org/10.1103/PhysRevE.105.044411>
- [17] Vaswani, A., Shazeer, N., Parmar, N., *et al.* (2017) Attention Is All You Need. *Proceedings of the 31st International*

Conference on Neural Information Processing Systems, Long Beach, 4-9 December 2017, 600-601.

- [18] 刘文婷, 卢新明. 基于计算机视觉的 Transformer 研究进展[J]. 计算机工程与应用, 2022, 58(6): 1-16.
- [19] Zhou, H., Zhang, S., Peng, J., *et al.* (2021) Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, **35**, 11106-11115. <https://doi.org/10.1609/aaai.v35i12.17325>
- [20] 张松慧, 熊汉江. 融合地理社交和时间序列信息嵌入排名位置推荐模型[J]. 计算机应用研究, 2019, 36(9): 2618-2624.
- [21] Xu, H., Peng, Q., Wang, Y., *et al.* (2023) Power-Load Forecasting Model Based on Informer and Its Application. *Energies*, **16**, Article 3086. <https://doi.org/10.3390/en16073086>
- [22] Liu, H., Dai, Z., So, D., *et al.* (2021) Pay Attention to MLPs. In: Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P.S. and Wortman Vaughan, J., Eds., *Advances in Neural Information Processing Systems 34*, MIT Press, Cambridge, 9204-9215.
- [23] 宋绍剑, 姜屹远, 刘斌. 一种 TCN 的改进模型及其在短期光伏功率区间预测的应用[J]. 计算机应用研究, 2023, 40(10): 3064-3069.
- [24] 周蜀杰, 曾园园, 江昊. 基于扩张因果卷积的城市客流量预测算法[J]. 武汉大学学报(工学版), 2023, 56(2): 218-225.