

# 基于GGSAGCformer的多区域短期电荷负载预测

贾艳红, 何利文

南京邮电大学物联网学院, 江苏 南京

收稿日期: 2024年6月20日; 录用日期: 2024年8月21日; 发布日期: 2024年8月28日

## 摘要

为确保电力系统的安全与稳定, 需对未来一天内各时段的电力需求进行精准预测。然而, 随着可再生能源的增加, 电荷预测变得更加复杂和不可预测。因此, 本文介绍了一个基于GCformer构建的新模型GGSAGCformer, 通过图卷积神经网络和门控循环单元提取数据中的空间和时序特征, 再引入多头注意力机制(Multi-head Attention Mechanism Layer)。在此层中加入了地理相似空间自注意力模块(Geographically Similar Spatial Self-Attention, GSSA)和气候自注意力模块(Climate Self-Attention, CSA), 旨在深入探索数据中的潜在关联, 输出层使用GCformer来处理预测结果, 以提升预测的准确性。实验结果显示, 本文模型在输出步长为192的情况下, 与传统模型GCformer、Informer和Reformer相比, MSE分别降低了15.3%、25%和29.3%。

## 关键词

时间序列预测, 多头自注意力机制, 多区域电荷短期预测, GCformer

# Multi-Region Short-Term Charge Load Prediction Based on GGSAGCformer

Yanhong Jia, Liwen He

School of Internet of Things, Nanjing University of Posts and Telecommunications, Nanjing Jiangsu

Received: Jun. 20<sup>th</sup>, 2024; accepted: Aug. 21<sup>st</sup>, 2024; published: Aug. 28<sup>th</sup>, 2024

## Abstract

In order to ensure the safety and stability of the power system, it is necessary to accurately predict

the power demand at each time of the day in the future. However, with the increase in renewable energy, charge prediction has become more complex and unpredictable. Therefore, this paper introduces a new model based on GCformer, GGSAGCformer, which extracts spatial and temporal features from the data through graph convolutional neural network and gated recurrent unit, and then introduces the multi-head attention mechanism layer. The Geographically Similar Spatial Self-Attention (GSSA) and Climate Self-Attention (CSA) modules are added to this layer to deeply explore the potential associations in the data, and the output layer uses GCformer to process the prediction results to improve the accuracy of the prediction. The experimental results show that the MSE of the proposed model is reduced by 15.3%, 25% and 29.3% compared with the traditional models GCformer, Informer and Reformer when the output step size is 192, respectively.

## Keywords

Time Series Prediction, Multi-Head Self-Attention Mechanism, Short-Term Prediction of Multi-Region Charges, GCformer

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

电力负荷预测对于电网的稳定运行和高效规划至关重要[1]。这一预测可以分为两大类方法：统计学方法与人工智能方法[2]。统计学方法如自回归移动平均模型(Autoregressive Integrated Moving Average Model, ARIMA) [3]依赖历史数据分析。而人工智能方法通过分析大量历史数据建立模型, 包括神经网络(Artificial Neural Network, ANN) [4]、支持向量回归(Support Vector Regression, SVR) [5] [6]和模糊推理系统(Fuzzy Inference System, FIS) [7] [8]。尽管这些方法各有优势, 但也存在限制, 如 SVR 在处理大数据集上能力受限, ANN 在时间特征选择上可能引入误差。

深度学习技术[9]如 LSTM [4] [10]已提高电荷负载预测的准确性, 而结合 CNN [11]和 LSTM 的方法虽能提取时序特征却缺乏空间特征处理[12]。图神经网络(GNN)及其与 GCN 和 TCN [13]的结合解决了这一问题, 提升了预测精度。K-means 方法[14]结合自适应时空同步图卷积神经网络也显示了潜力。尽管如此, LSTM 和 GRU 在处理长期数据上效率低, CNN 在全局上下文理解上有限, GNN 在计算复杂度且处理非图结构数据上能力有限。Transformer 模型[15]因其处理复杂依赖关系的能力而在时间序列预测中日益流行, 尽管面临长序列数据的高计算复杂度和高内存消耗的挑战。为此, 学者们已提出多种优化方案[16]-[18]以提升其性能, 其中改进的 Transformer 在实际应用中表现出减少计算负担和处理长短期依赖性问题的优势。因此, 本研究选择了性能优越的变体 GCformer [19]作为核心模型, 并结合多种注意力机制进行精确的多区域时空电荷负载预测。

## 2. 时空特征提取

本文提出了一种新的预测模型 GGSAGCformer (GCN GRU Self-Attention GCformer), 基于 GCformer 并加入了卷积神经网络和注意力模块, 以处理多地区、多因素影响下的电荷预测问题。该模型综合学习时空、气候及区域结构信息, 有效捕获短期时间、气候和空间特征。在卷积层使用 GCN 提取多区域空间特征和 GRU 提取多区域时间特征, 模型整合了不同时间尺度的电荷负载数据, 包括上周、昨天及当前时刻的数据, 以捕捉时间的连续性和周期性。在注意力机制层, 引入了地理相似空间、气候注意力模块,

深入挖掘数据的长期时空和气候相关性。通过卷积层和自注意力机制的综合, 模型能全面捕捉电荷负载数据的复杂性, 提高预测精度。

## 2.1. 基于 GCN 提取多区域空间特征

在多区域电荷负载预测中, 特征提取依赖于区域时空相关图, 该图由表示各区域(节点)和区域间时空相关性的链路(边)组成。这些节点和边具有独特的属性和连接顺序, 属于非欧式数据类型, 因而无法通过传统处理欧式数据的 CNN 进行处理。GCN 属于 GNN 中通过谱方法处理图数据的模型, 利用图卷积定理从谱域定义图卷积。特征提取过程主要通过聚合目标节点及其所有邻居节点的特征信息实现。如图 1 所示, 设节点 1 为目标节点, GCN 模型能够通过邻接矩阵汇聚其一跳邻居节点 2、3、5 的特征信息, 从而实现节点 1 电荷负载特征的提取。

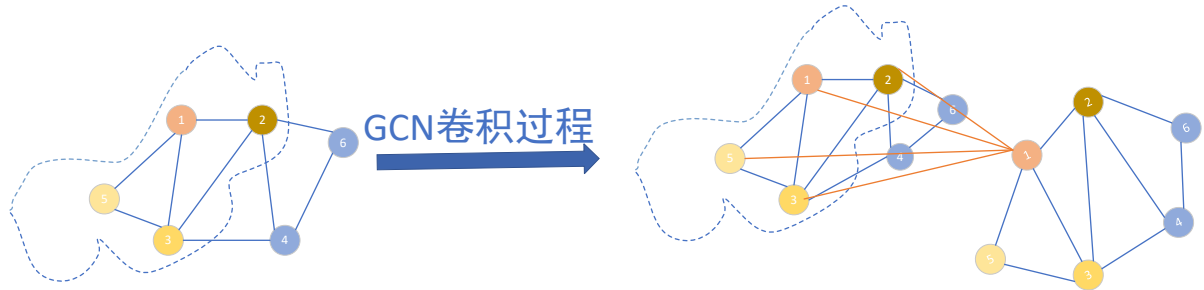


Figure 1. GCN convolution procedure  
图 1. GCN 卷积过程

使用  $l$  层 GCN 来对区域时空相关图  $G_T$  进行处理的过程为:

$$H_T^l = \text{ReLU} \left( D_T^{-\frac{1}{2}} A_T D_T^{-\frac{1}{2}} H_T^{l-1} W^l \right) \quad (1)$$

其中,  $A_T = A_T + I$ ,  $I$  为单位矩阵;  $D_T$  为  $A_T$  的度矩阵,  $\hat{d}_T^i = \sum_{j=1}^{|V|} \hat{a}_T^{ij}$ ;  $H_T^{l-1} \in \mathbb{R}^{|V| \times C}$  为第  $l-1$  层的特征矩阵,  $C$  为特征维度,  $H_T^0 = X_T \in \mathbb{R}^{|V| \times T}$ ;  $W^l \in \mathbb{R}^{C \times F}$  为第  $l$  层待训练的权重矩阵,  $F$  为输出的区域特征维度;  $\text{ReLU}(\bullet)$  为激活函数, 增加了非线性特征; 最终为了方便表示, 使用  $H_T = H_T^l$  来表示由 GCN 模型生成的输出矩阵。

## 2.2. 基于 GRU 提取多区域时间特征

在多区域电荷负载预测中, 目前常用的模型为 LSTM 和 GRU。LSTM 模型设计了内部的“门”机制, 而 GRU 是 LSTM 的简化版本, 将 LSTM 中的三个门控单元简化为两个, 并且没有设置记忆单元  $c_t$ 。与 LSTM 相比, GRU 能够更有效地避免梯度消失和爆炸的问题。对于 GCN 提取到的多区域空间特征  $H_T$ , 分别为其分配对应的 GRU 进行处理, 即每一个  $H_T^v, v_i \in V$  都分配一个 GRU 提取数据的时序特征。GRU 的计算过程如下:

$$\mathbf{r}_t = \text{sigmoid} \left( \mathbf{W}_{ir} \mathbf{H}_t + \mathbf{b}_{ir} + \mathbf{W}_{hr} \mathbf{h}_{(t-1)} + \mathbf{b}_{hr} \right) \quad (2)$$

$$\mathbf{z}_t = \text{sigmoid} \left( \mathbf{W}_{iz} \mathbf{H}_t + \mathbf{b}_{iz} + \mathbf{W}_{hz} \mathbf{h}_{(t-1)} + \mathbf{b}_{hz} \right) \quad (3)$$

$$\mathbf{n}_t = \tanh \left( \mathbf{W}_{in} \mathbf{H}_t + \mathbf{b}_{in} + \mathbf{r}_t \odot \left( \mathbf{W}_{hn} \mathbf{h}_{(t-1)} + \mathbf{b}_{hn} \right) \right) \quad (4)$$

$$\mathbf{h}_t = (1 - \mathbf{z}_t) \odot \mathbf{n}_t + \mathbf{z}_t \odot \mathbf{h}_{(t-1)} \quad (5)$$

其中,  $H_t$  为时刻  $t$  的输入数据特征;  $h_t$  为时刻  $t$  的 GRU 输出;  $h_{(t-1)}$  为时刻  $t-1$  的隐藏层状态;  $r_t$  和  $z_t$  分别为重置门和更新门;  $n_t$  为候选隐藏状态;  $\{W, b\}$  为对应的可训练参数;  $\text{sigmoid}(\cdot)$  和  $\sigma$  为激活函数;  $\odot$  代表 Hadamard 乘积。

### 3. 基于多头注意力机制的特征提取

多头注意力机制在处理序列数据时通过将输入分配到多个子空间, 并在每个子空间内独立实现注意力机制, 有效提高了模型对复杂模式的识别能力, 尤其适用于电荷负载预测。这种方法不仅能捕捉电荷负载数据中的动态变化, 从而优化预测的准确性和细致度, 而且在实际应用中能快速响应突发事件, 辅助识别异常模式, 为电网运营的稳定与资源配置提供支持。因此, 本研究引入了 2 种注意力机制层, 全面分析影响电荷负载的关键数据特征, 以增强系统的安全性和稳定性。

#### 3.1. 地理相似空间注意力机制层

考虑到相邻地区可能有相似的负荷模式, 本文采用论文[20]中的 DTW 算法, 构建一个延迟感知的电荷负载特征转换模块, 名为地理相似空间自注意力模块(Geographically Similar Spatial Self-Attention, GSSA)。该模块从每个节点的短期历史电荷负载数据中捕捉时间延迟, 并将这些信息集成到地理空间自注意力模块的关键矩阵中, 以精确模拟电荷信息的时间传播延迟。此方法以一维数据示例进行说明。

首先, 我们从历史电荷负载数据中识别出一组代表性的短期电荷模式。具体来说, 我们使用大小为  $S$  的滑动窗口对历史电荷数据进行切片, 获取一系列的电荷序列。接着, 我们应用  $k$ -Shape 聚类算法对这些电荷序列进行聚类,  $k$ -Shape 算法是一种保留时间序列形状并对缩放和移位具有不变性的时间序列聚类方法。我们用每个簇的质心  $p_i$  表示该簇,  $p_i$  是一个长度为  $S$  的时间序列。然后, 我们用集合  $\rho = \{p_i \in [1, \dots, N_p]\}$  表示聚类结果, 其中  $N_p$  是聚类总数。这样,  $\rho$  可以视为一组短期电荷模式。

类似的电荷模式可能对邻近节点的电荷状况产生相似的影响。因此, 我们将每个节点的历史电荷序列与提取的电荷模式集  $\rho$  进行比较, 将相似模式的信息融合到每个节点的历史电荷序列表示中。具体来说, 给定节点  $n$  从时间片  $(t-S+1)$  到  $t$  的  $S$  步历史电荷流序列  $x_{(t-s+1:t),n}$ , 我们首先使用嵌入矩阵  $W^u$  获得高维表示  $u_{t,n}$  为:

$$u_{t,n} = x_{(t-s+1:t),n} W^u \quad (6)$$

随后, 我们使用另一个嵌入矩阵  $W^m$  将电荷模式集合  $\rho$  中的每个电荷序列转换为记忆向量:

$$m_i = p_i W^m \quad (7)$$

将节点  $n$  的历史电荷流表示  $u_{t,n}$  与电荷模式记忆向量  $m_i$  进行比较, 得到相似向量为:

$$w_i = \text{softmax}(u_{t,n}^T m_i) \quad (8)$$

接下来根据相似度向量  $w$  对电荷模式集合  $P$  进行加权求和, 得到综合历史序列表示  $r_{t,n}$  为:

$$r_{t,n} = \sum_{i=1}^{N_p} \left( \text{softmax}(u_{t,n}^T m_i) \right) \cdot (p_i W^c) \quad (9)$$

其中,  $W^c$  是可学习的参数矩阵。综合历史序列表示  $r_{t,n}$  含有节点  $n$  的所有过去电荷信息, 时间区间为  $(t-S+1)$  到  $t$ 。最后, 我们使用所有节点的集合表示  $R_t$  更新关键  $K_t^{(S)}$ :

$$\tilde{K}_t^{(S)} = K_t^{(S)} + R_t \quad (10)$$

其中,  $R_t \in \mathbb{R}^{N \times d'}$  通过连接所有节点的集合表示得到  $r_{t,n}$ ,  $d'$  是维度。通过这种方式, 新的关键矩阵  $\tilde{K}_t^{(S)}$  能够全面包含从时间段  $(t - S + 1)$  至  $t$  内所有节点的历史电荷信息。在进一步计算空间依赖关系时, 我们利用查询矩阵与这一新的关键矩阵进行乘积运算。这一过程中, 查询矩阵能够充分考虑到其他节点的历史电荷状况, 从而精确模拟空间信息传播中可能存在的时间延迟效应。

### 3.2. 气候注意力机制层

鉴于电力负荷受气候因素的随机性影响较大, 分析电力负荷与气候因素之间的相关性对于选择适当的预测模型输入变量非常重要。因此, 使用皮尔逊相关系数来检验负荷值与气候数据之间的相关性, 公式如下:

$$\rho = \frac{\sum_{d=1}^a (x_d - \bar{x})(y_d - \bar{y})}{\sqrt{\sum_{d=1}^a (x_d - \bar{x})^2} \sqrt{\sum_{d=1}^a (y_d - \bar{y})^2}} \quad (11)$$

其中,  $a$  为数据点的数量;  $x_a$  和  $y_a$  分别是变量  $x$  和  $y$  第  $d$  个数据点的值;  $\bar{x}$  和  $\bar{y}$  为均值; 皮尔逊相关系数  $\rho$  的值介于-1到1之间。如图2相关性分析显示, 温度、湿度、气压和风速与电力负荷具有显著相关性, 因此被选为预测模型的输入参数。相比之下, 风向与电力负载的相关性较低, 因此在本次实验中未将其作为模型参数。

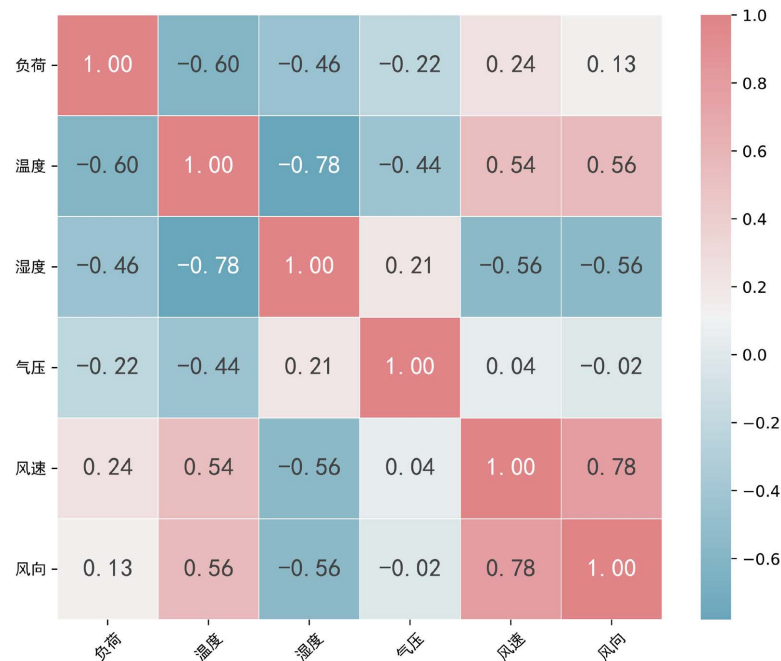


Figure 2. Correlation analysis of climatic factors

图2. 气候因素相关分析

所以文本设计了一种气候自注意力机制层(Climat Self-Attention, CSA), 该机制通过专注于与电荷负载密切相关的气候特征来提高预测精度。气候注意力机制层的实现, 其核心在于利用注意力机制的基本原理, 即动态地分配处理资源。这一机制确保了模型能够聚焦于气候因素中最为重要和相关的信息。给定一个输入特征集  $X \in \mathbb{R}^{N \times F}$ , 其中  $N$  是样本数量,  $F$  是特征数量, 该机制首先通过一个可学习的参数矩

阵  $W_c \in \mathbb{R}^{F \times F}$  来计算每个特征的加权重要性:

$$A_c = \text{softmax}(X)W_c \quad (12)$$

其中,  $A_c \in \mathbb{R}^{N \times F}$  是每个特征的注意力权重,  $\text{Softmax}$  函数确保了每个样本中所有特征权重的和为 1。接着, 通过与原始特征  $X$  的点乘, 模型加强了重要特征的影响, 同时减少了不那么重要特征的作用:

$$X' = A_c \odot X \quad (13)$$

其中,  $\odot$  表示元素间的乘法,  $X' \in \mathbb{R}^{N \times F}$  是加权后的特征矩阵。最后, 通过一个线性变换将加权特征矩阵  $X'$  转换为预测电荷负载所需的维度, 并通过  $\text{Dropout}$  进行正则化, 以防止过拟合:

$$Y = \text{Dropout}(\text{ReLU}(W_o X' + b_o)) \quad (14)$$

其中,  $W_o \in \mathbb{R}^{F \times K}$  和  $b_o \in \mathbb{R}^K$  是输出层的权重和偏置,  $K$  是输出的目标维度(如未来电荷负载),  $\text{ReLU}$  和  $\text{Dropout}$  是激活和正则化函数。

通过上述步骤, 气候注意力机制能够自动识别和聚焦于对预测任务最为关键的气候特征, 从而在考虑到天气影响的情况下, 有效提升电荷负载预测的准确性。此方法的一大显著优势在于其具备动态调整特征权重的能力, 这使得模型能够依据不同的输入和上下文条件灵活调整, 从而优化预测性能。

## 4. 基于 GCformer 的时序特征提取

GCformer 是一种用于多维时间序列预测的模型, 特别适用于短期和长期预测。该模型的主要创新之处在于它结合了全局卷积分支和本地分支变换, 有效地处理长度可变的时间序列。全局卷积分支专注于捕捉长期依赖, 而本地分支关注短期趋势, 确保模型对即时变化敏感。这种双分支结构既扩展了模型的应用范围, 也提升了预测的准确性和效率。

### 4.1. 全局卷积核

先前的研究已经证明了全局卷积核在捕获长短期依赖性方面的功效。具体来说, 原作者采用了跨越整个序列长度的冗长内核, 从而能够捕获长期依赖性。给定输入序列  $u \in \mathbb{R}^{n \times d}$ 、可学习的全局内核、 $k \in \mathbb{R}^{n \times d}$  以及输出  $y \in \mathbb{R}^{n \times d}$ 。全局卷积核表示为:

$$y = u * k \quad (15)$$

其中,  $*$  是卷积算子。虽然全局卷积运算的复杂度为  $O(n^2)$ , 但可以使用表示为  $\mathcal{F}$  的快速傅立叶变换快速实现, 其复杂度为  $O(N \log N)$ , 计算公式如下:

$$u * k = \mathcal{F}^{-1}(\mathcal{F}(u)\mathcal{F}(k)) \quad (16)$$

### 4.2. 高效参数化

为了控制参数数量仅次线性地缩放到序列长度, SGconv 通过组装一系列尺寸逐渐变大的子核来构造卷积核, 其中每个子核使用插值技术从相等数量的参数上进行采样。此外, 作者采用加权衰减的子内核组合, 为扩展序列建模提供有利的归纳偏差, 以提升模型性能。给定一个用多尺度子内核参数化的内核  $k_{mask}$ , 全局卷积模型  $Gconv_{mask}$  的过程定义为:

$$y = \mathcal{F}^{-1}(\mathcal{F}(u)\mathcal{F}(k_{mask})) \quad (17)$$

频域参数化提供了一种新的全局卷积核参数化方法, 利用时间序列数据的时频变换偏差。与时域内核 SGconv 不同, 此方法在频域中创建内核  $K_{freq} \in \mathbb{C}^{m \times d}$ , 且为保持次线性尺度, 约束  $m \ll n$ 。这种频域参数化的全局卷积模型称为( $Gconv_{freq}$ ), 过程定义为:

$$y = \mathcal{F}^{-1}(\mathcal{F}(u)\mathcal{F}(k_{freq})) \tag{18}$$

用状态空间模型来处理顺序数据这种方法为我们带来了一种新的思路, 可以将傅立叶变换之外的核参数化方法以一种不同于现有的方法来创新。S4 [21]使用状态空间建模程序:  $x_k = Ax_{k-1} + Bu_k$ ,  $y_k = Cx_{k-1} + Du_k$ , 其中  $u^k \in \mathbb{R}^a$  是时间步长  $k$  的输入信号,  $x^k \in \mathbb{R}^d$  是隐藏状态空间模型,  $y \in \mathbb{R}^a$  是输出信号。

状态转移矩阵  $A$  和输入矩阵  $B$  定义为:

$$A_{nk} = (2n+1) \begin{cases} (-1)^{n-k}, & \text{if } (k \leq n) \\ 1, & \text{if } (k \geq n) \end{cases} \tag{19}$$

$$B_n = (2n+1)(-1)^n$$

这里矩阵  $A \in \mathbb{R}^d$  和  $B \in \mathbb{R}^{d \times 1}$  是通过 LegT 度量导出的, 该度量为最近的历史分配统一的权重。  $C \in \mathbb{R}^{1 \times d}$  和  $D \in \mathbb{R}^{1 \times 1}$  是输出矩阵。状态空间模型的循环计算可以通过卷积方式快速计算。转换矩阵  $A$ 、 $B$  和  $C$  是预先定义的, 以便可以提前计算内核  $K$ :

$$y = u * K, K = (CB, CAB, \dots, CA^{N-1}B) \tag{20}$$

转移矩阵  $A$  源自勒让德多项式。因此,  $u * K$  可以视为一个将原始信号从时间 - 空间领域映射到 Legendre 空间的过程, 这与通过傅立叶变换将信号转换到频率空间的操作类似。在此框架下, 信号的映射从时间 - 空间到 Legendre 空间通过  $\bar{u} = \text{LegT.Project}(u)$  来实现, 而反之过程的信号重建则通过  $\bar{u} = \text{LegT.Project}(\bar{u})$  完成。此外, 设定  $K_{leg} \in \mathbb{R}^{m \times d}$  为内核, 其中  $m$  远小于  $n$ , 这定义了具备平移 Legendre 度量的全局卷积模型 ( $Gconv_{leg}$ ) 的操作过程:

$$y = \text{legt.Reconstruct}(\text{legt.Project}(u) * K_{leg}) \tag{21}$$

### 4.3. 全局与局部分支的协同融合

原作者提出的 GCformer 架构采用独特的双分支设计, 模型结构如图 3 所示。这种设计涉及同时通过两个独立的分支传递输入数据, 每个分支都专门设计用于捕获和提取不同类型的信息: 局部和全局。特定的解码器模块可以通过最大化其互补性的方式进行集成和融合这两种信息。

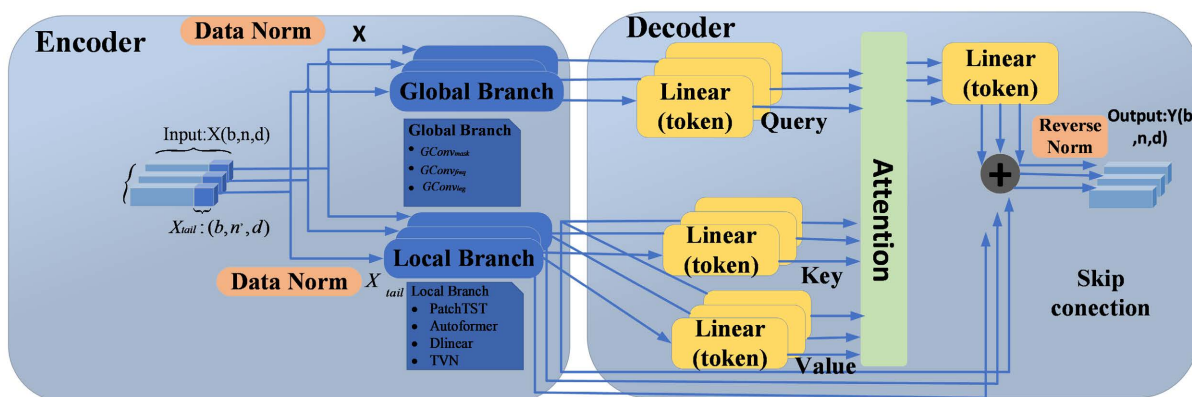


Figure 3. GCformer structure  
图 3. GCformer 结构

在编码器部分, 设置了两个并行分支。上分支用于提取全局信息 ( $Z_{global}$ ), 关注序列中的长期依赖关

系, 并通过馈送整个输入序列  $X \in \mathbb{R}^{N \times d}$  到全局分支处理, 其复杂性与序列长度成次线性关系, 解决了变压器内存要求高的问题。下分支则专注于捕获局部信息( $Z_{local}$ ), 涉及近期时间节点的依赖关系, 通过将序列的尾部  $X_{tail} \in \mathbb{R}^{N' \times d}$  (其中  $N' < N$ ) 馈送到 Transformer 分支, 旨在降低复杂性, 同时保持预测精度。公式如下:

$$z_{global} = Branch_{global}(X) \tag{22}$$

$$z_{local} = Branch_{local}(X_{tail}) \tag{23}$$

解码器模块通过融合全局信息( $Z_{global}$ )和局部信息( $Z_{local}$ )来提高信息利用效率, 并输出预测结果。该模块包含一个交叉注意力单元, 专门用于有效表现时间序列中的历史信息。在此单元中, 全局和局部特征首先被映射到令牌级别的潜在维度, 然后使用全局信息作为查询( $q$ ), 局部信息作为键( $k$ )和值( $v$ ), 以进行数据处理。这种结构确保了全局和局部信息的充分整合和有效利用。相关的计算公式为:

$$q = MLP(z_{global}) \tag{24}$$

$$k = MLP(z_{global}), v = MLP(z_{local}) \tag{25}$$

通过查询全局信息和本地信息, 可以有效集成全局和本地信息。公式为:

$$Atten(q, k, v) = \text{softmax}\left(\frac{qk^T}{\sqrt{d_q}}\right)v \tag{26}$$

通道独立性是指数据集中多个变量之间不存在相关性的假设。这个概念是在之前的研究中引入的, 并在多个数据集中展示了改进。作者也将这个想法融入到了工作中。

#### 4.4. GGSAGCformer 的多区域电荷负载预测模型

GGSAGCformer 是一个专为电荷负载预测设计的高效深度学习模型, 结合了 GCformer 的全局卷积分支, 并引入了卷积神经网络和多头注意力机制。此模型能有效捕捉电荷负载数据的复杂时间依赖性和多尺度变化。通过图卷积网络 GRU 和 GCN 提取多区域的时间空间信息, 再使用注意力模块提取相似空间信息和气候信息关联, 模型能同时关注序列数据的多个关键特性, 在不同的子空间中捕获更丰富的信息, 增强了数据解析能力。GCformer 的全局卷积分支能够捕捉长期依赖关系如季节性趋势, 而多头注意力机制则精细分析近期的日内波动和突发事件, 使得该模型在电荷负载预测中表现出色, 兼顾宏观和微观视角。

GGSAGCformer 结构流程图如图 4 所示。模型由卷积层、注意力机制层和输出层构成。卷积层嵌入 GCN 提取区域空间特征, GRN 提取区域时间特征。还嵌入了多头注意力机制层, 引入了 2 种自注意力机制, 并通过全连接层将其融合, 识别长期动态时空相关性。最后, 通过输出层使用 GCformer 进行预测。

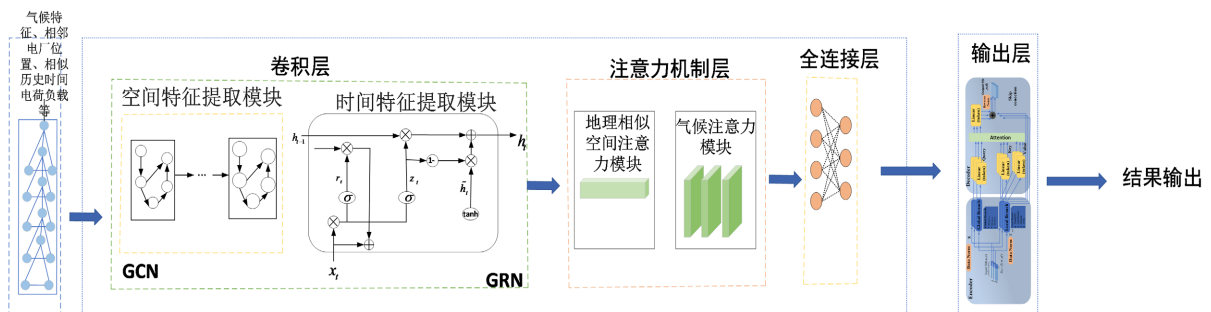


Figure 4. Structure flow of GGSAGCformer  
图 4. GGSAGCformer 结构流程



## 5. 实验设计

### 5.1. 数据集描述

本研究使用了 Trentino 区域多维城市生活数据集[22], 包含了 SET 电力和气候数据。为了便于管理 SET 能源数据, 该数据库将 Trentino 划分为多个网络, 并统计和存储每个网络中的客户数量和相应的配电线路数量。由于隐私要求, 客户类型隐藏在数据库中。为了提高模型的灵活性, 本研究随机选择了 Trentino 的特定区域, 并确定了为该区域提供服务的 10 条配电线路及其配送量和服务范围, 即 10 个对应区域。同时, 气候数据集包括该地区 36 个气象站的气温、降水量和风速/风向数据。该数据每 15 分钟收集一次, 没有空间聚合。

为应对训练数据与测试数据之间可能的分布偏移, 本文采用了数据归一化技术 RevIN 来增强模型的鲁棒性。具体操作如下: 首先计算输入数据每个实例  $x_k^{(i)} \in \mathbb{R}^T$  的平均值和标准差, 以此进行数据标准化处理。公式如下:

$$E_t[x_k^{(i)}] = \frac{1}{T} \sum_{j=1}^T x_{kj}^{(i)} \quad (27)$$

$$\text{Var}[x_{kj}^{(i)}] = \frac{1}{T} \sum_{j=1}^T (x_{kj}^{(i)} - E_t[x_k^{(i)}])^2 \quad (28)$$

利用这些统计数据, 我们通过应用可学习的仿射参数向量  $\gamma$  和  $\beta$ , 维度为  $\mathbb{R}^K$ , 对输入数据执行归一化处理公式为:

$$\hat{x}_{kt}^{(i)} = \gamma k \left( \frac{\hat{x}_{kt}^{(i)} - E_t[x_k^{(i)}]}{\sqrt{\text{Var}[x_k^{(i)}] + \varepsilon}} \right) \quad (29)$$

归一化后的数据输入到预测分析模型中, 模型输出后进行反归一化, 恢复到原始数值比例。数据分为训练集、验证集和测试集, 比例分别为 70%、10% 和 20%, 以提高模型的训练效率和泛化能力。

实验使用学习率为 0.001 的 ADAM 优化器。在本次实验中, 每次不同的输出步长都重复实验五次, 并将指标的平均值报告为最终结果。实验使用 NVIDIA GeForce RTX 3050 Laptop GPU, 使用 Pytorch 1.13.1 和 Python 3.9.16 实验环境。

### 5.2. 评价指标

本文选择均方误差(Mean Squared Error, MSE)和平均绝对误差(Mean Absolute Error, MAE)作为评价指标, 来评估所提出的电荷负载预测模型的性能。MSE 用于计算预测值与实际值的平均平方差, 它是进行回归分析时经常使用的一种评估指标。MAE 是预测值与实际值的平均绝对差异。公式为:

$$\text{MSE} = \frac{1}{T} \sum_{t=1}^T (\rho_{t+r} - \hat{\rho}_{t+r})^2 \quad (30)$$

$$\text{MAE} = \frac{1}{T} \sum_{t=1}^T |\rho_{t+r} - \hat{\rho}_{t+r}| \quad (31)$$

### 5.3. 超参数优化对比

在构建负荷预测模型的过程中, 超参数的选择不仅影响着模型的学习能力, 而且对泛化能力和预测精度起着决定性作用。本模型集中关注以下三个核心超参数:

1) 学习率设置为 0.01、0.001 和 0.0001。选择合适的学习率至关重要, 因为过高的学习率可能使模型在潜在最优解周围震荡而难以精确收敛; 而过低的学习率可能使训练速度极慢, 甚至停留在局部最优解。

2) 批次大小考虑为 6、64 和 128。批次大小影响模型的内存需求、训练速度及性能。小批次可以使模型更频繁地更新, 有助于精确模拟数据分布, 但可能增加计算开销和训练的不稳定性。相比之下, 大批次可以稳定训练过程, 但可能减少模型对数据细微变化的敏感性。

在本研究中, 超参数优化通过系统化的实验设计进行, 利用交叉验证方法在多种参数组合中进行评估。交叉验证结果如表 1 和表 2 所示。其中, 表 1 的对比批次大小设置为 32, 表 2 的对比参数学习率设置为 0.001。通过结果可以得知, 在所有测试组合中, 当学习率为 0.001, 批次大小为 32, 训练次数 100 时, 模型表现最佳, 实现了预测精度和计算效率的最佳平衡。

**Table 1.** Comparison of MAE and MSE between multiple models under different learning rates

**表 1.** 多种模型在不同学习率下 MAE 和 MSE 的对比

学习率	GGSAGCformer		GCformer		Informer		Reformer	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
0.01	0.141	0.266	0.162	0.278	0.171	0.301	0.182	0.331
0.001	0.111	0.234	0.131	0.252	0.148	0.288	0.157	0.302
0.000 1	0.116	0.238	0.136	0.255	0.149	0.292	0.159	0.307

**Table 2.** Comparison of MAE and MSE between multiple models in different batches

**表 2.** 多种模型在不同批次下 MAE 和 MSE 的对比

批次大小	GGSAGCformer		GCformer		Informer		Reformer	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
6	0.120	0.243	0.144	0.279	0.154	0.303	0.164	0.333
32	0.111	0.234	0.131	0.252	0.148	0.288	0.157	0.302
64	0.109	0.232	0.129	0.250	0.146	0.287	0.155	0.301

#### 5.4. 对比实验

本文还考虑了模型在不同区域间的同一时间段的预测精度。由于模型输出的结果包含了十个区域, 数据过多, 逐个分析太过于费时费力, 且工作量过于庞大和重复, 所以只选择其中两个区域的数据进行分析展示结果。在 SET 数据集中选取了 4 个有代表性的时间段进行实验, 来评估每个模型的预测性能, 这 4 个时间段分别为工作日的高峰用电时间段(2013 年 11 月 13 日星期三的 6:00~9:00, 18:00~22:00)以及休息日对应的时段(2013 年 12 月 15 日星期天的 6:00~9:00, 18:00~22:00)。实验结果如图 5~8 所示, 可以观察到, 在工作日期间, 所有模型基本都能拟合电荷预测的变化趋势。然而, 在休息日期间, 各模型的预测性能相对于工作日存在显著差异。这可能是因为周末电荷负载会受到更多复杂因素的影响, 如放假长时间休息、亲朋好友拜访等这些因素会导致电荷负载的突然增减, 从而表现出更加复杂和不稳定的特性。

从图 5~8 中可以看出, GGSAGCforme 表现出更好的预测效果, 是因为在嵌入层中加入了上周同一时间段的数据, 让模型能够周期性地学习这种不平稳数据的时空特征, 从而使模型能够更准确地预测电荷负载的动态变化, 因此表现得更加稳定。区域 1 和区域 2 在工作日和休息日的用电需求峰值相差较为明显, 在一定程度上说明了区域 1 和区域 2 之间用电需求的时空相关性较强。

为了验证所提模型的准确性和稳定性, 将 GGSAGCformer 与 GCformer、Informer、Reformer 在不同预测步长(96, 192, 336, 720)下的预测性能进行比较。实验结果如表 3 所示, 为了衡量预测性能, 选取了 MSE 和 MAE 作为评价指标。从表 3 中的数据可以看出, 随着预测步长的增加, 所有模型的实验误差均在增加, 预测准确性普遍降低, 有些模型甚至出现大幅波动的情况, 这表明这些模型不适合做长期预测。而 GGSAGCformer 在预测步长为 336 时, MSE 指 GGSAGCformer 的 MSE 在输出步长为 336 时, 相较于其他模型在预测步长为 336 时, MSE 指标比其他模型分别降低 17.1%、24.8%、27.1%, 所以 GGSAGCformer 表现出了更好的准确性。

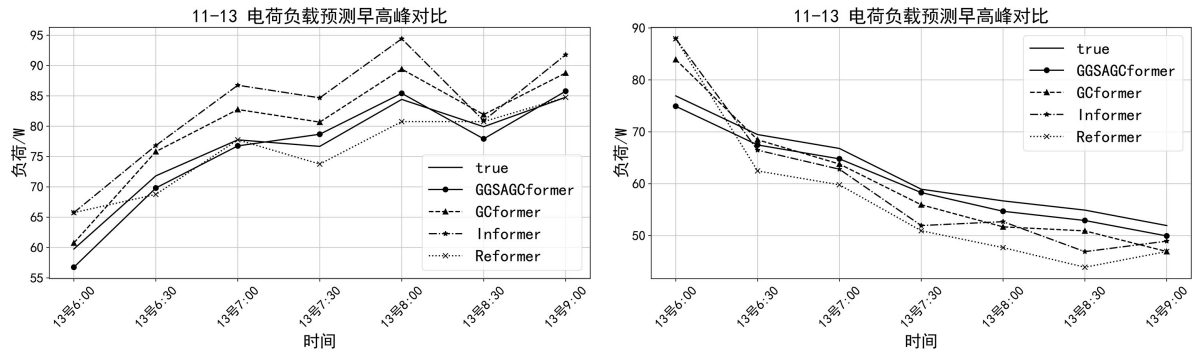


Figure 5. Comparison of the morning peak of charge prediction for regions 1 and 2 on November 13  
图 5. 11 月 13 日区域 1 和区域 2 的电荷预测早高峰对比

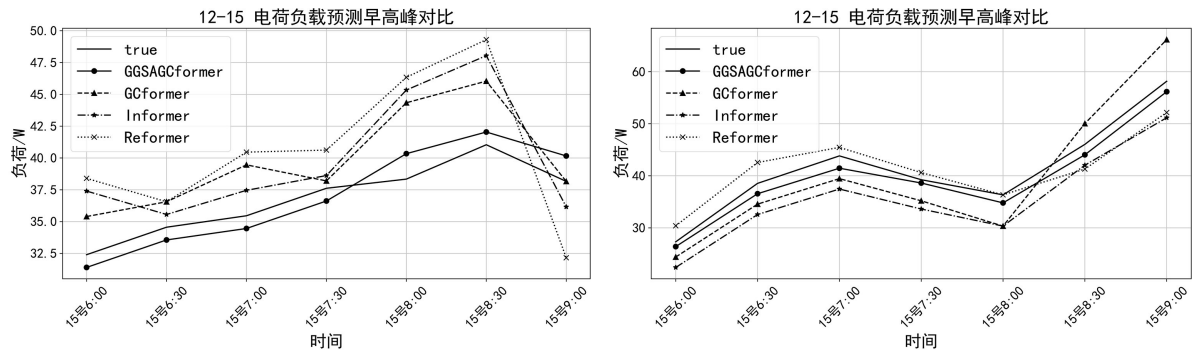


Figure 6. Comparison of the morning peak of charge prediction for regions 1 and 2 on December 15  
图 6. 12 月 15 日区域 1 和区域 2 的电荷预测早高峰对比

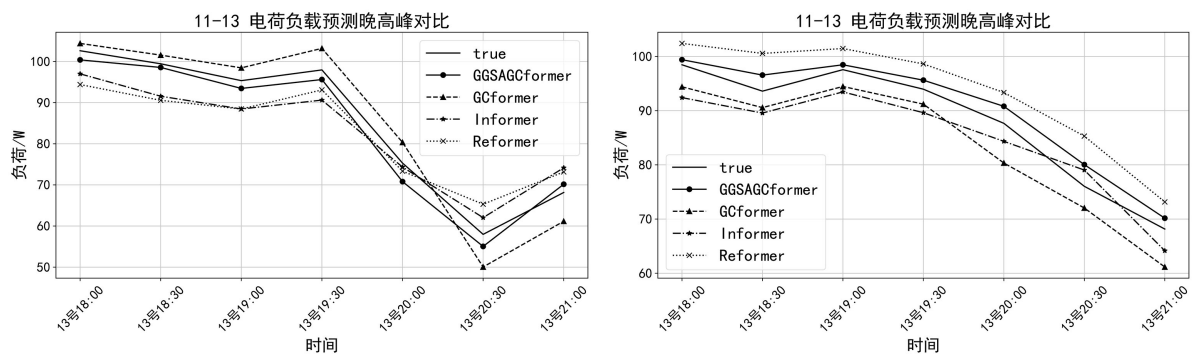


Figure 7. Comparison of the evening peak of charge prediction for regions 1 and 2 on November 13  
图 7. 11 月 13 日区域 1 和区域 2 的电荷预测晚高峰对比

为了深入对比各个模型的预测精度, 我们采用雷达图展示它们在同一输出节点下的预测结果。雷达图将每个模型的预测结果映射到不同的区域, 形成对应的多边形。在雷达图中, 每个区域代表一个坐标轴, 刻度从中心向外延伸, 数值越小越靠近中心, 代表预测效果越好。通过观察图 9、图 10 发现, GGSAGCformer 模型所构成的多边形范围相对较小, 显示出在每个节点预测结果方面具有更好的性能。综合考虑模型整体预测精度和各节点的表现, 我们的研究表明, GGSAGCformer 模型在提高电力系统短期负荷预测精度方面表现最佳。

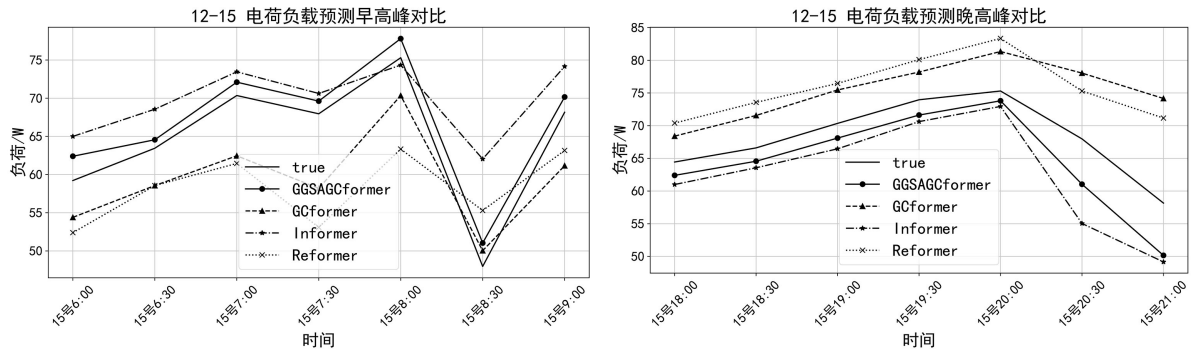


Figure 8. Comparison of the evening peak of charge prediction in regions 1 and 2 on December 15  
图 8. 12 月 15 日区域 1 和区域 2 的电荷预测晚高峰对比

Table 3. Comparison of models at different output steps  
表 3. 模型在不同输出步长下的对比

模型	GGSAGCformer		GCformer		Informer		Reformer	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
96	0.095	0.217	0.112	0.234	0.132	0.264	0.143	0.287
192	0.111	0.234	0.131	0.252	0.148	0.288	0.157	0.302
336	0.121	0.244	0.146	0.266	0.161	0.300	0.166	0.311
720	0.197	0.298	0.211	0.297	0.241	0.392	0.263	0.399

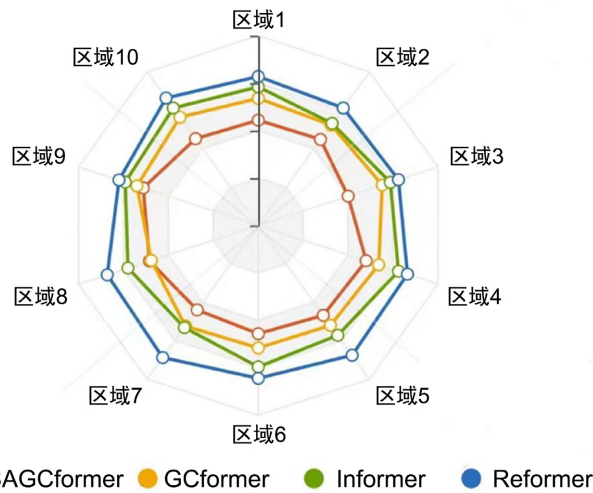


Figure 9. MSE comparison  
图 9. MSE 对比

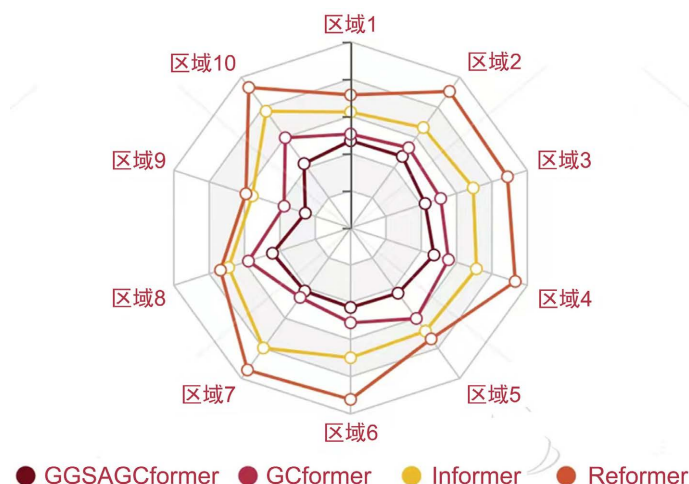


Figure 10. MAE comparison  
图 10. MAE 对比

## 6. 结论

在本文中, 本文设计了一个创新的多区域预测模型, 称为 GGSAGCformer, 它使用 GCN 提取多区域空间特征, GRN 提取多区域时间特征, 掌握每个区域节点的时空属性, 此外, 通过在注意力机制层中加入地理相似空间自注意力模块, 能够有效地捕捉邻近区域的属性及其时间滞后效应。这种机制能够利用每个节点的短期历史负荷数据来准确判断负荷传播的时间延迟, 从而增强对负荷变动的敏感度和预测精度。注意力层还加入了一个气候特征注意力模块, 用以捕捉气候变化对电荷负载的影响, 进一步提升多区域电荷负载预测的准确性。与传统的电荷负载预测模型如 GCformer、Informer 和 Reformer 相比, 本模型更有效地拟合了电荷负载的多种变化趋势。在未来的工作中, 本文计划继续优化模型的计算效率, 并研究其他因素对电荷负载的影响, 以及探索发掘动态时空相关性的其他方法。

## 参考文献

- [1] 闫群民, 邱绎同, 杨浩, 等. 基于大数据的电力系统短期负荷预测[J]. 电工技术, 2020(5): 16-18.
- [2] 陶娟, 邹红波, 周冬. 基于提升人工神经网络的短期负荷预测模型[J]. 电工材料, 2021(2): 53-56.
- [3] 兰华, 廖志民, 赵阳. 基于 ARMA 模型的光伏电站出力预测[J]. 电测与仪表, 2011, 48(2): 31-35.
- [4] 赵从杰, 潘文林. 基于深度学习的电网短期负荷预测[J]. 计算机与数字工程, 2024, 52(1): 219-222.
- [5] 徐良德, 郭挺, 雷才嘉, 等. 基于支持向量机的网格化电网负荷预测算法设计[J]. 电子设计工程, 2024, 32(3): 12-16.
- [6] 鲍伟强, 陈娟, 谢伟, 等. 一种基于自适应神经模糊推理系统的短期负荷预测方法[J]. 电气开关, 2019, 57(1): 8-11.
- [7] 陈江艳. 基于模糊推理算法的电力负荷预测系统的研究[J]. 电气应用, 2015, 34(13): 186-189.
- [8] 陈胜, 刘鹏飞, 王平, 等. 基于 LSTM 人工神经网络的电力系统负荷预测方法[J]. 沈阳工业大学学报, 2024, 46(1): 66-71.
- [9] 邹量. 基于深度学习的电力系统短期负荷预测研究与系统实现[D]: [硕士学位论文]. 南昌: 南昌大学, 2023.
- [10] 章家义, 龚圣辉, 聂堃. 基于 GRU 门控单元网络的电力负荷预测研究[J]. 粘接, 2024, 51(4): 145-148.
- [11] 徐虎, 刘新润, 周宣, 等. 基于 CNN 集成和非均匀量化的家庭负荷预测[J]. 湘潭大学学报(自然科学版), 2024, 46(1): 70-82.
- [12] 李诚. LSTM 与 CNN 相结合的多通道短期电力负荷预测研究[D]: [硕士学位论文]. 福州: 福建工程学院, 2023.
- [13] 陈柘, 刘嘉华, 赵斌, 等. 基于 GCN 和 TCN 的多因素城市路网出租车需求预测[J]. 控制与决策, 2023, 38(4):

- 1031-1038.
- [14] 孙林, 梁娜, 徐久成. 基于邻域互信息与 K-Means 特征聚类的特征选择[J]. 智能系统学报, 2024, 19(4): 983-996.
  - [15] 王萌. 基于 Transformer 的短期负荷预测方法研究[D]: [硕士学位论文]. 长沙: 湖南大学, 2022.
  - [16] Zhou, T., Ma, Z., Wen, Q., *et al.* (2022) Fedformer: Frequency Enhanced Decomposed Transformer for Long-Term Series Forecasting. arXiv: 2201.12740.
  - [17] Zhao, Y.J., Ma, Z., Zhou, T., *et al.* (2023) GCformer: An Efficient Framework for Accurate and Scalable Long-Term Multivariate Time Series Forecasting. *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, Birmingham, 21-25 October 2023, 3464-3473. <https://doi.org/10.1145/3583780.3615136>
  - [18] Li, Y., Cai, T., Zhang, Y., *et al.* (2022) What Makes Convolutional Models Great on Long Sequence Modeling? arXiv: 2210.09298. <https://arxiv.org/abs/2210.09298>
  - [19] Gu, A., Goel, K. and Ré, C. (2021) Efficiently Modeling Long Sequences with Structured State Spaces. arXiv: 2111.00396. <https://arxiv.org/abs/2111.00396>
  - [20] Yi, K., Zhang, Q., Fan, W., *et al.* (2024) Frequency-Domain MLPs Are More Effective Learners in Time Series Forecasting. arXiv: 2311.06184.
  - [21] Gu, A., Goel, K. and Ré, C. (2021) Efficiently Modeling Long Sequences with Structured State Spaces. arXiv: 2111.00396. <https://arxiv.org/abs/2111.00396>
  - [22] Barlacchi, G., De Nadai, M., Larcher, R., *et al.* (2015) A Multi-Source Dataset of Urban Life in the City of Milan and the Province of Trentino. *Scientific Data*, 2, Article No. 150055. <https://doi.org/10.1038/sdata.2015.55>