

# 基于改进 Autoformer 模型的短期电力负荷预测

范杏蕊, 李元诚

(华北电力大学 控制与计算机工程学院, 北京 102206)

**摘要:**针对短期电力负荷预测因受天气、温度、节假日等多重不确定性因素影响而造成精度低的问题,提出一种基于改进 Autoformer 模型的短期电力负荷预测模型。改变序列分解预处理的惯例,设计深度模型的内部分解模块,该模块提取模型中隐藏状态的内在复杂时序趋势,使得模型具有复杂时间序列的渐进分解能力;提出 Nystrom 自注意力机制,该机制利用 Nystrom 方法来逼近标准的自注意力机制。某地电力负荷预测实验结果表明,所提模型比基于标准 Autoformer 模型的短期电力负荷预测模型的时间复杂度更低,准确率更高。

**关键词:**短期电力负荷预测;时序分解模块;Nystrom 自注意力机制;Sdformer 模型

**中图分类号:**TM715

**文献标志码:**A

**DOI:**10.16081/j.epae.202305011

## 0 引言

准确的电力负荷预测可以提高电网调度水平,对整个电力系统的安全稳定运行有着至关重要的作用<sup>[1]</sup>。对短期电力负荷进行精准而高效的预测具有重要意义和实用价值<sup>[2]</sup>。

目前,常用的短期电力负荷预测方法主要包括统计学方法、机器学习方法和深度学习方法三大类。统计学方法主要包括广义加性模型<sup>[3]</sup>、卡尔曼滤波<sup>[4]</sup>、指数平滑法<sup>[5]</sup>等,该类方法虽然结构简单,训练速度快,但不能反映电力负荷序列的非线性特征。机器学习方法包括人工神经网络<sup>[6]</sup>、支持向量机<sup>[7]</sup>、随机森林<sup>[8]</sup>、小波神经网络<sup>[9]</sup>等,该类方法虽然在训练过程中可以提取输入、输出间的非线性关系,但对复杂电力系统数据特征的挖掘能力略显不足。随着计算机性能的提高,深度学习方法因具有高效数据表征和复杂非线性关系建模的能力而被广泛应用于电力负荷预测领域,该类方法主要包括循环神经网络(recurrent neural network, RNN)、卷积神经网络(convolutional neural network, CNN)、门控循环单元(gated recurrent unit, GRU)和长短期记忆(long short-term memory, LSTM)网络。文献<sup>[10]</sup>提出一种基于 CNN 以及 GRU 的混合神经网络模型,该模型在保持较快训练速度的同时,具有较高的预测精度。文献<sup>[11]</sup>提出一种基于特征筛选的 CNN 和双向长短期记忆(bi-directional long short-term memory, BiLSTM)网络组合模型的短期电力负荷预测方法,根据该方法构建的组合模型可提升多维电力负荷数据的短期预测精度。文献<sup>[12]</sup>引入注意力(Attention)机制,提出一种基于 CNN-BiLSTM-Attention 的新型超短期电力负荷预测方法,该方法能够有效减少历史信息的丢失以及突出关键历史时间点的信息,进

一步提高了短期电力负荷预测的精度。文献<sup>[13]</sup>提出基于 Attention-LSTM 网络的短期电力负荷预测模型,充分利用电力负荷的时序特征并采用 Attention 机制突出对电力负荷预测起关键作用的输入特征,该模型在预测精度和算法鲁棒性方面均优于 LSTM 网络、CNN 和支持向量机模型。

Transformer 模型最早是由 Google 团队提出的,主要用于完成机器翻译任务<sup>[14]</sup>,该模型抛弃了 RNN、CNN 等网络结构,因具有支持并行、训练速度快等优点而被广泛应用于电力负荷预测领域。文献<sup>[15]</sup>提出一种基于循环扩张机制的卷积门控循环单元-Transformer 短期电力负荷预测方法,该方法在大幅缩短训练时间的同时有效提高了预测精度。文献<sup>[16]</sup>提出一种改进 Transformer 的电力负荷预测方法,该方法将特征提取技术和深度学习相结合,提升了在进行多步预测时的预测性能。文献<sup>[17]</sup>提出一种融合极限梯度提升和改进 Transformer 模型的新型电力负荷预测方法,该方法考虑时间周期和气象因素对电力负荷的影响,显著提高了预测准确率和效率。

Transformer 模型中的解码器和编码器均采用基于自注意力(self-Attention)机制的网络,这使该模型存在二次时间复杂度和高内存利用率的问题。Autoformer 模型以自相关机制代替点向连接的 Attention 机制,实现了序列级连接和  $O(L\log_2 L)$  的时间复杂度<sup>[18]</sup>,其中  $L$  为序列长度。为进一步降低时间复杂度以及考虑影响电力负荷预测的气象因素、日期因素等外部因素,本文提出一种基于改进 Autoformer 模型,即 Sdformer 模型的短期电力负荷预测模型。首先,利用 Sdformer 模型内部分解模块将历史负荷分解为趋势性分量、季节性分量、节假日分量和随机分量;其次,利用 Sdformer 模型对各子序列进行建模,并通过加权融合将各子序列一起进行训练;然

后,对比分析 Sdformer 模型中的 Nystrom self-Attention 机制与 Autoformer 模型中的自相关机制;最后,基于澳大利亚某地区的电力负荷数据集进行实验,结果表明,相较于标准 Autoformer 模型,本文 Sdformer 模型的时间复杂度更低,训练速度更快,预测准确率更高。

## 1 电力负荷分析

### 1.1 电力负荷特征

电力负荷包含如下特征:周期性,电力负荷具有波动性,在波动的过程中表现出连续性的特征;随机性,电能是瞬时变化的,电力负荷也会表现出随机性的特征;条件性,电力负荷在不同条件下的变化具有不同的规律和特点。

### 1.2 短期电力负荷预测的影响因素

短期电力负荷预测受多种因素的影响,主要影响因素如下。

#### 1) 气象因素。

影响短期电力负荷预测的气象因素包括温度、湿度、风速等,其中温度对短期电力负荷的影响较大。电力负荷和温度随时间的变化如附录 A 图 A1 所示。由图可知,随着温度的变化,电力负荷需求在夏季达到高峰,而在春秋季节较低。因此,本文将温度作为短期电力负荷预测的影响因素之一。

#### 2) 日期因素。

短期电力负荷预测会受节假日、工作日以及不同月份等日期因素的影响。图 1 为某地区 6 月一周的电力负荷变化。图 2 为节假日对负荷的影响。相同月份中的日负荷变化趋势相似,而不同月份的日负荷变化趋势存在一定的差别。相较于工作日,节假日的负荷需求显著下降。

#### 3) 随机因素。

电力系统中不同用户会有不同的用电时间和用电方式,这使得电力负荷具有不确定性和随机性。另外,一些特殊事件,如台风、地震等自然灾害,也会对电力负荷产生较大影响。

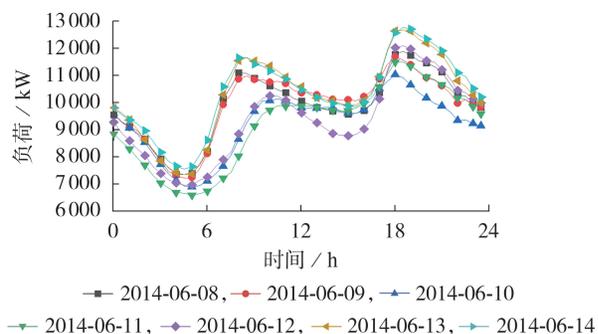


图1 某地区一周的电力负荷变化

Fig.1 Weekly power load change in a certain area

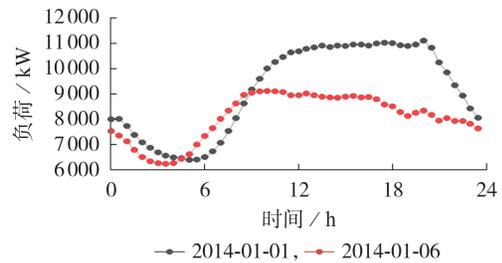


图2 节假日对电力负荷的影响

Fig.2 Influence of holiday on power load

## 2 基于 Sdformer 模型的短期电力负荷预测

### 2.1 Transformer 模型及其存在的问题

Transformer 模型由编码器和解码器两部分组成,模型结构如附录 A 图 A2 所示。

#### 1) 编码器。

编码器包含 2 个子层:一个是 multi-head Attention 层,该子层利用 self-Attention 学习序列内部的关系;另一个是 feed forward 层,该子层为简单的全连接网络,分别对每个位置向量进行相同的操作,feed forward 层包括 2 个线性变换和 1 个 ReLU 激活函数,将序列经过 feed forward 层产生的编码器输出传递给解码器。编码过程采用并行计算方式,极大地提高了效率。

#### 2) 解码器。

解码器包含 3 个子层,其中 1 个子层为多头 Attention 机制层,1 个子层为多头掩码 Attention 机制层。多头掩码 Attention 机制层先利用 self-Attention 机制学习目标句内部的关系,再将该层输出与接收的编码器输出一并输入多头 Attention 机制层,以学习序列之间的关系。经过多头 Attention 机制层的序列在经过一个与编码部分类似的负反馈层后,得到解码器的输出。

然而,由于 Transformer 模型的编码器和解码器均采用 Attention 机制,时间复杂度高,内存利用率随着输入长度的增加呈二次增长,这导致 Transformer 模型在应用于短期电力负荷预测时会出现内存不足的问题。

### 2.2 基于 Sdformer 模型的短期电力负荷预测模型

为了解决 Transformer 模型所存在的问题,结合 Nystromformer 中的 Nystrom self-Attention 机制所具有的线性优点,本文提出 Sdformer 模型,该模型具有深度分解架构,能够从复杂时间模式中分解出可预测性更强的成分,并且 Nystrom self-Attention 机制采用近似方式逼近标准的 self-Attention 机制,其时间复杂度和内存利用率均为  $O(L)$ 。本文提出的 Sdformer 模型结构如附录 A 图 A3 所示,主要包括输入层、时序分解层、编码器层、解码器层和输出层。将历史数据作为输入,通过位置编码对数据进行编

码;时序分解层将历史负荷分解为趋势性分量、季节性分量、节假日分量和随机分量;编码器层消除趋势性分量;解码器层对季节性分量、节假日分量和随机分量进行建模;在经过一步预测后得到预测结果,避免了在多步预测过程中造成的误差累积。

### 2.2.1 内部分解模块

为了在预测过程中学习复杂的时间关系,内部分解模块基于滑动平均和傅里叶变换思想将历史负荷分解为趋势性分量、季节性分量、节假日分量和随机分量,如式(1)所示。

$$\mathbf{y}(t) = \mathbf{g}(t) + \mathbf{s}(t) + \mathbf{h}(t) + \boldsymbol{\varepsilon}(t) \quad (1)$$

式中: $\mathbf{y}(t)$ 为时间 $t$ 的历史负荷; $\mathbf{g}(t)$ 为时间 $t$ 的趋势性分量; $\mathbf{s}(t)$ 为时间 $t$ 的季节性分量; $\mathbf{h}(t)$ 为时间 $t$ 的节假日分量; $\boldsymbol{\varepsilon}(t)$ 为时间 $t$ 的误差。 $\mathbf{g}(t)$ 、 $\mathbf{s}(t)$ 、 $\mathbf{h}(t)$ 的计算公式为:

$$\begin{cases} \mathbf{g}(t) = \text{AvgPool}(\text{Padding}(\mathbf{x})) \\ \mathbf{s}(t) = \sum_{i=1}^N \left( a_i \cos \frac{2\pi i t}{p} + b_i \sin \frac{2\pi i t}{p} \right) \\ \mathbf{z}(t) = [1, 1, \dots, 1] \quad t \in D_m \\ \mathbf{h}(t) = \mathbf{z}(t)R \end{cases} \quad (2)$$

式中:AvgPool( $\cdot$ )表示平均滤波卷积操作;Padding( $\cdot$ )为在特征图两侧的零填充宽度大小; $\mathbf{x}$ 为序列值; $a_i$ 、 $b_i$ 为系数; $N$ 为级数,一般取 $N=10$ ;  $p$ 为负荷序列的年度数据,一般取 $p=365.25$ ;  $\mathbf{z}(t)$ 为生成回归量矩阵; $D_m(m=1, 2, \dots, n)$ 为第 $m$ 个假期所有日期的集合, $n$ 为一整数; $R$ 为每个假期分配的参数。

### 2.2.2 Nystrom self-Attention 机制

Nystrom self-Attention 机制利用 Nystrom 方法近似标准 self-Attention 机制中的 Softmax 矩阵,本文以 Nystrom self-Attention 机制替代 Autoformer 网络中的自相关机制。

self-Attention 机制中的 Softmax 矩阵  $S$  的计算公式为:

$$\mathbf{S} = \text{Softmax} \left( \frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}} \right) \mathbf{V} \quad (3)$$

式中:Softmax( $\cdot$ )为按列进行归一化函数; $\mathbf{Q}$ 为查询向量; $\mathbf{K}$ 为键向量; $\mathbf{V}$ 为值向量; $d_k$ 为键向量的维度。

利用 Nystrom 方法中的正交技术对  $S$  进行近似,如式(4)所示。

$$\mathbf{S} \approx \text{Softmax} \left( \frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_q}} \right) = \begin{bmatrix} \mathbf{A}_s & \mathbf{B}_s \\ \mathbf{F}_s & \mathbf{C}_s \end{bmatrix} \quad (4)$$

式中: $\mathbf{A}_s \in \mathbf{R}^{r \times r}$ ,  $\mathbf{B}_s \in \mathbf{R}^{r \times (b-r)}$ ,  $\mathbf{F}_s \in \mathbf{R}^{(b-r) \times r}$ ,  $\mathbf{C}_s \in \mathbf{R}^{(b-r) \times (b-r)}$ ,  $\mathbf{A}_s$ 、 $\mathbf{B}_s$ 、 $\mathbf{F}_s$ 、 $\mathbf{C}_s$ 为通过对矩阵  $S$  采样得到的子矩阵, $r$ 为采样矩阵的列数, $b$ 为矩阵  $\mathbf{Q}$  的行数; $d_q$ 为模型维度。通过 Nystrom 分解得到的矩阵  $\widehat{S}$  的计算公式如式(5)

所示。

$$\widehat{S} = \begin{bmatrix} \mathbf{A}_s \\ \mathbf{F}_s \end{bmatrix} \mathbf{A}_s^+ \begin{bmatrix} \mathbf{A}_s & \mathbf{B}_s \end{bmatrix} = \mathbf{F}\mathbf{A}_s^+ \mathbf{B} \quad (5)$$

式中: $\mathbf{A}_s^+$ 为  $\mathbf{A}_s$  的穆尔-彭罗斯广义逆, $\mathbf{F}$ 、 $\mathbf{B}$ 为近似矩阵, $\mathbf{A}_s^+$ 、 $\mathbf{F}$ 、 $\mathbf{B}$ 的计算公式见文献[17]。最终,矩阵  $S$  的计算公式为:

$$\mathbf{S} \approx \text{Softmax} \left( \frac{\mathbf{Q}\tilde{\mathbf{K}}^T}{\sqrt{d_q}} \right) \text{Softmax} \left( \frac{\tilde{\mathbf{Q}}\tilde{\mathbf{K}}^T}{\sqrt{d_q}} \right)^+ \text{Softmax} \left( \frac{\tilde{\mathbf{Q}}\mathbf{K}^T}{\sqrt{d_q}} \right) \quad (6)$$

式中: $\tilde{\mathbf{Q}}$ 、 $\tilde{\mathbf{K}}$ 为矩阵  $\mathbf{A}$  通过平均池化获得的矩阵。

### 2.2.3 编码器

编码器由分解模块和前馈神经网络组成,如附录 A 图 A3 所示。编码器的输入包含季节性分量  $\mathbf{X}_{\text{ens}}$ 、趋势性分量  $\mathbf{X}_{\text{ent}}$ 、节假日分量  $\mathbf{X}_{\text{enh}}$  和随机性分量  $\mathbf{X}_{\text{enr}}$ ,如式(7)所示。

$$\begin{cases} \mathbf{X}_{\text{ens}}, \mathbf{X}_{\text{ent}}, \mathbf{X}_{\text{enh}}, \mathbf{X}_{\text{enr}} = \text{SeriesDecomp}(\mathbf{X}_{\text{en}}) \\ \mathbf{X}_{\text{des}} = \text{Concat}(\mathbf{X}_{\text{ens}}, \mathbf{X}_0) \\ \mathbf{X}_{\text{det}} = \text{Concat}(\mathbf{X}_{\text{ent}}, \mathbf{X}_0) \\ \mathbf{X}_{\text{deh}} = \text{Concat}(\mathbf{X}_{\text{enh}}, \mathbf{X}_0) \\ \mathbf{X}_{\text{der}} = \text{Concat}(\mathbf{X}_{\text{enr}}, \mathbf{X}_0) \end{cases} \quad (7)$$

式中: $\mathbf{X}_{\text{ens}}$ 、 $\mathbf{X}_{\text{ent}}$ 、 $\mathbf{X}_{\text{enh}}$ 、 $\mathbf{X}_{\text{enr}}$ 为从序列  $\mathbf{X}_{\text{en}}$  中分解的子序列;SeriesDecomp( $\cdot$ )为分解函数;Concat( $\cdot$ )为拼接函数; $\mathbf{X}_0$ 表示用0填充序列  $\mathbf{X}_{\text{en}}$  的填充符。

编码器侧重于对季节性分量的建模,其输出包含过去的季节性信息,将其作为交叉信息,以帮助解码器细化预测结果。编码器层的方程如附录 B 式(B1)所示。

### 2.2.4 解码器

解码器包含趋势性分量的累积结构和季节性分量的堆叠自相关机制两部分。每个解码器层均包含 Nystrom self-Attention 机制和分解模块,模型在解码期间从中间隐藏变量中提取潜在趋势,逐步细化趋势预测并消除 Nystrom self-Attention 机制中基于周期依赖关系发现的干扰信息。解码器的方程如附录 B 式(B2)所示。

## 2.3 基于 Sdformer 模型的短期电力负荷预测流程

基于 Sdformer 模型的短期电力负荷预测流程如下。

1)对电力负荷数据集进行预处理,将电力负荷数据按 3:1:1 进行划分,分别构成训练集、验证集和测试集,将预处理后的数据输入模型中。

2)设置 Sdformer 模型的超参数,构建和初始化 Sdformer 模型。

3)初始化模型的学习率,在训练集上训练 Sdformer 模型,计算模型预测结果误差。

4)更新模型的学习率,在训练集上进行多次迭

代(迭代次数大于 10)后终止迭代。

5)使用训练好的模型在测试集上进行预测,一步得到电力负荷预测结果,将模型的输出通过反正则化得到电力负荷序列最终的预测值。

### 3 实验分析

#### 3.1 实验数据集及预处理

本文将澳大利亚某地区 2006 年 1 月 1 日至 2010 年 12 月 31 日的电力负荷数据集的数据(每 30 min 进行 1 次采样,每日共 48 个采样点)作为实验数据。该数据集包括露点温度、干球温度、湿球温度、湿度、电价和电力负荷 6 个维度数据。由于本文不考虑电价对电力负荷预测的影响,而考虑日期的影响,对数据集进行预处理,标明工作日和休息日以及节假日和周,并重新构建数据集。重新构建后的数据集共包括 9 个维度特征参数,分别为露点温度、干球温度、湿球温度、湿度、季节、月份、周、节假日、电力负荷。特征参数如附录 B 表 B1 所示。

#### 3.2 数据预处理

为了便于模型网络的训练,采用 min-max 归一化方法将原始数据归一化至  $(-1, 1)$  区间,计算过程如附录 B 式(B3)所示。本文选取平均相对误差  $e_{MRE}$  和平均误差  $e_{ME}$  作为评价指标,表达式分别如附录 B 式(B4)、(B5)所示。将电力负荷数据按 3:1:1 进行划分,分别构成训练集、验证集和测试集,即将第 1—3 年的数据构成训练集,将第 4 年的数据构成验证集,将第 5 年的数据构成测试集。采用单步直接式的方法利用本文模型进行预测。与传统仅划分训练集和测试集的方法(记为传统方法)相比,本文划分验证集是为了调整模型的超参数,以获得模型最优预测结果。本文方法和传统方法的结果对比如表 1 所示。

表 1 模型训练方法的结果对比  
Table 1 Comparison of results between model training methods

方法	$e_{MRE}$	$e_{ME}$
本文方法	6.54	0.402
传统方法	6.95	0.434

#### 3.3 实验参数设置

Sdformer 模型中编码器的层数设置为 2,解码器的层数设置为 1,多头 Attention 机制的头数设置为 8,多头 Attention 机制的维度设置为 512,网络批量大小设置为 32。

Nystrom self-Attention 机制需要设置 num\_landmarks 参数  $V_{num\_landmarks}$  和卷积核尺寸参数  $V_{conv\_kernel\_size}$ ,其中  $V_{num\_landmarks}$  取值从  $\{6, 12, 24, 48\}$  中选择,  $V_{conv\_kernel\_size}$  取值为 35。  $V_{num\_landmarks}$  取值影响电力负荷预测的准

确率,  $V_{num\_landmarks}$  与  $e_{ME}$  的关系如图 3 所示,本文中  $V_{num\_landmarks}$  取值为 12。

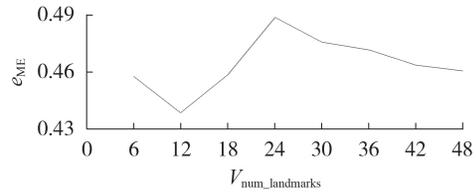


图 3  $V_{num\_landmarks}$  与  $e_{ME}$  的关系

Fig.3 Relationship between  $V_{num\_landmarks}$  and  $e_{ME}$

#### 3.4 特征参数对模型的影响

为了说明特征参数对负荷预测模型的影响,分别采用不同的特征参数组合作为模型的输入,负荷预测效果如表 2 所示。

表 2 不同特征参数的负荷预测效果

Table 2 Load forecasting effect of different characteristic parameters

特征参数	$e_{ME}$	$e_{MRE}$
露点温度+干球温度+湿球温度+湿度	0.440	6.163
干球温度+湿球温度+湿度	0.454	6.064
干球温度+湿度	0.472	6.620
湿度+湿球温度	0.463	6.349
干球温度	0.489	6.854
干球温度+湿球温度	0.479	6.689
气象因素+日期因素	0.438	6.200

由表 2 可以得出以下结论。

1)当仅用干球温度作为模型输入时,预测精度最低;在此基础上加上湿球温度作为模型输入时,预测模型的  $e_{ME}$ 、 $e_{MRE}$  分别减少了 2.04%、2.41%;在此基础上加上湿度作为模型输入时,相较于将干球温度和湿球温度作为模型输入,预测模型的  $e_{ME}$ 、 $e_{MRE}$  分别减少了 5.22%、9.34%。湿球温度高时,电力负荷高,湿球温度低时,电力负荷低,这导致温度对电力负荷预测的影响更明显。当将露点温度、干球温度、湿球温度和湿度作为模型输入时,相比仅将干球温度作为模型的输入,预测模型的  $e_{ME}$ 、 $e_{MRE}$  分别减少了 10.02%、10.08%,这表明湿度和温度综合影响电力负荷预测的精度。

2)当将日期因素和气象因素作为模型输入时,预测模型能够充分学习数据中的有效信息和隐含规律,预测效果最佳。

#### 3.5 Sdformer 模型的有效性验证

##### 3.5.1 Attention 机制对预测模型的影响

为了验证基于 Nystrom self-Attention 机制 Sdformer 模型的有效性,将该模型与基于 self-Attention 机制的 Sdformer+ 模型和不加 self-Attention 机制的 Noformer 模型进行对比,结果如表 3 所示。由表可以得出以下结论。

表3 使用不同 Attention 机制模型的结果  
Table 3 Results of models using different Attention mechanisms

模型	$e_{ME}$	$e_{MRE}$	训练时间 / s
Sdformer	0.452	6.386	38.146
Sdformer+	0.476	6.484	47.570
Noformer	0.541	6.718	30.456

1)相较于不加 self-Attention 机制的 Noformer 模型, Sdformer 和 Sdformer+ 模型的预测误差更小,  $e_{ME}$  分别减少了 16.45% 和 12.01%,  $e_{MRE}$  分别减少了 4.94% 和 3.48%, 这表明基于 Attention 机制的 Sdformer 模型可以显著提升模型的预测能力。

2)相较于 Sdformer+ 模型, Sdformer 模型的预测误差更小, 训练速度更快, 这表明本文基于 Nystrom self-Attention 机制的 Sdformer 模型可以显著提升模型的预测效果。

### 3.5.2 内部分解模块的有效性验证

为了验证内部分解模块的有效性, 对比原始负荷训练数据的 3 种处理方式: 方式 1, 对原始负荷训练数据不进行分解, 直接进行预测; 方式 2, 使用分解模块对原始负荷训练数据进行分解后再进行预测; 方式 3, 使用所提内部分解模块对原始负荷训练数据进行分解后再进行预测。对分解后的子序列进行预测, 将各序列预测结果进行叠加, 得到最终的负荷预测结果。3 种方式预测结果的评价指标如表 4 所示。

表4 3种方式预测结果的评价指标  
Table 4 Evaluation indicators of forecasting results with three dealing modes

方式	$e_{ME}$	$e_{MRE}$
1	0.514	6.598
2	0.485	6.489
3	0.449	6.354

由表4可以得出以下结论。

1)相较于方式 1, 方式 2 和方式 3 的预测误差更小,  $e_{ME}$  分别减少了 5.64%、12.65%,  $e_{MRE}$  分别减少了 1.65%、3.70%。可见, 对负荷训练数据进行分解可以提升预测精度。

2)相较于方式 2, 方式 3 的预测误差更小, 这表明利用内部分解模块对负荷训练数据进行分解更有利于提升预测精度。

### 3.6 不同预测模型的对比

为验证本文模型在短期电力负荷预测中的优势, 将本文模型同目前主流模型 LSTM、TCN、Reformer、Informer、标准 Autoformer 模型进行对比。各模型基于数据集的误差对比如附录 C 表 C1 所示, 由表可以得到以下结论。

1)相较于标准 Autoformer 模型, Sdformer 模型的  $e_{ME}$  在春季、夏季、秋季、冬季分别减少了 5.60%、6.13%、4.91%、6.30%, 这是由于 Sdformer 模型考虑特征序列内部的规律性, 充分发挥了模型内部分解模块处理时序序列的优势, 从而有效提高了模型的预测精度。此外, 由于夏季时气象因素与负荷序列间的相关性最强, 模型的预测精度最高。

2)在节假日的预测过程中, 相较于 Autoformer、Informer、LSTM、TCN 模型, Sdformer 模型的  $e_{ME}$  分别减少了 4.89%、21.02%、27.55%、25.62%,  $e_{MRE}$  分别减少了 14.62%、30.55%、54.42%、50.28%, 这表明 Sdformer 模型在节假日也表现出了更强的预测能力。

3)不同模型在 1 d、3 d、1 周的负荷预测结果如附录 C 图 C1 所示。由图 C1 可知本文所提模型的负荷预测精度更高。其中以春季 3 d 的测试结果为例, 相较于标准 Autoformer、Informer、LSTM、TCN 模型, Sdformer 模型的  $e_{ME}$  分别减少了 6.61%、23.39%、27.91%、23.2%,  $e_{MRE}$  分别减少了 14.93%、30.67%、54.33%、49.97%。

4)各种模型的预测效果图如图 4 所示。由图可知: 标准 Autoformer 模型的预测效果较好; Informer 模型是一种 Transformer 模型, 没有取得很好的预测效果; LSTM 模型的预测效果不佳, 不适合长序列的预测; Sdformer 模型取得了较好的预测效果。

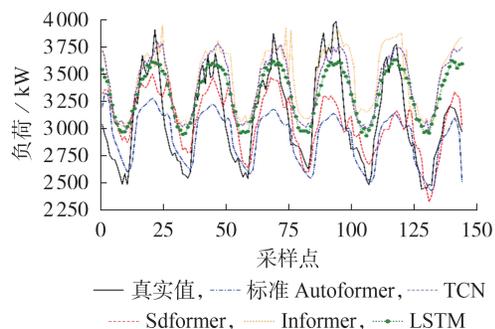


图4 各种模型的预测效果

Fig.4 Forecasting effect of various models

## 4 结论

针对短期电力负荷预测问题, 本文提出基于改进 Autoformer 模型的短期电力负荷预测模型, 即 Sdformer 模型。通过内部分解模块将负荷序列分解为趋势性分量、季节性分量、节假日分量和随机分量, 利用 Sdformer 模型对这 4 个子序列进行建模。基于澳大利亚某地区的电力负荷数据进行实验分析, 得到如下结论。

1)基于实验电力负荷数据, 重新构建包含日期因素、气候因素在内的多维特征集, 充分发挥了 Sdformer 模型在数据领域的优势。

2)根据负荷的时序特性, 结合气象因素、日期因

素和历史负荷序列,建立基于Sdformer模型的短期电力负荷预测模型。内部分解模块使得模型具有复杂时间序列的渐进分解能力。Nystrom self-Attention机制可以解决Transformer模型中二次时间复杂度和内存利用率高的问题。实验验证了本文模型可显著提高短期电力负荷预测精度。

附录见本刊网络版(<http://www.epae.cn>)。

## 参考文献:

- [1] 吴晨,姚菁,薛贵元,等. 基于MMoE多任务学习和长短时记忆网络的综合能源系统负荷预测[J]. 电力自动化设备,2022,42(7):33-39.  
WU Chen, YAO Jing, XUE Guiyuan, et al. Load forecasting of integrated energy system based on MMoE multi-task learning and LSTM[J]. Electric Power Automation Equipment, 2022,42(7):33-39.
- [2] 庞传军,刘金波,张波,等. 基于Shapley值的电力负荷预测结果溯源分析方法[J]. 电力自动化设备,2021,41(12):200-205.  
PANG Chuanjun, LIU Jinbo, ZHANG Bo, et al. Traceability analysis method of power load forecasting results based on Shapley value[J]. Electric Power Automation Equipment, 2021,41(12):200-205.
- [3] SONG K B, BAEK Y S, HONG D H, et al. Short-term load forecasting for the holidays using fuzzy linear regression method[J]. IEEE Transactions on Power Systems, 2005,20(1):96-101.
- [4] 肖白,高文瑞,李道明,等. 基于 $3\sigma$ -CEEMDAN-LSTM的空间负荷预测方法[J]. 电力自动化设备,2023,43(3):159-165.  
XIAO Bai, GAO Wenrui, LI Daoming, et al. Spatial load forecasting method based on  $3\sigma$ -CEEMDAN-LSTM[J]. Electric Power Automation Equipment, 2023,43(3):159-165.
- [5] 崔树银,汪昕杰. 基于最大信息系数和多目标Stacking集成学习的综合能源系统多元负荷预测[J]. 电力自动化设备,2022,42(5):32-39,81.  
CUI Shuyin, WANG Xinjie. Multivariate load forecasting in integrated energy system based on maximal information coefficient and multi-objective Stacking ensemble learning[J]. Electric Power Automation Equipment, 2022,42(5):32-39,81.
- [6] 张淑清,段晓宁,张立国,等. Tsne降维可视化分析及飞蛾火焰优化ELM算法在电力负荷预测中应用[J]. 中国电机工程学报,2021,41(9):3120-3130.  
ZHANG Shuqing, DUAN Xiaoning, ZHANG Liguang, et al. Tsne dimension reduction visualization analysis and moth flame optimized ELM algorithm applied in power load forecasting[J]. Proceedings of the CSEE, 2021,41(9):3120-3130.
- [7] 张明辉,周亚同,孔晓然. 基于深度混合储备池计算模型的短期电力负荷预测[J]. 电网技术,2022,46(12):4751-4762.  
ZHANG Minghui, ZHOU Yatong, KONG Xiaoran. Short-term electrical load prediction based on deep hybrid reservoir calculation model[J]. Power System Technology, 2022,46(12):4751-4762.
- [8] 臧海洋,许瑞琦,刘璟璇,等. 基于多维融合特征和卷积神经网络的多任务用户短期负荷预测[J/OL]. 电力系统自动化.(2022-11-30)[2023-07-05]. <https://kns.cnki.net/kcms/detail/32.1180.TP.20221130.0922.003.html>.
- [9] 孙辉,杨帆,高正男,等. 考虑特征重要性值波动的MI-BiLSTM短期负荷预测[J]. 电力系统自动化,2022,46(8):95-103.  
SUN Hui, YANG Fan, GAO Zhengnan, et al. Short-term load forecasting based on mutual information and bi-directional long short-term memory network considering fluctuation in importance values of features[J]. Automation of Electric Power Systems, 2022,46(8):95-103.
- [10] 姚程文,杨苹,刘泽健. 基于CNN-GRU混合神经网络的负荷预测方法[J]. 电网技术,2020,44(9):3416-3424.  
YAO Chengwen, YANG Ping, LIU Zejian. Load forecasting method based on CNN-GRU hybrid neural network[J]. Power System Technology, 2020,44(9):3416-3424.
- [11] 朱凌建,荀子涵,王裕鑫,等. 基于CNN-BiLSTM的短期电力负荷预测[J]. 电网技术,2021,45(11):4532-4539.  
ZHU Lingjian, XUN Zihan, WANG Yuxin, et al. Short-term power load forecasting based on CNN-BiLSTM[J]. Power System Technology, 2021,45(11):4532-4539.
- [12] 王继东,杜冲. 基于Attention-BiLSTM神经网络和气象数据修正的短期负荷预测模型[J]. 电力自动化设备,2022,42(4):172-177,224.  
WANG Jidong, DU Chong. Short-term load prediction model based on Attention-BiLSTM neural network and meteorological data correction[J]. Electric Power Automation Equipment, 2022,42(4):172-177,224.
- [13] 彭文,王金睿,尹山青. 电力市场中基于Attention-LSTM的短期负荷预测模型[J]. 电网技术,2019,43(5):1745-1751.  
PENG Wen, WANG Jinrui, YIN Shanqing. Short-term load forecasting model based on Attention-LSTM in electricity market[J]. Power System Technology, 2019,43(5):1745-1751.
- [14] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York, USA: ACM, 2017:6000-6010.
- [15] 逆宝中,李庚银,武昭原,等. 基于循环扩张机制的ConvGRU-Transformer短期电力负荷预测方法[J]. 华北电力大学学报(自然科学版),2022,49(3):34-43.  
TI Baozhong, LI Gengyin, WU Zhaoyuan, et al. A short-term load forecasting method based on recurrent and dilated mechanism of ConvGRU-Transformer[J]. Journal of North China Electric Power University, 2022,49(3):34-43.
- [16] 黄飞虎,赵红磊,弋沛玉,等. 一种改进Transformer的电力负荷预测方法[J]. 现代电力,2023,40(1):50-58.  
HUANG Feihu, ZHAO Honglei, YI Peiyu, et al. An improved power load forecasting method based on Transformer[J]. Modern Electric Power, 2023,40(1):50-58.
- [17] 董家富,万雄,王岩,等. 基于XGB-Transformer模型的短期电力负荷预测[J]. 电力信息与通信技术,2023,21(1):9-18.  
DONG Jiafu, WAN Xiong, WANG Yan, et al. Short-term power load forecasting based on XGB-Transformer model[J]. Electric Power Information and Communication Technology, 2023,21(1):9-18.
- [18] XIONG Y Y, ZENG Z P, CHAKRABORTY R, et al. Nystromformer: anystrom-based algorithm for approximating self-attention[J]. Proceedings of the AAAI Conference on Artificial Intelligence AAAI Conference on Artificial Intelligence, 2021,35(16):14138-14148.

## 作者简介:

范杏蕊(1997—),女,硕士研究生,主要研究方向为短期电力负荷预测(**E-mail**:fanxingrui123@163.com);

李元诚(1970—),男,教授,博士,通信作者,主要研究方向为信息安全(**E-mail**:ncepua@163.com)。

(编辑 王锦秀)

## Short-term power load forecasting based on improved Autoformer model

FAN Xingrui, LI Yuancheng

(College of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China)

**Abstract:** Aiming at the low accuracy problem of short-term power load forecasting caused by the influence of multiple uncertain factors such as weather, temperature and holiday, a short-term power load forecasting model based on an improved Autoformer model is proposed. By changing the pre-processing convention of sequence decomposition, an internal decomposition module of depth model is designed, which extracts the intrinsically complex time series trend of hidden state in the model, and makes the model have the ability to decompose complex time series asymptotically. The Nystrom self-Attention mechanism is proposed, which uses the Nystrom method to approximate the standard self-Attention mechanism. The experimental results of power load forecasting in a region show that the proposed model has lower time complexity and higher accuracy than the standard Autoformer model.

**Key words:** short-term power load forecasting; timing decomposition module; Nystrom self-Attention mechanism; Sdformer model

(上接第163页 continued from page 163)

## Imbalance correction method of transient stability assessment model based on gradient norm

HU Litao<sup>1</sup>, WANG Huaoyuan<sup>1</sup>, DANG Ran<sup>2</sup>, TONG Haoxuan<sup>3</sup>, ZHANG Yang<sup>1</sup>

(1. Fujian Key Laboratory of New Energy Generation and Power Conversion, College of Electrical Engineering and Automation, Fuzhou University, Fuzhou 350116, China;

2. Shaanxi Aircraft Industry Limited Liability Company, Hanzhong 723000, China;

3. Taining County Power Supply Branch of State Grid Fujian Electric Power Co., Ltd., Sanming 354400, China)

**Abstract:** In order to solve the problem of transient stability assessment deviation caused by the imbalance of power system sample quantity and quality, starting from the training process of assessment model, the gradient norm of samples to model parameters is obtained by the pre-training model, and the mean ratio of gradient norm is introduced to quantify the imbalance of samples. Compared with the prior information, the mean ratio of gradient norm comprehensively considers the imbalance between the sample quantity and quality. An imbalanced correction method based on the cost-sensitive method is proposed, which is used to improve the assessment preference of the model and realize a preferable correction effect. The simulative results of IEEE 39-bus system and East China Power System verify the effectiveness of the proposed method.

**Key words:** deep learning; transient stability assessment; cost-sensitive; gradient norm; stacked sparse auto-encoder; imbalanced sample

# 附录 A

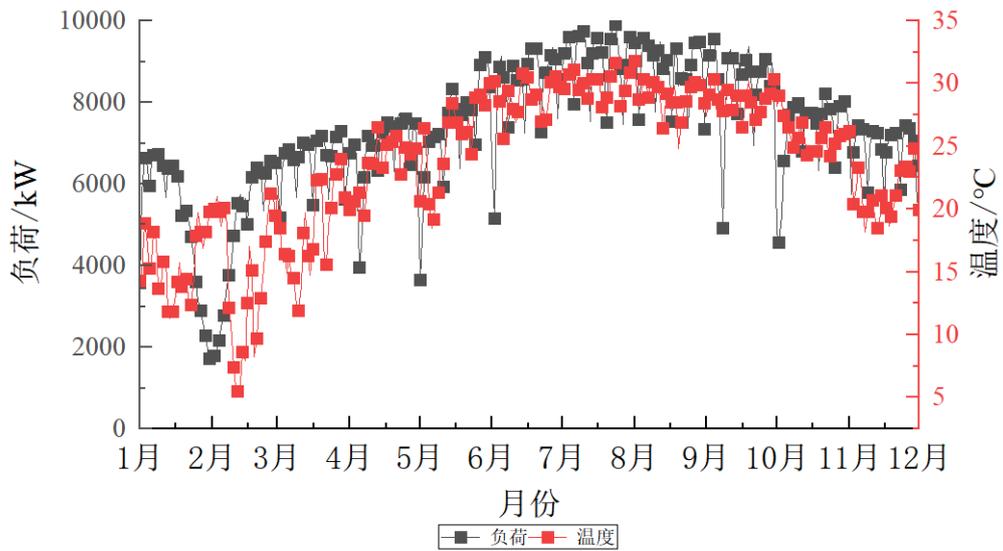


图 A1 负荷和温度随时间的变化图  
Fig.A1 Variation of load and temperature with time

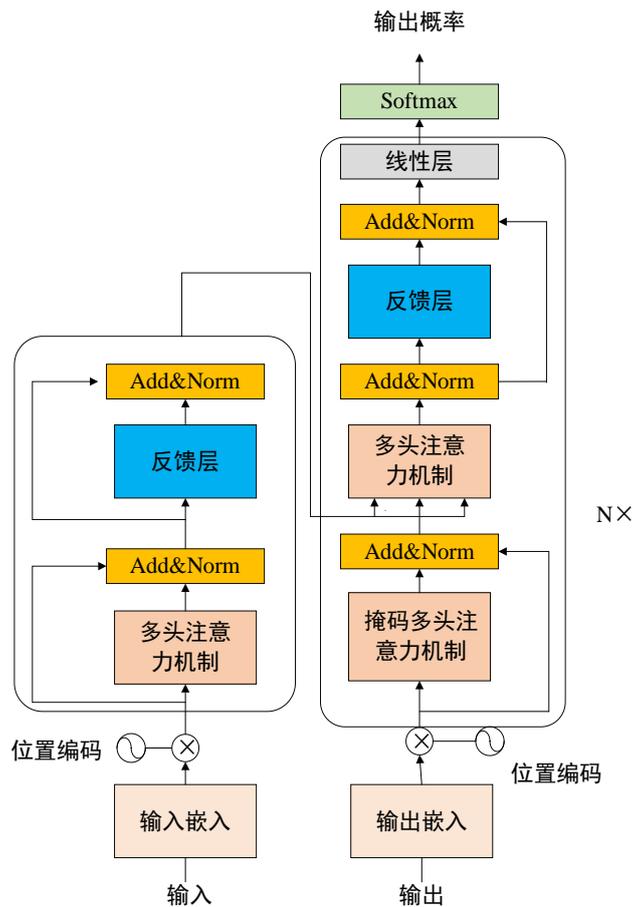


图 A2 Transformer 网络结构  
Fig.A2 Transformer network structure

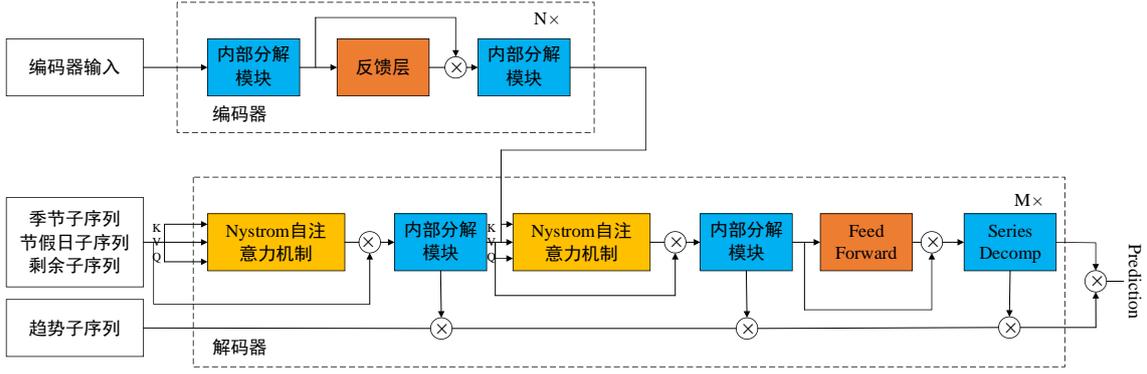


图 A3 Sdformer 网络结构  
Fig.A3 Sdformer network structure

## 附录 B

$$\begin{aligned}
 S_{en}^{l,1}, H_{en}^{l,1}, R_{en}^{l,1} &= \text{SeriesDecomp}(X_{en}^{l-1}) \\
 S_{en}^{l,2}, H_{en}^{l,2}, R_{en}^{l,2} &= \text{SeriesDecomp}(\text{FeedForward}(S_{en}^{l,1}, H_{en}^{l,1}, R_{en}^{l,1}) + S_{en}^{l,1}, H_{en}^{l,1}, R_{en}^{l,1})
 \end{aligned} \tag{B1}$$

式中：“\_”为被消除的趋势部分；L为第1个编码器层的输出。

$$\begin{aligned}
 S_{de}^{l,1}, T_{de}^{l,1}, H_{de}^{l,1}, R_{de}^{l,1} &= \text{SeriesDecomp}(\text{NystromSelfAttention}(X_{de}^{l-1}) + X_{de}^{l-1}) \\
 S_{de}^{l,2}, T_{de}^{l,2}, H_{de}^{l,2}, R_{de}^{l,2} &= \text{SeriesDecomp}(\text{NystromSelfAttention}(S_{de}^{l,1}, T_{de}^{l,1}, H_{de}^{l,1}, R_{de}^{l,1}, X_{en}^N) + S_{de}^{l,1}, T_{de}^{l,1}, H_{de}^{l,1}, R_{de}^{l,1}) \\
 S_{de}^{l,3}, T_{de}^{l,3}, H_{de}^{l,3}, R_{de}^{l,3} &= \\
 \text{SeriesDecomp}(\text{FeedForward}(S_{de}^{l,2}, T_{de}^{l,2}, H_{de}^{l,2}, R_{de}^{l,2}) + S_{de}^{l,2}, T_{de}^{l,2}, H_{de}^{l,2}, R_{de}^{l,2}) \\
 T_{de}^l &= T_{de}^{l-1} + W_{l,1} * T_{de}^{l,1} + W_{l,2} * T_{de}^{l,2} + W_{l,3} * T_{de}^{l,3}
 \end{aligned} \tag{B2}$$

表 B1 数据集特征参数  
Table B1 Dataset characteristic parameters

特征参数	特征参数类型	详细描述
日期因素	季节	春夏秋冬: 1, 2, 3, 4
	月份	一月-十二月: 1-12
	节假日	工作日: 1; 节假日: 0
	周	周一-周日: 1-7
天气因素	露点温度	水蒸气与水达到平衡状态的温度
	干球温度	被测量空气的真实热力学温度
	湿球温度	热力学绝热饱和温度
	湿度	表示空气中大气干燥程度
电力负荷	电力负荷	发电机组额定情况下单位时间发电量

$$Y^* = \frac{Y - Y_{\text{mean}}}{Y_{\text{max}} - Y_{\text{min}}} \tag{B3}$$

式中：Y为原始负荷数据；Y\*为归一化处理后的数据；Y<sub>mean</sub>、Y<sub>max</sub>和Y<sub>min</sub>分别为负荷数据的平均值、最大值和最小值。

$$e_{\text{MRE}} = \frac{1}{n} \sum_{i=1}^n \frac{\bar{y} - y}{y} \times 100\% \quad (\text{B4})$$

$$e_{\text{ME}} = \frac{1}{n} \sum_{i=1}^n |y - \hat{y}| \quad (\text{B5})$$

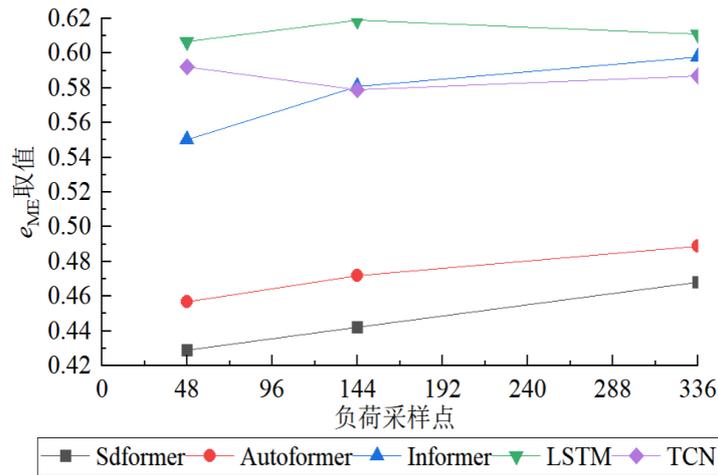
式中：y 为真实数据； $\bar{y}$  为预测值。

## 附录 C

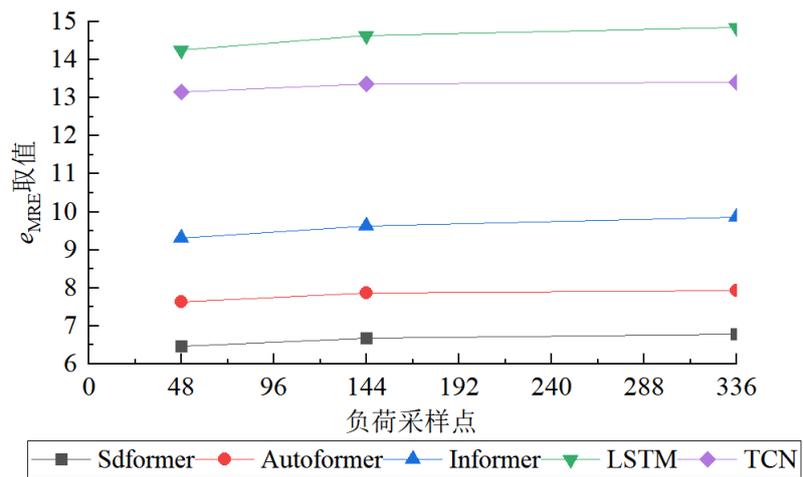
表 C1 实验结果对比表

Table C1 Comparison table of experimental results

预测模型	负荷采 样点	春季		夏季		秋季		冬季		节假日	
		$e_{\text{ME}}$	$e_{\text{MRE}}$								
Sdformer	48	0.438	6.542	0.429	6.473	0.445	6.534	0.431	6.536	0.447	6.547
	144	0.452	6.692	0.442	6.683	0.452	6.702	0.448	6.706	0.466	6.713
	336	0.476	6.793	0.468	6.783	0.479	6.808	0.470	6.812	0.484	6.817
Autoformer	48	0.464	7.654	0.457	7.643	0.468	7.658	0.460	7.663	0.470	7.668
	144	0.484	7.867	0.472	7.861	0.489	7.871	0.479	7.883	0.493	7.891
	336	0.498	7.935	0.489	7.923	0.477	7.942	0.469	7.945	0.484	7.953
Informer	48	0.559	9.365	0.550	9.312	0.564	9.328	0.554	9.335	0.566	9.427
	144	0.590	9.653	0.581	9.626	0.592	9.692	0.584	9.781	0.607	9.796
	336	0.604	9.875	0.598	9.862	0.609	9.979	0.601	9.991	0.604	9.998
LSTM	48	0.613	14.376	0.607	14.247	0.616	14.264	0.609	14.354	0.617	14.365
	144	0.627	14.654	0.619	14.631	0.629	14.829	0.622	14.818	0.632	14.896
	336	0.624	14.872	0.611	14.855	0.620	14.867	0.615	14.875	0.624	14.887
TCN	48	0.597	13.143	0.592	13.149	0.599	13.167	0.594	13.152	0.601	13.169
	144	0.589	13.375	0.579	13.363	0.603	13.408	0.586	13.372	0.610	13.392
	336	0.597	13.445	0.587	13.400	0.600	13.432	0.591	13.403	0.599	13.443



(a) ME 随负荷采样点的变化



(b) MRE 随负荷采样点的变化

图 C1 各模型误差随采样点的变化

Fig.C1 Error change of each model with sampling point