|  |  |
| --- | --- |
| **学 号： 2019210441** | **密 级： 公开** |



**Hefei University of Technology**

**本科毕业设计（论文）**

**UNDERGRADUATE THESIS**



**类 型：** 【论文】

**题 目：** 基于多头注意力机制的电力数据预

测方法研究

**专业名称：** 【信息管理与信息系统】

**入校年份：** 【2019级】

**学生姓名：** 孟熙冉

**指导教师：** 【台建玮】 【讲师】

**系 名 称：** 【信息管理与信息系统系】

**完成时间：** 【2023年05月】

合 肥 工 业 大 学

**本科毕业设计（论文）**

**基于多头注意力机制的电力数据预测方法研究**

学生姓名： 孟熙冉

学生学号： 2019210441

指导教师： 【台建玮】 【讲师】

专业名称： 【信息管理与信息系统】

系 名 称： 【信息管理与信息系统系】

2023年 05 月

**A Dissertation Submitted for the Degree of Bachelor**

**Research on power data prediction method based on multi-head attention mechanism**

By

Meng xiran

Hefei University of Technology

Hefei, Anhui, P.R.China

May, 2023

**毕业设计（论文）独创性声明**

本人郑重声明：所呈交的毕业设计（论文）是本人在指导教师指导下进行独立研究工作所取得的成果。据我所知，除了文中特别加以标注和致谢的内容外，设计（论文）中不包含其他人已经发表或撰写过的研究成果，也不包含为获得 合肥工业大学 或其他教育机构的学位或证书而使用过的材料。对本文成果做出贡献的个人和集体，本人已在设计（论文）中作了明确的说明，并表示谢意。

毕业设计（论文）中表达的观点纯属作者本人观点，与合肥工业大学无关。

毕业设计（论文）作者签名：　　　　签名日期：　　　年　　月　　日

**毕业设计（论文）版权使用授权书**

本学位论文作者完全了解 合肥工业大学 有关保留、使用毕业设计（论文）的规定，即：除保密期内的涉密设计（论文）外，学校有权保存并向国家有关部门或机构送交设计（论文）的复印件和电子光盘，允许设计（论文）被查阅或借阅。本人授权 合肥工业大学 可以将本毕业设计（论文）的全部或部分内容编入有关数据库，允许采用影印、缩印或扫描等复制手段保存、汇编毕业设计（论文）。

（保密的毕业设计（论文）在解密后适用本授权书）

|  |  |
| --- | --- |
| 学位论文作者签名： | 指导教师签名： |
| 签名日期：　　　年 月 日 | 签名日期：　　　年 月 日 |
|  |  |

**摘 要**

随着人工智能和大数据技术的发展，时间序列预测已经成为当下研究的热点之一。在电力领域，时序数据的分析对于电力系统的监控、预测和优化具有重要的意义，有利于电力系统预测用电高峰，合理分配电力资源，而对电力负荷进行长期预测是电力规划的关键需求。以往的基于Transformer的模型采用自注意力机制来发现依赖关系，这就会导致信息利用瓶颈，因此在长期预测时表现一般。本文设计了一种基于多头注意力机制的深度学习模型“PM-former”，以Autoformer模型为框架，设计了基于特征工程的预处理模块和混合专家分解模块，对电力数据进行了长期预测。使用实际电力数据进行预测，对比实验表明，在输入长度为96，预测长度为96，192，336的条件下，本文提出的模型在MAE和MSE方面均取得一定优势。

**关键词：**多头注意力机制；时间序列预测；Autoformer；电力；PM-former；

ABSTRACT

With the advancement of artificial intelligence and big data technologies, time series prediction has emerged as one of the hotspots in current research. In the field of electricity, the analysis of time-series data is crucial for the monitoring, forecasting, and optimization of power systems. It facilitates the prediction of power usage peaks and rational allocation of power resources, and long-term forecasting of power load is a key requirement for power planning.Past Transformer-based models have employed a self-attention mechanism to identify dependencies, leading to bottlenecks in information utilization, thus exhibiting mediocre performance in long-term prediction. This paper proposes a deep learning model, "PM-former", based on the multi-head attention mechanism. The model uses the Autoformer model as a framework and designs a preprocessing module based on feature engineering and a mixed expert decomposition module for long-term prediction of power data.When actual power data is used for prediction, comparative experiments show that under the conditions of an input length of 96 and prediction lengths of 96, 192, and 336, the model proposed in this paper achieves certain advantages in terms of MAE and MSE.

**KEYWORDS:** Multi-head attention mechanism; Electric power; Autoformer; Time series prediction；PM-former

**目录**

[1](#_Toc228386152) **[绪论](#_Toc228386152)** [1](#_Toc228386152)

[1.1 研究背景与意义 1](#_Toc932779475)

[1.2 时间序列预测电力领域研究现况 2](#_Toc594013225)

[1.2.1 时间序列预测 2](#_Toc2076281319)

[1.2.2 时间序列预测电力领域研究现况 2](#_Toc1598348330)

[1.3 模型创新点 4](#_Toc567441987)

[1.3.1 基于特征工程的预处理模块 4](#_Toc22599182)

[1.3.2 基于MOE的专家混合分解模块 5](#_Toc1867330002)

[1.4 论文组织架构 5](#_Toc889326356)

[2](#_Toc421882172) **[模型设计](#_Toc421882172)** [7](#_Toc421882172)

[2.1 多头注意力机制 8](#_Toc1730146057)

[2.2 时间序列分解 10](#_Toc1636199619)

[2.2.1 传统的时间序列分解 10](#_Toc1078896698)

[2.2.2 混合专家分解 12](#_Toc1812371665)

[2.3 基于Autoformer的PM-former框架设计 13](#_Toc622524607)

[2.3.1 编码器 13](#_Toc230741665)

[2.3.2 解码器 14](#_Toc1867180820)

[2.3.3 基于周期依赖关系的自相关机制 14](#_Toc529508129)

[2.4 预处理模块 16](#_Toc270890935)

[2.5 本章小结 16](#_Toc198612905)

[3](#_Toc897506897) **[实验结果与分析](#_Toc897506897)** [17](#_Toc897506897)

[3.1 数据集描述及可视化 17](#_Toc473281351)

[3.2 评价指标选择及阐述 18](#_Toc160237769)

[3.3 实验细节与参数设置 18](#_Toc171690245)

[3.4 对比实验 19](#_Toc1527409794)

[3.5 消融实验 23](#_Toc156891520)

[3.6 本章小节 23](#_Toc1913341771)

[4](#_Toc1115015019) **[总结与展望](#_Toc1115015019)** [25](#_Toc1115015019)

[4.1 本文工作及研究总结 25](#_Toc1115120611)

[4.2 后续研究分析与展望 25](#_Toc742321708)

**[参考文献](#_Toc1468440933)** [27](#_Toc1468440933)

**[致谢](#_Toc1204689607)** [30](#_Toc1204689607)

**插图清单**

[图 1.1 时间序列预测示例图 2](#_Toc2083620550)

[图 2.1 模型结构图 8](#_Toc1023592764)

[图 2.2 自注意力机制（左）与多头注意力机制（右） 9](#_Toc291864420)

[图 2.3 数据集部分数据可视化 11](#_Toc525034180)

[图 2.4 数据趋势项可视化 11](#_Toc1582182222)

[图 2.5 数据季节项可视化 11](#_Toc1978519675)

[图 2.6 数据残差项可视化 12](#_Toc1819425728)

[图 2.7 Autoformer模型结构 13](#_Toc2044168227)

[图 2.8 自相关机制 15](#_Toc369620074)

[图 3.1 数据可视化 17](#_Toc2120153544)

[图 3.2 PM-former-pl336 21](#_Toc1506792827)

[图 3.3 Autoformer-pl336 21](#_Toc1403527758)

[图 3.4 Informer-pl336 21](#_Toc1900216150)

[图 3.5 Reformer-pl336 22](#_Toc783666187)

[图 3.6 Transformer-pl336 22](#_Toc1891007311)

**表格清单**

[表 3.1 相关参数表 18](#_Toc652491861)

[表 3.2 实验数据表 20](#_Toc380873997)

[表 3.3 消融实验数据 23](#_Toc271929272)

# 绪论

1.1 研究背景与意义

时间序列是一组或多组随机变量，通常以相对固定的采样频率记录数据变化过程，比如按照每小时，每天，每15分钟等采样频率进行采样。由于时间序列数据的季节性和趋势-周期性，过去的时间序列数据可能包含了未来数据发展变化的规律，两者之间具有相关性。因此，时间序列数据预测研究就是在已有的时间序列数据基础上，通过对时序数据本身的季节性，趋势性等进行特征提取，从而找出未来时刻或时间段内某个或某些随机变量随时间变化的趋势和规律。不难发现，时间序列预测在我们的生活中应用是比较广泛的。

在电力领域中，时间序列的应用场景还是比较丰富的。比如，电力交易市场中的电力价格预测[1]：通过使用时间序列预测方法，对未来一段时间电力市场的电力价格进行预测，帮助用户更好的地制不同时间段的用电量，从而减少在高价时段减少用电量，控制用电支出；同时可以帮助供电企业更好的制定用电价格，获得更好的收益；也可以帮助发电企业进行发电量规划，优化电力的供需平衡。公寓的短期用电消耗预测[2]：可以帮助用户进行用电规划，了解自己的用电习惯。某个区域的生活生产的用电负荷预测[3]:优化电力系统供电与发电规划，促进电网智能化，更好地排运营计划，降低供电运营成本。

近些年来，随着智能电网的提出与发展，对于电力负荷的精确化预测提出了进一步要求，对未来一段时间的电力负荷预测有利于供电系统提前做出反应，如合理分配供电量等。通过时间序列预测可以帮助电力公司和政府制定相应的电力规划和调度策略，以确保电力系统的稳定运行和满足用户的用电需求。随着深度学习以及大数据技术的进一步发展，基于深度学习的时间序列预测方法[4-7]在电力领域的应用也十分丰富。当下对电力数据预测的迫切需求是如何延长预测时间，更长的预测时间代表着更多的规划冗余，因此本文研究的是对电力负荷的长期预测问题，输出长度比之输入长度，可以是其几倍。本文以多头注意力机制为基础，Autoformer[8]模型为框架，创新引入基于特征工程的数据预处理和专家混合分解[9]模块。相比于其他以Transformer[10]为基础的模型，本文所提出的模型在电力负荷的长期预测（输入长度为96，输出长度为96，192，336）中表现出了较好的精确度。

1.2 时间序列预测电力领域研究现况

1.2.1 **时间序列预测**

时间序列预测（Time Series Forecasting）是一种统计学和机器学习方法，用于分析和预测按时间顺序排列的数据点。时间序列数据是按时间顺序收集的数据，例如某只股票每天的价格、某地每小时的用电量或某地每天的气温。时间序列预测的主要目标是利用过去的数据来预测未来的趋势和模式。通常来讲，时间序列数据预测研究就是在已有的时间序列数据基础上，通过对时序数据本身的季节性，趋势-周期性等进行特征提取，从而找出未来时刻或时间段内某个或某些随机变量随时间变化的趋势和规律，可以通过图1.1对时间序列预测有一个简单的认识。

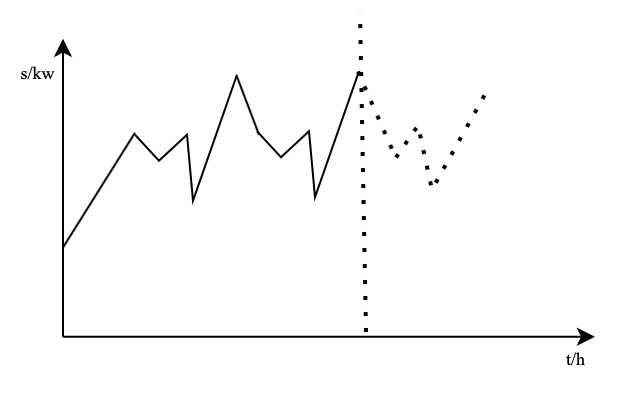


图1.1 时间序列预测示例图

在该示例图中，横轴表示时间，单位为小时，纵轴表示用电量，单位为千瓦时。竖直虚线之前的实折线表示过去的用电量数据，竖直虚线后的虚折线表示根据过去的用电数据，通过学习其中的特征，根据数据的趋势-周期性，季节性等对未来用电数据做出的预测值。

1.2.2 **时间序列预测电力领域研究现况**

时间序列预测一直是国内外研究重点，也经过了长时间的发展。从一开始的基于统计学的时间序列预测方法，到基于机器学习的时间序列预测方法，以及当下火热的基于深度学习的时间序列预测方法[11]。这三种类型的方法目前在时间序列预测的电力领域都有一定的研究占比，因此下文将对这三种类型的方法进行阐述。

（1）基于统计学的时序预测方法：时间序列预测中基于统计学的方法有很多，比如：自回归模型（AR）[12]，自回归移动平均模型（ARMA）[13]，自回归整合移动平均模型（ARIMA）[14],季节分解的自回归整合移动平均模型（SARIMA）[15]，广义自回归条件异方差模型（GARCH）[16]，季节性分解的局部线性趋势模型(TL)[17]。

基于统计学的时间序列预测方法通常应用于线性与非线性时序数据处理，比如前文提到的自回归模型可以处理线性时间序列数据，季节性分解的局部线性趋势模型可以处理非线性的时间序列数据。同时也常应用于平稳和非平稳时间序列数据处理，比如前文提到的自回归移动平均模型可以处理平稳的时间序列数据，同时也可以通过自回归整合移动平均模型进行差分操作，将非平稳数据转化为平稳数据进行分析。

基于统计学的时间序列预测方法经历了长足的发展，一开始统计学家们主要侧重于研究时间序列数据本身的性质，最早可以追溯到自回归模型（AR），使用线性加权方法，对过去短时间内的历史数据进行处理，对未来进行预测，效果一般，未能很好地捉数据所蕴含的特征，但这种加权思想一直影响着后续的时间序列预测方法，比如当下火热的多头注意力机制，本质也体现了加权思想。因为现实生活中的时间序列数据大多是非平稳的，因此后来Box和Jenkins提出了自回归整合移动平均（ARIMA）模型的方法。在电力领域，当下也有很多利用ARIMA模型进行时间序列预测的例子：赵滨滨等人[18]使用ARIMA模型对分布式光伏系统输出功率进行了预测；王斌[19]使用ARIMA模型对电力负荷进行了短期分析预测。

（2）基于统计学习的时序预测方法：时间序列预测中基于机器学习的方法有很多，比如支持向量机（SVM）[20]，循环神经网络（RNN）[21]，长短时记忆网络（LSTM）[22]，梯度提升决策树（GBDT）[23]，XGBoost[24]等等。这些方法主要通过学习时间序列数据中的模式以及趋势对其进行预测。机器学习模型从过去的数据中捕获潜在的数据结构以及依赖关系，从而对未来的数据点做出预测。

当前在时间序列预测领域，基于机器学习的方法使用较多的比如长短时记忆网络以及相关的改进模型：朱乔木等人[25]基于LSTM模型对风力发电厂的发电功率进行了短期预测；李少杰[26]以LSTM模型为基础，构建了一种组合神经网络的多步输出预测方法以提高短期负荷预测性能。XGBoost模型的使用也很多：陈振宇等人[27]将LSTM模型与XGBoost模型结合，对短期的电力负荷进行预测。

（3）基于深度学习的时序预测方法：时间序列预测中基于深度学习的方法有很多，比如时序卷积网络（TCN）[28],门控循环单元（GRU）[29]，Transformer模型[10]，Sequence-to-Sequence（Seq2Seq）模型[30]等等。相比于以上两种类型的时间序列预测方法，基于深度学习的方法在自动特征学习，高度非线性表达能力，长期依赖关系处理，端到端训练，多任务学习和迁移学习等方面表现出来了强大的优势。同时深度学习模型具有更强的泛化能力，可以应用于多种不同类型的时间序列数据。

主要通过搭建神经网络模型来对时序数据进行预测。通过搭建神经网络模型去学习时间序列数据中的潜在特征和结构。当前比较火热的深度学习时序预测方法主要是Transformer以及其变体，因为Transformer的核心：“多头注意力机制”在时序数据的长期依赖进行建模方面获得了很大的优势，这使得构建更强大的大模型成为可能。

在基于深度学习的时序预测方面，李磊等人[31]采用TCN与注意力机制相结合的方法对电力负荷进行了预测;徐武等人[32]以GRU为基础，对风电功率进行了短期预测。刘洪笑等人[33]以Informer模型为基础，对长序列的电力负荷进行了预测。

1.3 模型创新点

本文所提出的模型是以多头注意力机制为基础，以Autoformer模型为框架改进。Autoformer模型是在自相关的操作中体现多头注意力机制的（具体可参考论文第二章模型设计），这使得其在处理长序列时能同时捕捉全局信息与局部信息，从而获得了较好的性能。本文所提出的模型在Autoformer模型的编码器前加入基于特征工程的预处理模块，同时在用专家混合分解模块替换原模型中的序列分解模块。前者使得模型可以学习更多的特征，提高模型在长期预测问题中的预测准确度；后者可以能够更好地捕捉和处理复杂的周期性模式与趋势分量相结合的数据，在捕捉趋势、模型适应性、处理复杂性能力和预测准确性方面具有优势。

1.3.1 **基于特征工程的预处理模块**

数据预处理在深度学习领域发挥着巨大的作用，数据预处理一般是将原始数据进行清洗、转换和整理的过程，使得数据能更好地适应模型训练和分析。数据预处理的作用主要体现在以下几个方面：提高数据的质量，对数据进行编码，标准化或者归一化以及特征提取和降维等。Autoformer[8]模型本身具有一定的预处理功能，本文的模型在此基础上增加了基于特征工程的数据预处理模块，位于编码器之前。根据电力数据集的特征（详见3.1数据集描述及可视化）不难发现电力数据波动与季节性关联较大，因此可以根据日期对电力数据进行特征提取操作，从包含日期的数据集中提取出月份信息，并将月份信息存储到一个新的列表中。最后，将这个月份列表添加到原始数据表中作为一个新的列。主要原因是电力消耗可能会受到季节性因素的影响，比如夏季和冬季由于气候原因，空调和取暖设备的使用可能会导致电力需求的变化。通过将月份信息作为一个附加特征添加到数据集中，可以帮助模型捕捉这种季节性变化。通过该预处理操作，可以使得本文模型学习更多的特征，从而提高预测的精确度。

1.3.2 **基于**MOE**的专家混合分解模块**

专家混合模型（Mixture of Experts, MOE）是一种特殊的神经网络结构，不同于一般的神经网络的是它根据数据进行分离训练多个模型，各个模型被称为专家，而门控模块用于选择使用哪个专家，模型的实际输出为各个模型的输出与门控模型的权重组合。专家混合分解模块（MOE Decomp）是一种基于专家混合模型思想，设计的一种针对复杂时间序列数据的模块。专家混合分解模块中包括多个不同大小的平均滤波器，用以从输入信号中分离出多种趋势成分，从而更有效地提取趋势分量。同时，它还包含一组与数据相关的权重，用于将这些提取出的趋势成分整合为最终的趋势。相比于Autoformer模型中的序列分解模块，专家混合分解模块的主要优势在于能够更好地捕捉和处理复杂的周期性模式与趋势分量相结合的数据。

1.4 论文组织架构

本文旨在研究深度学习在时间序列预测的应用，提高在长期预测条件下模型预测的准确度。本文以电力数据为研究背景，采用的数据集来自于加州大学欧文分校机器学习网站里的电力数据集和清华大学机器学习里的电力数据集，研究基于多头注意力机制的深度学习方法在时间序列预测中长期预测的应用，根据时序数据特征，结合和改进深度学习模型，提高长期预测中的预测准确性。本文主要分为四大章节。

第一章节，绪论，主要分为四小节。第一节主要介绍时间序列预测在电力领域的研究背景与意义。第二节主要介绍当前国内外时间序列预测的研究现况，同时简单介绍了时间序列预测的基本含义。第三节则主要介绍了本文对Autoformer模型的改进以及创新点，包含两点：基于特征工程的预处理模块和基于MOE的混合专家分解模块。

第二章节，模型设计，主要分为五小节。首先简述里本文模型的设计，之后三小节是对相关模型和机制进行理论阐述。第一小节介绍了多头注意力机制以及在本文模型中是如何体现的；第二小节介绍了时间序列分解，从传统的时间序列分解和混合专家分解两个方向介绍，阐述了混合专家分解模块相较于Autoformer模型中序列分解模块的优势。第三小节主要阐述了本文的架构基础：Autoformer的核心组件。第四小节则主要阐述了与预处理模块的作用。第五小节则对本章工作进行了总结阐述。

第三章节，实验结果与分析，主要分为五小节。第一小节主要对数据集进行描述，如可视化以及缺失值处理等；第二节则对本文的评价指标进行了介绍；第三小节则主要进行对比实验描述，在对比实验的对比模型选用中，本文采用了当前时间序列预测领域使用较多的Transformer模型以及相关变体；第四节阐述了消融实验，分析两模块对模型的改进影响。第五节是对本章工作的总结。

第四章节，结论与展望，主要分为两小节。第一小节主要说明文模型在现实场景中的应用意义以及本文的研究结论。第二小节则主要对后续研究进行分析与展望。

# 模型设计

本文模型采用了多头注意力机制，以Autoformer模型为框架，设计了基于特征工程的预处理模块和混合专家分解模块。本文模型结构图如图2.1所示。

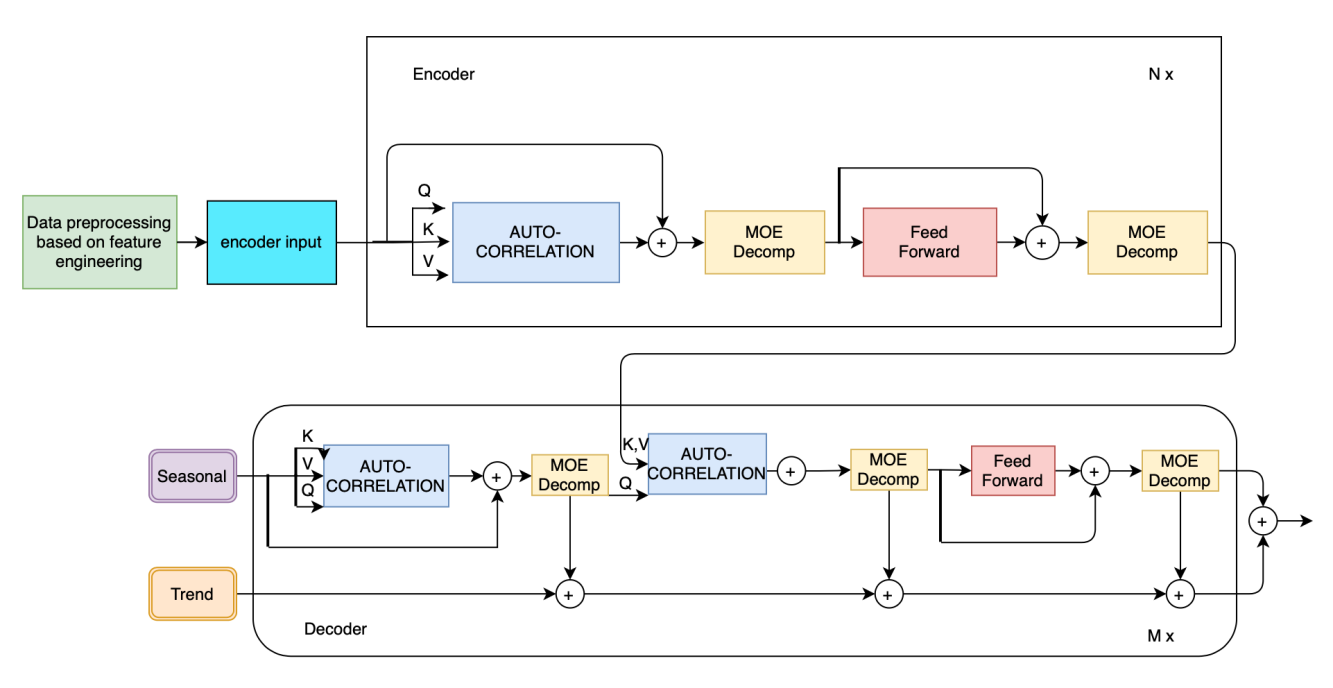
****

图2.1 模型结构图

本文模型通过基于特征工程的预处理模块，从包含日期的数据集中提取出月份信息，并将月份信息存储到一个新的列表中。最后，将这个月份列表添加到原始数据表中作为一个新的列。通过该预处理操作，可以使得本文模型学习更多的特征，从而提高预测的精确度。同时对Autoformer的内部架构进行改进，将原有的序列分解模块替换为混合专家分解模块，更好地捕捉和处理复杂的周期性模式与趋势分量相结合的数据。从而提高模型在长期预测中的准确度。

在本文模型中，数据集首先通过基于特征工程的预处理模块添加月份信息列，供模型学习更多特征。编码器的输入是过去的L个时间步长（在本模型中，一时间步长等于一小时），在进入自相关层前将输入序列编码为一个固定维度的向量序列，引入位置编码，形成包含位置信息的输入表示；该输入表示进入自相关层捕捉序列中的局部和全局依赖关系，然后通过混合专家分解模块消除长期趋势循环部分，进入前馈神经网络整合混合专家分解模块的输出信息，再通过混合专家分解模块进行输出，该输出包含过去的季节性信息，因此输入到编码器的自相关层被利用。对于第h个编码器层的总体方程如下所示（其中“\_”表示去除的趋势部分）：

编码器输入包含季节性部分和趋势-周期部分，对于季节性部分，取编码器输入中季节性部分的一半（后半部分）以表示最近信息，并在其后用标量填充的长度为0的占位符；对于趋势-周期部分，取编码器输入中趋势-周期性部分的一半（后半部分）以表示最近信息，并在其后用标量填充的长度为编码器输入均值的占位符，表示如下：

编码器部分包含两层：趋势性的积累机制和季节性的自相关机制，季节性输入经过自相关层捕捉局部和全局依赖关系，并通过混合专家分解模块分解出趋势性部分，该部分则与趋势性输入进行累积。分离出的季节性部分则与编码器的输出一起输入自相关层，并再次进行序列分解，分解后的季节项输入前馈神经网络，趋势项继续累积，最后经由混合专家分解模块进行分解，实现趋势项的累加，最后季节项与趋势项合并形成输出，对于第h层的解码器可由以下公式表示：

下文将阐述相关机制，框架和核心组件的理论。

2.1 多头注意力机制

多头注意力机制（Multi-Head Attention）是Transformer模型的核心组成部分之一，在自然语言处理和时间序列预测任务中应用广泛，是自注意力机制的扩展延伸。图2.2展示了自注意力机制与多头注意力机制的结构。由图2.2不难看出，多头注意力机制有h个相互独立的自注意力层，每层的权重矩阵WQ,WK,WV不同，具体计算过程可见以下公式：

*,,*

*, i = 1,2.....h*

*W0*

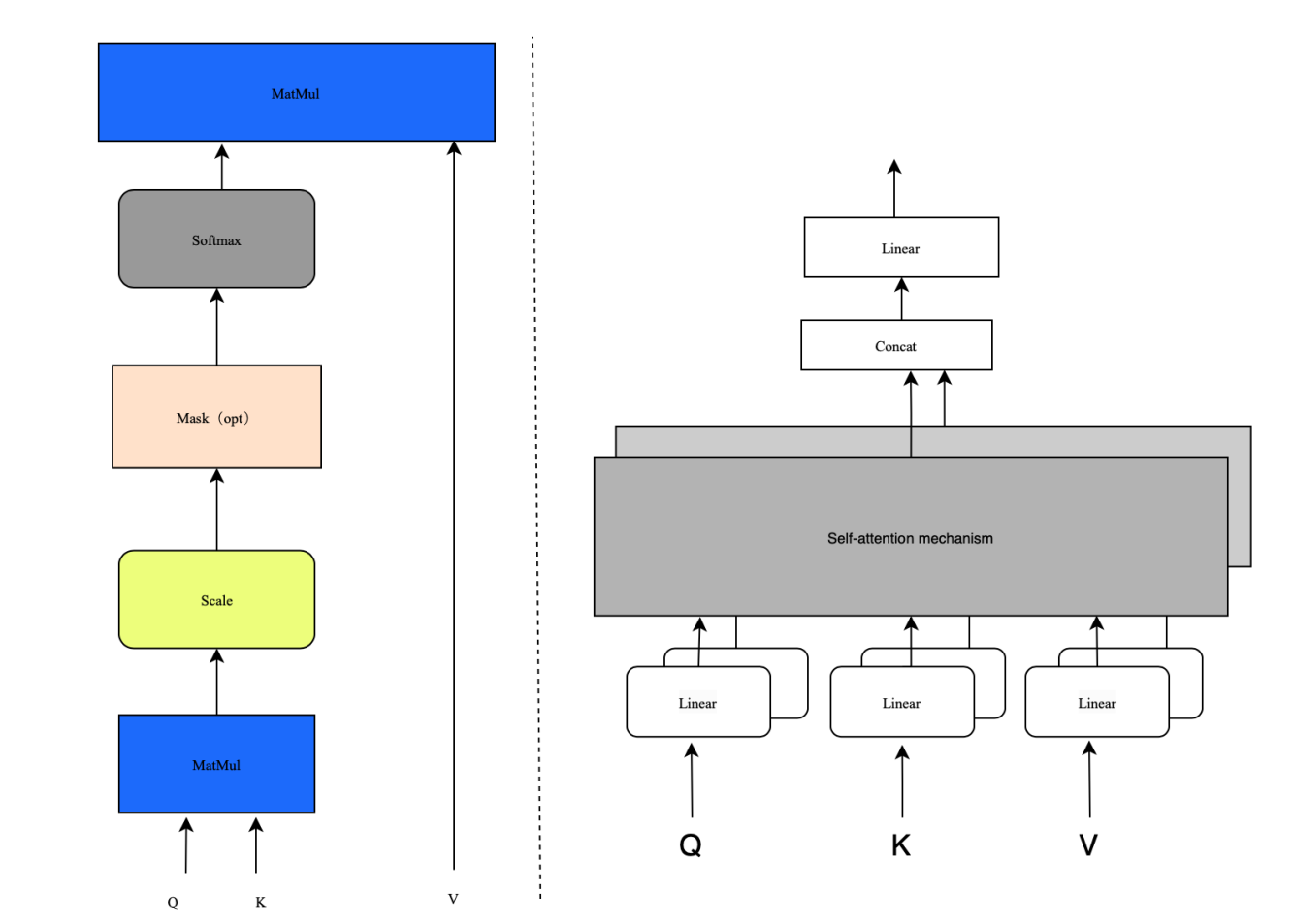


图2.2 自注意力机制（左）与多头注意力机制（右）

其中,第一个公式中中h表示多头注意力机制的注意力层数,Zi则表示每层的注意力头的输出向量，W0为输出的投影矩阵,其余三个为可学习的权重矩阵。第二个公式中 Qi、Ki和Vi可视为单头注意力中的Q、K、V在不同特征子空间下进行拆分,在不增加额外计算成本的情况下, 从多个角度提取特征之间的相关性，最后将各个自注意力层提取到的信息合并,得到更加丰富全面的特征信息[34]。

进一步，多头注意力机制的工作原理可以从以下三个步骤进行理解：（1）线性投影：对于输入序列中每个元素的嵌入向量，我们通过线性层将其映射到各个表示子空间。这些线性层通常由三个权重矩阵：查询矩阵、键矩阵和值矩阵组成。这样一来，我们便获得了每个子空间中元素的元素的查询（Query）、键（Key）和值（Value）表示。（2）计算注意力权重：在各个子空间中，我们则通过计算查询（Query）和键（Key）之间的点积来计算注意力权重，然后将结果除以缩放因子（通常为表示维度的平方根）。最后，我们对这些注意力权重应用softmax函数，从而确保这些注意力权重的和为1，具有概率分布的性质。（3）加权求和：最后，在每个子空间中，我们将计算出的注意力权重应用于值（Value）表示，通过加权求和得到输出表示。这意味着对于每个输入元素，其输出表示是值（Value）表示的加权组合，权重是由查询（Query）和键（Key）之间的关系决定的。因此，相比于自注意力机制，多头注意力机制使得模型能够在多个表示子空间中学习元素之间的依赖关系，同时在以下三个方面表现出了较强的优势：捕捉特征，并行计算，泛化能力。

在本文提出的模型中，多头注意力机制主要通过自相关机制体现出来，在本文模型的核心组件自相关机制中采用多头版本，生将输入序列分成多个子序列并输入不同层，自相关机制可以无缝替换注意力机制，PM-former使用多头自注意力机制来处理这些子序列，多头自注意力机制允许模型学习不同时间尺度上的依赖关系，从而更好地捕捉输入序列中的复杂模式和趋势。在自相关机制处理子序列后，本文模型PM-former将这些子序列重新组合以生成预测结果。这种方法允许模型在预测时考虑输入序列中的各种局部模式和全局趋势。自相关机制由基于周期的依赖关系和时间延迟聚合两部分组成，由时序数据的周期性可知不同周期的相同点位的数值是相似的，比如对于家庭用电的电力数据，在每周的相似的时间段表现出相似的耗电量，因此对于每一层而言，以每层输入的序列为基础自相关机制通过设置不同的时间延迟产生不同的子序列，来研究子序列与原序列的相关性，然后通过计算不同子序列与原序列的相关系数，经过softmax标准化置信度来聚合子序列，对于自相关机制在此就不多做赘述，关于自相关机制在2.3 章节里有更为详细的阐述。

2.2 时间序列分解

2.2.1 **传统的时间序列分解**

在时间序列预测研究中，时间序列分解[18]常用的一种分析时间序列数据的方法。时间序列分解将时序数据分解为三部分：季节项，趋势项，残差项。下面将以清华大学机器学习的开源电力数据集（因数据量问题，只取其中两百个数据点）进行时间序列分解示例：图2.3，图2.4，图2.5，图2.6（四图纵轴单位千瓦时，横轴单位为天）。

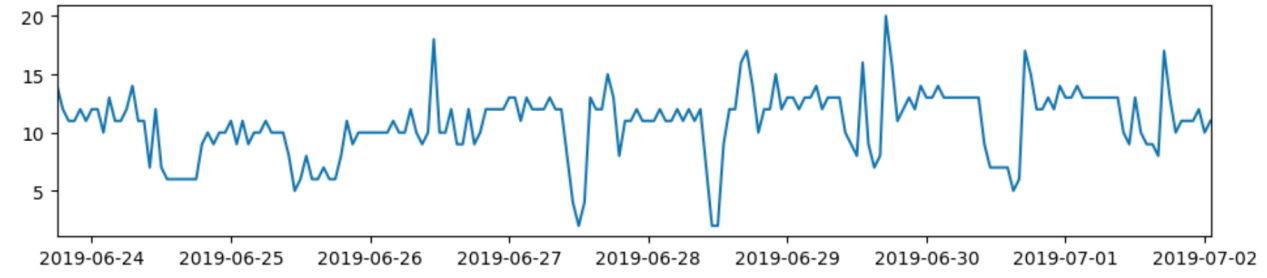


图2.3 数据集部分数据可视化

在图2.3中，我们不难观察到用电数据的周期性和趋势性都是比较明显的，因此通过时间序列分解的方法，我们可以得到该部分数据的季节项，趋势项和残差项。

在处理时序数据时，使用时间序列分解方法能够揭示数据的基本结构，提高预测准确性，同时也可以帮助进行异常检测，具有很高的灵活性和可解释性。这使得时间序列分解成为处理时序数据时非常有用的工具。

Autoformer在模型内部设置了序列分解模块，并将该模块置于编码器和解码器中，该模块将序列分解为季节项和趋势项两部分，进而反应序列的趋势发展和季节性变化。同时编码器与解码器中的时间序列分解模块起的作用不同：在编码器中，通过时间序列分解模块，去除了趋势项，只专注于季节项建模。在解码器中，解码器的输入为两部分，一部分是序列分解后的季节项，一部分是序列分解后的趋势项，因此解码器包括两部分：趋势性成分的积累结构和季节性成分的自相关机制。因为时间序列分解是作为一个内部模块存在于Autoformer中，每个解码器和编码器都有

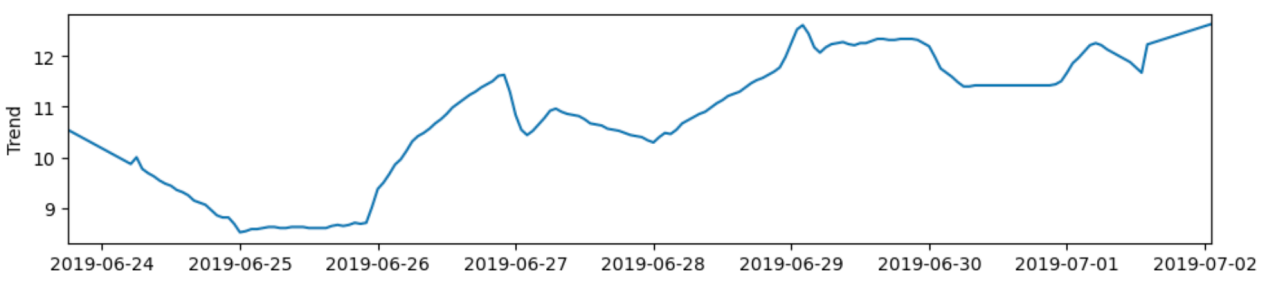


图2.4 数据趋势项可视化

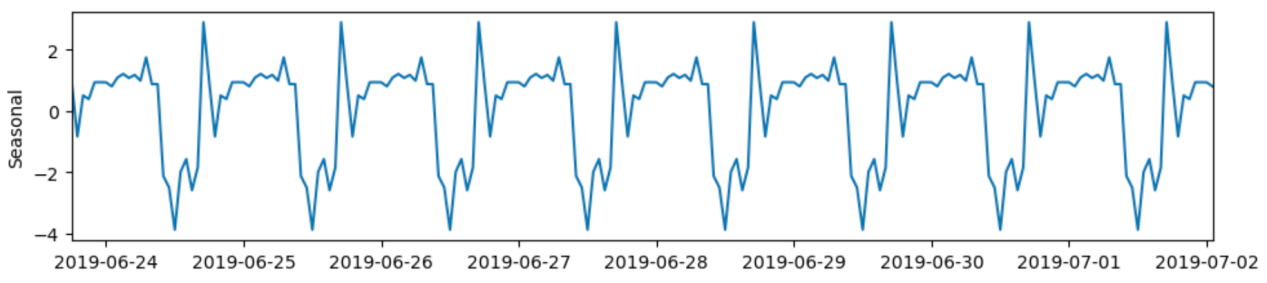


图2.5 数据季节项可视化

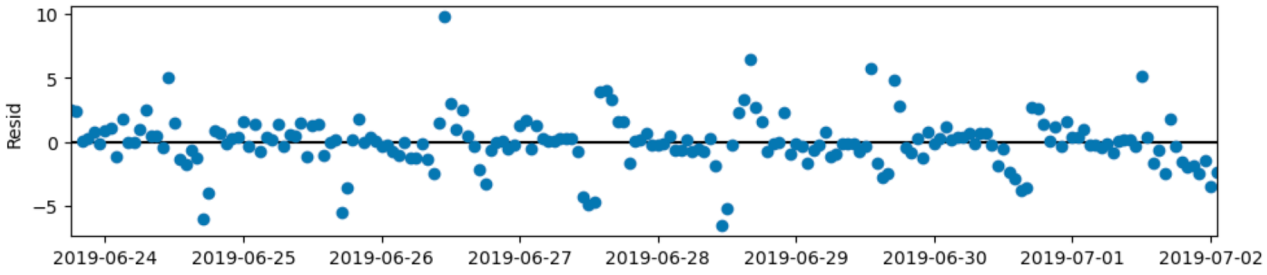


图2.6 数据残差项可视化

多个序列分解模块，所以Autoformer拥有渐进分解的能力，这保证了该模型在长期预测任务中保持较好的准确度。

2.2.2 **混合专家分解**

在Autoformer模型中，序列分解模块采用的是一种采用移动平均法来消除周期性波动并突显长期趋势的方法，将序列分为趋势-周期性和季节性，具体可如下列公式所示：

该公式通过Padding（）在输入数据的边缘添加额外的元素，通常是0值，以便在后续的池化操作中保持原始数据的形状。填充的目的是防止在进行池化操作时造成数据损失或边缘效应。AvgPool（）：平均池化操作，它将输入数据划分为多个大小相同的区域，并计算每个区域内元素的平均值，从而降低数据的维度和复杂性。这有助于减少计算量和内存需求，同时还可以提高模型的泛化能力。

然而，在现实世界的数据中，复杂的周期性模式和趋势成分通常是紧密相连的，这使得使用固定窗口的平均池化方法难以提取趋势。为了解决这一问题，本文使用混合专家分解模块替换序列分解模块。在混合专家分解模块中，括多个不同大小的平均滤波器，用以从输入信号中分离出多种趋势成分。同时，它还包含一组与数据相关的权重，用于将这些提取出的趋势成分整合为最终的趋势,可以从以下公式了解计算方法：

在公式中，F(X)表示通过一组平均池化滤波器把输入X分离出多种趋势成分，而Softmax(L(x))代表用于融合这些提取到的趋势的权重进行归一。

相比于序列分解模块，混合专家分解模块在多方面表现出了优势。混合专家分解模块可以提取多个趋势成分，在为这些成分分配权重时根据数据动态的分配相应权重，相比于序列分解模块能提取出更为丰富和详尽的特征，从而在时间序列预测任务中帮助模型提高预测准确度。同时，混合专家分解模块在处理现实世界数据中复杂的周期性模式与趋势成分时更具灵活性，能更好地应对复杂和多尺度的趋势。

2.3 基于Autoformer的PM-former框架设计

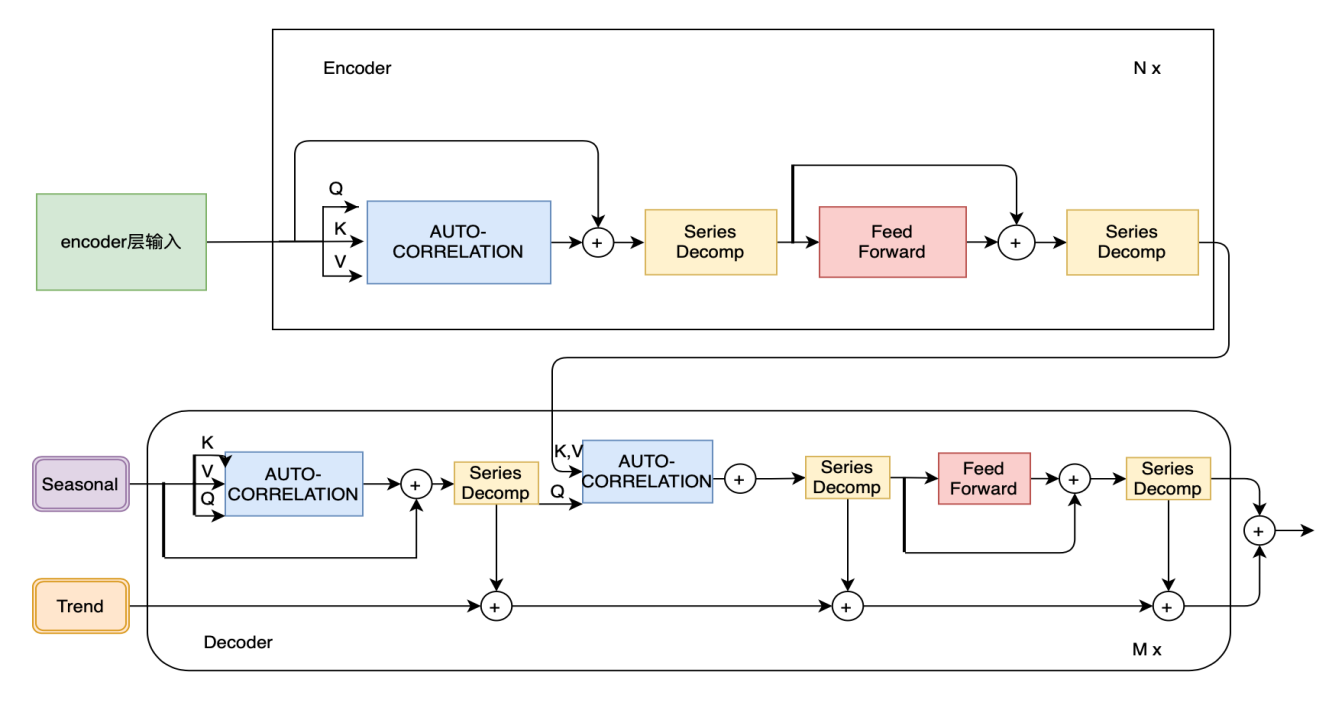


图2.7 Autoformer模型结构

Autoformer是一种用于处理长期时间序列预测任务的深度学习模型（如图2.7），提出了一种区别于自注意力机制的自相关机制，同时与自注意力机制（多头注意力

机制）相结合。在解决长期预测问题时，Autoformer通过捕捉长期依赖关系和复杂时间模式来提高预测的准确性。Autoformer的采用了一种深度分解结构，包含自相关机制，序列分解以及Ttransformer的经典架构：解码器和编码器。通过深度分解结构以及自相关机制，解决长期预测任务中的两大难题：复杂的时间模式，计算效率和信息利用的瓶颈。本文模型架构以Autoformer架构为基础，下面将从编码器，解码器，自相关机制三方面阐述Autoformer模型架构。

2.3.1 **编码器**

在Autoformer模型中，含有多层编码器，每层编码器都从上一层编码器的输出中捕捉更高级别的特征和信息。Autoformer中编码器的作用与传统Transformer模型中的编码器的作用类似，但在捕捉长期时间序列依赖关系和复杂模式方面进行了改进。Autoformer中的编码器包含自相关机制（详见2.3.3），序列分解模块（详见2.2）以及前馈神经网络。

在编码器中，原始时间序列数据被转换为模型可处理的内部表征形式，并将输入映射到高维空间中，从而使得Autoformer模型能够学习并捕获数据的抽象特性和隐藏结构。Autoformer中的编码器通过使用自相关机制和序列分解方法，进一步增强对长期依赖和复杂时间模式的捕捉能力，并通过前馈神经网络来捕获输入序列中的全局依赖以及非线性关系。编码器的输出是一个编码的内部表示，这个表示捕捉了输入时间序列的全局信息和长期依赖。由于时序数据的季节性起重要作用，在该模型中编码器通过序列分解模块消除了长期趋势循环部分，专注于季节性建模。同时由于编码器的输出包含过去的季节性信息，其被解码器的自相关模块所利用，从而使得解码器的预测结果更为准确。在PM-former中，本文将Autoformer编码器中的序列分解模块替换为基于MOE设计的专家混合分解模块。

2.3.2 **解码器**

在Autoformer模型中，含有多层解码器，一般情况下，层数越多，解码器能捕捉到越高层次的抽象信息。Autoformer中解码器的作用与传统Transformer模型中的解码器的作用类似，解码器基于编码器产生的内部表示，逐步生成预测序列，通常是未来n个时间步的预测值。与编码器一样，解码器也包含三项：自相关机制，序列分解模块，前馈神经网络；每一层解码器都含有内部自相关和编码器-解码器自相关（即编码器的输出包含过去的季节性信息被解码器的自相关模块所利用），这可以分别用于优化预测并充分利用过去的季节性信息。同时解码器的输入分为两部分：季节项输入和趋势项输入，因此解码器包括上下两部分：趋势性成分的积累结构和季节性成分的自相关机制。在PM-former中，本文将Autoformer解码器中的序列分解模块替换为基于MOE设计的专家混合分解模块。

2.3.3 **基于周期依赖关系的自相关机制**

自相关机制是Autoformer模型的主要创新点（如图2.8），可以无缝取代自注意力机制。自相关机制采用了多头版本，每层的自相关机制主要由两部分构成：基于周期的依赖关系和时间延迟聚合。自相关机制通过利用时间序列数据的周期依赖关系来发现长期依赖关系，根据随机过程理论[35]，对于离散的时间过程，通过自相关系数来反映该过程{Yt}本身与它滞后序列{Yt-τ}的相关性，通过比较自相关系数选出最优的k个周期，然后可以通过自相关系数的归一化处理进行加权处理。自相关系数计算如下：

自相关机制通过基于周期的依赖关系选择出的时延τk的大小来滚动序列，将时间序列前面长度为时延τk的大小的序列移出补充到该时间序列的后面，构成一个新的时间序列，并根据该时延序列与原时间序列之间的自相关系数，用softmax标准化置信度来聚合子序列。自相关机制由以往的点注意转换为现在的序列注意，提高了

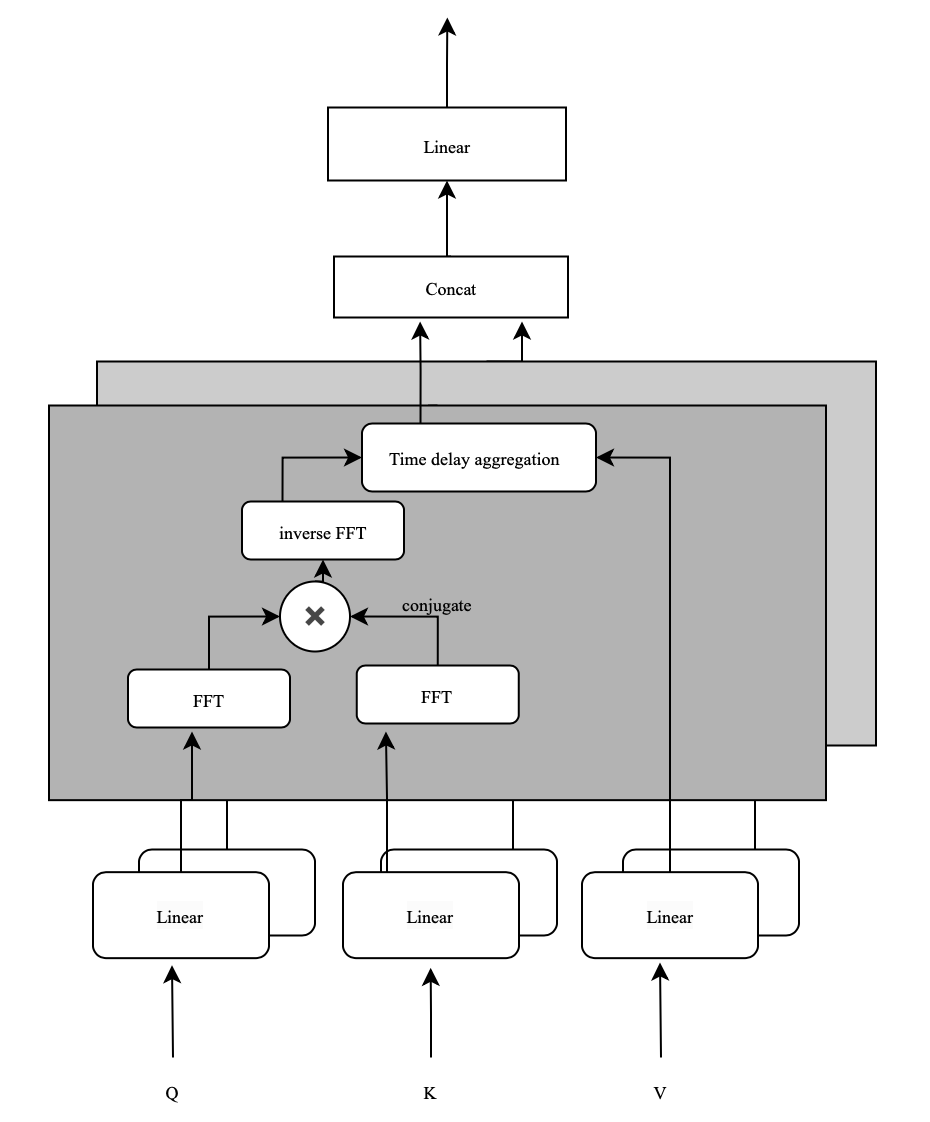


图2.8 自相关机制

信息的利用率，从而提高模型在长期预测问题中的预测准确性。自相关机制可由以下公式表示：

自相关机制相较于自注意力机制具有以下优势：

（1）发现周期性依赖关系：自相关机制通过计算序列的自相关来发现基于周期的依赖关系，这有助于捕获时间序列数据中的周期性模式。而自注意力机制主要关注点对点之间的相互关系，可能无法充分挖掘出这种周期性依赖关系。

（2）时间延迟聚合：自相关机制通过时间延迟聚合将相似的子序列对齐并聚合，这使得相同相位位置的周期性子序列得以有效整合。相比之下，自注意力机制采用点对点的点积聚合，可能无法充分捕捉这种周期性关联。

（3）更高效的计算：自相关机制仅关注选定的k个周期长度τ1.....τk，这可以减少计算量，提高模型的计算效率。而自注意力机制需要计算序列中所有元素之间的关系，计算复杂度相对较高。

（4）更好的泛化能力：自相关机制捕捉到的周期性依赖关系和时间延迟聚合有助于模型在面对新任务和不同领域的时间序列数据时，更好地泛化和适应。而自注意力机制可能在捕捉这类周期性关系方面存在局限。

自相关机制可以无缝替换自注意力机制，这意味着，在不改变整体模型结构的前提下，可以将自相关机制与自注意力机制进行替换。本文所提出的PM-former模型以自相关机制为核心挖掘时序数据的周期性依赖关系，提高了信息利用率，使得PM-former模型在长期预测中有出色的表现。

2.4 预处理模块

特征工程是机器学习和数据挖掘领域中的一个重要环节，它涉及到从原始数据中提取、选择、构建和转换特征，以便更好地表示数据，并提高模型的性能。通过之前介绍的序列分解和基于周期的依赖关系我们可以了解到，电力数据的季节性和趋势周期性很强。因此，我们可以根据日期对电力数据执行特征提取操作，从包含日期信息的数据集中抽取月份信息，然后将月份信息保存到一个新的列表中。最后，将这个月份列表作为一个新列添加到原始数据表格中。由于不同月份的用电差异是存在的，将月份信息作为一个特征可以帮助模型更好的进行预测。

2.5 本章小结

本章主要介绍了本文的模型设计，即对Autoformer模型进行改进和优化。通过分小节介绍相关模型和机制的方式，介绍了本文模型的核心组件的理论知识以及改进模块的优势和作用，第一部分介绍了多头注意力机制，第二部分介绍了时间序列分解方法以及采用混合专家分解模块的优势，第三节对本文模型的核心Autoformer进行了核心组件的介绍，第四节简述了添加的预处理模块的作用。总的来说，本章对本文的模型结构和模型改进进行了较为详细的描述。

# 实验结果与分析

3.1 数据集描述及可视化

本文在实验中所采用的数据集是基于开源数据集的自制数据集，来自于加州大学欧文分校机器学习网站里的单个家庭用电数据集和清华大学机器学习的电力负荷数据集，用处理过后的单个家庭用电数据集替换电力负荷数据集中缺失值较多的用户列，从而构成了一个涵盖321个用户，跨度三年的每小时电力负荷（单位时间内消耗的总电能，单位瓦时）的数据集，图3.1（纵轴单位为瓦时，横轴单位为天）是对数据集中某个用户的电力数据进行可视化展示（按时间顺序取最后200个数据点）。

其中在家庭用电数据集中，计数频率是1分钟，同时该数据集中存在数据缺失的问题，在进行数据处理时，考虑到短时间内的用电数据不会发生大范围波动，对于短时间内的缺失值采用了取缺失长度前后数值的平均值进行填充的方法，对于长时间内的缺失值，考虑到家庭用电的周期性，采用取对应缺失长度中数据点前后一周对应时间点的用电数据的平均值进行填充，最后通过合并操作使其汇总为每小时的数据。

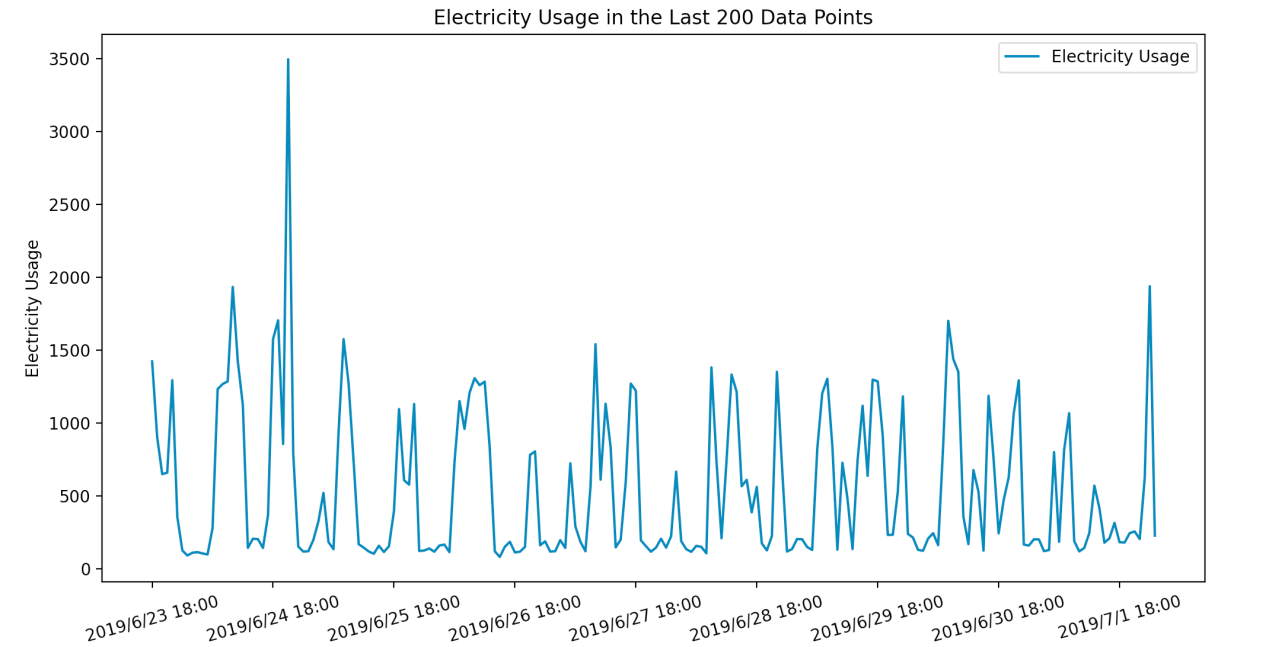


图3.1 数据可视化

在对数据集的划分中，本文采取了按照时间顺序进行索引，以训练集：校验集：测试集=7:1:2的比例进行划分的方法。

3.2 评价指标选择及阐述

本文的选择了MSE（均方误差）和MAE（平均绝对误差）作为评价指标，二者常被用来衡量时间序列预测任务中模型预测的准确性。MSE（均方误差）：首先计算预测长度内每个数据点的预测值与实际值之间误差，然后对每个误差平方后求平均值。MAE(平均绝对误差)：首先计算预测长度内每个数据点的预测值与实际值之间误差，然后对每个误差取绝对值后求平均值。

在以下公式中，样本数量记为n，实际值记为yi，预测值记为yi'。当MSE 和MAE的值越小时，说明该模型预测误差越小，其性能越好。但不难从公式看出，MSE对于较大的误差值具有较高的敏感性，因为较大的误差被平方后产生了放大效应。公式表示如下：

3.3 实验细节与参数设置

本文的实验程序采用Python编写，使用Pytorch搭建模型，使用CPU进行实验，CPU型号为“1.6 GHz双核Intel Core i5”。本文所有实验的优化器为ADAM，学习率设置为0.0001，训练时输入数据的批量大小设置为32。将自相关机制中的时延聚合超参数c设置为3，Autoformer和PM-former模型均包含2个编码器层和1个解码器层，基于多头注意力机制的自相关机制头数设置为8。以下是实验中的相关参数设置(以预测长度96为例，即“--pred\_len 96”)。

表3.1 相关参数表

|  |  |
| --- | --- |
| 模型名称 | 相关参数设置 |
| PM-former | --is\_training 1 \--root\_path ./dataset/electricity/ \--data\_path electricity.csv \--model\_id ECL\_96\_96 \ --model PM-former \ --data custom \--features M \--seq\_len 96 \ --label\_len 48 \ --pred\_len 96 \ --e\_layers 2 \ --d\_layers 1 \ --factor 3 \ --enc\_in 322 \--dec\_in 322 \--c\_out 322 \--des 'Exp' \--itr 1 |
| Autoformer | --is\_training 1 \--root\_path ./dataset/electricity/ \--data\_path electricity.csv \--model\_id ECL\_96\_96 \ --model Autoformer \ --data custom \--features M \--seq\_len 96 \ --label\_len 48 \ --pred\_len 96 \ --e\_layers 2 \ --d\_layers 1 \ --factor 3 \ --enc\_in 321 \--dec\_in 321 \--c\_out 321 \--des 'Exp' \--itr 1 |
| Transformer | --is\_training 1 \--root\_path ./dataset/electricity/ \--data\_path electricity.csv \--model\_id ECL\_96\_96 \ --model Transformer \ --data custom \--features S \--seq\_len 96 \ --label\_len 48 \ --pred\_len 96 \ --e\_layers 2 \ --d\_layers 1 \ --factor 3 \ --enc\_in 1 \--dec\_in 1 \--c\_out 1 \--des 'Exp' \--itr 1 |
| Informer | --is\_training 1 \--root\_path ./dataset/electricity/ \--data\_path electricity.csv \--model\_id ECL\_96\_96 \ --model Informer \ --data custom \--features S \--seq\_len 96 \ --label\_len 48 \ --pred\_len 96 \ --e\_layers 2 \ --d\_layers 1 \ --factor 3 \ --enc\_in 1 \--dec\_in 1 \--c\_out 1 \--des 'Exp' \--itr 1 |
| Reformer | --is\_training 1 \--root\_path ./dataset/electricity/ \--data\_path electricity.csv \--model\_id ECL\_96\_96 \ --model Reformer \ --data custom \--features S \--seq\_len 96 \ --label\_len 48 \ --pred\_len 96 \ --e\_layers 2 \ --d\_layers 1 \ --factor 3 \ --enc\_in 1 \--dec\_in 1 \--c\_out 1 \--des 'Exp' \--itr 1 |

3.4 对比实验

本文的对比实验模型选择了Transformer[10]以及当下较为火热的与其相关的改进模型如Informer[36]，Reformer[37]；同时与原模型进行了对比。因为本文旨在解决长期预测任务，因此本文对比实验以96为输入长度，以输出长度为96，192，336分别进行预测，预测长度分别是输入长度的1倍，2倍和3.5倍。本文的两个评价指标越低，说明模型的预测效果越好。三种预测长度的预测结果分别如表1所示(记本文模型名称为PM-former，最优结果已加粗显示)。

表3.2 实验数据表

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 预测长度 | 评价指标 | PM-former | Autoformer | Informer | Transformer | Reformer |
| 96 | MSE | **0.1941** | 0.2092 | 0.2581 | 0.2904 | 0.2933 |
| MAE | **0.3004** | 0.3189 | 0.3646 | 0.3863 | 0.3996 |
| 192 | MSE | **0.2216** | 0.2339 | 0.2765 | 0.3567 | 0.3231 |
| MAE | **0.3312** | 0.3418 | 0.3798 | 0.4354 | 0.4191 |
| 336 | MSE | **0.2541** | 0.2711 | 0.3283 | 0.4541 | 0.3616 |
| MAE | **0.3518** | 0.3739 | 0.4191 | 0.4946 | 0.4284 |

由表中的实验数据可以看出，本文在Autoformer基础上的改进模型PM-former在预测长度96，192和336上均表现出了优势。在输入长度为96，预测长度为96时：PM-former比之Autoformer在均方误差（MSE）上实现了7.2%的提升，在平均绝对误差（MAE）上实现了5.8%的提升；在输入长度为96，预测长度为192时：PM-former比之Autoformer在均方误差（MSE）上实现了5.3%的提升，在平均绝对误差（MAE）上实现了3.1%的提升；在输入长度为96，预测长度为336时：PM-former相比较于Autoformer在均方误差（MSE）上实现了6.3%的提升，在平均绝对误差（MAE）上实现了5.9%的提升。

由实验结果可以看出，本文提出的模型在长期预测任务中体现出了一定的优势性，通过预处理模块和混合专家分解模块实现了对Autoformer模型的改进，能够使得模型学习更多的特征，同时能更好地应对时序数据的复杂和多尺度的趋势。对比Autoformer，本文模型在两个评价指标下均实现了提升，同时相比于基于Transformer的改进模型以及Transformer模型本身，都实现了全面的超越。

为了更直观的反映出模型的预测准确度，本文用户的电力预测状况进行了可视化操作，因为本文中的实验模型采用的是多元预测，即分别对321个用户分别进行预测并求取MSE和MAE的平均值，且考虑到本文主要研究模型长期预测任务中的预测准确度，故在输入长度为96，预测长度为336的条件下以某位用户的电力预测情况进行可视化示例，如下图所示（标准化后无单位）。

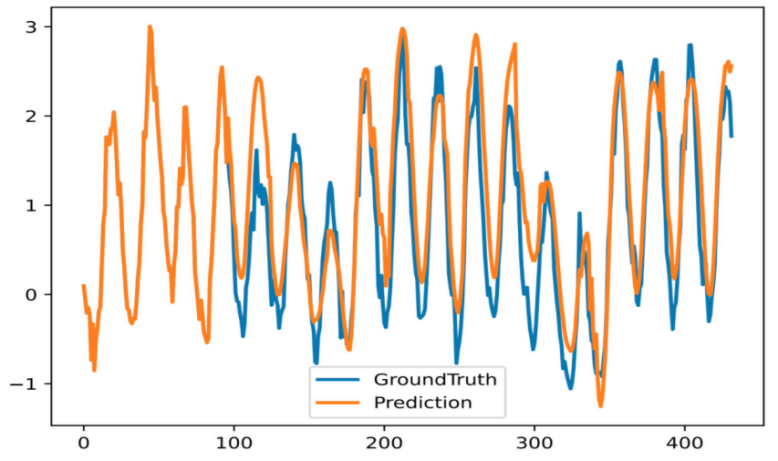


图3.2 PM-former-pl336

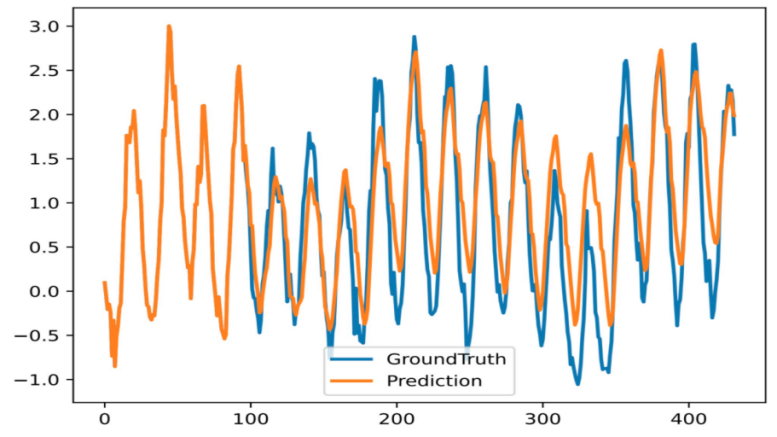


图3.3 Autoformer-pl336

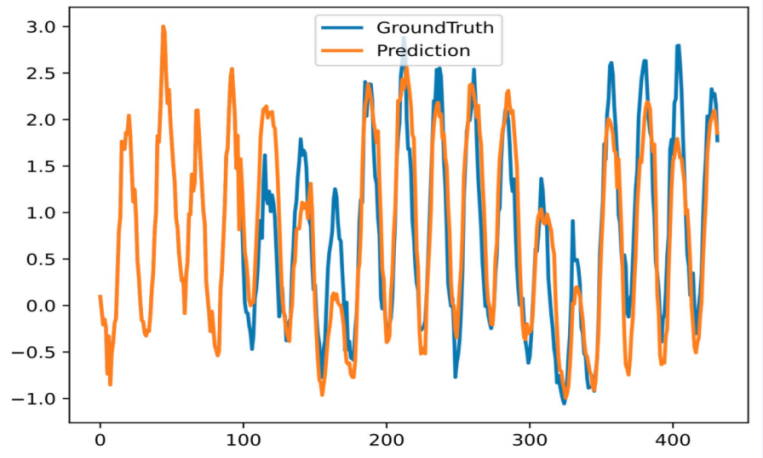


图3.4 Informer-pl336

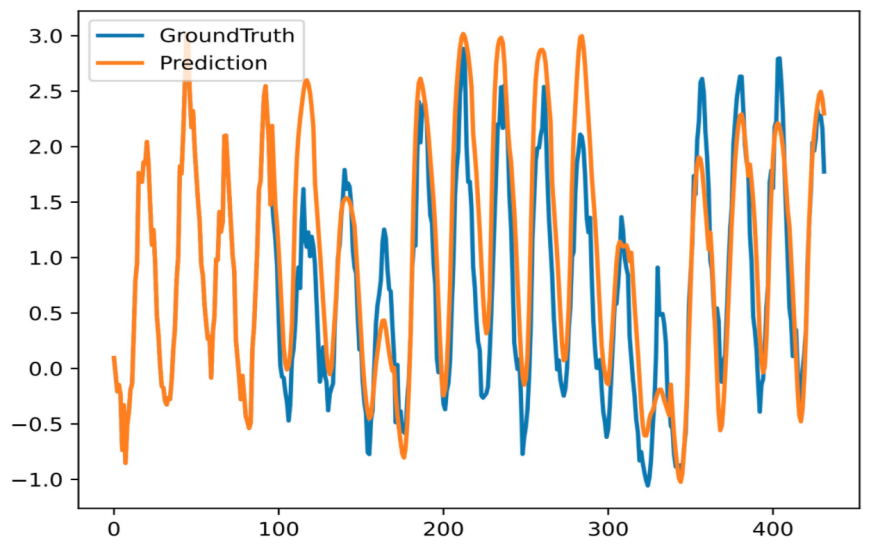


图3.5 Reformer-pl336

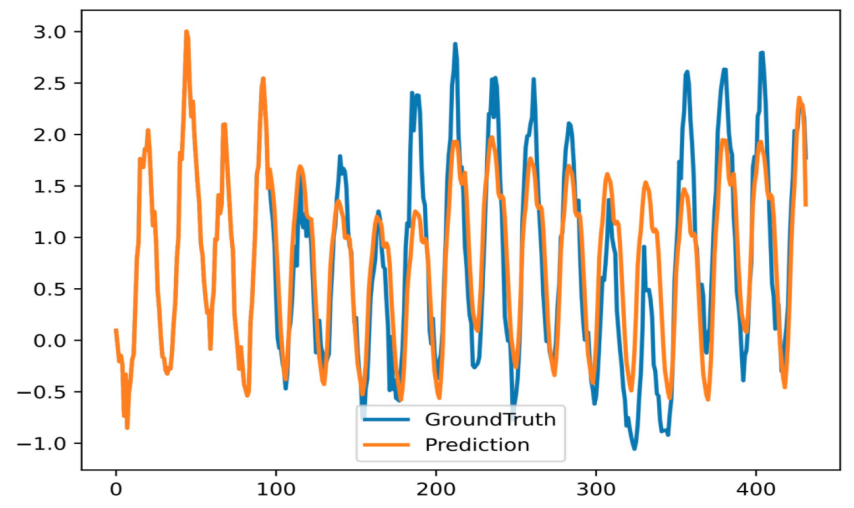


图3.6 Transformer-pl336

关于图像的纵坐标，因为在代码层面对输入的数据进行了标准化处理（均值为零，标准差为一），故存在负值和零值，考虑到可视化主要展示预测值与真实值的差异（曲线贴合程度），并未进行反标准化操作。

为了更直观的反映出模型的预测准确度，本文用户的电力预测状况进行了可视化操作（因为本文中的实验模型采用的是多元预测，即分别对321个用户分别进行预测并求取MSE和MAE的平均值，故在输入长度为96，预测长度为336的条件下以某位用户的电力预测情况进行可视化示例），如上图所示。

不难从实验结果看出，基于Ttransformer改进的Informer模型和Reformer模型在输入长度为96，预测长度为96，192和336时取得了较于Transformer更优的结果。总的来说，本文所提出的改进模型在时间序列预测任务的长期预测问题中体现出了强大的优势。

3.5 消融实验

为了验证两个模块的有效性，本文运用控制变量的思想进行了消融实验。为了便于理解，用以下代号代表相应的模型，对应实验结果如表3.3所示（最优结果已加粗显示）。

表3.3 消融实验数据

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 预测长度 | 评价指标 | PM-former | Autoformer | P-former | M-former |
| 96 | MSE | **0.1941** | 0.2092 | 0.2026 | 0.1982 |
| MAE | **0.3004** | 0.3189 | 0.3126 | 0.3054 |
| 192 | MSE | **0.2216** | 0.2339 | 0.2301 | 0.2265 |
| MAE | **0.3312** | 0.3418 | 0.3396 | 0.3347 |
| 336 | MSE | **0.2541** | 0.2711 | 0.2683 | 0.2581 |
| MAE | **0.3518** | 0.3739 | 0.3692 | 0.3542 |

PM-former：代表本文提出的基于Autoformer的改进模型。

P-former：代表在Autoformer模型的基础上添加基于特征工程的预处理模块后的模型。

M-former：代表将Autoformer模型中的序列分解模块替换为混合专家分解模块后的模型。

消融实验表明，基于特征工程的数据预处理模块和基于MOE的混合专家分解模块对Autoformer模型的预测准确度均有提升，其中使用混合专家分解模块替换序列分解模块对于Autoformer模型的测准确度的提升较大。因为混合专家分解模块在处理现实世界数据中复杂的周期性模式与趋势成分时更具灵活性，能更好地应对复杂和多尺度的趋势。

3.6 本章小节

本章首先对数据集进行了描述，可视化分析以及划分训练集的方式。然后选用当下较为火热的深度学习模型进行对比实验，并进行了部分结果的可视化，使不同模型的预测值与真实值对比更为直观，从结果上看，对比其他模型，本文所提出的模型在长期预测问题上表现出了较好的预测准确度。最后通过消融实验来分析每个改进模块是否起到优化预测准确度的作用，以及该模块改进作用的大小。

# 总结与展望

4.1 本文工作及研究总结

时间序列预测在现实生活中应用广泛，为资源最优分配，交通流预测，天气预测，能源消耗规划等提供了助力。需求侧响应[38]是智能电网重要组成部分，对家庭用户的用电数据进行预测是其基础，通过对家庭用户用电数据的长期预测，可以帮助电力公司和政府制定相应的电力规划和应急调度策略，以确保电力系统的稳定运行和满足用户的用电需求。本文以用户用电数据集为基础，通过序列分解技术分析用电数据的季节性和趋势-周期性，以Autoformer模型为基础，在其基础上进行模型改进工作，并分析改进模型在时间序列预测中的长期预测任务的预测准确度。与主流模型相比，本文所提出的改进模型在长期预测任务中取得了良好的表现，预测准确性有一定的提升。本文的创新点如下：

A.增加了基于特征工程的预处理模块，从日期中提取出月份信息作为新的特征列加入数据集中，是模型学习更多的特征，建立起用电数据与月份信息之间的联系。月份信息中通常蕴涵着季节性（气候）信息，季节（气候）也是影响家庭用电的因素，通过可以帮助模型捕捉这种季节性变化。通过该预处理操作，可以使得本文模型学习更多的特征，从而提高预测的精确度。

B.将序列分解模块替换为混合专家分解模块，混合专家分解模块通过一组不同大小的平均滤波器提取出多种趋势，并通过权重分配合成总趋势，在这种模式下，可以从输入的时间序列中提取出更为丰富和详尽的特征。

4.2 后续研究分析与展望

本文通过对Autoformer模型的改进，实现了在长期预测问题上的预测准确性提升，但是仍有一定的改进空间。

A.在基于特征工程的预处理模块中，只提取了月份信息，从消融实验结果上看，该模块对原模型预测准确性的提升效果有限，考虑到休息日与工作日家庭用电高峰不同，因此可以从日期信息中提取出工作日与休息日的相关信息，形成新的特征列，使得模型学习学习更多的特征，进而提高预测准确性。

B.在数据集的收集中，只包含了电力数据信息，没有考虑外部环境信息，比如气温和天气等信息，这些信息一定情况下也与用电信息构成联系，缺少这些信息会使得模型在预测时准确度受限。

C.本模型的核心架构是以Autoformer为基础的，其核心组件自相关机制以基于周期性的依赖关系为基础，因此，在对周期性不强以及噪声较多的时序数据进行预测时，模型预测效果可能会受限。

**参考文献**

1. Rafał Weron. Electricity price forecasting: A review of the state-of-the-art with a look into the future[J]. International Journal of Forecasting,2014,30(4).
2. Li L , Meinrenken C J , Modi V , et al. Short-term apartment-level load forecasting using a modified neural network with selected auto-regressive features[J]. Applied Energy, 2021, 287(147):116509.
3. Henrique Steinherz Hippert,Carlos Eduardo Pedreira,Reinaldo Castro Souza. Neural networks for short-term load forecasting: a review and evaluation[J]. IEEE Transactions on Power Systems: A Publication of the Power Engineering Society,2001,16(1).
4. Ru Chuanhong,Qin Jiansong,Lu Peng,Zhou Dongcheng,Zhang Yan. Research on Short-term Time Series Forecasting Method Based on Feature Fusion of Electric Power[J]. Journal of Physics: Conference Series,2023,2449(1).
5. Huang Xiaoqiao,Li Qiong,Tai Yonghang,Chen Zaiqing,Liu Jun,Shi Junsheng,Liu Wuming. Time series forecasting for hourly photovoltaic power using conditional generative adversarial network and Bi-LSTM[J]. Energy,2022,246.
6. Hussein Sharadga,Shima Hajimirza,Robert S. Balog. Time series forecasting of solar power generation for large-scale photovoltaic plants[J]. Renewable Energy,2020,150(C).
7. Wu Han,Liang Yan,Heng Jiani. Pulse-diagnosis-inspired multi-feature extraction deep network for short-term electricity load forecasting[J]. Applied Energy,2023,339.
8. Wu H,Xu J ,Wang J,et al. Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting[J]. 2021.
9. Li Zihan,Zhang Qingchuan,Dong Wei,Liu Yingjie,Wei Siwei,Zuo Min. FEDformer-Based Paddy Quality Assessment Model Affected by Toxin Change in Different Storage Environments.[J]. Foods (Basel, Switzerland),2023,12(8).
10. Sufeng Duan,Hai Zhao. Attention Is All You Need for Chinese Word Segmentation.[J]. CoRR,2019,abs/1910.14537.
11. 毛远宏,孙琛琛,徐鲁豫等.基于深度学习的时间序列预测方法综述[J/OL].微电子学与计算机,2023(04):8-17[2023-05-04].
12. Yule G. Udny.Why do we Sometimes get Nonsense-Correlations between Time-Series?--A Study in Sampling and the Nature of Time-Series[J]. Journal of the Royal Statistical Society,1926,89(1).
13. D. J. Bartholomew. Time Series Analysis Forecasting and Control[J]. Journal of the Operational Research Society,1971,22(2).
14. Contreras J,Espinola R, Nogales F J, et al. ARIMA Models to Predict Next-Day Electricity Prices[J]. IEEE Transactions on Power Systems, 2003, 18(3):1014-1020.
15. 孙玉环. 基于SARIMA模型的SARS对中国入境旅游收入影响的定量分析[J]. 旅游科学, 2006, 20(1):4.
16. Bollerslev T . Generalized autoregressive conditional heteroskedasticity[J]. Eeri Research Paper, 1986, 31(3):307-327.
17. 李飞宏,肖迎群.基于STL-LSTM-TCN模型的短期负荷预测方法[J].电子设计工程,2023,31(07):47-51+56.DOI:10.14022/j.issn1674-6236.2023.07.010.
18. 赵滨滨,王莹,王彬等.基于ARIMA时间序列的分布式光伏系统输出功率预测方法研究[J].可再生能源,2019,37(06):820-823.DOI:10.13941/j.cnki.21-1469/tk.2019.06.005.
19. 王斌.基于时间序列ARIMA模型的电力负荷短期预测分析[J].石河子科技,2019,No.245(03):43-47.
20. Vapnik V N . The Nature of Statistical Learning Theory[J]. Springer, 1995.
21. Jeffrey L. Elman. Finding structure in time[J]. Cognitive Science,1990,14(2).
22. Hochreiter S , Schmidhuber J . Long Short-term Memory[J]. 1997.
23. Jerome H. Friedman. Greedy Function Approximation: A Gradient Boosting Machine[J]. The Annals of Statistics,2001,29(5).
24. Tianqi Chen,Carlos Guestrin. XGBoost: A Scalable Tree Boosting System.[J]. CoRR,2016,abs/1603.02754.
25. 朱乔木,李弘毅,王子琪,陈金富,王博.基于长短期记忆网络的风电场发电功率超短期预测[J].电网技术,2017,41(12):3797-3802.DOI:10.13335/j.1000-3673.pst.2017.1657.
26. 李少杰. 基于FFT-CNN-LSTM的短期电力负荷预测研究[D].南昌大学,2021.DOI:10.27232/d.cnki.gnchu.2021.003155.
27. 陈振宇,刘金波,李晨,季晓慧,李大鹏,黄运豪,狄方春,高兴宇,徐立中.基于LSTM与XGBoost组合模型的超短期电力负荷预测[J].电网技术,2020,44(02):614-620.DOI:10.13335/j.1000-3673.pst.2019.1566.
28. Colin Lea,Michael D. Flynn,René Vidal,Austin Reiter,Gregory D. Hager. Temporal Convolutional Networks for Action Segmentation and Detection.[J]. CoRR,2016,abs/1611.05267.
29. Kyunghyun Cho,Bart van Merrienboer,Çaglar Gülçehre,Fethi Bougares,Holger Schwenk,Yoshua Bengio. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation.[J]. CoRR,2014,abs/1406.1078.
30. Ilya Sutskever,Oriol Vinyals,Quoc V. Le. Sequence to Sequence Learning with Neural Networks.[J]. CoRR,2014,abs/1409.3215.
31. 李磊,林珊,贾颉辉.基于TCN-Attention神经网络的短期负荷预测[J].电力信息与通信技术,2023,21(03):10-16.DOI:10.16543/j.2095-641x.electric.power.ict.2023.03.02.
32. 徐武,刘洋,沈智方等.基于改进局部自注意力机制的VMD-GRU模型短期风电功率预测[J].电网与清洁能源,2023,39(03):83-92.
33. 刘洪笑,向勉,周丙涛等.基于Informer的长序列时间序列电力负荷预测[J].湖北民族大学学报(自然科学版),2021,39(03):326-331.DOI:10.13501/j.cnki.42-1908/n.2021.09.015.
34. 李清格,杨小冈,卢瑞涛等.计算机视觉中的Transformer发展综述[J].小型微型计算机系统,2023,44(04):850-861.DOI:10.20009/j.cnki.21-1106/TP.2022-0504.
35. 杨晓虎,李丽辉. 随机过程一般方法理论[C]//中国运筹学会不确定系统分会.第十届中国不确定系统年会、第十四届中国青年信息与管理学者大会论文集.Global-Link Publisher,2012:86-91.
36. Gong Mingju,Zhao Yin,Sun Jiawang,Han Cuitian,Sun Guannan,Yan Bo. Load forecasting of district heating system based on Informer[J]. Energy,2022,253.
37. Nikita Kitaev, Lukasz Kaiser, and Anselm Levskaya. Reformer: The efficient transformer. In International Conference on Learning Representations, 2020.
38. 杨晓冉. 需求侧响应将成新型电力系统特征[N]. 中国能源报,2022-02-21(012).DOI:10.28693/n.cnki.nshca.2022.000332.

**致谢**

时光荏苒，四年悄然已过，回首大学生活，有欢笑有遗憾，感谢合肥工业大学对我的培养，“厚德，笃实，崇学，尚新”的校训我会铭记终身。

首先我要感谢我的指导老师：台建玮老师，在完成毕业设计的过程中，从题目选择，到模型的创新改进，台老师都给予我了很多建设性意见，并在我受挫时提供精神支持，本篇论文的完成离不开台老师的悉心指导，在此再次表达对台老师的感谢之情！

“爱子心无尽”，在此感谢我的父母，二十年如一日的对我无微不至的照顾，他们是我最坚实的后盾，在日后的生活中，我会不负期望，奋勇向前，迎难而上，砥砺前行！

同时，我也要感谢我的朋友和室友，大学四年中他们给予了我很多鼓励与支持，也正是与他们一起，让我感受到青春的美好，愿友谊永存！

至此落笔，愿直挂云帆济沧海！

作者：孟熙冉

2023年 05 月 20日