**第十届“泰迪杯”数据挖掘挑战赛**

——B题 电力系统负荷预测分析

目录

[摘要 3](#_Toc102074028)

[Abstract 3](#_Toc102074029)

[一、问题分析 4](#_Toc102074030)

[1.1 问题一分析 4](#_Toc102074031)

[1.2 问题二分析 4](#_Toc102074032)

[二、地方负荷的中短期预测分析 4](#_Toc102074033)

[2.1 数据预处理 4](#_Toc102074034)

[2.2 模型选择 7](#_Toc102074043)

[2.1方案一 回归预测 7](#_Toc102074044)

[2.2 方案二 ARIMA模型 9](#_Toc102074045)

[2.3 方案三 使用LSTM进行预测 13](#_Toc102074046)

[2.4 方案四 多元prophet模型 14](#_Toc102074047)

[三、行业负荷的中期预测分析 16](#_Toc102074048)

[3.1 分析各行业用电负荷突变的时间、量级和可能的原因 16](#_Toc102074049)

[3.2 各行业未来3个月日负荷预测 20](#_Toc102074054)

[3.2.1 数据预处理 21](#_Toc102074055)

[3.2.2 ARIMA 23](#_Toc102074056)

[3.3.2 LSTM 27](#_Toc102074057)

[3.3 对各行业的建议 33](#_Toc102074058)

[参考文献 33](#_Toc102074059)

# 摘要

电力系统负荷预测，是指充分考虑历史的系统负荷、经济状况、气象条件和社会时间等因素的影响，对未来一段时间的系统负荷作为预测。复杂多变的气象条件和社会事件等不确定因素会对电力系统负荷造成一定的影响。通过设置一定的模型可以实现电力系统负荷的预测，从而可以提前对相应的电力系统负荷进行预测，并根据预测结果进行一定程度的调节，以保障正常的电力系统负荷。

在问题一中，本次研究对地区电网进行中短期预测，在这个过程中，我们采用了回归模型、ARIMA模型、LSTM模型和Prophet模型进行电力系统负荷的预测。

在问题二中，本次研究分析了不同行业用电负荷突变的原因，并采用ARIMA模型和LSTM模型对该地区各行业未来3个月的日负荷最大值和最小值进行了预测，并根据相关信息提出了具体的建议。

对于各行业各地区的用电情况，考虑复杂多变的天气情况以及各地区的经济工业情况，对各地区各行业的用电负荷进行合理的处理，以便于为电网改造、扩建提供便利情况，也可以进一步提高电力系统的经济效益和社会效益。

**关键字**：ARIMA模型、Prophet模型、长短时记忆网络（LSYTM）

# Abstract

Power system load forecasting refers to fully considering the influence of historical system load, economic conditions, meteorological conditions, social time and other factors, and forecasting the system load for a period of time in the future. Uncertain factors such as complex and changeable meteorological conditions and social events will have a certain impact on the load of power system. By setting a certain model, the power system load can be predicted, so that the corresponding power system load can be predicted in advance, and adjusted to a certain extent according to the prediction results to ensure the normal power system load.

In question 1, this study makes medium and short-term prediction for regional power grid. In this process, we use regression model, ARIMA model, LSTM model and prophet model to predict power system load.

In question 2, this study analyzes the reasons for the sudden change of power load in different industries, forecasts the maximum and minimum daily load of various industries in the region in the next three months by using ARIMA model and LSTM model, and puts forward specific suggestions according to relevant information.

For the power consumption of various industries and regions, considering the complex and changeable weather conditions and the economic and industrial conditions of various regions, reasonably deal with the power load of various industries in various regions, so as to facilitate the transformation and expansion of power grid, and further improve the economic and social benefits of power system.

**Keywords**：ARIMA model, prophet model, long-term and short-term memory network (LSTM)

# 一、问题分析

## 问题一分析

问题一中，本次研究主要是根据附件中某地区的电网负荷数据，建立对地区电网的中短期负荷预测模型，并对该地区电网未来三个月的日负荷的最大值和最小值预测结果进行预测。对该地区未来10天间隔15分钟的负荷数据进行预测，并分析其精度。对该地区电网未来3个月的日负荷的最大值和最小值进行预测，并得出达到负荷最大值和最小值的时间，并进行预测精度的分析。

## 问题二分析

问题二中，主要是对不同行业的用电负荷进行预测，分析各行业用电负荷图变的时间、量级和可能的原因。分析该地区各行业未来三个月的日负荷最大值和最小值的预测结果，并对预测精度进行分析。根据各行业的具体情况，研究国家“双碳”目标下对各行业未来用电负荷可能产生的影响。

# 二、地方负荷的中短期预测分析

在问题一中，本次研究主要是针对附件中提供的信息对地区电网进行间隔15分钟的负荷数据进行检查分析，并进行预测。

## 2.1 数据预处理

在问题一中，本次研究主要利用附件一和附件三中的数据，考虑到数据的本身会存在一些问题，在本次研究中，主要采用了清楚重复值、缺失值填充、数据规范、添加新的特征、对指定数据编码等操作对数据进行处理，以便后续的分析整理。

### 2.1.1 清除重复值

综合考虑附件一和附件三中的数据，对附件中的数据进行了进一步的清理和研究，使数据可以更好的达到研究效果。下表为数据清洗具体信息：

|  |  |
| --- | --- |
| 附件信息 | 数量 |
| 附件一原始数据量 | 128156 |
| 附件一去除重复值后数量 | 128156 |
| 附件三原始数据量 | 1345 |
| 附件三去除重复值后数据量 | 1339 |

表1：去除重复值信息

### 2.1.2 填补缺失值

在进行数据研究中，本次研究发现数据中存在部分缺失值，为了保障数据完整性，本次实验选择对数据进行填充。在这一过程中，本次实验利用asfrep是重设频率15T，使索引对齐，之后采用interpolate填充缺失值。数据填充效果如下图所示：

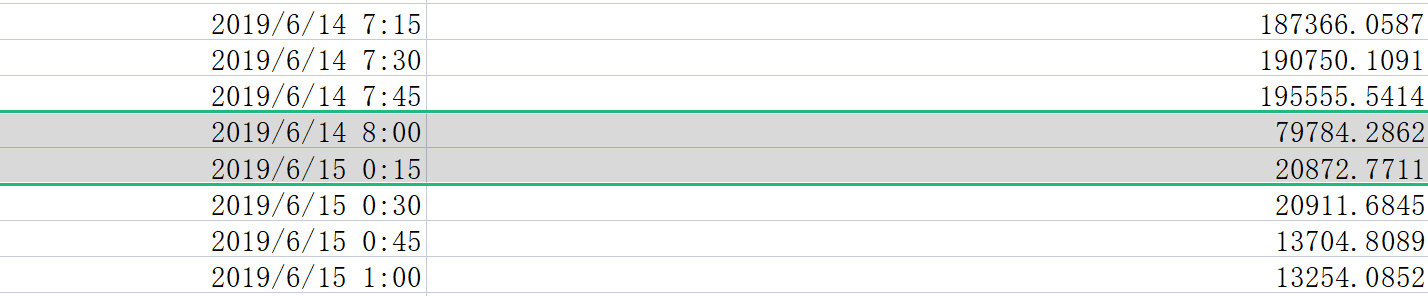


图1：缺失值填充

### 2.1.3 数据规范化

除了要对数据进行重复值清洗，填充空值，此外为了更好的实现数据效果还需要对附件三中的数据进行规范化处理。

1. 日期列处理为标准时间形式

|  |  |
| --- | --- |
| 处理前 | 处理后 |
| 2018年1月1日 | 2018-01-01 |
| 2018年1月2日 | 2018-01-02 |
| 2018年1月3日 | 2018-01-03 |
| 2018年1月4日 | 2018-01-04 |
| 2018年1月5日 | 2018-01-05 |

表2：时间列处理为标准时间形式

1. 处理温度列的特征

|  |  |  |  |
| --- | --- | --- | --- |
| 处理前 | | 处理后 | |
| 最高温度 | 最低温度 | 最高温度 | 最低温度 |
| 22℃ | 12℃ | 22 | 12 |
| 22℃ | 15℃ | 22 | 15 |
| 23℃ | 15℃ | 23 | 15 |
| 21℃ | 16℃ | 21 | 16 |
| 19℃ | 13℃ | 19 | 13 |
| 15℃ | 11℃ | 15 | 11 |

表3：处理温度列数据特征

### 2.1.4 添加新的特征

对天气情况进行分析，并添加新的特征。

|  |  |  |
| --- | --- | --- |
| 处理前 | 处理后 | |
| 天气状况 | 天气1 | 天气2 |
| 多云/多云 | 多云 | 多云 |
| 多云/多云 | 多云 | 多云 |
| 多云/阴 | 多云 | 阴 |
| 多云/小雨 | 多云 | 小雨 |
| 阴/小雨 | 阴 | 小雨 |

表4：对天气情况数据进行分析

### 2.1.5 对风力风向进行编码

对天气1，天气2，白天风力风向，夜晚风力风向进行Frequency编码，Frequency编码通过计算特征变量中每个值的出现次数来表示该特征的信息。

以白天风向为例进行描述（仅展示部分情况）

|  |  |
| --- | --- |
| 处理前白天风力风向 | 处理后白天风力风向 |
| 无持续风向<3级 | 139 |
| 北风4～5级 | 2 |
| 微风<3级 | 26 |
| 东北风3～4级 | 1 |
| 无持续风向1-2级 | 701 |

表5：风力编码

### 2.1.6 合并附件一和附件三

以时间为连接键，合并附件一和附件三，合并后效果如下图所示：

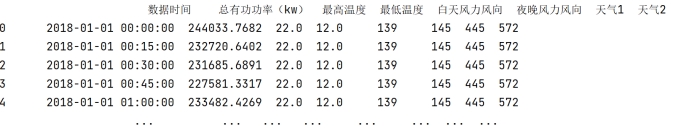


图1：附件一和附件三合并

### 2.1.7添加时间特征

基础时间周期的拆分，具体编辑信息如下图所示：

train\_data[**'年'**]=train\_data[**'数据时间'**].dt.year

train\_data[**'月'**]=train\_data[**'数据时间'**].dt.month

train\_data[**'日'**]=train\_data[**'数据时间'**].dt.day

train\_data[**'小时'**]=train\_data[**'数据时间'**].dt.hour

### 2.1.8 添加季节性特征

在数据研究的过程中，本次实验进行了添加季节性特征，下表为部分数据展示：

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 数据时间 | 一年中的第几周 | 月末 | 月初 | 季末 | 季初 | 周末 |
| 2018/1/1 0:00 | 1 | 0 | 1 | 0 | 1 | 0 |
| 2018/1/31 0:00 | 5 | 1 | 0 | 0 | 0 | 0 |
| 2018/3/31 0:00 | 13 | 1 | 0 | 1 | 0 | 1 |

表2：添加季节性特征

数据处理完毕，模型选择

## 2.2 模型选择

在模型选择中，本次研究采用多种方案进行研究以确定最终研究方案

### 2.1方案一 回归预测

#### 2.1.1 模型介绍

1. 使用树模型来选择特征，GBDT是基于CART树的集成学习

Boosting模型采用另一种形式，把基础模型组合起来——串联。这类模型的思想是，既然一个基础模型可以做出不完美的预测，那么我们可以用第二的基础模型，把“不完美的部分”补上。我们可以使用很多的基础模型，不断地对“不完美的部分”进行完善，以得到效果足够好的集成模型。Boosting的策略非常多，以GBDT为例，它会用第K个CART拟合前k-1个CART留下的残差，从而不断的缩小整个模型的误差

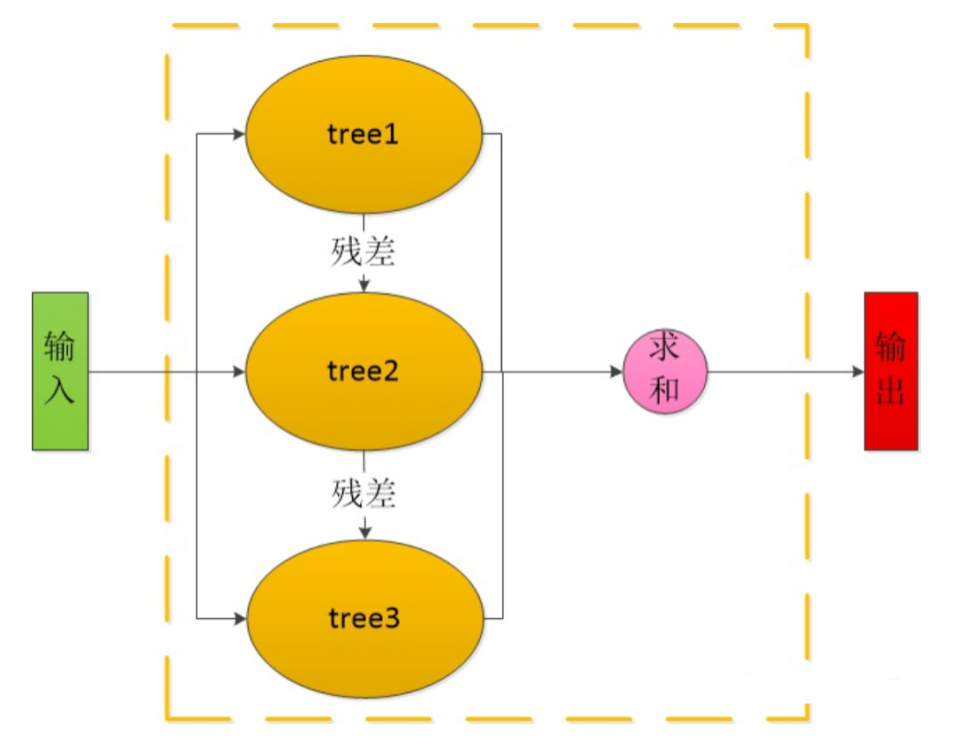


图1：CART树模型

我们清楚CART树是使用平方误差最小化标准拆分特征，会找到最优切分变量和最优切分点，我们使用这种特性来衡量特征的重要度，筛选出主要特征。

2. 使用lightGBM预测

LightGBM（Light Gradient Boosting Machine）是一个实现GBDT算法的框架，支持高效率的并行训练，并且具有更快的训练速度、更低的内存消耗、更好的准确率、支持分布式可以快速处理海量数据等优点。

#### 2.1.2模型检验

取训练集前百分之70的数据进行训练，后30%进行预测，预测结果如下

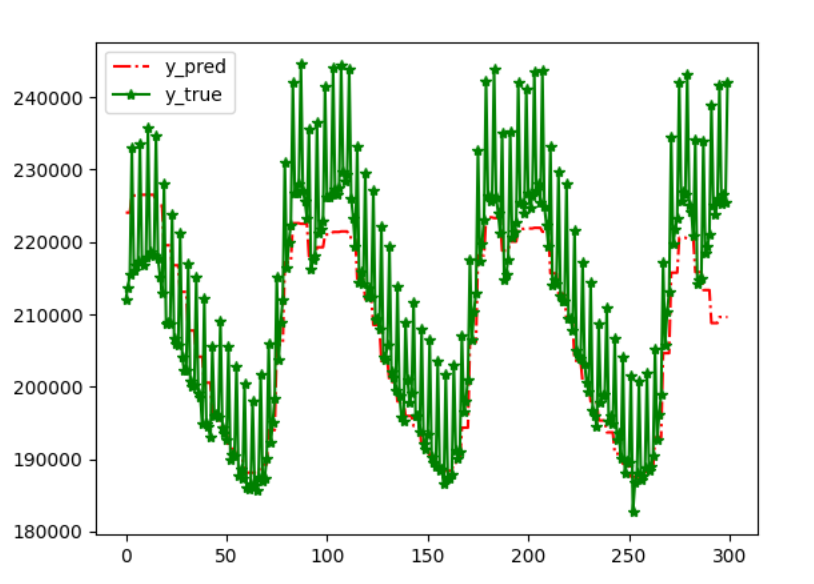
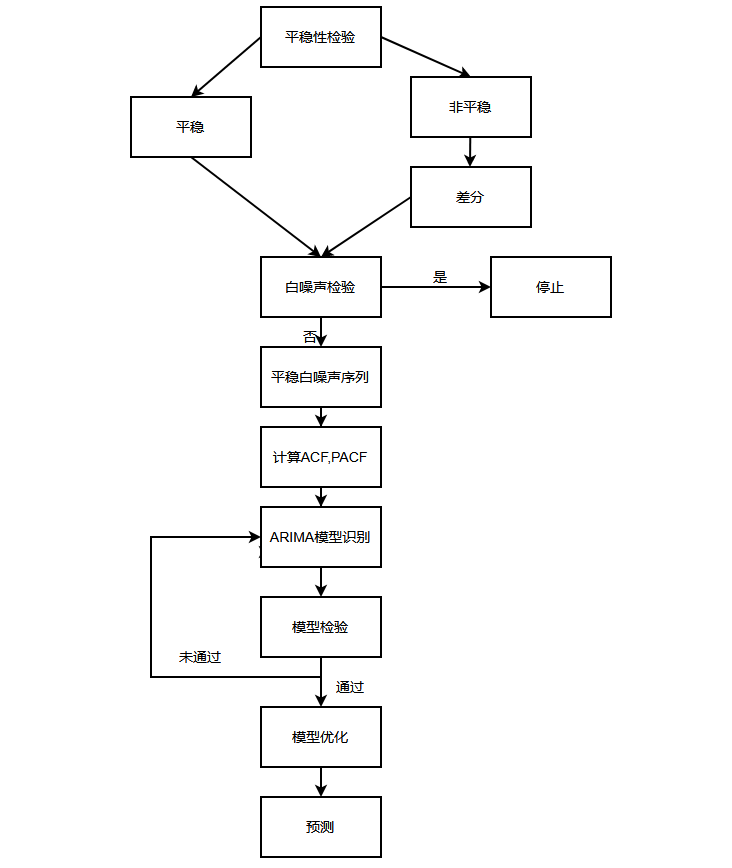


图2：回归预测结果展示

在这个模型中，可以明显观察到预测结果与实际结果有较大差距，使用回归模型预测时间序列的效果并不是很理想

### 2.2 方案二 ARIMA模型

使用ARIMA预测，预处理和填补缺失值的操作于以上类似，这里不再累述，模型的搭建过程如下图所示



1. 首先绘制时序图

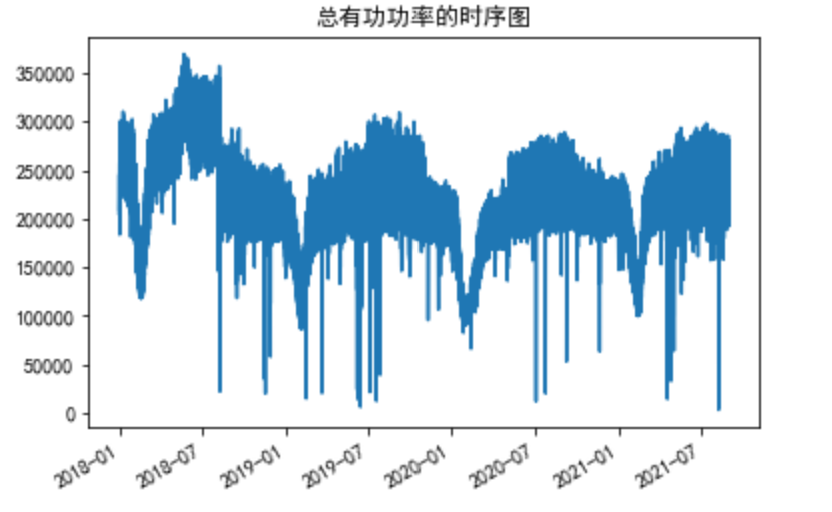


图1：ARIMA时序图

1. 按月进行重采样后绘制时序图

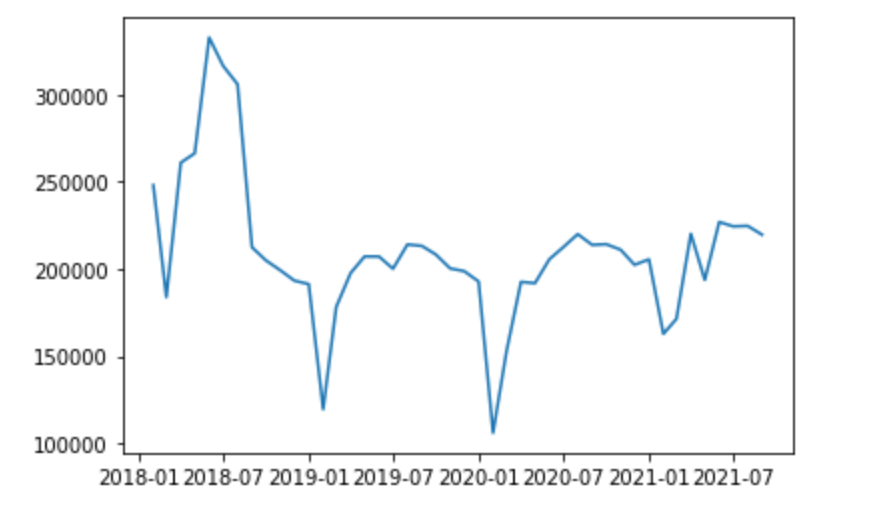


图2：总有功功率图

可以看出在总有功功率在18年有较大的波动，从2019-01开始，趋向于稳定的震荡（均值近似为常数）ARIMA 模型是通过寻找历史数据之间的自相关性，来预测未来（假设未来将重复历史的走势），要求序列必须是平稳的**。**因此我们取2019-01-01到2021-01-01的数据作为训练集，将2021-01-01到2021-08-31的数据作为测试集

1. 平稳性检验

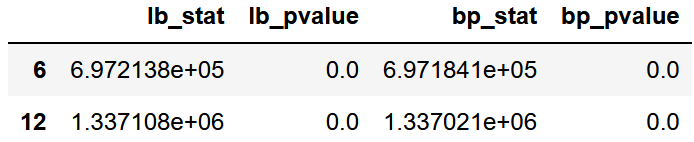
使用单位根检验中的ADF检验



1%、%5、%10不同程度拒绝原假设的统计值和ADF比较，ADF同时小于1%、5%、10%即说明非常好地拒绝该假设，本数据中，adf结果为-6.67， 小于三个level的统计值，说明数据是平稳的

1. 白噪声检验

使用QBP 和QLB 统计量进行序列的随机性检验



可以发现p值都很小，假设不成立，序列不是白噪声

1. 计算ACF,PACF

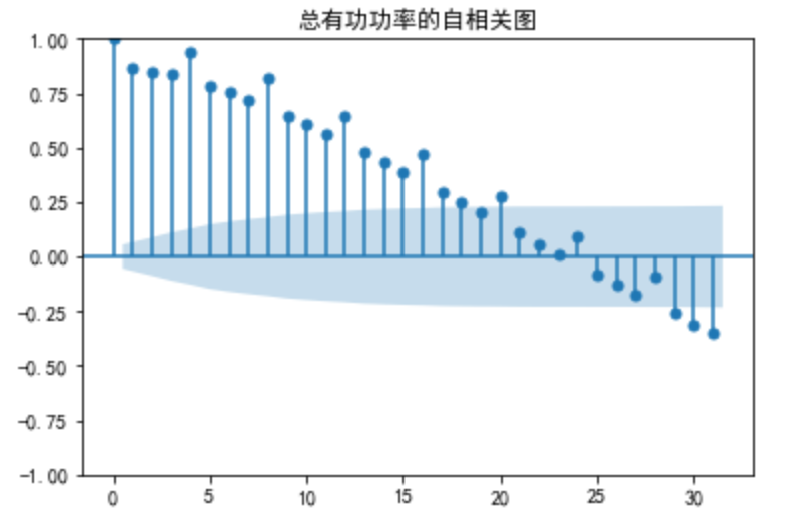


图1：总有功功率的自相关图

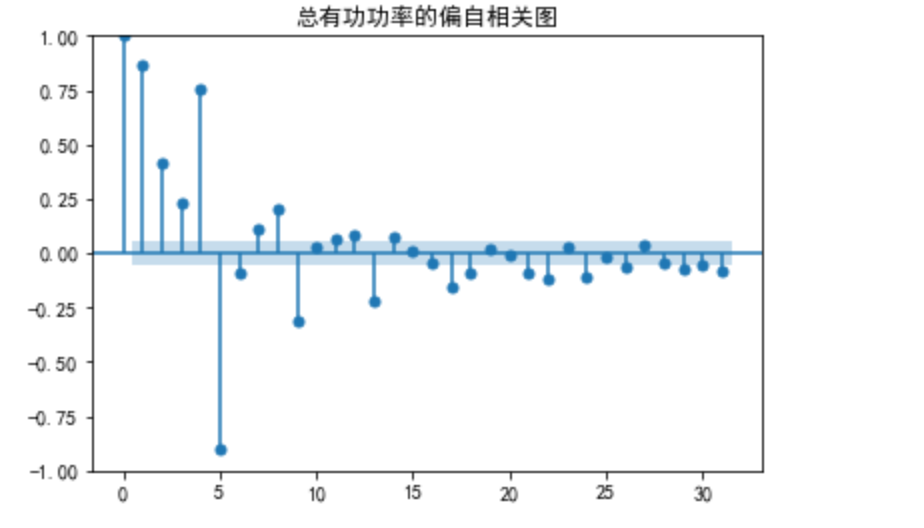


图2：总有功功率的偏自相关图

1. 模型拟合

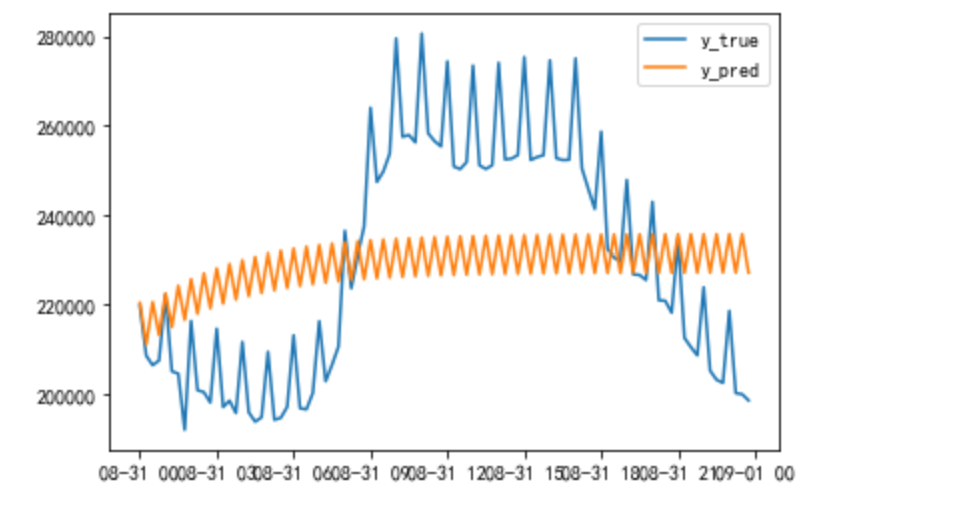
从ACF和PACF图的结果来看，p=7,q=4，因为看自相关图和偏自相关图有很大的主观性，因此，可以通过AIC或BIC来确定最合适的阶数



所以最终确定模型的阶数是(2,3)

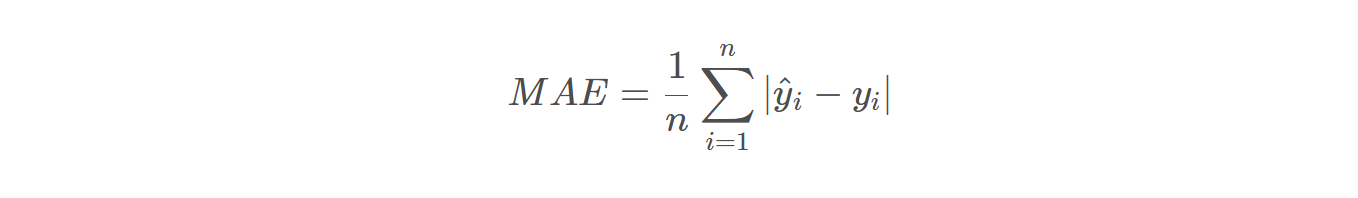
ARIMA(train,order=(2,0,3))

1. 模型预测



1. 模型评价

平均绝对误差（Mean Absolute Error）





1. 残差分析

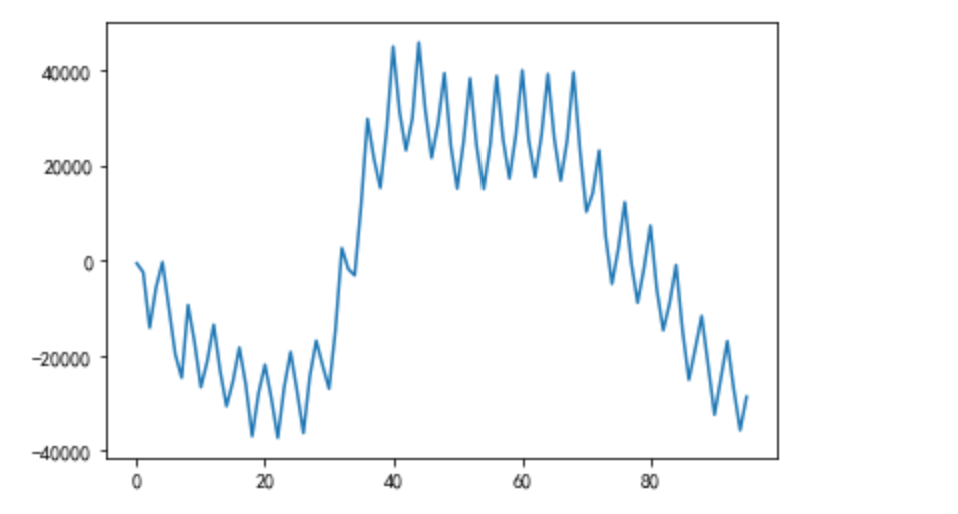


图1：残差图

1. 残差正态性检验

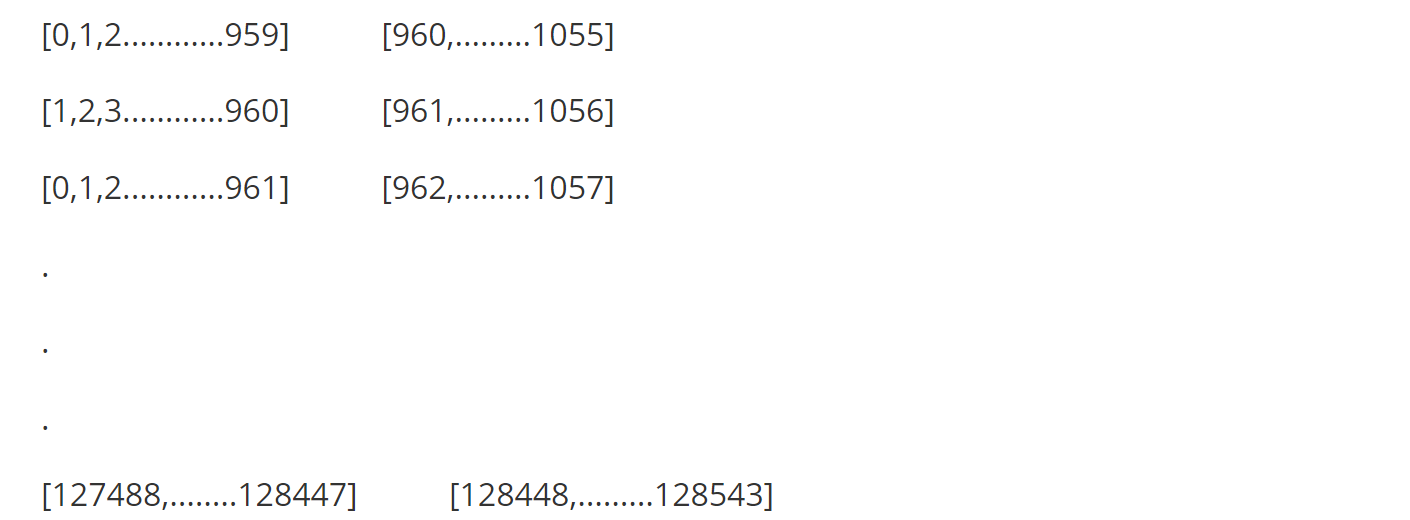


图1：残差正态检验

### 2.3 方案三 LSTM模型

长短期记忆（Long short-term memory, LSTM）是一种特殊的RNN，主要是为了解决长序列训练过程中的梯度消失和梯度爆炸问题。简单来说，就是相比普通的RNN，LSTM能够在更长的序列中有更好的表现。

首先把数据处理成下面这种形式(填补缺失值后数据总量128544)



0代表第一个数据(2018/1/1 0:00:00的总有功功率（kw）)，1代表第二个数据(2018/1/1 0:15:00的总有功功率（kw）)

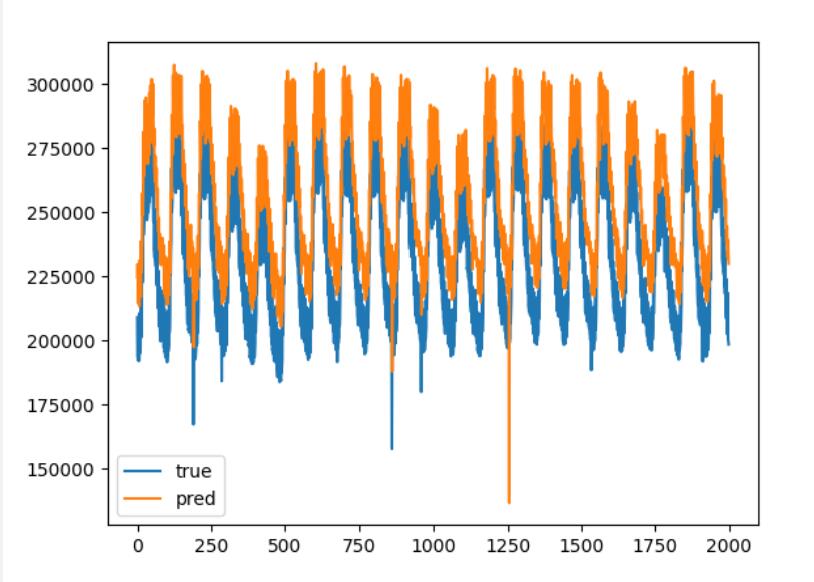


图1：LSTM预测模型

### 2.4 方案四 多元prophet模型

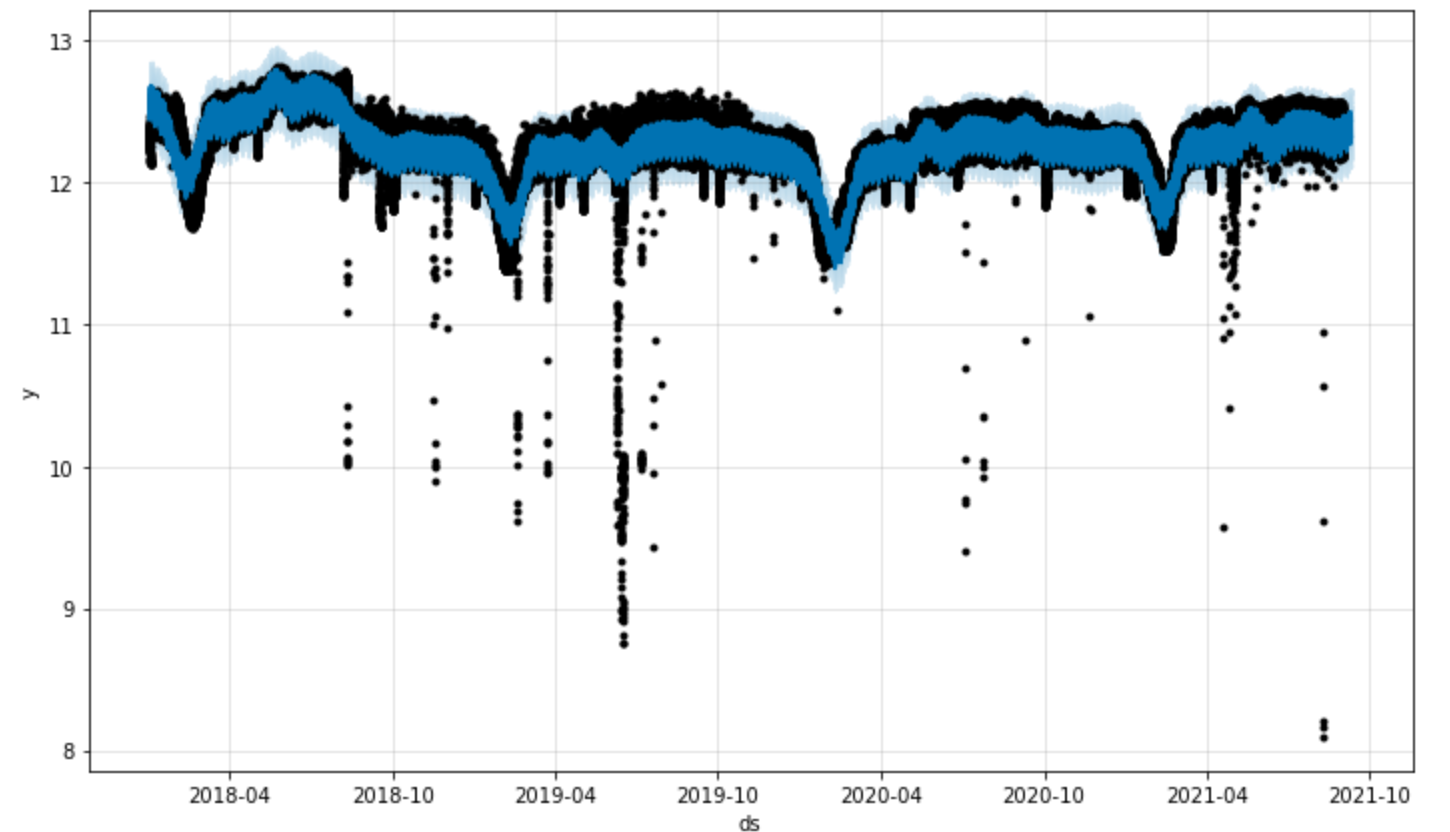


图1：多元prophet模型

1. 分析主成分

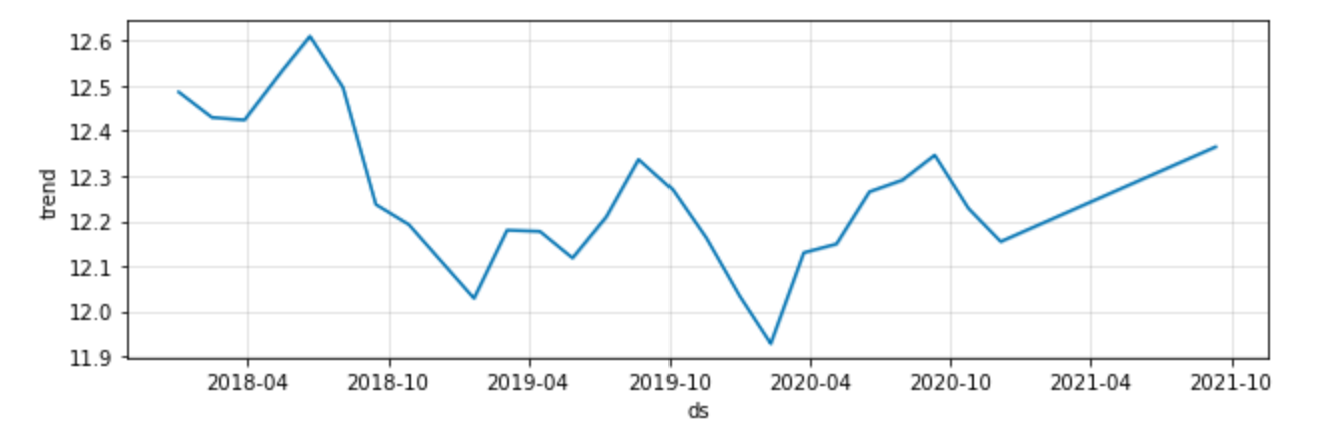


图1：主成分分析图1

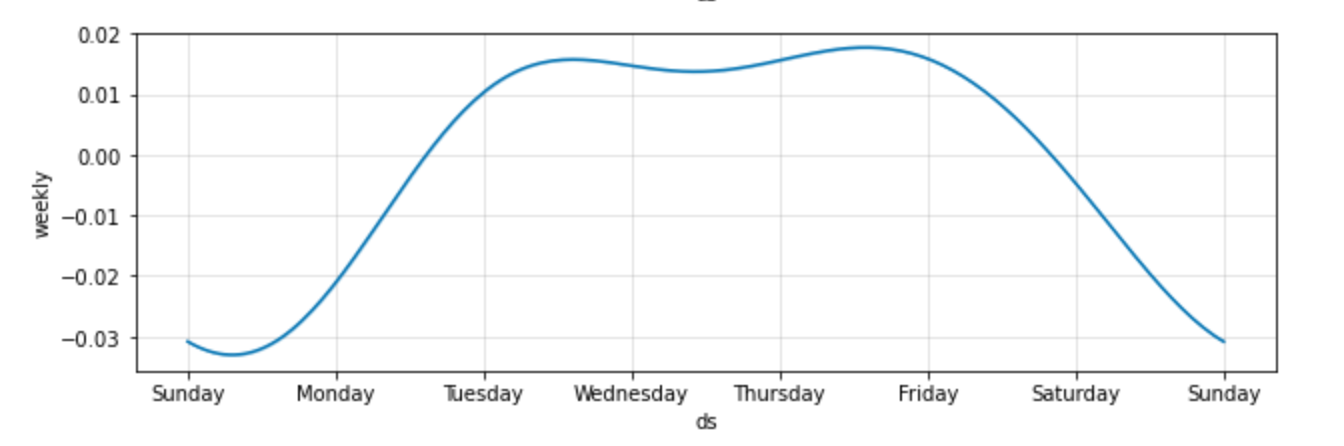


图1：主成分分析图2

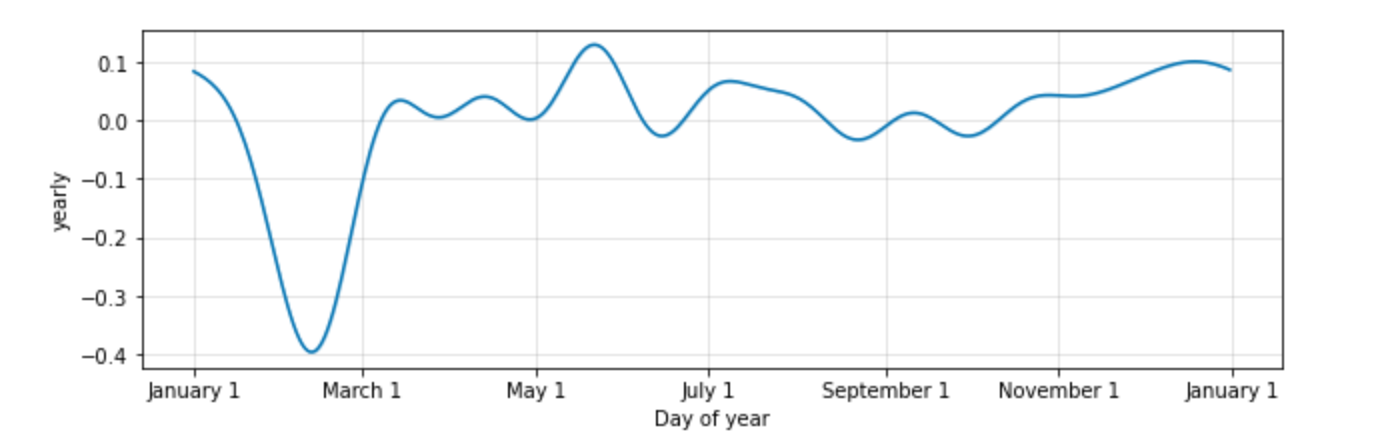


图1：主成分分析图3

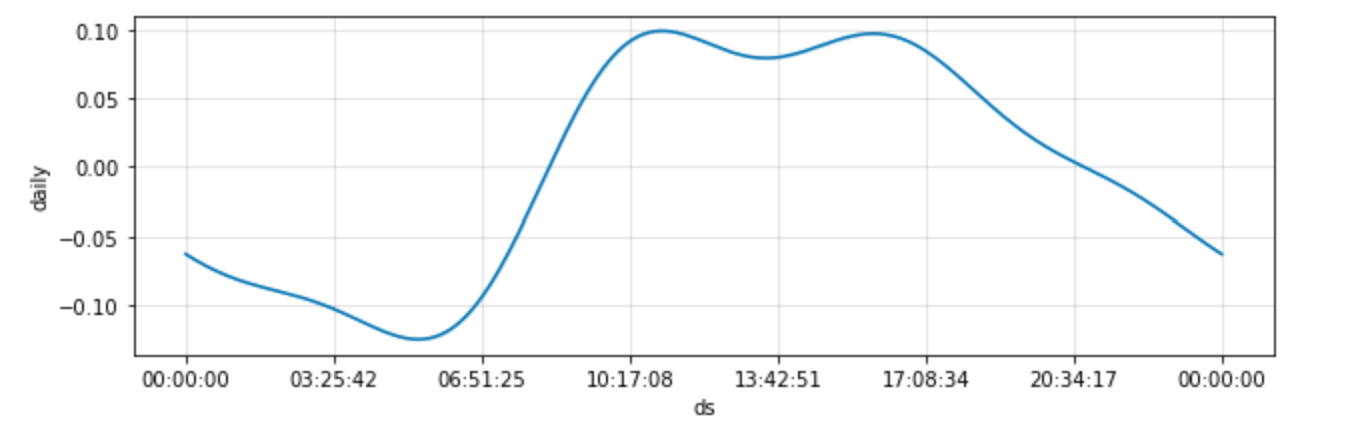


图1：主成分分析图4

1. 预测三个月的

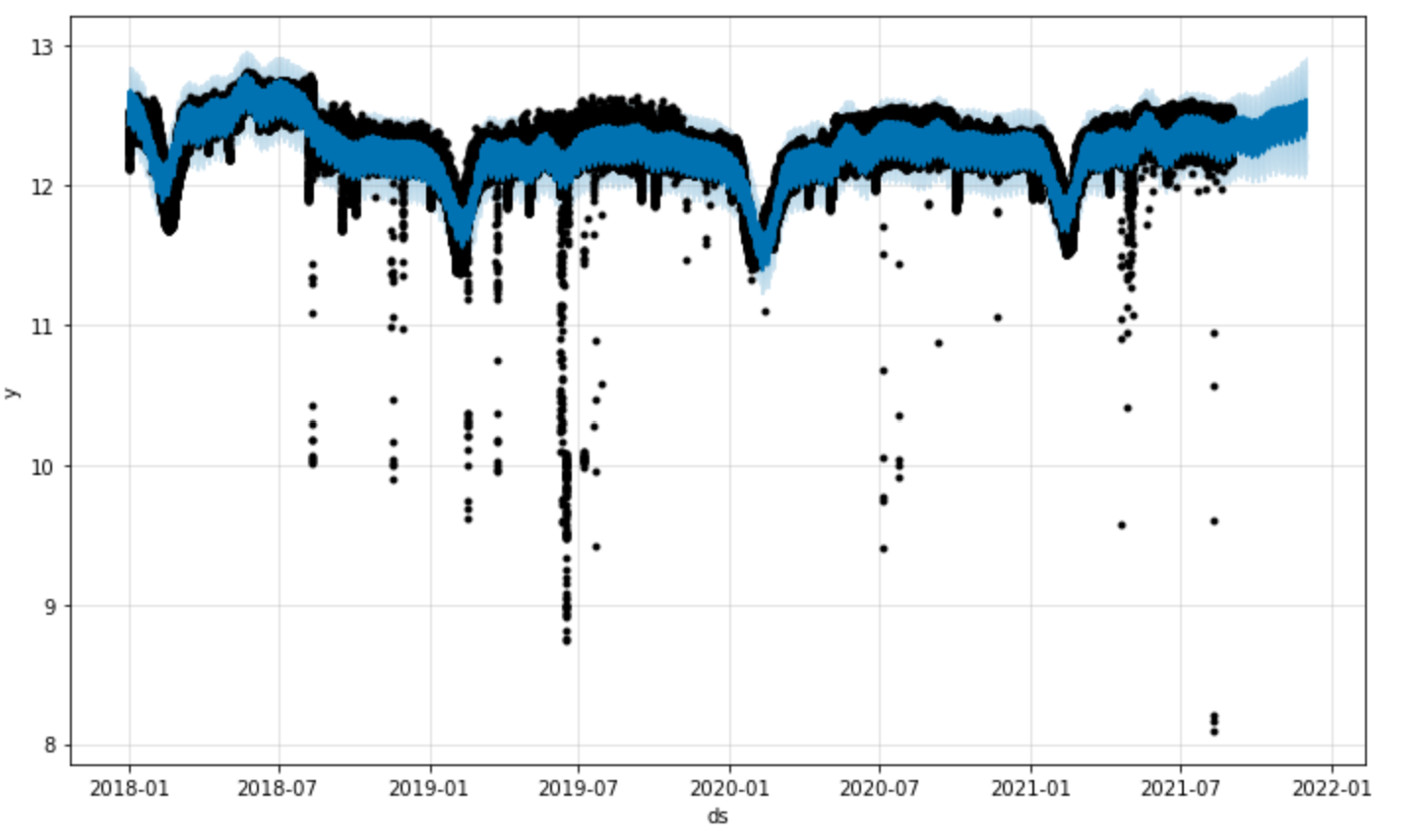


图4：prophet模型预测

# 三、行业负荷的中期预测分析

在进行行业负荷的中期预测分析过程中，本次实验首先采用了对附件二内的相关数据进行了数据预处理，以便后续的模型预测与整理。

## 分析各行业用电负荷突变的时间、量级和可能的原因

问题二中主要是对各行业用电负荷突变的时间、量级和可能的原因进行分析。

### 数据预处理

在数据预处理的过程中，本次研究考虑到此问题时针对用电负荷，用电负荷指的是用电设备和用户耗用功率得大小，考虑到附件二中得数据仅有有功功率的最大值和最小值，因此本次研究对附件二中的有功功率最大值和最小值进行整理，得到有功功率。用电负荷为有功功率和无功功率之和，无功功率是指电感或电容元件与交流电源往复的功率，考虑到在同一台设备上无功功率可以得之变换的可能较小，因此主要本次研究主要针对有功功率进行问题的探究。考虑到对各行业用电负荷突变的影响，因此，对附件三也进行处理（此处处理和问题一处处理类似，因此并不在此赘述），此后将附件二和附件三进行合并，并进行问题的整理与总结。

### 3.1.2 整体分析

针对此问题，本次实验首先通过绘制行业对比图进行观察对比。行业对比图可用于发现可能由于某天或某段时间出现了什么状况而导致所有行业功率最值均发生了突变，下图为观察大致趋势或者突变时间点或时间段

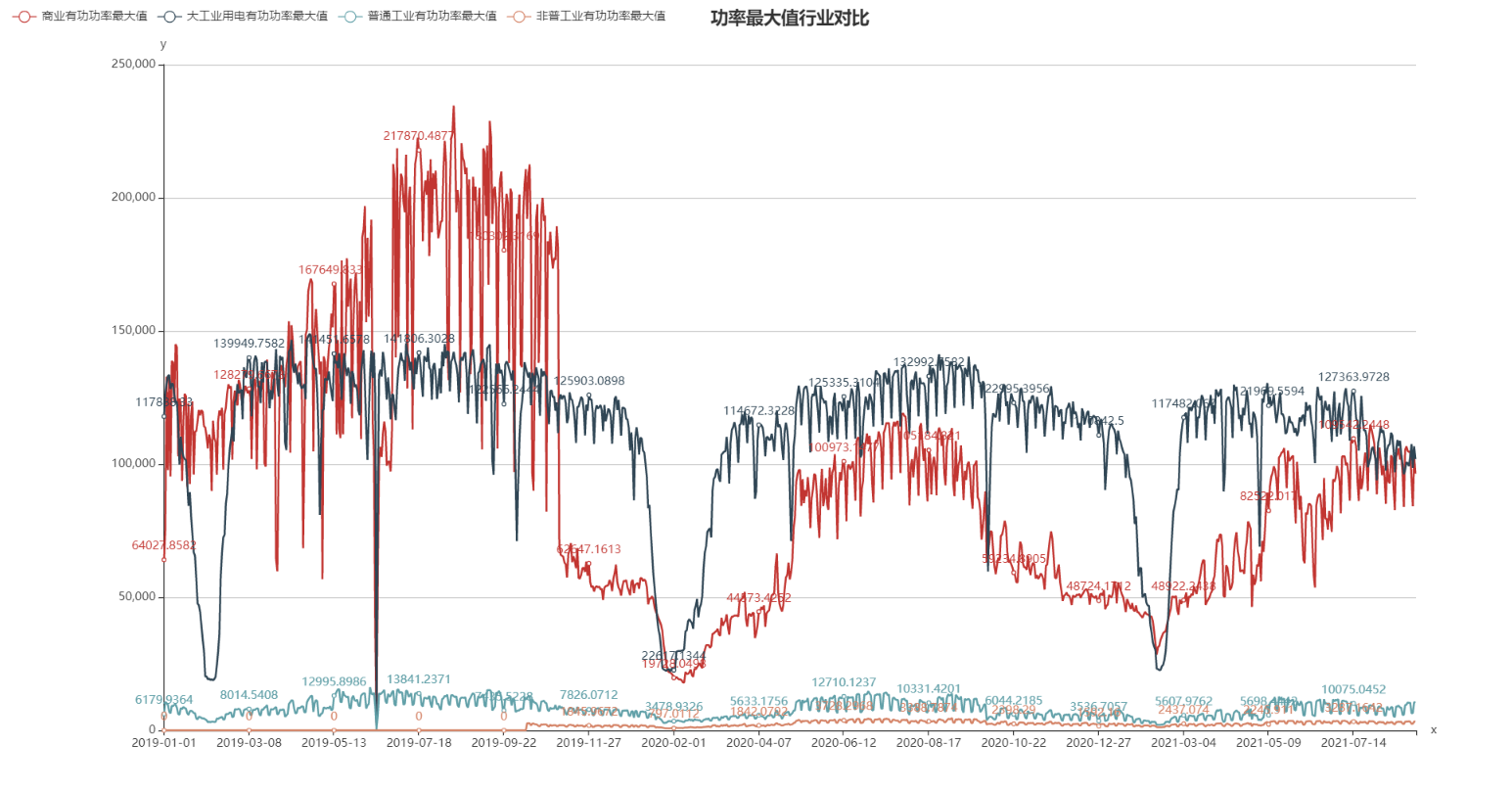


图1：不同行业有功功率最大值横向对比图

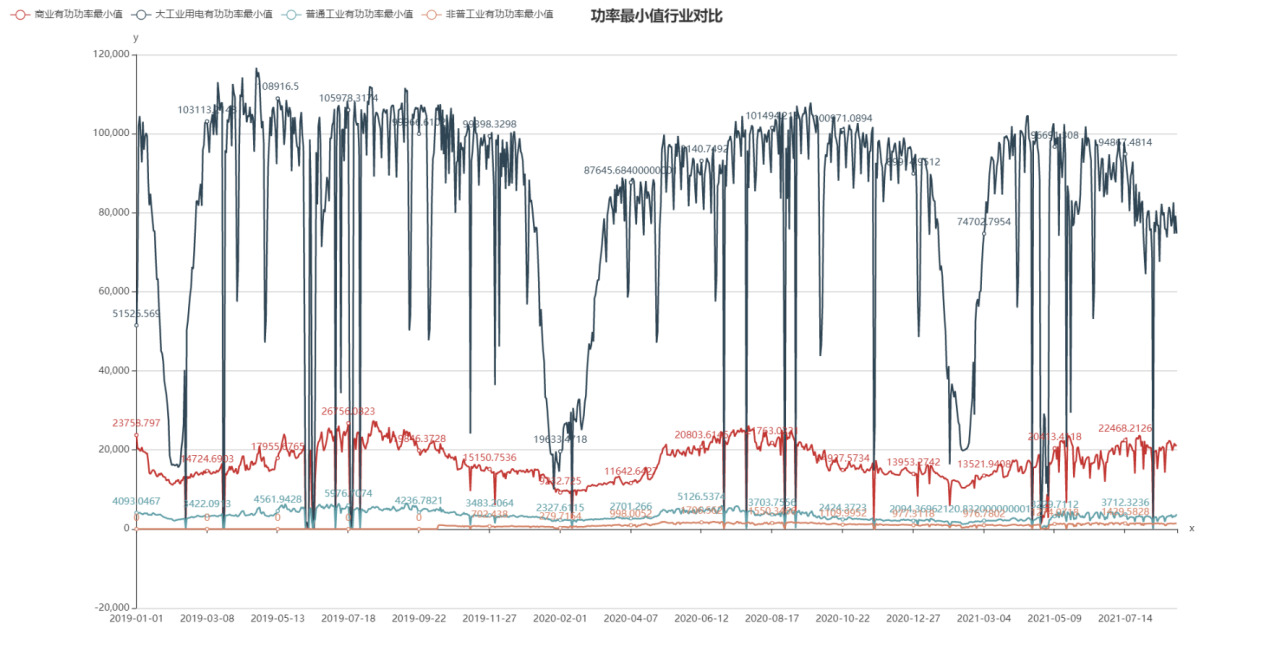


图2：不同行业有功功率最小值横向对比图

从图中可以注意到，商业有功功率最大值和大工业用电有功功率最大值有明显变化，而普通工业有功功率最大值和非普工也有功功率最大值无显著变化。

接下来，本次实验进行了各行业有功功率突变图的绘制进行分析。

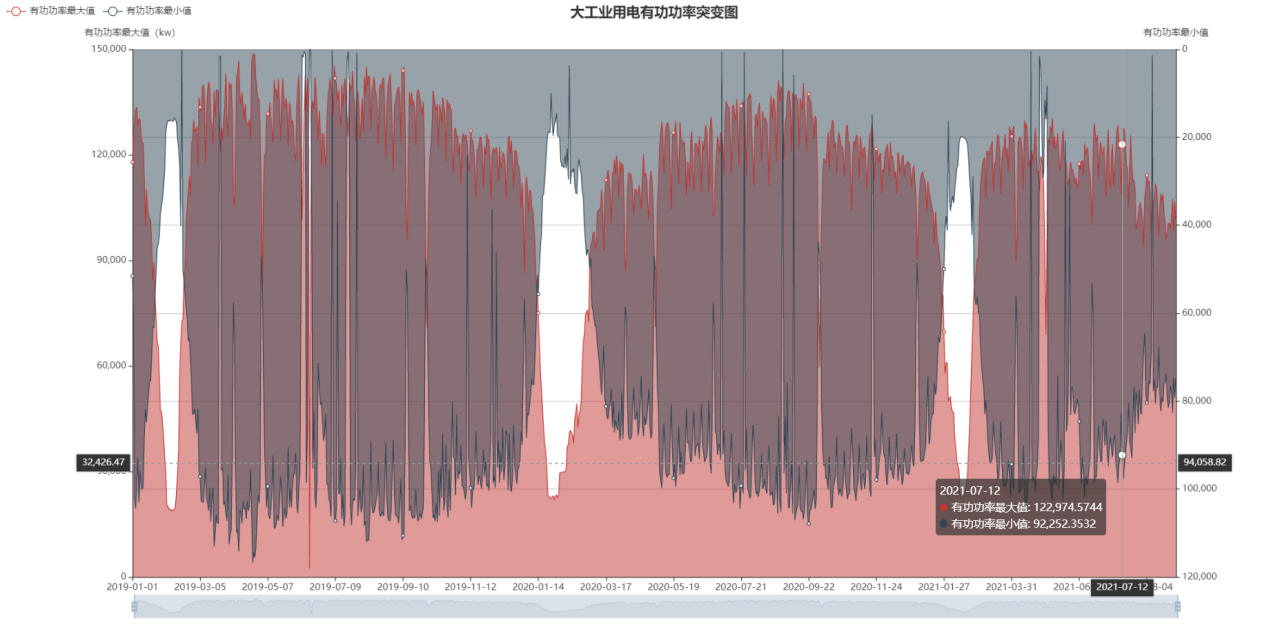


图3：大工业用电有功功率突变图



图4：非普工业有功功率突变图



图5：普通工业有功功率突变图

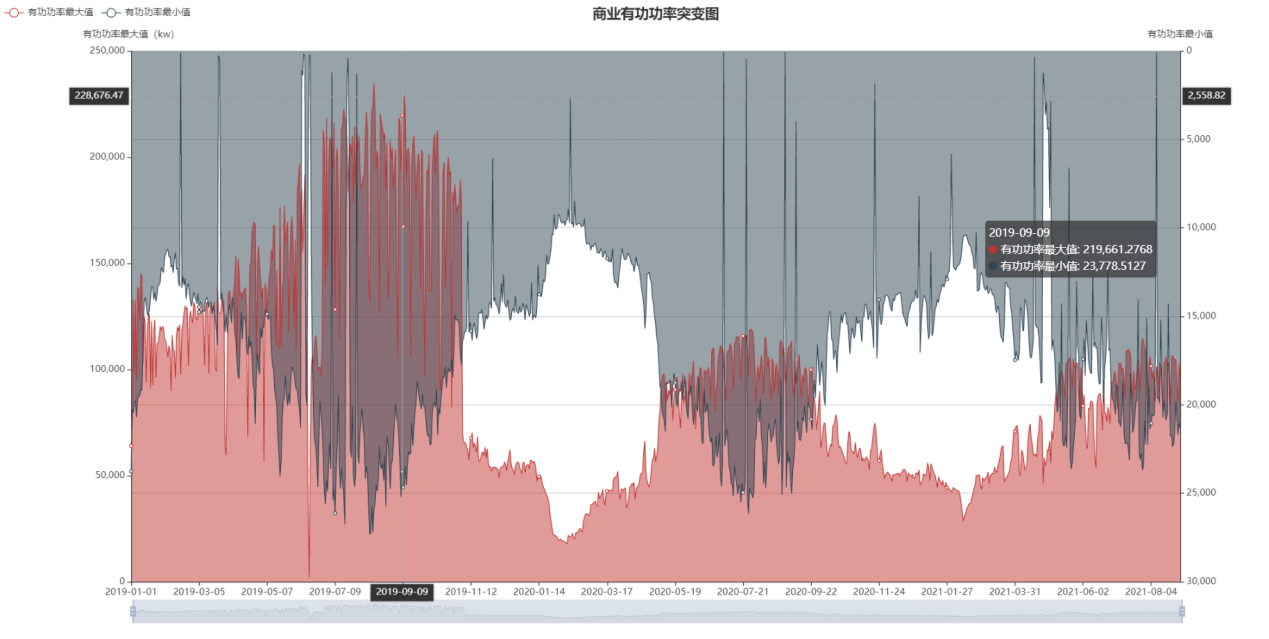


图6：商业有功功率突变图

通过对上图信息的分析，我们了解到各行业用电负荷突变的时间、量级和可能的原因为：天气原因。

### 突变的时间与量级

在这一过程中采用pettitt突变检测，测出突变时间点对应的时间的具体信息如下表所示：

|  |  |
| --- | --- |
| 时间 | 有功功率 |
| 2019-01-17 | 88667.8776 |
| 2019-03-06 | 117606.1401 |
| 2019-09-15 | 103311.1218 |

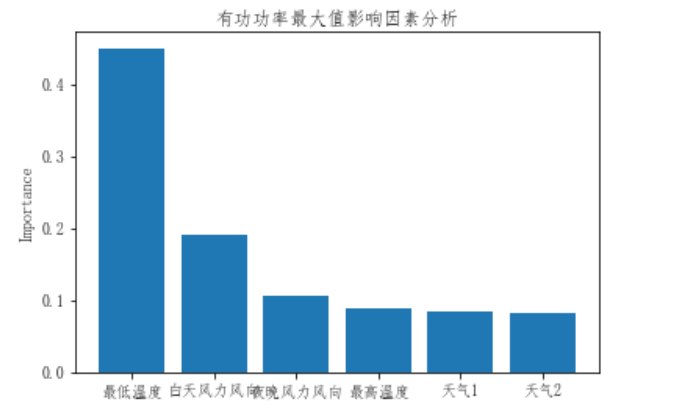
表1：突变时间与有功功率

根据突变的有功功率的处理，我们可以了解到的有功功率的突变点在有高有低，根据相关信息的整理，判断得出各行业用电负荷的突变量级为中级。

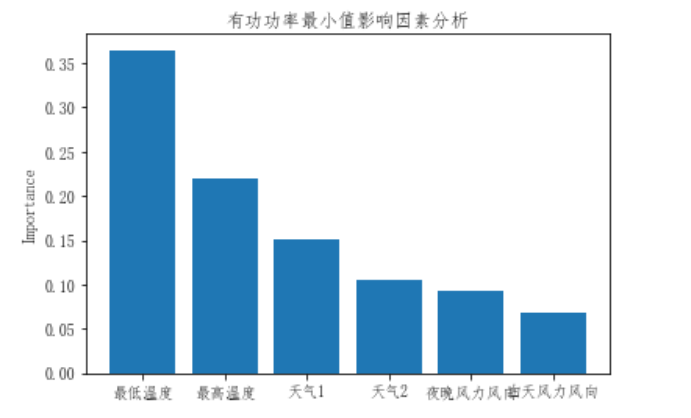
### 可能的原因

采用gbr\_reg对模型进行影响因素的分析

1. 对有功功率最大值进行分析影响因素分析



1. 对有功功率最小值进行分析影响因素分析



综合分析，温度对其的影响因素最大。

## 各行业未来3个月日负荷预测

在这一过程中，对ARIMA和LSTM的进行了模型的分析总结，以便更好的预测各行业未来3个月的日负荷的最大值和最小值的预测。

### 3.2.1 数据预处理

首先对数据进行预处理，这样可以首先更好的保障数据得使用效果。

#### 3.2.1.1 数据划分与整合

依据附件2将数据进行划分与整合，分离不同行业的有功功率最大值以及有功功率最小值数据，以得到时间序列分析格式化数据，为建立模型做准备。

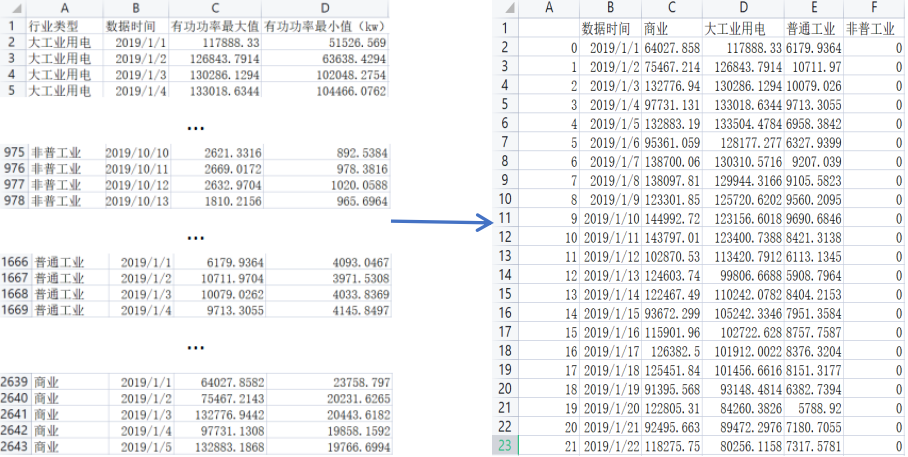
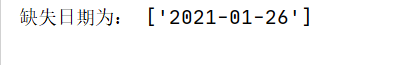


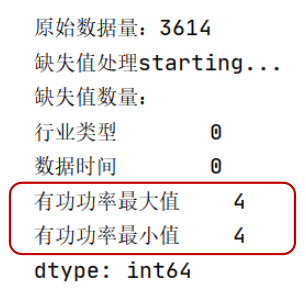
图1：附件二数据格式化

#### 3.2.1.2 日期连续性检测

防止数据日期本身存在缺失而造成的不连续情况，故对数据进行日期连续性检测。发现的确存在缺失日期：



由于仅仅存在一个缺失日期，故手动分别为四个行业加上日期“2021/1/26”，结果如下：



那么下一步只需对有功功率最大值和最小值分别进行填充即可。

#### 3.2.1.3 缺失值处理

1. 填充方法选择：

对缺失日期对应的数据值进行填充，常用的缺失值填充方法有：指定值填充、均值填充、中位数填充、众数填充、上下邻近数据值填充。

·指定值填充：对于该缺失值，并不能准确推测出一个量化的指定值进行填充。

·中位数填充：不能很好地考虑到时序影响。

·众数填充：有功功率并不像数量那般具有一定的说明性，众数填充意义不大。

·均值填充：均值中和掉了该数据本该有的特征，不适合用于此处。

·上下邻近数据填充：无法说明究竟上邻近还是下邻近数据填充哪一个更适合。

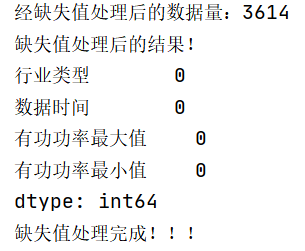
基于该数据的特点，又考虑到时间序列的影响，本报告欲结合均值填充和上下邻近数据填充，采用上下两行数据均值填充。既可以避免全部均值的中和特征缺点，又可以很好地利用时序特征。

1. 填充过程：

调用interpolate()函数进行缺失值填充，即：



1. 填充结果：

到此为止，数据缺失值处理完毕。

#### 3.2.1.4重复值检测

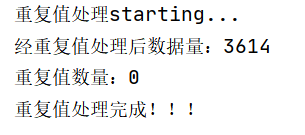
1. 处理方法：

·若存在重复值，则进行删除处理。

·若不存在重复值，则不必做出处理。

说明：基于该数据，本报告中的重复值是指完全重复的行数据。

（2）重复值检测：



由上，没有发现重复值，故不必做出处理。

数据处理完成，接下来进行数据分析与模型架构

### 3.2.2 ARIMA

#### 3.3.2.1 数据整合

由附件2将不同行业有功功率最大值和有功功率最小值分离开来，得到标准的时间序列分析格式的数据（即得到时间序列+某一行业的一列数据的数据格式），再对其分别进行分析与挖掘。

#### 3.3.2.2分出训练集和验证集

在这一过程中对数据进行训练集和验证集得划分，使之完成建模步骤。

#### 3.3.2.3 建模

（1）平稳性检验：绘制序列时序图，直观检测该序列的平稳性，发现无明显趋势或周期特征

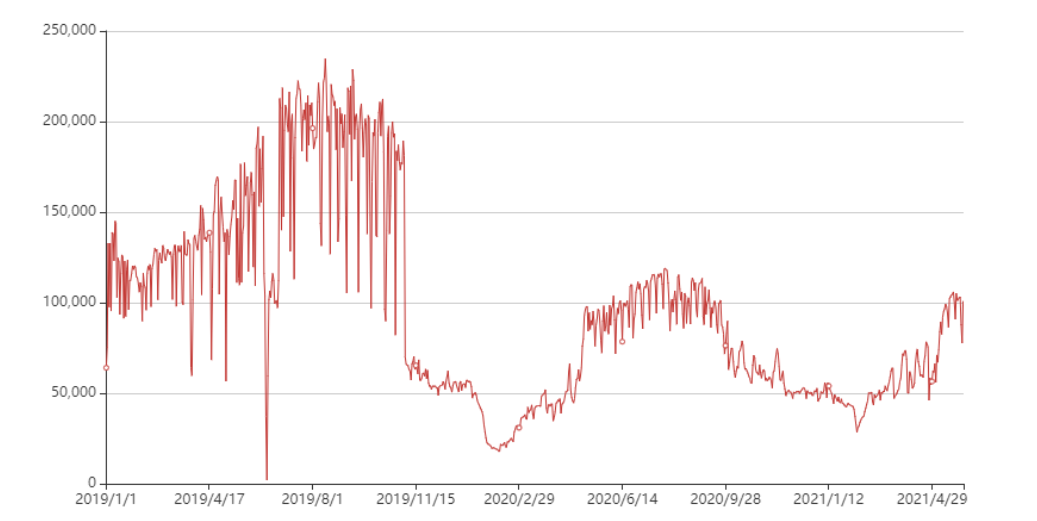
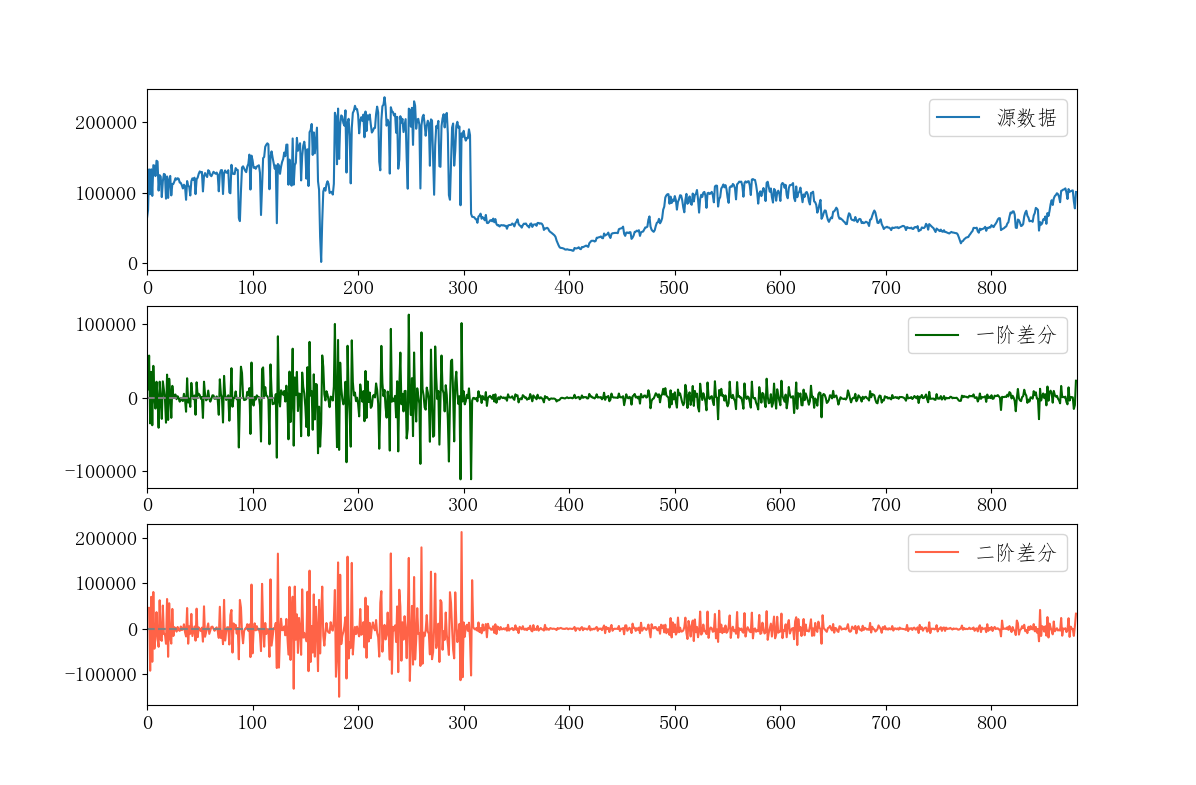


图1：序列时间图

1. 计算出该观察值序列训练样本的自相关系数（ACF）和偏自相关系数（PACF），并绘制ACF和PACF图
2. 模型识别与定阶：根据ACF和PACF的性质，进行模型定阶



得出p = q = r = 1

1. 模型检验
2. 模型所产生的残差做自相关图

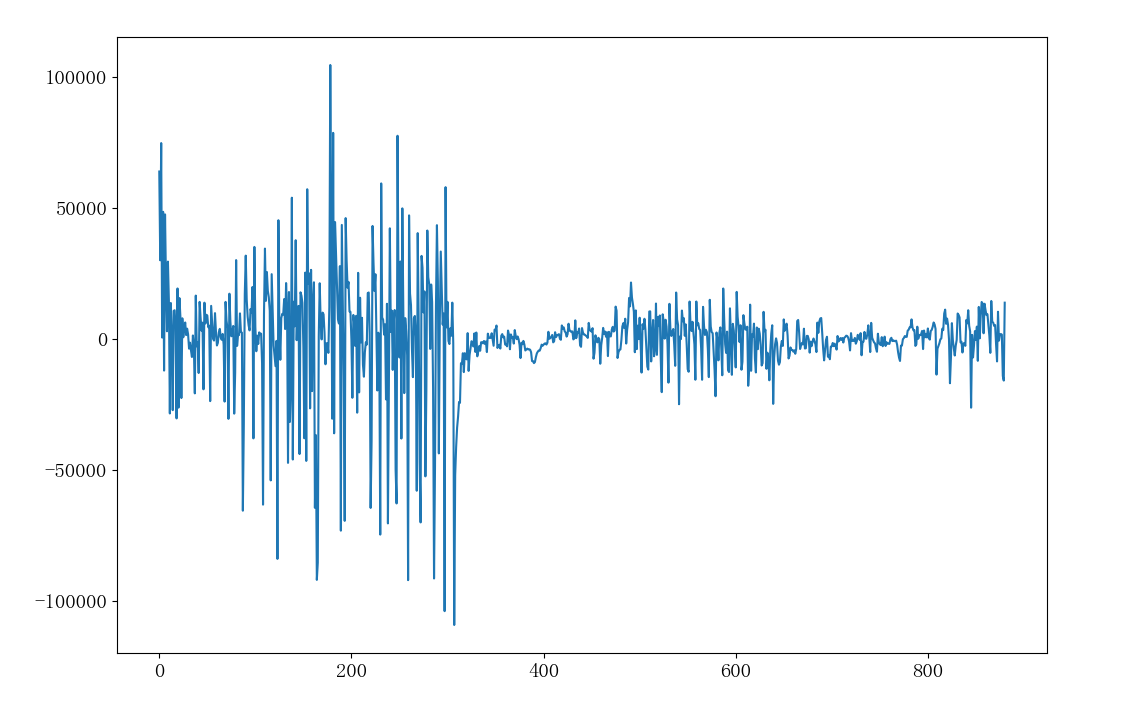


图1：残差相关图

1. 观察是否符合正态分布

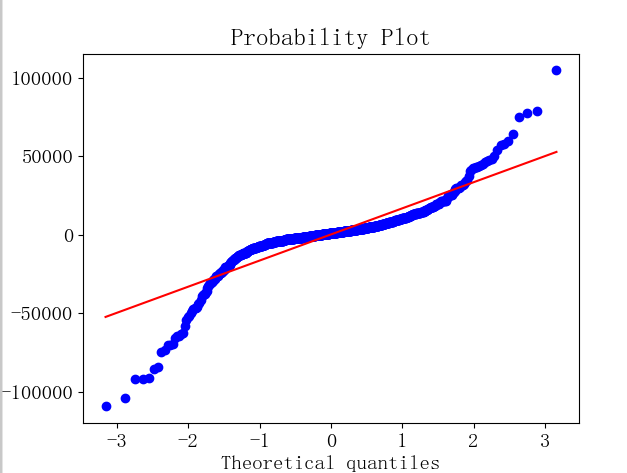
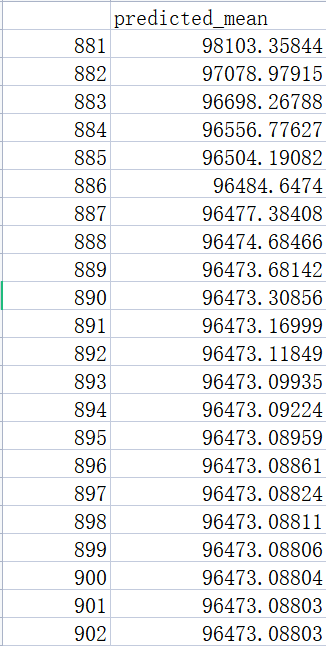


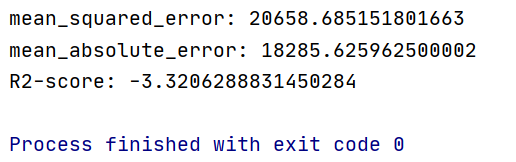
图2：观察正态分布

1. 模型预测

部分预测结果如下：



1. 模型评估及精度分析：



可以看到RMSE(平方根误差)、MAE（平均绝对误差）的值都比较大，说明该模型不够有效；其次，R2\_score为负值，预测效果不佳。

#### 3.3.2.4 分析总结

虽然ARIMA模型十分简单，只需内生变量而不需要借助其他外生变量，但其要求时序是稳定的，或者是通过差分后是稳定的，而针对该问题的用电负荷预测，常常受各种不确定性因素的影响而波动甚至产生突变，本质上只能捕捉线性关系，而不能捕捉非线性关系，所以在该问题上有其局限性，适合短期预测问题，当预测时间超过一定的天数，会自动以均值作为预测值填充，对中长期时间序列预测问题并不能发挥出其优势。所以本报告的ARIMA模型仅作为初始对数据的探索与模型初步尝试。

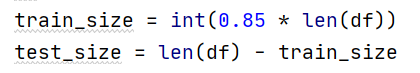
### 3.3.2 LSTM

#### 3.3.2.1 数据预处理

与ARIMA模型中的数据预处理基本一致，不再赘述。

#### 3.3.2.2划分训练集和验证集

将整合的数据集划分为训练集和验证集，将数据集中的85%作为训练集，剩下的作为验证集。即：



划分结果如下：



#### 3.3.2.3数据归一化与标准化

（1）说明

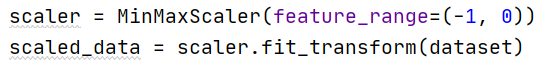
为了模型求解更加准确快速，提高模型的收敛速度，消除不同量纲带来的问题，避免数值过大等问题，本报告在训练LSTM之前，对数据进行归一化和标准化处理。

归一化：将数据变成(０，１)或者（1,1）之间的小数。便于数据处理，把数据映射到0～1范围之内处理，更加便捷快速。其次，把有量纲表达式变成无量纲表达式，便于不同单位或量级的指标能够进行比较和加权。

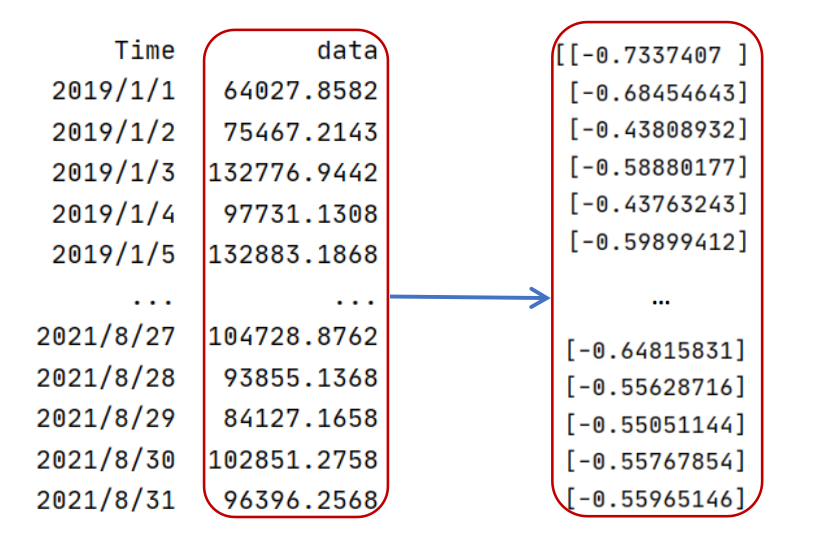
标准化：标准化后会使每个特征中的数值平均变为0(将每个特征的值都减掉原始数据中该特征的平均)、标准差变为1。

1. 归一化和标准化的实现：

直接调用sklearn.preprocessing中的MinMaxScaler即可：



结果如下：

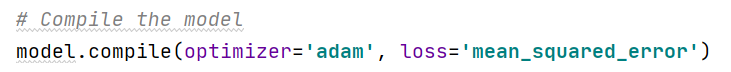


#### 3.3.2.4模型建立

首先使用Sequential()方法创造一个容器，进而描述神经网络的网络结构；然后构造LSTM层和全连接层，主要代码如下：

#### 3.3.2.5 设置优化器、[损失函数](https://so.csdn.net/so/search?q=%E6%8D%9F%E5%A4%B1%E5%87%BD%E6%95%B0&spm=1001.2101.3001.7020" \t "https://blog.csdn.net/Fwuyi/article/details/_blank)和准确率评测标准

1. 优化器选择：本报告在问题二中构建LSTM时，采用adam优化器，adam优化器结合了AdaGrad和RMSProp两种优化算法的优点。实现简单，计算高效；更新的步长能够被限制在大致的范围内，可以自动调整学习率；适合应用于本题中大规模的数据及参数的场景，且适用于不稳定的目标函数。故选择adam作为优化器。
2. 损失函数：为反映预测值与真实值之间的差异程度，这里选择均方误差作为度量标准。实现如下：

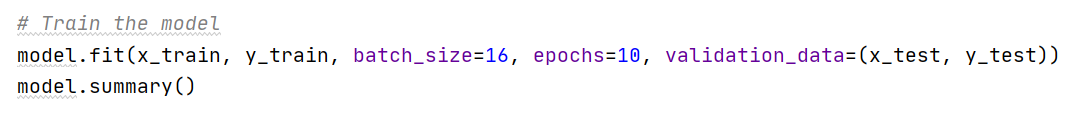


#### 3.3.2.6 模型训练：

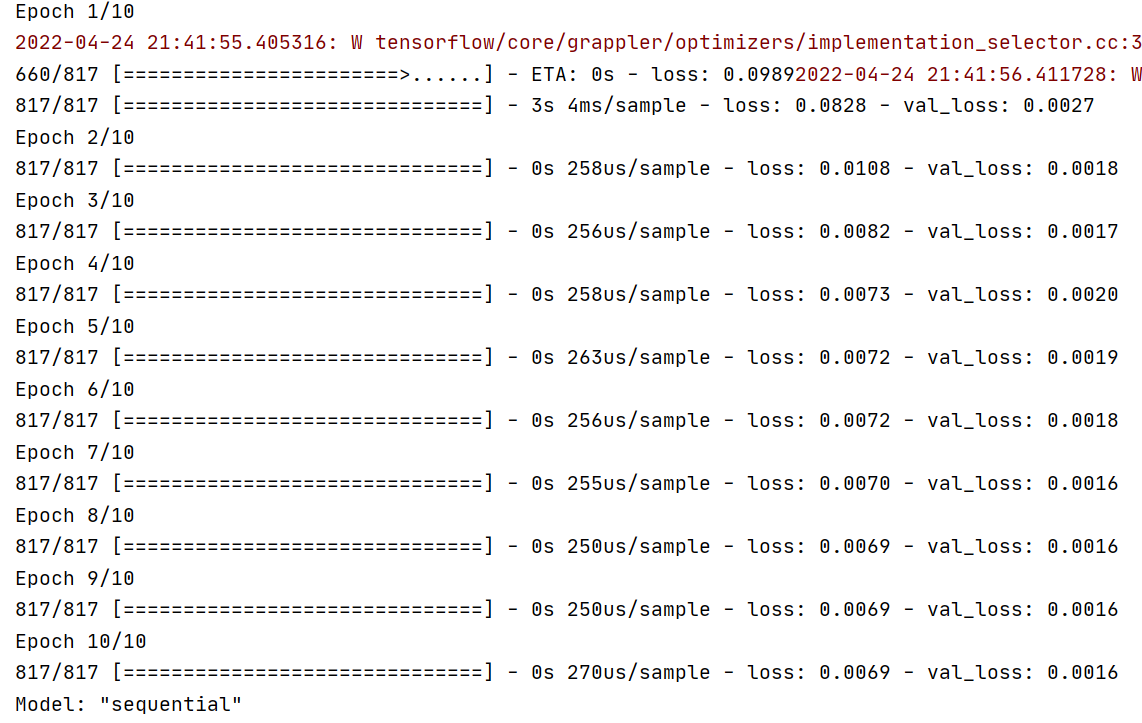
1. 设置batch\_size和epochs：

·batch\_size：一次训练所抓取的数据样本数量，其大小影响训练速度和模型优化，设置batch\_size可以使得内存的利用率提高，跑完一次训练数据的迭代次数减少，加快训练速度。在一定范围内，虽然加大batchsize的值可以使得下降方向变得准确，但是盲目加大batchsize的值，会因为数据集过大而受限于内存。因此，我们初始化为15，根据训练的loss再进一步进行调参。

·epochs：1个epoch相当于使用训练集中的全部样本训练一次。这里我们先设置为10。即：

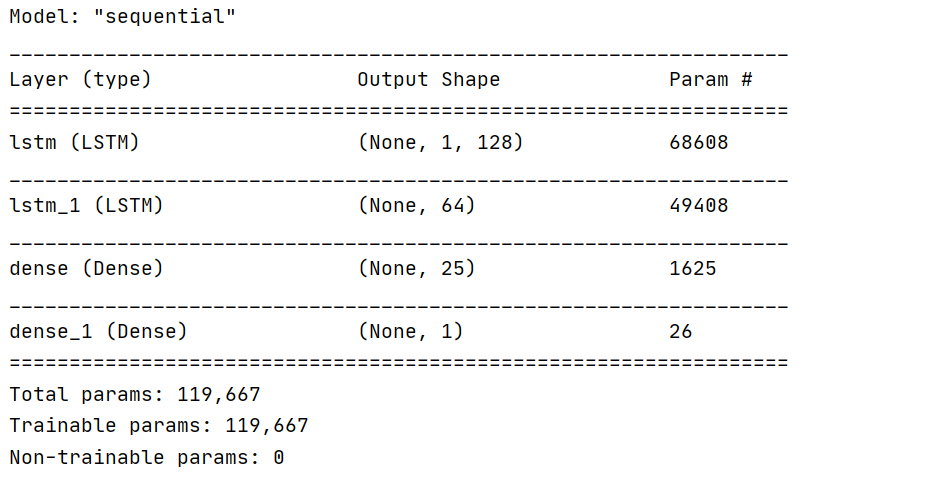


训练结果如下：



可以看到通过迭代，loss值逐渐减小且趋于不变，说明参数选择的还是较为合适的。

模型summary如下：



#### 3.3.2.7 模型预测及数据还原

1. 模型预测效果图：

·商业：

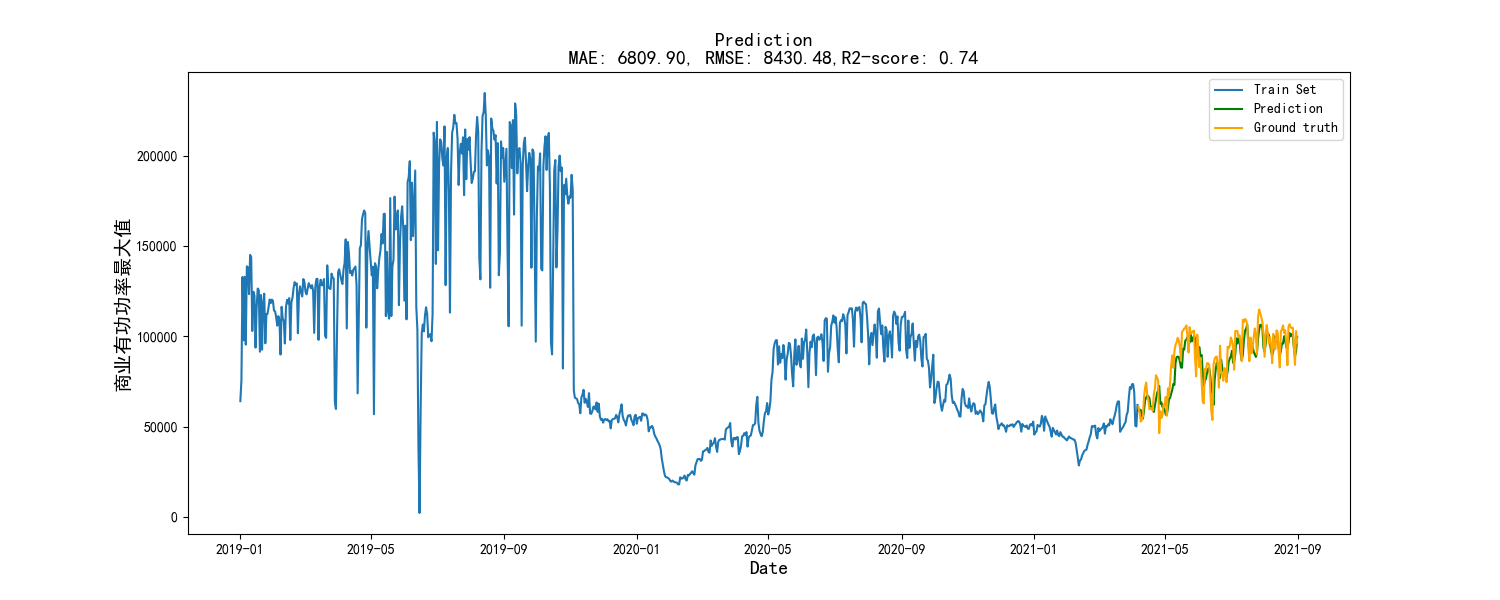
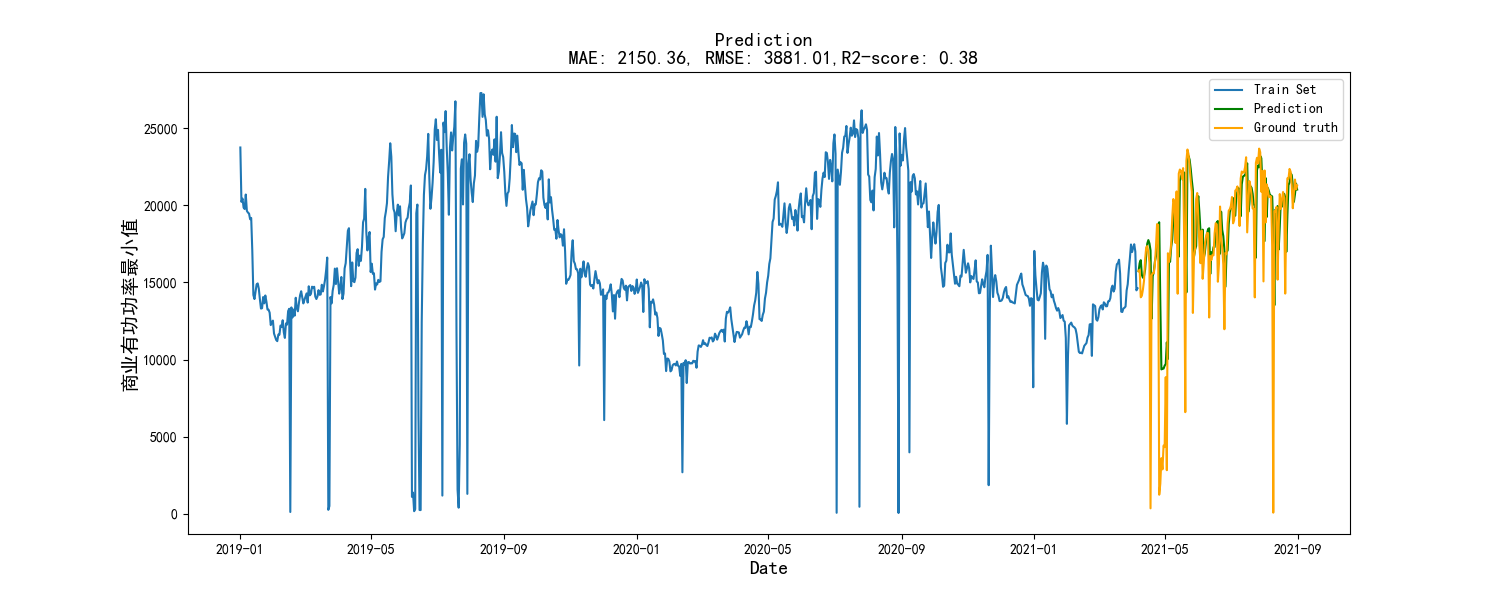


图1：商业预测效果图1

图1：商业预测模型图2

·大工业用电：

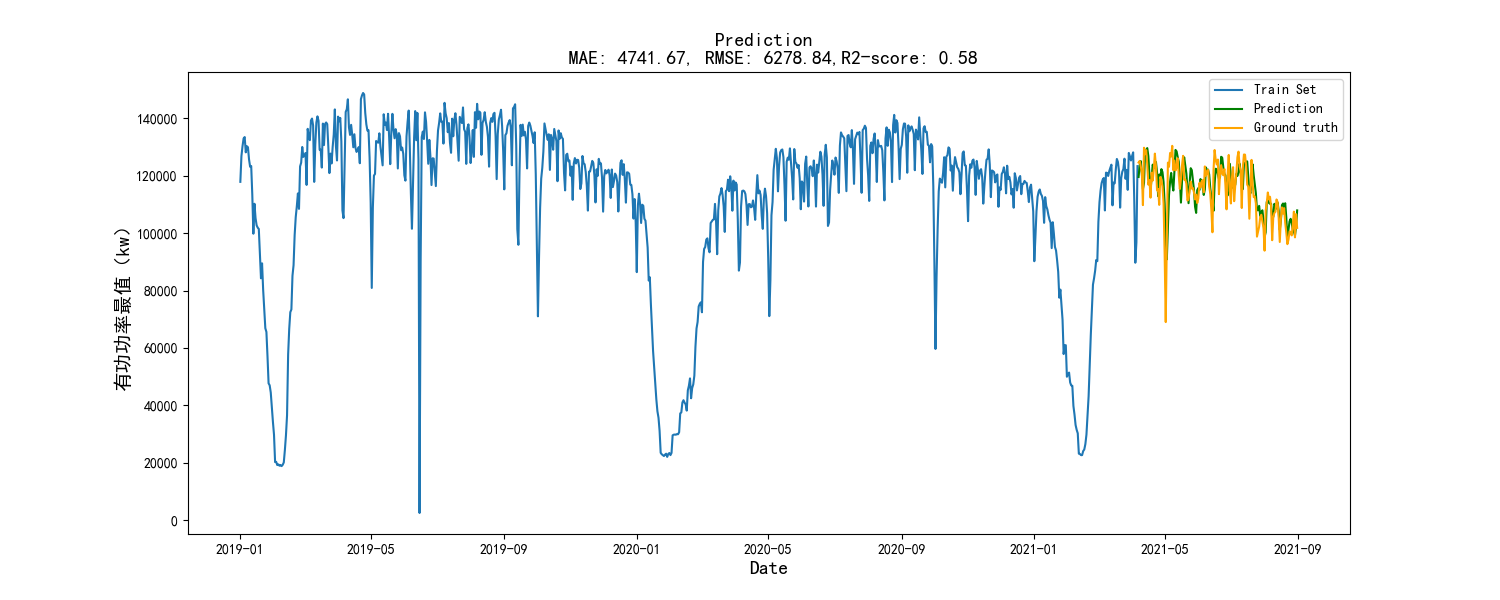


图1：大工业用电预测模型图1

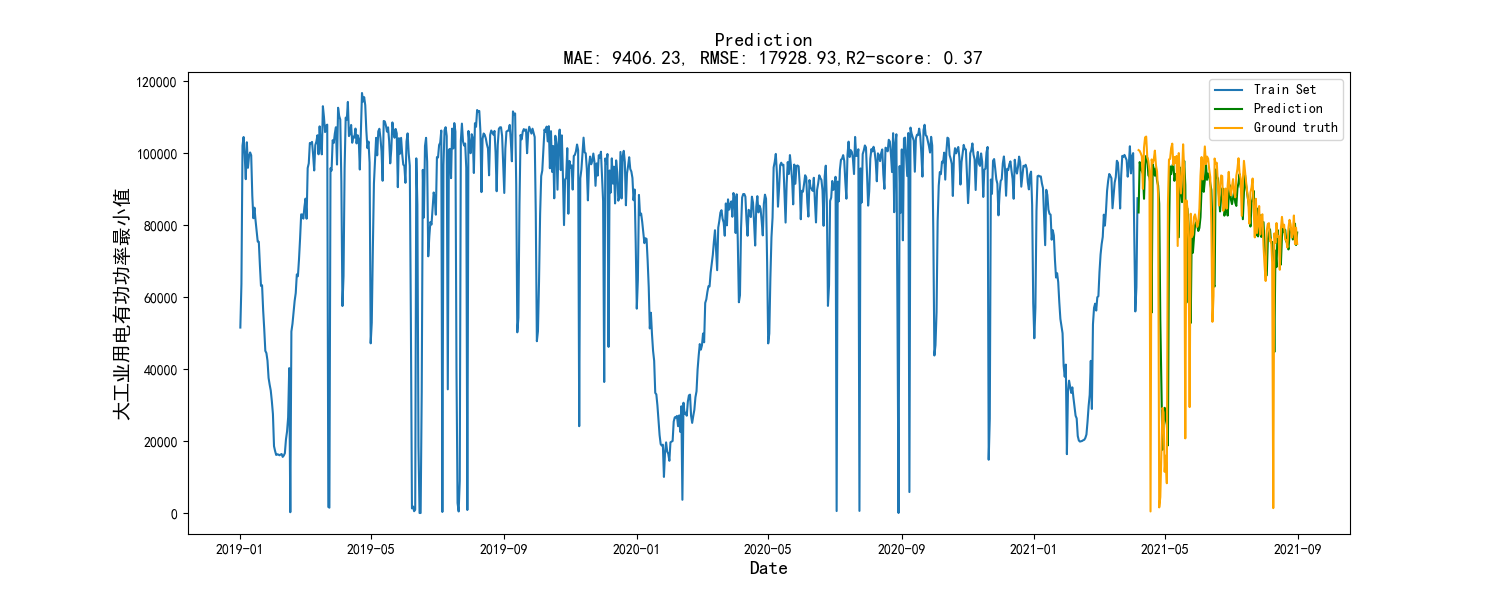
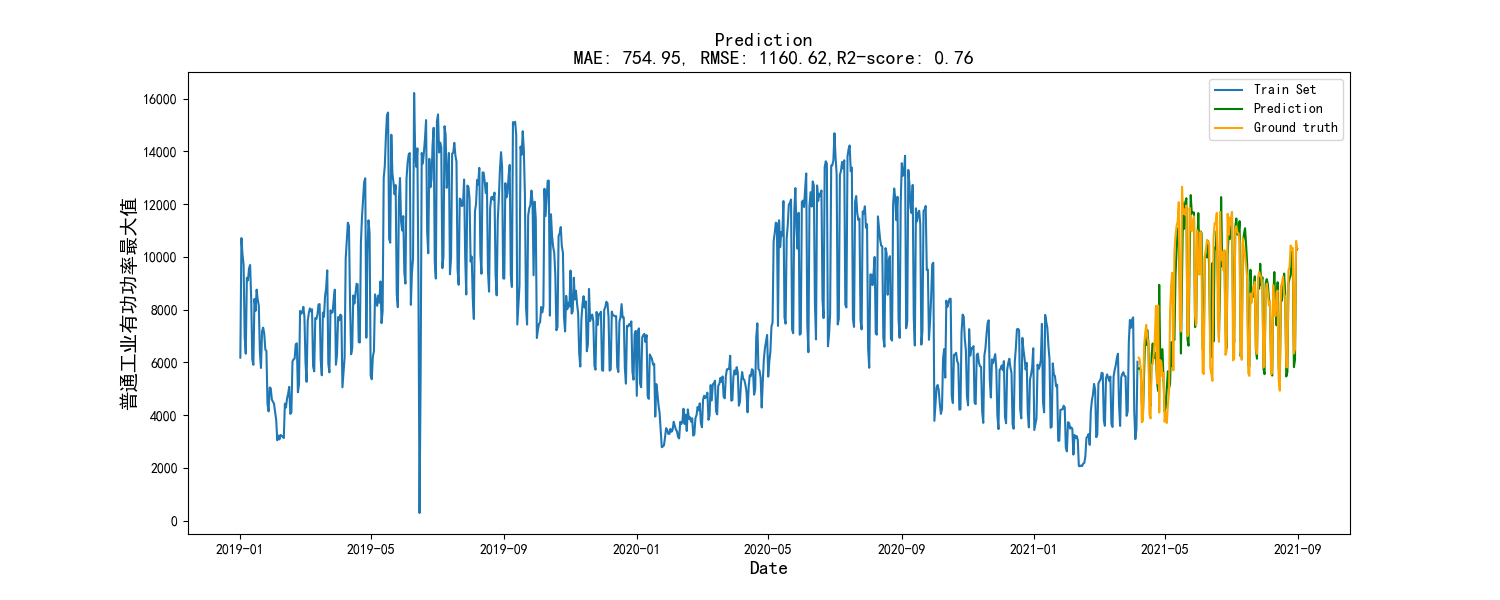


图1：大工业用电预测模型图2

·普通工业：

图1：普通工业用电预测模型图1

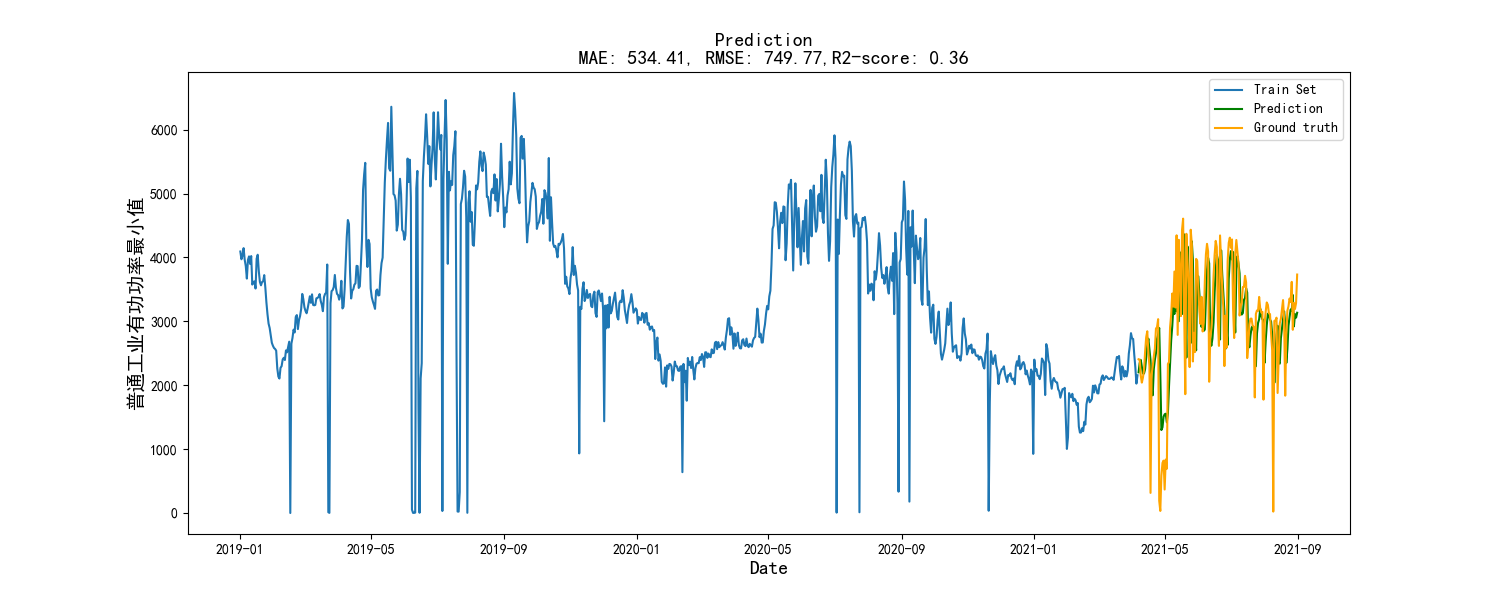


图1：普通工业用电预测模型图2

·非普工业：

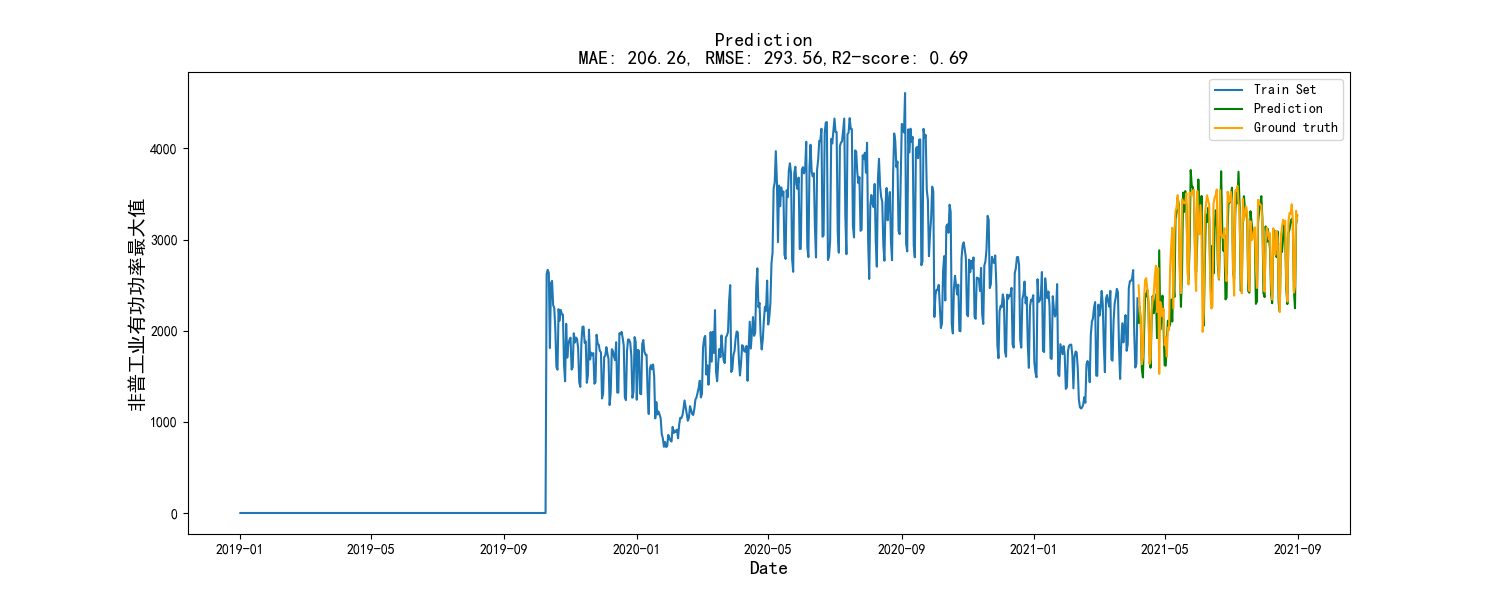


图1：非普工业用电预测模型图1

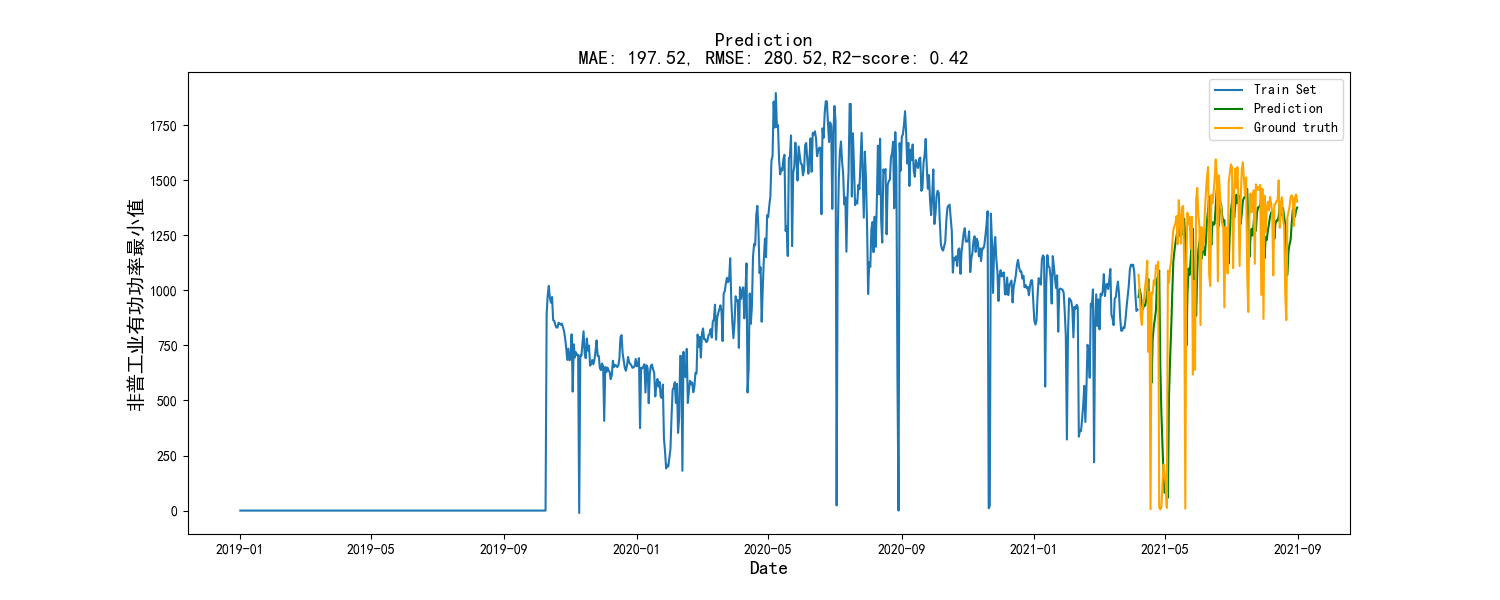


图1：非普工业用电预测模型图1

1. 数据还原输出至csv：将原本经过归一化和标准化的数据进行还原输出。

#### 3.3.2.8 模型评估

计算RMSE、MSE、R2-score对模型进行评估，输出值已展示于预测效果图上，不再一一列出。

#### 3.3.2.9 模型优化

通过调整batchsize、epochs、rolling\_window等不断训练，寻找较好的参数值，以得到更优的模型。

#### 3.3.2.10 模型评价

长短期记忆（Long short-term memory, LSTM）是一种特殊的RNN，就是相比普通的RNN，LSTM能够在更长的序列中有更好的表现。能够学习长期依赖性。

## 对各行业的建议

1. 针对大工业用电可虑到大工业企业为24小时持续运行，应根据天气情况进行急速调节，保证大工业用电的用电保障。
2. 对于商业用电，商业负荷主要用于商业部门的照明、空调、动力等用电负荷，覆盖面积较大，用电增长比较平稳，具有季节波动等情况，因此可为商业用电在节假日等人流较大的时候增加用电量，以保障商业用电的稳定性。
3. 对于普通工业，在日常情况保障用电是人民生活获得保障。
4. 对于非普工业，根据具体的事宜，对其用电量进行调节，以保障用电量的稳定性，保障生活水平。

# 参考文献