

# 第十届“泰迪杯” 数据挖掘挑战赛

## 优秀作品

作品名称：电力系统负荷预测分析

荣获奖项：一等并网宿创新奖

作品单位：汕头大学

作品成员：缪智伟 王璐 陈灿奕

指导老师：方睿

封面为后期添加，原作品没有此页。

# 电力系统负荷预测分析

## 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、气象条件和社会事件等因素的影响，对未来一段时间的系统负荷做出预测。

首先我们对所给数据进行洗清，包括重复值、异常值、缺失值进行识别，并用 KNN 近邻进行插补。考虑到有停电和跳高的情况，对负荷数据进行平滑处理，使数据更加平稳。日负荷数据具有 96 个时点，我们将以 96 时点的数据特征进行预测分析。

接下来，我们进行了负荷特性分析与特征工程集选择。我们针对地区 96 个时点负荷数据和行业负荷数据进行了负荷特性分析，分析出相邻日的负荷数据和延迟七天的负荷数据具有周期性，节假日的负荷数据比工作日负荷数据较为明显的异常，最高温度和最低温度与负荷数据具有明显的相关性。通过回归模型进行了特征工程集的选择，根据结果，我们剔除了风力风向因素对负荷数据的影响。

我们根据前面的负荷特性分析和特征工程集的结果，本文构建了短期负荷特征工程，考虑了预测日的工作日类型和月份，预测日前一天 96 个时点的负荷数据以及气象因素，预测日同周期延迟七天的所有负荷数据及气象数据。然后我们建立了基于 RBF 神经网络的电负荷预测模型，采用遗传算法确定最优参数。此外我们还将使用最小支持向量机回归 (LS-SVM) 和 LSTM 负荷预测模型，作为比较模型。针对问题一预测未来 10 天每隔 15 分钟电负荷数据，我们的模型计算得到总体绝对百分误差 TAPE 为 0.79%，总体平均准确度为 99.21%

考虑人体舒适度存在的惯性，多日连续高温时，即使温度变化不大，负荷也会增长。通过对预测日前几日的最高温度进行加权后，对待预测日的最高温度进行修正，从而体现出多日连续高温的留存影响。然后建立考虑温度累积效益的 RBF 神经网络中期电负荷预测模型。通过比较计算修正温度与未修正温度与电负荷相关系数，修正后的温度与负荷的相关系数比未修正的高，说明修正温度可以提高温度与负荷的相关性，这样有利于提高预测精度。针对问题一的中期预测，基于不同模型预测结果进行分析与比较，结果计算得到最大负荷预测平均准确度为 97.3%，最小负荷预测平均准确度 98.76%。除此之外，我们对最值负荷时间点的预测结果进行分析。

针对问题二，首先我们使用基于启发式分割算法的原理对不同行业的负荷数据进行突变点检测。我们认为，突变点可分为 3 个量级，分别为节假日突变点、季节性突变点和重大事件突变点，并对四个行业的突变点进行分析。然后，我们使用中期电负荷预测模型对各行业的最大值最小值电负荷进行预测。基于不同模型预测结果进行分析与比较，计算得到 RBF 的中期负荷预测模型预测效果最好。

最后，我们从行业负荷数据突变分析出发，针对各行业用电负荷规模和特性，分析了我国“双碳”目标对于大工业和普通工业等用电规模较大的行业可能产生较大冲击，但对非普工业等用电规模小的行业却可能存在“碳权”交易机遇。以及分析了新冠疫情以及其它突发事件的影响，并对“双碳”目标和新冠疫情下的各行业发展针对性提出多方面建议。

**关键词：**负荷特性分析，基于 RBF 中短期负荷预测，温度累积效应，启发式分割算法突变点检验

## 目录

1. 绪论	4
1.1 问题背景	4
1.2 问题重述	4
2. 数据清洗与预处理	5
2.1 数据清洗	6
2.1.1 重复值检测与处理	6
2.1.2 缺失值检测与处理	6
2.1.3 异常值检测与处理	7
2.2 数据预处理	10
2.2.1 负荷数据平滑	10
2.2.2 负荷数据标准化	10
2.2.3 影响因素数据特征数值化	11
3. 负荷特性分析与特征工程集选择	12
3.1 地区 96 时点负荷特性分析	12
3.1.1 时间特性	12
3.1.2 节假日	15
3.1.3 温度特性	17
3.2 行业负荷特性分析	18
3.2.1 时间特性	18
3.2.2 温度特性	19
3.2.3 工作日特性	19
3.3 基于回归模型特征工程集的选择	20
3.3.1 历史时点负荷因素	20
3.3.2 特征工程集的筛选	21
4. 机器学习模型介绍	23
4.1 RBF 算法	23
4.2 LS-SVMR 算法	23
4.3 LSTM 算法	23
5. 建立中短期电负荷预测模型	24
5.1 短期负荷预测模型	24
5.1.1 短期负荷特征工程构建	24
5.1.2 基于 RBF 神经网络的预测模型	24
5.2 中期负荷预测模型	26
5.2.1 温度累积效应	26
5.2.2 中期负荷特征工程构建	28
5.2.3 考虑温度累积效应的 RBF 神经网络的预测模型	28
5.3 效果评估指标	29
6. 地区负荷的中短期预测分析	31
6.1 地区负荷短期预测与分析	31

6.1.1 使用遗传算法优化的 RBF 神经网络预测模型 .....	31
6.1.2 确定最优参数 .....	31
6.1.3 地区负荷预测结果分析与比较 .....	32
6.2 地区负荷中期预测与分析 .....	36
6.2.1 地区中期最值负荷预测结果及分析 .....	36
6.2.2 地区中期日最大负荷对应时间点预测结果及分析 .....	38
6.2.3 地区中期日最小负荷对应时间点预测结果及分析 .....	39
6.2.4 地区负荷中期预测结果比较 .....	40
7. 行业负荷中期预测分析 .....	41
7.1 突变点分析 .....	41
7.1.1 突变点定义 .....	41
7.1.2 基于启发式的分割算法原理 .....	42
7.1.3 突变点检测过程 .....	43
7.1.4 突变点检测结果分析 .....	44
7.2 行业负荷中期预测与分析 .....	48
7.2.1 行业负荷模型中期预测 .....	48
7.2.2 行业极值负荷预测结果 .....	48
7.2.3 行业极值负荷预测模型比较 .....	50
8. 针对性建议 .....	54
8.1 结论 .....	54
8.2 针对性建议 .....	55
参考文献 .....	57

# 1. 绪论

## 1.1 问题背景

电力系统负荷（电力需求量，即有功功率）预测是指充分考虑历史的系统负荷、经济状况、气象条件和社会事件等因素的影响，对未来一段时间的系统负荷做出预测。负荷预测是电力系统规划与调度的一项重要内容。短期（两周以内）预测是电网内部机组启停、调度和运营计划制定的基础；中期（未来数月）预测可为保障企业生产和社会生活用电，合理安排电网的运营与检修决策提供支持。复杂多变的气象条件和社会事件等不确定因素都会对电力系统负荷造成一定的影响，使得传统负荷预测模型的应用存在一定的局限性。同时，随着电力系统负荷结构的多元化，也使得模型应用的效果有所降低，因此电力系统负荷预测问题亟待进一步研究。

## 1.2 问题重述

### 1. 地区负荷的中短期预测分析

通过挖掘题目所给的负荷数据和天气特征，构建一个具有天气影响因素和历史时点影响因素的特征工程模型，基于此，建议深度学习的中短期的负荷预测模型。首先预测出该地区电网未来 10 天间隔 15 分钟的负荷预测结果，并分析其预测精度；紧接着，预测出该地区电网未来 3 个月日负荷的最大值和最小值预测结果，以及相应达到负荷最大值和最小值的时间，并分析其预测精度。

### 2. 行业负荷的中期预测分析

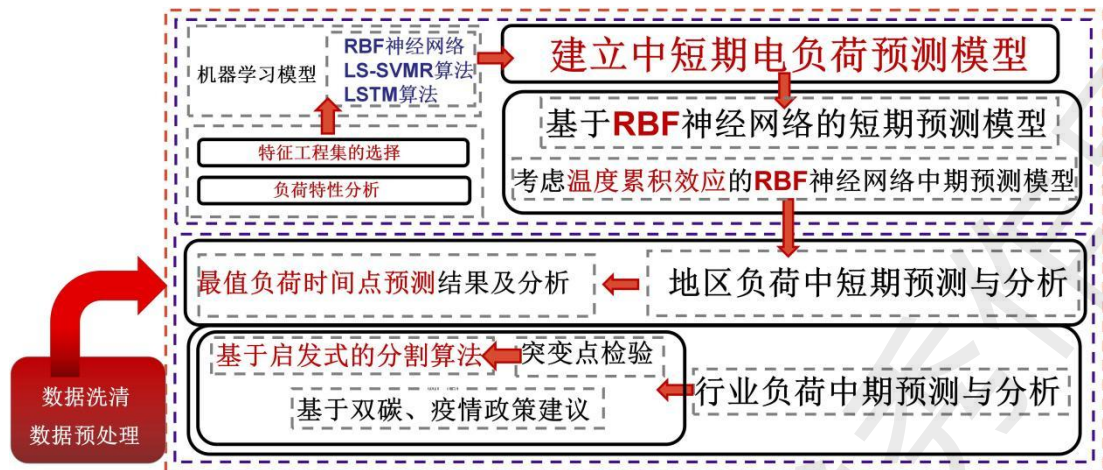
对不同行业的用电负荷进行中期预测分析，能够为电网运营与调度决策提供重要依据。特别是在新冠疫情、国家“双碳”目标等背景下，通过对大工业、非普工业、普通工业和商业等行业的用电负荷进行预测。我们继续根据联合题目所给的负荷数据与气象数据，构建一个具有天气影响因素和历史时点相关的特征工程的中期预测模型。利用该模型分别预测该地区各行业未来 3 个月日负荷最大值和最小值的预测结果，并对其预测精度做出分析。

除此之外，还需要对行业负荷数据的突变点进行识别，并结合天气数据和社会时事热点进行突变点分析。

### 3. 针对性建议

根据各行业的实际情况，结合前文的突变点挖掘，研究国家“双碳”目标对各行业未来用电负荷可能产生的影响，并对相关行业提出有针对性的建议。

针对该问题，我们的模型概述图如下所示。



## 2. 数据清洗与预处理

负荷预测依赖历史数据，而历史负荷数据大都存在误差或缺失数据，这些数据对预测结果影响较大。所以，在进行负荷预测前，需要对误差或缺失数据进行预处理。本文进行数据清洗和预处理的流程图如图 1 所示。

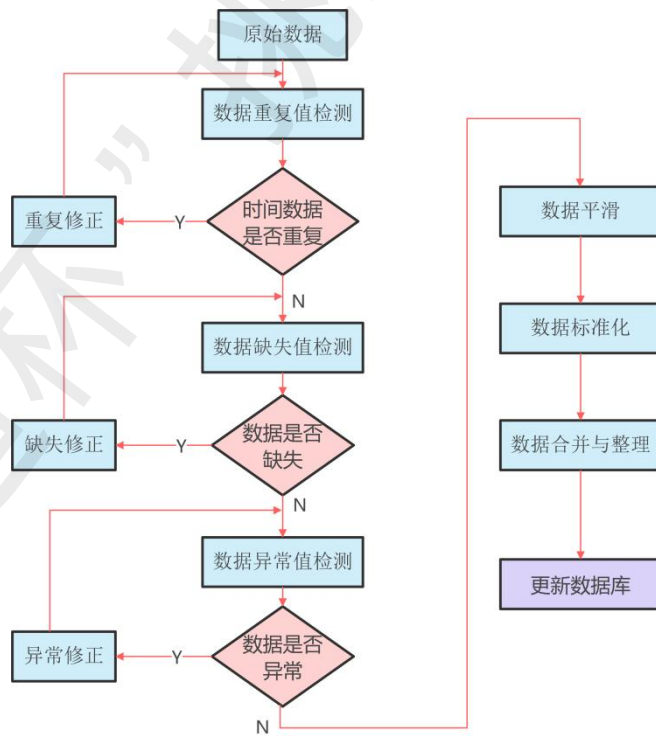


图 1 数据清洗与预处理流程图

## 2.1 数据清洗

### 2.1.1 重复值检测与处理

对于区域 15 分钟负荷数据、行业日负荷数据、气象数据，用 Python 中 pandas 包下 Dataframe 类的 duplicated() 方法，遵照如下规则进行重复值检查：

1. 从头逐行检查完全相同的数据：对于每个数据集，若出现 2 及 2 行以上每列都相同的数据，则称为有重复；
2. 从头逐行检查仅在时间上完全相同的数据：对于区域 15 分钟负荷数据，只要“数据时间”重复的行，便判定为重复；对于行业日负荷数据，只要“行业类型”、“数据时间”同时重复的行，便判定为重复；对于气象数据，只要“日期”重复的行，便判定为重复。

对于重复值，采用直接去除的策略。分析结果显示，区域 15 分钟负荷数据和行业日负荷数据按照两种规则均无重复值。气象数据按照两种规则均有相同的重复行，整理后得到需要去除的重复行如下表 1 所示，去除重复值的气象数据保存为 `气象数据_去重.csv`。

表 1 气象数据中需要去除的重复行

索引	日期
1	2018 年 1 月 1 日
917	2020 年 7 月 4 日
919	2020 年 7 月 5 日
1071	2020 年 12 月 3 日
1255	2021 年 6 月 4 日
1256	2021 年 6 月 4 日

注：气象数据从 1 开始索引

### 2.1.2 缺失值检测与处理

#### 1. 缺失值检测

区域 15 分钟负荷数据、行业日负荷数据、气象数据均为带有时间戳的数据集，所以对于缺失值有两个方面的检查。其一是普通的数据框缺失值检查，仅检查所有位置的数据是否有空缺；其二是时间序列的时间点缺失检查，即针对时间戳的前后连续性进行检查。

对于普通缺失值的检查，首先使用 Dataframe 类的 isnull().any() 两个方法链式调用，结果显示三个数据集均无普通缺失值。

对于时间序列的时间点缺失的检查，可以通过遍历时间戳来进行检查具体缺失时间。下面是对三个数据集检查的相关参数和结果。

表 2 时间点缺失检查记录

数据集	开始时间	结束时间	时间间隔	已有时间点个数	缺失时间点个数
区域 15 分钟 负荷数据	2018/1/1 00:00:00	2021/8/31 23:45:00	15 分 钟	128156	388
气象数据	2018 年 1 月 1 日	2021 年 8 月 31 日	1 日	1339	0
行业日负荷数 据（大工业用 电）	2019-01-01	2021-08-31	1 日	973	1
行业日负荷数 据（非普工业）	2019-10-10	2021-08-31	1 日	691	1
行业日负荷数 据（普通工业）	2019-01-01	2021-08-31	1 日	973	1
行业日负荷数 据（商业）	2019-01-01	2021-08-31	1 日	973	1

如上表 2，区域 15 分钟负荷数据缺失了 388 个时间点，气象数据没有缺失时间点，而行业日负荷数据中，四个行业均缺失了 2019 年 1 月 1 日的日负荷数据。

## 2. 缺失值处理

时点负荷数据的缺失主要是由于采集器发生故障或进行检修时设备暂停导致的数据缺失。面对缺失的负荷数据我们将根据历史近似日的时点负荷数据采用 KNN 最近邻插补，这基于一个假设，即排除其他外在因素的干扰的情况下，地区负荷的用电规律应该是一致的。

如上过程，我们使用 R 语言编写了 seqKNN，nmiss 对其进行了实现，其中函数 seqKNN 的实现，依赖于 nmiss。

通过找出 K 近邻样本，并根据距离确定权重，求取平均值对缺失值进行插补。

### 2.1.3 异常值检测与处理

如图 2 所示，通过分析题目所用的地区 96 点负荷数据并进行可视化后发现，负荷数据中存在较为明显的、异常的低负荷值，需要对这部分数据进行修正。此外，发现 2018 年 1 月到 8 月这段时间的负荷波动趋势相比后面三年出现断层现象，属于异常值。我们通过研究已给数据的分析，发现并非是已给数据造成的影响。因此我们假设可能是因为政策或地区管辖范围缩小等其他非自然因素导致 2018 年 9 月份之前的地区负荷数据相比往后数据有明显断层。从电负荷波动趋



势来看，该地区最近三年的电负荷波动趋势范围大致稳定，因此，为了减少该部分数据对负荷数据影响，我们将 2018 年 9 月份之前的数据去除，只考虑 9 月份开始往后的数据，如图 3 所示。

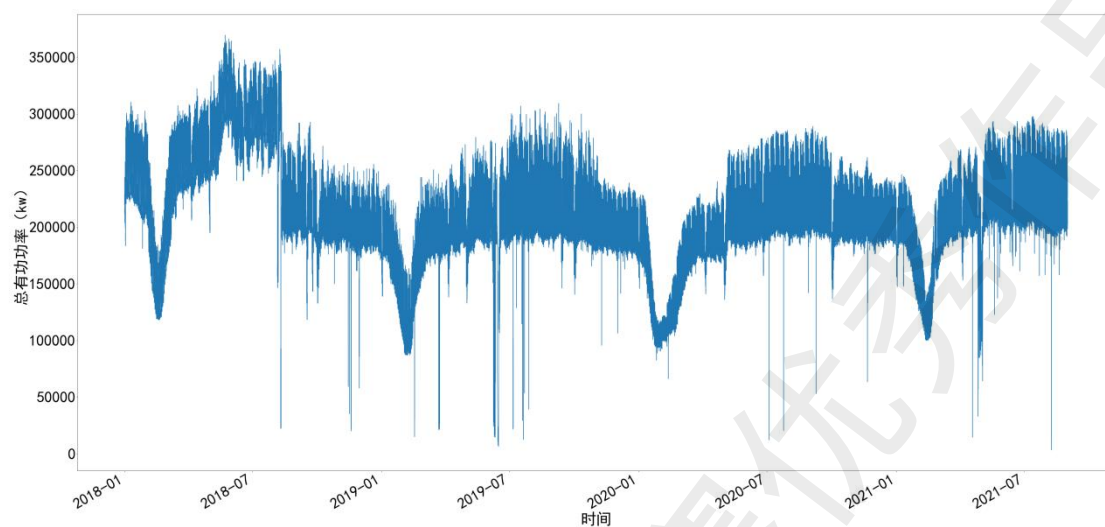


图 2 原始数据电负荷时间关系图

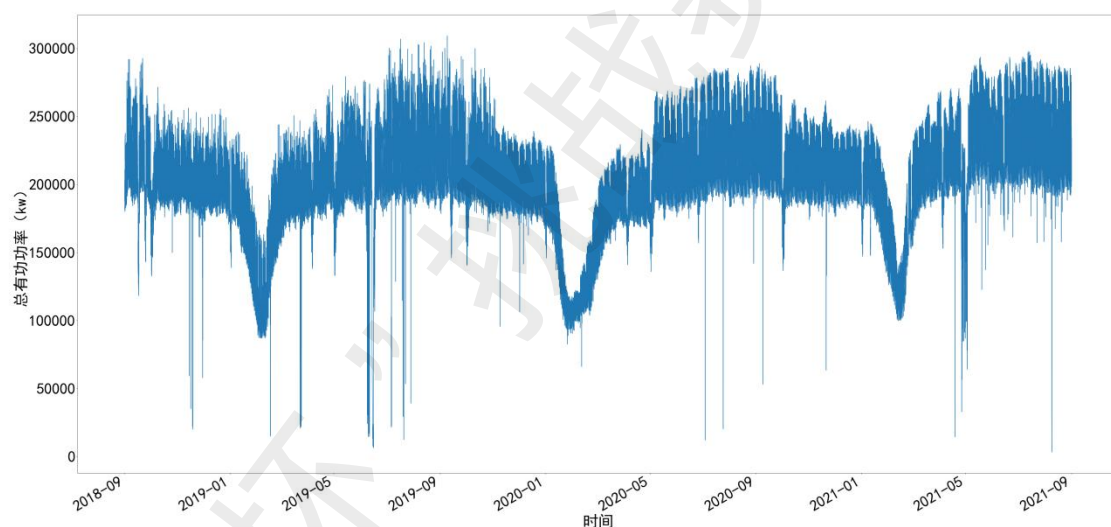


图 3 2018 年 9 月份开始电负荷时间关系图

从图 3 中可以看出，该地区出现三个低谷点。我们通过查阅谷点时间发现正是每年的春节，除此之外，我们发现法定节假日都会出现电负荷降低后上升的趋势，因此我们推导该地区电负荷可能受节假日影响较大。如果将节假日考虑进来进行异常值检测，很可能会导致节假日负荷被视为异常值。为了充分保留节假日所带来的电负荷变化，我们会人工进行筛选出节假日影响电负荷变化的时间段，然后对平常日进行异常值检测。

异常值检测常用方法有简单统计分析、3sigma 原则、箱线图、聚类算法等，针对平常日异常值我们选择箱线图进行异常值检测。箱线图是一种统计图，是通

过最小值、下四分位数、中位数、上四分位数和最大值来描述数据，它可以呈现数据大致的分布信息，这也是它的一个很大的优点，它还可以从宏观上展现总体样本的分布信息，同时发现异常值。

利用箱线图进行异常值检测时，考虑到不同时点可能会有不同的电负荷波动范围，因此我们将 96 时点分开检测其异常值，如图 4 所示。

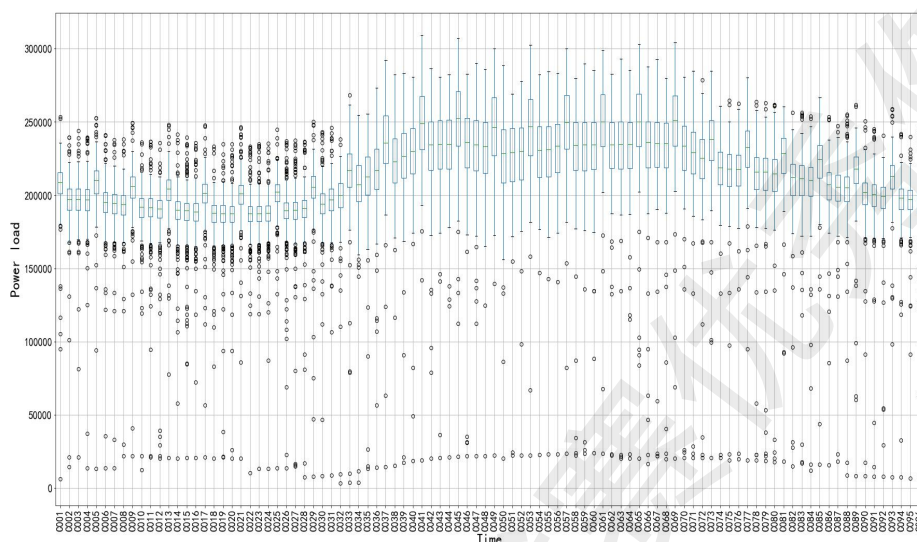


图 4 非节假日 96 时点箱线图

从图 4 中可以看出，不同时点均存在异常值。我们删除检测到的异常值，并视为缺失值进行 K 近邻插补处理，处理后结果如图 5 所示。可以看出 K 近邻插补法基本修正了检测到的异常值，这也侧面体现了我们选择的插补方法的有效性。

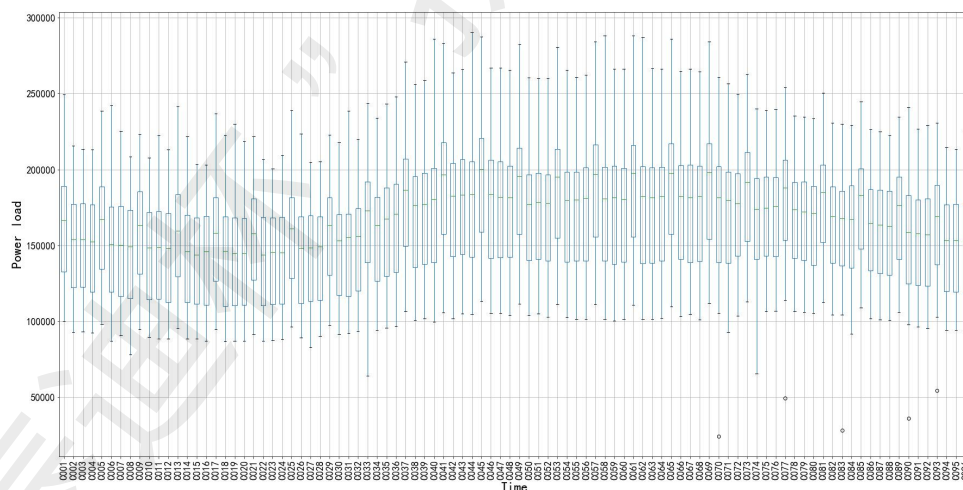


图 5 异常值修正后非节假日 96 时点箱线图

## 2.2 数据预处理

### 2.2.1 负荷数据平滑

通过下面的图我们可以看到，蓝色的线代表未平滑的地区负荷数据，可以明显地看到部分日负荷曲线出现了剧烈的波动，有的甚至直接降到 0，比如停电的情况，有时候突然上升到很高的负荷值。那么总的来说，对于这类情况需要消除离群点的影响，即停电和跳高的情况，另外还要对波动较大的日负荷曲线进行平滑处理。

如图 6 所示，橙色的线是数据经过平滑后的。经过平滑处理后，一些异常低的负荷数据被向上收缩，一定程度上改善了日负荷曲线的稳定性。

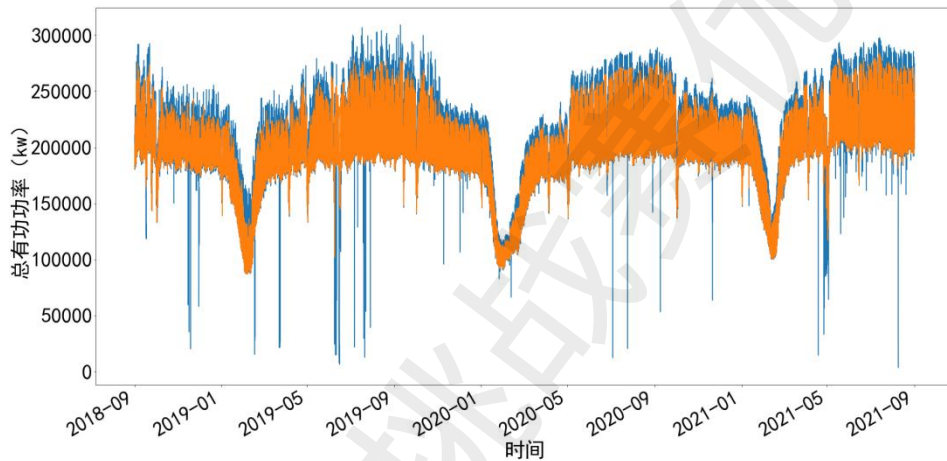


图 6 数据平滑前后对比图

### 2.2.2 负荷数据标准化

鉴于神经网络在训练过程中需要保证节点的输出值在激活函数的适合范围内，而电力负荷数据波动性较大，需要对历史负荷数据进行标准化处理。因此，我们在模型预测输入的数据时会对进行均值方差标准化。

其中，均值方差标准化是一种将数据转化为标准正态分布的标准化方法。在回归模型中，服从正态分布的自变量和因变量往往对应着较好的回归预测效果。均值方差标准化的计算公式为：

$$x' = \frac{x - \mu}{\sigma}$$

公式中， $x$  表示单个数据的取值， $\mu$  表示对应列的均值， $\sigma$  表示对应列的标准差。

## 2.2.3 影响因素数据特征数值化

### 1. 气象数据特征数值化

初步分析可以看到，气象数据多为文本类型数据，例如天气情况完全由文本进行表述。未来方便分析和利用气象数据，我们将气象数据进行了如下处理：

#### (1) 风力特征数值化

白天和夜晚风力方向中的风力等级可以经过一定匹配和处理，分别得到白天和夜晚的得到最大和最小风力级别。此级别遵照中国气象局划分的风力等级标准。

#### (2) 天气状况数值化

对天气状况进行切分，可以得到每日最好天气状况和最坏天气状况，例如“晴/小雨”分割后最好为“晴”，最坏为“小雨”。原本只有一个天气状况的则最好最坏天气相同。对分割后的所有天气统计得到 17 个天气状况，进行分组数值化编码，具体如下表。

表 4 分割后的天气状况和对应数值编码

天气	数值编码	天气	数值编码	天气	数值编码
晴	5	阵雨	4	小雪	2
晴间多云	5	雷阵雨	4	中雪	2
局部多云	5	小雨	4	暴雨	1
多云	5	中雨	3	大雪	1
雾	4	雨夹雪	3	暴雪	1
阴	4	大雨	2		

我们对分级后的天气进行打分，天气越好评分越高，这是因为夏季天气相对较好，且夏季负荷用量多一些。这样尽量减小组内的差距，而组与组之间则有明显的差异的做法，可以使数量较多，次序较为模糊的天气状况数据变为有序、层次分明的数据。而且每组能体现不同程度的天气状况特征。

#### (3) 温度数值化

最高温度和最低温度比较好处理，分析得知温度数据中只需去除摄氏度单位符号即可。

最终经过处理的气象数据保存为 `气象数据_特征数值化.csv`，特征说明如下表。

表 5 气象数据特征数值化处理后的数据字典

列名	数据说明
日期	%Y/%m/%d 格式的日期
最高温度	整数类型的温度值
最低温度	
白天最低风力	整数类型风力等级，范围为[0,12]

白天最高风力	
夜晚最低风力	
夜晚最高风力	
最坏天气值	整数类型的天气状况编码，范围[1,5]
最好天气值	

## 2. 节假日特征数值化

根据后面 3.1.2 节对节假日的特性分析，我们考虑不同节假日对负荷的影响，将节假日类型分为三类，分别为：大长假、小长假、短假，并依次用数值 1,2,3 量化。不同法定节假日的归类以及量化结果如表 6 所示。

表 6 节假日分类与量化

节假日类型	数值编码	节假日
大长假	1	春节
小长假	2	国庆节
短假	3	端午节、劳动节、清明节、中秋节
平常日	0	非节假日

## 3. 负荷特性分析与特征工程集选择

### 3.1 地区 96 时点负荷特性分析

电力系统负荷会随时间不断地变化，但这种变化不是毫无规律随机变化的，而是有一定的规律可循的，同时也受特定的因素影响的。具体来说，负荷的变化主要具有周期性和连续性，同时受多种因素影响，如经济水平和结构调整等，其中对负荷影响最为显著的是气象因素，同时一些特殊事件（如大型赛事、新冠疫情）对负荷也有较大的影响。在进行中短期负荷预测时，为保证预测的精度需要充分考虑负荷特性。下面我们会分成三个特性来挖掘地区 96 时点的负荷特性分析。

#### 3.1.1 时间特性

负荷变化规律实际上与人的生产和生活规律密切相关，而人们的生产和生活

规律具有周期性，具体表现为人们的工作和休息相互交替。因此，负荷在一段时间内呈现出重复变化，即周期性。但负荷变化并不是单一的周期，而是以日、周和年为周期的。

### 1. 日周期性

电力负荷以 24 小时为周期变化所体现出来的连续重复的规律称为电力负荷的日周期性。负荷的日周期性与人们的生产、生活习惯密切相关，现以该地区 96 时点中选取两天负荷进行分析。

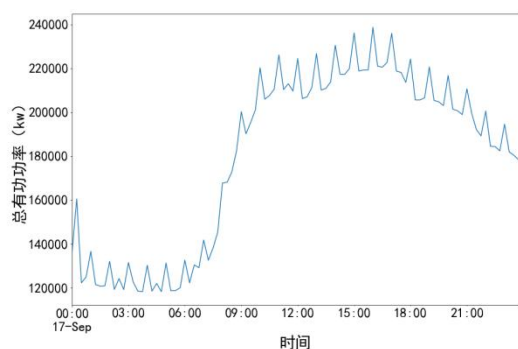


图 7 2018 年 9 月 17 日日负荷

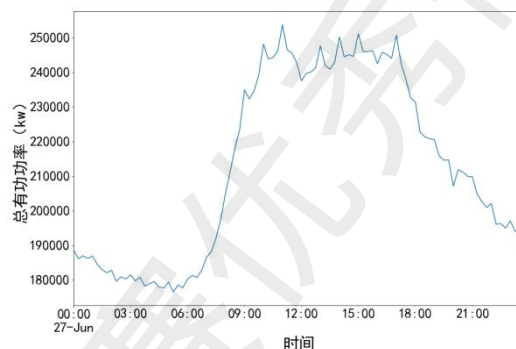


图 8 2019 年 7 月 27 日日负荷

如图 7-8 所示，具有典型的一峰两谷的特性，即有早上到下午一个高峰以及凌晨低谷和夜晚低谷。高峰一般出现在 10 点-12 点左右，早高峰和午高峰是因为工厂大规模生产和人们工作、学习引起的。之后会出现一定的负荷小低谷，是因为午休负荷下降导致。当日谷值在凌晨 3 点-6 点左右出现，是因为此时人们休息，一些照明、电视等电器设备停止使用。无论居民用电负荷、商业用电负荷都达到最低点，此时电力负荷主要是由一些连续生产的企业（如炼钢厂等）负荷构成的。

人们每天的生产和生活安排已经形成一种习惯，因此，每天的负荷曲线都是相似的。

### 2. 周周期性

电力负荷以 7 天为周期变化所表现出来的连续重复的规律称为电力负荷的周周期性。周一至周五（称为工作日），人们都在工作，大部分工厂都在生产；而在周末，大部分人都在休息，大部分工厂放假。由于电力负荷主要是由工业负荷构成的，因此，在工作日的负荷会明显高于周末的负荷。另外，在工作日，工厂的生产是具有规律性的，每天生产安排较为相似，因此，不同工作日之间有较高的相似性，如图 9 所示。



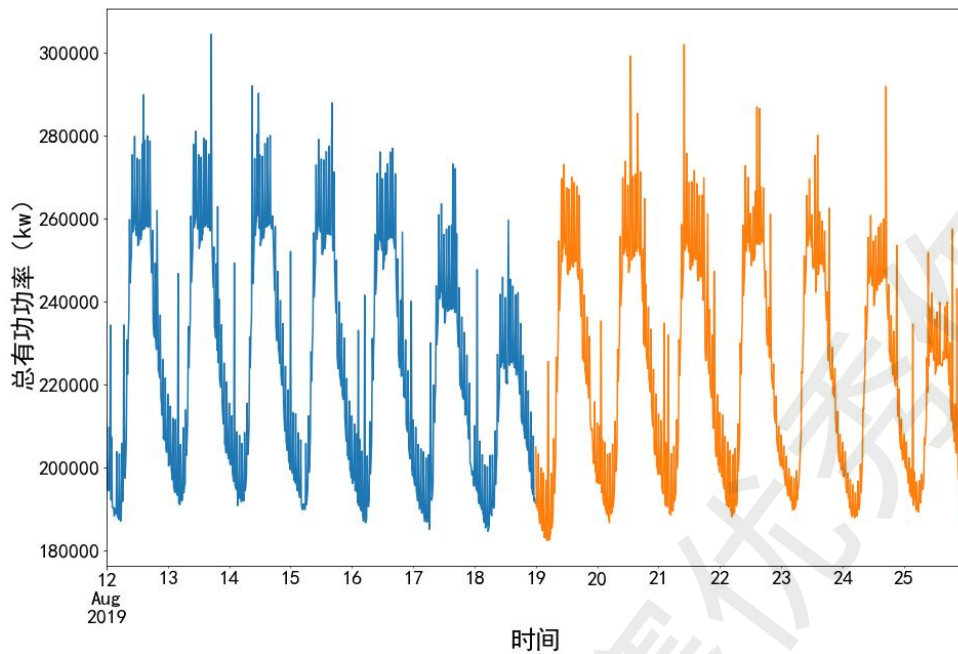


图 9 2019 年 8 月 12 日-8 月 25 日负荷曲线图

另外，为了进一步工作日分析与周末负荷的区别，我们选择第 40、50、60、70 个时点，按周一到周日的顺序绘制每种工作日类型的平均负荷水平，得到二维图表，如图 10 所示。从图中可以发现，周末的电力负荷相较于工作日会有明显下降趋势。而周一和周五由于邻近周末，也会表现得与周二到周四的用电规律有细微差别。我们初步判断每个时点在周末与工作日的用电水平有着明显区别，因此，工作日类型是我们建模考虑的一个维度。

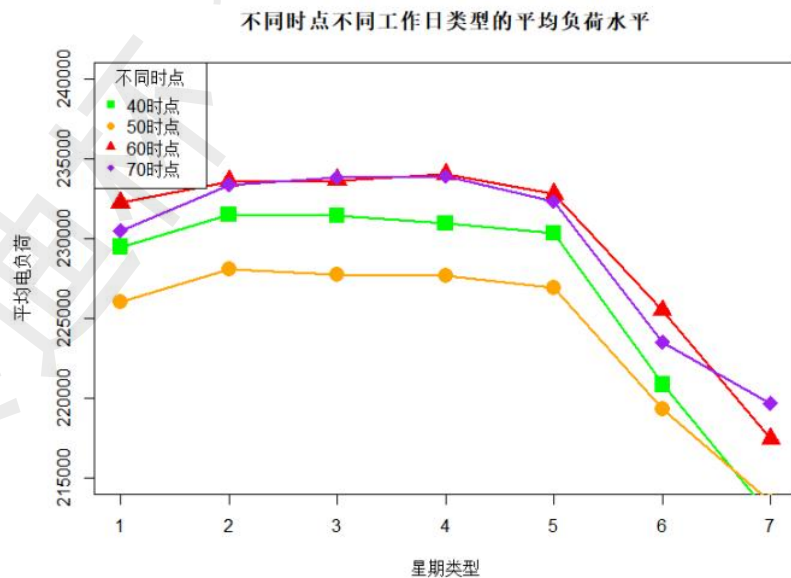


图 10 不同时段不同工作日类型平均负荷曲线图

### 3. 年周期性

电力负荷以 1 年为周期变化所表现出来的连续重复的规律称为电力负荷的年周期性。这主要是由于季节变化的周期性引起的。具体来说，在夏季，由于天气炎热，人们开始使用空调降温，空调负荷迅速增加；而在冬季，温度较低，基本不存在空调负荷，当温度降低到一定程度时，居民大量使用取暖设备，相应的取暖负荷大幅增加；在春季和秋季则基本没有空调负荷和取暖负荷。因此，负荷含有以年为周期变化的分量[1]。

下面我们选取地区 96 时点负荷，画出时间趋势图，如图 11 所示。

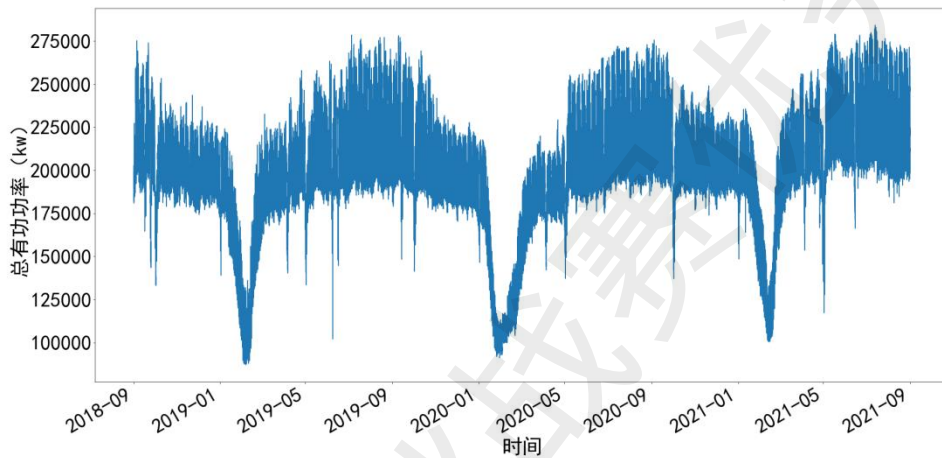


图 11 地区 96 时点负荷预测

整体上年负荷特性呈现季节性。夏季平均负荷相较春季负荷要高，是因为受到气温升高影响，天气炎热消暑负荷大幅增加。而冬季平均负荷相较夏季负荷较低，可以看出该地区冬季取暖负荷低于夏季空调负荷。因此，季节性负荷在该地区还是较为明显的，我们认为在负荷预测时也要考虑季节因素。

除了季节性的周期变化，我们还发现每年 2 月前后负荷存在最低谷趋势，经查询该时段为春节期。因此，我们认为节假日也会较大程度的影响负荷变化，需再进行详细分析。

#### 3.1.2 节假日

由前面年周期特性的负荷曲线可看出，节假日整体负荷量较平日相比平均负荷量较少，因为节假日工业负荷大幅降低。虽然节假日由于居民用电及商业用电负荷有所增加，但由于工业用电负荷所占该区域负荷总耗电量的比例很大，相比居民用电及商业用电负荷增加量，工业用电负荷量降低量对总负荷量影响显著，这是节假日整体负荷量比平日低的主要原因。

下面是来源于国务院对于节假日公休安排的通知：



## 2018年：

- 一、元旦：1月1日放假，与周末连休。
- 二、春节：2月15日至21日放假调休，共7天。2月11日（星期日）、2月24日（星期六）上班。
- 三、清明节：4月5日至7日放假调休，共3天。4月8日（星期日）上班。
- 四、劳动节：4月29日至5月1日放假调休，共3天。4月28日（星期六）上班。
- 五、端午节：6月18日放假，与周末连休。
- 六、中秋节：9月24日放假，与周末连休。
- 七、国庆节：10月1日至7日放假调休，共7天。9月29日（星期六）、9月30日（星期日）上班。

## 2019年：

- 一、元旦：2018年12月30日至2019年1月1日放假调休，共3天。2018年12月29日（星期六）上班。
- 二、春节：2月4日至10日放假调休，共7天。2月2日（星期六）、2月3日（星期日）上班。
- 三、清明节：4月5日放假，与周末连休。
- 四、劳动节：5月1日放假。
- 五、端午节：6月7日放假，与周末连休。
- 六、中秋节：9月13日放假，与周末连休。
- 七、国庆节：10月1日至7日放假调休，共7天。9月29日（星期日）、10月12日（星期六）上班。

## 2020年：

- 一、元旦：2020年1月1日放假，共1天。
- 二、春节：1月24日至30日放假调休，共7天。1月19日（星期日）、2月1日（星期六）上班。
- 三、清明节：4月4日至6日放假调休，共3天。
- 四、劳动节：5月1日至5日放假调休，共5天。4月26日（星期日）、5月9日（星期六）上班。
- 五、端午节：6月25日至27日放假调休，共3天。6月28日（星期日）上班。
- 六、国庆节、中秋节：10月1日至8日放假调休，共8天。9月27日（星期日）、10月10日（星期六）上班。

## 2021年：

- 一、元旦：2021年1月1日至3日放假，共3天。
- 二、春节：2月11日至17日放假调休，共7天。2月7日（星期日）、2月20日（星期六）上班。
- 三、清明节：4月3日至5日放假调休，共3天。
- 四、劳动节：5月1日至5日放假调休，共5天。4月25日（星期日）、5月8日（星期六）上班。
- 五、端午节：6月12日至14日放假，共3天。
- 六、中秋节：9月19日至21日放假调休，共3天。9月18日（星期六）上班。
- 七、国庆节：10月1日至7日放假调休，共7天。9月26日（星期日）、10月9日（星期六）上班。

由以上四个通知可知，全国公休假日安排一般将邻近的周末与节假日搭配安排，并形成了2个“七天假期”（春节、国庆）及5个“3天假期”（元旦、清明、

劳动、端午、中秋)。经过分析,我们发现除了春节之外,其他节假日对负荷的影响基本与放假天数相同,并且会在放假前一天电负荷开始下降。对于春节假期,对该地区负荷的影响不止7天假期,从电负荷开始下降到回升至同平日一样的负荷,其间隔从2018年-2021年来看都需要一至两个月。因此,在考虑节假日对负荷变化的影响时,我们需要考虑不同节假日对负荷变化的影响。

### 3.1.3 温度特性

从前面的时间特性分析可以得到,负荷随时间变化是具有周期性的,但是在周期性中又包含随机性,这是由一些因素引起的,而在这些因素中对负荷影响最大的是气象因素。影响负荷的气象因素主要包括温度、降雨量和台风等。利用附件2数据,我们对日最高温度、日最低温度与日平均负荷进行特性分析。我们以2019年、2020年为例,描绘日最高温度、最低温度与日平均负荷之间的关系,如图12-13所示。



图 12 2019 年温度与日平均负荷曲线图

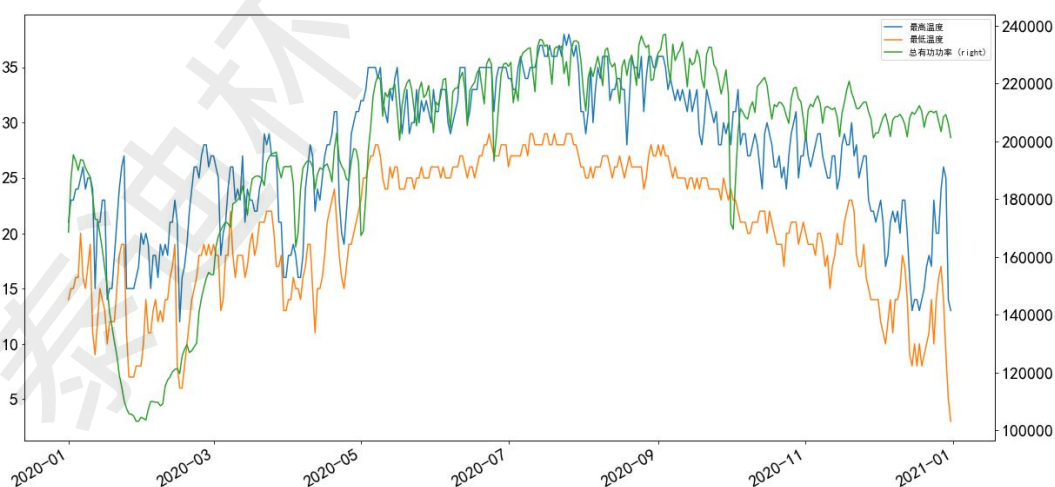


图 13 2020 年温度与日平均负荷曲线图

注:绿色线为日平均负荷,蓝色线为日最高温度,红色线为日最低温度  
左边纵坐标轴刻画温度,右边纵坐标轴刻画总有功率

由图 12 和图 13 可以看出，日最高温度与日最低温度的变化趋势大致一致：当最高温变大时，最低温也变大；最高温变小时，最低温也变小。因此可以大致判断日最高温度和日最低温度之间具有强相关性。

温度对负荷的影响是显而易见的，在夏季，随着气温升高，负荷也会上升，上升的这部分负荷主要是由降温负荷（也称空调负荷）构成的。在冬季，当气温降低时，降温负荷不再需要，日平均负荷会降低，由于该地区日最低温度范围在 $[5, 25]$ 之间，因此我们推测该地区应为四季宜人的南方地区，对南方地区来说冬季用大量取暖设备取暖的需求较低，因此取暖负荷在该地区也较少体现。随着人民生活水平的提高，空调已经成为人们生活的重要组成部分，相应的负荷也随之增加，负荷对温度的灵敏度也会增加，这将导致温度对负荷的影响越来越显著[1]。因此，在选择和建立短期负荷预测模型时，为保证负荷预测精度，必须要考虑温度的影响。

## 3.2 行业负荷特性分析

不同行业类型的负荷其变化特点也不尽相同，了解不同行业类型负荷的变化特点及其影响因素，对于负荷预测的研究工作有着重要的意义。

### 3.2.1 时间特性

以大工业区为例进行时间特性分析，如图 13-14 所示。

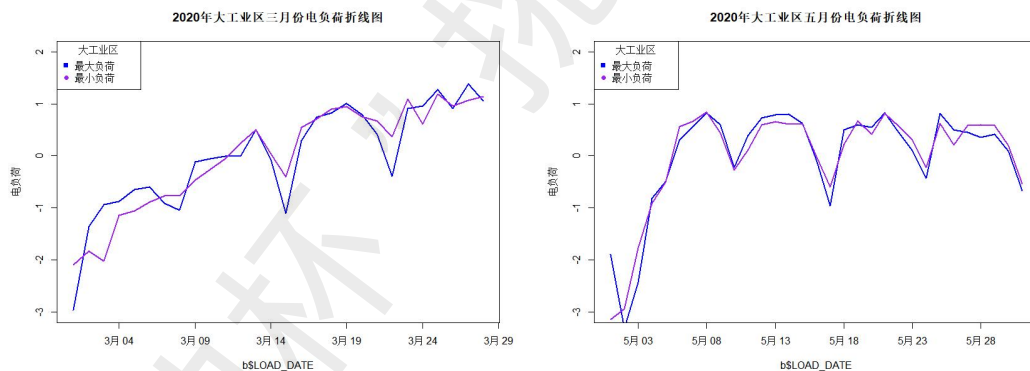


图 13 2020 年 3 月大工业区负荷曲线图 图 14 2020 年 5 月大工业区负荷曲线图

由上图我们可以分析出大工业区相邻日的最值负荷相差不是很大，都会在一定范围内波动。其次我们也可以观察出每隔一周，同一工作日类型的最值负荷具有一定的周期性，所以在接下来为了预测未来时点的最值负荷，我们将考虑到延迟一阶和延迟七阶的日负荷特性作为特征因子。

### 3.2.2 温度特性

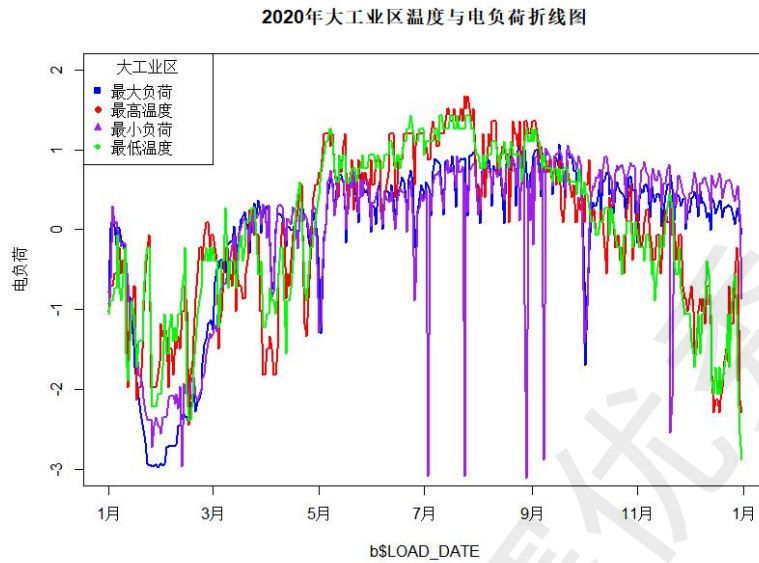


图 15 2020 年大工业区温度与电负荷折线图

通过上图我们可以观察出最大值负荷与最大值温度有明显的趋势相关性。所以我们将温度纳入特征工程里面。

### 3.2.3 工作日特性

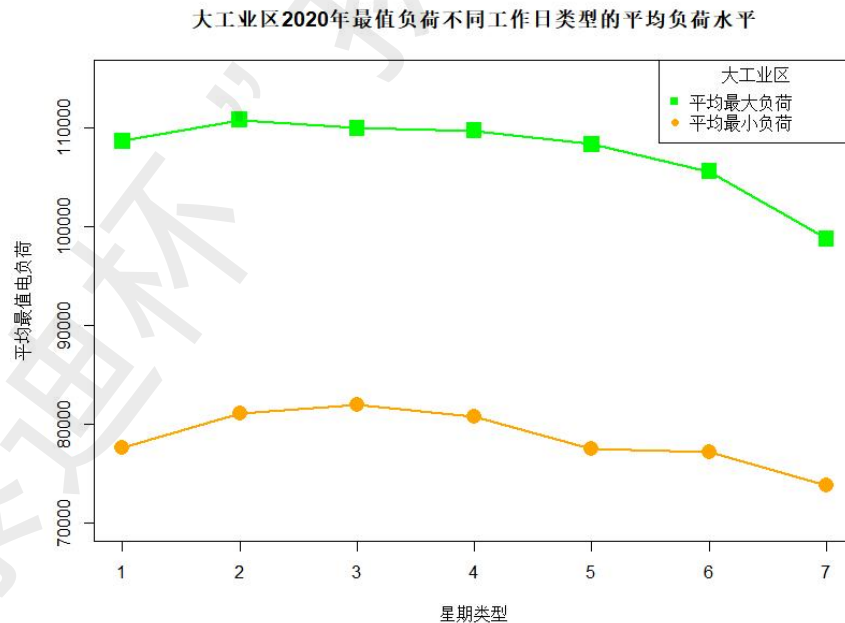


图 15 2020 年大工业区最值负荷平均负荷折线图

通过上图我们可以观察出不同的工作日类型，大工业区的负荷水平有一定区

别，特别是工作日和周末的平均最大值负荷有明显区别，周末的最值负荷水平都较低，而工作日比较高。

### 3.3 基于回归模型特征工程集的选择

#### 3.3.1 历史时点负荷因素

首先，我们分析了不同时点的自相关性，发现每个时点的自相关性差别很大，我们给出了第 10、40、60、80 个时点，对应的偏自相关分析图，如图 16(a)-(d) 所示。

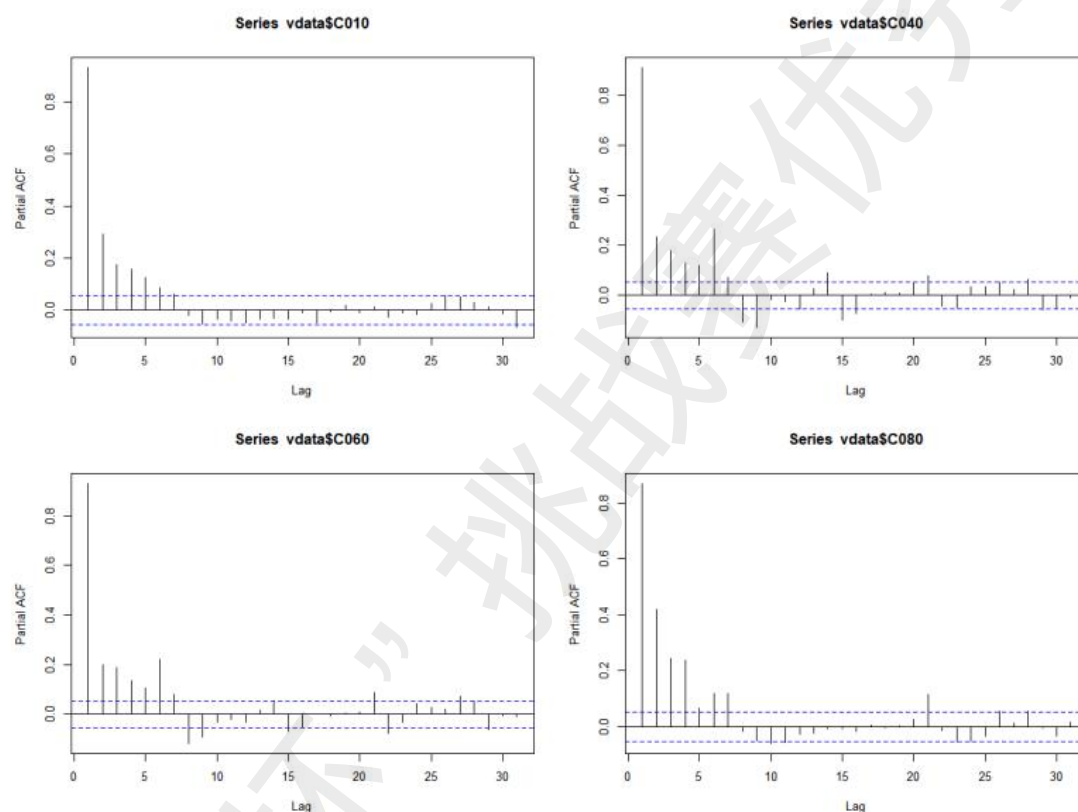


图 16(a)-(d) 第 10、40、60、80 个时点对应的偏自相关分析图

从上图中我们可以知道，这四个时点在不同的阶数截尾，因此可以建立不同阶数的自回归模型也就是说可以使用对应阶数的历史时点负荷构建预测特征。

此外，还可以分析不同时点之间的相关影响及延迟影响，这里给出了第 40 个时点分别与第 10、50、60、80 个时点的交互影响关系图，如图 17(a)-(d)所示。

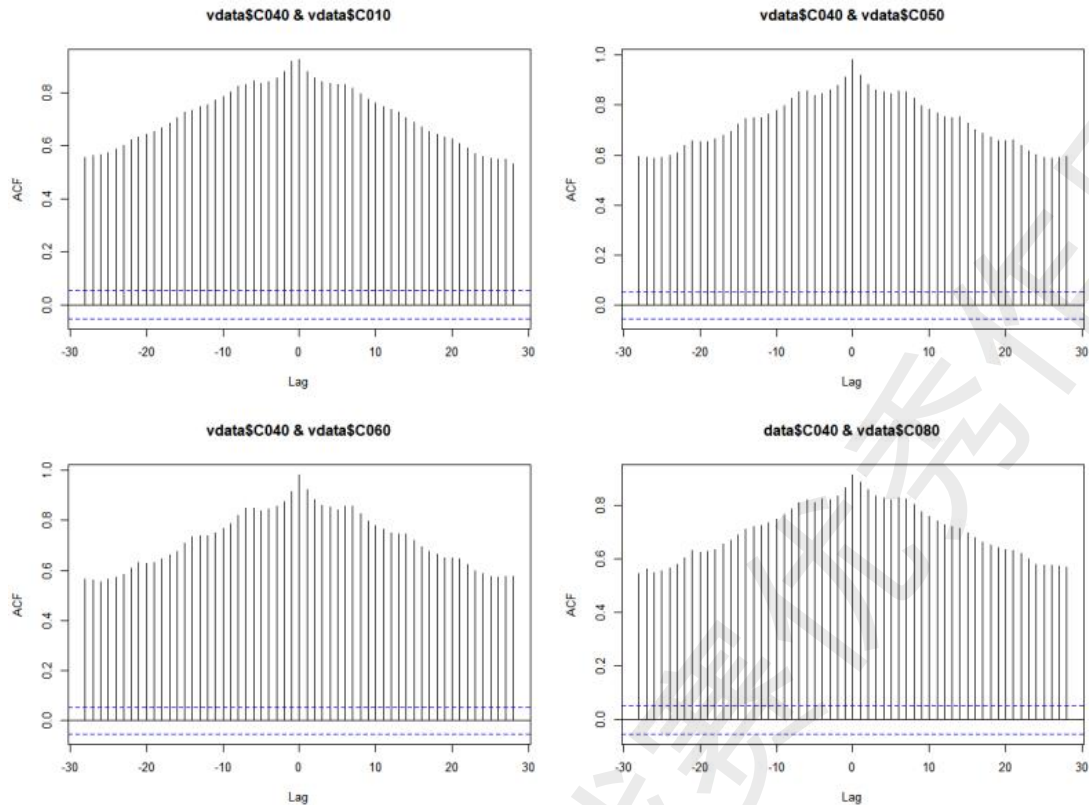


图 17(a)-(d) 第 10、40、60、80 个时点对应的交互影响图

由图 17(a)-(d)可知，第 40 个时点与这四个时点的影响关系非常对称， $Lag=0$  表示没有延迟，当期的相关性，在这四个关系图中都是  $Lag=0$  时相关关系最强，但是不能直接用于预测。我们可以使用第 10、80 个时点延迟一阶的值来构建特征，因为它与这两个时点延迟一阶的值相关性较强，同时使用第 50、60 个时点延迟一阶及一周的值来构建特征，如此等等，为所有时点构建起用于建立预测模型的特征集。

### 3.3.2 特征工程集的筛选

前面已对负荷特性进行了详尽的分析，分析结果表明电力负荷与温度、天气值、工作日类型、月份、是否节假日较为紧密，因此，负荷预测模型应考虑日最高温度和日最低温度、最高天气值、最低天气值、工作日类型、月份等作为输入变量。此外，由于数据集中还给了风力级别，但是为了接下来的负荷预测，我们需要确定具有较大影响力的特征，所以我们将考虑这些因素的影响，先将负荷数据按时段分组，形成 96 个负荷子序列，再对各个子序列分别建模。通过回归模型的方法选择对电负荷较有解释力的特征。基于上述考虑，可以对负荷建立如下模型：

$$P_{k,t} = a_t T_{k,1} + b_t T_{k,2} + c_t W_{k,1} + d_t W_{k,2} + e_t D_k + f_t M_k + g_t L_k + h_t F_k + \mu_{k,t} \quad (1 \leq k \leq n, 1 \leq t \leq 96)$$

式中， $n$  为历史样本天数； $T_{k,1}$ 和 $T_{k,2}$ 分别为历史样本中第  $k$  天的最高温度和最低温度； $P_{k,t}$ 为第  $k$  天时段  $t$  的负荷序列； $W_{k,1}$ 和 $W_{k,2}$ 分别为天气的最高天气值和最低天气值； $D_k$ 为第  $k$  天的工作日类型标识； $M_k$ 为月份， $L_k$ 为风力级别， $F_k$ 为是否为节假日； $\mu_{k,t}$ 为由其它各种因素引起的第  $k$  天时段  $t$  负荷的随机波动，回归常数由可以通过最小二乘求得。

接下来我们将对自变量进行筛选。最基本的选择准则是残差平方和 (Predicted Residual Sums of Squares, PRESS) 准则，这里采用 PRESS 准则对自变量进行筛选，它是衡量模型对样本数据推广能力的强弱的指标[2]。

自变量集筛选的基本过程是建立所有可能的回归方程，并用 PRESS 准则对各个模型进行检验，可以得到最优的回归方程。该方法的基本思想是：首先从  $n$  个样本数据中选择一个样本数据作为待预测的数据，然后对剩下的  $n-1$  个样本数据建立回归方程，最后用回归方程对待预测的数据进行预测，得到预测值，并将该预测值与样本实际值进行比较，得到预测残差。给定样本集的 PRESS 值定义为：

$$PRESS = \frac{1}{n} \sum_{t=1}^n \rho_t^2 = \frac{1}{n} \sum_{t=1}^n (y_t - \tilde{y}_t)^2$$

式中， $y_t$ 为样本中第  $t$  个数据，是被剔除并作为待预测数据； $\tilde{y}_t$ 采用回归方程获得的样本中第  $t$  个数据的预测值。

现以 2018 年 9 月 1 日至 2021 年 8 月 31 日的地区负荷数据为例，进行自变量筛选：将负荷、日最高温度、日最低温度和星期类型等标识进行线性变换，将其映射到[-1,1]区间。

计算各个回归方程的预测残差平方和。将所有特征进行随机组合，以负荷为因变量，分别以各个自变量集中的变量为自变量，建立回归方程，并计算各个回归方程的预测残差平方和，结果如表 7 所示。

表 7 回归方程的预测残差平方和

自变量集	预测残差平方和	自变量集	预测残差平方和
T1,T2,W1,L	5.32E+08	T1,T2,W1,W2,D,M	2.13E+09
T1,T2,W1,W2	9.83E+08	T,T2,W1,W2,D,M	2.34E+09
T2,W1,W2,D,M	2.13E+09	T1,T2,W1,W2,D,M,F	5.20E+06



由表 7 可知，以日最高温度、日最低温度和工作日类型、月份、是否节假日、天气值作为回归模型的自变量，预测残差平方和最小，即得到最优回归方程。这些特征将纳入接下来的特征工程。

## 4. 机器学习模型介绍

### 4.1 RBF 算法

RBF 神经网络是一种局部逼近的神经网络，它可以逼近任意的非线性函数，能够处理系统内的难以解析的规律性，具有良好的泛化能力，并有很快地学习收敛速度，同时由于 Cover 定理保证了 RBF 神经网络在数学上的相关合理性，目前已成功应用于非线性函数逼近、时间序列分析、数据分类、模式识别、信息处理、图像处理、系统建模、控制和故障诊断等。

RBF 神经网络的基本思想是：用径向基函数作为隐单元的“基”，构成隐含层空间，而输入层直接传递输入信号到隐含层，这样就可以将输入矢量直接（即不通过权值连接）映射到隐空间，使得隐含层单元输出即为网络输入的径向基函数映射（是非线性的），这种映射关系在 RBF 的变换中心确定之后也自然就确定了；另一方面，由隐含层空间到输出空间的映射是线性的，即网络的输出是隐含层单元输出的线性加权和，此处的权即隐含层单元与输出层单元的连接权，是网络的可调参数。由此可见，从总体上看，RBF 网络由输入到输出的映射是非线性的，而网络输出对可调参数而言却又是线性的，这样网络的权值就可以用最小方差算法、递推最小二乘法等计算，从而大大加快了学习速度并避免了局部极小问题。

### 4.2 LS-SVMR 算法

LS-SVMR 即最小二乘支持向量机回归，它是为了便于求解而对 SVM 的一种改进。与标准 SVM 相比，LS-SVMR 用等式约束替代 SVM 中的不等式约束，求解过程变成一组等式方程，避免求解耗时的二次规划问题，求解速度相对较快。

### 4.3 LSTM 算法

长短期记忆人工神经网络是一种改进的时间循环神经网络。LSTM 可以学习时间序列长短期依赖信息，由于神经网络中包含时间记忆单元，因此适用于处理和预测时间序列中的间隔和延迟事件。



LSTM 对于处理时间序列相关的数据有很好的效果。可以以负荷数据自身为训练数据和输出标签，通过迭代训练的方法建立负荷预测模型，该方法可充分利用负荷数据之间的相关性，从而提高负荷预测的精度，同时降低所需历史数据的维度大小。最后通过实际数据进行仿真实验，并与传统的负荷预测方法进行对比，证明该方法具有更高的精确度和更快的收敛速度。

## 5. 建立中短期电负荷预测模型

### 5.1 短期负荷预测模型

#### 5.1.1 短期负荷特征工程构建

根据前文的负荷特性与影响因素的相关性分析，我们构建短期负荷特征工程，考虑的因素以及延迟阶数如下表所示。

表 8 短期模型特征工程考虑因素

预测日因素	最高温度	最低温度	最坏天气值	最好天气值	周一	周二	周三	周四	周五	周六	月份	节假日
延迟一阶	最高温度	最低温度	最坏天气值	最好天气值	C001	C002	C003	...	C093	C094	C095	C096
延迟一阶	最高温度	最低温度	最坏天气值	最好天气值	C001	C002	C003	...	C093	C094	C095	C096

如上表所示，我们考虑了预测日前一天所有 96 个时点的负荷数据以及气象因素，还有预测日同周期延迟七天的所有 96 个时点负荷数据及气象数据，气象数据包括日最高温度、日最低温度、最坏天气值、最好天气值。

我们考虑了预测日的工作日类型和月份。对于预测日的工作日类型，在这里我们使用了哑变量，例如，将“周一”转换为“是否周一的属性”，其他日工作类型同理，还需要注意的是这里为了避免多重共线性，只生成了周一到周六的哑变量。

#### 5.1.2 基于 RBF 神经网络的预测模型

神经网络结构和学习算法的选择对负荷预测计算的速度和精度有着直接的

影响。现有的 ANN 方法一般采用误差反向传播算法的前向多层感知器网络，由于感知器网络权重初始化的随机性，难于根据负荷预测的实际情况确定一组较好的初始值，并且 BP 算法学习收敛速度较慢，容易陷于局部极小点，从而极大地限制了神经网络在负荷预测中的实际应用。后来发展起来的径向基函数 RBF 理论为前向网络的学习提供了一种新颖而有效的手段。

### 1. RBF 神经网络结构设计

在使用 RBF 神经网络进行预测之前，我们要对网络的结构进行设计。针对短期负荷预测特征工程的构建，输入神经元 212 个，隐含层神经元，需要根据数据特点进一步确认，输出神经元有 96 个如图 18 所示。

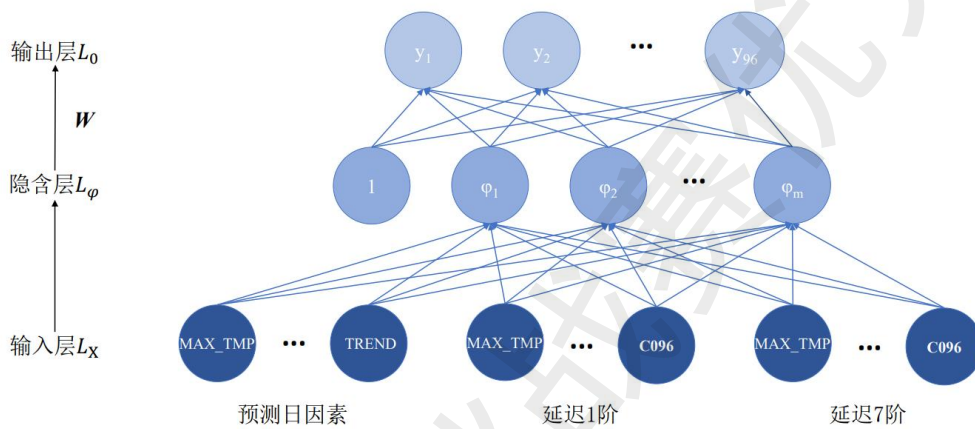


图 18 RBF 神经网络结构图

如上图，我们使用了预测日因素、延迟一阶及七阶的因素作为输入，输出为日负荷 96 个时点的负荷数据。RBF 神经网络的建模任务就是基于提供的训练集，求解出适合隐含层的神经元个数及扩展常数，并基于高维空间中的数据点使用最小二乘法求出权重矩阵  $W$ 。

### 2. 确定最优参数

构建 RBF 神经网络，最主要的就是根据训练数据确定隐含层神经元个数及扩展常数。如果隐含层神经元太多，容易使用网络陷入过拟合，而失去对新数据的泛化预测能力；如果隐含层的神经元太少，则会更丧失网络的学习能力。所以，确定合适数目的隐含层神经元数目对构建 RBF 神经网络非常重要。此外，扩展常数会影响对应高维空间中点的分布，如果设置扩展常数大大，则点的分布会很分散，适度的分布会对 RBF 神经网络的学习效果有影响，因此确定适当的扩展常数，对 RBF 神经网络也非常重要[3]。

本文的基于 RBF 神经网络电负荷预测模型是通过从训练数据中有偏选取一部分样本作为测试集，剩下的数据用于训练模型，同时结合遗传算法搜索 RBF 神经网络的最优隐含层神经元数目和扩展常数。

### (1) 有偏抽样

有偏抽样是对不同的样本设置了不同的权重,这样做的好处是有利于增加模型的泛化能力。对于给定的训练数据,使用有偏抽样,选取时间较新的数据用于测试,而使用时间较远的数据训练模型,如果基于测试集的效果较好,那么该模型在新数据上的泛化能力更能得到保证。

### (2) 使用遗传算法搜索最优参数

遗传算法是计算科学人工智能领域中用于解决最优化的一种搜索启发式算法,它也是一种进化算法,可用于生成有用的解决方案和搜索最优解。此处使用遗传算法,在有限的数值空间内搜索最优的  $\lambda$  和  $\alpha$  参数使得训练得到的模型具有较强的泛化能力和准确率。

## 5.2 中期负荷预测模型

### 5.2.1 温度累积效应

在中期负荷日最大值预测中,最高温度和日最大负荷的线性相关性可以通过温度累积效应的修正进一步提高。下面介绍如何采用文献[4]的方法进行日高温修正。

温度累积效应的主要思想是,考虑人体舒适度存在的惯性,多日连续高温时,即使温度变化不大,负荷也会增长。通过对预测日前几日的最高温度进行加权后,对待预测日的最高温度进行修正,从而体现出多日连续高温的留存影响。

$$T'_i = \sum_{j=0}^d k_j T_{i-j}$$

式中,  $T'_i$  为考虑温度累积效应后,经过修正的日最高温度,  $T_{i-j}$  为第  $i$  日前  $j$  天的日最高温度真实值,  $d$  为考虑最大的温度累积效应的天数,  $k_j$  为  $T_{i-j}$  对应的权重。

有研究表明,当  $d$  大于等于 3,即高温持续天数达 3 天及以上时,累积效应的强度对天数不再敏感。所以要考虑温度累积效应进行日最大温度修正,需要求解  $d$  和权重  $k_j$ ,具体的求解方法如下:

#### (1) 划分温度区间

因为日最高温度和对应的日最大负荷不是线性关系,故根据日最高温度划分为不同的区间,分段进行线性拟合。根据统计,选取  $T_{low}=25^{\circ}\text{C}$ ,  $T_{high}=37^{\circ}\text{C}$ 。

表 9 温度划分区间

区间序号	温度范围(°C)	$k_j$	区间序号	温度范围(°C)	$k_j$
0	>37	0	4	29,31]	$k_{j,4}$
1	(35,37]	$k_{j,1}$	5	(27,29]	$k_{j,5}$
2	(33,35]	$k_{j,2}$	6	(25,27]	$k_{j,6}$
3	(31,33]	$k_{j,3}$	7	<=25	0

## (2) 计算步骤

算法的核心是对于每一个不同的  $d$ ，将样本集划分入按温度划分好的  $n$  个区间中，形成  $n$  个方程组。每个方程组中有数量不定的等式，方程组一般为超定多元线性方程组。最终对方程组采用最小二乘法进行求解得到权值  $k_j$ 。计算流程如图 19 所示。

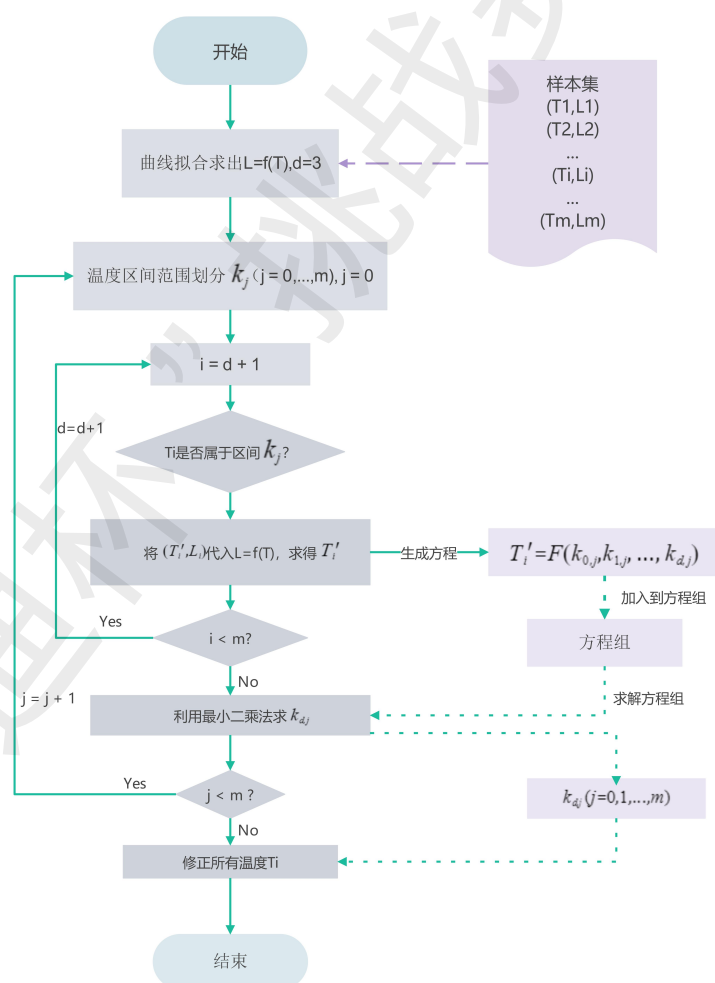


图 19 温度累积效应计算流程图

经过算法修正后，日最高温度与区域日最大负荷的皮尔逊相关性系数从

53.4324%提高到了 67.4324%。

### 5.2.2 中期负荷特征工程构建

首先，根据前文的负荷特性与影响因素的相关性分析，我们构建地区中期极值负荷特征工程，考虑的因素以及延迟阶数如下表所示。

表 10 中期模型特征工程考虑因素

预测日因素	最高温度	最低温度	最坏天气值	最好天气值	周一	周二	周三	周四	周五	周六	月份	节假日
延迟一阶	最高温度	最低温度	最坏天气值	最好天气值	C001	C002	C003	.....	C093	C094	C095	C096
延迟一阶	最高温度	最低温度	最坏天气值	最好天气值	C001	C002	C003	.....	C093	C094	C095	C096

此外我们构建行业中期负荷特征工程，考虑的因素以及延迟阶数如下表所示。

表 11 短期模型特征工程考虑因素

预测日因素	最高温度	最低温度	最坏天气值	最好天气值	周一	周二	周三	周四	周五	周六	月份	节假日
延迟一阶	最高温度	最低温度	最坏天气值	最好天气值	最大电负荷							
延迟一阶	最高温度	最低温度	最坏天气值	最好天气值	最低电负荷							

如上表所示，我们考虑了预测日前一天所有 96 个时点的负荷数据以及气象因素，还有预测日同周期延迟七天的所有 96 个时点负荷数据及气象数据，气象数据包括日最高温度、日最低温度、最坏天气值、最好天气值。

我们考虑了预测日的工作日类型和月份。对于预测日的工作日类型，在这里我们使用了哑变量，例如，将“周一”转换为“是否周一的属性”，其他日工作类型同理，还需要注意的是这里为了避免多重共线性，只生成了周一到周六的哑变量。

### 5.2.3 考虑温度累积效应的 RBF 神经网络的预测模型

首先基于前面的温度累积效应，我们将对最大温度进行修正，然后将修正好的最大温度结合构建的中期预测特征工程，建立考虑温度累积效应的 RBF 神经

网络预测模型。主要的 RBF 神经网络模型与前文的短期的负荷预测结构一致。这里不做重复的描述。

### 5.3 效果评估指标

针对短期预测模型，考虑到不同时点用电基数的差别，我们从三个不同时间粒度对预测误差和准确性进行评估，分别是 10 天、1 天、间隔 15 分钟的 1 个点 [3]。

从 10 天的粒度上，我们使用总体绝对百分误差（Total Absolute Percentage Error, 简记为 TAPE）来评估所有点预测值的总体误差情况，使用预测精度（Forecasting Accuracy, 简记为 FA）的平均值来评估总体预测精度；从 1 天的粒度上，使用 FA 的日均值评估精度；从间隔 15 分钟的 1 个点的粒度上，我们选取某一日，使用绝对百分误差（Absolute Percentage Error, 简记为 APE）来评估的一个数据点的预测误差。

TAPE 的定义如下：

$$TAPE = \frac{\text{绝对差异}}{\text{总体基数}} = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{\sum_{i=1}^n |y_i|}$$

其中， $n$  表示 96 乘以天数  $d$  的值，10 天粒度上  $d=10$ ，1 天粒度上  $d=1$ 。 $y_i$  表示按时间顺序对应第  $i$  个时点的电负荷真实值， $\hat{y}_i$  表示对应的第  $i$  个时点的预测值。拟合值与真实值完全相等的时候， $TAPE=0$ 。 $TAPE$  值越小，总体百分误差越小，预测效果越好。该指标可以形象化地用图 20 代表的面积及差异部分的占比来表示。

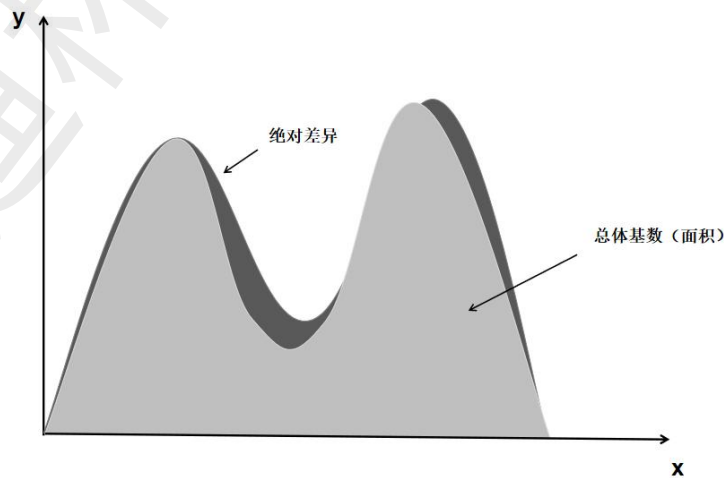


图 20 计算总体绝对百分误差

如上图，灰色深灰色的面积为差异部分，该部分的面积越小，其占总体基数的比例越小，预测的效果也就越好。

FA 的定义如下：

$$FA = \left( \frac{1 - |y_i - \hat{y}_i|}{y_i} \right) \times 100\%$$

APE 的定义如下：

$$APE = \frac{|y_i - \hat{y}_i|}{|y_i|}$$

对于中期预测模型，我们使用均方根误差（Root Mean Square Error, 简记为 RMSE）、平均绝对误差（Mean Absolute Error, 简记为 MAE）和来从三个月的时间粒度对负荷最大和最小值预测结果的稳定性和误差进行评估，使用 FA 均值对预测值结果的精度进行评估。另外，我们同样使用 RMSE 和 MAE 对预测的达到最大和最小负荷的时间点进行稳定性和误差评估。

RMSE 针对异常值会比较敏感，若出现预测值和真实值偏差较大，则 RMSE 值就会很大，可以用来衡量模型的稳定性。RMSE 的公式定义如下：

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - y)^2}$$

其中 m 表示时点个数。

MAE 对预测值和真实值的残差直接进行计算，且量纲与计算值相同，可以比较直观地体现模型预测结果的误差情况。MAE 的公式定义如下：

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - y|$$

## 6. 地区负荷的中短期预测分析

### 6.1 地区负荷短期预测与分析

#### 6.1.1 使用遗传算法优化的 RBF 神经网络预测模型

我们将运用构建好的 RBF 神经网络短期负荷预测模型来预测该地区电网未来 10 天间隔 15 分钟的负荷预测结果，并分析其预测精度。

我们选用 2018 年 9 月 1 日到 2021 年 8 月 21 日的数据作为训练集，选用 2021 年 8 月 22 号到 2021 年 8 月 31 号共 10 天的数据作为测试集，模型输入为已经构建好的短期负荷预测特征工程，输出为当天 96 时点负荷预测值。

首先我们利用训练集求解出 RBF 网络中的权重矩阵  $W$ ，然后基于测试集的特征工程和  $W$  得到测试集的预测值。通过与测试集的真实值进行比较，得到均方误差  $error$ 。进行训练之前，我们利用遗传算法找到合适的隐含层神经元数目和扩展常数使得在测试集上的均方误差  $error$  达到最小。

#### 6.1.2 确定最优参数

##### 1. 有偏抽样

遗传算法的适应性函数是根据所找的参数，计算测试集上的均方误差，该值越小，预测效果越好，参数也越优。而遗传算法中需要的训练数据和测试数据我们使用有偏抽样划分。划分结果如图 21 所示。横轴表示数据的序号，序号越大数据越新，可以看到随着序号的增加，数据的抽样频率越大。

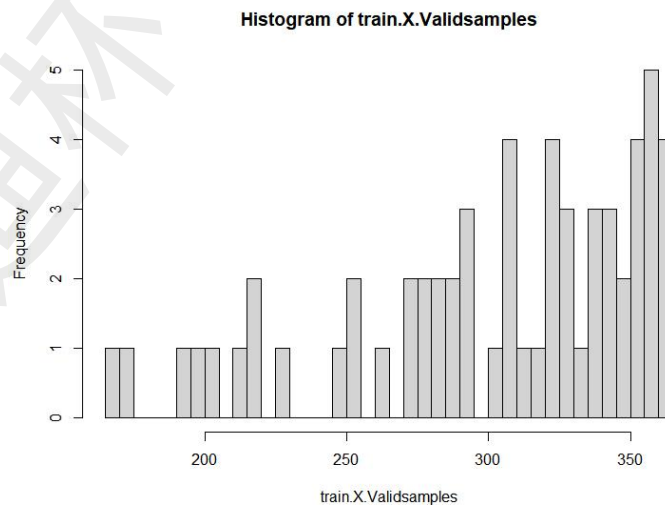


图 21 有偏抽样分布

##### 2. 遗传算法求最优参数



由于该模型最优参数涉及到多个参数，因此我们使用 R 语言 `genalg` 包中的 `rbga` 函数。`genalg` 包是基于 R 语言用于二元和浮点染色体的遗传算法。计算结果如下图所示。

```

> library (genalg)
> #定义监控函数
> monitor<-function (rbga0){
+   #打印种群中第一个个体的值population [1,]
+   print(rbga0$population[1, ])
+ }
> rbgaObj<-rbga(stringMin = c(1,1), stringMax = c (20,20), popSize =100,
+   iters = 7, mutationChance = 0.01, monitorFunc= monitor,
+   evalFunc = getError, verbose = TRUE)
Testing the sanity of parameters...
Not showing GA settings...
Starting with random values in the given domains...
Starting iteration 1
Calculating evaluation values... [1] "lambda: 11.1449986943044 alpha: 7.21490854443982
error: 71889339.5709493"
.[1] "lambda: 4.27745116339065 alpha: 7.40137808769941 error: 71486590.2890966"
.[1] "lambda: 7.91566271428019 alpha: 11.7125229206868 error: 66775416.0729682"
.[1] "lambda: 1.62637826055288 alpha: 19.0504838856868 error: 64713204.0930321"
.[1] "lambda: 14.4607487302274 alpha: 17.7081678446848 error: 74080368.9384509"
.[1] "lambda: 4.3163931122981 alpha: 4.75047083781101 error: 81176005.8735181"
.[1] "lambda: 18.6517079910263 alpha: 6.30550520378165 error: 95626537.8588888"
.[1] "lambda: 18.6730453304481 alpha: 19.3312089485116 error: 95619188.2503372"
.[1] "lambda: 1.37505367654376 alpha: 7.92707023303956 error: 75528203.9796574"
.[1] "lambda: 15.1796567044221 alpha: 7.38061125250533 error: 73407107.7359789"
.[1] "lambda: 19.4342523380183 alpha: 14.4268046510406 error: 97100604.6108579"

applying elitism...
applying crossover...
applying mutations... 2 mutations applied
Starting iteration 7
Calculating evaluation values... [1] "lambda: 2.24166035698727 alpha: 15.648837597575
error: 63218651.5226521"
.[1] "lambda: 2.24166035698727 alpha: 15.648837597575 error: 63218651.5226521"
.[1] "lambda: 2.24166035698727 alpha: 16.3187762356829 error: 63110107.7490545"
.[1] "lambda: 2.24166035698727 alpha: 16.7790798277128 error: 63070040.1315648"
.[1] "lambda: 2.40076329070143 alpha: 16.7790798277128 error: 63257216.0614523"
.[1] "lambda: 2.40076329070143 alpha: 13.5573476642543 error: 63475855.8253178"
.[1] "lambda: 2.24166035698727 alpha: 16.7790798277128 error: 63070040.1315648"
.[1] "lambda: 2.24166035698727 alpha: 16.7790798277128 error: 63070040.1315648"
.[1] "lambda: 2.24166035698727 alpha: 16.7790798277128 error: 63070040.1315648"
.[1] "lambda: 2.24166035698727 alpha: 16.7790798277128 error: 63070040.1315648"
.[1] "lambda: 2.40076329070143 alpha: 16.7790798277128 error: 63257216.0614523"
.[1] "lambda: 2.24166035698727 alpha: 16.7790798277128 error: 63070040.1315648"
.[1] "lambda: 2.24166035698727 alpha: 16.7790798277128 error: 63070040.1315648"
.[1] "lambda: 2.24166035698727 alpha: 16.3187762356829 error: 63110107.7490545"
.[1] "lambda: 9.37100780243054 alpha: 16.7790798277128 error: 71417607.2257793"
.[1] "lambda: 2.24166035698727 alpha: 15.648837597575 error: 63218651.5226521"
.[1] "lambda: 2.40076329070143 alpha: 16.7790798277128 error: 63257216.0614523"
.[1] "lambda: 2.40076329070143 alpha: 16.7790798277128 error: 63257216.0614523"
.[1] "lambda: 2.24166035698727 alpha: 15.648837597575 error: 63218651.5226521"
.[1] "lambda: 2.24166035698727 alpha: 16.7790798277128 error: 63070040.1315648"
.[1] "lambda: 2.24166035698727 alpha: 16.3187762356829 error: 63110107.7490545"
.[1] "lambda: 2.40076329070143 alpha: 16.7790798277128 error: 63257216.0614523"
.[1] "lambda: 2.40076329070143 alpha: 15.648837597575 error: 63322411.0797806"
. done.
Sending current state to rbga.monitor()...
[1] 2.24166 16.77908

```

图 22 遗传算法计算过程

经过 5 代进化，得到的最优  $\lambda=2.24166$ ， $\alpha=16.77908$ 。进一步可使用改参数基于训练数据得到预测模型并进行预测。

### 6.1.3 地区负荷预测结果分析与比较

对测试集 2021 年 8 月 22 日至 2021 年 8 月 31 日共 10 天 960 个时点的电力负荷进行预测，图 23(a)-(j)为本文 5.1.1 小节的基于 RBF 短期负荷模型的预测结果。

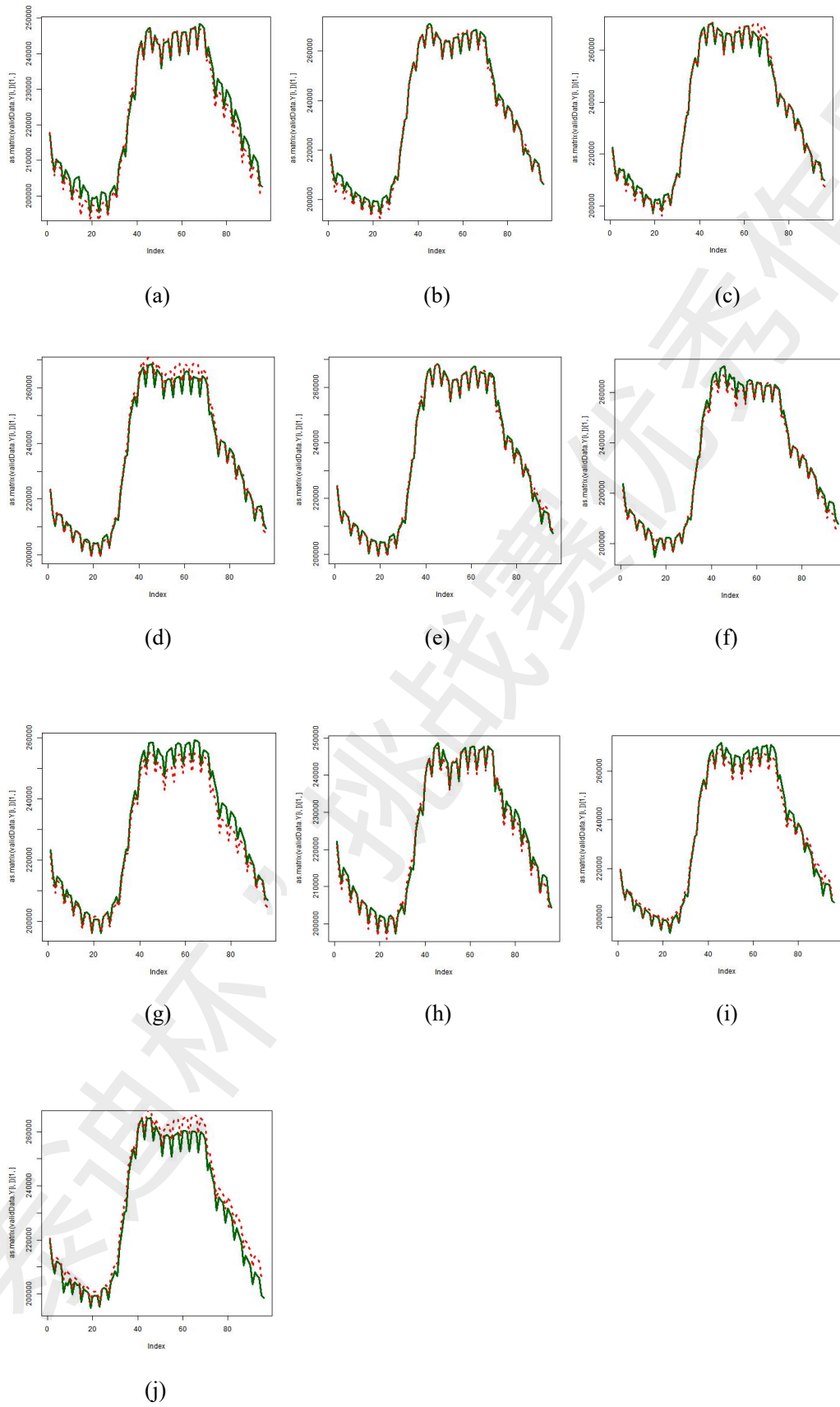


图 23 本文模型 1 十天间隔 15 分钟的短期负荷预测结果

直观地来看，预测效果很不错。从十天的粒度进行分析，计算得到十天的TAPE为0.79%，平均FA为99.21%，从准确度和误差两个角度来看都很优秀。

如下表12，选取2021年8月23日的一些具体时点进行展示和比较。可以看到基于RBF短期负荷模型在绝大多数时点的误差都要更小。

表12 2021年8月23日RBF模型短期负荷预测结果与其他模型比较

时段	实际负荷 (kW)	RBF		LSTM		LSSVM	
		预测值	APE	预测值	APE	预测值	APE
0:00:00	218239.2401	217251.6511	0.45	199283.5947	-8.6857 1818	209089.1109	5.55621 6626
0:15:00	210831.1059	207975.9999	1.35	196042.6725	-7.0143 50815	209105.6548	1.89110 1535
0:30:00	206633.8874	203214.322	1.65	202579.4505	-1.9621 35538	219541.4526	6.05689 6331
0:45:00	210764.7222	207209.4228	1.69	205549.7777	-2.4742 96657	226101.5383	7.88647 1436
1:00:00	210019.1595	206793.1602	1.54	209089.1109	-0.4428 39884	199283.5947	-4.1871 97165
1:15:00	209574.5151	205929.8726	1.74	209105.6548	-0.2237 20105	196042.6725	-10.523 72749
1:30:00	203542.9547	200314.1946	1.59	219541.4526	7.86001 0647	202579.4505	-10.175 72543
1:45:00	207136.9328	203858.6674	1.58	226101.5383	9.15558 8625	205549.7777	-11.442 41779
2:00:00	205703.6905	203149.3028	1.24	232909.9908	13.2259 6607	209089.1109	-10.271 10934
2:15:00	204910.1561	202522.887	1.17	234531.0163	14.4555 3544	209105.6548	-15.142 78545
2:30:00	199430.1807	197532.8779	0.95	197375.9099	-1.0300 70153	219541.4526	-12.736 81187
2:45:00	203141.3955	201514.0968	0.80	197139.1583	-2.9547 09031	226101.5383	-11.535 4163
3:00:00	202045.409	201028.8359	0.50	196778.6094	-2.6067 40554	232909.9908	-7.5602 12392
3:15:00	201568.2935	200054.0387	0.75	192260.8269	-4.6175 2515	198390.1241	-0.5090 17174
3:30:00	196304.3059	195047.9389	0.64	197351.4582	0.53343 3184	199283.5947	2.34580 3173
3:45:00	201073.0348	199063.63	1.00	-3.635093077	-3.6350 93077	196042.6725	-2.1128 46821

4:00:00	200237.5846	198858.6389	0.69	-1.008032343	-1.0080 32343	202579.4505	0.80090 455
4:15:00	199907.4326	198188.9466	0.86	-0.509017174	-0.5090 17174	205549.7777	1.89964 4285
4:30:00	194552.7333	193152.1256	0.72	2.345803173	2.34580 3173	198390.1241	-0.5090 17174
4:45:00	199513.3218	197375.9099	1.07	-2.112846821	-2.1128 46821	199283.5947	2.34580 3173
5:00:00	199361.0824	197139.1583	1.11	0.80090455	0.80090 455	196042.6725	-2.1128 46821
...	...	...	...	...	...	...	...
20:45:00	231969.4909	196778.6094	1.32	226101.5383	9.15558 8625	205549.7777	-2.4742 96657
21:00:00	229402.2201	192260.8269	1.26	232909.9908	13.2259 6607	209089.1109	-0.4428 39884
21:15:00	227465.0847	197351.4582	1.46	234531.0163	14.4555 3544	209105.6548	-0.2237 20105
21:30:00	217978.1837	198390.1241	1.28	197375.9099	-1.0300 70153	219541.4526	7.86001 0647
21:45:00	220945.0395	199283.5947	1.21	197139.1583	-2.9547 09031	226101.5383	9.15558 8625
22:00:00	218621.6059	196042.6725	1.03	196778.6094	-2.6067 40554	232909.9908	13.2259 6607
22:15:00	217939.7557	202579.4505	1.29	192260.8269	-4.6175 2515	234531.0163	14.4555 3544
22:30:00	211444.6888	205549.7777	0.70	197351.4582	0.53343 3184	197375.9099	-1.0300 70153
22:45:00	216306.5989	209089.1109	0.23	198390.1241	-1.3342 96597	197139.1583	-2.9547 09031
23:00:00	215575.0043	209105.6548	0.54	199283.5947	-0.4764 28966	196778.6094	-2.6067 40554
23:15:00	214393.6083	219541.4526	0.20	196042.6725	-1.9332 74839	198390.1241	-0.5090 17174
23:30:00	207530.5691	226101.5383	0.25	202579.4505	4.12572 8292	199283.5947	2.34580 3173

接着从 1 天的时间粒度进行分析，下表 13 中可以看到本文模型 1 的在十天的准确率均在 99.9%以上。

表 13 本文模型 1 负荷预测结果与其他模型准确率比较

日期	RBF	LSTM	LSSVM
2021.8.22	99.989%	93.99%	96.99%
2021.8.23	99.994%	91.99%	97.99%

2021.8.24	99.995%	86.00%	86.00%
2021.8.25	99.992%	87.99%	98.34%
2021.8.26	99.998%	93.44%	95.42%
2021.8.27	99.993%	91.33%	97.33%
2021.8.28	99.983%	92.44%	99.35%
2021.8.29	99.996%	92.78%	96.54%
2021.8.30	99.990%	93.89%	93.89%
2021.8.31	99.969%	91.69%	94.69%
平均准确率	99.990%	91.55%	95.65%

下面从间隔 15 分钟的时点的粒度进行分析。下图 24 为 RBF 负荷预测模型的预测结果的绝对百分误差分布直方图，可以发现，大多数时点的 APE 在 2.5% 的范围以内。

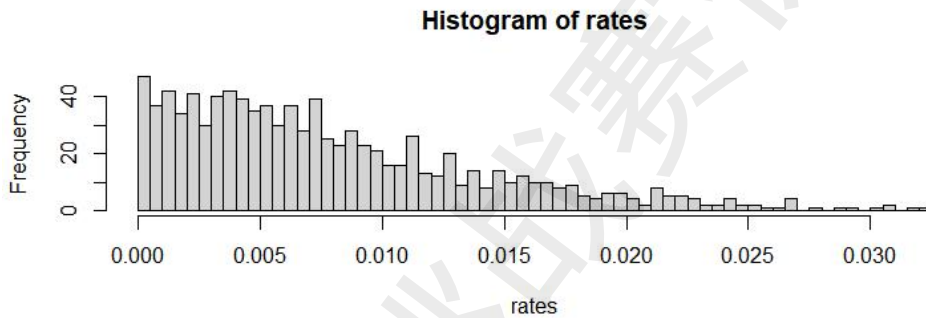


图 24 预测结果 960 个时点的 APE 分布直方图

## 6.2 地区负荷中期预测与分析

### 6.2.1 地区中期最值负荷预测结果及分析

我们将使用基于 RBF 的中期预测模型对测试集 2021 年 5 月 31 日至 2021 年 8 月 31 日（简记为测试集时间段 1）共 3 个月 93 天的电力负荷进行预测，下图 25 为其日最大电力负荷值预测结果，图 26 为其日最小电力负荷值预测结果。

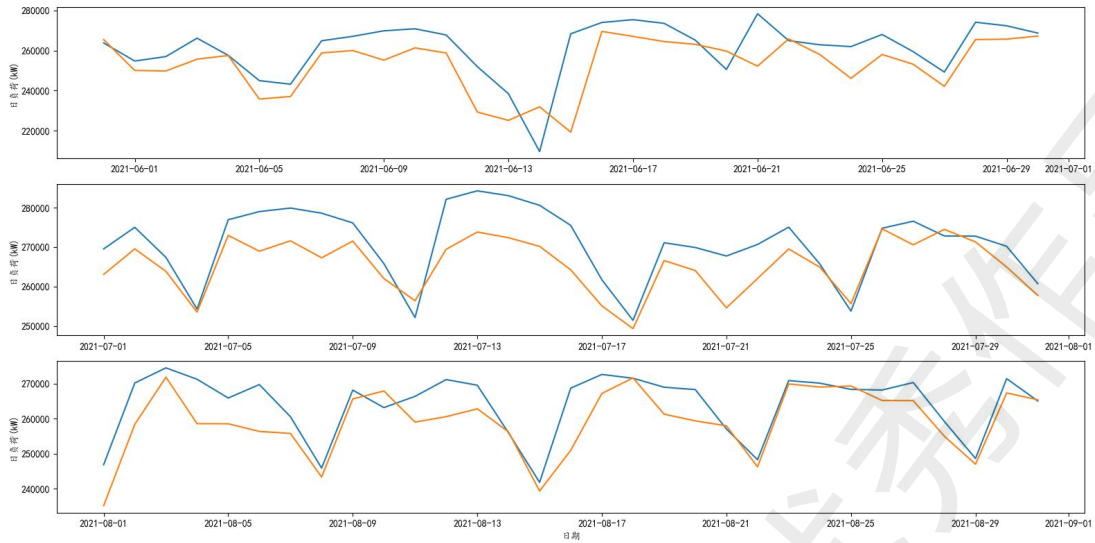


图 25 中期负荷预测模型的日最大负荷值预测结果

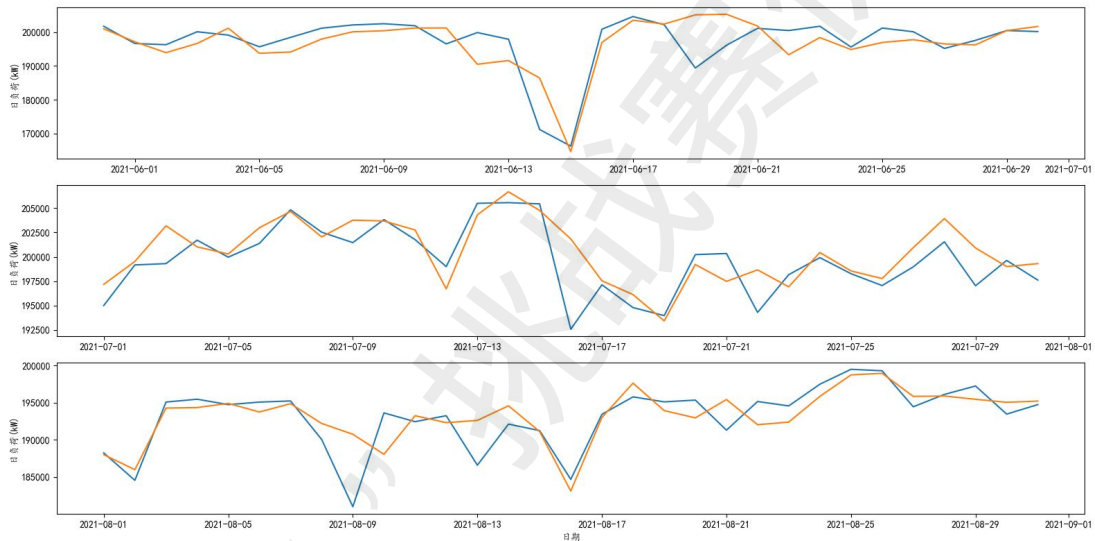


图 26 中期负荷预测模型的日最小负荷值预测结果

从上图 25 和图 26 可以直观地看到，本文模型 2 对日最大和最小负荷值的预测效果较为不错，没有出现较大偏差，变化趋势也基本正确。特别是 2021 年 6 月 12 日至 14 日的端午节影响下，日最小负荷值出现的低谷也被准确的识别和预测到，日最大负荷值出现的低谷也有一定体现。其次是，日最小负荷的整体误差相较于日最大负荷的整体误差要明显较小。下面对预测结果进行定量分析。

表 14 基于 RBF 中期负荷预测对日最大和最小负荷的预测结果的评估指标

评估指标	最大负荷的评估	最小负荷的评估
MAE	7146.8189	2406.3249
RMSE	9818.3639	3762.7893
平均 FA	97.30%	98.76%



如表 14，考虑温度累积效应 RBF 中期负荷预测模型对于测试集的日最大和日最小负荷的分别预测的准确率较高，均在 97%以上。其次是两者的 RMSE 均没有过大，说明中期预测模型对于日负荷最值的预测稳定性较好，预测值和实际值偏差较小。

其中日最小负荷的预测平均绝对误差比日最大负荷的预测平均绝对误差小 4740.7940，均方根误差要小 6055.5746。说明中期预测模型对于日最小负荷值的预测效果要比对于日最大负荷值的预测效果好，具体是在准确性、误差大小和误差稳定性上都要更优越。

### 6.2.2 地区中期日最大负荷对应时间点预测结果及分析

下面对测试集时间段 1 对应的达到负荷最大值和最小值的时间点的预测结果进行展示和分析。

首先对达到日最大负荷的时间点进行分析。如图 27，测试集时间段 1 上真实的日最大负荷出现时间主要集中在两个时间段：44 至 46（即 10:45 至 11:15，简记为上午最高型时段 1），61 至 68（即 15:00 至 16:45，简记为下午最高型时段 2）。而预测的日最大负荷时间点绝大多数情况也和实际值所在时间点所吻合。可以一定程度说明中期预测模型对于最大负荷所在时间点的预测的准确性。



图 27 中期负荷预测模型的预测及真实日最大负荷所在时点

注：时点编号从 1 到 96，顺序对应一天内 00:00 到 23:45 间隔 15 分钟的 96 个时点，下同

对上图 27 截取上午最高型时段 1 放大纵轴进行查看，得到下图 28。可以看到上午最高型时段 1 上实际值达到日最大负荷的时间点主要为 45 和 46（即分别为 10:45 和 11:00），而中期预测模型预测的日最大负荷的时间点主要集中在 10:45。

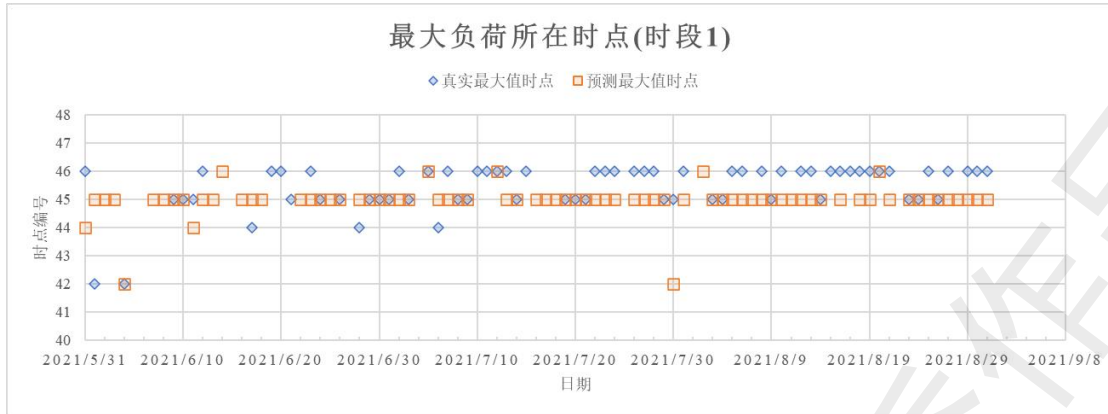


图 28 中期负荷预测模型的预测及真实日最大负荷所在时点(截取时段 1)

截取图 28 中下午最高型时段 2 放大纵轴进行查看，得到下图 29。可以看到下午最高型时段 2 内多天的日最高负荷出现时点较为离散，中期预测模型的预测对其时点预测效果也相对较差。

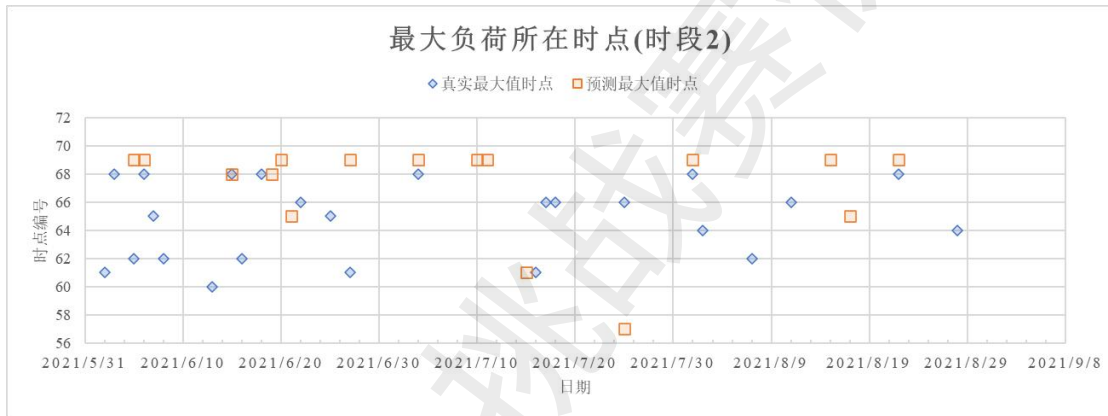


图 29 中期负荷预测模型的预测及真实日最大负荷所在时点(截取时段 2)

对测试集时间段 1 的日最大负荷对应时间点的预测结果进行定量评估，计算得到 MAE 为 94（分钟），即预测最大负荷到达时间与实际到达时间的误差平均为 94 分钟。计算得到 RMSE 为 169。

### 6.2.3 地区中期日最小负荷对应时间点预测结果及分析

接着对达到日最小负荷的时间点预测结果进行分析。图 30 中可以看到端午节假期后两天(2021 年 6 月 13 日至 14 日)的最小值均出现在了时点 96(即 23:45)，可以看到本文模型 2 也很好拟合出了这个节假日时的日最小负荷特征。





2021.5.31	263792.6115	265461.4705	4.057	260002.329	1.987	246122.2822	-3.273
2021.6.1	254758.3847	250042.9756	-2.731	255217.4297	-0.699	258029.5339	0.4038
2021.6.2	257000.6755	249798.0809	-6.371	261314.7169	-1.89	253149.7323	-5.067
2021.6.3	266173.5051	255691.4149	-0.718	258776.18	0.440	242113.0571	-5.819
2021.6.4	257603.1158	257510.769	4.866	229290.6038	-6.088	265508.8032	7.9715
2021.6.5	244973.8829	235773.4221	-3.018	225111.908	-7.370	265683.1082	9.1907
...	...	...	...	...	...	...	...
2021.8.30	243168.2084	237052.8337	-11.446	231886.3348	-13.570	267245.3539	0.9702
2021.8.31	264885.943	258795.3823	-3.132	219252.4333	-18.060	263075.6014	-1.516
平均准确率		<b>97.3%</b>		<b>96.896%</b>		<b>95.57%</b>	

## 2、地区中期最小值负荷模型比较

时段	实际最大负荷	RBF		LSTM		LSSVM	
		负荷预测值	相对误差/%	负荷预测值	相对误差/%	负荷预测值	相对误差/%
2021.5.31	201757.0459	201009.2635	-0.3706	202146.0321	0.19279	200878.1545	-0.435
2021.6.1	196638.4527	197150.0896	0.26019	202506.8026	2.98433	204635.6386	4.0669
2021.6.2	196286.2928	193925.9436	-1.2025	201907.2157	2.86363	202217.3771	3.0216
2021.6.3	200118.4672	196644.1189	-1.7361	196526.1144	-1.7951	189376.5132	-5.367
2021.6.4	199135.5438	201178.3926	1.02585	199891.2504	0.37949	196116.8379	-1.515
2021.6.5	195663.2838	193747.0829	-0.9793	197907.4099	1.14693	201138.9152	2.7984
...	...	...	...	...	...	...	...
2021.8.30	198456.5982	194110.2414	-2.1900	171136.9641	-13.766	200484.321	1.0217
2021.8.31	201186.5727	197963.2645	-1.6021	166215.7393	-17.382	201757.9495	0.2840
平均准确率		<b>98.76%</b>		<b>92.546%</b>		<b>91.347%</b>	

通过将 RBF 与 LSTM 和 LSSVM 模型的预测结果相比较，我们可以得出考虑温度累积效应基于 RBF 的负荷预测模型具有较好的预测精度较高。

## 7. 行业负荷中期预测分析

### 7.1 突变点分析

#### 7.1.1 突变点定义

电负荷的数值变化与日常生活息息相关，因此现实生活中的日电负荷时间序列往往是一种非平稳序列，因为负荷受许多因素影响，如天气、节假日等。从统

计学的角度而言, 突变现象可以定义为从一种统计特性到另一个统计特性的急剧变化, 如均值和方差。常见的突变有四种, 分别为[5]:

- (1) 均值突变: 从一个均值到另一个均值的急剧变化, 如图 32(a)所示。
- (2) 变率突变: 从一个方差状态到另一个方差状态的急剧变化, 其平均值却无明显变化, 如图 32(b)所示。
- (3) 跷跷板突变: 从一个递增(递减)变化到另一个递增(递减)变化, 交叉点存在数值突变, 如图 32(c)所示。
- (4) 转折突变: 某一时段持续减少(增加), 然后突然在某点开始持续增加(减少)。如图 32(d)所示。

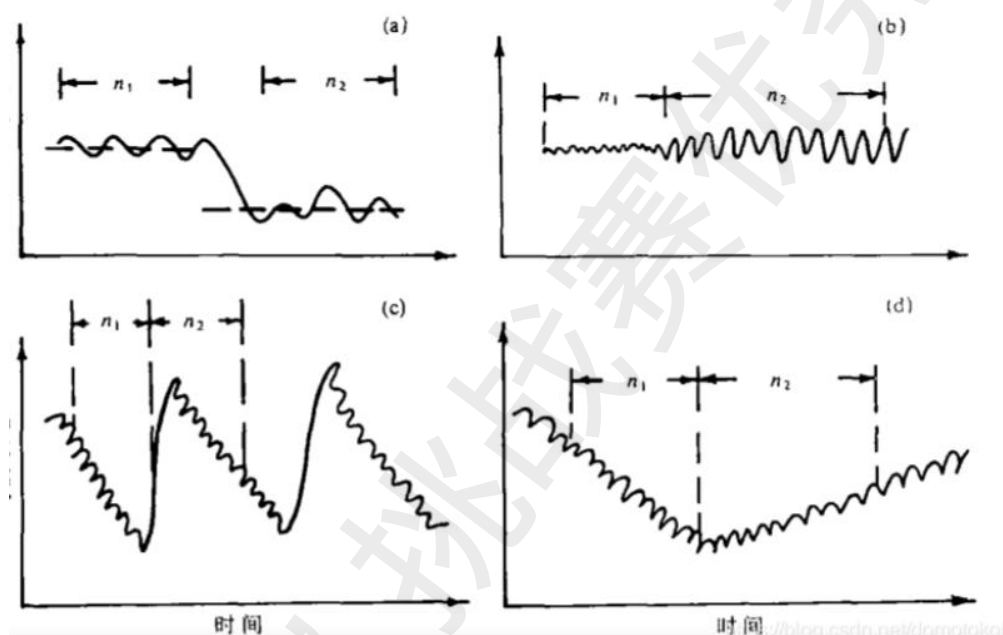


图 32 突变类型

检测突变点的方法有很多, 如 Mann-Kendall 方法、Pettitt 方法、滑动 T 检验、基于启发式的分割算法等。但是, 不同的突变检测方法有不同的适应范围, 比如, Mann-Kendall 方法和 Pettitt 方法对可能出现多个突变的时间序列检测不适用, 滑动 T 检验要求所测的序列满足正态分布、基于启发式的分割算法适用于多个突变点检测。通过对四大行业的日最大最小负荷进行可视化分析后, 我们发现行业日最大最小负荷序列存在的多个突变点, 因此, 本文选择基于启发式的分割算法进行突变点检测。

### 7.1.2 基于启发式的分割算法原理

基于启发式的分割算法是 Bernaola-Galvan 在 2001 年提出的一种突变检测的有效方法, 又称 BG 算法。该方法可以将非平稳序列分割为多尺度的自平稳子序列; 检测的尺度和精度具有可变性, 能够检测不同尺度和不同幅度的突变; 白噪

声和尖峰噪声的对该方法的影响较小，特别适合于处理类似气候、负荷的时间序列。其主要思想介绍如下：

对于包含  $N$  个样本的一时间序列数据  $X(t)$ ，通过第  $i(2 \leq i \leq N-1)$  个数据将  $X$  分割为左右两个子序列  $X_1$ （包含  $N_1$  个样本）和  $X_2$ （包含  $N_2$  个样本），分别计算两个子序列的均值  $\mu_1(i)$ 、 $\mu_2(i)$  和标准差  $s_1(i)$ 、 $s_2(i)$ ，以及  $t$ -检验统计值  $T(i)$ ，其计算公式为：

$$T(i) = \left| \frac{\mu_1(i) - \mu_2(i)}{S_D(i)} \right|$$

其中， $S_D(i)$  为第  $i$  个数据的合并偏差，定义为：

$$S_D(i) = \sqrt{\frac{(N_1-1) \times s_1(i)^2 + (N_2-1) \times s_2(i)^2}{N_1 + N_2 - 2}} \times \left( \frac{1}{N_1} + \frac{1}{N_2} \right)$$

对  $X(t)$  中的每一个点重复上述计算过程，得到与  $X(t)$  一一对应的检验统计值序列  $T(t)$ ， $T$  越大，表示该点左右两部份的均值相差越大。从中找到最大值  $T_{max}$  及其所对应的索引  $j$ ，如果  $T_{max}$  的统计显著性  $P(T_{max}) \geq P_0$  ( $P_0$  为给定的参数)，那么便可对序列  $X$  在第  $j$  个样本处进行分割，也就是突变点。 $P(T_{max})$  的计算可近似为：

$$P(T_{max}) \approx (1 - I_{\frac{v}{v+T_{max}^2}}(\delta v, \delta))^n$$

由蒙特卡洛模拟可以得到  $\eta = 4.19 \ln N - 11.54$ ， $\delta = 0.4$ ， $N$  为时间序列  $X(t)$  的长度， $v = N - 2$ ， $I_x(a, b)$  为不完全  $\beta$  函数。

同样的，当分割完后，可以对分割后的两个子序列重复进行上述操作，直到不可分割为止，便可以得到所有的突变点。为确保统计的有效性，当子序列的长度小于等于  $l_0$  ( $l_0$  为最小分割尺度) 时不再对其进行分割。

一般而言， $l_0$  的取值不小于 25， $P_0$  可取 0.5-0.95。可以根据不同取值的  $l_0$  和  $P_0$  来改变检测的尺度和精度，以检测不同尺度和幅度的突变。

### 7.1.3 突变点检测过程

将四个行业的日负荷最大值和最小值分别作为 8 个时间序列  $X(t)$ ，并用 BG 算法对该序列进行突变检测。我们根据 BG 算法定义，用 C++ 语言实现。由于日负荷最大最小值序列存在白噪声和尖峰噪声，因此我们通过对  $l_0$  和  $P_0$  进行灵敏性分析，得到当取  $l_0 = 150$ ， $P_0 = 0.85$  时，得到的突变点个数最满足实际个数。我们将统计分割点（突变点）数量以及绘制突变图，用来分析四个行业的突变量级和可能原因。

### 7.1.4 突变点检测结果分析

各行业突变图如图 33-40 所示。图中，蓝色线表示负荷的曲线，橙色线为均值线，均值线突变的位置即为突变点所在的位置。整体上看，三个行业的均值线呈“三峰三谷”的趋势，虽然非普工业在图中表示只有“两峰两谷”，但这是由于非普工业的数据是从 2019 年 10 月 10 日开始的，其他三个行业都是从 2019 年 1 月 1 日开始，因此一定程度上可以说明，四个行业的均值线变化趋势较一致。我们认为，大工业用电负荷的突变点根据原因可大致归为 3 类，分别为节假日突变点、季节性突变点和重大事件突变点。下面将分别对四个行业的突变点进行分析。

#### 1. 大工业用电负荷突变点分析

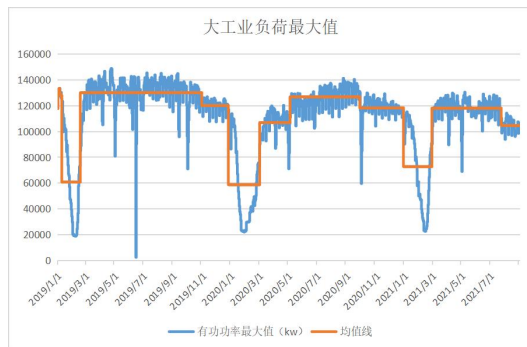


图 33 大工业负荷最大值突变图

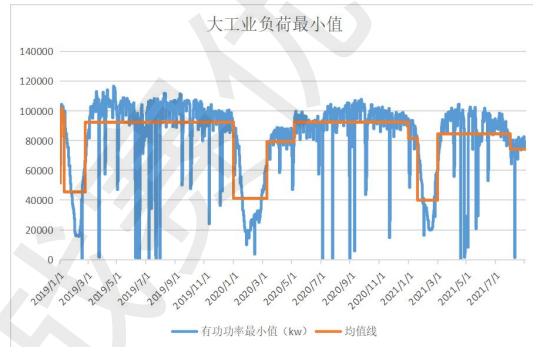


图 34 大工业负荷最小值突变图

从图 33 和图 34 的突变图可以看出，大工业负荷最大值和最小值突变的趋势与个数基本一致，且突变点出现位置具有一定的周期性。其中，负荷最大值有 10 个突变点，负荷最小值有 8 个突变点。表 16 为大工业负荷突变点出现的时间，为了更好地进行对比，表中我们将相似的突变点放在同一行。

表 16 大工业用电负荷突变点

最大值突变点时间	趋势	量级	最小值突变点时间	趋势	量级
2019-1-10	高->低	100MW	2019-1-10	高->低	100MW
2019-2-18	低->高	100MW	2019-2-23	低->高	100MW
2019-11-1	高->低	10MW			
2019-12-28	高->低	100MW	2019-12-30	高->低	100MW
2020-3-27	低->高	50MW	2020-3-9	低->高	50MW
2020-5-5	低->高	10MW	2020-5-5	低->高	10MW
2020-9-29	高->低	10MW			

由表 16 可知，负荷最大值突变点和最小值突变点出现位置基本相同。通过对比分析，我们可以将大工业突变分为以下两种类型的突变：

- 季节性突变点

此类突变点大多是由于季节气象变化所引起的负荷突变，突变量级一般在10MW，如日负荷最大值在2019年11月负荷降低（季节转冬）、负荷最大值和最小值在2020年5月升高（季节转夏），但由于突变变化不明显，我们认为季节温度变化对大工业用电负荷影响不大。而且从2019年不存在季节性突变、负荷变化曲线来看，也能侧面看出季节温度变化对大工业用电负荷影响不大。

- 节假日性突变点

节假日前，由于节假日公休，工人停业休息，会引起用电负荷从高用电突变为低用电，突变量级一般为10MW；在假期后，工厂会陆续开工，此时会引起用电负荷从低用电突变到高用电，突变量级一般为10MW。类似的假期有国庆节，劳动节。此类突变点如表中蓝色部分所示。

大工业用电负荷最大值最小值都会在每年春节期间负荷会先从突变到低，再突变为高，突变量级一般为100MW。特别地，2019年春节结束后负荷由低抬高的过程相较其它年份较为缓慢，较大可能是受COVID-19新型冠状病毒肺炎疫情（以下简称新冠疫情）爆发初期影响。

## 2. 非普工业用电负荷突变点分析

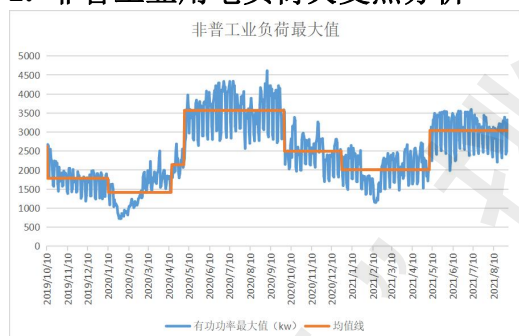


图 35 非普工业负荷最大值突变图

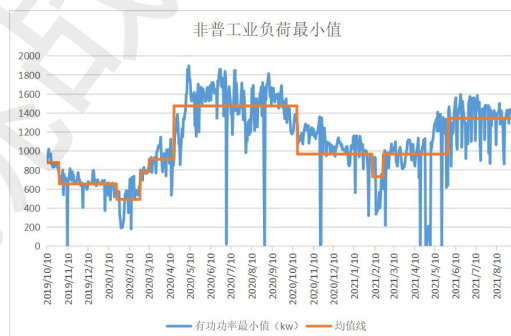


图 36 非普工业负荷最小值突变图

从图 35 和图 36 的突变图可以看出，大工业负荷最大值和最小值突变的趋势与个数基本一致，且突变点出现位置具有一定的周期性。其中，负荷最大值有 6 个突变点，负荷最小值有 7 个突变点。表 17 为非普工业负荷突变点出现的时间，为了更好地进行对比，表中我们将相似的突变点放在同一行。

表 17 非普工业用电负荷突变点

最大值突变点时间	趋势	量级	最小值突变点时间	趋势	量级
2020-1-10	高->低	1000kW	2020-1-20	高->低	100kW
			2020-3-4	低->高	100kW
2020-5-6	低->高	1000kW	2020-5-6	低->高	500kW
2020-9-19	高->低	1000kW	2020-10-10	高->低	500kW

2020-12-20	低->高	500kW			
			2021-2-2	高->低	100kW
			2021-2-25	低->高	100kW
2021-5-6	低->高	1000kW	2021-5-10	低->高	500kW

对于非普通工业区而言，主要是受到季节性因素的影响而造成的突变，在进入夏季前用电负荷的最大值会突然上升，量级为 1000kW，并且会达到和维持在一个新的用电水平。这也表明了非普通工业区在夏季的时候具有较大的用电趋势。而在 9 月、10 月降温后产生 500kW 到 1000kW 量级，从高到低的负荷突变。一定程度上表明了该地区普通工业用电对于高温比较敏感。

另一方面，我们同样观察到 2020 年年初时的第一个由低到高的突变的抬升过程相较 2021 年有所放缓，我们认为这同样是新冠疫情爆发初期的影响。

### 3. 普通工业用电负荷突变点分析

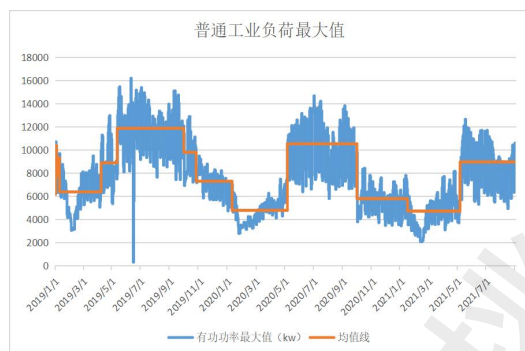


图 37 普通工业负荷最大值突变图

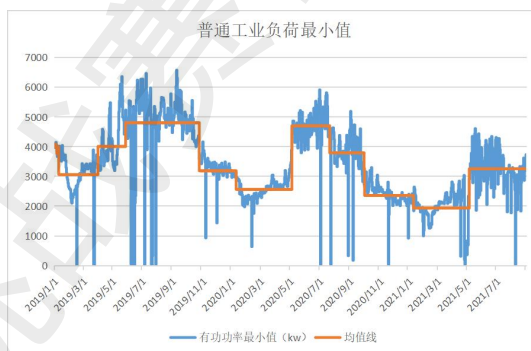


图 38 普通工业负荷最小值突变图

表 18 普通工业用电负荷突变点

最大值突变点时间	趋势	量级	最小值突变点时间	趋势	量级
2019-1-5	高->低	1MW	2019-1-5	高->低	1WM
2019-4-5	低->高	1MW	2019-4-5	低->高	1WM
2019-5-6	低->高	5MW	2019-5-9	低->高	1WM
2019-9-29	高->低	1MW	2019-10-2	高->低	1WM
2019-10-25	高->低	1MW			
2020-1-5	高->低	1MW	2020-1-5	高->低	1WM
2020-5-7	低->高	5MW	2020-5-7	低->高	5WM
			2020-7-10	高->低	1WM
2020-9-30	高->低	5MW	2020-9-29	高->低	1WM
2021-1-5	高->低	1MW	2021-1-5	高->低	0.5MW
2021-5-6	低->高	5MW	2021-5-4	低->高	1WM



对于普通工业而言，存在较多的季节性突变。可以通过观察图表，可以知道每当在春夏换季、夏秋换季的时候，普通工业最大用电负荷就会出现负荷突变，量级为 5MW。这一定程度上表明了该地区普通工业用电对于高温比较敏感。

#### 4. 商业用电负荷突变点分析



图 39 商业负荷最大值突变图



图 40 商业负荷最小值突变图

相较于普通工业和非普工业，商业该地区商业用电负荷水平相对较高，和大工业的总体用电水平在同一量级。但在 2019 年 11 月 1 日的 150MW 量级的突变后，该地区商业用电负荷日最大值的总体水平降低至一个新的水平，此次突变我们归类为突发事件导致的突变，例如电网调整等。另一方面，该地区的商业用电日最小负荷则基本没有受到影响。

表 19 商业用电负荷突变点

最大值突变点时间	趋势	量级	最小值突变点时间	趋势	量级
2019-1-6	高->低	50MW	2019-1-5	高->低	5MW
2019-4-20	低->高	10MW	2019-5-5	低->高	5MW
2019-6-17	低->高	50MW	2019-6-17	低->高	1MW
2019-11-1	高->低	150MW	2019-10-25	高->低	5MW
2020-1-5	高->低	10MW	2020-1-5	高->低	5MW
2020-5-7	低->高	50MW	2020-5-7	低->高	10MW
2020-9-30	高->低	50MW	2020-9-29	高->低	5MW
2021-3-20	低->高	10MW	2021-3-10	低->高	1MW
2021-5-6	低->高	50MW	2021-5-4	低->高	5MW

新型冠状病毒由 2019 年 12 月爆发，由图中，我们可以清楚的看到，在疫情发生的时候，商业的用电负荷存在突变点，并且随着疫情不断的持续，可以看到相较于前年，商业的用电负荷不断维持在一个较低的水平，这说明该地区的商业受到了疫情较大的冲击。



## 7.2 行业负荷中期预测与分析

### 7.2.1 行业负荷模型中期预测

我们将运用构建好的具有温度累积效应的 RBF 神经网络短期负荷预测模型来预测该地区各行业未来 3 个月日负荷最大值和最小值的预测结果，并对其预测精度做出分析。

我们选用 2018 年 9 月 1 日到 2021 年 5 月 30 日的数据作为训练集, 选用 2021 年 5 月 31 号到 2021 年 8 月 31 号共 93 天的数据作为测试集, 模型输入为已经构建好的行业的中期负荷预测特征工程, 输出为当天负荷最大值, 最小值。

首先我们利用训练集求解出 RBF 网络中的权重矩阵  $W$ , 然后基于测试集的特征工程和  $W$  得到测试集的预测值。通过与测试集的真实值进行比较, 得到均方误差 error。进行训练之前, 我们利用遗传算法找到合适的隐含层神经元数目和扩展常数使得在测试集上的均方误差 error 达到最小。

### 7.2.2 行业极值负荷预测结果

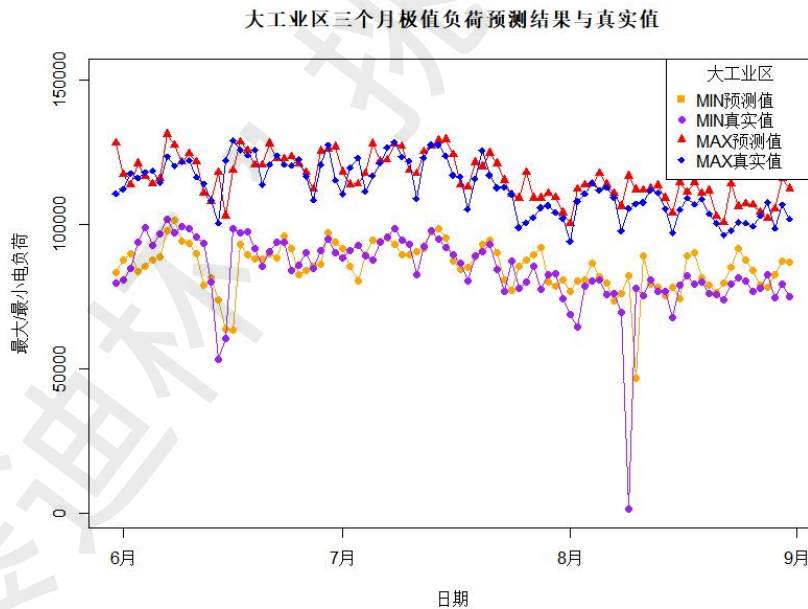


图 41 大工业区三个月极值负荷预测结果对比图

非普工业区三个月极值负荷预测结果与真实值

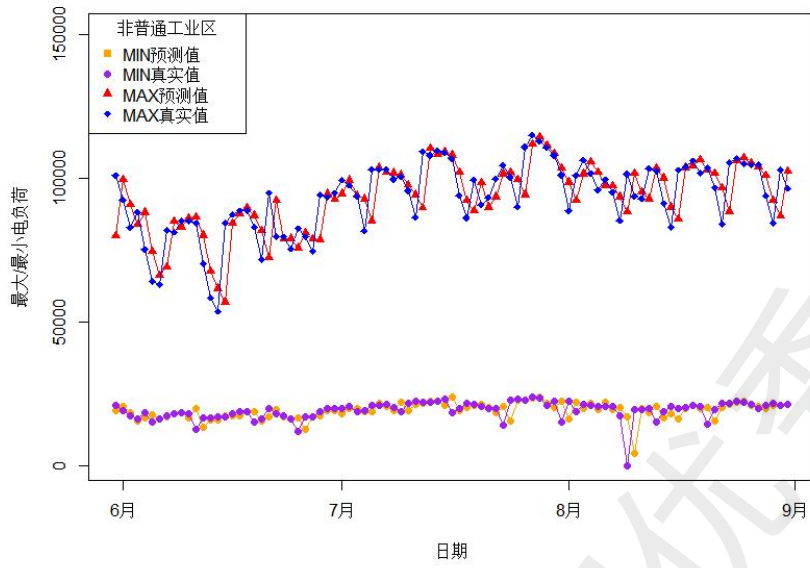


图 42 非普工业区三个月极值负荷预测结果对比图

普通工业区三个月极值负荷预测结果与真实值

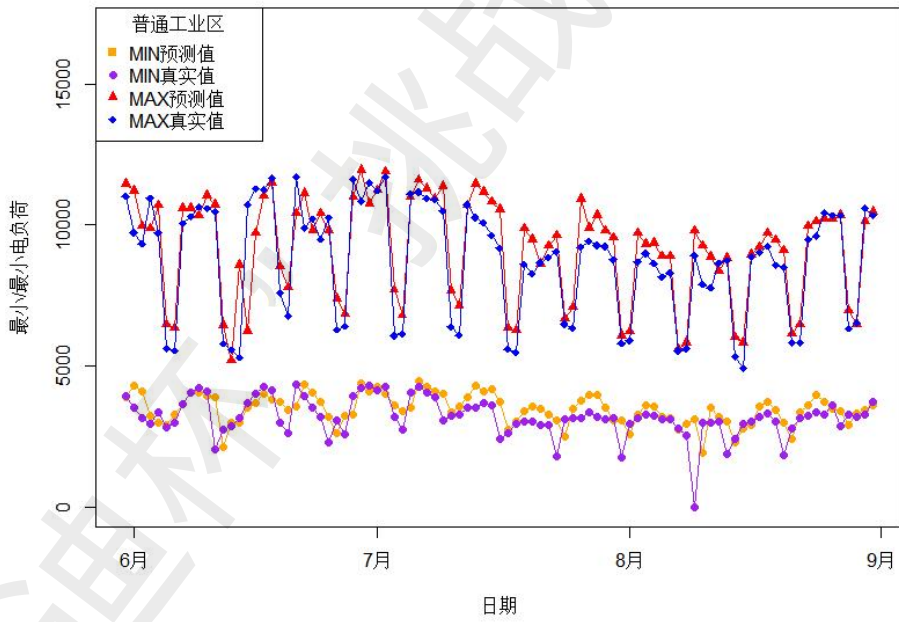


图 43 普通工业区三个月极值负荷预测结果对比图

商业三个月极值负荷预测结果与真实值

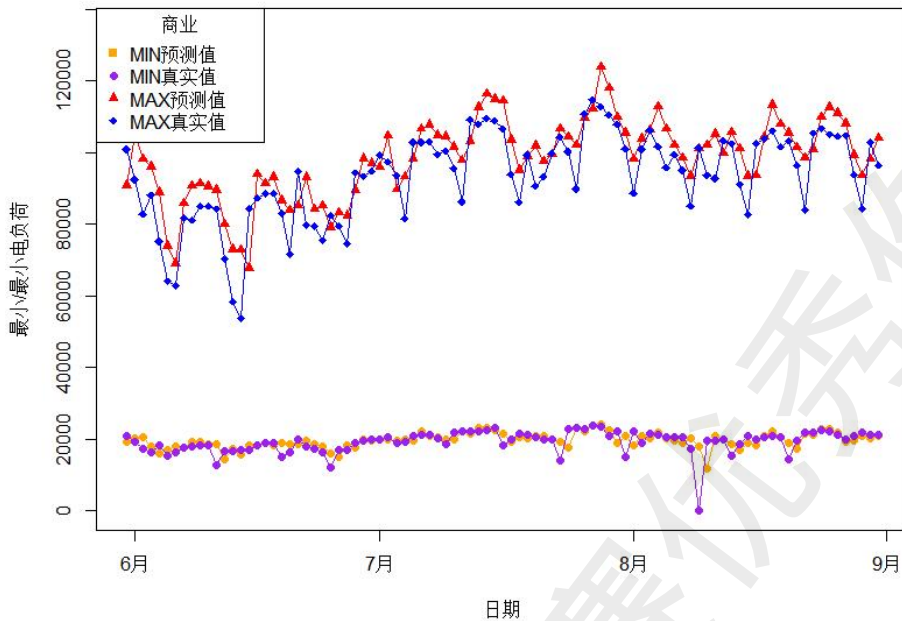


图 44 商业三个月极值负荷预测结果对比图

上述图 41-44 是基于 RBF 预测模型的四个行业电负荷日最大值、最小值与真实值可视化，可以很明显的看出预测的效果基本符合了实际负荷变化趋势，这说明了我们的特征工程特征的有效性。但是我们可以发现有一些预测点与真实值存在较大的误差，刚好这些点都出现了异常的跳动，即使停电和跳高的情况，这种情况往往是和特殊的事件有关，作为预测模型我们很难去预测出有些未知特殊事件的发生，但是总体上 RBF 在预测最值负荷总体上表现的效果还不错。

### 7.2.3 行业极值负荷预测模型比较

#### 1. 大工业区电负荷

表 20 大工业区电负荷预测结果

大工业区 时段	实际最大负 荷	实际最小 负荷	RBF				LSSVM			
			负荷最大值 预测值	相对误 差	负荷最小 值预测值	相对误 差	负荷最大值 预测值	相对误 差	负荷最小 值预测值	相对误 差
2021.5.31	110488.328	79503.118	127889.636	3.75%	83238.789	4.70%	113260.968	2.51%	63407.394	-20.25%
2021.6.1	112082.365	80809.612	117051.081	4.43%	87422.086	8.18%	108925.029	-2.82%	93040.460	15.14%
2021.6.2	117490.368	84615.709	113512.099	-3.39%	89588.492	5.88%	103700.493	-11.74%	89199.290	5.42%
2021.6.3	115925.471	93710.027	120893.868	4.29%	83514.930	-10.88%	114419.686	-1.30%	88076.370	-6.01%
2021.6.4	117920.147	98894.778	116365.647	-1.32%	85412.087	-13.63%	111025.790	-5.85%	87846.616	-11.17%

2021.6.5	118492.846	92684.845	113937.771	5.75%	87514.506	-5.58%	114332.511	-3.51%	89837.998	-3.07%	
...	...	...	...	...	...	...	...	...	...	...	
2021.8.30	106633.835	79252.946	115914.034	8.70%	87125.077	9.93%	117803.187	-5.85%	89199.290	-11.17%	
2021.8.31	101834.844	74755.666	112216.570	10.19%	86959.519	16.32%	102800.097	-3.51%	88076.370	-3.07%	
平均相对 误差				<b>4.436%</b>			<b>6.436%</b>			<b>9.635%</b>	<b>11.426%</b>

从上表可知，RBF 预测的最大负荷的绝对值平均相对误差为 **4.4356%**，最小负荷的绝对值平均相对误差为 **6.4356%**。除最后几个时段预测相对误差超过  $\pm 4\%$  范围，其他时段的相对误差都在  $\pm 6\%$  以内；

LSSVMR 预测的最大负荷的绝对值平均相对误差为 **9.634%**，最小负荷的绝对值平均相对误差为 **11.4356%**。除最后几个时段预测相对误差超过  $\pm 4\%$  范围，其他时段的相对误差都在  $\pm 6\%$  以内。

通过比较上述数据与分析，可以得出 RBF 预测模型在大工业区上的极值负荷预测效果比 LSSVMR 好。

## 2. 商业电负荷

表 21 商业区电负荷预测结果

商业区时 段	实际最大负 荷	实际最小 负荷	RBF				LSSVM			
			负 荷 最 大 值 预测值	相对误 差	负 荷 最 小 值 预测值	相对误 差	负 荷 最 大 值 预测值	相对误 差	负 荷 最 小 值 预测值	相对误 差
2021.5.31	100859.654	20803.531	90827.880	-9.95%	19093.722	-8.22%	89559.630	6.32%	18202.472	-1.54%
2021.6.1	92447.807	19063.055	105038.573	13.62%	20327.868	6.63%	79936.310	13.96%	12726.859	-31.28%
2021.6.2	82790.851	17410.132	98167.565	18.57%	20445.360	17.43%	72900.888	24.97%	16715.967	17.83%
2021.6.3	88042.790	16251.227	96087.794	9.14%	17761.443	9.29%	72690.352	35.57%	16456.622	-4.10%
2021.6.4	75160.770	18375.843	88903.003	18.28%	15893.737	-13.51%	67725.744	-19.72%	16793.878	7.41%
2021.6.5	63999.839	15238.965	73915.385	15.49%	16964.720	11.32%	93884.183	7.51%	16765.786	-7.88%
...	...	...	...	...	...	...	...	...	...	...
2021.8.30	62774.945	16247.799	68966.959	9.86%	17781.341	9.44%	91405.380	3.27%	18187.286	0.07%
2021.8.31	81796.158	17432.213	85808.080	4.90%	17598.921	0.96%	93147.681	5.04%	18845.178	-0.42%
平均相对 误差				<b>5.276%</b>		<b>7.475%</b>		<b>8.963%</b>		<b>8.283%</b>

RBF 模型预测最大负荷最小负荷准确率大于 LSSVMR 模型的预测准确率。通过比较上述数据与分析，可以得出 RBF 预测模型在商业上的极值负荷预测效果比 LSSVMR 好。

### 3. 非普工业区电负荷

表 21 非普工业区电负荷预测结果

非普工业 区时段	实际最大负 荷	实际最小 负荷	RBF				LSSVM			
			负荷最大 值预测值	相对误 差	负荷最小 值预测值	相对误 差	负荷最大 值预测值	相对误 差	负荷最小 值预测值	相对误 差
2021.5.31	79779.276	20803.531	100859.654	26.42%	19184.374	-7.78%	86319.549	8.20%	12726.859	-13.82%
2021.6.1	99431.085	19063.055	92447.807	-7.02%	20595.183	8.04%	79924.882	-19.62%	16715.967	-12.31%
2021.6.2	90752.204	17410.132	82790.851	-8.77%	18494.537	6.23%	67522.940	-25.60%	16456.622	-5.48%
2021.6.3	83704.545	16251.227	88042.790	5.18%	15359.741	-5.49%	61494.575	-26.53%	16793.878	3.34%
2021.6.4	88018.168	18375.843	75160.769	-14.61%	16657.202	-9.35%	56858.585	-35.40%	16765.786	-8.76%
2021.6.5	74378.561	15238.965	63999.839	-13.95%	17704.254	16.18%	84222.238	13.23%	18187.281	10.35%
...	...	...	...	...	...	...	...	...	...	...
2021.8.30	106633.835	79252.946	115914.034	8.70%	87125.077	9.93%	117803.187	-5.85%	89199.290	-11.17%
2021.8.31	101834.844	74755.666	112216.570	10.19%	86959.519	16.32%	102800.097	-3.51%	88076.370	-3.07%
平均相对 误差			<b>7.436%</b>		<b>9.436%</b>		<b>10.63%</b>		<b>9.426%</b>	

RBF 模型预测最大负荷最小负荷准确率大于 LSSVMR 模型的预测准确率。通过比较上述数据与分析，可以得出 RBF 预测模型在非普通工业区上的极值负荷预测效果比 LSSVMR 好。

#### 4. 普通工业区电负荷

表 22 普通工业区电负荷预测结果

普通工业 区时段	实际最大负 荷	实际最小 负荷	RBF				LSSVM			
			负荷最大 值预测值	相对误 差	负荷最小 值预测值	相对误 差	负荷最大 值预测值	相对误 差	负荷最小 值预测值	相对误 差
2021.5.31	11005.353	3945.925	11462.765	4.16%	3897.566	-1.23%	6246.401	-43.24%	3534.258	-10.43%
2021.6.1	9729.1236	3520.882	11203.602	15.16%	4312.535	22.48%	9715.490	-0.14%	3707.954	5.31%
2021.6.2	9326.973	3151.417	9954.145	6.72%	4121.877	30.79%	11028.617	18.24%	4040.475	28.21%
2021.6.3	10947.946	2959.081	9868.626	-9.86%	3222.684	8.91%	11491.106	4.96%	3831.298	29.48%
2021.6.4	9708.997	3383.842	10690.568	10.11%	2978.574	-11.98%	8523.846	-12.21%	3731.049	10.26%
2021.6.5	5619.949	2841.034	6464.630	15.03%	2928.208	3.07%	7779.096	38.42%	3454.416	21.59%
...	...	...	...	...	...	...	...	...	...	...
2021.8.30	106633.835	79252.946	115914.034	8.70%	87125.077	9.93%	117803.187	-5.85%	89199.290	-11.17%
2021.8.31	101834.844	74755.666	112216.570	10.19%	86959.519	16.32%	102800.097	-3.51%	88076.370	-3.07%
平均相对 误差			2.436%		2.846%		3.654%		4.23%	

RBF 模型预测最大负荷最小负荷准确率大于 LSSVMR 模型的预测准确率。通过比较上述数据与分析，可以得出 RBF 预测模型在普通工业区上的极值负荷预测效果比 LSSVMR 好。

## 8. 结论与建议

### 8.1 结论

区域负荷总体来看存在季节性规律、节假日规律、温度高相关性和高温留存效应规律。单日来看负荷曲线具有两峰两谷的特性，和人的普遍生产活动与作息规律高度相关。

行业负荷数据中，大工业用电日最大负荷存在节假日性、季节性以及突发事件导致的负荷突变，其中春节节点前后用电负荷突变量级最大，在 100MW 级别，而夏季开始和结束节点的突变量级则为 10MW。2020 年年初的负荷抬升由于起始有所放缓，突变量级只有 50MW 而不是正常的 100MW。大工业用电日最大和

最小负荷相近,说明日用电较为稳定。同时年均日最大和最小负荷数据存在递减趋势。

普通工业日最大负荷数据中,同样存在节假日性、季节性和突发事件导致的负荷突变。但影响级别有所不同。普通工业受季节性因素影响较大,特别是对高温比较敏感,每当在春夏换季、夏秋换季的时候,普通工业最大用电负荷就会出现负荷突变,量级为 5MW。而节假日性的突变量级为 1MW。同时普通工业年均日最大和最小负荷数据存在递减趋势。

非普工业日最大负荷数据中,同样存在节假日性、季节性和突发事件导致的负荷突变。非普工业受季节性因素影响较大,进入夏季前会产生 1000kW 量级的升变,并在夏季维持一个相对较高的用电水平。在 9、10 月降温后会产生 500kW 到 1000kW 量级的负荷降低突变。非普工业春节节点的突变量级为 1000kW,相较季节性突变没有特别突出,说明其对于节假日,特别是春节的敏感性不如大工业用电强。

商业负荷日最大数据中,突发事件导致的负荷突变量级最大,特别是 2019 年 11 月 1 日的负荷降低突变,量级为 150MW,存在大量级、跳变快的特点,突变前后总体用电水平差距大。而节假日性突变的量级相对较小,在 10MW 级别。季节性突变量级为 50MW。

## 8.2 针对性建议

习近平主席发表讲话指出,我国要力争在 2030 年前实现碳达峰,力争在 2060 年前实现碳中和,这就是所谓“双碳”目标。实现“双碳”目标,离不开产业结构的调整与经济发展方式的转型升级[6]。

针对该地区大工业用电,其存在用电上限规模大、负荷规律相对稳定和春节节点最大负荷突变量级大的特点。而我国为最终实现“双碳”目标,在降低化石能源终端消费比重方面限制了火力发电和火电使用。同时碳排放权交易和碳税政策等市场化减排政策推动下,进一步抬高了高耗能产业部门的能源使用成本。有研究指出[7],根据党的十九大报告和我国向国际社会做出的承诺,在三个减排模拟场景下,相比其它行业,高耗能行业的产业结构比例呈现出最大的降比。这意味着在碳减排政策下,大工业用电行业将受到较大程度的影响,特别是火电的使用成本将进一步提高。

对此我们首先建议大工业用电行业加快产业用电结构调整,但要注意不能盲目快速减少火电使用,而是应当因地制宜地充分考虑清洁能源的稳定性、负载能力和大量级负荷突变承受能力能否满足自身行业的用电需求,循序渐进地对火电使用进行替代。其次是建议推动节能减排技术创新,以控制生产成本和利益。因为碳达峰碳中和目标中是要尽可能减少碳排放,而不是不排放[8]。而且大工业



用电的计费中，容量电费部分是按最大用电需求来计费的，不受实际用电影响。所以通过改进优化火电的使用效率以及降低最大负荷需求是目前“碳达峰”目标下大工业行业切实可行的，对经济发展负面影响小的一种应对方式。如此一来，一方面可以响应“双碳”号召，另一方面可以控制生产成本。

针对该地区普通工业用电行业，其存在高温敏感和年均用电负荷逐年递减的特点。普通工业同样受“双碳”影响较大，我们首先建议应当针对夏季高温合理控制用电负荷。其次是可以参考大工业用电进行用电效率技术优化和能源结构改革。但在考虑清洁能源的进一步引入时，可以适当放宽对清洁能源稳定性和载荷能力的要求。

针对此地非普工业用电行业，其存在受季节性影响和突发事件影响较大的特点。同时相较其它用电行业，非普同业用电总体量级最小。对此，我们强烈建议非普工业各企业关注碳排放权交易和碳税政策。因为我国引入市场工具的碳排放控制后，用电规模较大的行业受限更大，而用电规模较小的行业则影响较小。而且碳排放少的企业可以出售自身的碳排放权，这可以为企业带来一定程度上的收益。其次是在碳排放政策的基础上，进一步压低碳排放和火电用电量，提高用电效率，可以带来更多经济效益和环境效益。

针对该地区商业用电，其存在用电规模上限较大，受突发事件影响大的特点。我们知道自新冠疫情爆发以来，我国政府秉持人民生命健康至上的理念，多次果断按下经济的“暂停键”，一定程度上减缓甚至阻止了疫情的进一步蔓延。此次新冠疫情对多个行业有不同程度的影响，其中服务业这类人员接触密集型产业受此次新冠疫情的冲击最大，而互联网等高新技术产业则从中获益良多[9]。

对此我们首先建议本地商业界对于疫情这种持续影响长的突发事件，需要关注疫情防控封锁性措施的正式实施和解除这两个节点，这两个节点前后的用电负荷突变较大[10]，商业界和供电行业都要做好充足准备进行应对。其次是时刻关注电网调整信息，提前对拉闸限电和大规模断电等突发事件准备应急方案。

## 参考文献

- [1] 万志宏. 基于时间序列的电力系统短期负荷预测研究[D].华南理工大学,2012.
- [2] 康重庆,夏清,张伯明.电力系统负荷预测研究综述与发展方向的探讨[J].电力系统自动化,2004(17):1-11.
- [3] 游浩麟, R 语言预测实战, 北京: 电子工业出版社, 2016: 415-441
- [4] 唐良艳. 电力系统负荷特性分析与负荷预测研究[D].华南理工大学,2010.
- [5] 符淙斌,王强.气候突变的定义和检测方法[J].大气科学,1992(04):482-493.
- [6] 史丹,李鹏.“双碳”目标下工业碳排放结构模拟与政策冲击[J].改革,2021(12):30-44.
- [7] 朱佩誉,凌文.不同碳排放达峰情景对产业结构的影响——基于动态 CGE 模型的分析[J].财经理论与实践,2020,41(05):110-118.DOI:10.16339/j.cnki.hdxbcjb.2020.05.015.
- [8] 能源发展网 <http://www.chinapower.com.cn/xw/zyxw/20210407/64285.html>
- [9] 寇宗来. 新冠疫情冲击对商业模式和社会治理的影响[J].东北财经大学学报, 2020, (5): 3-13.
- [10] 卢德龙,郭聚一,吴阳.新冠肺炎疫情影响下基于多源数据驱动的电力系统负荷预测方法[J].供用电,2022,39(01):74-80.DOI:10.19421/j.cnki.1006-6357.2022.01.11.