



同濟大學
TONGJI UNIVERSITY

硕士学位论文

城区短时电力负荷组合式预测算法研究

姓名：李梦西

学号：1333097

所在院系：机械与能源工程学院

职业类别：工程

专业领域：动力工程

指导教师：许鹏

二〇一七年四月



同濟大學
TONGJI UNIVERSITY

A dissertation submitted to

Tongji University in conformity with the requirements for the degree of

Master

**Study on the Combined Algorithm of urban
Short-term Load Forecasting**

Candidate: Li Mengxi

Student Number: 1333097

School/Department: School of
Mechanical Engineering

Discipline: Engineering

Major: Power Engineering

Supervisor: Xu Peng

April, 2017

城区短时电力负荷组合式预测算法研究

李梦西

同济大学

学位论文版权使用授权书

本人完全了解同济大学关于收集、保存、使用学位论文的规定，同意如下各项内容：按照学校要求提交学位论文的印刷本和电子版本；学校有权保存学位论文的印刷本和电子版，并采用影印、缩印、扫描、数字化或其它手段保存论文；学校有权提供目录检索以及提供本学位论文全文或者部分的阅览服务；学校有权按有关规定向国家有关部门或者机构送交论文的复印件和电子版；在不以赢利为目的的前提下，学校可以适当复制论文的部分或全部内容用于学术活动。

学位论文作者签名：

年 月 日

同济大学学位论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，进行研究工作所取得的成果。除文中已经注明引用的内容外，本学位论文的研究成果不包含任何他人创作的、已公开发表或者没有公开发表的作品的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明。本学位论文原创性声明的法律责任由本人承担。

学位论文作者签名：

年 月 日

摘要

电力是关系国计民生的基础能源，随着电力市场化改革的深入，对日前电力负荷基线预测的精确度有了更高的要求。准确的预测不仅是安全可靠供电的保障，也是实现电力市场调控和交易，调节市场电价的重要标杆。

本课题旨在通过组合式预测算法的搭建，实现更为准确的短时电力负荷预测，同时力求模型的简化性和良好的泛化性。本文借鉴暖通空调相关知识，在深入分析城区电力负荷特性，考察电力负荷与气象因素的相关性后，确定了负荷先拆分，后预测，再组合的预测方法论。对全年出现的两种日属性和三种季节属性均根据上述方法论建立出相应的预测模型，所对应的模型种类共计 6 类。通过与传统的直接以总电力负荷为目标值建立模型进行预测的方法相比，证明了该方法具有更高预测精度的优越性。其中对工作日日前逐 15 分钟电力负荷的预测误差由 3%-5% 下降到 2% 以内，对节假日日前逐 15 分钟电力负荷的预测误差由 4%-5% 下降到 4% 以内。

方法论的拆分依据是基于气象因素与电力负荷相关性的分析。得出不同季节周期内，影响电力负荷变化的主要气象因素不同，且过渡季电力负荷总体上和气候因素相关性较小的结论，从而定义过渡季节的电力负荷为基础负荷。借鉴度日数法的概念，细化季节属性划分规则。在供热季和制冷季中，首先通过对历史过渡季电力负荷的加权平均得到基础负荷，再通过与总负荷的差分，得到气象敏感型负荷，实现对总电力负荷的拆分。针对部分负荷的不同变化特性，采用不同算法分别建模预测，对于变化较为平稳的基础负荷采用自回归时间序列法，对于与天气非线性映射关系较为复杂的气象敏感型负荷采用支持向量机算法。最终两者加和即可得到对总电力负荷的估计。

本文在对海量电力负荷和气象数据进行处理时，根据统计学习理论对数据进行缺失值和离群值检验，利用决策树建立季节属性判定规则，通过变量间相关性分析和自回归函数确定模型输入参数，最后对拆分出的基础负荷和气象敏感型负荷分别采用时间序列法和支持向量机算法建立模型，实现了一套完整的基于拆分再预测理念的负荷预测方法，并取得较高的预测精度。

关键词：短时电力负荷预测，负荷拆分算法，组合式算法预测，支持向量回归机，时间序列法，决策分类树

ABSTRACT

Electricity is one of the most important basic energy for national developing. With the deepening of the electricity market reform in China, it attaches great importance to improving the accuracy of load forecasting. Accurate short-term load forecasting (STLF) guaranties not only the safe and reliable power supply, but also is an essential benchmark in electricity market regulation and trading.

The paper presents a load segment based approach to developing combined forecasting algorithm for predicting the next-day load, with the aim of improving the prediction accuracy, simplifying models and having a better generalization performance. Taking reference of the knowledge of HAVC the character of urban power load (PL) and the correlation between PL and meteorological factors are deeply analyzed. On this basis, the prediction method load segmented, partly forecasting and re-combining is proposed. According to two different day types and three air-conditioned season attributes in the whole year, six different corresponding types of prediction models are established. Compared with the traditional methods, which the total PL considered as the target value to predict, the accuracy of workday models for next day load in every 15 minutes, which measured by mean absolute percentage error are evidently improved from 3%-5% to under 2%, and holiday models improved from 4%-5% to under 4%.

On the basis of correlation analyze between PL and meteorological and the refining the rules of defining seasonal attribute, the PL in heating and cooling seasons can be divided into basic and meteorological sensitive PL. According to different character of two PL part, auto regression based on time series analyze and SVM are selected to forecast respectively. After re-adding two part of predictive PL together, the value of total PL can be estimated.

Keyword: next day load forecasting, load segment algorithm, combined forecasting algorithm, decision tree, time series analyze, support vector machine

目录

摘要.....	I
ABSTRACT.....	II
目录.....	III
第 1 章 引言.....	1
1.1 研究的背景和意义.....	1
1.2 研究目标及主要内容.....	2
1.3 技术路线及结构安排.....	3
第 2 章 短时电力负荷预测研究现状和常用方法.....	5
2.1 短时电力负荷预测研究现状.....	5
2.1.1 预测算法的选择与优化相关研究.....	5
2.1.2 输入变量选择的相关研究.....	10
2.1.3 组合式预测算法的相关研究.....	10
2.2 短时电力负荷预测存在的不足.....	11
第 3 章 电力负荷变化特性.....	13
3.1 城区电力负荷特点.....	13
3.1.1 典型日负荷特性.....	13
3.1.2 典型周负荷特性.....	14
3.1.3 典型月负荷特性.....	15
3.2 电力负荷影响因素.....	17
3.2.1 气候因素对短时电力负荷的影响.....	18
3.2.2 随机变量常用相关性分析方法.....	20
3.2.3 单一气象因素与短时电力负荷相关性分析.....	20
3.2.4 综合气象参数与短时电力负荷相关性分析.....	26
3.3 本章小结.....	31
第 4 章 电力负荷拆分.....	33
4.1 季节划分定义.....	33
4.1.1 决策树算法.....	33
4.1.2 供热制冷季节判断规则.....	35
4.1.3 当日季节判定.....	37
4.2 电力负荷拆分.....	38

4.2.1 电力负荷异常值检测.....	38
4.2.2 电力负荷拆分.....	40
4.2.3 特殊节假日基础负荷处理.....	43
4.2.4 气象敏感负荷变化特性.....	44
4.3 本章小结.....	46
第 5 章 基于时间序列法的基础负荷预测.....	47
5.1 时间序列法理论介绍.....	47
5.1.1 常用模型介绍.....	47
5.1.2 平稳性检验.....	48
5.1.3 模型识别和滞后阶数确定.....	49
5.1.4 模型检验.....	50
5.1.5 模型评估标准.....	50
5.2 实例分析.....	51
5.2.1 工作日实例分析.....	52
5.2.2 节假日实例分析.....	56
5.3 本章小结.....	57
第 6 章 基于支持向量机的气象敏感负荷预测.....	59
6.1 支持向量机理论介绍.....	59
6.1.1 结构风险最小化.....	59
6.1.2 回归支持向量机基本原理.....	60
6.1.3 回归支持向量机参数优化.....	62
6.2 气象敏感性负荷预测.....	63
6.2.1 模型输入参数选取.....	63
6.2.2 SVR_wh 模型参数优化.....	64
6.2.3 SVR_wh 模型性能分析.....	65
6.2.4 SVR_wc 模型性能评估.....	67
6.2.5 SVR_h 模型性能评估.....	69
6.3 电力负荷组合式模型评估.....	71
6.3.1 组合模型与基于总电力负荷模型对比.....	71
6.3.2 不同输入参数模型对比.....	75
6.4 本章小结.....	77
第 7 章 总结与展望.....	79
7.1 本文总结.....	79

7.2 存在不足与展望.....	80
致谢.....	82
参考文献.....	83
个人简历.....	87

第1章 引言

1.1 研究的背景和意义

电力是关系国计民生的基础能源,安全可靠的电力系统是国民经济增长和社会发展的重要保障。电力系统不同于一般的工业型产业,其产出电能是无法进行大规模存储的,这也就要求电能的生产和消耗达到实时平衡,即保持时刻的按需供应。负荷预测的实质是对电力系统中用户的需求进行预测,精准的负荷预测则与电力系统各个层面息息相关,在电力系统正常运行中起着举足轻重的作用。对需求侧电力负荷的预测是发电企业、输配电部门、调度中心等各部门在电力规划,制定发电计划,安排发电机组维修计划,规划跨区域电力交易等的基准,是确保供电安全性、可靠性和经济性的基础。

对需求侧负荷偏差过大的预测,无疑会给电力系统的可靠运行埋下安全隐患,这将给正常的社会生产生活及供电企业带来极大的损失。一方面,过小的需求侧负荷预测会造成供电系统的备用容量规划不足,从而在集中用电时段出现供电短缺的情况。以美国加州2001年大规模停电为例,由于电力系统备用容量只有1.5%左右,2001年一月中旬出现了发电容量无法满足高峰电力需求的窘境,从而导致了加州强制性的分区轮流停电,上百万人受到影响,致使一些公司损失了数千万美元^[1]。另一方面,电力负荷根据需求侧负荷特性呈现出极大的峰谷差。若为满足极大峰值,一味通过规划配置发电厂、变电站等增加系统冗余量,需要耗费大量资金,且建设周期较长,不仅造成用电谷段电力负荷的浪费,还将造成发电厂设备长期低负荷运行效率偏低,增加了运行成本等损失^[2-4]。因此,随着国民经济和社会发展对电力需求的不断提高,精准的负荷预测变得越来越重要。

近几年来,随着电力市场化改革的逐年深入,电力作为商品的市场特性越来越明显,电已逐步成为可供销售,购买,交易和结算的商品。我国正逐步放宽对电力市场的计划体制,使其向着市场化方面发展。电力市场随即变为以发电侧、配电侧、终端用户及市场和交易调度中心为主体,通过电网这个市场载体而构成的经济体制。电价从垄断下的统一定价向市场定价转变,这也对传统的负荷预测赋予了更多的意义,对其预测结果的准确性提出了更高的要求^[3]。负荷预测所提供的实时负荷需求基线,成为市场电价制定的标杆,同时也是进行需求响应,交易调度中心调控供需平衡的可靠依据^[5]。一切的批发交易、零售交易、需求响应

清算交易都将以更为精准的电力负荷预测为基准。对未来负荷增量、变化的走势决定着供应侧的发电机组配置、配电侧的购电容量、需求响应基线的制定以及市场电力价格的变化^[6]。由此可见，平衡供、配、需求三方的电力可靠供应和利益最大化需要准确把握市场电力变化的发展趋势。在此背景下，精确且具有实际操作价值的负荷预测算法更具意义。

1.2 研究目标及主要内容

基于负荷预测对于传统电力系统安全可靠运行的重要意义，以及对于蓬勃发展的电力市场的指导性作用，本课题将致力于研究组合式预测算法，实现更为准确的日前负荷预测。

电力负荷预测根据对未来预测时间的长短，分为以下四类：1. 超短期负荷预测，即对未来 1 小时以内的负荷进行预测，主要用于电力市场紧急调度和需求侧响应调节；2. 短期负荷预测，其预测目标为未来一天到几天的电力负荷，主要服务于优化供电侧发电计划，高效的安排发电机组运行，便于电力输配侧合理安排调度计划，同时为短期的电价波动提供依据；3. 中期负荷预测，是指预测未来一个月到一年的电力负荷，便于供电侧安排更长周期的发电计划，方便安排机组维护等计划；4. 长期负荷预测，指提前 3 至 5 年的负荷预测，主要用于电力系统的长期规划，便于规划供电容量扩充，电网建设等项目^[3]。本课题结合上海市需求响应工作要求，主要研究短期负荷预测，针对未来一天发生的电力负荷进行预测，为电力的需求调度提供日前需求基线。

本课题将以上海市 2014 年每 15 分钟电力负荷为例进行训练和预测，预测未来一天逐 15 分钟的电力负荷数据，该数据来源于上海市经济和信息委员会数据平台。2014 年气象参数来源于腾天数据平台和 Weather Underground 网站，所包含的气象参数有逐时干球温度，露点温度，相对湿度，含湿量，风速和空气焓值。

本课题的研究目标有三个：一是电力负荷特性研究，在众多学者对电力负荷预测算法的基础上，深入剖析海量电力数据在不同变化周期尺度上的变化特性，构成成分，各构成的变化趋势，以及影响电力负荷变化的影响因素。通过优化运用传统的数据挖掘算法更有效的将海量数据进行有效分类处理和归类；二是对总电力负荷数据拆分算法进行研究。在上述电力负荷变化特性分析的基础上，将电

力负荷拆分成基础负荷和气象敏感负荷，为后续提高预测精度奠定基础；三是基于对负荷数据的拆分进行组合式负荷预测算法研究。一方面结合专业知识分析不同影响因素对于拆分出的不同负荷数据的相关性及影响权重，从而优化预测模型的输入参数，简化模型的复杂程度。另一方面在对拆分的负荷数据的变化特性有了深入理解的基础上，采用合适的算法建立预测模型，在保障预测结果精准性的前提下，力求组合式算法的简易性，提高其普适性。

综上所述，本科题主要借助建筑能耗的背景知识，运用组合式机器学习算法，达到从大量电耗数据中提取有效的、潜在有用的、最终可被理解和输入模型完成预测的过程。首先完成针对上海地区的短时负荷实时较精准较快捷地预测，这对上海电力市场化改革，需求响应工作的基线预测都具有深远意义；其次将该算法工程化，将可操作的算法程序化，具有实际操作价值。力求达到电力系统提前掌握电力市场实施发展动向，完成机组的优化调节；对企业层面可在需求侧调控实现利益最大损失最小化。

1.3 技术路线及结构安排

针对上述研究目标及主要研究内容，本课题将按照下图所示的技术路线进行研究。本课题的章节也将按照如下的技术路线划分。

首先第二章将在课题研究前总结现阶段国内外对该课题的研究现状与存在的问题，通过归纳总结出目前针对短时负荷预测常用的算法，以及各个算法的适用范围和优缺点，为本课题后续算法的选择提供有力依据。概述当前针对本课题国内外学者所取得进展，着重针对研究存在的不足给出拟采取的解决方案，并提出本课题的方法论。

第三章是本课题电力负荷预测方法论的基础，将涵盖上图技术路线的前两列内容。本章节从分析负荷变化特性入手，结合其变化规律总结影响负荷变化的主要因素。负荷变化趋势根据观测时间的长短具有明显且不同的规律特性，通过分析负荷日、周、月和年尺度上的变化，将得出负荷变化具有周期性变化且受不同影响因素影响权重不同的结论。在分析各个影响因素时，着重研究了气象因素对电力负荷在不同时间观测尺度上的相关性，不仅对各个单一气象因素对负荷的影响，同时也对多个气象因素协同作用对电力负荷的影响进行了考察，为第四章的负荷拆分和预测模型输入变量的选择垫定了坚实的理论依据。

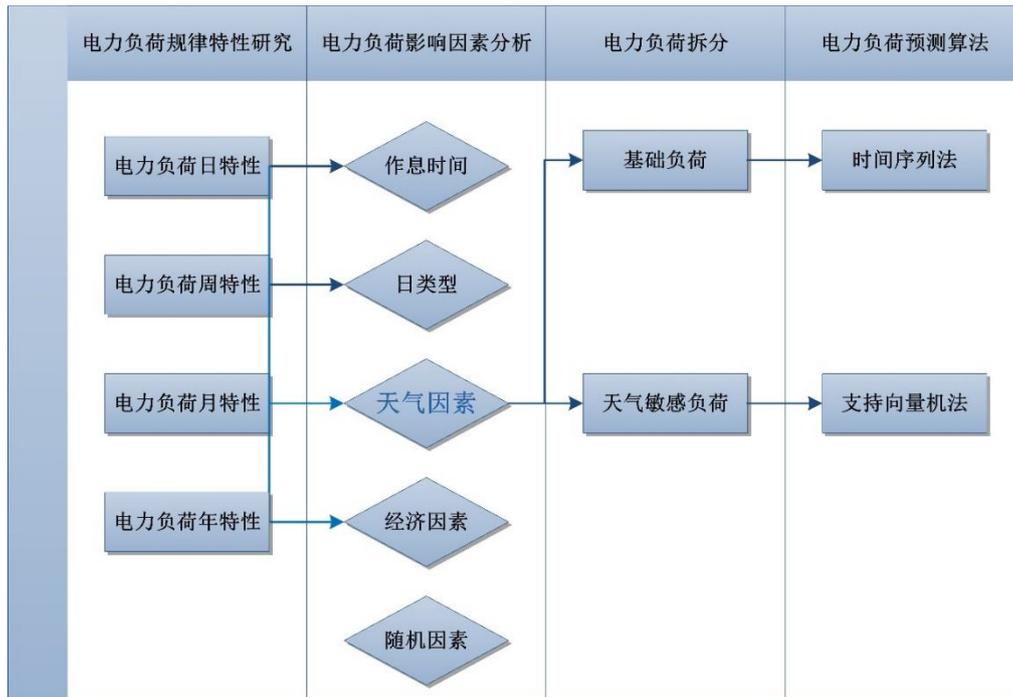


图 1.1 技术路线

第四章给出了技术路线中第三列的内容，通过设计拆分算法将总电力负荷数据拆分成基础负荷和气象敏感型负荷两个部分。该算法的实现依赖于第三章得出的结论，即不同时间段内电力负荷受不同影响因素的权重不同的结论。通过决策树算法定义了季节属性的概念，即不同时间段，并在此基础上通过差分将总电力负荷进行拆分，并就不同部分负荷的变化特性进行了分析。该章节是本课题组合式预测算法方法论中的重要内容，是第五章和第六章预测算法的选择和提高预测精度的基础。

第五章预测算法的研究对象为基础负荷部分，涵盖了技术路线第四列第一项内容。针对基础负荷的变化特性，采用简单的电力负荷自回归模型，即时间序列法建立模型，对第二天电力负荷进行预测。并在单一预测值的基础上，给出预测值的置信区间。

第六章的研究对象则为气象敏感型负荷，采用的预测算法为非线性关系拟合较好的支持向量机算法。通过实例建立不同季节属性和日属性的预测模型，将模型预测性能与传统非拆分算法建立的预测模型的性能进行比较，判定上述拆分组合预测的方法论是否具有优越性，模型精度是否提高。

第七章则对所得结论进行总结，并就课题中研究的不足提出有待改进的建议。

第2章 短时电力负荷预测研究现状和常用方法

本章节首先对国内外短时电力负荷研究现状和主要研究方向进行总结，并综述目前较为常用的预测算法，总结各个常用算法的使用范围和优缺点，为本课题后续算法的选择提供有力依据。其次针对现阶段短时电力负荷预测存在的问题进行总结，基于目前研究中存在的不足会在课题后续部分提出改进方法。

2.1 短时电力负荷预测研究现状

对于电力负荷的预测研究早在上世纪八十年代就已蓬勃发展，总结下来，主要的研究方向大体可分为以下三类：1) 预测算法的选择与优化；2) 输入变量的选择；3) 与其他算法的组合式应用。

2.1.1 预测算法的选择与优化相关研究

由于电力负荷影响因素的多样性和复杂性，各个因素对负荷变化的影响呈现复杂非线性的特点，目前并不存在一个准确性较高同时具有较高泛化性的预测算法，从而涵盖负荷变化中的种种情况^[7]。随着人工智能技术的不断发展，短时电力负荷预测仍是国内外学者们锲而不舍研究的课题。短时电力负荷预测方法随着科学技术对数据挖掘及处理能力的提升有了长足的发展，各种各样的短时电力负荷预测算法不断涌现。下面将对较为常用的短时电力负荷预测方法作综述性总结，归纳起来可分为两类：一是统计方法，二是人工智能算法。

统计方法需要建立一个明确的数学模型，建立负荷和不同输入变量之间的关系。通过给出预测时刻输入变量值，计算出预测时刻的电力负荷值。应用于短时电力负荷预测几种经典的统计方法有多元线性回归法^[8]，时间序列方法^[9, 10]，状态空间模型^[11]和卡尔曼滤波^[12]。通常统计方法对工作日变化较为平稳的负荷曲线的预测性能较好，但是由于结构的不灵活性，缺乏分析节假日等异常日的负荷特性的能力。本文着重分析时间序列法。

时间序列法 (Time Series Analyze)

时间序列法把负荷数据看作是一个按季节、按周、按天以及按小时周期性变化的时间序列。算法假设未来负荷数据是基于历史数据的时间延伸，反映在随机

过程理论中就是时间序列的平稳性或准平稳性，是一种线性自回归过程^[9]。较为常用的模型有移动平均模型（Moving Average, MA），自回归模型（Auto Regressive, AR），自回归移动平均模型（Auto Regressive Moving Average, ARMA）。上述模型均适用于平稳时间负荷序列，分别通过当前时刻随机干扰量和过去有限项干扰量的加权和，过去有限项负荷值的加权和，以及二者的加和对未来时刻负荷进行预测^[10]。在 ARMA 的基础上衍生而来的 ARIMA，是对非平稳时间负荷序列进行有限项差分得到平稳序列，进而建立 ARMA 模型进行预测。

时间序列法在短时负荷预测应用上具有如下优势：1）预测模型较为简单，易于理解；2）进行预测时所需数据仅为历史负荷数据，数据需求量较小，计算量较小。

但该方法的局限性也较为明显。1）影响因素与电力负荷之间规律挖掘较为片面。值得注意的是，历史负荷线性关系的假定通常不包括温度等天气因素的影响。在这种情况下一致认为，这种非线性关系已经在模型中得以保留^[13]。2）历史数据质量要求较高。时间序列法的本质特征是要求相邻数据具有较高的相关性，脏数据对预测误差有着较大影响。

经过数十年的研究，时间序列法用于短时电力负荷预测已较为成熟。在传统的模型参数确定时，主要采用最大似然估计或最小二乘法，这也就要求负荷序列的噪声为高斯噪声，即满足均值为 0，方差为 σ^2 的分布^[10]。Huang 和 Shih，通过引入累积量和傅里叶变化中双谱的概念检验负荷序列噪声的分布，对不服从高斯分布噪声的负荷序列建立 ARMA 模型，取得较高的预测精度，拓展了该方法的应用范围^[13]。此外 Eberhart 和 Kennedy 将粒子群优化法用于时间序列法模型阶数和参数确定^[14]。M. T. Hagan 等人在短时负荷预测中，在传统的单一以时间为变量的线性回归模型中，考虑了气温对负荷的影响，缩小了平均相对误差。但方法仍以线性模型为基础的，多种随机因素，如气候因素、突发事件等对负荷的走势影响都存在非线性关系，因此该方法无法做到对短时负荷的精准拟合^[15]。在时间序列法只能给出负荷线性关系且需要事先给出模型形式的基础上，雷绍兰运用混沌理论更为客观的探究电力负荷间复杂多层次的非线性关系，提出了基于混沌理论的时间序列预测算法^[16]。

人工智能方法试图模仿人类的思维方式和推理方式，从过去的经验中获取知识，并预测未来的电力负荷。在该领域最为常用的就是人工神经网络算法^[17, 18]。

机器学习算法作为人工智能算法的核心研究领域之一,其不断发展和优化实现了在海量的可利用的数据中发现有用的、面向任务的知识,并能挖掘出观测数据中暗含的潜在关系。其中在短时电力负荷预测领域应用最为广泛的是支持向量机法,小波分析法以及聚类法。

人工神经网络算法 (Artificial Neural Network, ANN)

ANN 是一种模仿人类大脑结构和对信息处理功能的数学模型。在对数据进行处理和学习的过程中,通过建立大量神经元,不断调整每个神经元上的激励函数与神经元间的权重,形成了一个具有自主学习的信息处理网络,实现输入变量与输出变量之间复杂非线性关系或未知函数的逼近。在电力负荷预测领域 ANN 具有如下优势,1) 自主学习能力自适应能力较强,ANN 网络为非线性系统,通过网络结构的训练,拟合出天气因素等与电力负荷的对应关系;2) ANN 网络具有较强的信息综合能力和信息记忆能力,容错率较高,通过各个神经元间权重的调整,实现对互相矛盾的输入信息的调整。3) ANN 模型的建立完全受数据驱动,可实现复杂映射关系的拟合。^[16-19]

ANN 的局限性也较为明显。1) ANN 网络结构,如隐藏层数量的确定较为复杂。需要在预测精度和模型泛化性间进行综合评判。根据 Hippert 等人的理论,随着隐藏节点的数量增多,模型预测精度越高,过拟合风险越大,模型泛化性越低^[20];2) ANN 模型训练过程收敛较慢,且易陷入局部最小值^[18];3) 缺乏有效的方法判断 ANN 模型过拟合或训练不足的问题^[13];4) ANN 模型的建立完全受数据驱动,脱离了待解决问题的物理意义,不易理解。

近几年,基于神经网络对短时电力负荷进行预测的应用研究不胜枚举。总体而言,研究方向可分为四类:1) 算法改进^[21];2) 隐藏层和隐藏神经元数量确定^[22, 23];3) 输入参数和训练集的选取^[24];4) 与其他算法的组合式应用^[21, 25]。在模型输入量研究方面,Abu-El-Magd 等人基于序列的相关系数对模型的输入参数进行选择,用神经网络对峰值负荷进行预测^[24]。

在算法改进和组合式应用方面,黄国栋利用遗传算法具有良好的自适应能力和优化能力,将适应度函数取反,以最小值为优化目标适应神经网络算法的要求,来优化神经网络的输入层与隐含层的连接权值、隐含层阈值、隐含层与输出层的连接权值和输出层阈值四部分。经过广东省某城市负荷数据训练和预测,误差相较原始按误差逆传播的前馈算法有所降低。但该算法是在给定隐含层数目的前提

下进行优化的^[21]。

陈耀武等人提出运用神经网络、模糊聚类分析和模式识别理论方法进行建模的预测方式。首先，采用模糊聚类分析方法，以每天的 24 点负荷数据、天气数据以及日属性数据为指标，将历史数据分成若干类别；其次，对每一类别建立相应的神经网络预测模型；预测时通过模式识别，找出与预测天相符的预测类别，利用相应的神经网络预测模型进行 24 小时的短期电力负荷预测^[24, 25]。在众多的利用神经网络进行预测算法中，均存在隐含层神经元数目无法确定，网络的最终权值受初始值影响较大、存在收敛到局部最小值的风险等。

支持向量机 (Support Vector Machine)

SVM 是由 Vapnik 于 1995 年首次提出的，常用于数据分类和回归。SVM 的提出是基于统计学习理论的机器学习方法，在解决小样本、非线性和高维问题中具有良好的适用性^[26]。SVM 通过将问题转换成求解二次凸规划的数学模型，求得全局最优解，有效避免了有些算法对初始值较为敏感，易陷入局部极值的问题。同时在处理非线性问题上，通过核函数的变化将其映射到高维特征空间，再利用构造最优超平面逼近目标函数，解决低维空间中的非线性问题，使其具有较好的泛化能力。在 Sapankevych 和 Sankar 的研究中表明，SVM 的全局寻优使得模型的泛化性高于其他非线性预测算法，包括 ANN 算法^[27]。在电力负荷预测领域的应用亦取得了令人满意的结果。

Zhang 在对电力负荷预测时，将电力负荷数据和相应的天气数据同时作为输入参数建立 SVM 模型，其预测精度明显高于传统的线性模型，卡尔曼滤波，AR 及 ARMA 模型。在采用相同输入参数的情况，同时建立反向传播神经网络进行预测。对比 SVM 和 ANN 模型，得出 SVM 模型的预测精度更高的结论^[28]。

然而，SVM 的泛化能力高度依赖于参数的适当设置，如惩罚系数，内核参数和损失函数的宽度。因此，选择最优参数对于利用 SVR 处理电力负荷预测任务的良好性能至关重要。在模型参数的选择优化上，学者们采用了最小二乘法^[29]，模拟退火算法^[30]，遗传算法等算法^[31]，均取得了较好的效果。

小波分析法 (Wavelet Analysis)

小波分析法是在傅里叶变换的基础上发展而来的，克服了傅里叶变换在时间尺度上无法给出信号的频率信息的缺点，能够根据信号频率自动调节采样的疏密，有效捕捉到不同频率组成的信号中的层层构成，并将其投影到不同频带尺度上^[32]。

对于电力负荷在不同时间尺度上具有不同的变化规律的特点,采用小波分析法能够更加清晰的拆分出负荷序列的周期性,并根据不同变化特性选择预测算法。

郇能灵等人利用小波变换法将负荷序列进行拆分,对高频率负荷分量采用常规预测方法,其他负荷分量则采用模糊神经网络处理技术,最后通过序列重构,得到完整的负荷预测结果^[33]。宋人杰等人则利用小波分析法把过去直接对负荷序列的预测替代为对小波系数的预测,并对小波细节系数作分层软阈值处理^[34]。

小波变换对电力负荷在频域和时域上的分解,在分解前需要选择适当的母小波函数和分解阶数。对于具有不同变化特性的电力负荷数据集,需要人为介入选择合适的母小波函数,因此模型的泛化性能较低。此外从电力负荷构成这个物理意义而言,依照频率分解得到的不同负荷分量,不具备此物理意义,从而不易理解。

集成学习法 (Ensemble Methods)

集成学习的基本理论是对同一组研究对象采用一系列预测算法进行学习,并使用一定的策略把各个不同的学习结果进行整合,从而获得比单个算法更好的预测结果^[35]。集成算法具备更高的预测精度和良好泛化性的原因有以下三点:首先,训练数据可能无法提供足够的信息来选择最佳模型。因此,不同模型的等效整合可能是一个更好的选择。其次,集成模型可以补偿单个模型在预测时的不完善。最后,在实际中对电力负荷的预测并不存在一个准确的目标函数。集成模型可以提供相对较好的近似,从而获得更好的泛化性能。^[36]

在追求更高预测精度的方向上,不少学者建立了基于不同预测算法的集成模型进行负荷预测。集成算法的建立主要考虑以下两个方面:首先,单个模型的预测性能应尽可能的更准确。其次,所包含的预测方法和模型种类应尽可能多样化,在此多样性主要表现在训练样本处理的不同,输入变量选择的不同,并采用多种学习算法等。^[36]

在 Fan, C. 等人的研究中,采用多元线性回归, ARIMA, SVM, 随机森林算法, 多层感知器, 决策树, 多变量自适应回归样条, k-最近邻居算法八种算法对香港商业建筑日前最大负荷和用电量进行了预测,采用遗传算法对各个算法的输出权重进行了优化,最终模型的平均相对误差小于 3%^[37]。

虽然集成算法因其高预测精度受到了青睐,但其建模过程较为复杂,且在基础模型数量和种类的选择上并不存在统一确定的理论指导,这也加大了该方法的

应用难度。

2.1.2 输入变量选择的相关研究

预测模型输入变量的选择,输入空间维度的确定与模型预测性能、模型简化性息息相关。针对电力负荷预测,输入变量选择的研究则从影响负荷的众多因素中入手,对输入空间维度的控制较为常用的方法为统计法和主元素分析法等。但目前为止,对于不同预测模型仍没有统一的理论衡量和判断输入模型的选择是否充分且合理。

在相关领域研究中,王鹏等人通过分析多种气象因子与电网负荷的相关性,找到气象因素与电网负荷的对应关系,在此基础上给出了经验性的负荷预测修正模型库建立方法。总体说来,6月到9月的日平均气温和日最高气温与用电量之间存在明显的正相关,其中7月份和8月份的相关系数最大,均达到0.8以上。统计分析发现,80%的降水过程均使日用电量减少,平均下降幅度为3%,最大下降幅度约为10%。而且降水对负荷的影响存在一定的延迟,一般影响以后几日的负荷。相对而言风向对电力负荷的影响作用较小^[38]。

上海交通大学学者谢宏等人就提出了一种基于粗糙集理论的解决方法,以此对各种影响因素进行识别,确定神经网络模型的输入变量,并在此基础上通过属性约简和属性值约简获得推理规则。但在阐述影响因素时,只是罗列出日平均温度、日最高温度、日最低温度和日平均湿度为四个影响变量,但并未给出上述影响因素对负荷的影响趋势和相关度的分析^[39]。

林辉等人基于节假日能耗特性与工作日不同,且该方面数据较少的情况,提出了修正节假日负荷预测结果的方法,并提出了气候因子、时间因子和综合因子的匹配系数的概念,在文中所述的校核模型具有较高精度^[40]。

在Fan, C.的研究中,对于输入参数的选择首先进行了特征提取,再通过考察各个特征变量的自相关函数和偏相关函数,判定该输入特征量的时滞量,从而判断其输入维数^[37]。

2.1.3 组合式预测算法的相关研究

在近几年的研究中,组合式预测算法,即对历史数据的预处理、分类拆分、预测模型的建立以及误差分析等步骤,根据其各自特点选用不同算法得到了广泛

关注。此外，在运用人工智能算法时，对模型结构和参数的确定往往采用计算智能技术（Computational Intelligence Techniques, CI），如遗传算法，粒子群寻优法，进化算法等，例如^[21]中基于遗传算法对 ANN 结构的优化确定。

Ko, C. N. 等人提出一套支持向量回归（SVR），径向基函数神经网络（RBFNN）和双向扩展的卡尔曼滤波算法（DEKF）的短时电力负荷组合式预测算法。该算法首先利用 SVR 确定 RBFNN 的结构和初始参数，在给定初值时，利用 DEKF 算法学习优化 RBFNN 的参数，最终确定的径向基函数神经网络算法用以实时负荷预测。经过台湾电力部门的数据测试证明该套组合式算法具有较高的精确性^[41]。

Bahram, S. 等人利用小波变换法结合灰色模型法进行负荷预测，并用粒子群寻优法进行模型结构和参数的优化。模型将气候信息，如平均温度、平均相对湿度、平均风速和前一天负荷值作为模型输入。小波变换法主要完成对前一天负荷数据的高频分量，再利用粒子群寻优对灰色模型生成系数优化，提高整个模型的预测精度^[42]。

Hooshmand, R. A. 等人则利用小波变换和自适应神经模糊推理系统将预测内核算法分解为两个步骤，其中小波变换主要用来提取历史数据和天气信息的低频分量。首先利用小波变换和神经网络法对未来 24 小时负荷进行预测，该步骤的主要输入量为天气参数。其次依靠小波变换法和自适应神经模糊推理系统对对应小时的预测结果进行细化改善。该结果经过伊朗电力负荷数据验证具有较好的精度^[43]。

2.2 短时电力负荷预测存在的不足

从上述的研究现状可以看出，电力负荷预测是一个庞大的研究领域。在从负荷数据的预处理，预测算法的选择，到输入变量的选择，最后到模型的建立与优化，每一步都包含着众多值得深入探讨的数学问题和物理问题。

在阅读相关文献并总结的过程中，作者发现了相关研究中存在的问题：

- i. 数据驱动的预测模型居多，忽略了电力负荷本身构成与变化特性等物理含义。数据驱动模型的建立依赖于数据样本的好坏。通过高质量的数据集，涵盖情况较多的数据集训练出来的模型往往具有较好的预测性能。而在实际情况中，数据样本量往往有限，这也就造成了文献给出的模型

算法泛化性较低的情况；

- ii. 对于电力负荷影响因素的相关分析不全面。主要探讨了气温与电力负荷的关系，且采用日特征量，如日最高温或日平均气温与日最大负荷的关系。对于其他天气因素的影响，以及天气因素对电力负荷延迟作用的研究较少；
- iii. 对短期电力负荷预测模型性能的好坏没有统一的评判标准。在作者阅读的文献中，以平均相对误差（MAPE）作为评判预测精度的居多，取值范围在 1.8%-20%左右。多数文献对短时电力负荷的预测精度集中在 2.5%-5%左右。部分文献在 MAPE 大于 15%时，仍判定预测模型精度较高，预测方法有效。对于该结论，难以让人信服；
- iv. 文献中给出的负荷预测情况和实例普遍较为片面，多集中在对夏季工作日电力负荷的预测，对于节假日和其他季节的负荷预测性能鲜有提及。

针对上述提出的问题，本文将进行如下改进：

- i. 基于暖通空调和建筑能耗知识，对电力负荷的构成和相关影响因素进行分析；
- ii. 探讨天气因素与电力负荷在不同时间观测周期上的相关性中，引入其他天气变量，并修正天气因素对电力负荷的滞后影响；
- iii. 本文提出的组合式模型和与其对比的基准模型，MAPE 精度要求均不高于 5%；
- iv. 通过对全年不同季节，不同日属性负荷建模，验证本文提出的预测方法论的全年适用性。

第3章 电力负荷变化特性

上一章节已对目前常用的短时电力负荷预测方法的优越性以及局限性进行了归纳总结,并指出了当下研究尚且存在的不足。在此基础上,只有深入剖析电力负荷的变化特性,明确其构成原理,在选择预测算法上才能做出更明智的判断。本章节将对城区电力负荷的变化特性,构成原理以及影响电力负荷变化特性的各项影响因素作深入分析。

3.1 城区电力负荷特点

本课题的研究对象为城区电力负荷。顾名思义,某一时刻的负荷即为该时刻下城区内不同功能类型用户的用电总和。根据 2016 年中国统计年鉴对用电量的统计,将用电负荷分为农、林、牧、渔、水利业,工业,建筑业,交通运输,餐饮住宿,其他行业以及生活消费,其中工业用电依据工艺流程不同进行了细分,而建筑业则包括办公和商业建筑^[44]。不同的用户类型具有不同的用电特性。例如办公建筑较为密集的用电负荷发生在工作日的 7 点至 18 点,商业建筑的用电能耗则主要集中在每天的 8 点至 21 点,同理第二产业用电则根据工艺和排班制的不同呈现出不同的用电特性且相较办公商业建筑变化较为平缓。同一时刻下城区负荷是所有用户不同用能特性的综合体现,因此相对于各类用户的用能特性,其占总负荷的比例则更为重要。

再将负荷自上而下拆分,负荷又可看作不同类型用户下各个分项能耗的累加,如照明负荷,设备负荷,空调负荷等。而影响各类分项能耗变化趋势的主要因素又不尽相同,例如空调负荷主要受天气因素影响,而照明负荷可受作息时间,日照时长及强度等影响。因此在分析各类分项能耗总和时,其变化特性是受多方面较为复杂的因素综合作用的结果。

本课题将以上海市 2014 年日 96 点负荷为例进行分析和预测,该数据来源于上海市经济和信息委员会数据平台。以下章节将从不同的观测周期对电力负荷变化特性做出分析。

3.1.1 典型日负荷特性

图 3.1 所示为上海市 2014 年全市夏季、过渡季及冬季典型日负荷曲线,从图中可看出电力负荷随时间序列呈现明显的峰谷波动,且各个季节的曲线特性差

别较大。以夏季典型日为例，定义小于等于日谷负荷的部分为基负荷，如图红色点划线所示。定义大于日谷负荷小于日间谷负荷的部分为腰负荷，大于腰负荷部分的为峰值负荷。夏季典型日负荷变化呈现出明显的两峰两谷的特点，其中日间高峰起止于日间典型的上下班时间，且至 15:00 左右用电负荷均呈现上升趋势，总体保持在较高水平。晚高峰出现在 19:00 至 21:00，相较于日间高峰差值较大。反观冬季典型日负荷，日负荷变化虽同样具有两峰两谷，但两个峰值的负荷差距较小，且日间谷负荷出现在午间 11:00 至 15:00，此外晚高峰的出现时段相对于夏季较早。过渡季的日负荷相较于冬夏两季明显偏低，且日间负荷变化更为平稳，总体呈现三峰三谷的特点，三个峰负荷的差距不明显，且晚间峰负荷的出现时段介于冬夏季的时间之间。由于上海地处夏热冬冷地区，气候的变化造成了不同季节间负荷特性变化的差异。

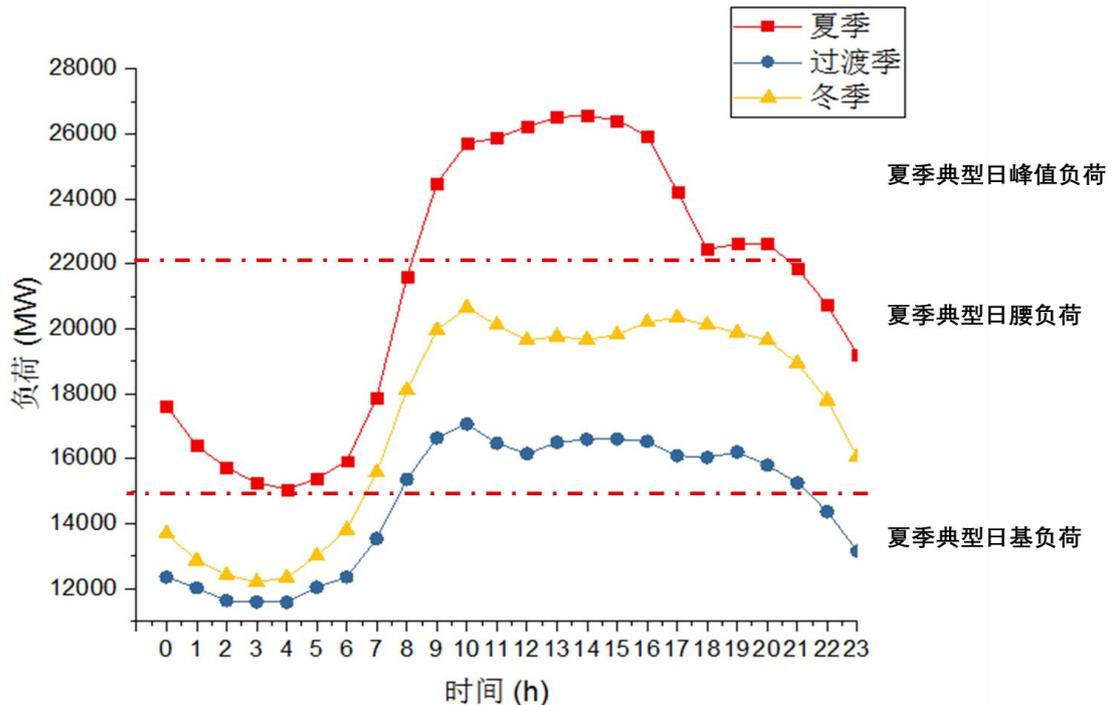


图 3.1 2014 年上海市典型日电力负荷变化曲线

3.1.2 典型周负荷特性

由图 3.2 所示的上海市典型周负荷变化曲线不难看出，电力负荷的变化在

任意连续两点间的变化是连续的，基本不存在突然地跃变，这是电网运行安全稳定运行的重要控制要素。此外，从负荷一周的变化特性来看，存在明显的周期性，日负荷的特性曲线相似。周六周日的负荷显著低于工作日负荷，这与社会生产生活作息相关。

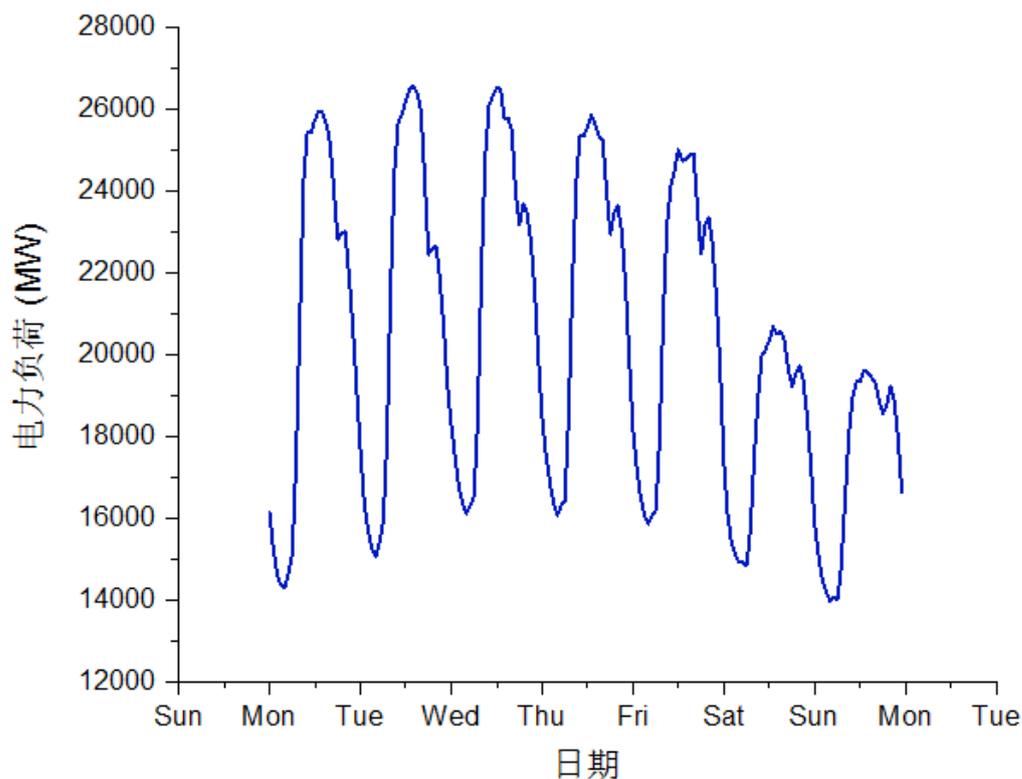


图 3.2 2014 年上海市典型周电力负荷变化曲线

3.1.3 典型月负荷特性

同理于负荷周特性，当遇到法定节假日时，电力负荷也会明显下降，且变化趋势与周末相似。当遇到较长的节假日，如春节、国庆，电力负荷总体上在该段周期内呈现出明显的线性下降再上升的趋势，且总体上最小日平均负荷出现在假期的中间一天。

日属性的不同直接影响电力负荷水平的大小，因此在统计月周期规律时将工作日和节假日分类，分别就其逐日最大负荷进行统计分析可得出电力负荷月变化特性，选择日最大负荷进行统计具有代表性。如图 3.3 所示为工作日日最大负

荷的时序图，日最大负荷随月份不同有着明显的波动。2014 年上海市用电高峰主要集中在 7、8 月份，为传统的空调制冷季节，其每日的负荷峰值远远高于其他月份，其中 8 月 5 日上海用电达到全年最大值，为 26,560 MW。其次日最高用电负荷较大的月份为 1、2、12 月份，该月份亦是供暖季，其中全年工作日用电高峰最低值也出现在二月份，这是由于春节长假后的第一周，还未完全恢复到正常的生产生活。制冷供暖月份内相邻几天的用电最大值随机波动较大，从暖通空调专业分析，是由气象因素的波动引起的。相对而言 4 至 6 月以及 11 月份的相邻日负荷波动则较为平缓。

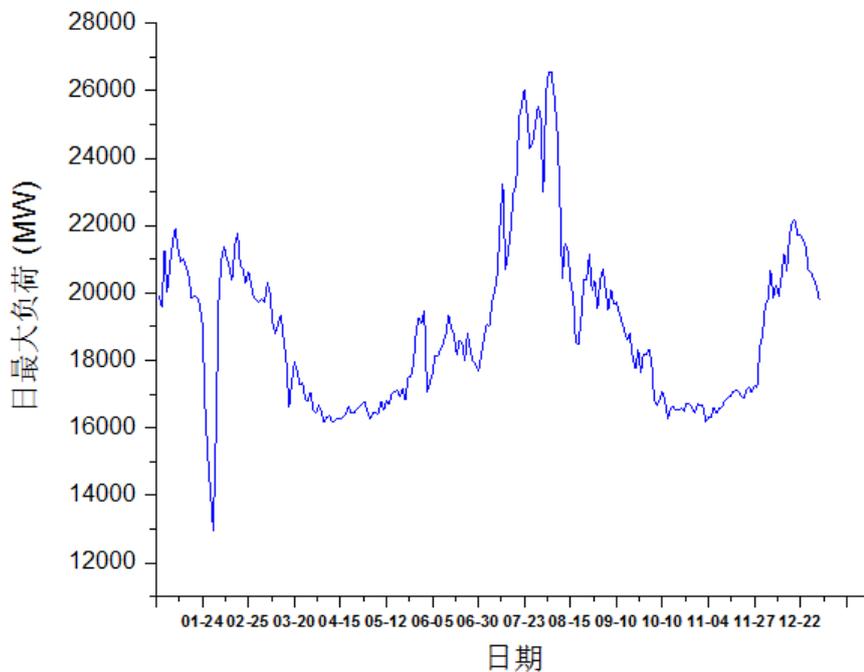


图 3.3 2014 上海工作日日最大负荷特性曲线

不同于单体建筑或同类型建筑群，城区电力负荷特性是各类功能建筑群特性综合累加的结果，可通过考察逐月每日峰值负荷时刻分布而知，如图 3.4 所示。经过对 2014 年上海电力负荷数据统计得出，逐日峰值负荷分布在 10 点至 20 点，且各月因气象参数不同导致用电行为不同，从而日峰值出现的时刻分布也不同，同时工作日和非工作日的分布也截然不同。工作日时段，除 7、8 月份外，各月用电峰值集中在 10 点至 11 点，11 月至 1 月部分用电尖峰时刻出现在 17:00 以后。相对而言，非工作日时段的峰值负荷主要集中在 17:00 以后，主要由于居民

生活用电负荷率在非工作时段比重较大。随着7、8月份制冷季的到来，尖峰负荷时段在此时也集中在了午间12:00至16:00气温较高的时候。

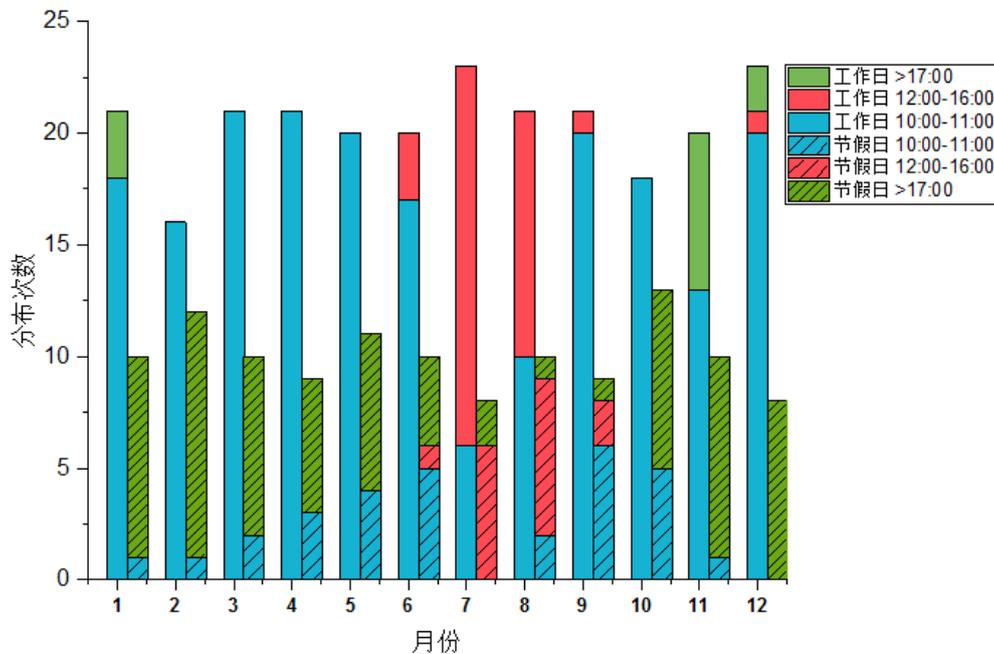


图 3.4 2014 上海市电力负荷逐月日峰值时刻分布

因此通过分析城区电力负荷月周期特性可知，气候条件的变化引起了各月负荷特性的不同，在制冷供暖季节日负荷明显增大，且存在明显波动。而在过渡季节负荷变化较为平稳。通过分析逐月日最大负荷的时刻分布，也能客观的说明气候对负荷变化的作用。在3.2章节着重剖析气候因素对负荷的影响关系。其次通过对比工作日和非工作日日峰值负荷时刻分布的不同，也说明了因社会生活作息不同造成的工作日和非工作日负荷特性的差异。

3.2 电力负荷影响因素

城区短时电力负荷具有波动较大，受多种因素影响叠加，呈现较为复杂的非线性特点。因此找出影响短时电力负荷变化的因素，往往是建立预测模型的首要步骤，也是避免预测模型选择盲目性的先决条件。通过上述分析可知，影响电力负荷变化特性的因素有如下几点：

- i. 作息时间因素：在电力负荷日变化（如图 3.1 所示）和周变化（如图 3.2

所示)周期中,随着作息时间的不同,每日负荷呈现日夜峰谷的周期性波动,且节假日的负荷明显小于工作日的用电负荷。

- ii. 气候因素:气候变化对电力负荷有着显著的影响,其影响关系呈现出较为复杂的非线性特性,至今无法通过数学模型准确的表达气候变化与短时电力负荷波动的关系。长期来看,季节的更替改变人们的用能习惯,导致电力负荷存在明显的季节性差异(如图 3.1 所示)。短期来看,每日气象因素的变化同样导致负荷的波动,该波动又是多种气象因素叠加的综合作用。根据 Feinberg, E.A 温度被公认为是影响电力负荷变化的最主要的气象因素,甚至在 20 世纪 30 年代就有学者提出了该概念^[18]。在目前的负荷预测研究中,主要将气候的影响归结为气温的影响,忽略了风速、湿度等参数的作用。
- iii. 需求响应调节:在高效节能的理念下,通过改变需求侧的用电行为以达到平缓电力负荷峰谷差,从而减少供电侧备用容量缓解供电压力的目的。随着电力市场的逐步完善,需求响应调节机制也逐步完善,通过电价杠杆和配合响应的补贴价格,越来越多的工业商业用户愿意参与到调节市场。因此,需求响应指令及电力价格也成为影响负荷特性的重要因素之一。由于本课题旨在研究的日前负荷预测作为需求响应调节的基线,因而该因素在本文中不做考虑。
- iv. 随机因素:对电力负荷的影响不具有周期性规律,如举办大型活动,大型用电设备故障等因素均可剧烈地改变电力负荷特性。

其他诸如经济增长,用能产业结构的调节对用电负荷的影响体现在长期变化中,针对本课题日前电力负荷预测认为这些因素是稳定不变的。

3.2.1 气候因素对短时电力负荷的影响

从暖通专业背景知识分析,气候变化引起的电力负荷波动是源于人体对于环境舒适度的感觉有所变化,从而影响了人们的用电行为。为了更准确地挖掘气候和电力负荷之间的关系,将从人体舒适度概念出发,通过引入表征热舒适性的综合气象参数,更全面地考察不同气象参数的耦合作用与电力负荷的关系。同时,考虑天气对负荷的累积效应,例如多日连续的高温天气也将改变城区的负荷特性。

相较于前人在此方面的研究,本课题在剖析气候因素对短时电力负荷的影响作用时,进行了以下四个方面的改进和创新:

- i. 采用逐时数据进行相关性分析：采用实时的气象数据和电力负荷数据，考察两者间逐时的相关关系。相较于前人只用历史日特征气象参数，如日平均温度，日最高温度等，与日平均电力负荷的相关关系，采用逐时数据的相关系数更能准确的反映一天内逐时天气对负荷的影响；
- ii. 考虑建筑热惰性：根据单体建筑能耗特性，由于围护结构存在热惰性，冷热负荷的出现相较于室外天气因素的变化具有延迟性。在分析待预测时刻的电力负荷和天气因素的关系时，天气参数的取值采用待预测时刻值和前三个小时的平均值，以更准确地反映延迟效应。
- iii. 空气焓值作为直接影响空调负荷的重要因素，将探讨其与电力负荷的相关性。
- iv. 借鉴人体热舒适性指标，引入综合气象参数，并同样考虑延迟效应，考察其与电力负荷的相关关系。该方法更能反映出不同气象因素之间对负荷影响的耦合作用。

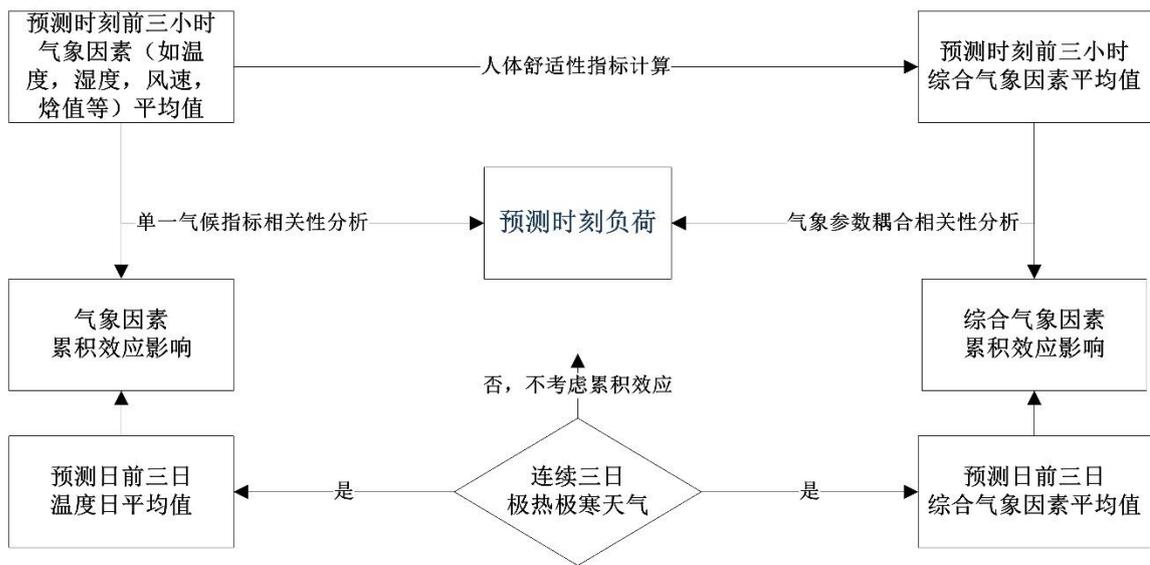


图 3.5 气候因素对短时电力负荷的影响分析技术路线

气候因素对短时电力负荷的影响分析思路如图 3.5 所示：

- i. 逐时单一气象因素与待预测时刻负荷的相关关系；
- ii. 逐时综合气象因素与待预测时刻负荷的相关关系；
- iii. 当连续三日出现极热或极寒天气，将考虑气象因素的累积效应。且参数考察范围由逐时扩展到日特征值，即考察前三日平均最高气象参数和预测日最高电力负荷的关系。

3.2.2 随机变量常用相关性分析方法

常用的分析随机变量间相关程度的方法有三种：散点图法，灰色关联度法以及 Pearson 相关系数法。

散点图法：通过两组数据在坐标系中的散点分布，判断两变量之间是否存在关联关系。该方法较为直观但无法定量说明。

灰色关联度法：通过比较随机变量序列与特征序列曲线的几何相似程度判断其关联关系是否紧密，曲线走势越相似，则相关程度越大。设特征序列为 x ，待比较的随机序列为 y ，则两序列的灰色关联度 r 可通过以下公式进行定量计算：

$$r(x, y) = \frac{1}{n} \sum_{k=1}^n \frac{\min_i \min_k |x(k) - y(k)| + \rho \max_i \max_k |x(k) - y(k)|}{|x(k) - y(k)| + \rho \max_i \max_k |x(k) - y(k)|} \quad 3-1$$

其中 ρ 为分辨系数， $\rho \in [0, 1]$ ； $\min_i \min_k |x(k) - y(k)|$ 为两序列的两级最小差值，同理 $\max_i \max_k |x(k) - y(k)|$ 为两序列两级最大差值。通过求得两序列各点关联系数的算术平均值，得到特征序列和待比较序列的灰色相关度。

Pearson 相关系数法：通过 Pearson 相关系数的计算，判断两个随机变量间的线性相关程度和相关的方向。相关系数介于 -1 至 1 之间，绝对值越接近 1 代表两组变量越线性相关。但其值为零不代表不存在相关关系，只能说明不存在线性关系。两随机序列相关系数 r 计算公式如下：

$$r = \frac{\sigma_{xy}^2}{\sqrt{\sigma_x^2 \sigma_y^2}} \quad 3-2$$

其中 $\sigma_{xy}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$ 为两个随机变量的协方差， $\sigma_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ 为随机变量 x 的方差，同理 $\sigma_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$ 为随机变量 y 的方差。

本课题在考察气象参数与电力负荷相关关系时采用 Pearson 相关系数法和灰色关联度进行讨论，在两者线性关系较小的情况下，利用灰色关联度佐证两者间的关联度大小。

3.2.3 单一气象因素与短时电力负荷相关性分析

2014 年气象参数来源于腾天数据平台和 Weather Underground 网站，所包

含的气象参数有干球温度，露点温度，相对湿度，含湿量，风速和空气焓值，数据采集频率为每30分钟数据。为了与电力负荷逐15分钟数据对应，需要对天气参数缺失值进行填补，所采用的方法为三次样条插值法。

根据第3章节的介绍，电力负荷呈现明显的周期性和季节性。因此选择制冷季7月，供暖季1月，过渡季4月为典型月份，分别计算以月为周期和以日为周期的电力负荷及各个天气参数序列的相关系数及灰色关联度。在以月为周期的统计中，将以日峰值电力负荷及各个天气因素日平均值作为统计序列，宏观上考察两者的相关性。在以日为周期的统计中，采用图3.5的技术路线，并将各个典型月份逐日相关性的中位数作为该月份典型日具有代表性的特征值。

为方便比较电力负荷与天气参数的关系，首先对数据进行归一化处理。计算公式如下：

$$x_i^* = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad x_i^* \in [0,1] \quad 3-3$$

表格3.1所示为典型月份日峰值电力负荷与单一气象因素相关性分析。由于灰色关联度的范围为0至1，无法反应两序列相关性的方向。但相关系数法和灰色关联度法所得结果的趋势较为吻合。

表格 3.1 典型月份日峰值电力负荷与单一气象因素日均值相关性表

		气温	风速	相对湿度	空气焓值
相关系数	制冷季七月	0.795	-0.083	0.718	0.769
	供暖季一月	-0.625	0.433	-0.668	-0.678
	过渡季四月	0.561	-0.174	0.255	0.366
灰色关联度	制冷季七月	0.703	0.588	0.729	0.716
	供暖季一月	0.683	0.734	0.652	0.666
	过渡季四月	0.594	0.648	0.532	0.540

整体上看，气温和电力负荷的相关性最大，且七月制冷季最高，其次为供热季，其中供暖季呈现明显的负相关性，即气温越低电力负荷越大。在相关系数法的计算中，相对于其他季节，风速在供暖季能够较大程度的影响电力负荷变化，这也和在冷环境中吹风感加重人体冷感觉的客观认知相符，此处数据可以说明风速越大，电力负荷所需越大。此外，相对湿度在制冷季和供暖季对电力负荷变化

也有着较大影响，制冷季呈正相关，供暖季呈负相关，在过渡季中相关性不大。灰色关联度方法计算结果中，尤其在制冷季和过渡季中与相关系数法的结果对比，风速对电力负荷的关联度明显增大。总体来说，空气焓值与电力负荷的相关性与气温相似，但过渡季节中相较气温其相关性更低。以一月份为例，根据日峰值电力负荷与气象参数日均值变化趋势说明两序列变化的相似度，如图 3.6 所示。

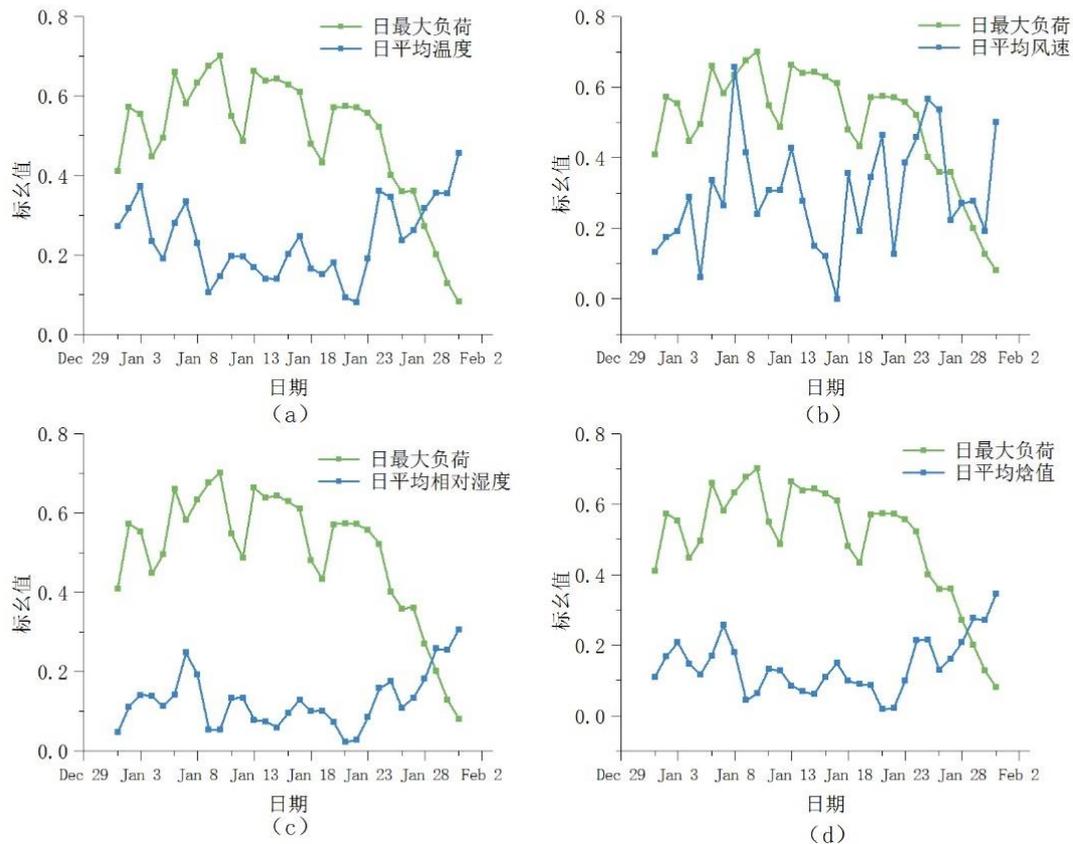


图 3.6 一月份日峰值电力负荷与气象参数日均值变化趋势

结合表格 3.1 计算结果，日最大负荷与日平均气温，日平均相对湿度以及日平均焓值较为相关，但呈现出明显的相反的变化趋势。即这些气象参数增大时，电力负荷反而减小。从图中可明显观察到，日平均气温、日平均相对湿度和日平均焓值的变化特性基本相同。三者之间存在较高的关联度。根据干空气状态方程，在一定大气压下，已知其中两个参数，可求得第三个状态。因此在后续预测模型输入参数的选择上，只选择其中的两个变量输入。日平均风速变化波动性较大且变化频率更大，灰色相关度计算显示其于负荷存在高相关性。

同理以七月为例，根据日峰值电力负荷与气象参数日均值变化时序图说明两序列变化的相似度，如图 3.7 所示。结合表格 3.1 计算结果，日最大负荷与四项气象因素均存在较大的关联度，最小的为日平均风速。从右上角的图中也可看出，逐日风速变化随机性较大，与负荷曲线相似度相对较低。其余三个因素不仅在曲线走势相似度上，在峰谷值与电力负荷峰谷值出现时刻拟合程度均较高。

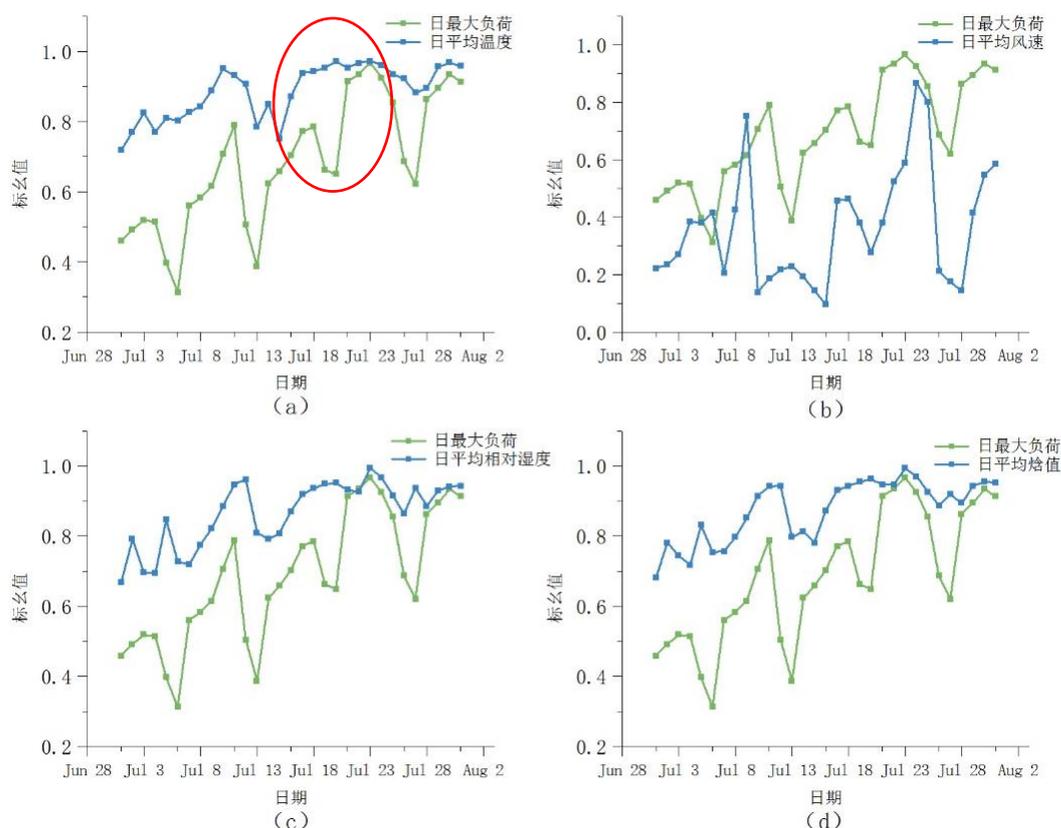


图 3.7 七月份日峰值电力负荷与气象参数日均值变化趋势

天气对短时电力负荷的累积效应表现在两个方面。一是天气对电力负荷作用的延时性上，这主要由于建筑等结构的热惰性所引起的，上文已通过采用气象因素的平均值进行补偿。另一方面，表现在临近几日连续极热或极寒的天气的出现，将造成后一日电力负荷特性的改变。值得注意的是日均气温和日最大负荷的图中（如图 3.7(a)红色圆圈内），7月16日至24日气温持续保持在了较高的水平，日平均温度达到了 30°C ，除去7月19和20日为双休日负荷水平下降，总体负荷呈现出了持续快速的增高趋势。该现象说明，由于连续几日气温，相对湿度和

焓值均维持在较高的水平，极热天气的出现带来了电力负荷的快速增长。根据相关性计算结果可知，制冷季电力负荷与温度相关性接近 0.8，累积效应可简化为通过对温度的修正表达。若连续三日出现极热天气，当日平均温度连续高于 30℃ 时，则当日温度的修正值等于当日实际温度乘以前一日平均电力负荷与再前一日平均电力负荷的比值。

接下来分析各个典型月内，以天为周期的电力负荷与单一气象因素逐时序列相关性，采用统计值中位数作为各个月份典型日的特征值。其中气象因素采用的是当前时刻和前三小时的平均值，以消除天气因素作用在电力负荷的延迟性，结果如表格 3.2。

和以月周期分析的结果类似，逐时电力负荷与气温序列按照过渡季，供暖季和制冷季的划分其相关性依次增大，不同的是在日周期内，供暖季气温与电力负荷间仍呈现正相关性，空气焓值统计结果存在类似的特性。风速在日周期内与电力负荷的相关性相较于上表明显增大，在供暖季中关联度最大，供冷季节风速与电力负荷存在负相关关系。相反，相对湿度对逐时电力负荷的影响程度减小。

表格 3.2 典型月份逐日电力负荷与单一气象因素相关性分析统计表

		气温	风速	相对湿度	空气焓值
相关系数	制冷季七月	0.835	-0.362	-0.402	0.778
	供暖季一月	0.711	0.566	-0.005	0.712
	过渡季四月	0.572	0.362	-0.389	0.696
灰色关联度	制冷季七月	0.614	0.645	0.622	0.622
	供暖季一月	0.605	0.683	0.514	0.521
	过渡季四月	0.554	0.490	0.310	0.613

相比于相关系数法计算的结果，采用灰色关联度法，各个气象因素在供热制冷季的关联度都较大，过渡季较小。以供暖季一月十日逐时电力负荷与气象参数变化趋势为例，更直观的考察各个序列灰色关联度，如下图所示。

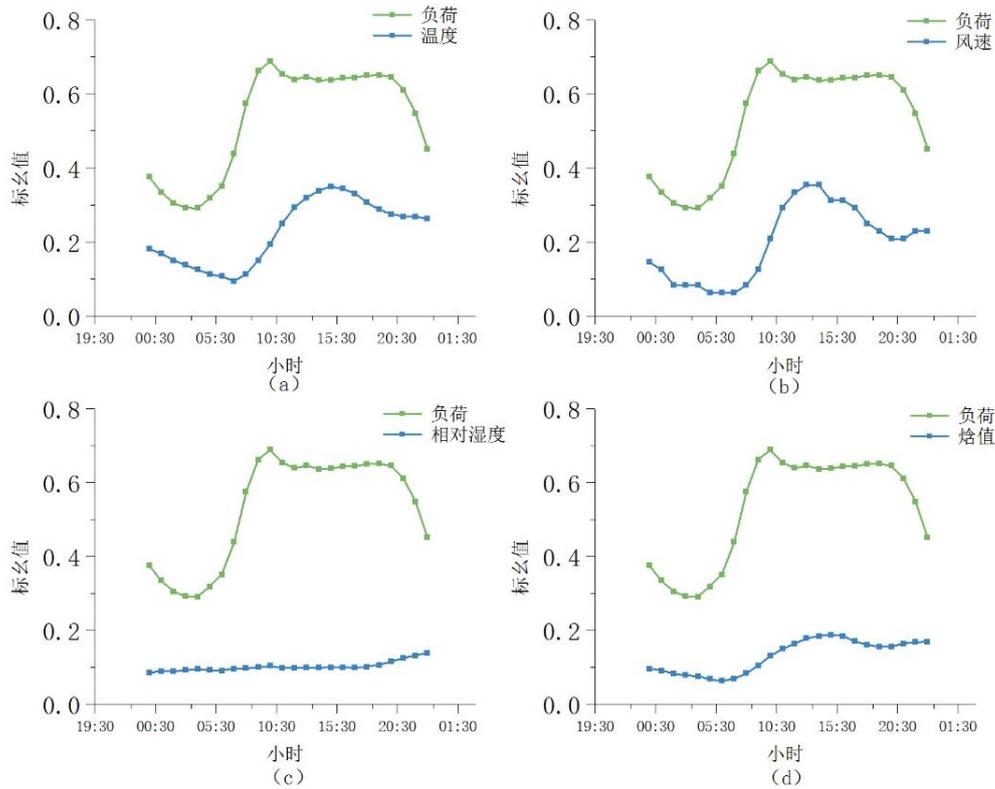


图 3.8 一月十日逐 15 分钟电力负荷与气象参数变化趋势

逐时气温与电力负荷首先表现出相同的下降和上升趋势，在 10:30 至 17:30 期间逐时气温和电力负荷变化趋势相反，至次日凌晨，两者变化趋势再次趋同。结合表格 3.2 的计算结果，逐时电力负荷与气温关联度较高。逐时风速的变化趋势类似于电力负荷，但风速峰值峰谷的出现相较于电力负荷存在滞后。逐时相对湿度和空气焓值变化平缓，同电力负荷的灰色关联度相对较低。

同理以制冷季七月十日逐时电力负荷与气象参数变化趋势为例，更直观的考察各个序列灰色关联度，如图 3.9 所示。制冷季中电力负荷变化与气温和焓值变化呈现出同增同减的特性。当日各项气象因素在日间变化较为缓和，而日间电力负荷较为平稳的维持在了一个较高的水平。

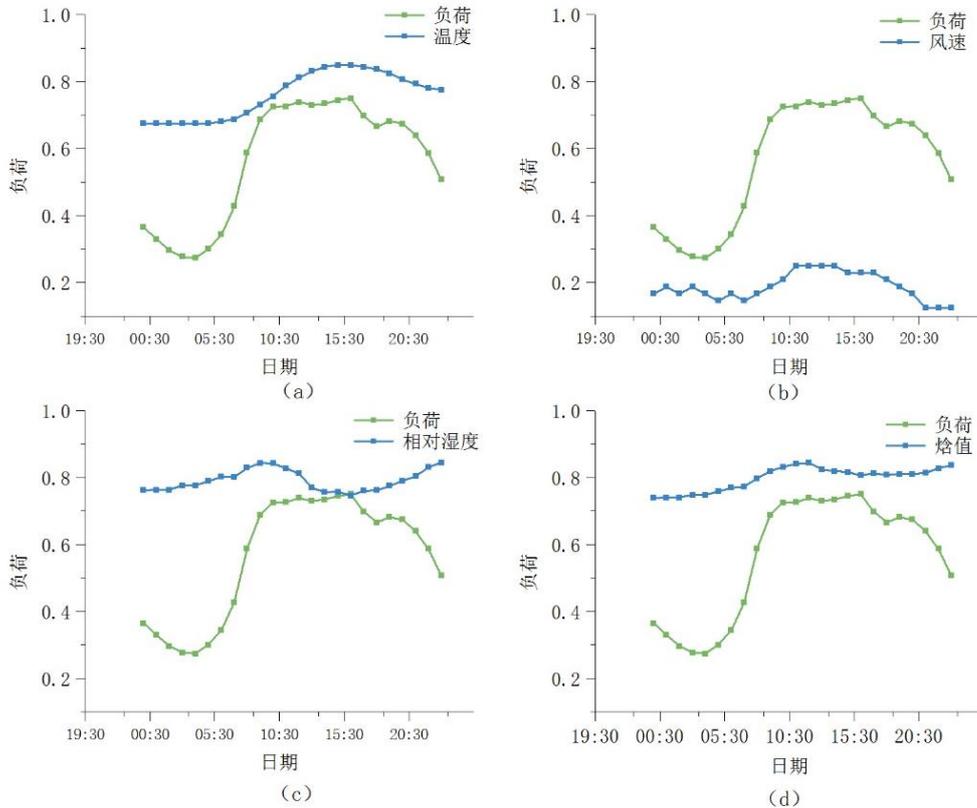


图 3.9 七月十日逐 15 分钟电力负荷与气象参数变化趋势

3.2.4 综合气象参数与短时电力负荷相关性分析

通过上述分析可知，不同气象参数在各个月份对电力负荷相关性是不同的，但分析仅局限于各个参数的单独作用。实际上，天气对电力负荷的影响是多个参数较为复杂的耦合性作用的结果。天气对电力负荷的影响是以人用电行为的改变作为媒介的，其中人体舒适度的变化起着决定性的作用。因此采用能够反映人体热舒适的综合气象指标，即以温度为输出单位，综合多个气象因素的热感觉指标，考察其与电力负荷相关性分析的随机序列是可行的。

通过单个天气因素的计算发现，制冷季温度和相对湿度同负荷相关性更大，供热季温度和风速同负荷相关性更大，过渡季温度同负荷相关性更大，风速和相对湿度影响力校对较弱，因此选取下列三个较为常用的热舒适性指标：

有效温度 (Effective Temperature)：表示在不同温度、湿度和风速下的当量温度值，冷热环境均可试用，计算公式如下^[45, 46]：

$$ET = 37 - \frac{37-T}{0.68-0.14RH+\frac{1}{1.76+1.4V^{0.75}}} - 0.29T(1-RH) \quad 3-4$$

其中，T 为干球温度（℃），V 为风速（m/s），RH 为相对湿度（%）

温湿指标 (Humidex): 表示在不同干球温度和湿球温度下的当量温度值，着重研究热环境下湿度对人体热舒适性的影响，计算公式如下^[45, 46]：

$$HI = T + 0.5555 \left[6.11 \times e^{5417.7530 \left(\frac{1}{273.16} - \frac{1}{T_{dew}} \right)} - 10 \right] \quad 3-5$$

其中，T 为干球温度（℃）， T_{dew} 为露点温度（K）。

风冷却指标 (Windchill Index): 表示在不同温度、风速下的当量温度值，着重研究冷环境下风速对人体热舒适性的影响，计算公式如下^[45, 46]：

$$WCI = 13.12 + 0.6215T - 11.37V^{0.16} + 0.3965TV^{0.16} \quad 3-6$$

其中，T 为干球温度（℃），V 为风速（km/h）。

对综合气象参数相关性的分析仍按照以月为周期和以日为周期的方法进行，其中风冷却指数只用于供暖季和过渡季，温湿指标只用于制冷季和过渡季。日峰值电力负荷及各个综合气象参数日平均值相关性计算结果如下：

表格 3.3：典型月份日峰值电力负荷与综合气象参数日均值相关性表

		ET	HI	WCI
相关系数	制冷季七月	0.815	0.823	
	供暖季一月	-0.721		-0.775
	过渡季四月	0.523	0.511	0.543
灰色关联度	制冷季七月	0.759	0.766	
	供暖季一月	0.732		0.682
	过渡季四月	0.551	0.464	0.492

供暖季有效温度和风冷却温度与负荷的相关系数相较于温度与负荷相关性明显增大，制冷季和过渡季改善效果不大。灰色关联度的分析则采用对比制冷季和供热季负荷与温度及各综合气象参数标幺值比值的方法，通过曲线的平稳程度判断其与负荷曲线走势的相似程度，如图 3.10 所示。

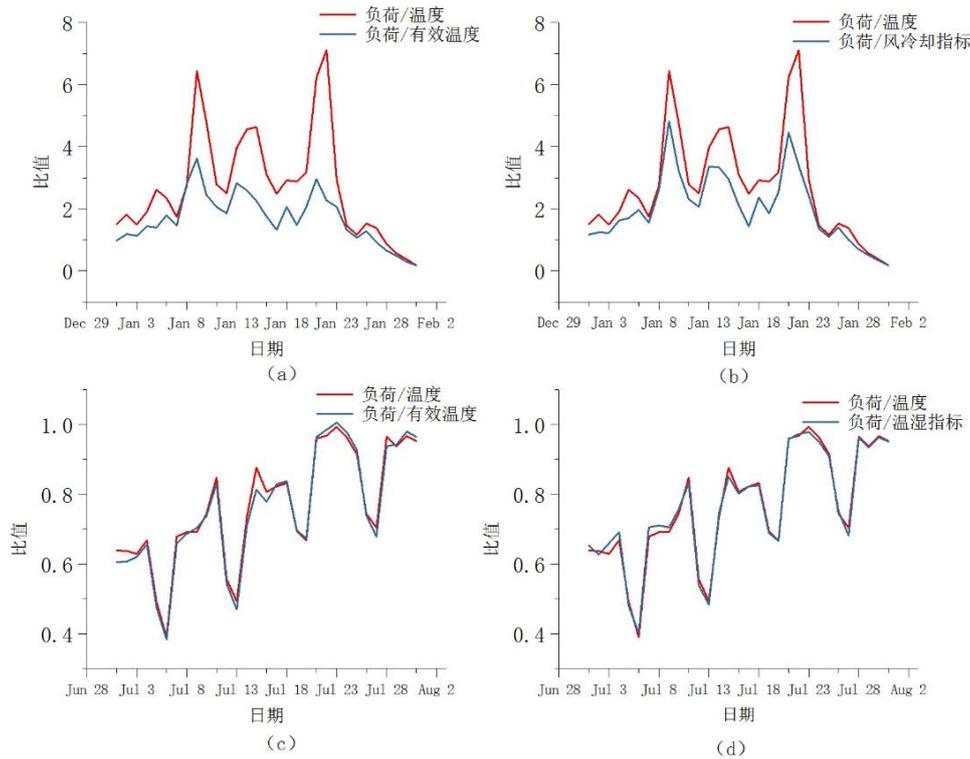


图 3.10 供暖季与制冷季日峰值负荷与日平均温度及各个综合气象参数比值对比图

供热季中（图 3.10 (a) (b)），电力负荷与有效温度的标么值比值变化最为平缓，说明两者变化趋势最为相似，相较于只考虑单一温度因素效果更好。制冷季中（图 3.10 (c) (d)），三条曲线走势基本吻合，说明采用单一温度尚且能够较好的反映出天气与电力负荷的相关性，因此采用综合气象参数改善效果不大。

以日为周期的电力负荷与综合气象因素逐时序列相关性，采用统计值中位数作为各个典型月的特征值。同样，综合气象因素采用的是当前时刻和前三小时的平均值，以消除天气因素作用在电力负荷的延迟性，结果如

表格 3.4 所示。相比于单一温度与电力负荷相关性的计算结果，供暖季采用有效温度或风冷却温度与电力负荷的相关性显著提高，制冷季和过渡季中改善效果不大。该结果与以月为周期的计算结果一致。

表格 3.4 典型月份逐日电力负荷与综合气象相关性分析统计表

		ET	HI	WCI
相关系数	制冷季七月	0.822	0.819	
	供暖季一月	0.748		0.755
	过渡季四月	0.523	0.511	0.543
灰色关联度	制冷季七月	0.821	0.842	
	供暖季一月	0.703		0.668
	过渡季四月	0.512	0.532	0.527

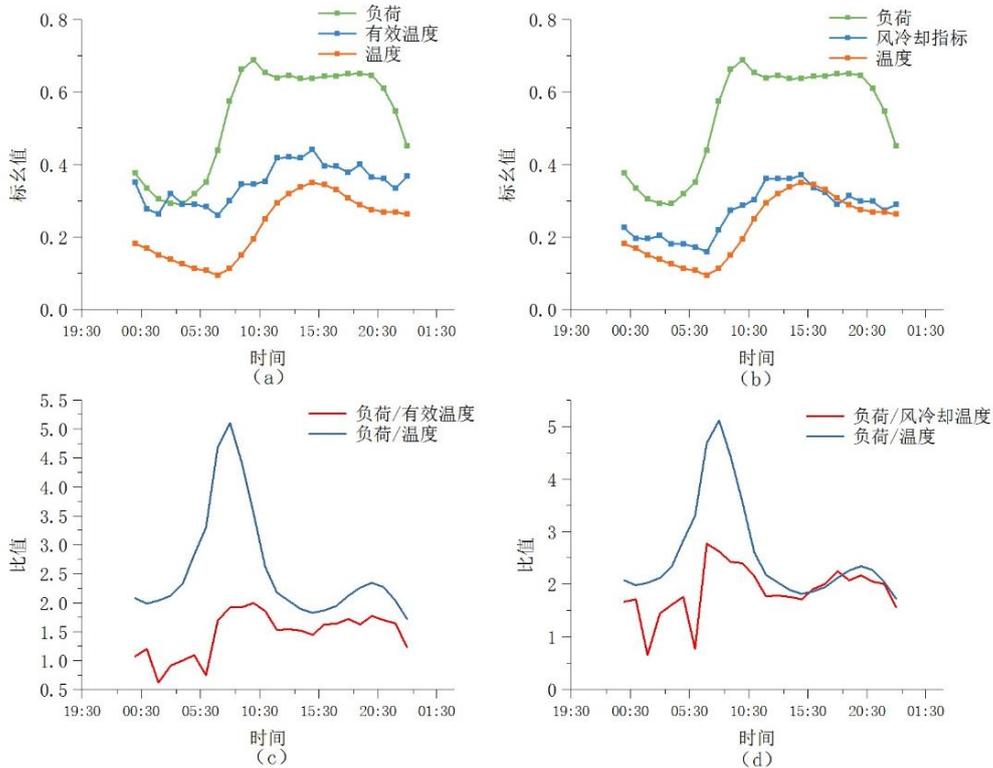


图 3.11 一月十日逐 15 分钟电力负荷与综合气象参数变化趋势以及逐 15 分钟电力负荷与温度及各个综合气象参数比值对比图

以一月十日逐时电力负荷与综合气象参数变化趋势（如图 3.11 (a) (b)), 以及电力负荷与温度及各个综合气象参数比值对比图为例(如图 3.11(c)(d)), 能更直观的考察各个序列灰色关联度。有效温度和风冷却温度午间峰值出现的时

间相较于干球温度较早,该时间与电力负荷的午间电力负荷谷值出现时刻更为吻合,说明综合气象参数对电力负荷表现出的延迟效果更小。电力负荷与有效温度标么值的比值曲线波动最为缓和,因此判断其与负荷曲线走势的最为相似,关联度最大。

同理直观考察供冷季综合气象参数与电力负荷的关联度,以七月十日为例,如图 3.12 所示。有效温度的变化比温度变化平缓,热湿指标与温度的变化趋势基本一致。但两者相对于温度的变化都存在波峰和波谷上的相对平移。温度的峰值与负荷峰值出现时刻更为吻合。反观各个参数与负荷比值的变化特性,综合气象参数的采用并未明显平滑比值曲线。综上两点,制冷季中通过综合气象参数反映多气象因素对负荷耦合作用的方法效果不佳。多气象因素间的协同作用仍不清晰,因此在后续模型输入时仍需将各单一气象因素同时输入。

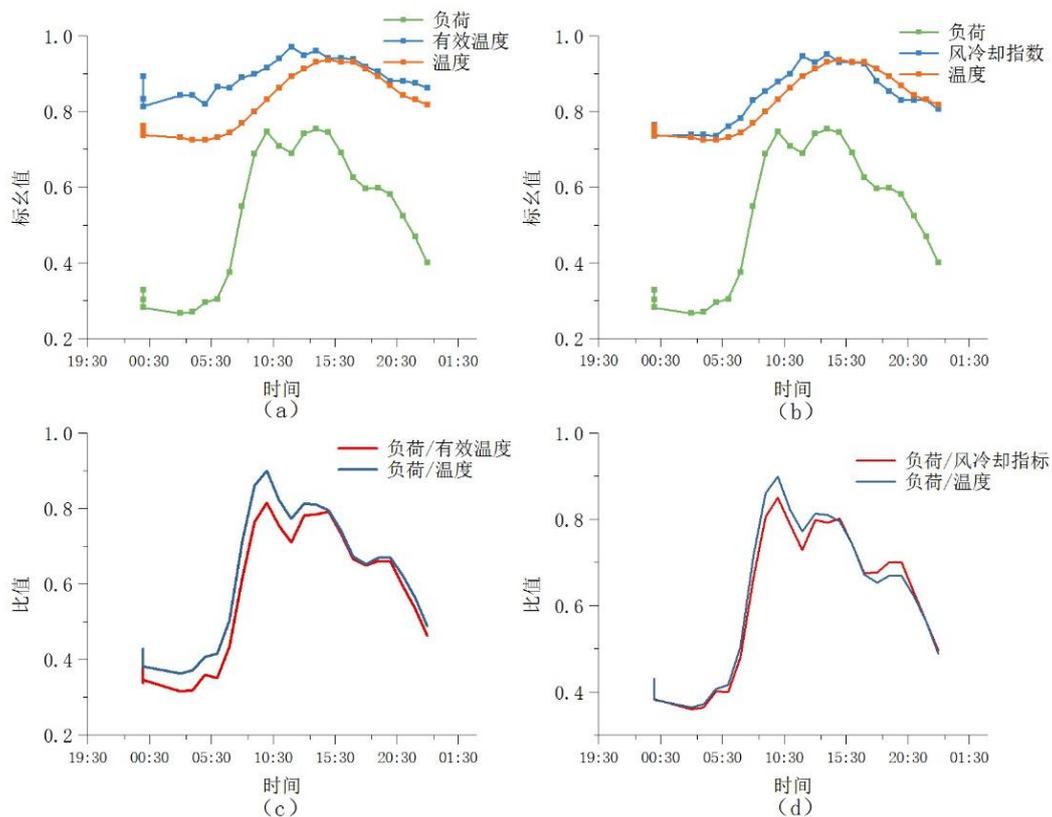


图 3.12 七月十日逐 15 分钟电力负荷与综合气象参数变化趋势以及逐 15min 电力负荷与温度及各个综合气象参数比值对比图

3.3 本章小结

事实上,气候因素对电力负荷有着密切且复杂的影响,目前尚未研究出数学公式表征其内在复杂的非线性关系,只可通过人工智能算法进行拟合。上述章节将两者具有非线性关系线简化成线性关系计算相关系数,并通过灰色关联度辅助校验,可客观的反映出电力负荷与哪些气象参数具有较高的相关性,为后续预测模型输入参数选择提供有力的依据。

同电力负荷相关性分析从单一气象参数和多个气象综合指标入手,分别在月周期和日周期两个观测时段上考察,可得以下结论:

- i. 不同季节周期内,影响电力负荷变化的主要气象因素不同。制冷季负荷主要与气温和相对湿度相关性较大,供暖季气温和风速为影响电力负荷的主导因素,过渡季电力负荷总体上和气候因素相关性不大。
- ii. 考察多个气象参数的综合作用,以人体热舒适指标为媒介,采用综合气象参数考察其与电力负荷的相关性。供暖季采用考虑风速和湿度的有效温度,其相关性相较于单一温度显著提高,说明在建立模型时可用有效温度作为输入参数,代替温度和风速及湿度的耦合作用。制冷季和过渡季综合气象指标与电力负荷的相关性同单一温度对负荷的关联度基本一致,相关性没有显著提高。说明综合气象参数不能准确表征气温,相对湿度及风速对负荷的耦合作用,建立模型时需将温度、相对湿度和风速同时作为输入参数输入。
- iii. 日周期变化内,温度仍是和电力负荷变化相关性最大的天气因素,但对负荷的影响呈现出延迟性。日周期内综合气象参数的应用结论与结论 ii 相同,供暖季采用有效温度相关性较好。

第4章 电力负荷拆分

通过上一章分析可知，电力负荷变化具有周期性和波动性的特点。其中日周期的波动主要与生活作息和天气变化有关，月周期的波动主要与日属性和天气因素有关。进而通过分析不同气象因素与负荷序列的相关性发现，在不同季节内，不同气象因素的影响权重并不相同，其中过渡季中负荷与气象参数的相关性最低，制冷季夏季和供暖季冬季则表现出较高的关联度。因此可认为过渡季节电力负荷对天气变化不敏感，负荷波动的主要原因是作息时间，本课题中称该部分负荷为基础负荷。而制冷季和供暖季的电力负荷则由基础负荷和天气敏感型负荷组成。本章节主要通过各个季节划分的探讨，将非过渡季电力负荷拆分为基础负荷和天气敏感型负荷两部分。

4.1 季节划分定义

上一章节的探讨中已经引入了供热季，过渡季和制冷季的概念，这是在暖通空调领域十分常见的季节划分概念，依据建筑是否需要冷热负荷定义一年中的不同时段。其中既不需要制冷也不需要制热的时段称为过渡季。

度日数法虽然是常用来确定建筑冷热负荷需求的，但是其确定的室外日平均温度基准可判断出建筑是否需要供冷和制热，因此也常用作季节划分的依据。根据《夏热冬冷地区居住建筑节能设计标准》上海地区采暖度日数的室外基准温度为 18°C ，即当室外日均温度低于 18°C 时，建筑需要供暖。上海地区制冷度日数的室外基准气温为 26°C ，即当室外日均温度高于 26°C 时，建筑需要制冷。

针对本课题采用日均值气温作为季节划分的依据并不合理，这主要由于夜间气温大幅下降，这会大大降低日平均气温值。此外冷热负荷需求主要出现在日间，即温度相对较高的时段，因此过低的日平均值无法准确划分当日是否需要供热或制冷，无法准确定义电力负荷的组成。因此本小节在度日数法的基础上，通过决策树分类法对气象参数进行信息分类，确定一套直观的规则对供热季，过渡季和制冷季进行划分。

4.1.1 决策树算法

决策树是机器学习中常用的监督式预测模型。它利用树的结构对数据进行分

类，树的一个节点就代表满足某个属性的数据子集，每个节点又可以因为不同决策向下引出不同分支，直到生成只具有同一属性的数据集，称为叶子节点。由根节点自上而下的不断重复的建立下节点和分枝，从而生成一棵决策树。与其他算法相比，决策树具有显著的优点，最重要的一点是使用白箱模型，其输出的结果通过决策数的结构易于理解。相反决策树亦可通过无限多的分支达到对训练数据的完美拟合，但由于规则过于复杂，导致无法很好的应用于训练集之外的数据，容易出现过拟合，泛化性降低。因此，理想的决策树分为三种：叶子节点少；叶子节点深度小；以及叶子节点少且叶节点深度小的结构。^[47]

决策树常用的分类算法有 ID3 算法，可处理连续属性的 C4.5 算法，CART 算法及 Assistant 算法等。下面就最为核心的 ID3 算法进行简单介绍。

ID3 算法是以信息熵和信息增益为衡量标准，在各级节点选择属性时，检测所有属性，选择信息增益最大的属性产生决策树节点，由该属性的不同取值建立分支，再对各分支的子集递归考察信息增益建立下层节点和分支，直到所有子集仅具备同一类别的数据^[48]。利用得到的决策树，对预测样本进行归类。

信息熵定义为分布概率为 $P = (p_1, p_2, \dots, p_n)$ 的变量，其所传递的信息量，计算公式如下：

$$I(P) = -\sum_{i=1}^n p_i \log_2 p_i \quad 4-1$$

若特征变量集合 A 已可按某规则被分成互相独立的类 $B_1, B_2 \dots B_k$ 类，则判断 A 中一个元素所属类别所需的信息熵为 $\text{Info}(A) = I(P)$ ，其中 P 为类别的概率分布。

若先根据非类别属性 X 的值将 A 分成集合 $A_1, A_2 \dots A_n$ 类，判定 A 中的一个元素类别的信息量通过 A_i 的加权平均值求得，称为条件熵，计算公式如下：

$$\text{Info}(X, A) = \sum_{i=1}^n [(|A_i|/|A|)\text{Info}(A_i)] \quad 4-2$$

信息增益是两个信息量的差值，其中一个信息量是确定 A 中一个元素的信息量，另一个信息量是在已得到属性 X 的值后确定 A 中一个元素的信息量，信息增益计算公式为：

$$\text{IG}(X, A) = \text{Info}(A) - \text{Info}(X, A) \quad 4-3$$

通过对所有属性信息增益比较，将最大增益值对应的属性作为分类的依据，构造出一级决策树。以此类推直到达到分类目的。

4.1.2 供热制冷季节判断规则

为准确判断预测当日是否需要冷热负荷,采用逐时气象参数代替度日数的日均值气温,对该时刻的供热制冷属性进行分类。再通过统计每日不同时段冷热需求的众数,判断该时段的供热制冷属性。最后对各时段统计值进行加权平均,判断当日的季节属性。分析思路如下图所示:

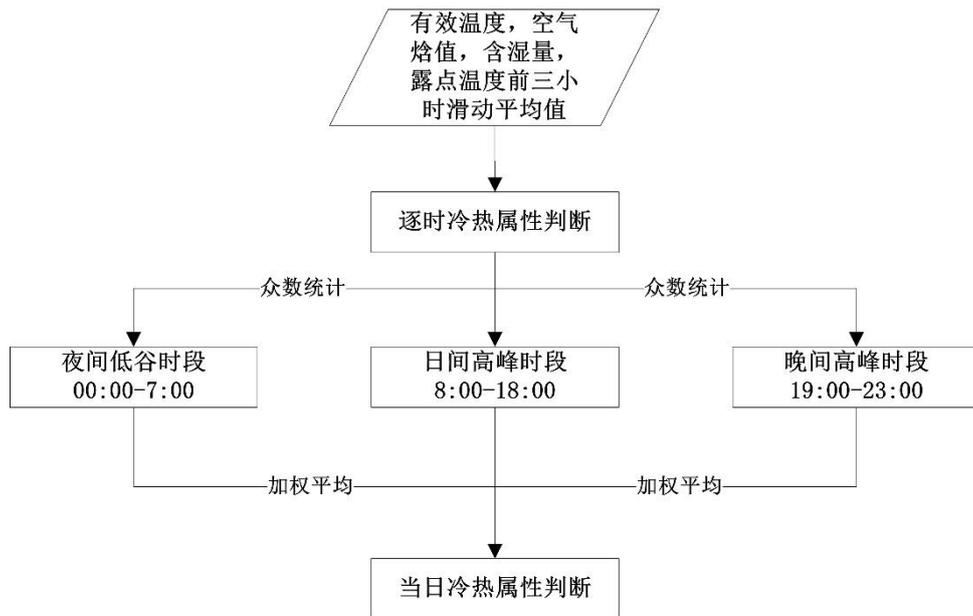


图 4.1 供热制冷季节判断分析思路

由于决策树每一节点只考虑单个属性的最大信息熵,因此决策树输入参数用有效温度代替干球温度,突出在同一节点上风速和相对湿度对人体实感温度的耦合作用。其余输入参数为空气焓值,含湿量和露点温度。同时考虑气象因素对电力负荷的延迟性,仍采用四个时刻的滑动平均值作为当前时刻取值。季节划分为离散属性,设供热季为1,制冷季为-1,过渡季为0。

为降低决策树的过拟合现象,在模型训练时采用 n 折交叉验证。将训练样本随机分割成 n 个子集,其中 $n-1$ 个子集用来训练,另外一个单独的样本用作验证模型的数据,交叉重复 n 次,综合 n 次结果优化模型。本章节决策树的训练采用 5 折交叉验证。

基于建筑能耗的背景知识,上海确定的采暖季为一月,确定的制冷季为七月,过渡季会出现在四、十月份。因此将一月数据季节属性赋值为1,七月数据季节属性赋值为-1,根据人体舒适度指标,当有效温度 ET 落在 $15-23^{\circ}\text{C}$ 时人体感觉

较为舒适^[45, 49]，因此可认为四、十月份处于该范围内的有效温度时段为过渡季。四、十月份有效温度不处于这个范围内的，暂认为季节属性无法判别。可通过后续决策树对训练数据的规则挖掘对该时段数据分类。

输入训练数据为一月、七月逐时数据，以及四、十月份有效温度落在 15-23℃ 的数据，通过训练，决策树生成的季节判别规则如下图所示：

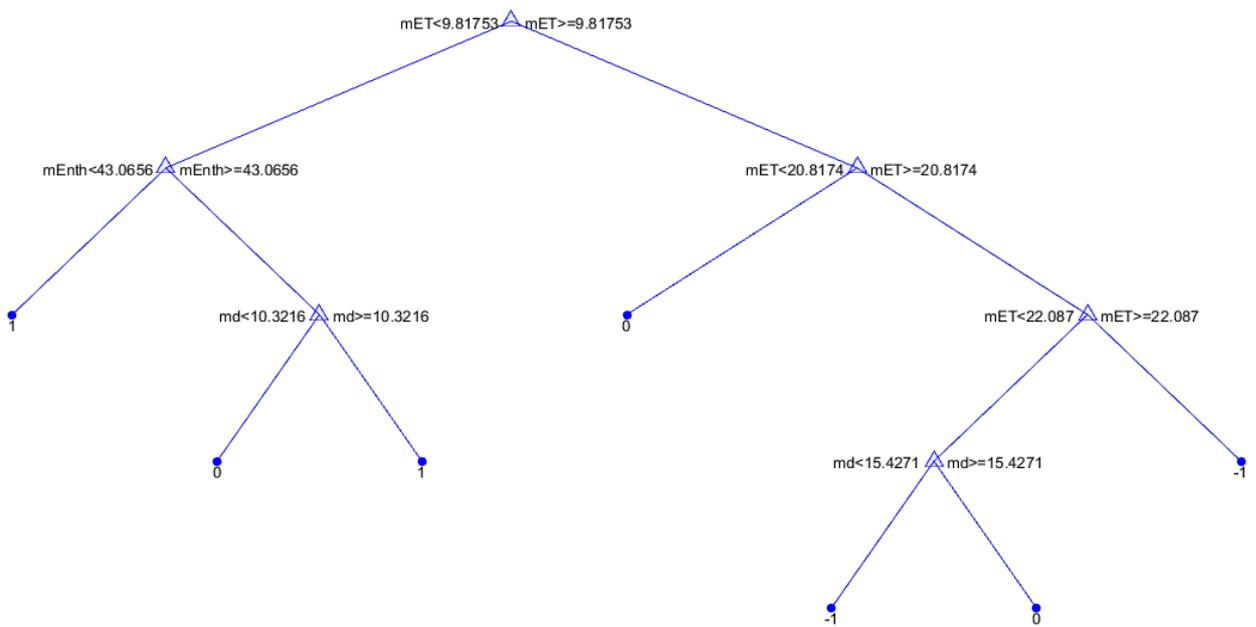


图 4.2 供热制冷季节判断规则

图中 mET 表示四个时刻滑动平均后的有效温度 (°C)，mEnth 表示平均焓值 (kJ/kg)，md 表示平均含湿度 (g/kg)。根节点根据权衡各个属性的信息增益度，以有效温度作为根节点属性，当前时刻滑动平均有效温度小于 9.8℃ 时，数据归向左分支，向下划分供热季和过渡季。根据人体热舒适性相关研究，当有效温度作小于 9℃ 时，人体会产生冷感觉^[45, 49]。二级节点考察空气焓值，左侧分支当焓值小于 43kJ/kg 时到达叶子节点，可判定当前时刻需要制热，为供热季。反之，通过三级节点含湿量大于还是小于 10g/kg 得出唯一属性类别，从而划分出供热季和过渡季。根节点数据有效温度大于 9.8℃，数据向右侧分支流动，从而最终划分出制冷季和过渡季。右侧二级节点信息增益度最大的也是焓值，当焓值小于 20.8kJ/kg 可判断为过渡季。反之继续向下到第三季节点，当有效温度大于

22℃时判断为制冷季。当有效温度小于 22℃时通过含湿量是否大于 15g/kg 将第四层节点的数据子集分为过渡季和制冷季。

经过训练，生成的决策树分类精准度为 92.5%。模型混淆矩阵如下图所示，该模型整体对过渡季分类的误差较大为 10%，其中过渡季和供热季的分类误差较大为 7%，主要由于训练数据中过渡季天气参数和供热季天气参数重合较多，且决策树左边分支级数较少。但为了防止数据过拟合，且针对逐时的季节判断，该模型精度已达到要求。

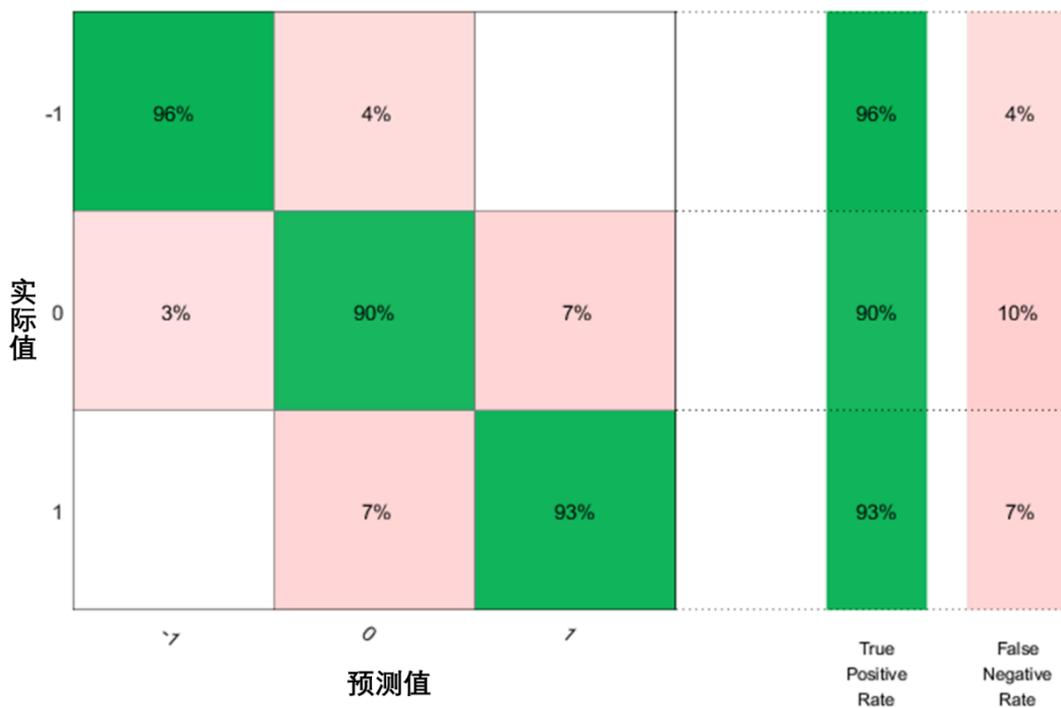


图 4.3 分类决策树混淆矩阵

4.1.3 当日季节判定

通过上述决策树模型已获得的逐时季节分类结果，如何对当日的季节属性做出判断，在此给出了统计不同时刻的季节属性众数，并进行加权平均确定当日的季节属性的分析思路，如图 4.1 所示。首先根据日峰谷变化，将一天分为夜间低谷时段（0:00-7:00），日间高峰时段（8:00-18:00）以及晚间高峰时段（19:00-23:00），此处的加权系数是根据人在该时段开空调的可能性确定的，类似于贝叶斯先验概率。这里设日间高峰时段和晚间高峰时段加权系数相同为 0.45，夜间

低谷时段为 0.1。日季节属性的划分为第 4.2 章节确定基础负荷和定义相似日打下基础。

表格 4.1 所示为度日数法和各时段季节属性众数加权平均法的结果对比，该方法对处于过渡季和制冷季，以及过渡季和供热季临界点的判定有了更细致的划分。例如表中第一行所示，4 月 9 日的日平均气温为 17.04℃，按照度日数法该温度小于 18℃，因此判定为供热季。但考察其日间温度较高为 17-24℃，根据上述规则该日日间并不需要供暖，只需夜间供热，因此判断其加权季节属性为 0.1。过渡季与制冷季临界点的判断类似，例如表中第三行所示，6 月 6 日的平均温度低于制冷度日数的判断基准温度 26℃，判定为过渡季。但因其日间和晚间峰值负荷时刻的温度均较高，因此日间存在极大可能性会出现冷负荷，加权季节划分为 0.9，且 6 月其他加权平均的季节属性偏向于制冷季。其他月份如 11、12、1、2 月和 3 月上旬两种方法判定结果一致，为供热季。7、8 月和 9 月上旬两种方法判定结果一致，为制冷季。

表格 4.1 日季节属性划分结果对比

日期	日平均温度 [°C]	度日数法季节划分	加权平均季节划分
4 月 9 日	17.04	1	0.1
5 月 9 日	18.28	0	0.55
6 月 6 日	25.64	0	-0.9
9 月 4 日	24.70	0	-0.45
10 月 28 日	17.67	1	0.1

4.2 电力负荷拆分

细化了季节属性分类规则后，可对电力负荷进行拆分。其主要目的是将制冷季和供热季的电力负荷拆分成基础负荷和相对天气变化较敏感部分。电力负荷异常值的存在会影响预测模型的精度，异常值检验包括数据缺失值和离群值的检验。在进行负荷拆分时，应首先消除历史数据中的异常值，提高数据质量。

4.2.1 电力负荷异常值检测

根据上一章节对电力负荷的特性分析可知，电力负荷相邻时刻变化较为平滑，

不存在突然地变大或变小。此外相同日类型的临近几日的电力负荷，其变化趋势相似。因此，电力负荷逐时的阶跃式变化，或相邻几日负荷特性明显的不同均将归为离群值。

缺失值检测：表现为某一时刻点没有负荷数据记录，可通过插值法弥补缺失数据。例如单一时刻负荷缺失，可通过前后两个时刻电力负荷的平均值代替该时刻的数据。若表现为连续几个时刻数据缺失，采用三次样条曲线进行插值，代替各时刻数据。对于2014年上海每15分钟负荷数据，检测到的均为单一时刻缺失值，采用线性差值进行了数据填补。

离群值检测：离群值被定义为明显偏离数据集中的其余部分，不服从相同统计模型的数据。离群值的存在具有相对性，正如电力负荷历史数据中，某一时刻负荷数据相较于相邻时刻变化平滑，但就临近几日相同时刻的负荷数据而言，该时刻数据明显偏离。局部观点而言判定该数据为非离群值，而全局观点该负荷数据为离群值。电力负荷根据工作日和节假日属性分成两组，分别在两组内对离群数据进行局部和全局检测。本课题采用基于统计的检测方法，即认为逐日数据的分布规律相同，而与模型不一致的分布则被认定为离群值。

工作日/节假日离群值局部检测：以每日96点数据为局部考察的子集，对每15分钟数据进行分析。设 $X(i)$ 为某日负荷序列且服从正态分布，其中 $i=1, 2, \dots, N$, $N=96$ 。

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad 4-4$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2} \quad 4-5$$

其中 \bar{x} 为电力负荷96点样本平均值， σ 为样本标准差。

根据随机变量概率分布，当 $|x_i - \bar{x}| > 3\sigma$ ，则认为 x_i 在 i 点负荷值的出现为小概率事件，判定为离群值。且该点负荷通过上下时刻线性平均值代替。

工作日/节假日离群值全局检测：全局检测分别在工作日和节假日两组数据中进行。设 $X(i, k)$ 为第 k 日第 i 时刻电力负荷，其中 $i=1, 2, \dots, 96$, $k=1, 2, \dots, n$ ，其中 n 根据不同月份工作日或节假日的天数而定。

$$\bar{x}(i, n) = \frac{1}{n} \sum_{k=1}^n x(i, k) \quad 4-6$$

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x(i, k) - \bar{x}(i, n))^2} \quad 4-7$$

全局检测的考察子集为 n 天同一时刻的电力负荷数据集，因此对于一个月份共进行 96 次检验。根据随机变量概率分布，当 $|x(i, k) - \bar{x}(i, n)| > 3\sigma$ ，则认为 $x(i, k)$ 在第 k 日 i 点负荷值的出现为小概率时间，判定为全局离群值。有趣的是，在对逐时电力负荷数据进行过离群值局部检测后，当 k 日第 i 时刻被判定为全局离群值后，k 日 96 点时刻大概率为全局离群值，呈现以日为单位的统一特性，总结如下表所示：

表格 4.2 全局离群值统计

日类型	样本数	离群值天数	所占比例
工作日	243	8	3.29%
节假日	122	3	2.45%
总计	365	11	3.01%

2014 年工作日天数共计 243 天，其中 8 天为全局离群值，分别为春节前的三日，春节后两日，以及国庆长假前的一日和长假后的两日。考察其负荷特性，明显低于临近工作日的用电能耗水平。节假日共计 122 天，其中 3 天为全局离群值，为 1 月 31 日-2 月 2 日，为春节假期的前三天。由此可见，长假期对于节前，节中和节后的电力负荷特性影响均较大。

4.2.2 电力负荷拆分

根据对电力负荷季节属性的判断，可对供暖季和制冷季的电力负荷进行基础负荷和气象敏感负荷部分的拆分。电力负荷拆分流程如图 4.4 所示。基础负荷定义为受生活作息影响较大，而与气象变化相关性较小的负荷。根据第三章不同季节与气象因素相关性分析可知，过渡季电力负荷与温度或综合气象温度相关性为 0.5 左右，因此可采用过渡季历史数据的各时刻加权平均的日负荷曲线作为电力负荷的基础负荷部分。

根据 4.1.3 章节中的季节属性判定方法，在计算基础负荷逐时电力负荷时，只采用各个时段众数统计为 0 的电力负荷进行加权平均。对于季节属性值的绝对值判定为小于 0.5 且不等于 0 时，判断当日为不完全过渡季。在进行平均基础负

荷基线计算时，将众数为0的时段纳入样本中。非0时段仍需拆分成基础负荷和气象敏感负荷部分。

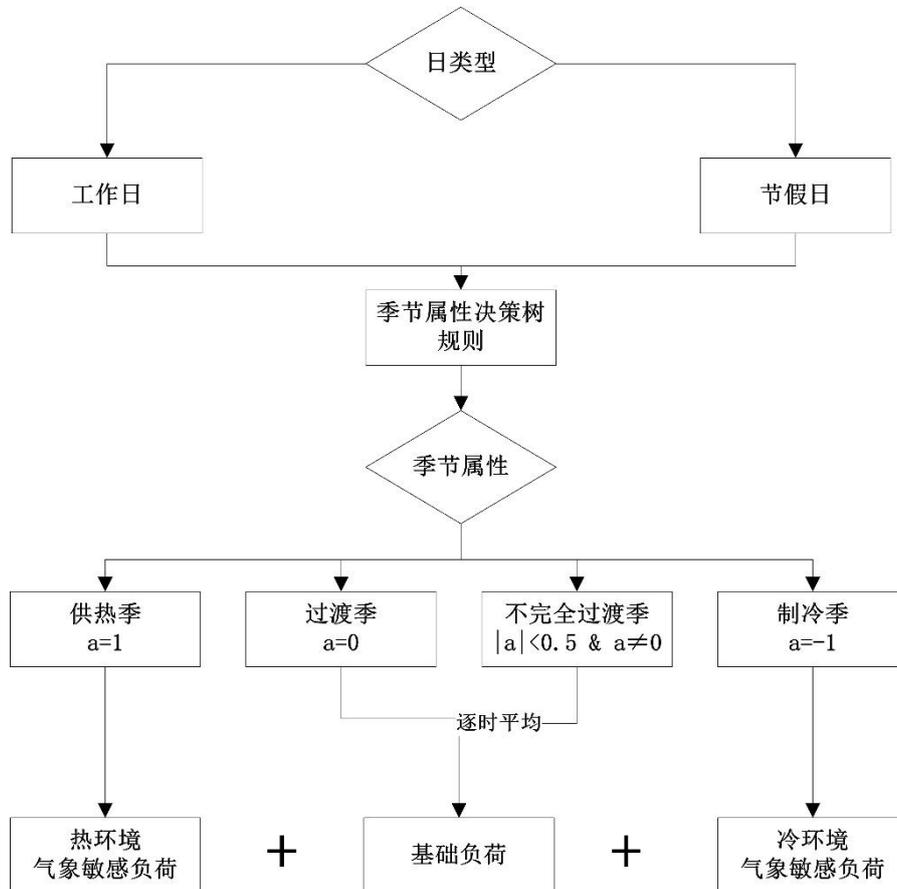


图 4.4 电力负荷拆分流程图

电力负荷拆分基本方法采用差分法，即在相同的日类型分组内，判断当日季节属性，若为供热季或制冷季，即用当日负荷减去基础负荷部分，多出部分为气象敏感负荷。

例如 2014 年 1、2、3 月供暖季，采用临近过渡季历史数据 2013 年 9 月、10 月各时刻加权平均的日负荷曲线作为基础负荷，由于缺乏 2013 年历史数据，本课题采用 2014 年 9、10 月份数据代替，并得出气象敏感负荷。

2014 年 6 月中的制冷季、7、8 月及 9 月上旬制冷季，采用临近过渡季历史数据 2014 年 4、5 月各时刻加权平均的日负荷曲线作为基础负荷，并得出气象敏感负荷。

2014 年 11、12 月供热季，采用 2014 年 4、5、9、10 月各时刻加权平均的

日负荷曲线作为基础负荷，并得到气象敏感负荷。

上述供暖季和制冷季中的基础负荷是由已发生的临近的过渡季逐时负荷加权平均得来的，因此不同月份相同日类型的基础负荷变化特性是固定的。

而对于日季节属性绝对值小于 0.5 的不完全过渡季或等于 0 的过渡季，由于统计的时段样本不断增加，因此基础负荷特性会发生微小变化，如图 4.5 所示

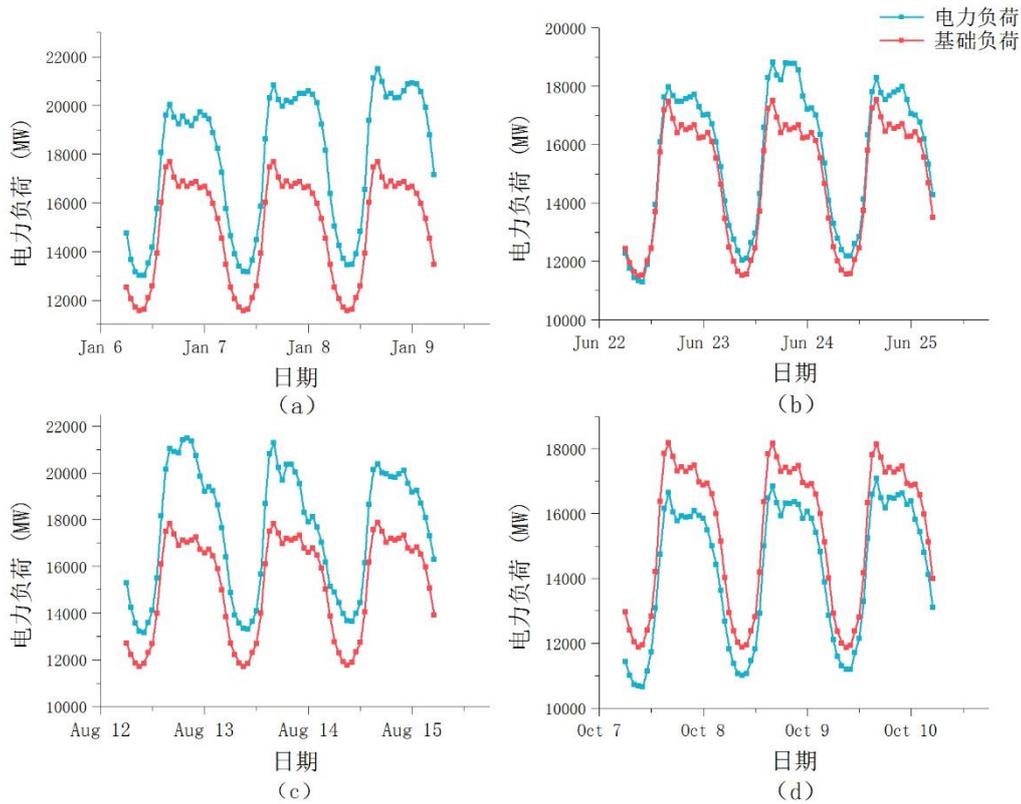


图 4.5 典型日逐 15 分钟负荷拆分示意图

图 4.5 选取不同季节属性工作日中典型日的基础负荷与气象敏感负荷的拆分图，从 (a) 至 (d) 分别为供暖季 1 月 7 日至 9 日，不完全过渡季 6 月 23 至 25 日，制冷季 8 月 13 日至 15 日，以及过渡季 10 月 8 日至 10 日。其中红色曲线为根据历史过渡季负荷计算出的基础负荷，蓝色曲线为当日的实际电力负荷，两者之间的差值即为气象敏感负荷。由计算可知，供暖季和制冷季相同日类型内基础负荷保持不变，如图中 (a) (c) 两幅图所示。

而不完全过渡季中的基础负荷逐日会存在微小变化，如图 4.5 (b) 图所示。

6月23至25日为不完全过渡季，即夜间时段判定为过渡季，而日间高峰时段划分为制冷季。基础负荷在夜间和晚高峰阶段和电力负荷基本重合，日间负荷存在气象敏感负荷。6月23日的基础负荷是以该日以前的过渡季逐时季节属性为0的序列的平均值计算而得。以此类推，6月24日在上述的子集中加入23日所求平均数而得。靠近与供热季的不完全过渡季，基础负荷和电力负荷的重叠部分刚好相反。日间两者表现为重合，而夜间和晚高峰期间电力负荷高于基础负荷，存在气象敏感负荷。

图4.5(d)所示为基础负荷高于当日电力负荷，10月8至10日正是工作日数据集中的全局离群值，为国庆长假后的三天工作日，整体负荷水平明显低于正常水平。通过上图也验证了异常值检验的准确性。在建立负荷预测模型时，应单独考虑长假前后的三天工作日。

电力总负荷与基础负荷的差值即为气象敏感负荷。通过计算气象敏感负荷占总体负荷的比例可知，供热季节所占比例居于6%-15%左右，制冷季所占比例在10%-26%之间。该比例与空调负荷占建筑能耗比例基本相符。判定差分法拆分负荷思路较为合理。

节假日的负荷拆分和工作日方法相同，这里不做赘述。

4.2.3 特殊节假日基础负荷处理

基于章节4.2.1检测出的全局离群点可知，需要对长假期及前后几日的基础负荷进行处理，否则将出现图4.5(d)的情况，即加权出的基础负荷高于总电力负荷，这是有悖于常识的。

以春节为例，经过分析，长假及前后几天电力负荷均有明显的线性减少和增加的趋势，采用电力负荷日均值为考察对象。分别统计全年相邻工作日和相邻节假日负荷日均值的变化率得出，相邻工作日负荷变化率均值为-0.06%，节假日为-0.28%。而在长假前后，相邻日负荷变化率显著增大，例如1月27日起，相较于前一个工作日1月24日负荷减小14%，春节第一天1月31日的电力负荷较前一日更是显著下降29%。因此在连续两个相邻同类型日负荷变化率绝对值大于4%时，即可认为因特殊节假日引起的负荷显著变化发生。在处理该时段的基础负荷时，可将逐时基础负荷按比例减小。图4.6所示为春节前后及节日期间基础负荷调整图。

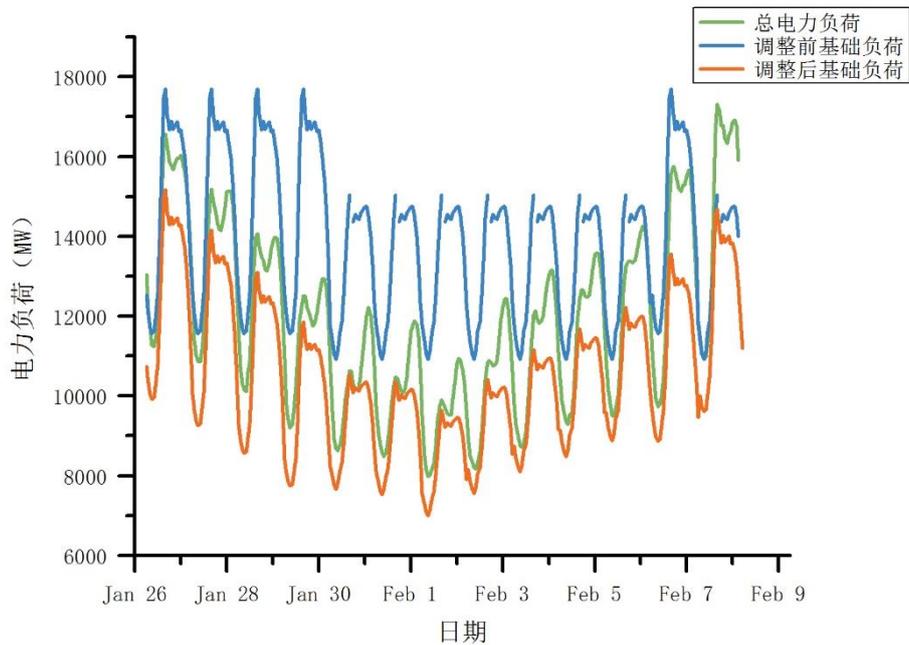


图 4.6 特殊节假日逐 15 分钟基础负荷调整结果

如图所示，春节期间基于日均负荷变化率调整的基础负荷小于总电力负荷，且可充分反映总电力负荷整体上的增加或减小。但是，由图中绿线所示可知，总电力负荷的变化特性与普通周末特性（图 3.2）并不相同，表现为夜间峰值更突出，且与日间的峰值差值更大。而逐时基础负荷的按比例增大或缩小是在相同日类型的历史数据，即普通周末居多的样本基础上完成，因此调整后的基础负荷走势无法表征出该时段负荷变化的特殊性，从而会导致拆分出来的气象敏感型负荷无法排除节假日对用电特性影响的因素。

4.2.4 气象敏感负荷变化特性

通过上述规则可将气候敏感型负荷拆分出来，该部分表现出与总电力负荷和基础负荷部分不同的变化特性。分别选取供热季和供冷季典型工作日作为观测对象，如图 4.7 所示。

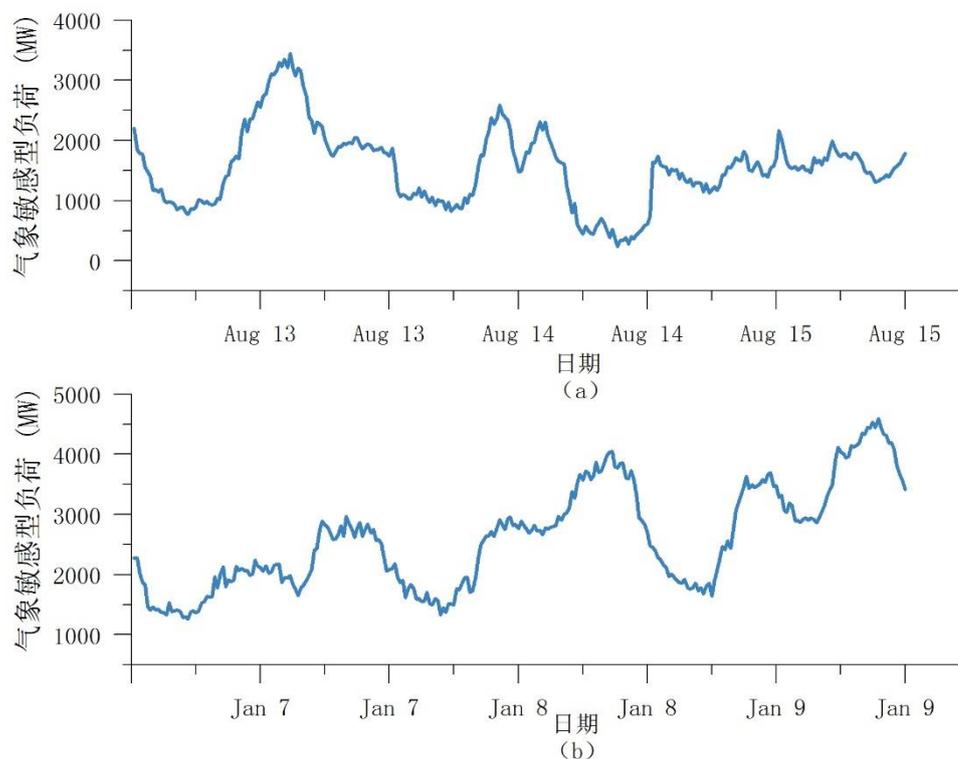


图 4.7 典型日逐 15 分钟气象敏感型负荷变化特性

气象敏感型负荷与总体负荷变化特性的不同表现在如下方面：

- i. 相邻日负荷变化周期性变弱。图中表现为相邻日负荷曲线走势差异性增大，且形状各异。这也突出了天气因素对电力负荷较为复杂的影响作用。而在研究总体负荷时，其主要变化趋势主要受到生活作息的影响，因此在同一日类型中呈现出一致的变化曲线。
- ii. 负荷变化平滑性减弱，随机变化性增强。图中表现为曲线整体波动较大，且短时间内上下抖动频率较高，曲线毛刺较多。负荷的拆分的作用类似于观测放大镜，将观测量级由总体的 10^4 缩小到 10^3 上，因此对于 4 次方量级上不明显的可被忽略的波动彰显在了 3 次方量级上。
- iii. 日负荷特性不同。例如峰值相较于总体负荷，出现时刻明显不同。图 3.2 冬季（供暖季）典型日负荷特性中日间高峰和晚间高峰峰值基本一致，而上图所示晚高峰峰值明显大于日高峰峰值。夏季（制冷季）两者表现较为一致，总体上日间高峰高于晚间高峰。

4.3 本章小结

本章节利用决策树算法根据逐时气象参数定义逐时的季节属性,并根据不同时段季节属性的众数加权得出当日季节属性,为电力负荷拆分奠定了基础。

基于第三章不同季节电力负荷与气象因素相关性不同的结论,定义了基础负荷与气象敏感型负荷的概念。并采用差分的方法,将供热季和制冷季的负荷拆分成两部分,为后续根据不同负荷特点分别建立预测模型打下了基础。

此外,着重探讨了特殊节假日中对基础负荷的处理,通过日均负荷的增大或缩小比例,相应放大或减小加权得出的基础负荷,有效消除了该时段基础负荷大于总电力负荷的情况。

最后针对拆分出来的气象敏感型负荷变化特性进行了探讨,并与第三章中总电力负荷特性结论进行了对比,明确了两者的异同。

第5章 基于时间序列法的基础负荷预测

上一章节定义了基础负荷的概念，并通过对待预测时刻、待预测日季节属性的划分，将电力负荷分解为基础负荷和气象敏感负荷。基础负荷具有变化周期性，相同日类型的电力负荷变化特性一致且较为稳定的特点。结合第二章对不同预测算法适用范围和优缺点的分析，本章节选用时间序列法对基础负荷进行预测，在保证模型精准度的前提下力求建模的简单化。

5.1 时间序列法理论介绍

时间序列法是发展最为成熟并被广泛应用于预测领域的算法之一。论其方法的成熟性，Box 和 Jenkins 早在 1976 年就建立了针对平稳时间序列预测模型的标准方法，基本可分为模型识别，参数估计和模型检验三个步骤^[9]。这套标准化的分析逻辑是建立在将时序数据看作一个随机过程的基础上，通过统计方法，从序列自相关的角度分析数据的变化规律，并推断出未来某时刻的数据。针对平稳时序数据，主要的预测模型有 MA, AR 和 ARMA. 对于非平稳的时序数据，则通过差分去除数据的趋势变化或季节性变化，对得出的新的平稳的时序样本建立 ARMA 预测模型，主要的预测模型有 ARIMA.

5.1.1 常用模型介绍

针对平稳时间序列有以下三个常用模型^[9, 10]：

MA (q): 该模型表示序列可由随机扰动和其 q 项滞后的线性组合构成，其数学模型表示如下，其中 $\theta(B) = 1 - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$ 为移动平均算子：

$$x_t = \theta(B)\varepsilon_t \quad 5-1$$

AR (p): 该模型表示序列可由 p 项序列本身滞后项的线性组合和随机扰动构成，其数学模型表示如下，其中 $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ 为自回归算子，且同样是 AR (p) 的系数特征多项式：

$$\phi(B)x_t = \varepsilon_t \quad 5-2$$

ARMA (p, q): 该模型是利用序列本身的滞后序列和随机扰动项及其滞后项，对数据的变化规律进行描述的方法。其数学模型表示为：

$$\phi(B)x_t = \theta(B)\varepsilon_t \quad 5-3$$

针对非平稳时间序列常用以下模型：

ARIMA (p, d, q)：该模型的核心对经过 d 阶差分后的平稳的时间序列建立 ARMA 模型，其数学模型与 ARMA 表达形式相同，不同的是序列本身转化为：

$$x_t = \nabla^d y_t \quad 5-4$$

其中 d 为差分级数， y_t 为原始时间序列， x_t 为差分后的时间序列。

5.1.2 平稳性检验

按照 Box 和 Jenkins 提出的步骤建立预测模型前，需要对研究对象的平稳性进行检验。检验平稳性的方法主要有两种，自相关系数法和单位根检验法。

自相关系数 (ACF, Autocorrelation Function) 是时间序列 $\{X(t)\}$ 各项之间的相关关系，其计算公式如下^[9]：

$$\gamma_{X_t X_{t-1}} = \frac{\sum_{t=2}^n (X_t - \bar{X}_t)(X_{t-1} - \bar{X}_{t-1})}{\sqrt{\sum_{t=1}^n (X_t - \bar{X}_t)^2} \sqrt{\sum_{t=1}^n (X_{t-1} - \bar{X}_{t-1})^2}} \quad 5-5$$

根据数理统计证明，随机序列自相关系数的分布近似于 $N(0, 1/\sqrt{n})$ 的正态分布，且该分布与样本容量 n 有关。根据统计检验，判断自相关系数是否与 0 显著不同，对于给定的显著水平，一般取 $\alpha = 0.05$ ，可构成一个随机区间。根据时序数据是否基本落入该随机区间内判断数据是否为随机序列。若时序数据时滞为 2 或 3 后自相关系数迅速趋于 0，则判断该样本是平稳的。反之若较多的自相关系数与 0 显著不同，即超出随机区间，则数据为非平稳的。

该方法虽然简单但不够精确，因此作为初步检测手段。单位根检验是较为准确的检测时间序列平稳性的统计检测方法^[50]。所谓单位根检验是对自回归模型对应的齐次线性差分方程 $\phi(B) = 0$ 的特征根是否为 1 进行检验，若特征根均在单位圆外，则序列平稳。

假如时间序列满足一阶自回归模型：

$$x_t = \alpha_1 x_{t-1} + \varepsilon_t \quad 5-6$$

为保证过程的稳定性，需满足系数 $|\alpha_1| < 1$ 。若 $\alpha_1 = 1$ ，随着时间的推移数据无法收敛，则序列不稳定。推广到高阶自回归模型，如公式 5-7 所示，在其基础上两边同时减去 x_{t-1} ，则 x_{t-1} 前系数变为 $\gamma = \sum \alpha_i - 1$ 。其原假设和备择假设则

变为:

$$H_0: \gamma = 0$$

$$H_1: \gamma < 0 \quad 5-8$$

通过计算检验统计量判定该假设接受或拒绝,从而判定序列特征方程是否存在单位根。上述单位根统计检测方法即为 ADF (Augmented Dickey-Fuller) 检测,由 Dickey 和 Fuller 在 1979 年提出的。^[51]

5.1.3 模型识别和滞后阶数确定

模型的识别主要通过时序数据自相关函数 (ACF) 和偏相关函数 (PACF, Partial Autocorrelation Function) 的变化特性进行判断^[9, 10]。偏相关性是考察时间序列 $\{X(t)\}$ 时刻 t 和时刻 $t-k$ 值的条件相关性,其中时刻 $t-1$ 到 $t-k+1$ 时刻的值已给定。在排除了其他后滞时刻的影响,只分析两时刻的关联程度。偏相关系数计算公式如下:

对由 $x_{t-1}, x_{t-2}, \dots, x_{t-k}$ 的线性自回归得出的 x_t 做线性最小方差估计,使得以下公式达到最小:

$$\delta = E(x_t - \sum_{j=1}^k \phi_{kj} x_{t-j})^2 (j = 1, 2, \dots, k) \quad 5-9$$

对 δ 对 ϕ_{kj} 求偏导数,使其为 0,可得到一下线性方程组,求解即可求得序列的偏相关函数。

$$\begin{bmatrix} 1 & r_1 & \cdots & r_{k-1} \\ r_1 & 1 & \cdots & r_{k-2} \\ \vdots & \vdots & \cdots & \vdots \\ r_{k-1} & r_{k-2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kj} \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_k \end{bmatrix} \quad 5-10$$

不同模型在自相关和偏相关函数的特征不同,由于算法证明非本课题重点,因此推导证明本课题不作具体阐述,仅给出结论如下表所示:

表格 5.1 平稳序列时间序列常用模型特征^[9, 10]

模型	自相关函数 (ACF)	偏相关函数 (PACF)
AR (p)	指数衰减	p 阶截尾
MA (q)	q 阶截尾	指数衰减
ARMA (p, q)	指数衰减	指数衰减

由上表可知, 可通过观测时序样本自相关函数和偏相关函数的特性, 结合上表不同模型的特征, 选择合适的模型并对滞后阶数进行初步估计。

在确定模型阶数时, 常采用建立在最大似然估计基础上的最小信息准则, 较为常用的是赤池信息准则 (AIC, Akaike information criterion) 对 p, q 进行估计^[52, 53], 得出最小 AIC 下的 p, q 为模型参数, 计算公式如下:

$$AIC = 2k - 2 \ln L \quad 5-11$$

其中, K 是参数的数量 $k=p+q$, L 是似然函数。假设预测负荷误差服从正态分布, 则公式如下, 其中 RSS 为残差平方和, 则 RSS/n 为残差极大似然估计的方差:

$$AIC = 2k + n \ln RSS/n \quad 5-12$$

AIC 的取值取决于残差的方差和 K, 两个参数的取值越小, AIC 越小。K 值小代表模型较为简化, 残差方差小则说明模型精确。因此在选用 AIC 最小值作为模型定阶依据, 可兼顾模型的精确性和简化性^[52]。

5.1.4 模型检验

在识别出模型类型及定阶后, 采用最小二乘法确定模型参数, 及特征多项式的系数。为判定回归模型的优劣, 还需要对模型的残差进行检验, 以便确定其是否为白噪声序列, 白噪声是纯随机序列。若判定为白噪声, 说明模型较为合理。否则需要对模型进行改进。最为常用的检验方法有残差自相关检验。

该检验方法简单, 即计算残差自相关系数, 判断各个系数是否与 0 显著不同。若各系数基本落在显著水平构成的随机区间内, 则判定残差序列为随机序列。说明建立的模型已包含原时间序列所有的变化规律, 模型较为合理。

5.1.5 模型评估标准

在判定建立的模型合理后,还需要对模型的预测性能进行评估,以下模型评估指标不仅适用于时间序列模型,同时适用于所有的预测算法模型。预测模型性能评估主要从以下两个方面进行,分别是拟合度和预测准确度。

拟合度用决定系数 (R^2 , Coefficient of determination) 表示, R^2 介于 0 至 1 之间,越接近 1 说明模型拟合程度越高,计算公式如下:

$$R^2 = 1 - \frac{\sum_{i=1}^n (x_i - \hat{x}_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad 5-13$$

度量预测准确度可用预测误差的标准差表示,均方根误差 (RMSE, Root mean square error) 越小,表明预测模型预测值距离实际值越近,模型准确越高,计算公式如下:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (x_i - \hat{x}_i)^2}{n}} \quad 5-14$$

平均绝对误差 (MAE, Mean absolute error) 表征模型预测值和实际值的误差,同样给出模型的准确度。采用误差绝对值进行计算是防止正负误差的相互抵消,计算公式如下:

$$\text{MAE} = \frac{\sum_{i=1}^n |x_i - \hat{x}_i|}{n} \quad 5-15$$

同时平均相对误差 (MAPE, Mean absolute percentage error) 可表征平均误差的相对大小,同样反映出模型的准确度。该指标可对不同量级数据建立的模型进行比较,计算公式如下:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \frac{|x_i - \hat{x}_i|}{x_i} \quad 5-16$$

其中 x_i 为实际值, \hat{x}_i 为预测值, n 为样本数量。

5.2 实例分析

本章节旨在采用时间序列法对过渡季的基础负荷建立预测模型,并判定该方法的可行性和优越性。通过选取过渡季不同日类型的基础负荷作为实例样本,分别建立相应的预测模型,并对模型性能进行评估。

5.2.1 工作日实例分析

根据第四章对电力负荷的拆分处理可知，基础负荷只有在过渡季和不完全过渡季会随着时间变化而改变，供热季和供冷季逐日基础负荷变化特性相同。在选取工作日样本建模时，只需考虑对过渡季和不完全过渡季的基础负荷。以3月17日至3月21日连续五个工作日数据为例进行训练建模，并对3月24日负荷进行预测，分析模型预测性能，其中间隔日期为节假日。

电力负荷平稳性检测与模型识别

按照上述建模方法首先对样本的平稳性进行检验，首先对原基础负荷数据进行ADF检验，其结果为 $h=0$ ，表示该测试不能拒绝 $\gamma = 0$ ，即存在单位根的假设，说明数据是不平稳的。考察序列的自相关系数变化趋势，如图5.1所示：

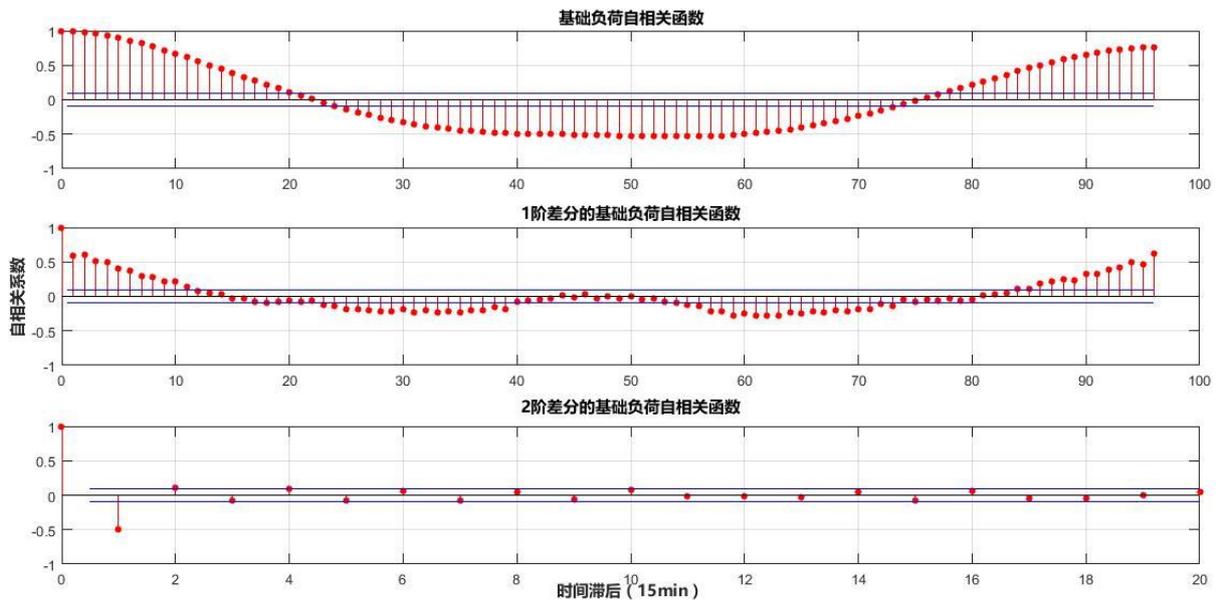


图 5.1 样本基础负荷及2阶差分内的基础负荷的自相关函数

如图所示，基础负荷的自相关系数呈现周期性变化，在对数据进行1阶差分后其周期相关性依然存在，且大部分系数落在随机区间之外，说明1阶差分后的基础负荷序列仍不平稳。分析2阶差分序列，自相关系数在时滞大于2后迅速趋于0，落在随机区间之内。经过ADF检验后，拒绝 $\gamma = 0$ 的假设，说明基础负荷经过2阶差分后已趋于平稳。

为确定预测模型的形式，考察基础负荷序列的偏相关函数，如图5.2所示，

其中横坐标代表以 15min 为采样间隔的基础负荷的时间滞后项个数。总体而言，序列的偏相关函数在数个滞后项后迅速趋于零，呈现出明显的指数衰减趋势。随着对基础负荷平稳化的处理，偏相关系数趋于 0 的速度越快。对于 2 阶差分的基础负荷，在 $k=3$ 后，偏相关系数基本为 0。

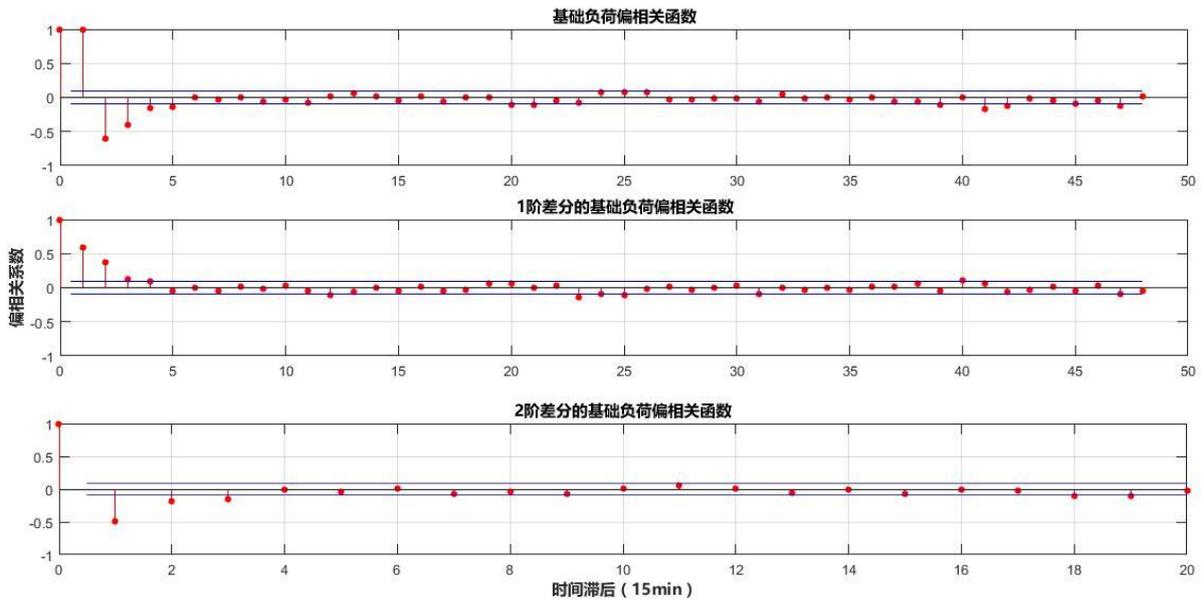


图 5.2 工作日样本基础负荷及 2 阶差分内的基础负荷的偏相关函数

根据 2 阶差分后的基础负荷自相关函数在 $k=2$ 和偏相关函数 $k=3$ 后呈现指数衰减特性，初步判定可对该序列建立 ARMA (2, 3) 的预测模型，对原基础负荷序列可采用 ARIMA (2, 2, 3) 的预测模型。

模型定阶和参数确定

模型定阶过程采用穷举法，组合 p, q 在^[1,5]范围内的所有组合，对平稳化的基础负荷分别建立模型，并比较 AIC 值。取 AIC 最小值时的 p, q 组合，该值即为预测模型的最佳时滞阶数值。图 5.3 所示为针对实例样本建立的预测模型 AIC 值寻优的三维图，其中 $p=5, q=3$ 时，AIC 达到最小。因此针对该训练样本，模型自回归时滞阶数为 5，误差时滞阶数为 3。表格 5.2 所示为 ARIMA (5, 2, 3) 预测模型各参数的取值和分布检验。

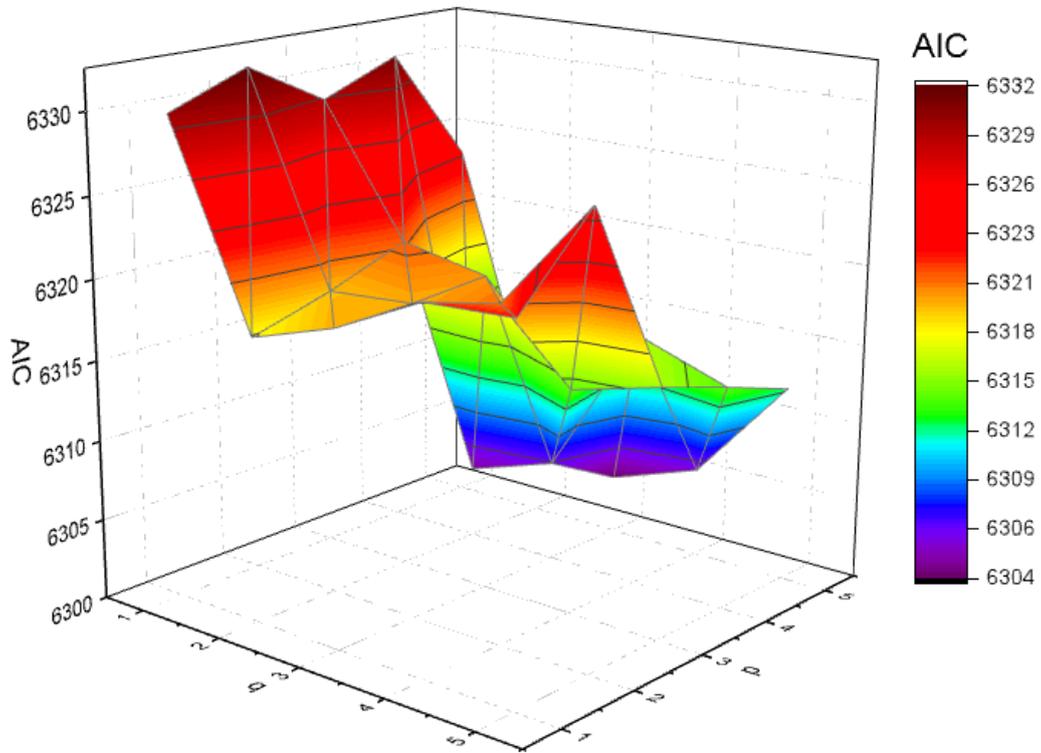


图 5.3 工作日样本时间序列模型时滞阶数确定

表格 5.2 工作日样本预测模型参数确定

参数	参数取值	标准误差	t-检验
常数项	0.0265	0.284	0.0933
AR{1}	-1.097	0.199	-5.500
AR{2}	0.203	0.155	1.308
AR{3}	0.741	0.135	5.494
AR{4}	0.414	0.106	3.883
AR{5}	0.104	0.0573	1.809
MA{1}	0.418	0.201	2.078
MA{2}	0.859	0.0439	-19.555
MA{3}	0.559	0.204	-2.734

预测模型评估

在计算出预测模型具体形式后,可对 3 月 24 日 96 点数据进行预测,并与该日实际基础负荷进行对比,以便评估预测模型性能。对于测试集的模型评估指标如表格 5.3 所示,整体而言,采用时间序列法对过渡季基础负荷的预测精准度较高,相对误差小于 2%,且模型的拟合度较高。表明该方法对于预测日间变化稳定的基础负荷性能较好。

表格 5.3 工作日样本测试集预测模型评估表

评估参数	数值
R^2	0.953
RMSE	343.390
MAE	265.370
MAPE	1.75%

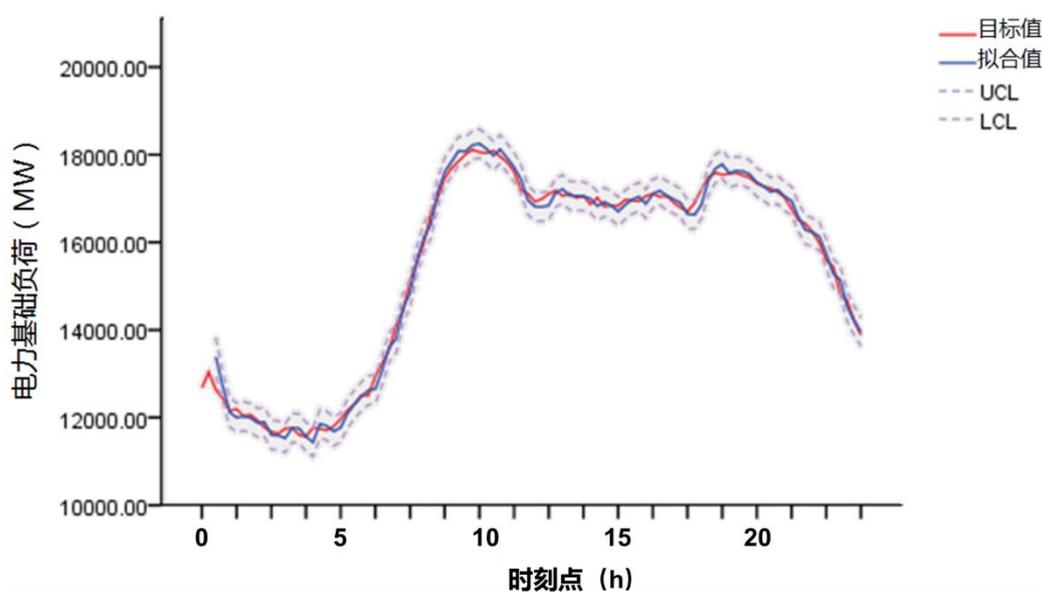


图 5.4: 3 月 24 日逐 15 分钟电力基础负荷预测

图 5.4 所示为对 3 月 24 日 96 点数据预测值和目标值的时序对比图,下图所示预测值与目标值高度拟合,较好的预测出基础负荷的变化特性。此外,在给出单一预测值的基础上,对预测值的 $\pm 95\%$ 置信区间进行了估计,预留出因未知

因素引起的负荷变动空间的估计。

5.2.2 节假日实例分析

对于过渡季节假日基础负荷模型的预测，其建模原理和思路同上述工作日实例一致。以 5 月 10 日-5 月 24 日期间的节假日数据作为训练集，其中节假日共计 5 天，对 5 月 25 日的电力基础负荷作为测试集，进行模型评估。在对数据进行平稳性判断，模型识别和定阶后，确定模型的形式为 ARIMA (3, 2, 2)，通过最小二乘法所得的系数如下表所示：

表格 5.4 节假日样本预测模型参数确定

参数	参数取值	标准误差	t-检验
常数项	-0.715	8.603	-0.0832
AR{1}	-0.998	0.135	-7.410
AR{2}	-0.150	0.194	-0.774
AR{3}	-0.0392	0.0868	-0.452
MA{1}	0.359	0.135	2.650
MA{2}	-0.383	0.119	-3.218

同理，在计算出预测模型具体形式后，可对 5 月 25 日 96 点数据进行预测，并与该日实际基础负荷进行对比，以便评估预测模型性能。对于测试集的模型评估指标如表格 5.5 所示，整体而言，采用时间序列法对过渡季基础负荷的预测精准度较高，相对误差小于 3%，且模型的拟合度较高。

表格 5.5 节假日样本测试集预测模型评估表

评估参数	数值
R^2	0.921
RMSE	312.450
MAE	255.180
MAPE	2.03%

图 5.5 所示为对 5 月 25 日 96 点数据预测值和目标值的时序对比图，下图所示预测值与目标值高度拟合，较好的预测出基础负荷的变化特性。此外，在给出单一预测值的基础上，对预测值的 $\pm 95\%$ 置信区间进行了估计，预留出因未知因素引起的负荷变动空间的估计。

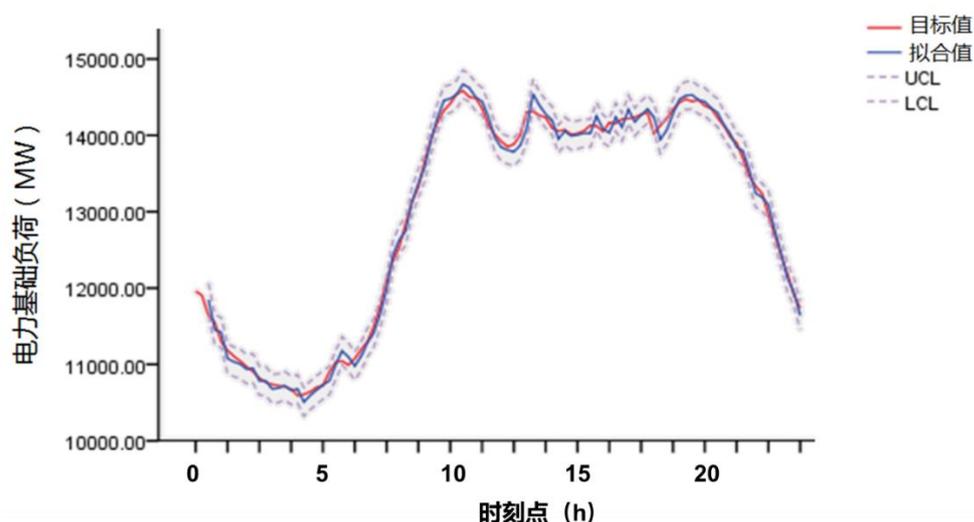


图 5.5 5 月 25 日逐 15 分钟电力基础负荷预测

将工作日模型与节假日模型预测性能进行对比发现，工作日模型的 R^2 更高，且预测相对误差更小。主要由于相邻节假日的时间跨度较长，无法保证间隔节假日的基础负荷的连续性，导致不同节假日间的基础负荷变化特性相似性减弱，从而削弱了预测精度。但总体而言，基于时间序列法对基础负荷的预测方法，可将相对误差控制在 3%，且对负荷变化趋势拟合较准，已可很好的满足预测需求。

5.3 本章小结

本章旨在对过渡季和不完全过渡季的基础负荷进行预测。根据基础负荷变化周期性强且变化特性稳定的特点，采用时间序列法通过历史负荷的自回归建立预测模型，保证预测精度的基础上力求预测方法的简单化。

本章首先对基于时间序列的预测原理进行了简单的介绍。在此基础上，分别选取不同日类型的基础负荷作为实例建立预测模型，并通过预测下一日的基础负荷对模型的预测性能进行评估。最后得出，该方法对于工作日和节假日基础负荷的预测均较为适用，平均相对误差小于 2.5%，预测精准度较高，且模型的拟合度 R^2 也较高，能够拟合出基础负荷的变化趋势。

第6章 基于支持向量机的气象敏感负荷预测

对于气象敏感负荷的预测,由于各类气象参数对电力负荷较为复杂的非线性关系,利用第五章时间序列法预测,即只利用历史负荷的自回归将产生较大误差。因此对于气象敏感负荷的预测选择能够将非线性关系逼近较好的机器学习算法。本课题因支持向量机算法具有易于训练,泛化能力强,不易陷入局部最小点和模型训练前不需要优先确定网络拓扑结构的优点^[54],采用 SVM 对该部分负荷进行训练和预测。

6.1 支持向量机理论介绍

支持向量机 (Support Vector Machine) 的核心是基于统计学习理论的机器学习方法,在解决小样本、非线性和高维问题中具有良好的适用性^[26]。SVM 通过将问题转换成求解二次凸规划的数学模型,求得全局最优解,有效避免了有些算法对初始值较为敏感,易陷入局部极值的问题。同时在处理非线性问题上,通过核函数的变化将其映射到高维特征空间,再利用构造最优超平面逼近目标函数,解决低维空间中的非线性问题,使其具有较好的泛化能力。^[54]

6.1.1 结构风险最小化

机器学习的目标是在有限的样本数目内,求得一个函数 $f(x, \beta)$, 其中 β 为函数参数集合,逼近输入 X 与输出 Y 之间确定但未知的映射关系 F , 即未知的联合概率分布 $F(x, y) = F(x)F(y|x)$ 。为了度量学习效果,通常采用风险泛函数 $R(f) = \int L(y, f(x, \beta)) dF(x, y)$ 表示逼近真实分布的程度,因为 F 未知,采用经验风险函数 $R_{emp}(f)$ 替代,计算公式如下:

$$R_{emp}(f) = \frac{1}{n} \sum_{i=1}^n L(y_i, f(x_i, \beta)) \quad 6-1$$

针对回归问题损失函数 $L(y, f(x, \beta)) = (y - f(x, \beta))^2$ 。通过寻找 $R(f)$ 最优值确定最佳的映射关系 $f(x, \beta)$, 即认为是 $F(x, y)$ 关系的最佳逼近。 $R(f)$ 由两部分组成,经验风险和置信界,三者随 VC 维的变化示意图如下图示。计算公式如下:

$$R(f) = R_{emp}(f) + \varepsilon(n, \delta, f) \quad 6-2$$

其中 $\varepsilon(n, \delta, f)$ 为针对某一机器学习算法拟合出的映射关系 $f(x, \alpha)$ 的置信风险，通常认为小于等于其粗糙界 $\varepsilon(n, \delta, f) \leq \mathcal{O}(n, \delta, f)$ 。其中评估 $R(f)$ 最优解基于两种理论，经验风险极小化（Empirical Risk Minimization, ERM）和结构风险最小化（Structural Risk Minimization, SRM）。

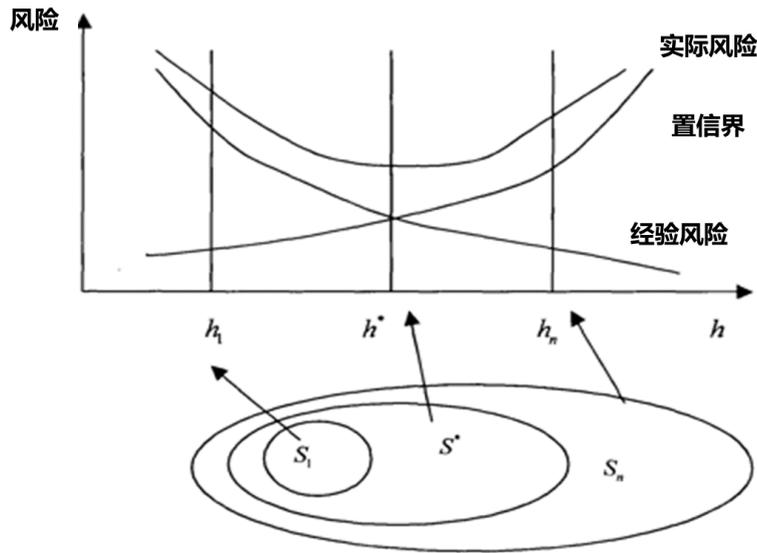


图 6.1 实际风险构成示意图^[55]

基于 ERM 的理论认为，当样本无穷大时，经验风险 $R_{emp}(f)$ 的最小值将逼近实际风险泛函数 $R(f)$ 的最小值，最小二乘法，BP 神经网络算法等均采用经验风险最小化理论。但针对现实问题，由于样本规模的限制，无法保证映射算法收敛到实际风险最小值。因而出现训练好的模型泛化能力较差的问题。

相反，基于 SVM 的理论则是通过判定 $R_{emp}(f) + \varepsilon(n, \delta, f)$ 为极小值，从而确定机器学习出来的函数 $f(x, \beta)$ 为最优。SVM 即基于结构风险最小化在经验风险，即函数的逼近能力，和置信界，即函数的复杂性中寻求平衡，使得总体风险达到最小。在保证精度的同时增强了其泛化能力。

6.1.2 回归支持向量机基本原理

SVM 方法最早由 Vapnik 提出用于模式识别问题，其基本特点是将对整个训练集数据的处理转化成对支持向量的处理，使得构造的超平面对支持向量的划分等同于对整个训练集的划分^[56]。随着 Vapnik 引入 ε 不敏感损失函数，该方法的应

用范围已被推广到非线性回归和曲线拟合中，即 ε -SVR 模型。 ε -SVR 的目标是通过模型训练寻找到函数 $f(x)$ ，使得该函数值与目标值 y 的距离小于 ε 。对于距离超出 ε 的点将导致回归算法的损失函数增大，损失函数公式如下：

$$L(y, f(x)) = \begin{cases} |y_i - f(x_i)| - \varepsilon, & |y_i - f(x_i)| > \varepsilon \\ 0, & |y_i - f(x_i)| \leq \varepsilon \end{cases} \quad 6-3$$

函数 $f(x)$ 模型如下：

$$f(x) = (\omega, \phi(x)) + b \quad 6-4$$

其中 (\cdot, \cdot) 代表向量 ω 与输入变量 x 的点积， $\phi(x)$ 表示从输入空间 x 映射的高维特征空间。最优函数转化成结构风险最小化的寻优问题，意味着最小化 VC 维的同时将训练集损失函数降到最低，数学公式如下：

$$\begin{aligned} & \min_{\omega, b} \left(\frac{1}{2} \|\omega\|_2^2 + C \sum_{i=1}^n \varepsilon_i \right) \\ \text{s. t. } & y_i \left((\omega, \phi(x_i)) - b \right) \geq 1 - \varepsilon_i, \quad \varepsilon_i \geq 0, \quad i = 1, 2, \dots, n \end{aligned} \quad 6-5$$

其中， C 为惩罚系数， C 值越大表示对超出 ε 范围的点惩罚越大。为解决上述高维度中的线性规划问题，可通过引入拉格朗日函数，解决其对偶问题完成，即：

$$L = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) - \sum_{i=1}^l (\eta_i \xi_i + \eta_i^* \xi_i^*) - \sum_{i=1}^l \alpha_i (\varepsilon + \xi_i - y_i + \langle \omega, \phi(x_i) \rangle + b) - \sum_{i=1}^l \alpha_i^* (\varepsilon + \xi_i^* + y_i - \langle \omega, \phi(x_i) \rangle - b) \quad 6-6$$

式中， L 是拉格朗日函数， η_i, α_i 是拉格朗日乘子向量。式中的对偶变量需满足非负数的约束条件：

$$\eta_i, \eta_i^*, \alpha_i, \alpha_i^* \geq 0 \quad 6-7$$

通过对函数 L 求解原始变量 $(\omega, b, \xi_i, \xi_i^*)$ 的偏导数，可得最优解：

$$\begin{cases} \partial_b L = \sum_{i=1}^l (\alpha_i^* - \alpha_i) = 0 \\ \partial_\omega L = \omega - \sum_{i=1}^l (\alpha_i - \alpha_i^*) \phi(x_i) = 0 \\ \partial_{\xi_i} L = C - \alpha_i - \eta_i = 0 \\ \partial_{\xi_i^*} L = C - \alpha_i^* - \eta_i^* = 0 \end{cases} \quad 6-8$$

消去 η_i, η_i^* 可得对偶优化问题：

$$\begin{aligned} \max_{\alpha_i, \alpha_i^*} & \begin{cases} -\frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) (\phi(x_i), \phi(x_j)) \\ -\varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) + \sum_{i=1}^l y_i (\alpha_i - \alpha_i^*) \end{cases} \\ \text{s. t.} & \quad \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0, \quad \alpha_i, \alpha_i^* \in [0, C] \end{aligned} \quad 6-9$$

最终结果可得，

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \phi(x_i) \cdot \phi(x) + b \quad 6-10$$

其中对于参数 $(\alpha_i - \alpha_i^*)$ 不为零的向量，称为该问题的支持向量。 $(\phi(x_i), \phi(x_j))$ 为将非线性数据投影到高维度空间的内积运算，由于其函数形式未知，可借助任何满足半正定条件的核函数 $K(x_i, x_j)$ 实现，巧妙的避免了需给出显式内积函数的问题。较为常用的核函数有径向基函数核（RBF），多项式核，神经网络核等。

对于时间序列数据的预测较为常用的核函数为径向基函数，公式如下：

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad 6-11$$

本章节将采用径向基函数核，但该核函数无法表征时序数据在时间上的相关性，因此在输入参数时将考虑历史负荷对当前预测负荷相关性因素。

6.1.3 回归支持向量机参数优化

在运用 SVR 模型对输入变量进行线性回归时，会出现即使映射到高维空间后，输入输出的现行关系仍不明确的情况，这需要通过调节 SVR 参数在模型复杂程度和泛化性上做出平衡。对 SVR 而言，需要调节的参数分为两类：一是模型固有参数，惩罚因子 C ，损失函数 ε 等。第二类是核函数中的参数，例如采用 RBF 把你的训练数据以高斯的方式投射到高维空间，所以好的分布，如核宽度 σ 反映出支持向量之间的相关程度^[57]，相关性过于松弛或过强均会影响训练结果。

参数的好坏决定着 SVR 模型的性能，选取过程也是一个寻优的过程，目前常用的参数优化方法有网格搜索法，遗传算法和粒子群优化算法等。

网格搜索法基本原理基于穷举法，组合给定参数范围内的所有组合，对训练集采用 k 次交叉验证，采用模型训练集中得到最小均方误差（MSE）所对应的参数组合为给定当前样本下的 SVR 模型最优参数。该算法原理简单，操作性较强，但对于训练样本数量过大时，寻优计算量大且时间过长。但立足于课题实现实时

预测的目的,且训练样本不大的情况下,可采用网格搜索法寻优。本章节针对 SVR 模型优化的参数有惩罚因子 C 和核函数参数 γ , 公式如下:

$$\gamma = \frac{1}{2\sigma^2} \quad 6-12$$

6.2 气象敏感性负荷预测

上述章节基于建筑能耗和暖通专业知识,根据电力负荷不同属性对其进行了分类。不同类别电力负荷所用模型归纳如表格 6.1 所示。本章节将就不同类别下的 SVR 模型分别进行训练,并通过示例评价模型性能。其中模型编号 w/h 分别表示不同日类别属性:工作日和节假日。h/c 则分别表示不同季节属性:供热季和制冷季。

表格 6.1 电力负荷预测模型汇总

模型	基础负荷		气象敏感负荷
工作日	供热季	过渡季历史负荷加权平均	SVR_wh
	制冷季	过渡季历史负荷加权平均	SVR_wc
	过渡季	时间序列法	SVR_wh/wc
节假日	供热季	过渡季历史负荷加权平均	SVR_hh
	制冷季	过渡季历史负荷加权平均	SVR_hc
	过渡季	时间序列法	SVR_hh/hc

6.2.1 模型输入参数选取

由于径向基函数核无法表征时序数据在时间上的相关性,在对当前时刻电力负荷进行预测时,无法表现出之前负荷对当前时刻的作用。在模型输入参数选取上,将考虑之前时刻负荷作为输入变量,借鉴时间序列法中 ACF 和 PACF 方程考察气象敏感负荷的相关性,以便确定当前负荷受此前几个小时负荷的影响,如图 6.2 所示,其中横坐标代表以 15min 为采样间隔的气象敏感型负荷的时间滞后项个数。

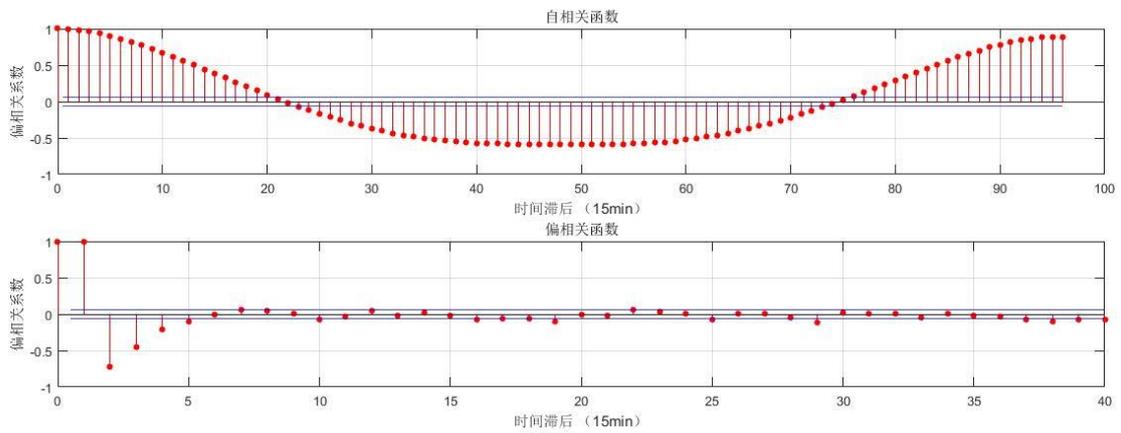


图 6.2 气象敏感负荷时间序列自相关和偏相关函数估计

从自相关函数可以看出，当前时刻负荷与后续负荷具有较强的相关性，且自相关函数呈现周期性，缓慢衰减的趋势。偏相关函数中，函数衰减趋势迅速，且在第四个时刻点后的负荷表现为与该点负荷不相关。选取当前预测负荷前四个时刻负荷作为模型输入参数。在最大程度的反映时序数据相互影响关系的前提下，减少输入参数维度，简化模型。同时将预测时刻的小时信息也作为输入变量。

此外，根据第三章不同季节电力负荷与气象参数相关性的分析，确定供热季中采用当前时刻四点滑动平均有效温度作为输入参数，供冷季采用当前时刻滑动平均温度、相对湿度和风速作为输入参数，过渡季中的部分制冷供热时段根据季节属性划分相应确定输入参数。综上，供热季模型输入空间为 6 维，制冷季模型输入空间为 8 维。

6.2.2 SVR_wh 模型参数优化

首先以搭建供暖季预测模型 SVR_wh 模型为例，为保证充足的训练样本数，选取 1 月 7 日-1 月 16 日十天 960 项数据作为训练集，以 1 月 17 日和 1 月 20 日数据为测试集，相隔日期为剔除的节假日。对于输入数据的离群值检测和修正已在 4.2.1 章节完成，数据归一化采用 3.3.3 章节中的方法。

图 6.3 所示为 SVR 模型参数惩罚因子 C 和核函数分布宽度 γ 的寻优结果图，通过网格搜索法以模型最小标准化的均方差 (MSE) 为目标，所得的 C 和 γ 即为该输入样本下的最优参数。针对上述训练样本最优方程参数 $C=2.828$ ， $\gamma=5.67$ 。

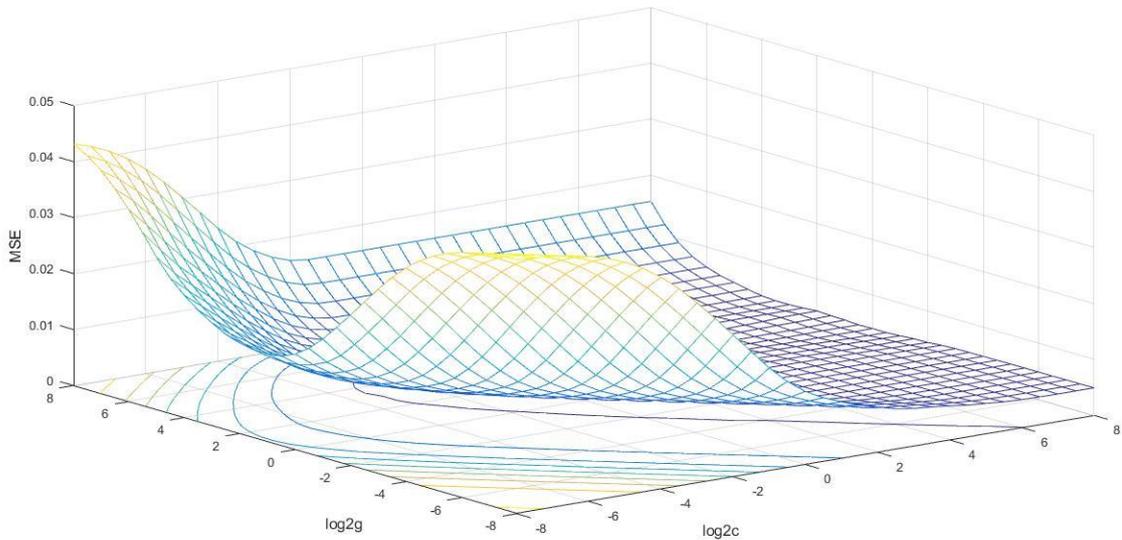


图 6.3 SVR_wh 模型参数寻优结果

6.2.3 SVR_wh 模型性能分析

和第 5 章模型评价指标相同，将从两个方面对预测模型性能做出评判，即决定系数 R^2 考察模型的拟合度，均方根差 RMSE 则表示的是残差的标准差，反映出模型的预测精准度，此外平均绝对 MAE 和平均相对误差 MAPE 可给出模型预测值的误差，反映模型的准确度。SVR_wh 模型基于训练集和预测集的性能参数汇总，第三列给出的是将测试集中的气象敏感负荷预测值与基础负荷相加，作为该模型对于总体负荷的间接预测，并与总体负荷目标值想比较得出的性质指标，定义为整体性能。

表格 6.2 给出 SVR_wh 模型基于训练集和预测集的性能参数汇总。训练集进行模型拟合时采用 5 折交叉验证的方法，训练出的模型具有较好的性能，首先 R^2 较高，具有较高的拟合程度。其次预测值和目标值的 RMSE 较小，说明残差分布较为集中，且 MAE 为 170.8，平均相对误差为 12%。相对而言，模型对于测试集的预测性能有所下降，但就平均相对误差的对比，结合输入参数维数较少且训练样本足够大，说明该模型具有较好的泛化性，排除了过拟合的情况。测试集中 RMSE 和 MAE 较大，主要由于气象敏感负荷的波动更为随机且周期变化性不强。但预测出的气象敏感负荷与基础负荷相加的数据，与总电力负荷序列比较得出 MAPE 显著降低，大大优化了对在整体预测集上的相对误差。

表格 6.2 SVR_wh 模型性能汇总

评估参数	训练集	测试集	整体性能
R^2	0.956	0.806	
RMSE	209.26	408.90	408.90
MAE	170.80	332.54	332.54
MAPE	12.06%	14.19%	1.91%
训练时间	1min33sec		

图 6.4 和图 6.5 分别展示了训练集和测试集上模型的拟合程度和偏差情况，总体而言模型可以较好的预测出电力负荷的变化走势。在图 6.5 测试集数据中可以看出，预测值波峰波谷，或较小的波动均滞后于预测函数。

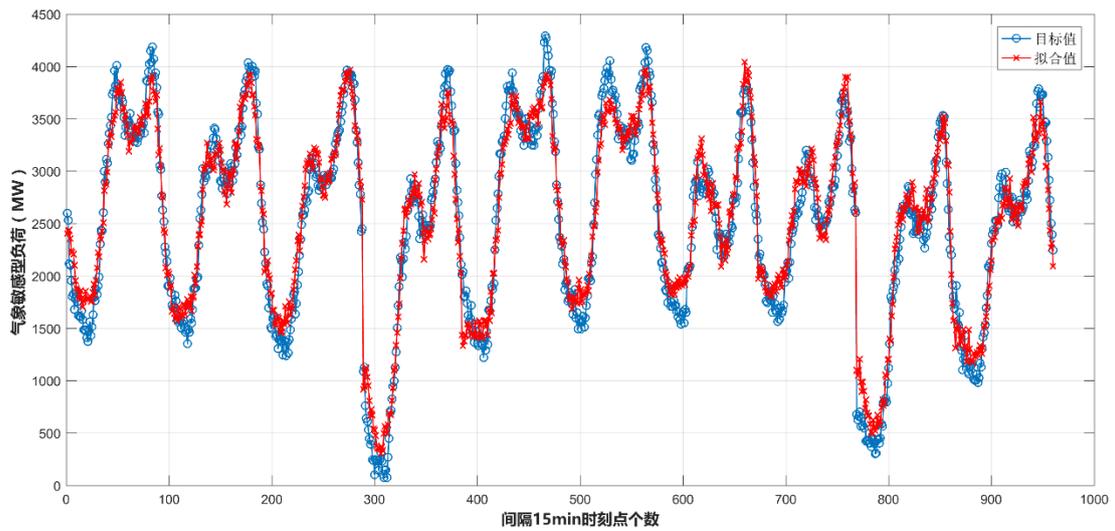


图 6.4 SVR_wh 模型训练集拟合结果

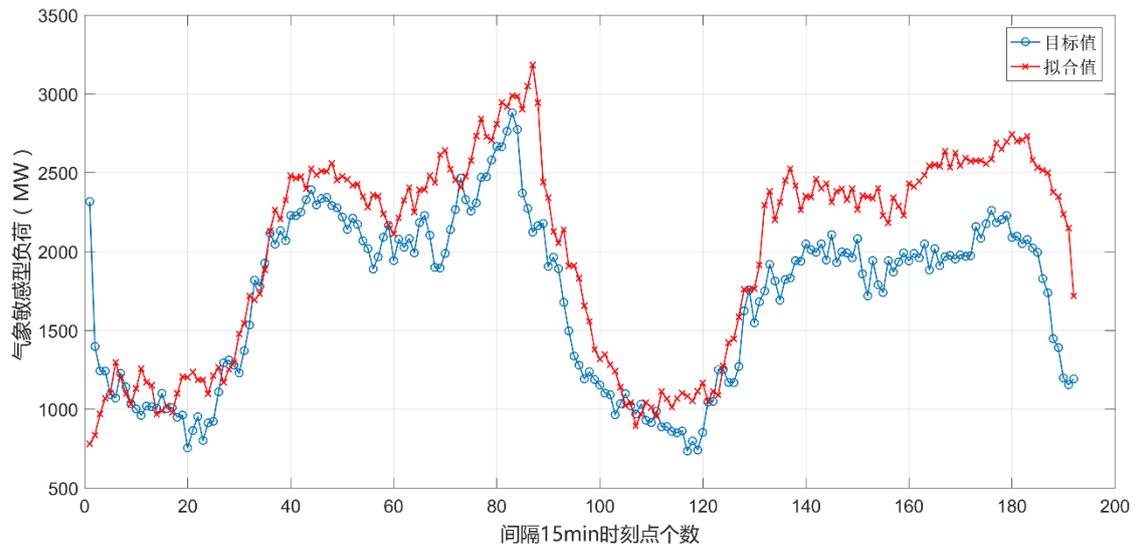


图 6.5 SVR_wh 测试集拟合结果

6.2.4 SVR_wc 模型性能评估

SVR_wc 模型的建立方法类似于 SVR_wh。由第三章对供冷季单一气象因素与电力负荷相关性分析中发现，温度，相对湿度和风速对负荷变化相关性较大。但通过采用体现这三个变量对人体热舒适性协同作用的有效温度分析，相关性并未有明显改善作用。因此在制冷季 SVR 模型中，为体现其对电力负荷潜在的非线性关系，将这三个气象参数均作为输入变量输入模型中。同时输入模型的变量有前四个时刻的负荷，以及预测时刻，相较于冬季模型，输入空间为 8 维。

选取 7 月 24 日-8 月 6 日中的十天工作日数据作为训练集，以 8 月 7 日和 8 月 8 日数据为测试集，相隔日期为剔除的节假日。同理通过网络寻优可得到针对上述样本模型的最优参数，惩罚因子 $C=2.48$ ，核函数宽度 $\gamma=2$ 。

训练集进行模型拟合时采用 5 折交叉验证的方法，整体来说该模型在训练集和测试集的表现均较佳。首先两者的 R^2 较高，具有较高的拟合程度。其次预测值和目标值的 RMSE 均较小，说明残差分布较为集中，且两者平均相对误差稳定在 5-6% 以内，说明该模型的泛化性较好。同样，将预测值与基础负荷相加作为该组合预测方法的预测值，同总体负荷实际值比较，其 MAPE 非常小，说明预测准确度很高。对比于供热季模型采用 6 维输入参数，供冷季输入参数增加到 8 维，最为明显的不同是训练时间变长了 15 秒。

表格 6.3 SVR_wc 模型性能汇总

评估参数	训练集	测试集	整体性能
R^2	0.957	0.931	
RMSE	353.91	464.82	464.82
MAE	292.58	362.31	362.31
MAPE	5.54%	5.91%	1.65%
训练时间	1min33sec		

图 6.6 和图 6.7 分别展示了训练集和测试集上模型的拟合程度和偏差情况，总体而言模型可以较好的预测出电力负荷的变化走势。在图 6.7 测试集数据中可以看出，预测值波峰波谷，或较小的波动滞后性相较供热季的模型并不明显，但对晚间高峰的预测结果不理想。

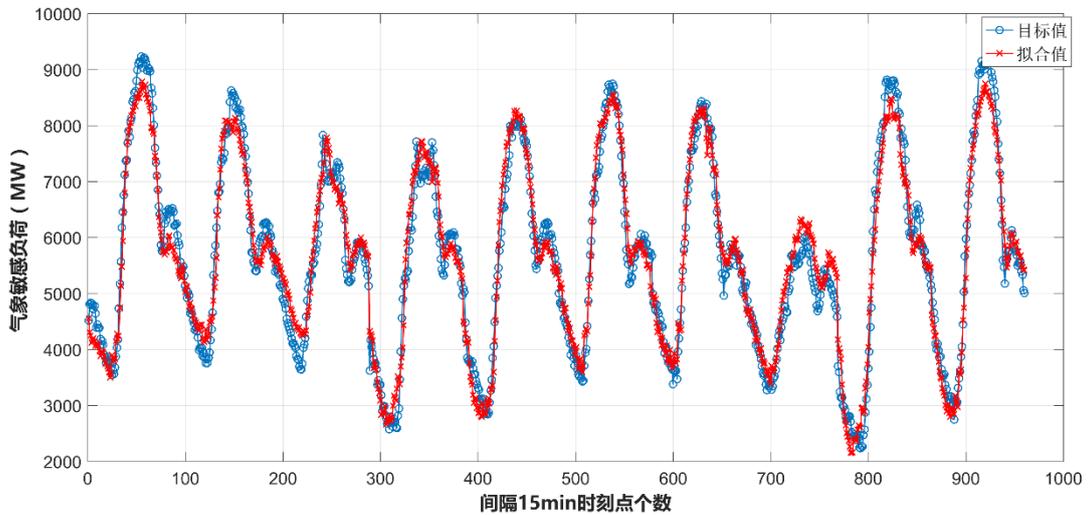


图 6.6 SVR_wc 模型训练集拟合结果

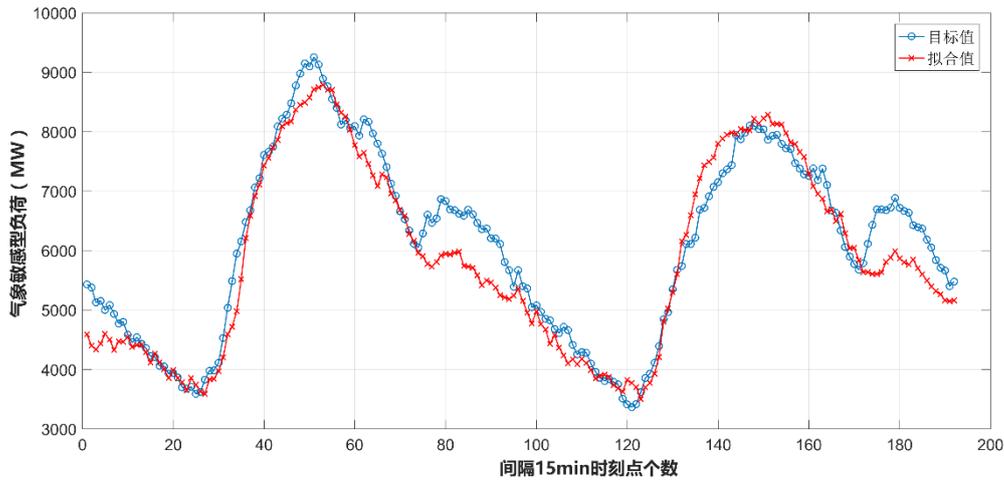


图 6.7 SVR_wc 测试集拟合结果

6.2.5 SVR_h 模型性能评估

节假日中针对气象敏感负荷建立的预测模型思路与上述完全相似,在此不再赘述。基于第四章对长假的分析可知,其电力负荷变化特性又与普通周末的变化有所区别。本小节将以春节长假为例,建立适用于该时段的 SVR_hh 模型,输入变量与 SVR_wh 模型相同,共 6 维。制冷季以普通周末为示例样本,建立 SVR_hc 模型,输入变量与 SVR_wc 模型相同,共 8 维。模型预测性能如表格 6.4 和表格 6.5 所示。

表格 6.4 特殊节假日 SVR_hh 模型性能汇总

评估参数	训练集	测试集	整体性能	
R^2	0.937	0.757	/	
RMSE	222.75	337.52		337.52
MAE	187.00	336.85		336.85
MAPE	8.17%	18.86%		3.98%
训练时间	1min09sec			

表格 6.5 普通周末 SVR_hc 模型性能汇总

评估参数	训练集	测试集	整体性能
R^2	0.959	0.798	
RMSE	383.05	558.81	558.81
MAE	326.67	442.54	442.54
MAPE	10.06%	14.19%	3.09%
训练时间	27sec		

横向对比供热季工作日预测模型，针对春节期间的预测模型性能较差。首先直观的表现测试集平均相对误差远远大于训练集相对误差，说明了该模型泛化性较差。其次测试集拟合度仅为 0.757，表明对预测负荷变化趋势把握能力不强。这主要由于春节相似日模型样本较少，且负荷变化特性较为特殊，推测模型对其学习能力不够充分。就预测结果与基础负荷相加并于总电力负荷比较，相对误差小于 4%，该误差范围可被接受。

供冷季普通周末的样本数量为 14 天，样本数量较少，且样本时间跨度较大。随着气象因素在此期间较大的波动变化，气象敏感型负荷变化的差异性也较大，下表所示是采用供冷季 5 个假期作为训练样本，预测下一个节假日的负荷。建立的模型性能如表格 6.5 所示。单就预测气象敏感型负荷，预测模型的误差较大，但在组合模型中，其预测精度可被接受。

整体而言，模型对于节假日负荷预测精准度显著低于对工作日负荷的预测。主要由于节假日样本较少，在同样选择 10 天数据作为训练样本时，其数据的时间跨度较大，造成了无论输入空间变量还是目标值的变化不连续，存在阶跃性变化。在减少训练样本数的情况下，模型训练时间明显缩短，但存在模型训练不充分的情况。反之，对于工作日负荷的预测，样本较多且连续，输入变量和目标值的连续性没有被破坏，因此训练模型性能较好。但就不同季节工作日模型和节假日模型而言，采用组合式方法对总电力负荷进行预测均获得较高精准度，平均相对误差较小，说明本文负荷拆分方法论较为合理。

6.3 电力负荷组合式模型评估

本课题负荷预测的主要思路是通过对季节属性的定义,将供热季和制冷季的负荷拆分成基本负荷和气象敏感负荷。通过对气象敏感型负荷部分的预测,累加已确定的基础负荷部分得到当日总电力负荷。以下章节将对比传统的负荷预测方法,即以当日总体负荷作为目标值训练模型进行预测,从而判断负荷拆分,并对部分预测,结果再组合方法论的优越性。

本课题方法论就电力负荷预测问题从暖通知识切入,进行了以下两个方面的改进,并就其改进分别与传统预测思路对比。

- i. 本课题通过对季节属性的定义规则,将全年数据清晰的划分为三组,并就不同分类建立不同预测模型。分类规则与预测方法清晰,且模型较为简单并具有良好的精准性。基于对电力负荷拆分的思路,对供热制冷季总电力负荷通过预测气象敏感型负荷再与基础负荷相加的方法得出。此处对比基准为不同季节属性下基于总电力负荷预测的模型。
- ii. 本课题模型输入参数中,天气因素考虑了对电力负荷作用的延迟性,采用了预测时刻前三个时刻的加权平均值。并在供热季模型中,引入有效温度降低输入变量的维度,达到简化模型的目的。此处对比基准为输入参数天气因素采用预测时刻实时值,基于气象敏感型负荷预测的模型。供热季模型输入变量用温度,相对湿度和风速替代有效温度。

6.3.1 组合模型与基于总电力负荷模型对比

本小节选取与上述供热季和制冷季相同日期的训练样本和测试样本进行模型训练。不同季节属性下,模型的输入参数与预测气象敏感型负荷模型相同,目标值采用总电力负荷。以模型对测试集的预测效果作为模型的评估指标。

供热季以1月7日-1月16日数据为训练集,预测1月17日和1月20日总电力负荷,模型输入参数与SVR_wh相同为6维,分别为滑动平均的有效温度,预测小时时刻以及预测时刻前四个时刻总电力负荷。基于相同样本容量训练出的模型性能对比如下表所示:

表格 6.6 供热季工作日不同建模方法预测总电力负荷性能对比

评估参数	总电力负荷 SVR_wc_1	气象敏感负荷 SVR_wc_2
R^2	0.933	
RMSE	740.83	408.90
MAE	664.88	332.54
MAPE	4.08%	1.91%
训练时间	1min48sec	1min33sec

通过两个模型的评估参数可知，基于气象敏感负荷的组合模型性能更佳。该模型在 RMSE 和 MAE 指标上均显著小于基于总电力负荷为目标值的 SVR 模型，表现出更好的精准性。其次组合模型可将总负荷的预测由 4.08% 缩小至 1.91%，准确度更佳。最后在对模型的训练方面，组合模型的训练时间略微短于基于总电力负荷的模型，这在实际实时预测中同样具有重要意义。但值得注意的是，基于总电力负荷的模型 R^2 拟合度较高，这是由于总负荷较强的周期变化特性。

下图所示为基于总电力负荷的预测模型对测试集 1 月 17 日和 1 月 20 日数据的预测。对比于图 6.5 更为直观的反映出该模型的预测偏差较大。

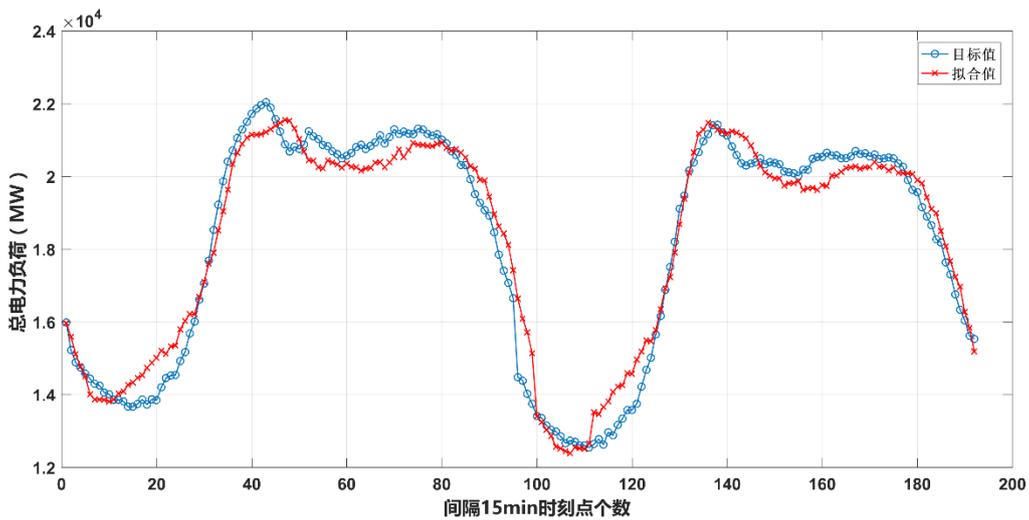


图 6.8 基于总电力负荷 SVR_wh 测试集拟合结果

制冷季仍以 7 月 24 日-8 月 6 日数据为训练集，预测 8 月 7 日和 8 月 8 日总

电力负荷，模型输入参数与 SVR_wc 相同为 8 维，分别为滑动平均的温度、相对湿度和风速，预测小时时刻以及预测时刻前四个时刻总电力负荷。基于相同样本容量训练出的模型性能对比如下表所示：

表格 6.7 制冷季工作日不同建模方法预测总电力负荷性能对比

评估参数	总电力负荷 SVR_wc_1	气象敏感负荷 SVR_wc_2
R^2	0.967	
RMSE	997.24	464.82
MAE	820.64	362.31
MAPE	3.66%	1.65%
训练时间	1min48sec	1min33sec

通过两个模型的评估参数可知，基于气象敏感负荷的组合模型性在制冷季的优越性更为显著。该模型在 RMSE 和 MAE 指标上均显著小于基于总电力负荷为目标值的 SVR 模型，表现出更好的精准性。其次组合模型可将总负荷的预测由 3.66% 缩小至 1.65%，准确度更佳。

下图所示为基于总电力负荷的预测模型对测试集 8 月 7 日和 8 月 8 日数据的预测。对比于图 6.7 更为直观的反映出该模型的预测在日间偏差很大，对于夜间总电力负荷的拟合度较高且偏差较小。可能原因是辐照度的影响。制冷季中辐照度的变化同时是影响电力负荷变化的重要因素，由于缺乏相关数据，因此在此不做考虑。但就相同输入参数情况下的组合式模型和基于总负荷预测的模型结果来看，组合式模型性能更优。

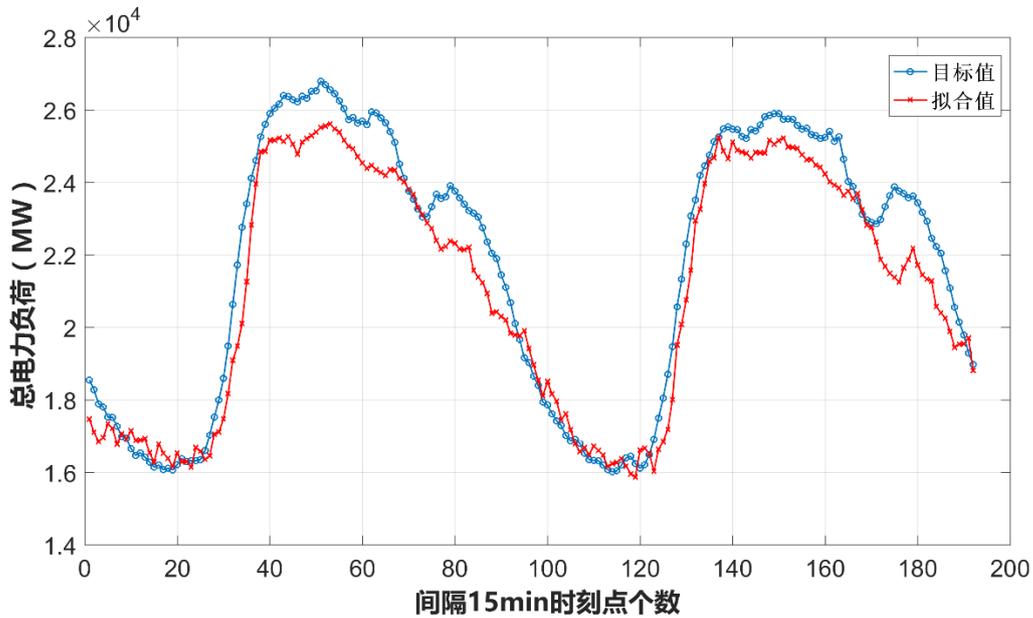


图 6.9 基于总电力负荷 SVR_wc 测试集拟合结果

同样建立春节期间基于总电力负荷的预测模型，并与组合模型性能进行对比，结果如下表所示。

表格 6.8 供热季特殊节日不同建模方法预测总电力负荷性能对比

评估参数	基于总电力负荷	组合模型
R^2	0.868	
RMSE	587.63	337.52
MAE	496.61	336.85
MAPE	4.30%	3.98%
训练时间	1min21sec	1min09sec

组合模型在 RMSE, MAE, MAPE 以及训练时间上均优于基于总电力负荷的预测模型。根据章节 4.2.3 中的分析，春季期间调整后的基础负荷无法表征出该时段较为特殊的负荷特性，即晚间高峰远远大于日间高峰的特点。在气象敏感型负荷预测模型拟合度不高的情况下，即同样无法准确展现出负荷变化趋势。因此组合

模型中两者相加的结果同样欠缺该特征。相反，在基于总电力负荷的模型中，拟合度 R^2 较高，且其 MAPE 为 4.3%，仅仅略低于组合模型的 3.98%，相对误差范围可被接受。综上所述，对于长假期样本较少且数据变化明显不同于其他日类型样本时，直接对周期性规律性变化较强的总电力负荷建立模型进行预测性能更佳。

相较而言，对于制冷季普通周末采用组合式模型对总电力负荷进行预测性能更佳，其 RMSE 和 MAE 指标远远小于直接以总电力负荷为目标函数的 SVR 模型，且相对误差为 3.09%，同样相较 4.36%有了明显的改善，如表格 6.9 所示。

表格 6.9 供冷季普通周末不同建模方法预测总电力负荷性能对比

评估参数	基于总电力负荷	组合模型
R^2	0.886	
RMSE	923.42	558.81
MAE	809.26	442.54
MAPE	4.36%	3.09%
训练时间	35sec	27sec

6.3.2 不同输入参数模型对比

供热季和制冷季模型输入气象参数均为预测时刻前三个小时的滑动平均值，采用该方法是考虑了其对电力负荷影响的延迟性，提高预测模型的准确性。如图 6.5 所示，供热季模型预测值波动变化相较目标值基本吻合。在制冷季模型中，两者负荷峰值和谷值的时间点拟合的较好，如图 6.7 所示。为考察该方法的有效性，选取制冷季 8 月份连续 10 天工作日作为训练集，后两天工作日为测试集，对气象敏感型负荷分别建立 SVR 模型进行对比，其中输入变量中的气象参数分别采用预测时刻实时值和滑动平均值。两预测模型的性能和测试集数据的拟合图如下所示。

表格 6.10 所示采用滑动平均的气象参数作为模型输入参数，其对气象敏感型负荷预测的精准度更高，且拟合度也更高。在得到总电力负荷的预测值时，平均相对误差为 1.32%，优于对比模型的 1.43%。图 6.10 更为直观的反映出以滑动平均气象参数为输入变量的模型对预测值的拟合度更好。

表格 6.10 供冷季工作日不同输入参数模型性能汇总

评估参数	实时气象参数输入模型		滑动平均气象参数输入模型	
	测试集	整体性能	测试集	整体性能
R^2	0.924	/	0.964	/
RMSE	373.74	373.74	326.29	326.29
MAE	295.51	295.51	261.77	261.77
MAPE	5.27%	1.43%	5.14%	1.32%
训练时间	1min41sec		1min41sec	

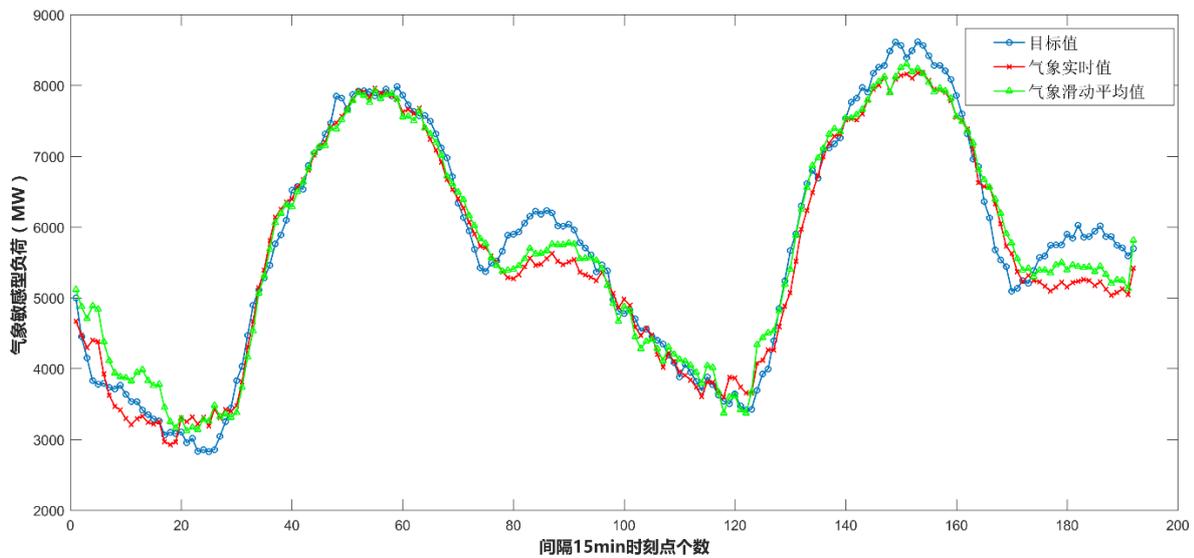


图 6.10 不同输入参数 SVR_wc 模型测试集拟合结果对比

选取不同于章节 6.2.4 的训练和测试样本建模,可同时验证针对不同情况下模型的稳定性。供热季对不同输入参数的模型对比方法与制冷季相同,对比结果为以滑动平均的有效温度作为输入变量,其模型的预测精准度略优于其对照模型。证明了本课题对预测模型输入参数的改进可行。

6.4 本章小结

本章节着重分析了基于支持向量机的针对气象敏感型负荷的建模方法,并评估了预测模型的性能。通过预测结果与该时段基础负荷相加,间接得到对总电力负荷的预测。继而通过与传统的负荷预测思路对比,即直接基于总电力负荷建立预测模型,得出本课题提出的组合式预测方法工作日和普通借节日具有较高的精准度的结论。在对特殊长假,如春节和国庆节的负荷进行与测试,虽然组合算法的精准度仍高于传统方法,但其拟合度较低,无法表征特殊长假期间特殊的负荷变化特点。因此判定该时段内基于总电力负荷建立的预测模型效果更佳。此外,通过改变模型的输入参数,得出采用滑动平均气象参数的模型预测精准度更高且拟合度更好的结论。

本章节给出的模型示例基本涵盖了全年的不同情况,分别就供热季、制冷季、工作日、普通周末和特殊长假等不同情况建立了模型,并对其组合预测算法的优越性进行了较为详尽的分析,结论夯实。

第7章 总结与展望

本章节主要对全文所得到的重要结论进行总结,同时就研究中出现的不足予以简要分析,并提出作者对此的研究展望。

7.1 本文总结

如何提高短时电力负荷的预测精度一直是海内外学者致力研究的重要课题,本文深入分析了城区总电力负荷的变化特性后,提出了负荷先拆分,后预测,再组合的预测思路。通过建模实例证明,该方法对全年出现的所有负荷变化情况,即不同季节工作日,长短节假日均具有较高精度的适用性。本文在搭建该预测方法论时主要结论如下:

- i. 在不同时间观测尺度上考察城区电力负荷的变化特性。分别考察日、周、月周期上的电力负荷时序图,从而得出负荷变化具有连续性,周期性和不同时段变化特征不同的结论。并就特性的不同,总结了影响电力负荷的主要因素;
- ii. 着重考察不同季节,不同时间观测尺度上单一气象因素以及多个气象因素耦合组成的综合气象参数与电力负荷的相关关系。得出不同季节周期内,影响电力负荷变化的主要气象因素不同。制冷季负荷主要与气温和相对湿度相关性较大,供暖季气温和风速为影响电力负荷为主导因素,过渡季电力负荷总体上和气候因素相关性不大。采用综合气象参数时,供热季与负荷的相关关系明显增强,制冷季和过渡季改善效果不明显;
- iii. 建立负荷拆分算法,可有效地将供热季和制冷季电力负荷拆分成基础负荷和气象敏感型负荷。借鉴暖通空调的度日数法,通过建立决策树,细化逐时季节属性的判断规则。并通过分时段加权平均的方法判断当日的季节属性。基于结论 ii,可将季节属性为过渡季节的电力负荷判定为基础负荷,供热季和制冷季期间的基础负荷则为同一年内过往过渡季的加权平均。进而通过与总负荷的差分,即可确定供热季和制冷季的气象敏感型负荷。通过计算该拆分算法得到的气象敏感负荷占总体负荷的比例可知,供热季节所占比例居于 6%-15%左右,制冷季所占比例在 10%-26%

之间，较为符合工程经验；

- iv. 根据过渡季和不完全过渡季基础负荷变化较为平稳且周期性强的特点，采用时间序列法建立自回归模型进行负荷预测。将分别对工作日和节假日建立实时预测模型，通过实例可知，模型预测精度较高，小于 2.0%；
- v. 对于供热季和制冷季气象敏感型负荷的预测，采用非线性关系逼近较强的支持向量回归算法。结合结论 ii，不同季节属性电力负荷的预测模型输入参数选取不同。供热季模型的输入参数包括四个时刻平均的有效温度，四个时刻的历史负荷以及预测时刻的时刻点，共计 6 维。而制冷季模型的输入参数则包括四个时刻平均的温度，湿度，风速，四个时刻的历史负荷以及预测时刻的时刻点，共计 8 维。预测的总电力负荷通过模型预测出的气象敏感型负荷与当日基础负荷相加得出。将该方法与传统预测算法，即以总电力负荷作为目标值建立模型进行预测相比，预测精度有了很大的提高。其中对工作日电力负荷的预测误差可由 3%-5%提高到 2%以内，对节假日电力负荷的预测误差可由 4%-5%提高到 4%以内。

本文在得到上述结论时主要的创新点如下：

- i. 天气参数与电力负荷相关性分析中，引入反映人体舒适度的综合气象参数，采用线性相关分析和灰色关联度相结合的方法，考察多个气象因素对电力负荷的耦合作用，达到降低模型输入维度的目的；
- ii. 搭建了先拆分，后预测，再组合的组合式预测算法。通过细化季节属性判定规则，从负荷构成的角度拆分总电力数据为基础负荷和气象敏感型负荷。对基础负荷和气象敏感型负荷分别采用时间序列法和支持向量回归机算法进行预测，预测结果相加得到总电力负荷的估计。

7.2 存在不足与展望

本文完成了从电力负荷数据和气象数据的收集，缺失值和离群值检验，电力负荷分类与拆分，到预测模型建立这样一套完整的短时电力负荷预测流程。每一个环节均在 Matlab 环境下编程完成，但并没有完成用户接口程序的编写。对于非专业人员，在缺少人机交互图形化的用户界面时，会对上述预测算法的应用感到生涩。在后续的推广开发中，可基于本课题的预测思路建立用户界面。

本文在对特殊节假日，如春节和国庆节的基础负荷进行处理时，采用的是根

据总负荷线性同比例增大或缩小的方法。但该方法得出的基础负荷无法准确表征总体负荷的变化特性。因此在后续的研究和应用中，若有历年节假日负荷数据，可在消除宏观经济因素的基础上，增加样本数量，对特殊节假日负荷进行更为精准地预测。

此外，本文对天气因素突变引起的电力负荷变化特性的改变并未进行探讨。采用规律变化的历史负荷数据作为训练集训练模型，对该日进行预测时将会引起较大误差。在后续的研究中，应搜集历年数据中天气因素突变的变相似日数据，可研究气象参数变化幅度与电力负荷变化特性的关系，从而可对预测结果进行修正。

天气参数的辐照度不仅具有季节性变化，同时也是影响人们用电行为的重要因素。由于本文缺乏辐照度数据的来源，并未考察该参数与电力负荷的相关性。在后续的研究中，可对此进行补充分析。

大数据时代对数据内在规律的深入挖掘，以及人工智能下利用数据驱动模型对未来发展趋势做出预测均是未来发展的方向。在此背景下，本文通过建筑能耗等专业背景知识，采用数据挖掘等一系列算法开创了针对城区短时电力负荷的组合式算法，提高了其预测精度。希望对后续相关研究有一点微薄的贡献。

致谢

时光荏苒，转眼已在同济大学度过了我最美好的青春。回想当年带着懵懂与一身稚气走入同济校园，在经历了本科阶段的学习与成长后，又如愿在这里完成进一步深造。不觉间，我已然成为具有初步学术造诣的研究生。为此，我衷心的感谢导师许鹏教授对我学术上的指导和生活上的关怀与鼓励。

本文就是在许鹏导师的悉心指导下完成的，选题上就能体现出导师对学术前沿发展方向的敏锐嗅觉，对于课题瓶颈问题的突破同样凝聚着导师的学术思想。在课余闲暇，导师总是提出最新颖的学术成果或工程难题让我们头脑风暴，在不断的讨论中开拓思路。他这种不断进取，对最新知识汲取的渴望是我所不具备的，也是非常值得我学习的地方。在生活中，许鹏导师亦师亦友，对每一位学生像孩子一样给予指导和点拨，他的鼓励让我们在踌躇不前时有了勇往直前的勇气。导师总会同我们谈谈理想聊聊生活情怀，偶然发现，有时自己确实闷头向前忘记了抬头看看方向，理想这个东西，是人生切勿丢失的大方向。三年的耳濡目染，许鹏导师这位儒雅学士教会了我很多。

此外，内心中我由衷的感谢同济大学暖通所每一位老师，本科和研究生期间严谨的教学，课下平易近人的关怀都让我对同济大学爱的更深沉。还有我可爱的同门们，有你们的陪伴稍显苦涩的学术研究增添了道道亮色。不管是学术讨论中争论的面红耳赤还是生活中的嬉笑怒骂，都是我这段美好经历无法遗忘的情怀。

最后，我最想感谢还有我的父母以及不服输不放弃的自己。父母的希冀是让我成为一个具有大格局的有抱负的人，他们对我的无私支持，对我无微不至的照顾以及对我点点滴滴的鼓励都是我继续前进的动力。我无以为报，唯有成为有用的人反哺爱护他们。还有那个有点不服输的我，希望你一直带着这种拼劲在以后的道路越战越勇。

特此鸣谢上海市经济与信息委员会肖明卫老师对本课题提供的技术支持。同时感谢国家科技支撑计划 2015BAL04B00 对本课题研究给予的资金支持。

李梦西

2017年4月11日 于同济大学

参考文献

- [1]. 文福拴与A.K. David, 加州电力市场失败的教训. 电力系统自动化, 2001. 25(5): 第1-5页.
- [2]. 吴杰康, 电力市场中生产成本与发电竞标策略研究, 2002, 华南理工大学.
- [3]. Kyriakides, E. and M. Polycarpou, Short term electric load forecasting: A tutorial, in Trends in Neural Computation, K. Chen and L. Wang, K. Chen and L. Wang^Editors. 2007, Spring. p. 391-418.
- [4]. Tzafestas, S. and E. Tzafestas, Computational Intelligence Techniques for Short-Term Electric Load Forecasting. 2001. 31(1): p. 7 - 68.
- [5]. Hahn, H., S. Meyer-Nieberg and S. Pickl, Electric load forecasting methods: Tools for decision making. 2009. 199(3): p. 902 - 907.
- [6]. 马莉等, 国外电力市场最新发展动向及其启示. 电力系统自动化, 2014. 0(13): 第1-9页.
- [7]. Almeshaie, E. and H. Soltan, A methodology for Electric Power Load Forecasting. Alexandria Engineering Journal, 2011. 50(2): p. 137-144.
- [8]. Papalexopoulos, A.D. and T.C. Hesterberg, A Regression Based Approach to Short Term Load Forecasting, in Transactions on Power System. p. 40-45.
- [9]. Box, G.E.P., G.M. Jenkins and G.C. Reinsel, Time_series_analysis_forecasting_and_con. 1990: A John Wiley & Sons, INC.
- [10]. Brockwell, P.J. and R.A. Davis, Time Series: Theory and Methods; Time Series: Theory and Methods. Springer Series in Statistics. 1991.
- [11]. Villalba, S.A. and C.Á. Bel, Hybrid Demand Model for Load Estimation and Short Term Load Forecasting in Distribution Electric Systems, in Transactions on Power Delivery. 2000, IEEE. p. 764-769.
- [12]. Al-Hamadi, H.M. and S.A. Soliman, Short-term electric load forecasting based on Kalman filtering algorithm with moving window weather and load model. Electric Power Systems Research, 2004. 68(1): p. 47-59.
- [13]. Huang, S.J. and K.R. Shih, Short-Term Load Forecasting Via ARMA Model Identification Including Non-Gaussian Process Considerations, in System, Transactions On Power . 2003, IEEE.
- [14]. Eberhart, R. and J. Kennedy, A New Optimizer Using Particle Swarm Theory, in Micro Machine and Human Science. 1995, IEEE.
- [15]. Hagan, M.T. and S.M. Behr, The time series approach to short-term load forecasting, in Transactions on Power Systems PWRS-2. 1989. p. 785-791.
- [16]. 雷绍兰, 基于电力负荷时间序列混沌特性的短期负荷预测方法研究, 2005, 重庆大学. 第111页.
- [17]. Hahn, H., S. Meyer-Nieberg and S. Pickl, Electric load forecasting methods: Tools for decision making. 2009. 199(3): p. 902 - 907.
- [18]. Feinberg, E.A. and D. Genethliou, Neural Networks for Short-Term Load Forecasting: A Review and Evaluation, in Applied Mathematics for Restructured Electric Power Systems, J.H. Chow, F.F. Wu and J.A. Momoh, J.H. Chow, F.F. Wu and J.A. Momoh^Editors. 2005. p. 269-285.

- [19]. A Review of ANN-based Short-Term Load Forecasting Models.
- [20]. Hippert, H.S., D.W. Bunn and R.C. Souza, Large neural networks for electricity load forecasting: Are they overfitted? *International Journal of Forecasting*, 2005. 21(3): p. 425-434.
- [21]. 基于遗传神经网络的短期负荷预测研究.
- [22]. Huang, S.C. and Y.F. Huang, Bounds on the Number of Hidden Neurons in Multilayer Perceptrons, in *Transactions on Neural Networks*. 1991, IEEE.
- [23]. Liu, Y.Y., J.A. Starzyk and Z. Zhu, OPTIMIZING NUMBER OF HIDDEN NEURONS IN NEURAL NETWORKS.
- [24]. Abu-Et-Magd, M.A. and R.D. Findla, A new approach using artificial neural network and time series models for short-term load forecasting. 2003.
- [25]. 陈耀武, 汪乐宇与龙洪玉, 基于组合式神经网络的短期电力负荷预测模型. *中国电机工程学报*, 2001. 21: 第79-85页.
- [26]. Vapnik, V.N., *The Nature of Statistical Learning Theory; The Nature of Statistical Learning Theory*. 1995.
- [27]. Sapankevych, N.I. and R. Sankar, Time Series Prediction Using Support Vector Machines A Survey, in *Computational Intelligence Magazine*. 2009, IEEE. p. 24-38.
- [28]. Zhang, M.G., Short-term load forecasting based on support vector machine regression, in 4th Int. Conf. on Machine Learning and Cybernetics. 2005. p. 4310-4314.
- [29]. Espinoza, M., J.A.K. Suykens and B. De Moor, Load Forecasting Using Fixed-Size Least Squares Support Vector Machines. 2005, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 1018 - 1026.
- [30]. Pai, P. and W. Hong, Support vector machines with simulated annealing algorithms in electricity load forecasting. *Energy Conversion and Management*, 2005. 46(17): p. 2669-2688.
- [31]. Hsu, C., et al., Dynamically Optimizing Parameters in Support Vector Regression: An Application of Electricity Load Forecasting. 2006, IEEE. p. 30c.
- [32]. 泰崔锦与正兴程, *小波分析导论*. 2003: 西安交通大学出版社. 45-49.
- [33]. 邵能灵与侯志俭, 小波模糊神经网络在电力系统短期负荷预测中的应用. *中国电机工程学报*, 2004. 24(1): 第25-35页.
- [34]. 宋人杰, 边奕心与闫淼, 基于小波系数和BP神经网络的电力系统短期负荷预测. *电力系统保护与控制*, 2009. 37(15): 第87-93页.
- [35]. 燕平张与铃张, *机器学习理论与算法*. 2012: 北京科学出版社. 121-123.
- [36]. Dietterich, T.G., *Ensemble Methods in Mashine Learning*.
- [37]. Fan, C. and S.W. Wang, Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. *Applied Energy*, 2014. 127: p. 1-10.
- [38]. 王鹏, 邵能灵与王波, 针对气象因素的短期负荷预测修正方法. *电力系统自动化*, 2008. 32(13): 第92-98页.
- [39]. 谢宏, 程浩忠与张国立, 基于粗糙集理论建立短期电力负荷神经网络预测模型. *中国电机工程学报*, 2003. 23(11): 第1-6页.
- [40]. 林辉, 刘晶与郝志峰, 基于相似日负荷修正的节假日短期负荷预测. *电力系统保护与控制*, 2010. 38(7): 第47-54页.
- [41]. Ko, C.N. and C.M. Lee, short-term load forecasting using SVR-based radial basis function neural

- network with dual extended Kalman filter. *Energy*, 2012(49): p. 413-422.
- [42]. Bahrami, S., R.A. Hooshmand and M. Parastegari, short term electric load forecasting by wavelet transform and grey model improved by pso algorithm. *Energy*, 2014(72): p. 434-442.
- [43]. Hooshmand, R.A., H. Amooshah and M. Parastegari, A hybrid intelligent algorithm based short-term load forecasting approach. *Electrical Power and Energy Systems*, 2013(45): p. 313-324.
- [44]. 中华人民共和国国家统计局, 中国统计年鉴2016, 2016.
- [45]. 张伟, 郗志与丁沃沃, 室外热舒适性指标的研究进展. *环境与健康杂志*, 2015(09): 第836-841页.
- [46]. 赵子健等, 基于标准有效温度和不舒适指标研究南京热舒适状况. *气象与环境科学*, 2013(04): 第16-21页.
- [47]. Dunham, M.H., 数据挖掘教程.第 1 版. 2005: 清华大学出版社.
- [48]. 罗雨滋与付兴宏, 数据挖掘ID3决策树分类算法及其改进算法. *计算机系统应用*, 2013. 22(10): 第136-139页.
- [49]. 何大四, 暖通空调中气象数据相关的几个问题探讨, 2006, 同济大学. 第 127页.
- [50]. 王沁, 时间序列分析及其应用. 2008: 西南交通大学出版社.
- [51]. Dickey, D.A. and W.A. Fuller, Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association*, 1979. 74(366): p. 427-431.
- [52]. Aho, K., D. Derryberry and T. Peterson, Model selection for ecologists: the worldviews of AIC and BIC. *Ecology*, 2014. 95: p. 631-636.
- [53]. Akaike, H., A new look at the statistical model identification, in *Transactions on Automatic Control*. 1974, IEEE. p. 716-723.
- [54]. 王文剑与门昌骞, 支持向量机建模及应用: 科学出版社 北京.
- [55]. Vapnik, V.N., *The Nature of Statistical Learning Theory; The Nature of Statistical Learning Theory. Statistics for Engineering and Information Science*. 2000.
- [56]. 王平与王文剑, 基于时序核函数的支持向量回归机. *计算机辅助工程*, 2006(03): 第35-38页.
- [57]. Cherkassky, V. and Y. Ma, Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks*, 2004. 17(1): p. 113-126

个人简历

个人简历:

李梦西，女，1990年12月生。

2013年7月毕业于同济大学建筑环境与设备工程(暖通空调)专业获学士学位。

2013年9月入同济大学动力工程专业，读硕士研究生。

2016年10月毕业于德国波鸿鲁尔大学能源与动力工程专业获理学硕士。