



声 明

本人郑重声明：此处所提交的硕士学位论文《基于局域波的地区负荷分析及其短期负荷预测研究》，是本人在华北电力大学攻读硕士学位期间，在导师指导下进行的研究工作和取得的研究成果。据本人所知，除了文中特别加以标注和致谢之处外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得华北电力大学或其他教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名：陈凤培 日期：2009.12.17

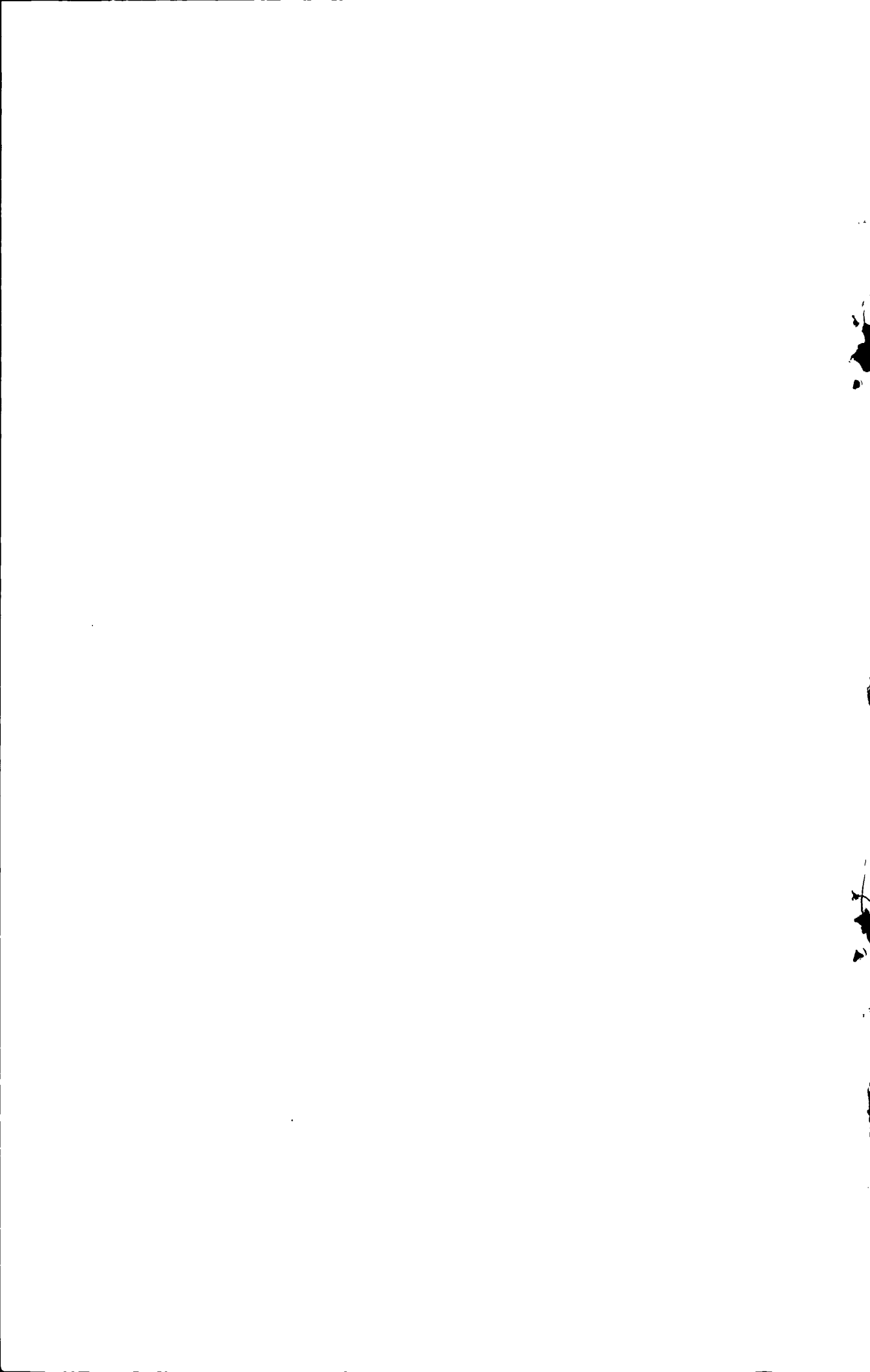
关于学位论文使用授权的说明

本人完全了解华北电力大学有关保留、使用学位论文的规定，即：①学校有权保留、并向有关部门送交学位论文的原件与复印件；②学校可以采用影印、缩印或其它复制手段复制并保存学位论文；③学校可允许学位论文被查阅或借阅；④学校可以学术交流为目的，复制赠送和交换学位论文；⑤同意学校可以用不同方式在不同媒体上发表、传播学位论文的全部或部分内容。

(涉密的学位论文在解密后遵守此规定)

作者签名：陈凤培
日期：2009.12.17

导师签名：栗然
日期：2009.12.26



摘 要

本文提出基于局域波的地区负荷分析及其短期负荷预测新方法。首先利用局域波和近似熵理论深入分析地区负荷特性,将负荷序列局域波分解,得到能反映负荷组成的分量及余量;对分量进行 Hilbert 变换得到 Hilbert 时频谱和功率谱,从而分析出各分量对总负荷的波动贡献;再将近似熵作为各分量及余量的特征参数,对比各分量和实际不同类型负荷的近似熵值,研究各分量的物理含义;采用实时气象因素对各负荷分量每天不同时段进行精细化分析,挖掘不同因素对各分量的影响状况。研究了电气化铁路的负荷特性及影响因素,提出基于灾变遗传算法和时序的 LS-SVM 的预测新方法,该法以时间序列模型与影响因素分析为据确定输入变量,利用灾变遗传优化算法得到模型的最优参数,建立 LS-SVM 预测模型。最后选择适合各负荷分量及余量的预测模型,然后进行合理重构,得出最后负荷预测结果。

关键词:电力系统,地区负荷分析,短期负荷预测,局域波

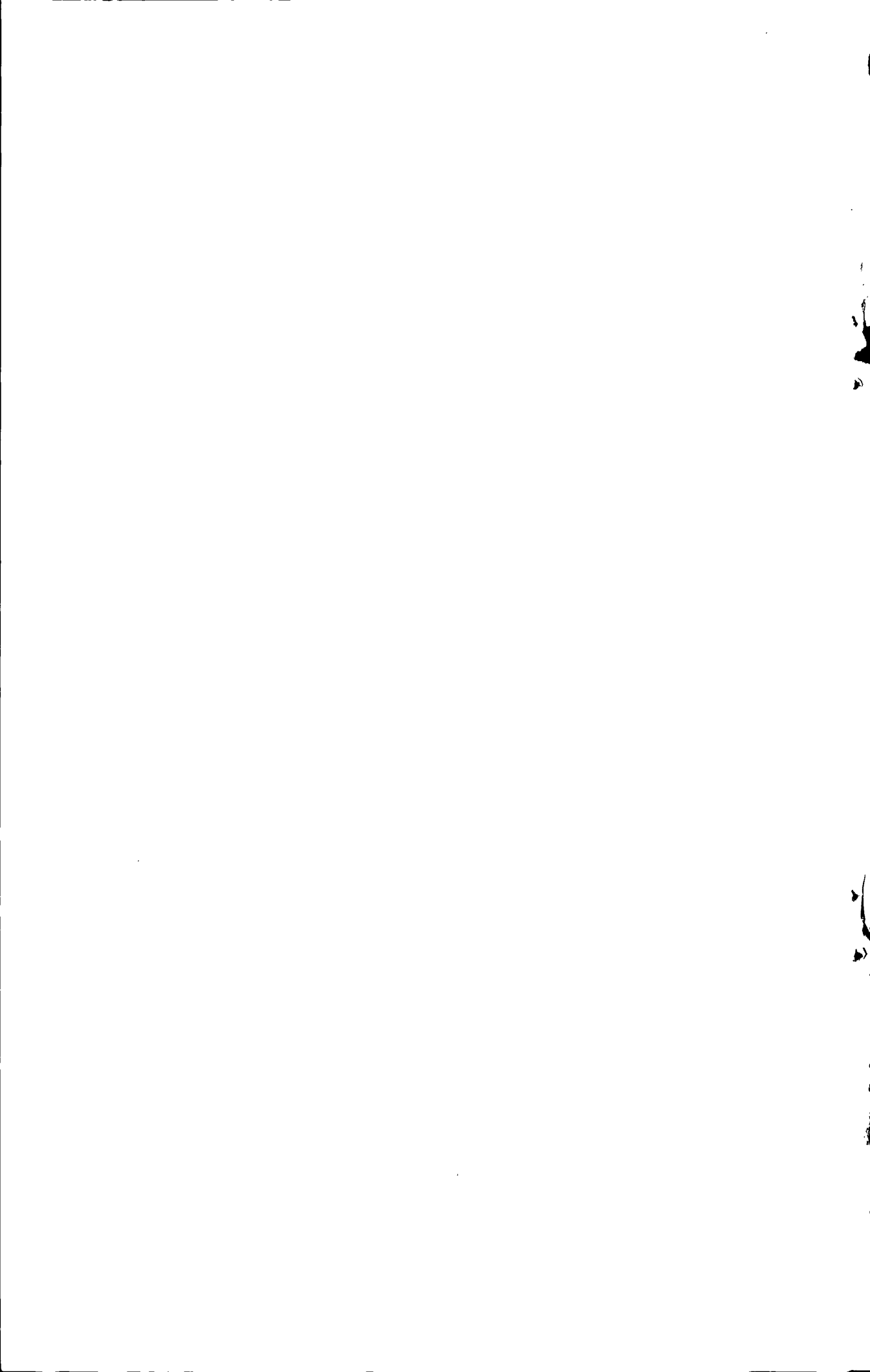
ABSTRACT

This paper presents area load analysis and short-term load forecasting based on Local Wave. Firstly, load characteristics are analyzed by using Local Wave and ApEn. Load is decomposed by local wave. Components reflecting load composition are obtained. Secondly, components are carried on Hilbert transformation obtaining Hilbert time-frequency and power spectrum. Next, ApEn is taken as characteristic parameter, then a comparison is made between components and actual load to learn the components' physical meaning. Then, real-time climatic factor is used to analyze load components of different time everyday, and influence of factors can be evaluated. Others, a forecasting method based on Cataclysmic GA and time-series LS-SVM is proposed. LS-SVM input variables are determined by influencing factor analysis and time-series model. It optimizes parameters using Cataclysmic GA, and establishes LS-SVM model. Finally, appropriate forecasting model for components and remainder are selected and reconstructed, and acquire final forecasting results.

Lu Fengyi(Electric Power System and its Automation)

Directed by Prof. Li Ran

KEY WORDS: power system, area load analysis, short-term load forecasting, local wave



目 录

中文摘要	
英文摘要	
第一章 绪论	1
1.1 课题研究意义及背景	1
1.2 短期负荷预测研究动向	2
1.2.1 基于全局模型预测方法	2
1.2.2 基于局部模型预测方法	3
1.3 目前电力系统短期负荷预测存在的问题	4
1.4 本文主要研究内容	5
第二章 局域波理论	7
2.1 局域波理论的起源	7
2.2 局域波理论与方法及其发展状况	8
2.2.1 瞬时频率的物理意义和局域波法	8
2.2.2 局域波分解算法的基本思想	10
2.2.3 局域波分解的几种算法	11
2.2.4 局域波分解的边界处理方法	12
2.2.5 局域波分解中的终止筛选标准	14
2.3 局域波时频谱	15
2.4 局域波法的工程应用现状	16
第三章 地区负荷特性分析	19
3.1 近似熵原理	19
3.2 实例分析	21
3.2.1 数据来源	21
3.2.2 负荷局域波分解	23
3.2.3 局域波时频分析	24
3.2.4 近似熵分析	25
3.2.5 实时气象因素影响分析	27
3.2.6 负荷分析	28
3.3 本章小结	29

第四章 含电铁负荷的地区负荷预测模型	31
4.1 电铁负荷特性及其影响因素分析	31
4.2 分量一预测模型	33
4.2.1 最小二乘支持向量机	33
4.2.2 时间序列	35
4.2.3 灾变遗传算法理论	36
4.2.3.1 遗传算法	36
4.2.3.2 灾变遗传算法	37
4.2.4 预测模型建立	38
4.2.4.1 数据来源	38
4.2.4.2 模型建立	39
4.2.5 预测结果分析	40
4.3 分量二预测模型	41
4.3.1 时间序列特征分析	41
4.3.2 模式识别	43
4.3.3 模型建立	43
4.4 分量三预测模型	43
4.5 分量四预测模型	44
4.5.1 多元线性回归理论	45
4.5.1.1 多元线性回归模型	45
4.5.1.2 未知参数估计	45
4.5.2 模型建立	47
4.6 余量预测模型	48
4.7 预测模型重构及模型评估	48
4.8 误差分析	50
4.9 本章小结	51
第五章 结论	52
参考文献	53
致 谢	59
攻读硕士学位期间发表的学术论文和参加科研情况	60

第一章 绪论

电力负荷预测是电力系统调度、用电、计划、规划等管理部门的重要工作之一。众所周知，电力系统有一个显著特点就是能量不能储存，随着电力的市场化，电力负荷预测成为一种亟需。提高负荷预测技术水平，有利于计划用电管理，有利于合理安排电网运行方式和机组检修计划，有利于节煤、节油和降低发电成本，有利于制定合理的电源建设规划，有利于提高电力系统的经济效益和社会效益。因此负荷预测已成为实现电力系统管理现代化的重要内容之一。

负荷预测作用的大小主要取决于精度，所以如何提高预测精度是目前研究的重点。如今我国的用电需求量大大增加，短期负荷预测已不再是省级或大地区供电局的工作，在拥有相对较小负荷的地级供电局也逐渐兴起并成为不可或缺的一部分。

1.1 课题研究意义及背景

短期负荷预测是指对未来一天或几天的负荷数据做出估计，它是电力系统调度运营部门的一项重要日常工作，是制定发电计划和输电方案的主要依据。短期负荷预测意义重大，主要体现在：

- (1) 帮助确定燃料供应计划；
- (2) 对运行中的电厂出力要求提出预告，使对发电机组出力变化事先得以估计；
- (3) 可以经济合理地安排本网内各机组的启停，降低旋转储备容量；
- (4) 可以在保证正常用电的情况下合理安排机组检修计划。

正因为短期负荷预测的重要性，因此有关短期负荷预测的研究一直是学术界热门研究课题，至今国内外的许多专家、学者在预测理论和方法方面做了大量的研究工作，取得了许多卓有成效的进展。但是，这些负荷预测的研究主要是针对电网容量较大，负荷变化比较平稳，规律性较强的地区，而对本课题所提出的地区级的短期负荷预测研究甚少。地区级负荷的主要特点是：

- (1) 电网的基础负荷相对较小，电网所能提供的负荷有限。某些地区的最大负荷和平均负荷是大负荷地区的十分之一甚至几十分之一；
- (2) 具有相当比重的冲击负荷。由于经济发展，地区积极发展地区经济，因此它们的日负荷中包括电气化铁路负荷（本文简称电铁负荷）、电化工业负荷、钢铁负荷等，这类负荷呈现出无规律的冲击变化，属于冲击负荷；
- (3) 各影响因素对负荷波动影响较大。由于它的负荷基数较小，因此各种影响因素在负荷波动上体现得更加明显；
- (4) 负荷曲线波动大，而且毛刺多。从地区级的日负荷曲线上可以发现，日负

荷的峰值与谷值相差很大，有的峰值可以是谷值的几倍，而且曲线出现许多无规律的毛刺。

随着科学的进步，经济的发展，地区负荷结构日趋复杂，负荷成分从单一化逐渐变得多样化，新类型的负荷，如电铁负荷占网内用电量越来越大，对负荷波动的影响也越来越明显。并且不同的影响因素，如气象因素中的温度、湿度、风速等对负荷的影响规律也不同。由于地区负荷的特殊性与复杂性，使得其负荷预测精度一直不是很高，而这类地区在国内大量存在，因此如果能够提高该类地区的负荷预测精度，那么对于国内负荷预测水平和精度的提高具有很大的作用。

1.2 短期负荷预测研究动向

短期负荷预测是电力系统领域一个传统的研究问题，预测的技术方法是负荷预测的核心问题，而预测精度是检验预测方法可行性的最重要的指标。对预测方法的研究和探索，一直以来是国内外学者关注的一个热点，各种预测方法层出不穷。总体上分为基于全局模型预测方法和基于局部模型的预测方法。

1.2.1 基于全局模型预测方法

基于全局模型的预测方法将所有的数据用于预测模型的建立，它包括线性预测模型和非线性预测模型。

(1) 线性预测模型

传统的预测方法主要采用线性预测模型，如自回归模型 (AR)^[1]、移动平均模型 (MA)、自回归移动平均模型 (ARMA)、混合自回归移动平均模型 (ARIMA)、Kalman 滤波器^[2]、递推最小二乘自适应滤波器 (RLS) 等。

(2) 非线性预测模型

而现实中的负荷序列一般是非线性和非平稳的，所以传统的线性模型显然不太适合进行预测，需要采用非线性来进行时间预测模型。文献[3]提出采用神经网络进行短期负荷预测，并用鱼群算法训练神经网络的权值，实验表明，该方法与传统的 BP 神经网络预测方法相比具有较强的自适应能力和较好的预测效果。文献[4]给出基于神经网络的短期负荷组合预测模型，该模型利用多种方法的预测结果与实际负荷数据的非线性关系，建立相应的神经网络模型，实验表明该方法具有很高的预测精度。文献[5]提出了一种将混沌和神经网络相结合的方法用于短期负荷预测，利用混沌方法重构相空间系统吸引子，用前向神经网络拟合吸引子上的全局整体映射，构成混合预测模型。文献[6]根据在相空间重构拓扑近邻点的时间演化原理，提出了优化近邻点 (ONP) 的短期负荷预测法，在此基础上，又提出 ONP 与递归性时延神经网络模型融合的 STLTF 法，这是一种能够对预测点跟踪智能辨识动态行为

的模型。文献[7]提出了一种基于粗糙集理论对各种影响负荷预测的因素变量进行识别,以此确定预测模型的输入量,在此基础上通过属性约简和属性值约简获得推理规则集,再以这些推理规则构筑神经网络预测模型,并采用加动量项的BP学习算法对网络进行优化。文献[8]提出了一种免疫聚类径向基函数神经网络模型来预测电力系统短期负荷,它比传统的神经网络具有更高的预测精度和更短的训练时间。文献[9]为了克服传统神经网络预测方法在网络结构设计、学习算法和收敛效果等方面存在的缺陷,通过借鉴免疫网络调节与免疫规划,提出了一种基于人工免疫网络的短期负荷预测模型,在网络设计中,创造性地融入了免疫调节原理,利用免疫规划来进化网络结构。文献[10]提出了一种基于支持向量机理论的电力系统短期负荷预测方法,通过和多层BP神经网络进行比较的试验,结果证明了其在短期负荷预测中的有效性。文献[11]提出了一种基于支持向量机的船舶电力负荷短期预测方法,对某大型集装箱船舶在不同工况下的电力负荷数据,分别用径向基核函数的支持向量机方法、多层BP网络和RBF网络方法进行训练和预测计算,取得较好效果。文献[12]将贝叶斯理论用于短期负荷预测中输入特征的自适应选取,同时建立了基于支持向量机在线学习的短期负荷预测模型,多个实际系统的预测算例表明了该方法在预测精度和预测速度方面的有效性。文献[13]利用支持向量机和遗传算法建立24个不同的混和模型来对夏季24点负荷进行滚动预测。文献[14]讨论了基于支持向量机的电力系统负荷预测模型建模方法,通过对模型结构的分析,提出了模型学习参数的选取方法,给出了基于最优方向搜索的高斯核函数参数优化算法。

1.2.2 基于局部模型预测方法

如今负荷结构日趋复杂,负荷成分从单一化逐渐变得多样化,一个变化复杂的非线性、非平稳信号难以使用单一的模型进行预测。因此人们开始进行基于局部模型的预测方法的研究。文献[15]指出基于局部模型的预测方法与基于全局模型的预测方法相比,具有更高的精度。时间序列预测的局部模型分为分解域预测模型和局部时间预测模型^[16]。

(1) 局部分解域预测模型

基于局部分解域模型的预测方法通过各种信号分解方法,如小波、Fourier分析等,将信号分解为一定频段的子序列,然后在各个子序列中应用各种建模手段进行预测建模。经过学者们的多年研究,已经研究了许多分解域预测模型。文献[17]提出了基于功率谱分解和实时气象因素的短期负荷预测方法,采用快速傅立叶变换对负荷序列进行变换得到功率谱,依据变换结果分析功率谱得出负荷基频、低频和高频分量的频率范围,采用有限脉冲响应滤波器从负荷中分离出各个负荷分量,分析各个分量的特点,针对各个分量分别设计预测模型。文献[18]采用小波变换对日负荷数据进行分解处理,使得数据信息相对集中,在此基础上将小波分量分解为受气

象因素影响的部分与不受气象因素影响的部分之和,对受气象因素影响部分与不受气象因素影响部分分别建立预测模型,提高了预测精度。文献[19]提出一种小波分析与神经网络相结合的预测方法,将历史电价和历史负荷都进行小波多分辨单尺度分解,分解成概貌电价、细节电价、概貌负荷、细节负荷,在此基础上,用历史概貌电价和概貌负荷预测训练 BP 神经网络,预测出未来的概貌电价,用历史细节电价和细节负荷序列训练 BP 神经网络,预测出未来的细节电价,将概貌电价和细节电价进行重构,得到最终的预测电价。文献[20]将多分辨率支持向量机用于短期负荷预测中节点负荷预测曲线的回归估计,其在保持曲线总体逼近能力的同时提高了局部区域的逼近能力。文献[21]提出了结合经验模式分解和支持向量机(SVM)的短期负荷预测法,运用 EMD 将负荷序列分解成若干个不同频率的平稳分量,分解后的分量突出了原负荷的局部特征,能更明显的看出原负荷序列的周期项、随机项和趋势项;根据各个分量的变化规律,选择合适的 SVM 参数和核函数构造不同的支持向量机分别预测,由 SVM 对各分量的预测值组合得到最终预测值。

(2) 局部时间域预测模型

基于局部时间域模型的预测方法通过模式分类和聚类等方法将时间序列数据划分成一系列局部区域,在每个区域中的数据具有相似的动力学特性,然后在每个局部区域采用各种建模手段进行预测建模。近年来,越来越多的人开始研究这种方法。文献[22]针对大电网的短期负荷预测,建立了按地域划分的网格化电力系统短期负荷预测模型,各子网根据自身的历史负荷和气象条件建立对网格更为有效的负荷预测模型,并采用了面向粗糙集的默认规则挖掘算法构成各自预测模型,从而获得更加准确的预测结果。文献[23]将负荷变化率数据、气象数据进行离散化,通过属性-值对的 2 次信息增益优化算法生成负荷预测树模型,在给出预测日气象及星期信息后可对特征负荷进行预测。文献[24]采用 SOM 聚类算法将数据划分为互不重叠的子区域,然后在各个子区域中采用多个 SVM 进行预测建模,其实验结果表明该方法不仅具有比单一 SVM 更高的预测精度,而且收敛速度更快、支持向量数目更少。

综上所述,短期负荷预测的局部建模是目前负荷预测研究领域的热点问题,关于这方面的研究还存在着诸多问题,需要进一步研究。

1.3 目前电力系统短期负荷预测存在的问题

如前所述,目前对于短期负荷的研究已经很多,也取得了很多研究成果,但仍旧存在一些尚未解决的问题,同时出现了新的问题,因此还有许多问题有待进一步研究。

(1) 新的负荷类型的出现,特别是电铁负荷占网内用电量越来越大,由于电铁

负荷变化剧烈，随机性较强，波动规律难以把握，因此为负荷预测带来了难度。为此，可以尝试对电铁负荷的研究，从它的变化规律，影响因素，对地区负荷的影响情况等入手，进一步提高负荷预测精度。

(2) 在电力系统目前的负荷预测研究中，预测方法应与实际情况相结合，尤其注重对气候条件，负荷构成等负荷的重要影响因素的细致研究。考虑到电力系统是一个多因素共同作用的系统，与相关因素相配合，找出影响预测精度的重要相关因素（或者称为主导因素）。这是一个有价值的发展方向。

(3) 影响因素是在目前经验基础上得到和划分的，随着认识的深入，可能还会再细分。

(4) 电力系统获得日天气气象数据将越来越详细，负荷预测可以根据不同气象数据提出相应的预测模型。

如果能从负荷构成入手，研究其变化规律及影响因素，然后在此基础上根据负荷的不同特点，运用适合各自自身特点的合理的预测方法进行预测，这样能够提高负荷预测的精度。

1.4 本文主要研究内容

本文重点研究基于局部波的地区负荷分析及其短期负荷预测，主要研究内容安排如下：

(1) 对某地区级负荷进行分层。

将某地区负荷分解成为若干个负荷分量，使得分解后所得的各层负荷分量能够突出原负荷的局部特征，并且各负荷分量包含了原负荷在不同时间尺度，不同频率上的局部特征。

(2) 负荷特征及其影响因素分析。

在对负荷进行合理分层的基础上，将各层负荷分量进行特征分析，对比它们与各种实际类型负荷的特征，找出各分量的物理意义。挖掘各种实时气象因素对各分量在不同时段的影响情况，为建立预测模型打下基础。

(3) 研究冲击负荷如电铁负荷的波动特性及影响因素。

研究了电铁负荷的波动特性，并分析了影响电铁负荷变化的主要因素，建立适合该类负荷特性的预测模型。

(4) 建立地区负荷预测模型。

在对各个负荷分量特性进行全面有效分析的基础上，根据各负荷分量的变化特点，影响因素以及影响规律的不同，以提高负荷预测精度为目标，探讨各自合适负荷分量的负荷预测模型，然后对各分量的负荷模型进行合理重构，得出最后负荷预测结果。

(5) 实例分析，评估模型的可行性。

本文主要以相对误差和平均相对误差为标准，分析模型的可行性。

第二章 局域波理论

目前人们已经提出了许多分解信号的方法,如傅立叶变换、Laplace 变换、Z 变换、Hartley 变换、小波变换、Chirplet 变换等。但现有的方法存在着如下几个缺点:

(1) 采用线性分解方法,对非线性序列进行分析时,破坏了信号本身的非线性性,产生了虚假频谱,影响对序列预测的精度;

(2) 某些分解出来的子序列不具物理意义,很难解释它的含义。

因此,应该采用非线性分解方法进行分解以保持序列本身的非线性,分解出来的子序列具有一定的物理意义,从而能够充分分析信号的组成和特点,进而提高预测的精度。本章提出利用局域波进行负荷序列的分解,该方法是一种处理非线性、非平稳信号的很有效的方法,由局域波分解信号后所得到的基本模式分量在很大程度上反映了信号的组成成分和形成过程,而且它不需要预先设定基函数,可根据信号自身的特征进行分解,具有良好的自适应性,而且所得的分量能够突出数据的局部特征。

2.1 局域波理论的起源

在传统的傅立叶分析中,频率被定义为整个分析数据长度中具有一定幅度的正、余弦函数,受这种固有观念的影响,人们在认识和接受瞬时频率的意义和概率时,总是从正、余弦函数的有关角度来分析。这样当人们定义局部频率值时就需要多于一个周期的正、余弦波动,基于这个逻辑,少于一个周期长度的信号将无法给出其频率的定义。而对于非线性和非平稳信号来说,其主要特征频率是时变的,即仅仅是在某一局部时间内存在或曾在某一时刻出现过,在描述频率随时间的变化关系上,傅立叶变换显然已经无能为力。为了弥补傅立叶变换对时变信号分析的不足,人们对原始信号进行加窗处理,认为在某个“窄带”内的信号是平稳的或者近似平稳的,然后再对窗内的信号进行分析,如短时傅立叶变换、小波分析等,这些方法从不同程度上对非线性和非平稳信号的时变性进行了描述,大大改进了傅立叶变换的不足。但是由于受 Heisenberg 测不准原理的制约,在时间和频率上的分辨率不能同时达到最小,因此,所得的结果是窗内信号的平均结果,同样也摆脱不了傅立叶变换的局限性。综合这些分析方法在时频域上的分析效果,可以用一个词来描述——全域波,即在信号分析中,信号频率的定义是基于分析数据整个时间历程上。也就是说,信号分析中获取的任意一个频率值,它所表达的一个周期信号存在于整个分析时间中。因此,利用这种全局意义上定义的频率分量对非平稳时变信号进行分析时,必然会产生额外的谐波频率来满足全局意义上的定义,从而使信号能量扩展到较宽的频带中。为了把信号的频谱分析精确到每一个时间点、频率点上,大连理

工大学振动工程研究所的马孝江教授提出了局域波理论。

局域波理论是受美国学者 Huang 的基于经验模式分解法的启发提出来的。1996 年,美国学者 Huang 在一次国际学术会议上首次提出一种适用于非平稳信号分析的新方法的设想——基于经验模式分解方法^[25]。Huang 认为,对于瞬态与非平稳现象,频率与能量一般都是时间的函数,因此需要给出瞬时频率和瞬时能量的定义。目前信号瞬时能量的概念已被广泛接受,但是瞬时频率的概念和意义却一直存有争议。当可以使数据解析化的 Hilbert 变换出现以后,人们根据 Hilbert 变换提供的,能够完全表达原始数据幅度和相位的函数,给出了瞬时频率的统一定义,从定义可以看出瞬时频率是时间的单值函数,即在任意时刻只存在一个振荡模式。所以在使用瞬时频率这一概念时,对应的数据受到了一定的限制。这主要因为任何一个时刻,数据中可能包含多个振荡模式,此时 Hilbert 变换不能给出该信号完全的频率内容,所得到的结果只是多个振荡模式的平均效果,从而瞬时频率的意义变得模糊。为了从复杂信号中得到有意义的瞬时频率,Huang 根据瞬时频率物理意义上的必要条件,提出把含有多个振荡模式的数据分解成满足一定条件的多个单一振荡模式分量的线性叠加,每个单一振荡模式分量又叫做一个基本模式分量,并提出了一种基于经验模式分解方法。每一个单一模式分量都满足 Hilbert 变换的必要条件,使得用 Hilbert 变换求解信号的瞬时频率成为可能。

2000 年马孝江教授通过对非平稳信号内在特征的深入分析,结合 Huang 的思想和 EMD 算法,首次提出局域波 (Local Wave) 的概念^[26]。局域波意义在于:在信号分析中信号频率的定义是基于波形的局部特征和瞬时特征,它能在信号数据的每个时间点上,从点与点之间的变化特征来给出瞬时频率值,而不是需要多个振荡周期的波形才能给出一个频率值。局域波分析中若存在一个频率,仅仅表示该频率对应的信息在某一局部时间内存在或者曾在某一时刻出现过。这样无论从概念上还是信号分析本质上来看,这种分析方法打破了传统频率思想,给出了一个全新的频率概念。从而把繁杂的信号划分为两大类:局域波和全局波。它们分别表现了信号的微观和宏观特征,代表着信号分析领域的两大阵营,对信号分析具有重大的意义。同时局域波理论的提出也为非平稳信号处理提供了新思路,开辟了新途径。

2.2 局域波理论与方法及其发展状况

2.2.1 瞬时频率的物理意义和局域波法

由于现在的大多数观点认为瞬时频率只有在特定的条件下才存在,比如单一振荡模式信号,但对单一振荡模式信号的定义目前还不明确,于是人们把“窄带”的概念引用到瞬时频率的限定上,以使瞬时频率有意义。虽然 W.K.Melville 成功地分析数据中滤出了满足带宽定义限制条件的信号,但却得到了许多没有物理意义的负

频率分量^[27]。为了得到有意义的瞬时频率，文献[28]讨论了更加严格的条件：任何一个函数要得到一个有意义的瞬时频率，其傅立叶变换的实部必须只有正的频率分量，这个限制条件虽可以在数学上被证明，但仍是一个全局性的定义。而全局性的定义对于频率瞬态变化的非平稳信号的分析将带来较大的偏差。

为了得到具有实际意义的瞬时频率，必须把基于信号全局性的限制条件修改为局部性的限制条件，于是 Huang 等提出了在物理意义上定义一个有意义的瞬时频率的必要条件^[29]：

- (1) 整个数据序列中，极值点的数量与过零点的数量应相等，或最多相差一个；
- (2) 在任一时间点上，信号的局部极大值和局部极小值定义的局部均值应该为零。

第一个限定条件近似于传统的平稳高斯过程的关于窄带的定义；第二个条件是一种新的概念，它把传统的全局限定为局部限定，可以消除由于波形的不对称而造成的瞬时频率波动。满足以上这两个条件的信号称为单一振荡模式（基本模式）信号，这样使得待分析信号在按过零点定义的每个周期中，只包含一个基本模式的振荡，不存在复杂的叠加波。

有了基本模式信号，就可以结合 Hilbert 变换给出瞬时频率的定义了。考虑任意一个具有基本模式的时间序列 $s(t)$ ，为了确定相位，消除信号分析中的负频率成分，把信号转换成解析的形式 $Z(t)$ ：

$$Z(t) = s(t) + j\tilde{s}(t) = A(t)e^{j\varphi(t)} \quad (2-1)$$

其中，

$$\tilde{s}(t) = H[s(t)] = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{s(\tau)}{t-\tau} d\tau \quad (2-2)$$

是信号 $s(t)$ 的 Hilbert 变换；

$$A(t) = \sqrt{s(t)^2 + \tilde{s}(t)^2} \quad (2-3)$$

$$\varphi(t) = \arctan \frac{\tilde{s}(t)}{s(t)} \quad (2-4)$$

$A(t)$ 为幅值函数； $\theta(t)$ 为相位函数。由此定义瞬时频率 ω 为：

$$\omega = \frac{d\varphi(t)}{dt} \quad (2-5)$$

从式 (2-2) 的 Hilbert 变换的定义上，可以看出信号的 Hilbert 变换是该信号与 $\frac{1}{t}$ 的卷积，因此它强调了信号的局部特性，能够从某种程度上给出瞬时频率。

对式 (2-1) 进行傅立叶变换可得：

$$F(\omega) = \int_{-\infty}^{+\infty} A(t)e^{j\varphi(t)-j\omega t} dt = \int_{-\infty}^{+\infty} A(t)e^{j(\varphi(t)-\omega t)} dt \quad (2-6)$$

由平稳相位方法可知，对 $F(\omega)$ 贡献最大的频率应满足以下条件：

$$\frac{d}{dt}(\varphi(t) - \omega t) = 0,$$

$$\text{即 } \omega = \frac{d\varphi(t)}{dt} \quad (2-7)$$

很显然这是一个比过零频率更好的瞬时频率的定义，且与经典的波形理论关于频率的定义是一致的。而且由于式(2-7)中的频率是通过局部正弦函数匹配的平稳相位条件得到的，因此它不需要整个振荡周期来定义一个频率值，而可以在每个点上，从点与点之间的变化来定义它。因此，基本模式的引入，使得瞬时频率这个概念具有了实际的物理意义；而且瞬时频率又不同于用很多谐波分量来描述复杂的非线性、非平稳信号的传统方法，如傅立叶变换，而是与频率的经典定义方法，即信号相位的导数相一致。

上面的分析表明，只有单一模式（基本模式）信号才能通过 Hilbert 变换得到有意义的瞬时频率。但现实存在的观测信号往往是非常复杂的，可能由多个基本模式分量叠加、调制、耦合而成，为了能够准确地求出信号的瞬时频率，必须把它分解成满足以上两个必要条件的基本模式分量，然后求出各分量的瞬时频率特征，通过各分量的瞬时频率特征来反映信号的时频特性，这就是局域波法要研究的内容，而相应信号分解过程叫做局域波分解。

2.2.2 局域波分解算法的基本思想

局域波法首先要做的就是分解信号。局域波分解的目的是把一个数据序列 $s(t)$ 分解成有限个基本模式分量 c_j 和一个趋势项（如果序列存在非零趋势）的和的过程，可表示成如下形式：

$$s(t) = \sum_{i=1}^n c_i(t) + r_n(t) \quad (2-8)$$

式中， $c_i(t)$ ——第 i 个基本模式分量， $r_n(t)$ ——趋势项。

$c_i(t)$ 可以是幅度、频率调制的。由于每个分量的每一个时间点上只有一个频率成分存在，他们符合 Hilbert 变换条件，从而可以通过 Hilbert 变换得到信号的瞬时频率，获得信号频率随时间变化的规律曲线，实现对时变信号完整的、准确的分析。

这种分解的关键是在信号的时间幅度特性上识别它的振荡模式，然后以此为依据分解数据。通过观察数据，可以从两方面直接地识别出信号的不同特性：

- (1) 通过局部极值点——局部极大值与局部极小值之间的时间间隔；
- (2) 通过连续的过零点之间的时间间隔。

这些交错的极值点与过零点构成了复杂的信号波动数据系列。一个波动骑在另一个波动之上，而它们又可能骑在其它的波动之上，如此等等。每一个起伏波动都定义了数据的一个特定尺度特征，因此用连续的极值点之间的时间间隔来作为一个局部振荡模式的时间尺度，它给出了局部振荡模式的更好的解析度。基于这些观点，要想得到满足瞬时频率的两个必要条件的基本模式分量，就要对信号不断地进行消

除骑波和直流成分的筛选操作，把组成信号的各基本模式分量一个一个地筛选出来，完成对信号的分解。

对于任何一个信号分解方法，我们都希望它具有完备性和正交性，这样就保证了分解后，各分量之间无信息重叠，并能够无缺失地重构原信号。而从(2-8)式可知，局域波分解必定是完备的；而文献[30、31]也验证了现有的局域波分解算法的完备性。同时也证明了，分解得到的所有基本模式分量在局部区域内是正交的，即保证了分解的正交性。

2.2.3 局域波分解的几种算法

自从马孝江教授和 Huang 提出局域波理论和 EMD 分解方法后，国内外学者都意识到此理论方法的重要性，积极投身到对它的研究中。到目前为止，已有多种分解算法产生，主要有下面几种：

(1) 经验模式分解方法 (EMD)

这种方法是 Huang 于 1998 年提出来的，也叫做包络均值法。它使用局部极大/极小值点和三次样条插值法计算信号的上/下包络，根据上、下包络求得信号的局部均值，从原信号中减去局部均值就可获得一个基本模式分量。EMD 方法有效地实现了对信号的分解，但是三次样条插值带来的过冲和欠冲问题却给分解带来了严重的误差；而且，三次样条插值使用的次数越多，误差越大。虽然可以用更复杂的样条插值方法来代替，但更高阶样条插值需要消耗更多的计算时间，并且改进效果并不明显^[32]。

(2) 抗模式混叠分解方法

由于局域波分解在任意时间点上总是首先提取频率高的模式，剩下频率低的模式，因此当信号中存在间歇不连续的模式时，就有可能造成连续的模式分量被这些间歇的模式分量割开，使得不同时间尺度（或空间尺度）的模式成分混叠在同一个模式分量中，而相同时间尺度（或空间尺度）的模式成分却被分割到不同的模式分量中，形成所谓的模式混叠现象。为了解决由于间歇波形的存在而使局域波分解产生的模式混叠的问题，Huang 进而提出了使用基于周期长度检验的间歇检验准则。这个准则就是限制包含在同一个模式分量中的波的周期长度上限，这里所说的周期不是传统意义上的周期，而是指两个连续极值点间的距离，它对应着传统谐波意义上的半个周期。间歇检验准则要求选择一个正数 n ， n 代表了半个周期长度，在给定的基本模式分量中，只有两个连续极值点间的距离小于 n 的波形存在，而距离大于 n 数据点被隔离到下面的分量中，从而保证了同一模式分量中只存在相近周期的波形。从本质上说，间歇检验准则类似对在信号分解过程中进行强制滤波，而且这种强制滤波会为分解带来较多的无意义的分量。

除了以上两种分解算法之外，Huang 本人还针对不同的情况提出了一种基于曲

率的分解算法,但由于算法本身使用的范围较窄,而且也存在很多问题,因此没有得到推广。

(3) 自适应时变滤波分解法(ATVFD)

马孝江教授的博士研究生盖强提出了一种基于自适应时变滤波的分解方法,这种方法的局部均值的计算不是基于极值的包络(EMD方法),而是在局部极值的基础上,通过时变滤波计算得到离散局部均值,再用三次样条对其它时间点的局部均值进行插值,得到各点的局部均值;其后的处理方法与EMD方法一致。由于每次分解过程只需要一次三次样条插值进行拟合,从而减少了过冲和欠冲的次数,同时减少了每次分解的重复次数,加速了分解过程,得到了更高的分解精度。

(4) 极值域均值模式分解法(EMMD)

盖强在前两种信号分解方法的基础上,还提出了一种极值域均值模式分解方法。该方法根据积分中值定理来计算相邻极值间的均值。由于该方法在求均值时利用了中值定理原理,并使用了局部信号中的所有数据,因此提高了局部均值计算的准确性。每次分解过程只需要一次三次样条插值进行拟合,从而减少了过冲和欠冲现象,同时也减少了每次分解的重复次数,加速了分解过程。

虽然以上几种分解算法各不相同,但所分解得到的分量都满足基本模式分量的定义。而且频率较高的基本模式先分解出来,随着分解分量的增加,频率逐渐降低,当然这里所说的频率的高低只在局部时间上区分,而不是在整个时间上区分。

由于信号的复杂程度不同,在实际分解过程中也存在着很多问题,比如边界处理和筛选终止规则问题,很多学者在研究分解算法的同时也在寻求解决边界问题和筛选终止规则问题,因此出现了多种不同的解决方案。

2.2.4 局域波分解的边界处理方法

在把信号分解成满足一定约束条件的基本模式分量的过程中,要不断计算信号各个时刻的均值,但由于信号边界的不确定性,边界处均值需近似估计。若估计的边界均值误差较大,容易引起信号各模式分量在两个端点处产生较大的摆动,造成时频域内信号在两个端点处谱能量的扩散;有时还可能向数据的中心扩散,破坏整个数据,因此精确估计边界极值和均值是局域波法的关键问题。针对边界问题,学者们研究出了以下几种方案:

(1) 经验模式分解方法(EMD)的边界处理方法

这种方案是Huang在EMD分解方法中提出的。Huang在数据序列的左右两端点处理使用了特征波,用来消除边界效应。即用与起始点最接近的两个极大值点和极小值点来估计起始点处的极大值和极小值。设起始点处前两个相邻的局部极大值在 t_2 时为 $\max(2)$, t_3 时为 $\max(3)$,局部极小值在 t_2' 时为 $\min(2)$,在 t_3' 时为 $\min(3)$,则上包络起始点用于插值的估计值为:

$$\max(1) = (\max(2) \times t_3 - \max(3) \times t_2) / (t_3 - t_2) \quad (2-9)$$

下包络起始点用于插值的估计值为：

$$\min(1) = (\min(2) \times t_3 - \min(3) \times t_2) / (t_3 - t_2) \quad (2-10)$$

同理也可求出终点处的上下包络的估计值。

这一方法限制了信号两端数据在分解过程中的摆动，不过还需要进一步证实边界效应是否会破坏数据特性，并传播到数据的中间段。

(2) 自适应滤波的边界处理方法

在自适应时变滤波分解法 (ATVFD) 中，为了消除边界效应，用信号两端的极值点数据和信号两端点的原始数据构造滤波函数，得到信号两端的局部均值。具体算法如下：

设所有局部极值点组成的一个时间序列为 $e(t_i)$ ， $i \approx 1, 2, \dots, n$ ，其中 t_i 代表第 i 个局部极值点的时间位置，则对两端点的局部均值可通过下式求出：

$$m(0) = 0.5 \times s(0) + 0.5 \times \frac{t_1}{2t_1 - t_2} \times [e(t_1) - e(t_2) + s(0)] + 0.5 \times \frac{t_2 - t_1}{2t_1 - t_2} \times e(t_1) \quad (2-11)$$

$$m(T) = 0.5 \times s(T) + 0.5 \times \frac{t_n - t_{n-1}}{T - t_{n-1}} \times e(t_n) + 0.5 \times \frac{T - t_n}{T - t_{n-1}} \times [e(t_n) - e(t_{n-1}) + s(T)] \quad (2-12)$$

其中， $s(0)$ 是原始信号的起始点， $s(T)$ 是原始信号的终点值， $m(0)$ ， $m(T)$ 分别是原始信号起点与终点的局部均值。

这种方法有效地减小了边界摆动，却容易引起基本模式分量边界处波形的摆动，并逐渐向数据中心扩散，因而破坏数据的某些特性。

(3) 波形匹配法^[33]

波形匹配边界处理方法是在边界点局部值的求取中，根据波形相关的特性，用信号中与边界三角波形最匹配的波形来预测边界点值。该方法通过匹配预测抑制了边界效应，使求取的边界点局部均值的精度大为提高。

(4) 数据序列端点延拓法

除了以上三种方法外，学者们还尝试把信号序列两端的数据进行延拓，即预测出信号端点以外的数据，并根据延拓的数据计算端点处的局部均值。端点数据延拓的方法主要有以下几种：

① 极值点对称延拓法^[34]

端点延拓的目的是确保上下包络都与端点相交，以便有与每个信号点相对应的局部均值，而上下包络是由极大值和极小值连接而成的，因此只要对极大值和极小值进行延拓，而不必对信号本身进行延拓。由于极大值和极小值是相间分布的，同时考虑到样条插值的要求，只要在信号两端分别对称延拓两个极大值和极小值即可，即把起始位置和终止位置处的两个极大值和两个极小值分别以起始点和终止点为中心对称到数据外，用它们计算起始位置和终止位置的局部均值。

② 基于神经网络的数据序列延拓^[35]

此方法利用神经网络的数据预测功能,在每一次计算三次样条前,对数据序列两端点外的数据进行神经网络预测,使数据的两端各增加两个附加的极大值点和极小值点,利用增加后的极值点进行三次样条插值,得到原始数据序列的上下包络和局部均值。

③ 基于 Chebyshev 数值逼近数据序列端点延拓^[36]

此方法利用 Chebyshev 正交多项式实现信号两端点以外数据预测。Chebyshev 正交多项式具收敛速度快,所需观测样本容量小,在线计算仅需计算两个向量的点积等特点,特别适合于高实时性要求的信号预测。

④ 镜像闭合延拓

镜像闭合延拓的思想是把序列以首尾两端某极值处为镜子,把镜子内的信号向外映射,得到两倍于镜内长度的周期信号(实际分解时只取一个周期),此时序列头尾相接,构成一个闭环的封闭曲线,因此序列不含端点,从而消除了边界效应。这种方法的缺点是:由于数据长度增加了一倍,所以计算量增加了一倍。

尽管各种各样的边界处理方法在一定程度上减少了边界的效应,但是却不能完全消除。而且,由于信号的复杂程度的不同,各种边界处理方法并不总是适用。因此边界效应仍是困扰国内外学者的一个难题。

2.2.5 局域波分解中的终止筛选标准

信号在实际的分解过程中,由于误差的存在,每分解一个模式分量需要多次减去局部均值,每减去一次局部均值叫做一次筛选,当达到某个标准时才可以认为得到的结果是一个模式分量,从而停止筛选。对于筛选终止标准的选择问题,不同的学者有不同的选择标准,使用的参数也各不相同。

(1) Huang 使用参数 (S_d 、 M) 作为停止筛选的标准。 S_d 称为标准差,它通过连续两次筛选结果 $h_{l(k-1)}$, h_{lk} 计算得出:

$$S_d = \sum_0^T \left[\frac{|h_{l(k-1)}(t) - h_{lk}(t)|^2}{h_{lk}(t)^2} \right] \quad (2-13)$$

当 S_d 小于某设定值(一般在 0.2-0.3 之间选择)时,即认为两次筛选的结果相等了,从而可以停止筛选操作。为了防止筛选过程不收敛,同时给定最大筛选次数 M 。

(2) 文献[37]使用参数 (S , Maxmean) 作为停止筛选的标准。即当连续 S 次筛选都产生相同数目的极值点和过零点时,停止筛选。为了使所分解的分量正交指数最小,可选择的范围为: $3 \leq S \leq 8$ 。为了防止满足条件时分解分量的局部均值较大,同时设定终止筛选时,所得模式分量的最大均值不能大于 Maxmean。

(3) Flandrin[38]在总结各种终止标准的基础上,提出了用三个阈值 (θ_1 , θ_2 ,

α) 参数来衡量信号筛选结果。Flandrin 使用的评价函数是:

$$\sigma(t) = \left| \frac{m(t)}{\alpha(t)} \right| \quad (2-14)$$

其中, $m(t) = \frac{e_{\max}(t) + e_{\min}(t)}{2}$ 是模式均值, $\alpha(t) = \frac{e_{\max}(t) - e_{\min}(t)}{2}$ 叫做模式幅值, $e_{\max}(t)$, $e_{\min}(t)$ 分别是模式分量的上下包络。

阈值 θ_1 , θ_2 , α 的意义是: 保证 $\sigma(t)$ 在大部分时间 $T(1-\alpha)$ 内变动较小的情况下 (小于 θ_1), 允许 $\sigma(t)$ 在局部 $T\alpha$ 时间内存在较大变动 (小于 θ_2)。即在 $T(1-\alpha)$ 时间内 $\sigma(t) < \theta_1$, 其余 $T\alpha$ 时间内 $\sigma(t) < \theta_2$ 。一般取: $\theta_1 \approx 0.05$, $\theta_2 \approx 10\theta_1$, $\alpha \approx 0.05$ 。

(4) 文献[39]从保证分解的正交性角度出发, 提出了两参数 ($\min(\Delta E)$, Maxmean) 的能量差跟踪法。其思想是: 当信号分解后的分量恰好是原信号中的一个正交分量时, 分量与剩余部分的总能量与原始信号的能量相等; 当被分解出去的分量不是原信号中的一个正交分量时, 两者能量之间存在差值 ΔE , ΔE 越小, 越能保证信号分解结果的完备性和正交性。因此, 在对信号进行分解时, 对 ΔE 跟踪, 当 ΔE 达到某一个最小值 $\min(\Delta E)$, 而且包络平均值足够小时 (小于 Maxmean), “筛选” 过程结束, 此时得到的分量是原始信号中的一个正交分量。

以上四种筛选终止标准侧重的出发点不同, 各具特点。使用不同的标准分解出的结果不同, 主要表现在分解出的分量个数不同, 但每个分量基本上都能满足求瞬时频率的两个必要条件。

2.3 局域波时频谱

信号经过局域波分解可以得到有限个分量, 通过 Hilbert 变换将分量信号 $C_i(t)$ 变成解析信号:

$$Z_{C_i}(t) = X_{C_i}(t) + jY_{C_i}(t) = a_{C_i}(t)e^{j\varphi_{C_i}(t)} = a_{C_i}(t)e^{j\omega_{C_i}(t)t} \quad (2-15)$$

其中

$$a_{C_i}(t) = [X_{C_i}^2(t) + Y_{C_i}^2(t)]^{\frac{1}{2}}, \quad \varphi_{C_i}(t) = \arctan \frac{Y_{C_i}(t)}{X_{C_i}(t)} \quad (2-16)$$

由此得到分量瞬时频率 $\omega_{C_i} = \dot{\varphi}_{C_i}$

由于分量信号 $C_i(t)$ 可表示为: $C_i(t) = \text{Re}[a_{C_i}(t)\exp(j\int\omega_{C_i}(t)dt)]$, 其中 Re 表示取实部, 则 $s(t)$ 可表示为:

$$s(t) = \text{Re}[\sum_{i=1}^n a_{C_i}(t)\exp(j\int\omega_{C_i}(t)dt)] \quad (2-17)$$

其中 n 为分量个数, 对于余量 Hilbert 变换可以把它当作一个长周期的低频振荡来处理, 这在能量上表现为一个微不足道的偏移量。考虑到这个长期趋势的不确定性, 以及人们为了更好的观察包含有用低频信息的局域波分量, 在进行 Hilbert 变换的时候通常都不考虑余量。

同样的数据用 Fourier 展开将是:

$$s(t) = \operatorname{Re}\left[\sum_{i=1}^{\infty} a_i e^{j\omega_i t}\right] \quad (2-18)$$

其中, a_i 、 ω_i 都是常量。由上述分析可知, 每一个局域波分量描述了一种广义的 Fourier 变换, 可变的幅度与瞬时频率不但在很大程度上改进了信号展开的效率, 而且使这种分解可以处理非平稳数据。这样一来, 已经打破了固定幅度和固定频率的 Fourier 展开的限制, 从而得到一个可变幅度、可变频率的信号描述方法。

可以把信号幅度在三维空间中表示成时间和频率的函数 $H(\omega, t)$, 称之为信号的局域波时频表示, 也可以称为局域波时频谱, 表示为:

$$H(\omega, t) = \operatorname{Re}\left[\sum_{i=1}^n b a_{c_i}(t) \exp(j\int \omega_{c_i}(t) dt)\right] \quad (2-19)$$

式中, n 为分量个数, 当 $\omega_{c_i}(t) = \omega$ 时, 变量 $b = 1$, 否则 $b = 0$ 。根据需要还可以得到分量的时频谱:

$$H_{c_i}(\omega, t) = \operatorname{Re}[b a_{c_i}(t) \exp(j\int \omega_{c_i}(t) dt)] \quad (2-20)$$

或者几个分量进行重构的时频谱:

$$H_{c_1 \dots c_k}(\omega, t) = \operatorname{Re}[b a_{c_1}(t) \exp(j\int \omega_{c_1}(t) dt)] + \dots + \operatorname{Re}[b a_{c_k}(t) \exp(j\int \omega_{c_k}(t) dt)] \quad (2-21)$$

在这样一个三维 $\{t, \omega, H(\omega, t)\}$ 平面中, 信号的幅度可以表示为时间频率平面上的等高线, 或者三维立体图。

局域波时频谱的表达形式可以有多种, 如灰度图形式, 等高线形式等。局域波时频谱能定量地描述时间和频率的关系, 实现了对时变信号完整的、准确的分析。

2.4 局域波法的工程应用现状

经过近几年的研究, 局域波法逐步形成了独立的理论体系。作为一种新的分析方法, 在非平稳信号分析中的作用和优势已初见端倪, 该方法打破了传统上定义频率的思想, 成功地给出了信号本质形象的描述和瞬时频率的物理意义, 可以说是整个信号分析领域的重大突破。在局域波理论发展的同时, 也得到了实际应用, 并且获得了丰硕的成果。

最早的局域波的应用是在 Huang 刚刚提出 EMD 分解算法的时候, Huang 把这种算法应用到非线性系统的数值分析、浪波数据处理、潮汐规律分析以及地震数据处理等方面, 都得到了意象不到的良好效果, 这也是这种算法迅速发展起来的原因。后来 Huang 又把它应用到其他领域, 如生物医药和经济领域, 用来研究肺部血压在缺氧状态下的响应^[40-41]和股票波动规律^[42]等现象; 其它的学者也在这方面做了一些尝试, 如 H.Liang^[43]、R.Balocchi^[44], 和 S.C.Phillips^[45]等人就用此方法来分析心、肺和胃部信号, 在病理研究上取得了良好的效果。同时, 局域波在土木工程领域也得到了应用, 文献[46]使用 EMD 算法分解地震波信号, 找到了地震波中 S 波检测

的新方法,求出了地震表面波的色散曲线,揭示了地震波的内在信息;而文献[47]则把这种方法应用到结构损伤检测中,成功地实现了对结构损伤位置、时刻、频谱等的监测分析。

局域波法与其它非平稳信号处理方法结合也取得了意想不到的效果,如文献[48]用局域波法分解信号后计算其 Wigner-Ville 分布,从而有效抑制了交叉干扰项。文献[49]提出了基于局域波法的非平稳信号的 AR 和 ARMA 模型分析方法,同样也得到了较好的效果。许多学者还试图通过局域波法这种非平稳、非线性信号的分解方法来解释科学中的许多随机现象,如 Stephen C. Phillips 用 Hilbert-Huang 变换成功地模拟出了分子的运动^[50]; Gabriel Rilling 用 EMD 算法研究分形高斯噪声 HURST 指数和分解分量序数之间的关系,认为可以根据分解分量序数估计出分形高斯噪声 HURST 指数^[51]; Zhaohua Wu 分析了白噪声的 EMD 分解特性,得到了白噪声 EMD 各分量的统计分布规律^[52],在此基础上 Patrick Flandrin 建立了信号被这种噪声污染时的去噪方法^[53]。

事实证明,局域波法解决了很多以前一些无法解决的问题,在旋转机械的故障诊断中,文献[54]通过对 EMD 和小波分析的比较,证明对于旋转机械振动信号,给予 EMD 的局域波时频方法比基于小波分析更有效;文献[55]应用 EMD 法对万能外圆磨床的砂轮架主轴振动信号进行了分析,指出了该方法对非平稳信号分析更具有直观物理意义;文献[56]通过局域波时频方法对转子系统的早期故障进行了分析,描述了这些故障振动信号的时频特性。文献[57]利用局域波时频谱实现了齿轮磨损故障的诊断;文献[58]等也利用局域波法对齿轮箱的故障进行了成功的诊断。局域波法也解决了很多往复机械的故障问题,如王珍^[59]、邹岩崑^[60]等利用局域波法深入分析了柴油机故障信号,得到了不同故障的频谱特征,为今后分析像柴油机这样复杂的往复机械提供了参考。局域波法还和其它的诸如 AR 模型^[61]、分形动力学^[62]、神经网络^[63]等方面的方法结合,提取出了代表不同故障特征的特征,实现了不同故障的分类,为模式识别提供了新思路。

前面所叙述的局域波方法和应用都是对一维信号的处理,在二维图像处理领域,学者们也提出了相应的局域波算法^[64],建立了图像纹理提取^[65]、滤波^[66]、去噪^[67]的新方法,并且应用在雷达图像的目标辨识^[68]、大气图像的降雨区分析^[69]以及故障频谱图^[70]的结果分析等方面,实现了图像数据的线性化、平稳化预处理与不同信息的尺度分离,为图像谱分析研究和目标识别研究提供了一种有效的数据源预处理新方法。

同时,瑞典学者 Anna Linderhed 还另辟蹊径,在比较小波和局域波特点的基础上,提出了利用局域波法进行数据压缩的新方法,为信号传输技术领域开创了新的前景^[71]。

局域波法从提出到现在只有几年的时间,已经有各种各样的成果涌现出来,可

见发展速度相当之快。特别应当指出的是，局域波法解决了非平稳信号分析中的一些困难，相信对它的深入研究可以给信号处理领域带来一次革命，并且必将会有力地推进工程应用的发展。

第三章 地区负荷特性分析

随着科学的进步,经济的发展,负荷结构日趋复杂,负荷成分从单一化逐渐变得多样化,新类型的负荷,如电气化铁路负荷占网内用电量越来越大,对负荷预测的影响也越来越明显。并且不同的影响因素,如气象因素中的温度、湿度、风速等对负荷的影响规律也不同,这些都对提高负荷预测的精度造成了很大的困难。如果单靠改进纯粹的数学方法来提高预测精度,而不是从电力系统的角度,重视负荷本身特性,显然是片面的。因此必须从负荷构成入手,研究其变化规律及影响因素,然后在此基础上根据负荷的不同特点,运用适合各自自身特点的合理的预测方法进行预测。

近似熵是从衡量信号复杂性的角度提供了一种表达信号特征的无量纲指标。它已被成功地应用到生理性时间序列分析^[72]、机械设备状态监测和故障诊断中^[73];近几年又被应用于电力系统的故障诊断领域,将近似熵作为特征向量进行故障判断,取得很好的效果^[74]。本文将近似熵引入负荷序列的分析中,用近似熵来度量各分量、余量和不同类型负荷的特征,通过对比各分量与不同类型负荷的近似熵值,进行分量的类型识别。

3.1 近似熵原理

相对于一些非线性动力参数(如关联维数、哥氏熵、李雅普诺夫指数等)而言,90年代初由Pincus提出的近似熵(Approximate Entropy,简记为ApEn)主要是从衡量时间序列复杂性的角度来度量信号中产生新模式的概率大小,产生新模式的概率越大,序列的复杂性越大,相应的近似熵也越大^[75]。

近似熵是用一个非负数来表示某时间序列的复杂性,越复杂的时间序列对应的近似熵越大。下面给出具体的算法步骤:

设采集信号的序列为 $\{x(i), i=0,1,\dots,N\}$,近似熵可以通过以下步骤计算得到:

(1) 给定模式维数 m ,用原数据组成一组 m 维矢量:

$$X(i)=[x(i),x(i+1),\dots,x(i+m-1)] \quad (3-1)$$

其中: $i=1\cdots N-m+1$;

(2) 定义 $X(i)$ 与 $X(j)$ 之间的距离:

$$d(X(i),X(j))=\max_{k=0\cdots m-1} [|x(i+k)-x(j+k)|] \quad (3-2)$$

(3) 给定阈值(相似容限) $r(r>0)$,对每个 i 值统计 $d(X(i),X(j))$ 小于 r 的数目及此数目与距离总数 $N-m+1$ 的比值,记作 $C_i^m(r)$,

$$C_i^m(r)=\frac{1}{N-m+1}\{[d(X(i),X(j))<r]_{num}\} \quad (3-3)$$

(4) 对 $C_i^m(r)$ 取对数, 再求其对所有的 i 的平均值, 记作 $\Phi^m(r)$:

$$\Phi^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \ln C_i^m(r) \quad (3-4)$$

(5) 把维数 m 加 1, 变成 $m+1$, 重复 (1)-(4) 步, 得 $\Phi^{m+1}(r)$;

(6) 理论上, 此序列的近似熵为:

$$ApEn(m, r) = \lim_{N \rightarrow \infty} [\Phi^m(r) - \Phi^{m+1}(r)] \quad (3-5)$$

此极限值以概率 1 存在, 但在实际工作中 N 不可能为 ∞ , 当 N 取有限值时, 可得到近似熵估计值:

$$ApEn(m, r, N) = \Phi^m(r) - \Phi^{m+1}(r) \quad (3-6)$$

$ApEn(m, r, N)$ 的值与参数 m, r, N 的选取有关。当 $m=2, r=0.1 \sim 0.2SD_x$ (SD_x 为原始序列 $\{x(i), i=0, 1, \dots, N\}$ 的标准差) 时, $ApEn(m, r, N)$ 的值对 N 的依赖程度最小, 即:

$$ApEn(2, r, N) \approx ApEn(2, r) \quad (3-7)$$

因此实际的计算中, 常常取 $m=2, r=0.1 \sim 0.2SD_x$ 。

可以看出, 近似熵实际上是在衡量当维数变化时该时间序列中产生新模式的概率的大小, 产生新模式的概率越大, 序列就越复杂, 对应的近似熵也就越大, 因此从理论上讲, 近似熵能够用来表征信号的不规则性和非线性程度。两个近似熵值相近的信号, 它们的非线性和非平稳性特征相近。文献[76]指出, 近似熵大致相当于维数变化时新模式出现在对数条件概率的均值, 在衡量时间序列的复杂性方面具有一定意义; 而且它不仅仅是一个非线性动力学参数, 对随机过程和确定性过程都适用, 文献[77]曾用近似熵对混沌信号进行分析。综合有关文献的论述, 可总结出近似熵的主要特点如下:

(1) 只需要比较短的数据就能估计出比较稳定的统计值, 因此比较适合工程应用;

(2) 近似熵值的大小和采样序列复杂程度成正比关系, 序列越复杂, 对应的近似熵值也就越大;

(3) 近似熵具有一定的抗噪能力, 特别对偶尔产生的瞬态强干扰有较好的承受能力。

下面通过一组简单的信号来验证近似熵算法在特征提取中所具有的良好能力。定义如图 3-1 所示的 3 组信号: $S_1 = \sin(2\pi f_1 t)$; $S_2 = 0.75\sin(2\pi f_2 t)$; $S_3 = \sin(2\pi f_1 t) + \sin(2\pi f_2 t) + N$ 。式中: $f_1 = 50\text{Hz}$, $f_2 = 70\text{Hz}$; N 为均值为 0、方差为 1 的均匀白噪声。采样 100 个数据点作为被分析信号序列, 在近似熵算法中取 $m=2, r=0.2std$ (std 表示信号的方差)。

分别计算 S_1 、 S_2 、 S_3 的近似熵值, 计算结果如表 3-1 所示。

可以发现, S_1 只含单一的基波正弦, 这是一种平稳而规则的信号, 故其近似熵较小; S_2 近似熵值与 S_1 相近, 由于 S_2 与 S_1 一样也只含单一的正弦波, 它们在线性

和非平稳特性上具有相似性；而 S_3 由于叠加了 3 次谐波和白噪声而变得较为复杂，其近似熵值较大。

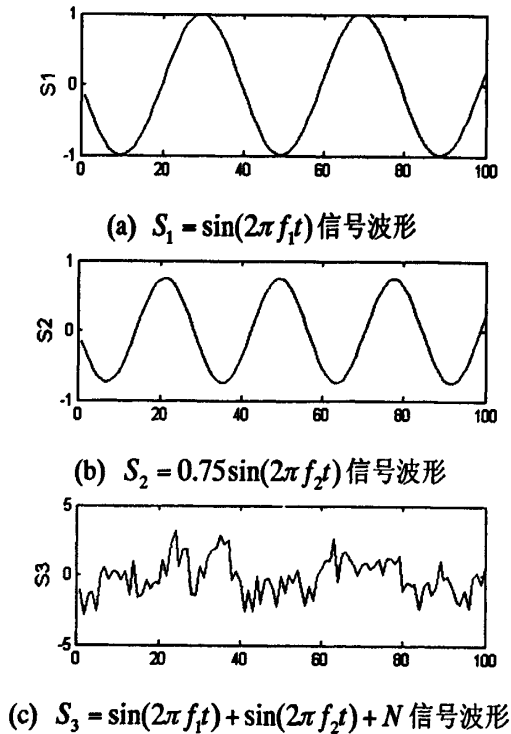


图 3-1 三组分析信号波形

表 3-1 三组分析信号的近似熵

信号序列	S_1	S_2	S_3
近似熵值	0.13303	0.13138	0.76167

因此，把近似熵用于负荷分析，以此为依据来识别各分量的负荷波动类型。

3.2 实例分析

3.2.1 数据来源

本文选取我国华北网某地区为分析对象，负荷数据为该地区 2009 年 3 月 16 日至 2009 年 3 月 27 日全网用电量数据，每日各 96 个数据点。该时段正是北方停止供暖，而气温相对较低，人们开始使用电器取暖，因此这段时间气象因素对负荷的影响较明显。考虑到周末放假期间用电情况的特殊性以及对前后的影响，从上述时间段的数据中去掉了周六（2009 年 3 月 21 日）和周日（2009 年 3 月 22 日）两天的数据。

由于原始数据往往从 EMS 实时系统采集的，动态的数据采集有时会出现通道故障、拥堵等现象，相应的数据采集程序就会中断，造成原始数据的缺失，因此对缺失数据利用三次样条插值法进行插值补充，对后面的数据同样用该方法进行预处

理。

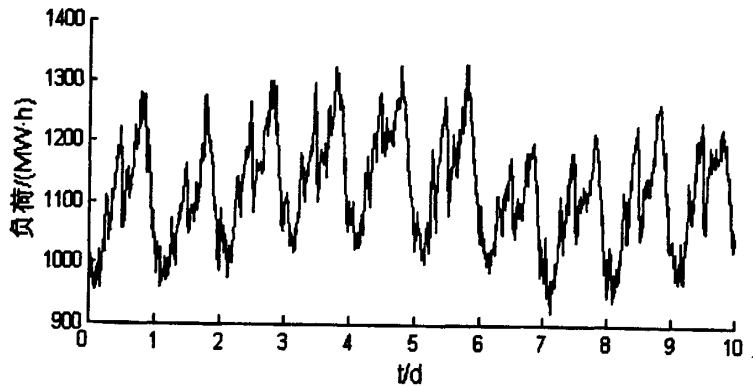


图 3-2 历史负荷数据

图 3-2 为所分析的历史负荷曲线, 从图中可看出负荷呈现以日为周期的波动, 且总体趋势保持稳定, 每日存在两个明显的高峰负荷, 约出现在正午 14:00 和晚上 20:00 时刻, 同时该曲线在局部时间段出现剧烈波动, 严重影响了负荷预测的精度。

对该地区进行数据统计, 发现该地区电铁、工厂负荷占总电网用电量较大, 而地区电网基数较小, 因此该地区负荷受电铁、工厂负荷的冲击性影响较突出。同时公共用电负荷(包括机关、政府、事业单位、商店用电等)也是该地区的主要负荷类型。因此从 SCADA 系统上采集 2009 年 3 月 16 日至 2009 年 3 月 27 日(除去 2009 年 3 月 21 日和 2009 年 3 月 22 日)的电铁负荷、工厂负荷和公共用电负荷数据作为本文分析对象, 每日各 96 个数据点。

图 3-3 为电铁负荷曲线, 该类负荷表现为无规律的冲击性, 且相隔时间短。图 3-4 为工厂负荷曲线, 该负荷波动也较剧烈。图 3-5 为公共用电负荷, 该类型负荷周期规律性较强。

为了更加精细地分析气象因素对负荷的影响情况, 本文采用该地区实时气象因素对负荷所分解出来的各个分量进行相关性分析, 包括 24 小时气温、24 小时湿度、24 小时风力。

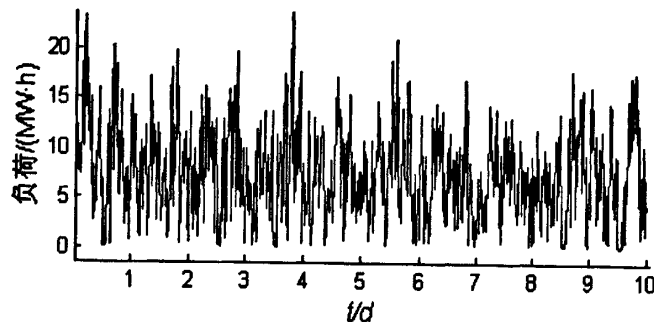


图 3-3 电铁负荷数据

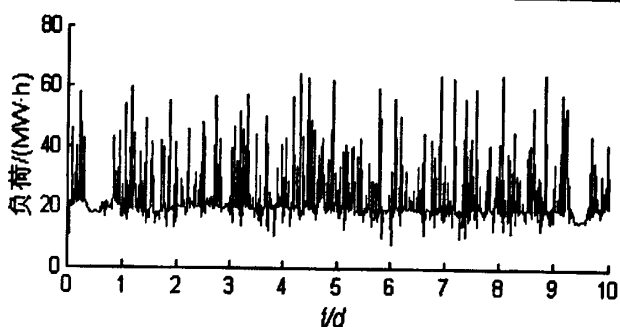


图 3-4 工厂负荷数据

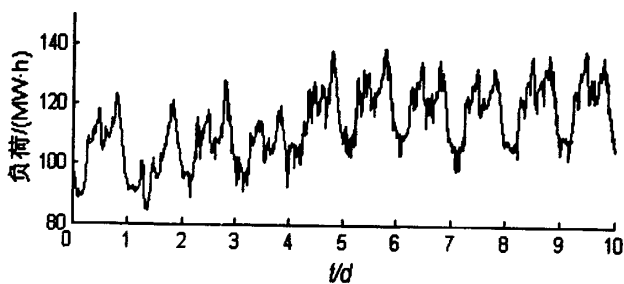


图 3-5 公共用电负荷数据

3.2.2 负荷局域波分解

本文利用局域波中自适应时变滤波分解法 (ATVFD) 和自适应滤波的边界处理方法, 使用参数 (S_d 、 M) 作为停止筛选的标准, 将负荷序列分解成多个基本模式分量, 来实现负荷特征的分离, 分解所得结果见图 3-6。

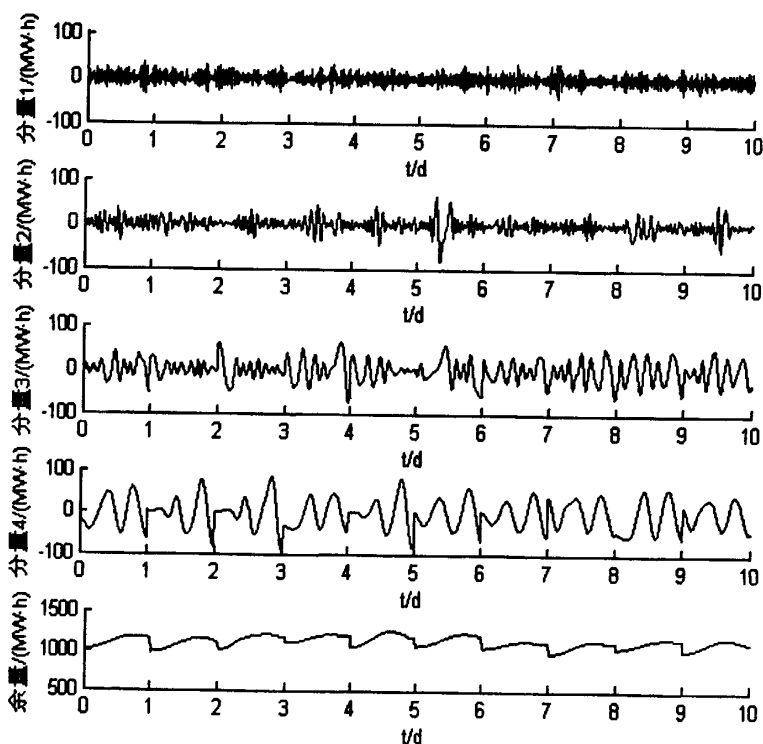


图 3-6 历史数据局域波分解得到的各分量和余量

分量一的曲线波动最剧烈, 有较强的冲击性; 分量二的频率小于分量一, 但是

波动规律性仍然不强，且仍具有一定的冲击性；分量三波动渐缓，具有较强的周期性，周期约为 4-5h；分量四具有一定的规律性，周期较长，变化较平缓，且每日都出现两个峰值；余量呈现明显以日为周期的周期特性，且每日的趋势为从低到高，再从高到低的变化。

3.2.3 局域波时频分析

在得到负荷分解基本分量后，对每个分量进行 Hilbert 变换，得到各分量的时频谱和功率谱，剖析各自对总负荷波动贡献程度。

具体步骤如下：由局域波分解得到

$$X(t) = \sum_{j=1}^n C_j(t) \quad (3-8)$$

对每一个基本内蕴模式函数 $C_j(t)$ 进行 Hilbert 变换，得到： $C_j(t)$ 的解析表达式 $a_j(t)e^{i\int \omega_j(t) dt}$ 。最后有原始数据 $X(t)$ 的表达式为其 Hilbert 变换后的实部：

$$X(t) = PR \sum_{j=1}^n a_j(t) e^{i\int \omega_j(t) dt} \quad (3-9)$$

经过这些处理后的时间频率平面上的幅度分布被称为 Hilbert 时频谱，而幅度与频率的分布称为 Hilbert 功率谱，功率谱代表了每个频率在全局上的幅度（或能量）贡献。可以通过计算幅度方差百分比，得到各个频率段波形对整体波形的波动贡献。

这里省去了余量的 Hilbert 变换，因为它或者是一个单调函数，或者是一个常量。Hilbert 变换可以把单调趋势当作一个长周期振荡的一部分，但是余量的能量可能过大，导致掩盖了其他分量的能量特点，因此省去余量的时频分析。

图 3-7 和图 3-8 为各分量经过 Hilbert 变换后的时频谱和功率谱，从图中可看出低频处分布较密集，且幅值大，高频处分布稀疏，且幅值小。计算幅度方差百分比，得到分量一为 15.73%，分量二为 15.34%，分量三为 28.05%，分量四为 40.88%。因此，分量三和分量四对总负荷波动性的贡献最大，即低频处分量是造成负荷波动的主要原因。而分量一与分量二对总负荷的波动贡献之和超过了三分之一，说明它们对负荷的波动造成了一定的影响，由于这两个分量的波动具有无规则性，导致该地区负荷局部的剧烈波动，影响该地区负荷预测的精度。图 3-9 为各分量对总体负荷波动贡献的程度。

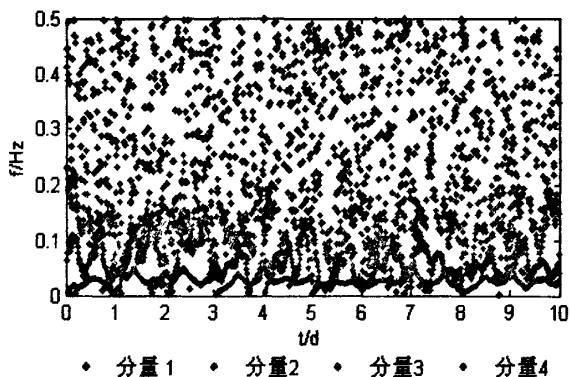


图 3-7 负荷的 Hilbert 时频谱

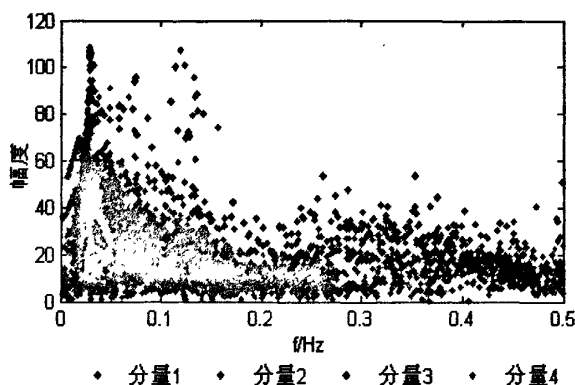


图 3-8 负荷的 Hilbert 功率谱

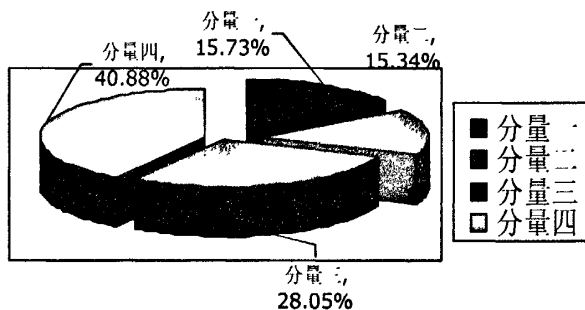


图 3-9 各分量对总体负荷波动贡献的百分比

3.2.4 近似熵分析

得到各分量和余量后，对其进行近似熵计算，并对比电铁、工厂、公共用电负荷的近似熵，以此为依据，对各分量进行负荷类型识别。所得结果见表 3-2 和表 3-3。

从表 3-2 中可以发现，各分量的近似熵随着分量阶数的增加而减小，即局域波分解后所得的各分量波形随着分量阶数的增加趋于规则，各分量的复杂程度由分量一到余量依次递减；表 3-3 说明了电铁负荷、工厂负荷与公共用电负荷的近似熵值逐个递减。这表明了近似熵确实可以很好地量化不同负荷类型的各自不同特点。

表 3-2 各分量及余量的近似熵

日期	分量一	分量二	分量三	分量四	余量
3.16	0.35156	0.30671	0.037272	0.01	0.0001
3.17	0.40366	0.34635	0.14138	0.00908	0.0001
3.18	0.43618	0.31909	0.025326	0.009341	0.0001
3.19	0.5106	0.19493	0.065221	0.01	0.0001
3.20	0.5165	0.23238	0.056528	0.009085	0.0001
3.23	0.38466	0.247487	0.020144	0.007883	0.0001
3.24	0.55322	0.34648	0.031285	0.004114	0.0001
3.25	0.51251	0.38478	0.068642	0.009443	0.0001
3.26	0.40733	0.19306	0.050227	0.0100	0.0001
3.27	0.49613	0.17125	0.023189	0.002591	0.0001
平均	0.457235	0.274252	0.051921	0.008154	0.0001

表 3-3 不同类型负荷的近似熵

日期	电铁负荷	工厂负荷	公共用电负荷
3.16	0.35114	0.27123	0.017342
3.17	0.63149	0.19251	0.022098
3.18	0.50801	0.30771	0.048913
3.19	0.44385	0.20798	0.084703
3.20	0.58832	0.14373	0.036816
3.23	0.53932	0.25079	0.042823
3.24	0.69762	0.18656	0.058693
3.25	0.83765	0.21433	0.053174
3.26	0.57315	0.16197	0.038162
3.27	0.407	0.213	0.027194
平均	0.557755	0.214981	0.044009

根据表 3-2、3-3，分量一近似熵均值约为 0.46，电铁负荷约为 0.56，两者较为接近，因此可以认为构成该分量的主要成分是以电铁为代表的冲击负荷，该分量在一定程度上反映了电铁负荷的变化情况。分量二的平均熵值约为 0.27，工厂负荷的平均熵值约为 0.21，两者较为相近，这说明了分量二与工厂负荷的波动特点相接近，能够在一定程度上反映工厂负荷变化的特点。分量三的平均熵值约为 0.052，公共用电负荷的平均熵值约为 0.044，发现其熵值相差不大，因此可以判断出该分量波动特性代表了公共用电的负荷特点。

3.2.5 实时气象因素影响分析

目前，对负荷的气象因素影响的研究绝大多数是依据每日的特征因素，如日平均温度、日平均湿度等。但是采用日特征气象因素无法深入分析气象对每日不同时间段负荷的影响情况，在进行预测时，只能从整体上对全天负荷进行修正。因此本文考虑实时气象因素对负荷的复杂影响。

由于该地区气象资料只能提供每日 24 点的实时数据，因此，需要对气象数据进行三次样条插值处理成为 96 点的曲线，与各负荷分量曲线相匹配。

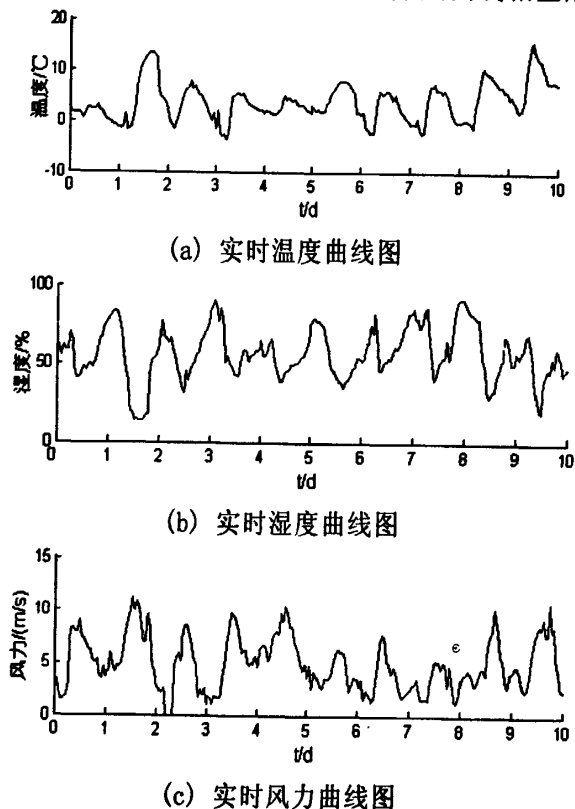
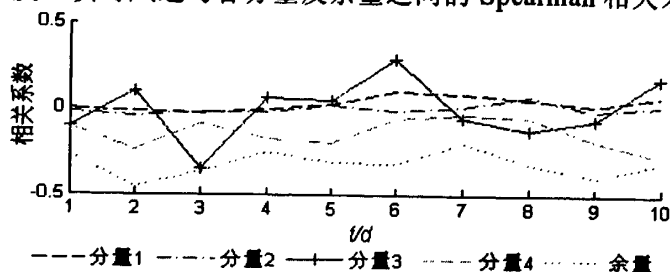
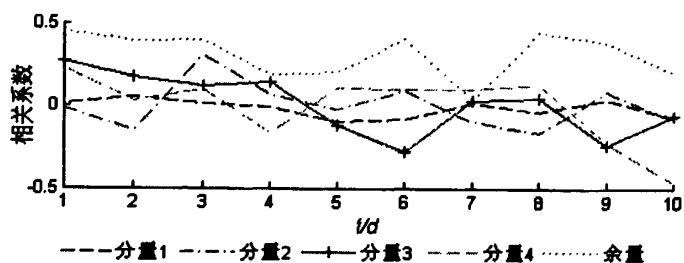


图 3-10 实时气象因素曲线图

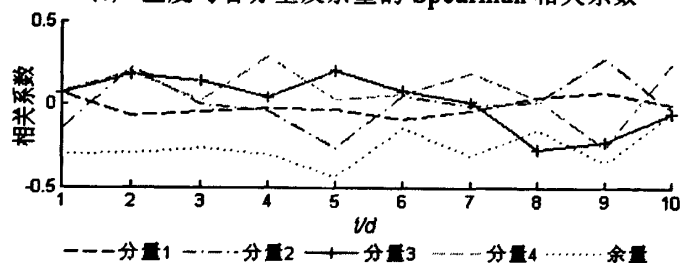
Spearman 相关系数对原始变量的分布不作要求，适用范围广，因此本文采用 Spearman 相关系数来分析各分量与实时气象因素之间的相关性。图 3-11 为每日实时温度、实时湿度、实时风速与各分量及余量之间的 Spearman 相关系数。



(a) 温度与各分量及余量的 Spearman 相关系数



(b) 湿度与各分量及余量的 Spearman 相关系数



(c) 风速与各分量及余量的 Spearman 相关系数

图 3-11 实时气象因素与各分量及余量的 Spearman 相关系数

图 3-11 显示了实时气象因素与每日的分量一、分量二相关系数接近于 0，分量三、分量四系数较大，余量的相关系数最大，这说明了气象对分量一、分量二的影响不明显，对分量三、分量四及余量的影响比较突出，从而体现了气象对日负荷的影响周期性较长。对比各相关系数值可以发现，温度和湿度与各分量及余量的相关系数值最大，这体现了实时温度和实时湿度是影响负荷的主要因素。

分量三与气象的关系较复杂，与各实时气象因素的相关系数时而为负，时而为正，规律性不强。

分量四受气象影响较大，规律性较强。温度对分量四呈负相关性，即该时间段的实时温度对负荷为负影响，温度越低，负荷越大；而湿度为正相关，即实时湿度对负荷为正影响，湿度越大，负荷越大；风速对分量四影响较不稳定。对比图 3-6 和图 3-10，每日的下半夜，温度低，湿度大，分量四较小；随着温度的回升，湿度的回落，分量四逐渐升高，并出现第一个峰值；下午温度开始下降，湿度明显上升，这时分量四开始回落；到了傍晚温度明显降低，湿度进一步回升，分量四出现一日中的第二个高峰；之后，分量四逐渐回落。

余量受气象影响最大。温度对余量也呈现负相关性；湿度也为正相关；而风速对余量呈现负相关。对比图 3-6 和图 3-10，从午夜到凌晨，余量较小；而后，温度升高，负荷值开始升高，到了傍晚，温度明显降低，余量又逐渐回落。

3.2.6 负荷分析

通过局域波近似熵对负荷进行分析，可以发现负荷由四个分量和一个余量组成。

分量一的曲线波动与电铁负荷的特点相近,且两者的近似熵也较为接近,因此认为分量一反映电铁负荷的波动特点。通过计算可知分量一对总体负荷的波动贡献约为 15.73%,说明该地区受电铁负荷的影响较为严重,导致该地区负荷曲线出现许多无规律的波动,大大影响负荷预测的精度。从该分量与实时气象的相关性来看,气象的变化对它影响较小,这也符合电铁负荷基本不受气象因素影响的实际特点。目前对电铁这类冲击负荷的研究还很少,如果能加强对该类负荷特性及预测的研究,对提高整体负荷预测具有很大的帮助。

分量二的近似熵与工厂负荷的较为相近,说明分量二能够在一定程度上反映工厂负荷变化特点。同时,该分量受实时气象因素影响也较小,这也与工厂用电的特点一致。由于该部分基本不受外界环境变化而产生波动,而波动的周期性不是很明显,所以可以运用时间序列对其进行预测。

分量三与公共用电负荷的近似熵相差不大,可判断出该分量波动特性代表公共用电的负荷特点。同时分量 3 对整体负荷波动的贡献相对较大,占 28.05%。该分量受一定气象因素影响,但影响规律较复杂。鉴于这部分负荷分量的周期性较好,但受气象因素影响的规律性差,因此可运用支持向量机进行预测。

分量四波动平滑,具有一定的规律性,每日都出现两个峰值,这与人们日常生活中白天和晚上两个用电高峰现象相符合,因此可以认为,该分量体现了受人们活动规律影响的用电负荷部分。由图 3-7、3-8 可见,分量四的能量较为密集,且幅值较大,所占能量比重较大,对整体负荷波动的贡献最显著,为 40.88%,这说明人们的活动规律是影响负荷波动的主要因素。同时,多种实时气象因素对该分量的影响较显著,其中实时温度对这部分影响最大且呈现出负相关性。由于该负荷的负荷规律性较强,因此对该部分的负荷预测相对比较容易,一般运用多元回归模型进行预测。

余量代表负荷的总体趋势,每日由低到高,再由高到低波动,这与电网的夜晚用电低谷和白天用电高峰特点相符,因此它代表了该地区整体负荷的变化趋势。各种实时气象因素对余量的影响最为明显,这也揭示了一天的负荷趋势与当天的气象因素的关联度很大,气象对负荷的影响周期较长。同时,温度对余量呈现较大的负相关性,湿度为较显著的正相关性,风速为负相关性。由于余量的周期性最强,受气象因素影响也最大,所以可以运用预测分量四的方法对其进行预测。

3.3 本章小结

精确的负荷预测是人们所期望的,而高精度的负荷预测既需要充分分析负荷的自身规律性,又要兼顾各种因素的影响。因此本文提出了一种局域波与近似熵相结合的负荷分析方法,经研究得出 3 点结论。

(1) 局域波能有效地处理非平稳信号，它能自适应地将信号逐级分解，分解后的分量能有效反映负荷的本质特征，通过 Hilbert 变换能分析出各分量对总体负荷的波动贡献。

(2) 近似熵具有很强的表征信号不规则性和复杂性的能力，通过计算各分量与实际各类型负荷的近似熵，作为它们的特征参数，为各分量的负荷类型识别提供有效的依据。

(3) 采用实时气象因素对负荷所分解出来的各分量及余量进行相关性分析，可以精细化挖掘出气象因素对它们的影响情况。

第四章 含电铁负荷的地区负荷预测模型

近年来,某些地区的电铁负荷占网内用电量越来越大,据秦皇岛供电公司调度所统计,该地区08年一季度电气化铁路负荷同比增长了65%,仅一个220kV站带的重载电铁线路的电铁负荷就有60MW。由于该类负荷瞬间变化大,具有很强的冲击性,使得该地区负荷变化比较剧烈,局部波动幅度较大,据该地区相关资料显示,电铁负荷影响地区负荷预测准确率约大于3.93个百分点,它对负荷造成的冲击性已经成为提高预测准确度的一个瓶颈。同时电铁负荷在大幅度上升过程中,经常在指标内用电地区短时超过计划用电指标,而超过负荷计划指标就要对其他电力用户实行限电来确保在指标内用电;在大幅度降低时,由于时间短暂及无规律性,不能组织其他用电生产,这样影响了其他用户用电,也干扰了正常的计划用电工作,给电力系统及用户用电带来了较大困难。由于电铁负荷变化剧烈,随机性较强,因此,研究电铁负荷的变化规律,以及电铁负荷的预测方法,对于提高提高负荷预测准确度和制定用电计划有着重要的现实意义。

通过第三章的研究,可以得出:分量一的曲线波动与电铁负荷的特点相近,根据该分量的特点,运用基于灾变遗传算法和时序分析的最小二乘支持向量机方法对其进行预测;分量二与工厂负荷的较为相近,受气象因素影响较小,因此可用时间序列进行预测;分量三与公共用电负荷类似,鉴于这部分负荷分量的周期性较好,受气象因素影响的规律性差,运用支持向量机进行预测;分量四波动平滑,与人们日常生活中白天和晚上两个用电高峰现象相符合,多种实时气象因素对该分量的影响较显著,因此对该部分的负荷预测相对比较容易,运用多元回归模型进行预测;余量代表负荷的总趋势,周期性最强,受气象因素影响也最大,所以可以运用预测分量4的方法对其进行预测。最后将各分量的预测结果合理整合,得到最终的预测结果。

4.1 电铁负荷特性及其影响因素分析

图4-1为我国华北网秦皇岛地区某条电铁线路2009年3月16日至2009年3月29日的用电量数据,每日各96个数据点。从图上可以看出,电铁负荷在短时间内的变化很快,呈现出无规律的大升和大降,并且具有一定的冲击性,波动剧烈。但是从每日的整体变化趋势看,也具有一定的规律性。

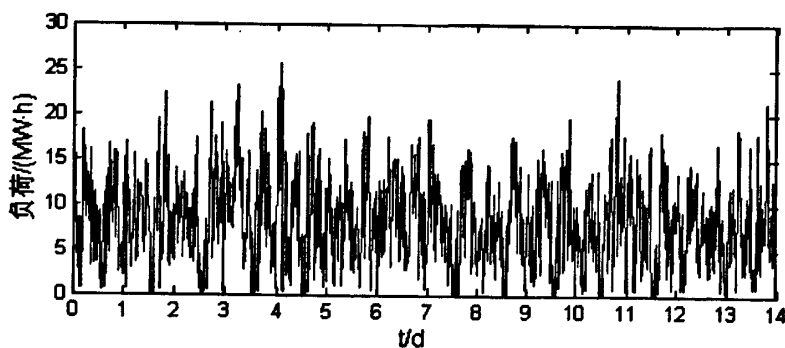
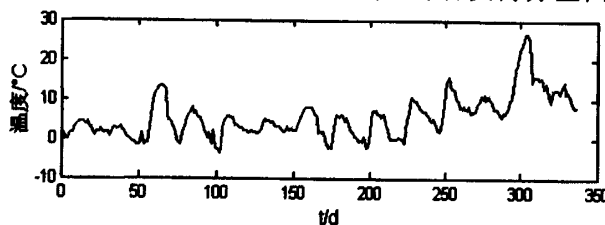


图 4-1 历史电铁负荷数据

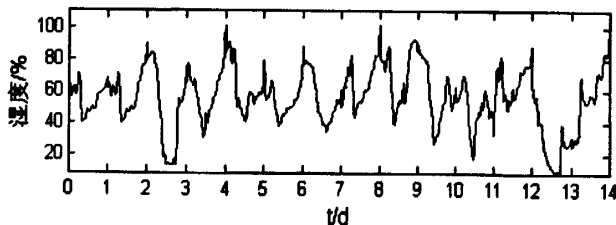
电铁负荷与许多的其他电力负荷有很多的不同之处，电铁负荷的电力需求主要由运输计划确定，对于不同的列车运行情况其网上取流变化是很大的，一般来说，在电气化铁路上电力机车的取流很难保持 30S 平稳不变，可以在短时间内从 0 变化到满负荷或过负荷，也可以从很大的取流快速降为 0。

通过统计计算，该线路的日平均负荷约为 $8\text{MW}\cdot\text{h}$ ；日最大峰谷差约为 $25.47\text{MW}\cdot\text{h}$ ，日最小峰谷差约为 $17.29\text{MW}\cdot\text{h}$ ，日平均峰谷差约为 $20.16\text{MW}\cdot\text{h}$ ；该条线路最大日负荷率约为 45.94%，最小日负荷率约为 31.98%，平均日负荷率约为 38.53%；半小时最大负荷与日平均负荷之比约在 1.6-2.9 之间，10min 最大负荷与日平均负荷之比约在 1.8-3.9 之间，1min 最大电流与半小时最大电流之比约为 2。由此可以发现电铁负荷波动很大，变化很快。

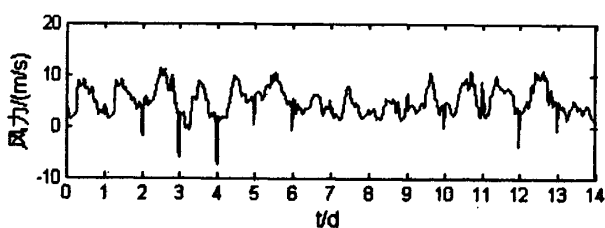
从影响负荷的基本影响因素考虑出发，本文选取了实时温度、实时湿度、实时风速三个气象因素，来分析其与电铁负荷的关系；同时考虑到电铁负荷的特殊性，又选取了每日客车流量来分析与电铁负荷的关系。由于该地区实时气象因素只能提供每日 24 点的数据，同时统计了该线路每日的客车流量，每日 24 点，因此需要对数据进行三次样条插值处理成为 96 点的曲线，与各负荷分量曲线相匹配。



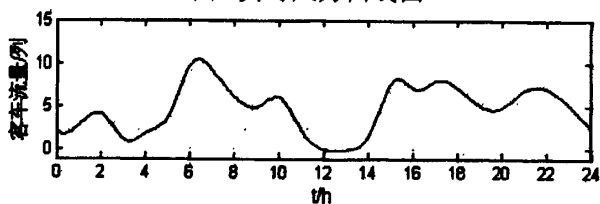
(a) 实时温度曲线图



(b) 实时湿度曲线图



(c) 实时风力曲线图



(d) 客车流量曲线图

图 4-2 各影响因素曲线图

采用 Spearman 相关系数来分析电铁负荷与各种因素之间的相关性。图 4-3 为每日实时温度、实时湿度、实时风速、客车流量与电铁负荷之间 Spearman 相关系数。

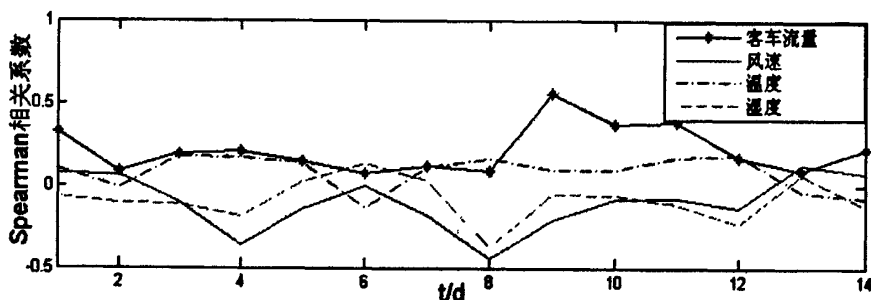


图 4-3 电铁负荷与各影响因素的 Spearman 相关系数

从图中可以发现，电铁负荷受实时气象因素影响较小，受列车运行计划的影响较大，因此在对电铁负荷进行预测时，将列车的运行计划作为考虑因素之一。

4.2 分量一预测模型

4.2.1 最小二乘支持向量机

支持向量机根据结构风险最小化 (SRM) 准则取得最小的实际风险，其拓扑结构由支持向量决定，克服了传统神经网络拓扑结构的选择在很大程度上依赖设计者经验的缺点，较好地解决了小样本、非线性、高位数和局部极小点等实际问题，具有很强的泛化能力^[78-80]。最小二乘支持向量机是标准支持向量机的一种扩展，优化指标采用平方项，并用等式约束代替标准支持向量机的不等式约束，即将二次规划问题转化为线性方程组求解，降低了计算复杂性，加快了求解速度，提高了实时性。其训练通过下式完成：

$$\min \frac{1}{2} W^T W + \frac{C}{2} \sum_{i=1}^M \xi_i^2 \quad (4-1)$$

约束条件为:

$$y_i(W^T g(x_i) + b) = 1 - \xi_i, i = 1, \dots, M \quad (4-2)$$

式中, 若 x_i 属于类 1, 则 $y_i = 1$; 若 x_i 属于类 2, 则 $y_i = -1$; W 为 1 维权重矢量; $g(x)$ 是将 x 从输入空间映射到特征空间的函数; ξ_i 是 x_i 的松弛系数; C 是边际系数。

需要指出的是, 与传统支持向量机不同, LS-SVM 中的松弛变量可以是负数, $\xi_i \geq 1$, 则 x_i 被错分; $\xi_i < 1$, x_i 被正确分类。

引入拉格朗日乘子 α_i , 将上式转化为无约束目标函数:

$$Q(W, b, \alpha, \xi) = \frac{1}{2} W^T W + \frac{C}{2} \sum_{i=1}^M \xi_i^2 - \sum_{i=1}^M \alpha_i \{y_i [W^T g(x_i)] + b - 1 + \xi_i\} \quad (4-3)$$

式中, $\alpha = (\alpha_1, \dots, \alpha_M)$, $\xi = (\xi_1, \dots, \xi_M)$ 。

令上式对 W , b , ξ 的偏导为零, 结合式(4-3)中的约束条件, 可得下列优化条件:

$$W = \sum_{i=1}^M \alpha_i y_i g(x_i) \quad (4-4)$$

$$\sum_{i=1}^M \alpha_i y_i = 0 \quad (4-5)$$

$$\alpha_i = C \xi_i, i = 1, \dots, M \quad (4-6)$$

$$y_i [W^T g(x_i)] + b - 1 + \xi_i = 0, i = 1, \dots, M \quad (4-7)$$

由式(4-6)可以发现, 与传统的支持向量机不同, α_i 可以为负数。将式(4-4)、式(4-5)代入式(4-7), 将它和式(4-6)用矩阵形式表示:

$$\begin{pmatrix} \Omega & y \\ Y^T & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ b \end{pmatrix} = \begin{pmatrix} I \\ 0 \end{pmatrix} \quad (4-8)$$

$$\Omega \alpha + y b = I \quad (4-9)$$

$$Y^T \alpha = 0 \quad (4-10)$$

式中, $I = (1, \dots, 1)^T$ 是一个 M 维向量, $y = (y_1, \dots, y_M)^T$, $\Omega_{ij} = y_i y_j g^T(x_i) g(x_j) + \frac{\delta_{ij}}{C}$,

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

为避免直接处理变量的空间映射, 令核函数 $H(x, x') = g^T(x) g(x')$ 。求解式(4-10)可解得 α 与 b , 考虑到 Ω 的对焦元素为 $1/C$, Ω 为正定阵, 因此:

$$\alpha = \Omega^{-1}(I - yb) \quad (4-11)$$

$$b = (Y^T \Omega^{-1} y)^{-1} Y^T \Omega^{-1} I \quad (4-12)$$

将式(4-11)代入式(4-12)即可解得 α 。

由上述过程可以发现,最小二乘支持向量机将 2 范数软间隔支持向量机中的不等式约束转换为等式约束,其训练过程也由二次规划问题求解转化为线性方程组的求解,简化了计算的复杂性。

4.2.2 时间序列

由美国学者 George Box 和英国统计学家 Gwilym Jenkins 在 1968 年提出的时间序列模型,被认为是最经典、最系统的一类预测方法,称为 Box-Jenkins 预测方法,广泛地应用于各类预测之中,也是短期负荷预测的常用方法。Box-Jenkins 预测方法把时间序列看作随机过程来研究和描述,其基本思想是:首先假设所分析的时间序列是由某个随机过程产生的,然后用时间序列的原始数据建立一个描述该过程的模型,并进行参数估计,运用所建立的模型,在已知时间序列在过去和现在的观测值的情况下,求得时间序列未来的预测值^[81]。

根据 Box-Jenkins 方法,可将随机时间序列的模型分为 3 类,它们分别是自回归模型(AR)、滑动平均模型(MA)和自回归-滑动平均模型(ARMA)。由 AR、MA、ARMA 模型描述的时间序列称为平稳时间序列。如果研究的时间序列是非平稳的,必须先进行平稳化处理,这主要通过下面两种方法来实现:

(1) 对原非平稳时间序列进行有序差分变换,差分后得到的平稳序列可以用 AR、MA、ARMA 模型来描述就是累积式自回归-滑动平均模型(ARIMA);

(2) 若研究的时间序列具有季节性变化趋势,可对其进行季节性差分变换,得到季节性 ARIMA 模型。

时间序列预测分为以下四步:

(1) 模型识别

根据 MA(q)、AR(p)、ARMA(p, q) 三种随机过程偏相关函数的特点进行模型识别。对时间序列 X_t 进行相关性分析。计算序列的样本自相关函数和样本偏相关函数,通过它们的截尾性和拖尾性初步选择模型的结构(表 4-1)。

表 4-1 ARMA(p, q)模型的序列特征表

项目	MA(q)	AR(p)	ARMA(p, q)
自相关函数	q 步截尾	拖尾	拖尾
偏相关函数	拖尾	p 步截尾	拖尾

(2) 模型定阶

ARMA(p, q) 中 p 、 q 的确定通过 AIC 信息准则来实现的。AIC 的英文全称是 Akaike's Information Criterion 译为赤池信息准则。AIC 准则是为了克服模型的过度

敏感而提出的。它利用似然函数估计值最大的原则来确定模型阶数。

对于 ARMA 序列的 AIC 准则定义为：

$$AIC(n,m) = \ln \hat{\sigma}_a^2 + 2(n+m+1)/N \quad (4-13)$$

设对于时间序列，其 p 的上界为 P ， q 的上界为 Q ，对于每一组 (n,m) , $0 \leq n \leq P, 0 \leq q \leq Q$ 计算 AIC 函数。若 $AIC(p,q) = \min_{0 \leq n, m \leq L} AIC(n,m)$ 则定 ARMA 模型的阶数为 (p,q) ，其中 $\hat{\sigma}_a^2$ 是相应的 ARMA 序列的 σ_a^2 的极大似然估计值。

(3) 模型参数的估计

模型的阶数确定后，运用矩估计、极大似然估计、最小二乘估计等计算模型的参数 $(\hat{a}_1, \dots, \hat{a}_p)$, $(\hat{b}_1, \dots, \hat{b}_q)$ 。

(4) 模型检验

计算模型的残差 $\hat{\varepsilon}_t$ ，如果残差序列可以通过白噪声检验，则认为模型是合理的，否则，进一步改进模型。

利用时间序列分析方法对电铁负荷序列进行分析，找出影响预测的主要因素，从而选择支持向量机的输入变量，克服了在选择输入变量时需要认为主观判断的缺点，进一步提高预测的精度。

4.2.3 灾变遗传算法理论

4.2.3.1 遗传算法

遗传算法是 JHOLLAND 于 1975 年受生物进化论的启发提出的，它是模拟现实生物界中“优胜劣汰，适者生存”的自然进化原理来进行全局寻优的。

传统的遗传算法包含以下四个部分：

(1) 初始种群的生成。初始种群一般由工程问题的本身来决定。可以由使用者开始给定，这样可以在算法的开始就得到较优的种群，可以大大缩短寻优的时间。但这种情况是比较少见的。大体上是给定的范围内随机生成的。一般采用随机数或伪随机数序列。生成的种群有采用二进制编码和浮点编码两种形式，在应用上依情况而定。二进制编码的优点在于遗传操作上比较清晰并有图示理论做引导。而浮点编码的计算精度较高^[82-83]。

(2) 选择操作。首先将上述二进制编码的个体进行解码。然后根据给定的适应值函数分别计算种群内基因的适应值。判断是否有符合终止条件的个体存在。采取一定的方法（一般采用轮盘赌方法）选择出两个适值较高的个体。

(3) 交叉操作。随机选择上述被选中个体中的一点，进行交叉操作（如图 4-4 所示），生成两个新的个体。

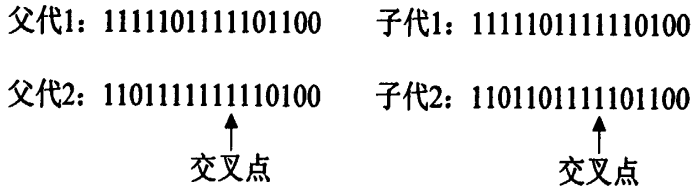


图 4-4 交叉操作示意图

(4) 变异操作。这是模仿自然进化中基因突变的操作。因为交叉操作不能产生新的基因段，如果初始种群中不包括这一段的基因，则基因算法的全局寻优不能成功。遗传算法中的突变是指随机的将所选中的个体的基因取反，以产生与原始种群不同的基因。而且变异操作也可以重新补充由于交叉操作而被破坏的有用基因。

由上述介绍的遗传算法的基本操作可见，“早熟”问题的产生是由于在遗传的后期，由于交叉操作和选择操作的作用，优良个体急剧的增加使种群失去多样性，而变异操作由于其发生的概率远远小于交叉操作发生的概率从而不能产生足够数量的有用的个体，从而造成程序陷入局部最优，达不到全局最优解的现象。而如果盲目地增大变异操作的概率则使遗传算法向随机算法的方向发展，最终丧失了遗传算法的优越性。另外如果所求问题在工程上要求的时间比较严格，则收敛时间长也是遗传算法的缺陷之一。

4.2.3.2 灾变遗传算法

为了进一步提高遗传算法的全局搜索性能，需要一种力量来改变初始种群集中分布情况。灾变往往是对绝大多数生物的灭顶之灾，只有适应能力特别强的物种或者个体才能得以生存，并进入下一代的繁衍。文献[84]指出，灾变后的最优解在新一轮进化中的作用与将灾变前的 s 个个体直接加入群体相当，因此，灾变可以在不扩大群体规模的情况下隐含扩大的群体规模，从而提高了遗传算法的全局性能。文献[85]提出了突然增大变异概率对不同个体实施不同的变异，以产生不同数目的后代，但变异率何时提高、提高多少才能达到要求，需要大量的经验和计算。文献[86]引入了尖点灾变模型，得到了灾变临界条件。

本文采用了锦标赛方法，它强调的是适应度的相对值，可以保证“灾变”之后，原先得到的局部最优解能被选到繁殖池中来，经过与重新产生的其他区域中的个体进行交叉之后，能以更大的概率跳出该局部最优解所在区域，找到全局最优。

程序流程图见下图

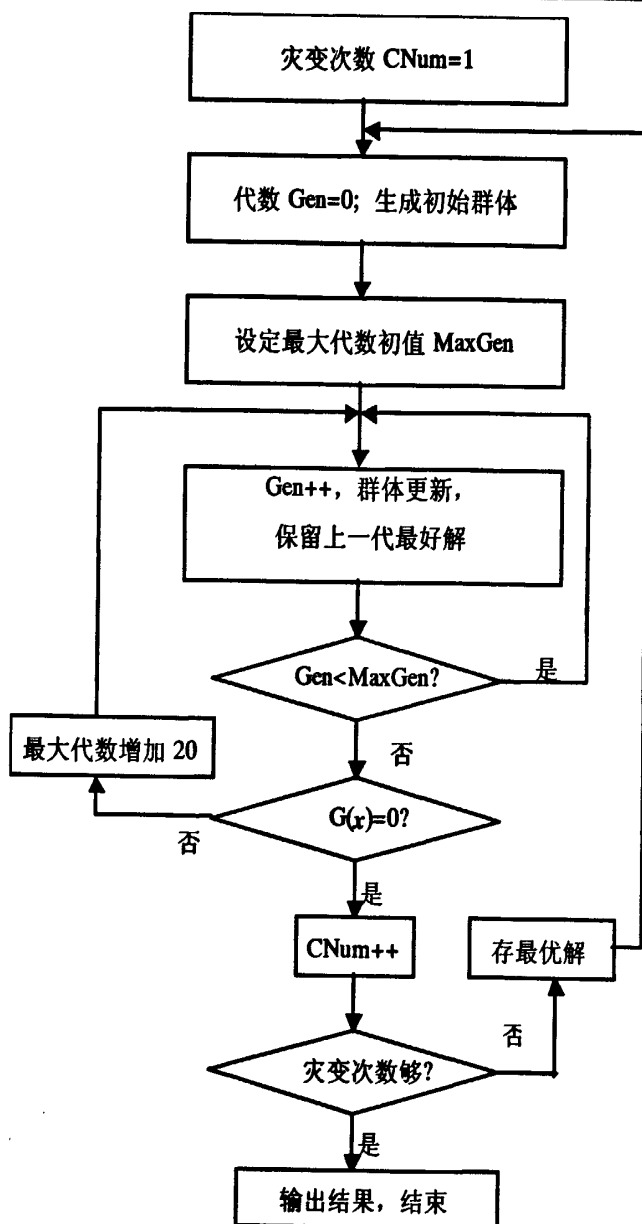


图 4-5 程序流程图

4.2.4 预测模型建立

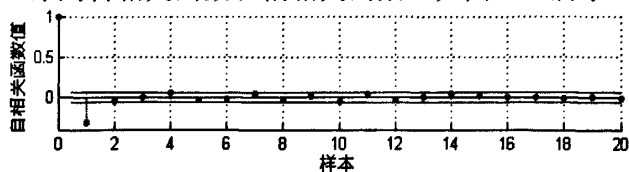
4.2.4.1 数据来源

根据上一章的分析，得出原始负荷序列所分解的分量一波动与电铁负荷的特点相近，认为分量一反映电铁负荷的波动特点，因此选取该段负荷分量序列作为分析对象。图 3-6 中可以发现该分量在短时间内的变化很快，呈现出无规律的大升和大降，并且具有一定的冲击性，波动剧烈。结合电铁负荷的影响因素特点，考虑每日客车流量对电铁负荷变化的影响。

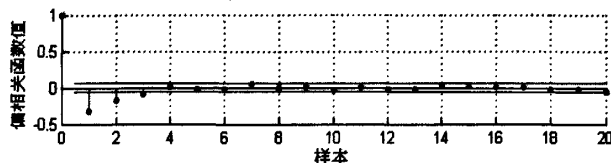
4.2.4.2 模型建立

(1) 确定最小二乘支持向量机输入变量。

首先计算分量一序列自相关函数和偏相关函数，如图 4-6 所示。



(a) 前 20 个自相关函数值



(b) 前 20 个偏相关函数值

图 4-6 前 20 个自相关和偏相关函数值

从图中可以看出该序列的自相关函数很快衰减到零，序列足够平稳。因此，可将序列识别为 $ARMA(p, q)$ 模型。由于模型识别具有很大的灵活性，为得到更合理的模型，对 $p=1,2,3; q=0,1$ 多组阶数进行参数估计和模型检验，检验结果如下：

表 4-2 模型检验

(p, q)	(3,1)	(3,0)	(2,1)	(2,0)	(1,1)	(1,0)
AIC	7.910876	8.377775	7.920764	8.361122	7.941739	8.565885
SC	7.933001	8.388827	7.937343	8.371287	7.951895	8.570964

通过 AIC 和 SC 准则确定模型为 $ARMA(3, 1)$ 。因此可以把最小二乘支持向量机的输入变量确定为电铁负荷序列 $ARMA(3, 1)$ 的最近 3 天历史值以及最近 1 天残差，同时考虑到列车运行计划对电铁负荷的影响较大，因此也将其作为输入变量，因此确定预测模型的输入变量共有 5 个。

(2) 确定最小二乘支持向量机的两个参数：边际系数 C 和核函数的控制半径 γ 。

考虑到 RBF 核函数具有一定的代表性，它在处理输入空间线性不可分数据时具有较好的效果，因此选择 RBF 型的高斯核函数 $H(x, x') = \exp(-\gamma \|x - x'\|^d)$ 。式中 d 为阶数； γ 是一个控制半径的正数。需要选择两个参数： d 、 γ 。但通常 d 一般设为 2，此时核函数值为两个样本的欧式距离相关， γ 控制核函数的宽度。因此共需要选择两个参数：边际系数 C 和核函数的控制半径 γ 。

利用灾变遗传算法进行优化的依据是适应度函数，适应度函数是个体空间到实数空间的一个映射这里一个个体代表最小二乘支持向量机算法的一组参数，个体对应的适应度函数则是该组参数下的算法性能。即适应度函数选取 LOO 误差，此时优化函数为：

$$\min_{C, \gamma} F(C, \gamma) = \varepsilon_{LOO} \quad (4-14)$$

$$\text{约束条件为} \begin{cases} C_{low} \leq C \leq C_{up} \\ \gamma_{low} \leq \gamma \leq \gamma_{up} \end{cases} \quad (4-15)$$

(3) 预测步骤。

① 利用时间序列和负荷影响分析结果确定最小二乘支持向量机输入变量, 并进行归一化处理, 然后将样本集分为训练样本集合预测样本集;

② 将染色体和适应度分别对应于最小二乘支持向量机的两个参数边际系数 C 、核函数的控制半径 γ 和目标函数, 固定码长 (边际系数 C 为 5 位, 核宽度系数 γ 为 3 位), 并确定参数变化范围为 $C \in [1, 9600]$, $\gamma \in [0.1, 10]$;

③ 染色体初始化, 形成初始种群;

④ 进行遗传操作: 选择、交叉、变异, 计算每一代染色体的适应值;

⑤ 判断是否进行灾变操作, 若是则进行灾变, 补全种群规模, 若否则进入下一步;

⑥ 判断是否收敛或达到迭代次数, 若是则输出结果并结束, 若否则转④;

⑦ 采用 RBF 核函数, 利用训练样本进行训练, 建立负荷预测模型;

⑧ 利用测试样本进行预测。

4.2.5 预测结果分析

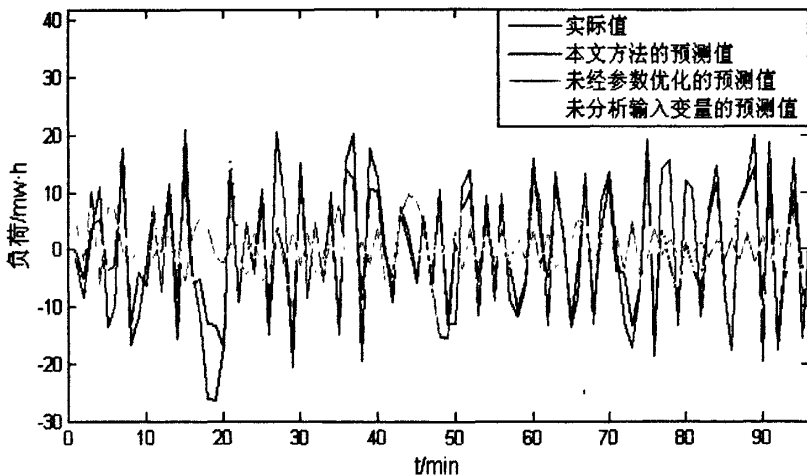


图 4-7 实际值和预测值比较图

本文采用相对误差、平均相对误差作为误差评价指标。

$$E_i = \frac{Y_i - \hat{Y}_i}{Y_i} \quad (4-16)$$

$$E_{MRE} = \frac{1}{N} \sum_{i=1}^N |E_i| \quad (4-17)$$

其中: Y_i 表示实际负荷序列; \hat{Y}_i 表示负荷预测值, N 表示预测数据的点数, E_i 表

示 i 点预测相对误差, E_{MRE} 预测的平均相对误差。利用上文介绍的方法与未进行输入变量分析和未进行参数优化分别对测试数据进行预测。预测发现, 利用本文方法的最小相对误差为 0.039%, 平均相对误差为 25.93%, 未进行输入变量分析时的最小相对误差为 1.01%, 平均相对误差为 87.61%, 说明通过时间序列分析和因素影响分析, 得出影响电铁负荷变化的重要因素, 确定输入变量, 较主观选择变量更加科学, 更加准确; 而未进行参数优化时的最小相对误差为 0.81%, 平均相对误差为 94.22%, 说明利用灾变遗传算法对模型参数进行优化可以有效选择最优参数, 进一步提高预测精度。

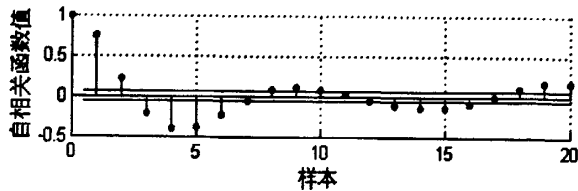
由于该实例的数据序列与电铁负荷序列的非线性特征类似, 波动规律相似, 影响因素也类似, 因此可以将该方法运用到实际的电铁负荷预测中。试验证明, 该方法运用到实际的电铁负荷预测也同样可以得到较好的效果。

4.3 分量二预测模型

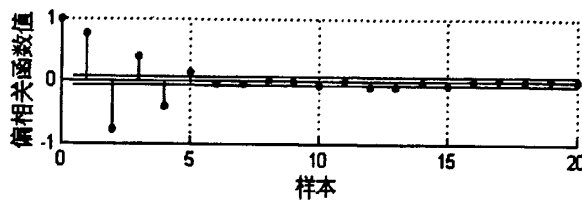
由于分量二受气象因素影响较小, 且本身波动规律性不强, 因此可用时间序列进行预测。

4.3.1 时间序列特征分析

分量二如 3-6 图所示, 可以发现分量二的波动规律性不强, 变化较为剧烈。计算序列自相关函数值和偏相关函数值, 如图 4-8 所示。



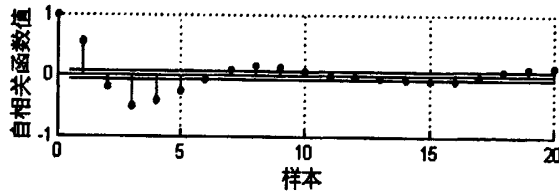
(a) 前 20 个自相关函数值



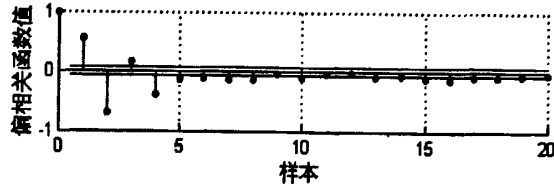
(b) 前 20 个偏相关函数值

图 4-8 前 20 个自相关和偏相关函数值

从图中可看出, 自相关函数不能很快衰减到零, 分量二序列非平稳, 需要进行平稳处理。因此对该序列进行一阶有序差分变换, 所得差分序列的自相关函数值和偏相关函数值如图 4-9 所示。



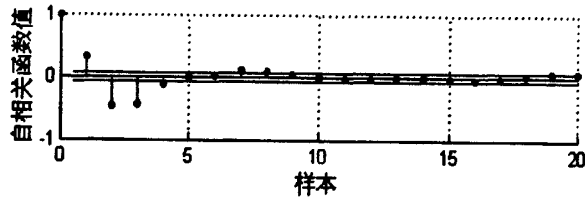
(a) 前 20 个自相关函数值



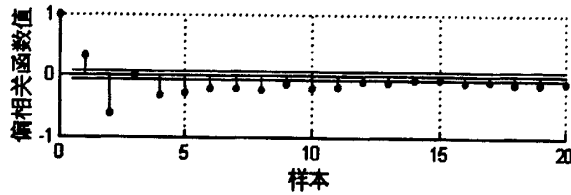
(b) 前 20 个偏相关函数值

图 4-9 前 20 个自相关和偏相关函数值

由图中可知，一阶差分不能完全消除序列的非平稳性，因此需要进行二阶差分，经过二阶差分后的自相关函数值及其偏相关函数值如图 4-10 所示。



(a) 前 20 个自相关函数值



(b) 前 20 个偏相关函数值

图 4-10 前 20 个自相关和偏相关函数值

此时该差分序列的均值趋近于 0，且自相关函数很快衰减到 0，序列变得足够平稳。由于 ARMA 模型形式都存在序列均值为 0 的假设，否则应做必要的转换。若序列的样本平均数为 \bar{y} ，均指标准误差为 $s_{\bar{y}}$ ，则当 \bar{y} 落入 $\pm 2s_{\bar{y}}$ 时，认为序列满足 0 均指假设。均值标准误差采用式 (4-18) 近似计算：

$$s_{\bar{y}} \approx \left[\frac{\sigma_y^2}{n} \left(1 + 2 \sum_{k=1}^M \hat{\rho}_k \right) \right]^{\frac{1}{2}} \quad (4-18)$$

其中， σ_y^2 是序列样本方差， $\hat{\rho}_k$ 为样本方差相关系数， $k=1$ 到 M 表示前 M 个显著不为 0 的样本自相关系数， n 是序列观测量。因此计算得到该序列二阶差分后样本平均数 \bar{y} 是 0.0064，均值标准差 $s_{\bar{y}}$ 为 0.41，序列均值与 0 无显著差异，因此分量二序列识别为 ARIMA(p, d, q) 模型。

4.3.2 模式识别

因为经过二阶逐期差分，故 $d=2$ 。观察二阶差分序列的偏、自相关函数图，如图 4-10， $p=2$ 比较适合，自相关图显示 $q=3$ 。考虑到 AR 模型是线性方程估计，相对于 MA 和 ARMA 模型的非线性估计容易，且参数意义便于解释，故实际建模时常希望用高阶的 AR 模型替换相应的 MA 或 ARMA 模型。综合考虑，可供选择的 (p,d,q) 组合有：(2, 2, 3)、(3, 2, 2)、(4, 2, 1)、(5, 2, 0)。将各模型进行参数估计和模型检验，检验结果汇总列入下表。

表 4-3 模型检验

(p, d, q)	(2, 2, 3)	(3, 2, 2)	(4, 2, 1)	(5, 2, 0)
AIC	6.478864	6.481063	6.350109	6.692475
SC	6.506521	6.508745	6.377816	6.720208

通过 AIC 和 SC 准则选择模型 ARIMA(4, 2, 1) 比较合适。

4.3.3 模型建立

将其通过检验的模型改写成为：

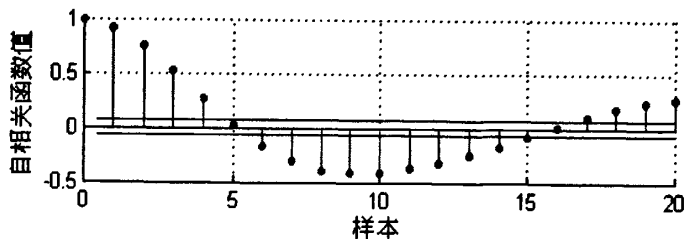
$$(1-1.101435B + 1.164467B^2 - 0.530394B^3 + 0.358291B^4)(1-B^2)y_t = (1+1.012414B)\alpha_t \quad (4-19)$$

4.4 分量三预测模型

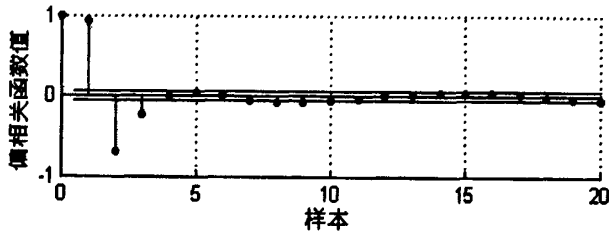
由于分量三的气象因素影响较为复杂，实时气象因素对该负荷分量影响不稳定，不能运用多元回归的预测方法，因此考虑运用最小二乘支持向量机的方法对其进行预测。

首先确定最小二乘支持向量机的输入变量。

计算分量三序列自相关函数和偏相关函数，如图 4-11 所示。



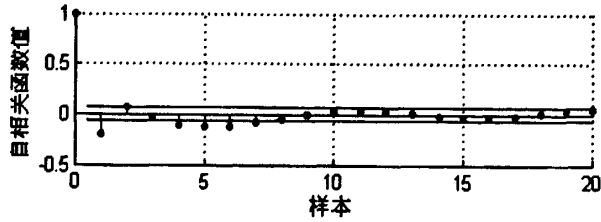
(a) 前 20 个自相关函数值



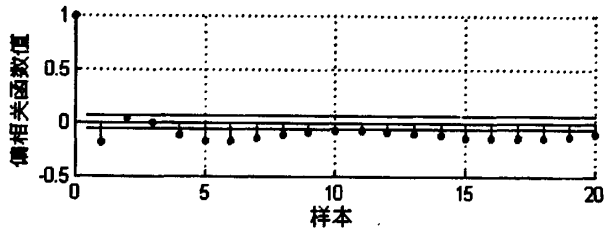
(b) 前 20 个偏相关函数值

图 4-11 前 20 个自相关和偏相关函数值

从自相关函数值可以看出，该序列不平稳，需要差分消除非平稳性。经过二阶差分以后的自相关系数和偏相关系数如下图。



(a) 前 20 个自相关函数值



(b) 前 20 个偏相关函数值

图 4-12 前 20 个自相关和偏相关函数值

从图中可以看出该序列的自相关函数很快衰减到零，序列足够平稳。因此，可将序列识别为 $ARIMA(p, q)$ 模型，其中 $p=1, 2; q=0, 1; d=2$ 。通过检验，确定 $p=1; q=1; d=2$ ，因此最小二乘支持向量机其中两个输入变量为最近 1 天历史值、最近 1 天残差。通过第三章的分析，分量三与实时气象因素关联度较大，因此将实时气象也作为输入变量。综上可以确定出最小二乘支持向量机的输入变量为最近 1 天历史值、最近 1 天残差以及实时气象因素。

然后利用遗传算法对模型的参数进行优化，最后采用 RBF 核函数，利用训练样本进行训练，建立负荷预测模型。

4.5 分量四预测模型

根据第三章的分析，分量四波动规律性较强，与实时气象因素的关联度较大，且关联规律较稳定，因此采用多元线性回归的方法对其建立预测模型。

4.5.1 多元线性回归理论

4.5.1.1 多元线性回归模型

设 x_1, x_2, \dots, x_p 是 $p (p > 1)$ 个线性无关的可控变量, y 是随机变量, 它们之间的关系为:

$$\begin{cases} y = b_0 + b_1 x_1 + \dots + b_p x_p + \varepsilon \\ \varepsilon \sim N(0, \sigma^2) \end{cases} \quad (4-20)$$

这里 $b_0, b_1, \dots, b_p, \sigma^2$ 都是与 x_1, x_2, \dots, x_p 无关的未知数参数, ε 是随机误差(或随机干扰), 这就是 p 元线性回归模型。式(4-20)两端去数学期望, 得

$$E y = b_0 + b_1 x_1 + \dots + b_p x_p \quad (4-21)$$

显然 $E y$ 是 x_1, x_2, \dots, x_p 的函数。式(4-21)称为回归平面方程, 其中 b_0, b_1, \dots, b_p 成为回归系数。

4.5.1.2 未知参数估计

对变量 x_1, x_2, \dots, x_p 和 y 作 n 次独立观察, 可得容量为 n 的一个样本 $(x_{i1}, x_{i2}, \dots, x_{ip}, y_i) (i = 1, 2, \dots, n)$ 。在负荷预测中, 这些常量是过去的历史资料。由式(4-20)可得

$$\begin{aligned} y_1 &= b_0 + b_1 x_{11} + b_2 x_{12} \dots + b_p x_{1p} + \varepsilon_1 \\ y_2 &= b_0 + b_1 x_{21} + b_2 x_{22} \dots + b_p x_{2p} + \varepsilon_2 \\ &\dots \\ y_n &= b_0 + b_1 x_{n1} + b_2 x_{n2} \dots + b_p x_{np} + \varepsilon_n \end{aligned} \quad (4-22)$$

这里各个 $\varepsilon_i \sim N(0, \sigma^2)$, 且 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 是相互独立的。

为了数学处理上的方便, 将式(4-22)可用矩阵形式来表示。记

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix},$$

$$B = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_p \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

于是, 式(4-22)可写为

$$Y = XB + \varepsilon \quad (4-23)$$

式中, ε 是 n 维随机向量, 它的分量是相互独立的。则式(4-23)为 p 元线性回归模型的矩阵形式。

下面用最小二乘法求未知参数 b_0, b_1, \dots, b_p 的估计量。作离差平方和

$$Q = \sum_{i=1}^n (y_i - b_0 - b_1 x_{i1} - \dots - b_p x_{ip})^2 \quad (4-24)$$

选取 $\hat{b}_0, \hat{b}_1, \dots, \hat{b}_p$, 使当 $b_0 = \hat{b}_0, b_1 = \hat{b}_1, \dots, b_p = \hat{b}_p$ 时, 离差平方和 Q 达到最小。依据高等数学中求多元函数极值的方法, 取 Q 分别关于 b_0, b_1, \dots, b_p 的偏导数, 并令它们都等于零, 得方程组

$$\begin{cases} \frac{\partial Q}{\partial b_0} = -2 \sum_{i=1}^n (y_i - b_0 - b_1 x_{i1} - \dots - b_p x_{ip}) = 0 \\ \frac{\partial Q}{\partial b_1} = -2 \sum_{i=1}^n (y_i - b_0 - b_1 x_{i1} - \dots - b_p x_{ip}) x_{i1} = 0 \\ \dots \\ \frac{\partial Q}{\partial b_p} = -2 \sum_{i=1}^n (y_i - b_0 - b_1 x_{i1} - \dots - b_p x_{ip}) x_{ip} = 0 \end{cases}$$

解此方程组得到的不是 b_0, b_1, \dots, b_p 的真实值, 而是估计值 $\hat{b}_0, \hat{b}_1, \dots, \hat{b}_p$, 故将此方程组化简并改写为

$$\begin{cases} n\hat{b}_0 + \left(\sum_{i=1}^n x_{i1}\right)\hat{b}_1 + \left(\sum_{i=1}^n x_{i2}\right)\hat{b}_2 + \dots + \left(\sum_{i=1}^n x_{ip}\right)\hat{b}_p = \sum_{i=1}^n y_i \\ \left(\sum_{i=1}^n x_{i1}\right)\hat{b}_0 + \left(\sum_{i=1}^n x_{i1}^2\right)\hat{b}_1 + \left(\sum_{i=1}^n x_{i1}x_{i2}\right)\hat{b}_2 + \dots + \left(\sum_{i=1}^n x_{i1}x_{ip}\right)\hat{b}_p = \sum_{i=1}^n x_{i1}y_i \\ \dots \\ \left(\sum_{i=1}^n x_{ip}\right)\hat{b}_0 + \left(\sum_{i=1}^n x_{ip}x_{i1}\right)\hat{b}_1 + \left(\sum_{i=1}^n x_{ip}x_{i2}\right)\hat{b}_2 + \dots + \left(\sum_{i=1}^n x_{ip}^2\right)\hat{b}_p = \sum_{i=1}^n x_{ip}y_i \end{cases} \quad (4-25)$$

这个方程组称为正规方程组。容易验证

$$X'X = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & \dots & x_{n1} \\ \dots & \dots & & \dots \\ x_{1p} & x_{2p} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \dots & \dots & & \dots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^n x_{i1} & \dots & \sum_{i=1}^n x_{ip} \\ \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i1}^2 & \dots & \sum_{i=1}^n x_{i1}x_{ip} \\ \dots & \dots & & \dots \\ \sum_{i=1}^n x_{ip} & \sum_{i=1}^n x_{ip}x_{i1} & \dots & \sum_{i=1}^n x_{ip}^2 \end{bmatrix}$$

$$X'Y = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & \dots & x_{n1} \\ \dots & \dots & & \dots \\ x_{1p} & x_{2p} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{i1}y_i \\ \vdots \\ \sum_{i=1}^n x_{ip}y_i \end{bmatrix}$$

又记 B 的估计值矢量为 \hat{B} 。于是, 正规方程组式 (4-25) 可写成矩阵形式为

$$X'X\hat{B} = X'Y \quad (4-26)$$

由于 x_1, x_2, \dots, x_p 线性无关, 所以, X 的列向量线性无关, 这样方程式 (4-26) 的系数矩阵 $X'X$ 是可逆的。故可得到式 (4-26) 的唯一解为

$$\hat{B} = \begin{bmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \vdots \\ \hat{b}_p \end{bmatrix} = (X'X)^{-1}X'Y \quad (4-27)$$

这就是回归模型参数估计表达式。

将得到的 $\hat{b}_0, \hat{b}_1, \dots, \hat{b}_p$ 代入式 (4-27) 的右端, 并记作

$$y = \hat{b}_0 + \hat{b}_1x_1 + \dots + \hat{b}_px_p$$

称这个方程为 p 元线性回归方程 (或称为回归平面方程), 简称回归方程。这里的 $\hat{b}_0, \hat{b}_1, \dots, \hat{b}_p$ 称为回归方程的系数。

4.5.2 模型建立

针对分量四的波动特点与影响因素特性, 分量四与实时温度、实时湿度、实时风速这三个气象因素相关性较大, 因此将该三个影响因素作为模型自变量, 同时考虑历史负荷对预测值的影响, 将预测日的前一天历史负荷作为自变量之一。因此该分量模型有四个自变量, 分别是: 前一天历史负荷 *load*、预测日实时温度预报值 *temp*、预测日实时湿度预报值 *humi*、预测日实时风速预报值 *wind*。然后建立多元线性回归模型, 可得到如下结果。

表 4-4 拟合模型情况

项目	相关系数 R	决定系数 R^2	校正后决定系数 \bar{R}^2	回归标准误差	F 值	P 值
数值	0.949	0.900	0.895	8.658	204.247	0.000

表 4-4 为所拟合模型的情况简报, 根据表中显示, R^2 和 \bar{R}^2 接近于 1, 表明模型的拟合效果非常好。通过模型检验, 得出所用的回归模型 F 值为 204.247, P 值为 0.000, 因此该回归模型是有统计学意义的。

表 4-5 模型系数及其检验

项目	系数 b	系数标准误差	T 值	P 值
常数	44.365	12.460	3.561	0.001
前一天历史负荷 <i>load</i>	0.806	0.037	21.748	0.000
实时温度预报值 <i>temp</i>	-6.334	1.078	-5.876	0.000
实时湿度预报值 <i>humi</i>	-0.311	0.161	-1.929	0.057
实时风速预报值 <i>wind</i>	3.211	0.857	3.747	0.000

上表 4-5 给出了包括常数项在内的所有系数的检验结果, 用的是 t 检验, 同时还给出系数。可见常数项和各变量都是有统计学意义的。根据表的结果可得到下面的估计方程:

$$\hat{IMF}_4 = 44.365 + 0.806 \times load - 6.334 \times temp - 0.311 \times humi + 3.211 \times wind \quad (4-28)$$

4.6 余量预测模型

由于余量的波动规律以及影响因素特点与分量四相类似，因此余量也采用多元回归进行预测建模。该分量模型同样为四个自变量，分别是：前一天历史负荷 *load*、预测日实时温度预报值 *temp*、预测日实时湿度预报值 *humi*、预测日实时风速预报值 *wind*。然后建立多元线性回归模型，可得到如下结果：

表 4-6 拟合模型情况

项目	相关系数 R	决定系数 R^2	校正后决定系数 \bar{R}^2	回归标准误差	F 值	P 值
数值	0.992	0.984	0.983	7.333	0.001	0.000

表 4-6 为所拟合模型的情况简报，根据表中显示， R^2 和 \bar{R}^2 接近于 1，表明模型的拟合效果非常好。通过模型检验，得出所用的回归模型 F 值为 0.001407，P 值为 0.000，因此该回归模型是有统计学意义的。

表 4-7 模型系数及其检验

项目	系数 b	系数标准误差	T 值	P 值
常数	206.150	31.477	6.549	0.000
前一天历史负荷 <i>load</i>	0.771	0.032	23.812	0.000
实时温度预报值 <i>temp</i>	4.522	0.861	5.253	0.000
实时湿度预报值 <i>humi</i>	-0.025	0.128	-0.197	0.844
实时风速预报值 <i>wind</i>	3.902	0.513	7.600	0.000

上表给出了包括常数项在内的所有系数的检验结果，用的是 t 检验，同时还给出标准化/未标准化系数。可见常数项和各变量都是有统计学意义的。根据表 4-7 结果可得到下面的估计方程：

$$\hat{r}_n = 206.150 + 0.771 \times load - 4.522 \times temp - 0.025 \times humi + 3.902 \times wind \quad (4-28)$$

4.7 预测模型重构及模型评估

设各分量的建模预测值分别为 \hat{IMF}_n ，余量的预测值为 \hat{r}_n 那么原始负荷的预测值为：

$$\hat{X} = \hat{IMF}_1 + \hat{IMF}_2 + \dots + \hat{IMF}_n + \hat{r}_n \quad (4-29)$$

按公式 (4-29) 求出负荷的预测值，并进行误差计算。

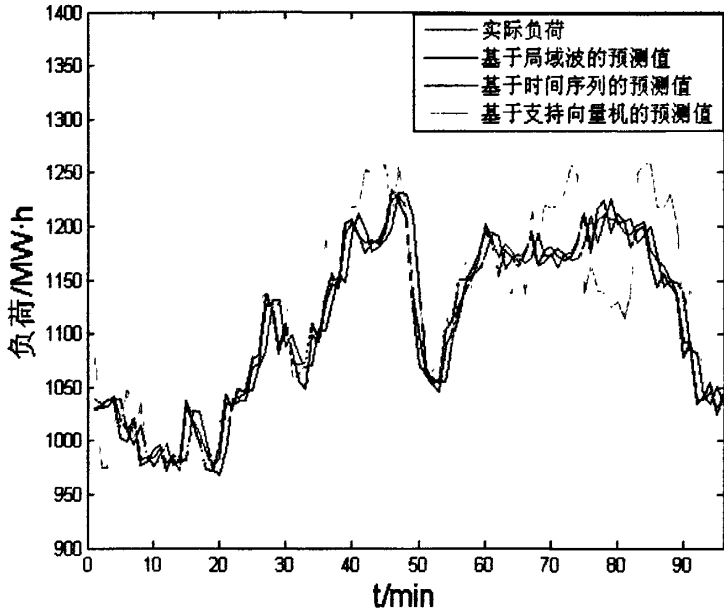
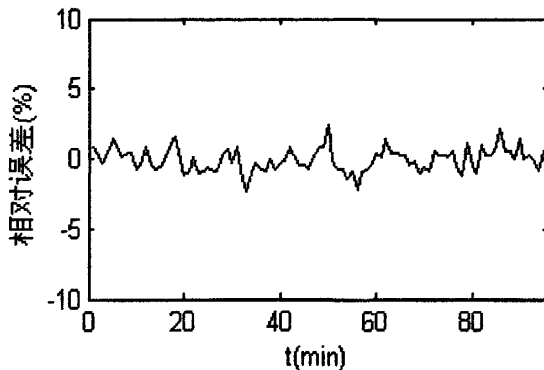


图 4-13 实际值和预测值比较图

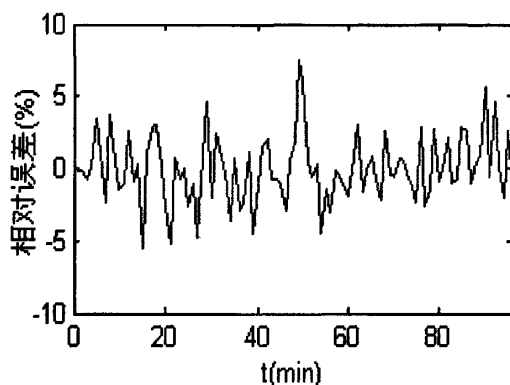
通过对预测数据的统计分析，采用本文基于局部建模的预测方法进行短期负荷预测的最大相对误差为 1.7%，最小相对误差为 0.30%，平均绝对百分误差为 0.68%；而同时本文利用同样的样本数据直接利用时间序列预测法和最小二乘支持向量机进行预测。利用时间序列预测法的最大相对误差为 7.48%，最小相对误差为 0.71%，平均相对误差为 1.79%；利用最小二乘支持向量机进行预测的最大相对误差为 7.97%，最小相对误差为 0.312%，平均相对误差为 2.99%。

通过对预测数据的误差计算及其统计分析，得出它们的误差曲线。

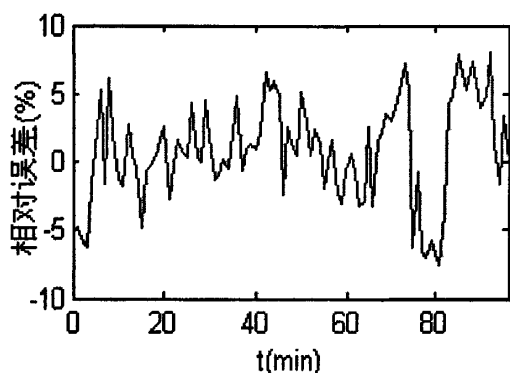
对比实际负荷曲线与误差曲线以及各中方法的误差曲线，可以看出，基于局域波的预测方法的误差总体较小，且误差波动很小，预测效果理想，而后两种基于全局建模的预测方法总体误差大，且波动大，预测效果较差。



(a) 基于局域波的预测相对误差曲线



(b) 基于时间序列的预测相对误差曲线



(c) 基于最小二乘支持向量机的预测相对误差曲线

图 4-14 预测相对误差曲线

4.8 误差分析

由于影响电力负荷波动的因素众多，同时随着国民经济的发展和人们消费思想观念的转变，使得预测的难度越来越大，下面就可能产生的误差原因进行分析。

(1) 冲击负荷的影响。由于冲击负荷波动性较大,变化规律难以把握。由预测理论知，任何预测方法的预测精度都会随着预测对象规律性强弱的变化而发生改变。若预测对象的规律性很强，大多数预测方法都可以取得较高的预测精度；相反，若预测对象的规律性很弱，任何预测方法都不可能突破预测对象本身的规律性所限定的预测精度极限。因此冲击负荷的预测精度难以提高，从而进一步影响了总体地区负荷的预测精度。

(2) 各影响因素对负荷波动影响较大。由于它的负荷基数较小，因此各种影响因素的在负荷波动上体现得更加明显，使得预测难度加大。

(3) 休息日的影响。虽然已经将周六、周日剔除，但是周六周日的气象情况对工作日的负荷肯定是会产生一定影响的，譬如一个普通的周末和一个降雨的周末对周一负荷肯定会有不同影响，周末的降雨往往会导致延续几天的凉爽天气，从而导致负荷较往常偏低。

(4) 其它实时气象因素的影响。气象原始数据中缺少降水、日照、气压等天气

因素，它们对负荷也有一定的影响。实际上，天气状况对负荷有一定影响，阴天负荷较大，晴天负荷较小；气压与负荷成反比关系，日平均气压越大时，负荷越小，平均气压越低，负荷越大；冬季白天云量与电力负荷有一定的正相关性。

4.9 本章小结

电铁负荷对电网冲击性越来越大，严重影响地区负荷预测的精度，研究含有电铁负荷的地区负荷对电网调度运行具有重要意义，本章在分析地区负荷特性的基础上，对该地区建立预测模型，得到结论如下：

(1) 利用基于灾变遗传算法的时序最小二乘支持向量机法对电铁负荷预测是有效的。时间序列法和因素影响分析出影响电铁负荷变化的主要因素，进而选择最小二乘支持向量机的输入变量；运用灾变遗传算法对模型的边际系数 C 和核宽度系数 γ 进行优化，明显提高了预测精度。

(2) 本章根据第三章的分析结果，对各分量分别进行建立了预测模型，对分量一采用灾变遗传算法的时序最小二乘支持向量机法；根据分量二的波动特点，运用了时间序列预测法；对分量三的预测，是结合其变化规律和影响因素特征，运用最小二乘支持向量机进行预测；分量四和余量由于其波动较有规律，变化规律较为简单，因此运用多元回归模型进行预测。

(3) 最后将各分量的预测结果进行重构，得出最后的负荷预测值。经过对比发现，基于局部建模的预测方法的精确度明显比基于全局建模的预测精确度高，说明基于局部建模的预测方法的有效性。

(4) 对产生误差的原因进行了分析，阐述了影响负荷预测的相关因素。

第五章 结论

影响负荷预测的因素是复杂多样、规律各异的。不同类型负荷的时间变化规律是不同的,影响情况也非相同,由各类负荷构成的系统负荷的变化规律就变得很复杂。因此在进行短期负荷预测时,既要充分分析、掌握并运用其自身规律性,又要兼顾各种因素的影响。目前对负荷分析的研究还不是很深入,对地区级负荷的分析更是少之又少。另外新类型负荷如电铁负荷占网内用电量的比例越来越大,由于其负荷的特殊性,难以把握负荷的变化规律,严重影响地区负荷预测的精度。因此研究这类具有电铁负荷的地区级负荷对电网调度运行具有重要意义。论文在阅读大量文献的基础上,提出了基于局域波的预测新方法,其中采用了局域波、近似熵、灾变遗传算法、最小二乘支持向量机等前沿理论,对某地区的负荷特性及其短期负荷预测作了研究,总的来说,本文取得了如下一些研究成果:

(1) 利用局域波能有效处理非平稳信号的能力,将负荷序列分解为若干个分量,分解后的分量能有效反映负荷的本质特征,通过 Hilbert 变换分析出各分量对总体负荷的波动贡献。

(2) 近似熵具有很强的表征信号不规则性和复杂性的能力,通过计算各分量与实际各类型负荷的近似熵,将其作为它们的特征参数,为各分量的负荷类型识别提供有效的依据。

(3) 应用 Spearman 相关系数来分析由局域波分解所得的各分量及余量与实时气象因素之间的相关程度,精细化挖掘出气象因素对它们的影响情况,并且直观的显示了各分量及余量与实时气象因素的关联情况,将此作为负荷预测的基础。

(4) 研究了电铁负荷的负荷特性并分析了其影响因素,针对电铁负荷特性提出了一种基于灾变遗传算法的时序最小二乘支持向量机的负荷预测方法。利用时间序列法和因素影响分析,选择出最小二乘支持向量机的输入变量;运用灾变遗传算法对模型的边际系数 C 和核宽度系数 γ 进行优化,明显提高了预测精度;利用最小支持向量机进行负荷预测,取得较好的效果。

(5) 通过对负荷分量特性以及影响因素的分析,选取适合的预测模型,分量一运用灾变遗传算法的时序最小二乘支持向量机的负荷预测方法,分量二运用时间序列预测模型,分量三同分量一的方法,分量四和余量采用多元回归预测模型,并对各分量的负荷模型进行重构,得出最后负荷预测结果。

冲击负荷是一种比较复杂,对地区负荷预测影响较大的负荷类型,对冲击负荷的研究还需要进一步探讨,需要全面深入地分析这类负荷的特性、波动规律,以及研究这类负荷的预测方法,从而提高整体地区负荷预测的精度。

参考文献

- [1] R. J. Bhansali. Autoregressive model selection for multistep prediction. *Journal of Statistical Planning and Inference*, 1999, 78(1-2): 295-305
- [2] Z. Deng, Y. Gao. New approach to information fusion steady-state Kalman filtering. *Automatica*, 2005, 41(10): 1695-1707
- [3] 马建伟, 张国立. 人工鱼群神经网络在电力系统短期负荷预测中的应用. *电网技术*, 2005, 29(11): 37-39
- [4] 谢开贵, 李春燕, 周家启. 基于神经网络的负荷组合预测模型研究. *中国电机工程学报*, 2002, 22(7): 86-89
- [5] 尤勇, 盛万兴, 王孙安. 一种新型短期负荷预测模型的研究及应用. *中国电机工程学报*, 2002, 22(9): 16-18
- [6] 张智晟, 孙雅明, 王兆峰, 等. 优化相空间邻近点与递归神经网络融合的短期负荷预测. *中国电机工程学报*, 2003, 23(8): 45-49
- [7] 谢宏, 程浩忠, 张国立, 等. 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): 2-4
- [8] 吴宏晓, 侯志俭, 刘涌, 等. 基于免疫聚类径向基函数网络模型的短期负荷预测. *中国电机工程学报*, 2005, 25(16): 54-56
- [9] 尤勇, 盛万兴, 王孙安. 基于人工免疫网络的短期负荷预测模型. *中国电机工程学报*, 2003, 23(3): 27-29, 98
- [10] 李元诚, 方廷健, 于尔铿. 短期负荷预测的支持向量机方法研究. *中国电机工程学报*, 2003, 23(6): 56-59
- [11] 王锡淮, 朱思锋. 基于支持向量机的船舶电力负荷预测. *中国电机工程学报* 2004, 24(10): 37-39
- [12] 赵登福, 庞文晨, 张讲社, 等. 基于贝叶斯理论和在线学习支持向量机的短期负荷预测. *中国电机工程学报*, 2005, 25(13): 9-13
- [13] 牛东晓, 刘达, 陈广娟, 等. 基于遗传优化的支持向量机小时负荷滚动预测. *电工技术学报* 2007, 22(6): 149-153
- [14] 谢宏, 魏江平, 刘鹤立. 短期负荷预测中支持向量机模型的参数选取和优化方法. *中国电机工程学报*, 2006, 26(22): 18-22
- [15] Zongwu Cai, Ram C. Tiwari. Application of a local linear autoregressive model to BOD time series. *Environmentrics*, 2000, 11(3): 341-450
- [16] Yuehui Chen, Bo Yang, Jiwen Dong. Time-series prediction using a local

- linear wavelet net work. Neurocomputing, 2006, 69: 449-465
- [17]张凯, 姚建刚, 李伟, 等. 基于功率谱分解和实时气象因素的短期负荷预测. 电网技术, 2007, 31 (23): 48-51
- [18]谢宏, 陈志业, 牛东晓, 等. 基于小波分解与气象因素影响的电力系统日负荷预测模型研究. 中国电机工程学报, 2001, 21 (5): 6-10
- [19]苏娟, 杜松怀, 李才华. 基于多因素小波分析的神经网络短期现货电价预测方法. 电力自动化设备, 2007, 27 (11): 27-33
- [20]畅广辉, 刘涤尘, 熊浩. 基于多分辨率 SVM 回归估计的短期负荷预测. 电力系统自动化, 2007, 31 (9): 37-41
- [21]祝志慧, 孙云莲, 季宇. 基于 EMD 和 SVM 的短期负荷预测. 高电压技术, 2007, 33 (5): 118-122
- [22]黎静华, 栗然, 顾雪平, 等. 网格化的电力系统短期负荷预测的 MDRBR 模型. 电力系统自动化, 2005, 29 (24): 27-31
- [23]栗然, 刘宇, 黎静华, 等. 基于改进决策树算法的日特征负荷预测研究. 中国电机工程学报, 2005, 25 (23): 37-41
- [24]L. Cao. Support vector machines experts for time series forecasting. Neurocomputing, 2003, 51: 321-339
- [25]Huang N E, Shen Z, Long S R, et al. The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc. Roy. London A, 1998, 454: 903-995
- [26]马孝江, 余伯, 张志新. 一种新的时频分析方法——局域波法. 振动工程学报, 2000, 13 (9): 24-29
- [27]Melville W K. Wave modulation and breakdown. Journal of Fluid Mech. 1983, 128: 489-506
- [28]盖强. 局域波时频分析的理论研究与应用: [博士学位论文]. 大连: 大连理工大学, 2001
- [29]Huang N E, Shen Z, Long S R, et al. The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc. Roy. Soc. London A, 1998, 454: 903-995
- [30]杨世锡, 胡劲松, 吴昭同, 等. 基于高次样条插值的经验模式分解方法研究. 浙江大学学报, 2004, 38 (3): 267-270
- [31]Huang N E, Shen Z, Long S R. A new view of nonlinear water waves: the Hilbert spectrum. A. Rev. Fluid Mech. 1999, 31: 417-457
- [32]盖强, 马孝江, 张海勇, 等. 一种消除局域波法中边界效应的新方法. 大连理工

- 大学学报, 2002, 42 (1): 115-117
- [33]陈忠, 郑时雄. EMD 信号分析方法边缘效应的分析. 数据采集与处理, 2003, 18 (1): 114-118
- [34]邓永军, 王伟, 钱成春, 等. EMD 方法及 Hilbert 变换中边界问题的处理. 科学通报, 2001, 46 (3): 257-263
- [35]王奉涛. 非平稳信号故障特征提取与智能诊断方法的研究及应用: [博士学位论文]. 大连: 大连理工大学, 2003
- [36]黄大吉, 赵讲平, 苏纪兰. 希尔伯特-黄变换的断点拓延. 海洋学报, 2003, 25 (1): 1-11
- [37]Huang N E, Shen Z, Long S R. A new view of nonlinear water waves: the Hilbert spectrum. *A. Rev. Fluid Mech.* 1999, 31: 417-457
- [38]Rilling G, Flandrin P, Goncalves P. On Empirical Mode Decomposition and it's algorithms. *IEEE Sig. Pro. Lett.*, 2003: 32-329
- [39]程军圣, 于德介, 杨宇. 经典模态分解方法中内禀模态函数判据问题研究. *中国机械工程*, 2004, 15 (20) 1861-1864
- [40]Huang W, Shen Z, Huang N E, et al. Use of intrinsic modes in biology: examples of indicial response of pulmonary blood pressure to step hypoxia. *Proc. Natl Acad. Sci. USA*, 1998, 95: 12766-12771
- [41]Huang W, Shen Z, Huang N E, et al. Nonlinear indicial response of complex nonstationary oscillations as pulmonary hypertension responding to step hypoxia. *Proc. Natl Acad. Sci. USA*, 1999, 96: 1834-1839
- [42]Huang N E, Wu M L, Qu Wendong, et al. Zhang. Applications of Hilbert-Huang transform to non-stationary financial time series analysis. *Applied Stochastic Models in Business and Industry*, 2003, 19: 245-268
- [43]Wu Ming Chya, Hu Chin Kun. Study the cardiorespiratory synchronization by using the empirical mode decomposition method http://www.phys.sinica.edu.tw/~statphys/publications/not_yet_published/
- [44]Liang H, Lin Z, McCallum R W. Artifact reduction in electrogastrogram based on the empirical mode decomposition method. *Medical and Biological Computation*, 2000, 38(1): 35-41
- [45]Balocchi R, Menicucci D, Santarcangelo E, et al. the respiratory sinus arrhythmia from the heartbeat time series using empirical mode decomposition, *Chaos, Solitons and Fractals*, 2004, 20(1): 171-177
- [46]Magrin-Chagnollean I, Baraniuk R G. Empirical mode decomposition based

- time-frequency attributes Proceedings of the 69th SEG Meeting, Houston, Texas, 1999
- [47] Robert F, Anastasi, Eric I. Application of Hilbert-Huang Transform for improved detection in terahertz NDE of shuttle tiles. SPIE International Symposia on Smart Structures & Materials, San Diego, California, SPIE:2005: 6-10
- [48] Zhang HY, Ma X J. A New Method for the Analysis of Non-stationary Signal: The Time-Varying AR Model Based on Intrinsic Mode Functions. ISTM, ShangHai, 2001. 6: 201-205
- [49] 张海勇, 马孝江, 盖强. 非平稳信号的一种 ARMA 模型分析方法. 电子与信息学报, 2002, 24 (7): 993-996
- [50] Stephen C P, Robert J G, Jonathan W E. Application of the Hilbert-Huang Transform to the Analysis of Molecular Dynamics Simulations. J. Phys. Chem. A 203, 107: 4869-4876
- [51] Gabriel Rilling, Patrick Flandrin, Paulo Goncalves. Empirical mode decomposition, fractional Gaussian noise and hurst exponent estimation. IEEE-ICASSP, Philadelphia, USA, 2005, March: 19-23
- [52] Wu Zhaohua, Huang N E. A Study of the Characteristics of White Noise Using the Empirical Mode Decomposition Method, 2003, 1: 27-31
- [53] Patrick Flandrin, Paulo Goncalves, Gabriel Rilling. Detrending and denoising with empirical mode decompositions. Pro. EUSIPCO, Wien, Austria, 2004
- [54] 杨世锡, 胡劲松, 吴昭同. 旋转机械振动信号基于 EMD 的希尔伯特变换和小波变换时频分析比较. 中国电机工程学报, 2003. 23 (6): 102-107
- [55] 钟佑明, 秦树人, 汤宝平. 一种振动信号新变换法的研究. 振动工程学报, 2002. 15 (2): 231-238
- [56] 马孝江, 王凤利, 蔡悦, 等. 局域波时频分布在转子系统早期故障诊断中的应用研究. 中国电机工程学报, 2004, 24 (3): 161-164, 168
- [57] Ga Q, Ma X J. The partial wave method for the analysis of non-stationary signals and its use in machine fault diagnosis. Proceedings of the International Symposium on Test and Measurement, IEEE, 2001, 6: 1465-1468
- [58] 胡劲松, 杨世锡. 基于 HHT 的转子横向裂纹故障诊断. 动力工程, 2004, 24 (2): 218-221
- [59] 王珍, 马孝江. 局域波时频法在柴油机缸套活塞磨损诊断中的应用研究. 内燃机

- 学报, 2002, 20 (2): 157-160
- [60] 邹岩崑, 马孝江. 局域波法的时频分析及应用. 机床与液压, 2003, 4: 190-192.
- [61] 张海勇, 马孝江, 盖强. 一种新的时变参数 AR 模型分析方法. 大连理工大学学报, 2002, 42 (2): 238-239
- [62] 王珍, 马孝江. 基于局域波相空间往复机械故障诊断方法的研究. 机床与液压, 2003, 2: 237-239
- [63] 杨宇, 于德介, 程军圣, 等. 基于经验模式分解的滚动轴承故障诊断方法. 中国机械工程, 2004, 15 (10): 908-911
- [64] Linderhed A. Adaptive image compression with wavelet packets and empirical mode decomposition: [dissertation]. Sweden: Linkoping University, 2004
- [65] Yang Zhihua, Qi Dongxu, Yang Lihua. Signal Period Analysis Based on Hilbert-Huang Transform and its Application to Textre Analysis. In: Proc. Of the 3rd Int' l Conf. on Image and Graphics. Hong Kong: IEEE Computer Society Press, 2004: 430-433
- [66] 岳焕印, 郭华东, 韩春明, 等. 经验模态分解技术在 SAR 干涉图滤波中的应用. 高技术通讯, 2001, 12: 37-40
- [67] Nunes J C, Bouaoune Y, Delechelle E et al. Image analysis by bidimensional empirical mode decomposition. Image and Vision Computing 2003, 21: 1019-1026
- [68] 韩春明, 郭华东, 王长林. 利用经验模态分解方法抑制 SAR 斑点噪声. 遥感学报, 2002, 6 (4): 266-271
- [69] Sinclair S, Pegram G S. Empirical Mode Decomposition in 2-D space and time: a tool for space-time rainfall analysis and nowcasting. Hydrol Earth Syst. Sci. Discuss. 2005, 2: 289-318
- [70] 郝志华. 基于局域波法和盲源分离的故障诊断方法应用研究:[博士学位论文]. 大连: 大连理工大学, 2005
- [71] Linderhed A. 2-d empirical mode decompositions-in the spirit of image compression. Wavelet and Independent Component Analysis Applications IXI, SPIE Proceeding, 2002: 1-8
- [72] Hornero R., Aboy M, Abasolo D., et al. Interpretation of approximate entropy: analysis of intracranial pressure approximate entropy during acute intracranial hypertension. Biornedical Engineering, IEEE Transactions on, 2005, 52(10): 1671-1680

- [73] Jin Tian, Jun-Jie Gu, Xue-Zhi Peng, et al. A fault diagnosis method based on wavelet approximate entropy for fan. *Machine Learning and Cybernetics, 2007 International Conference on*, 2007, 1: 519-523.
- [74] 符玲, 何正友, 麦瑞坤, 等. 近似熵算法在电力系统故障信号分析中的应用. *中国电机工程学报*, 2008, 28(28): 68-73
- [75] Steven M, Burton H. Randomness and degree of irregularity. *Proc. Nad. Acad. Sci. USA*, 1996, 93: 2083-2085
- [76] Pincus S M. Approximate entropy as a measure of system complexity. *Pro. Natl. Acad. Sci*, 1991, 88: 2297-2088
- [77] 胥永刚, 何正嘉. 分形维数和近似熵用于度量信号复杂性的比较研究. *振动与冲击*, 2003, 22 (3): 25-27
- [78] 吴琼, 杨以涵, 刘文颖. 基于最小二乘支持向量机的电力系统暂态稳定在线预测. *中国电机工程学报*, 2007, 9(27): 39-43
- [79] 焦尚彬, 刘丁, 郑岗等. 基于最小二乘支持向量机的绝缘子等值附烟密度预测. *中国电机工程学报*, 2006, 1(26): 150-153
- [80] 杨延西, 刘丁. 基于小波变换和最小二乘支持向量机的短期电力负荷预测. *电网技术*, 2005, 7(29): 61-64
- [81] Shyh-Jier Huang, Kuang-Rong Shih. Short-term load forecasting via ARMA model identification including non-Gaussian process considerations. *IEEE Transactions on Power systems*, 2003, 18(2): 205-215
- [82] 张经纬, 王雪帆, 熊飞等. 基于实验和遗传算法的无刷双馈电机参数估算. *中国电机工程学报*, 2008, 12(28): 103-107
- [83] 李良敏, 温广瑞, 王生昌. 基于遗传算法的改进径向基支持向量机及其应用. *系统仿真学报*, 2008, 11(20): 6066-6096
- [84] 张勇军, 任震, 钟红梅等. 基于灾变遗传算法的无功规划优化. *电力系统自动化*, 2002, 26 (23): 29-32
- [85] 金希东, 李治. 遗传-灾变算法及其在非线性控制系统中的应用. *系统仿真学报*, 1997, 9(2): 111-115
- [86] 余文, 李人厚. 一种基于灾变的多群体遗传算法. *计算机工程*, 2001, 27(7): 72-75

致 谢

本文是在我的导师栗然教授的精心指导和悉心关怀下完成的，两年多来导师对我的学业和研究工作倾注了无限的心血，每一次详细的讨论和耐心的指导都使我受益匪浅。栗老师严谨的治学态度、广博的学识、博大的胸怀、开阔的思路、敏锐的思维，给我留下了深刻的记忆，并将成为我以后工作、学习的力量源泉！栗老师是我学业的导师，也是我生活的导师。从导师身上我学到做人的道理、生活的态度。栗老师所给予我做人、做事、做学问的教导将使我终生受益。在论文完成之际，向栗老师表示深深的敬意和衷心的感谢！

在论文的研究和撰写过程中，师兄张烈勇给了我很大的帮助和鼓励，在此向张烈勇和教研室的各位同学表示感谢！

两年多来，我得到了许多人无私的关注与帮助。在论文完成之际，向你们致以诚挚的谢意！

最后衷心地感谢在百忙之中评阅论文和参加答辩的各位专家和老师！

攻读硕士学位期间发表的学术论文和参加科研情况

发表论文情况

- [1] 栗然, 陆凤怡, 徐宏锐, 张烈勇. 基于局域波与近似熵的负荷分析. 中国电机工程学报, 已录用
- [2] 栗然, 徐宏锐, 陆凤怡. 分层考虑气象因素的电力系统短期负荷预测. 华北电力大学学报, 已录用
- [3] 陆凤怡. 基于小波分解和 ID3 算法的短期日负荷最大值预测方法. 河北省电机工程学会 电工理论与新技术专业学术年会, 2008
- [4] 陆凤怡. 基于实时气象因素的小波最小支持向量机的短期负荷预测. 华北电力大学英采文化节优秀论文, 2009
- [5] 栗然, 陆凤怡, 徐宏锐. 一种基于局域波与近似熵的负荷分析方法. 第 25 届电力系统及其自动化年会, 2009
- [6] 栗然, 徐宏锐, 陆凤怡. 分层考虑气象因素的 PSO-SVM 短期负荷预测方法. 第 25 届电力系统及其自动化年会, 2009

参加科研情况

- [R-1] 参与华北电网有限公司“华北电网冀北五市气象因素与电力负荷相关性分析系统”, 2008, 项目参与人