

摘要

电力负荷预测,特别是短期负荷预测是电力部门一项重要的工作,经济意义重大。因此,追求精益求精的预测准确度成为电力行业的重大科研课题。近十几年来,研究者们提出了大量有效的预测算法,预测精度不断提高。然而,人们对预测精度的要求也越来越高,这就需要科研工作者探索更好的预测方法。

本文将模糊聚类和模式识别理论应用于负荷历史数据和气象数据的预处理,为负荷预测提供了更合理的原始预测数据集;在混沌相空间重构局域法的基础上提出了几种新的预测方法,并将气象因子引入到了混沌预测中。本文在研究过程中得到了以下一些初步的结果。

在数据预处理方面:通过对我国北方某地多年日峰值负荷与对应的9种气象数据进行相关性分析,确定了在夏季,“最大气温、最小气温、平均气温、降雨量、平均露点、平均海面气压”是影响负荷的主要气象因子;在冬季,“降雨量、平均海面气压、平均能见度、平均风速、最大持续风速”是影响负荷的主要气象因子。以2000年夏季的具体数据为例,对日峰值负荷数据和对应的主要气象因子数据进行了模糊聚类。在模糊隶属度的选择上,首先选取了相关系数和欧氏距离法,接着本文构造了“相关系数+欧氏距离”的隶属度函数,通过对三种方法聚类结果与原始数据的比较分析,认为在构造的隶属度函数下的聚类结果更具有合理性。

在负荷预测方面:在混沌重构相空间局域法中,首先采用多元二阶多项式回归法预测出未来时刻的负荷值。该方法与线性回归法相比,具有更高的预测精度,并且显著地抑制了12h周期性的误差增大现象。另外,在寻找嵌入相空间中与参考矢量邻近的矢量时,以指定邻近矢量的个数来代替与参考矢量的距离,不仅保证了在任何情况下都能找到足够的“邻近矢量”来进行负荷预测,还可以提高预测精度。其次,采用直接多步线性回归法进行预测,即从当前值直接回归预测出多个点的负荷,消除了原来一步法中的误差积累。在考虑天气因子的直接多步回归时,本文在“参考矢量”中增加各负荷记录对应的天气记录,以及待预测日的天气预报。这样“负荷记录、天气记录、天气预报”形成一个新型的“参考矢量”,从而实现了混沌时间序列中气象因子的使用。

关键词:模糊聚类分析,模式识别,气象因子,混沌相空间重构,负荷预测

ABSTRACT

Electric power load forecasting, especially short-term forecasting is an important routine of the electric power utility and of great economic significance. Therefore, to pursue higher forecasting precision becomes great research task for the electric power utility. In the past decade, researchers have advanced many forecasting methods which improve the forecasting precision, but it is still hard to meet the increasing demands of both utility and customers. More excellent load forecasting methods should be developed.

In this dissertation, fuzzy clustering and pattern recognition are used to the preprocessing of historical load and weather data to provide more effective data sets for forecasting. Then new load forecasting methods are developed based on the chaotic phase space reconstruction, in which climate factors are considered. The primary work in the dissertation is as follows:

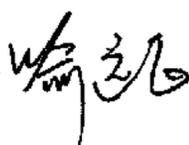
In the phase of load and weather data preprocessing: Firstly, by analyzing the correlation between the daily peak load data and the corresponding 9 kinds of weather data of certain northern area in china, it is confirmed that in summer, "maximum temperature, minimum temperature, mean temperature, precipitation, mean dew point and mean sea level air pressure" are the factors influencing the load most, while in winter, "precipitation, mean sea level air pressure, mean visibility, mean wind speed and maximum constant wind speed" are the main factors. Secondly, the specific data of year 2000 are used as demonstration to the fuzzy clustering for the load and weather data. On the choosing of fuzzy membership degree function, correlative coefficient function and Euclidean distance function are selected first. Then a new membership degree function "correlative coefficient + Euclidean distance" is constructed. Compared with the previous two on the basis of original data, it is approved that the result of the constructed method is more logical.

In the phase of load forecasting: multiple second order polynomial regression combining with the local approach of chaotic phase space reconstruction is used to forecast the future loads. Compared with the linear regression, it achieves better forecasting results and sharply restrains the 12h periodical error augmentation phenomenon. Besides, the "amount" of the neighboring vectors instead of Euclidean distance ϵ between them is used to gain the vectors near the reference vector in the embedded phase space, and this guarantees that in any case adequate neighboring vectors can be obtained as well as enhance the forecasting precision. Another improvement is directly forecasting multi-steps other than the existing one step to eliminate the accumulative errors. Further, weather records of corresponding load records and weather forecast of the day to be forecasted are added to the neighboring vectors which form the new type vectors "load records, weather records and weather forecasts" when consider the weather factors.

Key words: Fuzzy clustering analysis, Pattern recognition, Climate factor, Chaotic phase space reconstruction, Load forecasting.

独创性声明

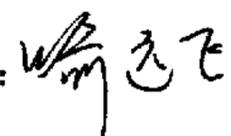
本人声明所呈交的学位论文是本人在导师指导下进行的研究工作和取得的研究成果，除了文中特别加以标注和致谢之处外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得 天津大学 或其他教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名： 签字日期：2004 年 1 月 1 日

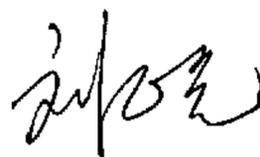
学位论文版权使用授权书

本学位论文作者完全了解 天津大学 有关保留、使用学位论文的规定。特授权 天津大学 可以将学位论文的全部或部分内容编入有关数据库进行检索，并采用影印、缩印或扫描等复制手段保存、汇编以供查阅和借阅。同意学校向国家有关部门或机构送交论文的复印件和磁盘。

(保密的学位论文在解密后适用本授权说明)

学位论文作者签名：

签字日期：2004 年 1 月 1 日

导师签名：

签字日期：2004 年 1 月 2 日

第一章 绪论

1.1 模糊聚类与混沌综述

1.1.1 模糊聚类分析的历史与研究进展

1965年, L. A. Zadeh 创立了模糊集合论。不久后, E. H. Ruspini 于 1969 年引入了模糊划分的概念进行模糊聚类分析。I. Gitman 和 M. D. Levine 提出了单峰模糊集方法用于处理大数据量和复杂分布的聚类。1974 年 J. C. Bezdek 和 J. C. Dunn 提出了模糊 ISODATA 聚类方法。随着模糊数学传入我国, 模糊聚类分析也传入我国。我国学者也给出了许多模糊聚类方法如文^[1-2]等, 同时我国学者对模糊聚类分析的数学基础也进行了研究, 得到了很多令人瞩目的成果, 如文^[3-4]等。另外有更多学者将模糊聚类方法广泛应用于各专业领域, 取得了满意的效果和可观的效益。1982 年方开泰等对常用数学聚类分析方法进行了系统总结。1994 年李相镐、李洪兴等对模糊聚类分析及其应用的有关文献做了系统总结。迄今为止, 研究者们已经提出了各种各样的模糊聚类算法, 比较常见的有: 基于模糊等价关系的传递闭包法, 模糊 ISODATA 聚类等。

利用传递闭包进行 Fuzzy 聚类, 就是由标定得到的 Fuzzy 相似矩阵 \tilde{R} , 求出包含矩阵 \tilde{R} 的最小模糊传递矩阵即 \tilde{R} 的传递闭包 $\tilde{R}^* = t(\tilde{R})$ 之后再依据 $t(\tilde{R})$ 进行聚类。这种方法是最基本的而且是常用的方法。该法得到的传递闭包 $t(\tilde{R})$ 是一个反映等价关系的模糊等价矩阵, 求传递闭包常用的方法是平方法。罗承忠教授在文^[5]中, 李相镐在文^[6]中都介绍了求传递闭包的简便方法, 这些方法都较好的提高了求解速度。模糊 ISODATA 方法是基于模糊划分的思想, 利用迭代方法, 在泛函极值意义下, 不断修正聚类中心的局部优化算法。该算法是 1974 年由 J. C. Bezdek 和 J. C. Dunn 提出。这种方法可以很快求出聚类中心, 但朱剑英在文^[7]中指出该方法必须事先确定类别个数, 且要求分类前能大体知道各样本隶属于各类的程度, 并提出了初始值如何选定和收敛判定问题。此外, 还有诸如模糊 C 均值聚类等算法。这些学者的研究不断地将模糊聚类分析理论推向前进。尽管如此, 要想获得一个好的模糊聚类算法, 仍有许多问题有待研究解决, 如聚类的形状和容量, 算法的初始化, 数据模式的分布, 聚类的数目等^[38]。

模糊聚类理论的发展推动了其在生产实践中的应用, 反过来实际应用的需求又促进了模糊聚类理论不断丰富和完善。随着理论的发展, 模糊聚类已经在众多的领域获得广泛的应用, 并取得了令人满意的效果和可观的效益。其应用范围涉及到通讯系统中的信道均衡、矢量量化编码中的码书设计、时间序列的预测、

神经网络的训练、参数估计、医学诊断、天气预报、食品分类、水质分析、模式识别、数据挖掘、模糊建模等领域^[8~11, 38]。模糊聚类分析在模式识别和图像处理两个领域中得到了相当成功的应用。在一些模式识别的具体应用中, 比如汉字字符识别中的字符预分类^[12]; 语音识别中的分类和匹配^[13]; 雷达目标识别中目标库的建立和新到目标的归类^[14]等等, 模糊聚类取得了较好的效果。图像处理是计算机视觉的重要组成部分, 由于人眼视觉的主观性使图像比较适合用模糊手段处理, 同时训练样本图像的匮乏又需要无监督分析, 而模糊聚类正好满足这两方面的要求, 因此成为图像处理中一个强大的研究分析工具。模糊聚类在图像处理中最为广泛的应用为图像分割, 由于图像分割问题可以等效为像素的无监督分类, 因此早在 1979 年 Coleman 和 Andrews^[15]就提出用聚类算法进行图像分割, 此后基于二维直方图^[16]、塔型结构^[17]、小波分析^[18]、分形分维^[19]等一系列新技术, 人们又相继提出了多种基于模糊聚类的灰度图像分割新方法, 并且在纹理图像分割^[20]、彩色图像分割、序列图像分割、遥感图像分割^[20]等方面也获得了很大的进展。基于模糊聚类的方法在边缘检测^[21, 22, 23]、图像增强^[24]、图像压缩^[25]、曲线拟合等众多方面的研究同样也取得了丰硕的成果。

特别值得一提的是, 近些年模糊聚类理论在电力负荷预测中也得到了大量的应用。我们知道, 电力负荷水平的影响因素很多, 传统的时间序列法只从负荷历史数据中发掘规律用来预测。这显然是有其局限性的, 它对满足电力负荷预测越来越高的精度要求显得力不从心。于是, 各种考虑影响因素的预测方法被提出, 模糊聚类方法便是其中的佼佼者。它将电力负荷与影响负荷变化的各种气候、社会、经济等环境因素联系起来考虑, 但同以往的预测方法不同, 它并不试图建立电力负荷与环境因素关系的数学表达式, 而是将负荷与环境因素作为一个整体进行数据加工和处理。首先应用模糊聚类分析的方法对历史数据进行提炼分类, 将负荷与其环境因素的历史样本分成若干典型类别。然后用合适的模糊集和模糊数描述出各类别环境因素的特征和负荷变化的模式。最后, 当给定未来环境因素状态时, 判定出未来负荷变化属于何种类型, 从而预测出未来负荷值。具体文献参见^[26~29]。

然而, 随着应用的发展, 各学科对模糊聚类理论又提出了许多新的要求, 比如聚类算法的快速实现在图像处理的应用中要求极为迫切; 必须把模糊聚类同新的技术相结合才能取得好的结果; 只有充分挖掘和利用实际应用中的先验知识, 并指导聚类才有望在速度和质量上同步提高; 另外, 现有的模糊聚类都是针对静态数据的, 还需研究数据动态变化或数据不断到达情况的分析方法。上述种种还需在理论上继续开拓和创新。

1.1.2 混沌理论概述^[40,41]

长期以来,人们在认识和描述运动时,总是将运动分为两种类型:确定性运动和随机性运动。在牛顿创立经典力学后的很长时间内,自然科学家都认为一个确定性的系统在确定性激励下,响应也是确定的。Newton 和 Laplace 指出,只要建立了方程,就可以依据初始条件来确定随后的运动。可是近四十年来,“混沌”的发现,冲破了这种传统观念。1963年,气象学家 Lorenz 在分析天气现象模型时得出气象不可长期预测的结论。从此,人们认识到即使确定性系统受确定性激励,响应也可能是不确定的。

下面我们回顾一下混沌学的发展简史。上世纪初的 1903 年,法国数学家 J. H. Poincare 在《科学与方法》一书中提出了 Poincare 猜想,它把动力学系统和拓扑学两大领域结合起来,提出了混沌存在的可能性。从 60 年代开始,人们开始探索科学上的那些莫测之谜,使混沌科学得到了飞速的发展。气象学家 Lorenz 取得了很大的成功,1963 年他发表了“决定性的非周期流”一文,指出气候在不能精确重演与长期天气预报者无能为力之间必然存在一种联系,是非周期与不可预见性之间的联系。这些研究清楚地描述了“对初始条件的敏感性”这一混沌基本性态,这就是著名的“蝴蝶效应”。1975 年,中国学者李天岩和美国数学家 J. Yorke 在《American Mathematical Monthly》杂志上发表了“周期三意味着混沌”的著名文章,揭示了从有序到混沌的演变过程。1976 年,美国生物学家 R. May 在自然杂志上发表“具有极复杂的动力学的简单数学模型”一文,他向人们表明了混沌理论的惊人信息:简单的确定论数学模型也可产生看似随机的行为。1977 年,第一次国际混沌会在意大利召开,标志着一门新的科学——混沌科学的诞生。在 80 年代,混沌科又得到进一步发展。1980 年,数学家 B. Mandelbrot 用计算机绘出了第一张 Mandelbrot 集的图像,从那时起, Mandelbrot 即成了混沌的一种公认的标志。1983 年,加拿大物理学家 L. Glass 在《物理学》杂志上发表著名文章“计算奇异吸引子的奇异程度”,开创了全世界计算时间序列维数的热潮。到了 90 年代,混沌科学与其他科学相互渗透,混沌在现代科学技术中起着十分重要的作用。正如混沌科学的倡导者之一,美国海军官员 M. Shlesinger 所说:“20 世纪科学将永远铭记的只有三件事,那就是相对论,量子力学与混沌。”

自八十年代以来,电子学领域出现了混沌的应用研究热潮。其中,以 L.O. Chua 的蔡氏电路为代表的混沌电路与系统的研究、混沌同步和控制通信中的应用、混沌在扩频通信中的应用、混沌信号处理、分形数据压缩和混沌神经网络成为各国学者研究的热点。进入九十年代后,混沌科学与其他科学相互渗透。无论是在数学、物理学、生命科学、地球科学、信息科学,还是在经济学、天文学等领域,混沌均得到了广泛应用。然而,混沌科学虽然在基础理论方面取得了很

大进展,但还没有取得根本性的突破,许多问题仍然没有解决。所以,混沌的研究及其应用主要仍是用数值方法。在混沌的应用上,根据混沌系统提取的非线性时间序列对系统的未来进行预测,是一个十分重要的方面。

1.2 电力负荷预测的基本原理及其意义^[30, 42]

电力系统负荷预测是指在充分考虑一些重要的系统运行特性、增容决策与自然条件的情况下,利用一套系统地处理过去与未来负荷的方法,在一定精度意义上、决定未来某特定时刻或某些特定时刻的负荷值。电力系统负荷预测,一般有短期负荷预测、中期负荷预测和长期负荷预测3种。短期负荷预测是指一年以内的预测;中期负荷预测一般是指1~5年以内的预测;而长期负荷预测是指5年以上的预测^[30]。对于供电部门来说,准确的负荷预测,特别是准确的短期负荷预测总是至关重要的,因为这决定了系统中大部分设备的运行特性。例如,对于一个电网,若负荷预测实际偏低,则该网实际上就不能履行向邻网供电的合同,甚至还可能缺电;反之,若负荷预测偏高,则会导致发电设备不能得到充分利用,从而引起投资的浪费。电力系统的任务是为各类用户尽可能经济地提供可靠性符合标准的电能,要求其能满足各种负荷需要,所以负荷的大小和特性对电力系统设计和运行都极为重要,因此,电力系统,特别是供电部门总是希望得到科学而精确的负荷预测来指导其制定计划和决策。然而,精确的负荷预测并不是容易做到的,这是因为:①未来各种可能引起负荷发生变化的情况,并不能事先确切地全部掌握;②某些复杂的因素,即使知道它们会对电力系统的负荷发生影响,然而,要定量地准确判定它们的影响,常常是困难的;③没有一种足够完善的理论方法适用于所有的负荷预测场合。

负荷预测具有不准确性、条件性、时间性和多方案性,因此必须科学的总结预测工作的基本原理,以指导预测工作。目前已经总结出来的原理有以下几个:

- (1) 可知性原理。即未来负荷的变化规律是可以为人们所知道的,这是由世界的可知性决定的。
- (2) 可能性原理。这是指事物的发展变化有多种可能,负荷预测存在多方案预测的情况。
- (3) 连续性原理。负荷预测从过去发展到现在,再从现在发展到未来,其中某些特征得以保持和延续,这一过程是连续变化的。
- (4) 相似性原理。即事物现在的发展状况和发展过程可能与过去一定阶段的发展状况和发展过程存在相似性,可以根据已知的发展状况来预测未来的发展状况。

- (5) 反馈性原理。预测的反馈性原理实际上是为了不断提高预测的准确性而进行的反馈环节。
- (6) 系统性原理。即认为预测对象是一个系统，预测对象的未来发展是系统整体的动态发展，各个组成部分和影响因素之间的相互作用和相互影响直接关系到整个系统的发展。

1.3 负荷预测研究的发展现状

电力负荷的组成多种多样且不断变化，其影响因素也复杂多变，因此负荷预测是一个系统工程，用单一理论研究负荷预测是困难的。长期以来，人们对电力系统负荷预测，特别是短期负荷预测进行了大量的研究，提出了许多有效的方法。近些年，随着科学技术的迅速发展，预测理论技术也取得了长足的进展，新的预测方法，尤其是属于人工智能与模式识别领域的新方法不断出现，这些方法为电力负荷预测问题的研究提供了有力工具。如灰色预测模型、优选组合预测、专家系统预测以及神经网络预测等等，这些预测方法都已经开始引入电力系统负荷预测，并取得了较好的效果。

尽管新的预测方法不断涌现，但从所用的理论来讲，负荷预测则可以分为传统预测方法和现代预测方法两个阶段。传统的预测方法的基础是传统数学工具，这类方法包括回归模型法、时间序列法等；现代预测方法是随着人工智能的发展而兴起的，它结合了人工智能领域里的模糊数学、神经网络、专家系统等学科的研究成果。这些方法主要分为四大类：（1）数理统计法；（2）模糊数学法；（3）神经网络法；（4）智能系统法。

数理统计法是一种经典的负荷预测方法，它主要是针对负荷历史记录内在的数学特性，利用统计理论进行未来时刻负荷的预测。如经常见到的回归分析法、指数平滑法、时间序列法、卡尔曼滤波法等。统计方法具有原理简单，计算量小，速度快等优点，但其缺点也是显而易见的，即预测精度不够高，对满足现代电力行业越来越高的负荷预测精度要求显得无能为力。在实际应用中，关于这类方法的研究文献并不少。如文献^[31]中提出了一种不同于传统时间序列模型（AR、MA、ARMA 等）的时间序列的传递函数模型，此模型的优点在于它克服了传统模型不能考虑气温等外界因素对负荷的非线性影响，并且保留了传统模型的计算量小，速度快的优点。文献^[32]和^[33]中分析了电力系统负荷中的混沌，对于重构混沌相空间，利用线性回归模型进行负荷预测。文献^[34]针对常规回归预测法的缺点分别研究了曲线参数的非线性直接估计法和基于“近大远小”原则的线性加权回归预测法，并给出了一般性的非线性加权回归预测的方法和步骤，改善了预测效

果。

模糊数学是美国加州大学教授 L. A. Zadeh 于 1965 年以模糊集的概念首次提出, 至今已经历了近 40 年的发展。作为现代智能技术中最重要的技术之一, 它是处理不确定性问题的有力武器。其中模糊聚类、模糊模式识别尤为热门。现代负荷预测越来越呈现出面对数据分析的特征, 它是信息时代发展影响负荷预测的结果。从预测的发展趋势看, 今后的预测方法会越来越多地建立在拥有大量的信息数据的基础上。这包含两方面的内容: 一方面是信息时代的发展, 预测素质和水平提高, 数据通信和采集手段的现代化, 使得负荷信息数据的量多; 另一方面, 由于信息不全面或其它因素, 使得许多数据的可靠性和正确性值得怀疑, 而数学模型首先要求初始的信息数据必须正确, 这使得必须首先对负荷数据进行必要的分析和处理, 然后再进行使用。于是模糊理论, 特别是模糊聚类分析和模式识别在电力负荷预测领域有了大放光彩的机会, 它们在电力负荷预测中的应用越来越多, 主要用于对数据进行预加工和处理, 使数据集的特征更明显和突出。例如, 文献^[26]依据模糊模式识别、模糊聚类理论, 提出了一种短期负荷预测新方法, 它根据模糊聚类参数与预测因子的前期特征值, 确定相应的类别变量的特征值, 应用建立的相关关系进行负荷预测; 文献^[27]采用一种组合神经网络模型, 将神经网络、模糊聚类分析和模式识别有机结合起来; 文献^[29]应用聚类分析方法描述由于相关因素的不同而导致的待预测日与历史日之间的差异程度, 实现了一种基于相关因素匹配的短期预测方法。

神经网络是近些年人们研究得比较多的热门课题。传统负荷预测的数学模型是用显式的数学表达式加以描述, 这就决定了传统的预测模型的局限性。因为负荷变化的自然规律很难用一个显式的数学公式表示。而神经网络是一个具有高度非线性的超大规模连续时间动力系统, 可以映射任意复杂的非线性关系, 且能够识别有噪声或变形的样本, 通过学习能把样本隐含的特征和规律分布于神经网络的连接权上, 因此, 近年来在电力负荷预测中得到了广泛应用。其缺点是计算复杂度随着样本特征维数的增加而急剧增长。关于神经网络在电力系统负荷预测中的应用, 文献^{[35][36][37][38]}都是很典型的研究。

智能系统法是随着人工智能发展而出现的一种方法, 它是基于 if/then 规则和经验数学表达式的, 一个完善的智能系统通常由知识库, 推理机, 数据库, 知识获取部分, 解释部分共五个部分组成。借助专家系统, 能识别预测日所属类型, 考虑天气因素对负荷的影响, 按一定的规则推理并进行预测。其缺点是知识库的形成过程复杂, 工作量大、把专家的知识 and 经验等精确地转化为一组规则往往难度较大, 从不同的专家得到的知识有可能不同。

1.4 本文的主要工作

如前所述,电力系统负荷预测是电力部门一项重要的日常工作,针对它的各种新技术和手段层出不穷,人们为了提高预测准确性可以说是挖空心思,但效果并不明显。可见,电力系统负荷预测是一项看似简单,实则困难的工作。本文以电力负荷预测为出发点,以提高预测准确性为目的,进行了以下几项研究:

(1) 在负荷预测数据的处理方面,应用模糊聚类和模式识别理论,详细分析了预测前期因素与负荷的关系。在求模糊相似矩阵过程中,提出用相关系数与距离的加权和作为模糊隶属度函数 $r(i,j)$ 。一般认为,相关系数主要反映的是各样本间的线性关系,即样本在形状上的相似性;而欧氏距离主要反映的是各样本在空间上距离的远近,距离较近的样本之间的相似性也较大。但是,对于电力负荷来说,时常出现的情况是,样本特征组成的曲线形状很相似而数值却存在较大的差别,或者是样本间各分量的总欧氏距离很接近但样本曲线形状差异很大。这些情况下,如果在模糊聚类时仅选择相关系数作为隶属度函数,则容易将样本特征组成的曲线形状相似而数值迥异的负荷日划为同一类。相反,如果在聚类时仅选择欧氏距离作为隶属度函数,则会将距离接近但形状迥异的负荷日划为同一类。在理论上聚类后,通过原始数据与聚类结果的对照分析认为,该方法求得的聚类结果具有更好的合理性,是对原始数据的成功分类处理。在确定聚类截水平 λ 方面,根据电力系统负荷所具有的规律和特性,认为属于同一类中的连续天数不超过 5 天,从而通过程序自动得出聚类水平 λ 。

(2) 在利用重构混沌相空间进行负荷预测时,将二阶多项式回归应用于负荷预测,该方法与线性回归法相比,具有更高的预测精度,并且显著降低了现有方法中存在的 12h 周期性误差增大现象。另外,在寻找嵌入相空间中与参考矢量邻近的矢量时,以指定邻近矢量的个数来代替与参考矢量的距离,不仅保证了在任何情况下都能找到足够的“邻近矢量”来进行负荷预测,还可以提高预测精度。

(3) 在利用重构混沌相空间进行负荷预测时,采用直接多步线性回归预测,即从当前点的负荷直接回归预测出未来多个点的负荷,这样就消除了原来一步法中的误差积累。从而提高了预测精度。另外,在考虑天气因子的回归时,本文在“参考矢量”中增加各负荷记录对应的天气记录,以及待预测日的天气预报。这样“负荷记录、天气记录、天气预报”形成一个新型的“参考矢量”,这种处理与考虑天气的相似日预测有某种类比性,从而实现了混沌时间序列中气象因子的使用。

第二章 理论基础

为了便于阅读，本章将介绍模糊聚类分析和混沌时间序列的有关理论知识。

2.1 模糊等价矩阵聚类分析法

聚类分析是根据事物间不同特征、亲疏程度和相似性等关系，对它们进行分类的一种数学方法。在现实世界中，一组事物根据其亲疏程度和相似性是否形成一个类群，或一个事物是否属于某一个类别，其界限往往不分明，具有很大程度的模糊性，模糊集合论正是刻画和解决这类聚类问题的数学方法。模糊聚类分析是依据客观事物间的特征、亲疏程度和相似性，通过建立模糊相似关系对客观事物进行分类的数学方法。用模糊聚类分析方法处理带有模糊性的聚类问题要更为客观、灵活、直观和计算更简捷。

在模糊聚类分析中，我们称所要进行分类的对象为样本。要对样本进行合理的分类，首先应考虑样本的各种特性指标（参测数据）。即被分类对象的集合为：

$$X = \{x_1, x_2, \dots, x_n\}$$

每一个样本 x_i 有 m 个特性指标，即样本 x_i 可表示为特性指标向量

$$x_i = (x_{i1}, x_{i2}, \dots, x_{im})$$

其中 x_{ij} 表示第 i 个样本的第 j 个特性指标。则 n 个样本的特性指标矩阵为：

$$\begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix}$$

2.1.1 数据规格化

一般来讲，由于 m 个特性指标的量纲和数量级都不相同，在运算过程中可能导致突出某些数量级特别大的特性指标对分类的作用，而降低甚至排除了某些数量级很小的特性指标的作用，致使对各特性指标的分类缺乏一个统一尺度。为了消除特性指标单位的差别和特性指标数量级不同的影响，必须对各指标值施行数据规格化，从而使得每一指标值统一于某种共同的数值特性范围。

数据规格化常用的方法有

1. 数据标准化

$$x'_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j} \quad (2-1)$$

$$\text{式中 } \bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}, \quad \sigma_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}$$

显然，施行变换后的每个指标值的平均值为零，方差为 1。

2. 极大值规格化

$$x'_{ij} = \frac{x_{ij}}{x_{j\max}} \quad (2-2)$$

式中 $x_{j\max} = \max(x_{1j}, x_{2j}, \dots, x_{nj})$

3. 均值规格化

$$x'_{ij} = \frac{x_{ij}}{\sigma_j} \quad (2-3)$$

σ_j 的意义同上。

4. 中心规格化

$$x'_{ij} = x_{ij} - \bar{x}_j \quad (2-4)$$

\bar{x}_j 的意义同上。

5. 对数规格化

$$x'_{ij} = \log x_{ij} \quad (2-5)$$

2.1.2 模糊相似关系矩阵

定义 2.1.2.1 称论域 X 上的模糊关系 \tilde{R} 是模糊相似关系，如果 \tilde{R} 满足：

- (1) 自反性，即 $\tilde{R}(x, x)=1$ ；
- (2) 对称性，即 $\tilde{R}(x, y) = \tilde{R}(y, x)$ 。

如果论域 X 有限，记 $X=\{x_1, x_2, \dots, x_n\}$ ，则模糊相似关系 \tilde{R} 可以用模糊矩阵 $\tilde{R} = (r_{ij})_{n \times n}$ 表示，易见，矩阵 \tilde{R} 是一个对角线元素为 1 的对称矩阵，称为模糊相似矩阵。

具体地说，即对于样本空间 $X = [x_{ij}]_{n \times m}$ ，设 x_{ij} 均已规格化，用多元分析的方法来建立样本与样本之间的相似关系（亲疏关系），即计算出衡量被分类对象

$$x_i = (x_{i1}, x_{i2}, \dots, x_{im})$$

$$x_j = (x_{j1}, x_{j2}, \dots, x_{jm})$$

之间的相似程度 r_{ij} ，使得

$$0 \leq r_{ij} \leq 1, \quad i, j = 1, 2, \dots, n$$

当 $r_{ij} = 0$ 时, 则表示样本 x_i 与样本 x_j 毫不相似; $r_{ij} = 1$ 表示 x_i 与 x_j 完全相似或者等同; 当 $i=j$ 时, r_{ii} 就是样本 x_i 自己与自己的相似程度, 恒取为 1, 即 $r_{ii} = 1, i=1, 2, \dots, n$

于是得到一个样本与样本之间的模糊相似关系矩阵

$$\tilde{R} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix}$$

对于模糊相似矩阵, 有以下两个重要的命题:

命题 2.1.2.1 设 \tilde{R} 是模糊相似矩阵, 则对任意自然数 k , \tilde{R}^k 也是模糊相似矩阵。

命题 2.1.2.2 设 \tilde{R} 是模糊相似矩阵, 则

$$\tilde{R} \leq \tilde{R}^2 \leq \cdots \leq \tilde{R}^k \leq \tilde{R}^{k+1} \leq \cdots$$

人们总结出了很多方法, 可以建立模糊相似关系, 从而获得模糊相似矩阵。样本 x_i 与样本 x_j 的相似程度 r_{ij} 可以根据实际的聚类问题选取下列方法之一:

1. 相似系数法

(1) 数量积法

$$r_{ij} = \begin{cases} 1 & i = j \\ \frac{1}{M} \sum_{k=1}^m x_{ik} x_{jk} & i \neq j \end{cases} \quad (2-6)$$

式中 $M = \max_{i \neq j} (\sum_{k=1}^m x_{ik} x_{jk})$ 。

(2) 夹角余弦法

$$r_{ij} = \frac{\sum_{k=1}^m x_{ik} x_{jk}}{\sqrt{\sum_{k=1}^m x_{ik}^2} \sqrt{\sum_{k=1}^m x_{jk}^2}} \quad (2-7)$$

(3) 相关系数法

$$r_{ij} = \frac{\sum_{k=1}^m (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)}{\sqrt{\sum_{k=1}^m (x_{ik} - \bar{x}_i)^2} \sqrt{\sum_{k=1}^m (x_{jk} - \bar{x}_j)^2}} \quad (2-8)$$

其中, $\bar{x}_i = \frac{1}{m} \sum_{k=1}^m x_{ik}$, $\bar{x}_j = \frac{1}{m} \sum_{k=1}^m x_{jk}$ 。

(4) 指数相似系数法

$$r_{ij} = \frac{1}{m} \sum_{k=1}^m e^{-\frac{(x_{ik} - x_{jk})^2}{s_k}} \quad (2-9)$$

其中 s_k 适当选择。

2. 距离法

可以用样本 x_i 与样本 x_j 间的距离 d_{ij} 来标定他们的相似程度 r_{ij} 。因为 x_i 与 x_j 的距离越大, 则它们的相似程度越小, 所以一般 r_{ij} 可由下式标定:

$$r_{ij} = 1 - c \times (d_{ij})^\alpha, \text{ 其中 } c \text{ 和 } \alpha \text{ 是两适当选择的常数。}$$

应用中, 常见的距离法有

(1) 海明距离法 (绝对值减数法)

$$r_{ij} = 1 - c \sum_{k=1}^m |x_{ik} - x_{jk}| \quad (2-10)$$

其中 c 适当选取, 使 r_{ij} 在 $[0,1]$ 中且分散开。

(2) 欧氏距离法

$$r_{ij} = 1 - c \times \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2} \quad (2-11)$$

(3) 绝对值指数法

$$r_{ij} = e^{-\sum_{k=1}^m |x_{ik} - x_{jk}|} \quad (2-12)$$

(4) 绝对值倒数法

$$r_{ij} = \begin{cases} 1 & i=j \\ \frac{M}{\sum_{k=1}^m |x_{ik} - x_{jk}|} & i \neq j \end{cases} \quad (2-13)$$

其中 M 适当选取, 使 r_{ij} 在 $[0,1]$ 中且分散开。

3. 贴近度法

当样本 x_i 的特性指标向量 $x_i = (x_{i1}, x_{i2}, \dots, x_{im})$ 为模糊向量, 其中 $x_{ik} \in [0,1]$, $i=1, 2, \dots, n$, $k=1, 2, \dots, m$, 即每个样本 x_i 是 X 上的模糊子集时, 样本 x_i 与样本 x_j 的相似程度实质上是模糊子集 x_i 与模糊子集 x_j 的贴近度。常用的贴近度法有

(1) 最大最小法

$$r_{ij} = \frac{\sum_{k=1}^m (x_{ik} \wedge x_{jk})}{\sum_{k=1}^m (x_{ik} \vee x_{jk})} \quad (2-14)$$

(2) 算术平均最小法

$$r_{ij} = \frac{2 \sum_{k=1}^m x_{ik} x_{jk}}{\sum_{k=1}^m (x_{ik} + x_{jk})} \quad (2-15)$$

(3) 几何平均法

$$r_{ij} = \frac{\sum_{k=1}^m (x_{ik} \wedge x_{jk})}{\sum_{k=1}^m \sqrt{x_{ik} \cdot x_{jk}}} \quad (2-16)$$

要求 $x_{ik} \cdot x_{jk} > 0$ 。

方法众多, 具体选择哪种方法, 则要根据实际问题而定。通过上述方法求解模糊矩阵的过程, 称作标定。

显然, 上述所有方法中 $|r_{ij}| \in [0, 1]$, 如果 r_{ij} 出现负值, 可以采用下述方法重新调整:

(1) $r'_{ij} \in [0, 1]$;

(2) $r'_{ij} = \frac{r_{ij} - m}{M - m}$ ($i \neq j$), 其中, $M = \max_{i \neq j} r_{ij}$, $m = \min_{i \neq j} r_{ij}$, 于是 $r'_{ij} \in [0, 1]$

2.1.3 模糊等价关系

我们知道, 集合 X 等价关系 \tilde{R} 可以将 X 分类, 每个类称为一个等价类, $x_0 \in X$ 所在的等价类记做 $[x_0]$,

$$[x_0] := \{x \mid R(x_0, x) = 1\}$$

对于模糊聚类问题, 我们同样需要借助经典等价关系。前面已经给出了 λ 截关系的定义, 下面给出模糊等价关系的定义。

定义 2.1.3.1 称模糊关系 \tilde{R} 是 X 的模糊等价关系, 当且仅当对 $\forall \lambda \in [0, 1]$, λ 截关系 R_λ 都是 X 的等价关系。

定理 2.1.3.2 模糊关系 \tilde{R} 是 X 的模糊等价关系, 当且仅当 \tilde{R} 满足

(1) 自反性, 即 $\tilde{R}(x, x) = 1$ ($\forall x \in X$)

(2) 对称性, 即 $\tilde{R}(x, y) = \tilde{R}(y, x)$;

(3) 传递性, 即 $\tilde{R}^2 \subseteq \tilde{R}$

若 $X = \{x_1, x_2, \dots, x_n\}$ 为有限论域, 则 X 的模糊等价关系可表示为一个 n 阶模糊矩阵 $\tilde{R} = \{r_{ij}\}_{n \times n}$, 且满足

(1) 自反性, 即 $r_{ii} = 1 \quad (\forall 1 \leq i \leq n)$

(2) 对称性, 即 $r_{ij} = r_{ji}$

(3) 传递性, 即 $\tilde{R}^2 \subseteq \tilde{R}$, 注意此时 $\bigvee_{k=1}^n (r_{ik} \wedge r_{kj}) \leq r_{ij}$

\tilde{R} 称为模糊等价矩阵。

由于自反性, $r_{ii} = 1$ 成立, 于是

$$\bigvee_{k=1}^n (r_{ik} \wedge r_{kj}) \leq r_{ii} \wedge r_{ij} = r_{ij}$$

又因为传递性成立, 于是

$$\bigvee_{k=1}^n (r_{ik} \wedge r_{kj}) = r_{ij}$$

即 $\tilde{R}^2 = \tilde{R}$ 成立。

根据模糊等价关系的定义, 可以通过给定适当的聚类水平 λ , 利用 λ 截矩阵 R_λ 来进行分类。

2.1.4 模糊聚类传递闭包法

我们知道, 一个模糊等价关系 (模糊等价矩阵) 可以确定一个模糊分类。通过标定所得的模糊相似关系矩阵 \tilde{R} 是模糊相容矩阵, 但未必是模糊等价矩阵。因此, 要进行基于模糊关系的模糊聚类, 由相应的模糊相似矩阵得到相应的模糊等价矩阵是关键。这通常需要对原矩阵进行一定的变换, 从而改变了原有的相似关系。我们通过“距离”来衡量这种改变的大小。

定义 2.1.4.1 模糊相似矩阵 $\tilde{A} = (a_{ij})_{n \times n}$ 和 $\tilde{B} = (b_{ij})_{n \times n}$ 之间的距离定义为

$$d(\tilde{A}, \tilde{B}) = \sqrt{\sum_{i=1}^{n-1} \sum_{j=i+1}^n (a_{ij} - b_{ij})^2} \quad (2-17)$$

求解模糊等价矩阵最常用的方法是传递闭包法。模糊相似矩阵 \tilde{R} 的传递闭包是包含 \tilde{R} 的与 \tilde{R} 的距离最小的模糊等价矩阵, 通常记做 $t(\tilde{R})$, 也可记做 \tilde{R}^* 。

\tilde{R} 的传递闭包 $t(\tilde{R})$ 满足

(1) $t(\tilde{R}) \supseteq \tilde{R}$

(2) 传递性: $t(\tilde{R})^2 \subseteq t(\tilde{R})$

(3) 最小性: $\forall \tilde{T} \supseteq \tilde{R}$, 满足 $\tilde{T}^2 \subseteq \tilde{T}$, 都有 $\tilde{T} \supseteq t(\tilde{R})$

定理 2.1.4.1 设 \tilde{R} 是 n 阶模糊相似矩阵, 则存在一个最小自然数 $k(k \leq n)$, 使得传递闭包 $t(\tilde{R}) = \tilde{R}^k$, 且对于一切大于 k 的自然数 l , 恒有 $\tilde{R}^l = \tilde{R}^k$ 。

此时, $\tilde{R}^k = t(\tilde{R})$ 为模糊等价矩阵。

该定理指出了求解传递闭包的方法，即矩阵 \tilde{R} 自乘，直到出现自然数 k ，满足 $\tilde{R}^{k+1} = \tilde{R}^k$ ，则 $\tilde{R}^{k+1} = t(\tilde{R})$ 。显然，利用平方法，依次计算 $\tilde{R}^2, \tilde{R}^4, \dots, \tilde{R}^{2m}, \dots$ ，可以减少计算量。

利用传递闭包法进行模糊聚类的步骤总结如下：

Step(1): 数据规格化；

Step(2): 利用前面确立模糊相似关系的方法建立模糊相似矩阵 \tilde{R} ；

Step(3): 使用平方法求 \tilde{R} 的传递闭包 \tilde{R}^* (模糊等价矩阵)；

Step(4): 给定适当的聚类水平 λ ，利用截矩阵 R_λ 进行聚类。

模糊聚类传递闭包法的分类过程框图，如图 2-1 所示。

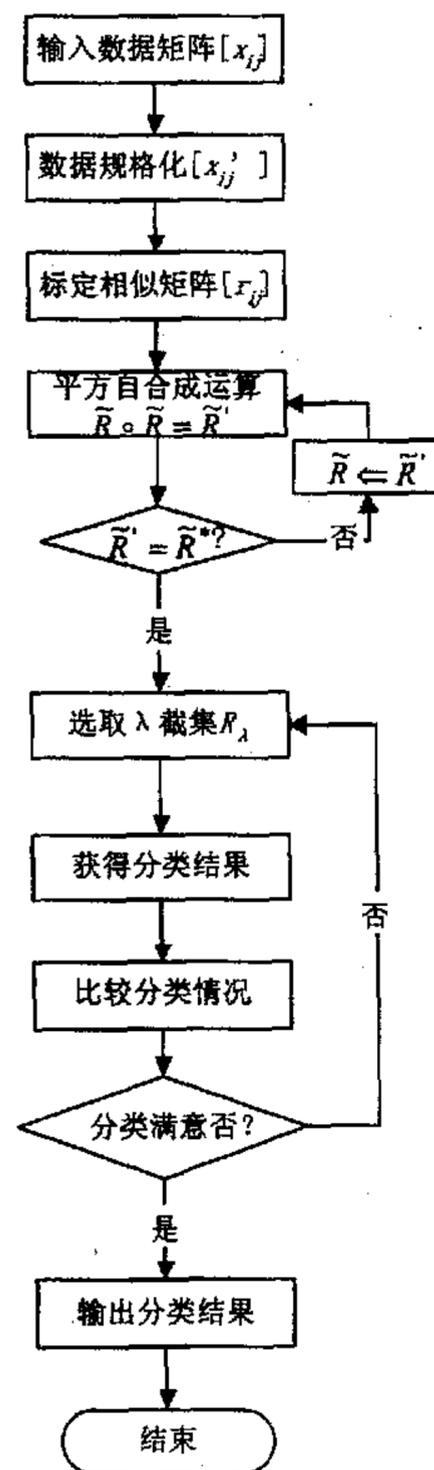


图 2-1 模糊聚类传递闭包法分类框图

2.2 混沌理论

2.2.1 混沌时间序列

从时间序列研究混沌,始于 Packard 等(1980)提出的重构相空间理论。我们知道,对于决定系统长期演化的任一变量的时间演化,均包含了系统所有变量长期演化的信息。因此我们可通过决定系统长期演化的任一单变量时间序列来研究系统的混沌行为。而吸引子的不变量——关联维(系统复杂度的估计)、Kolmogorov 熵(动力系统的混沌水平)、Lyapunov 指数(系统的特征维数)等在表征系统的混沌性质方面一直起着重要作用。

混沌时间序列的预测方法包括:全域法、局域法、加权零阶局域法、加权一阶局域法、基于 Lyapunov 指数的预测方法和基于神经网络的预测方法等。混沌时间序列的应用前景非常广阔,如电力系统短期负荷预测、股市行情预测、天气预报等。而且基于 Lyapunov 指数的时间序列预测是根据数据序列本身所计算出来的客观规律——Lyapunov 指数作预测,不是象传统方法通过主观建模预测,因此预测效果较好。

2.2.2 混沌时间序列的相空间重构

决定序列的可观测因素很多,而且相互作用的动力学方程往往是非线性的,甚至是混沌的。在确定性的基础上,对序列动力学因素的分析,目前广泛采用的是延迟坐标状态空间法。即对于给定的时间序列 $x_1, x_2, \dots, x_{n-1}, x_n, \dots$, 通常将其扩展到三维甚至更高维的空间中去,以便把时间序列中蕴藏的信息充分地显露出来,这就是延迟坐标状态空间重构法。

2.2.1.1 相空间重构方法要点

对于一个吸引子,包括混沌吸引子或其他吸引子,可以通过相空间重构来预测吸引子的时间行为。基本方法是将吸引子的 N 各观测值构成的时间序列 $\{x(i)\}$, $i=1,2,\dots,N$, 嵌入 m 维相空间中。假设当前值为 $x(T)$, 我们预测下一个值 $x(T+1)$ 。一种常用的方法为:

- (1) 首先需要构造 $N_m = N - (m-1)\tau$ 个 m 维相空间中的点(就是 m 维矢量):

$$Y(k) = [x(k), x(k+\tau), x(k+2\tau), \dots, x(k+(m-1)\tau)], \quad k=1, 2, \dots, N_m, \quad \tau > 1$$
 为延迟时间,其目的是使 $Y(k)$ 在吸引子上占据合适的轨道长度。
- (2) 用矢量 $Y(k)$ 中的各个分量的线性组合来表示出标量 $y(i) = A \times Y(k)$ 。但这里的系数 A (矢量) 还是未知的。其中,可以选取 $y(T+1) = A \times Y(N_m)$ 。

- (3) 用吸引子的性质来求出这些未知的系数 A 。具体方法是，找到 $Y(k)$ 中所有与 $Y(N_m)$ 的距离小于 ε 的那些“邻近向量” $Y(j)$ 。这些 $Y(j)$ 是以 $Y(N_m)$ 为中心的、与 $Y(N_m)$ 比较相似的矢量。从吸引子上看，它们围绕在 $Y(N_m)$ 周围。然后再找到各个 $Y(j)$ 对应的 $y(i)=A \times Y(j)$ ，运用最小二乘法估计出这些未知系数 A 。

2.2.1.2 重构相空间参数的选择

在重构相空间中，时间延迟 τ 和嵌入维数 m 的选择具有十分重要的意义，也是很困难得一件事。嵌入相空间上的距离 ε 对预测的效果也有影响。现有的混沌研究结论包括：

- (1) 嵌入定理：如果原来吸引子处在一个 d_A 维空间中，那么嵌入相空间的维数 $m \geq 2d_A + 1$ 。
- (2) 延迟时间 τ 的选择，方法很多。一般以 $x(i)$ 的自相关系数首次过 0 时的延迟时间来确定 τ 。
- (3) 嵌入相空间上的距离 ε 的确定，一般要经过试探。

2.2.3 相空间重构方法应用于负荷预测

目前常见的混沌预测方法有“ λ_1 预测”、“相空间重构”以及“人工神经网络 (ANN)”等。它们也是电力系统中短期负荷预测混沌方法中常用的基本方法。

其中“相空间重构”是比较适合于负荷预测的方法。因为它不仅对可以对混沌吸引子进行预测，还可以用来对周期和准周期等非混沌吸引子进行预测。正是这种适应性，使得它比较适合于行为类型比较复杂的负荷吸引子的预测。

其他的混沌预测方法，例如利用最大 Lyapunov 指数的“ λ_1 预测”，一般不太适合于负荷预测。“ λ_1 预测”存在 2 方面的问题：首先因为它是一种典型的混沌吸引子预测方法，只适应于混沌预测而不太适合其他吸引子的预测；并且只适用于只有一个正 λ_1 的低维混沌的预测。其次， λ_1 是一种“平均”性的参数，而不是一种瞬时参数，这就使得该预测方法存在某些必然的系统误差。此外，ANN 是一种智能方法，它不仅可以用来预测混沌，还可以引入天气和社会等因素的作用，因此 ANN 是一种比较有前途的方法。但 ANN 的训练还存在一定的问题。

从本质（即方法的结构）上看，“相似日”方法可以看成是“相空间重构法的局部线性法”的智能化升级版本。从概念上讲，“相似日”方法属于模式识别中聚类的概念与方法。它把日负荷曲线的形状按照一定的特征进行“相似性”分类，然后通过一定的（智能）处理后，得到负荷的预测值。而非线性动力系统“相空间重构法”，相空间重构后的系统与原非线性动力学系统微分同胚、与

原吸引子的拓扑结构完全相同。在相空间点的连线描述了系统在嵌入维 m 相空间中的演化轨迹，任一相点都是矢量。其中“局部线性法”在预测时，要找到与待预测负荷相关的 m 维矢量 $Y(N_m)$ (就是负荷预测值可由该矢量的线性组合表示) 的那些“邻近矢量” $Y(j)$ ，这些 $Y(j)$ 与 $Y(N_m)$ 的“形状”是很类似的。通过这些“邻近矢量” $Y(j)$ 找到吸引子的演化规律，将其应用到 $Y(N_m)$ 上，就可以得到 $Y(N_m)$ 的一个线性组合 (就是负荷预测值)。在负荷预测中，若取延时时间 $\tau=1h$ ， $m=24$ ，那么局部线性法中的各“邻近矢量” $Y(j)$ ，基本上就是“相似日”方法中根据待预测负荷聚类出的“相似日”。

简单地说，尽管 2 类方法分别属于“模式识别”、“非线性动力系统”，但具有结构上的对应性。根据当前特征聚类出的历史各“相似日”，对应于各“邻近矢量” $Y(j)$ ；由历史相似日经智能处理得到负荷预测值，对应于由“邻近矢量” $Y(j)$ 找到线性组合系数 A 矢量，并用它来得到负荷预测值。这 2 类方法结构上的一致性，也是我们采用“局部线性法”的依据之一。

第三章 基于模糊聚类的负荷与气象数据预处理

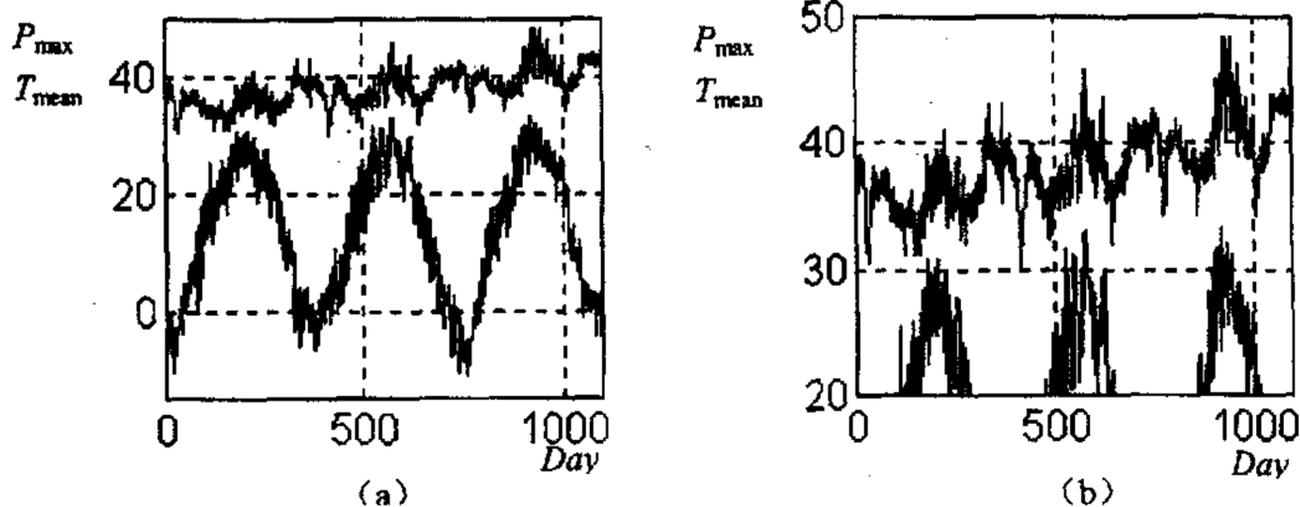
在电力负荷中，负荷的未来发展与过去一定阶段的发展存在相似性，可以根据已知的发展过程和状况来预测负荷的未来变化。然而，对于一定时期的历史日负荷，由于各负荷日的气象条件不同，因此直接根据历史日负荷预测未来负荷，显然不能满足越来越高的预测精度要求。在负荷预测中，可以选择一批与负荷水平有密切关系，对负荷影响较大的特征因子如气象状况、日类型等来进行预测。于是，可以考虑模糊聚类的方法，即将各历史负荷日按照其气象条件归入不同的类别，然后分别对不同的类别建立不同的预测模型。通过模式识别，将待预测日划入合理的类别，再利用已建立的模型进行预测。很显然在这个过程中，确定各历史负荷日的类别是最关键的一步。因此本章重点在于对负荷预测前期影响因子数据进行的预处理和加工，即用不同方法对其进行聚类，从而揭示负荷影响因子与负荷水平之间的内在关系，用以指导负荷预测。

3.1 负荷日特征因子的选取

本文使用的数据是我国北方某地的9种气象因子的日记录，即“最高气温，最低气温，平均气温，降雨量，平均露点，平均海面气压，平均能见度，平均风速，最大持续风速”的日记录。应该说，这九种数据比较全面的反映了负荷日的气象特征状况。但对于负荷预测而言，究竟哪些因素与日负荷有比较密切的关系，还需要作进一步详细的分析。

3.1.1 定性观察：气温、露点和气压的作用

从本文的算例中可以看到，气温不足以充分反映气候对电力负荷的影响。参见图3-1，它们是连续3年的日峰值负荷与同期的一些气象因子的形状对照。



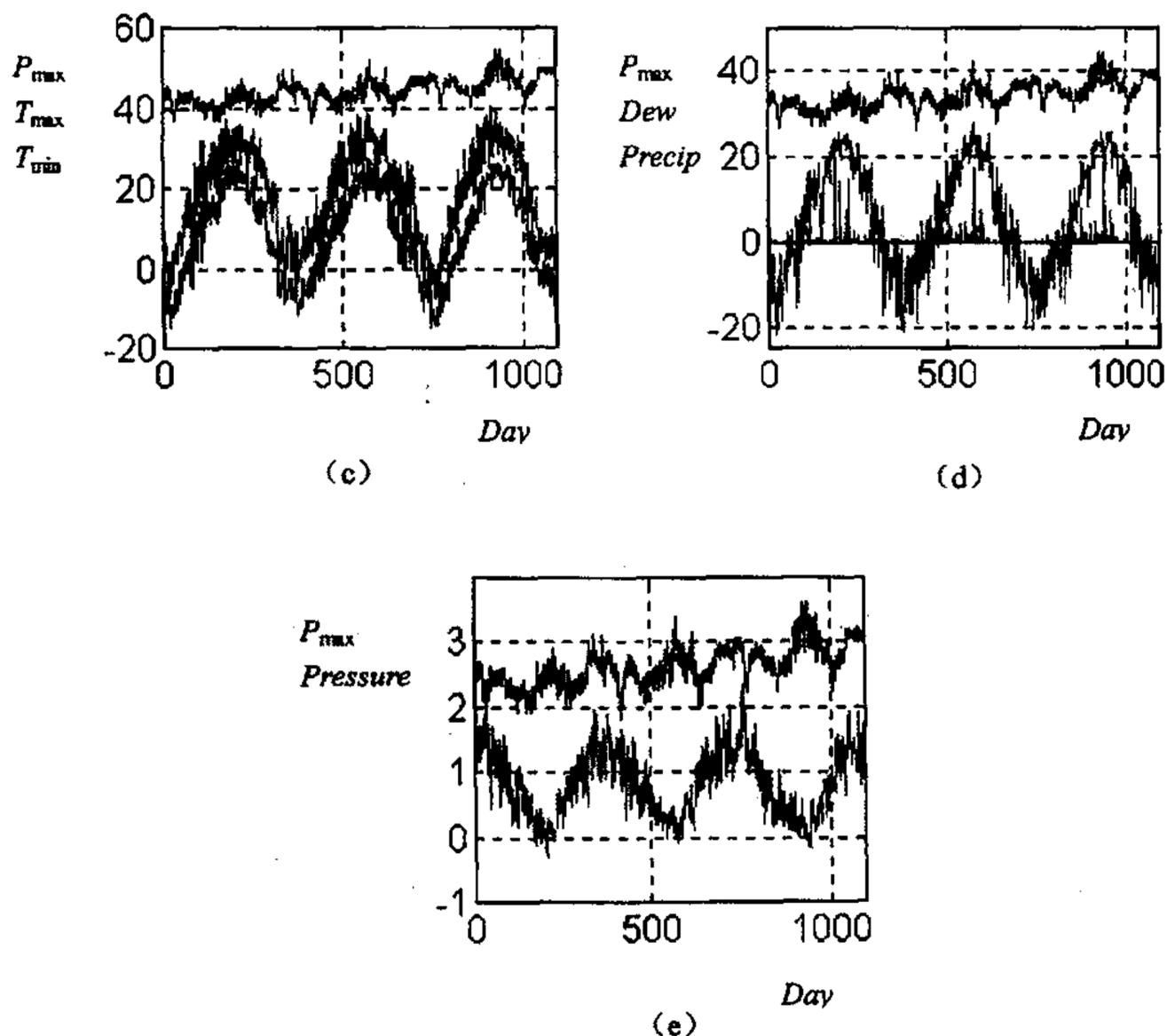


图 3-1 连续 3 年的日峰值负荷与日气象因子形状对照。
 (a) 日峰值负荷-平均气温。 (b) 夏季, 日峰值负荷-平均气温
 (c) 日峰值负荷-最大、最小气温。 (d) 日峰值负荷-平均露点、降雨量。
 (e) 日峰值负荷-平均海面气压。

从图 3-1 可以发现, 无论夏季、冬季, 仅考虑气温, 不足以很好地预测日峰值负荷。在夏季, 日峰值负荷滞后平均气温约 20 天; 而与“平均露点、降雨量”(反映湿度) 有较好的同步性。在冬季, 日峰值负荷与平均气温形状并不很相似(即使扣除春节期间)。平均海面气压在夏季和冬季都与负荷有明显的负、正相关性。从日平均负荷观察得到的结论类似。

这就要求我们从 9 种气象因子来分析负荷-气象关系。

图 3-2 (a) (b) 为该地连续 6 年的日平均气温与日峰值负荷、日平均负荷的散点图。显然气温-负荷是一种具有相当分散性的关系。用人们熟悉的 3 段或多段直线来代替, 是近似的一种方便处理。对负荷与 9 种气象因子的各种观察表明, 可以把平均气温作为划分季节的标准, 即 20°C 以上为夏季, 12°C 以下为冬季。

其合理性可以从图中看出。图 3-2 (c) 为平均气温与露点的散点图。在夏季露点与负荷具有较好的同步性，而气温则超前于负荷。因此“温度-湿度(露点)”是负荷预测必须考虑的。它们与日峰值负荷的关系见图 3-2 (d)。

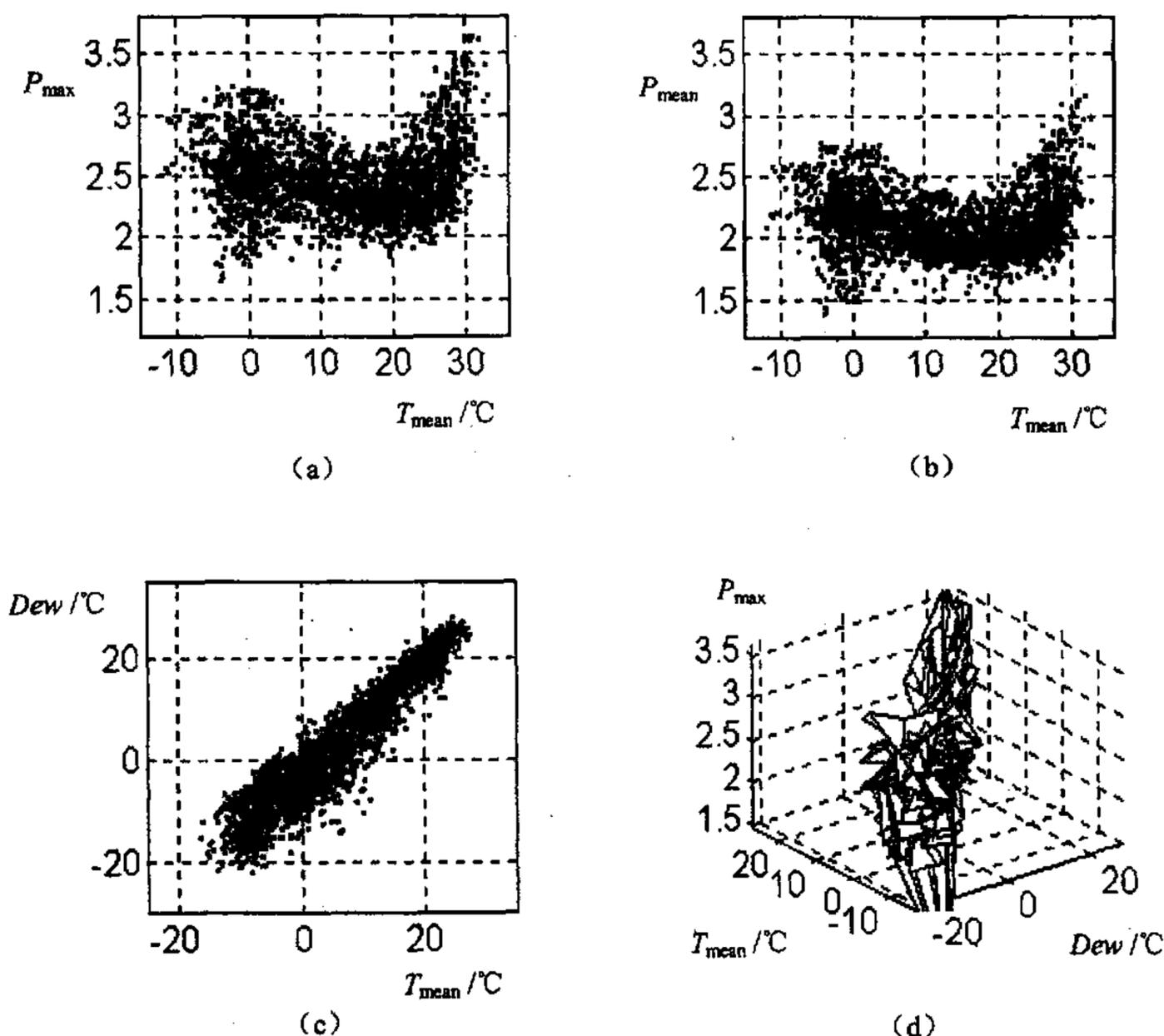


图 3-2 连续 6 年的日负荷、平均气温和平均露点的散点图。
 (a) 日峰值负荷与日平均气温。 (b) 日平均负荷与日平均气温。
 (c) 平均气温-平均露点。 (d) 日峰值负荷-平均露点、平均气温。

总之，一般认为负荷的特性来源于人们的用电行为。“在常压下气温 18~22 $^{\circ}C$ 、空气相对湿度 65%”是使人身体舒适、精力充沛和心旷神怡的“最佳环境指标”。因此夏季“温度-湿度”显然是需要考虑的，并且 20 $^{\circ}C$ 可以作为夏季界限。而夏季、冬季气压的影响以前很少提及，其影响机制也还不很清楚。

3.1.2 四季划分与全年负荷-气象因子相关系数

如上所述，20 $^{\circ}C$ 以上为夏季，12 $^{\circ}C$ 以下为冬季。以某年全年为例，即从 1 月 1 日起，第 90 天到第 134 天为春季（4 月初到 5 月中旬）；第 135 天到第 254 天

为夏季（5月中旬到9月中旬）；第255天到第300天为秋季（9月中旬到10月底）；其余天为冬季。各季节日峰值负荷与9种气象因子的相关系数见表1。

表1 四季日峰值负荷与9种气象因子间的相关系数

相关系数 r	夏季	冬季	春季	秋季
最大气温	0.6670	-0.1716	-0.3362	0.2276
最小气温	0.8088	0.0088	-0.4266	0.0887
平均气温	0.7689	-0.1073	-0.4016	0.1708
降雨量	-0.0170	0.1388	0.1715	-0.1676
平均露点	0.6921	0.1143	-0.3880	-0.0214
平均海面气压	-0.4879	0.2478	0.0817	-0.0159
平均能见度	-0.2005	-0.2760	-0.0839	0.1642
平均风速	-0.2359	-0.2333	0.1169	-0.1642
最大持续风速	-0.2161	-0.2416	0.0577	-0.1566

从上表可以看出，在夏季，“最大气温、最小气温、平均气温、降雨量、平均露点、平均海面气压”是影响负荷的主要气象因子；在冬季，“降雨量、平均海面气压、平均能见度、平均风速、最大持续风速”是影响负荷的主要气象因子。冬季电力负荷与气温关系不大的主要原因，应该是在我国北方，冬季有暖气供应（或其他非电力取暖方式）。春季、秋季的负荷并不是与天气无关，只是相比与夏季冬季关系小些。值得注意的是，在夏季，“降雨量”与负荷的相关系数并不大，这是由于降雨量不是连续量（时常为0）。但分析降雨后第2天峰值负荷与降雨量的相关系数，发现它并不比降雨当天的相关系数小，甚至还大于降雨当天的相关系数，这说明大的降雨量对后几天负荷是有影响的。这个结果符合人们的直观经验。

3.2 对负荷与气象数据的模糊聚类

从第一节分析得知，气象条件对夏冬两季的影响较大，而在负荷预测中研究中，尤以夏季的负荷预测更为困难，因此本节重点研究夏季的气象因子与负荷的关系。限于篇幅，冬季的情况不再叙述，其研究方法与夏季的方法类似。下文中将应用前面介绍的聚类方法对夏季数据进行聚类分析，并对聚类结果作出详细分析和讨论。

3.2.1 聚类样本的选取

聚类样本由负荷历史数据与对应的气象因子数据组成。对于夏季，根据前面

的分析,我们选取六种与负荷相关系数比较大的气象因子作为样本的预测前期因子特征值表,即“最大气温、最小气温、平均气温、降雨量、平均露点、平均海面气压”。而对于负荷数据,我们知道,日峰值负荷是决定电厂当天投入运行的机组设备规模和旋转储备的关键因素,意义非凡。故对未来日的峰值负荷预测是很重要的一项技术指标,所以本文重点考虑这种负荷,针对日峰值负荷数据附加六种与负荷相关性比较大的气象因子组成样本。即,对于日峰值负荷而言,每一个样本 u_i 由 7 种数据项组成。我们选择待预测日前 30 天的历史数据 U 作为预测前期数据集。在冬季,由前一节知,与负荷相关性大的气象因子为“降雨量、平均海面气压、平均能见度、平均风速、最大持续风速”。因此,日峰值负荷情况下的样本特征维数为 6。本文只针对天气影响最大的夏季负荷与气象样本数据进行分析。

为简单起见,下文中使用 2000 年第 181 到 210 天的数据进行聚类算例分析(太大的数据量处理导致运算时间快速增加),并简称“2000 年第 181 天到第 210 天”为“第 1 天到第 30 天”。下面对日峰值负荷数据与相应气象数据进行研究。

3.2.2 聚类结果与分析

应用第二章中介绍的传递闭包法聚类具体步骤如下:

(1) 数据的标准化:由于数据集 U 中各种数据之间的量纲、单位和数值大小各异,故需转化为相对值以消除量纲、单位等的影响,在第二章中介绍了几种数据规格化的方法,针对本文的具体实例,采用如下的标准化公式:

$$x_{ik} = \frac{x_{ik} - x_{k \min}}{x_{k \max} - x_{k \min}} \quad (3-1)$$

这里, $i=1,2,\dots,30$ 为样本数目; k 为样本维数。对于日峰值负荷, $k=1,2,\dots,7$; $x_{k \max}$ 、 $x_{k \min}$ 分别为 $x_{1k}, x_{2k}, \dots, x_{nk}$ 中的最大值、最小值; x_{ik} 为处理后的数据。显然有 $x_{ik} \in [0,1]$

(2) 标定:建立 U 上的模糊相似关系 R , R 的隶属度函数 $\mu_R(u_i, u_j)$ 表示 u_i 和 u_j 按其性质的相似程度。 R 可表示为相似矩阵 $R=[r_{ij}]$, 这里 $r_{ij} = \mu_R(u_i, u_j)$ 。在本文的第二章中我们给出了很多种隶属度函数,本文不打算一一采用,只选择两种较为典型的隶属度函数作为参考,以比较它们的不同隶属度函数的聚类效果,并根据已有的隶属度函数,针对负荷与气象历史数据这个特定的聚类对象,提出一种新的隶属度函数。

(3) 聚类:根据模糊理论,在求出相似矩阵 R 后,通过 R 的模糊等价关系矩阵 R^* , 用直接聚类法进行聚类。聚类水平 λ 的大小直接影响聚类结果,当 λ 从 1 降至 0 时,分类由细变粗逐渐归并,形成一个动态聚类图。

接下来的小节中, 首先介绍本文拟采用的隶属度函数, 即相关系数隶属度函数和欧氏距离隶属度函数, 并在此基础上构造“相关系数+欧氏距离”隶属度函数, 最后从理论上分析构造的隶属度函数对于负荷与气象数据的聚类具有的优点。

3.2.2.1 三种隶属度函数

(1) 相关系数法

相关系数法是最常见的隶属度函数, 它的表达式如下:

$$r_{ij} = \frac{\sum_{k=1}^m (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)}{\sqrt{\sum_{k=1}^m (x_{ik} - \bar{x}_i)^2} \sqrt{\sum_{k=1}^m (x_{jk} - \bar{x}_j)^2}} \quad (3-2)$$

由于在求出的相关系数中可能出现负值, 而建立模糊相似矩阵时要求元素都必须大于 0, 因此需要对相关系数作一定的处理, 即令:

$$r'_{ij} = \frac{r_{ij} + 1}{2} \quad (3-3)$$

这样处理后的相关系数都处于 0 到 1 之间。

(2) 欧氏距离法

欧氏距离法隶属度函数如下

$$r_{ij} = 1 - c \cdot \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2} \quad (3-4)$$

其中, m 为样本维数; x_{ik} 为第 i 个样本的第 k 个分量, x_{jk} 为第 j 个样本的第 k 个分量; c 为一适当的参数。

(3) “相关系数+欧氏距离”法

尽管相关系数法与欧氏距离法是模糊聚类中应用得很多的隶属度函数, 但对具体的聚类对象, 还应该根据聚类对象的物理本质寻求更好的隶属度函数。本文针对负荷与气象数据这种特定的聚类对象, 构造的“相关系数+欧氏距离”隶属度函数如下

$$r_{ij} = a \cdot r_{coeff} + (1-a) \cdot r_{eu} \quad (3-5)$$

其中, r_{ij} 为构造的隶属度函数, r_{coeff} 为相关系数隶属度函数, r_{eu} 为欧氏距离隶属度函数, a 为加权系数。

一般认为, 相关系数主要反映的是各样本间的线性关系, 即样本在形状上的相似性; 而欧氏距离主要反映的是各样本在空间上距离的远近, 距离较近的样本之间的相似性也较大。但是, 对于电力负荷来说, 时常出现的情况是, 样本特征组成的曲线形状很相似而数值却存在较大的差别, 或者是样本间各分量的总欧氏距离很接近但样本曲线形状差异很大。这些情况下, 如果在模糊聚类时仅选择相关系数作为隶属度函数, 则容易将样本特征组成的曲线形状相似而数值迥异的负荷日划为同一类。相反, 如果在聚类时仅选择欧氏距离作为隶属度函数, 则会将距离接近但形状迥异的负荷日划为同一类。这两种隶属度函数都不足以反映日负荷和气象数据的本质特征。正是基于此, 本文构造了上述第三种隶属度函数。在下面的小节中, 我们将讨论具体的聚类结果。

3.2.2.2 三种方法的聚类结果

由于模糊聚类的结果会因聚类截水平 λ 的不同而有所不同, 因此在聚类时一定要有限定条件来指导聚类。本文中, 限定聚类结果中同为一类的连续天数的最大值不超过 5, 这是由电力负荷的周周期性决定的。因为一般来讲, 周末的负荷与工作日的负荷不应该是一类。但因为气象因素的影响, 有时周末呈现的负荷同工作日也具有一定的相似性, 不能一概而论, 需具体分析。

(1) 相关系数法的聚类结果示例

在相关系数隶属度函数下, 对于日峰值负荷与气象数据, 根据限定条件由程序得出的聚类截水平为 $\lambda=0.96$, 聚类结果如图 3-3。

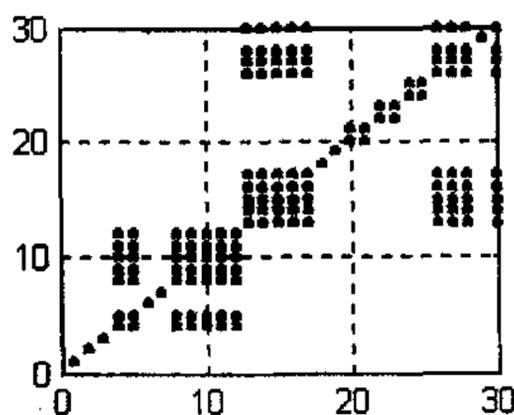


图 3-3 相关系数法日峰值负荷和气象数据的聚类结果
 $\lambda=0.96$

(2) 欧氏距离法的聚类结果示例

此时聚类结果见图 3-4, 聚类截水平为 $\lambda=0.968$ 。

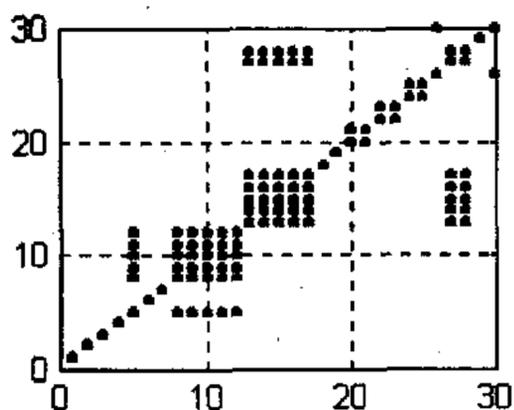


图 3-4 欧氏距离法日峰值负荷和气象数据的聚类结果图
 $\lambda = 0.97$

(3) “相关系数+欧氏距离”法

本文中，加权系数 a 的选择是这样的：分别求出由相关系数法和欧氏距离法得到的 r_{ij} 元素的平均值 m_1 、 m_2 ，取 $a = m_2 / (m_1 + m_2)$ 。而 c 则不能太大，否则将使 r_{ij} 为负值，这里取 $c = 0.1$ 。

此时由限定条件得到聚类截水平为 $i = 0.968$ 。聚类结果见图 3-5。

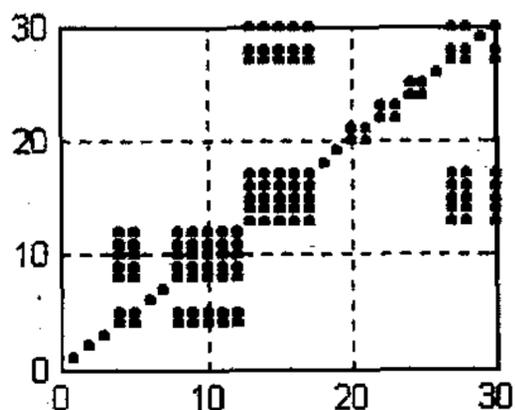


图 3-5 “相关系数+欧氏距离”法峰值负荷和气象数据聚类结果图
 $\lambda = 0.968$

3.2.2.3 聚类效果的比较和解释

对于上节中的不同方法的理论上的聚类结果，我们将结合原始数据进行分析和比较，并证实采用构造的隶属度函数进行聚类得到的结果确实优于前两种方法。

首先绘出原始样本数据的曲线图，见图 3-6。图中每条曲线为不同样本对应的相同特征指标形成的曲线，反映出同一指标的 30 天的变化情况。

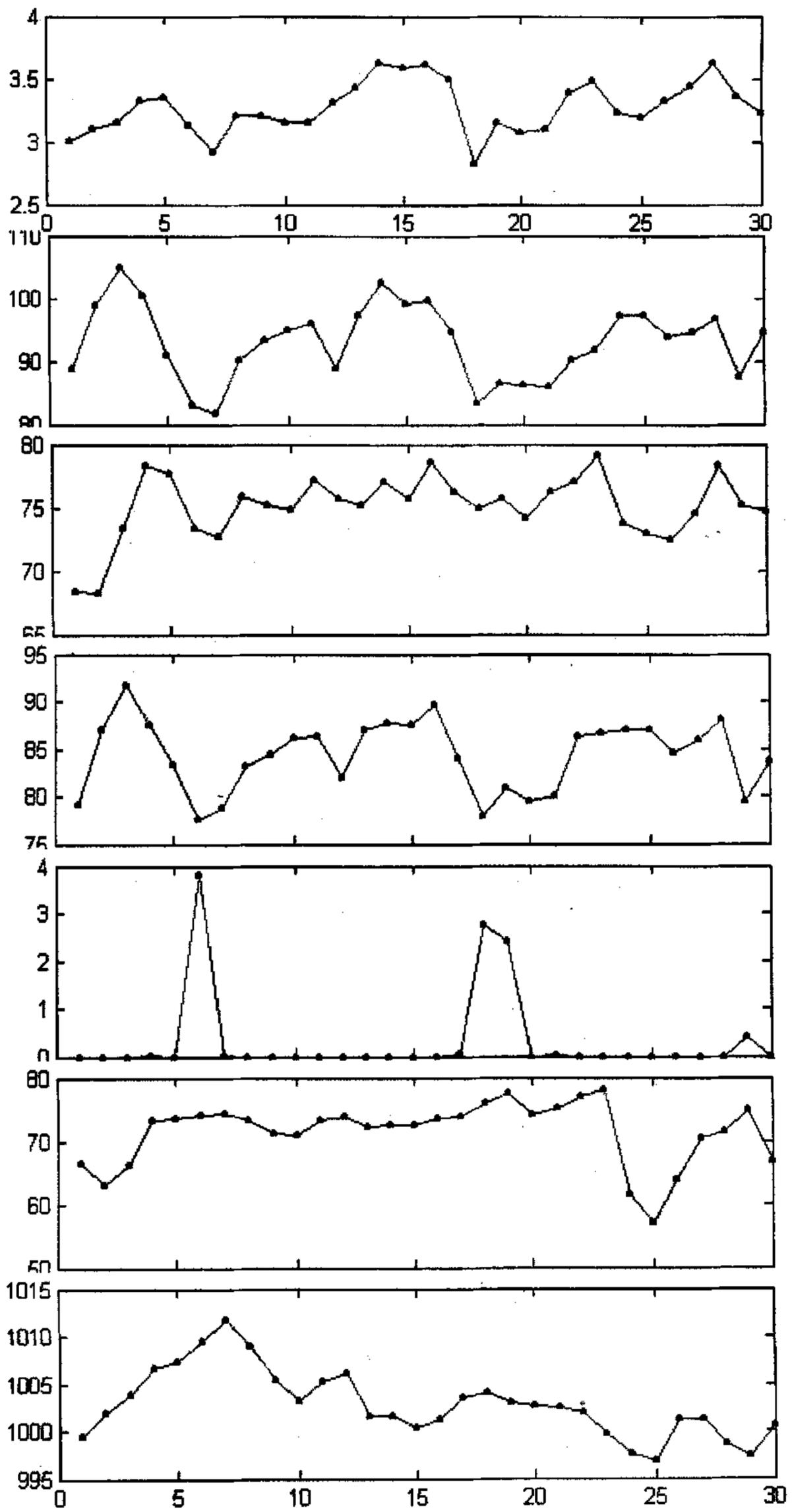


图 3-6 原始聚类数据曲线

从上至下依次为 30 天的日峰值负荷曲线、最大温度曲线、最小温度曲线、平均温度曲线、降雨量曲线、平均露点曲线、平均海平面压力曲线

为便于对照和比较,我们也将三种隶属度函数下得到的聚类结果图重新示于图 3-7。

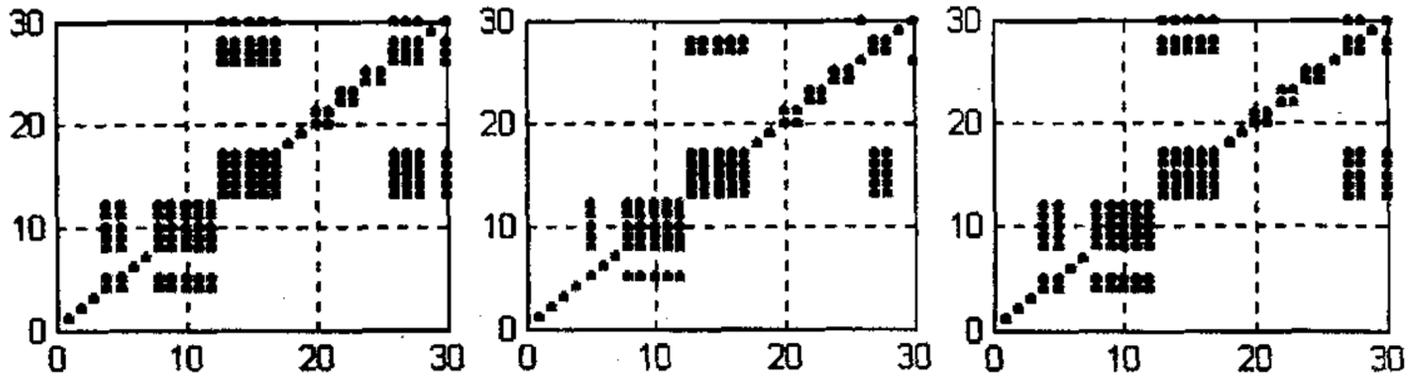


图 3-7 三种隶属度下的日峰值负荷与气象因子数据的聚类结果
左中右图依次为相关系数法、欧氏距离法、“相关系数+欧氏距离”法

首先比较一下图 3-7 中三幅聚类图的差异。从图中看到,三种方法总体聚类效果比较接近。即第 5, 8, 9, 10, 11, 12 天为一类设为 A 类;第 13, 14, 15, 16, 17, 27, 28 天为一类,设为 B 类;还有一些样本较少的类别,这里不另作讨论。分析原始数据图 3-6 可以看到,第 5, 8, 9, 10, 11, 12 天样本对应的各气象特征指标确实很接近;第 13, 14, 15, 16, 17, 27, 28 天样本对应的各气象特征指标也很接近。不同的是两类日峰值负荷水平不同,B 类日峰值负荷明显比 A 类大,这是酷暑日的体现,故将它们分开是合理的。

而三种方法的聚类差别主要体现在某几天,即对第 4, 26, 30 天的划分有所不同。相关系数法与“相关系数+欧氏距离”法的聚类差异在于:相关系数法将第 26 天划归入 B 类;而“相关系数+欧氏距离”法则认为第 26 天不应划归入 B 类。欧氏距离法与“相关系数+欧氏距离”法的聚类差异在于:欧氏距离法将第 4 天单独划分开,且认为第 26, 30 天单独为另一类。而“相关系数+欧氏距离”法则认为第 4 天划归入 A 类,第 26 天单独为一类,第 30 天划归入 B 类。

下面进一步结合原始数据分析存在差异的原因。由于原始数据数量级不一致,故此处采用的是原始数据标准化后的数据。本文将聚类结果有差异的几天的样本特征曲线和各类别中的“典型日”的样本特征曲线绘于图 3-8。各类别“典型日”的定义如下:将同一类别中各样本的对应特征指标的平均值作为该类别的“典型日”。

从图 3-8 中可以看到, 第 26 天和 B 类的“典型日”特征曲线存在一定的相似性, 然而它们之间的距离太大 (即特征因子值差太大), 这就使得按相关系数法聚类时将第 26 天划分到 B 类中, 而按加权后的“相关系数+欧氏距离”法聚类则没有将第 26 天归入 B 类中。类似, 第 4 天和 A 类的“典型日”如按欧氏距离法聚类不属同类, 因为它们之间距离仍较大, 但是可以看到, 它们的特征曲线相似性大, 趋势基本一致。因此按“相关系数+欧氏距离”法聚类时它们属于同一类。第 30 天的分析同理。

以上结合原始数据的分析进一步论证了前文在理论上给出的“相关系数+距离”法的优点, 其聚类结果具有更好的合理性。

因此, 对于本文的算例, 按“相关系数+欧氏距离”法得到的聚类结果是可以用于指导负荷预测的, 即第 4, 5, 8, 9, 10, 11, 12 天属一类, 第 13, 14, 15, 16, 17, 27, 28, 30 天属一类, 另外第 20, 21 天, 第 22, 23 天以及第 24, 25 天分别各为一类, 其他天各自自成一类。

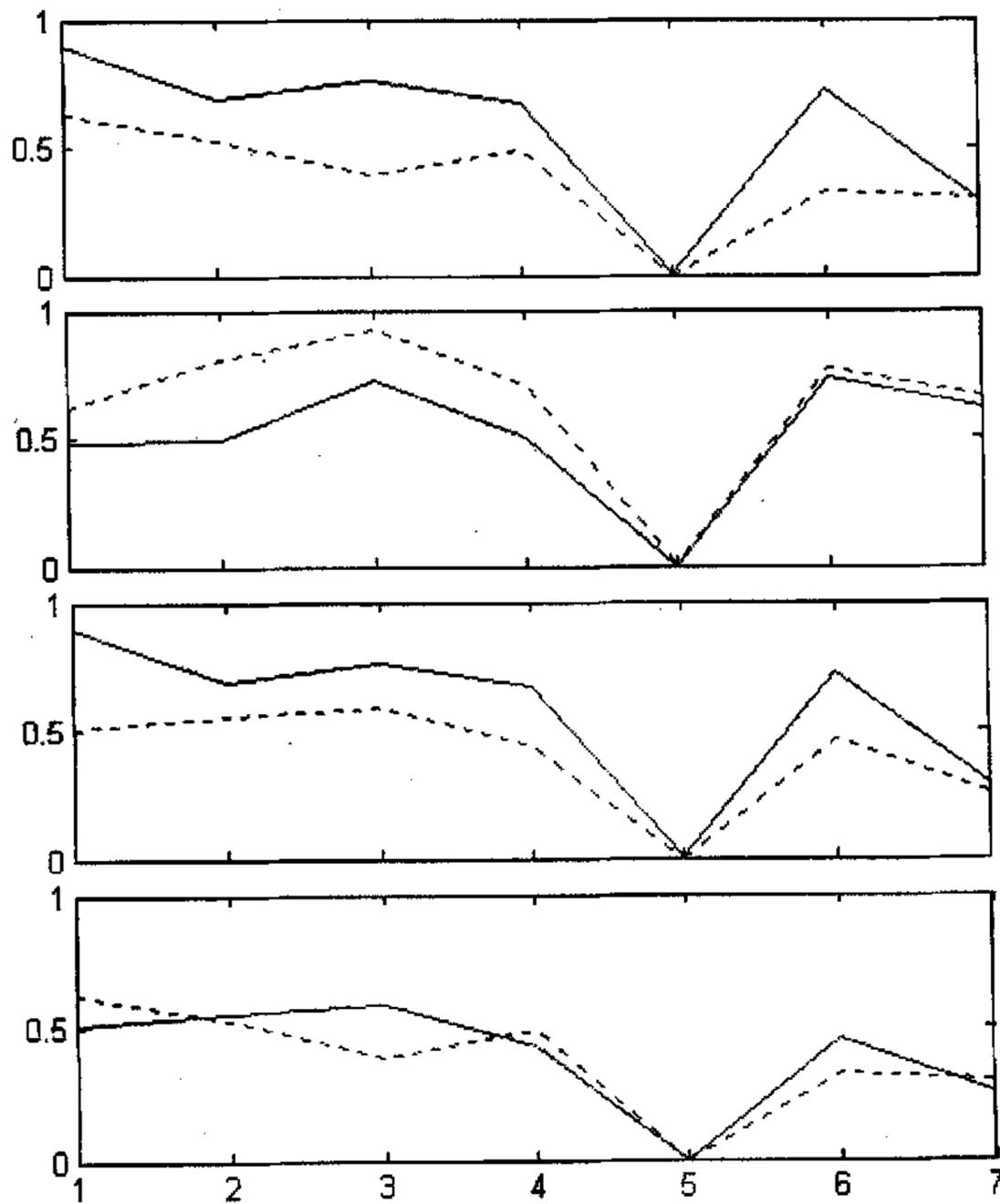


图 3-8 聚类结果有差异的样本曲线和各类别“典型日”曲线

从上到下依次为第 26 天（虚线）和 B 类的“典型日”曲线（实线）、第 4 天（虚线）和 A 类的“典型日”曲线（实线）、第 30 天（虚线）和 B 类的“典型日”曲线（实线）、第 26 天（虚线）和第 30 天曲线（实线）

3.3 待预测日的模式识别

经过聚类后，各历史负荷日划归入不同的类别后，就可以对待预测日的类别应用模式识别进行判断，在进行模式识别前，首先根据聚类情况，按下式求出各类的聚类中心

$$\bar{x}_{jk} = \frac{1}{m} \sum_{i=1}^m x_{ik} \quad k=1, 2, \dots, n \quad (3-6)$$

式中，包含 m 个样本的第 j 类的聚类中心为： $\bar{x}_{j1}, \bar{x}_{j2}, \dots, \bar{x}_{jn}$ ，是一个 n 维矢量；本文中对于日峰值负荷， $n=7$ 。

计算出个类别的聚类中心后，按下式逐一计算待预测日的特征矢量与各聚类中心的欧氏距离

$$d_{ij} = \sqrt{\frac{1}{n} \sum_{k=1}^n (x_{ik} - \bar{x}_{jk})^2} \quad (3-7)$$

比较各距离，将距离最短的类别作为待预测日的类别，再在相应的类别中应用现有的预测方法进行负荷预测，如神经网络，混沌相空间重构等。

这种先通过聚类方法把历史数据进行分类处理，然后由模式识别方法将待预测日划归到相应的类别中的预测方法，从本质上讲属于负荷预测中的相似日法，能充分考虑天气等相关因素的影响，具有极大的应用前途。如果能在聚类前进行坏数据的剔除工作，相信聚类结果会更佳。

3.4 小结

本章首先通过计算 9 种气象因子与对应天负荷的相关系数发现：在夏季，“最大气温、最小气温、平均气温、降雨量、平均露点、平均海面气压”是影响负荷的主要气象因子；在冬季，“降雨量、平均海面气压、平均能见度、平均风速、最大持续风速”是影响负荷的主要气象因子。进而以夏季第 181 天到 210 天的日峰值负荷附加气象因子数据为样本，对样本进行了详细的模糊聚类分析。采用的隶属度函数有：相关系数法，欧氏距离法。并根据样本的物理本质提出利用加权后的“相关系数+欧氏距离”法作为隶属度函数的聚类方法，通过对三种隶属度

函数下聚类结果和原始数据的具体分析, 论证了采用“相关系数+欧氏距离”法的聚类结果优于单独的相关系数和欧氏距离法。最后简单给出了待预测日特征量的模式识别方法, 从而完成了预测前数据的预处理过程。

第四章 负荷预测的混沌相空间二阶多项式回归法

大量研究表明,电力系统负荷记录存在混沌特性,并且可以利用混沌方法进行短期负荷预测^[43,44]。由于相空间重构法适用于各种吸引子的重构,如混沌、准周期等吸引子的重构,因此它经常成为负荷预测研究的重点之一。但是,在相空间重构中采用线性回归方法,会出现12h的周期性误差增大现象(原因是半天周期引起的延时相轨迹重叠现象),并且预测精度也受到一定的限制。于是相空间重构中采用非线性方法就成为值得研究的对象。本文在负荷(混沌)相空间重构中,采用二阶多项式回归方法。这种非线性方法具有比线性方法更好的预测精度,并且还显著抑制了12h的周期性误差增大现象^[45]。然后通过大量的数值仿真,初步探讨了相空间重构中的嵌入相空间的维数 m 、延时时间 τ ,以及选择邻近点时使用的距离 ε 的优化。

4.1 方法要点

对于一个吸引子,包括混沌吸引子或其它吸引子,可以通过相空间重构来预测它的时间行为。基本方法是将吸引子的 N 个观测值构成的时间序列 $\{x(i)\}$, $i=1, 2, \dots, N$, 嵌入 m 维相空间中。假设当前值为 $x(T)$,我们预测下一个值 $x(T+1)$ 。二阶多项式回归法(这是一种非线性方法)的要点为:

(1) 首先需要构造 $N_m = N - (m-1)\tau$ 个 m 维相空间中的点(就是 m 维矢量):

$$Y(k) = [x(k), x(k+\tau), x(k+2\tau), \dots, x(k+(m-1)\tau)], \quad k=1, 2, \dots, N_m,$$

$\tau > 1$ 为延时时间,其目的是使 $Y(k)$ 在吸引子上占据合适的轨道长度。

(2) 用矢量 $Y(k)$ 中的各个分量,按照二阶多项式方式来表示出标量 $x(k+1) = f(Y(k))$ 。待预测值可以表示为 $z(T+1) = f(Y(N_m))$, $Y(N_m)$ 有时被称为“参考矢量”。这里 $f(\bullet)$ 的形式为^[46]:

$$\begin{aligned} x(k+1) = & \beta_0 + \beta_1 x(k+\tau) + \beta_2 x(k+2\tau) + \dots + \beta_m x(k+(m-1)\tau) \\ & + \beta_{11} x(k+\tau)^2 + \beta_{22} x(k+2\tau)^2 + \dots + \beta_{mm} x(k+(m-1)\tau)^2 \\ & + \beta_{12} x(k+\tau)x(k+2\tau) + \beta_{13} x(k+\tau)x(k+3\tau) + \dots + \beta_{(m-1)m} \\ & x(k+(m-2))x(k+(m-1)). \end{aligned}$$

这里各系数 β_i 是未知的。 β_i 的个数是 $(1 + 2m + m(m-1)/2)$ 。

(3) 用吸引子的性质来估计各系数 β_i 。具体方法是,找到 $Y(k)$ 中与 $Y(N_m)$ 的距离小于 ε 的那些“邻近矢量” $Y(j)$ 。这些 $Y(j)$ 是与 $Y(N_m)$ 比较相似的矢量。从吸引子上看,它们是很靠近 $Y(N_m)$ 的。然后再找到各个 $Y(j)$ 对应的 $x(j+1)$,即矢

量形式的 $X(j+1)=f(Y(j))$ 。运用最小二乘法估计出各系数 β_s 。

更细节的过程，可以参见文献^[46]的第八章等。

4.1.1 3个参数的选择

相空间重构中需要 3 个参数：嵌入相空间的维数 m 、延时时间 τ ，以及选择“邻近矢量”时使用的距离 ε 。混沌研究中现有的观点有：

(1) 嵌入定理：如果原来吸引子处在一个 d_A 维空间，那么嵌入相空间的维数 $m \geq 2d_A + 1$ 。电力系统负荷记录的维数大于 2，根据嵌入定理， $m \geq 5$ 。有时也将该结果表示为“为描述该系统所需的最小独立变量数为 $\text{INT}(d_A + 1)$ 个，而最多（充分）的独立变量数为 $\text{INT}(2d_A + 1)$ 个。INT 为取整函数。”^[14] 依此， $m > 3$ 是必须的， $m > 5$ 是充分的。

(2) 延时时间 τ 的选择，方法较多^[47]。一般以 $x(i)$ 的自相关系数首次过 0 时的延时时间来确定 τ 。对于负荷记录， $\tau \approx 4 \sim 6$ 。

(3) 嵌入相空间上的距离 ε 的确定，一般要经过试探。

4.1.2 待研究的问题

由于电力系统负荷记录不是严格的混沌行为，而是包含混沌特性在内的一种复杂的混合行为，因此 m 、 τ 和 ε 这 3 个参数的选择，还需要通过数值测试来提供一定的优选方法。

本文对距离 ε 的选择，提出了一种改进方法：一般地，我们很难根据 ε 来事先估计出有多少个“邻近矢量” $Y(j)$ ，在一个预测点的预测中被选中。由于估计各系数 β_s 需要足够数量的 $Y(j)$ ，因此我们选用与 $Y(N_m)$ 最近的 $(2 + 2m + m(m-1)/2)$ 个“邻近矢量” $Y(j)$ ，来估计各系数 β_s 。这样不仅保证了在任何情况下都能找到足够的“邻近矢量”来进行负荷预测，数值测试还表明，这样还可以提高预测精度：过多的“邻近矢量”，通常反而降低预测精度；原因在于会包含一些离“参考矢量” $Y(N_m)$ 较远的“邻近矢量”，这些矢量与“参考矢量”的性质差异比较大。

4.2 二阶多项式回归法负荷预测示例

4.2.1 预测采用的负荷样本

图 4-1 (a) 为我国北方某年（近年）的负荷记录的相对值，图 4-1 (b) 为其

3 天窗口上的最大李雅普诺夫指数 λ_1 的分布。可见全年的负荷记录都存在混沌性质。已知吸引子的周期为 24h。采用上面的相空间重构的二阶多项式回归法, 对第 340、341、342 天连续 3 天进行预测。各预测点的绝对误差为该点的预测值与实际记录的差值。 ε_r 为某预测点的相对误差, E_r 为连续 3 天 (或 1 天) 预测相对误差的绝对值的平均值。

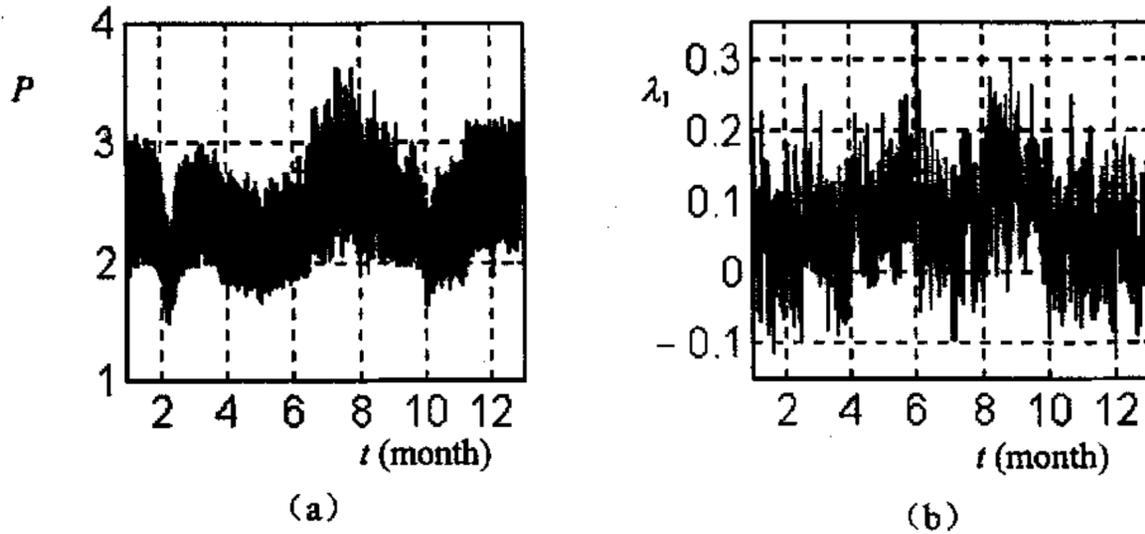


图 4-1 (a) 负荷记录的时间序列。(b) 该序列 3 天窗口上的 λ_1 。

4.2.2 预测结果示例

取 $m=11$ 、 $\tau=17h$, 从第 340 天开始进行连续 3 天的整点负荷预测 (离线预测), 预测结果、各点的相对误差分别如图 4-2 (a)、(b) 所示。

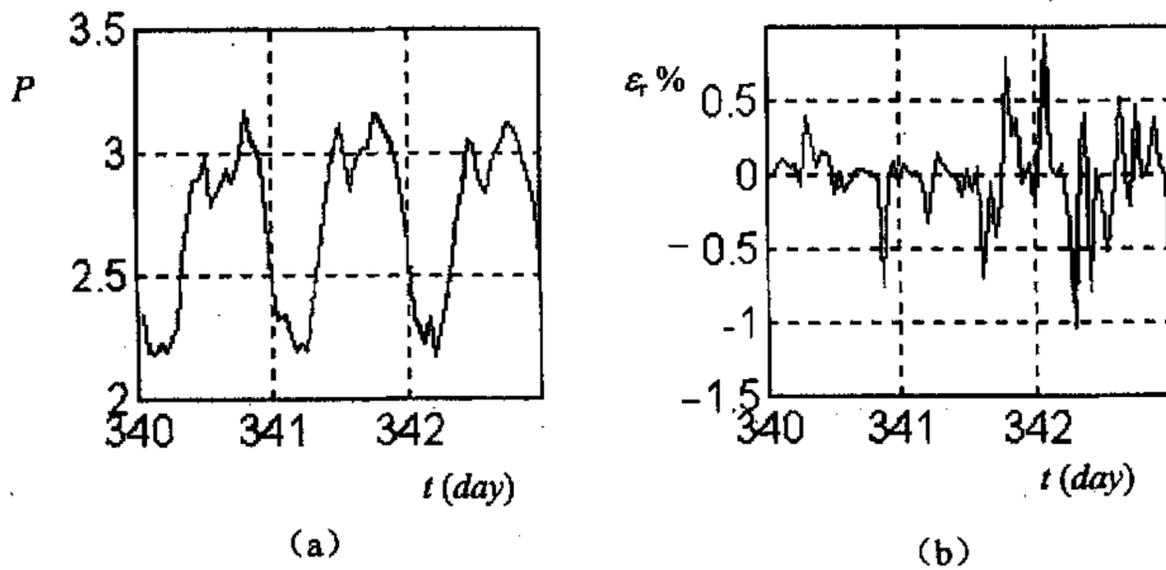


图 4-2 (a) $m=11$ 、 $\tau=17h$ 时的连续 3 天离线预测结果。
(b) 相应的各预测点相对误差。

连续 3 天离线预测的平均误差为 0.1933%, 最大误差 1.0323%。其中第 1 天离线预测平均误差 0.1113%, 最大误差为 0.7494%。在图 4-2 (a) 中, 由于预测

值和实际负荷记录差别较小，因此几乎重叠在一起。

可以看出，采用二阶多项式回归法进行混沌相空间重构来预测负荷，与线性回归法^[43, 44]相比，不仅提高了预测精度，还极大地消除了 12h 的周期性误差增大的现象。

4.2.3 3 个参数的数值测试与优选

采用什么样的嵌入相空间的维数 m 、延时时间 τ 和嵌入相空间上的距离 ε (本文改进为“邻近矢量” $Y(j)$ 的个数)，会得到优化的预测结果？

4.2.3.1 “邻近矢量” $Y(j)$ 的个数的优选

为了保证可靠地估计出各 β_i ， $(1 + 2m + m(m-1)/2)$ 个“邻近矢量” $Y(j)$ 是必须的，因为 β_i 有 $(1 + 2m + m(m-1)/2)$ 个。选 $m=11$ 、 $\tau=17$ h 进行测试，取 $k=-3, -2, -1, 0, 1, 2, \dots$ ，观察采用 $(k+1 + 2m + m(m-1)/2)$ 个“邻近矢量”时预测误差的变化，见图 4-3。

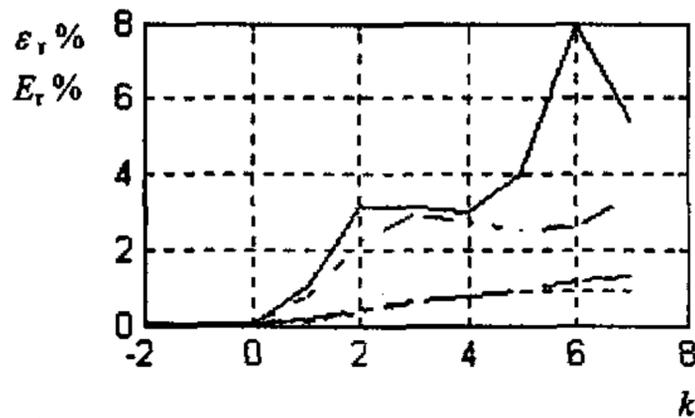
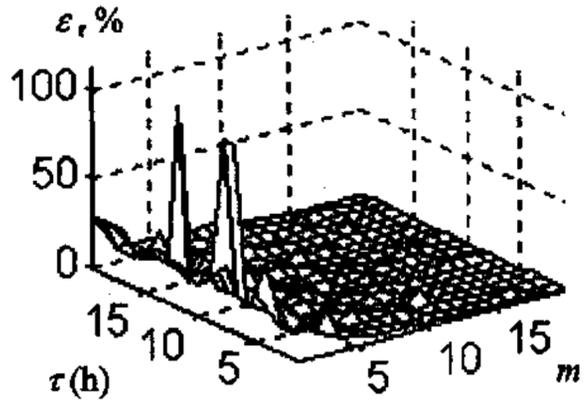


图 4-3 误差随 k 的变化情况：从上到下，4 条线分别为 3 天、1 天预测最大误差，3 天、1 天预测平均误差。条件 $m=11$ 、 $\tau=17$ h。

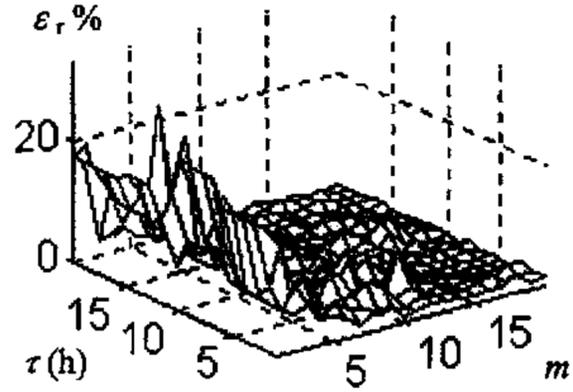
可见， $k=0$ 是比较优化的。为了保证预测的可靠性，本文一般采用 $k=1$ 。原因是取 $k=0$ 有时会在计算中出现接近奇异的矩阵。

4.2.3.2 m 、 τ 的优选

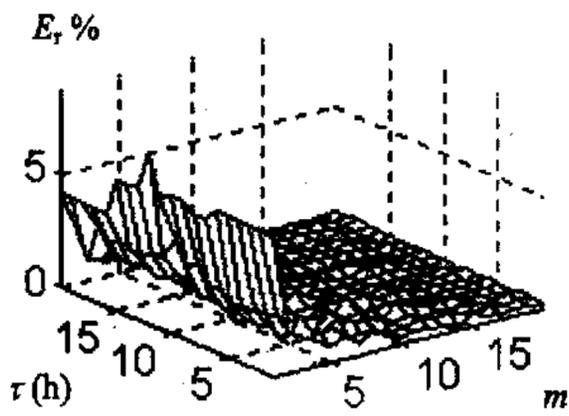
选最近的 $(2 + 2m + m(m-1)/2)$ 个“邻近矢量” $Y(j)$ 进行负荷预测， m 、 τ 的数值测试结果如图 4-4 所示。测试的参数范围 $m=1, 2, \dots, 20$ ， $\tau=1, 2, \dots, 20$ h。 $\varepsilon_r \%$ 为最大相对预测误差， $E_r \%$ 为平均相对预测误差。



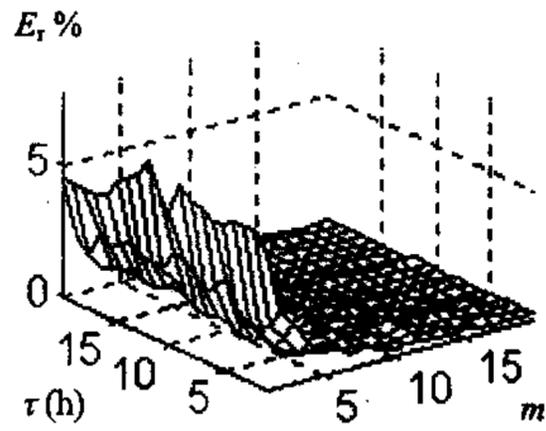
(a)



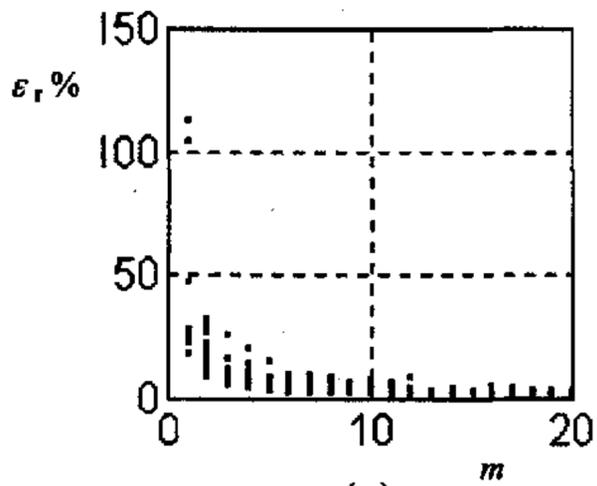
(b)



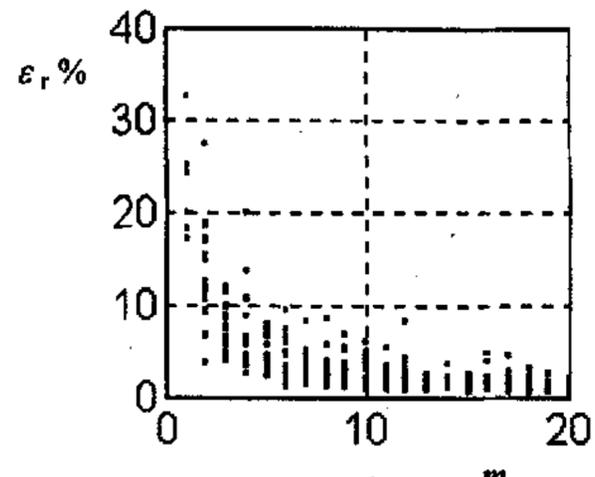
(c)



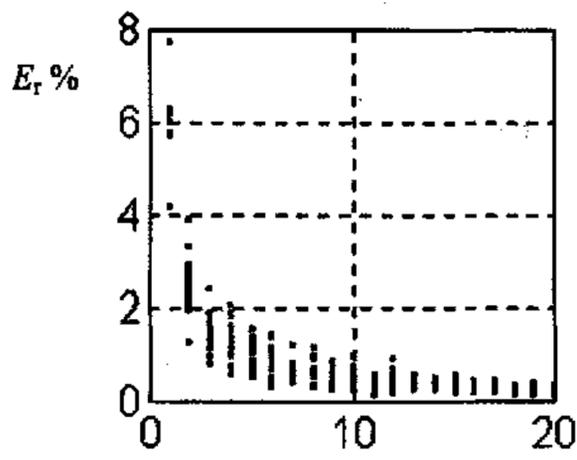
(d)



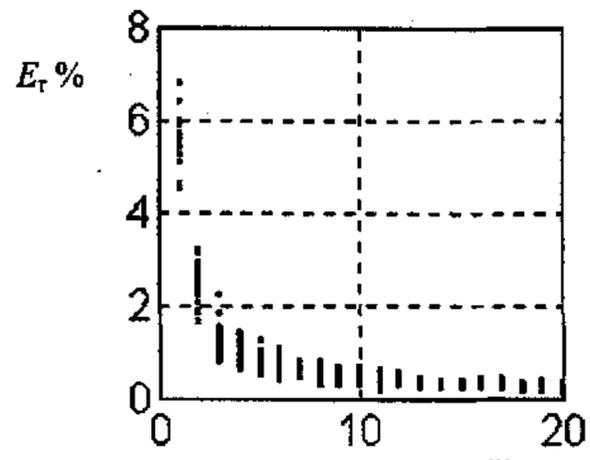
(e)



(f)



(g)



(h)

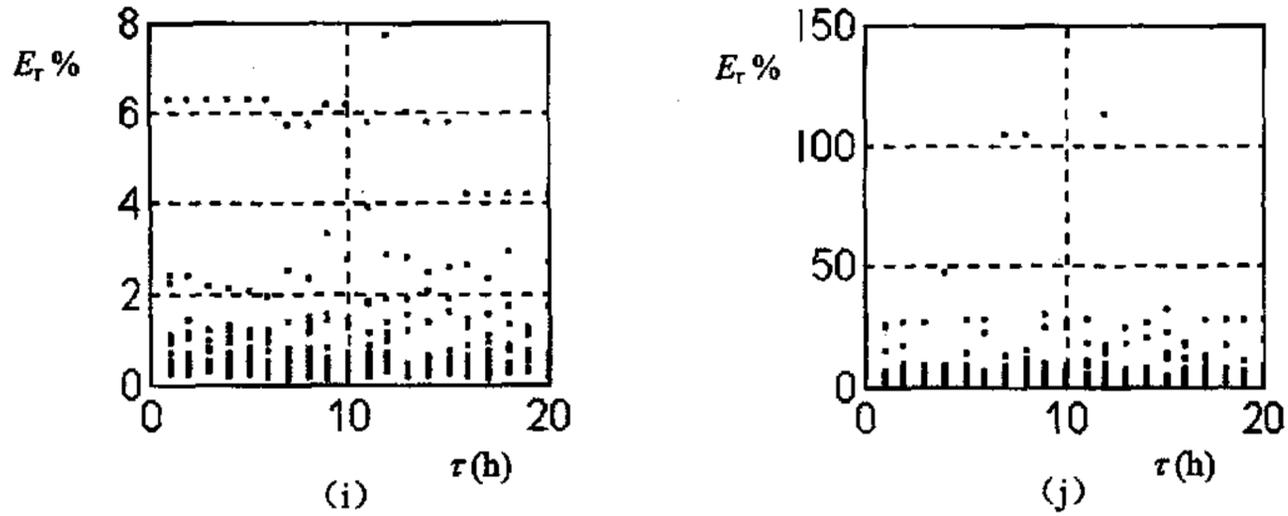


图 4-4 离线预测误差与 m 、 τ 的关系：(a) 3 天最大误差，(b) 1 天最大误差，(c) 3 天平均误差，(d) 1 天平均误差。
 预测误差在 m 上的投影：(e) 3 天最大，(f) 1 天最大，(g) 3 天平均，(h) 1 天平均。
 预测误差在 τ 上的投影：(i) 3 天平均，(j) 1 天平均。

可见， m 的取值对各误差的影响都很明显。一般地， m 增大，各误差都会减小。因此在实际预测中应该考虑选取一个比较大的 m 值。但是， m 的增大，也会使计算时间快速加长。 τ 的取值，对预测误差的影响相对较小。例如 $m=11$ 时各预测误差随 τ 的变化如图 4-5 (a) 所示。图 4-5 (b) 是一些误差较小的 m 、 τ 的搭配，其中主要是 m 增大的效果，而与 m 、 τ 的搭配规律性不很明显。

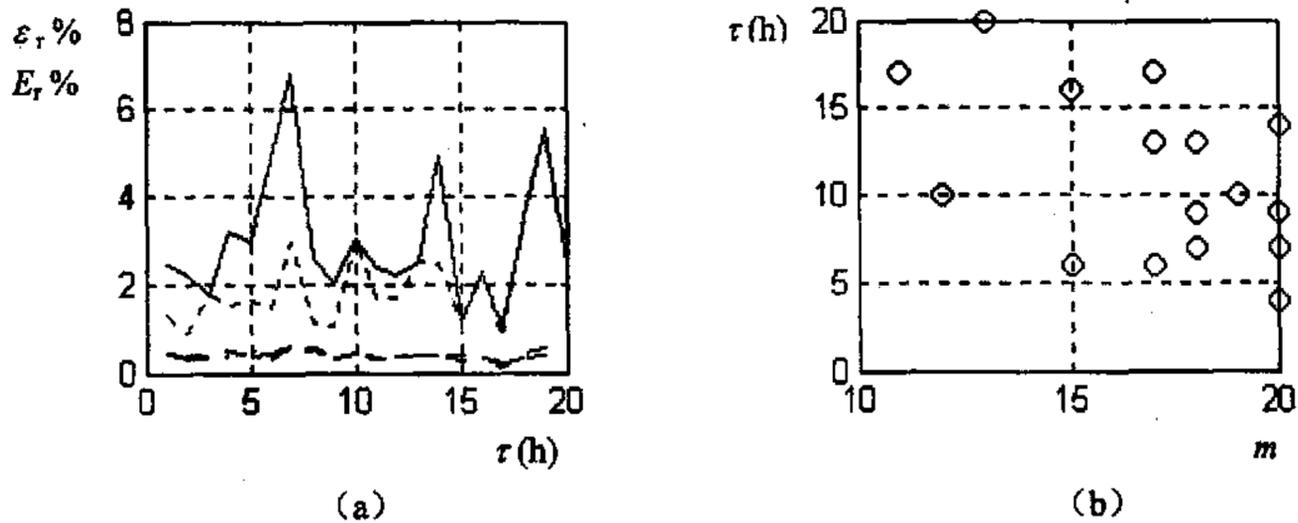


图 4-5 (a) $m=11$ 时预测误差随 τ 的变化：从上到下，4 条线分别 3 天、1 天预测最大误差，3 天、1 天预测平均误差。(b) 误差较小的一些 m 、 τ 的搭配。

4.3 测试结果成因的简单讨论

(1) “邻近矢量” $Y(j)$ 的个数，以 $(1+2m+m(m-1)/2)$ 个为下限。一般可取 $(2+2m+m(m-1)/2)$ 个。个数过多时，会由于这些“邻近矢量”与对应待预测

点的“参考矢量”的差异增大，反而造成预测误差增大。

(2) m 越大，预测误差越小。原因是相空间矢量对应的延时相轨迹的长度加长，从而二阶多项式回归法对相空间轨迹的回归就更准。但 m 增大会导致计算量的显著增加，因此应该权衡选择 m 的数值。

(3) τ 的优化。当 m 保持不变时，使预测误差较小的 τ 多出现在 8、16h 附近。这看来与线性回归方法中 $\tau=8$ 、16h 多见的的原因相同：此时负荷记录的“双周期”成分的延时相轨迹的重叠最小，或近似地说使负荷记录延时相轨迹的重叠最小。

4.4 小结

采用二阶多项式回归法进行混沌相空间重构来预测负荷，与线性回归法相比，不仅提高了预测精度，还极大地消除了 12h 的周期性误差增大现象。为了保证在每个点都能正常地预测负荷，本文采用与“参考矢量”（对应当前待预测负荷点）最近的 $(2+2m+m(m-1)/2)$ 个“邻近矢量” $Y(j)$ ，来估计各系数 β_i 。数值测试表明这样的处理是有效的。在这种“邻近矢量”个数条件下，对嵌入相空间的维数 m 、延时时间 τ 的数值测试表明， m 增大对降低预测误差有明显的意义，而 τ 的作用相对要小的多。依据混沌理论与方法（如嵌入定理等）选择的 m 、 τ ，对于本文采用的负荷记录，尽管有一定的指导意义，但一般不给出更优化的取值。这是由于负荷记录不是严格的混沌行为。从本文的数值测试结果看， m 取大，一般会降低预测误差；而 τ 的取值以 8、16h 附近为益（使负荷记录延时相轨迹的重叠最小）。但是， m 取大会显著增大计算量；而 τ 的取值则与计算量的关系不明显。对于具体的预测问题，应该权衡预测精度和计算量，来选取合适的 m 。

第五章 “双周期加混沌”预测中的直接多步法与气象因子使用

上一章中的混沌相空间重构法虽然取得了比较好的效果,但由于电力负荷记录并不是严格混沌的,这使得单纯的混沌相空间重构法^[48]受到一定的限制,为此,文献^[49]中提出了“双周期+混沌”方法,认为“双周期+混沌”拆分更接近电力负荷记录的数学性质。取得了不错的效果,但进一步分析发现,“双周期+混沌”法的预测精度有时仍不能满足要求^[49]。本文将对其进行改进:(1)采用直接多步线性回归预测;并且选择邻近矢量时,依据最接近参考矢量的邻近矢量的个数而不是距离。(2)在选择邻近矢量时,计及各负荷值对应的当天气象因子,实现了考虑气象条件的“双周期+混沌”预测。并且,在邻近矢量的选择和参数的优化上,按照上节提出的方法进行处理。

5.1 气象因子与短期负荷预测

过去的短期负荷预测,一般不考虑天气影响,或者只根据气象条件做简单的修正。然而由于电力负荷与天气有明显的关系,从而考虑天气的短期负荷预测成为近 10 多年来的主流。但是,为了可靠地提高负荷预测的准确率,“考虑气象条件”与“不考虑气象条件”的 2 类的研究都应该深入进行。原因在于:在实际的负荷预测中,使用的天气也是预报信息。一旦天气预报失误,将造成更大的负荷预测误差,例如可以造成 7% 以上的平均误差^[50]。天气预报的难度明显高于电力负荷预测,期望准确天气预报的理想是近期难以实现的。时常出现的情况是,即使不考虑天气影响的负荷预测准确率,也高于天气预报的准确率。

考虑与不考虑天气的 2 类负荷预测的结合,将为实际使用的综合负荷预测提供有益的素材。因此,本文从考虑与不考虑天气 2 方面对原来的“双周期+混沌”法进行改进。

5.2 “双周期+混沌”预测法及其 3 类改进的原理

5.2.1 “双周期+混沌”法要点

该方法根据某负荷记录的数学性质,将原始负荷记录 $P(t)$ 拆成“双周期 $P_D(t)$ ”和“剩余(混沌) $P_R(t)$ ”2 个子序列,即 $P(t) = P_D(t) + P_R(t)$ 。双周期成分 $P_D(t)$ 可以准确预测。混沌子序列 $P_R(t)$ 可以通过“相空间重构的线性回归法”进行预

测。由于双周期成分占据了负荷记录变化的绝大部分，因此混沌成分 $P_R(t)$ 的预测误差，对整个负荷预测的影响被显著缩小。这就是“双周期+混沌”法提高预测精度的原因，并且它还可以消除只采用混沌预测的各方法中 12h 周期性的误差增大现象：它起因自负荷性质，不是来自混沌预测方法。

5.2.2 “双周期+混沌”法 3 类改进的原理

(1) 混沌成分预测中 3 个参数 m 、 τ 和 ε 的优化：由于 $P_R(t)$ 、 $P(t)$ 都不是严格的混沌行为，而是以混沌特性为主，混合了其它性质的复合序列，因此现有混沌理论与方法给出的 m^* 、 τ^* ，一般不是负荷预测的优化参数。通过数值计算进行测试，可以找到合适的参数。

(2) 直接多步预测与“邻近矢量”的个数选择：①在常见的次日 24h 整点负荷预测中，现有的混沌预测法采用的是一步预测：在下一个点的预测中，要依据刚才的预测值。这会造成误差积累，限制了预测精度。本文采用了直接从当前“邻近矢量”回归出连续多个预测点的方法，克服了误差积累，并且有利于消除 12h 周期性的误差增大现象。②仅靠 ε 很难事先确定有多少个邻近矢量在当前预测点的预测中被选取，这隐含 2 方面不足：一是当邻近矢量的数目太少时，会造成预测失败；二是过多的邻近矢量，会造成预测精度下降，原因在于其与“参考矢量”（它对应待预测点）的差异较大。所以，我们采用最靠近“参考矢量”的若干个邻近矢量，很好地解决了这 2 方面的问题。

(3) 选择“邻近矢量”时，把气象因子作为参数考虑进去。这就实现了混沌预测中天气因素的使用。具体说，就是把每个负荷值，搭配当天的若干个气象因子。“参考矢量”除了包括负荷记录及同期气象记录外，还包括待预测日的天气预报。这样选择出的“邻近矢量”，不仅考虑了负荷记录，还考虑的若干天的天气记录。这种气象因子的处理，比采用气象特征量的“积累因子”，更方便、更客观。

5.2.3 直接多步预测、“邻近矢量”个数选择、天气因子使用的方法

此方法要点与上一章类似，所不同的是，在今后 S （多步）个值的预测中， $Y(N_m)$ 、 $Y(j)$ 是不变的；变化的是 $X(T+s)$ ，以及它引起的 $A(s)$ 的变化。这样就消除了原来一步法中的误差积累。在第 $S+1$ 个负荷值预测时，重新形成和计算“参考矢量”与各“邻近矢量”。

若要考虑天气，“参考矢量” $Y(N_m)$ 要增加其中各负荷记录对应的天气记录，以及待预测日的天气预报。这样“负荷记录、天气记录、天气预报”形成一个新

型的“参考矢量”。同样在相空间中的各矢量，也按相同的方式形成。为了方便计算各邻近矢量到参考矢量的距离，各天气因子可以采用标么值。当从“邻近矢量”回归各预测系数时，只考虑负荷值，而不考虑天气因子。这种处理，与考虑天气的相似日预测有某种类比性。

5.3 改进的预测结果示例

采用文献^[49]中的负荷历史记录以及同期的多种气象因子，这是我国北方空间跨度不大的某地区。从第 340 天开始进行连续 3 天离线预测，记 $E_r\%1d$ 、 $E_r\%3d$ 分别为第 1 天、连续 3 天预测相对误差的绝对值的平均值。

5.3.1 原来的“双周期+混沌”法中 3 个参数的数值优化

嵌入空间维数、延时时间和选择邻近矢量时采用的距离数值测试的结果见图 5-1。测试的参数范围 $m=1, 2, \dots, 20$, $\tau=1, 2, \dots, 10$ h。

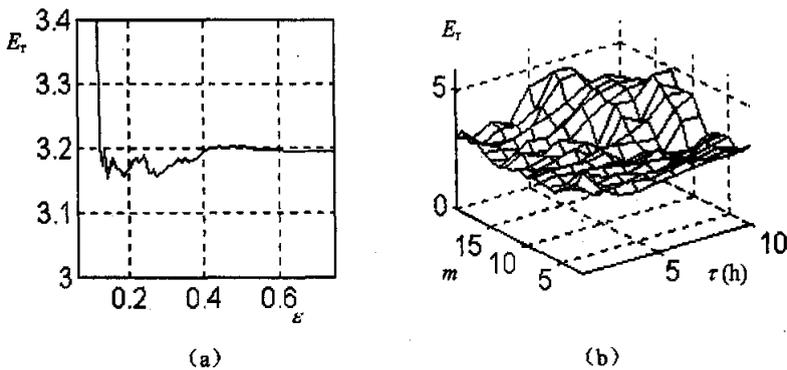


图 5-1 3 天离线预测。(a) ϵ 、 E_r 之间的关系，条件为 $m=9$ 、 $\tau=5$ 。
(b) m 、 τ 与 E_r 之间的关系，条件为 $\epsilon=0.6$ 。

可见，为了保证预测精度：(1) ϵ 不能太小，它取 30% 左右吸引子直径是比较合适的。(2) m 、 τ 不能同时取较大的数值。并且它们之间的搭配，对预测精度有明显的影晌。见图 5-2。

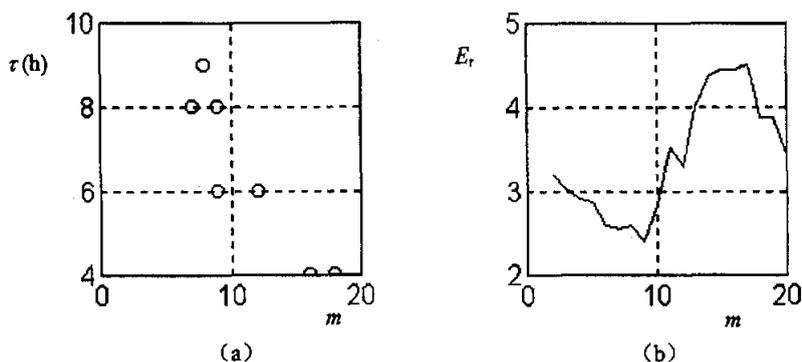


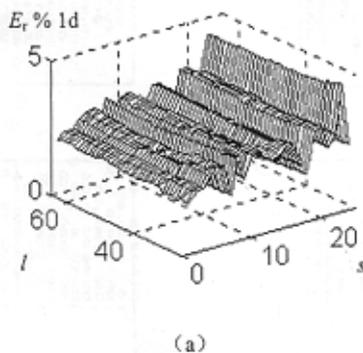
图 5-2 3 天离线预测, (a) 各小圆圈为 E_r 较小的一些 m 、 τ 的搭配。
(b) $\tau=8h$ 时, 预测误差 E_r 随 m 的变化。

上述结果表明, 通过参数优化可以改进“双周期+混沌”法预测精度。但更有意义的是下面 2 类改进。

5.3.2 直接多步回归法

在合适的参数下, 可以很直观地看到直接多步回归法对降低预测误差的好处。

(1) 图 5-3 为在 $m=7$ 、 $\tau=2h$ 条件下, 直接预测 s 步, 采用 $l+m$ 个最靠近“参考矢量”的“邻近矢量”时的 1 天、3 天预测平均绝对值误差 $E_r\% 1d$ 、 $E_r\% 3d$ 。 $l+m$ 个邻近矢量是回归所必须的。测试的参数范围 $l=25, 26, \dots, 65$; $s=1, 2, \dots, 24$ 。在 $l=39$ 、 $s=18$ 时, 24h 预测的平均误差为 1.6492%; 连续 3 天预测的平均误差为 2.6098%。在图 5-3 (b) 中可以看出, 在 24h 负荷预测中, 直接预测 7 步、18 步, 都比原来的一步法优越。



(a)

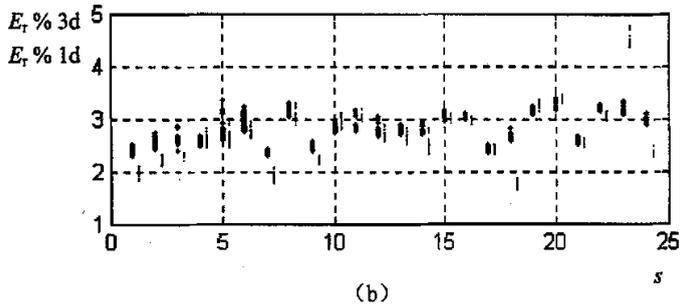


图 5-3 直接多步法中的预测误差与步数 s 、邻近矢量个数 l 的关系。

(a) 1 天误差与 s 、 l 的关系。

(b) 细线、粗线分别为 1 天、3 天误差在 s 上的投影。

(2) 图 5-4 为固定 $l=45$ 、 $s=18$ 时， m 、 τ 变化对预测误差的影响。测试的参数范围 $m=1, 2, \dots, 20$ ， $\tau=1, 2, \dots, 20h$ 。在 $m=10$ 、 $\tau=5h$ 时，24h 预测平均误差 1.6201%；连续 3 天预测平均误差 2.7484%。

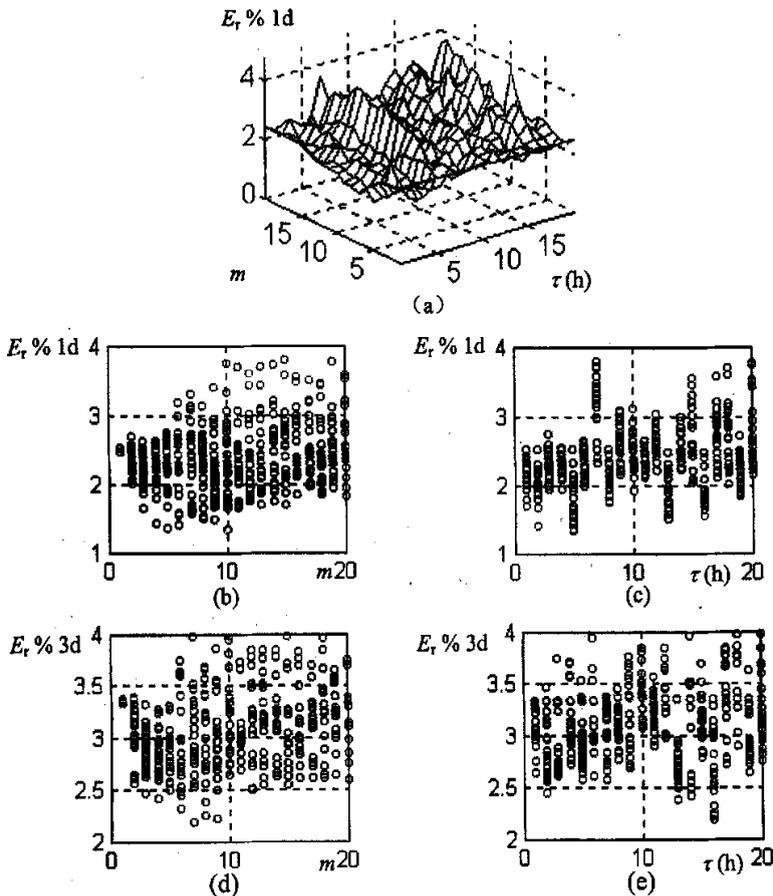


图 5-4 直接多步法中的预测误差与 m 、 τ 的关系。
 (a) 1 天误差与 m 、 τ 的关系 (b) 1 天误差在 m 上的投影
 (c) 1 天误差在 τ 上的投影 (d) 3 天误差在 m 上的投影
 (e) 3 天误差在 τ 上的投影

5.3.3 考虑气象因子的直接多步法

由于预测时间在冬季，故考虑日“最高气温、平均气温、平均海面气压、平均能见度、平均风速、最大持续风速”6 种气象因子（参数）。它们是从多种气象因子中筛选出来的与负荷相关系数较大的 6 个气象参数的日记录。这样，每个负荷值就有当天 6 个气象因子与其搭配。在选择邻近矢量时，共有连续 $(m \times \tau)$ / 24 天的气象因子记录被考虑到了。

在合适的参数下，可以很直观地看到引入气象因子对降低预测误差的好处。

(1) 图 5-5 为固定 $m=5$ 、 $\tau=13h$ 时， l 、 s 变化对预测误差的影响（邻近矢量有 $l+m$ 个）。测试的参数范围 $l=1, 2, \dots, 40$ ； $s=1, 2, \dots, 24$ 。在 $l=19$ 、 $s=18$ 时，24h 预测的平均误差为 1.3770%；连续 3 天预测的平均误差为 2.3426%。

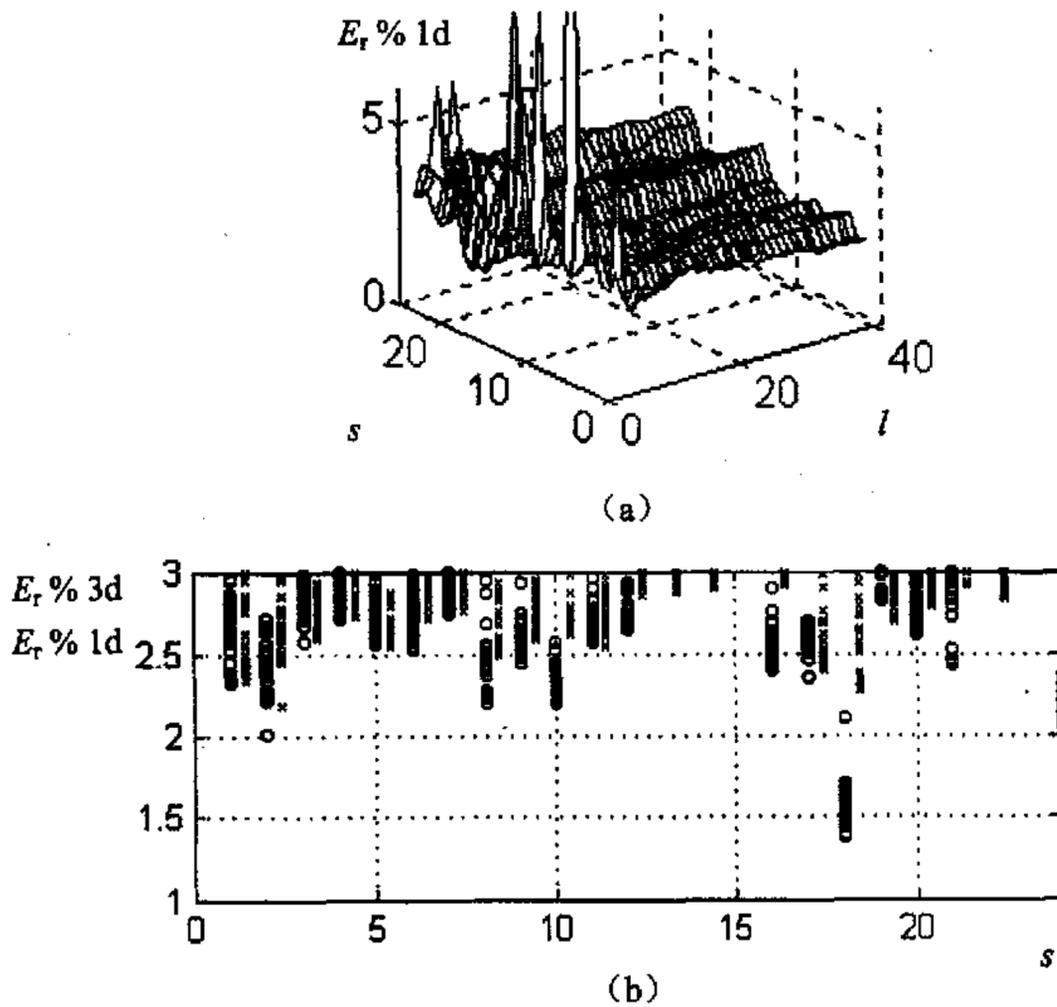


图 5-5 考虑 6 种气象因子的直接多步法的预测误差与步数 s 、邻近矢量个数 l 的关系。

(a) 1 天误差与 s 、 l 的关系。
 (b) 圆圈、 \times 号代表 1 天、3 天误差在 s 上的投影。

(2) 图 5-6 为固定 $l=20$ 、 $s=18$ 时, m 、 τ 变化对预测误差的影响。测试的参数范围 $m=1, 2, \dots, 20$, $\tau=1, 2, \dots, 20\text{h}$ 。在 $m=5$ 、 $\tau=13\text{h}$ 时, 24h 预测的平均误差为 1.3772%; 连续 3 天预测的平均误差为 2.3131%。

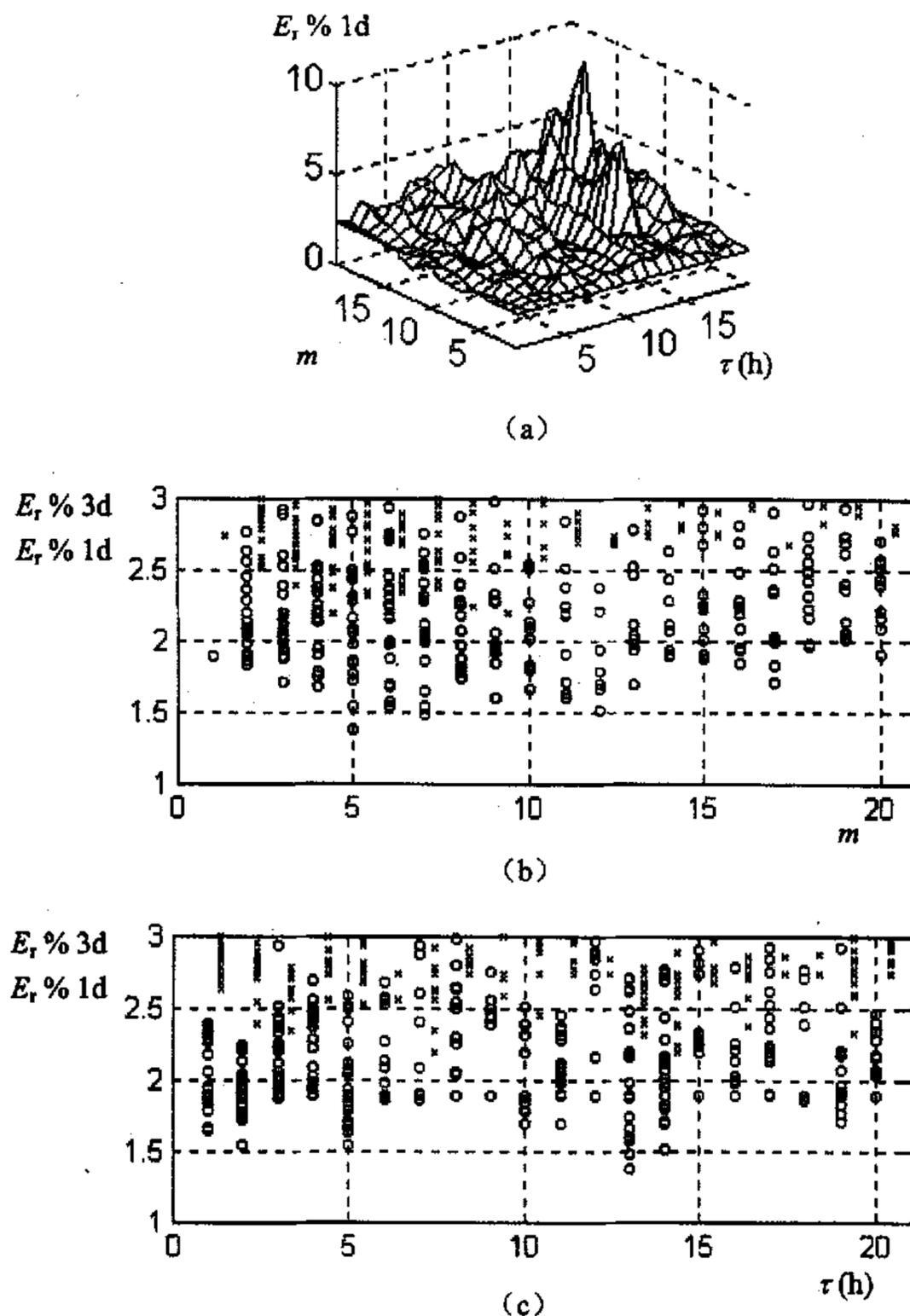


图 5-6 考虑 6 种气象因子的直接多步法中的预测误差与 m 、 τ 的关系。

(a) 1 天误差与 m 、 τ 的关系。

(b) 圆圈、 \times 号代表 1 天、3 天误差在 m 上的投影。

(c) 圆圈、 \times 号代表 1 天、3 天误差在 τ 上的投影。

5.4 数值测试结果的整理

(1) 在现有的混沌相空间线性回归法 (一步法) 中, 通过采用优化的 3 个参数, 可以提高预测精度。其中, m 、 τ 不能同时取较大的数值, 并且其搭配,

对预测精度有明显的影响。 ε 可取吸引子直径的 30%左右。

(2) 直接多步线性回归法比较明显地提高了预测精度, 它用最靠近参考矢量的若干个邻近矢量来进行回归预测。选择合适的直接预测步数 s 、延时时间 τ 对降低预测误差的作用比较明显。另外 2 个参数, 即邻近矢量个数 l 、嵌入空间维数 m 的影响要小。但各参数之间的搭配也较重要。

(3) 考虑气象因子后的直接多步法, 可以进一步降低预测误差。4 个参数的影响类似于 (2)。这里的关键是根据不同的季节选择合适数目的气象因子。

(4) 无论在哪种情况下, 混沌研究给出的 m^* 、 τ^* , 往往都不是优化的数值。这来自于负荷记录不是严格的混沌行为。两类直接多步法中 $\tau=13\text{h}$ 比较合适的原因, 是与该时混沌成分的自相关系数取极小值附近有关。

(5) 两类直接多步法中邻近矢量个数为 $l+m$ 个。当 l 取值比较大时, 误差随其变化很小。一般可以选择 l 在 20~40 之间。直接预测步数 $s=18$ 比较合适, 与该地冬季日峰值负荷经常出现在下午 6 点有关; 其后负荷记录呈明显下降趋势, 与上午 8 点左右到下午 6 点之间的负荷记录性质明显不同。实际上, 步数 $s \approx 7$ 、8 为另一个次优值, 它与经常出现的上午 7、8 点负荷记录从下降趋势变成上升趋势有关。

5.5 小结

某负荷记录是“双周期+混沌”的, 可以通过“双周期+混沌”法来进行短期负荷预测。为提高其预测精度, 本文的 3 类改进为: (1) 通过数值测试, 提供了混沌相空间重构的线性回归法的 3 个参数的优化值。如 $(m, \tau) = (7, 8\text{h})$ 、 $(9, 8\text{h})$ 、 $(8, 9\text{h})$ 等; ε 可取 30%左右吸引子直径。(2) 采用直接多步线性回归法, 并选取离“参考矢量”最近的若干个“邻近矢量”进行负荷预测。选择合适的直接预测步数和延时时间, 可以有效提高预测精度。(3) 再在直接多步法中选择邻近矢量时引入各负荷记录对应的当天若干气象因子, 在合适的直接预测步数和延时时间下, 可以进一步降低预测误差。两类直接多步法中 $\tau=13\text{h}$ 、直接预测步数 $s=18$ 比较合适, 分别与负荷记录自相关系数取极小值、负荷记录取日最大值的时间有关。其它 3 个参数的影响较小。混沌研究给出的 m^* 、 τ^* , 一般不是优化的参数。

第六章 总结与展望

电力系统负荷预测是电力部门一项重要又常规的工作。预测结果的好坏不仅直接关系到电力企业的经济利益,对用户来说也是很重要的。正因为如此,电力研究者们才会不断地对电力负荷预测进行研究,追求精益求精的效果。本文正是基于这一出发点,期望通过自己的工作,在负荷预测研究的工作中取得一些有益的结果。研究负荷预测的方法有许多种,这里就存在着一个方法论的问题,即采用什么方法进行研究。这是使研究独辟蹊径,事半功倍的关键所在。本文在阅读了大量参考文献的基础上,紧跟负荷预测研究的最新进展,采用属于人工智能领域的模糊聚类理论,对我国北方某地的负荷与气象数据进行了初步的研究,采用非线性动力学领域里的混沌理论,提出了几种负荷预测的新方法。总的说来,本文取得了以下一些初步结果:

1. 负荷预测历史数据的处理方面:应用模糊聚类理论,对负荷数据与气象数据组成的预测样本进行了详尽的分析。该理论是数据挖掘的一个分支,它适应了现代负荷预测的数据量和信息量急剧增长的形势,借助于计算机,能够实时而准确地发掘出数据内部的规律,从而满足预测精度和实时性的要求。模糊聚类方法也有多种,本文使用的是传统和常用的模糊等价传递闭包法,此法的关键之处有两个:一个是衡量样本间相似度的模糊隶属度函数 $r(i,j)$;另一个是聚类截水平 λ 。在模糊隶属度函数方面,针对实际样本,分别运用相关系数和欧氏距离法进行了聚类,并提出用相关系数与欧氏距离的加权和作为模糊隶属度函数进行聚类。最后对照原始数据,对聚类结果做了解释和分析,论证了本文构造的方法具有更合理的聚类效果。在确定聚类截水平 λ 方面,根据电力系统负荷所具有的规律和特性,认为属于同一类中的连续天数不超过 5 天,从而通过程序自动得出聚类水平 λ 。

2. 在负荷预测方面:在混沌重构相空间局域法中,首先采用多元二阶多项式回归法预测出未来时刻的负荷值。该方法与线性回归法相比,具有更高的预测精度,并且显著地抑制了 12h 周期性的误差增大现象。另外,在寻找嵌入相空间中与参考矢量邻近的矢量时,以指定邻近矢量的个数来代替与参考矢量的距离,不仅保证了在任何情况下都能找到足够的“邻近矢量”来进行负荷预测,还可以提高预测精度。其次,采用直接多步线性回归法进行预测,即从当前值直接回归预测出多个点的负荷,消除了原来一步法中的误差积累。在考虑天气因子的直接多步回归时,本文在“参考矢量”中增加各负荷记录对应的天气记录,以及待预测日的天气预报。这样“负荷记录、天气记录、天气预报”形成一个新型的“参考矢量”,从而实现了混沌时间序列中气象因子的使用。

通过论文期间的研究和探索,我体会到,聚类分析、模式识别以及混沌预测在电力负荷预测方面的应用具有强大的生命力。如何更好地处理预测前期数据使之更符合需要,如何将更多的预测技术融合起来都是值得进一步研究的问题。例如坏数据的剔除、丢失数据的模拟恢复,神经网络、灰色系统、聚类分析、模式识别等预测技术的结合等,这些都有待进一步探索。

负荷预测技术经历了开始、应用和发展各阶段直到今天,已经形成了现代负荷预测的分析思路、现代负荷预测方法和技术等较为成熟的负荷预测经验和现代化的负荷预测发展方向,也就是说,形成了现代负荷预测的特征。这些特征包括:预测定量化特征,面对数据分析特征,定量与定性预测结合特征,重视预测效益和质量特征,动态特征,综合预测特征,应用非线性预测的特征以及向智能化技术发展的特征。可以预见,今后的预测技术都将呈现出这些特征。这是指导我们进行负荷预测研究的指南针。

参考文献

- [1] 陈守煜, 模糊识别决策与聚类理论模型, 模糊系统与数学, 1991, 9(2)
- [2] 陈觉培, 姜立芝, 模糊数量化差别模型及其应用, 应用数学, 1988(4)
- [3] 李洪兴, 汪培庄, 基于摄动的 Fuzzy 聚类方法, 数学季刊, 1988, 3(1): 9~19
- [4] 沙丹, 绝对距离 $D(R, T)$ 最小的模糊等价矩阵, 模糊系统与数学, 1994(1): 47~54
- [5] 罗承忠, 模糊集引论(上册), 北京: 北京师范大学出版社, 1989
- [6] 李相镐, 李洪兴, 陈世权, 汪培庄, 模糊聚类分析及其应用, 贵阳: 贵州科学技术出版社, 1994
- [7] 朱剑英, 应用模糊聚类法应注意的若干关键问题, 模糊系统与数学, 1987, 1(1)
- [8] 何清, 模糊聚类分析理论与应用研究进展, 模糊系统与数学, 1998, 12(2): 89~94
- [9] Anderberg M R. Cluster Analysis for Applications. New York: Academic Press, 1973
- [10] Var Ryzin J. Classification and Clustering. New York: Academic Press, 1977
- [11] Dubes R C, Jain A K. Algorithms for Clustering Data. Englewood Cliffs, NJ: Prentice Hall, 1988
- [12] Wu Youshou, Ding Xiaoqing. A new clustering method for Chinese character recognition system using artificial neural networks. Chinese J. of Electronics, 1993, 2(3): 1~8
- [13] Huang Z, Kuth A. A combined selforganizing feature map and multilayer perception for isolated word recognition. IEEE Signal Processing, 1992, 40(11): 2651~2657
- [14] Stewart C, Lu Y C, Larson V. A neural clustering approach for high resolution radar target classification. Pattern Recognition, 1994, 27(4): 503~513
- [15] Coleman G B, Andrews H C. Image segmentation by clustering. Proc IEEE, 1979, 5(67): 773~785
- [16] 刘健庄, 基于二维直方图的图像模糊聚类分割方法, 电子学报, 1992, 20(9): 40~46
- [17] Trivedi M M, Bezdek J C. Low-level segmentation of aerial image with fuzzy clustering. IEEE SMC, 1986, 16(4): 589~598
- [18] Porter R, Canagarajah N. A robust automatic clustering scheme for image segmentation using wavelets. IEEE Image Processing, 1996, 5(4): 662~665
- [19] Chaudhuri B B, Sarkar N. Texture segmentation using fractal dimension. IEEE PAMI, 1995, 17(1): 72~77

- [20] Chen S W, Chen C F, Chen M S, et al. Neurfuzzy classification for segmentation of remotely sensed images. *IEEE Signal Processing*, 1997, 45(11): 2639~2654
- [21] Dave R N. Fuzzy shell clustering and applications to circle detection in digital image. *Inter J General System*, 1990, 16(4): 343~355
- [22] Dave R N. Generalized fuzzy c-shells clustering and detection of circular and elliptical boundaries. *Pattern Recognition*, 1992, 25(7): 713~721
- [23] Krishnapuram R, Frigui H, Nasraoni O. Fuzzy and possibilistic shell clustering algorithms and their application to boundary detection and surface approximation Part I, II. *IEEE Fuzzy System*, 1995, 3(1): 29~60
- [24] Shih F Y, Moh Jenlong, Chang Fuchun. A new artbased neural architecture for pattern classification and image enhancement without prior knowledge. *Pattern Recognition*, 1992, 25(5): 533~542
- [25] Lai Weichi. A VLSI neural processor for image data compression using selforganization networks. *IEEE Neural Networks*, 1993, 3(3): 506~517
- [26] 刘耀年, 伏祥运, 张文生, 曾令全, 基于模糊识别与模糊聚类理论的短期负荷预测, *电工技术学报*, 2002, 17(5)
- [27] 陈耀武, 汪乐宇, 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(4)
- [28] 吴娟, 费维刚, 黄伟, 模糊聚类预测法在电网负荷预测中的应用, *河北电力技术*, 1998, 17(2)
- [29] 康重庆, 程旭, 夏清, 一种规范化的处理相关因素的短期负荷预测新策略, *电力系统自动化*, 1999, 23(18)
- [30] 夏昌浩, 张毓哲, 电力系统负荷预报方法综述, *电力学报*, 2001, 16(2)
- [31] 王砚, 张少华, 一种应用时间序列技术的短期电力负荷预测模型, *上海大学学报(自然科学版)*, 2002, 8(2)
- [32] 吉国力, 程军, 米红, 应用混沌相空间线性回归模型研究短期负荷预报, *系统工程理论与实践*, 2001, (6)
- [33] 李天云, 刘自发, 电力系统负荷的混沌特性及预测, *中国电机工程学报*, 2000, 20(11)
- [34] 康重庆, 夏清, 刘梅, 相年德, 应用于负荷预测中的回归分析的特殊问题, *电力系统自动化*, 1998, 22(10)
- [35] 高山, 单渊达, 神经网络短期负荷预测输入变量选择新方法, *电力系统自动化*, 2001, 25(22)
- [36] 夏昌浩, 胡翔勇, 刘涤尘, 神经网络短期负荷预报模糊化改进, *电力学报*, 2001, 16(1)
- [37] 陈耀武, 汪乐宇, 龙洪玉, 基于组合式神经网络给的短期电力负荷预测模型, *中国电机工程学报*, 2001, 21(4)
- [38] 吴斌, 陈章潮, 包海龙基于人工神经元网络及模糊算法的空间负荷预测, *电网技术*. 1999, 23(11)

- [39] U.Kaymak, R.Babuska, M.Setnes, H.B.Verbruggen, and H.R.van Nauta Lemke. Methods for simplification of fuzzy models. In D. Ruan, editor, *Intelligent Hybrid Systems*, 1997. 91–108.
- [40] 吕金虎, 陆君安, 陈士华, 混沌时间序列分析及其应用, 武汉: 武汉大学出版社, 2002
- [41] 向小东, 基于神经网络与混沌理论的非线性时间序列预测研究: [博士学位论文], 成都: 西南交通大学, 2001
- [42] 刘晨晖, 电力系统负荷预报理论与方法, 哈尔滨: 哈尔滨工业大学出版社, 1987
- [43] Jae-Gyun Choi, Jong-Keun Park, Kwang-Ho Kim, Jae-Cheol Kim. Daily peak load forecasting system using a chaotic time series. *Proceedings of the International Conference on Intelligent Systems Applications to Power Systems, ISAP Jan 28-Feb 2 1996, Orlando, FL, USA*, pp: 283-287
- [44] H. G. Arango, A. C. Z. de Souza, G. Lambert – Torres, A. P. A. da Silva. Difference between regular and deterministic chaos processes based on time analysis of load: an example using CEMIG data [J]. *Electric Power Systems Research*, 2000, 56(1):35-41
- [45] Parker T S, Chua L O. *Practical numerical algorithms for chaotic systems* [M]. New York: Springer-Verlag, World Publishing Corp., 1989
- [46] 张小蒂, 应用回归分析, 杭州: 浙江大学出版社, 1991
- [47] H. S. Kim, R. Eykholt, J. D. Salas. Nonlinear dynamics, delay times, and embedding windows [J]. *Physic D*, 1999, 127(1):48-60
- [48] 李天云, 刘自发, 电力系统负荷的混沌特性及预测, *中国电机工程学报*, 2000, 20(11): 36-40
- [49] 杨正瓴, 林孔元, 余贻鑫, 短期负荷预报的“双周期加混沌”法中的子模型优选的理论探讨, *电网技术*, 2003, 27(5): 33-36
- [50] 胡子珩, 陈晓平, 刘顺桂等, 深圳电网自动运行的短期负荷预测系统, *电网技术*, 2003, 27(5): 21-25

作者发表论文及科研情况

学术论文:

1. 刘正光, 喻远飞, 吴冰, “利用云导风矢量的台风中心自动定位”, 气象学报, 2003, 61 (5): pp636~640
2. 杨正瓴, 喻远飞, 林孔元, “负荷预报的混沌相空间二阶多项式回归法”, 中国电力 (已录用)

科研经历:

《智能电池冲放电检测系统开发》上位机软件部分, 2002.12。

致 谢

本文的研究工作是在导师刘正光教授和学科组杨正瓴副教授悉心指导下完成的，刘老师严谨细致的工作作风使我终生受益，杨老师敏捷开阔的思维扩展了我的视野，在此向两位老师表示崇高的敬意和由衷的感谢！

在研究生期间，学科组各位老师：林孔元教授，王萍教授，路志英副教授，以及张军和薛俊涛老师在学习和生活上也给予了帮助和支持，在此一并表示感谢！另外，在论文的写作和校对过程中，硕士研究生田勇同学也给予了极大的帮助和支持，诚致谢意。

最后，感谢我的父母和家人两年多来给予我的关怀和鼓励。