

|                        |
|------------------------|
| 西安交通大学                 |
| 硕士学位论文                 |
| 变电设备的绝缘在线监测数据处理及状态模式分类 |
| 姓名：李辉                  |
| 申请学位级别：硕士              |
| 专业：电气工程                |
| 指导教师：严璋                |
| 2003. 3. 1             |

## 变电设备的绝缘在线监测数据处理 及状态模式分类

### 摘 要

随着电力系统电压等级的提高、设备容量的增大,对系统安全性、稳定性的要求也越来越高。研究电力设备在线监测技术是及早发现绝缘缺陷、确定设备绝缘状态、提高系统运行可靠性的重要手段之一。要实现设备的在线监测除了要完善监测系统外,还需对监测数据进行综合分析处理,挖掘数据蕴含的信息,丰富诊断专家系统的知识库,以提高诊断的可靠性。为了从这些监测的数据中获得设备的状态信息,论文采用了综合分析的方法对在线数据进行处理,其主要内容如下:

针对在线数据采集和传输过程中易受干扰的问题,论文分别对数据在时域和频域进行分析、处理,并提出了一种快速滤波算法——中值滤波快速算法。

为了及时、准确地发现设备潜伏性故障或绝缘缺陷,必须对在线监测数据进行预测。论文采用时间序列的思想对在线数据进行实时递推预测。

为了对被监测设备的状态作一个整体的评估,应通过对在线数据进行分析实现设备状态的分类。论文采用支撑向量机的思想设计了一个基于径向基函数的分类器,该分类器通过支撑向量机学习网络实现了在线数据的融合,达到对设备状态分类的目的。

关键词: 在线监测, 中值滤波快速算法, 时间序列, 支撑向量机, 径向基函数;  
~~论文类型: 应用研究~~

## **Data Processing of On-line Insulation Monitoring and Condition Pattern Classification for Power Equipments**

### **Abstract**

With the increase of rated voltage and capacity of power equipments, higher reliability of power system is expected. The on-line insulation monitoring is one of important methods for the fault detection of the equipments and the reliability of the power system. In order to realize the on-line monitoring, it is necessary to improve the monitoring system and analyze the detected data synthetically. So the information contained in the data should be mined, the diagnostic system of knowledge base be enriched, so as to improve accuracy of diagnosis. To realize real-time condition for substation equipments, several methods are developed for analyzing and processing on-line monitoring data synthetically. The main content of this paper is as following.

On-line monitoring data is easily to be interfered in the course of collection and transmission. In order to solve this problem, data filter analyzing and processing methods in time domain and frequency domain are introduced, and a new filtering algorithm, fast median filtering algorithm, is presented

In order to find potential fault or insulation defect of power equipments timely and accurately, on-line monitoring data should be forecasted. The method of time series is adopted in the paper, it is helpful to forecast on-line monitoring data tendency.

In order to evaluate monitored power equipment integrally, condition classification is realized based on the on-line monitoring data analysis. A classifier based on Radial Basis Function (RBF) is designed by Support Vector Machine (SVM). Based on the classifier, on-line monitoring data fusion is realized, and the aim of power equipment condition classification is achieved.

**Key Words:** On-line Monitoring   Fast Median Filtering Algorithm   Time Series  
Support Vector Machine   Radial Basis Function

492811

附件 1:

## 独创性声明

秉承学校严谨的学风与优良的科学道德, 本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知, 除了文中特别加以标注和致谢的地方外, 论文中不包含其他人已经发表或撰写过的研究成果, 不包含本人或他人已申请学位或其他用途使用过的成果。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致谢。

申请学位论文与资料若有不实之处, 本人承担一切相关责任。

论文作者签名: 李辉 日期: 03-3-21

## 保护知识产权声明

本人完全了解西安交通大学有关保护知识产权的规定, 即: 研究生在校攻读学位期间论文工作的知识产权单位属西安交通大学。本人保证毕业离校后, 发表论文或使用论文工作成果时署各单位仍然为西安交通大学。学校有权保留送交论文的复印件, 允许论文被查阅和借阅; 学校可以公布论文的全部或部分内容, 可以采用影印、缩印或其他复制手段保存论文。

(保密的论文在解密后应遵守此规定)

论文作者签名: 李辉 导师签名: 张辉 日期: 03-3-21

## 1 绪论

本章叙述了变电设备绝缘在线监测技术及其数据处理、模式识别的发展历程、基本概念以及目前存在的一些问题,最后介绍了本文的主要研究工作。

### 1.1 引言

随着经济的发展,电力系统对供电的可靠性和经济性提出了愈来愈高的要求,高压输变电设备的安全运行已成为影响电力系统安全、稳定、经济运行的重要因素。由于高压电力设备的绝缘系统在运行过程中长期处于电、机械、化学等应力作用下,不可避免地将逐渐劣化,在系统的薄弱环节会导致绝缘缺陷的出现,如未能及时发现并采取适当的修复措施,缺陷将不断发展,有可能引发设备绝缘击穿事故,造成多方面的损失。自六十年代以来,国内的某些大型电力网络出现了一些大的停电事故,给国民经济造成了巨大的损失<sup>[1]</sup>。“八五”期间,我国主要电网中由于设备故障而直接引发的电网事故约占事故总量的 26.3%<sup>[2]</sup>,可见提高设备的运行可靠性是保证电力系统安全运行的关键之一。及时了解电力设备的工作情况,对电力系统的故障作出先兆判断是十分重要的。尤其是在电力设备的运行过程中通过必要的监测和试验等手段有效地判断其运行状态,将会对减少事故发生、提高设备运行可靠性具有重要的意义。

### 1.2 变电设备检测技术的发展概况

我国开展电力设备绝缘在线监测技术的开发应用已有十几年了,此项工作对提高电力设备的运行维护水平,及时发现事故隐患,减少停电事故的发生起到了积极作用<sup>[3]</sup>。

我国从 50 年代开始,一直根据电力设备预防性试验规程对电力设备进行定期的停电试验、检修和维护。这些试验是过去长期的运行经验和试验研究的积累,对发现设备缺陷、减少事故的发生发挥了很重要的作用,但随着电压等级的提高、设备容量的增大也暴露出一些问题:

- 1) 预防性试验周期规定过死,有时不能及时发现设备绝缘的劣化。
- 2) 高压电力设备的试验电压低,难以真实反映运行电压下的设备绝缘状况。
- 3) “到期必修”没有充分考虑到设备的实际状态,超量维修不少,以致出现不必要的停电维修,造成大量人力、物力的浪费<sup>[4]</sup>。

4) 难以适应无人值班变电站自动化的要求。

随着电子技术的进步和传感器技术、光纤技术、计算机技术、信息处理技术等的发展以及其向各领域的渗透,使电力设备的在线监测成为了可能,且与传统的预防性试验相比在线监测具有很大优势:

1) 在线监测是设备在运行电压下进行的,能更真实地反映设备的绝缘状态。

2) 在线监测并不需要停电,有助于提高供电的可靠性。

3) 以设备的状态为依据,决定是否采用相应的维修策略,在很大程度上减少了维修费用,提高了经济效益。

4) 对变电设备进行在线监测,有助于加强对设备的管理和自动化监控。

我国对在线监测的重要性认识比较早,60年代时就提出过不少带电试验的方法,但行之有效的不多。近年来,国内外对变电设备绝缘状况的在线监测进行了大量的研究,取得了很大进展,也有不少形成了产品和系统。其发展大体经历了以下三个阶段<sup>[3]</sup>:

1) 带电测试阶段。主要对电气设备的某些绝缘参数进行直接测量(如泄漏电流等)。

2) 从80年代以来,采用各种专用带电测量仪器,使在线监测从传统的模拟测量转变到数字测量。

3) 从90年代开始,出现以数字信号采集和微处理技术为核心的微机多功能绝缘在线监测系统<sup>[5]</sup>。

目前,在国外在线监测的研究方向已经进一步发展到以可靠性为中心的状态维修,不但要保证设备正常运行,而且还要考虑经济效益、投入产出比等。但在线监测系统也还存在一些问题和不足之处:

1) 在线数据在采集和传输过程中易受周围环境的干扰,可能导致数据的波动和数据失真。

2) 在线监测系统通常以设备一个或几个特征参数为目标,构成一个监测诊断系统,因此对很难设备状态作出完整的评估。

3) 目前在线数据的诊断还没有一个完整的标准。

### 1.3 绝缘监测数据的处理及设备状态的模式分类

在线监测是以变电设备为监测对象,采用更高灵敏度的传感器以采集运行中设备绝缘状态的信息——主要是数据,然后对其进行加工处理获取知

识，并可参照过去积累的大量知识及经验(包括预防性试验规程等)加以识别，形成较为完善的系统，因此，在线监测系统中数据的处理、模式识别显得极为重要。

电力设备维修方式的发展使得用于模式分类的数据不再只是少量的单次数据，而是包括离线试验数据和大量连续监测的在线数据，这些大量数据背后隐藏着许多重要的信息，因此如何对数据进行更高层次的分析，以便更好地利用这些数据是一个重要的目标。虽然目前电力系统中广泛应用的数据库管理系统(MIS)可以高效地实现数据的录入、查询、统计等功能，但至今仍无法直接发现数据中存在的各种关系和规则，无法很好的根据现有的数据得知电力设备的绝缘状态并作出诊断决策<sup>[6]</sup>，因此，对进一步挖掘数据的信息十分必要。

数据是指一个有关事实 $F$ 的集合(如电力设备状态维修数据库中在线监测数据)，它是用来描述事物某方面的信息，是我们进一步对设备分析、诊断的原材料。经过数据挖掘出提取出诊断的模式，但其内容必须是新颖的、有效的，至少对监测系统来说应该如此。模式是否真实有效可以通过两个途径来衡量：其一是从得到的数据中，通过优化处理是否能反映设备真实状况，来判断该模式的真实程度；其二是通过其内部所包含的知识，即通过对比发现的模式与已有的模式的关系来判断。

综上所述，数据的处理、模式分类是一个多步骤的、循序渐进的处理过程，它们之间相互联系、相互影响，其图形如图 1-1 所示<sup>[6]</sup>。

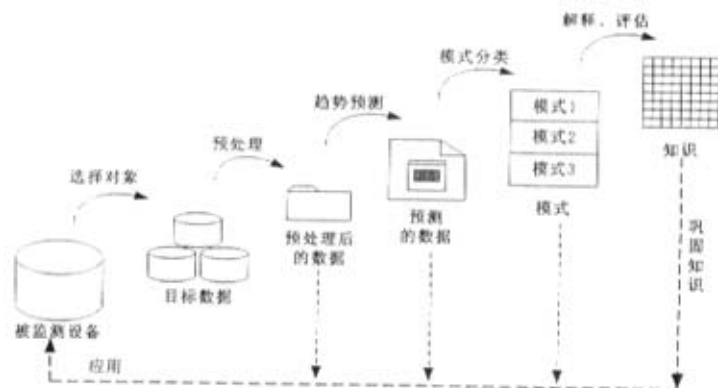


图 1-1 数据处理及模式分类过程示意图

在线数据的信息挖掘是以被监测的设备为对象，以设备的绝缘特性为依



据选择合理的参数,通过高灵敏度的传感器获取模拟信号,以数据采集卡转变为数字信号,以后的工作都是以采集的数字信号为基础进行。

由于被监测设备周围环境较复杂,且数字信号在传输过程中还可能存在着数据波动和数据失真的现象,因此数据采集和传输都难以保证其真实性,需对其进行合适的处理。目前最为常用的处理方法主要是滤波方法,虽然硬件和软件都可以实现,但基于经济效益的考虑,目前更多选择软件滤波,其中较为常用的有快速傅立叶变换、53H 算法<sup>[7]</sup>、滑动平均算法<sup>[5]</sup>及中值滤波快速算法<sup>[8]</sup>。

而进行在线监测主要的目的就是及时、准确地发现设备存在的隐患;应根据数据所反映信息为设备作一个“健康”报告。已获取的数据反映了设备当前信息,如果要对设备未来的趋势作出评估,还必须获取被监测对象的大量未知信息。通常获取未知信息的方法主要是数据预测,常用的预测方法如最小二乘法、时间序列模型(ARMA)<sup>[9]</sup>、灰色理论模型<sup>[10]</sup>、支撑向量机<sup>[10,11]</sup>等,本文将采用一种基于时间序列的最小二乘递推算法<sup>[9]</sup>进行数据预测。

在完成了数据预测步骤后,需要对数据进行整理、分析,以挖掘数据所蕴涵的信息。数据预测只是数据优化处理实现途径之一,其最终目标应是根据数据对设备状态进行模式分类。本文将设计一个基于径向基函数的支撑向量分类器<sup>[13,14]</sup>,该分类器能够有效地对数据进行模式识别。

由于各个步骤间存在着循序渐进关系,最后需要做的工作就是对数据进行综合分析即知识的积累和应用。在完成了对大量数据优化处理后,可以通过一个知识库的形式把知识积累下来,通过使用知识库来实现对被监测的设备的评估和解释。

#### 1.4 论文的研究工作

本文提出了以优化处理的方式对变电设备在线监测数据进行综合分析。优化处理的对象是在线数据,因此论文首先介绍了在线数据采集和数据的预处理,通过分析在线数据的特点同时结合变电站自动化监测的要求,提出了一种较为实用的处理算法—中值滤波快速算法;然后利用时序分析法,将预处理后的数据通过时序建模提取征兆,以变压器套管泄漏电流为例,采用最小二乘递推算法建立了相应的预测模型,并实现了对在线数据的预测;最后利用了支撑向量机可进行学习和模式识别的思想,以110kV变电站的金属氧化物避雷器全电流和阻性电流为例设计了一个基于径向基函数的支撑向量



分类器，达到了对被监测设备状态分类的目的。论文的主要结构如图 1-2 所示。

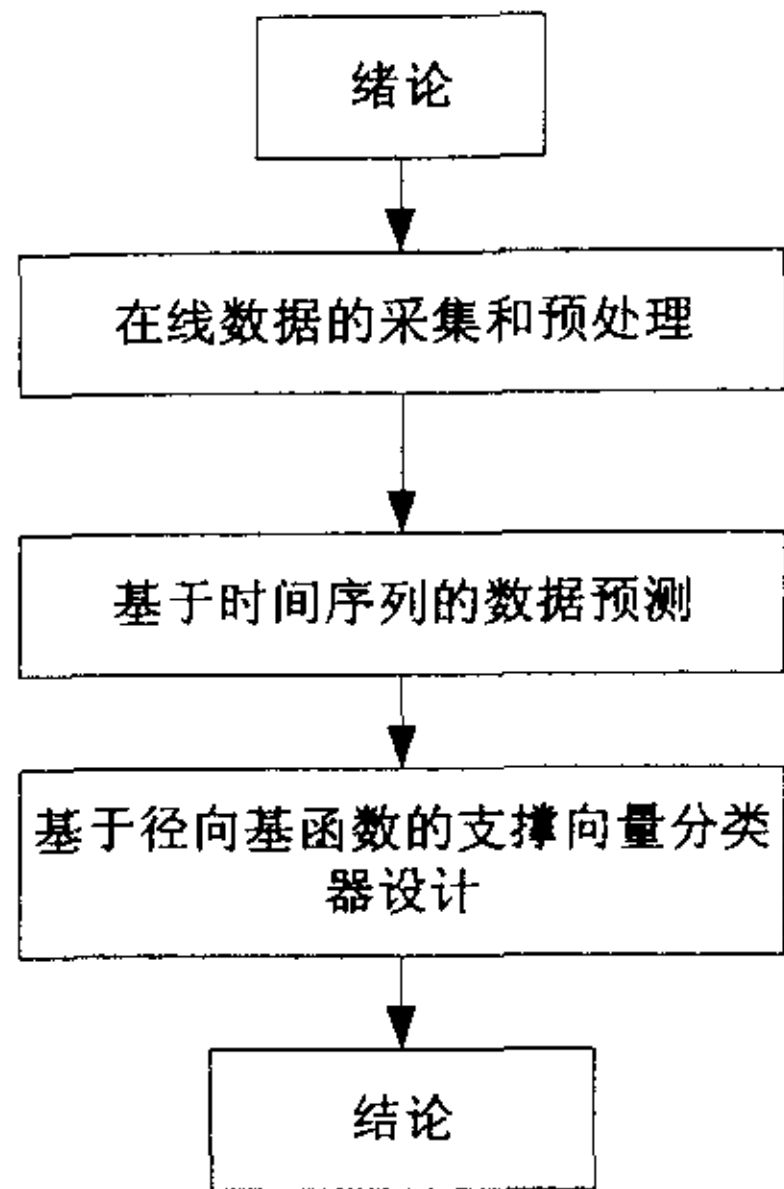


图 1-2 论文结构图

## 2 在线数据的采集和预处理

在线监测系统所获取的大量数据是有序的数据序列，它们反映了运行电压下设备的绝缘状况。要使获取的数据真实、有效，需从数据采集和预处理两方面解决，以提高其抗干扰的性能。因此本章主要针对这两方面进行研究，为下一步数据预测奠定基础。

### 2.1 数据的采集

由于大多数系统的输出信号是连续信号，要进行数据处理必须对连续信号进行离散化采样。而如何确定合适的采样时间间隔 $\Delta$ 与样本长度 $L$ ，以确保获取的信息真实、准确，是数据采集阶段必须解决的问题。

#### 2.1.1 采样时间间隔的确定

若假设被监测对象的信号为连续信号 $x(t)$ ，采样时间间隔 $\Delta$ 确定，连续信号 $x(t)$ 经离散采样后变成了 $\{x_i\}$ ，且两者的频谱发生了变化，在频域内可能出现高频分量和低频的分量即频率混迭，这种现象简称混频。

由傅立叶变换可知：对于一个连续信号 $x(t)$ 可分解成不同频率的谐波信号的迭加<sup>[15]</sup>。对于连续信号 $x(t)$ 我们感兴趣的只是某一频带范围，设为 $-\omega_{\max} \sim \omega_{\max}$ ，其中 $\omega_{\max}$ 为该频带中最高的频率。以某一连续信号 $x(t)$ 为例，按不同的采样间隔 $\Delta$ 采样后，对应的时域和频域图形如图 2-1 所示<sup>[9]</sup>。

其中图 2-1(a)是信号 $x(t)$ 的时域图形， $X(\omega)$ 为信号频谱图。图 2-1(b)是采样频率 $\omega_s > 2\omega_{\max}$ 时，时域和频域图形。图 2-1(c)为采样频率 $\omega_s < 2\omega_{\max}$ 时时域、频域图形。因此由图 2-1 可知，信号不发生混频的条件为

$$\omega_s \geq 2\omega_{\max} \quad (2-1)$$

由于采样频率 $f_s = 1/\Delta = \omega_s / 2\pi$ ，因此采样频率 $f_s$ 与该段频带中最高频率 $f_{\max}$ 对应的关系为

$$f_s \geq 2f_{\max} \quad (2-2)$$

$$\text{或} \quad \Delta \leq \frac{1}{2f_{\max}} \quad (2-3)$$

公式(2-2)和(2-3)为著名的 Shannon 采样定理，为了正确地获取连续信号中各种频率成分的信息，在最高频率谐波的一个周期内至少应采样两次。实践中可首先对连续信号进行带宽为 $0 \sim f_{\max}$ 的低通滤波，然后取

$\Delta = 1/2.5f_{\max}$  进行采样, 可在很大程度上防止了混频的发生。

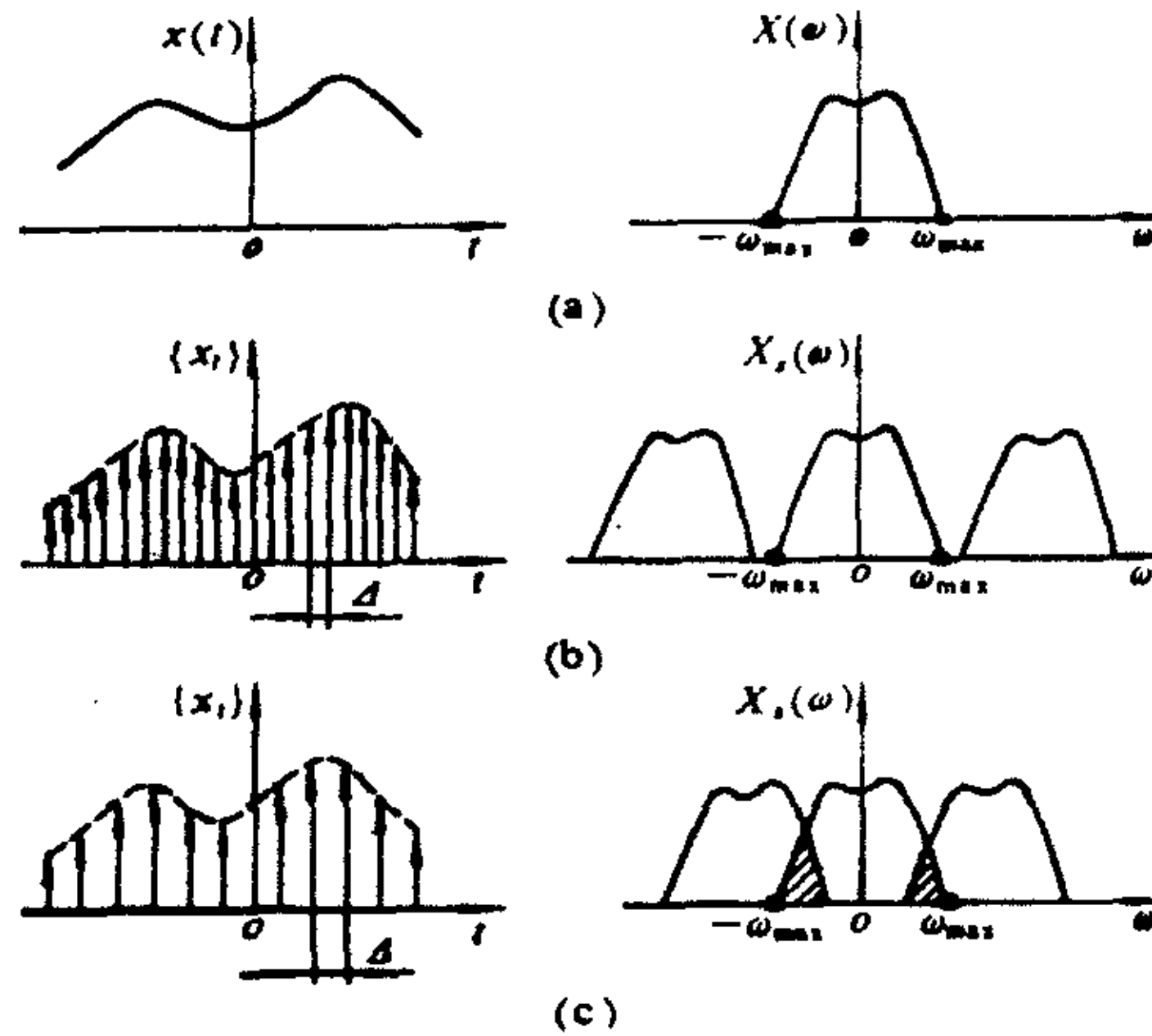


图 2-1 采样信号的频率混迭

(a)  $\omega_s = 2\omega_{\max}$  (b)  $\omega_s > 2\omega_{\max}$  (c)  $\omega_s < 2\omega_{\max}$

除了混频之外还需指出的是, 在确定  $\Delta$  时还应同时考虑两方面的问题: 第一, 当  $\Delta$  取得过大时,  $\{x_i\}$  中将丧失监测信号的真实信息; 第二, 当  $\Delta$  取得过小时, 可能将高频干扰也当作了有用信号, 干扰了正常的数据处理。

### 2.1.2 样本长度的确定

样本长度  $L$  的确定主要与信号在频域中的能量泄漏效应以及不同频率谐波的分辨率有关。从理论上来说, 要了解信号的统计特性, 需要样本长度无限大, 但对于工程应用来说显然是不切实际的。通常采用是加窗处理的方法截取有限长样本。

若假设输出的连续信号为  $x(t)$ , 用矩形窗  $b(t)$  对其进行加窗处理, 得到有限长的信号  $x_s(t)$ , 如图 2-2 所示。

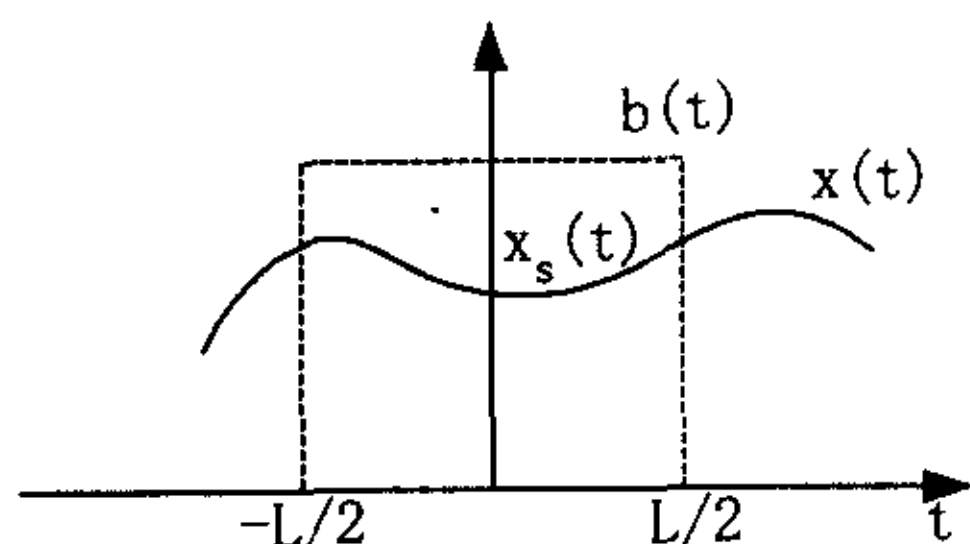


图 2-2 对连续信号进行加窗处理

若假设信号  $x(t)$  中相邻两频域成分的谐波频率分别为  $f_1$  和  $f_2$ ，在频域中的频谱如图 2-3(a) 所示，其频域的间隔为  $f_1 - f_2$ 。由于加窗截取的作用， $x_s(f)$  中该谐波成分的频谱如图 2-3(b) 所示。由于加窗作用，使样本频谱仅分布在某一频域范围，即相当于谱线被拉宽或扩散，这种现象成为能量泄漏，因此该信号对应的频谱将出现主瓣和旁瓣。对于能量泄漏，可根据信号的特征选择不同形状的窗函数以减弱其影响；而对于频率分辨率的问题，则可用窗函数的长度  $L$  加以解决。当窗的宽度  $L > 1/(f_1 - f_2)$  时， $X_s(f)$  呈双峰状，两谐波可很明确地被区分，如图 2-3(b) 所示；当  $L < 1/(f_1 - f_2)$  时， $X_s(f)$  汇合成单峰，两谐波不能被分辨，如图 2-3(c) 所示；当  $L = 1/(f_1 - f_2)$  的临界情况时，两谐波刚好能被分辨出，如图 2-3(d) 所示。

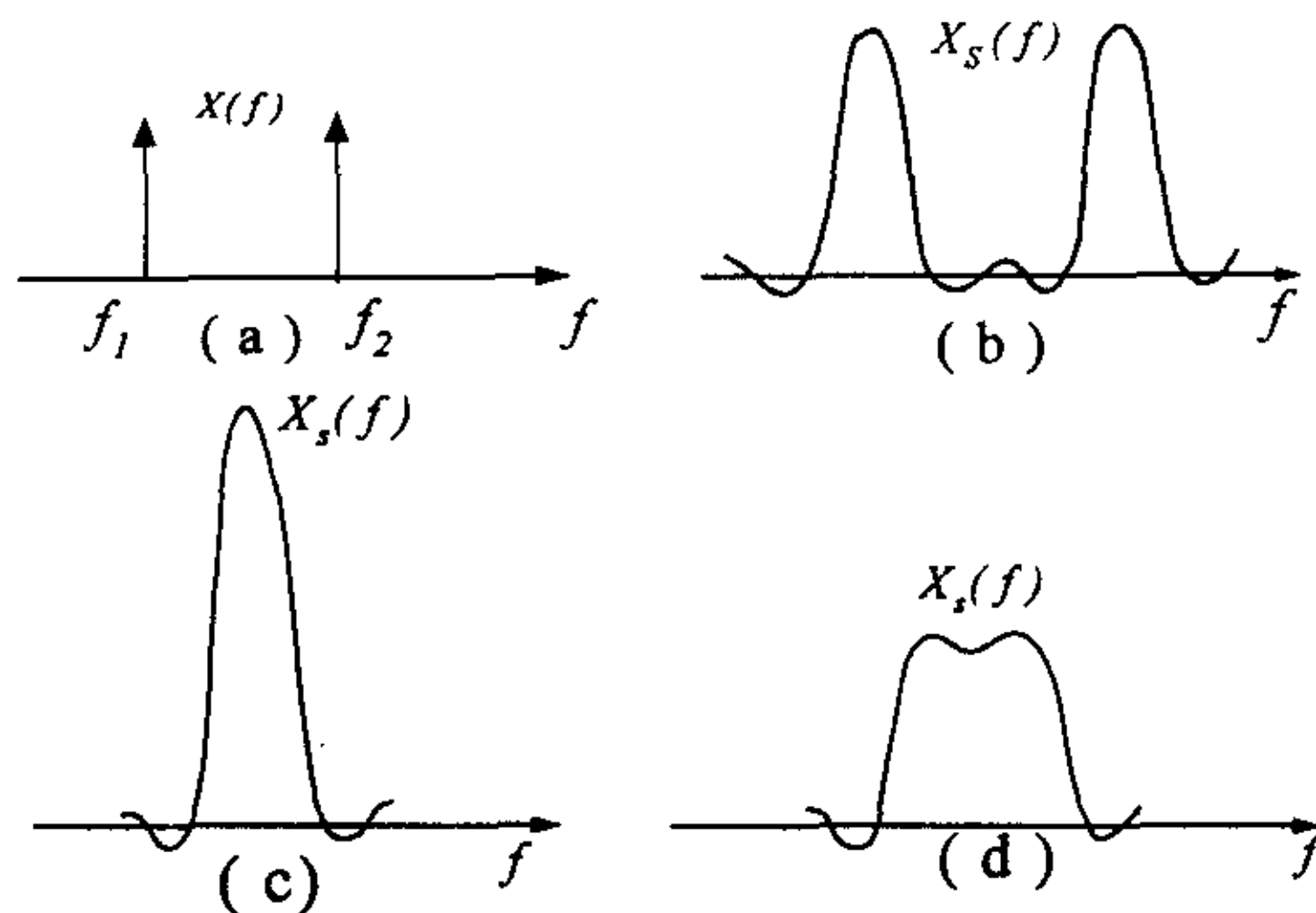


图 2-3 采样信号的频率分辨率

- (a) 信号两谐波的频谱    (b)  $L > 1/(f_1 - f_2)$   
 (c)  $L < 1/(f_1 - f_2)$     (d) 临界状况信号的频谱

由图 2-3 可知, 要保证信号中所需要的谐波频率分量能有效地被分辨, 样本长度  $L$  的选择应满足以下不等式

$$L > \frac{1}{f_2 - f_1} \quad (2-4)$$

由于  $L = N\Delta$ , 因此, 上式也可表达为

$$N > \frac{1}{\Delta(f_2 - f_1)} \quad (2-5)$$

## 2.2 中值滤波快速算法<sup>[8]</sup>

由于在线监测的各个特征量更能反映设备在运行电压下的绝缘状况, 从理论上来说在线测得的数据应该比离线测量的数据更加真实。对于在线测的数据, 通常可把它看作一个时间上连续的离散数据序列即时间序列:  $\{x_i\}$  中  $i=0, 1, 2, \dots$ , 若设备正常运行时,  $\{x_i\}$  为一个平稳的时间序列; 然而在线数据往往是在强噪声等多种干扰的现场环境下测得的, 而且在数据传输过程还可能引起数据失真, 因此, 对在线测得的数据分析诊断之前必须进行预处理。该预处理的目的是在于剔除“野点”, 以削弱环境等干扰因素的影响。预处理的原则是: 既能剔除“野点”, 又能保持数据的真实性。由于在线监测数据量大, 且要求对设备进行实时诊断, 因而还宜对预处理的算法进行优化。本文介绍了一种基于中值滤波的快速算法, 该方法已应用于电力设备在线监测, 且取得了较好的效果。

### 2.2.1 中值滤波的原理

在数字信号处理中, 经常遇到这样的问题: 原始数据、初始信号中包含噪声等干扰源的干扰, 在分析、处理之前需要对其进行预处理以确保数据和信号稳定、可靠, 因此预处理是数据和信号处理的基础, 在各种预处理方法中滤波是一种常用的预处理方法, 而中值滤波通常用于有高频干扰的随机信号处理, 特别是对于有随机脉冲或噪声叠加的信号尤为有效<sup>[16]</sup>。其基本思想是: 在一个滑动窗内对数据序列进行排序处理, 取其中间序号所对应的值作为输出值, 而由各输出值组成了另一个数据序列即滤波后的数据序列。因此中值滤波属于一种非线性的滤波方法。其具体的实现过程如下:

- 1) 由原始数据组成一个数据序列  $f(x_i)$ ;
- 2) 建立一个长度为  $N$  的滑动窗口, 该窗口能沿数据序号增加的方向移动, 即滑动窗  $g(i)$ ;

- 3) 利用  $G(x_i)$  对滑动窗内的数据进行排序;
- 4) 取出排序后的中间值放入或加入到另一数据序列  $h(x_i)$  中;

其数学表达式为

$$f(x_i) = [x_0, x_1, x_2, x_3, \dots, x_i, \dots] \quad i \in [0, +\infty] \quad (2-6)$$

$$g(i) = \begin{bmatrix} 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 1 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & \dots & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & \dots & 0 & \dots & 0 \end{bmatrix} \quad (2-7)$$

滑动窗的窗函数  $g(i)$  为  $n \times n$  矩阵, 其中  $n$  为数据序列的长度,  $i$  应为  $[0, n-1]$ 。矩阵  $g(i)$  从第  $i$  行、 $i$  列开始, 对角线的元素连续  $N$  个值为 1, 其它值为 0,  $N$  为窗函数中从数据序列取出数据的个数。

由公式 (2-6) 和 (2-7) 式得到窗函数中数据个数为  $N$  的数据序列  $F(x_i)$ 。其数学表达式为

$$F(x_i) = f(x_i) g(i) \quad (2-8)$$

令  $G(x_i)$  为排序函数, 由公式 (2-8) 可从数据序列取出  $N$  个数据, 在滑动窗内通过  $G(x_i)$  完成排序, 取出排序后的中间值, 由其组成的滤波后的数据序列  $h(x_i)$ 。该数据序列的数学表达式为

$$h(x_i) = G(F(x_i)) \delta \left( x_{i + \frac{N-1}{2}} \right) \quad (2-9)$$

利用公式 (2-9) 中  $G(F(x_i))$  可对窗口内  $N$  数据按升序或按降序排序。对排序后的序列取中间值后可得到滤波后的序列  $\{h(x_i)\}$ 。

由中值滤波算法及公式推理可知, 算法性能的好坏在很大程度上取决于滑动窗宽度  $N$  的选取。 $N$  的取值越大, 滤波的效果将越显著; 但考虑到中值滤波算法是以滑动窗中间值作为输出值, 因此将产生了  $\text{int} \frac{N}{2}$  的时延, 不利于发挥在线监测的优势, 特别是对突发性故障监测的优势。基于上述因素在滑动窗选取时, 既要保证能滤除干扰, 又要使  $N$  值尽可能的小。在滑动窗宽度选定后, 应确定一个快速的算法对在线数据进行处理。为了综合考虑算法



性能,引入了时间复杂度  $O(n)$  的概念以表征算法的运算量;而中值滤波算法的计算量主要集中在排序算法上,因此本文基于对各种排序算法的比较提出了快速中值滤波算法,且从理论及实例两方面说明其对于在线监测数据预处理的必要性及优越性。例如对  $N$  数据进行一次普通的冒泡排序的运算量为  $O(N^2)$ ,而对快速排序算法理论上的最小运算量为  $O(N \log_2 N)^{[17]}$ ,况且每次实时诊断时还需要进行相同数量级的移动运算。基于以上原因传统的算法不适宜于用作电力设备在线监测,为了实现电力设备在线监测,必须优化排序算法,因此论文中提出了快速滤波算法,其基本思想是:先把无序数据序列转变成有序数据序列,然后对其进行的二分法排序;因而避免了对无序数据序列的反复排序,实现了快速排序和中值滤波的功能。

### 2.2.2 中值滤波快速算法

该算法主要可分为两部分:

1 前  $N$  个数据可由传统排序或快速排序得到一个有序的数据序列。

假设  $\{x_i\}$  为无序数据序列,  $\{y_i\}$  是有序数据序列,  $\{h(x_i)\}$  是滤波后的序列。首先对  $\{x_i\}$  中前  $N$  个数据  $\{x(0), x(1), x(2), \dots, x(N-1)\}$  按升序或降序排序,既可采用快速排序也可采用普通的排序方法得到对应的有序数据序列  $\{y(0), y(1), y(2), \dots, y(N-1)\}$ ,再通过冲击函数取出其中值  $y((N-1)/2)$  作为滤波后序列的第一个值  $h(0)$ 。

2 在有序数据序列的基础上,用二分法快速的查询和内插,实现中值滤波。

对于无序数据序列  $\{x_i\}$  来说,首先由算法的第一部分获得了一个有序数据序列,以后的数据可先通过滑动窗从无序数据序列取出数据,然后与该有序数据序列的数据比较后进行排序、插值,这样无序数据序列就转变成了有序数据序列。若有序数据序列按降序排序,则数据  $x(N)$  先与中间值  $y((N-1)/2)$  比较;若数据  $x(N)$  大于中间值  $y((N-1)/2)$ ,则比较的范围可缩小到前  $((N-1)/2)$  个数据,在前  $((N-1)/2)$  个数据中取中间值与数据  $x(N)$  比较大小;如小于其中值则与后  $((N-1)/2)$  个数据比较,取后  $((N-1)/2)$  个数据中取中间值与数据  $x(N)$  比较大小;依此类推直到比较完成后记录下插入点的位置。若有序数据是按升序排列,则其比较过程也相同,但比较时范围选择的方向应相反。比较完成以后取出  $x(0)$ ,在插入点的位置插入数据  $x(N)$ ,从窗口函数输出中间值  $y(N/2+1)$  作为滤波后的序列  $\{h(x_i)\}$  的第二个值  $h(1)$ ,以后



由滑动窗口获取的数据均可按以上方法处理。

为了进一步分析本文中所用的快速中值滤波算法是否具有—般性,假设已知滑动窗口内数据  $\{x(i), x(i+1), \dots, x(i+N-1)\}$  及其对应的有序数据序列  $\{y(i), y(i+1), \dots, y(i+N-1)\}$ , 求数据序列  $\{x(i+1), x(i+2), \dots, x(i+N)\}$  所对应的有序数据序列及其中值, 具体步骤如下。

第一步, 找出数据  $x(i)$  在有序数据序列  $\{y(i), y(i+1), \dots, y(i+N-1)\}$  中的位置。对有序数据序列用二分法查找, 找出序号并记录下来。

第二步, 求出数据  $x(i+N)$  在有序数据序列中内插的位置。用二分法查找, 即按上面的数据  $x(N)$  的方法查找, 得出数据  $x(i+N)$  在窗口中的位置, 且把窗口位置赋值个一个变量。

第三步, 从有序数据序列中移出  $x(i)$ , 移进  $x(i+N)$ 。插入  $x(i+N)$  后构成了另一有序数据序列, 输出数据序列的中值  $y(i+(N-1)/2)$ , 并把它赋给  $h(i)$ ; 对应窗口中数据序列的序号均加 1, 构成另一新的有序数据序列  $\{y(i+1), y(i+2), \dots, y(i+N)\}$ 。令  $N=2M+1$ ,  $M$  属于整数,  $k, t$  分别代表窗口内数据序列的开始数据的位置、末尾数据的位置, 而  $m$  为滑动窗中值的位置。二分法查找和内插的流程图如图 2-4 所示。

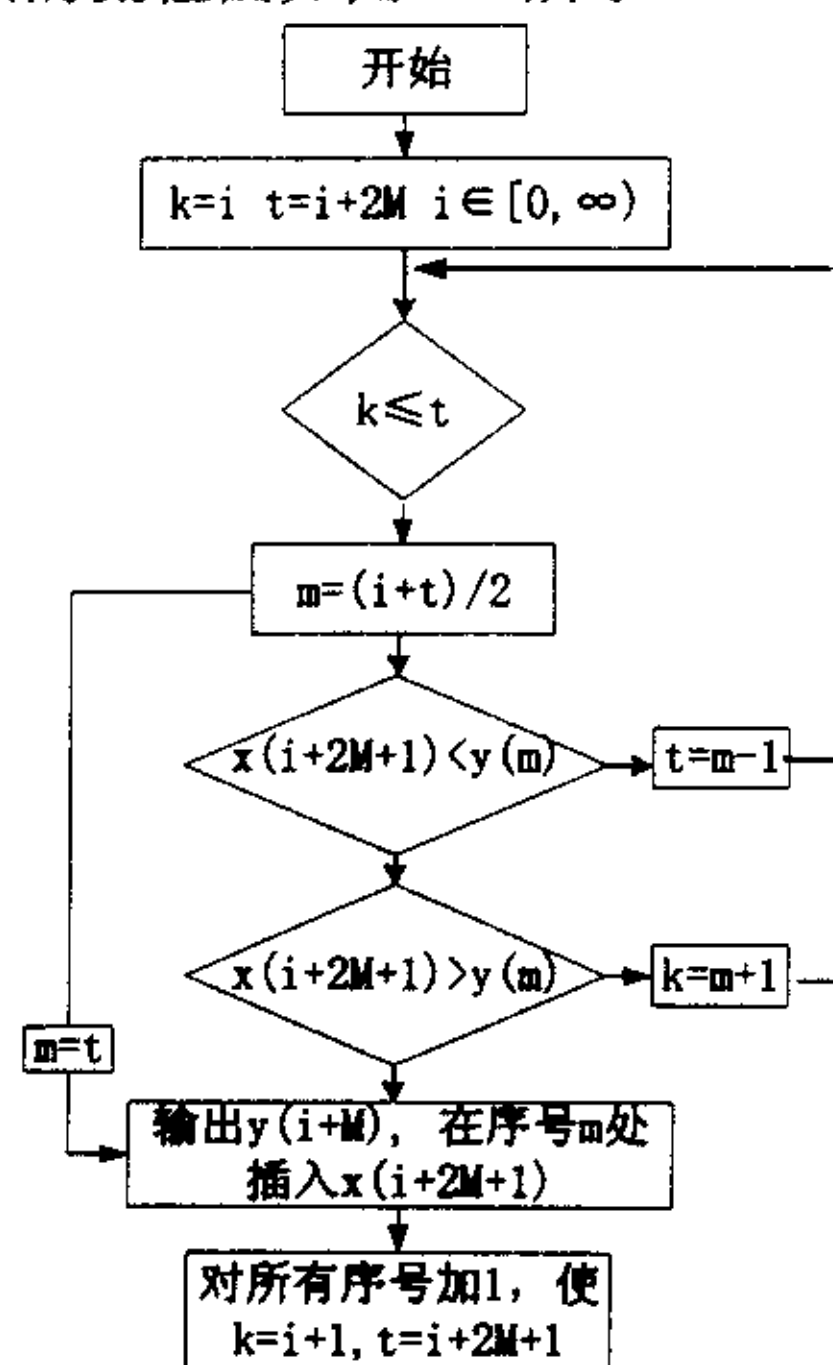


图 2-4 二分法查找及内插流程图

在线测量所获得的无序数据序列经滑动窗处理后可以转化为有序数据序列，然后通过中值滤波快速法实现快速排序。对于滑动窗内一个由  $N$  个数据组成的有序数据序列，用二分法查找序列中某一数据对应的序号，最多只需要进行  $O(1 + \lceil \log_2^N \rceil)$  次运算就能找到；对滑动的数据序列而言，平均查找的次数更少。可见本文提到的中值滤波快速算法的优势在于：运算量显著减小。求取数据序列的中值时，在数据序列中运用二分法比较查找时最大的运算量为  $O(2(1 + \lceil \log_2^N \rceil))$  [16]，此外，还应加上  $N$  次数据移位。而论文中提出的中值滤波快速排序算法无论是数据比较查询的次数还是数据移动的次数都比其它排序算法少得多。为了更进一步说明算法的优势，本文列举窗的宽度  $N=7、15、31、63、127、255$  时各种算法排序的运算量如表 2-1 所示。

表 2-1 不同排序算法运算量的比较

| 窗宽度<br>算法       | 7  | 15  | 31  | 63   | 127  | 255   |
|-----------------|----|-----|-----|------|------|-------|
| 冒泡法排序<br>(次)    | 21 | 105 | 465 | 1953 | 9954 | 32385 |
| 快速排序<br>(次)     | 10 | 59  | 154 | 377  | 888  | 2039  |
| 中值滤波快速<br>算法(次) | 3  | 4   | 5   | 6    | 7    | 8     |

由表 2-1 可以看出：当窗口宽度  $N$  逐渐增大时，冒泡排序算法的运算量呈  $N^2/2$  增长；快速排序算法的运算量呈  $N \log_2^N$  增长；本文的中值滤波快速算法运算量呈  $\log_2^N$  增长。由此可以看出随着  $N$  的不断增长本文中提出的快速算法优势更加明显。为了对不同的算法进行比较，用冒泡法和中值滤波快速算法分别在计算机(CPU 主频为 Pentium4 1.6G，内存 DDR265MB)编程，窗口宽度对应表 2-1，求出其机器时间如表 2-2 所示。

表 2-2 不同排序算法机器运算时间的比较

| 窗宽<br>度算法               | 7       | 15     | 31     | 63     | 127    | 255    |
|-------------------------|---------|--------|--------|--------|--------|--------|
| 冒泡法排序<br>(ms)           | 0.21790 | 1.0250 | 4.4457 | 18.578 | 75.610 | 308.91 |
| 中值滤波快<br>速算法( $\mu s$ ) | 36.125  | 47.031 | 55.657 | 65.578 | 75.578 | 85.688 |

由表 2-2 可知, 冒泡法的机器时间基本上在毫秒级水平, 而中值滤波快速算法的机器时间基本上在微秒水平, 两者机器时间相差很大; 且冒中值滤波快速算法机器运算时间随窗口的增加远不如冒泡法迅速, 因此就机器运算时间而言, 可以证明中值滤波快速算法比常用的排序算法快得多。

### 2.2.3 中值滤波快速算法的应用

变电设备在线数据在测量、传输过程中不可避免的受到了电晕、数据传输失真等因素的影响, 这些影响因素的作用相当于在原始数据序列中叠加了一些尖峰数据, 这时可运用中值滤波快速算法对数据预处理效果明显。为了验证算法的有效性、实用性, 应选用一些典型的数据进行分析, 通过对大量典型数据的分析, 如确实获得正确的结果, 进而可将此中值滤波的快速算法推广。例如对电容型设备 (高压套管、CT 等) 的介质损耗因数  $\tan\delta$  值往往仅千分之几, 因此在线监测时, 即使环境微小的变化或传输过程存在微弱的干扰也可能引起介损测量值的跳跃, 反映在介损数据序列中则表现为某些尖峰数据, 数据的这种跳跃严重影响了对设备的状态的诊断分析。本文选用了广东省某 110kV 变电站 C 相 CT 的介损在线数据, 并以此为例进行了分析、处理。通过选取不同宽度窗函数对这些介损的原始数据进行中值滤波快速处理, 处理前和处理后的图形如图 2-5 所示。

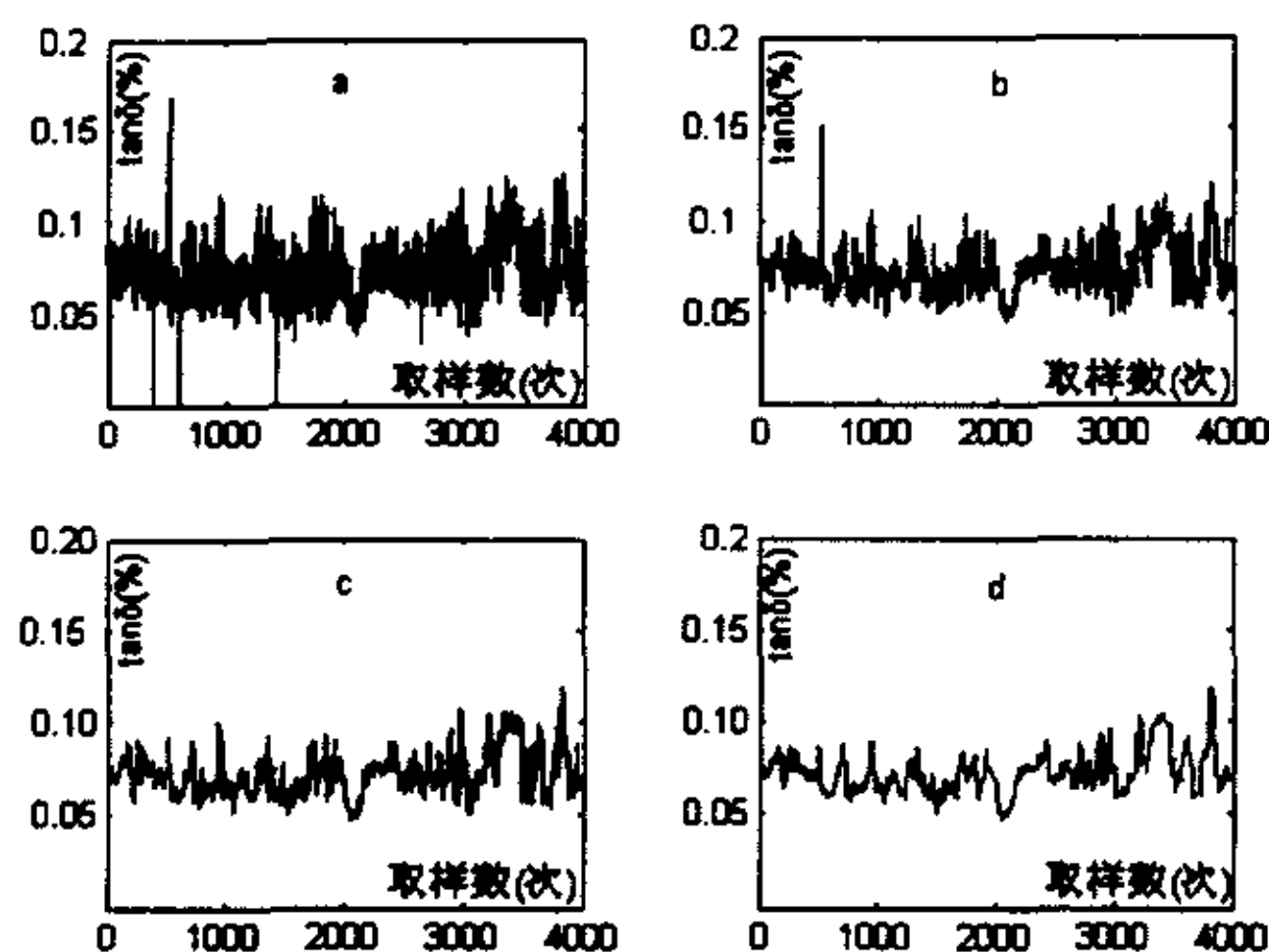


图 2-5 介损的原始数据及中值滤波后的图  
(a) 原始数据的图      (b) 滤波后的图 ( $N=7$ )  
(c) 滤波后的图 ( $N=15$ )      (d) 滤波后的图 ( $N=31$ )

## 2.3 预处理算法的选择

由于在线监测的目的在于准确、及时反映设备的运行状态,因此,对于滤波算法而言,不仅要考虑滤波后数据的真实性,而且也应考虑其实时性;同时对不同的对象所选用的方法应有所差别。

### 2.3.1 奇异点及处理方法

在采集信号和传输数据的过程中,由于噪声干扰、数据丢失、传感器瞬时失灵等原因,造成个别数据过高或者过低,这些点是应当剔除的,称为野点。通过对在线数据的分析、归纳,可将其野点(奇异点)归纳为如下几种类型:(1)孤立野点:一般在时间序列中是一个很大的尖峰,而野点邻近的数据都远小于(或大于)野点值,如图 2-6(a);(2)孤立的连续野点:该类型野点一般在时间序列中也是一个很大的尖峰,且连续出现的几个野点的左邻和右邻的数据都远小于(或大于)野点值,如图 2-6(b);(3)阵发噪音型野点:这种野点类似噪音,在时间序列里也像是有些规律,然而在每组野点内部却起伏很大,远离真实值的程度也大小不一,如果这样的野点很多,时间序列将会变得难以应用,如图 2-6(c)<sup>[5]</sup>所示。

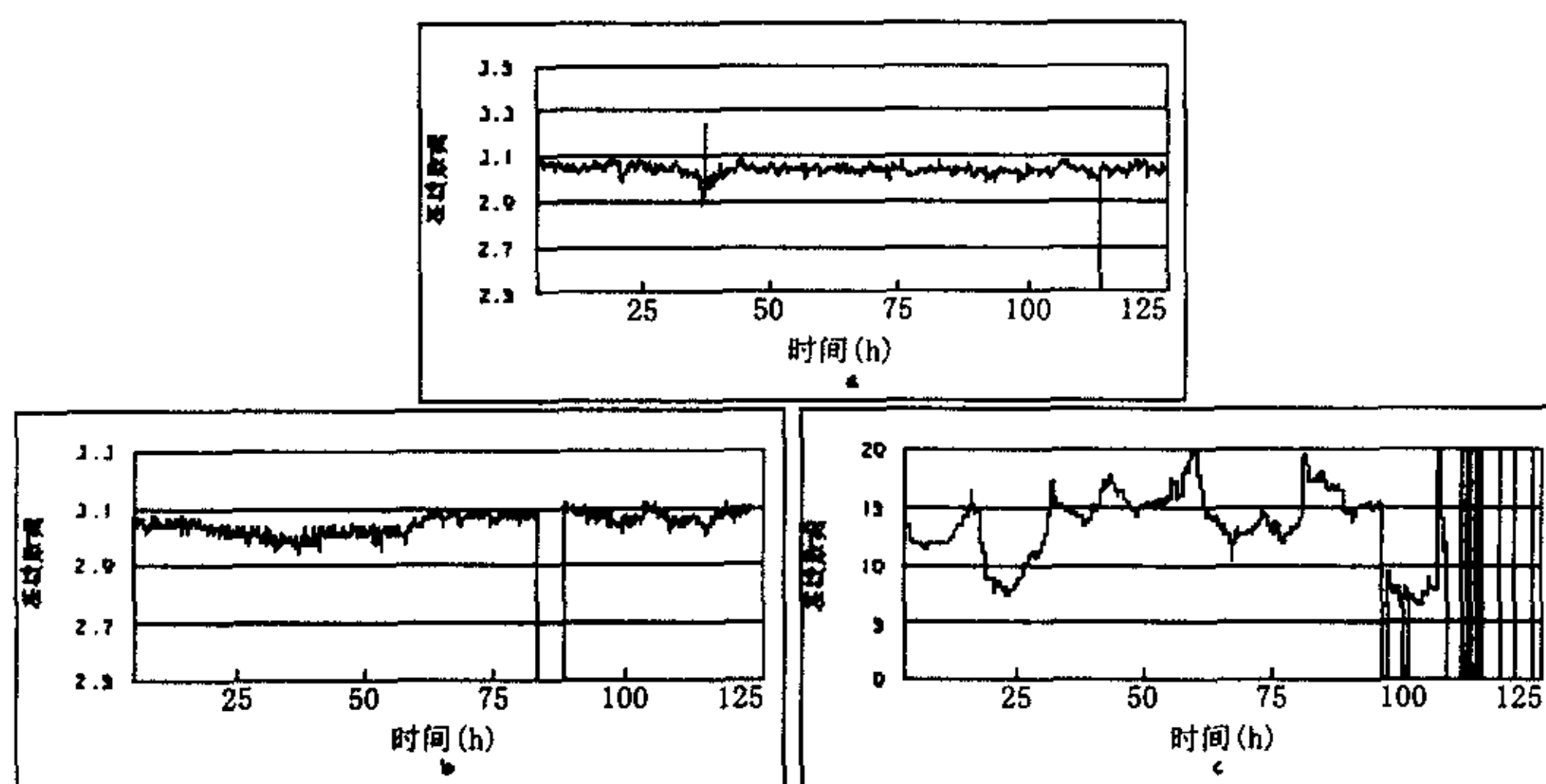


图 2-6 在线监测数据中野点的几种类型

(a)孤立野点; (b)孤立的连续野点; (c)阵发噪音型野点

对于 1、2 两类野点,常用差分的方法滤掉:构造一个新的时间序列  $s_n = a_n - a_{n-1}$ , 因为在野点的左邻和右邻各有一个很大的跳跃点并且方向是相反的,用这样的方法很容易滤掉这两类野点。而第 3 类野点不太容易滤

除,但由于这类情况往往强噪声源干扰而产生的,遇到此类情况时也可以在数据采集时采取一些适当的措施。

### 2.3.2 时域内几种常见的预处理算法

1. 53H 算法<sup>[5,6]</sup>:其基本思想是产生一个曲线的平滑估计,然后将测量值与这一估计值进行比较来识别野点。其步骤如下:

1) 设  $x(i)$  为测量的在线数据序列。为从  $x(i)$  构造一个新序列  $x_1(i)$ , 方法是取  $x(1), \dots, x(5)$  的中间值作为  $x_1(3)$ , 然后舍去  $x(1)$ 、加入  $x(6)$ , 取中间值得到  $x_1(4)$ ; 依此类推, 直到加入最后一个数据。显然,  $x_1(i)$  的项数比  $x(i)$  的项数少四项。

2) 用类似的方法在  $x_1(i)$  的相邻三个数中选取中间值而构成序列  $x_2(i)$ 。

3) 最后由序列  $x_2(i)$  按如下方式构成  $x_3(i)$

$$x_3(i) = 0.25x_2(i-1) + 0.5x_2(i) + 0.25x_2(i+1) \quad (2-10)$$

这是一个 Hanning 平滑滤波器。

4) 如果有下式成立, 则用  $x_3(i)$  代替  $x(i)$ ,

$$|x(i) - x_3(i)| > k \quad (2-11)$$

其中  $k$  为一预定值。

2. 滑动平均算法<sup>[15]</sup>: 将一个宽度为  $N$  个点的窗口逐点向前移动(滑动窗), 每次取窗口内所取点的算术平均值作为中间点的输出值。若  $N=2k+1$  且输入、输出分别为  $x(n)$  和  $y(n)$ , 则有以下式

$$y(n) = \frac{1}{2k+1} \sum_{i=-k}^k x(n+i) \quad (2-12)$$

3. 中值滤波快速算法: 其具体思想已如 2.2 节所述。

### 2.3.3 时域算法的实例分析

以广东省某 110kV 变电站 CT 介损的在线数据为对象, 分别采用 53H 算法、滑动平均算法和中值滤波快速算法进行预处理, 并对其处理后的数据比较分析和仿真处理, 处理前、处理后图形如图 2-7 所示。图 2-7(a) 为介损原始数据图, 由图可知其仿真图明显存在几个尖峰干扰; 图 2-7(b) 为经中值滤波快速算法处理后的图, 相对于图 2-7(a) 而言, 尖峰被抑制了, 数据的趋势基本得到了保留, 因此该中值滤波方法效果较好; 图 2-7(c) 和 2-7(d) 分别为经滑动平均算法和 53H 算法处理后的图, 与图 2-7(a) 比较, 他们在某种程度上对抑制尖峰干扰有一定作用, 但效果不如图 2-7(b) 明显。

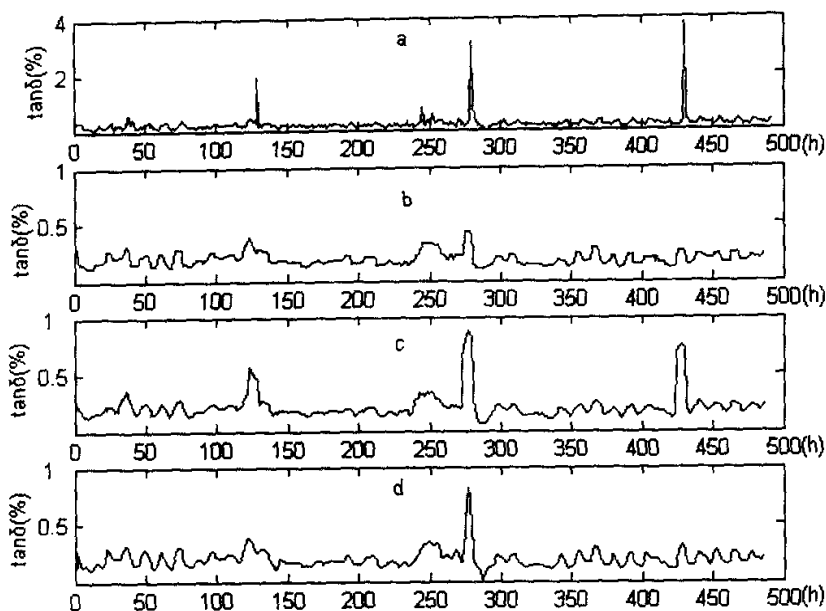


图 2-7 几种预处理算法的仿真结果

(a) 介损原始数据图 (b) 中值滤波快速处理图

(c) 滑动平均算法处理图 (d) 53H 算法处理图

以上几种算法的原理都是通过滑动窗的方式来实现滤波，其代价是以牺牲数据的实时性，因而滑动窗宽度的选择十分重要。53H 算法窗口宽度被限定，对于削弱阶跃性干扰效果较好，而对其它的干扰则不适应，因此，其灵活性较差。中值滤波快速算法和滑动平均算法窗口的选择较灵活，但对窗口选择依赖较大，若窗的宽度选择不合适，可能会导致两方面的问题：一方面窗口较小可能达不到滤波的目的，另一方面窗口过大可能因为时延较长很难实现真实意义上的实时监测。

相对于正常数据而言，奇异点的幅值一般较大，可能是正常数据的几倍或者更大，考虑到在线监测的实时性，滑动窗的宽度不宜取得太大；对于滑动平均算法而言，任何一个阶跃的奇异点，都可能使其输出值有较大的跳跃，因此有可能对以后数据的分析、处理造成很大干扰；而中值滤波快速算法只要滑动窗选择合适，则有可能完全避免。而当设备存在故障或故障隐患时，反映出来的征兆是输出信号变化，其表现形式通常是循序渐进的过程；而对

反映出来的征兆是输出信号变化，其表现形式通常是循序渐进的过程；而对于突发性故障而言，以上几种预处理算法都可反应其变化的征兆，其表现形式可能有所差别。滑动平均算法对奇异点和故障数据区分很困难，因此可能造成奇异点时，诊断误报，而故障数据时却不报；对于 53H 算法和中值滤波快速算法而言，若干扰为阶跃性的奇异点，则上述两种算法均可区分干扰数据和故障数据；但对于 53H 算法窗的宽度被限定，其适应的范围受到限制，而对于中值滤波快速算法而言，只要滑动窗选择较为灵活，在实际应用中选择范围较大，适合工程应用。综上所述中值滤波快速较适合于在线数据的预处理。

#### 2.3.4 频域处理方法

上述的分析处理均在时域进行的，通常变电站内存在很强的电晕干扰或其它放电源的干扰，这些因素都可能导致监测信号中包含高频干扰分量；而低频干扰分量则可能来源于传感器铁心致使监测信号的畸变，因此仅从时域的范围处理可能很难达到实际要求。以广东某变电站 CT 介损为例，分别作出其时域和频域的图形，如图 2-8 所示。

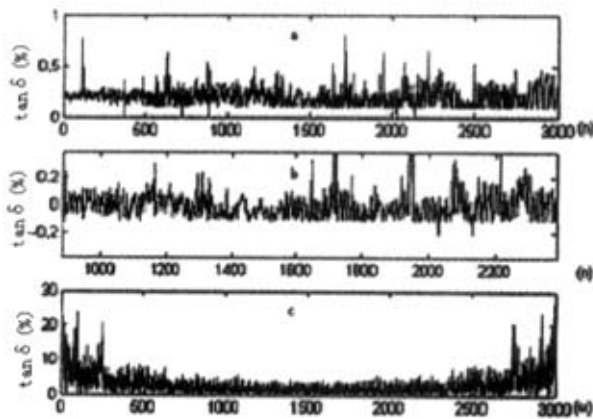


图 2-8 CT 介损时域和频域图

(a)原始数据时域图 (b)滤除直流分量后时域图 (c)频域图

对图 2-8(c)频谱分析可知介损在频域内低频分量和高频分量都较大，低频分量主要体现为设备绝缘特性，通常情况下它变化非常缓慢，变化周期也较长；而高频分量则反映了信号测量时的噪声污染。因此，如果将监测数据构成数据序列进行傅立叶变换，并将变换后的高频和低频分量置一个较低数



值或置零,然后对其进行反傅立叶变换从而实现滤除干扰的目的,即可通过设计一个带通滤波器来实现。采用傅立叶变换(FFT)可以保证在频域内信息完整性、准确性。如果监测数据序列的频率特性随时间的变化而变化,是一种时变序列,而傅立叶变换难以实现实时显示任意时刻数据的频谱信息,此时傅立叶变换很难达到要求。

### 2.3.5 小波奇异性处理

由于小波变换在时域和频域同时具有局部性,能够计算某一特定时刻的频率分布,因此能将各种交织在一起的不同频率组成的混合信号分解成不同频率的信号块<sup>[18]</sup>,一般情况下噪声信号的频率较高(局部放电信号除外),应存在于低尺度小波变换中,因而通过将在线监测数据用小波进行多尺度分析,将低尺度小波系数置零或选择合适的阈值量化,然后对其进行重构,可实现信号去噪目的。其实质是一种由小波构造带通滤波器的思想,可解决傅立叶变换只能处理全局信息的问题,同时能在较低信噪比的情况下灵活、有效地减弱噪声干扰。

以上的小波处理方法是一种全局的处理思想,无法有效地发现数据中的奇异点或不规则突变成分,且有可能丢失部分有用的信息。如能剔除数据中的奇异点,则可提高其处理效率并尽可能地保留数据的原始性。时域分析的方法可以在平稳的前提下有效地检测出因幅值发生突变而引起信号的非连续的奇异点,对于信号较平滑,幅值没有突变,但其一阶微分有突变的奇异点却缺乏行之有效的处理手段。小波的奇异变换为解决该问题提供了新的思想,信号小波变换系数模极大值的位置和幅度同信号的局部奇异性密切相关,可利用这些极大值以检测出信号的局部奇异性。

在小波变换中,局部奇异性可定义为:定义 1<sup>[19]</sup>设  $f(x) \in L^2(\mathbb{R})$  (平方可积空间),若  $f(x)$  对  $\forall \epsilon \in \delta x_0$ , 小波  $\psi(x)$  是满足连续可微的紧支集小波,并具有  $n$  阶消失矩,  $n < \alpha < n+1$ ,  $A$  为常数,有

$$|w_f(s, x)| \leq As^\alpha |w_f(s, x)| \leq As^\alpha \quad (2-13)$$

则称  $\alpha$  为  $x_0$  点的奇异指(即李普西兹指数)。如果函数在  $x_0$  处的李普西兹指数  $\alpha$  为正值,那么随着尺度的增加,小波变换模极大值的幅值也会变大;若李普西兹指数  $\alpha$  为负值,则情况相反。特别需要指出的是李普西兹指数  $\alpha$  刻画了函数在该点的正规性,其值越大,函数越光滑。函数在一点连续、可微,则该点的李普西兹指数  $\alpha \geq 1$ ; 在一点可导,而导数在有界但不连续时,

李普西兹指数  $\alpha$  仍为 1；如果函数在  $x_0$  的李普西兹指数  $\alpha < 1$ ，则称函数在  $x_0$  是奇异的；一个在  $x_0$  点不连续但有界的函数，则该点李普西兹指数  $0 \leq \alpha \leq 1$ ； $x_0$  点的冲击函数在该点李普西兹指数  $\alpha$  为 -1。

通过寻找尺度空间中的模极大值点，计算其李普西兹指数  $\alpha$ ，判断是否小于 1，可发现信号中的奇异点，因此通过小波分解和重构可实现去噪和平滑。但小波分解和重构的计算复杂且计算量较大，而时域分析去噪方法较多，且实现灵活、简单，但信号含有高频或低频分量很难处理。因此可把时域算法与小波算法结合起来应用，实现优势互补。利用小波变换检测奇异点的位置与性质，然后再根据奇异点的性质用时域分析的方法进一步平滑去噪，实现起来较为灵活，且可提高处理的效率与准确性。

## 2.4 小结

本章主要从在线数据的采集和预处理两方面进行了深入的探讨和研究。

1 数据采集时要注意时间间隔  $\Delta$  与样本长度  $L$  的选择，应确保采样频率  $f_s \geq 2f_{\max}$ ，样本数量  $N > \frac{1}{\Delta(f_2 - f_1)}$ 。

2 在对数据进行预处理时，分别从时域、频域等方面进行了讨论。在时域范围内处理时，首先提出了一种快速算法-中值滤波快速算法，该算法能有效地滤除奇异点，对平缓数据序列效果较好，其次从算法有效性和实用性两个角度对几种常见的算法进行了比较。而信号含有高频或低频分量时，时域算法难以实现；在频域内处理时，数据时变性和实时性又难以得到保障；而小波变换具有奇异性，因此可以考虑使用小波变换进行预处理。

### 3 基于时间序列的数据预测

第2章所研究的是数据的采集和预处理技术,但为了对设备的运行状态进一步的了解,需要对设备运行状态进行预测及其性能评估。而设备的状态信息通常是通过在线数据反映的,因此通过对在线数据的预测从而实现设备状态预测和性能评估。

目前,时间序列分析是一种常用的数据预测和分析的方法,本文以时间序列分析为工具,变电设备在线数据为实例,从预测的数学基础、建模、实例分析等方面进行研究。

#### 3.1 时间序列预测的数学基础<sup>[9]</sup>

时间序列是按时间顺序排列的一组数据,即一组有序的随机数据。时间序列分析主要是利用参数模型(ARMA模型、AR模型、MA模型等)对所获取的有序的数据进行分析和处理的一种数据处理方法。

监测信号(包括连续信号与离散信号数据序列)是被监测对象的信息载体。时间序列是被监测对象按时间或空间先后顺序排列的数据,其蕴含的信息主要通过两方面体现:一是依靠数据的顺序,二是依靠数据的大小。

##### 3.1.1 时间序列的统计特性

时间序列是一类随机过程,因此也具有统计特性,且这种特性能刻画随机过程的本质,从偶然性揭示必然性;其最根本的统计特性是分布函数,但求解时相当繁琐,在工程实际应用中通常用矩函数来反映其统计特性,如一阶矩函数(均值函数)、二阶矩函数(方差、自相关函数和协相关函数等),所有的这些函数都是确定性函数,便于应用。

设 $\{x_t\}$ 为离散随机过程,其中 $t=0, \pm 1, \pm 2, \dots$ ,则一阶矩函数(均值函数)定义为<sup>[9]</sup>

$$\mu_{x,t} \triangleq E[x_t] \quad (3-1)$$

$E$ 表示集合平均算子,即数据期望。

二阶矩函数中,自协相关函数定义为

$$C_{xx,k,t} \triangleq E[(x_t - \mu_{x,t})(x_{t-k} - \mu_{x,t-k})] \quad (3-2)$$

自相关函数定义为

$$\psi_{x,t} \triangleq E[x_t^2] \quad R_{xx,k,t} \triangleq E[x_t x_{t-k}] \quad (3-3)$$

方差函数定义为

$$\sigma_{x,t}^2 \triangleq E[(x_t - \mu_{x,t})^2] \quad (3-4)$$

均方值函数定义为

$$\psi_{x,t} \triangleq E[x_t^2] \quad (3-5)$$

在二阶矩函数中自协相关函数最为重要，其他的三个二阶矩函数是自协相关函数的特例，对于某些随机过程要刻划其本质，要求其三阶、四阶等高阶矩函数，求解过程相当繁琐，因此，在一般情况下采用特殊的随机过程以解决工程实际应用。

1 平稳随机过程。当所有描述随机过程的矩函数与统计起点无关时，则称之为严平稳随机过程；当只有一阶、二阶矩函数与统计起点无关时，则称宽平稳随机过程。

2 正态随机过程。若随机过程的各阶矩函数取决于一阶、二阶矩函数，则之为正态随机过程(Gauss 随机过程)。此随机过程一定是严平稳随机过程。

3 遍历性(各态历经)随机过程。若从随机过程中任取两个样本  $x_{i,t}, x_{j,t}$ ,  $i \neq j$ ,  $i, j = 1, 2, \dots$ , 又有  $x_{i,t} = x_{j,t-k}$ ,  $k$  为某一延迟步数，则称此随机过程为遍历性随机过程。遍历性随机过程一定是平稳随机过程，其对应样本的各阶矩函数时间平均可代替整个随机过程各阶矩函数的集合平均。

4 白噪声。一类特殊的平稳随机过程，且彼此独立不相关。其对应的均值为零，方差为某一常数。

### 3.1.2 ARMA 模型、AR 模型、MA 模型

1. ARMA(Autoregressive Moving Average)模型。对于一个平稳、零均值的时间序列  $\{x_t\}$ ,  $t=1, 2, 3, \dots, N$ , 一定能拟合成一个差分方程，具体形式如下

$$x_t - \varphi_1 x_{t-1} - \varphi_2 x_{t-2} - \dots - \varphi_n x_{t-n} = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_m a_{t-m} \quad (3-6)$$

其中  $x_t$  是时间序列  $\{x_t\}$  在  $t$  时刻的元素， $\varphi_i$  ( $i=1, 2, \dots, n$ ) 称为自回归参数， $\varphi_j$  ( $j=1, 2, \dots, m$ ) 称为滑动平均参数， $a_t$  为模型残差，其时序应为白噪声。公式(3-6)中左边  $n$  阶差分多项式称为  $n$  阶自回归部分，右边  $m$  阶差分多项式称为  $m$  阶滑动平均部分，因此该模型可记为 ARMA( $n, m$ ) 模型。

为了研究方便，引入了后移算子  $B$ ，定义如下：设时间序列  $\{w_t\}$ ，则有  $Bw_t = w_{t-1}$ ，以此类推公式(3-6)可改写为<sup>[9]</sup>

$$(1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_n B^n) x_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_m B^m) a_t \quad (3-7)$$

若设

$$\begin{aligned}\varphi(B) &= 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_n B^n \\ \theta(B) &= 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_m B^m \\ x_t \varphi(B) &= \theta(B) a_t\end{aligned}\quad (3-8)$$

2. AR(Autoregressive)模型。该模型可以认为是 ARMA 模型中  $\theta(B)=1$  的情况, 若存在  $n$  阶差分多项式, 则称之为 AR( $n$ ) 模型, 其表现形式可写为

$$x_t = \sum_{i=1}^n \varphi_i x_{t-i} + a_t \quad (3-9)$$

3. MA(Moving Average)模型。该模型可以认为是 ARMA 模型中  $\varphi(B)=1$  的情况, 若存在  $m$  阶差分多项式, 则称之为 MA( $m$ ) 模型, 其表现形式可写为

$$x_t = \sum_{i=1}^m \theta_i a_{t-i} + a_t \quad (3-10)$$

### 3.1.3 ARMA 模型的物理解释

建立 ARMA 模型所用的观测数据序列  $\{x_t\}$ , 可视为某一系统的输出, 而  $\varphi_i$ ,  $\theta_i$  和  $\sigma_a^2$  是基于序列  $\{x_t\}$  按某一方法估计出来的模型参数, 因此系统特性和系统工作状态的所有信息都蕴含在  $\{x_1, x_2, \dots, x_N\}$  这  $N$  个数据中, 且主要与数据的大小及其先后顺序有关, 因此其信息也就蕴含在  $\varphi_1, \varphi_2, \dots, \varphi_n$ ,  $\theta_1, \theta_2, \dots, \theta_m$  和  $\sigma_a^2$  这  $n+m+1$  个参数之中。这正是 ARMA 模型的一个最大的特点——信息的凝聚性, 即大量数据所蕴含的信息凝聚成为少数几个模型参数, 这中信息的凝聚称之为数据的压缩。

若对公式 (3-8) 进行变换可得

$$x_t = L(B) a_t \quad \text{其中 } L(B) = \frac{\theta(B)}{\varphi(B)} \quad (3-11)$$

若从系统的角度来看, 则  $a_t$  对应系统的输入信号, 而  $x_t$  对应系统的输出信号, 那么 ARMA( $n, m$ ) 模型描述了传递函数为  $L(B)$  的系统。可对  $\varphi(B)$  和  $x_t = L(B)a_t$  进行分解<sup>[9]</sup>

$$\begin{aligned}\varphi(B) &= (1 - \lambda_1 B)(1 - \lambda_2 B) \cdots (1 - \lambda_n B) = \prod_{i=1}^n (1 - \lambda_i B) \\ \theta(B) &= (1 - \eta_1 B)(1 - \eta_2 B) \cdots (1 - \eta_m B) = \prod_{j=1}^m (1 - \eta_j B)\end{aligned}\quad (3-12)$$

公式 (3-12) 中,  $\lambda_i$  和  $\eta_j$  分别对应 AR 部分和 MA 部分的特征根; 而从系统



的观点来看,  $1/\lambda_i$  对应系统传递函数的极点,  $1/\eta_i$  MA 的特征根对应系统的零点。系统传递函数的极点表征系统固有的、与外界无关的特性, 而其零点则表征系统与外界的相互联系。由于 ARMA 模型是基于数据序列  $\{x_t\}$  建立起来的, 不论系统的输入是否可观测, 都没有利用系统的任何信息, 而是将白噪声  $\{a_t\}$  的数据序列视为输入。设输入为某一平稳时间序列  $\{s_t\}$ , 则无法直接对其按时间序列的方法进行建模, 可将  $\{s_t\}$  视为某一线性环节在白噪声  $\{a_t\}$  驱动下的输出, 则有

$$s_t = \frac{\theta'(B)}{\varphi'(B)} a_t \quad (3-13)$$

因此可先将平稳时间序列  $\{s_t\}$  转换成白噪声输入再进行建模, 如下所示

$$x_t = \frac{\theta(B)}{\varphi(B)} \frac{\theta'(B)}{\varphi'(B)} a_t \quad (3-14)$$

若  $\varphi(B)$ 、 $\theta'(B)$  无公共的因子式, (3-14) 式中分母包含  $\varphi(B)$ , 这表明 ARMA 模型表示的等价系统的固定信息一定包含在实际的系统的固定信息中, 因此该模型不存在丢失信息的现象, 从拓宽了模型应用的范围。

### 3.2 AR 预测模型的建立

预测模型的建立就是对所测得的时间序列  $\{x_t\}$  ( $t=1, 2, \dots, N$ ) 拟合出合适的时序模型, 建模的内容包括数据的采集、数据的检验、数据的预处理、模型方式的选择、模型参数的估计和模型的实用性检验, 其中模型参数的估计最为重要。由于数据的采集和数据的预处理在第二章已进行了详细的阐述, 因而本节主要对其它部分进行研究, 其中模型参数的估计是研究的重点。

#### 3.2.1 数据的检验

时间序列 (ARMA, AR, MA) 模型均要求数据为平稳、正态、零均值的时间序列, 因此, 必须对预处理后的数据序列进行检验以确定其是否满足条件。

1 平稳性检验。根据平稳时间序列  $\{x_t\}$  的定义可知: 其均值  $\mu_x$  和方差  $\sigma_x^2$  为常数, 且其自协方差函数  $R_x$  仅与时间间隔有关。因此, 对  $\{x_t\}$  进行平稳性检验, 最根本的就是检验其是否具备这两个性质。采用的方法为: 逆序检验, 即将时间序列  $\{x_t\}$  ( $i=1, 2, \dots, N$ ) 均匀分成  $L$  子序列, 每个子序列的长度为  $M$ ,  $N=LM$ , 第  $i$  段序列的形式可表示为:  $\{x_{it}\} = \{x_{i1}, x_{i2}, \dots, x_{iM}\}$ 。对各段子序列而言, 可应用数理统计的理论分别计算出各子序列的均值、方差、自协相关系数; 以均值  $\mu_i$  为例, 这些均值构成了一个序列:  $\mu_1, \mu_2, \dots, \mu_L$ , 当

$i > j (j = 1, 2, \dots, L-1)$  时, 出现  $\mu_i > \mu_j$ , 定义为  $\mu_j$  的一个逆序; 对于某一  $\mu_i$ , 定义  $\mu_j$  的逆序数  $A_j$  为  $\mu_i > \mu_j (i > j)$  出现的次数, 而所有的逆序数  $A_j$  的总和称为序列逆序总数  $A$ 。

由于  $\{x_t\}$  是随机过程的一个样本, 则各子序列均值  $\mu_i$  的值也是随机的。当  $\{x_t\}$  是一个平稳序列时,  $\mu_1$  后面  $L-1$  个随机数  $\mu_i$  大于  $\mu_1$  和小于  $\mu_1$  几率是相等的, 故其逆序数的平均值  $E[A_1] = (L-1)/2$ , 同理也可求出  $\mu_i (i \neq 1)$  的逆序数的平均值, 因此总的逆序数的平均值及方差为<sup>[9]</sup>

$$E[A] = \sum_{j=1}^{L-1} E[A_j] = \frac{L(L-1)}{4} \quad (3-15)$$

$$\sigma_A^2 = \frac{L(2L^2 + 3L - 5)}{72}$$

构造一个统计量  $u$ , 对  $A$  进行正规化处理有

$$u = \frac{\left(A + \frac{1}{2} - E[A]\right)}{\sigma_A} \quad (3-16)$$

$\mu$  满足标准的正态分布:  $\mu \sim N(0, 1)$ 。因此当取其置信概率为 95% 时, 若有  $|u| \leq 1.96$ , 则认为  $\mu$  是平稳序列, 因此  $\{x_t\}$  也为平稳序列; 反之, 则认为  $\{x_t\}$  是非平稳序列。

2 零均值性检验。检验时间序列  $\{x_t\}$  的均值  $E(x_t)$  是否为零, 此处时间序列  $\{x_t\}$  是整个随机过程, 而不随机过程的样本。对于一个长度为  $N$  的样本可通过计算其算术平均来实现, 如  $\hat{\mu}_x = \left(\sum_{t=1}^N x_t\right) / N$ ,  $\hat{\mu}_x$  是  $\mu_x = E(x_t)$  估计值, 根据随机序列各态经历特性, 可用其算术平均值代替均值, 因此只需要检验  $\hat{\mu}_x$  是否满足  $|\hat{\mu}_x| < \xi$ ,  $\xi$  可定义较小的数。对于  $\hat{\mu}_x$  不满足条件的情况, 可其进行零均值处理, 从而得到处理后的时间序列  $\left\{\hat{x}_t\right\}$ , 即有  $\hat{x}_t = x_t - \mu_t$ 。

### 3.2.2 模型的选择<sup>[9]</sup>

ARMA( $n, m$ ) 模型既包含自回归部分, 又包含滑动平均部分, 因此能蕴含数据所有信息, 但求解时需要对其  $n+m+1$  个参数进行估计, 计算量和计算的难度很大。根据 ARMA 模型的解释可知, 模型的极点反映系统的固有的、与外



界无关的特性，而 ARMA 的极点又与 AR 模型的特征根有关，因此将 ARMA 模型转变成 AR 模型其系统固有的特性仍然得以维持，而且只需要求解  $n+1$  个参数，变换的证明过程如下<sup>[9]</sup>

将公式(3-8)改写，有： $\frac{\varphi(B)}{\theta(B)}x_t = a_t$ ，可对其进行进一步的分析，有

$$\left[ \varphi_1(B) + \frac{\varphi_2(B)}{\theta(B)} \right] x_t = a_t \quad (3-17)$$

当  $\varphi(B)$  的阶数高于  $\theta(B)$  的阶数 (即  $n > m$ )，可从  $\varphi(B)$  中分解一部分多项式算子即  $\varphi_1(B)$ ；而剩余部分为阶数低于  $\theta(B)$  的一部分算子。当  $\varphi(B)$  的阶数低于的阶次式 (即  $n < m$ )， $\varphi_1(B) = 0$ 。对公式(3-17)中  $\theta(B)$  进行分解，有

$$\left[ \varphi_1(B) + \frac{\varphi_2(B)}{\prod_{j=1}^m (1 - \eta_j B)} \right] x_t = a_t \quad (3-18)$$

对公式(3-18)中左边第二项进行分解，有

$$\varphi_1(B)x_t + \sum_{j=1}^m \frac{b_j}{1 - \eta_j B} x_t = a_t \quad \text{且当 } i \neq j \text{ 时 } \eta_i \neq \eta_j \quad (3-19)$$

$$\varphi_1(B)x_t + \sum_{j=1}^m b_j \left( \sum_{i=0}^{\infty} \eta_j^i B^i \right) x_t = a_t \quad \text{当 } |\eta_j| < 1 \text{ 时} \quad (3-20)$$

交换求和符号的位置，有

$$\varphi_1(B)x_t + \sum_{i=0}^{\infty} \left( \sum_{j=1}^m b_j \eta_j^i \right) B^i x_t = a_t \quad (3-21)$$

令  $\sum_{j=1}^m b_j \eta_j^i = -I_i$ ，有

$$\varphi_1(B)x_t + \sum_{i=0}^{\infty} (-I_i) B^i x_t = a_t \quad (3-22)$$

由于  $\varphi_1(B)$  是  $B$  的多项式，，因此可与公式(3-22)左边第二项相加，其结果如下

$$\left[ \sum_{i=0}^{\infty} (-I_i) B^i \right] x_t = a_t \quad (3-23)$$

且当  $i=0$  时,  $-I_0 \equiv 1$ , 这表明当  $|\eta_i| < 1$  时,  $\sum_{i=0}^{\infty} (-I_i)B'$  收敛与  $\frac{\varphi(B)}{\theta(B)}$ , 将公式(3-23)中  $i=0$  的一项保留在等式左边, 其余的部分移到等式右边, 并考虑到  $B$  算子的定义, 有

$$x_t = \sum_{i=1}^{\infty} I_i x_{t-i} + a_t \quad (3-25)$$

比较公式(3-24)和(3-25), 可知: 两者在形式上是完全一致, 而公式(3-25)是一个  $AR(\infty)$  模型, 证明过程表明一个有限阶的  $ARMA(n, m)$  模型等价于一个无限高阶的  $AR(\infty)$  模型。因此在工程应用中常用足够高阶的  $AR$  模型来取代  $ARMA$  模型, 以避免估计  $ARMA$  模型的困难。

### 3.2.3 AR 参数模型的建立

$AR(n)$  建模实际上是通过确定合适的模型以拟合所监测的数据序列。建模可从以下两方面进行: 1  $AR(n)$  阶数  $n$  的确定, 2 用最小二乘递推估计法进行参数估计。

#### 1. $AR(n)$ 阶数 $n$ 的确定

对于随机时序而言, 其自协方差函数  $R_k$  代表了数据序列  $\{x_t\}$  的所有二阶矩函数, 因此它可以在最大限度内描述  $\{x_t\}$  的统计特性, 在对数据序列进行建模时也可采用自协方差函数  $R_k$ 。 $AR$  模型可定义为<sup>[9]</sup>

$$x_t = \sum_{i=1}^n \varphi_i x_{t-i} + a_t \quad (3-26)$$

将等式两边均乘以  $x_{t-k}$  并取数学期望, 则有

$$E(x_t x_{t-k}) = \sum_{i=1}^n \varphi_i E(x_{t-i} x_{t-k}) + E(a_t x_{t-k}) \quad (3-27)$$

其中:  $E(x_t x_{t-k}) = R_k$ ,  $E(x_{t-i} x_{t-k}) = R_{k-i}$ ,  $E(a_t x_{t-k}) = \sigma_a^2 \delta_k$ , 因此可对公式(3-27)进行变换, 则有

$$R_k = \sum_{i=1}^n \varphi_i R_{k-i} + \sigma_a^2 \delta_k \quad (3-28)$$

当  $k \neq 0$  时, 有:  $R_k = \sum_{i=1}^n \varphi_i R_{k-i}$ 。

设现有一时间序列  $\{x_t\}$ , 它共有  $N$  个数据, 若对应的模型为  $AR(n)$ , 将时间序列  $\{x_t\}$  代入公式(3-26)中, 可得到一组线性方程<sup>[9]</sup>

$$\begin{aligned}
x_{n+1} &= \varphi_1 x_n + \varphi_2 x_{n-1} + \cdots + \varphi_n x_1 + a_{n+1} \\
x_{n+2} &= \varphi_1 x_{n+1} + \varphi_2 x_n + \cdots + \varphi_n x_2 + a_{n+2} \\
&\dots\dots
\end{aligned}
\tag{3-29}$$

$$\begin{aligned}
&x_N = \varphi_1 x_{N-1} + \varphi_2 x_{N-2} + \cdots + \varphi_n x_{N-n} + a_N \\
\text{设有: } y &= \begin{bmatrix} x_{n+1} & x_{n+2} & \cdots & x_N \end{bmatrix}^T, \varphi = \begin{bmatrix} \varphi_1 & \varphi_2 & \cdots & \varphi_n \end{bmatrix}^T, a = \begin{bmatrix} a_{n+1} & a_{n+2} & \cdots & a_N \end{bmatrix}^T, \\
x &= \begin{bmatrix} x_n & x_{n-1} & \cdots & x_1 \\ x_{n+1} & x_n & \cdots & x_2 \\ & \cdots & & \\ x_{N-1} & x_{N-2} & \cdots & x_{N-n} \end{bmatrix}
\end{aligned}$$

则可用矩阵的形式表示为

$$y = x\varphi + a \tag{3-30}$$

若求其自协相关系数  $R_k$ , 则在公式(3-30)两边均乘以  $x^T$ , 再进行变换后,  $\varphi$  矩阵参数的最小二乘估计为

$$\hat{\varphi} = (x^T x)^{-1} x^T y \tag{3-31}$$

由于最小二乘估计法非常简单, 不仅估计的精度高, 而且, 当已知  $\{x_i\}$  分布时, 还能估计出真值的置信区间。但其计算次数正比于  $Nn^2$ , 随着模型阶数的增高, 计算量将会剧增; 此外, 当  $x^T x$  中元素的绝对值相差较大时, 容易出现病态。对其改进的方法主要是采用自协相关系数, 如

$$\frac{x^T x}{N-n} = \begin{bmatrix} R_0 & R_1 & R_2 & \cdots & R_{n-1} \\ R_1 & R_0 & R_1 & \cdots & R_n \\ & \cdots & & & \\ R_{n-1} & R_{n-2} & R_{n-3} & \cdots & R_0 \end{bmatrix} \tag{3-32}$$

同理也可对公式(3-31)中矩阵  $x^T y$  展开

$$\frac{x^T y}{N-n} = \frac{1}{N-n} \begin{bmatrix} \sum_{t=n}^{N-1} x_t x_{t+1} \\ \sum_{t=n-1}^{N-2} x_t x_{t+1} \\ \vdots \\ \sum_{t=1}^{N-n} x_t x_{t+1} \end{bmatrix} = \frac{\sum_{t=n+1}^N \begin{bmatrix} x_{t-1} x_t \\ x_{t-2} x_t \\ \vdots \\ x_{t-n} x_t \end{bmatrix}}{N-n} = \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_n \end{bmatrix} \quad (3-33)$$

因此, 参数  $\varphi$  的最小二乘估计可写成

$$\hat{\varphi} = \begin{bmatrix} R_0 & R_1 & R_2 & \cdots & R_{n-1} \\ R_1 & R_2 & R_3 & \cdots & R_n \\ & & \cdots & & \\ R_{n-1} & R_{n-2} & R_{n-3} & \cdots & R_0 \end{bmatrix}^{-1} \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_n \end{bmatrix} \quad (3-34)$$

根据 BOX 建模方案, 从  $n=2$  开始对  $\{x_t\}$  拟合 AR( $n$ ) 模型, 由公式 (3-31) 先求出其参数  $\varphi$  的估计值  $\hat{\varphi}$ , 将  $\hat{\varphi}$  代替公式 (3-26) 中  $\varphi$ , 求出拟合的数据序列  $\{\hat{x}_t\}$ 。对数据序列  $\{x_t\}$  和  $\{\hat{x}_t\}$  进行相减处理, 求出对应的均值与方差。根据 F 准则检验拟合模型的适用程度。若模型不适用, 则令  $n=n+1$  再继续拟合, 直到模型适用。

对于不同  $n$ , 可应用信息准则 (An Information Criterion) 对其模型阶数进行确认。信息准则 (AIC) 函数的定义如下

$$AIC(P) = -2 \ln L + 2p \quad (3-35)$$

其中  $p$  为模型的阶数,  $L$  为时序  $\{x_t\}$  的似然函数, 对 (3-35) 式进行进一步的简化有

$$AIC(P) = -N \ln \sigma_a^2 + 2p \quad (3-36)$$

公式 (3-35) 中,  $N$  为给定数据的长度,  $\sigma_a^2$  残差的方差 (即上面 BOX 建模时求出的方差)。对于不同的阶数  $n$  及其对应的残差的方差可代入公式 (3-36), 得到不同的 AIC 值, 其最小值所对应的阶数为 AR 模型阶数。

## 2. 运用最小二乘递推估计法进行参数估计

当观测样本  $\{x_t\}$  的样本长度  $N$  增加, 参数估计精度提高, 其结果是计算机内存空间增大, 计算速度下降, 特别是对于工程系统的实时监测、故障诊断、在线控制, 普通算法难以达到要求, 因此, 对模型参数的估计应采用实

时估计的算法。实时估计算法通常具有以下优点：

- 第一，实时估计算法每步计算量小，占用机器内存小；
- 第二，可以在线建模，能及时对设备运行状况进行监测；
- 第三，实时估计算法中不断地引入新的系统信息，具有追踪时变参数地能力；

第四，能不断修正系统参数模型，从可实现用低阶模型去逼近高阶模型，用 AR 模型逼近 ARMA 模型，用线性模型逼近非线性模型，为复杂系统的在线监测提供了可能。

论文基于上述原因，采用最小二乘递推最小二乘算法作为实时预测的算法，具体的过程如下所示：

现将公式(3-30)中矩阵的形式改写为

$$\left. \begin{aligned} Y_N &= y \\ \varphi_N &= \varphi \\ X_N &= x = \begin{bmatrix} x_{(1)} & x_{(2)} & \cdots & x_{(N-n)} \end{bmatrix}^T \end{aligned} \right\} \quad (3-37)$$

则 AR 模型参数的最小二乘估计式为

$$\varphi_N = (X_N^T X_N)^{-1} X_N^T Y_N \quad (3-38)$$

将各矩阵与向量加以下标 N，指基于 N 个观测数据  $\{x_t\} (t=1, 2, \dots, N)$ 。而公式(3-38)的矩阵  $X_N$  中  $x_{(i)}$  是一个 n 维行向量，若令  $P_N = (X_N^T X_N)^{-1}$ ，则可将公式(3-38)改写为

$$\varphi_N = P_N X_N^T Y_N \quad (3-39)$$

设原有的观测时序为  $\{x_t\} (t=1, 2, \dots, N)$ ，根据公式(3-39)可求出模型参数  $\varphi_N$ ，若又测得一个新数据  $x_{N+1}$ ，这时可将  $x_{N+1}$  加入原序列，得到  $\{x_t\} (t=1, 2, \dots, N, N+1)$ ，再进行参数估计。根据上述推导，基于 N+1 个数据的参数估计为<sup>[9]</sup>

$$\left. \begin{aligned} \varphi_{N+1} &= P_{N+1} X_{N+1}^T Y_{N+1} \\ P_{N+1} &= (X_{N+1}^T X_{N+1})^{-1} \end{aligned} \right\} \quad (3-40)$$

$$\text{其中, } X_{N+1} = \begin{bmatrix} X_N \\ x_{(N-n+1)} \end{bmatrix}, Y_{N+1} = \begin{bmatrix} Y_N \\ x_{n+1} \end{bmatrix}$$

根据分块矩阵的乘法有

$$\left. \begin{aligned} X_{N+1}^T Y_{N+1} &= X_N^T Y_N + x_{(N-n+1)}^T x_{N+1} \\ P_{N+1} &= (P_N^{-1} + x_{(N-n+1)}^T x_{N+1})^{-1} = \left( I - P_N \frac{x_{(N-n+1)}^T x_{(N-n+1)}}{1 + x_{(N-n+1)}^T P_N x_{(N-n+1)}} \right) P_N \end{aligned} \right\} \quad (3-41)$$

其中,  $I$  为  $n$  阶单位阵。由公式(3-41)可知对模型参数进行新的估计  $\varphi_{N+1}$  时, 所用到矩阵  $X_{N+1}^T$ ,  $Y_{N+1}$ ,  $P_{N+1}$  均可由原矩阵  $X_N^T$ ,  $Y_N$ ,  $P_N$  递推估计, 而且求  $P_{N+1}$  时, 只需进行矩阵乘法运算, 其算法较最小二乘法简单。将公式(3-41)代入公式(3-40)并对其进行整理可得

$$\left. \begin{aligned} \varphi_{N+1} &= \varphi_N + K_{N+1} (x_{N+1} - x_{(N-n+1)} \varphi_N) \\ K_{N+1} &= \frac{1}{1 + x_{(N-n+1)}^T P_N x_{(N-n+1)}} P_N x_{(N-n+1)}^T \end{aligned} \right\} \quad (3-42)$$

公式(3-42)具有深刻含义, 它表示模型参数的新估计  $\varphi_{N+1}$  等于原参数估计  $\varphi_N$  与校正项之和。其中校正项为:  $K_{N+1} (x_{N+1} - x_{(N-n+1)} \varphi_N)$ , 它是新的数据  $x_{N+1}$  与新数据的估计值  $x_{(N-n+1)} \varphi_N$  之差的加权处理, 而加权系数为  $K_{N+1}$ , 因此该递推估计算法具有普遍的意义。

由于系统的过去行为也不一定能准确地反映系统的系统当前的特征, 递推最小二乘估计法引入了校正系数, 其本质就是既重视了原来  $N$  个数据所提供的信息, 即估计  $\varphi_N$ , 也重视新的信息  $x_{N+1}$ 。而新的信息又可能具有一定的随机性, 故估计式校正项中有  $x_{(N-n+1)} \varphi_N$ , 这是按原估计  $\varphi_N$  推算出的新数据的预测值。因此该预测值既包含了原来  $N$  个数据所蕴含的历史信息, 又包含了新数据的信息, 两者相结合, 从而可更好反映了系统的行为规律。

在递推最小二乘法中, 由于新数据的不断增加, 当数据长度  $N$  超过某一值后, 所求得的模型参数很难真实地反映系统的当前状况, 即导致数据饱和现象的出现。产生这一现象的原因在于: 矩阵  $P_N$ ,  $P_{N+1}$  一般为对称正定矩阵, 可从理论上证明, 当  $P_N \geq P_{N+1}$  时, 随着递推次数的增加,  $P_{N+1}$  将越来越小, 最后可能趋向零矩阵, 不但不能起到校正的作用, 反而导致模型参数的估计偏离真值。此外系统中新的信息和历史信息对其影响也不尽相同, 通常新的信息更能反映系统(被监测设备)当前行为特征, 因此, 更应侧重新数据的作用。

基于上述原因, 引入了遗忘因子  $\alpha$  对过去的数据进行加权处理, 以削弱

历史数据的影响，增强新数据的作用。分别对公式(3-41)和公式(3-42)引入遗忘因子，其模型参数的递推公式为<sup>[9]</sup>

$$\left. \begin{aligned} \varphi_{N+1} &= \varphi_N + K_{N+1} (x_{N+1} - x_{(N-n+1)} \varphi_N) \\ K_{N+1} &= \frac{1}{\alpha + x_{(N-n+1)}^T P_N x_{(N-n+1)}} P_N x_{(N-n+1)}^T \\ P_{N+1} &= \frac{1}{\alpha} \left( I - P_N \frac{x_{(N-n+1)}^T x_{(N-n+1)}}{\alpha + x_{(N-n+1)}^T P_N x_{(N-n+1)}} \right) P_N \end{aligned} \right\} \quad (3-43)$$

遗忘因子可根据模型的实际情况进行选择，当历史数据作用较小、或系统参数随时变动较大时， $\alpha$ 可取较小的值；当历史数据较大时， $\alpha$ 可取较大的值，通常情况下 $\alpha$ 的取值在 $[0.95 \ 0.99]$ 范围内。其建模的流程图如图 3-1 所示。

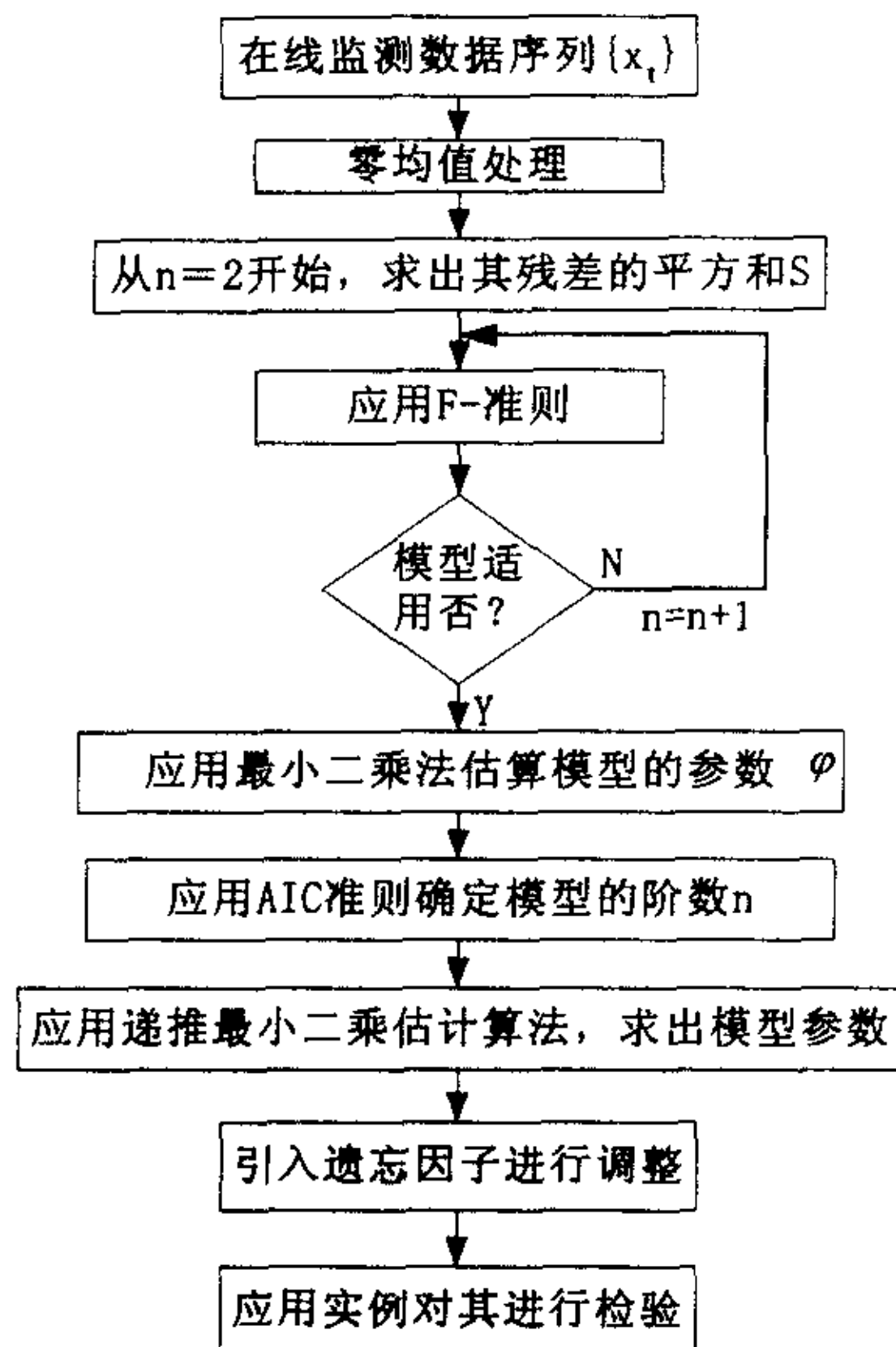


图 3-1 时序预测建模流程图



### 3.3 实例分析

在上节介绍了实时估计递推最小二乘算法。从中可以看出该实时估计算法不仅包含了系统新数据蕴含的新信息，而且还引入了遗忘因子，以减少历史数据的影响，增强了新数据的作用；同时还能对系统模型参数进行实时调整。论文以湖南某地 220kV 变压器 A 相套管泄漏电流的在线监测数据为例，采用该算法来预测套管泄漏电流。

论文选取 300 个样本(套管泄漏电流值)进行零均值处理和正规化处理，其具体的方法是：一阶差分处理。为了便于分析，对处理前、处理后样本比较，其图形如图 3-2 所示。

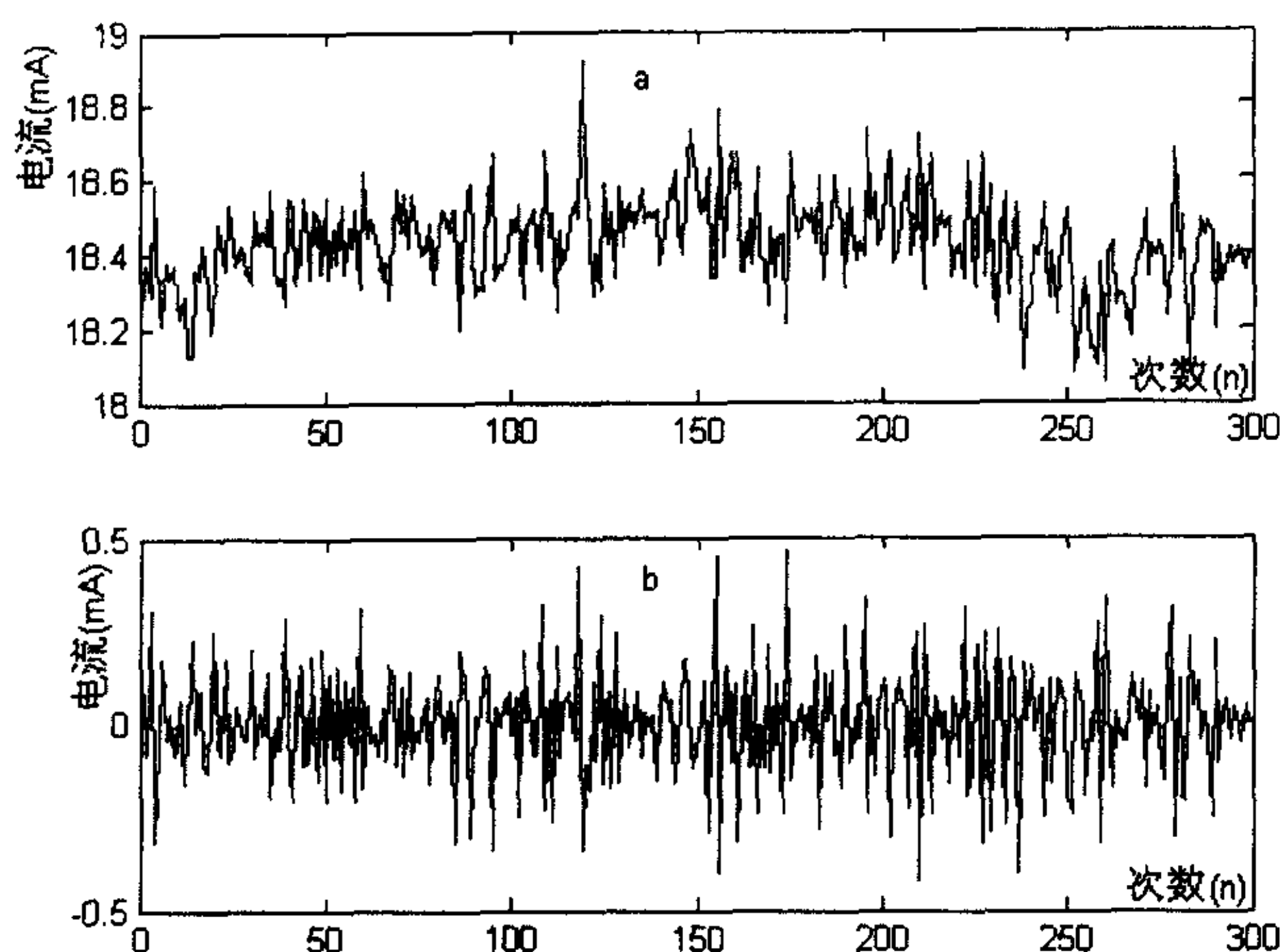


图 3-2 套管泄漏电流处理前后的图

(a) 样本原始据图 (b) 一阶差分处理后图

由图 3-2 可知，样本经一阶差分处理后，基本上能满足正态、零均值的要求，为以后模型参数的求解打下了基础。

从  $n=2$  开始，在 100 个样本作为训练样本，并求出其残差的平方和  $S$ ，应用 F-准则以确定  $n$  的取值范围。对于上述不同的  $n$ ，应用最小二乘法求出

其对应的模型参数  $\varphi$ ，并以此为依据求出其对应方差。其具体数据如表 3-1 所示。

表 3-1 在  $n$  不同的情况下对应模型参数及其相关性能的比较

| 阶数  | $\varphi_1$ | $\varphi_2$ | $\varphi_3$ | $\varphi_4$ | $\varphi_5$ | 最大误差 (%) | 方差     |
|-----|-------------|-------------|-------------|-------------|-------------|----------|--------|
| 2 阶 | 0.8100      | 0.1901      |             |             |             | 1.5370   | 0.1286 |
| 3 阶 | -0.8693     | -2.7710     | 4.6458      |             |             | 1.9374   | 0.1276 |
| 4 阶 | 0.8286      | 0.1666      | -0.2772     | 0.2820      |             | 1.4023   | 0.1291 |
| 5 阶 | 0.8825      | 0.1168      | -0.2296     | 0.4225      | -0.1920     | 1.6102   | 0.1307 |

表 3-1 列举了  $n$  在不同情况下系统的模型参数及其相关的性能。以此为依据应用 AIC 准则来确定  $n$  取值。将表 3-1 中模型的阶数和方差代入公式 (3-36) 有

$$AIC(2)=209.1273$$

$$AIC(3)=211.8854$$

$$AIC(4)=212.6922$$

$$AIC(5)=213.4878$$

而  $AIC(n)$  的最小值就是  $FPE(n)$  最小值，而当  $FPE(n)$  取得最小值时该模型为适用模型，即  $n=2$ 。

在确定了模型参数的阶数  $n=2$  后，选取了总体样本中后 200 个样本作为预测模型检验样本  $n'=200$ 。运用最小二乘法对训练样本进行训练，求出该套管泄漏电流的预测模型  $AR(2)$  为

$$x_{t+2} = 0.8100x_{t+1} + 0.1901x_t \quad (t=1, 2, \dots, 98) \quad (3-44)$$

由于在线数据需要实时预测、实时校正，所以根据检验样本运用最小二乘递推估计的思想，并引入遗忘因子  $\alpha$  对其预测模型进行递推、实时调整，其通用的表达式为

$$x_{t+2} = \varphi'_1 x_{t+1} + \varphi'_2 x_t \quad (t=99, 100 \dots 298) \quad (3-45)$$

$x_{t+2}$  表示第  $t+2$  个样本的值， $\varphi'$  表示经过第  $t-98$  次递推估计后系统的模型参数。令递推的次数为  $L$ ，分别对数据进行短期单步递推预测，并求出经过  $L$  次递推 ( $L=1, 20, 50, 80, 110, 140, 170, 190$ ) 后系统的参数  $\varphi$  如表 3-2 所示。

表 3-2 短期单步递推预测的模型参数

| 次数<br>$\varphi$ | 1       | 20      | 50      | 80      | 110     | 140     | 170     | 190     |
|-----------------|---------|---------|---------|---------|---------|---------|---------|---------|
| $\varphi_1$     | 0.81002 | 0.81139 | 0.81856 | 0.84751 | 0.85904 | 0.86088 | 0.89634 | 0.89571 |
| $\varphi_2$     | 0.19007 | 0.18737 | 0.18152 | 0.15254 | 0.14100 | 0.13914 | 0.10367 | 0.10430 |

由表 3-2 可知, 经过递推调整后系统的模型参数  $\varphi_1$ 、 $\varphi_2$  均有所改变。为了验证算法的准确性, 对预测模型检验样本  $n'$ 、预测数据及两者间误差进行了比较, 其结果如图 3-3 所示。

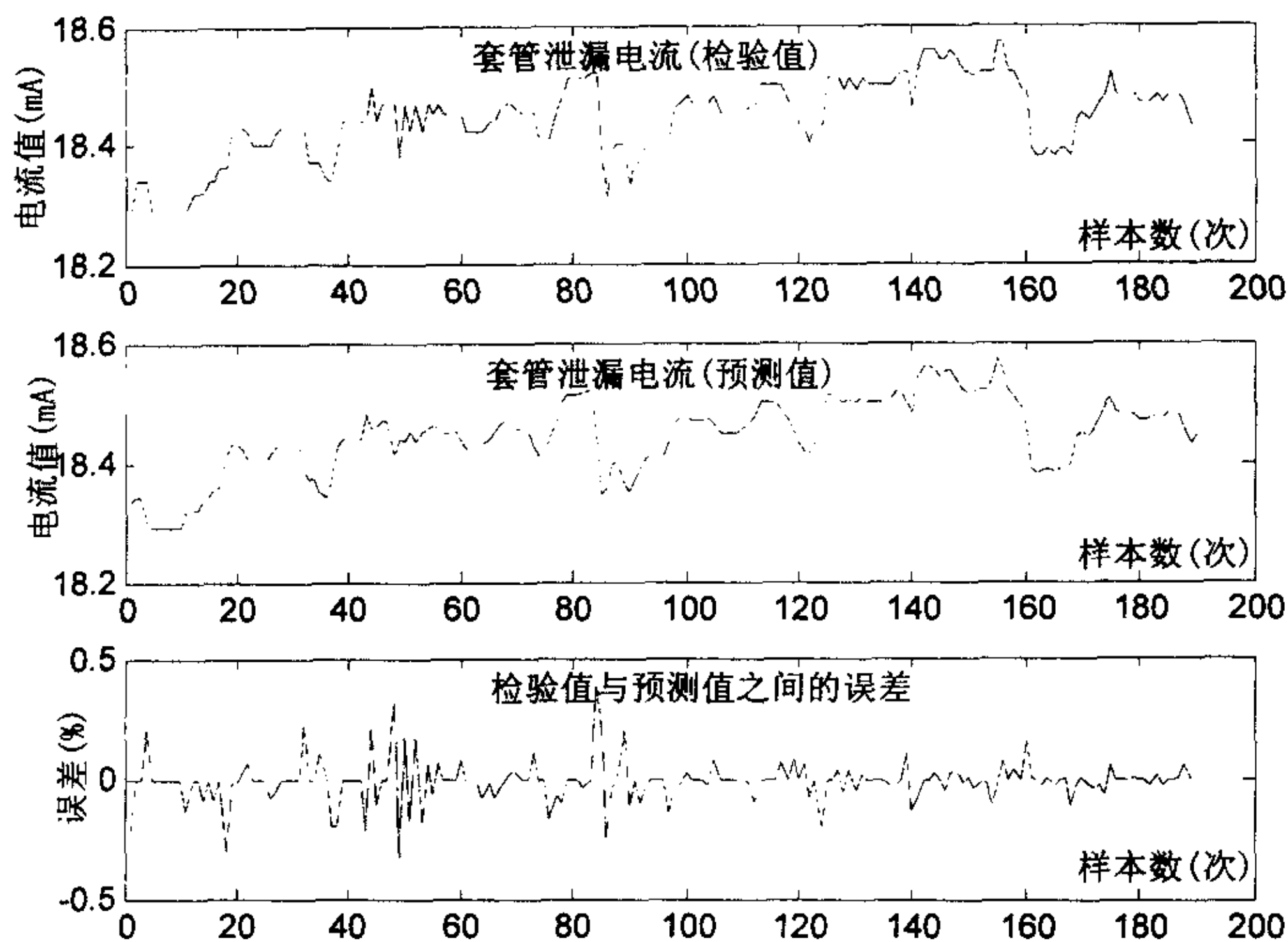


图 3-3 原始值、短期单步递推预测预测值及误差的比较

由图 3-3 可知, 原始值与短期单步递推预测值的趋势是一致的, 且误差很小, 基本上在  $[-0.5\%, +0.5\%]$  的范围内, 因此证明算法对短期内的数据预测的结果是可靠的、准确的。

为了对算法的实用进行推广, 论文以预测值作为新的样本, 通过遗忘因子对在线数据进行预测, 其经过多次递推后的模型参数如表 3-3 所示。

表 3-3 长期递推预测的模型参数

| 次数<br>$\varphi$ | 1       | 20      | 50      | 80      | 110     | 140     | 170     | 190     |
|-----------------|---------|---------|---------|---------|---------|---------|---------|---------|
| $\varphi_1$     | 0.81184 | 0.80783 | 0.80543 | 0.80723 | 0.80614 | 0.80897 | 0.81640 | 0.81327 |
| $\varphi_2$     | 0.18864 | 0.19639 | 0.20102 | 0.19747 | 0.19965 | 0.19378 | 0.18059 | 0.18592 |

对比表 3-2 可知, 经过相同次递推后, 短期单步递推预测的模型参数与长期递推的模型参数有明显的差异, 主要是在长期递推预测模型将预测值作为了新的样本。其对应检验值、预测值和误差的图形如图 3-4 所示。

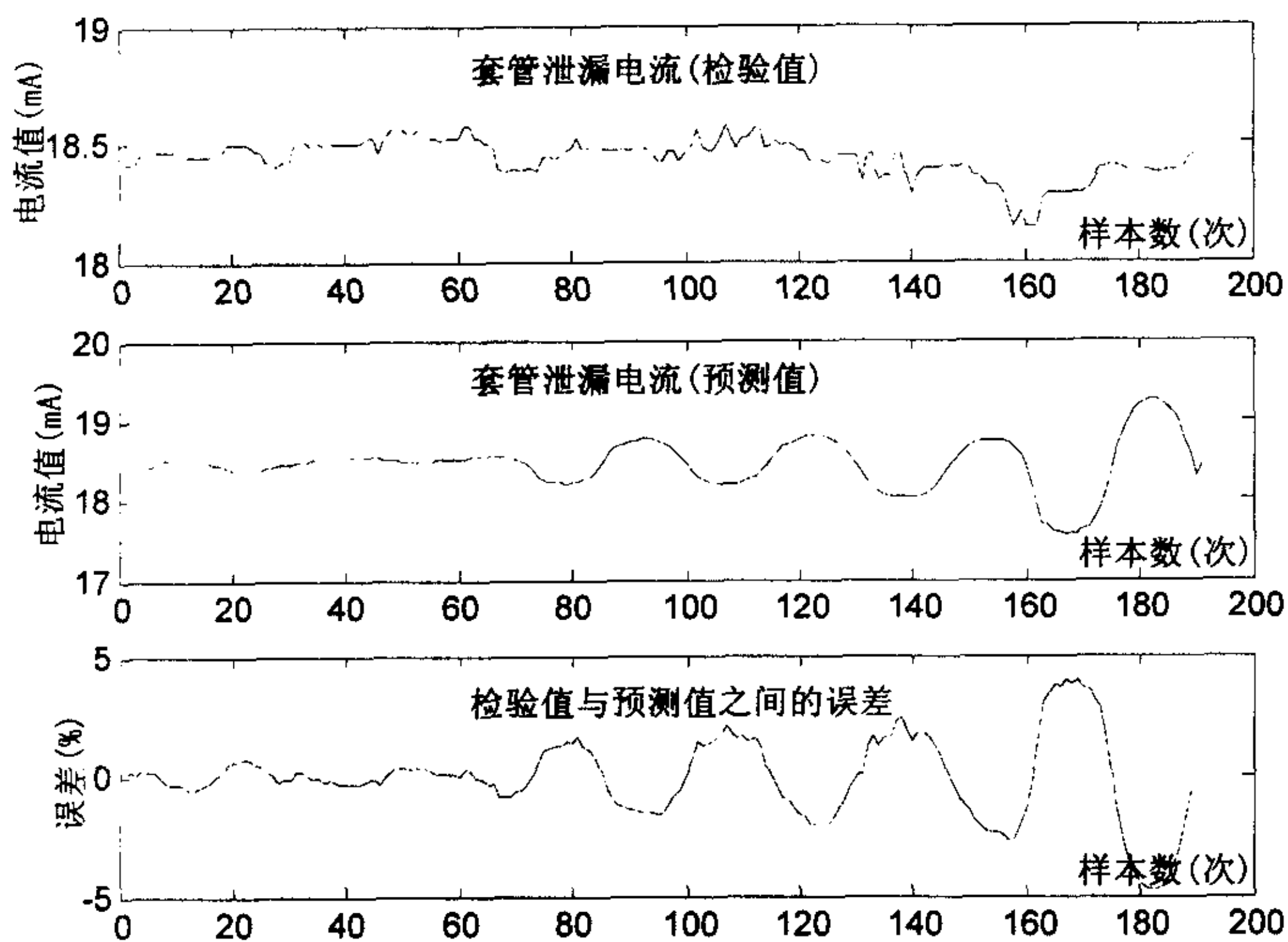


图 3-4 原始值、长期递推预测值及误差的比较

由图 3-4 可知, 当样本数小于 100 时, 套管泄漏电流检验值和预测值的趋势是基本一致的, 而且他们之间的误差很小, 基本上在  $[-2\%, +2\%]$  的范围内。但随着样本数量的增大, 预测值的误差逐渐增大, 而且检验值和预测值的偏差越来越大, 趋势的偏差也逐步增大。对于数据预测而言, 预测值与数据的真值之间必定存在误差, 若以预测值作为新的样本进行预测, 势必

造成误差的叠加使预测值偏离真值。数据预测的范围取决于训练样本的容量和算法的精度, 训练样本的容量越大, 算法的精度越高, 预测的效果越好, 预测的范围越大。论文通过引入遗忘因子, 提高了算法的精度, 拓宽了数据预测的范围。为了验证算法的递推性, 将该算法编写的程序在一台计算机(主频 1.6GHz, 外频 400MHz, 内存 256MB)上运行, 进行一次预测的时间为  $573.684 \mu s$ 。因此, 算法即可以保证预测的准确性又能满足实时监测的要求; 并且对于新的数据和历史数据进行调整, 能及时地反映被监测设备当前的真实状态。因此, 可证明基于时间序列的递推最小二乘估计算法可以用于在线数据的实时预测, 且预测的精度高、效果好。

### 3.4 讨论

对于零均值平稳时间序列或可零均化的平稳时间序列, 可以用 3.2 节的方法建立  $AR(n)$  进行拟合, 但以上方法均是假设“时间序列是平稳的”, 因此在建模前首先应检验其是否具有平稳性<sup>[20]</sup>。在实际问题中, 所考虑的时间序列并不都是平稳序列, 因此不能对其直接用  $AR(n)$  模型或  $ARMA(n,m)$  模型去拟合, 这种时间序列称为非平稳时间序列。而非平稳时间序列的情况相当复杂, 下面主要对具有趋势项的时间序列进行讨论。

在平稳时间序列中, 时间序列的均值为常数(零均值为特例, 但总可以通过零化处理得到零均值序列), 表现在时序的时间历程中: 则数据在某一水平线附近随机波动。而非时间序列的情况则与这种情况相反, 其均值为时序在时间上的函数, 表现为: 数据在某一趋向性曲线附近的随机波动。这种曲线有可能是线性函数、幂函数、指数函数、余弦函数或多种函数叠加。

非平稳时间序列的建模方法可分为两类, 一类称为直接剔除法, 它是通过某些处理方法(主要是差分)将确定性部分从时间序列中直接剔除, 如季节性  $ARIMA$ , 但这种方法仍然可以用  $AR(n)$  或  $ARMA(n,m)$  建模。另一类称为趋势项提取法, 它是从时间序列中提取出确定性的部分, 将确定性部分用明确的函数关系表达, 再对剩余的平稳时序按  $AR(n)$  或  $ARMA(n,m)$  的方法建模, 最后将确定性函数与平稳时序的函数组合, 得到非平稳时序模型, 如提取趋势项的逐步回归法。

1. 季节模型。该模型利用差分运算(称为季节差分)剔除非平稳时间序列周期性变动的趋势。模型的典型分解式<sup>[9][20]</sup>

$$X_t = m_t + S_t + Y_t \quad (3-46)$$

其中  $m_t$  是趋势项,  $S_t$  是季节项,  $Y_t$  是随机噪声。在许多实际问题中, 往往可根据系统的物理特性或数据的物理规律得知时间序列的周期。其具体作法是: 将某一时刻的测量数据  $x_t$  减去前一个周期相应时刻的测量数据  $x_{t-T}$ , 所得到的序列  $\{x_t - x_{t-T}\}$  将可能不再具有周期性趋向而成为平稳时间序列, 然后对其按 3.2 节方法建模。有时进行一次季节性差分尚不能得到平稳时序, 可进行多次差分, 直到数据序列平稳为止。对于在线监测的变电设备而言, 若以月来计算, 则数据序列的周期为 12, 对其进行趋势差分可得到数据序列  $\{x_t - x_{t-12}\}$ , 如果为平稳序列, 则可对其按 3.2 节方法建模; 若仍然不为平稳序列, 则可进行多次差分, 直到数据平稳。

根据上述思想, 记时间序列的周期为  $T$ , 则季节差分算子为:  $\nabla_T = (1 - B^T)$ , 非平稳时间序列  $\{x_t\}$  ( $t=1, 2, \dots, N$ ) 进行季节差分, 并对其进行建模, 其表达式为<sup>[9]</sup>

$$\phi(B^T) \nabla_T^D x_t = \theta(B^T) \varepsilon_t \quad (3-47)$$

若公式 (3-47)  $\varepsilon_t$  不一定为白噪声, 可能具有二项式趋向, 则应对其进行差分处理, 其表达式可改写为

$$\begin{aligned} \varphi(B) \nabla^d \varepsilon_t &= \psi(B) \alpha_t \\ \varphi(B) \phi(B^T) \nabla^d \nabla_T^D x_t &= \psi(B) \theta(B^T) \alpha_t \end{aligned} \quad (3-48)$$

其中  $\alpha_t$  为白噪声。

2. 提取趋势项的逐步回归法。非平稳时间序列是由确定性部分和平稳随机部分的叠加。若令确定性部分的函数为  $d_t$ , 平稳随机部分的函数为  $\varepsilon_t$ , 则其组合模型为<sup>[9]</sup>

$$x_t = d_t + \varepsilon_t \quad (t=1, 2, \dots, N) \quad (3-49)$$

在非平稳时间序列建模中, 由于平稳随机部分的函数为  $\varepsilon_t$ , 可按 3.2 节方法建模, 因此关键性的问题在于如何求解  $d_t$ 。由于趋势项  $d_t$  的表现形式存在差异, 可采用不同提取方法, 其中最常用的为逐步回归法。该方法的主要思想是: 根据趋势项  $d_t$  与时间  $t$  的关系, 建立组合模型, 如线性趋势组合模型、指数趋势组合模型、周期性趋势组合模型等, 采用回归估计的方法求出对应模型参数。

### 3.5 小结

本章采用时间序列的思想对在线数据进行实时预测, 首先介绍了时间序



列的数学基础，然后采用递推的思想对时序预测进行了较为深入的研究，最后进行实例分析和讨论。

1 预测模型的建立。由于 ARMA 模型参数较多，而且求解复杂，可将有限阶的 ARMA 模型等价的转化为无限高阶的 AR 模型，因此工程应用中常选用 AR 模型。先采用最小二乘算法和 F-准则，确定 AR 模型  $n$  的范围；再运用 CIA 准则确定  $n$  的取值；最后用递推最小二乘算法估计算法对时序模型的参数进行估计，并引入了遗忘因子对当前数据和历史数据的影响进行了调整。

2 实例分析和讨论。论文以 220kV 变压器 A 相套管泄漏电流为例，选取了 300 个接地电流值作为样本，其中前 100 样本为训练样本，后 200 为验证样本，对其进行建模和数据预测，其结果为：预测值曲线趋势走向与检验样本曲线趋势走向是一致的，且误差很小，而数据预测的范围在很大程度上取决于训练样本的容量和算法的精度。上述建模的对象为平稳时序或可转化为平稳时序的序列，而对于非平稳时序则情况相当复杂，论文主要对具有趋势项的非平稳时序进行了讨论，其建模的方法可分为直接剔除和趋势项提取两类，然后根据其特征可分别求解。

## 4 基于径向基函数的支撑向量分类器设计

在线数据的综合处理包含预处理、预测、分类等，而数据的预处理和数据的预测都是为设备的诊断和特征量的分类做准备的。一般来说，设备状态的模式识别主要包含两方面的内容：第一，有故障时，能够及时识别，并力争寻找故障的所在，分析故障原因；第二，能进行早期的故障诊断，即在故障发生之前能及时地把故障将要发生的征兆诊断出来，以便采用相应的措施，达到“防患于未然”的目的。在设备状态的模式识别时，系统的不同状态就表现为不同模式，而对其进行分类并加以识别，即模式分类。在进行模式识别和故障分类之前首先要确定参考模式，然后在将当前模式与参考模式比较，然后才能确定状态的类别。

本章研究内容是如何设计一个基于径向基函数的支撑向量分类器，以实现对被监测设备状态的模式分类。论文将从以下几方面展开：统计学习理论和支撑向量机理论的阐述、径向基函数神经网络的软竞争学习算法、分类器的构造及实例求解等。

美国贝尔实验室利用美国邮政标准手写数字库进行分类对比实验，其中训练集共 7300 个样本，测试集有 2000 个样本，用人工和几种传统的方法进行模式分类，其识别的结果如表 4-1 所示<sup>[13]</sup>。

表 4-1 不同分类方法对美国邮政手写体数字库的识别结果

| 分类器             | 误判率   |
|-----------------|-------|
| 人工分类            | 2.5%  |
| 决策树             | 16.2% |
| 两层神经网络          | 5.9%  |
| LeNet1          | 5.1%  |
| 基于径向基函数的支撑向量分类器 | 4.1%  |

由表 4-1 可知，人工分类的误判率最低，不能自动识别，效率低；基于径向基函数的支撑向量分类器的识别误判率比决策树和神经网络的正判率高，因此表明了用基于径向基函数的支撑向量分类器进行模式识别、状态分类是可行的。



## 4.1 支撑向量机理论的简述

传统的统计模式的方法都是在样本数目足够多的前提下进行研究的, 所提出的各种方法只有在样本数趋于无穷大时期性能才能得到理论的保证。而在多数情况下, 样本的数目通常时有限的, 因此很多方法都难以取得好的效果。而统计学习理论是一种专门的小样本理论, 支撑向量机是它的一种新的模式识别方法, 它能够较好地解决小样本的机器学习问题。

### 4.1.1 机器学习<sup>[13]</sup>

机器学习的可以用图 4-1 表示, 其中, 系统 S 是我们的研究对象, 它在给定一定输入  $x$  的条件下可得到输出  $y$ , LM 是我们所求学习机, 输出为  $\hat{y}$ 。机器学习的目的是: 根据给定的已知训练样本, 求取系统输入输出之间依赖关系的估计, 使它能够对未知输出做出尽可能准确的预测, 如图 4-1 所示。

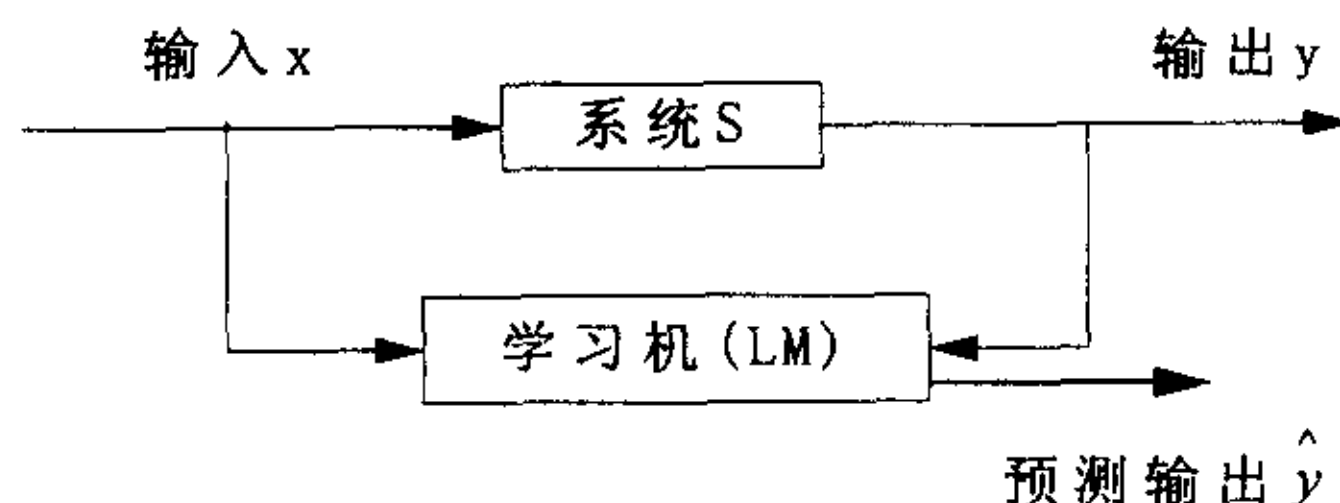


图 4-1 机器学习模型

机器学习的问题可以形式化地表示为: 已知变量  $y$  与输入  $x$  之间的存在一定的未知关系, 即存在一个未知联合概率分布  $F(x, y)$ , 根据多个独立同分布观测样本, 在一组函数  $\{f(x, \omega)\}$  中求一个最优的函数  $f(x, \omega_0)$  使其期望风险最小。

$$R(\omega) = \int L(y, f(x, \omega)) dF(x, y) \quad (4-1)$$

其中,  $\{f(x, \omega)\}$  为预测函数集,  $L(y, f(x, \omega))$  为由于  $f(x, \omega)$  对  $y$  进行预测而造成损失。通常情况下  $L(y, f(x, \omega))$  为  $\{-1, +1\}$  的二值函数, 其基本定义如下

$$L(y, f(x, \omega)) = \begin{cases} 0 & \text{if } y = f(x, \omega) \\ 1 & \text{if } y \neq f(x, \omega) \end{cases} \quad (4-2)$$

由于在实际应用中我们只能利用已知样本信息, 因此无法直接求解。根据大数定理的思想, 可用数学平均代替数学期望, 其表达式可定义为

$$R_{emp}(\omega) = \frac{1}{n} \sum_{i=1}^n L(y_i, f(x_i, \omega)) \quad (4-3)$$

由于  $R_{emp}(\omega)$  是用已知的训练样本(即经验数据)定义的, 因此称为经验风险。通过对参数  $\omega$  求经验风险  $R_{emp}(\omega)$  的最小值来代替求期望风险  $R(\omega)$  的最小值, 即所谓的经验风险最小化原则(Empirical Risk Minimization 简称 ERM)。

#### 4.1.2 支撑向量机的理论

支撑向量机(Support Vector Machine, 简称 SVM)是在统计学习理论的基础上发展出的一种新的模式识别的方法, 在解决小样本、非线性及高维模式识别问题中表现出许多特有的优势, 并能够推广到函数拟合和故障分类等机器学习的领域<sup>[11]</sup>。支撑向量机的基本思想可以概括为: 首先通过非线性变换将输入空间变换到一个高维空间, 然后在这个空间中求取最优线性分类面, 而这种非线性变换是通过定义适当的内积函数实现的<sup>[13]</sup>。

支撑向量机求得的分类函数形式上类似于一个神经网络, 其输出是若干中间节点对应于输入样本与一个支撑向量的内积, 因此也被称为支撑向量网络, 如图 4-2 所示。

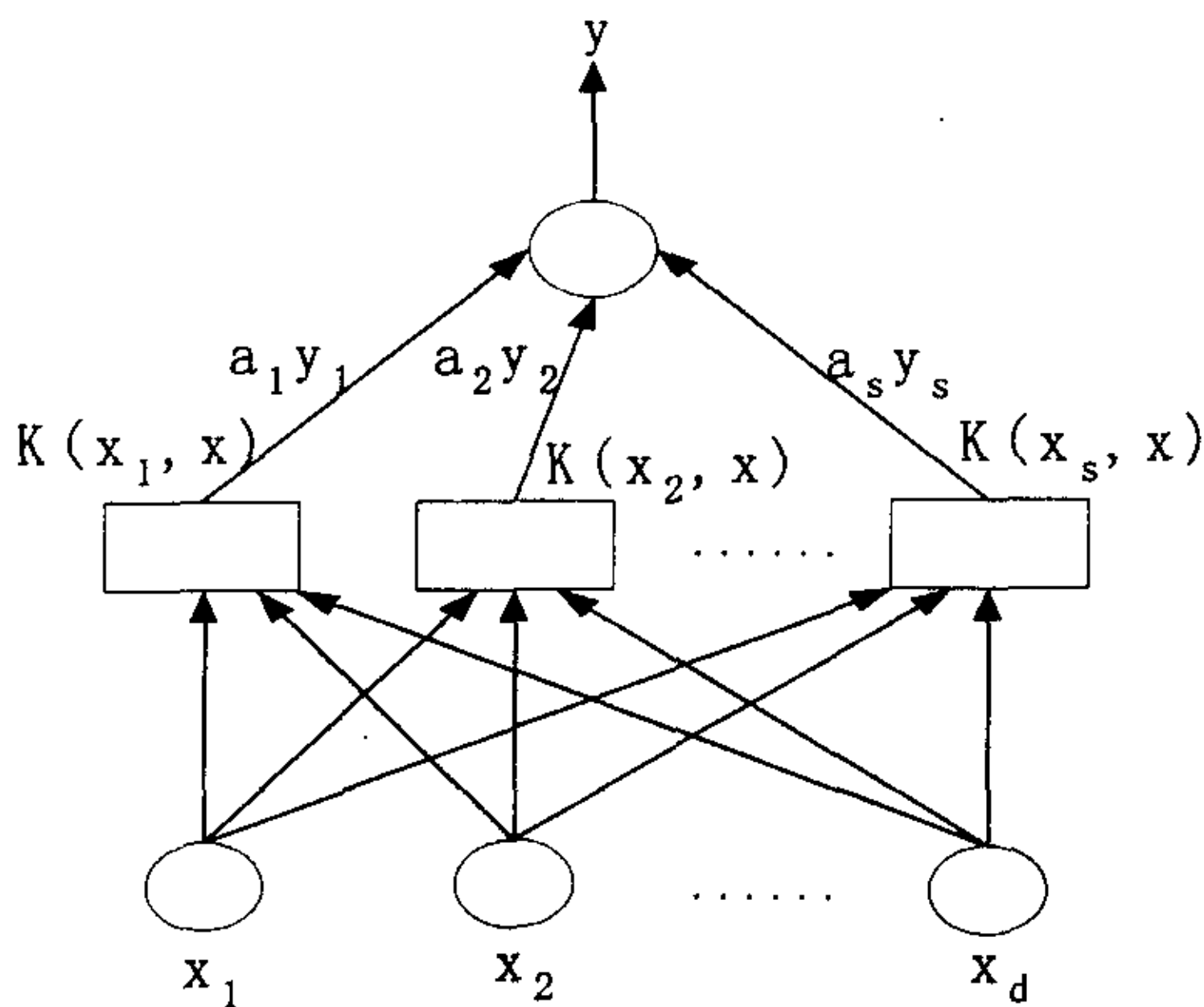


图 4-2 支撑向量机示意图

输入向量  $x = (x^1, x^2, \dots, x^d)$ , 基于  $S$  个支撑向量机  $x_1, x_2, \dots, x_s$  的非线性变

换(内积), 对应输出函数为

$$y = \text{sgn} \left( \sum_{i=1}^s a_i y_i K(x_i, x) + b \right) \quad (4-4)$$

其中权系数  $\omega_i = a_i y_i$ ,  $K(x_i, x)$  为内积函数。一个问题是, 由于最终的判别函数中实际只包含支撑向量机的内积和求和, 因此识别时的计算复杂度取决于支撑向量的个数。另一个问题是, 由于变换空间的维数可能很高, 在这空间中的线性判别函数的 VC 维因此也可能很大, 因此分类的性能并不一定好, 如何选择这种变换才能得到的判别函数具有比较好的推广性?

首先 SVM 方法是从线性可分情况下的最优分类面提出的, 若将所研究的对象分成二维线性可分, 分别用空心点和实心点表示两类训练样本,  $H$  为把两类对象正确地分开地分类线,  $H_1$ 、 $H_2$  分别为过各类样本中离分类线最近的点的且平行于分类线的直线,  $H_1$  和  $H_2$  之间的距离称为分类间隔。最优分类不但要求能将两类准确无误地分开, 而且要使两类的分类间隔最大。如图 4-3 所示。

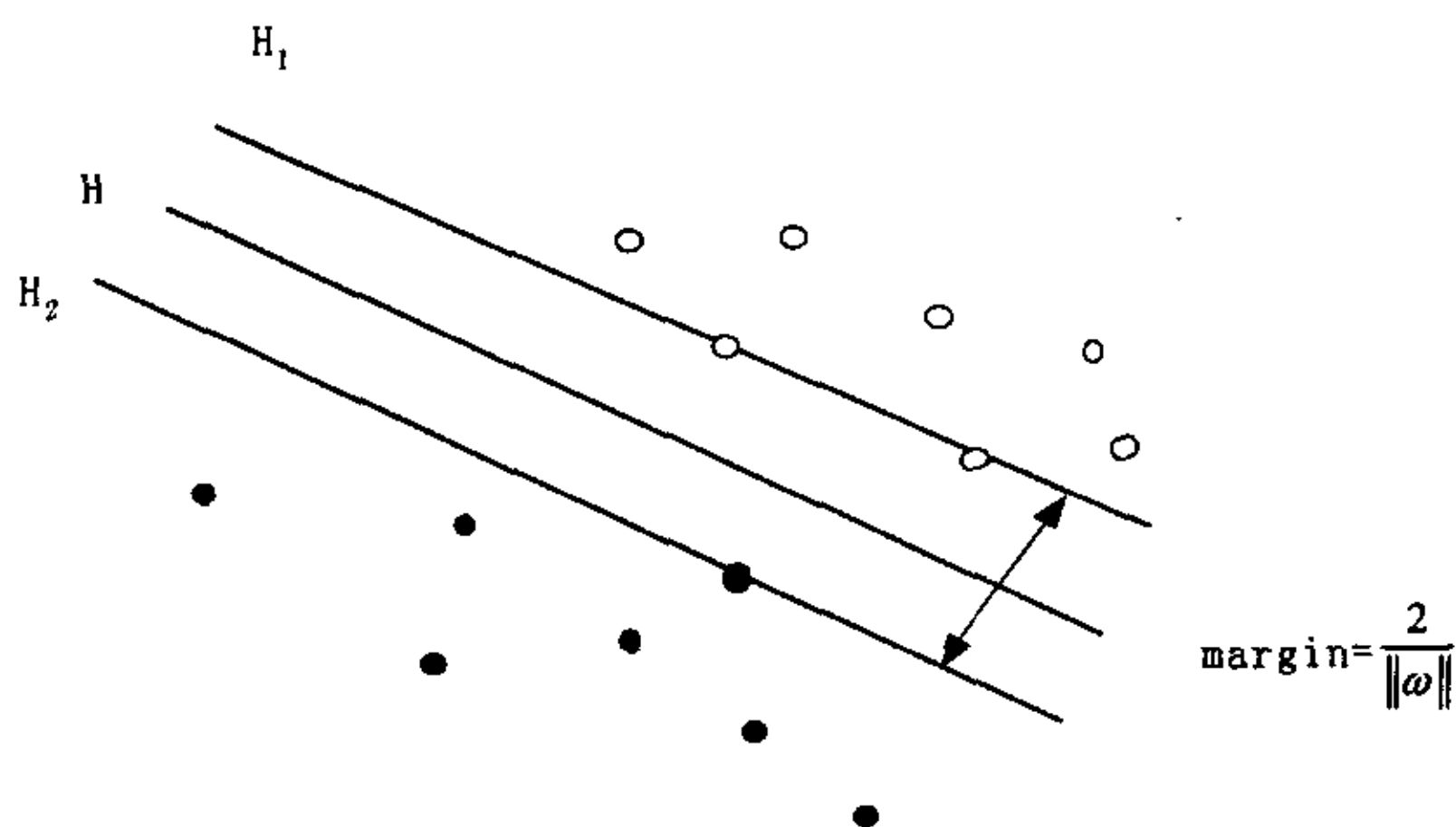


图 4-3 最优分类面示意图

设线性可分的样本集  $(x_i, y_i) i=1, 2, \dots, n$ ,  $x \in R^d$ ,  $y \in \{+1, -1\}$  是类别标号。 $d$  维空间中线性判别函数的一般形式为:  $g(x) = \omega \bullet x + b$ , 分类面方程为

$$\omega \bullet x + b = 0 \quad (4-5)$$

由图 4-3 可知要使分类间隔最大, 必须使  $\|\omega\|^2$  最小, 此时得到的分类面才是最优分类面。

此外, 还应该尽可能的减少 VC 维, 因此 VC 维的下界  $h$  满足以下不等式

$$h \leq \min([R^2 c], d) + 1 \quad (4-6)$$

设所有样本均在一个超球面内,  $R$  为其半径;  $c$  为最优分类面的条件即  $\|\omega\|^2 \leq c$ ;  $d$  为空间的维数。

## 4.2 应用径向基函数求解支撑向量

由于采用不同的内积函数, 因此求解支撑向量机的方法可能会有所差异。而本文采用径向基函数(Radial Basis Function 简称 RBF)作为内基函数  $K(x_i, x)$  的通用式为<sup>[13~14]</sup>

$$K(x, x_i) = \exp\left(-\frac{|x - x_i|^2}{\sigma^2}\right) \quad (4-7)$$

以径向基函数为内积函数, 将获得到一个基于径向基函数的支撑向量机分类器, 且每个基函数的中心对应一个支撑向量, 他们以及其输出权值都是由算法自动确定的。本章主要通求解基函数的中心向量进而实现求取支撑向量的目的。

### 4.2.1 径向基函数中心向量的学习

径向基函数(RBF)神经网络<sup>[21~23]</sup>是近几年来继多层感知机(Multi-Layer Perception, 简称 MLP)之后又一种十分有效的多层前向网络<sup>[24]</sup>。RBF 网络可通过输出节点计算隐含层节点, 并求出的输出函数(基函数)的线性组合, 而隐含层的输出函数通常由高斯函数定义。对于一个基函数类型及个数已选好的网络, 为了能使其正确地执行期望地数据处理, 有两个参数需要确定: 一基函数的中心在输入样本空间的位置; 二隐含层到输出层的权向量集合。而论文主要通过径向基函数确定其中心位置, 包含中心的数目和宽度, 从而求解出支撑向量机内积函数的参数, 因此, 本节的主要工作为: 1. 初始化网络结构及学习参数, 2. 训练权向量  $v_l$  计算  $\delta_l^2$  的值( $l=1, 2, \dots, c$ )。

第一个阶段通常的算法有  $K$ -均值聚类<sup>[24~27]</sup>、Kohonen 聚类<sup>[28]</sup>、正交最小二乘算法<sup>[29]</sup>等, 对初始值的选取非常敏感, 而且中心向量计算复杂, 致使学习的速度较慢。根据这一特点, 对于中心的调整采用了不同的模式, 即根据样本的不同响应程度, 中心的调整不同; 论文采用了一种网络初始化算法—网络增长技术<sup>[25]</sup>。而对第二阶段的学习, 引用了一种基于中心学习的软竞争学习策略<sup>[24~25][30]</sup>, 用来确定径向基函数中心的个数和位置, 其主要思想: 定义一个网络隐层结点数的增长的规则, 由此来自我增长 RBF 网络的隐节点。

### 4.2.2 学习算法

设训练样本集  $X = \{x_1, x_2, \dots, x_p\}$ ,  $p$  为样本的个数, 其中第  $i$  个输入样本  $x_i = \{x_{i1}, x_{i2}, \dots, x_{im}\}^T \in R^n$ ;  $y_i = \{y_{i1}, y_{i2}, \dots, y_{im}\} \in R^m$ ,  $o_i = \{o_{i1}, o_{i2}, \dots, o_{im}\} \in R^m$  分别为网络对应  $x_i$  的实际输出和期望输出。网络隐含层的节点个数设有  $c$  个。令  $w_{jl}$  表示第  $l$  个隐节点到第  $j$  个输出节点的权值;  $v_{lk}$  为第  $k$  个输入节点到第  $l$  个隐节点的权值。

RBF 网络隐层的输出函数采用高斯基函数, 它对输入激励产生一个局部化的响应, 这特点使高斯隐层对输入样本有一个聚类的作用, 隐层节点数代表聚类的类数, 隐层的中心为这一类的聚类中心点, 基函数的形式通常为

$$R_l(x) = \frac{1}{2} \exp \left[ -\frac{\|x - v_l\|^2}{2\sigma_l^2} \right] \quad (4-8)$$

其中  $v_l = (v_{l1}, v_{l2}, \dots, v_{ln})$ ,  $\sigma_l^2 (l=1, 2, \dots, c)$ , 分别是第  $l$  个隐层单元的中心向量和宽度, 当输入样本为  $x_i$  时, 网络的实际输出为

$$y_{ji} = f(x_i) = \sum_{l=1}^c w_{jl} R_l(x_i), \quad j=1, 2, \dots, m \quad (4-9)$$

#### 1. 初始化网络结构和学习参数

初始化网络结构和学习参数通常采用  $K$ -均值聚类来训练中心向量  $v_l$

$$v_l = \frac{1}{n_l} \sum_{x \in X_l} x, \quad l=1, 2, \dots, c \quad (4-10)$$

其中  $X = X_1 \cup \dots \cup X_l \cup \dots \cup X_c$  是由最近邻规则定义的一个硬  $K$ -分割,  $n_l = |X_l|$  表示第  $l$  类的容量, 下面的  $n_l$  的意义相同。在一般情况下, 宽度  $\sigma_l^2$  表示第  $l$  类的分散度

$$\sigma_l^2 = \frac{1}{n_l} \sum_{x \in X_l} (x - v_l)^T (x - v_l) \quad (4-11)$$

另外, 网络的权向量  $\omega_j = (\omega_{j1}, \omega_{j2}, \dots, \omega_{jc}) (j=1, 2, \dots, m)$  的学习是基于求下式定义的误差函数  $E_i$  最小值的梯度下降法<sup>[31]</sup>导出, 它的第  $t+1$  次学习规则由下面定义可知:

$$E_i = \frac{1}{2} \sum_{k=1}^m (y_{ik} - o_{ik})^2, \quad i=1, 2, \dots, p \quad (4-12)$$

$$\omega_{jl}(t+1) = \omega_{jl}(t) + \alpha(t)(y_{ij} - o_{ij})\phi_l(x_i)$$

其中,  $\frac{\partial E_i}{\partial \omega_{ji}} = (y_{ij} - o_{ij}) \phi_j(x_i)$ ,  $\alpha(t)$  为学习步长。

RBF 网络增长技术(Network Growth Technique 简称 NGT)的主要目的是为了克服以往凭经验选取 RBF 网络隐层节点个数的弊端,为隐层节点个数的选取找一定的理论依据;同时也可对网络结构和学习参数的初始化过程。其实现过程如下:

$$\text{NO. 1 计算 } v_1 = \frac{1}{p} \sum_{i=1}^p x_i \text{ 和 } \sigma_1^2 = \frac{1}{p} \sum_{i=1}^p (x_i - v_1)^2$$

令  $c = 1, \eta = 1, v_1 = v$  和  $\sigma_1 = \sigma$ 。

NO. 2  $K = 1 \cdots p$ ;

1. 如果  $\|x_k - v_l\|^2 = \min_{j=1 \cdots c} \|x_k - v_j\|^2 < \lambda \sigma_l^2$  ( $0 < \lambda < 1$ )

按下面的调整中心向量

$$n_l = n_l + 1;$$

$$v_l = v_l + \frac{1}{n_l} (x_k - v_l);$$

$$\sigma_l^2 = \frac{n_l - 1}{n_l} \sigma_l^2 + \frac{1}{n_l} (x_k - v_l)^2。$$

2. 否则按下面原则建立一个新的隐节点, 设  $c = c + 1$ 。

$$n_c = 1, v_c = x_k;$$

$$\sigma_c^2 = \|x_k - v_l\|^2。$$

NO. 3 如果节点不再增加, 则转入软竞争学习阶段。

2. RBF 网络的软竞争学习算法

$K$ -均值聚类算法虽然计算简单, 而且性能良好, 但使用它来训练 RBF 网络的中心向量时, 一方面对初值的选取敏感, 同时存在死结点(dead nod), 即无任和一个样本与其距离最近的问题。基于上述原因, 论文引用了一种新的算法-软竞争学习算法(Soft-competition Learning Algorithm, 简称 SCLA)。该算法采用两阶段分别训练中心向量和输出权向量  $\omega_j$ , 权向量  $\omega_j$  仍然用梯度下降法进行训练, 而中心向量的调整则利用一种软竞争方法。其学习的步骤如下:

1. 确定隐层的节点数  $c$  和最大的迭代次数  $T$ ;

2. 初始化中心  $v_l(0) = (v_{l1}(0), v_{l2}(0), \cdots, v_{ln}(0)) (l = 1, 2, \cdots, c)$ ;

3. For  $t=1$  to  $T$

a. 修正  $\beta(t)$ ;

b. For  $i=1$  to  $p$

For  $l=1$  to  $c$

计算  $P_{li}(t)$

$$P_{li} = \frac{\exp\left(-\beta(t)\|x_i - v_l(t-1)\|^2\right)}{\sum_{r=1}^c \exp\left(-\beta(t)\|x_i - v_r(t-1)\|^2\right)}$$

按下面的方法调整  $v_l(t)$

$$v_l(t) = v_l(t-1) + \eta(t) P_{li}(x_i - v_l(t-1))$$

4. 计算  $E1(t+1) = \sum_{j=1}^c \sum_{i=1}^n |v_{ji}(t+1) - v_{ji}(t)|$ ;

5. 计算  $E1(t+1) \geq \varepsilon_1$  且  $t < T$ , 令  $t = t+1$ , 转入第 3 步; 否则按照下式计算出

$$\sigma_l^2 = \frac{\sum_{i=1}^p P_{li}(t)(x_i - v_l)^T (x_i - v_l)}{\sum_{i=1}^p P_{li}(t)}$$

其中, 步长  $\eta(t)$  一般根据 Kohonen 网络的收敛性来选取, 通常根据模拟退后算法的冷却进度表,  $\eta(t)$  值的变化公式如下<sup>[32]</sup>

$$\eta(t) = \begin{cases} \eta_0 e^{-\alpha t} & \text{其中 } \alpha > 0 \\ \eta_0 e^{-\alpha} & \text{其中 } \alpha \leq 1 \\ \eta_0 (1 - \alpha t) & \text{其中 } 0 < \alpha < (\max\{t\})^{-1} \end{cases} \quad (4-13)$$

由上面学习步骤可知,  $0 < P_{li}(t) < 1$  且  $\sum_{i=1}^c P_{li}(t) = 1$ , 因此可以称  $P_{li}(t)$  为第  $i$  个样本  $x_i$  属于第  $l$  类的模糊隶属度函数,  $\beta(t)$  是表征其模糊程度的一种因子, 所以  $v_l$  的调整方法是一种模糊自适应方法, 在每次迭代时, 所有隐层节点的中心同时都进行调整。此外, 软竞争学习算法的关键在于如何选取  $\beta(t)$ , 论文基于模拟退后算法的思想来选取  $\beta(t)$ <sup>[32~33]</sup>: 开始设置高的温度随时间变化, 温度逐渐降低。而  $\beta(t)$  是迭代次数  $t$  的递增函数, 且满足  $\lim_{t \rightarrow +\infty} \beta(t) = \infty$ , 因此,  $\beta(t)$  可定义为



$$\beta(t) = \beta_0 + \frac{\beta_1 - \beta_0}{T} \quad (4-14)$$

上述的算法是一种序列训练方法，它的每次调整与当时输入样本有关，因而样本的输入顺序对其是有直接影响的。K-均值方法的优点在于它是一种批处理方式，同样考虑上述算法的批处理方式。按批处理方式求得  $v_i$  为

$$v_i(t) = v_i(t-1) + \eta(t) \frac{\sum_{l=1}^p P_{li}(t)(x_l - v_i(t-1))}{\sum_{l=1}^p P_{li}(t)} \quad (4-15)$$

### 4.3 分类器实例构造

由于该支撑向量机分类器的内积函数是径向基函数，因此每个径向基函数的中心向量对一个支撑向量，通过 4.2 节求取了中心向量  $v_i$  和  $\delta_i^2$ ，完成了分类器的设计，且对应的程序流程图 4-4。

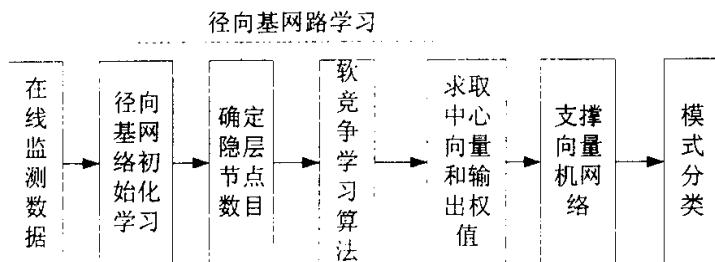


图 4-4 支撑向量基分类器的流程图

将中心向量和内积函数代入公式 (4-4) 得到分类器的表达式为：

$$y = \text{sgn} \left( \sum_{i=1}^s a_i y_i \exp \left( -\frac{|x - x_i|^2}{\sigma^2} \right) + b \right) \quad (4-16)$$

公式 (4-16) 为分类器函数，对于输出期望  $y_i$  及参数  $a_i$ 、参数  $b$  来说存在以下约束条件

$$\begin{aligned} \sum_{i=1}^s y_i a_i &= 0 & a_i &\geq 0, i=1, 2, \dots, s \\ a_i (y_i (\omega \bullet x_i) + b - 1) &= 0 & i &= 1, 2, \dots, s \\ y_i (\omega \bullet x_i) + b - 1 &= 0 \\ \omega &= \sum_{i=1}^s a_i y_i x_i \end{aligned} \quad (4-17)$$

对于非支撑向量对应的参数  $a_i$  均为 0，支撑向量对应的参数  $a_i$  大于 0；

而  $b$  为分类的域值，也可通过上式求解。

论文以某变电站金属氧化物避雷器全电流和阻性电流在线测量值为对象，把设备状态分成两类状态，即构建一个二维输入、二维输出、隐含层可求的支撑向量机分类器。

论文选取了 240 组数据作为样本，其中训练样本为 120 组，分类检验样本为 120 组，其中训练样本为正常数据，而分类样本中既包含正常数据又包含故障数据。用径向基神经网络训练前 120 组样本，可求出  $v_i$  和  $\delta_i^2$  和支撑向量的个数  $c$ ，通过软竞争学习算法求得  $c=3$ 。为了更有效的证明分类器的分类作用，选取的分类样本具有典型的意义，如图 4-5 所示。

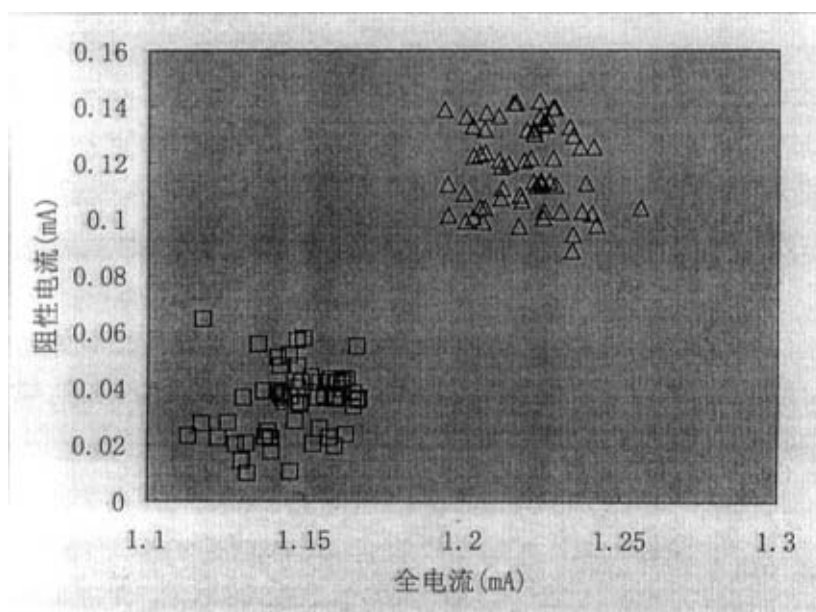


图 4-5 分类样本的原始数据

由图 4-5 可知，分类样本数据明显成两类，方框是避雷器在线监测的获取的在线数据，代表避雷器正常状态；而三角形则通过对避雷器故障仿真得到的仿真数据，代表避雷器故障状态。因此设计的分类器应能将避雷器两类状态正确地分开，且其错误分类的比例由参考文献<sup>[13]</sup>可知：如果一组训练样本能够被一个最优分类面或广义最优分类面分开，则对于检验样本错误的期

望的上界是训练样本中平均的支撑向量占训练样本数的比例，即

$$E(P(error)) \leq \frac{E[\text{支撑向量机}]}{\text{训练样本总数}-1} \quad (4-18)$$

因此，对于上述基于径向基函数的支撑向量分类器的模型的错误期望上界  $E = \frac{3}{120-1} \times 100 = 2.52\%$ 。应用上述基于径向基函数的支撑向量分类器进行分类，其分类结果如图 4-6 所示。

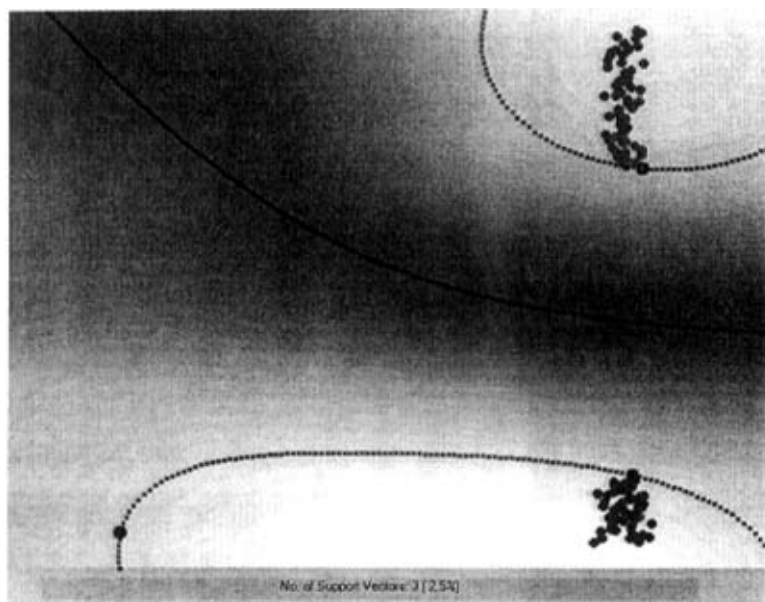


图 4-6 支撑向量机的分类图

由图 4-6 可知，上面的点对应图 4-5 三角所代表数据的状态，即避雷器的故障状态，下面的点则对应图 4-5 方框所代表数据的状态，即避雷器的正常状态；中间的弧线将图形分成两个区域，上面的区域表示避雷器故障状态的区域，下面的区域表示避雷器正常状态的区域。由此可见，分类器将避雷器正常数据与故障数据正确地分开，且没有出现分类错误的情况；在上下两个区域内各有一条虚线，该虚线是分类器最优线性分类面，两虚线之间的距离表示最大的分类间隔，虚线上的每个点对应该分类器的一个支撑向量，虚线上刚好有 3 个点，即支撑向量的个数  $c=3$ 。分类器对应的分类与避雷器监测数据之间存在对应关系：即正常数据对应正常状态，故障数据对应故障状

态。因此分类器用于设备的状态识别是可行的。

由 4.2 节的推理和 4.3 节的实例证明:基于径向基函数的支撑向量分类器能够通过在线数据对被监测对象进行准确的分类,因此表明了用该分类器对被监测设备进行模式分类是可行的。该分类器是通过在高维空间构造一个线性最优分类面实现准确地分类,而实现线性最优平面的关键在于选择一个容易求解的内积函数。论文中采用了径向基神经网络作为内积函数,而该内积函数有较为成熟的算法,因而实现起来很方便;但径向基网络学习算法复杂,涉及知识面宽,学习的速度受训练样本数量的约束。

#### 4.4 小结

本章设计了一个基于径向基函数的支撑向量分类器,该分类器以在线数据为对象通过构造内积函数和学习网络实现了设备状态的分类,从而达到了对在线监测设备进行分类识别的目的。论文从支撑向量机理论的阐述、径向基函数的学习算法以及分类器实例构造等几方面展开。

1 支撑向量机理论的阐述。支撑向量机是一种新的模式识别方法,能较好地解决小样本学习的。它通过非线性变换将输入空间变换到一个高维空间,然后在这个空间中求取最优线性分类面,而这种非线性变换是通过定义适当的内积函数实现的。其关键在于在高维空间构造一个最优分类平面以及选择在最优分类平面一个合适内积函数。

2 径向基函数的学习算法。论文采用网络增长技术对径向基函数网络结构和学习参数进行初始化处理,从而求得了中心向量的数目;然后运用软竞争学习算法分别训练中心向量和输出权向量,其中引入模拟退火算法来控制训练的步长。由于把径向基函数作为内积函数,由于其每个中心向量对应于一个支撑向量,从而实现了求解支撑向量的目的。

3 分类器实例构造。论文以 110kV 避雷器在线监测的全电流和阻性电流的测量值为输入节点,把监测设备分成两类不同的状态,且分类状态与监测数据之间存在对应关系。实例证明该分类器能根据在线数据正确地识别设备的状态,且准确率很高。

## 5 结论

论文以变电设备在线绝缘检测数据为对象, 对其进行综合分析与处理。主要从在线数据的采集和预处理、数据预测和设备状态的模式分类三个方面进行了研究。通过本文的工作得到以下结论:

1 论文提出了一种新的滤波算法—中值滤波快速算法, 且算法简单, 与其它滤波算法比较具有机器运算时间少的特点。实例分析证明该方法对抑制脉冲干扰效果明显。而对于频域内高频或低频的干扰, 可以考虑设计一个合适的滤波器或对其进行小波变换。

2 论文采用时间序列的思想对在线数据进行递推预测。先建立预测的模型, 然后用最小二乘递推算法求解模型参数; 最后引入了遗忘因子对其参数进行校正。模型既包含历史数据的信息又包含最新数据的信息, 遗忘因子的作用不仅减小历史数据的作用, 而且增强新数据的作用。以 220kV 变压器 A 相套管泄漏电流为例, 对数据进行了实时预测, 结果表明算法对于在线数据的预测是有效的, 且精度较高, 而数据预测的范围在很大程度上取决于训练样本的容量和算法的精度。

3 论文采用支撑向量机的思想设计了一个基于径向基函数的分类器。先通过对径向基网络的学习求取隐含层节点数目和中心向量, 然后由支撑向量机网络获得状态的分类, 其中关键的问题在于如何获取一个最优的线性分类面以及选择一个合适内积函数。分类器设置较灵活, 且对于设备状态的分类准确率较高, 因此该分类器可以用于在线设备的状态分类。

## 致谢

本论文是在严璋教授的悉心指导下完成的。在这两年多的研究生学习期间，作者在学习、生活等各个方面都得到了导师的关怀、爱护和帮助。导师渊博的知识、开阔的思路、敏锐的洞察力、严谨的治学态度、诲人不倦的长者风范以及崇高的敬业精神都给学生以深刻的启迪与影响，将使学生受益终生。在此论文完成之际，谨向导师对我关怀和培养表示衷心的感谢。

承蒙高压教研室董连文老师和张冠军老师对本文初稿进行了仔细审阅，并提出了许多宝贵的修改意见，在此谨向他们表示最诚挚的谢意。同时也感谢高压教研室的各位老师、同学两年多来对我的关心和帮助。

在二年多的学习期间，作者有幸结识了杨莉博士、王晓蓉博士、尚勇博士、董明博士生、李明华博士生、杨敏中硕士、郭宗军硕士、胡龙龙硕士、刘少宇硕士、赵文彬硕士生、周海洋硕士生、闫春江硕士生、王雪硕士生、孟源源硕士生、徐长响硕士生、李华硕士生、莫娟硕士生等一批志同道合、同舟共济的同学，在这个学术气氛活跃、团结友爱的课题组里，大家互相帮助，彼此探讨，共同提高，结下了深厚的友谊，作者将永远铭记在心。

二年多的学习生涯使我学到了很多，我会永远珍视这段难忘的岁月。回首这二年多走过的路，有欢乐，也有辛酸，但我无悔。感谢生活，感谢她给予我的一切。

最后，谨以此文献给我的家人和女友，没有他们的默默支持我将一事无成。

## 参考文献

1. 韩祯祥. 电力系统自动监视与控制. 北京: 水利电力出版社, 1989
2. 郭剑波. “八五”期间电网事故统计分析. 电网技术, 1998, 22(2): 72~74
3. 郭碧红、杨晓洪. 我国电力设备在线检测技术的开发应用状况分析. 电网技术, 1999, 23(8): 65-68
4. 严璋. 电气绝缘在线检测技术. 北京: 中国电力出版社, 1995
5. 郭宗军. 变电设备绝缘在线监测数据的分析及处理: [硕士学位论文]. 西安: 西安交通大学, 2002
6. 杨莉. 知识发现在变电设备状态监测及诊断专家系统中的应用: [博士学位论文]. 西安: 西安交通大学, 2001
7. 杨莉、张理等. 在线监测数据剔点处理算法的研究. 高压电器, 2000, 36(5): 3~6
8. 李辉、杨增辉、周海洋、严璋. 快速滤波算法于在线监测数据预处理中的应用. 高电压技术, 2002, 28(7): 31~33
9. 杨叔子、吴雅等. 时间序列分析的工程应用. 武汉: 华中理工大学出版社, 1994
10. 邓聚龙. 灰色系统理论教程. 武汉: 华中理工大学出版社, 1990
11. 张学工. 统计学习理论的本质. 北京: 清华大学出版社, 2000
12. A.J.Smola, B.Schölkopf,. A Tutorial on Support Vector Regression. Produced as part of the ESPRIT Working Group in Neural and Computation Learning II, NeuroCOLT2 27150 ,October 1998
13. 边肇祺、张学工等. 模式识别 北京: 清华大学出版社, 2000
14. B.Schölkopf, K.K.Sung, V.Vapnik etc. Comparing Support Vector Machine with Gaussian Kernels to Radial Basis Function Classifiers. IEEE Transaction on Signal Processing Vol 45, No.11, November 1997, 2758~2765
15. 郑南宁. 数字信号处理. 西安: 西安交通大学出版社, 1996



16. 刘松强. 数字信号处理系统及其应用. 北京: 清华大学出版社, 1996
17. 殷人昆、陶永雷、谢若杨、盛绚华. 数据结构. 北京: 清华大学出版社, 1999
18. 陈继东. 小波分析应用于在线监测中信噪分离的研究, 电网技术, 1999, 23(11): 54-57
19. 彭玉华. 小波变换与工程应用. 北京: 科学技术出版社, 2000
20. 田铮. 动态数据处理的理论与方法—时间序列分析. 西安: 西北工业大学出版社, 1995
21. 汪西莉、刘芳、焦李成. 基于大规模数据库的支撑向量机的训练和分类. 西安电子科技大学学报(自然科学报), 2002, 29(1): 123-127
22. 陈涛、蒋林、屈梁生. 基于正交最小二乘学习算法的径向基函数网络设计. 中国机械工程学报, 1997, 8(6): 95-98
23. 毕天姝、倪以信、吴复立、杨奇逊. 基于新型神经网络的电网故障诊断方法. 电机工程学报, 2002, 22(2): 73-78
24. 张志华、郑南宁、史罡. 径向基函数神经网络的软竞争学习算法. 电子学报, 2002, 30(1): 132-135
25. 史罡. 径向基函数神经网络的学习算法及其在控制中的应用: [硕士论文], 西安: 西安交通大学, 1999
26. J.C.Bezdek. Pattern recognition with fuzzy objective function algorithm. New York: Plenum, 1981
27. 岳喜才. 基于径向基网络的文本无关说话人识别研究: [博士论文], 西安: 西安交通大学, 1999
28. T.Kohonen. Self-organization and associative memory. Berlin, Germany: Springer-Verlag, 1989, 3rd. ed
29. S.Chen, C.F.N.Corvan, P.M.Grant. Orthogonal least squares learning algorithm for radial basis function networks. IEEE Trans. on Neural Network, 1991, 2(2): 302-309
30. E.Yair, K.Zeger, A.Gersho, Competitive learning and soft competitive for vector quantifier design, IEEE Trans. on Signal Processing, Vol.40, 294-308, 1992

31. B.Tarassenko, S.Roberts. Supervised and unsupervised learning in radial basis function classifier. Vision Image Signal Processing, 1994, 141(1): 210-216
32. 张智星、孙春在、水谷英二. 神经—模糊和软计算. 西安: 西安交通大学出版社 Prentice-Hall. 2000
33. 王凌. 智能优化算法及其应用. 北京: 清华大学出版社 施普林格出版社, 2000