

华北电力大学

专业硕士学位论文

基于大数据的设备状态在线监测与预警诊断 系统研究

Research on Equipment Condition Monitoring and Early Warning System Based on Big Data Technique

李子龙

2016 年 6 月

国内图书分类号：TP273

学校代码：10079

国际图书分类号：621.3

密级：公开

专业硕士学位论文

基于大数据的
设备状态在线监测与预警诊断系统研究

硕士研究生：李子龙

导师：田亮副教授

企业导师：李涛

申请学位：工程硕士

专业领域：控制工程

培养方式：在职

所在学院：控制与计算机工程学院

答辩日期：2016年6月

授予学位单位：华北电力大学

Classified Index: TP273

U.D.C: 621.3

Thesis for the Master Degree

Research on Equipment Condition Monitoring and Early Warning System Based on Big Data Technique

Candidate :	Li Zilong
Supervisor :	Associate Prof. Tian Liang
Academic Degree Applied for :	Master of Engineering
Speciality :	Control Engineering
School :	School of Control and Computer Engineering
Date of Defence :	June 2016
Degree-Conferring-Institution :	North China Electric Power University

华北电力大学硕士学位论文原创性声明

本人郑重声明：此处所提交的硕士学位论文《基于大数据的设备状态在线监测与预警诊断系统研究》，是本人在导师指导下，在华北电力大学攻读硕士学位期间独立进行研究工作所取得的成果。据本人所知，论文中除已注明部分外不包含他人已发表或撰写过的研究成果。对本文的研究工作做出重要贡献的个人和集体，均已在文中以明确方式注明。本声明的法律结果将完全由本人承担。

作者签名：

日期： 年 月 日

华北电力大学硕士学位论文使用授权书

《基于大数据的设备状态在线监测与预警诊断系统研究》系本人在华北电力大学攻读硕士学位期间在导师指导下完成的硕士学位论文。本论文的研究成果归华北电力大学所有，本论文的研究内容不得以其它单位的名义发表。本人完全了解华北电力大学关于保存、使用学位论文的规定，同意学校保留并向有关部门送交论文的复印件和电子版本，允许论文被查阅和借阅，学校可以为存在馆际合作关系的兄弟高校用户提供文献传递服务和交换服务。本人授权华北电力大学，可以采用影印、缩印或其他复制手段保存论文，可以公布论文的全部或部分内容。

本学位论文属于（请在以上相应方框内打“√”）：

保密☐，在 年解密后适用本授权书

不保密☐

作者签名：

日期： 年 月 日

导师签名：

日期： 年 月 日

摘 要

目前，各大发电企业均部署了发电企业控制系统、厂级监控系统（SIS）以及各种各样的信息系统。这些系统积累了设备运行的海量历史数据，如何借助大数据分析技术将这些数据转化为指导发电厂生产运行的决策信息，是现代火力发电厂走向信息数据化、状态在线化和决策智能化的关键。

本文介绍的设备状态在线监测与预警诊断系统就是基于大数据技术，从设备运行的海量历史数据中分析出设备的各种运行工况及参数关联关系建立状态动态监测模型，在设备运行过程中更准确、更灵敏地发现设备的异常变化，从而比DCS 和人工点检更早的发现设备问题，实现企业和厂级生产数据挖掘和设备故障早期预警。

本文详细介绍了基于大数据的设备状态在线监测系统在国内某300MW 机组的应用。通过对大数据技术的深入研究，结合火力发电厂设备的实际运行情况，选用合适的数据挖掘算法建立电厂关键设备模型。在建模过程中，无论是数据集成、数据清理还是数据挖掘，都体现了大数据技术与电厂专业领域的深入结合。同时，本文重点介绍了氧化风机、凝结水泵和磨煤机等电厂关键设备的建模过程以及模型构建后的实际监测情况，并将这些设备的监测预警情况以案例的形式进行了重点分析。

关键词：火电厂；大数据；状态监测；故障预警；诊断分析；

Abstract

Currently, most power plants have deployed power plant control systems, supervisory information system (SIS) and a wide variety of manager information systems. These systems have accumulated mass data of equipments. How to translate these data into the decision information of the production and operation of the power plant by means of the big data analysis technology is the key to the information data, the state and the intelligent decision of the modern thermal power plant.

This paper describes the equipment status online monitoring and early warning diagnosis system is based on big data technology. From the massive historical data of equipment operation, the various operation conditions and parameters of the equipment were analyzed to establish the dynamic state monitoring model, The abnormal change of equipment is more accurate and more sensitive in the process of equipment operation, To discover the DCS and artificial inspection equipment, The realization of enterprise and plant production data mining and fault early warning device.

In this paper, the application of equipment condition monitoring system based on large data in a 300MW unit in China is introduced in detail. Through in-depth study of large data technology, combined with the actual operation of thermal power plant equipment, select the appropriate data mining algorithm to establish the model of the key equipment of the power plant. In the modeling process, whether it is data integration, data cleaning or data mining, has reflected the combination of large data technology and power plant professional. At the same time, this paper mainly introduces the modeling process of the key equipment of the power plant, such as the oxidation fan, the condensate pump and the coal mill, and the actual monitoring situation after the model construction, And the monitoring and early warning of these devices in the form of a case analysis of the key analysis.

Keywords: power plant; big data; condition monitoring; fault prognosis; diagnostic analysis;

目 录

摘 要	I
Abstract.....	II
第 1 章 绪 论	1
1.1 课题背景及研究的目的和意义	1
1.2 国内外研究状态	1
1.3 本文研究的思路及主要内容	3
1.3.1 研究思路	3
1.3.2 研究内容	3
第 2 章 大数据概述	4
2.1 大数据分析技术	4
2.2 大数据挖掘	4
2.2.1 定义	4
2.2.2 一般过程	5
2.2.3 数据挖掘常用方法	6
第 3 章 基于大数据的设备状态在线监测与预警诊断系统理论研究	8
3.1 关联度的计算	8
3.1.1 关联度性质	8
3.1.2 关联度的计算	9
3.2 运行模式的计算	10
3.2.1 基本聚类挖掘方法	10
3.2.2 函数型数据的相似性指标	12
3.2.3 基于基函数展开的函数型数据聚类方法	13
3.2.4 利用相似系数聚类	13
3.3 本章小结	14
第 4 章 电厂关键设备氧化风机模型的构建	15
4.1 模型创建	15
4.2 模型优化及调整	21
4.3 模型运行情况分析	22
4.4 预警分析	24
4.5 本章小结	26

第 5 章 电厂关键设备凝结水泵模型的构建	27
5.1 模型创建	27
5.2 模型优化及调整	28
5.3 模型运行情况分析	29
5.4 预警分析	30
5.5 本章小结	31
第 6 章 电厂关键设备磨煤机模型的构建	32
6.1 模型创建	32
6.2 模型优化及调整	33
6.3 模型运行情况分析	34
6.4 预警分析	34
6.5 本章小结	35
第 7 章 结论与展望	36
7.1 结论	36
7.2 展望	36
参考文献	37
攻读硕士学位期间发表的论文及其它成果	40
攻读硕士学位期间参加的科研工作	41
致 谢	42
作者简介	43

第 1 章 绪 论

1.1 课题背景及研究的目的和意义

自工业革命以来，设备故障的监测及诊断就应运而生。在很久之前，工人们判断设备发生故障与否，主要是通过长时间累积的经验或是对于设备表象的观察而得出的结论^[1-3]。比如说，维修工人在设备巡检的时候，常常采用敲击的方法，通过声音的不同，确定故障类型，并对故障位置进行定位，进而以此为基础制定相应的维护方案。从 20 世纪 60 年代开始，美国宇航局将设备故障的监测及诊断列为一门学科进行系统性的研究^[4]。1961 年起，美国实施了阿波罗计划，该计划以后，便开始出现设备故障问题，为了解决令人棘手的问题，“美国故障预防”小组顺势成立，其主要任务是对故障问题的深入研究，从此故障诊断成为了一项技术被提上日程上来^[5]。

伴随着工业发展的突飞猛进，工业设备也日益发展，向智能、大型、高速、分布式的方向发展。设备结构相较于传统更加复杂，与此同时功能也愈加强大，进而增加了维护的难度^[6]。继而，设备发生故障的概率较之以前大幅提高，设备故障的信息数据呈现指数型增长，所产生的海量数据，人工已经无法负荷，无法完成正确、有效的故障分析^[7-9]。此外，工业设备结构复杂，各部分之间存在相互作用，导致故障的发生也会产生交集，故障机理便因此变得非常复杂。为此，对设备所产生的海量故障信息的分析尤为关键，将有效确定故障类型，定位故障位置，有效提高设备维护效率，减少维修费用^[10-13]。

随着信息技术的高速发展，特别是发电企业控制系统、SIS 系统和各种信息系统的大规模建设促使发电厂积累的数据量迅猛增长，为发电企业关键设备的大数据分析研究工作奠定了基础。通过特有、高效的实时数据挖掘技术，新一代的设备状态在线监测与故障预警系统能够帮助用户实现设备状态的智能管理，进而充分发挥设备管理者的专业效率，变故障事后处理为事前预防，实时掌握设备运行中的整体动态变化，在每台设备的生命周期内，极大地提高设备的运行安全水平和效率，降低由于设备原因引起的非计划停机与事故，减少设备运维成本，为发电企业创造更多的效益^[14-18]。

1.2 国内外研究状态

故障诊断由三个不同的阶段组成，在第一个阶段中，所涉及的工业设备是

比较简单的，故障也不是那么复杂，设备的运行监测和故障诊断都可以完全借助于经验进行确认，除此之外还可以应用机器仪表；进入第二个阶段，由于各项技术的迅猛发展，和在故障诊断方面的渗透，故障诊断技术也得到了大幅度的进步，虽然该阶段取得了不少研究成果，但是仍然不能缺少人工决策。发展到第三阶段，各项技术向故障诊断中充分渗透，使得复杂设备的故障诊断变得简单化，其中发展最为成熟的例子是人工智能技术与故障诊断的结合，包括神经网络诊断、模糊诊断等。伴随着网络技术的快速发展，诞生了基于网络的故障监测手段，这种手段可以有效连接各部分设备，将所有设备信息进行传递、处理，最终构建起服务网络，实现所有设备的故障诊断^[19-24]。

目前，包括美国、英国、日本在内的诸多国家研究机构把工作的中心进行转移，重点进行有关故障检测的理论和应用研究^[25]。

在火电厂，汽轮机振动监测与故障诊断最早触及状态监测领域。目前，美国的美国电力研究协会、西屋公司等都在从事着电厂故障诊断技术方面的开发与研究。西屋公司所开发的发电机、汽轮机人工智能诊断系统，目前连接了大约有十个电厂，运行超过十多年，该系统实施改善了七台机组的运行情况，有效降低了非计划停机率，并把机组平均可用率从 95.2% 提高到了 96.1%^[26-27]。

在锅炉故障监测及诊断方面，国外相关研究机构现已研发出了多种监测系统，并获得了成功应用，为机组的经济运行做出了巨大贡献。加拿大某高校通过在炉膛水冷壁安装热流计，建立了灰污诊断系统，并在 Boundary Dam 电厂 3 号机组上成功应用^[28]。

美国电力研究协会也成功开发了一套锅炉维护工作站系统，并得到了成功的应用。其主要功能模块有四个，该系统不仅可以准确地对水冷壁的实时运行工况进行分析，还能够进行换热器蠕变寿命的计算，其中包括过热和再热器，同时，还具备管道故障诊断功能。该系统的核心是 ESCARTA 专家系统，在美国，它是能够实现锅炉事故诊断的首个系统。此外，本系统还可以辅助并指导运行人员正确的分析锅炉四管泄露发生的原因^[29]。

随着对机组安全运行要求的提高，辅机状态监控也受到重视，例如空预器漏风、凝汽器脏污、风机、泵、阀门等辅机设备运行状态监控的技术发展十分迅速，这些辅机设备运行状态监测主要技术手段种类繁多。

近年来，在自动化、信息化技术的推动下，发电设备在线监测与故障预警系统正向着全方位、系统化的方向发展。美国电力研究协会组建了专门的监测与诊断中心，目前，该中心已经成功研发了涉及锅炉、汽轮机、发电机等方面的设备监测及故障诊断系统。该中心设有运维服务中心，可通过互联网获取各个电站的实时运行情况，完成对设备运行状态的准确评估并未电站提供必要的

运行及维修建议^[30-33]。与此同时，设备监测及故障诊断的相关理论研究也发展迅速，主要包括：（1）声、光测量技术、软测量技术等新技术；（2）大数据分析技术；（3）采用智能化方法对设备进行监测与诊断，如专家系统、神经网络等。

中国在设备监测及故障诊断方面的研究起步较晚，然而经过广大学者、工程技术人员探索和研究，与国外相关技术的差距日益缩小，甚至对于某些理论方面，赶超国外技术水平。现阶段，中国各大院校已经成功研制出了故障诊断系统，如西安交通大学、哈尔滨工业大学等^[34-36]。与此同时，一些大型发电企业正与国内外高校、研究所等单位合作，推行设备状态检修，该种检修方式是以设备的可靠性作为指标，旨在保证系统最优化运行。

1.3 本文研究的思路及主要内容

传统的设备在线监测及预警诊断技术的研究已经趋于完善。本文针对基于大数据的设备状态监测与预警诊断技术进行深入研究，从大数据技术入手，与电力企业生产实际情况想结合，探究采用大数据技术建立发电设备模型的方法。在此基础上，建立关键设备运行状态动态模型，并对模型不断进行优化调整，满足设备在线监测的需求。当设备发出故障早起预警时，对预警情况进行分析，并制定相应维护策略。具体研究内容包括：

1.3.1 研究思路

- （1）深入研究大数据技术，结合关键设备自身特点，建立设备故障预警模型的；
- （2）根据设备模型的运行情况，验证设备模型的准确性，并对模型进行相应的优化调整；
- （3）实时监测设备运行状况，并对其健康状况进行分析；
- （4）利用故障预警分析工具，分析设备预警原因；

1.3.2 研究内容

- （1）如何选取建立设备动态模型需要的测点；
- （2）如何根据测点数据，利用大数据技术建立设备动态监测模型；
- （3）如何优化设备动态模型，使监测模型趋于完善；
- （4）如何对设备的运行状态进行实时评估，以及对已发生的预警信息进行诊断分析。

第 2 章 大数据概述

2.1 大数据分析技术

大数据分析，就是对数目庞大的各种数据实施综合分析的过程。目前，各种信息数据增长已经很难想象。据保守预计，在未来的数年里，数据会以 50% 每年的速度飞速提高。在电力领域，各大发电厂均部署了发电企业控制系统、SIS 系统和各种信息系统，这些系统积累了关于过去生产与资产的海量历史数据，电力领域早已进入大数据时代。

对于大数据技术，其难点主要在于数据挖掘算法的选取。不管是孤立点分析、分割还是集群等其他的算法，这些算法都能够帮助挖掘出数据背后所蕴含的巨大价值。通常，这些算法必须要以最快的速度处理海量的数据，以满足需要，为从海量的数据中带来最有用的信息。在电力领域，生产控制系统产生的数据量都是秒级甚至毫秒级的，数据量之大恐难想象。要想通过数据挖掘，攫取这些生产数据的核心价值，所以首先要选取合适算法。

本文所研究的设备状态在线监测与预警诊断系统就是采用大数据技术，从发电设备运行的海量历史数据中，挖掘出设备运行的历史规律和参数间的关联关系等有价值的信息，根据设备运行的历史规律分析当前设备的运行状态，实现设备故障状态的监测和提前报警。

2.2 大数据挖掘

2.2.1 定义

数据挖掘的定义有很多种，其中最被广泛接受的是：数据挖掘是一个从不完整的、不明确的、大量的并且包含噪声，具有很大随机性的，在实际应用数据中提取出隐含其中、事先未被人们获知、却潜在有用的知识或模式的过程^[37]。由上述定义，即可可以得出以下结论：

(1) 数据源必须取自真实应用，不能包含任何虚假数据，这些数据中可能不可避免的包含噪声数据；

(2) 对于所挖掘到的信息，是必须能够被有效利用的；

(3) 挖掘出的信息不可能适用于所有领域。

具体到所研究的设备在线监测与预警诊断系统，通过对数量众多的、且没

有规律的数据处理、分析，挖掘出所有数据背后有价值的规律，进而为设备状况评估供决策依据。

2.2.2 一般过程

首先要需要完成的任务是，确定数据挖掘的最终目标，也是关键任务，然后依据既定的步骤进行即可。本文进行数据挖掘的目的是挖掘设备运行的规律和模式，建立设备健康状态模型，以此为基础判断设备的运行状态进行故障早期预警。

具体来说，数据挖掘可分为三个过程，分别是数据准备、数据挖掘以及结果的解释和评价阶段。其中，第一个阶段在挖掘过程中最为繁杂，它包括数据集成、数据选择、数据清理和数据变换等步骤^[38]，如图 2-1 所示。

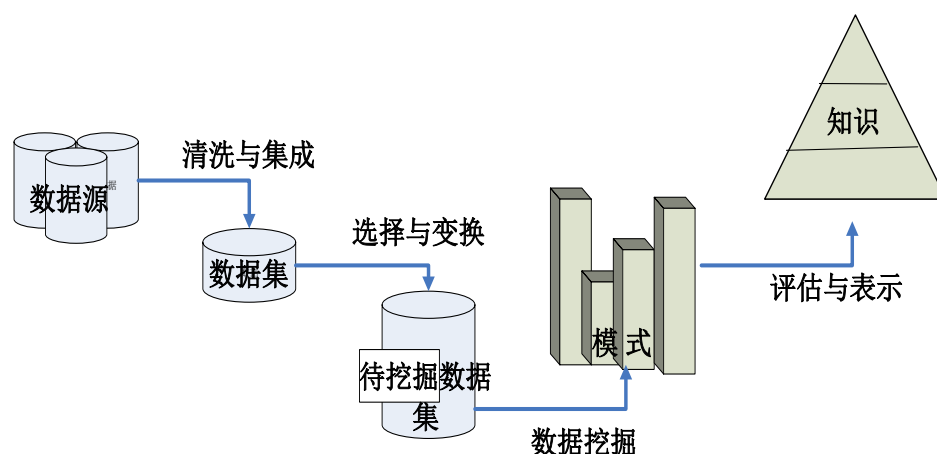


图 2-1 数据挖掘基本过程

（1）数据清理

数据清理指的是，对数量众多的相对表征不明确的，同时夹带噪声的一类数据，视不同情况进行相应清洗的过程。一般要运用到不同领域知识进行初步地判断，然后依据情况的差异，选取最为合适的清洗方法。

（2）数据集成

通常需要把众多极具差异性的数据进行集中，为后续数据处理提供方便，数据集成的关键其实就是解决由于类型、平台等条件不一样所引起的数据差异。

（3）数据选择

数据选择指的是依据数据挖掘的目标，从步骤 2 所得到的数据中，抽取目标数据对象。

（4）数据变换

数据变换的实质即为数据类型转换，以便完成数据挖掘过程。例如，如果进行神经网络计算，需要将数据自离散型转换为连续型，或者进行相反的转换

过程。此外，数据变换具备其他作用，即降低数据维数，缩减数据量，确定出真正有价值的信息。

（5）数据挖掘

数据挖掘是一个多学科交叉技术，不断进行多学科的融合。从统计学角度进行分类，当前主流的数据挖掘方法包括线性和非线性分析、回归分析、最近序列分析，以及时间序列分析等。通过上述方法通常可以确定出异常数据，然后运用数学方法进行分析，就能够挖掘出数据中的潜在信息，这也是基于大数据分析技术实现设备状态监测及预警诊断的核心。

选择合适的算法是这个环节的关键，而选择算法的依据，是由数据特点和最终功能所决定的，只有确定了正确算法，才能从海量的数据当中有效地提取出隐含模式。

（6）知识表示

知识表示是指，对已经被挖掘出的信息予以说明、解释，同时把该信息转换成可以被清晰理解的知识，其中相应知识表示方法包含了两种。

经历了上述的过程，就能够得到有用的信息，如何合理的运用这些信息极其关键。通常来说，有两种方法，第一种，得到的信息能够清晰地描述目标结果，这种情况下便可以直接给出决策支持；另外一种方法，以挖掘出的信息为基础，将这些信息运用到新的数据中，会有发现新问题的情况出现。通常情况下，数据挖掘是要经过循环执行的过程，某一个过程出现了任何状况，均要跳回至前一个过程，等待调整完善后，重新执行。

2.2.3 数据挖掘常用方法

数据挖掘有很多种方法，对于每一种方法其数据挖掘角度均有其自身的特点^[39]。

（1）回归分析

回归分析主要用于对数据序列的预测及相关关系进行研究，它所表征的是数据属性特性，其中运用了数学函数，用其将数据之间的联系清晰地表示，并从中确定数据属性关系。

（2）神经网络方法

神经网络方法具有非线性处理、分布存储等一系列特点，基于其特有的优点，适用于处理表述模糊、不完全表征的数据，所以说神经网络方法有一种解决数据挖掘问题的天赋。

（3）关联规则

关联规则就是基于某个数据项的出现，在此基础上予以推导，得出其他项出现的概率。可以分为两个阶段：（1）从众多原始数据中，搜索出全部高频项目组；（2）从发现的高频项目组中，找出关联规则。

（4）聚类

聚类就是对数据进行分类，其分类依据是数据之前的差别性，对于同类数据而言，差别很小，而对于不同类的数据，差别性非常大。

（5）分类

分类主要应用于应用分类、趋势预测。找出众多数据的共同特性，并将其进行分类，根据不同的分类模型，可以在众多数据中，将每一个数据都对应到一个类别。

（6）Web 数据挖掘

是一种集众多技术的综合性技术，目前常用的 Web 数据挖掘算法有：PageRank 算法，HITS 算法以及 LOGSOM 算法。

第3章 基于大数据的设备状态在线监测与预警诊断系统理论研究

基于大数据的设备状态在线监测与预警诊断系统，其实质是利用大数据技术分析设备长年运行所存储的海量历史数据，找出参数间的关联关系和设备运行的规律模式，建立设备在线监测模型；利用监测模型实时计算设备运行状态，实现设备故障的早期预警。据此分析，设备状态在线监测与预警诊断系统的核心包括两部分：一是参数关联度计算方法；二是获取设备运行的规律模式。

3.1 关联度的计算

一个设备具有很多的参数，如风机，它的参数可能包括电机电流、电机线圈温度、轴承温度、振动值、进出口介质温度和流量等，这些参数间是有复杂的关联关系的。在实际应用中，采用关联度算法对模型进行降维度处理是非常有效的。基于大数据分析的设备状态在线监测系统就是利用计算参数间的关联度这种手段剔除一些对设备运行状态影响不大的测点，对于设备建模测点的选取具有很高的参考价值。本系统采用的是灰色关联度算法，计算电厂关键设备运行参数间的关联。

3.1.1 关联度性质

以字母 S 代表所要研究的系统，以字母 x 和字母 $y \dots$ 代表系统中的序列，同时，这里的系统中不存在总体因素，以下是规定的基本条件^[40]。

(1) 这里采用 $\gamma(x, y)$ 表示序列 x 和序列 y 之间的相互关联度， $\gamma(x, y)$ 的取值范围是 0 和 1 之间，并包含两端数值。考虑到关联度是表征序列之间相互关联程度的，所以既可能出现正值，也可能出现负值，因此结果应该体现出正相关和负相关两种情况。鉴于以上，引入关联符号函数

$$\text{sgn } \gamma(x, y) = \begin{cases} 1 & \text{当 } x \text{ 与 } y \text{ 正相关时} \\ 0 & \text{当 } x \text{ 与 } y \text{ 无关时} \\ -1 & \text{当 } x \text{ 与 } y \text{ 负相关时} \end{cases} \quad (3-1)$$

(2) $\gamma(x, x+b) = 1, \gamma(x, -x) = 1$ ，由此可得，若序列之间存在相互平行的关系，则表示他们必定是相似的，换言之， x 与 $x+b$ 序列应该是完全正相关的，

因此, $\gamma(x, x+b)=1$; 类似地, x 与 $-x$ 之间存在着完全负相关, 因此, $\gamma(x, -x)=1$ 。

(3) $\gamma(x, B)=0$ (B 序列是一个同常数序列; x 序列则是一个异常数序列)。由于无论序列 x 发生何种变化, 都不能改变序列 B 的状态, 所以两序列完全不相关。

(4) 设 $x_0, x_1 \in S$, 则 $\gamma(x_1, x_0)=\gamma(x_0, x_1)$ 。关联作用是相互的, 序列 x 改变时, 序列 x_1 会发生相应的变化; 同样地, 序列 x_1 变化时, 序列 x 也应当发生相应的变化; 也就是说关联度具有对称性。

(5) 设 $x, x' \in S, \gamma(x, x')=1$ 的充要条件是, $y \in S$, 有 $\gamma(x, y)=\gamma(x', y)$, 也就是说完全相关具有传递性。

3.1.2 关联度的计算

(1) 假定要计算关联度的两序列分别为 x, y 。

(2) 对 x, y 两序列进行无量纲处理。由于 x, y 两序列可能代表着不同的意义, 所以量纲未必相同, 以至于比较起来不是很方便, 甚至无法得到准确结论。所以, 当计算关联度之前, 需要对数据进行预处理, 即通常所说的无量纲化。

(3) 求 x, y 两序列之间的关联系数 $\xi(x_i)$, 如公式 (3-2) 即为其计算公式。其中, ρ 表示的是分辨系数, 一般取值范围位于 0 和 1 之间, 通常取为 0.5 即可。

$$\xi(x_i) = \frac{\Delta(\min) + \rho\Delta(\max)}{\Delta_{0i}(k) + \rho\Delta(\max)} \quad (3-2)$$

式中 $\xi(x_i)$ ——序列关联系数

$\Delta\min$ ——第二级最小差

$\Delta\max$ ——两级最大差

$\Delta_{0i}(k)$ ——各比较数列 X_i 曲线上的每一个点与参考数列 X_0 曲线上的每一个点的绝对差值

(4) 求关联度 r_i

该参数的计算应用了公式 (3-3), 计算结果表示的是不同序列之间在某时刻的关联度。

$$r_i = \frac{1}{N} \sum_{k=1}^N \xi_i(k) \quad (3-3)$$

式中 r_i ——序列 x 对 y 的灰关联度。

r_i 值越接近 1，则代表序列 x 与 y 的相关性越好^[41-43]。

3.2 运行模式的计算

电力生产领域的数据大都是时序数据，该类型可以说是一种特殊但也极其重要的数据。将大数据分析技术与设备状态在线监测技术相结合，其目的其实就是利用该技术挖掘出设备运行潜在规律，以此为基础实现设备运行状态的监测。在此过程中所挖掘的数据就是时序性数据，可以对设备的运行规律进行了解，从而对其规律进行深入研究，

针对时序数据挖掘，当前主流研究有 3 方面，而本文采用了聚类的方法，借助函数型数据分析的基本思想和方法，进行深入研究。它是以数据之间的差异性作为基础，将众多数据进行分类，被分为一类的数据之间的差异性是很小的，换言之就是有着很高的相似性，而对于不同类别数据之间，彼此之间差别很大，即有着非常低的相似性。聚类分析没有特定的分类标准，往往根据数据自身的特点分类，同时，其分类方法种类众多，因此，具有很高的灵活性，能够有效探索数据之间潜在关系^[44]。

将聚类方法应用于函数型数据的研究进展相对缓慢。运行模式的计算就是以基函数展开作为工具，对函数型数据予以聚类。在数学中，有这样一个被证明了的事实，只要将原始函数在标准正交基函数上展开，则原始函数之间的欧式距离与基展开系数向量之间的欧式距离是一致的；如果在非正交的基函数上展开，则只需对系数向量之间的距离进行一定修正，依然可以得到一致的结论。以上表明，为数众多的聚类方法是能够被运用到时序数据中去的，进行相应的分析工作。同时，该方法还被应用到多变量函数型的数据中去，把函数型数据聚类分析方法，运用到设备在线监测领域，借助聚类结果实施设备监测，能够有效降低设备运行的风险。

3.2.1 基本聚类挖掘方法

通常来说，算法可以被分为五类，分别为划分法、层次法、基于网格的方法、基于密度的方法和基于模型的方法等^[45]。

(1) 划分法

假定有数据集 S ，它有 N 条数据记录。划分法会将这 N 条数据记录进行划分，得到 K 个数组，每个数组都可称为一个聚类。并且，这 K 个数组满足以下条件：第一，每个数组包含的数据记录不能为空；第二，一个数据记录不可能分属于不同的数组。

当提前设置要分组的个数 K 之后，划分算法就会进行初始分组；然后，利用迭代方法不断改进分组情况，最终达到最佳状态。这个所谓的最佳状态就是：同一数组内，数据记录应该保持足够小的距离；而对于不同数组之间，应该保持足够大的距离。

划分有其自身的优点，具有优良的伸缩性和很高的效率，当数据分布情况比较均匀时，所能达到的效果是非常可观的。但是，该方法需要人工设定 k 值，这就使得在某些情况下效果并不可观。与此同时，该类算法对噪声表现出了非常高的敏感性，离群数据对其也有很大的影响，对聚类效果产生的影响很大。因此，在实际应用中，应用极少。

（2）层次法

其中所涉及的算法非常多，本文重点介绍一下 CURE 算法。

该算法的关键步骤如下：

第一步：随机样本的抽取，样本以字母 S 表示；

第二步：进行样本划分；

第三步：针对于每个划分，采用某种方法实施聚类；

第四步：排除离群点，并去除增速缓慢的类；

第五步：重复步骤 3，实施局部再聚类；

第六步：进行数据标记。

相较于其他算法，本方法优势明显，算法较容易，在离群点的数据处理方面更加强大，不仅有着很好的伸缩性，还具备很好的效率，对于噪声不敏感。同时，也存在一定的不足，即在分类属性问题上没有优势，不具备该种处理能力。

（3）基于密度的方法

相较于其他方法具有本质上的不同，因为该方式数据聚类的依据是并非以距离作为基础，而是密度。也正因为如此，可以很好的解决前者的缺点。该方法也派生出了众多算法，DBSCAN 算法是应用最广泛的一种算法。

其优点在于无须提前确定目标聚类个数，有着很好的噪声抵抗力，与此同时，还可以进行过滤，利用该方法可以不受聚类形状、大小的限制，同时也存在一定的缺点和不足，即有着非常高的时间复杂度，并且受设定参数的影响非常大。

（4）基于网格的方法

该方法是以单元为基础，把众多数据进行划分，成为网络结构。包含多种算法，本文将重点介绍 STING 算法，该方法主要是把数据划分为矩形单元。本方法运行速度是非常快的，其速度不受数据数量的影响，而是同划分的单元总

数有关，其中计算和查询相互独立，互不影响，其结构形式有助于进行并行处理。与此同时，该方法也存在缺点和不足，聚类质量受最底层粒度的影响较大，并且聚类精度不是非常可观。

（5）基于模型的方法

首先对每个聚类给定一个假设的模拟，以该模型为基础，在众多数据中进行搜索，确定出于模型匹配的所有数据。最常用的方法是两类，分别利用了神经网络和统计学。在众多的算法中，COBWEB 算法相对来说最具代表性，它是基于增量概念的方法，简单、清晰，采用此方法会产生树状图。采用此种算法的优点是，有着非常可观的聚类效果，同时由于是依靠树状图进行聚类，所以能够清晰、直观的了解聚类的整个过程和最终结果。与此同时，本算法也存在一定的缺点，结果受原始数据质量的影响较大，当数据量过于庞大，该算法表现出较高的复杂度。

3.2.2 函数型数据的相似性指标

对于所有的聚类问题，对象之间存在一个非常关键的指标需要确定，即距离。给定了 $x(t)$ 、 $z(t)$ ，衡量距离常用指标包括：

（1）差异上确界

$$D_{xz}^{(2)} = \sup \{ |x(t) - z(t)| \} \quad (3-4)$$

（2）一致差异

$$D_{xz}^{(2)} = \int_0^T |x(t) - z(t)| dt \quad (3-5)$$

（3）欧式距离

$$D_{xz}^{(3)} = \int_0^T [x(t) - z(t)]^2 dt \quad (3-6)$$

鉴于欧式距离有着非常优异的性质，所以也是用做衡量函数相似的非常普遍的参数。然而该种参数只能进行曲线差异的表征，对于曲线形态的差异性并不能有效的进行表征，所以，便提出了基于形态的相似性指标和基于秩相关的相似性指标，但是在运用以上两种方法进行计算时，需要完成大量的积分运算，从来限制了运算速度，因此这两种算法在聚类算法中有着先天不足，很难在聚类运算中采用。因此，欧式距离虽然存在着缺陷，但依然是首选。而且它的这些缺陷可以通过对函数事先进行标准化或者通过对函数特定阶导数进行聚类等方式来克服。综上所述，仍然选择欧式距离作为函数聚类主要的相似性度量指标。

3.2.3 基于基函数展开的函数型数据聚类方法

计算过程中，如果直接进行欧式距离的计算，则过程会变得非常的繁琐，原因是引入了大量积分运算，不仅增加了完成计算的时间，同时也加大了复杂程度[46]。为此，这里采用了基函数展开系数向量的距离代替原函数之间的距离，将问题进行了转换。于是，函数型数据的聚类问题就被转化成低维空间中基函数展开系数向量的聚类分析，使问题更加简单化。其主要计算步骤是：

- (1) 采用某种基函数，对数据进行展开，最终得到系数向量；
- (2) 选取某种聚类方法，依据加权或未加权的欧式距离，进行系数向量聚类。

以上，就是解决时序数据聚类问题的方式，主要在两个方面表现出了其一般性，首先是选取基函数具有很大的灵活性，可以根据实际情况采用不同的函数进行数据展开；其次，以欧式距离作为基础的聚类方法具有很大的适应性，能够被有效作为时序数据聚类分析的手段。

3.2.4 利用相似系数聚类

相似性理论算法概述：设向量 $X = \{x_1, x_2, \dots, x_n\}$ 、 $Y = \{y_1, y_2, \dots, y_n\}$ ，当其相似性函数 $\text{sim}(X, Y) < \varepsilon$ ，（ ε 为给定的相似性阈值），称向量 X ， Y 相似，并记为 $X \approx Y$ ^[47]。 $\text{sim}(X, Y)$ 满足正定性、对称性和三角不等式。

- (1) 正定性， $\text{sim}(X, Y) \geq 0$ ，当且仅当 $X = Y$ 时， $\text{sim}(X, Y) = 0$ ；
- (2) 对称性， $\text{sim}(x, y) = \text{sim}(y, x)$ ；
- (3) 满足三角不等式， $\text{sim}(a, c) = \text{sim}(a, b) + \text{sim}(b, c)$ 。

当取 $\text{sim}(X, Y)$ 为 X 、 Y 之间的欧氏距离 $d(X, Y)$ ，即 $\text{sim}(X, Y) = d(X, Y)$ ，当 $d(X, Y) < \varepsilon$ 时，称向量 Y 在欧氏距离下以 ε 为界相似。

对于两个向量，经过计算得出的欧式距离能够表征向量的差异性，程度与数值大小成正比，最大数值为 1。在很多情况中，向量之间相似程度是具备一定的比较难度的，所以提出了相似度的概念对向量接近程度进行表征，取值范围 0-1，基于该方法可以实现设备测点状态估计。

任意两向量 $X = \{x_1, x_2, \dots, x_n\}$ 、 $Y = \{y_1, y_2, \dots, y_n\}$ ， $\text{sim}(X, Y)$ 为 X 与 Y 之间的相似性函数，给定阈值 $\varepsilon > 0$ ，若 $X \approx Y$ ，定义 $s(X, Y) = 1 - \text{sim}(X, Y)/\varepsilon$ 是 X 、 Y 的相似度。

3.3 本章小结

本章主要介绍了设备状态在线监测与预警诊断系统核心算法的理论性研究，提出采用灰色关联度算法计算参数间的关联关系，运用相似系数聚类方法，获取了设备运行模式。这两大理论算法的提出，将有效解决设备测点选取和运行规律重新划分的关键问题，为在线监测核心算法提供了理论依据。

第 4 章 电厂关键设备氧化风机模型的构建

4.1 模型创建

(1) 初选测点

氧化风机安装于机房中，采用了罗茨风机，该设备的作用是，将氧化空气送入到吸收塔浆池中，氧化风机结构较为复杂，由众多部件组成。基于其组成部分，将整个氧化风机的测点进行初步统计，去除开关量测点，初步选择的测点包括：氧化风机电流、氧化风机电机 A 相线圈温度 1、氧化风机电机 A 相线圈温度 2、氧化风机电机 B 相线圈温度 1、氧化风机电机 B 相线圈温度 2、氧化风机电机 C 相线圈温度 1、氧化风机电机 C 相线圈温度 2、氧化风机电机前轴承温度、氧化风机电机后轴承温度、吸收塔底部液位 1、吸收塔氧化空气压力、吸收塔底部液位 2、3 等。

(2) 数据抽取

数据抽取是指，将源数据源系统当中的有用数据提取出来，依照目的数据源系统的需要进行提取。实际情况下，数据源较多采用了关系数据库。具体到火电厂，生产设备运行的历史数据并非存储在关系型数据库中，而 DCS、PLC 等控制系统，仅会存储少量的历史数据。通过了解，设备长年运行的历史数据大都存储在电厂 SIS 数据库中，要想获取设备长年运行的历史数据进行分析，可以从 SIS 数据库中采集历史数据。

系统采集的历史数据需要符合如下要求：第一，历史数据必须涵盖了一段足够长的运行时间（一般为一到五年时间），保证涵盖机组不同负荷、不同环境下的运行状况。第二，对于每组采样值，各变量必须保证是在同一时刻的采样值，这样才能有效反应多变量在不同运行情况下的统计关联关系。

基于历史数据采集的要求，以某厂所用 SIS 数据库 PI 为例，进行数据抽取的过程如下：

第一：开发数据采集接口

通常，数据库厂家都会提供开放的数据采集规范 API 文档。按照文档说明，编写数据采集接口程序即可。

第二：抽取历史数据

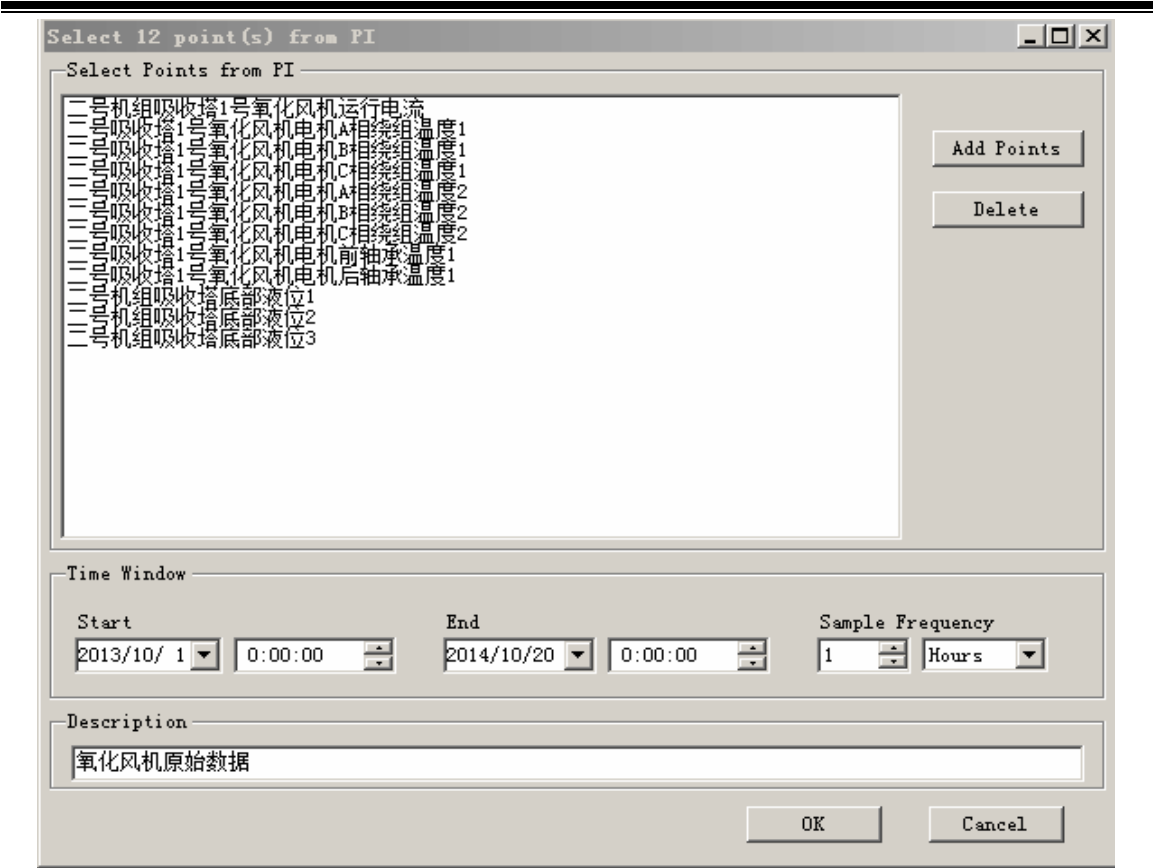


图 4-1 历史数据采集界面

以上是开发好的数据采集界面，将历史数据采集的相关参数设置好，程序即可自动采集数据分析所用的历史数据。需要设置的参数包括：设备测点、数据采集起止时间、间隔和数据集名称等。其中数据采集间隔设置为 1 小时即可。参数设置好，系统即可采集到历史数据。

第三：数据展示

Row	TimeStamp	一号机组吸收塔1号氧化风机运行电流	一号机组吸收塔底部液位1	一号机组吸收塔底部液位2	一号机组吸收塔底部液位3	一号吸收塔1号氧化风机电动机A相绕组温度1	一号吸收塔1号氧化风机电动机B相绕组温度1	一号吸收塔1号氧化风机电动机C相绕组温度1	一号吸收塔1号氧化风机电动机A相绕组温度2
1	2013/10/1 0:00:00	18.079269409...	71.908630371...	74.506225585...	73.977531433...	61.201808929...	61.206897735...	62.331890106...	60.050949096...
2	2013/10/1 1:00:00	18.360414505...	73.303245544...	76.083892822...	75.499183654...	61.443658921...	61.485794067...	62.671474456...	60.236894335...
3	2013/10/1 2:00:00	18.079570770...	72.322921752...	74.880531311...	74.266235351...	60.796249389...	60.841655731...	62.037784576...	59.707706451...
4	2013/10/1 3:00:00	17.784284591...	70.442092895...	72.899513244...	72.341323852...	59.804492950...	59.949787139...	61.040142059...	58.713657379...
5	2013/10/1 4:00:00	17.615983963...	68.907516479...	71.219284057...	70.888465881...	58.164043426...	58.246910095...	59.164478302...	57.061672210...
6	2013/10/1 5:00:00	19.031240463...	77.215682983...	79.676391601...	79.619796752...	60.158489227...	60.241115570...	61.385123748...	58.974811553...
7	2013/10/1 6:00:00	18.343690872...	73.072074890...	75.278274536...	74.722595214...	60.321060180...	60.360549926...	61.524486541...	59.174938201...
8	2013/10/1 7:00:00	18.268489837...	72.239990234375	74.540153503...	74.167610168...	59.923652648...	59.919509887...	61.034942626...	58.776363372...
9	2013/10/1 8:00:00	18.352825164...	72.405105590...	74.720886230...	74.426504943...	60.196842193...	60.358914520...	61.387585612...	59.068302154...
10	2013/10/1 9:00:00	18.404220581...	73.267471313...	75.386520385...	75.023735046...	60.082637786...	60.244148254...	61.388584136...	58.962238311...
11	2013/10/1 10:00:00	18.201646804...	72.207443237...	74.574966430...	74.039184570...	60.584285736...	60.594421386...	61.630542755...	59.470397949...
12	2013/10/1 11:00:00	18.074081420...	71.837852478...	74.239364624...	73.801177978...	61.020320892...	61.084995269...	62.257900238...	59.948341369...
13	2013/10/1 12:00:00	17.919193267...	70.849388122...	73.464485168...	72.838088989...	61.283325195...	61.313377380...	62.410072326...	60.174938201...
14	2013/10/1 13:00:00	17.695390701...	69.653457641...	71.898086547...	71.319869995...	61.190464019...	61.265842437...	62.368778228...	60.172332763...
15	2013/10/1 14:00:00	17.752542495...	70.406036376...	72.654083251...	72.095199584...	61.888755798...	61.966217041...	63.031265258...	60.767562866...
16	2013/10/1 15:00:00	17.837560653...	70.707046508...	72.910293579...	72.325576782...	61.987545013...	61.933933258...	63.231124877...	60.784519195...
17	2013/10/1 16:00:00	18.133537292...	72.696838378...	74.989952087...	74.494689941...	63.197280883...	63.191593170...	64.174659729...	61.993686676...
18	2013/10/1 17:00:00	17.947950363...	71.738311767...	74.119178771...	73.534736633...	63.299686431...	63.304588317...	64.255838122...	62.057830810...
19	2013/10/1 18:00:00	17.998741149...	71.980979919...	74.540596008...	73.916435241...	62.585528869...	62.586135864...	63.527309417...	61.315929412...
20	2013/10/1 19:00:00	17.780591964...	70.073478698...	72.615364074...	71.891418457...	61.753711700...	61.760822296...	62.981212615...	60.637645721...
21	2013/10/1 20:00:00	17.516229629...	69.140625	71.625007629...	70.956863403...	60.983844757...	61.029265430...	62.155017852...	59.865608215...
22	2013/10/1 21:00:00	18.764915466...	74.972595214...	77.821525573...	77.443138122...	62.226474761...	62.111038208...	63.385620117...	60.943130493...
23	2013/10/1 22:00:00	18.348661422...	73.400276184...	75.856201171875	75.392250061...	62.290065765...	62.237369537...	63.428966522...	60.996871948...

图 4-2 历史数据展示

从图中可以看出，采集到的是以时间为基准的历史数据组，满足数据分析要求。

（3）数据清洗

指发现能够被有效识别的错误，并进行相应的纠正，它是过程中的最后程序，包括清除噪声、推导计算填补缺省和不完整数据、修正异常数据和清除重复数据。具体到在线监测系统，需要基于设备正常运行时的数据建立模型，但是通过 SIS 数据库采集到的历史数据可能包含了设备非正常运行状态的数据、噪声数据以及设备停运数据，故需对数据进行分析处理，挖掘出建模所需要的准确的、科学的有效数据，确保所建模型的准确性。

基于工程实践，系统决定采用两种数据清洗方式并用的方法清理掉大量无用数据，这两种方法可以命名为测点趋势法和关联关系比较法。

方法一：测点趋势法

测点趋势法就是查看每个测点的历史数据趋势变化，若某些样本严重偏离大部分数据所呈现的统计特性，则这些样本数据将被清洗掉。

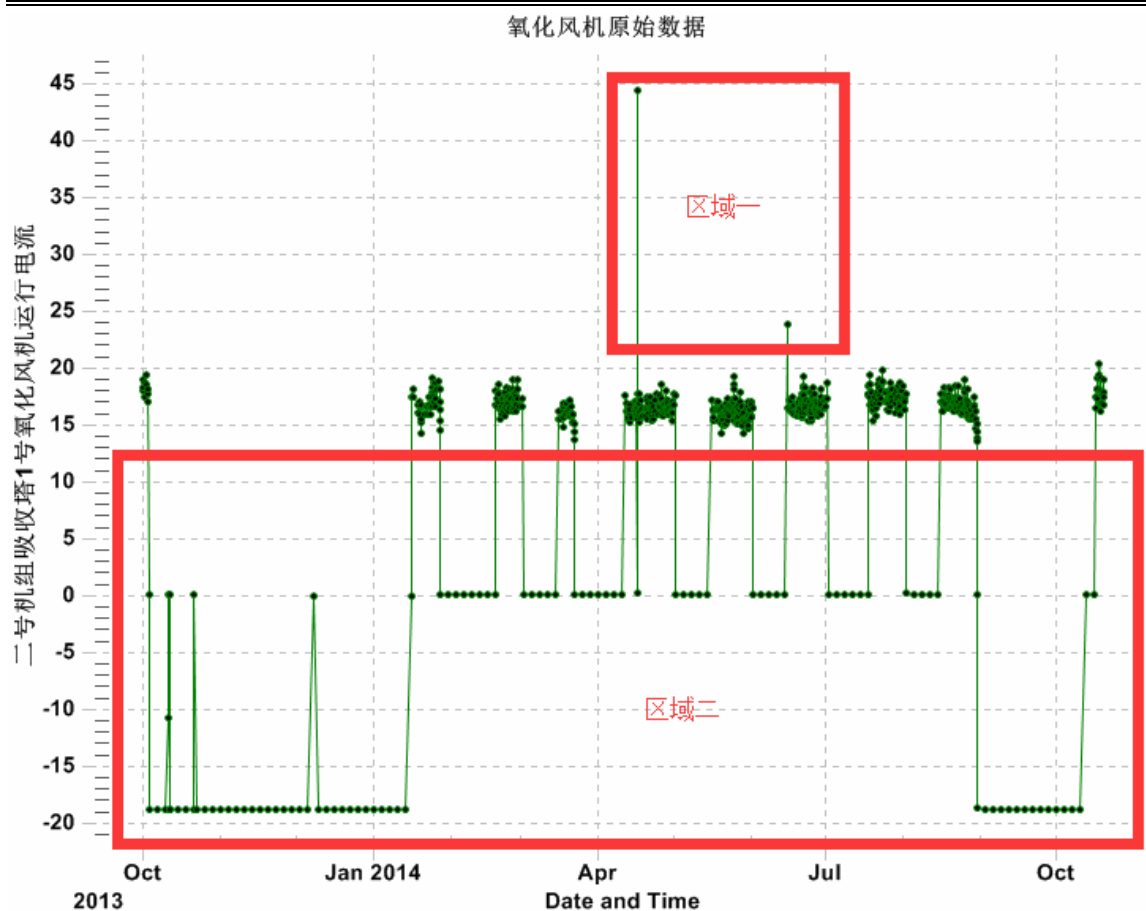


图 4-3 氧化风机电机电流趋势图

上图是氧化风机电流的历史趋势，从图中可以看出，氧化风机是定期轮换运行的，区域二的数据大都是设备停运阶段的数据，区域一的数据属于噪声数据，可以将这些数据清洗掉。同样地，利用此方法查看其它测点的历史趋势，清洗掉这些噪声、错误、不符合要求的数据。

方法二：关联关系比较法

关联关系比较法就是将每两个测点的历史数据投射到二维空间，直观的展现设备运行状态的情况；所实现的是对于存在耦合和相关性的参数，历史测点数据中的某些样本的相关性系数若严重偏离大部分数据所呈现出的相关性特性，则这些样本数据将被清洗掉。

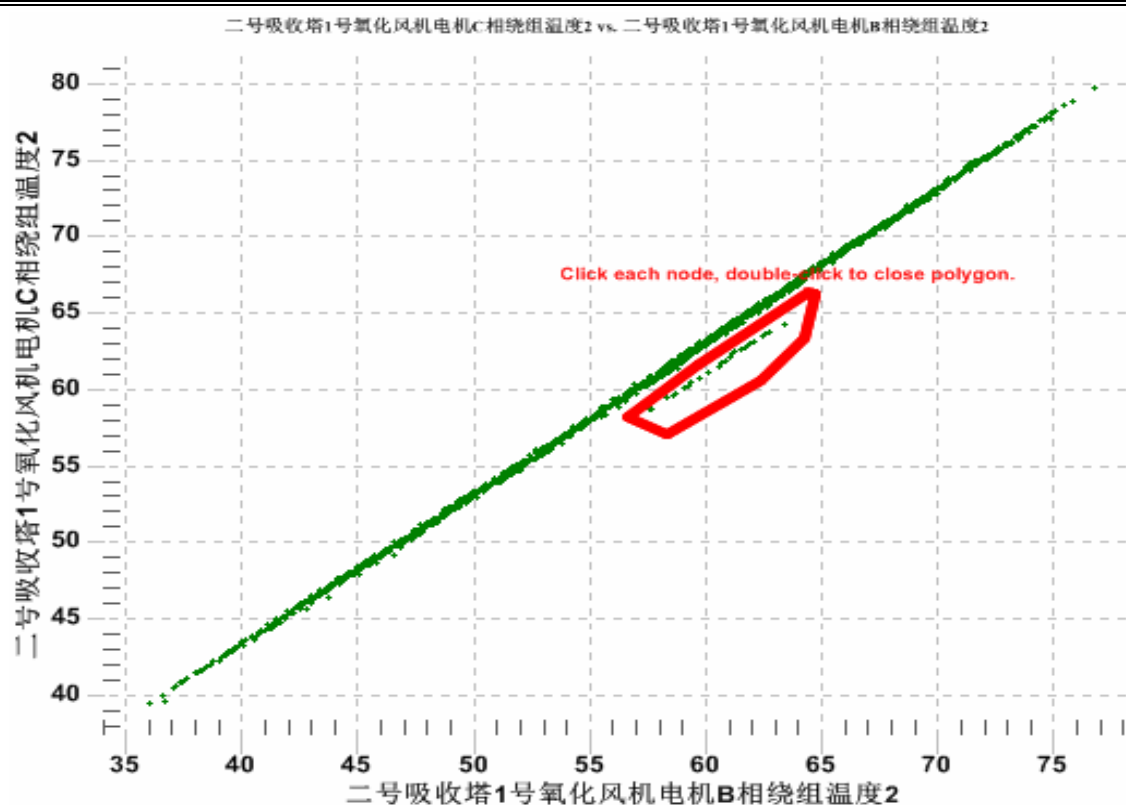


图 4-4 关联关系对比图

上图中，将氧化风机 C 相绕组温度和氧化风机 B 相绕组温度做了关联关系对比。从中可以看出，设备包括两种运行情况，其中红色区域的运行情况较少发生，通过查看设备运行历史记录发现这些运行情况是由于设备异动产生的，故认为是设备非正常运行状态的历史数据，数据清洗时需要将这些数据清除。通过关联关系的比较，可以方便的看出设备历史上不同的运行情况，包括检修前后运行情况的改变、设备异动和设备故障时运行情况的不同等。

（4）数据无量纲化处理

对于所有的设备模型，所表征的规律是需要不受量纲影响的，为了消除这种影响作用，就需要对数据进行处理，全部处理成无量纲的形式，以方便对于设备规律的正确描述。根据实际需要，可以采用直线型处理方法，对所有数据进行处理。其中可以采用极值法或者标准差标准化法，本文采用了极值法。

（5）数据关联度分析

数据清洗完成后，就得到了设备正常运行状态的数据，然后对这些数据进行无量纲化处理，以便分析测点关联度。而分析测点关联度的最终目的是，清除设备建模的冗余测点。通过大量的工程实践，即可发现如果某个测点与其他测点的关联度都低于 20%，则这个测点在设备建模时，可以不予考虑。测点间的关联度如下表 4-1 和 4-2 所示。

表 4-1 设备测点关联度分析表

机组设备测点	运行电流	A 相绕组温度 1	B 相绕组温度 1	C 相绕组温度 1	A 相绕组温度 2
运行电流	100%	20.90%	20.84%	21.55%	20.04%
吸收塔液位 1	89.26%	38.68%	38.56%	39.27%	38.02%
吸收塔液位 2	89.74%	35.64%	35.63%	36.53%	35.23%
吸收塔液位 3	90.23%	34.15%	34.06%	34.73%	33.40%
A 相绕组温度 1	20.90%	100.00%	100.00%	99.99%	99.98%
A 相绕组温度 2	20.04%	99.98%	99.98%	99.98%	100.00%
B 相绕组温度 1	20.84%	99.98%	100.00%	99.99%	99.96%
B 相绕组温度 2	21.52%	99.99%	99.98%	99.99%	99.96%
C 相绕组温度 1	21.55%	99.99%	99.99%	100.00%	99.95%
C 相绕组温度 2	21.40%	99.95%	99.95%	99.95%	99.95%
电机前轴承温度	6.01%	98.73%	98.78%	98.62%	98.85%
电机后轴承温度	10.01%	99.02%	99.03%	98.93%	99.07%

表 4-2 设备测点关联度分析表

机组设备测点	B 相绕组 温度 2	电机前轴承 温度	电机后轴承 温度	吸收塔 液位 1	吸收塔 液位 2	吸收塔 液位 3
运行电流	21.04%	21.44%	10.01%	89.28%	89.74%	90.23%
吸收塔液位 1	39.25%	39.15%	29.58%	100.00%	99.73%	99.34%
吸收塔液位 2	36.49%	36.37%	27.05%	99.73%	100.00%	99.54%
吸收塔液位 3	34.65%	34.65%	25.11%	99.24%	99.54%	100.00%
A 相绕组温度 1	99.98%	99.95%	99.02%	38.68%	35.94%	34.15%
A 相绕组温度 2	99.99%	99.95%	99.07%	38.02%	35.23%	33.40%
B 相绕组温度 1	99.97%	99.95%	99.03%	38.56%	35.83%	36.06%
B 相绕组温度 2	99.97%	99.95%	99.91%	39.25%	36.48%	34.65%
C 相绕组温度 1	100.00%	99.95%	98.84%	39.27%	37.13%	34.73%
C 相绕组温度 2	99.96%	100.00%	98.95%	39.15%	37.78%	34.65%
电机前轴承温度	99.96%	98.56%	99.62%	25.14%	38.43%	21.40%
电机后轴承温度	99.95%	98.84%	100.00%	29.58%	39.08%	25.11%

(6) 创建模型

数据清洗的主要任务是获取正常运行状态的历史数据；关联度分析的目的则是清除设备建模时的冗余测点；这两个过程是建模前最为关键和重要的部分，完成后即可进行设备建模。设备建模的过程如下：

将设备正常运行时的所有历史数据，对应的状态点均映射到一个状态空间中，基于此中的边界点，构造外接的超球，便能够包括历史数据中的全部正常工作状态点，即为初级设备对象模型。

根据生成模型的尺度要求，系统筛选出差别大于某一个相似性指标的状态点，并将其作为关联参照点，通过对参照点的设置，能够对设备动态模型实现高度的细致划分，最终形成一个精确模型。

精准的数学模型其实可以理解为一个个的运行剖面，如图 4-5 所示。其中，运行剖面可以理解成一种极坐标，每个测点代表一个极，每个测点在运行剖面中都有一个范围，并可在相应的极坐标中标注出最大最小值。将所有测点的最大值连接，即为图中的绿线，将所有测点的最小值连接，即为图中的红线。红线和绿线之间的范围就代表了设备的一种运行剖面。氧化风机模型由 3.2 节算法计算得出，与设备模型建模所用历史数据，设备运行模式等有关。针对本氧化风机，系统计算出 300 余种运行剖面作为氧化风机模型的一部分。

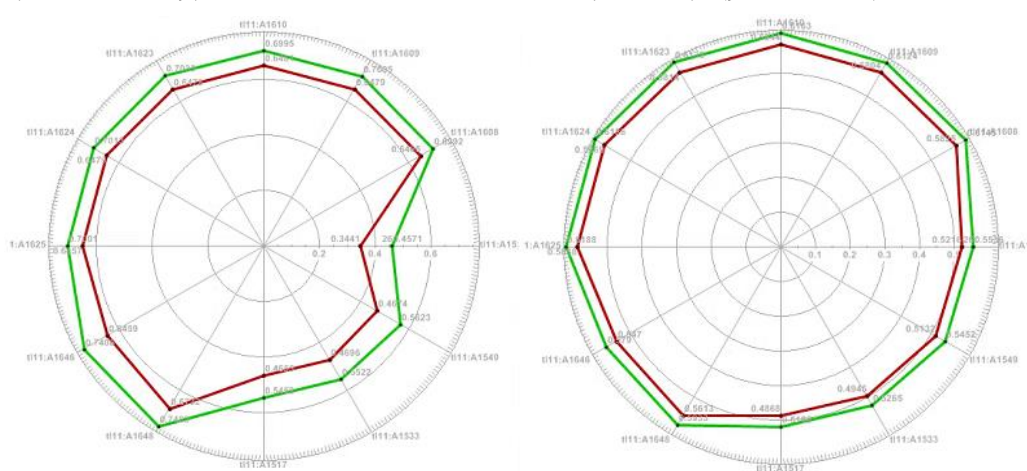


图 4-5 模型剖面图

4.2 模型优化及调整

(1) 数据过滤

设备模型主要用于监测设备运行状态，对设备启停阶段不做监测，故需要设置过滤值，以过滤设备非运行状态的值。过滤点应该选择在运行状态变化较少的测点，比如电流、转速等测点。氧化风机所用测点为电机电流，过滤值条件为电机电流大于 13。也就是说，只有电机电流大于 13 的时候，设备模型才会对设备运行状态就行评估，当电机电流小于 13 的时候，模型认为设备在启停阶段，不需要进行状态评估。

(2) 历史数据回放

历史数据回放就是将设备某段时间的历史数据与模型数据进行对比，回放设备的运行状态，其主要目的是验证模型是否可用。拿设备正常运行的数据与模型对比，模型不应出现误报警；拿设备异动或者故障时的数据与模型对比，

模型应该提前发出预警，这是验证模型是否可用的最好方式，也是状态在线监测与预警诊断的终极目的。如下图，拿建模时间的原始数据与模型对比，由于原始数据并未出现大量的异常数据，设备运行正常，只是有少量的噪声数据，所以模型不应该出现报警，而回放的最终结果也是模型并未发出预警，证明模型可用。

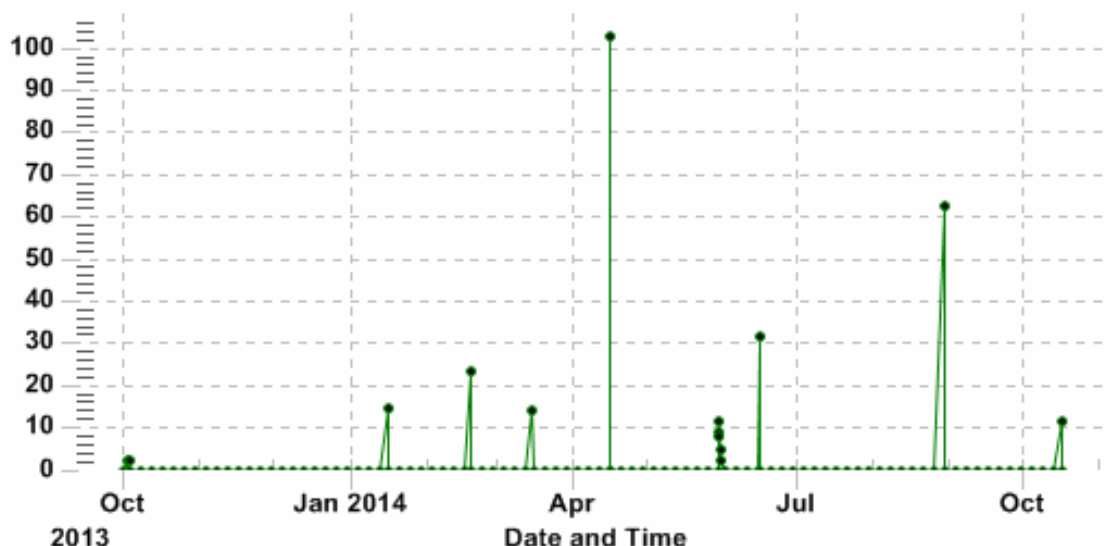


图 4-6 历史回放数据

4.3 模型运行情况分析

模型创建完成之后即可接入实时数据，在线评估设备运行情况。在线监测评估应包括两部分：单个测点值的评估及设备整体运行状态的综合评估。

(1) 单点评估

系统针对每个测点会产生动态预测带，随生产实时运行状态不断变化。如图 4-7，是电机电流的单点评估曲线，其中红色曲线代表实际测量值，蓝色曲线代表预测值上限，绿色曲线代表预测值的下限，蓝线和绿线之间的动态带即为设备理想运行状态带。如果实际测量值一直在理想运行状态带里运行，则表明设备运行状态良好；如果实际测量值超出理想运行状态带运行，则表明设备正偏离正常运行状态，当偏离过大时可能造成设备异动甚至故障。

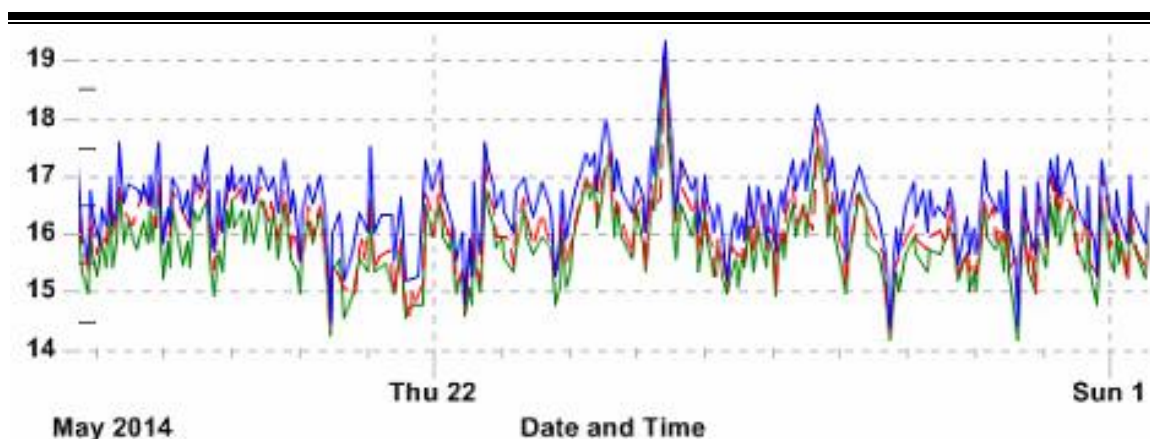


图 4-7 单点运行曲线

(2) 健康度评估

采用健康偏离度的概念评估设备整体运行状态，健康偏离度采用欧式距离的方式计算，即将各个测点的偏离情况进行欧式计算，得出设备健康状态的偏离，如下图所示。由于设备在运行过程中的波动时难以避免的，需要根据历史数据确定设备运行状态的波动情况，以此设置健康偏离度的报警值。

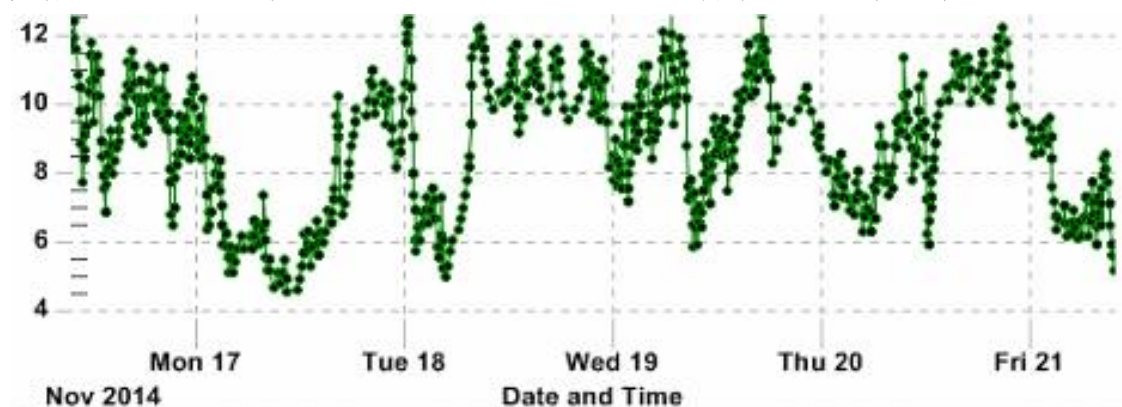


图 4-8 健康偏离度曲线

此外，还要进行报警规则设置，可以根据不同设备选择不同的报警规则，减少误漏报等现象的发生。报警规则包括但不限于以下几种情况：

- 第一种情况：连续 N 次超出报警阈值报警；
- 第二种情况：连续 N 次事件中，有 $N-1$ 次事件超出报警阈值报警；
- 第三种情况：超出报警阈值即报警；
- 第四种情况：持续超出报警阈值达一段时间后的报警；
- 第五种情况：混合报警方式。

针对氧化风机，选择第二种报警方式，即设置连续 5 次监测结果，有 4 次超出报警阈值，模型才会发出警报，这样能大大减少预报警等问题的产生。

4.4 预警分析

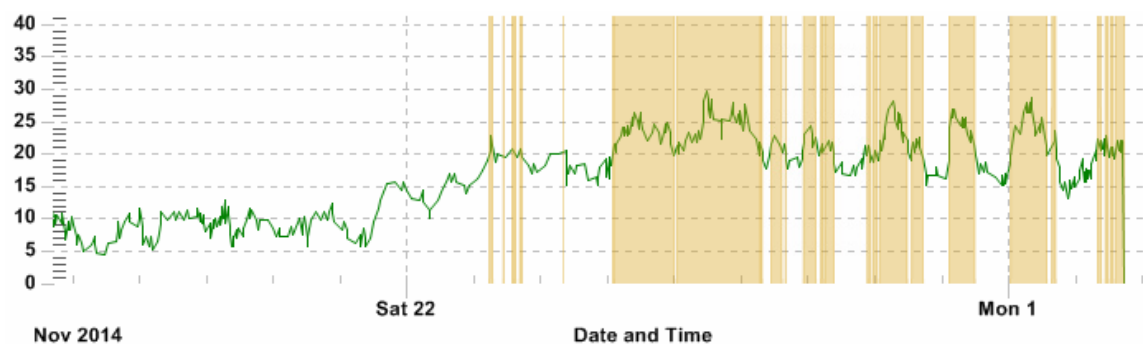


图 4-9 氧化风机设备健康偏离度

氧化风机的设备模型从 2014 年 10 月 25 日开始运行,用于评估设备实时运行状态。在 2014 年 11 月 22 日,氧化风机开始发出设备异动早期预警,如上图所示,绿色曲线为氧化风机模型健康偏离度曲线,当设备健康度高于 15 且满足预警规则引擎时,系统即发出预警信息。

从健康偏离度的计算方法,可知只有设备的某些测点出现较大偏离时,才会引起设备异动早期预警。因此,可以通过找出偏离较大的测点来逐步确认异动预警原因。为了快速查找出这些测点,建立了一套寻找偏离测点的简单算法,并直观展现出来,如下图所示,可以看出,引起当前报警的测点为氧化风机机电电流和吸收塔氧化空气压力。

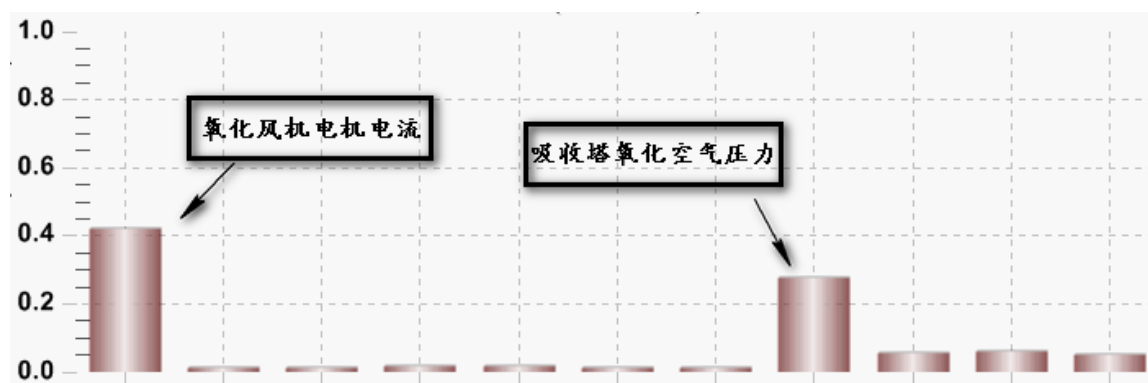


图 4-10 分析工具

找到引起设备异动预警的主要影响测点后,就可以分别查看这几个测点的实际值、预测值及偏差值情况,进一步分析设备异动的原因。氧化风机电流的偏差值如下图。通过偏差值得知,氧化风机电流高于正常值。同样地,吸收塔氧化空气压力高于正常值。基于此,简单列出氧化风机模型异动预警的可能原因:氧化风机出口支路有个别管路堵塞;氧化风机进口过滤器需要清洁;氧化风机出口门未全开或者堵塞;吸收塔液位过高;吸收塔密度过高;氧化风机出口消声器可能有杂物堵塞。

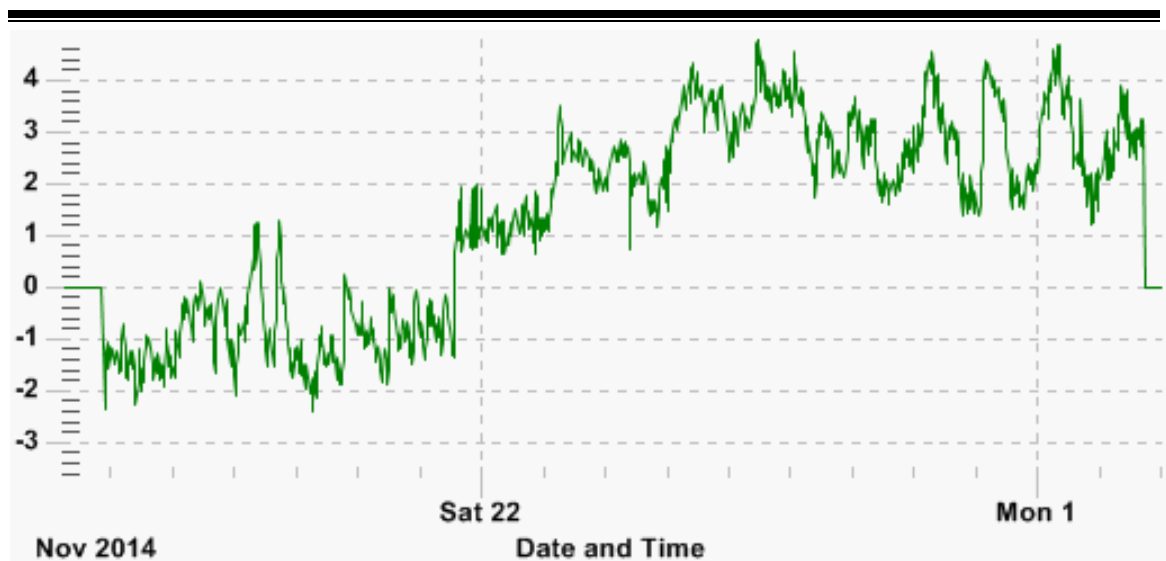


图 4-11 氧化风机电流偏差值

初步诊断，找出可能的原因之后，需要通过进一步的专业分析，排除部分可能性，以锁定预警的原因。

首先，查看过去一段时间吸收塔液位曲线，如下图。从图中可以看出：从 10 月起，吸收塔液位均保持较高水平，而模型从 11 月末才开始预警，排除吸收塔液位过高引起模型预警的可能。

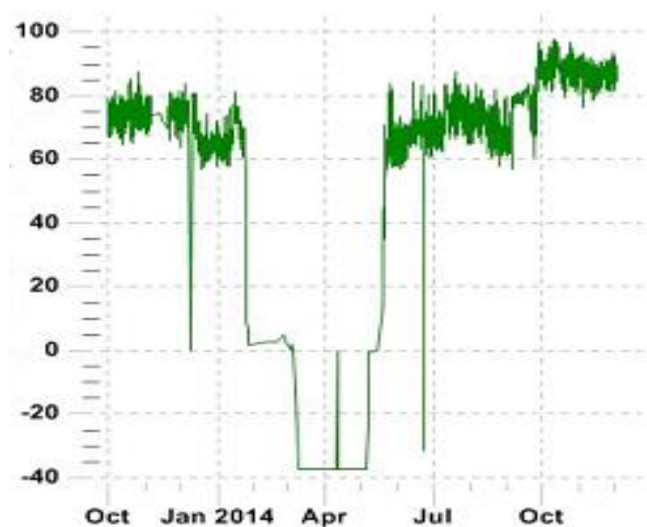


图 4-12 吸收塔液位曲线

然后，查看过去一段时间吸收塔密度曲线。由于吸收塔近期密度与之前相比处于较低水平，故排除吸收塔密度过高引起模型报警的可能。

最后，锁定故障。设备定期轮换时，2 号氧化风机运行，1 号氧化风机停止，但是 2 号氧化风机模型产生同样的预警信息，表明 1、2 号氧化风机模型预警的原因为公用部分故障，排除氧化风机独立部分故障的可能，故锁定预警原因为氧化风机出口支路有个别管路堵塞。

4.5 本章小结

本章以某电厂氧化风机为例，介绍了氧化风机基于大数据分析技术进行建模的详细过程，包括时序数据的采集、时序数据的清洗、参数关联度的计算和设备模型展示等；然后采用数据回放的模式验证模型的可用性，进行对模型进行初步的优化调整；最后以氧化风机支路堵塞为例，介绍了预警后的分析过程。氧化风机的案例可以得出结论：在电厂中，基于大数据分析技术进行设备状态在线监测及故障预警是切实可行的。

第 5 章 电厂关键设备凝结水泵模型的构建

5.1 模型创建

(1) 初选测点

凝结水泵为立式筒袋型双层壳体结构，由四部分组成，分别为泵筒体、工作部、出水部分以及推力装置部分。基于凝结水泵组成部分，将整个凝结水泵的测点进行初步统计，去除开关量测点，初步选择的测点包括：凝结水泵电流、凝结水泵上轴承温度、凝结水泵下轴承温度、凝结水泵推力轴承温度 1、2、3 和 4，凝结水泵线圈温度 1、凝结水泵线圈温度 2、凝结水泵线圈温度 3、凝结水泵线圈温度 4、凝结水泵线圈温度 5、凝结水泵线圈温度 6、凝结水母管压力等。

(2) 数据抽取

按照第四章氧化风机模型数据抽取的方式抽取凝结水泵历史数据，所得结果如下图所示。

Row	TimeStamp	1号凝结泵电 流	1号凝结水泵 上轴承温度	1号凝结水泵 推力轴承温 度1	1号凝结水泵 推力轴承温 度2	1号凝结水泵 推力轴承温 度3	1号凝结水泵 推力轴承温 度4	1号凝结水泵 下轴承温度	1号凝结水泵 线圈温度1
1	2015/1/1 0:00:00	-0.056900288...	-0.066846762...	-0.064166850...	-0.063079610...	-0.048127550...	-0.035636406...	-0.035575125...	-0.058175005...
2	2015/1/1 1:00:00	-0.056736417...	-0.067106582...	-0.066256165...	-0.065479710...	-0.055864062...	-0.033620808...	-0.034379236...	-0.057641409...
3	2015/1/1 2:00:00	-0.056572545...	-0.067566409...	-0.068345680...	-0.067879810...	-0.063600577...	-0.031605206...	-0.033183343...	-0.057107813...
4	2015/1/1 3:00:00	-0.056408677...	-0.068026237...	-0.070435188...	-0.070279911...	-0.071337081...	-0.029589608...	-0.031987454...	-0.056574217...
5	2015/1/1 4:00:00	-0.056244805...	-0.068486064...	-0.072524704...	-0.072680011...	-0.079073593...	-0.027574008...	-0.030791563...	-0.056040622...
6	2015/1/1 5:00:00	-0.056080933...	-0.068966545...	-0.074405118...	-0.075026683...	-0.084831893...	-0.026289047...	-0.030507726...	-0.055463004...
7	2015/1/1 6:00:00	-0.055775906...	-0.069671772...	-0.074508473...	-0.077114038...	-0.077945053...	-0.031440474...	-0.032638475...	-0.054497573...
8	2015/1/1 7:00:00	-0.055256921...	-0.070377007...	-0.074611827...	-0.079201392...	-0.071058206...	-0.036591898...	-0.034769222...	-0.053532145...
9	2015/1/1 8:00:00	-0.054737940...	-0.071082234...	-0.074715182...	-0.081288747...	-0.064171368...	-0.041743326...	-0.036899968...	-0.052566718...
10	2015/1/1 9:00:00	-0.054218955...	-0.071787469...	-0.074818536...	-0.083376102...	-0.057284526...	-0.046894751...	-0.039030712...	-0.051601290...
11	2015/1/1 10:00:00	-0.053699970...	-0.072492696...	-0.074921891...	-0.085463456...	-0.050397886...	-0.052046176...	-0.041161458...	-0.050635859...
12	2015/1/1 11:00:00	-0.053180988...	-0.073197923...	-0.075025245...	-0.087550811...	-0.043510843...	-0.057197604...	-0.043292205...	-0.049670431...
13	2015/1/1 12:00:00	-0.052662003...	-0.073903158...	-0.075128600...	-0.089638166...	-0.036624003...	-0.062349028...	-0.045422952...	-0.048705004...
14	2015/1/1 13:00:00	-0.052143022...	-0.074318744...	-0.076396204...	-0.089624322...	-0.032454203...	-0.066410630...	-0.046983428...	-0.048153162...
15	2015/1/1 14:00:00	-0.052447233...	-0.072443015...	-0.080776385...	-0.085449622...	-0.034590747...	-0.060900740...	-0.047036260...	-0.051233686...
16	2015/1/1 15:00:00	-0.053389549...	-0.070567287...	-0.085158567...	-0.081275321...	-0.036727290...	-0.055390849...	-0.047089096...	-0.054314214...
17	2015/1/1 16:00:00	-0.054331861...	-0.068691559...	-0.089536748...	-0.077100820...	-0.038863833...	-0.049880959...	-0.047141928...	-0.057394739...
18	2015/1/1 17:00:00	-0.055274177...	-0.066815830...	-0.093916930...	-0.072926327...	-0.041000377...	-0.044371068...	-0.047194760...	-0.060475263...
19	2015/1/1 18:00:00	-0.056216469...	-0.064940102...	-0.098297111...	-0.068751826...	-0.043136920...	-0.038861177...	-0.047247592...	-0.063555791...
20	2015/1/1 19:00:00	-0.057158801...	-0.063064374...	-0.102677293...	-0.064577326...	-0.045273467...	-0.033351287...	-0.047300424...	-0.066636316...
21	2015/1/1 20:00:00	-0.058101117...	-0.061188649...	-0.107057474...	-0.060402825...	-0.047410011...	-0.027841394...	-0.047353256...	-0.069716840...
22	2015/1/1 21:00:00	-0.059043429...	-0.060378670...	-0.109018318...	-0.060121621...	-0.047605458...	-0.023268079...	-0.047370549...	-0.071738570...
23	2015/1/1 22:00:00	-0.059550080...	-0.062413945...	-0.107775755...	-0.064723491...	-0.045504320...	-0.022333046...	-0.047341413...	-0.070785567...

图 5-1 历史数据展示

(3) 数据关联度分析

按照第四章的方法完成数据的清洗和无量纲处理后，就得到了设备正常运行状态的数据，以便析测点关联度。

(4) 创建模型

数据清洗任务是获取正常运行状态的历史数据；关联度分析的目的则是清除设备建模时的冗余测点；这两个过程是建模前最为关键和重要的部分，完成后即可进行设备建模。设备的动态模型最终被划分成一个个的运行剖面，如下所示：



图 5-2 模型剖面图

5.2 模型优化及调整

(1) 数据过滤

凝结水泵所用测点为电机电流，过滤值条件为电机电流大于 20。也就是说，只有电机电流大于 13 的时候，设备模型才会对设备运行状态就行评估，当电机电流小于 20 的时候，模型认为设备在启停阶段，不需要进行状态评估。

(2) 历史数据回放

如图 5-3 所示，拿建模时间的原始数据与模型对比，由于原始数据并未出现大量的异常数据，设备运行正常，只是有少量的噪声数据，所以模型不应该出现报警，而回放的最终结果也是模型并未发出预警，证明模型可用。

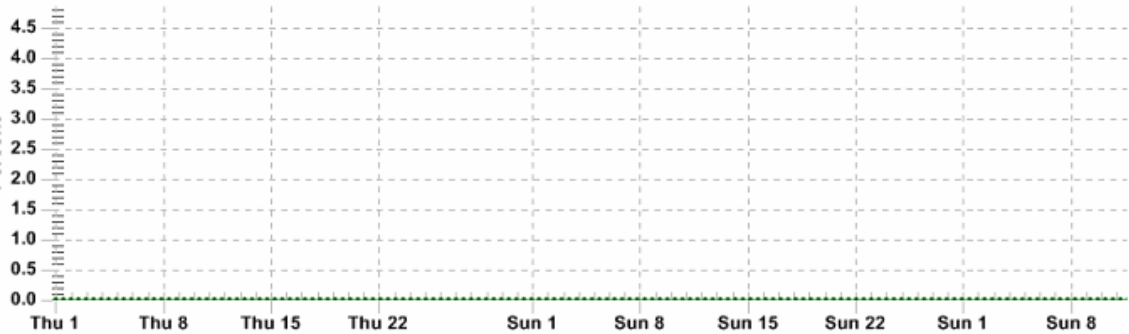


图 5-3 历史回放数据

5.3 模型运行情况分析

(1) 单点评估

系统针对每个测点会产生动态预测带，随生产实时运行状态不断变化。系统大约每 5 分钟采集一次实时数据，计算出单个测点的理想的动态预测值上下限。如果采集到的实际值在预测值上下限内，则代表测点运行正常；反之，则测点可能发生了异常，这可能是设备异动引起的，也可能是测点本身故障引起的。将测点每个时间点的实际值连接，就形成了测点实际值曲线；将测点预测值上限连接即形成预测值上限曲线；将测点预测值下限连接即形成预测值下限曲线。图 5-4 是凝结水泵下轴承的单点评估曲线，其中红色曲线代表实际测量值，蓝色曲线代表预测值上限，绿色曲线代表预测值的下限，蓝线和绿线之间的动态带即为设备理想运行状态带。从图中可以看出，三条线基本处于重合状态，证明凝结水泵下轴承温度的预测情况与实际运行情况高度吻合，设备运行状态良好。

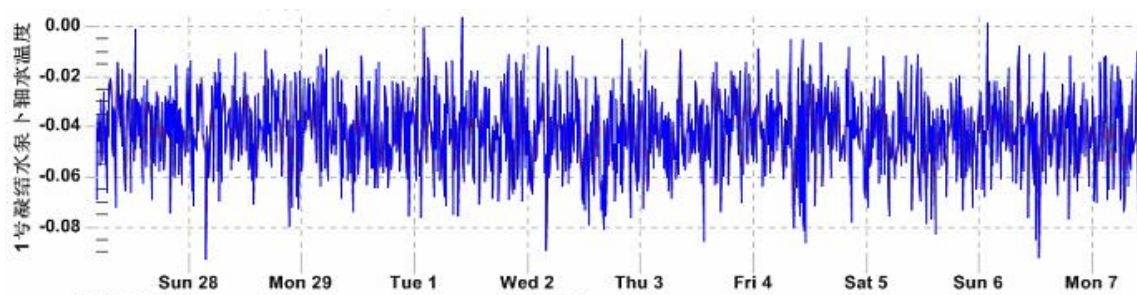


图 5-4 单点运行曲线

(2) 健康度评估

采用健康偏离度的概念评估设备整体运行状态，健康偏离度采用欧式距离的方式计算，即将各个测点的偏离情况进行欧式计算，得出设备健康状态的偏离，如图 5-5 所示。

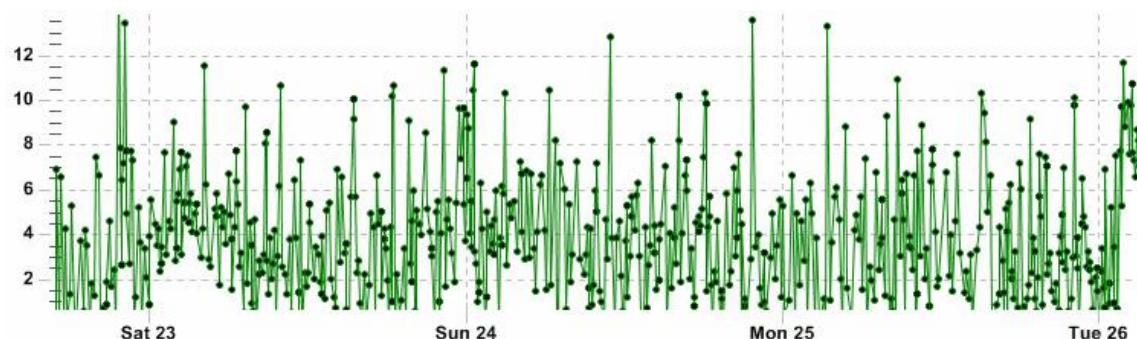


图 5-5 健康偏离度曲线

5.4 预警分析

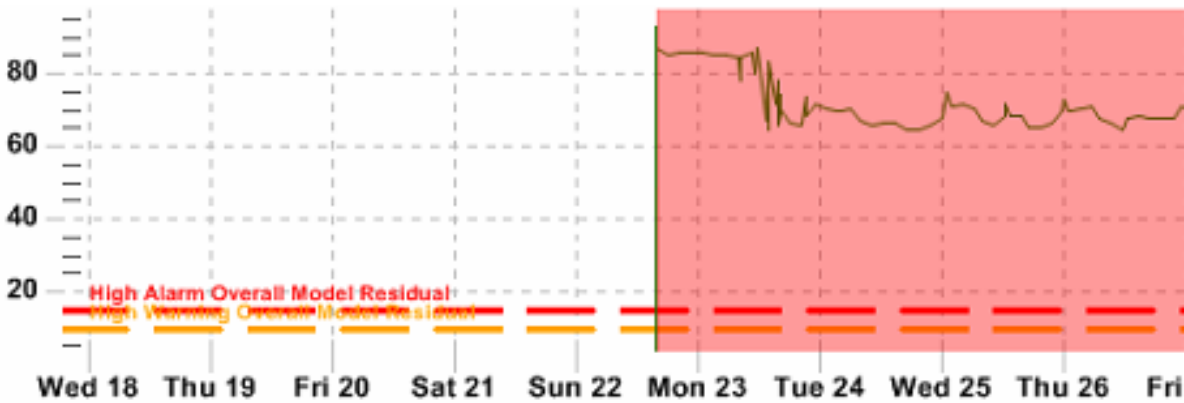
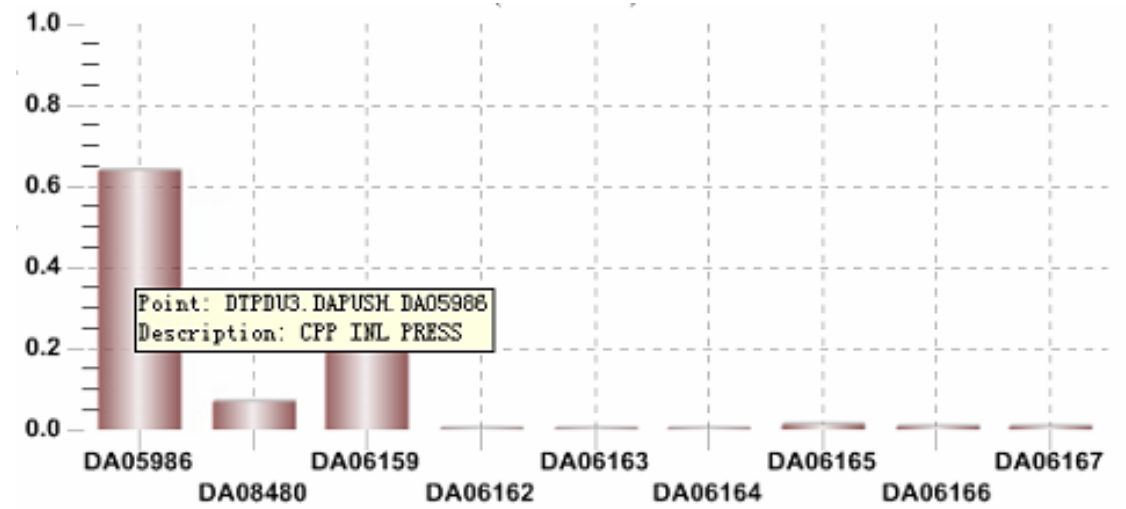


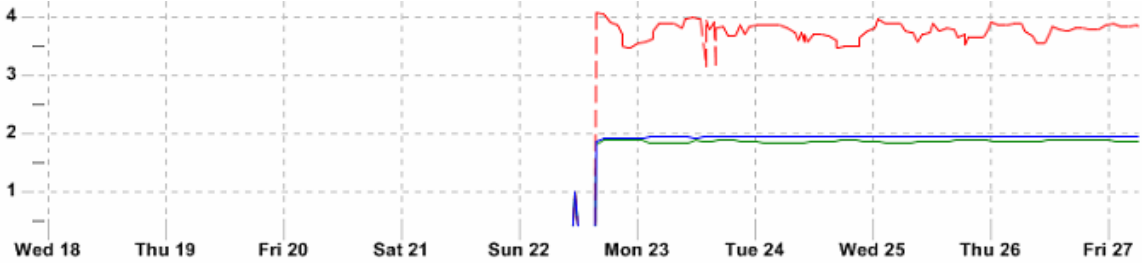
图 5-6 凝结水泵设备健康偏离度

凝结水泵自启机后即出现持续报警情况，具体时间为 2015 年 03 月 22 日。利用分析工具，查看模型报警测点贡献度，如图 5-7 所示。

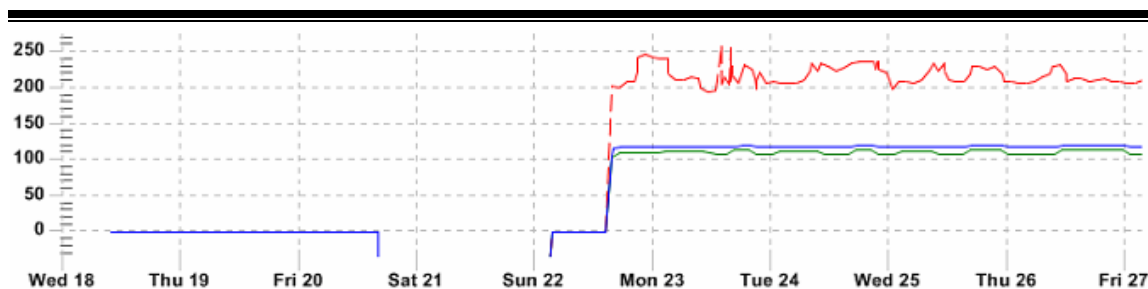


5-7 凝结水泵模型报警测点贡献度

从图中可以看出，引起凝结水泵模型报警的测点为凝结水泵电流和凝结水母管压力两个测点。查看以上两个测点的预测曲线图，如图 5-8、5-9 所示。



5-8 凝结水母管压力预测曲线



5-9 凝结水泵电机电流预测曲线

从图中可以看出，凝结水泵电机电流及凝结水母管压力的实际测量值与预测值有明显的偏差，且在偏离理想状态较高的水平运行。根据两测点的变化情况，判断报警可能的原因为凝结水泵为工频启动工况。

通常电厂凝结水泵目前大都采取变频运行模式。工频启动后，流量调节变为节流调节，节流调节会导致较大的能量损失，而工频启动电机电流较大，提升了厂用电率，导致电机发热，绝缘老化加快，厂用电耗提高，这对于节能降耗是非常不利的。凝泵变频运行可降低厂用电率至少 0.1%，提高经济效益 15 万元/月左右。

5.5 本章小结

本章以某电厂凝结水泵为例，简单介绍了凝结水泵基于大数据分析技术进行建模的过程；然后采用数据回放的模式验证模型的可用性，进而对模型进行初步的优化调整；最后以凝结水泵监测的预警为例，介绍了预警后的分析过程，进而证明，此系统可以为电厂的节能降耗做出一定的贡献。

第 6 章 电厂关键设备磨煤机模型的构建

6.1 模型创建

(1) 初选测点

磨煤机初步选择的测点包括：磨电机定子线圈温度 1、磨电机定子线圈温度 2、磨电机定子线圈温度 3、磨电机定子线圈温度 4、磨电机定子线圈温度 5、磨电机定子线圈温度 6、磨电机后轴承温度、磨电机前轴承温度、磨电流、磨轴承温度 1、磨轴承温度 2、磨轴承温度 3、磨轴承温度 4 等。

(2) 数据抽取

按照第四章氧化风机模型数据抽取的方式抽取磨煤机历史数据，所得结果如下图所示。

Row	TimeStamp	2号给煤机给煤率	2号磨电流	2号磨电机定子线圈温度1	2号磨电机定子线圈温度2	2号磨电机定子线圈温度3	2号磨电机定子线圈温度4	2号磨电机定子线圈温度5
1	2013/10/1 0:00:00	30.769096374...	34.497024536...	3.8085851669...	46.509960174...	68.887886047...	64.112236022...	64.727684020...
2	2013/10/1 1:00:00	29.442424774...	34.517372131...	16.753372192...	46.030364990...	68.371421813...	63.690505981...	64.231338500...
3	2013/10/1 2:00:00	25.968324661...	33.449775695...	-24.05347442...	45.550769805...	67.854957580...	63.268775939...	63.734989166...
4	2013/10/1 3:00:00	26.233856201...	33.612293243...	-16.08392524...	44.800319671...	67.206405639...	62.414234161...	62.951206207...
5	2013/10/1 4:00:00	25.879121780...	33.650310516...	0.6259646415...	43.787307739...	66.268379211...	61.404972076...	61.979320526...
6	2013/10/1 5:00:00	25.400516510...	33.508850097...	5.7255053520...	43.190753936...	65.330352783...	60.395709991...	61.007434844...
7	2013/10/1 6:00:00	26.300329208...	33.568817138...	37.247871398...	43.280132293...	65.297477722...	60.490108489...	61.112258911...
8	2013/10/1 7:00:00	26.323612213...	33.628784179...	31.975385665...	43.369510650...	65.472343444...	60.702606201...	61.288520812...
9	2013/10/1 8:00:00	29.387344360...	33.668704986...	35.275634765625	43.458889007...	65.647209167...	60.915103912...	61.464786529...
10	2013/10/1 9:00:00	30.687870025...	33.948722839...	42.967681884...	43.700969696...	65.822067260...	61.127605438...	61.641052246...
11	2013/10/1 10:00:00	31.128427505...	34.112491607...	39.161003112...	44.202331542...	65.996932983...	61.340103149...	61.817317962...
12	2013/10/1 11:00:00	28.650875091...	34.473983764...	35.547000885...	45.387767791...	66.869911193...	62.147506713...	62.832092285...
13	2013/10/1 12:00:00	29.077646255...	35.305769947...	32.969997406...	46.573204040...	68.123870849...	63.403186798...	64.095375061...
14	2013/10/1 13:00:00	28.024856567...	34.784229278...	39.314769744...	47.758640289...	69.377838134...	64.658866882...	65.358657836...
15	2013/10/1 14:00:00	29.303524017...	35.916088104...	35.185222625...	48.944076538...	70.631797790...	65.91455078125	66.621940612...
16	2013/10/1 15:00:00	29.980678558...	36.598091125...	25.000663757...	49.664737701...	71.885757446...	67.170227050...	67.885223388...
17	2013/10/1 16:00:00	30.039077758...	35.945224761...	30.311668395...	50.289001464...	72.054061889...	67.353363037...	68.075263977...
18	2013/10/1 17:00:00	25.219352722...	33.615619659...	10.893070220...	48.879360198...	70.871498107...	66.422782897...	67.023056030...
19	2013/10/1 18:00:00	32.404113769...	36.420269012...	25.044586181...	48.397205352...	70.412147521...	65.933197021...	66.475349426...
20	2013/10/1 19:00:00	31.699762344...	35.772693634...	2.3158516883...	48.669742584...	70.762451171875	66.137397766...	66.705429077...
21	2013/10/1 20:00:00	32.044776916...	35.932392120...	-0.570470511...	48.201953887...	70.305801391...	65.706748962...	66.337615966...
22	2013/10/1 21:00:00	32.654872894...	36.306648254...	17.657581329...	47.734165191...	69.849151611...	65.276107788...	65.858024597...
23	2013/10/1 22:00:00	31.205440521...	34.922973632...	15.254354476...	47.266376495...	69.351211547...	64.781784057...	65.372825622...

图 6-1 历史数据展示

(3) 数据关联度分析

按照第四章的方法完成数据的清洗和无量纲处理后，就得到了设备正常运行状态的数据，以便分析测点关联度。

(4) 创建模型

数据清洗的主要任务是获取正常运行时的历史数据；关联度分析的目的则是清除设备建模时的冗余测点；这两个过程是建模前最为关键和重要的部分，完成后即可进行设备建模。设备的动态模型最终被划分成一个个的运行剖面，如下所示：



图 6-2 模型剖面图

6.2 模型优化及调整

(1) 数据过滤

凝结水泵所用测点为电机电流，过滤值条件为电机电流大于 30。也就是说，只有电机电流大于 30 的时候，设备模型才会对设备运行状态就行评估，当电机电流小于 30 的时候，模型认为设备在启停阶段，不需要进行状态评估。

(2) 历史数据回放

如图 6-3 所示，拿建模时间的原始数据与模型对比，由于原始数据并未出现大量的异常数据，设备运行正常，只是有少量的噪声数据，所以模型不应该出现报警，而回放的最终结果也是模型并未发出预警，证明模型可用。

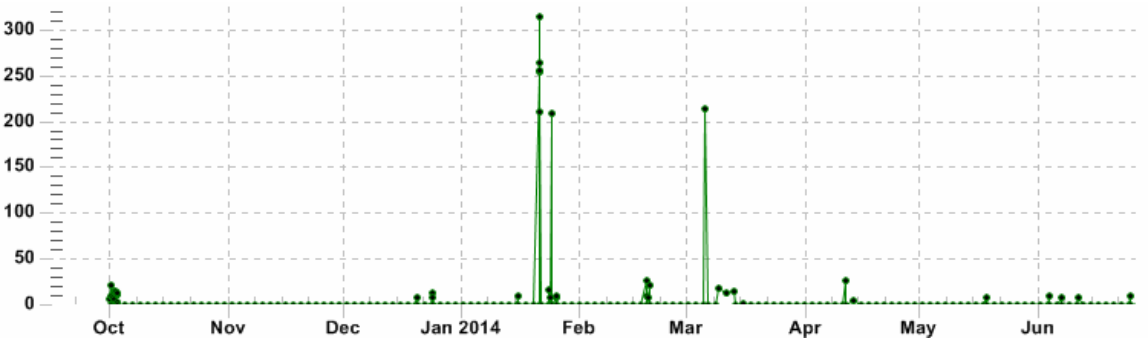


图 6-3 历史回放数据

6.3 模型运行情况分析

(1) 单点评估

系统针对每个测点会产生动态预测带，随生产实时运行状态不断变化。图 6-4 是磨煤机后轴承温度单点评估曲线，其中红色曲线代表实际测量值，蓝色曲线代表预测值上限，绿色曲线代表预测值的下限，蓝线和绿线之间的动态带即为设备理想运行状态带。从图中可以看出，实际运行曲线在预测值的上下限之间，证明磨煤机后轴承温度的预测情况与实际运行情况高度吻合，设备运行状态良好。

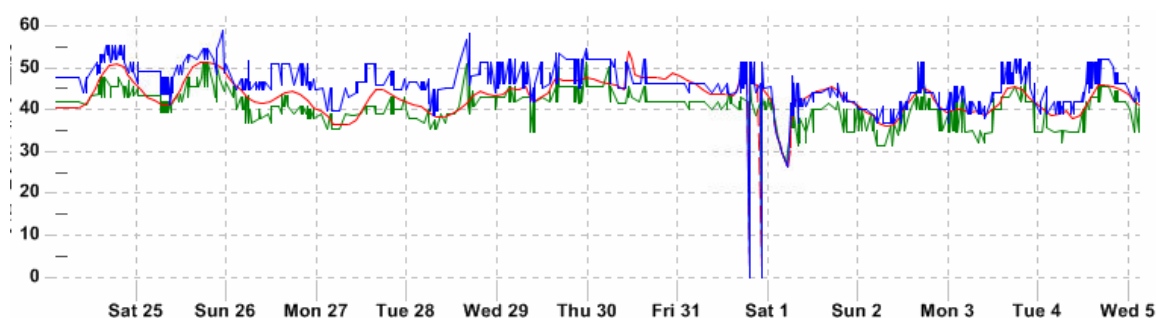


图 6-4 单点运行曲线

(2) 健康度评估

采用健康偏离度的概念评估设备整体运行状态，健康偏离度采用欧式距离的方式计算，即将各个测点的偏离情况进行欧式计算，得出设备健康状态的偏离，如图 6-5 所示。

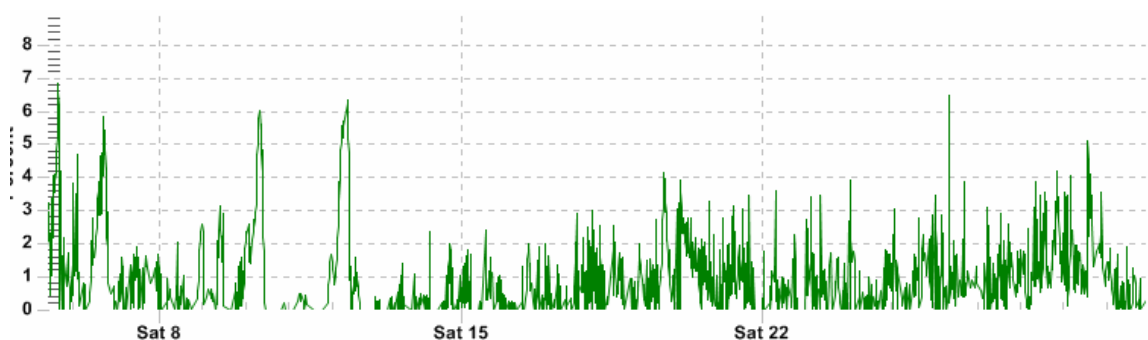


图 6-5 健康偏离度曲线

6.4 预警分析

2015 年 1 月 11 日，2 号磨煤机模型出现报警，报警曲线如图 6-6 所示：

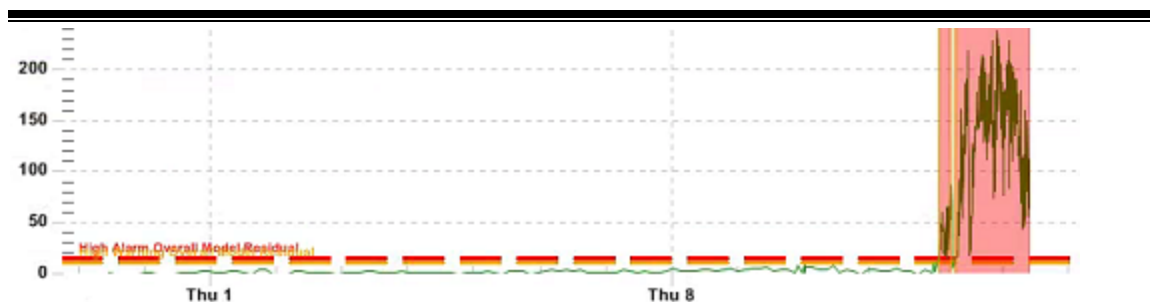


图 6-6 磨煤机电机模型报警曲线

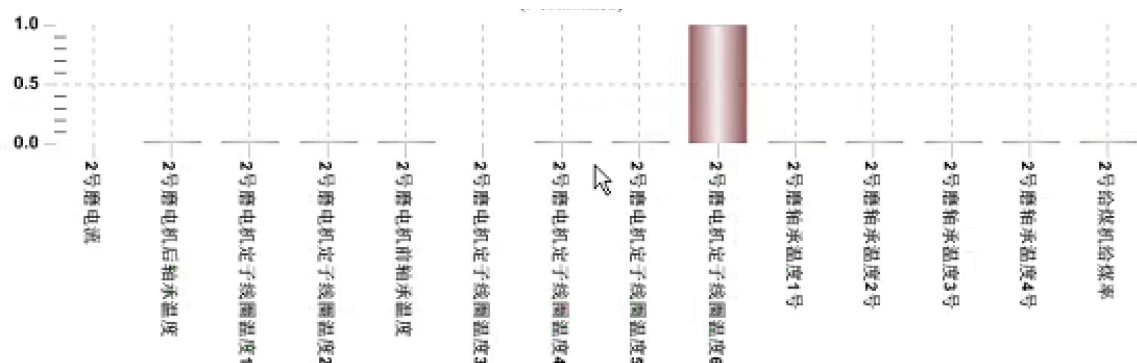


图 6-7 磨煤故障测点贡献度

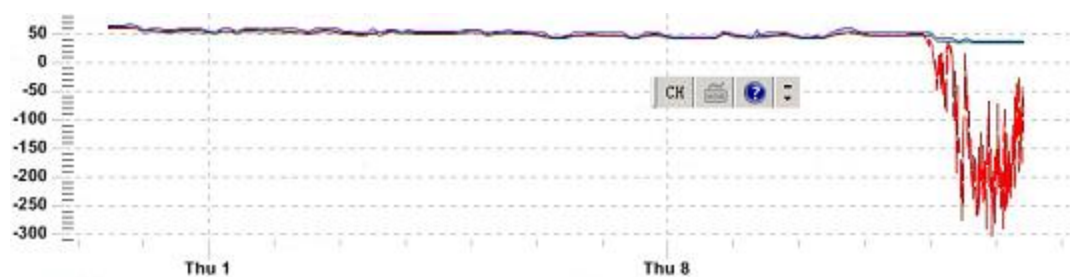


图 6-8 磨煤机定子线圈温度 6 测点预测值曲线

利用分析工具，查看报警具体测点贡献度，如图 6-7 所示。从图中得知，导致该报警的测点为电机定子线圈温度 6，进而查看该测点预测值与实际值，如图 6-8 所示。图中红线为该测点实际值，蓝线为该测点预测值。从图中可以看出，该测点实际值为负值，故判断为测点本身问题。

6.5 本章小结

本章以某电厂磨煤机为例，简单介绍了磨煤机基于大数据分析技术进行建模的过程；然后采用数据回放的模式验证模型的可用性，进而对模型进行初步的优化调整；最后以磨煤机测点故障预警为例，介绍了如何判断设备测点故障引起的报警。通过磨煤机的案例，得知设备状态在线监测及预警诊断系统还可以为电厂热工人员提供及时准确的测点故障信息。

第 7 章 结论与展望

7.1 结论

大数据分析技术是未来的发展方向，电力企业早就积累了海量的时序型历史数据，如何运行这些数据一直是电力人关注的热点。本文就是通过将大数据分析技术应用在设备状态在线监测与预警诊断领域，探索出设备海量历史数据应用的新模式，结论如下：

(1) 介绍了大数据分析技术的一些基本概念，常用的数据挖掘算法等。

(2) 介绍了大数据分析技术在在线监测与预警诊断领域的应用。通过采用灰色关联度算法计算出每两个参数间的关联关系，确定设备模型所需的测点；通过采用相似性聚类算法，发现设备运行规律。

(3) 以某电厂氧化风机、磨煤机、凝结泵为例，清晰地描述了设备从建模到在线监测再到预警诊断的过程。在建模时，采用了多种大数据分析的思路，比如时序数据的采集、时序数据的清洗工作，其中清洗工作还需要专业领域人员的判断；在在线监测的过程中，系统会根据设备的运行情况，与历史运行规律做对比，计算出每个测点的预测带及设备整体的健康偏离度；在预警诊断时，找出影响设备预警的测点，并对测点运行变化情况进行详细分析，推理出设备故障的可能。氧化风机的案例表明，基于大数据分析技术实现在线监测及预警诊断功能是可行的，必将推动电厂向数字化、智能化方向发展，能够有效提升设备的管理水平。

7.2 展望

研究发现，64%的大企业已经计划实施大数据项目，但是 85%的项目将面临失败的尴尬。在国外，采用此模式实现设备在线监测及预警诊断已经证明是可行的，包括美国的 Exlon、法国的 EDF、意大利的 Edipower 都已经通过此模式实现设备的管理。而在国内，包括大唐集团在内的诸多大小型发电集团已经计划或者已经开始实施此类项目。该项目在电厂的试点实施工作已经初见成效，相信能逐步推广到其他电厂，并在发电集团逐步推广应用。

参考文献

- [1]袁静. 面向设备故障诊断的数据挖掘关键技术研究[与实现[D]. 西安电子科技大学硕士学位论文. 西安: 2012
- [2]侯捷. 数据挖掘在旋转机械故障诊断中的应用研究[D]. 大连理工大学硕士学位论文. 大连: 2012
- [3]王昌长,高胜友,李福棋. 电力设备的在线监测与故障诊断[M]. 北京:清华大学出版社, 2008
- [4]吴明强. 故障诊断专家系统研究的现状与发展[J]. 计算机测量与控制. 2005, 13(12): 1301-1304
- [5]刘小明.基于故障树和神经网络的运载火箭故障诊断关键技术研究[D]. 电子科技大学硕士学位论文. 西安: 2012
- [6]陈鸣山,程芳真,赵建. 火电厂单元机组经济运行专家系统的研制[M]. 中国电力2006, 30(9): 63-66
- [7]刘有才,刘增良. 模糊专家系统原理与设计[M]. 北京:北京航空航天大学出版社2006, 1-69: 121-158
- [8]Odgaard P.F, Bao Lin, Jorgensen. Observer and Data-Driven-Model-Based Fault Detection in Power Plant Coal Mills[J].IEEE Transactions on Energy Conversion.2008, 23(2): 659-668
- [9]Christina A,Vasilis C. Intelligent system for identification and replacement of faulty sensor measurements in thermal power plants[J]. Expert Systems with Applications.2009, 36(5) : 8750-8757
- [10]Yinsong Wang,Xinghuo Yu. New Coordinated Control Design for Thermal-Power-Generation Units[J]. IEEE Transactions on Industrial Electronics. 2010, 57(11) : 3848-3856
- [11]杨叔子,郑晓军. 人工智能与诊断专家系统[M] . 西安:西安交通大学出版社. 2003
- [12]许婧,王晶,高峰,束洪春. 电力设备状态检修技术研究综述[J]. 电网技术. 2005
- [13]白建青. 供电设备从定期维修制向状态检修制过渡 [J]. 青海电力. 2002, 24(4): 66-69
- [14]邵胜利. 开展状态检修实现可靠性和经济性相统一[J]. 电力建设. 2001, 14(3): 25-28
- [15]杨以涵,唐国庆,高曙. 专家系统及其在电力系统中的应用[M]水利电力出版社. 2005
- [16]K F Thang,R K Aggarwal,D G Esp. Statistical and Neural Network Analysis of Dissolved Gases in Power[J]. 2004
- [17]Blair J. Shirkhodaie. A Diagnosis and Prognosis of Bearing Using Data Mining and Numerical Visualization Techniques[J]. 2004
- [18]Blair J. Shirkhodaie. A Bearing Fault Identifications Using Agent-based Data Mining and Numerical Visualization Techniques[J]. 2007

-
- [19] Z. Huaichun, L. Chun, C. Qiang, et al. Experimental investigations on visualization of three-dimensional temperature distributions in a large-scale pulverized-coal-fired boiler furnace[J]. Proceedings of the Combustion Institute, 2005, 30(1): 1699-1706
 - [20] Kwon W. H, Kim S. W., Park P. G. On the multivariable robust controller of a boiler-turbine system[C]. IFAC Symposium. Power Systems and Power Plant Control 1989. Oxford, UK: Pergamon, 1989: 219-223.
 - [21] 杨璋. 大型汽轮发电机组振动在线状态监测与故障诊断系统[D]. 浙江大学硕士学位论文. 杭州: 2010
 - [22] 王明锐. 汽轮机及主要辅机在线监测诊断系统研究与应用[D]. 天津大学硕士学位论文. 天津: 2012
 - [23] 李玉荣. 模糊专家系统在火电厂故障诊断中的应用与研究[D]. 华北电力大学硕士学位论文. 北京: 2006
 - [24] 黄文虎, 夏松波. 设备故障诊断原理技术与应用[M]. 北京: 北京科学出版社. 2004
 - [25] 冯西桥, 陈进. 现代信号处理在机械故障诊断中的应用与展望[J]. 振动与冲击 2014, 33(3): 45-48
 - [26] Lu H. Effective data mining using neural net works[J]. IEEE Trans. on Knowledge and Data Eng. 2006, 18(6): 125-127
 - [27] Daniel E. Miller. A New Approach to Model Reference Adaptive Control[J]. IEEE Transactions on automatic control. 2003, 48(5): 743-757
 - [28] Park J. Using a hash-Based method with transaction trimming for mining association rules[J]. IEEE Trans. on Knowledge and Data Eng. 1997, 9(5): 224-228
 - [29] Fu Y, J Han. Meta-rule-guided mining of association rules in relational databases, KDOOD'95. Singapore. 1995: 39-46
 - [30] Chen M et al. Data mining for path traversal patterns in a Web environment In Proc. 16th Int'l Conf[C]. Distributed Computing Systems. 1996: 382-392
 - [31] Venter F. Knowledge discovery in databases using lattices[J]. Expert systems with Applications. 1999, 13(4): 259-264
 - [32] Dhar V, Tuzhilin A. Abstract-driven pattern discovery in databases[J]. IEEE Trans Knowledge and Data Eng. 1997: 926-938
 - [33] Han J. Mining knowledge at multiple concept levels[J]. Information and Knowledge Management. 2000: 19-24
 - [34] Elder J. A statistical perspective on knowledge discovery in databases[J]. Advances in Knowledge Discovery and Data Mining. AAAI/MIT Press. 2000: 83-115
 - [35] Cheeseman P. Bayesian classification (AutoClass): Theory and results. advances in knowledge discovery and data mining. AAAI/MIT Press. 2001: 153-180
 - [36] Lu H. NeuroRule: A connectionist approach to data mining In Proc. 21st Int. Conf. very Large Databases. 2001: 478-489
 - [37] 苏剑飞, 赵捍东. 神经网络在旋转机械故障诊断中的应用[J]. 微计算机

- 息. 2012: 221-225
- [38] 石荣德,赵廷弟.故障诊断专家系统[J].北京航空航天大学学报.1999, 12(3): 224-226
- [39] 张彦铎,姜兴渭.多传感器信息融合及在智能故障诊断中的应用[M].传感器技术.2004
- [40] 苏剑吃,赵捍东.神经网络在旋转机械故障诊断中的应用[J].微计算机信息. 2011,35(6): 37-41
- [41] 卢学军,岳忠宇.基于知识与模糊神经网络的故障诊断专家系统[J].机械强度2008, 22(1): 4-6
- [42] 付阶辉.基于 Petri 网的故障诊断方法研究[D].东南大学硕士学位论文. 南京: 2012
- [43] 吴烨.液压挖掘机状态监测与故障诊断系统的研究[D].东南大学硕士学位论文. 南京: 2008
- [44] A Moradian, E D Thompson, M A Jenkin, G H Tomlinson. New idea in on-line diagnostics improves plant performance[J]. Power, May 1998
- [45] 丁艳军,王培红,吕震中.生产过程早期故障诊断检测与诊断的一种新方法[J].中国电机工程学报. 2011, 11(5): 36-40
- [46] 常澍平,郭江龙,吕玉坤.非线性状态估计(NSET)建模方法在故障预警系统中的应用[J]. 软件. 2012, 7(2): 21-26
- [47] 马少平,朱小燕.人工智能[M].北京:清华大学出版社.2012
- [48] 陆汝铃.人工智能[M].北京:科学出版社.1999

攻读硕士学位期间发表的论文及其它成果

（一）发表的学术论文

攻读硕士学位期间参加的科研工作

- [1] 脱硫系统浆液 PH 值在线监测缓冲测量筒装置研制，并获得国家专利一项。
- [2] 负责张家口电厂“基于大数据的事故预警与故障诊断系统应用”科技项目。

致 谢

一转眼硕士研究生的三年学习时间就要过去了，在这三年的学习过程中，经历了很多，也学习到了很多。忘不了老师的严谨和博学，也忘不了同学们之间的欢声笑语。

首先，衷心感谢我的导师田亮副教授对我本人的关怀和指导。从论文的开题到最终完成，田老师总是热情、真诚、细致、不厌其烦的对我进行点拨和鼓励，并及时提醒我的写作进度。在此，我谨向田老师表示衷心的感谢，并祝福田老师身体健康，工作顺利！

其次，要感谢刘鑫屏老师在我的科研工作方面提出了许多宝贵的意见。同时，还要感谢专业各位同学及同事的帮助和支持！

最后，感谢家人对我的理解和支持，让我顺利的完成了学业，谢谢！

作者简介

1986 年 09 月 02 日出生于辽宁省喀左县。

2004 年 07 月考入东北电力大学自动化学院（系）自动化专业，2008 年 07 月本科毕业并获得工学学士学位。

获奖情况：

2012 年获中国大唐集团公司 2012-2013 年生产技能类 C 级“112 人才”称号

2013 年获得“脱硫系统浆液 PH 值在线监测缓冲测量筒装置”国家专利。

2014 年获大唐国际“火电厂自动化专业点检定修制创新与实践”管理创新成果一等奖

2014 年获大唐国际“运用动态控制思想和绩效评价手段实现机组检修全过程管理创新实践”管理创新成果二等奖

2014 年获大唐国际“机组检修工艺标准知识库体系创建与系统应用”大唐国际管理创新成果二等奖

工作经历：

2008 年 7 月到 2012 年 6 月工作于大唐国际张家口电厂检修公司热控车间。。

2012 年 7 月至今工作于大唐国际张家口电厂设备部。