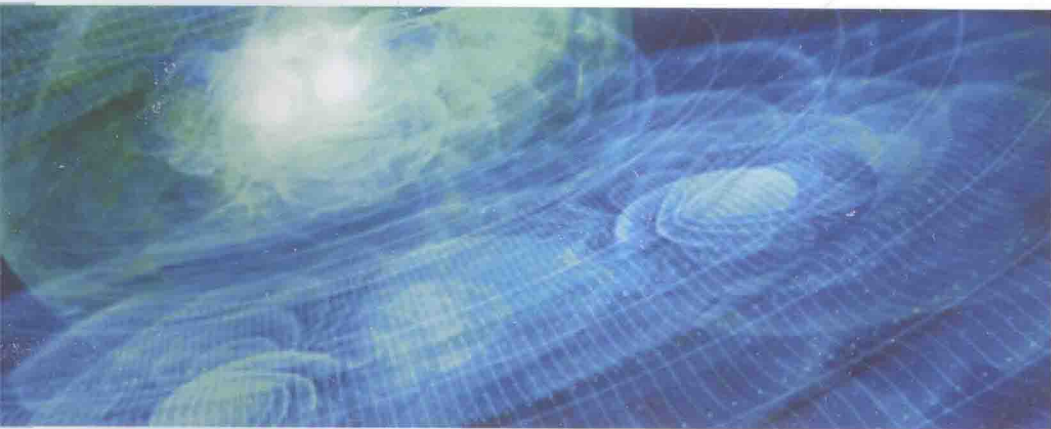


学术研究专著系列 · 机电工程



基于系统图谱的复杂机电系统

状态分析方法

孙 锴◎著

JIYU XITONG TUPU DE
FUZA JIDIAN XITONG
ZHUANGTAI FENXI
FANGFA

西北工业大学出版社

学术研究专著系列·机电工程

JIYU XITONG TUPU DE FUZA JIDIAN XITONG ZHUANGTAI FENXI FANGFA

基于系统图谱的复杂机电系统状态分析方法

孙 锴 著

西北工业大学出版社

图书在版编目(CIP)数据

基于系统图谱的复杂机电系统状态分析方法/孙锴著. —西安:西北工业大学出版社,
2016.8

ISBN 978-7-5612-5079-2

I. ①基… II. ①孙… III. ①机电系统—状态分析—研究 IV. ①TM7

中国版本图书馆 CIP 数据核字 (2016) 第 217131 号

策划编辑: 雷 军

责任编辑: 雷 军

出版发行: 西北工业大学出版社

通信地址: 西安市友谊西路 127 号 邮编: 710072

电 话: (029)88493844 88491757

网 址: www.nwpup.com

印 刷 者: 陕西天意印务有限责任公司

开 本: 787 mm×1 092 mm 1/16

印 张: 9 插页: 16

字 数: 209 千字

版 次: 2016 年 8 月第 1 版 2016 年 8 月第 1 次印刷

定 价: 58.00 元

前言

流程工业生产系统是通过流体、能量和信息,多介质传递耦合,由多种动力装置、反应装备和控制设备构成的分布式复杂机电系统。近年来,随着事故类型和数量的增加,我们发现以往被忽略的因素,如微小设备或部件异常引发的故障和事故,带来的损失甚至超过了关键设备和典型故障,系统状态分析问题实质上已经演变为一个安全系统工程问题。

现代企业通过以分布式控制系统(Distributed Control System,DCS)为核心的监控系统对整个生产过程进行监测和记录,保存了系统运行的所有信息。通过分析DCS监测数据集分析系统的运行状态是一种切实可行的系统安全分析方法。由于复杂机电系统自身的特性,其监测数据集是具有海量性、非线性、多源性、层次性的含有噪声的数据包,传统的基于数据降维的多元统计分析方法难以从全局角度分析系统运行状态,量化某个故障对整个系统造成的危害程度。

为了充分利用系统监控数据集,分析系统数据间普遍存在的复杂关联关系,对系统进行全面的状态分析,本书以数据可视化理论为依据,提出系统图谱的概念,建立了一整套基于系统图谱的系统运行状态分析及故障诊断的理论和算法。^①

书中首次提出系统图谱的概念。将数据可视化技术引入分布式复杂机电系统安全监控领域,利用DCS数据集自身所具有的数据间高度的关联性和耦合性,制定数据着色规则,将DCS数据投影到色彩相空间中,构造系统图谱作为之后面向人机交互和自动分析的基础。利用数字图像像素自身特有的非线性、关联性和耦合性展示DCS监测数据间复杂的关系,避免了数据降维所带来的复杂的数学运算,并且消除了数据海量性、非线性和耦合性对系统分析造成的影响。

为了有效地构造系统图谱,需要对DCS数据进行预处理,包括对DCS数据集的分割、归一化和降噪。本书提出时间域数据分割方法。利用单变量时间序列的自相关性分析DCS监测数据集中每个监测变量的周期,统计所有变量周期性的共同规律,从而得出系统的统计平均周期。以系统的统计平均周期为基本单位,对DCS监测数据集在时间域进行分割,消除时间海量性对系统分析的影响;提出通过数据归一化消除数据的多源性的方法。针对由不同类型的监测变量的量纲不同而导致的DCS监测数据集的各个变量的绝对值的差异问题,利用数据标准化算法,对DCS监测数据集先做归一化处理,再根据数据着色规则构造归一化系统彩色图谱,从而消除了数据多源性造成的系统彩色图谱中横向联系被隔断的问题,并且进一步分析了归一化系统彩色图谱中不同的颜色与系统稳定性之间的关系;提出二维

^①本研究受到国家863计划(项目编号:2006AA04Z441,2007AA04Z432)和国家自然科学基金(项目编号:51175402,61104025)的支持。

全阈值小波降噪消除归一化系统彩色图谱中的白噪声。针对小波变化可以区分噪声信息和系统高频信息的特点,利用全阈值小波降噪法消除归一化系统彩色图谱中的噪声后得到降噪后的系统彩色图谱,作为系统分析和故障诊断的基础。

本书以 DCS 数据集构成的系统图谱为研究对象,首先提出面向人机交互的系统状态分析和故障诊断方法。分析不同类型系统图谱与系统运行健康状态的关系,利用人眼对于色彩变化的敏感度远远高于数字的生理特点,建立一种依赖人眼挑出代表系统运行异常的系统图谱的规则。然后,利用数字图像处理技术分析系统图谱,提出系统危险度指标、正常程度函数、危险能量函数等定量描述系统运行健康状态的参数,提出面向计算机自动分析的企业级系统安全运行程度量化方法和故障模式识别技术,同时还给出一套基于监测数据集的企业健康状态评级方法。

为了说明基于系统图谱分析系统运行健康状态的有效性,本书以某煤化工企业的空气压缩机组的监测数据为实例,验证本研究所提出的系统分析和故障诊断方法。同时,依提出的系统分析和故障诊断方法为核心算法,开发了一套面向流程工业系统安全监控原型软件。

本书著者孙锴,西安交通大学控制科学与工程专业博士毕业,现任教于西安建筑科技大学机电工程学院。多年来一直从事质量与可靠性、制造系统状态监测与分析以及化工系统安全分析等方面的研究,参与多项国家自然科学基金的研究工作,发表多篇学术论文,并获得多项国家发明专利。

本书在撰写过程中查阅了大量文献,在此向文献作者表示衷心感谢!

由于水平有限,书中难免有不足欠妥之处,恳切希望读者予以批评指正。

著者

2016年5月

主要符号表

DCS	分布式控制系统(Distributed Control System)
X	数据矩阵
$X(i, j)$	采样值
Π	色彩相空间
χ	着色规则
Γ	系统彩色图谱
X	数据矩阵向量
Γ	彩色图谱序列
Γ^N	正常运行状态的彩色图谱集合
Γ^A	异常运行状态的彩色图谱集合
Γ_{std}	基准彩色图谱
Λ_k	异常图谱
$W_k(\Lambda_k)$	危险度指标
W	危险度指标序列
memFun	正常程度函数
H	系统空间
X_{up}	系统上阈值面
X_{down}	系统下阈值面
X_{th}	系统阈值空间
B	系统边界
X_f	系统异常模式矩阵
P	系统故障图谱
B_f	故障相似度函数
PBW	系统故障图谱库
W	权重矩阵
X_{fw}	系统异常模式权重矩阵
F	危险能量函数
$D(X)$	系统健康状态分级函数

目 录

主要符号表

第 1 章 复杂机电系统运行健康状态分析的传统理论	(1)
1.1 复杂机电系统的安全问题	(1)
1.2 过程监控系统在生产企业中的现状	(3)
1.3 数据驱动系统状态分析方法及国内外研究现状	(6)
1.3.1 数据驱动系统状态分析的研究对象和内容	(6)
1.3.2 系统状态分析研究现状	(7)
1.3.3 数据驱动系统运行状态分析	(8)
1.3.4 目前研究中存在的问题	(11)
第 2 章 复杂机电系统监测数据图谱的构造方法研究	(13)
2.1 DCS 监测数据集的特点以及面临的问题	(13)
2.1.1 海量性	(13)
2.1.2 多源性	(14)
2.1.3 层次性	(14)
2.1.4 关联性	(16)
2.1.5 非线性	(16)
2.2 数据可视化技术	(20)
2.3 田纳西-伊斯曼化工仿真系统	(24)
2.4 系统图谱构造方法	(29)
2.4.1 数据矩阵	(29)
2.4.2 系统彩色图谱构造方法	(29)
2.4.3 系统故障图谱构造方法	(32)
2.5 本章小结	(36)

第 3 章 消除时间海量性对系统健康运行状态分析的影响	(37)
3.1 离散序列的相关性 ^[79]	(38)
3.2 单变量时间序列 $X[n]$ 的自相关性	(40)
3.3 计算单变量时间序列的周期	(41)
3.4 系统的运行周期	(43)
3.5 本章小结	(52)
第 4 章 分布式复杂机电系统监测数据集的数据预处理方法研究	(53)
4.1 数据归一化	(53)
4.1.1 DCS 监测数据集的多源性	(54)
4.1.2 数据归一化理论	(57)
4.1.3 监测集的数据归一化	(57)
4.2 数据降噪	(62)
4.2.1 小波的概念	(62)
4.2.2 小波用于信号降噪的原理	(66)
4.2.3 图像降噪的理论基础	(67)
4.2.4 归一化系统彩色图谱的二维小波降噪	(71)
4.3 本章小结	(73)
第 5 章 面向人机交互的系统故障识别方法研究	(75)
5.1 基于系统彩色图谱的故障识别方法	(75)
5.2 基于系统故障图谱的故障识别方法	(82)
5.3 各种故障辨识方法的对比	(85)
5.4 本章小结	(87)
第 6 章 面向计算机自动分析的系统健康状态分析方法研究	(88)
6.1 系统运行健康状态分析的基本内涵	(88)
6.2 数字图像的逻辑运算	(89)
6.3 基于数字图像处理技术的系统安全定量分析方法	(92)
6.3.1 基于系统彩色图谱的系统异常度定量分析	(92)
6.3.2 基于系统故障图谱的企业级故障模式识别技术	(98)
6.3.3 企业级系统健康状态评级	(100)
6.4 本章小结	(103)

第 7 章 企业实际案例	(104)
7.1 实例对象简介	(104)
7.1.1 空压机子系统	(105)
7.1.2 增压机子系统	(106)
7.1.3 汽轮机及蒸汽冷凝子系统	(107)
7.1.4 润滑油路子系统	(108)
7.2 系统的运行周期	(109)
7.3 面向人机交互的系统故障识别	(113)
7.3.1 基于彩色图谱的系统故障识别	(113)
7.3.2 基于故障图谱的在线监控方法	(116)
7.4 系统运行健康状态分析	(117)
7.4.1 彩色图谱分析系统动态特性	(117)
7.4.2 故障图谱分析系统动态特性	(119)
7.5 本章小结	(120)
第 8 章 分布式复杂机电系统健康运行状态分析原型软件	(121)
8.1 需求分析和系统架构设计	(121)
8.2 系统实现	(123)
8.2.1 系统的运行环境	(123)
8.2.2 生产状态监控模块	(124)
8.2.3 生产数据长效分析模块	(126)
8.3 本章小结	(127)
参考文献	(128)

第1章 复杂机电系统运行健康状态分析的传统理论

以能源重化工为代表的现代流程工业系统由一系列机电设备通过物质流、能量流和信息流紧密耦合而成,各单元要素之间呈现复杂非线性耦合关系,具有分布地域广泛、组成单元众多、耦合关系复杂等特点,称为分布式复杂机电系统^[1]。随着计算机技术与控制技术的飞速发展,一方面,分布式复杂机电系统不断向大型化、智能化、集约化方向发展,复杂程度不断提高;另一方面,对系统安全性、可靠性的要求也不断增加。因此,系统生产运行过程的长周期平稳性和可靠性成为现代流程工业生产的重要目标和任务^[2]。

1.1 复杂机电系统的安全问题

以能源重化工为代表的现代流程工业系统由一系列机电设备通过物质流、能量流和信息流紧密耦合而成,各单元要素之间呈现复杂非线性耦合关系,具有分布地域广泛、组成单元众多、耦合关系复杂等特点,称为分布式复杂机电系统^[1]。随着计算机技术与控制技术的飞速发展,一方面,分布式复杂机电系统不断向大型化、智能化、集约化方向发展,复杂程度不断提高;另一方面,对系统安全性、可靠性的要求也不断增加。因此,系统生产运行过程的长周期平稳性和可靠性成为现代流程工业生产的重要目标和任务^[2]。

在能源重化工为代表的现代流程工业生产系统中,由于生产过程和加工对象属于高温、高压和易燃、易爆等物质,因此发生火灾、爆炸、污染泄露、人员中毒等安全事故的概率远远高于其他产业,停产一天的经济损失可达数百万元,并造成重大的人员伤亡。例如,1984年的印度博帕尔灾难被称为人类历史上最惨的中毒事件,共导致了2.5万人丧生,55万人间接致死,20余万人永久残废,并造成孕妇流产和死产122例,77名新生儿出生后不久死去,9名婴儿畸形,对环境更造成难以补救的破坏。至今当地居民的患癌率及儿童夭折率仍因这场灾难而远远高于印度其他城市。1950年11月21日,墨西哥波查·里加镇,一所新建的从天然气中回收硫磺的工厂,由于硫化氢泄漏,造成22人死亡,320人被送进医院。1964年9月14日,日本富山市,由于氯气贮罐至三氯化磷生产工段约90米长的1英寸管道,在液氯气化时发生破裂,氯气猛烈喷出,造成9167人受害,其中533人患病,47人入医院治疗。

在国内,据国家安全生产监督管理总局的报告,2014年1至5月份,全国共发生9起危险化学品和化工重大事故,导致35人死亡、52人受伤。其中,山东滨洲滨阳燃化有限公司“1·1”中毒事件造成4人死亡、3人受伤;安徽亳州康达化工有限公司“1·9”中毒事件造成4人死亡、2人轻伤;吉林通化化工股份有限公司“1·18”爆炸事故造成3人死亡、5人轻伤;四川攀枝花天亿化工有限公司“3·1”泄露着火中毒事故造成3人死亡;内蒙古乌海泰和煤焦化公司“4·8”爆炸事故造成3人死亡、1人重伤、1人轻伤;江苏如皋双马化工公司“4·16”爆炸事故造成8人死亡、9人受伤(其中3人危重,Ⅲ度烧伤分别达到91%,96%,98%);山西临汾永鑫焦化有限公司“4·26”煤气爆炸事故造成4人死亡、31人受伤(其中8人重伤);辽宁灯塔北方有限

公司“4·24”中毒事故造成3人死亡；四川广元天森煤化集团有限公司“5·2”爆炸事故造成3人死亡。在上述9起事故中，5起爆炸事故，4起中毒窒息事故；7起涉及直接作业换接，其中动火3起，进入受限空间2起，检维修2起。

流程工业生产系统中各类设备的小型事故常常导致整个系统停车，给企业造成巨大的经济损失。以2014年1至5月发生的9起较大事故为例，事故往往起因于一些被人们忽略的细节。同时，据我们在某煤化工集团企业的实际调研，该集团每年由于各种小型故障导致的系统全线跳车事故达10余起，造成年直接经济损失2000万元以上。而这些跳车事故的起因绝大多数都是由于非核心设备的局部微小故障引起的。以企业双甲车间空气压缩机组的两次典型的跳车事故为例。第一次是因为转速突然下降而导致的紧急停车事故。事故原因调查结果为涡轮轴封泄露导致高温气体辐射电缆，使得电液转换器无法正常传输信号，以控制高位阀正常动作，最终因蒸汽流量下降而导致系统停车。另一次跳车事故是由于涡轮控制柜内的一个继电器的触电氧化使得接触电阻增加，无法正常送出闭合信号。这些事故表明，流程工业系统作为一个典型的分布式复杂机电系统，包含机、电、液、高温、动力传输和控制信号传递等多种因素，系统设备间紧密关联、高度耦合、结构复杂，任何微小的故障或扰动都有可能被系统放大和传递，形成级联式“雪崩”，造成系统崩溃。若不进行预防的话，任何局部性的故障都有可能造成全局性的影响。因此，流程工业系统的安全生产问题不仅仅要关注核心设备，而是一个系统性的、全局性的问题，必须对其中的所有设备都给予足够的重视，站在系统层面考虑系统运行的健康状态，才能更好地保障生产安全。

以印度博帕尔灾难为例，因为系统整体安全性的极度恶化，必然会出现重大安全事故。1984年12月3日凌晨，美国联合碳化物属下的联合碳化物(印度)有限公司设于贫民区附近的一所农药厂发生甲基异氰酸酯(MIC，一种易燃、易爆且具有挥发性的剧毒液体)泄露，引发著名的印度博帕尔灾难。MIC的毒性极强，对人体的毒性表现为：眼和上呼吸道的刺激表现；低浓度引起咳嗽、流泪，高浓度致角膜溃烂，失明；刺激还可致上呼吸道感染及嗅觉丧失；高浓度的MIC可通过无损的皮肤吸收，引起皮肤水肿，组织坏死；可引起严重的急性肺水肿而致人死亡，并继发感染、肺功能损伤及肺纤维化。MIC的长期影响可导致失明、不育、智力迟钝等。

从表面上看，造成事故的直接原因是负责清洗管道和过滤器的一名新工人在清洗过程中，由于疏忽忘了插入一个防止水渗入MIC储藏罐的专用的碟片，结果水流进了储藏罐并导致罐内的压力激增，致使一个减压阀打开，40吨MIC被释放到了城市的上空。但是，麻省理工学院安全工程专家Nancy Leveson教授通过深入分析事故原因，认为管道清洗工在整个事故中只是一个微不足道的角色。管道清洗工作应该在主管的监督下进行，但是这个职位却因为压缩成本而取消了。而管道清洗工的职位较低，也不应该负责插入碟片。退一步说，即使MIC被水污染，如果能按照设计要求将MIC冷藏，也不至于引起压力激增。然而，冷冻设备一直没有运行，而且洗涤器也应该阻挡MIC从管道泄漏出来。但不幸的是，洗涤器也停止了工作。因此，整个事故真正的原因在于长期积累所造成的安全状况恶化，以及对这些状况的熟视无睹。一系列错误的决策将工厂推向深渊的边缘，以至于任何轻微的失误都可能导致重大事故^[3]。可以说，事故的发生是必然的，造成事故的原因是偶然的，整个系统的健康运行状态的极度恶化才是导致事故发生的真正原因。

流程工业系统的核心监控系统称为分布式控制系统(Distributed Control System, DCS)，包含数以千记的各种类型的传感器，遍布整个系统。随着计算机和自动控制技术的发展与进步，DCS中的监测点不断增加，采样周期不断缩短，导致DCS监测数据日益积累，

形成了海量 DCS 监测数据集,以数据变量的形式实时记录了整个系统的运行信息。DCS 监测数据集反映了整个系统在全生命周期的运行状态、寿命状态与工作状态,从系统层面记录了整个系统的健康状态信息,是整个系统在不同时刻、不同工况下的下的真实记录,蕴含了系统运行状态的内在演化规律与本质。因此,如何从庞大的 DCS 监测数据集中挖掘有用信息,发现系统内在运行规律,判断系统整体运行态势,快速准确地定位和排除故障,评定系统安全运行等级,对于减少生产系统重大事故、避免人员伤亡、减少财产损失和保证系统平稳健康运行具有重要的社会和经济意义。

1.2 过程监控系统在生产企业中的现状

流程工业系统企业级实时过程监控将企业中所有生产要素上的控制信号、监测信号和设备状态信号集中上传到中央控制系统的数据库中,并通过计算机的显示设备以一定的形式(见图 1-1),展示给操作工人、调度人员、工程师和企业管理人员,作为各类人员保障企业正常生产的技术依据。

设备状态监测与故障诊断技术(Condition Monitoring and Faults Diagnosis, CMFD)是美国在 20 世纪 60 年代提出的一门涵盖多个学科领域的综合性技术。它涵盖了包括传感器、嵌入式技术、测试技术、信号处理技术、模式识别、现代通信技术、高性能计算机技术、人工智能等多个学科领域,是一门交叉学科。

设备故障诊断技术和人类对设备的维护维修方式紧密相连。在工业生产早期,生产规模较小,设备的技术水平和复杂度较低,设备间关联关系少,产品的维护维修更多地依赖于熟练技术工人的经验,以及对事故案例的总结。由于当时的生产设备简单,设备间依赖程度低,一台设备的损坏不会对整个企业生产造成太大的影响,设备的维护维修成本相对较低,没有引起人们足够的重视。大约在 20 世纪 60 年代,美国军方认识到了定期维护维修的一系列问题,从而将被动定期维护修改为基于设备过程监控的主动预知维修。这种主动预知维修通过对设备运行状态的过程监控,可以发现设备的潜在故障源,从而可以将事故扼杀在萌芽状态,避免灾难性事故的发生,同时还可以降低因为过度维修所造成的维修成本,具有良好的社会效益和经济效益,因此很快被工业生产的各个部门采纳。

以流程工业为代表的分布式复杂机电系统内部的设备空间位置分散、数量庞大、种类众多、各个设备间关系复杂,在连续的生产过程中持续产生物质、能量和信息的交换,具有海量性、多源性、分布性、层次相关性和整体性。其对应的 DCS 监测数据集难以建立严格的解析表达式,具有复杂性和动态性。而且,由于流程工业对于生产安全有着较高的要求,对于事故的容忍度较低,使得 DCS 监测数据集中正常状态样本数量远远大于异常状态样本数量,具有非平衡性。同时,系统中的故障或异常状态通常不具有代表性,重复出现的概率极小,具有非典型性。

故障通常是指系统出现异常现象导致其难以完成所规定的功能。流程工业系统中故障的含义根据问题的出发点不同分为两类:一类是指控制回路部分的故障,如传感器执行机构或过程部件失灵等(Frank et al. 2000);另一类是指生产过程中的故障,包括控制回路、无控制回路的监测变量、系统硬件以及人员误操作(MacGregor, et al. 1996; Dauia and Qin, 1998; Qin, 2003)。

第一类故障可以通过操作人员实时观察如图 1-1 所示的监控数据,结合专业知识和工

作经验做出合理的判断,随时调整控制回路的数据,保证生产的正常进行。但是,流程工业系统中的设备数量使得需要监测和控制的回路数动辄数以千计,数据的采样周期以分钟或秒为单位。如果按照如图 1-1 所示方式,将系统中的设备以组态的形式在计算机中存储和显示,则需要多个显示屏才能完全显示信息。

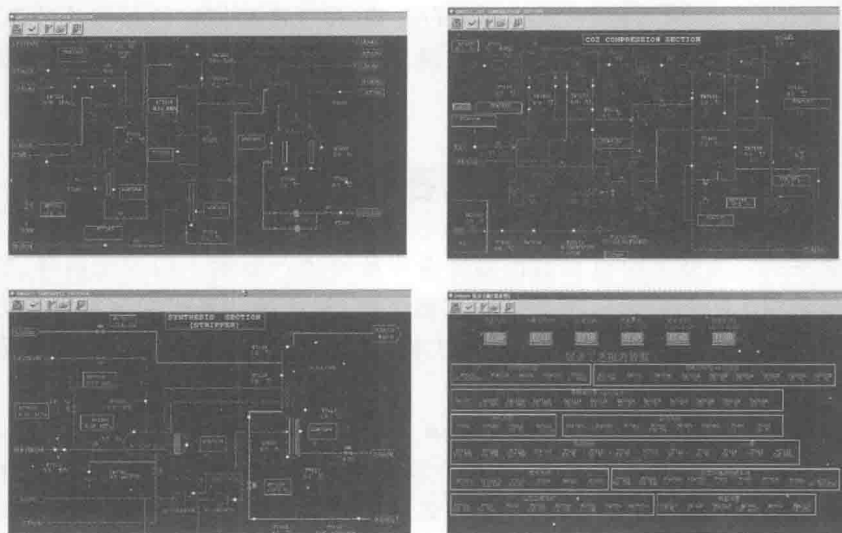


图 1-1 DCS 监测屏幕

心理学的研究表明,人类在同一时刻最多只能关注 7 件不同的事。即使同时设置多块显示屏,将系统中所有设备信息同时显示,也没有人能在一个采样周期内同时关注上千个数据,并给出准确的判读。因此,流程工业系统的企业中都是采取多人同时监控系统、协同操作的方式来保证企业生产的正常进行,如图 1-2 所示。



图 1-2 DCS 控制室

流程工业系统按照生产流程分为多个工段,每个工段设定为一个车间,每个车间都有一个 DCS 控制室,如图 1-2 所示。在控制室中设有多个终端与 DCS 系统相连,将该工段中的系统状态信息实时反映到监控屏幕上,由一线的操作工人负责监控。在每个工段,DCS 需要监测的点位都少则数十,多则上百,需要显示页面数远远大于控制室终端的数目。一般情况下,每一个操作员需要负责十几个系统页面上的数据。因此,即使一线的操作工人在 8 小时的工作时间内目不转睛地盯着计算机屏幕,也无法顾及所有的系统页面。

在某化工企业调研中,我们对一线的操作工人进行了访谈。根据这些操作工人的讲述,虽然工厂要求他们每人负责十几个组态页面,上百个传感器信号的监测,但是他们实际上只是根据经验,挑选比较重要的十几个,甚至是几个点位加以关注,其他大多数监测数据实际上是被忽略的。而什么点位是重要点位并没有相关的技术手册予以指导,完全是依靠老工人传授经验以及自己的摸索。至于监测数据后出现异常如何处理,则绝大多数依靠以往的案例和工人自身的经验。相关的技术指导手册只是收集整理了不多的典型事故案例,并不能应付稍微复杂的系统故障。在一线生产监控中,人的因素非常重要,操作人员的工作态度和工作经验决定着企业生产能否正常进行。

第二类故障在生产过程中发生的概率远远高于第一类故障,且无法用仪器仪表直接测量,需要对监测数据和控制数据进行进一步的分析,结合专业知识,才能做出准确的判断。流程工业系统自身固有的复杂性和层次关联性,使得故障在系统中发生传递,导致故障源往往远离故障发生点。在实际生产企业中表现出来的就是故障源所对应的监测数据和故障发生点所对应的监测点数据并不是由同一个工人甚至同一个车间负责,如图 1-3 所示。这就导致了在某一个操作工人看来在监测范围允许内可以被忽略的数据偏离却正是导致另外一个操作工人监测到的异常数据的原因。

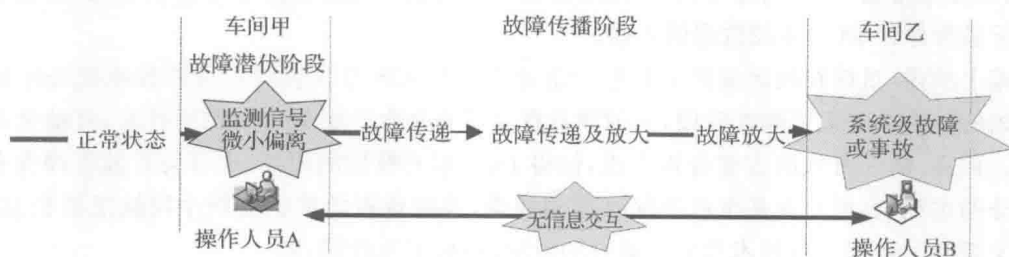


图 1-3 信息不良导致事故

由于一个 DCS 系统中的操作员工数以百计,并且分布在不同的车间。若要满足操作人员之间点对点的传播,则需要的信息交互网络结构如图 1-4 所示。

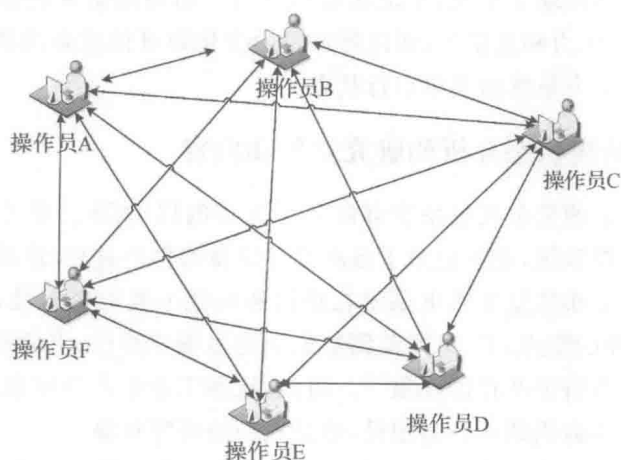


图 1-4 操作人员需要的信息交互网络结构

如图 1-4 所示, n 个操作人员要想进行有效的信息交互, 需要的通信网络数目的数量级为 n^n 。随着操作人员的增加, 信息交互渠道的数量规模以 n 阶增长。现有的 DCS 控制室并没有一个有效的信息交流机制保证操作员工之间可以流畅地交流信息, 因此导致故障溯源异常困难。在实际生产中, 很多事故发生后, 难以找到故障源。从企业的生产事故报告中可以发现, 70% 以上事故并没有找到真正的事故原因。

在企业调研中我们还发现, DCS 中虽然对于监控数据的简单分析, 但是仅限于给出单点的时序图像, 以及报警信息等, 从未涉及点与点之间的相关性分析。而且缺乏对于历史数据的分析。DCS 本身自带的数据存储系统极小, 最多只能保留 3 天的监测数据。超过时间的监测数据将被自动清除。如果想要保留监测数据, 需要企业花费巨额资金另行购买实时数据库。虽然企业购买了一些国外知名企业的故障分析系统, 但是由于流程工业系统故障本身的非平衡性和非典型性, 使得这些故障分析系统在实际生产中对于减少事故发生率所起的作用很小。企业对于 DCS 监控数据的应用始终停留在人工监控阶段, 基于数据驱动的系统运行健康状态自动分析技术还处于理论研究阶段, 并没有一套完整的满足企业安全生产监控需求的解决方案。DCS 监控系统积累下来的宝贵历史数据没有为企业创造任何经济效益, 被企业视为无用的“数据垃圾”。加之存储介质的成本以及数据管理的成本, 使得企业普遍缺乏妥善保存 DCS 监控数据的意愿。

综上所述, 目前我国流程工业生产企业中, 以 DCS 为代表的过程监控系统还停留在最原始的单人单点人工操作阶段, 人的因素在企业安全生产中所占的比例过大, 自动化程度极低。而且, 缺乏有效的数据分析手段, 使得 DCS 监测数据的积累、保存与有效管理没有得到充分的重视, 有的企业甚至从不保存监测数据, 导致现有的数据分析手段缺乏数据基础, 反之又阻碍了数据分析技术与方法研究的发展, 造成了恶性循环。

1.3 数据驱动系统状态分析方法及国内外研究现状

系统状态是一个用于描述系统动态特性的概念, 包括系统相互关联的各组成单元及其进行的动作。对于小单元或小系统, 系统状态通常由一组物理量来表征, 如电机电流、轴承震动、转子转速、温度、压力和流量等, 通过物理量的变化即可描述系统状态的变化^[4]。本书所涉及状态主要是指整个系统的健康运行状态。

1.3.1 数据驱动系统状态分析的研究对象和内容

以流程工业为代表的复杂机电系统都配备了完善的以 DCS 为核心的生产设备故障诊断和生产运行状态监控系统, 完整记录了系统所有设备的生产过程, 形成了一个庞大的高维海量数据信息。由于分布式复杂机电系统在结构和功能上具有整体性、层次性、分布性、相关性、复杂性和动态性的特点, 其 DCS 监测数据集也呈现非线性、多源性、非平衡性、非典型性和非线性等多种复杂特征并存的现象^[5]。因此, 流程工业生产系统积累下来的 DCS 监测数据集是整个系统全生命周期的真实记录, 也是我们的研究对象。

系统状态监测与分析的目的就是为了对系统的状态做出科学合理的判断, 尽早发现系统状态的潜在变化, 实现故障的早期诊断, 并对系统当前状态和未来走势做出科学判断, 实

现系统状态的评估与预测。对于单元或设备来说,故障是指因某种原因不能完成规定功能或危害安全的现象^[6]。由于复杂机电系统的各个部分高度耦合,紧密联系,任何局部微小的故障或扰动都会通过系统介质迅速传播,造成大面积的数据异常,湮没真正的故障源。需要在企业层面实现故障模式识别和溯源。同时,如何深度挖掘数据集内部蕴含的系统动态演化规律,得出系统的健康运行趋势,对系统设备给出科学的大修预警,制定完善的安全等级评定体制,也是状态分析需要研究的内容。因此,研究基于多特征数据的复杂系统健康状态的评估、预测理论与方法具有非常重要的理论价值和实际应用价值。

1.3.2 系统状态分析研究现状

分布式复杂机电系统的运行状态分析方法可分为基于解析模型、基于知识和基于数据分析等三类方法^[7,8]。

基于解析模型的状态分析方法又称为解析冗余法,需要已知系统的解析表达式。该方法以系统的数学模型为基础,利用状态观测器、卡尔曼滤波器、参数估计辨识和等价空间状态方程等方法产生残差,根据残差的产生方式可细分为状态观测法、参数估计法和等价关系法等。然后基于某种准则或阈值对残差进行分析与评价,以实现状态和故障诊断,与系统的机理模型紧密结合,可以方便地实现监控、容错控制、故障重构等。然而,对于具有大量的影响因素和变量的分布式复杂机电系统来说,这些变量之间只有少数在设计阶段具有确定的解析表达式,在运行阶段它们之间的影响关系是复杂的,实际应用中很难建立解析表达式^[9]。

基于知识的状态分析方法是领域专家的启发式经验或模型知识为核心,找到局部故障和系统异常状态之间的因果关系,有专家系统、模糊推理、神经网络、符号有向图^[10]、故障树^[11]和定性趋势分析^[12]等方法。这类方法在复杂机电系统状态分析中存在两个主要问题:①这类方法依赖于系统的先验知识,其有效性依赖于知识的准确度和完整性,而实际应用中往往难以实现;②为了实现知识的表达,在建模过程中往往采用系统的静态知识,无法反应系统的动态过程,或者动态过程的知识急速扩大了搜索空间,超过了实际的计算和存储能力^[13]。

基于数据分析的方法,也称作数据驱动的方法。与前两类方法相比,这类方法对于系统的解析表达式和先验知识没有严格的要求,而是从系统的大量历史数据和实时数据中得到系统变量间的影响关系或知识,避免了基于解析模型的方法难以获取系统精确解析模型的问题,以及基于知识的方法在分析推理过程中的先验知识的获取问题和动态知识的表达问题。基于数据驱动的复杂系统状态分析方法目前已经在化工、冶金、机械、物流等多个行业得到了广泛的应用。美国 University of Minnesota 于 2002 年主办了题为“IMA hot topics workshop: Data-driven control and optimization”的研讨会。IEEE 从 2008 年开始举办 The IEEE Int Workshop on Defect and Data-driven Testing,专注于各类基于数据驱动的数据异常检测和故障诊断技术。在国内,国家自然科学基金委员会于 2008 年召开了“基于数据的控制、决策、调度和故障诊断”研讨会。2009 年第三届全国技术过程故障诊断与安全学术会议的三场大会报告中有两场是关于基于数据驱动的故障检测、诊断与辨识技术。

分布式复杂机电系统的运行过程中积累了海量多态的数据。然而,随着系统规模的扩

大和复杂程度的提高,未经处理的原始数据根本无法说明系统运行状态。这些海量数据可能蕴含着系统状态为正常或异常的所有征兆,目前已有许多的理论和方法将这些征兆从数据中提取或分离出来。从数据分析的角度看,经过转换而对状态分析有价值的数据被称为特征数据^[14],将这类从原始数据转换为特征的一大类方法从广义上归入特征提取的概念,实际上包含了特征检测、特征选择、特征抽取、特征分类等一系列理论和方法,这些方法又可以进一步区分为统计类方法和非统计类方法。它们都是通过某种映射来实现对原始数据某种意义上的约简。

1.3.3 数据驱动系统运行状态分析

在传统的基于数据驱动的系统分析方法中,主要分为定性分析与定量分析两类方法:定性分析主要是包含专家系统和定性趋势模型 QTA(Qualitative Trend Modeling)两类方法;定量分析可以分为非统计方法和统计方法两大类。本书采用的是基于数据驱动的系统状态定量分析方法。

1. 多元统计方法

以主元分析、偏最小二乘分析、独立分量分析(Independent Component Analysis, ICA)等为代表的多元统计分析方法,其主要思想是通过采用多变量投影的方法将数据降维。这类方法能够建立特征或由特征组成的统计量的控制图,以及时发现系统故障,在流程工业生产中应用很广。PCA 方法是其中较有代表性的方法^[15],其主要思想是将数据通过坐标的平移和旋转变换,找到原始数据变异的最大的几个方向及其投影,从而达到维数约简的目的。PCA 方法对于含有噪声和高度线性相关的多变量监测数据特别有效,它能从原系统变量中提取具有最佳解释能力的综合变量,克服了变量之间的多重相关性造成的信息重叠,并能有效地区分信息与噪声,降低异常点和错误样本对建模的影响。

以 PCA 方法为例,其用于系统状态监测和评估的主要步骤为:①从历史的正常和故障状态数据中分别建立参考 PCA 模型;②对每个参考 PCA 模型计算实时监测数据的多元统计量,常见的有 SPE 或 T^2 统计量;③判别所有模型中最小的多元统计量,其对应的状态作为当前系统状态的评估结果。

在实际应用中,PCA 方法通过将原始数据从高维信息投影到低维子空间,能提供展示直观信息的主元图;如果关心对原始数据的解释程度,可以用主元的累积贡献率或平均特征值来衡量;如果关心模型的预测能力,可以用交叉有效性检验或 Akaike 信息准则等确定需要保留的主元数;对于过程监控和故障检测,一些广泛应用的多元统计量如 SPE 或 T^2 等^[16,17]还可以从特定 PCA 模型中直接计算出来以用于过程在线监控。此外,多尺度 PCA 方法(Multiscale PCA, MSPCA)^[16,17]将传统 PCA 与小波变换的特点相结合,实现了在多尺度上进行故障的监测。

多元统计分析方法的主要问题是正在进行数据降维过程中可能会发生信息丢失,并且其提取的特征或特征统计量与故障类型相关,可辨别的历史故障库难以用这种方法一一建立。传统的多元统计分析方法只对线性问题有效,例如 PCA 方法本质上是对历史数据集所构成的输入空间作线性变换。在很多情况下,数据集具有任意分布,特别是具有非线性关系而不能用线性分类时,传统的 PCA 方法无法使用。以核主元分析(Kernel PCA, KPCA)方法为