

密级:\_\_\_\_\_



**中国科学院大学**  
University of Chinese Academy of Sciences

## 博士学位论文

化工过程中故障检测与故障诊断方法研究

作者姓名: 彭 艺

指导教师: 焦建彬 教授

学位类别: 工学博士

学科专业: 计算机应用技术

研究所: 中国科学院大学电子电气与通信工程学院

二零一四年四月



**Research on Fault Detection and Fault Diagnosis**

**Methods in Chemical Process**

**By**

**Yi Peng**

**A Dissertation Submitted to**

**The University of Chinese Academy of Sciences**

**In partial fulfillment of the requirement**

**For the degree of**

**Doctor of Computer Application Technology**

**School of Electronic, Electrical and Communication Engineering**

**April, 2014**



**中国科学院大学直属院系  
研究生学位论文原创性声明**

本人郑重声明：所呈交的学位论文是本人在导师的指导下独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明或致谢。

作者签名：

日 期：

**中国科学院大学直属院系  
研究生学位论文原创性声明**

本人完全了解并同意遵守中国科学院有关保存和使用学位论文的规定，即中国科学院有权保留送交学位论文的副本，允许该论文被查阅，可以公布该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、汇编本学位论文。

涉密的学位论文在解密后适用本声明。

作者签名：

日 期：

导师签名：

日 期：



## 摘 要

化学工业作为高安全隐患行业，危险事故时有发生，给生产带来巨大损失。行之有效的过程监控和故障诊断方法是预防事故发生、保证系统安全可靠运行的有效手段。针对目前故障检测与故障诊断存在的问题，本文提出了相应的方法并在实验仿真模型中进行了验证。

本文主要工作和创新点如下：

(1) 提出了一种基于稀疏化非负矩阵分解进行故障检测的方法。稀疏化非负矩阵分解对分解结果加入非负限制，因此提取的成分向量能很好地反映样本的局部特征；稀疏化非负矩阵分解对分解矩阵添加稀疏性约束条件，因此能更好地发现稳定、直观的局部特征，并能自由地控制分解后矩阵的稀疏度，提高求解收敛速度等。由于稀疏化非负矩阵分解算法对基矩阵和系数矩阵的初始化是随机的，可能会导致算法不稳定，因此，本文提出采用主成分分析法来对稀疏化非负矩阵分解算法进行初始化。针对化工过程的观测数据序列相关性，提出采用指数加权滑动平均对采样数据进行动态化处理。

(2) 提出了一种基于结构化支持向量机进行故障诊断的方法。由于结构化支持向量机可以根据所处理的输入输出的内部结构构造相应的结构化特征函数，因此可以用来解决工业过程中的多种故障判别问题。为了从特征全集中选取出能够提供较大信息量的特征来建立故障诊断模型，本文提出了一种采用决策树提取特征的方法。为了对故障数据的动态特性进行提取并有效地加以表示，从而实现更有效地故障分类，本文引入过程变量的延时测量值，延时测量值的均值和方差扩展样本的特征，组合成新的特征矩阵。

(3) 提出了一种基于深度学习进行故障诊断的方法。当前化工业采用的大多监督学习算法一般需要获取较多的已标记样本，因此成本较高。本文提出了一种基于深度学习进行故障诊断的算法。该算法具有多层非线性映射的深层结构，可以完成复杂的函数逼近。通过逐层获取输入数据的分布式表示，该算法可以自动学习获取鲁棒的故障特征表达。自动学习的特征加入原有的特征后，可以大大提高原特征的精确度，从而提升分类或预测的准确性。

本文对所提出的方法在田纳西-伊斯曼实验仿真化工过程上进行了验证。实

验结果验证了所提出算法的有效性。

**关键词：**故障检测，故障诊断，稀疏化非负矩阵分解，结构化支持向量机，深度学习



## Abstract

**Yi Peng (Computer Application Technology)**

Directed by: Jian-bin Jiao (Professor)

In the chemical industry, accidents occur, causing disasters and tremendous losses. Effective process monitoring and fault diagnosis can prevent the accidents from happening and guarantee safe and reliable operations of the system. To improve the effectiveness of fault detection and diagnosis, the dissertation proposes some novel methods and verifies them.

The main contributions of this dissertation are as follows:

(1) Non-negative Matrix Factorization with Sparseness Constraints (NMFSC) for fault detection is presented. Because of the non-negative constraints of NMFSC, the extracted component of the vector can reflect the parts-based feature of the sample. Because of the sparseness constraints of NMFSC, the stable and intuitive parts-based feature can be found, and the degree of sparseness can be freely controlled. The initializations of the basis vectors and the coefficient vectors of NMFSC are random and therefore the algorithm may be unstable. We propose to use the principal component analysis to initialize the NMFSC algorithm. In view of the relativity of the observation data in the chemical process, Exponentially Weighted Moving Average (EWMA) is used for dynamic processing.

(2) Structural Support Vector Machines (Structural SVMs) for fault diagnosis is proposed. Because Structural SVMs structure the combined feature representation of inputs and outputs, it can be used to identify multiple kinds of faults in chemical process. In order to select the features, which can provide a large amount of information to construct fault diagnosis model, we present a feature extraction method using a decision tree. Besides, the dynamic characteristics of the fault data are extracted and represented effectively, and more effective fault classification can be achieved. Based on this, the mean and variance of the delay measurements of process variable are used to augment the sample feature.

(3) Deep Learning (DL) for fault diagnosis is put forward. Currently, most supervised learning algorithms need to get more labeled samples in the chemical process, which leads to the higher cost. Base on this, this thesis proposes to use a Deep Learning algorithm for fault diagnosis. The algorithm has the deep structure of multiple nonlinear mapping, and can complete complex function approximation. By obtaining the distributed representations of input data by layers, the algorithm can automatically learn robust fault features. That automatic learned features are added to the original features, can greatly improve the accuracy of the original features, so as to enhance the accuracy of classification or prediction.

The proposed methods and approaches are verified on a public benchmark: Tennessee Eastman Process. The results from the experiment show the effectiveness of the methods.

**KEY WORDS: Fault detection, Fault diagnosis, NMFSC, Structural SVMs, Deep Learning**

## 目 录

摘 要.....	I
ABSTRACT.....	III
目 录.....	V
图目录.....	VIII
表目录.....	IX
<b>第一章 绪论 .....</b>	<b>1</b>
1.1 研究背景与意义 .....	1
1.1.1 故障检测与故障诊断的主要内容.....	3
1.1.2 故障检测与故障诊断的方法.....	4
1.2 国内外研究现状及研究难点 .....	9
1.2.1 故障检测研究现状及难点.....	9
1.2.2 故障诊断研究现状及难点.....	11
1.3 本文的主要研究内容 .....	12
1.4 本文的组织结构 .....	13
<b>第二章 相关基本理论内容.....</b>	<b>15</b>
2.1 非负矩阵分解 .....	15
2.1.1 引言.....	15
2.1.2 非负矩阵分解的基本思想.....	15
2.1.3 非负矩阵分解的目标函数.....	16
2.1.4 非负矩阵分解的迭代规则.....	17
2.2 支持向量机 .....	19
2.2.1 机器学习与统计学习理论.....	19
2.2.2 支持向量机.....	22
2.2.3 支持向量机多分类算法.....	26
2.3 本章小结 .....	28
<b>第三章 基于稀疏化非负矩阵分解的故障检测 .....</b>	<b>31</b>
3.1 稀疏化非负矩阵分解的初始化 .....	31
3.1.1 稀疏化非负矩阵分解算法.....	31

3.1.2 基于 PCA 的 NMFSC 初始化 .....	33
3.2 基于动态特性与 NMFSC 的故障检测 .....	35
3.2.1 基于动态特性的故障表示 .....	35
3.2.2 基于动态化 NMFSC 的故障检测 .....	35
3.3 TEP 过程介绍 .....	37
3.3.1 TEP 介绍 .....	37
3.3.2 TEP 过程变量和故障描述 .....	39
3.3.3 仿真数据描述 .....	42
3.4 TEP 故障检测应用 .....	43
3.4.1 基于 TEP 数据的 NMFSC 算法分析 .....	43
3.4.2 实验结果分析与讨论 .....	44
3.5 本章小结 .....	48
<b>第四章 基于结构化支持向量机的故障诊断 .....</b>	<b>49</b>
4.1 结构化支持向量机 .....	49
4.1.1 结构化机器学习 .....	49
4.1.2 结构化支持向量机算法 .....	51
4.2 基于结构化支持向量机的故障诊断 .....	54
4.2.1 结构化特征函数构造 .....	55
4.2.2 基于 Structural SVMs 的故障诊断算法 .....	55
4.2.3 基于 Structural SVMs 的方法在 TEP 故障诊断中的应用 .....	56
4.3 基于特征优化的结构化支持向量机的故障诊断 .....	59
4.3.1 基于决策树的特征选取 .....	60
4.3.2 基于动态特性的故障表示方法 .....	63
4.3.3 基于 FO Structural SVMs 的方法在 TEP 故障诊断中的应用 .....	65
4.4 本章小结 .....	68
<b>第五章 基于深度学习的故障诊断 .....</b>	<b>71</b>
5.1 深度学习算法 .....	71
5.1.1 引言 .....	71
5.1.2 稀疏自动编码器 .....	72
5.1.3 Softmax 回归分类器 .....	74
5.2 基于深度学习的故障诊断方法 .....	76
5.3 基于深度学习的方法在 TEP 故障诊断中的应用 .....	77
5.3.1 TEP 故障诊断应用 .....	77

5.3.2 与其他方法的实验比较分析.....	79
5.4 本章小结 .....	83
<b>第六章 总结与展望 .....</b>	<b>85</b>
6.1 本文工作总结 .....	85
6.2 未来工作展望 .....	86
<b>参考文献.....</b>	<b>89</b>
<b>附录：主要符号对照表 .....</b>	<b>99</b>
<b>作者简历与论文发表 .....</b>	<b>101</b>
<b>致 谢.....</b>	<b>103</b>



## 图目录

图 1-1 兰州石化公司爆炸着火事故和河北赵县化工厂爆炸事故.....	2
图 1-2 故障诊断过程.....	3
图 1-3 故障检测与故障诊断的实现过程.....	3
图 1-4 数据驱动方法分类图.....	5
图 2-1 结构风险最小化示意图.....	22
图 2-2 最优分类超平面.....	23
图 2-3 特征空间的非线性映射示意图.....	25
图 3-1 基于 NMFSC 的故障检测框架.....	36
图 3-2 TEP 过程工艺流程图.....	38
图 3-3 反应器压力.....	43
图 3-4 反应器温度.....	43
图 3-5 基矩阵的示意图.....	44
图 3-6 基于 PCA 方法故障 1 的检测结果.....	45
图 3-7 基于 NMFSC 方法故障 1 的检测结果.....	46
图 3-8 基于 PCA 方法故障 4 的检测结果.....	46
图 3-9 基于 NMFSC 方法故障 4 的检测结果.....	47
图 4-1 基于 STRUCTURAL SVMs 的故障诊断框架.....	54
图 4-2 故障 4、9 和 11 在变量 9 和 51 的散点图.....	57
图 4-3 决策树示意图.....	62
图 4-4 故障 4、9 和 11 在变量 9 上的示意图.....	64
图 4-5 故障 4、9 和 11 在变量 51 上的示意图.....	65
图 4-6 混淆矩阵示意图.....	67
图 4-7 不同特征子集的 STRUCTURAL SVMs 故障诊断模型的错分率.....	68
图 5-1 自编码神经网络.....	72
图 5-2 基于深度学习的故障诊断框架.....	76
图 5-3 故障诊断率趋势图.....	82





## 表目录

表 3-1 TEP 过程操作模式参数.....	39
表 3-2 连续测量变量.....	40
表 3-3 成分变量.....	41
表 3-4 控制变量.....	41
表 3-5 过程故障描述.....	42
表 3-6 TEP 故障检测率.....	47
表 4-1 二组实验描述.....	58
表 4-2 故障 4、9 和 11 的错分率比较.....	59
表 4-3 故障 2、10、13 和 14 的错分率.....	59
表 4-4 故障 2、10、13 和 14 的实验混淆矩阵.....	59
表 4-5 故障 4、9 和 11 的错分率比较.....	66
表 5-1 故障 4、9 和 11 的实验混淆矩阵.....	78
表 5-2 故障 2、10、13 和 14 的实验混淆矩阵.....	78
表 5-3 故障诊断方法比较.....	81



## 第一章 绪论

随着现代科学技术的进步，化工工业生产系统日趋大型化、复杂化。一旦系统内发生异常状况，若不能及时发现并处理，将会造成巨大的经济损失甚至是人员伤亡。为了预防事故的发生，保证系统安全可靠的运行，必须实施有效的过程监控和故障诊断手段。因此，故障检测和故障诊断对于化工安全具有十分重要的实际应用意义。

### 1.1 研究背景与意义

伴随着国民经济的逐年快速发展，社会各界对工业产品的质量和数量的要求不断上升。作为国民经济的重要组成部分，过程工业，如以化工业为代表的流程工业，由于其特殊的重要性和潜在的危险性，越来越受到人们的重视。大多数化工生产过程具有连续性强、工艺复杂、设备管线阀门繁多的特点，如果存在操作不当、检测不到位或一些不受控制的破坏，很有可能产生故障，从而引起事故发生。另外，很多化学品具有毒性和腐蚀性，一旦发生事故，极有可能造成巨大的财产损失和恶劣的环境影响。

多年来，化工反应器升压爆炸、锅炉断裂、有毒气体泄漏等事故常有发生，不仅给生产带来巨大损失，而且对周边环境造成严重污染和破坏，导致影响非常大的社会公共安全事故。据统计，2001-2008年我国化工企业共发生较大及其以上级别事故119起，其中，死亡510人，重伤105人，轻伤377人[1]。此外，2009年山东临沂金兰物流“9·10”爆炸事故造成18人死亡。2010年1月7日，中国石油天然气集团公司兰州石化公司303厂316烃类罐区一裂解碳四储罐阀门处突然发生泄漏，现场可燃气体浓度达到极限，在当班操作人员紧急处理时发生爆炸，爆炸事故造成了6人遇难，1人重伤，5人轻伤。2010年“7·28”南京丙烯管道爆燃事故造成22人死亡[2]。2011年4月13日，黑龙江大庆市让胡路区喇嘛甸镇富鑫化工厂非法生产偶氮二异丁腈过程中发生爆炸燃烧，现场作业人员共计14人，9人当场死亡。2012年2月28日，位于河北石家庄赵县生物产业园

的河北克尔化工有限公司一号车间发生爆炸，该事故共造成 25 人死亡、4 人失踪和 46 人受伤。2013 年 7 月 21 日，甘肃省张掖市一化工有限公司硫化碱车间发生一起一氧化碳中毒事故，造成 4 人死亡，10 人受伤。2013 年 11 月 22 日，发生在山东青岛的中石化管道公司输油管线破裂，造成原油泄漏，抢修中起火爆炸，事故造成 62 人遇难，136 人受伤，直接经济损失 7.5 亿元。这些事故的发生警示着我们，确保化工工业的安全至关重要。



图 1-1 兰州石化公司爆炸着火事故和河北赵县化工厂爆炸事故

为了保证化工工业生产过程的安全平稳运行，对生产过程进行安全管理极具重要性。而且能够有效地对过程进行监控，对异常状况及时处理，具有非常重要的意义。这不仅可以减少停产时间、增加系统运行的安全性和减少制造成本，还可以使企业避免人员和财产的巨大损失。

过程监控就是监控过程运行，在过程偏离正常工况时及时汇报故障发生，也即是故障检测。对异常状况及时处理就是分析故障并判定故障来源，消除故障影响，实现过程的稳定运行，即是故障诊断。故障检测与故障诊断针对生产中存在的非正常状态采取一系列的措施，是化工过程连续、稳定、安全运行的有力保障。

化工工业生产过程中广泛使用集散控制系统（Distributed Control System, DCS），这一方面加强了对工艺的自动检测和调节能力，扩大了操作人员的信息来源，另一方面则导致系统复杂程度增加，以至于当生产状况发生异常时，面对大量的信号警报，操作人员无从下手，造成了过程中的异常和故障难以及时获得监控，导致所谓的“Data overload, but Information poor”现象时有发生。如何从海量数据中挖掘出有价值的过程信息，确保过程运行状况满足预先设定的各项性能指标，进一步提高过程监控能力，已经成为学术界和工业界亟需解决的问题。

### 1.1.1 故障检测与故障诊断的主要内容

一般说来,化工过程的故障诊断系统包括故障检测、故障识别、故障诊断和过程恢复4个基本过程,构成一个闭环控制回路[3],如图1-2所示。一旦检测到有故障发生,故障识别、故障诊断以及过程恢复程序就会按照各自的顺序执行,否则就只重复执行故障检测程序。但是,在实际的过程监控中,有时并不需要包括上述所有步骤,如不需要故障识别也可实现故障诊断[4]。本文研究的重点是故障检测和故障诊断两个环节。

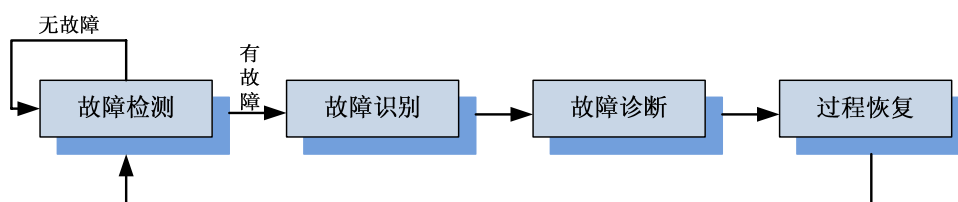


图 1-2 故障诊断过程

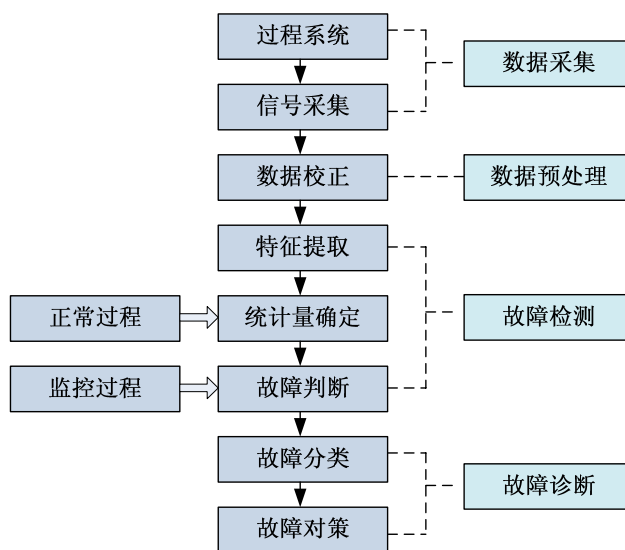


图 1-3 故障检测与故障诊断的实现过程

故障检测和故障诊断不仅直接关系到生产产品的质量,更重要的是它通过对生产过程的监控,可以有效地检测出什么地方、什么设备发生了故障。这些信息不仅能够使系统操作人员和维护人员及时了解生产过程的运行状态,而且能够帮助他们做出适当的补救措施,消除过程的不正常行为,从而大大减少生产过程的危险,提高生产安全性。本文对故障检测的研究主要集中在根据正常过程的统计

信息判断系统运行过程中故障是否发生,对故障诊断的研究主要集中在通过模式识别的方法进行故障分类,确定故障发生的类型。故障检测与故障诊断实现过程如图 1-3 所示。

### 1.1.2 故障检测与故障诊断的方法

随着统计学、信号分析、人工智能理论的引入,故障检测与诊断技术得到极大发展,出现了各种各样的理论与方法。从不同的角度出发,可以把这些方法分为不同的类型。传统的故障检测与诊断技术分为三大类[5]:基于解析模型的方法、基于信号处理的方法和基于知识的方法。随着理论研究的深入、新技术的涌现和相关领域的发展,各种新的故障诊断方法相继出现,促使故障检测与诊断技术的分类方法不断更新。目前常见的分类为基于解析模型的方法、基于知识的方法和基于数据驱动的方法。对于复杂的和多变量的系统,基于解析模型的方法建立精确的数据模型是非常困难的;同时,对于变量彼此耦合的大系统,建立模型也需要付出巨大的成本。基于知识的方法使用定性的模型来获得过程监控的量度,在复杂系统中的应用有一定的局限性。例如,诊断的准确程度依赖于知识库中专家经验的丰富程度和知识水平的高低。基于数据驱动的方法通过监控系统的运行状态,不断检测系统运行过程的变化和故障信息;故障发生后,能够准确识别故障源,有效地监控设备安全可靠运行,迅速地隔离并消除故障。它的优势在于能够将过程数据从高维数据空间投影到低维特征空间,所得到的特征变量保留了原始数据的特征信息,去除了冗余信息,大大降低了工作量。所以,基于数据驱动的故障检测与诊断方法较其他两类方法有着明显的优势,目前已成为研究的热点。这类方法该主要包括:基于机器学习的方法、基于多元统计的方法及基于信号处理的方法等[6-8](图 1-4)。

(1) 机器学习方法:神经网络和支持向量机等机器学习算法通过过程的正常数据以及故障数据来训练各类机器学习算法,进而确定故障的类型,实现故障诊断的目的。

(2) 信息融合类方法:通过利用不同传感器信息之间的互补和冗余关系,对多个信息源加以分析和综合,从而获得较为可靠的监控结果。按照融合信息的不同层面可以分为数据层融合方法、特征层融合方法以及决策层融合方法。

(3) 多元统计类方法:对过程的历史数据进行统计分析,为每个样本计算对

应的监控统计量值,根据正常样本估算出的监控指标置信限来分析当前样本的运行状态。主要分为单变量统计监控方法和多变量统计监控方法。

(4) 粗糙集方法:基本思想是进行属性约简操作,通过删除不重要的过程变量实现过程数据的降维操作,降低过程监控系统的规模和复杂程序,从而提高故障的检测精度。

(5) 信号处理方法:主要是利用各种信号处理方法,如谱分析方法、小波变换方法等,对过程或系统的测量信号进行分析或处理,提取与故障相关的时域或频域特征信息,进行故障检测和诊断。

本文的研究主要是利用基于多元统计分析和机器学习的方法进行故障检测和故障诊断。

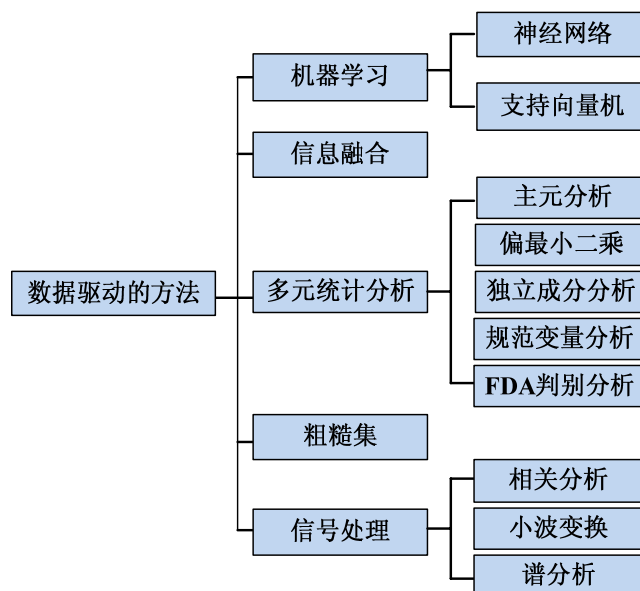


图 1-4 数据驱动方法分类图

### 1.1.2.1 故障检测方法

常用的故障检测方法有主元分析 (Principal Components Analysis, PCA)、偏最小二乘 (Partial Least Square, PLS)、独立成分分析 (Independent Component Analysis, ICA) 等[9-11], 这些方法在理论方面已经有了大量的研究成果, 并得到了广泛的工业应用。

#### (1) 主元分析[12-14]

主元分析是一种应用广泛的多元统计分析方法, 最早由 Pearson 于 1901 年提

出[15]。其基本原理是采集处于正常工况下的过程数据，对其进行归一化处理，得到均值为 0、方差为 1 的数据矩阵  $X$ ，然后进行主元分析建立统计分析模型：

$$X = TP^T + \tilde{T}\tilde{P}^T = TP^T + E \quad (1.1)$$

其中， $T \in R^{N \times k}$ ， $P \in R^{m \times k}$  分别为主元的得分矩阵(score matrix)和载荷矩阵(loading matrix)， $k \leq m$  为主元数，通常采用累积方差贡献率方法和交叉检验法来确定；

$\tilde{T} \in R^{N \times (m-k)}$ ， $\tilde{P} \in R^{m \times (m-k)}$  分别为残差的得分矩阵和载荷矩阵， $E \in R^{N \times m}$  为残差矩阵。

主元及残差的得分和载荷矩阵可以通过对数据矩阵  $X$  的协方差矩阵

$\Sigma = \frac{1}{n-1} X_{n \times m}^T X_{n \times m}$  进行奇异值分解得到，即：

$$\Sigma = [P\tilde{P}]\Lambda[P\tilde{P}]^T \quad (1.2)$$

$$T = XP \quad (1.3)$$

$$\tilde{T} = X\tilde{P} \quad (1.4)$$

式中， $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m)$ ， $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$  为  $\Sigma$  的特征值。经过 PCA 分解之后，原来的变量空间分解为主元空间和残差空间，而且各个得分变量之间是不相关的。PCA 通常采用定义在主元和残差空间的 Hotelling  $T^2$  统计量和平方预测误差 SPE 统计量（又称  $Q$  统计量）对新的过程采样数据进行故障检测。 $T^2$  统计量表征模型内部变化的一种测度，其定义为：

$$T^2 = tD^{-1}t^T = \tilde{X}PD^{-1}P^T\tilde{X}^T = \sum_{i=1}^d \frac{t_i^2}{\lambda_i} \sim \frac{d(n^2-1)}{n(n-d)} F(d, n-d) \quad (1.5)$$

$$D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m) \quad (1.6)$$

$$t = \tilde{X}P \quad (1.7)$$

SPE 统计量表示每次采样在变化趋势上与统计模型之间的误差，是模型外部数据变化的一种测度，其定义为：

$$SPE = ee^T = \tilde{X}(1-PP^T)\tilde{X}^T \quad (1.8)$$

其中， $T^2$  统计量服从  $F$  分布，其控制限可以根据  $F$  分布确定，SPE 统计量服从  $\chi^2$  分布。当  $T^2$  统计量和 SPE 统计量超过控制限时，我们就认为系统中出现了故



障。

### (2) 偏最小二乘[16-17]

偏最小二乘是一种将降维空间的每个元素的输入数据矩阵  $X \in R^{m \times n}$  和输出数据矩阵  $Y \in R^{m \times p}$  的协方差最大化的降维技术，其中  $m$  是样本数， $n$  是预测变量个数， $p$  是响应变量个数。

输入变量矩阵  $X$  和输出变量矩阵  $Y$  的 PLS 分解可表达为：

$$X = \sum_{i=1}^r t_i p_i^T + E = TP^T + E \quad (1.9)$$

$$Y = \sum_{i=1}^r u_i q_i^T + F = UQ^T + F \quad (1.10)$$

其中， $T$  为得分向量 (score)， $P$  为对  $X$  的负荷向量 (loading)， $Q$  为对  $Y$  的负荷向量， $E$  和  $F$  是残差矩阵。通过  $T$  对  $U$  的线性回归得到的内部关系为：

$$\hat{U} = TB \quad (1.11)$$

最终建立的 PLS 模型为：

$$Y = UQ^T + F = TBQ^T + F \quad (1.12)$$

式中， $F$  为模型误差。

应用 PLS 的一种通用方法与 PCA 相同，用于 PCA 的故障检测方法适用于 PLS，建立统计量，判断新的样本统计里是否超出其对应的控制限。

### (3) 独立成分分析[18-20]

相比于 PCA 方法，ICA 提取出来的信息不仅是非高斯的，而且是相互独立的。此外，PCA 只能处理过程数据的二阶统计量信息，而 ICA 提取高阶统计量信息。选择正常工况的过程数据  $X_{normal} \in R^{n \times k}$  ( $n$  为传感器数， $k$  为采样点数)，经过白化（一般利用 PCA），独立成分分析有

$$S_n = W_n X_{normal} \quad (1.13)$$

式中， $S_n \in R^{m \times k}$  是正常工况下  $m$  个独立成分分量， $W_n \in R^{m \times n}$  表示正常工况下的分离矩阵。独立分量数  $m$  由个分量的非高斯度方法来确定。

对独立成分的单变量监控，需对每个独立分量构造一张统计控制图。统计控

制限采用样本分布的  $\alpha/2$  和  $1-\alpha/2$  分位点来确定。其中,  $\alpha$  为显著水平, 一般取 0.01-0.05。

由于独立分量的各个分量满足统计独立, 所以其联合概率密度等于每个独立分量的边缘概率密度函数的乘积, 即:

$$p(s) = \prod_{i=1}^r p(s_i) \quad (1.14)$$

经过 ICA 分析后, 联合概率密度的估计可以转换为每个单变量的概率密度估计问题, 此问题已有很多成熟的算法, 如核密度估计。独立成分的联合概率密度函数得到之后, 可以通过等概率密度曲线来确定独立成分分量联合分布的统计控制置信限。

当在线监控时, 新的测量数据  $X_p$  经

$$\tilde{S}_p = W_n X_p \quad (1.15)$$

可以得到独立元  $\tilde{S}_p$ , 每一个在当前运行条件下得到的独立成分同它们各自独立分量在正常工况下获得的置信限进行对比, 如果独立成分超出了置信限, 表明此过程有故障发生。

### 1.1.2.2 故障诊断方法

故障被检测出来后, 接下来的工作就需要对故障进行识别和诊断, 判断出引起故障的原因。如果过程变量的数目比较大, 并且过程高度集成化, 故障诊断的任务就具有挑战性。故障诊断方法主要包括以下几种:

#### (1) 贡献图法

贡献图 (Contribution Plots), 是指将每个过程变量对统计量的贡献计算出来, 即定义各变量对统计量的贡献, 当检测到故障发生时比较统计量中各个变量的贡献大小, 由此判断故障变量[21]。用于多变量统计监控模型的两种最通用的贡献图是基于  $SPE$  和 Hotelling's  $T^2$  统计量的贡献图, 在故障发生时虽然贡献图表征了变量与正常工况的差异, 但是由于故障变量通过投影会影响到其它变量对隐变量的贡献大小, 这种拖尾效应可能会导致误判。另外, 由于变量之间存在线性相关性, 可能使得多个贡献值同时增加, 从而导致错误的诊断。

## (2) 模式分类法

模式分类即机器学习，是模式识别的主要内容。它的基本思想是通过映射把数据从测量空间映射到决策空间，然后在决策空间进行分类。基于机器学习的方法进行故障诊断[22]，就是对已经存在的故障数据进行分类，将故障诊断问题转化为把系统现有的工作状态归入哪一类的问题，亦即转化为模式分类和识别问题。可用于模式分类的方法很多，比较经典的方法如 Bayesian 分类[23-26]、Fisher 判别[27-29]、K 近邻法（K Nearest Neighbor, KNN）[30]等，近来应用比较广泛的方法有神经网络(Artificial Neural Networks, ANN)[31-32]、支持向量机(Support Vector Machine, SVM) [33-35]、支持向量数据描述（Support Vector Data Description, SVDD）[36-37]，随机森林（Random Forests, RF）[38-39]等。

## (3) 结构化残差法

结构化残差是在基于解析模型的过程监控系统中广泛应用的故障隔离工具[40]。在文献[41]中，Gertler 等首先构造一系列结构化残差，且每个残差只和某些故障相关，而其余故障和该残差无关；然后通过监测这些残差的变化，来确定故障发出的原因。

## 1.2 国内外研究现状及研究难点

### 1.2.1 故障检测研究现状及难点

实际的工业过程存在动态、非线性、大规模、多尺度等过程特性。这些过程特性对于采集到的数据会带来数据自相关、数据均值方差漂移、共线性、变量间非线性关系等数据特点以及数据的非正态。另外，DCS 进行数据采集存储时，可能引起数据缺失、不对齐等。这些数据特性都使得基本的多元统计方法难以直接使用或结果不可靠，易出现漏报或误报等问题。基于此，很多学者提出了改进的多元统计方法[42]。

大多数工业过程都存在动态特性，测量变量并不是序列无关的，当前时刻的测量变量与过去若干时刻的测量变量都有关系。然而，经典 PCA 等方法都是以“样本观测相互独立”作为假设前提的。针对于此，一些学者提出改进的算法来解决动态性问题。Ku[43]等在 1995 年提出动态主元分析法(Dynamic Principal Component Analysis, DPCA)，当前时刻的样本数据用滞后若干时刻的数据扩展，

形成数据的动态性表示, 采用类似的方法还有文献[44]等。为了使动态方法适于变量非高斯的情况, Lee 等[45]提出了动态独立主元方法 (DICA), 通过 ICA 提取测量值不服从高斯分布的部分, 并使用时间延迟变量来增广矩阵, 提取自相关、互相关的独立信号。

PCA, PLS 及 ICA 都是线性化的建模方法, 而实际工业过程本质上都是非线性的。很多学者提出算法来解决非线性问题, 其中得到最广泛应用的是基于核函数的算法。如 Lee[46]等提出了一种新的 KPCA 方法对非线性连续过程进行监测, 通过采用积分算子和非线性核函数使 KPCA 有效地计算高维特征空间的主成分, 证明了该方法与 PCA 监测方法相比具有较大的优越性。

由于存在外界环境变化、设备磨损等因素, 化工工业过程具有缓慢时变特性。此时变特性, 使得过程变量的样本均值、方差以及变量间的关系等都在发生变化。基于此, 一些学者提出了基于递归 (Recursive) 的自适应算法[47-49]对传统的算法进一步改进以克服工业过程的非静态性。基于递归的算法的基本原理是将新的测量数据以一定的权值 (一般是指指数减小的) 包含到待处理的数据矩阵中。由于递归模型可以看成是系统在不同操作点的线性化模型, 因此, 基于递归的自适应算法在一定程度上可以克服非线性的影响。

实际工业过程一般包括多个子系统且包括复杂的耦合关系[42], 涉及较大规模的过程变量, 如果用较远距离的变量进行降维可能对诊断结果不利。此时, 可利用已有的过程知识对过程变量进行有意义的分块, 从而更好地检测故障。这些方法主要包括: 层次 PCA(Hierarchical PCA)[50], 多块 PCA(Multiblock PCA)[51] 和一致 PCA(Consensus PCA)[52]。工业过程数据本质上是多尺度的, 而在单一尺度上建立的模型对于某些尺度上的事件可能并不灵敏。基于此, 文献[53]提出了多尺度主元分析 (Multi-scale PCA), 利用小波将每一信号分解为多个尺度上的信号, 在每一尺度上分别建立相应的模型, 剔除相应于噪声部分的小波系数, 然后在此基础上重构出原始信号。文献[54][55]提出多尺度偏最小二乘法 (Multi-Scale PLS) 对工业过程进行监控。

此外, 为了适应工业过程中同时包含非线性、非平稳、强耦合等多种复杂特性, 研究者进一步取得了如下进展。Zhang 和 Qin[56]基于 Mutilway KICA 提出一种新的故障检测方法, 能够捕捉过程数据的非线性部分。Choi[57]等在数据进

行小波变换及逆变换的基础上, 采用 KPCA 对非线性的变量数据进行降维, 从而降低了误报率和漏报率。Wang[58]等提出的多模态单动态核主元分析法 (Multi-model Single Dynamic Kernel Principal Component Analysis, M-SDKPCA), 能够检测出微弱的故障。

### 1.2.2 故障诊断研究现状及难点

随着对化工工业系统安全性要求越来越高, 基于机器学习的故障诊断, 日渐成为各国研究热点。Li 等[59]介绍基于 SVM 的在线故障诊断方法, 并将其应用于化学流程工业当中。Chiang 等[60]提出了一种 FDA 和支持向量机 (Support Vector Machine, SVM) 相结合的故障诊断方法。该方法首先通过 FDA 降维方法进行特征提取, 剔除不相关特征, 然后通过特征扩展将系统的动态特性加入其中, 最后采用 SVM 分类器进行故障诊断。针对实际的工业过程中, 过程变量极其复杂, 不是绝对的高斯或非高斯分布的情况, Zhang[61]提出 KICA 用于非高斯分布, KPCA 用于高斯分布, 形成非线性动态方法在线检测故障, 然后采用 SVM 对故障类型进行诊断。针对故障数据重叠的多分类问题, 为了提高通用的 FDA 故障诊断性能, He[62]等提出一种变量加权 Fisher 判别分析方法 (Variable-Weighted Fisher Discriminant Analysis, VW-FDA), 它是根据最大化正常和故障数据集的间隔, 用加权变量找出每一个故障的权重向量。Zhu 等[63]人在 2011 年提出采用 KFDA (Kernel FDA) 对故障数据进行特征提取, 然后采用高斯混合模型 (Gaussian Mixture Model, GMM) 和 KNN 两种分类器对故障数据进行分类。该方法在 TEP (Tennessee Eastman Process) 数据集的正常数据、故障 4、故障 8 和故障 14 上得到了验证。Monroy 等人[64]在 2010 年提出了一种半监督的故障诊断方法。该方法首先利用 ICA 进行特征提取, 之后利用具有 Bayesian Information Criterion(BIC)信息的 GMM 和 SVM 进行组合用于最终的故障识别。

在以往的研究成果中, 人工神经网络具有自学习能力, 使得它在故障诊断领域得到广泛应用。当样本数量有限时, 神经网络算法会出现“过学习” [65]现象, 即训练效果良好的算法, 推广能力却可能很差, 对新的未知样本不能做出正确的判断。SVM 将最大分界面分类器思想和核的方法结合在一起, 表现出很好的推

广能力。它利用核函数降低了计算的复杂度，有效地克服了局部极小和维数灾难等传统算法所不可避免的问题，所以受到了越来越多研究人员的关注。但是该算法在计算上存在着一些问题，如训练速度慢、算法复杂而难以实现以及检测阶段运算量大等。此外，SVM 在处理多类问题时基本采用两种策略：**one-against-one** 和 **one-against-all**[66]，使多类问题转化为二分类问题。若采用 **one-against-one** 需要训练 $(n(n-1)/2)$ 个分类器（假设  $n$  为类别个数），采用 **one-against-all**，需要训练  $n$  个分类器。当  $n$  较大时候，需要建立很多分类器，训练起来相对麻烦，分类时需经过多个分类器。采用 **one-against-all** 时还会出现训练数据不平衡的问题导致训练出来的分类器不准确[67]。针对如此等等的问题，近年来关于 SVM 方法的研究，陆续提出了包括算法本身的改进和算法的实际应用。

除此之外，故障检测与诊断领域还存在一些其他亟需解决的问题。例如，大多数学者主要将注意力集中在可认知故障的分类上，对于不可认知故障的诊断、可认知故障与不可认知故障的检测，尚存在一定的空白。

### 1.3 本文的主要研究内容

结合前文所介绍的化工过程中故障检测与故障诊断所存在的研究难点问题，本文的主要研究内容如下：

★ 提出一种基于稀疏化非负矩阵分解（**Non-negative Matrix Factorization with Sparseness Constraints, NMFSC**）进行故障检测的方法。由于 NMFSC 对分解结果加入非负限制，基于 NMFSC 提取的成分向量能很好地反映样本的局部特征。另一方面，NMFSC 对分解矩阵添加稀疏性约束条件能更好地发现稳定、直观的局部特征，并能自由地控制分解后矩阵的稀疏度，具有求解收敛速度快等特点。此外，考虑到化工过程的时序性，本文提出采用指数加权滑动平均对采样数据进行处理；由于 NMFSC 算法对基矩阵  $W$  和系数矩阵  $H$  的初始化是随机的，可能导致每次 NMFSC 的结果都有可能不同，影响算法的稳定性，所以，本文采用 PCA 来对 NMFSC 进行初始化。

★ 提出一种基于结构化支持向量机（**Structural SVM**）进行故障诊断的方法。通过构建一个模型解决化工过程中多类故障判别问题，突破了传统的支持向量机需要通过某种方式构造一系列两类分类器并将它们组合在一起来实现多类分类

的情况。另外，由于对于故障特征全集而言，并不是每一个特征都能够对故障诊断做出同等的贡献，本文提出从特征全集中选取出能够提供较大信息量的特征来建立故障诊断模型。在大多数的工业过程中，测量变量并不是序列无关的，当前时刻的测量变量与过去若干时刻的变量都有关系，都是具有一定的动态特性的，并且不同的故障产生时，同一变量所具有的动态特性也是不一样的。因此，本文引入过程变量的延时测量值，延时测量值的均值和方差扩展样本的原始特征，组合成新的数据矩阵，来表征故障的动态特性。

★ 提出一种基于深度学习（Deep Learning）进行故障诊断的方法。化工过程中故障诊断大多是采用监督学习的算法，总是需要获取更多的已标记样本，然而实际中要想找到这种有利于分类的比较合理的表示方法的数据是比较困难的，深度学习通过建立类似人脑的分层模型结构，对未标记的输入数据逐级提取从底层到高层的特征进行非监督学习，从而能很好地建立底层信号到高层语义的映射关系。基于深度学习的这些特性，本文提出了一种采用深度学习进行故障诊断的方法。此方法是首先对化工数据进行预处理，然后经过稀疏自编码器进行非监督学习，学习出能较好表征原数据的特征，最后采用 Softmax 回归模型进行分类。该模型是 Logistic 回归模型的一般化，可以用来解决类标签  $y$  的可能取值多于两种的情况。除此之外，介绍了其他几种用于故障诊断的方法，以便进行实验对比分析。

#### 1.4 本文的组织结构

第一章，绪论。主要论述故障检测与故障诊断的研究背景和意义，总结当前故障检测与故障诊断的主要内容及方法，分析国内外的研究现状并指出当前研究存在的主要难题。

第二章，非负矩阵分解与支持向量机的相关理论。详细介绍非负矩阵分解的基本思想以及理论推导等，论述机器学习与统计学习理论以及支持向量机的相关内容。

第三章，基于稀疏化非负矩阵分解的故障检测研究。首先论述稀疏化非负矩阵分解算法并对其进行初始化处理。然后，设计动态特性的故障表示以及如何对动态化 NMFSC 进行故障检测。最后，介绍本文的实验平台“田纳西-伊斯曼实

验仿真化工过程”，并在该仿真化工过程上验证本章所提出的算法。

第四章，基于结构化支持向量机的故障诊断研究。论述结构化支持向量机的算法原理，提出一种基于结构化支持向量机的故障诊断算法并基于 TEP 数据进行实验论证。介绍决策树算法以及如何使用此算法进行特征选取。对于测量变量的动态性，提出新的特征表达方式。最后，对所提出的算法进行验证。

第五章，基于深度学习的故障诊断研究。介绍深度学习的基本思想及其常用方法等，设计基于深度学习的故障诊断算法，基于 TEP 数据进行了实验验证。然后，与其他几种有代表性的故障诊断方法进行了实验对比。

最后是对现有工作的总结和对未来工作的展望。



## 第二章 相关基本理论内容

### 2.1 非负矩阵分解

#### 2.1.1 引言

基于数据驱动故障检测、描述多维观测数据的变换方法主要有主元分析、线性判别、因子分析、独立主元分析等。它们共同的特点是：1) 允许负的分解量存在；2) 实现线性的维数约减。

然而，一种好的变换方法应具备两个基本特性：1) 能够使数据的某种隐性的结构变得清晰；2) 能够使数据的维数得到一定程度的约减[68]。主元分析等方法不能很好地描述数据的隐性结构，而且大多是使数据维数线性的约减。

D.D.Lee 和 H.S.Seung 于 1999 年在 Nature[69]上提出了非负矩阵分解 (Non-negative Matrix Factor, NMF)，然后，在 2000 年证明了该算法的收敛性[70]。NMF 是基于乘性迭代规则的，使得分解后的所有分量均为非负值，即纯加性描述，而且在一定程度上实现了非线性的维数约减，即一种好的变换方法应具备的特征之一。非负性限制使得原始数据基于部分表示，从而能更好地反映原始数据的局部特征，而且它更能体现数据的本质特性[69]，由于很多物理数据并不存在负性部分，所以对数据的解释更合理。此外，非负性约束会使 NMF 分解的最终结果在一定程度上呈现出稀疏性，这对数据的解释会变得方便，在组成方式上变得更加直观和清晰，并且稀疏的特性能够抑制外界条件带来的许多不利影响。正是由于具有这些特性，NMF 正逐渐成为各种研究领域所受欢迎的多维数据降维处理的算法之一，如文本分析与聚类、图像分析、盲信号分离、网络安全等。

#### 2.1.2 非负矩阵分解的基本思想

NMF 解决的问题可以描述为[69]：给定非负矩阵  $V \in R^{m \times n}$ ，求解两个非负子矩阵  $W \in R^{m \times r}$  和  $H \in R^{r \times n}$ ，使得

$$V \approx WH \quad (2-1)$$

分解后的两个矩阵组合表达原来的矩阵，分解矩阵  $V$  中的列向量为对矩阵  $W$

中的所有列向量的加权和，而加权系数为矩阵  $H$  中对应列向量中的元素。其中称  $W$  为基矩阵， $H$  为系数矩阵。一般情况下， $r$  的选取需满足  $(m+n)r < mn$ ，即  $r$  远小于  $\min(m, n)$ ，使得  $W$  和  $H$  的维数小于原矩阵  $V$  的维数。 $H$  就是原始矩阵  $V$  在线性子空间  $W$  上的投影，所以用系数矩阵可以表征原始数据矩阵，从而达到降低数据维数的目的。

NMF 问题实际上是一个最优化问题，它利用乘性迭代的方法求得  $W$  和  $H$ 。

### 2.1.3 非负矩阵分解的目标函数

为了寻找一个近似分解过程  $V \approx WH$ ，我们需要先定义目标函数来保证逼近的效果。这样的目标函数有很多种，一般最常用的有两种：基于 KL 散度 (Kullback-Leibler Divergence) 的目标函数和基于欧几里德距离的平方 (Square of Euclidean Distance, SED) 的目标函数。这两个函数都是刻画函数相似度的，NMF 分解后的矩阵要求和原矩阵越相似越好，所以在满足函数最小时求函数的局部最优解。这两种目标函数的形式如下：

(1) KL 散度函数：

$$D(V \| WH) = \sum_{ij} (V_{ij} \log \frac{V_{ij}}{(WH)_{ij}} - V_{ij} + (WH)_{ij}) \quad (2-2)$$

(2) Euclidean distance 函数

$$\|V - WH\|^2 = \sum_{i,j} (V_{ij} - (WH)_{ij})^2 \quad (2-3)$$

此时， $H$ 、 $W$  和  $V$  的所有元素均为非负。两者的度量的下限都是零。当且仅当  $V=WH$  时，该度量的值为零。NMF 的目的是选取最优的  $W$  和  $H$ ，使  $V$  和  $WH$  之间的重构误差尽可能小。

分别对式(2-2) (2-3)进行简化后，NMF 问题转化为以下最优化问题：

(1) KL 散度

$$\begin{aligned} \min \sum_{i=1}^m \sum_{j=1}^n [V_{ij} \log(WH)_{ij} - (WH)_{ij}] \\ \text{s.t. } W \geq 0, H \geq 0, i=1, 2, \dots, m, j=1, 2, \dots, n \end{aligned} \quad (2-4)$$

(2) Euclidean distance

$$\begin{aligned} \min & \sum_{i=1}^m \sum_{j=1}^n [V_{ij} - (WH)_{ij}]^2 \\ \text{s.t.} & W \geq 0, H \geq 0, i=1, 2, \dots, m, j=1, 2, \dots, n \end{aligned} \quad (2-5)$$

其中,  $V_{ij}$  为待分解矩阵  $V$  中的元素,  $(WH)_{ij}$  为分解后因子矩阵  $W$ 、 $H$  乘积矩阵中的元素。

上述的目标函数对于单独的  $W$  和  $H$  均是凸函数, 但是同时对于  $W$  和  $H$  却不是凸函数。因此要找到一个解决上述两个问题的全局最优解是不现实的。所以, 有不少数学优化的方法寻找局部最优解。在实现局部最优解时, 梯度下降法也许是比较容易实现的方法, 但其收敛速度慢。其他的一些方法例如共轭梯度法有更快的收敛性, 但在实际运用中比较复杂。

#### 2.1.4 非负矩阵分解的迭代规则

乘性迭代算法是一个比较好的折中方法, 收敛速度快且算法复杂度低, 解决上述两问题时能很好地处理收敛速度和实际运用之间的矛盾。由于使用的目标函数不同, 相应的更新规则也有所差别。其迭代分解规则为: 算法中  $W$  和  $H$  每次迭代的新值由当前值乘以某个系数而得到, 这样一个系数是由 KL 散度函数或 SED 函数中两次迭代的近似程度来决定的。按照这样的规则不断地迭代, 可以确保  $W$  和  $H$  收敛到一个局部最优的矩阵分解。采用 KL 散度为目标函数时的乘性迭代规则为:

$$H_{\alpha j} \leftarrow H_{\alpha j} \frac{\sum_i W_{i\alpha} V_{ij} / (WH)_{ij}}{\sum_i W_{i\alpha}} \quad (2-6)$$

$$W_{i\alpha} \leftarrow W_{i\alpha} \frac{\sum_j H_{\alpha j} V_{ij} / (WH)_{ij}}{\sum_j H_{\alpha j}} \quad (2-7)$$

其中,  $i=1, 2, \dots, m, j=1, 2, \dots, n, a=1, 2, \dots, r$ 。在  $W, H \geq 0$  的限制条件下, 求  $W$  和  $H$ , 使得  $\|V - WH\|^2$  的值最小,  $W$  和  $H$  的初始化一般都是采用随机初始化方法。

采用 Euclidean distance 为目标函数时, 乘性迭代规则是:

$$H_{\alpha j} \leftarrow H_{\alpha j} \frac{(W^T V)_{\alpha j}}{(W^T W H)_{\alpha j}} \quad (2-8)$$

$$W_{i\alpha} \leftarrow W_{i\alpha} \frac{(V H^T)_{i\alpha}}{(W H H^T)_{i\alpha}} \quad (2-9)$$

其中,  $i=1, 2, \dots, m, j=1, 2, \dots, n, a=1, 2, \dots, r$ 。在  $W, H \geq 0$  的限制条件下, 求  $W$  和  $H$ , 使得  $D(V||WH)$  的值最小,  $W$  和  $H$  的初始化一般都是采用随机初始化方法。

NMF 迭代规则理论推导如下[71], 首先求(2-2)式的梯度, 分别固定  $W$  和  $H$  对  $f(W, H)$  求偏导可得

$$\nabla_H f(W, H) = \sum_i W_{i\alpha} - \sum_i \frac{W_{i\alpha} V_{ij}}{(WH)_{ij}} \quad (2-10)$$

$$\nabla_W f(W, H) = \sum_j H_{\alpha j} - \sum_j \frac{V_{ij} H_{\alpha j}}{(WH)_{ij}} \quad (2-11)$$

沿  $f(W, H)$  的负梯度方向按照一定权重系数递进, 产生迭代公式

$$H_{\alpha j} \leftarrow H_{\alpha j} - \eta_{\alpha j} [\nabla_H f(W, H)] \quad (2-12)$$

$$W_{i\alpha} \leftarrow W_{i\alpha} - \xi_{i\alpha} [\nabla_W f(W, H)] \quad (2-13)$$

其中,  $\eta_{\alpha j}$ 、 $\xi_{i\alpha}$  为对应梯度元素的权重, 文献[71]的权重设置为

$$\eta_{\alpha j} = \frac{H_{\alpha j}}{\sum_i W_{i\alpha}}, \quad \xi_{i\alpha} = \frac{W_{i\alpha}}{\sum_j H_{\alpha j}} \quad (2-14)$$

分别代入(2-12) (2-13)可得

$$H_{\alpha j} \leftarrow H_{\alpha j} \frac{\sum_i W_{i\alpha} V_{ij} / (WH)_{ij}}{\sum_i W_{i\alpha}}, \quad W_{i\alpha} \leftarrow W_{i\alpha} \frac{\sum_j H_{\alpha j} V_{ij} / (WH)_{ij}}{\sum_j H_{\alpha j}} \quad (2-15)$$

求(2-3)式的梯度, 分别固定  $W$  和  $H$  对  $f(W, H)$  求偏导, 可得

$$\nabla_W f(W, H) = WHH^T - VH^T \quad (2-16)$$

$$\nabla_H f(W, H) = W^T WH - W^T V \quad (2-17)$$

由 KKT(Karush-Kuhn-Tucker)优化条件[72]得:

$$\begin{aligned} W_{i\alpha} &\geq 0, H_{\alpha j} \geq 0, \\ \nabla_W f(W, H) &\geq 0, \nabla_H f(W, H) \geq 0, \\ W_{i\alpha} \nabla_W f(W, H) &= 0, H_{\alpha j} \nabla_H f(W, H) = 0 \end{aligned} \quad (2-18)$$

沿  $f(W, H)$  的负梯度方向按照一定权重系数递进, 产生迭代公式

$$H_{\alpha j} \leftarrow H_{\alpha j} - \eta_{\alpha j} [\nabla_H f(W, H)] = H_{\alpha j} + \eta_{\alpha j} (W^T V - W^T WH)_{\alpha j} \quad (2-19)$$

$$W_{i\alpha} \leftarrow W_{i\alpha} - \xi_{i\alpha} [\nabla_W f(W, H)] = W_{i\alpha} + \xi_{i\alpha} (VH^T - WHH^T)_{i\alpha} \quad (2-20)$$

其中  $\eta_{\alpha j}$ 、 $\xi_{i\alpha}$  为对应梯度元素的权重，文献[71]的权重设置为

$$\eta_{\alpha j} = \frac{H_{\alpha j}}{(W^T W H)_{\alpha j}}, \quad \xi_{i\alpha} = \frac{W_{i\alpha}}{(W H H^T)_{i\alpha}} \quad (2-21)$$

分别代入(2-19)、(2-20)得到迭代式

$$H_{\alpha j} \leftarrow H_{\alpha j} \frac{(W^T V)_{\alpha j}}{(W^T W H)_{\alpha j}}, \quad W_{i\alpha} \leftarrow W_{i\alpha} \frac{(V H^T)_{i\alpha}}{(W H H^T)_{i\alpha}} \quad (2-22)$$

文献[71]严格证明了以上两种迭代规则的收敛性。乘性迭代规则保证了迭代过程中的每一步结果的非负性。

## 2.2 支持向量机

### 2.2.1 机器学习与统计学习理论

#### (1) VC 维[73][74]

为了研究学习过程收敛的速度和推广能力，统计学习理论定义了一系列有关函数集学习性能的指标，其中最重要的是 VC 维(Vapnik-Chervonenkis Dimension) [75]。对一个指示函数集，如果存在  $m$  个样本能够被函数集中的函数按照所有可能的  $2^m$  种形式分开，则称函数集能够把  $m$  个样本打散；函数集的 VC 维就是它能打散的最大样本数目  $m$ 。若对任意数目的样本都有函数能将它们打散，则函数集的 VC 维是无穷大。由于 VC 维是建立在一个有限点集被“打散”的基础上，我们首先引出“打散”的概念。

**定义 1** ( $Z_l$  被  $F$  打散) 设  $F$  是一个决策函数候选集，即由在空间  $R^n$  上取值为 1 或-1 的若干函数组成的集合。记  $Z_l = \{x_1, \dots, x_l\}$  为  $R^n$  中的  $l$  个点组成的集合。称  $Z_l$  能被  $F$  打散，或  $F$  能打散  $Z_l$ ，如果对  $Z_l$  中各点  $x_i$  给予任意标号  $y_i = 1$  或-1 ( $i=1, \dots, l$ )，并组成训练集

$$T = \{(x_1, y_1), \dots, (x_l, y_l)\} \quad (2-23)$$

$F$  中总存在一个函数  $f$ ，它能完全正确把该训练分开，即满足

$$f(x_i) = y_i, \quad i=1, \dots, l. \quad (2-24)$$

**定义 2** (VC 维) 设  $F$  是一个定义在空间  $R^n$  上取值为 1 或-1 的函数组成的决策函数候选集。定义  $F$  的 VC 维为它能打散的  $R^n$  中的点的最大个数。换句话说，

若存在  $l$  个点组成的集合  $Z_l$  能被  $F$  打散, 且任意  $l+1$  个点组成的集合  $Z_{l+1}$  都不能被  $F$  打散, 则  $F$  的 VC 维就是  $l$ ; 若任给正整数  $l$ , 都存在  $l$  个点组成的集合  $Z_l$  能被  $F$  打散, 则  $F$  的 VC 维就是  $\infty$ 。

VC 维理论是统计学习理论的最重要的理论基础, 是目前为止对函数集学习性能的最好描述指标。它反映了函数集的学习能力, VC 维越大, 学习机器的学习能力就越强, 同时学习机器也会变得越复杂。

## (2) 推广能力的界[76]

VC 维描述了学习机器的复杂度, 并由此为出发点导出了学习机器推广界的理论。所谓推广性的界是描述经验风险和实际风险之间的关系, 用以刻画经验风险和实际风险的接近程度。它们反映了根据经验风险最小化原则得到的学习机器的推广能力。

### ① 经验风险最小化

统计学习的一个要求是使期望风险最小化, 但是, 可利用的信息只有样本, 不知其分布, 期望风险是无法计算的。因此, 传统的学习方法中采用了“经验风险最小化原则”解决此问题。

**定义 3** (经验风险) 设给定训练集  $T = \{(x_i, y_i), \dots, (x_l, y_l)\} \in (R^n \times Y)^l$ , 其中  $x_i \in R^n, y \in Y = \{-1, 1\}, i = 1, \dots, l$ , 并且给定损失函数  $c(x, y, f(x))$ 。决策函数的经验风险是指

$$R_{emp}[f] = \frac{1}{l} \sum_{i=1}^l c(x_i, y_i, f(x_i)). \quad (2-25)$$

**定义 4** (经验风险最小化原则) 设给定训练

$$T = \{(x_i, y_i), \dots, (x_l, y_l)\} \in (R^n \times Y)^l, \quad (2-26)$$

其中  $x_i \in R^n, y \in Y = \{-1, 1\}, i = 1, \dots, l$ , 又设定一个损失函数  $c(x, y, f(x))$ 。求解分类问题的经验风险最小化原则是: 取定一个由定义在  $R^n$  上, 取值在  $Y = \{-1, 1\}$  上的若干函数组成的决策函数候选集  $F$ , 在集合  $F$  中选取使其经验风险达到最小的函数  $f$  作为决策函数。

### ② 结构风险最小化

统计学习理论系统地研究了对于各种类型的函数集, 经验风险和实际风险之间的关系。根据统计学习理论, 对于函数集中的所有函数, 经验风险  $R_{emp}[f]$  和期望风险  $R[f]$  之间至少以  $1-\delta$  的概率满足如下关系:

$$R[f] \leq R_{emp}[f] + \sqrt{\frac{8}{l} \left( h \left( \ln \frac{2l}{h} + 1 \right) + \ln \frac{4}{\delta} \right)} \quad (2-27)$$

其中  $h$  为函数集的 VC 维,  $l$  为样本数, 并且,

$$l > h, \quad h \left( \ln \frac{2l}{h} + 1 \right) + \ln \frac{4}{\delta} \geq \frac{1}{4} \quad (2-28)$$

式(2-27)中右端第 1 项是经验风险, 第 2 项称为置信区间, 这两项之和称为结构风险; 左端是期望风险, 并可知结构风险是期望风险的上界。由此, 最小化期望风险, 就可以考虑转化为最小化作为期望风险上界的结构风险。

函数集的大小对期望风险的影响, 分析起来是困难的, 由上可知, 可以分析其对结构风险的影响。如图 2-1 所示, 横轴描述函数集的大小, 纵轴描述式(2-27)右端相应的值。显然, 函数集增大时, 经验风险会减少, VC 维  $h$  会增大。由式(2-27), 置信区间是  $h$  的递增函数, 那么此时置信区间会增大。所以要使结构风险达到最小, 应该兼顾函数集对经验风险和置信区间两方面的影响。这就导致了如下的结构风险最小化原则:

**定义 5** (结构风险最小化原则) 设给定训练集

$$T = \{(x_1, y_1), \dots, (x_l, y_l)\}, \quad (2-29)$$

其中  $x_i \in R^n, y \in Y = \{-1, 1\}, i = 1, \dots, l$ 。又设定一个损失函数  $c(x, y, f(x))$ 。结构风险最小化原则是: 选取一个依赖于参数  $t$  的函数集  $F(t)$ , 它随着  $t$  的增加而增大

$$F(t_1) \subset F(t_2), \quad \forall t_1 < t_2 \quad (2-30)$$

对于每个  $t$ , 可以在函数集  $F(t)$  中找到一个使经验风险达到最小的函数  $f$ , 与此对应有一个结构风险的值。选择使结构风险达到最小的  $t = \hat{t}$ , 并以相应的函数  $f^{\hat{t}}$  作为决策函数。

结构风险最小化原则本质就是寻找一个适当大小的决策函数候选集, 然后用经验风险最小化原则在该候选集中选出决策函数。

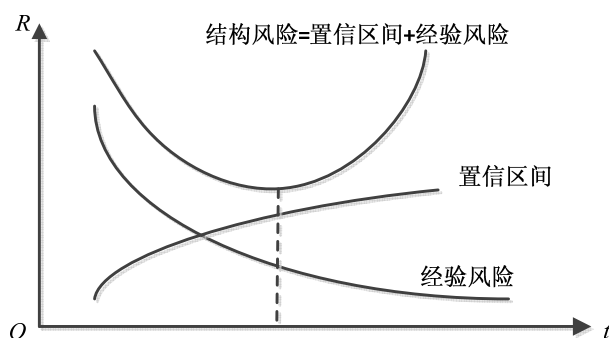


图 2-1 结构风险最小化示意图

### 2.2.2 支持向量机

支持向量机是一种建立在统计学习理论基础上的新型学习机器,其核心思想是通过最小化结构风险,解决学习机器的学习能力和泛化能力之间的矛盾,为有限样本情况下学习能力和泛化能力之间的矛盾提供一个解决方案,概括地说,支持向量机就是通过核函数,将输入空间映射到一个高维特征空间,然后在特征空间中通过最大化分类间隔来构造最优分类超平面。

#### (1) 最优分类超平面的构造

对于两类的分类问题,设训练集  $T = \{(x_1, y_1), \dots, (x_l, y_l)\} \in (R^n \times Y)^l$ , 其中  $x_i \in R^n, y_i \in Y = \{-1, 1\}, i = 1, \dots, l$ , 可以被一个超平面分开,即存在分类超平面  $w \cdot x + b = 0$ , 使得

$$y_i((w \cdot x_i) + b) \geq 1 \quad (2-31)$$

其中,  $w$  是超平面的法线方向,  $b$  为阈值。

对于两类线性可分情况,构建最优线性分类超平面如图 2-2 所示,图中实心点和空心点分布表示两类样本,  $H$  为分类超平面,  $H_1$ 、 $H_2$  为通过各类样本中离超平面最近的点,而且与超平面平行的超平面,  $H_1$ 、 $H_2$  之间的距离  $2/\|w\|$  称为分类间隔(Margin)。所谓最优超平面就是使两类样本无错误地分开,而且使分类间隔最大。从而保证经验风险和置信范围同时达到最小,提高泛化和推广能力。



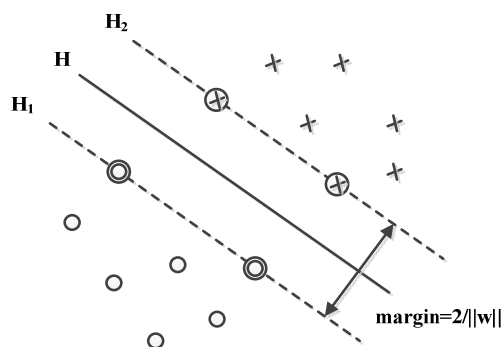


图 2-2 最优分类超平面

## (2) 支持向量机的分类算法

支持向量机基本分类算法，包括线性可分算法、线性不可分及非线性可分算法，具体算法如下：

### ① 线性可分情况

由最大间隔原则，使分类间隔  $2/\|w\|$  最大就等价于最小化  $\|w\|$ ，那么可得如下的最优化问题：

$$\begin{aligned} \min_{w,b} \quad & \frac{1}{2} \|w\|^2, \\ \text{s.t.} \quad & y_i((w \cdot x_i) + b) \geq 1, i = 1, \dots, l \end{aligned} \quad (2-32)$$

上式的最优化问题本质是一个求解不等式约束条件的二次规划问题，为求解这个二次优化问题，引入 Lagrange 函数，可得

$$L(w, b, a) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l a_i (y_i((w \cdot x_i) + b) - 1) \quad (2-33)$$

其中  $a = (a_1, \dots, a_l)^T$  为 Lagrange 乘子向量。

$L(w, b, a)$  是  $w$  的严格凸二次函数，它取最小值时，满足

$$\begin{cases} \frac{\partial L}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^l a_i y_i x_i \\ \frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^l a_i y_i = 0 \end{cases} \quad (2-34)$$

把式(2-34)代入式(2-33)，可得

$$L(w, b, a) = \sum_{j=1}^l a_j - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j (x_i \cdot x_j) a_i a_j \quad (2-35)$$

利用对偶原理, 以及用等价的最小化问题取代最大化问题, 把原优化问题变为:

$$\begin{aligned} \min \quad & \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j (x_i \cdot x_j) a_i a_j - \sum_{j=1}^l a_j \\ \text{s.t.} \quad & \sum_{i=1}^l a_i y_i = 0, a_i \geq 0, i = 1, \dots, l \end{aligned} \quad (2-36)$$

若  $a^*$  为式 (2-36) 的最优解, 则由式 (2-34) 得到最优分类超平面的系数

$w^* = \sum_{i=1}^l a_i^* y_i x_i$ , 即最优分类面的权系数是训练样本向量的线性组合。

式(2-32)是一个不等式约束条件下的二次规划问题, 存在唯一解, 根据KKT(Karush-Kuhn-Tucker)条件, 这个优化问题的解满足:

$$a_i [y_i ((w \cdot x_i) + b) - 1] = 0, i = 1, \dots, l \quad (2-37)$$

多数  $a^*$  为零, 取值不为零的  $a^*$  使公式(2-37)等号成立的样本就是支持向量。根据式(2-37)由  $a^*$ 、 $w^*$  和任意一个支持向量可求出  $b^*$ 。由此, 求得决策函数

$$f(x) = \text{sgn} \left( \sum_{i=1}^l y_i a_i^* (x_i \cdot x) + b^* \right) \quad (2-38)$$

其中  $l$  是支持向量的个数,  $b^*$  为分类阈值。

## ② 线性不可分情况

上面的讨论是在线性可分的前提下进行的, 对于线性不可分的情况下, 训练样本不能满足  $y_i ((w \cdot x_i) + b) \geq 1$  的条件, 可以在约束条件中引入一个松弛变量  $\xi_i$  来实现, 即:

$$y_i ((w \cdot x_i) + b) \geq 1 - \xi_i \quad (2-39)$$

显然, 当  $\xi_i$  充分大时, 样本总可以满足上述约束条件。但是应该设法避免  $\xi_i$  取太大的值。为此在目标函数中对它们进行惩罚, 原始最优化问题改为:

$$\begin{aligned} \min \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i \\ \text{s.t.} \quad & y_i ((w \cdot x_i) + b) \geq 1 - \xi_i, i = 1, \dots, l \\ & \xi_i \geq 0, i = 1, \dots, l \end{aligned} \quad (2-40)$$

其中,  $\xi_i$  大于零表示划分出现错误,  $\sum_{i=1}^l \xi_i$  为总偏离量, 表示训练集中错误分的上界;  $C$  大于零表示惩罚参数, 用来使分类间隔和分类错误达到折中。

当训练样本为线性不可分时, 求解方法和上节线性可分问题相近。对偶问题中, 用来确定 Lagrange 乘子的公式是一样的, 只是现在的乘子需要满足更加严格的条件限制:

$$0 \leq a_i \leq C, i=1, \dots, l. \quad (2-41)$$

### ③ 非线性可分情况

为了在样本非线性可分情况下实现最优分类面, 为此通过非线性变换  $\phi(x)$  将输入空间的数据映射到某个高维空间, 然后在高维空间进行线性分类, 如图 2-3 所示, 从而用线性方法进行求解。

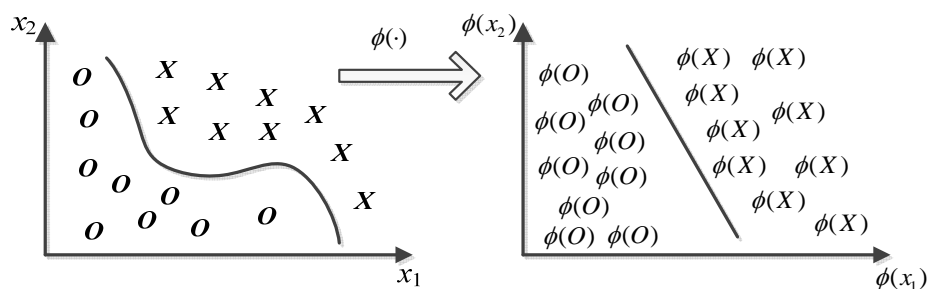


图 2-3 特征空间的非线性映射示意图

这种非线性变换是通过定义适当的核函数  $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$  来实现的。根据泛函有关理论, 只要一个核函数满足 Mercer 条件, 它就对应某一变换空间的内积, 无需知道变换的具体形式。因此, 可以采用适当的内积函数  $K(x_i, x_j)$  实现非线性变换后的线性分类, 而计算复杂度却没增加。为此, 非线性可分的支持向量机的优化问题为:

$$\begin{aligned} \min_{w, b, \xi} \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i \\ \text{s.t.} \quad & y((w \cdot \phi(x_i)) + b) \geq 1 - \xi_i, i=1, \dots, l \\ & \xi_i \geq 0, i=1, \dots, l \end{aligned} \quad (2-42)$$

用核函数  $K(x_i, x_j)$  代替内积  $\phi(x_i) \cdot \phi(x_j)$ , 可得其对偶问题及其转化为:

$$\begin{aligned} \min_a \quad & \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j K(x_i, x_j) a_i a_j - \sum_{j=1}^l a_j \\ \text{s.t.} \quad & \sum_{i=1}^l y_i a_i = 0 \\ & 0 \leq a_i \leq C, i = 1, \dots, l \end{aligned} \quad (2-43)$$

相应的决策函数为：

$$f(x) = \text{sgn} \left( \sum_{i=1}^l y_i a_i^* K(x_i, x) + b^* \right) \quad (2-44)$$

核函数、核参数的选取直接影响 SVM 的性能，目前常见的核函数主要有以下几种：

$$\text{多项式核函数:} \quad K(x, y) = ((x \cdot y) + c)^d \quad \text{其中 } c \geq 0 \quad (2-45)$$

$$\text{径向基核函数:} \quad K(x, y) = \exp(-\|x - y\|^2 / \sigma^2) \quad (2-46)$$

$$\text{傅里叶核函数:} \quad K(x, y) = \frac{1 - q^2}{2(1 - 2q \cos(x - y) + q^2)} \quad \text{其中 } 0 < q < 1 \quad (2-47)$$

$$\text{Sigmoid 核函数:} \quad K(x, y) = \tanh(v(x \cdot y) + c) \quad \text{其中 } v \geq 0, d < 0 \quad (2-48)$$

### 2.2.3 支持向量机多分类算法

支持向量机主要是针对二类分类问题提出来的，不能直接用来解决多类分类问题，如何有效地将其推广到多类分类问题还是一个正在研究的问题。当前主要有如下算法：

#### (1) 一对多 (one-against-all) 算法[77]

该方法对于  $k$  类分类问题，通过构造  $k$  个两类分类器，转化为  $k$  个两类支持向量机的问题来解决。对于在第  $i$  个 SVM 中，第  $i$  类的训练样本的标识为+1，其它的训练样本的标识为-1，最后输出的是两类分类器中输出为最大的那一类。设有  $l$  个训练样本  $(x_1, y_1), \dots, (x_l, y_l)$ ，其中  $x_i \in R^n$ ， $i = 1, \dots, l$ ， $y_i \in \{1, \dots, k\}$ ，则第  $i$  个支持向量机需要解下列问题：

$$\begin{aligned}
& \min_{w^i, b^i, \xi^i} \frac{1}{2} \|w\|^2 + C \sum_{j=1}^l \xi_j^i (w^i)^T \\
& \text{s.t.} \quad (w^i)^T \phi(x_j) + b^i \geq 1 - \xi_j^i, y_j = i \\
& \quad \quad (w^i)^T \phi(x_j) + b^i \leq -1 + \xi_j^i, y_j \neq i \\
& \quad \quad \xi_j^i \geq 0, j = 1, 2, \dots, l
\end{aligned} \tag{2-49}$$

其中， $\phi(x_j)$ 是原样本数据空间映射到高维空间的核函数， $w^i$ 是第*i*类的权重向量， $b^i$ 是第*i*类阈值，可得决策函数为：

$$f(x) \equiv \arg \max_{i=1, \dots, k} ((w^i)^T \phi(x) + b^i) \tag{2-50}$$

该方法该方法需要构造的两类分类支持向量机的数量相对较少，在相同类别的情况下，需要构造的决策函数也少，分类速度较快。但是，如果出现训练数据不平衡的问题，导致训练出来的分类器不准确；而且算法的复杂度较高，实现比较困难。

#### (2) 一对一 (one-against-one) 算法[78]

最初由 Knerr[79]提出的 SVM 多类分类算法，该方法对于*k*类分类问题，任意两类样本构造 1 个支持向量机，那么共需构造  $k(k-1)/2$  个两类支持向量机。对于第*i*和*j*类的训练样本，建立如下的分类问题：

$$\begin{aligned}
& \min_{w^{ij}, b^{ij}, \xi^{ij}} \frac{1}{2} \|w\|^2 + C \sum_t \xi_t^{ij} (w^{ij})^T \\
& \text{s.t.} \quad (w^{ij})^T \phi(x_t) + b^{ij} \geq 1 - \xi_t^{ij}, y_t = i \\
& \quad \quad (w^{ij})^T \phi(x_t) + b^{ij} \leq -1 + \xi_t^{ij}, y_t = j \\
& \quad \quad \xi_j^i \geq 0
\end{aligned} \tag{2-51}$$

其中， $\phi(x_j)$ 是原样本数据空间映射到高维空间的核函数， $w^{ij}$ 是第*i*和*j*类的权重向量， $b^{ij}$ 是第*i*和*j*类阈值，可得决策函数为：

$$f_{ij}(x) = \text{sgn}((w^{ij})^T \phi(x) + b^{ij}) \tag{2-52}$$

在分类过程中采用“投票法”，所有分类器均对测试样本进行分类，在第*i*和*j*类之间进行分类时，若该分类器判定*x*属于第*i*类，则意味着第*i*类获得一票。得票数最多的类别就是最终判定*x*所属类别。该方法优点是：每个 SVM 只考虑两类问题，分类精度大大提高。但是，如果单个两类分类器不规范化，则整个*K*类分类器将趋向于过学习；分类器的数目  $K(K-1)/2$  随着类别数*K*急剧增加，

导致判决性能降低, 而且投票过程中票数相同的两类会存在误分的情况。

### (3) 有向无环图 (DAG-SVM) 算法[81]

该算法是 Platt J.C. 提出的有向无环图 SVM 算法[80] (Directed Acyclic Graph SVM, 简称 DAG-SVM), 该算法的本质是一对一算法思想的具体化。

DAG-SVM 算法将多个两类分类器组合起来形成一个多类分类器, 对  $K$  类分类问题, 需要  $K(K-1)/2$  个分类器, 每个分类器对应两类。对测试样本进行分类时, 首先从顶部的根节点开始, 通过分类结果确定一条遍历路径, 寻找到下一层的邻节点, 而后继续分类, 直到遍历到叶节点, 该叶子节点所表示的类别即为测试样本的类别。

该算法的优点是决策速度比一对多和一对一方法快。缺点是不同的根节点的选择直接影响着分类的精度, 不同的分类器作为根节点, 其分类结果可能会不同, 因此产生了分类结果的不确定性。

### (4) 二叉树 SVMs(Binary Tree SVMs) [82]

二叉树 SVMs 的基本思想是先将所有类别分成为二类分类问题, 再对子分类进行进一步分类, 直至子类中只包含一个类别为止。对于  $K$  类的训练样本, 训练  $K-1$  个分类器, 每个二叉树节点都是一个支持向量机分类器。对测试样本进行分类时, 首先遍历二叉树根节点, 进行第一次分类, 而后依据分类结果继续遍历, 实现样本的进一步分类, 直到叶节点为止, 叶节点所对应的类别即为测试样本类别。

该算法的优点是能将所有样本进行分类, 并且对于一个  $K$  类分类问题, 只需要训练  $K-1$  个两类支持向量机分类器, 分类时不必遍历所有的支持向量机分类器, 因此, 具有较高的分类速度较快。缺点是在分类过程中, 初始分类错误具有遗传性。另外, 基于二叉树的多分类算法中二叉树的结构选取很重要, 哪一个类别先被分出来对分类效果影响很大。

## 2.3 本章小结

非负矩阵分解作为一种新的变换方法, 在各个研究领域受到广泛的关注。本章主要介绍了非负矩阵的基本思想及理论推导过程等。支持向量机由于具有很好的全局优化能力和较好的泛化能力, 在机器学习领域, 受到很多研究者的推崇。本章对支持向量机做了详细的数学描述, 给出了原始问题如何转化为最优化问题

的数学推导过程，之后介绍了支持向量机的几种多分类算法。本章内容为后续章节算法的提出做好了背景知识的铺垫。





### 第三章 基于稀疏化非负矩阵分解的故障检测

虽然 PCA 等传统的方法已广泛应用于化工过程的故障检测，但是，它们大多只是使数据维数线性的约减，而化工工业是非线性过程，同时它们也忽略了对过程数据中重要特征的提取。由于对分解结果加入非负限制，稀疏化非负矩阵分解提取的成分向量能很好地反映样本的局部特征，同时，也接近人们“由局部构成整体”的认知习惯，具有很高的可解释性和预测性。例如，对于化工过程，它可以解释故障的发生或许是仅由几个变量的变化导致的。此外，稀疏化非负矩阵分解对分解矩阵添加稀疏性约束条件能更好地发现稳定、直观的局部特征，并且能自由地控制分解后矩阵的稀疏度。

考虑到化工过程的动态性、算法的不稳定性以及故障产生的偶然性等问题，本文提出首先基于 PCA 进行初始化处理，然后对初始化的故障数据进行动态化表示，最后基于 NMFSC 进行故障检测。

#### 3.1 稀疏化非负矩阵分解的初始化

##### 3.1.1 稀疏化非负矩阵分解算法

稀疏化非负矩阵分解 (Non-negative Matrix Factorization with Sparseness Constraints, NMFSC) [83] 是对基矩阵  $W$  或系数矩阵  $H$  添加稀疏性约束条件而发展的一种非负矩阵分解算法。和传统的 NMF 相比，该算法能够更好地发现稳定、直观的局部特征，并能自由地控制分解后矩阵的稀疏度，具有求解收敛速度快、左右非负矩阵存储空间小等特点。

在介绍 NMFSC 之前，首先简单介绍下  $L_1$ ,  $L_2$  范数以及稀疏的概念。

$L_1$  范数是指向量中各个元素绝对值之和， $L_1$  范数距离公式为：

$$L_1(x, y) = \sum_{i=1}^k |x_i - y_i| \quad (3-1)$$

$L_2$  范数也称欧几里德距离的平方根， $L_2$  范数距离公式为：

$$L_2(x, y) = \sum_{i=1}^k (x_i - y_i)^2 \quad (3-2)$$

所谓“稀疏”是指在大量的数据中，仅使用少量的数据单位来有效表示典型的数据向量。这意味着具有稀疏性的数据大部分为零。基于  $L_1$  和  $L_2$  范数，NMFSC 的稀疏因子定义为：

$$sparseness(x) = \frac{\sqrt{n} - (\sum |x_i|) / \sqrt{\sum x_i^2}}{\sqrt{n} - 1} \quad (3-3)$$

其中， $n$  表示非负向量  $x$  的维数。如果  $x$  中只包含一个非零分量，则稀疏因子等于 1；当且仅当所有元素都相等时，稀疏因子等于 0。稀疏因子在  $[0,1]$  区间内滑动表示稀疏度。 $sparseness(x)$  越大，意味着向量越稀疏，反之越稠密。

稀疏化非负矩阵分解定义为：

假定非负矩阵  $V \in R^{m \times n}$ ，将其分解为两个非负子矩阵  $W \in R^{m \times r}$  和  $H \in R^{r \times n}$ ，并使得目标函数值：

$$E(W, H) = \|V - WH\|^2 \quad (3-4)$$

最小，并满足如下约束条件：

$$sparseness(w_i) = S_w, \forall i \quad (3-5)$$

$$sparseness(h_i) = S_h, \forall i \quad (3-6)$$

其中， $w_i$  表示  $W$  的第  $i$  列， $h_i$  表示  $H$  的第  $i$  行； $S_w$  和  $S_h$  分别表示基矩阵  $W$  和系数矩阵  $H$  期望的稀疏度。

NMFSC 具体实现步骤如下：

- (1) 初始化  $W$  和  $H$  为随机正矩阵；
- (2) 如果要对  $W$  进行稀疏约束，那么投影  $W$  的每一列为非负， $L_2$  范数不变，但  $L_1$  范数达到要求的稀疏度；
- (3) 如果要对  $H$  进行稀疏约束，那么投影  $H$  的每一行为非负， $L_2$  范数单位化，并使  $L_1$  范数达到要求的稀疏度；
- (4) 迭代
  - (a) 如果要对  $W$  进行稀疏约束
    - i. 设置  $W := W - \mu_w (WH - V)H^T$ ；
    - ii. 那么投影  $W$  的每一列为非负， $L_2$  范数不变，但  $L_1$  范数达到要求的稀疏度；

否则，进行标准乘法运算  $W := W \otimes (VH^T) \oslash (WHH^T)$ ;

(b) 如果要对  $H$  进行稀疏约束

i. 设置  $H := H - \mu_w W^T (WH - V)$ ;

ii. 那么投影  $H$  的每一行为非负， $L_2$  范数单位化，并使  $L_1$  范数达到要求的稀疏度；

否则，进行标准乘法运算  $H := H \otimes (W^T V) \oslash (W^T WH)$ ;

注：这里的  $\otimes$  和  $\oslash$  都是基于矩阵元素的乘除法。相应的， $\mu_w, \mu_H$  是非常小的正数，用于控制迭代的进行，保证算法收敛。

在上述的算法中，基于  $L_1$  和  $L_2$  范数的非负稀疏投影是其中一个重要步骤，其用于将矩阵的某一行（或列）对应的向量  $x$  在  $L_1$  和  $L_2$  范数的要求限制下找到与该向量最近邻的非负向量  $s$ ，其算法如下：

(1) 设定  $s_i = x_i + (L_1 - \sum x_i) / \dim(x), \forall i$

(2) 设定  $Z := \{\}$

(3) 迭代

(a) 设定  $m_i := \begin{cases} L_1 / \dim(x) - \text{size}(Z), & i \notin Z \\ 0, & i \in Z \end{cases}$

(b) 设定  $s := m + \alpha(s - m)$ ，这里  $\alpha \geq 0$ ， $s$  满足  $L_2$  范数。

(c) 如果  $s$  中所有元素都是非负的，则返回  $s$ ，结束。否则，算法继续。

(d) 设定  $Z := Z \cup \{i; s_i < 0\}$

(e) 设定  $s_i := 0, \forall i \in Z$

(f) 计算  $c := (\sum s_i - L_1) / (\dim(x) - \text{size}(Z))$

(g) 设定  $s_i := s_i - c, \forall i \notin Z$

(h) 返回(a)

### 3.1.2 基于 PCA 的 NMFSC 初始化

由于 NMFSC 算法对  $W$  和  $H$  的初始化是随机的，可能导致每次 NMFSC 的

结果不同，影响算法的稳定性 [84]。当前，解决 NMFSC 初始化问题的研究主要分两类：一类是利用 PCA 等分解数据集，将分解矩阵中负元素改为零后作为 NMFSC 的初始化矩阵；另一类是利用聚类方法将数据集分类，将聚类中心作为初始化矩阵。然而，聚类个数的确定目前还没完全解决，聚类个数直接影响分类的结果。所以，我们采用主成分分析来对稀疏化非负矩阵分解进行初始化。PCA 是被广泛采用的一种用于数据降维和成分提取的多变量统计方法，它是指基于生产过程的历史数据，通过“统计投影”方法找到能够表达各过程变量之间关系的低维主成分。对过程数据  $X_{n \times m}$  标准化处理

$$\tilde{X} = \frac{X_i - \bar{X}}{S} \quad (3-7)$$

其中， $\bar{X}$  表示均值， $S$  表示方差。对协方差矩阵  $\tilde{X}^T \tilde{X}$  进行奇异值分解，得到其特征值和特征向量  $P$ 。特征向量  $P_r$  (对应前  $r$  个最大的特征值  $\lambda_i$ ) 初始化 NMFSC 的分解矩阵  $W$ 。 $r$  的确定通过累积贡献率确定，若前  $r$  个主成分的方差和占了总方差的很大一部分 (如 80%-90%)，则可以用前  $r$  个主成分代替原  $n$  个变量。累计贡献率按下式计算

$$\frac{\sum_{i=1}^r \lambda_i}{\sum_{i=1}^n \lambda_i} \geq \alpha \quad (3-8)$$

在我们的实验中，设  $\alpha$  为 0.85。

得分矩阵表示如下

$$T_r = \tilde{X} P_r \quad (3-9)$$

由  $T_r$  初始化 NMFSC 分解矩阵  $H$ 。

由于 NMF 算法的非负性约束，为了能用  $T_r$  和  $P_r$  对  $W$  和  $H$  进行初始化，定义如下：

$$P(W) = \max(0, W) \quad (3-10)$$

$$P(H) = \max(0, H) \quad (3-11)$$

然后，由  $P(W)$ ， $P(H)$  分别初始化 NMFSC 分解矩阵  $W$  和  $H$ 。

## 3.2 基于动态特性与 NMFSC 的故障检测

### 3.2.1 基于动态特性的故障表示

对于大多数工业过程而言，测量变量并不是序列无关的，当前时刻的测量变量与过去若干时刻的测量变量都有关系。由于观测数据序列相关性是实际过程数据普遍具有的特点，探讨适合于序列相关数据的描述方法是非常必要的。最常用的一种方法是用前  $l$  时刻的观测数据扩展当前时刻  $t$  的样本数据，扩展后的数据矩阵为[43]：

$$V = [V_t \ V_{t-1} \ \cdots \ V_{t-l}]^T \quad (3-12)$$

这种方法虽然解决了样本数据之间的动态相关性，但是输入数据的维数也增加了原来的  $l$  倍。基于此，我们采用一种简单的方法：指数加权滑动平均对采样数据进行处理。过程如下：

$$\begin{aligned} \hat{V}_t &= w_1 V_t + w_2 V_{t-1} + \cdots + w_N V_{t-N+1} = \sum_{i=1}^N w_i V_{t-(i-1)} \\ s.t \quad &w_i > 0, \sum_{i=1}^N w_i = 1 \end{aligned} \quad (3-13)$$

$V_t$  为变量  $V$  的第  $t$  个采样点； $w_i$  为权系数。 $N$  为滑动窗口的宽度，在我们的实验中经验取值为 3。

### 3.2.2 基于动态化 NMFSC 的故障检测

基于 NMFSC 进行故障检测，将过程数据分解成基向量，用基于部分的特征表示出来，能有效地提取过程数据的局部特征。使用 PCA 进行初始化，可以解决算法不稳定的问题。使故障动态化表示，可以表达系统变量间的相关性信息。本文基于动态化 NMFSC 进行故障检测的框架如图 3-1 所示。

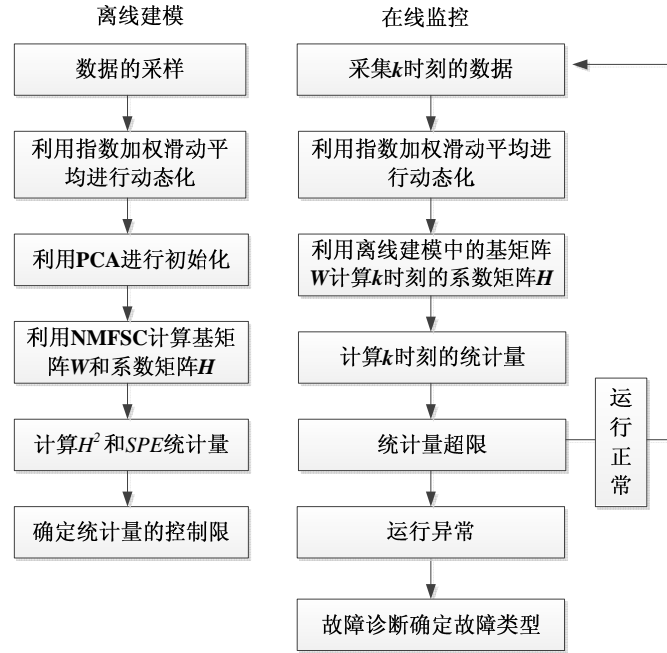


图 3-1 基于 NMFSC 的故障检测框架

经过指数加权滑动平均算法的预处理以及利用 PCA 对 NMFSC 初始化后，下面构造基于 NMFSC 的故障检测模型。首先，定义两个统计量  $H^2$  和  $SPE$ ，分别表征模型内部变化的测度和模型外部变化的测度：

$$\begin{aligned} H^2 &= \hat{H}^T \hat{H} \\ \hat{H} &= (W^T W)^{-1} W^T V \end{aligned} \quad (3-14)$$

其中， $W$  是由正常数据经过 NMFSC 分解得到的， $(W^T W)^{-1} W^T$  表示  $W$  的伪逆矩阵； $V$  是异常数据； $\hat{H}$  表示对  $H$  的估计。

$$\begin{aligned} SPE &= e^T e = (V - \hat{V})^T (V - \hat{V}) \\ \hat{V} &= W \hat{H} = W ((W^T W)^{-1} W^T V) \end{aligned} \quad (3-15)$$

其中， $\hat{V}$  表示对异常数据  $V$  的估计。

在主成分分析中，通常假设信号为正态分布，故控制限可以通过正态分布的置信区间来确定。而在 NMFSC 检测中，经过分解所得到的信号并不一定服从正态分布。因此，统计量的控制限需要采用其他方法来确定。目前针对这种情况，大多学者采用非参数化的方法。文献[85]采用非参数化方法—核密度估计 (Kernel Density Estimation, KDE) 来确定过程的控制限。本文借鉴此思想来求取控制限，核密度估计公式为：

$$\hat{p}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) \quad (3-16)$$

其中,  $n$  为样本数,  $h$  为平滑参数,  $x$  是待求的控制界限值,  $x_i$  是由样本数据集中得到的观测值。  $K$  为核函数, 我们选取高斯核:

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \quad (3-17)$$

则核密度估计为:

$$\hat{p}(x) = \frac{1}{\sqrt{2\pi}nh} \sum_{i=1}^n e^{-\frac{(x-x_i)^2}{2h^2}} \quad (3-18)$$

基于初始化的稀疏化非负矩阵分解建立化工过程中的故障检测模型的详细流程如下:

(1) 模型训练过程:

- ① 对正常条件下的数据进行指数加权滑动平均法处理, 并使其标准化;
- ② 对数据用主元分析初始化稀疏化非负矩阵分解;
- ③ 采用 NMFSC 算法分解正常数据矩阵, 得到基矩阵  $W$  和系数矩阵  $H$ ;
- ④ 计算正常数据的  $H^2$  和  $SPE$  统计量;
- ⑤ 利用核密度估计确定  $H^2$  和  $SPE$  的控制限。

(2) 在线故障检测:

- ① 对发生故障时的新数据进行如上①的处理;
- ② 计算故障数据的检测统计量  $H^2$  和  $SPE$  ;
- ③ 检测故障数据统计量是否超出控制限。如果超出控制限, 说明有故障发生; 否则, 无故障发生。

### 3.3 TEP 过程介绍

为了验证本文所提出算法的性能, 本文选取该领域内公认的数据集合 TEP (Tennessee Eastman Process) 作为数据源进行故障检测与诊断算法的验证。下面对 TEP 过程进行详细的介绍。

#### 3.3.1 TEP 介绍

田纳西-伊斯曼过程 (TEP) 是基于实际的化工工业过程的仿真案例。该过

程是由美国 Eastman 化学公司过程控制小组的 Downs 和 Vogel 在 1993 年创建的 [86]。建立该过程的目的在于为评价过程控制和监控方法建立一个现实的工业平台 [87]。大量文献引用它作为数据源来进行控制、优化、过程监控、故障诊断等方面的研究。

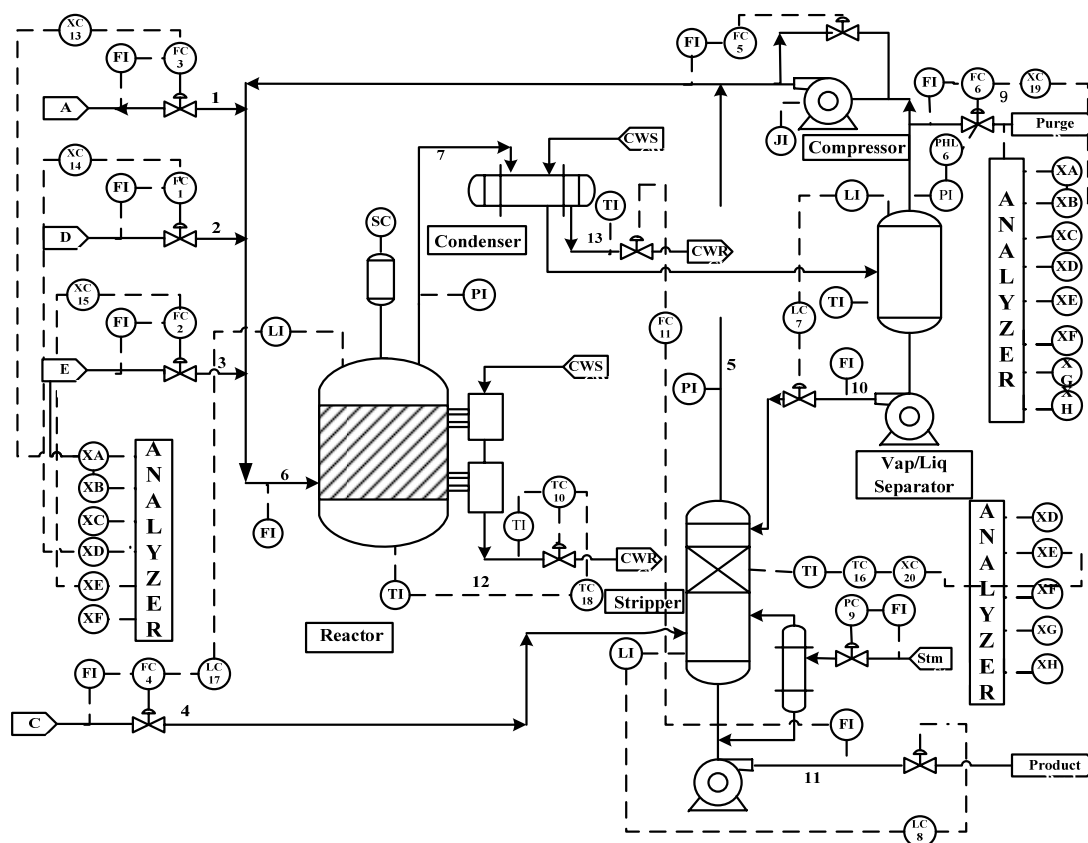
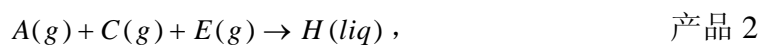
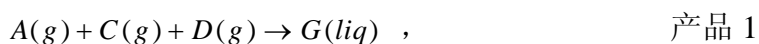


图 3-2 TEP 过程工艺流程图

TEP 过程流程图如图 3-2 所示。该过程由五个反应单元组成：反应器 (Reactor)、冷凝器 (Condenser)、分离器 (Separator)、汽提塔 (Stripper) 和压缩机 (Compressor)；共有八种成分：A、B、C、D、E、F、G 和 H，其中 A、C、D、E 为四种气态反应物，B 为催化剂，F 为副产物，G 和 H 为最终产物。在反应器中进行的化学反应如下：







其中,  $g$  代表气体,  $liq$  代表液体。所有的反应都是不可逆的、放热的, 并且对反应物浓度来说, 近似为一阶系统。反应速率是温度的函数, 其中生成  $G$  的反应比生产  $H$  的反应有更高的活化能, 因此对温度更加敏感。

反应物在反应器中经过催化剂催化作用生成液态产物。之后产品进入冷凝器进行冷却, 紧接着送入到冷/液分离器。从分离器出来的蒸汽通过压缩机再循环进入反应器。为了防止过程中惰性成份和反应副产物产品的积聚, 必须排放一部分再循环流。来自分离器的冷凝成份(流 10)被泵送入汽提塔。流 4 用于流 10 中的剩余反应物, 这些剩余反应物通过流 5 与再循环流结合。从汽提塔底部出来的产品  $G$  和  $H$  被送到下游过程。

根据  $G/H$  质量比的不同, TEP 过程具有六个操作模式, 本文采用的数据为基本工况下产生的。见表 3-1:

表 3-1 TEP 过程操作模式参数

模式	G/H 质量比	生产率
1	50/50	G:7038 kgh-1; H: 7038 kgh-1 (基本工况)
2	10/90	G:1408 kgh-1; H: 12669 kgh-1 (优化基工况)
3	90/10	G:10000 kgh-1; H: 1111 kgh-1 (优化操作)
4	50/50	最大生产率
5	10/90	最大生产率
6	90/10	最大生产率

### 3.3.2 TEP 过程变量和故障描述

TEP 过程包括 12 个控制变量和 41 个测量变量, 测量变量包含 19 个非连续测量组分变量和 22 个连续变量, 其中 19 个非连续测量成份变量由 3 个成份分析仪获得[86]。表 3-2 为 22 个连续测量值描述表, 用 XMEAS 进行标示定义。表 3-3 中, XMEAS(23)~XMEAS(41)为 19 个成分测量值。12 个控制变量用 XMV 来定义, 取值范围为 0~100, 初始值为基础值, 如表 3-4。本文实验采用的训练集和测试集中的数据包含了 41 个测量变量和 11 个控制变量。除了反应器中搅拌器的搅拌速率 XMV(12)外, 共有过程变量 52 个。假定在某特定时刻的观测向量

可表示为： $X = [XMEAS(1) XMEAS(2) \dots XMEAS(41) XMV(1) \dots XMV(11)]$ 。

TEP 仿真模型一共包括 21 个预设故障（如表 3-5 所示），其中 16 个是已知的，5 个是未知的[88]。故障类型主要分为五种：故障 1~7 为阶跃型故障，指的是由过程变量的阶跃变化的故障，如冷却水的入口温度或者进料成分的变化；故障 8~12 为随机变量型，这是与一些过程变量的可变性增大有关；故障 13 是反应动力学中的缓慢漂移；故障 14、15 和 21 是与粘滞阀有关的；故障 16~20 是未知的。

表 3-2 连续测量变量

变量名	变量符号	基础值	单位
A 物料流量	XMEAS1	0.25052	kscmh
D 物料流量	XMEAS2	3664.0	Kgh <sup>-1</sup>
E 物料流量	XMEAS3	4509.3	Kgh <sup>-1</sup>
A、C 混合物料流量	XMEAS4	9.3477	kscmh
回收流量	XMEAS5	26.902	kscmh
反应器进料率	XMEAS6	42.339	kscmh
反应器压力	XMEAS7	2705.0	kPa gauge
反应器液位	XMEAS8	75.00	%
反应器温度	XMEAS9	120.40	°C
放空率	XMEAS10	0.3372	kscmh
产品分离器温度	XMEAS11	80.109	°C
产品分离器液位	XMEAS12	50.000	%
产品分离器压力	XMEAS13	2633.7	kPa gauge
产品分离器出口流	XMEAS14	25.160	m <sup>3</sup> h <sup>-1</sup>
汽提塔液位	XMEAS15	50.000	%
汽提塔压力	XMEAS16	3102.2	kPa gauge
汽提塔出口流量	XMEAS17	22.949	m <sup>3</sup> h <sup>-1</sup>
汽提塔温度	XMEAS18	65.731	°C
汽提塔蒸汽流量	XMEAS19	230.31	kgh <sup>-1</sup>
压缩机工作功率	XMEAS20	341.43	kw
反应器冷却水出口	XMEAS21	94.599	°C
分离器冷却水出口	XMEAS22	77.297	°C

表 3-3 成分变量

组分	变量名	基础值	单位
A	XMEAS23	32.188	Mol%
B	XMEAS24	8.8933	Mol%
C	XMEAS25	26.383	Mol%
D	XMEAS26	6.8820	Mol%
E	XMEAS27	18.776	Mol%
F	XMEAS28	1.6567	Mol%
A	XMEAS29	32.958	Mol%
B	XMEAS30	13.823	Mol%
C	XMEAS31	23.978	Mol%
D	XMEAS32	1.2565	Mol%
E	XMEAS33	18.579	Mol%
F	XMEAS34	2.2633	Mol%
G	XMEAS35	4.8436	Mol%
H	XMEAS36	2.2986	Mol%
D	XMEAS37	0.01787	Mol%
E	XMEAS38	0.83570	Mol%
F	XMEAS39	0.09858	Mol%
G	XMEAS40	53.724	Mol%
H	XMEAS41	43.828	Mol%

表 3-4 控制变量

变量名	变量符号	基础值 (%)	最低值	最高值	单位
D 进料量	XMV1	63.053	0	5811	$\text{Kgh}^{-1}$
E 进料量	XMV2	53.980	0	8354	$\text{Kgh}^{-1}$
A 进料量	XMV3	24.644	0	1.017	kscmh
总进料量	XMV4	61.302	0	15.25	kscmh
压缩机再循环阀	XMV5	22.210	0	100	%
排放阀	XMV6	40.064	0	100	%
分离器罐液流量	XMV7	38.100	0	65.71	$\text{m}^3\text{h}^{-1}$
汽提器液体产品流量	XMV8	46.534	0	49.10	$\text{m}^3\text{h}^{-1}$
汽提器水流阀	XMV9	47.446	0	100	%
反应器冷却水流量	XMV10	41.106	0	227.1	$\text{m}^3\text{h}^{-1}$
冷凝器冷却水流量	XMV11	18.114	0	272.6	$\text{m}^3\text{h}^{-1}$
搅拌速率	XMV12	50.000	150	250	rpm

表 3-5 过程故障描述

编号	故障描述	分类
1	A/C 进料流量比变化, 组分 B 含量不变 (流 4)	阶跃
2	组分 B 含量发生变化, A/C 进料流量比不变 (流 4)	阶跃
3	物料 D 的温度发生变化 (流 2)	阶跃
4	反应器冷却水入口温度发生变化	阶跃
5	冷凝器冷却水入口温度发生变化	阶跃
6	物料 A 损失 (流 1)	阶跃
7	物料 C 压力损失 (流 4)	阶跃
8	物料 A,B,C 的组成发生变化 (流 4)	随机变量
9	物料 D 的温度发生变化 (流 2)	随机变量
10	物料 C 的温度发生变化 (流 4)	随机变量
11	反应器冷却水入口温度发生变化	随机变量
12	冷凝器冷却水入口温度发生变化	随机变量
13	反应动力学特性发生变化	慢漂移
14	反应器冷却水阀门	黏住
15	冷凝器冷却水阀门	黏住
16	未知	未知
17	未知	未知
18	未知	未知
19	未知	未知
20	未知	未知
21	流 4 的阀门固定在稳态位置	恒定位置

### 3.3.3 仿真数据描述

本文实验中采用的数据集由 MATLAB 仿真程序获得。训练集和测试集的数据包含 41 个测量变量和 11 个控制变量, 采样间隔取 3 分钟, 包括正常数据和前 20 类故障数据。训练集中正常数据和 20 类故障数据都具有 480 个样本数据; 测试集中它们的样本数据个数为 960。测试集中故障数据从第 160 个样本处引入故障。所有仿真数据可以从网站 <http://web.mit.edu/braatzgroup/links.html> 下载。正常状态下某些测量值随时间变化的情况如图 3-3 和图 3-4 所示。

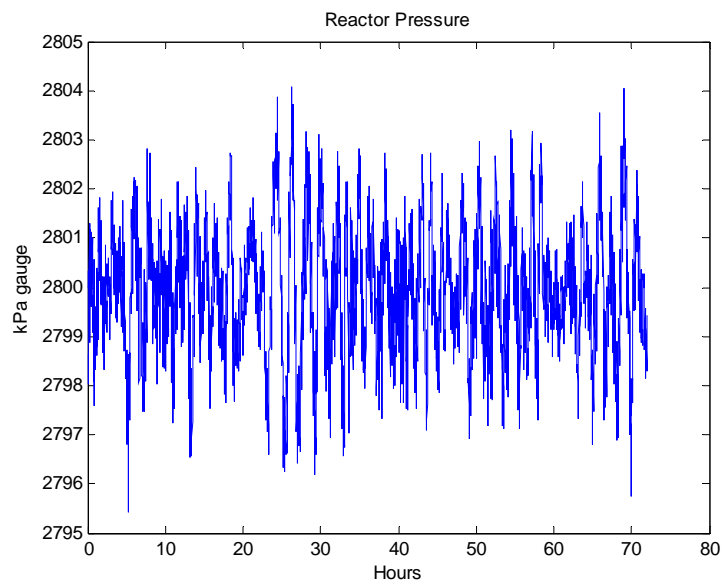


图 3-3 反应器压力

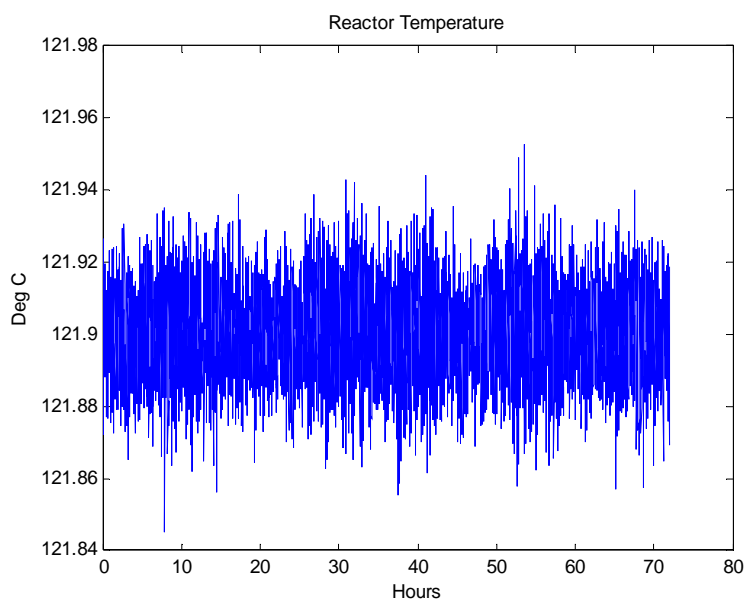


图 3-4 反应器温度

### 3.4 TEP 故障检测应用

#### 3.4.1 基于 TEP 数据的 NMFSC 算法分析

稀疏表示是一种有效的表示方法，NMF 算法本身的基于局部的表示在某种程度上也是一种稀疏表示方法，但是基本 NMF 算法的稀疏能力和程度还是比较弱的，并且难以控制。稀疏化非负矩阵分解，通过同时在  $L_1$  范数和  $L_2$  范数上实

现对稀疏性的控制，即同时在超平面和超球面上的投影，实现对矩阵的非负性和稀疏性的控制。在对原始矩阵进行分解时，为了达到特定的应用要求可以对基矩阵  $W$  或系数矩阵  $H$  进行稀疏度的控制。对于是对基矩阵还是系数矩阵进行稀疏度的控制，则根据具体需要进行选择。当采用稀疏化非负矩阵分解进行故障检测时，希望能够使基矩阵稀疏化，这样可以使基矩阵能够产生故障的特征非常突出，从而容易被检测到。因此，可以采用稀疏化非负矩阵分解的方法进行故障检测。

为了更好地说明 NMFSC 算法的特性，我们对 TEP 数据（维数为 52 维）进行稀疏化非负矩阵分解。基矩阵  $W$  的稀疏度设为 0.8，系数矩阵  $H$  的稀疏度不受限，图 3-5 是得到的基矩阵的示意图。其中， $x$  表示基的数目，根据经验设为 15； $y$  表示 52 维变量， $z$  表示基的值。从图中可以看出，它是局部表示的（仅一部分值是有意义的，而大多值为 0）。

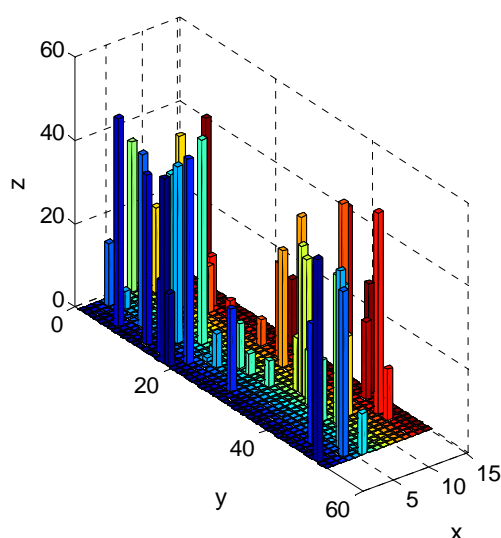


图 3-5 基矩阵的示意图

### 3.4.2 实验结果分析与讨论

在 TEP 过程的 21 种故障中，本研究以故障 1 和 4 为例，和经典的 PCA 方法进行对比以分析 NMFSC 的检测效果。实验结果如图 3-6~图 3-9 所示，图中的虚线是统计量控制限，实线表示统计量，超过控制限的统计量则表明检测到故障。

故障 1 为 A/C 进料流量比发生变化，组分 B 含量保持不变。由于反应物 A 和 C 的比例发生变化，与物质平衡有关的变量如压力、液位、成份的分配关系

也会随之改变。因此，超出一半的监控变量都会明显偏离它们正常运行的状态，从而这种故障很容易地被检测出来。图 3-6、图 3-7 分别是基于 PCA 和 NMFSC 的方法故障 1 的检测结果。在图 3-6 中，(a)统计量为 Hotelling  $T^2$ ，(b)统计量为 SPE。在图 3-7 中，(a)统计量为  $H^2$ ，(b)统计量为 SPE。在很显然，这两种方法都比较容易地检测出故障的发生。其中，前 8h 是在正常运行状况下运行的，那么，前 8h 的所有统计量应该在控制限下方，从图中可知结果正如此。

故障 4 为反应器冷却水入口温度发生变化。当这个故障发生时，其他 50 个变量仍保持稳定。故障 4 与正常工况下相比，每个变量的均值和标准差的变化小于 2%。所以，这个故障的检测比故障 1 具有更大的挑战性。图 3-8、图 3-9 分别是基于 PCA 和 NMFSC 的方法故障 4 的检测结果。在图 3-8 中，(a)统计量为 Hotelling  $T^2$ ，(b)统计量为 SPE。在图 3-9 中，(a)统计量为  $H^2$ ，(b)统计量为 SPE。从 8h（第 160 个数据）后引入故障，图 3-8(a)在控制限下的数据显然比图 3-9(a)的要多很多，也即是图 3-8(a)中有更多的数据未被检测出来。同理，图 3-8(b)也比 3-9(b)有更多的数据未被检测出来。因此，基于两种不同统计量，NMFSC 方法比 PCA 有更高的检测率。

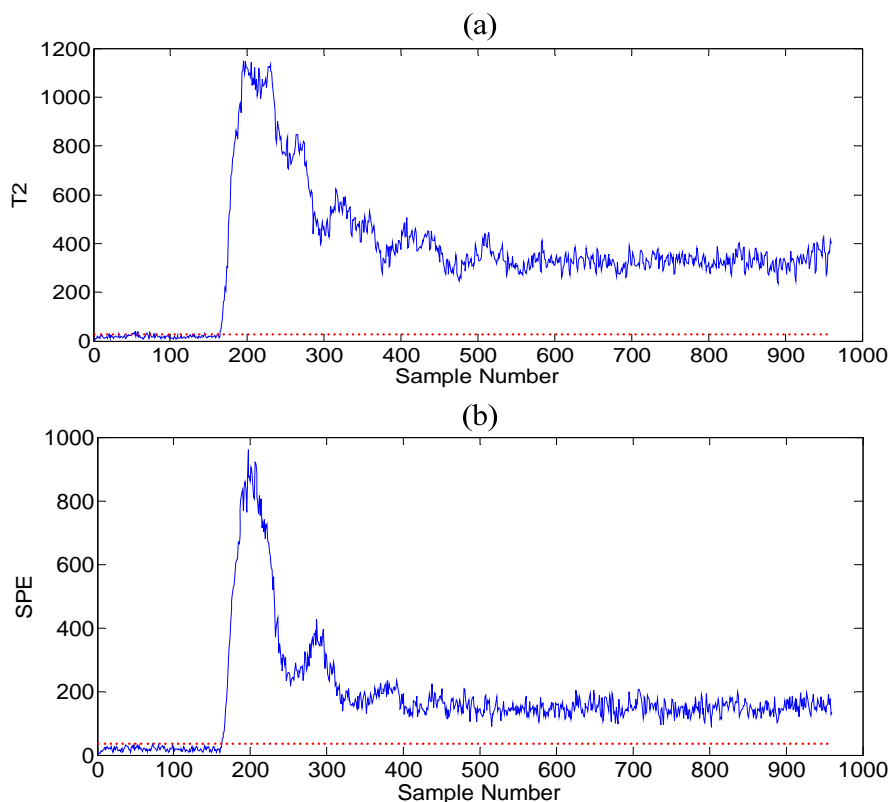


图 3-6 基于 PCA 方法故障 1 的检测结果

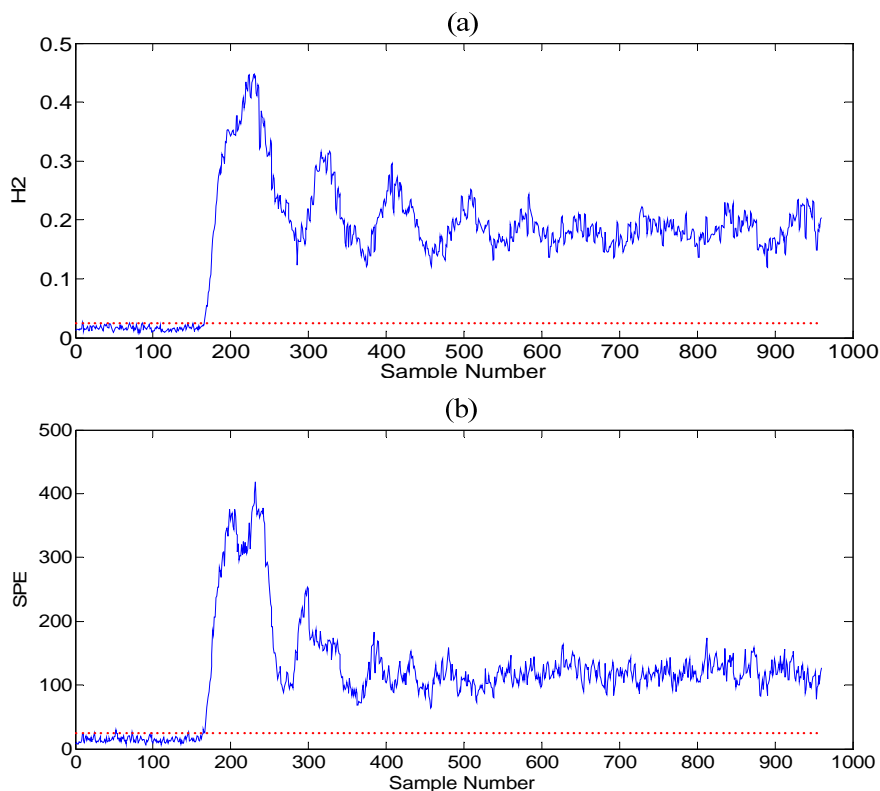


图 3-7 基于 NMFSC 方法故障 1 的检测结果

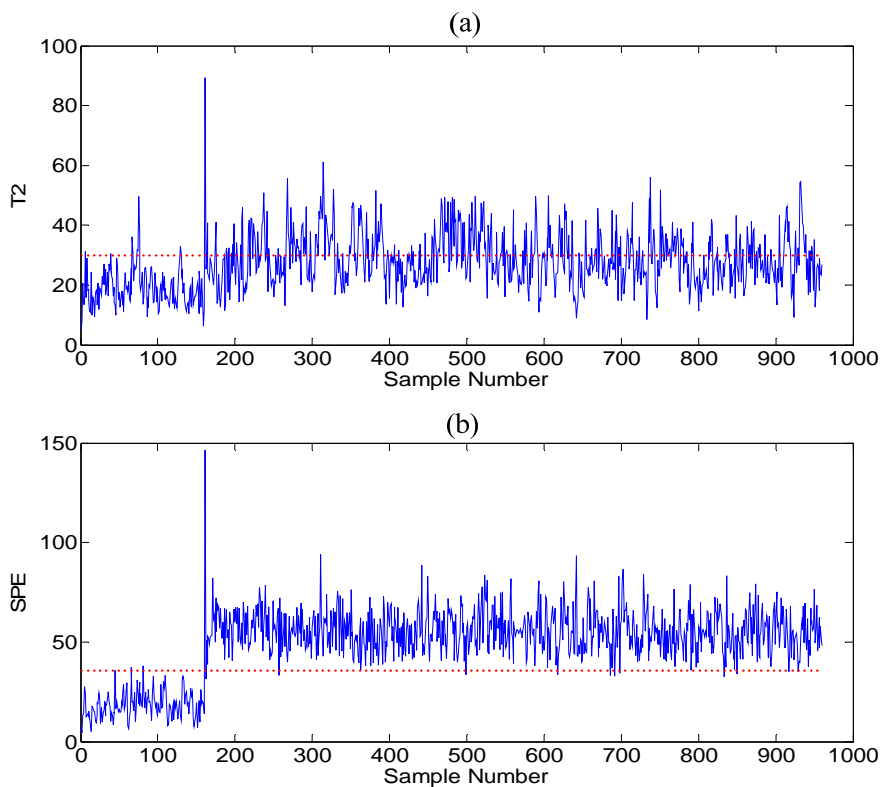


图 3-8 基于 PCA 方法故障 4 的检测结果



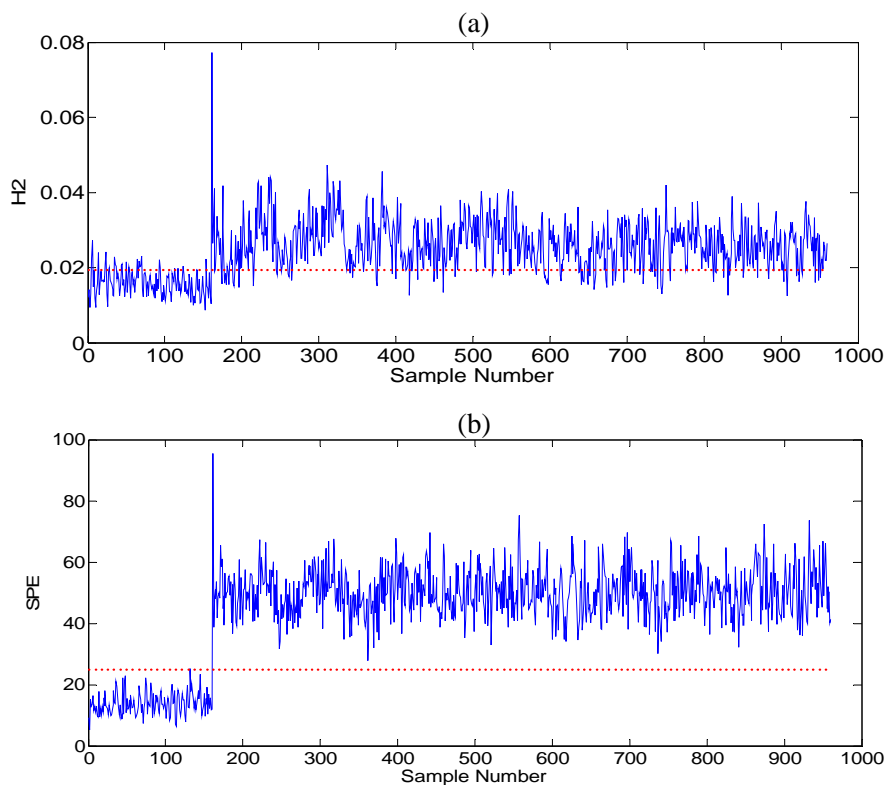


图 3-9 基于 NMFSC 方法故障 4 的检测结果

表 3-6 TEP 故障检测率

Fault	PCA $T^2$	PCA Q	NMFSC $H^2$	NMFSC Q
1	99.2	99.8	<b>99.8</b>	<b>100</b>
2	98.0	98.6	<b>98.5</b>	<b>99.3</b>
4	4.4	96.2	<b>74.2</b>	<b>100</b>
5	22.5	25.4	<b>28.6</b>	<b>52.4</b>
6	98.9	100	<b>99.4</b>	<b>100</b>
7	91.5	100	<b>96.0</b>	<b>100</b>
8	96.6	97.6	<b>97.0</b>	<b>98.4</b>
10	33.4	34.1	<b>36.8</b>	<b>50.4</b>
11	20.6	64.4	<b>35.3</b>	<b>69.9</b>
12	97.1	97.5	<b>97.3</b>	<b>98.1</b>
13	94.0	95.5	<b>91.0</b>	<b>95.9</b>
14	84.2	100	<b>70.4</b>	<b>83.8</b>
16	16.6	24.5	<b>24.1</b>	<b>45.0</b>
17	74.1	89.2	<b>65.8</b>	<b>71.4</b>
18	88.7	89.9	<b>90.1</b>	<b>93.4</b>
19	0.4	12.7	<b>9.1</b>	<b>33.0</b>
20	29.9	45.0	<b>31.2</b>	<b>49.0</b>
Average	61.7	74.7	<b>67.3</b>	<b>78.8</b>

同时，本文对于 TEP 数据集中各个故障的检测率及不同统计量的平均检测

率，和主成分分析法也进行了比较，结果如表 3-6 所示。在表中，加粗的两列为本文提出基于 NMFSC 方法对 TEP 数据集中各种故障的检测率。可以看出，NMFSC 统计量  $H^2$  的平均检测率 67.3% 比 PCA 统计量  $T^2$  平均检测率 61.7% 提高了 5.6 个百分点；NMFSC 统计量 SPE 的平均检测率 78.8% 比 PCA 统计量 SPE 平均检测率 74.7% 提高了 4.1 个百分点。TEP 数据集的实验结果表明，对于故障的检测率，本研究所设计的方法和现有经典方法相比有较为明显的提升，也就是提高了化工过程中出现异常状况的判断率。

### 3.5 本章小结

考虑到化工过程的时序性，本章提出采用指数加权滑动平均对采样数据进行处理。由于 NMFSC 算法对  $W$  和  $H$  的初始化是随机的，可能会导致每次 NMFSC 的结果不同，影响算法的稳定性，因此，本章提出采用 PCA 来对稀疏化非负矩阵分解进行初始化。基于 NMFSC 把过程数据分解成基向量，用基于部分的特征表示出来，从而有效地提取过程数据的局部特征。此外，本章详细介绍了本文算法的验证平台 TEP，包括过程的工艺流程、过程变量、生产模式、过程故障和过程仿真数据等，为后续实验部分奠定了基础。TEP 过程是一个典型的化工过程，具有现代化工过程各方面特点，是故障诊断领域内公认的实验验证平台。利用它作为我们实验的验证平台能全面地检验算法的可靠性。最后，TEP 过程的仿真结果验证了本章所提出的方法可以有效地检测故障的发生。

## 第四章 基于结构化支持向量机的故障诊断

故障被检测出来后,接下来的工作就是对故障进行识别和诊断,判断出引起故障的原因。支持向量机由于遵循了结构风险最小化原理,并可以将非线性问题转为线性问题而能得到全局最优解,避免了神经网络等方法中的过学习、欠学习、局部最优等问题,因而近年来在故障诊断领域得到广泛的应用。然而,传统的支持向量机解决的是二类分类问题,对于多类分类问题,多是通过某种方式构造一系列两类分类器并将它们组合在一起来实现多类分类的,其本质仍是二类分类问题,对于多类问题处理非常繁琐。但是,在工业过程中,由于系统的复杂性,产生故障的种类是很多的,而且对分类速度和精度要求也比较高。基于此,本文提出一种基于结构化支持向量机进行故障诊断的方法。

对于故障特征全集而言,并不是每一个特征都能够对故障诊断做出同等的贡献。因此,从特征全集中选取出能够提供较大信息量的特征来建立故障诊断模型是非常必要的。同时,测量变量并不是序列无关的,当前时刻的测量变量与过去若干时刻的变量可能都有关系,都是具有一定的动态特性的。因此,如何对故障数据的动态特性进行提取并有效地加以表示后用于模式分类,对于提高故障诊断有着重要的作用。基于此,本文提出采用决策树算法进行特征选取;引入过程变量的延时测量值,用延时测量值的均值和方差扩展样本的特征,得到的特征能动态性地表达故障,从而使特征优化的结构化向量机进行故障诊断。

### 4.1 结构化支持向量机

#### 4.1.1 结构化机器学习

与传统的机器学习方法相比,结构化机器学习能更充分地利用上下文信息改善局部信息的不足。结构化机器学习的研究是基于判别分析模型(Discriminative Model)的,只需要对输出分布作假设,因此可以利用大量冗余的特征函数来拟合分布。下面从一般的判别分析模型对结构化机器学习作简单的介绍。

监督学习是通过已有标记的训练样本进行学习,对测试样本的标记以尽可

能正确地进行预测[89]。实际上，就是学习一个从输入到输出类标记的函数。这个函数是通过研究输入  $x$  与输出  $y$  样本的内部结构而构造的，用来表示样本的特征，所以亦称为特征函数，它为一个依赖于  $x, y$  特征向量：

$$f \in F : (x, y) \rightarrow \begin{pmatrix} f_1(x, y) \\ \vdots \\ f_n(x, y) \end{pmatrix} \quad (4-1)$$

机器学习是在特征函数的基础上对样本  $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$  进行训练，从而得到预测结果。常见的线性分类器形式为：

$$\hat{y} = \arg \max_{y \in Y} \sum_{i=1}^n w_i f_i(x, y) \quad (4-2)$$

其中  $w_i$  表示权重，其值越大，则代表所对应的特征越重要； $Y$  是  $y$  所有可能的取值，对于结构化的输出， $Y$  中元素的数量可能是很大的。机器学习的训练过程，就是对于给定的训练样本集合自动估计参数  $w$  的过程。结构化机器学习中常见的训练算法有条件随机场（CRFs）[90]、结构化支持向量机（Structured Support Vector Machines, Structured SVMs）[91]、平均感知器（Averaged Perceptron, AP）[92]、最大边际马尔科夫网络（Max Margin Markov Networks, M3Ns）[93]等。

在监督机器学习问题中，对于数据样本一般采用给定类别标签的标注方式。然而，对于更复杂的、数据中包含某些“结构化”信息的学习问题，数据的标注形式往往具有复杂的结构，即类标签集合  $Y$  为表现数据中特定结构信息的形式，而非单一的标量标签。结构化学习本质上为多类问题，与传统的多类问题区别在于，可能的类标签集合  $Y$  通常非常大。对于多类问题，传统的支持向量机通常采用“一对多”和“一对一”两种策略。如果用“一对多”的多类分类方法，由于每个两类分类器的训练都是将全部的样本作为训练样本，所以训练样本数量的增加会导致需要求解的二次规划问题的次数也急剧增加，进而使分类器的训练时间急剧增加。“一对一”的多类分类方法，虽然训练时间较“一对多”快，但需要构造的两类分类器数目较多，当类别较多时，训练速度较低。此外，这两种方法都存在不可分区域，带来无法判别类别的情况。然而，采用结构化机器学习方法能够解决这些问题。

### 4.1.2 结构化支持向量机算法

在实际应用中,大部分要处理的数据都比较复杂,而且彼此之间存在相互依赖,具有特定的结构关系,如果采用传统的支持向量机很难去处理。根据数据内部的结构性,结构化支持向量机[94-97]提出采用结构化特征函数对传统的支持向量机进行改进,能有效地处理结构化数据。

结构化数据分析问题的目的是要找到样本的输入与输出之间对应的函数。假定函数  $f$  形式为:

$$f_w(x) = \arg \max_{y \in Y} [w \cdot \Psi(x, y)] = \arg \max_{y \in Y} \langle w, \Psi(x, y) \rangle \quad (4-3)$$

其中,特征函数  $\Psi(x, y)$  是结合输入与输出数据特性,描述输入输出之间特性联合的一个向量,它的形式取决于解决的具体问题; $w$  是特征函数的权重,其值越大,则代表所对应的特征越重要。下面从四个方面详细介绍结构化支持向量机。

#### (1) 损失函数和经验风险最小化

在传统的分类中所采用的具有代表性的0-1损失函数已经不适应于各种结构化输出数据。例如,在自然语言处理中,与正确的分析结果基本相同或仅有一两个节点不同的分析树,应该和完全不同的分析树有所区别。

为了量化预测的正确性,需要考虑损失函数  $\Delta: Y \times Y \rightarrow R$  [96]。这里  $\Delta(y, \hat{y})$  指预测输出结果  $\hat{y}$  与其正确输出结果  $y$  之间的损失。当预测准确无误时,损失值为零;当预测有误差时,损失值不为零。(  $\Delta(y, y) = 0; \Delta(y, y') > 0, y \neq y'$  )。

对于监督学习,输入输出对  $(x, y)$  与确定的某个分布  $P(x, y)$  相关,所以监督学习的目的是找到决策函数  $f$  使期望风险

$$R_p^\Delta(f) = \int_{X \times Y} \Delta(y, f(x)) dP(x, y) \quad (4-4)$$

最小化。然而,由于  $P$  是未知的,所以期望风险是无法计算的。假定有限个点组成的训练集根据已知的概率  $P$  生成独立同分布,在训练集  $S$  上的函数  $f$  的表现被经验风险描述为:

$$R_S^\Delta(f) = \frac{1}{n} \sum_{i=1}^n \Delta(y_i, f(x_i)) \quad (4-5)$$

## (2) 最优化问题的导出

假设存在函数  $f$  使经验风险为 0, 那么 0 风险的训练误差的条件可以写为如下的非线性约束:

$$\forall i \in \{1, \dots, n\}: \max_{y \in Y \setminus y_i} \{\langle w, \Psi(x_i, y) \rangle\} \leq \langle w, \Psi(x_i, y_i) \rangle \quad (4-6)$$

这里损失函数是独立的, 由于假定了  $\Delta(y, y) = 0, \Delta(y, y') > 0$  for  $y \neq y'$ 。

非线性不等式(4-6)可以用  $n$  个线性约束表示为:

$$\forall i \in \{1, \dots, n\}, \forall y \in Y \setminus y_i: \langle w, \Psi(x_i, y_i) - \Psi(x_i, y) \rangle \geq 0 \quad (4-7)$$

定义  $\delta\Psi_i(y) \equiv \Psi(x_i, y_i) - \Psi(x_i, y)$ , 式(4-7)可表示为  $\langle w, \delta\Psi_i(y) \rangle \geq 0$ 。可知, 满足式(4-7)的  $w$  不至于一个。为了解决  $w$  唯一性, 提出使最小间隔  $r$  最大化, 根据此原则选择  $w$ 。最小间隔  $r$  是指正确分类  $y$  与最容易错分的  $\hat{y}$  的间隔。其中,  $\hat{y}$  为  $\arg \max_{y \neq y_i} \langle w, \Psi(x_i, y) \rangle$ 。

通过  $L_2$  范数约束  $w$ , 问题可以转化为下面的优化问题

$$\begin{aligned} \max_{r, w: \|w\|=1} r \\ \text{s.t. } \forall i \in \{1, \dots, n\}, \forall y \in Y \setminus y_i: \langle w, \delta\Psi_i(y) \rangle \geq r \end{aligned} \quad (4-8)$$

其中,  $\|w\|=1$  是对分离超平面的法向量  $w$  进行约束, 使其规范化, 进而使间隔是确定的。

式(4-8)由式(4-9)转化的,

$$\begin{aligned} \max_{r, w} r \\ \text{s.t. } \forall i \in \{1, \dots, n\}, \forall y \in Y \setminus y_i: \left\langle \frac{w}{\|w\|}, \delta\Psi_i(y) \right\rangle \geq r \end{aligned} \quad (4-9)$$

其中,  $\|w\|$  为  $w$  的  $L_2$  范数。此优化问题希望最大化超平面关于训练数据集的间隔  $r$ , 约束条件表示的是超平面关于每个训练样本点的间隔至少为  $r$ 。

式(4-9)可以改写为

$$\begin{aligned} \max_{r, w} \frac{r}{\|w\|} \\ \text{s.t. } \forall i \in \{1, \dots, n\}, \forall y \in Y \setminus y_i: \langle w, \delta\Psi_i(y) \rangle \geq r \end{aligned} \quad (4-10)$$

间隔  $r$  的取值并不影响上述最优化问题的解。这样, 可以取  $r = 1$ , 将其代入

上面的最优化问题（最大化  $\frac{1}{\|w\|}$  和最小化  $\frac{1}{2}\|w\|^2$  是等价的），此问题可以等价地表示为一个标准的凸二次规划问题

$$\begin{aligned} \min_w \quad & \frac{1}{2}\|w\|^2 \\ \text{s.t.} \quad & \forall i, \forall y \in Y \setminus y_i : \langle w, \delta\Psi_i(y) \rangle \geq 1 \end{aligned} \quad (4-11)$$

为了允许不满足约束条件  $\langle w, \delta\Psi_i(y) \rangle \geq 1$  的训练样本存在，通过引入松弛变量  $\xi_i$ ，“软化”约束条件  $\langle w, \delta\Psi_i(y) \rangle \geq 1 - \xi_i$  [98-99]。当  $\xi_i$  很大时，训练样本总可以满足此约束条件。但是，应该设法避免  $\xi_i$  取太大的值。所以在目标函数里对它们进行惩罚，则软间隔最优化问题为：

$$\begin{aligned} \text{SVM} : \quad & \min_{w, \xi} \frac{1}{2}\|w\|^2 + \frac{C}{n} \sum_{i=1}^n \xi_i, \\ \text{s.t.} \quad & \forall i, \xi_i \geq 0, \forall y \in Y \setminus y_i : \langle w, \delta_i\Psi(y) \rangle \geq 1 - \xi_i \end{aligned} \quad (4-12)$$

其中， $C$  为惩罚参数，用来控制错分样本点对目标函数的影响。

### (3) 约束条件调整

由于实际问题中  $|Y|$  往往很大，以上形式不适合解决结构化问题，所以在约束条件中引入损失函数。引入损失函数的方法主要有两种，一是对每个约束条件，由损失函数调整其松弛变量（Slack-Rescaling）：

$$\begin{aligned} \text{SVM}^{\Delta_s} : \quad & \min_{w, \xi} \frac{1}{2}\|w\|^2 + \frac{C}{n} \sum_{i=1}^n \xi_i, \\ \text{s.t.} \quad & \forall i, \xi_i \geq 0, \forall y \in Y \setminus y_i : \langle w, \delta_i\Psi(y) \rangle \geq 1 - \frac{\xi_i}{\Delta(y_i, y)} \end{aligned} \quad (4-13)$$

另一种方法是 Taskar 提出的用损失函数调整间隔（Margin-Rescaling）[100]，间隔约束条件被调整为：

$$\begin{aligned} \text{SVM}^{\Delta_m} : \quad & \min_{w, \xi} \frac{1}{2}\|w\|^2 + \frac{C}{n} \sum_{i=1}^n \xi_i, \\ \text{s.t.} \quad & \forall i, \xi_i \geq 0, \forall y \in Y \setminus y_i : \langle w, \delta_i\Psi(y) \rangle \geq \Delta(y_i, y) - \xi_i \end{aligned} \quad (4-14)$$

### (4) 切割平面算法

对于大规模的实际问题，训练 Structural SVMs 是个挑战性问题。同时，解决二次规划(Quadratic Programming, QP)问题的最大困难是约束条件规模很大。

为了解决这些问题，提出切割平面算法（Cutting-plane algorithm），用损失函数总和的单个切割平面模型代替损失函数（每个训练样本对应一个）的  $n$  个切割平面模型。可以节约大量的训练时间。那么，式(4-13) (4-14)分别转化为：

$$\begin{aligned}
 SVM_1^{\Delta s} : \quad & \min_{w, \xi} \frac{1}{2} \|w\|^2 + C\xi \\
 s.t. \quad & \forall (\bar{y}_1, \dots, \bar{y}_n) \in Y^n : \frac{1}{n} w^T \sum_{i=1}^n \Delta(y_i, \bar{y}_i) [\psi(x_i, y_i) - \psi(x_i, \bar{y}_i)] \geq \frac{1}{n} \sum_{i=1}^n \Delta(y_i, \bar{y}_i) - \xi
 \end{aligned} \tag{4-15}$$

$$\begin{aligned}
 SVM_1^{\Delta m} : \quad & \min_{w, \xi} \frac{1}{2} \|w\|^2 + C\xi \\
 s.t. \quad & \forall (\bar{y}_1, \dots, \bar{y}_n) \in Y^n : \frac{1}{n} w^T \sum_{i=1}^n [\psi(x_i, y_i) - \psi(x_i, \bar{y}_i)] \geq \frac{1}{n} \sum_{i=1}^n \Delta(y_i, \bar{y}_i) - \xi
 \end{aligned} \tag{4-16}$$

对于公式(4-13)、(4-14)的优化问题对于  $n$  个训练样本需要不同的松弛变量，即有  $n$  个松弛变量，这被称为“ $n$ -松弛”问题。割平面方法用一个松弛变量，称为“1-松弛”问题。

#### 4.2 基于结构化支持向量机的故障诊断

结构化支持向量机可以通过输入与输出的内部结构构造特征函数。由于特征函数可以设计为适合于多类问题的函数，因此，多类故障的诊断就可以通过统一的模型进行判别。基于结构化支持向量机进行故障诊断的框架如图 4-1 所示。

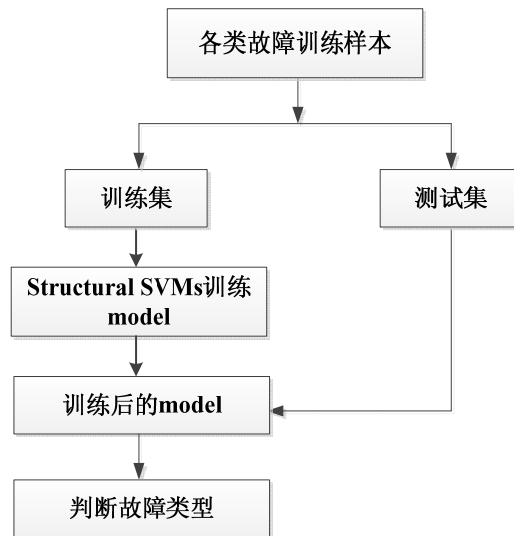


图 4-1 基于 Structural SVMs 的故障诊断框架



### 4.2.1 结构化特征函数构造

结构化支持向量机可以处理各种不同的复杂数据，针对不同的应用领域，可以根据所处理的输入和输出的内部结构，构造出相应的结构化特征函数。特征函数选择是否适当，是分类器是否成功的重要因素。针对化工工业多类故障的诊断问题，借鉴文献[94]的思想，本文构造特征函数  $\Psi(x, y)$  的定义如下：

$$\Psi(x, y) \equiv \Phi(x) \otimes \Lambda(y) \quad (4-17)$$

$$\Lambda(y) \equiv (\delta(y_1, y), \delta(y_2, y), \dots, \delta(y_K, y))' \in \{0, 1\}^K \quad (4-18)$$

其中， $\otimes$  表示张量积； $\Phi(x)$  表示输入数据，也可以由核函数定义； $\Lambda(y)$  是单位向量，描述标号  $y$ 。 $\otimes$  的定义如下：

$$\otimes: R^D \times R^K \rightarrow R^{D \cdot K}, \quad (a \otimes b)_{i+(j-1)D} \equiv a_i \cdot b_j \quad (4-19)$$

$\delta$  的定义为：

$$\delta(y_k, y) = \begin{cases} 1 & y_k = y \\ 0 & y_k \neq y \end{cases} \quad (4-20)$$

例如，对于 3 类问题，那么  $K=3$ 。当输出  $y=1$  时， $\Lambda(1) = (100)'$ 。

### 4.2.2 基于 Structural SVMs 的故障诊断算法

综上所述，基于 Structural SVMs 的故障诊断算法步骤如下。在该算法中，首先初始化工作集  $S_i$  为空集，然后通过反复迭代在训练数据集中找最违反约束条件的  $\hat{y}$ ，来不断更新调整工作集  $S_i$ ，然后求解最优化问题。

#### 基于 Structural SVMs 的故障诊断算法

- 1: 输入训练样本  $(x_1, y_1), \dots, (x_n, y_n)$ ，设置参数  $C, \xi$ 。
- 2: 初始化工作集  $S_i$  为空集， $i=1, \dots, n$ 。
- 3: 选定损失函数，即为 0-1 损失或者  $\Delta(y_i, y)$
- 4: 计算  $H(y)$ ， $SVM_1^m: H(y) \equiv \Delta(y_i, y) - \langle \delta \psi_i(y), w \rangle$ ，

$$\text{其中 } w \equiv \sum_j \sum_{y' \in S_j} a_{(jy')} \delta \psi_j(y')$$

- 5: 计算  $\hat{y} = \arg \max_{y \in Y} H(y)$

- 
- 6: 计算  $\xi_i = \max\{0, \max_{y \in S_i} H(y)\}$ , 如果  $H(\hat{y}) > \xi_i + \varepsilon$ , 那么
- $S_i \leftarrow S \cup \{\hat{y}\}$ ,  $S = \cup_i S_i$  在  $S$  上由  $SVM_1^m$  的对偶问题进行二次优化更新  $a_S$ , 并继续返回到第 4 步进行迭代优化, 直至  $S_i$  在迭代中不再变化。然后转到第 7 步。
- 7: 输入测试样本集根据训练得到的分类器进行分类测试并输出结果。
- 

该算法采用分解思想, 将原问题转化为对偶问题后[95], 其求解方法与传统的支持向量机类似。对于大规模的实际分类问题, 通常只有不多的支持向量, 而最优化问题的解与支持向量对应的那些训练点有关, 所以只需求解规模较小的最优化问题。

#### 4.2.3 基于 Structural SVMs 的方法在 TEP 故障诊断中的应用

本研究引入结构化支持向量机多类判别方法, 用于对多类故障之间的诊断。该方法有效地简化了分类器模型, 只使用一个单一模型就能进行多类故障的诊断。

##### 4.2.3.1 TEP 数据描述

为了研究 Structural SVMs 在故障诊断中的应用, 验证其分类效果, 选取田纳西过程中的故障 2、4、9、10、11、13 和 14。这些故障不但包括了 TEP 仿真数据中所有的故障类型, 而且很好地呈现了故障数据间的交错性。为了验证本文提出的算法, 我们设置了两组实验, 其中组合 1 包括故障 4、9 和 11, 组合 2 包括故障 2、10、13 和 14。每一组故障分别都产生两组数据, 训练样本数据用来建立模型, 测试样本数据用来验证模型。其中, 训练样本数据包含 480 组数据, 测试样本数据包含 800 组数据。

实验 1 包括故障 4、9 和 11, 分别命名为类别 1、2 和 3。具体描述如表 4-1。故障 4 和故障 11 都是与反应器冷却水入口温度有关, 具有相同的故障变量; 只是故障类型不同, 故障 4 是阶跃变化, 故障 11 是随机变化。当这两种故障发生时, 偏离正常值的变量之间有重叠区域, 具有很强的不可分性。故障 9 是物料 D 的温度的随机变化, 不同于故障 4 和故障 11。这三种故障呈现出数据间是相互重叠的,

能有效地验证所提出的分类方法的有效性。在每组故障数据中包括 52 个过程变量。其中，变量 9 为反应器温度，变量 51 为反应器冷却水流量，两者对于区分故障 4、9 和 11 相对比较重要。因为故障 4 和故障 11 都只与反应器有关，而故障 9 与进料 D 的温度变化有关。其他 50 个变量对于这三种故障的区分没有明显的作用。由此可知，故障 4 和故障 9 可以容易地根据这两个变量区分，但这两个变量在故障 11 发生时与故障 4 和 9 有很大程度的重叠，因此对于分类存在重叠区的数据会存在一定错分率。图 4-2 列出了该组训练数据集关于变量 9 和 51 的二变量散点图。从图中可以看出，故障 4 和故障 9 的训练数据是相互分离的，但是这两种故障都和故障 11 重叠在一起。

实验 2 包括故障 2、10、13 和 14，分别命名为类别 1、2、3 和 4。具体描述如表 4-1。它们涉及了故障的所有类型，分别为阶跃、随机变量、慢漂移和黏住。其不同的故障类型对于 TEP 的所有故障具有很好的代表性。另外，故障 10 为物料 C 的温度发生变化（流 4），故障 2 为组分 B 含量发生变化，A/C 进料流量比不变（流 4），两者都是流 4 且故障 10 影响故障 2。故障 13 和 14 都和反应器有关，而反应器影响流 4。因此这四种故障区分有一定的难度，可以有效地验证算法。

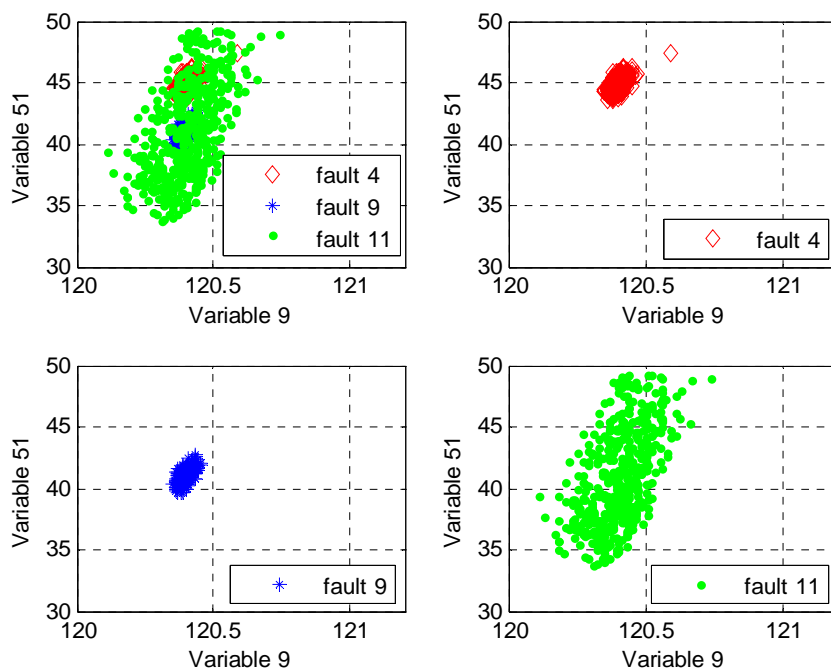


图 4-2 故障 4、9 和 11 在变量 9 和 51 的散点图

表 4-1 二组实验描述

实验 1					
类别	模式	故障描述	类型	训练数据	测试数据
1	故障 4	反应器冷却水入口温度发生变化	阶跃	480	800
2	故障 9	物料 D 的温度发生变化 (流 2)	随机变量	480	800
3	故障 11	反应器冷却水入口温度发生变化	随机变量	480	800
实验 2					
1	故障 2	组分 B 含量发生变化, A/C 进料流量比不变 (流 4)	阶跃	480	800
2	故障 10	物料 C 的温度发生变化 (流 2)	随机变量	480	800
3	故障 13	反应动力学特性发生变化	慢漂移	480	800
4	故障 14	反应器冷却水阀门	黏住	480	800

#### 4.2.3.2 实验结果分析

传统的支持向量机只考虑了二类分类这一简单的问题,组合多个两类分类支持向量机来解决多类分类问题。我们提出的基于结构化支持向量机的方法,只使用一个单一模型就能进行多类分类,大大提高了分类效率。

故障 4、9 和 11 分别命名为类别 1、2 和 3, 分别有 480 组训练样本数据训练分类器, 然后由 800 组测试样本数据验证构建的分类器。表 4-2 为故障 4、9 和 11 采用本文提出的方法与其他类型的 SVM 方法的错分率比较结果。其中, 第一列是每一组训练或测试数据由所有的变量 (52 个) 组成的情况, 第二列是每一组训练或测试数据中仅由变量 9 和 51 组成的情况。很显然, 相对于普通 SVM、偏最小二乘 PSVM [101]以及排列不变性 Invariant SVM [102], Structural SVMs 的错分率已有很大程度的降低。

故障 2、10、13 和 14 分别命名为类别 1、2、3 和 4, 分别有 480 组训练样本数据训练分类器, 然后由 800 组测试样本数据验证构建的分类器。表 4-3 为故障 2、10、13 和 14 采用 Structural SVMs 方法, 选用不同核函数的实验结果。选用多项式核函数 (polynomial kernel)、径向基核函数 (RBF kernel) 和线性核 (linear kernel) 所对应的错分率分别是 8.22%、8.47% 和 37.34%。可以看出, 多项式核函数有较好的表现, 在训练中, 我们设参数  $d$  为 3, 惩罚系数  $C$  为 540。下面基于此核函数构建的分类器的分类情况做进一步的分析。表 4-4 列出了错分率为 8.22% 的多项式核函数对应的混淆矩阵, 800 个故障 2 的测试样本中, 784 个样本

被正确分为故障 2，16 个测试样本被错分为故障 10；故障 10 的测试数据完全被正确分类；故障 13 中，555 个样本被正确分类，241 个测试样本被错分为故障 10，较少的被错分为故障 2 和 14；故障 14 中，798 个样本被正确分类，仅有 2 个测试样本被错分为故障 10。由于故障 10 与故障 2、13、14 有着紧密的联系，所以故障 2、13、14 被错分为故障 10 的较多。

表 4-2 故障 4、9 和 11 的错分率比较

Method \ features	All	{9,51}
SVM	44%	6.5%
PSVM	35%	6.0%
Invariant SVM	29.86%	6.0%
<b>Structural SVMs</b>	<b>23.82%</b>	<b>6.0%</b>

表 4-3 故障 2、10、13 和 14 的错分率

Classifier	Classification error
Structural SVMs (polynomial kernel)	<b>8.22 %</b>
Structural SVMs (RBF kernel)	8.47 %
Structural SVMs (linear kernel)	37.34 %

表 4-4 故障 2、10、13 和 14 的实验混淆矩阵

class	Fault 2	Fault 10	Fault 13	Fault 14
Fault 2	<b>784</b>	0	1	0
Fault 10	16	<b>800</b>	241	2
Fault 13	0	0	<b>555</b>	0
Fault 14	0	0	3	<b>798</b>
Total	800	800	800	800

通过两组实验结果可以看出，我们所提出的方法不仅解决了化工过程中多类故障需要训练大量分类器的问题，而且相比于其他类型的 SVM 方法取得了更好的结果。

### 4.3 基于特征优化的结构化支持向量机的故障诊断

在机器学习分类算法中，分类器的产生受到很多因素的影响。其中，样本数

据的质量及表现形式是很重要的。理论上说, 样本数据的特征属性越多, 产生的分类器越有效。然而, 实际应用中, 很多算法往往由于一些无关特征而对训练集中的某些类别的评估产生偏差。所以, 在分类器构建之前, 对训练样本进行特征选取是十分必要的。基于此, 本文提出采用决策树从特征全集中选取出能够提供较大信息量的特征来建立故障诊断模型。我们进一步考虑到过程数据样本间的时序相关性, 引入过程变量的延时测量值。利用延时测量值的均值和方差扩展样本的特征, 组合新的数据矩阵来动态地表达故障。这样, 在原有特征空间上进行特征选取, 并加入动态特性信息, 利用结构化支持向量机进行多类判别来诊断故障, 我们将该方法定义为特征优化的结构化支持向量机 (Feature-Optimized Structural SVMs, FO Structural SVMs)。

#### 4.3.1 基于决策树的特征选取

每个故障特征对故障诊断的贡献可能是不同的, 一些特征能够对故障分类提供一些有用信息, 但同时也带来了一定的噪声; 还有一些特征对故障分类不能提供或仅能提供一点有用信息, 但却引入较大的噪声, 增加了故障诊断的错分率。因此, 为了降低故障诊断错误率, 应该选取能够提供较大信息量的特征, 摒弃带来噪声的特征来建立故障诊断模型。

决策树的过程, 其实就是找出分类能力最好的属性, 把数据分成多个子集, 每个子集再用分类能力最好的属性进行划分。那些在每个子集里被评价为分类能力最好的属性, 如果用来分类能够提供较大的信息量。所以, 本文采用决策树进行特征选取。

##### 4.3.1.1 决策树算法

决策树算法主要有: ID3 算法、C4.5 算法和 CART 算法。其中 ID3 算法是决策树算法的代表, 它采用分治策略, 在决策树各级结点上选择属性时, 用信息增益作为属性的选择标准, 以便在每一个非叶子结点上进行测试时, 能获得关于被测试记录最大的类别信息。具体方法是: 检测所有的属性, 选择信息增益最大的属性产生决策树结点, 由该属性的不同取值建立分支, 再对各分支的子集递归调用该方法建立决策树结点的分支, 直到所有子集仅包含同一类别的数据为止, 最后得到一棵决策树[103]。

给定一个类标记的样本训练集  $S$ ， $|S|$  表示其样本个数。假设类标号具有  $m$  个不同值，定义  $m$  个不同类  $C_i$  ( $i=1, \dots, m$ )。设  $|C_i|$  是类  $C_i$  中的样本数，对  $S$  中样本分类所需的期望信息为：

$$Info(S) = -\sum_{i=1}^m p_i \log_2(p_i) \quad (4-21)$$

其中， $p_i = |C_i|/|S|$  是  $S$  中任意样本属于类  $C_i$  的概率。使用以 2 为底的对数函数，因为信息用二进制编码。 $Info(S)$  是识别  $S$  中样本的类标号所需要的平均信息量。

设属性  $A$  具有  $v$  个不同值  $\{a_1, a_2, \dots, a_v\}$ ，可以用属性  $A$  将  $S$  划分为  $v$  个子集  $\{S_1, S_2, \dots, S_v\}$ ， $|S_j|$  为  $S_j$  的样本数，其中  $S_j$  中的样本在属性  $A$  上具有相同的值  $a_j$ 。那么由  $A$  划分子集的信息期望为：

$$Info_A(S) = \sum_{j=1}^v \frac{|S_j|}{|S|} \times Info(S_j) \quad (4-22)$$

其中， $|S_j|/|S|$  是第  $j$  个划分的权重。 $Info_A(S)$  是基于按  $A$  划分对  $S$  的样本分类所需要的期望信息。 $Info(S_j)$  表示子集  $S_j$  的信息期望。在属性  $A$  上获得的信息增益为：

$$Gain(A) = Info(S) - Info_A(S) \quad (4-23)$$

#### ID3 算法[104]

输入：训练数据集 $S$ ，特征集 $A$ ，阈值 $\varepsilon$		输出：决策树
1:	若 $S$ 中所有样本属于同一类 $C_k$ ，则 $T$ 为单结点树，并将类 $C_k$ 该结点的类标记，返回 $T$ ；	
2:	若 $A = \emptyset$ ，则 $T$ 为单结点树，并将 $S$ 中样本数最大的类 $C_k$ 作为该结点的类标记，返回 $T$ ；	
3:	否则，计算 $A$ 中各个特征对 $S$ 的信息增益，选择信息增益最大的特性 $A_g$ ；	
4:	如果 $A_g$ 的信息增益小于阈值 $\varepsilon$ ，则置 $T$ 为单结点树，并将 $S$ 本数最大的类 $C_k$ 作为该结点的类标记，返回 $T$ ；	
5:	否则，对 $A_g$ 的每一可能值 $a_i$ ，依 $A_g = a_i$ 将 $S$ 分割为若干非空子集 $S_i$ ，将 $S_i$ 中样本数最大的类作为标记，构建子结点，由结点及其子结点构成树 $T$ ，返回 $T$ 。	
6:	对第 $i$ 个子结点，以 $S_i$ 为训练集，以 $A - \{A_g\}$ 为特征集，递归地调用步 1~5，得到子树 $T_i$ ，返回 $T_i$ 。	

### 4.3.1.2 特征选取算法

在 TEP 过程中, 共有 20 种类型的典型故障。本研究从中选择 3 种典型故障: 故障 4、9 和 11。因为这 3 类故障, 其过程观测数据发生空间重叠, 不易被分开。这三种故障体现了很好的交错性, 能够验证算法的有效性。我们以这三种故障为例, 分析如何根据决策树选取特征的。

对故障 4、9 和 11 这三种故障采用决策树进行训练, 得到如图 4-3 的决策树。可以看出, 浅节点处的特征 51 和 9 具有很高的信息增益, 能够对故障分类提供较多的有用信息。这正好和前面分析的特征 51 和 9 对判别此三种故障有至关重要的作用相吻合。还有一些特征如 29、36 和 42 等也具有较高的信息增益。另外一些深节点处的特征如 3 和 26 等则具有较低的信息增益, 对分类不能或仅能提供很少的有用信息, 但却引入较大的噪声。因此, 我们可根据信息增益的高低选取特征。例如, 特征 51 和 9 具有很高的信息增益, 我们选取作为特征点; 而特征 3 和 26 等则具有较低的信息增益, 我们舍弃这些特征。最后, 根据所得到的决策树, 选取的特征点为 {51, 9, 29, 36, 42, 28, 21, 38, 5, 46, 8, 17, 37, 6, 19, 13, 41, 32, 4, 24, 34, 31, 33, 1, 39, 22, 3, 26}。

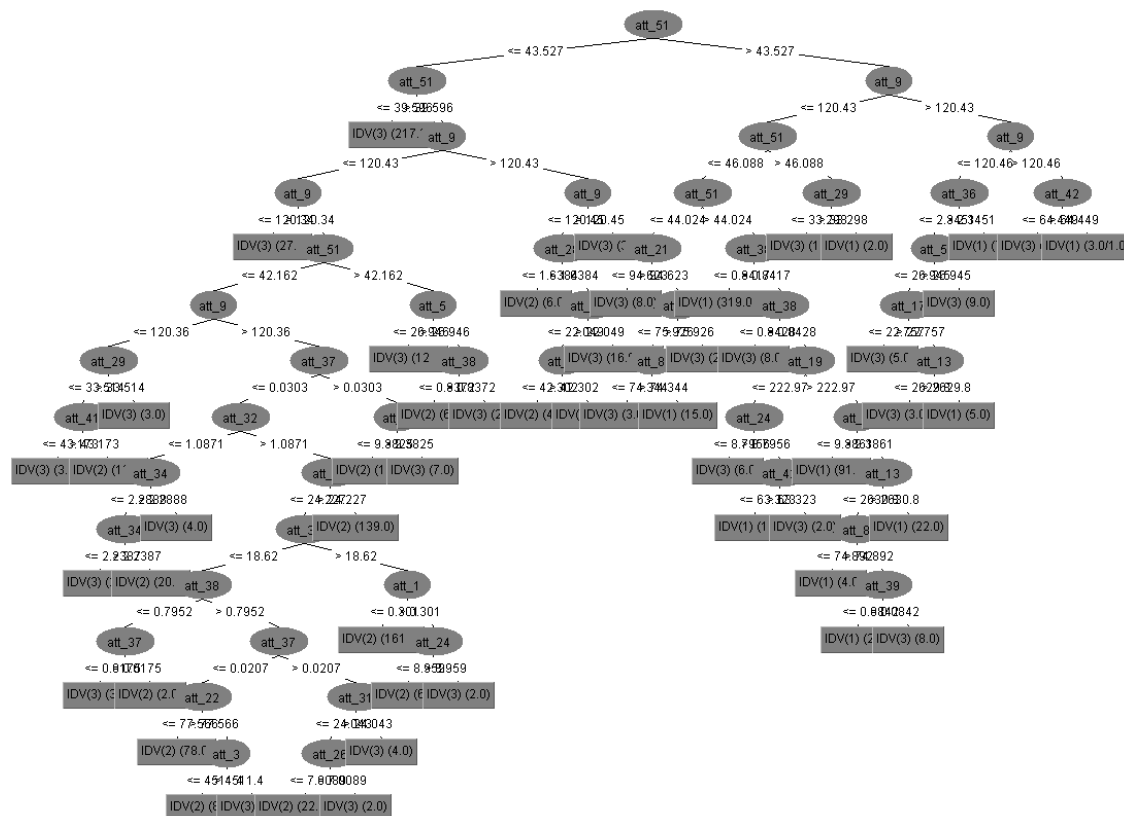


图 4-3 决策树示意图



在决策树构建的过程中，采用信息增益确定决策树的每个节点。每个节点即特征属性，它的信息增益的高低能够很好地表达分类能力的高低。

#### 4.3.2 基于动态特性的故障表示方法

现代企业中，高度自动化的流程生产系统通常由多个相互连接的设备组成，各设备之间的生产流程及相互影响关系复杂。为保证设备的无故障运行，需要利用各种传感器对设备的各种参数信息进行实时监控，由此积累了大量带有时间标记的检测数据。如何从这些时间序列数据库中挖掘出故障所特有的动态特性信息，对于复杂系统的诊断与生产预警都是非常必要的。

另外，在模式识别领域内，有效的特征表达对于分类效果起着至关重要的作用。在当前流行的故障诊断方法中，数据特征空间基本都是由相关传感器的测量值组成的，对于每维数据在时间轴上都是假设独立的，然而这种假设在大多数的工业过程中都是不现实的，测量变量并不是序列无关的，当前时刻的测量变量与过去若干时刻的变量都有关系，都是具有一定的动态特性的。并且不同的故障产生时，同一变量所具有的动态特性也是不一样的。因此，如何对故障数据的动态特性进行提取并有效地加以表示出来用于模式分类，是提高故障诊断效果的有效途径之一[105-107]。

在故障诊断领域对于动态特性的研究一直都在进行。早在 1995 年，Ku 等人将动态行为与 PCA 算法相结合，提出了 DPCA 算法[108]。Li 和 Qin 在此基础上提出了一种基于 EIV-SMI(Error-In-Variables Subspace Model Identification)的一致 DPCA 方法[109]。2003 年 Lee 等人结合 Sensor Validity Index(SVI)和 DPCA 提出了一种用于故障检测的 SVI[110]。前人在故障特性方面的研究表明数据动态特性的提取对于改善故障诊断效果起到促进作用。因此如何建立一种动态特性的特征表达方式是很重要的。基于此，本文引入过程变量的延时测量值，延时测量值的均值和方差扩展样本的特征，组合成新的数据矩阵，来表达故障的动态性：

$$X = \begin{bmatrix} x_k^d & M_k^d & V_k^d \\ x_{k+1}^d & M_{k+1}^d & V_{k+1}^d \\ \vdots & \vdots & \\ x_{k+n}^d & M_{k+n}^d & V_{k+n}^d \end{bmatrix} \quad (4-24)$$

$$M_k^d = \frac{1}{l}(x_k^d + x_{k-1}^d + \dots + x_{k-l}^d) \quad (4-25)$$

$$V_k^d = \frac{1}{l-1}[(x_k^d - M_k^d)^2 + (x_{k-1}^d - M_k^d)^2 + \dots + (x_{k-l}^d - M_k^d)^2]$$

其中,  $x_{k-l}^d$  表示比当前时刻  $k$  滞后  $l$  个时刻的数据,  $l$  为时滞常数, 本实验经验地设为 10;  $M_k^d$  表示从  $k-l$  时刻到  $k$  时刻的样本均值,  $V_k^d$  表示其方差。  $X$  表示原数据  $x_{k+i}^d$  和均值  $M_{k+i}^d$ 、方差  $V_{k+i}^d$  组合成新的特征数据。  $d$  表示特征(属性)个数。

提取方差和均值的意义在于体现测量值在故障发生时所表现出的波动幅度信息和平均取值信息。这样做的好处是: 一方面方差信息很好地表示出某一维度数据在时间轴上的波动信息; 另一方面均值信息结合方差信息较好地反映测量值波动大致范围。因 TEP 中的故障 4、9 和 11 之间较强的不可分性, 具有代表性。所以, 仍以故障 4、9 和 11 为例进行分析。变量 9 是反应器温度, 变量 51 是反应器冷却水流量。故障 4 和 11 都只与反应器有关, 而故障 9 与 D 的温度有关并且会影响反应器, 所以, 变量 9 和 51 对区分故障 4、9 和 11 起很大的作用。因此, 我们基于这两个变量来分析加入动态性对区分故障 4、9 和 11 的影响。

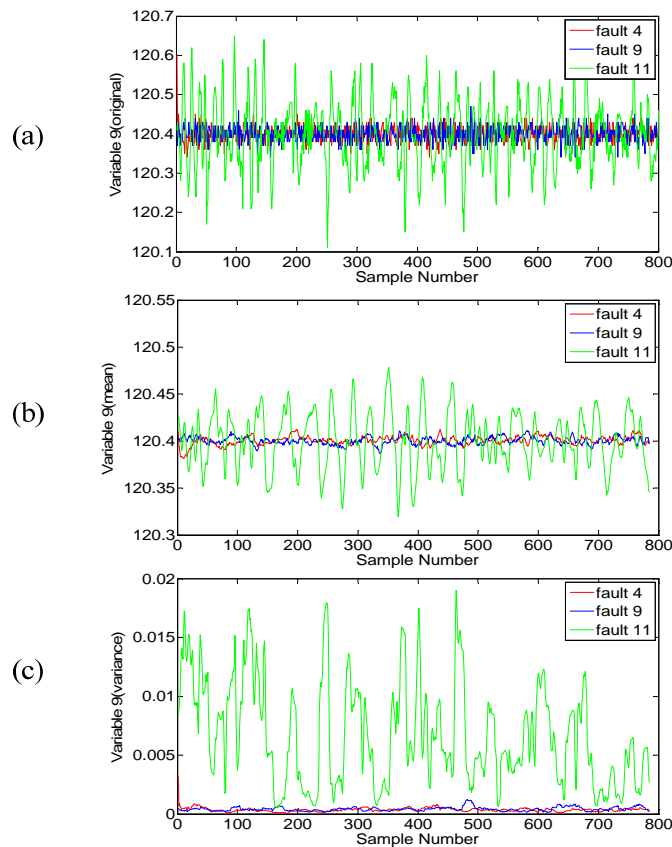


图 4-4 故障 4、9 和 11 在变量 9 上的示意图

图 4-4 和 4-5 分别是变量 9 和 51 在故障 4、9 和 11 上的示意图，(a)(b)(c)依次为原数据、均值和方差，其中红色为故障 4，蓝色为故障 9，绿色为故障 11。在此过程中，我们经验地设时间窗口大小为 10。从图 4-4-a 中可以清楚地看出故障 4、9 和 11 的原数据在变量 9 上是不可分的，而在图 4-4-c 中故障 4 和 9 的方差与故障 11 是可分的。在图 4-5 中，故障 4、9 和 11 在变量 51 上的均值和方差在某种程度上均可分。很显然，这两个统计特征结合之后能够很好地区别故障类型。因此，将此方法提取的动态特性作为新的特征扩展到原始特征空间，为故障诊断提供了更多的可判别信息，可以提高分类器的效果。

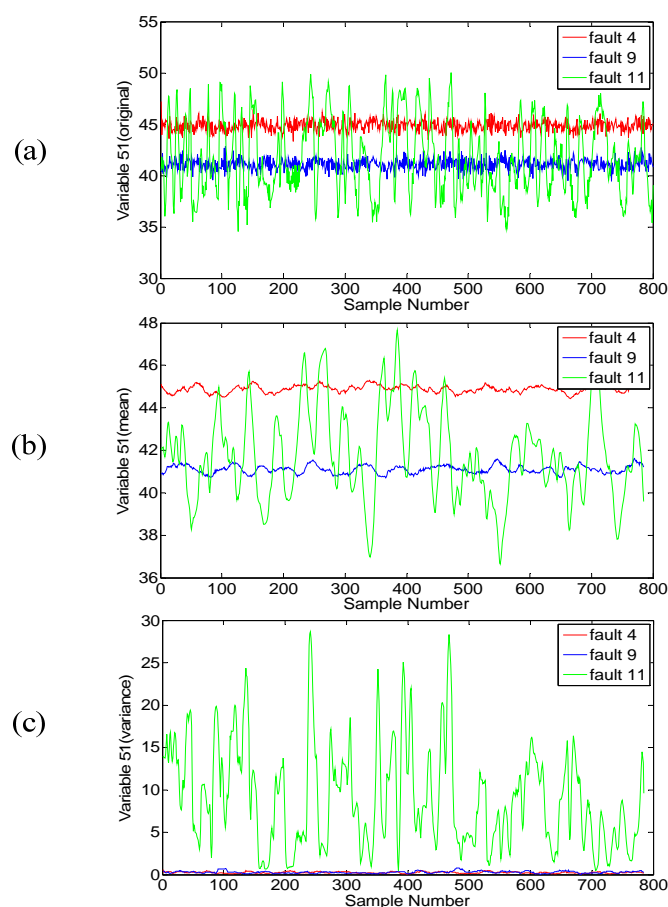


图 4-5 故障 4、9 和 11 在变量 51 上的示意图

### 4.3.3 基于 FO Structural SVMs 的方法在 TEP 故障诊断中的应用

基于特征优化的结构化支持向量机进行故障诊断：首先，采用决策树从特征全集中选取能够提供较大信息量的特征；然后，利用过程变量的延时测量值的均值和方差来扩展样本的特征，进而产生新的数据矩阵，形成新的样本；最后，将

新的样本作为结构化支持向量机训练模型的输入,故障测试样本即可根据所得模型进行故障诊断。下文将此方法应用于标准化工过程 TEP 的故障诊断。

#### 4.3.3.1 基于特征优化的故障诊断方法实验

研究中,从 TEP 数据集中选择了具有较强不可分性的故障 4、9 和 11 进行了实验。为了验证采用决策树提取特征以及引入动态特性作为特征补充的诊断效果,我们对故障 4、9 和 11 训练集和测试集提取特征并加入动态特性信息,然后根据训练集构建的模型,对测试集进行预测。此结果与上一章未进行这些操作的实验进行了对比。表 4-5 为 FO Structural SVMs 方法和其他方法的比较结果。从表 4-5 可以看出,我们提出的采用决策树提取特征并基于时序性扩展这些特征的方法,比 Structural SVMs 基于原始特征全集的方法故障诊断率有了明显的提升。图 4-6 为我们所提出方法的实验效果的混淆矩阵示意图,对应错分率为 20.25%。(a)(b)(c)分别是故障 4(红色)、9(蓝色)和 11(黑色)被正确分类以及其他故障被错分该类的情况。可以看出,在(a)中故障 4 大多被正确分类,故障 9 和 11 被错分为故障 4 的样本很少;在(b)中,故障 11 被错分为 9 的样本显然比故障 4 要多很多;在(c)中,故障 4 和 9 被错分为故障 11 的样本都很多。这正好验证了前面对这三类故障的分析:故障 4 和 9 有一定的可分性,两者都与故障 11 有较强的不可分性。

本研究所提出的方法不仅解决了化工过程中多类故障需要训练大量分类器的问题,而且相比于未特征优化的 Structural SVMs 以及其他类型的 SVM 方法(普通 SVM, 偏最小二乘 PSVM [101], 排列不变性 Invariant SVM[102])均取得了更好的结果。因此,从特征全集中选取出能够提供较大信息量的特征,然后引入动态特性信息,为故障的判别提供了更可靠的信息,从而提高了故障诊断的能力。

表 4-5 故障 4、9 和 11 的错分率比较

Method	Misclassification rate
SVM	44%
PSVM	35%
Invariant SVM	29.86%
<b>Structural SVMs</b>	<b>23.82%</b>
<b>FO Structural SVMs</b>	<b>20.25%</b>

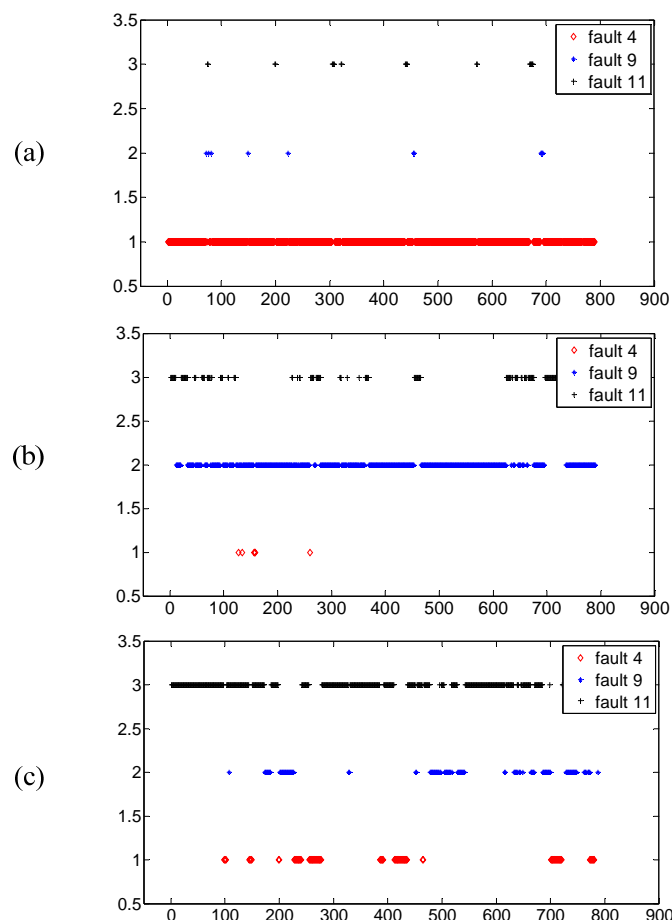


图 4-6 混淆矩阵示意图

#### 4.3.3.2 不同特征集实验的讨论

本节研究继续以 TEP 过程的故障 4、9 和 11 为例。在前面采用决策树选取特征得到特征的次序，即信息增益从高到低的次序，为{51, 9, 29, 36, 42, 28, 21, 38, 5, 46, 8, 17, 37, 6, 19, 13, 41, 32, 4, 24, 34, 31, 33, 1, 39, 22, 3, 26}。然而，将所有这些特征构成的特征集用于构建故障诊断的分类器，或许不一定是故障分类的最优特征集。因为，一些特征对故障分类提供有用信息的同时，也会带来了一定的噪声。

根据特征的选择次序，我们构造含有不同特征数目的特征子集：{51}，{51, 9}，{51, 9, 29}，{51, 9, 29, 36} ... {51, 9, 29, 36, 42, 28, 21, 38, 5, 46, 8, 17, 37, 6, 19, 13, 41, 32, 4, 24, 34, 31, 33, 1, 39, 22, 3, 26}，共 28 组。对这些特征子集分别用前面的方法，对故障数据进行动态表示，即用延时测量值的均值和方差扩展原特征矩阵。然后，在这些特征子集上，分别训练相应的分类器，建立相应

的故障诊断模型。图 4-7 为 TEP 过程的测试数据集对于不同特征数目的特征子集所得到的 Structural SVMs 故障诊断模型的错分率，横轴为特征数目，纵轴为在各个不同特征子集构建的分类器上测试数据集的错分率。

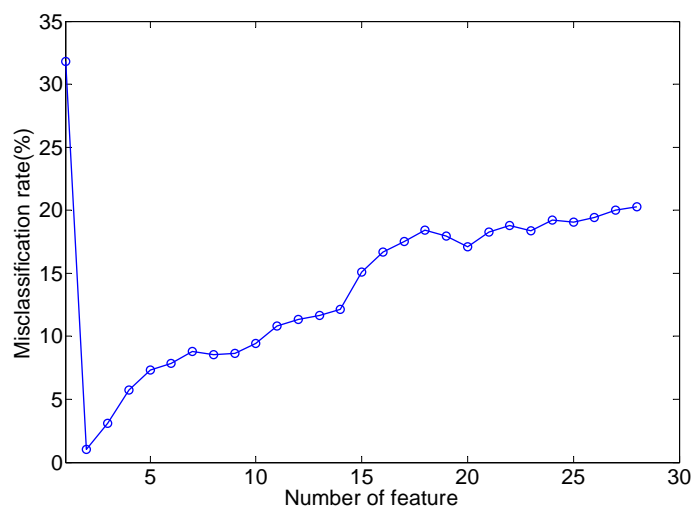


图 4-7 不同特征子集的 Structural SVMs 故障诊断模型的错分率

从图中可以看出，特征子集{51, 9}用于故障诊断时，测试数据集的错分率仅为 1.01%，并且所构建的故障诊断模型也比较简单。然而，当仅使用特征{51}进行故障诊断时，错分率达到 31.81%，这说明特征的选取并不是越少越好。采用决策树选取的全部特征建立的故障诊断模型在测试数据上的错分率最高，为 20.25%。所以，特征的选取也不是越多越好。当选取特征{51, 9, 29, 36, 42, 28, 21, 38, 5, 46, 8, 17, 37, 6}时，所用的训练测试时间不比特征集{51, 9}多很多，但比决策树选取的所有 28 个特征要少很多。另外，它在测试数据上的错分率为 12.11%，相比决策树选取的所有特征的错分率大幅下降。

#### 4.4 本章小结

本章提出基于结构化支持向量机进行故障诊断的方法，以结构信息为基础建立空间度量，拓展并列特征，对于多个故障用统一的模型进行判别故障的类型，突破了传统的支持向量仅能区分两类故障或者要构建多个模型判别多类故障的情况，有效地简化了分类器模型，节约了训练时间，而且降低了错分率。此外，本章还提出基于特征优化的结构化支持向量机进行故障诊断的方法。对于故障特征全集而言，并不是每一个特征都能够对故障诊断做出同等的贡献。选取对于分

类能够提供较大信息量的特征来建立故障诊断模型,可以摒弃对故障分类不能提供较多有用信息,但却引入较大噪声的特征。针对于此,提出采用决策树提取特征。另外,针对过程变量序列相关性,提出引入过程变量的延时测量值,延时测量值的均值和方差扩展原始特征,组合成新的特征矩阵。新的特征矩阵能够对故障数据的动态特性进行有效地表示。最后,在标准化工过程 TEP 数据集上证明了本章所提出方法的有效性。





## 第五章 基于深度学习的故障诊断

在上一章提出的基于结构化支持向量机进行故障诊断的方法中,故障样本必须有类标号,其分类问题是对给定有类标号的数据训练分类器。然而,在实际中要想找到这种有利于分类的比较合理的表示方法的样本数据是比较困难的。深度学习通过建立类似人脑的分层模型结构,对输入数据逐级提取从底层到高层的特征,从而能很好地建立底层信号到高层语义的映射关系,非监督、多层次地进行特征学习。另外,当前大多数分类、回归等机器学习方法为浅层结构算法,其局限性在于由于非线性特征表示能力有限,针对复杂分类问题其表示能力受到一定制约;由简单的单层架构转换原始输入特征为特定问题特征空间时,其过程是不可观察的。深度学习可通过学习一种深层非线性网络结构,实现复杂函数逼近,表征输入数据分布式表示,从而学习到更有用的特征表征输入数据,即从样本集中学习数据集本质特征,进而最终提升分类或预测的准确性[111]。

### 5.1 深度学习算法

#### 5.1.1 引言

深度学习(Deep Learning, DL)的概念是由 Geoffrey Hinton 等在 2006 年人提出[112],主要通过神经网络来模仿人的大脑学习过程,通过组合较低层的特征形成更为抽象的高层表示,从而发现观测数据的分布式特征。通过特征自下而上的学习形成多层的抽象表示,并且多层次的特征学习是一个自动地无人工干预的过程。

深度学习的基本思想如下。假设有一个系统  $S$ , 它有  $n$  层 ( $S_1, \dots, S_n$ ), 输入是  $I$ , 输出是  $O$ , 形象地表示为:  $I \Rightarrow S_1 \Rightarrow S_2 \Rightarrow \dots \Rightarrow S_n \Rightarrow O$ , 如果输出  $O$  等于输入  $I$ , 即输入  $I$  经过这个系统变化之后保持不变, 没有任何的信息损失, 这意味着输入  $I$  经过每一层  $S_i$  都可以当做原有信息的另一种表示。深度学习的精髓也就是, 在一个  $n$  层的神经网络中, 任何一层的输入  $I$  和输出  $O$  是“相等”的。学习过程中, 希望能够自动地学习对象的特征。给定一组原始的输入  $I$ , 经过一个包含  $n$  层的系统  $S$ , 通过调整系统中的参数, 使得输出  $O$  仍然是输入  $I$ , 这样就

获取了输入  $I$  的一系列的层次特征，记为  $S_1, \dots, S_n$ 。

深度学习的常用方法主要有受限玻尔兹曼机（Restricted Boltzmann Machine, RBM）和自动编码器（AutoEncoder）两种方法。本文运用“深度学习”中的稀疏自动编码器[113]自动选取样本特征，然后结合 Softmax 回归分类器对训练集进行分类。

### 5.1.2 稀疏自动编码器

给定一个神经网络，假设有一个无标签的训练样本集  $\{x^{(1)}, x^{(2)}, x^{(3)}, \dots\}$ ，其中  $x^{(i)} \in R^n$ ，并使目标值等于输入值，如  $y^{(i)} = x^{(i)}$ 。然后使用反向传播算法训练调整其参数，得到每一层中的权重，这样就得到输入的几种不同表示，这些表示就是输入数据的特征。如果在原有的特征中加入这些自动学习得到的特征可以大大提高精确度。这种方法称为自动编码器（AutoEncoder, AE）。图 5-1 是一个自编码神经网络，带 3 个隐藏单元。如果隐藏层的节点数目小于输入输出节点的数目，输入数据中隐含一些特定结构（例如某些输入特征是彼此相关的），那么这一算法就可以发现输入数据中的这些相关性。这就可以“自动地”得到输入数据降维后的特征，即隐藏节点。如果隐藏层的节点数目较大（可能大于输入节点数），可以对隐藏节点加入稀疏性限制，即约束每一层中的节点大部分都为 0，只有少数不为 0，这就得到稀疏自动编码器（Sparse AutoEncoder, SAE）方法[114-115]。

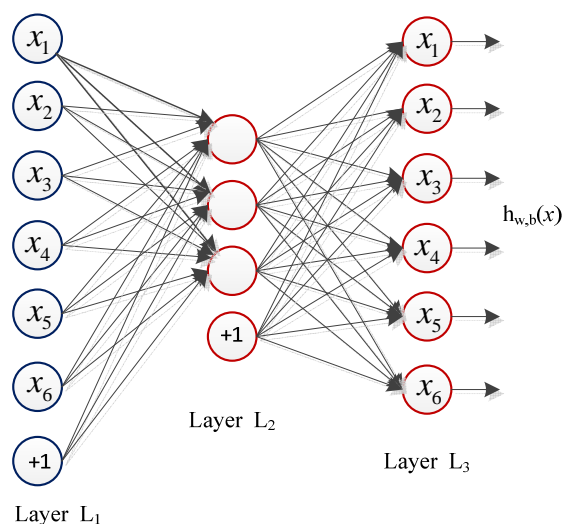


图 5-1 自编码神经网络

为求稀疏自编码的参数，首先根据网络的结构确定损失函数及损失函数偏导函数的表达式，然后利用优化算法获取网络最优的参数。无稀疏约束时，给定  $m$

个样例的数据集，损失函数定义为：

$$\begin{aligned} J(W, b) &= \left[ \frac{1}{m} \sum_{i=1}^m J(W, b; x^{(i)}, y^{(i)}) \right] + \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_l+1} (W_{ji}^{(l)})^2 \\ &= \left[ \frac{1}{m} \sum_{i=1}^m \left( \frac{1}{2} \| h_{w,b}(x^{(i)}) - y^{(i)} \|^2 \right) \right] + \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_l+1} (W_{ji}^{(l)})^2 \end{aligned} \quad (5-1)$$

其中， $x$ 、 $y$  分别表示输入和输出， $W$  表示权重， $b$  表示偏置项， $\lambda$  为权重衰减参数。公式中的第一项是一个均方差项，第二项是一个规则化项（也叫权重衰减项），其目的是减少权重的幅度，防止过度拟合。对式中  $h_{w,b}$  的定义如下

$$\begin{aligned} h_{w,b}(x) &= A^{(3)} = f(W^{(2)}A^{(2)} + B^{(2)}) \\ A^{(2)} &= f(W^{(1)}X + B^{(1)}) \end{aligned} \quad (5-2)$$

式中， $f(z) = \frac{1}{1+e^{-z}}$ 。

添加稀疏编码约束条件后，隐含节点输出的平均值应尽量为 0，那么稀疏自动编码器的损失函数表示为：

$$J_{sparse}(W, b) = J(W, b) + \beta \sum_{j=1}^{s_2} KL(\rho \| \hat{\rho}_j) \quad (5-3)$$

其中， $\beta$  控制稀疏性惩罚因子的权重， $\rho$  是稀疏性参数；后一项为 KL 距离，即是以  $\rho$  为均值和以  $\hat{\rho}_j$  为均值的两个伯努利随机变量之间的相对熵，其表达式为：

$$KL(\rho \| \hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1-\rho) \log \frac{1-\rho}{1-\hat{\rho}_j} \quad (5-4)$$

当  $\hat{\rho}_j = \rho$  时， $KL(\rho \| \hat{\rho}_j) = 0$ ，并随着  $\hat{\rho}_j$  和  $\rho$  差异的增大而单调递增。隐含节点输出平均值  $\hat{\rho}_j$  的表达式为：

$$\hat{\rho}_j = \frac{1}{m} \sum_{i=1}^m [a_j^{(2)}(x^{(i)})] \quad (5-5)$$

其中， $a_j^{(2)}(x)$  表示给定输入  $x$ ，自编码神经网络隐藏神经元  $j$  的激活度。此表达式说明，如果隐含层的每个节点输出平均值接近 0.05，则达到稀疏的目的。

求出损失函数后，便可以求其偏导数。假定输出层为  $n_l$ ，那么在输出层每个节点的误差表达式为：

$$\delta_i^{(n_l)} = -(y_i - a_i^{(n_l)}) \cdot f'(z_i^{(n_l)}) \quad (5-6)$$

其中,  $a_i^{(n_l)}$  表示第  $n_l$  层第  $i$  单元的激活值,  $z_i^{(l)}$  表示第  $n_l$  层第  $i$  单元输入加权和,  $f(\cdot)$  为激活函数。

在隐含层  $l=n_l-1, n_l-2, n_l-3, \dots, 2$ , 每个节点的误差表达式为:

$$\delta_i^{(l)} = \left( \sum_{j=1}^{s_{l+1}} W_{ji}^{(l)} \delta_j^{(l+1)} \right) + \beta \left( -\frac{\rho}{\hat{\rho}_i} + \frac{1-\rho}{1-\hat{\rho}_i} \right) f'(z_i^{(l)}) \quad (5-7)$$

其中,  $\beta$  控制稀疏性惩罚因子的权重,  $\hat{\rho}_i$  是隐藏神经元  $i$  的平均激活度。

最终计算得到的偏导数为:

$$\frac{\partial}{\partial W_{ij}^{(l)}} J_{sparse}(W, b) = a_j^{(l)} \delta_i^{(l+1)} \quad (5-8)$$

$$\frac{\partial}{\partial b_{ij}^{(l)}} J_{sparse}(W, b) = \delta_i^{(l+1)} \quad (5-9)$$

得到损失函数及其偏导数后, 可以采用梯度下降法求网络最优化的参数。

### 5.1.3 Softmax 回归分类器

Softmax 回归模型[116]是 Logistic 回归模型在多分类问题上的推广, 在多分类问题中, 类标签  $y$  可以取两个以上的值。当为二分类时, Softmax 回归模型就退化为 Logistic 回归模型。

在 Softmax 回归中, 类标签  $y$  可以取  $k$  个不同的值 ( $k$  也可以等于 2)。因此, 训练集  $\{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$  便有  $y^{(i)} \in \{1, 2, \dots, k\}$ 。对于给定的测试样本  $x$ , 假设函数针对每一个类别  $j$  估算出概率值  $p(y=j|x)$ , 即估计  $x$  的每一种分类结果出现的概率。那么, 假设函数将输出一个  $k$  维的向量 (向量元素和为 1) 来表示  $k$  个估计的概率值, 假设函数  $h_\theta(x)$  形式如下:

$$h_\theta(x^{(i)}) = \begin{bmatrix} p(y^{(i)} = 1 | x^{(i)}; \theta) \\ p(y^{(i)} = 2 | x^{(i)}; \theta) \\ \vdots \\ p(y^{(i)} = k | x^{(i)}; \theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^k e^{\theta_j^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \vdots \\ e^{\theta_k^T x^{(i)}} \end{bmatrix} \quad (5-10)$$

其中  $\theta_1, \theta_2, \dots, \theta_k \in R^{n+1}$  是模型的参数。为了方便起见, 将  $\theta$  用一个  $k \times (n+1)$  的矩阵来表示, 将  $\theta_1, \theta_2, \dots, \theta_k$  按行罗列起来, 即

$$\theta = \begin{bmatrix} \theta_1^T \\ \theta_2^T \\ \vdots \\ \theta_k^T \end{bmatrix} \quad (5-11)$$

### 5.1.3.1 代价函数

Softmax 回归算法的代价函数定义为

$$J(\theta) = -\frac{1}{m} \left[ \sum_{i=1}^m \sum_{j=1}^k 1\{y^{(i)} = j\} \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^k e^{\theta_l^T x^{(i)}}} \right] \quad (5-12)$$

其中， $1\{\cdot\}$  是示性函数，其取值规则是： $1\{\text{值为真的表达式}\}=1$ ， $1\{\text{值为假的表达式}\}=0$ 。

在 Softmax 回归中将  $x$  分类为类别  $j$  的概率是：

$$p(y^{(i)} = j | x^{(i)}, \theta) = \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^k e^{\theta_l^T x^{(i)}}} \quad (5-13)$$

对于  $J(\theta)$  的最小化问题，借助于梯度下降法或 L-BFGS 算法等迭代优化算法。

经过求导，得到梯度公式如下：

$$\nabla_{\theta_j} J(\theta) = -\frac{1}{m} \sum_{i=1}^m [x^{(i)} (1\{y^{(i)} = j\} - p(y^{(i)} = j | x^{(i)}; \theta))] \quad (5-14)$$

其中， $\nabla_{\theta_j} J(\theta)$  是一个向量，它的第  $l$  个元素  $\frac{\partial J(\theta)}{\partial \theta_{jl}}$  是  $J(\theta)$  对  $\theta_j$  的第  $l$  个分量的偏

导数。根据偏导数公式，可以将它代入梯度下降法等算法中，使  $J(\theta)$  最小化。

### 5.1.3.2 权重衰减

为了能惩罚参数值过大的情况，通过添加一个权重衰减项  $\frac{\lambda}{2} \sum_{i=1}^k \sum_{j=0}^n \theta_{ij}^2$  来修改代价函数，代价函数变为：

$$J(\theta) = -\frac{1}{m} \left[ \sum_{i=1}^m \sum_{j=1}^k 1\{y^{(i)} = j\} \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^k e^{\theta_l^T x^{(i)}}} \right] + \frac{\lambda}{2} \sum_{i=1}^k \sum_{j=0}^n \theta_{ij}^2 \quad (\lambda > 0) \quad (5-15)$$

权重衰减项使代价函数变成了严格的凸函数，这样可以保证得到唯一解。此时，Hessian 矩阵变为可逆矩阵，由于  $J(\theta)$  是凸函数，使用梯度下降和 L-BFGS 等算

法可以保证收敛到全局最优解。

为了使用优化算法，需要求得新函数  $J(\theta)$  的导数公式如下：

$$\nabla_{\theta_j} J(\theta) = -\frac{1}{m} \sum_{i=1}^m [x^{(i)} (1\{y^{(i)} = j\} - p(y^{(i)} = j | x^{(i)}; \theta))] + \lambda \theta_j \quad (5-16)$$

通过对参数  $\theta$  进行迭代，最小化  $J(\theta)$  函数，实现 Softmax 回归模型。

## 5.2 基于深度学习的故障诊断方法

目前，用于化工工业的监督学习算法往往需要获取很多已标记样本，造成很高的成本。如果算法能够从未标记数据中自动地进行特征学习，将得到的特征加入到原有的特征中，将可以大大提高精确度，使算法取得更好的效果。深度学习可以对无标记输入数据非监督地逐级提取从底层到高层的特征，自动学习到更有用的特征表征输入数据。基于此，本文提出一种基于深度学习进行故障诊断的方法。

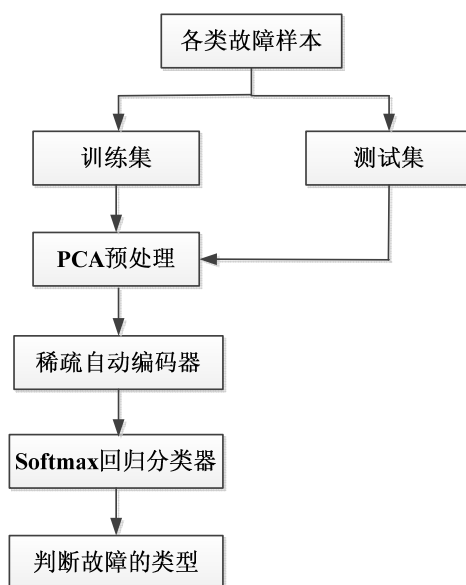


图 5-2 基于深度学习的故障诊断框架

基于深度学习的故障诊断算法的基本思想是：在给出化工过程的无标记样本数据后，用稀疏自动编码器对数据逐层进行无监督学习，然后通过监督学习进行微调，最后诊断出故障的类型。基于深度学习的故障诊断方法主要包含三个模块：数据预处理模块、深度学习模块和识别模块。数据预处理模块主要是对输入的化

工数据进行 PCA 白化处理。深度学习模块是通过稀疏自编码器算法进行非监督学习，学习出能较好表征原数据的特征，然后使用 Softmax 回归算法构建分类模型，通过有监督的微调使全局网络参数优化。识别模块是新的样本数据根据构建的深度网络诊断出故障的类型。基于深度学习进行故障诊断的框架如图 5-2 所示。

基于深度学习的故障诊断算法如下。在该算法中，对无标签故障数据训练两个级联的稀疏自编码器，通过这两个稀疏自编码器，学习出较好的特征表示。基于第二个稀疏自编码器学习得到的特征表示作为 Softmax 分类器的输入。然后，对全局网络进行有监督的微调，使全局网络参数最优。最后，对新的故障样本通过训练得到的深度网络进行分类。

#### 基于深度学习的故障诊断算法

- 1: 对输入的无标签故障样本集  $\{x^{(1)}, x^{(2)}, \dots\}$  进行 PCA 预处理得到  $\{X^{(1)}, X^{(2)}, \dots\}$
- 2: 以  $\{X^{(1)}, X^{(2)}, \dots\}$  作为输入特征训练一个 AutoEncoder1
- 3: 将样本集  $\{X^{(1)}, X^{(2)}, \dots\}$  通过 AutoEncoder1 进行非线性变换，得到新的特征表示集  $\{S^{(1)}, S^{(2)}, \dots\}$
- 4: 以  $\{S^{(1)}, S^{(2)}, \dots\}$  作为输入特征训练一个 AutoEncoder2
- 5: 将  $\{S^{(1)}, S^{(2)}, \dots\}$  通过 AutoEncoder2 进行非线性变换，得到新的特征表示集  $\{R^{(1)}, R^{(2)}, \dots\}$
- 6: 以有标签的样本集  $\{R^{(1)}, R^{(2)}, \dots\}$  作为输入特征训练 Softmax 分类器
- 7: 进行一次前馈传递，计算  $L_2$  层、 $L_3$  层直至输出层  $L_{n_l}$  的激活值
- 8: 对输出层  $n_l$ ，令  $\delta^{(n_l)} = -(\nabla_{a^{(n_l)}} J) \cdot f'(z^{(n_l)})$
- 9: 对  $l = n_l - 1, n_l - 2, n_l - 3, \dots, 2$ ，令  $\delta^{(l)} = ((W^{(l+1)})^T \delta^{(l+1)} \cdot f'(z^{(l)}))$
- 10: 计算所需的偏导数： $\nabla_{w^{(l)}} J(W, b; x, y) = \delta^{(l+1)} (a^{(l)})^T, \nabla_{b^{(l)}} J(W, b; x, y) = \delta^{(l+1)}$
- 11: 采用梯度下降法进行参数更新。如果不满足停止条件，返回 7；否则，结束。
- 12: 无标签的测试样本集根据训练得到的深度网络进行分类测试并输出结果。

### 5.3 基于深度学习的方法在 TEP 故障诊断中的应用

#### 5.3.1 TEP 故障诊断应用

为了验证本章节所提出算法的有效性，我们选取 TEP 过程中的故障 2、4、9、

10、11、13 和 14，因为这些故障不但包括 TEP 仿真数据中所有的故障类型，而且很好地呈现故障数据间的交错性，详细的分析在前面章节已经介绍。基于这些数据，我们设置了两组实验，其中组合 1 包括故障 4、9 和 11，组合 2 包括故障 2、10、13 和 14。

故障 4、9 和 11 采用本章所提出的方法，其错分率为 27.90%。表 5-1 列出其组合故障所对应的混淆矩阵。由于故障 4 和 11 都与反应器冷却水入口温度有关，具有相同的故障变量，只是故障类型不同（故障 4 是阶跃变化，故障 11 是随机变化），所以具有很强的不可分性。故障 9 是物料 D 的温度的随机变化，完全不同于故障 4，但和故障 11 有较强的不可分性。从表 5-1 可以看出，故障 4 和 9 之间没有错分，而故障 4 和 11，故障 9 和 11 之间分别被错分。故障 2、10、13 和 14 采用本章所提出的方法，其错分率为 9.50%。表 5-2 列出了此组合故障所对应的混淆矩阵。组合 1 由于故障数据间的交错性，相对于组合 2 有一定的不可分性，因此其错分率要高一些。

表 5-1 故障 4、9 和 11 的实验混淆矩阵

Class	Fault 4	Fault 9	Fault 11
Fault 4	<b>736</b>	0	185
Fault 9	0	<b>590</b>	211
Fault 11	64	210	<b>404</b>
Total	800	800	800

表 5-2 故障 2、10、13 和 14 的实验混淆矩阵

class	Fault 2	Fault 10	Fault 13	Fault 14
Fault 2	<b>786</b>	4	1	0
Fault 10	12	<b>775</b>	242	0
Fault 13	2	20	<b>535</b>	0
Fault 14	0	1	22	<b>800</b>
Total	800	800	800	800

通过这两组实验结果可知，由于使用了稀疏自编码能够学习出较好表征原数据的特征以及 Softmax 回归模型能够解决类标签多于两种的情况，使得基于深度



学习的故障诊断方法对于较强不可分性的三类故障以及包含 TEP 数据集所有类型的四类故障具有不错的诊断率。

### 5.3.2 与其他方法的实验比较分析

为了进一步对所提出的故障诊断方法进行验证,本节对 TEP 数据集的 20 类故障数据进行实验。该数据集的每个样本数据有 52 维属性,每种类别分别包含有 480 个训练样本数据和 800 个测试样本数据。基于此数据集,对本章所提出的方法,并分别与基于支持向量机、结构化支持向量机、稀疏表示和随机森林的故障诊断方法进行实验比较分析。

#### 5.3.2.1 用于对比的各方法

由于支持向量机满足结构风险最小化原理,同时能减少训练误差且推广能力较强,所以在解决小样本和非线性问题上具有明显的优势,尤其适合于样本不容易得到的化工过程的故障诊断模型的建立。所以, SVM 在故障诊断领域得到广泛的应用。对于 20 种类别的故障样本数据, SVM 通常采用“一对一”或者“一对多”的分类器构造方法。

然而, SVM 采用“一对一”或者“一对多”策略构造分类器,对于 20 类故障数据需要构建 200 或者 20 个分类器,还可能造成样本的类别不确定性。所以,本文提出基于 Structural SVMs 构建统一的模型进行故障诊断,此算法在前面章节已详细介绍,本章不再赘述。

随机森林是由多个决策树组成的分类器,由所有决策树综合投票决定输入向量的最终类标签。基于随机森林方法进行故障诊断,就是对一个新的输入样本,每个分类器给出自己的诊断类别,成为“投票”,随机森林选择票数最多的类别作为最终结果。与其他组合分类技术比较,随机森林方法不会随着决策树的增加而产生过度拟合,并且泛化误差趋于一个上界[117]。另外,随机森林作为决策树基础上的提升算法,具备很多优越性:运行速度快,能高效处理大样本数据,能够并行处理;适用于数据集中存在大量未知特征;当数据集中存在部分变量值缺失时也不影响随机森林算法的使用等。但同时也发现一些不完善的地方,和不同算法相比优劣势也不相同:和 Boosting 相比,随机森林的执行速度要快很多,

当样本噪声很大时，随机森林模型优于 Boosting，但当样本噪声很小或可以忽略不计时，随机森林的分类准确率低于 Boosting[118]等。

稀疏表示的基本假设是：某一类的未知样本可以被该类的若干样本在某子空间有效地进行线性表示。基于稀疏表示的故障诊断方法，实质就是将化工过程中故障诊断理解为每一时刻测试样本与预先标定的能代表化工故障典型的训练样本进行匹配分类进而判断的过程。在现实化工过程中由于系统相对比较复杂，故障的类型往往很多，导致使用故障样本数据集进行新样本稀疏表示的系数会很稀疏，故障样本数据集中的多数类别的样本对应重构系数趋近于零。基于稀疏表示分析获取样本稀疏重构系数和重构误差的方法在获得故障样本数据集的基础上，对样本集中的所有样本通过重构系数和重构误差进行评估，从而选择出最具有代表性的一些样本来重构新样本。这些样本能够揭示新样本数据的主要特征和内在结构，通过对新样本的重构，从而使得故障诊断具有较强的鲁棒性[119]。

### 5.3.2.2 实验结果比较分析

表 5-3 列出了基于深度学习、Structural SVMs、稀疏表示、随机森林以及 SVM 进行故障诊断的实验结果。

从表中可以看出，本文提出的基于深度学习的故障诊断方法显然优于其他方法，故障诊断率平均达到 68.23%。在实际的化工过程中要想找到有利于分类的比较合理的表示方法的数据是比较困难的，而深度学习模仿人的大脑的学习过程，通过组合较低层的特征形成更为抽象的高层表示，从而发现观测数据的分布式特征，进而学习更有用的特征表征输入数据。同时，深度学习的非监督特征学习与人工规则构造特征的方法相比，利用数据多层次地学习特征，更能够刻画数据的丰富内在信息。因此，非监督、多层次的特征学习使深度学习在分类或预测上有了更高的精确度。

结构化支持向量机本质是一种浅层机器学习模型，此模型的结构可以看成带有一层隐藏节点，其训练方法需要很多经验和技巧。深度学习可以设计模型结构的深度并自动地进行特征学习，具有较强的表达能力。也就是说对于复杂的样本数据，深度模型相比较于浅层模型有更好的对非线性函数的表示能力。因此，对于化工过程中复杂的 20 类故障数据进行判别时，基于深度学习的故障诊断方法

优于基于结构化支持向量机的方法。基于结构化支持向量机进行故障诊断的方法突破了传统的支持向量仅能判别两类故障或者要构建多个模型判别多类的情况。它以结构信息为基础建立空间度量，拓展了并列特征的方法，对于多个故障用统一的模型进行判别故障的类型，有效简化了分类器模型。所以，与传统的 SVM 相比，它节约了训练时间，而且降低了错分率。从表中可知，结构化支持向量机平均故障诊断率 64.09% 仅次于深度学习方法，具有不错的表现。

表 5-3 故障诊断方法比较

编号 \ 方法	深度学习	结构化SVM	随机森林	稀疏表示	SVM
1	99.00%	99.79%	98.5%	94.3%	97.7%
2	98.40%	99.36%	97.9%	93.3%	95.8%
3	29.00%	22.13%	25.8%	7.7%	17.3%
4	93.13%	49.57%	89.3%	33.8%	85.8%
5	99.38%	60.00%	57.1%	75.3%	92.6%
6	54.25%	100%	99.1%	100%	100%
7	100%	100%	98.1%	94.7%	100%
8	58.13%	96.17%	46.3%	41.3%	50.0%
9	24.13%	14.04%	18.8%	9.3%	15.0%
10	73.38%	69.15%	37.6%	51.7%	19.6%
11	44.13%	42.77%	50.2%	12.1%	13.3%
12	80.13%	98.94%	57.9%	42.4%	57.5%
13	35.00%	93.83%	16.6%	35.9%	39.8%
14	96.25%	100%	94.0%	56.9%	73.6%
15	27.88%	42.55%	17.8%	8.0%	16.1%
16	77.00%	6.81%	32.0%	27.4%	26.7%
17	89.50%	75.32%	73.4%	69.9%	79.2%
18	28.88%	78.30%	81.9%	84.6%	89.2%
19	87.63%	100%	65.5%	48.8%	55.3%
20	77.39%	73.40%	55.8%	64.2%	52.7%

Average	68.23%	64.09%	60.7%	52.6%	56.68%
---------	--------	--------	-------	-------	--------

随机森林模型如果在建立过程中可供选择的样本很少，就不能很好地发挥 Bagging 的优势。但是，随机森林作为决策树基础上的提升算法，具备很多优越性。随机森林自身就是一种多分类算法，可以一次性对多种类型数据进行建模和分类测试，运行速度快。所以，相对于传统的 SVM，有比较高的平均故障诊断率。稀疏表示的方法通过故障训练数据组成重构样本字典用于测试样本的重构，从而获得用于故障诊断的重构系数和重构残差。它对变量间的差异性要求比较高。由于样本的 52 个变量间具有很强的相关性，所以它的诊断率不高。

图 5-3 显示了 20 种故障诊断率的大致趋势。可以看出，无论采用那种方法对这 20 种故障的诊断趋势大致是一致的。对于较易判别的故障，这些方法都有高的诊断率，例如故障 1、故障 2。故障 1 为 A/C 进料流量比变化，组分 B 含量不变。由于反应物 A 和 C 的比例发生变化，那么与物质平衡有关的变量如压力、液位、成份的分配关系也会随之改变。因此，超出一半的监控变量都会明显偏离它们正常运行的状态，从而这种故障很容易和别的故障区分出来。故障 2 为组分 B 含量发生变化，A/C 进料流量比不变。同上，故障发生后，超过一半数量的变量都明显偏离其正常运行状态，所以较容易被诊断出故障的类型，诊断率相对都比较高。对于一些不易判别的故障，这些方法大多诊断率都相对较低，例如故障 3。故障 3 为 D 的进料温度，当故障发生后观测变量与正常情况下的变量相比没有明显变化，其故障类型不易被诊断，因此该故障的诊断率大多比较低。

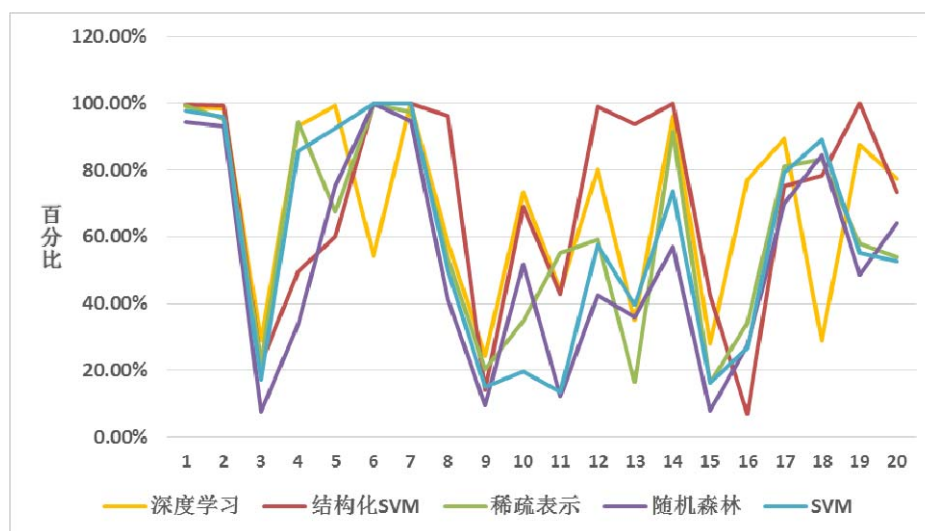


图 5-3 故障诊断率趋势图

#### 5.4 本章小结

本章阐述了一种基于深度学习的故障诊断算法的方法,把化工过程故障数据进行 PCA 白化预处理,再利用稀疏自动编码器对故障数据进行无监督的学习,最后用 Softmax 对故障数据进行监督的学习,进行化工故障的识别。在 TEP 数据集上的实验结果表明,基于深度学习的方法与其他几种有代表性的故障诊断方法相比,故障诊断准确率有了明显提高。



## 第六章 总结与展望

### 6.1 本文工作总结

本文针对目前化工过程中故障检测与故障诊断存在的问题，提出一些新方法，旨在能够及时地检测故障的发生并诊断出故障的根源。本文的研究成果如下：

(1) 传统的方法，比如 PCA, ICA 等，已广泛应用于化工过程的故障检测。但是它们只是实现对高维数据的降维，而忽略对过程数据中重要特征的提取。本文提出了一种基于 NMFSC 进行故障检测的方法，一方面由于对分解结果加入非负限制，基于非负矩阵分解提取的成分向量能很好地反映样本的局部特征，所以它可以很好地解释化工过程中故障的发生或许仅由几个变量的变化所导致的；另一方面，NMFSC 对分解矩阵添加稀疏性约束条件能更好地发现稳定、直观的局部特征，并能自由地控制分解后矩阵的稀疏度，求解收敛速度快。由于 NMFSC 算法对基矩阵  $W$  和系数矩阵  $H$  的初始化是随机的，因此导致每次 NMFSC 的结果有可能不同，影响算法的稳定性。所以，本文进一步提出采用 PCA 来对 NMFSC 进行初始化来解决该问题。另外，大多数工业过程，都存在动态特性，当前时刻的测量变量与过去若干时刻的测量变量都有关系。但是，NMFSC 是一种静态的矩阵分解算法，不能有效地监控动态过程。基于此，本文还提出采用指数加权滑动平均对采样数据进行处理。实验结果表明，基于 NMFSC 的故障检测与经典的 PCA 方法相比，故障检测率有了明显的提高。

(2) 在化工工业过程中，由于系统的复杂性，产生故障的种类往往很多，而且对分类速度和精度要求也比较高。然而，传统的支持向量机解决的是二类分类问题，对于多类分类问题，多是通过某种方式构造一系列两类分类器并将它们组合在一起来实现多类分类，其本质仍是二类分类问题。针对于此，本文提出了一种使用 Structural SVMs 进行故障故障诊断的方法，不仅解决了化工过程中多类故障需要训练大量分类器的问题，节约了训练时间，而且还降低了错分率，提高了故障分类效率。此外，本文还提出了一种基于特征优化的结构化支持向量机进行故障诊断的方法。

对于故障特征全集而言，并不是每一个特征都能够对故障诊断做出同等贡献

的。从特征全集中选取出能够提供较大信息量的特征来建立故障诊断模型是非常必要的。因此,本文提出采用决策树提取对于区分故障能够提供较大信息的特征。针对化工过程中观测数据的序列相关性,本文引入过程变量的延时测量值,用延时测量值的均值和方差扩展样本的原始特征,组合成新的数据矩阵,对故障数据的动态特性进行提取并有效地加以表示,这样能够有效地进行故障分类。在实验仿真化工过程 TEP 数据集上的实验证明了所提出方法的有效性。

(3) 当前化工工业的大多监督学习算法,一般需要获取更多的已标记样本,这样做成本通常很高。本文提出了一种基于深度学习进行故障诊断的算法。该算法利用输入与输出之间的非线性关系拟合观测样本,从未标记数据中采用稀疏自编码器进行特征学习,自动获取鲁棒的故障特征表达,然后通过构建能够解决多类问题的 Softmax 回归模型进行分类,判别化工故障的类型。与其他几种有代表性的故障诊断方法相比,基于深度学习的故障诊断方法具有较高的诊断率。

## 6.2 未来工作展望

故障检测与故障诊断发展近二十年,虽然已取得不少的研究成果,但将这种技术直接应用于化工工业生产,在理论和应用上都还存在一些问题有待于研究和完善。结合本文研究中所遇到的问题,关于未来的研究工作提出以下展望:

(1) 本文提出的 NMFSC 进行故障检测,只是在单尺度建模,然而实际工业过程中采集的数据本质上是多尺度的。所以,下一步工作可以考虑将 NMFSC 的单尺度建模方法推广到多尺度,可以采用小波分析方法等。由于小波分析具有多分辨率分析的能力,能够有效地对信号进行时频分解,这样可以模拟实际工业过程中所获取的海量数据,其故障可能发生在不同的频段上。

(2) 目前大多分类器假定变量对分类具有相同的贡献,它们具有相同的权值。然而,当所有变量使用相同的权值时,故障数据集将遭受无关信息的影响,从而使分类器遭受更多的重叠数据的影响,也就较难从这些故障数据中抽取有效的分类信息。变量加权可以通过选取的特征对分类的贡献不同,对变量增加不同的权重。在各故障数据分别被相应的加权向量加权后,所有变量将使用不同的权值,通过构造组合故障数据来放大各故障本身的特征,这可以有效地改善 Structural SVMs 从相互重叠的数据中抽取判别信息。所以,下一步研究工作中考



虑将变量加权融入基于 **Structural SVMs** 进行故障诊断方法中。

(3) 基于 **Structural SVMs** 进行故障诊断，关于输出的表达，本文只是用简单的类别表示  $(1, 2, \dots, k)$ 。下一步研究工作可以尝试用决策树表示来代替类别，能够更形象地表达某一故障。在此基础上，并设计相应的结构化特征函数。

(4) 目前在化工工业领域，非监督特征学习目前研究还比较少，如何自动学习获取鲁棒的故障特征表达对于化工工业应用具有至关重要的意义。由于深度学习能够从未标记数据中很好地进行非监督特征学习，本文创新性地把深度学习引入化工过程故障诊断领域，不过只是初步的试验，还需要进一步的深入研究。



## 参考文献

- [1] 马杰, 宋建池. 近 8 年我国化工事故统计与分析[J]. 工业安全与环境, 2009, 28(7): 37-38.
- [2] 吴宗之等. 2006-2010 年我国危险化学品事故统计分析研究. 中国安全生产科学技术, 2011, 7(7).
- [3] 田文德, 张军. 化工安全分析中的过程故障诊断. 北京: 冶金工业出版社, 2008.
- [4] 何小斌. 基于统计学方法的自适应过程监控与故障诊断. 上海交通大学博士学位论文, 2008.
- [5] 李蓉一. 基于核主元分析的故障检测与诊断研究. 南京师范大学硕士学位论文, 2011.
- [6] 李晗, 萧德云. 基于数据驱动的故障诊断方法综述[J]. 控制与决策, 2011, 26(1): 1-9.
- [7] 周东华, 胡艳艳. 动态系统的故障诊断技术[J]. 自动化学报, 2009(6): 748-758.
- [8] 马贺贺. 基于数据驱动的复杂工业过程故障检测方法研究. 华东理工大学博士学位论文, 2012.
- [9] Kruger U., Kumar S., Littler T. Improved principal component monitoring using the local approach[J]. Automatica, 2007, 43:1532-1542.
- [10] Lee J. M., Qin S.J., Lee I. B. Fault detection of non-linear processes using kernel independent component analysis [J]. The Canadian Journal of Chemical Engineering, 2007, 85(4): 526-536.
- [11] Li Y.F., Wang Z.F., Yung J. Q. On-line Fault Detection Using SVM-based Dynamic MPLS for Batch Processes [J]. Chinese Journal of Chemical Engineering, 2006, 14(6): 754-758.
- [12] Nguyen V. H., Golival J. C. Fault detection based on kernel principal component analysis [J]. Engineering Structures, 2010, 32(11): 3683-3691.
- [13] Shams B., Budman H. M., Duever T. A. Fault detection, identification and

- diagnosis using CUSUM based PCA [J]. *Chemical Engineering Science*, 2011, 66(20): 4488-4498.
- [14] Grbovic M., Li W., Xu P., et al. Decentralized fault detection and diagnosis via sparse PCA based decomposition and Maximum Entropy decision fusion [J]. *Journal of Process Control*, 2012, 22(4): 738-750.
- [15] Pearson K. LIII. On lines and planes of closest fit to systems of points in space[J]. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 1901, 2(11): 559-572.
- [16] Malthouse E. C., Tamhane A. C., Mah R. S. H. Nonlinear partial least squares. *Computers & Chemical Engineering*, 1997, 21(8): 875-890.
- [17] MacGregor J. F., Jaeckle C., Kiparissides C., Koutoudi M. Process monitoring and diagnosis by multiblock PLS methods[J]. *AIChE*, 2004, 40(5): 826-838.
- [18] Kano M., Tanaka S., Hasebe S., Hashimoto I., Ohno H. Monitoring independent components for fault detection[J]. *AIChE*, 2003, 49(4): 969-976.
- [19] Lee J. M., Qin S.J., Lee I. B. Fault detection of non-linear processes using kernel independent component analysis [J]. *AIChE*, 2006, 52:3501-3514.
- [20] Zhang Y, Zhang Y. Fault detection of non-Gaussian processes based on modified independent component analysis [J]. *Chemical engineering science*, 2010, 65(16): 4630-4639.
- [21] Westerhuis J. A., Gurden S. P., Smilde A. K. Generalized contribution plots in multivariate statistical process monitoring[J]. *Chemometrics and Intelligent Laboratory Systems*, 2000, 51(1): 95-114.
- [22] 陈玉东. 动态系统的故障诊断方法综述[J]. *化工自动化仪表*, 2001, 28(3): 1-14.
- [23] Lerner U., Parr R., Koller D., et al. Bayesian fault detection and diagnosis in dynamic systems[C]. *AAAI/IAAI*, 2000: 531-537.
- [24] Barford L., Kanevsky V. and Kamas L. Bayesian fault diagnosis in large-scale measurement systems. *Instrumentation and Measurement Technology Conference*, 2004. *IMTC 04. Proceedings of the 21st IEEE*, 2004, 2: 1234-1239.
- [25] Dearden R. Bayesian fault diagnosis: Common approaches and challenges[C].

- Cognitive Information Processing (CIP), 2010 2nd International Workshop on. IEEE, 2010: 215-220.
- [26] Dos Santos D A, Yoneyama T. A Bayesian solution to the multiple composite hypothesis testing for fault diagnosis in dynamic systems [J]. *Automatica*, 2011, 47(1): 158-163.
- [27] He Q. P., Qin S. J., Wang J. A new fault diagnosis method using fault directions in Fisher discriminant analysis[J]. *AIChE journal*, 2005, 51(2): 555-571.
- [28] Zhu Z. B., Song Z. H. Fault diagnosis based on imbalance modified kernel Fisher discriminant analysis [J]. *Chemical Engineering Research and Design*, 2010, 88(8): 936-951.
- [29] Zhu Z. B., Song Z. H. A novel fault diagnosis system using pattern classification on kernel FDA subspace [J]. *Expert Systems with Applications*, 2011, 38(6): 6895-6905.
- [30] Zhang M. L., Zhou Z.H. ML-KNN: A lazy learning approach to multi-label learning [J]. *Pattern Recognition*, 2007, 40(7): 2038-2048.
- [31] Zhang J. Improved on-line process fault diagnosis through information fusion in multiple neural networks[J]. *Computers & chemical engineering*, 2006, 30(3): 558-571.
- [32] Saravanan N., Ramachandran K. I. Incipient gear box fault diagnosis using discrete wavelet transform (DWT) for feature extraction and classification using artificial neural network (ANN)[J]. *Expert Systems with Applications*, 2010, 37(6): 4168-4181.
- [33] Chiang L. H., Kotanchek M. E., Kordon A. K. Fault diagnosis based on Fisher discriminant analysis and support vector machines [J]. *Computers & chemical engineering*, 2004, 28(8): 1389-1401.
- [34] Mahadevan S., Shah S. L. Fault detection and diagnosis in process data using one-class support vector machines [J]. *Journal of Process Control*, 2009, 19(10): 1627-1639.
- [35] Liang S. Y., Lv J. H. Least Squares Support Vector Machine for Fault Diagnosis Optimization [J]. *Applied Mechanics and Materials*, 2013, 347: 505-508.

- [36] Zhang Y., Liu X. D., Xie F. D., et al. Fault classifier of rotating machinery based on weighted support vector data description [J]. *Expert Systems with Applications*, 2009, 36(4): 7928-7932.
- [37] Ge Z., Gao F., Song Z. Batch process monitoring based on support vector data description method [J]. *Journal of Process Control*, 2011, 21(6): 949-959.
- [38] Breiman L. Random forests [J]. *Machine Learning*, 2001, 45(1):5-32.
- [39] Statnikov A., Wang L., Aliferis C. F. A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification[J]. *BMC bioinformatics*, 2008, 9(1): 319-327.
- [40] Staroswiecki M., Comtet-Varga G. Analytical redundancy relations for fault detection and isolation in algebraic dynamic systems [J]. *Automatica*, 2001, 37(5): 687-699.
- [41] Gertler J., Li W., Huang Y., et al. Isolation enhanced principal component analysis [J]. *AIChE Journal*, 1999, 45(2): 323-334.
- [42] 刘强, 柴天佑, 秦泗钊, 赵立杰. 基于数据和知识的工业过程监视及故障诊断综述. *控制与决策*, 2010, 25(6): 801-807.
- [43] Ku, W., Storer, R.H., and Georgakis C. Disturbance detection and isolation by dynamic principal component analysis[J]. *Chemometrics and Intelligent Laboratory Systems*, 1995, 30:179-196.
- [44] Tsung F. Statistical monitoring and diagnosis of automatic controlled process using dynamic PCA[J]. *International Journal of Production Research*, 2000, 38(3): 625-637.
- [45] Lee J. M., Yoo C. K., Lee I. B. Statistical monitoring of dynamic processes based on dynamic independent component analysis [J]. *Chemical Engineering Science*, 2004, 59(14): 2995-3006.
- [46] Lee J. M., Yoo C. K., Choi S. W., etc. Nonlinear process monitoring using kernel principal component analysis[J]. *Chemical Engineering Science*, 2004, 59(1): 223-234.
- [47] Ouyang S., Bao Z., and Liao G. S. Robust recursive least squares learning algorithm for principal component analysis[J], *IEEE Trans. Neural Netw.*, 2000,

10(1):215-221.

[48] Xie L., Wang S.Q., Zhang J.M. Block Recursive MPCA and its application in Batch Process Monitoring[C], IEEE International Conference on System, Man & Cybernetic, Washington, USA,2003.

[49] Zhang Y.W., Li S., Teng Y.D. Dynamic process monitoring using recursive kernel principal component analysis [J], Chemical Engineering Science, 2012, 72:78-86.

[50] Wold S., Kettaneh N., Tjessem K. Hierarchical multiblock PLS and PC models for easier model interpretation and asan alternative to variable selection [J]. Journal of Chemometrics, 1996, 10(5-6): 463-482.

[51] Cherry G. A., Qin S. J. Multiblock principal component analysis based on a combined index for semiconductor fault detection and diagnosis [J]. Semiconductor Manufacturing, IEEE Transactions on, 2006, 19(2): 159-172.

[52] Lee D. S., Vanrolleghem P. A. Adaptive consensus principal component analysis for on-line batch process monitoring [J]. Environmental monitoring and assessment, 2004, 92(1-3): 119-135.

[53] Bakshi B. R. Multiscale PCA with application to multivariate statistical process monitoring [J]. AIChE, Journal, 1998, 44(7):1596-1610.

[54] Lee H.W., Lee M.W., Park J.M. Multi-scale extension of PLS algorithm for advanced on-line process monitoring [J]. Chemometrics and Intelligent Laboratory Systems, 2009, 98:201-212.

[55] Roodbali M.S.E., Shahbazian M. Multi-Scale PLS Modeling for Industrial Process Monitoring [J]. International Journal of Computer Applications, 2011, 26(6):26-33.

[56] Zhang Y. and Qin S. J. Fault Detection of Nonlinear Processes Using Multiway Kernel Independent Component Analysis [J]. Industrial & Engineering Chemistry Research, 2007, 46:7780-7787.

[57] Choi S. W., Morris J. L., Lee I. B. Nonlinear multiscale modelling for fault detection and identification [J]. Chemical Engineering Science, 2008, 63(8): 2252-2266.

[58] Wang Y. J., Jia M. X., and Mao Z. Z. Weak fault monitoring method for batch

process based on multi-model SDKPCA [J]. *Chemometrics and Intelligent Laboratory Systems*, 2012, 118: 1-12.

[59] Li Y.F., Wang Z.F., Yuan J.Q. On-line Fault Detection Using SVM-based Dynamic MPLS for Batch Processes [J]. *Chinese Journal of Chemical Engineering*, 2006, 14(6): 754-758.

[60] Chiang L. H., Kotanchek M. E., Kordon A. K. Fault diagnosis based on Fisher discriminant analysis and support vector machines [J]. *Computers & chemical engineering*, 2004, 28(8): 1389-1401.

[61] Zhang Y. W. Enhanced statistical analysis of nonlinear process using KPCA, KICA and SVM[J]. *Chemical Engineering Science*, 2009, 64(5): 801-811.

[62] He X. B., Wang W., Yang Y. P., Yang Y. H. Variable-weighted Fisher discriminant analysis for process fault diagnosis [J]. *Journal of Process Control*, 2009, 19(6): 923-931.

[63] Zhu Z. B., Song Z. H. A novel fault diagnosis system using pattern classification on kernel FDA subspace [J]. *Expert Systems with Applications*, 2011, 38(6): 6895-6905.

[64] Monroy I., Benitez R., Escudero G., et al. A semi-supervised approach to fault diagnosis for chemical processes [J]. *Computers & Chemical Engineering*, 2010, 34(5): 631-642.

[65] Barabino N., et al. Support vector machines vs multi-layer perceptrons in particle identification[C]. In *ESANN*, 1999: 257-262.

[66] Wu L., Chen X., Peng Y., et al. Fault Detection and Diagnosis based on Sparse Representation Classification (SRC)[C]. *Robotics and Biomimetics, ROBIO 2012, IEEE International Conference on. IEEE*, 2012: 926-931.

[67] Zhu Z. B., Song Z. H. Fault diagnosis based on imbalance modified kernel Fisher discriminant analysis [J]. *Chemical Engineering Research and Design*, 2010, 88(8): 936-951.

[68] 李乐, 章毓晋. 非负矩阵分解算法综述. *电子学报*, 36(4): 737-743, 2008.

[69] Lee D. D., Seung H. S.. Learning the parts of objects by non-negative matrix factorization[J]. *Nature*, 1999, 401(6755):788-791



- [70] Lee D. D., Seung H. S. Algorithms for non-negative matrix factorization [C]. In: *Advances in Neural Information Processing Systems 13*, Denver, 2000. 556-562
- [71] Seung D, Lee L. Algorithms for non-negative matrix factorization [J]. *Advances in neural information processing systems*, 2001, 13: 556-562.
- [72] Chu M., Diele F., Plemmons R., et al. Optimality, computation, and interpretation of nonnegative matrix factorizations [C]. *SIAM Journal on Matrix Analysis*, 2004.
- [73] 邓乃扬, 田英杰. 支持向量机—理论、算法与拓展 [M]. 科学出版社, 2009.
- [74] 邓乃扬, 田英杰. 数据挖掘中的新方法—支持向量机 [M]. 科学出版社, 2004.
- [75] Vapnik V., Levin E., Le Cun Y. Measuring the VC-dimension of a learning machine [J]. *Neural Computation*, 1994, 6(5): 851-876.
- [76] Bousquet O., Boucheron S., Lugosi G. Introduction to statistical learning theory [M]. *Advanced Lectures on Machine Learning*. Springer Berlin Heidelberg, 2004: 169-207.
- [77] Hsu C. W., Lin C. J. A comparison of methods for multiclass support vector machines [J]. *Neural Networks, IEEE Transactions on*, 2002, 13(2): 415-425.
- [78] Weston J, Watkins C. Support vector machines for multi-class pattern recognition [C]. *ESANN*, 1999, 99: 61-72.
- [79] Knerr S., Personnaz L., Dreyfus G. Single-layer learning revisited: a stepwise procedure for building and training a neural network [M]. *Neurocomputing*, Springer Berlin Heidelberg, 1990: 41-50.
- [80] Platt J. C., Cristianini N., Shawe-Taylor J. Large Margin DAGs for Multiclass Classification [C]. *NIPS*, 1999, 12: 547-553.
- [81] 陈铭. 基于 LS-SVM 的复杂工业过程故障诊断方法研究. 南京理工大学硕士学位论文, 2009.
- [82] 苏海. 基于 SVM 的邮件内容分类方法研究. 大连理工大学硕士学位论文, 2013.
- [83] Hoyer P. O. Non-negative matrix factorization with sparseness constraints [J]. *The Journal of Machine Learning Research*, 2004, 5: 1457-1469.

- [84] Zheng Y., Yang J. and Zhu Y. Initialization enhancer for non-negative matrix factorization. *Engineering Applications of Artificial Intelligence*, 2007, 20: 101-110.
- [85] Hsu C. C., Chen L. S., Liu C. H. A process monitoring scheme based on independent component analysis and adjusted outliers, *International Journal of Production Research*, 48(6), 2010.
- [86] Downs J. J. and Vogel E. F. A plant-wide industrial process control problem[J]. *Computers & Chemical Engineering*, 17(3): 245-255, 1993.
- [87] Lyman P. R. and Georgakist C. Plant-wide control of the Tennessee Eastman problem. *Computers & Chemical Engineering*, 19(3): 321-331, 1995.
- [88] Laukonen E. G., Passino K. M., Krishnaswami V., et al. Fault detection and isolation for an experimental internal combustion engine via fuzzy identification [J]. *Control Systems Technology*, IEEE Transactions on, 1995, 3(3): 347-355.
- [89] Duda R. O., Hart P. E., et al. 模式分类 [M]. 机械工业出版社, 2012.
- [90] Lafferty J., McCallum A., Pereira F. C. N. Conditional random fields: Probabilistic models for segmenting and labeling sequence data [J]. 2001.
- [91] Tsochantaridis I., Hofmann T., Joachims T., et al. Support vector machine learning for interdependent and structured output spaces[C]. *Proceedings of the twenty-first international conference on Machine learning*, ACM, 2004: 104.
- [92] Collins M. Discriminative training methods for hidden markov models: Theory and experiments with perceptron algorithms[C]. *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, Association for Computational Linguistics, 2002: 1-8.
- [93] Roller B. Max-margin Markov networks [J]. *Proc. Advances in Neural Information Processing Systems*, MIT Press, 2004: 25.
- [94] Tsochantaridis I., et al, Support Vector Machine learning for interdependent and structured output spaces. *Proceeding of the 21th International Conference on Machine Learning*, 2004.
- [95] Joachims T., Finley T., and Yu C.J. Cutting-plane training of Structural SVMs. *Machine Learning*, 2009, 77(1): 27-59.
- [96] Tsochantaridis I., et al. Large margin methods for structured and interdependent

- output variables. *Journal of Machine Learning Research*, 6:1453-1484, 2005.
- [97] Yu C., and Joachims T. learning structural SVMs with latent variables. *Proceeding of the 26th International Conference on Machine Learning*, 2009.
- [98] Crammer K., Singer Y. On the algorithmic implementation of multiclass kernel-based vector machines [J]. *The Journal of Machine Learning Research*, 2002, 2: 265-292.
- [99] Har-Peled S., Roth D., Zimak D. Constraint classification for multiclass classification and ranking [J]. *Urbana*, 2002, 51: 61801.
- [100] Tasker B., Guestrin C., Koller D. Max-margin Markov networks [J]. In *Advances in Neural Information Processing Systems*, 2004.
- [101] Chiang L. H., Kotanchek M. E. and Kordon A. K. Fault diagnosis based on Fisher discriminant analysis and support vector machines. *Computer & Chemical Engineering*, 28(8): 1389-1401, 2004.
- [102] Kulkarni A., Jayaraman V. K. and Kulkarni B.D. Knowledge incorporated support vector machines to detect faults in Tennessee Eastman Process. *Computers & Chemical Engineering*, 29(10): 2128-2133, 2005.
- [103] 范明, 孟小峰(译). 数据挖掘概念与技术. 北京: 机械工业出版社, 2007.
- [104] 李航. 统计学习方法. 北京: 清华大学出版社, 2012.
- [105] Iri M., Aoki K., et al. An algorithm for diagnosis of system failures in the chemical process[J]. *Computers and Chemical Engineering*, 1979, 3(1): 489-493.
- [106] Caceres S., Henley E. J. Process failure analysis by block diagrams and fault trees[J]. *Industrial and Engineering Chemistry Fundamentals*, 1976, 15(2): 128-134.
- [107] Leung D, Romagnoli J. Dynamic probabilistic model based expert system for fault diagnosis[J], *Computers and Chemical Engineering*, 2000, 24(11):2473-2492.
- [108] Ku W., Storer R. H. and Georgakis C. Disturbance detection and isolation by dynamic principal component analysis[J]. *Chemometrics and intelligent laboratory systems*, 1995, 30(1): 179-196.
- [109] Li W., Qin S.J. Consistent dynamic PCA based on error-in-variables subspace identification[J]. *Process Control* 11 (2001) 661-678.

- [110] Lee C., Choi S. W., Lee I. B. Sensor fault identification based on time-lagged PCA in dynamic processes [J]. Chemometrics and intelligent laboratory systems, 2004, 70(2): 165-178.
- [111] 孙志军, 薛磊, 许阳明, 王正. 深度学习研究综述. 计算机应用研究, 29(8): 2806-2810, 2012.
- [112] Hinton G. E., Salakhutdinov R. R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507.
- [113] Huval B., Coates A., Ng A. Deep learning for class-generic object detection[J]. arXiv preprint arXiv:1312.6885, 2013.
- [114] Ng A. Sparse autoencoder[J]. CS294A Lecture notes, 2011: 72.
- [115] Ngiam J., Coates A., Lahiri A., et al. On optimization methods for deep learning [C]. Proceedings of the 28th International Conference on Machine Learning (ICML-11), 2011: 265-272.
- [116] [http://deeplearning.stanford.edu/wiki/index.php/UFLDL\\_Tutorial](http://deeplearning.stanford.edu/wiki/index.php/UFLDL_Tutorial)
- [117] 刘艳丽. 随机森林综述[D]. 南开大学硕士学位论文, 2008.
- [118] Dietterich T. An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting and Randomization [J]. Machine Learning, 2000, 40(2).
- [119] 武利军. 基于动态特性与稀疏表示的故障诊断方法研究. 中国科学院大学硕士学位论文, 2013.

## 附录：主要符号对照表

SPE	平方预测误差 (Square Predicted Error)
PCA	主元分析 (Principal Components Analysis)
DPCA	动态主元分析 (Dynamic Principal Components Analysis)
FDA	判别分析 (Fisher Discriminant Analysis)
VW-FDA	变量加权判别分析 (Variable-Weighted Fisher Discriminant Analysis)
ICA	独立分量分析 (Independent Component Analysis)
PLS	偏最小二乘 (Partial Least Square)
ANN	神经网络 (Artificial Neural Networks)
KNN	K 邻法 (K Nearest Neighbor)
SVM	支持向量机 (Support Vector Machine)
SVDD	支出向量数据描述 (Support Vector Data Description)
RF	随机森林 (Random Forests)
GMM	高斯混合模型 (Gaussian Mixture Model)
TEP	田纳西-伊斯曼过程 (Tennessee Eastman Process)
NMF	非负矩阵分解 (Non-negative Matrix Factor)
NMFSC	稀疏化非负矩阵分解 (Non-negative Matrix Factorization with Sparseness Constraints)
KDE	核密度估计 (Kernel Density Estimation)
SED	欧几里德距离 (Square of Euclidean distance)

DAG	有向无环图(Directied Acyclic Grap)
AP	平均感知器(Averaged Perceptron)
M3Ns	最大边际马尔科夫网络(Max Margin Markov Networks)
DL	深度学习(Deep Learning)
RBM	受限玻尔兹曼机(Restricted Boltzmann Machine)
AE	自动编码器(AutoEncoder)
SAE	稀疏自动编码器(Sparse AutoEncoder)

## 作者简历与论文发表

2006.09-2009.07	中国科学院研究生院	计算机应用技术	硕士
2009.08-2011.02	中国国信信息总公司		
2011.03-2014.06	中国科学院大学	计算机应用技术	博士

### 已发表文章目录

1. **Yi Peng**, Xiaodan Zhang, Zhenjun Han and Jianbin Jiao, "Dynamic Fault Diagnosis in Chemical Process Based on SVM-HMM," Proc.of IEEE International Conference on Mechatronics and Automation, 2013.(EI)

2. **Yi Peng**, Xiaogang Chen, Qixiang Ye, Jianbin Jiao, "Fault Detection and Classification in Chemical Processes Using NMFSC and Structural SVMs", The Canadian Journal of Chemical Engineering [J]. 92(6):1016-1023, 2014. (SCI)

3. **Yi Peng**, Qixiang Ye, Jianbin Jiao, Xiaogang Chen and Lijun Wu, "Fault Diagnosis via Structural Support Vector Machines," Proc.of IEEE International Conference on Mechatronics and Automation, 2012.(EI)

4. Lijun Wu, Xiaogang Chen, **Yi Peng**, Qixiang Ye, Jianbin Jiao, "Fault Detection and Diagnosis based on Sparse Representation Classification (SRC)", Proc.of IEEE International Conference on Robotics and Biomimetics, 2012.(EI)

5. **彭艺**, 叶齐祥, 黄钧, 焦建彬. 一种内容完整的视频稳定算法. 中国图象图形学报, 2010, 15(9): 1384-1390. (核心期刊)

6. **Y. Peng**, Q.x. Ye, and J.b. Jiao, "Shaking Video Stabilization with Content Completion", IS&T/SPIE Visual Communication and Image Processing, 2009, (VCIP2009), San Jose, CA, USA.(EI)





## 致 谢

转眼间博士学习生涯即将告一段落，经历了诸多坎坷，付出了艰辛的努力，同时也得到很大的收获，这无疑成为我人生旅途中最为重要和难忘的时光。借毕业论文完成之际，谨向所有关心和帮助过的老师、同学、朋友和家人表示深深的谢意！

首先，要特别感谢我的恩师焦建彬教授！焦老师治学严谨求实，一丝不苟，在论文选题、研究方法以及论文撰写上，每一环节无一不倾注着他的心血和指导。感谢焦老师对我的严格要求，使我在写作、演讲和工作能力等多个方面受到了教育，取得了进步。此外，他从生活上和思想上都一直给予我悉心的指导和关心，再次表示万分的感谢！焦老师对事物孜孜不倦的追求和宽厚坦荡的为人给我留下了深刻的印象，这在我今后的工作学习中必将受益终生。

其次，衷心感谢实验室叶齐祥副教授！叶老师在我的理论学习和实验研究中给予了耐心的指导和帮助，他对于研究领域精深的理解以及严谨的治学态度，开阔了我的研究视野，启发了我在研究中发现新问题。此外，叶老师还充当兄长的角色，在生活上给予我极大的关心和帮助，再次表示感谢！

感谢实验室的师弟师妹们，感谢一起学习生活中结下的深厚友谊，感谢一起度过的愉快而短暂的美好时光，同窗之谊弥足珍贵。特别感谢对本文有所贡献的师弟陈孝罡、武利军以及师妹张晓丹、李策、魏朋旭！

再次，深深地感谢我的父母和哥哥姐姐，感谢你们为提供的一切和最伟大的爱，是你们在我面临困难时给我指引方向，在我徘徊迷茫时教会我懂得去摸索。感谢我的先生苏飞博士，不但从科研上给予我很多帮助，在生活上给我耐心的照顾，而且从精神上给我很大的支持，使我能够顺利完成博士学业。

最后，感谢在百忙之中抽出宝贵时间参加论文评审和答辩的各位老师！

**彭艺**  
**2014年4月**