江南大学

全日制专业硕士学位论文

题　　目：多块建模策略下的 kNN 故障检测方法研究

英文并列题目：Research on kNN fault detection method under multi-block modeling strategy

研　究　生：郑静　　专 业 领 域 ：控制工程

研 究 方 向：控制工程及应用

导 师 ① 姓 名：熊伟丽　　职　　称：教授

导 师 ② 姓 名：张保平　　职　　称：高级工程师

学位授予日期：2022 年 6 月

答辩委员会主席：潘丰

江 南 大 学

地址：无锡市蠡湖大道 1800 号

二〇二二 年 六 月

独 创 性 声 明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含本人为获得江南大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

签名：郑静　　日期：2022 年 5 月 31 日

关于论文使用授权的说明

本学位论文作者完全了解江南大学有关保留、使用学位论文的规定：江南大学有权保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅，可以将学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文，并且本人电子文档的内容和纸质论文的内容相一致。

保密的学位论文在解密后也遵守此规定。

签名：郑静　　导师签名：

日期：2022 年 5 月 31 日

摘 要

现代工业过程日趋复杂，故障类别日益增多，一旦发生故障，不仅会降低经济效益，还可能造成严重的人员伤亡事故。随着计算机技术和数据采集设备的发展，一些工业过程积累了丰富的过程数据，使得基于数据驱动的故障检测技术不断进步，成为保证工业安全运行的重要关键技术。多块建模策略通过对整个工业过程建立多个子块检测模型，能够及时有效地检测规模庞大和具有复杂工况的工业过程，相较于全局模型具有一定的优势。本文在多块建模策略下，基于 k 近邻(k-Nearest Neighbor, kNN)算法进行故障检测方法的改进研究，主要内容如下：

（1）针对传统基于 kNN 的故障检测方法不考虑过程的局部信息、只建立一个全局模型而导致报警率低的问题，提出一种基于互信息的多块 kNN 故障检测方法。本方法利用变量间的互信息进行子块构建来提取过程的局部信息，使得子块内的变量拥有更多相同的信息。在此基础上，对每个变量子块建立基于 kNN 的故障检测模型，并利用贝叶斯推断方法将各子块的检测结果融合，使得整体的检测效果更为直观。进一步采用基于马氏距离的故障诊断方法，通过计算样本中各变量与其均值的马氏距离，找出引发故障的源变量并对其隔离。最后通过田纳西-伊斯曼(Tennessee-Eastman, TE)过程仿真实验，验证了该方法较传统故障检测方法具有更高的报警率。

（2）针对基于变量分块的故障检测方法对微小偏移、脉冲振荡等故障报警率低的问题，提出一种基于双层信息提取的多块 kNN 故障检测方法。第一层利用典型相关分析计算变量间的相关系数构建变量子块以获取局部信息；第二层对各变量子块分别提取观测信息、累计信息和变化率信息数据作为信息子块，通过提取特征信息放大故障差异。同时考虑到数据的分布特征和局部区域样本的稀疏程度，对各信息子块建立基于马氏距离的 kNN 故障检测模型，并将所有子块的检测结果通过贝叶斯推断方法进行融合，整合各子块的优势，提升了整体的故障检测性能。

（3）针对传统基于 kNN 的故障检测方法中引发故障的异常信息易被正常信息淹没，导致故障检测不及时和报警率低的问题，利用自编码器和多块建模策略提出一种基于重构误差的 kNN 故障检测方法。本方法利用正常工况数据集训练自编码器模型，基于该模型进行重构误差提取以解决异常信息易被淹没的问题。进一步考虑微小偏移和振荡等故障特征，采用多块建模策略，对各子块分别计算统计量并融合检测。通过数值仿真与 TE 过程仿真实验进行分析，结果验证了所提方法的可行性和优越性。

关键词：故障检测；k 近邻；信息提取；多块模型；重构误差

Abstract

Modern industrial processes are becoming increasingly complex and fault types are increasing. Once a fault occurs, it will not only reduce economic benefits, but also cause serious casualties. With the development of computer technology and data acquisition equipment, some industrial processes have accumulated rich process data, which makes the data-driven fault detection technology continue to progress and become an important key technology to ensure the safe operation of the industry. Multi-block modeling strategy can effectively monitor large-scale and complex industrial processes by establishing multiple sub-block detection models for the whole industrial process, which has certain advantages compared with the global model. In this paper, the improvement of fault detection method based on k-Nearest Neighbor algorithm is studied under multi-block modeling strategy. The main contents are as follows :

(1) Aiming at the problem that the traditional kNN-based fault detection method does not consider the local information of the process and only establishes a global model, resulting in the low alarm rate, a multi-block kNN fault detection method based on mutual information is proposed. This method uses mutual information between variables to construct sub-blocks to extract local information of the process, so that the variables in the sub-blocks have more same information. On this basis, the fault detection model based on kNN is established for each variable sub-block, and the Bayesian inference method is used to integrate the detection results of each sub-block, so that the overall detection effect is more intuitive. Further, the fault diagnosis method based on Mahalanobis distance is adopted. By calculating the Mahalanobis distance between each variable and its mean value in the sample, the source variable causing the fault is found and isolated. Finally, the Tennessee-Eastman (TE) process simulation experiment verifies that this method has higher alarm rate than the traditional fault detection method.

(2) Aiming at the problem that the fault detection method based on variable block has a low alarm rate for faults such as small offset and pulse oscillation, a multi-block kNN fault detection method based on two-layer information extraction is proposed. The first layer uses canonical correlation analysis to calculate the correlation coefficient between variables to construct variable sub-blocks to obtain local information; the second layer extracts observation information, cumulative information and the rate of change information data from each variable sub-block as information sub-blocks, and amplify fault differences by extracting feature information. Considering the distribution characteristics of data and the sparse degree of local area samples, the kNN fault detection model based on Mahalanobis distance is established for each information sub-block, and the detection results of all sub-blocks are fused by Bayesian inference method to integrate the advantages of each sub-block and improve the overall detection performance.

(3) Aiming at the problem that abnormal information causing faults in the traditional kNN-based fault detection method is easily overwhelmed by normal information, resulting in untimely fault detection and a low alarm rate, a kNN fault detection method based on reconstruction error is proposed by using the auto-encoder and multi-block modeling strategy.

In this method, the auto-encoder model is trained using the normal working condition data set, and the reconstruction error is extracted based on this model to solve the problem that the abnormal information is easily submerged. Further considering the fault characteristics such as small offset and oscillation, the multi-block modeling strategy is adopted to calculate the statistics of each sub-block and fuse the detection. The feasibility and superiority of the proposed method are verified by numerical simulation and TE process simulation experiments.

Keywords: Fault detection; k-nearest neighbor; information extraction; multi-block model; reconstruction error

目 录

第一章 绪论

1.1 课题研究背景及意义

随着工厂智能化水平的大幅度提升，现代工业过程逐渐规模化和自动化，创造更多经济效益的同时，对生产设备的性能和系统的状态提出了越来越高的要求。工业生产过程中若使用存在安全隐患或未达标的容器、管道等生产设备，其产生的偏差将不可避免地导致整个控制系统的性能恶化，甚至严重威胁工作人员的生命安全[1]。在近几年的工业生产中，因未对工业过程进行故障检测而导致严重事故的案例屡见不鲜。2019 年 4 月内蒙古东兴化工公司因未对氯乙烯气柜进行故障检测与诊断，使得氯乙烯气体泄漏产生特大爆炸，造成 40 人受伤，直接经济损失四千万元。2020 年 8 月湖北省仙桃市蓝化有机硅有限公司在清理分层积液时，由于缺乏温度检测和故障检测步骤，造成静置槽内丁酮肟盐酸盐发生分解爆炸，导致 6 人死亡。2021 年 4 月河北省赤城县安宸公司在进行民用爆炸物品销毁作业时，由于未及时检测物品的挤压变形情况导致爆炸物爆炸，导致 9 人死亡，直接经济损失 1614 万元。这些严峻的安全事故证明了故障检测技术是保证现代工业过程安全稳定运行的关键[2]，因此需要不断发展过程监控的技术研究以保证工业过程的安全性和可靠性。

早期的故障检测技术主要针对一些操作单元少、复杂度低的过程装置，依靠专家的知识或检测对象的机理模型进行决策和诊断[3]。然而随着现代工业智能化水平的不断提升，工业生产规模日益庞大且模型的复杂度越来越高，难以凭借系统的内部机理建立精确的数学模型。因而基于专家知识的检测方法应运而生，但是该方法受限于专家经验，具有一定的局限性，且建立一个权威的专家系统需要付出大量的人力和物力[4]。此外，由于计算机技术和传感器技术的迅猛发展，大量过程数据被收集和存储，使得设备或系统数据得到进一步的分析和利用。利用数据进行故障检测，不需要精准的机理模型和专家系统，只需要采集足够规模的离线数据，利用多元统计方法快速处理并分析数据间的有效信息，从而对过程进行在线检测。但是，由于工业数据之间存在非线性、非高斯、数据缺失等问题，使得传统的故障检测方法愈加受到挑战，无法满足检测需求。因此，考虑到实际工业过程的多样化和复杂化，建立合适的故障检测模型对传统故障检测方法进行优化和改进，具有重要的意义。

1.2 工业过程检测技术的研究现状

1.2.1 工业过程检测概述

工业过程故障检测技术是保证工业安全运行的关键技术，故障的准确识别与定位可以使操作人员及时发现故障并快速采取适当措施来避免系统产生不可控情况。所谓故障，是指过程中一个或多个变量发生了偏移正常工况范围的情况，导致系统无法完成预期成果，影响产品质量和运行安全。目前，工业过程检测技术根据处理顺序一般可分为四个部分，分别为：故障检测、故障识别、故障诊断和过程修复[5]，如图 1-1 所示。故障检

测是整个流程的关键环节，是研究的重点内容。

(1)故障检测：为准确判断系统是否发生故障而采取的一系列手段和措施，若检测到故障，及时报警。目前任何的故障检测系统都无法做到完全准确的检测，因此提高系统的故障检出率(即报警率)是改进故障检测技术的核心内容。

(2)故障识别：检测到故障后，及时确定导致故障产生的一个或多个变量及其所在位置，以便操作人员找到故障并对故障进行分离操作。

(3)故障诊断：完成检测后，判断引发故障的源变量，并对故障类型、时间、成因等情况进行分析，以便准确分离故障。

(4)过程恢复：为修复过程而采取的一系列补救措施，最大化降低故障对系统的影响，使系统尽快恢复到正常运行状态，并继续检测过程。

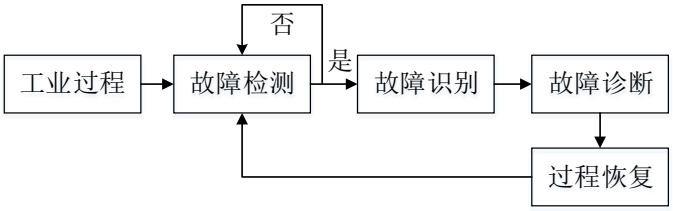


图 1-1 工业过程检测技术理论组成部分

20 世纪 70 年代初，美国国家宇航局创立故障预防小组，奠定了故障检测技术的基础。得益于计算机信息技术的发展，工业过程的故障检测技术已在多个领域取得突破性进展。通过对权威专家 P.M.Frank 教授[6]和 Venkat Subramanian 教授[7]的研究成果进行归纳，将现有的故障检测方法分为定量分析法和定性分析法[8]，定量分析法主要包括基于机理模型方法和基于数据驱动方法两类，具体关系如图 1-2 所示。

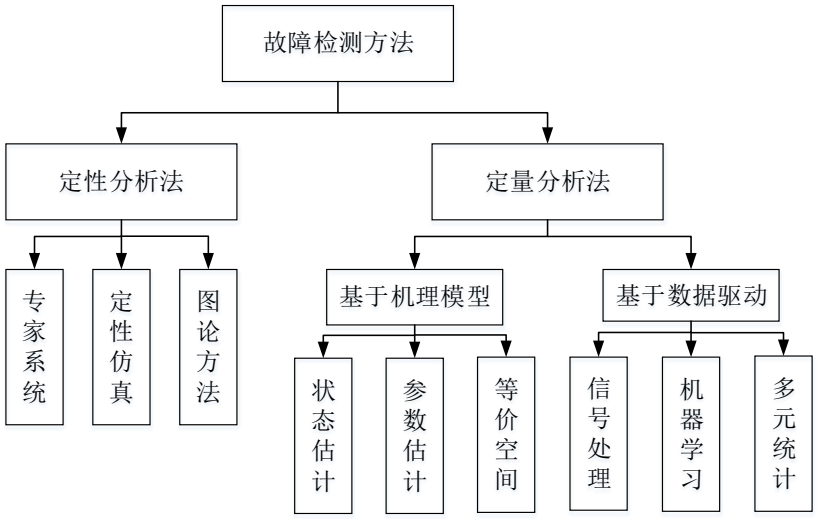


图 1-2 故障检测方法分类框图

1.2.2 基于机理模型的方法

基于机理模型的检测方法是一种定量分析方法，即针对待测对象的内部机理关系建立数学模型。通过分析比较采集的信息与数学模型之间的残差来实现故障检测。基于机理模型的检测方法主要包括参数估计法[9]、状态估计法[10]、等价空间法[11]和分析冗余法[12]等。但是随着工业系统不断复杂化和生产结构的多层次化，操作人员无法精准掌握检

测对象内部的机理结构，仅利用物理和化学原理无法很好地表现工业过程的特征和状态，导致建立数学模型愈加困难。且基于机理模型的方法无法实时完成故障检测，限制了该方法在复杂工业过程的故障检测中得到广泛应用。

1.2.3 基于知识的方法

基于定性知识的方法无需建立精准的机理模型，主要通过人工经验和逻辑辩证推理，定性描述各个单元之间关系并进行融合整理，有效挖掘待测对象的故障信息。基于定性知识的方法主要包括专家系统[13]、有向图法[14]、故障树法[15]等。该方法适用于有大量人工经验的情况，依赖于专家或专业人员的知识及经验。但是专家容易受到心理、生理等因素的影响，做出与实际情况差异较大的分析。而建立一个专家知识库则需要投入大量的人力与物力，因此，基于定性知识的检测方法难以实现复杂系统的故障检测，具有一定的局限性。

1.2.4 基于数据驱动的方法

随着新型传感器和数据采集设备的迅速发展，工业过程收集并存储了大量的过程数据，使得多元统计过程监控方法得到了广泛发展。该方法不需要掌握过程精准的机理知识，也无需建立昂贵的专家知识库，只需要对采集到的过程数据进行处理、加工与分析，提取出具有代表性的特征信息，并以此建立相应的故障检测模型。基于数据驱动的检测方法主要包括信息处理、机器学习和多元统计方法。基于信息处理的方法包括小波变换、谱分析等，根据信号在不同频域的特征进行故障检测。常用的机器学习方法有神经网络、支持向量机等。多元统计方法则包括主元分析(Principal Component Analysis, PCA)[16]，偏最小二乘(Partial Least-Square, PLS)[17]、独立主元分析(Independent Component Analysis, ICA)[18]、k 近邻[19]、线性判别分析(Linear Discriminant Analysis, LDA)[20]等。当过程数据满足静态、高斯分布、线性分布等假设条件时，这些传统的多元统计方法可以取得较好的检测效果。但是在现代复杂的工业过程中，数据往往存在动态性、非线性、非高斯、缺失、多模态等特征，不可避免地降低了这些方法的检测性能。因此，如何根据所收集的数据的特征，建立合适且能够及时准确预警的检测模型，成为学者们广泛研究的重点。

本文在多块模型的框架下，以经典的 kNN 方法为基础，对基于数据驱动的故障检测方法开展改进研究，并与其他多元统计方法进行分析比较。

（1）过程呈现非线性问题

对于现代工业过程和控制系统而言，非线性是普遍存在的特性。为提取系统的非线性的特征，一般使用线性模型去拟合非线性结构，其中核方法是解决非线性问题最常用的方法。文献[21]借助于神经网络的非线性逼近能力，提出一种基于神经网络的 PCA 检测方法，但是该方法耗时大且复杂。文献[22]提出基于核主元分析(Kernel Principal Component Analysis, KPCA)故障检测方法，利用核函数的非线性映射，将核方法与 PCA 方法结合，有效提高了传统 PCA 模型对非线性数据的检测效率。由于核方法的非线性映射效果显著，文献[23]提出了核独立元分析(Kernel Independent Component Analysis, KICA)，文献[24]提出了核偏最小二乘方法以解决非线性问题。文献[25]考虑到非线性的

非平稳过程，利用数据块和可变窗口快速更新 KPCA 模型，有效提高模型检测效率。文献[26]为识别非线性故障中的微小不易检出故障，通过核主元分析方法计算得分向量与特征值之间的残差并加权样本距离，放大了微小故障从而提高检出率。文献[27]通过比较主元统计量与残差统计量之间的相关性，提出一种改进的 KPCA 方法解决核主元分析方法无法识别故障变量的问题，相比于原始 KPCA，故障检测效率更高。文献[28]利用 PCA 将非线性空间分解构造成多个线性子空间，有效地对非线性过程进行故障检测。

（2）过程的动态性问题

在实际过程控制系统中，变量通常会在噪声和随机干扰的作用下，发生动态变化过程，传统的检测方法大多只针对静态过程，无法达到理想的检测效果。文献[29]提出一种动态独立主元分析算法，通过构造时延样本的动态矩阵，建立当前时刻样本与过去时刻样本之间的联系，从而捕捉动态特征。文献[30]使用部分动态主元分析方法处理过程的动态性，从而提升检测性能。文献[31]将核方法引入动态主元分析，提出一种基于动态核主元分析的故障检测算法，以解决动态过程中的非线性问题。文献[32]考虑到动态过程中质量变量的相关特征信息，提出了一种有监督形式的线性动态系统，相较于传统方法，更容易描述过程的动态特征。针对动态非线性系统故障检测率低的问题，文献[33]提出一种改进的动态核主元分析故障检测方法，该方法通过提取数据变量的时序特征，能够间接实现非线性映射。文献[34]将滑动窗口技术与局部离群因子算法相结合，提出了一种新的动态多向局部离群因子算法，实现工业过程的动态检测。文献[35]将动态 PCA 和 kNN 相结合，先建立主元模型，再利用 kNN 获取样本的 k 个近邻，明显提高了故障的报警率。文献[36]针对 kNN 模型不能及时更新的问题，采用基于样本距离的更新规则自适应更新检测模型，有效提高模型的实时检测能力。

（3）数据的非高斯性问题

在实际化工过程中大部分数据不满足高斯分布的假设，由于传统 PCA、PLS 方法检测指标控制限的确定需要数据满足高斯分布假设，因此它们无法很好地检测到故障。文献[37]最先提出使用独立主元分析(Independent Component Analysis, ICA)来处理非高斯过程的故障检测问题，ICA 能够分解数据并消除数据之间的独立性，提取出高斯特征。文献[38]首先使用 ICA 提取非高斯信息，再对 ICA 的残差空间使用 PCA 方法，ICA 与 PCA 的结合很大程度上提高了非高斯数据的故障检出率。文献[39]提出一种基于 KICA 和高斯混合模型的故障检测方法，采用 KICA 提取独立元并计算独立元监控计量均值来判断非高斯性，从而选择不同的检测方法进行单独监控。文献[40]提出了一种基于集成学习和贝叶斯推断的 ICA 方法，解决了传统 ICA 方法性能不稳定、独立元数量选择不准确问题，对 ICA 方法进行了有效地改进。文献[41]采用 Jarque-Bera 检测方法并利用变量间的 Hellinger 距离获得高斯和非高斯子块，然后分别采用 ICA 和 PCA 方法进行建模，并加权计算每个子块的统计量得到一个最终联合指标，改善了模型的监控能力。

尽管上述所提方法在一定程度上解决了数据的非线性、非高斯、动态性等问题，从而提升传统多元统计方法的检测性能，但是当其中的一些问题同时存在时，它们便难以取得理想的检测效果。为此，文献[42]提出了一种基于 kNN 规则的故障检测算法(Fault

Detection Method based on k Nearest Neighbor Rule, FD-kNN)，该算法核心思想是待分类样本的类别由其近邻的 k 个样本投票决定，对非线性、非高斯和多模态过程有很好的检测效果。FD-kNN 的故障检测依据是异常工况样本相较于正常工况样本会产生明显大于正常工况范围的偏移量，通过比较正常样本与故障样本在训练集中的前 k 个最近邻样本距离平方和判断是否发生故障。为提高 kNN 故障检测方法的效率和精度，并针对传统的 kNN 故障检测算法计算量比较大，近邻样本 k 值不容易确定等问题，学者们进行了大量的改进研究。文献[43]利用局部保持投影方法将高维数据投影到低维空间，然后根据样本的近邻权重计算权重统计量实现故障检测，不仅提高了检测率还降低了计算复杂度；文献[44]针对多模态数据的方差差异，运用局部相对概率密度对数据进行预处理，实现了 kNN 方法对多模态数据的有效检测；文献[45]通过深层网络结构将原始数据空间转换为不同的特征子空间，并在数据空间中分别建立 kNN 模型，解决了 kNN 模型难以有效利用高阶数据信息的问题。

1.3 基于多块建模策略的故障检测方法研究现状

1.3.1 多块建模策略的研究内容

传统的故障检测方法大多采用全局建模策略，但是由于现代工业结构复杂度提升，操作单元的数量不断增加以及变量间相关关系不断多样化，全局建模策略容易忽略局部信息，导致模型的监控能力不佳。而近年来提出的局部、多块等建模策略的优势得到了充分发挥[46]，多块建模策略能更加充分地利用过程变量之间的相关关系，提取出过程的局部特征，最大化利用数据的信息，并降低监控的复杂度。采用多块建模策略的故障检测方法主要对子块划分、子块检测模型建立和决策融合各子块检测结果这三部分内容进行研究。

（1）子块划分方法

根据过程变量的不同特性将变量划分成适当的若干子块是多块建模流程的第一个步骤，也是多块建模的关键研究内容。相较于全局模型，通过划分子块可以将全局的计算量分配到不同的子块中，有效减小各子块模型的计算负担，节省了运行时间。同时，多块模型框架更具有变通性，若在建模框架中改动部分单元，后续的数学模型配置和更新相比于传统的全局模型更加简易。此外，对过程进行子块划分可以提高模型的容错能力，将故障缩小到一个特定的子块范围，一个或多个子块的异常情况几乎不会影响其他子块的检测模型。目前常用的方法是根据专家知识和机理模型来划分子块，分块原则并不固定。基于变量关系的划分方法仅需要正常工况数据而不依赖于故障信息，同时能够放大某些变量的特征信息以更好建立检测模型。基于数据的划分方法不包含对故障信息的划分，具有较强的随机性且缺少一定的解释性和合理性，因此如何划分子块仍然是多块建模故障检测方法的重点研究内容。

（2）子块模型的检测方法

在划分好子块的基础上，对各子块分别建立故障检测模型，选取不同的建模方法将很大程度上影响最终决策融合的检测结果。常用的子块建模方法是一些经典的建模方法，

比如 PCA、PLS、kNN 等。当工业过程复杂度高、数据规模大时，通过划分子块，根据子块数据的不同特征信息，建立合适的检测模型。这样不仅充分利用各子块的数据特性，还能最大程度地挖掘数据信息，有效提高整体模型的检测精度。

（3）子块检测结果的决策融合方法

建立各子块检测模型后，由于子块数目不唯一且产生多个检测结果，无法得到一个直观的最终决策，因此，需要提出一种决策融合方法将多个结果融合成新的检测指标。决策融合方法能将多种不同类型的数据信息进行融合并作出决策判断，包括视频、图像、声音等信息。为提升检测系统的稳健性，需要集成各种类型的数据模型来达到复杂工业过程检测的需求。在决策融合时，需要考虑各子块的差异性以及不同类型的故障。

1.3.2 多块建模策略的应用

（1）面对多模态过程的故障检测

在现代复杂工业中，由于受到设备老化、气候变更、技术更新等因素的影响，工业生产中往往存在多个稳定运行的模态或正确的操作工况，若只建立单一的全局模型对多模态过程进行监控会产生较高漏报率的情况。使用多块建模策略将很好地解决多模态过程的故障检测问题，比如利用聚类方法将原始数据集划分为多个样本子集，每个子集代表不同模态，再分别建立相应的故障检测模型。文献[47]针对多模态数据容易因聚类不当而降低检测性能的问题，利用贝叶斯学习将多模态数据集聚类，提出一种基于贝叶斯 ICA 算法的多工况过程故障检测方法，建立不同模态的 ICA 模型。文献[48]考虑到不同模态切换时稳定特性的过渡，利用过渡数据差分得到变量相对变化信息，提出一种差分分段主元分析多模态过程故障检测方法，确保了模态间的平稳过渡。文献[49]分析数据集中样本之间的邻域信息，采用局部邻域标准化策略，将不同工况的数据同步至同一尺度下，对多模态过程建立了统一模型。文献[50]结合传统 PCA 和高斯混合方法，并通过构建稀疏编码和局部保持投影的混合框架来提取局部特征，结合传统 PCA 方法和高斯混合模型，改善了多模态过程中的非高斯问题。文献[51]考虑到多模态过程数据具有多中心、方差差异大等特点，通过构造标准距离，实现了 kNN 方法对多模态数据的有效检测。

（2）面向厂级的故障检测

近年来鉴于现代工业的快速发展，厂级过程的控制与检测获得越来越多的关注。厂级过程一般具有多个不同的操作单元，借助于发达的工业信息化水平，不同操作单元的流水线地理位置不再受到限制。传统的单模型多元统计方法无法很好地完成监控任务。文献[52]首次提出使用多区块投影法，为整个流程建立多区块监控图表，并建立偏最小二乘故障检测模型。文献[53]提出了一种用于厂级过程监控的分布式主元分析检测模型，利用过程变量在主元方向上的贡献度划分子块。文献[54]采用故障相关变量构建子块，并结合贝叶斯推理实现故障检测。文献[55]采用了多块建模策略，使用互信息将不同的变量划分到各个子块中，并将其扩展到多模态监控问题。文献[28]利用主元分析方法将非线性空间构造成线性子空间，实现 PCA 对非线性的检测。文献[56]根据 Hellinger 距离将概率分布相似的变量划分在一个子块中，然后用独立主成分分析方法对每个子块建

立故障检测模型。文献[57]在子块构建后对每个子块分别进行相对变换独立主元分析处理，实现故障排查和识别。

1.4 论文的主要研究内容

本论文的主要工作是多块建模策略下的 kNN 故障检测方法研究。考虑到过程数据的非线性和非高斯等特性，以经典的 kNN 方法为基础开展研究。同时针对全局模型未能利用局部信息和特征信息造成信息资源浪费，以及引发故障的异常信息易被正常信息淹没等问题，采用多块建模策略展开进一步的故障检测方法研究。本文在结构上共分为五个章节，各章节的内容安排如下：

第一章为绪论，首先阐述了本文所涉及课题的研究背景和意义，提出了工业过程检测的概念；随后描述了工业过程检测技术的研究现状并重点介绍了基于数据驱动的故障检测方法；最后对基于多块建模策略的故障检测方法和应用进行详细介绍。

第二章首先介绍了 kNN 的基本原理及基于 kNN 的故障检测方法，随后针对传统基于 kNN 的故障检测方法不考虑过程的局部信息，只建立一个全局模型，提出一种基于互信息的 kNN 故障检测方法。首先对过程变量进行互信息计算，根据互信息值的大小将变量分成多个子块；然后对每个子块建立 kNN 故障检测模型，并利用核密度估计方法求出控制限；最后利用贝叶斯推断方法整合各子块的统计量，使得整体的检测效果更为直观。并进一步采用基于马氏距离的故障诊断方法，通过计算样本中各变量与其均值的马氏距离，找出引发故障的源变量并对其隔离。利用 TE 过程数据对所提方法进行了仿真，验证了所提方法的性能。

第三章进一步从变量子块的特征信息考虑，提出了一种基于双层信息提取的多块 kNN 故障检测方法，提升了 kNN 检测方法对多种不易检出类型的故障的检测性能。第一层，利用典型相关分析计算变量间的相关系数，将相关性高的多个变量放在一起组成子块，从而提取过程的局部信息；第二层，对各个变量子块提取观测信息、累计信息和变化率信息数据，利用特征信息数据构建多个信息子块，进一步挖掘了子块的有效信息。同时考虑到数据的分布特征和局部区域样本的稀疏程度，采用基于马氏距离的 kNN 故障检测方法进行检测；最后利用贝叶斯方法进行融合检测。通过 TE 过程和实际高炉炼铁过程的仿真实验，验证了所提方法的有效性。

第四章针对基于 kNN 的故障检测方法中，引发故障的异常信息易被正常信息淹没，导致故障检测不及时和报警率低的问题，利用自编码器和多块建模策略提出一种基于重构误差的 kNN 故障检测方法。为了解决统计量计算过程中异常信息易被淹没的问题，先利用自编码器模型进行正常工况数据的重构还原，再基于该模型求取异常工况数据的重构误差，抽离出异常信息数据；并将该重构误差作为观测信息，进一步提取累计信息和变化率信息构建 3 个信息子块，对每个信息子块建立相应的 kNN 模型，利用核密度估计方法确定各子块中的控制限。最后利用贝叶斯方法将待测样本在各个子块上的检测结果融合。将所提方法进行数值仿真和 TE 过程仿真，均取得较高的故障报警率，表明了方法的优越性。

第五章为总结与展望，对课题的研究工作做了回顾和总结，并对基于多块建模策略的故障检测方法的发展趋势进行了展望。

第二章 基于互信息的多块 kNN 故障检测方法

随着新型传感器和数据采集设备的迅速发展，一些先进工业过程积累了丰富的过程数据，使得多元统计过程监控技术不断进步，其中 kNN 方法是最常用的方法之一，得到了广泛的应用。然而传统的过程监控方法大多只建立一个全局模型，由于现代工业结构复杂度提升，操作单元的数量不断增加以及变量间相关关系的多样化，全局建模策略无法准确地对过程建模，因此多块建模策略成为有效的解决方案。文献[41]通过分析过程变量的不同特性，将变量分成若干子块，并针对子块的不同特性采用不同的方法进行建模检测，其最终检测结果优于单一全局模型的检测效果。文献[28]利用主元分析法将非线性空间近似为多个线性子空间，对子空间分别进行检测并融合，实现了对非线性故障的检测。

为了更加充分地对复杂过程变量之间的关系进行描述，并提取过程的局部特征，利用多块建模策略，本章提出一种基于互信息的 kNN 故障检测方法。首先对过程变量进行互信息计算，根据互信息值的大小进行子块构建；然后对每个子块建立基于 kNN 的故障检测模型并确定控制限；最后利用贝叶斯推断方法将各子块的检测结果融合，得到一个最终的检测指标。并进一步采用基于马氏距离的故障诊断方法，通过计算样本中各变量与其均值的马氏距离，找出引发故障的源变量。利用 TE 过程数据进行了仿真实验，并与几种传统检测方法进行对比，验证了所提方法的有效性。

2.1 kNN 算法

2.1.1 kNN 算法的基本原理

在机器学习中，kNN 算法是一种被广泛使用的分类和回归方法[42]。kNN 算法，也叫做 k 近邻算法，根据近邻样本距离特征确定待测样本的相似度从而对其进行分类，即通过计算待测样本与训练集中的 k 个近邻样本的距离来确定待测样本所属类别[36]。kNN 算法的三个基本要素分别为距离度量、近邻个数 k 值的选择和分类决策规则。

（1）距离度量

在 kNN 算法中，以不同距离度量函数来确定样本之间的距离标准，常用的距离度量函数包括欧氏距离、曼哈顿距离、马氏距离等。在建模过程中，往往根据实际的样本分布特性和数据特征信息进行距离度量函数的选择，通常情况下选取欧氏距离作为距离计算指标。

（2）近邻个数 k 值的选择

在实际建模过程中，k 值大小的选择对模型的预测结果产生重要作用。若选择的近邻个数 k 值过小，则训练模型的领域样本数量过少而导致模型的预测精度偏低；当 k 值选取过大时，则容易产生欠拟合的现象。因此，在实际的应用中，通过对模型结果的对比分析进行参数调整。在多数情况下，通过网格搜索、交叉验证等方法可以选取到合适的 k 值。

（3）分类决策规则

寻找到待测样本的 k 个最近邻后，根据少数服从多数的投票法则，将待测样本与最邻近样本中所属类别占比最多的归为一类。

2.1.2 基于 kNN 的故障检测方法

基于 kNN 的故障检测方法，其基本思想是通过计算样本点与近邻距离来度量样本间的相似度，在故障检测过程中，若样本点与训练集中前 k 个样本的欧式距离平方和大于正常样本的相应距离，则判断该样本点为故障点，反之正常。检测过程包括两部分，分别为建立模型和故障检测。

步骤一：建立模型

首先在训练集样本中，寻找样本 x_i 的 k 个最近邻样本，记做 $A(x_i,k)=\{x_i^1,x_i^2,\cdots,x_i^j,\cdots x_i^k\}$，其中 x_i^j 表示样本 x_i 的第 j 个近邻样本。然后计算每个样本 x_i 与其 k 个近邻样本的欧氏距离平方和作为统计量，如公式(2.1)所示。其中 d_{ij}^2 表示样本 x_i 与它的第 j 个近邻样本的欧氏距离平方。接着，结合核密度估计方法(Kernel Density Estimation, KDE)和置信度 α 确定检测模型的控制限 D_α^2。

$$D_i^2=\sum_{j=1}^{k}d_{ij}^2,\quad d_{ij}^2=\left\|x_i-x_i^j\right\|^2 \tag{2.1}$$

其中，D_i^2 表示样本 x_i 的统计量。

步骤二：在线故障检测

首先，在训练集样本中，确定待测样本 x 的前 k 个近邻样本。然后，计算 x 与其 k 个近邻样本的欧式距离平方和作为统计量，记做 D_x^2。最后，比较 D_x^2 与 D_α^2 的大小，若 $D_x^2\geq D_\alpha^2$，则判定待测样本 x 为故障点，反之为正常点。

2.2 基于互信息的多块建模 kNN 故障检测及诊断

2.2.1 基于互信息的多块策略及阈值确定

在实际的工业过程中，变量之间大多是线性、非线性共存，高斯、非高斯混合分布，传统的 kNN 方法从全局的角度出发，系统的本质特征无法得到充分的展示。在概率论领域，互信息是一种相对成熟的统计分析技术，可以通过信息熵度量两个随机变量之间的依赖性，表示出两个变量共享的信息，反映两个变量的相关性[58]。因此，对过程变量进行互信息计算，将互信息值大的多个变量放在一起组成子块，使得子块内的变量拥有更多相同的信息，最大化地反映变量的一个或者多个局部特征。令随机变量 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的联合概率分布及边缘概率分布分别为 $p(x,y)$，$p(x)$ 和 $p(y)$，其中 $x\in\boldsymbol{X}$，$y\in\boldsymbol{Y}$，$\boldsymbol{X}$ 的熵定义如式(2.2)所示，联合熵如式(2.3)所示。

$$H(\boldsymbol{X})=-\sum_{x\in\boldsymbol{X}}p(x)\log p(x) \tag{2.2}$$

$$H(\boldsymbol{X},\boldsymbol{Y})=-\sum_{x\in\boldsymbol{X}}\sum_{y\in\boldsymbol{Y}}p(x,y)\log p(x,y) \tag{2.3}$$

则变量 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 之间的互信息计算公式如式(2.4)所示。

$$I(\boldsymbol{X};\boldsymbol{Y})=\sum_{y\in\boldsymbol{Y}}\sum_{x\in\boldsymbol{X}}p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \tag{2.4}$$

若变量 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 完全独立，则 $\boldsymbol{X}$ 不对 $\boldsymbol{Y}$ 提供任何信息，此时互信息值为零；反之，两个变量的相关性越高，互信息值越大。

对于训练集 $\boldsymbol{X}_{train}\in\mathbb{R}^{n\times m}$，$x_i\in\boldsymbol{X}_{train}$，$x_j\in\boldsymbol{X}_{train}$，计算变量 x_i 与变量 x_j 之间的互信息 I_{ij}，即

$$I_{ij}=I(x_i,x_j)(i=1,2,\ldots m;\ j=1,2,\ldots m) \tag{2.5}$$

确定阈值 I_{il}，若 $I_{ij}\geq I_{il}$，则把变量 x_j 与变量 x_i 放到相同的子块中。其中，I_{il} 一般根据经验获得，本章结合互信息针状图为了更好地划分变量，I_{il} 取 $1.3I_{iM}$，其中 I_{iM} 是 I_{ij} 的中值。基于互信息的多块建模方法如图 2-1 所示。

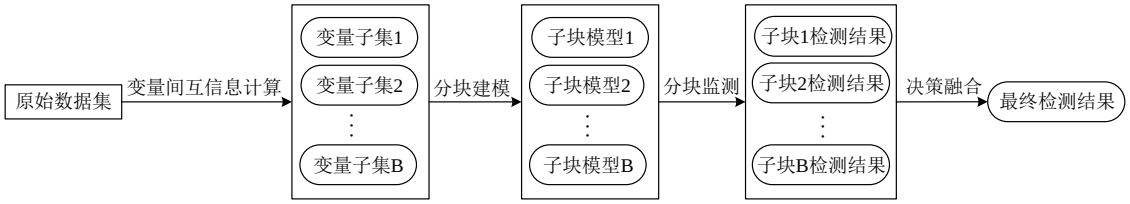


图 2-1 基于互信息的多块模型建立步骤

2.2.2 故障在线检测及诊断过程

首先对变量进行互信息的计算，利用阈值将变量划分成各子块。针对划分好的子块，建立 kNN 故障检测模型。寻找各子块中样本的 k 个近邻样本集，记做

$$T(x_{ib},k)=\{x_{ib}^1,x_{ib}^2,\ldots,x_{ib}^k\}(b=1,2,\ldots,B) \tag{2.6}$$

其中 x_{ib} 表示第 b 个子块中的变量 x_i，x_{ib}^k 表示样本 x_{ib} 的第 k 个近邻样本。

然后计算每个子块中样本与其 k 个最近邻样本的距离平方和作为子块的统计量，即 $D_{ib}^2=\sum_{j=1}^{k}d_{ibj}^2$，其中 D_{ib}^2 表示第 b 个子块的统计量。通过核密度估计确定每个子块中统计量的控制限，并基于贝叶斯融合策略[59]，得到最终的检测结果。

假设测试样本 $x_{\text{test}}\in\boldsymbol{X}_{\text{test}}$，根据第 i 个子块的 D^2 统计量和控制限可得该样本在第 i 个子块中的故障条件概率为 $P_{D^2}(\text{F}\,|\,x_{\text{test}},i)$，如公式(2.7)所示。

$$P_{D^2}(\text{F}\,|\,x_{\text{test}},i)=\frac{P_{D^2}(x_{\text{test},i}\,|\,\text{F})P_{D^2}(\text{F})}{P_{D^2}(x_{\text{test},i}\,|\,\text{N})P_{D^2}(\text{N})+P_{D^2}(x_{\text{test},i}\,|\,\text{F})P_{D^2}(\text{F})} \tag{2.7}$$

其中，$x_{\text{test},i}$ 表示第 i 个子块中的测试样本。$P_{D^2}(\text{N})$ 和 $P_{D^2}(\text{F})$ 分别为正常样本和故障样本的先验概率，置信度为 α 和 $1-\alpha$。条件概率 $P_{D^2}(x_{\text{test},i}\,|\,\text{N})$ 和 $P_{D^2}(x_{\text{test},i}\,|\,\text{F})$ 如公式(2.8)定义。

$$\begin{aligned}P_{D^2}(x_{\text{test},i}\,|\,\text{N})&=e^{-D_{i,new}^2/D_{i,\text{lim}}^2}\\P_{D^2}(x_{\text{test},i}\,|\,\text{F})&=e^{-D_{i,\text{lim}}^2/D_{i,new}^2}\end{aligned} \tag{2.8}$$

其中，$D^2_{i,new}$表示新样本在第 i 个子块中的统计量；$D^2_{i,\lim}$表示第 i 个子块中由核密度估计方法得到的控制限。最终，以条件概率 $P_{D^2}(x_{\text{test},i}\,|\,\text{N})$ 和 $P_{D^2}(x_{\text{test},i}\,|\,\text{F})$ 为权重，加权计算各子块中待测样本 x_{test} 为故障样本的概率。融合后的 BIC_{D^2} 统计量即为待测样本 x_{test} 发生故障的概率，如公式(2.9)所示

$$BIC_{D^2}=\sum_{i=1}^{b}\left\{\frac{P_{D^2}(x_{\text{test},i}\,|\,\text{F})P_{D^2}(\text{F}\,|\,x_{\text{test},i})}{\sum_{j=1}^{3}P_{D^2}(x_{\text{test},j}\,|\,\text{F})}\right\} \tag{2.9}$$

融合后的统计量控制限为$1-\alpha$，其中，α 为先验知识，即生产过程中正常样本的先验概率，$1-\alpha$ 为异常样本的先验概率，本文取$1-\alpha$ 为 0.01。当 BIC 统计量大于$1-\alpha$ 时，则判断待测样本为故障样本，否则为正常样本。

当检测到故障后，需要找出引发故障的源变量并对其进行分离。计算数据样本中各变量与其均值的马氏距离[55]，即加权计算数据样本中各变量相较于其均值的偏移量，偏移量越大，说明该变量对于故障贡献越大。该方法可以有效辨识引发故障的源变量，即发生故障的根本原因。

综上所述，基于互信息的多块 kNN 故障检测方法(Multi-block kNN Fault Detection Method Based on Mutual Information, MI-MBkNN)流程如图 2-2 所示，具体步骤描述如下。

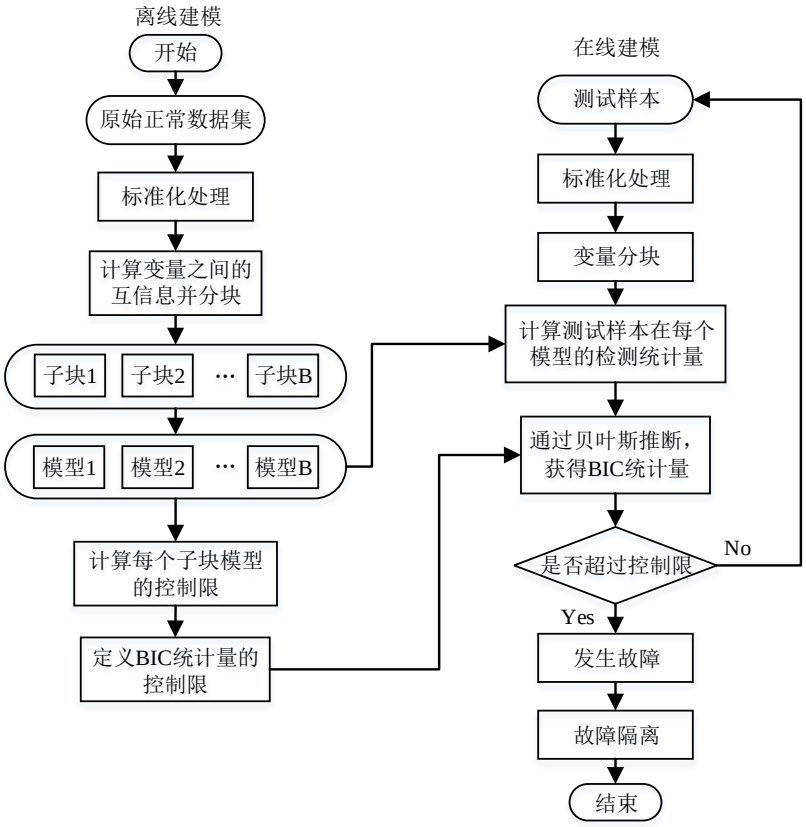


图 2-2 基于 MI-MBkNN 的故障检测流程

Step 1:获取正常训练数据，并对其进行标准化处理；

Step 2:计算两两变量间互信息，根据 2.1 节所述方法对变量进行分块，得到各个子

块；

Step 3:对每个子块分别建立 kNN 模型，利用核密度估计方法确定各自的故障控制限；

Step 4:对于待测样本，对其标准化处理和互信息计算进行变量子块构建，得到多个不同的子块；

Step 5:对各子块采用 kNN 故障检测方法，获得每个子块的检测结果；

Step 6:通过贝叶斯推断方法，将各个子块的统计量组合成为一个新的 BIC 统计量，并根据置信度确定控制限，当 BIC 超过控制限时则判断发生了故障，否则正常；

Step 7:检测到故障后计算数据样本中各变量与其均值的马氏距离，确定故障变量及故障块，分离出对故障影响最大的变量。

2.3 仿真实验

2.3.1 田纳西-伊斯曼过程介绍

TE 过程是一个基于实际化工过程的仿真模拟系统，它包括 5 个主体部分：冷凝器、汽提塔、压缩机、反应器以及分离器[60-61]。TE 过程的具体流程图如图 2-3 所示。

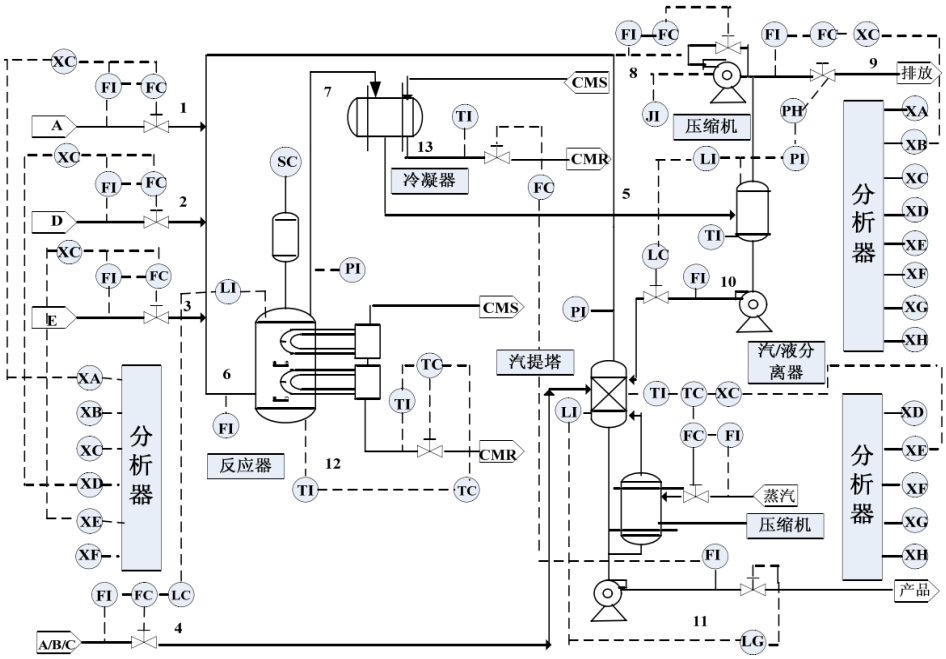


图 2-3 田纳西-伊斯曼过程流程图

TE 过程一共包含八种成分，分别为成分 A、B、C、D、E、F、G 和 H，其中 A、C、D、E 为气体进料，B 为惰性不可溶进料，G 和 H 为反应产物，F 为反应副产物。各反应器之间的会发生以下一系列化学反应：

$$\mathrm{A(g)+C(g)+D(g)\rightarrow G(liq)};$$

$$\mathrm{A(g)+C(g)+E(g)\rightarrow H(liq)};$$

$$\mathrm{A(g)+E(g)\rightarrow F(liq)};$$

$$\mathrm{3D(g)\rightarrow 2F(liq)};$$

该过程主要包含 41 个测量变量(如表 2-1 所示)，12 个操作变量(如表 2-2 所示)，并设定 21 种不同的故障(包括 16 种已知故障和 5 种未知故障，如表 2-3 所示)。本章选取 22 个过程测量变量和 11 个操作变量(不包括搅动速度)用于故障检测方法建模和性能测试。对于每种故障，训练集用于训练建立模型，测试集用来检验模型的故障检测性能。训练集和测试集均采用 960 个样本，测试集中故障从第 161 个样本点引入。

表 2-1 TE 过程测量变量

变量编号	描述	变量编号	描述
1	物料 A 流量(流 1)	22	分离器冷却水出口温度
2	物料 D 流量(流 2)	23	成分 A(流 6)
3	物料 D 流量(流 2)	24	成分 B (流 6)
4	总进料量(流 4)	25	成分 C (流 6)
5	再循环流量(流 8)	26	成分 D (流 6)
6	反应器进料流度(流 6)	27	成分 E (流 6)
7	产品反应器压力	28	成分 F (流 6)
8	产品反应器等级	29	成分 A (流 9)
9	产品反应器温度	30	成分 B (流 9)
10	排空速度(流 9)	31	成分 C (流 9)
11	分离器温度	32	成分 D (流 9)
12	分离器液位	33	成分 E (流 9)
13	分离器压力	34	成分 F (流 9)
14	分离器塔底流量(流 10)	35	成分 G (流 9)
15	汽提器等级	36	成分 H (流 9)
16	汽提器压力	37	成分 D (流 11)
17	汽提器塔底压力(流 11)	38	成分 E (流 11)
18	汽提器温度	39	成分 F (流 11)
19	汽提器流量	40	成分 G (流 11)
20	压缩机功率	41	成分 H (流 11)
21	反应器冷却水出口温度		

表 2-2 TE 过程操作变量

变量编号	描述	变量编号	描述
42	D 进料量(流 2)	48	分离器罐液流量(流 10)
43	E 进料量(流 3)	49	分离器液体产品流量(流 11)
44	A 进料量(流 1)	50	汽提器水流阀
45	总进料量(流 4)	51	反应器冷却水流量
46	压缩机再循环阀	52	冷凝器冷却水流量
47	排放阀(流 9)	53	搅拌器速度

表 2-3 TE 过程故障描述

故障编号	故障描述	故障类型
1	A/C 组分进料比发生变化，B 组分不变（流 4）	阶跃
2	B 成分发生变化，A/C 组分进料比不变（流 4）	阶跃
3	D 进料温度发生变化（流 2）	阶跃
4	反应器冷却水的入口温度发生变化	阶跃
5	冷凝器冷却水的入口温度发生变化	阶跃
6	A 组分进料损失（流 1）	阶跃
7	C 出现压力损失——可用性降低（流 4）	阶跃
8	A、B、C 组分进料变化（流 4）	随机
9	组分 D 进料温度发生变化（流 2）	随机
10	组分 C 进料温度发生变化（流 2）	随机
11	反应器冷却水的入口温度发生变化	随机
12	冷凝器冷却水的入口温度发生变化	随机
13	反应动态常数变化	缓慢漂移
14	反应器冷却水阀门发生变化	阀门黏滞
15	冷凝器冷却水阀门发生变化	阀门黏滞
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	阀门黏滞

2.3.2 故障检测性能指标

评估故障检测模型的性能好坏通常以两个数据指标为依据，分别为报警率(Fault Detection Rate, FDR)和误报率(False Alarm Rate, FAR)，报警率是指模型成功检测到的故障样本占所有故障样本的比例，误报率是指将正常样本误报为故障样本的比例。报警率和误报率的高低代表模型的性能好坏。它们的关系如混淆矩阵表 2-4 所示：

表 2-4 混淆矩阵

混淆矩阵		真实值	
		正常	异常
预测值	正常	TP	FP
	异常	FN	TN

$$FAR = \frac{FN}{TP + FN} \times 100\% \tag{2.10}$$

$$FDR = \frac{TN}{FP + TN} \times 100\% \tag{2.11}$$

2.3.3 检测结果与分析

为了建立多块模型，对选取的过程变量和操作变量进行互信息计算并分块，分块结

果如表 2-5 所示。图 2-4 分别展示了变量 18、变量 19、变量 31 与其他 32 个变量间的互信息，图中阈值虚线表示的数值大小为 1.3 倍互信息中值，该阈值根据经验和多次对比仿真实验的检测结果获得，互信息值超过虚线的变量即为与该变量具有较大互信息的变量。因此将变量 18、变量 19 和变量 31 放到相同的子块中。图 2-5 分别展示了变量 10、变量 17、变量 28、变量 33 与其他 32 个变量间的互信息，通过与阈值虚线的比较，将它们组成一个子块。

表 2-5 变量分块结果

子块编号	变量	子块编码	变量
1	x_1, x_{25}	5	$x_{10}, x_{17}, x_{28}, x_{33}$
2	x_{12}, x_{29}	6	$x_7, x_{13}, x_{16}, x_{20}, x_{27}$
3	x_{15}, x_{30}	7	$x_2, x_3, x_4, x_5, x_6, x_8, x_9, x_{11}, x_{14}, x_{21}, x_{22}, x_{23}, x_{24}, x_{26}, x_{32}$
4	x_{18}, x_{19}, x_{31}		

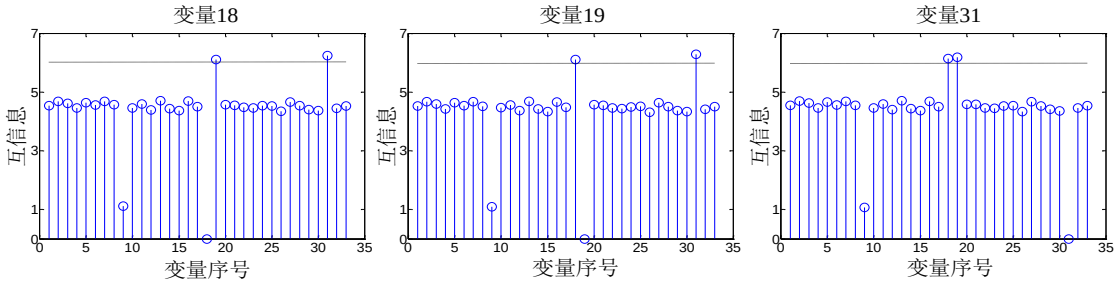


图 2-4 子块 4 中各变量间的互信息

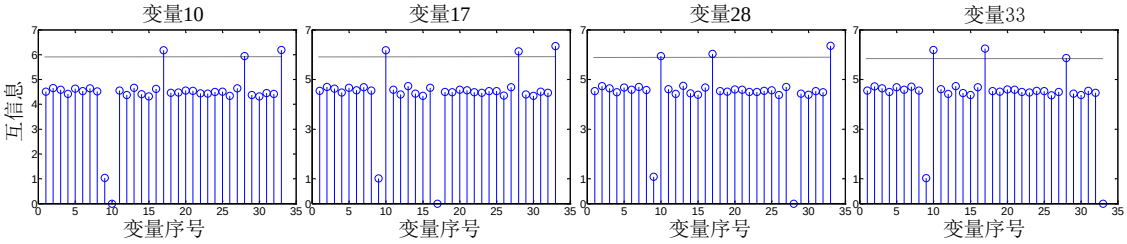


图 2-5 子块 5 中各变量间的互信息

表 2-6 给出了 7 个子块对 21 种故障的报警率和平均误报率。通过网格搜索算法，取近邻个数 k 为 14。从报警率来看，对于大多数故障类型，子块 7 的检测结果要优于其他六个子块。子块 5 的平均报警率很低，但是对于某些故障(如故障 5)，子块 5 的报警率达 96%，对整个的检测起到了关键的作用。对于不同的故障，由于某些子块拥有较高的报警率和较低的误报率，使得最终融合的 BIC 统计量表现了良好的检测性能。从对 21 种故障的检测结果来看，对于大部分故障，融合后的报警率有了明显的提高。

表 2-7 给出了 TE 过程 21 种故障在不同传统检测方法下的报警率和误报率，主要方法包括基于 PCA 的检测方法[53]、基于支持向量数据描述(Support Vector Data Description, SVDD)的检测方法[62]、基于 kNN 的检测方法、基于信息提取的 kNN 故障检测方法(kNN Fault Detection Based on Multi-block Information Extraction, MBIkNN)[63]和本章提出的 MI-

MBkNN。为方便比较，使用加粗字体表示所提方法的最优检测结果及相应的故障编号。其中各方法的控制限均采用核密度估计方法确定，阈值设置为 99%，通过网格搜索算法确定近邻 k 取值 14。从仿真结果可以看出，对于绝大多数故障类型，MI-MBkNN 能取得优越于其他三种方法的检测结果，尤其是对故障 5、故障 10、故障 16、故障 19 的检测。图 2-6 以故障 5 为例展示了详细的检测过程与结果。

表 2-6 TE 过程各故障报警率

故障编号	子块 1	子块 2	子块 3	子块 4	子块 5	子块 6	子块 7	BIC
1	0.989	0	0	0.97	0.103	0.373	0.98	**0.998**
2	0.054	0	0	0.898	0.986	0.729	0.951	**0.986**
3	0.003	0	0.001	0.003	0.008	0.029	0.011	**0.033**
4	0.003	0	0	0	0.008	0.019	1	0.994
5	0.090	0	0	0.196	0.96	0.23	0.194	0.951
6	1	0	0	0.966	0.944	0.996	0.99	**1**
7	0.208	0	0	0.311	0.153	0.415	1	**1**
8	0.580	0	0.003	0.72	0.721	0.936	0.839	**0.976**
9	0.003	0	0	0.001	0.009	0.023	0.013	0.021
10	0.079	0	0	0.715	0.004	0.348	0.141	**0.758**
11	0.005	0	0.001	0.02	0.011	0.046	0.735	0.660
12	0.446	0	0.004	0.879	0.42	0.916	0.968	**0.991**
13	0.448	0	0	0.933	0.574	0.875	0.89	**0.949**
14	0.005	0	0.001	0	0.003	0.004	1	**1**
15	0.013	0	0	0.034	0.003	0.074	0.01	**0.089**
16	0.019	0	0	0.801	0.005	0.18	0.07	**0.804**
17	0.010	0	0	0.073	0.011	0.103	0.943	0.900
18	0.828	0	0	0.881	0.859	0.88	0.896	**0.899**
19	0.008	0	0.001	0.011	0.004	0.541	0.105	0.468
20	0.018	0	0.003	0.05	0.008	0.651	0.114	0.630
21	0.003	0	0	0.49	0.003	0.376	0.334	0.449
平均故障报警率	0.229	0	0.001	0.426	0.276	0.416	0.58	**0.741**
平均故障误报率	0.002	0.001	0	0.002	0.005	0.017	0.003	0.017

表 2-7　几种传统故障检测方法性能比较

故障编号	PCA	SVDD	kNN	MBIkNN	MI-MBkNN
	报警率	报警率	报警率	报警率	报警率
1	0.999	0.993	0.996	0.998	0.998
2	0.983	0.984	0.983	0.986	**0.986**
3	0.025	0.037	0.013	0.033	0.033
4	1	0.793	0.975	0.994	0.995
5	0.244	0.275	0.260	0.951	**0.951**
6	1	1	1	1	**1**
7	1	1	1	1	**1**
8	0.968	0.975	0.976	0.976	**0.976**
9	0.017	0.03	0.020	0.021	**0.021**
10	0.299	0.448	0.418	0.758	**0.759**

故障编号	PCA	SVDD	kNN	MBIkNN	MI-MBkNN
11	0.759	0.599	0.683	0.869	0.660
12	0.986	0.985	0.989	0.993	**0.991**
13	0.954	0.945	0.946	0.951	0.949
14	1	1	1	0.998	**1**
15	0.031	0.062	0.029	0.106	**0.089**
16	0.274	0.283	0.289	0.549	**0.805**
17	0.954	0.878	0.919	0.966	0.901
18	0.902	0.897	0.896	0.899	0.899
19	0.126	0.047	0.099	0.134	**0.468**
20	0.497	0.458	0.495	0.643	**0.630**
21	0.476	0.419	0.425	0.566	0.449
平均报警率	0.643	0.624	0.639	0.701	**0.741**
平均误报率	0.004	0.007	0.006	0.015	0.017

TE 过程中的故障 5 涉及冷凝器冷却水入口温度的变化情况，使用传统的 PCA、kNN、SVDD 检测方法和本章提出的 MI-MBkNN 的对故障 5 的检测结果如图 2-6 所示。从图 2-6(a)~(c)可以发现，在故障开始时就可以检测出故障，但是在大约第 350 个样本处，统计量出现低于控制限的情况，导致故障的漏报。由于该故障是局部故障，因此很难在全局模型中检测到，为了更好的找出故障的原因，图 2-7 给出了数据样本在第 161 个样本点(故障最开始处)的各变量与其均值中心的马氏距离。图 2-7(a)给出了基于 kNN 的检测结果，图 2-7(b)给出了 MI-MBkNN 的结果。可以看出这两个模型都能正确识别变量在

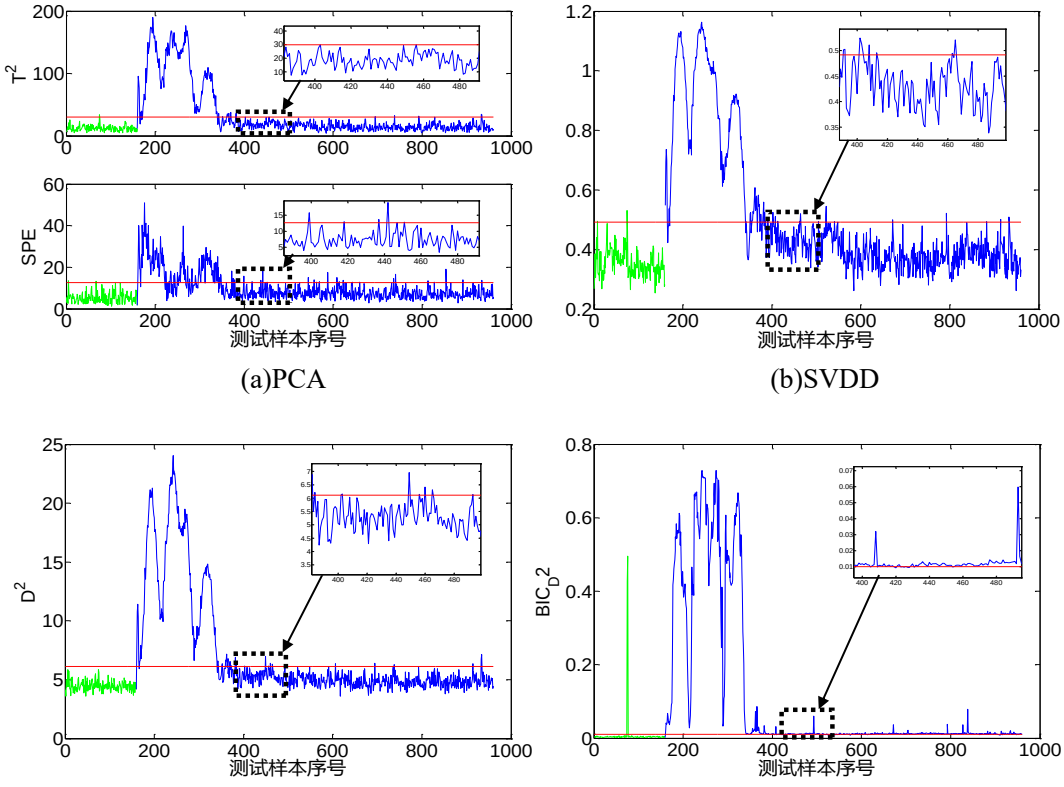
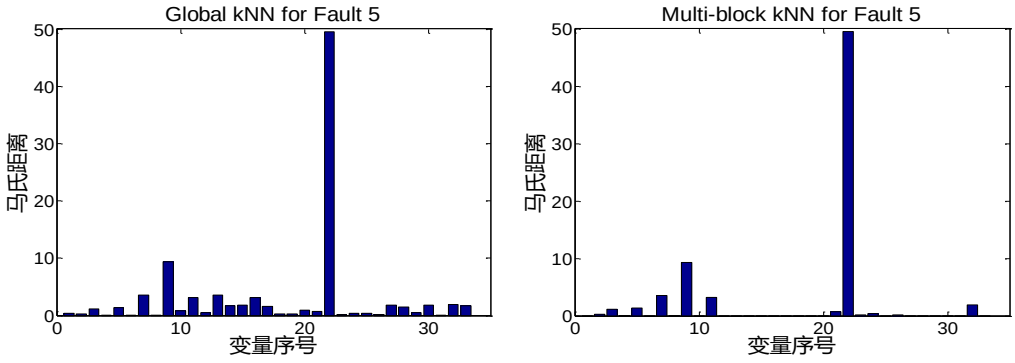


图 2-6 TE 过程故障 5 检测结果

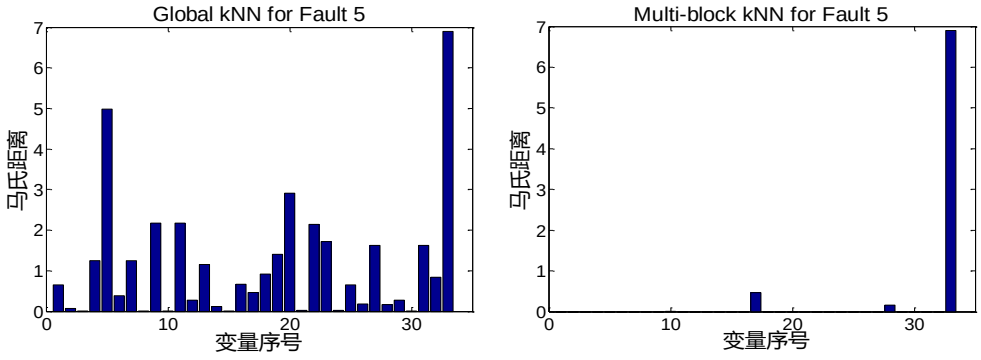


(a) (b)

图 2-7 第 161 样本点故障 5 的变量识别结果：(a)kNN 和(b)MI-MBkNN

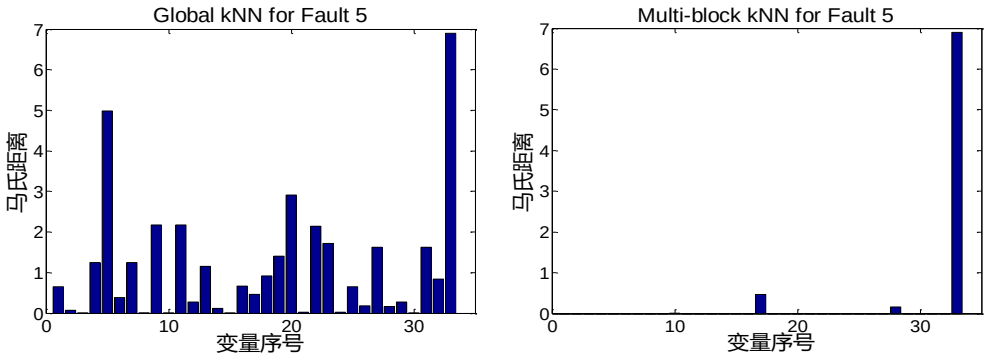
过程中的变化，如分离器冷却水出口温度的变化(变量 22)，反应器温度(变量 9)，产品分离器温度(变量 11)和反应器冷却水流量(变量 32)。但是在 350 个样本点后，从图 2-8(第 400 个样本点)可以看出，kNN 无法识别出冷凝器冷却水流量的变化(变量 33)，但是 MI-MBkNN 模型可以成功识别，因此 MI-MBkNN 对故障 5 表现出了优越的检测效果。

故障 10 是 C 进料中温度的随机变化情况，从图 2-9 中可以看出 350 到 650 样本之间，传统的检测方法难以检测到故障，但是 MI-MBkNN 却能很好地检测出来。为了更好的找出引发故障的源变量，图 2-10(a)和(b)分别给出了使用 kNN 和 MI-MBkNN 方法时数据样本在第 400 样本点处的各变量与其均值中心的马氏距离，可以发现 MI-MBkNN 在寻找故障源变量方面提供更重要的指导，即汽提塔温度(变量 18)，汽提塔蒸汽流量(变



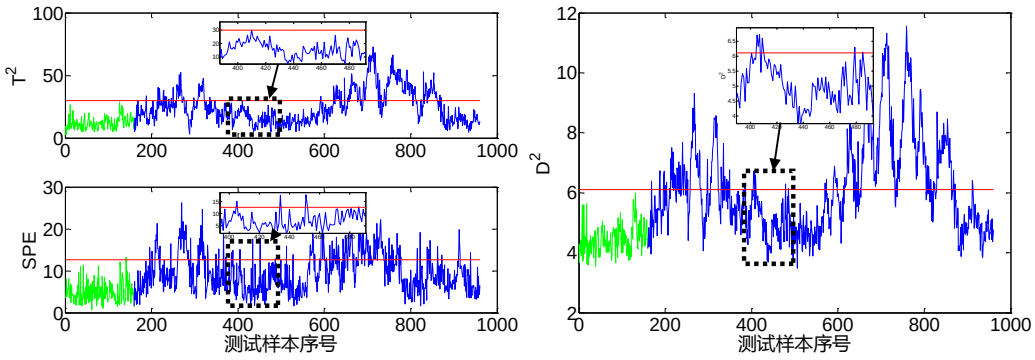
(a) (b)

图 2-7 第 161 样本点故障 5 的变量识别结果：(a)kNN 和(b)MI-MBkNN

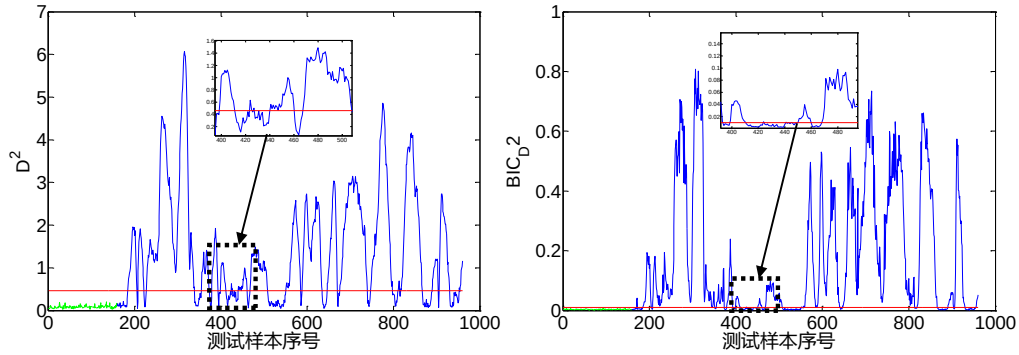


(a) (b)

图 2-8 第 400 样本点故障 5 的变量识别结果：(a)kNN 和(b)MI-MBkNN

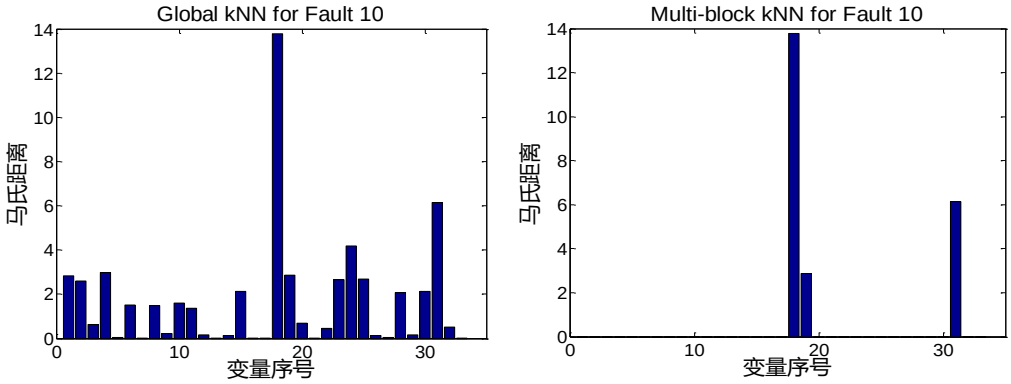


(a)PCA (b)kNN



(c)block 4 (d)MI-MBkNN

图 2-9 TE 过程故障 10 检测结果



(a) (b)

图 2-10 变量识别结果：(a)kNN 和(b)MI-MBkNN

量 19)，汽提塔蒸汽阀的变量(变量 31)是引起故障 10 的原因，因此子块 4 对故障 10 的敏感程度远远大于其他子块，表现为该子块的检测效果优于其他子块。最终通过贝叶斯融合后提升了整体的检测性能。因此本章所提方法相较于传统的全局模型检测方法有更好的检测效果，验证了方法的有效性。

2.4 本章小结

本章出了一种基于互信息的多块 kNN 故障检测方法，使用互信息对过程变量进行划分，把结构相似且对故障敏感程度相似的变量放在同一子块中，使得子块内的变量拥有更多相同的信息。并在每个子块中建立基于 kNN 的故障检测模型，所提方法反映了过程的更多局部信息，所以更易于故障的检测和诊断。将所提方法应用于 TE 过程仿真实验中，取得了较好的检测效果。

第三章 基于双层信息提取的多块 kNN 故障检测方法

第二章通过分析变量间的相关关系构建子块模型，可以有效解决只建立单一的全局模型而导致局部信息被淹没的问题，从而在一定程度上改善故障检测性能。然而只利用过程变量的观测数据进行建模，容易忽略数据中隐含的其他有效信息，模型的检测性能无法进一步提升，尤其当故障表现为微小偏移或振荡等不易检出特征时，模型的检测效果不佳。文献[63]根据原始数据定义累计误差信息和变化率信息进行故障特征提取，结合实际观测信息将原始数据扩充为 3 个子块，并进行分块检测，提升了传统检测方法对微小偏移或振荡等不易检出故障的检测性能。

为了更好地提取过程的局部信息并挖掘观测数据的特征信息，提出一种基于双层信息提取的多块 kNN 故障检测方法，改进传统 kNN 算法对微小偏移等不易检出类型故障的检测性能。首先利用典型相关分析计算变量间的相关系数对变量进行分块，获取局部信息；其次对各个变量子块分别提取观测信息、累计信息和变化率信息数据作为信息子块，通过提取特征信息放大故障差异。同时考虑数据的分布特征和局部区域样本的稀疏程度，采用基于马氏距离的 kNN 故障检测方法对各个信息子块建立检测模型；最后利用贝叶斯方法融合所有子块的检测结果，得到一个最终决策。通过 TE 过程和实际高炉炼铁过程的仿真实验，验证了所提方法的有效性。

3.1 典型相关分析

典型相关分析最早由 Hotelling 提出[64]，能有效提取变量之间的最大相关性。为研究变量之间的相关关系，定义变量 $\boldsymbol{X}^{\mathrm{T}}=(x_1,x_2,\ldots,x_p)$，$\boldsymbol{Y}^{\mathrm{T}}=(y_1,y_2,\ldots,y_q)$，定义系数矩阵 $\boldsymbol{A}=(a_1,a_2,\cdots,a_p)^{\mathrm{T}}$，$\boldsymbol{B}=(b_1,b_2,\cdots,b_q)^{\mathrm{T}}$，再分别定义变量 $\boldsymbol{X}$ 与变量 $\boldsymbol{Y}$ 的线性相关变量 $\boldsymbol{U}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}=a_1x_1+a_2x_2+\cdots+a_px_p$ 和 $\boldsymbol{V}=\boldsymbol{B}^{\mathrm{T}}\boldsymbol{Y}=b_1y_1+b_2y_2+\cdots+b_qy_q$，此时可通过研究 $\boldsymbol{U}$ 与 $\boldsymbol{V}$ 的相关性来确定 $\boldsymbol{X}$ 变量与 $\boldsymbol{Y}$ 变量之间的相关关系。

定义相关系数 ρ 如公式(3.1)所示。

$$\rho(\boldsymbol{X},\boldsymbol{Y})=\frac{\mathrm{E}[(\boldsymbol{X}-\mu_{\mathrm{X}})(\boldsymbol{Y}-\mu_{\mathrm{Y}})]}{\sigma_{\mathrm{X}}\sigma_{\mathrm{Y}}}=\frac{\mathrm{E}[(\boldsymbol{X}-\mu_{\mathrm{X}})(\boldsymbol{Y}-\mu_{\mathrm{Y}})]}{\sqrt{\sum_{i=1}^{n}(\boldsymbol{X}_i-\mu_{\mathrm{X}})^2}\sqrt{\sum_{i=1}^{n}(\boldsymbol{Y}_i-\mu_{\mathrm{Y}})^2}} \tag{3.1}$$

相关系数 ρ 的输出范围为-1 到 1，ρ 的绝对值越接近于 1，则表示变量 $\boldsymbol{X}$ 与变量 $\boldsymbol{Y}$ 的相关性越高；越接近零，则相关性越小。

定义数据矩阵 $\mathbf{Z}=[\boldsymbol{X}_{n\times p},\boldsymbol{Y}_{n\times q}]_{n\times(p+q)}$ 来合并上述变量，将 $\mathbf{Z}$ 的协方差矩阵和相关系数矩阵记做 $\boldsymbol{R}$，如公式(3.2）所示。

$$\boldsymbol{R}=\begin{bmatrix}\boldsymbol{R}_{11} & \boldsymbol{R}_{12}\\ \boldsymbol{R}_{21} & \boldsymbol{R}_{22}\end{bmatrix}=\begin{bmatrix}\mathrm{cov}(\boldsymbol{X},\boldsymbol{X}) & \mathrm{cov}(\boldsymbol{X},\boldsymbol{Y})\\ \mathrm{cov}(\boldsymbol{X},\boldsymbol{Y}) & \mathrm{cov}(\boldsymbol{Y},\boldsymbol{Y})\end{bmatrix} \tag{3.2}$$

因此，$\boldsymbol{U}$ 与 $\boldsymbol{V}$ 的相关系数可如公式(3.3)表示。

$$\rho(\boldsymbol{U},\boldsymbol{V})=\frac{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{12}\boldsymbol{B}}{\sqrt{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{11}\boldsymbol{A}}\sqrt{\boldsymbol{B}^{\mathrm{T}}\boldsymbol{R}_{22}\boldsymbol{B}}} \tag{3.3}$$

典型相关分析的目标是寻找最优的正交投影方向，因此可以转化为优化问题，如公式(3.4)所示。

$$J=\max(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{12}\boldsymbol{B}) \tag{3.4}$$

其中，$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_{11}\boldsymbol{A}=\boldsymbol{B}^{\mathrm{T}}\boldsymbol{R}_{22}\boldsymbol{B}=1$为约束条件。

定义$\boldsymbol{\gamma}=\boldsymbol{R}_{11}{}^{-1/2}\boldsymbol{R}_{12}\boldsymbol{R}_{22}{}^{-1/2}$，对$\boldsymbol{\gamma}$进行 SVD 分解，如公式(3.5)所示。

$$\boldsymbol{\gamma}=\boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Delta}^{\mathrm{T}} \tag{3.5}$$

其中，$\boldsymbol{\Gamma}=(\gamma_1,\cdots,\gamma_s)$，$\boldsymbol{\Delta}=(\delta_1,\cdots,\delta_m)$，$\boldsymbol{\Lambda}_k=diag\{\lambda_1,\cdots,\lambda_k\}$。

因此，相关性系数矩阵如公式(3.6)所示。

$$\boldsymbol{A}=\boldsymbol{R}_{11}{}^{-1/2}\boldsymbol{\Gamma}(:,1:k),\ \boldsymbol{B}=\boldsymbol{R}_{22}{}^{-1/2}\boldsymbol{\Delta}(:,1:k) \tag{3.6}$$

3.2 基于双层信息提取的多块建模故障检测方法

3.2.1 双层信息提取的分块策略

考虑到生产过程的局部信息以及信息的多样性，采用双层信息提取方式。

第一层：为更好提取和放大系统的局部特征，利用典型相关分析求出变量间的典型相关系数，根据经验设置相关系数的阈值，超过阈值的两两变量放在一起组成子块，使子块内的变量拥有更多相同的信息。

对于训练集$\boldsymbol{X}\in\mathbb{R}^{n\times m}$，$x_i\in\boldsymbol{X},x_j\in\boldsymbol{X}$，计算变量$x_i$与变量$x_j$之间的相关系数，即$\rho_{ij}=(x_i,x_j)(i=1,2,\cdots,m;\ j=1,2,\cdots,m)$，根据多次实验对比和经验，若$\rho_{ij}$的绝对值超过0.6，则把变量$x_i$与变量$x_j$放到相同子块。最终，将原始数据集分为$b$个子块，即$\boldsymbol{X}=[\boldsymbol{X}_1,\boldsymbol{X}_2,...,\boldsymbol{X}_b]\in\mathbb{R}^{n\times m}$。

第二层：对每个变量子块分别提取观测信息、累计信息和变化率信息数据，将提取的特征信息数据作为信息子块，然后对构建的信息子块分别建立相应的故障检测模型。

累计信息是指将一段时间内的原始数据通过相加计算而得到的数据信息。通过提取累计信息可以扩大故障变量的弱小偏移和缓变，从而提高故障的检出率。

假设标准化后的变量数据集为$\boldsymbol{X}\in\mathbb{R}^{n\times m}$，由于构造新特征应用了数据集前$\boldsymbol{T}$个时刻的样本，因此得到的累计信息数据集会损失$\boldsymbol{T}$个样本，即$\boldsymbol{X}_{\mathrm{I}}\in\mathbb{R}^{(n-T)\times m}$。第$t$时刻的累计信息如公式(3.7)所示。

$$x_{\mathrm{I}}(t)=\sum_{l=0}^{T}(x(t-l))\ ;\ x_{\mathrm{I}}(t)\subset\boldsymbol{X}_{\mathrm{I}} \tag{3.7}$$

其中，$x_{\mathrm{I}}(t)$为t时刻的累计信息，$x(t)$为t时刻的观测数据样本。

变化率信息反映变量的变化速率，当故障的类型表现为振荡而非单一的偏移，传统的检测方法难以检测。但利用其变化速率超过正常范围，可达到检测振荡类型故障的目的。

假设标准化后的变量数据集为 $\boldsymbol{X}\in\mathbb{R}^{n\times m}$，由于累计信息数据集损失前 T 个样本，为了保证各子块数据集样本和变量数量相同，令提取变化率信息后的数据集为 $\boldsymbol{X}_{\mathrm{D}}\in\mathbb{R}^{(n-T)\times m}$。第 t 时刻的变化率信息如公式(3.8)所示。

$$x_{\mathrm{D}}(t)=x(t)-x(t-1)\ ;\ x_{\mathrm{D}}(t)\subset\boldsymbol{X}_{\mathrm{D}} \tag{3.8}$$

其中，$x_{\mathrm{D}}(t)$ 表示 t 时刻变化率信息，$x(t)$ 表示观测数据中 t 时刻的样本。

通过对划分好的变量子块进一步提取观测信息、累计信息和变化率信息数据，得到信息子块 $\boldsymbol{X}_b$、$\boldsymbol{X}_{b\mathrm{I}}$ 和 $\boldsymbol{X}_{b\mathrm{D}}$，充分挖掘了各子块数据中隐含的有效信息。本章采用的基于双层信息提取的多块模型建立步骤如图 3-1 所示。

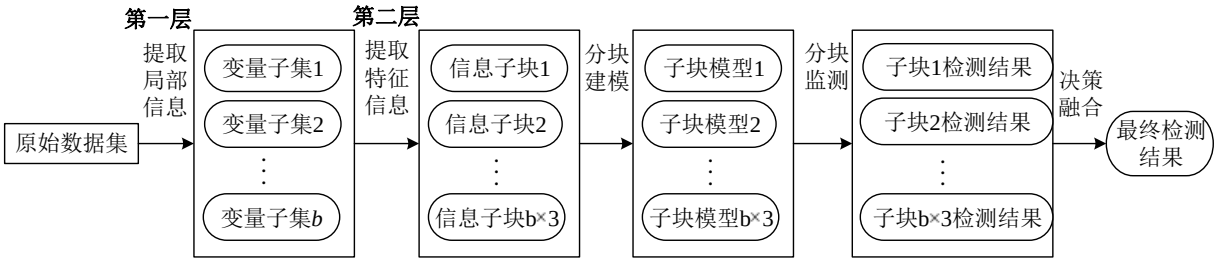


图 3-1 基于双层信息提取的多块模型建立步骤

3.2.2 基于马氏距离的 kNN 故障检测方法

马氏距离(Mahalanobis Distance, MD)是一种常用的距离指标，由印度统计学家马哈拉诺比斯提出[65]。它通过计算数据的协方差距离，能有效表示样本集的分布特征，并且不受量纲的影响，这使它优于欧氏距离。若协方差矩阵为单位矩阵，则马氏距离简化为欧氏距离。

记 $\mathbb{D}=\left\{\boldsymbol{g}_i\,\middle|\,\boldsymbol{g}_i\in\mathbb{R}^{m\times 1},\ \ i=1,2\cdots n\right\}$ 为样本数据集，其中 n 为样本数量，m 为样本维度，$\boldsymbol{\Sigma}\in\mathbb{R}^{m\times m}$ 表示协方差矩阵。则样本 $g_1,g_2\in\mathbb{D}$ 的马氏距离 d^M 如公式(3.9)所示。

$$d^M(g_1,g_2)=\sqrt{(g_1-g_2)^{\mathrm{T}}\Sigma^{-1}(g_1-g_2)} \tag{3.9}$$

其中，$\boldsymbol{\Sigma}$ 的计算公式如公式(3.10)所示。

$$\boldsymbol{\Sigma}=\frac{1}{n-1}\sum_{g_i\in\mathbb{D}}(g_i-\bar{g})(g_i-\bar{g})^{\mathrm{T}},\ \bar{g}=\frac{1}{n}\sum_{g_i\in\mathbb{D}}g_i \tag{3.10}$$

马氏距离考虑了数据集中样本的分布特征以及局部区域样本的稀疏程度，对样本在维度上的差异性更加敏感，因此非常适用于故障检测。

基于马氏距离的 kNN 故障检测方法(kNN Fault Detection Based on Mahalanobis Distance, MDkNN)相较于传统 kNN 故障检测方法在距离度量上做了改进。假设数据集 $\boldsymbol{X}\in\mathbb{R}^{n\times m}$，$x_i\in\boldsymbol{X},i=1,2,\cdots,n$，通过计算马氏距离寻找样本点 x_i 在训练集中前 k 个近邻，记做 $F(x_i,k)=\{x_i^1,x_i^2,\ldots,x_i^j,\ldots,x_i^k\}$，其中，$x_i^j$ 表示样本 x_i 的第 j 个近邻样本。定义 D_i^2 为样本 x_i 与 k 个近邻样本的马氏距离的平方，如公式(3.11)所示，通过根据置信度 α 确定控制限 D_α^2

$$D_i^2 = \sum_{j=1}^{k} d^{M2}(x_i, x_i^j) \tag{3.11}$$

故障检测部分首先计算待测样本 x 的统计量 D_x^2，若 $D_x^2 \geq D_\alpha^2$，则判定样本 x 为故障样本，否则为正常样本。

3.2.3 故障在线检测过程

对故障进行在线检测时，根据 3.2.1 节所述方法构造信息子块，并将每个信息子块分别作为单独的子块。子块划分结果如表 3-1 所示。

表 3-1 子块划分结果

子块编号	训练数据集		待测样本
1		$\boldsymbol{X}_1$	x_{1test}
2	$\boldsymbol{X}_1$	$\boldsymbol{X}_{1I}$	x_{1test}^I
3		$\boldsymbol{X}_{1D}$	x_{1test}^D
4		$\boldsymbol{X}_2$	x_{2test}
5	$\boldsymbol{X}_2$	$\boldsymbol{X}_{2I}$	x_{2test}^I
6		$\boldsymbol{X}_{2D}$	x_{2test}^D
…	…	…	…
$3\times b-2$		$\boldsymbol{X}_b$	x_{btest}
$3\times b-1$	$\boldsymbol{X}_b$	$\boldsymbol{X}_{bI}$	x_{btest}^I
$3\times b$		$\boldsymbol{X}_{bD}$	x_{btest}^D

针对构建好的信息子块，分别建立基于马氏距离的 kNN 故障检测模型，得到相应的统计量与控制限。由于子块数目较多且产生了多个检测结果，难以得到一个最终决策。本章采用贝叶斯融合策略融合各子块的检测结果，得到一个最终的 BIC 检测指标。BIC 指标的具体计算方式与第二章相同。

3.2.4 基于双层信息提取的多块 kNN 故障检测方法流程

基于双层信息提取的多块 kNN 故障检测方法(kNN Fault Detection Based on Two-layer Information Extraction and Mahalanobis Distance, TIE-MDkNN)流程如图 3-2 所示，算法的具体实施过程描述如下。

Step 1:获取正常工况数据集 $\boldsymbol{X}$，并对其进行标准化处理；

Step 2:利用典型相关分析方法，根据两两变量间的相关系数构建变量子块，生成 b 个变量子块；

Step 3:对每个变量子块进一步提取观测信息、累计信息和变化率信息 3 种特征信息数据，构建 $3\times b$ 个信息子块；

Step 4:对每个子块分别建立 kNN 模型，利用核密度估计方法确定各自的故障控制限；

Step 5:对于新来的标准化测试样本 x_{test}，按照步骤 2 和步骤 3 的方法得到新的测试样本 $x_{1test}, x_{1test}^{I}, \cdots, x_{btest}^{D}$；

Step 6:对步骤 5 中的各信息子块建立基于马氏距离的 kNN 故障检测模型；

Step 7:融合各子块的检测结果，得到 BIC 统计量，根据置信度确定控制限，当 BIC 统计量超过控制限时则判断发生了故障，反之正常。

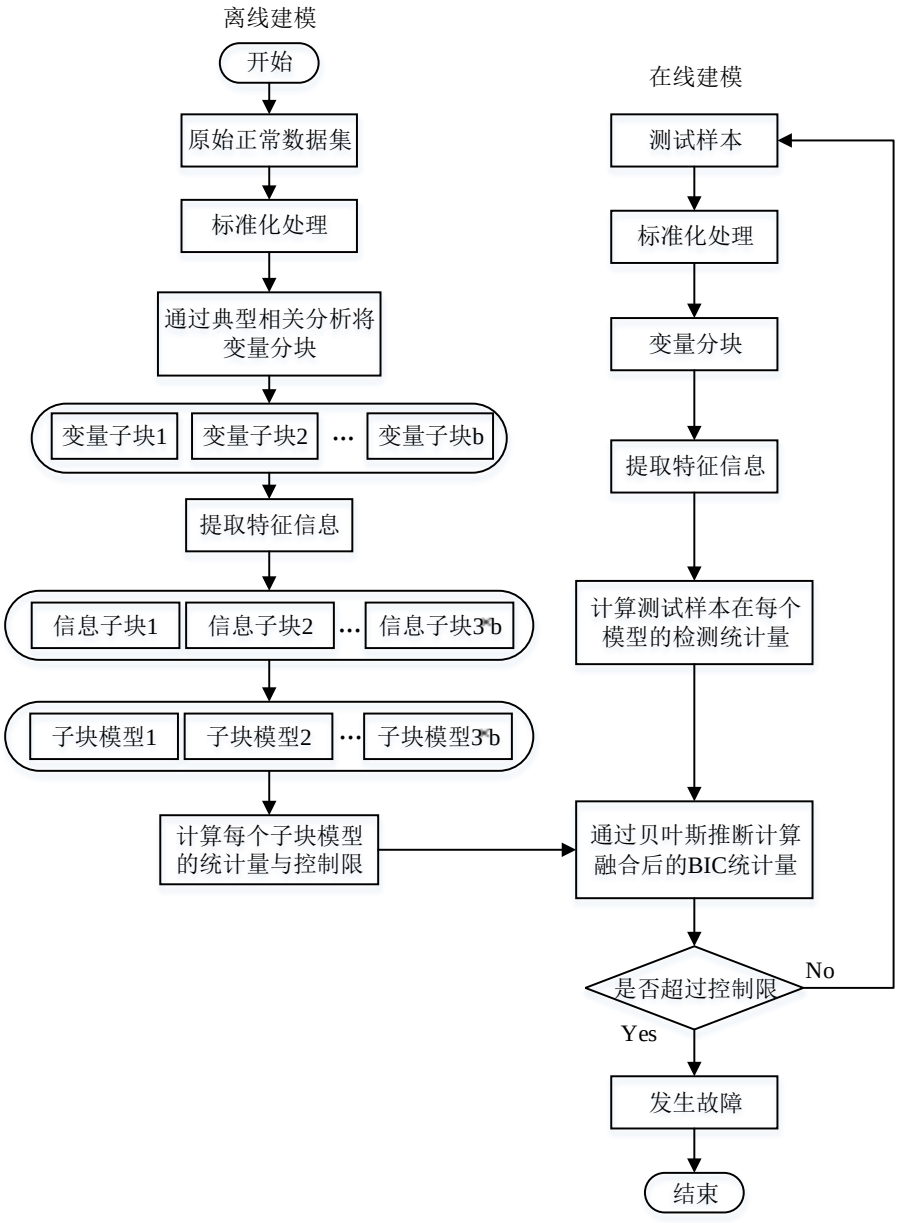


图 3-2 基于 TIE-MDkNN 的故障检测算法流程

3.3 仿真实验

3.3.1 TE 过程仿真

本章选取 TE 过程中 22 个过程测量变量和 11 个操作变量用于故障检测方法建模和性能测试。图 3-3 展示了 33 个变量之间的相关系数值，根据多次实验比较最优结果，将相关系数的阈值设定为绝对值大于 0.6。若变量之间相关系数超过阈值，则变量间相关性较高，受到的故障影响接近，对故障的敏感程度相似，因此将其放入同一个子块更

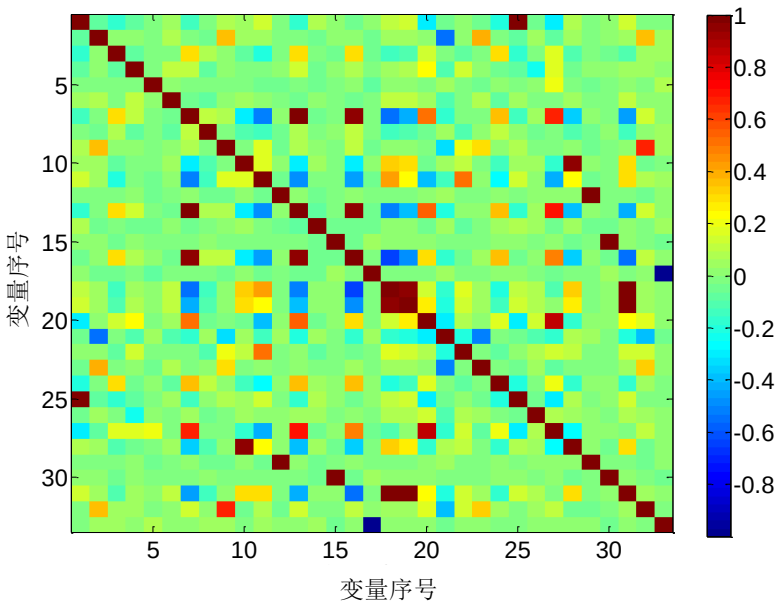


图 3-3 变量之间的相关系数值

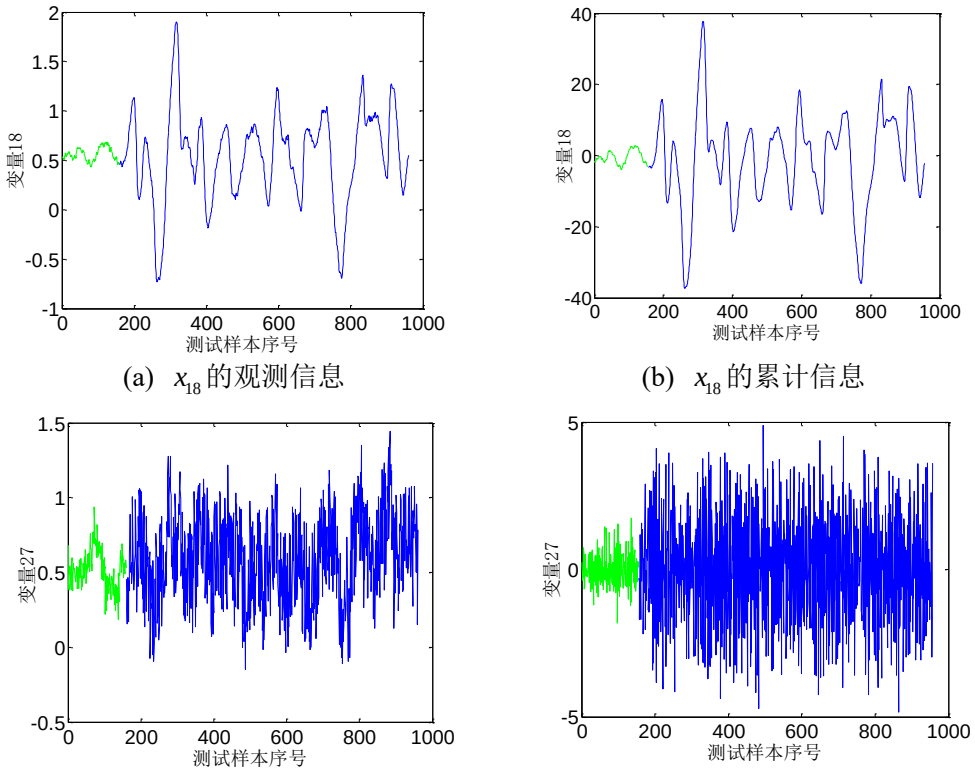
容易检测到故障。具体的变量分块结果如表 3-2 所示。

表 3-2 TE 过程变量分块结果

变量子块编号	变量序号	变量子块编号	变量序号
1	1,25	6	17,33
2	9,32	7	18,19,31
3	10,28	8	7,13,16,20,27
4	12,29	9	2,3,4,5,6,8,11,14,21,22,23,24,26
5	15,30		

经过第一层信息提取后，获得 9 个变量子块，再对每个变量子块分别提取观测信息、累计信息和变化率信息数据，并根据特征信息构建信息子块。最终，经过双层信息提取后，9 个变量子块可以得到 27 个特征信息子块。

表 3-3 给出了 9 个子块对 21 种故障的报警率和平均误报率。为了便于观察与比较各方法的检测效果，用加粗的字体表示本章方法下最好的检测结果及相应的故障序号。可以看出，对于不同的故障，由于某些子块经过特征信息提取后，放大了正常样本与故障样本的差异性，最终提升了整体的检测效果。从对所有故障的平均报警率来看，融合后的检测性能有了明显的提高。对于故障 10 和故障 19，最优检测子块为变量子块 7(包括变量 18,19,31)所扩展的累计信息子块和变量子块 8(包括变量 7,13,16,20,27)所扩展的变化率信息子块。图 3-4 展示了变量 18 的观测信息和累加信息的特征曲线图以及变量 27 的观测信息和变化率信息的特征曲线图。通过对变量 18 提取累计信息，放大了故障的偏移量，使故障样本明显偏离正常工况，因此累计信息子块表现了良好的检测性能。故障 19 表现为振荡类型故障，通过对变量 27 提取变化率信息，放大了正常与故障样本之间的差异性，使得变化率信息子块对故障更为敏感，最后融合的检测结果融合了该信息子块的优势。



(a) x_{18} 的观测信息　(b) x_{18} 的累计信息

(c) x_{27} 的观测信息　(d) x_{27} 的振荡信息

图 3-4 变量 18 和变量 19 的特征信息曲线图

表 3-3 TE 过程各变量子块的故障报警率

故障编号	子块 1	子块 2	子块 3	子块 4	子块 5	子块 6	子块 7	子块 8	子块 9	BIC
1	0.986	0.188	0.423	0.003	0.001	0.260	0.981	0.531	0.998	**0.998**
2	0.048	0.114	0.984	0.001	0.008	0.946	0.920	0.985	0.986	0.984
3	0.011	0.013	0.030	0.009	0.005	0.003	0.069	0.115	0.405	0.093
4	0.009	0.999	0.023	0.006	0.009	0.011	0.033	0.081	0.410	**0.999**
5	0.110	0.155	0.205	0.021	0.009	0.995	0.264	0.318	0.583	**0.999**
6	0.999	0.975	0.991	0.003	0.004	0.959	0.966	0.999	0.995	**0.999**
7	0.264	0.269	0.333	0.009	0.011	0.270	0.434	0.503	0.999	**0.999**
8	0.661	0.555	0.906	0.010	0.015	0.666	0.908	0.978	0.985	0.979
9	0.014	0.021	0.026	0	0.009	0.046	0.049	0.148	0.429	0.133
10	0.119	0.120	0.141	0.003	0.005	0.148	0.865	0.493	0.658	**0.849**
11	0.010	0.966	0.009	0.005	0.006	0.025	0.093	0.190	0.476	**0.971**
12	0.509	0.839	0.743	0.049	0.080	0.965	0.988	0.993	1.000	0.999
13	0.524	0.570	0.833	0.033	0.015	0.900	0.946	0.921	0.973	0.956
14	0.011	0.999	0.005	0.006	0.009	0.003	0.015	0.249	0.995	**1**
15	0.024	0.023	0.010	0.006	0.009	0.039	0.070	0.166	0.410	0.109
16	0.041	0.049	0.091	0.004	0.003	0.051	0.970	0.313	0.549	**0.974**
17	0.021	0.924	0.074	0.006	0.005	0.028	0.265	0.228	0.986	0.981
18	0.845	0.863	0.881	0.013	0.013	0.890	0.898	0.908	0.958	0.912
19	0.006	0.043	0.005	0.008	0.009	0.006	0.628	0.996	0.593	**0.993**
20	0.046	0.041	0.079	0.006	0.016	0.584	0.259	0.904	0.751	**0.913**
21	0.005	0.011	0.055	0.005	0.004	0.529	0.643	0.449	0.793	0.630
平均故障报警率	0.251	0.416	0.326	0.010	0.012	0.396	0.536	0.546	0.759	**0.832**

平均故障误报率	0.006	**0.005**	0.012	0.010	0.007	0.008	0.075	0.089	0.321	0.058

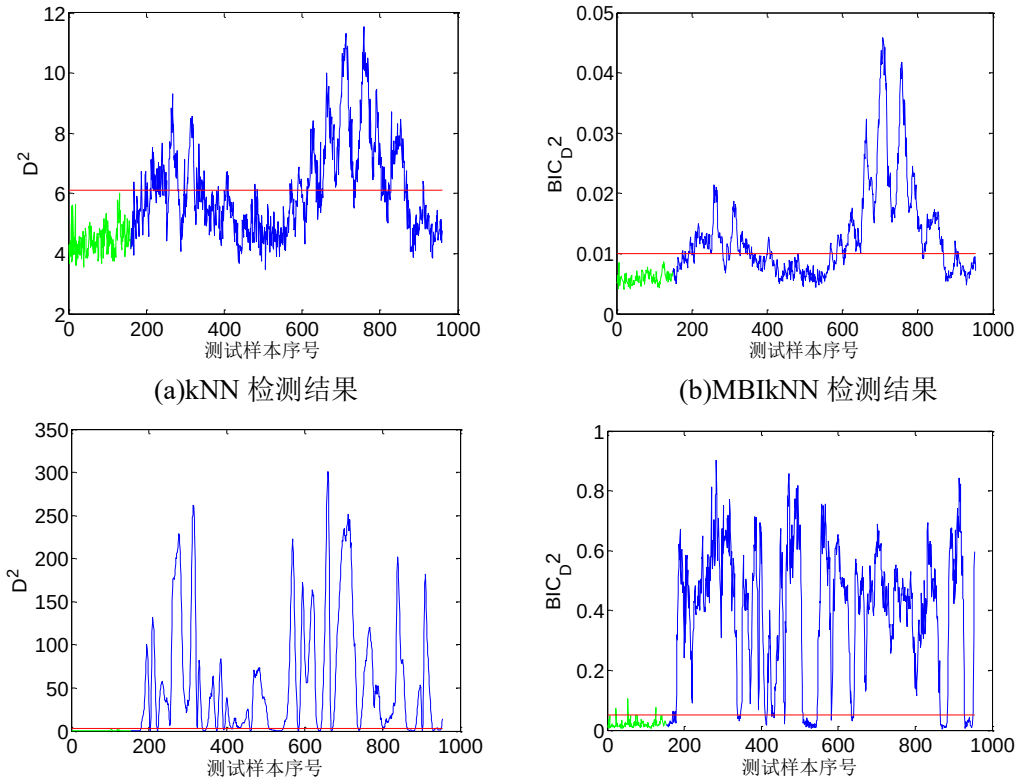
表 3-4 给出了 TE 过程在不同检测方法下的检测结果，方法包括基于 PCA、kNN、MI-MBkNN、MDkNN 的检测方法，基于信息提取的 kNN 故障检测方法(kNN Fault Detection Based on Multi-block Information Extraction, MBIkNN)、基于多块信息提取和马氏距离的 kNN 故障检测方法(kNN Fault Detection Based on Multi-block Information Extraction and Mahalanobis Distance, MBI-MDkNN)以及本章所提的 TIE-MDkNN。通过网格搜索算法，取近邻个数 k 为 19，累计信息宽度 T 为 5。为了便于观察，将本章方法下最好的检测结果及相应的故障序号以加粗的字体表示。从仿真结果来看，对于大多数故障，TIE-MDkNN 方法的检测结果要优于其他 6 种方法。

表 3-4 各种检测方法性能比较

故障编号	PCA	kNN	MI-MBkNN	MDkNN	MBIkNN	MBI-MDkNN	TIE-MDkNN
	报警率	报警率	报警率	报警率	报警率	报警率	报警率
1	0.999	0.996	0.998	1	0.998	0.999	0.998
2	0.983	0.983	0.986	0.985	0.986	0.983	**0.984**
3	0.025	0.013	0.033	0.035	0.033	0.028	**0.093**
4	1	0.975	0.994	1	0.994	1	0.999
5	0.244	0.260	0.951	1	0.951	1	0.999
6	1	1	1	1	1	1	0.999
7	1	1	1	1	1	1	0.999
8	0.968	0.976	0.976	0.984	0.976	0.984	0.979
9	0.017	0.020	0.021	0.035	0.021	0.025	0.133
10	0.299	0.418	0.758	0.892	0.758	0.849	**0.918**
11	0.759	0.683	0.660	0.808	0.660	0.951	**0.970**
12	0.986	0.989	0.991	0.999	0.991	0.999	**0.999**
13	0.954	0.946	0.949	0.952	0.949	0.955	**0.956**
14	1	1	1	1	1	0.999	**1**
15	0.031	0.029	0.089	0.066	0.089	0.056	**0.109**
16	0.274	0.289	0.804	0.925	0.804	0.971	**0.974**
17	0.954	0.919	0.900	0.972	0.900	0.976	**0.981**
18	0.902	0.896	0.899	0.904	0.899	0.904	**0.911**
19	0.126	0.099	0.468	0.936	0.468	0.991	**0.993**
20	0.497	0.495	0.630	0.912	0.630	0.921	0.913
21	0.476	0.425	0.449	0.565	0.449	0.628	**0.630**
平均报警率	0.643	0.639	0.741	0.808	0.741	0.824	**0.832**
平均误报率	0.004	0.006	0.017	0.02	0.017	0.004	0.058

为说明方法的有效性，以故障 10 为例展示详细的检测过程。故障 10 是流 2(C 进料)中温度的随机变化情况，图 3-5 展示了该故障在不同方法下的检测结果。对于传统 kNN 检测方法，在 350 到 650 样本间和 900 至 960 样本间的检测效果不佳，整体报警率只有 41.8%。MBIkNN 经过一层的信息提取后，放大了故障样本的偏移量和对振荡故障的敏感程度，检测效果得到了改善，报警率达 75.8%。而在本章方法中，通过对数据进行双

层信息提取以及采用基于马氏距离的 kNN 故障检测方法，模型的检测性能再次得到提升。TIE-MDkNN 方法下的最优子块能够很好地检测到故障，使得融合后的检测结果报警率达到 91.8%，明显高于其他两种方法。



(a)kNN 检测结果　(b)MBIkNN 检测结果

(c)最优子块检测结果　(d)TIE-MDkNN 检测结果

图 3-5 TE 过程故障 10 检测结果

3.3.2 高炉炼铁过程应用

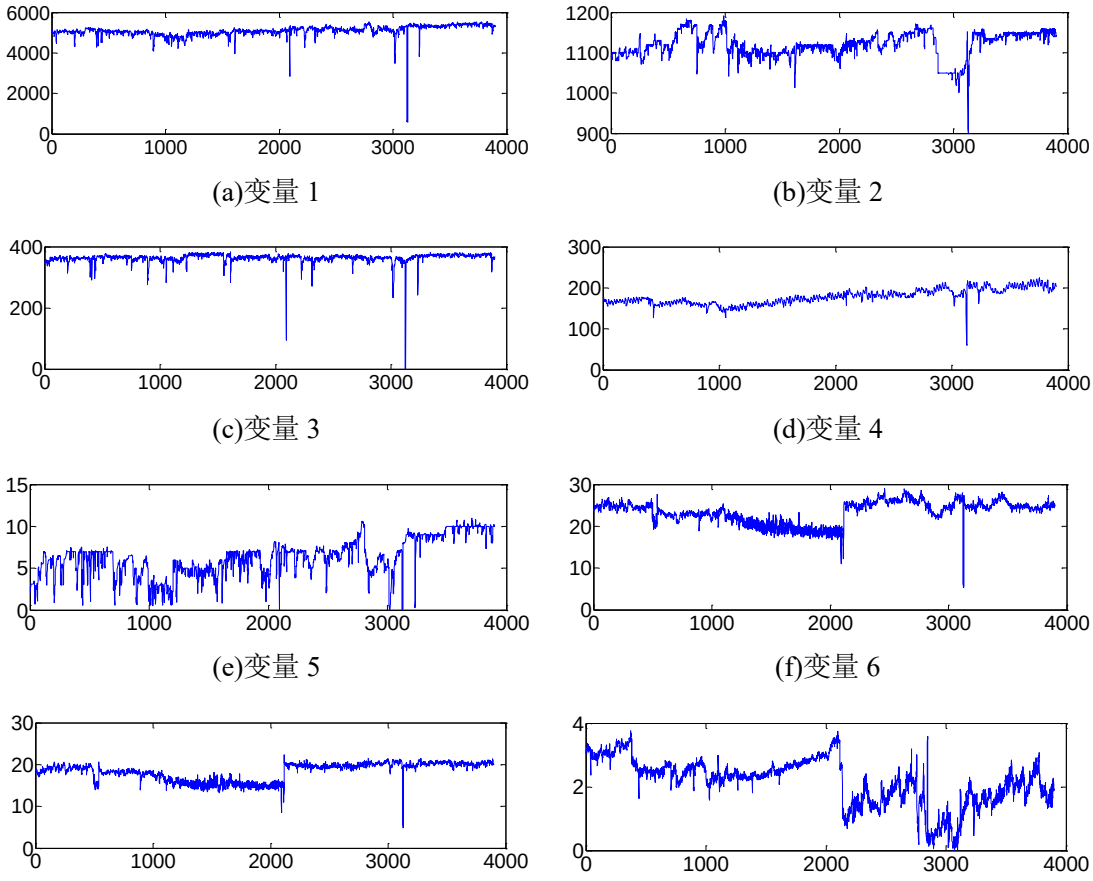
本章基于实际高炉炼铁过程进行仿真实验，用以验证所提方法的有效性。为了达到高炉炼铁过程节能降耗的目的，必须保证铁水的生产质量和产量。当气体流动不稳定时会影响碳的燃烧，最终导致炉腹架空，产生悬挂故障。若没有及时检测出悬挂故障，将会导致热应力和内部的气体压力过大，使得顶部结构受到严重的破损。本章考虑了实际情况中悬挂故障的存在，选择 8 个高炉过程变量进行仿真实验，将正常工况下的 2000 个样本作为训练样本，悬挂故障下的 1900 个样本作为测试样本。在悬挂故障下，炉内的温度和压力增加，炉顶的一氧化碳和二氧化碳浓度上升，氢气的浓度下降。为了更好的表现变量的特性，表 3-5 给出了 8 个过程变量的描述，图 3-6 给出了 8 个变量的变化曲线图。

利用 3.2.1 节所述分块方法将 8 个变量分成两个变量子块，变量子块 1 为$u1$、$u3$，

表 3-5 悬挂故障检测中选择的过程变量

高炉过程变量	变量描述	高炉过程变量	变量描述
$u1$	风量	$u5$	富阳量
$u2$	风温	$u6$	炉顶煤气 CO_2 含量
$u3$	风压	$u7$	炉顶煤气 CO 含量
$u4$	炉顶压力	$u8$	炉顶煤气 H_2 含量

变量子块 2 为$u2$、$u4$、$u5$、$u6$、$u7$、$u8$，并对各变量子块进一步提取观测信息、累计信息和变化率信息。构成 6 个信息子块并建立基于马氏距离的 kNN 故障检测模型，并采用贝叶斯方法融合各子块检测结果。表 3-6 给出了不同检测方法的检测结果，图 3-7 展示了变量子块 2 的各信息子块的检测结果。可以看出累计信息子块由于放大了故障样本偏移量，其检测效果明显好于变化率信息子块，由于本章所提方法对变量进行了合理分块，把结构相似且对故障最为敏感的变量放在同一个子块中，提升了检测效果。并再次对子块数据进行特征提取，充分挖掘数据的有效信息，同时考虑到局部区域样本的稀疏程度，采用基于马氏距离的 kNN 故障检测方法，使得最终的检测性能得到了显著的提升，再次验证本章所提方法的有效性和优越性。



(a)变量 1　(b)变量 2

(c)变量 3　(d)变量 4

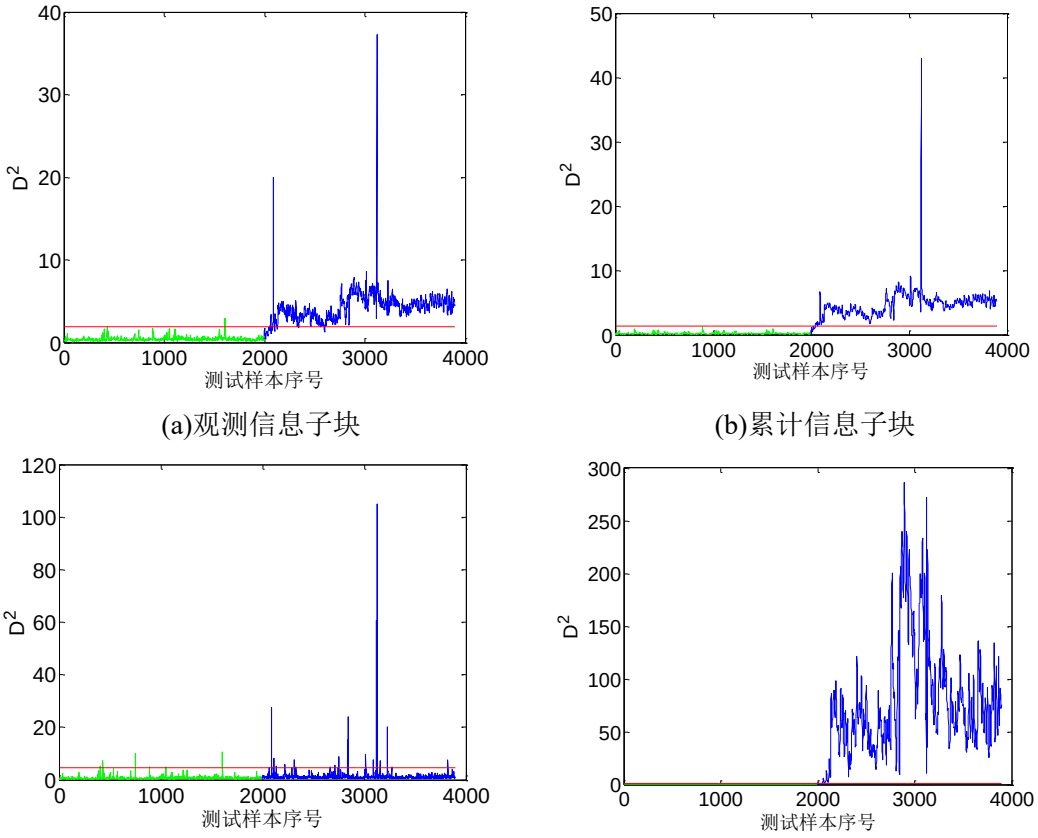
(e)变量 5　(f)变量 6

(g)变量 7　(h)变量 8

图 3-6 高炉过程各变量曲线图

表 3-6 不同方法的检测性能比较

	PCA	kNN	TIE-MDkNN
报警率	0.941	0.935	**0.992**
误报率	0.018	0.001	**0**



(a)观测信息子块　(b)累计信息子块

(c)变化率信息子块　(d)TIE-MDkNN

图 3-7 高炉过程检测结果

3.4 本章小结

本章提出了一种基于双层信息提取的多块 kNN 故障检测方法，通过对变量进行合理分块以获得局部信息，并再次对子块数据进行特征提取，充分挖掘子块数据的有效信息。利用典型相关分析计算变量间的相关系数进行变量子块构建，对各个变量子块分别提取观测、累计和变化率 3 种特征信息数据，并将其作为信息子块。同时考虑到样本的分布特征以及局部区域的稀疏程度，采用 MDkNN 算法对各个信息子块建立故障检测模型，最后采用贝叶斯方法融合各子块的检测结果。将所提方法应用于 TE 过程和实际高炉炼铁过程中，均取得了较高的故障检出率，验证了方法的有效性。

第四章 基于重构误差和多块建模策略的 kNN 故障检测

在传统基于 kNN 的故障检测算法中，引发故障的异常信息易被正常信息淹没，导致故障检测不及时和报警率低问题。文献[66]利用近邻距离之差将样本映射到距离空间，通过捕捉样本之间的差异性来解决故障样本信息易被其他样本掩盖的问题。文献[67]采用加权策略构建故障数据与正常数据之间的距离残差，并对含有较强微小故障信息的样本赋予较大权值，避免了故障样本信息被其他样本淹没。传统的故障检测方法大多采用全局建模策略，但是由于现代工业结构复杂度提升，操作单元的数量不断增加以及变量间相关关系不断多样化，建立的全局模型检测准确度有待提高。而近年来提出的局部、多块等建模策略的优势得到了充分发挥。

基于以上分析，本章在多块建模策略框架下，提出一种基于自编码器重构误差的 kNN 故障检测方法。为了解决统计量计算过程中异常信息易被淹没的问题，先基于自编码器模型进行正常工况数据的重构还原，再基于该模型求取异常工况数据的重构误差，抽离出异常信息数据；并将该重构误差作为观测信息，进一步提取累计信息和变化率信息构建 3 个信息子块，对每个信息子块建立相应的 kNN 模型。再利用核密度估计方法确定各子块中的控制限，最后将待测样本在各个子块上的检测结果通过贝叶斯融合进行故障检测。将所提方法进行数值仿真和 TE 过程仿真，均取得了较好的检测效果，验证了所提方法的性能。

4.1 自编码器

自编码器是一种三层架构的无监督学习模型，由编码器和解码器组成。通过将输入数据进行编码和解码得到重构项，然后以最小化重构项与原始输入间的误差为目标更新模型参数[68]。自编码器模型的预测输出称为重构输出(Reconstruction Output, RO)，故模型输入与其重构输出的差值可称为重构误差。

从结构上看，自编码器非常类似于多层感知器，包括输入层、隐含层和输出层[69]。由于自编码器目标在于提取隐变量空间或对输入进行降噪重构，而不是根据给定输入 $\boldsymbol{X}$ 进行传统分类或回归任务，因此自编码器中的输出层与其输入层具有相同数量的节点，其结构图如图 4-1 所示。

自编码器将输入 $\boldsymbol{X}\in\mathbb{R}^{n\times1}$ 通过函数 f 映射得到新特征输出 $\boldsymbol{H}\in\mathbb{R}^{m\times1}$(通常 $m<n$)，去除输入中冗余信息的同时，最大程度地保留对输入数据解释性最好的信息。编码函数如公式(4.1)所示。

$$H(\boldsymbol{X})=f(\boldsymbol{W}_1*\boldsymbol{X}+\boldsymbol{b}_1) \tag{4.1}$$

其中 $\boldsymbol{H}$ 为输入的新特征表示； $f(\cdot)$ 通常为非线性激活函数，本章选用 sigmoid 函数，即 $f(x)=\dfrac{1}{1+e^{-x}}$； $\boldsymbol{W}_1$ 和 $\boldsymbol{b}_1$ 分别为编码器的权重矩阵和偏差向量。

解码器通过非线性激活函数 g 将 H 映射回原始数据空间，得到 $\boldsymbol{X}'\in\mathbb{R}^{n\times1}$。解码函数如式(4.2)所示。

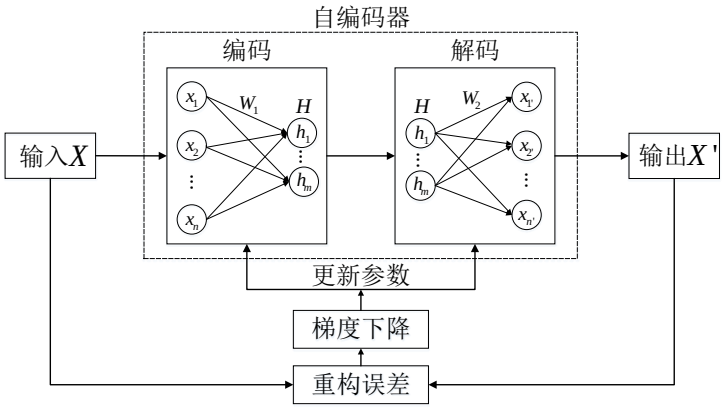


图 4-1 自编码器结构图

$$\boldsymbol{X}' = g(\boldsymbol{W}_2 * \boldsymbol{H} + \boldsymbol{b}_2) \tag{4.2}$$

其中 $\boldsymbol{X}'$ 为输入的重构表示； $g(\cdot)$ 也为 sigmoid 激活函数，$\boldsymbol{W}_2$ 和 $\boldsymbol{b}_2$ 分别为解码器的权值矩阵和偏差向量。

参数集 $\boldsymbol{\theta} = \{\boldsymbol{W}_1, \boldsymbol{b}_1, \boldsymbol{W}_2, \boldsymbol{b}_2\}$ 以最小化重构误差 $L(\boldsymbol{X}, \boldsymbol{X}')$ 为目标进行参数更新，从而使得输出尽可能接近输入，故优化目标如公式(4.3)所示。

$$\begin{aligned} \boldsymbol{\theta}^* &= \arg\min_{\theta} L(\boldsymbol{X}, \boldsymbol{X}') \\ s.t. \boldsymbol{X} &\approx \boldsymbol{X}' \end{aligned} \tag{4.3}$$

本章以重构输出与输入的欧式距离平方为指标计算重构误差，误差函数如公式(4.4)所示。

$$\begin{aligned} L(\boldsymbol{X}, \boldsymbol{X}') &= \left\| \boldsymbol{X} - \boldsymbol{X}' \right\|_2^2 = \left\| \boldsymbol{X} - g(\boldsymbol{W}_2 * \boldsymbol{H} + \boldsymbol{b}_2) \right\|_2^2 \\ &= \left\| \boldsymbol{X} - g(\boldsymbol{W}_2 * f(\boldsymbol{W}_1 * \boldsymbol{X} + \boldsymbol{b}_1) + \boldsymbol{b}_2) \right\|_2^2 \end{aligned} \tag{4.4}$$

4.2 一种基于重构误差和多块建模策略的 kNN 故障检测方法

4.2.1 基于自编码器重构误差的 kNN 故障检测

测试集样本数据实际上可以看做正常信息数据加异常信息数据的和集，如公式(4.5)所示。

$$\boldsymbol{X}_{test} = \boldsymbol{X}_{test}^{n} + \boldsymbol{X}_{test}^{un} \tag{4.5}$$

其中， $\boldsymbol{X}_{test}^{n}$ 为正常信息数据， $\boldsymbol{X}_{test}^{un}$ 为异常信息数据。

本章构建的自编码器模型以 $\boldsymbol{X}_{train}$ 作为训练集，$\boldsymbol{X}_{train}$ 为正常工况下记录得到的数据，因此 $\boldsymbol{X}_{train}$ 也称为正常信息数据，其重构输出 $\hat{\boldsymbol{X}}_{train}$ 可如式(4.6)所示。

$$\hat{\boldsymbol{X}}_{train} = g(\omega * \boldsymbol{H}_{train} + \boldsymbol{b}) \tag{4.6}$$

其中， $\boldsymbol{H}_{train}$ 表示 $\boldsymbol{X}_{train}$ 经过编码后提取出的潜隐变量，$\boldsymbol{b}$ 为偏差向量。假设 $\boldsymbol{X}_{train}$ 服从 D 分布，则 $\boldsymbol{H}_{train}$ 服从 D' 分布， $\hat{\boldsymbol{X}}_{train}$ 服从 D'' 分布。

自编码器的损失函数为重构误差即 $\boldsymbol{X}_{train} - \hat{\boldsymbol{X}}_{train}$，当训练过程结束时，损失函数近似为零，可以得出 $\hat{\boldsymbol{X}}_{train} \approx \boldsymbol{X}_{train}$。

测试集 $\boldsymbol{X}_{test}$ 基于该自编码器的重构输出如式(4.7)所示。

$$\begin{aligned}\hat{\boldsymbol{X}}_{test} &= g(\omega*(\boldsymbol{H}_{test}^{n}+\boldsymbol{H}_{test}^{un})+\boldsymbol{b}) \\ &= g(\omega*\boldsymbol{H}_{test}^{n}+\boldsymbol{b})+g(\omega*\boldsymbol{H}_{test}^{un}+\boldsymbol{b}) \\ &= \hat{\boldsymbol{X}}_{test}^{n}+\hat{\boldsymbol{X}}_{test}^{un}\end{aligned} \tag{4.7}$$

其中，$\boldsymbol{H}_{test}^{n}$ 为 $\boldsymbol{X}_{test}$ 中正常信息数据的潜隐变量，则其服从 D' 分布，$\hat{\boldsymbol{X}}_{test}^{n}$ 服从 D'' 分布，结合 $\boldsymbol{X}_{train}$ 服从 D 分布，$\hat{\boldsymbol{X}}_{train}$ 服从 D'' 分布可以推断出 $\hat{\boldsymbol{X}}_{test}^{n} \approx \boldsymbol{X}_{test}^{n}$。

将测试集与其在该自编码器模型的重构输出做差，由式(4.5)和式(4.7)可得重构误差如式(4.8)所示。

$$\boldsymbol{X}_{test}-\hat{\boldsymbol{X}}_{test} \approx \boldsymbol{X}_{test}^{un}-\hat{\boldsymbol{X}}_{test}^{un} \tag{4.8}$$

因此可以发现重构误差仅与异常信息数据有关，基于重构误差建立故障检测模型，避免了统计量计算过程中，由于正常信息数据量级远超异常信息数据而淹没异常信息数据的问题，提升了故障检测的效果。图 4-2 给出了基于重构误差的 kNN 故障检测方法(kNN Fault Detection Based on Reconstruction Error, RE-kNN)与基于传统 kNN 检测方法对比图。

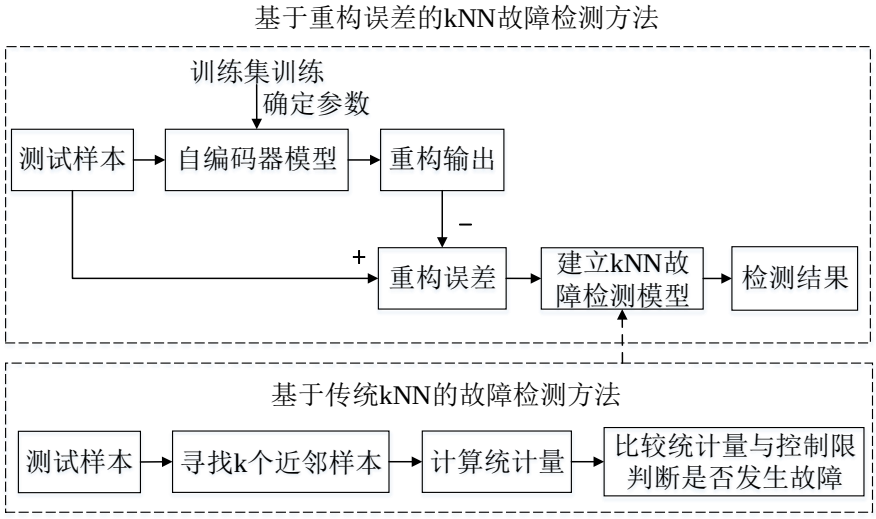


图 4-2 基于重构误差的 kNN 故障检测方法与传统 kNN 对比

4.2.2 子块模型的建立

传统多块建模方法根据不同的变量选取准则将过程划分成若干子块，本章将自编码器重构误差作为观测信息，提取出累计信息和变化率信息构成三个信息子块，进行分块检测，关于累计信息和变化率信息的具体描述如 3.2.1 节所述。

将原始数据集经自编码器得到的重构误差作为原始观测值信息 $\boldsymbol{X}$，对 $\boldsymbol{X}$ 提取累计信息和变化率信息，分别得到信息子块 $\boldsymbol{X}_{\mathrm{I}}$ 和 $\boldsymbol{X}_{\mathrm{D}}$，从而得到三个信息子块。本章所采用的基于自编码器重构误差的多块建模方法如图 4-3 所示。

划分好子块的基础上，对各子块建立 kNN 检测模型，得到相应的统计量与控制限。再进一步采用贝叶斯融合策略融合所有子块的检测结果，得到 BIC 检测指标。BIC 指标的具体计算方式与第二章相同。

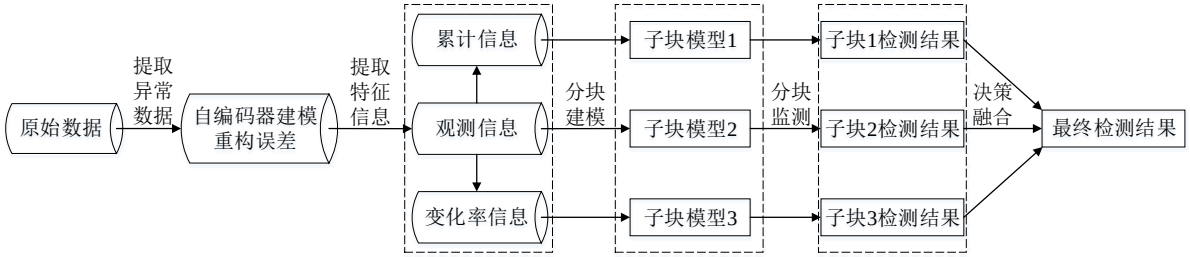


图 4-3 基于自编码器重构误差的多块建模方法

4.2.3 基于重构误差和多块建模策略的故障检测方法流程描述

基于重构误差和多块建模策略的 kNN 故障检测方法(kNN Fault Detection Based on Reconstruction Error and Multi-block Modeling Strategy, RE-MBIkNN)的流程如图 4-4 所示，具体描述如下。

Step 1:基于标准化后的正常工况数据集 $\boldsymbol{X}$，训练自编码器模型，得到重构误差 R_{e}；

Step 2:将重构误差作为观测信息，进一步提取累计信息和变化率信息，构成三个信息子块 R_{e}、R_{e}^{I}、R_{e}^{D}；

Step 3:对每个子块分别建立 kNN 检测模型并根据核密度估计方法求取故障控制限；

Step 4:对于新来的标准化测试样本 x_{test}，在训练好的自编码器模型上得到重构项并求取重构误差，按步骤 2 提取特征信息，得到信息子块；

Step 5:对步骤 4 中的各子块建立 kNN 故障检测模型；

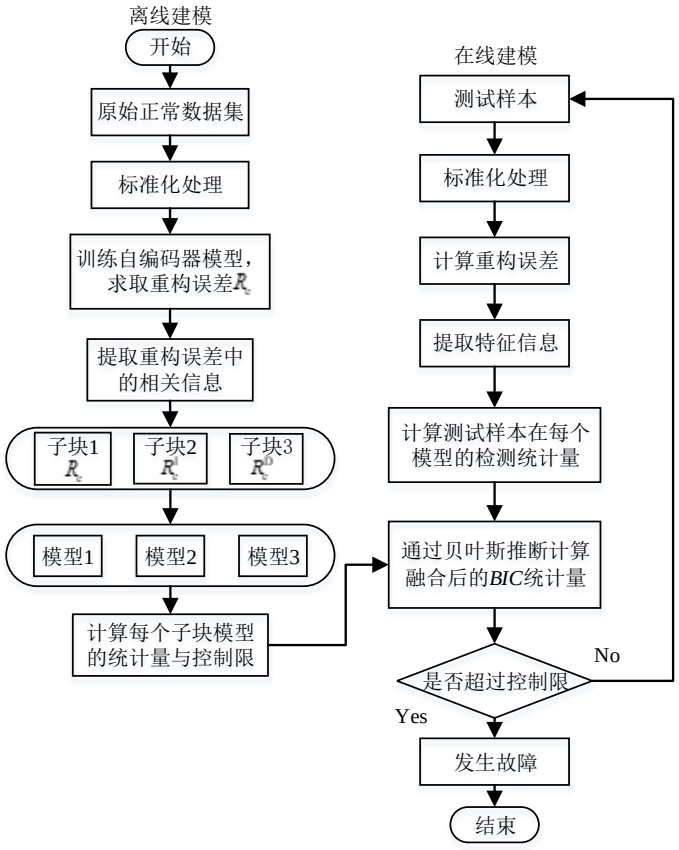


图 4-4 基于重构误差和多块建模策略的 kNN 故障检测流程

Step 6:采用贝叶斯融合方法，将各子块的统计量组合成一个新的 BIC 统计量，并根据置信度确定控制限，当 BIC 统计量超过控制限时则判断发生了故障，反之正常。

4.3 仿真实验

4.3.1 数值仿真

采用文献[28]中的数值例子进行仿真验证，其具体结构如式(4.5)：

$$\begin{aligned} x_1 &= y + e_1, \\ x_2 &= y^2 - 3y + e_2, \\ x_3 &= -y^3 + 3y^2 + e_3, \\ x_4 &= y^4 - 4y^3 + 2y + e_4, \\ x_5 &= -2y^5 + 6y^4 - 3y^3 + y + e_5 \end{aligned} \tag{4.5}$$

其中，$x_i(i=1,2,3,4,5)$为变量，y为服从[0.01,2]上均匀分布的源变量，$e_j(j=1,2,3,4,5)$为噪声变量，$e_j \sim N(0,0.01^2)(j=1,2,3,4,5)$。

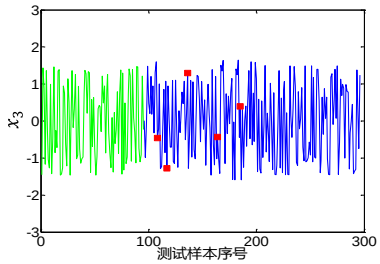
该数值例子共产生 300 个训练样本，故障设置如下：

故障 1：在第 100 到第 200 处样本给变量 x_3 加入幅值为 0.2 的振荡信号；

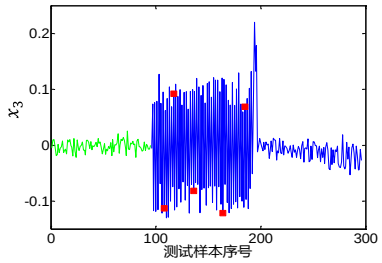
故障 2：在第 201 到第 300 处样本给变量 x_5 加入幅值为 0.005 的斜坡信号。

为了验证所提方法的有效性，图 4-5 给出了数值样本及处理后数值样本的变量曲线图。其中图 4-5(a)、4-5(b)、4-5(c)分别为数值样本、重构误差样本及对重构误差样本提取变化率信息后样本在变量 x_3 上的曲线图，图 4-5(d)、4-5(e)、4-5(f)分别为数值样本、重构误差样本及对重构误差样本提取累计信息后样本在变量 x_5 上的曲线图。由于累计信息宽度 T 取值为 5，因此在检测中会损失前 5 个样本点。从图中可以看出，对于数值仿真中设计的两种特殊类型故障，故障样本曲线与正常样本的变量曲线差异性不大(图 4-5(a)和图 4-5(d))，而重构误差的变量曲线图很好地还原了故障信息特征(图 4-5(b)和图 4-5(e))，可以明显地区分开正常样本与异常样本。为进一步深入分析，在曲线图中用“■”表示变量 x_3 的第 108、117、136、164、185 个样本，用“▲”表示变量 x_5 的第 210、226、236、264、285 个样本。原始数据中由于上述几个故障样本点在变量 x_3 上的值位于正常工况样本变量值域内而漏报，当经过重构误差后，变量值均明显偏离了正常工况样本的变量值域，因此被准确地检测出来。

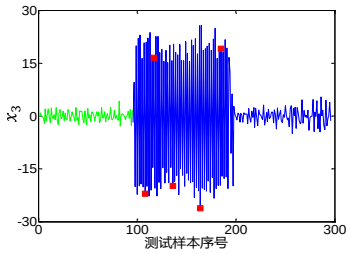
由图4-5(c)和图4-5(f)可以看出，合适的信息提取方式可以显著地提高故障的检出率，



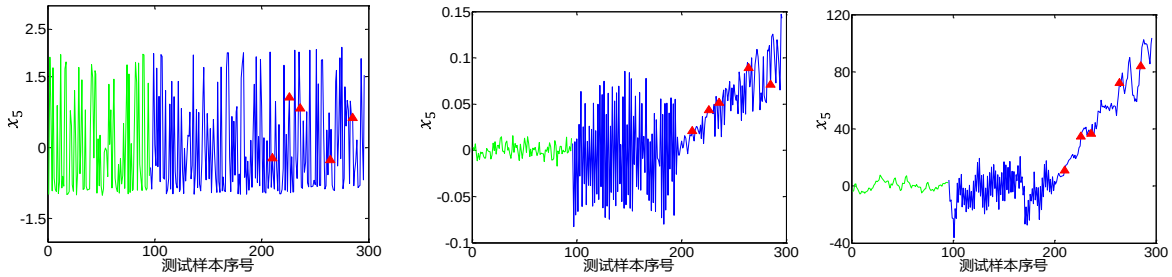
(a) 变量 x_3



(b) 变量 x_3 的重构误差



(c)子块3中变量 x_3 的重构误差

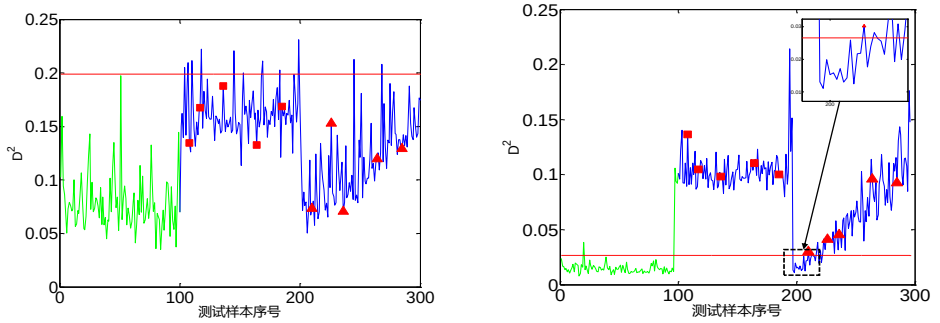


(d) 变量 x_5　　(e) 变量 x_5 的重构误差　　(f)子块2中变量 x_5 的重构误差

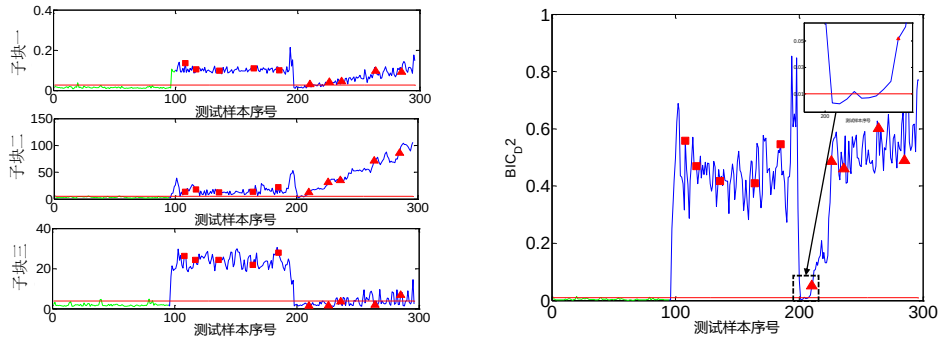
图 4-5 数值样本变量曲线图

比如，对于变量 x_3 的振荡类型故障，由于提取了变化率信息，使得故障发生前后的差异更加明显；对于变量 x_5，提取的累计信息将异常数据的微小偏移放大，使故障可以及时地被检测出来。

基于 kNN，RE-kNN，RE-MBIkNN 三种方法进行仿真与对比，采用网格搜索方法确定近邻个数 k 为 19。检测结果如图 4-6 和表 4-1 所示。由图 4-6 可知，传统 kNN 方法无法有效检测到故障的发生，报警率仅有 9%。而基于重构误差的 kNN 可以有效提高报警率，报警率达 87%。对比图 4-6(b)和图 4-6(c)可以发现，通过提取累计信息将异常信息的微小偏移再次放大，提升了子块二对微小偏移类型故障的检测效果。子块三通过提取异常信息的变化率，使得对振荡类型的故障更为敏感。最后融合的检测结果融合了子块二和子块三的优势，对这两种不易检出的特殊故障类型都做到了很好的检测，最终报警率达到 98%。采用加权贡献图[62]方法对过程进行诊断，图 4-7 给出



(a) kNN 检测结果　　(b) RE-kNN 检测结果



(c) 各子块的检测结果　　(d) RE-MBIkNN 检测结果

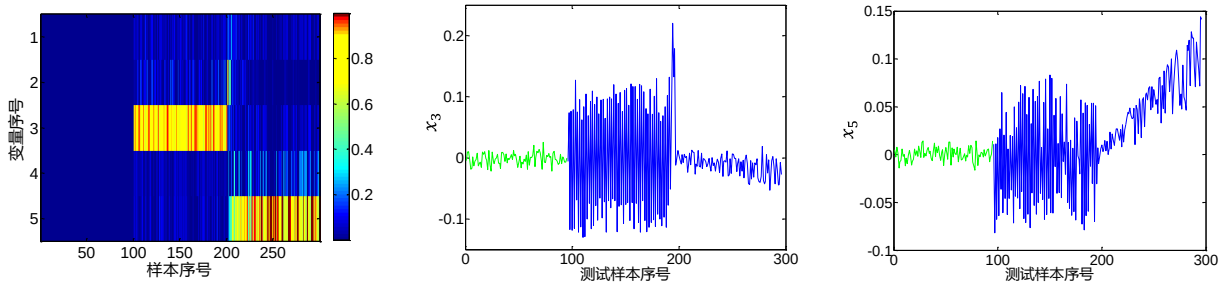
图 4-6 kNN、RE-kNN 和 RE-MBIkNN 的检测结果

了故障的诊断结果，可以看出变量 x_3 和 x_5 是引发故障的源变量。

表 4-1 数值仿真故障报警率及误报率比较

	kNN	RE-kNN	RE-MBIkNN
报警率	0.090	0.872	**0.981**
误报率	0	0.04	**0**

为更好地说明本文方法的有效性，对提取重构误差后的样本绘制变量 x_3 和 x_5 的曲线图，图 4-7(b)第 100 至第 200 个样本处变量曲线呈现出了与数值仿真设置的故障一类型一致的振荡信号，图 4-7(c)第 200 至第 300 个样本处变量曲线呈现出了与仿真设置故障二类型一致的斜坡信号。由于数据经过了预处理故重构误差提取的异常信息与实际设置的故障信号幅值并不相等，但变量变化趋势和数据分布都大致相同，可见通过求取自编码器重构误差确实有效地从正常工况数据中抽离出了异常信息数据，解决了异常信息易被淹没的问题，改善了检测效果。



(a) RE-MBIkNN (b)变量 x_3 (c)变量 x_5

图 4-7 数值仿真故障诊断结果

4.3.2 TE 过程仿真

利用正常工况数据训练自编码器模型，基于该模型进行重构误差提取以解决异常信息易被淹没的问题，并将该重构误差作为观测信息，进一步提取累计信息和变化率信息构建 3 个信息子块，对每个信息子块建立相应的故障检测模型并融合得到一个最终的检测指标。

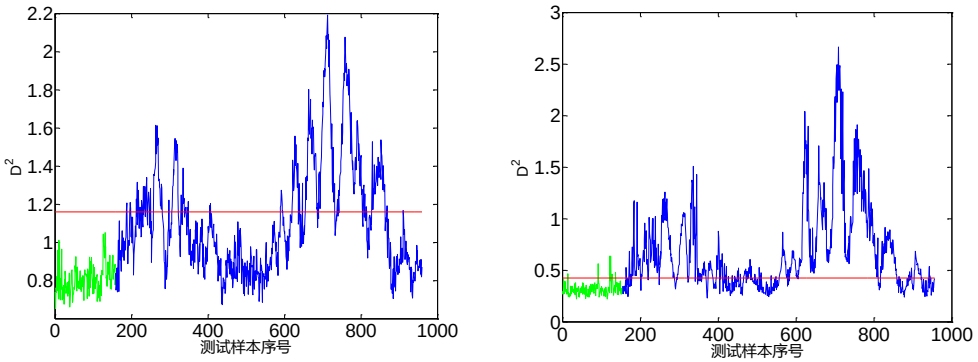
表 4-2 给出了各子块分别以原始数据为观测信息和以重构误差为观测信息时对 21 种故障的报警率和平均误报率。对比各子块的报警率，其中以重构误差为观测信息时的报警率要高于以原始数据为观测信息时的报警率，可见通过重构误差提取异常信息，改善了传统基于 kNN 的故障检测方法中故障源信息易被正常信息淹没导致检测效果不理想的问题。此外，由于子块二提取了累计信息，放大了故障信号和故障偏移量，使得子块二相较于其他两个子块更易检测到故障，但同时也放大了噪声干扰信息，导致误报率的提高。子块三虽然报警率低，但是对于某些故障检出率高且对降低融合后的误报率起着重要的作用。从对 21 种故障的检测结果来看，融合后的检测性能有了明显的提高，以重构误差为观测信息的报警率达到 81.6%。

为更好地说明所提方法的有效性，选取故障 10、故障 11 的检测结果做详细的说明。图 4-8 展示了故障 10 的检测结果，从图 4-8 可以看出，传统 kNN 方法只能在第 700 到

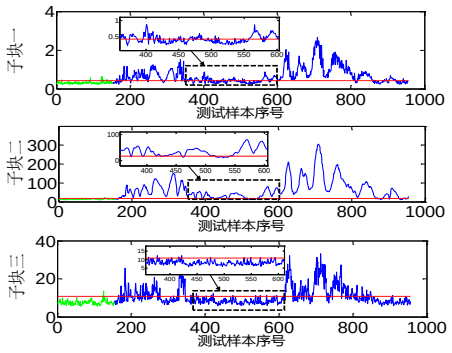
表 4-2 以原始数据、重构误差为观测信息时 TE 过程检测结果对比

故障编号	以原始数据为观测信息				以重构误差为观测信息			
	子块一	子块二	子块三	BIC	子块一	子块二	子块三	BIC
1	0.996	0.998	0.001	0.995	0.996	0.995	0.976	0.995
2	0.983	0.988	0	0.984	0.986	0.989	0.978	0.984
3	0.013	0.326	0.001	0.054	0.092	0.195	0.119	0.150
4	0.975	0.999	0.005	0.999	0.999	0.999	0.117	0.999
5	0.260	0.809	0.002	0.381	0.999	0.999	0.336	0.999
6	1	0.999	0.004	0.999	0.999	0.999	0.911	0.999
7	1	0.999	0.007	0.999	0.999	0.999	0.554	0.999
8	0.976	0.993	0.006	0.979	0.979	0.984	0.969	0.976
9	0.020	0.290	0.005	0.051	0.084	0.162	0.135	0.132
10	0.418	0.815	0.000	0.618	0.737	0.860	0.393	0.848
11	0.683	0.941	0.047	0.868	0.758	0.930	0.553	0.913
12	0.989	1	0.199	0.993	0.995	1	0.981	1
13	0.946	0.963	0.031	0.951	0.950	0.958	0.944	0.953
14	1	0.878	0.998	0.998	0.999	0.998	0.999	0.999
15	0.029	0.326	0.004	0.105	0.101	0.220	0.119	0.181
16	0.289	0.820	0.002	0.549	0.729	0.886	0.291	0.893
17	0.919	0.974	0.149	0.966	0.946	0.975	0.864	0.976
18	0.896	0.940	0.041	0.899	0.906	0.914	0.898	0.913
19	0.099	0.412	0.114	0.135	0.669	0.644	0.697	0.779
20	0.495	0.848	0.039	0.642	0.669	0.843	0.452	0.811
21	0.425	0.699	0.001	0.566	0.576	0.664	0.583	0.638
平均报警率	0.639	0.810	0.079	0.701	0.770	0.820	0.613	0.816
平均误报率	0.006	0.159	0.003	0.015	0.046	0.079	0.067	0.059

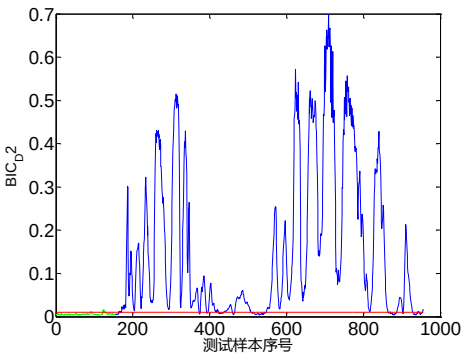
第 800 样本间较好地检测到故障，基于重构误差的 kNN 已经可以在故障初期和故障后期做到大范围的报警，而采用信息提取和多块建模策略融合检测后，报警率进一步提升，实现了更准确及时的报警。从图 4-9(a)中可以看出引发故障 10 的源变量为变量 x_{18}、x_{19} 和 x_{31}。故障的主要原因是故障变量在第 300 和 800 样本点附近发生了过大的波动，而提取的重构误差放大了正常与故障样本之间的差异性，使得 RE-kNN 可以很好地检测到由于变量不正常波动产生的故障。但是变量 x_{19} 在第 400 到 600 样本点之间有一个幅值较小的偏移，kNN 和 RE-kNN 方法均难以检测到。而本章节所提方法在求取重构误差的



(a) kNN 检测结果　　(b) RE-kNN 检测结果

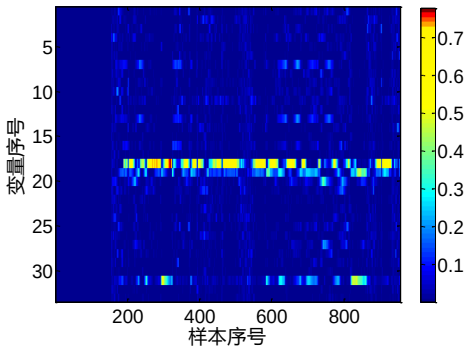


(c) 各子块的检测结果

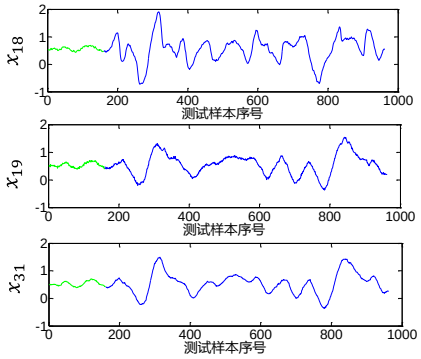


(d) RE-MBIkNN 检测结果

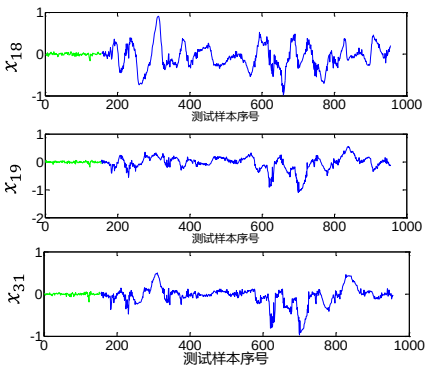
图 4-8 kNN、RE-kNN 及 RE-MBIkNN 对故障 10 的检测结果



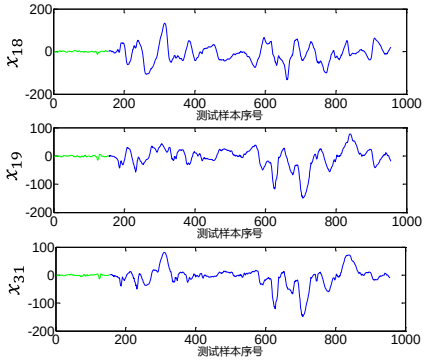
(a) 故障 10 诊断结果



(b) 故障变量曲线图



(c) 重构误差故障变量曲线图(子块一)

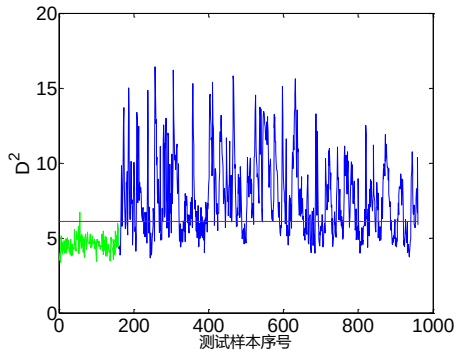


(d) 重构误差故障变量曲线图(子块二)

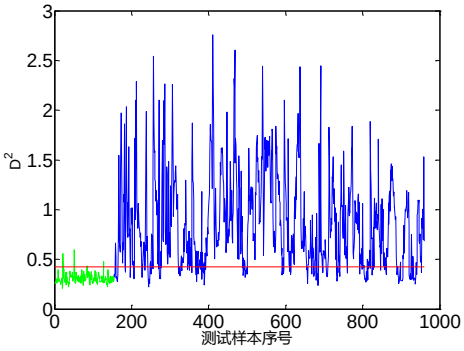
图 4-9 故障 10 诊断结果和故障变量曲线图

基础上进行信息提取，将该偏移再次放大，因此能够很好地检测到该故障。

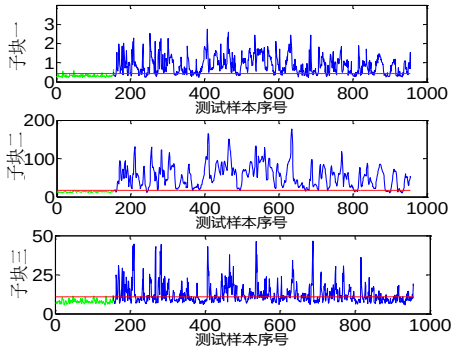
传统 kNN 方法对故障 11 的报警率只有 68%，而本章所提方法从故障引入点处开始就能做到大范围的持续报警，报警率达 91.3%，具体检测结果如图 4-10 所示。图 4-11(a)对故障提供了清晰的诊断结果，即引发该故障的源变量为变量 x_9 与 x_{32}。在原始 TE 数据集、重构误差数据集及重构误差数据集提取累计信息后数据集上分别对变量 x_9 与 x_{32} 绘制变量曲线图，如图 4-11(b)、4-11(c)、4-11(d)所示，可以看出引发该故障的故障变量相较于重构误差的故障变量，有更多的故障样本变量值处于正常工况样本的变量值域内，这些样本点计算所得统计量与正常工况样本统计量相差不大，因此以重构误差为统计量



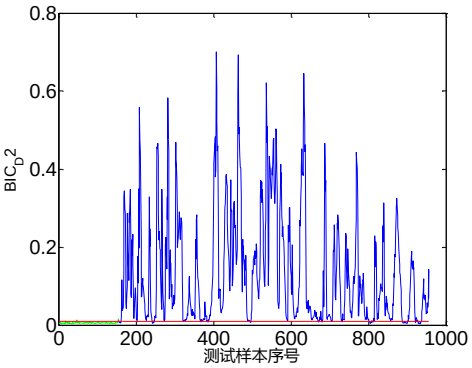
(a) kNN 检测结果



(b) RE-kNN 检测结果

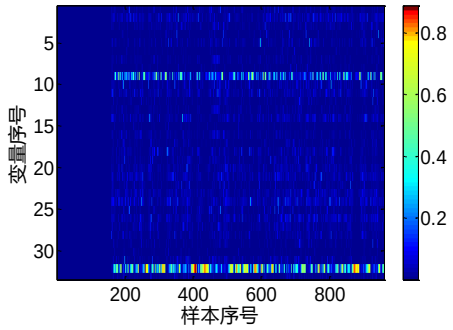


(c) 各子块检测结果

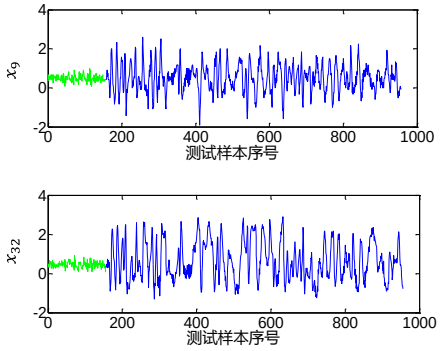


(d) RE-MBIkNN 检测结果

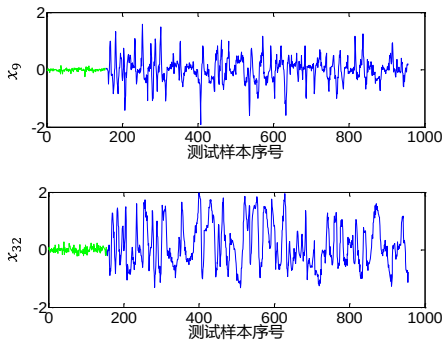
图 4-10 kNN、RE-kNN 及 RE-MBIkNN 对故障 11 的检测结果



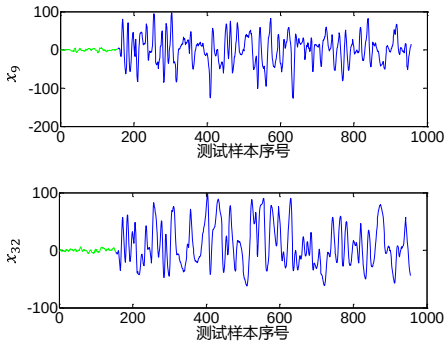
(a) 故障 11 诊断结果



(b) 故障变量曲线图



(c)重构误差故障变量曲线图（子块一）



(d) 重构误差故障变量曲线图（子块二）

图 4-11 故障 11 诊断结果和故障变量曲线图

计算时的检测效果要优于传统 kNN 检测方法。对比图 4-11(c)和图 4-11(d)可以看出提取累计信息后，放大了变量之间的差异，使得子块二表现了良好的检测性能，最终提升了

整体的检测效果。

表 4-3 给出了六种不同检测方法对 TE 过程 21 种故障的报警率和平均误报率，包括传统的基于 PCA 的故障检测方法、基于支持向量数据描述(SVDD)的检测方法、基于 kNN 的故障检测方法、基于信息提取的 kNN 故障检测方法(MBIkNN)、基于自编码器的 kNN 故障检测(RE-kNN)和本章所提方法(RE-MBIkNN)，其中各方法控制限均采用核密度估计方法确定，核宽系数为 0.5，通过网格搜索算法，取近邻个数 k 为 19。为了便于观察与比较，用加粗的字体表示本章方法下最好的检测结果及相应故障序号。从表 4-3 中可以看出，RE-kNN 相比于传统检测方法，报警率有了一定的提升。而引入多块建模策略后，在部分传统方法及 RE-kNN 不易检出的故障上，报警率获得了显著的提升。本章所提方法综合了 MBIkNN 和 RE-kNN 的优势，在大多数故障下报警率最高，平均报警率达 81.6%，进一步验证了本章所提方法的检测性能。

表 4-3 PCA、SVDD、kNN、MBIkNN、RE-kNN 和 RE-MBIkNN 的 TE 过程检测结果

故障编号	PCA	SVDD	kNN	MBIkNN	RE-kNN	RE-MBIkNN
	报警率	报警率	报警率	报警率	报警率	报警率
1	0.999	0.993	0.995	0.995	0.996	0.995
2	0.983	0.984	0.983	0.984	0.986	0.984
3	0.025	0.037	0.013	0.054	0.092	**0.150**
4	1	0.792	0.974	0.999	0.999	0.999
5	0.244	0.275	0.260	0.382	0.999	**0.999**
6	1	1	1	0.999	0.999	0.999
7	1	1	1	0.999	0.999	0.999
8	0.968	0.974	0.977	0.979	0.979	0.976
9	0.017	0.03	0.021	0.051	0.084	**0.132**
10	0.299	0.448	0.418	0.618	0.738	**0.848**
11	0.759	0.599	0.683	0.869	0.758	**0.913**
12	0.986	0.984	0.988	0.993	0.995	**1**
13	0.954	0.945	0.946	0.951	0.950	**0.953**
14	1	1	1	0.998	0.999	0.999
15	0.031	0.062	0.029	0.106	0.102	**0.181**
16	0.274	0.283	0.279	0.549	0.729	**0.893**
17	0.954	0.878	0.919	0.966	0.946	**0.976**
18	0.902	0.897	0.898	0.899	0.906	**0.913**
19	0.126	0.047	0.089	0.134	0.668	**0.779**
20	0.497	0.458	0.485	0.643	0.669	**0.810**
21	0.476	0.419	0.426	0.566	0.576	**0.639**
平均报警率	0.643	0.624	0.637	0.701	0.771	**0.816**
平均误报率	0.004	0.007	0.006	0.015	0.046	0.059

4.4 本章小结

针对基于 kNN 的故障检测算法中，引发故障的异常信息易被正常信息淹没，导致故障检测不及时和报警率低的问题，提出一种基于重构误差和多块建模策略的 kNN 故障检测方法，通过求取自编码器的重构误差解决异常信息数据易被正常工况数据淹没的

问题，再对该重构误差提取观测、累计和变化率信息，建立 3 个信息子块进行多块建模，最终通过贝叶斯方法融合各子块结果做出决策。所提方法与其他几种故障检测方法相比，在仿真实验中取得更好的检测效果，表明了方法的优越性。

第五章 总结与展望

5.1 工作总结

本文在多块建模策略下，对基于数据驱动的故障检测方法展开了进一步的研究，主要内容如下：

（1)针对基于 kNN 的故障检测方法只建立一个全局模型,不考虑过程的局部信息，提出一种基于互信息的多块 kNN 故障检测方法。根据变量间的互信息值大小进行子块构建，使得子块内的变量拥有更多相同的信息，最大化地反映变量的一个或者多个局部特征。随后对每个子块采用 kNN 方法进行建模并监控，再利用贝叶斯推断方法将各子块的检测结果融合，最后采用基于马氏距离的故障诊断方法找出引发故障的源变量。TE 过程的仿真实验表明了所提方法是有效的多块检测方法。

（2）针对基于变量分块的故障检测方法对微小偏移、脉冲振荡等故障报警率低的问题，提出一种基于双层信息提取的多块 kNN 故障检测方法。第一层利用典型相关分析计算变量间的相关系数构建变量子块，将相关性高的变量放在一起组成子块，从而提取局部信息；第二层对各个变量子块分别提取观测信息、累计信息和变化率信息 3 种特征信息，并建立信息子块，充分挖掘子块数据的特征信息。考虑到数据的分布特征和局部区域样本的稀疏程度，采用 MDkNN 故障检测方法对各信息子块进行检测，然后采用贝叶斯方法进行融合，通过 TE 过程和实际高炉炼铁过程的仿真实验验证了所提方法的有效性。

（3）针对基于 kNN 的故障检测方法中，引发故障的异常信息易被正常信息淹没，导致故障检测不及时和报警率低的问题，提出一种基于重构误差和多块建模策略的 kNN 故障检测方法。首先利用正常工况数据集训练自编码器模型，基于该模型进行重构误差提取以解决异常信息易被淹没的问题。考虑微小偏移和振荡等故障特征，对重构误差提取观测、累计和变化率信息，建立 3 个信息子块进行多块建模并融合检测。最后通过一个数值例子和 TE 过程进行仿真与分析，表明通过求取自编码器重构误差确实能够有效地抽离出异常信息数据，从而提升了检测效果，所提方法具有一定的有效性和优越性。

5.2 前景展望

本文以经典的 kNN 方法为基础，开展面向复杂工业过程的故障检测方法研究，并提出了三种新的基于多块建模策略的方法。但是在诸多方面还需要进一步研究，未来将从以下方面进行更加深入的讨论分析。

（1）本文提出的基于自编码器重构误差的故障检测算法可有效抽离出导致故障的异常信息数据，从而提高检测效果，但是在仿真实验中发现，训练自编码器模型时常出现过拟合问题，如何有效改进自编码器结构使其更好地适用于基于 kNN 的故障检测，可作进一步研究。

（2）kNN 方法需要计算待测样本与每个样本之间的距离并查询近邻样本，计算量

大，尤其是数据量较多时，基于 kNN 的故障检测算法计算负担重，如何降低计算复杂度并提高运算速率值得进一步研究。

（3）在实际的复杂工业生产过程中，数据往往同时存在多种类型的问题，例如非线性、非高斯、含噪声、时序性、数据缺失等，且检测方法大多基于历史数据进行离线建模研究，如何改进检测模型使其能够在线实时处理不同问题共存的过程数据，有待于研究分析。

致谢

时光荏苒，流年似水。转眼之间，马上就要从学校毕业。三年的研究生时光，说长不长说短不短，终将成为我宝贵的回忆。从 2019 年那个炎热夏天开始，第一次踏入江南大学的大门，便被学校宏伟大气的图书馆大楼所吸引，这是江南学子补充精神食粮、学习知识的地方。学校深厚的知识底蕴激励着我好好利用研究生这三年的时光，争取有所成长和突破。在即将完成学位论文的这一刻，我突然意识到，自己即将要离开校园，离开这生活了三年的地方，离开朝夕相处的老师和朋友们。在这里，我由衷地感谢我的父母，我的老师，以及每个帮助过我和陪伴过我的朋友们。

首先我要感谢我的导师——熊伟丽教授。熊老师是一位博学多才、保持严谨治学态度的老师。在她的带领下，我们实验室保持着良好的学术氛围，有丰硕的科研成果。在学习上，老师给了我很多建设性的意见，研一刚进入课题组时，熊老师循循善诱，让我快速地对自己的研究方向有个清晰的认识。在撰写论文的过程中，熊老师认真严格地把关，在语句的表达上和研究的内容上给出指导意见，使我对课题有更深的思考。熊老师这种尽心竭力、严谨细致的态度从内而外地感染到我。除了学习，老师也经常关注我们的生活情况，教我们做人的道理以及如何适应竞争激烈的社会生活，当我们难以解决问题的时候即时给予帮助和辅导，同时也会参与实验室的聚餐活动，跟学生们打成一片。熊老师对我的教诲以及她的科研态度将使我受用终身。

其次，非常感谢我的父母，我的弟弟。我们这个幸福的四口之家，是我最坚强的后盾，是我停靠心灵的港湾。在我学习遇到困难时，在我找工作迷茫时，是你们的关心和照顾给了我力量，让我勇往直前。

同时我要感谢 C412 的各位小伙伴们。感谢孙文心、刘文韬、王佳宇三位博士师兄和翟超、顾炳斌、盛晓晨、费秋玲、代学志等师兄师姐们，给了我学习的榜样和指导，感谢吴晓东、赵杨、周博文等同门，一起攻克难题的时光将不会忘记，非常荣幸能够遇见你们，一起度过这三年的学习生活。感谢周阅昇、刘传玉、王新月、何罗苏阳等师弟师妹们，感谢你们的陪伴和共同努力，使我们实验室越来越好。感谢我的好朋友胡小曼同志，给了我很多陪伴，带给我很多快乐，祝愿我们都有一个美好的未来。

此外，我要感谢控制工程专业的各位领导和老师，给我们创造了良好的学习环境。

离别之际，将最好的祝福送给每一个人。

参考文献

[1] Qin S J. Survey on data-driven industrial process monitoring and diagnosis[J]. Annual Reviews in Control, 2012, 36(02): 220-234.

[2] 刘强, 卓洁, 郎自强, 等. 数据驱动的工业过程运行监控与自优化研究展望[J]. 自动化学报, 2018, 44(11): 1944-1956.

[3] 周东华, 胡艳艳. 动态系统的故障诊断技术[J]. 自动化学报, 2009, 35(06): 748-758.

[4] Yang J D, Zhang Y. Research of fault detection system combining with fault tree analysis with artificial neural network in large scale Die-Forging Process[J]. Advanced Materials Research, 2011, 317-319: 661-666.

[5] 周东华, 刘洋, 何潇. 闭环系统故障诊断技术综述[J]. 自动化学报, 2013, 39(11): 1933-1943.

[6] Frank P M. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy: A survey and some new results[J]. Automatica, 1990, 26(03): 459-474.

[7] Venkatasubramanian V, Rengaswamy R, Yin K, et al. A review of process fault detection and diagnosis: Part I: Quantitative model-based methods[J]. Computers & Chemical Engineering, 2003, 27(03): 293-311.

[8] Ding S X. Model-based fault diagnosis techniques: design schemes, algorithms, and tools[M]. Springer Science & Business Media, 2008.

[9] 周东华, 孙优贤, 席裕庚. 一类非线性系统参数偏差型故障的实时检测与诊断[J]. 自动化学报, 1993(02): 184-189.

[10] Demetriou M A, Polycarpou M M. Incipient fault diagnosis of dynamical systems using online approximators[J]. IEEE Transactions on Automatic Control, 2002, 43(11): 1612-1617.

[11] Frank P M. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy: A survey and some new results[J]. Automatica, 1990, 26(03): 459-474.

[12] Gertler J. Analytical redundancy methods in fault detection and isolation-survey and synthesis[J]. IFAC Proceedings Volumes, 1991, 24(06): 9-21.

[13] 罗天洪, 杨彩霞, 孙冬梅. 基于故障树的汽车起重机液压故障诊断专家系统[J]. 机械科学与技术, 2013, 32(04): 538-544.

[14] Chang C C, Cheng C Y. On-Line Fault Diagnosis Using the Signed Directed Graph[J]. Industrial & Engineering Chemistry Research, 1990, 29(07): 1290-1299.

[15] Lapp S A , Powers G J . Computer-aided synthesis of fault-trees[J]. IEEE Transactions on Reliability [J]. 2009, 26(1): 2-13.

[16] Jiang Q, Yan X, Zhao W. Fault detection and diagnosis in chemical processes using sensitive principal component analysis[J]. Industrial & Engineering Chemistry Research, 2013, 52(04): 1635–1644.

[17] Gunther J C, Conner J S, Seborg D E. Process monitoring and quality variable prediction utilizing PLS in industrial fed-batch cell culture[J]. Journal of Process Control, 2009, 19(05): 914-921.

[18] Li R F, Wang X Z. Dimension reduction of process dynamic trends using independent component analysis[J]. Computers and Chemical Engineering, 2002, 26(03): 467-473.

[19] Zhou Z, Wen C, Yang C. Fault detection using random projections and k-nearest neighbor rule for semiconductor manufacturing processes[J]. IEEE Transactions on Semiconductor Manufacturing, 2014, 28(01): 70-79.

[20] Jin X, Zhao M, Chow T W S, et al. Motor bearing fault diagnosis using trace ratio linear discriminant analysis[J]. IEEE Transactions on Industrial Electronics, 2013, 61(05): 2441-2451.

[21] Kramer M A. Nonlinear principal component analysis using autoassociative neural networks[J]. AIChE journal, 1991, 37(02): 233-243.

[22] Schölkopf B, Smola A, Müller K R. Nonlinear component analysis as a kernel eigenvalue problem[J]. Neural computation, 1998, 10(05): 1299-1319.

[23] Zhang Y, Ma C. Decentralized fault diagnosis using multiblock kernel independent component analysis[J]. Chemical Engineering Research and Design, 2012, 90(05): 667-676.
[24] Godoy, José L, Zumoffen D A , et al. New contributions to non-linear process monitoring through kernel partial least squares[J]. Chemometrics and Intelligent Laboratory Systems, 2014, 135: 76-89.
[25] Khediri I B, Limam M, Weihs C. Variable window adaptive kernel principal component analysis for nonlinear nonstationary process monitoring[J]. Computers & Industrial Engineering, 2011, 61(03): 437-446.
[26] 邓佳伟, 邓晓刚, 曹玉苹, 张晓玲. 基于加权统计局部核主元分析的非线性化工过程微小故障诊断方法[J]. 化工学报, 2019, 70(07): 2594-2605.
[27] 周卫庆, 司风琪, 徐治皋, 黄葆华, 仇晓智. 基于 KPCA 残差方向梯度的故障检测方法及应用[J]. 仪器仪表学报, 2017, 38(10): 2518-2524.
[28] Ge Z, Zhang M, Song Z. Nonlinear process monitoring based on linear subspace and Bayesian inference[J]. Journal of Process Control, 2010, 20(05): 676-688.
[29] 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.
[30] Rongyu L I, Gang R. Fault isolation by partial dynamic principal component analysis in dynamic process[J]. Chinese Journal of Chemical Engineering, 2006, 14(04): 486-493.
[31] Choi S W, Lee I B. Nonlinear dynamic process monitoring based on dynamic kernel PCA[J]. Chemical Engineering Science, 2004, 59(24): 5897-5908.
[32] Ge Z , Chen X. Supervised linear dynamic system model for quality related fault detection in dynamic processes[J]. Journal of Process Control, 2016, 44:224-235.
[33] 翟坤, 杜文霞, 吕锋, 辛涛, 句希源. 一种改进的动态核主元分析故障检测方法[J]. 化工学报, 2019, 70(02): 716-722.
[34] 李元, 马雨含, 郭金玉. 基于动态多向局部离群因子的在线故障检测[J]. 计算机应用研究, 2017, 34(11): 3259-3261+3266.
[35] 魏域琴, 宋丹丹, 翁正新. 基于 DPCA-KNN 的工业过程故障诊断方法研究[C]//第三十八届中国控制会议论文集(7). 广州, 2019: 14−19.
[36] 郭小萍, 徐月, 李元. 基于特征空间自适应 k 近邻工业过程故障检测[J]. 高校化学工程学报, 2019, 33(02): 453-461.
[37] Li R F, Wang X Z. Dimension reduction of process dynamic trends using independent component analysis[J]. Computers and Chemical Engineering, 2002, 26(03): 467-473.
[38] Xu Y, Shen S Q, He Y L, et al. A novel hybrid method integrating ICA-PCA with relevant vector machine for multivariate process monitoring[J]. IEEE Transactions on Control Systems Technology, 2018, 27(04): 1780-1787.
[39] 田学民, 蔡连芳. 一种基于 KICA-GMM 的过程故障检测方法[J]. 化工学报, 2012, 63(09): 2859-2863.
[40] Ge Z Q, Song Z H. Performance-driven ensemble learning ICA model for improved non-Gaussian process monitoring[J]. Chemometrics and Intelligent Laboratory Systems, 2013, 123:1-8.
[41] 王振雷, 江伟, 王昕. 基于多块 MICA-PCA 的全流程过程监控方法[J]. 控制与决策, 2018, 33(02): 269﹣274.
[42] He P Q，Wang J. Fault detection using the k-nearest neighbor rule for semiconductor manufacturing processes[J].IEEE Transactions on Semiconductor Manufacturing, 2007, 20(04) : 345-354.
[43] 张成, 郭青秀, 冯立伟, 等. 基于局部保持投影-加权 k 近邻规则的多模态间歇过程故障检测策略[J]. 控制理论与应用, 2019, 36(10): 1682-1689.
[44] 郭金玉, 刘玉超, 李元. 基于局部相对概率密度 kNN 的多模态过程故障检测[J]. 高校化学工程学报, 2019, 33(01): 159-166.

[45] 刘腾飞. 基于深层自编码器的发酵过程故障监测[D]. 北京：北京工业大学, 2020: 54-64.
[46] Jiang Q, Huang B. Distributed monitoring for large-scale processes based on multivariate statistical analysis and Bayesian method[J]. Journal of Process Control, 2016, 46: 75-83.
[47] 邓晓刚, 徐莹. 基于贝叶斯 ICA 的多工况非高斯过程故障检测[J]. 控制工程, 2018, 25(03): 402-407.
[48] 谭帅, 王福利, 常玉清, 王姝, 周贺. 基于差分分段 PCA 的多模态过程故障监测[J]. 自动化学报, 2010, 36(11): 1626-1636.
[49] Ma H, Hu Y, Shi H. A novel local neighborhood standardization strategy and its application in fault detection of multimode processes[J]. Chemometrics and Intelligent Laboratory Systems, 2012, 118: 287-300.
[50] Peng X, Tang Y, Du W. Multimode process monitoring and fault detection: a sparse modeling and dictionary learning method[J]. IEEE Transactions on Industrial Electronics, 2017, 64(06): 4866-4875.
[51] 冯立伟, 张成, 李元, 等. 基于标准距离 k 近邻的多模态过程故障检测策略[J]. 控制理论与应用, 2019, 36(04):553-560.
[52] MacGregor J F, Jaeckle C, Kiparissides C, et al. Process monitoring and diagnosis by multiblock PLS methods[J]. AIChE Journal, 1994, 40(5): 826-838.
[53] Ge Z Q, Song Z H. Distributed PCA model for plant-wide process monitoring[J]. Industrial & Engineering Chemistry Research, 2013, 52(05): 1947-1957.
[54] Jiang Q, Yan X, Huang B. Performance-driven distributed PCA process monitoring based on fault-relevant variable selection and Bayesian inference[J]. IEEE Transactions on Industrial Electronics, 2015, 63(01): 377-386.
[55] Jiang Q, Yan X. Plant-wide process monitoring based on mutual information-multiblock principal component analysis[J]. ISA transactions, 2014, 53(05): 1516-1527.
[56] Jiang Q, Wang B, Yan X. Multiblock independent component analysis integrated with hellinger distance and Bayesian inference for non-Gaussian plant-wide process monitoring[J]. Industrial & Engineering Chemistry Research, 2015, 54(09): 2497-2508.
[57] 石怀涛, 王雨桐, 李颂华, 等. 基于多块相对变换独立主元分析的故障诊断方法 [J]. 控制与决策, 2018, 33(11):2009-2014.
[58] 童楚东, 史旭华. 基于互信息的 PCA 方法及其在过程监测中的应用[J]. 化工学报, 2015, 66(10): 4101-4106.
[59] Ge Z, Song Z. Multimode process monitoring based on Bayesian method[J]. Journal of Chemometrics: A Journal of the Chemometrics Society, 2009, 23(12): 636-650.
[60] Yin S, Ding S X, Haghani A, et al. A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process[J]. Journal of Process Control, 2012, 22(09): 1567-1581.
[61] 孔祥玉, 解建, 罗家宇, 李强. 基于改进高效偏最小二乘的质量相关故障诊断[J]. 控制理论与应用, 2020, 37(12): 2645-2653.
[62] 王晓慧, 王延江, 邓晓刚, 张政. 基于加权深度支持向量数据描述的工业过程故障检测[J]. 化工学报, 2021, 72(11): 5707-5716.
[63] 顾炳斌, 熊伟丽. 基于多块信息提取的 PCA 故障诊断方法[J]. 化工学报, 2019, 70(02): 316-329.
[64] 董洁, 孙瑞琪, 彭开香, 唐鹏. 自动编码器与典型相关分析方法联合驱动的工业过程质量检测[J]. 控制理论与应用, 2019, 36(09): 1493-1500.
[65] Mei J, Liu M, Wang Y F, et al. Learning a mahalanobis distance-based dynamic time warping measure for multivariate time series classification[J]. IEEE transactions on Cybernetics, 2015, 46(6): 1363-1374.
[66] 马贺贺, 胡益, 侍洪波. 基于距离空间统计量分析的多模态过程无监督故障检测[J]. 化工学报, 2012, 63(03): 873-880.

[67] 邓佳伟, 邓晓刚, 曹玉苹, 张晓玲. 基于加权统计局部核主元分析的非线性化工过程微小故障诊断方法[J]. 化工学报, 2019, 70(07): 2594-2605.

[68] Shen B, Ge Z. Supervised Nonlinear Dynamic System for Soft Sensor Application Aided by Variational Auto-encoder[J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(09): 6132-6142.

[69] Chen Z, Li W. Multisensor Feature Fusion for Bearing Fault Diagnosis using Sparse Autoencoder and Deep Belief Network[J]. IEEE Transactions on Instrumentation and Measurement, 2017, 66(07): 1693-1702.

附录：作者在攻读硕士学位期间发表的论文

[1] 郑静, 熊伟丽. 基于互信息的多块 k 近邻故障监测及诊断[J]. 智能系统学报, 2021, 16(04): 717-728. (CSCD 中文核心)

[2] 郑静, 熊伟丽, 吴晓东. 基于重构误差和多块建模策略的 kNN 故障监测[J]. 系统仿真学报, 已录用.(CSCD 中文核心)

[3] 熊伟丽, 郑静. 国家发明专利(CN202011060648.X): 基于互信息的多块k近邻故障监测及诊断. (实质审查)