

DOI:10.19651/j.cnki.emt.2518657

基于 KmeansSMOTE 与领域自适应的 化工过程故障诊断*

卫亚楠 谢俊 吕凯 朱荣 李财年

(新疆大学机械工程学院 乌鲁木齐 830017)

摘要: 针对化工过程中由于多变量传感数据的时间依赖性导致的特征耦合、工况变化带来的数据分布偏移以及样本数据不平衡等问题,提出了将 K 均值合成少数类过采样技术与条件对抗域适应相结合的化工过程故障诊断框架。首先,通过时间窗口分割技术将原始一维数据转换为多个二维时间窗口数据,在这些窗口内采用 KmeansSMOTE 方法对少数类故障样本进行扩充,扩充后得样本能够保留完整的时序故障特征,同时该算法还能减少生成噪声样本的数量;然后,使用领域自适应技术对齐源域与目标域的特征分布,减少两者间的分布差异,使得基于源域训练的故障诊断模型能够在新工况下有效识别故障类别;最后,通过使用田纳西——伊斯曼过程的故障数据进行诊断实验,并通过与 CDAN、DANN 以及 JDA 等模型进行诊断率对比,验证了所提方法的有效性。

关键词: 合成采样;迁移学习;故障诊断;化工过程;田纳西伊斯曼数据集

中图分类号: TP277;TQ086;TN911 **文献标识码:** A **国家标准学科分类代码:** 520.2040

Fault diagnosis of chemical processes based on KmeansSMOTE and domain adaptation

Wei Yanan Xie Jun Lyu Kai Zhu Rong Li Cainian

(School of Mechanical Engineering, Xinjiang University, Urumqi 830017, China)

Abstract: A fault diagnosis framework for chemical processes is proposed, which combines K-means synthetic minority oversampling technique with conditional adversarial domain adaptation to address issues such as feature coupling caused by the temporal dependence of multivariate sensing data, data distribution shift caused by changes in operating conditions, and imbalanced sample data in chemical processes. Firstly, the original one-dimensional data is converted into multiple two-dimensional time window data using time window segmentation technology. Within these windows, the Kmeans SMOTE method is used to expand the minority class fault samples. The expanded samples can retain the complete temporal fault features, and this algorithm can also reduce the number of generated noise samples; then, domain adaptation techniques are used to align the feature distributions of the source domain and the target domain, reducing the distribution differences between the two and enabling the fault diagnosis model trained on the source domain to effectively identify fault categories under new operating conditions; finally, diagnostic experiments were conducted using fault data from the Tennessee Eastman process, and the effectiveness of the proposed method was validated by comparing its diagnostic rates with models such as CDAN, DANN, and JDA.

Keywords: synthetic sampling; transfer learning; fault diagnosis; chemical process; Tennessee Eastman dataset

0 引言

安全生产是化工企业的生命线。随着化学过程的复杂性和规模不断增加,事故频发给行业带来重大损失^[1]。在化工过程中,故障通常表现为过程变量与正常操作的偏差,

这些偏差可能由设备故障、操作错误或参数变化引起^[2-3]。尽管故障的根本原因可能涉及少数因素或警报变量,但其负面影响往往会迅速蔓延至整个制造系统^[4]。因此,在发生故障时能迅速锁定关键变量,且根据变量及时且准确地识别故障,成为故障诊断的核心任务^[5]。然而,化工过程的

收稿日期:2025-04-22

* 基金项目:新疆维吾尔自治区重点研发计划项目(2023B01031-2)资助

复杂性和故障的传播使得这一任务变得异常困难^[6]。

为应对这些挑战,智能故障诊断技术正日益成为识别关键干扰因素、防止重大事故升级的重要手段^[7-8]。近年来基于数据驱动的故障诊断方法在工业过程中展现出优异的性能^[9]。与传统的基于模型的诊断方法相比,数据驱动方法在不依赖于精确的数学模型的情况下,通过分析大量的过程数据来检测和诊断故障^[10]。这些方法可以分为统计方法和机器学习和深度学习^[11],其中深度学习由于其强大的挖掘数据特征的能力,在大规模数据集上取得优异的性能^[12]。例如,Wang等^[13]提出了使用深度学习多模型融合的故障诊断方法,该方法使用长短期记忆网络(long short term memory network, LSTM)和卷积神经网络(convolutional neural network, CNN)分别提取特征并进行特征融合,使用多层分类器(multi-layer perceptron, MLP)作为进一步的特征压缩和提取输入,最后输出诊断结果。Han等^[14]提出了一种基于正交自关注卷积自编码器(orthogonal self attention convolutional autoencoder, OSCAE)的故障诊断方法,使用卷积自编码器和正交注意力机制相结合,实现故障状态的精确检测。Cao等^[15]提出了基于双层动态的循环神经网络(independently recurrent neural network, INDRNN)的故障诊断方法,该方法旨在使用两层分类器来识别且提取动态特征,避免了传统的INDRNN的故障诊断方法无法提取动态特征的问题,在Tennessee Eastman(TE)过程上的平均诊断率可达90%。尽管基于深度学习的模型在化工故障诊断中表现优异,但其泛化能力不足仍然是限制其实际应用的主要障碍。这些模型都会要求训练和测试数据具有相同的概率分布,但事实上经常会因为一些情况而出现数据分布差异过大的问题,导致已有模型的诊断性能变差^[16]。

为了解决这些问题,迁移学习被引入到化工故障诊断领域^[17]。如Wu等^[18]提出了基于迁移学习的多模式化工过程诊断方法,有效解决了一些模式下的故障数据稀有性和无标签等问题并通过TE过程展示了所提方法的有效性。Qin等^[19]提出了一个多元域自适应网络,通过多项式核诱导的最大均值差异的多通道域自适应实现了多个域联合分布对齐,提高模型的泛化性能。Jiang等^[20]提出了一种基于潜在时间特征的迁移学习方法,该方法利用自编码器和强化学习优化特征选择,显著提高在目标域诊断模型的准确性。

上述基于迁移学习的化工故障诊断研究都建立在目标域和源域中存在着大量的数据的基础上。但是在真实的工业条件下,因为故障发生是小概率事件,因此会导致新工况下的故障数据与正常数据的比例失衡,此外,由于多变量传感数据的时间依赖性,不同传感变量之间可能存在复杂的特征耦合关系,直接进行数据扩充可能会导致生成数据失真。同时,新工况与原工况之间的工况变化可能导致数据分布偏移,使得模型在新工况下的泛化能力受限导致上述

方法失效。因此如何使用新工况下庞大的正常数据和稀少的故障数据来建立有效的故障诊断模型,并且该模型可以成功有效的在新工况中进行应用是一个具有挑战性的问题。化工过程数据因其动态性和强时序依赖性使得完成特征提取这一任务变得艰难,对此,Wang等^[21]提出了使用LSTM+CNN的方法来对时序特征进行提取,该方法虽然可以有效的提取数据的时序特征,但是所需的计算量过于庞大,不利于进行快速的进行化工故障诊断的任务。为了解决这些问题,本研究提出了一种新的故障诊断方法,采用KmeansSMOTE的数据处理方法和条件对抗域适应的故障诊断方法(KSCDAN)的新模型,该模型的具体描述为两部分:首先增加样本数量,故障数据样本的数量增加可以使诊断模型轻易的学习到故障特征,便于进行故障的有效诊断。然后使用条件域自适应的方法来进行诊断,可以适应不同情况下的故障类型,减少建立不必要的故障网络,大大降低工作量。最后,通过田纳西-伊斯曼化工过程进行模拟实验,验证了所提出方法的有效性。

1 滑动窗口构建与基于KmeansSMOTE的数据增强方法

1.1 滑动窗口方法

化工过程中的数据往往是随着时间而采集的,因此这些数据自带时间顺序,如果随意的对其进行处理,可能会丢失其时间特征,因此采用时间滑动窗口的方法来保证所提取出的数据的完整,运用时间窗口时会将原本的一维数据转变为二维数据,但是TE数据中的变量之间存在着较大的差异,这些差异可能会导致某些变量在数值上主导特征提取,从而影响下一部分工作,因此需要对源域数据按照式(1)所示来进行标准化。

$$\mathbf{X}_{norm} = \frac{\mathbf{X} - \mu}{\sigma} \quad (1)$$

其中, \mathbf{X} 代表原始数据向量(长度为 N ,对应某一变量的 N 个采样点), μ 代表该变量的均值, σ 表示该变量的标准差, \mathbf{X}_{norm} 表示标准化后的数据向量。

在进行标准化以后,执行如图1所示的滑动窗口采样操作。形成的二维数据一维为变量维,长度为50;另一维为时间维,长度为4 800,即80 h。

1.2 KmeansSMOTE合成过采样方法数据扩充

在复杂的化工生产过程中,由于实际生产的特性,难以获得故障工况下的数据,因此如何在小样本和不平衡数据下进行化工故障诊断,引发了思考。

为了解决这个问题,数据增强吸引了人们的注意,在2002年,Chawla等^[22]提出了一种新的过采样数据处理方法,称为合成少数过采样技术(SMOTE),该方法已成为针对当前不平衡数据开发过采样方法的先驱。SMOTE可以通过减少或最小化过拟合来改进过采样方法。SMOTE的必要过程是创建引用相邻少数类实例的新合成数据。新的

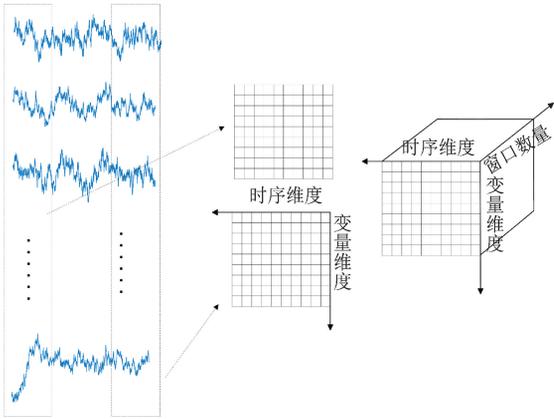


图 1 滑动窗口采样

Fig. 1 Sliding window sampling

合成数据用于确定最接近的少数实例,并在样本数据和选定的少数实例之间建立插值线。沿线合成的数据可以通过在邻域中引入新的少数类实例来增加少数类实例的数量,从而帮助分类器提高泛化能力。但是这种简单的合成方法可能会在不知情的情况下进行随机选择样本,产生的样本中可能会随机产生一些噪音样本来污染原始样本数据,导致模型出现过拟合的问题^[23],为了处理这种问题,在过采样过程前使用 Kmeans 方法替代原来的进行聚类,形成一个 KmeansSMOTE 的数据处理框架。

在 KmeansSMOTE 框架中,摒弃了传统的 SMOTE 中随机选择少数类样本邻近的 K 个样本,而是使用 Kmeans 聚类算法将少数类样本分成多个簇,每个簇代表一类相似样本,簇内的样本具有相似的特征。在这个过程中, Kmeans 算法会根据样本间的距离将它们划分到不同的簇中,目标是让簇内的样本尽可能相似,而簇与簇之间的差异尽可能大。在选定的簇中,使用 SMOTE 方法生成新的合成样本。对于选定的簇 C_i , KmeansSMOTE 的步骤如下:

步骤 1) 首先输入数据集 \mathbf{X} 并且对数据进行标准化处理。

步骤 2) 将得到的数据利用 Kmeans 算法将数据集分为若干个簇,同时使用式(2)来最小化簇内平方误差。

$$SSE = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (2)$$

其中, C_i 是第 i 个簇, μ_i 是簇 i 的中心, $\|x - \mu_i\|$ 是样本 x 到簇中心的欧氏距离。

步骤 3) 对于得到的若干个簇,通过计算不平衡率来考虑选择哪一些簇来进行过采样,不平衡率的公式如下:

$$IR_i = \frac{N_{maj}^i}{N_{min}^i} \quad (3)$$

其中, N_{maj}^i 表示第 i 个簇中数量较多的样本, N_{min}^i 表示第 i 个簇中数量较少的样本。

挑选出不平衡率小于 1 的簇进行过采样,得到一个新的合成样本 x_{new} :

$$x_{new} = x + \lambda(x' - x) \quad (4)$$

其中, λ 是一个随机数, $\lambda \in [0, 1]$, 用于控制插值的位置; $x' - x$ 表示样本 x 和 x' 之间的差异变量

该算法原理如图 2 所示。

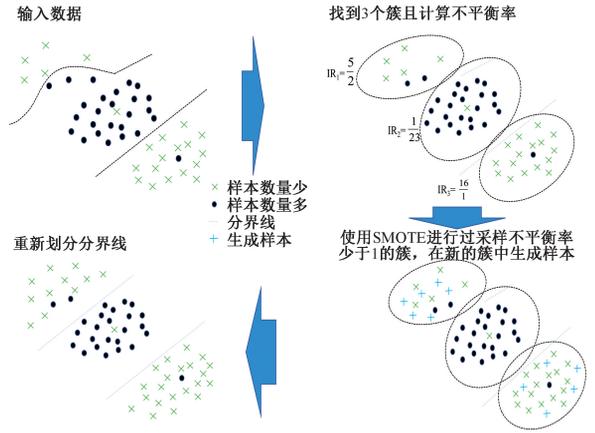


图 2 合成采样算法

Fig. 2 Composite sampling algorithm

2 基于 KSCDAN 方法的故障诊断方法

2.1 条件对抗域适应方法

条件对抗域适应的目标域是将从源域学到的知识迁移到目标域,对于分布不同的两个域:源域 D_S 和目标域 D_T , 其中, $D_S = \{(x_i^s, y_i^s)\}_{n_i^s=1}$, 其中 x_i^s 是源域样本, y_i^s 是对应的标签; 目标域数据为 $D_T = \{(x_i^t, y_i^t)\}_{n_i^t=1}$ 其中 x_i^t 是目标域样本。通过对这些数据进行训练出一个模型 f , 使其在目标域上的性能接近源域。

领域自适应的关键点在于引入了条件信息来进行增强对抗训练,使用下述操作来进行条件对抗训练:

1) 需要对划分好的数据进行特征的提取与分类,对于输入样本 x , 特征提取器生成特征表示 $h = G(x)$ 。分类器 C 基于特征 h 预测类别概率 $p = C(h)$ 。

2) 将特征 h 和类别预测 p 结合起来,形成条件特征表示 $h \otimes p$ 其中 \otimes 表示外积。条件特征表示为:

$$h \otimes p = h \cdot p^T \quad (5)$$

3) 将得到的特征送入领域判别器 D 中,领域判别器 D 接收条件特征表示 $h \otimes p$, 并预测特征来自源域还是目标域。其损失函数为:

$$\mathcal{L}_{adv} = \mathbf{E}_{x \sim D_S} [\log D(h \otimes p)] + \mathbf{E}_{x \sim D_T} [\log(1 - D(h \otimes p))] \quad (6)$$

4) 考虑到可能会出现模型在源域上的分类性能可能不佳的问题,引入分类损失:

$$\mathcal{L}_{cls} = \mathbf{E}_{(x, y) \sim D_S} [\text{CrossEntropy}(C(G(x)), y)] \quad (7)$$

故 CDAN 的总损失函数分为分类损失和对抗损失的加权和:

$$\mathcal{L}_{total} = \mathcal{L}_{cls} + \gamma \mathcal{L}_{adv} \quad (8)$$

其中, γ 是权衡参数。

为了进一步评估模型对源域与目标域分布的对齐效果,本文引入最大均值差异(maximum mean discrepancy, MMD)作为度量指标,该指标的定义

$$MMD^2(\mathbf{D}_s, \mathbf{D}_t) = \frac{1}{n(n-1)} \sum_{i \neq j}^n k(x_i, x_j) + \frac{1}{m(m-1)} \sum_{i \neq j}^m k(y_i, y_j) - \frac{2}{nm} \sum_{i,j} k(x_i, y_j) \quad (9)$$

其中, $k(\cdot)$ 是核函数, x_i 和 y_i 分别属于 \mathbf{D}_s 和 \mathbf{D}_t , n 和

m 分别是两个数据集的样本数。

2.2 KSCDAN 故障诊断模型

本文提出的 KSCDAN 的化工过程故障诊断模型框架如图 3 所示。在域适应网络中使用双层卷积神经网络来提取特征,并且通过引入条件分类器和梯度反转层来实现域自适应,该方法通过对抗学习机制对其源域与目标域在特征和类别预测的联合分布,有效保证了特征表示的域无关性与类别判别性,提高了模型在目标域上的泛化性能。

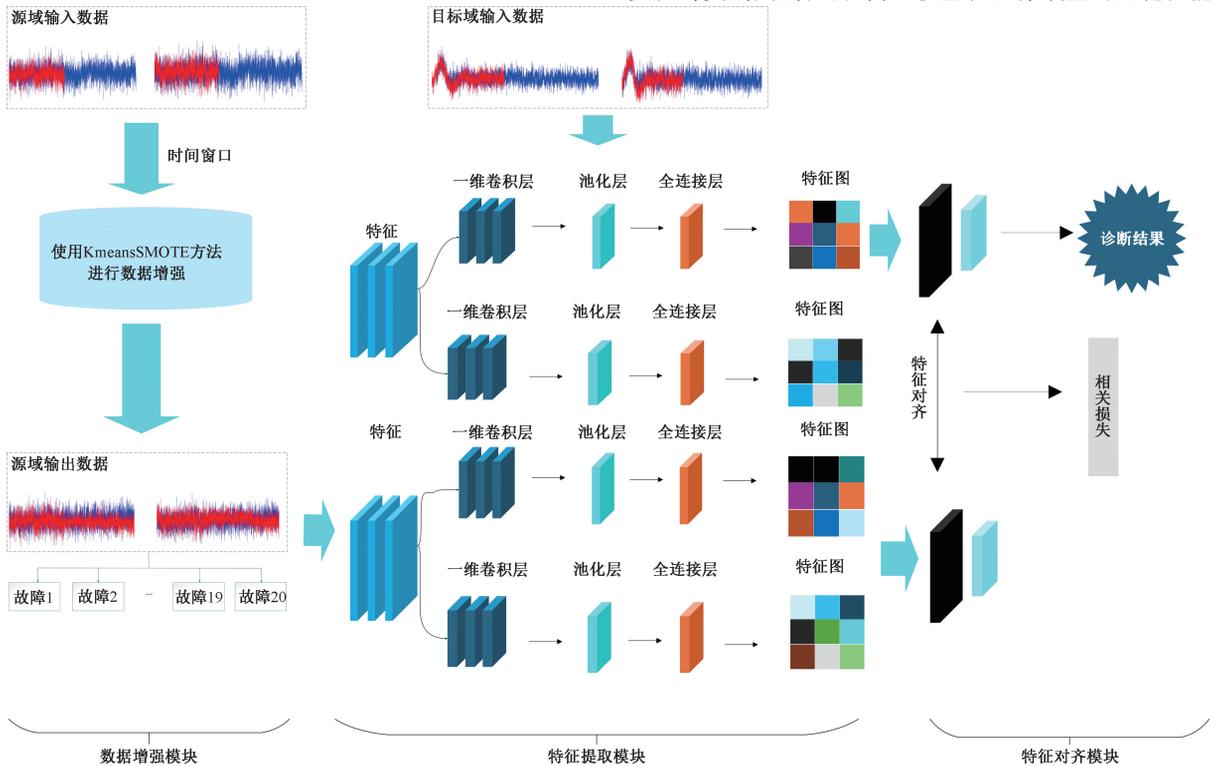


图 3 KSCDAN 模型框架

Fig. 3 KSCDAN model framework

总体的故障诊断流程如图 4 所示,该流程分为两个阶段:离线建模和在线阶段。

离线建模步骤如下:

步骤 1) 收集一个包含多模态条件下正常和故障数据的训练数据集,或者从类似化工过程的不同操作单元中获取数据,获取到的数据集应包括源域和目标域的多种工况,以确保模型有足够多的信息来进行特征的提取和学习。

步骤 2) 评估获取到的数据集中的故障数据是否少于正常数据,如果故障样本稀缺,应用时间窗口技术,将数据分为若干个时间窗口,在每个时间窗口内使用 KmeansSMOTE 方法来进行数据增强以扩大数据集。

步骤 3) 使用源域数据的均值和标准偏差来进行数据归一化,保证所有变量的数值范围一致,避免某些特征对模型影响过大。

步骤 4) 采用 CNN 来提取时序特征,找到最具区分性

的特征,提高模型的分类能力。

步骤 5) 通过调整模型的超参数选择出最优参数,找到实现最好性能的模型。

步骤 6) 构建出最优模型并保存训练后的模型。

在线诊断步骤如下:

步骤 1) 针对目标域的状况进行实时数据采集。

步骤 2) 执行与离线阶段相同的数据归一化步骤,保证在线数据与离线数据分布一致。

步骤 3) 加载离线阶段训练好的模型并使用目标域的标签信息进行有监督的领域自适应,进一步优化模型。

步骤 4) CDAN 模型基于提取到的特征对目标域进行无监督分析,输出故障诊断结果。

3 案例分析

为了验证所提化工过程故障诊断与识别的方法的有效性,使用 TE 过程数据集来进行实验验证。

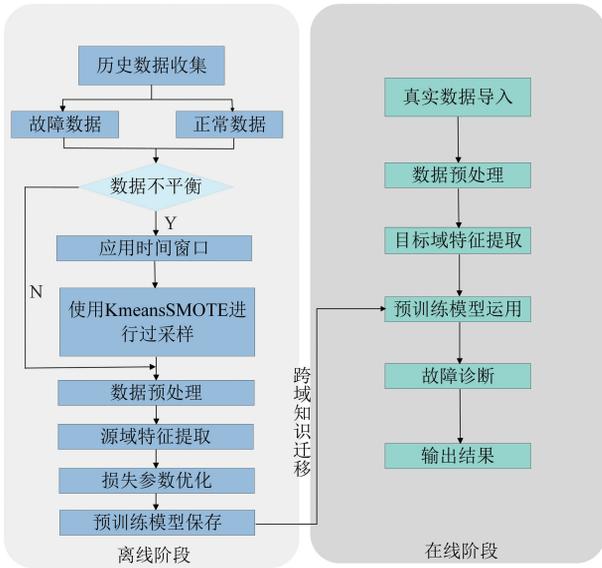


图 4 诊断算法流程

Fig. 4 Diagnostic algorithm process

3.1 数据采集

TE 过程是由美国田纳西公司所开发的一个经典的可用于化工过程的仿真模型,近年来被广泛的用来进行对化工生产中的故障诊断与检测^[24-25] TE 过程流程图如图 5 所示。

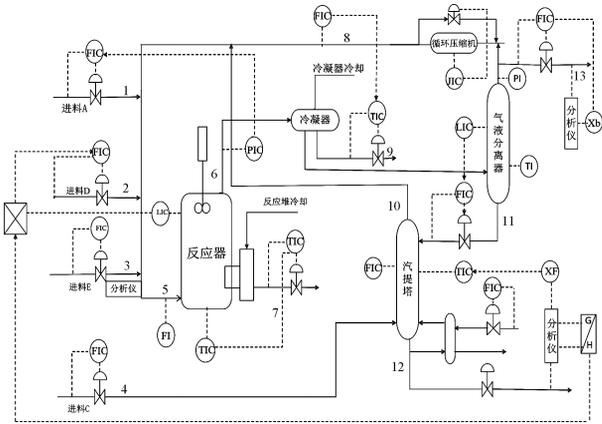


图 5 TE 过程工艺流程图

Fig. 5 TE process flow chart

本实验的数据在 MATLAB 平台上的 simulink 模块中进行仿真得到,该流程主要包括五个操作单元:反应器、冷凝器、循环压缩机、气液分离器和解析塔,整个过程包括 12 个操作变量和 41 个测量变量,各种故障类型如表 1 所示。

本实验中选取模态一作为源域,模态二作为目标域。不同的模态如表 2 所示。数据集获取在 MATLAB 仿真平台下采集,将采集到的数据按照 8:2 的比例划分为训练集和测试集,数据集的采样间隔为 1 min,训练集运行 80 h 获得,测试集中运行 12 h 获得,其中正常样本正常运行 12 h,

表 1 故障类型

Table 1 Fault types

故障编号	故障状态	故障类型
故障 1	A/C 进料比例变化, B 成分不变	阶跃型
故障 2	B 成分变化, A/C 进料比例不变	阶跃型
故障 3	D 进料温度变化	阶跃型
故障 4	反应器冷却水入口温度变化	阶跃型
故障 5	冷凝器冷却水入口温度变化	阶跃型
故障 6	A 进料损失	阶跃型
故障 7	C 存在压力损失, 可用性降低	阶跃型
故障 8	A、B、C 进料成分变化	随机型
故障 9	D 进料温度变化	随机型
故障 10	C 进料温度变化	随机型
故障 11	反应器冷却水入口温度变化	随机型
故障 12	冷凝器冷却水入口温度变化	随机型
故障 13	反应动力学常数变化	缓慢偏移型
故障 14	反应器冷却水阀门	粘滞型
故障 15	冷凝器冷却水阀门	粘滞型
故障 16~20	未知故障	未知类型

故障样本中先正常运行 2 h, 在引入故障继续运行 10 h 获得。因为采样频率为每分钟一次, 将时间窗口设置为 10, 可以覆盖故障发生前后连续十分钟的传感器变化情况, 因此, 十分钟的窗口长度能够在保证模型及时响应的同时, 充分捕捉到故障初期的动态变化特征。因此输入数据的大小为 10×50。同时使用故障诊断率 (FDR) 来判断模型的诊断性能。

表 2 模态信息

Table 2 Modal information

模态	G/H 质量比	目标产量
1	50/50	14 076
2	10/90	14 077

3.2 实验结果分析

为验证实验结果的有效性, 将目标域的测试集数据作为故障样本进行预处理后输入到已建立的诊断模型中, 所得到的结果如图 6 所示。虽然本文采用了 CDAN 结构在模型中隐式优化领域对齐目标, 但为了进一步验证特征对齐效果, 显示计算中间层 MMD 值作为补充验证指标, 所得到的结果如图 7 所示。

采用传统的 CDAN、域对抗神经网络 (domain adversarial neural network, DANN) 模型以及联合分布适应 (joint distributed adaptation, JDA) 不同的诊断方法与本文所提方法对 TE 过程中的故障进行的诊断结果如表 3 所示, 由表 3 可知, KSCDAN 模型的整体故障诊断准确率为 96.8%, 与原始的 CDAN 方法相比, 故障诊断率提高了

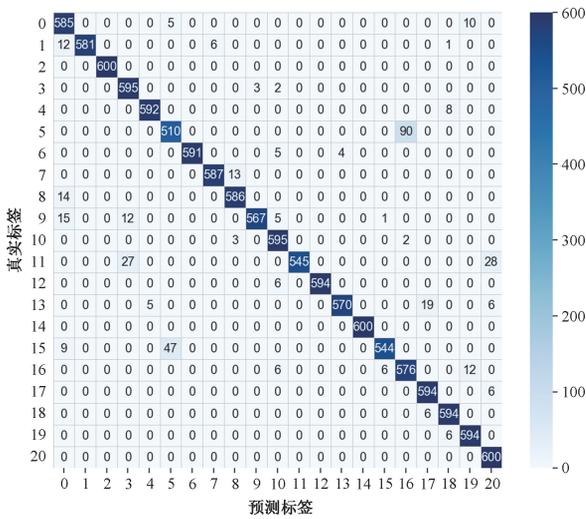


图6 混淆矩阵

Fig. 6 Confusion matrix

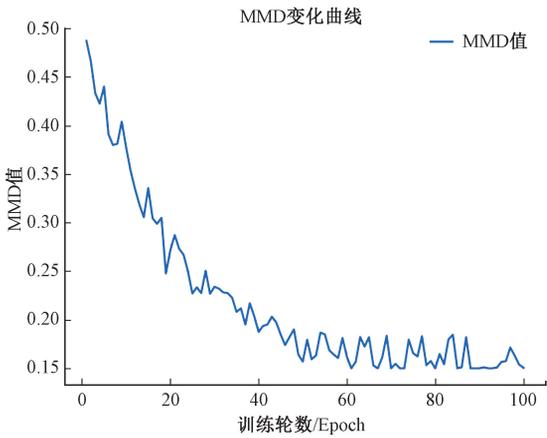


图7 MMD图

Fig. 7 MMD diagram

6.4%,其中对于故障3、9和15等以往难以准确诊断的故障,对于故障3和15的诊断率较高,对故障9的诊断率低于CDAN方法但是高于DANN与JDA方法。因这几种故障是与温度相关联的工艺参数之间的相互作用也存在滞后性,比如故障3为D的温度变化,该故障类型为阶跃型,参数会随着温度的变化而变化,反应物D和反应器之间的温度变化存在滞后性,当反应物D的温度发生变化时,并不能及时的反映反应器的各种参数变化情况,这样就会导致对该故障的诊断率低和不准确等问题,但是使用研究所提出的方法可达到99.2%的诊断率。同时故障2、14和20故障诊断率达到了100%,说明模型对这3种故障有较高的敏感度。

在实际的化工生产过程中,故障诊断时间是一个关键指标。诊断时间越短,巡检人员就能越及时地发现并处理故障,有助于提高生产效率、降低风险,并尽可能减少故障对系统运行的影响。表4展示了每个故障的诊断时间。

表3 实验结果

Table 3 Experimental results

故障模式	KSCDAN	JDA	CDAN	DANN
故障1	96.8	90.0	99.2	83.3
故障2	100	98.6	99.2	98.3
故障3	99.2	81.2	88.2	91.7
故障4	98.7	88.0	99.2	97.6
故障5	84.9	85.7	96.0	87.5
故障6	98.4	91.0	82.8	97.5
故障7	97.8	83.3	92.3	85.7
故障8	97.6	92.8	91.2	92.2
故障9	94.5	60.3	98.4	68.4
故障10	99.2	99.0	99.2	95.4
故障11	90.9	86.1	81.2	98.4
故障12	99.2	85.2	99.2	95.6
故障13	95.0	88.7	88.0	95.7
故障14	100	91.5	99.2	98.7
故障15	90.6	80.3	82.4	28.0
故障16	96.0	92.8	87.5	44.2
故障17	99.2	87.5	85.1	83.3
故障18	99.2	80.9	81.2	93.9
故障19	99.2	90.9	82.0	98.6
故障20	100	80.6	92.3	93.3
平均	96.8	86.7	90.4	91.8

表4 故障诊断时间

Table 4 Fault diagnosis time

故障模式	KSCDAN	CDAN	JDA	DANN
故障1	98	113	231	92
故障2	49	133	151	140
故障3	79	140	149	72
故障4	51	157	130	146
故障5	45	131	264	134
故障6	97	151	330	89
故障7	71	74	261	94
故障8	57	133	193	186
故障9	115	136	345	160
故障10	92	58	137	107
故障11	90	91	171	132
故障12	72	53	309	92
故障13	106	61	174	191
故障14	96	63	194	199
故障15	92	147	284	130
故障16	96	130	153	194
故障17	81	97	270	198
故障18	118	131	147	148
故障19	61	97	154	116
故障20	79	139	252	112
平均	82.17	111.72	215.02	136.43

结果表明,相较于其他方法,所提出的 KSCDAN 方法有着出色的诊断速度。

4 结 论

针对传统诊断模型在面对样本不平衡问题,往往由于故障样本和正常样本的数量比例差异过大,导致特征不能被模型有效的提取出来,因此提出了使用 KmeansSMOTE 的数据增强方法来保证正常样本和故障样本数量的均衡,大大增强了诊断模型提取特征的能力,同时考虑到频繁建立诊断模型所产生的工作量以及在改变工况时,模型可能会出现适应性差的问题,引入迁移学习的思想,通过将某一数据丰富领域(源领域)中训练好的模型迁移到数据相对稀缺的新领域(目标领域),可以在较少样本的条件下迅速构建高效的故障诊断模型。实验结果表明该模型对于多故障模式,KSCDAN 模型有着较高的诊断率,该方法的平均诊断率高于其他 3 种方法,采用该方法可以明显提高故障诊断率,从而为安全化工生产带来有效的依据。

参考文献

- [1] RAJEEVAN A, FAISAL K, SAILM A, et al. An analysis of process fault diagnosis methods from safety perspectives[J]. *Computers & Chemical Engineering*, 2021, 145: 107197.
- [2] MOHAMMAD A, FAISAL K, SYED I, et al. A robust neural network model for fault detection in the presence of mislabelled data[J]. *The Canadian Journal of Chemical Engineering*, 2024, 102(4): 1368-1380.
- [3] BAI M, QI M, SHU C M, et al. Why do major chemical accidents still happen in China: Analysis from a process safety management perspective[J]. *Process Safety and Environmental Protection*, 2023, 176: 411-420.
- [4] CHEN Y, TONG C D, LAN T. Fault monitoring for chemical processes using neighborhood embedding discriminative analysis [J]. *Process Safety and Environmental Protection*, 2022, 164: 109-118.
- [5] CHIANG L H, RUSSELL E L, BRAATZ R D. *Fault detection and diagnosis in industrial systems* [M]. Springer Science & Business Media, 2012.
- [6] HARINARAYAN R R A, SHALINIE S M. XFDDC: eXplainable fault detection diagnosis and Correction framework for chemical process systems[J]. *Process Safety and Environmental Protection*, 2022, 165: 463-474.
- [7] BI X, ZHAO J. A novel orthogonal self-attentive variational autoencoder method for interpretable chemical process fault detection and identification[J]. *Process Safety and Environmental Protection*, 2021, 156: 581-597.
- [8] KOPBAYEV A, FAISL K, YANG M, et al. Fault detection and diagnosis to enhance safety in digitalized process system[J]. *Computers & Chemical Engineering*, 2022, 158: 107609.
- [9] CHEN Z W, DING S X, PENG T, et al. Fault detection for non-Gaussian processes using generalized canonical correlation analysis and randomized algorithms [J]. *IEEE Transactions on Industrial Electronics*, 2017, 65(2): 1559-1567.
- [10] 张佳鑫, 张森, 戴一阳, 等. 面向实际化工过程故障诊断的强化深度卷积神经网络模型构建与应用[J]. *化工进展*, 2024, 43(9): 4833-4844.
- [10] ZHANG J X, ZHANG M, DAI Y Y, et al. Construction and application of enhanced deep convolutional neural network model for practical chemical process fault diagnosis [J]. *Advances in Chemical Engineering*, 2024, 43(9): 4833-4844.
- [11] 杨晓岗, 夏涛. 基于多尺度融合模型的化工故障诊断[J]. *电子测量技术*, 2023, 46(13): 8-16.
- [11] YANG X G, XIA T. Chemical fault diagnosis based on multi-scale fusion model[J]. *Electronic Measurement Technology*, 2023, 46(13): 8-16.
- [12] YUAN X F, WANG Y L, YANG C H, et al. Weighted linear dynamic system for feature representation and soft sensor application in nonlinear dynamic industrial processes[J]. *IEEE Transactions on Industrial Electronics*, 2017, 65(2): 1508-1517.
- [13] WANG N, YANG F, ZHANG R D, et al. Intelligent fault diagnosis for chemical processes using deep learning multimodel fusion[J]. *IEEE Transactions on Cybernetics*, 2020, 52(7): 7121-7135.
- [14] HAN S B, YANG L, DUAN D W, et al. A novel fault detection and identification method for complex chemical processes based on OSCAE and CNN[J]. *Process Safety and Environmental Protection*, 2024, 190: 322-334.
- [15] CAO Y P, LI P H, DENG X G. Chemical process fault diagnosis based on bi-level dynamic IndRNN[J]. *Chemical Engineering Science*, 2025, 307: 121335.
- [16] 王翔, 柯颀挺, 任佳. 样本重构多尺度孪生卷积网络的化工过程故障检测[J]. *仪器仪表学报*, 2019, 40(11): 181-188.
- [16] WANG X, KE L T, REN J. Sample reconstruction of multi-scale twin convolutional networks for chemical process fault detection [J]. *Journal of Instrumentation*, 2019, 40(11): 181-188.
- [17] WANG Y L, WU D X, YUAN X F. LDA-based deep transfer learning for fault diagnosis in industrial

- chemical processes [J]. Computers & Chemical Engineering, 2020, 140: 106964.
- [18] WU H, ZHAO J S. Fault detection and diagnosis based on transfer learning for multimode chemical processes[J]. Computers & Chemical Engineering, 2020, 135: 106731.
- [19] QIN R S, LYU F Y, YE H W, et al. Unsupervised transfer learning for fault diagnosis across similar chemical processes [J]. Process Safety and Environmental Protection, 2024, 190: 1011-1027.
- [20] JIANG H, YANG W Y, SUN Z B, et al. Reinforcement aided latent temporal feature transfer learning: time series prediction with insufficient labeled data for industrial chemical process[J]. IEEE Access, 2024, 12: 165094-165104.
- [21] WANG N, YANG F, ZHANG R D, et al. Intelligent fault diagnosis for chemical processes using deep learning multimodel fusion[J]. IEEE Transactions on Cybernetics, 2020, 52(7): 7121-7135.
- [22] CHAWLA N V, BOWYER K W, HALL L O, et al. Smote: Synthetic minority over-sampling technique [J]. Journal of Artificial Intelligence Research, 2002, 16: 321-357.
- [23] BATISTA G E, PRATI R C, MONARD M C. A study of the behavior of several methods for balancing machine learning training data [J]. ACM SIGKDD Explorations Newsletter, 2004, 6(1): 20-29.
- [24] ADELIM M, MAZINA A H. High efficiency fault-detection and fault-tolerant control approach in Tennessee Eastman process via fuzzy-based neural network representation [J]. Complex & Intelligent Systems, 2020: 199-212.
- [25] ISURU A, UDUGAMA K V, GERNAEY M A, et al. A novel use for an old problem: The Tennessee Eastman challenge process as an activating teaching tool [J]. Education for Chemical Engineers, 2020, 30: 20-31.

作者简介

卫亚楠, 硕士研究生, 主要研究方向为化工过程故障诊断。

E-mail: wei1171857157@163.com

谢俊(通信作者), 博士, 教授, 博士生导师, 主要研究方向为化工过程故障诊断。

E-mail: xiejun@xju.edu.cn