



基于两阶段多教师知识蒸馏的工业过程故障检测方法

陈光捷 张洪海 刘毅 周乐

Two-stage Multi-teacher Knowledge Distillation for Industrial Process Fault Detection

CHEN Guang-Jie, ZHANG Hong-Hai, LIU Yi, ZHOU Le

在线阅读 View online: <https://doi.org/10.16383/j.aas.c250617>

您可能感兴趣的其他文章

基于拓扑一致性对抗互学习的知识蒸馏

Topology-guided Adversarial Deep Mutual Learning for Knowledge Distillation

自动化学报. 2023, 49(1): 102–110 <https://doi.org/10.16383/j.aas.c200665>

广义余弦二维主成分分析

Generalized Cosine Two-dimensional Principal Component Analysis

自动化学报. 2022, 48(11): 2836–2851 <https://doi.org/10.16383/j.aas.c190392>

独立慢特征分析建模方法及其在动态故障检测中的应用

Independent-Slow Feature Analysis Modelling Method and Its Application in Dynamic Fault Detection

自动化学报. 2025, 51(11): 2520–2533 <https://doi.org/10.16383/j.aas.c250134>

一种基于信息熵迁移的文本检测模型自蒸馏方法

Self-distillation via Entropy Transfer for Scene Text Detection

自动化学报. 2024, 50(11): 2128–2139 <https://doi.org/10.16383/j.aas.c210598>

多模态动态核主成分分析的气液两相流状态监测

Flow State Monitoring of Gas-liquid Two-phase Flow Using Multiple Dynamic Kernel Principle Component Analysis

自动化学报. 2022, 48(3): 762–773 <https://doi.org/10.16383/j.aas.c210690>

基于改进结构保持数据降维方法的故障诊断研究

Research on Fault Diagnosis of Data Dimension Reduction Based on Improved Structure Preserving Algorithm

自动化学报. 2021, 47(2): 338–348 <https://doi.org/10.16383/j.aas.c180138>

基于两阶段多教师知识蒸馏的工业过程故障检测方法

陈光捷¹ 张洪海¹ 刘毅² 周乐¹

摘 要 现代工业过程数据具有大容量、高维度及复杂相关性等特征,单一多元统计监测方法难以兼顾不同类型特征的监测需求. 现有多模型融合方法与深度学习技术虽能提升故障检测性能,但前者依赖模型库构建,难以统一建模,后者存在结构复杂与参数冗余问题. 针对上述问题,提出一种基于两阶段多教师知识蒸馏的工业过程建模与故障检测方法. 该方法通过蒸馏框架将核主成分分析与独立成分分析提取的异构知识内化至学生自编码器模型中,实现非线性与非高斯特征的统一建模,并通过两阶段蒸馏协同优化特征空间与重构空间. 第一阶段在特征层蒸馏以引导学生模型学习教师模型的特征分布,第二阶段在重构层蒸馏以提升模型对过程变化的表征与重构能力. 在田纳西-伊斯曼仿真过程及合成氨实际过程上的实验结果表明,该方法能够有效提升故障检测的准确性与鲁棒性,并通过离线知识蒸馏实现在线阶段的统一建模与高效监测.

关键词 独立成分分析; 核主成分分析; 知识蒸馏; 多教师知识蒸馏; 故障检测

引用格式 陈光捷, 张洪海, 刘毅, 周乐. 基于两阶段多教师知识蒸馏的工业过程故障检测方法. 自动化学报, 2026, 52(7): 1426–1437

DOI 10.16383/j.aas.c250617

CSTR 32138.14.j.aas.c250617

Two-stage Multi-teacher Knowledge Distillation for Industrial Process Fault Detection

CHEN Guang-Jie¹ ZHANG Hong-Hai¹ LIU Yi² ZHOU Le¹

Abstract Modern industrial process data are characterized by large scale, high dimensionality, and complex correlations, making it difficult for a single multivariate statistical monitoring method to simultaneously address diverse monitoring requirements. Although multi-model fusion methods and deep learning techniques can improve fault detection performance, the former relies on the construction of model libraries and lacks unified modeling capability, while the latter suffers from complex structures and parameter redundancy. To address these issues, a two-stage multi-teacher knowledge distillation method for industrial process modeling and fault detection is proposed. In this framework, heterogeneous knowledge extracted by kernel principal component analysis and independent component analysis is embedded into a student autoencoder model, enabling unified modeling of nonlinear and non-Gaussian characteristics. A two-stage distillation strategy is further adopted to collaboratively optimize the feature space and reconstruction space. In the first stage, feature-level distillation guides the student model to learn the feature distributions of the teacher models. In the second stage, reconstruction-level distillation is performed to enhance the model's capability in representing and reconstructing process variations. Experiments on the Tennessee-Eastman simulation process and a real ammonia synthesis process demonstrate that the proposed method can effectively improve fault detection accuracy and robustness, while achieving unified modeling and efficient online monitoring through offline knowledge distillation.

Keywords independent component analysis; kernel principal component analysis; knowledge distillation; multi-teacher knowledge distillation; fault detection

Citation Chen Guang-Jie, Zhang Hong-Hai, Liu Yi, Zhou Le. Two-stage multi-teacher knowledge distillation for industrial process fault detection. *Acta Automatica Sinica*, 2026, 52(7): 1426–1437

收稿日期 2025-11-10 录用日期 2026-03-19

Manuscript received November 10, 2025; accepted March 19, 2026

国家重点研发计划 (2025YFE0204600), 浙江省自然科学基金 (LR26F030005), 国家自然科学基金 (U23A20328) 资助

Supported by National Key Research and Development Project of China (2025YFE0204600), Zhejiang Provincial Natural Science Foundation of China (LR26F030005), and National Natural Science Foundation of China (U23A20328)

本文责任编辑 罗彪

Recommended by Associate Editor LUO Biao

1. 浙江科技大学自动化与电气工程学院 杭州 310023 2. 浙江工业大学机械工程学院 杭州 310023

1. School of Automation and Electrical Engineering, Zhejiang University of Science and Technology, Hangzhou 310023 2. College of Mechanical Engineering, Zhejiang University of Technology, Hangzhou 310023

现代工业中, 物联网、云计算、大数据分析和人工智能等技术的进步正驱动着智能制造的发展. 随着生产需求日趋灵活多样、工业过程的复杂性不断提升, 对生产的安全性及可靠性也提出更高要求. 过程监控是保障工业系统稳定运行、安全生产和产品质量的关键技术, 尤其在庞大复杂的系统中, 微小故障也可能引发重大损失. 故障检测作为过程监控的核心环节, 旨在及时识别系统偏离正常运行状态的异常, 从而避免潜在的经济损失与安全事故.

然而, 在面对工业数据普遍存在的非线性、非高斯性及高维相关性特征时, 依赖单一原理或视角

的传统监测模型往往难以提供完整且有效的特征表示。例如, Wang 等^[1]为突破单一网络在时空特征提取上的局限性, 构建复杂的 CNN 与 LSTM 组合; Zhang 等^[2]则针对高维样本的特征提取难题, 采用多个差异化子网络协同工作的策略。这些已有研究从侧面印证了单一模型在表征复杂工业数据特性方面存在固有局限, 凸显了发展更先进方法的必要性。为使模型具备应对多种数据特性的综合建模能力, 知识蒸馏 (knowledge distillation, KD) 提供一种富有前景的途径, 其核心思想是让学生模型学习一个或多个更强大的教师模型所蕴含的知识。

知识蒸馏是一种高效的模型压缩技术, 旨在将大型教师模型的知识迁移至轻量级学生模型。该概念最早由 Buciluă 等^[3]提出, 用于压缩集成模型; Hinton 等^[4]将其扩展至深度神经网络, 通过让学生模型模仿教师模型的输出分布, 即软标签, 实现知识迁移, 并正式确立“知识蒸馏”框架。随着研究的深入, 可蒸馏的知识来源不断扩展。最初的蒸馏使用软标签^[4]作为教师输出, 目前已演变为包括中间层激活^[5]、神经元与特征^[6-7]、通道^[8]、实例^[9]及类别^[10]之间的关系等多层次信息。据此, 知识源可分为基于软标签、基于特征和基于相似性三类。根据师生训练模式, 蒸馏方案主要分为离线、在线和自蒸馏。为实现高效的知识传递, 研究者还提出多样化的蒸馏框架, 包括基于注意力的蒸馏、对抗蒸馏、多教师蒸馏、跨模态蒸馏、自适应蒸馏和对比蒸馏等^[11]。

知识蒸馏已在计算机视觉、自然语言处理和语音识别等领域取得显著成功^[11]。近年来, 该技术逐渐应用于工业过程监控领域。例如, Zhang 等^[12]利用知识蒸馏技术将复杂教师网络压缩为高效学生模型以实现模型轻量化; Petrosian 等^[13]结合解耦蒸馏与 LoRA (low-rank adaptation) 微调, 构建参数量极低的轴承故障检测模型 DKDL-Net; Ai 等^[14]通过蒸馏将基于手工特征的领域知识迁移至基于深度特征的网络; Liu 等^[15]构建基于蒸馏的零样本学习 (zero-shot learning) 框架, 实现未知故障的检测与诊断。

同时, 以自编码器 (autoencoder, AE) 为代表的深度学习方法, 因其模型结构简单、方法成熟且网络结构可根据不同需求灵活修改等优势, 已成为工业故障诊断领域的研究热点^[16]。然而, 尽管知识蒸馏为提升轻量级模型性能提供了新思路, 但在复杂工业过程监控场景中, 尤其是结合自编码器结构时, 仍面临若干挑战。首先, 现有研究多基于单一教师框架, 难以通过单一视角全面捕捉工业过程数据的非线性与非高斯异构特性。其次, 标准的端到端

单阶段训练方式在 AE 蒸馏框架中存在局限。传统的 AE 训练通过最小化重构误差同时优化编码器与解码器, 但这种联合优化策略容易导致特征学习与重构任务的耦合, 限制了模型的收敛性能。正如 Hinton 等^[17]在 Deep Belief Nets 的研究中指出, 逐层预训练能有效缓解复杂网络的优化难题; Romero 等^[18]提出的 FitNets 进一步证实, 在蒸馏框架下先约束中间层特征再进行整体训练, 能引导学生模型更好地逼近教师性能。这表明, 缺乏解耦的单阶段策略难以在特征表示与重构能力之间取得最佳平衡。

为解决上述问题, 本文提出一种基于两阶段多教师知识蒸馏的工业过程故障检测方法。不同于多模型统计融合 (multi-model statistic fusion, MMSF) 的在线集成策略, 本文旨在通过知识蒸馏将核主成分分析 (kernel principal component analysis, KPCA) 与独立成分分析 (independent component analysis, ICA) 提取的异构判别知识内化至单一学生模型中, 从而通过离线阶段的知识迁移实现在线阶段的单模型高效统一监测。具体而言, 该方法利用异构教师体系提取非线性与非高斯特征, 并提出两阶段解耦蒸馏策略, 依次通过特征层与重构层蒸馏分别优化编码器与解码器。该策略有效规避了联合优化的任务冲突, 实现特征学习与数据重构的协同提升。

综上所述, 本文的主要贡献与创新点如下:

- 1) 设计 KPCA-ICA 异构教师体系, 分别提取非线性结构特征与非高斯统计特征, 并通过多层蒸馏实现异构特征的有效融合;
- 2) 构建两阶段分离蒸馏机制, 分别针对编码器与解码器进行优化, 以提升学生模型的特征表达与重构能力;
- 3) 提出一种基于两阶段多教师知识蒸馏的工业过程故障检测方法, 突破传统多模型融合依赖多模型在线并行的局限, 实现非线性与非高斯特征的单模型统一监测。

1 预备知识

1.1 自编码器

自编码器通过编码器将包含 N 个样本、 M 个变量的输入数据 $X \in \mathbf{R}^{N \times M}$ 映射到低维潜在空间 z , 并通过解码器实现数据重构。本文采用不包含非线性激活函数的线性自编码器, 其结构在原理上与主成分分析具有相似性^[16]。其编码与重构过程如式 (1) 和式 (2) 所示:

$$z = W_e X + b_e \quad (1)$$

$$\hat{X} = W_d z + b_d \quad (2)$$

其中, $\hat{X}$ 为输入数据 X 的重构结果; W_e , b_e 与 W_d , b_d 分别表示编码器与解码器的参数. 模型参数通过最小化如式 (3) 所示的均方重构误差进行训练:

$$L_{\text{Recon}} = \frac{1}{N} \sum_{n=1}^N \|X^n - \hat{X}^n\|_2^2 \quad (3)$$

基于训练完成的模型, 引入 T^2 和 Q 统计量用于故障检测. 对于新的测试样本 x_k , 首先计算其潜变量 $z_k = W_e x_k + b_e$ 及重构结果 $\hat{x}_k$. T^2 统计量用于衡量样本在潜变量空间中的偏离程度, 其定义如式 (4) 所示:

$$T^2 = (z_k - \mu_z)^T \Sigma_z^{-1} (z_k - \mu_z) \quad (4)$$

其中, μ_z 和 Σ_z 分别为正常工况下潜变量的均值与协方差矩阵. Q 统计量用于检测残差空间的异常, 定义为重构误差的平方欧氏范数, 如式 (5) 所示:

$$Q = \|x_k - \hat{x}_k\|_2^2 \quad (5)$$

1.2 核主成分分析

核主成分分析利用非线性映射 $\Phi(\cdot)$ 将原始数据 $X \in \mathbf{R}^{N \times M}$ 投影至高维特征空间, 并在该空间内执行线性 PCA, 从而提取数据中的非线性结构特征^[19-20]. 在核技巧的作用下, 无需显式构造映射函数, 仅通过核函数 $\text{ker}(\cdot, \cdot)$ 即可完成特征提取.

设 $K \in \mathbf{R}^{N \times N}$ 为中心化后的核矩阵, 其元素由核函数计算. 通过特征值分解

$$K a = N \lambda a \quad (6)$$

可获得 $\tilde{N}$ 个非零特征值及对应的特征向量. 其中, λ 为特征值, a 为对应的特征向量. 选取前 K 个主特征向量构成模型参数 $A_p = [a_1, \dots, a_K] \in \mathbf{R}^{N \times K}$.

对于新的测试样本 x_k , 其在第 i 个核主成分上的得分 $t_{k,i}$ 如式 (7) 所示:

$$t_{k,i} = \mathbf{k}_k^T \alpha_i \quad (7)$$

其中, $\mathbf{k}_k = [\text{ker}(x_k, x_1), \dots, \text{ker}(x_k, x_N)]^T$ 为测试样本的核向量; α_i 为核矩阵对应的第 i 个特征向量. 构建用于过程故障检测的监控统计量, 如式 (8) 和式 (9) 所示:

$$T^2 = \mathbf{t}_k^T \Lambda_K^{-1} \mathbf{t}_k \quad (8)$$

$$Q = \|\Phi(x_k) - \hat{\Phi}(x_k)\|_2^2 = \sum_{j=1}^{\tilde{N}} t_{k,j}^2 - \sum_{j=1}^K t_{k,j}^2 \quad (9)$$

其中, $\Lambda_K = \text{diag}\{\lambda_1, \dots, \lambda_K\}$ 为前 K 个核主成分对应的特征值构成的对角矩阵; $\mathbf{t}_k = [t_{k,1}, \dots,$

$t_{k,K}]^T$ 为得分向量; $\hat{\Phi}(x_k)$ 表示样本 x_k 在特征空间中的重构映射.

1.3 独立成分分析

独立成分分析假定观测数据由 K 个相互独立的非高斯源信号 $S \in \mathbf{R}^{K \times N}$ 经线性混合矩阵 $A \in \mathbf{R}^{M \times K}$ 生成. 对于原始数据矩阵 $X \in \mathbf{R}^{N \times M}$, 该关系如式 (10) 所示:

$$X^T = A S \quad (10)$$

ICA 通过求解混合矩阵 W_u 对观测数据进行逆向解混, 以恢复源信号.

为消除数据相关性, 对数据进行白化处理. 记 $E(\cdot)$ 为期望算子, 对协方差矩阵 $E(X^T X)$ 进行特征值分解, 得到正交矩阵 P 及对角矩阵 Λ , 构造白化矩阵 $V = \Lambda^{-\frac{1}{2}} P^T$, 白化后的数据如式 (11) 所示:

$$\tilde{X}^T = V X^T \quad (11)$$

此时满足 $E(\tilde{X}^T \tilde{X}) = \mathbf{I}$, 其中 $\mathbf{I}$ 为单位矩阵. 在白化空间中, 混合过程可视为正交变换, 因此通过最大化统计独立性求解正交矩阵 B , 得到独立成分估计 $\hat{S} = B \tilde{X}^T$. 结合白化过程, $W_u = B V$, 混合矩阵可估计为 $A = W_u^{-1} = V^{-1} B^T$.

对于新的测试样本 x_k , 通过 W_u 得到独立成分得分向量 $s_k = W_u x_k$. 此外, 为计算在白化空间中的重构误差, 同步求取该样本的白化向量 $\tilde{x}_k = V x_k$. 基于独立成分构建监控统计量, 如式 (12) 和式 (13) 所示:

$$I^2 = s_k^T \Sigma_s^{-1} s_k \quad (12)$$

$$Q = \|\tilde{x}_k - \hat{\tilde{x}}_k\|_2^2 \quad (13)$$

其中, Σ_s 为独立成分的协方差矩阵; I^2 为独立成分统计量; $\hat{\tilde{x}}_k = B^T s_k$ 表示由独立成分经逆变换得到的白化空间重构.

2 多教师蒸馏

2.1 整体框架概述

本文提出一种两阶段多教师知识蒸馏的过程监控框架, 其总体结构如图 1 所示. 该框架以自编码器作为学生模型, 通过离线蒸馏将多源教师模型的监控知识迁移至统一模型中, 从而在保证检测性能的同时, 实现在线阶段的单模型高效统一监测. 为排除学生模型自身非线性结构的干扰, 确保其非线性建模能力完全源于教师知识的迁移, 本文采用不含非线性激活函数的线性自编码器作为学生模型.

在教师端, 选取 KPCA 与 ICA 作为异构教师

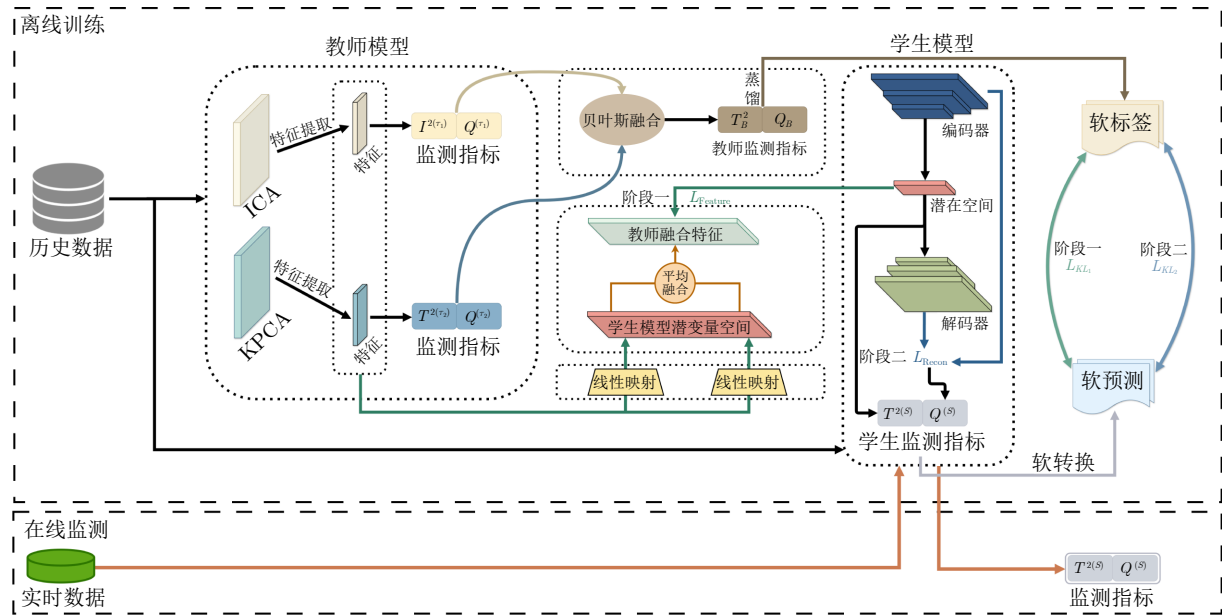


图1 两阶段多教师知识蒸馏过程监测框架

Fig.1 Framework of two-stage multi-teacher knowledge distillation for process monitoring

模型, 分别用于表征过程数据中的非线性相关结构与非高斯独立结构. 两类教师模型在正常工况数据上独立建模, 并分别输出对应的监控统计量. 为综合利用不同教师所提供的多视角信息, 本文基于贝叶斯推理对各教师的监控统计量进行融合, 构建反映过程整体状态的融合监控特征, 为后续学生模型的知识蒸馏提供统一目标.

在学生端, 采用分离式两阶段蒸馏策略以实现教师知识的有序迁移. 第一阶段固定解码器, 仅对编码器进行训练, 通过潜在特征约束与基于潜在空间监控统计量的知识蒸馏相结合, 引导学生模型在潜在空间中学习教师模型所刻画的结构特征, 从而建立与多教师一致的表示能力; 第二阶段固定编码器, 仅训练解码器, 在保持潜在特征稳定的基础上, 通过重构损失与基于重构残差统计量的知识蒸馏相结合, 约束学生模型在重构空间中对齐教师融合知识所表征的监控特性, 从而增强其对过程异常的敏感性.

经过上述两阶段蒸馏训练后, 学生模型能够在在线监测阶段独立运行, 利用单一模型同时表征多种过程特性, 实现对复杂工业过程的高效故障检测.

2.2 异构教师模型构建与贝叶斯知识融合

数据驱动过程监测中亟需应对的数据特性主要包括非线性相关结构^[21]与非高斯统计分布^[22]. 基于单一统计假设的监测模型通常只能从某一特定视角刻画过程状态, 难以充分反映复杂工业过程的多样

性特征. 为此, 本文构建一个由不同统计原理驱动的异构教师系统, 选取 KPCA 与 ICA 作为教师模型. 其中, KPCA 教师模型用于刻画过程变量之间的非线性相关结构, 而 ICA 教师模型侧重于提取数据中的非高斯独立统计成分. 两类模型基于不同的统计建模机理, 对同一过程数据进行特征提取, 从而为过程状态提供具有差异性的多视角统计描述.

通过引入上述异构教师模型, 学生模型在蒸馏阶段能够同时接收来自不同统计视角的知识约束, 从而避免受限于单一建模假设. 然而, 由于 ICA 与 KPCA 教师模型在特征提取机理上的显著差异, 其对应的监控统计量在数值尺度与统计分布特性上并不一致, 若直接进行线性叠加或平均, 容易导致尺度失配与概率语义缺失, 从而影响融合结果的可靠性.

为此, 本文引入一种基于贝叶斯推断的概率融合策略^[21-22], 在统一的概率空间内对多源监控统计量进行转换与加权整合, 构建反映过程整体状态的融合监控特征. 首先, 基于核密度估计 (kernel density estimation, KDE) 方法, 在显著性水平 δ 下分别确定各教师模型监控统计量的控制限. 为统一表示形式, 将教师模型 $\tau \in \{\tau_1, \tau_2\}$ 的某类主空间监控统计量记为 $S^{2(\tau)}$ (对应 $I^{2(\tau_1)}$ 或 $T^{2(\tau_2)}$), 其控制限记为 $S_{lim}^{2(\tau)}$.

对于测试样本 x_k , 以其在教师模型 τ 下计算得到的监控统计量 $S^{2(\tau)}$ 为输入, 分别构建样本在故障状态 f 与正常状态 n 下的条件概率, 并进一步通过贝叶斯公式计算样本属于故障状态的后验概率,

如式 (14) 所示:

$$P_{S^2}^{(\tau)}(f|x_k) = \frac{P_{S^2}^{(\tau)}(x_k|f)P_{S^2}^{(\tau)}(f)}{P_{S^2}^{(\tau)}(x_k)} \quad (14)$$

其中, $P_{S^2}^{(\tau)}(x_k)$ 为样本的边缘概率, 由全概率公式计算得到, 如式 (15) 所示:

$$P_{S^2}^{(\tau)}(x_k) = P_{S^2}^{(\tau)}(x_k|f)P_{S^2}^{(\tau)}(f) + P_{S^2}^{(\tau)}(x_k|n)P_{S^2}^{(\tau)}(n) \quad (15)$$

式中, $P_{S^2}^{(\tau)}(f) = \delta$ 与 $P_{S^2}^{(\tau)}(n) = 1 - \delta$ 分别为故障与正常状态的先验概率. 以主空间统计量 $S^{2(\tau)}$ 为例, 其条件概率采用指数形式定义, 如式 (16) 和式 (17) 所示:

$$P_{S^2}^{(\tau)}(x_k|f) = \exp\left(-\mu \frac{S_{lim}^{2(\tau)}}{S^{2(\tau)}}\right) \quad (16)$$

$$P_{S^2}^{(\tau)}(x_k|n) = \exp\left(-\mu \frac{S^{2(\tau)}}{S_{lim}^{2(\tau)}}\right) \quad (17)$$

其中, μ 为可调系数, 用于平衡异常值对概率估计的影响. 同理可得残差空间统计量 Q 的后验概率 $P_Q^{(\tau)}(f|x_k)$.

在获得各教师模型的后验概率后, 采用加权平均方式对多教师统计信息进行融合, 构建全局概率监控指标 T_B^2 与 Q_B , 如式 (18) 和式 (19) 所示:

$$T_B^2 = \frac{\sum_{\tau} w_{S^2}^{(\tau)} P_{S^2}^{(\tau)}(f|x_k)}{\sum_{\tau} w_{S^2}^{(\tau)}} = \sum_{\tau} \bar{w}_{S^2}^{(\tau)} P_{S^2}^{(\tau)}(f|x_k) \quad (18)$$

$$Q_B = \frac{\sum_{\tau} w_Q^{(\tau)} P_Q^{(\tau)}(f|x_k)}{\sum_{\tau} w_Q^{(\tau)}} = \sum_{\tau} \bar{w}_Q^{(\tau)} P_Q^{(\tau)}(f|x_k) \quad (19)$$

其中, $w_{S^2}^{(\tau)}$ 和 $w_Q^{(\tau)}$ 为原始权重; $\bar{w}_{S^2}^{(\tau)}$ 和 $\bar{w}_Q^{(\tau)}$ 为其归一化形式, 定义为 $\bar{w}_{S^2}^{(\tau)} = w_{S^2}^{(\tau)} / \sum_{\tau} w_{S^2}^{(\tau)}$ 和 $\bar{w}_Q^{(\tau)} = w_Q^{(\tau)} / \sum_{\tau} w_Q^{(\tau)}$.

权重分配同时考虑当前时刻及其历史窗口内的后验概率变化, 以增强对持续性故障的响应. 具体而言, 以主空间统计量为例, 其原始权重 $w_{S^2}^{(\tau)}$ 的更新规则如式 (20) 所示:

$$w_{S^2}^{(\tau)} = \begin{cases} \frac{1}{\varepsilon}, & P_{S^2}^{(\tau)}(f|x_k) \geq \delta \text{ 且} \\ & \bar{P}_{S^2}^{(\tau)}(f|x_k) \geq \delta \\ \varepsilon, & \text{其他} \end{cases} \quad (20)$$

其中, $0 < \varepsilon < 1$ 为常数; $\bar{P}_{S^2}^{(\tau)}(f|x_k)$ 表示样本在最近 H 个时刻内的平均后验概率, 如式 (21) 所示:

$$\bar{P}_{S^2}^{(\tau)}(f|x_k) = \frac{1}{H} \sum_{j=k-H+1}^k P_{S^2}^{(\tau)}(f|x_j) \quad (21)$$

在本文的实现中, 各参数设置如下: 可调参数 $\mu = 0.2$, $\varepsilon = 0.01$, 历史窗口长度设为 $H = 6$. 最终, 若融合后的概率监控指标超过显著性水平 δ , 则判定过程处于故障状态; 否则认为过程运行正常.

值得指出的是, 尽管本文在教师端采用了多模型统计量的加权融合形式, 但其设计逻辑与传统的多模型融合监测方法存在本质差异. 传统融合监测通常在线阶段并行运行多个监测模型, 并将融合统计量作为在线监测判据, 其监测流程高度依赖多个子模型的同步输出与后续融合步骤.

相比之下, 本文提出的框架以知识内化为核心目标, 融合统计量仅作为离线训练阶段的软标签, 用于引导学生模型学习多教师一致反映的异常统计特性. 通过知识蒸馏, 不同教师模型的多视角信息被压缩并迁移至单一学生模型中, 使其在在线阶段仅需独立运行即可实现对复杂过程特性的综合刻画. 因此, 本文方法本质上是一种以融合统计特性为导向的模型能力迁移, 而非传统意义上的指标级融合监测.

2.3 两阶段分离蒸馏策略

为清晰描述所提出的两阶段多教师知识蒸馏过程, 其整体训练流程总结于算法 1.

算法 1. 两阶段多教师知识蒸馏框架训练流程

输入. 训练数据集 X , 预训练教师模型 (ICA、KPCA), 最大训练轮数 E_1 、 E_2 , 学习率 η_1 、 η_2 , 收敛阈值 ϵ_1 、 ϵ_2 .

输出. 训练完成的学生模型参数: 编码器 θ_e , 解码器 θ_d .

初始化. 随机初始化编码器参数 θ_e 与解码器参数 θ_d .

1. 阶段一: 潜在空间蒸馏 (优化编码器)
2. 冻结解码器参数 θ_d ;
3. **for** epoch = 1 **to** E_1 **do**, epoch 表示当前训练轮次
4. **for** $x_k \in X$ **do**
5. $z^{(\tau_1)} \leftarrow \text{ICA}(x_k)$, $z^{(\tau_2)} \leftarrow \text{KPCA}(x_k)$;
6. $\tilde{z}^{(\tau_1)} \leftarrow G^{(\tau_1)} z^{(\tau_1)}$, $\tilde{z}^{(\tau_2)} \leftarrow G^{(\tau_2)} z^{(\tau_2)}$;
7. $z^{(\tau)} \leftarrow \frac{1}{2}(\tilde{z}^{(\tau_1)} + \tilde{z}^{(\tau_2)})$;
8. 由式 (18) 计算 T_B^2 ;
9. $z^{(S)} \leftarrow f_{\theta_e}(x_k)$;
10. $L_{\text{Stage1}} \leftarrow \alpha L_{\text{Feature}} + (1 - \alpha) L_{KL1}$
11. $\theta_e \leftarrow \theta_e - \eta_1 \nabla_{\theta_e} L_{\text{Stage1}}$;
12. **end for**
13. **if** $\Delta L_{\text{Stage1}} \leq \epsilon_1$ **then break**;
14. **end for**
15. 阶段二: 重构空间蒸馏 (优化解码器)

```

16. 冻结编码器参数  $\theta_e$ ;
17. for epoch = 1 to  $E_2$  do
18.   for  $x_k \in X$  do
19.     由式 (19) 计算  $Q_B$ ;
20.      $z^{(S)} \leftarrow f_{\theta_e}(x_k)$ ;
21.      $\hat{x}_k \leftarrow g_{\theta_d}(z^{(S)})$ ;
22.      $L_{Stage2} \leftarrow \beta L_{Recon} + (1 - \beta)L_{KL2}$ ;
23.      $\theta_d \leftarrow \theta_d - \eta_2 \nabla_{\theta_d} L_{Stage2}$ ;
24.   end for
25.   if  $\Delta L_{Stage2} \leq \epsilon_2$  then break;
26. end for
返回.  $\theta_e, \theta_d$ .

```

2.3.1 第一阶段: 编码器训练 (潜在空间蒸馏)

如算法 1 第 1 ~ 14 行所示, 第一阶段在固定解码器参数的条件下, 仅对编码器进行优化. 该阶段旨在通过特征层蒸馏, 引导学生模型在潜在空间中学习与教师模型一致的结构表征.

针对算法 1 第 5 ~ 7 行所涉及的特征提取与融合过程, 考虑到异构教师模型在建模机理上的差异, 来自 ICA 与 KPCA 的潜在特征 $z^{(\tau_1)}$ 与 $z^{(\tau_2)}$ 在尺度与分布上通常不一致, 难以直接进行融合. 为此, 本文首先引入线性映射矩阵 $G^{(\tau)}$ 对教师特征进行空间对齐, 如式 (22) 所示:

$$\tilde{z}^{(\tau)} = G^{(\tau)} z^{(\tau)}, \tau \in \{\tau_1, \tau_2\} \quad (22)$$

进而采用平均策略获得算法中所需的融合目标特征 $z^{(\tau)}$, 如式 (23) 所示:

$$z^{(\tau)} = \frac{1}{2} (\tilde{z}^{(\tau_1)} + \tilde{z}^{(\tau_2)}) \quad (23)$$

随后, 定义特征层蒸馏损失如式 (24) 所示. 其中, $f_{\theta_e}(\cdot)$ 表示参数为 θ_e 的学生模型编码器, $z^{(S)}$ 表示学生模型提取的潜在特征.

$$L_{Feature} = \left\| z^{(S)} - z^{(\tau)} \right\|_2^2 \quad (24)$$

此外, 为约束学生模型在统计监控层面与教师模型保持一致, 引入基于 KL 散度的统计蒸馏损失项 L_{KL1} . 设 $T^{2(S)}$ 为学生模型的主空间统计量, T_B^2 为教师模型的融合统计量, 则在温度参数 γ 调节下, 其定义如式 (25) 所示:

$$L_{KL1} = \text{KL} \left(\sigma \left(\frac{T^{2(S)}}{\gamma} \right) \parallel \sigma \left(\frac{T_B^2}{\gamma} \right) \right) \quad (25)$$

其中, $\sigma(\cdot)$ 表示 Softmax 函数. 综合上述两部分, 第一阶段的总体优化目标为:

$$L_{Stage1} = \alpha L_{Feature} + (1 - \alpha) L_{KL1} \quad (26)$$

其中, $\alpha \in [0, 1]$ 为权衡系数. 此外, $\Delta L_{Stage1} =$

$|L_{Stage1}^{(t)} - L_{Stage1}^{(t-1)}|$ 表示相邻两个训练轮次之间的损失变化量, 当其小于预设阈值 ϵ_1 时, 认为模型已收敛并提前终止训练.

需要指出的是, 该分阶段训练策略实现了特征学习与重建建模的有效解耦. 在第一阶段中, 通过冻结解码器参数, 使编码器能够专注于学习与教师模型一致的潜在表示, 从而为后续解码阶段提供稳定且具有判别性的特征基础.

2.3.2 第二阶段: 解码器训练 (重构空间蒸馏)

如算法 1 第 15 ~ 26 行所示, 第二阶段在固定编码器参数的前提下对解码器进行优化. 其中, $g_{\theta_d}(\cdot)$ 表示参数为 θ_d 的学生模型解码器, 以在保持潜在特征结构稳定的同时, 实现从特征空间到重构空间的知识迁移, 并使学生模型具备与教师一致的残差建模能力.

在训练目标设计上, 第二阶段同时考虑重构精度与统计一致性两个方面. 首先, 引入重构损失以约束学生模型输出与原始输入之间的偏差, 从而保证其对于正常工况数据的重构能力.

进一步地, 为将教师模型在残差空间中的统计判别信息有效地迁移至学生模型, 引入基于 KL 散度的统计蒸馏损失 L_{KL2} . 设 $Q^{(S)}$ 为学生模型的残差统计量, Q_B 为由多教师融合得到的残差统计量, 则在温度参数 γ 的调节下, 统计蒸馏损失定义为:

$$L_{KL2} = \text{KL} \left(\sigma \left(\frac{Q^{(S)}}{\gamma} \right) \parallel \sigma \left(\frac{Q_B}{\gamma} \right) \right) \quad (27)$$

综合重构损失与统计蒸馏损失, 第二阶段的总体优化目标定义为:

$$L_{Stage2} = \beta L_{Recon} + (1 - \beta) L_{KL2} \quad (28)$$

其中, $\beta \in [0, 1]$ 为权衡系数, 用于平衡重构损失与残差统计蒸馏损失之间的贡献. 当本阶段相邻两个训练轮次的损失变化量 $\Delta L_{Stage2} = |L_{Stage2}^{(t)} - L_{Stage2}^{(t-1)}|$ 小于预设阈值 ϵ_2 时, 认为模型已收敛并提前终止训练.

3 案例研究

为验证所提出的两阶段多教师知识蒸馏过程监测方法的有效性与泛化能力, 本文选取两个典型案例进行研究: 田纳西-伊斯曼 (Tennessee-Eastman, TE) 过程与实际合成氨生产过程. 在两个实验中, KPCA 与 ICA 两种教师模型均基于正常工况数据建模, 并采用平均特征值法确定主成分和独立成分的数量. 对于 KPCA 模型, 核函数选用高斯核, 其宽度参数设置为 $c = 500m$, 其中 m 为过程变量的数量. 各监控统计量的控制限通过核密度估计方法

确定. 在后续的实验结果中, 本文将该蒸馏后的学生模型简称为多教师蒸馏模型 (multi-teacher distillation model, MTDM); 此外, 在所有监控图中, 统一使用虚线表示控制限, 实线表示监控统计量曲线.

在 TE 过程实验中, 显著性水平设定为 $\delta = 5\%$. 监测性能通过以下两项指标进行评估: 故障检测率 (fault detection rate, FDR), 即超过控制限的故障样本占全部故障样本的比例; 故障检测时间 (fault detection time, FDT), 即首次连续六个样本超过控制限的样本序号.

在合成氨生产过程实验中, 考虑到工业系统的安全与稳定性要求更高, 显著性水平设定为 $\delta = 1\%$. 监测性能除了采用故障检测率外, 还采用误报率 (false alarm rate, FAR) 作为评价指标, 后者表示被误判为故障的正常样本比例.

3.1 TE 过程实验

TE 过程^[23]是过程监测领域的经典基准过程, TE 过程的流程图如图 2 所示, 它被广泛用于评估各类过程监测模型的性能^[15, 21–22, 24]. 本文选取 52 个过程变量作为监测对象, 关于这 52 个监测变量的详细信息可参见文献^[23]. 在实验中, 设置 1 种正常工况和 21 种故障工况. 正常工况数据集包含 1460 个样本, 其中前 500 个样本用于模型训练, 其余 960 个样本用于验证并计算各统计量的置信限. 每一种故障工况的数据集均包含 960 个样本, 且故障在第 160 个样本之后被引入.

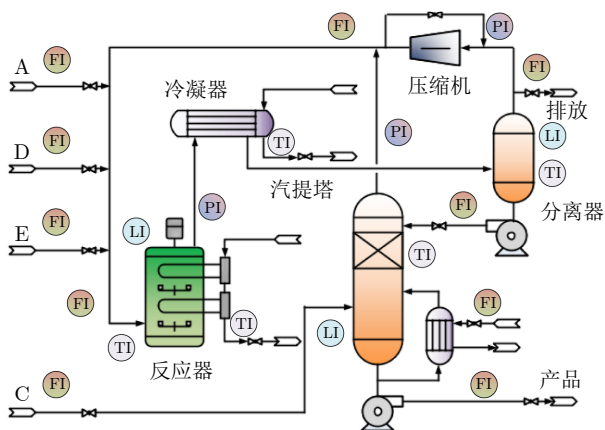


图 2 TE 过程流程图

Fig. 2 Flowchart of the TE process

本文所采用的学生模型为三层线性自编码器, 其编码器结构为 [128, 64, 32], 解码器与其对称. 在 TE 实验中, 学生模型及蒸馏训练过程的参数设置均保持一致: 训练轮数 (epoch) 为 600 次, 学习率为 0.002. 蒸馏过程中两阶段训练各占 300 轮并且

温度参数 γ 设为 5. 式 (26) 和式 (28) 中, 两阶段损失函数的权重系数分别设置为 $\alpha = 0.6$ 与 $\beta = 0.3$.

故障 F5 的监测结果如图 3 所示. 虽然蒸馏前的学生 AE 模型在 $T^{2(S)}$ 统计量上能够于第 160 个样本处检测到故障, 但其检测率较低, 仅为 34.88%; 同样, $Q^{(S)}$ 统计量的 FDT 为 162, 对应检测率为 34.50%. 从监控图可明显观察到, MTDM 的检测性能显著提升, 两个统计量的 FDT 均为 160, 且对应的 FDR 均达到了 100%.

图 4 展示了故障 F10 下蒸馏前后模型的监测性能对比结果. 蒸馏前的学生 AE 模型在 $T^{2(S)}$ 与 $Q^{(S)}$ 统计量下的 FDT 分别为 214 与 204, 对应的

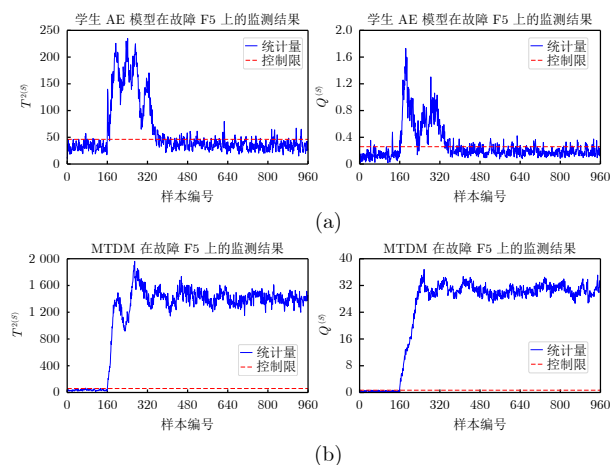


图 3 TE 过程故障 F5 的 AE 蒸馏前后监测结果 ((a) 学生 AE 模型; (b) MTDM)

Fig. 3 AE monitoring results before and after distillation for fault F5 in the TE process ((a) Student AE model; (b) MTDM)

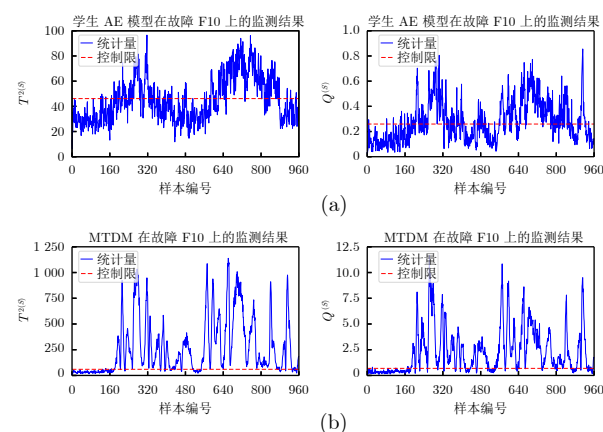


图 4 TE 过程故障 F10 的 AE 蒸馏前后监测结果 ((a) 学生 AE 模型; (b) MTDM)

Fig. 4 AE monitoring results before and after distillation for fault F10 in the TE process ((a) Student AE model; (b) MTDM)

FDR 分别为 49.12% 与 56.38%; 而 MTDM 的两个统计量的 FDT 分别提前至 181 与 185, 对应 FDR 分别提高至 88.00% 与 81.50%. 实验结果表明蒸馏过程显著增强了模型的故障响应速度与检测准确性.

最后, 故障 F19 进一步验证了蒸馏的有效性, 故障 F19 监测结果如图 5 所示. 蒸馏前的学生 AE 模型在整个故障期间, $T^{2(S)}$ 统计量未能检测到故障, 其 FDR 仅为 25.50%, $Q^{(S)}$ 统计量虽在第 623 个样本处触发报警, 但检测率也仅为 25.37%; 而 MTDM 的两个统计量 $T^{2(S)}$ 和 $Q^{(S)}$ 分别在第 169 个样本和第 168 个样本检测到该故障, 对应 FDR 分别为 86.62% 和 73.88%.

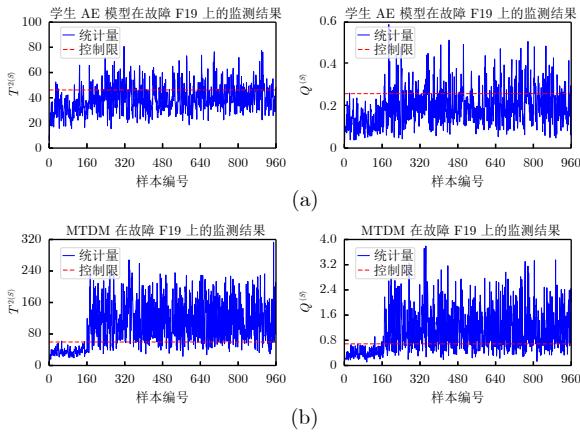


图 5 TE 过程故障 F19 的 AE 蒸馏前后监测结果 ((a) 学生 AE 模型; (b) MTDM)

Fig. 5 AE monitoring results before and after distillation for fault F19 in the TE process ((a) Student AE model; (b) MTDM)

表 1 展示了本文所提方法、各教师模型 (ICA、KPCA) 以及蒸馏前的学生 AE 模型在 21 个 TE 过程故障上的故障检测率结果. 从总体平均性能来看, MTDM 在检测率上优于教师 KPCA 和蒸馏前的学生 AE 模型, 但略低于教师 ICA.

对检测率结果作进一步分析, 可发现其主要呈现出以下三种情形:

一是模型性能显著提升的情况. 如故障 F5、F10、F19 等表明, MTDM 在故障检测率上相比蒸馏前有明显提高, 验证了多教师蒸馏策略在知识迁移和监控能力增强方面的有效性.

二是性能相近的情况, 对于 F3、F9 和 F15 故障, 四种方法的检测率均较低, 这是由于这三个故障数据的均值与方差变化不显著^[25-26], 使得数据驱动的监测方法难以有效识别. 除这三个难检故障外, 其余故障 F1、F2、F6 ~ F8、F12 ~ F14、F18 总体上较易检测, 大多表现出轻微的性能提升, 少数故

表 1 21 个 TE 过程故障的 FDR——基于 ICA、KPCA、AE 与多教师蒸馏方法 (%)

Table 1 FDR of 21 TE process faults detected by ICA, KPCA, AE, and multi-teacher distillation method (%)

故障 编号	ICA		KPCA		AE		MTDM	
	$T^{2(\tau_1)}$	$Q^{(\tau_1)}$	$T^{2(\tau_2)}$	$Q^{(\tau_2)}$	$T^{2(S)}$	$Q^{(S)}$	$T^{2(S)}$	$Q^{(S)}$
F1	99.75	99.75	99.75	99.25	99.50	99.88	99.88	99.50
F2	98.62	98.88	98.75	98.12	98.75	98.88	98.88	98.38
F3	8.75	8.50	7.88	9.62	10.12	5.62	7.88	6.25
F4	100.00	100.00	100.00	9.62	72.88	100.00	99.50	96.38
F5	100.00	100.00	28.62	31.25	34.88	34.50	100.00	100.00
F6	100.00	100.00	99.50	99.50	99.25	100.00	100.00	100.00
F7	100.00	100.00	100.00	66.75	100.00	100.00	100.00	100.00
F8	98.38	98.12	98.12	97.25	97.50	95.12	98.38	97.50
F9	8.75	6.50	6.38	7.88	9.38	5.50	6.00	7.75
F10	90.50	90.62	54.75	49.25	49.12	56.38	88.00	81.50
F11	78.38	78.25	79.25	27.62	64.38	61.38	73.25	66.88
F12	99.88	99.88	99.12	97.25	99.00	97.12	99.75	99.62
F13	95.38	95.50	95.50	94.25	95.62	95.25	95.62	94.62
F14	100.00	100.00	100.00	93.00	100.00	97.88	100.00	99.88
F15	19.88	17.50	13.25	15.62	12.38	8.75	14.00	8.75
F16	92.12	93.88	37.00	34.88	32.00	51.88	90.62	83.38
F17	96.25	96.25	96.00	76.12	85.50	96.50	95.38	91.62
F18	90.38	91.00	91.25	89.50	90.38	90.38	91.00	91.50
F19	86.50	90.75	18.88	1.75	25.50	25.37	86.62	73.88
F20	90.38	90.75	68.38	41.38	50.25	61.00	78.38	75.88
F21	64.88	61.62	54.37	30.12	48.38	54.87	57.63	37.00
均值	81.85	81.80	68.89	55.71	65.47	68.39	80.04	76.68

障则出现微弱下降. 总体而言, 这类情况是可以接受的, 因为模型在大多数故障下仍能保持较高的检测性能.

三是检测性能明显下降的情况, 即故障 F21 的 $Q^{(S)}$ 统计量表明检测性能有所下降. 在 TE 仿真过程中, 故障 F21 常被用于验证动态故障检测方法的有效性^[27-28]. 本文教师模型均基于静态假设, 其前提是数据在时间上不相关. 而 F21 的时序自相关性与此假设相悖, 导致教师模型难以捕捉其特征, 学生模型在蒸馏中也难以习得针对此类动态故障的知识, 从而监测效果下降.

3.2 合成氨实验

本文的实验数据取自一个真实的合成氨 (ammonia synthesis) 生产过程中的一段转化炉 (primary reformer) 单元, 其工艺流程如图 6^[29] 所示. 该单元的核心功能是将原料甲烷在高温下与蒸汽混合, 通过化学反应将其转化为氢气. 在这个过程中, 操作条件的稳定 (特别是燃烧温度) 对氢气产率至

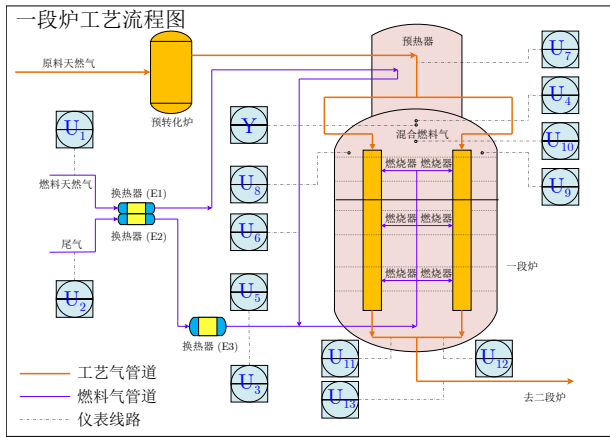


图 6 一段炉流程图
Fig.6 Flowchart of the primary reformer

关重要，而燃烧的稳定性又与炉内的氧气含量等关键参数密切相关。

炉内氧气含量的动态变化（例如，操作均值从一个水平偏移至另一个水平）即代表了生产工况的实质性切换。因此，本实验的核心目的在于利用故障检测的方式，验证所提方法能否及时并准确地检测这种微妙的“工况切换”，而非局限于传统意义上的设备故障检测。

本文选取 13 个过程变量作为监测对象，关于这 13 个监测变量的详细信息可参见文献 [30–31]。为验证本文所提方法的有效性，从历史数据集中分别选取两段连续数据，每段包含 2000 个样本，作为本次实验的 Case1 和 Case2。

通过观察两个案例的氧气含量曲线，确定各自的初始稳定工况 (State1)。Case1 的 State1 对应 0 ~ 700 样本，Case2 的 State1 对应 0 ~ 1000 样本，而其后的样本则被视为工况 2 (State2)。本实验基于 State1 进行训练集和验证集的划分，并构建测试集。具体的样本划分如表 2 所示，其中工况切换点指明了在测试集中，工况从 State1 切换至 State2 的样本点。

学生模型为线性 AE，针对合成氨实验的 13 个过程变量，其编码器结构为 [32, 16, 8]，解码器结构与其对称。为客观评估模型性能，蒸馏前的学生 AE 模型与本文所提蒸馏方法采用相同的训练参数：总训练轮数设为 600，学习率设为 0.001。蒸馏过程

表 2 合成氨实验案例的数据划分设置
Table 2 Data division settings for ammonia synthesis experimental cases

案例编号	训练集	验证集	测试集	工况切换点
Case1	1 ~ 300	301 ~ 560	561 ~ 2000	140
Case2	1 ~ 550	551 ~ 800	801 ~ 2000	200

中，两阶段的训练轮数各为 300 次。阶段一与阶段二的损失函数权重系数分别设为 $\alpha = 0.8$ 与 $\beta = 0.25$ ，知识蒸馏的温度参数 γ 设为 6。

根据表 3 所示结果，实验基于两位教师模型、蒸馏前的学生 AE 模型以及本文所提的 MTDM，在合成氨过程一段炉的 Case1 和 Case2 下进行模型性能评估。总体来看，四种方法均在 FDR 指标上表现良好，其中 ICA 和 KPCA 两个教师模型在两个案例中都维持了较高的检测能力。

值得注意的是，蒸馏前的学生 AE 模型的 $T^{2(S)}$ 统计量在两个案例中均取得了 100.00% 的检测率，表面上看似达到了理想的故障检测效果。然而，从图 7(a) 和图 8(a) 的监控曲线中可以直观地看到，蒸馏前的学生 AE 模型的 $T^{2(S)}$ 统计量从监测开始时刻就已远超其控制限。这一现象与表 4 的 FAR 结果（其平均误报率高达 98.14%）完全一致。因此，尽管蒸馏前的学生 AE 模型的 $T^{2(S)}$ 统计量在检测率方面表现突出，但其极高的误报率使其完全丧失了监测的可靠性，难以应用于实际工业场景。

表 3 2 个合成氨过程一段炉案例的 FDR——基于 ICA、KPCA、AE 与多教师蒸馏方法 (%)
Table 3 FDR of 2 primary reformer cases in the ammonia synthesis process detected by ICA, KPCA, AE, and multi-teacher distillation method (%)

案例编号	ICA		KPCA		AE		MTDM	
	$T^2(\tau_1)$	$Q(\tau_1)$	$T^2(\tau_2)$	$Q(\tau_2)$	$T^{2(S)}$	$Q^{(S)}$	$T^{2(S)}$	$Q^{(S)}$
Case1	72.77	91.92	90.77	96.46	100.00	14.69	89.54	86.46
Case2	82.50	83.60	86.70	92.80	100.00	82.70	84.10	85.20
均值	77.64	87.76	88.74	94.63	100.00	48.70	86.82	85.83

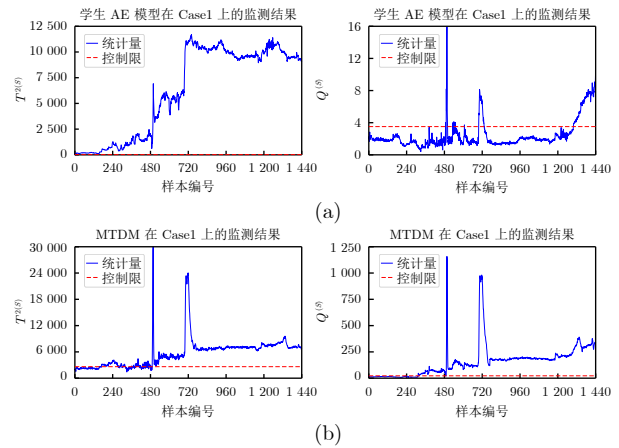


图 7 合成氨过程一段炉 Case1 的 AE 蒸馏前后监测结果
(a) 学生 AE 模型; (b) MTDM
Fig.7 AE monitoring results before and after distillation for Case1 of the primary reformer in the ammonia synthesis process (a) Student AE model; (b) MTDM

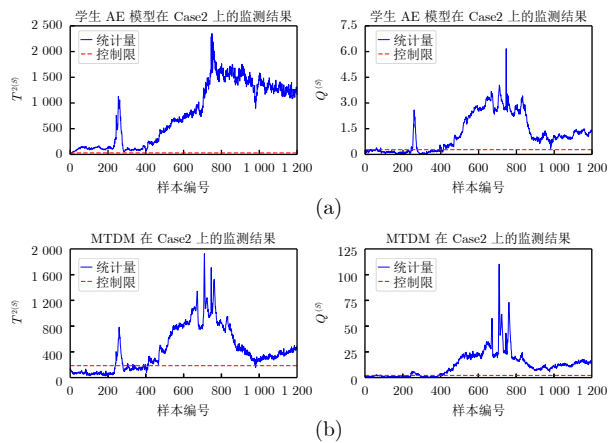


图 8 合成氨过程一段炉 Case2 的 AE 蒸馏前后监测结果 ((a) 学生 AE 模型; (b) MTDM)

Fig.8 AE monitoring results before and after distillation for Case2 of the primary reformer in the ammonia synthesis process ((a) Student AE model; (b) MTDM)

表 4 2 个合成氨过程一段炉案例的 FAR——基于 ICA、KPCA、AE 与多教师蒸馏方法 (%)

Table 4 FAR of 2 primary reformer cases in the ammonia synthesis process detected by ICA, KPCA, AE, and multi-teacher distillation method (%)

案例 编号	ICA		KPCA		AE		MTDM	
	$T^2(\tau_1)$	$Q(\tau_1)$	$T^2(\tau_2)$	$Q(\tau_2)$	$T^2(S)$	$Q(S)$	$T^2(S)$	$Q(S)$
Case1	0	0	0.70	0	99.28	0	0	0
Case2	0	0	10.00	4.00	97.00	5.00	5.00	5.50
均值	0	0	5.35	2.00	98.14	2.50	2.50	2.75

相比之下, MTDM 在保持较高检测率的同时显著降低了误报率. 如图 7(b) 和图 8(b) 所示, MTDM 的统计量在工况切换点之前均能稳定地保持在控制限以下, 仅在工况切换后迅速响应. 表 4 的结果也印证了这一点: MTDM 的 FAR 在两个案例中均保持在 0% ~ 5.50% 区间, 相比蒸馏前的学生 AE 模型大幅下降. 尽管在 Case2 的 $Q(S)$ 指标上, MTDM 的误报率出现了约 0.50% 的轻微增加, 但仍处于可接受范围内.

综合来看, 本文所提方法有效提升了学生模型的稳定性与可靠性, 使其在保持敏感性的同时避免了误报率显著升高, 从而验证了所提蒸馏框架在工业过程监测任务中的有效性.

3.3 讨论与分析

为深入剖析所提方法的内在机理并验证其实际应用价值, 本节将重点围绕多教师蒸馏与多模型统计量融合的实验差异、学生模型的鲁棒泛化机制以及教师特征组合的充分必要性展开深入讨论.

首先, 从图 9 展示的 TE 过程平均故障检测率结果可以观察到, MTDM 与 MMSF 在整体检测性能上表现较为接近, 这从实验层面表明所采用的知识蒸馏策略具有一定有效性, 即学生模型能够在一定程度上内化多模型集成的监测能力. 尽管两者在平均意义上的表现趋于一致, 但在针对若干具体故障的监测特性上仍存在差异.

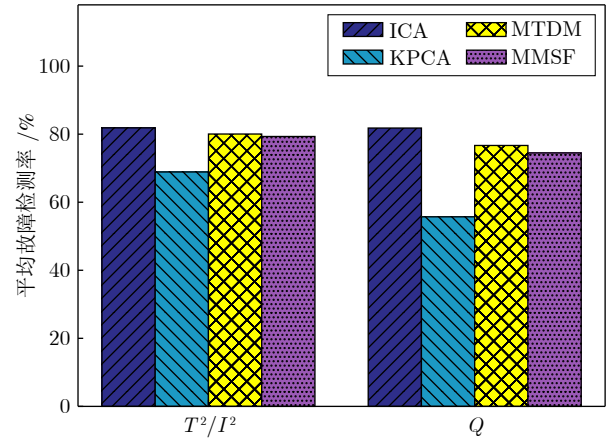


图 9 四种方法在 TE 过程上的平均故障检测率
Fig.9 Average fault detection rates of four methods on the TE process

本文选取三个具有代表性的故障案例进行对比分析, 相关结果汇总于表 5, 其中 $T^{2(F)}$ 与 $Q^{(F)}$ 表示 MMSF 所得到的融合统计量. 可从中发现, 在故障 F4 中, MTDM 的残差统计量 $Q^{(S)}$ 显著优于融合统计量 $Q^{(F)}$; 在故障 F10 中, MTDM 的主空间统计量 $T^{2(S)}$ 同样优于由多模型统计量融合得到的 $T^{2(F)}$; 而在故障 F19 上, MTDM 在主空间与残差空间的监测性能均明显优于 MMSF.

表 5 MTDM 与 MMSF 在 TE 过程代表性故障上的 FDR 对比 (%)

Table 5 FDR comparison between MTDM and MMSF on representative TE process faults (%)

故障编号	MTDM		MMSF	
	$T^{2(S)}$	$Q^{(S)}$	$T^{2(F)}$	$Q^{(F)}$
F4	99.50	96.38	99.88	64.62
F10	88.00	81.50	86.00	81.88
F19	86.62	73.88	73.00	51.50

类似的实验现象亦可在合成氨过程数据中观察到. 在所选取的两个案例中, MTDM 与 MMSF 均展现出较高的故障检测率与较低的误报率, 整体性能表现稳定. 尽管在此情形下, 相较于直接的多模型融合, 知识蒸馏所带来的检测率提升幅度相对有

限,但其核心优势在于通过单一学生模型实现了对多模型监测能力的有效迁移,从而在保持较高监测性能的同时,显著简化了在线监测流程。

综合 TE 与合成氨过程的实验结果可以看出,尽管 MTDM 在平均意义上的整体性能与 MMSF 基本相当,但在部分具有代表性的故障类型中,蒸馏后的学生模型能够更有效地吸收教师模型所蕴含的统计判别信息,从而在特定故障工况下展现出更强的检测敏感性与稳定性。

实验结果表明,多教师蒸馏约束下的学生模型在不同数据集与故障类型上均保持了较为稳定的检测性能,体现出良好的鲁棒性与泛化能力。这表明蒸馏过程有助于学生模型学习更加平衡的统计表征,从而在不同监测场景和工况条件下维持稳定的判别行为。与此同时,针对工业过程数据中普遍存在的非线性与非高斯特性,单一统计模型的局限性进一步凸显了异构教师组合的必要性;而学生模型在多类故障上的一致表现,则从实验层面支持了该特征组合在过程状态刻画中的合理性与充分性。此外,鉴于本文研究主要聚焦于静态建模框架,引入动态时序模型^[32]以刻画更复杂的过程演化特性,将作为后续研究的重要方向。

4 结束语

本文针对工业过程数据特征复杂与单一监测模型表征能力不足的问题,提出一种基于两阶段多教师知识蒸馏的故障检测方法。该方法通过构建异构多教师模型与设计分阶段蒸馏策略,成功将教师模型对非线性与非高斯特征的判别知识迁移至单一学生模型中,实现了多源过程特性的协同表征。

在 TE 过程与合成氨生产过程上的实验结果表明,所提方法显著提升了学生模型在复杂工况下的故障敏感性与监测鲁棒性。该方法通过将多模型监测能力内化至单一模型,使在线监测阶段无需依赖多模型并行运行与统计量融合即可刻画复杂过程特性,为工业过程的高效在线监测提供了一种可行思路。未来工作将进一步引入动态时序信息,以增强模型对时变工业场景的适应能力。

参考文献

- Wang N, Yang F, Zhang R, Gao F. Intelligent fault diagnosis for chemical processes using deep learning multimodel fusion. *IEEE Transactions on Cybernetics*, 2020, **52**(7): 7121–7135
- Zhang X, Huang T, Wu B, Hu Y, Huang S, Zhou Q, et al. Multimodel ensemble deep learning method for intelligent fault diagnosis with high-dimensional samples. *Frontiers of Mechanical Engineering*, 2021, **16**(2): 340–352
- Bucilua C, Caruana R, Niculescu-Mizil A. Model compression. In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA: ACM, 2006. 535–541
- Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network. arXiv preprint arXiv: 1503.02531, 2015.
- Song J, Chen Y, Ye J W, Song M L. Spot-adaptive knowledge distillation. *IEEE Transactions on Image Processing*, 2022, **31**: 3359–3370
- Li Y C, Wang X Y, Xu W C, Wang H Z, Qi Y N, Dong J H, et al. Feature distillation is the better choice for model-heterogeneous federated learning. arXiv preprint arXiv: 2507.10348, 2025.
- Mansourian A M, Jalali A, Ahmadi R, Kasaei S. Attention-guided feature distillation for semantic segmentation. arXiv preprint arXiv: 2403.05451, 2024.
- Dai T, Lin Y, Guo H, Wang J B, Zhu Z X. DCSF-KD: Dynamic channel-wise spatial feature knowledge distillation for object detection. In: Proceedings of the AAAI Conference on Artificial Intelligence. Philadelphia, USA: AAAI Press, 2025. 2627–2635
- Mishra D, Uikey R. Unified knowledge distillation framework: Fine-grained alignment and geometric relationship preservation for deep face recognition. In: Proceedings of the IEEE International Joint Conference on Biometrics (IJCB). Osaka, Japan: IEEE, 2025. 1–10
- Karine A, Napoléon T, Jridi M. I2CKD: Intra-and inter-class knowledge distillation for semantic segmentation. *Neurocomputing*, 2025, **649**: Article No. 130791
- Mansourian A M, Ahmadi R, Ghafouri M, Babaei A M, Golezani E B, Ghamchi Z Y, et al. A comprehensive survey on knowledge distillation. arXiv preprint arXiv: 2503.12067, 2025.
- Zhang W F, Biswas G, Zhao Q, Zhao H B, Feng W Q. Knowledge distilling based model compression and feature learning in fault diagnosis. *Applied Soft Computing*, 2020, **88**: Article No. 105958
- Petrosian O, Li P Y, He Y L, Liu J R, Sun Z R K, Fu G F, et al. DKDL-Net: A lightweight bearing fault detection model via decoupled knowledge distillation and LoRA fine-tuning. arXiv preprint arXiv: 2406.06653, 2024.
- Ai M, Xie Y, Ding S X, Tang Z, Gui W. Domain knowledge distillation and supervised contrastive learning for industrial process monitoring. *IEEE Transactions on Industrial Electronics*, 2022, **70**(9): 9452–9462
- Liu Y, Huang J J, Jia M W. Knowledge distillation-based zero-shot learning for process fault diagnosis. *Advanced Intelligent Systems*, 2025, **7**(6): Article No. 2400828
- Qian J C, Song Z H, Yao Y, Zhu Z R, Zhang X M. A review on autoencoder based representation learning for fault detection and diagnosis in industrial processes. *Chemometrics and Intelligent Laboratory Systems*, 2022, **231**: Article No. 104711
- Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets. *Neural Computation*, 2006, **18**(7): 1527–1554
- Romero A, Ballas N, Kahou S E, Chassang A, Gatta C, Bengio Y. FitNets: Hints for thin deep nets. arXiv preprint arXiv: 1412.6550, 2014.
- Schölkopf B, Smola A, Müller K R. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 1998, **10**(5): 1299–1319
- Lee J M, Yoo C K, Choi S W, Vanrolleghem P A, Lee I B. Nonlinear process monitoring using kernel principal component analysis. *Chemical Engineering Science*, 2004, **59**(1): 223–234
- Deng X G, Tian X M, Chen S, Harris C J. Deep principal component analysis based on layerwise feature extraction and its application to nonlinear process monitoring. *IEEE Transactions on Control Systems Technology*, 2018, **27**(6): 2526–2540
- Kong X Y, Ge Z Q. Deep learning of latent variable models for industrial process monitoring. *IEEE Transactions on Industrial Informatics*, 2021, **18**(10): 6778–6788
- Downs J J, Vogel E F. A plant-wide industrial process control problem. *Computers & Chemical Engineering*, 1993, **17**(3): 245–255

- 24 Zheng J H, Yang Z Y, Ge Z Q. Deep residual principal component analysis as feature engineering for industrial data analytics. *IEEE Transactions on Instrumentation and Measurement*, 2024, **73**: 1–10
- 25 Chiang L H, Russell E L, Braatz R D. *Fault Detection and Diagnosis in Industrial Systems*. London: Springer Science & Business Media, 2012.
- 26 Ku W F, Storer R H, Georgakis C. Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 1995, **30**(1): 179–196
- 27 Bao D, Wang Y J, Li S H. Dynamic graph embedding PCA to extract spatio-temporal information for fault detection. *IEEE Transactions on Industrial Informatics*, 2025, **21**(2): 1714–1723
- 28 Chen Y Q, Zhang R D. Deep multiscale convolutional model with multihead self-attention for industrial process fault diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2025, **55**(4): 2503–2512
- 29 Shen B B, Jiang X Y, Yao L, Zeng J S. Gaussian mixture TimeVAE for industrial soft sensing with deep time series decomposition and generation. *Journal of Process Control*, 2025, **147**: Article No. 103355
- 30 Wang J B, Shao W M, Song Z H. Student's-t mixture regression-based robust soft sensor development for multimode industrial processes. *Sensors*, 2018, **18**(11): Article No. 3968
- 31 Zheng J H, Zhou L, Lyu Y T, Yang Z Y, Ge Z Q. Multi-rate data distillation for deep process monitoring. *IEEE Transactions on Instrumentation and Measurement*, 2025, **74**: 1–10
- 32 Zheng J, Zhao C H, Gao F. Retrospective comparison of several typical linear dynamic latent variable models for industrial process monitoring. *Computers & Chemical Engineering*, 2022, **157**: Article No. 107587



陈光捷 浙江科技大学自动化与电气工程学院讲师. 主要研究方向为工业过程监控, 数据驱动建模和故障检测与诊断.

E-mail: chenguangjie@zust.edu.cn
(CHEN Guang-Jie Lecturer at the

School of Automation and Electrical Engineering, Zhejiang University of Science and Technology. His research interests include industrial process monitoring, data-driven modeling, and fault detection and diagnosis.)



张洪海 浙江科技大学自动化与电气工程学院硕士研究生. 主要研究方向为工业过程监控, 数据驱动建模, 知识蒸馏和故障检测与诊断.

E-mail: zhanghonghai@zust.edu.cn

(ZHANG Hong-Hai Master student

at the School of Automation and Electrical Engineering, Zhejiang University of Science and Technology. His research interests include industrial process monitoring, data-driven modeling, knowledge distillation, and fault detection and diagnosis.)



刘毅 浙江工业大学机械工程学院教授. 主要研究方向为数据智能及其在工业过程建模、控制、优化中的应用.

E-mail: yliuzju@zjut.edu.cn

(LIU Yi Professor at the College of Mechanical Engineering, Zhejiang

University of Technology. His research interests include data intelligence and its applications to modeling, control, and optimization of industrial processes.)



周乐 浙江科技大学自动化与电气工程学院教授. 主要研究方向为工业过程建模、监控与故障诊断, 软传感器建模和深度学习. 本文通信作者.

E-mail: zhoule@zust.edu.cn

(ZHOU Le Professor at the School

of Automation and Electrical Engineering, Zhejiang University of Science and Technology. His research interests include industrial process modeling, monitoring and fault diagnosis, soft sensor modeling, and deep learning. Corresponding author of this paper.)