Computer Science and Application 计算机科学与应用, 2026, 16(11), 12-27
Published Online November 2026 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2026.1611322

Hans汉斯

基于跨运行稳定统计特征与类别感知自适应融合的流程工业故障诊断方法研究

王钰凯，辜小花，陈　余，冉志勇

重庆科技大学电子与电气工程学院，重庆

收稿日期：2026年9月30日；录用日期：2026年10月28日；发布日期：2026年11月4日

摘　要

工业时序文本化有助于利用过程变量的语义信息，但会压缩原始信号中的频域结构和局部变化特征；与此同时，辅助数值信息对不同故障类别及不同样本的补偿作用并不一致。针对上述问题，本文构建基于跨运行稳定统计特征的数值分支，并提出类别感知自适应融合方法(Class-Aware Adaptive Fusion, CAAF)。该方法首先根据故障效应方向与幅度在两个独立开发运行间的一致性筛选主频幅值占比、主频位置和单调性三类统计特征，通过变量身份映射与轻量编码器生成数值类别得分；随后，将传统样本级融合权重扩展为样本－类别级权重，采用类别系数刻画数值证据对不同候选类别的修正作用，并利用文本置信度门控调节其样本介入程度。TEP 21类故障诊断结果表明，CAAF的准确率达到94.51%，较纯文本分支、静态融合和QMF分别提高1.83、1.14和0.75个百分点，在IDV-16至IDV-20五类困难故障上的平均准确率达到89.35%。消融实验、多种子重复训练和数值信息扰动实验进一步表明，性能增益主要来自稳定统计信息对文本分支类别相关错判的选择性补偿。

关键词

流程工业，故障诊断，预训练语言模型，文本－数值融合，类别感知自适应融合

Process Industry Fault Diagnosis Based on Cross-Run Stable Statistical Features and Class-Aware Adaptive Fusion

Yukai Wang, Xiaohua Gu, Yu Chen, Zhiyong Ran

School of Electronic and Electrical Engineering, Chongqing University of Science and Technology, Chongqing

Received: September 30, 2026; accepted: October 28, 2026; published: November 4, 2026

文章引用: 王钰凯, 辜小花, 陈余, 冉志勇. 基于跨运行稳定统计特征与类别感知自适应融合的流程工业故障诊断方法研究[J]. 计算机科学与应用, 2026, 16(11): 12-27. DOI: 10.12677/csa.2026.1611322

Abstract

Converting industrial time series into structured text facilitates the use of process semantics but compresses frequency-domain structures and local variations in the original signals. Meanwhile, the complementary value of auxiliary numeric information varies across fault classes and samples. To address these issues, this paper constructs a numeric branch based on cross-run stable statistical features and proposes class-aware adaptive fusion (CAAF). Dominant-frequency amplitude ratio, dominant-frequency location, and monotonicity are first selected according to the consistency of fault-effect directions and magnitudes across independent development runs. Variable-identity mapping and a lightweight encoder then transform these statistics into numeric class scores. CAAF further extends conventional sample-wise fusion to sample-class-specific weighting: class-wise coefficients characterize class-dependent numeric corrections, while a text-confidence gate controls sample-dependent intervention. Experiments on the 21-class Tennessee Eastman Process benchmark show that CAAF achieves an accuracy of 94.51%, exceeding the text-only branch, static fusion, and QMF by 1.83, 1.14, and 0.75 percentage points, respectively. Its average accuracy on the five difficult classes from IDV-16 to IDV-20 reaches 89.35%. Ablation, multi-seed repeated training, and numeric-information perturbation experiments further indicate that the improvement mainly arises from selective compensation for class-dependent errors of the text branch.

Keywords

Process Industry, Fault Diagnosis, Pre-Trained Language Model, Text-Numeric Fusion, Class-Aware Adaptive Fusion

Copyright © 2026 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

1. 引言

流程工业故障诊断是保障化工、冶金和能源等连续生产系统安全运行的重要技术。流程装置具有高维、强耦合和动态时变等特征，局部异常传播后可能在多个变量上形成相近响应。Venkatasubramanian 等[1]综述了定量模型驱动方法，但此类方法依赖准确的过程模型与先验知识。Yin 等[2]针对数据驱动方法缺乏统一比较的问题，在 TEP 上评价了多种统计学习方法，但线性投影对非线性关系的描述能力有限。Downs 和 Vogel [3]构建了 TE 过程模型，Rieth 等[4]进一步发布了扩展仿真数据。Lomov 等[5]针对长时依赖与传感器关联建模问题提出 CNN1D2D，但仍主要依赖数值序列的端到端表征。Zhang 和 Qiu [6]针对标注不足与故障定位困难提出半监督 LSTM-LAE，但未讨论数值证据与语义表示的协同。Peng 等[7]针对噪声干扰和可解释性不足提出 DSAE-SIG，但其归因分析未刻画类别间的差异化补偿。Lin 等[8]针对变量动态特性差异提出 DSGNN，但仍以单一数值模态完成监测。因此，如何获得跨运行稳定的数值表示，并对易混淆类别实施有约束的补偿，仍需进一步研究。

Transformer 的自注意力机制为长距离依赖建模提供了基础[9]。Devlin 等[10]针对上下文语义表示不足提出 BERT，但其预训练对象仍是自然语言文本。工业传感信号属于连续数值序列，预训练语言模型的语义能力难以直接迁移至流程故障诊断。Jin 等[11]针对时序与语言模型输入空间不一致提出 Time-LLM，但该方法主要面向预测任务。Zhou 等[12]针对专用时序模型迁移性不足提出 FPT，但未利用工业变量的物理语义。Gruver 等[13]针对零样本时序预测提出 LLMTime，但其性能受数值分词和概率校准影响。

Ansari 等[14]针对跨数据集概率预测提出 Chronos，但尚未面向工业故障类别判别进行设计。时序文本化可以缓解模态差异，却会压缩原始波形中的频域结构和局部变化特征，使文本分支在部分困难故障上仍存在稳定错判。

多模态融合为重新引入原始数值信息提供了可行途径。Han 等[15]针对多视图预测可靠性不足提出 TMC，但未考虑强、弱分支间的非对称补偿关系。Zhang 等[16]针对低质量模态导致的融合失效提出 QMF，但样本级权重难以区分数值证据对不同候选类别的贡献。Zeng 等[17]针对强模态主导提出模态贡献调制方法，但未刻画同一样本内的类别级可靠性。Wang 等[18]针对模态质量差异提出相关性感知融合框架，但未涉及多类别故障得分的定向修正。McKinney 等[19]针对异构传感信息整合提出无监督融合方法，但未限制辅助信息对可靠主分支的干扰。Peng 等[20]针对模态优化失衡提出在线梯度调制方法，但其调节仍停留在模态层面。Wei 和 Hu [21]针对多模态与单模态目标的梯度冲突提出 MMPareto，但未形成面向候选类别的样本级修正。现有方法大多从模态或样本层面估计整体可靠性，同一样本的不同候选类别通常共享融合权重，难以描述数值信息在类别间分布不均的补偿作用。

针对上述问题，本文研究强文本分支与弱数值分支之间的非对称互补关系及数值信息的选择性介入机制。首先，根据故障效应方向与幅度在两个独立开发运行间的一致性筛选稳定统计特征，并通过变量身份映射和轻量编码器构建数值分支，以形成具有明确物理对应关系的数值证据。随后，以文本错判样本的可救回率刻画两分支的类别相关互补空间，提出 CAAF，将融合权重由样本级 $\alpha(x)$ 扩展为样本－类别级 $\alpha(x,c)$，分别利用类别系数和文本置信度门控调节数值证据的类别修正幅度与样本介入程度。

2. 跨运行稳定统计特征与类别感知自适应融合方法

2.1. 问题描述与总体框架

流程工业故障诊断中的文本－数值融合面临两个相互关联的问题。结构化故障描述能够支持文本分支形成较强的语义表示，但时序文本化会压缩原始波形中的频域结构和局部变化特征；统计数值分支可以保留上述信息，却受输入特征规模限制，整体判别能力弱于文本分支，且补偿效果随故障类别和样本状态变化。前一问题要求重新引入数值证据，后一问题要求约束数值证据的作用范围。基于此，本文构建由固定文本分支、稳定统计数值分支和 CAAF 融合模块组成的故障诊断框架，其总体流程及 CAAF 内部结构如图 1 所示。

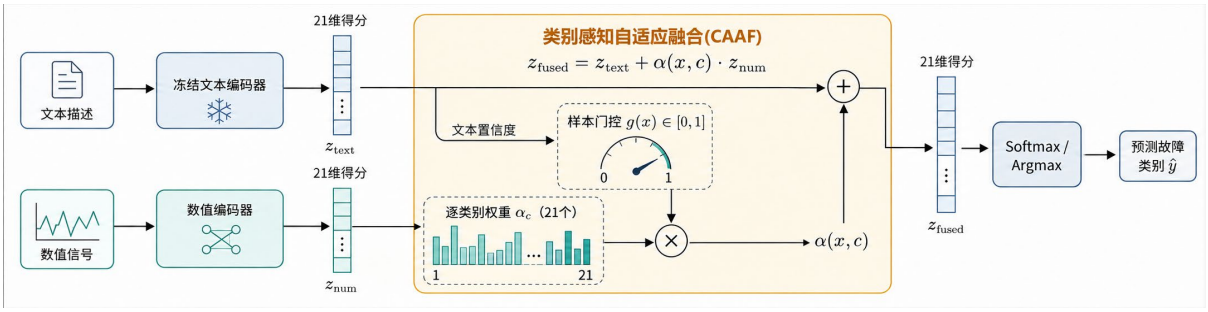


Figure 1. Overall fault diagnosis framework and CAAF fusion module
图 1. 本文故障诊断总体框架及 CAAF 融合模块

图 1 所示框架包括固定文本分支、稳定统计数值分支和 CAAF 融合模块。对于样本 x，工业多变量时序经语义化转换后形成结构化故障描述，固定文本编码器与分类头据此输出 21 维文本类别得分 z_{text}；数值分支从入选变量的原始波形中提取稳定统计特征，经轻量编码器得到 21 维数值类别得分 z_{num}。CAAF 以文本得分间隔估计样本置信度，以类别系数表征数值证据对不同候选类别的修正幅度，并由二者共同

确定样本－类别权重 $\alpha(x,c)$。加权后的数值得分作为增量项与 z_{text} 相加，得到最终融合得分 z_{fused}，并据此输出故障类别。

该框架采用非对称得分修正结构，以适应两个分支在判别能力上的差异。文本分支已具备较强的整体分类能力，若同时重标定文本与数值得分，可能削弱文本分支在简单类别上的稳定优势。数值分支主要补充文本化过程中损失的频域结构和局部变化信息，其作用范围因而同时受类别系数和文本置信度约束。

2.2. 跨运行稳定统计特征构建

流程数据的均值、终值和变化幅度容易随运行初值及工况基线发生偏移，单次运行中具有区分能力的统计量未必能在另一运行中保持相同的故障指向。若直接将此类特征用于融合，数值分支可能把运行条件差异误判为故障证据。为提高数值特征在独立运行间的一致性，本文利用跨运行保持率筛选稳定统计量。稳定特征筛选及后续数值分支构建流程如图 2 所示。

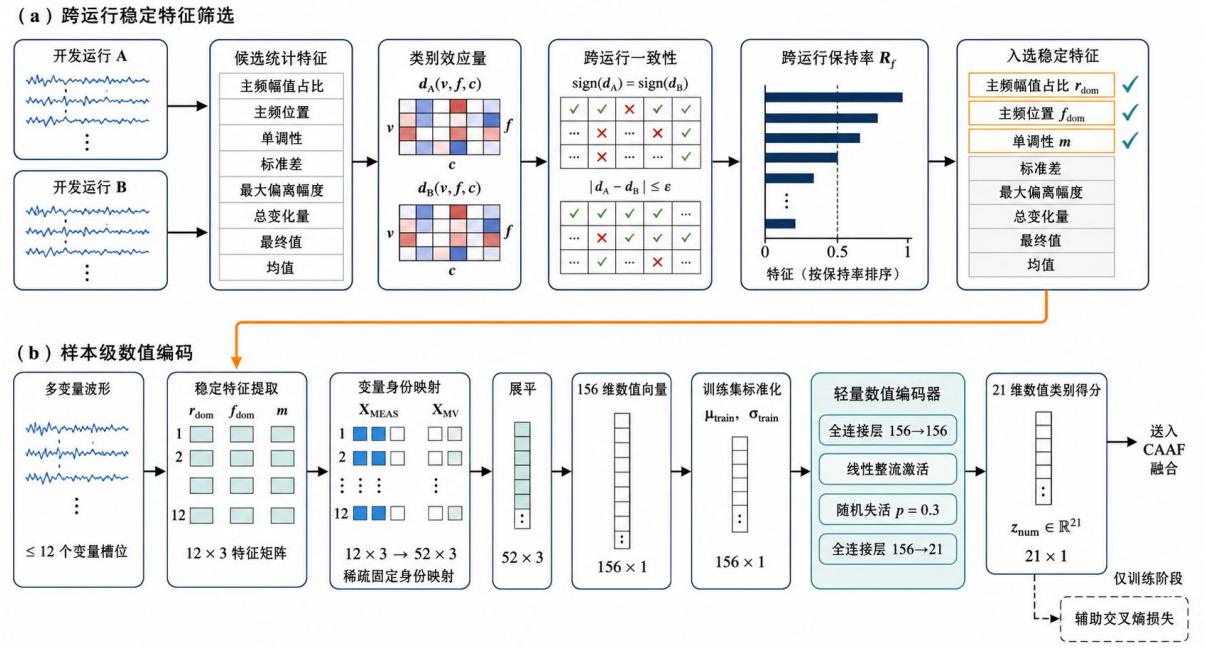


Figure 2. Cross-run stable statistical feature screening and numeric branch construction
图 2. 跨运行稳定统计特征筛选与数值分支构建流程

为避免将仅在单次运行中有效的统计量引入融合，本文从训练数据中按完整仿真批次划分出两个相互独立的开发运行 A 和 B。两个运行均覆盖正常工况及 20 类故障，但采用独立的初始状态和扰动轨迹；验证集与测试集不参与统计特征筛选。对于开发运行 $r\left(r\in\{A,B\}\right)$、变量 v、统计特征 f 和故障类别 c，分别计算类别 c 与其余类别之间的标准化差异。效应量能够在考虑类内波动的同时减小样本规模对差异判断的影响[22]，其定义为：

$$d_r(v,f,c)=\frac{\mu_r(v,f,c)-\mu_r(v,f,\neg c)}{s_{r,\text{pool}}(v,f,c)}. \tag{1}$$

式中，$\mu_r(v,f,c)$ 和 $\mu_r(v,f,\neg c)$ 分别表示类别 c 和其余类别在变量 v、特征 f 上的均值，$s_{r,\text{pool}}$ 为合并标准差。当开发运行 A 与 B 的效应量符号一致，且幅值差不超过容差 ε 时，将对应的“变量－类别”单元

判定为跨运行保持。Ω_f 表示特征 f 上合并标准差有效且两次效应量均可计算的变量–类别组合集合，$I[\cdot]$ 为指示函数。特征 f 的跨运行保持率定义为：

$$R_f=\frac{1}{|\Omega_f|}\sum_{(v,c)\in\Omega_f} I\left[\operatorname{sign}(d_A)=\operatorname{sign}(d_B)\wedge|d_A-d_B|\le\varepsilon\right]. \tag{2}$$

跨运行保持率同时考察故障效应的方向与幅度。仅比较两个开发运行的均值差异会忽略类内波动，仅比较效应量符号则可能将接近零的偶然一致误判为有效保持。本文设置特征级入选阈值 $\tau=0.5000$，当 $R_f\ge\tau$ 时将候选统计特征纳入数值分支；主实验取 $\varepsilon=0.2000$。ε 控制跨运行单元判定的宽松程度，τ 用于将特征级保持率转换为最终入选集合。

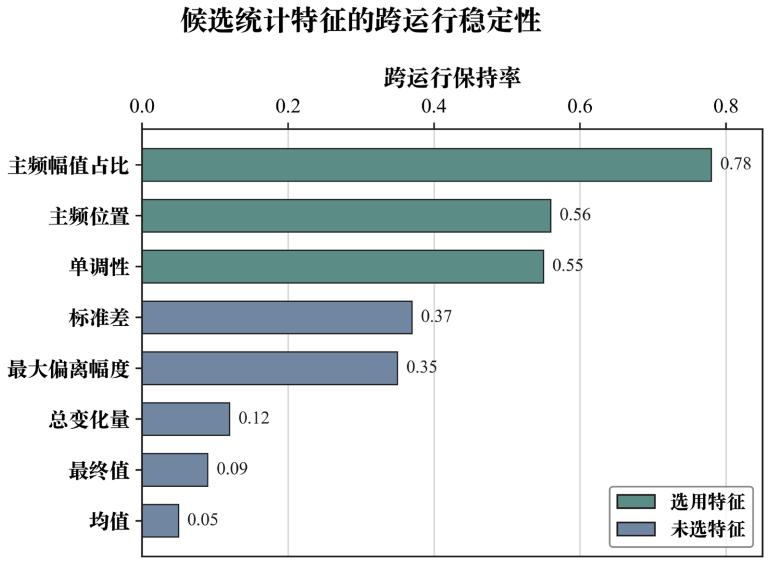


Figure 3. Comparison of cross-run stability among candidate statistical features
图 3. 候选统计特征的跨运行稳定性对比

在 $\varepsilon=0.2000$ 时，主频幅值占比、主频位置和单调性的跨运行保持率分别为 0.7806、0.5625 和 0.5521，均达到 $\tau=0.5000$；标准差与最大偏离幅度分别为 0.3732 和 0.3523，总变化量、终值与均值分别为 0.1232、0.0930 和 0.0580。该排序与图 3 的近似展示一致。结果表明，绝对水平、终值及变化幅度对运行基线较为敏感，而振荡集中程度、主振荡位置和变化方向在开发运行 A 与 B 之间具有更高的一致性。因此，本文选取主频幅值占比、主频位置和单调性构建数值分支。

2.3. 稳定统计数值分支建模

跨运行稳定性筛选确定了主频幅值占比、主频位置和单调性三类特征。为保证不同样本采用一致的计算方式，三类特征均依据固定定义提取。设单个变量的去均值序列为 $\tilde{x}_t=x_t-\bar{x}$，其单边离散傅里叶幅值谱为 $A_k=\left|\mathrm{FFT}(\tilde{x})_k\right|$。去除直流分量后，以最大谱幅值对应位置作为主频索引 k^*，主频位置和主频幅值占比分别定义为：

$$f_{\mathrm{dom}}=\frac{k^*}{T},\quad r_{\mathrm{dom}}=\frac{A_{k^*}}{\sum_k A_k}. \tag{3}$$

主频幅值占比反映谱幅值向主频位置集中的程度，取值越大，说明信号的周期性成分越集中。为描述相邻采样点变化方向的一致性，单调性定义为：

$$m=\frac{\left|\sum_{t=1}^{T-1}\operatorname{sign}\left(x_{t+1}-x_t\right)\right|}{(T-1)}. \tag{4}$$

当序列严格单调上升或下降时，$m=1$；当相邻变化方向频繁交替时，正负符号相互抵消，m 趋近于 0。对于常数序列或总差分绝对值不超过 10^{-8} 的序列，将 m 置为 0。主频幅值占比和单调性的取值位于 $[0,1]$，归一化主频位置位于 $[0,0.5]$。

每个样本按照结构化描述中的变量出现顺序保留前 12 个互不重复的过程变量，不足 12 个时采用空槽位补齐；变量选择过程不使用真实故障标签。依据 XMEAS 和 XMV 变量的固定身份，将 12 × 3 统计特征矩阵散射至 52 × 3 全变量空间，并展平为 156 维向量。相同输入维度在不同样本中始终对应同一物理变量，从而避免槽位变化造成语义错位。各维均值和标准差仅由训练集估计，并固定应用于验证集和测试集；未选中变量对应位置在标准化后重新置零。

标准化后的 156 维特征输入轻量数值编码器。编码器由 156 维全连接层、ReLU 激活、Dropout 层和 21 维输出层构成，输出数值类别得分 $z_{\text{num}}\in\mathbb{R}^{21}$。数值分支不重建文本分支已经获得的完整故障语义，而是将振荡集中程度、主振荡位置和局部变化方向转换为可参与融合的补充证据。辅助分类损失用于维持 z_{num} 的独立判别能力，固定变量身份映射用于保持输入维度与物理变量之间的一致对应关系。

以反应器温度变量的 IDV-19 样本为例，开发运行 B 中该类主频幅值占比的组内均值和标准差分别为 0.6185 和 0.0350，其余类别对应数值分别为 0.4835 和 0.1059，且开发运行 A 与 B 上的效应方向一致。该变量的频域结构在独立运行间保持稳定的类别指向，可为文本描述未充分保留的局部动态差异提供补充证据。

2.4. 类别感知自适应融合

静态融合以单一全局系数控制数值分支，无法描述类别间的补偿差异；样本级动态融合虽然能够判断当前样本是否需要辅助信息，但同一样本的全部候选类别仍共享相同权重，难以同时刻画类别差异与样本置信度。针对上述粒度限制，本文从文本错判样本出发定义两分支互补性，并通过类别系数与文本置信度门控构造样本–类别级融合权重，使数值证据主要作用于存在互补空间的样本和候选类别。

2.4.1. 文本与数值信息互补性

文本分支依据结构化故障描述形成类别表示，其整体判别能力高于仅使用三类统计量的数值分支。然而，辅助分支的融合价值不仅取决于其独立准确率，还取决于其能否覆盖主分支的错误样本。Wang 等[18]的研究表明，辅助信息的有效性与其相对主模态的互补关系密切相关。基于此，本文以文本错判条件下数值分支的正确覆盖比例刻画类别互补性。

设文本分支和数值分支对样本 x 的预测结果分别为 $\hat{y}_t$ 和 $\hat{y}_n$，类别 c 的可救回率定义为：

$$\text{Rescue}(c)=\text{P}\left(\hat{y}_n=y\mid\hat{y}_t\neq y,y=c\right). \tag{5}$$

在此基础上，将全局互补空间定义为：

$$\text{Headroom}=\text{P}\left(\hat{y}_t\neq y\wedge\hat{y}_n=y\right). \tag{6}$$

可救回率衡量文本分支对类别 c 发生错判时，数值分支能够正确识别的样本比例；全局互补空间表示理想选择器相对于纯文本分支可能获得的准确样本比例增量。与数值分支的独立准确率相比，两项指标更直接地反映其对文本错误集合的边际补偿能力。上述统计仅用于分析分支互补关系，不参与模型参数学习。

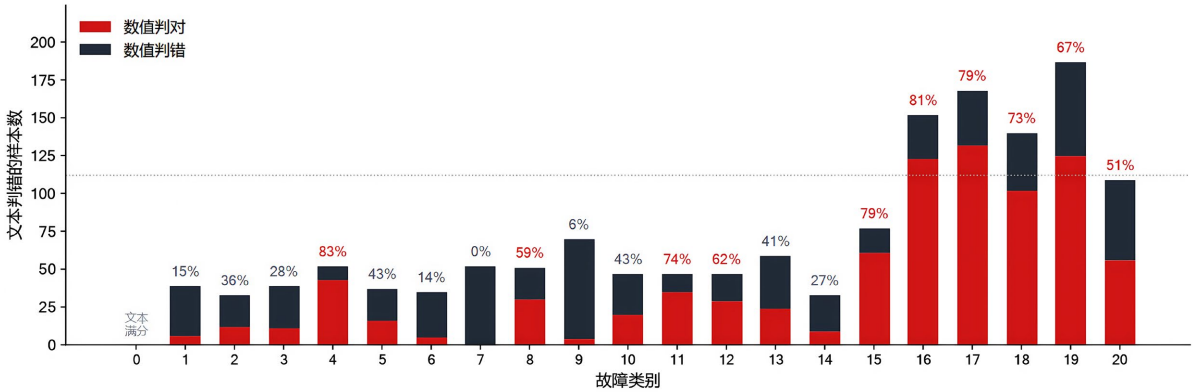


Figure 4. Class-wise recoverable proportion of the numeric branch among samples misclassified by the text branch
图 4. 文本错判样本中数值分支的逐类可救回比例

图 4 给出了文本错判样本中数值分支的逐类可救回率。不同故障类别之间存在明显差异：文本分支在部分类别上的诊断结果接近饱和，数值分支可提供的补偿空间较小；在部分困难类别上，可救回率超过 50%。其中，IDV-16 的可救回率为 81%；IDV-20 的数值分支独立准确率虽低于文本分支，其可救回率仍达到 51%。这表明，较低的独立准确率并不等同于缺乏融合价值，数值分支的有效信息主要分布在文本分支未正确覆盖的样本中，并呈现类别差异。因此，融合机制既需要在类别维度区分数值证据的修正幅度，也需要在样本维度限制其对高置信文本判断的影响。

2.4.2. 类别感知数值得分校正

传统静态后期融合可表示为 $z_f = z_{\text{text}} + \alpha z_{\text{num}}$，其中全部样本和类别共享标量 α。样本级动态融合将权重扩展为 $\alpha(x)$，但同一样本内 C 个候选类别仍采用相同系数。为描述数值证据在不同候选类别上的补偿差异，本文将融合得分定义为：

$$z_{\text{fused}}(c \mid x) = z_{\text{text}}(c \mid x) + \alpha(x,c) \cdot z_{\text{num}}(c \mid x). \tag{7}$$

式(7)保持文本类别得分的原始尺度，仅将数值类别得分作为增量项。当 $\alpha(x,c) = 0$ 时，模型退化为纯文本分支；$\alpha(x,c)$ 增大时，数值分支对候选类别 c 的修正作用相应增强。该非对称结构与文本分支整体判别能力较强、数值分支主要提供局部补充的特点一致，可避免同时重标定两个分支而削弱文本分支已有优势。

为保证类别系数与“数值介入强度”的含义一致，引入全局基础参数 a_0 与类别偏移 δ_c，并通过 Softplus 函数施加非负约束：

$$\alpha_c = \text{softplus}(a_0 + \delta_c), \quad c = 0,1,\cdots,20. \tag{8}$$

式中，α_c 表示数值分支对候选类别 c 的基础介入强度。为限制类别偏移对训练样本的过度拟合，引入 L_2 收缩项 $\lambda_s \|\delta\|^2$，并根据验证集融合交叉熵从 $\{0,10^{-3},10^{-2},10^{-1}\}$ 中选择 λ_s。该参数化方式仅增加 $C+1$ 个标量参数，即可在全局介入强度的基础上学习类别相关修正。

2.4.3. 文本置信度驱动的样本级门控

类别系数能够表征同一故障类别的总体互补倾向，但不能区分该类别内文本判断可靠的样本与存在歧义的样本。若同类样本均采用相同 α_c，数值分支仍可能改变文本高置信样本的正确判别。为此，本文以文本分支最大 logit 与次大 logit 之差描述样本置信度：

$$m(x)=z_{\text{text}}^{(1)}(x)-z_{\text{text}}^{(2)}(x). \tag{9}$$

式中，$z_{\text{text}}^{(1)}$ 和 $z_{\text{text}}^{(2)}$ 分别表示排序后的最大与次大文本得分。相较于容易饱和的最大 Softmax 概率，原始 logit 间隔能够保留困难样本之间的置信度差异。利用训练集间隔均值 μ_m 和标准差 σ_m 进行标准化，并令 $w=-\text{softplus}(\eta)$ 以保证门控值随文本得分间隔增大而单调下降，样本门控定义为：

$$g(x)=\text{sigmoid}\left[-\text{softplus}(\eta)\cdot\frac{m(x)-\mu_m}{\sigma_m}+b\right]. \tag{10}$$

参数 η 和 b 由融合目标联合学习。较小的文本得分间隔对应较大的门控值，使数值分支在文本判断存在歧义的样本上获得更高介入强度；较大的文本得分间隔对应较小门控值，从而降低数值证据对高置信文本判断的影响。样本门控 $g(x)$ 和类别系数 α_c 分别表征当前样本对数值补偿的需求及候选类别的基础介入强度。

2.4.4. 样本–类别联合权重与优化目标

将类别系数与样本门控相乘，得到 CAAF 的样本–类别权重：

$$\alpha(x,c)=\alpha_c\cdot g(x)=\text{softplus}(a_0+\delta_c)\cdot g(x). \tag{11}$$

式中，α_c 表示数值分支对候选类别 c 的基础介入强度，$g(x)$ 表示当前样本对数值补偿的需求，二者共同限定数值证据的作用范围。模型训练目标由融合分类损失、数值分支辅助分类损失和类别偏移收缩项构成：

$$L=\text{CE}(z_{\text{fused}},y)+\lambda_n\text{CE}(z_{\text{num}},y)+\lambda_s\|\delta\|^2. \tag{12}$$

实验中取 $\lambda_n=1$，λ_s 依据验证集融合交叉熵确定。数值分支辅助分类损失用于保持其独立判别能力，避免数值得分退化为仅适用于训练样本的任意修正；类别偏移收缩项用于限制各类别参数相对全局参数的偏离。训练过程中冻结文本编码器与文本分类头，联合优化 156 维数值头和 CAAF 参数。文本类别得分预先计算并在各融合方法之间共享；当 $\alpha(x,c)=0$ 时，融合结果与纯文本分支一致。

3. 实验与结果分析

3.1. 数据集与评价指标

为评价不同融合方法的诊断性能，实验采用统一的数据划分、预处理流程和评价指标。TEP 任务包含正常工况与 20 类故障，仅报告总体指标可能掩盖困难类别的识别差异。因此，除整体分类性能外，本文进一步统计 IDV-16 至 IDV-20 的逐类准确率及五类平均值，以考察方法对复杂故障的诊断能力。

本文采用 TEP 基准数据集验证所提方法。Downs 和 Vogel [3]构建的 TE 过程模型包含反应器、冷凝器、汽液分离器、汽提塔和压缩机等主要单元，为流程故障诊断提供了标准测试对象。Rieth 等[4]进一步发布扩展仿真数据，用于不同诊断方法在统一运行条件下的比较。本文使用的数据包含 41 个测量变量和 11 个操纵变量，分类任务由正常工况 IDV-0 与 20 类故障 IDV-1 至 IDV-20 构成。训练侧共 10,500 个样本，按 90%和 10%划分为 9450 个训练样本与 1050 个验证样本；独立测试集包含 20160 个样本，每类 960 个，类别分布均衡。开发运行 A 和 B 仅从训练数据的完整运行批次中划分，用于统计特征筛选。

评价指标包括总体准确率(Accuracy)、加权精确率(Precision)、加权召回率(Recall)和加权 F1。由于测试集 21 个类别的样本数相同，总体准确率与宏平均召回率数值相等。为分析复杂故障的细粒度识别性能，实验同时报告 IDV-16 至 IDV-20 的逐类准确率及五类平均值。逐类准确率表示给定真实类别条件下的正确预测比例，与该类别召回率等价。

实验围绕四个层次展开：首先比较 CAAF 与代表性融合方法的整体性能和困难故障识别结果；随后通过组件消融区分类别系数与样本门控的独立贡献；进一步结合多种子重复和跨运行容差 ε 敏感性实验，检验性能差异及特征筛选结果的稳定性；最后利用数值信息破坏和融合权重分析，验证性能增益的信息来源及其与类别相关互补机制的对应关系。

3.2. 实验设置与对比方法

后期融合结果可能受到文本得分尺度、数值头容量、优化配置和模型选择准则等因素影响。为使性能差异主要归因于融合策略，所有方法共享相同的文本类别得分、统计特征、数值头结构、数据划分和验证集模型选择规则。

工业多变量时序首先转换为结构化故障描述，并采用 Qwen2.5-7B-Instruct [23]进行编码。对模型最后一层隐藏状态进行已训练的加权汇聚，得到 3584 维文本表示，再由预先训练的分类头映射为 21 维文本类别得分 z_{text}。后续融合实验冻结文本编码器、汇聚模块和分类头，所有方法共享同一组文本得分。数值输入由 52 个固定变量位置与 3 类统计量展开为 156 维向量，数值头采用“156 维全连接层-ReLU-Dropout-21 维全连接层”结构，Dropout 为 0.3。数值头与融合模块采用 Adam 优化器，学习率为 1×10^{-3}，权重衰减为 1×10^{-4}，批大小为 512，最大训练轮数为 300，早停耐心值为 30。各方法根据验证集目标函数选择最优模型检查点。

Table 1. Comparison methods and fusion mechanisms
表 1. 对比方法及其融合机制

方法	核心机制	权重粒度	说明
纯文本	仅使用固定文本分支类别得分	无	固定基准
Static	$z_{\text{text}}+\alpha z_{\text{num}}$	全局	可学习标量
TMC	证据不确定性与可信组合	样本–视图	证据融合
PDF	面向低质量模态的动态融合	样本	动态质量
MNL	强模态引导弱模态负学习	样本	强模态引导的负学习
QMF	质量与能量置信度加权	样本	能量置信度加权
CAAF	$z_{\text{text}}+\alpha(x,c)z_{\text{num}}$	样本–类别	本文方法

对比方法包括纯文本分支、静态后期融合、TMC、QMF、PDF 和 MNL。纯文本分支仅使用固定文本分类器的输出，静态融合采用全局可学习标量调制数值得分。TMC [15]通过证据理论实现可信多视图组合；QMF [16]依据能量置信度生成样本级模态权重；PDF [24]联合单模态置信度、跨模态置信度与相对校准形成动态权重；MNL [25]利用强模态预测引导弱模态抑制非目标类别。考虑到文本与数值得分尺度存在差异，本文在不改变各方法核心机制的前提下进行决策级适配：TMC 分别为两侧证据引入可学习温度，QMF 对能量置信度进行标准化并设置可学习融合尺度，PDF 和 MNL 沿用相同的固定文本输入及数值头。各方法的融合机制如表 1 所示。

3.3. 诊断性能对比

本节以纯文本分支、静态融合及 4 种代表性动态融合方法为比较对象，在统一设置下依次考察 21 类总体性能与困难五类识别结果。总体指标用于判断数值证据能否形成有效增益，逐类结果用于定位增益的类别分布；两者结合可避免仅依据平均指标推断定向补偿效果。

Table 2. Overall performance comparison of different methods
表 2. 不同方法的整体性能对比

方法	Accuracy/%	较纯文本提升/百分点	困难五类平均准确率/%
纯文本	92.68	-	84.21
Static	93.37	+0.69	87.25
TMC	92.81	+0.13	84.64
PDF	93.63	+0.95	86.80
MNL	93.68	+1.00	87.27
QMF	93.76	+1.08	87.97
CAAF	94.51	+1.83	89.35

表 2 列出了不同方法在 TEP 21 类任务上的整体结果。纯文本分支的准确率为 92.68%，构成统一的文本基准。静态融合将准确率提高至 93.37%，较纯文本分支增加 0.69 个百分点，表明稳定统计特征能够提供有效补充。TMC 的准确率为 92.81%，较纯文本分支提高 0.13 个百分点；PDF 和 MNL 分别达到 93.63%和 93.68%；QMF 取得 93.76%，为对比融合方法中的最优结果。

CAAF 的整体准确率达到 94.51%，较纯文本分支、静态融合和 QMF 分别提高 1.83、1.14 和 0.75 个百分点。与采用全局权重或样本级权重的方法相比，CAAF 进一步在候选类别维度控制数值证据的介入程度，能够利用类别分布不均的互补信息。困难五类平均准确率由纯文本分支的 84.21%提高至 89.35%，说明整体增益主要伴随困难类别识别能力的改善。

上述结果说明 CAAF 在总体指标上具有稳定优势，但平均性能只能反映方法间的总体差异，不能确定增益是否集中于文本分支的薄弱类别。为明确性能改善的类别分布，随后进一步比较 IDV-16 至 IDV-20 的逐类结果。

进一步以 IDV-16 至 IDV-20 的逐类准确率和五类平均值为比较对象。上述故障的变量响应重叠程度较高，文本分支在这些类别上的识别性能低于总体水平，因而能够更直接地反映不同融合策略是否利用数值信息补偿了文本分支的集中错判。

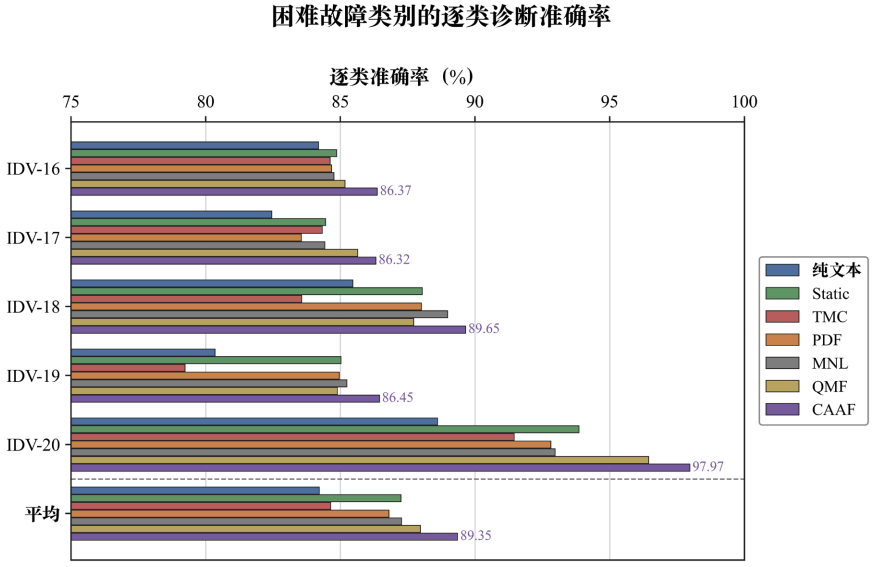


Figure 5. Comparison of per-class diagnostic accuracy on difficult fault classes
图 5. 困难故障类别的逐类诊断准确率对比

图5给出了不同方法在IDV-16至IDV-20上的逐类准确率。纯文本分支的五类平均准确率为84.21%，较其 21 类总体准确率低 8.47 个百分点，表明文本表示在困难故障上的判别能力明显下降。静态融合、PDF、MNL 和 QMF 分别将五类平均准确率提高至 87.25%、86.80%、87.27%和 87.97%。各方法均获得不同程度的改善，说明稳定统计信息能够补充结构化文本未充分保留的局部动态特征。

CAAF 在五类困难故障上的平均准确率达到 89.35%，较纯文本分支和 QMF 分别提高 5.14 和 1.38 个百分点。在 IDV-16、IDV-17、IDV-18、IDV-19 和 IDV-20 上，CAAF 的准确率分别为 86.37%、86.32%、89.65%、86.45%和 97.97%，均高于纯文本分支。其中，IDV-20 的数值分支独立准确率并不占优，但图 4 显示该类文本错判样本仍具有较大的可救回空间。类别系数提高了有效数值证据在该类候选通道上的修正作用，样本门控则限制该修正对高置信文本样本的影响。逐类结果与互补性分析方向一致，说明 CAAF 的总体增益主要伴随困难类别识别能力的改善。该结论限于 TEP 中的 5 类困难故障，两个调节组件的独立贡献仍需结合消融结果判断。

3.4. 融合组件消融分析

为区分类别系数和样本门控的独立作用，本节以 Static 为参照，依次比较仅保留类别系数的 Ours-A、仅保留样本门控的 Ours-B 和完整 CAAF。比较同时覆盖整体准确率、加权 F1 和困难五类平均准确率，以判断各组件的贡献是否在总体任务与困难类别上保持一致。

消融实验比较 Static、Ours-A、Ours-B 和完整 CAAF。Static 仅学习全局标量 α；Ours-A 保留类别系数 α_c，不使用样本门控；Ours-B 仅保留文本置信度门控 $g(x)$，各类别共享同一基础系数；完整 CAAF 同时使用 α_c 和 $g(x)$。各变体采用相同的数值头、训练轮数、数据划分和模型选择规则。

Table 3. Ablation results of CAAF components
表 3. CAAF 模块消融结果

方法	类别系数	样本门控	Accuracy/%	加权 F1/%	困难五类平均准确率/%
Static	否	否	93.3700	93.3500	87.2500
Ours-A	是	否	94.2176	94.1983	88.9365
Ours-B	否	是	93.9284	93.9071	88.2742
CAAF	是	是	94.5100	94.5000	89.3500

表3给出了 CAAF 各组件的消融结果。Ours-A 和 Ours-B 的整体准确率分别为 94.2176%和 93.9284%，较 Static 分别提高 0.8476 和 0.5584 个百分点，说明类别系数与样本门控均能独立改善融合性能。完整 CAAF 的准确率达到 94.5100%，较 Ours-A 和 Ours-B 分别提高 0.2924 和 0.5816 个百分点。CAAF 相对 Ours-B 的增益反映类别维度建模的贡献，相对 Ours-A 的增益则表明样本门控能够在类别修正基础上进一步改善样本级选择。

困难五类的消融结果与整体性能排序一致。Static、Ours-B、Ours-A 和 CAAF 的五类平均准确率依次为 87.2500%、88.2742%、88.9365%和 89.3500%。CAAF 较 Ours-B 提高 1.0758 个百分点，较 Ours-A 提高 0.4135 个百分点，类别系数带来的增益在困难故障上更为明显。加权 F1 与 Accuracy 的排序一致且数值接近，与测试集类别均衡的设置相符。结果表明，类别系数和样本门控分别从候选类别与当前样本两个维度调节数值证据，二者共同构成样本–类别级融合。该消融结论建立在相同数值头和固定数据划分下，不用于推断不同模型容量或跨数据集条件下的性能变化。

3.5. 统计稳定性及融合机制验证

单次性能比较和组件消融仍不足以排除随机初始化、启发式阈值设置、新增参数或权重统计偶合等替代解释。本节首先以纯文本分支、Static、QMF 和 CAAF 为比较对象，通过 10 次独立重复检验性能差异的稳定性；随后改变跨运行容差 ε，观察候选特征保持率、最终入选集合和诊断性能；最后结合数值信息破坏与权重－互补性对应关系，分析性能增益的信息来源及所学权重是否符合方法设计。

各可训练融合方法采用相同的 10 个随机种子重复训练，每次实验共享数据划分和文本类别得分，仅改变数值头与融合模块的随机初始化。95%置信区间依据样本均值、样本标准差和 t 分布临界值计算。CAAF 与 Static、QMF 之间基于相同随机种子下的结果进行配对 t 检验；与固定文本分支比较时，对逐次准确率差值执行相对于 0 的单样本 t 检验。多重比较采用 Holm 方法校正。

Table 4. Multi-seed repeated experiments and significance tests
表 4. 多种子重复与显著性检验

方法	Accuracy 均值/%	标准差/%	95%置信区间半宽/%	困难五类平均准确率/%	与 CAAF 比较的 p 值
纯文本	92.6800	0.0000	0.0000	84.2100	<0.0001
Static	93.3681	0.0685	0.0490	87.2479	<0.0001
QMF	93.7649	0.0553	0.0396	87.9688	<0.0001
CAAF	94.5139	0.0454	0.0325	89.3479	-

表 4 列出了多种子重复实验结果。Static、QMF 和 CAAF 的平均准确率分别为 93.3681%、93.7649%和 94.5139%，与单次实验的性能排序一致；三者的标准差依次为 0.0685%、0.0553%和 0.0454%，95%置信区间半宽分别为 0.0490%、0.0396%和 0.0325%。纯文本分支采用固定类别得分，不受融合模块随机初始化影响，其准确率保持为 92.6800%。CAAF 在取得最高平均准确率的同时未表现出更大的训练波动。

差值检验结果表明，CAAF 相对纯文本分支、Static 和 QMF 的校正后 p 值均小于 0.0001，平均提升分别为 1.8339、1.1458 和 0.7490 个百分点。上述提升幅度与单次实验结果基本一致，说明 CAAF 相对于主要基线的性能优势在不同随机初始化下保持稳定。该结果反映固定数据划分下的训练稳定性，不代表方法已经具备跨工况或跨数据集泛化能力。

3.5.1. 跨运行容差敏感性分析

为检验稳定特征筛选对容差设置的依赖，在保持特征入选阈值 $\tau = 0.5000$ 不变的条件下，将 ε 依次设为 0.1000、0.1500、0.2000、0.2500、0.3000 和 0.3500，并重新计算 8 类候选统计特征的跨运行保持率。对于每个 ε，依据 $Rf \geq \tau$ 确定数值分支的输入特征，在相同数据划分和 10 个随机种子设置下独立重跑训练流程。模型性能报告 Accuracy 均值、标准差、95%置信区间半宽和困难五类平均准确率。

Table 5. Cross-run retention rates of candidate statistical features under different ε values
表 5. 不同 ε 下候选统计特征的跨运行保持率

ε	主频幅值占比	主频位置	单调性	标准差	最大偏离幅度	总变化量	最终值	均值
0.1000	0.6118	0.4321	0.4185	0.2182	0.2074	0.0711	0.0522	0.0280
0.1500	0.7062	0.5136	0.4964	0.2945	0.2805	0.0946	0.0691	0.0391
0.2000	0.7806	0.5625	0.5521	0.3732	0.3523	0.1232	0.0930	0.0580
0.2500	0.8283	0.6159	0.6033	0.4288	0.4068	0.1445	0.1122	0.0674
0.3000	0.8608	0.6533	0.6415	0.4761	0.4517	0.1762	0.1331	0.0816
0.3500	0.8879	0.6827	0.6719	0.5136	0.4899	0.1990	0.1549	0.0965

表 5 显示，8 类候选特征的保持率均随 ε 放宽而单调增加，但特征间的相对排序保持稳定。$E=0.1000$ 时仅主频幅值占比达到入选阈值；$\varepsilon=0.1500$ 时主频位置进一步入选；当 ε 处于 0.2000 至 0.3000 时，入选集合稳定为主频幅值占比、主频位置和单调性。ε 增至 0.3500 后，标准差的保持率达到 0.5136 并进入数值分支。该结果说明，ε 主要影响临界特征是否通过筛选，三个核心统计特征在中等容差区间内具有稳定的跨运行优势。

Table 6. Effects of ε on selected features and CAAF performance
表 6. ε 对入选特征和 CAAF 性能的影响

ε	入选特征	特征数	Accuracy 均值/%	标准差/%	95%置信区间半宽/%	困难五类平均准确率/%	相对 $\varepsilon=0.2000$ 变化/百分点
0.1000	主频幅值占比	1	94.0137	0.0720	0.0515	88.1996	−0.5335
0.1500	主频幅值占比、主频位置	2	94.2744	0.0589	0.0421	88.7615	−0.2728
0.2000	主频幅值占比、主频位置、单调性	3	94.5472	0.0451	0.0323	89.3456	0.0000
0.2500	主频幅值占比、主频位置、单调性	3	94.5472	0.0451	0.0323	89.3456	0.0000
0.3000	主频幅值占比、主频位置、单调性	3	94.5472	0.0451	0.0323	89.3456	0.0000
0.3500	主频幅值占比、主频位置、单调性、标准差	4	94.2483	0.0629	0.0450	88.8199	−0.2989

表 6 进一步表明，ε 过小时筛选条件较严格，数值分支因缺少主频位置或单调性而出现性能下降。$\varepsilon=0.1000$ 和 0.1500 时，Accuracy 均值分别为 94.0137%和 94.2744%，较 $\varepsilon=0.2000$ 分别降低 0.5335 和 0.2728 个百分点。当 ε 处于 0.2000 至 0.3000 时，入选特征集合不变，Accuracy 均值稳定为 94.5472%，标准差为 0.0451%，95%置信区间半宽为 0.0323%，困难五类平均准确率为 89.3456%。$\varepsilon=0.3500$ 时引入标准差，Accuracy 均值降至 94.2483%，困难五类平均准确率降至 88.8199%，说明放宽容差后纳入的弱稳定特征会引入额外运行差异。敏感性实验中的 $\varepsilon=0.2000$ 结果与表 4 的独立多种子实验相差 0.0333 个百分点，差异小于两组置信区间半宽，且不改变性能排序。因此，本文保留 $\varepsilon=0.2000$，它是稳定区间内满足三类核心特征入选条件的最小取值。

3.5.2. 数值信息有效性验证

融合性能的提高可能来自新增参数、类别偏置或训练过程中的间接正则化作用，而不一定来自统计数值内容本身。为检验数值证据与性能增益之间的对应关系，本节固定同一 CAAF 模型参数，仅在测试阶段破坏样本对应关系或变量语义，并观察整体及困难类别性能的变化。

数值信息破坏实验设置三种输入：原始统计特征、沿样本维度随机打乱的统计特征，以及保持样本不变但随机置换 52 个物理变量与三类统计量对应位置的统计特征。样本维打乱破坏数值内容与当前样本及其标签之间的对应关系；变量对应打乱保留样本级统计分布，但破坏输入维度所承载的变量身份和统计语义。两种置乱均使用 10 个随机种子重复，实验过程中固定模型参数，不重新训练受损输入。

图 6 给出了数值信息破坏前后的性能变化。使用原始特征时，Accuracy 与加权 F1 分别为 94.5100%和 94.5000%。样本维打乱后，两项指标分别降至 92.5873%和 92.5486%，其中 Accuracy 下降 1.9227 个百分点，并略低于纯文本分支的 92.6800%。这表明，当数值内容与当前样本失配时，错误的数值类别得分会抵消原有融合收益。变量身份与统计量对应关系打乱后，Accuracy 降至 93.2384%，较原始特征下降 1.2716 个百分点。该结果说明，变量身份是数值证据形成类别指向的重要依据。

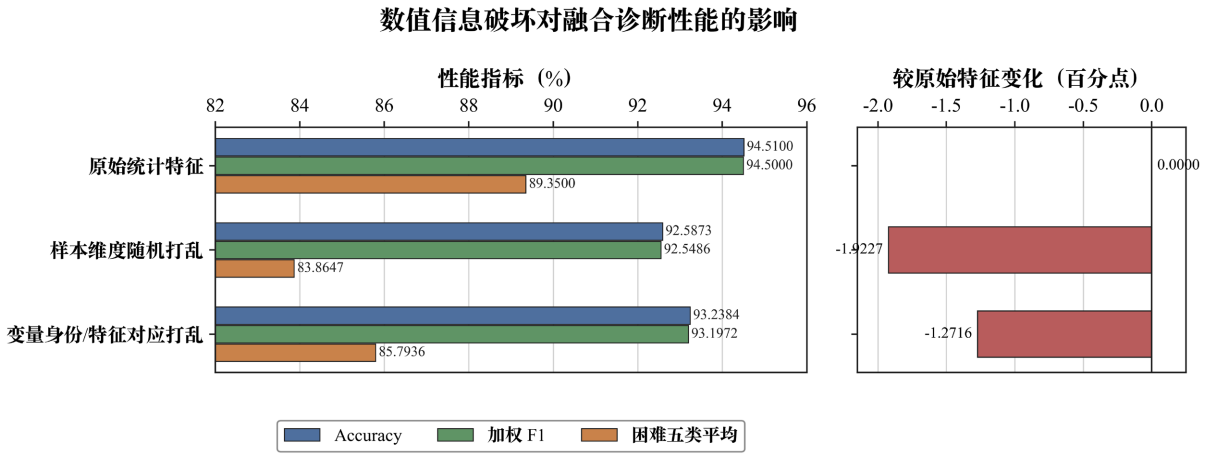


Figure 6. Effects of numeric information corruption on fusion diagnosis performance
图 6. 数值信息破坏对融合诊断性能的影响

困难五类对数值信息破坏表现出更高敏感性。样本维打乱和变量对应打乱后的五类平均准确率分别为 83.8647%和 85.7936%，较原始特征下降 5.4853 和 3.5564 个百分点。前者低于纯文本分支的 84.2100%，后者低于 Static 的 87.2500%。两种破坏方式均使困难类别性能明显回落，且样本对应关系破坏造成的损失更大。由此可见，CAAF 的增益依赖与当前样本匹配的数值内容及变量身份所提供的物理对应关系，不能仅由参数规模增加或类别偏置解释。该实验验证的是模型对人为对应关系破坏的响应，不等同于真实噪声、传感器漂移或变量缺失条件下的鲁棒性结论。

3.5.3. 自适应融合权重与互补性分析

数值信息破坏实验能够确认性能增益依赖有效数值内容，但不能说明所学权重是否与类别互补空间和文本样本不确定性一致。本节比较全部类别与困难类别的类别系数、样本门控和有效权重，并进一步分析类别系数与可救回率的相关关系，以检验参数变化方向是否符合样本－类别级调节机制。

以真实类别通道上的 $\alpha(x,y)=\alpha_y\cdot g(x)$ 表示有效数值介入强度，并计算类别系数与图 4 所示可救回率之间的相关系数。由于式(8)采用非负参数化，α_c 和 $\alpha(x,y)$ 均表示介入强度，而不解释为概率。

Table 7. Correlation analysis between fusion weights and complementarity
表 7. 融合权重与互补性的相关分析

分析对象	平均 α_c	平均有效权重 $\alpha(x,y)$	平均门控 $g(x)$	与可救回率的 Pearson r
全部类别	1.1837	0.5604	0.4681	0.7416
IDV-16 至 IDV-20	1.4618	0.7913	0.5413	0.5037
困难五类中文本正确样本	-	0.7290	0.4987	-
困难五类中文本错误样本	-	1.1232	0.7684	-

表 7 列出了融合权重与互补性的统计结果。全部 21 类的平均 α_c 为 1.1837，困难五类为 1.4618，后者提高 23.4941%。全部类别与困难五类的平均有效权重分别为 0.5604 和 0.7913，平均门控值分别为 0.4681 和 0.5413。困难类别具有更大的类别系数和平均门控值，二者共同作用后，其有效权重较全部类别平均值提高 41.2027%。该结果与图 4 所示困难类别具有较大可救回空间的现象一致。

在困难五类中，文本正确样本与文本错误样本的门控均值分别为 0.4987 和 0.7684，后者提高 0.2697；对应有效权重由 0.7290 增至 1.1232。按照困难五类中文本正确与错误样本的比例(文本准确率为 0.8421)

加权后，平均门控值和有效权重分别为 0.5413 和 0.7913，与表 7 的汇总结果一致。文本错误样本获得更高的数值介入强度，表明样本门控的变化方向与文本置信度调节目标一致。

在可定义可救回率的 20 个类别中，α_c 与可救回率的 Pearson 相关系数为 0.7416，表明两者总体呈较强正相关。困难五类内的相关系数为 0.5037，但由于样本点仅为 5 个，该结果仅用于描述变化趋势，不作为显著性证据。类别系数、样本门控和有效权重的分析结果均与互补性统计方向一致，为 CAAF 的样本–类别级调节机制提供了参数层面的解释。相关分析本身不构成因果证据，其结论范围受类别数量和当前数据集限制。

4. 结论

本文针对工业时序文本化造成频域结构和局部变化信息压缩，以及弱数值分支的补偿作用随故障类别和样本状态变化的问题，提出基于跨运行稳定统计特征与类别感知自适应融合的流程工业故障诊断方法。首先，根据故障效应方向与幅度在两个独立开发运行间的一致性构建跨运行保持率，并筛选主频幅值占比、主频位置和单调性，通过变量身份映射与轻量编码器构建数值分支。随后，以文本错判样本的可救回率刻画两分支的类别相关互补空间，并将融合权重扩展为 $\alpha(x, c)$，利用类别系数和文本置信度门控分别调节数值证据的类别修正幅度与样本介入程度。

TEP21 类故障诊断结果表明，CAAF 的单次实验准确率达到 94.51%，较纯文本分支、静态融合和 QMF 分别提高 1.83、1.14 和 0.75 个百分点；在 IDV-16 至 IDV-20 五类困难故障上的平均准确率达到 89.35%，较纯文本分支提高 5.14 个百分点。模块消融结果表明，类别系数与样本门控均对性能提升产生贡献。跨运行容差敏感性实验进一步表明，当 ε 处于 0.2000 至 0.3000 时，入选特征集合保持不变，10 次重复训练的 Accucy 均值稳定为 94.5472%，标准差为 0.0451%；ε 过小会遗漏互补特征，ε 过大则会引入弱稳定特征。数值信息有效性验证和融合权重分析表明，性能增益与有效数值内容、变量身份及类别相关互补性存在明确对应关系。

本文实验仍以 TEP 单一基准为主，稳定统计特征的筛选范围限定于预设候选集合，敏感性分析也仅覆盖两个开发运行和有限的 ε 取值。当前模型尚未系统评估真实工业噪声、传感器漂移、变量缺失、新故障类型和跨装置迁移条件下的适应能力；固定变量身份映射及重复训练还会增加在线部署的存储与计算成本。后续将结合真实工业数据开展跨装置验证，研究不依赖结构化文本变量出现顺序的在线自适应特征选择方法，并进一步探索面向开放集故障识别的融合策略、噪声与缺失变量鲁棒训练以及轻量化部署。

基金项目

重庆科技大学研究生创新计划项目，项目编号：YKJCX2520515。

参考文献

[1] Venkatasubramanian, V., Rengaswamy, R., Yin, K. and Kavuri, S.N. (2003) A Review of Process Fault Detection and Diagnosis: Part I: Quantitative Model-Based Methods. *Computers & Chemical Engineering*, **27**, 293-311. https://doi.org/10.1016/s0098-1354(02)00160-6

[2] Yin, S., Ding, S.X., Haghani, A., Hao, H. and Zhang, P. (2012) A Comparison Study of Basic Data-Driven Fault Diagnosis and Process Monitoring Methods on the Benchmark Tennessee Eastman Process. *Journal of Process Control*, **22**, 1567-1581. https://doi.org/10.1016/j.jprocont.2012.06.009

[3] Downs, J.J. and Vogel, E.F. (1993) A Plant-Wide Industrial Process Control Problem. *Computers & Chemical Engineering*, **17**, 245-255. https://doi.org/10.1016/0098-1354(93)80018-i

[4] Rieth, C.A., Amsel, B.D., Tran, R., *et al*. (2017) Additional Tennessee Eastman Process Simulation Data for Anomaly Detection Evaluation. Harvard Dataverse.

[5] Lomov, I., Lyubimov, M., Makarov, I. and Zhukov, L.E. (2021) Fault Detection in Tennessee Eastman Process with Temporal Deep Learning Models. *Journal of Industrial Information Integration*, **23**, Article ID: 100216. https://doi.org/10.1016/j.jii.2021.100216

[6] Zhang, S. and Qiu, T. (2022) Semi-Supervised LSTM Ladder Autoencoder for Chemical Process Fault Diagnosis and Localization. *Chemical Engineering Science*, **251**, Article ID: 117467. https://doi.org/10.1016/j.ces.2022.117467

[7] Peng, P., Zhang, Y., Wang, H. and Zhang, H. (2022) Towards Robust and Understandable Fault Detection and Diagnosis Using Denoising Sparse Autoencoder and Smooth Integrated Gradients. *ISA Transactions*, **125**, 371-383. https://doi.org/10.1016/j.isatra.2021.06.005

[8] Lin, Z., Hu, Z., Peng, J. and Zhao, H. (2022) Dynamic-Scale Graph Neural Network for Fault Detection. *Process Safety and Environmental Protection*, **168**, 953-970. https://doi.org/10.1016/j.psep.2022.10.036

[9] Vaswani, A., Shazeer, N., Parmar, N., *et al*. (2017) Attention Is All You Need. arXiv: 1706.03762.

[10] Devlin, J., Chang, M., Lee, K. and Toutanova, K. (2019) BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the* 2019 *Conference of the North American Chapter of the Association for Computational Linguistics*: *Human Language Technologies*, *Volume* 1 (*Long and Short Papers*), Minneapolis, 2-7 June 2019, 4171-4186. https://doi.org/10.18653/v1/n19-1423

[11] Jin, M., Wang, S., Ma, L., *et al*. (2024) Time-LLM: Time Series Forecasting by Reprogramming Large Language Models. arXiv: 2310.01728.

[12] Zhou, T., Niu, P., Wang, X., Sun, L. and Jin, R. (2023) One Fits All: Power General Time Series Analysis by Pretrained LM. *Advances in Neural Information Processing Systems* 36, New Orleans, 10-16 December 2023, 43322-43355. https://doi.org/10.52202/075280-1877

[13] Gruver, N., Finzi, M., Qiu, S. and Wilson, A. (2023) Large Language Models Are Zero-Shot Time Series Forecasters. *Advances in Neural Information Processing Systems* 36, New Orleans, 10-16 December 2023, 19622-19635. https://doi.org/10.52202/075280-0861

[14] Ansari, A.F., Stella, L., Turkmen, A.C., *et al*. (2024) Chronos: Learning the Language of Time Series. arXiv: 2403.07815.

[15] Han, Z., Zhang, C., Fu, H. and Zhou, J.T. (2023) Trusted Multi-View Classification with Dynamic Evidential Fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **45**, 2551-2566. https://doi.org/10.1109/tpami.2022.3171983

[16] Zhang, Q., Wu, H., Zhang, C., *et al*. (2023) Provable Dynamic Fusion for Low-Quality Multimodal Data. *Proceedings of the* 40*th International Conference on Machine Learning*, Honolulu, 23-29 July 2023, 41753-41769.

[17] Zeng, Y., Mai, S. and Hu, H. (2021) Which Is Making the Contribution: Modulating Unimodal and Cross-Modal Dynamics for Multimodal Sentiment Analysis. *Findings of the Association for Computational Linguistics*: *EMNLP* 2021, Punta Cana, 7-11 November 2021, 1262-1274. https://doi.org/10.18653/v1/2021.findings-emnlp.109

[18] Wang, Y., Lei, Y., Li, N., Li, X. and Yang, B. (2025) Multimodal Correlation-Aware Fusion Framework for Enhanced Machinery Health Prognosis with Unlabeled and Low-Quality Data Exploitation. *IEEE Transactions on Neural Networks and Learning Systems*, **36**, 12040-12051. https://doi.org/10.1109/tnnls.2024.3453604

[19] McKinney, M., Garland, A., Cillessen, D., Adamczyk, J., Bolintineanu, D., Heiden, M., *et al*. (2025) Unsupervised Multimodal Fusion of In-Process Sensor Data for Advanced Manufacturing Process Monitoring. *Journal of Manufacturing Systems*, **78**, 271-282. https://doi.org/10.1016/j.jmsy.2024.12.003

[20] Peng, X., Wei, Y., Deng, A., Wang, D. and Hu, D. (2022) Balanced Multimodal Learning via On-the-Fly Gradient Modulation. 2022 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), New Orleans, 18-24 June 2022, 8238-8247. https://doi.org/10.1109/cvpr52688.2022.00806

[21] Wei, Y. and Hu, D. (2024) MMPareto: Boosting Multimodal Learning with Innocent Unimodal Assistance. *Proceedings of the 41st International Conference on Machine Learning*, Vienna, 21-27 July 2024, 52559-52572.

[22] Lakens, D. (2013) Calculating and Reporting Effect Sizes to Facilitate Cumulative Science: A Practical Primer for T-Tests and ANOVAs. *Frontiers in Psychology*, **4**, Article 863. https://doi.org/10.3389/fpsyg.2013.00863

[23] Qwen Team (2024) Qwen2.5 Technical Report. arXiv: 2412.15115.

[24] Cao, B., Xia, Y., Ding, Y., *et al*. (2024) Predictive Dynamic Fusion. *Proceedings of the* 41*st International Conference on Machine Learning*, Vienna, 21-27 July 2024, 5608-5628.

[25] Gong, B., Gao, X., Zhu, P., Hu, Q. and Cao, B. (2025) Multimodal Negative Learning. *Advances in Neural Information Processing Systems* 38, San Diego, 2-7 December 2025, 170087-170118. https://doi.org/10.52202/085713-5125