

多模态数据融合管道中的质量异常传播 与根因定位技术研究

李薇

广东 广州 510000

DOI:10.61369/ME.2025090033

摘 要： 多模态数据融合管道中，质量异常传播与根因定位技术至关重要。本文从机理研究、技术创新与工程实践三个维度，提出综合解决方案，包括分析异构数据流耦合强度、构建 DAG 拓扑表征、设计动态权重因果推理等算法，并通过故障注入测试、构建基准数据集等验证。方案有效提升了智能制造系统可靠性，还探讨了未来技术融合方向。

关 键 词： 多模态数据融合；质量异常根因定位；智能制造

Research on Quality Anomaly Propagation and Root Cause Localization Technology in Multimodal Data Fusion Pipeline

Li Wei

Guangzhou, Guangdong 510000

Abstract： Quality anomaly propagation and root cause localization techniques are crucial in multimodal data fusion pipelines. This article proposes a comprehensive solution from three dimensions: mechanism research, technological innovation, and engineering practice, including analyzing the coupling strength of heterogeneous data streams, constructing DAG topology representation, designing dynamic weight causal inference algorithms, and verifying them through fault injection testing and constructing benchmark datasets. The plan effectively improves the reliability of intelligent manufacturing systems and explores future directions for technology integration.

Keywords： multimodal data fusion; root cause localization of quality abnormalities; smart manufacturing

引言

随着《智能制造发展规划（2021 - 2025 年）》的颁布，智能制造领域备受关注，多模态数据融合管道在其中作用关键。多模态数据特征空间关联复杂，其数据质量评估、异构流耦合等对异常传播影响重大。基于此，本文从机理研究、技术创新与工程实践维度，提出面向多模态数据融合管道的质量异常治理综合解决方案，涵盖构建融合管道拓扑表征、设计跨模态根因定位技术框架等。这不仅提升智能制造系统可靠性，也为实际生产提供路径，顺应政策对智能制造发展的推动方向。

一、多模态数据融合管道的异常传播机理分析

（一）多模态数据特征耦合机制

多模态数据涵盖视觉、文本、时序等不同类型，其在特征空间中存在复杂关联特性。视觉数据可能通过图像特征与文本数据中的语义描述相关联，文本数据又可能和时序数据在事件发生的时间序列上存在联系。数据质量评估指标体系是关键，例如通过数据的完整性、准确性、一致性等指标来衡量数据质量。异构数据流耦合强度会显著影响异常传播，当耦合强度较高时，一个模态数据的异常可能更容易传播至其他模态。例如，图像数据的噪声干扰可能因与文本描述强耦合，导致文本文义理解出现偏差^[1]。量化这种耦合强度对异常传播的影响，有助于深入理解多

模态数据融合过程中异常传播的内在机制，为后续根因定位提供理论基础。

（二）管道拓扑结构的异常扩散模型

基于图神经网络构建融合管道的有向无环图（DAG）拓扑表征，为研究异常在各处理节点的级联扩散规律奠定基础。在多模态数据融合管道里，数据清洗、特征提取、决策推理等处理节点相互关联形成特定的拓扑结构^[2]。异常在这样的结构中扩散，从某个初始异常节点开始，沿着拓扑结构中的连接关系，向相邻节点传播。例如在数据清洗节点若出现数据缺失或错误的异常，经清洗的数据流入特征提取节点，可能导致特征提取不准确，从而将异常进一步传递至决策推理节点，影响最终决策结果。通过对这一异常扩散模型的深入分析，能清晰把握异常在多模态数据融

合管道拓扑结构中的传播路径与级联扩散规律，为后续质量异常根因定位提供关键依据。

二、跨模态根因定位技术框架设计

（一）动态权重因果推理算法

在跨模态根因定位技术框架设计中，动态权重因果推理算法是关键环节。该算法结合贝叶斯网络与注意力机制，旨在通过模态间特征重要性动态评估实现跨层级的根因追溯。贝叶斯网络能够有效描述多模态数据间的因果关系，而注意力机制则可精准聚焦于对根因定位起关键作用的特征。在算法运行过程中，随着数据的流动与融合，各模态特征的重要性并非一成不变，而是根据数据特点及融合需求进行动态调整。通过这种动态评估，能够在复杂的多模态数据融合管道中，敏锐捕捉质量异常传播路径，并准确锁定根因。此算法为跨模态根因定位提供了一种可解释的有效途径，有助于提高多模态数据融合质量^[3]。

（二）增量式定位效率优化方法

在多模态数据融合管道中，为提升跨模态根因定位效率，开发面向在线场景的增量式定位算法是关键。设计基于管道处理日志的异常传播路径剪枝策略，能有效减少不必要的计算。通过对管道处理日志的深度分析，识别出那些不太可能是根因的传播路径并予以剪枝^[4]。同时，增量式定位算法可避免重复计算，仅对新出现的异常数据进行增量处理。当有新的异常数据流入时，算法依据剪枝后的路径，快速定位根因，大大降低计算复杂度，使得在在线场景下，能够高效且实时地对多模态数据融合过程中的质量异常进行根因定位，提升系统对异常情况的响应速度与处理能力。

三、工程化实现与验证体系

（一）分布式计算架构实现

1. 微服务化质量监控组件

为实现多模态数据融合管道中的质量异常传播与根因定位，构建容器化部署的异常检测微服务是关键。该微服务能实现对数据处理节点的细粒度监测与异常预警。通过将异常检测功能封装为微服务，利用容器技术进行部署，可提升系统的灵活性与可扩展性。微服务能精准感知每个数据处理节点的运行状态和数据质量，一旦发现异常，立即触发预警机制。例如，对图像、文本等多模态数据处理环节中的数据缺失、格式错误等问题及时察觉并发出警报。这一微服务化质量监控组件的实现，基于分布式计算架构，确保了多模态数据融合过程中数据质量的实时把控。通过实际应用场景的验证^[5]，证明其能有效提升异常检测效率与准确性，为根因定位提供有力支持。

2. 弹性资源调度机制

在多模态数据融合管道的根因定位任务面临突发负载时，开发基于 Kubernetes 的动态资源分配策略以保障服务质量至关重要。借助 Kubernetes 强大的容器编排能力，能够实时感知根因定位任务的负载变化^[6]。通过对任务资源需求的动态监测，依据

负载的波动情况，自动且灵活地为根因定位任务分配计算资源，如 CPU、内存等。当突发负载出现，可迅速增加资源供给，确保任务处理效率不受影响，维持较高的服务质量；而在负载降低时，及时回收多余资源，避免资源浪费，提升资源利用率。这种弹性资源调度机制，基于 Kubernetes 实现了动态且精准的资源分配，有力保障了根因定位任务在复杂负载场景下的稳定运行与服务质量。

（二）全生命周期验证体系

1. 故障注入测试平台

在多模态数据融合管道的工程化实现与验证体系中，故障注入测试平台具有关键作用。搭建能模拟数据污染、节点宕机等典型故障模式的仿真环境，此环境需尽可能真实地反映多模态数据融合管道在实际运行中可能遭遇的异常情况^[7]。通过在该仿真环境中注入不同类型和程度的故障，对异常传播模型的准确性展开验证。密切观察故障注入后的数据流向与变化，分析异常在多模态数据融合过程中的传播路径、影响范围和程度。依据实际观测结果与异常传播模型的预测进行对比，评估模型的准确性与可靠性，若存在偏差，及时对模型进行优化调整，从而确保多模态数据融合管道在面对各类故障时，异常传播模型能准确反映实际情况，为后续根因定位奠定坚实基础。

2. 工业级基准数据集构建

在多模态数据融合管道的工程化实现与全生命周期验证体系中，工业级基准数据集构建至关重要。首先，从智能制造产线采集多源异构运维数据，这些数据涵盖设备运行参数、环境监测数据、故障报警信息等不同类型，其异构性为后续融合带来挑战。对采集的数据进行预处理，包括数据清洗以去除噪声和错误值，数据标准化使不同尺度的数据具备可比性。之后，通过专业领域知识与实际运维经验，为数据标注根因，明确质量异常的根源。最终构建出包含标注根因的测试基准数据集，此数据集将为多模态数据融合管道中的质量异常传播与根因定位技术验证提供可靠依据，助力相关技术在工业场景下的有效应用^[8]。

四、工程管理应用与实证研究

（一）智能工厂质量控制实践

1. 工艺参数异常溯源案例

在汽车焊接产线这一智能工厂质量控制场景中，多模态数据融合管道的质量异常传播与根因定位技术发挥了关键作用。产线出现批次质量缺陷问题，通过运用根因定位技术，对各类多模态数据进行深入分析。视觉传感器数据作为其中重要的一环，经过与其他工艺参数数据融合分析，发现视觉传感器发生畸变是导致此次批次质量缺陷的根源。这一成功案例表明，精准的根因定位技术能够有效追溯工艺参数异常，帮助产线迅速锁定问题源头，及时采取措施解决，避免更多质量问题产生，为智能工厂高效稳定生产提供有力保障，也进一步验证了多模态数据融合管道中根因定位技术在智能工厂质量控制实践中的可行性与有效性^[9]。

2. 设备健康管理应用场景

在智能工厂质量控制实践的设备健康管理应用场景中，通过

振动传感器与过程数据的跨模态分析实现 CNC 机床早期故障精准定位意义重大。振动传感器可捕捉机床运行时的振动信号，反映机械部件的运行状态；过程数据则记录加工过程中的各类参数信息。利用多模态数据融合技术，将两者结合分析，能深入挖掘设备潜在问题。例如，通过对振动频率与加工参数关联分析，若振动频率异常且对应加工参数出现波动，可初步判断故障区域。进一步深度融合分析，依据数据特征建立故障模型，精准定位早期故障位置与原因^[10]。这一应用场景可有效提高设备运行稳定性，保障产品质量，为智能工厂的高效生产提供有力支撑。

（二）供应链协同优化方案

1. 物流信息流异常处理

在供应链协同优化方案中，物流信息流异常处理至关重要，尤其是解决 RFID 与 GIS 数据异步导致的库存预测偏差问题，对提升供应链响应速度意义重大。RFID 数据主要用于实时追踪货物位置与数量，而 GIS 数据提供地理空间信息，二者异步易使库存预测出现偏差。为解决该问题，可构建统一的数据集成平台，使 RFID 与 GIS 数据能实时同步与交互。通过优化数据采集频率与传输机制，确保两种数据在时间维度上的一致性。同时，运用数据清洗与预处理技术，对异步数据进行规整。经实证研究，该处理方式有效纠正库存预测偏差，供应链响应速度显著提升，货物调配更及时、准确，降低了库存积压与缺货风险，实现供应链各环节的高效协同。

2. 供应商质量追溯系统

建立覆盖多级供应商的质量数据融合管道，能有效缩短原材料质量问题追溯周期。在实际工程管理应用中，通过该质量数据融合管道，可将来自不同层级供应商的各类多模态质量数据，如原材料的物理特性数据、生产流程中的参数数据以及检测报告等进行整合。这样一来，当出现质量异常时，能够快速且精准地沿着数据管道追溯到问题根源，极大减少追溯所需的时间与人力成本。以某大型制造业企业为例，在构建该质量数据融合管道前，原材料质量问题追溯往往需要数周时间，且定位不够准确。而实施后，追溯周期缩短至数天，显著提升了供应链协同效率，保障了产品质量，为企业带来了可观的经济效益，充分验证了这一供应商质量追溯系统在供应链协同优化中的重要价值。

（三）综合效益量化评估

1. 质量成本降低分析

在多模态数据融合管道中，对质量成本降低进行分析，主要

通过统计应用前后废品率、返工工时等关键指标来实现。废品率直接反映产品质量状况，若应用技术后废品率显著下降，意味着生产中浪费的原材料、人力等资源减少，直接降低了生产成本。返工工时同样关键，返工意味着额外的人力投入和时间成本，若返工工时减少，表明产品一次合格率提高，质量成本随之降低。通过对比技术应用前后这些指标的变化，能清晰验证该技术在质量成本降低方面带来的直接经济效益，为工程管理的多模态数据融合管道场景下的质量成本管控提供有力的量化依据。

2. 系统可靠性提升验证

在多模态数据融合管道中，系统可靠性至关重要。通过 MTBF（平均故障间隔时间）、故障恢复时间等指标评估技术对系统鲁棒性的改善效果，以此验证系统可靠性的提升。MTBF 指标可直观反映系统两次故障之间的平均时间，数值越高表明系统越稳定，能长时间正常运行，减少故障发生频率。故障恢复时间则体现系统从故障状态恢复到正常工作状态所需时长，该时间越短，系统在面对故障时的自我修复能力越强，可快速恢复正常工作，保障数据融合工作的连续性。通过对这些指标的量化评估，能够精准判断技术在提升系统可靠性方面的实际成效，为多模态数据融合管道的优化提供有力依据，确保系统在复杂多变的环境中稳定运行。

五、总结

多模态数据融合管道在智能制造等领域发挥着关键作用，而其中质量异常传播与根因定位技术至关重要。本文提出的面向多模态数据融合管道的质量异常治理综合解决方案，从机理研究、技术创新与工程实践三个维度着手，有效提升了智能制造系统的可靠性。这不仅深化了对多模态数据融合中质量异常问题的理解，也为实际生产提供了可行的技术路径。同时，联邦学习在跨企业数据融合中的应用前景广阔，能突破数据孤岛，实现更广泛的数据价值挖掘；数字孪生技术支持下的预测性维护发展方向，有望进一步提升系统的运行稳定性与维护效率。未来，持续探索这些技术的融合与创新，将为多模态数据融合管道的质量异常治理带来新的突破与发展。

参考文献

- [1] 吴愚. 多模态融合感知技术研究与设计 [D]. 电子科技大学, 2021.
- [2] 王玉珏. 工业物联网中的多模态数据融合方法研究 [D]. 华中科技大学, 2022.
- [3] 郑德重. 基于多模态数据融合的肺部肿瘤智能分析技术研究 [D]. 中国科学院大学, 2021.
- [4] 郭朔. 基于多模态数据融合的图像描述算法研究与应用 [D]. 郑州大学, 2023.
- [5] 王剑. 多模态数据融合的肿瘤预后分析研究 [D]. 南方医科大学, 2023.
- [6] 刘沛然. 基于智能教育领域的多模态数据融合应用研究 [J]. 智能城市, 2022, 8(11): 13-15.
- [7] 黄美琳. 基于多模态数据融合的大学生英语听说教学改革研究 [J]. 文教资料, 2023(11): 110-113.
- [8] 孟凡会, 王玉亮, 汪雷. 基于深度学习的农村电商多源多模态数据融合 [J]. 滁州学院学报, 2023, 25(2): 52-61.
- [9] 李晓娜, 雷娜. 多模态数据融合下智能教育伦理问题及应对策略 [J]. 中国信息技术教育, 2023(8): 91-94.
- [10] 张莉. 基于多模态数据融合的社区智慧停车系统设计 [J]. 无线互联科技, 2023, 20(16): 12-17.