

# 基于大数据分析的制丝车间烟机设备机械故障模式识别 与预防策略研究

廖综岑 贺文懂

红塔烟草（集团）有限责任公司昭通卷烟厂 云南 昭通 657000

**【摘要】**：烟草生产要求烟机设备长期保持稳定运行状态，但在复杂工况下，设备易发各种机械故障，进而对生产效率和产品质量造成直接影响。针对故障问题，定期检修、事后处理自有其局限性，难以早期预警故障，也难以精准干预。应用数据分析技术则可以实现以上功能，弥补传统维护策略的不足。本文研究基于大数据分析的制丝车间烟机设备机械故障模式识别方法，探索相应的主动预防策略，旨在探索实现预测性维护的路径，提升运维智能化水平。

**【关键词】**：大数据分析；制丝车间烟机设备；机械故障；故障模式识别；预防策略

DOI:10.12417/2705-0998.25.24.056

在现代工业生产中，机械设备的可靠性、稳定性会直接影响企业的生产效率与经济效益。烟草工业是我国传统支柱产业之一，整个生产过程对机械设备有较高的依赖性。然而，烟草加工机械种类繁多、工艺复杂，加之所处环境复杂，难免出现设备故障问题，严重影响生产进度，影响产品质量。为提高设备可靠性，减少故障停机时间，烟草企业亟需先进的故障预测与维护技术支持。基于大数据分析等现代信息技术则可以采集、分析、建模设备运行数据，及早发现设备潜在故障，制定针对性地预防、维护策略，实时监控设备状态，实现预测性维护<sup>[1]</sup>。

## 1 制丝车间烟机设备概述

在烟草工业生产流程中，制丝车间的烟机设备承担烟叶到烟丝定型过程中的多项任务。例如，切丝机负责将烟叶、烟梗切成规定尺寸，烘丝机去除多余水分、提升感官品质，加香（料）机用于均匀施加料液。这些设备不仅结构精密，协同作业，也需要克服高温、高湿、多振动的复杂工况，保持连续稳定的运行状态。烟机设备机械故障模式多样，切丝机可能出现刀门堵塞、刀片磨损、主轴振动异常等问题；烘丝机可能出现筒体传动系统失效、热风系统故障以及出现各类设备普遍存在的轴承失效、齿轮磨损、传送带打滑等问题。各类故障会增加非计划停机的可能性，一旦发生非计划停机，会严重影响产量，造成巨大经济损失，也会对烟丝的宽度、温度、含水率等质量指标造成负面影响，使产品不同批次间出现较大质量差异。所以，有必要精准识别烟机设备故障模式，加强故障预测。

## 2 适用于故障诊断的机器学习与深度学习算法简介

在设备故障诊断领域，采用机器学习与深度学习算法具有较高价值，可以从海量、高维的监测数据中自动学习并提取故障特征，识别故障模式。机器学习方法以专家经验为依托进行特征工程，例如支持向量机（SVM），可用于分类小样本、高维度故障，处理非线性问题的能力较强；随机森林（RF）抗过

拟合能力强，可评估特征重要性，用于多故障分类有较多优势；梯度提升决策树则在预测精度上具有优异的表现。深度学习方法可以自动从振动信号、时序数据等原始数据中学习层次化特征表达，在处理复杂、高维的工业大数据时有优越表现。深度学习方法典型算法较多，卷积神经网络（CNN）是代表之一，该算法可从振动信号的频谱图、时频图中提取出空间局部特征，在识别轴承、齿轮故障时收集这些特征有积极意义<sup>[2]</sup>；长短时记忆网络（LSTM）可有效捕捉传感器时序数据中的长期依赖关系，可将之用于描述设备性能的退化趋势，也可以用于故障预测<sup>[3]</sup>；两者也可以结合应用，建立 CNN-LSTM 混合模型，同时提取时空特征，可发挥二者的优势。

## 3 制丝车间烟机设备大数据平台构建与数据处理

### 3.1 面向故障分析的大数据平台逻辑架构与技术选型

大数据平台逻辑架构采用分层、解耦的设计思想，自下而上可设置为数据采集与接入层、数据存储与计算层、数据分析与建模层以及智能应用与可视化层。技术选型时，可基于开源大数据生态构建平台核心。采集层实时流数据接入与初加工选用 Flink 或 Spark Streaming。存储层采用混合架构，使用 HDFS 或对象存储存放原始数据与模型文件；使用 Kafka 缓冲实时数据；设备传感器数据使用时序数据库存储、查询；使用关系型数据库存储业务与维护数据。计算分析层选用 Spark MLlib 和 Flink ML 进行大规模特征工程，训练机器学习模型；深度神经网络模型采用 TensorFlow 或 PyTorch 深度学习框架构建，依赖 Kubernetes 调度弹性资源。应用层通过 Web 服务封装模型 API，应用可视化工具实现健康状态仪表盘显示、预警看板等功能。

### 3.2 数据预处理流程与方法

原始工业数据必须预处理后再使用，本身含有大量噪声、缺失值、量纲不一等问题。预处理需要先清洗数据，对于因传感器失效、通信中断导致的数据缺失，修复工作可采用前后向填充、线性插值或基于 K 近邻（KNN）的算法；明显的异常

值可基于 $3\sigma$ 准则、箱线图或业务规则识别，之后剔除。数据集成与对齐需要以统一时间戳为依据，时间对齐、关联来自不同系统的多源数据，建立以设备-时间为主键的宽表。对于关键的振动与电流信号，其中的故障特征频率成分可采用数字滤波或小波阈值去噪方法分离出来。最后是特征工程，需要将具有物理意义的时域特征、频域特征以及能同时反映时频特性的时频域特征从清洗后的时序数据中提取出来，为模型输入做好准备。

### 3.3 故障数据标签化与样本库构建策略

工业场景下，常面临故障样本稀少、标注困难的情形，要想成功建立模型，必须构建高质量的标签化样本库。在实际工作中，需基于设备维修工单（EAM），结合 SCADA 异常报警记录、操作工日志以及对应时间段的传感器数据，发挥领域专家的作用，由领域专家确认故障发生时刻、类型及严重程度。针对样本不平衡问题，在处理时可采用合成少数类过采样技术（SMOTE）或其变种在特征空间生成模拟的故障样本；海量的正常样本则降采样，通过这一方法平衡正负样本比例。样本库采用分层设计，在标注故障的同时也标注设备所处的健康状态下的数据，健康状态可分为“健康”“退化”“亚健康”等几种，与具体部件关联，形成结构化知识。面对不断变化发展的环境，需要持续学习、增量更新，设计样本库管理机制，模型误报以及新发故障案例均由专家确认，之后纳入样本库，使知识库始终与模型匹配，协同进化。

## 4 基于大数据分析的烟机设备机械故障模式识别模型构建

### 4.1 基于多源数据融合的设备健康状态评估指标构建

为全面量化设备的健康状况，有必要融合多源数据构建分层次的健康状态评估指标体系。指标体系可分为核心性能指标、工艺质量关联指标和历史状态演化指标三个方面。在核心性能指标方面，通过设备传感器收集数据，之后处理信号，提取信息，以此判断机械与电气系统的实时物理状态。工艺质量关联指标可揭示设备状态对产品质量的潜在影响，此类指标通过关联 MES 工艺参数与设备状态计算。基于时序数据计算历史状态演化指标，关键性能指标的滑动窗口均值变化趋势、退化指数等均属于历史状态演化指标。最后，通过熵权法、专家打分确定权重进行多指标融合或采用深度学习模型自动加权<sup>[4]</sup>，构建综合的“设备健康指数（EHI）”，通过该指数反映设备健康状态的连续退化过程，为后续预警、预测提供量化依据。

### 4.2 故障模式识别模型选择与设计

遵循“分而治之，集成决策”的原则。因数据特性不同，故障模式也不同，所以可采用组合模型架构。高频振动以及声音信号本身富含时空局部特征，可设计特征提取器提取特征，

例如采用一维卷积神经网络（1D-CNN）或二维 CNN，二者均可自动学习故障的冲击性、周期性模式，应用效果较好。对于温度、压力、流量序列等多通道、多变量的工艺时序数据可用长短期记忆网络（LSTM）捕捉动态时序依赖关系、退化趋势，也可以采用门控循环单元（GRU）。对于已构建的统计特征集，分类时可将之输入鲁棒性强的梯度提升决策树或随机森林等模型处理。以上工作完成后，设计模型融合层，将上述多个机器学习器输出的故障概率或特征向量作为元特征，使用逻辑回归等元学习器综合决策，以便高质量识别不同故障模式，提升泛化能力。

### 4.3 模型训练、优化与验证方法

训练模型采用“滚动时间窗口”策略，明确划分训练集、验证集以及测试集，保证时间序列的因果关系。在训练过程中，针对样本不平衡，可采用代价敏感学习为少数类故障样本赋予更高错分权重，也可以结合应用批内过采样技术。优化模型分两步进行，先通过网格搜索或贝叶斯优化自动调优超参数，为防止过拟合，引入早停法；提升模型泛化能力可应用 Dropout、L2 正则化等技术。对于验证工作，采用分层 K 折交叉验证，按故障类别分层，可对模型稳定性进行评估，在独立的、代表未来时间段的测试集上提供性能报告。

## 5 基于故障模式识别的主动预防策略

### 5.1 建立从故障识别到预防性维护的决策映射

准确“识别”是实现“预防”的基础，要发挥故障模式识别的作用，驱动主动决策，就必须建立精准的决策映射机制。制定映射规则需考虑故障的严重性、发展速度、备件可用性 & 生产计划。例如，“轴承早期点蚀”属于低严重性、慢发展的范畴，对此，可映射为“加强监测，结合下次计划停机更换”；“刀片严重磨损”属于高严重性、直接影响质量的范畴，直接映射为“建议在 24 小时内安排短时维护”较好。映射过程与设备知识库、实时生产上下文深度集成，提供具体的行动指令，即检查什么、何时做、做什么，为现场工程师做出决策提供支持，实现预测性维护。

### 5.2 构建分级预警机制

为管理风险、避免预警疲劳，设置多级预警机制很有必要。机制的判断依据为综合“设备健康指数”以及关键单一指标，设置“注意”“预警”“报警”三级动态阈值，分别对应蓝色、黄色、红色三种颜色指示。对于健康指数轻微偏离基准或单一指标出现早期异常波动的情形，归入“注意”（蓝色），由系统自动记录，提示点检关注，不中断生产。对于健康指数处于持续恶化状态或通过模型识别出存在特定故障模式但置信度中等的情形，归入“预警”（黄色），由系统自动向维修班组推送工单，提示在下一个计划维护窗口检查、维护，做好相应准备。报警级：对于健康指数已经超限、故障模式置信度高或

已经发生突发性严重异常的情形，归入“报警”（红色），系统触发声光报警，向各级管理人员发送信息，要求紧急介入，提示立即评估是否需停机处理。实际应用中基于历史数据统计与机器学习动态调整各级阈值。

### 5.3 基于故障预测的维护决策优化

基于故障预测的决策优化可以超越实时诊断，全局最优调度维护活动。为实现该功能，可利用 LSTM、Transformer 等时序模型预测设备健康指数或关键性能指标的发展趋势，据此估算设备的剩余使用寿命或者获得衰退曲线。在此基础上，以“综合成本最小化”或“设备可用性最大化”为目标，综合考虑预测性维护成本、非计划停机损失、预防性维护固定成本以及生产计划约束建立决策优化模型，通过优化算法求解最佳维护时机与维护组合<sup>[5]</sup>。例如，实际应用中预测认为在 7 天后 A 烘丝机风机轴承将达到失效阈值，3 天后有既定的工艺停机窗口，

则此时优化模型会提示维修人员做出“优化行动”，“在 3 天后工艺停机时同步更换轴承”，此举减少了额外的专门停机次数。

### 5.4 动态维护策略库与知识库构建思路

为固化、传承专家经验，提高系统的持续学习能力，维护策略库与知识库需要动态化。构建时，通过专家经验初始化知识图谱和策略模板；同时，对于系统运行中产生的新案例，由维修人员验证、工程师审核并转化为新知识，更新入策略库与知识图谱，通过以上机制迭代优化系统。

## 6 结语

应用大数据分析可显著提升设备运维精准性、主动性，减少非计划停机。随着不断积累数据，迭代算法，故障预测模型必将获得更多技术的支持，不断优化，推动烟草工业设备运维更加智能化。

### 参考文献：

- [1] 代磊,吕晶,曹明辉.基于大数据分析的烟草机械设备状态感知与预警技术研究[J].电子元器件与信息技术,2025,9(1):103-106.
- [2] 钱周,付敏韬,荆健霖.烟草加工机械中的故障预测与维护技术研究[J].南方农机,2025,56(4):57-59,73.
- [3] 李家贵.烟机设备故障智能诊断与维护[J].设备管理与维修,2022(12):73-75.
- [4] 钟东怀.烟机设备故障诊断智能化流程及日常维护策略研究[J].中国设备工程,2023(20):178-180.
- [5] 孟祥策,王如璇.烟草机械设备故障诊断技术及其智能化发展[J].机械与电子控制工程,2025,7(4).