

大数据驱动的燃气输配系统故障预测与诊断方法研究

严仲武 (萍乡港华燃气有限公司, 江西 萍乡 337244)

摘要: 随着燃气输配系统的复杂性增加, 故障预测与诊断的准确性对保障系统安全至关重要。本文构建了适用于燃气输配系统故障预测与诊断的评估指标体系, 重点关注准确率、召回率、F1 值及精度等指标, 全面评估模型性能。通过混淆矩阵、交叉验证和敏感性分析等方法, 深入分析模型的优缺点, 为模型优化和改进提供依据。评估结果可用于调整模型结构、选择最适合的模型并设定预警阈值, 从而提高系统故障检测与响应能力, 确保燃气输配系统的安全稳定运行。

关键词: 燃气输配系统; 故障预测; 模型评估; 准确率; 召回率

0 引言

燃气输配系统的故障预测与诊断在保障系统安全运行中起着关键作用。随着系统规模和复杂性的增加, 传统的故障检测方法已难以满足实时性和准确性的要求。因此, 发展高效的预测与诊断模型, 并通过科学的评估指标体系对模型性能进行量化分析, 成为提升系统可靠性的重要途径。准确率、召回率、F1 值等指标是评估模型性能的核心, 能够反映模型在不同场景下的实际应用效果。

1 大数据驱动的燃气输配系统故障预测与诊断概述

大数据于燃气输配系统故障预测与诊断的运用, 对燃气供应安全性与稳定性意义非凡。燃气输配系统极为复杂, 包含众多环节与设备, 其运行期间产生的数据量庞大且多样。数据类型涉及压力、流量、温度、液位等参数, 压力数据通常以 1-5min 的间隔进行采集, 流量数据的采集频率相近且精度达立方米每小时量级。这些数据是故障预测与诊断工作的核心依据。故障预测依赖于对历史数据的深度挖掘与多源数据融合^[1]。从历史数据来看, 存在特定的参数变化规律, 在故障发生前的一定时间内, 压力波动幅度可达 0.1-0.2MPa, 流量异常变化比例在 5%-10% 之间, 这为故障预测提供了重要的量化线索。多源数据融合方面, 将燃气输配数据与气象等外部数据相结合, 能拓展故障预测的维度, 更精准地捕捉潜在故障风险。

故障诊断主要有基于规则和基于模型两种方式。基于规则的诊断依赖于长期积累的专家经验和专业领域知识构建的规则体系, 然而其应对复杂故障场景的能力存在局限性。基于模型的诊断则凭借预先构建的故障预测模型, 当模型输出的故障发生概率超过设定阈值时, 进一步综合多方面数据详细分析故障类型与位置。故障预测与诊断模型通过准确率、召回率、

F1 值等指标进行评估, 当前部分模型的准确率能达到 85%, 召回率达 75%, F1 值为 0.8, 这些数据反映了模型的性能现状, 也为后续的优化改进工作明确了目标与方向, 促使模型在燃气输配系统故障预测与诊断中发挥更高效能。

2 燃气输配系统数据采集与预处理现状

2.1 数据采集现状

压力、流量、温度和液位数据通过管道关键节点的传感器采集, 用于监测燃气输配系统的运行状态。压力传感器采集频率为 1-5min, 其数据准确性对于检测管道泄漏、堵塞或用气不均衡至关重要。流量计反映燃气输送量, 精度为立方米每小时级别, 直接影响分配与计费。温度数据有助于评估燃气状态及能量变化, 液位数据则对于储存环节至关重要。然而, 数据采集存在问题, 如传感器布局不系统、监测点稀疏、兼容性差、数据标准不统一, 以及长期运行中的性能衰减和环境影响, 导致数据可靠性降低, 影响系统的精准调控与安全保障。

2.2 数据噪声问题

从数据特征来看, 部分燃气企业压力数据的噪声比例可达 5% 左右。噪声成因复杂, 一方面环境中的电磁干扰会使传感器接收信号异常, 如周边电气设备运行产生的电磁场叠加于传感器信号上; 另一方面传感器内部元件的热噪声、零点漂移等自身缺陷也会引入噪声。噪声的存在严重扭曲了数据的真实性, 使原本稳定的压力、流量等数据曲线出现不规则波动。在故障预测与诊断流程中, 噪声数据会干扰基于数据模型的判断, 导致模型误判故障发生概率, 无法准确识别故障早期的微弱信号, 延误故障排查时机, 降低燃气输配系统运行的安全性与稳定性, 增加运维成本与资源浪费。

2.3 数据缺失状况

在流量数据方面,因网络通信故障、传感器临时故障或数据传输链路中断等因素,会出现数据缺失片段。数据缺失破坏了数据序列的完整性与连续性,使基于数据序列的分析方法失效。在故障预测中,缺失数据使模型无法完整学习数据规律,预测结果偏差增大,可能遗漏潜在故障预警;在故障诊断环节,由于关键数据缺失,无法准确对比正常与异常数据特征,难以精准定位故障类型与位置,阻碍及时有效的故障修复,对燃气输配系统的稳定供气、设备维护及安全运营构成显著威胁,影响整体运行效率与服务质量。

3 燃气输配系统故障特征提取方法

3.1 历史数据深度挖掘法

通过建立数据挖掘模型,分析过往压力、流量、温度等数据记录,识别潜在故障模式。针对压力数据,分析不同时间段、管道区域的压力波动规律,如峰值、谷值及波动频率等;流量数据则分析季节性变化、用气时段的流量波动及突变情况。利用聚类分析,将具有相似变化模式的数据归类,识别与故障相关的数据模式。例如,将长期低压且波动小、流量持续低于正常值的数据进行关联分析,揭示可能存在的管道堵塞或气源供应不足等故障特征。此方法为故障预测与诊断提供了数据支持,提高了系统故障识别的精准性。

3.2 多源数据关联分析法

此方法除了采集燃气输配系统内部的压力、流量等数据,还引入气象和地理信息等外部数据。结合气象数据分析气温变化对燃气需求的影响,气温骤降时需求增加,若此时压力急剧下降且流量上升,可能预示系统调节能力不足或潜在故障风险。地理信息数据用于分析管道地理位置、海拔和地质环境对压力和流量的影响,特别是经过不稳定区域时,压力波动可能与地质活动相关。通过建立多源数据关联模型(如多元线性回归或神经网络),挖掘隐藏的故障特征,提高故障预测与诊断的准确性和全面性,增强系统的安全保障^[2]。

3.3 动态数据实时监测法

通过流式数据处理技术,实时接收并处理压力、流量数据,计算移动平均值、标准差等统计特征,监测数据的稳定性。压力数据的移动平均窗口设为10分钟,若计算结果超出正常范围且标准差过大,表明可能存在管道泄漏或用气设备故障等异常。与此同时,实时监测流量变化率,当变化率异常增大或减小时,结合压力数据判断是否为阀门故障或管道破裂等问

题,及时识别流量异常。此方法能够捕捉故障发生的动态特征,为系统故障响应提供精准的数据支持,提升燃气输配系统的安全性与稳定性。

3.4 数据特征变换与降维法

该方法针对燃气输配系统中的高维数据,采用特征变换与降维处理技术,提取关键的故障特征。通过主成分分析(PCA)等算法,将多个压力、流量、温度等数据特征进行线性变换,降低数据维度。PCA算法根据数据在不同维度上的方差贡献率,选取方差较大的主成分,保留主要信息并减少冗余。处理压力数据时,PCA会突出反映出关键变化趋势,同时降低噪声干扰;流量数据也进行类似降维处理,以简化特征空间。在降维后的空间中,分析哪些低维特征组合在故障发生时呈现明显的变化模式,如主成分的数值比例关系变化,从而提取出与故障相关的特征。该方法提高了故障预测与诊断模型的计算效率和准确性,有助于提升燃气输配系统的安全性与稳定性。

4 燃气输配系统故障预测与诊断模型构建

4.1 基于支持向量机(SVM)的模型构建

支持向量机核心原理是通过构建一个超平面对数据进行分类或回归分析。对于燃气输配故障预测,设训练数据集为 $D=\{(x_1,y_1),(x_2,y_2),\dots,(x_n,y_n)\}$,其中 x_i 表示第 i 个样本的特征向量(如压力、流量等数据组成), y_i 表示对应的故障类别或故障程度。SVM的目标是找到一个最优超平面 $w^T x + b = 0$,使得两类数据间隔最大。通过引入松弛变量 ξ_i 来处理线性不可分情况,优化目标函数:

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i$$

其中 C 为惩罚参数,约束条件为 $y_i(w^T x_i + b) \geq 1 - \xi_i$, $\xi_i \geq 0$ 。利用拉格朗日乘子法求解得到决策函数:

$$f(x) = \text{sign}(\sum_{i=1}^n \alpha_i y_i K(x_i, x) + b)$$

其中 $K(x_i, x)$ 为核函数,常用的核函数有线性核、多项式核、高斯核等。在燃气输配系统中,通过对大量历史数据的训练,确定模型参数,从而根据输入的实时数据特征预测故障是否发生以及故障类型。

4.2 决策树模型构建

决策树模型构建基于对燃气输配数据的特征分析与划分。以燃气输配系统中的压力、流量等数据特征为节点构建树状结构。设 S 为训练样本集,首先计算

样本集 S 的信息熵:

$$H(S) = -\sum_{k=1}^{|y|} p_k \log_2 p_k$$

其中 P_k 是样本集 S 中属于第 $\backslash(k)$ 类的样本比例。然后选择信息增益最大的特征作为当前节点的分裂特征, 信息增益:

$$G(S, A) = H(S) - \sum_{v \in V(A)} \frac{|S^v|}{|S|} H(S^v)$$

其中 A 为特征, $V(A)$ 为特征 A 的取值集合, S^v 为特征 A 取值为 v 的样本子集。不断重复这个过程, 直到满足停止条件, 如节点中的样本数小于某个阈值或信息增益小于某个阈值等。在燃气输配故障诊断中, 从根节点开始, 根据输入数据的特征值沿着树的分支逐步判断, 最终到达叶子节点, 叶子节点对应的类别即为预测的故障类别, 根据压力是否异常、流量是否稳定等特征判断是管道泄漏、设备故障还是其他类型故障^[3]。

4.3 基于深度学习的模型构建

深度学习模型如长短期记忆网络 (LSTM) 在燃气输配故障预测中有独特优势。对于燃气输配系统的时间序列数据, 如连续的流量数据序列 $X=(x_1, x_2, \dots, x_T)$, LSTM 网络通过记忆单元和门控结构来处理序列中的长期依赖关系。LSTM 单元中的遗忘门 $f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$ 决定上一时刻记忆单元信息的保留程度, 输入门 $i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$ 控制当前输入信息的更新, 输出门 $o_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$ 确定记忆单元信息的输出。记忆单元 $c_t = f_t * c_{t-1} + i_t * \tanh(W_c[h_{t-1}, x_t] + b_c)$, 隐藏状态 $h_t = o_t * \tanh(c_t)$ 。通过对大量历史流量数据的训练, 网络学习到数据中的规律, 在预测时, 根据输入的前序流量数据序列, 预测后续流量数据的变化趋势, 若预测值与实际值偏差过大, 则可能预示着燃气输配系统存在故障风险, 从而实现故障的早期预警与预测。

5 燃气输配系统故障预测与诊断模型评估

5.1 评估指标体系构建

准确率是关键指标之一, 用于衡量模型预测正确结果在总预测结果中的比例, 反映模型整体的正确性。召回率则专注于模型正确预测出故障情况的能力, 尤其在故障检测场景中, 能体现对实际故障的捕捉程度。F1 值综合了准确率与召回率, 平衡两者关系, 更全面地评估模型性能。此外, 还可考虑精度指标, 衡量模型预测为正样本且实际也为正样本的比例。这些指标相互关联又各有侧重, 从不同角度对模型在燃气输配

故障预测与诊断中的表现进行量化评估, 为判断模型优劣提供依据。

5.2 模型性能分析方法

基于混淆矩阵的分析可直观呈现模型预测结果与实际情况的对应关系, 清晰展示真正例、假正例、真反例、假反例的数量分布, 从而深入了解模型在不同类别预测上的表现。交叉验证技术通过将数据集划分为训练集、验证集和测试集, 多次训练与验证模型, 有效避免过拟合或欠拟合现象对评估结果的干扰, 确保评估的可靠性与稳定性。同时, 可进行模型的敏感性分析, 研究模型对不同输入特征的敏感程度, 确定哪些燃气输配数据特征对模型预测结果影响较大, 以便针对性地优化数据采集与特征提取过程, 提升模型性能。

5.3 评估结果应用

依据评估结果可确定模型的改进方向, 若准确率较低, 可能需要调整模型结构或增加训练数据量; 若召回率不足, 则需优化故障特征提取方法以增强对故障的识别能力。在模型选择方面, 通过对比不同模型的评估指标, 选择最适合燃气输配系统的故障预测与诊断模型进行部署应用。此外, 评估结果还可用于设定模型预警阈值, 根据模型在不同阈值下的性能表现, 结合燃气输配系统对故障容忍度和安全性要求, 确定合适的预警触发条件, 保障燃气输配系统安全稳定运行。

6 结语

燃气输配系统故障预测与诊断模型的评估为提升系统安全性提供了重要依据。通过构建科学的评估指标体系, 结合多种分析方法, 可以全面评估模型的优缺点, 并为模型的优化和应用提供指导。准确率、召回率等指标的平衡, 能够有效提高故障检测的准确性和及时性, 确保燃气输配系统的稳定运行。未来的研究可以进一步深化模型的性能分析和优化, 推动智能化监测与预警系统的发展。

参考文献:

- [1] 李思颖. 基于多源异构数据的燃气输配场站综合预警技术研究 [D]. 北京邮电大学, 2024.
- [2] 陈伟. 燃气管网输配系统中调压器故障诊断和日负荷预测技术 [J]. 今日制造与升级, 2023, (10): 26-28.
- [3] 赵英新. 燃气工程中燃气输配技术探析 [J]. 石化技术, 2022, 29(06): 253-255.

作者简介:

严仲武 (1974-), 男, 本科, 现有职称: 中级, 研究方向: 燃气输配与储运。