

文章编号:0253-4339(2025)02-0145-10

doi:10.12465/j.issn.0253-4339.2025.02.145

基于关键特征的制冷剂泄漏故障软测量研究

凌敏彬¹ 杨钰婷¹ 韩华¹ 徐玲² 崔晓钰¹

(1 上海理工大学能源与动力工程学院 上海 200093; 2 开利空调冷冻研发管理(上海)有限公司 上海 200436)

摘要 针对制冷剂泄漏难以直接测量的问题,建立基于数据挖掘和关键特征的制冷剂泄漏故障软测量研究。通过随机森林重要性排序和距离相关系数对制冷剂泄漏故障的表征特征进行筛选,建立支持向量回归(SVR)软测量模型对泄漏进行定量测量。经一台额定制冷量为1440 kW、充注量为330 kg螺杆式冷水机组泄漏实验验证,基于3个表征特征建立的SVR软测量模型在测试集上的均方根误差(RMSE)和平均绝对误差(MAE)分别为0.844 kg和0.734 kg,软测量性能较其它3个特征子集显著提升。

关键词 制冷剂泄漏;特征选择;软测量;随机森林;支持向量回归

中图分类号:TB61⁺2; TB657

文献标识码: A

Soft Measurement of Refrigerant Leakage Based on Key Features

Ling Minbin¹ Yang Yuting¹ Han Hua¹ Xu Ling² Cui Xiaoyu¹

(1. School of Energy and Power Engineering, University of Shanghai for Science and Technology, Shanghai, 200093, China; 2. Carrier Air Conditioning & Refrigeration R&D Management (Shanghai) Co., Ltd., Shanghai, 200436, China)

Abstract Refrigerant leakage is a frequent and costly fault that deteriorates the normal operation of a chiller; however, it is difficult to measure directly. This study proposes a data mining- and key-feature-based approach for the soft measurement of refrigerant leakage. Random forest importance ranking and distance correlation coefficients were used to select the characteristic features, and a support vector regression (SVR) soft measurement model was established to measure leakage quantitatively. The proposed model was validated through a leakage experiment conducted on a screw chiller with a rated cooling capacity of 1440 kW and a refrigerant charge of 330 kg. The results showed that the SVR soft measurement model established on the three selected key features achieved significantly improved performance. The model had a root mean square error (RMSE) of 0.844 kg and a mean absolute error (MAE) of 0.734 kg, outperforming the other three feature subsets.

Keywords refrigerant leakage; feature selection; soft measurements; random forest; support vector regression

制冷剂泄漏故障是制冷系统中最常发生的故障之一^[1-2],占总维护成本的17%,排在制冷系统控制故障和风扇故障之后^[3]。此外,当泄漏量(充注量减少)为25%时,导致制冷系统的COP(性能系数,coefficient of performance)降低15%,制冷量降低20%^[4]。制冷剂泄漏故障不仅会直接影响机组的制冷效果,导致机组制冷量降低,制冷效果变差^[5],且泄漏的制冷剂会污染大气环境,增加能源消耗,造成温室效应^[6-7]。由于制冷剂在故障发生的初始阶段泄漏量较小,其作为全局故障的影响将随着制冷剂循环迅速蔓延至整个系统,因此可以认为是常见故障中最难以检测和最容易误诊的故障^[8-9]。因此,制冷剂泄漏的研究对于指导节能环保的维护干预具有重要意义,近

年来尤其是中国“双碳”目标后,受到广泛关注。

在判别制冷剂泄漏故障的研究中,涉及的特征参数种类繁多,但针对不同类型制冷系统的泄漏实验常以制冷剂充注量不足代替,通过数据拟合来获得参数与制冷剂充注量的关系,用其中灵敏度最高的参数来表征制冷剂泄漏。刘杰等^[10]在焓差实验台中研究了不同换热器下汽车空调系统的充注量情况,结果表明,制冷剂充注量不足和过充均会导致压缩机排气温度过高,造成压缩机内润滑油失效,对系统稳定运行产生不利影响。王海峰等^[11]基于COP最大原则研究了不同工况下多功能空调热水器的最佳充注量,结果表明,不同工况下所需制冷剂最佳充注量存在较大差异。Sun Shaobo等^[12]在变流量制冷系统中提出—

基金项目:国家自然科学基金(51506125)资助项目。(The project was supported by the National Natural Science Foundation of China (No. 51506125).)

收稿日期:2023-09-10;修回日期:2023-12-27;录用日期:2024-01-18

种独立成分分析-反向传播神经网络 (independent component analysis-back propagation neural network, ICA-BPNN) 的混合故障检测模型检测制冷剂充注量, 结果表明, 融合 ICA-BPNN 检测模型具有更好的检测性能。Liu Jianguan 等^[13] 将主元分析 (principal component analysis, PCA) 和指数加权移动平均 (exponentially-weighted moving average, EWMA) 方法相结合检测变流量制冷系统制冷剂充注量故障, 结果表明, 在故障程度较低时, PCA-EWMA 模型获得较 PCA (T^2 和 Q 统计量) 更好的故障检测效率。但通过简单统计分析某些特征参数来对比正常工况与泄漏工况, 所得结果存在极大的局限性, 并不能明确是否存在其他干扰工况造成类似结果^[14]。

随着机器学习技术的兴起, 数据驱动的故障检测和诊断策略已广泛应用于制冷剂泄漏故障^[15]。S. A. Tassou 等^[16] 一组 10 个人工神经网络 (artificial neural network, ANN) 作为预测模块来识别故障或无故障操作, 并为其生成一组残差来匹配泄漏和过充注条件。韩华等^[17] 利用主元分析 (PCA) 法提取制冷系统特征向量, 对典型人工智能方法所建故障诊断模型的性能进行理论与应用分析, 确定了以支持向量机 (support vector machine, SVM) 算法为基础的故障诊断模型, 综合讨论了包括制冷剂泄漏等多个冷水机组故障的诊断性能。王江宇等^[18] 针对多联机系统运行过程发生的制冷剂充注量故障, 结合主成分分析与决策树优点, 提出了基于主成分分析与决策树 (principal component analysis-decision tree, PCA-DT) 的制冷剂充注量故障诊断方法。A. Rai 等^[19] 应用高斯混合模型, 得到了用于监测管道状态的泄漏检测指标, 可有效区分非泄漏和泄漏状态。这些研究通过生成残差和最大似然概率作为泄漏指标, 定性检测、诊断制冷剂是否发生泄漏, 但对泄漏的定量研究很少。

本文针对冷水机组制冷剂泄漏量难以直接测量的问题, 提出利用软测量, 即通过建立计算原理之间的关系, 将无法直接测量的量与其他易于测量的量连接起来完成间接测量^[20]。常用的软测量方法有人工神经网络 (ANN)^[21]、模糊逻辑^[22]、支持向量回归 (support vector regression, SVR) 等, 其中 SVR 被认为是小样本统计估计和测量学习的最佳方法^[23]。G. Maksimović 等^[24] 对比研究了 SVR、ANN 在医疗支出预测问题, 结果表明 SVR 较 ANN 具有更好的预测准确性。Wang Wenchuan 等^[25] 在水文预报领域通过 SVM、ANN 和模糊逻辑等人工智能技术对水电站的月流量进行时间序列预测, 发现 SVM 在不同评价指

标下均能获得更好的预测精度。本文首先通过随机森林对故障运行参数进行重要性排序, 结合距离相关系数剔除强相关参数, 获得 3 个泄漏表征特征, 再基于 SVR 建立制冷剂泄漏软测量模型, 并对比分析了不同特征子集下软测量值与实测值的误差, 验证了所选表征特征的有效性。

1 制冷剂泄漏故障模拟实验

为解决制冷剂随着泄漏时间而带来的影响和变化, 选择一台带四管制热回收的双机头螺杆式冷水机组作为制冷剂泄漏故障实验的研究对象。图 1 所示为制冷剂泄漏实验的系统原理, 机组额定制冷量为 1 440 kW (409.5 冷吨), 额定功率为 270.8 kW, 制冷剂为 R134a, 额定充注量为 330 kg, 节流装置为电子膨胀阀 (electronic expansion valve, EEV)。为保证制冷剂泄漏实验安全进行, 泄漏点设在低温低压侧蒸发器出口处, 连接装有流量控制阀的软管, 制冷剂通过软管泄漏至放置在电子秤上的储液罐中, 电子秤精度为 ± 10 g, 泄漏量由控制阀控制。在制冷模式下运行, 制冷剂从额定充注量的 330 kg 均匀泄漏至 264 kg, 相当于额定充注量的 20%, 每隔 5 s 采集一组数据, 每一组数据由 25 个特征参数组成 (附录 A), 共采集 2.5 h, 获得 1 800 条时间序列的数据。

为消除不同特征变量在数据的量纲和量级产生巨大的差异, 如 EER 和冷凝器换热量这两个参数的单位不同, 取值范围相差 1 000 倍, 需对数据进行标准化消除量纲差异。采用 min-max 方法 (即离差标准化) 对数据进行线性压缩, 使结果落入 $[0, 1]$ 区间内。计算式如下:

$$x_{\text{normalization}} = \frac{x - x_{\text{Min}}}{x_{\text{Max}} - x_{\text{Min}}} \quad (1)$$

式中: x 为变量的值; x_{Min} 为该类变量中的最小值; x_{Max} 为该类变量中的最大值; $x_{\text{normalization}}$ 为归一化后的值。

2 基于随机森林与去相关性的故障特征筛查

2.1 基于随机森林的特征重要性排序

随机森林是 L. Breiman 于 2001 年提出的利用多棵决策树组合进行分类与回归的集成算法^[26]。在选取样本过程中, 采用有放回随机采样 (Bootstrapping), 包括训练数据的随机选取和待选特征的随机选取。随机森林原理如图 2 所示, 从总的原始训练数据 N 中, 采用 Bootstrapping 方法随机选取 z 个样本作为单棵决策树的训练样本。每一棵单独的决策树需要继

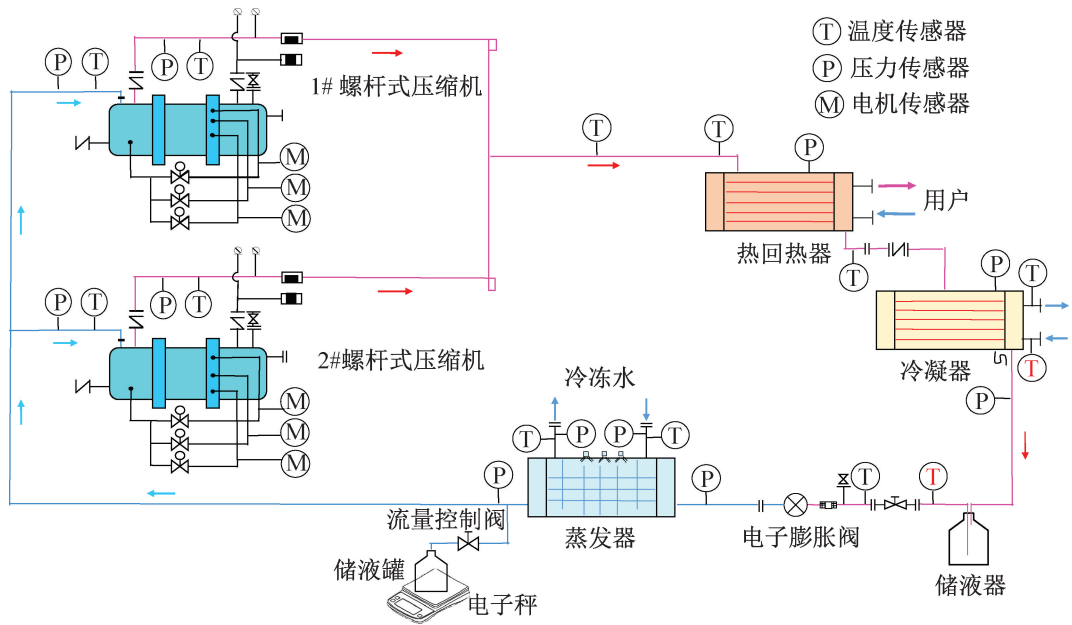


图1 制冷剂泄漏故障实验装置

Fig.1 Refrigerant leakage fault experimental setup

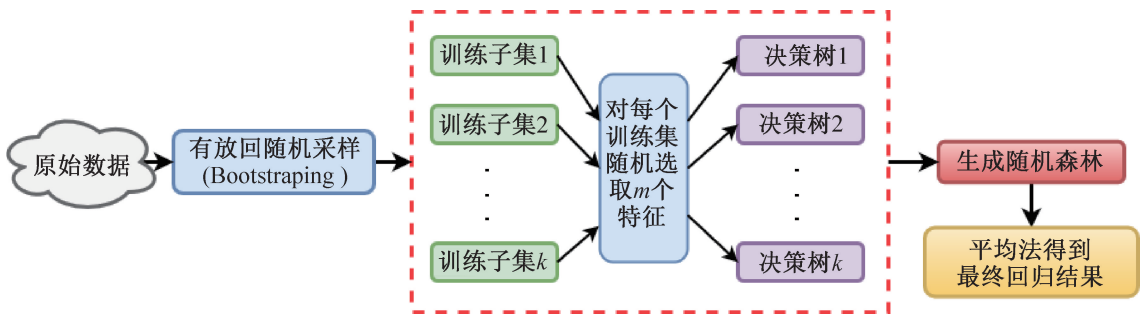


图2 随机森林原理

Fig.2 Schematic of random forest

续向下分裂,若每个样本有 M 个特征,则每个分裂节点处,都将在 M 个特征组成的特征空间里随机选择 $m(m \leq M)$ 个特征,通过计算每个特征的信息量选择具有分类能力的特征继续节点分裂,直到无法再生长为止。重复上述步骤 k 次,得到 k 棵决策树的 k 个训练样本集,最终的输出结果为 k 棵决策树结果的预测平均值。

随机森林通过对变量重要性进行度量(variable importance measure, VIM)实现变量的排名,评判不同特征变量对目标变量的影响程度大小。较常见的重要性度量指标为方差增量(increase in mean squared error, IncMSE)或节点纯度增量(increase in node purity, IncNodePurity)^[27]。IncMSE(式(2))主要是通过随机扰动的方法依次改变特征前后的顺序,并测试模型在扰动后特征上的效果。IncNodePurity(式(3))根据随机森林中决策树节点分裂时的不纯度变化作为衡量指标。

$$\Delta E = \frac{1}{k} \sum_{i=1}^k (E_i^n - E_i) \quad (2)$$

$$\Delta P = \frac{1}{k} \sum_{j \in N_i} (G_j I_j - G_l I_l - G_r I_r) \quad (3)$$

式中: ΔE 为方差增量; ΔP 为节点纯度增量; k 为构造的决策树数量; E_i^n 和 E_i 分别为对第 n 种统计参量扰动后的袋外数据和未添加扰动的袋外数据在第 i 棵决策树下的预测误差; G_j, G_l, G_r 分别代表该节点 j 及其左子节点和右子节点的不纯度; I_j, I_l, I_r 分别表示节点 j 及其左边节点和右边节点中训练样本数占总训练样本数的比重; N_i 表示按特征 i 分割的节点。

RF 变量选择模型训练过程如图3所示,将泄漏实验的1800组时间序列数据作为RF模型的输入,每一个数据包包含25个特征变量,基于Python系统构建随机森林变量选择模型。对RF模型中的2个重要参数:决策树的个数 $n_estimators$ 和最大节点个数 $max_features$ 进行网格搜索和五折交叉验证寻优。经

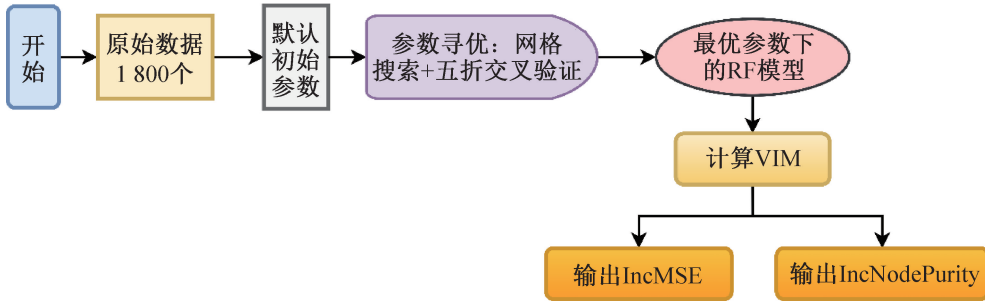


图3 RF 变量选择模型训练过程

Fig.3 RF variable selection model training process

过 1 877. 698 s 的寻优过程, 获得随机森林模型的最佳参数。寻优范围和最佳参数如表 1 所示。

表 1 模型参数优化结果

Tab.1 Model parameter optimization results

参数名称	寻优范围	步长	最佳参数
n_estimators	(0, 300)	3	25
max_features	(1, 8)	1	3

注: 硬件环境基于 intel (R) Core (TM) i7-8700 CPU@3. 20 GHz, 8G RAM 及 intel (R) UHD Graphics 630 显卡配置。

在此基础上, 分别用 2 个评价指标 (IncMSE 和 IncNodePurity) 计算各特征参数对制冷剂泄漏故障的重要性, 进行重要性排序, 如图 4 所示。排名第 11 以后的 IncMSE 权值系数 (图 4(a)) 均小于 0. 009, 甚至靠后的特征参数方差增量几乎为 0。在图 4(b) 中, 冷冻水进出水温度之间的节点纯度很接近, 约为 0. 006, 蒸发温度与液相温度 2 的节点纯度很相似, 约为 0. 039, 排名第 11 以后的 IncNodePurity 值均小于 0. 03。将 2 种排序方法的前 11 个特征列于表 2 中,

发现变量名称一致, 仅前后顺序有差别, 因此初步选用该 11 个变量。分析可知, 靠前的大多为物理参数与性能参数, 因重要性排序而剔除的变量大部分为信号参数。

2.2 制冷剂泄漏高耦合变量独立性筛查

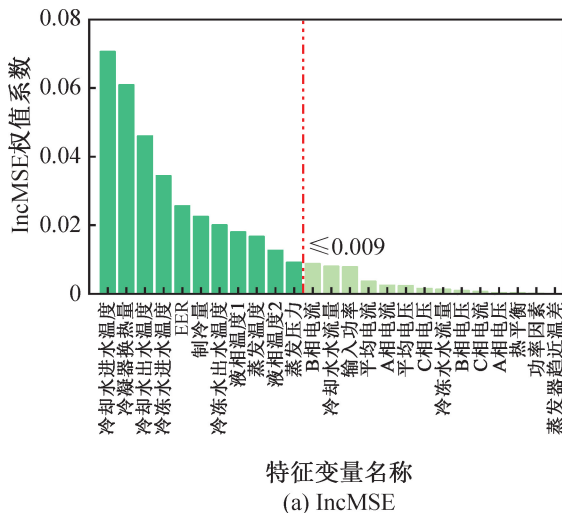
在上述基于随机森林重要度排序初选变量中, 考虑到变量与变量之间的耦合性, 选择距离相关系数作为衡量变量间相关性的指标, 进行相关性分析, 进一步剔除制冷剂泄漏故障的冗余信息。

对于 2 个随机向量 $x \in R^p, y \in R^q$, 记 $(x, y) = \{(x_i, y_i), i = 1, \dots, n\}$ 为观察到的随机样本, x, y 间的距离相关系数 (d_{Cor}) 可以定义为:

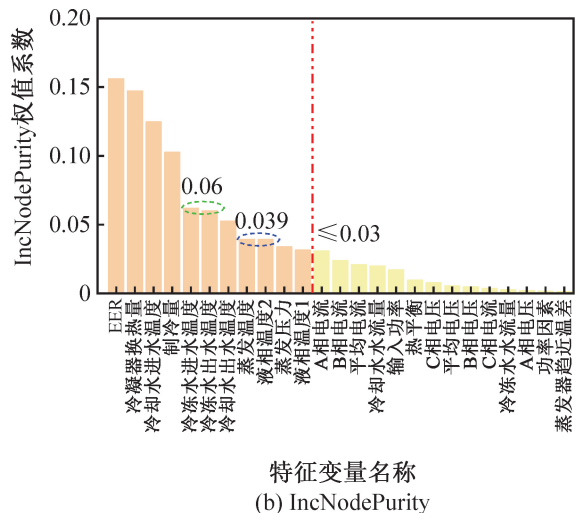
$$d_{Cor}(x, y) = \frac{v^2(x, y)}{\sqrt{v^2(x, x) \cdot v^2(y, y)}} \quad (4)$$

式中: $v(x, y)$ 为 x 和 y 的距离协方差; $v(x, x)$ 和 $v(y, y)$ 分别为 x 和 y 的距离方差。

图 5 所示为表 2 中 11 个参数经过两两变量之间的距离相关系数计算后的结果呈现, 对角线为各特征参数的名称, 黄色部分表示两变量间呈现较强的相关



特征变量名称
(a) IncMSE



特征变量名称
(b) IncNodePurity

图4 基于随机森林特征重要性排序

Fig.4 Feature importance ranking based on RF

表 2 基于 IncMSE 和 IncNodePurity 随机森林特征重要性排序
Tab.2 Feature importance ranking based on IncMSE and IncNodePurity RF

排名	IncMSE	权值系数	IncNodePurity	权值系数
1	冷却水进水温度	0.070 7	EER	0.156 1
2	冷凝器换热量	0.061 0	冷凝器换热量	0.147 2
3	冷却水出水温度	0.046 1	冷却水进水温度	0.124 8
4	冷冻水进水温度	0.034 5	制冷量	0.102 6
5	EER	0.025 7	冷冻水进水温度	0.061 9
6	制冷量	0.022 6	冷冻水出水温度	0.060 0
7	冷冻水出水温度	0.020 1	冷却水出水温度	0.052 5
8	液相温度 1	0.018 1	蒸发温度	0.039 3
9	蒸发温度	0.016 8	液相温度 2	0.039 2
10	液相温度 2	0.012 7	蒸发压力	0.033 9
11	蒸发压力	0.009 2	液相温度 1	0.031 5

性,紫色表示相关性较弱,第 1 行为各参数与制冷剂泄漏量之间的距离相关系数。在统计学中,相关系数大于 0.6 的属于强相关和极强相关的关系,因此,首先要保证变量和制冷剂泄漏量之间的强相关,即大于 0.6,如图 5 第 1 行中绿色的变量均小于 0.6,便不能被选择。同时,所选特征变量中两两之间的相关性要尽可能小,以减小信息冗余,若某 2 个变量之间相关系数过大,则删除其中贡献值靠后的那个,例如:冷凝器换热量与制冷量、EER 及冷冻水进水温度的距离相关系数高于 0.8,则选择与制冷剂泄漏量相关性最

强的冷凝器换热量作为表征特征之一,将其他 3 个均舍弃。

按照上述选择方法,最终选择冷凝器换热量、冷却水进水温度、液相温度 1(即储液器出口温度)作为制冷剂泄漏故障的关键特征。其中,冷凝器换热量与制冷剂泄漏量的距离相关系数为 0.820,冷却水进水温度与制冷剂泄漏量的距离相关系数为 0.711,液相温度 1 与制冷剂泄漏量的距离相关系数为 0.631。与传统制冷剂泄漏故障的定性诊断不同,软测量的关键参数集中在高压侧。

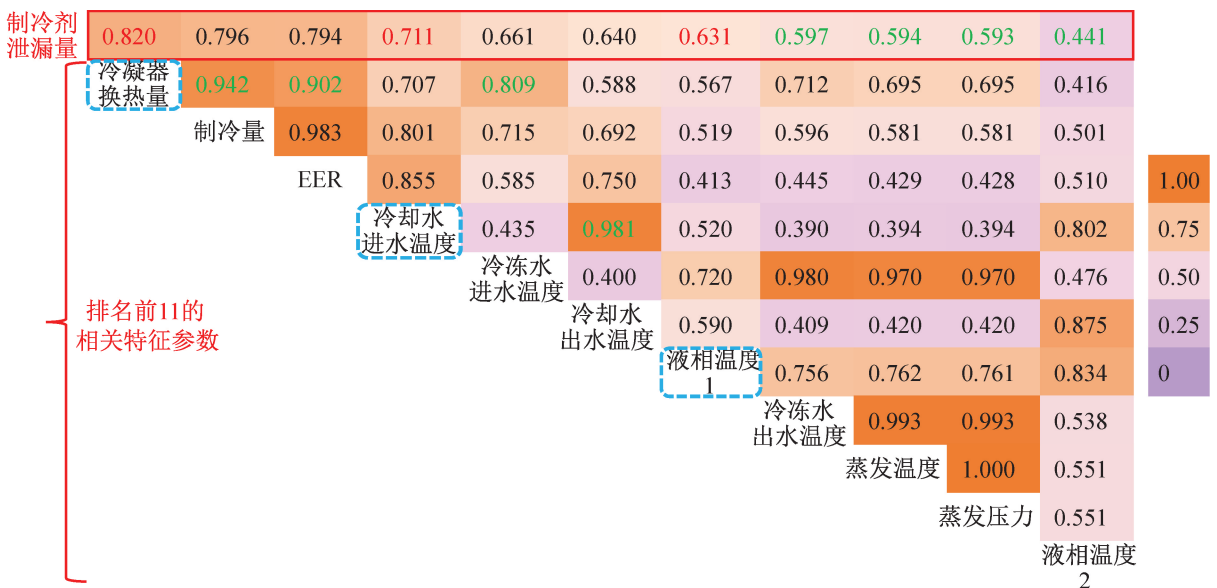


图 5 11 个特征参数之间的距离相关性系数
Fig.5 Distance correlation coefficient among 11 feature parameters

3 基于故障特征的制冷剂泄漏软测量

3.1 SVR 原理

支持向量回归(SVR)^[28]是一种经典的机器学习算法,尤其在小样本数据集的场景中被广泛应用。SVR 是寻找一个超平面,使得该超平面沿着纵轴方向上下平移 ε (误差间隔带)后,扫过的 ε -不敏感损失区域包含所有的样本点。具体为设置一个阈值 ε , 仅计算 $|f(x) - y| > \varepsilon$ 的数据点的误差损失。假设给定样本集 $D = \{(x_1, y_1), \dots, (x_n, y_n)\} \in (x, y)^n$, 其中, $x_i \in x = R^n, y_i \in R, i = 1, 2, \dots, n$, 并给定 $\varepsilon > 0$ 。超平面的表达式为:

$$f(x) = \omega \cdot \psi(x_i) + b \quad (5)$$

式中: ω 和 x 均为 n 维列向量; b 为偏置; $\psi(x_i)$ 表示引入的核函数,核函数 $\psi(x_i, x_j)$ 的表达式为:

$$\psi(x_i, x_j) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \quad (6)$$

式中: σ 为核宽度函数, σ 越小,核函数的宽度越小,越有选择性。

SVR 问题可以形式化为求解最优化问题,其中定义损失函数为:

$$l = \min_{\omega, b} \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n l_{\varepsilon}(f(x_i) - y_i) \quad (7)$$

式中: C 为惩罚参数; l_{ε} 为 ε -不敏感损失函数。

SVR 求解时通过引入拉格朗日乘子 a_i 来解决该约束最优化的问题:

$$\max \sum_{i=1}^n a_i y_i - \sum_{i=1}^n \sum_{j=1}^n a_i a_j y_i y_j \cdot \psi(x_i, x_j) \quad (8)$$

3.2 软测量结果与讨论

选取制冷剂随时间泄漏实验获得的 1 800 组数据,将数据的前 1 560 组(0~7 800 s)数据作为软测量模型的训练样本,剩余 240 组(7 805~9 000 s)数据作为测试数据。如表 3 所示,建立不同特征子集的软测量模型,分别为:基于上述两阶段选出的 3 个特

征参数组成的特征子集,用 S-3 表示, Yang Yuting 等^[29]研究得出的基于直接用皮尔逊相关性系数选取的 4 个参数(S_{su4})、RF 重要性排序前 11 的参数(S_{rank11})以及最原始的 25 个特征参数(S_{all25}),对制冷剂泄漏进行软测量研究,验证所选关键特征(S-3)的有效性。

表 3 不同特征子集

Tab.3 Different feature subsets

特征子集	特征描述
S-3	冷凝器换热量,冷却水进水温度,液相温度 1
S_{su4}	冷却水出水温度,蒸发器趋近温度 ^① ,液相温度 1,冷冻水出水温度
S_{rank11}	详见表 2
S_{all25}	详见附录 A

^①蒸发器趋近温度是冷冻水出水温度与蒸发温度的差值。

基于 Python 系统在上述 4 个特征子集中构建 SVR 制冷剂泄漏软测量模型,其训练过程如图 6 所示。参数设置对于机器学习方法的模型训练至关重要,SVR 选择的核函数类型为径向基函数(radial basis function, RBF), 2 个关键变量为惩罚系数 C 和核函数的宽度 σ ,采用网格搜索和五折交叉验证实现参数对寻优,寻优范围如表 4 所示。由不同特征子集构成的泄漏数据训练集分别输入模型,SVR 首先采用默认初始参数实施软测量,根据五折交叉验证结果计算均方根误差(root mean squared error, RMSE),按照设定的范围和步长进行 C 和 σ 的网格搜索,重新训练 SVR 并计算 RMSE,与前一次对比,保留 RMSE 较小的参数和模型,直到搜索范围结束, RMSE 达到最小,输出最佳参数(表 4)和训练好的 SVR 软测量模型;然后将相应的测试集输入模型进行测试,根据模型输出的制冷剂泄漏量(即软测量值)与实测泄漏量的差

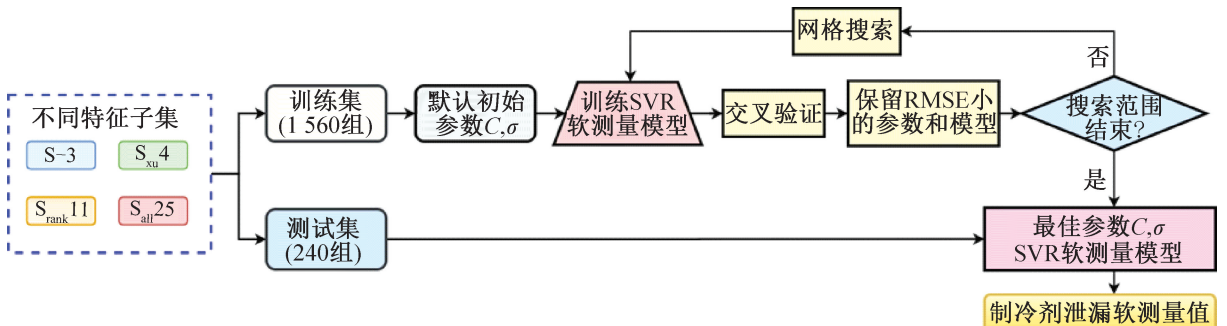


图 6 SVR 软测量模型训练过程

Fig.6 SVR soft measurement model training process

异(图 7),对模型进行评价。评价指标为 RMSE 和平均绝对误差(mean absolute error, MAE),如式(9)、式(10)所示。

$$Y_{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_s)^2} \quad (9)$$

$$Y_{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - y_s| \quad (10)$$

式中: y_i 为实验测得的泄漏量; y_s 为 SVR 软测量模型得到的制冷剂泄漏软测量值; n 为样本数。

表 4 不同特征子集下的 SVR 模型最优参数

Tab.4 Optimal parameters of SVR model with different subsets of features

特征子集	参数寻优范围		最佳参数	
	C	σ	C	σ
S-3	$(10^2, 10^6)$	(0.24, 0.25)	10^4	0.246
$S_{xu}4$		(0.34, 0.35)		0.341
$S_{rank}11$		(0.19, 0.20)	10^5	0.192
$S_{all}25$		(0.01, 0.02)		0.013

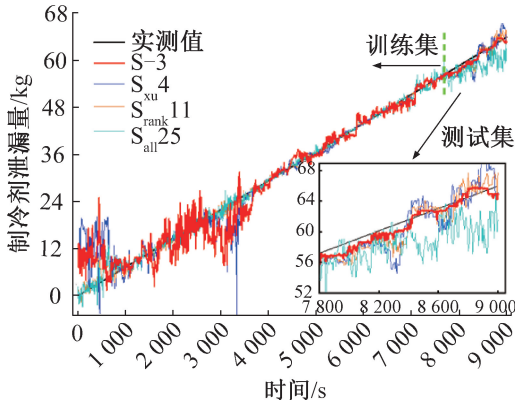


图 7 不同特征子集下的 SVR 软测量结果

Fig.7 SVR soft measurements with different subsets of features

图 7 中前 7 800 s 为模型训练阶段,4 个模型在 4 000 s 之前的波动均很大,由于制冷剂泄漏故障发生时触发系统的自我调节,出现参数回升,各软测量曲线围绕制冷剂实测值上下波动。随着模型训练的继续,软测量结果越来越接近实测值曲线,说明模型在不断地从数据中学习规律和知识。其中,特征参数较多的 $S_{rank}11$ 和 $S_{all}25$ 模型在前 4 000 s 的波动较小,S-3 和 $S_{xu}4$ 两个模型则起伏较大,表明参数多时初期跟踪性更强。

图 7 中将 7 800~9 000 s 的测试阶段放大,其结果可以反映软测量模型的测试效果。整体来看 4 个

模型均能输出与实际泄漏量相似的泄漏趋势,S-3 模型相对于其他 3 个模型效果更佳,软测量值与实测值曲线较为接近,且波动较小。 $S_{xu}4$ 模型围绕实测值曲线上波动范围较大, $S_{rank}11$ 模型在最后时间内的软测量较好, $S_{all}25$ 软测量模型整体在实测值曲线之下,虽然趋势相同,但偏差较大,测试结果不佳。

图 8 所示为软测量结果的 RMSE 与 MAE,误差值越小表示与实测值越接近,性能越好。S-3 和 $S_{xu}4$ 在训练过程中的 RMSE 与 MAE 均较大,相当于 $S_{rank}11$ 和 $S_{all}25$ 的 3 倍,原因在于训练初期软测量值波动较大。但在测试阶段,S-3 呈现出最佳性能, RMSE 和 MAE 分别仅为 0.844 kg 和 0.734 kg,其次是 $S_{rank}11$,误差值约为 S-3 的 2 倍; $S_{xu}4$ 次之,最差的是 $S_{all}25$ 模型,误差值约为 S-3 的 5 倍。由此分析,由冷凝器换热量、冷却水进水温度、液相温度 1 这 3 个关键变量建立的 S-3 软测量模型外推性最佳,测试集 RMSE 分别较 $S_{xu}4$ 、 $S_{rank}11$ 和 $S_{all}25$ 下降 61.5%、48.9%、71.4%, MAE 分别下降 59.2%、48.1%、75.1%。

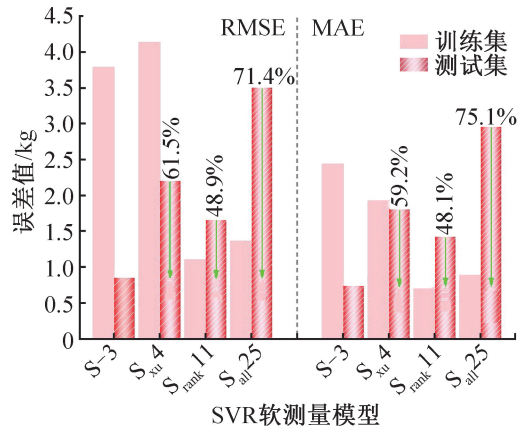


图 8 测试集不同特征子集下的 SVR 软测量结果

Fig.8 SVR soft measurements under different feature subsets of the test set

针对训练时的波动,S-3 软测量模型能很好地调整进而呈现出越来越好的效果,模型的学习能力与泛化能力相对其他 3 个模型最好,根本原因是上述 3 个关键参数的选择,能够很好地捕捉到制冷剂泄漏且与泄漏量强相关,模型能高效且低噪声地提取与泄漏相关的重要信息。 $S_{xu}4$ 模型和 S-3 有同样的关键变量——液相温度 1,其中冷却水出水温度与冷冻水出水温度在 RF 重要性排序和距离相关性系数中排在前列,但蒸发器趋近温度在本文研究中处于后 11 名中,因此与 S-3 相比, $S_{xu}4$ 引入了一个相关度较差的变量,使其性能相对较差。在 RF 重要性排序中,若前面的参数重要性较大,则后续与之相关的参数可能

对模型的重要性减弱。 $S_{\text{rank}11}$ 和 $S_{\text{all}25}$ 均包含 S-3 中所有的参数变量,其中 $S_{\text{rank}11}$ 是经过 RF 重要性排在前 11 的特征参数子集,且一定程度上削减了参数相关性,因此呈现出比 $S_{\text{su}4}$ 与 $S_{\text{all}25}$ 更好的性能。 $S_{\text{all}25}$ 包括所有参数,在训练过程呈现出较好的性能,但由于存在大量冗余信息,在测试中干扰软测量结果,性能较差。

4 结论

本文针对螺杆式冷水机组的制冷剂泄漏量难以直接测量的问题,以一台双机头螺杆式冷水机组作为研究对象,开展泄漏量软测量研究。基于随机森林重要性排序剔除重要度较低特征,并结合距离相关性系数进行独立性筛查,如此经复合层级甄选获得制冷剂泄漏故障关键特征;然后基于 SVR 建立制冷剂泄漏软测量模型,对比分析不同特征子集下,模型软测量值与泄漏量实测值的误差,验证了所选关键特征及软测量模型的有效性。得到如下结论:

1) 采用 RF 算法和距离相关系数,筛选出的对制冷剂泄漏故障较为重要且独立性较强的 3 个关键特征为:冷凝器换热量、冷却水进水温度、液相温度 1 (储液器出口温度)。与传统制冷剂泄漏故障的定性诊断不同,软测量的关键参数集中在高压侧。

2) 基于 3 个关键特征(S-3)建立的 SVR 制冷剂泄漏软测量模型,与基于皮尔逊相关系数得到的 $S_{\text{su}4}$ 、RF 重要性排名前 11 ($S_{\text{rank}11}$) 及原始特征参数 ($S_{\text{all}25}$) 建立的 SVR 软测量模型相比,虽然在训练过程前期波动较大,进而使训练集上的均方根误差 (RMSE) 和平均绝对误差 (MAE) 较大,却在测试阶段呈现出最佳性能, RMSE 分别比 $S_{\text{su}4}$ 、 $S_{\text{rank}11}$ 和 $S_{\text{all}25}$ 下降 61.5%、48.9%、71.4%, MAE 分别下降 59.2%、48.1%、75.1%,具有良好的学习和泛化能力。

实际应用中,若缺乏流量传感器,则 S-3 中的冷凝器换热量无法获得,此时可退而采用 $S_{\text{su}4}$ 中的 4 个温度参数进行制冷剂泄漏软测量。

参考文献

[1] BEHFAR A, YUILL D, YU Yuebin. Supermarket system characteristics and operating faults (RP - 1615) [J]. Science and Technology for the Built Environment, 2018, 24(10): 1104-1113.

[2] ZHANG Yun, LIU Cichong, WANG Tianying, et al. Leakage analysis and concentration distribution of flammable refrigerant R290 in the automobile air conditioner system [J]. International Journal of Refrigeration, 2020, 110: 286-294.

[3] MADANI H, ROCCATELLO E. A comprehensive study on the important faults in heat pump system during the warranty period[J]. International Journal of Refrigeration, 2014, 48: 19-25.

[4] KIM W, BRAUN J E. Evaluation of the impacts of refrigerant charge on air conditioner and heat pump performance [J]. International Journal of Refrigeration, 2012, 35(7): 1805-1814.

[5] 钱长华, 于林. 空调系统中制冷剂泄漏浅析[J]. 暖通空调, 2008, 38(3): 122-124. (QIAN Changhua, YU Lin. Refrigerant leakage of air conditioning systems[J]. Journal of HV & AC, 2008, 38(3): 122-124.)

[6] CALM J M. Emissions and environmental impacts from air-conditioning and refrigeration systems [J]. International Journal of Refrigeration, 2002, 25(3): 293-305.

[7] MCCULLOCH A, MIDGLEY P M, ASHFORD P. Releases of refrigerant gases (CFC - 12, HCFC - 22 and HFC - 134a) to the atmosphere [J]. Atmospheric Environment, 2003, 37(7): 889-902.

[8] YAO Wanli, LI Donghui, GAO Long. Fault detection and diagnosis using tree-based ensemble learning methods and multivariate control charts for centrifugal chillers [J]. Journal of Building Engineering, 2022, 51: 104243.

[9] HAN Hua, ZHANG Zhan, CUI Xiaoyu, et al. Ensemble learning with member optimization for fault diagnosis of a building energy system[J]. Energy and Buildings, 2020, 226: 110351.

[10] 刘杰, 赵宇, 祁照岗, 等. 制冷剂充注量对新型换热器汽车空调的影响[J]. 制冷学报, 2011, 32(1): 12-15. (LIU Jie, ZHAO Yu, QI Zhaogang, et al. Impact of refrigerant charge on mobile air conditioning system with new heat exchanger [J]. Journal of Refrigeration, 2011, 32(1): 12-15.)

[11] 王海峰, 张守兵, 原惠惠, 等. 多功能空调热水器最佳制冷剂充注量的实验与分析[J]. 制冷学报, 2015, 36(3): 87-91. (WANG Haifeng, ZHANG Shoubing, YUAN Huihui, et al. Experimental study on optimal refrigerant charge of multi-functional air-conditioning hot water heater [J]. Journal of Refrigeration, 2015, 36(3): 87-91.)

[12] SUN Shaobo, LI Guannan, CHEN Huanxin, et al. A hybrid ICA-BPNN-based FDD strategy for refrigerant charge faults in variable refrigerant flow system [J]. Applied Thermal Engineering, 2017, 127: 718-728.

[13] LIU Jiangyan, HU Yunpeng, CHEN Huanxin, et al. A refrigerant charge fault detection method for variable refrigerant flow (VRF) air-conditioning systems [J]. Applied Thermal Engineering, 2016, 107: 284-293.

[14] 于仙毅, 巫江虹, 高云辉. 基于主成分分析与支持向量

- 机的热泵系统制冷剂泄漏识别研究[J]. 化工学报, 2020, 71(7): 3151-3164. (YU Xianyi, WU Jianghong, GAO Yunhui. Research on refrigerant leakage identification for heat pump system based on PCA-SVM models[J]. *CIESC Journal*, 2020, 71(7): 3151-3164.)
- [15] FAN Yuqiang, CUI Xiaoyu, HAN Hua, et al. Feasibility and improvement of fault detection and diagnosis based on factory-installed sensors for chillers[J]. *Applied Thermal Engineering*, 2020, 164: 114506.
- [16] TASSOU S A, GRACE I N. Fault diagnosis and refrigerant leak detection in vapour compression refrigeration systems [J]. *International Journal of Refrigeration*, 2005, 28(5): 680-688.
- [17] 韩华, 谷波, 任能. 基于主元分析与支持向量机的制冷系统故障诊断方法[J]. 上海交通大学学报, 2011, 45(9): 1355-1361. (HAN Hua, GU Bo, REN Neng. Fault diagnosis for refrigeration systems based on principal component analysis and support vector machine [J]. *Journal of Shanghai Jiao Tong University*, 2011, 45(9): 1355-1361.)
- [18] 王江宇, 陈焕新, 刘江岩, 等. 基于PCA-DT的多联机制冷剂充注量故障诊断[J]. 华中科技大学学报(自然科学版), 2016, 44(7): 1-4. (WANG Jiangyu, CHEN Huanxin, LIU Jiangyan, et al. PCA-DT based refrigerant charge fault diagnosis for variable refrigerant flow system [J]. *Journal of Huazhong University of Science and Technology (Natural Science Edition)*, 2016, 44(7): 1-4.)
- [19] RAI A, KIM J M. A novel pipeline leak detection approach independent of prior failure information [J]. *Measurement*, 2021, 167: 108284.
- [20] 俞金寿. 软测量技术及其应用[J]. 自动化仪表, 2008, 29(1): 1-7. (YU Jinshou. Soft sensing technology and its application [J]. *Process Automation Instrumentation*, 2008, 29(1): 1-7.)
- [21] HECHT-NIELSEN R. Theory of the backpropagation neural network [M]//*Neural Networks for Perception*. US: Academic Press, 1992: 65-93.
- [22] POIRIER P J, MEECH J A. Using fuzzy logic for on-line trend analysis [C]//*Proceedings of IEEE International Conference on Control and Applications*. Vancouver, BC, Canada: IEEE, 1993: 83-86.
- [23] LÁZARO M, SANTAMARÍA I, PÉREZ-CRUZ F, et al. Support Vector Regression for the simultaneous learning of a multivariate function and its derivatives [J]. *Neurocomputing*, 2005, 69(1-3): 42-61.
- [24] MAKSIMOVIĆ G, JOVIĆ S, JOVANOVIĆ R, et al. Management of health care expenditure by soft computing methodology[J]. *Physica A: Statistical Mechanics and Its Applications*, 2017, 465: 370-373.
- [25] WANG Wenchuan, CHAU K W, CHENG Chuntian, et al. A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series[J]. *Journal of Hydrology*, 2009, 374(3/4): 294-306.
- [26] BREIMAN L. Random forests [J]. *Machine Learning*, 2001, 45: 5-32.
- [27] ECHEVERRY-GALVIS M A, PETERSON J K, SULO-CACERES R. The social network: tree structure determines nest placement in Kenyan weaverbird colonies [J]. *PLoS ONE*, 2014, 9(2): e88761.
- [28] STEINWART I, CHRISTMANN A. Support vector machines[M]. New York: Springer, 2008.
- [29] YANG Yuting, XU Ling, HAN Hua, et al. Soft measurement and prediction of refrigerant leakage based on SVR-LSTM [J]. *International Journal of Refrigeration*, 2023, 152: 303-314.

通信作者简介

韩华,女,副教授,上海理工大学能源与动力工程学院,13611880360,E-mail:happier_han@126.com。研究方向:制冷空调系统的故障诊断及优化,AI在制冷系统中的应用,新型制冷方式。

About the corresponding author

Han Hua, female, associate professor, School of Energy and Power Engineering, University of Shanghai for Science and Technology, 86-13611880360, E-mail:happier_han@126.com. Research fields: fault diagnosis and optimization of refrigeration and air conditioning system, application of AI in refrigeration system, new refrigeration methods.

附录

表 A 基于时间序列制冷剂泄漏故障实验的 25 个特征参数

序号	特征参数	单位	序号	特征参数	单位
1	冷冻水进水温度	℃	14	EER	
2	冷冻水出水温度	℃	15	热平衡	%
3	冷却水进水温度	℃	16	A 相电压	V
4	冷却水出水温度	℃	17	B 相电压	V
5	液相温度 1	℃	18	C 相电压	V
6	液相温度 2	℃	19	平均电压	V
7	蒸发温度	℃	20	A 相电流	A
8	蒸发器趋近温差	℃	21	B 相电流	A
9	蒸发压力	kPa	22	C 相电流	A
10	冷冻水流量	m ³ /h	23	平均电流	A
11	冷却水流量	m ³ /h	24	输入功率	kW
12	制冷量	kW	25	功率因数	
13	冷凝器换热量	kW			