

# 基于 PCA-DT 的多联机制冷剂充注量故障诊断

王江宇 陈焕新 刘江岩 李冠男

(华中科技大学能源与动力工程学院, 湖北 武汉 430074)

**摘要** 针对多联机制冷剂充注量故障,提出了一种基于主成分分析-决策树(PCA-DT)算法的制冷剂充注量故障检测与诊断方法.该方法先通过数据预处理进行数据清理,然后利用 PCA 对原始数据做降维处理,最后将新的数据划分为训练集与测试集,以训练集建立决策树模型进行故障检测与诊断.实验数据测试集的检测与诊断结果验证了该模型的可行性.采集了某多联机在 3 个地区的实际运行数据,对建立的决策树模型进行进一步的验证.结果表明,该方法对于多联机的制冷剂充注量故障有良好的检测与诊断效果,且检测与诊断效果整体上要优于 DT 算法.

**关键词** 多联机;故障;检测与诊断;主成分分析;决策树;制冷剂;充注量

中图分类号 TB65 文献标志码 A 文章编号 1671-4512(2016)07-0001-04

## PCA-DT based refrigerant charge fault diagnosis for variable refrigerant flow system

Wang Jiangyu Chen Huanxin Liu Jiangyan Li Guannan

(School of Energy and Power Engineering, Huazhong University of Science and Technology, Wuhan 430074, China)

**Abstract** In order to solve the problem that variable refrigerant flow refrigerant charge breaks down, a fault diagnosis method based on principal component analysis-decision tree (PCA-DT) algorithm was put forward. The raw data was preprocessed at first. Then the PCA method was employed to reduce the dimension of the data. Further, the data was divided into two sets, i. e. training set and test set. The decision tree model was developed based on training set to achieve fault detection and diagnosis. Through detection and diagnosis result, the practicability of the decision trees model was confirmed. Other further test was taken out to check its feasibility, and practical running data from 3 typical regions was collected. The results show that the proposed method performs better at detection and diagnosis to VRF refrigerant charge problem compared to the decision tree model.

**Key words** variable refrigerant flow; fault; detection and diagnosis; principal component analysis; decision trees; refrigerant; charge

多联机一般应用在公寓、商场、写字楼等场所调节环境,一旦出现故障会导致舒适性下降,经济损失,甚至安全问题,要求维修人员观察机组数据与物理特征,结合自身的经验来判断机组故障的根源,因此维修人员必须具备丰富的专家知识以及对多联机结构与运行的深入了解.不需要专家

知识的数据驱动的故障检测与诊断算法对诊断人员的专业性要求较低,而且诊断效率更高.目前,制冷系统故障的检测与诊断方法<sup>[1]</sup>主要可分为基于知识的方法、基于模型的方法和基于信号处理的方法.基于知识的方法包含有专家系统<sup>[2]</sup>、神经网络<sup>[3]</sup>、模糊推理<sup>[4]</sup>等,须建立在对制冷系统有较

收稿日期 2015-12-16.

作者简介 王江宇(1993-),男,博士研究生;陈焕新(通信作者),教授, E-mail: chenhuanxin@tsinghua.org.cn.

基金项目 国家自然科学基金资助项目(51576074, 51328602).

为深入了解的基础上,基于模型的方法有状态估计法、参数估计法等,稳定性较差,对于非线性的系统地处理有所欠缺.基于信号处理的方法无须建立系统数学模型,有主元分析法、小波变换等.

数据挖掘算法从大量的数据中发现知识,获取自身想要的信息,不需要先验知识.决策树算法是应用得最为广泛的数据挖掘算法之一,具有计算速度快、准确率高的优点,在金融<sup>[5]</sup>、计算机技术<sup>[6]</sup>、医学<sup>[7]</sup>等领域都有应用.多联机结构复杂、传感器多、测算的数据量大,可以尝试使用决策树算法进行故障诊断.本研究提出了一种基于主成分分析-决策树(PCA-DT)多联机制冷充注量故障检测与诊断方法,利用充注量实验数据进行决策树模型建模,并使用三组机组实际运行的数据验证,结果证明该方法有良好的检测与诊断效果.

### 1 PCA-DT 方法

PCA-DT 方法的基本思路是:先根据正常的制冷剂充注量水平实验数据建立主元模型;然后利用主元模型对非正常的制冷剂充注量水平实验数据(故障数据)进行处理,得到降维后的故障数据;最后利用决策树模型对经过主元模型降维后的网络数据进行故障检测与诊断.

#### 1.1 PCA 算法

假设样本观测数据的历史数据集为  $X(X \in \mathbf{R}^{n \times p})$ , 每一列代表一个变量,每一行代表一个观测值.  $X$  可以写作  $X = TP^T$ , 式中:  $T \in \mathbf{R}^{n \times k}$  为主元向量;  $P \in \mathbf{R}^{p \times p}$  为负载向量. 选取包含信息量较大的前  $k$  个主元, 得出主元矩阵  $T_k$ , 相应的负载矩阵  $P_k$ , 以及  $p-k$  个主元构成的主元矩阵  $\tilde{T}$ , 相应的负载矩阵  $\tilde{P}$ , 于是有

$$\begin{aligned} \tilde{X} &= T_k P_k^T = \sum_{i=1}^k t_i p_i^T; \\ E &= \tilde{T} \tilde{P}^T = \sum_{i=k+1}^p t_i p_i^T, \end{aligned} \tag{1}$$

式中:  $\tilde{X}$  为对于原始数据集  $X_p$  的估计;  $E$  为残差空间;  $t_i$  为  $T$  的第  $i$  列;  $p_i$  为  $P$  的第  $i$  列.

原始数据集可表示为  $X = \tilde{X} + E$ . 可以看出:  $X$  可分为投影到主元空间和残差空间的信息, 利用  $\tilde{X}$  可以很好地估计原始数据  $X$ . 值得注意的是, 在本研究的数据中, 建立 PCA 模型须采用正常充注量水平的数据. 对于实验数据中非正常水平制冷剂充注量数据以及网络数据均要经过模型处理, 达到数据降维的目的. 记新的数据为  $X_{new}$ , 样本主元值为  $T_{new}$ , 新的负载向量为  $P_{new}$ , 由式

(1)可知  $X_{new}$  的估计值为

$$\tilde{X}_{new} = P_{new} T_{new} = P_{new} P_{new}^T X_{new}.$$

处理后得到的估计值  $\tilde{X}_{new}$  作为下一步决策树模型建模的数据来源.

#### 1.2 DT 算法

分类回归树(classification and regression tree, CART)算法是最普遍的 DT 算法之一. CART 算法采用自顶向下的方式建立一棵二叉树, 在每一个分支节点上进行属性的比较判断, 其中应用最广的节点选择原则是 GINI 不纯度尽量低<sup>[8]</sup>. 节点  $A$  的 GINI 不纯度定义为

$$\alpha_{GINI}(A) = 1 - \sum_{m=1}^C p_m^2,$$

式中:  $C$  为属性中因变量的数目;  $p_m$  为观测点中属于第  $m$  类的概率. 建树的终止条件包含两种: 叶节点中的样本数为 1 或者被划分为同类; 决策树的高度达到了设定值.

### 2 基于 PCA-DT 的故障检测与诊断

基于 PCA-DT 算法进行多联机制冷剂充注量故障诊断方法包括数据预处理、训练建模与故障检测三个部分. 数据预处理通过剔除异常值、死值等影响故障检测效率的值, 并选取所需要的特征变量, 达到提高实验数据质量, 降低数据维数的目的<sup>[9]</sup>. 训练建模通过 PCA 进行实验数据的进一步处理, 选取数据的主元空间, 并抽取其中的部分数据作为决策树的训练集输入, 建立决策树模型. 故障检测利用训练建模所得到的决策树模型, 对实验数据及网络数据进行故障的检测与诊断.

#### 2.1 数据来源

实验数据来自某多联机制冷充注量实验, 多联机的结构图参看文献[10]. 该多联机由 1 台室外机和 5 台室内机构成.

如表 1 所示, 引入 3 种不同的制冷剂充注程度, 即不足、适中及过量, 共 9 种充注量水平. 表 2

表 1 制冷剂充注量水平

编号	充注量水平/%	充注程度
1	65	不足
2	75	不足
3	80	不足
4	85	适中
5	95	适中
6	105	适中
7	110	适中
8	120	过量
9	130	过量

2 给出了多联机实验工况与实际运行工况,其中网络数据为多联机在实际运行过程中所采集的数据,实验与实际运行的多联机型号相同。

表 2 多联机运行实验工况与实际运行工况

数据来源	多联机编号	模式	室外环境温度/℃
实验数据	VRF0	制冷	31/35/40
		制热	-7/2/7
网络数据	VRF1	制冷	22~35
	VRF2	制冷	22~33
	VRF3	制冷	22~38

### 2.2 数据预处理

多联机实际的运行数据采样间隔短、时间长、测量点较多,数据量庞大.庞大的数据一方面存在各种死值、缺失值、高度线性相关的变量及很多对于故障检测贡献较低的量,会导致后期故障检测与诊断的准确率降低;另一方面会导致算法运行时间周期长、效率低下.实际上,检测与诊断结果的准确性很大程度上依赖数据的质量,而由于不可靠的测量和动态过程的存在,数据质量较低,因此数据预处理环节非常重要。

数据预处理包含数据清洗和特征变量选取.过程如下:a.剔除数据中的死值变量(该变量的观测值保持为常数,不随着时间发生变化);b.剔除异常值,如温度、压力等变量中超出正常范围的观测值;c.选取室外环境温度、冷凝温度、蒸发温度、压缩机进气温度、压缩机排气温度、压缩机壳顶温度、压缩机运行频率、压缩机电流、化霜温度、过冷器气体出口温度、过冷器液体出口温度、气液分离器进口温度、气液分离器出口温度、室外机膨胀阀开度 14 个变量进行后续的建模工作;d.添加数据标签,将 9 种不同充注量水平的数据划分为 3 个水平并分别添加充注程度不足、适中和过量;e.由于数据之间有量纲的差异会对后续的故障检测造成影响,因此对正常充注量与故障充注量的数据集分别采用标准化来消除该差异,数学模型为  $X_i^* = [X_i - E(X_i)] / \sqrt{\text{var}(X_i)}$  ( $i=1, 2, \dots, n$ ),式中  $X_i$  为数据集; $E(X_i)$  为数据集的期望; $\text{var}(X_i)$  为  $X_i$  的方差。

### 2.3 训练建模

训练建模过程如下:a.选取标准化后的多联机正常充注量水平下的实验数据,求取其负荷向量矩阵,并建立 PCA 模型;b.对于经过 PCA 处理的多联机数据,在其正常数据与故障数据中分别进行随机抽样,选取 3/4 的数据汇总后作为之后决策树算法的训练集;c.利用决策树算法对训练集建立决策树模型。

### 2.4 故障检测与诊断

利用决策树模型对实验数据和网络数据进行故障检测,得出实验数据与网络数据的检测率,即诊断正确率  $R = (N_{\text{pre}}/N) \times 100\%$ ,式中: $N_{\text{pre}}$  为模型正确检测出故障的样本个数; $N$  为样本总数.图 1 为基于 PCA-DT 的多联机制冷剂充注量故障诊断流程图,包含数据预处理及故障诊断。

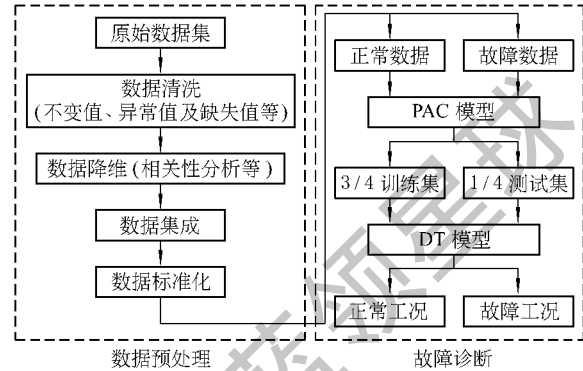


图 1 基于 PCA-DT 的多联机制冷剂充注量故障诊断流程图

## 3 故障检测与诊断结果及分析

### 3.1 实验数据

对于使用 PCA 降维处理后的数据,其主元个数由设定的累积贡献率来决定,主元个数为 1~14 时,累积贡献率分别为 0.521, 0.758, 0.874, 0.921, 0.950, 0.967, 0.976, 0.983, 0.989, 0.993, 0.996, 0.998, 0.999 和 1.000。

图 2 为 DT 和 PCA-DT 算法的诊断正确率.可以看出:当  $k=4$  时,PCA-DT 算法的诊断正确率最低;当  $k=6$  时,两种算法的诊断正确率大致相同; $k>6$  时,PCA-DT 算法的诊断正确率较 DT 算法有所提高,提高了约 3.32%。

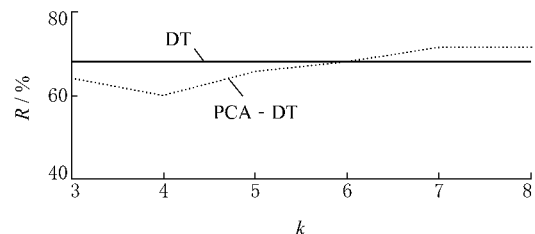


图 2 DT 和 PCA-DT 算法的诊断正确率

### 3.2 网络数据

对三台同型号机组(VRF1, VRF2 和 VRF3)在不同地区的网络数据进行故障检测,得到的诊断正确率如图 3 所示.从图 3 中可以看出:PCA-DT 算法的诊断正确率高于 DT 算法, $k=4$  时的差距最大。

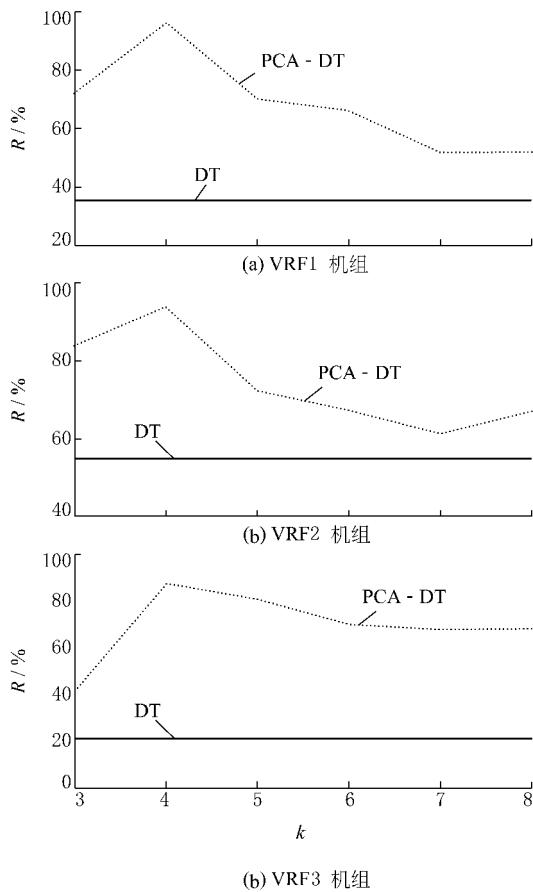


图 3 三台同型号机组的网络数据故障检测诊断正确率

使用 PCA-DT 算法对多联机的制冷剂充注量故障进行诊断,在数据预处理中引入了主元分析法,对挑选出来的变量进行降维,并将得到的新变量利用决策树算法建立决策树模型,对实际运行数据进行故障诊断.主要结论如下:

a. 对于实验数据,主元分析算法的引入对于决策树算法的诊断正确率有较大提高,但主元个数的取值还须进一步探讨,在本研究中主元个数大于 6 时,PCA-DT 算法的诊断正确率高于 DT 算法;

b. 对于网络数据,PCA-DT 算法相较于 DT 算法,较大幅度地提高了诊断正确率.

## 参 考 文 献

- [1] Frank P M. Analytical and qualitative model-based fault diagnosis—a survey and some new results[J]. European Journal of control, 1996, 2(1): 6-28.
- [2] Xiao F, Wang S, Xu X, et al. An isolation enhanced PCA method with expert-based multivariate decoupling for sensor FDD in air-conditioning systems[J]. Applied Thermal Engineering, 2009, 29(4): 712-722.
- [3] 卞荷洁,谷波,黎远光. 基于概率神经网络的制冷空调系统故障诊断分析[J]. 上海交通大学学报, 2004, 38(10): 1613-1616.
- [4] 谢慧,李德英,周国兵. 模糊推理在空调系统故障诊断中的应用[J]. 计算机工程, 2006, 32(4): 24-26.
- [5] Al Nasser A, Tucker A, de Cesare S. Quantifying StockTwits semantic terms' trading behavior in financial markets: an effective application of decision tree algorithms[J]. Expert Systems with Applications, 2015, 42(23): 9192-9210.
- [6] Uehara Y, Endo S, Shiitani S, et al. A computer-aided visual exploration system for knowledge discovery from images [C] // Proc of MDM/KDD. San Francisco: Quality Color Press, 2001: 102-109.
- [7] Zaiane O R, Antonie M L, Coman A. Mammography classification by an association rule-based classifier [C] // Proc of MDM/KDD. Edmonton: Quality Color Press, 2002: 62-69.
- [8] 冯少荣. 决策树算法的研究与改进[J]. 厦门大学学报: 自然科学版, 2007, 46(4): 496-500.
- [9] Xiao F, Fan C. Data mining in building automation system for improving building operational performance[J]. Energy and Buildings, 2014, 75: 109-118.
- [10] Li G, Hu Y, Chen H, et al. Extending the virtual refrigerant charge sensor (VRC) for variable refrigerant flow (VRF) air conditioning system using data-based analysis methods[J]. Applied Thermal Engineering, 2016, 93: 908-919.