

# 故障诊断分析在柴油机运行故障检测中的应用

## 1 问题背景

随着科学和生产技术的发展，现代设备大多数集机电液于一体，结构越来越复杂，自动化程度越来越高。在工作时，出现故障的概率就越来越高，连带着其他部分也会出现故障，危害极大。柴油机由于本身结构复杂，加上系统的输入输出不明显，难以用比较精确的机理模型对其进行准确的描述，因此给故障诊断带来了很大的麻烦。但是近年来智能算法的出现，为故障诊断问题过了一种新的解决途径，尤其是随机森林。随机森林方法具有分类精度高和泛化能力强的特点，对噪声和异常值的稳健性较好。该方法在数字识别、图像处理和数据挖掘等众多领域受到广泛关注。所以这里利用随机森林对燃油系统输出的压力波形特征进行分析挖掘，进而判断柴油机运行的故障原因。

## 2 问题描述

对于燃油压力波形来说，最大压力 ( $P_1$ )、次最大压力 ( $P_2$ )、波形幅度 ( $P_3$ )、上升沿宽度 ( $P_4$ )、波形宽度 ( $P_5$ )、最大余波的宽度 ( $P_6$ )、波形面积 ( $P_7$ )、起喷压力 ( $P_8$ ) 等特征最能体现柴油机运行的情况。

燃油系统常见的故障有供油不足，针阀门卡死致油孔阻塞，针阀门泄露，出油阀失效等几种故障。本例里诊断的故障也是基于上述故障，主要有 100%供油量 ( $T_1$ )，75%供油量 ( $T_2$ )，25%供油量 ( $T_3$ )，怠速油量 ( $T_4$ )，针阀门卡死（小油量  $T_5$ )，针阀卡死（标定油量  $T_6$ )，针阀泄露 ( $T_7$ )，出油阀失效 ( $T_8$ ) 8 种故障。

## 3 方法与结果

随机森林 (RF) 是在 Bagging 方法基础上引入了随机属性，这个随机主要体现在随机选取样本和随机选取特征属性，从而提高模型的不确定性和多样性，具体的两个随机过程如下：

(1) 随机样本：从原样本集中有放回的随机抽取  $K$  个训练样本集  $\{T_k, k = 1, 2, \dots, K\}$ ，未被抽取的样本数据作为袋外数据；

(2) 随机属性：从  $M$  个特征中等概率的随机选取  $m$  个特征作为特征子集。各决策树进行分裂时，在相应的子集中选择一个最优属性分裂节点。（ $m$  通常取

$\sqrt{M}$ ，M 为特征总数）。

随机森林是以随机的方式构建多棵决策树，各决策树的训练相互独立，最后采用投票法根据基决策树的结果决策最终输出的分类结果。RF 的训练过程实际上是训练各个决策树的过程，该过程将每一个子样本集，采用 CART 决策树方法生长为单棵分类树。采用随机森林进行分类，需要将样本输入到每棵决策树中进行分类。

采用故障诊断率(即准确率 FDR)作为诊断模型的分类评价指标，从而对诊断结果进行分析。

$$FDR = \frac{TP}{TP+FP} \tag{1}$$

其中，TP、FN、FP、TN 由混淆矩阵可得，如表 1 所示。

表 1 混淆矩阵			
混淆矩阵 Confusion Matrix		预测值 Predict	
		正	负
		Positive	Negative
真实值 Real	正 True	TP	FN
	负 False	FP	TN

调用 Python 中 sklearn 库中封装好的 RandomForestClassifier，将训练集的各特征值作为输入，故障类型对应的标签作为输出。经过一定的参数寻优之后确定了决策树的数量 K 为 4，随机属性的个数 m 为 3。

建立好模型之后利用测试集进行测试，得到精确率即故障诊断率为 95.83%。其预测值与真实值的比较如图 1 所示。该模型的精确度可以进一步提高，本案例的不足之处在于训练的数据样本太少，可以增加数据样本进行进一步的试验。

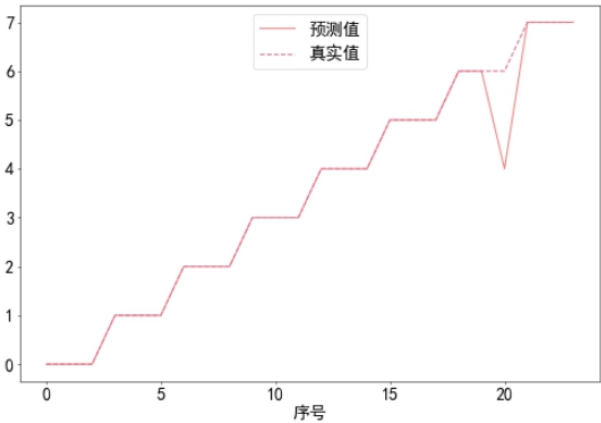


图 1 预测值与真实值比较