

# 控制与决策

Control and Decision

## 基于相似度的燃气轮机故障可诊断性分析方法

韩道豫, 曹云鹏, 屈东生, 冯伟兴

引用本文:

韩道豫, 曹云鹏, 屈东生, 等. 基于相似度的燃气轮机故障可诊断性分析方法[J]. *控制与决策*, 2025, 40(9): 2859–2867.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2024.1142>

## 您可能感兴趣的其他文章

### Articles you may be interested in

#### 基于Fisher Score与最大信息系数的齿轮箱故障特征选择方法

[Fault feature selection method of gearbox based on Fisher Score and maximum information coefficient](#)

控制与决策. 2021, 36(9): 2234–2240 <https://doi.org/10.13195/j.kzyjc.2019.1770>

#### 基于局部信息增量与MPLS的质量相关故障检测方法

Quality-related fault detection method based on local information increment and MPLS

控制与决策. 2021, 36(7): 1647–1654 <https://doi.org/10.13195/j.kzyjc.2019.1402>

#### 基于广义主成分分析的重构故障子空间建模方法

Reconstructed fault subspace modelling method based on generalized principal component analysis

控制与决策. 2021, 36(4): 808–814 <https://doi.org/10.13195/j.kzyjc.2019.0818>

#### 基于向量集逆区间滤波的故障观测器设计

Vector set inversion interval filtering based fault observer design

控制与决策. 2021, 36(12): 2973–2981 <https://doi.org/10.13195/j.kzyjc.2020.0212>

#### $l_p$ -范数约束下MKL-OC-ELM的装备故障检测

MKL and OC-ELM fault detection based on  $l_p$ -norm constraint

控制与决策. 2021, 36(10): 2379–2388 <https://doi.org/10.13195/j.kzyjc.2020.0443>

# 基于相似度的燃气轮机故障可诊断性分析方法

韩道豫<sup>1</sup>, 曹云鹏<sup>2†</sup>, 屈东生<sup>3</sup>, 冯伟兴<sup>1</sup>

- 哈尔滨工程大学 智能科学与工程学院, 哈尔滨 150006;
- 哈尔滨工程大学 动力与能源工程学院, 哈尔滨 150006;
- 杭州汽轮动力集团股份有限公司 先进动力研究院, 杭州 310000)

**摘要:** 针对燃气轮机信号监测过程中存在噪声干扰及特征冗余所导致的数据稳定性差、诊断效率低的问题, 提出一种基于相似度的燃气轮机可诊断性分析方法. 首先, 通过多项式拟合得到运行数据的整体趋势, 消除局部突变导致的噪声, 计算得到噪声阈值, 解决监测信号噪声干扰问题; 然后, 利用动态时间规整算法对燃气轮机不同运行状态下的数据进行相似度计算, 得到独立性特征, 构建可诊断性评价矩阵与故障特征矩阵, 筛选燃气轮机最优特征组合, 解决监测信号特征冗余问题; 最后, 将筛选得到的最优特征组合作为深层卷积神经网络的输入, 建立故障诊断模型, 利用燃气轮机典型气路故障数据对所提出的方法进行实验验证, 结果表明, 所提出的方法有效去除了冗余特征, 具备良好的故障隔离性, 准确率达到 99.98%, 相对原始特征诊断效率提升 32.61%.

**关键词:** 燃气轮机; 可诊断性; 时序相似度; 特征提取; 故障诊断

中图分类号: U664.131 文献标志码: A

DOI: 10.13195/j.kzyjc.2024.1142

引用格式: 韩道豫, 曹云鹏, 屈东生, 等. 基于相似度的燃气轮机故障可诊断性分析方法 [J]. 控制与决策, 2025, 40(9): 2859-2867.

## A similarity-based approach for diagnosability analysis of gas turbine faults

HAN Xiao-yu<sup>1</sup>, CAO Yun-peng<sup>2†</sup>, QU Dong-sheng<sup>3</sup>, FENG Wei-xing<sup>1</sup>

- College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin 150006, China;
- College of Power and Energy Engineering, Harbin Engineering University, Harbin 150006, China;
- Advanced Power Research Institute, Hangzhou Steam Turbine Power Group Co., Ltd., Hangzhou 310000, China)

**Abstract:** A similarity-based diagnosability analysis method for gas turbines is proposed to aim at the problems of poor data stability and low diagnostic efficiency caused by noise interference and feature redundancy in the gas turbine signal monitoring process. Firstly, the overall trend of the operation data is obtained by polynomial fitting, the noise caused by local mutation is eliminated, and the noise threshold is calculated to solve the noise interference problem in the monitoring signal. Then, the time-series similarity algorithm is used to calculate the similarity of the data under different operating conditions of the gas turbine, obtain the independence features, construct the diagnosability evaluation matrix and the fault feature matrix, and screen the optimal features of the gas turbine to combine the fault features, to solve the problem of redundancy of the monitoring signal features. Finally, the optimal combination of filtered features is used as the input of the convolutional neural network to establish the fault diagnosis model. The proposed method is experimentally verified using typical gas turbine fault data, and the results show that the proposed method effectively removes the redundant features, has good fault isolation, and achieves an accuracy of 99.98%, which improves the diagnostic efficiency by 32.6% relative to the original features.

**Keywords:** gas turbine; diagnosability; temporal similarity; feature extraction; fault diagnosis

## 0 引言

在传统工业领域和现代工业 4.0 背景下, 对设备

进行可诊断性分析成为实现设备故障诊断的重要环节. 故障特征提取作为可诊断性分析的一个重要组

收稿日期: 2024-09-25; 录用日期: 2025-02-20.

基金项目: 基础产品创新科研基金项目 (KY10300240082).

责任编辑: 叶丹.

†通信作者. E-mail: caoyunpeng@hrbeu.edu.cn.

成部分,对故障诊断的效果具有决定性作用<sup>[1]</sup>.

故障可诊断性是系统在故障诊断过程中所展现出的识别和隔离故障的能力,涵盖了故障能否被识别、识别的难度及所需的代价等方面<sup>[2]</sup>.它不仅解决了故障是否能被检测的问题,还评估了故障诊断的复杂性.可诊断性研究包含两个重要方向:可诊断性评价和设计<sup>[3]</sup>.评价主要用于分析系统当前状态下的故障检测与隔离能力,并为后续设计提供理论基础.当系统可诊断性不足时,通常通过优化测试点配置来提升检测信息量,进而提高诊断能力<sup>[4]</sup>.

故障特征提取是可诊断性分析的关键环节,旨在从系统或设备数据中提取能够反映故障状态的特征,以实现故障的准确诊断与预测.其核心在于选择最具代表性和区分度的特征,从而提高故障识别的精度与效率.现代工业中,特征提取方法主要包括相似度计算、特征选择算法、机器学习技术、信息论方法、基于物理原理的分析及频域分析等.这些方法不仅提升了系统的故障诊断能力,还增强了设备的可维护性与可靠性.具体而言,基于相似度的方法通过衡量数据样本间的相似性提取代表性特征或降维,如相位相似度<sup>[5]</sup>和波形相似度<sup>[6]</sup>;特征选择算法利用卡方检验、信息增益、基尼指数、Lasso 回归等方法,根据特征与故障类别的相关性进行排序与选择<sup>[7]</sup>;机器学习方法通过主成分分析 (PCA)<sup>[8]</sup>、线性判别分析 (LDA)、决策树<sup>[9]</sup>及遗传算法<sup>[10]</sup>等实现特征选择与降维;基于信息论的方法(如互信息、最大信息系数)通过量化特征与故障状态间的信息量进行特征选取<sup>[11]</sup>;基于物理原理的方法则利用系统或设备的物理特性选择相关参数作为特征<sup>[12]</sup>;频域分析方法通过频谱分析、小波变换等技术提取频域特征,如田野等<sup>[13]</sup>采用滤波技术减少低频干扰,结合频谱信号和时频图像提升特征区分度,姚欣等<sup>[14]</sup>利用最大重叠离散小波变换实现高精度信号特征提取.

尽管上述方法在故障诊断领域取得了一定成效,但仍存在一些不足之处.首先,监测信号易受噪声干扰,导致数据稳定性差,影响故障特征的准确提取;其次,特征冗余问题尚未得到有效解决,上述方法在提取特征时未能充分考虑特征间的相关性,导致冗余特征较多,降低了诊断效率;最后,上述方法往往依赖于单一的数据驱动或物理原理,缺乏对领域知识的有效结合,导致特征选择的准确性和鲁棒性不足.

为此,本文提出一种基于相似度的燃气轮机可诊断性分析方法.主要工作如下:

1) 设计一种基于多项式拟合的噪声消除机制,

通过提取运行数据的整体趋势并计算噪声阈值,有效消除了局部突变噪声,增强了监测信号的稳定性;

2) 提出一种基于动态时间规整算法的独立性特征提取方法,通过计算不同运行状态数据的相似度,构建可诊断性评价矩阵与故障特征矩阵,筛选最优特征组合,解决了高维数据中的特征冗余问题,为故障诊断模型提供了高质量输入;

3) 设计一种融合数据驱动与领域知识的故障诊断模型,将最优特征组合输入至诊断模型中,实验结果表明,该方法在保证诊断精度的同时,显著缩短了计算时间,提升了诊断效率.

## 1 可诊断性分析方法

### 1.1 多项式拟合降噪

多项式拟合是一种用于拟合数据的方法,通过构造一个多项式来逼近真实数据.多项式的次数取决于数据的复杂性,通常是1次方到6次方.在多项式拟合中经常使用最小二乘法来优化多项式的参数,以求得最佳的拟合效果<sup>[15]</sup>.

对于数据集  $y = f(x)$ , 多项式拟合的目的是求解出次数为  $n$  的多项式  $p(x)$  的系数,该阶数是数据  $x$  对数据  $y$  映射关系的最佳拟合.  $p$  中的系数按降幂排列,  $p$  的长度为  $n + 1$ , 计算过程如下:

$$p(x) = p_1 x^n + p_2 x^{n-1} + \dots + p_n x + p_{n+1}. \quad (1)$$

对于一维向量  $x$ , 构造具有  $n + 1$  列和  $m$  行的范德蒙矩阵  $V$  并生成线性方程组, 计算过程如下:

$$\begin{bmatrix} x_1^n & x_1^{n-1} & \dots & 1 \\ x_2^n & x_2^{n-1} & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ x_m^n & x_m^{n-1} & \dots & 1 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_{n+1} \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}, \quad (2)$$

其中  $m$  为数据  $x$  长度.

拟合多项式  $p(x)$  转换为求解线性方程组, 先对矩阵  $V$  进行正交三角 ( $QR$ ) 分解, 然后对多项式矩阵  $p$  进行求解, 计算过程如下:

$$\begin{aligned} Vp &= y, \\ QRp &= y, \\ Rp &= Q^T y, \\ p &= R^{-1} Q^T y. \end{aligned} \quad (3)$$

对于含有噪声的数据  $x$ , 通过低阶多项式拟合降低数据复杂度, 达到数据降噪的目的.

### 1.2 动态时间规整

时间序列相似性度量方法总体可分为锁步度量和弹性度量两类. 锁步度量是时间序列进行一对一的比较, 弹性度量允许时间序列进行一对多的比较<sup>[16]</sup>. 针对时序的数据进行相似度计算, 提出使用 DTW

(dynamic time warping) 算法, DTW 算法是一种计算时间序列之间距离的方法, 主要用于处理序列在时间轴上并不完全对齐的情况. DTW 算法的基本思想是, 在两个时间序列之间选取一个合适的路径, 使得路径上每个点的纵坐标之间的距离最小. 这个距离称为 DTW 距离, 它可以作为两个序列之间的相似度量.

给定两个时间序列  $A$  和  $B$ , 假设  $A$  的长度为  $n$ ,  $B$  的长度为  $m$ . DTW 算法将两个时间序列看作二维平面上的点集, 其中一个时间序列被视为横坐标轴上的点, 另一个时间序列被视为纵坐标轴上的点. 通过动态规划的方式计算得到一个最优路径, 该路径上的所有点构成了一个最优的匹配序列, 实现两个时间序列之间的相似性信息的捕捉.

定义一个二维矩阵  $DTW[N, M]$ ,  $DTW[i, j]$  表示在  $(i, j)$  位置的 DTW 距离. DTW 矩阵的计算过程如图 1 所示.

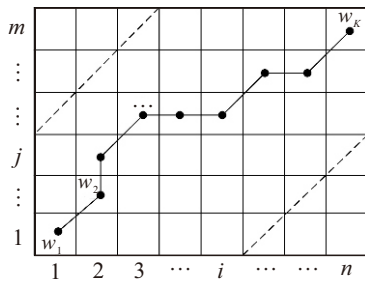


图1 DTW 选取过程

step1: 初始化.  $DTW[0, 0], DTW[i, 0] = \infty$ ,  $DTW[0, j] = \infty$ , 其中  $i = 1, 2, \dots, n, j = 1, 2, \dots, m$ .

对于每个格子  $DTW[i, j]$ , 计算 3 个距离值  $d_0, d_1, d_2$ , 距离计算方法如图 2 所示.

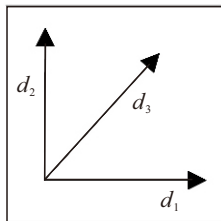


图2 距离计算方法

step2: 计算  $A[i]$  与  $B[j]$  之间的距离 (欧几里得距离)  $d_0 = \text{distance}(A[i], B[j])$ .

step3: 计算从左边的格子到当前格子的距离  $d_1 = DTW[i - 1, j] + d_0$ .

step4: 计算从下边的格子到当前格子的距离  $d_2 = DTW[i, j - 1] + d_0$ .

step5: 计算从左下角的格子到当前格子的距离

$$d_3 = DTW[i - 1, j - 1] + d_0.$$

step6: 取 3 个距离值的最小值  $DTW[i, j] = \min(d_1, d_2, d_3)$ .

在计算过程中, 可以采用动态规划的方式, 从左下角开始逐行或逐列地计算 DTW 矩阵的每个元素. 通过填充整个 DTW 矩阵, 即可得到两个时间序列之间的 DTW 距离.

### 1.3 相似度

考虑到数据中特征数量众多且可能存在内部关联, 这可能导致多个特征在反映设备相同故障时表现出相似的变化趋势. 因此, 在进行特征降维之前, 应先对特征进行聚类分析, 筛选出具有较强独立性的特征, 并排除不必要的冗余特征.

已知正常数据信号  $F_0$  与故障数据信号  $F_1 \sim F_n$ . 采用多项式拟合对正常数据进行滤波操作求解拟合多项式  $p(x)$ , 计算过程如下:

$$p(x) = p_1x + p_2. \tag{4}$$

将式 (4) 代入原式数据  $F_0$  得到滤波信号  $F'_0$ , 将滤波信号看作标准运行数据, 计算原始信号与标准信号之间的 DTW 相似度. 即可得到由信号噪声引起的相似度差异. 此相似度为噪声阈值  $ini$ , 即相似度差异大于该阈值时算法才考虑信号变化是由故障引起, 计算过程如下:

$$ini = DTW(F_0, F'_0). \tag{5}$$

1) 包含冗余特征的信号相似度  $d'_{tk}$ .

记通道  $t_k$  在发生故障  $F_i$  时所检测到的信号为  $\bar{s}_k$ , 在发生故障  $F_j$  时所检测到的信号为  $\hat{s}_k$ . 此时测试信号  $\bar{s}_k$  与  $\hat{s}_k$  的相似度为

$$d'_{tk} = d(\bar{s}_k, \hat{s}_k) = \begin{cases} DTW(\bar{s}_k, \hat{s}_k), & DTW(\bar{s}_k, \hat{s}_k) > ini; \\ 0, & DTW(\bar{s}_k, \hat{s}_k) \leq ini. \end{cases} \tag{6}$$

2) 信号相似度  $d_{tk}$ .

计算故障  $F_i$  与故障  $F_j$  所有通道信号相似度, 对所有信号相似度进行分类处理, 计算过程如下:

$$\Phi = \sum_{i=1}^p \Phi_i, \quad |d'_{tr} - d'_{ty}| \leq I, \quad (d'_{tr}, d'_{ty}, \dots, d'_{tu}) \rightarrow \Phi_i. \tag{7}$$

其中:  $p$  为类别数;  $r, y, \dots, u$  为一个类别中的信号通道;  $I$  为检测阈值, 本文基于不同故障类型数据的试验结果和经验, 设定  $I$  为 10%.

同类相似度中可检测性相似, 为避免特征冗余故同类相似度中保存第 1 个信号相似度, 同时令其

他相似度为 0, 计算过程如下:

$$\begin{cases} d_{tr} = d'_{tr}, \\ d_{ty}, \dots, d_{tu} = 0. \end{cases} \quad (8)$$

#### 1.4 可诊断性评价

建立可诊断性指标用于评估故障之间是否能够被准确区分, 确定不同故障模式( $F_i, \dots, F_j$ )之间的可隔离性. 当两个故障共享类似的信号模式时, 它们的可隔离性较低; 当信号通道彼此独立时, 故障模式之间更容易区分.

1) 可诊断性评价矩阵.

故障 $F_i$ 与 $F_j$ 的可隔离性指标计算过程如下:

$$d_{F_i, F_j} = \sum_{k=1}^n d_{ik} i, \quad j = 0, 1, \dots, m. \quad (9)$$

其中:  $n$ 为信号通道总数,  $m$ 为故障模式数,  $k = 1, 2, \dots, n$ .

构建故障可诊断性评价矩阵 $\Delta$ , 计算过程如下:

$$\Delta = \begin{matrix} & \begin{matrix} F_0 & F_1 & F_2 & \dots & F_m \end{matrix} \\ \begin{matrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{matrix} & \begin{bmatrix} d_{F_1, F_0} & * & d_{F_1, F_2} & \dots & d_{F_1, F_m} \\ d_{F_2, F_0} & d_{F_2, F_1} & * & \dots & d_{F_2, F_m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{F_m, F_0} & d_{F_m, F_1} & d_{F_m, F_2} & \dots & * \end{bmatrix} \end{matrix} \quad (10)$$

其中:  $d_{F_i, F_j} = d_{F_j, F_i}$ ; 对于任意故障, 不存在与其自身的可隔离性, 用\*表示, 值为 0. 故障可诊断性评价矩阵, 明确了系统中每个故障对其他故障的可隔离性信息, 展示了故障模式在不同信号通道下的诊断性能.

2) 故障特征矩阵.

通过分析不同故障模式间的信号相似度, 构建故障特征矩阵 $\Theta$ , 从而量化故障之间的差异. 这一过程对于确定哪些信号通道对不同故障模式有诊断意义至关重要. 故障 $F_i$ 与故障 $F_j$ 中信号相似度不为 0 的信号通道为故障间诊断特征

$$\Theta_{i,j} = [x, y, \dots, z] d_{t(x,y,\dots,z)} \neq 0. \quad (11)$$

构建故障特征矩阵

$$\Theta = \begin{matrix} & \begin{matrix} F_0 & F_1 & F_2 & \dots & F_m \end{matrix} \\ \begin{matrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{matrix} & \begin{bmatrix} \Theta_{1,0} & * & \Theta_{1,2} & \dots & \Theta_{1,m} \\ \Theta_{2,0} & \Theta_{2,1} & * & \dots & \Theta_{2,m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Theta_{m,0} & \Theta_{m,1} & \Theta_{m,2} & \dots & * \end{bmatrix} \end{matrix} \quad (12)$$

整个系统的故障特征 $\theta$ 计算过程如下:

$$\theta = \bigcup_{i=0}^m \bigcup_{j=1}^{m-1} \Theta_{i,j} \Theta_{i,j+1}, \quad i \neq j. \quad (13)$$

通过式 (13) 对整个系统的故障特征进行量化, 从全局角度评价系统在故障诊断方面的能力. 这一评价可以为系统设计者或维护者提供重要的参考, 帮助优化故障诊断策略.

## 2 燃气轮机可诊断性分析算法流程

基于相似度的燃气轮机故障可诊断性分析方法, 在燃气轮机故障诊断的特征提取过程中具体分为 3 个步骤, 可以有效提高故障诊断的准确性和效率.

**step 1: 数据预处理阶段.** 在这个阶段, 首先选取燃气轮机的正常运行数据, 并对其进行降噪处理. 通过多项式拟合, 对原始监测数据进行平滑和降噪, 得到无噪声的理想运行数据. 然后, 利用动态时间规整算法计算原始数据与理想运行数据之间的相似度, 并将此相似度作为噪声阈值. 阈值用于后续判断监测数据中是否存在故障信号, 只有当某个测点的异常波动超过该噪声阈值时, 才认为该波动可能是由故障引起.

**step 2: 可诊断特征选取阶段.** 在确定噪声阈值后, 需要计算不同运行状态下燃气轮机监测数据的 DTW 相似度. 在这个阶段, 将所有测点的相似度与噪声阈值进行对比. 对于那些未超过噪声阈值的测点, 认为其波动可能是由噪声引起, 而非故障信号, 因而可以暂时忽略这些测点. 相反, 超过噪声阈值的测点则被认为可能含有故障信息, 将其视为待选特征. 这些待选特征将用于进一步分析和特征筛选.

**step 3: 去冗余特征阶段.** 在这个阶段, 对 step 2 中确定的待选特征进行归类处理, 目的是去除冗余特征. 通过对不同测点的相似度进行比较, 如果两个测点的相似度差异小于阈值, 则认为它们属于同类测点, 可能含有冗余信息. 为了避免特征冗余, 只保留每一类中相似度最大的测点, 其他相似度较低的测点将被视为冗余测点而去除. 对于相似度差异大于阈值的特征, 则保留为独立特征. 最终, 所有的最佳检测特征和最佳隔离特征将被聚合, 形成用于燃气轮机故障诊断的最优特征集合.

**step 4: 典型故障诊断验证.** 将计算得到的最优特征集合作为诊断模型的输入, 输出故障诊断结果.

针对不同故障, 通过该方法可计算得到不同故障间的最佳隔离特征. 对所有的最佳检测特征与最佳隔离特征进行聚合, 可得到燃气轮机故障诊断的最优特征. 通过结合样本数据、相似度分析、去冗余处理等方法, 最终提取出最优的故障诊断特征. 可诊断性分析过程如图 3 所示.

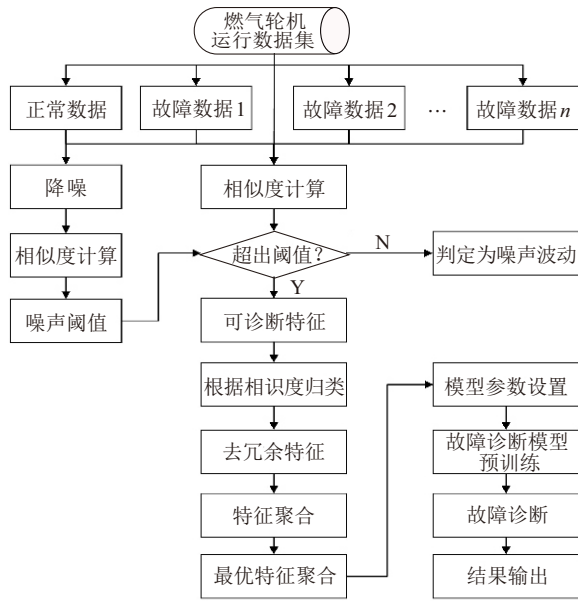


图3 可诊断性分析过程

### 3 实验验证

#### 3.1 数据集描述

以船舶燃气轮机为研究对象, 开展实验验证. 该设备包含低压压气机、高压压气机、低压涡轮、高压涡轮、动力涡轮等部件<sup>[17-18]</sup>, 收集整理典型故障模式, 如表1所示. 数据集中的监测测点数量共70个, 包含正常及13种典型故障数据, 每种类别样本数为1000.

表1 故障模式表

故障部件	模式	编号	故障部件	模式	编号
低压压气机	积垢	$F_1$	高压涡轮	积垢	$F_8$
	损伤	$F_2$		损伤	$F_9$
高压压气机	积垢	$F_3$	动力涡轮	磨损腐蚀	$F_{10}$
	损伤	$F_4$		积垢	$F_{11}$
低压涡轮	积垢	$F_5$	—	损伤	$F_{12}$
	损伤	$F_6$		磨损腐蚀	$F_{13}$
	磨损腐蚀	$F_7$	—	正常	$F_0$

使用主成分分析 (PCA) 对数据集进行降维可视化, 将输入特征降至2个主成分 PC1 和 PC2, 分别表示主要和次要的变异趋势. 同时, 使用核密度估计叠加了类别密度曲线, 展示各类数据在主成分轴上的分布趋势及重叠情况. 分布情况如图4所示, 14种数据类型在 PC1 和 PC2 空间中已经具有一定的区分性, 但部分类型在边缘密度图上仍表现出区间分布重叠, 说明了在某些故障类型上区分性较差.

由于基于数据的可诊断性评价方法的依据是传感器中的信号, 而信号的提取一定会包含噪声, 同时众多部件热力学参数之间存在一定的物理耦合, 对监测参数的噪声干扰及冗余性需重点考虑.

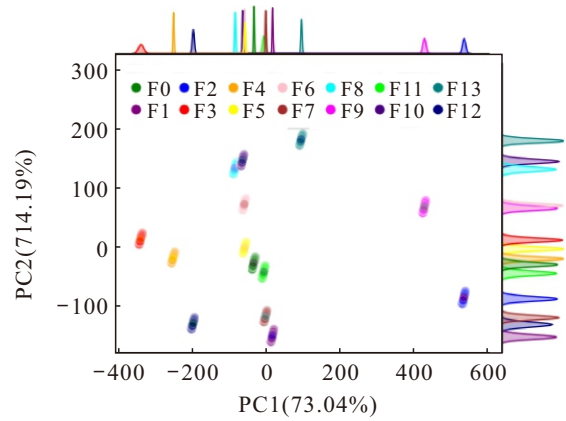


图4 原始数据分布情况

#### 1) 噪声干扰.

由于燃气轮机在信号监测过程中受到外界环境、传感器本身等多种因素产生噪声的干扰, 这种干扰会影响监测数据的准确性和稳定性, 从而对故障诊断的分析造成干扰. 以0.95工况下的低压压气机出口温度正常监测数据为例 (数据经归一化处理), 对其进行降噪处理, 降噪结果如图5所示.

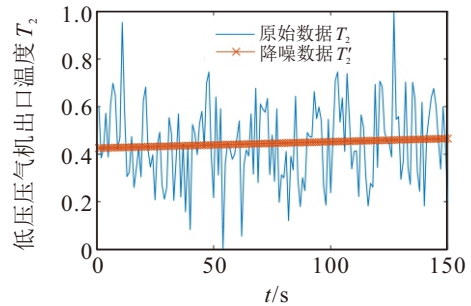


图5 基于相似度的噪声阈值

通过信噪比对降噪效果进行评估, 通常情况下, 信噪比的数值越大, 表示信号相对于噪声的能量越大, 降噪效果越好. 经过计算, 0.95工况下的低压压气机出口温度正常监测数据降噪后的信噪比为46.12 dB, 降噪效果优异.

对正常运行数据集  $F_0$  进行降噪处理, 得到理想运行数据  $F'_0$ . 计算噪声阈值  $ini$ , 燃气轮机正常数据下监测测点的噪声阈值如图6所示.

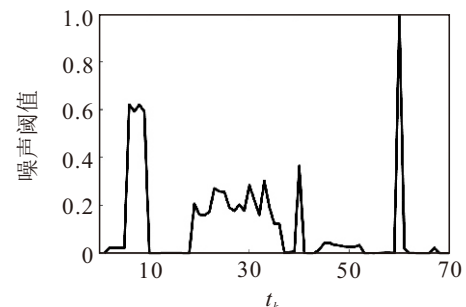


图6 基于相似度的噪声阈值

#### 2) 冗余性. 燃气轮机热力学监测测点在监测过

程中可能提供相似或重复的信息, 导致数据冗余. 这种冗余性会影响故障诊断的效率和准确性, 因为冗余特征可能掩盖真正的故障特征或引入不必要的复杂性. 以燃气轮机低压涡轮积垢为例, 低压涡轮后排

气温度点 1~16 的实际值如图 7 所示, 监测信号的多个测点显示出明显的相似性, 不同的测点可能监测到的是同一热力学过程的相似响应或相关联的物理现象, 这表明测点之间存在数据冗余.

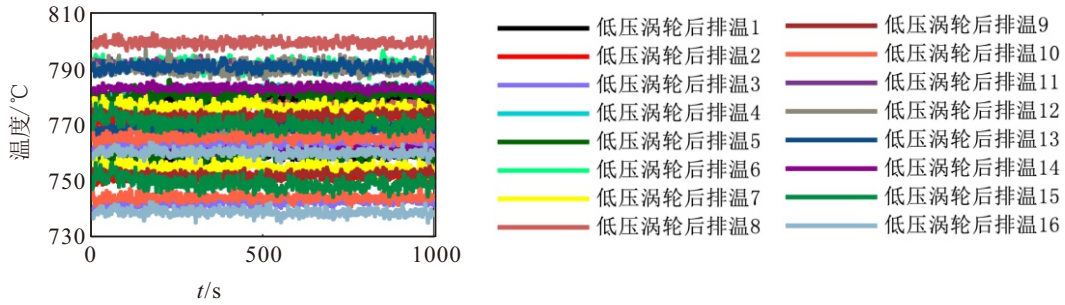


图7 排温监测测点冗余性

3.2 特征选取结果分析

根据表 1 对数据集  $F = \{F_0, F_1, \dots, F_{13}\}$  进行故障可诊断性分析, 通过式 (9) 和 (10) 计算得到可诊断性评价矩阵如图 8 所示. 其中, 矩阵中第 1 列为对应故障检测度, 第  $i$  行第  $j$  列 ( $j > 1$ ) 表示故障  $F_i$

与故障  $F_{j-1}$  之间的故障隔离度, 数值越大表示该故障引起的可诊断数据变化越多, 故障可诊断性越好; 反之, 则表示故障引起的数据变化越小, 故障特征越不明显, 故障隔离度为 1 时表示只有一个有效故障特征.

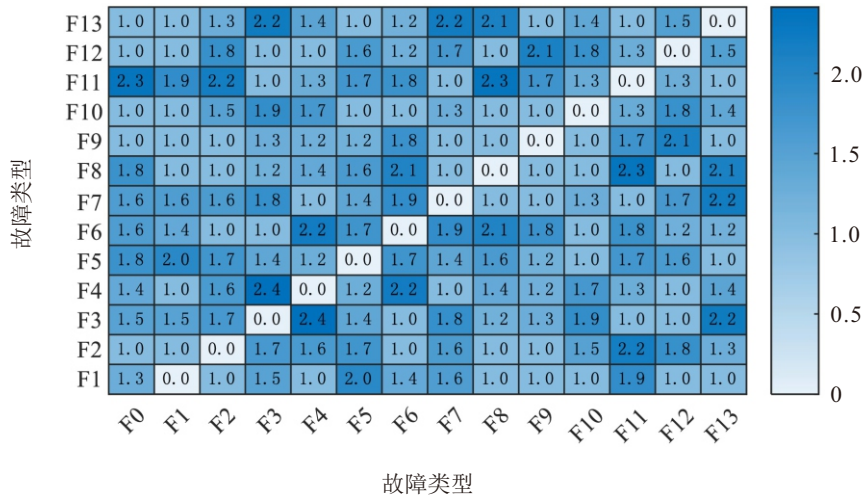


图8 可诊断性评价矩阵

通过式 (11) 和 (12) 计算得到故障特征矩阵  $\Theta$  如图 9 所示, 第  $i$  行第  $j$  列 ( $j > 1$ ) 表示故障  $F_i$  与故障

$F_{j-1}$  之间的故障隔离特征, 0 表示故障之间不存在特征.

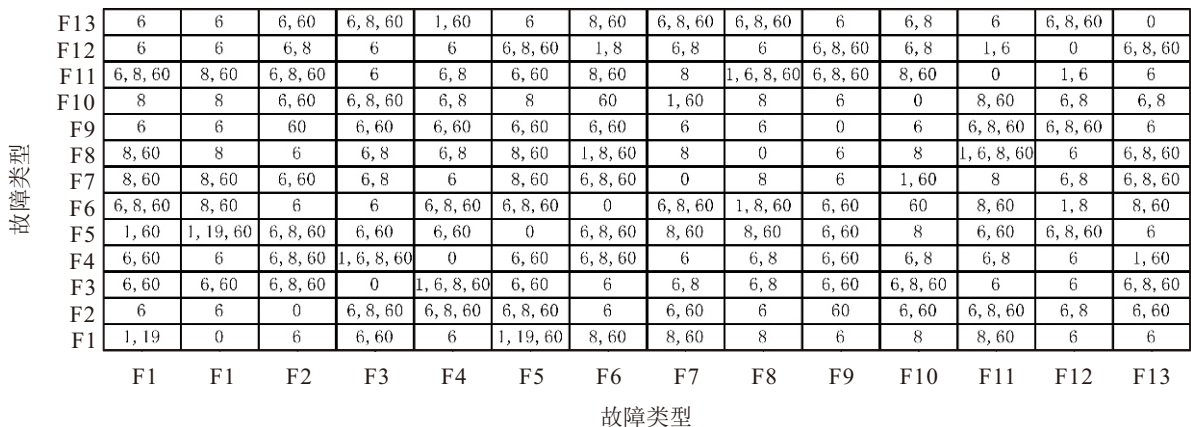


图9 故障特征矩阵

由上述故障特征矩阵得到燃气轮机的最优特征集合 $\theta$ : 测点 1 燃机进口温度、测点 6 高压转速、测点 8 低压转速、测点 19 低压涡轮后燃气温度、测点 60 燃机进口燃油温度, 共计 5 个特征, 原始特征共计 70 个, 仅占比 7.14%。经过优化选取的特征涵盖温度参数和转速参数, 并去除冗余特征, 对燃气轮机具有良好的可诊断性。

### 3.3 故障特征与测试点相关性验证

对于上述故障特征 $\theta$ 中各测点对不同故障的相关性也可验证故障可诊断性, 一个故障对于单个测点的相关性表示为该故障对所有故障在对应测点的隔离度的标准差, 计算过程如下:

$$\delta_{i,k} = \text{sqrt}\left(\sum_{j=1, j \neq i}^m (x_j - \bar{x})^2\right). \quad (14)$$

其中:  $i$ 表示对应故障,  $k$ 表示对应测点。

计算所有故障相关性, 得到的故障与测试点相关性关系如图 10 所示。图 10 中横坐标为故障特征, 纵坐标为故障种类, 对应坐标下数值为该状态对于故障特征的故障相关性。同一个故障特征下, 数值相差越小说明两种状态在该特征下更难区分, 需要借助

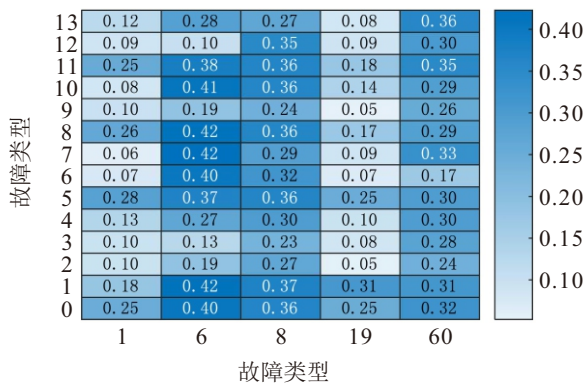


图10 故障与测试点相关性关系

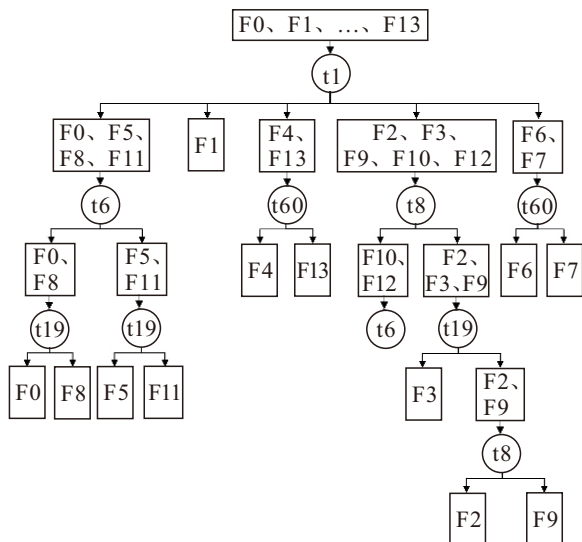


图11 故障隔离流程

其他特征辅助诊断; 反之数值相差越大说明该特征下两种状态具有更好的可隔离性。

根据故障与测试相关性关系可以得到故障诊断流程, 根据标准差的差异体现出数据的差异化, 从而对不同故障进行隔离, 图 11 是一种隔离流程, 借助所选的 5 种特征可以将 14 种状态隔离开。

### 3.4 整体验证

利用卷积神经网络 (CNN) 搭建故障诊断模型, 对燃气轮机 14 种状态 (正常及 13 种故障类型) 进行故障诊断验证, 每种类别样本数量为 1 000, 数据集样本总数为 14 000 个, 模型的输入为计算得到的故障特征集合 $\theta$ , 故障诊断模型的结构参数如表 2 所示。

表2 故障诊断模型结构参数

编号	神经层	大小	步长	数量	零填充
1	卷积层1	2×1	1×1	16	是
2	池化层1	2×1	1×1	16	否
3	卷积层2	2×1	1×1	32	是
4	池化层2	2×1	1×1	32	否
5	卷积层3	2×1	1×1	32	是
6	池化层3	2×1	1×1	32	否
7	卷积层4	2×1	1×1	32	是
8	池化层4	2×1	1×1	32	否
9	全连接层	—	—	14	—
10	Softmax	—	—	14	—

实验将随机抽取全部样本的 70% 划分为训练样本, 剩余的 30% 作为测试样本, 划分完后训练样本共计 9 800 个, 测试样本共计 4 200 个。利用搭建好的故障诊断模型对数据集进行验证, 模型的迭代次数设为 50 次。

为分析经可诊断性分析得到的最优组合特征对典型故障诊断的效果, 本文进行了不同比例输入特征的实验。实验分别以 100%、80%、60%、40%、20% 特征及最终最优特征集合 $\theta$ 作为故障诊断模型的输入。为排除 CNN 自身特征提取能力对结果的影响, 还对浅层和深层 CNN 的可诊断性分析效果进行了对比验证, 实验结果如表 3 所示。由表 3 可知, 当输入为最优特征集合时, 该集合仅占原始特征数量的 7.14%, 但诊断精确率仍与输入原始特征时相近。同时, 单步迭代计算速度从 1.14 下降至 0.81(浅层 CNN), 从 1.38 下降至 0.93(深层 CNN), 速度提升幅度分别为 28.95% 和 32.61%。速度提升的主要原因在于最优特征集合通过降低特征维度, 减少了计算量, 同时去除了冗余信息, 优化了特征分布, 从而提高了计算效率。此外, 在相同特征输入情况下, 浅层 CNN 的诊

断精度仅略低于深层 CNN, 这表明本文提出的故障可诊断性分析方法不仅能有效提取关键特征, 还能在保证诊断精度的同时提升计算效率.

表3 故障诊断结果

输入特征	精确率/%		F1分数均值/%		速度/(s/epoch)	
	2层	4层	2层	4层	2层	4层
100%特征	99.79	100	99.79	100	1.14	1.38
80%特征	96.72	98.88	96.72	98.88	1.07	1.33
60%特征	79.98	84.24	79.98	84.24	0.99	1.26
40%特征	77.93	80.45	77.93	80.45	0.92	1.18
20%特征	55.96	63.36	55.96	63.36	0.85	1.02
特征集合 $\theta$	98.14	99.98	98.14	99.98	0.81	0.93

本文同时将所提出的方法与传统特征提取方法(如主成分分析(PCA)、线性判别分析(LDA)、独立成分分析(ICA))进行实验对比, 展示其在相同任务中的优势. 实验结果如表4所示, 表明所提出方法在

保证诊断精度的同时, 有效缩短了计算时间.

表4 故障诊断结果

方法	精确率/%	F1分数均值/%	计算时间/s
PCA + CNN	100	100	71
LDA + CNN	98.76	98.76	73
ICA + CNN	99.34	99.34	85
本文方法	99.98	99.98	47

为了更好地理解本文所提出方法的特征提取过程, 利用  $t$ -SNE 技术分别对原始数据、最优特征集合及 Softmax 层处理后的结果进行降维可视化, 结果如图12所示. 由图12可知, 原始数据的初始特征聚类较为紧密, 聚类内的点分布较为集中, 说明不同类别数据之间的相似性较高. 在经过可诊断性分析进行特征筛选以后, 对比数据初始分布具有更加明显的可分性. 经过 Softmax 层处理后, 14 种类别的数据被很好地分隔, 验证了方法的有效性.

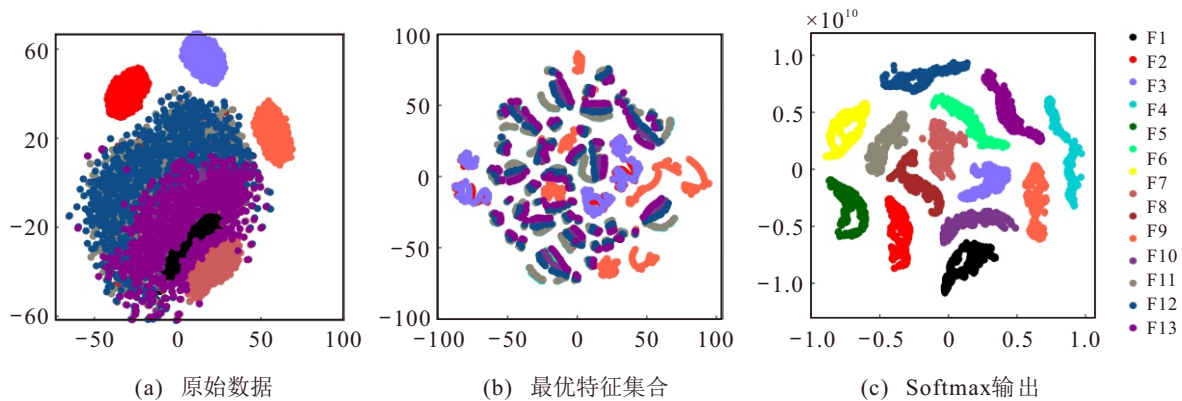


图12 数据降维可视化

## 4 结论

为了解决燃气轮机监测信号存在噪声干扰及冗余性问题, 本文提出了一种基于相似度的燃气轮机故障可诊断性分析方法, 并得到以下结论:

1) 通过多项式拟合方法有效消除了局部噪声干扰, 并采用动态时间规整算法对故障数据进行相似度计算, 成功提取了关键故障特征, 且经诊断验证了该方法具备可靠性.

2) 本文还针对燃气轮机最优故障特征的筛选进行深入探讨, 并基于故障可诊断性量化评价, 制定了故障隔离流程. 结果表明, 优化的特征选择方案显著提高了系统在故障检测与隔离过程中的效率.

3) 最后, 通过燃气轮机故障数据验证了所提出方法的有效性. 实验结果显示, 所提方法能够精确识别多种故障类型, 并显著提升了故障诊断的速度.

## 参考文献 (References)

[1] 张志强, 储昭碧, 陈立平, 等. 一种用于旋转机械故障

特征提取的簇紧凑自编码器[J]. 控制与决策, 2024, 39(7): 2251-2258.

(Zhang Z Q, Chu Z B, Chen L P, et al. A cluster compact auto-encoder for rotating machinery fault feature extraction[J]. Control and Decision, 2024, 39(7): 2251-2258.)

[2] 蒋栋年, 李炜, 王君, 等. 基于故障可诊断性量化评价的传感器优化配置方法研究[J]. 自动化学报, 2018, 44(6): 1128-1137.

(Jiang D N, Li W, Wang J, et al. Research on sensor optimal placement method using quantitative evaluation of fault diagnosability[J]. Acta Automatica Sinica, 2018, 44(6): 1128-1137.)

[3] Pérez-Zuñiga G, Chanthery E, Travé-Massuyès L, et al. Near-optimal decentralized diagnosis via structural analysis[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2022, 52(12): 7353-7365.

[4] 秦玉峰, 史贤俊. 基于 MMD 的故障可诊断性定量评价方法[J]. 控制与决策, 2023, 38(10): 2925-2933.

(Qin Y F, Shi X J. Quantitative evaluation approach of fault diagnosability based on maximum mean discrepancy[J]. Control and Decision, 2023, 38(10):

- 2925-2933.)
- [5] 李江涵, 张钧益, 熊庆, 等. 采用特征频率电流相位相似度的配电网接地故障选线方法[J]. 西安交通大学学报, 2025, 59(4): 81-92.  
(Li J H, Zhang J Y, Xiong Q, et al. A grounding fault line selection method for distribution networks based on feature frequency current phase similarity [J]. Journal of Xi'an Jiaotong University, 2025, 59(4): 81-92.)
- [6] 郑涛, 沈文韬, 陈冉, 等. 基于柔性互联装备主动控制的配电网故障定位方法[J]. 电力自动化设备, 2024, 44(12): 35-41.  
(Zheng T, Shen W T, Chen R, et al. Fault location method for distribution network based on active control of flexible interconnected equipment[J]. Electric Power Automation Equipment, 2024, 44(12): 35-41.)
- [7] Peng H C, Long F H, Ding C. Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and Min-redundancy[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 27(8): 1226-1238.
- [8] Jolliffe I T. Principal component analysis[M]. The 2nd Edition. New York: Springer, 2002.
- [9] Quinlan J R. Induction of decision trees[J]. *Machine Learning*, 1986, 1(1): 81-106.
- [10] Goldberg D E. Genetic algorithms in search, optimization, and machine learning[M]. Boston: Addison-Wesley Pub. Co. , 1989.
- [11] Tibshirani R. Regression shrinkage and selection via the lasso[J]. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 1996, 58(1): 267-288.
- [12] Pagán Rubio J A, Vera-García F, Hernandez Grau J, et al. Marine diesel engine failure simulator based on thermodynamic model[J]. *Applied Thermal Engineering*, 2018, 144: 982-995.
- [13] 田野, 陈姚节, 张莉, 等. 频域特征提取联合双流 CNN 的轴承故障诊断方法[J]. 噪声与振动控制, 2024, 44(4): 167-173.  
(Tian Y, Chen Y J, Zhang L, et al. Bearing fault diagnosis based on frequency domain feature extraction and combined dual-stream CNN[J]. Noise and Vibration Control, 2024, 44(4): 167-173.)
- [14] 姚欣, 邢砾云, 辛平. 基于小波特征提取与深度学习的微电网故障诊断与分类方法[J]. 智慧电力, 2021, 49(12): 17-24.  
(Yao X, Xing L Y, Xin P. Fault diagnosis and classification of microgrid based on wavelet feature extraction and deep learning[J]. Smart Power, 2021, 49(12): 17-24.)
- [15] Trefethen L N. Approximation theory and approximation practice[M]. Philadelphia: Society for Industrial and Applied Mathematics, 2013.
- [16] 陈海燕, 刘晨晖, 孙博. 时间序列数据挖掘的相似性度量综述[J]. 控制与决策, 2017, 32(1): 1-11.  
(Chen H Y, Liu C H, Sun B. Survey on similarity measurement of time series data mining[J]. Control and Decision, 2017, 32(1): 1-11.)
- [17] 徐鹏, 曹云鹏, 欧惠宇, 等. 船用三轴燃气轮机气路故障建模与聚类诊断技术[J]. 中国舰船研究, 2014, 9(3): 88-92.  
(Xu P, Cao Y P, Ou H Y, et al. Gas path fault modeling and cluster diagnosis for three-shaft marine gas turbines[J]. Chinese Journal of Ship Research, 2014, 9(3): 88-92.)
- [18] 张会生, 周登极. 燃气轮机可靠性维护理论及应用[M]. 上海: 上海交通大学出版社, 2016.  
(Zhang H S, Zhou D J. Reliability maintenance theory and its applications for gas turbine engine[M]. Shanghai: Shanghai Jiao Tong University Press, 2016.)

## 作者简介

韩道豫 (1996-), 男, 博士生, 主要研究方向为燃气轮机运行状态监测技术, E-mail: 2545256317@hrbeu.edu.cn;

曹云鹏 (1979-), 男, 副教授, 博士, 主要研究方向为动力系统故障诊断及健康管理, E-mail: caoyunpeng@hrbeu.edu.cn;

屈东生 (1998-), 男, 硕士, 主要研究方向为燃气轮机健康管理, E-mail: 852896422@qq.com;

冯伟兴 (1971-), 男, 教授, 博士, 主要研究方向为模式识别及其智能算法, E-mail: fengweixing@hrbeu.edu.cn.