

基于特征增维和近邻成分分析的民航发动机故障分类方法

孔祥兴¹, 刘凯伟², 莫李平², 王奕首², 卿新林²

(1. 中国航空发动机研究院, 北京 101304; 2. 厦门大学 航空航天学院, 厦门 361005)

摘要:为提高航空发动机故障诊断准确度,提出了一种从快速存取记录器(QAR)数据中提取最合适故障特征的方法。对原始QAR数据进行缺失值填补和巡航点提取操作,选择部分发动机性能参数差值作为初始特征值;再采用特征增维方法挖掘隐藏特征信息,进而采用近邻成分分析算法进行特征筛选优化,将所提方法与朴素贝叶斯等4种分类算法相结合,对某航空公司CFM56-7B发动机的QAR数据进行试验验证。结果表明:从QAR数据中提取最合适故障特征的方法能有效地提高发动机故障分类算法的准确率,且适用于不同的诊断算法,准确率优于80%。

关键词:特征增维;近邻成分分析;快速存取记录器数据;故障诊断;航空发动机

中图分类号: V263.3

文献标识码: A

doi: 10.13477/j.cnki.aeroengine.2022.05.002

Aeroengine Fault Classification Method Based on Feature Expand and Neighbourhood Components Analysis

KONG Xiang-xin¹, LIU Kai-wei², MO Li-ping², WANG Yi-shou², QING Xin-lin²

(1. AEAC, Beijing 101304, China; 2. School of Aerospace Engineering, Xiamen University, Xiamen 361005, China)

Abstract: In order to improve the accuracy of aeroengine fault diagnosis, a method for extracting the most suitable fault features from QAR data was proposed. Firstly, the missing value filling and cruising point extraction were performed on the original QAR data, and some engine performance parameter differences were selected as the initial characteristic values. Then the method of feature expand was used to excavate the hidden feature information and furthermore, feature selection and optimization were carried out by using the nearest neighbor component analysis algorithm (NCA). Finally, the method proposed in this paper was combined with four classification algorithms such as Naive Bayes to carry out experimental verification on the CFM56-7B engines' QAR data of an airline. The results show that the optimal features can effectively improve the accuracy of engine fault diagnosis algorithm and are applicable to different diagnosis algorithms, accuracy of diagnosis exceeds 80%.

Key words: feature expand; neighbourhood components analysis; QAR data; fault diagnosis; aeroengine

0 引言

航空发动机结构复杂且长时间工作在高温高压等恶劣环境下,其安全性和稳定性问题一直备受关注^[1]。数据表明,民用航空发动机的维修费用占航空公司整体运营支出的15%以上^[2]。因此,开展航空发动机故障诊断研究对保证飞行安全、降低维护成本起着关键作用。

快速存取记录器(Quick Access Recorder, QAR)作为一种带保护装置的机载飞行数据记录设备,其记录数据可连续完整地反映飞机各系统在整个航程中的实际运行状况^[3]。通过分析和挖掘QAR数据的隐含相关关系和知识,可以实现对发动机故障的诊断^[4-5]。目前,从QAR等飞行数据中提取特征主要有主成分分析、线性判别分析以及卷积神经网络等方法。徐萌等^[6]依据发动机制造商提供的故障报告选

收稿日期: 2020-09-21 基金项目: 装发联合基金(6141B090301)资助

作者简介: 孔祥兴(1982),男,博士,自然科学研究员,从事航空发动机健康管理技术研究工作;E-mail: 94589431@qq.com。

引用格式: 孔祥兴, 刘凯伟, 莫李平, 等. 基于特征增维和近邻成分分析的航空发动机故障分类方法[J]. 航空发动机, 2022, 48(5): 40-44. KONG Xiangxin, LIU Kaiwei, MO Liping, et al. Aeroengine fault classification method based on feature expand and neighbourhood components analysis[J]. Aeroengine, 2022, 48(5): 40-44.

择4种关键气路参数,结合Stacking集成学习模型,实现了航空发动机典型气路故障的智能诊断;顾彬等^[7]给出基于正交约束闭包球的最大间隔QAR数据特征提取方法,有效解决了QAR数据大样本特征提取问题;戴婧睿等^[8]利用深度置信网络算法提取QAR数据中的特征,并与主成分分析法对比验证了其提取的特征对提高分类识别准确率的有效性;张鹏等^[9]将卷积神经网络和长短时记忆网络通过注意力机制融合,使模型能同时表达QAR数据在空间维度和时间维度上的特征;Jiang等^[10]提出了一种基于主成分分析和深度置信网络相结合的航空发动机气路故障诊断方法,并验证了该方法的有效性;Cui等^[11]采用动态主成分分析算法对原始数据进行去噪、降维和消除相关性处理,并输入改进的支持向量机中进行发动机故障诊断;Cao等^[12]利用单位向量、比值系数和相关系数对数据进行降维处理,以支持向量机为基本分类器建立了多分类AdaBoost算法。

目前发展的多种常用的特征选择方法均具有一定的局限性。例如基于神经网络方法提取的特征可解释性差,缺乏明确的物理意义;主成分分析算法作为线性的降维方法,其应用具有局限性,部分情况下得到特征并不是最优的。

本文基于常用的5种特征,结合特征增维方法和近邻成分分析算法,提出了一种最合适的特征提取方法,挖掘隐藏信息并剔除冗余特征,并利用朴素贝叶斯和决策树等分类算法验证该方法对提升算法准确度的有效性。

1 近邻成分分析算法

近邻成分分析算法(Neighbor component analysis, NCA)是一种非参数的特征优化方法。其通过构建包含正则项的目标函数,将最优特征组合的选取问题转换为求目标函数最小值的问题^[13]。NCA在医疗检测、人脸识别、机械轴承故障诊断等研究方向上应用广泛^[14-16]。以 n 个训练样本的情况进行说明,训练样本空间为

$$S = \{(x_i, y_i), i = 1, 2, \dots, n\} \quad (1)$$

式中: $x_i \in R^p$ 为特征向量, p 为特征向量的维度; y_i 为特征向量所对应的标签。

NCA的具体实现步骤如下:

步骤1:设 w_r 为特征的距离权重,定义距离函数 d_w

$$d_w(x_i, x_j) = \sum_{r=1}^p w_r^2 |x_{ir} - x_{jr}| \quad (2)$$

步骤2:设 k 为相似函数,定义样本 x 的参考样本为 x_j 的概率为

$$P(\text{Ref}(x) = x_j | S) = \frac{k(d_w(x, x_j))}{\sum_{j=1}^n k(d_w(x, x_j))} \quad (3)$$

这里NCA假设了 x 与 x_j 的距离较小时, x_j 较大的概率作为 x 的参考样本。当 x_j 是 x 的参考样本时,使用 x_j 的标签 y_j 作为 x 的标签。

步骤3:去除样本 (x_i, y_i) 后,计算 x_i 的参考样本为 x_j 的“去一概率”

$$p_{ij} = P(\text{Ref}(x) = x_j | S^{-i}) = \frac{k(d_w(x, x_j))}{\sum_{j=1, j \neq i}^n k(d_w(x, x_j))} \quad (4)$$

步骤4:计算参考样本 x_j 的标签与 x_i 标签相同情况下的概率

$$\begin{cases} p_i = \sum_{j=1, j \neq i}^n P(\text{Ref}(x_i) = x_j | S^{-i}) I(y_i = y_j) = \sum_{j=1, j \neq i}^n p_{ij} y_{ij} \\ y_{ij} = I(y_i = y_j) = \begin{cases} 1 & \text{if } y_i = y_j, \\ 0 & \text{otherwise.} \end{cases} \end{cases} \quad (5)$$

显然当 p_i 的值越大时,说明样本之间的距离越小的情况,有更大的概率作为参考样本,即有更大的概率是同一类。

步骤5:计算所有样本的“去一概率”的平均值

$$F(w) = \frac{1}{n} \sum_{i=1}^n p_i \quad (6)$$

同时通过加入 λ 组成的正则项构造目标函数

$$J = -(F(w) - \lambda \sum_{r=1}^p w_r^2) = -(\frac{1}{n} \sum_{i=1}^n p_i - \lambda \sum_{r=1}^p w_r^2) \quad (7)$$

构造目标函数后,将特征选择的问题转换为求最小值问题

$$w_r^* = \underset{w_r}{\operatorname{argmin}} J(w_r) \quad (8)$$

根据目标函数选择合适的正则系数得到各特征的权重值,筛选部分权重最大的特征作为最终特征。

2 QAR数据处理及特征提取

对原始QAR的处理过程主要分为数据预处理和特征提取2部分,具体流程如图1所示。

2.1 QAR数据预处理

某航空公司提供的4台CFM56-7B发动机的

QAR数据主要包括发动机、飞行控制、液压控制以及环境控制等4个系统的重要参数,如:计算空速 C_{AS} 、马赫数 Ma 、飞行高度(Altitude, ALT)、燃油流量 F_F 、低压涡轮转速 N_1 、高压涡轮转速 N_2 、排气温度 E_{CT} 等。

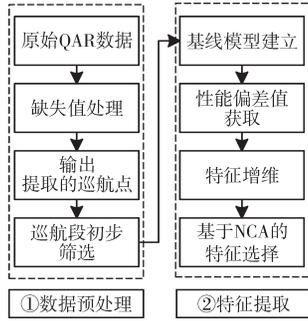


图1 原始QAR数据处理流程

一方面,由于不同传感器采样频率不同导致传感器信号周期性缺失,因此根据发动机QAR数据的缺失值与附近值相似的特性,采用“最近邻填补”方法对所有的缺失值进行填补;另一方面,根据 C_{AS} 、 Ma 、 F_F 等参数按初步提取飞机巡航段规则(见表1)初步提取飞机巡航段数据,如图2中步骤1所示。

表1 初步提取飞机巡航段规则

	C_{AS}/kn	Ma	$F_F/(kg/s)$	$N_1/\%$	$N_2/\%$
筛选条件	[45, 500]	[0.44, 0.9]	[0, +∞]	[75, 95]	[75, 95]

步骤2,设置ALT的偏差不得超过10feet提取到的巡航段数据如图2红色线段所示。依次判断所提取到的巡航段数据的跨度,

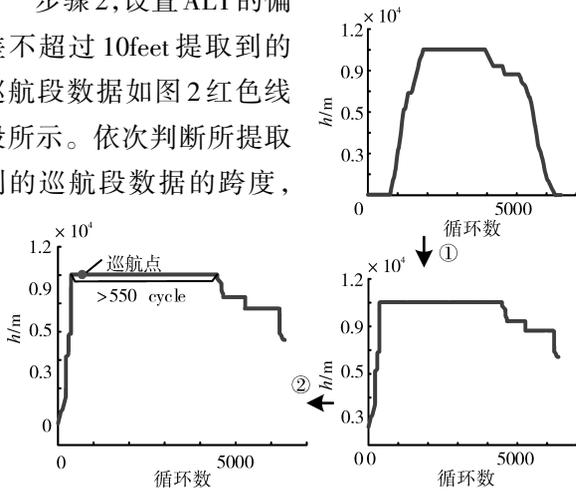


图2 数据预处理

选取跨度超过500cycle的第1个巡航段数据,随机取其中1个点的的数据作为巡航点数据(图2中绿点)。

2.2 特征提取

传统的民航飞机通常是使用 ΔE_{CT} 、 ΔN_2 和 ΔF_F 这3个性能参数偏差作为特征,结合相应的诊断算法来进行发动机的故障诊断。由于实际的发动机故障种类较多,仅仅使用以上3个故障特征无法进行准确的故障诊断。因此,本文通过增加故障特征与故障特征优化的方法提高发动机故障诊断的精度。

根据航空公司提供的维修记录,选取水洗之后的一段时间内的QAR数据用于建立 E_{CT} 、 F_F 、 N_2 、 T_3 、 T_{25} 这5个参数的基线模型^[17-18]。本文采用多元线性的方法进行基线建模,选取 N_1 、 T_{AT} 、 Ma 和 A_{LT} 作为自变量。在获取基线方程后,通过将测量值减去基线值得到性能参数偏差值。

特征增维是在已有特征的基础上,以提高故障诊断精度为目的增加故障特征的方法。对已有的故障特征做4种类型的特征增维,包括指数特征增维、对数特征增维、组合特征增维以及差分特征增维,具体的故障特征增维方法见表2,共有31个增维特征。

表2 故障特征增维方法

特征类型	增维特征
指数特征	ΔN_2^{-1} 、 ΔF_F^{-1} 、 ΔE_{CT}^{-1} 、 ΔT_3^{-1} 、 ΔT_{25}^{-1} 、 ΔN_2^{-2} 、 ΔF_F^{-2} 、 ΔE_{CT}^{-2} 、 ΔT_3^{-2} 、 ΔT_{25}^{-2} 、 ΔN_2^2 、 ΔF_F^2 、 ΔE_{CT}^2 、 ΔT_3^2 、 ΔT_{25}^2
对数特征	$\log(\Delta N_2 + C_2)$ 、 $\log(\Delta F_F + C_1)$ 、 $\log(\Delta E_{CT} + C_3)$ 、 $\log(\Delta T_3 + C_1)$ 、 $\log(\Delta T_{25} + C_2)$
差分特征	$\text{Diff}(\Delta E_{CT})$ 、 $\text{Diff}(\Delta N_2)$ 、 $\text{Diff}(\Delta F_F)$ 、 $\text{Diff}(\Delta T_3)$ 、 $\text{Diff}(\Delta T_{25})$
组合特征	$\Delta F_F/\Delta N_2$ 、 $\Delta E_{CT}/\Delta F_F$ 、 $\Delta E_{CT}/\Delta N_2$ 、 $\Delta F_F/\Delta T_3$ 、 $\Delta F_F/\Delta T_{25}$ 、 $\Delta E_{CT}/\Delta T_{25}$

按照表中列出的先后顺序对31个特征编号,再采用近邻主成分分析的方法对特征进行特征优化,从中选取出最优的特征组合。构造目标函数与正则系数 $\lambda = \text{Lambda}$ 关系,正则系数选取如图3所示。

从图中可见,当选取正则系数 $\lambda^* = 8.0091 \times 10^{-5}$ 时(图中实心点),目标函数值最低,此时各特征的权重如图4所示。

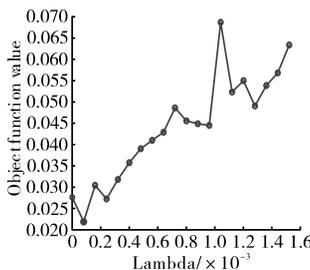


图3 正则系数选取

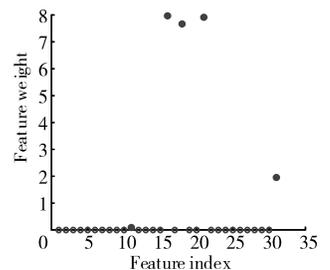


图4 不同特征的特征权重

从图4中选取权重最大的5个特征(图中实心点),分别为第11个特征 ΔN_{22} 、第16个特征 $\log(\Delta N_2 + C_2)$ 、第18个特征 $\log(\Delta E_{CT} + C_3)$ 、第21个特征 $\text{diff}(\Delta E_{CT})$ 和第31个特征 $\Delta E_{CT}/\Delta T_{25}$ 。结合5个原始性能参数偏

差作为最终特征。NCA 特征保留结果见表 3。

表 3 NCA 特征保留结果

原始特征	最终保留
ΔE_{CT} 、 ΔF_F 、 ΔN_2 、 ΔT_3 、 ΔT_{25}	ΔE_{CT} 、 ΔF_F 、 ΔN_2 、 ΔT_3 、 ΔT_{25} 、 ΔN_{22} 、diff (ΔE_{CT}) $\log(\Delta N_2+C_2)$ 、 $\log(\Delta E_{CT}+C_3)$ 、 $\Delta E_{CT}/\Delta T_{25}$

3 结果与讨论

在特征增维与优化的基础上,利用常见的几种分类算法实现故障诊断,并对比了特征优化前后对分类算法结果的影响。

3.1 样本提取与处理

由于故障样本有限,本文当前涉及 3 种常见的故障类型,包括发动机叶片积垢、鸟撞、VSV 作动筒工作故障。根据维修记录,选择维修当天航段及前 20 个航段构造故障样本数据集,如图 5 所示。

所得样本数量分布见表 4。

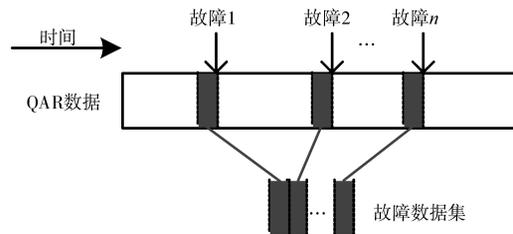


图 5 故障样本数据集构造

表 4 样本数量分布

无故障样本健康	叶片积垢故障 1	鸟撞故障 2	VSV 作动筒工作故障 3
3426	290	45	33

从表中可见,故障样本的数量远小于健康样本的数量,由于正负样本数量不均衡的现象会导致算法倾向于诊断出拥有较多样本的状态结果,且影响模型的泛化性与准确性,因此采用重复采样的方法对少数的样本进行扩充,从而达到样本均衡的目的,并采用 Z-score 方法对均衡后的样本进行归一化处理。

3.2 近邻成分分析算法验证

利用高斯朴素贝叶斯、二次判别、提升树、高斯支持向量机、共 4 种算法验证该特征提取方法对分类结果的提升程度。为防止模型产生过拟合,采用 5 折交叉验证的方法对模型进行验证。分别对比仅使用 5 个原始特征、使用原始特征加所有增维特征共 36 个特征、使用原始特征加 NCA 算法提取后的增维特征共 10

个特征用于训练和构造模型的结果,如图 6 所示。

从图中可见,相比只利用 5 个维度的特征去训练模型,5 个原始特征加 31 个增维特征对分类算法准确率提升效果不明显,特别在二次判别算法

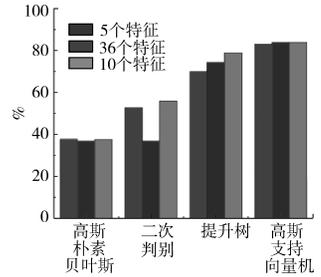


图 6 选择不同维度的特征分类结果

上算法准确率反而降低,而 5 个原始特征加上 NCA 算法提取的 5 个增维特征总体上对分类算法的准确率有明显提高,在二次判别和提升树算法上尤为明显,特别是在提升树算法准确率的提升达到 15.1%。

利用不同维度的特征训练和测试算法的运行时间见表 5。从表中可见,采用特征增维方法增加了特征的维度后,算法的训练时间相应延长,在大部分情况下利用 36 个特征构建算法模型所花费的运行时间最长,而采用 NCA 算法降低增维特征的维度后,可以显著缩短算法的运行时间。

表 5 利用不同维度的特征训练和测试算法的运行 s

特征数	高斯朴素贝叶斯	二次判别	提升树	高斯支持向量机
5	0.6281	0.52326	4.4182	20.882
36	1.2704	1.2746	9.2605	28.014
10	1.3746	0.95291	6.2385	23.811

通过特征增维方法构建新的特征,再利用近邻成分分析算法从中提取最优特征的方法,一方面,相比仅使用原始特征构建模型,能明显提高模型的准确率;另一方面,相比只使用特征增维的方法,运算量大减小,大幅缩短了运行时间,且具有更高的准确率。

4 结论

(1)对于数据驱动的诊断方法,不同的算法对于诊断准确率影响较大,诊断准确率最高相差达 40%,其中基于高斯核支持向量机算法的诊断准确率达到 83%;

(2)采用 ΔE_{CT} 、 ΔN_2 、 ΔF_F 、 ΔT_3 、 ΔT_{25} 参数偏移量增维特征进行故障诊断能够显著提升诊断准确率,相较于 5 个特征参数偏移量的诊断精度最高提升了 15.1%;

(3)采用近邻成分分析算法通过主元分析实现故障特征参数向量的降维和优化,有利于提高诊断准确

率和计算效率。

参考文献:

- [1] 郑波,马昕.改进的Kohonen网络在航空发动机分类故障诊断中的应用[J].航空发动机,2020,46(2):23-29.
ZHENG Bo, MA Xin. Application of improved kohonen network in aeroengine classification fault diagnosis[J]. Aeroengine, 2020, 46(2): 23-29. (in Chinese)
- [2] 王翔宇.民用航空发动机维修市场特点探析[J].航空动力,2018(3):59-62.
WANG Xiangyu. Analysis of commercial aero engine MRO market[J]. Aerospace Power, 2018(3):59-62. (in Chinese)
- [3] Gao X, Cheng Z, Huo W. Anomaly location method for QAR data based on principal component analysis hierarchical clustering[J]. IOP Conference Series Materials Science and Engineering, 2020, 790(3): 012085.
- [4] 曹惠玲,高扬. QAR数据的认识及民航应用研究[J].机械工程与自动化,2018,10(1):24-26.
CAO Huiling, GAO Yang. Research on QAR data knowledge and application in civil aviation field[J]. Mechanical Engineering & Automation, 2018, 10(1):24-26. (in Chinese)
- [5] Lei W, Chang X W, Rui S S. An analysis of flight Quick Access Recorder (QAR) data and its applications in preventing landing incidents [J]. Reliability Engineering and System Safety, 2014, 127(5):11-12.
- [6] 徐萌,席泽西,王雍赞,等.基于集成学习的航空发动机故障诊断方法[J].中国民航大学学报,2019,37(2):29-33,42.
XU Meng, XI Zexi, WANG Yongyun, et al. Aeroengine fault diagnosis based on ensemble learning algorithm[J]. Journal of Civil Aviation University of China, 2019, 37(2):29-33, 42. (in Chinese)
- [7] 顾彬,王建东.一种新颖的QAR数据特征提取方法[J].四川大学学报(工程科学版),2011,43(3):113-117.
GU Bin, WANG Jiandong. A novel feature extraction method for QAR data[J]. Journal of Sichuan university (Engineering Science Edition), 2011, 43(3):113-117. (in Chinese)
- [8] 戴婧睿,吴奇,仁和,等.基于深度置信网络的QAR飞行数据特征提取分析[J].电光与控制,2017(11):82-86.
DAI Jingrui, WU Qi, REN He, et al. DBN based feature extraction for flight data of quick access recorder[J]. Electronics Optics & Control, 2017(11):82-86. (in Chinese)
- [9] 张鹏,杨涛,刘亚楠,等.基于CNN-LSTM的QAR数据特征提取与预测[J].计算机应用研究,2019(10):35.
ZHANG Peng, YANG Tao, LIU Yanan, et al. Feature extraction and prediction of QAR data based on CNN-LSTM[J]. Application Research of Computers, 2019(10):35. (in Chinese)
- [10] 蒋丽英,栗文龙,崔建国,等.基于PCA与DBN的航空发动机气路系统故障诊断[J].沈阳航空航天大学学报,2019(1):57-62.
JIANG Liying, LI Wenlong, CUI Jianguo, et al. Fault diagnosis of aeroengine gas path system based on PCA and DBN[J]. Journal of Shenyang Aerospace University, 2019(1):57-62. (in Chinese)
- [11] 崔建国,严雪,蒲雪萍,等.基于动态PCA与改进SVM的航空发动机故障诊断[J].振动、测试与诊断,2015,35(1):94-99.
CUI Jianguo, YAN Xue, PU Xueping, et al. Aeroengine fault diagnosis based on dynamic PCA and improved SVM[J]. Zhendong Ceshi Yu Zhenduan/Journal of Vibration, Measurement and Diagnosis, 2015, 35(1):94-99. (in Chinese)
- [12] 曹惠玲,高升,薛鹏.基于多分类AdaBoost的航空发动机故障诊断[J].北京航空航天大学学报,2018,44(9):1818-1825.
CAO Huiling, GAO Sheng, XUE Peng. Aeroengine fault diagnosis based on multi-classification AdaBoost[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(9):1818-1825. (in Chinese)
- [13] Qin C, Song S, Huang G, et al. Unsupervised neighborhood component analysis for clustering[J]. Neurocomputing, 2015, 168(30):609-617.
- [14] 赵咏旺,刘毅慧,黄伟.基于近邻成分分析算法的原发性肝癌精确放疗后HBV再激活分类预测[J].生物信息学,2018,16(3):163-169.
ZHAO Yongwang, LIU Yihui, HUANG Wei. Classification and prediction of HBV reactivation after radiotherapy of primary liver cancer based on neighborhood component analysis[J]. Chinese Journal of Bioinformatics, 2018, 16(3):163-169. (in Chinese)
- [15] Liu J, Chen S. Discriminant common vectors versus neighbourhood components analysis and laplacianfaces: a comparative study in small sample size problem[J]. Image & Vision Computing, 2006, 24(3):249-262.
- [16] Zhou H T, Chen J, Dong G M, et al. Bearing fault recognition method based on neighbourhood component analysis and coupled hidden Markov model[J]. Mechanical Systems and Signal Processing, 2016(66-67):568-581.
- [17] 王奕首,余映红,卿新林,等.基于KPCA和DBN的航空发动机排气温度基线建模研究[J].航空发动机,2020,46(1):54-62.
WANG Yishou, YU Yinghong, QING Xinlin, et al. Exhaust gas temperature baseline model of aeroengine based on kernel principal component analysis and deep belief network[J]. Aeroengine, 2020, 46(1): 54-62. (in Chinese)
- [18] 黄曦,卿新林,王奕首,等.基于SDAE的航空发动机燃油流量基线模型构建[J].航空发动机,2019,45(2):29-33.
HUANG Xi, QING Xinlin, WANG Yishou, et al. Construction of aeroengine fuel flow baseline model based on stacked denoising autoencoders[J]. Aeroengine, 2019, 45(2):29-33. (in Chinese)

(编辑:程海)