文章编号：1006-1355(2019)03-0001-06

一种基于流形学习和KNN算法的柴油机工况识别方法

江志农[1]，赵南洋[1]，夏 敏[2]，赵飞松[2]，高佳丽[2]，张进杰[3]

（1. 北京化工大学 高端机械装备健康监控与自愈化北京市重点实验室，北京 100029；
2. 中石化重庆天然气管道有限责任公司，重庆 408000；
3. 北京化工大学 压缩机技术国家重点实验室压缩机健康智能监控中心，北京 100029）

摘 要：不同负荷状态下的柴油机振动、温度、转速等信号显著不同，而机组故障信号特征往往被淹没在随负荷变化而剧烈变化的信号中，因此变负荷状态下的柴油机故障监测诊断难度较大，一直困扰着柴油机的实际故障诊断工作。提出一种基于流形学习和KNN算法的柴油机工况识别方法，为柴油机变负荷工况下故障监测预警打下基础。方法融合机组的多源信号特征构建特征向量，通过流形学习t-分布邻域嵌入算法（t-SNE）实现特征向量的维数约简和敏感特征提取，采用K最近邻分类算法（KNN）完成柴油机运行负荷状态的自动分类。正常及故障状态下多组柴油机监测数据的处理结果验证了方法的有效性和实用性。

关键词：振动与波；柴油机；变负荷；流形学习；KNN；敏感特征

中图分类号：TK42；TH165$^+$.3；X839 **文献标志码**：A **DOI编码**：10.3969/j.issn.1006-1355.2019.03.001

Operating Mode Identification Method for Diesel Engines based on Manifold Learning and KNN Algorithm

JIANG Zhinong[1], *ZHAO Nanyang*[1], *XIA Min*[2], *ZHAO Feisong*[2], *GAO Jiali*[2], *ZHANG Jinjie*[3]

(1. Beijing Key Laboratory of Health Monitoring Control and Fault Self-recovery for High-end Machinery, Beijing University of Chemical Technology, Beijing, 100029, China;
2. Sinopec Chongqing Natural Gas Pipeline Co., Ltd., Chongqing 408000, China;
3. Compressor Health and Intelligent Monitoring Center of National Key Laboratory of Compressor Technology, Beijing University of Chemical Technology, Beijing 100029, China)

Abstract : Under different load conditions, the signals of vibration, temperature, speed, etc., of diesel engines are significantly different, and the fault signal characteristics of the engine unit are often submerged in signals that change drastically with the load change. Therefore, the diesel engine's fault monitoring and diagnosing under variable load conditions is difficult and the drastically changed signals always troubles the actual fault diagnosis. This paper presents an operating mode identification method for diesel engines based on manifold learning and KNN algorithm, which lays a foundation for fault monitoring and early warning of diesel engines under variable load conditions. The method combines the multi-source signal features of the unit to construct the feature vector. The feature reduction and sensitive feature extraction of the feature vector is achieved through the manifold learning t-distributed Stochastic Neighbor Embedding (t-SNE) algorithm. The K-Nearest Neighbor (KNN) classification algorithm is used to complete the automatic classification of diesel engine's operating load status. The diesel engine signal under normal and fault conditions verifies the effectiveness and practicality of this method.

Keywords : vibration and wave; diesel engine; variable load; manifold learning; KNN; sensitive feature

收稿日期：2018-06-11

基金项目：国家"863"计划资助项目（2014AA041806）；国家重点研发计划资助项目（2016YFF0203305）；中央高校基本科研业务费专项资金资助项目（JD1815）；双一流建设专项经费资助项目（ZD1601）。

作者简介：江志农（1967–），男，辽宁省丹东市人，研究员，主要研究方向为机械故障监测诊断技术，故障自愈调控技术。

通信作者：张进杰（1987–），男，安徽省马鞍山市人，副教授，主要研究方向为往复机械故障监测诊断技术、节能降耗技术。

E-mail: zjj87427@163.com

柴油机作为一种典型的内燃机械，在石油钻井、舰船动力、应急发电、工程机械等等工业领域应用十分广泛。但由于柴油机机械结构复杂、易损件多、工作条件恶劣，实际运行中故障频发，如失火、气门磨损、拉缸、连杆断裂等，易造成严重的经济损失并影响安全稳定生产。因此，柴油机故障监测诊断技术一直以来都是研究热点。在离线测试方面，可通过定期采集油品，进行光谱、铁谱分析，对润滑油内的各类元素含量进行测试[1-2]。应用振动分析方法、噪声分析方法进行柴油机故障诊断已有一定的研究，梅检民等提出了一种基于高阶累积量的振动信号分析方法，采用高阶累积量的AR双谱，对连杆轴承故障加速振动信号进行分析[3]，沈虹等提出了一种基于角域四阶累积量切片谱的柴油机连杆轴承故障特征提取方法[4]，程利军等提出了一种基于阶比跟踪及共振解调的连杆轴承故障诊断方法[5]。

但是，由于柴油机工况复杂，如功率、转速及其他工作环境的变化，导致机组监测信号多变，给故障监测诊断带来了较大的难度。例如，柴油机故障监测系统的报警线设置，若按照单一报警线设置，在变负荷工况下将会产生较多的误报、漏报；若根据机组运行状态设置，则需要对柴油机运行状态数据开展深入的分析研究，以机组运行状态的自动识别为基础。

因此，本文提出了一种基于流形学习和KNN算法的柴油机工况识别方法，可实现对柴油机运行状态的自动分类。方法融合了机组的振动、温度、压力等多源信号特征构建特征向量；通过流形学习t-分布邻域嵌入算法（t-SNE）实现特征向量的维数约简和敏感特征提取，降低数据分析处理难度；最后通过K最近邻分类算法（KNN）构建了自动分类模型，完成柴油机运行负荷状态的自动分类。

1 流形学习与KNN算法

1.1 流形学习方法

流形学习（Maniflod Learning）是一种非线性降维方法[6]，首次由Bregler和Omohundro提出，假设数据是均匀采样于一个高维欧氏空间，其采样点的分布具有低维流形的结构特征，流形学习就是从高维的采样数据中提取出低维流形结构，并求出相应的嵌入映射，从而实现数据特征维数的约简及数据可视化[7]。

目前，其算法包括t-分布邻域嵌入（t-distributed Stochastic Neighbor Embedding，t-SNE）[8]、等距映射（Isometric Feature Mapping，Isomap）[9-10]、局部线性嵌入（Local Linear Embedding，LLE）[11]、随机邻域嵌入（Stochastic Neighbor Embedding，SNE）[12]等。其中t-SNE算法是一种利用概率形式的降维算法，来缓解在低维空间中映射数据点之间拥挤问题。t-SNE算法用*KL*散度（Kullback–Leibler divergence）来度量两个分布之间的相似性，使数据在低维空间上有聚类的效果，且不同类别的数据簇之间存在一定的距离，这对使用分类算法来实现柴油机运行工况识别提供了良好条件，因此本文选用t-SNE算法来对柴油机运行工况特征向量进行降维及可视化处理。

t-SNE算法的步骤如下：

（1）原始数据，记为：$X=\{x_1,x_2,\cdots,x_n\}$；

（2）设置迭代次数T，学习速率η，动量$\alpha(t)$；

（3）计算成本函数参数（perplexity Perp）

$$P_{j|i}=\frac{\exp\left(-\left\|x_i-x_j\right\|^2/2\sigma_i^2\right)}{\sum_{k\neq i}\exp\left(-\left\|x_i-x_k\right\|^2/2\sigma_i^2\right)} \tag{1}$$

上式中σ_i是数据点x_i的高斯方差；

（4）设置$P_{ij}=\frac{P_{j|i}+P_{i|j}}{2n}$，用$N\left(0,10^{-4}I\right)$随机初始化$y^{(0)}=\{y_1,y_2,\cdots,y_n\}$，其中$y$为初始化向量；

（5）计算低维相似度q_{ij}

$$q_{ij}=\frac{\left(1+\left\|y_i-y_j\right\|^2\right)^{-1}}{\sum_{k\neq l}\left(1+\left\|y_k-y_l\right\|^2\right)^{-1}} \tag{2}$$

（6）计算*KL*散度（Kullback-Leibler divergence）

$$C=KL\left(P\|Q\right)=\sum_i\sum_j p_{ij}\log\frac{p_{ij}}{q_{ij}} \tag{3}$$

上式中*KL*散度是Q和P之间的相似度的度量；

（7）计算梯度最小化*KL*散度

$$\frac{\delta C}{\delta y_i}=4\sum_j\left(p_{ij}-q_{ij}\right)\left(y_i-y_j\right)\left(1+\left\|y_i-y_j\right\|^2\right)^{-1} \tag{4}$$

（8）更新

$$y^{(t)}=y^{(t-1)}+\eta\frac{\delta C}{\delta y}+\alpha(t)\left(y^{(t-1)}-y^{(t-2)}\right) \tag{5}$$

（9）循环迭代步骤（5）—步骤（8），迭代T次；

（10）最终得到低维数据

$$y^{(T)}=\{y_1,y_2,\cdots,y_n\} \tag{6}$$

1.2 KNN算法

*K*最近邻分类算法（KNN）被广泛应用于数据挖掘和数据分类。如果待分类样本的样本数据点在空间中有*K*个距离最近样本数据点，其中某一类别的数据点占绝大多数，则KNN算法判定该样本数据点属于这个类别。

KNN算法在待识别数据点与样本数据点之间的距离度量主要有欧式距离，曼哈顿距离和闵科夫斯基距离，其中闵科夫斯基距离定义如下

$$\mathrm{Dis}_p(x,y)=\left(\sum_{j=1}^{d}|x_j-y_j|^p\right)^{1/p}=\|x-y\|_p \quad (7)$$

其中：d为维数，p为范数。当$p=1$时为曼哈顿距离；当$p=2$时为欧式距离。

本文选用欧氏距离作为KNN算法待识别数据点与样本数据点之间的空间距离的度量。欧式距离定义如下

$$\mathrm{Dis}(x,y)=\sqrt{\sum_{j=1}^{d}(x_j-y_j)^2} \quad (8)$$

2　柴机油运行负荷状态识别方法

2.1　方法处理流程

本方法的处理流程图如图1所示。

首先根据柴油机监测信号进行多信号工况敏感特征参数提取，形成数据样本、测试样本和待识别工况数据，一起进行流行学习t-SNE算法降维。设定K值区间，使用降维后的样本数据和测试数据对KNN工况识别模型进行K值优化，确定识别率最高的K值的KNN模型进行工况识别结果输出。

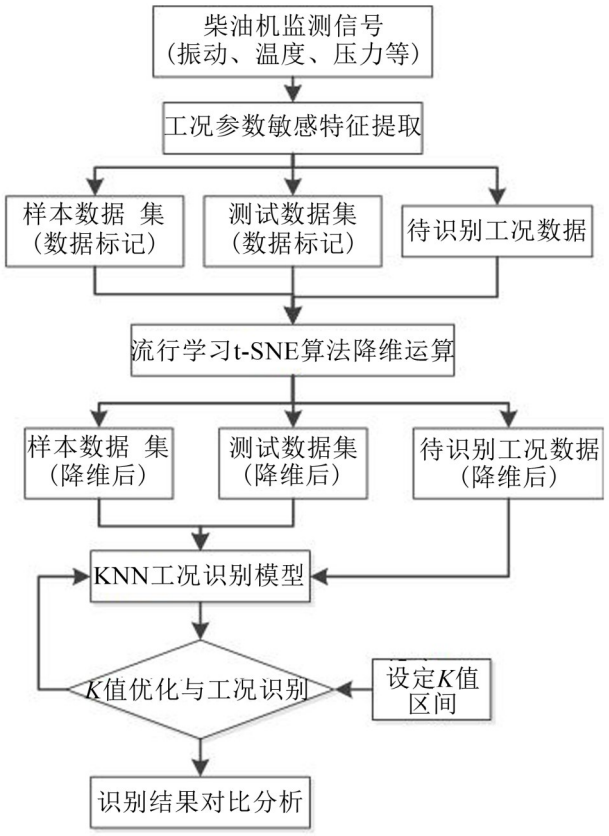


图1　方法处理流程图

2.2　柴油机实验台应用

（1）实验台机组介绍

实验台柴油机为TBD234V12柴油机，机组如图2所示，其主要技术参数如表1所示。

表1　TBD234V12缸柴油机主要技术参数

气缸数量	压缩比	工作转速/(r·min^{-1})	气缸点火顺序
12	15	0～2 300	B1-A1-B5-A5-B3-A3-B6-A6-B2-A2-B4-A4

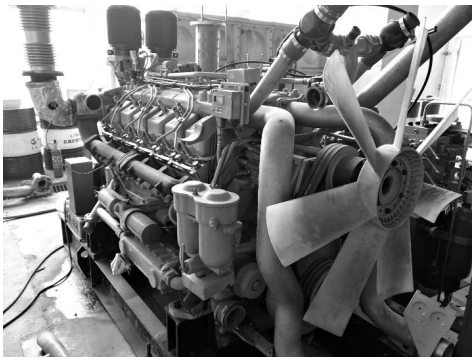


图2　实验台TBD234V12型柴油机

该机组已自带部分温度、压力传感器，可实现对总排气温度、冷却水温度、润滑油压力等参数的在线监测。同时，在该机组上新增了振动、瞬时转速、活塞上止点信号传感器，对缸盖振动、曲轴箱振动、齿轮箱振动、涡轮增压器振动、瞬时转速、活塞上止点信号进行在线监测。从上述测点信号中，构建了如表2所示的信号特征集。

表2　多源信号监测参数特征集

序号	测点类型	信号特征参数
1	缸盖振动测点	振动速度有效值
2		点火角域振动速度有效值
3		振动背景噪声有效值
4	曲轴箱振动测点	振动速度有效值
5	瞬时转速测点	瞬时转速峰峰值
6	排气温度测点	排气温度峰值
7	齿轮箱振动测点	振动速度有效值
8	增压器振动测点	振动速度有效值
9	润滑油测点	润滑油压力值
10		润滑油温度值
11	冷却水温测点	冷却水温度值
12	机组转速测点	机组平均转速值

（2）正常状态下工况分类

对机组正常状态下不同负荷、不同转速状态的数据进行测试应用，共计12种工况。

Ø 转速1 500 r/min，扭矩700 Nm，1 000 Nm，1 300 Nm。

Ø 转速1 800 r/min，扭矩700 Nm，1 000 Nm，1 300 Nm，1 600 Nm。

Ø 转速2 100 r/min，扭矩700 Nm，1 000 Nm，1 300 Nm，1 600 Nm，2 200 Nm。

分类结果如下图4－图6所示。

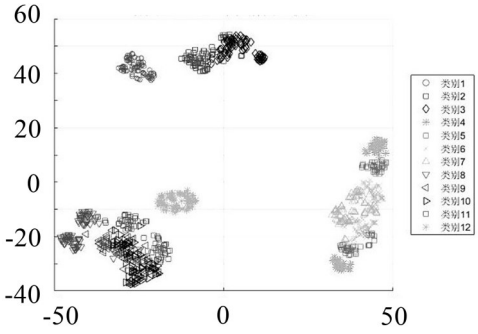


图4 t-SNE算法二维映射降维结果

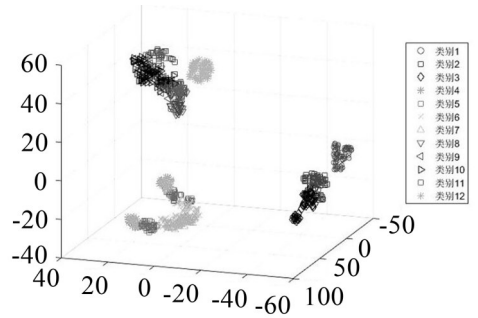


图5 t-SNE算法三维映射降维结果

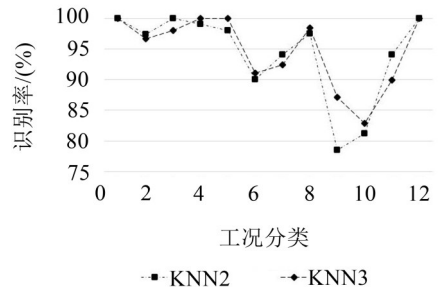


图6 KNN算法识别结果

（横坐标为12种工况，纵坐标为识别率，KNN2代表二维KNN分类结果，KNN3代表三维KNN分类结果）

图4和图5分别为二维和三维映射下的12种工况数据的分布情况，3种转速下的数据在空间3个地方集聚，同转速不同负荷的数据在集聚区团聚分布。图5为最优K值下的KNN算法二维和三维各工况识别率图，由图显示，三维下的KNN识别率优于二维。

（3）故障状态下工况分类

由于柴油机机械结构复杂，激励众多，常常发生故障。在柴油机发生故障的情况下，其运行工况的识别准确率必然受到影响。本文针对柴油机常见的气门间隙故障，使用耦合了故障的监测数据来研究故障对柴机油运行工况识别结果的影响。

实验工况共计6种：

Ø 转速1 500 r/min，扭矩700Nm，1 300 Nm；

Ø 转速1 800 r/min，扭矩700 Nm，1 300 Nm；

Ø 转速2 100 r/min，扭矩700 Nm，1 300 Nm。

分类结果如下图7－图8所示。

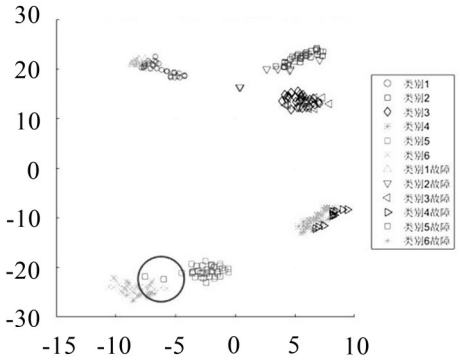


图7 t-SNE算法二维映射降维结果

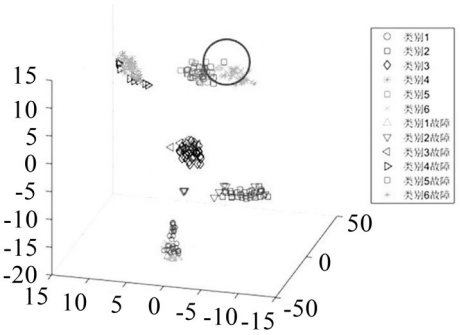


图8 t-SNE算法三维映射降维结果

表3 气门磨损状态下工况三维识别结果

工况类别	识别率/(%)
转速1 500 r/min，扭矩700 Nm	100
转速1 500 r/min，扭矩1 300 Nm	62.18
转速1 800 r/min，扭矩700 Nm	100
转速1 800 r/min，扭矩1 300 Nm	100
转速2 100 r/min，扭矩700 Nm	84.09
转速2 100 r/min，扭矩1 300 Nm	100

图7和图8分别为二维和三维映射下的6种工况故障耦合数据的分布情况，6种工况数据分别在空间集聚，部分故障数据点偏移数据集。由表3显示为气门磨损状态下工况三维识别正确率略有降低。

2.3 结果分析

实验台机组正常状态下的12种运行工况数据，分别使用t-SNE算法向二维和三维空间映射，采用KNN算法识别，二维和三维空间的工况数据均得到了较好的分类，且三维空间上KNN综合识别率优于二维。

气门磨损故障数据集的工况识别结果显示，部

分故障数据点偏移正常数据集，KNN识别率在工况转速1 500 r/min，扭矩1 300 Nm和工况转速2 100 r/min，扭矩700 Nm降低，但整体识别准度仍较好。

3　方法的进一步验证与对比

3.1　工业柴油机测试结果与分析

以某20缸大型柴油机监测数据为基础，利用5种负荷状态（0 %，50 %，85 %，100 %，110 %）监测数据进行方法验证，结果如下所示。

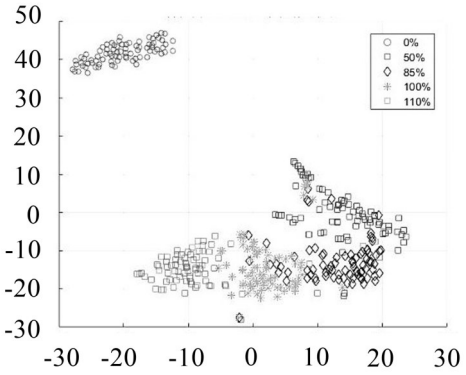


图9 t-SNE算法二维映射降维结果

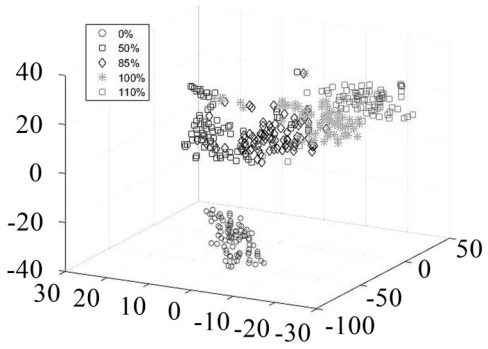


图10 t-SNE算法三维映射降维结果

图9和图10分别为某20缸柴油机五种负荷状态的特征向量样本和测试数据集经过流形学习t-SNE算法降维至二维和三维的结果，其中0 %负荷的数据点离其他四种负荷数据点较远。50 %，85 %，100 %，110 %的数据点集之间的界限明显。

之后使用KNN算法对测试数据点进行识别，将K值的区间设定为5～20，分别计算不同K值的KNN识别率。由表4和表5显示，二维情况下，K值为6时，各工况平均识别率94.15 %最高；三维情况下，K值为7时，各工况平均识别率94.16 %最高。因此，在二维时选择K值为6的KNN进行识别；在三维时选择K值为7的KNN进行识别。

3.2　与其他方法的对比与分析

为了进一步验证本文方法的有效性，利用20缸大型柴油机监测数据，对比应用了主成分分析（PCA）方法、BP神经网络方法处理上述数据，结果如下所示。

表4　二维映射下不同K值KNN识别率/(%)

K值/负荷	0 %	50 %	85 %	100 %	110 %
5	100.00	95.90	87.13	91.09	90.00
6	100.00	98.98	89.06	95.96	86.73
7	100.00	94.74	86.78	91.83	89.30
8	100.00	92.22	82.48	89.18	89.66
9	100.00	95.52	81.22	85.79	89.34
10	100.00	95.07	78.64	83.33	89.27
11	100.00	96.79	80.98	86.19	89.35
12	100.00	97.16	82.70	90.05	89.86
13	100.00	95.77	74.19	80.10	90.66
14	100.00	92.06	78.87	85.42	89.78
15	100.00	93.85	76.59	82.38	90.53
16	100.00	99.45	81.44	80.36	90.31
17	100.00	93.60	76.92	84.58	91.46
18	100.00	90.14	76.76	83.42	90.05
19	100.00	91.79	72.68	82.27	89.60
20	100.00	91.51	82.70	84.41	85.92

表5　三维映射下不同K值KNN识别率/(%)

K值/负荷	0 %	50 %	85 %	100 %	110 %
5	100.00	96.60	83.51	88.99	89.53
6	100.00	94.51	86.93	90.37	94.15
7	100.00	95.45	87.23	96.06	92.08
8	100.00	94.24	85.85	88.78	88.17
9	100.00	96.74	82.49	91.48	89.01
10	100.00	98.10	79.23	92.27	89.30
11	100.00	95.85	74.75	91.53	91.05
12	100.00	95.17	81.60	89.16	88.48
13	100.00	97.10	77.39	87.69	92.57
14	100.00	92.63	78.70	90.10	89.06
15	100.00	90.00	79.62	89.05	83.84
16	100.00	91.05	77.09	87.72	87.55
17	100.00	87.68	80.42	87.50	89.40
18	100.00	90.00	80.28	81.73	85.39
19	100.00	88.17	75.00	86.98	89.94
20	100.00	90.05	77.72	85.78	91.54

图10和图11分别为某20缸柴油机5种负荷状态的特征向量样本和测试数据集经过PCA算法降维至二维和三维的结果。数据经过PCA降维后，从数据分布看，数据混淆的程度比t-SNE严重。

图13为二、三维的BP和KNN算法的5种工况识别率比较图。

从上述结果中，可看出：

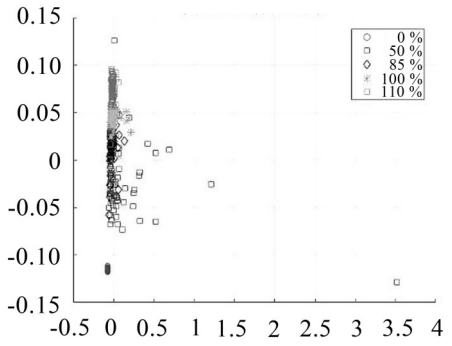


图 11 PCA算法二维映射降维结果

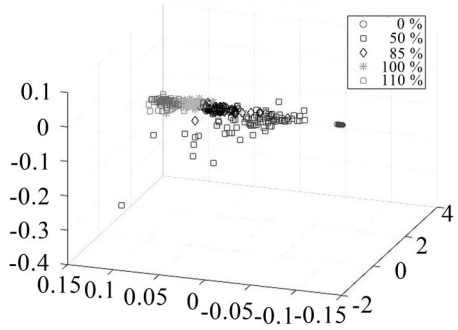


图 12 PCA算法三维映射降维结果

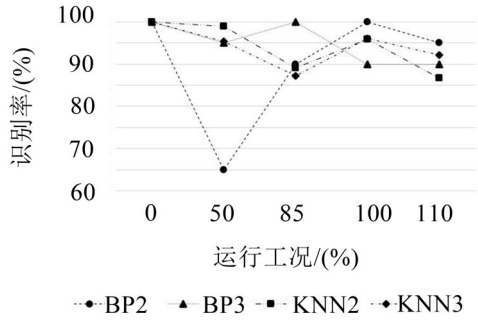


图 13 BP神经网络与KNN算法分类结果比较

（横坐标为运行工况，纵坐标为识别率，BP2为二维BP识别结果，BP3为三维BP识别结果，KNN2为二维KNN识别结果，KNN3为三维KNN识别结果）

（1）t -SNE算法比PCA方法在数据降维和聚类效果上具有优势；

（2）针对本文20缸柴油机数据样本，KNN算法对柴油机运行工况识别效果优于BP神经网络；

（3）数据特征向量降维至三维的工况识别率优于二维，因此合理选择降维维度对提高分类准确度有重要作用。

4 结 语

本文提出了一种基于流形学习和KNN算法的柴油机工况识别方法，提取了对工况敏感的特征参数，组成特征向量，使用流形学习t-SNE算法对特征向量数据降维，最后进行KNN算法工况识别。利用不同机组数据验证了方法的实用性；并针对气门故障条件下的柴油机运行状态识别开展了研究。

得到的结论如下：

（1）基于流行学习和KNN算法的柴油机工况识别方法对正常状态下的机组运行状态具有较高的识别准确率；

（2）柴机油故障对工况的识别产生了不利影响，降低了识别准确性，但是总体仍达到了较好的识别准确率。后续将针对故障特征与负荷状态特征进行深入研究，一方面提高柴油机工况识别准确度，另一方面为变负荷工况故障监测诊断奠定基础。

参考文献：

[1] 周平，刘东风，吴千. 柴油机油液监测综合评价方法研究[J]. **润滑油**，2010，25(02)：35-38.

[2] YADAV G, GANAI P, TIWARI S, et al. An investigation for prior failure of engine component through spectroscopy oil analysis program[J]. **Applied Mechanics & Materials**, 2014, 592-594(19): 1362-1365.

[3] 梅检民，张玲玲，肖云魁，等. 基于高阶累积量的轴承并发故障振动信号分析[J]. **内燃机学报**，2011，29(4)：327-331.

[4] 沈虹，赵红东，梅检民，等. 基于角域四阶累积量切片谱的柴油机连杆轴承故障特征提取[J]. **振动与冲击**，2014，33(11)：90-94.

[5] 程利军，张英堂，李志宁，等. 基于阶比跟踪及共振解调的连杆轴承故障诊断研究[J]. **内燃机工程**，2012，33(5)：67-73.

[6] 马发民，张林，王锦彪. 维数约简算法简述[J] . **软件工程**，2017，20(8)：7-13.

[7] BREGLER C, OMOHUNDRO S M. Nonlinear manifold learning for visual speech recognition[C]. //Int. Conf. Computer Vi-sion, 1995.

[8] VAN DER MAATEN L, HINTON G. Visualizing data using t-SN-E[J]. **Journal of Machine Learning Research**, 2008, 9(2579-2605): 85.

[9] VAN DER MAATEN L. Accelerating t-SNE using tree-based algorithms[J]. **Journal of Machine Learning Research**, 2014, 15: 1-21.

[10] SILVA V D, TENENBAUM J B. Global versus local method-s in nonlinear dimensionality reduction[C]. // Proc. of C-onference on Advances in Neural Information Processing Systems, 2003.

[11] ROWEIS S T, SAUL L K. Nonlinear dimensionality reduction by locally linear embedding[J]. **Science**, 2000, 290(5500): 2323-2326.

[12] HINTON G E, ROWEIS S T. Stochastic neighbor embedding[C]. //Advances in Neural Information Processing Systems, 2002: 833-840.