

基于 HFSGE 的轴承故障数据集降维方法^{*}

丁 华^{1,2}, 董晓鑫^{1,2}, 郑广禹^{1,2}, 薛 岩^{1,2}, 李亮亮^{1,2}

(1. 太原理工大学机械工程学院 太原, 030024)

(2. 煤矿综采装备山西省重点实验室 太原, 030024)

摘要 针对多源传感器监测构建的轴承故障数据集存在高维非线性特性、故障模式难以有效辨识的问题,提出一种基于超特征空间图嵌入(hyper feature space graph embedding,简称 HFSGE)的故障数据降维算法。首先,该算法在每个样本点的近邻域内确定特征空间的组合元素,构造样本点间的多元邻接关系;其次,在特征空间基础上引入超边模型构建超特征空间邻接图,保留数据分布的拓扑结构信息;然后,建立类内、类间超特征空间邻接嵌入图,利用最大化散度差值矩阵准则,在嵌入空间实现类内聚集、类间分散,降低后续故障分类决策难度;最后,基于帕德伯恩大学的轴承故障模拟实验对所提算法各项性能进行验证。结果表明:HFSGE算法可从低价值密度的海量监测数据中剔除冗余信息,并挖掘出有效的故障决策知识,为降低数据维度、提升辨识准确率提供理论参考。

关键词 故障诊断;降维;特征空间;超特征空间图嵌入
中图分类号 TP18;TH165

0 引 言

轴承是大型高端装备的关键零部件,对其实施必要的智能监测、诊断与运维,是在保障工业生产过程安全稳定的基础上实现智能高效生产目标的必要途径^[1]。如何从海量的监测数据中挖掘设备状态的表征知识,已成为传统制造业与新一代人工智能技术实现融合的关键^[2]。数据驱动的轴承智能故障诊断决策技术,可通过解析信号组成、提取轴承故障特征,挖掘出“隐喻”在大量监测数据背后的机械装备故障演化规律^[3]。设备无人化、智能化与集群化的发展趋势促使多源多通道的故障监测方式成为行业主流,但随之产生的数据“维数灾难”问题,给后续故障识别带来挑战。因此,利用有效的降维方法提取表征故障本质的关键特征,更利于提升智能决策技术的可靠性。

近年来,以局部保持投影(locality preserving projections,简称 LPP)^[4]、边缘 Fisher 分析(marginal Fisher analysis,简称 MFA)^[5]、双邻接判别投影(double graphs-based discriminant projections,简称 DGDGP)^[6]等为代表的流形学习算法通过建立嵌入在高维空间中的低维流形图,保持样本点在高维空

间中的邻接几何拓扑关系,可在降低故障数据维度的同时提取有利于判别的信息。该类方法也被归纳为图嵌入方法^[5]。

传统图嵌入框架是在节点之间建立的二元邻接图,可归属于二阶图范畴。然而,监测数据在高维空间中表现出的复杂非线性的多元结构往往被建立的普通邻接关系所忽视^[6]。针对上述问题,学者们提出了可表征高维复杂数据结构的高阶多元邻接嵌入图模型。例如,近邻特征空间嵌入^[7]、加权特征线嵌入^[8]、中值特征空间判别投影(median feature space discriminant projection,简称 MFSDP)^[9]等特征空间嵌入算法,以及以判别超-拉普拉斯投影(discriminant hyper-Laplacian projections,简称 DHLP)^[10]、局部与全局标准超图嵌入^[11]等为代表的超图嵌入方法。上述两类降维方法,以实现低维空间中敏感特征数据重构、增加模型鲁棒性与泛化能力为目标,通过构建融合领域知识与标签信息的高阶多元嵌入图结构模型,可为多分类数据的模式识别提供便利。然而,特征空间邻接图嵌入算法中近邻特征空间的建立过程较为复杂,如何建立简单高效的特征空间邻接图是亟待解决的问题^[9]。此外,超图嵌入算法模型精度直接取决于超边的选取质量,而超边的选

^{*} 国家自然科学基金面上资助项目(52174148);山西省科技重大专项计划“揭榜挂帅”资助项目(202101010101018);山西省科技合作交流专项资助项目(202104041101003);山西省科技创新人才团队资助项目(202204051001017);山西省重点研发计划资助项目(202202100401013)

收稿日期:2023-10-25;修回日期:2023-12-06

择与建立尚没有统一且明确的规则。

综上,为解决特征空间嵌入图建立困难的问题,以及提高超边选取质量,本研究拟将超图理论引入特征空间图嵌入过程,并通过构建类内、类间 2 个超特征空间嵌入图,以最大化散度差值矩阵^[9]的方式实现有效特征的提取。由此,本研究提出一种基于 HFSGE 的故障数据降维算法,为基于工业大数据资源的机械装备智能决策技术提供一种数据运算理论依据。

1 中值特征空间嵌入与超图嵌入相关原理

1.1 中值特征空间嵌入算法

特征空间嵌入算法 MFSDP 属于高阶图嵌入框架,该算法通过改进特征空间度量方式、加速近邻特征空间选择进程以及最大化散度差分矩阵等手段,设法寻找更有效的投影矩阵,以实现数据降维。该算法原理为:假设原始高维故障特征集为 $X=[x_1, x_2, \dots, x_n]$, $X \in \mathbb{R}^F$, 经过 MFSDP 的投影变换 $Y=A^T X$ 实现样本特征降维,降维后的故障特征集为 $Y=[y_1, y_2, \dots, y_n]$, $Y \in \mathbb{R}^f$, 其中 $f \leq F$ 。MFSDP 算法具体步骤如下:

1) 建立中值特征空间邻接图。特征空间的结构取决于空间参数 P , 当 $P=2$ 时,建立的是特征线,当 $P=3$ 时为特征面,当 $P \geq 4$ 时为特征体。以图 1 所示的中值特征线为例,首先,待测样本 x_i 依据近邻规则 $k=4$,划定近邻域范围为红色虚线以内;其次,将近邻域样本点 x_1, x_2, x_3, x_4 任意两点连接共组成 C_4^2 个特征线 FL;最后,建立样本点 x_i 与每个特征线中值点的连接关系,构建中值特征空间邻接图。

2) 类内、类间散度矩阵计算。类内散度矩阵 $S_w^{(MP)}$ 由 k_1 个距离最近同类中值特征空间计算得到,类间散度矩阵 $S_b^{(MP)}$ 由 k_2 个距离最近的异类中值特

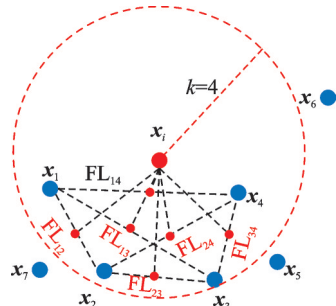


图 1 中值特征线

Fig.1 Median feature line

征空间计算得到,具体计算式为

$$S_w^{(MP)} = \sum_{i=1}^N \left\| \sum_{w=1}^{k_1} (y_i - f_w^{(MP)}(y_i)) \omega_w^{(MP)}(y_i) \right\|^2 \quad (1)$$

$$S_b^{(MP)} = \sum_{i=1}^N \left\| \sum_{b=1}^{k_2} (y_i - f_b^{(MP)}(y_i)) \omega_b^{(MP)}(y_i) \right\|^2 \quad (2)$$

3) 构建目标函数 J

$$\arg \max_A A^T (S_b^{(MP)} - S_w^{(MP)}) A$$

$$\text{s.t. } A^T A = I \quad (3)$$

利用拉格朗日乘子法,将式(3)转化为特征值求解问题,即

$$(S_b^{(MP)} - S_w^{(MP)}) A = \lambda A \quad (4)$$

对式(4)进行特征值分解,其中待求降维投影矩阵 A 由前 f 个最大的特征值对应的特征向量组成。

1.2 超图嵌入框架

超图嵌入框架是超图算法的核心,超图嵌入框架可包含多个顶点的超边,在嵌入空间中保留更多样本间的复杂多元结构信息。超图模型 G^H 为

$$G^H = \{X^H, E^H, W^H\} \quad (5)$$

其中: X^H 为顶点集; E^H 为超边集; W^H 为权重集,任一超边 $e_j \in E^H$, 则对应的权重为 $\omega(e_j) \in W^H$ 。

超边与顶点的关联矩阵、超边的度与顶点的度分别被定义为

$$H = \begin{cases} h(x_i, e_j) & (x_i \in e_j) \\ 0 & (x_i \notin e_j) \end{cases} \quad (6)$$

$$d(e_j) = \sum_{i=1}^n h(x_i, e_j) \quad (7)$$

$$d(x_i) = \sum_{j=1}^{n_e} \omega(e_j) h(x_i, e_j) \quad (8)$$

其中: $h(x_i, e_j)$ 表示顶点 x_i 与超边 e_j 之间的关系; e 为边。

超图嵌入如图 2 所示。图 2(a) 为超图模型的实例化展示,观察可知该超图模型由 8 个顶点和 5 条超边组成。上述信息可通过如图 2(b) 所示的关联矩阵 H 得以保留。

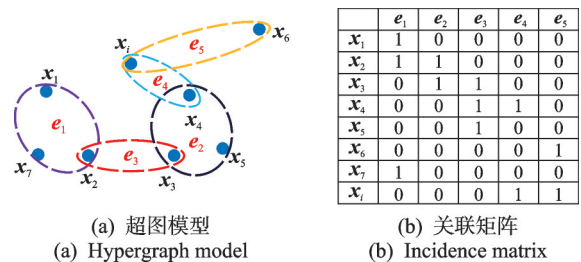


图 2 超图嵌入

Fig.2 Hypergraph embedding

2 HFSGE 算法

2.1 构建特征空间超边准则

中值特征空间算法具有明确的多元样本数目,可建立清晰的高阶嵌入结构。此外,还可以通过近邻域样本的重新分配组合,建立一种更具鲁棒性的多元关系^[9]。然而,特征空间的建立过程较为复杂。超图嵌入模型通过建立超边更好地保留多元关系,但该方法尚缺乏统一、固定的超边确立准则。

针对上述问题,提出一种基于特征空间的超边建立准则,具体为:首先,设定特征空间参数 P ,并根据1.1小节中步骤1划定每个待测样本点的近邻特征空间;其次,以每个近邻特征空间为单元,将其中的所有样本点与待测样本构建一条超边。超特征空间图如图3所示,即当 $P=2$ 时所建立的 $P+1=3$ 阶超图。

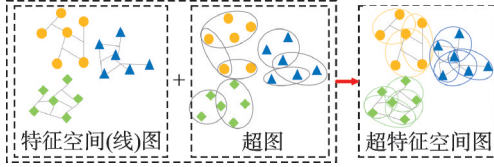


图3 超特征空间图

Fig.3 Hyper feature space graph

需要注意的是,由于特征空间超边的构建依赖于待测样本与近邻特征空间的组合关系,且相似样本更大概率存在于相同邻域空间中,因而可能存在重合的超边。因此,在建立特征空间超边集 E_{FS}^H 的过程中,可跳过包含相同元素的超边,减少计算空间与算力的浪费。

2.2 建立类内、类间散度矩阵

散度矩阵是度量样本团簇聚合分散程度的关键。在2.1小节中给出了近邻域中基于特征空间模型的超边确立法则。在此基础上,融合故障样本数据标签以及样本余弦相似度度量信息,得到的类间、类内超特征空间嵌入图如图4所示,并以此求解类内散度矩阵,实现同类数据的近邻域保持。构建异类数据的类间超特征空间图,依据超图建立类间散度矩阵,实现异类数据的近邻域破坏。

$$\sum_{e_j \in E_{FS}^{Hw}(x_i, x_k)} \sum_{e_j \in X_{FS}^{Hw}} \frac{\omega(e_j)_{FS}^w h(x_i, e_j)_{FS}^w h(x_k, e_j)_{FS}^w}{d(e_j)_{FS}^w} \|A^T x_i - A^T x_k\|^2 =$$

$$\begin{aligned} & \sum_j \sum_{i,k} \frac{\omega(e_j)_{FS}^w H_{FS}^w(i,j) H_{FS}^w(k,j)}{d(e_j)_{FS}^w} \text{tr}(A^T x_i x_i A^T - A^T x_k x_i A^T) = \left[\sum_j \sum_i \frac{\omega(e_j)_{FS}^w H_{FS}^w(i,j)}{d(x_i)_{FS}^w} \text{tr}(A^T x_i x_i A^T) \times \right. \\ & \left. \sum_k \frac{H_{FS}^w(k,j) d(x_k)_{FS}^w}{d(x_k)_{FS}^w} \right] - \left[\sum_j \sum_{i,k} \frac{\text{tr}(A^T x_k \omega(e_j)_{FS}^w H_{FS}^w(i,j) H_{FS}^w(k,j) x_i^T A)}{d(x_k)_{FS}^w} \right] = [\text{tr}(A^T X D(X)_{FS}^w X^T A)] - \\ & [\text{tr}(A^T X H_{FS}^w W_{FS}^w D(X)_{FS}^w - H_{FS}^w W_{FS}^w D(E)_{FS}^w - H_{FS}^w T^T X^T A)] = \text{tr}\{A^T X [D(X)_{FS}^w - H_{FS}^w W_{FS}^w D(E)_{FS}^w - H_{FS}^w T^T X^T A]\} = \\ & \text{tr}(A^T X L_{FS}^w X^T A) \end{aligned} \quad (12)$$

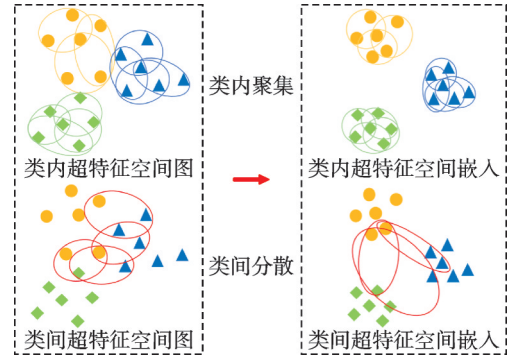


图4 类内、类间超特征空间嵌入图

Fig.4 Inter-class and intra-class hyper-feature space embedding graphs

建立类内散度的关键在于保持同类数据间近邻几何拓扑结构关系,同时实现同类数据聚合。具体为:首先,设定近邻参数 k_1 ,依据欧式度量准则确立每个样本点 x_i 的 k_1 个同类样本;其次,将近邻域中的样本按照 $C_{k_1}^P$ 进行排列组合,划定样本点 x_i 的 $C_{k_1}^P$ 个近邻特征空间元素;最后,将样本 x_i 分别与 $C_{k_1}^P$ 个近邻特征空间利用 $C_{k_1}^P$ 个超边进行连接,构建类内超特征空间图 $G_{FS}^{Hw} = \{X_{FS}^{Hw}, E_{FS}^{Hw}, W_{FS}^{Hw}\}$,其中: X_{FS}^{Hw} 、 E_{FS}^{Hw} 、 W_{FS}^{Hw} 分别为类内特征空间的顶点集、超边集以及超边权重集。

依据类内超特征空间图建立关联矩阵 H_{FS}^w ,即

$$H_{FS}^w = \begin{cases} h(x_i, e_i)_{FS}^w & (x_i \text{ 或 } x_{n,n \neq i} \in FS_P^w(x_i)) \\ 0 & \text{其他} \end{cases} \quad (9)$$

其中: $FS_P^w(x_i)$ 为样本点 x_i 的近邻域中的 P 阶类内特征空间。

类内超边的度、顶点的度分别定义为

$$d(e_j)_{FS}^w = \sum_{i=1}^n h(x_i, e_j)_{FS}^w \quad (10)$$

$$d(x_i)_{FS}^w = \sum_{j=1}^{ne} \omega(e_j)_{FS}^w h(x_i, e_j)_{FS}^w \quad (11)$$

其中:类内特征空间超边权重 $\omega(e_j)_{FS}^w$ 的求解方式为超边中任意2个元素余弦相似度之和。

由此建立类内超特征空间图散度矩阵 S_{FS}^{Hw} 为

建立类间散度的关键在于破坏异类数据间的近邻几何拓扑结构关系,同时实现异类团簇分散。该建立过程与类内散度矩阵相似,具体为:首先,设定参数 k_2 并确立每个样本点 x_i 的 k_2 个异类样本;其次,将近邻域中的样本按照 $C_{k_2}^P$ 进行排列组合,划定样本点 x_i 的 $C_{k_2}^P$ 个近邻特征空间;最后,利用 $C_{k_2}^P$ 个超边,将样本 x_i 分别与 $C_{k_2}^P$ 个近邻特征空间进行连接,构建类间超特征空间图 $G_{FS}^{Hb} = \{X_{FS}^{Hb}, E_{FS}^{Hb}, W_{FS}^{Hb}\}$,其中: X_{FS}^{Hb} 、 E_{FS}^{Hb} 、 W_{FS}^{Hb} 分别为类间特征空间超图中的顶点集、超边集和超边权重集。

依据类间超特征空间图建立关联矩阵 H_{FS}^b ,即

$$\begin{aligned} \sum_{e_j \in E_{FS}^{Hb}(x_i, x_k)} \sum_{x_i, x_k \in X_{FS}^{Hb}} \frac{\omega(e_j)_{FS}^b h(x_i, e_j)_{FS}^b h(x_k, e_j)_{FS}^b}{d(e_j)_{FS}^b} \|A^T x_i - A^T x_k\|^2 = \\ \sum_j \sum_{i, k} \frac{\omega(e_j)_{FS}^b H_{FS}^b(i, j) H_{FS}^b(k, j)}{d(e_j)_{FS}^b} \text{tr}(A^T x_i x_i A^T - A^T x_k x_k A^T) = \\ \left[\sum_j \sum_i \frac{\omega(e_j)_{FS}^b H_{FS}^b(i, j)}{d(x_i)_{FS}^b} \text{tr}(A^T x_i x_i A^T) \sum_k \frac{H_{FS}^b(k, j) d(x_k)_{FS}^b}{d(x_k)_{FS}^b} \right] - \\ \left[\sum_j \sum_{i, k} \frac{\text{tr}(A^T x_k \omega(e_j)_{FS}^b H_{FS}^b(i, j) H_{FS}^b(k, j) x_i^T A)}{d(x_k)_{FS}^b} \right] = \\ [\text{tr}(A^T X D(X)_{FS}^b X^T A)] - [\text{tr}(A^T X H_{FS}^b W_{FS}^b D(X)_{FS}^b {}^{-1} H_{FS}^b{}^T X^T A)] = \\ \text{tr}\{A^T X [D(X)_{FS}^b - H_{FS}^b W_{FS}^b D(E)_{FS}^b {}^{-1} H_{FS}^b{}^T] X^T A\} = \text{tr}(A^T X L_{FS}^b X^T A) \end{aligned} \quad (16)$$

2.3 建立目标函数

构建目标函数时,为实现类内样本聚集以及类间样本分散,需要最小化类内超特征空间图散度矩阵 S_w^{Hrs} ,同时最大化类间超特征空间图散度 S_b^{Hrs} 。因此,为使算法更利于故障分类,构建最大化散度矩阵差值形式的目标函数为

$$\begin{aligned} \underset{A}{\text{argmax}} \quad A^T (S_b^{Hrs} - S_w^{Hrs}) A \\ \text{s.t.} \quad A^T A = I \end{aligned} \quad (17)$$

利用拉格朗日乘子法,式(17)可等价于特征值求解问题,即

$$(S_b^{Hrs} - S_w^{Hrs}) A = \lambda A \quad (18)$$

将特征值降序排列,取最大的 f 个特征值对应的特征向量组成降维投影矩阵 A 。

2.4 HFSGE 算法步骤

HFSGE 算法具体实现步骤如下:

- 1) 输入:原始特征数据集 X ,空间参数 P 、同类近邻域参数 k_1 、异类近邻域参数 k_2 和目标维数 f ;
- 2) 构建类内、类间超特征空间嵌入图,并依据具体数据选定类内、类间近邻域参数 k_1 、 k_2 ;

$$H_{FS}^b = \begin{cases} h(x_i, e_j)_{FS}^b & (x_i \text{ 或 } x_{n, n \neq i} \in FS_P^b(x_i)) \\ 0 & \text{其他} \end{cases} \quad (13)$$

其中: $FS_P^b(x_i)$ 为样本点 x_i 的近邻域中的 P 阶类间特征空间。

类间超边的度、顶点的度分别定义为

$$d(e_j)_{FS}^b = \sum_{i=1}^n h(x_i, e_j)_{FS}^b \quad (14)$$

$$d(x_i)_{FS}^b = \sum_{j=1}^{ne} \omega(e_j)_{FS}^b h(x_i, e_j)_{FS}^b \quad (15)$$

其中: $\omega(e_j)_{FS}^b$ 为类间特征空间超边权重,其具体求解方式为超边中任意 2 个元素余弦相似度之和。

由此建立类间超特征空间图散度矩阵 S_b^{Hrs} 为

3) 根据式(12)、(16)分别计算类内、类间散度矩阵;

4) 对目标函数式(18)进行特征值分解,取前 f 个最大的特征值对应的特征向量组成投影矩阵 A ;

5) 基于 $Y = A^T X$ 计算样本在低维空间中的投影。

6) 输出:投影矩阵 A ,低维特征集 Y 。

3 基于 HFSGE 算法的轴承故障诊断流程

将 HFSGE 降维算法应用于轴承的故障诊断,具体实施步骤如下:

1) 从预处理后的多源监测信号中提取多域特征,构建高维特征数据集 X ;

2) 将 X 归一化处理后划分训练集 $\text{train}X$ 与测试集 $\text{test}X$;

3) 设定 HFSGE 算法参数,按照式(18)构建目标函数,将 $\text{train}X$ 输入算法进行训练得到投影矩阵 A ;

4) 根据 $Y = A^T X$ 对 $\text{train}X$ 和 $\text{test}X$ 分别进行降

维投影,得到低维特征集 trainY 与 testY;
5) 将投影后的 trainY 与 testY 输入至 K-近邻分

类器进行故障分类,并输出辨识结果。
故障诊断流程图如图 5 所示。

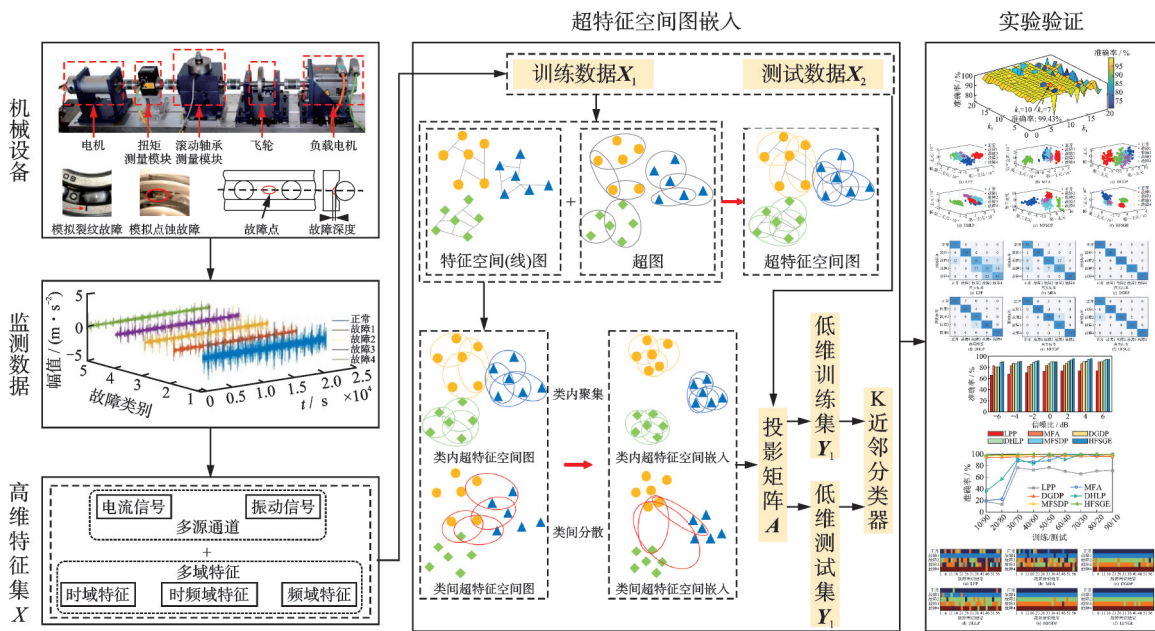


图 5 故障诊断流程图
Fig.5 Procedure of fault diagnosis

4 实验说明

4.1 故障模拟实验平台

为评估所提出 HFSGE 方法对不同故障识别和分类的有效性,采用帕德博恩大学轴承故障模拟实验台的故障数据集^[12]进行分析验证,转子实验台如图 6 所示。

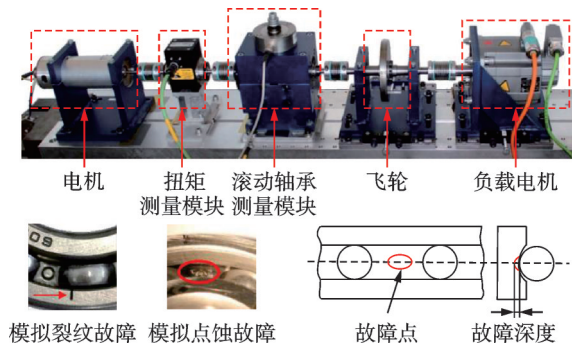


图 6 转子实验台
Fig.6 Rotor test bench

实验共采集 2 个通道的电机电流信号和 1 个通道的振动信号。在采样频率为 64 kHz、转速为 900 r/min 的条件下,模拟 5 种故障类型:K₀₀₁为健康状态;KA₀₁为外圈电火花加工故障,1 级损伤程度;KA₀₆为外圈点蚀,2 级损伤程度;KI₀₈为内圈点蚀故

障,2 级损伤程度;KI₁₇为内圈重复故障,2 级损伤程度。每种状态采样 100 组样本,按照 30:70 划分训练与测试数据集,从 2 个通道的电机电流信号以及 1 个通道的振动信号中提取出如表 1 所示的特征参数,共得到 38×2+38=114 个特征。

表 1 特征参数
Tab.1 Characteristic parameters

序号	特征	序号	特征	序号	特征
1	最大值	9	方根幅值	17	均值频率
2	最小值	10	均方根值	18	频谱二阶矩
3	峰峰值	11	绝对均值	19	标准偏差率
4	方差	12	波形指标	20	峭度频率
5	均值	13	峰值指标	21	均方根频率
6	歪度	14	脉冲指标	22	中心频率
7	峭度	15	裕度指标	23~	四层小波包分解
8	均方值	16	峭度指标	38	频带能量特征

1~16 为时域特征,17~22 为频域特征,23~38 为时频域特征。

4.2 参数设定

需设定的参数为:空间参数 P 、目标维数 f ; 式(12)中类内散度矩阵近邻数 k_1 ; 式(16)中类间散度矩阵近邻数 k_2 。为简化计算,设定空间参数 $P=2$ 。将目标维数 f 设定为故障的类别数减 1^[13],即 $f=5-1=4$ 。参数 k_1 与 k_2 对准确率的影响如图 7 所

示, k_1 与 k_2 的取值范围分别为 $2 \leq k_1$ 与 $k_2 \leq 20$ ^[14-15], 通过图 7 的网格搜寻选取最优解, 本研究将 k_1 和 k_2 分别取为 10 和 7。

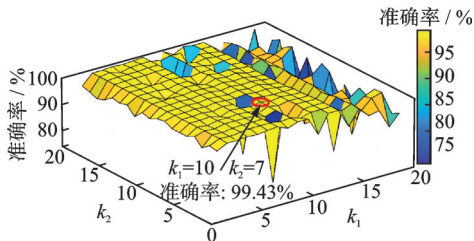


图 7 k_1 与 k_2 对准确率的影响
Fig.7 Effect of k_1 and k_2 on accuracy

4.3 可分性指标

为评价低维测试集各个故障类别间的可分性, 引入可分性指标 δ 评价指标来量化降维效果^[9], 即

$$O_B = \frac{1}{k} \sum_{i=1}^k (m_i - m)(m_i - m)^T \quad (19)$$

$$O_{w_i} = \frac{1}{c} \sum_{i=1}^c (y_n^{(i)} - m_i)(y_n^{(i)} - m_i)^T \quad (20)$$

$$\delta = O_B / \sum_{i=1}^c O_{w_i} p_i \quad (21)$$

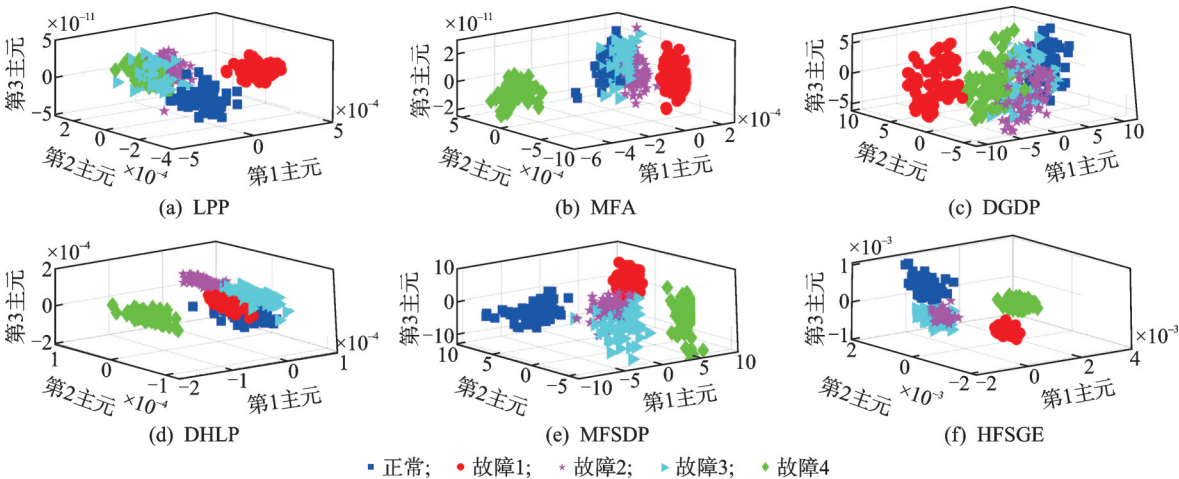


图 8 三维可视化效果图
Fig.8 Graph of 3D visualization

为定量描述各算法降维后的数据可分性, 依据 4.3 节式 (19)~(21) 计算各算法的降维可分性指标, 算法可分性指标结果如表 2 所示。由表可知, HFSGE 算法表现出更高的可分性。

表 2 算法可分性指标

Tab.2 Algorithm separability index

降维算法	MFA	LPP	DGDP	DHLP	MFSDP	HFSGE
δ	12.51	8.12	2.18	20.75	3.59	23.48

降维后数据的分类准确率也是判断降维效果的关键, 因此利用 K-近邻分类器对准确率各算法降维

其中: z 为故障类别数; m_i 为第 i 类均值; m 为总体均值; c 为每类中的样本个数; $p_i = n_i/n$, 为第 i 类的先验概率; $y_n^{(i)}$ 为第 i 类所包含的样本; O_{w_i} 为第 i 类的散度; δ 越大, 表示同类聚集更加紧凑, 异类更加分散, 算法降维效果越好。

5 实验结果与分析

5.1 降维效果分析

为验证 HFSGE 算法的降维效果, 采用 LPP^[4]、MFA^[5]、DGDP^[6]、MFSDP^[9] 以及 DHLP^[10] 算法进行对照分析, 并在各算法降维后的测试数据集中选取前 3 个特征主元绘制三维散点图, 三维可视化效果图结果如 8 所示。观察图 8 可知, 经 DGDP 算法降维后的同类样本聚合效果较弱, 进而导致异类团簇间的故障区分效果不明显。而经 HFSGE 算法降维后的同类故障样本聚集成团, 异类样本边界明显。经 LPP、MFA、DHLP、MFSDP 等算法降维后的故障样本均表现出不同程度的混叠现象, 且可视化效果介于 DGDP 与 HFSGE 算法之间。

后的测试样本进行分类辨识, 不同状态识别准确率如表 3 所示。观察可知, 采用 HFSGE 算法的准确率结果在分类辨识效果最好, 除故障 3 外的其余类别均被完全识别。采用 LPP 算法的故障分类精度最低, 平均准确率仅为 76.57%; MFA、DGDP、DHLP、HFSDP 算法识别准确率介于 LPP 与 HFSGE 之间。

观察图 8 并结合表 2、3 综合分析可知: LPP、MFA、DGDP 等通过任意两点建立邻接关系的普通图嵌入算法的降维综合效果欠佳。而包含更多元数

表3 不同状态识别准确率

Tab.3 Recognition rate of different states %

降维 算法	正常	故障1	故障2	故障3	故障4	平均识别 准确率
LPP	100	100	91.43	21.43	70.00	76.57
MFA	100	100	78.57	62.86	100	88.29
DGDP	95.71	100	94.29	87.14	100	95.43
DHLP	84.29	100	100	77.14	97.14	91.71
MFSDP	100	100	94.29	94.29	100	97.71
HFSGE	100	100	100	97.14	100	99.43

据样本特征空间、超图等高阶图嵌入方法可容纳多元数据点之间更为复杂的空间集合拓扑结构信息,易于提取并保留分类的敏感特征成分,表现出更好的降维效果。

5.2 不同训练比例实验

为进一步验证HFSGE算法的稳定性,本研究分别按照10:90、20:80、30:70、40:60、50:50、60:40、70:30、80:20和90:10的比例划分训练数据与测试测数,并将经各算法降维后的测试集放入K-近邻分类器中进行分类识别,不同比例数据的识别准确率如图9所示。

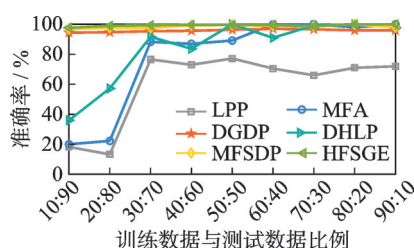


图9 不同比例数据的识别准确率

Fig.9 Recognition accuracy corresponding to different proportions of data

由图9可知,随着训练样本数量的增加,各算法的识别准确率逐渐上升,二者呈现出明显的正相关关系。具体来说,LPP整体准确率较低,原因在于其属于无监督算法,无法有效利用数据标签。在训练数量低于30时,MFA、DHLP算法识别精度十分

正常	57	0	3	0	0
故障1	0	70	0	0	0
故障2	12	0	46	9	7
故障3	1	0	21	39	18
故障4	0	0	0	22	45
真实标签	正常	故障1	故障2	故障3	故障4

(a) LPP

正常	50	1	2	5	2
故障1	1	69	0	0	0
故障2	8	0	61	12	0
故障3	10	0	7	53	0
故障4	1	0	0	0	68
真实标签	正常	故障1	故障2	故障3	故障4

(b) MFA

正常	70	0	3	0	0
故障1	0	70	0	0	0
故障2	0	0	67	3	0
故障3	0	0	3	62	3
故障4	0	0	0	5	67
真实标签	正常	故障1	故障2	故障3	故障4

(c) DGDP

有限,原因在于这2种算法的目标函数采用最大化Fisher准则,该准则在训练样本不足时,可能导致类内散度矩阵病态或奇异,进而使得模型求解不稳定,识别性能下降。DGDP、MFSDP、HFSGE算法则均表现出稳定的高识别精度,其中所提HFSGE算法表现最佳。

5.3 抗噪性实验

真实设备状态监测环境中存在大量干扰噪声,针对此问题,在采集的振动信号中分别加入信噪比为-6、-4、-2、0、2、4和6 dB的随机噪声,故障辨识结果如图10所示。由图可知,随着信噪比的增大,整体诊断准确率逐渐上升。DHLP、MFSDP、HFSGE等高阶图嵌入算法抗噪性能优于LPP、MFA、DGDP等普通图嵌入算法。其中,HFSGE在不同程度的噪声环境下均保持良好的鲁棒性能。

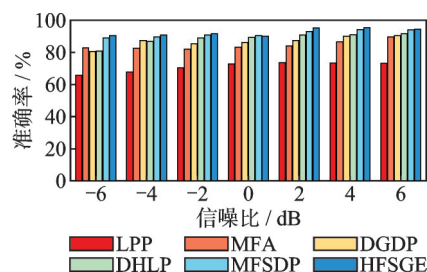


图10 抗噪性实验结果

Fig.10 Anti-noise experiment results

5.4 负载波动实验

负载的波动与变化是实际工况中的典型特征。为验证各算法的适用性,在转速为1500 r/min、采样频率为64 kHz、扭矩为0.7 N·m的条件下,分别采集负载为400与1000 N的样本各50组。将总计所得的100组样本混合后按照30:70划分训练与测试比例,输入K-近邻分类器中进行分类识别。图11为负载波动实验中故障识别混淆矩阵。其中,主对角线数值越大代表该类故障识别精度越高,对比图中数据可知,所提HFSGE算法在变负载实验中表现出最佳效果。

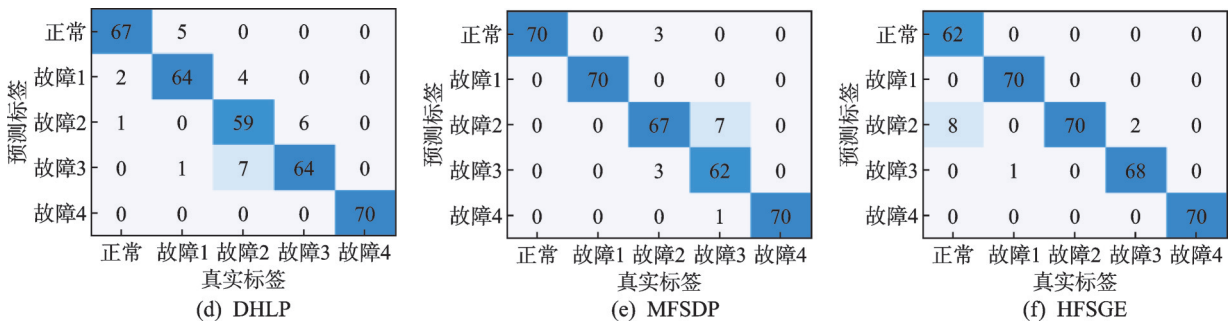


图 11 负载波动实验中故障识别混淆矩阵

Fig.11 Confusion matrix for fault identification in load fluctuation experiments

5.5 转速波动实验

为验证所提算法在转速波动变化工况中的表现,在采样频率为 64 kHz、负载为 1 000 N、扭矩为 0.7 N·m 的条件下,分别采集转速为 900 与 1 500 r/min 的样本各 50 组,将总计所得的 100 组样本混合后按照 30:70 划分训练与测试比例,输入 K-近邻分类器

中进行分类识别。图 12 为转速波动实验故障识别颜色梯度图。图中同行色块的一致性越高,代表该类故障识别精度越高。观察可知,LPP、MFA、DHLP 等算法色块一致性较差,DGDP、MFSDP、HFSGE 算法一致度更好,其中,HFSGE 算法在变速转速实验中表现出最佳效果。

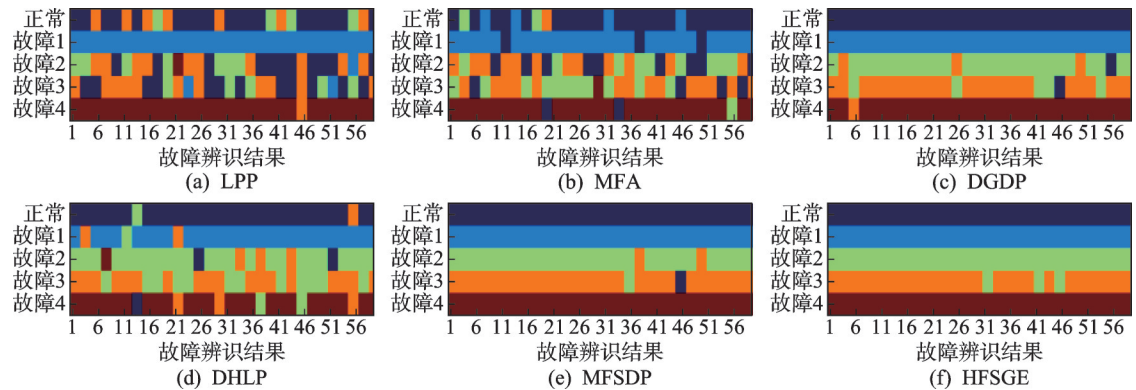


图 12 转速波动实验故障识别颜色梯度图

Fig.12 Color gradient diagram for fault identification in rotational speed fluctuation experiments

6 结 论

- 1) 通过新提出的特征空间超边建立准则与去重合超边的关联矩阵,更高效地保留了样本间的高阶邻接关系;利用最大化类间、类内超特征空间散度矩阵差,降低了故障分类难度。
- 2) 基于帕德博恩轴承故障模拟实验对算法进行了验证,结果表明,相较于 LPP、MFA、DGDP、MFSDP 以及 DHLP 等对比算法,所构建的 HFSGE 算法在降维效果方面的数据可分性指标最高为 23.48,且在变比例实验和抗噪实验中表现更为稳定可靠。此外,所提算法有效将故障辨识精度提高到 99.43%。
- 3) 所提 HFSGE 算法属于高阶图嵌入模型在故

障诊断领域的应用,可解决智能故障诊断方法受限于过高的故障数据维数这一问题,针对工业大数据技术诱发的旋转机械智能决策需求,提供了一种新的理论参考依据。

参 考 文 献

[1] WANG Z J, HE X X, YANG B, et al. Subdomain adaptation transfer learning network for fault diagnosis of roller bearings[J]. IEEE Transactions on Industrial Electronics, 2022, 69(8): 8430-8439.

[2] 陈是钊,彭志科,周鹏. 信号分解及其在机械故障诊断中的应用研究综述[J]. 机械工程学报, 2020, 56(17): 91-107.

CHEN Shiqian, PENG Zhike, ZHOU Peng. Review

- of signal decomposition theory and its applications in machine fault diagnosis[J]. *Journal of Mechanical Engineering*, 2020, 56(17): 91-107. (in Chinese)
- [3] 赵孝礼, 姚建勇, 邓文翔, 等. 基于GTCNs的航空发动机高速轴承智能故障诊断[J]. *振动、测试与诊断*, 2026, 46(1): 70-76.
- ZHAO Xiaoli, YAO Jianyong, Deng Wenxiang, et al. Intelligent fault diagnosis of aero-engine high-speed bearings using graph transfer convolutional networks under fluctuating working conditions[J]. *Journal of Vibration, Measurement & Diagnosis*, 2026, 46(1): 70-76. (in Chinese)
- [4] HE X F, NIYOGI P. Locality preserving projections [C]//*Proceeding of the 17th International Conference on Neural Information Processing Systems*. MA, USA: 2003: 153-160.
- [5] YAN S C, XU D, ZHANG B Y, et al. Graph embedding and extensions: a general framework for dimensionality reduction[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, 29(1): 40-51.
- [6] GOU J P, XUE Y, MA H X, et al. Double graphs-based discriminant projections for dimensionality reduction[J]. *Neural Computing and Applications*, 2020, 32(23): 17533-17550.
- [7] CHEN Y N, HAN C C, WANG C T, et al. Face recognition using nearest feature space embedding[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(6): 1073-1086.
- [8] IMANI M, MONTAZER G A. Phishing website detection using weighted feature line embedding[J]. *The ISC International Journal of Information Security*, 2017, 9(2): 147-159.
- [9] DONG X X, ZHAO R Z, YUAN J H, et al. Dimensionality reduction method based on multiple feature-space collaborative discriminative projection for rotor fault diagnosis[J]. *Measurement Science and Technology*, 2023, 34(5): 055007.
- [10] HUANG S, YANG D, GE Y X, et al. Discriminant hyper-Laplacian projections and its scalable extension for dimensionality reduction[J]. *Neurocomputing*, 2016, 173: 145-153.
- [11] YUAN J H, ZHAO R Z, CHEN P F, et al. Dimensionality reduction using local-global standard hypergraph embedding for rotor fault diagnosis[J]. *Measurement Science and Technology*, 2022, 34(3): 034006.
- [12] LESSMEIER C, KIMOTHO J K, ZIMMER D, et al. Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: a benchmark data set for data-driven classification[C]//*Proceedings of the European Conference of the PHM Society 2016*. Bilbao, Spain: PHM, 2016: 1577.
- [13] 原健辉, 赵荣珍, 邓林峰. 半监督超边界Fisher分析的转子故障数据降维算法[J]. *振动、测试与诊断*, 2025, 45(3): 543-550.
- YUAN Jianhui, ZHAO Rongzhen, Deng Linfeng. Dimensionality reduction algorithm of rotor fault dataset based on semi-supervised hyper-marginal fisher analysis [J]. *Journal of Vibration, Measurement & Diagnosis*, 2025, 45(3): 543-550. (in Chinese)
- [14] ZHAO X L, JIA M P. Fault diagnosis of rolling bearing based on feature reduction with global-local margin fisher analysis[J]. *Neurocomputing*, 2018, 315: 447-464.
- [15] 陈芳军, 赵荣珍, 邓林峰. 基于LFPCDA的转子故障数据集降维算法[J]. *振动、测试与诊断*, 2025, 45(6): 1151-1156.
- CHEN Fangjun, ZHAO Rongzhen, Deng Linfeng. A dimensionality reduction algorithm of rotor fault dataset based on LFPCDA [J]. *Journal of Vibration, Measurement & Diagnosis*, 2025, 45(6): 1151-1156. (in Chinese)



第一作者简介:丁华,女,1979年2月生,博士、教授、博士生导师。主要研究方向为煤机装备故障诊断与寿命预测等。曾发表《基于简化区间核全局-局部特征融合的采煤机智能故障诊断》(《煤炭学报》2024年第49卷第11期)等论文。
E-mail: dinghua2002@163.com