Unified discriminant manifold learning for rotating machinery fault diagnosis

Changyuan Yang[1,2,3,4] · Sai Ma[1,2,3,4] 🔗 · Qinkai Han[5]

Abstract

Fault diagnosis is an important technology for performing intelligent manufacturing. To simultaneously maintain high manufacturing quality and low failure rate for manufacturing systems, it is of great value to accurately locate the fault element, evaluate the fault severity and find the fault root cause. In order to effectively and accurately perform fault diagnosis for rotating machinery, a novel feature selection method named unified discriminant manifold learning (UDML) is proposed in this research. To be specific, the local linear relationship, the distance between adjacent points, the intra-class and inter-class variance are unified in UDML. Based on these, the local structure, global information and label information of high-dimensional features are effectively preserved by UDML. Through this dimension reduction method, homogeneous features become more concentrated while heterogeneous features become more distant. Consequently, mechanical faults could be diagnosed accurately with the help of proposed UDML. More importantly, local linear embedding algorithm, locality preserving projections algorithm, and linear discriminant analysis algorithm could be regarded as a special form of UDML. Moreover, a novel weighted neighborhood graph is constructed to effectively reduce the interference of outliers and noise. The corresponding model parameters are dynamically adjusted by the gray wolf optimization algorithm to find a subspace that discovers the intrinsic manifold structure for classification tasks. Based on the above innovations, a fault diagnosis method for rotating machinery is proposed. Through experimental verifications and comparisons with several classical feature selection algorithms, rotating machinery fault diagnosis can be more accurately performed by the proposed method.

Keywords Feature selection · Manifold learning · Feature extraction · Dimension reduction · Fault diagnosis

Introduction

Nowadays, with the maturity of technologies such as the internet of things and big data, predictive maintenance (Ding et al., 2020) has emerged for mechanical and electrical equipment. This technology can not only perform big data analysis, monitor equipment in real time, and perceive equipment failures, but also can troubleshoot potential failures in advance. Predictive maintenance makes maintenance more intelligent, operation more reliable, and both more economic. Predictive maintenance has become a general trend in industry (Ma et al., 2019). And for which, fault diagnosis is an important part. This technology collects the state signals of each mechanical part of the equipment through sensors, then extracts the features, and finally performs fault identification (Alavi et al., 2022). As is known to all, rotating machinery play a key role in many equipment and industrial fields.

✉ Sai Ma
massana@163.com

1 School of Mechanical Engineering, Shandong University, Jinan 250061, People's Republic of China

2 Key Laboratory of High-Efficiency and Clean Mechanical Manufacture at Shandong University, Ministry of Education, Jinan 250061, People's Republic of China

3 National Demonstration Center for Experimental Mechanical Engineering Education at Shandong University, Jinan 250061, People's Republic of China

4 Shandong Key Laboratory of Brain Function Remodeling, Qilu Hospital of Shandong University, Jinan 250061, People's Republic of China

5 Department of Mechanical Engineering, Tsinghua University, Beijing 100084, People's Republic of China

Once the rotating machinery fails, it may lead to the failure of the entire mechanical system or even accident. The intelligent fault diagnosis method of rotating machinery can generally be divided into three stages: data acquisition, feature selection and fault type identification (Li et al., 2022). It is a critical stage to select representative fault features from redundant state features considering the data volume and dimension is becoming larger and larger as the big data challenge evolves. Vibration signal is usually used as input data for mechanical fault diagnosis because it contains a wealth of fault information. But the original vibration signal is usually non-stationary and redundant, and contains complicated components, so feature extraction should be firstly performed for further processing. Wavelet transform inherits the localization ability of short-time Fourier transform, and overcomes the adaptivity inefficiency of the traditional signal processing methods. But this method has limitations such as difficulty in selecting wavelet basis and constant resolution problem. Empirical mode decomposition (Unver & Sener, 2021) could adaptively decompose the vibration signal into several intrinsic mode functions. Each component represents different meaningful physical information. However, it has modal confusion, end effect, over- and under-envelope problems. Gilles combined the idea of empirical mode decomposition and the wavelet analysis, and proposed the empirical wavelet transform (Gilles, 2013). In this algorithm, the frequency spectrum of the signal is adaptively divided by designed empirical orthogonal wavelet filter bank to extract Amplitude Modulation-Frequency Modulation (AM-FM) components. Since empirical wavelet transform is a method established under the wavelet framework, its theory is solid and the problems in empirical mode decomposition could be avoided. With feature extraction methods, high-dimensional feature set containing fault features can be obtained. But these fault features still contain a lot of noise and redundant information generated by the coupling of different features. Consequently, it's essential and beneficial to perform feature selection to obtain a fault feature set with lower redundancy and better clustering characteristics.

Feature selection and dimension reduction methods could be classified into supervised/unsupervised or linear/nonlinear types. Principal component analysis (Lee et al., 2020) and linear discriminant analysis (Yang et al., 2019) are classical linear dimensionality reduction methods. Principal component analysis can preserve the global information of the dataset by finding orthogonal bases and maximizing the total variance. Linear discriminant analysis considers the label information of input data. With the label information, linear discriminant analysis simultaneously minimizes intra-class variance and maximizes inter-class variance to produce the optimal discriminant projection. But when the number of samples is smaller than the dimensionality of samples, the corresponding intra-class scatter matrix is not reversible

which makes the algorithm unable to solve the problem (Li et al., 2006). And it is difficult for these traditional dimensionality reduction methods to find the nonlinear structure or local features of the high-dimensional dataset.

As an important part of the thriving brain-inspired artificial intelligence algorithms (Nieh et al., 2021), Manifold learning algorithms are utilized for dimensionality reduction recently (Siblini et al., 2021;). Representative manifold learning methods include isomap (Anowar et al., 2021), local linear embedding (Liu et al., 2021), locality preserving projections (He et al., 2005), local tangent space alignment (Kumar & Kumar, 2016), etc. Manifold learning has been widely used in mechanical fault diagnosis. Ding and He (2016) proposed a new type of feature extraction method based on time–frequency manifold learning for fault diagnosis, in which the part of dimensionality reduction method is local tangent space alignment algorithm and it has achieved good results. Xu et al. (2021) proposed the multi-manifold joint projections to reflect the essential characteristics within and between different patterns. Li et al. (2008) proposed the locally linear discriminant embedding algorithm. It combined the constraints of local linear embedding and maximum margin criterion to achieve high recognition accuracy. Sun et al. (2019) proposed an enhanced manifold learning method to reduce the dimension of fault features. The number of data neighbors and the connection weight are adaptively determined by the kernel sparse representation. However, these methods also have several shortcomings:

(1) Conventional manifold learning algorithms tend to be disturbed by noise and outliers, which could affect the feature selection performance for further fault diagnosis.

(2) Generally, only single constraint is considered in conventional manifold learning algorithms, for instance: the goal of local linear embedding algorithm is to preserve the local linear relationship, but the algorithm does not consider local features such as distance (Sha & Saul, 2005). Locality preserving projection algorithm preserves local information by maintaining adjacent distance, but the global information is not considered. Zhu et al. (2018) proposed the local and global structure preservation algorithm. But in this algorithm, the label information is not utilized.

(3) Label information could be utilized to improve the feature selection performance such as in linear discriminant analysis (Yang et al., 2019), locally linear discriminant embedding (Li et al., 2008) and so on.

Accordingly, a novel weighted neighborhood graph construction method and unified discriminant manifold learning (UDML) algorithm are proposed in this research. With this method, local linear relationship and local distance as well

as label information could be effectively utilized for feature selection. And it is also worth noting that local linear embedding, locality preserving projections, and linear discriminant analysis algorithms could be considered as special form of proposed UDML algorithm.

In summary, the main contributions of this work could be summarized as follows:

(1) A novel weighted neighborhood graph construction method is proposed based on q-Rényi kernel. As q-Rényi density function (Zhang et al., 2020) has the ability to suppress the disturbance of both Gaussian and non-Gaussian noise, it is utilized for nearest neighbor distance calculation, and the interference of outliers and noise is effectively restrained.

(2) A novel manifold learning algorithm is proposed for feature selection and fault diagnosis. The local linear reconstruction coefficient, the distance between adjacent points, intra-class and inter-class variance are simultaneously constrained in proposed UDML algorithm. With this operation, the local structure, global information and label information of high-dimensional features are effectively preserved by UDML.

(3) A rotating machinery fault diagnosis method based on the novel neighborhood graph and proposed UDML algorithm is proposed. The vibration signal of rotating machinery is firstly decomposed by empirical wavelet transform and features are extracted to form a high-dimensional feature set. Then, the fault features are selected by UDML. During this stage, the parameters of UDML are optimized by gray wolf optimization algorithm (Mirjalili et al., 2014) to improve its generalization performance. Finally, the low-dimensional fault feature sets are input to the k-nearest neighbor classifier (KNN) for fault type identification. As demonstrated by the experimental verifications, the fault diagnosis model proposed in this paper is suitable and effective for rotating machinery fault diagnosis.

This paper is organized as follows: The fault extraction method is described in Sect. 2. The proposed novel neighborhood graph, UDML algorithm and the rotating machinery fault diagnosis approach are shown in Sect. 3. Experimental results are shown in Sect. 4. Finally, conclusions are given in Sect. 5.

Fault feature extraction

In order to effectively extract fault feature, multi-component signals are conventionally decomposed into several components. Among the widely utilized methods, the adaptability of empirical mode decomposition and the theoretical framework of wavelet analysis are combined in empirical wavelet transform (Gilles, 2013). Considering its superior ability to obtain the condition related information for rotating

Table 1 Time-domain and frequency-domain features

Time-domain features		Frequency-domain features
$T_1 = \frac{\sum_{t=1}^{T} c_i(t)}{T}$	$T_5 = \frac{\sum_{t=1}^{T} (c_i(t))^4}{N T_4^4}$	$F_1 = \frac{\sum_{k=1}^{K} s(k)}{K}$
$T_2 = \max\lvert c_i(t)\rvert$	$T_6 = \frac{T_3}{\lvert T_1\rvert}$	$F_2 = \frac{\sum_{k=1}^{K} f_k s(k)}{\sum_{k=1}^{K} s(k)}$
$T_3 = \sqrt{\frac{\sum_{t=1}^{T}(c_i(t))^2}{T}}$	$T_7 = \frac{T_2}{\lvert T_1\rvert}$	$F_3 = \sqrt{\frac{\sum_{k=1}^{K} f_k^2 s(k)}{\sum_{k=1}^{K} s(k)}}$
$T_4 = \sqrt{\frac{\sum_{t=1}^{T}(c_i(t)-T_1)^2}{T}}$		$F_4 = \sqrt{\frac{\sum_{k=1}^{K}(f_k - F_2)^2 s(k)}{\sum_{k=1}^{K} s(k)}}$

machinery under instantaneous working conditions, the empirical wavelet transform is utilized in this research for feature extraction.

After the signal is decomposed, the fault features need to be extracted. The state of the system can be reflected by multi- domain distribution information of the vibration signal. As illustrated in Table 1 and Eq. (1)-(4), 7 time domain statistical features, 4 frequency domain statistical features (Gilles, 2013), 4 autoregressive coefficients and Shannon entropy are considered, which means altogether 16 multi-domain features are calculated and utilized for feature extraction in this study.

As shown in Table 1, $c_i(t)$ are the signal components extracted by the empirical wavelet transform($i = 1,…,N$). $s(k)$ is the spectrums($k = 1,…,K$). f_k is the frequency value. The time domain features are represented by T_1-T_7. The frequency domain features are represented by F_1-F_4 (Su et al., 2015).

Given that the autoregressive coefficients (Al-Bugharbee & Trendafilova, 2016) can reflect the characteristics of the system and are sensitive to the condition change of impact characteristics, they are also used for feature extraction. The autoregressive model can be established as follows:

$$c_i(t) = \sum_{j=1}^{m} \varphi_{ij} c_i(t-j) + e_i(t) \tag{1}$$

φ_{ij} ($j = 1,…,m$) are m order coefficients. $e_i(t)$ is the residual error. In this research, $A = [\varphi_1, \varphi_2, \varphi_3, \varphi_4]$ is extracted as 4 fault features.

Instantaneous amplitude Shannon entropy (Su et al., 2015) is a common information entropy used to evaluate signal

uncertainty. Fault feature can be represented by Shannon entropy because it reflects the characteristics and distribution of the vibration signal. The instantaneous amplitude $a_i(t)$ is shown as follows:

$$a_i(t) = \sqrt{c_i^2(t) + \hat{c}_i^2(t)} \tag{2}$$

$$\hat{c}_i(t) = \frac{1}{\pi} \int\limits_{-\infty}^{\infty} \frac{c_i(t)}{(t - \pi)} \mathrm{d}\tau \tag{3}$$

$\hat{c}_i(t)$ is the Hilbert transformation of $c_i(t)$. The Shannon entropy of the instantaneous amplitude is shown as follows:

$$S_i = \sum_{t=1}^{N} \left(|a_i(t)|^2 \log(|a_i(t)|^2) \right) \tag{4}$$

Unified discriminant manifold learning

The conventional manifold learning methods such as local linear embedding and locality preserving projection achieve feature selection by retaining the local linear relationship or adjacent distance on the data manifold. However, they fail to consider these constraints simultaneously. More importantly, these methods are local and unsupervised algorithms, and they ignore global and label information during the dimensionality reduction process. On the other hand, when the neighborhood graphs are constructed for these manifold learning algorithms, the relationship between adjacent points could be easily disturbed by noise and outliers, which may lead to the failure of the local relationship extraction.

In order to improve the feature selection performance for further fault diagnosis, a novel supervised manifold learning algorithm named unified discriminant manifold learning (UDML) is proposed in this research: Firstly, a new weighted neighborhood graph is designed. The q-Rényi kernel function is used to improve the neighborhood graph, and the interference of outliers and noise is effectively reduced. Then, the local linear relationship, the distance between adjacent points, intra-class and inter-class variance are unified in the proposed discriminant manifold learning (UDML) model. This model could effectively preserve both the local structure (linear relationship and adjacent point distance) and global structure (label information) of high-dimensional features. Then the homogeneous features become more concentrated while heterogeneous features become more distant. The conventional manifold learning algorithms such as local linear embedding, locality preserving projections and linear discriminant analysis could be regarded as special case of proposed UDML with proper parameter setting. Gray wolf

optimization algorithm is used to adjust the model parameters to improve the generalization ability. To cope with different data distribution, the weights of local linear relationship, nearest neighbor distance and global relationship (label information) could be adjusted adaptively. The notations used in this article are shown in Table 2.

Novel weighted neighborhood graph

Constructing neighborhood graph is the key step to establish the point-to-point relationship for high dimensional datasets. In order to constrain the influence of noise and outliers, a novel weighted neighborhood graph construction method is proposed in this research.

Gaussian kernel function is usually used to measure the distance information on the nearest neighbor graph. When the data set contains a variety of noise points and pseudo neighbors, the performance of neighborhood graph algorithm with conventional kernel functions, such as Gaussian kernel, could be affected. When $q = 1$, the q-Rényi distribution becomes Gaussian distribution. When q increases from small to large, the q-Rényi distribution changes from pulse shape to Gaussian distribution, and finally to uniform distribution (Zhang et al., 2020). q-Rényi kernel is defined as:

$$\kappa_{q,\sigma}(x_i, x_j) = \left[1 - \left(\frac{q-1}{3q-1} \right) \frac{\|x_i - x_j\|^2}{\sigma^2} \right]^{\frac{1}{q-1}} \tag{5}$$

The shape of the kernel is determined by q. σ is the kernel width. With the change of q, quadratic kernel, tricube kernel, Epanechnikov kernel and uniform kernel could be expressed by the q-Rényi kernel. The q-Rényi kernel function is used to define the edge of the nearest neighbor graph, which could effectively reduce the interference of noises and abnormal outliers. The edge of the nearest neighbor graph is defined as:

$$z_{ij} = \left[1 - \left(\frac{q-1}{3q-1} \right) \frac{\|x_i - x_j\|^2}{\sigma^2} \right]^{\frac{1}{q-1}}, \ x_j \in N_i(x_i) \tag{6}$$

When $x_i = [x_{i1}, x_{i2}]^T$, $x_j = [0,0]^T$, the surface of the z_{ij} with x_{i1} and x_{i2} is shown in Fig. 1. The closer x_i and x_j are, the value of the edge z_{ij} between them tends to be 1. When z_{ij} is close to the optimal value, the gradient could be reduced by adjusting q to reduce the disagreement. When z_{ij} is further away from the optimal value, the gradient can be adjusted by different q to avoid fluctuations caused by abnormal values.

Therefore, the novel weighted neighborhood graph with q-Rényi kernel is more robust and generalized. Distance information could be accurately retained and the disturbance of noise or outliers could be effectively restrained.

Table 2 Notations and descriptions

| Notation	Description	Notation	Description
X	High dimensional dataset	x	The single sample of X
Y	Low dimensional dataset	y	The single sample of Y
m	Dimension of X	d	Dimension of Y
A	Transformation matrix	α, β, γ	Weight factor of constraint
z	Edge of nearest neighbor graph	w	Neighborhood reconstruction coefficient
S_b	Inter-class variance matrix	S_w	Intra-class variance matrix
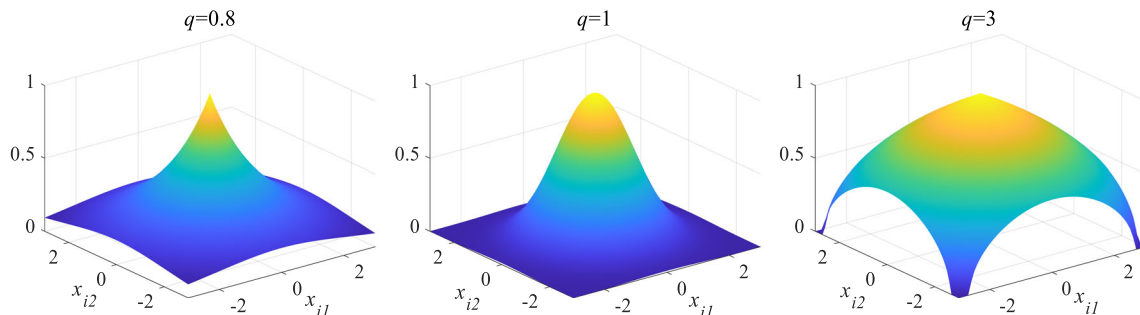
Fig. 1 The surface of the z_{ij} with x_{i1} and x_{i2}

The goal of UDML

Assume the m dimensional feature set X $(x_1, \ldots, x_n) \in R^m$ is composed of n vectors. The manifold learning algorithms are utilized to calculate the optimal transformation matrix A that maps the n feature vectors to feature set Y $(y_1, \ldots, y_n) \in R^d$ ($d < m$). This operation stands for the features selection process to obtain feature set with better intra-class clustering and inter-class discrimination characteristics. The features with higher representativeness could be retained and the redundant information will be removed.

To effectively perform feature selection for further fault diagnosis, a novel manifold learning cost function is designed in this research, the designed pluralistic cost function is composed of multiple constraints including the local linear relationship, the neighbor points distance, the intra-class variance and the inter-class variance. The corresponding manifold learning cost function is presented as follows:

$$\min \left\{ \alpha \sum_i \left\| y_i - \sum_{j=1}^{k} w_{ij} y_j \right\|^2 \right.$$

$$\left. + \beta \sum_{i,j} \left\| y_i - y_j \right\|^2 z_{ij} - \lambda(S_b - S_w) \right\}$$

$$s.t., \ YY^T = nI, \ 0 < \alpha, \beta, \gamma < 1, \alpha + \beta + \gamma = 1 \quad (7)$$

where w_{ij} is the reconstruction coefficients between node i and node j (Liu et al., 2021). z_{ij} is distance information of

the nearest neighbor (Shikkenawis & Mitra, 2016). S_b is the inter-class variance matrix and S_w is the intra-class variance matrix (Yang et al., 2019).

In this cost function, the first term is used to maintain the local linear relationship on the data manifold (Li et al., 2008). By minimizing the following loss function, the weights on the edges are obtained:

$$\min \sum_i \left\| x_i - \sum_{j \in N_i(x_i)} w_{ij} x_j \right\|^2$$

$$s.t., \begin{cases} \sum_{j=1}^{k} w_{ij} = 1, \text{ if } x_j \in N_i(x_i) \\ w_{ij} = 0, \text{ if } x \notin N_i(x_i) \end{cases} \quad (8)$$

where $N_i(x_i)$ denotes the k nearest neighbors of point x_i.

With the obtained weights w_{ij}, the local linear feature could be preserved by maintaining the linear representation relationship (Li et al., 2008). The objective function is as follows:

$$J_1(Y) = \min \sum_i \left\| y_i - \sum_{j=1}^{k} w_{ij} y_j \right\|^2$$

$$s.t., \begin{cases} \sum_{i=1}^{n} y_{ij} = 0 \\ Y_{d \times n} Y_{d \times n}^T = nI_{d \times d} \end{cases} \quad (9)$$

The second term of the proposed cost function is used to preserve the neighborhood distance, where z_{ij} is obtained from Eq. (6). The local information could also be maintained by nearest neighbor point distance (Shikkenawis & Mitra, 2016), which is modeled by the following constraint:

$$J_2(Y) = \min \sum_{i,j} \| y_i - y_j \|^2 z_{ij} \tag{10}$$

With Eq. (10), the nearest neighbor points are kept close after dimensionality reduction.

The third term of the cost function (S_b-S_w) is used to maintain the label information and global structure. Specifically, two reliable measures: inter-class variance S_b and intra-class variance S_w are used to ensure the smallest intra-class distance and largest inter-class distance. With this, homogeneous features could be concentrated while heterogeneous features become distant after dimensionality reduction.

In this research, the n samples $x_1, \ldots, x_n$ are assumed to belong to c classes. The number of samples in the i-th class is n_i. x_{ij} denotes the i-th sample in the j-th class, $i = 1,\ldots,n_j$, $j = 1,\ldots,c$. The inter-class variance matrix S_b and intra-class variance matrix S_w are as follows (Yang et al., 2019):

$$\begin{cases} S_b = \dfrac{1}{n} \sum_{j=1}^{c} n_j (\mu_j - \mu)(\mu_j - \mu)^T \\ S_w = \dfrac{1}{n} \sum_{j=1}^{c} \sum_{i=1}^{n_j} \left(x_i^j - \mu_j \right)\left(x_i^j - \mu_j \right)^T \end{cases} \tag{11}$$

$$s.t., \ \mu = \frac{1}{n} \sum_{i=1}^{n} x_i, \mu_i = \frac{1}{n_j} \sum_{i=1}^{n_j} x_i^j$$

Based on the inter-class and intra-class variance matrix, the objective function for utilizing discriminant and global information is given by:

$$J_3(Y) = \max(S_b - S_w) \tag{12}$$

As shown in Eq. (7), to leverage the ability of the aforementioned constraints, three loss functions J_1, J_2 and J_3 are unified in the proposed UDML method, which means the local linear reconstruction coefficients, adjacent points distance, intra-class and inter-class variance are simultaneously considered, therefore the local–global-label information of high-dimensional features are effectively preserved.

Mapping matrix construction

In order to construct the mapping matrix for the proposed UDML model, the objective functions J_1 and J_2 could be transformed into appropriate forms as follows, derivation of

which can be found in (He et al., 2005; Li et al., 2008).

$$\begin{aligned} J_1(Y) &= \min \left(y_i - \sum_{j=1}^{k} w_{ij} y_j \right) = \min (Y - WY)^T (Y - WY) \\ &= \min \operatorname{tr} \left(Y M Y^T \right) \end{aligned} \tag{13}$$

where $M = (I\text{-}W)^T (I\text{-}W)$, $I = diag(1, \ldots,1)$.

$$\begin{aligned} J_2(Y) &= \min \sum_{i,j} \| y_i - y_j \|^2 z_{ij} \\ &= \min \left(\operatorname{tr} \left(Y D Y^T \right) - \operatorname{tr} \left(Y Z Y^T \right) \right) = \min \operatorname{tr}(Y L Y^T) \end{aligned} \tag{14}$$

where $Z = [z_{ij}]_{n \times n}$, $D = diag\{D_{11},\ldots,D_{nn}\}$, $D_{ii} = \sum n\, j = 1 z_{ij}$, $L = D\text{-}Z$.

Then the proposed novel manifold learning model could be utilized to calculate the embedding feature set as: $Y = A^T X$. To find the optimal transformation matrix A, according to Eq. (9), Eq. (10) and Eq. (12), the following conditions should be satisfied:

$$\begin{cases} \min \operatorname{tr}(A^T X M X^T A) \\ \min \operatorname{tr}(A^T X L X^T A) \\ \max \operatorname{tr}\left(A^T (S_b - S_w) A \right) \end{cases} \tag{15}$$

$$s.t. A^T X X^T A = nI$$

Equation (15) could be transformed to the following constrained problem:

$$\min \operatorname{tr}\left\{ A^T \left(\alpha X M X^T + \beta X L X^T - \gamma (S_b - S_w) \right) A \right\} \tag{16}$$

$$s.t., \ A^T X X^T A = nI, \ 0 < \alpha, \beta, \gamma < 1, \alpha + \beta + \gamma = 1$$

The impact of different constraints can be adjusted by the weights α, β and γ. Then, Lagrange multipliers are used to solve the corresponding optimization problem:

$$\frac{\partial}{\partial A} \operatorname{tr} \left\{ \begin{array}{l} A^T \left(\alpha X M X^T + \beta X L X^T - \gamma (S_b - S_w) \right) A \\ -\lambda \left(A^T X X^T A - nI \right) \end{array} \right\} = 0 \tag{17}$$

Then Eq. (17) could lead to the following equation:

$$\left(\alpha X M X^T + \beta X L X^T - \gamma (S_b - \mu S_w) \right) A = \lambda X X^T A \tag{18}$$

where λ_i is the generalized eigenvalue of Eq. (18), A_i is the corresponding eigenvector. Therefore, the optimal mapping

matrix A could be obtained with the eigenvectors corresponding to the first d smallest eigenvalue.

The proposed UDML is a novel generalized model as it constrains multiple objectives including the ones used in local linear embedding, locality preserving projection and linear discriminant analysis, which makes these conventional methods special cases of proposed UDML.

Parameter optimization

In order to enhance the adaptivity of the proposed UDML model, the model parameters could be adjusted for specific applications, such as the kernel parameter q, the number of nearest neighbors k, the weight of constraints α, β, and λ. With this, the generalization ability and robustness of the proposed UDML method could be ensured.

Gray wolf optimization algorithm is a swarm intelligence multi-objective optimization algorithm based on gray wolf's rank and group hunting behavior (Mirjalili et al., 2014). During the hunting (optimization) process, α wolves, β wolves and δ wolves are in charge of guiding ω wolves to track and hunt prey. The main hunting processes of wolves includes: tracking and approaching, chasing and harassing, surrounding and attacking. The candidate solutions are distributed in a random circle defined by the three levels of wolves. Firstly, different levels of wolves evaluate the location of the prey, and then the rest of the individuals in the group use this as a reference and randomly update their positions around the prey. The process is repeated until the optimization result is achieved.

The outline of UDML

The steps of performing the proposed UDML method are shown as follows:

Unified Discriminant Manifold Learning

Input: high-dimensional data set X, d, q, α, β, γ, k

Output: Mapping matrix A, low-dimensional data set Y

1: Establish weighted neighborhood graph

2: W and Z are obtained by weighted neighborhood graph

3: M is obtained according to $M = (I-W)^T(I-W)$

4: L is obtained according to $L = D-Z$

5: S_b and S_w are obtained according to Eq. (11)

6: Matrix XMX^T, XLX^T and S_b-S_w are computed

7: A are obtained based on Eq. (18)

8: d dimensional embedding $Y = A^T X$ is obtained

Rotating machinery fault diagnosis based on the proposed method

A novel rotating machinery fault diagnosis method based on UDML is proposed, as is shown in Fig. 2. The vibration signal collected by the sensor is complicated, and different frequency bands contain various fault characteristic information. Therefore, the vibration signal is firstly processed by empirical wavelet transform and decomposed into several components. The aforementioned statistical features, autoregressive coefficients and Shannon entropy are extracted from the N components and the high-dimensional feature set is obtained.

As there is abundant redundancy information in the high-dimensional feature set, they may disturb the fault diagnosis approach, the high-dimensional feature set is input to UDML for feature selection and dimension reduction. To achieve accurate fault diagnosis, the kernel parameter q, the number of nearest neighbors k, the weight of constraints α, β, and λ of UDML are optimized by gray wolf optimization algorithm. When the diagnostic accuracy reaches 99.9% or the maximum number of iterations is limited, the optimization is completed. Then, the low dimensional feature set is obtained through UDML with optimized parameters. KNN is a classical classification algorithm with strong robustness and is often used in fault diagnosis (Bustillo et al., 2022). Finally, the low-dimensional feature set is input to the KNN for classification. In this way, an accurate fault diagnosis model could be automatically obtained for different situations.

Application of rotating machinery fault diagnosis

As two typical and important components of rotating machinery, bearings and gears would be affected by various impact loads during operation, which makes them most prone to failure in all components. In addition, long-term friction, corrosion, wear and other factors would also lead to bearing and gear failure. Therefore, effective fault diagnosis of bearings and gears could reduce the failure rate of mechanical equipment and effectively improve production efficiency. Many researchers have carried out fault diagnosis research on bearings and gears (Medina et al., 2022). Accordingly, the fault diagnosis of bearings and gears based on our proposed novel manifold learning method is performed in this paper.

Case 1

The rolling bearing experimental data comes from Paderborn University (Lessmeier et al., 2016; Hoang & Kang, 2020). The experimental ball bearing type is 6203. The spin-
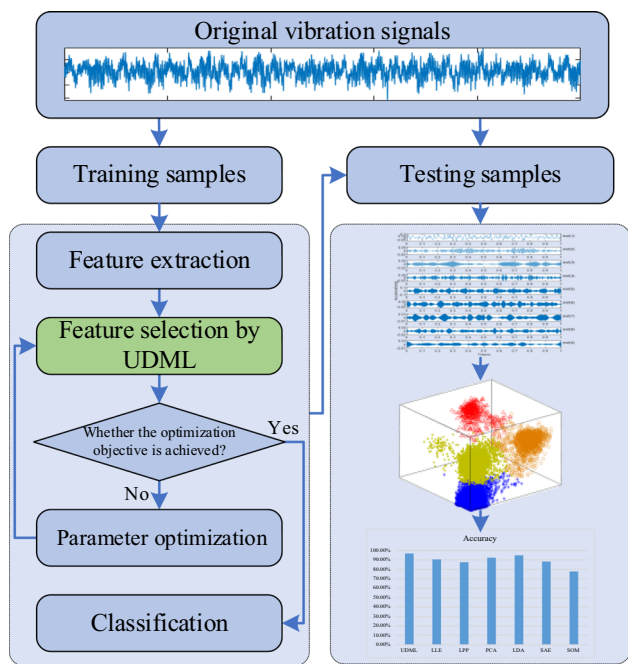
Fig. 2 The flowchart of the proposed rotating machinery fault diagnosis method

dle speed of the test-stand is 900 rpm, the sampling frequency is 64 kHz, load torque is 0.7 N·m and radial force is 1000 N. As shown in Fig. 3, the experiment device consists of 5 parts. Different types of bearings are installed in the rolling bearing test device to obtain experimental data.

A total of 3 operating states of bearings: (I) normal state; (II) outer race fault; (III) inner race fault are considered in this research. The faults of the bearing inner ring and outer ring are processed by electrical discharge machining, about 2 mm in size. There are 400 samples of vibration signals in each state (a total of 1200 samples), among which 900 samples are considered as training dataset and 300 samples are considered as testing dataset. In order to avoid overfitting, the whole dataset is divided into five parts for cross validation and cross validation is repeated five times, then the accuracy values are averaged.
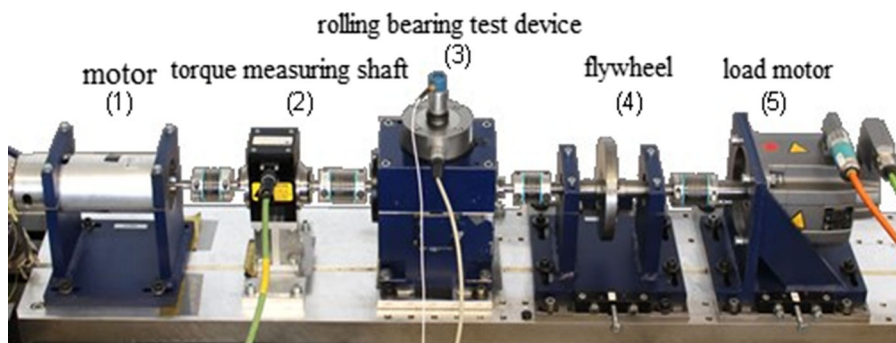
Firstly, the original vibration signals of bearings are decomposed by the empirical wavelet transform algorithm. Altogether 11 statistical features, 4 autoregressive coefficients and Shannon entropy are extracted. Then the proposed UDML is utilized to select features from high-dimensional feature set. In the parameter optimization process, the number of the search agents is set to 50 and the number of the iterations is 300. After optimization with the training dataset, $q = 0.2$, $\alpha = 0.3$, $\beta = 0.3$, $\gamma = 0.4$ and $k = 6$ are determined for UDML. The parameter k is set to 3 for the KNN. The embedded dimension d is set to 13. With these parameters, the fault diagnosis accuracy reaches 99.5% on the testing dataset. The standard deviation of each cross validation and each fault type are shown in Table 2. To demonstrate the superiority of the proposed method, it is compared with several conventional dimensionality reduction algorithms, including local linear embedding (LLE), locality preserving projection (LPP), principal component analysis (PCA), linear discriminant analysis (LDA), stacked autoencoder (SAE) (Pang et al., 2020) and self-organizing maps (SOM) (Moehrmann et al., 2011). The parameters of the comparison dimension reduction method are determined by grid search and the optimal parameters are as follows: The number of nearest neighbor points k in LLE and LPP is set to 12. The layer of SAE is set to 13. Dimension size is set to [4 3] for SOM. The comparison results are also shown in Table 3.

IR: inner-race fault; OR: outer-race fault.

It can be known from Table 3 that when UDML is used as the feature selection algorithm, the bearing fault could be accurately distinguished from each other. When performing dimensionality reduction on the bearing fault dataset in this experiment, more weights are imposed on the distance of the intra-class and inter-class constraints (the third term in cost function). In each verification, the standard deviation of UDML is also relatively low, which shows that UDML could stably and accurately select fault features.

In the experimental results, the recognition accuracy of the data set through LDA dimensionality reduction is higher than LPP and LLE, which shows that considering intra-class distance and inter-class distance is helpful (Su et al., 2015).

Fig. 3 Rolling bearing experiment device

Table 3 Comparison result of proposed method and baseline methods

Method	Normal	IR	OR	Accuracy
UDML	98.8 ± 0.7	100.0	99.8 ± 0.4	**99.5 ± 0.3**
LLE	94.4 ± 2.3	90.8 ± 1.9	98.8 ± 0.4	94.7 ± 1.3
LPP	75.4 ± 5.0	84.8 ± 2.6	98.4 ± 1.0	86.2 ± 1.2
PCA	79.0 ± 2.7	92.8 ± 3.2	92.8 ± 0.4	88.2 ± 1.6
LDA	99.0 ± 0.6	98.4 ± 1.0	100.0	99.1 ± 0.3
SAE	94.8 ± 1.5	97.4 ± 2.4	99.0 ± 0.9	97.1 ± 1.2
SOM	99.6 ± 0.4	99.6 ± 0.4	36 ± 5	78.4 ± 4.8

However, the fault diagnosis accuracy through the proposed UDML is higher than LDA, which also demonstrates the importance of retaining local structural information during dimensionality reduction. The global features, label information and specific local information are not preserved by LPP and LLE, so the accuracy of fault diagnosis is affected (Li et al., 2008). The local information and label information of the data are not preserved by PCA, so the fault features are not accurately selected, which results the low accuracy of fault diagnosis (Li et al., 2015). SAE is an unsupervised neural network algorithm with multiple hidden layers. When selecting fault features through SAE, label information is ignored, which results lower fault diagnosis accuracy of SAE than LDA and UDML. SAE could not construct explicit mapping between input and output dataset, so it is difficult to generalize the results of training samples to new samples (Pang et al., 2020). SOM is an unsupervised algorithm based on neural network, which is composed of cell grid neurons on the map. The algorithm is too sensitive to the initial data when dealing with small sample problems. And because of the lack of label information, the features of bearing outer ring fault are not accurately selected by SOM algorithm. The generalization ability of SOM is poor as this algorithm also could not construct explicit mapping between input and output dataset (Moehrmann et al., 2011).

Case 2

The experimental data was gathered from a two-stage gearbox experiment system (Cao et al., 2018; Shao et al., 2019), as shown in Fig. 4. The first stage input shaft consists of 32-tooth pinion and 80-tooth gear. The 48-tooth pinion and 64-tooth gear are mounted on the second stage input shaft. The gear speed is controlled by the motor. The sampling frequency is 20 kHz.

The monitoring signal of different pinion gear states on the input shaft is collected. The states of gear include 5 types, as shown in Fig. 5. 208 samples are collected for each state and a total of 1040 samples are collected, among which 800 samples are training samples and 240 samples are testing samples. The dataset is also divided into five parts for cross

Fig. 4 Gearbox experiment system

validation and the cross validation is repeated five times, then the accuracy values are averaged.

The original vibration signals of gears are firstly decomposed by the empirical wavelet transform algorithm. Then 11 statistical features, 4 autoregressive coefficients and Shannon entropy are extracted. The number of the search agents is set to 50 and the iterations number is 300 for the optimization algorithm. After optimization, $q = 1.6$, $\alpha = 0.1$, $\beta = 0.1$, $\gamma = 0.8$ and $k = 11$ are determined. The embedded dimension d is set to 16 by trial-and-error test. Then the proposed UDML is utilized to select features from high-dimensional feature set. And the low-dimensional feature set is input to the KNN ($k = 3$) for classification. It is verified that the fault diagnosis accuracy is 96.8%. UDML is compared with other algorithms, including NPE, LPP, PCA, LDA, SAE and SOM. And the parameters of the comparison dimension reduction method are determined by grid search and the optimal parameters are as follows: The k of LLE and LPP is set to 5 and 21, respectively. The layer number of SAE is set to 8. Dimensions size is set to [8 8] for SOM. The comparison results are shown in Table 4.

State 1: health status; State 2: missing teeth; State 3: tooth root cracks; State 4: spalling; State 5: chipping tip.

Fig. 5 Five gears with different health conditions

Table 4 Comparison result of proposed method and baseline methods

Method	State1	State2	State3	State4	State5	Average
UDML	95.4 ± 3.7	96.3 ± 2.3	97.9 ± 1.5	97.9 ± 2.9	96.3 ± 1.7	**96.8 ± 1.4**
LLE	92.5 ± 2.4	85.8 ± 7.3	95.4 ± 2.3	100	78.7 ± 1.7	90.5 ± 1.2
LPP	88.3 ± 5.4	90.8 ± 3.5	80.4 ± 9.8	89.2 ± 5.4	89.2 ± 5.6	87.6 ± 4.5
PCA	92.9 ± 2.4	89.6 ± 5.3	95.4 ± 2.7	99.6 ± 0.9	84.2 ± 6.4	92.3 ± 1.3
LDA	93.3 ± 2.3	91.2 ± 4.5	95.4 ± 4.6	100	94.6 ± 4.8	94.9 ± 1.0
SAE	92.1 ± 0.9	82.1 ± 9.2	94.2 ± 1.1	99.2 ± 0.8	73.3 ± 6.7	88.2 ± 3.1
SOM	92.9 ± 4.8	58.8 ± 7.8	93.3 ± 5.4	97.9 ± 1.5	45.8 ± 8.6	77.7 ± 3.6

It can be shown that the fault diagnosis accuracy of the approach with the proposed UDML method is the highest among all methods. For all fault types, fault features could be accurately preserved by UDML. As the multiple constraint weights of UDML can be adjusted, the dimension reduction stability of UDML is better than the compared methods.

By using the label information, the fault features of gears could be accurately selected by LDA. Local information is ignored by LDA, so the accuracy of LDA is lower than that of UDML (Zhao & Jia, 2018). LLE algorithm performs poorly for chipping tip fault detection as this method ignores the neighborhood distance information and label information (Li & Zhang, 2011). Compared with other manifold learning algorithms, the overall accuracy of LPP is relatively low, which is because the local linear relationship and label information is ignored (Shikkenawis & Mitra, 2016). Because PCA is unsupervised, the accuracy of fault diagnosis through PCA is lower than LDA (Yang et al., 2019). Because of the lack of label information, the missing tooth fault features and chip tip fault features could not be accurately selected by SAE and SOM. SAE and SOM could not construct explicit mapping between input and output dataset, so it is difficult to generalize the training results to new samples in fault diagnosis (Moehrmann et al., 2011; Pang et al., 2020). In summary, the proposed UDML can adaptively maintain local and global structure, as well as the label information, which improves the feature selection ability for gear fault diagnosis approach.

Conclusion

A novel rotating machinery fault diagnosis method based on the novel weighted neighborhood graph construction method and unified discriminant manifold learning (UDML) model is proposed in this paper. The novel weighted neighborhood graph is constructed to effectively reduce the interference of outliers and noise. The proposed unified discriminant manifold learning algorithm can simultaneously preserve the local linear relationship, neighborhood distance, intra-class and inter-class information for datasets. Therefore, it could be used for rotating machinery fault diagnosis to accurately select and preserve representative fault features. Local linear embedding algorithm, locality preserving projections algorithm, and linear discriminant analysis algorithm could be regarded as special form of proposed UDML. Combined with the swarm intelligence multi-objective optimizer, the corresponding parameters can be adjusted adaptively, which makes the proposed method applicable for various types of fault datasets. As demonstrated by the experiments, the proposed method is the most accurate one for rotating machinery fault diagnosis. The UDML could also be used for fault diagnosis of other industrial systems. UDML is essentially a manifold learning algorithm, which could effectively extract low dimensional features from high-dimensional complex data sets. It could also be used in: process monitoring (Tong

et al., 2016; Xu et al., 2021), diesel engine fault diagnosis (Wang et al., 2021; Xi et al., 2018), nuclear power plant fault diagnosis (Li et al., 2018) and so on. In future research, UDML will be used for wear detection of turning tools and defect monitoring of additive manufacturing. UDML could also be used to extract hybrid fault features and perform multi-sensor information fusion. Moreover, considering the linearity of the UDML algorithm, the incremental learning could be achieved by the proposed method.

Author contributions All authors contributed to the study conception and design, data collection and analysis were performed by CY, SM and QH. The first draft of the manuscript was written by CY and SM. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose. The authors have no competing interests to declare that are relevant to the content of this article. All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript. The authors have no financial or proprietary interests in any material discussed in this article. The authors declare that they have no conflict of interest.

Research involving human participants or animals This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

Alavi, H., Ohadi, A., & Niaki, S. T. (2022). A novel targeted method of informative frequency band selection based on lagged information for diagnosis of gearbox single and compound faults. *Mechanical Systems and Signal Processing, 170*, 108828. https://doi.org/10.1016/j.ymssp.2022.108828

Al-Bugharbee, H., & Trendafilova, I. (2016). A fault diagnosis methodology for rolling element bearings based on advanced signal pretreatment and autoregressive modelling. *Journal of Sound and Vibration, 369*, 246–265. https://doi.org/10.1016/j.jsv.2015.12.052

Anowar, F., Sadaoui, S., & Selim, B. (2021). Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE).

Computer Science Review, 40, 100378. https://doi.org/10.1016/j.cosrev.2021.100378

Bustillo, A., Reis, R., Machado, A. R., & Pimenov, D. (2022). Improving the accuracy of machine-learning models with data from machine test repetitions. *Journal of Intelligent Manufacturing, 33*, 203–221. https://doi.org/10.1007/s10845-020-01661-3

Cao, P., Zhang, S., & Tang, J. (2018). Gear fault data. Figshare. https://doi.org/10.6084/m9.figshare.6127874.v1

Ding, H., Gao, R. X., Isaksson, A. J., Landers, R. G., Parisini, T., & Yuan, Y. (2020). State of AI-based monitoring in smart manufacturing and introduction to focused section. *IEEE/ASME Transactions on Mechatronics, 25*(5), 2143–2154. https://doi.org/10.1109/TMECH.2020.3022983

Ding, X., & He, Q. (2016). Time–frequency manifold sparse reconstruction: A novel method for bearing fault feature extraction. *Mechanical Systems and Signal Processing, 80*, 392–413. https://doi.org/10.1016/j.ymssp.2016.04.024

Gilles, J. (2013). Empirical wavelet transform. *IEEE Transactions on Signal Processing, 61*(16), 3999–4010. https://doi.org/10.1109/TSP.2013.2265222

He, X., Yan, S., Hu, Y., Niyogi, P., & Zhang, H. (2005). Face recognition using Laplacianfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27*(3), 328–340. https://doi.org/10.1109/TPAMI.2005.55

Hoang, D. T., & Kang, H. J. (2020). A motor current signal-based bearing fault diagnosis using deep learning and information fusion. *IEEE Transactions on Instrumentation and Measurement, 69*(6), 3325–3333. https://doi.org/10.1109/TIM.2019.2933119

Kumar, A., & Kumar, R. (2016). Manifold learning using linear local tangent space alignment (LLTSA) algorithm for noise removal in wavelet filtered vibration signal. *Journal of Nondestructive Evaluation, 35*, 50. https://doi.org/10.1007/s10921-016-0366-4

Lee, W. J., Mendis, G. P., Triebe, M. J., & Sutherland, J. W. (2020). Monitoring of a machining process using kernel principal component analysis and kernel density estimation. *Journal of Intelligent Manufacturing, 31*, 1175–1189. https://doi.org/10.1007/s10845-019-01504-w

Lessmeier, C., Kimotho, J. K., Zimmer, D., & Sextro, W. (2016). 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. In *3rd European Conference of the Prognostics and Health Management Society* (pp. 1–17).

Li, B., & Zhang, Y. (2011). Supervised locally linear embedding projection (SLLEP) for machinery fault diagnosis. *Mechanical Systems and Signal Processing, 25*(8), 3125–3134. https://doi.org/10.1016/j.ymssp.2011.05.001

Li, B., Zheng, C., & Huang, D. (2008). Locally linear discriminant embedding: An efficient method for face recognition. *Pattern Recognition, 41*(12), 3813–3821. https://doi.org/10.1016/j.patcog.2008.05.027

Li, F., Wang, J., Chyu, M. K., & Tang, B. (2015). Weak fault diagnosis of rotating machinery based on feature reduction with Supervised Orthogonal Local Fisher Discriminant Analysis. *Neurocomputing, 168*, 505–519. https://doi.org/10.1016/j.neucom.2015.05.076

Li, H., Jiang, T., & Zhang, K. (2006). Efficient and robust feature extraction by maximum margin criterion. *IEEE Transactions on Neural Networks, 17*(1), 157–165. https://doi.org/10.1109/TNN.2005.860852

Li, W., Peng, M., Liu, Y., Jiang, N., Wang, H., & Duan, Z. (2018). Fault detection, identification and reconstruction of sensors in nuclear power plant with optimized PCA method. *Annals of Nuclear Energy, 113*, 105–117. https://doi.org/10.1016/j.anucene.2017.11.009

Li, Y., Wang, S., Li, N., & Deng, Z. (2022). Multiscale symbolic diversity entropy: a novel measurement approach for time-series analysis and its application in fault diagnosis of planetary gearboxes. *IEEE Transactions on Industrial Informatics, 18*(2), 1121–1131. https://doi.org/10.1109/TII.2021.3082517

Liu, Y., Hu, Z., & Zhang, Y. (2021). Bearing feature extraction using multi-structure locally linear embedding. *Neurocomputing, 428*, 280–290. https://doi.org/10.1016/j.neucom.2020.11.048

Ma, S., Chu, F., & Han, Q. (2019). Deep residual learning with demodulated time-frequency features for fault diagnosis of planetary gearbox under nonstationary running conditions. *Mechanical Systems and Signal Processing, 127*, 190–201. https://doi.org/10.1016/j.ymssp.2019.02.055

Medina, R., Macancela, J. C., Lucero, P., Cabrera, D., Sánchez, R., & Cerrada, M. (2022). Gear and bearing fault classification under different load and speed by using Poincaré plot features and SVM. *Journal of Intelligent Manufacturing, 33*, 1031–1055. https://doi.org/10.1007/s10845-020-01712-9

Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey wolf optimizer. *Advances in Engineering Software, 69*, 46–61. https://doi.org/10.1016/j.advengsoft.2013.12.007

Moehrmann, J., Burkovski, A., Baranovskiy, E., Heinze, GA., Rapoport, A., Heidemann, G. (2011). A discussion on visual interactive data exploration using self-organizing maps. In Laaksonen, J., Honkela, T. (Eds.), *Advances in self-organizing maps* (pp. 178–187). Springer, Berlin. https://doi.org/10.1007/978-3-642-21566-7_18.

Nieh, E. H., Schottdorf, M., Freeman, N. W., Low, R. J., Lewallen, S., Koay, S. A., Pinto, L., Gauthier, J. L., Brody, C. D., & Tank, D. W. (2021). Geometry of abstract learned knowledge in the hippocampus. *Nature, 595*(7865), 80–84. https://doi.org/10.1038/s41586-021-03652-7

Pang, S., Yang, X., Zhang, X., & Lin, X. (2020). Fault diagnosis of rotating machinery with ensemble kernel extreme learning machine based on fused multi-domain features. *ISA Transactions, 98*, 320–337. https://doi.org/10.1016/j.isatra.2019.08.053

Sha, F., & Saul, L. K. (2005). Analysis and extension of spectral methods for nonlinear dimensionality reduction. In *ICML 2005- Proceedings of the 22nd International Conference on Machine Learning* (pp. 784–791). https://doi.org/10.1145/1102351.1102450

Shao, S., McAleer, S., Yan, R., & Baldi, P. (2019). Highly accurate machine fault diagnosis using deep transfer learning. *IEEE Transactions on Industrial Informatics, 15*(4), 2446–2455. https://doi.org/10.1109/TII.2018.2864759

Shikkenawis, G., & Mitra, S. K. (2016). On some variants of locality preserving projection. *Neurocomputing, 173*, 196–211. https://doi.org/10.1016/j.neucom.2015.01.100

Siblini, W., Kuntz, P., & Meyer, F. (2021). A review on dimensionality reduction for multi-label classification. *IEEE Transactions on Knowledge and Data Engineering, 33*(3), 839–857. https://doi.org/10.1109/TKDE.2019.2940014

Su, Z., Tang, B., Liu, Z., & Qin, Y. (2015). Multi-fault diagnosis for rotating machinery based on orthogonal supervised linear local tangent space alignment and least square support vector machine. *Neurocomputing, 157*, 208–222. https://doi.org/10.1016/j.neucom.2015.01.016

Sun, C., Wang, P., Yan, R., Gao, R. X., & Chen, X. (2019). Machine health monitoring based on locally linear embedding with kernel sparse representation for neighborhood optimization. *Mechanical Systems and Signal Processing, 114*, 25–34. https://doi.org/10.1016/j.ymssp.2018.04.044

Tong, C., Shi, X., & Lan, T. (2016). Statistical process monitoring based on orthogonal multi-manifold projections and a novel variable contribution analysis. *ISA Transactions, 65*, 407–417. https://doi.org/10.1016/j.isatra.2016.06.017

Unver, H. O., & Sener, B. (2021). A novel transfer learning framework for chatter detection using convolutional neural networks. *Journal of Intelligent Manufacturing*. https://doi.org/10.1007/s10845-021-01839-3

Wang, R., Chen, H., Guan, C., Gong, W., & Zhang, Z. (2021). Research on the fault monitoring method of marine diesel engines based on the manifold learning and isolation forest. *Applied Ocean Research, 112*, 102681. https://doi.org/10.1016/j.apor.2021.102681

Xi, W., Li, Z., Tian, Z., & Duan, Z. (2018). A feature extraction and visualization method for fault detection of marine diesel engines. *Measurement, 116*, 429–437. https://doi.org/10.1016/j.measurement.2017.11.035

Xu, X., Ding, J., Liu, Q., & Chai, T. (2021). A novel multi manifold joint projections model for multimode process monitoring. *IEEE Transactions on Industrial Informatics, 17*(9), 5961–5970. https://doi.org/10.1109/TII.2020.3036676

Yang, A., Wang, Y., Zi, Y., & Chow, T. W. S. (2019). An enhanced trace ratio linear discriminant analysis for fault diagnosis: an illustrated example using HDD data. *IEEE Transactions on Instrumentation and Measurement, 68*(12), 4629–4639. https://doi.org/10.1109/TIM.2019.2900885

Zhang, Y., Peng, L., Li, X., & Xie, Y. (2020). A sparse robust adaptive filtering algorithm based on the q-Rényi Kernel function. *IEEE Signal Processing Letters, 27*, 476–480. https://doi.org/10.1109/LSP.2020.2978408

Zhao, X., & Jia, M. (2018). Fault diagnosis of rolling bearing based on feature reduction with global-local margin Fisher analysis. *Neurocomputing, 315*, 447–464. https://doi.org/10.1016/j.neucom.2018.07.038

Zhu, X., Zhang, S., Hu, R., Zhu, Y., & Song, J. (2018). Local and global structure preservation for robust unsupervised spectral feature selection. *IEEE Transactions on Knowledge and Data Engineering, 30*(3), 517–529. https://doi.org/10.1109/TKDE.2017.2763618