ATLANTIS PRESS International Journal of Computational Intelligence Systems, Vol. 11 (2018) 428–437

Fault diagnosis of sucker rod pumping systems based on Curvelet Transform and sparse multi-graph regularized extreme learning machine

Ao Zhang , Xianwen Gao *

[1] *College of Information Science and Engineering, Northeastern University, Shenyang 110819, China*

Received 14 November 2017

Accepted 24 December 2017

Abstract

A novel approach is proposed to complete the fault diagnosis of pumping systems automatically. Fast Discrete Curvelet Transform is firstly adopted to extract features of dynamometer cards that sampled from sucker rod pumping systems, then a sparse multi-graph regularized extreme learning machine algorithm (SMELM) is proposed and applied as a classifier. SMELM constructs two graphs to explore the inherent structure of the dynamometer cards: the intra-class graph expresses the relationship among data from the same class and the inter-class graph expresses the relationship among data from different classes. By incorporating the information of the two graphs into the objective function of extreme learning machine (ELM), SMELM can force the outputs of data from the same class to be as same as possible and simultaneously force results from different classes to be as separate as possible. Different from previous ELM models utilizing the structure of data, our graphs are constructed through sparse representation instead of K-nearest Neighbor algorithm. Hence, there is no parameter to be decided when constructing graphs and the graphs can reflect the relationship among data more exactly. Experiments are conducted on dynamometer cards acquired on the spot. Results demonstrate the efficacy of the proposed approach for faults diagnosis in sucker rod pumping systems.

Keywords: curvelet transform; extreme learning machine; sparse representation; sucker rod pumping systems; fault diagnosis

1. Introduction

Sucker rod pumping systems are the most common artificial lift methods used for oil production. Nearly 90% artificially lifted wells in the world[1] and nearly 94% artificially lifted wells in China[2] adopt sucker rod pumping systems. In production practice, sucker rod pumping systems are not stable and many kinds of faults can cause reduction of oil production, stop production and even damage to equipments. Therefore, diagnosing the faults of suck rod pumping systems automatically has been a very important research subject.

Many advanced methods have been adopted to overcome this problem. J.P. Wang, Z.F. Bao[3] used a rough classifier to diagnosis faults. Y.H. He *et al*[4] transferred the time domain signals into the frequency domain signals and used a fuzzy mathematical recognition model for diagnosis of failures in pumping wells. P. Xu *et al*[5] constructed a self-organizing competitive neural network model to achieve automatization of fault diagnosis. W. Wu *et al*[6] decomposed cards to get eight energy eigenvectors by using three layers of wavelet packet, and then regard them as the input of the RBF networks. K. Li *et al*[7] used the moment curve method to extract the

* Corresponding author: Xianwen Gao, E-mail: gaoxianwen@mail.neu.edu.cn

features of typical dynamometer cards and then used the improved SVM method for pattern classification. These researches show that fault diagnosis of sucker rod pumping systems can be regarded as a process of pattern recognition.

Feature extraction and the pattern classification are two important factors in pattern recognition problems. For sucker rod pumping systems, different working conditions can be represented by the shapes of dynamometer cards. Furthermore, the differences among dynamometer cards are mainly reflected in the scales and the directions. Curvelet Transform is a multi-resolution method which has been widely used to handle feature extraction problems[8,9,10]. It is not only a multi-scale but also a multi-direction transform, so features extracted by Curvelet Transform are very sensitive to the shapes of dynamometer cards. In this paper, Fast Discrete Curvelet Transform (FDCT) is adopted to extract features of dynamometer cards.

After feature extraction, an appropriate model should be chosen to classify the dynamometer cards. Extreme Learning Machine (ELM) proposed by G.B. Huang *et al*[11] is a non-iterative method that has been proven to be an efficient model for classification. Many efforts have been made to improve the performance of ELM. Recently, using the data's inherent structure detected by manifold learning algorithms[12,13,14,15] to improve the performance of existing machine learning methods has drawn much attention[16,17,18,19]. Y. Peng *et al*[20] used the K-nearest Neighbor algorithm (KNN) to construct the graphs and introduce the discriminative information into the ELM model to improve the performance, but how to choose the value of the parameter k in KNN is still a problem. Motived by Sparsity preserving projection[21] and Graph regularized sparsity discriminant analysis[22], we propose a sparse multi-graph regularized extreme learning machine (SMELM) for classification. In SMELM, two graphs: intra-class graph and inter-class graph are constructed to utilize the inherent structure of the training data more effectively. The intra-class graph can reflect the similarity among data in the same class; the inter-class graph can reflect the relationship among data from different classes. These two graphs help to improve the performance by forcing the output results of data from same class to be as same as possible and meanwhile forcing results from different classes to be as different as possible. Different from discriminative graph regularized extreme learning machines[23], the sparse representation algorithm is used instead of the KNN algorithm, so there is no parameter to be decided when constructing graphs and the elements in graph matrix are not simplify set as 0 or 1. The graph matrix calculated by sparse representation can reflect the relationship among data more exactly so that the proposed SMELM can achieve better performance.

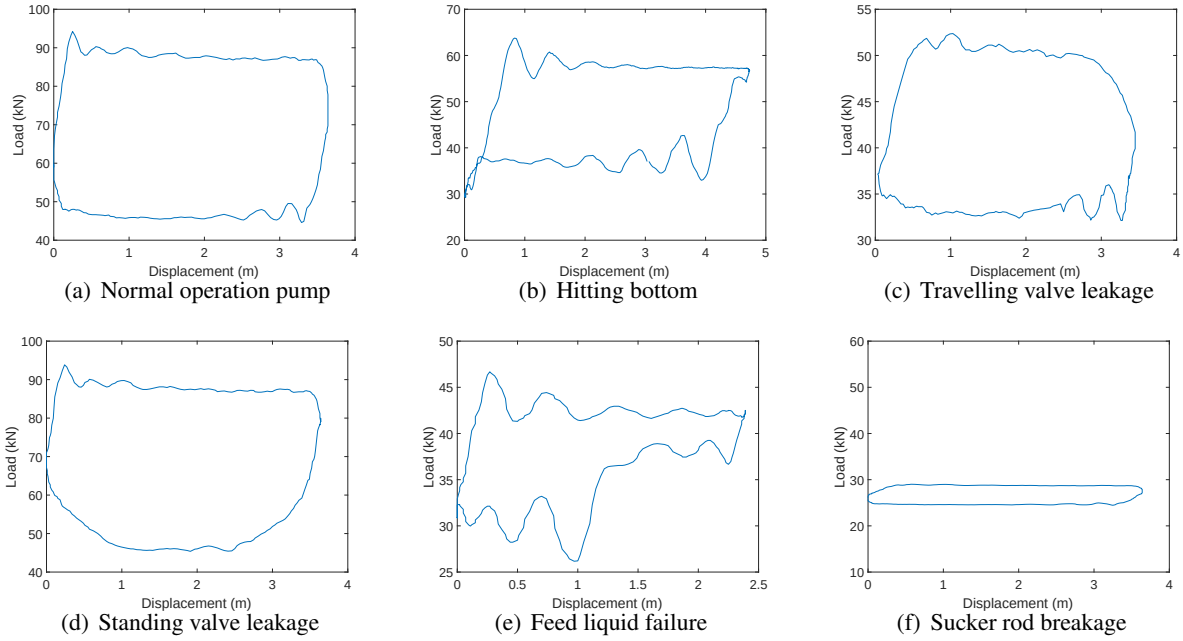
2. Dynamometer card feature extraction based on Discrete Curvelet

2.1. Dynamometer card and its properties

The dynamometer card is a closed curve of load versus displacement. Different working conditions of sucker rod pumping systems can be represented by the shapes of the dynamometer cards. For examples, the condition "normal operation pump" is reflected by a closed curve of approximate parallelogram; the lower left corner of "hitting bottom" has an extra circular pattern; for "travelling valve leakage", the upper portion of the curve is like a parabola; on the contrary, for "standing valve leakage", its shape is a downward arch; The curves of "feed liquid failure" lack the right-bottom corner; and for "sucker rod breakage", its shape is a flat strip curve. Some main reference patterns of conditions are shown in Fig. 1. It can be seen that the differences among dynamometer cards of different conditions are obvious. Therefore, recognizing the shapes of dynamometer cards can distinguish the normal and fault conditions.

2.2. Fast Discrete Curvelet Transform

Curvelet Transform proposed by J.C. Emmanuel, D.L. Donoho[24] is one kind of Multi-resolution methods, it has been widely used for feature extraction because of its ability to represent images at different scales and angles. The discrete curvelet transform provides a decomposition of the frequency space



(a) Normal operation pump (b) Hitting bottom (c) Travelling valve leakage

(d) Standing valve leakage (e) Feed liquid failure (f) Sucker rod breakage

Fig. 1. Classes of dynamometer card patterns.

into "wedges". A wedge is defined by the support of the radial window and the angular window[25]. The radial window is given by:

$$\widetilde{W}_j(\omega) = \sqrt{\Phi^2_{j+1}(\omega) - \Phi^2_j(\omega)}, j \geqslant 0 \qquad (1)$$

where Φ is defined as the inner product of low-pass 1D window:

$$\Phi_j(\omega_1, \omega_2) = \varphi(2^{-j}\omega_1)\varphi(2^{-j}\omega_2), 0 \leqslant \varphi \leqslant 1 \quad (2)$$

The angular window is defined as follows:

$$V_j(\omega) = V(2^{[j/2]}\omega_2/\omega_1) \qquad (3)$$

By introducing the set of equispaced slopes as $\tan\theta_l = l \cdot 2^{-[j/2]}, l = -2^{[j/2]}, ..., 2^{[j/2]} - 1$, the wedge in the cartesian coordinate system can be described as:

$$\widetilde{U}_{j,l}(\omega) = \widetilde{W_j}(\omega) V_j(S_{\theta_L}\omega) \qquad (4)$$

where S_{θ_L} is the shear matrix of the form:

$$S_{\theta_L} = \begin{pmatrix} 1 & 0 \\ -\tan\theta & 1 \end{pmatrix} \qquad (5)$$

Thus, the expression of Discrete Curvelet transform can be presented as followed:

$$C(j,l,k) = \int \widehat{f}(\omega)\widetilde{U}_j(S^{-1}_{\theta_l}\omega)e^{i<b,\omega>}d\omega \qquad (6)$$

where $b \approx (k_1 2^{-j}, k_2 2^{-j/2})$, $\widehat{f}$ is the 2D FFT of the image and j, l, k are the parameters of the scale, the direction and the position, respectively.

In this paper, the Fast Discrete Curvelet transform (FDCT) via Unequispaced Fast Fourier Transform[26] is utilized to extract the features of dynamometer cards.

2.3. *Features extracted by Fast Discrete Curvelet transform*

A five-level scale fast discrete curvelet transform is used for dynamometer cards. Every card is a 256×256 gray image. After the transformation, the image composed of the curvelet coefficients at first four scales is shown in Fig. 2. The cartesian concentric coronas show the coefficients at different scales. The low frequency coefficients are stored in the center of the image while the high frequency coefficients are stored in the outer corona[27]. There

are four strips in each corona and the strips are further subdivided in angular wedges. The numbers of wedges of all five scales are 1, 32, 32, 64 and 1. Every wedge can represent coefficients at a specified scale and orientation.

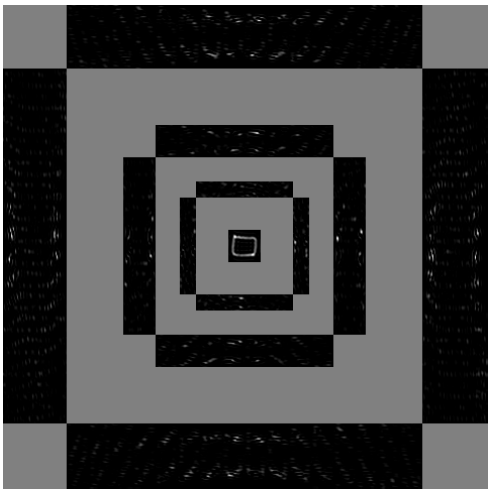


Fig. 2. Curvelet coffients of a dynamometer card at scale $j = 0,1,2,3$ in multiple directions.

The energy of each wedge[28] can be adopted as features:

$$E_{j,l} = \sum_{k} |C(j,l,k)|^2 \tag{7}$$

Features of six typical dynamometer cards are shown in Fig. 3.

3. Sparse multi-graph regularized extreme learning machines

3.1. Extreme Learning Machine

Extreme Learning Machine is a kind of single-hidden layer feed-forward neural networks whose input weights and the hidden layer biases are randomly generated. Assuming that the number of hidden neurons is L, and the output function for an input $\mathbf{X}$ is expressed as:

$$f_L(\mathbf{X}) = \sum_{i=1}^{L} \beta_i \mathbf{h}_i(\mathbf{X}) = \mathbf{h}(\mathbf{X})\beta \tag{8}$$

where $\beta = [\beta_1, \ldots, \beta_L]$ is the output weight vector between the hidden layer and the output layer. $\mathbf{h}(\mathbf{X}) = [h_1(\mathbf{X}), ..., h_L(\mathbf{X})]$ is a nonlinear mapping function, each $h_i(\mathbf{X})$ is defined as:

$$h_i(\mathbf{X}) = G(\mathbf{a}_i, b_i, \mathbf{X}) \tag{9}$$

where $G(\cdot)$ denotes an active function such as a sigmoid function or a Gauss function. $\mathbf{a}_i$ is the input vector between the ith hidden node and the inputs nodes, b_i is the bias term.

Let $\mathbf{Y}$ denotes the label matrix, traditional ELM aims to minimize the following objective function:

$$\min_{\beta} ||\mathbf{H}\beta - \mathbf{Y}||^2 \tag{10}$$

where $\mathbf{H}$ is the hidden layer output matrix denotes as:

$$\mathbf{H} = \begin{bmatrix} h_1(\mathbf{x}_1) & h_2(\mathbf{x}_1) & \cdots & h_L(\mathbf{x}_1) \\ h_1(\mathbf{x}_2) & h_2(\mathbf{x}_2) & \cdots & h_L(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ h_1(\mathbf{x}_N) & h_2(\mathbf{x}_N) & \cdots & h_L(\mathbf{x}_N) \end{bmatrix} \tag{11}$$

The output weight vector can be obtained as the optimal solution of (10):

$$\widehat{\beta} = \mathbf{H}^{\dagger}\mathbf{Y} \tag{12}$$

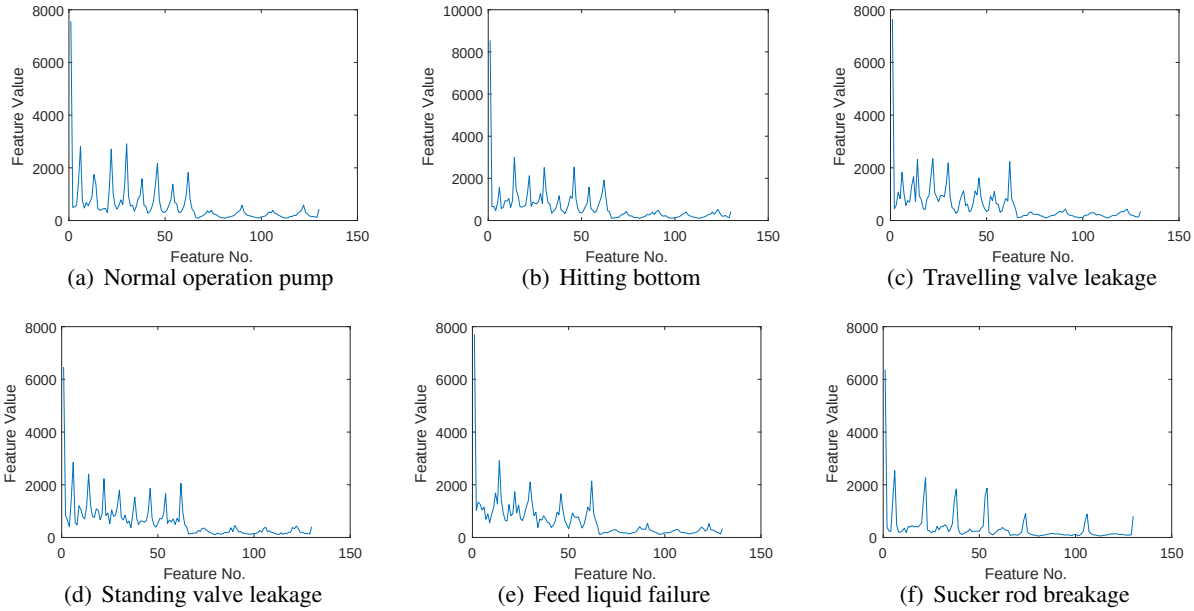
where $\mathbf{H}^{\dagger}$ is the MoorePenrose generalized inverse of $\mathbf{H}$.

A regularization term is introduced to avoid the singularity problem when calculating $\mathbf{H}^{\dagger}$. Therefore, the objective function of regularized ELM can be rewritten as:

$$\begin{aligned} &\min_{\beta} \tfrac{1}{2}||\beta||^2 + \tfrac{\lambda}{2}\sum_{i=1}^{N}||\xi_i||_2^2 \\ &\text{s.t.} : \xi_i = \mathbf{y}_i - \mathbf{h}(\mathbf{x}_i)\beta \end{aligned} \tag{13}$$

where ξ_i is the error vector and $\lambda > 0$ is a regularization parameter. The output weight vector β in regularized ELM can be estimated as:

$$\widehat{\beta} = (\mathbf{H}^T\mathbf{H} + \frac{1}{\lambda})^{-1}\mathbf{H}^T\mathbf{Y} \tag{14}$$



(a) Normal operation pump (b) Hitting bottom (c) Travelling valve leakage

(d) Standing valve leakage (e) Feed liquid failure (f) Sucker rod breakage

Fig. 3. Features of six typical dynamometer cards

3.2. *Sparse representation*

Traditional signal representations such as Fourier representation and Wavelet representation assume that an observation signal y can be modeled as a linear combination of a set of basis functions ψ_i:

$$y = \sum_{i=1}^{N} \alpha_i \psi_i \tag{15}$$

where α_i is the combination coefficient.

As an extension to these classical methods, sparse representation constructs the signal with an over-complete set of basis functions and most coefficients are set to zero. J. Wright *et al*[29] utilized the training samples $\mathbf{A} = [x_1, x_2, \ldots, x_N] \in R^{d \times N}$ as the basis functions to represent the observation sample as:

$$y = \mathbf{A}\alpha \tag{16}$$

where $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_N] \in R^{N \times 1}$ is the coefficient vector.

The sparsest coefficient vector can be obtained by solving the following optimization problem:

$$\begin{aligned} &\widehat{\alpha} = \arg\min ||\alpha||_1 \\ &\text{s.t.} : y = \mathbf{A}\alpha \end{aligned} \tag{17}$$

In practice, the dimension of data is always much larger than the number of samples, so it is hard to get the over-complete dictionary. In this case, the sparsest coefficient vector can be obtained by simultaneously minimizing both the reconstruction error and the norm of coefficient vector, and turned to solve the following optimization problem:

$$\begin{aligned} &\widehat{\alpha'} = \arg\min ||\alpha'||_1 \\ &\text{s.t.} : y = \mathbf{A}'\alpha' \end{aligned} \tag{18}$$

where $\mathbf{A}' = [\mathbf{A}, \mathbf{I}] \in R^{d \times (N+d)}$.

3.3. *Sparse multi-graph regularized extreme learning machines*

M. Belkin *et al*[30] proposed a manifold framework to estimate unknown functions:

$$\widehat{f} = \arg\min[V + \theta||f_H||^2 + \eta||f_I||^2] \tag{19}$$

where V is the loss function, f_H represents the complexity of the unknown functions and f_I represents the inherent structure of the samples. θ and η are the tradeoff parameters. Suppose that we have a training set $\mathbf{X} = \{\mathbf{X}^1, \mathbf{X}^2, \ldots, \mathbf{X}^C\} = \{x_1, x_2, \ldots, x_N\}$ of N samples, where $x_i \in R^d, i = 1, 2, \ldots, N$ and d is the

dimensionality. There are C classes and each class has N_k samples, where $k = 1,2,\ldots,C$. In traditional methods, The inherent structure can be expressed by constructing one graph $\mathbf{G} =< \mathbf{V},\mathbf{E},\mathbf{W} >$, where $\mathbf{V}$ is the vertex set in which each vertex represents a data x_i, $\mathbf{E}$ is the edge set in which each edge means the relationship between two vertexes, and the weights of the edges form the weight matrix of $\mathbf{W}$.

To improve the discriminative ability of ELM, motived by manifold learning methods[21,22], we construct two graphs: the intra-class graph $\mathbf{G}_{\text{intra}} =< \mathbf{X},\gamma,\mathbf{R} >$ and the inter-class graph $\mathbf{G}_{\text{inter}} =< \mathbf{X},\delta,\mathbf{D} >$.

The intra-class graph can reflect the similarity among data of same class. For every sample x_i^m in the mth class, its corresponding over-complete dictionary includes all the samples in $\mathbf{X}^m$ except x_i^m i.e. $\mathbf{X}_i^m = [x_1^m, x_2^m, \ldots, x_{i-1}^m, x_{i+1}^m, \ldots, x_{N_k}^m] \in R^{d\times(N_m-1)}$, x_i^m can be represented as:

$$\begin{aligned} &\widehat{\gamma}_i^m = \arg\min ||\gamma_i^m||_1 \\ &\text{s.t.} : x_i^m = \mathbf{X}_i^m \gamma_i^m \end{aligned} \tag{20}$$

where $\gamma_i^m = [\gamma_{i,1}^m, \gamma_{i,2}^m, \ldots, \gamma_{i,i-1}^m, \gamma_{i,i+1}^m, \ldots, \gamma_{i,N_m}^m] \in R^{(N_m-1)\times 1}$ is the coefficient vector. The intra-class weight matrix $\mathbf{R}$ can be calculated as:

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}^1 & 0 & \cdots & 0 \\ 0 & \mathbf{R}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{R}^C \end{bmatrix} \tag{21}$$

where the matrix $\mathbf{R}^m$ denotes the similarity among samples in mth class. Elements of $\mathbf{R}^m$ can be calculated as:

$$r_{ij}^m = \begin{cases} \widehat{\gamma}_{i,j}^m, & i > j \\ 0, & i = j \\ \widehat{\gamma}_{i,j-1}^m, & i < j \end{cases} \tag{22}$$

To improve the discriminative ability, we hope that the output results of similar samples from the same class are as close as possible:

$$\begin{aligned} \min f_{intra} &= \frac{1}{2}\sum_{i,j} r_{ij} ||f(x_i) - f(x_j)||^2 \\ &= \sum_i f(x_i)^2 t_{ii} + \sum_{i,j} f(x_i) f(x_j) r_{ij} \\ &= \mathbf{f}^T\mathbf{T}\mathbf{f} - \mathbf{f}^T\mathbf{R}\mathbf{f} = \mathbf{f}^T\mathbf{L}_{\text{intra}}\mathbf{f} \end{aligned} \tag{23}$$

where $\mathbf{L}_{\text{intra}} = \mathbf{T} - \mathbf{R}$, $\mathbf{T}$ is a diagonal matrix that can be calculated as:

$$t_{ii} = \sum_j r_{ij} \tag{24}$$

The inter-class graph can represent the relationship among data in different classes. For every x_i^m from the mth class, its corresponding over-complete dictionary includes all data except the samples from mth class i.e. $\mathbf{X}^{-m} = \{\mathbf{X}^1, \mathbf{X}^2, \ldots, \mathbf{X}^{m-1}, \mathbf{X}^{m+1}, \ldots, \mathbf{X}^C\} \in R^{d\times(N-N_m)}$, x_i^m can be represented as:

$$\begin{aligned} &\widehat{\delta}_i^m = \arg\min ||\delta_i^m||_1 \\ &\text{s.t.} : x_i^m = \mathbf{X}^{-m}\delta_i^m \end{aligned} \tag{25}$$

where $\delta_i^m = [\delta_{i,1}^{m,1}, \delta_{i,2}^{m,1}, \ldots, \delta_{i,N_1}^{m,1}, \delta_{i,1}^{m,2}, \delta_{i,2}^{m,2}, \delta_{i,3}^{m,2}, \ldots, \delta_{i,N_2}^{m,2}, \ldots, \delta_{i,1}^{m,m-1}, \delta_{i,2}^{m,m-1}, \delta_{i,3}^{m,m-1}, \ldots, \delta_{i,N_{m-1}}^{m,m-1}, \delta_{i,1}^{m,m+1}, \delta_{i,2}^{m,m+1}, \delta_{i,3}^{m,m+1}, \ldots, \delta_{i,N_{m+1}}^{m,m+1}, \ldots, \delta_{i,1}^{m,C}, \delta_{i,2}^{m,C}, \delta_{i,3}^{m,C}, \ldots, \delta_{i,N_C}^{m,C}] \in R^{(N-N_m)\times 1}$ is the coefficient vector. The inter-class weight matrix $\mathbf{D}$ can be calculated as:

$$\mathbf{D} = \begin{bmatrix} 0 & \mathbf{D}^{2,1} & \mathbf{D}^{3,1} & \cdots & \mathbf{D}^{C,1} \\ \mathbf{D}^{1,2} & 0 & \mathbf{D}^{3,2} & \cdots & \mathbf{D}^{C,2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{D}^{1,C} & \mathbf{D}^{2,C} & \mathbf{D}^{3,C} & \cdots & 0 \end{bmatrix} \tag{26}$$

where $\mathbf{D}^{m,n}$ denotes the relationship among samples in mth class and samples in nth class. Elements in $\mathbf{D}^{m,n}$ can be calculated as:

$$d_{i,j}^{m,n} = \widehat{\delta}_{i,j}^{m,n} \tag{27}$$

To improve the discriminative ability, we hope that the output results of similar samples from different classes are as different as possible:

$$\begin{aligned} \max f_{\text{inter}} &= \frac{1}{2}\sum_{i,j} d_{ij} ||f(x_i) - f(x_j)||^2 \\ &= \sum_i f(x_i)^2 v_{ii} + \sum_{i,j} f(x_i) f(x_j) d_{ij} \\ &= \mathbf{f}^T\mathbf{V}\mathbf{f} - \mathbf{f}^T\mathbf{D}\mathbf{f} = \mathbf{f}^T\mathbf{L}_{\text{inter}}\mathbf{f} \end{aligned} \tag{28}$$

where $\mathbf{L}_{\text{inter}} = \mathbf{V} - \mathbf{D}$, $\mathbf{V}$ is a diagonal matrix that can be calculated as:

$$v_{ii} = \sum_j d_{ij} \tag{29}$$

However, the elements in the coefficient vector calculated by sparse representation are not guaranteed to be greater than zero. For the intra-class graph, the negative element will transform the objective function (23) into:

$$\max f_{intra} = \frac{1}{2}\sum_{i,j}|r_{ij}|||f(x_i)-f(x_j)||^2 \tag{30}$$

Eq. (30) means that, a larger $|r_{ij}|$ will make the output results $f(x_i)$ and $f(x_j)$ to be more different, although x_i and x_j are from the same class.

Similarly, for the inter-class graph, the negative element will transform the objective function (28) into:

$$\min f_{\text{inter}} = \frac{1}{2}\sum_{i,j}|d_{ij}|||f(x_i)-f(x_j)||^2 \tag{31}$$

Eq. (31) means that, a larger $|d_{ij}|$ will make the output results $f(x_i)$ and $f(x_j)$ to be more similar, although x_i and x_j are from different classes. The objective function (30) and (31) will obviously reduce the discriminative ability.

In classical methods, when x_j belongs to the set of k nearest neighbors of x_i, the edge between x_j and x_i is set to 1; otherwise, the edge between x_j and x_i is set to 0. The sparse coefficients calculated through sparse representation can reflect the relationship among samples. The greater coefficient corresponds to the more similar pairwise samples. When constructing graphs via sparse representation, a positive coefficient means that x_j and x_i are similar, so x_j can be regarded in the set of k nearest neighbors of x_i; a negative coefficient means that x_j and x_i are not similar so we can deem that x_j is not in the set of k nearest neighbors of x_i. Therefore, elements in intra-class graph and inter-class graph can be recalculated as:

$$r_{ij}^m = \begin{cases} p(\widehat{\gamma}_{i,j}^m), & i > j \\ 0, & i = j \\ p(\widehat{\gamma}_{i,j-1}^m), & i < j \end{cases} \tag{32}$$

and

$$d_{i,j}^{m,n} = p(\widehat{\delta}_{i,j}^{m,n}) \tag{33}$$

where

$$p(x) = \begin{cases} x, x > 0 \\ 0, x \leqslant 0 \end{cases} \tag{34}$$

Combine (23) and (28), $||f_I||^2$ in (19) has the following expression:

$$||f_I||^2 = \text{Tr}\left(\mathbf{f}^T(\mathbf{L}_{\text{inter}}{}^{-1/2})^T\mathbf{L}_{\text{intra}}(\mathbf{L}_{\text{inter}}{}^{-1/2})\mathbf{f}\right) \tag{35}$$

According to (8) and (10), choose the following loss function V:

$$V = ||\mathbf{H}\beta - \mathbf{Y}||^2 \tag{36}$$

where $\mathbf{Y} = [y_1, y_2, \ldots, y_N] \in R^{C\times N}$ is the label matrix, and $||f_H||^2$ in (19) has the following expression:

$$||f_H||^2 = ||\beta||^2 \tag{37}$$

Substituting (35) (36) (37) into (19):

$$\begin{aligned}\widehat{f} = \arg\min[&||\mathbf{H}\beta - \mathbf{Y}||^2 + \theta||\beta||^2 \\ &+ \eta\text{Tr}\left(\beta^T\mathbf{H}^T(\mathbf{L}_{\text{inter}}{}^{-1/2})^T\mathbf{L}_{\text{intra}}(\mathbf{L}_{\text{inter}}{}^{-1/2})\mathbf{H}\beta\right)]\end{aligned} \tag{38}$$

Differentiating it with respect to β, and the output weight vector β can be estimated as:

$$\begin{aligned}\widehat{\beta} = (&\eta\mathbf{H}^T(\mathbf{L}_{\text{inter}}{}^{-1/2})^T\mathbf{L}_{\text{intra}}(\mathbf{L}_{\text{inter}}{}^{-1/2})\mathbf{H} \\ &+ \mathbf{H}^T\mathbf{H} + \theta\mathbf{I})^{-1}\mathbf{H}^T\mathbf{Y}\end{aligned} \tag{39}$$

Based on the above discussion, the SMELM algorithm is summarized as **Algorithm 1**.

4. Experiments

4.1. Experiments results

In order to verify the correctness and validity of the proposed approach, experiments are carried out on dynamometer cards that sampled from LiaoHe oil fields, China. The dynamometer cards set includes six classes which are: "normal operation"; "hitting bottom"; "travelling valve leakage"; "standing valve leakage"; "feed liquid failure" and "sucker rod breakage". Each class has 30 samples. FDCT is adopted on each dynamometer card to extract features. We randomly select 25 samples per class as training samples and the rest samples are used for testing. Repeat the segmentation process for 15 times to make better estimates of the accuracy. The sigmoid function is selected as the active function

Algorithm 1 Sparse multi-graph regularized extreme learning machine algorithm

Input: Training set $\mathbf{X} = \{\mathbf{X}^1, \ldots, \mathbf{X}^C\} = \{x_1, \ldots, x_N\}$,$x_i \in R^d, i = 1, 2, \ldots, N$; the activation function G;the number of hidden nodes L; the parameters θ and η;

Output: Output weight vector β;

1: Randomly initialize the input vector α_i and the bias term b_i, $i = 1, 2, \ldots, L$;
2: Compute the hidden layer output matrix $\mathbf{H}$;
3: For every sample in training set $\mathbf{X}$, calculate its sparse representation according to (20) and (25);
4: Construct the intra-class weight matrix $\mathbf{R}$ according to (21) and (22);
5: Construct the inter-class weight matrix $\mathbf{D}$ according to (26) and (27);
6: Compute β according to by (39).

for SMELM. The whole procedure of fault diagnosis in sucker rod pumping systems is shown in Fig. 4.

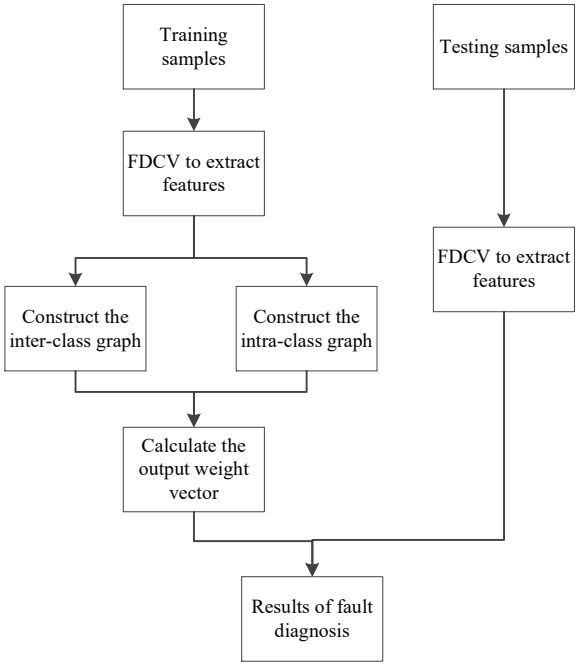


Fig. 4. The flow chart of fault diagnosis in sucker rod pumping systems.

The proposed method is compared with the algorithms of ordinary ELM, discriminative manifold extreme learning machine (DMELM), Support Vector Machine (SVM) and BP neural network (BPNN). Fig. 5 shows the accuracy curve of different algorithms over different splittings. It can be found the proposed SMELM gets the best accuracy. More specifically, with the help of the information of data's inherent structure, the accuracy curve of SMELM and DMELM are both higher than the one of ordinary ELM, SVM and BPNN. Utilizing the sparse representation method to construct graphs, SMELM achieves better discriminative ability than DMELM because the sparse coefficients can reflect the relationship among samples more effectively.

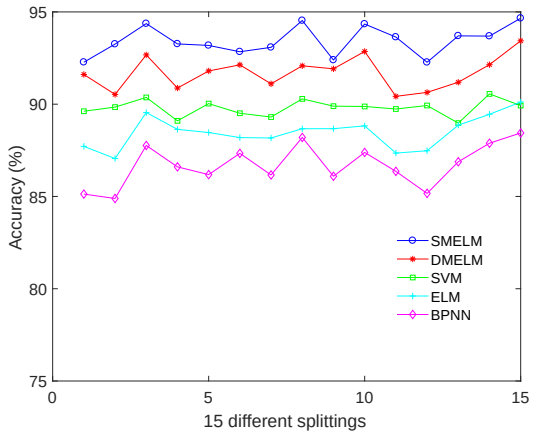


Fig. 5. Performance of different algorithms.

4.2. *Parameter sensitivity analysis*

There are three parameters L, θ and η in the proposed SMELM algorithm. Fig. 6 shows the performance of SMELM with respect to different number of hidden neurons L.

It can be concluded that in the case of less number of hidden neurons, the accuracy arises quickly when the number of the hidden neurons increases; in the case of large number of hidden neurons, the

accuracy is no more sensitive to the increasing of hidden neurons. Therefore, in our experiments, the number of hidden neurons is simply set as 5 times of the feature dimension i.e. $L = 650$.

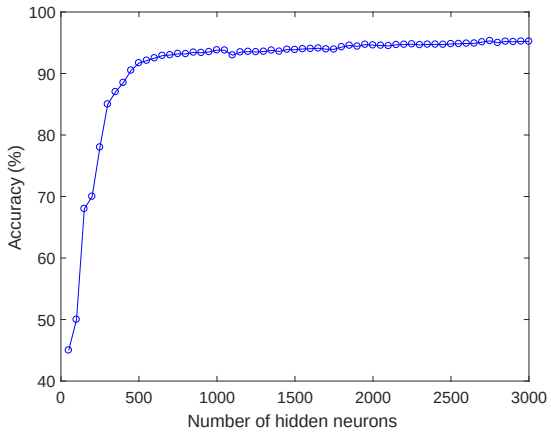


Fig. 6. Performance according to the number of neurons.

As for the remaining two parameters: θ and η, we vary their values from the following exponential sequence $\{10^{-10},\ldots,10^{10}\}$. Fig. 6 shows performance with regard to different combinations of θ and η. It can be found that the optimal values of the parameters are near $\eta = 1$ and $\theta = 10^{-4}$, and there is a large flat area near the optimal values. This means SMELM is not very sensitive to the combination of parameters θ and η. A relatively large η can emphasize the samples' inherent structure information to achieve better performance.

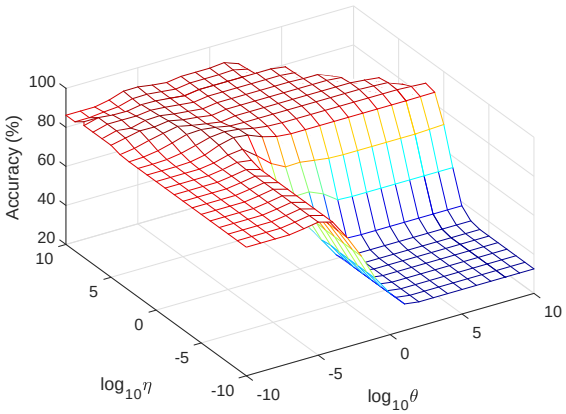


Fig. 7. Performance with regard to different combinations of θ and η.

In this paper, we fixed $L = 650$, $\eta = 1$ and $\theta = 10^{-4}$ in the previous experiments.

5. Conclusion

In this paper, a novel approach is proposed for faults diagnosis in sucker rod pumping systems. First, FDCV is employed to extract features of dynamometer cards. Then, a novel model named sparse multi-graph regularized extreme learning machine is proposed and applied as a classifier. In SMELM, two graphs are constructed to explore the inherent structure of data more exactly. The inter-class graph shows the similarity among data from the same class, and the inter-class graph shows the relationship among data from different classes. With the help of these two graphs, the discriminative ability of the ordinary ELM is enhanced. Each graph is obtained through the sparse representation algorithm to avoid the difficulty of choosing the appropriate parameter. The experimental results demonstrate the effectiveness of our method for fault diagnosis in sucker rod pumping systems.

Acknowledgements

This work is partially supported by the National Natural Science Foundation of China (No. 61573088, No. 61573087 and No. 61433004).

References

1. H.A.Tripp, A review: analyzing beam-pumped wells, *Journal of Petroleum Technology.* **41** (05) (1989) 457–458.
2. F. Ding and D.X.Zhou, Neural network expert system for rod pumping diagnosis, *China Petroleum Machinery.* **30** (1996) 1–4.
3. J.P. Wang and Z.F. Bao, Study on Pump Fault Diagnosis Based on Rough Sets Theory, in *The 3rd International Conference on Innovative Computing Information and Control*, (IEEE, Harbin, 2008), pp. 288-288.
4. Y.F. He, X.D. Wu, W. Xiao, W.C. Li and X.L. Yu, Frequency spectrum analysis method for recognition of dynamometer card, *Acta Petrolei Sinica.* **29** (4) (2008) 619–623.
5. P. Xu, S.J. Xu and H.W. Yin, Application of self-organizing competitive neural network in fault diagnosis of suck rod pumping system, *Journal of*

Petroleum Science & Engineering. **58** (12) (2007) 43–48.
6. W. Wu, W.L. Sun and H.X. Wei, A Fault Diagnosis of Suck Rod Pumping System Based on Wavelet Packet and RBF Network, *Advanced Materials research.* 189-193 (2011) 2665–2669.
7. K. Li, X.W. Gao, Z.D. Tian and Z.X. Qiu, Using the curve moment and the PSO-SVM method to diagnose downhole conditions of a sucker rod pumping unit, *Petroleum Science.* **10** (1) (2013) 73–80.
8. D.O. Bruno, M.Z.D. Nascimento, R.P. Ramos, V.R. Batista, L.A. Neves and A.S. Martins, LBP operators on curvelet coefficients as an algorithm to describe texture in breast cancer tissues, *Expert Systems with Applications*. **55** (2016) 329–340.
9. F. Li, F. Fang and G. Zhang, Unsupervised change detection in SAR images using curvelet and L1-norm based soft segmentation, *International Journal of Remote Sensing.* **37** (14) (2016) 3232–3254.
10. Y. Li, Q. Yang and R. Jiao, Image compression scheme based on curvelet transform and support vector machine, *Expert Systems with Applications.* **37** (4) (2010) 3063–3069.
11. G.B. Huang, Q.Y. Zhu and C.K. Siew, Extreme learning machine: Theory and applications, *Neurocomputing.* **70** (13) (2006) 489–501.
12. S.T. Roweis and L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, *Science.* **290** (5500) (2000) 2323–2326.
13. J.B. Tenenbaum, V.D. Silva and J.C. Langford, A Global Geometric Framework for Nonlinear Dimensionality Reduction, *Science.* **290** (5500) (2000) 2319–2323.
14. M. Belkin and P. Niyogi, Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering, *Advances in Neural Information Processing systems.* **14** (6) (2002) 585–591.
15. Z.Y. Zhang and H.Y. Zha, Principal Manifolds and Nonlinear Dimensionality Reduction via Tangent Space Alignment, *Journal of Shanghai University.* **8** (4) (2005) 406–424.
16. G. Huang, S.J. Song, J.N. Gupta and C. Wu, Semi-supervised and unsupervised extreme learning machines, *IEEE Transactions on Cybernetics.* **44** (12) (2014) 2045–2417.
17. J. Liang, F.Y. Zhang, X.X. Xiong, X.B. Chen, L. Chen, and G.H. Lan, Manifold Regularized Proximal Support Vector Machine via Generalized Eigenvalue, *International Journal of Computational Intelligence Systems*. **9** (6) (2016) 1041–1054.
18. S.L. Sun and X.J. Xie, Semisupervised Support Vector Machines With Tangent Space Intrinsic Manifold Regularization, *IEEE Transactions on Neural Networks & Learning Systems.* **27** (9) (2016) 1827–1839.
19. L. Zhang and W.D. Zhou, Fisher-regularized support vector machine, *Information Sciences.* 343-344 (2016) 79–93.
20. Y. Peng and B.L. Lu, Discriminative manifold extreme learning machine and applications to image and EEG signal classification, *Neurocomputing.* **174** (PA) (2015) 265–277.
21. L.S. Qiao, S.C. Chen and X.Y. Tan, Sparsity preserving projections with applications to face recognition, *Pattern Recognition.* **43** (1) (2010) 331–341.
22. S.J. Lou, X.M. Zhao, Y.L. Chuang, H.T. Yu and S.Q. Zhang, Graph Regularized Sparsity Discriminant Analysis for face recognition, *Neurocomputing.* **173** (2) (2016) 290–297.
23. Y. Peng, S.H. Wang, X.Z. Long and B.L. Lu, Discriminative graph regularized extreme learning machine and its application to face recognition, *Neurocomputing.* **149** (PA) (2015) 340–353.
24. J.C. Emmanuel and D.L. Donoho, Curvelets - A Surprisingly Effective Nonadaptive Representation For Objects with Edges, *Astronomy & Astrophysics.* **283** (3) (2000) 1051–1057.
25. A.D. Rahulkar, D.V. Jadhav and R.S. Holambe, Fast discrete curvelet transform based anisotropic iris coding and recognition using k-out-of-n: A, fused post-classifier, *Machine Vision and Applications.* **23** (6) (2012) 1115–1127.
26. E. Cands, L. Demanet, D. Donoho and L. Ying, Fast Discrete Curvelet Transforms, *Multiscale Modeling & Simulation.* **5** (3) (2006) 861–899.
27. C.V. Rao, J.M. Rao, A.S. Kumar, D.S. Jain, V.K. Dadhwal, Satellite Image Fusion using Fast Discrete Curvelet Transforms, in *Advance Computing Conference*, (IEEE, Gurgaon, 2014), pp. 952–957.
28. K. Feng, Z Jiang, W. He and B. Ma, A recognition and novelty detection approach based on Curvelet transform, nonlinear PCA and SVM with application to indicator diagram diagnosis, *Expert Systems with Applications.* **38** (10) (2011) 12721-12729.
29. J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry and Y. Ma, Robust face recognition via sparse representation, *IEEE transactions on pattern analysis and machine intelligence.* **31** (2) (2009) 210–227.
30. M. Belkin, P. Niyogi and V. Sindhwani, Manifold Regularization: A Geometric Framework for Learning from Labeled and Unlabeled Examples, *Journal of Machine Learning Research.* **7** (1) (2006) 2399–2434.