
Degradation state mining and identification for railway point machines

Chong Bian^a, Shunkun Yang^{b,*}, Tingting Huang^b, Qingyang Xu^c, Jie Liu^b, Enrico Zio^{d,e,f}

^aSchool of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China.

^bSchool of Reliability and Systems Engineering, Beihang University, Beijing 100191, China.

^cInfrastructure Inspection Research Institute, China Academy of Railway Sciences, Beijing 100081, China.

^dChair on system science and energetic challenges, EDF Foundation, Laboratoire Genie Industriel, CentraleSupélec, Université Paris-Saclay, 3 rue Joliot Curie, 91190 Gif-sur-Yvette, France.

^eEnergy Department, Politecnico di Milano, Piazza Leonardo da Vinci, 32, 20133 Milano, Italy.

^fSino-French Risk Science and Engineering Lab, Beihang University, Beijing 100191, China.

*Corresponding Author. Tel.: +86-13641236614; E-mail address: ysk@buaa.edu.cn

Abstract

Critical point machine failure in railway-signal systems can lead to fatal accidents. Hence, early identification of anomalies is vital in guaranteeing reliable and safe transportation. However, most of the existing early fault diagnosis methods can only estimate the degradation trend under a specific fault mode. How to analyze the diversified degradation conditions under multiple fault modes is still a key problem. Considering the diversity of fault modes, this study proposes an early fault diagnosis

method based on self-organizing feature map network and support vector machine, focusing on the use of non-fault data to simultaneously mine and accurately identify degradation states under different fault modes, to provide guidance for proactive machine maintenance. The experimental results obtained via application of this scheme to field data for railway point machines demonstrate that the proposed methodology can effectively mine and accurately identify degradation states with different machine characteristics.

Keywords: Degradation state mining, degradation state identification, early fault diagnosis, railway point machine, self-organizing mapping, support vector machine

1. Introduction

Railway point machines are the critical electromechanical apparatus used in signal systems for changing the train directions by switching and locking turnouts. During a field investigation of one railway station in China, point machine operation faults were observed to account for 30% of all signal equipment failures. The average time required for point machine repair was approximately 23.4 min, which seriously affected the reliability and availability of railway operation. Many railway companies employ condition-monitoring systems for point machines to manage the measurement, centralization, and analysis of data collected from the sensors installed in field equipment to detect failures [1]. However, these systems only sound alarms when failure either has already occurred or is about to, which is too late for effective proactive

intervention [2]. Therefore, many early diagnostic techniques are employed to detect, locate, and identify potential abnormalities and faults to minimize performance degradation and avoid overall system damage and collapse [3]. Based on monitoring data, these methods can be used to capture hidden fault information through data analysis techniques for diagnosis and provide trade-offs in terms of complexity, cost, and accuracy [4–6]. Machine performance degradation is often considered to achieve early fault diagnosis, especially in some data-driven methods [7–11].

Considerable work for railway point machines has been published in the last decade to ensure that defects can be detected in advance. These techniques can be divided into statistical analysis [12–14], classification [15–17], and model-based methods [18–20]. Some early fault diagnosis studies have also been proposed regarding point machine performance degradation. Atamuradov et al. [21–23] performed a series of investigations on fault detection, diagnosis, and degradation assessment for point machine sliding chair failure. First, the effective time-domain statistical features of point machine sensing signals are selected based on the evaluation metrics. A fusion algorithm is then used to construct a generic health indicator for the point machine. Finally, different learning algorithms are used to analyze the indicator data to achieve degradation monitoring and early fault identification. Ardakani et al. [24] proposed a health assessment method to analyze degradation and diagnose abnormalities in advance. Principal component analysis (PCA) and the Hotelling T^2 coefficient were used to select an energy-related feature set from point machine electrical signals to effectively evaluate the health state. Letot et al. [25] proposed a degradation trend

assessment model by utilizing the statistical features of active power signals for point machine reliability calculation and faulty behavior diagnosis.

However, the abovementioned methods can only estimate the degradation trends and determine the degradation levels under a specific fault mode. How to analyze the diversified degradation conditions under different fault modes is still a key problem. Therefore, this paper proposes a new early fault diagnosis method based on self-organizing feature map (SOM) network and support vector machine (SVM), focusing on the use of non-fault data to mine and identify degradation states under different fault modes, to provide guidance for proactive machine maintenance. The main contributions of this research can be summarized as follows:

- 1) A feature selection scheme is proposed for effective signal feature processing. Discriminant- and correlation-based feature selections are integrated to obtain feature indicators that can effectively represent a machine's health state.
- 2) A SOM-based degradation state mining method is proposed for different degradation conditions. SOM is employed to automatically perform multiple clustering analysis on the non-fault feature set to simultaneously label the degradation states of a machine under different fault modes.
- 3) A particle swarm optimization (PSO)-SVM-based classification model is proposed to accurately identify the different degradation states for early fault diagnosis. In this model, the SVM is optimized by PSO to build the classifier and realize optimized degradation state identification.
- 4) The abovementioned methods were applied to the SIEMENS S700K point machine,

which is widely used for high-speed railway, to prove its effectiveness and applicability.

The rest of this paper is organized as follows. Section 2 explains the proposed methods. The case study on a real industry railway point machine is presented in Section 3. The observed experimental results are discussed in Section 4. Section 5 summarizes the conclusions and the future work.

2. Proposed methodology

A machine usually goes through various degradation states, from initial deterioration to complete failure. Targeted maintenance can be implemented to prevent machine failure if the degradation state is identified in a timely manner. However, most of the current early fault diagnosis methods usually determine the degradation states under a specific fault mode [26]. In the actual situation, the machine has multiple fault modes, resulting in the diversity of the degradation. Therefore, for early fault diagnosis, the degradation trend and level under a specific fault mode should be horizontally estimated, and the degradation conditions under different fault modes should be vertically analyzed (Fig. 1). How to simultaneously mine the degradation states under different fault modes and accurately identify them is a key problem to be solved.

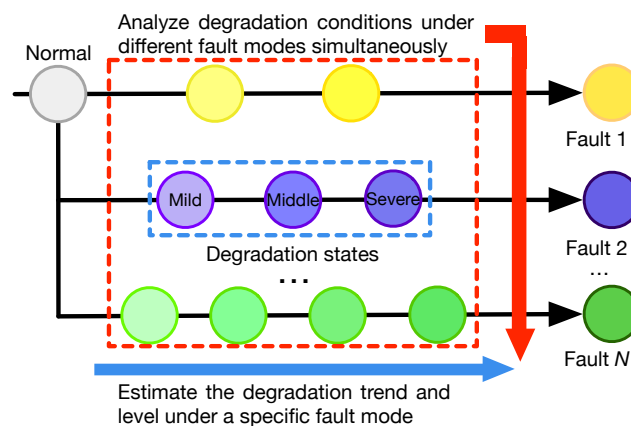


Fig. 1. Diagram of the degradation analysis aspect.

To solve the above problem, a method based on SOM and SVM is proposed in this study for early fault diagnosis. This method consists of four steps, as shown in Fig. 2.

(1) Data acquisition: The monitoring system is used to acquire various signal sample sets of the machine. (2) Feature processing: The non-fault data set is processed to obtain a low-dimensional feature set representing machine degradation. (3) State mining: Based on SOM, multiple clustering analysis of the non-fault feature set is conducted to mine degradation states under different fault modes simultaneously. (4) State identification: The classification model is built based on PSO-SVM to identify degradation states with different characteristics.

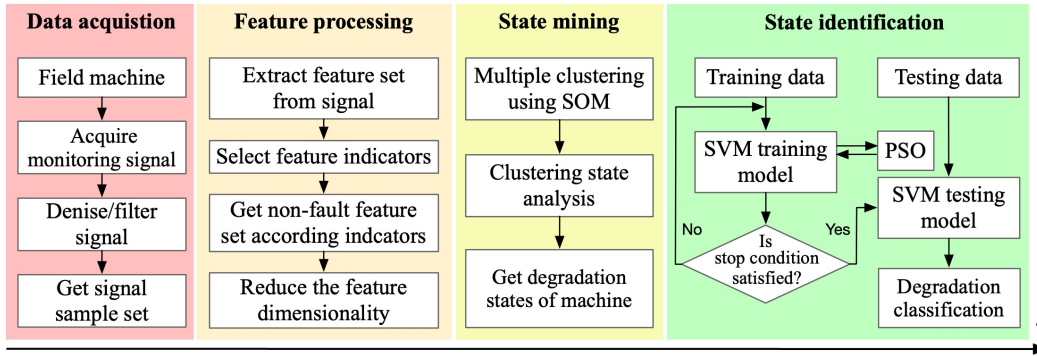


Fig. 2. Framework of the proposed method.

2.1 Feature processing for non-fault data

Concise and effective feature indicators should be used to represent the degradation information of non-fault samples. In this section, a combined strategy with three main parts is proposed (i.e., feature extraction, selection, and dimensionality reduction) for feature processing of non-fault data.

2.1.1 Feature extraction from time- and value-domain

In this study, ten signal parameters, including the out-to-in value, maximum difference, mean value, root-mean-square difference, variance, sum of difference,

kurtosis, crest factor, form factor, and impulse factor, are extracted from different time phases as the statistical features; some of these parameters were used in [14,23,27]. $t_{i,j}$ ($i = 1, \dots, P; j = 1, \dots, 10$) is the j th statistical feature extracted from the i th time phase of a signal sample, and P is the number of time phases of the signal. However, for some degenerative and failure symptoms of the machine, the features are concentrated in specific value segments. If signal features are extracted only from the time domain, important features describing the machine state will be omitted. To solve this problem, we projected the signal curve into the value domain and analyzed the different value domains in segments. Eight parameters were extracted within each segment to represent statistical value-domain features that completely describe machine operating conditions: maximum time value, mean value, number of data points, maximum difference, median, maximum value, time median, and mode. $v_{m,n}$ ($m = 1, \dots, S; n = 1, \dots, 8$) is the n th statistical feature extracted from the m th value-domain segment of a signal sample, and S is the number of value-domain segments of the signal. A D -dimensional feature vector x can be extracted from the signal sample and expressed as

$$x = (t_{1,1}, \dots, t_{i,j}, \dots, t_{P,10}, v_{1,1}, \dots, v_{m,n}, \dots, v_{S,8}), \quad (1)$$

where $D = 10P + 8S$.

2.1.2 Discriminant- and correlation-based feature selection

The degradation state of a machine eventually shifts to a fault state as the runtime increases. Meanwhile, the differences between the degradation and fault features gradually diminish and eventually disappear. The fault features can, therefore, be considered indicators for measuring degradation. By extracting the statistical features

of the machine fault samples, a D -dimensional feature set can be obtained. However, some features within such a set do not effectively represent machine fault state. Consequently, one must first select features from the set. Compared to the normal state, the fault state has some features that tend to deviate from their normal values, and the greater the deviation, the more obvious the discriminability of features corresponding to that fault. In this study, the features between machine fault and the normal state were distinguished based on the Fisher criterion [28], and the redundancy between features was measured based on the correlation coefficient. By eliminating the irrelevant and redundant features of the fault states, the features indicating machine abnormalities were selected to provide the basis for feature processing of non-fault data. The steps of feature selection are as follows.

(1) Calculate the intra- and inter-class variances of statistical features between the fault and normal states:

$$S_W^{(d)} = (m_F^{(d)} - m_N^{(d)})^2, \quad d = 1, \dots, D \quad (2)$$

$$S_B^{(d)} = (\sigma_F^{(d)})^2 + (\sigma_N^{(d)})^2, \quad d = 1, \dots, D, \quad (3)$$

where $S_W^{(d)}$ and $S_B^{(d)}$ are the intra- and inter-class variances of the d th-dimensional feature, respectively; $m_F^{(d)}$ is the mean value of the d th-dimensional feature of the fault state; $m_N^{(d)}$ is the mean value of the d th-dimensional feature of the normal state; $\sigma_F^{(d)}$ is the standard deviation of the d th-dimensional feature of the fault state; and $\sigma_N^{(d)}$ is the standard deviation of the d th-dimensional feature of the normal state.

(2) Calculate the criterion between the fault and normal states:

$$J^{(d)} = S_B^{(d)} / S_W^{(d)}, \quad d = 1, \dots, D, \quad (4)$$

where $J^{(d)}$ denotes the criterion value of the d th-dimensional feature between the fault

and normal states.

(3) For the obtained D criterion values, half of the maximum values are considered to be standard, and accordingly, features with criterion values larger than the corresponding standard are selected as effective features.

(4) Calculate the correlation coefficient between the selected features of the fault state:

$$\rho_{pq} = \frac{\sum_i (f_{S_i}^{(p)} - m_F^{(p)})(f_{S_i}^{(q)} - m_F^{(q)})}{\sqrt{(\sigma_F^{(p)})^2} \sqrt{(\sigma_F^{(q)})^2}}, \quad (5)$$

where ρ_{pq} is the correlation coefficient between the p th- and q th-dimensional features of the fault state and $f_{S_i}^{(p)}$ is the p th-dimensional feature value of the i th fault sample.

(5) If the correlation coefficient between pairwise features exceeds 0.95, the feature with a higher criterion value is selected. Eventually, all of the remaining features translate into selected effective features.

2.1.3 Feature dimensionality reduction based on KPCA

Regarding the SOM and SVM models, the use of a high-dimensional dataset for learning would lead to overfitting problems, thereby affecting the recognition accuracy [29]. Hence, the feature set should be further reduced to an appropriate dimensionality. Kernel PCA (KPCA) is a dimensionality reduction method, and its basic principle involves nonlinearly mapping the input vector onto the high-dimensional feature space and using PCA to calculate the principal component (PC) of the data within this feature space [30]. For high-dimensionality and non-linear characteristics of the non-fault feature set, KPCA was used in this study to extract PCs to reduce the dataset to low dimensionality and retain the original feature information to the maximum extent.

2.2 Degradation state mining based on SOM

Different fault modes have different degradation conditions. Each degradation condition can be subdivided into several degradation states according to the degradation level. Therefore, for degradation state mining, an effective clustering method that cannot only discriminate the feature differences of the degradation state, but also group the degradation states with similar features should be employed to organize and represent degradation clusters. In this section, the SOM was used to cluster the non-fault feature set to mine degradation states with different characteristics.

The SOM is a self-organizing neural network consisting of input and competitive layers. These two neuron layers are fully connected by variable weights. Through an unsupervised learning mechanism, the SOM makes neurons within the competitive layer sensitive to the features of input vectors, demonstrating that neurons act as recognizers of input vectors [31]. After learning, the input data are divided into different regions of the competitive layer. The data features within the same region are similar, while those between separate regions are different; thus, data clustering can be realized. In machine diagnosis studies, SOM is used for anomaly detection [32,33] and clustering analysis [34,35] because of its topology-preserving property.

A low-dimensional feature set $X_{nf} = \{x_i | i = 1, \dots, M\}$ can be obtained by performing feature processing on non-fault samples, where M denotes the number of samples. This low-dimensional feature set X_{nf} can then be further clustered by SOM to mine machine degradation states with different characteristics. The detailed steps of feature set clustering are as follows.

(1) The number of neurons within the input layer is set equal to the number of dimensions of the feature vector, and competitive-layer neurons are arranged in a 2D array with order $(ss \times ss)$. X_{nf} is then provided as input to the first clustering. After clustering, the label number of the competitive-layer neuron corresponding to each sample in X_{nf} is counted to obtain the clustering-category set $clusA = \{a_{1,i} \mid i = 1, \dots, N\}$, where $a_{1,i}$ denotes the label number corresponding to the i th non-fault sample.

(2) Competitive-layer neurons are arranged in a 2D array of order $(ss+1) \times (ss+1)$. X_{nf} is provided as input for the second clustering, after which the clustering category is counted and the set $clusB = \{b_{1,i} \mid i = 1, \dots, N\}$ is obtained.

(3) Competitive-layer neurons are arranged in a 2D array of $(ss+2) \times (ss+2)$, and X_{nf} is provided as input for the third clustering, after which the clustering category is counted and the set $clusC = \{c_{1,i} \mid i = 1, \dots, N\}$ is obtained.

(4) Sets $clusA$, $clusB$, and $clusC$ are integrated to obtain the clustering sequence set $Seq_1 = \{(a_{1,i}, b_{1,i}, c_{1,i}) \mid i = 1, \dots, N\}$, where $(a_{1,i}, b_{1,i}, c_{1,i})$ denotes the sequence corresponding to the i th non-fault sample.

By analyzing the neuron distribution within the competitive layer, samples are selected, merged, and removed to obtain the dataset consisting of degradation state samples only. The steps for this mining strategy can be described as follows.

(1) Count the labels of neurons within the $ss \times ss$ competitive layer, whose clustering sample number is not less than $M/(ss \times ss)$, and create the neuron label set $nNum_1$. In accordance with first-dimensional data concerning Seq_1 , select samples correspond to the label $nNum_1$ in X_{nf} . Next, create the dataset $Sel_1 = \{x'_i \mid i = 1, \dots, P\}$, along with the corresponding clustering sequence set $Seq_2 = \{(a_{2,i}, b_{2,i}, c_{2,i}) \mid i = 1, \dots, P\}$, by using selected samples, where P denotes the number of selected samples.

(2) Follow step 1 for the $(ss+1) \times (ss+1)$ layer and proceed to build the label set $nNum_2$, dataset $Sel_2 = \{x''_i \mid i = 1, \dots, Q\}$, and corresponding clustering sequence set $Seq_3 = \{(a_{3,i}, b_{3,i},$

$c_{3,i} | i = 1, \dots, Q$, where Q denotes the number of selected samples.

(3) Follow step 1 for the $(ss+2) \times (ss+2)$ layer and build the label set $nNum_3$, dataset $Sel_3 = \{x_i''' | i = 1, \dots, R\}$ and corresponding clustering sequence set $Seq_4 = \{(a_{4,i}, b_{4,i}, c_{4,i}) | i = 1, \dots, R\}$, where R denotes the number of selected samples.

(4) Based on the distance distribution of neurons within the $(ss+2) \times (ss+2)$ competitive layer, analyze the neighborhood of each neuron in $nNum_3$. Merge the closest neighboring neurons with labels consistent with the number of neurons with the most clustering samples. By referring to the merged-neuron labels, modify the third-dimensional data of Seq_4 to update the corresponding clustering sequences of Sel_3 samples and obtain $Seq_5 = \{(a_{5,i}, b_{5,i}, c_{5,i}) | i = 1, \dots, R\}$.

(5) Follow step 4 for the $(ss+1) \times (ss+1)$ layer and modify the second-dimensional Seq_5 data to update the corresponding clustering sequence of Sel_3 samples and obtain $Seq_6 = \{(a_{6,i}, b_{6,i}, c_{6,i}) | i = 1, \dots, R\}$.

(6) According to Seq_6 , sort samples within Sel_3 according to their respective clustering states. By analyzing the sample within each state, find and remove normal-state samples. Finally, the degradation state dataset can be obtained. Fig. 3 depicts the overall mining strategy workflow.

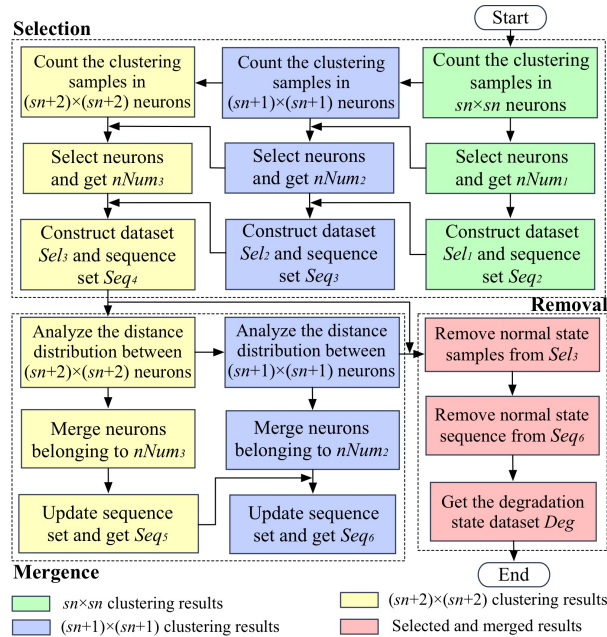


Fig. 3. Flowchart of mining strategy.

2.3 Degradation state identification based on the PSO-SVM

Because of the multiple types of degradation states, the number of samples corresponding to each degradation state is relatively small under the condition that the total number of degradation samples is constant. Therefore, the degradation state identifier should be built using a model with good classification effect for multi-class small samples. In this section, a PSO-SVM model is proposed to identify different degradation states.

The SVM is a margin-based classifier that can perform classification for linear and nonlinear samples [17,36]. SVMs are used as classification tools in machine diagnosis research because of their high classification accuracy and good generalization ability, even with few samples [16,18,37]. The basic principle of SVM involves constructing an optimum hyperplane linearly separating two classes of samples with the maximum margin [38]. Let $S = \{(x_i, y_i) \mid x_i \in R^N, y_i \in \{-1, 1\}, i=1, \dots, n\}$ represent the training set, wherein x_i denotes an input vector and y_i refers to the label of x_i . The optimum hyperplane function can be expressed as

$$f(x) = \text{sgn}(\sum_{i=1}^n \alpha_i y_i K(x, x_i) + b), \quad (6)$$

where α_i represents a Lagrange multiplier and $K(x_i, x_j)$ is the kernel function. In this study, the radial basis function (RBF) was used as the SVM kernel function. The RBF can be defined as follows:

$$K(x, x_i) = \exp(-\|x - x_i\|^2 / 2\sigma^2), \quad i = 1, \dots, n, \quad (7)$$

where σ is a parameter that sets the kernel “spread.”

In order to identify multiple degradation states, a multi-class SVM classifier is

applied. Multi-class SVM is mainly implemented by combining multiple binary SVMs. The “one-against-one” strategy was used in this study to build the SVM classifier [39]. Regarding the sample data of k classes, a set of $k \cdot (k-1)/2$ binary SVMs was built. Each of these SVMs considered only two sample classes, and they classified and voted for the input sample, i.e., the sample belonging to the class with the most votes.

Regarding these SVMs, the parameters C and σ describe impacts on the classification performance, where C is a penalty parameter used to control the degree of penalty associated with misclassified samples. Optimum parameters must be selected to improve the classification accuracy of the SVMs. PSO is a population-based search method that can efficiently find the optimal or near-optimal solutions in search space. In this study, PSO was employed to select the best value for C and σ of SVM, because it has been demonstrated to be a powerful parametric optimization technique [40].

The detailed process of the degradation state identification based on PSO-SVM is depicted in Fig. 4 and consists of the following three steps. (1) Data processing: Degradation state samples are categorized into training and test sets, on which feature processing is performed to obtain corresponding low-dimensional datasets tr and te , respectively. (2) Training: The SVMs are trained using tr , and the PSO algorithm is used to optimize the SVM parameters iteratively until the termination criterion is met. The process of PSO optimizing SVM parameters were described in [41,42]. Based on the “one-against-one” multi-class classification strategy, the classifier model is built using optimum parameter values thus obtained. (3) Testing: The te dataset was used to test the classification accuracy of the proposed model to prove the feasibility of employing this model for degradation state identification.

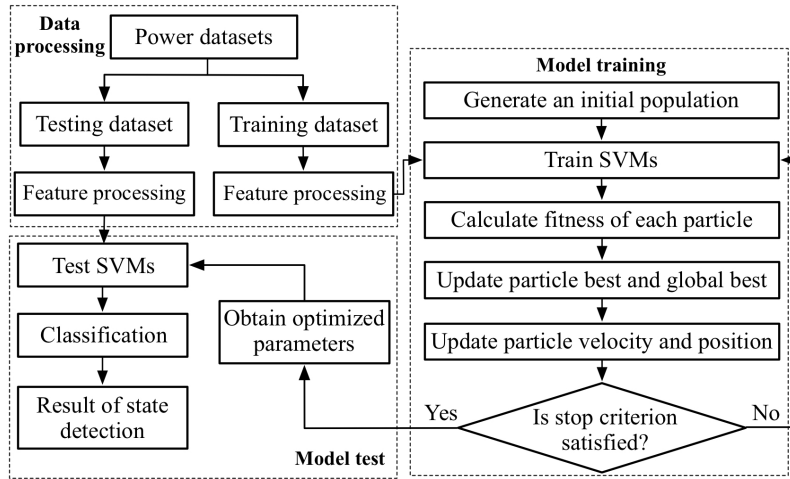


Fig. 4. Identification workflow based on PSO-SVM.

3. Case study

3.1 Point machine monitoring

A centralized signaling monitoring system is an integrated platform for monitoring railway signal equipment. This system can monitor point machine operation according to the turnout relay condition. Fig. 5 depicts the process through which a monitoring system acquires machine power data. The switch-amount acquisition module obtains the machine start and end times by judging the state of the turnout relay 1DQJ. During this period, the Hall current sensor collects current data from the open-phase protector, while the output terminals of the Hall sensor are connected to the power collector. The power collector collects voltage data during each phase from the open-phase protector. In addition, the current and voltage sampling frequency is 25 Hz. After power data are obtained by performing calculations using current and voltage data, the power collector transmits power data to the communication front-end processor via the RS-485 bus and finally sends the data to the monitoring host through the switch.

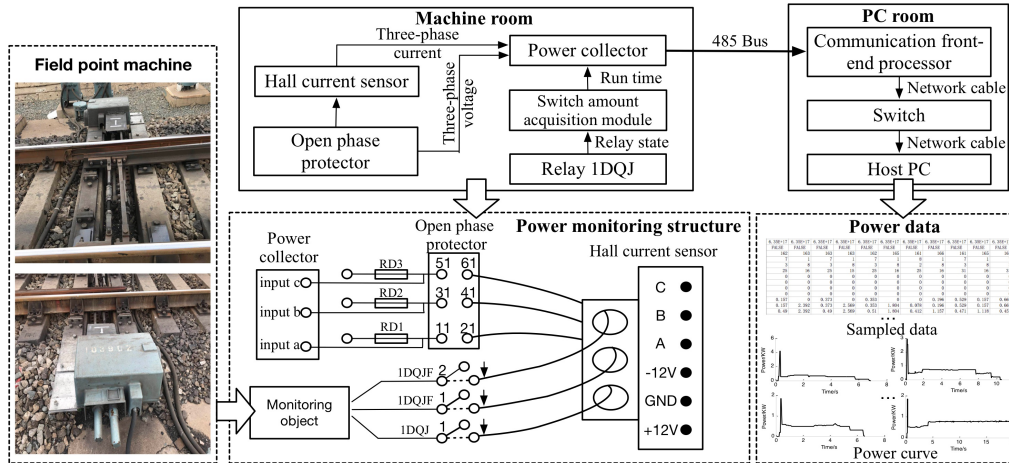


Fig. 5. Flowchart of point machine monitoring.

Fig. 6 presents the power signal acquired by the centralized signaling monitoring system. As the turnout power source, the changes in the point machine signal curve are related to the process of switching and locking the turnout. The power curve is divided into four phases according to the point machine operation characteristics: (1) a start-release phase (0–1 s), (2) a switch phase (1–4 s), (3) a lock phase (4–5 s), and (4) an indication phase (≥ 5 s). Similar partitioning methods were used in [2,20,24].

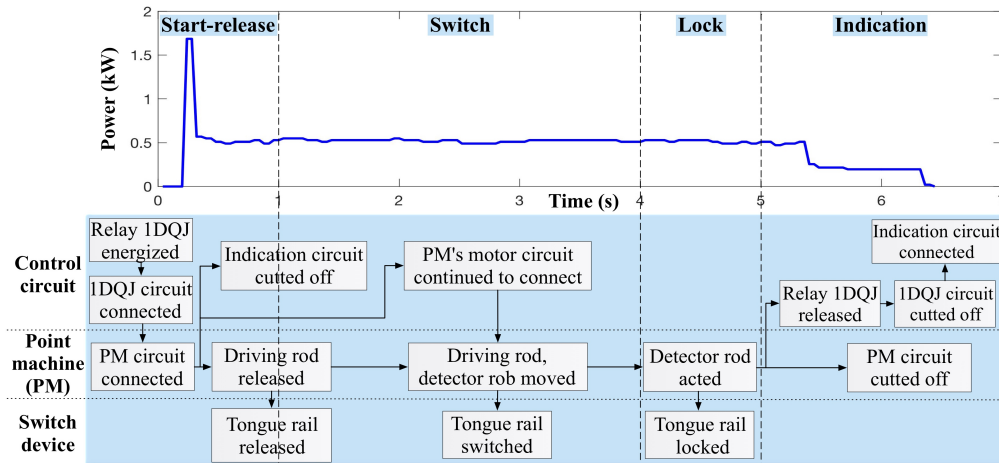


Fig. 6. Normal power curve of an S700K point machine.

3.2 Fault modes of S700K point machine

Six fault modes of the S700K point machine were identified by investigating the actual working conditions of onsite equipment and reviewing maintenance log. Fig. 7

depicts the power curves corresponding to these fault modes. Their corresponding characteristics can be summarized in the time-domain phase (Table 1). The four time-domain phases are represented as phases 1 to 4 for convenience.

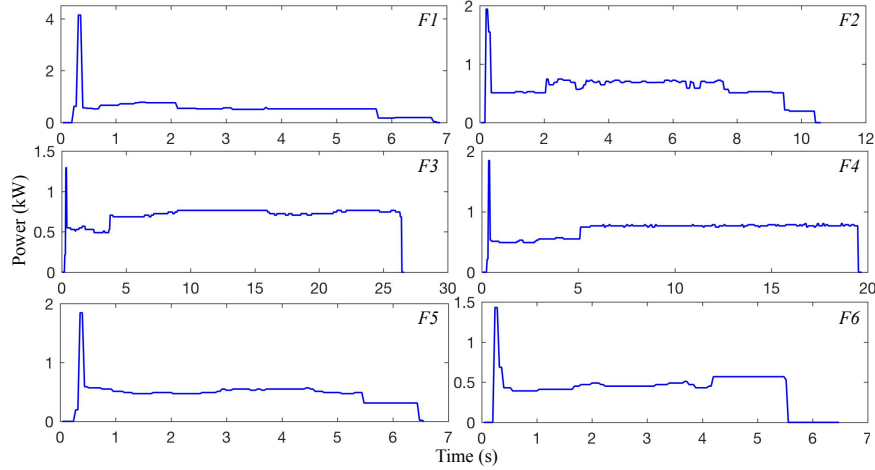


Fig. 7. Power curves corresponding to six fault modes.

Table 1

Descriptions of fault characteristics within time-domain phases

Fault	Characteristics
<i>F1</i>	In phase 1, the observed power peak is rather large, and the overall power is larger than that under normal operating conditions
<i>F2</i>	In phase 2, the degree of power fluctuations is large
<i>F3</i>	In phase 2, the power rises sharply and remains constant until it reaches the time limit; subsequently, it drops to zero
<i>F4</i>	In phase 3, the power rises sharply and fluctuates abnormally until it reaches the time limit; subsequently, it drops directly to zero
<i>F5</i>	In phase 4, the power is approximately twice that under normal conditions
<i>F6</i>	In the later part of phase 4, the power drops and remains at zero

The six power curves were subsequently projected onto the value domain. Fig. 8

demonstrates that significant differences exist between the power distributions corresponding to the above-mentioned fault modes. The power peak of $F1$ is rather large—approximately 4.2 kW—and was caused by the high point machine power required to release the turnout during phase 1. The power distribution of $F2$ is relatively discrete due to large power fluctuations during switching operation of the turnout performed by the point machine during phase 2. $F3$ and $F4$ possess more power points due to the long point machine idling time during phases 2 and 3, respectively. In addition, the power values corresponding to these two faults are mostly between 0.5 kW and 1 kW, indicating that the point machine operating power is rather high in these two states. The power corresponding to $F5$ has a specific distribution between 0 kW and 0.5 kW due to abnormal power changes in the slow-release region during phase 4. $F6$ exhibits a greater “zero value” of power due to all of the power values within the slow-release region being zero during phase 4.

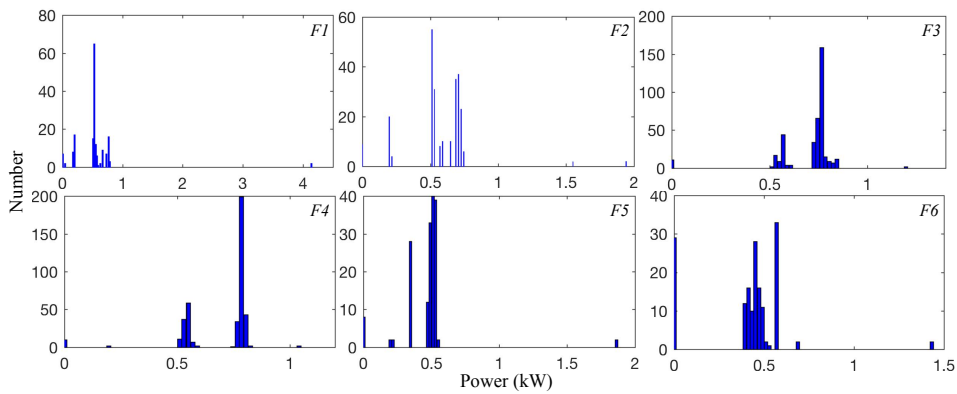


Fig. 8. Value-domain projections of six fault modes.

3.3 Point machine non-fault power data

Fig. 9 depicts the first PC distributions of 600 power samples. As can be seen, the first PC values of the non-fault samples are between 0 and 0.08, while those of the fault

samples are between 0.08 and 0.37. The non-fault samples include both the normal- and latent-state samples. The attribute values of the normal samples are mostly zero, while a few slightly fluctuate around zero. The attribute values of the latent samples are mainly between 0.02 and 0.08. One can observe that the attribute value distributions corresponding to the normal and latent states differ significantly, demonstrating that these state samples can be distinguished and used for degradation state mining.

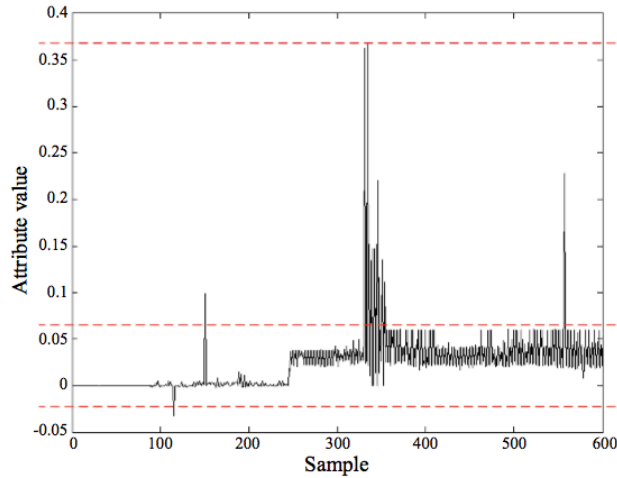


Fig. 9. First PC distributions of fault and non-fault samples.

4. Experimental results

In this study, a 52-day power dataset obtained from the S700K point machines stationed at the turnouts—No. 1 (J1, J2, J3), No. 2 (J1, J2, J3), No. 3 (J1), and No. 4 (J1)—was acquired by centralized signaling monitoring system of the Chenzhou West station. The acquired dataset served as experimental data to verify the proposed method. This dataset contained 20 normal samples, 1,000 non-fault samples, and 120 fault-state samples. MATLAB 2017a was used to program the simulation, which was implemented on a laptop equipped with an Intel Core i7-5600U HQ processor (4 MB cache, up to 3.20 GHz) and 8 GB memory.

4.1 Power data feature processing results

Twenty normal samples (N) and 20 samples corresponding to each fault state were selected to form a standard dataset containing 140 samples. By performing feature processing on this dataset, the effective features and settings of the KPCA parameters were obtained. Further, with reference to the abovementioned process, feature processing of the non-fault dataset was also performed.

By extracting the statistical parameters of the standard dataset from the time and value domains, a 64-dimensional feature set was obtained. Subsequently, the criterion values and correlation coefficients of this feature set were calculated and compared, and effective features— $t_{1,3}$, $t_{2,1}$, $t_{2,2}$, $t_{2,8}$, $t_{4,3}$, $t_{4,5}$, $t_{4,8}$, $t_{4,9}$, $v_{1,2}$, $v_{1,7}$, and $v_{3,3}$ —were selected. These features were used to form a new 11-dimensional feature set, which was normalized and utilized to obtain a distribution of the attribute values of the samples within each feature dimension, as depicted in Fig. 10. These 11 features were effectively used to distinguish between the different sample-data states. Furthermore, KPCA was applied to reduce the dimensionality of the feature set.

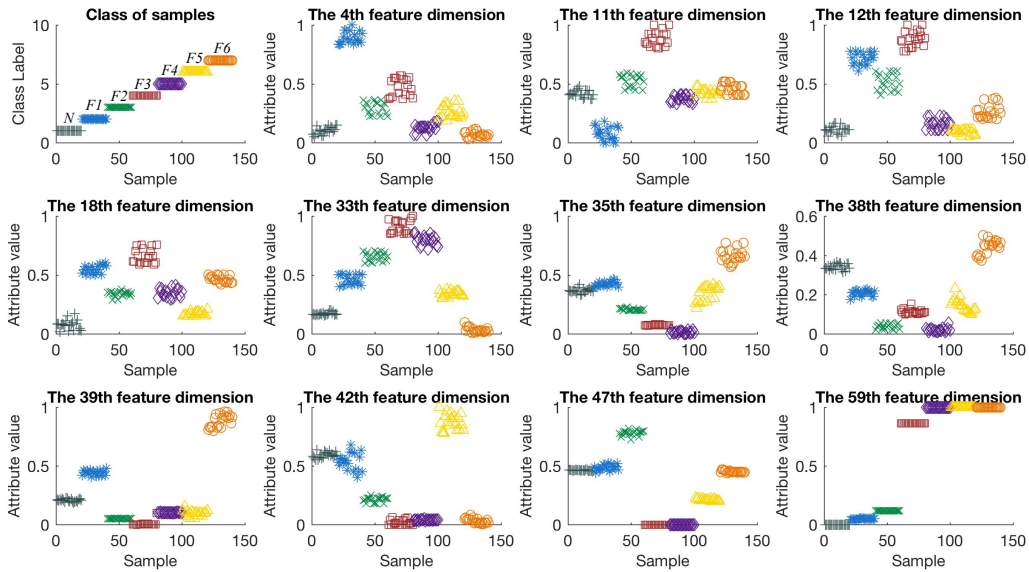


Fig. 10. Distributions of normal and fault samples within the effective feature dimensions.

The RBF was also set as the KPCA kernel function with a radial basis parameter

$\sigma = 2$. The number of PCs varies between 3 and 9. By comparing the seven-dimensional reduction results, it can be seen that when the number of PCs is set to 6 (i.e., the dataset is reduced to six dimensions), the best distinction among different states is achieved and more than 99% of the original information can be retained. The distributions of samples in different PCs are depicted in Fig. 11. Feature processing was performed on the non-fault dataset consisting of 1,000 samples, as per the abovementioned procedure, and a six-dimensional feature set was obtained. Fig. 12 depicts the distributions of non-fault samples within different PCs. As shown, the aforementioned sample distribution differs in all six dimensions.

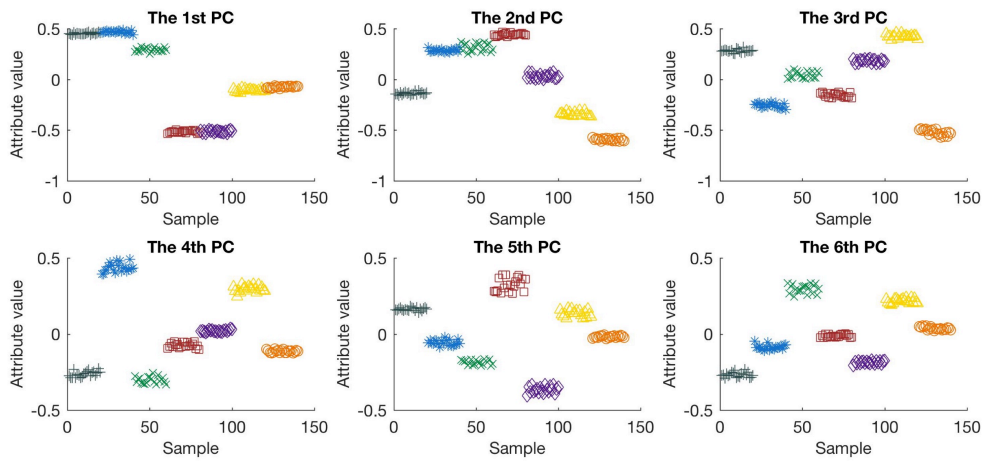


Fig. 11. Distributions of normal and fault samples in six PCs.

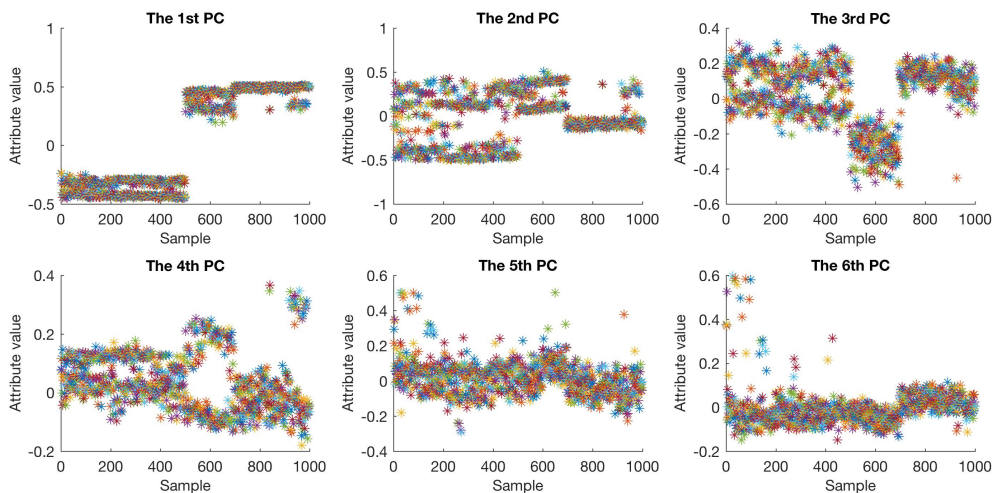


Fig. 12. Distributions of non-fault samples in six PCs.

4.2 Degradation mining results

The competitive-layer neurons of SOM were arranged in arrays with dimensions of 4×4 , 5×5 , and 6×6 . The connection weights were randomly initialized, and the neuron neighborhood was observed to shrink to the shape of a hexagonal grid. The number of iterations was set to 1000. The loss function was represented as a mean square error. After setting the SOM parameters, cluster analysis was performed on the non-fault feature set. The computational times required for SOM with competitive-layer sizes of 4×4 , 5×5 , and 6×6 were 309 s, 312 s, and 317 s, respectively. The sample distributions after clustering are depicted in Fig. 13, wherein the hexagonal lattice represents neurons within the competitive layer, whereas the number within the lattice represents the sample size clustered by neurons. Fig. 14 presents the distance distribution of neurons within the competitive layer after clustering, wherein the connection band between adjacent neurons was used to measure the distances. A lighter connection band color represents a shorter distance, whereas a darker color denotes a longer distance.

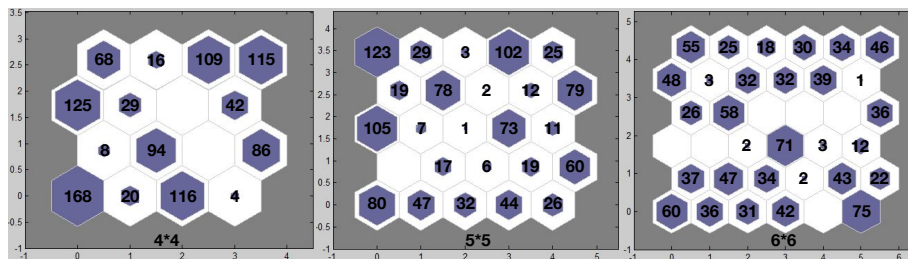


Fig. 13. Distributions of clustering samples within competitive-layer neurons.

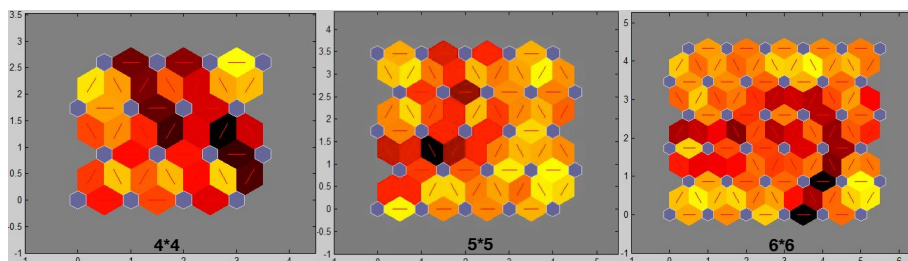


Fig. 14. Distributions of neuron distance within competitive layers.

By selecting, merging, and removing the SOM clustering results, six degradation states— $D1$ – $D6$ —were obtained. Table 2 lists the clustering sequences and sample sizes corresponding to each state. The total number of degradation samples is 386, accounting for 38.6% of the non-fault dataset. Most of the samples within the dataset were not divided into degradation states. These undivided samples include normal- and intermediate-state samples. During transition from the normal to the degradation state, a point machine experiences several intermediate states. The difference between the intermediate and normal states is fuzzy, and compared to a degradation state, an intermediate state does not typically reflect machine degradation. Consequently, these intermediate states are not divided into degradation states.

Table 2

Statistical information concerning six degradation states

Clustering sequence	$D1$	$D2$	$D3$	$D4$	$D5$	$D6$
SOM 4×4	1	6	8	9	3	15
SOM 5×5	21	20	17	10	24	11
SOM 6×6	7	20	16	29	31	1
Sample number	73	58	71	69	55	60

Fig. 15 depicts power curves corresponding to the six degradation states. The power values for $D1$, $D2$, and $D3$ are evidently normally distributed in phases 1, 3, and 4, respectively. However, in phase 2, compared to N , the power values and fluctuation degrees of the three states increase successively. The characteristics of $D1$, $D2$, and $D3$ are similar to those of $F2$, which could be because point machines are subject to abnormal resistances during the tongue rail switching process. The power values of $D4$

and $D5$ are normally distributed in phases 2 and 4, respectively. However, during phase 1, the observed power peaks of the two states are lower than that of N . Moreover, during phase 3, compared to N , the power and fluctuation degree of $D4$ and $D5$ increase successively; these characteristics are consistent with those of $F4$. This behavior may result from the detector rob moving abnormally during the process of locking the point machine tongue rail. The overall power value of the $D6$ state is lower than that of N , and its degradation characteristics are not consistent with the existing six fault modes. Referring to the field maintenance log, this abnormal state is caused by imperfect debugging of the point machine before operation or after maintenance.

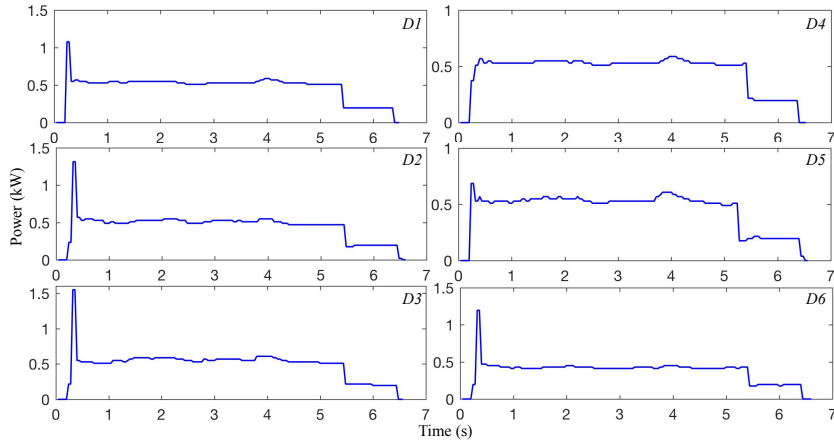


Fig. 15. Power curves corresponding to six degradation states.

4.3 Degradation identification results

With regard to $D1$ – $D6$, a degradation dataset consisting of 330 samples was created, with 55 samples in each state. Each state sample in this dataset was randomly divided in a 4:1 ratio to create tr and te sets. The PSO-SVM parameters were set as follows: the inertia weights w_{\min} and w_{\max} were set to 0.4 and 0.9, respectively; the number of particles was 20; the maximum iteration number was set to 200; and the acceleration constants c_1 and c_2 were set to 1.5 and 1.7, respectively. The search ranges of SVM parameters C and σ were set to $[0, 100]$ and $[0, 1000]$, respectively, and the

two parameters were randomly initialized. The 5-fold cross validation was used to evaluate the performances of the SVM parameters. The computational time required to train and test the PSO-SVM model using tr and te was about 54.5 s.

Five simulation experiments were performed, and the average classification accuracy of the model was considered to be the experimental result. Fig. 16a depicts the fitness curve of PSO searching for the optimum parameter values in one of the experiments, wherein the best fitness is 100% and the average fitness is approximately 96%. The optimum SVM parameter values (C , σ) obtained in this experiment were (4.894, 20.954). Fig. 16b depicts the classification results of the abovementioned experiment. During this experiment, the classifier mistakenly identified two $D6$ samples as $D2$ and $D3$ samples. Because the degradation characteristics of $D6$ are not obvious and its power distributions during phases 1, 2, and 4 are similar to those of $D2$ and $D3$, the model failed to identify $D6$ samples accurately.

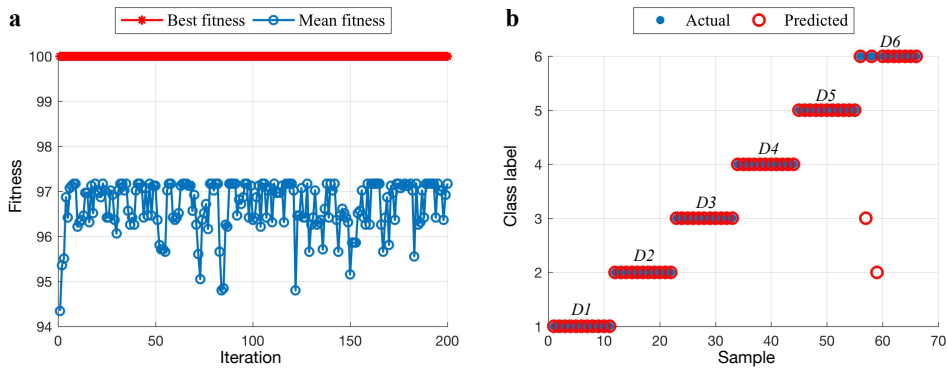


Fig. 16. PSO-SVM training and testing. (a) PSO fitness curve; (b) classification results.

The experimental results obtained using different classification methods, including back propagation (BP) neural network, decision tree, and naïve Bayes methods, are listed in Table 3. The average classification accuracy of the proposed model is evidently superior to those of these methods. The average classification accuracy of the PSO-SVM model without KPCA is low, proving the necessity of further

dimensionality reduction in terms of the selected feature data. Unlike the PCA-based method, the proposed method uses KPCA to extract nonlinear power information from signal data and retains more original data for learning and classification. Although the classification accuracy of genetic algorithm (GA) parameter optimization is lower than that of PSO, it can yield higher classification results than the other methods. Unlike the method without a heuristic algorithm, optimum parameters are used to build the SVM model in the proposed method, thereby improving its classification performance.

Table 3

Experimental results obtained using different methods

State	KPCA+ PSO-SVM	KPCA+ BP network	KPCA+ Decision tree	KPCA+ Naïve Bayes	KPCA+ GA- SVM	PSO- SVM	PCA+ PSO- SVM	KPCA+ SVM
<i>D1</i> (%)	100	90.91	100	81.82	100	90.91	90.91	95.46
<i>D2</i> (%)	100	100	90.91	72.73	95.46	100	100	100
<i>D3</i> (%)	100	90.91	63.67	72.73	90.91	90.91	100	100
<i>D4</i> (%)	100	100	100	86.37	100	81.82	81.82	86.37
<i>D5</i> (%)	100	100	90.91	86.37	100	100	100	100
<i>D6</i> (%)	86.37	72.73	81.82	86.37	81.82	72.73	81.82	72.73
Acc (%)	97.73	92.43	87.89	81.07	94.70	89.40	92.42	92.43

5. Conclusions

This study proposed an early fault diagnosis method for the railway point machine that focuses on the use of non-fault data to mine and identify degradation states under different fault modes. A concise feature set indicating machine degradation was obtained by using the proposed combination strategy to process the non-fault data of the machine. A SOM-based degradation state mining method was proposed for different

degradation conditions. Moreover, a PSO-SVM-based classification model was proposed to accurately identify different degradation states. The feasibility and validity of the proposed method were verified by examining the field point machine power signals. The experimental results showed that this method can be used to mine different degradation level states under multiple fault modes and that the classifier can identify these states with 97.73% accuracy.

In the future, this proposed method will also be used for other machine types to further extend its application domains.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China [grant number 61672080], JWYY [grant number 41402020502], and NASFC [grant number 2016ZD51031].

References

- [1] Tavner PJ. Review of condition monitoring of rotating electrical machines. *IET Electr Power App* 2008;2(4):215-47.
- [2] Márquez FPG, Roberts C, Tobias AM. Railway point mechanisms: Condition monitoring and fault detection. *Proc Inst Mech Eng F J Rail Rapid Transit* 2010;224(1):35-44.
- [3] Gao Z, Cecati C, Ding SX. A survey of fault diagnosis and fault-tolerant techniques—Part I: Fault diagnosis with model-based and signal-based approaches. *IEEE Trans Ind Electron* 2015;62(6):3757-67.
- [4] An D, Kim NH, Choi JH. Practical options for selecting data-driven or physics-based prognostics algorithms with reviews. *Reliab Eng Syst Safe* 2015; 133:223-36.
- [5] Cecati C. A survey of fault diagnosis and fault-tolerant techniques—Part II: Fault diagnosis with

-
- knowledge-based and hybrid/active approaches. *IEEE Trans Ind Electron* 2015.
- [6] Tobon-Mejia DA, Medjaher K, Zerhouni N, Tripot G. A data-driven failure prognostics method based on mixture of Gaussians hidden Markov models. *IEEE Trans Reliab* 2012;61(2):491-503.
- [7] Ali JB, Fnaiech N, Saidi L, Chebel-Morello B, Fnaiech F. Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals. *Appl Acoust* 2015; 89:16-27.
- [8] Jiang H, Chen J, Dong G, Liu T, Chen G. Study on Hankel matrix-based SVD and its application in rolling element bearing fault diagnosis. *Mech Syst Signal Process* 2015; 52:338-59.
- [9] Li Y, Liang X, Yang Y, Xu M, Huang W. Early fault diagnosis of rotating machinery by combining differential rational Spline-based LMD and K–L divergence. *IEEE Trans Instrum Meas* 2017;66(11):3077-90.
- [10] Huang W, Cheng J, Yang Y. Rolling bearing fault diagnosis and performance degradation assessment under variable operation conditions based on nuisance attribute projection. *Mech Syst Signal Process* 2019,114:165-88.
- [11] Lv Y, Yuan R, Wang T, Li H, Song G. Health degradation monitoring and early fault diagnosis of a rolling bearing based on CEEMDAN and improved MMSE. *Materials* 2018;11(6):1009.
- [12] Adachi H, Kikuchi M, Watanabe Y. Electric switch machine failure detection using data-mining technique. *Quarterly Report RTRI* 2006;47(4):182-6.
- [13] Panja SC, Ray PK. Reliability analysis of a ‘point-and-point machine’ of the Indian railway signaling system. *Qual Reliab Eng Int* 2007;23(7):833-48.
- [14] Márquez FPG, Muñoz JMC. A pattern recognition and data analysis method for maintenance management. *Int J Syst Sci* 2012; 43(6):1014-28.

-
- [15] Aguiar EPD, Fernando MDA, Amaral RPF, Fabri DF, Sergio CDA, Ferreira JG, Vellasco MMBR, Tanscheit R, Vellasco PCGDS, Ribeiro MV. EANN 2014: A fuzzy logic system trained by conjugate gradient methods for fault classification in a switch machine. *Neural Comput Appl* 2016;27(5):1175-89.
- [16] Asada T, Roberts C, Koseki T. An algorithm for improved performance of railway condition monitoring equipment: Alternating-current point machine case study. *Transport Res C Emer Technol* 2013;30(5):81-92.
- [17] Atamuradov V, Camci F, Baskan S, Sevkli M. Failure diagnostics for railway point machines using expert systems. In *2009 IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives*, 2009 (pp. 1-5). IEEE.
- [18] Lee J, Choi H, Park D, Chung Y, Kim HY, Yoon S. Fault detection and diagnosis of railway point machines by sound analysis. *Sensors* 2016;16(4):549.
- [19] Vileiniskis M, Remenyte-Prescott R, Rama D. A fault detection method for railway point systems. *Proc Inst Mech Eng F J Rail Rapid Transit* 2016;230(3):18.
- [20] García FP, Pedregal DJ, Roberts C. Time series methods applied to failure prediction and detection. *Reliab Eng Syst Safe* 2010;95(6):698-703.
- [21] Atamuradov V, Medjaher K, Camci F, Dersin P, Zerhouni N. Degradation-level assessment and online prognostics for sliding chair failure on point machines. *IFAC-PapersOnLine*. 2018;51(24):208-13.
- [22] Atamuradov V, Medjaher K, Lamoureux B, Dersin P, Zerhouni N. Fault detection by segment evaluation based on inferential statistics for asset monitoring. In *Annual Conference of the Prognostics and Health Management Society*, 2017 (pp. 58-67).
- [23] Atamuradov V, Medjaher K, Camci F, Dersin P, Zerhouni N. Railway point machine prognostics based on feature fusion and health state assessment. *IEEE Trans Instrum Meas* 2018; 99:1-14.

-
- [24] Ardakani HD, Lucas C, Siegel D, Chang S, Dersin P, Bonnet B, Lee J. PHM for railway system—
A case study on the health assessment of the point machines. In IEEE Conference on Prognostics and
Health Management (PHM), 2012 (pp. 1-5). IEEE.
- [25] Letot C, Dersin P, Pugnaroni M, Dehombreux P, Fleurquin G, Douziech C, La-Cascia P. A data
driven degradation-based model for the maintenance of turnouts: A case study. IFAC-PapersOnLine.
2015;48(21):958-63.
- [26] Lei Y, Li N, Guo L, Li N, Yan T, Lin J. Machinery health prognostics: A systematic review from
data acquisition to RUL prediction. Mech Syst Signal Process 2018 May 1; 104:799-834.
- [27] Li Y, Liang X, Lin J, Chen Y, Liu J. Train axle bearing fault detection using a feature selection
scheme based multi-scale morphological filter. Mech Syst Signal Process 2018; 101:435-48.
- [28] Yin L, Ge Y, Xiao K, Wang X, Quan X. Feature selection for high-dimensional imbalanced data.
Neurocomputing 2013; 105:3-11.
- [29] Kuo FY, Sloan IH. Lifting the curse of dimensionality. Notices of the AMS 2005;52(11):13201328.
- [30] Shao R, Hu W, Wang Y, Qi X. The fault feature extraction and classification of gear using principal
component analysis and kernel principal component analysis based on the wavelet packet transform.
Measurement 2014;54(6):118-32.
- [31] Ritter H, Martinetz T, Schulten K. Neural computation and self-organizing maps: An introduction.
Neurocomputing 1992;5(4-5):243-4.
- [32] Yu J. A hybrid feature selection scheme and self-organizing map model for machine health
assessment. Appl Soft Comput J 2011;11(5):4041-54.
- [33] Germen E, Başaran M, Fidan M. Sound based induction motor fault diagnosis using Kohonen self-
organizing map. Mech Syst Signal Process 2014;46(1):45-58.

-
- [34] Hu J, Zhang L, Liang W. Dynamic degradation observer for bearing fault by MTS–SOM system. *Mech Syst Signal Process* 2013;36(2):385–400.
- [35] Jin W, Shi Z, Siegel D, Dersin P, Douziech C, Pugnaroni M, La Cascia P, Lee J. Development and evaluation of health monitoring techniques for railway point machines. In *IEEE Conference on Prognostics and Health Management (PHM)*, 2015 (pp. 1-11). IEEE.
- [36] Kuang F, Xu W, Zhang S. A novel hybrid KPCA and SVM with GA model for intrusion detection. *Appl Soft Comput J* 2014;18(C):178-84.
- [37] Tran VT, Hong TP, Yang BS, Nguyen TT. Machine performance degradation assessment and remaining useful life prediction using proportional hazard model and support vector machine. *Mech Syst Signal Process* 2012;32(4):320-30.
- [38] Adankon MM, Cheriet M. Support vector machine. *Comput Sci* 2009;1(4):1-28.
- [39] Huang CL, Dun JF. A distributed PSO–SVM hybrid system with feature selection and parameter optimization. *Appl Soft Comput J* 2008;8(4):1381-91.
- [40] Zhao W, Tao T, Zio E. System reliability prediction by support vector regression with analytic selection and genetic algorithm parameters selection. *Appl Soft Comput J* 2015;30(4):792-802.
- [41] Zhang X, Qiu D, Chen F. Support vector machine with parameter optimization by a novel hybrid method and its application to fault diagnosis. *Neurocomputing* 2015 Feb 3; 149:641-51.
- [42] Li Y, Xu M, Wei Y, Huang W. A new rolling bearing fault diagnosis method based on multiscale permutation entropy and improved support vector machine based binary tree. *Measurement* 2016 Jan 1; 77:80-94.