KSME

Springer

Journal of Mechanical Science and Technology 37 (6) 2023

Original Article
DOI 10.1007/s12206-023-0501-y

Keywords:
· Deep reinforcement learning
· Dual-experience pool
· Unbalanced data
· Rolling bearing
· Fault diagnosis

Correspondence to:
Guo Chen
cgnuaacca@163.com

Citation:
Kang, Y., Chen, G., Pan, W., Wei, X., Wang, H., He, Z. (2023). A dual-experience pool deep reinforcement learning method and its application in fault diagnosis of rolling bearing with unbalanced data. Journal of Mechanical Science and Technology 37 (6) (2023) 2715~2726.
http://doi.org/10.1007/s12206-023-0501-y

Received June 11th, 2022
Revised January 15th, 2023
Accepted February 18th, 2023

† Recommended by Editor
No-cheol Park

A dual-experience pool deep reinforcement learning method and its application in fault diagnosis of rolling bearing with unbalanced data

Yuxiang Kang [1], Guo Chen[2], Wenping Pan[1], Xunkai Wei[3], Hao Wang[3] and Zhiyuan He[1]

[1]College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China, [2]College of General Aviation and Flight, Nanjing University of Aeronautics and Astronautics, Liyang 213300, China, [3]Beijing Aeronautical Engineering Technical Research Center, Beijing 100076, China

Abstract A dual-experience pool deep reinforcement learning (DEPDRL) model is proposed for rolling bearing fault diagnosis with unbalanced data. In this method, a dual-experience pool structure is designed to store the sample data of majority and minority classes. A parallel double residual network model is established to extract deep features of the majority and minority input samples, respectively. In the process of training, the proposed balanced cross-sampling technique is used to randomly select samples from dual-experience pool in a certain proportion to realize the training of a double residual network model. We show the effectiveness of our method on three standard data sets, and compared with Resnet18, DCNN, DQN and DQNimb methods, the results show that DEPDRL has the best performance. Finally, with wavelet time-frequency graph as input, DEPDRL is applied to rolling bearing fault diagnosis with unbalanced test data. The results show that on a variety of unbalanced data sets, both the diagnostic accuracy and the G-means value of the DEPDRL are more than 5 % higher than other algorithms, which fully indicates that the DEPDRL has a very high fault diagnosis ability of rolling bearing with unbalanced data.

1. Introduction

The rolling bearing is an important supporting part in rotating machinery that is widely used in various mechanical equipment. Rolling bearing failure caused by overload, wear and other reasons may cause the shutdown of production of the entire mechanical equipment, resulting in economic losses or even casualties. Therefore, it is of great significance to effectively implement rolling bearing fault diagnosis [1].

At present, a deep learning method has been widely used in rolling bearing fault diagnosis because of its natural advantages of feature extraction and the ability to automatically establish nonlinear mapping from feature to type. Lei et al. [2] used deep transfer learning to accurately diagnose the fault state of rolling bearings. Wen et al. [3] proposed an improved deep transfer autoencoder and realized accurate diagnosis of rolling bearing faults. Zhang et al. [4] proposed a convolutional neural networks (CNN) diagnosis method based on casing detection points, which can well identify the fault states of rolling bearings. Guo et al. [5] proposed an improved convolutional neural network based on adaptive learning rate and applied it to rolling bearing fault diagnosis, achieving good results in fault size recognition and fault type diagnosis. Zhang [6] et al. proposed an improved deep belief network (DBN) algorithm and achieved good results in the fault diagnosis test of rolling bearings. Wang et al. [7] proposed a multi-layer supervised autoencoder model, which effectively improved the accuracy of fault diagnosis. Huang et al. [8] proposed a multi-scale cascaded convolutional neural network for bearing fault diagnosis. Hou et al. [9] proposed an improved stack denoising autoencoder for fault classification of rolling bearings, which has the characteristics of high detection accuracy and fast convergence.

The above rolling bearing fault diagnosis methods based on deep learning are all based on the premise of balanced data, that is, the amount of sample data of all classes participating in model training is basically the same. However, in practical work, the running time of rolling bearings in normal state is much longer than that in fault state, and normal samples are often easier to obtain, which leads to serious imbalance between normal and fault sample data in model training. In the diagnosis of unbalanced data samples, it is sensitive to most of the normal state data, but difficult to identify a few of the fault state data, resulting in a large recognition error. Therefore, studying the fault diagnosis of rolling bearings under unbalanced data has become a key technology that needs to be solved urgently [10].

At present, the classification methods for unbalanced data are mainly as follows: 1) Data level methods to achieve inter-class data balance by changing data distribution; 2) Algorithm-level methods that focus on a few types of data samples [11]. In the method research at the data level, Zhou et al. [12] proposed an improved generative adversarial network (GAN) model, which extracts the features of the fault data through the generator and then inputs the features into the discriminator. The global optimization mechanism was adopted to realize the update of GAN. The fault diagnosis ability under unbalanced data is improved. Han et al. [13] put forward the algorithm of SMOTE (synthetic minority oversampling technique), which is a oversampling method that generates new samples by linear interpolation between a few kinds of samples. This method can solve the problem of low accuracy of rolling bearing fault diagnosis caused by unbalanced data [14]. Wang et al. [15] adopted down-sampling technology to reduce the sample size of normal data and achieve accurate diagnosis of a few types of fault samples by balancing the amount of data between classes. However, blind oversampling technique can lead to sample aliasing, which leads to oversampling problem. Similarly, the downsampling method will cause data loss of most classes [16].

Algorithms mainly include integrated learning, cost sensitive learning and deep learning [10, 16]. Wu et al. [17] proposed an LSTM model with undersampling strategy and weighted cost-sensitive loss function, and verified the effectiveness of the proposed model on unbalanced data sets. Qian et al. [18] proposed a class imbalance robust network for bearing fault diagnosis, which can effectively solve the class imbalance problem in feature extraction and classification. Lin et al. [19] proposed a bearing fault diagnosis method based on the concept of class 1 classification and random forest, and applied this method to bearing fault diagnosis with unbalanced data, achieving good results. It is difficult to design an appropriate cost-sensitive function and deal with large data samples because of the above method of processing unbalanced data. Therefore, Lin et al. [20] regarded the classification problem as a continuous decision problem of deep reinforcement learning and designed a return function for unbalanced data, which was fully verified on MNIST, CIFAR-10 and other data sets. Inspired by Lin et al., Kang et al. [21] proposed the reward function of deep reinforcement learning based on K-means algorithm under unbalanced data, and applied the proposed deep reinforcement learning to rolling bearing fault diagnosis with unbalanced data. This method is based on the idea that minority data have large reward and punishment values in cost-sensitive learning. Based on the sample imbalance ratio and manual experience, the model can be trained by designing a specific reward function. However, the classification method at the algorithm level is difficult to determine the imbalance ratio, and the cost sensitive value needs to be obtained through manual experience [22].

In view of the shortcomings of classification methods in processing unbalanced data at the present stage, a dual-experience pool structure is proposed [23] (Experience pool structure is proposed by Google DeepMind team, which mainly stores the state, action, reward and other result data generated during the interaction between an agent and the environment). These data are used for the classification of unbalanced data using the dual-experience pool deep reinforcement learning (DEPDRL) model. The model is designed to solve the problem of unbalanced data classification from algorithm level and data level simultaneously. At the data level, the dual-experience pool structure and balanced cross-sampling method are used to classify the majority and minority samples. At the algorithm level, the parallel dual-residual network model is established to extract the deep features of majority and minority samples. Three sets of standard data sets of the DEPDRL model and two sets of rolling bearing data sets are verified, which shows the correctness and effectiveness of the proposed method.

2. Deep reinforcement learning

2.1 Deep reinforcement learning method based on classified Markov decision process

Combining the advantages of reinforcement learning (RL) and deep learning (DL), It has been favored by many scholars since it was proposed [23]. The purpose of DRL is to maximize the benefit of an agent's trial and error with the environment. The unbalanced data classification task can be regarded as the sequential decision problem of an agent, which is actually the classification Markov process decision (CMDP) [24]. The process usually consists of state set S, action set A, decision reward R, state transition matrix P and discount factor γ. First, we assume that the unbalanced training dataset is $D = \{(x_1, l_1), (x_2, l_2), \ldots, (x_N, l_N)\}$, where (x_i,l_i) represents the i-th sample x_i and the corresponding sample label l_i, and N represents the total number of samples. If an agent correctly judges the category (label l_i) of the current state (sample x_i) during its interaction with the environment, the environment will give a positive reward; otherwise, the environment will punish it negatively. Fig. 1 shows the specific flow of CMDP.

The specific definition of CMDP and related variables in Fig. 1 is as follows:

State set *S*: The state of the environment is determined by the training sample. State s_t at time t corresponds to the t sample x_t in the training set. When the T-step training was completed, the sample set D was randomly shuffled to start the new training.

Action set *A*: The action of agent is determined by the label of the sample, A = {0, 1, 2, …, K-1}, where K is the number of class. The action is the class label of the output judged by the agent after receiving the current state s_t.

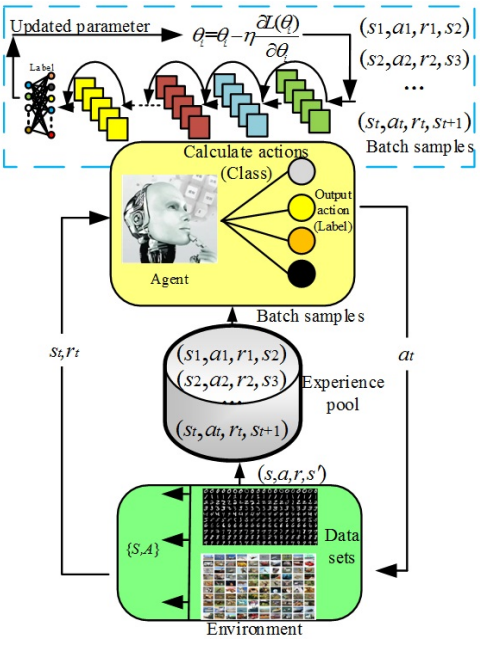

Fig. 1. Classification Markov process decision.

Decision reward *R*: Reward r_t represents the feedback given by the environment after the agent performs the action at in state s_t, that is, $S \times A \rightarrow R$, it is used to reflect whether the action performed by the agent is correct.

State transition matrix *P*: State transition probability p represents the probability of state s_t transition to the next state s_{t+1} according to the existing training sample D after the agent performs the action a_t under state s_t, and is expressed as $p(s_{t+1}|s_t, a_t)$. In CMDP, as the training sample set D is determined, the state transition matrix p is determined.

Discount factor γ: $\gamma \in [0, 1]$ is used to balance future and current rewards. The larger the γ, the more the agent pays attention to long-term return; The smaller the γ, the more the agent cares about immediate interests.

Episode: Indicates the transition trajectory from the initial state s_1 to the termination state s_T. Episode = {s_1, a_1, r_1, s_2, a_2, r_2, …, s_T, a_T, r_T}.

Policy $\pi(\theta)$: Indicates the mapping between the state s_t and the corresponding action a_t, $S \rightarrow A$. In this article, agents with parameters θ are represented by $\pi(\theta)$.

To achieve more accurate fault diagnosis of unbalanced data, it is necessary to find the optimal mapping strategy $\pi(\theta)$ from state to action, and the process of exploring the optimal strategy $\pi(\theta)$ can be realized through deep reinforcement learning.

2.2 Deep reinforcement learning method based on classified Markov decision process

The specific flow of DEPDRL for the classification and diagnosis with unbalanced data is shown in Fig. 2. DEPDRL consists of the following parts:

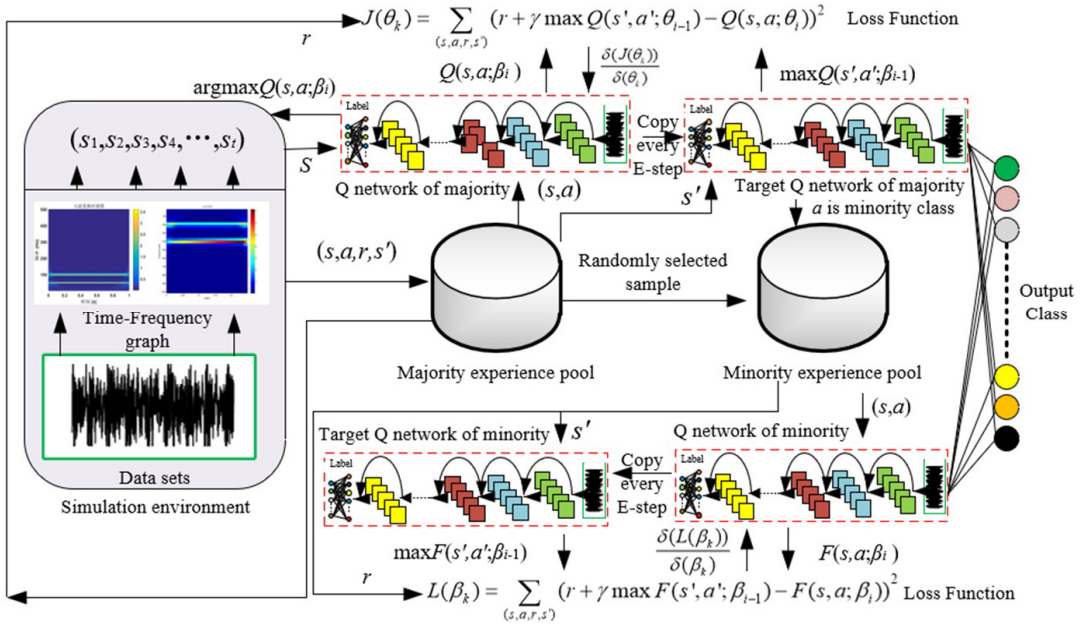

Fig. 2. DEPDRL.

1) Dual-experience pool structure (DEPS). The DEPS is mainly used to store <*s, a, r, s'*> samples generated by the environment. One of the experience pool is called majority experience pool (Maj-EP) ,and it is used to store all the samples from the simulated environment. The samples diagnosed as minority classes and the samples incorrectly diagnosed are stored in minority experience pool (Min-EP), Meanwhile, the Min-EP stores part of the majority class samples according to a certain probability. The DEPS and the balanced cross-sampling in the training process can decompose the classification of unbalanced data into the recognition process of balanced data, thus improving the diagnostic recognition ability of the model.

2) Double parallel residual network. The dual parallel residual network mainly serves DEPS, and trains the double parallel residual network, respectively, by sampling in the DEPS. The majority class network is mainly used for the recognition of the majority class, and the minority class network is mainly used for the classification of the minority class and the diagnosis of the error recognition samples of the majority class network. The final classification result is the sum of double parallel residual networks.

3) Simulation environment. In the simulation environment, the unbalanced sample data set is converted into a state value s_t through sampling, and the reward and punishment function is set in the simulation environment to realize the reward and punishment of the classification results of agents (double parallel residual network).

In DEPDRL, the agent obtains the optimal strategy $\pi(\theta)$ during the interaction with the environment. In this process, the strategy π learned by the agent can be expressed as the probability of executing the action at in the current state s_t:

$$\pi(a \mid s) = P(a = a_t, s = s_t) . \tag{1}$$

The purpose of an agent is to achieve target classification as accurately as possible under the guidance of the strategy π. The more accurate the classification is, the more rewards the agent will get, namely, the greater the cumulative reward G_t. Therefore, the merits and demerits of the strategy π can be measured by the cumulative reward G_t.

$$G_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k} . \tag{2}$$

In the process of judging the value of G_t, the expectation of G_t can reflect how well the agent performs the action a_t in state s_t, which is called the state action value function $Q(s, a)$.

$$Q^{\pi}(s,a) = \mathrm{E}[\mathrm{G}_t | s_t = s, a_t = a; \pi] . \tag{3}$$

According to Bellman equation [25], Eq. (3) can be simplified as:

$$Q^{\pi}(s,a) = \mathrm{E}_{\pi}[r_t + \gamma Q^{\pi}(s_{t+1}, a_{t+1}) | s_t = s, a_t = a] . \tag{4}$$

To obtain the optimal strategy π^*, it is necessary to calculate the optimal action value function $Q^*(s,a)$:

$$\pi^*(s,a) \to Q^*(s,a) = \max_{\pi} Q^*(s,a) . \tag{5}$$

The optimal action value function can be determined by Eq. (6):

$$Q^*(s,a) = \mathrm{E}_{\pi}[r_t + \gamma \max_{\pi} Q^*(s_{t+1}, a_{t+1}) | s_t = s, a_t = a] . \tag{6}$$

The deep residual network can be used to calculate the optimal action value function shown in Eq. (6). The experience replay technology is mainly used to sample data from the experience pool to realize network training and parameter updating. In this process, loss function as shown in Eq. (7) is used:

$$L(\theta_i) = \mathrm{E}_{s,a,r,s'}[(y_i - Q(s,a;\theta_i))^2] \tag{7}$$

where, y_i is the estimated value of the target, as shown in Eq. (8); θ_i is network parameters.

$$y_i = r + \gamma \max_{a'} Q(s', a'; \theta_i') \tag{8}$$

where, s', a' are the state and action at the next moment. Since the network adopts the asynchronous update strategy in the training process (the parameters of the Q network are copied to the target Q network in step E of the Q network training), the network parameter values of the target Q network and the Q network are different. Therefore, the target Q network parameter in Eq. (8) is expressed as θ_i'.

According to the loss function, the gradient of network parameters is calculated to complete the update of network parameters θ_i, as shown in Eq. (9):

$$\frac{\partial L(\theta_i)}{\partial \theta_i} = -2 \mathrm{E}_{s,a,r,s'}[(y_i - Q(s,a;\theta_i)) \frac{\partial Q(s,a;\theta_i)}{\partial \theta_i}] . \tag{9}$$

2.3 Reward function

At present, the reward function for deep reinforcement learning is mainly sparse reward, as shown in Eq. (10).

$$r = \begin{cases} 1 & a_t = l_t \\ -1 & a_t \neq l_t \end{cases} \tag{10}$$

where, at is the predictive output of the agent; it is the class label.

By using the reward function shown in Eq. (10), the environment gives the same reward to all the correctly classified samples of the agent, and the same punishment to the incorrectly classified samples. As a result, the same punishment reward value is always used in the training process, which slows the

convergence speed of the algorithm and even makes it difficult to converge. In view of this, a new reward and punishment function with disturbance factor is designed on the basis of Eq. (10), as shown in Eq. (11):

$$r=\begin{cases}1-\eta_1 & a_t=l_t\\ -1-\dfrac{|a_t-l_t|}{\max(a_t,l_t)}-\eta_2 & a_t\neq l_t\end{cases} \tag{11}$$

where, η_1 and η_2 are random numbers between [0, 0.1].

2.4 Balanced cross sampling

In the course of training, Maj-EP stores all samples generated by the environment, while the Min-EP stores the minority samples. If the sample in Min-EP is directly used for the training of minority class network, this will undoubtedly lead to network model over-fitting, so as to make the network generalization ability insufficient. We propose a method called balanced cross-sampling to solve this problem. The main steps of this method are as follows:

1) Set the batch sample size to *BatchSize*.

2) Calculate the ratio of minority class to the total number of class:

$$w=\frac{U}{J} \qquad U<J$$

where, J and U are the total number of class and the number of minority class in this data set.

3) Randomly select N_1 = f (*BatchSize*w*) from a Min-EP. Where *f*(*) is the rounded function.

4) The number of randomly selected samples from the Maj-EP is $N_2 = BatchSize - N_1$.

5) Recombine samples N_1 and N_2 into batch samples with a new *BatchSize*.

Through the above process, the batch samples of minority class network are sampled. The collected batch samples balance minority class and majority class, so that the network can be trained according to the normal algorithm process.

2.5 Deep residual network

Residual structure can effectively solve the problem of over-fitting model with the increase of network depth [26, 27]. Therefore, deep residual network is chosen as the recognition network of agent. Meanwhile, to meet the design requirements of $Q(s, a)$, the output part of deep residual network is redesigned. The improved network structure is shown in Fig. 3.

In Fig. 3, the action value at is added to the full-connection layer, and then the squash compression function is added to the output layer to prevent gradient mutation caused by excessive output value, as shown in Eq. (12). Then the output of $Q(s,a)$ value is carried out.

$$\begin{aligned} \mathrm{v}_j &= \frac{\|\mathrm{s}\|^2}{1+\|\mathrm{s}\|^2}\frac{\mathrm{s}_j}{\|\mathrm{s}\|}\\ \mathrm{v}_j &\approx \|\mathrm{s}\|\mathrm{s}_j \quad \|\mathrm{s}\|\to 0\\ \mathrm{v}_j &\approx \frac{\mathrm{s}_j}{\|\mathrm{s}\|} \quad \|\mathrm{s}\|\to +\infty \end{aligned} \tag{12}$$

where, $\boldsymbol{s}$ is the one-dimensional vector transformed by the convolution result; $\boldsymbol{s}_j$ is the *j-th* element in the one-dimensional vector transformed by the convolution result; $\boldsymbol{v}_j$ is the *j-th* element after compression.

To improve the training efficiency and prevent the problem of low training accuracy caused by the network falling into the local optimal solution in the training process, the Adam optimization algorithm is adopted, and the learning rate shown in Eq. (13) is adopted in the training process.

$$\alpha=\begin{cases}0.01 & \text{epoch}\le 20\\ 0.001 & 20<\text{epoch}\le 50\\ 0.0001 & 50<\text{epoch}\end{cases} \tag{13}$$

Table 1. Model parameter values.

The parameter name	Value
Batch size	64
Epoch	500
The size of the Maj-EP	10000
The size of the Min-EP	1000
γ for majority networks	0.98
γ for minority class networks	0.90
ε -greedy	0.1
E	100

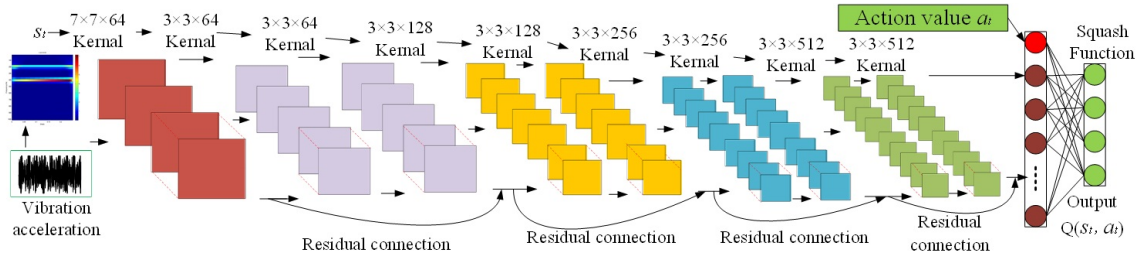

Fig. 3. Deep residual network.

where α is the learning rate, and epoch is the number of iterations.

In this paper, the GPU is NVIDIA GTX1660 6G; I5-9600k processor; 8G of memory; The operating system is Win10. The programming language is Python3.7. The deep learning framework is Pytorch1.18.0. The specific parameter values of DERDRL are shown in Table 1, where E is to transfer the parameters of the current value network to the target value network after the current value network has been trained for a certain number of times. The parameter settings of double parallel residual network are similar to Resnet18. See Fig. 3 for specific values.

3. Unbalanced data verification

3.1 Experimental data set

To verify the classification effectiveness of the proposed DERDRL model on unbalanced data sets, the validation is carried out on MNIST, CIFAR-10 and fashion-MNIST data sets [28], where, both MNIST and fashion-MNIST data sets (fashion) contain 60000 grayscale images with 10 class of size of 28×28. The CIFAR-10 dataset contains 60000 RGB images of 10 class of size of 32×32×3. The class labels of all three datasets are integers from 0 to 9. Results of DERDRL, Resnet18, DQNimb [20], DCNN [22] and DQN models in this paper are compared under the same experimental conditions to verify the effectiveness of the algorithm. As shown in Table 2 is on the basis of the existing data, in order to facilitate comparison, the method of unbalanced data structure [20] is used (in accordance with the ratio imbalance ρ implementation to extract data from the original data set as a minority class label, ρ is the goal of a single minority class sample size ratio of the number of samples and the original data set), to generate a variety of unbalanced data sets of custom implementations. We sampled a few samples of classes with unbalanced ratio ρ from the samples of the class labeled 0 and 1, all the class with odd labels, and all the class with even labels, respectively.

Table 2. Custom data sets.

Data set	ρ (%)	Training set		Test set	
		Negative	Positive	Negative	Positive
MNIST	100	60000		10000	
Cifar-10	100	50000		10000	
Fashion	100	50000		10000	
MNIST (0 and1)	10	1200	49200	2000	8000
	25	3000	50500	2000	8000
Cifar-10 (0 and1)	10	1000	41000	2000	8000
	25	2500	42500	2000	8000
Fashion (0 and1)	10	1200	48000	2000	8000
	25	3000	48000	2000	8000
MNIST (odd)	10	2543	30000	5000	5000
	25	6358	30000	5000	5000
Cifar-10 (odd)	10	2500	25000	5000	5000
	25	6250	25000	5000	5000
Fashion (odd)	10	3000	30000	5000	5000
	25	7500	30000	5000	5000
MNIST (even)	10	2543	30000	5000	5000
	25	6358	30000	5000	5000
Cifar-10 (even)	10	2500	25000	5000	5000
	25	6250	25000	5000	5000
Fashion (even)	10	3000	30000	5000	5000
	25	7500	30000	5000	5000

In the verification process, the output of the model is all 10 class, that is, the unbalanced multi-classification problem, and the test accuracy on the test set is used to judge the classification performance of the model. At the same time, the unbalanced multi-classification index G-means [29] is used to evaluate the classification performance of the model. G-means takes any two categories c_i and c_j from all class, calculates the G-mean(c_i, c_j) values of the two class, and then weights and sums all G-mean(c_i, c_j) values to get the G-means, as shown in Eq. (14).

$$\begin{aligned} G\text{-mean} &= \sqrt{Recall * Specificity} \\ Recall &= \frac{TP}{TP+FN} \\ Specificity &= \frac{TN}{TN+FP} \\ G\text{-means} &= \frac{2}{k(k-1)} \sum_{1\le i\le j} G\text{-mean}(c_i, c_j) \end{aligned} \tag{14}$$

where, TP is the number of true positive samples, TN is the number of true negative samples, FP is the number of false positive samples, FN is the number of false negative samples; k is the number of class.

3.2 Verification results

Table 3 shows the test results of DEPDRL algorithm and other algorithms. The results in Table 3 show that DEPDRL performs well in all data sets and is superior to other methods in classification accuracy. The second method is DQNimb, which improves the reward function according to the imbalance ratio and enhances the reward intensity of the algorithm for a few classes. However, compared with DEPDRL, the classification accuracy of DQNimb is lower. The classification accuracy of Resnet18 and DQN is almost the same, because Resnst18 algorithm model is also used in DQN. Compared with other algorithms, the accuracy of DCNN model is slightly lower, indicating that convolutional network is not suitable for unbalanced data classification directly.

By comparing the experimental results on multiple unbalanced data sets, it is easy to see that DEPDRL has a high classification accuracy on multiple unbalanced data sets. In

Table 3. Unbalanced data test results.

Data set	ρ (%)	Test accuracy (%)				
		Resnet18	DQNimb	DCNN	DQN	DEPDRL
MNIST	100	99.7	99.6	99.3	99.2	99.7
Cifar-10	100	91.6	88.7	83.5	85.4	94.7
Fashiont	100	95.3	92.1	92.3	93.7	96.4
MNIST (0/1)	10	95.3	96.6	94.3	95.1	98.1
	25	98.2	99.2	97.1	97.1	99.6
Cifar-10 (0/1)	10	53.9	58.17	50.4	53.62	66.9
	25	59.6	62.4	52.6	58.2	69.6
Fashion (0/1)	10	81.4	87.65	78.6	80.27	91.1
	25	87.3	91.5	86.7	88.6	93.6
MNIST (odd)	10	97.3	97.9	96.8	97.4	98.7
	25	98.1	98.9	97.9	98.6	99.3
Cifar-10 (odd)	10	45.7	46.6	41.3	43.8	51.7
	25	55.9	57.8	50.2	52.9	63.8
Fashion (odd)	10	85.6	86.9	82.2	84.4	91.8
	25	90.3	91.6	87.1	88.7	93.2
MNIST (even)	10	97.2	98.0	97.6	97.2	98.6
	25	97.7	98.9	97.5	97.9	99.2
Cifar-10 (even)	10	46.9	47.0	44.1	45.3	52.4
	25	57.1	58.1	53.2	54.4	65.3
Fashion (even)	10	85.3	86.2	83.6	84.3	90.6
	25	89.7	90.5	85.8	88.4	92.5

Cifar-10(0 and 1) data set when ρ = 25 %, the accuracy of DEPDRL algorithm is more than 6 % higher than that of DQNimb, reaching 69.93 %. When ρ = 10 %, the accuracy of the algorithm is improved by more than 8.7 % and reaches 66.9 %. When ρ = 25 % in the fashion-MNIST (0 and 1) data set, the accuracy of DQNimb algorithm with better classification accuracy is improved by 2 %, reaching 93.62 %. When ρ = 10 %, it increases by 3.45 %. At this time, the classification accuracy of MNIST data set is similar to that of other algorithms, but it also keeps ahead.

In Cifar-10 (odd) data set, when ρ = 25 %, DEPDRL improves the accuracy by more than 6 %, reaching 63.8 %. When ρ = 10 %, it increases by more than 5 % and reaches 51.7 %. Similarly, when ρ = 25 % in the fashion-MNIST (odd) data set, the accuracy of DEPDRL algorithm is improved by 1.6 % compared with that of DQNimb algorithm with better classification accuracy, reaching 93.2 %. When ρ = 10 %, it increases by 4.9 %. At this time, the classification accuracy of MNIST data set is similar to that of other algorithms, but it also has the highest classification accuracy.

In Cifar-10(even) data set, when ρ = 25 %, the accuracy of DEPDRL algorithm is 64.5 % more than 6 % higher than that of DQNimb algorithm. When ρ = 10 %, the accuracy is improved by 5.4 % to 52.7 %. In the same fashion-MNIST (even) data set, when ρ = 25 %, the accuracy is improved by 2 %, reaching 92.5 %. When ρ = 10 %, the increase is 4.4 %. At this time, the classification accuracy of MNIST data set is similar to that of other algorithms, but it also has the highest classification accuracy.

We report the G-means scores of different algorithms in Table 4. The results in the table show that the G-means evaluation index of DEPDRL is significantly higher than the results of other methods, which further indicates that DEPDRL can effectively extract the features of a few types of samples in the process of unbalanced data classification and can accomplish the classification task well.

4. Rolling bearing fault diagnosis example

To verify the effectiveness of DEPDRL in rolling bearing fault diagnosis, verification is carried out on the rolling bearing fault diagnosis data set of Case Western Reserve University [29] and the rolling bearing fault test data set of aeroengine rotor tester with casing [29] of Nanjing University of Aeronautics and Astronautics, respectively, and the results were compared with those of the above methods.

4.1 Rolling bearing test data diagnosis of case western reserve university

The driving end data set of rolling bearings from Case Western Reserve University was selected. The corresponding bearing model was SKF6205, and the data sampling frequency was 12 kHz. There are four bearing states in total, including three

Table 4. Unbalanced datasets results of G-means.

Data set	ρ (%)	The results of the G-means				
		Resnet18	DQNimb	DCNN	DQN	DEPDRL
MNIST	100	0.996	0.992	0.993	0.992	0.995
Cifar-10	100	0.898	0.883	0.859	0.853	0.951
Fashion	100	0.947	0.924	0.926	0.938	0.966
MNIST (0/1)	10	0.943	0.964	0.941	0.949	0.981
	25	0.981	0.992	0.969	0.965	0.997
Cifar-10 (0/1)	10	0.523	0.572	0.495	0.513	0.687
	25	0.582	0.633	0.513	0.582	0.694
Fashion (0/1)	10	0.812	0.865	0.788	0.798	0.905
	25	0.861	0.918	0.865	0.846	0.929
MNIST (odd)	10	0.973	0.983	0.954	0.971	0.984
	25	0.981	0.981	0.972	0.982	0.987
Cifar-10 (odd)	10	0.449	0.453	0.441	0.413	0.498
	25	0.554	0.453	0.516	0.519	0.625
Fashion (odd)	10	0.837	0.864	0.837	0.816	0.901
	25	0.906	0.905	0.891	0.862	0.921
MNIST (even)	10	0.971	0.977	0.962	0.972	0.981
	25	0.978	0.981	0.973	0.980	0.988
Cifar-10 (even)	10	0.458	0.465	0.412	0.441	0.512
	25	0.576	0.578	0.523	0.543	0.646
Fashion (even)	10	0.842	0.853	0.826	0.831	0.901
	25	0.889	0.894	0.839	0.878	0.916

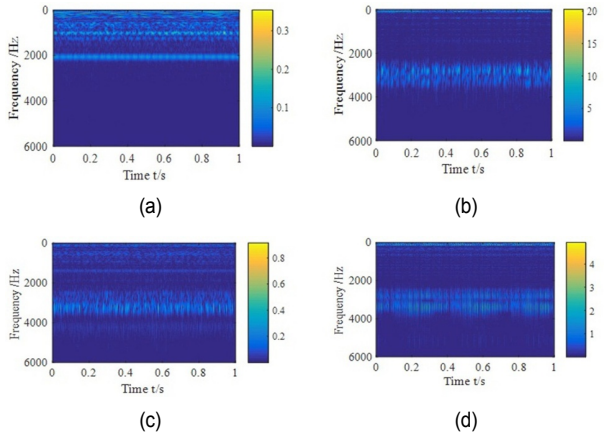

Fig. 4. Wavelet time-frequency diagram: (a) normal sample; (b) inner failure; (c) rolling failure; (d) outer failure.

processing defects fault and normal state of inner ring, outer ring and rolling body. In the process of data processing, the loading and non-loading conditions were considered together, and the training and test sets were divided in a 4:1 ratio. Then, the wavelet time-frequency diagram (using "Morse" wavelet) of the original vibration acceleration signal with sampling time of 0.04 s (longer than the time per revolution) was directly kept as a PNG graphic file with size of 128×128×3, which serves as the input of each classification model. Fig. 4 shows the wavelet time-frequency diagrams of different fault categories. To simulate the problem of data imbalance caused by too much normal sample data and insufficient fault sample data in real vibration, the unbalanced data set was constructed by using the unbalanced ratio ρ. Table 5 shows the number of unbalanced data samples constructed according to the imbalance ratio ρ and the sample size under normal conditions (ρ = 100).

Table 5. Unbalanced test datasets.

ρ	Train set				Test set	
	Inner ring	Outer ring	Rolling fault	Normal	Fault	Normal
100	3234	2431	3231	2828	2225	707
25	809	608	808	2828	2225	707
20	647	486	646	2828	2225	707
15	485	365	485	2828	2225	707
10	323	243	323	2828	2225	707
5	162	122	162	2828	2225	707
1	32	24	32	2828	2225	707
Lable	0	1	2	3	Test	

We show the classification and diagnosis results of unbalanced fault data and normal conditions by different methods in Table 6. The results show that DEPDRL has the highest classification diagnosis accuracy relative to other models in the customized unbalanced data sets. Taking ρ = 10 as an example, the classification accuracy of DEPDRL reaches 94.72 %, which is 5 % higher than that of DQNimb. In addition, compared with

Table 6. Diagnosis result.

ρ	Test result (%)			
	Resnet18	DQNimb	DQN	DEPDRL
100	99.62	98.92	98.65	99.68
25	89.44	92.37	87.34	96.78
20	87.65	91.34	84.86	95.84
15	87.36	89.94	80.97	95.71
10	86.57	89.34	80.16	94.72
5	82.37	85.64	78.82	91.19
1	69.76	73.87	65.35	86.67

Table 7. G-means value of diagnosis results.

ρ	G-means			
	Resnet18	DQNimb	DQN	DEPDRL
100	0.994	0.985	0.982	0.996
25	0.897	0.916	0.882	0.968
20	0.869	0.921	0.847	0.961
15	0.853	0.897	0.815	0.951
10	0.848	0.874	0.793	0.939
5	0.817	0.728	0.767	0.912
1	0.679	0.716	0.603	0.849

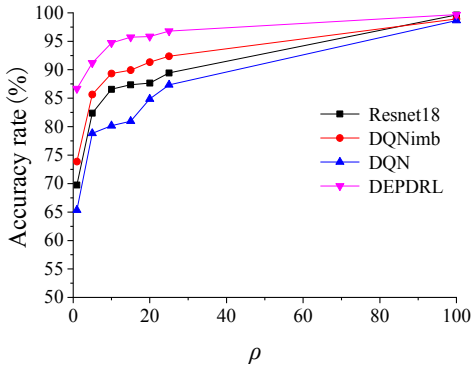

Fig. 5. Classification accuracy under different ρ.

Resnet18 and DQN, the classification accuracy of DQNimb on unbalanced data is improved, because the algorithm increases the reward and punishment values of a few types of samples in the classification of unbalanced data.

We report the statistical results of G-means evaluation indicators in Table 7. The results show that DEPDRL has higher diagnostic accuracy in fault diagnosis of rolling bearing unbalanced data. It also further illustrates the validity of DEPDRL.

We report the changes of classification accuracy of each model when ρ changes in Fig. 5. The results show that with the increase of the imbalance ratio range (ρ value decreases), the diagnostic accuracy of all models decreases, and the DQN has the largest range of change, indicating that with the increase of the imbalance ratio, the recognition ability of DQN model decreases for a few types of samples. The reduction range of

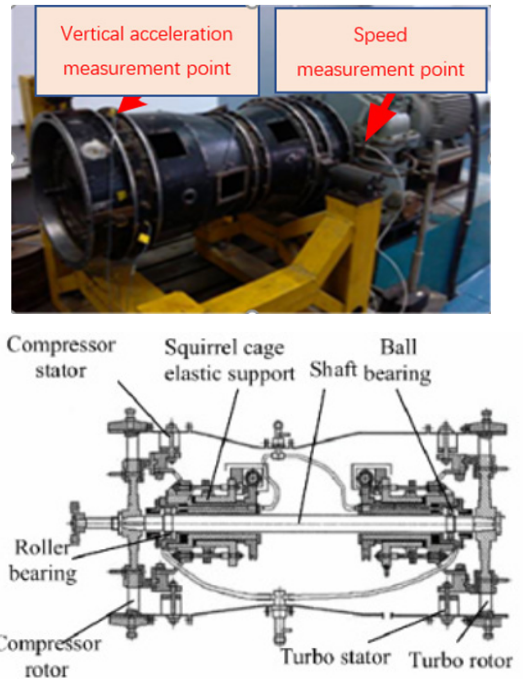

Fig. 6. Acro-engine rotor tester.

DEPDRL is small, because the model considers the improvement of the algorithm level and the balanced cross-sampling scheme at the data level, which enhances the recognition ability of the algorithm for minority classes and thus improves the classification accuracy of the algorithm.

4.2 Rolling bearing fault diagnosis test of aero-engine rotor tester with casing

The Laboratory of Intelligent Diagnosis and Expert System of Nanjing University of Aeronautics and Astronautics completed the rolling bearing fault test on the aero-engine rotor tester with casing as shown in Fig. 6 (the test platform is a 1:3 scale imitation of a real engine). The test platform has the ability to simulate the attenuation characteristics of aero-engine vibration signal during transmission. The single row deep groove ball bearing with bearing type 6206, vibration acceleration sensor (B&K4805) and NI USB9234 data collector were used in the test. The sampling frequency was 10240 Hz, the single sample data point was 8192, and the test speed was 1500, 1800, 2000 and 2400 (r/min). The sensor installation position is shown in Fig. 6. To simulate rolling bearing faults, 6mm wide cracks were processed on the outer and inner rings, respectively, by means of edm cutting during the test, and a depression with a radius of 0.5 mm and a depth of 2 mm was processed on the rolling body [4, 29]. The specific defects are shown in Fig. 7.

The original vibration signals were divided into test set and training set in a ratio of 1:4, and the time-frequency graph was also converted by wavelet time-frequency decomposition as the input of each classification model. The sample conversion time was 0.04 s. The unbalanced data are divided on the fault

Table 8. Failure unbalance test datasets.

ρ	Train set				Test set	
	Inner ring	Outer ring	Rolling fault	Normal	Fault	Normal
100	6896	7376	7808	8272	5520	2068
25	1724	1844	1952	8272	5520	2068
20	1379	1475	1562	8272	5520	2068
15	1034	1106	1171	8272	5520	2068
10	690	738	781	8272	5520	2068
5	345	369	390	8272	5520	2068
1	69	74	78	8272	5520	2068
Lable	0	1	2	3	Test	

Table 9. Diagnosis result.

ρ	Test result (%)			
	Resnet18	DQNimb	DQN	DEPDRL
100	97.56	97.32	96.54	98.86
25	91.37	93.27	89.62	95.12
20	90.65	90.97	88.36	95.23
15	87.23	88.33	84.68	93.28
10	84.49	86.27	78.39	90.64
5	79.92	81.68	73.89	88.83
1	65.39	72.18	60.67	78.23

Fig. 7. Fault 6206 ball bearings used in the experiment.

data set by using multiple imbalance ratios ρ. A variety of unbalanced data sets are shown in Table 8. DEPDRL, Resnet18, DQNimb and DQN are used to diagnose aero-engine bearing faults based on casing signals, and the fault diagnosis ability of DEPDRL on rolling bearing unbalance fault data set was verified by comparing various algorithms. Table 9 and Fig. 8 show the diagnostic results of each model under different imbalance ratios. The results show that DEPDRL has higher diagnostic accuracy than other algorithms in fault diagnosis of rolling bearing unbalance data based on casing detection points. Taking ρ = 5 as an example, the classification accuracy of DEPDRL is 88.83 %, which is 7.15 % higher than that of DQNimb. The results in Table 9 and Fig. 8 also reflect that the diagnostic accuracy of each model decreases with the increase of the range of the imbalance ratio (ρ value decreases). As a whole, DEPDRL has higher classification accuracy in unbalanced fault data. The advantages of the proposed model in fault diagnosis are demonstrated, and the accuracy of the proposed model is further illustrated.

Table 10. Diagnosis result.

ρ	G-means (%)			
	Resnet18	DQNimb	DQN	DEPDRL
100	0.969	0.965	0.963	0.985
25	0.905	0.923	0.887	0.952
20	0.901	0.906	0.863	0.941
15	0.873	0.869	0.840	0.919
10	0.832	0.844	0.772	0.895
5	0.784	0.813	0.730	0.881
1	0.636	0.718	0.582	0.767

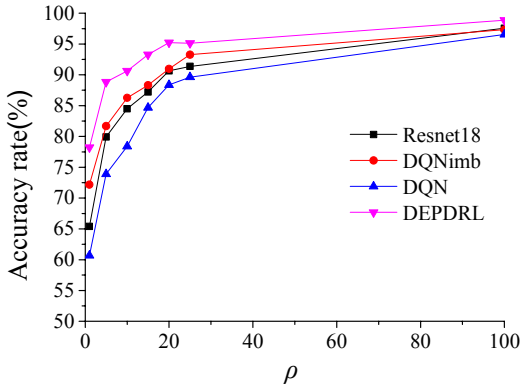

Fig. 8. Classification accuracy under different ρ.

We report the statistical results of G-means scores of different algorithms in Table 10. The results show that for DEPDRL, the range of G-means scores is 0.7677-0.952, which is higher than other algorithms within the defined ρ values. For example, the value of ρ = 10 % is 0.8926, compared with 0.8437 of DQNimb model. An increase of 0.0489. The results show that relative to the rest of the model calculation results, DEPDRL G-means scores higher, and further illustrates the DEPDRL method based on the site of casing in the fault diagnosis of rolling bearing unbalance data has more obvious advantages, can effectively improve the accuracy of fault diagnosis, and the algorithm has good generalization ability.

5. Conclusions

This paper proposes a new model, named DEPDRL, for unbalanced data classification, which has advantages both in algorithm and data level. A double parallel residual network model is proposed. At the data level, a DEPS is proposed to store unbalanced data, and then the unbalanced data is transformed into balanced data by balanced cross-sampling. After introducing the basic principle of the proposed model, we compared and verified the algorithm on three kinds of data sets. Finally, the proposed algorithm was applied to the standard data set of bearing fault diagnosis and the experimental data set of engine casing detection points, showing the advantages of fault diagnosis under unbalanced data. This also fully proves

that DEPDRL has high diagnostic accuracy and fast convergence speed, and also indicates that the network has a good application prospect.

Acknowledgments

This research is sponsored by the. National Science and Technology Major Project of China (J2019-IV-004-0071), Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX20_0211).

References

[1] Y. G. Lei, X. F. Xu and X. Cai, Research on data quality assurance for health condition monitoring of machinery, *Journal of Mechanical Engineering*, 57 (4) (2021) 1-9.

[2] Y. G. Lei, B. Yang and Z. J. Du, Deep transfer diagnosis method for machinery in big data era, *Journal of Mechanical Engineering*, 55 (7) (2019) 1-8 (in Chinese).

[3] L. Wen, L. Gao and X. Y. Li, A new deep transfer learning based on sparse auto-encoder for fault diagnosis, *Journal of Technology*, 49 (1) (2019) 136-144.

[4] X .Y. Zhang, G. Chen and T. F. Hao, Rolling bearing fault convolutional neural network diagnosis method based on casing signal, *Journal of Mechanical Science and Technology*, 34 (6) (2020) 2307-2316.

[5] X. J. Guo, L. Chen and C. Q. Shen, Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis, *Measurement*, 93 (7) (2016) 490-502.

[6] X. Zhang, H. K. Jiang and X. Zhang, Rolling bearing fault diagnosis using an optimization deep belief network, *Measurement Science and Technology*, 26 (11) (2015) 115002.

[7] Y. L. Wang, H. B. Yang and X. F. Yuan, Deep learning for fault-relevant feature extraction and fault classification with stacked supervised auto-encoder, *Journal of Process Control*, 92 (5) (2020) 79-89.

[8] W. Y. Huang, J. S. Cheng and Y. Yang, An improved deep convolutional neural network with multi-scale information for bearing fault diagnosis, *Neurocomputing*, 359 (24) (2019) 77-92.

[9] W. Q. Hou, M. Ye and W. H. Li, Rolling element bearing fault classification using improved stacked de-noising auto-encoders, *Journal of Mechanical Engineering*, 54 (7) (2018) 87-96.

[10] Z. Meng, Y. Guan and Z. Z. Pan, Fault diagnosis of rolling bearing based on secondary data enhancement and deep convolutional network, *Journal of Mechanical Engineering*, 57 (23) (2021) 106-115.

[11] S. H. Khan, M. Hayat and M. Bennamoun, Cost-sensitive learning of deep feature representations from imbalanced data, *IEEE Transactions on Neural Networks and Learning Systems*, 29 (8) (2017) 3573-3587.

[12] F. Zhou, S. Yang and H. J. Fu, Deep learning fault diagnosis method based on global optimization GAN for unbalanced data, *Knowledge-Based Systems*, 187 (2020) 104837.1-1048 37.19.

[13] H. Han, W. Y. Wang and B. H. Mao, Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning, *Proceedings of the 2005 International Conference on Advances in Intelligent Computing*, Springer (2005) 878-887.

[14] P. Soltanzadeh and M. Hashemzadeh, RCSMOTE: range-controlled synthetic minority over-sampling technique for handling the class imbalance problem, *Information Sciences*, 542 (2020) 92-111.

[15] H. Wang, Y. Ke and G. Luo, Compressed sensing of roller bearing fault based on multiple down-sampling strategy, *Measurement Science and Technology*, 27 (2) (2016) 025009.

[16] Z. Zheng, Z. Yu and Y. Wu, Generative adversarial network with multi-branch discriminator for imbalanced cross-species image-to-image translation, *Neural Networks*, 141 (2021) 355-371.

[17] Z. Wu, Y. Guo, W. Lin, S. Yu and Y. Ji, A weighted deep representation learning model for imbalanced fault diagnosis in cyber-physical systems, *Sensors*, 18 (4) (2018) 1096.

[18] W. Qian and S. Li, A novel class imbalance-robust network for bearing fault diagnosis utilizing raw vibration, *Measurement*, 156 (2020) 107567.

[19] L. Lin, B. Wang and J. Qi, Bearing fault diagnosis considering the effect of imbalance training sample, *Entropy*, 21 (4) (2019) 386.

[20] E. L. Lin, Q. Chen and X. M. Qi, Deep reinforcement learning for imbalanced classification, *Applied Intelligence*, 50 (2020) 2488-2502.

[21] S. Q. Kang, Z. Liu and Y. J. Wang, A fault diagnosis method of rolling bearing based on the improved DQN network, *Chinese Journal of Scientific Instrument*, 42 (3) (2021) 201-212.

[22] J. M. Johnson and T. M. Khoshgoftaar, Survey on deep learning with class imbalance, *Journal of Big Data*, 6 (1) (2019) 1-54.

[23] M. Volodymyr, K. Koray and S. David, Human-level control through deep reinforcement learning, *Nature*, 518 (7540) (2015) 529-533.

[24] M. A. Wiering, H. V. Hasselt and A. D. Pietersma, Reinforcement learning algorithms for solving classification problems, *2011 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL)*, Paris, France (2011) 91-96.

[25] A. K. Dixit and J. J. Sherrerd, *Optimization in Economic Theory*, Oxford University Press (1990).

[26] K. M. He, X. Y. Zhang, S. Ren and J. Sun, Deep residual learning for image recognition, *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, Las Vegas, USA (2016) 770-778.

[27] K. M. He, X. Y. Zhang and S. Q. Ren, Identity mappings in deep residual networks, *Proceedings of the European Conference on Computer Vision*, Amsterdam, Netherlands (2016) 630-645.

[28] H. D. Shao, H. K. Jiang and H. Z. Zhang, Rolling bearing fault feature learning using improved convolutional deep belief network with compressed sensing, *Mechanical Systems and Signal Processing*, 100 (1) (2018) 734-765.

[29] Y. X. Kang, G. Chen, X. K. Wei and L. Zhou, Deep residual hedging network and its application in fault diagnosis of roll-ing bearings, *Acta Aeronautica et Astronautica Sinica*, 43 (8) (2022)

625201, doi: 10.7527/S1000-6893.2021.25201.

Y. X. Kang received his Master's from Shenyang University of Aeronautics and Astronautics, ShenYang, P. R. China, in 2018. Now he is a Ph.D. student at the College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, P. R. China. His current research interests include rotating-machine fault diagnosis, deep learning signal analysis and processing

G. Chen received a Ph.D. in Mechanical Engineering from the Southwest Jiaotong University, Chengdu, P. R. China, in 2000. He works at the College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, P. R. China. His current research interests include the whole aero-engine vibration, rotor-bearing dynamics, rotating machine fault diagnosis, pattern recognition and machine learning, signal analysis and processing.