Research on Fault Diagnosis Technology Based on Deep Learning

Haisheng Wang¹*, Jian Wei¹, Pengjin Li²
¹Department of Weapons and Control, Army Academy of Armored Forces, Fengtai, Beijing 100072, China;
²Unit 66109 of PLA, Qinhuangdao, Hebei, 066000, China
*Corresponding author’s e-mail: 448912668@qq.com

Abstract. This paper introduces the basic theory, research status and challenges of fault diagnosis technology based on deep learning, and expounds the great application prospect of fault diagnosis technology based on deep learning.

1. Introduction

With the rapid development of science and technology, production efficiency and human life have been greatly improved, but the failure of equipment has caused more serious damages. Therefore, fault diagnosis has become an essential safety measure. The traditional fault diagnosis methods are difficult to deal with the problems of equipment with complex connection and massive heterogeneous data. The fault diagnosis technology based on deep learning adopts the method of artificial intelligence and applies data mining technology, which greatly improves the efficiency and accuracy of fault diagnosis.

2. Basic theory

2.1. Meaning of fault diagnosis

Fault diagnosis is the process of extracting and analyzing the characteristic data of the equipment to obtain the state of the equipment and judge whether the equipment is abnormal. Generally, fault diagnosis technology includes three tasks: the first is fault detection, which is used to detect equipment faults; the second is fault isolation, which is used to locate and classify faults; the third is fault estimation, which is used to determine the nature and level of faults.

2.2. Classification of fault diagnosis

There are three kinds of fault diagnosis methods: experience based method, model driven method and data driven method[1].

The experience based method is used in systems that lack information and are not easy to model, and requires a lot of expert experience; the model driven method relies on accurate mathematical model and requires a lot of expert experience too.

The data driven method mainly adopts various data mining technologies to obtain the key information hidden in the data, analyzes the state of the equipment, and achieves the goal of fault detection, diagnosis and isolation. Compared with the experience based and model driven methods, the data driven method solves the problem of relying too much on expert experience and mathematical model. With the rapid development of information technology, data-driven fault diagnosis technology...
has been developing rapidly and been widely used. Fault diagnosis based on deep learning is an important data-driven fault diagnosis method.

2.3. Fault diagnosis based on deep learning
Deep learning is a new research in the field of machine learning. It makes machine learning closer to the original goal -- artificial intelligence (AI). With the help of strong learning ability, deep learning transforms the low-level features of the input data into high-level features, simulates very complex functions, and obtains the inherent laws of sample data.

Fault diagnosis technology based on deep learning takes feature learning as the main purpose. With the development and application of Convolutional Neural Network (CNN), Deep Belief Network (DBN), Generative Adversarial Network (GAN), Transfer Learning and other algorithms, the application of fault diagnosis based on deep learning is more and more common.

3. Research status
According to the classification of deep learning, the fault diagnosis technology based on deep learning can be divided into: fault diagnosis based on DBN, fault diagnosis based on autoencoder (AE), fault diagnosis based on CNN, fault diagnosis based on Long-Short Term Memory (LSTM), etc.

3.1. Fault diagnosis based on DBN
DBN is a kind of probabilistic generative model. By training the weights between neurons, it makes the neural network generate training data with the maximum probability. The model is composed of multi-layer random hidden variables. The top two layers are connected without direction, the other lower layers are connected by top-down direction, and the bottom unit is used to input data vector. DBN contains multiple restricted Boltzmann machines (RBM), and adds classifiers at the last layer for data monitoring, recognition and classification. DBN with powerful high-dimensional nonlinear data processing ability and self-learning ability, does good at feature recognition and classification, and can realize effective and accurate fault location.

Li R Y et al. propose an improved DNN model[2]. The model introduces the "centering trick" in the pre-training process of network, and combines kernel principal component analysis (KPCA) with CDBN. Compared with KPCA-CDBN and KPCA-DBN, this method has better stability and prediction performance in fault diagnosis.

Zhang X et al. propose a method based on the combination of variational autoencoder and DBN (VAE-DBN)[3]. This method increases the constraints on the encoder to limit the hidden variables and reparameterize to do back propagation. This method not only reduces the amount of calculation, but also avoids the complex dimension problem, and it is easy to extract deep features from the data. Experiments show that the fault classification of this method is better than that of stacked autoencoder (SAE), and the maximum improvement is about 20%.

Wei Y Q et al. propose a fault diagnosis and identification method of nonlinear process based on DBN-dropout[4]. Compared with the traditional neural network model, in the fine-tuning stage of DBN network, dropout is added to each hidden layer to improve the generalization of the model.

Tian W D et al. propose a GAN-SRCC-DBN fault diagnosis method to solve the problems of insufficient data, missing data, measurement noise, redundant process variables and high coupling of data in fault diagnosis[5]. As shown in table 1, the average fault recognition accuracy of this method in 21 kinds of faults is 89.7%, which is 6.23% higher than that of DBN.
Table 1. Fault diagnosis rate (FDR) of TE process

| IDV  | FDR      | DBN | GAN-DBN | GAN-SRCC-DBN | GAN-PCC-DBN | GAN-MI-DBN |
|------|----------|-----|---------|--------------|-------------|------------|
| 0    | 0.941    | 0.989 | 0.997   | 0.995        | 0.994       |
| 1    | 0.978    | 0.997 | 0.997   | 0.996        | 0.993       |
| 2    | 0.951    | 0.981 | 0.998   | 0.984        | 0.985       |
| 3    | 0.208    | 0.241 | 0.33    | 0.402        | 0.283       |
| 4    | 0.972    | 0.983 | 0.992   | 0.988        | 0.977       |
| 5    | 0.936    | 0.938 | 0.967   | 0.974        | 0.971       |
| 6    | 0.992    | 0.999 | 0.999   | 0.997        | 0.991       |
| 7    | 0.957    | 0.971 | 0.973   | 0.981        | 0.971       |
| 8    | 0.934    | 0.943 | 0.95    | 0.954        | 0.935       |
| 9    | 0.191    | 0.201 | 0.422   | 0.41         | 0.28        |
| 10   | 0.946    | 0.962 | 0.965   | 0.974        | 0.961       |
| 11   | 0.948    | 0.948 | 0.971   | 0.955        | 0.97        |
| 12   | 0.965    | 0.971 | 0.993   | 0.99         | 0.994       |
| 13   | 0.917    | 0.931 | 0.957   | 0.94         | 0.955       |
| 14   | 0.927    | 0.934 | 0.944   | 0.948        | 0.92        |
| 15   | 0.471    | 0.51  | 0.633   | 0.552        | 0.603       |
| 16   | 0.775    | 0.907 | 0.924   | 0.919        | 0.902       |
| 17   | 0.78     | 0.822 | 0.935   | 0.901        | 0.926       |
| 18   | 0.914    | 0.943 | 0.955   | 0.961        | 0.949       |
| 19   | 0.945    | 0.959 | 0.982   | 0.97         | 0.981       |
| 20   | 0.881    | 0.9   | 0.952   | 0.932        | 0.96        |

Average 83.47% 85.86% 89.70% 89.15% 88.10%

Yu J B et al. propose a fault diagnosis method based on MI, DBN, KDE and SVDD (M-DBN)[6]. MI is responsible for automatically analysing variables, DBN is responsible for extracting high-order feature information, SVDD is responsible for integrating the functional information of each scheme, and KDE is used for monitoring and control.

Zhang Z P et al. propose a fault diagnosis model based on scalable DBN[7]. The model extracts the features in each domain, takes the output of the sub network of DBN as the input, and constructs a three-layer back propagation neural network as the global hierarchical network. The model generates a comprehensive fault diagnosis result according to the result of each DBN. The model trains the double-layer back propagation network for fault classification.

Yu J B et al. analyze each feature captured by DBN[8]. This study proposes an active feature (AF), which is used to represent the fault information of the original input data. This study selects the feature from each sample for on-line monitoring, uses Euclidean distance to calculate the dissimilarity between test samples and training ones, analyzes the impact of each feature extracted by DBN on the monitoring results, and selects the features used to fault detection. Experiments verify the feasibility and superiority of this method.
Ye Z W uses the DBN model for the fault identification of analog circuit[9]. In this study, the capability entropy of wavelet packet is extracted from the raw data, principal component analysis (PCA) method is used to reduce the dimension to obtain the eigenvector, and the eigenvector is input into the DBN model. Experiments show that this research has higher fault recognition rate than the common fault diagnosis methods of analog circuits.

Zhao C B proposes a multi-layer neural network for gearbox fault diagnosis[10]. In order to retain the local feature information and global information at the same time, the skipping-layer structure and the last convolution layer are used as the input of multi-scale feature layer in the deep convolutional neural network, and a polybasic feature fault diagnosis method is realized.

3.2. Fault diagnosis based on AE

AE uses the back propagation algorithm to restore the output of the decoder to the input data. AE can realize the efficient representation of input data with unsupervised learning. The efficient representation of input data is called coding, and its dimension is generally much smaller than that of input data, so AE can be used to reduce dimension. Moreover, AE can be used as a powerful feature detector for the pretreatment of deep neural networks. In addition, AE can also randomly generate data similar to the training data, which is called a generative model. AE has gradually evolved into a variety of networks, such as undercomplete autoencoder (UAE), regular autoencoder (RAE), sparse autoencoder (SAE), denoising autoencoder (DAE), contractive autoencoder (CAE), stacked autoencoder (SAE), etc.

Yu J B et al. propose a stacked denoising autoencoder (SDAE) model for process pattern recognition (PPR) in manufacturing processes[11]. This model gets useful information from process signals, then identifies faults with deep network, and visually presents the feature representation of SDAE. The effectiveness of the proposed PPR method is verified through a big simulation dataset and Tennessee Eastman process (TEP).

In order to solve the problem of imbalanced data, Zhou F N et al. combine GAN with SAE to generate more qualified data for fault diagnosis[12]. This research designs new generator and discriminator of GAN to generate more discriminant fault samples using a scheme of global optimization. The generator is used to generate fault feature extracted from a few fault samples via AE instead of fault data sample. The discriminator is used to filter the unqualified generated samples so that qualified samples are useful for improving the accuracy of fault diagnosis.

Zheng S D et al. propose a state recognition model based on unsupervised learning[13]. The model is used to process unlabeled data, and the dimension is reduced to a reasonable number by DAE. Experiments show that the recognition rate of some faults is 100%, but there are some unrecognizable faults.

Jiang L et al. propose a semi supervised fault classification method based on dynamic sparse stacked autoencoder (DSSAE), which is used to classify dynamic fault data. In this method, a hierarchical sparse artificial neural network is used to classify faults in dynamic process[14]. SAE is used to extract different faults. This method combines dynamic time windows into SAE to construct DSSAE. Experiments show that the performance of DSSAE method is better than SAE and SSAE.

Zhang Z et al. propose a stacked sparse autoencoder (SSAE) model, which extracts the deep feature vector from the original fault data[15]. In this study, a real-time fault diagnosis model is proposed. The trained model can diagnose faults in only a few milliseconds. As shown in table 2, experiments based on TEP show that the average FDR of the method is 90.11% and is higher than other methods.
Table 2. Fault diagnosis rate comparison

| Fault | PCA   | PLS   | NMFSC | SSAE |
|-------|-------|-------|-------|------|
| 1     | 99.88%| 99.88%| 100%  | 100% |
| 2     | 98.75%| 98.63%| 99.3% | 100% |
| 4     | 100%  | 99.5% | 100%  | 100% |
| 5     | 33.63%| 33.63%| 52.4% | 99.79%|
| 6     | 100%  | 100%  | 100%  | 100% |
| 7     | 100%  | 100%  | 100%  | 100% |
| 8     | 98%   | 97.88%| 98.4% | 83.96%|
| 10    | 60.5% | 82.63%| 50.4% | 76.77%|
| 11    | 78.88%| 78.63%| 69.9% | 78.85%|
| 12    | 99.13%| 99.25%| 98.1% | 93.70%|
| 13    | 95.38%| 95.25%| 95.9% | 78.65%|
| 14    | 100%  | 100%  | 83.8% | 98.96%|
| 16    | 55.25%| 68.38%| 45.0% | 86.04%|
| 17    | 95.25%| 94.25%| 71.4% | 88.33%|
| 18    | 90.5% | 90.75%| 93.4% | 88.75%|
| 19    | 41.13%| 26%   | 33.0% | 81.25%|
| 20    | 63.38%| 62.75%| 49.0% | 76.77%|
| Average| 82.74%| 83.97%| 79.72%| 90.11%|

Wen J T et al. propose a SSAE model for fault diagnosis[16]. Because the vibration signal is sparse matrix, the random Gaussian matrix is used to transform and project the signal in frequency domain. The model compresses fault information to low dimensions and achieves good results.

3.3. Fault diagnosis based on CNN

CNN was initially used to minimize preprocessing operations. CNN is of characteristic invariance. When the convolution kernel with the same parameters is applied to different positions of the previous layer, it can obtain different features. There are five layers in CNN: input layer, convolution layer, pooling layer, full connection layer and output layer. CNN performs convolution operation on each layer to obtain local features, and integrates local features within the layer. CNN performs well in classification, detection and recognition.

Wu H et al. propose a fault diagnosis method based on DCNN[17], which includes convolution layer, pooling layer, dropout layer and full connection layer. This method extracts the features of space domain and time domain for fault diagnosis. The accuracy of fault diagnosis in TEP can reach 88.2%, and the accuracy of general fault diagnosis can reach more than 91%, but this method is not suitable for the case of little historical data.

Zhao Y K et al. use VGG16 transfer learning model for fault diagnosis[18]. The structure and weight of VGG16 CNN model do not need training, so it saves time. In this study, the vibration signal is converted into image as the data input of VGG16 network. This study performs well in different working conditions and classification tasks.

Zhang C L et al. combine complementary ensemble empirical mode decomposition (CEMMD) with CNN to extract and analyse fault feature data of bearing[19]. This method effectively improves the fault identification rate of bearing under multiple working conditions.

Leren E directly processes the original signal in time domain using CNN processing one-dimensional data to achieve the fault diagnosis of vibration signal of motor bearing[20]. In this research, the raw vibration data is fed into the proposed system, so that there is no need to use a separate feature extraction algorithm every time for classification.

3.4. Fault diagnosis based on LSTM

LSTM is a timing-cycle neural network. It is mainly used to solve the problem of long-term dependence of general RNN (cyclic neural network). Due to its unique design structure, LSTM is suitable for processing and predicting important events with very long interval and delay in time series.
Zheng S D et al. propose an on-line fault diagnosis method of unsupervised data mining[21]. CSAE extracts low dimensional features from the original high-dimensional data in convolution layer or LSTM layer. This method can be used for feature visualization. This method obtains data mining results from visualized features by clustering.

Dou S et al. propose a method based on LSTM time series reconstruction[22]. This method uses one-layer LSTM network to represent the time series of data, and uses another one-layer LSTM network to reconstruct the time series in reverse order. This method achieves the fault detection of equipment with maximum likelihood estimation.

Xu P et al. propose a pipeline leakage prediction model based on GAN and LSTM[23]. This model applies sliding window and key feature selection method, uses a data enhancement method based on GAN to process fault data to solve the imbalance problem, uses the classifier method based on LSTM to learn the time correlation of data, and classifies the state of pipeline to predict leakage. Experiments show that this method has higher prediction accuracy than the existing methods.

Zhao H T et al. propose a fault diagnosis method based on LSTM[24]. This method can classify the raw data directly, and adaptively learn the dynamic information in the raw data. Experiments based on TEP show that this method can distinguish faults well.

Kim T Y et al. propose a C-LSTM neural network to effectively process the spatiotemporal information contained in the data[25]. This method combines CNN, LSTM and DNN to effectively extract the characteristics of data. CNN extracts spatial information to reduce the change of frequency, LSTM extracts temporal information, and DNN maps data. This method can detect anomalies in web traffic data very well.

Wei X L et al. propose a fault diagnosis method based on the combination of LSTM and one-dimensional CNN[26]. This method identifies faults by vibration signals with different pressures at the inlet. Experiments show that this method can accurately identify four kinds of faults, and the accuracy is as high as 99.5%. The visualization of dimensionality reduction of each layer in this method is shown in figure 1.

![Figure 1. Visualization of dimensionality reduction.](image)

Cheng Q Z et al. propose a fault diagnosis method based on CNN and LSTM[27]. In this study, CNN is used to extract horizontal features to obtain the spatial and temporal characteristics and extract vertical features to solve the problem of single feature type and slow training, and LSTM is used for fault diagnosis.
3.5. Other relevant studies

In order to solve the problems that it is difficult to get fault data and data is of imbalance in industrial processes, Gao X et al. propose a GAN model based on gradient penalty[28]. This method generates data samples to supplement the missing fault data and improve the generalization and accuracy of the classifier.

Zhou Y Y preprocesses the truck image, extract the image features by PCA, and uses radial basis function neural network to automatically identify the fault. Based on image processing and neural network, the accuracy of fault diagnosis reaches 90%[29].

Aiming at the phenomenon of aliasing of some fault categories in analog circuits, Zhang C L et al. propose an analog circuit fault diagnosis method based on quantum neural network algorithm[30].

Shi P M et al. propose an intelligent diagnosis model based on mathematical statistics[31]. Combined with the feature extraction of deep learning and the state recognition of particle swarm and support vector machine, the model can accurately identify the fault of gear.

4. Challenges

Deep learning can be understood as a hypothesis: the search of the objective function is equivalent to the optimal problem of the objective function. Therefore, a lot of optimization work is needed to correct the error between hypothesis and reality. Generally, there are only two ways to solve this problem: more effective data and stronger computing power. Therefore, fault diagnosis based on deep learning is mainly subject to the above two points.

4.1. The processing of raw data is the focus

Fault diagnosis based on deep learning is useless without the processing of massive raw data. In terms of raw data processing, the following problems are usually faced:

4.1.1. Data missing. The lack of raw data will seriously affect the effect of fault diagnosis based on deep learning. In the existing research, this problem is mainly solved by data filling. First, data filling is carried out by traditional algorithms such as expectation maximization algorithm, mean imputation algorithm, hot deck imputation algorithm and multiple imputation algorithm. Second, data filling is carried out by deep learning, which have been well explored. A MEWT-KELM-CS-MIMO model to fill sensor fault data by deep learning is proposed by [32], the performance is shown in table 3 and improvement means how much better this model is than other models, and the biggest improvement is up to 76.75%. But the existing research only improves the raw data, and this problem is not resolved in data collection.

| Table 3. Prediction performance indexes of four prediction models |
|---------------------------------------------------------------|
| **Experiment** | **Model** | **MAE** | **Improv.** | **MAE** | **Improv.** |
| 1 | EWT-KELM-CS | 0.438 | 8.22% | 1.330 | 46.69% |
| | NA-MEMD-KELM-CS | 1.729 | 76.75% | 2.147 | 66.98% |
| | MEWT-KELM-CS | 0.471 | 14.65% | 0.977 | 27.43% |
| | MEWT-KELM-CS-MIMO | 0.402 | — | 0.709 | — |
| 2 | EWT-KELM-CS | 0.469 | 0.211 | 0.910 | 18.35% |
| | NA-MEMD-KELM-CS | 1.658 | 0.777 | 2.090 | 64.45% |
| | MEWT-KELM-CS | 0.542 | 0.317 | 0.948 | 21.62% |
| | MEWT-KELM-CS-MIMO | 0.370 | — | 0.743 | — |
| 3 | EWT-KELM-CS | 0.392 | 0.084 | 0.713 | 12.20% |
| | NA-MEMD-KELM-CS | 1.401 | 0.744 | 1.674 | 62.60% |
| | MEWT-KELM-CS | 0.458 | 0.216 | 0.929 | 32.62% |
| | MEWT-KELM-CS-MIMO | 0.359 | — | 0.626 | — |
4.1.2. Data exception. The error of raw data will inevitably affect the effect of deep learning. In the process of fault diagnosis, the raw data are prone to be with errors or space-time disorder in the process of measurement, transmission and storage. If there is no special preprocessing mechanism or human intervention, the wrong raw data will seriously affect the effect of fault diagnosis based on deep learning.

4.1.3. Data imbalance. In machine learning, if the probability distribution of training data set and test data set is consistent, the model will achieve good results. If the probability distribution is inconsistent, the result may be wrong. Existing studies mostly use knowledge transfer to solve such problems. By finding the similarity between different data, knowledge transfer can reduce the distribution difference between the two fields as much as possible in a certain space to achieve the accurate classification of data. Specific methods of knowledge transfer include: domain adaptive methods based on data distribution[33-34], feature selection[35-36] and feature transformation[37-38].

As shown in figure 2, the JDA method proposed by [34] achieves much better performance than the five baseline methods with statistical significance, and this method is one of domain adaptive methods based on feature selection.

Figure 2. Accuracy (%) on the 20 cross-domain image datasets.

As shown in figure 2, the JDA method proposed by [34] achieves much better performance than the five baseline methods with statistical significance, and this method is one of domain adaptive methods based on feature selection.

Figure 3. Classification accuracy of LR, SVM, CDSC, TCA, LMTTL, SSKM, ARRLS and ARRLS+. 
As shown in figure 3, ARTL (such as ARRLS and ARRLS+) proposed by [36] generally outperforms all baseline methods, and this method is one of domain adaptive methods based on data distribution.

| Dataset | LR   | TCA  | DAM  | TJM  | uSCA | SCA  |
|---------|------|------|------|------|------|------|
| C → A   | 92.28| 92.17| 92.38| 93.53| 93   | 93.11|
| C → W   | 81.02| 83.39| 86.1 | 89.15| 83.05| 85.42|
| C → D   | 89.17| 87.26| 89.81| 89.81| 89.17| 89.81|
| A → C   | 85.75| 86.46| 86.38| 86.82| 92   | 92   |
| A → W   | 77.29| 85.76| 82.71| 89.49| 81.02| 86.1 |
| A → D   | 87.9 | 87.26| 85.35| 88.54| 88.4 | 89.17|
| W → C   | 73.91| 82.37| 76.49| 82.64| 81.74| 84.42|
| W → A   | 77.14| 90.4 | 79.65| 89.46| 88.94| 89.24|
| W → D   | 100  | 100  | 98.09| 98.73| 100  | 100  |
| D → C   | 79.52| 82.99| 81.21| 82.73| 85.75| 86.02|
| D → A   | 86.95| 90.5 | 88.94| 90.19| 91.54| 91.75|
| D → W   | 98.98| 98.98| 98.64| 97.63| 99.66| 99.66|
| Avg.    | 85.83| 88.96| 87.15| 89.89| 89.52| 90.56|

As shown in table 4, SCA proposed by [38] achieves the highest average accuracy over 12 cross-domain cases at lower computational cost than its closest competitor TJM, with 8 best and 2 second best cross-domain performance, and this method is one of domain adaptive methods based on feature transformation.

4.2. The setting of model parameters is difficult

The model based on deep learning is very complex, which involves many factors, such as the selection of training data, the division of distributed computing, the design of neural network structure, the setting of learning efficiency and so on. For the same model, gradient explosion, gradient disappearance and other problems will occur due to different settings of parameters, resulting in great differences in results. The fault diagnosis technology based on deep learning needs a major expenditure of time and effort to adjust parameters, which is facing great challenges.

4.3. The support of relevant technologies is very important

Fault diagnosis based on deep learning depends on the timely acquisition of accurate data and the rapid processing of massive data. The former needs the support of stable and reliable sensor technology and high real-time deterministic network technology; The latter needs the support of high-capacity storage technology and high-speed processing technology. The development of these technologies also affects the effect of fault diagnosis based on deep learning.

5. Conclusion

With the development of science and technology, the continuous improvement of data acquisition, the continuous maturity of data mining, the continuous improvement of computing power and the continuous optimization of algorithms, it is possible to apply artificial intelligence analysis for fault diagnosis. Fault diagnosis technology based on deep learning has become a popular research and continues to mature. It will solve the difficulty of modeling, identification, and positioning of traditional fault diagnosis, and ensure the safety of production and life.

References

[1] Li, H., Xiao, D. (2011) Survey on data driven fault diagnosis methods. Control and Decision, 26(1): 1-16.
[2] Li R Y, Ni J H, Qin L, et al. Fault detection and diagnosis based on KPCA-CDBNs model. 2017 Chinese Automation Congress (CAC).

[3] Zhang, X., Cui, Z., Dong, Y., et al. (2018) Application of Fault Classification Method Based on VAE-DBN in Chemical Process. The Chinese Journal of Process Engineering, 18(3): 590-594.

[4] Wei, Y., Weng, Z. (2020) Research on TE process fault diagnosis method based on DBN and dropout. Canadian Journal of Chemical Engineering, 98(6): 1293-1306.

[5] Tian, W., Liu, Z., Li, L., et al. (2020) Identification of abnormal conditions in high-dimensional chemical process based on feature selection and deep learning. Chinese Journal of Chemical Engineering, 28(7): 1875-1883.

[6] Yu, J., Yan, X. (2020) Modeling large-scale industrial processes by multiple deep belief networks with lower-pressure and higher-precision for status monitoring. IEEE Access, 8: 20439-20448.

[7] Zhang, Z., Zhao, J. (2017) A deep belief network based fault diagnosis model for complex chemical processes. Computers and Chemical Engineering, 107: 395-407.

[8] Yu, J., Yan, X. (2019) Active features extracted by deep belief network for process monitoring. ISA Transactions, 84:247-261.

[9] Ye, Z. (2021) Intermittent fault detection of analog circuits based on deep belief network. M. S. thesis. Jiangxi University of Science and Technology. Jiangxi. China. pp: 1-56.

[10] Zhao, C. (2021) Research on Fault Diagnosis Method Based on Deep Learning and Image Processing. M. S. thesis. Shenyang Ligong University. Liaoqing. China. pp: 1-51.

[11] Yu, J., Zheng, X., Wang, S. (2020) A deep auto-encoder feature learning method for process pattern recognition. Process Control, 79: 1-15.

[12] Zhou, F., Yang, S., et al. (2020) Deep learning fault diagnosis method based on global optimization GAN for unbalanced data. Knowledge-Based Systems 187: 1-19.

[13] Zheng, S., Zhao, J. (2020) States identification of complex chemical process based on unsupervised learning. Computer Aided Chemical Engineering, 44: 2239-2244.

[14] Jiang, L., Ge, Z., Song, Z. (2017) Semi-supervised fault classification based on dynamic Sparse Stacked auto-encoders model. Chemometrics and Intelligent Laboratory Systems, 168: 72-83.

[15] Zhang, Z., Ren, X., Lv, H. (2018) Fault Diagnosis with Feature Representation Based on Stacked Sparse Auto Encoder. In: 2018 33rd Youth Academic Annual Conference of Chinese Association of Automation. Nanjing. China. pp. 18-20.

[16] Wen, J., Yan, C., Sun, J., et al. (2018) Bearing fault diagnosis method based on compressed acquisition and deep learning. Chinese Journal of Scientific Instrument, 39(01): 171-179.

[17] Wu, H., Zhao, J. (2018) Deep convolutional neural network model based chemical process fault diagnosis. Computers and Chemical Engineering, 115: 185-197.

[18] Zhao, Y., Xu, G., Liu, M. (2020) Method for fault diagnosis of bearing based on transfer learning with VGG16 model. Spacecraft Enviroimeng Engineering, 37(05): 446-451.

[19] Zhang, C., Fan, Y. (2019) Fault Diagonsis of Bearing using Feature Extraction Method Based on CEEMD Algorithm and CNN. Mechanical Science and Technology for Aerospace Engineering, 38(2): 178-183.

[20] Levent, E. (2017) Bearing Fault Detection by One-Dimensional Convolutional Neural Networks. Mathematical Problems in Engineering, 10(3): 1-9.

[21] Zheng, S., Zhao, J. (2020) A new unsupervised data mining method based on the stacked auto-encoder for chemical process fault diagnosis. Computers and Chemical Engineering, 135: 106755.

[22] Dou, S., Zhang, G., Xiong, Z. (2019) Anomaly detection of process unit based on LSTM time series reconstruction. CIESC Journal, 70(2): 481-486.

[23] Xu, P., Du, R., Zhang, Z. (2019) Predicting pipeline leakage in petrochemical system through GAN and LSTM. Knowledge-Based Systems, 175: 50-61.
[24] Zhao, H., Sun, S., Jin, B. (2018) Sequential Fault Diagnosis Based on LSTM Neural Network. IEEE Access, 6: 12929-12939.
[25] Kim, T., Cho, S. (2018) Web traffic anomaly detection using C-LSTM neural networks. Expert Systems With Applications, 106: 66-76.
[26] Wei, X., Chao, Q., Tao, J., et al. (2020) Cavitation fault diagnosis method for high-speed plunger pumps based on LSTM and CNN. Acta Aeronautica et Astronautica Sinica, 42(3): 423876.
[27] Chen, Q., Chen, Z., Zhang, Y., et al. (2020) Research on fault diagnosis of solar photovoltaic module based on CNN-LSTM. Computer Technology and Its Applications, 46(4): 66-76.
[28] Gao, X., Deng, F., Yue, X. (2020) network with gradient penalty. Neurocomputing, 396: 487-494.
[29] Zhou, Y. (2007) The Study of Machine Diagnosis Based on the Theory About Image Analysis. Ph. D. dissertation. Huazhong University of Science & Technology. Wuhan. China. pp: 1-110.
[30] Zhang, C., He, Y., Yuan, L., et al. (2015) An Analog Circuit Fault Diagnostics Approach Based on QNN. Journal of Liaoning Shihua University, 35(02): 58-61.
[31] Shi, P., Liang, K., Zhao, N., et al. (2017) Intelligent Fault Diagnosis for Gears Based on Deep Learning Feature Extraction and Particle Swarm Optimization SVM State Identification. Chinese Journal of Mechanical Engineering, 28(09):1056-1061+1068.
[32] Li, C., Zhang, J. (2019) Sensor fault prediction based on MEWT. Journal of Vibration, measurement & Diagnosis, 39(1): 197-208+230-231.
[33] Tahmoresnezhad, J., Hashemi, S. (2016) Visual domain adaptation via transfer feature learning. Knowledge & Information Systems, 50(2): 1-21.
[34] Long. M., Wang. J., Ding. G., et al. (2013) Transfer feature learning with joint distribution adaptation. In: Proceedings of the IEEE international conference on computer vision. Sydney. Australia. pp: 2200-2207.
[35] Nannan, L., Fei, C., Haoran, Q., et al. (2018) A new domain adaption algorithm based on weights adaption from the source domain. IEEE Transaction on Electrical and Electronic Engineering, 13(12): 1769-1776.
[36] Long. M., Wang. J., Ding. J., et al. (2014) Adaptation regularization: A general framework for transfer learning. IEEE TKDE, 26(5): 1076-1089.
[37] Sun, B., Saenko, K. (2016) Deep coral: Correlation alignment for deep domain adaptation. In European Conference on Computer Vision, 16(7):443-450.
[38] Ghifary. M., Balduzzi. D., Kleijn. W., et al. (2017) Scatter Component Analysis: A Unified Frame work for Domain Adaptation and Domain Generalization. IEEE transactions on pattern analysis and machine intelligence. 39(7):1414-1430.