A Regularized Deep Clustering Method for Fault Trend Analysis

Yongzhi Qu[1], Yue Zhang[2], David He[3], Miao He[4], Zude Zhou[5]

[1,2,5]*School of Mechanical and Electronic Engineering, Wuhan University of Technology, Wuhan, Hubei, China, 430070*
[1]*Department of Mechanical and Industrial Engineering, University of Minnesota Duluth, Duluth, MN 55812*
[3]*Department of Mechanical and Industrial Engineering, University of Illinois at Chicago, Chicago, IL, 60607*
[4]*National Oilwell Varco, Houston, Tx, 77036*
yongzhi@d.umn.edu
davidhe@uic.edu
brainwilde@gmail.com
zudezhou@whut.edu.cn

ABSTRACT

Effective fault feature extraction is the key of fault diagnosis. In previous works, it is shown that some embedding methods and unsupervised deep learning methods have the ability to extract fault features from raw signals directly, such as PCA and deep autoencoder. Particularly, deep autoencoder has been shown in relevant research that it can effectively extract the hidden 'trend' associated with machinery health states which can be used directly for online anomaly detection and prediction. However, in practical online fault diagnosis, the discrimination between successive signals is small due to the slow degradation progress and the external noise. Therefore, it is important to optimize the feature extraction process to achieve better online fault tracking. In this paper, a regularized deep clustering algorithm is proposed to guide the optimization process of feature extraction which combines embedding method and semi-guided learning. A regularization term for the cluster center points is proposed to make the feature optimization converge in a monotonic linear trend. In order to verify the effectiveness of the method, an accelerated gearbox run-to-failure experiment is carried out. The result shows that the feature optimization method can optimize the fault features on the basis of the deep autoencoder algorithm in two aspects: a better distinction of the fault features in short term and a more consistent trend of the gear wear in the long term.

1. INTRODUCTION

Gears are widely used in mechanical power transmission, as one of the most important rotary machines. The research of gear fault diagnosis is a popular topic. Many different gear fault analysis methods have been researched and proposed in

literature. In the past few, many research were focused on time frequency domain analysis and traditional signal processing methods. Several preeminent methods such as EMD(empirical mode decomposition), Wigner-Ville distribution, ICA (Independent Component Analysis), cyclic statistical processing, have been proved to be effective for processing gear signals, which are highly non-linear and nonstationary (Ricci et al., 2012 and Žvokelj et al., 2016). However, in practice, the application of these methods often requires complex professional knowledge and is not conducive to the automation of gear fault diagnosis.

In the past few years, many intelligent methods based on data-driven methods have been proposed, which only require data collection with sensors form different periods of machinery running stages. Those methods do not require the professional knowledge of the specific mechanical systems. Therefore, it is possible to achieve on-line health monitoring of gear status. Among the intelligent methods, deep learning methods have shown exceptional performance. Sun et al. (2018) utilized stack sparse autoencoders-based deep neural network for network establishing and then performed with a supervised fine-tuning process for classification of bearing faults. Similar types of methods with unsupervised network initialization and supervised fine-tune have been proved very successful in many fault diagnosis tasks (Chen et al., 2017; Lu et al., 2017). Shao et al. (2018) utilized deep autoencoder and proposed ensemble deep auto-encoders (EDAEs) for intelligent fault diagnosis of bearings. It trained DAEs model with different activity functions and the learned features are fed into Softmax classifiers for accurate and stable fault classification based on a combination strategy. With the use of ensemble learning, the method was more effective and robust than the existing intelligent diagnosis methods. Yu et al. (2018) combined the merit of different activation functions in a joint learning fashion and propose a joint multiple reconstructions autoencoder (JMRAE) for fault diagnosis. It was shown that more discriminative and robust

scale-specific feature representations at different scales were learned. All of the above researches used Autoencoder to get the feature representation and utilized classifiers for fault classification. Jiang et al. (2018) proposed an intelligent fault diagnosis method to automatically identify different health conditions of wind turbine gearbox. The method focused on the multiscale feature learning of complex vibration signals. Shao et al. (2018) proposed a convolutional deep belief network (CDBN) with Gaussian visible units to learn the representative features of bearing fault and utilized exponential moving average to improve the performance of the constructed model. Wen et al. (2018) proposed a new CNN based on LeNet-5, and applied the CNN model to the fault diagnosis. The proposed methods were tested on three datasets, including motor bearing dataset, self-priming centrifugal pump fault diagnosis dataset, and axial piston hydraulic pump fault diagnosis dataset. The results outperformed other DL and traditional methods. Pan et al. (2018) proposed a new network called LiftingNet for fault classification from raw mechanical data which also utilized CNN model. In addition to the CNN model, the application of RNN model is also popular in the field. Zhao et al. (2018) proposed local feature-based gated recurrent unit networks (LFGRU) for machine health monitoring. The method combined GRU network and Softmax/Regression Layer for bearing fault classification. Besides GRU, the LSTM network has also used in the relative researches (Li et al., 2018; Hinchi & Tkiouat, 2018). In the above studies with CNN or RNN, the methods all belong to supervised deep learning method. Such methods usually require large number of training data for different systems and different types of faults. In many cases, the faulty historical data is hard to obtain. Moreover, supervised models generally have poor generalization capability. The model obtained by network training upon one dataset can hardly be useful for other dataset or other machines. Although transfer learning has also been studied for this purpose (Sun et al., 2019), the application is still quite limited.

The study of gear condition detection and fault diagnosis is mainly divided into three parts: (1) Acquisition of condition information, (2) Feature extraction and feature representation, (3) Pattern recognition and classification. It is found that in the past research, the pattern recognition and classification is the main research direction, as shown in the above literature review. However, the application of supervised deep learning model is limited in practical situations. On the contrary, existing research on the other key point -- feature extraction are quite limited. In the paper, unsupervised feature learning method will be investigated and feature extraction is the main task in our work.

It is proved in our previous work (Qu et al., 2019) that autoencoder has the ability to effectively extract the hidden 'trend' associated with machinery health state which can be used directly for online anomaly detection and prediction. However, in practical online fault diagnosis, the discrimination between successive signals is subtle due to the slow evolving progress of the fault and the masking noise. Therefore, it is important to optimize the features extracting process to achieve online fault feature learning. Clustering is one of the major data mining methods. It aims to measure the spatial distance between samples. As an unsupervised method, it explores the intrinsic connection of data.

In this paper, deep embedded clustering method (DEC) (Xie et al., 2016) is firstly used to extract and optimize gear fault features unsupervisedly and a regularization of the cluster center points is proposed to make the feature optimization process converge in certain trends. The intuitive of the proposed improved DEC method with a regularization term lies in that the fault deterioration should contain an intrinsic trend. Assuming the trend is linear, we can potentially use this information with unsupervised methods to learn a more accurate embeddings for the fault features, and therefore enable unsupervised online fault diagnosis and prognosis.

The rest of the paper is organized as follows. Section 2 introduces the method of DEC and the regularization method. Section 3 presents the experimental work. Section 4 gives the results and discussion. Section 5 concludes the paper.

2. DEEP EMBEDDED CLUSTERING (DEC)

For the purpose of completeness, we will give a brief review of the DEC method first. DEC is a method combining deep embedding and clustering. Embedding is a nonlinear mapping from high dimension space to low dimension space. In the embedding space, the relationship can be more obvious, which is the merit of using embedding method. In this paper stack Autoencoder is chosen as the embedding method, which is shown in Figure 1. In the DEC method (Xie et al., 2016), the soft clustering and self-learning are the key points.
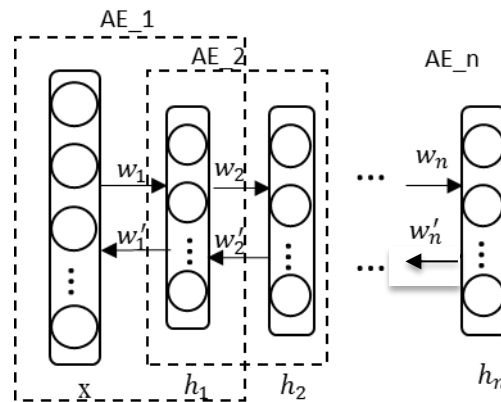

Figure 1. The structure of Stack Autoencoder.

2.1. Soft Clustering

In the DEC method, K-means is opted as the clustering method. Soft clustering method, as its name implies, do not use hard label criteria, but allocated label according to

percentage probability (weight). K-means can also be used with soft clustering method. The essence is the Gaussian mixture model, and the method for its training process is EM (Expectation Maximization) algorithm. The distinct feature of soft clustering is that the distance between sample and each center point is calculated by a similarity method. It should also be pointed out that in deep clustering method, the clustering process is conducted on the embedded space, not on the original space. Embedded space belongs to low-dimensional space. In low-dimensional space, t-distributed stochastic neighbor embedding (t-SNE) method proposed by Hinton et al. (2006) is used to measure the distance between points y, as shown in Eq. (1):

$$q_{i,j} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq l}(1 + \|y_k - y_l\|^2)^{-1}} \tag{1}$$

Generally, the probability distribution of high-dimensional data is Gauss distribution, so the distance between points x in original space $p_{i,j}$ as in Eq. (2):

$$p_{i,j} = \frac{\exp\left(-\|x_i - x_j\|^2/2\sigma^2\right)}{\sum_{k \neq l}\exp\left(-\|x_k - x_l\|^2/2\sigma^2\right)} \tag{2}$$

In soft clustering, Eq. (1) is changed to Eq. (3) when t-SNE is used to calculate the distance between the embedding point and the center point.

$$q_{i,j} = \frac{(1 + \|z_i - \mu_j\|^2)^{-1}}{\sum_{j'}(1 + \|z_i - \mu_j\|^2)^{-1}} \tag{3}$$

where z_i represents i[th] embedded space sample point, or mapping space sample point, μ_j represents the jth clustering center.

2.2. DEC Target and Training

Another key point of DEC is the application of self-learning, and the focus of self-learning is to find a target distribution p, which is related to the current distribution q as in Eq. (3), optimize q according to p, and then re-calculate p, and continue to optimize until the termination condition is met, thus forming a self-learning method. According to the need, target p should achieve the following three goals: (1) the accuracy of clustering needs to be improved; (2) put more emphasis on data points assigned with high confidence; (3) making the probability sum of 1, that is to say, it needs to be normalized. Thus, the target distribution p is written as follows (Xie et al., 2016):

$$p_{i,j} = \frac{q_{i,j}^2/f_j}{\sum_{j'} q_{i,j'}^2/f_{j'}} \tag{4}$$

where $f_j = \sum_i q_{ij}$.

The Eq. (4) utilizes $q_{i,j}^2$ square term, according to function property of quadratic function in range [0, 1], if $q_{i,j}$ is small,

then $q_{i,j}^2$ will be smaller, if $q_{i,j}$ is large, then $q_{i,j}^2$ will be larger, and the larger $q_{i,j}$ lead the change of greatly, which can achieve the goal of improving the accuracy of clustering.

Target distribution p is recalculated after each iteration of training, so termination conditions need to be set: iteration stops when less than 1 % of the points change the category. The Kullback-Leibler (KL) divergence is used to measure the loss function. The equation is as follows:

$$L = KL(p\|q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \tag{5}$$

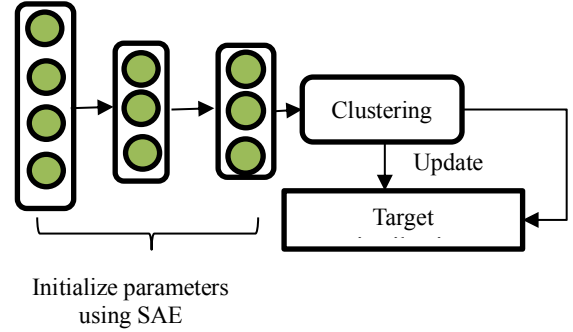The structure of DEC is shown in Figure 2 (Xie et al., 2016):

Figure 2. The structure of DEC.

The whole network training process is divided into two parts: the first part is the initialization process of embedded layer parameters, which is initialized by stack autoencoder to form the embedded network; the second part is clustering optimization by using the self-learning method.

Detailed illustration of the DEC method can be found in (Xie et al., 2016).

2.3. Regularization of the Cluster Center Points

The main purpose of DEC method is to improve the accuracy of clustering, by increasing the degree of discrimination between sample categories. It is often impossible to guarantee the original trend of features. However, unsupervised methods often need to extract features with a certain regularity for analysis, rather than just for differentiation. Therefore, adding regularization term could enforce the clustering optimization to be carried out in a certain direction. In this paper, a regularized constraint method for clustering center points is proposed, which is illustrated next.

In the training process, if the clustering center can present a linear relationship, the embedding feature can show progressive relationship in at least some dimensions, because the clustering centers determine the general position of the sample points of each cluster it contains. It should be noted that if the dimension of the last embedded layer is too high, the network training might fail or be difficult to be constrained due to large number of constraints during

3

training. In order to simplify the difficulty of constraints and prevent the failure of training network parameters, the dimension of the last layer in the autoencoder network layer is set to 2. The loss function of DEC is redefined as follows:

$$L = \sum_i \sum_j p_{ij} \, log \frac{p_{ij}}{q_{ij}} + \alpha \cdot C \quad (6)$$

$$C = \sum_{i=1}^{N-1} \left(\frac{\mu_{i+1,1}}{\mu_{i,1}} - \frac{\mu_{i+1,2}}{\mu_{i,2}} \right)^2 \quad (7)$$

In Eq. (6), the first term is the loss function of the original DEC, the second term is the regularization term of the cluster center. α is the regularization parameter of clustering center, N is the number of clustering clusters. By the regularization of the second term, the clustering center point can be trained toward a linear relationship.

The gradient equations for the parameters of the training network is given below. The Stochastic Gradient Descent (SGD) is used for training. The first part is the optimization of DAE parameters, which is not discussed in details here. The second part is the process of clustering optimization. The parameters need to be initialized by a standard K-means clustering algorithm. The gradients of parameters (z_i and μ_j) are shown as follows:

$$\frac{\partial L}{\partial z_i} = \sum_j \left(1 + \|z_i - \mu_j\|^2 \right)^{-1}$$
$$\cdot (p_{ij} - q_{ij})(z_i - \mu_j) \quad (8)$$

$$\frac{\partial L}{\partial \mu_j} = -\sum_i \left(1 + \|z_i - \mu_j\|^2 \right)^{-1}$$
$$\cdot (p_{ij} - q_{ij})(z_i - \mu_j) + \alpha \frac{\partial C}{\partial \mu_j} \quad (9)$$

The added constraint has no effect on Eq. (8) because it contains only the parameters of the central point, and in Eq. (9), μ_j should be calculated. In the study, the deep learning framework of Python 3 + Keras is used. Keras can calculate and update gradient directly without manual calculation.

3. EXPERIMENT SETUP

In order to validate the improved DEC method, gear fatigue wear experiment was carried out.

3.1. Experiment Test Rig

The experiment was performed on the test rig shown in Figure 3. It consists of a single-stage gearbox and two 45Kw Siemens PH8 series motors. One motor is used as a driving motor and the other one is used as a load to form a closed-loop experimental device, which is more efficient than the traditional open-loop gear experimental test rig. In the experiment, industrial lubricating oil was added into lubrication based cooling system.

The transmission ratio of the gear system is 1.8:1, the key parameters of the gear are shown in Table 1 and the 3D figure of the gear is shown in Figure 4.
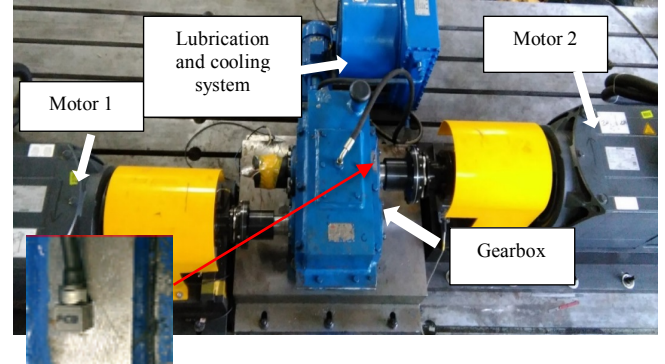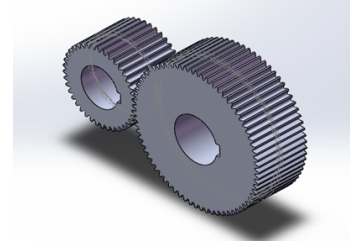

Figure 3. The gearbox Test Gig.

Figure 4. 3D model of the Gear.

Table 1. The main parameters of gears

Gear Parameter	Driving Gear	Driven Gear
Tooth number	40	72
Module	3mm	3mm
Pitch diameter	120mm	120mm
Base circle diameter	112.763mm	202.974mm
Pressure angle	20°	20°
Tooth width	85mm	85mm

3.2. Fatigue Wear Experiment

In the experiment, gearbox was run about 8 hours a day with 44 days of data collection. We collected the gear vibration signals periodically. More frequent data collections were made in late period of the experiment. There are a total of 101 data acquisition made. In order to speed up the wear properly, the total amount of lubricating oil was reduced by 1/3 in the gearbox, and experiment was carried out with an input speed of 1200 rpm and an output torque of 200Nm. The vibration signals were collected with a sampling rate of 20.48 KHz.

For gear health state tracking, data acquisition and analysis are generally carried out in a consecutive manner. Generally, when the fault is small, it is difficult to distinguish the data collected within short intervals. In practical application, intelligent diagnosis is needed to reduce human intervention. Therefore, this paper proposes to use DEC to learn features intelligently to gain insight into online fault trending.

Throughout 44 days of experiment, it was obvious to observe the failure frequencies from the spectrum of the signals in the later stage of the experiment,. A selected 3 spectrums (acquisition No. 5, 38, and 61) of the z direction vibration signals were shown in Figure 5. It can be seen from the Figure 5 that there is no significant change in the spectrums of the first two spectrums, but there are clear fault frequencies in the spectrum of acquisition #61, acquired on the 37th day. In order to validate the proposed method in early wear detection, four acquisitions were selected as test data with acquisition # 18, 27, 30, 35, as shown in Table 2. The data acquisitions were made on 18th, 27th, 29th, 31st day, respectively.

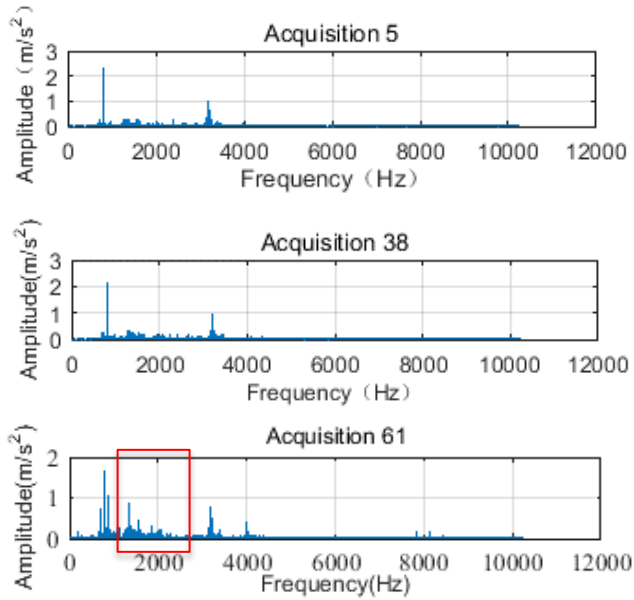Figure 5. Spectrum of 3 acquisitions.

Table 2. Details of acquisitions used in evaluation

Time	Acquisition #	Total duration (unit: h)
1	18	179
2	27	274
3	30	300
4	35	328

To enhance the performance of the unsupervised method, we use the frequency domain signals obtained by Fast Fourier Transform (FFT), to generate input samples for DEC, and each sample has1000 dimensions.

4. RESULTS AND DISCUSSION

4.1. Result Analysis

In this session, we present the results of the improved DEC method as well as the shallow method PCA and the Deep Autoencoder method. The results for 4 selected data are marked as #1 to #4 in chronological order, each data including 670 samples.

Figure 6 shows the result of PCA, with the first two principal components as the selected features. The proportion of the

two principal components represent 27% of the full representations based on eigenvalues. It can be seen from Figure 6 that PC1 has a certain upward trend, but there is little difference between cases #1 and #2, while PC2 is a chaotic feature, which has no distinction for #1 to #4 cases.
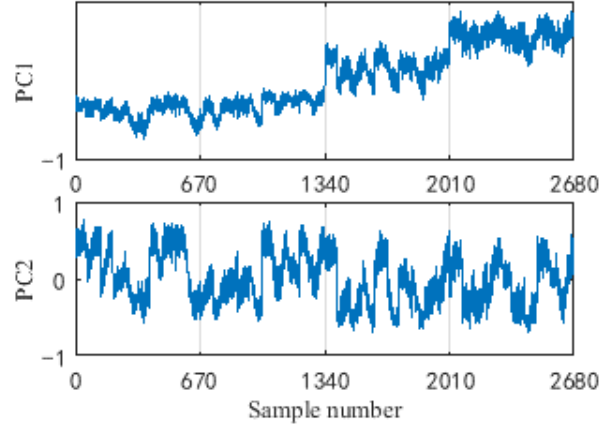

Figure 6. Two dimensions of features obtained from PCA.

In the next test, a stacked autoencoder was designed with structure of 1000-1000-20-2. Figure 7 shows the results of stacked autoencoder. Feature1 has an obvious grading regularity while Feature 2 has low differentiating degree on case #1 to #4.

The result of the original DEC is shown in Figure 8. It can be seen that the features exacted from each cluster is generally more stable than sparse autoencoder (SAE), since it optimized toward clustering performance.

The result of the improved DEC was shown in Figure 9, it can be seen that Feature 1 present a much cleaner feature compared with SAE and original DEC. If we look at the visual separation capability of case 1 and case 2 (corresponding to sample number 1-670, and 671-1340, respectively), the trend obtained from improved DEC is better than that of the original DEC. Samples from 4 different cases can be more clearly separated by looking at feature 1. While for Feature 2, DEC method also obtained an obvious trend for cases #1 to #4. In terms of fault level separation, cases #1 and cases #2 can be differentiated from cases #3 and cases #4, which have more severe wear faults. Therefore, feature 2 could potentially help to achieve more robust fault diagnostic results. Quantitative results will be discussed later.

It should be noted that the features extracted by the embedding algorithm are of no explicit physical meaning, they represent the evolving trend of some hidden features in the signals. While the machine operation condition remains the same, the time evolving trend indicate the health status of the machines. Therefore, although the features extracted from SAE and DEC generally show a downward trend, they still represent an increasing wear of gearbox components.
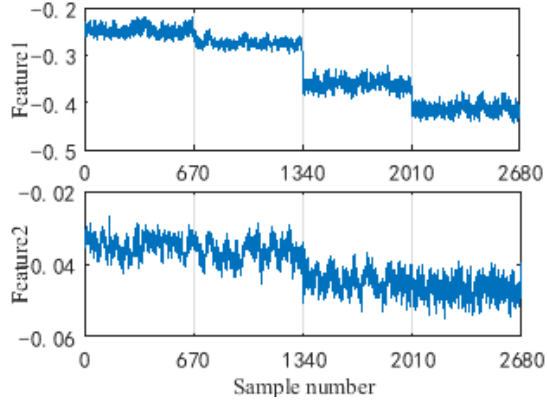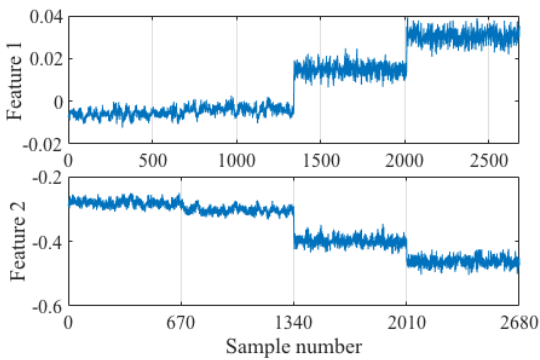
Figure 7. Features obtained from deep SAE.

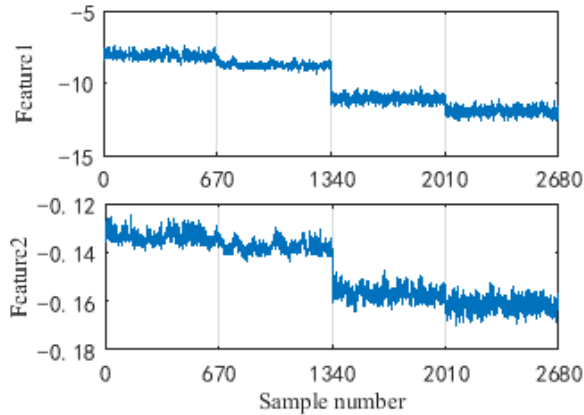Figure 8. Features obtained from original DEC.

Figure 9. Features obtained from improved DEC.

Overall, the improved DEC method has a better performance on feature optimization than other unsupervised deep learning method. To further quantify the performance of DEC, the final results of K-means are shown in Table 3. A standard unsupervised evaluation metric – accuracy is used for the evaluation and comparison. The accuracy is defined as the follows (Ahmed & Khan, 2012):

$$ACC = \max_{m} \frac{\sum_{i=1}^{n} 1\{l_i = m(c_i)\}}{n} \qquad (5)$$

where l_i is the ground truth label, c_i is the cluster assignment produced by the K-means, and m ranges over all possible one-to-one mappings between clusters and labels.

Table 3 gives the average accuracy over 5 runs. It can be seen that the K-means accuracy on the original sample is low, which means the original datasets are hard to be distinguished. The result of PCA is 60.41%, while the accuracy of SAE is about 83%, which proves that the unsupervised deep learning method is superior to shallow network. The accuracy of original DEC is 85.14%, while the accuracy of DEC is over 90%. Also, it is found the performance of the improved DEC is more stable over different runs, while the result of original DEC have slightly more fluctuations with regards to the accuracy.

Table 3. K-means clustering results on 5 methods

Methods	K-means	PCA+ K-means	AE+ K-means	Original DEC	Improved DEC
ACC(%)	50.67	60.41	83.47	85.14	90.34

4.2. Fault Discussion

After the run-to-failure experiment, the gearbox was disassembled, it is found that the fault occurred on the bearing of the driving shaft, which is shown in Figure 10. Scattered ball wear and severe inner case wear can be observed after the experiment.
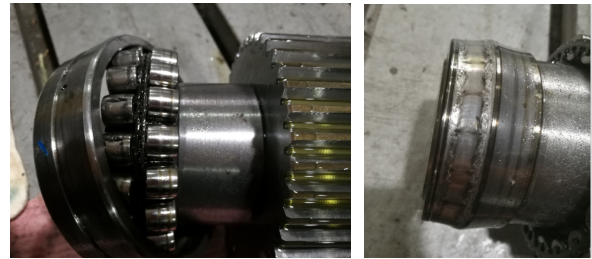

Figure 10. The fault occurred on the bearing

In fact, abnormal noise can be heard from the gearbox starting from around the 40th days. The system has been completely worn out by then. The experiment lasted until the 44th day. However, the fault type and fault information is assumed to be unknown and was not used in this paper.

5. CONCLUSION

It is important to monitor mechanical health status intelligently. In the field of intelligent manufacturing, it is necessary to detect the degree or trend of failure for the safe and effective operation of machinery. In this paper, an unsupervised deep embedded clustering method was used to extract and optimize gear fault features and a regularization of the cluster center points was proposed to make the feature optimization process converge in certain trends. An accelerated run-to-failure experiment on gearbox system was

conducted to verify the effectiveness of the method. The feature extracted from the optimized DEC method can better represent the degradation process compared with those from PCA and deep autoencoder. In practice, the extracted features which indicate the fault trend can be used directly for fault level diagnostics on real-time diagnostic system. It can also be used for health prediction and remaining useful life prediction.

REFERENCES

Ricci, R., & Pennacchi, P.. (2011). Pennacchi, p.: diagnostics of gear faults based on emd and automatic selection of intrinsic mode functions. *Mechanical Systems and Signal Processing*, vol. 25, no. 3, pp. 821-838. doi: 10.1016/j.ymssp.2010.10.002

Žvokelj, Matej, Zupan, S., & Prebil, I.. (2016). Eemd-based multiscale ica method for slewing bearing fault detection and diagnosis. *Journal of Sound & Vibration,* vol. 370, pp. 394-423. doi: 10.1016/j.jsv.2016.01.046

Sun, J., Yan, C., & Wen, J.. (2018). Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning. *IEEE Transactions on Instrumentation & Measurement*, PP (99), 1-11.doi: 10.1109/TIM.2017.2759418

Chen, R., Chen, S., He, M., He, D., & Tang, B. (2017). Rolling bearing fault severity identification using deep sparse auto-encoder network with noise added sample expansion. Proceedings of the Institution of Mechanical Engineers, *Part O: Journal of Risk and Reliability*, 231(6), 666–679. doi:10.1177/1748006x17726452

Lu, C., Wang, Z.-Y., Qin, W.-L., & Ma, J. (2017). Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification. *Signal Processing*, 130, 377–388. doi:10.1016/j.sigpro.2016.07.028

Shao, H., Jiang, H., Lin, Y., & Li, X. (2018). A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders. *Mechanical Systems and Signal Processing*, 102, 278–297. doi: 10.1016/j.ymssp.2017.09.026

Yu, H., Wang, K., & Li, Y. (2018). Multi-scale Representations Fusion with Joint Multiple Reconstructions Autoencoder for Intelligent Fault Diagnosis. *IEEE Signal Processing Letters*, 1–1. doi:10.1109/lsp.2018.2878356

Jiang, G., He, H., Yan, J., & Xie, P. (2018). Multiscale Convolutional Neural Networks for Fault Diagnosis of Wind Turbine Gearbox. *IEEE Transactions on Industrial Electronics*, vol: 66, no: 4. doi:10.1109/tie.2018.2844805 Shao, H., Jiang, H., Zhang, H., & Liang, T. (2018). Electric Locomotive Bearing Fault Diagnosis Using a Novel Convolutional Deep Belief Network. *IEEE Transactions on Industrial Electronics*, 65(3), 2727–2736. doi:10.1109/tie.2017.2745473

Wen, L., Li, X., Gao, L., & Zhang, Y. (2018). A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method. *IEEE Transactions on Industrial Electronics*, 65(7), 5990–5998. doi:10.1109/tie.2017.2774777

Pan, J., Zi, Y., Chen, J., Zhou, Z., & Wang, B. (2018). LiftingNet: A Novel Deep Learning Network With Layerwise Feature Learning From Noisy Mechanical Data for Fault Classification. *IEEE Transactions on Industrial Electronics*, 65(6), 4973–4982. doi:10.1109/tie.2017.2767540

Zhao, R., Wang, D., Yan, R., Mao, K., Shen, F., & Wang, J. (2018). Machine Health Monitoring Using Local Feature-Based Gated Recurrent Unit Networks. *IEEE Transactions on Industrial Electronics*, 65(2), 1539–1548. doi:10.1109/tie.2017.2733438

Li, X., Duan, F., Loukopoulos, P., Bennett, I., & Mba, D. (2018). Canonical variable analysis and long short-term memory for fault diagnosis and performance estimation of a centrifugal compressor. *Control Engineering Practice*, 72, 177–191. doi:10.1016/j.conengprac.2017.12.006

Hinchi, A. Z., & Tkiouat, M. (2018). Rolling element bearing remaining useful life estimation based on a convolutional long-short-term memory network. *Procedia Computer Science*, 127, 123–132. doi:10.1016/j.procs.2018.01.106

Sun, C., Ma, M., Zhao, Z., Tian, S., Yan, R., & Chen, X. (2019). Deep transfer learning based on sparse auto-encoder for remaining useful life prediction of tool in manufacturing. *IEEE Transactions on Industrial Informatics*, PP(99), 1-1.doi: 10.1109/TII.2018.2881543

Qu, Y., Zhang, Y., He, M., He, D., Jiao, C., & Zhou, Z. (2019). Gear pitting fault diagnosis using disentangled features from unsupervised deep learning. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability.* https://doi.org/10.1177/1748006X18822447

Xie J, Girshick R, Farhadi. (2016). Unsupervised Deep Embedding for Clustering Analysis. *ICML*, 2016.

Hinton, G. E. (2006). Reducing the Dimensionality of Data with Neural Networks. *Science*, 313(5786), 504–507. doi:10.1126/science.1127647

Ahmed, N., Jalil, A., & Khan, A. (2012). Performance improvement in image clustering using local discriminant model and global integration. *Proceedings of 2012 9th International Bhurban Conference on Applied Sciences & Technology (IBCAST).* doi:10.1109/ibcast.2012.6177530