Adversarial flow-based model for unsupervised fault diagnosis of rolling element bearings

Jun Dai, Jun Wang*, Linquan Yao, Juanjuan Shi, Lei Wang and Guolei Wang

1 School of Rail Transportation, Soochow University, Suzhou, Jiangsu, China
2 Xiangyang BOYA Precision Industrial Equipments Co., Ltd., Xiangyang, Hubei, China

*Corresponding author: junking@mail.ustc.edu.cn

Abstract. Nowadays, numerous supervised deep learning models have been applied to bearing fault diagnosis. However, labelling the health states of the bearing vibration data is a time-consuming work and dependent on expert experience. In order to tackle this problem, a novel unsupervised bearing fault diagnosis method named adversarial flow-based model is explored in this paper. Flow-based model is a type of generative models that is proved to be better than other types in many aspects. This paper introduces the flow-based model into the field of machinery fault diagnosis, and designs an appropriate model architecture so as to train the model in unsupervised and adversarial ways. The proposed model contains an autoencoder (AE), a flow-based model, and a classifier. Firstly, the AE maps the vibration data from signal space to latent vector space. Then, the flow-based model aligns the distributions of the latent vectors of different bearing states with specific prior distributions. Finally, the classifier tries to discriminate the aligned latent vectors from the vectors sampled from the prior distributions. With the help of distinguishable prior distributions and the adversarial training mechanism between the classifier and the flow-based model together with the AE, the bearing data with the same health states are clustered into the same areas. The good clustering performance of the adversarial flow-based model is verified by a dataset with 10 health states from a bearing test rig.

Keywords: Fault diagnosis, flow-based model, adversarial training, unsupervised learning, rolling element bearings.

1. Introduction
As a critical rotating part of mechanical systems, rolling element bearings are widely applied in transportation systems, industrial systems, and electrical systems. Owing to the harsh operating environments, such as high speed and heavy dynamic load, bearings are prone to failure that are difficult to be recognized [1]. Bearing fault may affect the normal operation of mechanical systems and lead to serious accidents. To ensure the healthy operation of rolling bearings, massive data are collected, which promoted the evolution of data-driven based fault diagnosis techniques [2].
Nowadays, many supervised learning models, such as deep neural networks (DNN) [3], convolutional neural networks (CNN) [4], deep belief networks (DBN) [5], have been used to bearing fault diagnosis. Latent features in the bearing vibration signals can be extracted adaptively by these models, and used for recognition of the bearing health states. However, these models are all supervised model, a large amount of historical data and corresponding labels are required for model training. In practical diagnosis task, labelling massive bearing vibration signals is time consuming and requires expert experience. Hence, unsupervised learning models, such as autoencoder (AE) [6], K-nearest neighbor (KNN) [7], and generative model [8], have attracted much attention to tackle the absence of labels in bearing fault diagnosis.

Generative adversarial network (GAN) and variational autoencoder (VAE) are typical generative models which are designed to learn the data distributions in an unsupervised way, and have been gradually employed in fault diagnosis. Chen et al. [9] constructed a sliding-window convolutional variational autoencoder (SWCVAE) to realize real-time anomaly detection. However, the AE is only an approximate method for estimation of data distributions. Dai et al. [10] proposed a novel GAN to capture the distribution of normal samples, which adopts adversarial training mechanism between a generator constructed via the AE and a discriminator for identification of data authenticity. A health indicator was established for monitoring the real-time health states of a machine. The results have demonstrated impressive performance on unsupervised fault diagnosis. However, training the GAN model is a challenging task owing to the training instability, such as mode collapse, gradient vanishing, and gradient explosion.

Flow-based model is a new type of generative models, which is popularized by Dinh et al. [11] for accurately learn of data distributions through a sequence of invertible transformations. The flow-based model has been utilized in many fields, including image generation [12], voice conversion [13], and speech synthesis [14]. Yang et al. [15] introduced the flow-based model to the AE to construct a sequence-to-sequence learning model to realize remaining useful life prediction of a mechanical system. State-of-the-art results have been achieved by the flow-based models in these literatures. However, the flow-based models have not been used to recognize the health states of mechanical systems until now.

In this paper, an adversarial flow-based model is constructed for unsupervised fault diagnosis, by embedding the flow-based model into the AE and classification networks. The proposed model aligns the latent vectors obtained by the AE from the bearing vibration signals of each health states with a specific prior distribution through the flow-based model and a classifier, which is trained in unsupervised and adversarial ways. The AE and flow-based model try to trick the classifier into obtaining a certain category assignment for the aligned latent vectors, while the classifier tries to discriminate the aligned latent vectors from the prior distributions and gives vague category judgments. After fully trained, the proposed model will reach a balance, which means the distribution characteristics of the bearing vibration signals have been fully captured by the AE and flow-based model, and the classifier cannot discriminate the aligned latent vectors from the prior distributions. As a result, the aligned latent vectors of each health state strictly follow one specific prior distribution, which exhibits great clustering and unsupervised classification performance. A bearing vibration dataset with 10 health states is analyzed in this paper. The experiment verified the superiority of the proposed adversarial flow-based model over traditional AE model.

2. Flow-based model
Flow-based model is a kind of generative networks, which aims to accurately capture the probability distribution of input samples. Given an observed data $x \in X$ that follow the distribution $p_X$, and a prior probability distribution $p_Z$ on a variable $z \in Z$, the probability density follows the formula below:

$$\int p_X(x)dx = \int p_Z(z)dz = 1$$

(1)
The flow-based model designs a bijection function \( f : X \rightarrow Z \) with \( g = f^{-1} \). The function \( f(\cdot) \) maps the \( x \) from original latent distribution \( p_X \) to the prior distribution \( p_Z \). The change of variable formula defines a model distribution on \( X \) by:

\[
p_x(x) dx = p_x(f(x)) df(x)
\]

\[
\Rightarrow p_x(x) = p_x(f(x)) \frac{df(x)}{dx}.
\]

Generalization to higher dimensions, the probability density can be expressed as:

\[
p_x(x) = p_Z(f(x)) \left| \frac{\partial f(x)}{\partial x} \right|
\]

\[
\log(p_x(x)) = \log(p_Z(f(x))) + \log\left( \left| \frac{\partial f(x)}{\partial x} \right| \right)
\]

where \( \frac{\partial f(x)}{\partial x} \) is the Jacobian matrix of \( f(\cdot) \) at \( x \). To simplify the calculation of the Jacobi matrix, a flow network named real-valued non-volume preserving (Real NVP) [16] is applied, whose architecture is shown as figure 1.

![Real NVP Architecture](image)

**Figure 1.** Architecture of the Real NVP.

The architecture of the Real NVP is stacked by several flow blocks. In each block, the input data are divided into two parts, \( x_{k1} \) and \( x_{k2} \). The first part \( x_{k1} \) is transmitted directly to the output, i.e., \( x'_{k1} = x_{k1} \). For the second part, \( x'_{k2} \) is calculated according to the following formula:

\[
x'_{k2} = w \cdot x_{k2} + b
\]

where \( w, b \) represent the weight and bias, respectively, which are output by the fully connected networks. The Jacobi matrix of this structure is a diagonal matrix and can be formulated as the following equation:

\[
\text{Jacobi matrix} = \begin{bmatrix} E & O \\ M & W \end{bmatrix}
\]

where \( E, O, W \) represent the identity matrix, null matrix, and weight matrix, respectively. \( M \) is a matrix which will not influence the determinant. The determinant of the Jacobi matrix is calculated by successive multiplication of the elements at the main diagonal.

### 3. Adversarial flow-based model

The flow-based model described above is trained through an unsupervised approach. That is to say, the model is trained without using label information. In order to enhance the clustering performance in unsupervised rolling element bearing fault diagnosis, an adversarial flow-model is proposed, whose architecture is illustrated in figure 2. The proposed model contains an AE, a flow-based model, and a classifier. In the forward propagation process, the signal \( x \) after simple preprocessing using Fourier transform is first input to the AE, by which the latent feature \( z \) of the signal in latent space are learned by reconstructing the signal \( x \) as \( x' \) in the frequency domain. Then the latent vector \( z \) is input into a flow-based model, by which the vector will be mapped to another space after aligning to a specific prior distribution. Thereafter, the vector \( k \) sampled from one of the \( K \) prior distributions and the
aligned latent vector $z'$ are input into a classifier, which is trained in an adversarial way and predicts the labels for both of the vectors.

During the training process, the vectors sampled from the prior distributions are treated as real samples. The classifier will give these vectors certain classification assignments. On the contrary, the aligned latent vectors are considered as fake samples and vague judgments will be given by the classifier. However, the AE and flow-based model are trained to align the latent vector to follow the prior distributions as much as possible, so that the classifier could not tell the real and fake samples apart. This is an adversarial training process. To further enhance the adversarial performance in classification, both the encoder and the classifier are required to assign classification equally to the samples (assume the amount of samples per class is the same for all categories). After fully trained, the original signal with the same healthy state will be clustered by the proposed model in both the latent space and the space of the designed prior distributions.

**Figure 2.** Architecture of the proposed adversarial flow-based model.

In this paper, three parts of functions, including reconstruction function, flow-based model function and adversarial function are applied to train the proposed model. The reconstruction function measures the dissimilarity between the original signal and the reconstructed signal, as formulated by:

$$F_{rec} = \min_{En,De} \|x - x'\|_2$$  \hspace{1cm} (7)

The flow-based model function measures the probability density of the latent vectors, as defined by:

$$F_{flow} = \max_{\text{flow}} \left\{ \log p_{\text{latent}}(z) \right\} = \max_{\text{flow}} \left[ \log p_{\text{prior}}(f(z)) + \log \det \left( \frac{\partial f(z)}{\partial z^T} \right) \right]$$  \hspace{1cm} (8)

The adversarial function is used to optimize the encoder and classifier respectively. For the classifier, the function aims to discriminate the source of input vector and assign classification equally. The objective function for the classifier is shown as:

$$F_{\text{adversarial-Classifier}} = \max_C \left\{ H_{\text{prior}} \left[ p(Y|C) \right] - E_{x\sim\text{prior}} \left[ H \left[ p(Y|k, C) \right] \right] + E_{x\sim\text{data}} \left[ H \left[ p(Y|f(En(x)), C) \right] \right] \right\}$$  \hspace{1cm} (9)

where $H[-]$ represents the Shannon entropy. $C$ represents the classifier. $En$ represents the encoder.

For the encoder, the adversarial function is optimized to obtain certain and equal category assignment from the classifier. The objective function is shown as:

$$F_{\text{adversarial-Encoder}} = \min_{En} \left\{ -H_{f} \left[ p(Y|C) \right] + E_{x\sim\text{data}} \left[ H \left[ p(Y|f(En(x)), C) \right] \right] \right\}$$  \hspace{1cm} (10)

4. Experiment validation
In this section, a bearing dataset containing 10 types of health states is analyzed by proposed method. The experimental platform is shown in figure 3. The test bearing (Type: NSK 6205) was installed at the right end of a shaft driven by an AC motor. The rotating speed was kept at 900 RPM and the radial load was kept at 1 kN. The sampling frequency was 10 kHz. The complete dataset contains a normal state (N) and three fault states, including inner race fault (IR), outer race fault (OR), and ball element fault (B). According to the width of the fault slit, each fault state is divided into three sub- states, corresponding to the width of 0.2 mm, 0.3 mm, and 0.4 mm, respectively. For convenience, one fault state is denoted by the fault location combining the slit width. For example, IR02 represents the fault state with a 0.2 mm width of slit located in the bearing inner-race way. Each fault state contains 500 samples and each sample has 2048 points in the time domain.

![Experimental platform for data acquisition of rolling element bearings with 10 types of health states.](image)

In the experiment, the designed prior distributions are \( K (=10) \) the two dimensional Gaussian distributions, which corresponds to 10 types of the bearing health categories. Each sub-distribution is generated from an initial Gaussian distribution \( X \sim N \left( \begin{bmatrix} a \\ 0 \end{bmatrix}, \begin{bmatrix} b_1 & 0 \\ 0 & b_2 \end{bmatrix} \right) \) by rotating an angle \( \theta \) around the origin and can be formulated as:

\[
X \sim N \left( \begin{bmatrix} a \cos \theta \\ a \sin \theta \end{bmatrix}, \begin{bmatrix} b_1 & 0 \\ 0 & b_2 \end{bmatrix} \right) \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \Theta = \frac{2\pi i}{K}, \{i|0 \leq i \leq K-1, i \in N\}
\]

where \( \begin{bmatrix} a \cos \theta \\ a \sin \theta \end{bmatrix} \) is the mean value matrix of each sub-Gaussian distribution, \( \begin{bmatrix} b_1 & 0 \\ 0 & b_2 \end{bmatrix} \) is the covariance matrix, where \( \begin{bmatrix} b_1 & 0 \\ 0 & b_2 \end{bmatrix} \) is the covariance matrix of initial Gaussian distribution and \( \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \) is the rotation matrix.

In this paper, \( a = 5, b_1 = 0.5, \) and \( b_2 = 0.02 \). The 10 prior distributions are displayed in the lower left quarter of figure 2. The time domain samples are transformed into frequency domain by fast Fourier transform (FFT) at first. Then, the min–max scaling is performed on the frequency domain samples. The samples are all scaled within the range of \([0,1]\). The model is optimized by Adam optimization algorithm (Learning rate = 0.0005). The structure details of the proposed adversarial flow-based model are summarized in table 1.
After fully trained by the bearing dataset, the aligned latent vectors by the flow-based model in the proposed method are showed in figure 4(a). In the figure, the samples belonging to the same bearing health state are assembled at the same area in the two dimensional space, which strictly follows a specific prior distribution shown in the lower left quarter of figure 2. Therefore, the 10 types of bearing health states are accurately classified to different categories in an unsupervised way.

Table 1. The structure details of the proposed adversarial flow-based model

| Modules       | Parameter name          | Parameters | Activation function |
|---------------|-------------------------|------------|---------------------|
| Encoder       | Encoder-layer-1         | 1024 to 512| Tanh                |
|               | Encoder-layer-2         | 512 to 128 | Tanh                |
|               | Encoder-layer-3         | 128 to 2   | /                   |
|               | Decoder-layer-1         | 2 to 128   | Tanh                |
|               | Decoder-layer-2         | 128 to 512 | Tanh                |
|               | Decoder-layer-3         | 512 to 1024| Tanh                |
| Flow-based    | Number of flow layer    | 4          | /                   |
| model         | Dimension of hidden layer| 8          | /                   |
| Classifier    | Classifier-layer-1      | 2 to 100   | Tanh                |
|               | Classifier-layer-2      | 100 to 100 | Tanh                |
|               | Classifier-layer-3      | 100 to $K$ | Softmax             |

Figure 4. Visualization of (a) the aligned latent vectors in the proposed method and (b) the latent vectors of the AE.

As a comparison, an AE model with the same structure as that of the proposed model is constructed to analyse the same dataset. The latent vectors obtained by simple AE model are displayed in figure 4(b). Some bearing health states cannot be discriminated from another by the latent vectors of the AE, especially the states of N, B02, and B03. And the clustering performance of every health state is not satisfactory, because the assembled area has much larger covariance value than proposed method. The comparison results prove the outstanding clustering ability of the adversarial flow-based model in unsupervised rolling element bearings fault diagnosis.

5. Conclusion
In this paper, a novel unsupervised bearing fault diagnosis model called adversarial flow-based model is presented. The contribution is that the flow-based model is embedded in the AE and classification networks for mapping of the latent vectors on prior distributions, which is implemented in unsupervised and adversarial ways. Two dimensional Gaussian distributions with different mean and
covariance values were designed as the prior distributions, which benefits the clustering performance of the proposed model. The signals with the same health states can be clustered in the same areas, whereas those with different health states can be separated distinctly. A bearing vibration dataset containing 10 types of health states was used to validate the clustering effectiveness of the proposed adversarial flow-based model. The results proved the superiority of the proposed model in clustering and unsupervised classification of bearing health states. More dataset will be applied to validate the proposed model in the future work.

Acknowledgments
This work was financially supported in part by the National Natural Science Foundation of China under Grant 51805342, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20180842, in part by the China Postdoctoral Science Foundation under Grant 2021M692354, and in part by the Postgraduate Research & Practice Innovation Program of Jiangsu Province under Grant KYCX20_2661.

References
[1] Liu R, Yang B, Zio E and Chen X 2018 Artificial intelligence for fault diagnosis of rotating machinery: A review Mech. Syst. Signal Process. 108 33–47
[2] Lei Y, Yang B, Jiang X, Jia F, Li N and Nandi A K 2020 Applications of machine learning to machine fault diagnosis: A review and roadmap Mech. Syst. Signal Process. 138 106587
[3] Teng W, Cheng H, Ding X, Liu Y, Ma Z and Mu H 2018 DNN-based approach for fault detection in a direct drive wind turbine IET Renew. Power Gener. 12 1164–71
[4] Wen L, Li X, Gao L and Zhang Y 2018 A new convolutional neural network-based data-driven fault diagnosis method IEEE Trans. Ind. Electron. 65 5990–8
[5] Yu J and Liu G 2020 Knowledge extraction and insertion to deep belief network for gearbox fault diagnosis Knowledge-Based Syst. 197 105883
[6] Shao H, Jiang H, Zhao H and Wang F 2017 A novel deep autoencoder feature learning method for rotating machinery fault diagnosis Mech. Syst. Signal Process. 95 187–204
[7] Lu J, Qian W, Li S and Cui R 2021 Enhanced K-nearest neighbor for intelligent fault diagnosis of rotating machinery Appl. Sci. 11 919
[8] Liu H, Zhou J, Xu Y, Zheng Y, Peng X and Jiang W 2018 Unsupervised fault diagnosis of rolling bearings using a deep neural network based on generative adversarial networks Neurocomputing 315 412–24
[9] Chen T, Liu X, Xia B, Wang W and Lai Y 2020 Unsupervised anomaly detection of industrial robots using sliding-window convolutional variational autoencoder IEEE Access 8 47072–81
[10] Dai J, Wang J, Huang W, Shi J and Zhu Z 2020 Machinery health monitoring based on unsupervised feature learning via generative adversarial networks IEEE/ASME Trans. Mechatronics 25 2252–63
[11] Dinh L, Krueger D and Bengio Y 2014 NICE: Non-linear independent components estimation arXiv Prepr. arXiv1410.8516
[12] Lugmayr A, Danelljan M, Van Gool L and Timofte R 2020 SRFlow: Learning the super-resolution space with normalizing flow European Conference on Computer Vision pp 715–32
[13] Serrà J, Pascual S and Segura C 2019 Blow: A single-scale hyperconditioned flow for non-parallel raw-audio voice conversion arXiv Prepr. arXiv1906.00794
[14] Kim H, Lee H, Kang W H, Cheon S J, Choi B J and Kim N S 2020 WaveNODE: A continuous normalizing flow for speech synthesis arXiv Prepr. arXiv2006.04598
[15] Yang H, Ding K, Qiu R C and Mi T 2020 Remaining useful life prediction based on normalizing flow embedded sequence-to-sequence learning IEEE Trans. Reliab. Preprint 10.1109/TR.2020.3010970
[16] Dinh L, Sohl-Dickstein J and Bengio S 2016 Density estimation using real NVP arXiv Prepr. arXiv1605.08803