K-Means clustering GAN based Fault Diagnosis Approach for Imbalanced Dataset

Huifang Li[*1], Rui Fan[*2], Qisong Shi[*3], and Zijian Du[*4]

[*1] School of Automation, Beijing Institute of Technology, Beijing, China
E-mail: huifang@bit.edu.cn
[*2] School of Automation, Beijing Institute of Technology, Beijing, China
E-mail: 1148215410@qq.com
[*3] School of Automation, Beijing Institute of Technology, Beijing, China
E-mail: 992974590@qq.com
[*4] School of Automation, Beijing Institute of Technology, Beijing, China
E-mail: 616590807@qq.com

Abstract: Due to the different occurring frequencies of various faults, class distribution of fault data is often imbalanced. However, most existing machine learning based diagnosis methods, which didn't take this imbalance into consideration, tend to be biased toward the majority classes and result in poor accuracy for minority ones. To solve this problem, we propose a K-Means clustering GAN (KM-GAN) based fault diagnosis approach which can reduce the imbalance of fault data and improve the diagnostic accuracy for minority classes. First, we design a new oversampling method based on K-Means clustering algorithm and GAN, to generate diverse minority-class samples which have similar distribution with the original minority data. The K-Means clustering algorithm is adopted to divide minority-class samples into K-clusters, while GAN is applied to learn the distribution of the resulting clusters and then generate a certain number of minority samples as a supplement to the original dataset. Second, we construct a Deep Neural Network (DNN) and Deep Belief Network (DBN) based heterogeneous ensemble model as a fault classifier to improve generalization, while DNN and DBN models are trained separately on the resulting dataset, and then the outputs from DNN and DBN are averaged as the final diagnostic results. A series of comparative experiments are conducted to testify the effectiveness of our proposed method, and the experimental results show that our method can improve the diagnostic accuracy of minority samples.

Keywords: Class Imbalance, Fault Diagnosis, Machine Learning, Deep Learning

1. INTRODUCTION

With the rapid development of science and technology, modern industrial equipment has become more and more sophisticated, leading to the increase of equipment maintenance cost and diagnostic difficulty [1]. Machine learning- and deep learning-based fault diagnosis methods have been proved as a promising way to achieve accurate fault diagnosis. However, most of the existing machine learning based diagnosis methods seldom consider or even ignore the imbalance of the actual fault data, so there are great limitations in solving the class-imbalance fault diagnosis problems. For example, it is easy to identify minority-class faults as majority-class ones, and the diagnosis accuracy is significantly reduced [2].

To this end, the class-imbalance classification problem has attracted much attention from scholars in recent years [3]. Many research efforts have been conducted to solve this problem [4]. Typical method is resampling, which tries to reduce the imbalance of dataset from data perspective. It can be mainly divided into the following three types: under-sampling, over-sampling, and the combination of both under- and over-sampling. The under-sampling methods reduce the number of majority-class samples to improve the classification accuracy of minority classes. Random under-sampling (RUS) is the simplest under-sampling method, which can balance datasets through randomly deleting a part of the majority-class samples. However, randomly discarding may remove some potential useful information from majority classes and result in performance degradation of the classifier. To solve this, Mani [5] used a K-nearest neighbors (KNN) classifier to remove samples based on the distance between majority class and minority class. Barandela et al. [6] removed majority-class samples from class boundaries based on an KNN rule that removes misclassified samples from the training set. In contrast to under-sampling, the over-sampling focuses on increasing the number of minority-class samples to improve the classification performance of minority class. The simplest over-sampling method is to increase the number of minority-class samples by random replication or simple rotation [7]. However, this approach will increase the risk of overfitting due to repeated replication. In order to solve this problem, Chawla et al. [8] proposed a Synthetic Minority Oversampling Technique (SMOTE) which increases the number of minority-class samples by randomly inserting synthetic samples. It works well but may cause over-generalization and the introduction of noises. Thus, Zhu et al. [9] introduced weights for different samples when selecting the nearest neighbor samples, which can effectively deal with over-generalization.

Though all the above-mentioned methods have achieved certain performance in classification from imbalanced data, they still need to be further researched for application in the real world, e.g. the lack of adaptability. These methods are difficult to learn the distribution characteristics of data samples automatically. With the development of deep learning techniques, new data generation methods have been presented. Generative Adversarial Network (GAN), proposed by Goodfellow et al. [10], can learn the data distribution characteristics of the original samples and then generate new synthetic samples with similar distribution. Douzas et al. [11] used Conditional GAN (CGAN) to generate minority-class samples and then effectively improve the imbalance extent of data set. Wang et al. [12] combined GAN with Stacked Denoising Auto Encoder (SDAE) to realize fault diagnosis for gear box. GAN was used to expand the number of samples while SDAE was taken as a discriminator of GAN to extract deep features adaptively and diagnose fault types as well. Liu et al. [13] trained an Auto Encoder through an adversarial training process and imposed a prior distribution on the latent coding space for fault diagnosis. Han et al. [14] adopted adversarial learning into Convolution Neural Network (CNN) as a regularization, which improves the robustness of feature representations and enhances its model generalization.

However, the diversity of samples generated is poor due to the mode collapse of GAN [15], which may result in within-class imbalance and poor model generalization. Therefore, we propose a K-Means clustering GAN based fault diagnosis approach. First, we design a new oversampling method based on K-Means clustering algorithm [16] and GAN to generate diverse minority-class samples that have similar distribution with original data. Second, we construct a Deep Neural Network (DNN) and Deep Belief Network (DBN) based heterogeneous ensemble model as a fault classifier to improve the generalization of classifier to different samples.

The remainder of the paper is organized as follows. We introduce some basic concepts of GAN and K-Means as well as DNN and DBN in Section 2. Section 3 is our proposed method. Section 4 is the experiments conducted and its result analysis. We conclude the paper in Section 5.

2. Basic Principles

In this section, we will briefly introduce some relevant principles of clustering algorithms, generative adversarial networks, deep neural networks and so on, to facilitate the following introduction of our proposed method.

2.1. Clustering Algorithms

Clustering is a task to divide data into groups, which are called clusters. The goal is to partition the data, so that the data points in a cluster are extremely similar while the data points in different clusters are extremely different. Similar to classification algorithms, clustering algorithms assign (or predict) a number for each data point to indicate which cluster the point belongs to.

K-means clustering is one of the simplest and most commonly used clustering algorithms, which aims to find some cluster centers which can represent different areas of the data. This algorithm alternately performs the following two steps:

1) Assign each data point to the nearest cluster center;

2) Set each cluster center to the average value of all the assigned data points.

If the cluster allocation is no longer changed, the algorithm will terminate.

2.2. Generative Adversarial Networks

Generative Adversarial Networks (GAN) consists of a generator and a discriminator. The generator is mainly used to generate samples with the same distribution as original samples through estimating the joint probability distribution, while discriminator is applied to judge whether the input is the real data or the one produced by the generator. After the combination of generator and discriminator, the generator will produce more realistic samples during many training iterations, and the discriminator will obtain a more accurate ability to identify the true and false data. The final identification result of discriminator for the generator data is that the correct and error rate are both 50%.

Specially, GAN could effectively estimate the distribution of high-dimensional data and can be applied in mathematics and engineering, even used as a technical means to represent the state of model in reinforcement learning.

2.3. Deep Neural Network

As the deep learning model can fit complex nonlinear relationships, deep neural network (DNN) is widely used in classification, recognition, and other fields. DNN is derived from the perceptron that was proposed by Rosenblatt [17]. Perceptron is a neural network composed of two layers of neurons, and its most basic structure is with several inputs and a single output.

The forward calculation between the input and output of perceptron is illustrated as follows:

$$y = \sum_{i=1}^{n} w_i x_i + b \tag{1}$$

where x_i represent the input variable, and each input variable corresponds to a weight w_i. b denotes the offset. The number of input variables is n, and y is the intermediate result of the above calculation. Then the intermediate result y is fed into an activation function for nonlinear transformation, and the final output is obtained.

Activation function imitates the threshold activation characteristics of human brain neurons, and also introduces nonlinear features into neurons to realize the transformation from simple linear space to high-dimension nonlinear space. Generally, the commonly used activation function is Sigmoid, and its mathematical form is shown as follows:

$$f(x) = \frac{1}{1+e^{x}} \tag{2}$$

Perceptron with single layer is relatively simple structured, which only deals with simple linear classification tasks and cannot solve the XOR problem. While DNN is a generalized extended model on the basis of perceptron, by adding hidden layers, neurons and expanding activation functions, the structure of DNN is shown in Fig. 1. The neural network layer in DNN can be divided into three types: input layer, hidden layer, and output layer. Generally, the first layer is the input layer, the middle layers are the hidden layers, and the last layer is the output layer.

Compared with shallow networks, the advantage of DNN is the multiple hidden layers, which results in its stronger ability to deal with complex classification tasks.

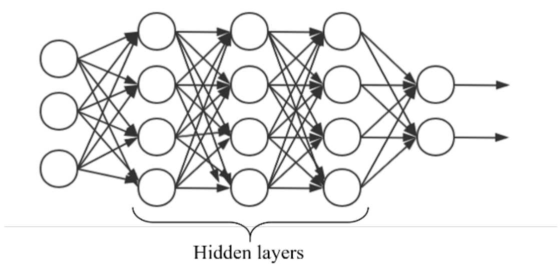

Fig. 1 The structure of DNN

Since the forward calculation cannot learn the optimal parameters (weight and bias) based on learning samples, Rumelhart et al. [18] proposed the error back propagation algorithm (BP). This algorithm adjusts the parameters of each layer from the error between the predicted and actual value of the output layer. When BP algorithm is used to optimize parameters, the loss function is generally selected as cross entropy loss for classification tasks.

2.4. Deep Belief Network

The basic component of Deep Belief Network (DBN) is Restricted Boltzmann Machine (RBM), which consists of a visible layer and a hidden layer, and there is no connection between the visible and hidden layer.

The visible layer is used to obtain input data, while the hidden layer extracts features from the input data through training procedure. For RBM, the training process aims to figure out the optimal weights between the visible and hidden layer. The most commonly used algorithm for training RBM is the contrast divergence algorithm, which exploits Gibbs sampling to update the weights during the process of gradient descent, which is similar to the BP algorithm in training feed forward neural networks.

DBN is constructed by stacking several RBMs together, and these RBMs are trained from bottom to top. The training steps are illustrated as follows:

1) The bottom-up unsupervised learning method is used to train the parameters of neurons layer by layer. To begin with training, the training data is input into the bottom RBM. Then the output of the underlying hidden layer is taken as the input of the upper RBM. Through this layer by layer training, the optimal connection weights of each neuron are found. This step belongs to unsupervised training process, so it is called feature-learning process;

2) After getting the parameters of neurons in each layer, the selected classifier is added to the top layer of network in the top-down supervised learning process, and the structural parameters of network are fine-tuned according to the labeled data. This step belongs to supervised training process, so it is called fine-tuning process.

3. PROPOSED METHOD

To address the problem of imbalanced fault diagnosis, this paper presents a new oversampling method named KM-GAN which is based on K-Means clustering algorithm and GAN. Firstly, the K-Means clustering algorithm is adopted to divide minority-class samples into K clusters so that K modes of minority-class samples are obtained. Then, with the help of GAN, the method learns the distribution of K clusters respectively and generates K modes of similar samples to alleviate the imbalance between minority-class and majority-class. Finally, an DNN and DBN based heterogeneous ensemble classifier is trained by the resulting balanced dataset and applied for the diagnosis. The flowchart of the proposed method is shown in Fig. 2.

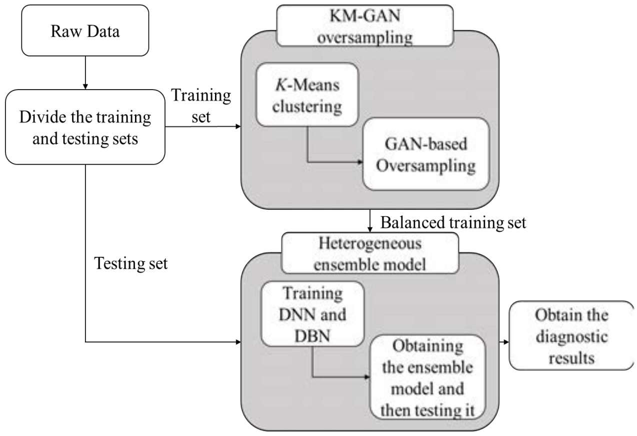

Fig. 2 Flow chart of the proposed method

3.1. KM-GAN Oversampling Method

Oversampling has been rigorously studied to alleviate the problem of class imbalance as the most popular technique. GAN based over-sampling technique is widely applied to balance the dataset, thanks to its ability to learn the distribution of data, and adaptively generate samples obeying the distribution. However, the diversity of samples generated by GAN is lower than the original data, resulting from the mode collapse problem of GAN. Considering that K-Means clustering algorithm could capture different modes of samples and divide them into different clusters without supervision while minimize the within-cluster sum of squares, which could make up the lack of diversity for synthetic samples caused by the mode collapse to some extent. We propose a novel over-sampling method, in which GAN is applied after K-Means clustering algorithm as shown in Fig. 3.

Firstly, K-Means clustering algorithm is adopted to minority-class fault data from original training set to divide the samples into K modes of clusters. And the specific steps are illustrated as follows.

Step 1: Input minority class sample set A;

Step 2: Randomly select K samples from set A as the initial clustering centers;

Step 3: The distances of each sample in set A to different clustering centers are calculated, and each sample in set A is added to the cluster where its nearest clustering center is located;

Step 4: According to the newly generated clusters, the clustering centers of the clusters are updated;

Step 5: Repeat *Step* 3 to 4 until each cluster no longer changes;

Step 6: Obtain K clusters Cluster_1, Cluster_2, ..., Cluster_K.

Secondly, the resulting clusters are sent to K discriminators respectively, to compare with samples randomly generated by the K generators and train the discriminators and generators alternately.

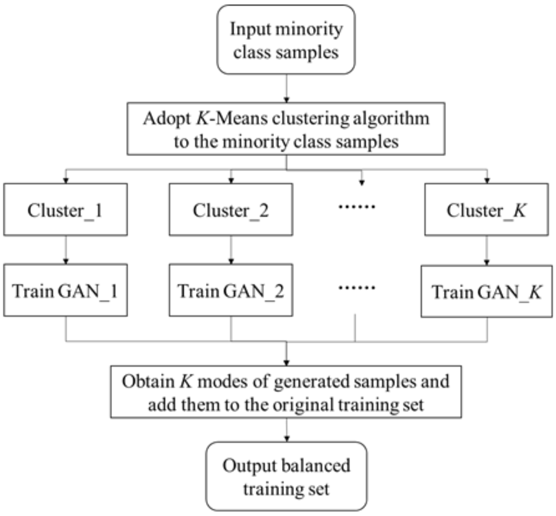

Fig. 3 Flow chart of the KM-GAN

The Sketch map of GAN is shown in Fig. 4, where X_t represents the real samples, that is, the minority-class samples in the original training set. Z represents a set of random noises which obey uniform distribution. Specially, Z is sent to generator G as the input of G and the output $G(Z)$ is taken as generated samples. And then X_t and $G(Z)$ are input into discriminator D at the same time to judge the true and false attributes of each sample respectively. The purpose of discriminator D is to judge X_t to be true with probability 1, and to judge the generated samples to be false with probability 1. The purpose of generator G is to "confuse" the discriminator D, so that the probability that the discriminator D judge $G(Z)$ to be true is about 0.5. Finally, through the antagonistic training, the generator can generate the false sample $G(Z)$ that the discriminator cannot distinguish true from false, and the generated false sample $G(Z)$ can be used to expand the original training set to form a new training set.

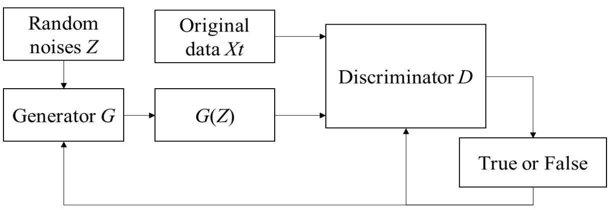

Fig. 4 Sketch map of the GAN

3.2. Heterogeneous Ensemble Fault Classifier

In order to improve generalization of our approach, we construct a heterogeneous ensemble model after oversampling minority-class samples by KM-GAN. First, we train and build two classification models based on DNN and DBN respectively. Then, we integrate these two models to combine their advantages by calculating the average value of the output probabilities from DNN and DBN. Finally, the classification results of faults are determined through the average probabilities. The specific process is shown in Fig. 5.

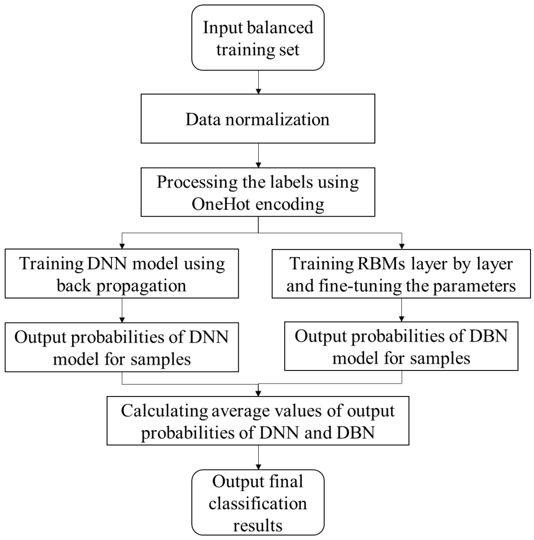

Fig. 5 Specific implementation plan of the heterogeneous ensemble model

4. EXPERIMENT

In this section, our proposed method is empirically evaluated on the challenging public dataset for industrial fault diagnosis, to verify its feasibility and effectiveness. We first introduce the implementation details of our experiment configurations, i.e. dataset, parameter settings and evaluation metrics. Then a series of experiments are conducted, and all the experimental results are demonstrated and analyzed.

4.1. Implementation details

Our experiments are conducted on the UCI dataset called Dataset for Sensorless Drive Diagnosis Data Set, whose all attribute values are real numbers. This dataset has 11 kinds of samples and the number of characteristic dimensions is 49, from which we select four kinds of samples to form the original unbalanced dataset, and the number of samples for each category is 248, 1178, 1219 and 1312 respectively. In our experiments, the first class with a sample number of 248 is regarded as the minority-class, and the remaining three classes are the majority-class.

In our experiments, firstly the original unbalanced dataset is pre-processed and all the data is normalized to [0, 1]. Afterwards, this dataset is randomly divided into the original training set and testing set, then the minority-class samples in the original training set are taken out for K-means clustering. Specifically, K is set to 3 in this paper. Next, we will introduce the parameters of classifiers.

The DNN in our experiments has only one hidden layer and the number of its neurons is 12. Moreover, the number of neurons in the input layer and output layer is 49 and 4 respectively. Additionally, we select ReLU as the activation function, and set the learning rate and batch size to 0.001 and 200 respectively, the number of iterations is 40. While for DBN that has two hidden layers, the number of neurons in the first and second hidden layer is 8 and 6 respectively. Moreover, the number of neurons in both the input and output layer are the same as that of DNN. Otherwise, we still select ReLU as the activation function, and the learning rate and batch size is set to 0.001 and 150 respectively, the number of iterations is 20.

For evaluation metrics, generally, typical classification methods take the *Accuracy* into account. When the samples are imbalanced, the influence of minority-class samples on the overall accuracy is small, so even if the classification algorithm regards all samples as majority-class, it can still obtain a high accuracy. That is to say that, *Accuracy* only reflects the overall performance of classifier over the dataset but cannot represent its performance on minority-class samples well. Therefore, in our experiments, we also use other three evaluation metrics *Recall*, *Precision* and *F-measure* for its performance on minority-class samples. The *Accuracy*, *Recall*, *Precision* and *F-measure* are denoted as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

$$Precision = \frac{TP}{TP+FP} \tag{5}$$

$$F-measure = Recall \times Precision \tag{6}$$

where *TP* denotes the minority-class samples which are predicted to be minority class. *FN* means the minority-class samples that are predicted to be majority class. *FP* denotes the majority-class samples that are predicted to be minority class. *TN* refers to the majority-class samples which are predicted to be majority class. It can be found that *Accuracy* reflects the ability of the classifier to recognize all types of samples. *Recall* represents the proportion of samples which are classified as the minority-class in the total minority-class samples. *Precision* reflects the proportion of the samples which are correctly predicted as the minority-class in all the samples classified as minority class. *F-measure* considers *Recall* and *Precision* comprehensively.

4.2. Experimental results and analysis

For performance comparison purpose, we evaluate Raw data+DNN, Random Over-Sampling+DNN (ROS+DNN), SMOTE+DNN, GAN+DNN, KM-GAN+DNN and our proposed method on UCI dataset. Our experiments report the results on four evaluation metrics for all the methods. Through the comparison experiments of the first five methods, the oversampling performance of KM-GAN can be evaluated. While comparing KM-GAN+DNN with our approach, we can evaluate the classification performance of the heterogeneous ensemble model, since the same oversampling method is adopted in these two approaches. The experimental results on *Accuracy* (ACC), *Recall* (REC), *Precision* (PRE) and *F-measure* (F-M) are shown in Fig. 6-9 respectively, and more specific results are summarized in Table. 1.

As shown in Table. 1, our method obtains the best results on four evaluation metrics. And as Fig. 6 shows, the ACC of the proposed approach is about 12% higher than that of Raw data+DNN. It probably results from that Raw data+DNN has none pre-processing for the original imbalanced dataset, while we adopt K-means clustering GAN and heterogeneous ensemble model to focus more attention on the minority-class samples within dataset. Therefore, our approach can finally achieve significantly better performance.

Table. 1 Specific results on four evaluation metrics.

Method	ACC	REC	PRE	F-M
Raw data+DNN	0.68	0.53	0.65	0.59
ROS+DNN	0.66	0.60	0.90	0.72
SMOTE+DNN	0.71	0.98	0.87	0.92
GAN+DNN	0.75	0.83	1.00	0.91
KM-GAN+DNN	0.78	0.95	1.00	0.97
Our Method	**0.82**	**0.98**	**1.00**	**0.99**

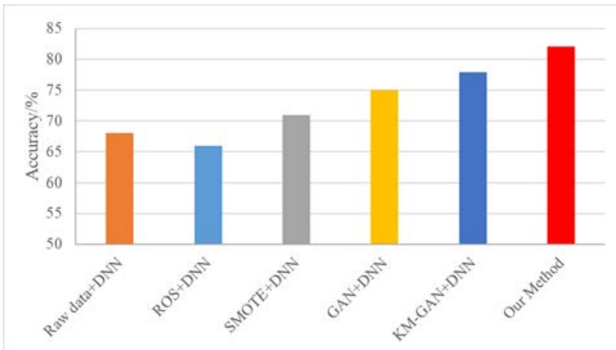

Fig. 6 Results on Accuracy

It is easy to observe that the REC of our approach is much higher than that of ROS+DNN in Fig. 7. This may be principally ascribed to that ROS+DNN produce new samples only through copying the existing data, which is easy to cause over fitting. However, our approach exploits GAN to generate diverse samples with K modes, reducing the risk of over fitting. In addition, we can find that our approach also outperforms GAN+DNN in Fig. 7. This may be mostly owing to that GAN+DNN is easy to give rise to insufficient diversity for the new samples, which is also called model collapse. Unlike GAN+DNN, we employ K-means clustering that can create new samples under guarantee for the diversity of categories. Hence, KM-GAN+DNN greatly weakens the model collapse and achieves better performance.

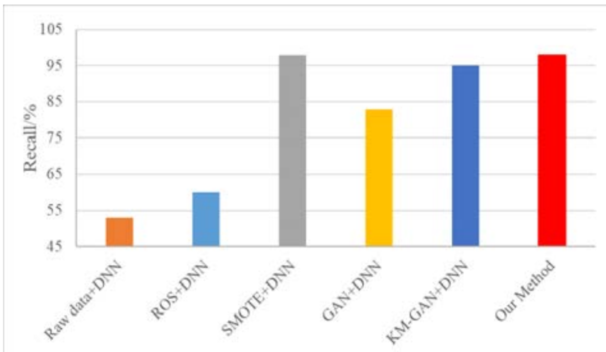

Fig. 7 Results on Recall

It can be seen in Fig. 8 that our method surpasses SMOTE+DNN about 15% in terms of PRE. It is presumably because that SMOTE+DNN can insert a new sample between the two existing ones, which reduces the risk of over fitting but may lead to more noises for data, resulting in misclassification for minority class. Nevertheless, our approach utilizes K-means clustering GAN to fit the distribution of all the original data while creating new samples, thus can obtain better results.

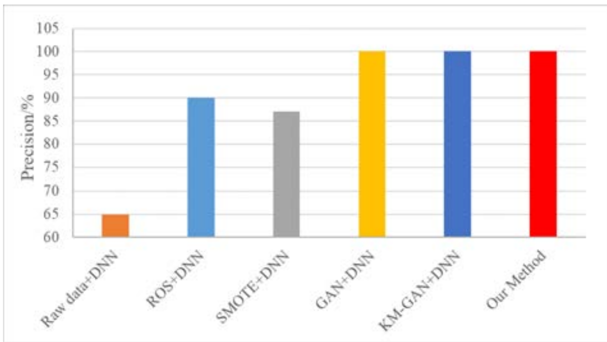

Fig. 8 Results on Precision

Moreover, it can be found in Fig. 9 that our proposed method obtains the best performance on F-M. Leveraging the preferences of different model for different classes, the recognition performance of our approach for majority class is not reduced much while improving the identification rate of minority class. And compared with KM-GAN+DNN, the F-M of the proposed approach is little higher. It perhaps because our approach applies heterogeneous ensemble to combine the advantages of both DNN and DBN, and obtain a unified ensemble learning model and finally achieve more accurate and stable results.

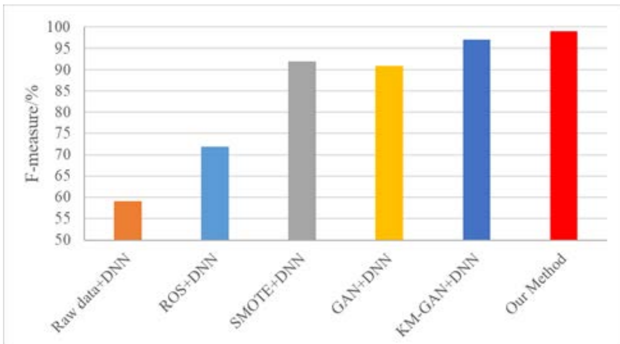

Fig. 9 Results on F-measure

5. Conclusion

As one of the great challenges in fault diagnosis area, class imbalance has attracted much attention from scholars recently. In this work, aiming at improving diagnostic accuracy for minority-class faults and constructing a classifier with better generalization, we present a K-Means clustering GAN based fault diagnosis approach for imbalanced dataset. First, a novel K-Means clustering algorithm and GAN based oversampling method is designed to generate diverse minority-class samples which obey the similar distribution with the original data. Second, a heterogeneous ensemble model based on DNN and DBN is constructed as a fault classifier to further improve the generalization of our diagnosis model as well as the accuracy of its fault recognition. Experimental results demonstrate that the proposed approach outperforms the compared ones. However, we only testify our approach on one dataset, in the near future, we will try more datasets and continue our study on class-imbalance data driven deep learning methods so as to further improve the identification accuracy for minority-class samples.

Acknowledgements

This work was supported by the National Natural Science Foundation of China [grant number 61836001].

References:

[1] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li and A. K. Nandi, “Applications of machine learning to machine fault diagnosis: A review and roadmap,” Mechanical Systems and Signal Processing, vol. 138, pp. 106587, April 2020.

[2] H. Choi, S. Choi, S. Han and J. Oh, “Fault diagnosis of planetary gear carrier packs: A class imbalance and multiclass classification problem,” International Journal of Precision Engineering and Manufacturing, vol. 20, no. 2, pp. 167-179, February 2019.

[3] Y. He, J. Zhou, Y. Lin and T. Zhu, “A class imbalance-aware Relief algorithm for the classification of tumors using microarray gene expression data,” Computational Biology & Chemistry, vol. 80, pp. 121-127, March 2019.

[4] H. X. Guo, Y. J. Li, J. Shang, M. Y. Gu, Y. Y. Huang and B. Gong, “Learning from class-imbalanced data: review of methods and applications,” Expert Systems with Applications, vol. 73, pp. 220-239, May 2017.

[5] J. Zhang and I. Mani, “KNN approach to unbalanced data distributions: A case study involving information extraction,” International Conference on Machine Learning (ICML 2003), Workshop on Learning from Imbalanced Datasets, Washington DC, America, August 2003.

[6] R. Barandela, R. M. Valdovinos, J. S. Sánchez and F. J. Ferri, “The imbalanced training sample problem: under or oversampling?” Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition and Structural and Syntactic Pattern Recognition, Lisbon, Portugal, pp. 806–814, 2004.

[7] G. E. A. P. A. Batista, R. C. Prati and M. C. Monard, “A study of the behavior of several methods for balancing machine learning training data,” ACM Special Interest Group on Knowledge Discovery and Data Mining Explorations Newsletter, vol. 6, no. 1, pp. 20-29, June 2004.

[8] N. V. Chawla, K. W. Bowyer, L. O. Hall and W. P. Kegelmeyer, “SMOTE: Synthetic minority over-sampling technique,” Journal of Artificial Intelligence Research, vol. 16, pp. 321-357, June 2002.

[9] T. Zhu, Y. Lin and Y. Liu, “Synthetic minority oversampling technique for multiclass imbalance problems,” Pattern Recognition, vol. 72, pp. 327-340, December 2017.

[10] I. J. Goodfellow, “Generative adversarial nets,” International Conference on Neural Information Processing Systems, Montreal, Canada, 2014, pp. 2672-2680.

[11] G. Douzas and F. Bacao, “Effective data generation for imbalanced learning using conditional generative adversarial networks,” Expert Systems with Applications, vol. 91, pp. 464-471, January 2018.

[12] Z. Wang, J. Wang and Y. Wang, “An intelligent diagnosis scheme based on generative adversarial learning deep neural networks and its application to planetary gearbox fault pattern recognition,” Neurocomputing, vol. 310, no. 8, pp. 213-222, October 2018.

[13] H. Liu, J. Zhou, Y. Xu, Y. Zheng, X. Peng and W. Jiang, “Unsupervised fault diagnosis of rolling bearings using a deep neural network based on generative adversarial networks,” Neurocomputing, vol. 315, no. 13, pp. 412-424, July 2018.

[14] T. Han, C. Liu, W. Yang and D. Jiang, “A novel adversarial learning framework in deep convolutional neural network for intelligent diagnosis of mechanical faults,” Knowledge-Based Systems, vol. 165, pp. 474-487, December 2018.

[15] M. Arjovsky, S. Chintala and L. Bottou, “Wasserstein GAN,” 34th International Conference on Machine Learning, Sydney, Australia, pp. 214-223, December 2017.

[16] J. A. Hartigan and M. A. Wong, “A K-Means clustering algorithm,” Applied Statistics, vol. 28, no. 1, pp. 100-108, January 1979.

[17] F. Rosenblatt, “The perceptron: a probabilistic model for information storage and organization in the brain,” Psychological Review, vol. 65, pp. 386-408, November 1958.

[18] D. E. Rumelhart, G. E. Hinton and R. J. Williams, “Learning representations by back propagating errors,” Nature, vol. 323, no. 6088, pp. 533-536, October 1986.