SEVERAL APPROACHES FOR ANOMALY DETECTION FROM SOUND

Technical Report

Yaoguang Wang[1],Yaohao Zheng[2], Yunxiang Zhang[2], Ying Hu[2], Minqiang Xu[3], Liang He[1,2]

[1]Tsinghua University,Department of Electronic Engineering
[2]Xinjiang University, School of information science and engineering
[3]SpeakIn Technology

ABSTRACT

The task2 of IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events mainly research unsupervised anomalous sound detection for machine condition monitoring under domain shifted conditions, three methods are proposed to solve this problem: principal component analysis (PCA), outlier classifier and contarative learning. Firstly, PCA is used for anomaly detection with three components. Secondly, outlier classifier is used by selecting the normal sound of other section as outlier samples. At last, contrastive learning is used by taking normal samples of other sections as negative examples. We present results obtained by each kind of models separately,as well as,a results of an esemble obtained by averaging anomaly scores computed by individual models.

Index Terms— Anomaly Detection, PCA, Outlier classifier, Comparative learning

1. INTRODUCTION

Unsupervised Anomalous Sound Detection for Machine Condition Monitoring under Domain Shifted Conditionsnder Domain Shifted Conditions, This task is the follow-up task of task 2 in dcase2020. The task is divided into two parts. One is to detect unknown abnormal sound while only providing normal sound clips as training data. Second, the task is performed under the condition that the acoustic characteristics of training data and test data are different (i.e. domain shift).

The data set provided by the task includes seven kinds of normal and abnormal operation sounds of machines, including fan, gearbox, pump, slider, toy car, toy train and valve. The data set is divided into the development, additional training, and evaluation data sets, and each machine is divided into three sections. The same machine may appear in different sections, and different machines may appear in the same section. Development data sets contains section00,01,02, and each section is a complete set of training and test data. There are about 1000 normal sound clips for training in the source domain of the development data set, only three normal sound clips for training in the target domain, and about 100 normal and abnormal sound clips in the test target domain. Additional training data sets and evaluation data sets contain section03,04,05, and there are 1000 normal sound clips in the source domain of the additional training set, There are only three normal sound clips in the target domain, in the test set, the sound clips in the source domain and the target domain are the same, and there are no conditional tags.

This task was evaluated using Area under curve(AUC) and pAUC. AUC is the area enclosed with abscissa under the Receiver Operating Characteristic(ROC). The abscissa of ROC curve is false positive rate or false alarm rate, which is related to the probability of normal voice being recognized as abnormal voice; The ordinate is the true positive rate or detection rate, that is, the probability of abnormal sound being accurately identified.The pAUC is calculated as the AUC over a low false-positive-rate (FPR) range [0,p] [1] [2] [3].

2. PROPOSED METHOD

2.1. principal component analysis

PCA is often used to reduce the dimension of high-dimensional data, and can be used to extract the main feature components of data. The main idea of PCA is to map high-dimensional features to low-dimensional features. This low-dimensional feature is a new orthogonal feature, also known as principal component [4]. It is a low dimensional feature reconstructed on the basis of the original high-dimensional features. In this method, each sample given as a time series signal is converted into a log-mel-spectrogram.

We follow [5] for anomaly detection. In this paper, three components are used:mean component, basis component, latent component. The anomaly for the mean component is calculated by a simple Mahalanobis distance.The anomaly for the basis component is calculated with k-nearestneighbor based on the distance between the subspace spanned by the basis component.The anomaly for latent component is calculated by matrix density estimation [5].

2.2. outlier classifier

When the normal audio samples of the machine to be tested can be obtained, sometimes the normal samples of other machines of the same type can also be obtained easily. At this time, the model can use a large amount of additional data to improve its performance. This section will design a variety of models to explore the impact of different application methods of section information on the accuracy of anomaly detection. Although different section machines belong to the same type, their samples are still regarded as different categories, so unsupervised learning is transformed into supervised learning. The supervised training model can be directly used for anomaly detection, and can also be used as feature extractor as the upstream model of anomaly detection. The sound of the machine to be tested can be regarded as normal samples, and the normal samples of all other machines can be regarded as outlier samples (abnormal samples).

The input of the above model can be divided into two categories, one is the sound of the target machine as a positive sample (the label is set to 0), the other is the sound of all other machines as a negative sample (the label is set to 1). The purpose of network training is to identify whether the sound belongs to the target machine, so that the binary classification network can be trained by using the cross entropy loss function. The biggest advantage of the network is that the output of the network can be used as the abnormal score directly in the reasoning stage, which saves the trouble of calculating the abnormal score. Its essence is anomaly detection based on outlier detection. The hyperplane represented by the two classifiers separates the normal mode of the target machine from the normal mode of the non target machine, which is different from the traditional classifier. If the input is the normal sound of the target machine, the output probability should be closer to 0, otherwise the output probability is farther away from 0. The normal and abnormal samples can be distinguished by taking the output probability as the abnormal score directly. In addition, because the normal sound of different section machines is very close in acoustic mode, the decision surface of the classifier can well ”surround” the normal data distribution of the target machine, so the two classifiers have a strong ability to identify anomalies.

2.3. contrastive learning

Contrastive learning uses the powerful ability of Siamese Neural Network (SNN) in extracting invariant features to replace the similarity calculation module, which makes the feature extraction and similarity calculation process naturally integrate into a whole network and greatly reduces the network complexity [6]. The SNN consists of a pair of parallel twin networks. The network structures of the two branches are identical and the parameters of the training process are completely shared. The input is a sample pair, which is a pair of samples randomly selected from the normal samples and a small number of abnormal samples. A small number of abnormal samples are first simply enhanced to expand their scale. The data enhancement method used here is the random cyclic shift of audio in the waveform domain.

The definition of sample pair includes two types. If both are normal samples, it is defined as positive sample pair, and its label is set to 0; If one is a normal sample and the other is an abnormal sample, it is defined as a negative sample pair, and its label is set to 1. Convolutional neural network can extract the discriminative information of sample pairs, and then calculate the similarity of their high-level semantic features. Because the same network structure is used, the samples of phase are mapped to the similar positions in the high-level space, and the sample pairs with great differences are mapped to the positions far away. In order to achieve this goal, the contrast energy function can be used to reduce the contrast energy of similar sample pairs and increase the contrast energy of dissimilar sample pairs in the training process.

3. EXPERIMENTS RESULTS

We use the following four systems to detect anomalies in the data set of DCASE 2021.

1. Principal component analysis method
2. Outlier classifier
3. Contrastive Learning method
4. Ensemble method, choose the best results among seven machines from the first three methods.

On the test data in the development dataset, we archieve the following AUC and pAUC results as shown in Table 1.

4. CONCLUSION

In this technical report, we have used three methods to detect the abnormality of machine monitoring. But we did not adopt the method of domain migration. For each machine type, we use three models to detect anomalies in the source domain and the target domain. We present AUC and pAUC results for the challenge baseline model and our 3 method for all 7 machines in table 1.

5. REFERENCES

[1] Y. Kawaguchi, K. Imoto, Y. Koizumi, N. Harada, D. Niizumi, K. Dohi, R. Tanabe, H. Purohit, and T. Endo, “Description and discussion on DCASE 2021 challenge task 2: Unsupervised anomalous sound detection for machine condition monitoring under domain shifted conditions,” In arXiv e-prints: 2106.04492, 1–5, 2021.

[2] R. Tanabe, H. Purohit, K. Dohi, T. Endo, Y. Nikaido, T. Nakamura, and Y. Kawaguchi, “MIMII DUE: Sound dataset for malfunctioning industrial machine investigation and inspection with domain shifts due to changes in operational and environmental conditions,” In arXiv e-prints: 2006.05822, 1–4, 2021.

[3] N. Harada, D. Niizumi, D. Takeuchi, Y. Ohishi, M. Yasuda, and S. Saito, “ToyADMOS2: Another dataset of miniature-machine operating sounds for anomalous sound detection under domain shift conditions,” arXiv preprint arXiv:2106.02369, 2021.

[4] https://blog.csdn.net/program_developer/article/details/80632779.

[5] Y. Sakamoto and N. Miyamoto, “Anomaly calculation for each components of sound data and its integration for dcase 2020 challenge task2,” DCASE2020 Challenge, Tech. Rep., July 2020.

[6] X. Chen and K. He, “Exploring simple siamese representation learning,” 2020.

Table 1: AUC (%) and pAUC (%) for each machine

	ToyCar	ToyTrain	fan	gearbox	pump	slider	valve
	AUC(pAUC)	AUC(pAUC)	AUC(pAUC)	AUC(pAUC)	AUC(pAUC)	AUC(pAUC)	AUC(pAUC)
Baseline	62.49(52.36)	61.71(53.81)	63.24(53.38)	65.97(52.76)	61.92(54.41)	66.74(55.94)	53.41(50.54)
System 1	56.85(56.14)	58.95(55.27)	68.99(56.15)	61.19(63.84)	67.80(56.17)	64.17(54.96)	71.49(59.62)
System 2	60.15(53.96)	62.28(58.90)	57.64(66.16)	70.56(61.32)	67.33(57.52)	63.96(58.60)	66.07(59.72)
System 3	65.07(57.85)	58.66(58.15)	63.45(62.25)	67.97(58.00)	63.70(55.50)	60.59(57.13)	62.88(57.42)