ATTRIBUTE CLASSIFIER WITH IMBALANCE COMPENSATION FOR ANOMALOUS SOUND DETECTION

Technical Report

Yifan Zhou

Shanghai Normal University
Shanghai, China
1000511937@smail.shnu.edu.cn

Yanhua Long

Shanghai Normal University
Shanghai, China
yanhua@shnu.edu.cn

ABSTRACT

This paper proposes an Attribute Classifier with Imbalance Compensation (ACIC) for DCASE 2023 Challenge Task 2. The goal is to perform anomalous sound detection by exploiting prior knowledge about machine attributes. First, we propose to use the weak prior knowledge provided by attribute for anomaly detection. Then, we design the Imbalance Compensation (IC) strategy to address the class imbalance problem of attributes. Finally, we propose a score fusion method based on ACIC to enhance the robustness of the model. Experimental results show that compensating for attribute class imbalance improves the exposure of anomalies.

Index Terms— Anomalous sound detection, attribute classification, imbalance compensation, data augmentation

1. INTRODUCTION

In recent years, with the rapid development of artificial intelligence technology, anomalous sound detection (ASD) based on machine sounds has become an important part of Industry 4.0. In real-world scenarios, abnormal samples cannot be obtained by damaging the machine. Therefore, anomaly detection with only normal samples is the biggest challenge. Moreover, the complex engineering environment introduces a lot of noise into the sound samples, and the operating settings of different machines are different, which also puts forward the requirement of domain generalization for algorithms.

In previous works of the Detection and Classification of Acoustic Scenes and Events (DCASE) Challenge Task 2 [1], generative model-based anomaly detection systems such as AutoEncoders (AEs) have been widely used due to their simple design and efficient inference. AE-based anomaly detection algorithms require the hypothesis that the model has poor reconstruction capability for anomaly samples. However, AE reconstruction itself has a certain denoising characteristic. If the representation ability of AE neural networks is excessively improved, it will eliminate the significance of anomaly samples, leading to a decrease in anomaly detection performance. Therefore, discriminative model-based systems have been designed and achieved excellent performance [2, 3, 4]. Such models utilize powerful feature extraction networks of deep learning to perform the machine ID classification task. During inference, abnormal samples are exposed due to the difficulty of classification, achieving anomaly detection.

It is undeniable that anomaly detection algorithms based on machine ID classification have achieved excellent performance in previous DCASE TASK2. The success of this algorithm relies on the high-quality classification boundaries provided by the strong prior knowledge of machine ID. Unfortunately, in real production, we cannot obtain such high-quality prior knowledge about machine ID. We can't help but ask, how to apply discriminative model-based anomaly detection algorithms under limited prior knowledge?

To solve such problems, we need to design anomaly detection algorithms under weak prior knowledge conditions. According to the task setting of DCASE 2023 Task2, we cannot obtain high-quality prior knowledge such as machine ID. However, we can obtain the attribute information of each audio clip, such as the microphone number, machine running speed, machine load condition, etc. We define the attribute information of the audio clips as weak prior knowledge. The labels of such attributes are extremely unbalanced, complex categories, and unable to form clear classification boundaries.

In this paper, we propose the Attribute Classifier with Imbalance Compensation (ACIC) method, aiming to overcome the disadvantages of weak prior knowledge and use attribute labels to build discriminative models for anomaly detection. Our main contributions are as follows: (1) We use the weak prior knowledge provided by attribute information for anomaly detection, making the application of discriminative models possible when machine ID labels are limited. (2) We propose the Imbalance Compensation (IC) strategy to solve the common problem of extreme sample imbalance in attribute labels. (3) We propose a score fusion method based on ACIC to enhance the robustness of the model.

2. PROPOSED METHOD

The ACIC framework we propose contains an imbalance compensation module, an attribute classifier, and an anomaly score fusion calculator. The imbalance compensation module aims to solve the problem of extreme sample imbalance in attributes. In addition, the novelty of the attribute classifier lies in breaking through the limitations of strong prior knowledge such as machine ID. The overview of the overall framework is shown in Fig.1.

2.1. Attribute Classifier

Although the previous work using machine IDs for classification achieved good results, it is often more limited in practical applications, such as only one machine is working. In this case, strong prior

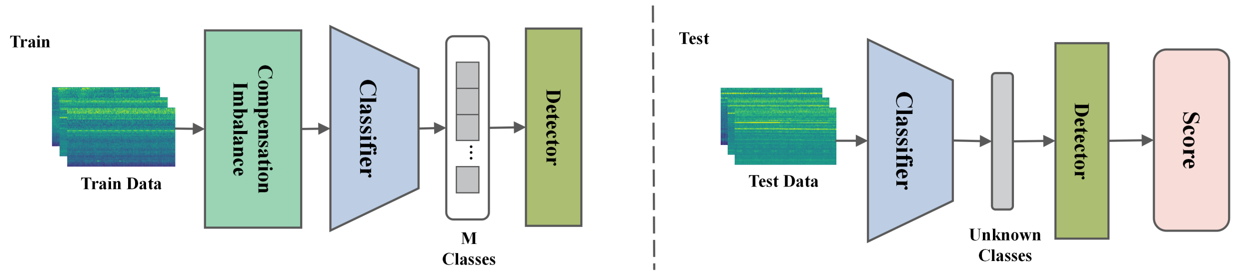

Figure 1: The framework of proposed ACIC. During training, by classifying the attribute labels, the embedding of the samples after IC is extracted and used to train an anomaly detector. During testing, the embedding of the test samples is extracted through the trained classifier, and the trained anomaly detector is used to calculate the score.

knowledge such as machine IDs cannot be used. Nevertheless, machines still have weak prior knowledge that is easy to obtain, such as attributes in Table 1. Therefore, we train a discriminative model using such weak prior knowledge as attributes.

Table 1: Attributes of different machines.

	Attribute 1	Attribute 2	Attribute 3
ToyCar	Car model	Speed	Mic
ToyTrain	Train model	Speed	Mic
Fan	Mixing of machine	N/A	N/A
Gearbox	Voltage	Weight	N/A
Bearing	Velocity	Mic	N/A
Slider	Velocity	Acceleration	N/A
Valve	Open/close	N/A	N/A

Taking ToyCar as an example, Table 2 shows the label information of different attributes. Among them, Car model includes 10 categories from A1 to E2, Speed includes 5 categories from 28V to 40V, and Mic includes 2 categories, 1 and 2. The number of samples is expressed in the form of 'Category: Number'. Obviously, machine attributes are weaker prior knowledge than machine IDs for the following reasons: (1) The number of attributes and the number of labels for each attribute of each type of machine are inconsistent. (2) The number of samples of different labels in the same attribute is extremely unbalanced. (3) The machine attributes are unknown during testing. Therefore, we design an Imbalance Compensation (IC) module and a score fusion calculator to solve the above challenges caused by attributes.

After the IC module, we train a discriminative model for each attribute of each machine to distinguish which label the samples come from in the attribute. In particular, during training, we use ResNet18 as the backbone of the classifier to perform the classification task and obtain the embedding of each attribute. Then, we use these embeddings to train an anomaly detector, here we use KNN. This anomaly detector can characterize the spatial relationship between each category in the machine's attributes and form a better classification boundary. During testing, we obtain the embedding of the test samples through the trained classifier. At this time, the embedding of abnormal samples is difficult to be distinguished by the classifier and exhibits outlying and low-density characteristics in the feature space, which is easy to achieve anomaly detection.

Table 2: Different labels of attributes in ToyCar dataset.

	Car model	Speed	Mic
Label 1	C1: 215	31V: 350	1: 990
Label 2	D1: 214	40V: 350	2: 10
Label 3	B1: 166	34V: 290	N/A
Label 4	B2: 164	28V: 5	N/A
Label 5	D2: 116	37V: 5	N/A
Label 6	C2: 115	N/A	N/A
Label 7	A1: 3	N/A	N/A
Label 8	E2: 3	N/A	N/A
Label 9	A2: 2	N/A	N/A
Label 10	E1: 2	N/A	N/A

2.2. Imbalance Compensation

To solve the problem of extremely unbalanced samples among different labels shown in Table 2, we propose the IC module. The module mainly includes two parts: (1) Maximum expansion-uniform sampling (MEUS). (2) Robust data transformation (RDT), as shown in Fig.2.

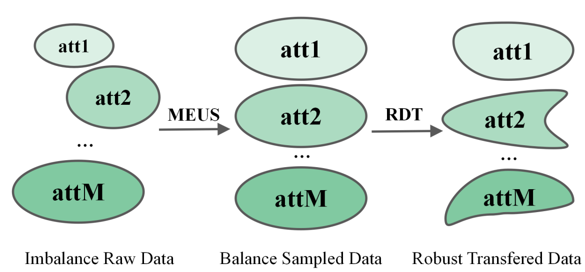

Figure 2: A schematic diagram of the effects of imbalance compensation on data.

When applying MEUS, in the first step, we obtain M labels of the current attribute. At the same time, we can know the number of samples in each category and take the maximum number of samples in the class as N. Then, we perform uniform sampling in each class respectively, and the sampling times are also N. In this way, the

Table 3: Performance of AUC(source domain) / AUC(target domain) / pAUC(%) (p=0.1) comparison on seven different ASD tasks.

	AE(MSE)	AE(MAHA)	AE-IC(MSE)	AE-IC(MAHA)	ACIC(MSP)	ACIC(KNN)
ToyCar	70.1/46.89/52.47	74.53/43.42/49.18	59.64/64.16/51.05	70.22/58.26/49.10	68.28/51.42/51.47	72.23/50.05/53.89
ToyTrain	57.93/57.02/48.57	55.98/42.45/48.13	55.48/58.88/48.36	53.26/45.4/48.47	68.86/49.59/52.84	63.4/48.68/51.42
Bearing	65.92/55.75/50.42	65.16/55.28/51.37	63.98/62.8/51.26	62.14/64.8/51.89	74.44/55.92/49.26	79.86/54.64/57.47
fan	80.19/36.18/59.04	87.1/45.98/59.33	85.86/62.4/63.78	75.98/87.2/63.36	64.36/70.34/61.42	65.18/79.62/63.73
gearbox	60.31/60.69/53.22	71.88/70.78/54.34	65.64/64.8/54.78	74.14/71.08/55.63	55.39/61.26/54.26	52.4/60.56/54.15
slider	70.31/48.77/56.37	84.02/73.29/54.72	63.22/47.04/54.94	81.08/73.06/53.26	80.8/55.26/54.26	82.64/56.44/54.31
valve	55.35/50.69/51.18	56.31/51.4/51.08	51.02/47.98/51.47	55.64/49.32/51	69.32/22.58/48.84	71.38/34.85/49.78

number of samples in each category is sampled to N. In the second step, due to the previous sampling, the number of samples in each class reaches balance. At this time, uniform sampling is performed between classes. When the number of samples is large, according to the law of large numbers (LLN), the samples after sampling satisfy the same distribution as the samples before sampling. Therefore, in this step, K samplings are performed to expand the samples in each category to K, while satisfying the balance between samples between categories. When applying RDT, we perform Waveform transforms on K samples after MEUS respectively. In particular, we adopted AddGaussianNoise, TimeStretch, PitchShift, Shift four data augmentation methods for each sample, and the probability of each augmentation method is 0.5. After RDT, we discard the original data and only use the audio clip after data augmentation for training. The combination of MEUS and RDT solved the problem of extreme imbalance of attribute samples. By transforming, certain disturbances are introduced to the expanded samples to improve the robustness of the model and give the model domain generalization ability.

2.3. Anomaly Detector

Considering that different machines have different attributes, such as ToyCar with 3 attributes while Fan only has 1 attribute, we train a model for each attribute and train an anomaly detector for each attribute, as shown in Fig.3. Here, our anomaly detector uses KNN.

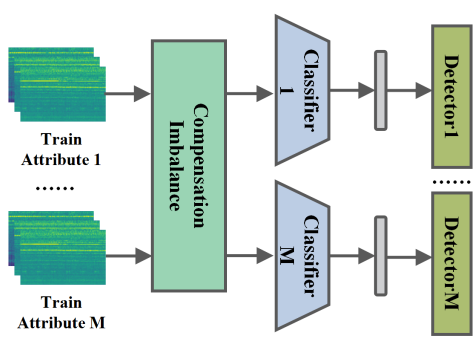

Figure 3: The schematic diagram of the anomaly detector training stage.

During testing, we input the embedding obtained by each classifier for the test sample into each anomaly detector respectively, and take the average score, as shown in Fig.4. Therefore, anomaly detection from the perspective of each attribute classifier improves the robustness of the anomaly detection system.

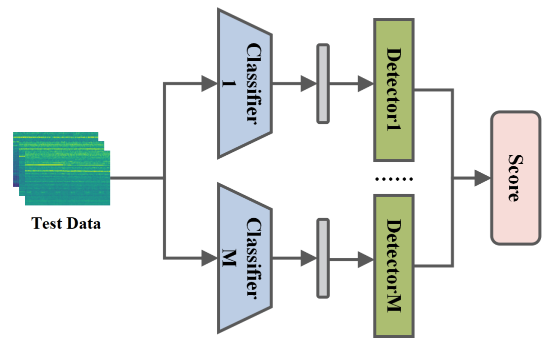

Figure 4: The schematic diagram of the anomaly detector testing stage.

3. EXPERIMENTS AND RESULTS

Our systems are developed on the development dataset released in DCASE2023 Task2. The dataset contains 7 machines, each with different attributes, and different attributes contain data with different labels [5, 6]. Related experimental settings are described in this section.

In this study, we use 128-dimensional log-mel-spectrogram as input features for the classifier, with a frame length of 64ms and a hop size of 50%. The classifier uses ResNet18, the Adam optimizer, a learning rate of 0.0001, trains for 15 epochs, and a batch size of 128. In addition, the data augmentation used in the IC module uses the audiomendation toolkit [7], and the KNN training uses the pyod toolkit [8]. Same as the baseline system, we use AUC and pAUC as evaluation metrics.

For anomaly scores, we compare the Maximum Softmax Probability (MSP) [9] of the classifier output embedding and the KNN score respectively. In addition, to verify the effectiveness of the IC module, we applied the IC module to the AE baseline. The results are shown in Table 3, which shows that the proposed ACIC has significantly improved performance.

Finally, we submit four systems trained on the evaluation dataset, which are: (1) ACIC-AE selected: Different models are selected for different machines (2) ACIC: The ACIC is used for all machines (3) AE-IC: AE is used for all machines, and the data processing uses the IC module. (4) ACIC-AE ensemble: The average

Z-score of ACIC and AE scores is used for all machines.

4. REFERENCES

[1] K. Dohi, K. Imoto, N. Harada, D. Niizumi, Y. Koizumi, T. Nishida, H. Purohit, R. Tanabe, T. Endo, and Y. Kawaguchi, "Description and discussion on dcase 2023 challenge task 2: First-shot unsupervised anomalous sound detection for machine condition monitoring," *arXiv preprint arXiv:2305.07828*, 2023.

[2] Y. Zeng, H. Liu, L. Xu, Y. Zhou, and L. Gan, "Robust anomaly sound detection framework for machine condition monitoring," DCASE2022 Challenge, Tech. Rep., 2022.

[3] I. Kuroyanagi, T. Hayashi, K. Takeda, and T. Toda, "Two-stage anomalous sound detection systems using domain generalization and specialization techniques," DCASE2022 Challenge, Tech. Rep., 2022.

[4] Y. Deng, J. Liu, and W.-Q. Zhang, "Aithu system for unsupervised anomalous detection of machine working status via sounding," DCASE2022 Challenge, Tech. Rep., 2022.

[5] 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.

[6] K. Dohi, T. Nishida, H. Purohit, R. Tanabe, T. Endo, M. Yamamoto, Y. Nikaido, and Y. Kawaguchi, "Mimii dg: Sound dataset for malfunctioning industrial machine investigation and inspection for domain generalization task," *arXiv preprint arXiv:2205.13879*, 2022.

[7] https://github.com/iver56/audiomentations.

[8] Y. Zhao, Z. Nasrullah, Z. Li, S. Han, X. Hu, H. Huang, M. Jiang, and Y. Zhao, "Pyod: A python toolbox for scalable outlier detection," *Journal of Machine Learning Research*, vol. 20, pp. 1–7.

[9] D. Hendrycks, M. Mazeika, and T. Dietterich, "Deep anomaly detection with outlier exposure," in *International Conference on Learning Representations*, 2019.