TWO-STEP ANOMALY DETECTION: INTEGRATING ATTRIBUTE CLASSIFICATION AND GENERATIVE MODELING WITH ATTRIBUTE INFERENCE FOR DIVERSE MACHINE

Technical Report

Lei Wang[1], Mingqi Cai[2], Jia Pan[2], Tian Gao[2], Xin Fang[2]

[1]iFLYTEK, Hefei, China
{leiwang32, mqcai, jiapan, tiangao5, xinfang}@iflytek.com

ABSTRACT

This study presents a novel approach to address the DCASE2024 Challenge Task2[1], focusing on First-Shot Unsupervised Anomalous Sound Detection for Machine Condition Monitoring. The task presents a unique challenge of training models with and without attribute information, necessitating robust performance under both scenarios.

Our proposed method combines attribute classification and generative modeling to address the variability in attribute information across different machine types. Initially, we perform attribute prediction or clustering to infer labels for machines without explicit attribute information. Using these inferred labels, we then apply comprehensive attribute classification across all machine types. Concurrently, we integrate an Autoencoder (AE) to analyze reconstruction losses, enhancing anomaly detection. This two-step approach − first predicting or clustering attributes, and then merging attribute classification with generative model results − ensures robust and effective anomaly detection, maintaining high performance regardless of the presence or absence of attribute information across various scenarios.

Experimental results demonstrate the effectiveness of our approach, achieving a competitive harmonic mean of AUC and PAUC(p = 0.1) of 63.09% on the development set. This innovative framework offers a versatile solution for anomaly detection in machine condition monitoring, accommodating real-world data variability and enhancing overall system robustness.

Index Terms— Anomalous sound detection, Attribute classification, Generative modeling, Attribute prediction, Machine condition monitoring, Two-step approach

1. INTRODUCTION

Anomalous sound detection (ASD) is the task of identifying whether the sound emitted from a target machine is normal or anomalous. In DCASE challenge 2024 Task 2 ″ First-shot unsupervised anomalous sound detection for machine condition monitoring″, the ″first-shot″ ASD task which was designed as a scenario in which ASD needs to be performed on completely new machine types with data collected from only a single identification (ID) of that machine, without machine-specific hyperparameter tuning. The task was similar to DCASE challenge 2023 Task 2 but has a few modifications. First, the evaluation dataset is updated with new machine types unseen in the previous DCASE ASD challenges. Second the attribute information such as the machine operation conditions of some machine types are concealed for several machine types, which means we cannot use machine attribute information as auxiliary information to detect anomalies.

In this work, we propose a two-step anomaly detection approach. First, we employ an unsupervised attribute analysis method to generate pseudo-attribute labels for machines lacking attribute information. Second, we use an integrated attribute classification approach to model all machines in the development and additional sets, while also conducting separate group modeling for machines with low noise levels. Additionally, we utilize an attribute-conditioned AE model (CAE) for modeling single machine. Eventually, these subsystems are ensembled.

This paper is organized in the following manner: Section 2 describes the proposed attribute classification approach and the generative approach. In the final of this section we show the experimental results. Section 3 contains the conclusions based on our report.

2. PROPOSED ASD SYSTEM

2.1 Unsupervised Attribute Analysis Method

In task2 this year, a few machines only have domain information, the attribute informations are unavailable, so that some effective methods such as attribute classification methods cannot be used.

Unsupervised clustering algorithms are a type of machine learning technique used to automatically group data into different categories without the need for prior labeling or supervision. These algorithms analyze the similarities and differences between data points to divide them into groups with similar characteristics. The K-Means clustering algorithm[6] is a widely used unsupervised learning method for dividing a dataset into K clusters, such that each data point belongs to the cluster with the nearest mean point (cluster center). In this subsection, we use the k-means algorithm to perform unsupervised clustering on machines with missing attribute information, generating pseudo attribute information.

Among the seven machines in the development dataset, Toycar、 Fan 、Bearing and Valve have attribute informations , so we utilized the k-means algorithm to validate the performance on these machines. Through experimentation, we determined the optimal configuration: we initially cut the audio into 2-second clips and extracted mel-spectrogram features. During Extracting the spectrum ,we set the lowfreq to 200HZ and the higfreq to

7800HZ. We determined the hyperparameters K in k-means based on empirical experience.

2.2 Classification Method

Based on the unsupervised attribute analysis method described in the previous section, we obtained pseudo-attribute information for machines without attributes. In this section, we introduce a classification scheme based on machine ID, domain information, and attribute information.

Our method consists of two steps. The first step is to train a feature extraction model. The second step involves using this model to extract embeddings from audio recordings and then calculating an audio anomaly score through a unsupervised algorithm based backend.

2.2.1 Feature Extraction Model

We train a feature extractor based on the SeResNet18[13] , and we employ the AdaCos loss which has been shown to be more robust to label noise compared to softmax-based losses. For the classification task, the total number of classes corresponds to all unique combinations of machine IDs, domain labels, and attribute labels that have appeared in the dataset. For machines without attribute information, we use the unsupervised method in section 2.1 to generate pseudo-attribute labels.

Kevin Wilkinghoff proposed an effective self-supervised learning (SSL) method for ASD[11], in which he combines mixup[7], StatEx[8], and FeatEx to augment new categories, thereby increasing model generalization and enabling the model to extract better audio representations. To expand more diverse categories, we filtered machine-related audios from AudioSet[9] and mixed it with DCASE task2 data for training.

During the experiments, we discovered that modeling machines with lower noise levels as a separate group leads to improved performance. Therefore, we conducted manual analysis on the noise levels of each machine, categorizing the noise levels into low, medium, and high. Machines with low noise levels were modeled in separate groups, while machines with medium and high noise levels maintained the original scheme.

2.2.2 KNN based Backend

After obtaining the feature extractor, we train a KNN-based backend to distinguish between normal and abnormal states. We train the KNN according to attribute categories, with hyperparameters K set to 1, 2, 4, 8, using Euclidean distance as the distance metric. We pass the audio through the KNN model across all attribute categories and hyperparameters. Then, perform z-score normalization on the scores of all audios based on the same model, and take the minimum score as the anomaly score. Therefore, a higher anomaly score indicates an anomalous sample.

2.3 Conditional AutoEncoder based ASD Method

The baseline AE model[4] does not perform well on the target domain of some machines, so we made improvements to the baseline AE. We changed the linear layer to a convolutional layer and used attribute information as conditions to train Conditional Autoencoders (CAE). Also, to address the imbalance of data in the source and target domains, we used SMOTE[10] for data augmentation. Finally, we used MAHALA distance to calculate the anomaly score.

2.4 Ensemble System

We employ the ensemble learning strategy [12] to integrate the methods proposed above. In order to balance the various systems, we apply the zscore normalization based on the scores distribution of classification model and CAE model. Then, we use weighted sum of them.

2.5 Results

We compare our systems with the baseline systems of DCASE 2024challenge task 2, the AE-MSE and the AE-MAHALA. Our system outperform the baseline systems, the AUC scores of each machine is shown in Table 1.

Table 1: AUCs and pAUCs per machine type obtained on the development set

		baseline		our system
		mse	mahala	
bearing	AUC(source)	62.01%	54.43%	**71.65%**
	AUC(target)	61.40%	51.58%	**71.45%**
	pAUC	57.58%	58.82%	**58.90%**
fan	AUC(source)	67.71%	**79.37%**	77.42%
	AUC(target)	**55.24%**	42.70%	42.85%
	pAUC	**57.53%**	53.44%	57.35%
gearbox	AUC(source)	70.40%	**81.82%**	68.32%
	AUC(target)	69.34%	**74.35%**	68.44%
	pAUC	55.65%	**55.74%**	53.86%
slider	AUC(source)	66.51%	75.35%	**79.25%**
	AUC(target)	56.01%	**68.11%**	65.64%
	pAUC	51.77%	49.05%	**55.67%**
ToyCar	AUC(source)	**66.98%**	63.01%	63.64%
	AUC(target)	33.75%	37.35%	**57.10%**
	pAUC	48.77%	51.04%	**51.13%**
ToyTrain	AUC(source)	76.63%	61.99%	**81.56%**
	AUC(target)	46.92%	39.99%	**57.06%**
	pAUC	47.95%	48.21%	**55.32%**
valve	AUC(source)	51.07%	55.69%	**87.94%**
	AUC(target)	46.25%	53.61%	**72.67%**
	pAUC	52.42%	51.26%	**67.34%**
All(hmean)	AUC(source)	65.00%	65.77%	**74.89%**
	AUC(target)	50.28%	49.51%	**60.36%**
	pAUC	52.84%	52.28%	**56.71%**

3. CONCLUSION

In tis work, a two-step anomaly detection approach has been presented. We proposed an unsupervised attribute analysis method to generate pseudo-attribute labels of machines with no attribute information. Then we proposed SeResNet18 based classification and convolution based CAE ASD models. After ensemble, our system outperforms the baseline systems and achieves a competitive harmonic mean of AUC and PAUC (p = 0.1) of 63.09% on the development set.

4. REFERENCES

[1] Nishida, Tomoya and Harada, Noboru and Niizumi, Daisuke and Albertini, Davide and Sannino, Roberto and Pradolini, Simone and Augusti, Filippo and Imoto, Keisuke and Dohi, Kota and Purohit, Harsh and Endo, Takashi and Kawaguchi, Yohei. Description and Discussion on DCASE 2024 Challenge Task 2: First-Shot Unsupervised Anomalous Sound Detection for Machine Condition Monitoring. In arXiv e-prints: 2406.07250, 2024.

[2] Harada, Noboru and Niizumi, Daisuke and Takeuchi, Daiki and Ohishi, Yasunori and Yasuda, Masahiro and Saito, Shoichiro. ToyADMOS2: another dataset of miniature-machine operating sounds for anomalous sound detection under domain shift conditions. In Proceedings of the Detection and Classification of Acoustic Scenes and Events Workshop (DCASE), 1 – 5. Barcelona, Spain, November 2021.

[3] Dohi, Kota and Nishida, Tomoya and Purohit, Harsh and Tanabe, Ryo and Endo, Takashi and Yamamoto, Masaaki and Nikaido, Yuki and Kawaguchi, Yohei. MIMII DG: Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection for Domain Generalization Task. In Proceedings of the 7th Detection and Classification of Acoustic Scenes and Events 2022 Workshop (DCASE2022). Nancy, France, November 2022.

[4] Harada, Noboru and Niizumi, Daisuke and Ohishi, Yasunori and Takeuchi, Daiki and Yasuda, Masahiro. First-Shot Anomaly Sound Detection for Machine Condition Monitoring: A Domain Generalization Baseline. In 2023 31st European Signal Processing Conference (EUSIPCO), 191-195. 2023.

[5] A. B. Smith, C. D. Jones, and E. F. Roberts, “A sample paper in journals,” *IEEE Trans. Signal Process.*, vol. 62, pp. 291-294, Jan. 2000.

[6] Mc Queen, John. "Some methods for classification and analysis of multivariate observations." Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, 1967, pp. 281-297.

[7] Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, and David Lopez-Paz, ″ Mixup: Beyond empirical risk minimization,″ in ICLR, 2018.

[8] Han Chen, Yan Song, Zhu Zhuo, Yu Zhou, Yu-Hong Li, Hui Xue, and Ian McLoughlin, ″An effective anomalous sound detection method based on representation learning with simulated anomalies,″ in ICASSP. IEEE, 2023.

[9] J. F. Gemmeke, D. P. Ellis, D. Freedman, A. Jansen, W. Lawrence, R. C. Moore, M. Plakal, and M. Ritter, ″ AudioSet: An ontology and human-labeled dataset for audio events,″ in Proc. of the International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2017, pp. 776–780.

[10] Fernández A, Garcia S, Herrera F, et al. SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary[J]. Journal of artificial intelligence research, 2018, 61: 863-905.

[11] Wilkinghoff, Kevin. "Self-supervised learning for anomalous sound detection." ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2024.

[12] R. L. Sagi Omer, “Ensemble learning: A survey,” Wiley interdisciplinary reviews. Data mining and knowledge discovery, vol. 8, 2018.

[13] Hu, Jingcao, Liangzhi Shen, and Saad Albanie. "SeResNet: Deep Spatial Feature Reconstruction for Image Recognition." IEEE Transactions on Multimedia, vol. 24, no. 3, 2021, pp. 985-996.