UNSUPERVISED ABNORMAL SOUND DETECTION METHOD BASED ON CAUSAL SEPARATION

Technical Report

Zhang Yunxiang

Xinjiang University
School of Information Science and Engineering
Urumqi, 830046, China
zhangyunxiang@stu.xju.edu.cn

He Liang*

Xinjiang University
School of Information Science and Engineering
Urumqi, 830046, China
heliang@mail.tsinghua.edu.cn

ABSTRACT

Anomalous sound detection (ASD) is the task of identifying if a sound is normal or anomalous with respect to a given reference. In this report we present a solution for the DCASE2023 task 2 (First-Shot Unsupervised Anomalous Sound Detection for Machine Condition Monitoring), which aims to address the ASD problem under domain generalization and First-shot problem. We use the method consists of the classification method and stable learning. The proposed systems achieve in the harmonic mean of all machine types, sections, and domains for the area under the curve (AUC) and partial AUC (p = 0.1) on the development set.

Index Terms— first-shot anomaly sound detection, domain generalization, machine condition monitoring stable learning

1. INTRODUCTION

This report describes unsupervised anomalous sound detection (ASD) methods developed for the DCASE 2023 Challenge Task 2[1]. This task aims for machine condition monitoring and requires detecting unknown anomalous data using only normal data. In DCASE 2023 task, a new condition that the set of machine types in the development dataset and evaluation dataset are completely different and each machine type contains only one section (i.e., first-shot problem) is newly added.

This year, developing the dataset and evaluating the machine type set in the dataset are completely different. The methods used in previous years to train a model for each machine type are not available. Instead, we use all machine types to train a classifier to recognize exceptions. However, different types of machines have different audio characteristics. Therefore, we add the method of stable learning to the classification network, and use stable learning for generalization, so that the variables in the deep learning scene have strict independence, so that the classification network can extract the essential characteristics of data

We conduct an experimental evaluation of the developed system using the DCASE 2023 Task 2 Challenge development datasets. Here, the datasets [2, 3] consist of sounds from seven types of real/toy machines. The development dataset consists of normal/anomalous operating sounds of seven types of machines and each type of machine consists of one “section”. The additional training and evaluation datasets also consist of normal/anomalous operating sounds of machines, but the sets of machine types are completely different from the development dataset.

Experiments on the datasets show that all of the created systems significantly outperformed the official baseline system in the evaluation metric, the harmonic mean of the area under the curve (AUC), and partial AUC (p = 0.1) for all machines types, and section IDs (all / har-mean).

*Corresponding Author

2. PROPOSED METHOD

The machine types in the developing data set and the evaluating data set are completely different and each machine type contains only one part. In order to extract the essential characteristics of audio data, our model is divided into two parts: In the first part, we added the method of stable learning to solve the distribution shift problem by globally weighting the samples, so as to directly remove all the features of each input sample and eliminate the statistical correlation between the relevant features and irrelevant features. In the second part, the output prediction results of the classifier are reweighted with those of the first part to make different samples as independent as possible. MobileNetV2 model was adopted to extract Features for Random Fourier Features (RFF) mapping, sample weights were obtained to make cross covariance between different samples as small as possible, and the classifier obtained prediction results to calculate cross loss entropy, which was multiplied by weights to obtain final loss. The model structure is shown in Figure 1.

3. LOCAL OUTLIER FACTOR)

You are allowed a total of up to 5 pages for your DCASE 2022 Challenge technical report. The report can also be shorter and does not need to include a literature review. If however you plan to submit the same report as a regular paper to DCASE 2022 Workshop, please structure it as a scientific paper and respect the rules given for the workshop paper formatting. For the workshop submission you will also have to use the workshop paper template, which has different header.

4. DATASET

We use Audioset [4], a large-scale unsupervised pre-training dataset provided by Google in 2017, for pre-training. This dataset contains 632 audio categories and 2,084320 manually labeled sound clips (including 527 labels) with each 10-second length. Tags cover human voice, environment sound, machine sound and other fields. This dataset has been used for many times in large-scale unsupervised audio pre-training tasks [5, 6] to evaluate model effects in multiple audio tasks. The dataset for our detection is the official dataset published in the Task 2 of DCASE 2020 and DCASE 2022, which are both composed of two-part datasets [7]. The data of DCASE 2022 consists of the normal/abnormal operating sounds of seven types of machines. Each recording is a single-channel 10-sec length audio clip that includes both the sounds of the target machine and environmental sounds, each machine type provides data in multiple different domains for training and validation in the training set and validation set. Of course, the domain to which the test data belongs is unseen. The data of DCASE 2020 consists of the normal/abnormal operating sounds of six types of toy/real machines, the other parameters are the same as the dataset of DCASE 2022 Task 2, and they all provide two parts: training set and validation set.

5. REFERENCES

[1] N. Harada, D. Niizumi, D. Takeuchi, Y. Ohishi, and M. Yasuda, “First-shot anomaly detection for machine condition monitoring: A domain generalization baseline,” In arXiv e-prints: 2303.00455, 2023.

[2] 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 ggeneralization task,” in Proceedings of the 7th Detection and Classification of Acoustic Scenes and Events 2022 Workshop (DCASE2022), Nancy, France, November 2022.

[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,” in Proceedings of the Detection and Classification of Acoustic Scenes and Events Workshop (DCASE), Barcelona, Spain, November 2021, pp. 1–5.

[4] 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 ICASSP. IEEE, 2017, pp. 776–780.

[5] D. Niizumi, D. Takeuchi, Y. Ohishi, N. Harada, and K. Kashino, “Byol for audio: Self-supervised learning for general-purpose audio representation,” in IJCNN. IEEE, 2021, pp. 1–8.

[6] A. Saeed, D. Grangier, and N. Zeghidour, “Contrastive learning of general-purpose audio representations,” in ICASSP. IEEE, 2021, pp. 3875–3879.

[7] K. Dohi, K. Imoto, N. Harada, D. Niizumi, Y. Koizumi, T. Nishida, H. Purohit, T. Endo, M. Yamamoto, and Y. Kawaguchi, “Description and discussion on DCASE 2022 challenge task 2: Unsupervised anomalous sound detection for machine condition monitoring applying domain generalization techniques,” In arXiv e-prints: 2206.05876, 2022.