DCASE 2020 TASK 2: UNSUPERVISED DETECTION OF ANOMALOUS SOUNDS FOR MACHINE CONDITION MONITORING

Technical Report

Duc H. Phan *

University of Illinois at Urbana-Champaign
Illinois, USA
ducphan2@illinois.edu

Douglas L. Jones

University of Illinois at Urbana-Champaign
Illinois, USA
dl-jones@illinois.edu

ABSTRACT

A multiple layer neural network predictor is proposed for anomalous sound detection instead of a traditional auto-encoder approach. The network operates on the log-mel-spectrogram, predicting the log-mel feature vector given the previous and future feature vectors. The prediction error is used as the anomaly score measure. The proposed system outperforms the baseline system [1] on Detection and Classification of Acoustic Scenes and Events 2020 (DCASE2020) Task 2 development data set [2, 3].

Index Terms— Predictor Network, anomaly detection, unsupervised learning, machinery monitoring

1. INTRODUCTION

Anomalous sound detection (ASD) tries to identify abnormal sound emitted from a target machine. The main challenge of the task is that anomalies are unknown because only normal sound samples are given for training. Particularly, DCASE 2020 Task 2 provides data from ToyADMOS [2] and MIMII Dataset [3] including normal and anomalous sound clips of six types of toy and real machines. Only normal sound instances are available for training.

In the age of deep learning, autoencoder (AE) is a standard technique for ASD. In an AE-based anomaly detection system, the anomaly score is the reconstruction error of the observed sound. During the training phase, only normal sound clips are presented, and the goal of AE is minimizing the reconstruction error over the training set. Hence, AE assumes that it cannot reconstruct anomalies as well as normal singals. However, if target sounds are immersed in background noises, AE could try to minimize the background in the training set. As a consequence, it may produce a very noisy anomaly score in the prediction phase.

Machines are engineer-designed for consistency. In a normal operating mode, they should produce predictable sound patterns. Based on these assumptions, we proposed a predictor network in which the network tries to predict log mel-band energy feature vectors at a given time from nearby past and future feature vectors. The predictor error will be treated as an anomaly score; the normal instances are expected to be more predictable with smaller anomaly scores. In this report, we studied a simple predictor network on the development data set. The rest of this report is organized as follows: First, we briefly describe the development data set before introducing our proposed model. Next, the performance of the proposed method against the baseline autoencoder system is shown,followed by conclusion.

*This work utilizes resources supported by the National Science Foundation's Major Research Instrumentation program, grant #1725729, as well as the University of Illinois at Urbana-Champaign.

2. DATA SET AND PREPROCESSING STEP

Each recording in the dataset is a single-channel 10-sec long audio file of the target operating machine's sound and environmental noise. There are six types of machines: toy car, toy conveyor, valve, pump, fan and slide rail; the first two are from ToyADMOS while the rest are from MIMII Dataset. The sampling rate of test and training samples is 16 kHz. Each machine type has three or four machine IDs in the development data set. Each machine ID comprises approximately 1000 samples of normal instances for training and 100-200 samples each of normal and anomalous recordings for test.

In the data preprocessing step, we convert raw audio recordings into log mel-band energy feature vectors with 64 bands; a 1024-bin fast Fourier transform (FFT) with hop size of 512 is used. Each audio recording is converted to a 64×313 frequency-time matrix. The matrix is then split into overlapped windows of size 128×7 as inputs into the proposed predictor network.

3. MODEL AND TRAINING

In the proposed network, each recording is transformed into a log-mel-spectrogram $X \in \mathbf{R}^{F \times T}$ where F is the number of mel filters and T is the number of time frames. Given the frames at time t, and P frames before and after time t, the anomaly score for a given clip is defined to be

$$A_\theta = \frac{1}{T-2P} \sum_{P+1}^{T-P-1} ||X_t - f(\phi_t)||_2^2 \tag{1}$$

where $\phi_t = (X_{t-P}, \ldots, X_{t-1}, X_{t+1}, \ldots, X_{t+P})$, $f_\theta(\cdot)$ is the predictor network, and θ is the hyper-parameter set of the system.

The proposed predictor network has one fully-connected network (FCN) layer with 256 units as the input layer followed by three hidden FCN layers and one FCN output layer. Each hidden layer contains 256 units, and the output dimension is 64. Batch normalization [4] followed by Rectified linear unit (ReLU) and dropout [5] is used after FCN except for the output layer. The network was trained for 70 epochs with a batch size of 512. The ADAM optimizer [6] was used with a learning rate of 0.0001. The training model that has the smallest prediction error over the validation set is selected. Our validation set is 10 percent of the training set. Note that each machine ID has its own set of trained weights, and the dropout rate is set at 0.2. The performance of the proposed model is presented in the next section.We submit outputs from independent training to DCASE 2020 Task 2.

4. PERFORMANCE ON DEVELOPMENT DATA SET

The area under the receiver operating characteristic (ROC) curve (AUC) [1] and the partial-AUC (pAUC) [1] are performance metrics for this tasks. For comparison, we train our proposed system 10 times independently and report the averages and standard deviations of the measurements. The performances of the baseline system and the proposed predictor network are given in Tables 1 and 2 respectively. Overall, our system outperforms the baseline in all machine types except toy car.

5. CONCLUSION

The proposed predictor network is a very simple multi-layer FCN network, it actually mimics the encoder part of the baseline system. However, the network shows significant performance improvement from the baseline. Among machine types, valve and fan have very high jumps in performance above the baseline with increases of 23.22% and 8.5% on AUC metric, respectively. The results of the proposed network suggests that predictor network potentially becomes a powerful tool for machinery sound anomaly detection.

6. REFERENCES

[1] Y. Koizumi, Y. Kawaguchi, K. Imoto, T. Nakamura, Y. Nikaido, R. Tanabe, H. Purohit, K. Suefusa, T. Endo, M. Yasuda, and N. Harada, “Description and discussion on DCASE2020 challenge task2: Unsupervised anomalous sound detection for machine condition monitoring,” in *arXiv e-prints: 2006.05822*, June 2020, pp. 1–4. [Online]. Available: https://arxiv.org/abs/2006.05822

[2] Y. Koizumi, S. Saito, H. Uematsu, N. Harada, and K. Imoto, “ToyADMOS: A dataset of miniature-machine operating sounds for anomalous sound detection,” in *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, November 2019, pp. 308–312. [Online]. Available: https://ieeexplore.ieee.org/document/8937164

[3] H. Purohit, R. Tanabe, T. Ichige, T. Endo, Y. Nikaido, K. Suefusa, and Y. Kawaguchi, “MIMII Dataset: Sound dataset for malfunctioning industrial machine investigation and inspection,” in *Proceedings of the Detection and Classification of Acoustic Scenes and Events 2019 Workshop (DCASE2019)*, November 2019, pp. 209–213. [Online]. Available: http://dcase.community/documents/workshop2019/proceedings/DCASE2019Workshop_Purohit_21.pdf

[4] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” *arXiv preprint arXiv:1502.03167*, 2015.

[5] P. Baldi and P. J. Sadowski, “Understanding dropout,” in *Advances in neural information processing systems*, 2013, pp. 2814–2822.

Toy Car		
Machine ID	AUC(Ave.± Std.)	pAUC(Ave.± Std)
1	81.36 ± 1.15	68.40 ± 0.92
2	85.97 ± 0.58	77.72 ± 0.90
3	63.30 ± 1.03	55.21 ± 0.37
4	84.45 ± 1.87	68.97 ± 2.37
Average	78.77 ± 1.03	67.58 ± 1.04
Toy Conveyor		
Machine ID	AUC(Ave.± Std.)	pAUC(Ave.± Std)
1	78.07 ± 0.79	64.25 ± 0.99
2	64.16 ± 0.53	56.01 ± 0.71
3	75.35 ± 1.39	61.03 1 ± 1.00
Average	72.53 ± 0.67	60.43 ± 0.74
Fan Machine		
Machine ID	AUC(Ave.± Std.)	pAUC(Ave.± Std)
0	54.41 ± 0.47	49.37 ± 0.10
2	73.40 ± 0.58	54.81 ± 0.34
4	61.61 ± 1.08	53.26 1 ± 0.40
6	73.92 ± 0.54	52.35 1 ± 0.51
Average	65.83 ± 0.53	52.45 ± 0.21
Pump Machine		
Machine ID	AUC(Ave.± Std.)	pAUC(Ave.± Std)
0	67.15 ± 0.87	56.74 ± 0.82
2	61.53 ± 0.97	58.10 ± 0.93
4	88.33 ± 0.66	67.10 1 ± 1.09
6	74.55 ± 1.24	58.02 1 ± 1.21
Average	72.89 ± 0.70	59.99 ± 0.77
Slider Machine		
Machine ID	AUC(Ave.± Std.)	pAUC(Ave.± Std)
0	96.19 ± 0.43	81.44 ± 1.89
2	78.97 ± 0.28	63.68 ± 0.72
4	94.30 ± 0.64	71.98 1 ± 2.20
6	69.59 ± 1.45	49.21 ± 0.41
Average	84.76 ± 0.29	66.53 ± 0.62
Valve Machine		
Machine ID	AUC(Ave.± Std.)	pAUC(Ave.± Std)
0	68.76 ± 0.65	51.70 ± 0.19
2	68.18 ± 0.86	51.83 ± 0.31
4	74.30 ± 0.71	51.97 1 ± 0.20
6	53.90 ± 0.38	48.43 ± 0.20
Average	66.28 ± 0.49	50.98 ± 0.15

Table 1: Baseline system results; values in percentage

[6] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

Toy Car		
Machine ID	AUC(Ave.± Std.)	pAUC(Ave.± Std)
01	79.91 ± 0.33	71.72 ± 0.79
02	81.81 ± 0.67	77.54 ± 0.65
03	62.81 ± 0.98	52.8 ± 0.38
04	75.50 ± 0.95	65.58 ± 0.81
Average	75.01 ± 0.46	66.91 ± 0.26
Toy Conveyor		
Machine ID	AUC(Ave.± Std.)	pAUC(Ave.± Std)
01	77.21 ± 0.47	63.68 ± 0.35
02	66.25 ± 0.61	55.67 ± 0.33
03	77.67 ± 0.56	59.48 ± 0.22
Average	73.71 ± 0.35	59.61 ± 0.11
Fan Machine		
Machine ID	AUC(Ave.± Std.)	pAUC(Ave.± Std)
00	57.05 ± 0.64	49.61 ± 0.21
02	83.72 ± 0.44	65.24 ± 0.71
04	66.30 ± 0.8	54.46 ± 0.53
06	90.26 ± 0.44	67.62 ± 1.00
Average	74.33 ± 0.43	59.23 ± 0.21
Pump Machine		
Machine ID	AUC(Ave.± Std.)	pAUC(Ave.± Std)
00	63.03 ± 1.0	52.64 ± 0.80
02	55.90 ± 0.77	56.81 ± 0.90
04	97.73 ± 0.34	88.89 ± 1.50
06	78.18 ± 0.38	60.54 ± 0.52
Average	73.71 ± 0.36	64.72 ± 0.59
Slider Machine		
Machine ID	AUC(Ave.± Std.)	pAUC(Ave.± Std)
00	97.56 ± 0.22	87.44 ± 1.08
02	84.83 ± 0.22	65.46 ± 1.04
04	94.77 ± 0.35	72.51 ± 1.77
06	79.27 ± 0.66	52.31 ± 0.56
Average	89.11 ± 0.26	69.43 ± 0.62
Valve Machine		
Machine ID	AUC(Ave.± Std.)	pAUC(Ave.± Std)
00	96.59 ± 0.31	82.72 ± 1.37
02	87.34 ± 0.77	59.92 ± 1.15
04	92.49 ± 0.42	67.71 ± 0.91
06	81.59 ± 1.16	56.34 ± 1.14
Average	89.50 ± 0.37	66.67 ± 0.59

Table 2: The proposed system results; values in percentage