UNSUPERVISED ANOMALOUS DETECTION BASED ON RIEMANNIAN GEOMETRY

Technical Report

Or Cohen, Yahav Vinokur, Asaf Arad, Dolev Vaknin, Shahaf-Yaron Peleg, Alon Amar

Technion, Israel Institute of Technology,
The Andrew and Erna Viterbi Faculty of Electrical and Computer Engineering
Haifa, Israel
{or.cohen, yahavvinokur, asaf.arad, dolev-vaknin}@campus.technion.ac.il

ABSTRACT

This technical report presents our proposed algorithms for the task 2 of the DCASE2022 challenge, which is unsupervised anomalous sound detection for machine condition monitoring by applying domain generalization techniques. We suggest two methods for feature extraction. The first method is based on extracting features using the latent space of an Autoencoder, and the second method is based on using the Mel-frequency cepstral coefficients (MFCC) to represent the signal. We represent the features using symmetric positive-definite (SPD) matrices. As there maybe a domain shift between the train data and the test data, we first performed spectral clustering given the Riemannian distances between the SPD matrices. A one class SVM is then trained on each of the centers of the clusters and is used to detect the anomalies in the data.

Index Terms— Unsupervised Anomaly Detection, Autoencoder feature extraction, Spectral feature extraction, Machine Condition Monitoring, Riemannian geometry, One-class SVM.

1. INTRODUCTION

In this technical report we consider task 2 of the DCASE 2022 challenge which is about *unsupervised anomalous sound detection for machine condition monitoring applying domain generalization techniques*.

Anomalous detection in audio signals is an important task in several applications. Many factories and craftsman are operating machinery on daily basis. When a fault occurs, the sound that the machinery emit might change. This attribute can be processed and exploited to report the fault. The costs of such acoustic anomalous detection is very low, and does not necessarily require heavy or expensive equipment. The main challenge in this task is the fact that most of the machinery anomalous behavior have not been recorded, which makes this task unsupervised - detection of an anomaly, when no anomalies are in the training data-set. Moreover, most of the machinery can operate in several modes.

In this task we approached a problem in which there where many recordings of one operating mode (source domain) and a small amount of another (target domain) in the training set, whereas at the test set, the probability that a fault occurs is equal between the domains. So, a data-based solution must lean on the major domain, but also be able to give good results on the smaller domain data. This challenge is called domain-shift and it suggests that the attributes to solve one problem can be shifted to solve the other.

We consider two types of feature extraction. The first is based on Autoencoders (AE) which are commonly used in deep learning methods. AE's are able to learn the latent features need to obtain a reliable reconstruction of the data. The second is based on Mel-scale frequency cepstral coefficients, which are usually used in as features to represent an audio signal, especially in audio signals containing poor spectral information.

The domain-shift problem lead us to tackle it while combining classical methods in our approach which can be easily transformed between domains, without many data samples needed. The next step was processing the features using Riemannian geometry, as described in [1]. Lastly we detect anomalies on the processed features using one-class support vector machine (OC-SVM) as suggested by [4].

In this report we will describe our methods in details in section 2. We will present the results on the development data-set in section 3, and our conclusion will be in section 4.

2. PROPOSED APPROACHES

Our approaches are composed from three parts: feature extracting, feature processing, and anomaly detection.

2.1. Feature extraction

We used two different methods for feature extraction: the latent layer of an AE, and cepstral coefficients.

Extracting AE features is performed by training a simple fully connected AE. The AE is trained to minimize the Itakura-Saito spectral distance [2] between five, concatenated in the time domain log-Mel-Spectrum, frames. The AE latent space is one $z \in R^{1\times 8}$ vector for five concatenated frames, and overall, we result in $N \times 8$ matrix output for N segments of five frames in a recording. The Autoencoder architecture is as described in table 1. Also, the AE training hyper-parameters are as described in table 2.

We used the Itakura-Saito spectral distance [2] as loss function between the input log-Mel-Spectrum to the AE output. The distance is defined as

$$\mathrm{IS}_{\text{loss}} = \int_{-\pi}^{\pi} \frac{P(\omega)}{\hat{P}(\omega)} - \log\left(\frac{P(\omega)}{\hat{P}(\omega)}\right) - 1 \; d\omega \qquad (1)$$

Where P represents the input spectrum, $\hat{P}$ represents the reconstructed spectrum, and ω denotes the angular frequency.

The cepstral coefficients are extracted as follows. The data is recorded with sampling frequency $f_s = 16[kHz]$, and it was split into overlapping segments using window with length of $0.03 \cdot f_s$,

Layer name	Output Shape	Parameters
input 1	[(None, 640)]	0
dense 0	(None, 128)	82048
batch normalization 0	(None, 128)	512
activation 0	(None, 128)	0
dense 1	(None, 128)	16512
batch normalization 1	(None, 128)	512
activation 1	(None, 128)	0
dense 2	(None, 128)	16512
batch normalization 2	(None, 128)	512
activation 2	(None, 128)	0
dense 3	(None, 128)	16512
batch normalization 3	(None, 128)	512
activation 3	(None, 128)	0
dense 4	(None, 8)	1032
batch normalization 4	(None, 8)	32
activation 4	(None, 8)	0
dense 5	(None, 128)	1152
batch normalization 5	(None, 128)	512
activation 5	(None, 128)	0
dense 6	(None, 128)	16512
batch normalization 6	(None, 128)	512
activation 6	(None, 128)	0
dense 7	(None, 128)	16512
batch normalization 7	(None, 128)	512
activation 7	(None, 128)	0
dense 8 (Dense)	(None, 128)	16512
batch normalization 8	(None, 128)	512
activation 8	(None, 128)	0
dense 9	(None, 640)	82560

Table 1: AE architecture details

parameter name	value
batch size	512
max epochs	50
learning rate	0.001
early stop	true
patience (early stop)	5

Table 2: AE training hyper-parameters

and overlap length of $0.02 \cdot f_s$. The number of coefficients extracted from the mel filter bank is 40.

2.2. Feature processing

We utilized a Riemannian distance measure, which is defined by

$$d_R(C_1, C_2) = \|\text{logm}(C_2^{-\frac{1}{2}} C_1 C_2^{-\frac{1}{2}})\|_F^2 \tag{2}$$

where logm is the matrix logarithm operator, and C_1, C_2 are symmetric and positive definite matrices (SPD). We transformed our input features to SPD matrices by eq. (3), to form an SPD matrix. Let z_i be the latent vector of the AE (deep method), or the MFC coefficients (classical method) associated with the ith segment. Let $\mathbf{Z} = [\mathbf{z_1}, \mathbf{z_2}, \ldots, \mathbf{z_N}]$ where N is the number of signal segments. We then define the SPD matrix as

$$C = \frac{1}{N}\mathbf{Z}^T\mathbf{Z} \tag{3}$$

We calculate the Riemannian mean, $\bar{C}$, of all the SPD matrices in the training data. This mean is a matrix that minimizes the Riemannian distance from all samples matrices space $C_i \in C_{samples}$ and is given by solving the following optimization problem,

$$\bar{C} = \underset{\bar{C}}{\text{argmin}} \sum_{C_i} d_R(\bar{C}, C_i) \tag{4}$$

We then compute the Riemannian distance matrix between each of the SPD matrices to the mean matrix $\bar{C}$ as

$$S_i = log(\bar{C}^{-\frac{1}{2}} C_i \bar{C}^{-\frac{1}{2}}) \tag{5}$$

The feature vector for sample i will be: $s_i = vec(S_i)$, where $vec(S_i)$ is the column stack vector representation of the upper triangle of the symmetric matrix S_i.

In order to adjust our model to the domain-shift problem, we computed several Riemannian means. These means where computed on clusters of feature matrices that where constructed by spectral clustering [3]. The affinity metric is defined by the Riemannian distances. Then, each sample matrix S_i is calculated using (5), and its nearest cluster mean is obtained using (4).

2.3. Anomaly detection

An anomaly score is assigned to each data sample using a one-class SVM model with a RBF kernel. The resulting support vectors separate the data from the origin in the high-dimensional kernel space. Once the SVM is trained, the anomaly score of a new incoming data point (an SPD matrix) C_i is defined as the distance of its respective feature s_i from the trained separating hyperplane.

3. RESULTS

The development data-set results are described in Table 3 (using the AE features), and in Table 4 (using the MFCC features). The system based on AE feature extraction is assigned in submission 1, and the one based on MFCC feature extraction is assigned in submission 2.

machine	$section$	$AUC(source)$	$AUC(target)$	$pAUC$
bearing	00	0.6229	0.4949	0.5075
	01	0.7376	0.5273	0.5386
	02	0.5426	0.5906	0.5678
fan	00	0.5352	0.3785	0.4997
	01	0.6660	0.4764	0.5000
	02	0.8986	0.4945	0.4997
gearbox	00	0.6720	0.5952	0.5710
	01	0.7270	0.6474	0.5426
	02	0.7744	0.5908	0.5668
slider	00	0.9460	0.7550	0.6689
	01	0.8642	0.8004	0.5931
	02	0.8380	0.6928	0.6436
ToyCar	00	0.7528	0.3534	0.5026
	01	0.7048	0.5434	0.5084
	02	0.9846	0.6822	0.6589
ToyTrain	00	0.7122	0.2518	0.5000
	01	0.8474	0.3152	0.5015
	02	0.8486	0.4260	0.4878
valve	00	0.5116	0.4934	0.5126
	01	0.5354	0.5392	0.5042
	02	0.6198	0.5788	0.4978

Table 3: development dataset results using AE

machine	$section$	$AUC(source)$	$AUC(target)$	$pAUC$
bearing	00	0.5497	0.7430	0.5148
	01	0.5184	0.7573	0.5071
	02	0.4759	0.5672	0.5077
fan	00	0.6418	0.7102	0.5831
	01	0.6898	0.4116	0.5200
	02	0.8782	0.5596	0.5531
gearbox	00	0.8180	0.6972	0.6010
	01	0.6874	0.5886	0.5237
	02	0.7462	0.6674	0.5889
slider	00	0.9264	0.7464	0.6084
	01	0.8524	0.7644	0.5578
	02	0.8268	0.7387	0.5995
ToyCar	00	0.8940	0.5756	0.5142
	01	0.8108	0.6462	0.5142
	02	0.9942	0.7960	0.6479
ToyTrain	00	0.7148	0.2710	0.4942
	01	0.8408	0.2930	0.5000
	02	0.8570	0.4424	0.5026
valve	00	0.5544	0.5248	0.5078
	01	0.5658	0.5676	0.5100
	02	0.5500	0.5072	0.4926

Table 4: development dataset results using MFCC

4. CONCLUSION

We presented an approach to detect anomalies in machinery recordings using Riemannian feature processing based on features extracted from training an AE, and based on the classical MFCC features. Further research may focus on determining the number of clusters in the spectral clustering step, as the affinity matrix that construct it is not sparse.

5. ACKNOWLEDGMENT

We would like to thank Prof. Ronen Talmon for his valuable and helpful suggestions during the development of the algorithms.

6. REFERENCES

[1] P. Lifshits and R. Talmon, ”Unsupervised Acoustic Condition Monitoring with Riemannian Geometry,” 2020 IEEE 30th International Workshop on Machine Learning for Signal Processing (MLSP), 2020, pp. 1-6.

[2] P. Chu and D. Messerschmitt, ”A frequency weighted Itakura-Saito spectral distance measure,” in IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 30, no. 4, pp. 545-560, August 1982.

[3] Shi, J., and J. Malik. “Normalized cuts and image segmentation.” IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 22, 2000, pp. 888–905.

[4] Li, Kun-Lun, et al. ”Improving one-class SVM for anomaly detection.” Proceedings of the 2003 international conference on machine learning and cybernetics (IEEE Cat. No. 03EX693). Vol. 5. IEEE, 2003.

[5] Kota Dohi, Tomoya Nishida, Harsh Purohit, Ryo Tanabe, Takashi Endo, Masaaki Yamamoto, Yuki Nikaido, and Yohei Kawaguchi. MIMII DG: sound dataset for malfunctioning industrial machine investigation and inspection for domain generalization task. In arXiv e-prints: 2205.13879, 2022.

[6] Noboru Harada, Daisuke Niizumi, Daiki Takeuchi, Yasunori Ohishi, Masahiro Yasuda, and Shoichiro Saito. ToyADMOS2: another dataset of miniature-machine operating sounds for anomalous sound detection under domain shift conditions. In Proceedings of the 6th Detection and Classification of Acoustic Scenes and Events 2021 Workshop (DCASE2021), 1–5. Barcelona, Spain, November 2021.

[7] Kota Dohi, Keisuke Imoto, Noboru Harada, Daisuke Niizumi, Yuma Koizumi, Tomoya Nishida, Harsh Purohit, Takashi Endo, Masaaki Yamamoto, and Yohei 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.