ANOMALY CALCULATION FOR EACH COMPONENTS OF SOUND DATA AND ITS INTEGRATION FOR DCASE 2020 CHALLENGE TASK2

Technical Report

Yuya Sakamoto and Naoya Miyamoto

Fixstars Corporation
1-11-1 Ohsaki Sinagawa-ku, Tokyo 141-0032, Japan
{yuya.sakamoto, naoya.miyamoto}@fixstars.com

ABSTRACT

This paper is a technical report on our method that we submitted to the DCASE2020 Challenge Task 2. In our method, we first convert one sample into a log-mel-spectrogram, as in the baseline system. Next, the log-mel-spectrogram is decomposed into mean component, basis component and latent component by principal component analysis, and anomaly score is calculated for these each components. Then, the final anomaly score was determined by integrating the calculated anomaly score of each components. Each anomal score is calculated using Mahalanobis distance, k-nearest neighbor based on subspace distance, and distance based on matrix normal distribution.

Index Terms— Matrix Decomposition, Subspace Distance, Matrix Normal Distribution, Integration of anomaly

1. INTRODUCTION

DCASE2020 Challenge Task 2 is a task to find abnormal data in the test data set when only normal data is given as training data [1], [2], [3]. In order to find the abnormal data under such a situation, it is necessary to detect that the data is in some different state from the normal data. Therefore, we decompose the log-mel-spectrogram of the input signal into several components based on the principal component analysis, and calculate the distance between each element and the normal state as the degree of anormality. Finally, the anormaly score is integrated and the final anormaly score is caliculated. In addition, all the methods described here learn each Machine Type and ID individually, and do not use the information of other Machine Type and ID and other datasets. In the following, we describe the details of the method adopted in this study. In section 2, We describe the transformation applied to each time series signal and the mathematical expression after the transformation. In section 3 we describe a specific method for calculating anomalies. Since we are allowed to submit up to four different systems, we describe a variation of the system we have actually submitted in section 4.

2. SIGNAL CONVERSION

In this method, each sample given as a time series signal is converted into a log-mel-spectrogram in the same way as the baseline system [3]. log-mel-spectrogram can be expressed in a matrix as $\mathbf{X} \in \mathbb{R}^{F \times T}$. where F is the number of mel-filters and T is the number of time-frames. Also, the column vector of the t th column of the $\mathbf{X}$ is defined as $\mathbf{x}(t) \in \mathbb{R}^F$. Thus, $\mathbf{X} = [\mathbf{x}(1) \cdots \mathbf{x}(T)]$.

Then, the training data are N_{train} samples and the test data are N_{test} samples, the training and test datasets are $\{\mathbf{X}_1, \cdots, \mathbf{X}_{N_{train}}\}$, $\{\mathbf{X}_1, \cdots, \mathbf{X}_{N_{test}}\}$. Unlike the baseline system, this method does not combine multiple frames to form one vector.

3. CALCULATION OF ANOMALY SCORE

This method takes the following steps to calculate the anormality score.

1. Decomposition of the log-mel-spectrogram matrix.
2. Calculate the degree of anomaly for each component.
3. Calculate the final anomaly score by integrating the anomalies.

We describe a deteil of method for each items.

3.1. Decompoosition of the log-mel-spectrogram

If we consider the log-mel-spectrogram matrix $\mathbf{X}_i$ in the sample i as a set of T column vectors, each column vector is reduced to the $f < F$ dimension using principal component analysis and $\mathbf{X}_i$ is approximated as follows [4].

$$\mathbf{x}_i(t) \sim \mathbf{W}_i \mathbf{z}_i(t) + \mathbf{m}_i \qquad (1)$$

where $\mathbf{m}_i \in \mathbb{R}^F$ is the mean vector of $\{\mathbf{x}_i(1), \cdots, \mathbf{x}_i(T)\}$(hereafter, this is called "mean component"). $W_i \in \mathbb{R}^{F \times f}$ is a matrix of orthonormal basis of low-dimensional subspaces spanned by column vectors (hereafter, this is the "basis component"). $\mathbf{z}_i(t) \in \mathbb{R}^f$ is the low-dimensional latent vector in the t th frame. Then, the matrix $\mathbf{Z}_i = [\mathbf{z}_i(1) \cdots \mathbf{z}_i(T)]$ which $\mathbf{z}_i(t)$ is arranged in chronological order as a column vector, is defined as "latent component". Thus, the log-mel-spectrogram matrix is divided into three components: the mean component, the basis component, and the latent component. Once broken down into components, these are used to calculate each of the following anomalies.

3.2. Anormalies for each component

3.2.1. Anormaly for mean component

The mean component $\mathbf{m}$ can be taken as the strength of each frequency component. The anomaly for the mean component is calculated by a simple Mahalanobis distance. Specifically, since the training dataset yields a set of mean component $\{\mathbf{m}_1, \cdots, \mathbf{m}_{N_{train}}\}$, and the average vector of this set is $\boldsymbol{\mu}$, the covariance matrix is $\boldsymbol{\Sigma}$,

then the mean component anomaly $a_m(\mathbf{X}_*)$ for sample $\mathbf{X}_*$ can be calculated as follows.

$$a_m(\mathbf{X}_*) = \sqrt{(\mathbf{m}_* - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{m}_* - \boldsymbol{\mu})} \qquad (2)$$

3.2.2. Anomaly for basis component

The anomaly for the basis component $\mathbf{W}$ is calculated with k-nearest-neighbor based on the distance between the subspace spanned by the basis component. The distance between the subspaces is based on the principal angle[5], [6], [7]. Specifically, the distance between the low-dimensional subspace of the sample $\mathbf{X}_i$ and the sample $\mathbf{X}_j$ is calculated as follows.

$$\text{SubspaceDist}(\mathbf{X}_i, \mathbf{X}_j) = 1 - \| \mathbf{W}_i^T \mathbf{W}_j \|_2^2 \qquad (3)$$

In (3), $\| \cdot \|_2$ is the $p = 2$ matrix norm and is equal to the maximum singular value of the matrix. Now that the distances between the subspaces using the basis components are defined, we use them to calculate the anomaly for the sample $\mathbf{X}_*$ in k-nearest-neighbor. From $\{\mathbf{X}_1, \cdots, \mathbf{X}_{N_{train}}\}$, we find the top k values for which $\text{SubspaceDist}(\mathbf{X}_*, \mathbf{X}_i)$ is smaller. Then, let the average of the distances be the basis component anomaly $a_b(\mathbf{X}_*)$. It is expressed as follows.

$$a_b(\mathbf{X}_*) = \frac{1}{k} \sum_{i \in K} \text{SubspaceDist}(\mathbf{X}_*, \mathbf{X}_i) \qquad (4)$$

K is the set of indices near k for $\mathbf{X}_*$.

3.2.3. Anomaly for latent component

Since $\mathbf{Z}$ is a matrix, the distance based on Matrix Normal Distribution is used [8]. The Matrix Normal Distribution is a probability density function for matrix data. Thus, probability density function for the latent component $\mathbf{Z}$ is as follow.

$$p(\mathbf{Z}|\mathbf{M}, \mathbf{U}, \mathbf{V}) = \frac{\exp\left(-\frac{1}{2}\text{tr}\left[\mathbf{V}^{-1}(\mathbf{Z} - \mathbf{M})^T \mathbf{U}^{-1}(\mathbf{Z} - \mathbf{M})\right]\right)}{(2\pi)^{fT/2}|\mathbf{V}|^{T/2}|\mathbf{U}|^{f/2}} \qquad (5)$$

$\mathbf{M} \in \mathbb{R}^{f \times T}$, $\mathbf{U} \in \mathbb{R}^{f \times f}$, $\mathbf{V} \in \mathbb{R}^{T \times T}$ $\mathbf{M}$ is the mean matrix, while $\mathbf{U}$, $\mathbf{V}$ can be considered as the covariance matrix in the direction of frequency and time, respectively. In our method, these are estimated from the latent component set $\{\mathbf{Z}_1, \cdots, \mathbf{Z}_{N_{train}}\}$ of the training data with maximum likelihood. The direction of the frequency is fixed to $\mathbf{U} = \mathbf{I}$ because the latent component is expected to be uncorrelated since the target is a latent component derived from the principal component analysis. Thereby, the maximum likelihood estimated solution $\hat{\mathbf{M}}$ of $\mathbf{M}$ and $\mathbf{V}$ is analytically obtained by the following formula.

$$\hat{\mathbf{M}} = \frac{1}{N_{train}} \sum_{i=1}^{N_{train}} \mathbf{Z}_i \qquad (6)$$

$$\hat{\mathbf{V}} = \frac{1}{f N_{train}} \sum_{i=1}^{N_{train}} (\mathbf{Z}_i - \hat{\mathbf{M}})^T (\mathbf{Z}_i - \hat{\mathbf{M}}) \qquad (7)$$

Using these maximum likelihood estimation, the anormaly of latent component $a_l(\mathbf{X}_*)$ for the sample $\mathbf{X}_*$ is calculated as follows.

$$a_l(\mathbf{X}_*) = \sqrt{\text{tr}\left(\hat{\mathbf{V}}^{-1}(\mathbf{Z}_* - \hat{\mathbf{M}})^T (\mathbf{Z}_* - \hat{\mathbf{M}})\right)} \qquad (8)$$

3.2.4. Anomaly for components that mixtured basis and latent components

In addition to the anomaly scores for each of the basic components described above, another anomaly score for a component in which the basis component and the latent component are mixed can be considered. Specifically, the distance based on Matrix Normal Distribution of $\mathbf{X}' = \mathbf{W}\mathbf{Z} = \mathbf{W}\mathbf{W}^T \mathbf{X}$ is the degree of anomaly. In this case, unlike the previous term, it is not appropriate to consider $\mathbf{U}$ in (5) as an identity matrix, so $\mathbf{U}$ is also estimated by maximum likelihood estimation. In such a case, the maximum likelihood solution cannot be analytically obtained. Therefor Fix one of $\mathbf{U}$ and $\mathbf{V}$ and get the optimal solution of the other. Moreover, this process is alternately repeated to obtain a solution. See [8] for details.

Let this maximum likelihood solution be $\hat{\mathbf{M}}'$, $\hat{\mathbf{U}}'$, $\hat{\mathbf{V}}'$, respectively. The anomalous score $a_{bl}(\mathbf{X}_*)$ of the basis and latent mixed components for the sample $\mathbf{X}_*$ is calculated as follows.

$$a_{bl}(\mathbf{X}_*) = \sqrt{\text{tr}\left(\hat{\mathbf{V}}'^{-1}(\mathbf{X}_*' - \hat{\mathbf{M}}')^T \hat{\mathbf{U}}'^{-1}(\mathbf{X}_*' - \hat{\mathbf{M}}')\right)} \qquad (9)$$

3.3. combination of anomalies

Four types of degree of anomaly were defined, one for each component and one for mixed components. The final anomaly score is obtained by combining some of these. If it is truly abnormal, this degree of anomaly also deviates from the tendency of the learning data set. In other words, considering a vector in which the calculated degree of anomaly are connected, it is considered that the abnormality becomes high when the anomaly vector of the test data is far from the distribution of the anomaly vector of the learning data. However, it is not considered as anomalous if it deviates from the distribution in the direction of decreasing anomaly. The specific procedure for integrating the anomaly is as follows.

1. Calculate each degree of anomaly for all learning data

2. Training data with all components of anomaly smaller than the median are retained (decimation)

3. Calculate anomaly of each component for test data

4. For each anomaly of test data, if it is smaller than the average value of the anomaly of the training data left in step 2, replace it with that average value.

5. Calculate the Mahalanobis distance from the decimated learning data set for the anomaly vector of the test data subjected to step 4, and use that as the final anomaly score.

Step 2 is a decimation process for reducing the influence of variations due to outliers included in the learning data set. Depending on the system submitted, this step may not be performed, as described below.

4. SUBMITTED SYSTEM VARIATION

Since 4 submissions are allowed for this task, We submitted 4 patterns that changed combination of anomaly ((2), (4), (8), (9)), with or without decimation, number of dimensions after reduction f, k-nearest neighbor method k. Specific combinations and hyperparameter settings are shown in Table 1.

Table 1: submit system list

No.	Abnormality combination	decimation	f	k
1	a_m , a_b , a_l	yes [1]	20	5
2	a_m , a_b , a_l	no	20	5
3	a_m , a_b	yes	20	5
4	a_m , a_{bl}	yes	1	-

5. CONCLUSION

We adopted the method of calculating the degree of anomaly for each component of log-mel-spectrogram and integrating them. This method is relatively simple then it work with few training time and small train data size. The inference time is also short, so it can be used for real-time anomaly detection systems, which makes it practical. Furthermore, combining with other methods can be considered (such as other preprocessing, other feature extraction, using external data and more), which may further improve accuracy.

6. REFERENCES

[1] 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

[2] 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

[3] 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

[4] C. M. Bishop, *Pattern recognition and machine learning.* springer, 2006.

[5] K. Fukui and O. Yamaguchi, "Face recognition using multi-viewpoint patterns for robot vision," in *Robotics Research. The Eleventh International Symposium.* Springer, 2005, pp. 192–201.

[6] T. Idé and K. Tsuda, "Change-point detection using krylov subspace learning," in *Proceedings of the 2007 SIAM International Conference on Data Mining.* SIAM, 2007, pp. 515–520.

[7] J. Hamm and D. D. Lee, "Grassmann discriminant analysis: a unifying view on subspace-based learning," in *Proceedings of the 25th international conference on Machine learning*, 2008, pp. 376–383.

[8] H. Glanz and L. Carvalho, "An expectation-maximization algorithm for the matrix normal distribution," *arXiv preprint arXiv:1309.6609*, 2013.

---

[1] We use only a_m and a_b for decide decimated sample because using all anomaly scores occur too many decimation.