XV International Conference on Durability of Building Materials and Components
DBMC 2020, Barcelona
C. Serrat, J.R. Casas and V. Gibert (Eds)

Development of a Damage Detecting Method for RC Slabs by Means of Machine Learning

Yutaka Tanaka and Takahiro Nishida

Port and Airport Research Institute, 3-1-1, Nagase, Yokosuka, 239-0826, JAPAN,
tanaka-yu@p.mpat.go.jp, nishida-ta@p.mpat.go.jp

Abstract. *It is beneficial to understand damage conditions of RC structural members by non-destructive methods. In this study, K-means clustering method was applied to the AE wave data in order to develop a damage detection method. From the result of this study, the damaged area had the relatively large number of AE hits and AE energy. In addition, from the result of the K-means clustering of the power spectral density of AE waves, the data could be separated into the obviously damaged area and other areas.*

Keywords: *AE, Non-Destructive Method, Machine Learning, Damage Detection, RC Slab.*

1 Introduction

It is beneficial to understand damage conditions of RC structural members by non-destructive methods for long term use of RC structures. Capturing AE signals is one of the effective methods to detect the damaged position and area in the structural member.

In the previous research, Nishida *et al*. (2019) developed an aggregation detector. The detector is equipped AE sensors to capture AE waves generated by the internal damage of the RC bridge deck. Nishida et al. focused on the AE energy in order to detect the damaged area and they proposed the method to decide the threshold value of AE energy to judge the damaged area.

In this study, authors applied a clustering method to the AE wave data that was obtained in previous research. The aim of this study is to develop a damage detecting method for RC slabs with machine learning.

2 Analysis of AE Waves

2.1 The Outline of Analysis of AE Waves

In this study, the AE wave data obtained in the previous research (Nishida *et al*., 2019) was analyzed. In the previous research, Nishida et al. focused on the total AE energy and did not focus on the number of AE hits. In general, the number of AE hits in damaged area is larger than that in sound area because the damaged area has a lot of AE source. Therefore, authors focused on not only the AE energy but also the number of AE hits. The AE wave data was separated into a group by 1.0 m for each measurement line, and then, the number of AE hits was counted in each group, and the average of AE energies was calculated in each group.

Authors also focused on the frequency domain of AE waves. The power spectral density (PSD) of each AE wave was calculated and averaged in each group. In this study, the averaged

PSDs were regarded as the representative values of each group and selected as the input data for the clustering method.

The K-means clustering method (MacQueen, 1967; Arthur and Vassilvitskii, 2007) was applied in this study. This method is a non-hierarchical clustering method and an unsupervised machine learning method. This method is used to divide automatically a data into k groups. The objective function of the K-means clustering that should be minimized is;

$$J = \sum_{i=1}^{N} \sum_{k=1}^{K} w_{ik} \|x_i - c_k\|_2 \tag{1}$$

where N is the number of a data set, K is the number of the cluster, x_i is the i-th data in a data set, c_k is the centroid of the k-th cluster, $w_{ik} = 1$ if the data x_i belongs to cluster k, otherwise $w_{ik} = 0$ and $\|v\|_2$ is the Euclidean norm of a vector v.

The algorithm of the K-means is shown as follows.

1. Choose k initial centroids $\boldsymbol{C} = \{c_1, ..., c_k\}$.
2. For each $i \in \{1, ..., k\}$, set the cluster C_i to be the set of points in X that are closer to c_i than they are to c_j for all $i \neq j$.
3. For each $i \in \{1, ..., k\}$, set c_i to be the center of mass of all points in C_i.
4. Repeat Steps 2 and 3 until $\boldsymbol{C}$ no longer changes.

In this study, the number of the cluster k was changed to be from 2 to 5 due to find the optimal k for the clustering of the data.

The calculation of the K-means clustering and the power spectral density was conducted with Python, and the Scikit-learn (Pedregosa *et al.*, 2011) was used for the K-means clustering and the SciPy (Virtanen *et al.*, 2019) was used to calculate the power spectral density of the AE wave data.

3 Results and Discussions

Figure 1 shows the damaged area that was the result of the visual and the hammering inspection in the previous research. The red colored area is the damaged area where the crack or the delamination were observed.

3.1 The Number of AE Hits and the Average of AE Energies

Figure 2 and Figure 3 show contour maps of the number of AE hits and the average of AE energies in each group respectively. In Figure 2, the number of AE hits may have a close relation with the damaged area as shown in Figure 1, and the average of AE energies also have a relation with the damaged area.

In general, the defect such as the crack and the delamination become the source of AE signal. Therefore, the area where the number of AE hits is larger than the other areas has potential to be the damaged area.

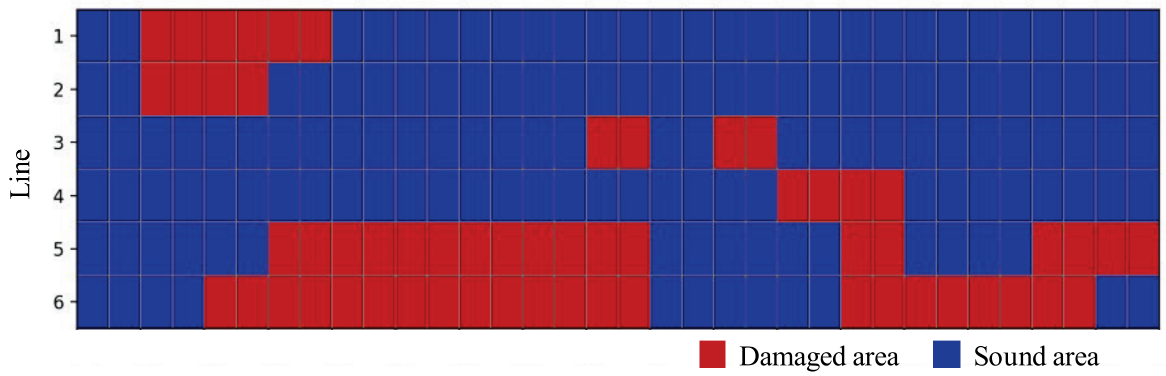


Figure 1. The result of the visual and the hammering inspection.

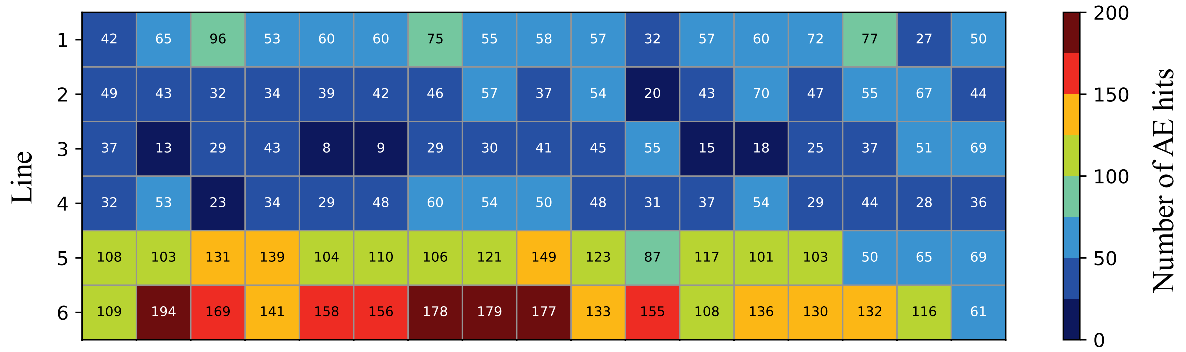


Figure 2. The number of AE hits in whole groups.

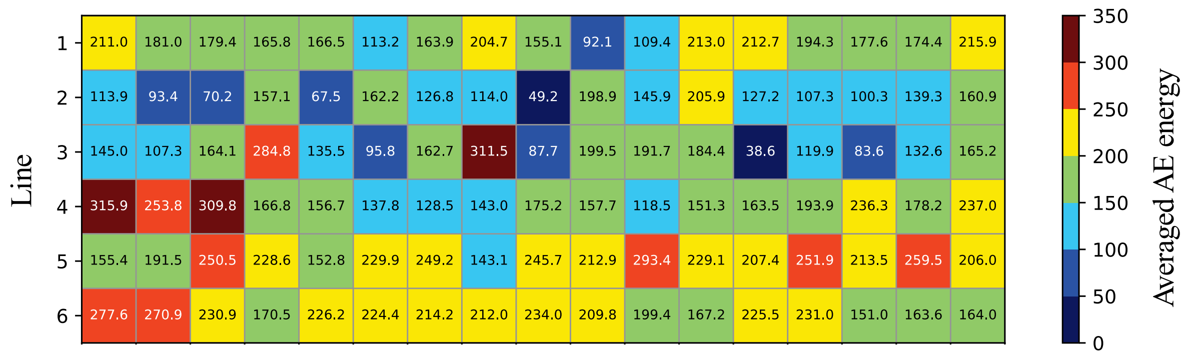


Figure 3. The average of AE energies in whole groups.

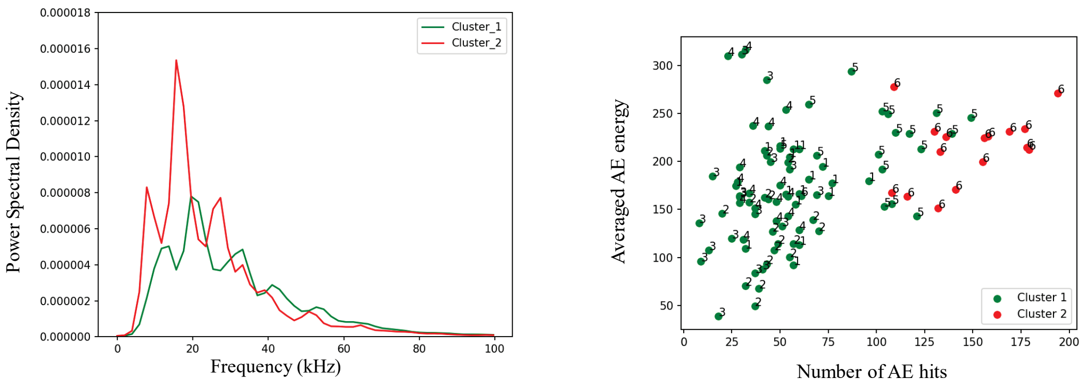
3.2 Clustering the PSD Data

Figure 4 (a) shows the PSD of the centroid of each cluster when the number of clusters was 2. The peak frequency and value of each PSD were different. Figure 4 (b) shows the scatter plot of each group. The number located beside each plot is the line number. In this figure, the all data belong to the cluster 2 was in the line 6 where a lot of defects were observed. From this result, the data was separated into the obviously damaged group and the other group, and the peak frequency of the PSD in the damaged area became lower than that in the other area.

Figure 5 (a) shows the PSD of the centroid of each cluster when the number of clusters was 3, and Figure 5 (b) shows the scatter plot of each group. From these figures, the cluster 1 in Figure 4 (b) was divided into two groups based on the average AE energy. In Figure 5 (a), the peaks of frequency and the forms of the PSD was also almost the same, but the peak values of the PSD were different in clusters 1 and 3.

Figure 6 (a) shows the PSD of the centroid of each cluster when the number of clusters was 4, and Figure 6 (b) shows the scatter plot of each group. From these figures, the cluster 1 in Figure 4 (b) was divided into three groups based on the average AE energy. In Figure 6 (a), the peaks of frequency and the forms of the PSD was also almost the same, but the peak values of the PSD were different in clusters 1, 3 and 4.

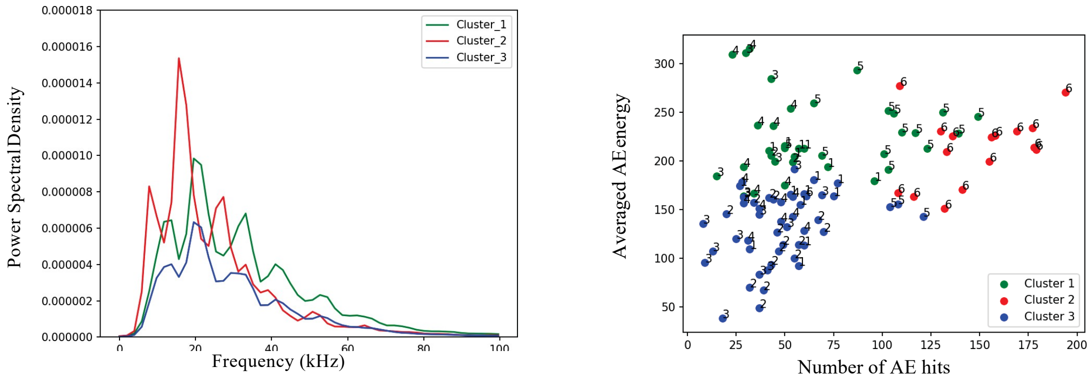
Figure 7 (a) shows the PSD of the centroid of each cluster when the number of clusters was 5, and Figure 7 (b) shows the scatter plot of each group. From these figures, the cluster 1 in Figure 4 (b) was divided into three groups and the cluster 2 in Figure 4 (b) divided into two groups based on the average AE energy. In Figure 7 (a), clusters 2, 3 and 4 were the same as clusters 2, 4, and 3 in Figure 6 (a) respectively. In addition, the peaks of frequency and the forms of the PSD was also almost the same, but the peak values of the PSD were different in clusters 1 and 5.



(a) PSD of each cluster centroid

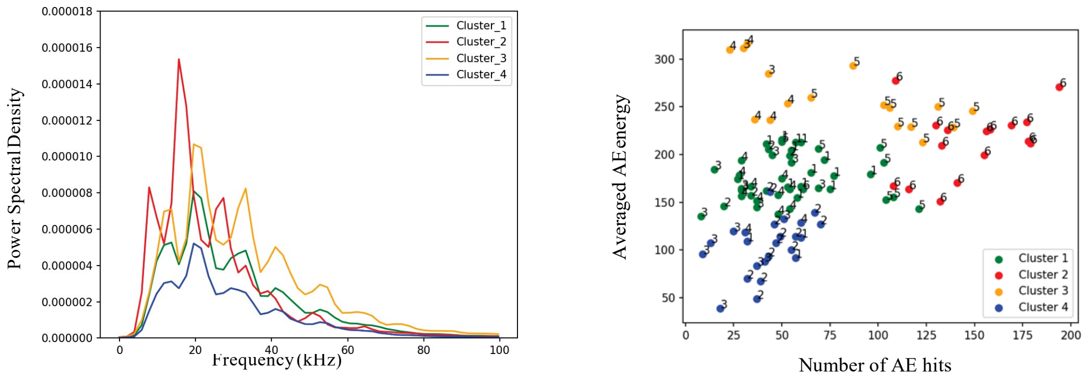
(b) Scatter plot of each group

Figure 4. Clustering result of the data ($k = 2$).



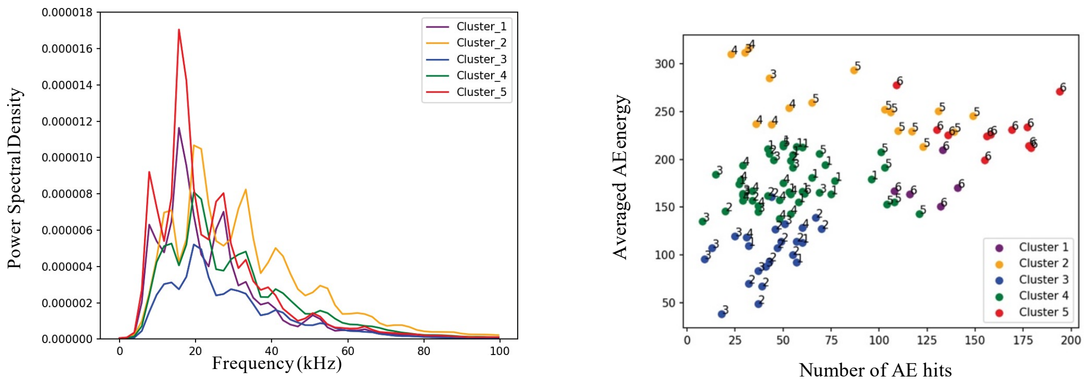
(a) PSD of each cluster centroid (b) Scatter plot of each group

Figure 5. Clustering result of the data ($k = 3$).



(a) PSD of each cluster centroid (b) Scatter plot of each group

Figure 6. Clustering result of the data ($k = 4$).



(a) PSD of each cluster centroid (b) Scatter plot of each group

Figure 7. Clustering result of the data ($k = 5$).

To evaluate the optimal k for the clustering, silhouette analysis (Rousseeuw, 1987) was conducted. The silhouette score of the data point i is:

$$s(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}} \tag{2}$$

$$a(i) = \frac{1}{N_i - 1} \sum_{j \in C_i, i \neq j} d(i, j) \tag{3}$$

$$b(i) = \min_{k \neq i} \frac{1}{N_k} \sum_{j \in C_k} d(i, j) \tag{4}$$

where N_i is the number of data points in the cluster C_i, C_i is the cluster to whom the data point i belongs and $d(i, j)$ is the distance between data points i and j.
The range of the silhouette score is from -1 to +1, and a high value means that the data point is well matched to its cluster. From the plotting of the silhouette scores, the optimal number of the clusters can be determined graphically.

Figure 8 show the plotting of the silhouette score of each data. The black dotted line means the average of the whole silhouette score, and the thickness of the silhouette score of each cluster means the number of the data. In Figure 8 (a), the number of the data that belong to cluster 1 was larger than that to cluster 2, therefore, $k = 2$ may not be the optimal value for the clustering of the data. In Figures 8 (b) and 8 (d), some silhouette score was negative, therefore, $k = 3$ and 5 may not be also the optimal value for the clustering of the data. In Figure 8 (c), the thickness of the silhouette score of the cluster 1 was slightly thicker than those of other clusters, however, there was no negative silhouette score and many of silhouette scores exceeded the average silhouette score. Thus, $k = 4$ may be the optimal value for the clustering of the data.

The optimal k might be determined by the silhouette score, however, it is necessary to evaluate the accuracy of the clustered groups with the actual damage condition of the target structure. This is the future work for our study.

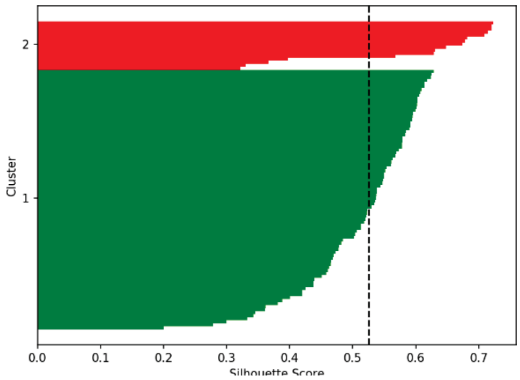
4 Conclusions

Following conclusions were derived from this study.

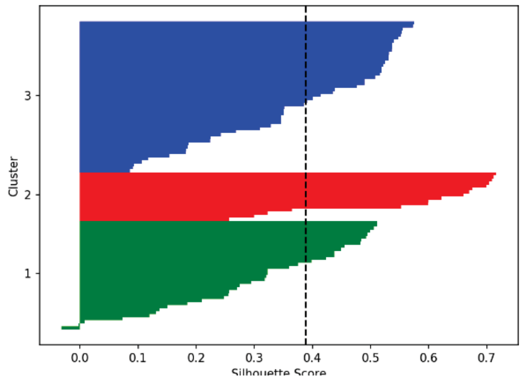
- The area where the number of AE hits and the average of AE energies were relatively lager than other areas might be the damaged area.
- From the result of the K-means clustering, the data was separated into the obviously damaged area and other areas. In addition, the characteristic of the power spectral density of the centroid of the damaged cluster was different from those of other clusters.
- As the increase of the number of centroids k, the data was separated into some groups based on the average of the AE energies.

ORCID

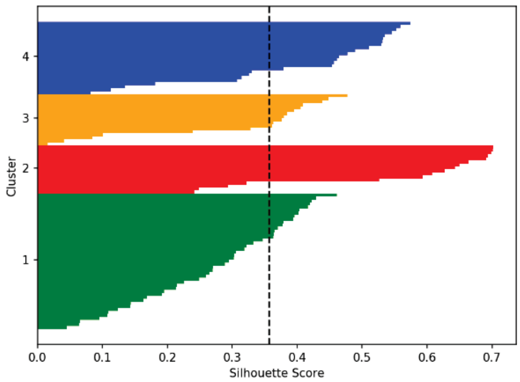
Yutaka Tanaka: https://orcid.org/0000-0002-9685-1330
Takahiro Nishida: https://orcid.org/0000-0002-2018-6928



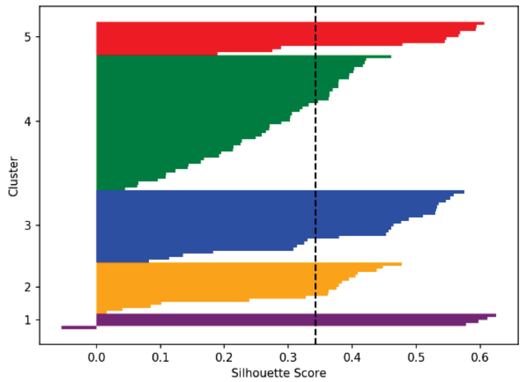
(a) $k = 2$



(b) $k = 3$



(c) $k = 4$



(d) $k = 5$

Figure 8. Silhouette score plots.

References

Arthur, D. and Vassilvitskii, S. (2007). *k*-means++: the advantages of careful seeding. *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, Society for Industrial and Applied Mathematics Philadelphia, PA, USA, 1027-1035.

MacQueen, J. B. (1967). Some Methods for classification and Analysis of Multivariate Observations. *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 281-297.

Nishida, T., Hashimoto, K., Yasuzato, T. and Kawada, K. (2019). Data Analysis for Damage Area Detection of Reinforced Concrete Decks, *Proceedings of The 4th International Symposium on Concrete and Structures for Next Generation*, 371-376.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python, *JMLR 12*, 2825-2830.

Rousseeuw J. P. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, *Journal of Computational and Applied Mathematics, 20*, 53-65.

Virtanen, P. *et al* (2019). SciPy 1.0 –Fundamental Algorithms for Scientific Computing in Python, arXiv:1907.10121.