Unsupervised Learning Approach to Detection of Void-type Defects in Concrete Structure Using Hammering and Clustering

Jun Younes Louhi (Kasahara)*1, Hiromitsu Fujii*1
Atsushi Yamashita*1, Hajime Asama*1
*1 Graduate School of Engineering, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656 Japan

In this paper we present an online unsupervised method based on clustering to find void-type defects in concrete structures using hammering. The dataset of sound samples is clustered in order to find the regular model for the hammering sound, which is assumed to be the non-defective sound model. The algorithm is fast and reliable enough to allow efficient diagnosis by running it each time a new sample is acquired.

1 Introduction

Concrete structures are extremely common and can be greatly affected by aging and environmental conditions, those often lead to structural failure. A fast and reliable diagnosis method to ensure their proper maintenance is needed.

The popular hammering test consists of hitting the structure with a hammer and assessing the presence of defects from sound. Its performance largely relies on the operator's skill and thus its automation is demanded.

Previous works consists of [1] and [2], using supervised learning to correctly spot defective areas. However these approaches need to train the algorithm first using a training set and, due to various factors, concrete can greatly differ from one structure to another, therefore making these methods difficult to use in practice. In [3], we presented a new approach to this task using unsupervised learning based on clustering.

In [3], crack-type defects only were examined. In this paper, we expanded our previous work to the detection of void-type defects and implemented it online.

2 Method

2.1 Concept

It assumed that most of the tested structure is non-defective. Therefore, if a large group of similar sounds can be found in the dataset, it can be inferred that it is the regular sound of the structure and a model representing the regular, non-defective sound of that structure can be made. Simple clustering algorithms, such as k-means, can effectively accomplish this task. This regular model can then be used as a reference to conduct a diagnosis of the whole dataset.

Having a fast enough algorithm, online implementation can simply be done as in Algorithm 1.

2.2 Feature vector and dissimilarity measure

Fourier spectrum is used as feature vector for a hammering sound sample and we need to define an appropriated metric. To do this, we use the sample Pearson correlation coefficient r_{AB}. Given two Fourier spectrum A and B, respectively defined by $(a_0, ..., a_{N-1})$ and $(b_0, ..., b_{N-1})$, our metric is defined as in Eq. (1).

The defined distance is ranging in [0.0,1.0], returning small values the more the compared sounds are alike and zero if the sounds are identical. Cases of negative correlation are located in the [0.5,1.0] range since negative correlation is in our case not a similarity.

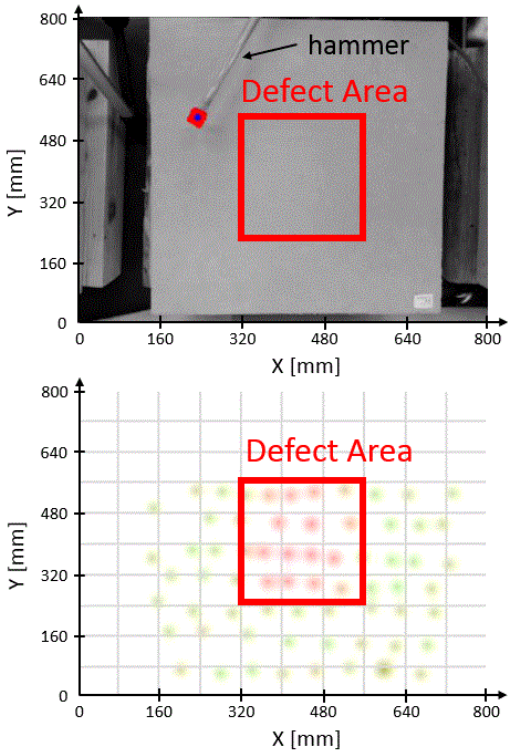

Fig. 1 Online diagnosis being conducted: image of the test block and the detected hammer head in red (upper), results with distance shown using a colorscale, defective spots appear red (bottom).

$$d(A,B) = \frac{1-r_{ab}}{2} = \frac{1}{2}\left(1-\frac{\sum_{l=0}^{N-1}[(a_l-\bar{a})(b_l-\bar{b})]}{\sqrt{\sum_{l=0}^{N-1}(a_l-\bar{a})^2}\sqrt{\sum_{l=0}^{N-1}(b_l-\bar{b})^2}}\right) \quad (1)$$

2.3 The algorithm

The k-means++ seeding procedure described in [4]. In our case, the first seed is chosen randomly following an uniform distribution. For the second seed, a probability distribution to reflect the similarity to the first seed is devised: each sample X_i has a probability $P(X_i)$, as defined in Eq. (2), based on the previously defined metric to the first seed S_1, $d(X_i, S_1)$, to be chosen. Unlike the regular seeding process where the seeds are simply chosen randomly,this procedure allows the

Data: dataset of hammering samples
Result: distance of each sample to the regular model
initialization;
while $N_{sample} < N_{initial}$ **do**
  keep collecting samples;
end
run k-means++ with k=2;
find biggest cluster $Cl_{biggest}$;
get its centroid $C_{biggest}$;
foreach X_i **do**
  calculate $d(X_i, C_{biggest})$;
end
while *testing is ongoing* **do**
  if *new sample is acquired* **then**
    add new sample to the dataset;
    run k-means++ with k=2;
    find biggest cluster $Cl_{biggest}$;
    get its centroid $C_{biggest}$;
    foreach X_i **do**
      calculate $d(X_i, C_{biggest})$;
    end
  end
end

Algorithm 1: Pseudo algorithm for the proposed method

seeds to be spread trough the dataset and therefore close to the final centroids location.

$$P(X_i) = \frac{d(X_i, S_1)^2}{\sum_{i=1}^{N_{sample}} d(X_i, S_1)^2} \tag{2}$$

The clustering algorithm is applied using the correlation distance defined in Eq. (1) to classify the sound dataset in two clusters. Then, the cluster containing the majority of samples, $Cl_{biggest}$, is assumed to be the one containing all the non-defective samples. Its centroid $C_{biggest}$ is used as a model for the non-defective sound: each sound sample is compared to it using the distance defined in Eq. (1). Since the model represents the most regular sound shape in the dataset, irregularities i.e. distant sound samples can be recognized as characteristic of defects on the structure.

3 Experiments

Testing was conducted on a concrete test block containing a 200x200 mm square void on its center at a depth of 30 mm (Fig. 1). The used hammer was a KTC UDHT-2 (head diameter 16 mm, length 380 mm, weight 160 g) and sound was recorded at 44.1 kHz using a Behringer ECM8000 microphone coupled with a Roland UA-25EX sound board and a laptop with an Intel core i7-4500U @ 1.80 GHz for data analysis.

Receiver Operating Characteristic (ROC) curve was computed and we obtained a value of 0.92 for the area under the curve. This value close to 1 shows that this method's efficiency is high for the detection of void-type defects.

Time taken for the algorithm to return a result at each new sample was measured. After stabilization, when around 200 samples were collected, the average was 454 ms. This allows the hammer to hit the structure twice in one second and therefore could be considered satisfying.

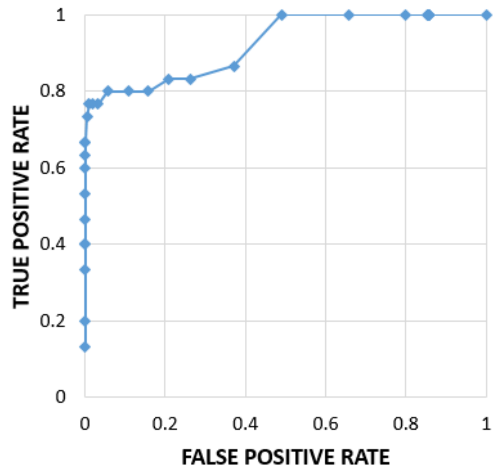

Fig. 2 ROC curve of the proposed method, area under the curve of 0.92 shows high efficiency.

4 Conclusion

We proposed an online implementation of the method proposed in the previous work and achieved satisfactory results for the detection of void-type defects. In future work, we would like to increase the robustness of this method, especially at the regular model generation stage.

Acknowledgements

This work was supported in part by the Cross-ministerial Strategic Innovation Promotion Program (SIP) of the New Energy and Industrial Technology Development Organization (NEDO), Grant-in-Aid for JSPS Fellows 269039, and Institute of Technology, Tokyu Construction Co., Ltd.

References

[1] Hiromitsu Fujii, Atsushi Yamashita and Hajime Asama: "Automated Diagnosis of Material Condition in Hammering Test Using a Boosting Algorithm", *Proceedings of the 2014 IEEE Workshop on Advanced Robotics and its Social Impacts*, pp. 101–107, 2014.

[2] Jiaxing Ye, Masaya Iwata, Takumi Kobayashi, Masahiro Murakawa, Tetsuya Higuchi, Yuichi Kubota, Toshiya Yui and Kiyoshi Mori: "Statistical Impact-Echo Analysis Based on Grassmann Manifold Learning: Its Preliminary Results for Concrete Condition Assessment", *Proceedings of the 7th European Workshop on Structural Health Monitoring*, pp. 1349-1356, 2014.

[3] Jun Younes Louhi (Kasahara), Hiromitsu Fujii, Atsushi Yamashita and Hajime Asama: "Clustering-based Automated Diagnosis of Concrete Condition Using Hammering", *Proceedings of the 33th Annual Conference of the Robotics Society of Japan*, RSJ2015AC2G2-08, 2015.

[4] David Arthur and Sergei Vassilvitskii: "k-means++: The advantages of careful seeding", *Proceedings of the 18th annual ACM-SIAM symposium on Discrete algorithms*, pp. 1027-1035, 2007.