

Periodicity Estimation in Mechanical Acoustic Time-Series Data

Yongbo Zhu, Zhipan Hong, Guoliang Lu ^a

Key Laboratory of High-efficiency and Clean Mechanical Manufacture of MOE
School of Mechanical Engineering, Shandong University, Jinan, 250061, China

Abstract. Periodicity estimation in mechanical acoustic time-series data is a well-established problem in data mining as it can be applicable in variety of disciplines either for anomaly detection or for prediction purposes in industry. In this paper, we develop a new approach for capturing and characterizing periodic patterns in time-series data by virtue of the dynamic time warping (DTW). We have conducted extensive experiments to evaluate the proposed approach with synthetic data and our collected data in practice. Experimental results demonstrated its effectiveness and robustness on periodicity detection in highly noised data.

1 Introduction

Data mining in time series is a well-established problem in signal processing as it can be applicable in a wide range of industrial disciplines which acquire and record large amounts of periodic time-series data. For the analysis of such data, periodicity estimation, namely, discovering the potential periodic patterns in a given time-series data, is the first and essential step which can: (1) allow for more effective and accurate signal processing in many high-level purposes such as seen in anomaly detection and prediction tasks; (2) allow for more meaningful and accurate processing, e.g., clustering and classification of given data, and (3) also be used for interactive exploration and visualization of massive periodic database.

Let us consider one specific area where periodicity can be applicable in industrial field:

Periodic analysis is an indispensable tool in automotive, aviation and manufacturing industries for machine monitoring and diagnostics [1]. Based on a widely accepted fact that mechanical acoustic data can be a good indicator of running-state of a machine, we can identify whether a structural system has changed from its preceding running-state, depending on the estimated periodicity, through a statistical and mathematical comparison of the data in the past periods with the one in the current period, and then take appropriate actions to maintain the operation and to avoid accidents.

In the past several decades, we have witnessed a lot of research achievements to address this problem, e.g., [2, 3], but the majority of these approaches generally require manual re-calibration or intervention in execution. Therefore, automatic approaches for periodicity detection/estimation are still in progress until now. In this study, we proposed a new approach for capturing and characterizing periodic patterns in mechanical acoustic

time-series data. Our main contributions of the study are summarized as follows.

—We develop a computation approach to accurately detect the potential periodic patterns in a given time-series data. This approach does not require any manual re-calibration or intervention in execution.

—We conducted extensive experiments to evaluate the proposed approach with synthetic data and our collected data in practice. Experimental results demonstrated its effectiveness and robustness on periodicity detection in highly noised data.

2 Periodicity detection problem

Let us assume that a sequence of n time-stamped measured values has been collected in a time series Y . “A time series Y is said to be periodic with a period p if it can be divided into equal length segments, each of length p , that are almost similar” [4]. One simple instance is a time series $Y=\{abcabcabcd\}$ which is obviously periodic with $\ell =3$. Hamming distance can be a good indicator to measure the degree/level of similarity of two segments by comparing the time-stamped values position-wisely [5], which can be defined as follows:

$$H(a,b) = \sum_{j=0}^{m-1} \begin{cases} 1 & a_j \neq b_j \\ 0 & a_j = b_j \end{cases} \quad (1)$$

$$S(a,b) = 1 - H(a,b)/m$$

Where a, b are two segments with equal length m , and a_j, b_j are corresponding to j th data in these two segments a and b respectively. H is the hamming distance, while S is the similarity measure.

^a Corresponding author: luguoliang@sdu.edu.cn

2.1 Major issue of periodicity detection for mechanical acoustic time series data

The higher value length m in hamming distance is, the more similar the two segments a , b are, i.e., is more approaching to 1. However, in mechanical acoustic time-series data, there are often relatively-large variations in both amplitude and timing among segments in two periods due to the unavoidable noise in capturing and transmission in real-world workshop, which takes a serious effect on periodicity detection. Taking $Y = \{abcabbcababd\}$ for example, Y can be also periodic in spite that the second segment $\{abbc\}$ and the fourth segment $\{abd\}$ are not the same as others due to time shift and amplitude variation. As for this kind of situation, the hamming distance fails to measure the similarity of two segments, e.g., $\{abbc\}$ and $\{abc\}$ when they are out of the phase. Therefore, for effective and reliable data mining, we need the periodicity estimation to be accurate and robust.

3 Proposed scheme: periodicity estimation in mechanical acoustic time-series data

To absorb the variations in both amplitude and timing in collected data and the existed noise, we employ to use the *Dynamic Time Warping* (DTW) [6] to measure the similarity of two segments for periodicity estimation.

The DTW finds the optimal alignment between two segments of numerical measurement-values. Let us assume that two segments A and B have been extracted time-stamped features, i.e., $A = \{a_i\}$, $i=1, 2, \dots, N$ and $B = \{b_i\}$, $i=1, 2, \dots, M$, where a_i and b_i are corresponding to i th frame in A and B . DTW calculates the distances between every pair of (a_i, b_i) in terms of their corresponding feature values. Various distance metric p can be used in calculation. The best match between A and B , namely the optimal warping path, is the one with the lowest distance given by $DTW_p(A, B) = D(a_N, b_M)$ where matrix D is dynamically computed by:

$$D(a_i, b_j) = |a_j, b_j|^p + \min \begin{cases} D(a_{i-1}, b_{j-1}) \\ D(a_i, b_{j-1}) \\ D(a_{i-1}, b_j) \end{cases} \quad (2)$$

Then, our proposed approach for periodicity detection in mechanical acoustic time series data is executed as follows.

3.1 Segment comparison

To capture and characterize the potentially periodic patterns in a given mechanical acoustic time-series data, we need to synchronize one candidate segment with the given mechanical acoustic time series. The candidate segment should contain the typical patterns of periodicity in time series and should have the same or at least very close time length as one periodic data. Supposing that we

have obtained one time series data, i.e., $Y = \{y_1, y_2, y_3, \dots, y_n\}$ where n is the data number of Y , we then perform the following procedures for segment comparison :

- Determine the lengths of candidate segments to be from l_1 to l_2 . l_1 and l_2 can be estimated empirically or estimated with prior knowledge in practice;
- For each candidate segment Y^i , i.e., $Y^i = \{y_{l_1}, y_{l_2}, y_{l_3}, \dots, y_{l_i}\}$ where $l_i \in [l_1, l_2]$, we compared it with the given time series Y iteratively on the entirety of Y , and the comparison cost C_i is defined as:

$$C_i = \sum_j DTW(Y^i, \tilde{Y}_j) \quad (3)$$

Where \tilde{Y}_j is the subsequence of Y with the equal data number of Y^i , i.e., $\tilde{Y}_j = \{y_{j_l}, y_{j_{l+1}}, \dots, y_{j_{l+i-1}}\}$ where j_l is the index of first data in \tilde{Y}_j ;

- Repeat the step b. We last have the comparison cost $\{C_i\}$ for all candidate segments $\{Y^i\}$.

Using the above segment comparison, the given time series Y is synchronized with all candidate segments.

3.2 Periodicity determination

Since the resulted comparison cost can measure the compatibility of candidate segments, specifically, a lower cost indicates a more compatibility for one compared segment, we can determine the expected period-length ℓ by seeking the minimum comparison cost, as:

$$\ell \leftarrow l^*, \text{ where } l^* = \arg \min_i C_i \quad (4)$$

4 Experiment

We conducted two experiments in order to evaluate the effectiveness of our proposed approach. In the first experiment, we first use a synthetic sine signal with a random noise for a preliminary evaluation, and then, add white noise with different SNR in the sine data to evaluate the robustness of the approach to noise. In the second experiment, we use our collected acoustic signal of gear-shift of engine to evaluate the effectiveness of our approach.

The detailed description of the experiments and results are given in the following.

4.1 Experiment I: synthetic data

To preliminarily evaluate the effectiveness of the approach, we synthetically generate a sine signal, i.e., $y = a \sin(\omega t + \varphi)$, with a random noise and we set its parameters as *Amplitude*: $a = 1$, *Frequency*: $\omega = 1/5$, *Phase*: $\varphi = 0$, *Sampling*: 10.

Clearly, the period length of this sine signal is $2\pi/\omega$, that is, 31.42. Under a sampling rate of 10, the detected period length is approximately to 314. In the detection of this experiment, we set the lengths of candidate segments $\in [280, 350]$ empirically. The detected result in this signal is shown in Fig.1. More specifically, from

Fig.1 we can see that the detected periodic length is 315, which is approximately to be 314. So, the approach is feasible in this synthetic sine signal. Note that, in this test data, the SNR (signal-to-noise ratio) is 3.0095dB. However, it is uncertain that whether the algorithm is still feasible when the SNR is higher or lower than it.

Therefore, in order to evaluate the feasibility of the approach with the influence of *differently-level* noise, we re-generated the *sine* signal by adding noise from a SNR range of -50dB to 50dB with an interval 5dB. The detected results are shown in Fig.2, where the line with “*” indicates the detected results and the dotted line indicates the real period. For SNR higher than 0 dB that indicates more signal than noise, our approach can be operated successfully. When SNR is lower than 0 which indicates more noise than signal, our approach degrades in the overall trend. In Fig.3, we give such an example (SNR -35dB). The noise is much higher than the original signal which make it very hard to discriminate the actual signal. For this case, our approach fails to the estimation due to the large noise. Overall, from Fig.2, we can conclude that our approach can feasible for periodicity estimation for a SNR at least of -20dB.

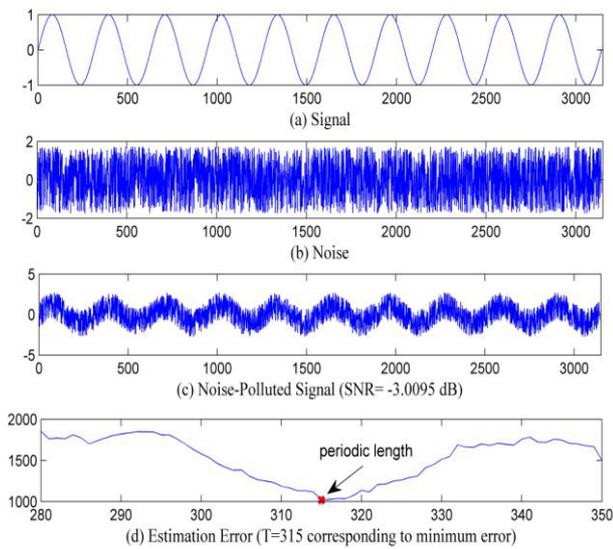


Fig.1 The detected result in the *sine* signal

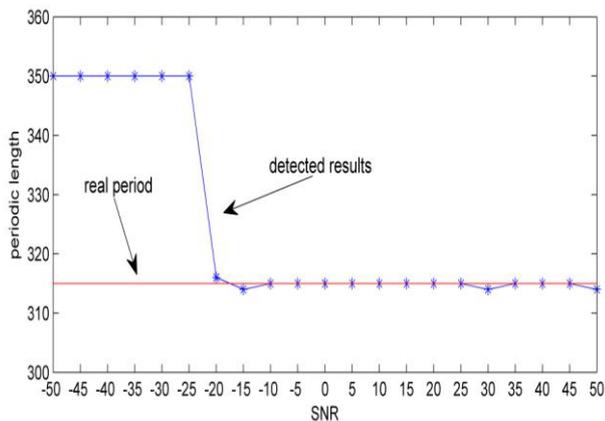


Fig.2 The detected results with the SNR of -50dB to 50dB

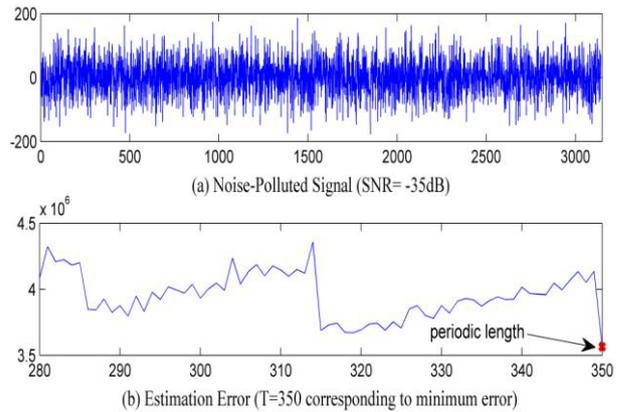


Fig.3 Periodicity detection in a highly-noise polluted data (SNR -35dB).

4.2 Experiments II: Real signal

In this session, we will evaluate the feasibility of the proposed approach on our collected acoustic signal. This test signal was first acquired by a microphone which was mounted on an automobile transmission gearbox; then the signal was amplified and sent to a sound level meter for periodicity estimation in the experiment.

Because the acoustic signal of engine gearbox is a very large data and with an unknown noise, we sampled it down to eliminate the impact of noise and decrease the burden in computation. The down-sampling rate is 1/50 , and we choose the beginning 1000 data of resampled signal to estimate its period . In estimation, we set the lengths of candidate segments $\in [100, 200]$ in this data. Fig.4 shows the result. It can be seen that the estimated period-length is approximately equivalent to the desired result it should be, which indicates that the approach is feasible in actually-collected mechanical acoustic time-series data.

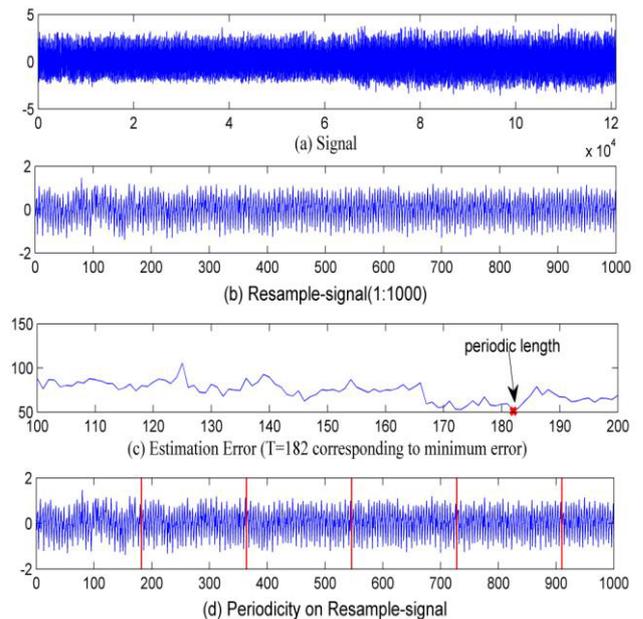


Fig.4 Periodicity detection in the acoustic signal of an automobile transmission gearbox.

5. Conclusion

In this paper, we have proposed a new approach for capturing and characterizing periodic patterns in mechanical acoustic time-series data. The approach does not require any manual re-calibration or intervention in execution. We have conducted extensive experiments to evaluate the proposed approach with synthetic data and the collected data in practice. Experimental results demonstrated its effectiveness and robustness on periodicity detection in highly noised data.

In the future work, we will apply the proposed approach for mechanical anomaly detection and prediction task in the workshop.

Acknowledgement

This work is supported in part by National Natural Science Foundation of China (61403232), Natural Science Foundation of Shandong Province, China (ZR2014FQ025).

Reference

- [1] ZeFeng WANG, Jean-Luc ZARADER, Sylvain ARGENTIERY, "A Novel Aircraft Engine Fault Diagnostic and Prognostic System based on SVM," 2012 IEEE International Conference on Condition Monitoring and Diagnosis 23-27 September 2012, Bali, Indonesia. pp. 723–728.
- [2] Yuan Li, Zhongjie Xie, "The wavelet detection of hidden periodicities in time," Elsevier Science B.V. Statistics & Probability Letters 35 (1997) 9-23, Received July 1996; revised October 1996.
- [3] Feiyan Hu, Alan F. Smeaton, Eamonn Newman. "Periodicity Detection in Lifelog Data with Missing and Irregularly Sampled Data," in International Conference on Bioinformatics and Biomedicine. IEEE , 2014, pp. 16–23.
- [4] Elfeky, Mohamed G.; Aref, Walid G.; and Atallah, Mikhail J., "Periodicity Detection in Time Series Databases" (2002). Computer Science Technical Reports. Paper 1547.
- [5] Stan Z. Li, Anil Jain, "Encyclopedia of Biometrics," Springer US, 2009.
- [6] Alan Bundy, Lincoln Wallen. "Catalogue of Artificial Intelligence Tools," Symbolic Computation 1984, pp 32-33.