Electromagnetic wave pattern detection using cepstral features in the manufacturing field

Ayano Ohnishi¹, Michio Miyamoto¹, Yoshio Takeuchi¹, Toshiyuki Maeyama¹,², Akio Hasegawa¹, and Hiroyuki Yokoyama¹

¹ Advanced Telecommunications Research Institute International, 2–2–2 Hikaridai, Seika-cho, Sorakugun, Kyoto 619–0288, Japan
² Takushoku University, 815–1 Tatemachi, Hachioji-shi, Tokyo 193–0985, Japan

a) ayano@atr.jp

Abstract: In manufacturing fields such as factories, multiple wireless communication systems often operate in the same area simultaneously. Moreover, it is known that industrial equipment emits electromagnetic noise over channels for wireless communications [1]. In order to ensure reliable communications under such an environment, monitoring radio wave environments specific to each manufacturing field and finding channels and timing which enable stable communications are required. The authors have studied technologies to efficiently analyze a large amount of monitoring data including signals which show unknown spectrum such as wide band electromagnetic noise [2, 3]. This paper proposes performing machine learning using cepstrum vectors as features to grasp types of noise and signals from data measured under environments in which electromagnetic noise and communication signals coexist. By using this features, the authors demonstrate that reduction of computational loads and improvement of detection accuracy can be expected.

Keywords: wireless communication, factory, electromagnetic noise, machine learning, clustering, cepstrum

Classification: Sensing

References

[1] IEEE 802 Nendica Report, “Flexible factory IoT: Use cases and communication requirements for wired and wireless bridged networks,” IEEE SA Industry Connections White Paper, April 2020.
[2] A. Ohnishi, M. Miyamoto, Y. Takeuchi, A. Hasegawa, and H. Yokoyama, “A study of noise pattern detection for frequency sharing among various communications under high electromagnetic noise in the manufacturing field,” Proc. 2020 IEICE General Conference, B-17-28, 2020 (in Japanese).
[3] M. Miyamoto, A. Ohnishi, Y. Takeuchi, A. Hasegawa, and H. Yokoyama, “Development of measurement technology for frequency sharing among various
1 Introduction

In recent years, sampling and recording data for a long time have become possible when monitoring and analyzing radio wave environments because software defined radio receivers and storage media of a large capacity and high speed performance have become widely available. Therefore, methods to extract the necessary information by efficiently analyzing types of signals and their generation timing from a large amount of data are needed.

Electromagnetic noise emitted from industrial equipment especially does not have known spectrum and time wave forms such as communication signals. In addition to this, the spectrum and time wave forms are not fixed because kinds of equipment or their layouts in each factory are different. Therefore, algorithms to detect known spectrum and known wave forms are not applicable.

Machine learning is useful when generally grasping types of signals from data measured in environments in which electromagnetic noise and communication sig-
signals coexist. An assumed classification process is as follows: a classifier is built by learning prepared supervised data. After that, examination data is input to the classifier and classified by types of signals. In order to do this, firstly, it needs to be prepared supervised data efficiently by performing clustering of unclassified data.

In this paper, cepstrum vectors which are calculated and extracted from IQ samples are used as features when performing clustering process of unclassified data including electromagnetic noise emitted by industrial equipment.

Cepstral analysis is a widely known technique in the field of speech recognition and vibration analysis. It is used to extract envelopes of spectrum and detect periodicity of spectrum and so on [4, 5]. Cepstral features have been used to classify data in their field. Cepstral features calculated from vibration data of on-vehicle motion sensors are used to detect some events in [6]. In the wireless communication fields, cepstral features are used to analyze the channel utilization state of primary users in [7] and identify modulation technique [8] but there are few examples using cepstral features for detection of electromagnetic noise. Focusing on features for detection of noise and signal, if data is investigated previously and information about characteristics of frequency is not needed, only using average and deviation values of power spectrum will work [9]. However, features including information about spectrum pattern should be used when analyzing data containing signals which have obvious characteristics related to frequency such as narrow band communications. Alternatively, it is necessary to examine computation speed and reduce the amount of data when using spectrum as features to keep information about frequency [10]. We propose using cepstral features aiming to reducing the amount of data and improving the accuracy of clustering.

We describe advantages of using cepstral features in Section 2. Then, we show the effectiveness of using them to detect electromagnetic noise and communication signal by showing clustering results with the actual data recorded in a factory in Section 3. Finally, we conclude in Section 4.

2 Calculation of cepstral features

Clustering process is performed to grasp various signal patterns such as communication signals and electromagnetic noise. Cepstral vectors are used as the features in this paper. The cepstrum is defined by performing inverse Discrete Fourier Transform (DFT) of the sequence after calculating logarithm of magnitude of the DFT of a signal. The cepstrum is given by

$$\text{cepstrum} = \text{DFT}^{-1}(\log |\text{DFT}(x_n)|) \quad (1)$$

where \(\log(\cdot)\) denotes the logarithm with any base and \(x_n\) is a signal sequence obtained in a certain period and \(\text{DFT}(\cdot)\) is DFT and \(\text{DFT}^{-1}(\cdot)\) is inverse DFT.

2.1 Dimension of cepstral features

Extracting low-dimensional elements of a cepstral vector calculated from monitoring data (IQ samples) of radio wave environments corresponds to applying low-pass filter to a spectrum pattern. This facilitates discriminating of spectrum patterns. Fig. 1 shows a part of cepstrum vectors extracted from 15000 vectors calculated using
actual data. Elements of cepstrum vectors are close to zero in higher dimensions. 16-dimensional cepstrum of signal and noise have different characteristics which enable to classify them. Therefore, we decided to extract 16-dimensional cepstrum and use as features.

2.2 Advantage of cepstral features in clustering process

(1) Reducing the number of dimensions

When extracting low-dimensional elements from cepstral vectors and using as features, it is possible to reduce computational loads of clustering process and memory consumption by reducing the number of dimensions of features drastically while keeping characteristics of spectrum. It is also possible to reduce the amount of data recording in storage media keeping their characteristics.

(2) Suppression of variations in clustering results

When using spectrum as features, some results are not appropriate depending on the random seed because of ambiguity of spectrum patterns including white noise and randomness found in initial settings of clustering algorithms (k-means++ [11], which can improve initial settings in k-means algorithm, is used in this paper). On the other hand, when using cepstrum as features, generating inappropriate results are suppressed by reducing white noise. (An example of clustering results is shown in Chapter 3.)

(3) Estimation of the number of clusters

As explained in (2), because variations in clustering results are suppressed, the same results are generated regardless of the values of random seed as long as values below a value of the proper number of clusters are set to the parameter for clustering analysis. Here the proper number of clusters refers to the situation in which each classified pattern has different characteristics and multiple signal patterns are classified properly. When clustering process is performed setting values exceeding the proper number of clusters, the results are different depending on the random seed because each cluster is classified in an inappropriate boundary. It is possible to simply estimate the proper number of clusters using this.
3 Clustering of electromagnetic wave pattern

As shown in Fig. 2 [2], a pattern of narrow band communication signals and two patterns of electromagnetic noise are contained in the actual data (IQ samples) analyzed in this paper.

![Fig. 2. (a) Spectrogram of analyzed data. (b) Communication signal. (c) Type1 of electromagnetic noise. (d) Type2 of electromagnetic noise (modified based on [2] Figs. 1, 2)](image)

We investigated whether generation timing of electromagnetic noise and communication signals are detected properly by performing clustering process. We extracted and used 16-dimensional cepstral features from a 15000-dimensional spectrum vector at every millisecond and analyzed them for one second in total. A 15000-by-1000 array is input to clustering process at every second when using spectrum. In contrast, a 16-by-1000 array is input when using cepstral features. Therefore, it is performed the reduction of data of approximately one-thousandth when running clustering process at every one second. The k-means algorithm have a computation complexity of O(ndk) [12], where n is the number of data points and d is the number of dimensions, and k is the number of clusters. The number of dimensions is reduced by using cepstral features.

The clustering results when using cepstral features and setting the number of clusters within the range of 3 to 5 are shown in Fig. 3. The ordinate of the graph is cluster Id and it is shown the cluster Id into which signals have been classified at every millisecond. When setting the number of clusters to 4, communication signals and two types of noise patterns are classified properly in both random seeds.

The clustering process was performed ten times changing the value of random seed to investigate suppression of variations in clustering results and possibility of estimating the number of clusters. As an example, Fig. 3 shows the clustering results in two different seeds while using cepstral features. When setting the number of clusters to 3 or 4 which are less than the appropriate number, no matter what random seed was used, the results were the same and appropriate. However, when using...
spectrum as features, even if the number of clusters less than the appropriate number was set, these results were not the same and inappropriate in some cases. On the other hand, when setting the number of clusters to values exceeding the appropriate number while using cepstral features, the different results are shown depending on the values of random seed. From this, it is possible to estimate the proper number of clusters by seeing if the results are variable while changing the value of a random seed.

4 Conclusion

This paper demonstrated that the reduction of data of approximately one-thousandth and detecting of multiple type of signals with high accuracy were performed by using cepstral features when analyzing data measured under environments in which electromagnetic noise and communication signals coexist. We also confirmed that estimation of the proper number of clusters was possible by comparing clustering results while changing the value of random seeds.

Acknowledgments

This work includes results of the project entitled “R&D on Adaptive Media Access Control for Increasing the Capacity of Wireless IoT Devices in Factory Sites,” which is supported by the Ministry of Internal Affairs and Communications as part of the research program “R&D for Expansion of Radio Wave Resources (JPJ000254)”.

The authors would like to thank to Murata Machinery, Ltd. for supporting experiments in their factory.