2023 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (RISP NCSP 2023), Feb. 28 - Mar. 3, 2023

Investigation of Sound Data with Three Types of Noise-Mixing Effects by Neural Network Using Autocorrelation Function

Takuya Nakamura, Ryosuke Shimizu, Yoko Uwate and Yoshifumi Nishio

Dept. of Electrical and Electronic Engineering, Tokushima University
2-1 Minami-Josanjima, Tokushima 770-8506, Japan
E-mail: {takuya, shimizu, uwate, nisho}@ee.tokushima-u.ac.jp

Abstract

One-dimensional convolutional neural networks (1D-CNN) are used for time series analysis. However, noise mixed in the data can interfere with time series analysis. Therefore, we compare the classification accuracy of two types of patterns. One is a pattern learned by mixing three kinds of noise (white noise, pink noise, and red noise) into the data, and the other is a pattern learned by replacing the original data with an autocorrelation function (ACF).

1. Introduction

There are many phenomena in the world that have time series, such as temperature and seismic waves. In addition, audio and video can also be considered time series. Its features are difficult for the human eye to recognize and judge. However, by using a neural network (NN), it is possible to discriminate even the smallest details. The academic field of NN was established in 1958, and much research has been conducted to date [1]. The advantage of NN is that it can learn from input data and can be used for pattern recognition and data classification. NN include well-known models such as Recurrent neural networks (RNN) and Convolutional neural networks (CNN) [2]. In recent years, many researcher have been investigated on time series analysis, in which neural network is trained to analyze and classify time series [3].

However, the data often contain information that is not needed for analysis, called noise. If the amount of noise mixed in the data is too large, time series analysis becomes complicated and very troublesome. In this study, three types of noise (white noise, pink noise, and red noise) are mixed into the training data of speech, and the classification accuracy of the NN is investigated.

2. 1D-CNN

CNN is a type of NN with a convolution layer and a pooling layer added [4]. CNN is also used for natural language processing and image recognition. Figure 1 shows the example of image recognition using CNN. However, CNN has the disadvantage of being computationally expensive and time-consuming since they are mainly used for image recognition [5]. Therefore, the use of 1D-CNN can reduce the computational cost. In addition, it is possible to automatically extract features through learning. The one-dimensional Residual Network (1d-ResNet) is used as the model for the NN. This is a model proposed by He of Facebook AI Research in 2015 [6]. This model can be used to prevent the gradient loss problem, which prevents learning from progressing in the NN.

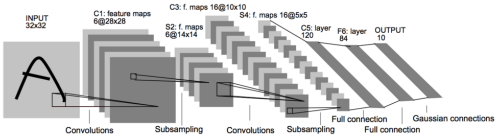

Figure 1: The example of image recognition using CNN.

3. Dataset

In this study, three sound data sets are used: fireworks, thunderstorms, and vacuum cleaner. Table 1 shows the number of training and test data used in this study. These numbers are performed under the same conditions for all three noises.

Table 1: Number of train data and test data.

Data	Original	ACF
Train data	1500	1500
Test data	500	500

Figures 2, 3, and 4 show examples of fireworks, thunderstorms, and vacuum cleaner sounds.

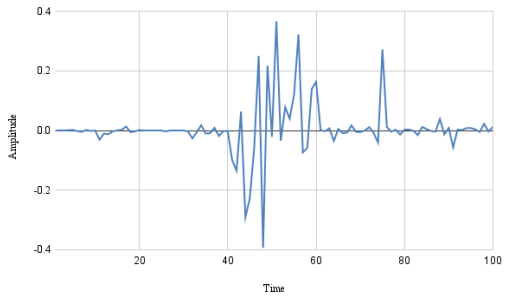

Figure 2: The example of fireworks sound.

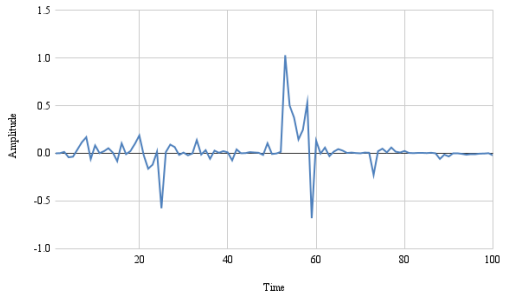

Figure 3: The example of thunderstorms sound.

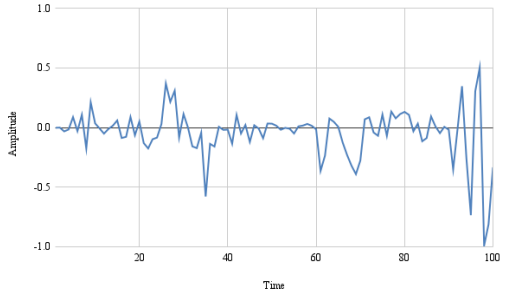

Figure 4: The example of vacuum cleaner sound.

4. Proposed Method

In this study, the following method is used to confirm the accuracy of the classification of noise in the data.Noises are added to both the original data and ACF, and the accuracy of the original data is compared to the using ACF.

Step 1. The data to be trained is replaced with ACF from the original data.

Step 2. ACF is trained on the 1D-CNN.

Step 3. ACF is trained on the 1d-ResNet.

4.1 Autocorrelation Function

The autocorrelation function (ACF) is defined as the product of time t and data shifted by k from t in time series data. The correlation between the current data and the data shifted by k in the past is examined. ACF measures the relationships between a lagged version of itself over successive time intervals [7]. ACF is expressed by the following Eq. (1).

$$r_k = \frac{\sum_{i=k+1}^{n}(x_t - \overline{x})(x_{t-k} - \overline{x})}{\sum_{i=1}^{n}(x_t - \overline{x})^2} \tag{1}$$

n is the sampling number and k is the time lag. Figure 5 shows the example of the original data. Figure 6 shows ACF of the original data. These images also show amplitude on the vertical axis and time on the horizontal axis.

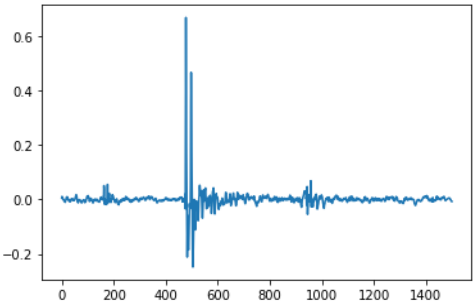

Figure 5: Original data.

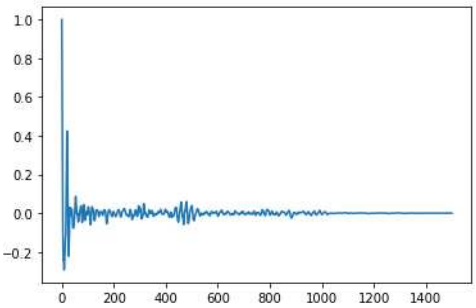

Figure 6: The autocorrelation function (ACF).

4.2 Noise

In the field of time series analysis, noise refers to not necessary data that is not subject to analysis. The greater the amount of noise, the less accurate the analysis becomes. In this case, we used three types of noise: white noise, pink noise, and red noise. These noises are added to the training data to train the 1d-ResNet.

4.2.1 White Noise

White is noise mixed into data such as voice, and its energy is uniformly mixed in all frequency bands. This noise is added to the training data to train the 1d-ResNet. Figure 7 shows the example of White noise.

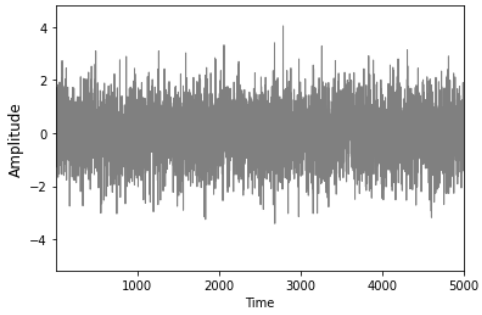

Figure 7: White Noise.

4.2.2 Pink Noise

Pink noise is noise whose energy is inversely proportional to its frequency, and is often represented by the sound of rain or a television sandstorm. Figure 8 shows the example of Pink noise.

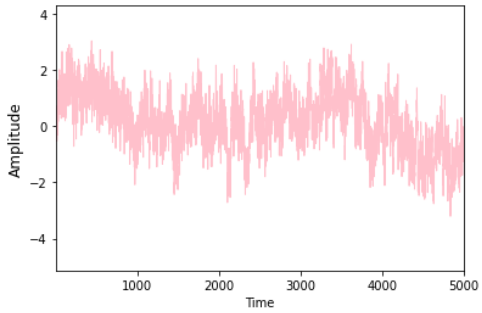

Figure 8: Pink Noise.

4.2.3 Red Noise

Red noise is also known as Brownian noise. A noise that is strong in energy at low frequencies and becomes less powerful at higher frequencies inversely proportional to the square of the frequency. Figure 9 shows the example of Red noise.

5. Verification Structure

In this study, two types of classification accuracy are tested with each of the three types of noise using the original data for training data and using ACF, and the original data for test

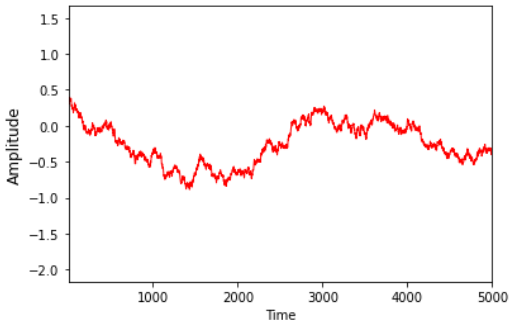

Figure 9: Red Noise.

data. In addition, a dropout layer is used to prevent overtraining. 1d-ResNet is used as the classification model. 1d-ResNet solves the gradient loss problem, which is a problem of multi-layer networks, by residual learning. Figure 10 shows the structure of ResNet. Convolution(1, 128) means convolution layer with 1 filter and 128 channel.

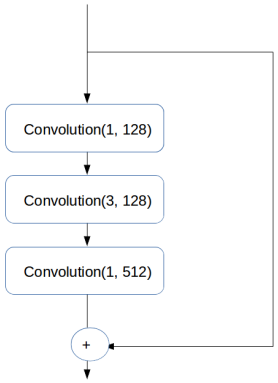

Figure 10: Structure of 1d-ResNet.

6. Simulation Results

Tables 2, 3 and 4 show the classification accuracy of ACF and original data whose three kinds of noise rate is verified from 0.0 to 0.5 in increments of 0.1. These Tables show that for white noise, the use of ACF increased the classification accuracy for all rates. For pink noise, the classification accuracy is higher only at the rate of 0.0 and 0.4, and lower at other rates. For red noise, the classification accuracy is lower than the original data expect for noise rate of 0.3. From the above, it is found that the use of ACF to improve the classification accuracy of data contaminated with noise is effective

for white noise, but not for pink and red noises.

Table 2: Test accuracy of original and ACF(White Noise).

Rate	Original[%]	ACF[%]
0.0	85.33	85.68
0.1	75.80	80.18
0.2	73.41	77.35
0.3	67.08	77.80
0.4	63.52	71.13
0.5	62.04	68.88

Table 3: Test accuracy of original and ACF(Pink Noise).

Rate	Original[%]	ACF[%]
0.0	81.65	83.00
0.1	71.24	71.44
0.2	68.92	64.53
0.3	62.84	59.11
0.4	59.95	61.52
0.5	58.62	56.48

Table 4: Test accuracy of original and ACF(Red Noise).

Rate	Original[%]	ACF[%]
0.0	66.88	57.04
0.1	61.07	55.57
0.2	60.07	58.68
0.3	57.20	57.76
0.4	59.61	54.09
0.5	57.19	55.02

7. Conclusions

In this study, 1d-ResNet is trained with training data containing white noise, pink noise, and red noise using an ACF. The results showed that ACF is effective in increasing the classification accuracy for white noise, while the use of ACF decreased the classification accuracy for pink and red noise. In future studies, We will seek ways to increase classification accuracy even with pink and red noise.

References

[1] F. Rosenblatt,“ The preception : a Probablistic Model for Information Storage and Organization in the Brain ”,*Psychol. Rev*, Vol.65, no.6, pp. 386-408, 1958.

[2] A. Sehgal, N. Kehtarnavaz, “ A Convolutional Neural Network Smartphone App for Real-Time Voice Activity Detection ”, *Access IEEE*, vol. 6, pp.9017-9026, 2018.

[3] T. Frugaki, Y. Uwate and Y.Nishio, “ Time Series Classification Using Autocorrelation Function as Training Data in 1D-CNN ”, *Journal of Shikoku-Section Joint Convention of the Institutes of Elecrical and Related Engineers*, no. 1-6, p.6, Sep. 2021.

[4] M. Matsugu, K. Mori, Y. Mitari and Y. Kaneda,“ Subject Independent Facial Expression Recognition with Robust Face Detection Using a Convolutional Neural Network ”, *Neural Networks 16*, vol.5, pp.555-559, 2003.

[5] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner,“ Gradient-Based Learning Applied to Document Recognition ”, *Proceedings of the IEEE*, vol.86, pp.2278-2324, Nov. 1998.

[6] K. He, X. Zhang, S. Ren and J. Sun, “ Deep Residual Learning for Image Recognition ”,*IEEE International Conference on Computer Vision and Pattern Recognition*, arXiv:1512.03385, 10 Dec. 2015.

[7] T. Furugaki, Y. Uwate and Y. Nishio, “ Time Series Classification Using Autocorrelation Function as Training Data in 1D-CNN ”, *Jouenal of Shikoku-Section Joint Convention of the Institutes of Electrical and Related Engineers*, 21, Sep. 2018.