1-16

Voice Classification Using Surrogate Data with 1D-Convolutional Neural Networks

Tomiyuki FURUGAKI　Tomoya TAKATA
Yoko UWATE　Yoshifumi NISHIO
(Tokushima University)

1. Introduction

Convolutional Neural Network (CNN) was devised based on neurophysiological on the visual cortex of the brain of the creature. It has attracted attention mainly for use in image recognition. However, in recent years, CNN has also been use to audio signal processing. In this research, we used 1D-CNN as a classification model, we focus on which part of time series data is effective for using 1D-CNN by using surrogate data.

2. Proposed Method

In this research, we classify three types of voices. Each time series data is sampled at a sampling frequency 3000. These data are classified using CNN. Three types of surrogate data are created. The following explanations (a), (b) and (c) describe how to create three types of surrogate data.

(a) Random Shuffle Surrogate Data (RSSD)

$x(n)$ means time function. n means time. It is RSSD data that changes the order of n at random.

(b) Fourier Transform Surrogates Data (FTSD)

$$X(\omega) = \sum_{n=1}^{n} x(n) e^{-i\frac{2\pi kn}{N}} \tag{1}$$

$$x(n) = \frac{1}{N} \sum_{n=1}^{n} X(\omega) e^{i\frac{2\pi kn}{N}} \tag{2}$$

Equations (1) and (2) show discrete Fourier Transform (DFT) and Inverse Discrete Fourier Transform (IDFT). k means frequency. N (= 15000) means the number of the samples.

Step 1. Calculate DFT $X(\omega)$ of $x(n)$.

Step 2. Randomize the phase of $X(\omega)$.

Step 3. Calculate IDFT randomized $X(\omega)$.

FTSD is made in this way.

(c) Amplitude Adjusted Fourier Transform Surrogates Data (AAFTSD)

Step 1. Prepare random numbers $R(n)$ according to the standard normal distribution.

Step 2. Sorting $R(n)$ in the same size relation as $x(n)$.

Step 3. Create $R'(n)$ which is FTSD of sorted $R(n)$.

Step 4. Sorting $x(n)$ in the same size relation as $R'(n)$.

AAFTSD is made in this way.

3. Simulation Result

Figure 1 shows the structure of CNN which we use in this study. It should be noted that the dimension of the input layer is one dimension.

Forty pieces of data for 6 seconds each are prepared. After sampling them, surrogate is created. Furthermore, each data is divided into three pieces. The number of the training time series data is 90. The number of the test time series data is 30. Each data is classified with 1D-CNN. Table 1 shows the results of our research. We investigate the average of ten times of test accuracy. Test accuracies of RSSD and AAFTSD are lower than test accuracy of the original time series data. Test accuracy when using FTSD is similar to that when using the original time series data.

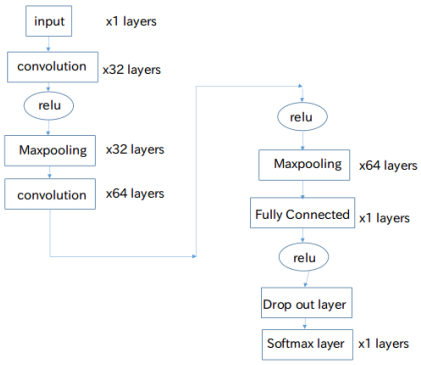

Figure 1: Structure of CNN.

Table 1: Test accuracy

	test accuracy(%)
original data	70.9
RSSD	58.2
FTSD	71.1
AAFTSD	60.2

4. Conclusion

In this study, we carried out 3-value classification with surrogate data. Then, we understood that test accuracies of RSSD and AAFTSD were lower than test accuracy when using the original time series data. RSSD and AAFTSD did not store power spectrum. Therefore, it was understandable that 1D-CNN recognized power spectrum.

However, we do not know how much the accuracy is reduced by the similarity to the power spectrum of original data. In the future, we will choose more similar power spectrum of surrogate data than that of AAFTSD.