Influence of Surrogate Data on Time Series Classification by Neural Network

Tomiyuki Furugaki†, Tomoya Takata†,
Yoko Uwate†, Yoshifumi Nishio† and Hidekuni Takao‡
Tokushima University, 2-1 Minami-Josanjima, Tokushima-shi, Tokushima-ken 770-8506, Japan†
Email: {furugaki, tomoya, uwate, nishio}@ee.tokushima-u.ac.jp†
Kagawa University, 2217-20 Hayashi-Cho, Takamatsu-shi, Kagawa-ken 761-0396, Japan‡
Email: takao@eng.kagawa-u.ac.jp‡

***Abstract*—Neural Networks (NN) can classify time series data. One of the challenges is understanding what the NN depend on for classification. In this study, 1-Dimensional Convolutional Neural Network (1D-CNN) is used as a classification models of surface shape of material. In this study, we focus on the relationship between autocorrelation, frequency distribution and test accuracy for using 1D-CNN with surrogate data.**

I. Introduction

Neural Networks (NN) are a system modeled on neurons of the human brain nervous system. Among them, 1-Dimensional Convolutional Neural Network (1D-CNN) is used for time series classification. 1D-CNN needs to learn time series data in advance to classify. There are various characteristics depending on each time series source. 1D-CNN classifies by finding them. However, the specific judgment part is unknown. Therefore we need to find these.

In this study, surrogate data method is used. The surrogate data method is creating surrogate data. Surrogate data is data that preserves some of the statistical properties of time series data and destroys other properties. After that, it is indicated that there is a significant difference between the statistical properties of time series data and the surrogate data. In this way, the method proves the importance of destroyed properties.

In this study, the data that the 1D-CNN learns is replaced from the original data with surrogate data. In this way, we can find out which part of the time series data is important.

II. Convolutional Neural Network

The research on CNN was established as an academic field in 1956. Since then, it has repeated the ice ages and booms many times and now reaches the present. Currently, CNN is diverse in medical fields, car fields, home electronics fields and so on. The beginning of these booms was image recognition. CNN is inspired from the biological process and conceived from the arrangement of the visual cortex of animals. In the field of image recognition, CNN has achieved tremendous performance with many tasks. In addition, CNN is attracting attention. In particular, the intermediate layer of CNN extracts high versatility and splendid feature quantities. The network structure of CNN is divided into an input layer, an intermediate layer and an output layer. The intermediate layer includes convolution layers, pooling layers and fully connected layers. Features of inputs are extracted in the convolution layer, and position invariance is acquired in the pooling layer. Next, it becomes the 1-Dimensional array in fully connected layers and it changes to probability. Finally, CNN outputs classification results by the probability. However, in recent years, CNN has also been use to time series data. It is 1D-CNN. To date, Recurrent Neural Network (RNN) has been the mainstream for learning time series data. In this study, CNN is used for time series data that is one-dimensional data.

III. Surrogate data method

The surrogate data method was proposed in 1992 for chaos time series analysis. There are no necessary and sufficient conditions for chaos. Therefore, the only way to determine chaos is to determine that there is chaos. In many cases, chaos is determined by spectral continuity, strange attractors, Lyapunov exponents, bifurcations, and so on. However, it has been pointed out that even with random noise alone. The Lyapunov exponent is positive and noise and chaos cannot be distinguished. Therefore, the surrogate method was proposed to test whether it is noise or data generated from a deterministic system. hypothesis test, it is difficult to say that it is noise if a data passes the test. However, it cannot be asserted that it is chaos because, the surrogate method is based on hypothesis testing in statistics. In this study, surrogate data was created and compared the accuracy of learning surrogate data with the accuracy of learning original data. In this way, we researched that characteristics of the original data were important.

IV. Dataset

In this study, time series data of surface profile of the material were classified. Three types of surface profile were prepared. In this study, they were called data 1, 2 and 3. Fifteen pieces of data for 20 second were prepared. Each time series data was sampled at a sampling frequency 3000 [Hz]. Next, the data was agumentationed. There were two types of augmentation. Figure 1 shows the examples of the original time series data. Figure 2 and 3 show the examples of the time series data about the data was added time shift and time stretch.

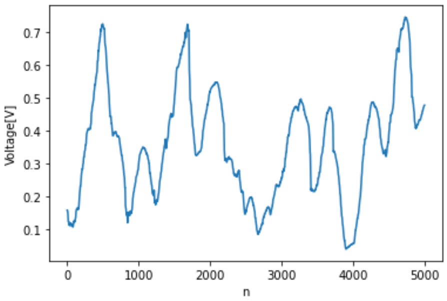

Fig. 1. Original data.

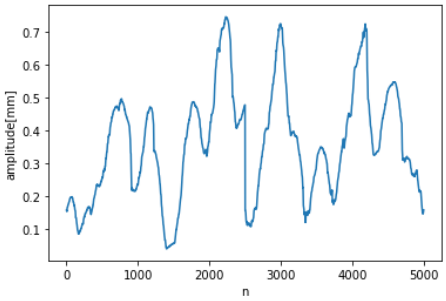

Fig. 2. Time shift data.

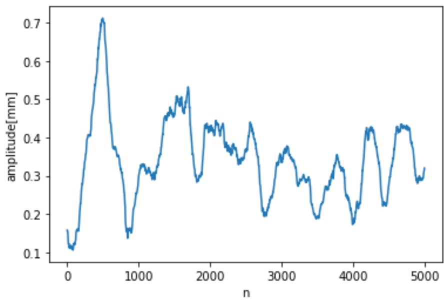

Fig. 3. Time stretch data.

V. Proposed Method

Four types of surrogate data were created. Surrogate data was destroyed some information. The following explanations (a), (b), (c) and (d) describe how to create four types of surrogate data. (a) Random Shuffle Surrogate Data (RSSD) $x(n)$ means time function. n means time. The order of n is changed randomly by the RSSD data. RSSD has broken the autocorrelation of the data. Figure 4 shows RSSD of the Fig. 1

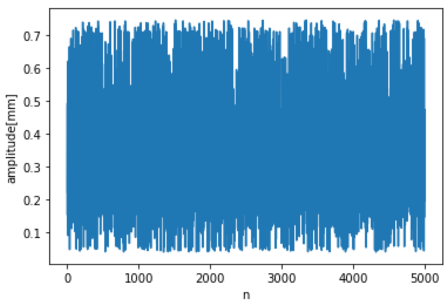

Fig. 4. RSSD.

(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 (= 5000) 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 was implemented using the previous steps. FTSD has broken the frequency distribution of the data. Figure 5 shows FTSD of the Fig. 1.

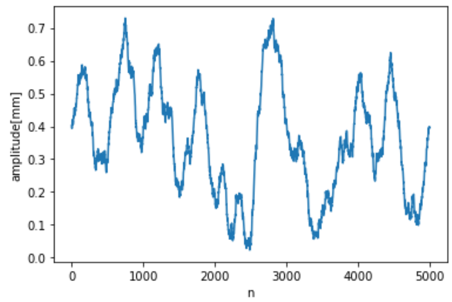

Fig. 5. FTSD.

(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 was implemented using the previous steps. AAFTSD has broken the autocorrelation. However AAFTSD has similar autocorrelation than that of original. Figure 6 shows AAFTSD of the Fig. 1.

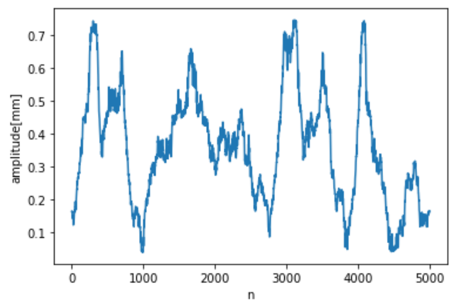

Fig. 6. AAFTSD.

(d) Iterated Amplitude Adjusted Fourier Transform Surrogates Data (IAAFTSD)

Step 1. Prepare $s^{(0)}$ which is RSSD of original data as the initial value.
Step 2. Calculate DFT $S^{(i)}$ of $s^{(i)}$.
Step 3. Replace amplitude of $S^{(i)}$ with amplitude of original. Put it as $\overline{S}^{(i)}$
Step 4. Calculate IDFT $\overline{s}^{(i)}$ of $\overline{S}^{(i)}$.
Step 5. Sorting $\overline{s}^{(i)}$ in the same size relation as original data.
Step 6. Add 1 to i.
Step 7. Increase the number of i.

IAAFTSD was implemented using the previous steps. IAAFTSD saves the frequency distribution and has similar autocorrelation than that of original than that of AAFTSD. The more i, the closer the autocorrelation of the IAAFTSD approaches that of the original. Figure 7 shows IAAFTSD of the Fig. 1.

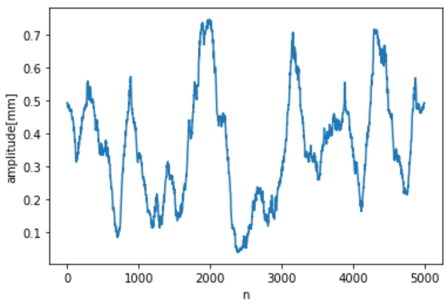

Fig. 7. IAAFTSD.

VI. Architecture

1D-CNN was used for convention architecture. Figure 8 shows the structure of 1D-CNN we used.

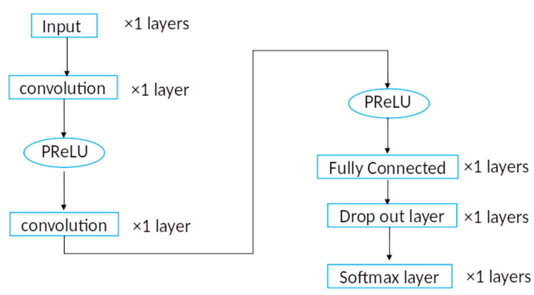

Fig. 8. Structure of 1D-CNN.

Two convolutional layers were used. Drop out layer was prepared to prevent over learning. Classification results is derived the probability by the softmax activation function. Parametric Rectified Linear Unit (PReLU) function was prepared for the activation function. Equation (3) shows softmax function. Equation (4) shows PReLU function.

$$\rho(x) = \frac{exp(x_j)}{\sum_{i=1}^{n} exp(x_i)} \tag{3}$$

$$f(\alpha, x) = \begin{cases} \alpha x & (x < 0) \\ x(x \geq 0) \end{cases} \tag{4}$$

$\rho(x)$ was the probability of being classified as j. n was the total number of classes. α was a parameter that is determined by learning. The PReLU function learns values during training. Therefore, it can better adapt to weights and biases. Figure 9 shows PReLU function when $\alpha = 0.02$.

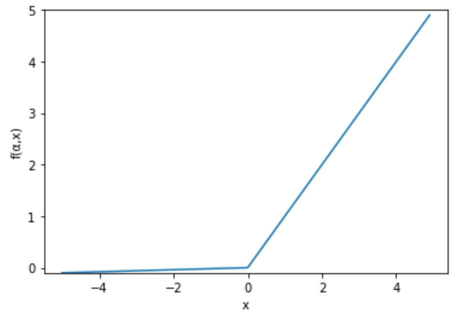

Fig. 9. PReLU function.

VII. Simulation Results

Fifteen pieces of data were prepared. After sampling them, each data was divided into four pieces. Furthermore, surrogate data was created. Table 1 shows the number of the train data and test data. Table 2 shows the results of our research.

TABLE I
The number of the train data and test data

	train data	test data
data 1	336	84
data 2	336	84
data 3	336	84

TABLE II
Test accuracy

	test accuracy
original data	47.77778
RSSD	39.72210
FTSD	54.49840
AAFTSD	52.14286
IAAFTSD($i = 7$)	54.12698

Table 2 shows ten-times-averaged test accuracy results. Table 2 shows test accuracy of RSSD is lower than that of the FTSD. Test accuracies of IAAFTSD, AAFTSD and FTSD are higher than that of original.

Figure 10 shows the relationship between test results and autocorrelation relative error of surrogate data.

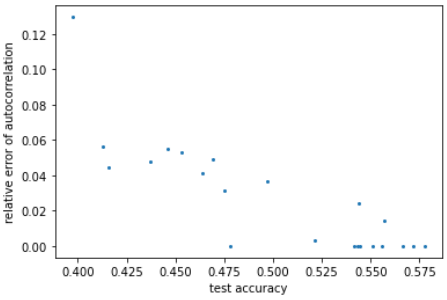

Fig. 10. Relative autocorrelation and test result.

$$D_r = \frac{\sum_{f=0}^{N/2} (Y_2(f) - Y_1(f))^2}{\sum_{f=0}^{N/2} Y_1(f)^2} \tag{5}$$

Equation (5) denotes autocorrelation relative error of surrogate data. $Y_1(f)$ means fourier transform of original data. $Y_2(f)$ means fourier transform of surrogate data. First, create surrogate data of the original data. Next, create the surrogate data of the previous surrogate data. In this way, the data was created with only the autocorrelation gradually destroyed. Figure 10 shows test accuracy will increase when the autocorrelation of the original data is broken just a little. However, test accuracy will decrease when the autocorrelation of the original data is broken larger.

VIII. Conclusion

In this study, three types classification was carried out with surrogate data. Then, we understood that test accuracy of RSSD is lower than that of the FTSD. RSSD has broken the autocorrelation of the data. FTSD has broken the frequency distribution of the data. Therefore, it was understandable that autocorrelation is more important than than frequency distribution for 1D-CNN. Furthermore, test accuracy will increase when the autocorrelation of the original data is broken just a little. However, test accuracy will decrease when the autocorrelation of the original data is broken larger.

The future challenge is to find new indicators that relate to the test accuracy. Test accuracy will increase when the autocorrelation of the original data is broken just a little. However, it is not obvious that why has test accuracy improved.

References

[1] James Theiler, Paul Linsay, and David M.Rubin. ”Detecting Nonlinearity in Data with Long Coherence Times”. In Andreas S.Weigend and Neil A.ershenfeld, editors, ”Time Sereies Prediction Forecasting the Future and Understandung the Past”, pp.429-445. Addison Wesley,1993. Proceedings Volume In The Santa Fe Institute Studies In The Science Of Complexity.

[2] James Theiler, Bryan Galdrikian, Andr Longtin, Stephen Eubank, and J. Doyne Farmer. ”Using Surrogate Data to Detect Nonlinearity in Time Series”. In Martin Casdagli and Stephen Eubank, editors, ”Onlinear Modeling And Forecasting”, pp. 163-188. Addison Wesley, 1992. A Proceed- Ings Volume In The Santa Fe Institute Studies In The Science Of Complexity.

[3] Daniel T. Kaplan and Richard J. Cohen. ”Is Fibrillation Chaos?”. Circulation Research, Vol. 67, No. 4, pp. 886-892, October 1990.

[4] James Theiler. ”Some Comments on the correlation dimension of $1/f^{\alpha}$ noise Physics LetterS A, Vol. 155, No. 8,9, pp. 480-493, May 1991.

[5] James Theiler, Stephen Eubank, Andre Longtin, Bryan Galdrikian, and J. Doyne Farmer. ”Testing for nonlinearity in time series: the method of surrogate data” Physica D, Vol. 58, pp. 77-94, 1992.

[6] Dean Prichard and James Theiler. ”Generating Surrogate Data for with Several Simultaneously Measured Variables”. Physical Review Letters, Vol. 73, No. 7, pp. 951-954, August 1994.

[7] Thomas Schreiber and Andreas Schmitz.”Improved Surrogate Data for Nonlin- earity Tests”. Physical Review Letters, Vol. 77, No. 4, pp. 635-638, 1996.