

Performance of Neural Networks with Surrogate Data for Classifying time series data

Tomiyuki FURUGAKI[†] Tomoya TAKATA[†] Yoko UWATE[†]
Yoshifumi NISHIO[†] Hidekuni TAKAO[‡]
[†](Tokushima University) [‡](Kagawa University)

1. Introduction

Neural Networks (NN) can classify time series data. One of the challenges is understanding what the NN depend on for classification. In this study, NN is used as a classification models. The relationship between autocorrelation, frequency distribution and test accuracy for using NN with surrogate data focus on.

2. Proposed Method

In this research, three types of time series data are classified using NN. Each time series data is sampled at a frequency of 1000[Hz]. Four types of surrogate data are created. The following methods (a), (b), (c) and (d) describe how to build these 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.

(b) Fourier Transform Surrogates Data (FTSD)
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.

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

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

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. It breaks the frequency distribution of the samples.

(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 also breaks the autocorrelation of the data but to a lesser extend than the RSSD method.

(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 $\bar{S}^{(i)}$

Step 4. Calculate IDFT $\bar{s}^{(i)}$ of $\bar{S}^{(i)}$.

Step 5. Sorting $\bar{s}^{(i)}$ in the same size relation as original data.

Step 6. Add 1 to i .

Step 7. Repeat until $i = \infty$.

Using this method, the data stays autocorrelated to a better degree than with the AAFTSD method. Even better results are achieved as i increases.

3. Simulation Result

In this study, three time series data are classified. Six hundred pieces of data that have 5 seconds each are prepared. Table 1 shows ten-times-averaged test accuracy results. It can be seen that RSSD test accuracy is lower than that of the other time series data, while the other methods present test accuracies similar with the original time series data.

Table 1: Test accuracy

	test accuracy(%)
original data	88.9
RSSD	34.5
FTSD	87.5
AAFTSD	88.8
IAAFTSD	89.0

Figure 1 shows the relationship between test results and autocorrelation relative error of surrogate data. While the AAFTSD method leads to an autocorrelation relative error of 0.0017, the RSSD method shows an autocorrelation relative error of 0.1200. In order to obtain data with autocorrelation relative error in between those two values, the AAFTSD method is consecutively applied several times to the samples to obtain desired deterioration of the autocorrelation. Figure 1 shows if the relative error exceeds about 0.02, the accuracy of the test is significantly reduced.

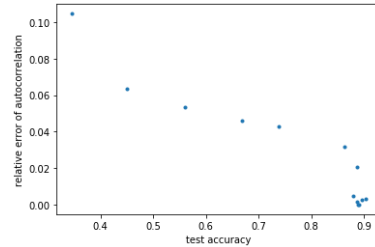


Figure 1: relative autocorrelation and test result

4. Conclusion

In this study, we carried out a 3-value classification of surrogate data. We understood that test accuracy of FTSD was about the same as that of original time series data. Therefore, it was understood that NN didn't recognize frequency distribution. In addition, if the relative error exceeds about 0.02, the accuracy of the test will be significantly reduced.