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

Time Series Classification Using Neural Networks with Chaotic Feature Extraction For Multiple Data

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: {shimizu, uwate, nisho}@ee.tokushima-u.ac.jp

Abstract

Time series classification is an important and challenging problem in data analysis. Recently, time series analysis using neural networks (NN) has attracted much attention. However, the analysis of time series data with complex oscillations is difficult. Therefore, it is important to search for effective features of the data. In this study, we transform the dimensionality of the data and search for features suitable for NN classification. The effect of the appropriate dimension on classification accuracy and the characteristics of the data are being investigated.

1. Introduction

In recent years, there has been a great deal of research on the analysis of time series data. This real-world data has various characteristics, such as not only periodic oscillation but also random oscillation. Among them, chaos theory deals with data that obey deterministic laws. However, in order to observe chaos, it is necessary to transform the data. One of the methods is the time-delay coordinate system, which is a kind of chaos theory [1]. Using this method, one dimensional data (1d-data) can be transformed into multi-dimensional data called attractors. Many phenomena in the real world are represented by attractors, which can be treated as mathematical models. Attractors are also applied to systems, physical phenomena, and economic phenomena in the living body [2]. In recent years, there is a lot of research on time series data analysis using neural networks (NN) such as 1-dimensional convolutional neural networks (1d-CNN) and recurrent neural networks (RNN). The advantage of NN is that they can capture features by themselves. However, even with these models, it is difficult to learn numerical and time-series features of time series data with very complex oscillations. Therefore, data preprocessing is important in NN. In this research, the data are extended to multi-dimensional data using a time-delay coordinate system and compressed to 1d-data using a dimensionality reduction method. The goal is to find features that are easy for NN to learn from time series data with complex oscillations by selecting the optimal dimension.

2. Convolutional Neural Network

Convolutional Neural Network (CNN) is one of the NN and is mainly used in image recognition [3]. In 1982, Fukushima and Miyake proposed ”New Recognition TRON”, the predecessor network of CNN [4]. However, CNN are required high computational power. To compensate for these shortcomings, in 2015, Kiranyaz proposed the first compact and adaptive 1d-CNN that works directly with patient-specific ECG data [5]. Today, 1d-CNN have quickly achieved state-of-the-art performance levels in several applications such as biomedical data classification and early diagnosis, structural health monitoring, anomaly detection and identification in power electronics and electrical [6]-[8]. CNN is a model with a hierarchical structure that overlaps layers called convolutional layers, and the accuracy has been improved by deepening this hierarchical structure. In 1998, Le Cun Laba LeNet-5 has five layers. Categorizing handwritten characters using CNN. In 2012, eight-layer AlexNet won the ILSVRC image classification competition In the 2014 competition, VGG Net 19. GoogleNet 22 layers have been layered to further improve accuracy. In the 2015 competition, the Residual Network (ResNet) layered 152 layers and won the competition of the year. Now ResNet is popular as one of the standard CNN models [9]. In this study, 1d-ResNet is used for time series classification.

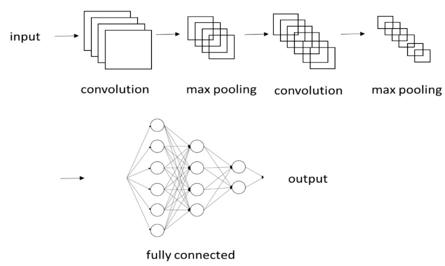

Figure 1: An example of CNN structure.

3. Dataset

The UCR repository [10] data used in this study is the EMG (EMG) of hand movements [11]. The participants were asked to repeat the following three movements, which are considered basic hand movements. Data (a) is an action for grasping with the palm of the hand. Data (b) are movements for holding thin and flat objects. Data (c) is for holding a cylindrical tool. An example of the EMG time series data for these three movements is shown in Fig. 2.

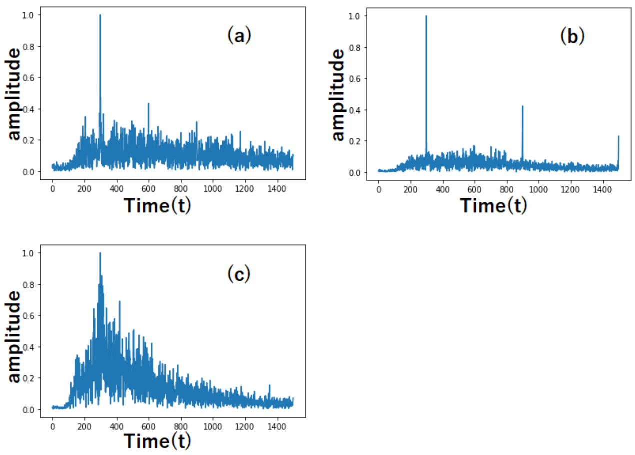

Figure 2: Three hand movements.

4. Proposed Method

In this study, we propose a method of feature translation for using multidimensional features of time series data. Figure 3 shows the flow of the proposed method of this study.

1. Calculate the optimal dimension.
2. Transformed into three dimensional attractor.
3. Compressed to extract the necessary features.

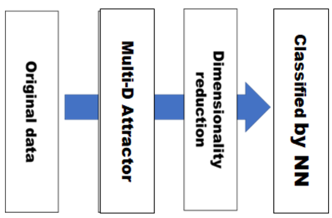

Figure 3: Procedure of the proposed method.

4.1 Attractor construction

Time-delayed coordinate systems are commonly known as dimensional dilation methods by Takens ’ Embedding Theorem. It is a method of dilation of dimensions. Let the value of the data at a certain time be $x(n)$. Further, let the value of time delay be τ. This system is represented by Eq. (1) and is shown in Fig. 4 [12]. In this study, we extended the data to three dimensions.

$$f(x) = [x(n),\ x(n+\tau),\ x(n+2\tau)\ ...] \tag{1}$$

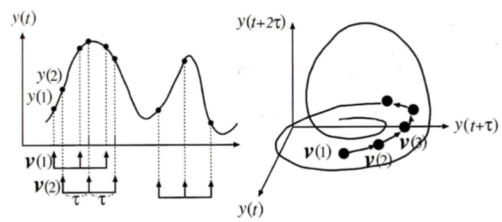

Figure 4: Time delay coordinate system.

4.2 False Nearest Neighbor

The attractor has arbitrary values of dimension D and time delay τ. By setting each of these to an appropriate value, it is possible to extract chaotic features. This study calculates the optimal dimension D using False Nearest Neighbor (FNN).

The basic idea underlying the estimation of embedding dimension using FNN was proposed by Kennel. Suppose two data points in the one-dimensional time series are close together then they are neighbors. Their difference in magnitude provides us with the distance of those neighbors. If we embed the time series once using some time delay τ, then we can use the coordinates of those data points to examine whether the distance between them has changed appreciably. If embedding changes the distance between the neighbors appreciably, then they are dubbed false neighbors, and this indicates that the data need to be embedded further. Usually, this method requires a threshold, but Cao’s proposed averaging method does not threshold [13].

$$E1(\tau) = \frac{E^a(\tau+1)}{E^a(\tau)} \tag{2}$$

$$E^b(\tau) = \frac{1}{N-\tau+1}\sum_{t=0}^{N-\tau} b(t,\tau) \tag{3}$$

A point that satisfies either of these two Eq. (2) and Eq. (3) are an FNN. we choose a value for D at the point where the number of FNN drops to 0, or subsequent embeddings do not change the number of FNN or the point before which the number of FNN starts to increase again.

4.3 Dimensionality reduction

Dimensionality reduction refers to the reduction of high-dimensional feature vectors to low-dimensional vectors while retaining the distribution information of the original dataset. The procedure of Principal Component Analysis (PCA) method is shown in the following steps.

1. Determine the axis that can maximize the variance
2. Determine the axis of the second that is orthogonal to the first
3. Get the subspace as eigenvectors by singular value decomposition
4. Compute lower-dimensional embedding

PCA is one of the most widely used low-dimensional embedding methods. As the name implies, it is an algorithm that searches for the important element principal component in the observed data.

5. Verification Structure

In this study, the 1d-ResNet is used for classification model.

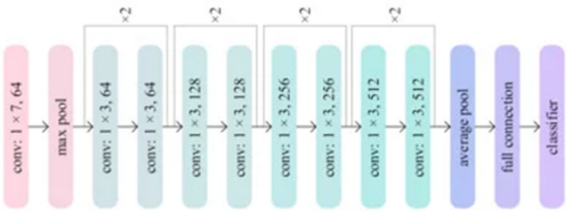

Figure 5: 1d-ResNet structure used in this study.

ResNet is commonly used for the classification of image data. ResNet is the winner network of ILSVRC 2015, and it is the network that enable to have deep layered networks without vanishing gradient problem by using residual learning that is shown in Fig. 5. By inserting the shortcut connection that is expressed by the longest arrow over 3 convolutional layers, previous information is learned, and it avoid vanishing gradient problem. In Fig. 6, Convolution (1×1, 64) indicates convolution layer with 1×1 filter and 64 channel.

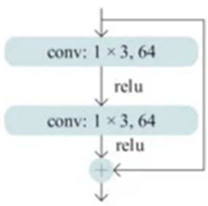

Figure 6: The structure of residual learning.

Table 1 shows the batch size, epochs, and learning rate in this study. Batch size is the amount of data processed at once. The epoch is the number of the learnings. Simulation terminates when val loss no longer decays during 1500 epochs. The learning rate is the degree of learning progress. By setting as shown in Table 1, the training accuracies reach 100%.

Table 1: Learning parameters.

	Batch size	Learning rate	Epoch
1d-ResNet	42	0.0001	Earlystopping

6. Simulation Results

The results of the FNN method at τ= 10 and 25 are shown in the following fig. 7.

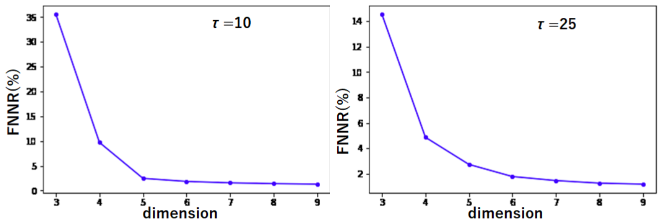

Figure 7: Relationship between FNN ratio and dimension.

It is appropriate for the FNN method to deal with dimensions where the FNN ration (FNNR) decrease begins to converge. The figure 7 shows that at $\tau = 10$, dimensions 5 and 6 are appropriate, and at $\tau = 25$, dimensions 7 and 8 are appropriate.

The test accuracy is the average of five test accuracies. 1d-ResNet is used to investigate the accuracy of 3-value classification. Table 2 shows the accuracy of the original data without processing and the test accuracy of the proposed method with different dimensions.

Table 2: Test accuracy of the proposed method.

	test accuracy	
original	0.871	
dimension	D=3	D=5
$\tau = 10$	0.849	0.839
$\tau = 25$	0.868	0.905

This is consistent with the appropriate value of 8 dimensions by the FNN method. On the other hand, the appropriate values at τ=10 are 5 and 6 dimensions, but the accuracy did not increase. In this study, $\tau = 25$ is considered suitable.

7. Conclusions

In this study, we extended time series data to Multi Dimensional space using time-delay coordinate method and compressed the data. We checked how the accuracy of time series classification by 1d-ResNet changes by optimizing the number of parameters during dimensionality expansion. We also compared the accuracy of 1d-ResNet without processing and with the proposed method. The results showed that dimensionality expansion using the appropriate dimension calculated by FNN extracted important features and improved the accuracy of the test accuracy. In addition, by appropriately compressing the higher dimensional features of the attractors, the features could be used in low dimensional classification. We will use these results to investigate whether they are valid for other data as well. We will also use models such as RNN that are specialized for analyzing the flow of time series or construct optimal extended dimensions.

References

[1] F. Takens,“Detectingstrange attractors intur bulance”, LectureNotes in Mathematics, Vol.898(Springer-Verlag.), pp.366-381, 1981.

[2] C. Oestreicher,“A history of chaos theory”, doi: 10.31887/DCNS.2007.9.3/coestreicher, Dialogues Clin Neurosci, pp.279289, 2007.

[3] M. Matsugu, K. Mori, Y. Mitari and Y. Kaneda, “Subject independent facial expression recognition with robust face detection using a convolutional neural network”, Neural Networks16, vol.5, pp.555-559, 2003.

[4] K. Fukushima, S. Miyake,“A new algorithm for pattern recognition tolerant of deformations and shifts in position”, Pattern Recognition, Neocognitron15, pp.455-469, 1982.

[5] S. Kiranyaz, T. Ince, R. Hamila, M. Gabbouj, “Convolutional Neural Networks for patient specific ECG classification”, in Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS, pp.2608-2611, 2015.

[6] S. R. Bajare, V. V. Ingale, “ECG Based Biometric for Human Identification using Convolutional Neural Network”, Computing Communication and Networking Technologies (ICCCNT) 2019 10th International Conference on, pp.1-7, 2019.

[7] P. Sharma, S. Chandan, B. P. Agrawal, “Vibration Signal-based Diagnosis of Defect Embedded in Outer Race of Ball Bearing using 1-D CNN”, Computational Performance Evaluation (ComPE) 2020 International Conference on, pp.531-536, 2020.

[8] I. Mitiche, A. Nesbitt, S. Conner, “Philip Boreham, Gordon Morison, “1D-CNN based real-time fault detection system for power asset diagnostics”, Generation Transmission Distribution IET, vol. 14, no. 24, pp.5766-5773, 2020.

[9] K. He, X. Zhang, S. Ren, J. Sun, “Deep residual learning for image recognition”, InProceedings of the IEEE, conference on computer vision and pattern recognition, pp.770-778, 2016.

[10] H. A. Dau, A. Bagnall, K. Kamgar, C. M. Yeh, Y. Zhu, S. Gharghabi, C. A. Ratanamahatana, E.Keogh, “The UCR Time Series Archive”, IEEE/CAA J. Autom. Sinica, vol. 6, no. 6, pp.1293-1305, Nov. 2019.

[11] C. Sapsanis, G. Georgoulas, A. Tzes and D. Lymberopoulos, “Improving EMG based Classification of basic hand movements using EMD”, 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Osaka, Japan, pp.5754-5757, 2013.

[12] K. Aihara, T. Ikeguchi, Y. Yamada and M. Komuro, “Fundamentals and applications of chaotic time series analysis”, Industrial book, p.23, 2000.

[13] C. Li, “Practical method for determining the minimum embedding dimension of a scalar time series”, PVLD 110, pp.43-50, 1997.