

Reconstruction of time series with missing value using 2D representation-based denoising autoencoder

TAO Huamin, DENG Qiuqun^{*}, and XIAO Shanzhu

National Key Laboratory of Science and Technology on Automatic Target Recognition, College of Electronic Science,
National University of Defense Technology, Changsha 410073, China

Abstract: Time series analysis is a key technology for medical diagnosis, weather forecasting and financial prediction systems. However, missing data frequently occur during data recording, posing a great challenge to data mining tasks. In this study, we propose a novel time series data representation-based denoising autoencoder (DAE) for the reconstruction of missing values. Two data representation methods, namely, recurrence plot (RP) and Gramian angular field (GAF), are used to transform the raw time series to a 2D matrix for establishing the temporal correlations between different time intervals and extracting the structural patterns from the time series. Then an improved DAE is proposed to reconstruct the missing values from the 2D representation of time series. A comprehensive comparison is conducted amongst the different representations on standard datasets. Results show that the 2D representations have a lower reconstruction error than the raw time series, and the RP representation provides the best outcome. This work provides useful insights into the better reconstruction of missing values in time series analysis to considerably improve the reliability of time-varying system.

Keywords: time series, missing value, 2D representation, denoising autoencoder (DAE), reconstruction.

DOI: 10.23919/JSEE.2020.000081

1. Introduction

Chronological sequences of observations, which are referred to as time series, are widely found in biology, medicine, finance, astronomy and engineering [1]. A time series contains rich information about dynamic systems that can provide essential knowledge for decision makers. Time series analysis has been increasingly used in data mining tasks, such as classification, prediction, clustering, anomaly detection and segmentation [2]. However, missing values are problems frequently encountered during data observation because of technical fault or human

errors, thereby immensely increasing the uncertainty of classification or prediction results. Improper processing of missing values also increases the computing time and complexity of the follow-up algorithm. The data completeness of time series with missing values is crucial to be solved and should be carefully performed without disturbing the time series' statistical properties.

Existing handling methods for missing values [3–7] can be mainly classified into two categories, namely, deletion and imputation. The deletion method is simply conducted by deleting the missing data, resulting in the loss of useful information. The imputation method, which includes statistical and learning methods, is a procedure that replaces the missing values by some certain values. Different imputation methods have different performance in completing missing values. Statistical methods, such as mean imputation, are effective in small amounts of missing values. However, the drawback of this method is the high bias caused by the newly imputed values that are similar with the observed data. Several other methods, referred as learning methods, e.g. genetic algorithm, support vector machine, expectation maximisation algorithm and multi-layer perceptron, have been developed to overcome this problem. The multi-objective genetic algorithm imputation proposed by Lobato et al. [8] is relatively easy to implement and adapt to different domains. The disadvantage of this method is that the fitness calculation may demand high computation time. The expectation-maximisation algorithm proposed in [9] performs well on large fraction of missing data and small samples. However, the parameters need to be initialised and their selection has a vital effect to the convergence. As a classic neural network, multilayer perceptron is widely used to model the complex relationships between data and in time series imputation [10].

Recently, deep neural networks, especially convolutional neural networks (CNNs), have achieved impres-

Manuscript received March 15, 2020.

^{*}Corresponding author.

sive results in various domains [11–14] and many deep learning based methods for missing data imputation have been proposed [15–19]. The most popular ones are generative adversarial nets (GANs) and denoising autoencoders (DAEs). Yoon [15] proposed a novel GAN framework for missing data imputation. However, the network modelling is complicated and the training is difficult. The GAN-based methods often require complete data during training. Different from GANs, DAEs are designed to recover the outputs from noisy inputs through unsupervised learning [20–22], which makes it suitable for data without labels. Missing data is a special case of noisy input, making DAEs ideal as an imputation model. However, missing data can depend on interactions/latent representations that are unobservable in the input dataset space. Silva et al. [23,24] and Wang et al. [25] reformulated the time series as the image for time series classification, thereby providing a useful perspective for the processing of missing values.

In this research, we propose an overcomplete DAE framework using the 2D representation combined with a DAE as a solution to address the abovementioned problem. Missing information can be accurately recovered by projecting the input data into a high-dimensional subspace. The main contributions of this work are summarised in three points.

Firstly, an effective 2D time series representation method is adopted to transform the raw time series into a 2D matrix. The latent relationships between each time stamp can be explicitly represented by projecting the 1D time series into the 2D image.

Secondly, we propose a 2D representation-based DAE to tackle the missing values in the time series, for fully utilising the correlations between observations in the learning process of DAE.

Finally, extensive evaluations are performed to validate that the proposed framework can achieve a significant improvement in missing value estimation and time series classification.

The remainder of this paper is organised as follows. Section 2 describes the proposed methodology in detail. Section 3 presents the experiments and results, and Section 4 provides the discussion. Section 5 summarises the conclusions of this study.

2. 2D representation-based DAE for time series reconstruction

In Subsections 2.1 and 2.2, we describe the concepts of two time series data representations, namely, recurrence plots (RPs) [26] and Gramian angular field (GAF) [25]. The details of the proposed method are presented in Sub-

section 2.3.

2.1 RPs

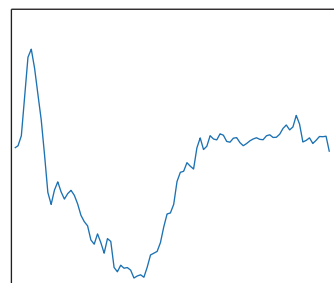
Representing the time series as RPs is proposed by Marwan et al. [27, 28]. Considering that time series reflects the dynamical characteristics of a system that generates it, RPs are a suitable tool for analyzing the recurrences of the states of a dynamic system and for measuring the recurrences of a trajectory $\mathbf{x}_i \in \mathbf{R}^m$ in the phase space, where m is the dimension of the phase space trajectory, which can be formulated as

$$\mathbf{R}_{i,j} = \theta(\varepsilon - \|\mathbf{x}_i - \mathbf{x}_j\|), \quad i, j = 1, \dots, N, \quad (1)$$

$$\theta(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0, \end{cases} \quad (2)$$

where N is the number of measured points \mathbf{x}_i , ε is a threshold distance, $\theta(x)$ is the Heaviside function, and $\|\cdot\|$ is a norm. The \mathbf{R} -matrix contains textures, which are single dots, diagonal lines and vertical and horizontal lines, and typology information, which is characterised as homogeneous, periodic, drift and disrupted. For instance, fading to the upper left and lower right corners indicates that the process contains a trend or a drift, or vertical and horizontal lines/clusters show that some states remain constant or slowly change for some time, which can be interpreted as laminar states [29].

Fig. 1 shows the RP representations for the two datasets obtained from the University of California, Riverside (UCR) time series archive [30], which is a public dataset for time series analysis and includes time series sets from different application domains. Fig. 1(a) and Fig. 1(b) present the raw time series “ECG200” and “Coffee” respectively. Fig. 1(c) and Fig. 1(d) illustrate the scaled colour image visualisation of the \mathbf{R} -matrix, which is calculated on the basis of the closeness of the states in the phase space, using phase space, dimension $m = 3$ and embedding time delay $\tau = 10$.



(a) ECG200

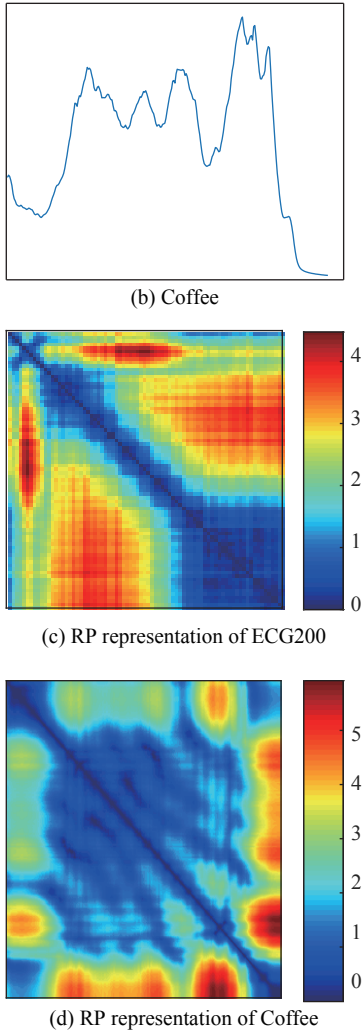


Fig. 1 RP representations of time series

2.2 GAF

GAFs, including Gramian angular summation field (GASF) and Gramian angular difference field (GADF), are techniques used to transform the time series into a 2D matrix [25]. Given a time series $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$ of n point observations, the first step for GAF transformation is to rescale the value into the interval $[-1, 1]$, which can be expressed as

$$\tilde{x}_i = \frac{(x_i - \max \mathbf{X}) + (x_i - \min \mathbf{X})}{\max \mathbf{X} - \min \mathbf{X}}. \quad (3)$$

The second step is to represent the rescaled time series $\tilde{\mathbf{X}}$ in the polar coordinates using (4), where t_i is the time stamp, and N is a constant factor regularising the span of the polar coordinate system.

$$\begin{cases} \phi = \arccos(\tilde{x}_i), & -1 \leq \tilde{x}_i \leq 1; \tilde{x}_i \in \tilde{\mathbf{X}} \\ r = \frac{t_i}{N}, & t_i \in N \end{cases} \quad (4)$$

Fig. 2 illustrates the “Coffee” time series represented in different coordinates, including Cartesian and polar coordinate systems.

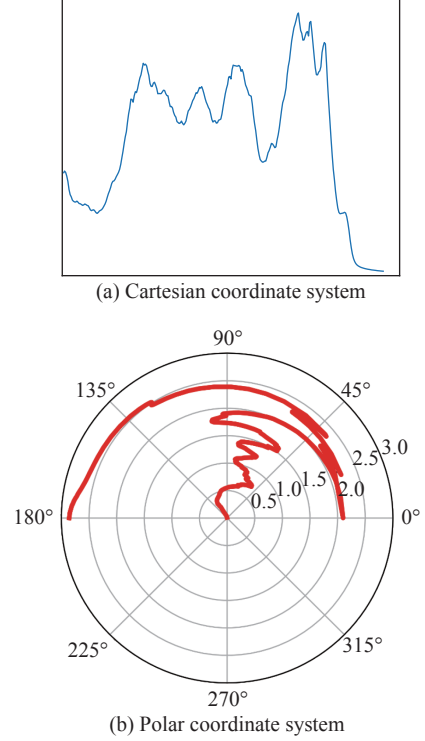


Fig. 2 “Coffee” time series represented in different coordinates

In the polar coordinate system, the temporal correlation between different time intervals can be easily exploited by considering the trigonometric sum/difference between each point. GASF and GADF are defined as follows:

$$\mathbf{GASF} = \begin{bmatrix} \cos(\phi_1 + \phi_1) & \cdots & \cos(\phi_1 + \phi_n) \\ \cos(\phi_2 + \phi_1) & \cdots & \cos(\phi_2 + \phi_n) \\ \vdots & \ddots & \vdots \\ \cos(\phi_n + \phi_1) & \cdots & \cos(\phi_n + \phi_n) \end{bmatrix} = \tilde{\mathbf{X}}' \cdot \tilde{\mathbf{X}} - \sqrt{\mathbf{I} - \tilde{\mathbf{X}}^{2'}} \cdot \sqrt{\mathbf{I} - \tilde{\mathbf{X}}^2}, \quad (5)$$

$$\mathbf{GADF} = \begin{bmatrix} \sin(\phi_1 - \phi_1) & \cdots & \sin(\phi_1 - \phi_n) \\ \sin(\phi_2 - \phi_1) & \cdots & \sin(\phi_2 - \phi_n) \\ \vdots & \ddots & \vdots \\ \sin(\phi_n - \phi_1) & \cdots & \sin(\phi_n - \phi_n) \end{bmatrix} = \tilde{\mathbf{X}} \cdot \sqrt{\mathbf{I} - \tilde{\mathbf{X}}^{2'}} - \tilde{\mathbf{X}}' \cdot \sqrt{\mathbf{I} - \tilde{\mathbf{X}}^2}, \quad (6)$$

where \mathbf{I} is the unit row vector. Defining the inner product $\langle x, y \rangle = x \cdot y - \sqrt{1-x^2} \cdot \sqrt{1-y^2}$ and $\langle x, y \rangle = y \cdot \sqrt{1-x^2} - x \cdot \sqrt{1-y^2}$, GASF and GADF can be expressed as quasi-Gramian matrices $|\langle \tilde{\mathbf{x}}_n, \tilde{\mathbf{x}}_n \rangle|$.

As shown in (7), the GAF matrix is actually a symme-

tric matrix of the inner product. GADF is similar to GASF, except that the former is constructed by using the trigonometric difference of the inverse sine. Therefore, we only consider GASF in this study.

$$\mathbf{GASF} = \begin{bmatrix} \langle \tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_1 \rangle & \cdots & \langle \tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_n \rangle \\ \langle \tilde{\mathbf{x}}_2, \tilde{\mathbf{x}}_1 \rangle & \cdots & \langle \tilde{\mathbf{x}}_2, \tilde{\mathbf{x}}_n \rangle \\ \vdots & \ddots & \vdots \\ \langle \tilde{\mathbf{x}}_n, \tilde{\mathbf{x}}_1 \rangle & \cdots & \langle \tilde{\mathbf{x}}_n, \tilde{\mathbf{x}}_n \rangle \end{bmatrix} \quad (7)$$

Fig. 3 shows the GASF representations for the two

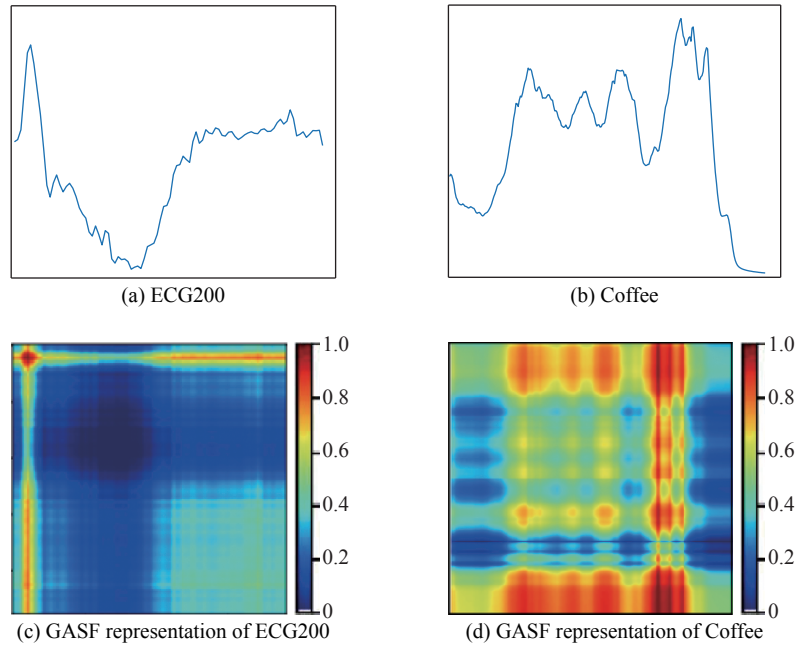


Fig. 3 GASF representations of time series

2.3 2D representation-based DAE

A DAE is widely used for image denoising and recovery, and usually consists of an encoder and a decoder [31]. Unlike autoencoders, the DAE receives the corrupted data as input and outputs the completed data. It can be trained to learn high-level representation of the feature space in an unsupervised manner.

Given input $\tilde{\mathbf{x}}$ and damaged input \mathbf{x} , the encoder maps from an input domain \mathbb{R}^{D_x} to a hidden representation in \mathbb{R}^{D_z} , and the decoder maps from \mathbb{R}^{D_z} back to \mathbb{R}^{D_x} . Therefore, the encoding and decoding functions can be defined as

$$\begin{cases} \mathbf{z} = \phi(\mathbf{x}; \boldsymbol{\theta}_E) \\ \mathbf{y} = \psi(\mathbf{z}; \boldsymbol{\theta}_D) \end{cases} \quad (9)$$

where $\mathbf{x} \in \mathbb{R}^{D_x}$, $\mathbf{z} \in \mathbb{R}^{D_z}$ and $\mathbf{y} \in \mathbb{R}^{D_x}$ denote a sample from the input space, its hidden representation and reconstructed representation, respectively. Encoder $\phi(\cdot)$ and decoder $\psi(\cdot)$ are usually implemented by stacked dense lay-

ers of neurons with the sigmoid activation function. In this study, we replace the dense layers by convolutional layers. Compared with the fully connected layer, the convolutional encoder and decoder can effectively reduce the parameters and computation costs. $\boldsymbol{\theta}_E$ and $\boldsymbol{\theta}_D$ are the trainable parameters of the encoder and the decoder, respectively. DAEs are trained to minimise the variance between input $\tilde{\mathbf{x}}$ and its reconstruction \mathbf{y} . Given that the objective is to reconstruct the missing values, the mean squared error (MSE) is adopted as the loss function, which is represented as

$$E(\boldsymbol{\theta}) = \frac{1}{N} \cdot \|\tilde{\mathbf{x}} - \mathbf{y}\|^2. \quad (10)$$

After training, the network can be applied to unknown data denoising and recovery.

The proposed framework for the recovery of time series with missing values is shown in Fig. 4. Firstly, we transform the raw time series into 2D matrices, namely,

datasets from the UCR time series archive. Fig. 3(a) and Fig. 3(b) present the raw time series “ECG200” and “Coffee” respectively. Fig. 3(c) and Fig. 3(d) illustrate the corresponding GASF representations.

From the main diagonal of GASF, $\{G_{ii}\} = \{\cos(2\phi_{ii})\}$, we can easily reconstruct the original time series using the following equation:

$$\cos \phi = \sqrt{\frac{\cos(2\phi) + 1}{2}}, \quad \phi \in \left[0, \frac{\pi}{2}\right]. \quad (8)$$

RP and GASF representations. Secondly, the 2D representations are used as the input of the DAE, which learns middle representation through the encoder and outputs the reconstructed 2D representation through the decoder. The encoder contains two convolution and two max pooling layers, and the decoder has an architecture that is

symmetric to the encoder. Finally, we reconstruct the original time series from the 2D representations to compare the imputation performance. For the GASF representation, we can reconstruct the recovered time series by extracting the main diagonal.

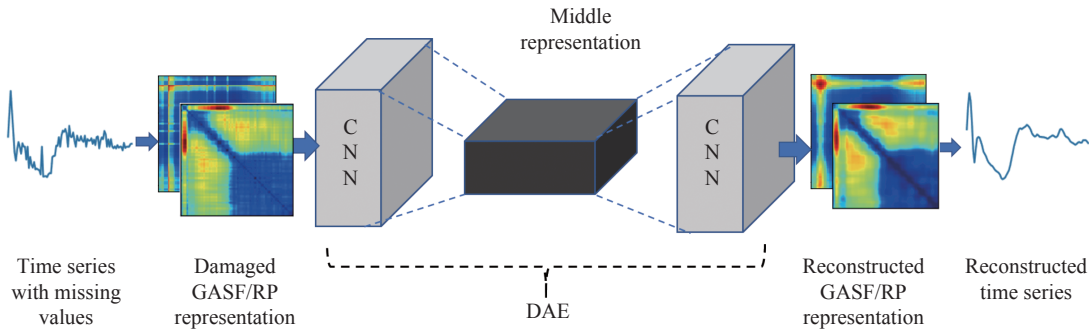


Fig. 4 Framework for recovery of time series with missing values

We downsample the 2D representation to the same size because different types of time series have different lengths. The input of the encoder is a 64×64 matrix. The number of kernels of the first and second convolutional layers are 16 and 32, respectively. Each convolutional layer has a 3×3 receptive field with a stride fixed to 1 pixel. Each max pooling layer has a 2×2 receptive field with a stride fixed to 2 pixels. The decoder operates in reverse direction. The output has the same size with the input. The adaptive moment estimation optimisation algorithm [32] is applied to train the DAE models by minimising the MSE of the inputs and reconstructions. The batch size is 20, the initial learning rate is 0.01, and the learning rate reduces when the MSE remains unchanged for 10 iterations. The decay rate is 0.1, and each experiment runs for 10 times.

3. Experimental results

3.1 Results of missing value estimation

To demonstrate the effectiveness of the 2D representation on missing value imputation, we select the standard DAE with raw time series data as the input to recover the missing values for comparison. The performance of different methods is evaluated on the datasets from the UCR time series classification archive.

Table 1 provides the detailed information about each dataset used in this study, including the number of classes, the time series length and the size of training/validation/test set. The datasets division in Table 1 is set for the classification tasks in Subsection 3.3. We train the DAE using “ECG200” and test it on the remaining sequences. Different types of datasets are chosen for test-

ing, including “Face all”, “Swedish leaf”, “OSU leaf”, “Wafer”, “50 words” and “Coffee”. The damaged inputs are generated by randomly setting 20% of the original sequences amongst specific time series to zero. The proposed method is implemented on TensorFlow and run on NVIDIA GeForce GTX 1080 Ti graphic card.

Table 1 Summary of datasets

Dataset	Class	Length	Training/Validation/Test
ECG200	2	96	100/50/50
Face all	14	131	560/845/845
Swedish leaf	15	128	500/313/312
OSU leaf	6	427	200/121/121
Wafer	2	152	1 000/3 082/3 082
50 words	50	270	450/272/273
Coffee	2	287	28/14/14

Fig. 5 shows the reconstruction of “ECG200” time series using the standard DAE. The first, second and third rows denote the raw time series, time series with missing values and reconstructed time series, respectively. The standard DAE recovers the general tendency of the time series, and most of the missing values are imputed. However, the local fluctuations are smoothed.

Fig. 6 and Fig. 7 illustrate the reconstruction of GASF and RP representations of “ECG200” using DAE, respectively. All the 2D matrices are presented in colour maps. The first, second and third rows denote the

GASF/RP representations of the raw time series, time series with missing values and reconstructed time series,

respectively. They can accurately reconstruct the missing values to a certain degree.

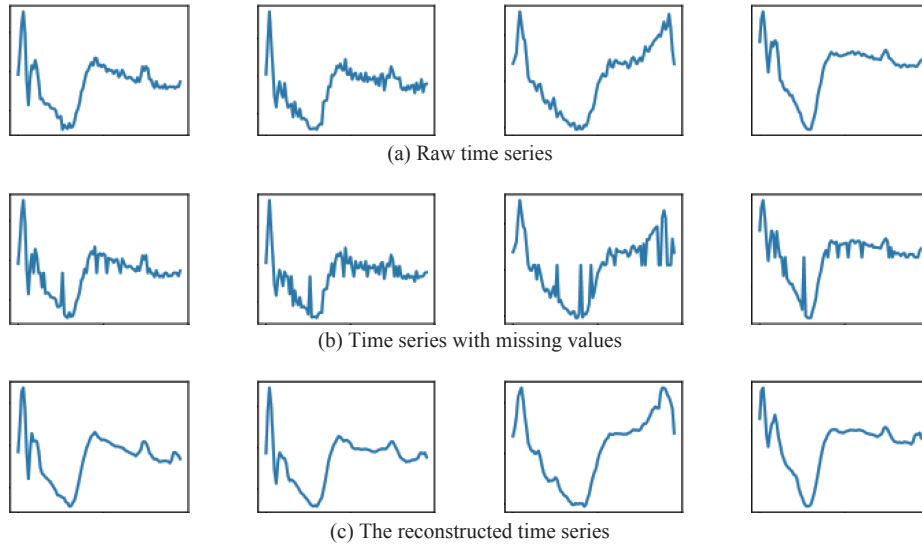


Fig. 5 Reconstruction of “ECG200” time series

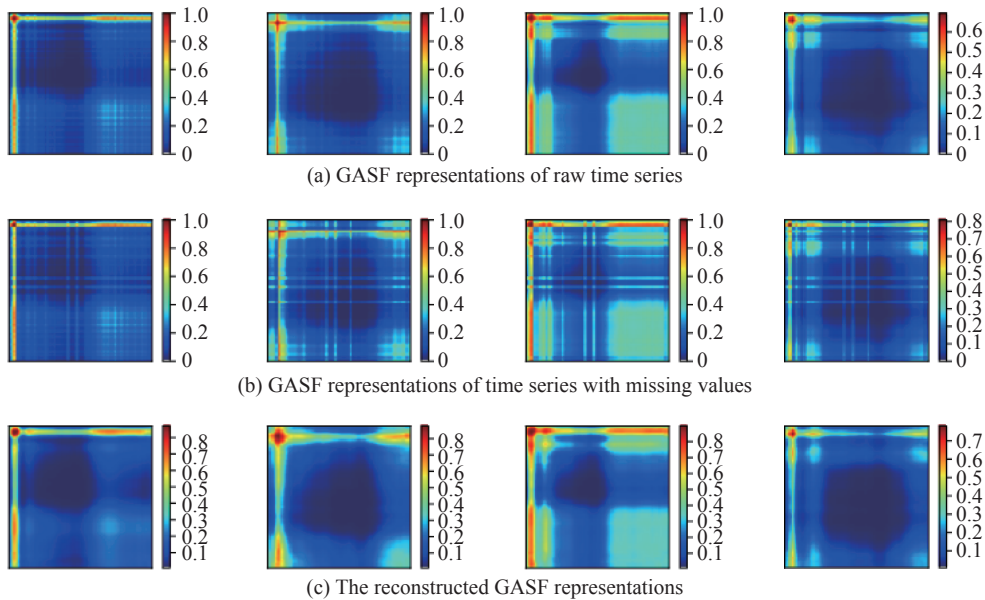


Fig. 6 Reconstruction of GASF representations of “ECG200”

We compute the MSE of the recovered and raw time series of each method to compare the imputation accuracy of three different representations. Given that RP and GASF representations can be compared at the 2D level, we only reconstruct the time series from the restored GASF matrix for comparison with the raw time series. We also compare our proposed method with other reconstruction methods including mean and auto-regression (AR). Mean

imputation replaces the missing values with the mean value of the adjacent observations. AR imputation uses the auto-regression model ($p = 3$) to fill in missing values. The results are displayed in Table 2 and Table 3. The GASF time series in Table 2 denotes the time series restored from the GASF matrix. The GASF and RP time series in Table 3 represent the comparison of the 2D representations. The low reconstruction errors are displayed in bold.

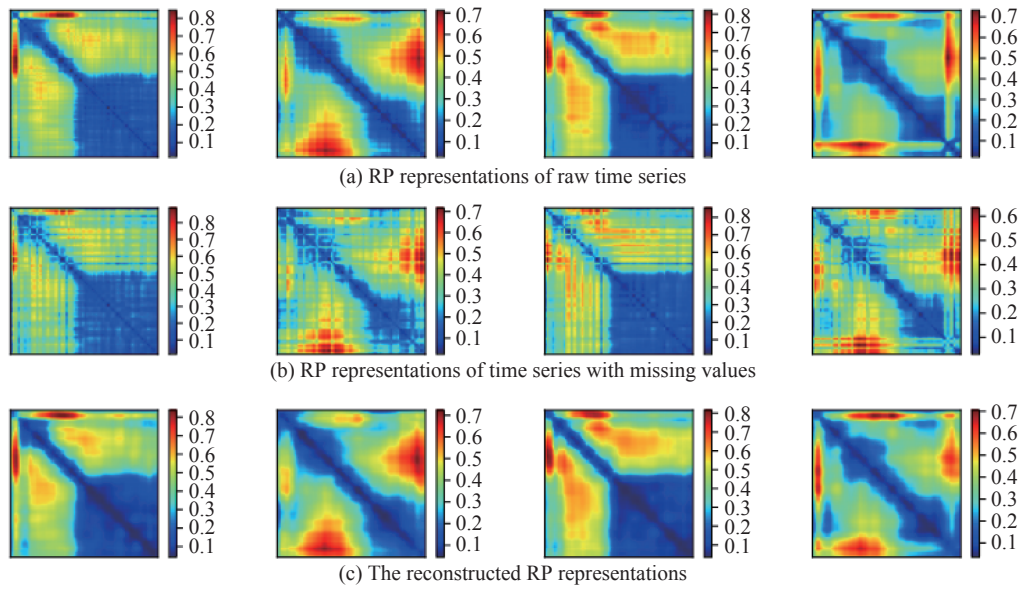


Fig. 7 Reconstruction of RP representations of “ECG200”

Table 2 Comparison on MSE of different reconstruction methods

Dataset	MSE			
	Raw time series	GASF time series	Mean	AR
ECG200	0.000 5	0.000 3	0.018 6	0.031 4
Face all	0.000 7	0.001 7	0.029 9	0.059 2
Swedish leaf	0.001 8	0.000 9	0.014 3	0.040 3
OSU leaf	0.001 0	0.000 6	0.018 8	0.026 9
Wafer	0.000 2	0.002 4	0.031 3	0.031 4
50 words	0.001 9	0.001 1	0.018 5	0.052 3
Coffee	0.057 9	0.004 0	0.013 1	0.006 1

Table 3 MSE of reconstruction based on GASF representations and RP representations

Dataset	MSE	
	GASF	RP
ECG200	0.004 8	0.003 7
Face all	0.013 4	0.022 1
Swedish leaf	0.009 8	0.009 2
OSU leaf	0.007 7	0.012 1
Wafer	0.024 0	0.006 9
50 words	0.010 1	0.010 8
Coffee	0.033 6	0.023 4

As shown in Table 2, the GASF representation performs better than the raw sequences on most of the datasets, except for “Face all” and “Wafer”. The mean and AR methods have the largest MSE in missing values reconstruction due to the insufficiency information used in prediction.

As shown in Table 3, the RP representation performs better than the GASF representation. The 2D representa-

tions have better performance in recovering missing values than the raw time series, and the RP representation shows a slightly better performance than the GASF representation. This finding indicates the efficiency of the proposed method.

3.2 Influence of different missing data ratios

Increasing missing data ratios deteriorates reconstruction performance. To verify the influence of varying missing data ratio on the proposed method, we introduce missing values in three UCR time series datasets with missing ratios set to 20%, 30%, 40%, 50%, 60%, 70% and 80% using different representation methods. The three datasets are “ECG200”, “50 words” and “Swedish leaf”. We trained the model using the “ECG200” dataset and tested it on “50 words” and “Swedish leaf” datasets. MSE is adopted as the evaluation criterion for the comparison of reconstruction errors of different representation methods. The results are shown in Fig. 8 and Fig. 9. Similar to the previous section, we compare the three methods in two

dimensions, namely, 1D and 2D. Fig. 8 displays the reconstruction errors of the raw and GASF-restored time series under different missing ratios. The MSEs increase with the increase in the missing ratio. However, the reconstruction performance of the GASF-restored time series is relatively stable, while the reconstruction performance of the method based on raw time series deteriorates

considerably. Fig. 9 displays the reconstruction errors of the GASF and RP representations in 2D under different missing ratios. The results of two datasets show that the reconstruction performance of RP representation is more superior than GASF representation, thereby demonstrating the effectiveness of the proposed method under varying missing ratios.

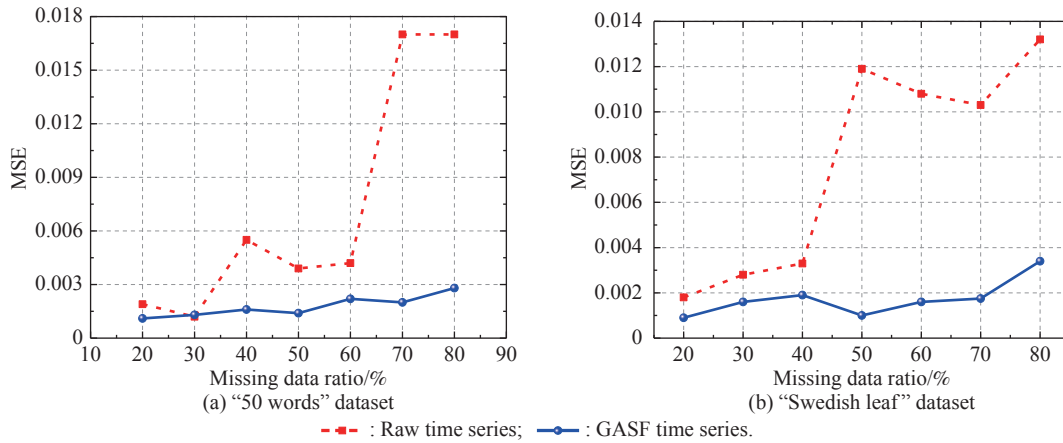


Fig. 8 Reconstruction error comparison of raw time series and GASF time series on two UCR datasets with varying missing data ratio

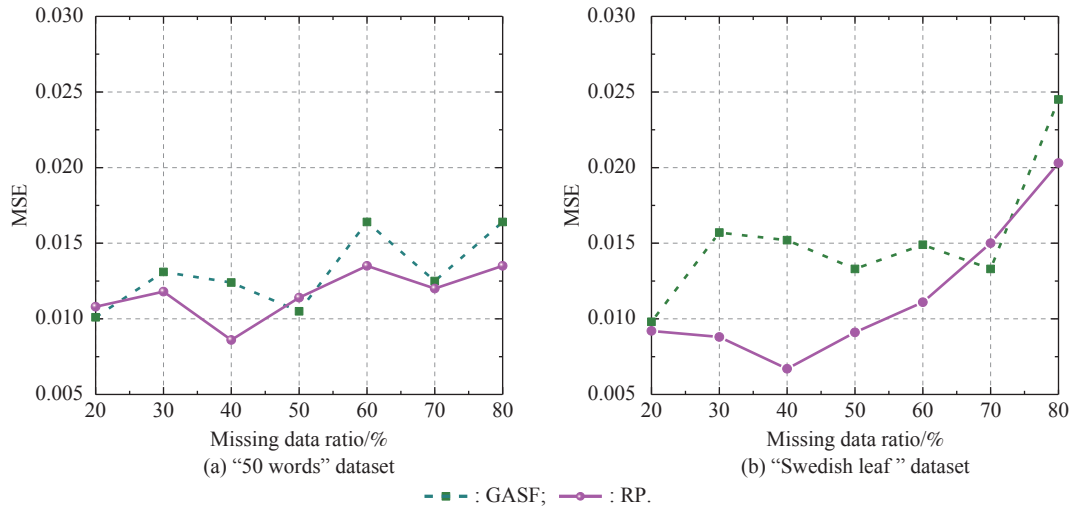


Fig. 9 Reconstruction error comparison of GASF and RP representation on two UCR datasets with varying missing data ratio

3.3 Results of time series classification

To validate the efficiency of the 2D representation, we use the CNN proposed in [14] as a classifier to test the influence of different representation methods before and after the imputation of missing values on time series classification. Table 4 shows the classification accuracy of different methods on the UCR time series. 1D-CNN denotes the CNN classifier based on the reconstructed 1D time series. 1D-CNN-0.2 denotes the CNN classifier based on the time series with a missing ratio of 20%. GASF-CNN denotes the CNN classifier based on the re-

constructed GASF. GASF-CNN-0.2 denotes the CNN classifier based on the GASF representation of the time series with a missing ratio of 20%. RP-CNN denotes the CNN classifier based on the reconstructed RP. RP-CNN-0.2 denotes the CNN classifier based on the RP representation of the time series with a missing ratio of 20%. Considering that different types of time series in the UCR archive have different lengths, we downsample the size of the GASF/RP matrix to a fixed value, which is denoted in the bracket of Table 4. The optimal results are displayed in bold.

Table 4 Classification accuracy of different methods on part of UCR time series

Dataset	1D-CNN	1D-CNN-0.2	GASF-CNN	GASF-CNN-0.2	RP-CNN	RP-CNN-0.2
Swedish leaf	0.913 4	0.330 1	0.923 0(64)	0.900 6(64)	0.942 3(32)	0.916 6(32)
Face all	0.830 7	0.801 1	0.757 3(64)	0.738 4(64)	0.757 0(64)	0.738 4(64)
ECG200	0.859 9	0.840 0	0.920 0(20)	0.910 0(20)	0.930 0(20)	0.890 0(20)
Wafer	0.994 4	0.993 8	0.994 4(64)	0.990 9(64)	0.997 4(20)	0.995 0(20)
OSU leaf	0.454 5	0.446 2	0.590 0(64)	0.545 4(64)	0.595 0(64)	0.573 0(64)
Coffee	1.000	1.000	1.000(64)	1.000(64)	1.000(64)	1.000(64)

As shown in Table 4, the classification accuracy is significantly improved after the imputation of missing values, and RP-CNN exhibits the best performance on most of the datasets. For the ‘‘Coffee’’ dataset, all methods perform well mainly because of the simple pattern of the data. The classification performance of ‘‘Swedish leaf’’ dataset remarkably deteriorates by using 1D-CNN. The performance is relatively stable using 2D-CNN. Therefore, reconstructing the missing data before classification can effectively improve the stability of the classification algorithm. The result is consistent with the abovementioned conclusion, demonstrating the superiority of the proposed method.

4. Discussion

The results in the previous section demonstrate the effectiveness of the 2D representations. The better performance of the 2D representations on the imputation of missing values compared with raw time series can be explained as follows. On the one hand, the transformations of time series into 2D matrices are equivalent to a kernel trick. It achieves data augmentation by increasing the dimensionality of the raw time series. On the other hand, 2D representations can preserve the temporal and spatial correlations. The DAE can utilise rich information for prediction by considering the temporal and spatial dependencies amongst the missing values and other time intervals. Thus, the 2D representation has more stable performance than the raw time series.

The RP and GASF representations are similar because they consider the temporal and spatial dependencies. The main difference is that the RP matrix preserves the relative temporal relations, whereas the GASF matrix represented in polar coordinates preserves the absolute temporal relations. The relative relations are robust to the translation variance of inputs. Therefore, the RP representation is more stable than the GASF representation in reconstructing the missing values.

However, we only consider the effectiveness of the method for univariate time series. The analysis of multivariate time series is more complicated, which requires a further study.

5. Conclusions

In this study, we propose a 2D representation-based DAE framework to analyze the time series with missing values. The performance is evaluated through extensive experiments. The results show that using 2D representations significantly improves the imputation and classification performance than the raw time series, and the RP representation is more robust than the GASF representation. Unlike previous missing value imputation methods, 2D representation-based DAE can utilise both temporal and spatial dependencies of the time series in predicting the missing values. In future work, we will apply the proposed method in more specific time series tasks, such as medical diagnosis and weather forecasting, to improve the reliability of decision-making systems.

References

- [1] BOX G E P, JENKINS G M. Time series analysis: forecasting and control. *Journal of Time*, 1977, 19(3): 343–344.
- [2] CHATFIELD C. The analysis of time series: an introduction. Boca Raton: CRC Press, 2016.
- [3] YOZGATLIGIL C, ASLAN S, IYIGUN C, et al. Comparison of missing value imputation methods in time series: the case of Turkish meteorological data. *Theoretical and Applied Climatology*, 2013, 112(1/2): 143–167.
- [4] PRATAMA I, PERMANASARI A E, ARDIYANTO I, et al. A review of missing values handling methods on time-series data. *Proc. of the International Conference on Information Technology Systems and Innovation*, 2016: 1–6.
- [5] LEE K J, SIMPSON J A. Introduction to multiple imputation for dealing with missing data. *Respirology*, 2014, 19(2): 162–167.
- [6] PENG L. A comparison study of missing value processing methods. *Computer Science*, 2004, 31(10): 155–156,174. (in Chinese)
- [7] LITTLE R J A, RUBIN D B. *Statistical analysis with missing data*. Hoboken: Wiley, 2019.
- [8] LOBATO F, SALES C, ARAUJO I, et al. Multi-objective genetic algorithm for missing data imputation. *Pattern Recognition Letters*, 2015, 68(15): 126–131.
- [9] BANBURA M, MODUGNO M. Maximum likelihood estimation of factor models on data sets with arbitrary pattern of missing data. *Journal of Applied Econometrics*, 2014, 29(1): 133–160.
- [10] HAYKIN S. *Neural networks: a comprehensive foundation*. Upper Saddle River: Prentice-Hall, 1999.
- [11] LECUN Y, BENGIO Y, HINTON G. Deep learning. *Nature*, 2015, 521(7553): 436–444.

- [12] BENGIO Y, COURVILLE A, VINCENT P. Representation learning: a review and new perspectives. *IEEE Trans. on Pattern Analysis & Machine Intelligence*, 2013, 35(8): 1798–1828.
- [13] FAWAZ H I, FORESTIER G, WEBER J, et al. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, 2019, 33: 917–963.
- [14] DENG Q W, LU H Z, HU M F, et al. Exo-atmospheric infrared objects classification using recurrence-plots-based convolutional neural networks. *Applied Optics*, 2019, 58(1): 164–171.
- [15] YOON J. GAIN: missing data imputation using generative adversarial nets. *Proc. of the 35th International Conference on Machine Learning*, 2018: 1–8.
- [16] GUO Z J, WAN Y M, YE H. A data imputation method for multivariate time series based on generative adversarial network. *Neurocomputing*, 2019, 360: 185–197.
- [17] ASSENDORP J P. Deep learning for anomaly detection in multivariate time series data. Hamburg: Hochschule für Angewandte Wissenschaften, 2017.
- [18] ESTEBAN C, HYLAND S L, RTSCH G. Real-valued (medical) time series generation with recurrent conditional gans. *arXiv*, 2017: arXiv170602633E.
- [19] VINCENT P, LAROCHELLE H, BENGIO Y, et al. Extracting and composing robust features with denoising autoencoders. *Proc. of the 25th International Conference on Machine Learning*, 2008: 1096–1103.
- [20] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks. *Science*, 2006, 313(5786): 504–507.
- [21] KAMPFFMEYER M, LOKSE S, BIANCHI F M, et al. Deep kernelized autoencoders. SHARMA P, BIANCHI F, ed. *Image Analysis*. Cham: Springer, 2017: 419–430.
- [22] BIANCHI F M, LIVI L, Mikalsen K, et al. Learning representations for multivariate time series with missing data using temporal kernelized autoencoders. *arXiv*, 2018: arXiv1805.03473.
- [23] SILVA D F, SOUZA V M A D, BATISTA G E A P A. Time series classification using compression distance of recurrence plots. *Proc. of the IEEE International Conference on Data Mining*, 2013: 687–696.
- [24] SOUZA V M A D, SILVA D F, BATISTA G E A P A. Extracting texture features for time series classification. *Proc. of the 22nd International Conference on Pattern Recognition*, 2014: 1425–1430.
- [25] WANG Z, OATES T. Imaging time-series to improve classification and imputation. *Proc. of the 24th International Joint Conference on Artificial Intelligence*, 2015: 3939–3945.
- [26] ECKMANN J P, KAMPHORST S O, RUELLE D. Recurrence plots of dynamical systems. *Europhysics Letters*, 1987, 4(9): 973–977.
- [27] MARWAN N, DONGES J F, ZOU Y, et al. Complex network approach for recurrence analysis of time series. *Physics Letters A*, 2009, 373(46): 4246–4254.
- [28] MARWAN N, ROMANO M C, THIEL M, et al. Recurrence plots for the analysis of complex systems. *Physics Reports*, 2007, 438(5/6): 237–329.
- [29] HATAMI N, GAVET Y, DEBAYLE J. Bag of recurrence patterns representation for time-series classification. *Pattern Analysis & Applications*, 2019, 22: 877–887.
- [30] DAU H A, BAGNALL A, KAMGAR K, et al. The UCR time series classification archive. *IEEE/CAA Journal of Automatica Sinica*, 2019, 6(6): 1293–1305.
- [31] LI J H, STRUZIK Z, ZHANG L Q, et al. Feature learning from incomplete EEG with denoising autoencoder. *Neurocomputing*, 2015, 165: 23–31.
- [32] KINGMA D P, BA J. Adam: a method for stochastic optimization. *arXiv*, 2014: arXiv1412.6980K.

Biographies



TAO Huamin was born in 1972. She received her M.S. degree from National University of Defense Technology, Changsha, China, in 1997. She is currently a researcher at the National Key Laboratory of Science and Technology on Automatic Target Recognition, National University of Defense Technology. Her research interests include signal processing and real-time system design.

E-mail: Taohmpeach@163.com



DENG Qiuqun was born in 1991. She received her Ph.D. degree in College of Electronic Science from National University of Defense Technology, Changsha, China, in 2019. She is currently a lecturer at the National Key Laboratory of Science and Technology on Automatic Target Recognition, National University of Defense Technology. Her research interests include optical guidance,

target recognition and machine learning.

E-mail: dengqiuqun75@163.com



XIAO Shanzhu was born in 1978. He received his M.S. degree from National University of Defense Technology, Changsha, China, in 2002. He is currently a researcher at the National Key Laboratory of Science and Technology on Automatic Target Recognition, National University of Defense Technology. His research interests include automatic target recognition, integrated circuit and real time system.

E-mail: mounbamboo@vip.163.com