Large-Span Bridge Strain Reconstruction Based on Bidirectional LSTM and ESN

YAN-KE TAN[1,2,3], YU-LING WANG[1,2], YI-QING NI[1,2], QI-LIN ZHANG[3] and YOU-WU WANG[1,2]

[1]Department of Civil and Environment Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China
[2]National Rail Transit Electrification and Automation Engineering Technology Research Center (Hong Kong Branch), The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China
[3]College of Civil Engineering, Tongji University, No. 1239 Siping Rd. Yangpu District, Shanghai, China

ABSTRACT

Partly missing and anomalous of the data collected from structural health monitoring (SHM) systems are inevitable due to the failure of sensor and data acquisition equipment, which lead to misjudgment of the target structure state. The data integrity demands for guaranty using reconstruction algorithms before signal processing. Recurrent neural networks (RNN) has been proved effective of reconstruction issue by learning from the historical and future signal segments. The gated RNN represented by long short-term memory (LSTM) networks and reservoir computing represented by echo state networks (ESNs) show superiority on accuracy or training efficiency than standard RNN method. In addition, bidirectional concept can be introduced into these two methods to further improve their reconstruction precision. In this paper, models built by bidirectional LSTM and ESN are used to reconstruct strain data measured by the SHM system of Tsing Kau bridge, during which performances in both time and frequency domains are compared and evaluated. Furthermore, hyperparameters including number of layers, number of hidden units, scale of reservoir, and leaky rate have been optimized to improve the structures of the proposed models.

INTRODUCTION

As the modern structural forms and functions becomes more complex, but thanks to the maturity of non-destructive testing technology, the structural health monitoring (SHM) technology that acquires real-time status information of the target structure and diagnoses its damage by installing sensors on it is booming [1, 2]. Considering the economic cost and the difficulties of data transmission and storage, the types, quantities and distributions of sensors in each SHM system are strictly designed. System operators can only perform status assessment, damage location, and accident warning based on the signals sent back by sensors if the signals are complete and accurate. However, the service duration of the SHM system is always the entire life cycle of the target structure, and during this period, the occasional sudden abnormalities of the sensors are unavoidable, which will cause the data corresponding to the signal channel to be missing or unreliable.

For continuous structures, there are always potential correlations between the response signals collected by sensors at various locations. These correlations come from the similarity of the loads on adjacent positions at the same time, and also from the fact that the vibration of the structure can often be decomposed into a linear combination of a few fixed modes. Additionally, even for missing channels, the data before and after the missing segment are good raw materials for us to infer the missing segment. This family of methods for recovering missing segments from the remaining signal data is called signal reconstruction algorithms.

Some traditional signal reconstruction algorithms focus on the statistical parameters of the signals [3, 4]. But today, the accuracies of these algorithms are completely lower than those of various machine learning-based algorithms, especially the neural network models.Owing to the existence of the additional delay loop, the recurrent neural network (RNN) has become a type of deep learning model which is aimed at processing time-series data. The amount and quality of the historical information determine the memory ability of the RNN model. To improve the memory capacity of RNN models, several “gates” are added to control and select the information that flow through each neuron.

Among these gated RNNs, the long-short term memory (LSTM) [5, 6] network has the best performance and the widest range of applications. But the sophisticated gating system also makes LSTM more difficult to train than standard RNN, which inherently has difficulty of training because of its recursive structure. In contrast, some researchers have also proposed a more lightweight RNN model, called the echo state network (ESN) [7, 8], which has a simple structure and is easy to train, and is therefore very suitable for online tasks. ESN replaces the hidden layer of RNN with a reservoir with predefined parameters, and finally simplifies the training process of RNN into a linear regression problem.

In the previous research of the authors, an LSTM model introducing Bayesian optimization has been proposed to generate railway track slab vibrations, while a modified ESN model with bidirectional conception has put forward used to reconstruct the acceleration data of super high-rise buildings. However, as typical representatives of reconstruction algorithms, the performance of ESN and LSTM models and the choice of algorithms for specific problems are not conclusive. In this paper, various performances of ESN and LSTM models on reconstruction problems will be compared, including training efficiency, inference speed, accuracies in both time and frequency domains, etc. At the same time, the improvement brought by the bidirectional concept to the two types of algorithms will also be quantitatively studied. This study uses the field measuring data of the Tsing Kau Bridge (TKB) in Hong Kong as a verification.

METHODOLOGY

In this section, both LSTM and ESN are introduced. As a kind of RNNs respectively, both of them share some basic structure and notations. It is assumed that the input signals being $\boldsymbol{X} \in \mathbb{R}^{T\times N_i}$, where T is the number of time steps and N_i is the number of input channels. So that, the value of signal in channel i at time step t can be denoted as $X_i(t)$. The state signals stored in the hidden layer is $\boldsymbol{R} \in \mathbb{R}^{(T+1)\times N_r}$, where N_r is the number of neurons in the hidden layer and $\boldsymbol{R}(0)$ is the random set initial value of the state signals. The output of the network which is called $\boldsymbol{Y} \in \mathbb{R}^{T\times N_o}$ is generate by a certain type of connection between $\boldsymbol{R}$. Here, N_o is the number of output channels. In the training process of these two models, the lost function can be expressed as,

$$L(\boldsymbol{Y_i}, \widehat{\boldsymbol{Y}}_{\boldsymbol{i}}) = \frac{1}{T}\sum_{t=1}^{T}\left(Y_i(t) - \hat{Y}_i(t)\right)^2 + \eta\|\boldsymbol{W_i}\|_2, \tag{1}$$

where $\widehat{\boldsymbol{Y}}_{\boldsymbol{i}}$ is the reconstructed signal of the channel i, η is the regularization parameter, $\boldsymbol{W}$ is the weight matrixes illustrated below, and $\|\cdot\|_2$ stands for the second normal form.

LSTM and Bi-LSTM

For the standard RNNs, when calculating $\boldsymbol{R}(t)$, both $\boldsymbol{R}(t-1)$ which is the representative of historical information and $\boldsymbol{X}(t)$ which provides present excitation are completely considered. However, in LSTM, a changing vector of equal length (called internal state) is added into the transfer process along time dimension to control the memory of the network, that is, $\boldsymbol{C} \in \mathbb{R}^{(T+1)\times N_r}$. Thus, the $\boldsymbol{R}(t)$ can be calculated by,

$$\boldsymbol{R}(t) = \boldsymbol{O}(t)\odot\tanh(\boldsymbol{C}(t)), \tag{2}$$

where $\boldsymbol{O} \in \mathbb{R}^{T\times N_r}$ is the output gate and $\odot$ is the Hadamard product operation. The mentioned $\boldsymbol{R^o}$ is obtained as,

$$\boldsymbol{O}(t) = \sigma\left(\boldsymbol{W}_{\boldsymbol{L}}^{\boldsymbol{in,o}}\boldsymbol{X}(t) + \boldsymbol{W}_{\boldsymbol{L}}^{\boldsymbol{r,o}}\boldsymbol{R}(t-1) + \boldsymbol{b}_{\boldsymbol{L}}^{\boldsymbol{o}}\right), \quad (3)$$

where $\sigma(\cdot)$ is the sigmoid function. On the other hand, $\boldsymbol{C}(t)$ can be seen as the combination of the partially forgotten historical information and the selected current input information, as,

$$\boldsymbol{C}(t) = \boldsymbol{F}(t) \odot \mathbf{C}(\mathrm{t}-1) + \mathbf{I}(\mathrm{t}) \odot \tanh\left(\boldsymbol{W}_{\boldsymbol{L}}^{\boldsymbol{in,c}}\boldsymbol{X}(t) + \boldsymbol{W}_{\boldsymbol{L}}^{\boldsymbol{r,c}}\boldsymbol{R}(t-1) + \boldsymbol{b}_{\boldsymbol{L}}^{\boldsymbol{c}}\right), \quad (4)$$

where $\boldsymbol{F}, \boldsymbol{I} \in \mathbb{R}^{T \times N_r}$ are forgetting gate and input gate. They are determined by,

$$\boldsymbol{F}(t) = \sigma(\boldsymbol{W}_{\boldsymbol{L}}^{\boldsymbol{in,f}}\boldsymbol{X}(t) + \boldsymbol{W}_{\boldsymbol{L}}^{\boldsymbol{r,f}}\boldsymbol{R}(t-1) + \boldsymbol{b}_{\boldsymbol{L}}^{\boldsymbol{f}}), \quad (5)$$

$$\boldsymbol{I}(t) = \sigma(\boldsymbol{W}_{\boldsymbol{L}}^{\boldsymbol{in,i}}\boldsymbol{X}(t) + \boldsymbol{W}_{\boldsymbol{L}}^{\boldsymbol{r,i}}\boldsymbol{R}(t-1) + \boldsymbol{b}_{\boldsymbol{L}}^{\boldsymbol{i}}) \quad (6)$$

And finally, the output can be calculated as given by,

$$\boldsymbol{Y}(t) = \sigma(\boldsymbol{W}_{\boldsymbol{L}}^{\boldsymbol{out}}\boldsymbol{R}(t) + \boldsymbol{b}_{\boldsymbol{L}}) \quad (7)$$

Weight matrixes $\boldsymbol{W}_{\boldsymbol{L}}^{\boldsymbol{in,o}}, \boldsymbol{W}_{\boldsymbol{L}}^{\boldsymbol{in,c}}, \boldsymbol{W}_{\boldsymbol{L}}^{\boldsymbol{in,f}}, \boldsymbol{W}_{\boldsymbol{L}}^{\boldsymbol{in,i}} \in \mathbb{R}^{N_i \times N_r}$, $\boldsymbol{W}_{\boldsymbol{L}}^{\boldsymbol{r,o}}, \boldsymbol{W}_{\boldsymbol{L}}^{\boldsymbol{r,c}}, \boldsymbol{W}_{\boldsymbol{L}}^{\boldsymbol{r,f}}, \boldsymbol{W}_{\boldsymbol{L}}^{\boldsymbol{r,i}} \in \mathbb{R}^{N_r \times N_r}$, $\boldsymbol{W}_{\boldsymbol{L}}^{\boldsymbol{out}} \in \mathbb{R}^{N_r \times N_o}$, $\boldsymbol{b}_{\boldsymbol{L}}^{\boldsymbol{o}}, \boldsymbol{b}_{\boldsymbol{L}}^{\boldsymbol{c}}, \boldsymbol{b}_{\boldsymbol{L}}^{\boldsymbol{f}}, \boldsymbol{b}_{\boldsymbol{L}}^{\boldsymbol{i}} \in \mathbb{R}^{N_r}$, and $\boldsymbol{b}_{\boldsymbol{L}} \in \mathbb{R}^{N_o}$ are the parameters to be trained.

It is clear that the three gates control the updating process of the state and internal state signals by selecting the input and forgetting part of the history. The structure of the LSTM is also shown in Figure 1(a).

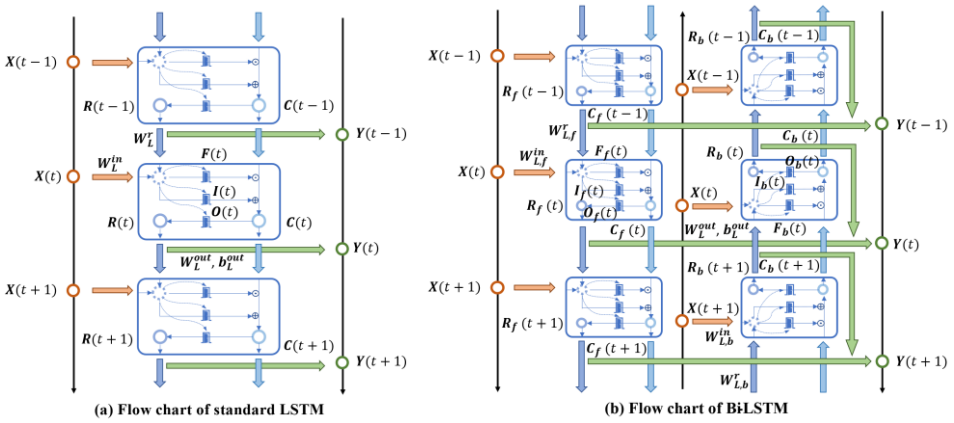


Figure 1. The structures of LSTM and Bi-LSTM

Some researchers claim that the future information can be valuable for the network to precisely determine the state signals. By reversing the input signals, another information flow will be generated to form the bidirectional LSTM (Bi-LSTM). In this improved network (see Figure 1(b)), the Equation (6) will be changed into,

$$\boldsymbol{Y}(t) = \sigma\left(\boldsymbol{W}_{\boldsymbol{L,f}}^{\boldsymbol{out}}\boldsymbol{R}_{\boldsymbol{f}}(t) + \boldsymbol{W}_{\boldsymbol{L,b}}^{\boldsymbol{out}}\boldsymbol{R}_{\boldsymbol{b}}(t) + \boldsymbol{b}_{\boldsymbol{L}}\right) \quad (8)$$

ESN and Bi-ESN

In ESN, a large, nonlinear, and sparsely connected reservoir is used to replace the hidden layer. The input signals will be transformed into a higher dimension space by the reservoir. Due to the sparsity of the reservoir, the shapes of the state signals are always simplex, that they will be employed as a group of basis signals to generate complex signals. If the scale of the reservoir is large enough, it will be possible to fit the output signals $\boldsymbol{Y}(t)$ by linear combination of the $\boldsymbol{R}(t)$, while $\boldsymbol{R}(t)$ generating weight matrixes can be fixed without training, that is,

$$\boldsymbol{R}(t) = (1-\alpha)\boldsymbol{R}(t-1) + \alpha\tanh\left(\boldsymbol{W}_{\boldsymbol{E}}^{\boldsymbol{in}}\boldsymbol{X}(t) + \beta_w\boldsymbol{W}_{\boldsymbol{E}}^{\boldsymbol{r}}\boldsymbol{R}(t-1) + \beta_b\boldsymbol{b}_{\boldsymbol{E}}^{\boldsymbol{r}}\right), \quad (9)$$

where α is the leaky rate which can be used to improve the memory capacity of the ESN models, β_w and β_b are the hyperparameters being responsible for controlling the ratio of input and history, and $\boldsymbol{W_E^{in}}, \boldsymbol{W_E^r} \sim N[0,1]$ are the preset mapping matrixes.

To further improve the memory ability of ESN, a series of delay signals of input, and state signals are added to the output layer, that is,

$$\boldsymbol{Y}(t) = \sigma(\boldsymbol{W_E^{out}}[\boldsymbol{X}(t); \dots; \boldsymbol{X}(t-t_i); \boldsymbol{R}(t); \dots; \boldsymbol{R}(t-t_r)] + \boldsymbol{b_E}), \quad (10)$$

where $\boldsymbol{W_E^{out}} \in \mathbb{R}^{(t_i N_i + t_r N_r) \times N_o}$ and $\boldsymbol{b_E} \in \mathbb{R}^{N_o}$ are the only parameters to be trained, t_i and t_r are the delay time steps of the input and state signals, and $[\cdot;\cdot]$ stands for the vector concatenation operation. The training process will be only conducted in the output layer, and it is a simple linear regression problem which can be solved by standard gradient decent algorithm.

Similarly, the bidirectional concept can be also introduced into ESN models, as given by Equation (10). The structures of ESN and Bi-ESN are shown in Figure 2.

$$\boldsymbol{Y}(t) = \sigma\begin{pmatrix} \boldsymbol{W_{E,f}^{out}}[\boldsymbol{X}(t); \dots; \boldsymbol{X}(t-t_i); \boldsymbol{R_f}(t); \dots; \boldsymbol{R_f}(t-t_r)] + \\ \boldsymbol{W_{E,b}^{out}}[\boldsymbol{R_b}(t); \dots; \boldsymbol{R_b}(t-t_r)] + \boldsymbol{b_E} \end{pmatrix} \quad (11)$$

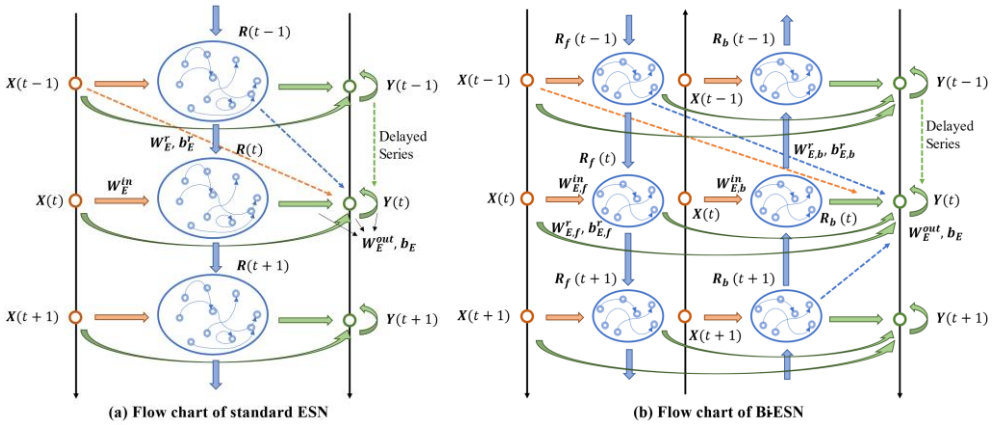


Figure 2. The structures of ESN and Bi-ESN

DATASET SELECTION AND PREPROCESS

A dataset collected from the SHM system of TKB in Hong Kong is sued to verify the proposed methods and compare their performance. The SHM system consists of 42 strain gauges and 107 other types of sensors. Ten channels contain strain data are selected to build the dataset. The locations of these measuring points are shown in Figure 3. The ten sensors locate in two cross beams of the bridge, one is near the main tower, and another is at the midspan of the third span. Each cross beam has five measuring points, corresponding to three sections which are beneath the lanes. In one cross beam, the left and right sections possess two strain gauges on the top and bottom flanges respectively, while only one strain gauge is installed on the middle section. The sampling rate of the strain gauges is 51.2 Hz, and the length of the signal is therefore determined as 10240 (200 s). The dataset can be denoted as $\boldsymbol{S} \in \mathbb{R}^{10240 \times 10}$.

The dataset is normalized for convenience of training. The normalization process is expressed in Equation (11), and the ten normalized signals have the same mean value (0) and standard deviation (1).

$$\boldsymbol{S_{i,norm}} = (\boldsymbol{S_i} - E[\boldsymbol{S_i}])/\sqrt{V[\boldsymbol{S_i}]}, \quad (12)$$

where $E[\cdot]$ stands for determine the mean value and $V[\cdot]$ stands for determine the variance.

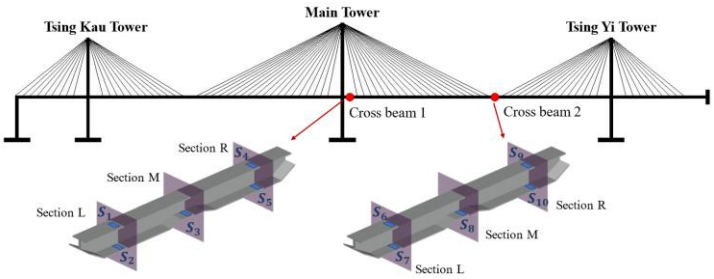


Figure 3. The strain monitoring system of TKB

Table 1. The condition setting for the reconstruction task

Condition group	Condition label	Missing channel	Known channel
Single channel missing	$R_{sig,1}$~$R_{sig,10}$	$\boldsymbol{S_i}$	$\boldsymbol{S_j}, (j \neq i)$
Double channel missing	$R_{sec,1}$	$\boldsymbol{S_1, S_2}$	$\boldsymbol{S_3}$~$\boldsymbol{S_{10}}$
	$R_{sec,2}$	$\boldsymbol{S_4, S_5}$	$\boldsymbol{S_1}$~$\boldsymbol{S_3, S_6}$~$\boldsymbol{S_{10}}$
	$R_{sec,3}$	$\boldsymbol{S_6, S_7}$	$\boldsymbol{S_1}$~$\boldsymbol{S_5, S_8}$~$\boldsymbol{S_{10}}$
	$R_{sec,4}$	$\boldsymbol{S_9, S_{10}}$	$\boldsymbol{S_1}$~$\boldsymbol{S_8}$
Multi-channel missing	$R_{beam,1}$	$\boldsymbol{S_1}$~$\boldsymbol{S_5}$	$\boldsymbol{S_6}$~$\boldsymbol{S_{10}}$
	$R_{beam,2}$	$\boldsymbol{S_6}$~$\boldsymbol{S_{10}}$	$\boldsymbol{S_1}$~$\boldsymbol{S_5}$

In the reconstruction tasks, part of the channels in $\boldsymbol{S}$ is assumed lost, three groups of conditions are taken into consideration in this paper. The condition settings are summarized in Table 1.

RESULTS AND DISCUSSIONS

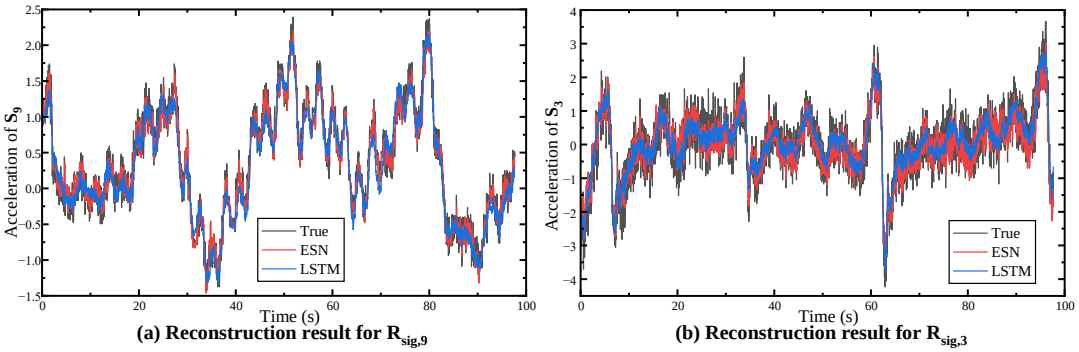
In this section, the results of two-dimensional comparison analysis are shown, one is the comparison between the two types of models, and another is between models inducing the bidirectional structure or not. The results are obtained by feeding the dataset described in Section 3 into the models (LSTMs and ESNs) introduced in Section 2. The ratio of the length of training set to that of testing set is 0.5, and in order to fairly compare their performance, both LSTM and ESN models experience a 200-epoch training with the same learning rate of 0.0002 and the same η of 0.001. The reconstruction accuracy will be evaluated by the mean square error (MSE).

Comparison analysis of standard LSTM and ESN

In the group of single channel missing, each channel is reconstructed by the model with signals of other nine channels as input. Figure 4(a) and 4(b) shows the reconstructing results of the conditions with the highest and the lowest accuracy. The reconstruction MSE for other channels are listed in Table 2. The reconstructing accuracy order of these ten conditions are the same regardless of the LSTM or ESN model being used. The reason for this phenomenon might be that the signal similarity between $\boldsymbol{S_9}$ and other channels is high, especially being similar to $\boldsymbol{S_{10}}$ (on the same section), while the shape of $\boldsymbol{S_3}$ is quite different from others, because there are no adjacent sensors

nearby. However, although the apparent similarity between the channels is divergent, the intrinsic correlations between the responses of the strain gauges on the same bridge under the same external excitation always exists objectively. This will lead to possibility of reconstructing part of them by the others if the used reconstructing model can learn the potential relationships. From the results, both the ESN and LSTM models have completed this task, and the reconstruction MSE of each condition fluctuates between 0.0201~0.1508 and 0.0252~0.1342 respectively, in other words, for all channels, both LSTM and ESN models have qualified reconstruction capabilities.

In details, it is clear that, in most of conditions within this group, ESN and LSTM show almost the same learning performance, but in majority, the MSEs of ESN are slightly lower. In these conditions where the ESN performance is better, the signal to be reconstructed has obvious fluctuation characteristics, which is specifically manifested in the signal with fewer high-frequency components and more low-frequency components. Taking the TKB's natural frequency into consideration, these channels can be generalized as signals with a large signal-to-noise ratio (SNR). In contrast, when reconstructing the signals with lower SNR, the simple "fitting" structure of ESN is not acting as well as before. This is perhaps caused by the interference of the noise.



(a) Reconstruction result for $R_{sig,9}$ **(b) Reconstruction result for $R_{sig,3}$**

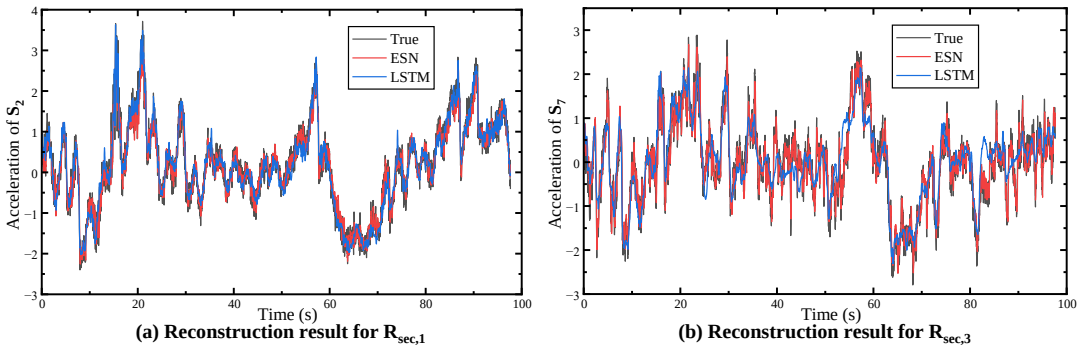
Figure 4. The reconstruction results of the first group of conditions via LSTM and ESN

Table 2. The reconstruction MSE for the conditions in the group of single channel missing

Condition label	Reconstruction MSE		Condition label	Reconstruction MSE	
	ESN	LSTM		ESN	LSTM
$R_{sig,1}$	0.0405	0.0701	$R_{sig,6}$	0.0834	0.0589
$R_{sig,2}$	0.0410	0.0699	$R_{sig,7}$	0.0412	0.0561
$R_{sig,3}$	0.1508	0.1342	$R_{sig,8}$	0.1220	0.1081
$R_{sig,4}$	0.0424	0.0498	$R_{sig,9}$	0.0201	0.0252
$R_{sig,5}$	0.0396	0.0365	$R_{sig,10}$	0.0347	0.0456

Similarly, we also selected the channels with the best and worst reconstruction effects in the second group of working conditions and showed them in Figure 5. For each reconstruction segment in Figure 5, there is no supplementary information from the same section, resulting in that the reconstruction MSE for each channel in this group of conditions is higher than that in the first group. In addition, from the results of this group, the performance of ESN began to surpass LSTM in an all-round way. The average reconstruction MSE for ESN is around 0.04, while that for LSTM reaches about 0.09. The author speculates that the reason for this result might be that no matter how many input channels there are, the "raw material" involved in constructing the output signals in the ESN network is always the basis consisting of all simple basic signals in

the high-dimensional space. Therefore, this very large reservoir makes the impact from the reduction of input channels on ESN reduced to a very low level. This inference is further confirmed by the results of the third group of working conditions. The reconstruction MSEs for the channels on Section L and Section R are not much lower than those of the second group. There is also just a negligible rise of the error happened in reconstruction channel in Section M. All the facts mentioned above confirmed that the ESN model has better performance in the reconstruction problem facing the synchronized missing of several channels.



(a) Reconstruction result for $R_{sec,1}$ **(b) Reconstruction result for $R_{sec,3}$**

Figure 5. The reconstruction results of the second group of conditions via LSTM and ESN

All work described in this paper is performed on an NVIDIA GeForce RTX3060. The training of the single-channel ESN reconstruction model on this device takes about 240 s (200 epochs). Since the ESN model is trained for an optimal $\boldsymbol{W_E^{out}}$, the workload is proportional to the number of output channels, that is, the time cost for double-channel model is 470 s, and that for five-channel model is about 15 min. But the above results are all based on $N_{res} = 4000$, which is manually selected for this specific task. According to its principle, the complexity of the training process of ESN model is $O({N_{res}}^3)$. For the LSTM model. The training of the output layer is only a part of the training process, and the number of neurons in the hidden layer of the LSTM does not have to be as large as that of the ESN model. Therefore, the training time of the LSTM model hardly changes with the number of output channels, and has always fluctuated from 15 to 17 min.

Improvement of introducing bidirectional concept

The bidirectional concept is a method which can improve the performance of the model processing time series by introducing future information into the state signals.

After the model is modified according to the bidirectional structure, the reconstruction performance of both ESN and LSTM is improved. For the first group of conditions, the highest MSE is reduced from 0.1342 to 0.1091 for LSTM and 0.1508 to 0.1207 for ESN, while the lowest MSE is reduced by about 0.005 for both LSTM and ESN models. In addition, for both models, adopting a bidirectional structure will also double the computational cost (both training process and testing process). In order to further study the performance of the two reconstruction algorithms, we also performed Fourier transform on the response signals.

Obviously, when the time domain reconstruction results are basically accurate, the results of both the ESN and LSTM models can also achieve the same frequency

spectrum as the original signal in the frequency domain. After the introduction of the bidirectional concept, the reconstruction accuracy of ESN and LSTM will be further improved in both time and frequency domains. The frequency domain characteristics of vibration signals are often important indicators for structural state assessment and damage detection. The superiority of the proposed algorithms in frequency domain reconstruction makes the signal reconstructed by the algorithm not only complete the dataset, but also It can also provide significant assistance to the realization of the ultimate goal of SHM.

CONCLUSIONS

In this paper, ESN and LSTM are adopted as representatives of RNNs to reconstruct missing signal segments of SHM systems. According to the results based on a dataset collected from Hong Kong TKB SHM system, the conclusions can be drawn as below.

(1) The ESN model is sensitive enough to the change mode of the vibration signals, and its reconstruction accuracy will increase with the similarity of the input and output channels. But it will also be confused by noise, at this time, LSTM can better find the true value of the missing signals.

(2) As the number of missing channels increases, the reconstruction accuracy of LSTM will be greatly reduced because it is difficult to learn the internal relationship between signals, but the sufficient basic signals stored in the reservoir in ESN can basically help it maintaining the original accuracy.

(3) The LSTM and ESN models have acceptable frequency domain reconstruction performance, which will be further improved with the introduction of the bidirectional structure. High level of frequency reconstruction accuracy will provide a basis for the signals reconstructed by the models to be used in complex problems such as structural damage detection and state assessment.

REFERENCES

1. J.M. Ko, Y.Q. Ni. 2005. “Technology developments in structural health monitoring of large-scale bridges,” *Eng. Struct.,* 27: 1715-1725.
2. Q.L. Zhang, B. Yang, T. Liu, H. Li, J. Lv. 2015. “Structural health monitoring of Shanghai Tower considering time-dependent effects,” *International Journal of High-Rise Buildings*, 4 (1): 39-44.
3. C.H. Liu, S. Hoi, P.L. Zhao, J.L. Sun. 2016. “Online ARIMA algorithms for time series prediction,” in *Proc. AAAI Conf. on Artificial Intelligence,* pp. 1867-1873.
4. M. Hoshiya, E. Saito. 1984. “Structural identification by extended Kalman filter,” *Jour. Of Eng. Mech.,* 110 (12): 1757-1770.
5. S. Hochreiter, J, Schmidhuber. 1997. “Long short-term memory,” *Neural Comput.,* 9 (8): 1735-1780.
6. A. Sherstinsky. 2020. “Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network,” *Physica D: Nonlinear Phenomena*, 404: 132306.
7. H. Jager, H. Haas. 2004. “Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication,” *Science,* 304 (5667): 78-80.
8. A. Rodan, P. Tino. 2010. “Minimum complexity echo state network,” *IEEE transactions on neural networks,* 22 (1): 131-144.