

Deep learning-based recovery method for missing structural temperature data using LSTM network

Hao Liu^{1,3a}, You-Liang Ding^{*1,2}, Han-Wei Zhao^{1,2b},
Man-Ya Wang^{1,2c} and Fang-Fang Geng^{4d}

¹School of Civil Engineering, Southeast University, Nanjing 210096, China

²Key Laboratory of C&PC Structures of the Ministry of Education,
Southeast University, Nanjing 210096, China

³Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology,
Clear Water Bay, Kowloon, HKSAR

⁴School of Architecture Engineering, Nanjing Institute of Technology, Nanjing 211167, China

(Received August 13, 2019, Revised November 10, 2019, Accepted November 15, 2019)

Abstract. Benefiting from the massive monitoring data collected by the Structural health monitoring (SHM) system, scholars can grasp the complex environmental effects and structural state during structure operation. However, the monitoring data is often missing due to sensor faults and other reasons. It is necessary to study the recovery method of missing monitoring data. Taking the structural temperature monitoring data of Nanjing Dashengguan Yangtze River Bridge as an example, the long short-term memory (LSTM) network-based recovery method for missing structural temperature data is proposed in this paper. Firstly, the prediction results of temperature data using LSTM network, support vector machine (SVM), and wavelet neural network (WNN) are compared to verify the accuracy advantage of LSTM network in predicting time series data (such as structural temperature). Secondly, the application of LSTM network in the recovery of missing structural temperature data is discussed in detail. The results show that: the LSTM network can effectively recover the missing structural temperature data; incorporating more intact sensor data as input will further improve the recovery effect of missing data; selecting the sensor data which has a higher correlation coefficient with the data we want to recover as the input can achieve higher accuracy.

Keywords: structural health monitoring (SHM); structural temperature; deep learning; LSTM network; missing data recovery

1. Introduction

Nowadays, many key infrastructures are deployed with structural health monitoring (SHM) systems, the real-time monitoring data provides opportunities for the assessment of structural state

*Corresponding author, Professor, E-mail: civilchina@hotmail.com

^a Graduate Student, E-mail: h.liu@connect.ust.hk

^b Ph.D. Candidate, E-mail: wudzihw_0@126.com

^c Ph.D. Student, E-mail: 836383567@qq.com

^d Associate Professor, E-mail: gengfangfang_1983@163.com

(Qu *et al.* 2019, Pei *et al.* 2019, Huang *et al.* 2020). However, in the field monitoring, the data collected by the monitoring system is often missing due to sensor faults, power failure and many other reasons, which poses a barrier to the analysis of monitoring data and the mastery of the structural state. Therefore, some scholars began to study the recovery method for missing SHM data. Chen *et al.* (2019) proposed an LQD-RKHS-based distribution-to-distribution regression methodology to restore the probability distribution of missing data from the perspective of probability distribution. Wan and Ni (2018) reconstructed SHM data using Bayesian multi-task learning methodology with multi-dimensional Gaussian process prior. Ye *et al.* (2017) suggested that based on wavelet multi-resolution analysis and support vector machine method, the missing data of the health monitoring system of a prestressed concrete cable-stayed bridge could be imputed effectively. Ni and Li (2016) used neural network technology to reconstruct the missing wind pressure monitoring data of a high-rise building. In addition, the compressed sensing algorithm and the improved algorithms based on it were widely used in the recovery of missing data in SHM systems (Thadikemalla and Gandhi 2018, Huang *et al.* 2016, Bao *et al.* 2018).

Meanwhile, with the development of artificial intelligence, deep learning has made breakthroughs in speech recognition, image recognition, and other fields (Schmidhuber 2015). As a data-driven method, deep learning can self-evolve and self-correct according to big data, and continuously improve the accuracy of the model without human intervention. In recent years, deep learning method has been introduced into the SHM technology by some scholars and has achieved good results in the identification of structural diseases such as cracks (Cha *et al.* 2017, Zhang *et al.* 2017) and corruptions (Atha and Jahanshahi 2018), detection of structural monitoring data anomaly (Bao *et al.* 2019), detection of sensor faults (Li *et al.* 2018), and assessment of structural health condition (Rafiei and Adeli 2018). However, there are few studies that use deep learning method to recover missing health monitoring data directly. Therefore, based on the massive characteristic and time series correlation of SHM data, this paper aims to recover the missing monitoring data in the actual environment by using deep learning method and discuss its practical effect. Among many deep learning algorithms, LSTM algorithm is very good at processing time series data due to its unique design of network units (Zhao *et al.* 2019). According to this feature of LSTM network, this paper mainly discusses the application of LSTM network in the missing data recovery, and the data recovery object is selected from the structural temperature data of Nanjing Dashengguan Yangtze River Bridge.

The rest of this paper is organized as follows. The second section briefly introduces the Nanjing Dashengguan Yangtze River Bridge and its structure temperature monitoring system. In the third section, the support vector machine (SVM), wavelet neural network (WNN), and short-short time memory (LSTM) network are used to predict the structural temperature data, and the comparison is made to prove the superior performance of LSTM network in processing time series data such as structural temperature. In the fourth section, the application of LSTM network in the recovery of missing structural temperature data is discussed in detail.

2. Bridge description and temperature monitoring system

2.1 Nanjing Dashengguan Yangtze River Bridge

Located in Nanjing City, Jiangsu Province, China, the Nanjing Dashengguan Yangtze River Bridge is one of the control projects of the Beijing-Shanghai high-speed railway. The main bridge

is a six-span continuous steel truss arch bridge, and the main span (2×336 m) is the maximum span of the same type of bridges in the world. The elevation view of the bridge is shown in Fig. 1.

2.2 Structural temperature monitoring of the bridge

There are 116 monitoring points in Nanjing Dashengguan Yangtze River Bridge, the main monitoring contents include: (1) wind speed, temperature, and humidity related to the bridge site environment; (2) structure temperature, bridge deformation, working condition of bearings, structure vibration, driving condition, structure stress, fatigue condition, etc. related to bridge maintenance requirements. And the bridge health monitoring system is equipped with 12 temperature sensors at section 1-1 of the main girder shown in Fig. 1. W_i represents the i^{th} temperature sensor. $W_1 - W_8$ are arranged on both sides of the side truss arch in order of height, while $W_9 - W_{12}$ are arranged transversely on the steel bridge deck and the transverse stiffening beam respectively. The detailed positions of the temperature sensors are shown in Figs. 2(a)-2(f) respectively. The sampling frequency of sensors is 1 Hz.

3. Long short-term memory network and temperature data prediction

3.1 Recurrent neural network and long short-term memory network

The LSTM network is a variant of the recurrent neural network (RNN). Therefore, before introducing the LSTM network, it is necessary to briefly introduce RNN. As shown in Fig. 3, different from the traditional artificial neural network (ANN), each node in the hidden layers of RNN connects the information input at the current time point and the information transmitted from the last time point simultaneously, so that the network has the function of "memory" and is suitable for processing time series data with context.

From Fig. 3 we can see that, theoretically, RNN can process the information in time series data of any length. But actually, it is impossible. The underlying cause is that after many stages of propagation, the gradients in the network tend to disappear or explode, making it difficult to capture the long-term correlation of sequence data (Bengio *et al.* 1994). Aiming at this long-term dependence problem in RNN, Hochreiter and Schmidhuber (1997) proposed the LSTM network. The difference between the LSTM network and RNN is the composition of the network unit, as shown in Fig. 4.

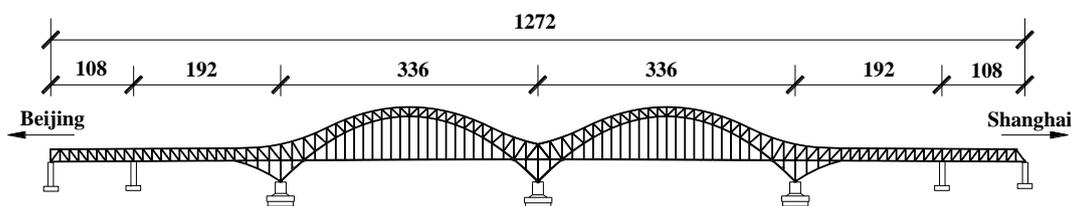


Fig. 1 Elevation view of the bridge (unit: m) (Ding *et al.* 2017)

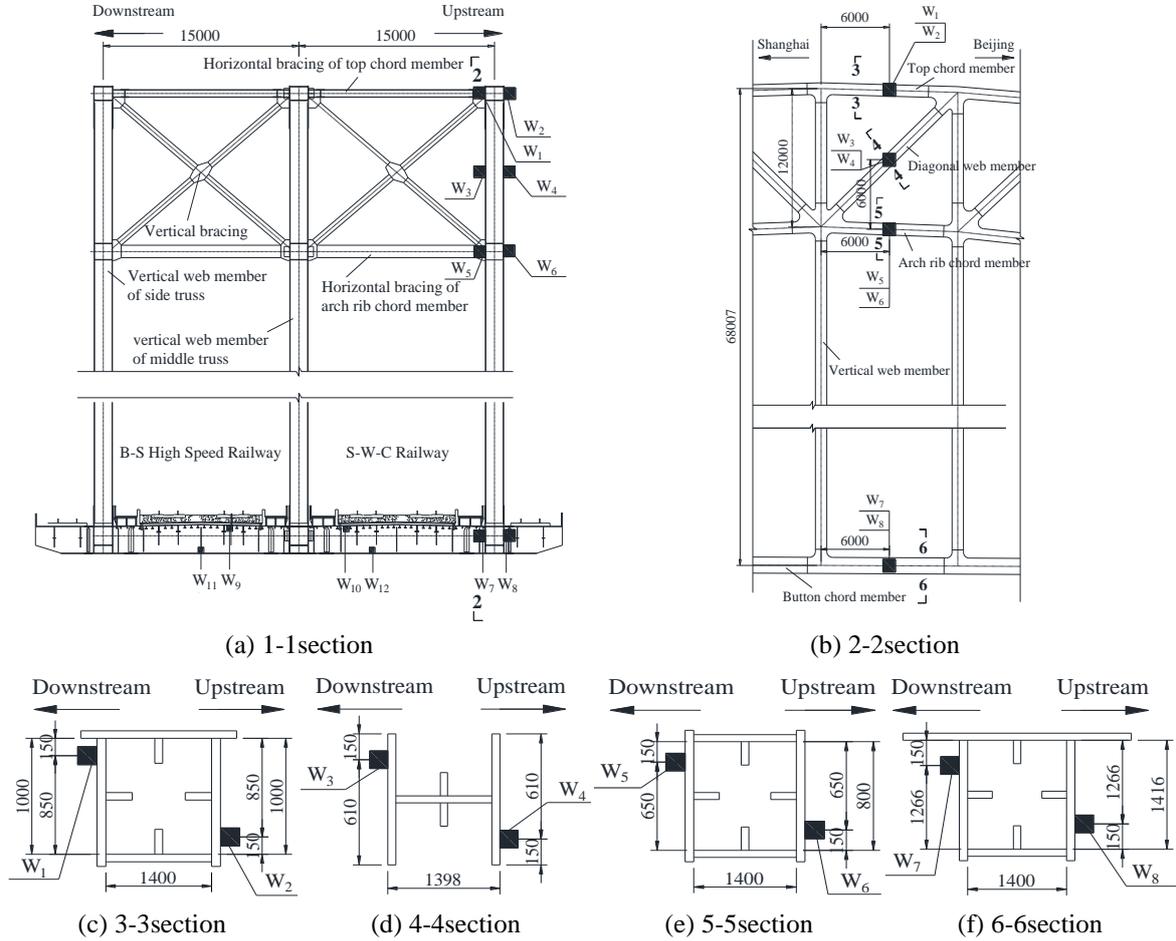


Fig. 2 Locations of temperature sensors of Nanjing Dashengguan Yangtze River Bridge (unit: mm)

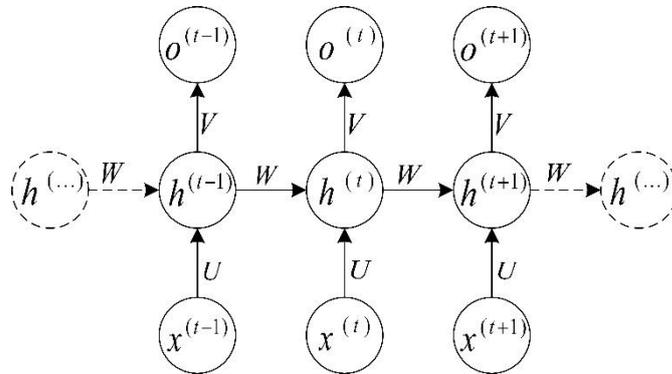


Fig. 3 Structure of RNN

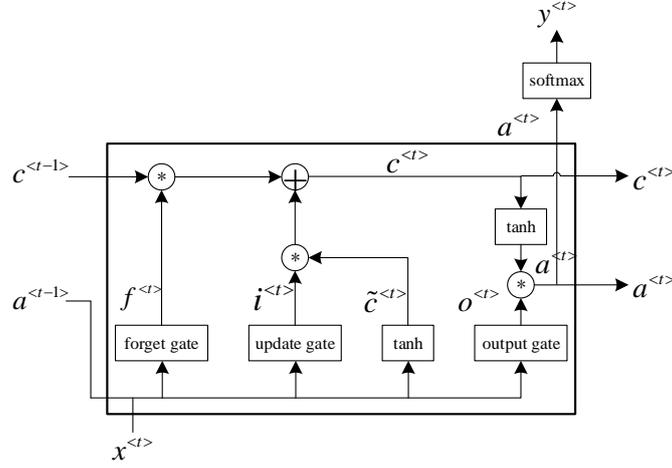


Fig. 4 Structure of LSTM network unit

LSTM network introduces cell state and three gating units ("forget gate", "update gate", and "output gate") to control the process of information processing. The cell state combines information about the current input $x^{(t)}$ with information about the cell state at the previous moment $c^{(t-1)}$. A gate is a metaphor for a structure that combines a Sigmoid neural network layer with point multiplication. The value range of the Sigmoid function is between 0 and 1. When the output value is 0, any vector multiplied by it will be 0, which is equivalent to a closed door. When the output value is 1, any vector multiplied by it will not change, which is equivalent to an open door. When the output value is between 0 and 1, the door is half-closed and the retention of "memory" can be regulated. The "forget gate" controls the integration of information about the last cell state $c^{(t-1)}$ into the current cell state $c^{(t)}$. The "update gate" controls the integration of the current input $x^{(t)}$ into the current cell state $c^{(t)}$. The "output gate" generates the hidden layer unit state $a^{(t)}$ from the current cell state $c^{(t)}$. For each input, the LSTM network unit performs the following calculations

$$\tilde{c}^{<t>} = \tanh(W_c[a^{<t-1>}, x^{<t>}] + b_c) \quad (1)$$

$$\Gamma_u = \sigma(W_u[a^{<t-1>}, x^{<t>}] + b_u) \quad (2)$$

$$\Gamma_f = \sigma(W_f[a^{<t-1>}, x^{<t>}] + b_f) \quad (3)$$

$$\Gamma_o = \sigma(W_o[a^{<t-1>}, x^{<t>}] + b_o) \quad (4)$$

$$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + \Gamma_f * c^{<t-1>} \quad (5)$$

$$a^{<t>} = \Gamma_o * \tanh c^{<t>} \quad (6)$$

Where Γ_u , Γ_f , Γ_o represent "update gate", "forget gate", and "output gate" respectively; W_u , W_f , W_o and b_u , b_f , b_o represent the weight and bias between "update gate", "forgetting gate", and "output gate", respectively; $x^{<t>}$, $a^{<t>}$, $c^{<t>}$ represent the current input, hidden layer state, and cell state, respectively; W_c and b_c represent the weight and bias between cells, respectively; σ is the Sigmoid function.

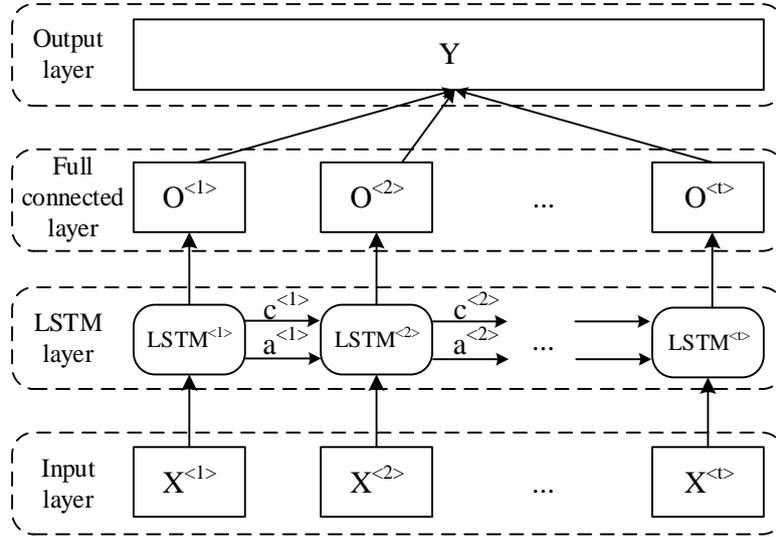


Fig. 5 Construction of LSTM network

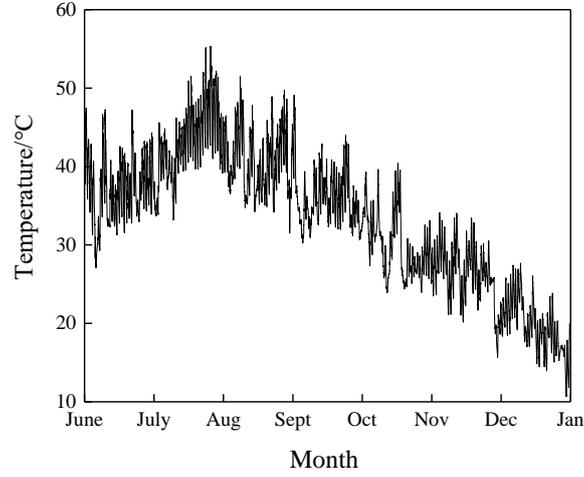
Fig. 5 shows the construction of the LSTM network in this paper, which includes an input layer, LSTM layer, fully connected layer, and output layer. The input layer is responsible for the input of the prepared time series data. The LSTM layer is constructed with LSTM units shown in Fig. 4. Finally, the LSTM layer's output is connected to the output layer by the linear connection of the fully connected layer to generate predictions.

3.2 Temperature data prediction using long short-term memory network

The temperature monitoring data from June 1, 2017 to December 31, 2017 of Nanjing Dashengguan Yangtze River Bridge is collected in this paper. Considering that the temperature values within 10 minutes change little, the average value within 10 minutes is regarded as the representative value within 10 minutes. In this case, there are 144 representative temperature values each day. For the convenience of representation, the temperature data collected from W_i ($i = 1, 2, \dots, 12$) are denoted as T_i ($i = 1, 2, \dots, 12$) And T_1 was selected as the training and testing sample of models, which is shown in Fig. 6.

Before training models, the data should be labeled and transformed into a supervised learning problem. After removing the Not a Number (NaN) data, considering the demand for time series data prediction, the data is preprocessed by the rolling method. The 31st value is predicted by the 1st to 30th values, the 32nd value is predicted by the 2nd to 31st values, the 33rd value is predicted by the 3rd to 32nd values, and so on. In other words, the model takes the temperature values of the first 5 hours as the input and predicts the temperature value at the next moment.

Then, the obtained data is normalized. The normalization process converts the data range into (0, 1), which is beneficial to improve the prediction accuracy of the model and accelerate the convergence speed (Sola and Sevilla 1997). The *Max - Min* normalization method is adopted in this paper, as shown in Eq. (7), in which X_{max} and X_{min} are the maximum and the minimum in the sample, respectively. Finally, the data is divided into the training set and testing set according to


 Fig. 6 Selected temperature sample T_1

the ratio of 3:1, which means that the first 75% and last 25% of the sample are regarded as the training set and testing set, respectively.

$$X_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (7)$$

The temperature prediction problem is a kind of regression problem, in which "mean square error (MSE)" is often used as the performance measure (Zhou 2016), as shown in Eq. (8). Here, it is adopted to evaluate model performance. The smaller the MSE value is, the more precise the model is

$$E(f; D) = \frac{1}{m} \sum_{i=1}^m (f(x_i) - y_i)^2 \quad (8)$$

In this equation, $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ is the sample set, y_i is the true value, and $f(x_i)$ is the predicted value.

The performances of the support vector machine (SVM), wavelet neural network (WNN), and long short-term memory (LSTM) network in predicting temperature data are compared using selected data. The kernel function, penalty coefficient, loss distance measurement, and kernel function parameters of SVM were selected by the grid search method and set as Gaussian kernel function, 10, 0.05, and 0.01, respectively. The commonly used wavelet basis function Morlet function was adopted as the activation function of WNN. The learning rate and regularization coefficient of WNN and LSTM network were determined by the grid search method and set as 0.0001 and 0, respectively. The number of hidden layer unit and was set as 64. 50 iteration epochs were performed in total, from which the optimal situation with minimum test MSE was selected. Compared with other random optimization methods, Adaptive Moment Estimation (Adam) algorithm performs better generally in practical applications (Kingma and Ba 2015). So Adam algorithm was adopted to optimize network parameters. Finally, to provide an intuitive and clear presentation with the real data, the predicted data that is between 0 and 1 is converted to typical temperature range using Eq. (9) that can be deduced from Eq. (8). The results are shown in Table 1 and Fig. 7.

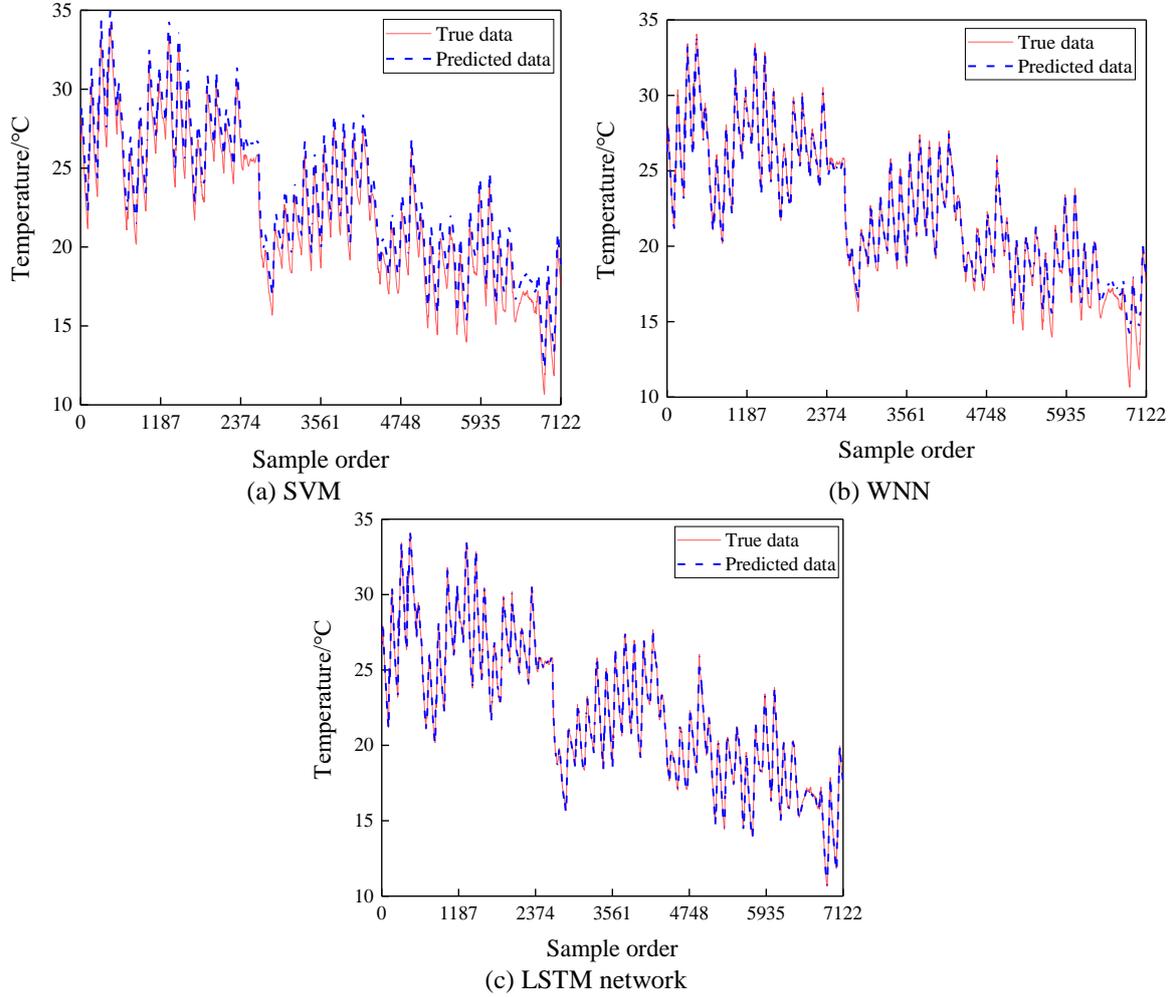


Fig. 7 Prediction effect of the three models on the testing set

Table 1 Precision of the three models

Prediction model	MSE value of the testing set
SVM	1.23
WNN	0.26
LSTM	6.27×10^{-3}

$$X_{typical} = X_{norm}(X_{max} - X_{min}) + X_{min} \quad (9)$$

As shown in Table 1, the prediction accuracy of the LSTM network is the highest, which is only about 1/196 and 1/40 of that of SVM and WNN, respectively. Meanwhile, the prediction accuracy of the neural network (WNN and LSTM network) is also higher than that of traditional machine learning model (SVM).

It can be seen from Fig. 7 that the prediction effect of SVM in the whole testing set is relatively consistent, while there is a significant deviation from the true values compared with WNN and LSTM network. WNN has a good prediction effect on the first 2/3 of the testing set, the prediction data is generally consistent with the real data, while the prediction effect has a significant decline on the last 1/3 of the testing set, it can be seen easily that the deviation is obvious. In the whole testing set, the predicted data of the LSTM network is highly consistent with the real data, and the two curves almost coincide, reflecting the good prediction performance.

The experimental results show that the LSTM network achieves the highest prediction accuracy among the three models, which fully demonstrates its advantages in time series data prediction.

4. Missing data recovery for structural temperature using LSTM network

Based on the above analysis, it can be seen that compared with WNN and SVM, LSTM network does have certain advantages in processing time series data such as temperature data. Therefore, the superior performance of the LSTM network in time series data prediction is utilized here to discuss its application in missing structural temperature data recovery. Assuming that T_1 is missing, T_2 , T_3 , and T_5 , which are near T_1 , were selected as the input to predict T_1 , in other words, to recover T_1 . And the recovery effects by using one sensor and multiple sensors were discussed. Similar to section 3, the learning rate and regularization coefficient were set as 0.0001 and 0, respectively. The previous 30 values of the intact sensors were used to predict the current value of the faulty sensor. The first 75% and last 25% of the sample were regarded as the training set and testing set, respectively. 50 iteration epochs were performed in total, from which the optimal situation with minimum test MSE was selected. Adam algorithm was adopted to optimize network parameters.

4.1 Missing data recovery using data of a single sensor

T_2 , T_3 , and T_5 were used as the input to recover T_1 , respectively. The temperature data was selected randomly to test the effectiveness of the model. The differences (absolute value) between recovered T_1 and real T_1 , mean square error between recovered T_1 and real T_1 , as well as the Pearson correlation coefficients between the input (T_2 , T_3 , and T_5) and T_1 were calculated. The results are shown in Table 2, Figs. 8 and 9. In general, this method recovers the missing data effectively, and the model accuracy increases with the increase of correlation coefficient. At the same time, it can be seen from Fig. 8 and Fig. 9 that most predictions are relatively accurate. The maximum and minimum differences are listed in Table 2, presenting the worst and best cases and therefore giving a comprehensive perspective on the recovery effect. When using T_5 which has a higher correlation coefficient as the input, the fluctuation of peak values is small, the maximum difference is only 5.25°C, much lower than those when using T_2 (10.89°C) and T_3 (10.00°C) as the inputs. Thus, using T_5 as the input has better recovery effect not only for general cases but also for extreme cases.

It is noteworthy here that the MSE value in Table 2 increases significantly than that in Table 1 using the same LSTM model, even greater than those of the SVM and WNN model in Table 1. Such

results are believed to be caused by different inputs with different correlation coefficients. The higher the correlation coefficient between two variables, the more accurate it is to predict another variable from one variable. This is because the higher the correlation coefficient is, the more covariant parts of the two variables would have, and the more one variable can learn from the other. In Section 3, T_1 itself is used as the input, making the correlation coefficient to be 1, which means the best and ideal case. So models using T_1 itself as the input can achieve much higher accuracy.

Besides, Fig. 10 is the daily change of MSE value on the testing set, namely the recovery effect of the model on different days. It can be seen that the MSE value of the last 1/3 part is higher and the prediction effect is poor. Pearson correlation coefficients between the first 2/3 and the last 1/3 of T_2 , T_3 , T_5 and T_1 were calculated respectively, and the results are shown in Table 3. It can be seen that the correlation coefficient between the last 1/3 of T_2 , T_3 and T_1 is lower than that between the first 2/3 of them, so does the prediction effect. While the correlation coefficient between the last 1/3 of T_5 and T_1 is close to that between the first 2/3 of them, and the prediction effect is relatively stable as well. It is further proved here that the larger the correlation coefficient is, the better the prediction effect will be.

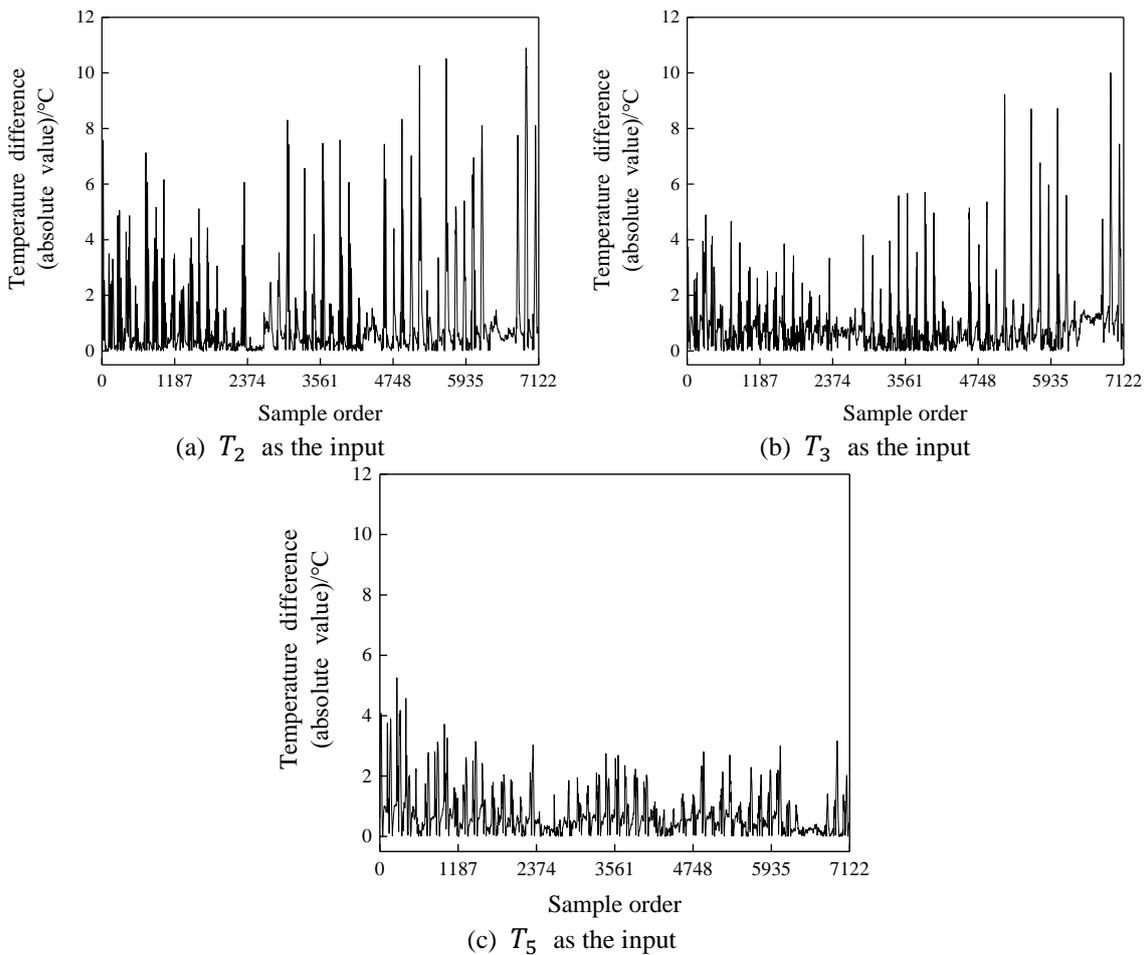


Fig. 8 Difference between recovery data and true data (absolute value) using a single sensor

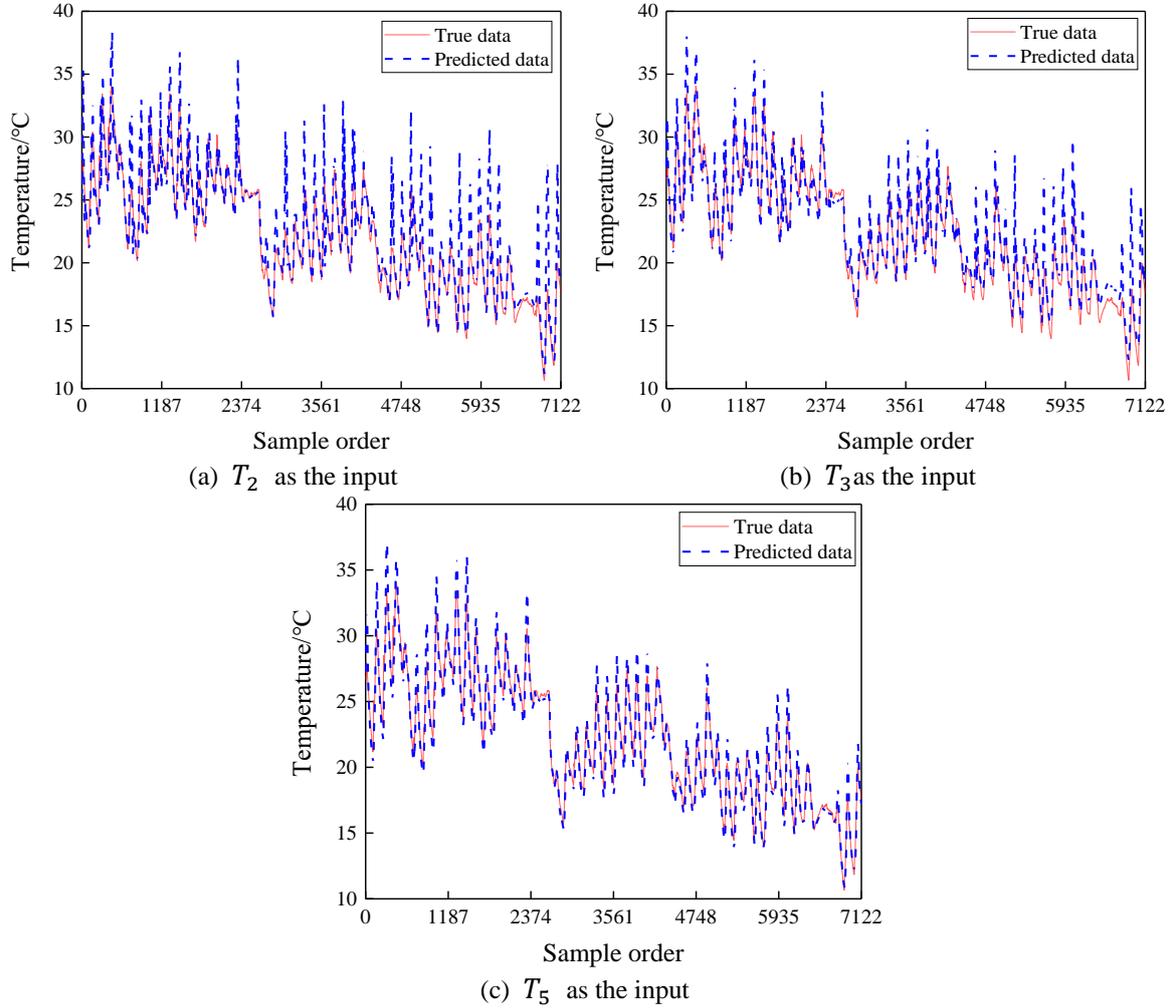


Fig. 9 Comparison between recovery data and true data using a single sensor

Table 2 Recovery effect using a single sensor

	T_2	T_3	T_5
Minimum difference/ $^{\circ}\text{C}$	5.00×10^{-6}	6.67×10^{-6}	1.05×10^{-4}
Maximum difference / $^{\circ}\text{C}$	10.89	10.00	5.25
MSE value	3.78	2.18	1.01
Correlation coefficient	0.945	0.979	0.991

Therefore, based on the LSTM network, structural temperature data of intact sensors can be used to recover the missing one of the faulty sensors. At the same time, selecting the data which has a higher correlation coefficient with the missing data as the input can achieve higher recovery accuracy.

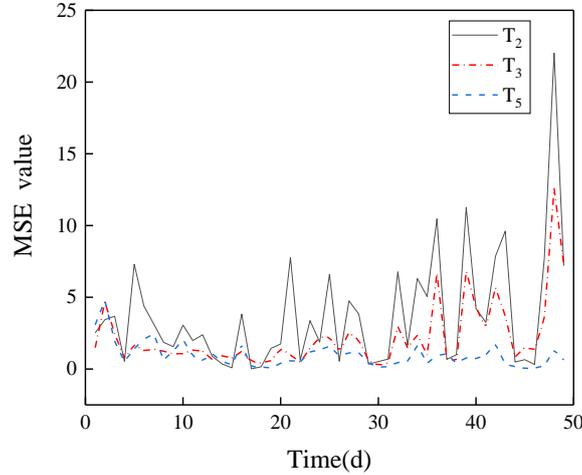


Fig. 10 Recovery effect on different days

Table 3 Correlation coefficients between different parts of T_2 , T_3 , T_5 and T_1

		T_2	T_3	T_5
Correlation coefficient	The first 2/3	0.929	0.974	0.988
	The last 1/3	0.812	0.852	0.971

4.2 Missing data recovery using data of multiple sensors

T_2 , T_3 , and T_5 were used together as the input to recover T_1 . The minimum and maximum differences (absolute value) between recovered T_1 and real T_1 , as well as the mean square error between recovered T_1 and real T_1 were calculated. The results are shown in Table 4. Compared with the recovery effect using a single sensor in Table 2, using multiple sensors can effectively improve the recovery accuracy. As the number of sensors that act as inputs increases, the recovery accuracy increases. For example, using $[T_2, T_3]$ as input reduces the MSE value by about 78% compared to using T_2 as input only, and using $[T_2, T_3, T_5]$ as input reduces the MSE value by about 57% further compared to using $[T_2, T_3]$ as input. At the same time, as shown in Table 4, when the input is $[T_2, T_3]$, the MSE value is 0.84, once the input contains T_5 which has higher correlation coefficient with T_1 , the MSE value is significantly reduced to 0.36. And the maximum difference between the predicted data and the true data is only about 3 °C. In addition, it can be seen from Fig. 11 and Fig. 12 that the combination of inputs containing T_5 which has a higher correlation coefficient with T_1 and the inclusion of more sensors as inputs can improve the recovery effect of peak values.

Therefore, incorporating more intact structural temperature sensors as inputs can further improve the missing data recovery effect of the faulty sensor. At the same time, including sensor data which has a higher correlation coefficient with data of the faulty sensor in the input combination can achieve higher recovery accuracy.

Table 4 Recovery effect using multiple sensors

	$[T_2, T_3]$	$[T_2, T_5]$	$[T_3, T_5]$	$[T_2, T_3, T_5]$
Minimum difference/°C	2.64×10^{-4}	4.33×10^{-6}	1.09×10^{-4}	6.70×10^{-5}
Maximum difference /°C	6.00	3.49	3.79	3.06
MSE value	0.84	0.42	0.52	0.36

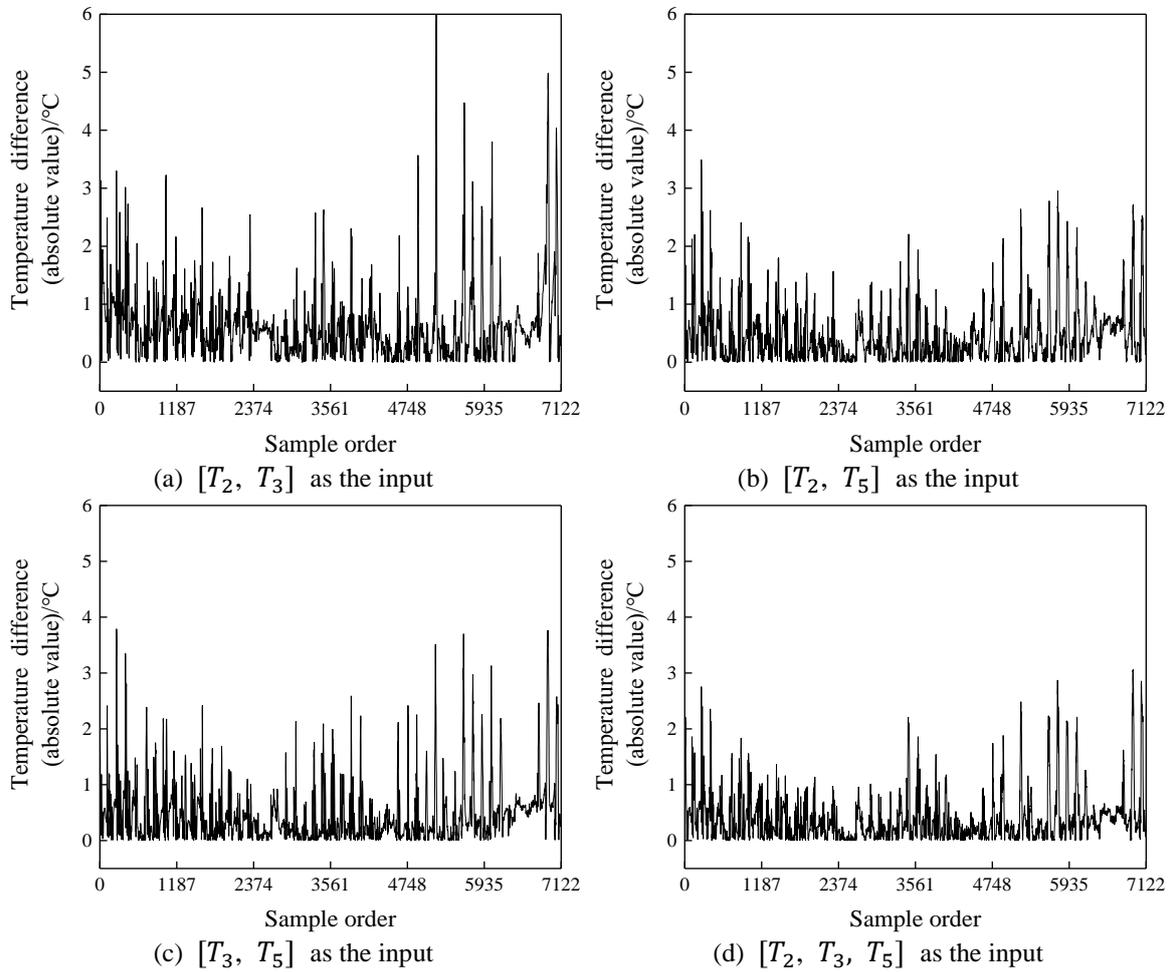


Fig. 11 Difference between recovery data and true data (absolute value) using multiple sensors

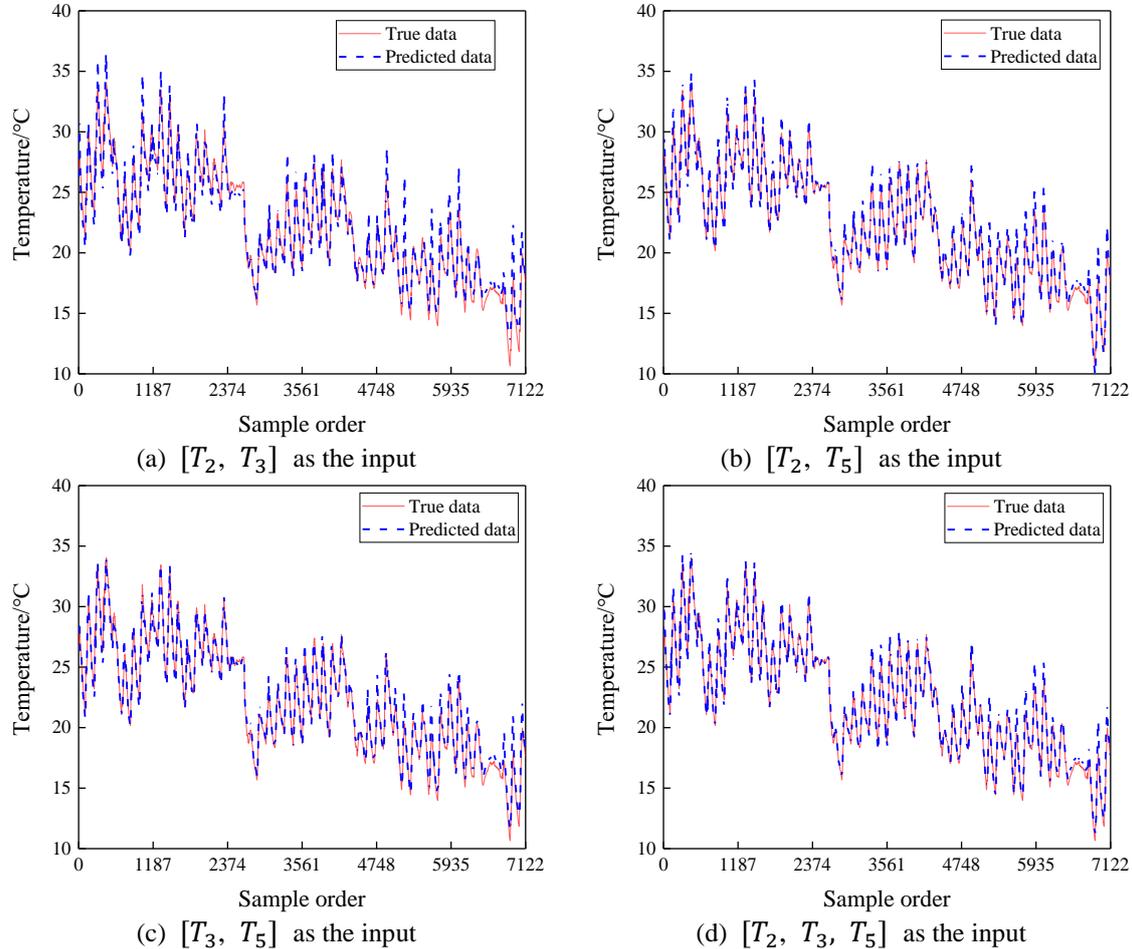


Fig. 12 Comparison between recovery data and true data using multiple sensors

5. Conclusions

This paper mainly discusses the effectiveness of LSTM network in the recovery of missing structural temperature data, with which more assessments for structures based on analysis between different kinds of monitoring data can be implemented. Experiments show that the LSTM network performs better in terms of accuracy than SVM and WNN in predicting time series data such as structural temperature. And LSTM network is an effective recovery method for the missing structural temperature data. In the recovery using a single sensor, selecting the sensor data which has a higher correlation coefficient with the missing data as the input can achieve higher recovery accuracy. Besides, incorporating more intact structural temperature sensors as inputs can further improve the recovery effect of missing data. At the same time, in the recovery using multiple sensors, including sensor data which has a higher correlation coefficient with data of the faulty sensor in the input combination can achieve higher recovery accuracy.

Finally, the SHM data itself is a kind of time series data. In view of the good effect of LSTM network in processing structural temperature data verified in this paper, its promotion and application in other SHM data, such as deflection and displacement, are also worth discussing.

Acknowledgments

The authors gratefully acknowledge the support of the Distinguished Young Scientists of Jiangsu Province (Grant. BK20190013), the National Natural Science Foundation of China (Grants. 51978154 and 51608258).

References

- Atha, D.J. and Jahanshahi, M.R. (2018), "Evaluation of deep learning approaches based on convolutional neural networks for corrosion detection", *Struct. Health Monit.*, **17**(5), 1110-1128. <https://doi.org/10.1177/1475921717737051>.
- Bao, Y., Shi, Z., Wang, X. and Li, H. (2018), "Compressive sensing of wireless sensors based on group sparse optimization for structural health monitoring", *Struct. Health Monit.*, **17**(4), 823-836. <https://doi.org/10.1177/1475921717721457>.
- Bao, Y., Tang, Z., Li, H. and Zhang, Y. (2019). "Computer vision and deep learning-based data anomaly detection method for structural health monitoring", *Struct. Health Monit.*, **18**(2), 401-421. <https://doi.org/10.1177/1475921718757405>.
- Bengio, Y., Simard, P. and Frasconi, P. (1994), "Learning long-term dependencies with gradient descent is difficult", *IEEE T. Neural Networ.*, **5**(2), 157-166. <https://doi.org/10.1109/72.279181>.
- Cha, Y.J., Choi, W. and Büyüköztürk, O. (2017), "Deep learning-based crack damage detection using convolutional neural networks", *Comput.-Aided Civil Infrastruct. Eng.*, **32**(5), 361-378. <https://doi.org/10.1111/mice.12263>
- Chen, Z., Bao, Y., Li, H. and Spencer, B.F. (2019), "LQD-RKHS-based distribution-to-distribution regression methodology for restoring the probability distributions of missing SHM data", *Mech. Syst. Signal Pr.*, **121**, 655-674. <https://doi.org/10.1016/j.ymssp.2018.11.052>.
- Ding, Y.L., Zhao, H.W. and Li, A.Q. (2017), "Temperature effects on strain influence lines and dynamic load factors in a steel-truss arch railway bridge using adaptive FIR filtering", *J. Perform. Constr. Fac.*, **31**(4), 04017024. [https://doi.org/10.1061/\(ASCE\)CF.1943-5509.0001026](https://doi.org/10.1061/(ASCE)CF.1943-5509.0001026).
- Hochreiter, S. and Schmidhuber, J. (1997), "Long short-term memory", *Neural Comput.*, **9**(8), 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>.
- Huang, H.B., Yi, T.H., Li, H.N. and Liu, H. (2020). "Strain-Based Performance Warning Method for Bridge Main Girders under Variable Operating Conditions", *J. Bridge Eng.*, **25**(4), 04020013. [https://doi.org/10.1061/\(ASCE\)BE.1943-5592.0001538](https://doi.org/10.1061/(ASCE)BE.1943-5592.0001538).
- Huang, Y., Beck, J.L., Wu, S. and Li, H. (2016), "Bayesian compressive sensing for approximately sparse signals and application to structural health monitoring signals for data loss recovery", *Probabilist. Eng. Mech.*, **46**, 62-79. <https://doi.org/10.1016/j.probengmech.2016.08.001>.
- Kingma, D.P. and Ba, J. (2015), "Adam: A method for stochastic optimization", *ICLR 2015*, San Diego, USA, May.
- Li, L., Liu, G., Zhang, L. and Li, Q. (2018). "Deep learning-based sensor fault detection using S-Long Short Term Memory Networks", *Struct. Monit. Maint.*, **5**(1), 51-65. <https://doi.org/10.12989/smm.2018.5.1.051>.
- Ni, Y.Q. and Li, M. (2016), "Wind pressure data reconstruction using neural network techniques: A comparison between BPNN and GRNN", *Measurement.*, **88**, 468-476. <https://doi.org/10.1016/j.measurement.2016.04.049>.

- Pei, X.Y., Yi, T.H., Qu, C.X. and Li, H.N. (2019). “Conditional information entropy based sensor placement method considering separated model error and measurement noise”, *J. Sound Vib.*, **449**: 389–404. <https://doi.org/10.1016/j.jsv.2019.02.035>.
- Qu, C.X., Yi, T.H. and Li, H.N. (2019). “Mode identification by eigensystem realization algorithm through virtual frequency response function”, *Struct. Control Health Monit.*, **26**(10), e2429. <https://doi.org/10.1002/stc.2429>.
- Rafiei, M.H. and Adeli, H. (2018). “A novel unsupervised deep learning model for global and local health condition assessment of structures”, *Eng. Struct.*, **156**, 598-607. <https://doi.org/10.1016/j.engstruct.2017.10.070>.
- Schmidhuber, J. (2015), “Deep learning in neural networks: An overview”, *Neural networks.*, **61**, 85-117. <https://doi.org/10.1016/j.neunet.2014.09.003>.
- Sola, J. and Sevilla, J. (1997), “Importance of input data normalization for the application of neural networks to complex industrial problems”, *IEEE T. Nuclear Sci.*, **44**(3), 1464-1468. <https://doi.org/10.1109/23.589532>.
- Thadikemalla, V.S.G. and Gandhi, A.S. (2018), “A data loss recovery technique using compressive sensing for structural health monitoring applications”, *KSCE J. Civil Eng.*, **22**(12), 5084-5093. <https://doi.org/10.1007/s12205-017-2070-z>.
- Wan, H. and Ni, Y.Q. (2019), “Bayesian multi-task learning methodology for reconstruction of structural health monitoring data”, *Struct. Health Monit.*, **18**(4), 1282-1309. <https://doi.org/10.1177/1475921718794953>.
- Ye, X.W., Su, Y.H., Xi, P.S. and Liu, H. (2017), “Structural health monitoring data reconstruction of a concrete cable-stayed bridge based on wavelet multi-resolution analysis and support vector machine”, *Comput. Concrete.*, **20**(5), 555-562. <https://doi.org/10.12989/cac.2017.20.5.555>.
- Zhang, A., Wang, K.C., Li, B., Yang, E., Dai, X., Peng, Y., Fei, Y., Liu, Y., Li, J. and Chen, C. (2017), “Automated pixel-level pavement crack detection on 3D asphalt surfaces using a deep-learning network”, *Comput. - Aided Civil Infrastruct. Eng.*, **32**(10), 805-819. <https://doi.org/10.1111/mice.12297>.
- Zhao, H.W., Ding, Y.L., Li, A.Q., Ren, Z.Z. and Yang, K. (2019), “Live-load strain evaluation of the prestressed concrete box-girder bridge using deep learning and clustering”, *Struct. Health Monit.*, 1475921719875630. <https://doi.org/10.1177/1475921719875630>.
- Zhou, Z.H. (2016), *Machine Learning*, Tsinghua University Press, Beijing, China.