Article

Long-Term Structural State Trend Forecasting Based on an FFT–Informer Model

Jihao Ma [†,‡] and **Jingpei Dan** [*,‡]

College of Computer Science, Chongqing University, Chongqing 400044, China
* Correspondence: danjingpei@cqu.edu.cn
† Current address: No. 174 Shazhengjie, Shapingba, Chongqing 400044, China.
‡ These authors contributed equally to this work.

Abstract: Machine learning has been widely applied in structural health monitoring. While most existing methods, which are limited to forecasting structural state evolution of large infrastructures. forecast the structural state in a step-by-step manner, extracting feature of structural state trends and the negative effects of data collection under abnormal conditions are big challenges. To address these issues, a long-term structural state trend forecasting method based on long sequence time-series forecasting (LSTF) with an improved Informer model integrated with Fast Fourier transform (FFT) is proposed, named the FFT–Informer model. In this method, by using FFT, structural state trend features are represented by extracting amplitude and phase of a certain period of data sequence. Structural state trend, a long sequence, can be forecasted in a one-forward operation by the Informer model that can achieve high inference speed and accuracy of prediction based on the Transformer model. Furthermore, a Hampel filter that filters the abnormal deviation of the data sequence is integrated into the Multi-head ProbSparse self-attention in the Informer model to improve forecasting accuracy by reducing the effect of abnormal data points. Experimental results on two classical data sets show that the FFT–Informer model achieves high and stable accuracy and outperforms the comparative models in forecasting accuracy. It indicates that this model can effectively forecast the long-term state trend change of a structure and is proposed to be applied to structural state trend forecasting and early damage warning.

Keywords: structural health monitoring; time series forecasting; FFT; Informer model; FFT–Informer

1. Introduction

Infrastructure structures play a vital role in maintaining the well-being of people, protecting significant capital investments, and promoting regional and national prosperity [1,2]. Structural health monitoring (SHM) [3–5], which collects a large amount of data that can reflect the states and changes of a structure, has attracted increasing attention in both academic and industrial communities to ensure the safe and reliable operation of infrastructure systems. Compared with short-term forecasting, long-term structural state forecasting is more focused on disaster early warning and damage repair for large infrastructures that actually is an LSTF task. However, infrastructure structures suffer from diverse types of potential damages during their service lives [6–8]. The nonlinear change process of structural state greatly increases the difficulty of structural state prediction.

Most machine learning forecasting methods have been widely applied in structural state forecasting. Ren et al. [9] established an incremental Bayesian matrix/tensor learning scheme to achieve efficient prediction of structural response. Zhang [10] applied ConvLSTM for three-year structural response forecasting. Suryanita [11] predicted the structural response using a backpropagation neural network (BPNN) method. Bahrami [12] applied an encoder–decoder architecture with gated recurrent unit (GRU) and long short-term memory (LSTM) for bridge response forecasting. Li [13] proposed a novel deep recurrent neural network

(RNN) model for long-term prediction. Yang [14] proposed a Bayesian modeling approach embedding model class selection for dynamic forecasting purposes. However, the trend of structural state forecasting is an LSTF task, while these methods in a step-by-step manner cannot predict the long-term changing of structural state efficiently. Although some statistical regression models [8,15–17] fully consider various constraints in specific situations, in order to achieve good results, these methods must fully consider the professional knowledge and influence factors of the prediction object, which is impossible to comprehensively consider under limited data collection conditions.

As the application of various prediction models continues in various domains, such as energy [18], economics [19], and disease propagation analysis [20], researchers are paingy more attention to the LSTF problems, which analyze a large amount of time series data and capture its potential relevance to make long-term forecasts. In order to successfully apply machine learning to an LSTF task, it is crucial to capture long-term dependence between input and predicted values. Recently,a Transformer model [21] with an encoder–decoder system has shown better performance in obtaining long-range dependency than recurrent neural network (RNN) models [22,23]. It brings the analysis, processing, and prediction of sequential data into a new stage. The Informer model [24], a new kind of LSTF model based on Transformer, is able to forecast long sequences in a one-forward operation instead of in a step-by-step manner, which achieves higher inference speed of prediction by introducing an encoder–decoder system. Meanwhile, the Informer model has lower computational complexity with the Multi-head ProbSparse self-attention mechanism compared to Transformer, which has been applied for LSTF tasks in many domains, such as wind power [25] and load [26]. By simply employing the Informer model in long-term structural state forecasting, however, it is easy to fall into the problem of invalid attention mechanism and an unobvious forecasting trend because of the complex characteristics of sensing data. Especially, due to the error of data collection and the influence of external factors, the deviation of some data will not affect the development of the data overall trend. Because the Multi-head ProbSparse self-attention focuses on the deviation point, this particularly affects the efficiency and accuracy of prediction.

In many LSTF tasks, feature extraction methods [27–29] have been applied to time series data, which are able to explain sequence relationships and help forecasting models learn the nonlinear characteristics of structural state data. Moreover, extracted features, which reflect the state of the structure to a certain extent, are also widely used in the SHM [30,31], especially the amplitude and phase of data [32,33]. Among them, FFT has been widely recognized for its characteristics of decomposing data to make them stable, fast convergence, and reflecting the trend to a certain extent [29,34].

Therefore, a long-term structural state trend forecasting method based on LSTF with the improved Informer model integrated with FFT is proposed, namedthe FFT–Informer model. Specially, by using FFT, the structural state trend feature is represented by extracting amplitude and phase of a certain period of a data sequence. Structural state trend, a long sequence, can be forecasted in a one-forward operation by the Informer model that can achieve high inference speed and accuracy of prediction based on the Transformer model. Furthermore, a Hampel filter that filters the abnormal deviation of a data sequence, is integrated into the Multi-head ProbSparse self-attention in the Informer model to improve forecasting accuracy by reducing the effect of abnormal data points. Experimental results on two classical data sets show that the FFT–Informer model achieves high and stable accuracy and outperforms the comparative models in forecasting accuracy. It indicates that this model can effectively forecast the long-term state trend change of a structure and is proposed to be applied to structural state trend forecasting and early damage warning.

The main contributions of our work are as follows:

(i)   An efficient LSTF method combined FFT with an improved Informer model is proposed, which successfully applies machine learning to long-term structural state forecasting in a one-forward operation rather than a step-by-step way.

(ii) The Multi-head ProbSparse attention in the Informer model is integrated with a Hampel filter, named the Multi-head ProbHamSparse attention, focuses on filtering abnormal deviations by setting the deviation upper limit, which reduces the impact of abnormal data items and obtains more accurate trends and dependency.

(iii) Experimental results on two classical data sets show that the FFT–Informer model achieves high and stable accuracy and outperforms the comparative models in forecasting accuracy. It indicates that the proposed method can be applied in real structural state forecasting.

The rest of this article is organized as follows. In Section 2, the proposed FFT–Informer model is elaborated in detail. The experiments and results are presented and discussed in Section 3. Finally, a conclusion and potential direction for future research works is given.

2. Long-Term Structural State Trend Forecasting Based on FFT–Informer Model

2.1. Overview

In this article, a long-term structural state forecasting method combined Fast Fourier transform (FFT) with an improved Informer model is proposed, as shown in Figure 1. The model mainly includes two parts: feature extraction by FFT and LSTF by the improved Informer model. Firstly, structural state data are processed into equally sized data blocks through sliding windows, which are not overlapping, and feature vectors (amplitude and phase) of each window are extracted by the FFT method. Then, using the improved Informer model, the future features are predicted in a one-forward operation. What is more, the forecasted feature vectors can be transformed back into time series data with a trigonometric function to obtain predicted long-term structural state in feature analysis.
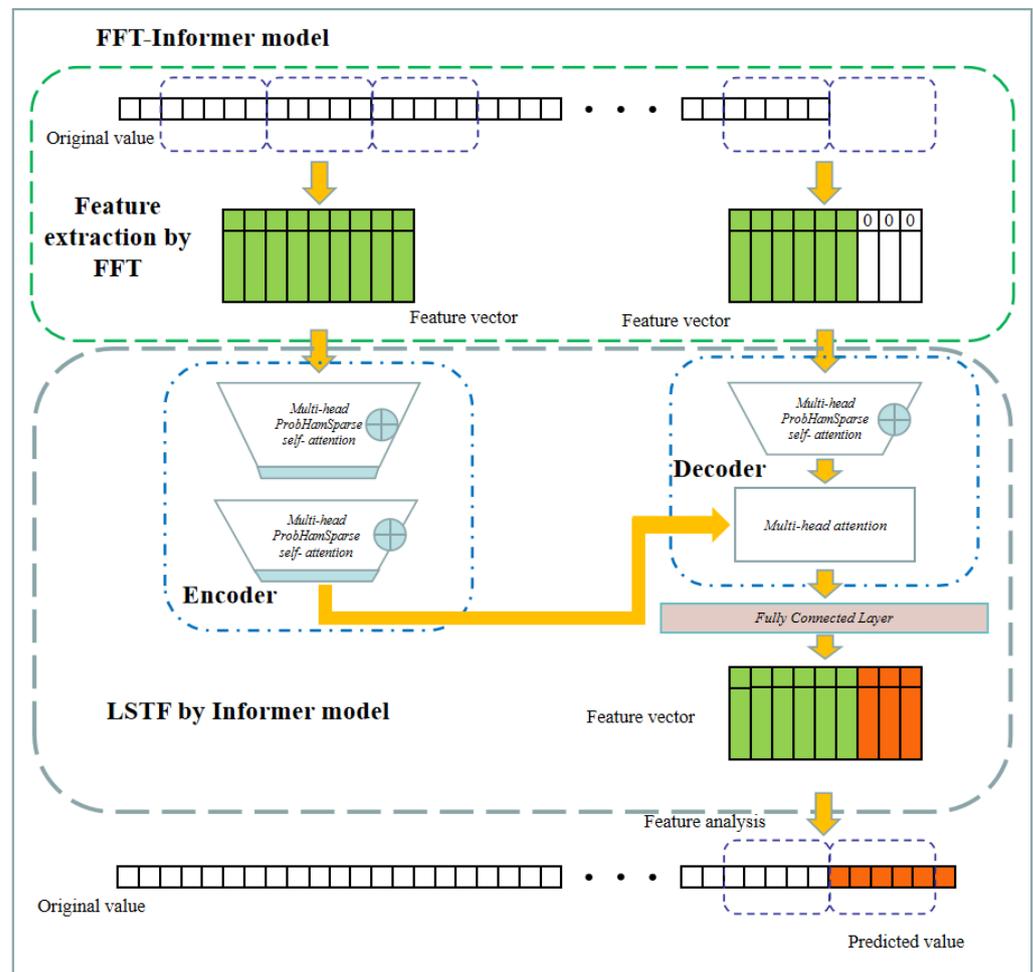

Figure 1. Framework of the FFT–Informer model.

2.2. Feature Extraction by FFT

Extracting appropriate features is essential to ensure the performance of machine learning algorithms. Fast Fourier transform (FFT) [35] is a feature extraction method often used to pre-process time series. As a time–frequency domain signal feature extraction method, the resulting approximation reflects the amplitude, vibration frequency, trend, and other characteristics of the original data to a certain extent and is suitable for further processing of the data. Because of the fluctuation characteristics of structural state data, FFT is applied to extract maximum amplitude and its phase key amplitude and phase of the series as the data characteristics.

The Fourier transform is applied widely in the signal processing field and can be regarded as a linear transformation operator. Fast Fourier transform (FFT) is a general term for applying computers to calculate Discrete Fourier transform (DFT), which is more efficient and faster. Let Z be a time series of length N. Using sliding windows, Z is divided into m data segments with length n. Its Discrete Fourier transform (DFT) is:

$$X(n) = \sum_{k=0}^{N-1} c_k e^{-j\frac{2\pi}{N}nk} \tag{1}$$

Amplitude and phase are extracted by FFT as features. Since the trigonometric function reflects the amplitude and phase well, the extracted features can be fitted with the original data through the trigonometric function as Equation (3). Vector A_m represents the feature information (maximum amplitude a_{m1}, its phase a_{m2}, and moving a_{m3}) obtained by trigonometric function fitting.

$$A_m = [a_{m1}, \cdots, a_{m3}] \tag{2}$$

$$\rho(x_i) = a_{m1} \sin\left(a_{m2}\frac{X\pi}{n/2}\right) + a_{m3} \tag{3}$$

The original structural state data Z is transformed into the feature vector A as shown in Figure 2. Equation (3) is used to fit the window data, in which a_{m1} and a_{m2} are fixed by maximum amplitude and its phase. The parameter vectors A_m as Equation (2), as the feature vectors, are the input of the Informer model.
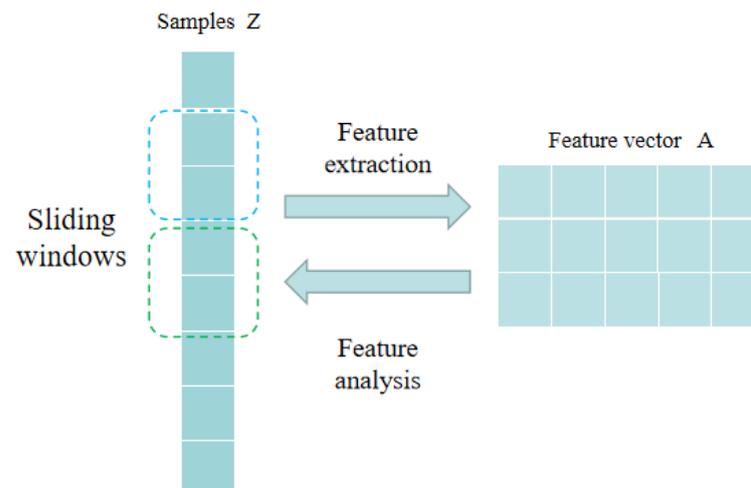

Figure 2. Feature extraction based on FFT.

In addition, the least-squares approximation solution is the optimal approximation of a time series minimalizing the sum of the squared distances between the data points (samples) and an approximating polynomial. So, the least-squares approximation is applied to fit the original data and Equation (3). The solution of the least-squares approximation

fits the coefficients of the approximating polynomial, minimizing the approximation error σ between the data, samples, and the polynomial.

$$\sigma = \frac{1}{n} \sum_{i=1}^{n-1} (p(x_i) - y_i)^2 \qquad (4)$$

where $p(x)$ is the approximating of the fitted data. The fitted coefficients of Equation (4) are then used as features of the data in this window.

2.3. LSTF by the Improved Informer Model

An Informer [24] with an encoder–decoder structure already has the ability to accomplish long-term sequence prediction. It is a non-autoregressive predictor, which greatly improves the inference performance in a forward operation. The prediction error rises steadily and slowly within the growing prediction range, which distinguishes it from the existing machine learning and statistical models. However, the Informer model is easily affected by abnormal deviation because of the deviation attention of the Multi-head ProbSparse self-attention.

Therefore, the Informer model based on an encoder–decoder is improved by changing the self-attention mechanism, as shown in Figure 3. In particular, a Hampel filter is integrated with the Multi-head ProbSparse self-attention, named the ProbHamSparse self-attention. This Informer model mainly consists of two parts. The left part in Figure 3 is the encoder, and the right part is the dncoder . In the encoder, the white trapezoid represents the Multi-head ProbHamSparse self-attention, and the blue trapezoid represents the self-attention distilling operation. This distilling operation is applied to extract dominating attention and reduce the network size sharply. In dncoder, the Multi-head ProbHamSparse self-attention is duplicated to improve the robustness. Then the Multi-head attention measures the weighted attention composition of the feature map. In addition, encoder and dncoder both receive massive long sequence inputs, which is depicted in green. However, dncoder pads the target elements into zero and finally predicts output elements in a generative style.
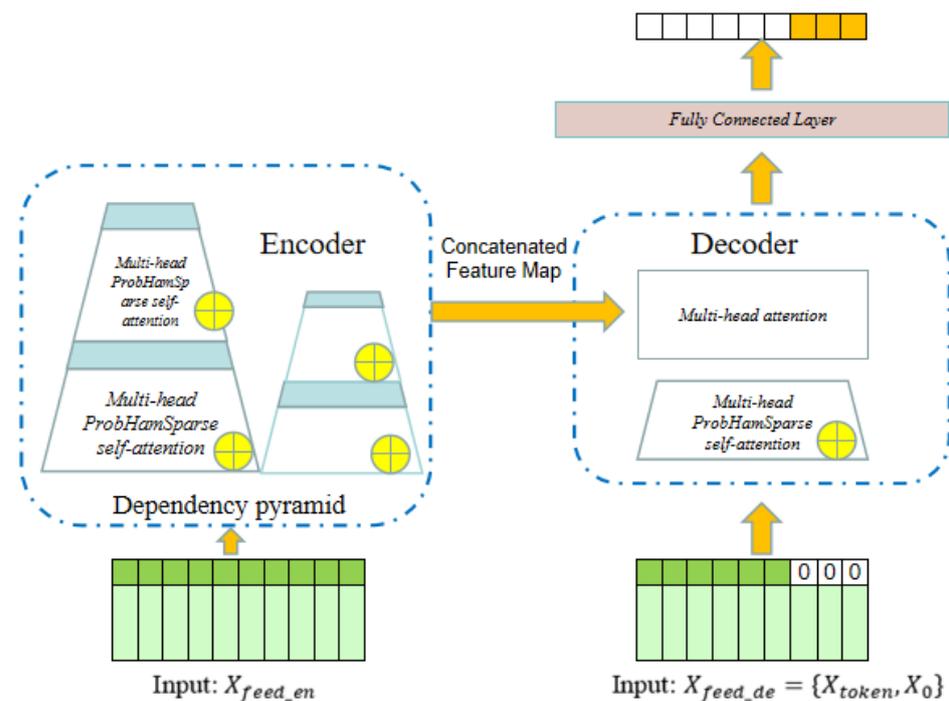

Figure 3. Improved Informer model.

2.3.1. Multi-Head ProbHamSparse Attention

Considering sparsity in the distribution of self-attention probability and the influence of abnormal data on the attention mechanism, the Multi-head ProbHamSparse self-attention is proposed to reduce computation consumption and make prediction more accurate. The ProbHamSparse self-attention is a development based on canonical self-attention and ProbSparse self-attention. Instead of performing a single attention function, the Multi-head attention is a progress from a single side to a multiple side in the Multi-head ProbHamSparse self-attention, as shown in Figure 4 which is applied in the attention mechanism of the Informer model and allows the model to jointly attend to information from different representation subspaces at different positions.

Canonical self-attention: The canonical self-attention [21] is defined on receiving the tuple input (query, key, value). It performs the scaled dot-product as Equation (5):

$$\varphi(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V$$

$$(Q \in R^{L_Q \times d}, K \in R^{L_K \times d}, V \in R^{L_V \times d}, \text{Softmax}(t) = \frac{1}{1 + e^{-t}})$$

(5)

where d is the input dimension; Q, K, and V represent query, key, and value and Softmax$(-)$ is a normalized exponential function. Equation (5) computes the dot products of the query with all keys, divides each by $\sqrt{d}$, and applies the Softmax function to obtain the weights on the values. To further discuss the canonical self-attention mechanism, q_i, k_i, v_i can represent the i-th row in Q, K, V; the i-th query's attention can be defined as a kernel smoother in a probability form:

$$\varphi(q_i, K, V) = \sum_j \frac{k(q_i, k_i)}{\sum_l k(q_i, k_l)} v_i$$

(6)

where $p(k_j|q_i) = \frac{k(q_i, k_i)}{\sum_l k(q_i, k_l)}$, and $k(q_i, k_j)$ selects the asymmetric exponential kernel $exp(\frac{q_i k_i^T}{\sqrt{d}})$. The relationship between k_i and q_i can be obtained based on computing the probability $p(k_j|q_i)$, which can be appiled to improve the prediction ability of the self-attention mechanism. This potential sparsity of distribution of self-attention is recognized by some studies [36,37]. It reveals that the performance can effectively reduce the computational burden if the irrelevant $p(k_j|q_i)$ can be distinguished.

ProbHamSparse self-attention: ProbHamSparse self-attention is proposed to improve the ProbSparse self-attention by combining a Hampel filter. ProbSparse self-attention is an improvement on the basis of canonical self-attention, which considers the sparsity of time series and the dependence of data points. Main contribution points and negligible points can be distinguished by ProbSparse self-attention. So, Informer can focus on the main contribution of a small number of dot products. The i-th query's attention on all the keys is defined as a probability $p(k_j|q_i)$, and the output is its composition with values V. The deviation points, as dominant dot-product pairs, encourage the corresponding query's attention probability distribution away from the uniform distribution. If $p(k_j|q_i) = 1/L_k$, $p(k_j|q_i)$ is close to a uniform distribution, the query is redundant to the attention. The ProbSparse self-attention mechanism measure the "likeness" through the Kullback–Leibler divergence:

$$KL(q||p) = ln \sum_{j=1}^{L_K} e^{\frac{q_i k_j^T}{\sqrt{d}}} - \frac{1}{L_k} \sum_{j=1}^{L_K} \frac{q_i k_j^T}{\sqrt{d}} - lnL_K$$

(7)

Dropping the constant, the sparsity measurement of i-th query is defined as:

$$\overline{M}(q_i, K) = \max_j \left(\frac{q_i k_j^T}{\sqrt{d}}\right) - \frac{1}{L_k} \sum_{j=1}^{L_K} \frac{q_i k_j^T}{\sqrt{d}}$$

(8)

where $\max\limits_{j}\left(\frac{q_i k_j^T}{\sqrt{d}}\right)$ is the Log–Sum–Exp (LSE) of q_i on the all keys, and $\frac{1}{L_k}\sum_{j=1}^{L_K}\frac{q_i k_j^T}{\sqrt{d}}$ is the arithmetic mean on them. The dominant dot-product pairs, which gain a larger $M(q_i, K)$, can be obtained in the header field of the long tail self-attention distribution. The attention mechanism can pay more attention to these points to reduce the computational burden.

Due to the ProbSparse self-attention focusing on distribution away from the uniform distribution, some outliers will still be selected as main contribution points through KL divergence, which will have an irreversible impact on the attention. Therefore, a Hampel filter is combined to filter the outliers by setting the upper bound of deviation. The Hampel filter is calculated by Formula (9):

$$|X - m_i| < 3\sigma \tag{9}$$

where X is the median value, and σ is the standard deviation.

Multi-head attention: Multi-head attention, which uses different and learned linear projections to project query, key, and value to K, Q, and V dimensions h times, respectively, is a progress from a canonical single side to multiple side. It allows the model to jointly attend to information from different representation subspaces at different positions, which is more effective than using the model query, key, and value to execute a canonical single attention function, as depicted in Figure 4.

Multi-head attention is defined as:

$$\text{Multihead}(Q, K, V) = \text{Concat}(head_1, \cdots, head_h)W^O \tag{10}$$

$$head_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \tag{11}$$

where the projection is parameter matrices $Q_i^Q \in R^{d_{model} \times d_k}$, $K_i^K \in R^{d_{model} \times d_k}$, $V_i^V \in R^{d_{model} \times d_k}$, $W^O \in R^{hd_v \times d_{model}}$. ProbHamSparse attention also uses this projection method, which changes from single attention to multiple side, named Multi-head ProbHamSparse attention.
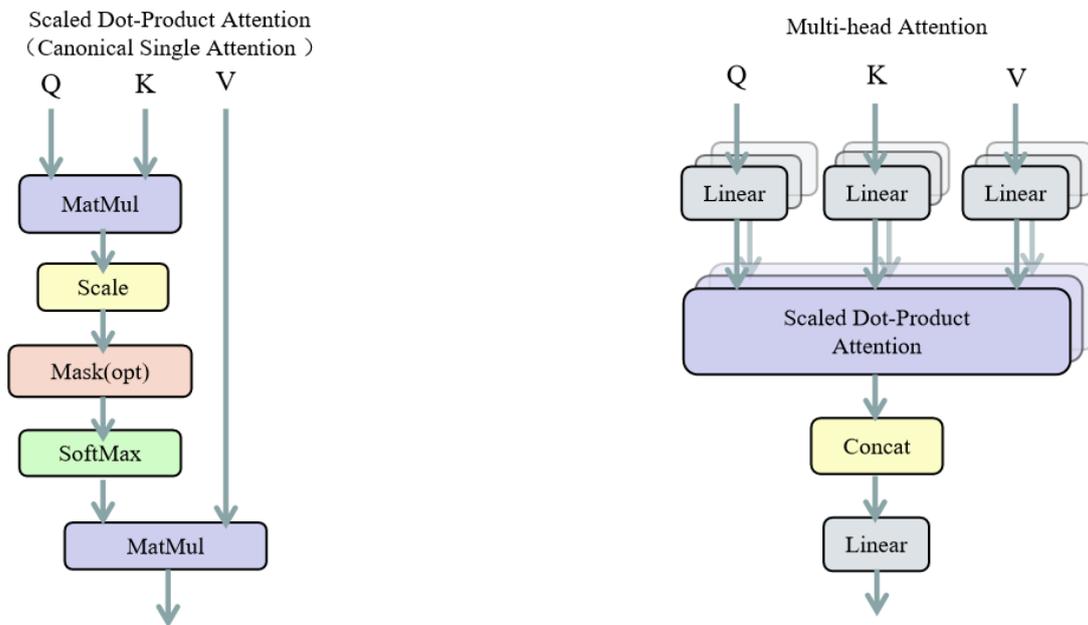

Figure 4. Canonical single self-attention function and Multi-head attention.

2.3.2. The Unified Input

In the LSTF task, global information is required to capture long-range dependence such as hierarchal time stamps. However, canonical self-attention hardly leverages this global information, which results in query–key mismatches between the encoder and the

decoder and impact the forecasting performance. In order to solve this problem, a unified input representation is applied in the Informer models, as shown in Figure 5.
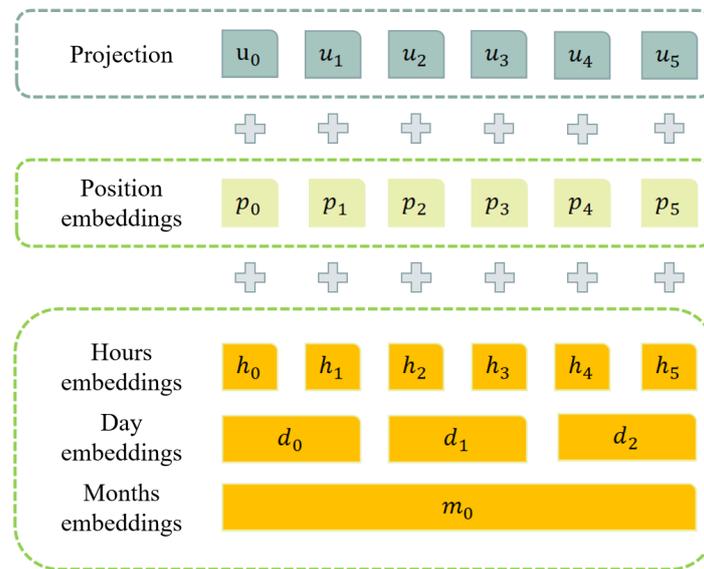

Figure 5. The input representation of the Informer model.

The local context is preserved by using a fixed position embedding first:

$$PE_{(pos,2j)} = \sin(pos/2L_x^{2j/d_{model}}) \tag{12}$$

$$PE_{(pos,2j+1)} = \cos(pos/2L_x^{2j/d_{model}}) \tag{13}$$

where d_{model} represents the feature dimension. Assuming the p types of global time stamps of t-th sequence input X^t are given, a learnable time stamp embedding SE is employed in each global time stamp with limited vocabulary size (taking second as the finest granularity). The similarity computation consumption can have access to a global context and capture long-range dependence. Then, the scalar context X_i^t is projected into d_{model}-dim vector u_i^t with a 1D convolution filter (kernel width = three, stride = one) to align the dimension.

$$X_{feed[i]}^t = \alpha u_i^t + PE_{(L \times (t-1)+i)} + \sum_p \left[SE_{(L \times (t-1)+i)} \right]_p \tag{14}$$

where α is the magnitude between the scalar projection and local/global embedding.

2.3.3. Encoder and Decoder

The encoder–decoder system is designed to extract the robust long-range dependency of long sequential inputs. As shown in Figure 3, in each layer of the encoder, the input is progressively decreased by the Multi-head ProbHamSparse self-attention. In order to avoid losing attentional points, encoder consists of the two same layers, and each layer has two sublayers. The first sublayer is the Multi-head ProbHamSparse self-attention mechanism, and the second sublayer is a simple fully connected feedforward network ELU. In order to facilitate the remaining connections, all sublayers and embedded layers in the model output the results with the same dimension.

However, a natural result of the ProbHamSparse self-attention mechanism is a redundant combination of values V. In order to solve this problem and ensure the connection between layers, a self-attention distilling operation is applied. This distilling operation favors dominating features over inferior ones and makes a focused self-attention feature

map in the next layer, which trims the input's time dimension sharply. The distilling operation forward from *j*-th layer into $(j + 1)$-th layer can be defined as:

$$X_{j+1}^t = \text{Maxpool}(\text{ELU}(\text{Conv1d}([X_j^t]_{AB}))).$$ (15)

where $[-]_{AB}$ represents the Multi-head ProbHamSparse self-attention; $\text{Conv1d}(-)$ performs a 1D convolutional filter on the time dimension and the $\text{ELU}(-)$ is the activation function.

The decoder also consists of a stack of *n* = 2 different layers, which is a standard decoder structure in the Transformer model. Different from the encoder, in the decoder, the first layer is composed of the ProbHamSparse self-attention and a full connection sublayer, and the second layer is composed of the Multi-head self-attention and a full connection sublayer. The Informer model provides the decoder with the following vectors:

$$X_{in}^t = \text{Concat}(X_{token}^t, X_O^t).$$ (16)

which ensures the accuracy of the decoder in predicting *X* through dependencies. Encoder–decoder is a forward conversion operation. Input and output are still unified form through the encoder and the decoder.

3. Experiments and Discussion

In order to verify the efficiency of our method on structural state data, extensive experiments on two typical data sets were conducted by comparing with different feature extraction and long-term structural state forecasting methods. All the experiments were conducted on AMD Ryzen 4600U CPU, 2.5 GHz, 16 GB RAM.

3.1. Datasets and Evaluating Measurements

The SMC data set [38] is a real data set. It is provided by SMC of Harbin Institute of Technology. More than 150 sensors were installed onto the Yonghe bridge, including 14 uniaxial accelerometers on the deck and one biaxial accelerometer at the top of the south tower. From 2005 to 2007, a large amount of data was collected, as shown in Figure 6a.

The ASCE benchmark data set [39] is an official data set. The data set was provided by the International Society for Structural Control and the American Society of Civil Engineering. The ASCE benchmark structural model was built in the Earthquake Engineering Research Laboratory at the University of British Columbia, Canada, as shown in Figure 6b. Three types of excitation were used to test it, including an electrodynamic shaker, an impact hammer, and ambient vibration. In our experiments, the data under ambient vibration was used for evaluation, which is the natural structural state without external force.

(a)

(b)
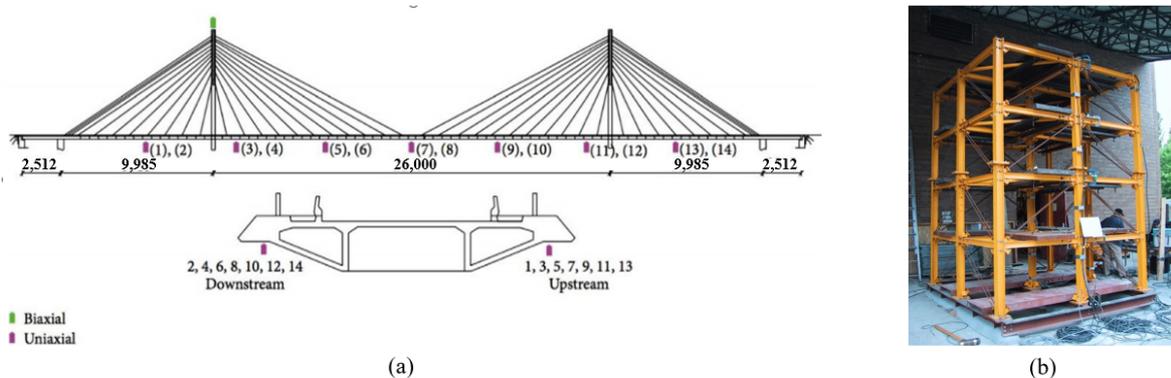
Figure 6. (**a**) Arrangement of accelerometers on Yonghe Bridge; (**b**) diagram of ASCE benchmark structural model.

In order to keep the structure in a progressive state, segment data were randomly selected from a sensor as the experimental data. Multiple segment data from multiple sensors were extracted to train and test the prediction performance, respectively. There are 150,000 data points in each segment of the SMC data set and 60,000 points in each segment of the ASCE data set. A total of 70% of each segment data set was used as the training set and the remaining 30% was used as the testing set.

In order to evaluate the forecasting accuracy and facilitate comparison, two classical evaluation metrics were applied, mean absolute error (MAE) and mean square error (MSE), as defined in (17) and (18), respectively.

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \bar{y}_i| \tag{17}$$

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y}_i)^2 \tag{18}$$

where y_i is the actual value, and $\bar{y}$ is the estimated forecasting value of the i-th test value.

3.2. Parameter Setting and Analysis

Since the FFT–Informer model is a combination of FFT and the improved Informer model, the main parameter settings of the FFT–Informer model were analyzed in two steps, as shown in Table 1. First, the parameter settings of feature extraction were analyzed by FFT. In order to ensure the flexibility of the experiments and verify the effectiveness of FFT, window lengths and sliding distances were fixed, which can be adjusted. For the SMC data set, the window length was set to 100 and the sliding distance to 100. However, due to the frequent vibration amplitude of the ASCE data set, it was difficult to fit the function with a long feature window, so the window length in the ASCE data set was 50.

Secondly, the parameters of the Informer model were analyzed. In this model, the attention window was used to capture the relationship between feature vectors in the window and long-term dependence. The training window length, which can also be adjusted, should be set beyond three times the forecasting length to ensure the correlation between the forecast and historical data. The training window length was initially set to 96 by trial and error, and the value of the predicted length was adjusted to verify the performance of long-term prediction in a one-forward operation. It is worth noting that feature vectors are predicted by the Informer model and can be transformed into prediction values by Equation (3), which greatly extends the prediction length as shown in the last line of Table 1. In addition, the number of Multi-heads was set to 8, and the dimensions of q_i, k_i, v_i were set to 512, which achieved the best results [24]. When the encoder–decoder layers are two, the model runs faster and the prediction effect is better.

Table 1. Parameter setting of FFT–Informer model.

		SMC	**ASCE**
	Feature window length	100	50
Feature extraction	Window sliding distance	100	50
	Feature vector length	3	3
	Attention window	96	96
	Predicted length	5 (5 × 100 = 500)	5 (5 × 50 = 250)
Informer model	Multi-head number	8	8
	Dimension of the vectors q_i, k_i, v_i	2	2
	Encoder–Decoder layers	2	2

3.3. Feature Extraction by FFT

In our experiment, the feature vectors were extracted from the data in sliding by FFT. These features can be interpreted as the best estimation of the amplitude and phase of the

structural state data. The fitted curve is directly interpreted as the characteristic estimator of the window data, and its parameter vectors can be regarded as the feature vectors of the windows. The SMC data set is taken as an example to illustrate this step in Figure 7.

As shown in Figure 7, under a window length of 100, the fitting curve of feature extraction basically simplifies and fits the data features of each window, and the amplitude, vibration frequency, and other features of structural state data are preserved by the simplification.
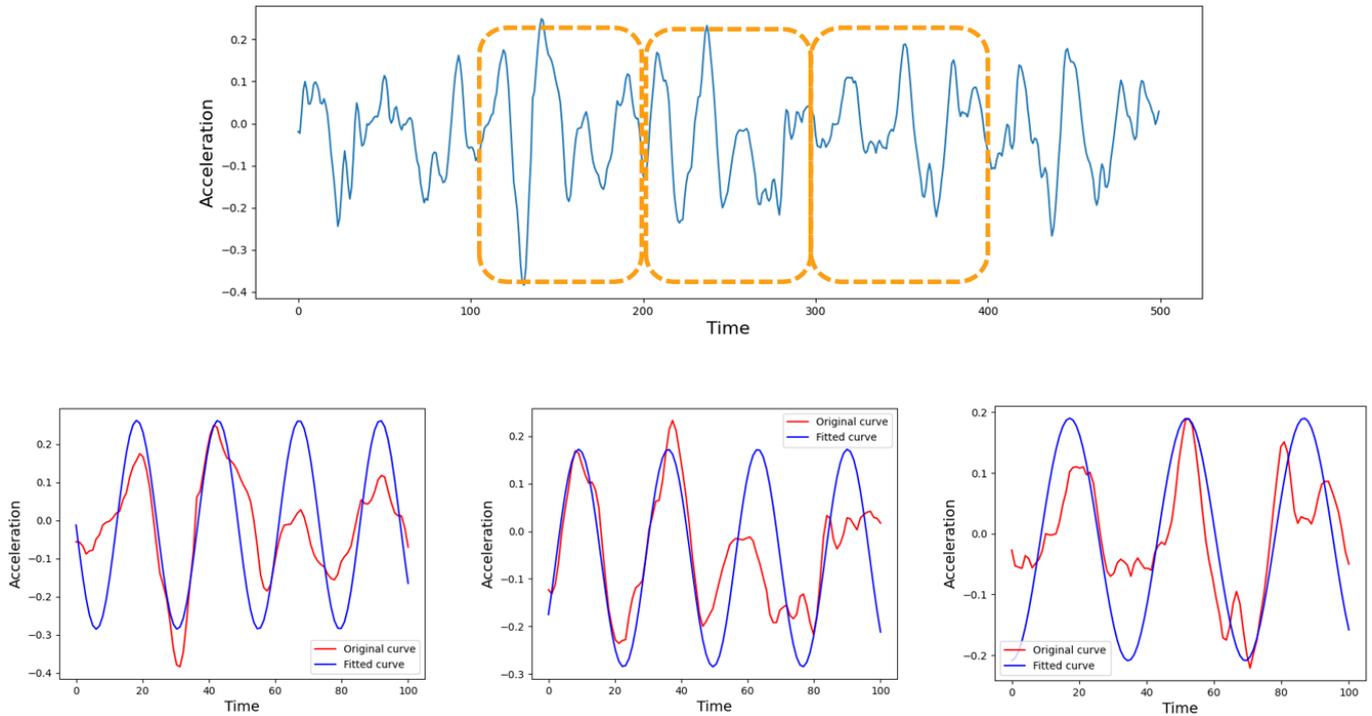
Figure 7. Fitting curve of feature vectors by FFT.

To verify the performance of FFT for trend feature extraction, it is compared with several time-frequency domain feature extraction methods that can also capture the characteristics of amplitude and frequency and have the ability to assist the forecasting algorithm, as shown in Table 2. Among them, direct signal extraction does not simplify the state characteristics between segments, stabilize the data, and reduce the randomness of the data. Each feature (extreme value, disturbance, etc.) still exists independently. The time–frequency domain feature extraction methods (Fourier polynomial, wavelet transform, and least-squares approximation) need more items to achieve the efficiency function fitting, which greatly increases the complexity of prediction.

Table 2. Number of features collected in unit window for each feature extraction method.

Feature Extraction by FFT	Least-Squares Approximation	Direct Signal Extraction	Fourier Polynomial	Wavelet Transform
3	5	1 or 1	20	12

3.4. Long-Term Structural State Trend Forecasting

The Informer model obtains the correlation of data in and between windows through the overall training of windows. The prediction length can be set to different values, but it needs to be less than the training window length. The one-forward prediction is carried out through the window.

First, feature vectors of the data set are obtained through the feature extraction by FFT in the previous step, and they are converted into data tensor. The next step is to add location information in the attention window, which can capture a correlation between

the data in the training windows. Finally, the model parameters with an Adam optimizer are optimized. Adam optimizer [24] is a gradient descent algorithm with an adaptive learning rate, which can not only adapt to a sparse gradient, but also alleviate the problem of gradient oscillation. The training error decreases rapidly and becomes stable as training proceeds and parameter adjustment in epochs as shown in Figure 8.
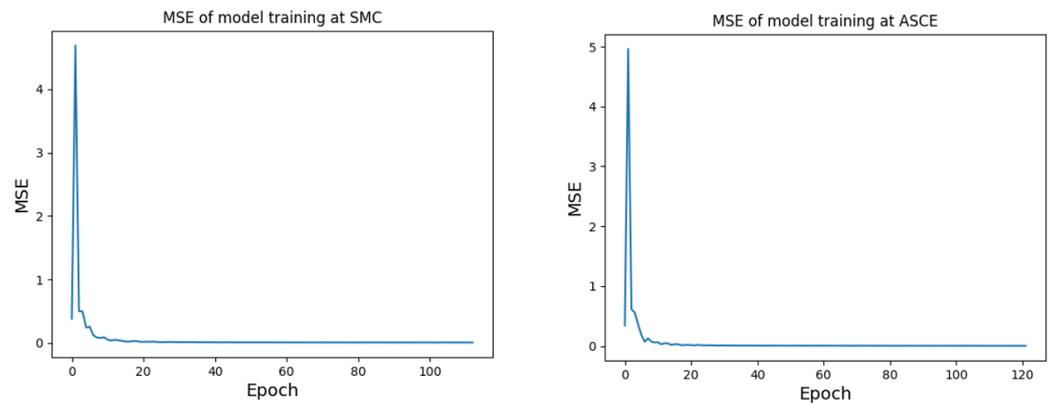

Figure 8. Training error change during the training phase.

By the early data processing and parameters optimization, the parameters of the model are determined before the prediction. Like training, the data required by the Informer model forecasting is still in tensor form, and the size of the forecasting windows is the same as that of the training windows. The forecasting feature vectors can be obtained by adjusting the length of the model in a one-forward operation. In this experiment, the vibration state data collected by SMC sensors are predicted directly, and the prediction length is set to five eigenvectors, as shown in Figure 9. The feature forecasting is carried out section by section in the form of windows, and the change of actual value is well captured. In order to verify its flexibility and consider the impact of noise, the data of the same sensor separated by half a year are selected to learn the model, which predicts different structural states under different structural condition, as shown in Figure 10, the change of structural state can still be captured well. The forecasting windows fit the changes of data well and achieve good forecasting results.
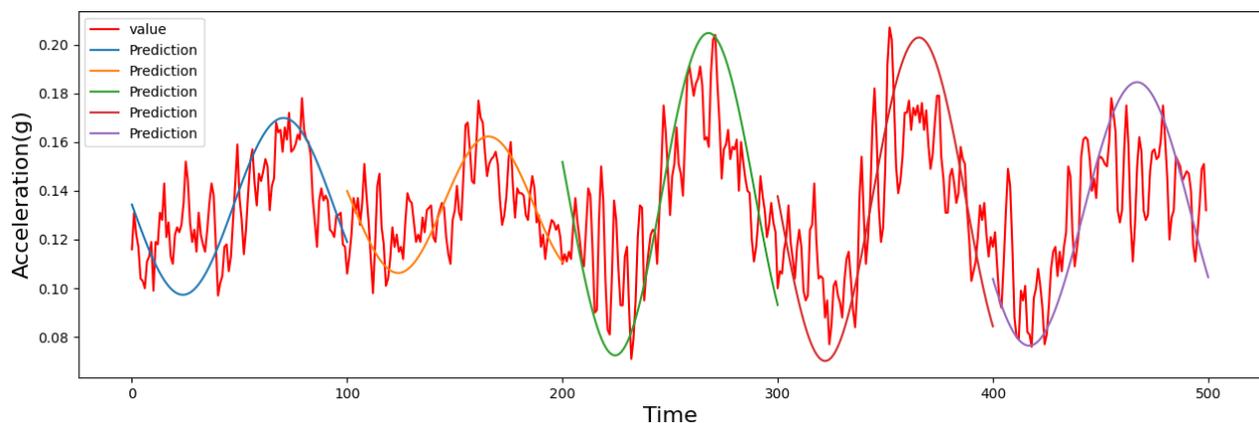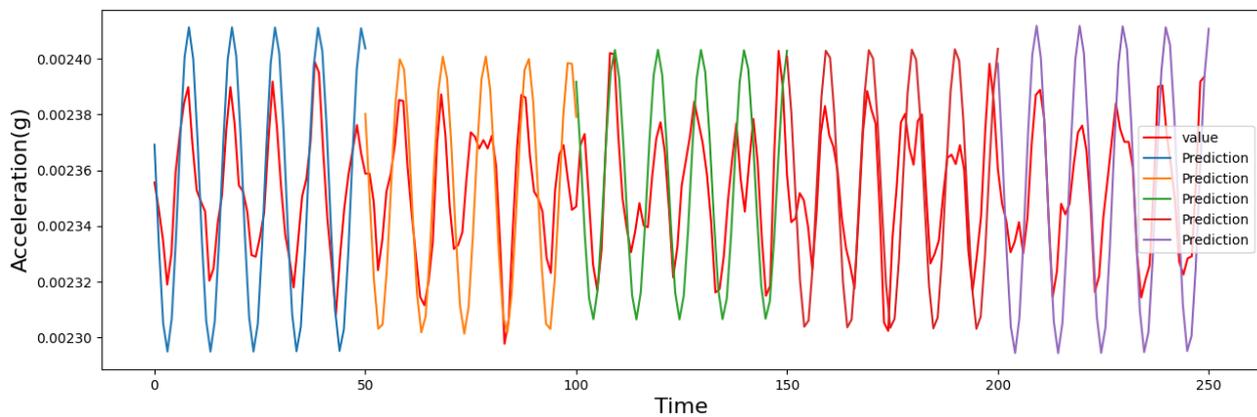

Figure 9. Structural state trend forecasting based on the FFT–Informer on the SMC data set. (Time represents the number of the predicted value).
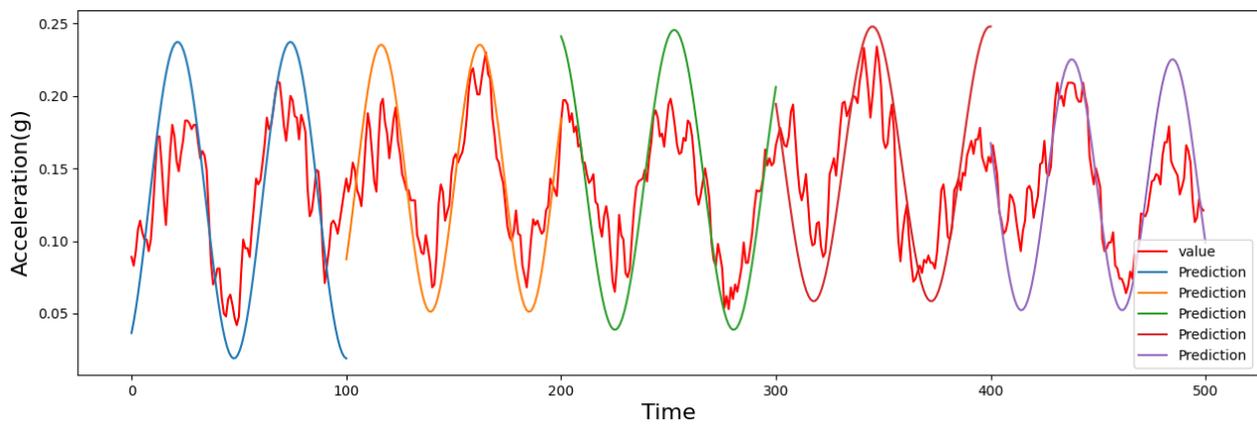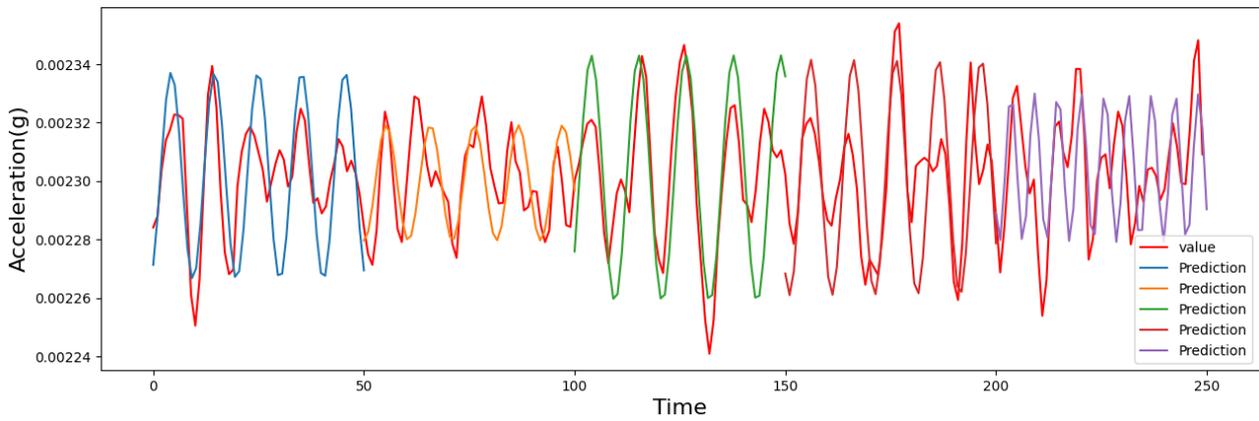
Figure 10. Structural state trend forecasting separated by half a year at different structure status.

In order to fully verify the effectiveness of the model, we also forecast the structural state trend using ambient vibration data in the ASCE data set. Different from the data in the SMC, the data in the ASCE fluctuate more obviously, so the windows size of feature extraction is reduced to 50 to better fit trend features. The forecasting length is still five window lengths, as shown in the Figure 11. In order to effectively prove our model applicable to general cases, the predicted position is also adjusted as in the previous experiment to verify that the prediction is not sensitive to the change of structural state, as shown in Figure 12. Experimental results show that the model can be well-applied for a general case and have good forecasting performance.
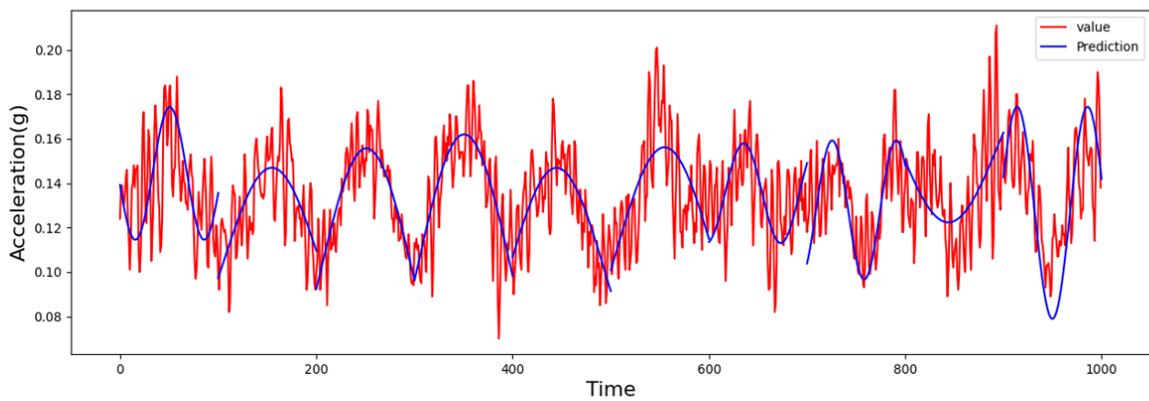

Figure 11. Structural state trend forecasting based on long FFT–Informer on the ASCE benchmark data set.

The Informer model is a nonautoregressive predictor through a forward operation. The length of prediction can also be adjusted directly to predict long time series. As shown in Figure 13, in order to show long-term forecasting performance, the forecasting length is adjusted to 10 and 15, in which each represents the feature vectors of 100 values. When changing the data set, the size of the feature extraction window and the forecasting length, the FFT–Informer model still performs well in long time series forecasting.

Figure 12. Structural state trend forecasting on the ASCE benchmark data set separated by half a year at different structure status.

(a) 10 forecasting lengths
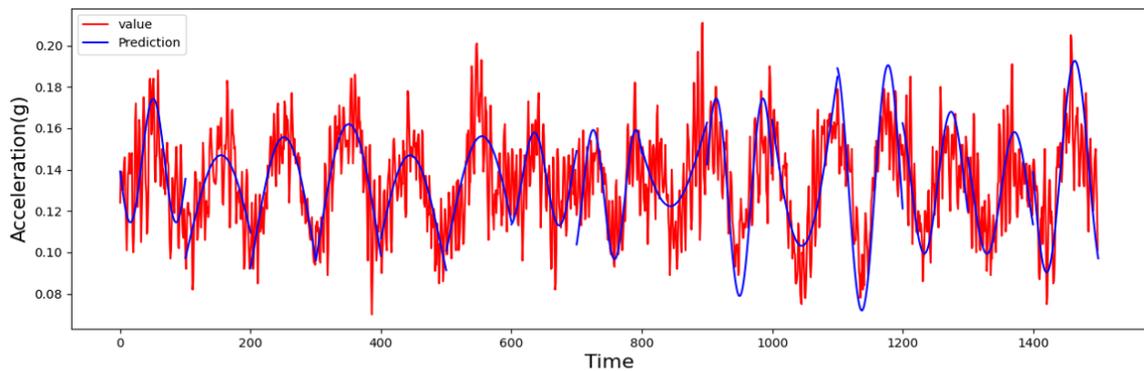
(b) 15 forecasting lengths

Figure 13. Structural state forecasting based on the FFT–Informer on 10 and 15 forecasting lengths.

In order to conduct a comparative experiment, these feature extraction methods are combined with Informer to compare with FFT–Informer forecasting, which also shows the advantages and disadvantages of each feature extraction method combined with Informer. Table 3 indicates that: (1) compared with the Informer model, the MAE of the FFT–Informer model is reduced by 17%, and the MSE of the FFT–Informer model is reduced by 29% on average on two data sets, respectively. It indicates that FFT has a better ability to improve the prediction ability of the Informer model. (2) Compared with other feature extraction methods-based Informer models, the MAE and MSE of the FFT–Informer is the smallest, which indicates that FFT has a better ability to extract features and aid in prediction.

In addition, forecast length is increased to 15, and the predicting accuracy of the FFT–Informer model is compared with the Informer model and some commonly used forecasting models, including LSTM, the echo state network (ESN), and the autoregressive integrated moving average model (ARIMA) ARIMA. Table 4 indicate that: (1) Compared with the ARIMA model, which is a statistical regression model, the MAE of the FFT–Informer model is reduced by 61% ,and the MSE of the FFT–Informer model is reduced by 59% on average. (2) Compared with the ESN and LSTM model, which are machine learning forecasting methods, the MAE of the FFT–Informer model is reduced by 43% and 37%, and the MSE of the FFT–Informer model is reduced by 49% and 41% on average. (3) With the increase of prediction length, the prediction ability of the FFT–Informer can remain relatively stable compared with others. These indicate the FFT–Informer has good prediction performance and stability.

Table 3. Comparison of the prediction results of the combination of each feature extraction method for the long-term forecasting algorithm and Informer. (* represents the combination with Informer model. 5 and 10 represents five and ten feature windows. Bold value is the minimum value).

Method	Metric	SMC		ASCE	
		5	10	5	10
FFT–Informer	MAE	**0.084**	**0.119**	**0.091**	**0.098**
	MSE	**0.017**	**0.028**	**0.029**	**0.031**
Least-squares approximation *	MAE	0.095	0.131	0.034	0.041
	MSE	0.101	0.126	0.03	0.039
Direct signal extraction *	MAE	0.123	0.167	0.146	0.175
	MSE	0.049	0.062	0.067	0.053
Fourier *	MAE	0.114	0.127	0.114	0.132
	MSE	0.048	0.052	0.054	0.062
Wavelet transform *	MAE	0.121	0.123	0.110	0.118
	MSE	0.044	0.052	0.031	0.056
Informer	MAE	0.111	0.135	0.117	0.121
	MSE	0.024	0.029	0.027	0.031

Table 4. Comparison of long-term forecasting methods. (Bold value is the minimum value).

Method	Metric	SMC			ASCE		
		5	10	15	5	10	15
FFT–Informer	MAE	**0.084**	**0.119**	**0.132**	**0.091**	**0.098**	**0.113**
	MSE	**0.017**	**0.028**	**0.042**	**0.021**	0.036	**0.048**
Informer	MAE	0.111	0.135	0.147	0.117	0.121	0.134
	MSE	0.024	0.029	0.046	0.027	**0.031**	0.049
ARIMA	MAE	0.216	0.181	0.298	0.196	0.165	0.234
	MSE	0.053	0.062	0.069	0.055	0.062	0.072
ESN	MAE	0.174	0.197	0.213	0.186	0.192	0.226
	MSE	0.046	0.052	0.072	0.051	0.059	0.079
LSTM	MAE	0.142	0.191	0.231	0.163	0.212	0.246
	MSE	0.037	0.041	0.062	0.045	0.057	0.069

4. Conclusions

In order to construct machine-learning-based long-term forecasting methods to obtain the long-term dependence between historical data and prediction while maintaining high forecasting accuracy, a long-term structural state trend forecasting method based on LSTF with the improved Informer model integrated with FFT is proposed, named the FFT–Informer model. In this method, FFT extracts amplitude and phase, and the improved Informer model forecasts trends, fully considering trend development and the impact of

abnormal deviation. The method has a good performance to capture long-term dependence between input and predicted values. Experimental results on two classical data sets show that the FFT–Informer model achieves high and stable accuracy and outperforms the comparative models in forecasting accuracy, which indicates that this model can effectively forecast the long-term state trend change of structures and is proposed to be applied to structural state trend forecasting and early damage warning.

Abbreviations

The following abbreviations are used in this manuscript:

LSTF	Long sequence time-series forecasting
FFT	Fast Fourier transform
SHM	Structural health monitoring
LSTM	Long Short-term Memory
RNN	Recurrent Neural Network
BPNN	Backpropagation Neural Network
DFT	Discrete Fourier transform

References

1. Liu, Y.F.; Nie, X.; Fan, J.S. Image-based crack assessment of bridge piers using unmanned aerial vehicles and three-dimensional scene reconstruction. *Comput.-Aided Civ. Infrastruct. Eng.* **2020**, *35*, 511–529. [CrossRef]
2. Rafiei, M.H.; Adeli, H. A novel machine learning-based algorithm to detect damage in high-rise building structures. *Struct. Des. Tall Spec. Build.* **2017**, *26*, e1400. [CrossRef]
3. Mishra, M.; Lourenco, P.B.; Rammana, G.V. Structural health monitoring of civil engineering structures by using the internet of things: A review. *J. Bulid. Eng.* **2022**, *48*, 103954. [CrossRef]
4. Sony, S.; Laventure, S.; Sadhu, A. A literature review of next-generation smart sensing technology in structural health monitoring. *Struct. Control Health Monit.* **2019**, *26*, e2321. [CrossRef]
5. Han, B.G.; Ding, S.Q.; Yu, X. Intrinsic self-sensing concrete and structures: A review. *Measurement* **2015**, *59*, 110–128. [CrossRef]
6. Wang, X.Y.; Yang, K.; Shen, C.S. Study on MPGA-BP of gravity dam deformation prediction. *Math. Probl. Eng.* **2017**, *6*, 2586107. [CrossRef]
7. Hong, S.; Park, C.; Cho, S. A Rail-temperature-prediction Model Based on Machine Learning: Warning of Train-Speed Restrictions Using Weather Forecasting. *Sensors* **2021**, *21*, 4606. [CrossRef]
8. Wang, H.; Zhang, Y.M.; Mao, J.X. Modeling and forecasting of temperature-induced strain of a long-span bridge using an improved Bayesian dynamic linear model. *Eng. Struct.* **2019**, *192*, 220–232. [CrossRef]
9. Ren, P.; Chen, X.Y.; Sun, L.J. Incremental Bayesian matrix/tensor learning for structural monitoring data imputation and response forecasting. *Mech. Syst. Signal Process.* **2021**, *158*, 107734. [CrossRef]
10. Zhang, R.Y.; Meng, L.; Mao, Z.; Sun, H. Spatiotemporal Deep Learning for bridge Response Forecasting. *J. Struct. Eng.* **2021**, *147*, 04021070. [CrossRef]
11. Suryanita, R.; Maizir, H.; Firzal, Y.; Jingga, H.; Yuniarto, E. Response prediction of multi-story building using backpropagation neural networks method. In Proceedings of the ICANCEE, Bali, Indonesia, 24–25 October 2018; Volume 276.
12. Bahrami, O.; Hou, R.; Wang, W.; Lynch, J.P. Time series forecasting to jointly model bridge responses. In *Bridge Maintenance, Safety, Management, Life-Cycle Sustainability and Innovations*; CRC Press: Boca Raton, FL, USA, 2021; pp. 299–307.
13. Li, T.; Pan, Y.X.; Tong, K.T.; Ventura, C.E.; de Silva, C.W. Attention-Based Sequence-to-Sequence Learning for Online Structural Response Forecasting Under Seismic Excitation. *IEEE Trans. Syst. Man Cybern.-Syst.* **2022**, *52*, 2184–2200. [CrossRef]

14. Wang, Y.W.; Ni, Y.Q. Bayesian dynamic forecasting of structural strain response using structural health monitoring data. *Struct. Control Health Monit.* **2020**, *27*, e2575. [CrossRef]

15. Yang, N.; Bai, X.B. Forecasting structural strains from long-term monitoring data of a traditional Tibetan building. *Struct. Control Health Monit.* **2019**, *26*, e2300. [CrossRef]

16. Buckley, T.; Pakrashi, V.; Ghosh, B. A dynamic harmonic regression approach for bridge structural health monitoring. *Struct. Health Monit.-Int. J.* **2021**, *20*, 3150–3181. [CrossRef]

17. Fan, X.P.; Liu, Y.F. Dynamic extreme stress prediction of bridges based on nonlinear mixed Gaussian particle filtering algorithm and structural health monitoring data. *Adv. Mech. Eng.* **2016**, *8*, 1–10. [CrossRef]

18. Somu, N.; Raman, M.R.G.; Ramamritham, K. A hybrid model for building energy consumption forecasting using long short term memory network. *Appl. Energy* **2020**, *261*, 114131. [CrossRef]

19. Sezer, O.B.; Gudelek, M.U.; Ozbayoglu, A.M. Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. *Appl. Soft Comput.* **2020**, *90*, 106181. [CrossRef]

20. Abbasimehr, H.; Paki, R. Prediction CoVID-19 confirmed cases combining deep learning methods and Bayesian optimization. *Chaos Solitons Fractals* **2021**, *142*, 110511. [CrossRef]

21. Vaswani, A.; Shazeer, N. Attention is All You Need. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.

22. Raffel, C.; Shazeer, N.; Roberts, A. Exploring the Limits Transfer Learning with a Unified Text-to-Text Transformer. *J. Mach. Learn. Res.* **2020**, *21*, 5485–5551.

23. Gao, S.; Alawad, M.; Young, M.T. Limitation of Transformers on Clinical Text Classification. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 3596–3607. [CrossRef]

24. Zhou, H.Y.; Zhou, S.H.; Peng, J.Q. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. *AAAI* **2021**, *35*, 11106–11115. [CrossRef]

25. Huang, X.H.; Jiang, A.H. Wind Power Generation Forecast Based on Multi-Step Informer Network. *Energies* **2022**, *15*, 6642. [CrossRef]

26. Gong, M.J.; Zhao, Y.; Sun, J.W.; Han, C.T.; Sun, G.N.; Yan, B. Load forecasting of district heating system based on Informer. *Energy* **2022**, *253*, 124179. [CrossRef]

27. Perez-Ramirez, C.A.; Amezquita-Sanchez, J.P. Recurrent neural network model with Bayesian training and mutual information for response prediction of large buildings. *Eng. Struct.* **2019**, *178*, 603–615. [CrossRef]

28. Li, L.; Wu, Y. Research on machine learning algorithms and feature extraction for time series. In Proceedings of the 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Montreal, QC, Canada, 8–13 October 2017.

29. Kong, Z.; Zhang, C.; Lv, H.; Xiong, F.; Fu, Z. Multimodal Feature Extraction and Fusion Deep Neural Networks for Short-Term Load Forecasting. *IEEE Access* **2020**, *8*, 185373–185383. [CrossRef]

30. Li, M.W.; Xu, D.Y.; Geng, J.; Hong, W.C. A hybrid approach for forecasting ship motion using CNN-GRU-AM and GCWOA. *Appl. Soft Comput.* **2021**, *114*, 108084. [CrossRef]

31. Gharehbaghi, V.R.; Farsangi, E.N.; Yang, T.Y.; Hajirasouliha, I. Deterioration and damage identification in building structures using a novel feature selection method. *Structures* **2021**, *29*, 458–470. [CrossRef]

32. Li, Z.H.; He, J.B.; Liu, D.K.; Liu, N.X.; Long, Z.L.; Teng, J. Influence of Uniaxial Stress on the Shear-Wave Spectrum Propagating in Steel Members. *Sensors* **2019**, *19*, 492. [CrossRef]

33. Chen, Z.C.; Li, H.; Bao, Y.Q. Analyzing and modeling inter-sensor relationships for strain monitoring data and missing data imputation: A copula and functional data-analytic approach. *Struct. Health Monit.-Int. J.* **2019**, *18*, 1168–1188. [CrossRef]

34. Shi, J.Q.; Si, G.Q.; Li, S.W.; Oresanya, B. Feature extraction based on the fractional Fourier transform for vibraton signals with application to measuing the load of a tumbling mill. *Control Eng. Pract.* **2019**, *84*, 238–246. [CrossRef]

35. Duhamel, P.; Vetterli, M. Fast Fourier-Transforms. *Signal Process.* **1990**, *19*, 259–299. [CrossRef]

36. Li, S.Y.; Jin, X.Y.; Xuan, Y.; Zhou, X.Y.; Chen, W.H.; Wang, Y.X. Enhancing the Locality and Breaking the Memory Bottleneck of Transformer on Time Series Forecasting. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 1–11.

37. Beltagy, I.; Peters, M.; Cohan, A. Longformer: The Long-Document Transformer. *arXiv* **2020**, arXiv:2004.05150.

38. Li, S.; Li, H.; Liu, Y. SMC structural health monitoring benchmark problem using monitored data from an actual cable-stayed bridge. *Struct Control Health Monit* **2013**, *21*, 156–173. [CrossRef]

39. Dyke, S.J.; Bernal, D.; Beck, J. Experimental phase II of the structural health monitoring benchmark problem. In Proceedings of the 16th ASCE Engineering Mechanics Conference, Seattle, MA, USA, 16–18 July 2003.