Temporal information embedding neural network for structural seismic response prediction

Chengbo Wang
Tianjin University
wangchb@tju.edu.cn

Abstract

Neural networks have been used to predict the nonlinear response of structures on seismic excitation due to their excellent fitting ability. However, with the complexity of neural network models and limited training data, it is crucial to efficiently train accurate models using limited ground motion data. This paper introduces a novel temporal information embedding neural network. It utilizes specified feature generators to perform data augmentation and feature information embedding. It enables the network to learn distinct feature spaces and facilitates mutual correction between feature data through feature information embedding. Moreover, the temporal information embedding model is not reliant on additional physical information, enhancing the generalizability of the information embedding neural network. This study employs independent repetitive tests based on Taylor's Formula of the temporal information embedding neural network to predict the response of structures. The results demonstrate that the temporal information embedding model exhibits excellent performance in prediction accuracy and efficiency.

1 Introduction

The dynamic response of structures is critically important in all phases of building design, maintenance, and emergency assessment, particularly in earthquake-prone regions. Accurate calculation of the dynamic response is essential for evaluating the safety and reliability of structures. As a result, calculating structural dynamic responses has been extensively studied in the field of structural engineering over the past few decades.

Researchers have investigated the linear and nonlinear response of structural dynamics using physical models. The finite element method (FEM) is the most accurate numerical simulation technique for analyzing dynamic time courses, [Mahendran, 2007; Reddy and Gartling, 2010]. However, the computational cost can increase significantly as the complexity and scale of structures increase. Fine-grained modeling of complex, nonlinear systems or large-scale buildings can result in an exponential increase in computational cost, whereas low-fidelity reduced-order models may compromise computational accuracy, [Lucia *et al.*, 2004; Caiazzo *et al.*, 2014]. Therefore, it is essential to maintain a balance between computational accuracy and efficiency when evaluating the performance of structures.

As artificial intelligence technology has advanced, Artificial Neural Networks (ANNs) have become capable of establishing non-linear relationships between input and output, disregarding the complex structure between them, [Chen and Billings, 1992; Molas and Yamazaki, 1995]. ANNs can approximate the real structure of complex non-linear systems through iterative optimization. However, traditional Backpropagation (BP) neural networks are inadequate for modeling complex non-linear systems due to their simplicity, [Ying *et al.*, 2009]. To overcome this limitation, researchers have introduced novel architectures for modeling structural dynamic responses using deep learning techniques. For example, [Wu and Jahanshahi, 2019] predicted the dynamic response of single-degree-of-freedom and multi-degree-of-freedom steel frames using Convolutional Neural Networks (CNNs) [Gu *et al.*, 2018] to extract local features and suppress noise components.

Recurrent Neural Networks (RNNs) incorporate memory units that enable them to selectively retain historical information, thereby effectively extracting relationships between sequence data contexts. [Zhang *et al.*, 2019] developed a stacked Long Short-Term Memory (LSTM) [Hochreiter and Schmidhuber, 1997] network and reduced the training time and validation loss by down-sampling. [Xu *et al.*, 2022] introduced a recursive mechanism based on LSTM to make full use of prediction data in the time dimension while controlling the accumulation of errors by controlling the sampling interval. [Liao *et al.*, 2023] achieved a more powerful fitting capability by combining the attention mechanism with the LSTM layer. It should be noted that the complexity of models increases with the amount of data required for training. Conversely, insufficient data can result in significant overfitting [Lever *et al.*, 2016; Ying, 2019].

[Raissi *et al.*, 2019] introduced physics-informed neural networks (PINNs), a novel neural network framework that integrates physics principles with neural networks to improve the accuracy of predictions with limited data. This approach conditions the neural network to satisfy the underlying physics principles, significantly improving model perfor-

mance. PINNs have seen rapid application in physical simulations, but embedding prior knowledge becomes challenging when the physical equation cannot be fully identified due to the system's complexity. The core of the PINNs optimization process involves identifying a mathematical equation that can constrain the relationship between physical quantities and adding the equation's residuals to the loss function during back-propagation.

This study proposes a novel approach to embed temporal information into neural networks using feature generation and feature information embedding directly from data, inspired by PINNs and data augmentation techniques. The primary contribution of this research is to enable neural networks to incorporate temporal feature information constraints when an explicit physical formulation is not available. The neural network with temporal information embedding offers the advantages of PINNs, such as the ability to effectively capture data features and generalize well with limited data.

This paper is organized as follows: Section 2 introduces the fundamental concepts of temporal information embedding. Section 3.1 provides details on the experimental scheme, including data processing and model configuration. Section 4 reports the implementation of temporal information embedding in multiple datasets to verify its impact on model performance. Section 5 presents the research findings on temporal information embedding.

2 Temporal information embedding

Physics-informed neural networks (PINNs) infer a continuous potential function $u(x,t)$ as the solution of a system of nonlinear partial differential equations (PDEs). In the back-propagation phase of the optimized neural network, the loss function not only calculates the residual loss of the neural network prediction, as in Equation 2. The equation $PDE = 0$ used to describe the physical law is also taken into account: the residual of the PDE is calculated as the physical loss L_{Physic}, as in Equation 3. The final loss function used to optimize the neural network is calculated by weighting the two loss terms, which is finally expressed as:

$$Loss = \omega_{Data} L_{Data} + \omega_{Physic} L_{Physic} \tag{1}$$

$$L_{Data} = MSE(u(x_i,t_i) - \hat{u}(x_i,t_i)) \tag{2}$$

$$L_{Physic} = MSE(PED) \tag{3}$$

Where $MSE(\cdot)$ denotes the mean squared error operation, ω_{Data} and ω_{Physic} are the weighting coefficients used to balance two loss terms, respectively. The selection of weighting coefficients directly affects the efficiency of PINNs [Wang *et al.*, 2021].

2.1 Temporal information embedding neural network

Temporal information-embedding neural networks are similar to PINNs. The information expressed as mathematical formulas is found in the time series to replace the physical formulas in PINNs, and the data relationship between the features of the data is added to the loss function of the neural network.

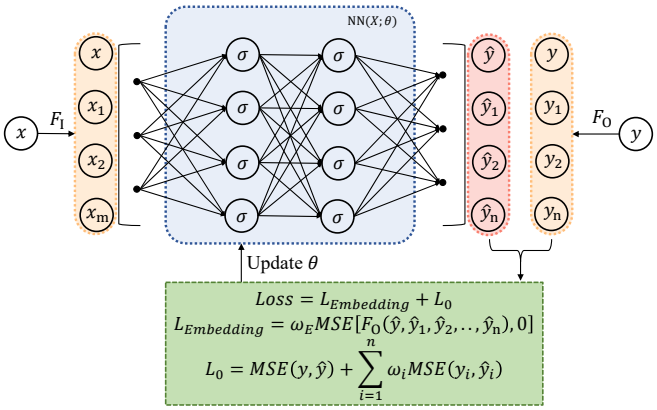


Figure 1: Temporal information embedding neural network

Figure 1 illustrates the fundamental concept of information embedding neural networks. $Y = NN(X;\theta)$ is a neural network with X and Y as the input and output with the network parameters θ. The features are generated from original input data x and y using specific feature generators F_I and F_O. Where $F_I(x, x_1, x_2, \ldots, x_m) = 0$, $F_O(y, y_1, y_2, \ldots, y_n) = 0$. The embedding loss $L_{Embedding}$ is calculated according to the feature generator F_O, and it is weighted with the data residual L_0 to obtain the loss function $Loss$ used to update the parameters θ.

2.2 Taylor's Formula based information embedding neural network

Temporal information embedding for computing structural responses

The dataset of the structural response model consists of the ground motion time history X and the response time history Y for the multi-degree-of-freedom model:

$$X = \{x_1, x_2, \cdots, x_t, \cdots, x_T\}^{\mathrm{T}}, X \in \mathbb{R}^{T\times 1} \tag{4}$$

$$x_t = \{x_{t1}\} \tag{5}$$

$$Y = \{y_1, y_2, \cdots, y_t, \cdots, y_T\}^{\mathrm{T}}, Y \in \mathbb{R}^{T\times D} \tag{6}$$

$$y_t = \{y_{t1}, y_{t2}, \cdots, y_{td}, \cdots, y_{tD}\} \tag{7}$$

Where T denotes the time step of the ground motion time history and D denotes the number of degrees of freedom of the structural response.

At any moment t, the structural response of any dth degree of freedom is $y_{td} = Y(t,d)(d = 1, 2, \cdots, D)$, which is given by the Taylor expansion:

$$\begin{aligned} Y(t,d) =& Y(t_0,d) + \frac{Y^{(1)}(t_0,d)}{1!}\Delta t + \frac{Y^{(2)}(t_0,d)}{2!}\Delta t^2 \\ &+ \cdots + O(\Delta t^n) \end{aligned} \tag{8}$$

Where $\Delta t = t - t_0$. The paper takes $\Delta t = 0.1$, at which point $Y^{(i)}(t_0,d)$ is $(\frac{t-t_0}{0.1})^i$ times the approximate estimate of the ith order derivative of the structural response of the dth degree of freedom at the moment t_0.

Figure 2 illustrates the detailed process of predicting the structural response using the Taylor expansion-based information embedding neural network. The feature generators F_I

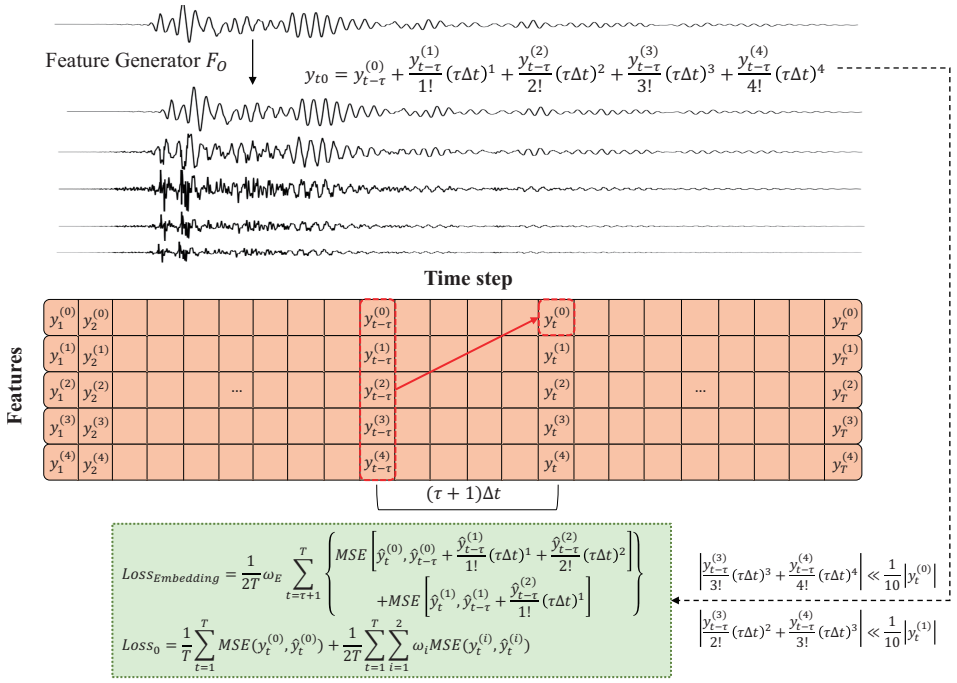


Figure 2: Taylor's Formula based information embedding neural network

and F_O based on the Taylor expansion are used to calculate the first 5 order derivatives of all sequences at $\Delta t = 0.1$. It should be noted that the 0th order derivative should be equal to the values of the original data. As in Figure 2, $L_{Embedding}$ embeds Taylor's Formula by calculating the bias between the data at the $(t-\tau)$ time step and the data at the tth time step. Considering heavy calculations of the loss function and the decreasing value of the higher order derivatives in the Taylor expansion, the first 3 order derivatives of output data are extracted as the embedding loss.

The datasets are stacked according to the LSTM-s model [Zhang *et al.*, 2019], and the stacking time step is set to 10 corresponding to which τ is set to 10. The transformed data are:

$$\begin{aligned} Y_{train} &= \{y_1', y_2', \cdots, y_i', y_{i+1}', \cdots\}^{\mathrm{T}} \\ Y_{train} &\in \mathbb{R}^{\lfloor \frac{T}{10} \rfloor \times 3D} \end{aligned} \tag{9}$$

$$y_i' = \left\{ \begin{array}{c} y_{10i,1}^{(0)}, y_{10i,1}^{(1)}, y_{10i,1}^{(2)}, \\ y_{10i,2}^{(0)}, y_{10i,2}^{(1)}, y_{10i,2}^{(2)}, \\ \cdots, \\ y_{10i,D}^{(0)}, y_{10i,D}^{(1)}, y_{10i,D}^{(2)} \end{array} \right\} \tag{10}$$

Calculate the loss function used to optimize the model according to Figure 2:

$$Loss = L_0 + \omega_{Embedding} L_{Embedding} \tag{11}$$

$$L_0 = \frac{1}{3D\lfloor \frac{T}{10} \rfloor} \sum_{t=1}^{\lfloor \frac{T}{10} \rfloor} \sum_{d=1}^{D} \sum_{i=0}^{2} \omega_i (\hat{y}_{10td}^{(i)} - y_{10td}^{(i)})^2 \tag{12}$$

$$\begin{aligned} &L_{Embedding} = \\ &\frac{1}{2D\lfloor \frac{T}{10} \rfloor} \sum_{t=1}^{\lfloor \frac{T}{10} \rfloor} \sum_{d=1}^{D} \left\{ \left[\hat{y}_{10td}^{(0)\prime} - \hat{y}_{10td}^{(0)}\right]^2 + \left[\hat{y}_{10td}^{(1)\prime} - \hat{y}_{10td}^{(1)}\right]^2 \right\} \\ &\hat{y}_{10td}^{(0)\prime} = \hat{y}_{(10t-\tau)d}^{(0)} + \frac{\hat{y}_{(10t-\tau)d}^{(1)}}{1!}(\tau\Delta t)^1 + \frac{\hat{y}_{(10t-\tau)d}^{(2)}}{2!}(\tau\Delta t)^2 \\ &\hat{y}_{10td}^{(1)\prime} = \hat{y}_{(10t-\tau)d}^{(1)} + \frac{\hat{y}_{(10t-\tau)d}^{(2)}}{1!}(\tau\Delta t)^1 \end{aligned} \tag{13}$$

Where ω_i and $\omega_{Embedding}$ are the weights used to balance the individual loss terms, which can be artificially selected constants or automatically adjusted with other algorithms[Karniadakis *et al.*, 2021; Wang *et al.*, 2022].In this paper, we calculate the losses using constant weights.

Interpretation of temporal information embedding

In terms of the dynamical equilibrium equation, when a multi-degree-of-freedom system is subjected to an external excitation vector $p(t)$, the dynamical equilibrium equation for the multi-degree-of-freedom system can be formulated as:

$$f_S + f_D + f_I = p(t) \tag{14}$$

Where the elastic force vector f_S, the damping force vector f_D, and the inertia force vector f_I are all forces generated by the motion of the object and can be calculated:

$$\begin{aligned} f_S &= kx \\ f_D &= c\dot{x} \\ f_I &= m\ddot{x} \end{aligned} \tag{15}$$

Where k is the structural stiffness matrix, c is the damping matrix, m is the mass matrix, and all three are the inherent parameters of the structure, determined by the structure.

Dataset	Time step	Sampling interval	Number of samples	Data distribution		
				Training	Validation	Test
Bouc-Wen	10001	0.005 s	100	25	25	50
MRFDBF	10001	0.005 s	548	50	50	100

Table 1: Dataset distribution

	Basic Model	Multi-Feature Model	Temporal embedding model
Data	Original Data	Multi-Feature Data	Multi-Feature Data
Input Shape	[100, 10001, 1]	[100, 10001, 1, 5]	[100, 10001, 1, 5]
Output Shape	[100, 10001, 1]	[100, 10001, 1, 3]	[100, 10001, 1, 3]
Loss Function	L_0	L_0	$L_0 + L_{Embedding}$

Table 2: Differences among the models

Parameter	Value
Number of LSTM layers	3
Number of neurons	100
Activation	ReLU
Batch size	8
Optimizer	Adam
Initial learning rate	0.001
Epochs	10000
Platform	Pytorch 1.13.0
GPU	NVIDIA GeForce RTX 3060

Table 3: Model configuration

The close relationship between the inherent parameters of all degrees of freedom suggests that a neural network model capable of accurately calculating these three responses can serve as a good proxy model for predicting the structural response.

3 Experimental scheme

3.1 Data processing

The datasets are obtained from the public data of [Zhang *et al.*, 2019], namely Bouc-Wen and MRFDBF. Relevant information on the structure will be presented in Section 4.

Zero Padding is implemented before each data sample in the seismic excitation-response dataset, and the padding time steps are all set to 1000. The Zero Padding contributes to the structural response state before/without seismic excitation. LSTM-s [Zhang *et al.*, 2019] is adopted to accelerate the operation and improve the performance of the benchmark model, and the stacking time step is set to 10.

To effectively test the generalization ability of the model and eliminate stochastic errors introduced by data set partitioning and parameter initialization, we conduct multiple independent trials for each dataset. In each trial, we sample the training data, validation data, and test data within a specific dataset using a simple random sampling method, ensuring that the training and validation data do not leak information to the test set. Table 1 presents the distribution of the dataset. By conducting multiple randomized experiments, we minimize the impact of training data sampling and parameter initialization on the comparison model and demonstrate the stability performance of the temporal information embedding model compared to the baseline model.

3.2 Model configuration

To compare the effect of information embedding on neural network performance, we establish a benchmark model and compare its performance with that of the temporal information embedding model.

Table 3 shows the basic structure of the benchmark model and the temporal information embedding model, with the ReLU activation function [Agarap, 2018] added after the LSTM layer. Each independent experiment trains both models for 10,000 epochs. The learning rate decays to 0.97 of the previous value every 100 epochs.

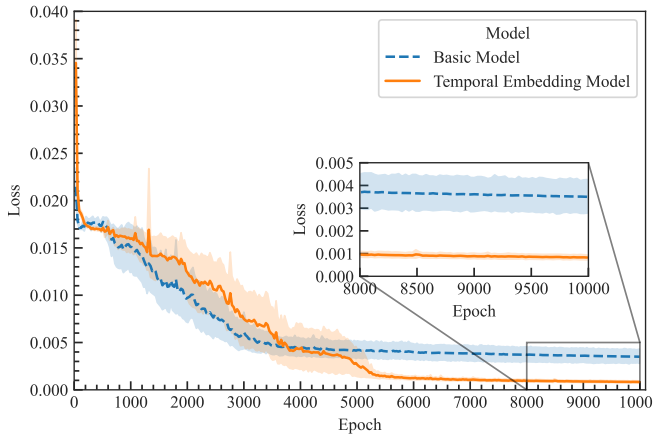
The loss function of the benchmark model is set as the MSE residual between the true and predicted values, using Equation 12. The loss function for the temporal information embedding model is calculated by weighting the MSE loss and the embedding loss with Equation 11.

4 Experimental validation

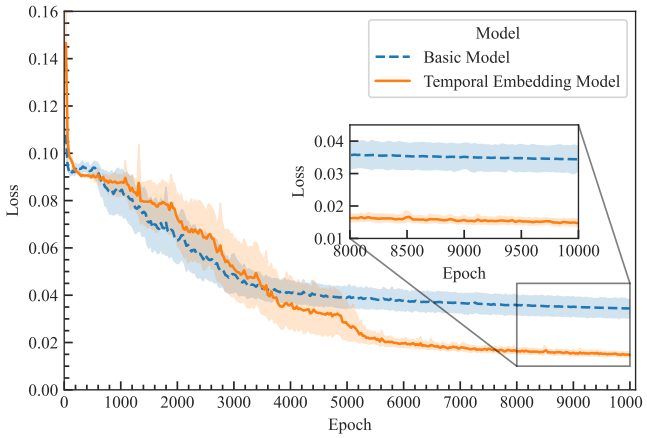
4.1 1 DOF Bouc-Wen nonlinear hysteretic system

The Bouc-Wen dataset used in this study is obtained from [Zhang *et al.*, 2019], which uses a LSTM model to calculate the seismic response of a five-degree-of-freedom nonlinear hysteresis system based on the Bouc-Wen numerical model [Sun *et al.*, 2013]. The dataset consists of ground acceleration data replaced by random limited-band white noise sequences of different amplitudes to excite the numerical model, with inter-story displacements used as output data. The LSTM-s model's prediction accuracy for DOF 1 data is found to be the lowest. Therefore, the DOF 1 data is extracted to create a new dataset with an output data size of [100, 10001, 1].

Data processing, including zero-padding, stacking, random data set partitioning, and normalization, is performed on the original data before inputting it into the models. In the case of the temporal information embedding model, multi-order



(a) Test MSE Loss



(b) Test MAE Loss

Figure 3: Test MSE loss and MAE loss of 1 DOF Bouc-Wen

Model	Test loss		Time Spent	Pearson correlation coefficient
	MSE	MAE		
Basic Model	3.51E-3	3.45E-2	12.65 min	0.936
Temporal embedding Model	8.33E-4	1.49E-2	13.31 min	0.992

Table 4: Model performances with 1 DOF BoucWen

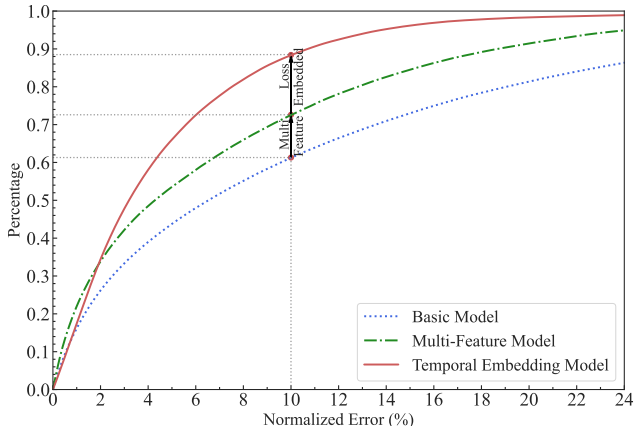


Figure 4: Cumulative distribution function of Bouc-Wen

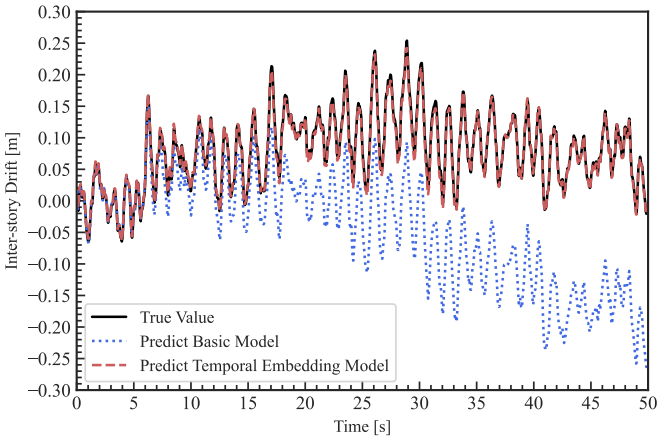


Figure 5: A seismic response comparison of BoucWen

derivative estimators were required to generate features for the original data and calculate. The $0 \sim 4$th order derivatives of the input and output data were calculated to obtain data sizes [100, 10001, 1, 5], [100, 10001, 1, 5].

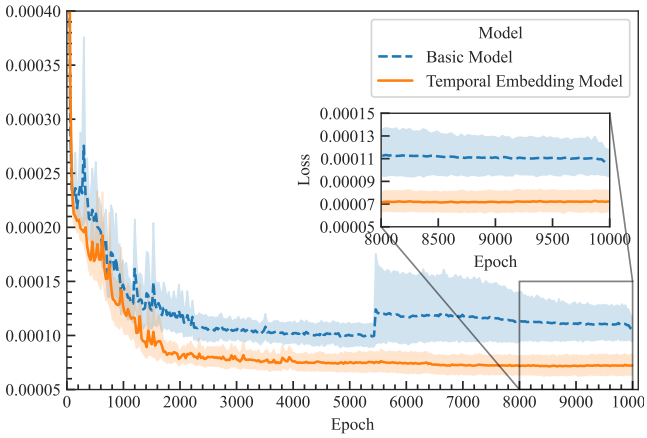
The performance indicators of the benchmark model and the temporal information embedding model were calculated based on 20 independent experiments, as presented in Table 4. The loss-epoch error band plots at the 95% confidence level for both models are illustrated in Figure 3. It is calculated that the MSE test error is improved by 76.27% and the MAE test error is improved by 56.78% after 10,000 epochs.

For more detailed quantification of the improvement of the benchmark model by multi-order derivative data augmentation and information embedding, the multi-feature model without information embedding is added, and the detailed network differences are shown in Table 2.

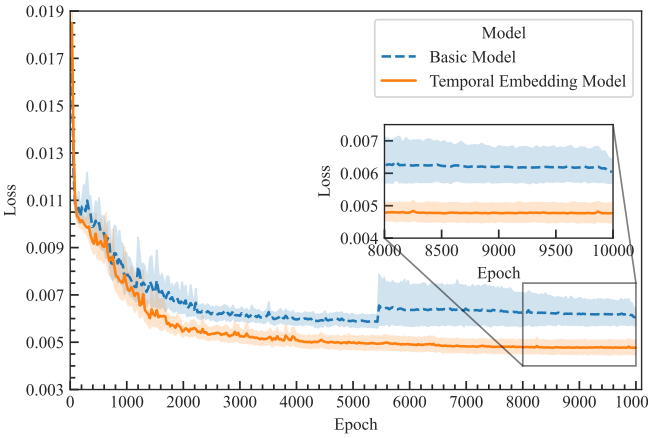
The cumulative distribution functions(CDF) of the test errors for the benchmark model, the multi-feature model, and the temporal information embedding model for 20 independent experiments are statistically calculated, respectively. The normalized error distribution of each group of responses is counted for the purpose of reducing the effect of response amplitude on the error:

$$C_i = CDF\left\{\frac{|\hat{y}_i - y_i|}{max(|y_i|)}\right\} \tag{16}$$

Where y_i denotes the true value of the ith response signal, $\hat{y}_i$ denotes the predicted value, C_i denotes the CDF.



(a) Test MSE Loss



(b) Test MAE Loss

Figure 6: Test MSE loss and MAE loss of 3 DOF MRFDBF

Model	Test loss		Time Spent	Pearson correlation coefficient
	MSE	MAE		
Basic Model	1.07E-4	6.12E-3	17.12 min	0.947
Temporal embedding Model	6.92E-5	4.74E-3	17.42 min	0.965

Table 5: Model performances with 3DOF MRFDBF

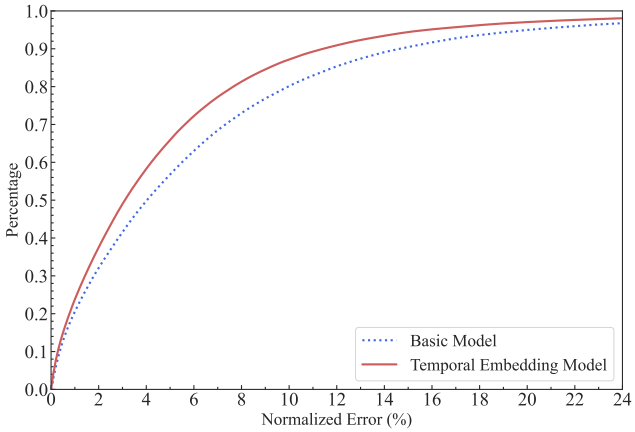


Figure 7: Cumulative distribution function of MRFDBF

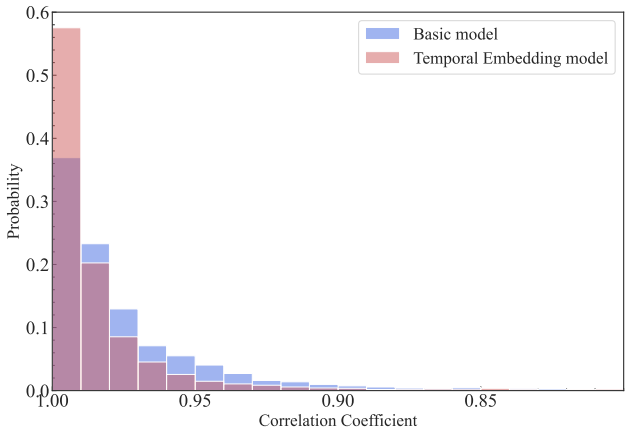


Figure 8: Pearson correlation coefficient of MRFDBF

Figure 4 presents that the multi-feature model improves 11.3% over the benchmark model within 10% of the cumulative distribution of normalization error, while the temporal information embedding model improves 27.3%.

4.2 3 DOF steel frame

Figure 5 illustrates the comparison between the predicted values of the benchmark model and the temporal information embedding model.

The dataset MRFDBF is derived from the paper and the prototype building is a 3-story damped supported steel frame. It consists of ground acceleration as the input data and inter-story displacement as the output data.The number of training samples, validation samples and test samples is 50-50-100.

Table 5 shows the statistical results of 20 independent replicate experiments.

The MSE test error is reduced by 35.3% and the MAE test error is reduced by 22.5%. The results reveal that the temporal information embedding neural network exhibits more efficient optimization and stable prediction accuracy, despite the random selection of training data. The significant enhancement of the models can be observed in the CDF, shown in Figure 7.

In Figure 8, the probability of the correlation coefficient in the interval of 1 to 0.99 is greatly improved, and the probability of the correlation coefficient of the predicted sample

below 0.9 is greatly reduced.

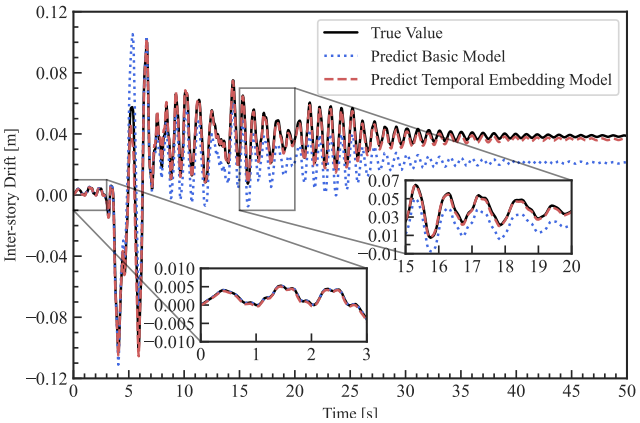


Figure 9: A seismic response comparison of MRFDBF

Figure 9 illustrates the improvement of the temporal information embedding model compared with the benchmark model. Within $0 \sim 3$s, both models maintain an extremely high-quality fitness. In most of the other periods, such as $15 \sim 20$s, the temporal information embedding model corrects the error substantially and reduces the error between the predicted and true responses.

5 Conclusions

In this paper, a new neural network model for information embedding is proposed. The temporal information embedding model based on Taylor's Formula is implemented based on the concept of information embedding. The paper validates the structural nonlinear response under ground motion excitation.

The experiments were conducted on two datasets to analyze and compare the performance between the temporal information embedding neural network and the benchmark model. The experimental results show that the embedding of temporal information improves model performance significantly, reduces the prediction probability in a large error range, and enhances the generalization ability of the benchmark model. The reduction of the error band bandwidth of the test set indicates that the embedding of information between features strengthens the model's ability to extract features from the dataset.

Information embedding is a neural optimization method. The temporal information embedding model based on Taylor's Formula can be used not only in structural dynamic response prediction but also be applied in other time series prediction tasks in different domains when suitable feature generators are selected.

References

[Agarap, 2018] Abien Fred Agarap. Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375*, 2018.

[Caiazzo *et al.*, 2014] Alfonso Caiazzo, Traian Iliescu, Volker John, and Swetlana Schyschlowa. A numerical investigation of velocity–pressure reduced order models for incompressible flows. *Journal of Computational Physics*, 259:598–616, 2014.

[Chen and Billings, 1992] SABS Chen and Stephen A Billings. Neural networks for nonlinear dynamic system modelling and identification. *International journal of control*, 56(2):319–346, 1992.

[Gu *et al.*, 2018] Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Gang Wang, Jianfei Cai, et al. Recent advances in convolutional neural networks. *Pattern recognition*, 77:354–377, 2018.

[Hochreiter and Schmidhuber, 1997] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[Karniadakis *et al.*, 2021] George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

[Lever *et al.*, 2016] Jake Lever, Martin Krzywinski, and Naomi Altman. Points of significance: model selection and overfitting. *Nature methods*, 13(9):703–705, 2016.

[Liao *et al.*, 2023] Yuchen Liao, Rong Lin, Ruiyang Zhang, and Gang Wu. Attention-based lstm (attlstm) neural network for seismic response modeling of bridges. *Computers & Structures*, 275:106915, 2023.

[Lucia *et al.*, 2004] David J Lucia, Philip S Beran, and Walter A Silva. Reduced-order modeling: new approaches for computational physics. *Progress in aerospace sciences*, 40(1-2):51–117, 2004.

[Mahendran, 2007] Mahen Mahendran. Applications of finite element analysis in structural engineering. In *Proceedings of the International Conference on Computer Aided Engineering*, pages 38–46. Indian Institute of Technology Madras, 2007.

[Molas and Yamazaki, 1995] Gilbert L Molas and Fumio Yamazaki. Neural networks for quick earthquake damage estimation. *Earthquake engineering & structural dynamics*, 24(4):505–516, 1995.

[Raissi *et al.*, 2019] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

[Reddy and Gartling, 2010] Junuthula Narasimha Reddy and David K Gartling. *The finite element method in heat transfer and fluid dynamics*. CRC press, 2010.

[Sun *et al.*, 2013] Hao Sun, Hilmi Luş, and Raimondo Betti. Identification of structural models using a modified artificial bee colony algorithm. *Computers & Structures*, 116:59–74, 2013.

[Wang *et al.*, 2021] Sifan Wang, Yujun Teng, and Paris Perdikaris. Understanding and mitigating gradient flow

pathologies in physics-informed neural networks. *SIAM Journal on Scientific Computing*, 43(5):A3055–A3081, 2021.

[Wang *et al.*, 2022] Sifan Wang, Xinling Yu, and Paris Perdikaris. When and why pinns fail to train: A neural tangent kernel perspective. *Journal of Computational Physics*, 449:110768, 2022.

[Wu and Jahanshahi, 2019] Rih-Teng Wu and Mohammad R Jahanshahi. Deep convolutional neural network for structural dynamic response estimation and system identification. *Journal of Engineering Mechanics*, 145(1):04018125, 2019.

[Xu *et al.*, 2022] Zekun Xu, Jun Chen, Jiaxu Shen, and Mengjie Xiang. Recursive long short-term memory network for predicting nonlinear structural seismic response. *Engineering Structures*, 250:113406, 2022.

[Ying *et al.*, 2009] Wang Ying, Wang Chong, Li Hui, and Zhao Renda. Artificial neural network prediction for seismic response of bridge structure. In *2009 International conference on artificial intelligence and computational intelligence*, volume 2, pages 503–506. IEEE, 2009.

[Ying, 2019] Xue Ying. An overview of overfitting and its solutions. In *Journal of physics: Conference series*, volume 1168, page 022022. IOP Publishing, 2019.

[Zhang *et al.*, 2019] Ruiyang Zhang, Zhao Chen, Su Chen, Jingwei Zheng, Oral Büyüköztürk, and Hao Sun. Deep long short-term memory networks for nonlinear structural seismic response prediction. *Computers & Structures*, 220:55–68, 2019.