Frequency domain deep learning

Preprint · May 2024

DOI: 10.13140/RG.2.2.27531.30249/1

1 author:

Gary Nan Tie
Mu Risk LLC

Multivariate time series learning via frequency domain tensors

Gary Nan Tie, May 26, 2024

Abstract

A 2-Dimensional Fourier Transform takes a multivariate time series in the time domain into a (complex) matrix in the frequency domain, that captures auto and cross correlation information. This is followed by a tensor product dimension enlargement, akin to a word embedding. For the purposes of forecasting, this enlarged matrix is then input to a multilayer perceptron to learn in the frequency domain, before returning to the time domain.

Let $[x_1, x_2, \ldots, x_T] \in \mathbb{R}^{N \times T}$ be a regularly sampled multivariate time series with N channels and T timestamps, $x_t \in \mathbb{R}^N$.

Let $X_t = [x_{t-L+1}, x_{t-L+2}, \ldots, x_t] \in \mathbb{R}^{N \times L}$ be a lookback window of length L at timestamp t

and $Y_t = [x_{t+1}, x_{t+2}, \ldots, x_{t+\tau}] \in \mathbb{R}^{N \times \tau}$ be a horizon window of length τ at timestamp t

The multivariate time series forecasting problem is to use historical observations X_t to predict values $\hat{Y}_t$ via a MLP $\mathcal{Y}_\theta$ parameterized by θ so that $\hat{Y}_t = \mathcal{Y}_\theta(X_t)$.

Notation:

$\mathcal{F}$ 2D Discrete Fourier Transform

$X_t \in \mathbb{R}^{N \times L}$ in the time domain

$\mathcal{X} \triangleq \mathcal{F}(X_t) \in \mathbb{C}^{N \times L}$ in the frequency domain

I identity matrix

$\mathbb{1}$ matrix of 1's

Frequency domain MLPs $\mathcal{Y}$

Let $\mathcal{W}$ denote a learnable weight matrix and

$\mathcal{B}$ denote a learnable bias matrix.

$$\mathcal{Y}^0 \triangleq \underset{d \times d}{\mathcal{W}^0} \otimes \underset{N \times L}{\mathcal{H}} \in \mathbb{C}^{dN \times dL}$$

$$\mathcal{Y}^l \triangleq \sigma\left[\left(\underset{d \times d}{\mathcal{W}^l} \otimes \underset{N \times N}{I} \right) \mathcal{Y}^{l-1} + \underset{d \times d}{\mathcal{B}^l} \otimes \underset{N \times L}{\mathbb{1}} \right] \in \mathbb{C}^{dN \times dL}$$

- a $d \times d$ partitioned matrix with $N \times L$ cells in $\mathbb{C}$.

$\mathcal{Y}^L \in \mathbb{C}^{dN \times dL}$ is the final output, l denotes the l-th layer,

and σ is the activation function applied elementwise.

Let $\mathcal{Y}_t \triangleq \mathcal{F}^{-1}(\mathcal{Y}^L) \in \mathbb{R}^{dN \times dL}$ in time domain,

where $\mathcal{F}^{-1}$ is applied cellwise to $\mathcal{Y}^L$.

FFN projection $\hat{Y}_t$

$$\underset{N\times T}{\hat{Y}_t} \triangleq \left[\sigma\left(\underset{N\times dN}{\varphi_L}\, \mathcal{Y}_t\, \underset{dL\times d_h}{\varphi_R} + \underset{N\times 1}{\mathbb{1}} \otimes \underset{1\times d_h}{b}\right)\right] \underset{d_h\times T}{\chi} + \underset{N\times 1}{\mathbb{1}} \otimes \underset{1\times T}{c} \in \mathbb{R}^{N\times T}$$

where $\mathcal{Y}_t \in \mathbb{R}^{dN\times dL}$ is the output of the frequency learner,

σ is the activation function applied elementwise,

$\varphi_L \in \mathbb{R}^{N\times dN}$, $\varphi_R \in \mathbb{R}^{dL\times d_h}$, $\chi \in \mathbb{R}^{d_h\times T}$ are the weights,

$b \in \mathbb{R}^{1\times d_h}$, $c \in \mathbb{R}^{1\times T}$ are the biases, and

d_h is the inner layer dimension size.

Learning schematic:

$$X_t \xrightarrow{\mathcal{F}} \mathcal{X} \longrightarrow \mathcal{W}^{\circ} \otimes \mathcal{X} \xrightarrow[]{\overset{\text{MLP}}{\mathcal{G}}} \mathcal{Y}^{l} \xrightarrow{\mathcal{F}^{-1}} \mathcal{Y}_t \xrightarrow{\text{FFN}} \hat{Y}_t$$

time domain   Frequency domain

Summary

The 2D Fourier transform captures the inter and intra series dependencies in the frequency domain. The MLP first enlarges the dimension of the spectral input, akin to word embedding, then learns multivariate time series patterns globally with energy compaction. Finally a feed forward network uses the learned channel and temporal dependencies to make predictions for future time steps. Neural networks make nonlinear multivariate time series forecasting possible.

References

'Frequency-domain MLPs are More Effective Learners in Time Series Forecasting'
Kun Yi, Qi Zhang, Wei Fan, Shoujin Wang, Pengyang Wang, Hui He, Defu Lian, Ning An, Longbing Cao, Zhendong Niu
37th Conference on Neural Information Processing Systems (NeurIPS 2023)
arXiv:2311.06184v1 [cs.LG] 10 Nov 2023

'Constructing the Matrix Multilayer Perceptron and its Application to the VAE'
Jalil Taghia, Maria Bankestad, Fredrik Lindsten, Thomas B. Schon
arXiv:1902.01182v1 [stat.ML] 4 Feb 2019