Fast Temporal Wavelet Graph Neural Networks

Duc Thien Nguyen*
Tokyo Institute of Technology, Japan

Manh Duc Tuan Nguyen*
Toyo University, Japan

Truong Son Hy*
Indiana State University, USA

Risi Kondor
University of Chicago, USA

* Equal contribution. Correspondence to TruongSon.Hy@indstate.edu

December 2023

Summary

We propose:

- An **time & memory efficient** temporal graph neural network model for spatio-temporal forecasting on multivariate timeseries.
- Leverage **Multiresolution Matrix Factorization** (MMF) & **Wavelet theory**.
- Extensive experiments with **traffic & brain signals forecasting**, show competitive performance.

Prior arts

Traditional methods:

- Historical Average
- ARIMA with Kalman filter
- Vector Auto-regressive VAR
- Linear Support Vector Regression SVR

Prior arts

Traditional methods:

- Historical Average
- ARIMA with Kalman filter
- Vector Auto-regressive VAR
- Linear Support Vector Regression SVR

Deep Learning:

- Feed-forward neural network FNN
- Fully-connected LSTM
- Spatio-Temporal Graph Convolutional Networks (STGCN)
- GWaveNet
- Diffusion Convolutional RNN (DCRNN)

Wavelet bases replace Fourier bases

Limitations of GFT (Bruna et al., 2014)

- High computational cost:
    - EVD of the graph Laplacian has complexity $O(n^3)$
    - GFT involves multiplying with a **dense** matrix of eigenvectors
- The graph convolution is **not localized** in the vertex domain, even if the graph itself has well defined local communities.

Wavelet bases replace Fourier bases

Limitations of GFT (Bruna et al., 2014)

- High computational cost:
  - EVD of the graph Laplacian has complexity $O(n^3)$
  - GFT involves multiplying with a **dense** matrix of eigenvectors
- The graph convolution is **not localized** in the vertex domain, even if the graph itself has well defined local communities.

Is there an alternative to Fourier transform/basis?

Wavelet bases replace Fourier bases

Limitations of GFT (Bruna et al., 2014)

- High computational cost:
  - EVD of the graph Laplacian has complexity $O(n^3)$
  - GFT involves multiplying with a **dense** matrix of eigenvectors
- The graph convolution is **not localized** in the vertex domain, even if the graph itself has well defined local communities.

Is there an alternative to Fourier transform/basis?

MMF of a graph Laplacian $\tilde{\boldsymbol{L}}$ (Kondor et al., 2014)

$$\tilde{\boldsymbol{L}} = \boldsymbol{U}_1^T \boldsymbol{U}_2^T \dots \boldsymbol{U}_L^T \boldsymbol{H} \boldsymbol{U}_L \dots \boldsymbol{U}_2 \boldsymbol{U}_1$$

MMF gives us a total of N wavelets:

- L mother wavelets $\overline{\psi} = \{\psi^1, .., \psi^L\}$,
- $N - L$ father wavelets $\overline{\phi} = \{\phi_m^L = \boldsymbol{H}_{m,:}\}_{m \in \mathbb{S}_L}$.

Wavelet Neural Networks (WNN)

Similar to GFT-based convolution, each layer $k = 1, .., K$ transforms a $|V| \times F_{k-1}$ input $\boldsymbol{f}^{(k-1)}$ into a $|V| \times F_k$ output $\boldsymbol{f}^{(k)}$ as

$$\boldsymbol{f}^{(k)}_{:,j} = \sigma\left(\boldsymbol{W} \sum_{i=1}^{F_{k-1}} \boldsymbol{g}^{(k)}_{i,j} \boldsymbol{W}^T \boldsymbol{f}^{(k-1)}_{:,i}\right) \quad \text{for} \quad j = 1, \dots, F_k,$$

where $\boldsymbol{W} = [\overline{\phi}, \overline{\psi}]$ is wavelet basis matrix, $\boldsymbol{g}^{(k)}_{i,j}$ is a parameter/filter in the form of a diagonal matrix, and σ is an element-wise nonlinearity.

Fast Wavelet Transform

Sparse wavelet basis → Sparse matrix multiplication for wavelet transform

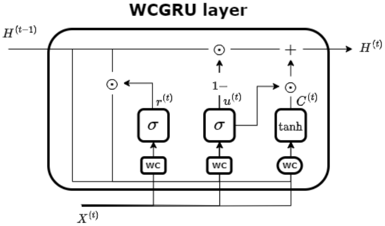
Fast Temporal Wavelet Graph Neural Networks

Spatial Dependency by diffusion process on an undirected graph $G = (\boldsymbol{X}, \boldsymbol{A})$ $\frac{d\boldsymbol{X}(t)}{dt} = (\tilde{\boldsymbol{A}} - \boldsymbol{I})\boldsymbol{X}(t)$

Fast Temporal Wavelet Graph Neural Networks

Spatial Dependency by diffusion process on an undirected graph $G = (\boldsymbol{X}, \boldsymbol{A})$ $\quad \frac{d\boldsymbol{X}(t)}{dt} = (\tilde{\boldsymbol{A}} - \boldsymbol{I})\boldsymbol{X}(t)$

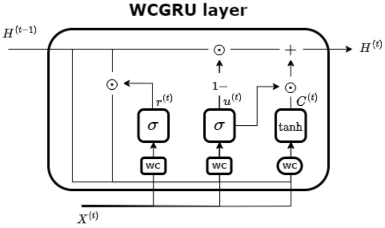
Temporal Dependency



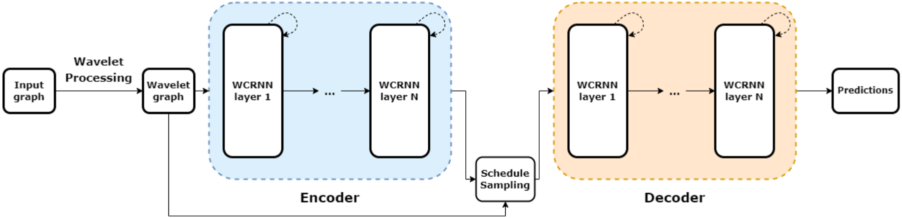
Fast Temporal Wavelet Graph Neural Networks

Spatial Dependency by diffusion process on an undirected graph $G = (\boldsymbol{X}, \boldsymbol{A})$ $\frac{d\boldsymbol{X}(t)}{dt} = (\tilde{\boldsymbol{A}} - \boldsymbol{I})\boldsymbol{X}(t)$

Temporal Dependency



FTWGNN uses sparse wavelet bases instead of Fourier bases



Experiments (1)

Traffic network

Datasets:

- METR-LA
- PEMS-BAY

Adjacency matrix $\boldsymbol{A}$ depends on physical distances between sensors.

Experiments (1)

Traffic network

Datasets:

- METR-LA
- PEMS-BAY

Adjacency matrix $\boldsymbol{A}$ depends on physical distances between sensors.

Brain network

Dataset:

- AJILE12

Adjacency matrix $\boldsymbol{A}$ depends on correlation between sensors.

Experiments (2)

Dataset	T	Metric	HA	ARIMA$_{kal}$	VAR	SVR	FNN	FC-LSTM	STGCN	GWaveNet	DCRNN	FTWGNN
METR-LA	15 min	MAE	4.16	3.99	4.42	3.99	3.99	3.44	2.88	**2.69**	2.77	2.70
		RMSE	7.80	8.21	7.89	8.45	7.94	6.30	5.74	**5.15**	5.38	**5.15**
		MAPE	13.0%	9.6%	10.2%	9.3%	9.9%	9.6%	7.6%	6.9%	7.3%	**6.8%**
	30 min	MAE	4.16	5.15	5.41	5.05	4.23	3.77	3.47	3.07	3.15	**3.02**
		RMSE	7.80	10.45	9.13	10.87	8.17	7.23	7.24	6.22	6.45	**5.95**
		MAPE	13.0%	12.7%	12.7%	12.1%	12.9%	10.9%	9.6%	8.4%	8.8%	**8.0%**
	60 min	MAE	4.16	6.90	6.52	6.72	4.49	4.37	4.59	3.53	3.60	**3.42**
		RMSE	7.80	13.23	10.11	13.76	8.69	8.69	9.40	7.37	7.59	**6.92**
		MAPE	13.0%	17.4%	15.8%	16.7%	14.0%	13.2%	12.7%	10.0%	10.5%	**9.8%**
PEMS-BAY	15 min	MAE	2.88	1.62	1.74	1.85	2.20	2.05	1.36	1.3	1.38	**1.14**
		RMSE	5.59	3.30	3.16	3.59	4.42	4.19	2.96	2.74	2.95	**2.40**
		MAPE	6.8%	3.5%	3.6%	3.8%	5.2%	4.8%	2.9%	2.7%	2.9%	**2.3%**
	30 min	MAE	2.88	2.33	2.32	2.48	2.30	2.20	1.81	1.63	1.74	**1.50**
		RMSE	5.59	4.76	4.25	5.18	4.63	4.55	4.27	3.70	3.97	**3.27**
		MAPE	6.8%	5.4%	5.0%	5.5%	5.43%	5.2%	4.2%	3.7%	3.9%	**3.2%**
	60 min	MAE	2.88	3.38	2.93	3.28	2.46	2.37	2.49	1.95	2.07	**1.79**
		RMSE	5.59	6.5	5.44	7.08	4.98	4.96	5.69	4.52	4.74	**3.99**
		MAPE	6.8%	8.3%	6.5%	8.0%	5.89%	5.7%	5.8%	4.6%	4.9%	**4.1%**

Dataset	T	Metric	HA	VAR	LR	SVR	LSTM	DCRNN	FTWGNN
AJILE12	1 sec	MAE	0.88	0.16	0.27	0.27	0.07	0.05	**0.03**
		RMSE	1.23	0.25	0.37	0.41	**0.09**	0.45	0.35
		MAPE	320%	58%	136%	140%	38%	7.84%	**5.27%**
	5 sec	MAE	0.88	0.66	0.69	0.69	0.39	0.16	**0.11**
		RMSE	1.23	0.96	0.92	0.93	0.52	0.24	**0.15**
		MAPE	320%	221%	376%	339%	147%	64%	**57%**
	15 sec	MAE	0.88	0.82	0.86	0.86	0.87	0.78	**0.70**
		RMSE	1.23	1.15	1.13	1.13	1.14	1.01	**0.93**
		MAPE	320%	320%	448%	479%	330%	294%	**254%**

FTWGNN outperforms other baselines by roughly **10%**

Experiments (3)

Dataset	T	DCRNN	FTWGNN	Speedup
METR-LA	15 min	350s	**217s**	1.61x
	30 min	620s	**163s**	3.80x
	60 min	1800s	**136s**	13.23x
PEMS-BAY	15 min	427s	**150s**	2.84x
	30 min	900s	**173s**	5.20x
	60 min	1800s	**304s**	5.92x
AJILE12	1 sec	80s	**35s**	2.28x
	5 sec	180s	**80s**	2.25x
	15 sec	350s	**160s**	2.18x

FTWGNN's training time is faster than DCRNN's by **5 times** on average.

Dataset	**Fourier basis**	**Wavelet basis**
METR-LA	99.04%	**1.11%**
PEMS-BAY	96.35%	**0.63%**
AJILE12	100%	**1.81%**

FTWGNN provides a **remarkable compression** by wavelet bases.

Conclusion

In summary:

- An **time & memory efficient** temporal graph neural network model for spatio-temporal forecasting on multivariate timeseries.
- Leverage **Multiresolution Matrix Factorization** (MMF) & **Wavelet theory**.
- Extensive experiments with **traffic & brain signals forecasting**, show competitive performance.

Software:

`https://github.com/HySonLab/TWGNN`

Thank you very much for your attention!