Graph Neural Flows for Unveiling Systemic Interactions Among Irregularly Sampled Time Series

The 38th Annual Conference on Neural Information Processing Systems
(NeurIPS 2024)

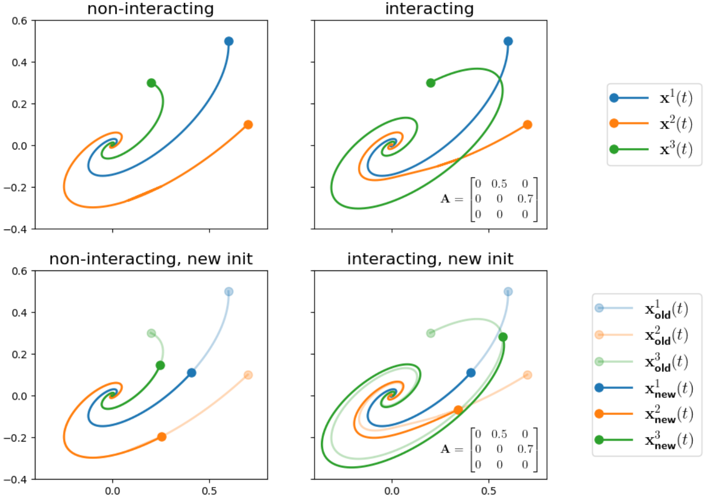
Giangiacomo Mercatali†
HES-SO Genève

Andre Freitas
University of Manchester
Idiap Research Institute

Jie Chen
MIT-IBM Watson AI Lab
IBM Research

†Work done at the University of Manchester

Modeling Interacting Time Series via a DAG



DAG-Based ODE Model

Let $\mathbf{A}$ be the DAG adjacency matrix of n time series $\mathbf{X}(t) : \mathbb{R} \to \mathbb{R}^{n \times d}$.

$$\dot{\mathbf{X}} = f(t, \mathbf{X}, \mathbf{A}) \quad \text{with} \quad \mathbf{X}(0) = \mathbf{X}_0, \tag{1}$$

where the right-hand side f is unknown.

Problem: Given $\mathbf{X}(t)$ at irregular time points $t_0, \ldots, t_N$, predict $\mathbf{X}(t)$ for any $t \geq t_0$. Additionally, find $\mathbf{A}$.

Graph neural flow framework:

$$\mathbf{X}(t) = F(t, \mathbf{X}_0, \mathbf{A}), \tag{2}$$

where $F(t, \mathbf{X}, \mathbf{A})$ is a neural network satisfying:

1. $F(0, \mathbf{X}_0, \mathbf{A}) = \mathbf{X}_0$;
2. $F(t, \mathbf{X}, \mathbf{A})$ is invertible in $\mathbf{X}$ for any t and $\mathbf{A}$; equivalently, the streamline $F(t, \mathbf{X}_0, \mathbf{A})$ given any $\mathbf{X}_0$ and $\mathbf{A}$ is not self-intersecting.

Graph Encoder

$$\widetilde{\mathbf{X}} = \mathrm{GCN}(\mathbf{A}, \mathbf{X}) = \widehat{\mathbf{A}}\,\mathrm{ReLU}(\widehat{\mathbf{A}}\mathbf{X}\mathbf{W})\mathbf{U}, \tag{3}$$

$$\widehat{\mathbf{A}} = \mathbf{I} - \mathbf{A}^\top/\gamma, \quad \text{where} \quad \gamma = \max_j \left\{ \sum_{i \neq j} |\mathbf{B}_{ij}| \right\} \ \text{and} \ \mathbf{B} = \mathbf{A} + \mathbf{A}^\top. \tag{4}$$

Theorem (Contractiveness of GCN)

For any DAG adjacency matrix $\mathbf{A}$, the matrix $\widehat{\mathbf{A}}$ admits $\|\widehat{\mathbf{A}}\|_2 \leq 2$.

Parameterization of F (1/3): ResNet Flow

$$F(t, \mathbf{X}, \mathbf{A}) = \mathbf{X} + \varphi(t) \cdot g(t, \mathbf{X}, \mathbf{A}), \tag{5}$$

$$g(t, \mathbf{X}, \mathbf{A}) = \mathrm{MLP}^1(\mathbf{X}||\widetilde{\mathbf{X}}||t) \odot \mathrm{MLP}^2(\mathbf{X}||t), \qquad \widetilde{\mathbf{X}} = \mathrm{GCN}(\mathbf{A}, \mathbf{X}). \tag{6}$$

Parameterization of F (2/3): GRU Flow

$$F(t, \mathbf{X}, \mathbf{A}) = \underbrace{\mathbf{X} + \varphi(t) \cdot h^1(t, \mathbf{X})}_{\text{base form}} \odot h^2(t, \widetilde{\mathbf{X}}), \qquad \widetilde{\mathbf{X}} = \mathrm{GCN}(\mathbf{A}, \mathbf{X}), \qquad (7)$$

where h^k, $k = 1, 2$, is computed by

$$r^k(t, \mathbf{X}) = \beta \cdot \mathrm{sigmoid}(f_r^k(t, \mathbf{X})), \qquad c^k(t, \mathbf{X}) = \tanh(f_c^k(t, r^k(t, \mathbf{X}) \odot \mathbf{X})),$$
$$z^k(t, \mathbf{X}) = \alpha \cdot \mathrm{sigmoid}(f_z^k(t, \mathbf{X})), \qquad h^k(t, \mathbf{X}) = z^k(t, \mathbf{X}) \odot (c^k(t, \mathbf{X}) - \mathbf{X}).$$

Theorem

If $f_z^k(t, \cdot)$, $f_r^k(t, \cdot)$, $f_c^k(t, \cdot)$, and $\mathrm{GCN}(\mathbf{A}, \cdot)$ are contractive, the function $F(t, \cdot, \mathbf{A})$ is invertible whenever $\alpha(5\beta + 6) \leq 2$.

Parameterization of F (3/3): Coupling Flow

$$F(t, \mathbf{X}, \mathbf{A})_U = \mathbf{X}_U \odot \exp\left(\varphi_u(t) \cdot u(t, \mathbf{X}_V, \widetilde{\mathbf{X}}_V)\right) + \left(\varphi_v(t) \cdot v(t, \mathbf{X}_V, \widetilde{\mathbf{X}}_V)\right)$$
$$F(t, \mathbf{X}, \mathbf{A})_V = \mathbf{X}_V, \qquad \underbrace{\widetilde{\mathbf{X}}_V = \mathrm{GCN}(\mathbf{A}, \mathbf{X}_V)}_{\text{graph conditioned}}, \tag{8}$$

where

$$u(t, \mathbf{X}_V, \widetilde{\mathbf{X}}_V) = \mathrm{MLP}^3\left(\mathrm{MLP}^1(\mathbf{X}_V || t) \;||\; \mathrm{MLP}^2(\widetilde{\mathbf{X}}_V || t)\right)$$
$$v(t, \mathbf{X}_V, \widetilde{\mathbf{X}}_V) = \mathrm{MLP}^4\left(\mathrm{MLP}^1(\mathbf{X}_V || t) \;||\; \mathrm{MLP}^2(\widetilde{\mathbf{X}}_V || t)\right).$$

Learning the Graph

$$\min_{\mathbf{A},\boldsymbol{\theta}} \quad \mathcal{L}(\mathbf{A}, \boldsymbol{\theta}) \quad \text{s.t. } \mathbf{A} \text{ corresponds to a DAG.} \tag{9}$$

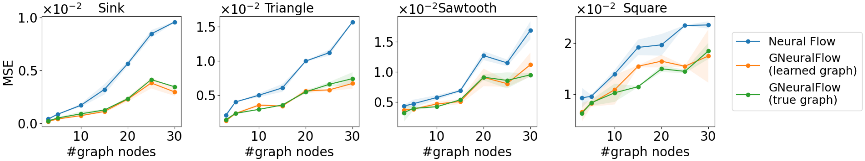
Theorem (Zheng et al, 2018)

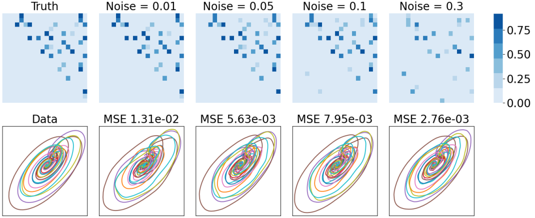
$\mathbf{A}$ is a DAG adjacency matrix iff $\mathrm{tr}(\mathrm{expm}(\mathbf{A} \odot \mathbf{A})) = n$.

Theorem (Yu et al, 2019)

$\mathbf{A}$ is a DAG adjacency matrix iff $\mathrm{tr}((\mathbf{I} + \alpha\mathbf{A} \odot \mathbf{A})^n) = n$ for any $\alpha \neq 0$.

Synthetic Systems





Synthetic Systems

		Sink MSE ($\times 10^{-4}$)	Triangle MSE ($\times 10^{-3}$)	Sawtooth MSE ($\times 10^{-3}$)	Square MSE ($\times 10^{-3}$)
No Graph	Neural ODE	10.6 ($\pm$ 0.03)	8.32 ($\pm$ 0.24)	9.32 ($\pm$ 0.36)	16.8 ($\pm$ 0.39)
	Neural flow (ResNet)	8.41 ($\pm$ 0.05)	4.01 ($\pm$ 0.52)	4.73 ($\pm$ 0.06)	9.61 ($\pm$ 0.02)
	Neural flow (GRU)	10.9 ($\pm$ 0.43)	10.3 ($\pm$ 0.45)	16.1 ($\pm$ 0.41)	17.2 ($\pm$ 0.51)
	Neural flow (Coupling)	9.31 ($\pm$ 0.23)	12.2 ($\pm$ 0.41)	14.2 ($\pm$ 0.24)	13.0 ($\pm$ 0.63)
	GRU-D	12.3 ($\pm$ 0.23)	11.3 ($\pm$ 0.32)	17.6 ($\pm$ 0.53)	18.7 ($\pm$ 0.31)
Graph ODE	GDE	10.4 ($\pm$ 0.20)	3.99 ($\pm$ 0.05)	7.65 ($\pm$ 0.03)	15.89 ($\pm$ 0.81)
	LG-ODE	8.57 ($\pm$ 0.06)	3.58 ($\pm$ 0.21)	7.07 ($\pm$ 0.02)	13.99 ($\pm$ 0.73)
	CF-GODE	8.60 ($\pm$ 0.14)	7.19 ($\pm$ 0.02)	8.19 ($\pm$ 0.03)	13.53 ($\pm$ 0.11)
Graph Learn	NRI	5.25 ($\pm$ 0.02)	3.96 ($\pm$ 0.16)	4.99 ($\pm$ 0.12)	9.39 ($\pm$ 0.45)
	dNRI	5.40 ($\pm$ 0.04)	3.39 ($\pm$ 0.09)	4.97 ($\pm$ 0.21)	9.78 ($\pm$ 0.21)
Our Method	GNeuralFlow (ResNet)	**3.95** ($\pm$ 0.32)	**2.32** ($\pm$ 0.11)	**3.84** ($\pm$ 0.06)	**8.24** ($\pm$ 0.64)
	GNeuralFlow (GRU)	6.83 ($\pm$ 0.23)	5.41 ($\pm$ 0.23)	5.11 ($\pm$ 0.13)	9.14 ($\pm$ 0.61)
	GNeuralFlow (Coupling)	4.45 ($\pm$ 0.51)	3.21 ($\pm$ 0.34)	4.25 ($\pm$ 0.09)	8.33 ($\pm$ 0.23)

Latent Variable Modeling: Smoothing

		Activity Accuracy	Physionet AUC	Activity MSE ($\times 10^{-2}$)	Physionet MSE ($\times 10^{-3}$)	MujoCo MSE ($\times 10^{-3}$)
No Graph	ODE-RNN	0.785 (± 0.003)	0.781 (± 0.004)	6.050 (± 0.10)	4.52 (± 0.03)	**2.540** (± 0.12)
	Neural flow (ResNet)	0.760 (± 0.004)	0.784 (± 0.010)	6.279 (± 0.09)	4.90 (± 0.12)	8.403 (± 0.14)
	Neural flow (GRU)	0.783 (± 0.008)	0.788 (± 0.008)	5.837 (± 0.07)	5.04 (± 0.13)	4.249 (± 0.07)
	Neural flow (Coupling)	0.752 (± 0.012)	0.788 (± 0.004)	6.579 (± 0.04)	4.86 (± 0.07)	4.217 (± 0.14)
Graph ODE	GDE	0.721 (± 0.014)	0.757 (± 0.010)	6.491 (± 0.011)	4.83 (± 0.38)	5.220 (± 0.42)
	LG-ODE	0.743 (± 0.023)	0.748 (± 0.018)	5.738 (± 0.089)	4.87 (± 0.27)	6.699 (± 0.83)
	CG-ODE	0.768 (± 0.048)	0.783 (± 0.082)	6.241 (± 0.012)	4.73 (± 0.07)	4.312 (± 0.17)
Our Method	GNeuralFlow (ResNet)	0.786 (± 0.009)	0.800 (± 0.009)	5.947 (± 0.03)	4.31 (± 0.06)	2.916 (± 0.21)
	GNeuralFlow (GRU)	0.804 (± 0.003)	**0.812** (± 0.001)	**5.169** (± 0.05)	**4.23** (± 0.15)	4.112 (± 0.13)
	GNeuralFlow (Coupling)	**0.808** (± 0.005)	0.808 (± 0.008)	5.431 (± 0.10)	4.59 (± 0.23)	3.849 (± 0.07)

Latent Variable Modeling: Filtering

		MSE	NLL
No Graph	GRU-ODE-Bayes	0.379 (± 0.005)	0.748 (± 0.045)
	Neural flow (ResNet)	0.379 (± 0.005)	0.774 (± 0.059)
	Neural flow (GRU)	0.364 (± 0.008)	0.734 (± 0.054)
	Neural flow (Coupling)	0.366 (± 0.002)	0.675 (± 0.003)
Graph ODE	GDE	0.342 (± 0.001)	0.657 (± 0.007)
	LG-ODE	0.349 (± 0.002)	0.649 (± 0.005)
	CG-ODE	0.372 (± 0.011)	0.825 (± 0.018)
Our Method	GNeuralFlow (ResNet)	0.356 (± 0.0007)	0.663 (± 0.008)
	GNeuralFlow (GRU)	**0.335** (± 0.003)	**0.606** (± 0.001)
	GNeuralFlow (Coupling)	0.350 (± 0.004)	0.662 (± 0.008)

Conclusions and Discussions

Summary

- learning the systemic interactions of time series
- learning the graph structure in tandem with the system dynamics

Limitation

- number of parameters on the $\mathbf{A}$ part grows quadratically with the number of time series (nodes)
- mitigation is to introduce structures into $\mathbf{A}$ (such as low-rankness),

More info

Poster session 2.

Wed 11 Dec 4:30 p.m. PST — 7:30 p.m. PST