Number: 0140

ICONIP 2023
November 20-23, 2023, Changsha, China

Stony Brook University

AGGDN: A Continuous Stochastic Predictive Model for Monitoring Sporadic Time Series on Graphs

Yucheng Xing[1], Jacqueline Wu[2], Yingru Liu[1], Xuewen Yang[1,3], Xin Wang[1]

[1] Stony Brook University, [2] New York University, [3] InnoPeak Technology Inc.

Motivation

- Accurate prediction of the networked system states based on collected data is important for timely and effective control of the system;
- Challenges:
  - Spatial-temporal interactions among dynamic signals are difficult to capture;
  - The collected data could be sparse and irregular in both spatial and temporal domains due to constraints such as unreliable communications and device malfunctions;
  - The collected data are usually influenced by measurement noise and other process uncertainties within the system;

Objective

- We propose a new continuous-time stochastic method to provide accurate and timely predictions of the system states based on sporadic observations with system uncertainties
  - Model spatial-temporal interaction of dynamic signals across graph;
  - Model underlying stochastic process with irregular data;
  - Capture complicated data distribution with process uncertainty and measurement noise;

Methods

- **Hybrid ODE-SDE Structure**
  - **ODE Module**
    - Encoding the stable parts of the signals and the topological information into the hidden features H_t;
    - Providing an approximation of system states to modulate SDE;
  - **SDE Module**
    - Embedding the system uncertainties into the stochastic latent states Z_t;
    - Refining the coarse predictions from ODE to get the final accurate output;
- **Component Details – dyn-ODE Module**
  - **Soft-masking mechanism:** Hidden features H_t is composed of two factors

$$H_t = \rho(H_{t,m}W_m + b_m) \odot H_{t,f}$$

    - $H_{t,f}$: the feature factor to extract the information of the system;
    - $H_{t,m}$: the masking factor to modulate the values in features
  - **Dynamic Diffusion Convolution:** replace the fixed binary adjacency matrix A with $W_A \odot A$, where W_A is a learnable matrix capturing neighbor impact;
    - During the interval (t_{n-1}, t_n), using Euler Method

$$H_{t,f} = H_{t-\delta t,f} + F_t\left(H_{t-\delta t,f}, W_A \odot A\right)\delta t$$
$$H_{t,m} = H_{t-\delta t,m} + F_m(H_{t-\delta t,m}, W_A \odot A)\delta t$$

      - $F_t(\cdot)$ and $F_m(\cdot)$ are Dynamic Diffusion Convolution Networks (dynDCNs)
    - At an observation time t_n

$$H_{t_n,f} = G_t\left(H_{t_n-\delta t,f}, \mathcal{O}_{t_n}, W_A \odot A\right)$$
$$H_{t_n,m} = G_m\left(H_{t_n-\delta t,m}, \mathcal{O}_{t_n}, W_A \odot A\right)$$

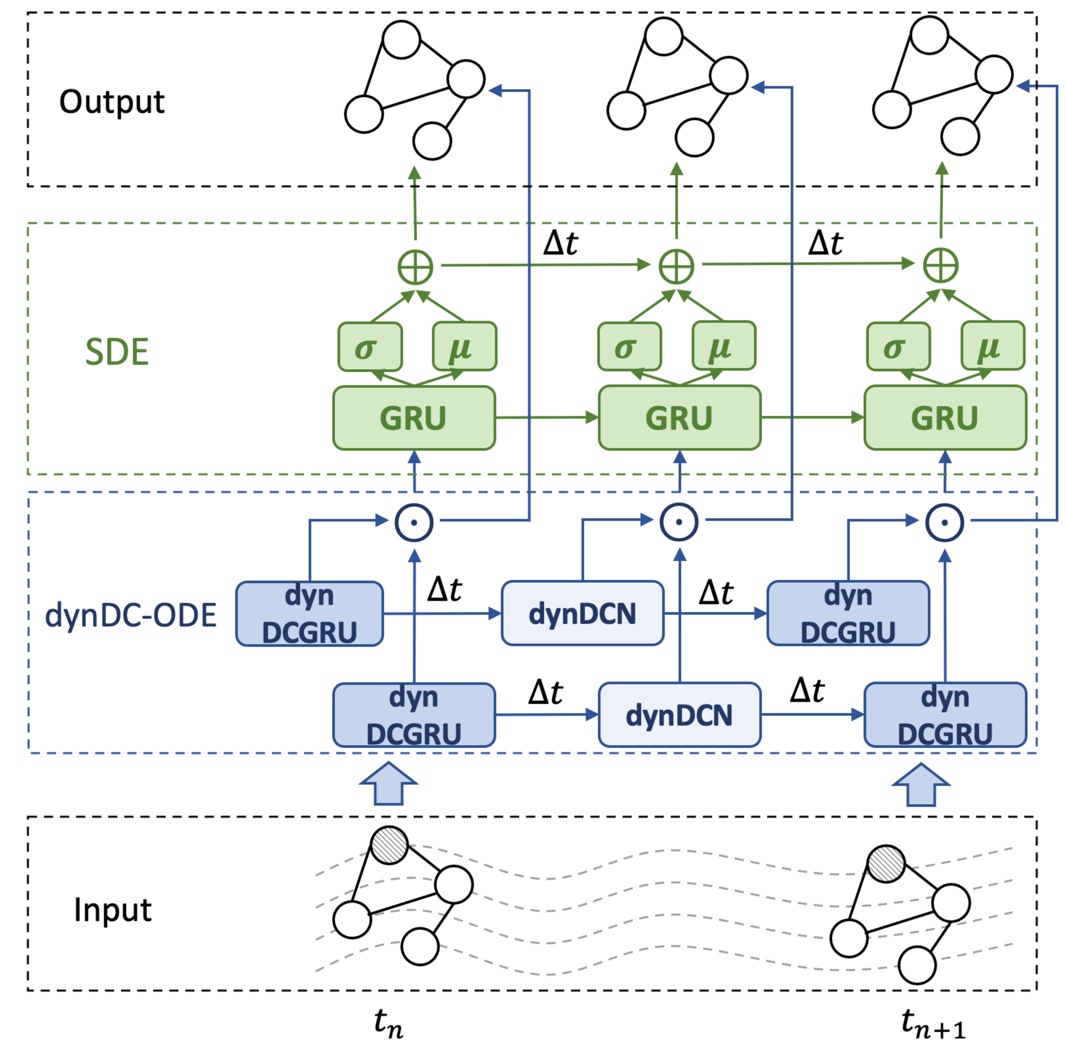
      - $G_t(\cdot)$ and $G_m(\cdot)$ are Dynamic Diffusion Convolution Gate Recurrent Units (dynDCGRUs)
- **Component Details – SDE Module**
  - Computing the integration by Euler-Maruyama Method

$$Z_t = Z_{t-\delta t} + \mu(Z_{t-\delta t}, H_{\leq t})\delta t + \sqrt{\delta t}\,\sigma(H_{\leq t})\epsilon_t$$

- **Component Details – Output Module**

$$\hat{\mathcal{X}}_t = N_{\hat{\mathcal{X}}}(H_t) + N_{\hat{\mathcal{X}}}^{(res)}(H_t, Z_t)$$

  - $N_{\hat{\mathcal{X}}}(\cdot)$: predict the smooth and stable trend of signals;
  - $N_{\hat{\mathcal{X}}}^{(res)}(\cdot)$: estimate the residual stochastic variations;



- **Training Details – Wasserstein Adversarial Training**
  - Simplified log-likelihood:

$$\mathcal{L}_{cond} = \mathbb{E}_{\{\mathcal{X}_{t_n}, \mathcal{M}_{t_n}\} \in \mathbb{D}} \sum_{n=1}^{N} \mathcal{M}_{t_n} \otimes \log P_{\mathcal{G}}(\mathcal{X}_{t_n} | Z_{t_n}, \mathcal{O}_{t_0:t_{n-1}}, A)$$

  - Adversarial training loss:

$$\mathcal{L}_{adv} = \mathbb{E}_{\{\mathcal{X}_{t_n}, \mathcal{M}_{t_n}\} \in \mathbb{D}}\left[\mathcal{F}\left(\{\mathcal{X}_{t_n} \odot \mathcal{M}_{t_n}\}\right) - \mathbb{E}_{\epsilon_t \in \mathcal{N}(0,1)} \mathcal{F}\left(\{\hat{\mathcal{X}}_{t_n} \odot \mathcal{M}_{t_n}\}\right)\right]$$

  - Total objective:

$$\mathcal{G}_* = arg\min_{\mathcal{G}}(\lambda \mathcal{L}_{adv} - \mathcal{L}_{cond})$$
$$\mathcal{F}_* = arg\max_{\mathcal{F}} \mathcal{L}_{adv}$$

Results

- **Performance Comparison**

During experiments, we assume only partial of data are available due to occlusions:

  - p_t: ratio of the frames within the data sequence are observable;
  - p_s: ratio of the nodes available within each observable frame;

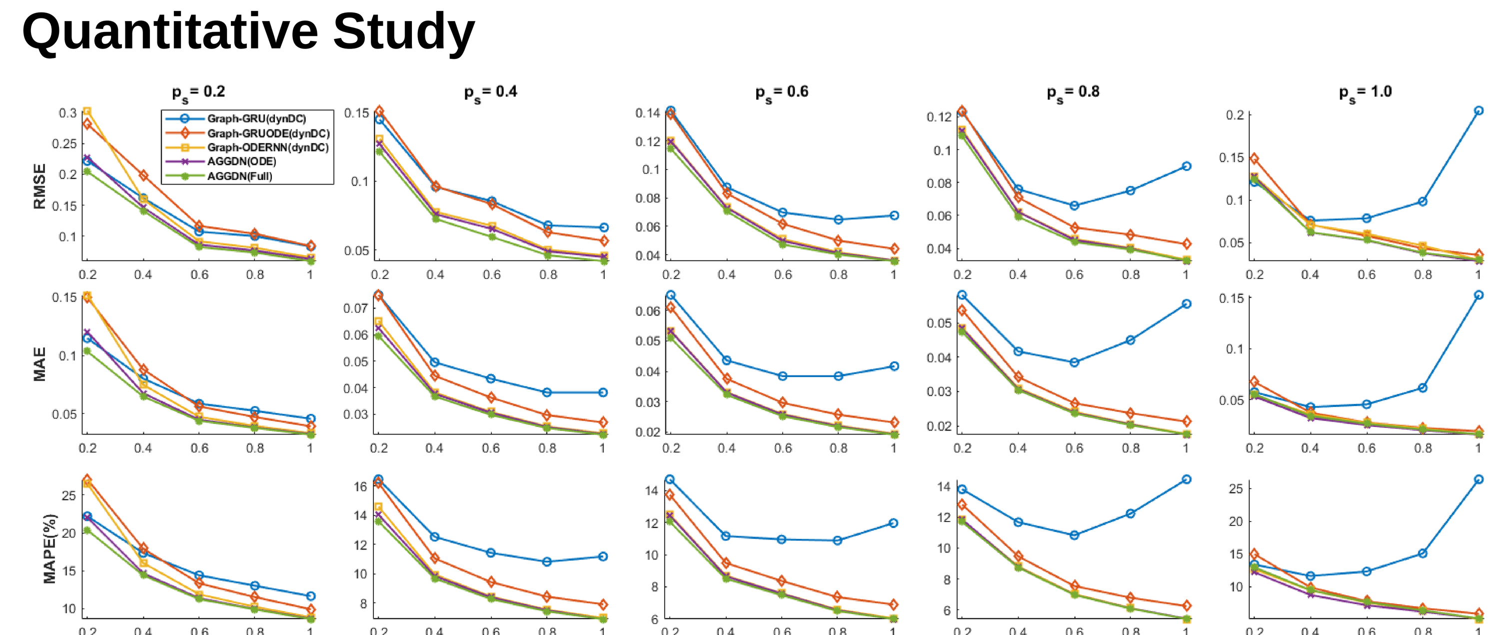
Datasets	IEEE33-Nodes			METR-LA			PEMS-BAY		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
Discrete									
STGCN	0.0812	0.1273	18.18%	0.2290	0.4375	37.58%	0.2499	0.4468	49.63%
Graph-GRU	0.0349	0.0643	9.65%	0.1978	0.4226	33.45%	0.1925	0.3725	41.17%
Continuous									
Graph-ODE-RNN	0.0306	0.0578	8.36%	0.1918	0.4217	32.73%	0.1774	0.3605	37.34%
Graph-GRU-ODE	0.0313	0.0607	8.49%	0.1947	0.4322	32.90%	0.1726	0.3488	37.44%
Ours									
AGGDN	**0.0243**	**0.0457**	**7.03%**	**0.1612**	**0.3480**	**30.41%**	**0.1489**	**0.2739**	**35.32%**

Table 1. Performance of different models on various datasets ($p_t = 0.5, p_s = 0.8$)

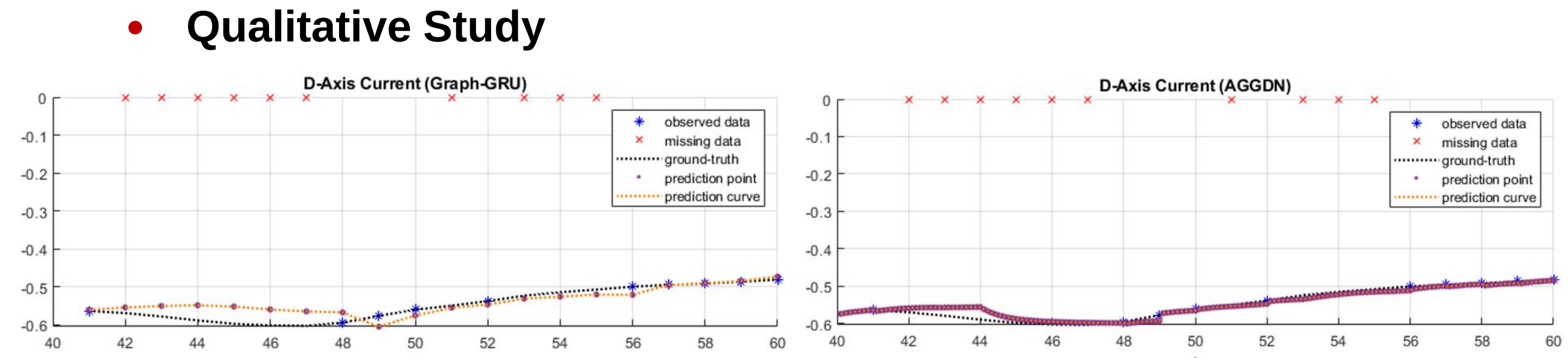
Datasets	IEEE33-Nodes			METR-LA			PEMS-BAY		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
Discrete									
STGCN	0.0925	0.1598	19.52%	0.2497	0.4681	39.64%	0.2826	0.5084	53.15%
Graph-GRU	0.0522	0.1068	12.39%	0.2195	0.4460	35.96%	0.2219	0.4345	44.78%
Continuous									
Graph-ODE-RNN	0.0458	0.0992	11.02%	0.2174	0.4536	35.57%	0.2026	0.4147	40.65%
Graph-GRU-ODE	0.0463	0.1018	11.03%	0.2160	0.4552	35.36%	0.1998	0.4090	40.17%
Ours									
AGGDN	**0.0322**	**0.0706**	**8.51%**	**0.1719**	**0.3576**	**31.71%**	**0.1715**	**0.3331**	**37.91%**

Table 2. Performance of different models on various datasets ($p_t = 0.4, p_s = 0.6$)

- **Quantitative Study**



- **Qualitative Study**



Corresponding Author
Yucheng Xing
yucheng.xing@stonybrook.edu