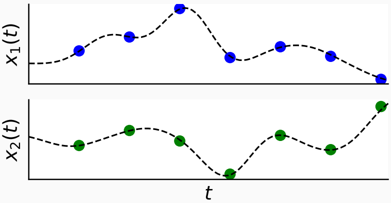
mTAN: Multi-Time Attention Networks

For Irregularly Sampled Time Series

Satya Narayan Shukla, Benjamin Marlin

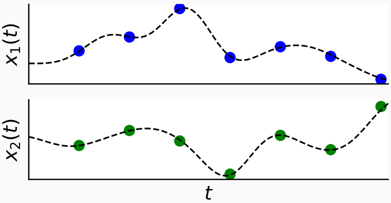
University of Massachusetts Amherst

Irregularly Sampled Time Series

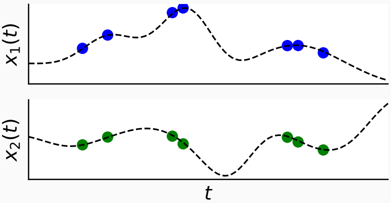


Multivariate regularly sampled

Irregularly Sampled Time Series

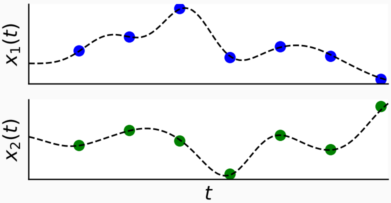


Multivariate regularly sampled

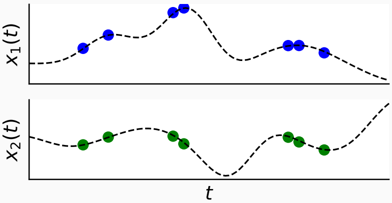


Multivariate irregularly sampled(aligned)

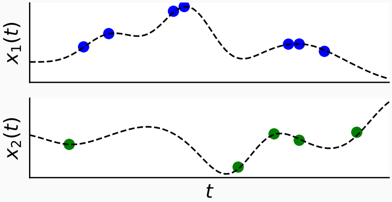
Irregularly Sampled Time Series



Multivariate regularly sampled



Multivariate irregularly sampled(aligned)



Multivariate irregularly sampled (unaligned)

Challenges

1. Irregular spacing between observation time points
2. Variable numbers of observations
3. Lack of alignment of observation time points

Contributions

- A flexible approach to modeling multivariate, sparse and irregularly sampled time series data by leveraging a time attention mechanism.

Contributions

- A flexible approach to modeling multivariate, sparse and irregularly sampled time series data by leveraging a time attention mechanism.
- Our approach uses a temporally distributed latent representation to better capture local structure in time series data.

Contributions

- A flexible approach to modeling multivariate, sparse and irregularly sampled time series data by leveraging a time attention mechanism.
- Our approach uses a temporally distributed latent representation to better capture local structure in time series data.
- Improved interpolation and classification performance than current state-of-the-art methods with significantly reduced training times.

mTAN: Multi-Time Attention Networks

- Continuous-time interpolation-based models

$$\hat{x}(t) = \frac{\sum_i \kappa_\theta(t, t_i)\, x(t_i)}{\sum_i \kappa_\theta(t, t_i)}$$

where $\kappa_\theta()$ is a similarity kernel, e.g. squared exponential kernel[1].

[1] Satya Narayan Shukla and Benjamin Marlin. Interpolation-prediction networks for irregularly sampled time series. In International Conference on Learning Representations, 2019.

mTAN: Multi-Time Attention Networks

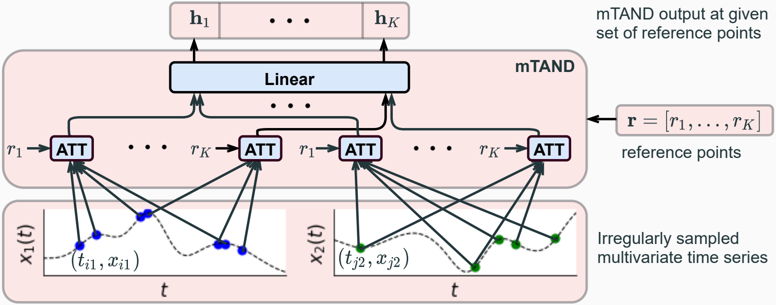
- Continuous-time interpolation-based models

$$\hat{x}(t) = \frac{\sum_i \kappa_\theta(t, t_i)\, x(t_i)}{\sum_i \kappa_\theta(t, t_i)}$$

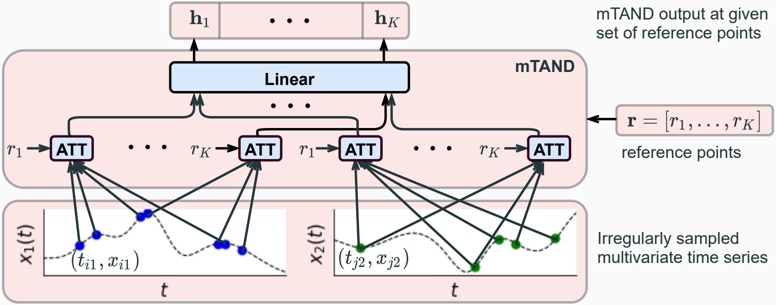
  where $\kappa_\theta()$ is a similarity kernel, e.g. squared exponential kernel[1].
- Continuous-time embedding coupled with Time Attention
- Replace the use of a fixed similarity kernel
- More representational flexibility than previous interpolation-based models

[1] Satya Narayan Shukla and Benjamin Marlin. Interpolation-prediction networks for irregularly sampled time series. In International Conference on Learning Representations, 2019.

mTAN: Multi-Time Attention Networks

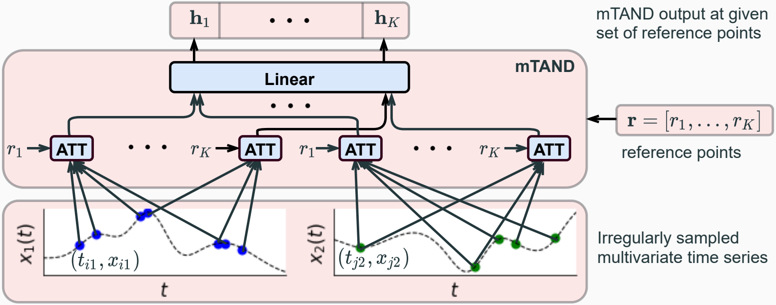


mTAN: Multi-Time Attention Networks



$$\hat{x}_{hd}(t, s) = \sum_{i=1}^{L_d} \mathrm{softmax}\left(\frac{\phi_h(t) w v^T \phi_h(t_{id})^T}{\sqrt{d_k}}\right) x_{id}$$

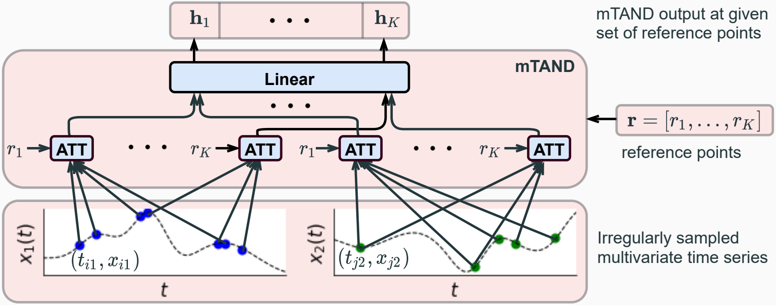
mTAN: Multi-Time Attention Networks



$$\hat{x}_{hd}(t, s) = \sum_{i=1}^{L_d} \mathrm{softmax}\left(\frac{\phi_h(t) w v^T \phi_h(t_{id})^T}{\sqrt{d_k}}\right) x_{id}$$

$$\phi_h(t)[i] = \begin{cases} \omega_{0h} \cdot t + \alpha_{0h}, & \text{if } i = 0 \\ \sin(\omega_{ih} \cdot t + \alpha_{ih}), & \text{if } 0 < i < d_r \end{cases}$$

mTAN: Multi-Time Attention Networks

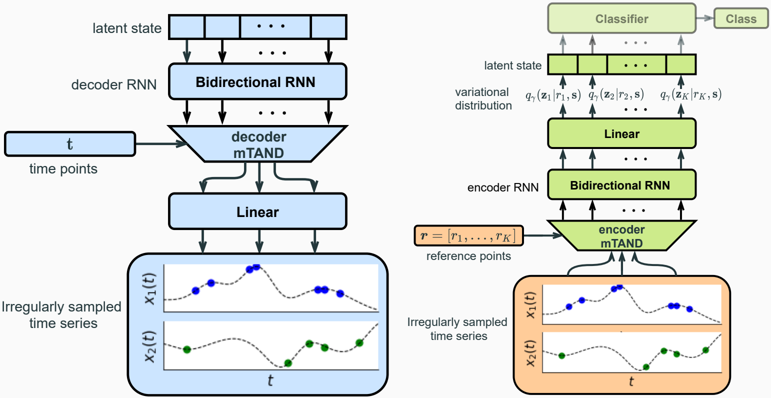


$$\hat{x}_{hd}(t, \mathbf{s}) = \sum_{i=1}^{L_d} \operatorname{softmax}\left(\frac{\phi_h(t) w v^T \phi_h(t_{id})^T}{\sqrt{d_k}}\right) x_{id}$$

$$\mathrm{mTAN}(t, \mathbf{s})[j] = \sum_{h=1}^{H} \sum_{d=1}^{D} \hat{x}_{hd}(t, \mathbf{s}) \cdot U_{hdj}$$

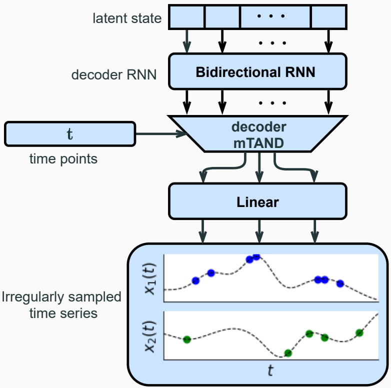
- mTAND (discretized mTAN) materializes mTAN's output at a set of reference time points

Encoder-Decoder Framework

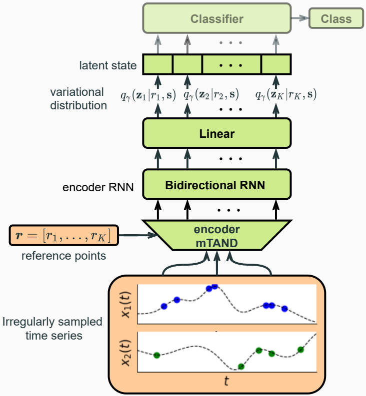


(a) Generative Model (Decoder) (b) Inference Network (Encoder)

Encoder-Decoder Framework



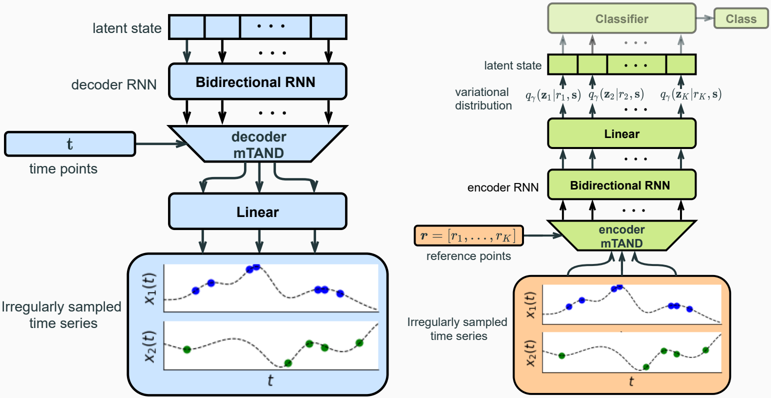
(a) Generative Model (Decoder)



(b) Inference Network (Encoder)

$$\mathcal{L}_{\text{NVAE}}(\theta,\gamma) = \sum_{n=1}^{N} \frac{1}{\sum_d L_{dn}} \left(\mathbb{E}_{q_\gamma(\mathbf{z}|\mathbf{r},\mathbf{s}_n)}[\log p_\theta(\mathbf{x}_n|\mathbf{z},\mathbf{t}_n)] - D_{\mathrm{KL}}(q_\gamma(\mathbf{z}|\mathbf{r},\mathbf{s}_n)||p(\mathbf{z})) \right)$$

Encoder-Decoder Framework

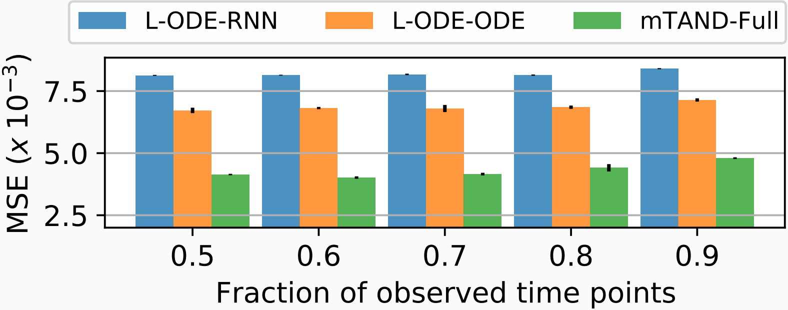


(a) Generative Model (Decoder) (b) Inference Network (Encoder)

$$\mathcal{L}_{\text{NVAE}}(\theta,\gamma) = \sum_{n=1}^{N} \frac{1}{\sum_d L_{dn}} \left(\mathbb{E}_{q_\gamma(\mathrm{z}|\mathrm{r},\mathrm{s}_n)}[\log p_\theta(\mathrm{x}_n|\mathrm{z},\mathrm{t}_n)] - D_{\mathrm{KL}}(q_\gamma(\mathrm{z}|\mathrm{r},\mathrm{s}_n)||p(\mathrm{z})) \right)$$

$$\mathcal{L}_{\text{sup}}(\theta,\gamma,\delta) = \mathcal{L}_{\text{NVAE}}(\theta,\gamma) + \lambda \mathbb{E}_{q_\gamma(\mathrm{z}|\mathrm{r},\mathrm{s}_n)} \log p_\delta(y_n|\mathrm{z})$$

Interpolation Experiments: PhysioNet



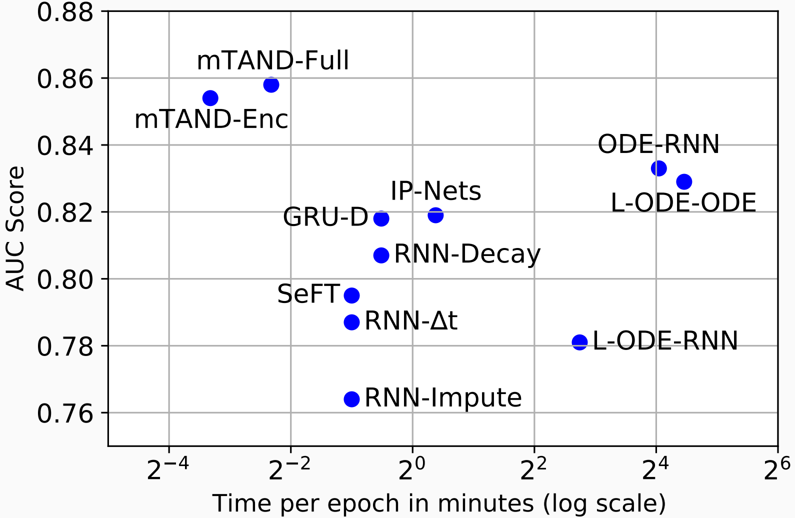
- **L-ODE-RNN**: Latent ODE with RNN encoder[2].
- **L-ODE-ODE**: Latent ODE with ODE-RNN encoder[3].

[2] T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud. Neural ordinary differential equations. In Advances in Neural Information Processing Systems. 2018.

[3] Y. Rubanova, R. T. Q. Chen, and D. K. Duvenaud. Latent ordinary differential equations for irregularly-sampled time series. In Advances in Neural Information Processing Systems. 2019.

Classification Experiments: PhysioNet

Classification Experiments: PhysioNet



Thank You.

- Poster Session 2, 3rd May 2021, 9 am - 11 am PST.
- Code: `https://github.com/reml-lab/mTAN`
- Paper: `https://arxiv.org/pdf/2101.10318.pdf`
- Contact: `snshukla@cs.umass.edu`