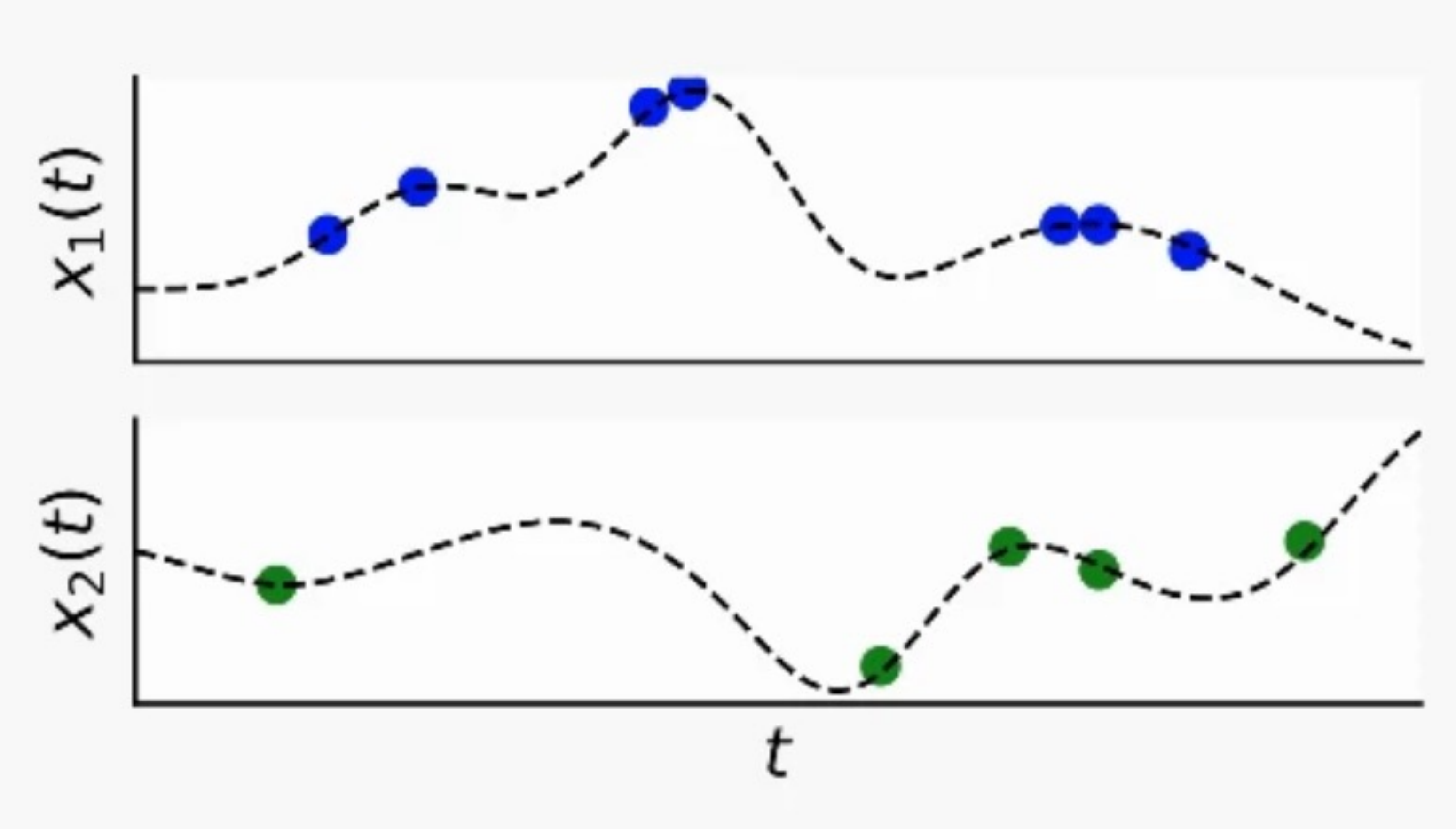
MULTI-TIME ATTENTION NETWORKS FOR IRREGULARLY SAMPLED TIME SERIES

Charles

Problem

- Irregularly sampling

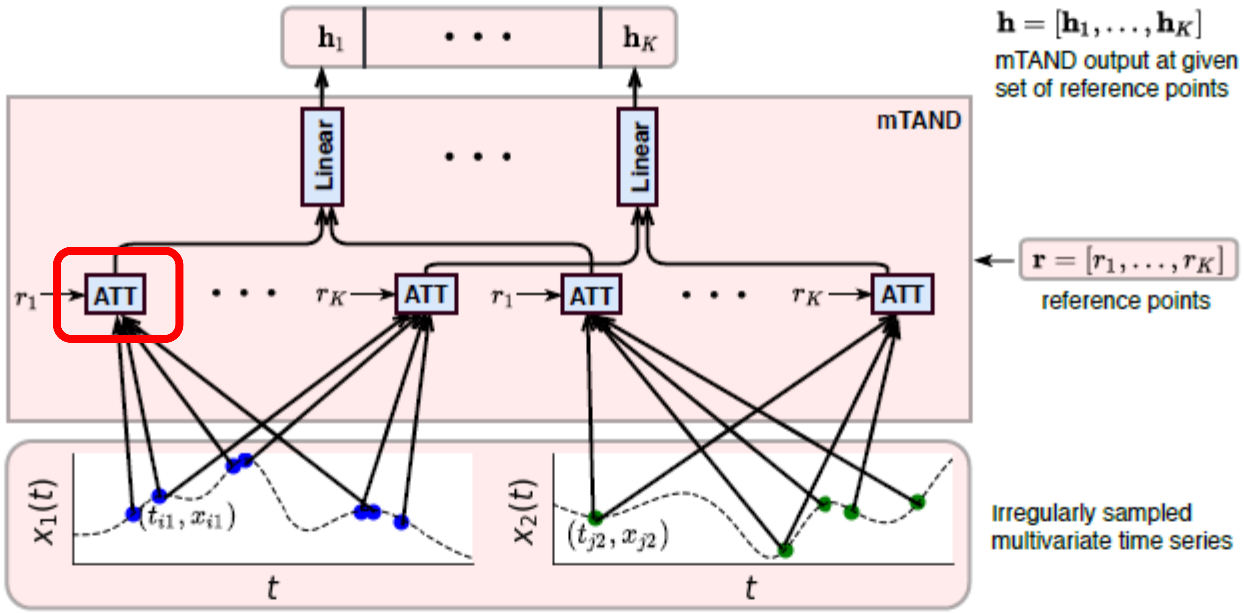


Multi-Time Attention networks -- mTANs

- 1) Flexible approach to model multivariate, sparse and irregularly sampled time series data by using time attention mechanism
- 2) Temporally distributed latent representation to better capture local structure in time series data.
- 3) Have good interpolation and classification performance and significantly reduce training times.

mTANS architecture

- ATT -> attention blocks, perform a scaled dot product attention over the observed values using time embedding of query and key time points



$$\mathrm{mTAN}(t, \mathrm{s})[j] = \sum_{h=1}^{H} \sum_{d=1}^{D} \hat{x}_{hd}(t, \mathrm{s}) \cdot U_{hdj} \quad (2)$$

$$\hat{x}_{hd}(t, \mathrm{s}) = \sum_{i=1}^{L_d} \kappa_h(t, t_{id})\, x_{id} \quad (3)$$

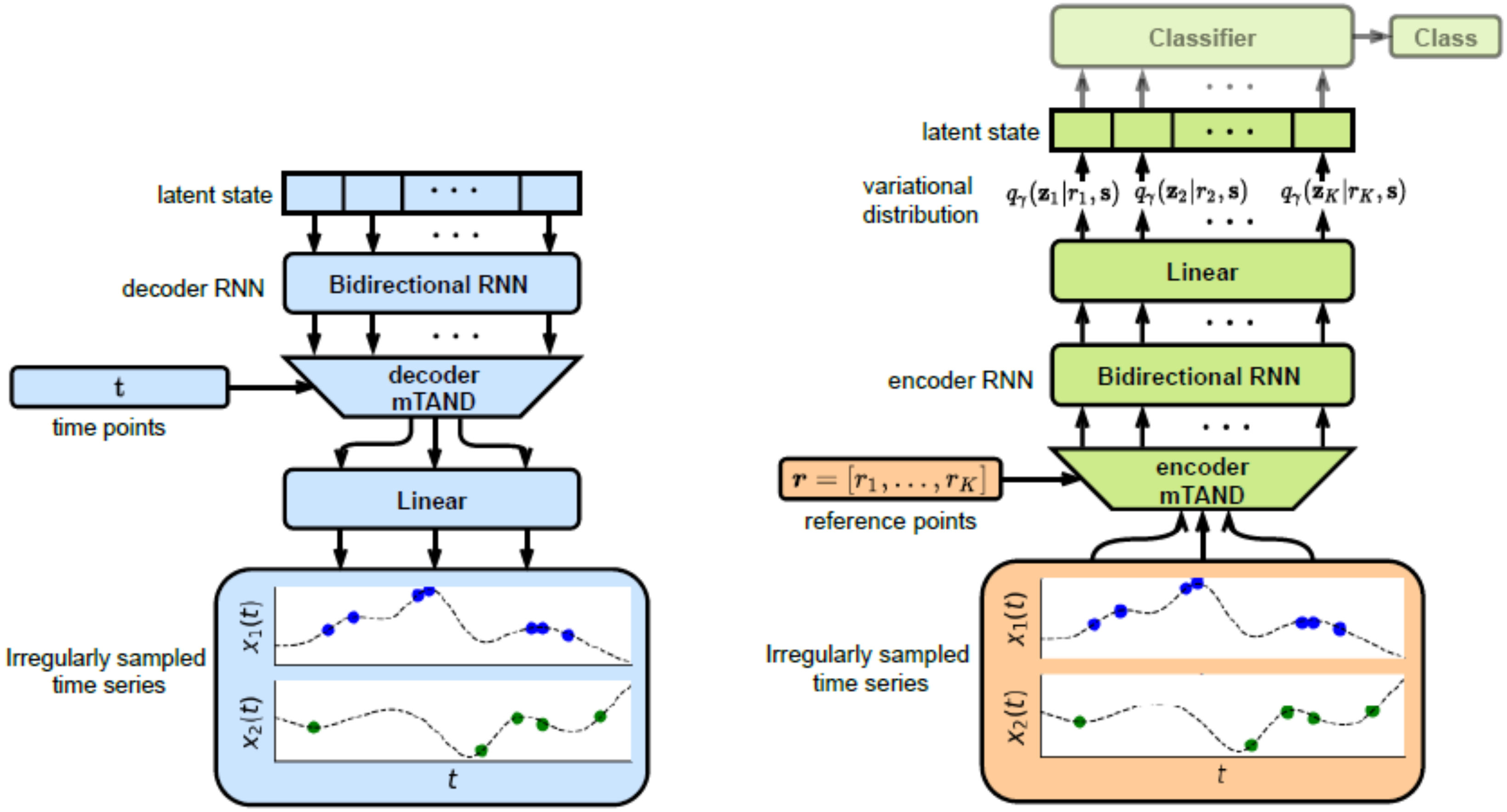
$$\kappa_h(t, t_{id}) = \frac{\exp\left(\phi_h(t)\mathbf{w}\mathbf{v}^T \phi_h(t_{id})^T / \sqrt{d_k}\right)}{\sum_{i'=1}^{L_d} \exp\left(\phi_h(t)\mathbf{w}\mathbf{v}^T \phi_h(t_{i'd})^T / \sqrt{d_k}\right)} \quad (4)$$

$$\phi_h(t)[i] = \begin{cases} \omega_{0h} \cdot t + \alpha_{0h}, & \text{if} \quad i = 0 \\ \sin(\omega_{ih} \cdot t + \alpha_{ih}), & \text{if} \quad 0 < i < d_r \end{cases} \quad (1)$$

Time embedding

Encoder-decoder framework, for classification and interpolation tasks

- Developed within VAE, variational autoencoder framework



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

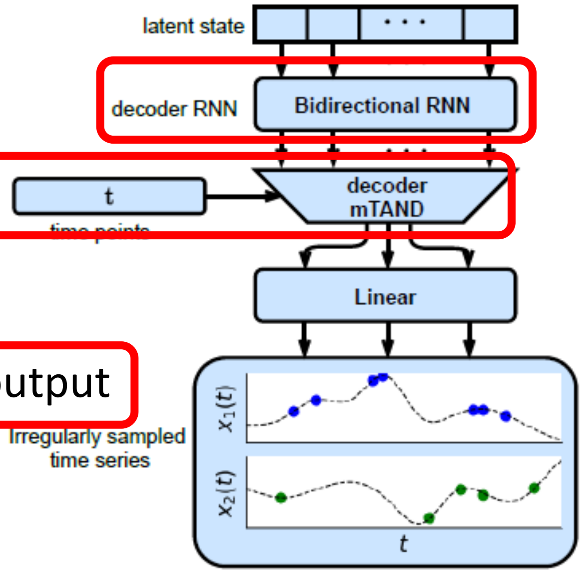
Decoder

$$\mathbf{z}_k \sim p(\mathbf{z}_k)$$

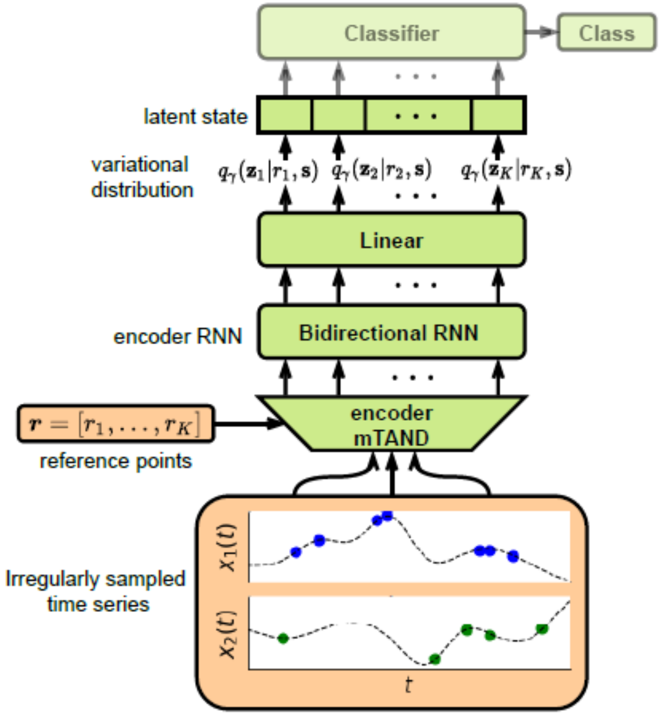
$$\mathrm{h}_{RNN}^{dec} = \mathrm{RNN}^{dec}(\mathbf{z})$$ RNN decoder output

$$\mathrm{h}_{TAN}^{dec} = \mathrm{mTAND}^{dec}(\mathrm{t}, \mathrm{h}_{RNN}^{dec})$$ mTAN decoder output

$$x_{id} \sim \mathcal{N}(x_{id}; f^{dec}(\mathrm{h}_{i,TAN}^{dec})[d], \sigma^2 \boldsymbol{I})$$



(a) Generative Model (Decoder)

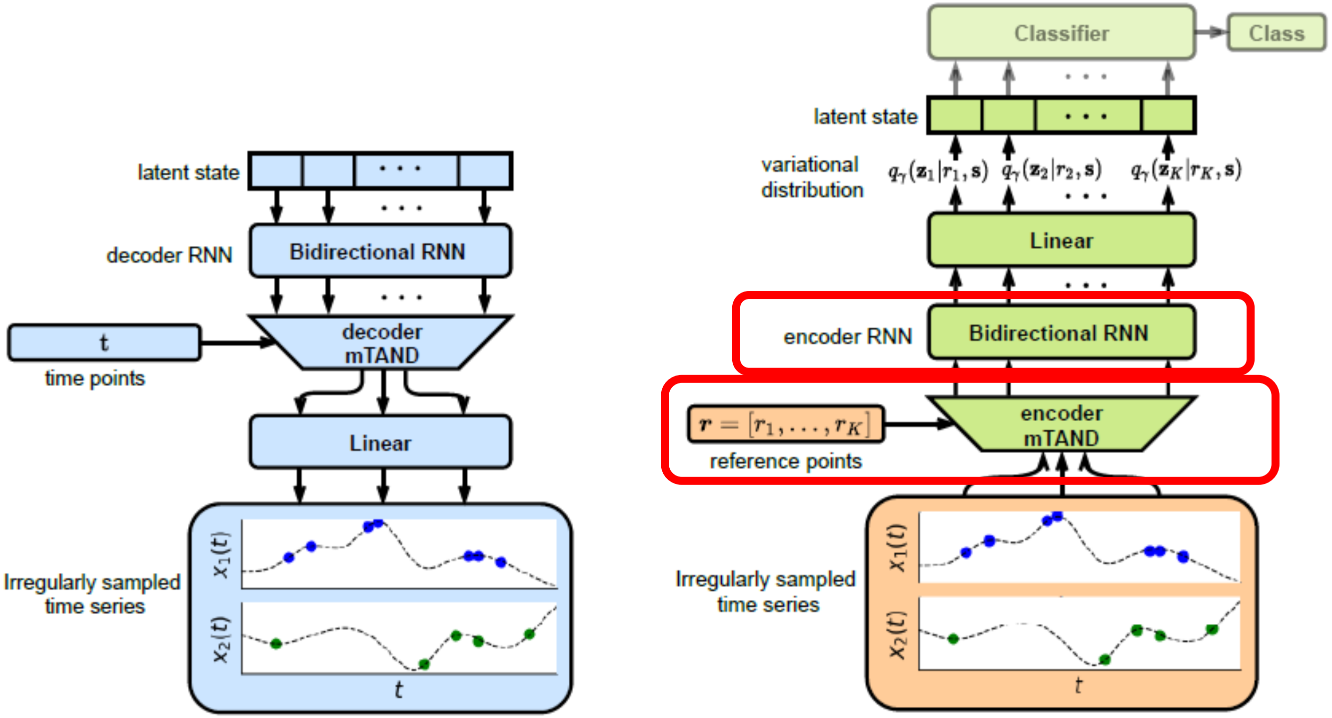


(b) Inference Network (Encoder)

$$\mathrm{h}_{RNN}^{dec} = [\mathrm{h}_{1,RNN}^{dec}, \ldots, \mathrm{h}_{K,RNN}^{dec}].$$

$$\mathrm{h}_{TAN}^{dec} = [\mathrm{h}_{1,TAN}^{dec}, \ldots, \mathrm{h}_{T,TAN}^{dec}]$$

Encoder



(a) Generative Model (Decoder)

(b) Inference Network (Encoder)

$$\mathbf{h}^{enc}_{TAN} = \text{mTAND}^{enc}(\mathbf{r}, \mathbf{s})$$ mTAND encoder output

$$\mathbf{h}^{enc}_{RNN} = \text{RNN}^{enc}(\mathbf{h}^{enc}_{TAN})$$ RNN encoder output

$$\mathbf{z}_k \sim q_\gamma(\mathbf{z}_k|\boldsymbol{\mu}_k, \boldsymbol{\sigma}^2_k), \quad \boldsymbol{\mu}_k = f^{enc}_\mu(\mathbf{h}^{enc}_{k,RNN}), \quad \boldsymbol{\sigma}^2_k = \exp(f^{enc}_\sigma(\mathbf{h}^{enc}_{k,RNN}))$$

Evaluation

Table 1: Interpolation performance versus percent observed time points on PhysioNet

Model	Mean Squared Error ($\times 10^{-3}$)				
RNN-VAE	13.418 ± 0.008	12.594 ± 0.004	11.887 ± 0.005	11.133 ± 0.007	11.470 ± 0.006
L-ODE-RNN	8.132 ± 0.020	8.140 ± 0.018	8.171 ± 0.030	8.143 ± 0.025	8.402 ± 0.022
L-ODE-ODE	6.721 ± 0.109	6.816 ± 0.045	6.798 ± 0.143	6.850 ± 0.066	7.142 ± 0.066
mTAND-Full	$\mathbf{4.139 \pm 0.029}$	$\mathbf{4.018 \pm 0.048}$	$\mathbf{4.157 \pm 0.053}$	$\mathbf{4.410 \pm 0.149}$	$\mathbf{4.798 \pm 0.036}$
Observed %	50%	60%	70%	80%	90%

Evaluation

Table 2: Classification Performance on PhysioNet, MIMIC-III and Human Activity dataset

Model	**AUC Score**		**Accuracy**	**time**
	PhysioNet	**MIMIC-III**	**Human Activity**	**per epoch**
RNN-Impute	0.764 ± 0.016	0.8249 ± 0.0010	0.859 ± 0.004	0.5
RNN-Δ_t	0.787 ± 0.014	0.8364 ± 0.0011	0.857 ± 0.002	0.5
RNN-Decay	0.807 ± 0.003	0.8392 ± 0.0012	0.860 ± 0.005	0.7
RNN GRU-D	0.818 ± 0.008	0.8270 ± 0.0010	0.862 ± 0.005	0.7
Phased-LSTM	0.836 ± 0.003	0.8429 ± 0.0035	0.855 ± 0.005	0.3
IP-Nets	0.819 ± 0.006	0.8390 ± 0.0011	0.869 ± 0.007	1.3
SeFT	0.795 ± 0.015	0.8485 ± 0.0022	0.815 ± 0.002	0.5
RNN-VAE	0.515 ± 0.040	0.5175 ± 0.0312	0.343 ± 0.040	2.0
ODE-RNN	0.833 ± 0.009	$\mathbf{0.8561 \pm 0.0051}$	0.885 ± 0.008	16.5
L-ODE-RNN	0.781 ± 0.018	0.7734 ± 0.0030	0.838 ± 0.004	6.7
L-ODE-ODE	0.829 ± 0.004	$\mathbf{0.8559 \pm 0.0041}$	0.870 ± 0.028	22.0
mTAND-Enc	0.854 ± 0.001	0.8419 ± 0.0017	$\mathbf{0.907 \pm 0.002}$	**0.1**
mTAND-Full	$\mathbf{0.858 \pm 0.004}$	$\mathbf{0.8544 \pm 0.0024}$	$\mathbf{0.910 \pm 0.002}$	0.2

Summary

- Problem: irregular sampled time series data
- Introduce mTAN module and the encoder-decoder framework
- Have good interpolation and classification performance

- mTAND -> discretized mTAN $\text{mTAND}(\mathbf{r}, \mathbf{s})[i] = \text{mTAN}(r_i, \mathbf{s}).$
- latent states $\mathbf{z} = [\mathbf{z}_1, ..., \mathbf{z}_K]$ at K reference time points.
- VAE: https://en.wikipedia.org/wiki/Variational_autoencoder