Interpolation-Prediction Networks for Irregularly Sampled Time Series

Satya Narayan Shukla and Benjamin Marlin

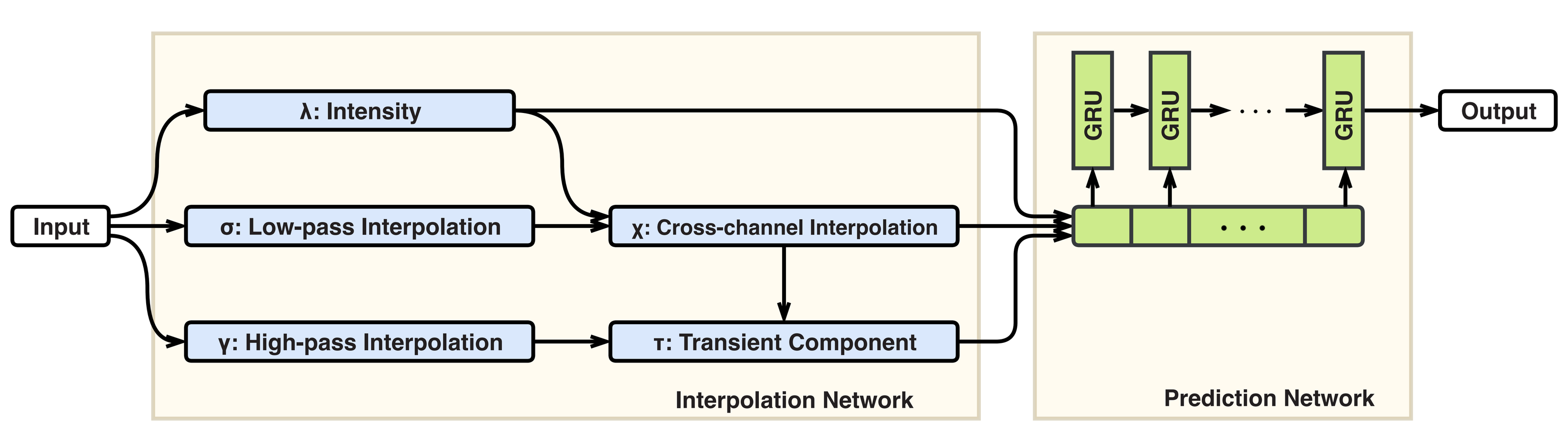
Introduction

Problem: We consider the problem of learning supervised machine learning models for sparse and irregularly sampled multivariate time series.

Setting:
$\mathcal{D} = \{(\mathbf{s}_i, y_i) | i = 1, ..., N\}$: dataset containing N data cases
$\mathbf{s}_{dn} = (\mathbf{t}_{dn}, \mathbf{x}_{dn})$: time series d for data case n as a tuple
$\mathbf{t}_{dn} = [t_{1dn}, ..., t_{L_{dn}dn}]$: list of time points at which observations are defined
$\mathbf{x}_{dn} = [x_{1dn}, ..., x_{L_{dn}dn}]$: corresponding list of observed values

Contributions:
1. We propose a novel method to handle irregularly sampled time series directly without any adhoc preprocessing.
2. Our approach also allows to compute an explicit multi-timescale representation of an irregularly sampled data.
3. The interpolation network serves the same purpose as the multivariate Gaussian process, but remove the restrictions associated with the need for a positive definite covariance matrix.
4. Our approach is fully modular as in any standard deep learning network for fixed length inputs can be used as prediction network.

Interpolation-Prediction Network



Interpolation Network interpolates the multivariate, sparse, and irregularly sampled input time series against a set of reference time points $\mathbf{r} = [r_1, ..., r_T]$.

$$Z(r, \mathbf{t}, \alpha) = \sum_{t \in \mathbf{t}} w(r, t, \alpha), \quad w(r, t, \alpha) = \exp(-\alpha(r - t)^2)$$

Intensity Function: $\lambda_{kd} = h_\theta^\lambda(r_k, \mathbf{t}_d, \mathbf{x}_d) = Z(r_k, \mathbf{t}_d, \alpha_d)$

Low-pass: $$\sigma_{kd} = h_\theta^\sigma(r_k, \mathbf{t}_d, \mathbf{x}_d) = \frac{1}{Z(r_k, \mathbf{t}_d, \alpha_d)} \sum_{j=1}^{L_{dn}} w(r_k, t_{jd}, \alpha_d)\, x_{jd}$$

High-pass: $$\gamma_{kd} = h_\theta^\gamma(r_k, \mathbf{t}_d, \mathbf{x}_d) = \frac{1}{Z(r_k, \mathbf{t}_d, \kappa\alpha_d)} \sum_{j=1}^{L_{dn}} w(r_k, t_{jd}, \kappa\alpha_d)\, x_{jd}$$

Cross-channel: $$\chi_{kd} = h_\theta^\chi(r_k, \mathbf{s}) = \frac{\sum_{d'} \rho_{dd'} \lambda_{kd'} \sigma_{kd'}}{\sum_{d'} \lambda_{kd'}}$$

Transients: $\tau_{kd} = h_\theta^\tau(r_k, \mathbf{s}) = \gamma_{kd} - \chi_{kd}$

Output of interpolation network :-
1. smooth, cross-channel interpolants χ_d to capture smooth trends
2. transient components τ_d to capture transients
3. intensity functions λ_d to capture information about where observations occur

Prediction Network can be any standard supervised neural network architecture such as fully-connected feedforward, convolutional or recurrent network.

Learning Objective

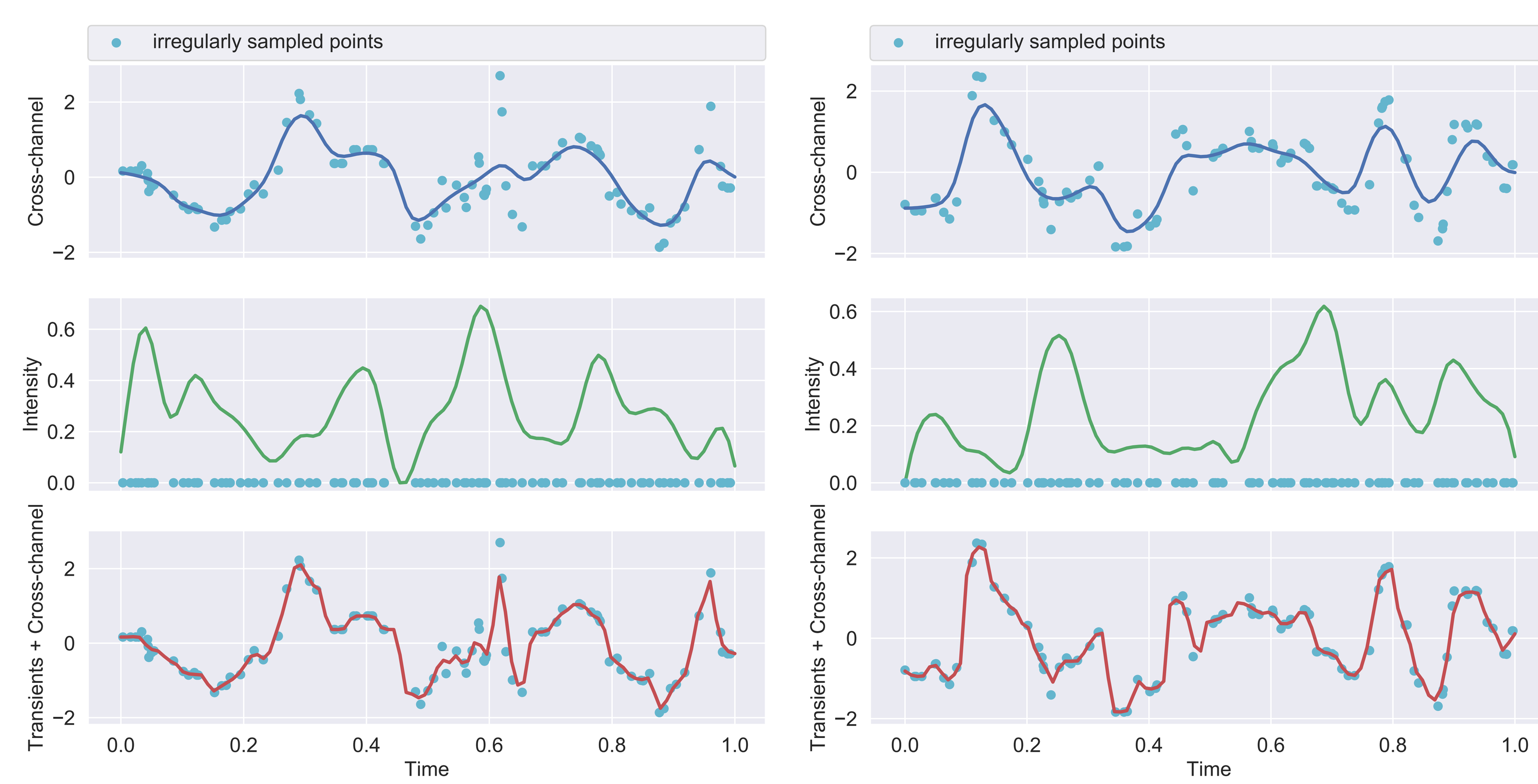
The parameters of the interpolation-prediction network are learned end-to-end via a composite objective function consisting of supervised and unsupervised components.

$$\theta_*, \phi_* = \underset{\theta, \phi}{\operatorname{argmin}} \sum_{n=1}^{N} \ell_P(y_n, g_\phi(f_\theta(\mathbf{s}_n))) + \delta_I \|\theta\|_2^2 + \delta_P \|\phi\|_2^2$$
$$+ \delta_R \sum_{n=1}^{N} \sum_{d=1}^{D} \sum_{j=1}^{L_{dn}} m_{jdn} \ell_I(x_{jdn}, h_\theta^\chi(t_{jdn}, (1 - \mathbf{m}_n) \odot \mathbf{s}_n))$$

where $\mathbf{m}$ denotes mask, y_n is the label for data case n, f and g are interpolation and prediction network respectively. We use mask $\mathbf{m}$ to hold out some observed data points during learning to compute the reconstruction loss.

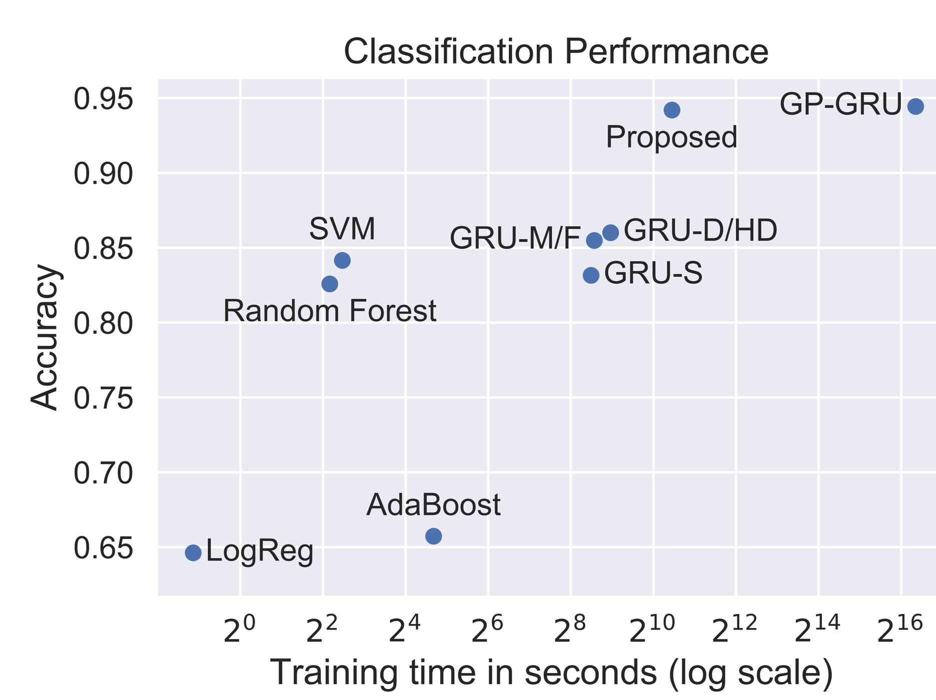
Quality of Interpolation

Output of interpolation network on UWave dataset:



Experiments: Univariate Data

We use the UWaveGesture dataset to assess the training time and classification performance of our model.

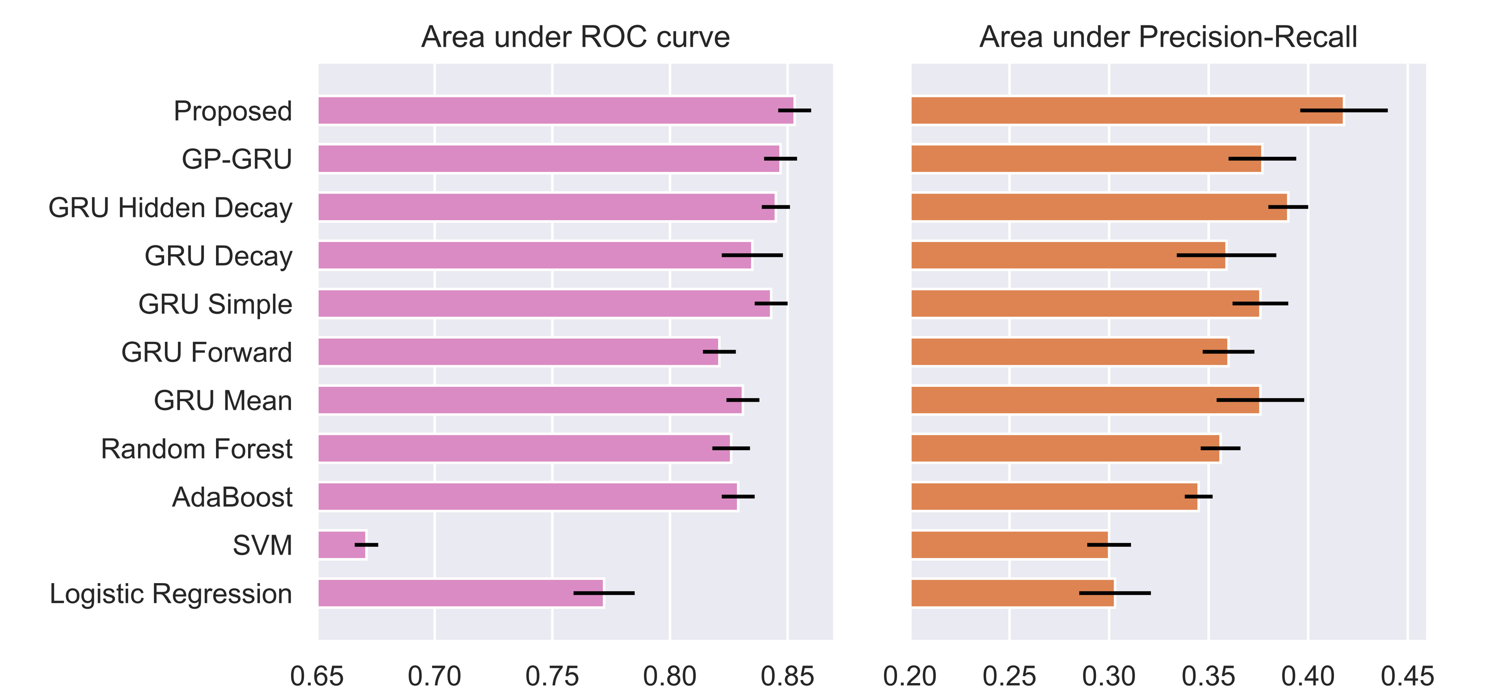


Code is available at `https://github.com/mlds-lab/interp-net`

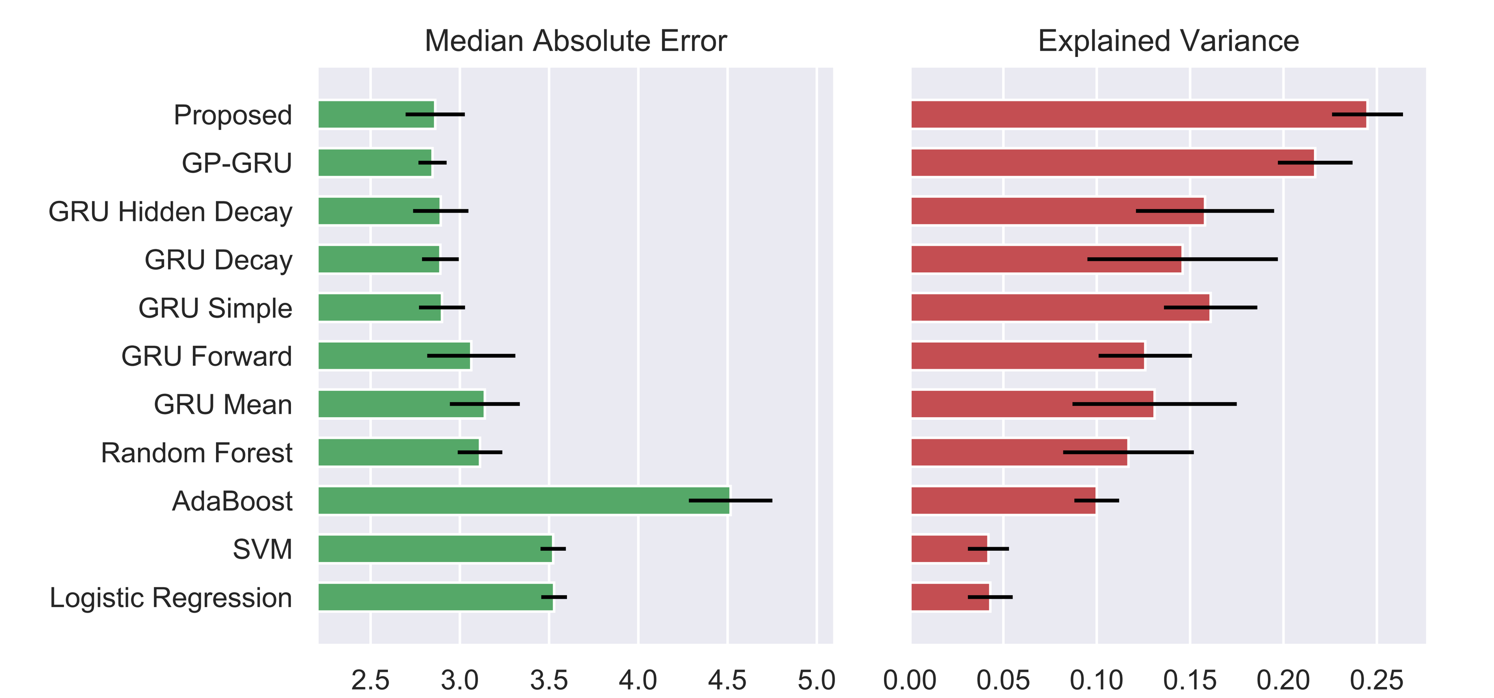
Experiments: Multivariate Data

Our experiments are based on the publicly available MIMIC-III dataset.

Classification: In-hospital mortality prediction

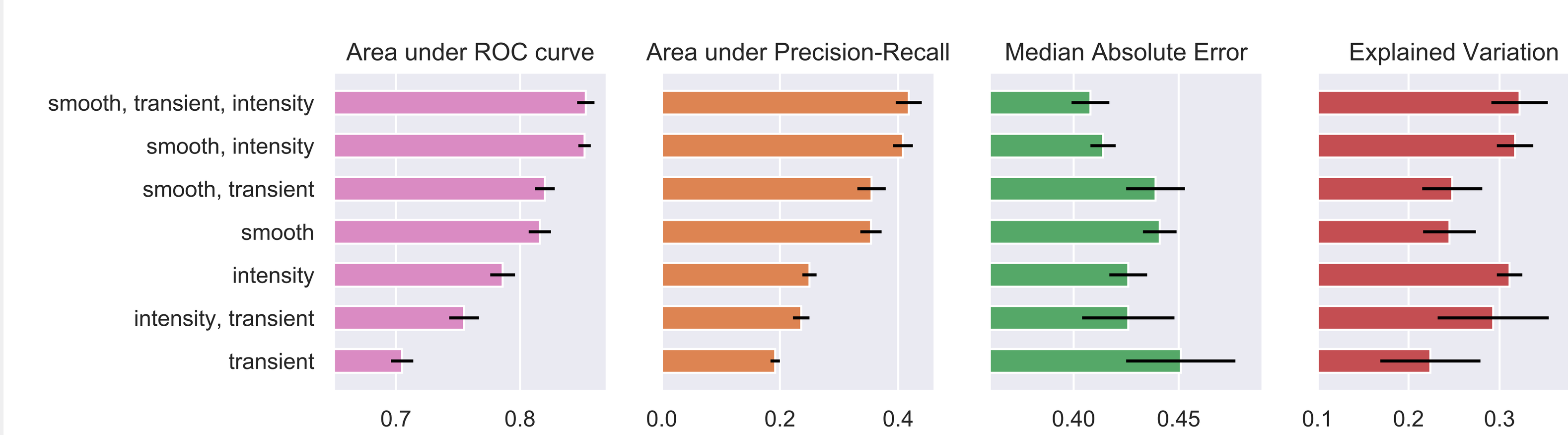


Regression: Length of stay prediction



Ablation Study

We conduct a set of ablation experiments of the internal structure of the interpolation network using all subsets of outputs.



Dataset

We use the first 48 hours of data for the prediction tasks.

Table: Features extracted from MIMIC III for our experiments

feature	#Missing	Sampling Rate	feature	#Missing	Sampling Rate
SpO2	31.35%	0.80	TGCS	87.94%	0.14
HR	23.23%	0.90	CRR	95.08%	0.06
RR	59.48%	0.48	UO	82.47%	0.20
SBP	49.76%	0.59	FiO2	94.82%	0.06
DBP	48.73%	0.60	Glucose	91.47%	0.10
Temp	83.80%	0.19	pH	96.25%	0.04

{snshukla, marlin}@cs.umass.edu