On Principles, Models and Methods for Learning from Irregularly Sampled Time Series:

From Discretization to Attention and Invariance

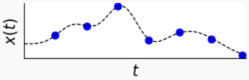
Satya Narayan Shukla, Benjamin Marlin

University of Massachusetts Amherst

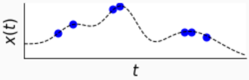
Motivation

Irregularly sampled time series commonly occur in several domains such as healthcare, climate science, ecology, astronomy, biology and others.

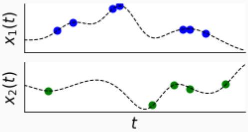
Irregularly Sampled Time Series



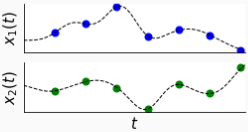
Univariate regularly sampled



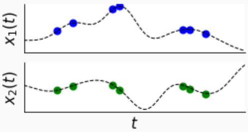
Univarite irregulalry sampled



Multivariate irregularly sampled (unaligned)



Multivariate regularly sampled



Multivariate irregularly sampled (aligned)

Challenges

- Irregular spacing between observation time points
- Different data cases may have different numbers of observations
- Lack of alignment of observation time points
- Most machine learning models typically assume fully-observed, fixed-size feature representations

Categorization

We survey 40 papers and categorize methods in terms of:

Categorization

We survey 40 papers and categorize methods in terms of:

- Data Representation
    - Series-based
    - Vector-based
    - Set-based

Categorization

We survey 40 papers and categorize methods in terms of:

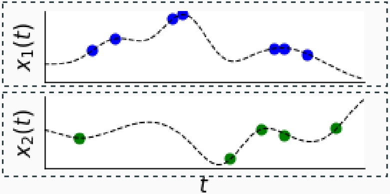
- Data Representation
  - Series-based
  - Vector-based
  - Set-based
- Inference Tasks
  - Detection
  - Prediction
  - Filtering
  - Smoothing
  - Interpolation
  - Forecasting

Categorization

We survey 40 papers and categorize methods in terms of:

- Data Representation
  - Series-based
  - Vector-based
  - Set-based
- Inference Tasks
  - Detection
  - Prediction
  - Filtering
  - Smoothing
  - Interpolation
  - Forecasting
- Modeling Primitives
  - Discretization
  - Interpolation
  - Recurrence
  - Attention
  - Invariance

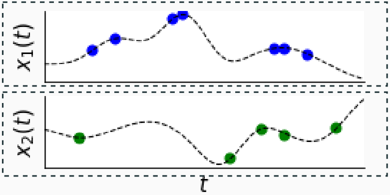
Series-based



Multivariate irregularly sampled (unaligned)

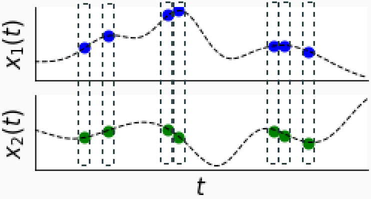
Data Representation

Series-based

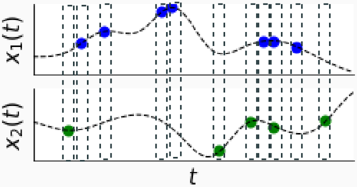


Multivariate irregularly sampled (unaligned)

Vector-based



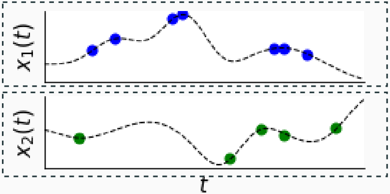
Multivariate irregularly sampled (aligned)



Multivariate irregularly sampled (unaligned)

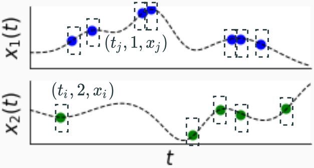
Data Representation

Series-based



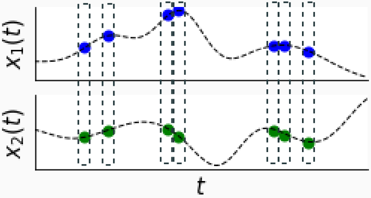
Multivariate irregularly sampled (unaligned)

Set-based

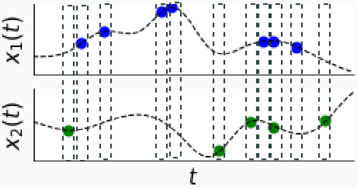


Multivariate irregularly sampled (unaligned)

Vector-based

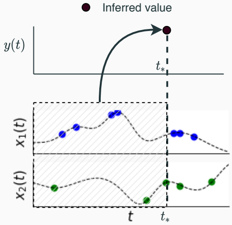


Multivariate irregularly sampled (aligned)



Multivariate irregularly sampled (unaligned)

Figure 1: Detection



Inference Tasks

Figure 1: Detection

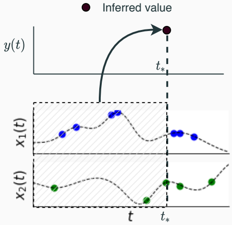
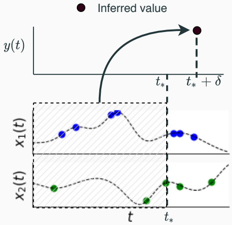


Figure 2: Prediction



Inference Tasks

Figure 1: Detection

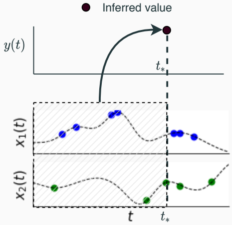


Figure 2: Prediction

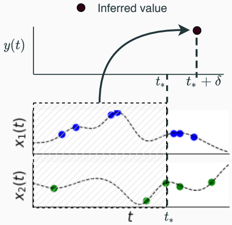


Figure 3: Filtering

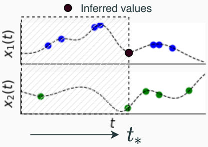
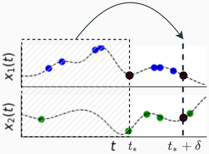


Figure 4: Forecasting



Inference Tasks

Figure 1: Detection

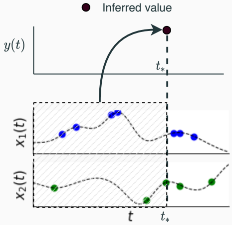


Figure 2: Prediction

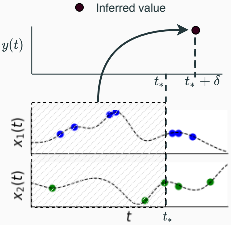


Figure 3: Filtering

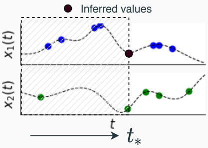


Figure 4: Forecasting

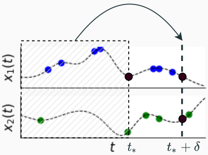
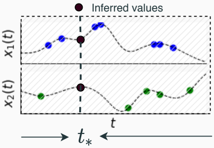


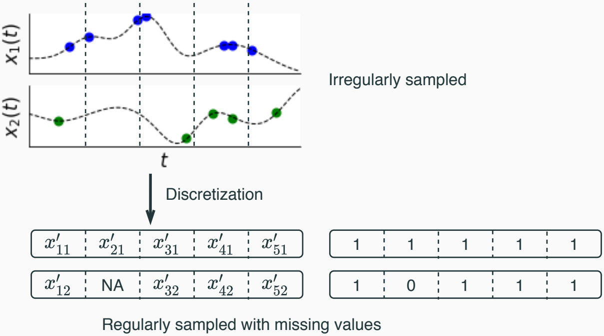
Figure 5: Interpolation



Modeling Primitives

Discretization

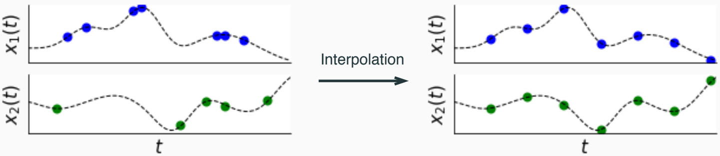
Reduction from irregularly sampled time series to regularly sampled time series that may contain missing data (Lipton et al. 2016)[1].



- Aggregation and Imputation methods required
- Leads to information loss

[1] Zachary C Lipton, David Kale, and Randall Wetzel. Directly modeling missing data in sequences with rnns: Improved classification of clinical time series. In Machine Learning for Healthcare Conference, pages 253–270, 2016.

Interpolation



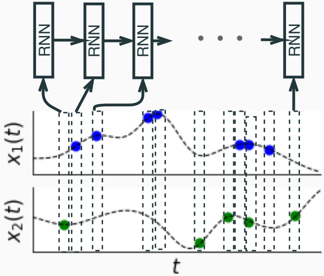
- **Deterministic Interpolation** (Kernel Smoother)
  - Shukla and Marlin (2019)[2]
- **Probabilistic Interpolation** (Gaussian Process Regression)
  - Li and Marlin (2016)[3]
- **Similarity kernel between irregularly sampled time series**
  - Lu et al. (2008)[4]

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

[3] Steven Cheng-Xian Li and Benjamin M Marlin. A scalable end-to-end gaussian process adapter for irregularly sampled time series classification. In Advances In Neural Information Processing Systems, pages 1804–1812, 2016.

[4] Zhengdong Lu, Todd K. Leen, Yonghong Huang, and Deniz Erdogmus. A reproducingkernel hilbert space framework for pairwise time series distances. In Proceedings of the25th International Conference on Machine Learning, ICML '08, pages 624–631, NewYork, NY, USA, 2008.

Recurrence

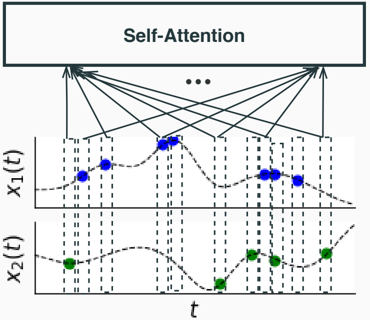


Multivariate irregularly sampled (unaligned)

- Append the time points or inter-observation intervals to the vector-valued observations (Che et al. 2018)[5]
- Using Ordinary Differential Equations to evolve the hidden state between continuous time observations (Rubanova et al. 2019)[6]

[5] Zhengping Che, Sanjay Purushotham, Kyunghyun Cho, David Sontag, and Yan Liu. Recurrent neural networks for multivariate time series with missing values. Scientific Reports, 8(1):6085, 2018.

[6] Yulia Rubanova, Ricky T. Q. Chen, and David K Duvenaud. Latent ordinary differential equations for irregularly-sampled time series. In Advances in Neural Information Processing Systems 32, pages 5320–5330, 2019.



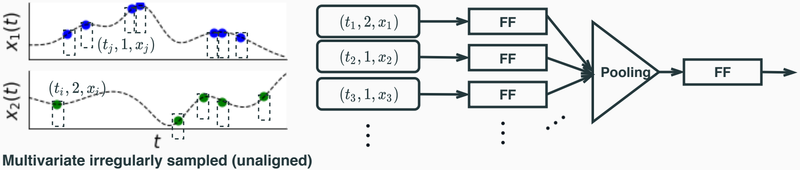
Multivariate irregularly sampled (unaligned)

$$\mathrm{Attn}(Q, K, V) = \mathrm{softmax}\left(\frac{Q\,K^T}{\sqrt{C}}\right)V$$

- Time embedding can be used to deal with irregular sampling (Horn et al. 2020)[7]
- Imputation required to deal with missing values

[7] Max Horn, Michael Moor, Christian Bock, Bastian Rieck, and Karsten Borgwardt. Set functions for time series. In Proceedings of the 25th International Conference on Machine Learning, 2020.

Structural Invariance



$$\mathbf{h} = g_\phi(\text{pool}(\{f_\theta(t_{in}, d_{in}, x_{in}) | 1 \leq i \leq L_n\}))$$

- Ordering of observations not required
- Supports variable length sequences, partially observed vectors and irregular intervals between observations (Horn et al. 2020)[8]
- Avoids discretization, imputation and interpolation

[8] Max Horn, Michael Moor, Christian Bock, Bastian Rieck, and Karsten Borgwardt. Set functions for time series. In Proceedings of the 25th International Conference on Machine Learning, 2020.

Thank You!