Dynamic RNN State Updates for Irregular Multivariate Time Series Classification

Tom Hartvigsen, Cansu Sen, Xiangnan Kong, Elke Rundensteiner

Data Science, Worcester Polytechnic Institute

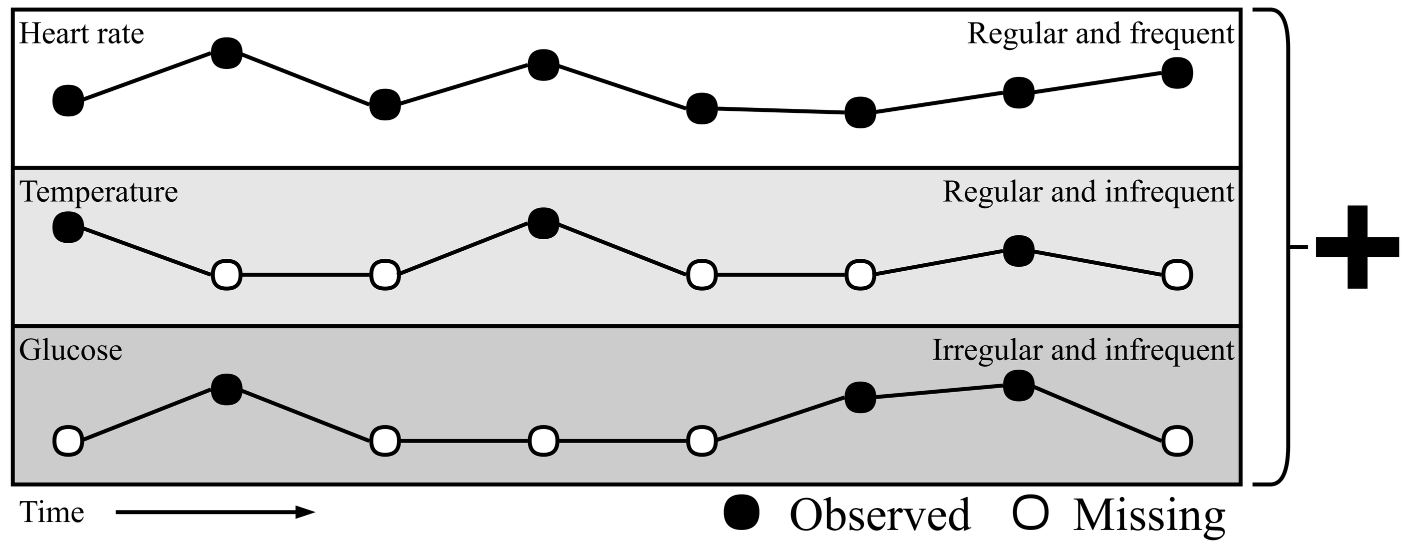
Irregular Multivariate Time Series

Variables have different sampling rates. Empty spaces between observations can be informative. Prior to classification, timestamps must be aligned.

Three common challenges with such data:

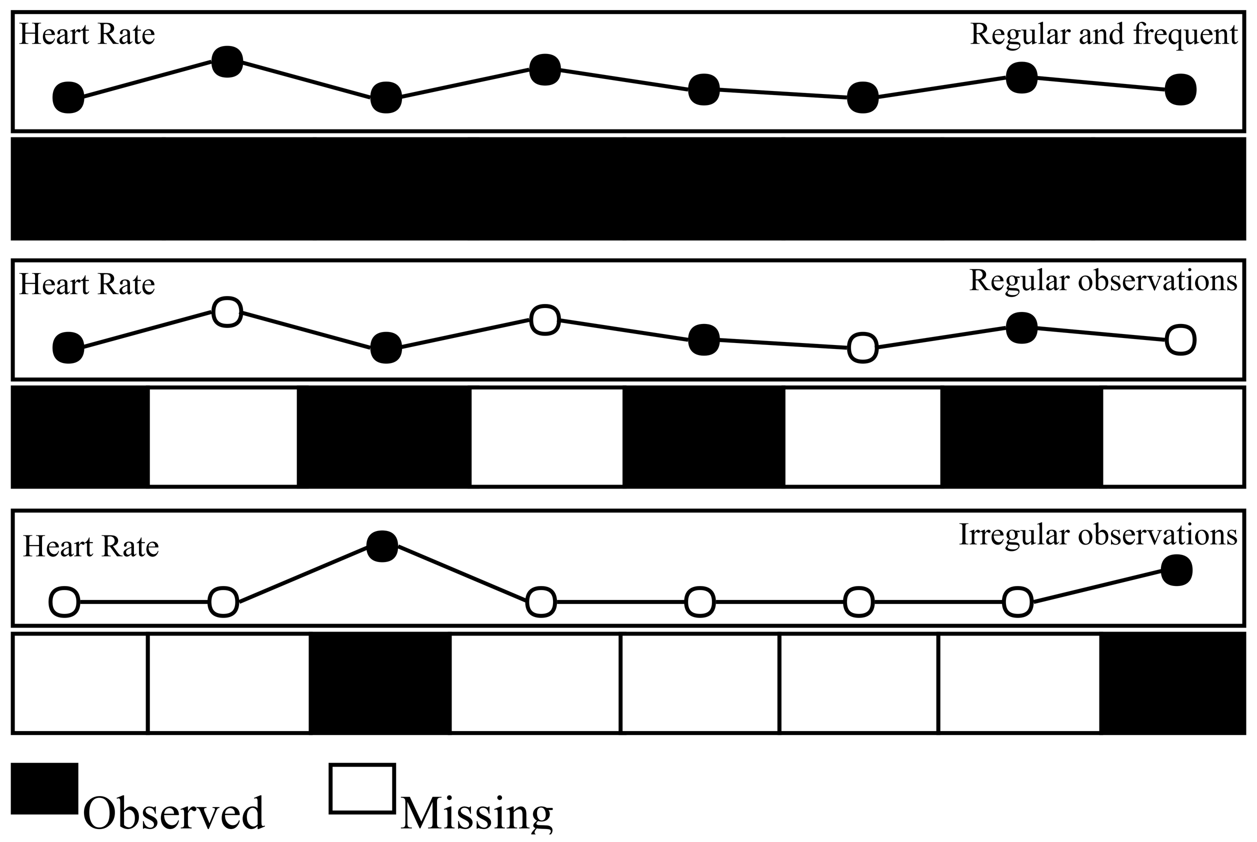
- Irregular observation rates require time-alignment.
- Missing values may or may not be informative.
- High sampling-rate lead to long sequences. As dependencies grow over time, it becomes more challenging to remember only useful information.

Problem Setting



Traditional approaches

- Up-sampling slow variables vs. down-sampling fast variables.
- Combinations of different information sources (time delta, masking vector) are fed to RNNs.
- Some approaches also try to find meaningful values to impute.



Missingness-Informed State-Skipping RNN

Intuition: Update the cell memory only when useful information is observed.

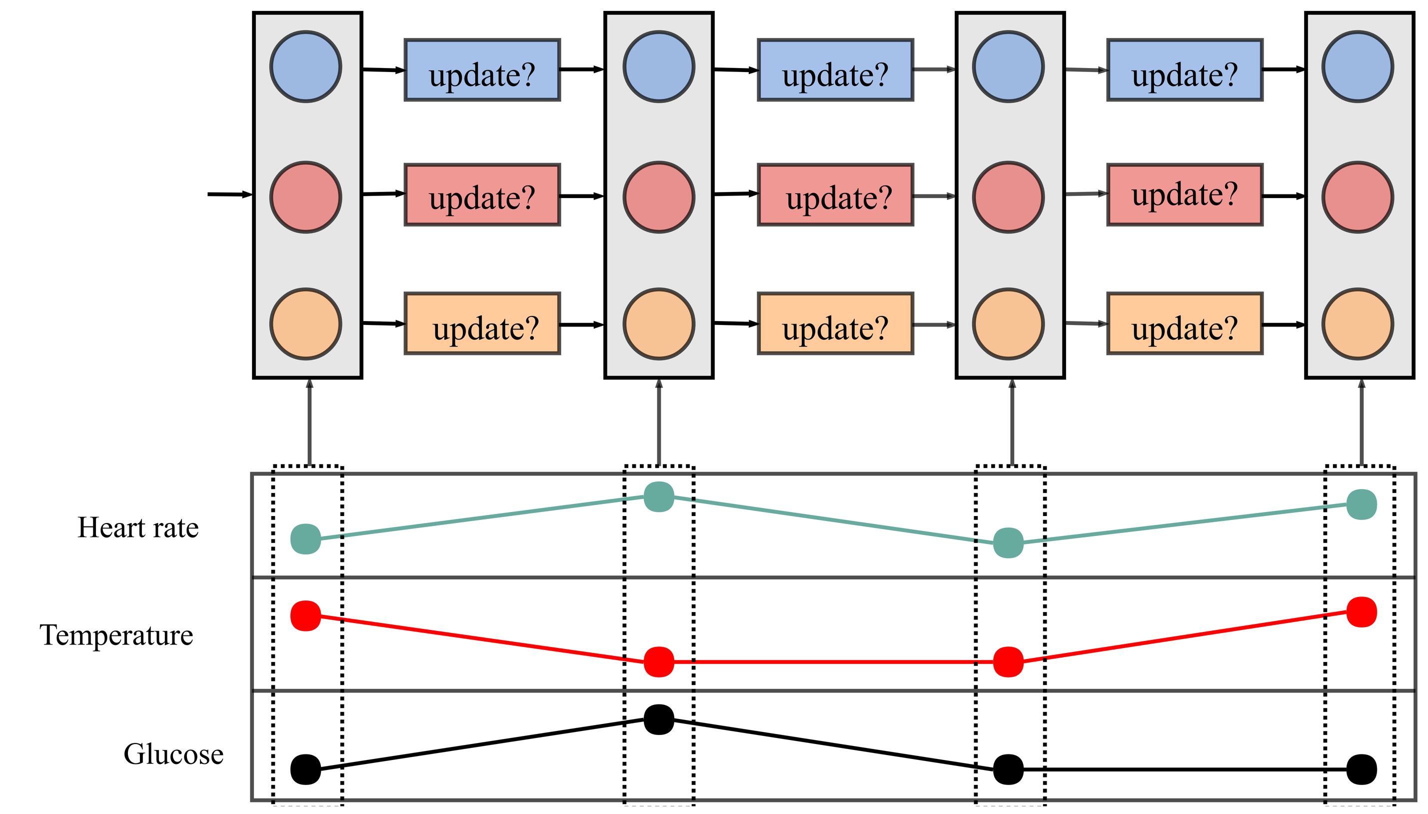


Figure 1: Mechanics of partial state-updates.

Representation generation in MISS-RNN:

$$\begin{aligned}
\gamma_t &= -\exp\{-max(0, W_\gamma \cdot \delta_t + b_\gamma)\} \\
\hat{x}_t &= m_t x_t + (1-m_t)(\gamma_{x_t} x_{t'} + (1-\gamma_{x_t})\tilde{x}) \\
\hat{S}_{t-1} &= \gamma_{S_t} * S_{t-1} \\
\tilde{S}_t &= \mathrm{GRU}([\hat{x}_t, m_t], \hat{S}_{t-1}) \\
u_t &= \mathrm{binarize}(\sigma(W_s \cdot \tilde{S}_t + b_s)), \text{ s.t. } u_t \in \mathbb{R}^h \\
S_t &= u_t \odot \tilde{S}_t + (1-u_t) \odot S_{t-1}
\end{aligned}$$

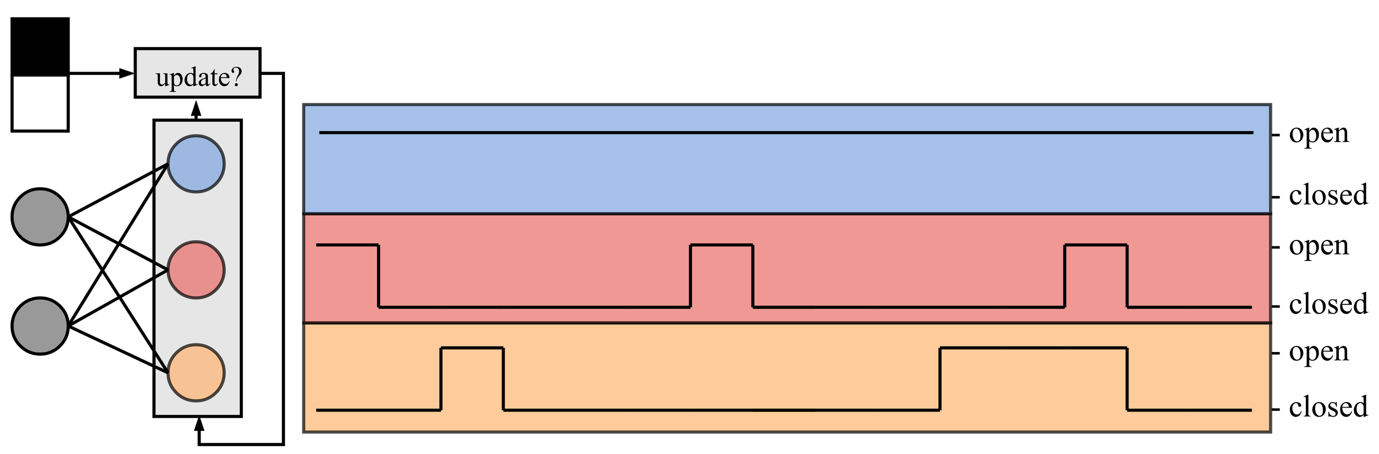


Figure 2: Resulting dynamics in hidden state updates patterns.

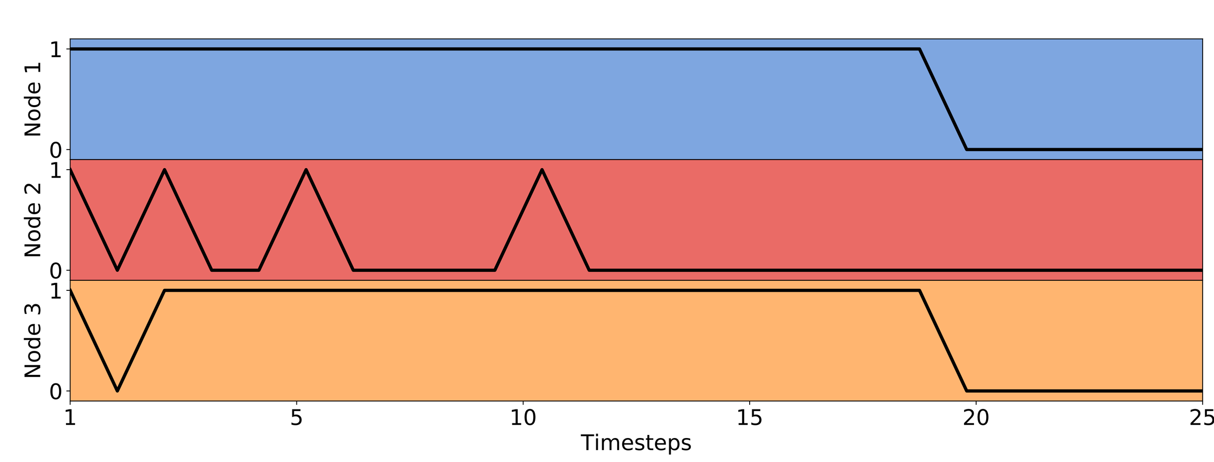


Figure 3: Disentangled updates in sub-representations.

Evaluation Data

Synthetic dataset:

- Balanced dataset of 1000 time series, each is 100 timesteps long.
- Each time series is a sequence of 0's. Positive examples have a 1 at a random location. Gaussian noise is added.
- Remove 2 values from uniform locations from positive examples, remove 4 from negative.
- Impute surrogate values for missing values.

Results

Method	Accuracy
Mean Imputation	59 ± 0.0
Zero Imputation	59 ± 0.0
Forward Imputation	59 ± 0.0
SkipRNN [1]	59.0 ± 0.0
GRU-D [2]	80.24 ± 22.26
Mask + Mean Imp. [3]	72.4 ± 26.66
PhasedLSTM [4]	59 ± 0.0
Mask and diff input	72.40 ± 26.66
SkipRNN + Mask	71.6 ± 22.83
Mask as updates	59 ± 0.0
Input mask and full-skip	81.5 ± 20.95
Mask informs skipping	$59 \pm + 0.0$
Input-mask + mask-inform	83.6 ± 22.0
Full-skipping w/decay impute	91.6 ± 5.04
MISS [proposed]	$\mathbf{94.72 \pm 2.69}$

References

[1] V. Campos, B. Jou, X. Giro-i Nieto, J. Torres, and S.-F. Chang, "Skip rnn: Learning to skip state updates in recurrent neural networks," in *International Conference on Learning Representations*, 2018.

[2] Z. Che, S. Purushotham, K. Cho, D. Sontag, and Y. Liu, "Recurrent neural networks for multivariate time series with missing values," *Scientific reports*, vol. 8, no. 1, p. 6085, 2018.

[3] Z. C. Lipton, D. C. Kale, and R. Wetzel, "Modeling missing data in clinical time series with rnns," in *Machine Learning for Healthcare*, 2016.

[4] D. Neil, M. Pfeiffer, and S.-C. Liu, "Phased lstm: Accelerating recurrent network training for long or event-based sequences," in *Advances in Neural Information Processing Systems*, pp. 3882–3890, 2016.

Acknowledgements

Thank you to the U.S. Dept. of Ed. for sponsoring this research and to the DSRG at WPI for an enriching research community.