Learning Representations for Time Series in Healthcare

Topics in Machine Learning for Healthcare CSC2541 - guest lecture

Sana Tonekaboni

UNIVERSITY OF TORONTO

Time Series in Healthcare

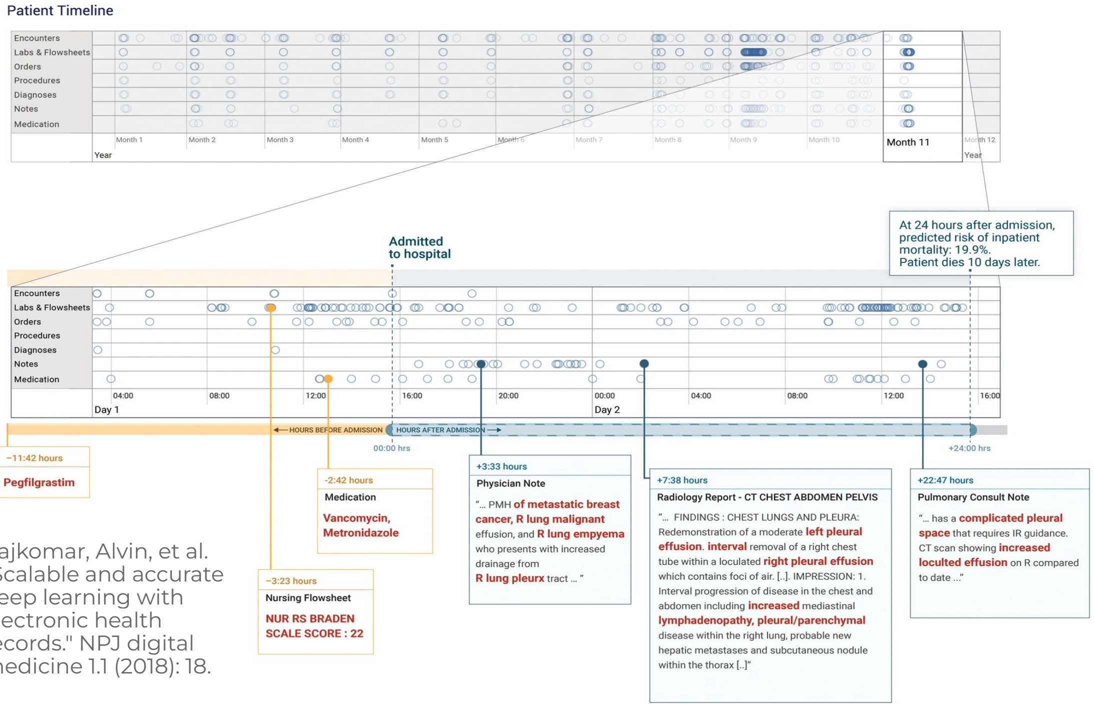
Wearables

Physiological signals

Patient records

Time series is a data modality rich in information that captures historical context and trajectory of events over time.

Challenges with real-world Time Series



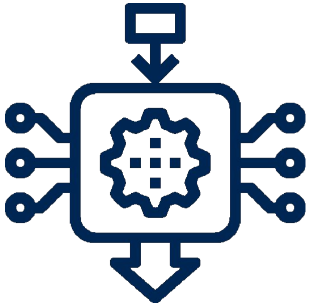
Rajkomar, Alvin, et al. "Scalable and accurate deep learning with electronic health records." NPJ digital medicine 1.1 (2018): 18.

- Irregular measurements
- Multi-modality
- Lack/quality of labels
- Multiple confounders
- Missing measurements
- ...

The complexity of time series becomes a barrier to building generalizable ML models.

UNIVERSITY OF TORONTO

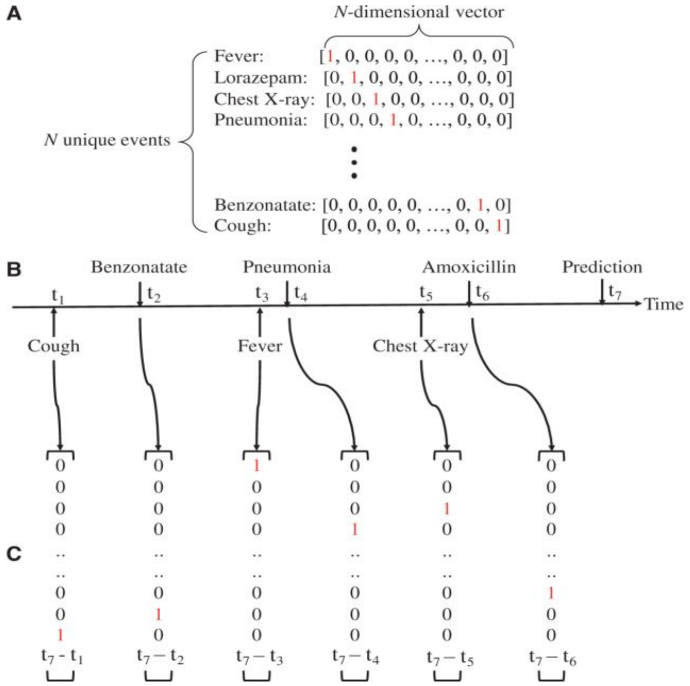
Representation learning for Time series



Representation Learning

Create representations of the data that summarizes key characteristics of the data for machine learning models

- **Embedding**
- Feature Engineering
- Task-specific
- Self-supervised



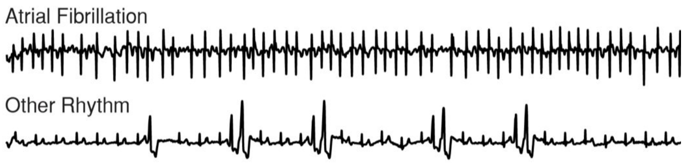
Choi, E., Schuetz, A., Stewart, W. F., & Sun, J. (2017). Using recurrent neural network models for early detection of heart failure onset. Journal of the American Medical Informatics Association, 24(2), 361-370.

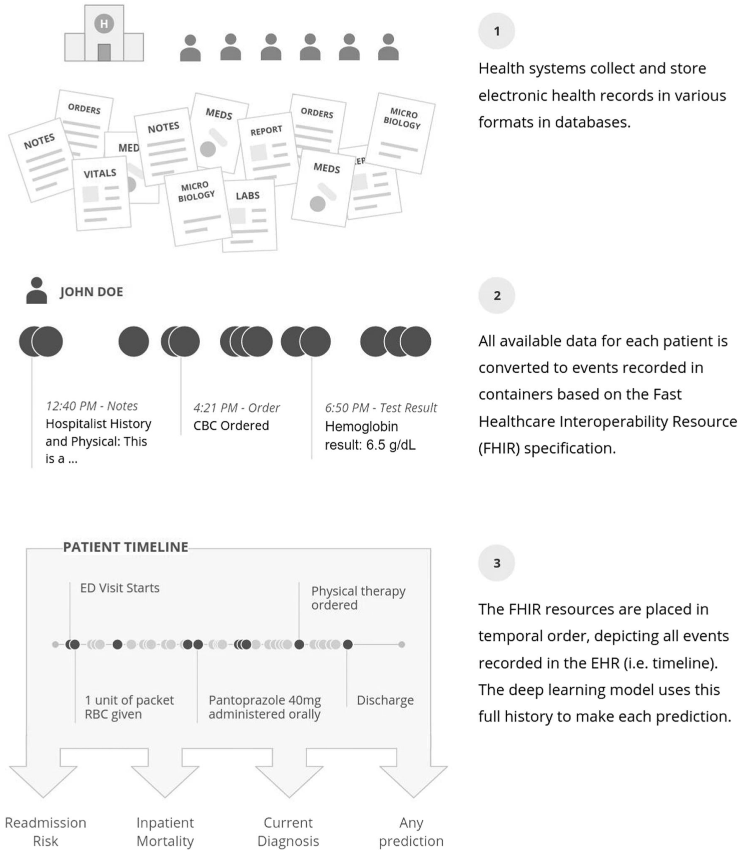
UNIVERSITY OF TORONTO

Representation Learning

Create representations of the data that summarizes key characteristics of the data for machine learning models

- Embedding
- **Feature Engineering**
- Task-specific
- Self-supervised



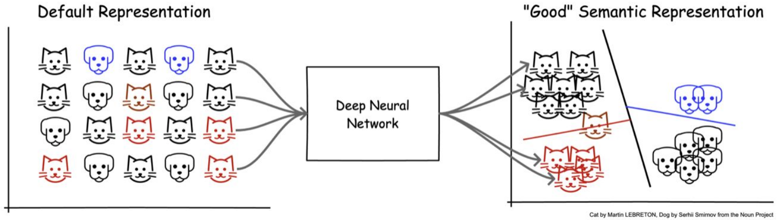


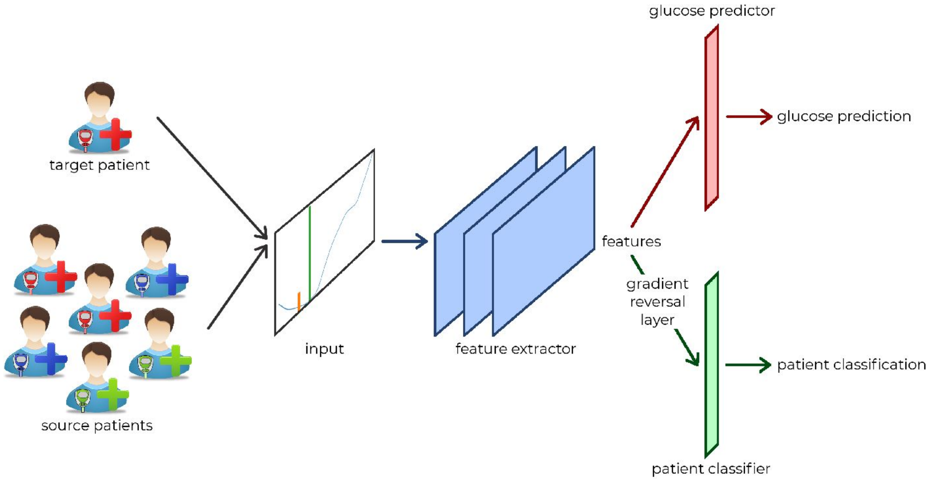
Rajkomar, Alvin, et al. "Scalable and accurate deep learning with electronic health records." NPJ digital medicine 1.1 (2018): 18.

Representation Learning

Create representations of the data that summarizes key characteristics of the data for machine learning models

- Embedding
- Feature Engineering
- **Task-specific**
- Self-supervised





De Bois, M., El Yacoubi, M. A., & Ammi, M. (2021). Adversarial multi-source transfer learning in healthcare: Application to glucose prediction for diabetic people. *Computer Methods and Programs in Biomedicine*, *199*, 105874.

UNIVERSITY OF TORONTO

Representation Learning

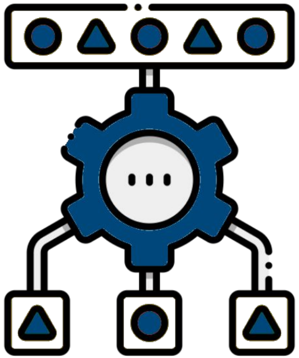
Create representations of the data that summarizes key characteristics of the data for machine learning models

- Embedding
- Feature Engineering
- Task-specific
- **Self-supervised**

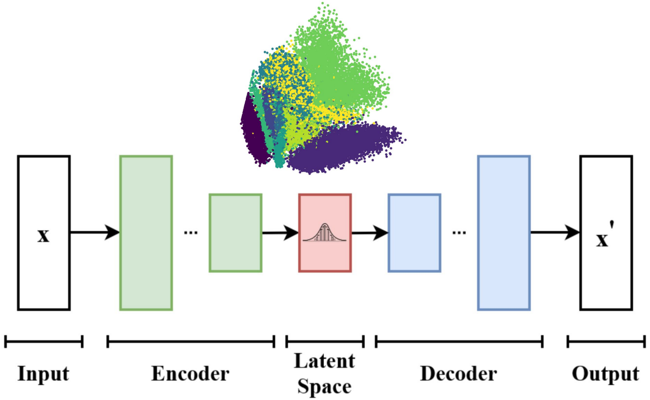
- High-dimensional/high-frequency
- Limited prior knowledge
- Expensive to label over time
- Non-stationary with multiple confounders

UNIVERSITY OF TORONTO

Self-Supervised Representation learning for Time series

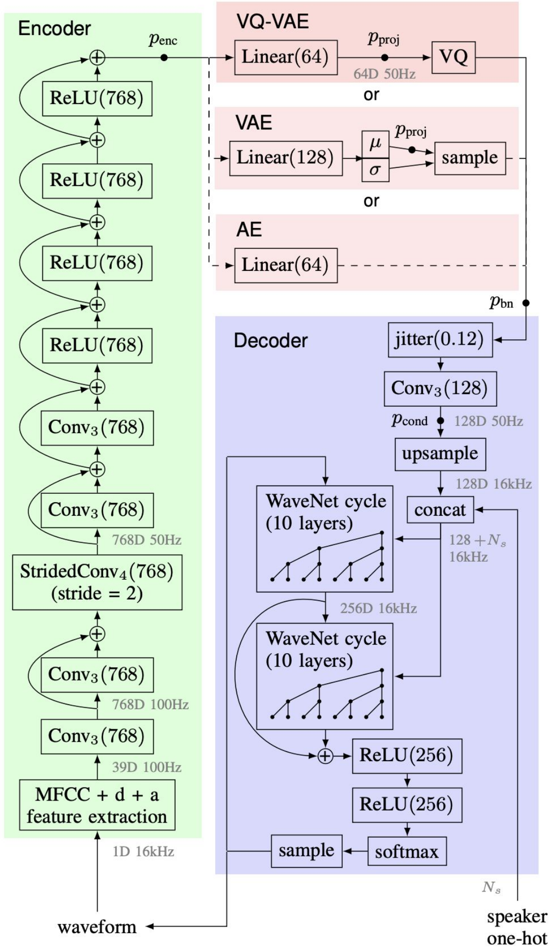


Reconstruction-based methods



ELBO: $$\mathcal{L}(\lambda) \triangleq \mathrm{E}_{q_\lambda(z)}[\log p(x,z) - \log q(z)]$$

Chorowski, J., Weiss, R. J., Bengio, S., & Van Den Oord, A. (2019). **Unsupervised speech representation learning using wavenet autoencoders**. *IEEE/ACM transactions on audio, speech, and language processing*, *27*(12), 2041-2053.

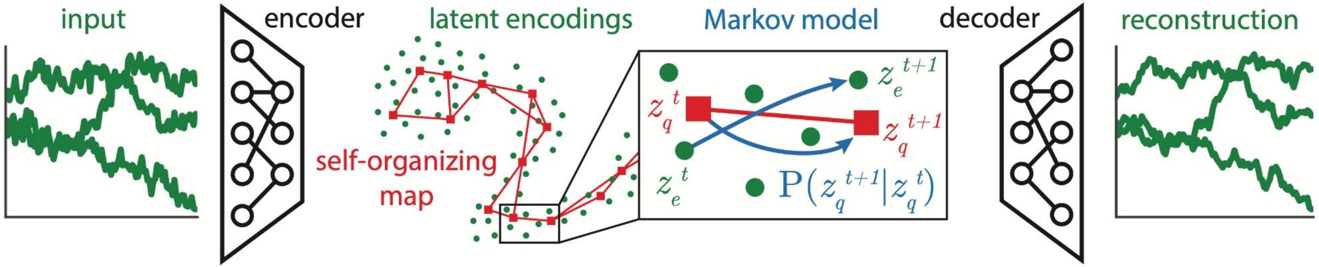


UNIVERSITY OF TORONTO

Reconstruction-based methods

SOM-VAE

Enforces topological structure in a lower dimensional space through Self-Organizing Maps such that the representations retain their smoothness in that space

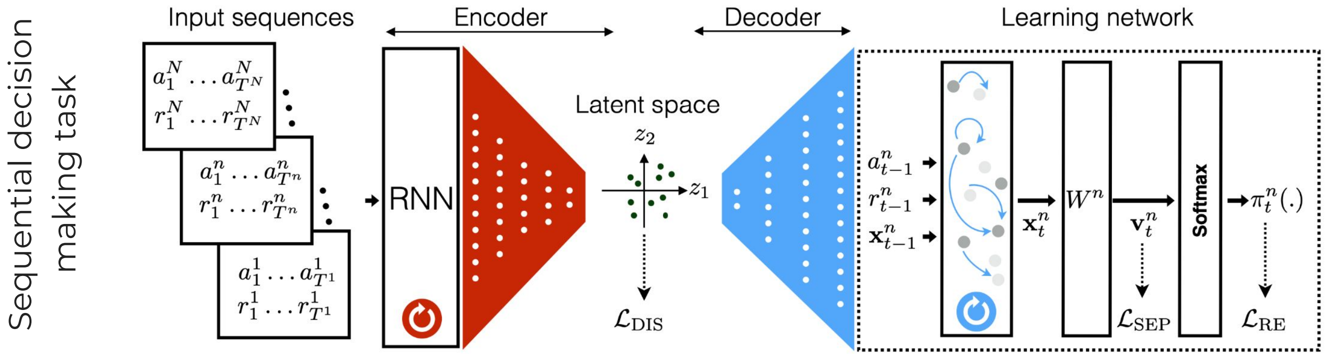


$$\mathcal{L}_{\text{SOM-VAE}}(x, \hat{x}_q, \hat{x}_e) = \mathcal{L}_{\text{reconstruction}}(x, \hat{x}_q, \hat{x}_e) + \alpha\, \mathcal{L}_{\text{commitment}}(x) + \beta\, \mathcal{L}_{\text{SOM}}(x)$$

$$\mathcal{L}(x^{t-1}, x^t, \hat{x}_q^t, \hat{x}_e^t) = \mathcal{L}_{\text{SOM-VAE}}(x^t, \hat{x}_q^t, \hat{x}_e^t) + \gamma\, \mathcal{L}_{\text{transitions}}(x^{t-1}, x^t) + \tau\, \mathcal{L}_{\text{smoothness}}(x^{t-1}, x^t)$$

Fortuin, V., Hüser, M., Locatello, F., Strathmann, H., & Rätsch, G. (2018, September). **SOM-VAE: Interpretable Discrete Representation Learning on Time Series**. In *International Conference on Learning Representations*.

Reconstruction-based methods



$$\mathcal{L} = \mathcal{L}_{\text{RE}} + \lambda_2 \mathcal{L}_{\text{DIS}} + \lambda_3 \mathcal{L}_{\text{SEP}}$$

$$\mathcal{L}_{\text{RE}} = -\frac{1}{N}\sum_{n=1}^{N}\sum_{t=1}^{T^n} \log \pi_t^n(a_t^n; a_{1\ldots t-1}^n, r_{1\ldots t-1}^n)$$ Reconstruction

$$\mathcal{L}_{\text{DIS}} = \lambda_1 \text{MMD}(\hat{q}(\mathbf{z}), p(\mathbf{z})) + \text{KL}(g(\mathbf{z}) \| p(\mathbf{z}))$$ Disentanglement

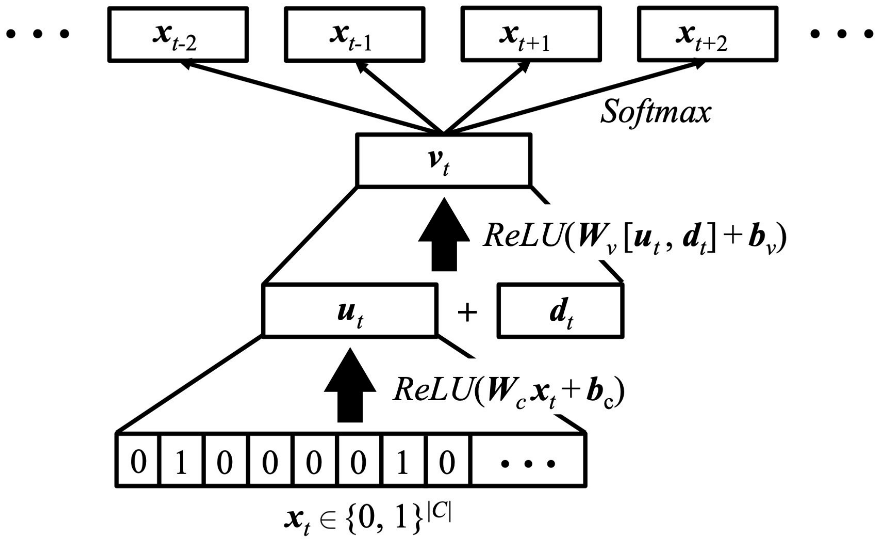
$$\hat{\mathcal{L}}_{\text{SEP}} = \frac{1}{N}\sum_{n=1}^{N}\sum_{t=1}^{T^n} \left| \frac{\partial^2 u_t^n}{\partial z_1 \partial z_2} \right|$$ Separation

Dezfouli, A., Ashtiani, H., Ghattas, O., Nock, R., Dayan, P., & Ong, C. S. (2019). **Disentangled behavioural representations.** *Advances in neural information processing systems, 32*

Predictive methods

Med2Vec

Trains an encoder neural network to learn representations that are predictive of future medical codes of visits within a context window.



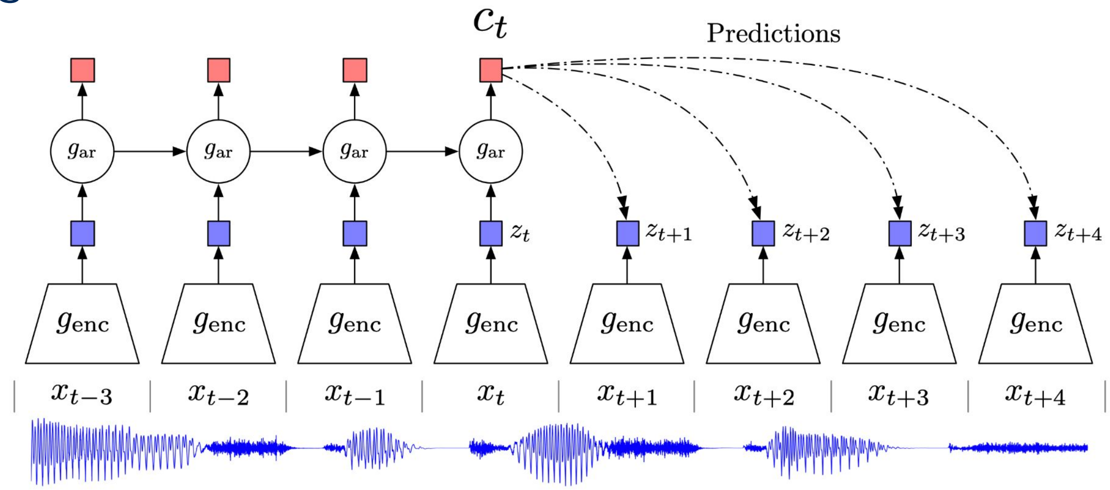
Choi, E., Bahadori, M. T., Searles, E., Coffey, C., Thompson, M., Bost, J., ... & Sun, J. (2016, August). **Multi-layer representation learning for medical concepts.** In *proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1495-1504).

Predictive/contrastive methods

Contrastive Predictive Coding

$$f_k(x_{t+k}, c_t) = \exp\left(z_{t+k}^T W_k c_t\right)$$

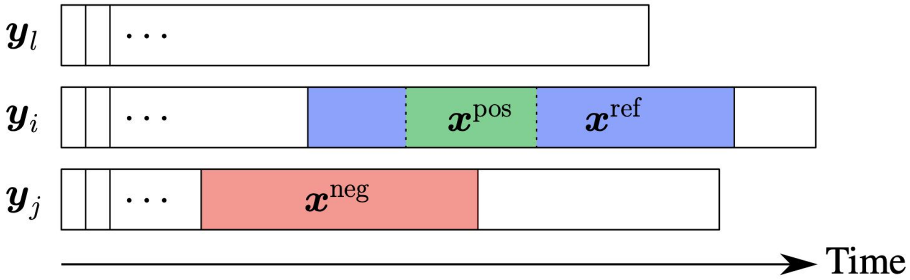
$$\underbrace{\mathcal{L}_\mathrm{N} = -\mathop{\mathbb{E}}_{X}\left[\log \frac{f_k(x_{t+k}, c_t)}{\sum_{x_j \in X} f_k(x_j, c_t)}\right]}_{\text{Contrastive Loss}}$$



Oord, A. V. D., Li, Y., & Vinyals, O. (2018). **Representation learning with contrastive predictive coding**. *arXiv preprint arXiv:1807.03748.*

Contrastive methods

Representation learning with triplet loss

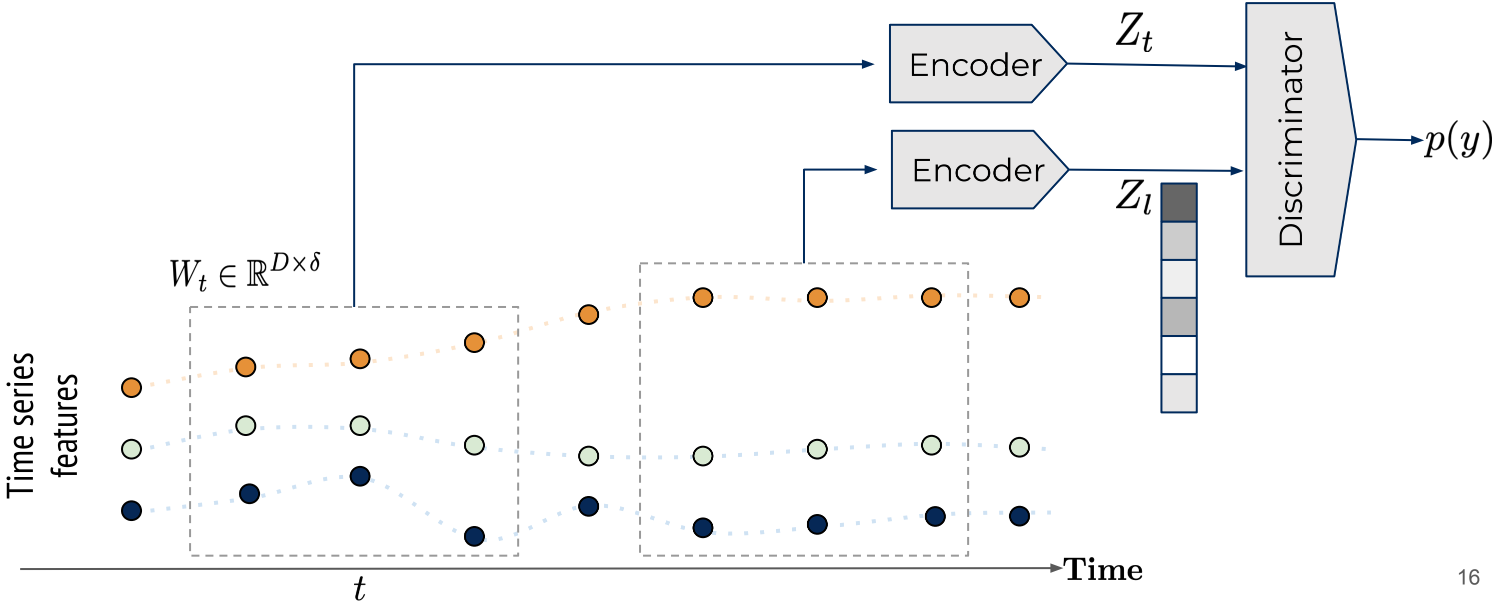


$$\underbrace{-\log\left(\sigma\left(\boldsymbol{f}(\boldsymbol{x}^{\mathrm{ref}},\boldsymbol{\theta})^\top \boldsymbol{f}(\boldsymbol{x}^{\mathrm{pos}},\boldsymbol{\theta})\right)\right)-\sum_{k=1}^{K}\log\left(\sigma\left(-\boldsymbol{f}(\boldsymbol{x}^{\mathrm{ref}},\boldsymbol{\theta})^\top \boldsymbol{f}(\boldsymbol{x}_k^{\mathrm{neg}},\boldsymbol{\theta})\right)\right)}_{\text{Triplet Loss}}$$

Franceschi, J. Y., Dieuleveut, A., & Jaggi, M. (2019). **Unsupervised scalable representation learning for multivariate time series**. *Advances in neural information processing systems, 32.*

Temporal Neighborhood Coding (TNC)

TNC is a self-supervised framework for learning representations of multivariate, non-stationary time series.

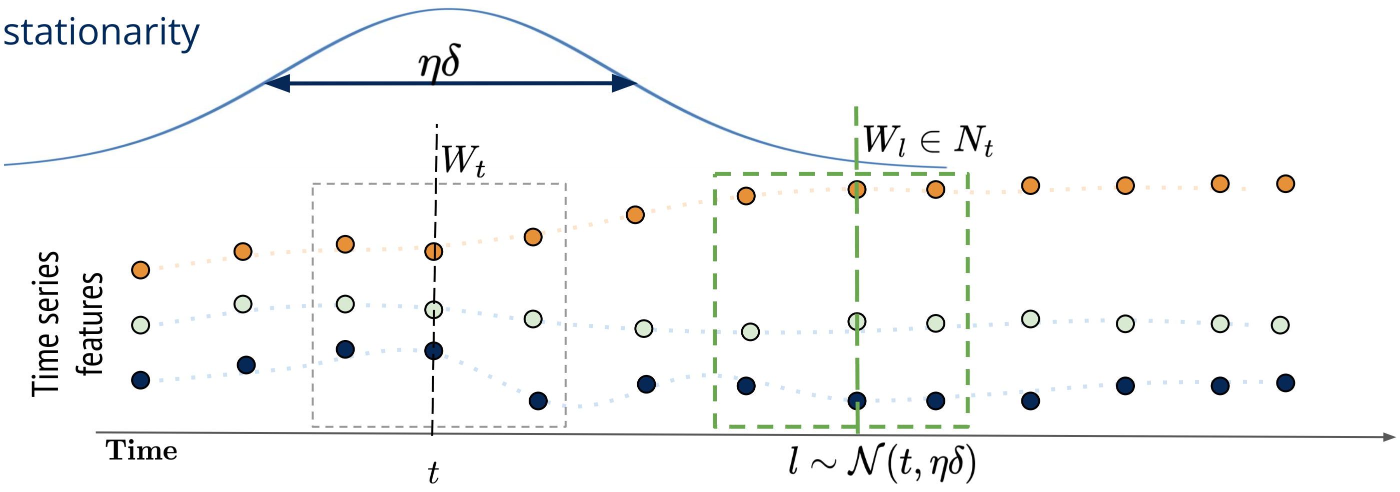


S. Tonekaboni, D. Eytan, A. Goldenberg. ***Unsupervised Representation Learning for Time Series with Temporal Neighborhood Coding***. In *International Conference on Learning Representations (ICLR 2020).*

Temporal Neighborhood

Temporal neighborhood N_t of window W_t is the set of windows with centroids $l \sim \mathcal{N}(t, \eta\delta)$ representing the region of time series with similar underlying state.

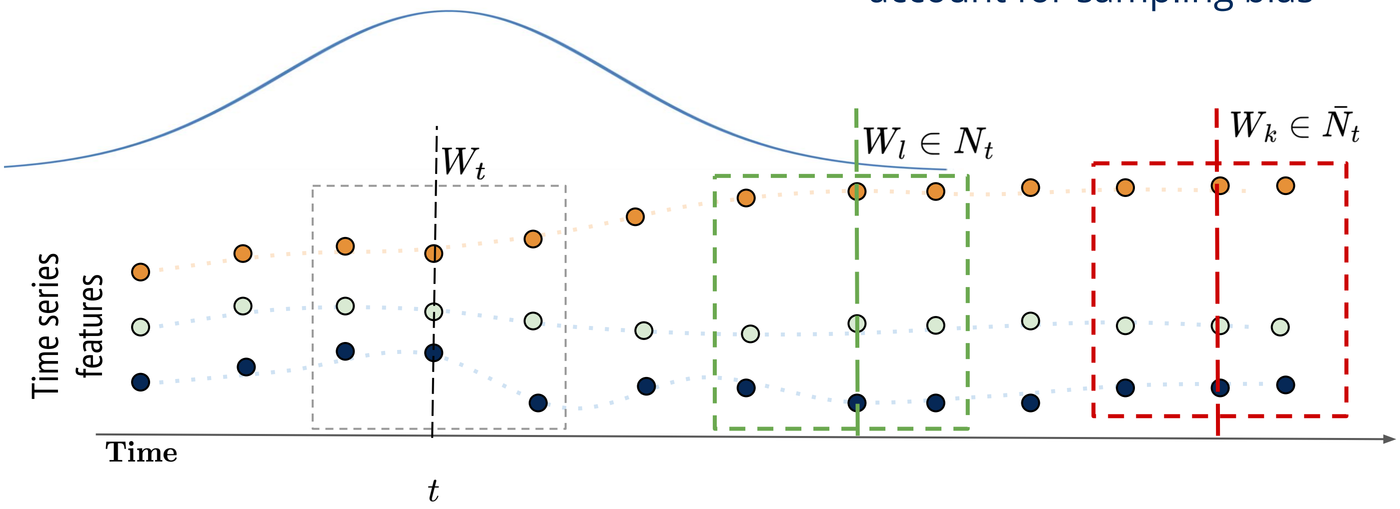
neighborhood range using ADF test of stationarity



Contrastive Objective

$$-\mathbb{E}_{W_t \sim X}\left[\boxed{\mathbb{E}_{W_l \sim N_t}[\log \mathcal{D}(Z_t, Z_l)]} + \boxed{\mathbb{E}_{W_k \sim \bar{N}_t}[(1-w_t) \times \log(1-\mathcal{D}(Z_t, Z_k)) + w_t \times \log \mathcal{D}(Z_t, Z_k)]}\right]$$

Weight adjustment from **PU learning** to account for sampling bias

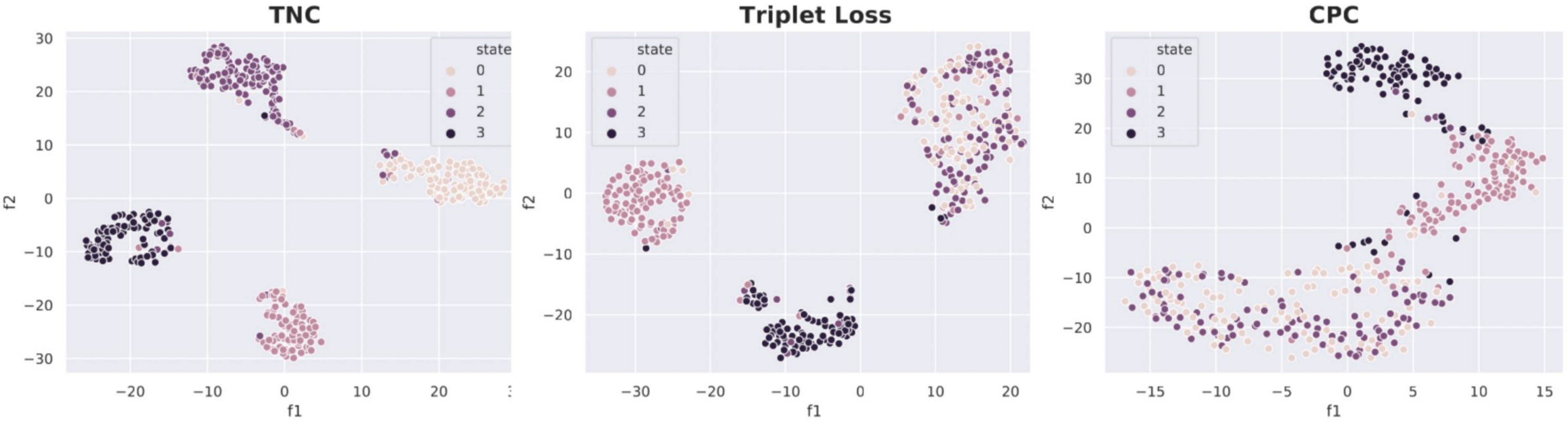


What makes a representation a good representations?

What makes a good representations?

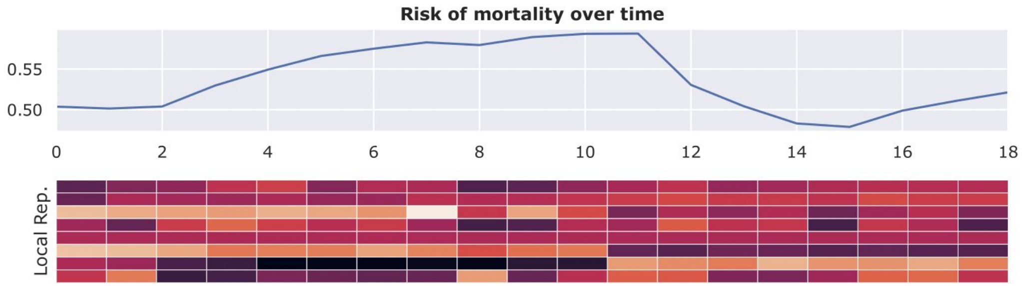
1) **clusterability** of the underlying states.

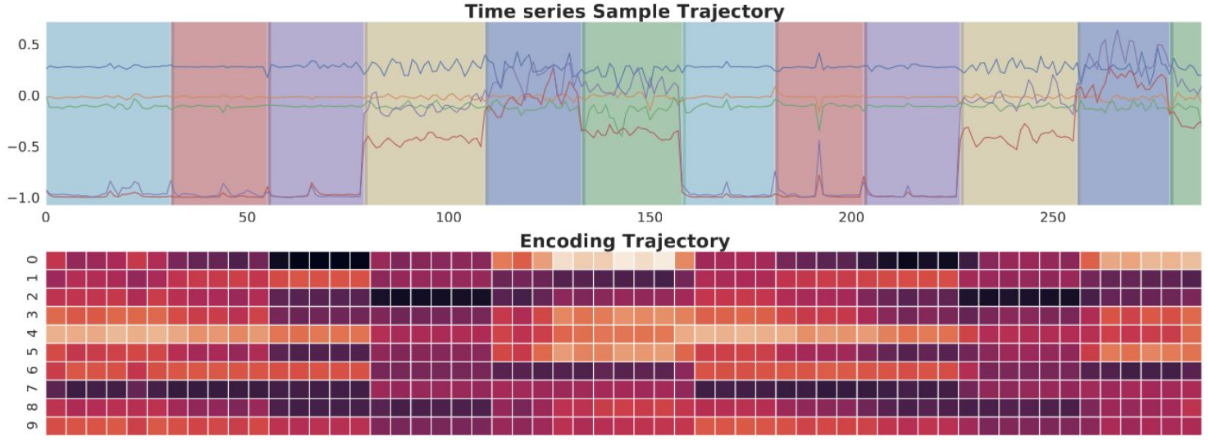
2) Generalizability for downstream **classification.**



What makes a good representations?

2) Generalizability for downstream tasks.

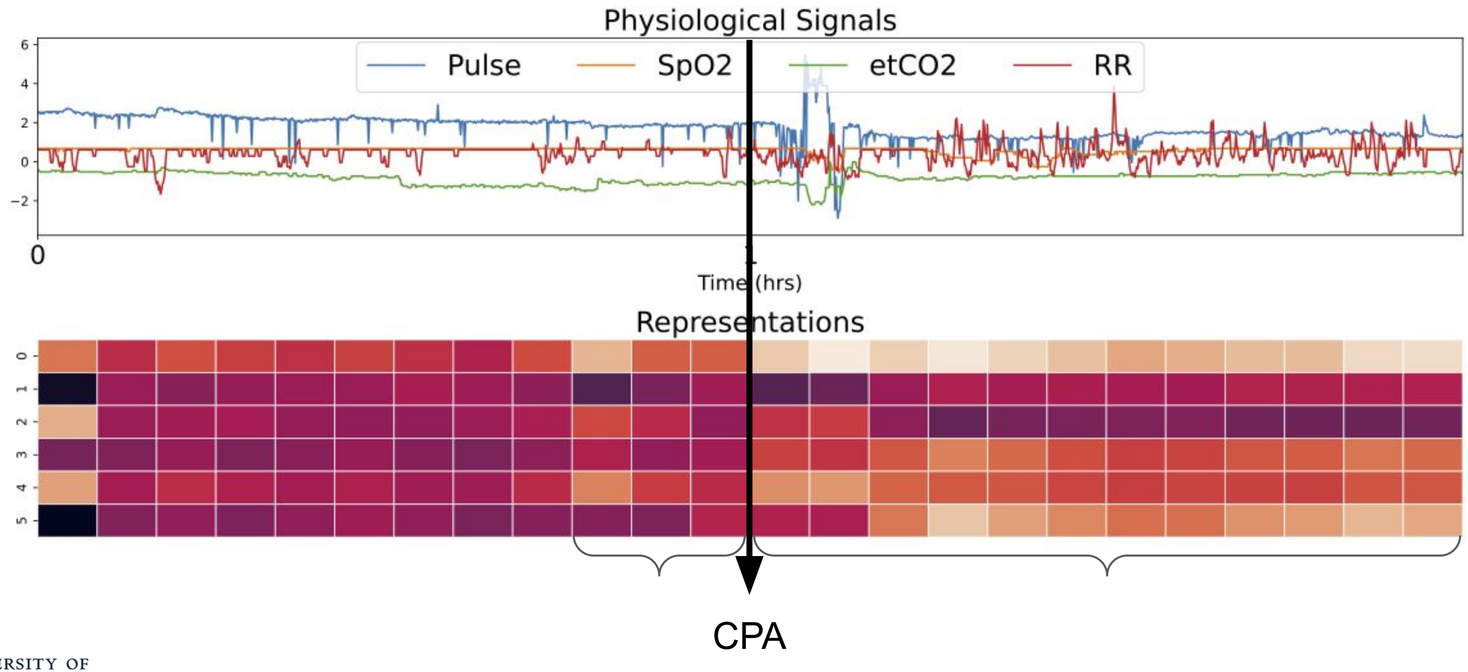




UNIVERSITY OF TORONTO

What makes a good representations?

3) Identifies changes in time



Where do you see potential for unsupervised representation learning in healthcare?

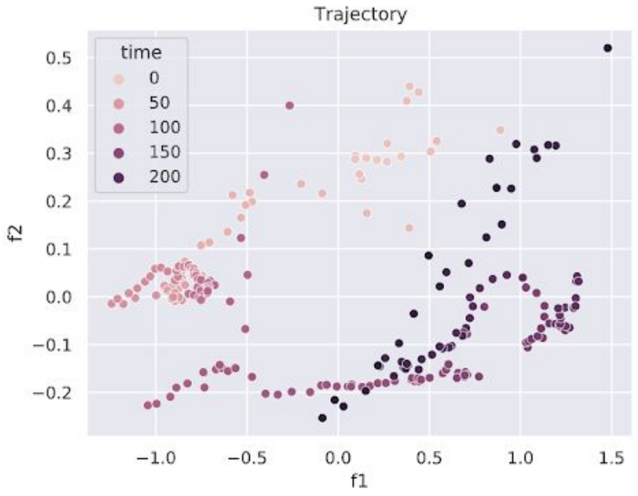
UNIVERSITY OF TORONTO

Tracking and gaining insight into individual's health

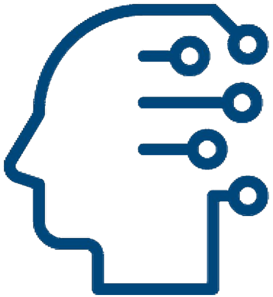
Knowledge discovery

Disease subtypes

Understanding population heterogeneity



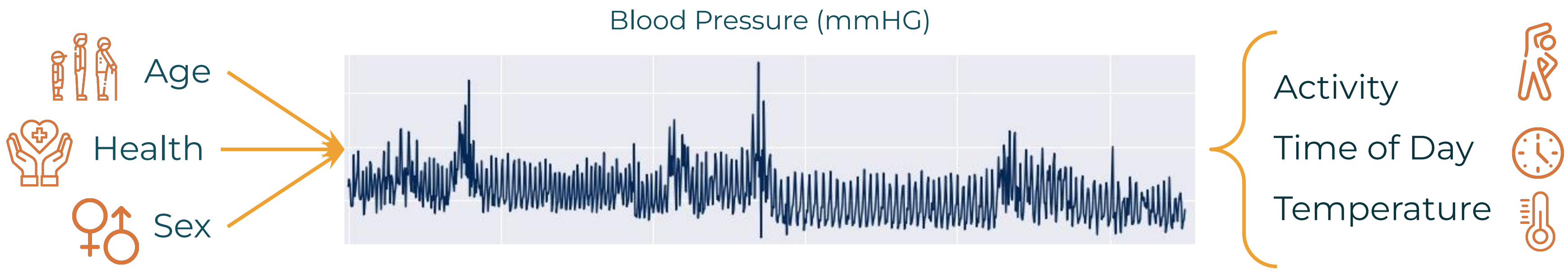
Better understanding representations



Decoupled Representation Learning

Time series data is generated from two independent sources of variation:

- **Global variable**: representation of the global behaviour of a time series
- **Local variable:** characterizes within sample variability and behaviours such as non stationarity

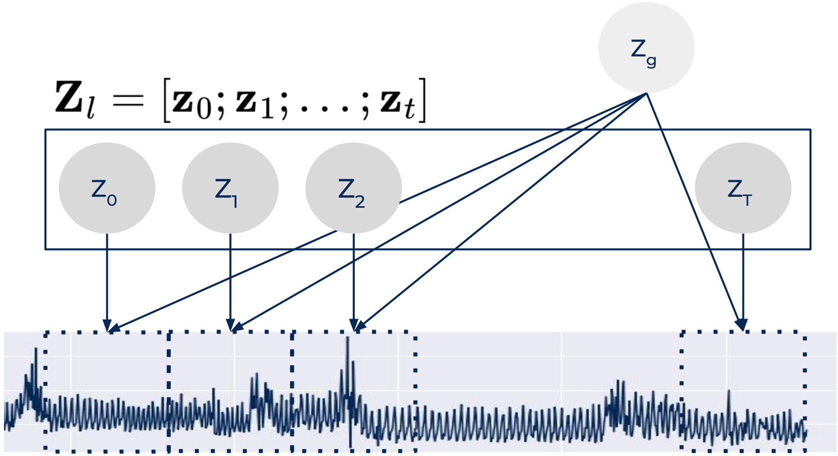


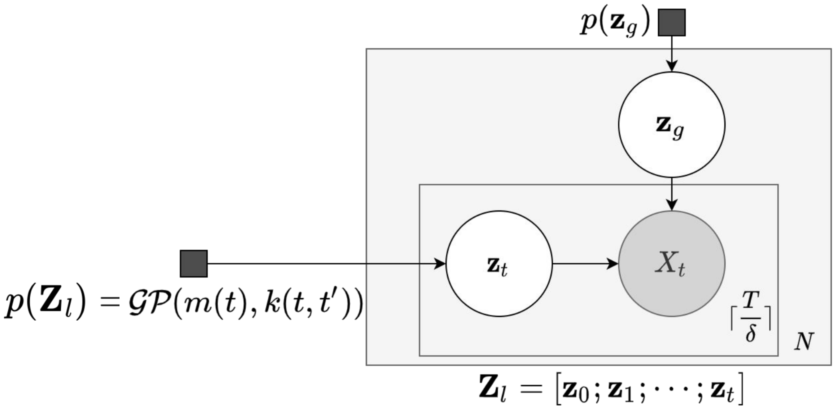
Tonekaboni, S., Li, C. L., Arik, S. O., Goldenberg, A., & Pfister, T. (2022, May). Decoupling local and global representations of time series. In *International Conference on Artificial Intelligence and Statistics (AISTATS)* (pp. 8700-8714). PMLR.

UNIVERSITY OF TORONTO

Generative Model

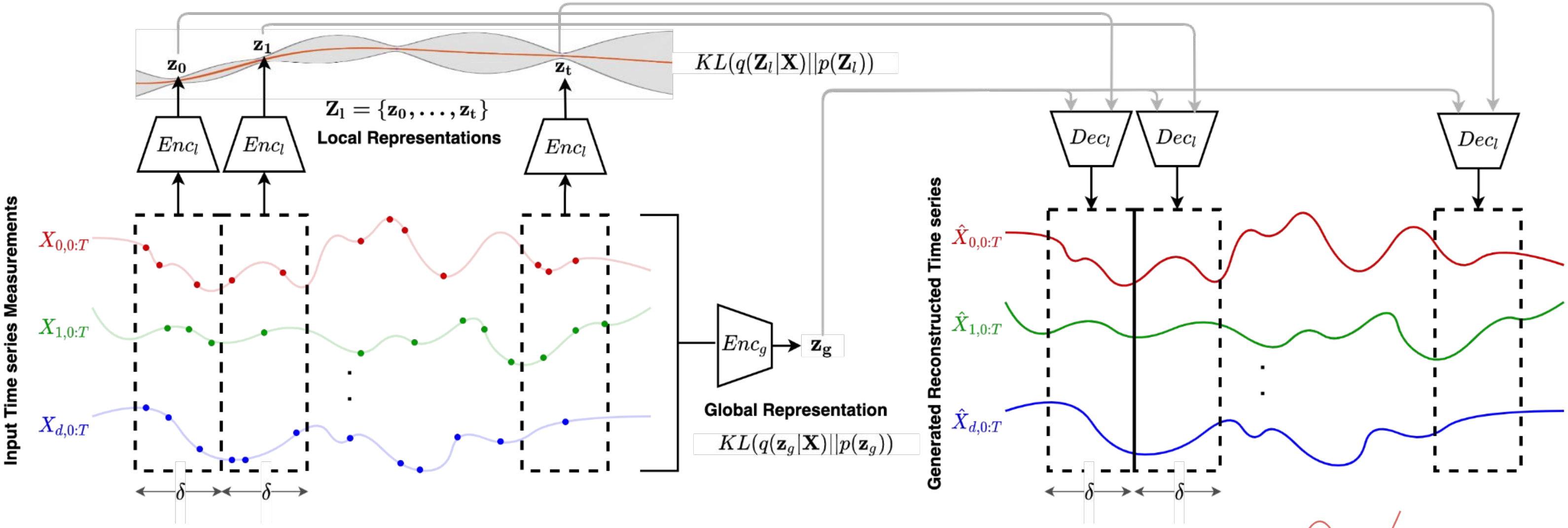
Every window of time series is generated from an underlying global and local variable. $p(X_{t:t+\delta} | \mathbf{z}_t, \mathbf{z}_g)$





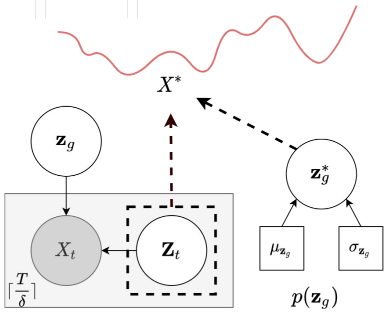
UNIVERSITY OF TORONTO

Variational Inference:



$$ELBO \boxed{+\lambda \mathbb{E}_{z_g,Z_l} \frac{q(z_g|X^*)}{q(z_g^*|X^*)}}$$

Counterfactual Regularization:
Disentangle global and local factors



UNIVERSITY OF TORONTO

Why understanding representations is important?