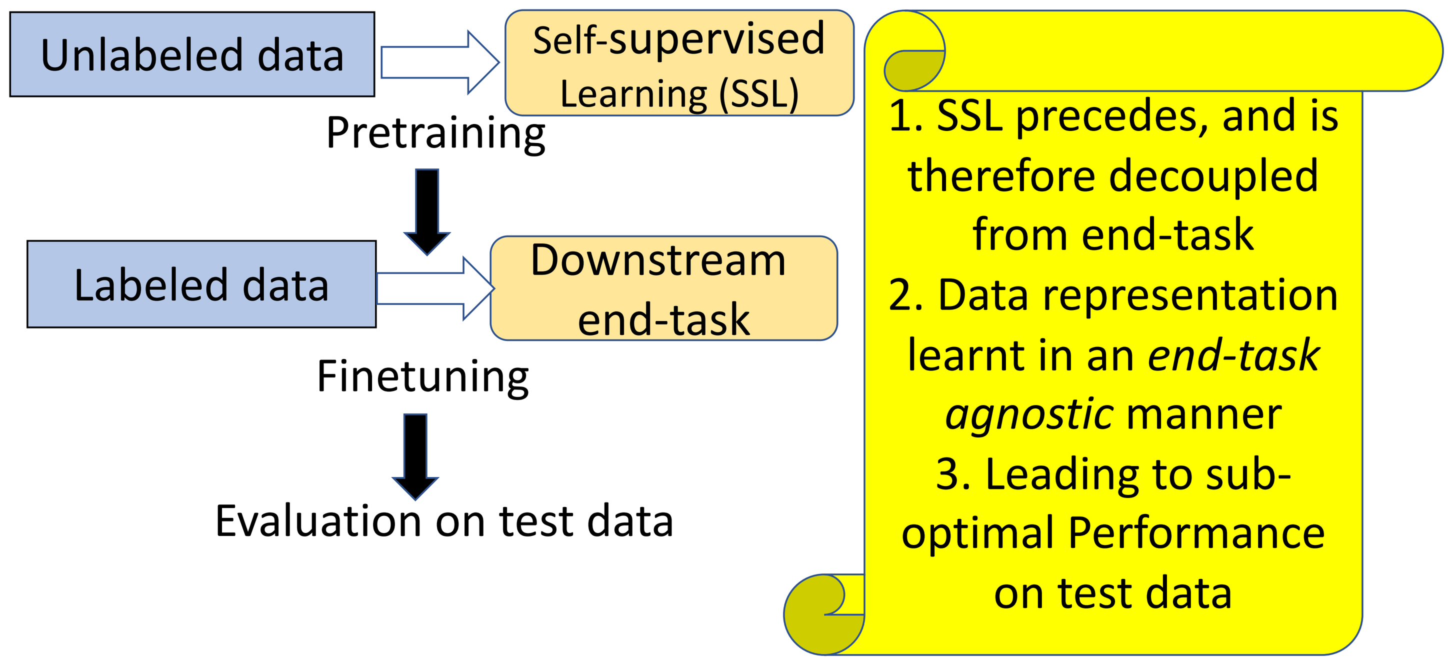
TARNet: Task-Aware Reconstruction for Time-Series Transformer

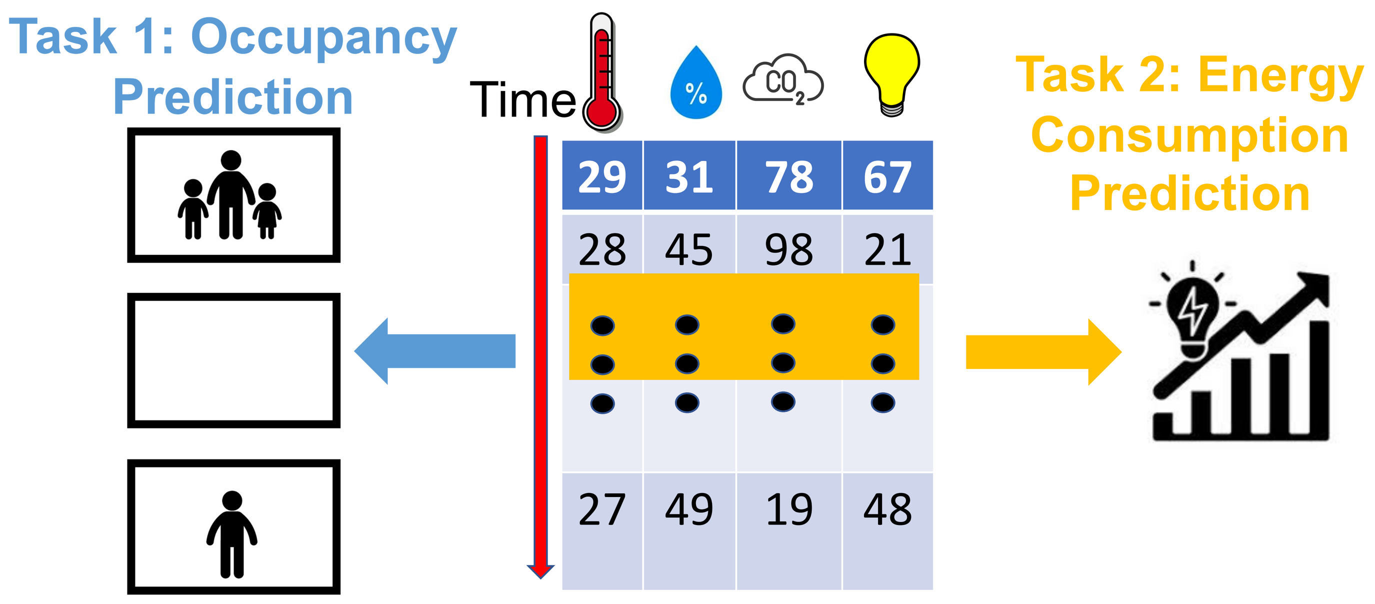
Ranak Roy Chowdhury, Xiyuan Zhang, Jingbo Shang, Rajesh K. Gupta, Dezhi Hong

rrchowdh@eng.ucsd.edu, xiyuanzh@eng.ucsd.edu {jshang, gupta, dehong}@eng.ucsd.edu

University of California San Diego, La Jolla, CA, USA

Motivation





- **Goal:** How can we learn a more *task-aware* data representation through SSL?
- **Hypothesis**: Using end-task specific knowledge to customize the learnt representation towards the end task may improve performance on end-task.

Related Work

- Statistical Methods: Distance-based, Shapelets, ROCKET
- Deep Learning Methods:
  - Using labeled data: CNN, LSTM, Attention
  - Using both unlabeled and labeled data: Negative Sampling, Contrastive Loss, Data Reconstruction[1]

Proposed Method

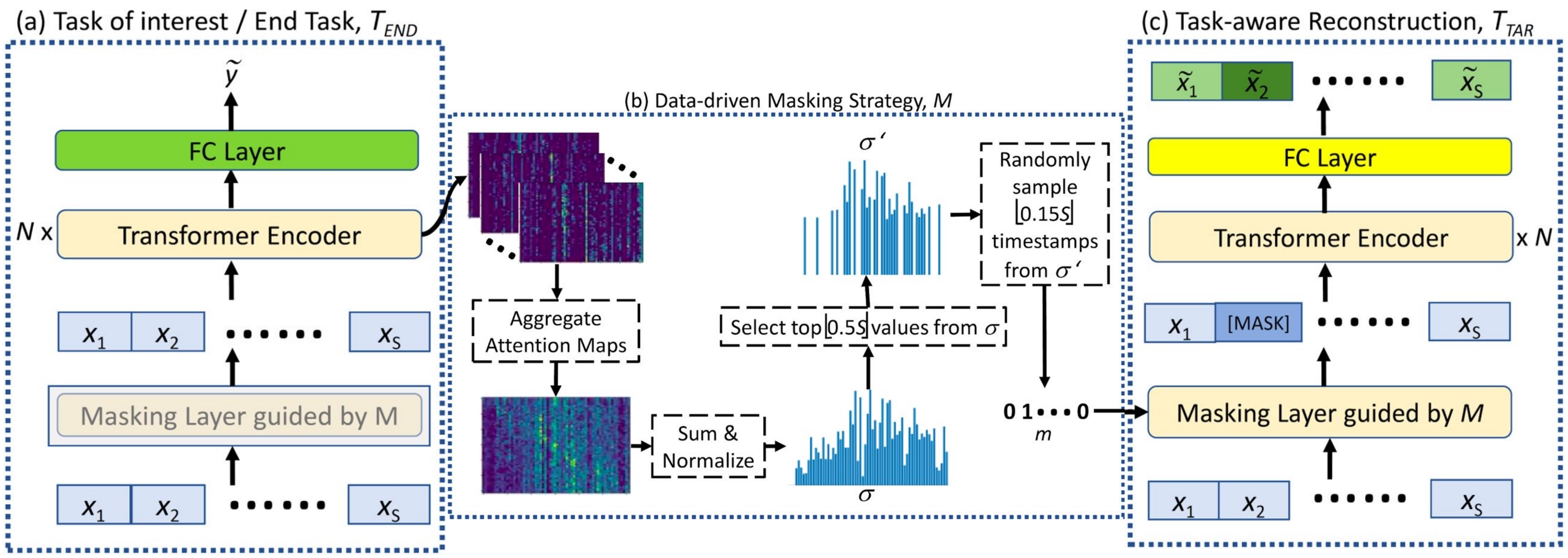
- Input: Uni-/multi-variate time-series X, Output: label y
- We use Transformer Encoder as the backbone model.
- T_{END} generates attention scores that is fed to M.
- M selects a set of most important timestamps, and randomly samples a subset of those times to produce m.
- Generated mask m decides which timestamps to mask during reconstruction, T_{TAR}.

$$\mathcal{L}_{TAR} = \lambda \mathcal{L}_{masked} + (1-\lambda)\,\mathcal{L}_{unmasked}$$

$$\mathcal{L}_{Total} = \eta \mathcal{L}_{TAR} + (1-\eta)\mathcal{L}_{END}$$

[1] George Zerveas, Srideepika Jayaraman, Dhaval Patel, Anuradha Bhamidipaty, and Carsten Eickhoff. [n. d.]. A Transformer-based Framework for Multivariate Time Series Representation Learning. In KDD, pages=2114–2124, year=2021.

Proposed Method



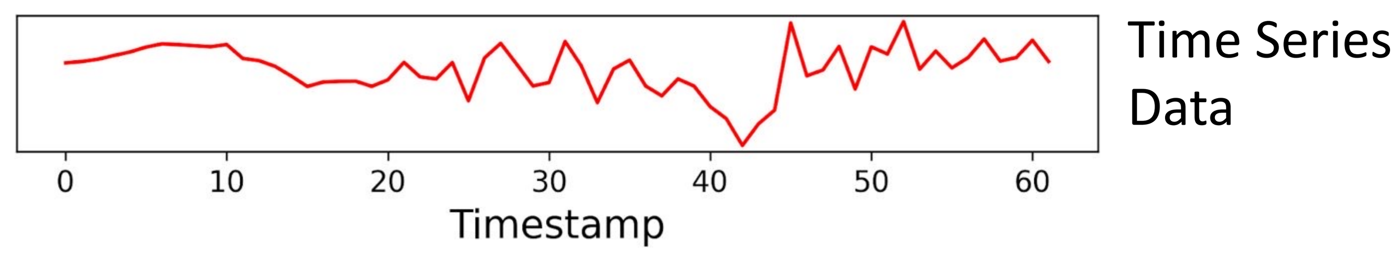
Algorithm

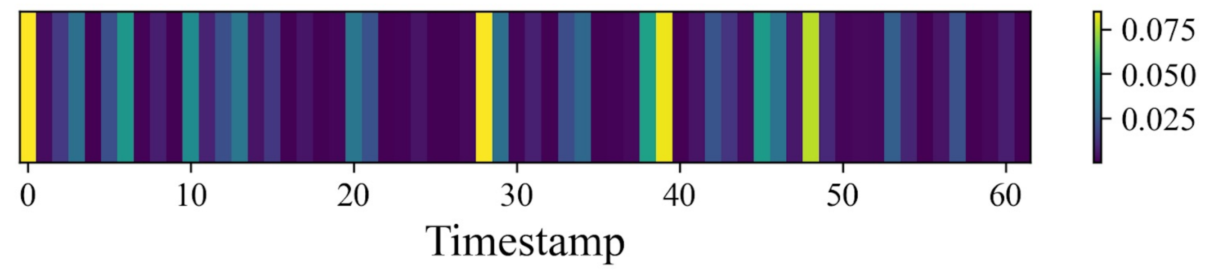
Algorithm 1 Training of TARNet

Input: X, y
Hyper-parameters: μ, β, λ, η
Output: *Model*

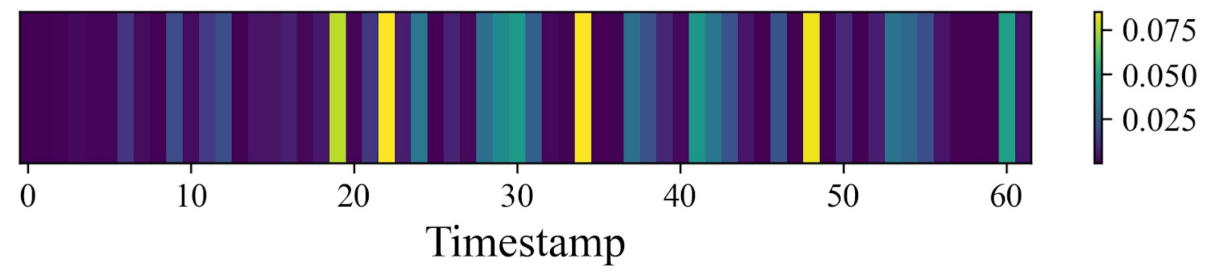
1: σ initialized randomly
2: *Model* = TransformerEncoder()
3: **while** training **do**
4: σ' = top $\lfloor \beta S \rfloor$ values from σ
5: $m \sim$ Randomly sample $\lfloor \mu S \rfloor$ timestamps without replacement from σ'
6: $\tilde{\mathrm{X}}, \tilde{y}, \mathrm{A} = Model.\mathrm{train}(\mathrm{X}, m)$ # A $\leftarrow$ Self-Attention Scores
7: Compute $\mathcal{L}_{TAR}(\tilde{\mathrm{X}}, \mathrm{X}, \lambda)$ and $\mathcal{L}_{END}(\tilde{y}, y)$
8: $\mathcal{L}_{Total} = \eta\mathcal{L}_{TAR} + (1-\eta)\mathcal{L}_{END}$
9: σ = add_and_normalize(A)
10: **end while**
11: **return** *Model*

Case Study





Random Data Reconstruction



Task-aware Data Reconstruction

Experimental Results

- Classification:
  - 34 datasets from UEA Time Series Classification archive.
  - 14 baselines – statistical and deep learning-based.
  - 2.7% higher average accuracy, 1.74-point lower average rank, and best results on 17 datasets compared to 7 by 2nd best baseline, Time Series Transformer (TST)[1].
- Regression:
  - 6 datasets from UEA Time Series Regression archive.
  - 12 baselines – statistical and deep-learning based.
  - 0.67-point lower average rank, and best results on 3 datasets compared to 2 by 2nd best baseline, TST.

Conclusion

- Task-agnostic SSL may produce sub-optimal performance
- Learn *task-aware* representation customized to end-task
- End-task and reconstruction task trained alternately.
- Data-driven masking strategy uses attention score distribution to find timestamps deemed important by end-task and mask them out for reconstruction.
- TARNet outperforms 26 baselines on 40 datasets.
- Case study shows task-aware method captures domain-specific inherent properties from data.

Code is publicly available at: https://github.com/ranakroychowdhury/TARNet