Self-supervised learning for time series

Borbély Bernárd

Supervisor: Csiszárik Adrián

Eötvös Loránd University

2024 May 30

Motivation

- Time series everywhere
  - Medical data such as: ECG, EMG, EEG
  - Stock market
  - Boring and drill failure
- A lot of data (mostly unlabeled)
- Self-supervised learning
- Foundation model for time-series

Introduction

- Zhang et al: Pre-training method for time series datasets
- Merging multiple dataset for pre-train failed
- Investigate the reasons behind this phenomena

Time-Frequency Consistency framework [2]

- Zhang et al: Pre-training method for time series datasets
- Uses Fourier-transform
- Method:
  - Embedding from Time-series: Time domain
  - Embedding from Fourier transform: Frequency domain
  - Cast both into a common domain: Time-Frequency domain

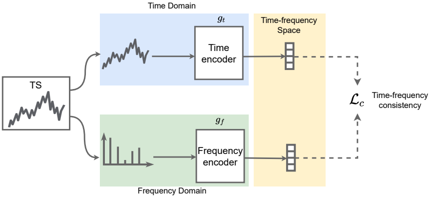
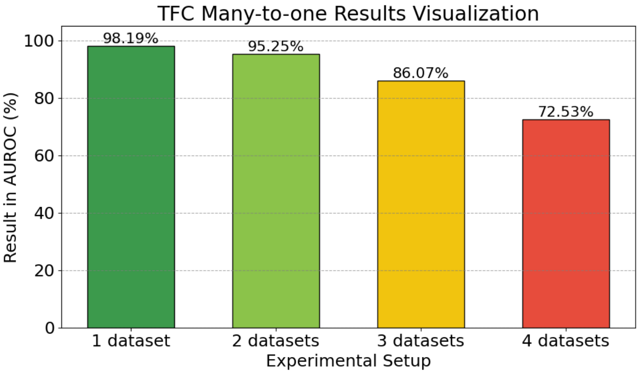


Figure: TFC contrastive learning [2]

TFC results

- Testing embedding quality:
  1. Pre-train on one dataset
  2. Fine-tune on a different dataset
- Results:
  - One-to-one: success
  - One-to-many: success
  - Many-to-one: fail
  - *More pre-train dataset → Worse fine-tuning*



Questions

- Why does combining multiple pre-training datasets hurt performance in TFC fine-tuning?
- Unusal in machine learning (more data is better)
- Interesting implications for time-series domain
- It this specific for dataset composition?

Methodology

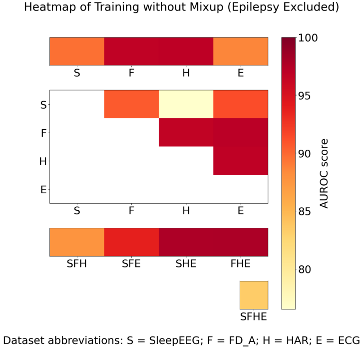
- Mixup augmentation
  - For data points x_i, x_j and their respective labels y_i, y_j
  - Augmented data point $(\hat{x}, \hat{y})$:

$$\hat{x} = \lambda \cdot x_i + (1 - \lambda) \cdot x_j$$

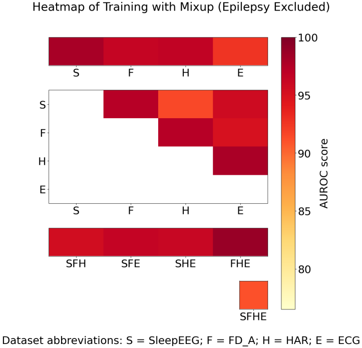
$$\hat{x} = \lambda \cdot y_i + (1 - \lambda) \cdot y_j,$$

  where $\lambda \in Beta(\alpha)$, and α is a fixed.
- Adding fine-tuning dataset to pre-train dataset
- (Never Train from Scratch: Fair Comparison of Long-Sequence Models Requires Data-Driven Priors, ICLR 2024) [1]

Experiments and results 1.



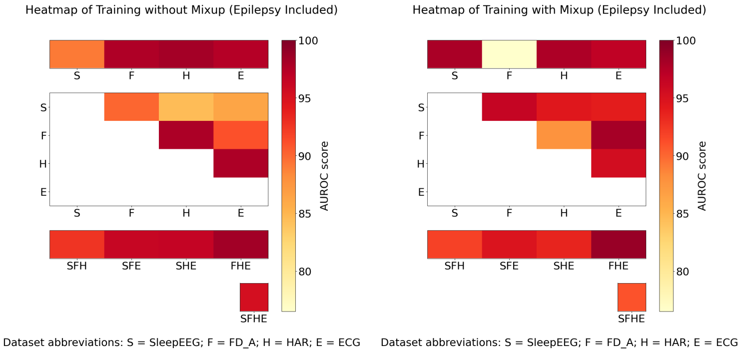
(a) No mixup, Epilepsy excluded from pre-train



(b) Using mixup, Epilepsy excluded from pre-train

Figure: No matter the pre-train dataset composition, using mixup yields success.

Experiments and results 2.



(a) No mixup, Epilepsy included in pre-train

(b) Using mixup, Epilepsy included in pre-train

Figure: Adding the target dataset to pre-training.

Discussion and Future research

- Mixup achieves improvement
- Adding fine-tune to pre-train: doesn't achieve considerable improvement
- Future research:
  - Repeat the experiment
  - Other augmentation techniques
  - A plot comparing dataset similarity and fine-tune AUC
  - Loss function expansion
  - New networks

Thank you for your attention!