Self-supervised learning for time series

Borbély Bernárd

Supervisor: Csiszárik Adrián

Eötvös Loránd University

2023

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

- Good for unlabeled data
- unlabeled $\rightarrow$ labeled
- Computer Vision and NLP domains
- Time-series domain?
- Goal: Create embedding

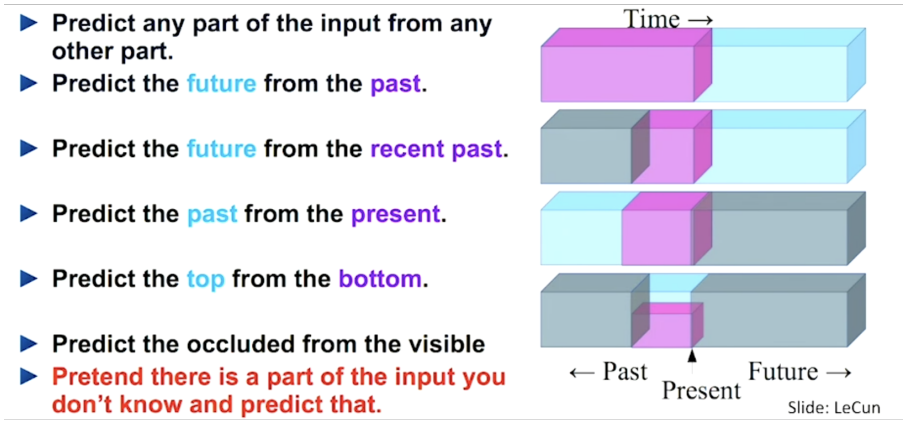

Figure: Yann LeCun's famous figure showing different kinds of realization of self-supervised learning

Time Frequency Consistency framework [1]

- 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
  - (Contrastive learning)

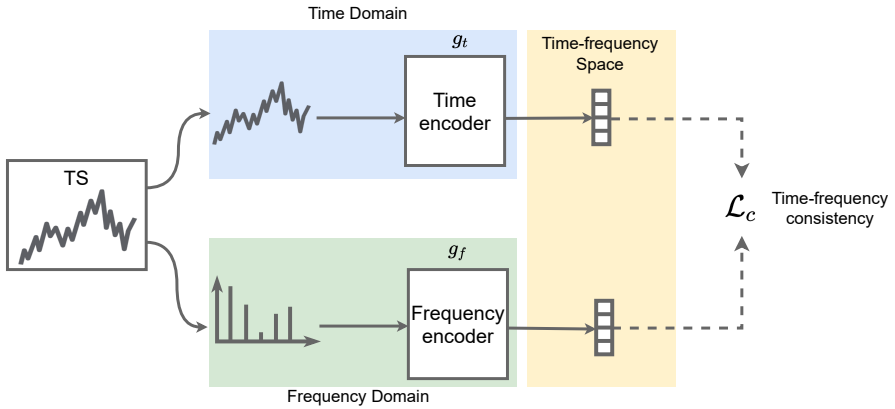

Figure: TFC contrastive learning from [1]

Time Frequency Consistency results

- Testing embedding quality:
  1. Pre-train on one of the dataset
  2. Fine-tune on a different dataset
- One-to-one: success
- One-to-many: success
- Many-to-one: fail
- *The more datasets were combined for pre-training, the worse the results became on the fine-tuning datasets.*
- Why?

Initial hypothesis

- Pre-train datasets are too diverse
- Too easy to tell apart
- Dataset similarity measure
- Expectations: Closeness $\rightarrow$ Higher fine-tune accuracy

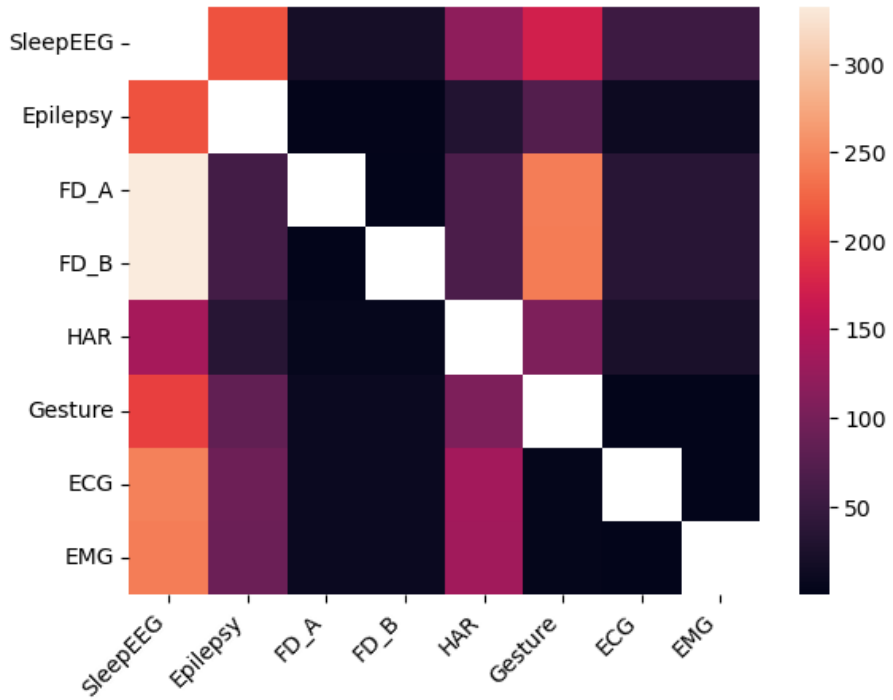

Figure: Dataset comparison

Future research

- Extend the code so every dataset could be pre-train and fine-tune dataset
- One-to-one measurements for each dataset pair
- Plot comparing dataset similarity and fine-tune efficiency
- Loss function expansion
- New networks

Thank you for your attention!