Optimizing Sleep Stage Classification

Course: STAT 4830 – Numerical Optimization

Date: February 7, 2025

Project Code Name: STAT-4830-MOIRA

1. Project Objectives

- Develop an optimized **classifier** for sleep staging from **scalp EEG**.
- Improve **wake vs. N1 sleep classification**, particularly during **sleep onset transitions**.
- Address **limitations of existing classifiers**, such as:
  - Over-reliance on **in-bed intervals**.
  - **High sensitivity** to sleep epochs but **low specificity** for wake epochs.
  - **Generalization issues** across different subjects and time-of-day variations.

2. Dataset Description

- **Data Source:** EEG recordings from **29 subjects**, stored in **EDF format**.
- **Sleep Stage Labels:** Provided in **text files**, with **one label per 30-second epoch**.
- **Preprocessing:**
  - **Bandpass filtering (0.5–40 Hz)**
  - **Epoch segmentation (30s windows)**
  - **Normalization per subject**
- **Feature Extraction:**
  - **Power Spectral Density (PSD)** features for **delta, theta, alpha, beta** bands.
  - **Catch22 time-series features** capturing autocorrelation, entropy, and complexity.

3. Methodology

3.1 Mathematical Formulation of the Optimization Problem

Problem Setup:

Given an EEG time series Y with associated covariates Z, we seek to predict the sleep stage S for future time steps using a learned function:

$$\hat{S}_t = f_\theta(Y_{t-l:t}, Z_{t-l:t+h})$$

Objective Function:

- We train f_θ to maximize the **log-likelihood** of the predicted sleep stage distribution:

$$\max_\theta \mathbb{E}_{(Y,Z)\sim p(D)} \mathbb{E}_{(t,l,h)\sim p(T|D)} \log p(S_{t:t+h} \mid \hat{\phi})$$

where $ \hat{\phi} $ are the learned parameters of the predictive distribution.

3.2 Implementation using MOIRA

MOIRA: Masked Encoder-based Universal Time Series Forecasting Transformer

- Why MOIRA?
  - **Handles multivariate time series** effectively using a **masked encoder architecture**.
  - **Learns from unobserved data (zero-shot learning)** to improve generalizability.
  - **Uses multi-patch projections** for capturing hierarchical temporal structures.
  - **Outputs probabilistic distributions** rather than single-point predictions.

MOIRA-Based Classification Pipeline

1. **Preprocess EEG Data**
   - Load EDF recordings and segment into **5-second windows**.
   - Apply **bandpass filtering (0.5–40 Hz)**.
   - Extract **PSD features + Catch22 features** for each epoch.
2. **Feature Transformation for MOIRA**
   - Apply **multi-patch projection layers** to generate vector embeddings.
   - Use **learnable masked embeddings** to encode future sleep stage predictions.
3. **Training Objective**
   - Train the model using **cross-entropy loss**:

$$\text{loss} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} y_i^k \log \hat{y}_i^k$$

where:

4. What We'Ve Learned So Far

- We have started to understand what models our goals necessitate. Timeseries forecasting models like the Google timesFM approach do not fit our classification problem.
- We learned about the MOIRA model's masked encoder approach and how it effectively captures temporal dependencies, which is applicable to our problem.
- We now understand catch22 on a deeper level and think we can go back to our initial idea of using it together with the MOIRA model.
- Additionally we learned processing full 24-hour EEG recordings is currently infeasible without significant optimization or additional hardware (e.g., GPUs). We expect data aggregation to be necessary in completing this project.
- We will need to determine the optimal downsampled sampling rate that does not result in significant loss of information.

5. Current Progress

7. Planned Improvements

- **Parallelization Strategies:**
  - Implement `joblib` or `multiprocessing` to speed up **feature extraction**.
- **Deep Learning Extensions:**
  - Evaluate performance of **TCNs, CNNs, and GNNs** and potentially deploy ensemble methods on raw EEG signals.
- **Hyperparameter Tuning:**
  - Optimize **epoch length, learning rate, embedding dimensions**.

Conclusion

- **Further tuning and parallelization** are needed to scale to 24-hour EEG recordings.
- **Next steps involve deep learning integration** to enhance robustness and generalizability.