Semester Project

AI for Medical Time Series - Spring 2022

Introduction

As of this week, you will be working on the course project. The project will take 3 weeks in total and every week one part of the project has to be submitted together with a progress report. In the last lecture of the semester, you will present your projects in the class. The aim of this project is to let you gain experience on designing and applying supervised and unsupervised learning methodologies on medical time series data. While doing that you will learn how to build an analysis pipeline for your classification problem based on the course material. For your project we ask you to choose one time-series dataset and work with that. There are two options for the dataset: (a) choosing one from 3 datasets that we have curated and you can find on Ilias in "Project Material" folder (2 scalp electroencephalography (EEG) and 1 intracranial EEG (iEEG) datasets), or (b) finding another medical time series dataset that you want to work with. If you want to use your own dataset (other than one of the three recommended ones), you may need to additionally curate and clean it. This week, you should submit code and project report for **only 1st week**.

The first part of the project will be marked as OK if you get **19** / **24** points or more.

The solutions must be handed in via **ILIAS**. Deliver your submission as a compressed file (zip) containing one .py or jupyter notebook file. Please make sure to name the zip file as follows:

project_part1_surname_name.zip.

If you are working as a group, then indicate the two names in the file name as

project_part1_surname1_name1_surname2_name2.zip.

The project can be handed in by two students working together. Copying code or solutions of individuals outside the group (e.g. submitting the code of other individuals as your own) will result in 0 points.

Deadline: 16:00, **May 4**.

Datasets

1. **Dataset-1: time series of EEG responses to sounds** .....

   This dataset consists of 18 subjects, 60 channels, 301 time points, and 3 conditions representing standard, target and novel sounds. For more detail about the dataset please check the original publication: https://www.frontiersin.org/articles/10.3389/fninf.2017.00067/fullh3. One example classification task for this dataset is differentiating responses to standard vs target sounds, or standard vs novel. On Ilias you can find a script named LoadingData_dataset_1-sounds.ipynb in Project Material folder, for loading the data and accessing the labels of all trials as 'Standard', 'Novel' and 'Target'.

2. **Dataset-2: EEG time series of responses to neutral vs unpleasant images** .....

   This dataset consists of 22 subjects, 64 channels, 384 time points. There are 6 classes representing EEG responses to images with different familiarity and pleasantness (familiar and neutral (FN), familiar and pleasant (FP), familiar and unpleasant (FU), novel and neutral (NN), novel and pleasant (NP), novel and unpleasant (NP) pictures). For more detail about the dataset please check https://zenodo.org/record/197404.Ymf7HS0RrRY.

   There are different options for the classification task for this dataset such as (a) familiar and neutral (FN) vs. familiar and unpleasant (FU) images, (b) familiar and pleasant (FP) vs. familiar and unpleasant (FU) and (c) familiar (FN, FP and FU together) vs novel (NN, NP and NU together). On Ilias you can find a script named LoadingData_dataset_2-images.ipynb in Project Material folder, that includes code for loading the data and accessing the labels of all trials as 'FU', 'FP', 'FN', 'NU', 'NP' and 'NN'.

3. **Dataset-3: intracranial EEG responses during a virtual game** .....

   This dataset consists of 9 patients with epilepsy, each of them has a variable number of channels (5-61), and 57 time points. For more detail about the dataset please check https://www.sciencedirect.com/science/article/pii/S0960982218309758. The dataset is available on Ilias. Each patient has two files: one numpy file for the iEEG data and another one for the labels (1: win, 0: loss) of the trials. For each patient the data has (trials x time points x channels) structure. There are 2 conditions representing loss vs. win in the behavioral task. The classification task for this dataset is classifying intracranial EEG responses to wins vs. losses.

4. **Exemplar ideas for your project** .....

-Decoding identity of participants or patients based on their EEG responses
-Classify EEG responses to different sounds (e.g. standard vs. novel) or images (e.g. pleasant vs. unpleasant) or to the virtual game (win vs. loss)
-Examine cross-participant generalization of time-domain classifiers (e.g. how well does a classifier trained on participant perform on others?)
-Evaluate how classification parameters and choices affect final classification performance and conclusions (e.g. evaluating different metrics of performance, dealing with biased/unbalanced classes, evaluating different CV ratios etc)

Project plan for the next three weeks

1. **1st week** .....24 points

   In the first week you will need to (a) find a time series dataset and define a classification task that you want to work on and (b) work on feature extraction.

   (a) In your report please describe the dataset that you choose to work with (e.g. the number of classes and the data size of each class) and the exact classification task that you plan to work on. Based on your choice of data and classification task please describe in your report how you plan to analyse these data, step by step. You can describe how you will apply machine learning techniques, and which ones, training scheme, the metrics for evaluating performance of the models and feature selection (time-frequency/frequency transforms, dimensionality reduction etc.). Please hand in your project report on Ilias.

   (b) After planning your analysis and classification task, please start working on the feature selection part. Please include your python code for feature selection in your submission for this week on Ilias, plots of the features (e.g. sub-sets of electrodes / time-frequency plots, etc), together with your report from step (a).

2. **2nd - 3rd week** .....

   In the second and 3rd week you will apply machine learning models for your classification task. More details for that to come in the next 2 weeks!