TP1: Data Analysis and Classification

Objective

Implement a multi-step data analysis pipeline in MATLAB to classify human emotions based on brain activity, utilizing raw electroencephalogram (EEG) recordings.

At the end must be submitted the code in MATLAB and a report with the relevant explanations and observations requested (max. of 12 pages).

Groups: 2 students
Execution time: 8 weeks
Deadline: 06/04/2025.

Motivation

The ability to autonomously detect human emotions has valuable applications in fields such as brain-computer interfaces, therapy, and neuroscience. One effective method for capturing diverse emotional states is through EEG, which measures electrical brain activity. However, EEG signals are inherently complex, non-stationary, and subject to randomness and noise contamination, requiring careful analysis and processing.

Database

Participants. Fifteen subjects participated in the study. Each participant was submitted to three experiments on different days. On each experiment the participants watched fifteen film clips (trials) corresponding to portions of Chinese movies, intended to elicit three different emotional states: happy (5 clips), neutral (5 clips) or sad (5 clips). On https://bcmi.sjtu.edu.cn/home/seed/seed.html can be found more details about the experimental protocol.

To access the dataset, you need to **register** and send a signed form to the SEED manager, following the instructions in : https://bcmi.sjtu.edu.cn/home/seed/downloads.html#seed-access-anchor (only one student per group needs to register).

Each group will work with a **raw dataset** from **one participant**, including the **three experiments** (the participants each group must work will be distributed in the laboratorial class). The records from each experiment are provided in individual files, with the name structure “Nº participant_Nº experiment.cnt”. The “cnt” files contain a header and the data itself (EEG records). From the header you can obtain information about the channels’ distribution, the sample frequency, etc. Additional information is also available in individual files, like the film clip sequence, start and end times of each clip, etc.

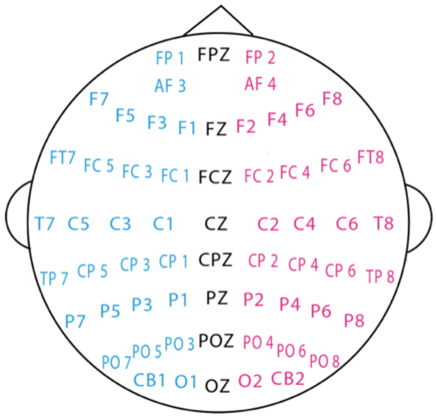

Fig. 1: System 10-20 of EEG electrode placement (62 channels). Obtained from: https://bcmi.sjtu.edu.cn/home/seed/seed.html.

Tasks

Main Tasks. The main steps to be implemented are represented in Fig. 2.

Fig. 2. Data analysis pipeline.

Task 1

Load data (week 1). Load the *.cnt* files of the corresponding group.

- To open the .cnt files you can use the *loadcnt*.m function available in https://github.com/sccn/neuroscanio/blob/master/loadcnt.m
- Exclude the signals that do not corresponds to EEG channels (based on the information contained in the header).
- Reduce the sampling frequency by a factor of 4 (doc *downsample*).
- Based on the information contained in the *time.txt* file from the SEED database, segment each experiment in the 15 trials.
- Plot the EEG record from each channel as function of time (based on the sampling frequency).

ICA (weeks 2-3). From the EEG records extract the independent components using the FastICA algorithm available at: https://www.mathworks.com/matlabcentral/fileexchange/38300-pca-and-ica-package.

- Obtain the independent components from all the EEG channels, and for each trial. Test the methods "negentropy" and "kurtosis" and select that with better results.

- Plot the independent components.
- Analyze components containing artifact and exclude it (i.e. make the component equal to zero).
- Reconstruct the EEG signals following the instruction in the fastICA function. Compare the reconstructed signals with the original ones.

📝 Report the method that presented the best result, the types of artifacts that can (or cannot) be separated in the independent components, and any other relevant approach/ information about this task.

Task 2

Windowing. To avoid introducing spectral artifacts related to signal truncation, apply a window function (example: doc ***hamming***, doc ***blackman***, doc ***hann***, etc.) to each reconstructed EEG segment (trial).

📝 Report the selected window types.

Task 3

Spectrum. Obtain the amplitude spectrum of each windowed EEG signal (doc ***fft***).

📝 Report what you observe from the signal's spectrum.

Task 4

Filtering. Considering that the frequencies outside the range 1 Hz - 100 Hz are not of interest for the study, identify the type of filters (LPF, HPF, BPF and/or BSF) and their cut-off frequencies to be applied for excluding the undesired spectral components.

📝 Report the filter(s) to be applied.

4.1 Considering only Finite Impulse Response (FIR) filters:

- Design the filters using the "Filter Design" app of MATLAB.
    - Evaluate the filter responses for different filter orders and types.
- Filter the signals using the function filter (doc ***filter***) and considering different window types.
- Evaluate the results through the spectrum of the filtered signals.

4.2 Considering Infinite Impulse Response (IIR) filters:

- Design the filters using the "Filter Design" app of MATLAB.
    - Evaluate the filter responses for different filter orders and types.

  - For IIR filters, the structures SOS and G are returned by the Filter Desing. These structures must be converted to Num and Den, using the sos2tf function (doc ***sos2tf***), before filtering the signals using the function ***filter***.
  - Evaluate the results through the spectrum of the filtered signals.

- Report the observations for different window and filter types and order, and specify the optimal choice for applying to the dataset.

Task 5

Spectral features. Investigate EEG spectral features that can be representative of emotional states.

  - Extract two features from the entire dataset, that will be used in the next tasks.

- Report the features extracted and the criteria for your choice.

Task 6

Wavelet. One way to observe the spectral changes over time is the application of wavelet transform.

6.1 Continuous wavelet transform (CWT): investigate the EEG channels that may be more representative in the assessment of emotional states.

  - Apply CWT to at least two of the signals from the identified channels (doc ***cwt***).

Ler o papper

- Report the observations.

6.2 Discrete wavelet transform (DWT).

  - Define the appropriated decomposition levels (decomposition tree) considering the frequency bands characteristics of brain activity (Table 2). Generate the filter banks for different wavelet types using the function dwtfilterbank (doc ***dwtfilterbank***). Using MATLAB functions, you can visualize the passband for each decomposition level (doc ***dwtpassbands***), the wavelets (doc ***wavelets***) and the corresponding frequency response (doc ***freqz***).
  - Performs the wavelet analysis using the wavelet decomposition filters (doc ***wavedec***) and extract the detail coefficients from the wavelet decomposition structure (doc ***detcoef***).
  - Extract features from the DWT, i.e. the power on each frequency band.

Table 2: EEG frequency bands.

Band	Frequency Range [Hz]
Delta	1 – 3
Theta	4 – 7
Alpha	8 – 12
Beta	13 – 30
Gamma	30 - 100

Task 7

Dimensionality reduction. After feature extraction a high dimensionality space will be obtained. On this sense, dimensionality reduction is useful for excluding irrelevant or redundant information. Also, reducing dimensionality makes possible data visualization.

PCA. Dimensionality reduction must be first implemented using PCA.

- For PCA, the data must be previously normalized (doc ***normalize***). Test at least two methods for features normalization (e.g. Z-Score, Rescaling, Interquartile Range).
- Apply PCA to the normalized data (doc ***pca***).

Report the results using the diverse normalization methods, and discuss the result obtained with PCA.

MDS. Reduce dimensionality by multidimensional scaling.

- The first step consists of distances measurements between pairs of observations (doc ***pdist***). Test at least ~~three~~ 2 different distance methods. Note: use the normalized data.
- Apply MDS using non-classical methods (doc ***mdscale***). Test at least ~~two~~ 1 metrics and one non-metric criterion.
- Consider two and three dimensions.

Report the results obtained with the diverse implemented approaches.

Task 8

Clustering. Apply different clustering types introduced in the theoretical lectures, considering that we are interested on distinguish three situations: happy, neutral and sad emotional states.

- Test (at least) the following algorithms: k-means (doc ***kmeans***), k-medoids (doc ***kmedoids***), agglomerative hierarchical clustering (doc ***cluster***), and fuzzy c-means (doc ***fcm***, from Fuzzy Logic Toolbox).
- Evaluate the clustering performance by constructing the confusion matrix (doc ***confusionmat***) and computing some metric like macro true positive rate.

Report the results obtained using different: wavelets, normalization methods, dimensionality reduction approaches, dimensions, and clustering algorithms