Ecole Polytechnique Fédérale de Lausanne

CS-433: Machine Learning 2021-2022
INL - Integrated Neurotechnologies Laboratory

Project 2: Fear Decoding in Rodents

Authors:
Brants Tomas (SV, SCIPER 342478)
Deleu Arthur (IC, SCIPER 342491)
Soria Gómez Alicia (IC, SCIPER 342122)

Supervisors:
Prof. Flammarion Nicolas (CS-433)
Prof. Jaggi Martin (CS-433)
Prof. Mahsa Shoaran (INL)
Bingzhao Zhu (INL)
Mohammad Ali Shaeri (INL)

Abstract

A deep learning model is built to try to decode fear in rodents. A combined approach was employed, consisting of a CNN and a MLP which have power spectra and phase-amplitude coupling matrices as input. The overall accuracy on the imbalanced data set is 0.71. On the balanced data set, the accuracy is 0.18. The rather poor results are a consequence of inadequate feature extraction.

1 Introduction

Deep Brain Stimulation (DBS) is a device-based therapy for Parkinson's disease, and has recently shown promise in treating patients with mental disorders [1]. However, the classical open-loop DBS has shown variable effect across patients, because the DBS outcome depends on the behavioural and clinical state of the patient at the fixed times of the DBS treatments.

A solution to personalize the DBS treatment for every individual and make DBS targeted and effective is to use closed-loop systems. In such a treatment, the system will be able to adapt to the state of the individual and provide the DBS treatment in response to changes in neuronal activity [2]. This motivates the work on developing efficient machine learning models to decode psychiatric disorders in real time for a closed loop DBS system.

Toward this direction, the EPFL research laboratory 'Integrated Neurotechnologies Laboratory (INL)' aims to design a model to decode the emotion of fear from rodents in real time, as a common symptom in anxiety disorders. Emotions like fear, anxiety and depression are abstract and not restricted to one specific brain area, which makes data preprocessing challenging from a machine learning perspective. The main challenge is to identify the correct features in the time-domain electrode recordings. This study focused on identifying features of Local Field Potential (LFP) data, obtained by placing 4 electrodes in two brain areas mediating fear: the IL and the BLA. These brain area's are known to be regions in the brain which are strongly involved in fear expression and hence, they can be used for determining fear emotion in rodents through behavioural experiments for fear decoding [3].

2 Data collection

The INL supplied a data set with the following measurements for 16 different rodents:

Local Field Potentials (LFPs): correspond to brain measurements represented as time signals of 1200 seconds. These are obtained with invasive intracortical electrodes in two areas of the brain.

- IL measurements: 4 channels sampled at 30 kHz
- BLA measurements: 4 channels sampled at 30 kHz

Fear quantification parameters: rodents were conditioned to fear by training them with electric shocks after playing a sound signal. The rodents then displayed particular behaviours before, during and after fear that can be quantified with the following measurements:

- Bar press data to receive a reward: 30 kHz sampling
- Acceleration data: 20 kHz sampling
- Freezing data: 20 kHz sampling

The fear quantification parameters were smoothed by binning in a .05 second window, resulting in $1200/0.05 = 2400$ time-windows. For each of these time-windows, a Gaussian kernel with size 400 was applied in the bar press (Figure 1), acceleration, and freezing measurements to produce softened curves, displaying the smoothed frequency of the measurement in the y-axis.

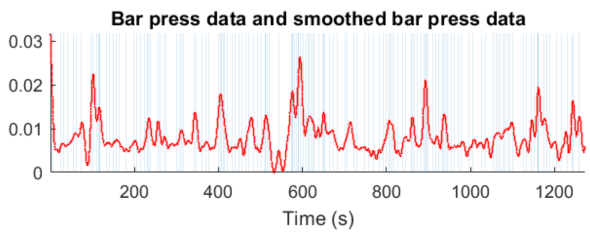

Figure 1: Raw (blue) and smoothed bar press (red) data

3 Data preprocessing

The preprocessing of data in order to extract the features and appropriate labels has been an important part of this project, as the provided raw data consisted of multiple time signals and measurements without much straightforward information or evident characteristics. Therefore, more than half of the effort devoted to the project has been oriented towards data preprocessing to try to extract adequate features. All preprocessing was done in Matlab.

3.1 Data selection:

The raw data was only partially employed to extract relevant features. Based on literature and guidance of the INL, the feature extraction was performed on only a part of the data set.

<u>OB9 rodent:</u> only measurements of the particular rodent with ID 'OB9' were considered, on recommendation of the INL researchers.

<u>IL measurements:</u> the infralimbic cortex is associated with fear suppression, memory, and training.

<u>BLA measurements:</u> the BLA is located in the prelimbic cortex (PLC), which is generally associated to fear expression, creation, and acquisition. LFP signals of the BLA are characterized by a rather small bandwidth, between few Hz up to a maximum of a few hundred Hz. They are divided according to their range:

- *Theta frequencies*: low frequency signals, between 4–12 Hz. Their behaviour is synchronous with brain impulses and normal functioning, and their objective in general terms is to coordinate different brain areas such as the synchronization of the BLA with the hippocampus and prefrontral portex [4].

- *Gamma frequencies*: these are higher frequency signals, which coordinate neural coding when special events occur. They can be divided into slow-gamma (between 40–70 Hz); and fast-gamma (between 70–120 Hz) [5].

One of the interesting characteristics of BLA LFPs during conditioned fear, displays itself in the interaction between theta and gamma frequencies. In particular, the phase of the theta signals affects the gamma frequency amplitude in a phenomenon known as amplitude modulation. Therefore, the 'theta phase – gamma amplitude coupling' (PAC) results in the most relevant features of this data set [6].

<u>Smoothed bar press data:</u> to quantify the expression of fear, only the smoothed bar press is considered. In this context, the lower the bar press rate, the more fearful the mouse is considered to be. Since the raw bar press rate is continuous, it was chosen to quantize the rate to three discrete levels each with corresponding interval values:

Label	Bins	Interpretation
0	0 - 0.01	**No fear apparent in rodents brain activity.**
1	0.01 - 0.02	**A small amount of fear could be noticed, but too noisy to be sure.**
2	0.02 - 0.03	**Rodent is fearful.**

Table 1: Label levels and their meaning.

3.2 Extraction of features:

Based on the previously selected data, the time signals were preprocessed and features were extracted to later train the model. For each of the features, the time signals were then divided into time-windows of 2 seconds, which resulted in 601 samples for the 1200 available seconds. The extracted features are:

1. *BLA Phase-Amplitude Coupling:* measured with its modulation index (MI)

2. *BLA power spectrum:* measured in normalized Watts

3. *IL power spectrum:* measured in normalized Watts

4. *Labels:* smoothed bar press rate average every 2 seconds

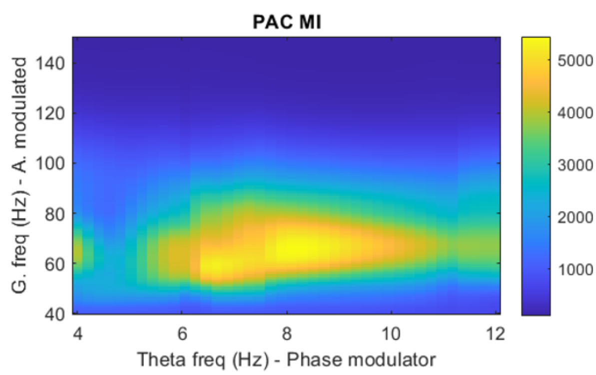

Figure 2: Visualization of the PAC MI feature

The detailed procedures carried out to obtain each of the features are explained in Annex 6.

4 Machine learning models

All machine learning models were coded in Python using the Pytorch, Pandas, NumPy and SciKit-learn libraries. Ray Tune was used to tune the hyperparameters of the neural networks.

4.1 Baseline model: majority voting

It became quickly clear that the data is unbalanced. The intermediate '1'-label was far more abundant than the other two classes. It was thus chosen to implement a very crude baseline model as a benchmark. The baseline will just label all test values equal to the most occurring training test label (here label '1'). The accuracy on the test set of this model is 0.64. A decent machine learning model should at least perform as well as this.

4.2 Multichannel Convolutional Neural Network

As an initial attempt, a convolutional neural network (CNN) was proposed with three input channels: one channel corresponds to the raw BLA Phase-Amplitude Coupling represented as the matrix of shape (110, 40), displayed in Figure 2, the other two contain the BLA and IL power spectra. Since these spectra are represented as vectors of size (100,), it was chosen to append these vectors with zeros until the total length equals (110,) and to then naively repeat them 40 times to create a matrix of shape (110, 40). It became clear rather quickly that this approach would lead to sub-optimal results due to the fact that the 3 channels are not spatially correlated with each other. Convolutions were executed at the same locations in each layer and patterns were discovered where in reality there aren't any.

4.3 Single Channel Convolutional Neural Network and Multi-Layer Perceptron

As a result, a second approach was considered. This approach is a two-staged rocket where the PAC-matrix is first processed by a CNN to extract features. These extracted features are then combined with the IL- and BLA-bandpower feature vectors and fed to a multi-layer perceptron (MLP), which will take care of the final classification.

4.3.1 CNN Architecture

The CNN architecture [7] was based on literature and can be found in Figure 3. The first step of this approach is to extract the features from the PAC-matrix, return them as a vector of shape (100,) and to later concatenate them with the IL- and BLA-bandpower, which have the same shape. The network first consists of 3 convolutional layers each of kernel size 3x3. Stride is set to its default value 1 and padding is initialised to 1 as well. After each convolutional layer firstly a rectified linear unit activation function is applied and secondly, a max pooling layer of kernel size 2x2 is used. After the last convolutional layer, the output is flattened and fed into a first fully connected neural network layer where a tanh activation function is applied its the output. After this a random dropout with probability of dropping 0.25 is performed. Finally a batch normalization is performed. After this the resulting tensor gets returned as a feature vector, which is fed along with other features to the MLP.

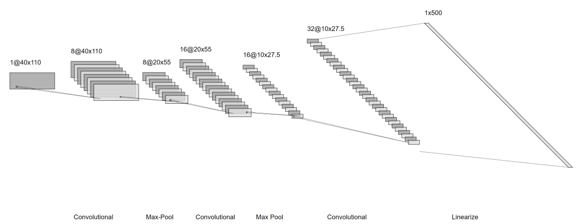

Figure 3: CNN architecture.

4.3.2 MLP architecture

After extracting relevant features from the PAC matrices with the CNN, these features are combined with the IL- and BLA-bandpowers by pasting the (100,)-shaped vectors all after each other to form a (300,)-shaped vector. This vector is then fed to a fully connected neural network. The MLP consists of 3 fully connected layers, each followed by a rectified linear activation function and a final fully connected layer. Both networks employed Cross Entropy

Loss as a loss function and the Adam optimizer to train the model.

4.3.3 Crossvalidation and Hyperparameter Tuning

5-fold crossvalidation was used to measure model performance using the SciKit-learn toolkit. Hyperparameter tuning was performed with the Ray Tune library in order to find the optimal learning rate, batch size and amount of epochs for both the CNN as the MLP. Optimal results are shown in Table 1.

Hyperparameter	**CNN**	**MLP**
Learning Rate	0.0094	0.173
Batch Size	10	10
Number of epochs	10	5
Accuracy	0.64	0.68

Table 2: Hyperparameters

The final accuracy of the whole model cascade is at most 0.71. This seemed like a very reasonable result given the fair amount of noise in the data.

5 Conclusion and reflections

5.1 Unbalanced pitfall

Although it was postponed until the very end of the project, it was deemed informative to also report the accuracy per label class of the bar press rate.

	Label 0	**Label 1**	**Label 2**
Accuracy	0.0	1.0	0.0

Table 3: Per class accuracy of unbalanced model.

From Table 3 it is immediately clear something went crucially wrong. Although the model had an averaged accuracy of 0.71, the model used exactly the same algorithm as the naive baseline. This was noticed far too late into the project since the per-class accuracy was only researched when wrapping up the report.

5.2 Results after balancing

As a last resort, it was attempted to balance the amount of data in each bin. This led to a training accuracy of 0.36 averaged over 5 folds and a testing accuracy of 0.18. Per-class accuracy is reported in Table 4, where the amount of samples per class has also been reported. An averaged total accuracy of 0.18 is almost twice as bad as a random class assignment accuracy of 0.33 and signifies a worthless model.

	Label 0	**Label 1**	**Label 2**
Accuracy	0.20	0.32	0.48
Datacount	200	204	196
Bins	0.0-0.013	0.013-0.017	0.017-0.03

Table 4: Per class accuracy of balanced model.

6 Further work

Due to the crucial mistake of not checking the per-class accuracy and taking into account the imbalance of the data set, the final model performance is poor. This is mostly because of the lack of both knowledge in the field of neurology and guidance in the preprocessing of the data. As mentioned, preprocessing was a big chunk of the project's workload and too little time was left to explore different ML approaches after the models flaws were discovered. Nevertheless, several steps could be taken to obtain better model performances.

First of all, feeding more data into the algorithms will improve the performance of the ML-models. By not only considering one rodent which is suitable for analysis (i.e. OB9), the models can be generalized to multiple rodents which is desired from scientific point of view. Generating more samples per rodent in Matlab is also a way to get more data, but this would require a high performance computing cluster.

Within the scope of this project, only the bar press rate was used as a fear quantifier. In further work, one could also label the input data according to freezing scores and acceleration data to get a more detailed idea of the fearfulness of the rodent at a certain moment.

Finally, one could extract other features useful in fear decoding from the LFP recordings then the PAC and the BLA/IL power spectra. As can be read in the conclusion paragraph, these features do not seem very relevant, so by extracting other features the model performance will go up.

Annexes

Preprocessing of data details

PAC MI image extraction:

1. Bandpass each channel to obtain theta frequencies (4-12 Hz) with the Matlab function `bandpass`, with parameters 'iir', and 'Steepness'=1
2. Bandpass each channel to obtain gamma frequencies (50-100 Hz) with the Matlab function `bandpass`, with parameters 'iir', and 'Steepness'=1
3. Average the 4 channels – time signals for theta frequencies
4. Average the 4 channels – time signals for gamma frequencies
5. Obtain the PAC Modulation Index as an image using the Matlab toolbox for calculating PAC , and specifically the function: `find_pac_shf(gamma, fs=30000, 'MI', theta, theta_f, gamma_f)`
6. The analyzed frequencies for the PAC MI analysis were defined as: `theta_f = 4:0.2:12` Hz, `gamma_f = 40:1:150` Hz, in order to have a larger visualization

BLA normalized power extraction:

1. Bandpass each channel to obtain the interest frequencies (4-100 Hz) with the Matlab function bandpass, with parameters 'iir', and 'Steepness' = 1
2. Average the 4 channels in the time domain
3. Find the normalized power spectrum with the Matlab function `abs(fft(.))/length(.)`

IL normalized power extraction:

1. Bandpass each channel to obtain the interest frequencies (4-100 Hz) with the Matlab function bandpass, with parameters 'iir', and 'Steepness' = 1
2. Average the 4 channels in the time domain
3. Find the normalized power spectrum with the Matlab function `abs(fft(.))/length(.)`

Smooth bar press rate: the 1200 seconds time signal containing 2400 samples was divided into 600 time-windows of 2 seconds and the average of the 100 samples of each time-window was calculated. That average would be the label for each of the window.

6.1 Notes on Matlab bandpass function:

- The passband ripple is 0.1 dB
- The stopband attenuation is 60 dB
- 'iir' parameter: the function designs a minimum-order infinite impulse response (IIR) filter
- 'Steepness' = 1 ensures an almost ideal filter with fstop = fpass*(1+/-0.01), which implies a 1% of un-ideality

References

1. Sullivan, C. R., Olsen, S. & Widge, A. S. Deep brain stimulation for psychiatric disorders: From focal brain targets to cognitive networks. *NeuroImage* **225,** 117515. ISSN: 1053-8119 (Jan. 2021).
2. Fleming, J. E., Dunn, E. & Lowery, M. M. Simulation of Closed-Loop Deep Brain Stimulation Control Schemes for Suppression of Pathological Beta Oscillations in Parkinson's Disease. *Frontiers in Neuroscience* **14,** 166. ISSN: 1662453X (Mar. 2020).
3. Bloodgood, D. W., Sugam, J. A., Holmes, A. & Kash, T. L. Fear extinction requires infralimbic cortex projections to the basolateral amygdala. *Translational Psychiatry 2018 8:1* **8,** 1–11. ISSN: 2158-3188. `https://www.nature.com/articles/s41398-018-0106-x` (Mar. 2018).
4. Schönfeld, L. M. & Wojtecki, L. Beyond emotions: Oscillations of the amygdala and their implications for electrical neuromodulation. *Frontiers in Neuroscience* **13,** 366. ISSN: 1662453X (Apr. 2019).
5. Buzśaki, G. & Wang, X. J. Mechanisms of gamma oscillations. *Annual review of neuroscience* **35,** 203–225. ISSN: 1545-4126. `https://pubmed.ncbi.nlm.nih.gov/22443509/` (July 2012).
6. Stujenske, J. M., Likhtik, E., Topiwala, M. A. & Gordon, J. A. Fear and safety engage competing patterns of theta-gamma coupling in the basolateral amygdala. *Neuron* **83,** 919. ISSN: 10974199. `/pmc/articles/PMC4141236/%20/pmc/articles/PMC4141236/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4141236/` (Aug. 2014).
7. Lucasius, C. Noninvasive EEG-Based Prediction of Infantile Spasms Using Deep Learning Networks with Phase Amplitude Cross-Frequency Coupling. *Master Thesis University of Toronto* (2020).