Selection of relevant features for alpha rhythm classification

Matjaž Zupančič Muc
Human Computer Interaction 22/23, FRI, UL
mm1706@student.uni-lj.si

***Abstract*—In this work, we implement a feature selection technique utilizing R^2 feature maps to identify the most significant frequencies and channels used for classification of alpha rhythms. Our methodology was applied to data from a subject in the EEG MMIDB database, the resulting R^2 map is visualized and analysed. We show that the most informative set of frequencies lie around the frequency of 9.9Hz in the visual cortex of the brain.**

I. Introduction

Classification of alpha rhythms in electroencephalography (EEG) signals is an important task in the field of neuroscience and clinical research. However, it can be challenging to identify the most relevant features in the signal for accurate classification. In this work, we implement a feature selection method based on R^2 feature maps to select the most informative set of frequencies and channels in EEG signals for classifying alpha rhythms. We demonstrate the effectiveness of our method on a subject from the EEG MMIDB database.

II. Related Work

The authors of [1] propose using a feature selection method based on R-square coefficients to identify the EEG features that are most relevant for classifying different types of motor imagery. They evaluate the performance of their method using a dataset of EEG signals recorded from subjects who were instructed to imagine moving their left or right hand. They found that their method led to a significant improvement in classification accuracy compared to traditional methods such as the common spatial patterns (CSP) algorithm.

III. Methodology

The input to our system is a set of raw EEG baseline signals, in particular we used subject S001 (records 1, 2) from the EEGMMI Database ([2]). Record 1 corresponds to an experiment where the subject kept their eyes opened the entire time, while record 2 corresponds to an experiment where the subject kept their eyes closed the entire time. Signals are sampled at a frequency of $F_s = 160$Hz. Each signal in the database is composed of 64 channels.

When loading in the records we are left with two matrices X_1 (corresponding to the signal where subject kept their eyes opened) and X_2 (corresponding to the signal where subject kept their eyes closed) of shape $(64 \times M)$, where M is the number of data points in each record.

Using a band-pass filter we remove noise from the input signals. We chose a band-pass filter which cuts off frequencies outside the $(3-30)$Hz band.

We split X_1 and X_2 into intervals that are roughly 3.5 seconds long. The interval length is denoted with $m = 3.5F_s$. Lets denote the interval tensors as $\tilde{X}_1$ and $\tilde{X}_2$, they are both of the following shape $(\#\text{intervals} \times C \times m)$.

To compute the R^2 feature map we iterate over all channels C and compute the power spectrum for each interval in $\tilde{X}_1$ and $\tilde{X}_2$. The R^2 signal for the current channel C is computed using the following equation:

$$R^2 = \frac{(s_1 - s_2)^2}{2N(q_1 + q_2) - (s_1 + s_2)^2 + \epsilon} \tag{1}$$

Where $s_1 = \sum_{i=0}^{N} x_i^{(1)}$, $s_2 = \sum_{i=0}^{N} x_i^{(2)}$, $q_1 = \sum_{i=0}^{N} x_i^{2(1)}$, $q_2 = \sum_{i=0}^{N} x_i^{2(2)}$. N is the number of intervals and $\epsilon = 10e^{-10}$ (used for numeric stabilization). Note that $x_i^{(1)}$ and $x_i^{(2)}$ denote the power spectrum of record 1 and the record 2 computed for the i-th interval. Once we compute R^2 for all channels C we are left with a matrix of shape $(C \times \frac{F_s}{2})$.

IV. Background

A band-pass filter is a type of filter that allows only a specific range of frequencies to pass through while attenuating (reducing the amplitude of) frequencies outside of this range. The range of frequencies that are allowed to pass through is called the ”passband” of the filter. They are used to eliminate noise from a signal, by eliminating unwanted frequencies outside of the desired passband.

The power spectrum of a signal is a representation of the signal’s frequency content. It shows the distribution of power among the various frequencies that are present in the signal. It is calculated by taking the squared magnitude of the signal’s Fourier transform. If signal corresponding to closed eyes is different to the signal corresponding to opened eyes, its corresponding power spectrum’s are also different. Therefore we can use the power spectrum in the calculation of R^2 signal.

In the context of EEG feature extraction, the R^2 metric can be used as a way to identify the EEG features that are most closely related to the imagined mental task, by evaluating the correlation between the EEG features and the corresponding mental task. The intuition behind the use of R^2 as a metric is that a high R^2 value indicates that a large proportion of the variance in the outcome variable can be explained by the predictor variables (EEG features) and, thus, that the EEG features are strongly correlated with the mental task. Conversely, a low R^2 value indicates that there is not a strong correlation between the EEG features and the mental task.

The R^2 feature map is a function of frequency and channel number. Each row in the feature map is an R^2 signal (computed using Equation 1). In the context of EEG feature extraction it is used to select the set of most informative channels and frequencies.

V. Experiments

Using subject S001 (record 1 and 2) from the EEGMMI Database ([2]), we first split signals of both records into 15 intervals, which are roughly 3.5 seconds long. We compute and display mean power spectrum's and the corresponding R^2 signals for a subset of channels. We compute, display and analyse the R^2 map.

VI. Results and Discussion

A. Results

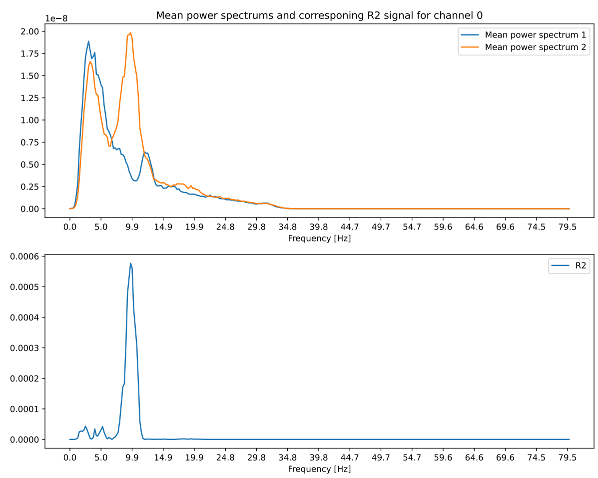

Fig. 1. Mean power spectrum and the corresponding R^2 signal for channel 0. Blue curve on the top corresponds to the mean power spectrum for signals where the subject kept their eyes opened. Orange curve corresponds to the mean power spectrum for signals where the subject kept their eyes closed.

B. Discussion

Figure 1 shows the mean power spectrum and the corresponding R^2 signal for channel number 0. We can see that the R^2 signal is the highest where the difference between power spectrum's is the largest (and the variation of interval signals is the smallest), that happens at a frequency of around 9.9Hz. Figure 2 and Figure 3 show the same plot for channel number 20 and 40 respectively. We can see that the corresponding R^2 signal is almost identically to the one in channel number 0. Figure 4 shows the same plot for channel number 60, we

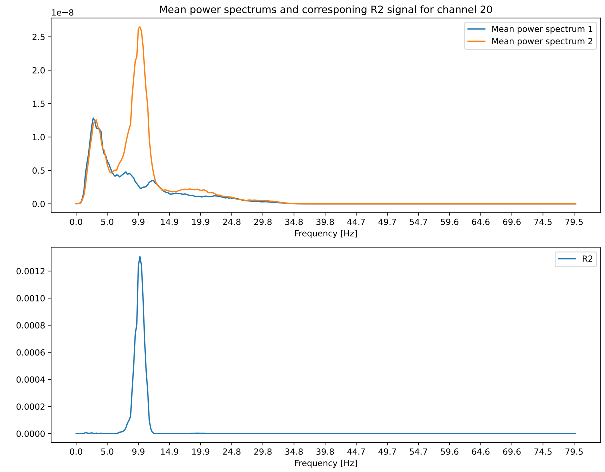

Fig. 2. Mean power spectrum and the corresponding R^2 signal for channel 20. Blue curve on the top corresponds to the mean power spectrum for signals where the subject kept their eyes opened. Orange curve corresponds to the mean power spectrum for signals where the subject kept their eyes closed.

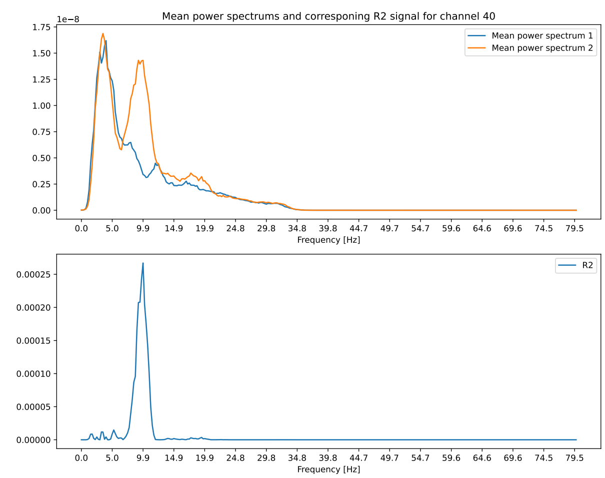

Fig. 3. Mean power spectrum and the corresponding R^2 signal for channel 40. Blue curve on the top corresponds to the mean power spectrum for signals where the subject kept their eyes opened. Orange curve corresponds to the mean power spectrum for signals where the subject kept their eyes closed.

can see that amplitude of the R^2 is substantially higher than in other channels, the overall difference in the mean power spectrum's is also the highest. We can see that the blue curve (corresponding to opened eyes) is almost constant and of a low amplitude, while the orange curve (corresponding to closed eyes) has a large amplitude around frequency of 9.9Hz.

Figure 6 shows the R^2 map as a function of channel number and frequency, we can see that the most informative channels are channels $60, 61, 62, 63, 64$ and frequencies around 9Hz. From Figure 5 we can see that channels $60, 61, 62, 63, 64$ correspond to the visual cortex of the brain.

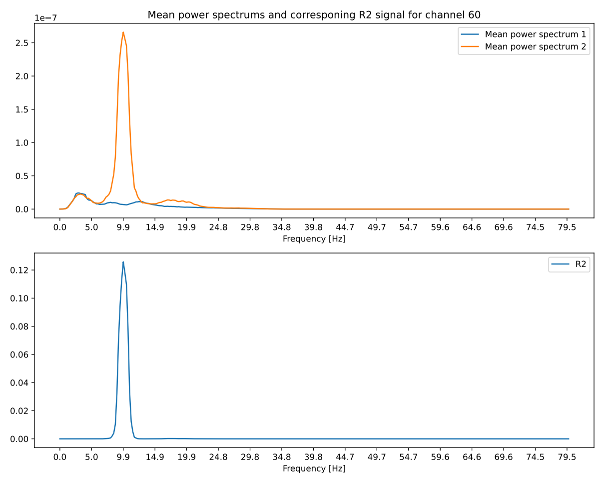

Fig. 4. Mean power spectrum and the corresponding R^2 signal for channel 60. Blue curve on the top corresponds to the mean power spectrum for signals where the subject kept their eyes opened. Orange curve corresponds to the mean power spectrum for signals where the subject kept their eyes closed.

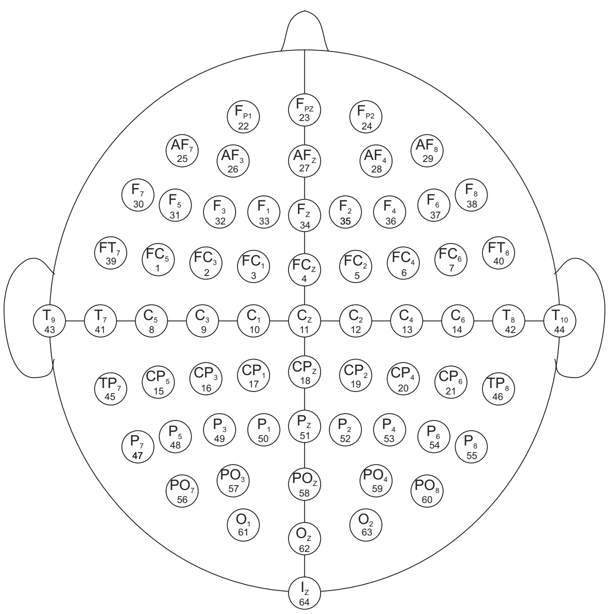

Fig. 5. Physical location of electrodes for each channel (https://physionet.org/content/eegmmidb/1.0.0/64_channel_sharbrough.pdf)

VII. Conclusion

We implemented the R^2 feature extraction method. We evaluated the method on a subject from the EEGMMI Database using the signals which correspond to the opened and closed eyes. We show that the most informative set of frequencies lie around the frequency of 9.9Hz in the visual cortex of the brain. Further work could evaluate the method over a larger number of subjects and check if the observations still hold.

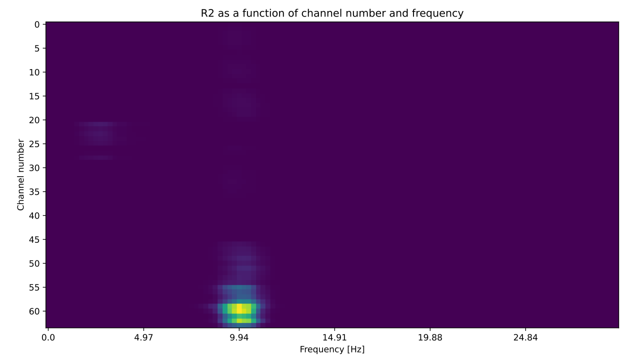

Fig. 6. R^2 feature map as a function of frequency and channel number. We can see that the most informative channels are channels $60, 61, 62, 63, 64$ and frequencies around 9 Hz.

References

[1] P. Chum, S.-M. Park, K.-E. Ko, and K.-B. Sim, "Optimal eeg feature extraction based on r-square coefficients for motor imagery bci system," in *2012 12th International Conference on Control, Automation and Systems*. IEEE, 2012, pp. 754–758.

[2] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov, R. Mark, J. Mietus, G. Moody, C.-K. Peng, and H. Stanley, "Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.