Graph models for constructing the connectivity map of functional groups

Panchenko S. K., Strijov V. V., Varenik N. V.

The problem of constructing the model of human brain activity with respect to the spatial structure of the signal is considered. The brain activity is described by multidimensional temporal series extracted from electrodes, arranged on the patient's head in accordance with one of the universal arrangement schemes. Classical convolutional neural networks can not be utilised effectively to account for the spatial information due to absence of a regular definition of a neighbourhood on a spherical surface of the brain. A graph representation of the signal for identifying the interactions between different activity zones in space and for neurobiological interpretation of the functional connections of the brain is proposed. Different methods of constructing the connectivity matrix defining the graph structure are studied. Deterministic methods of estimating linear relationship between time series based on cirrelations, spectral analysis, autoregression approach and nonlinear phase synchronisation are employed for evaluating the connectivity matrix. A composition model of graph convolution for aggregating spatial information and recurrent block for time series processing is proposed for decoding.

Keywords: *spatio-temporal structure, multidimensional temporal series, electroencephalography, graph representation of the signal, graph convolution, spectral coherence, частично направленная когерентность, phase synchronization, emotion recognition.*

1 Introduction

Intent-oriented decoding of human brain signals is crucial for the technology of Brain Computer Interface (BCI), introduced by McFarland and Wolpaw (2011). Works of Chaudary et al. (2021), Collinger et al. (2013), Gemein et al. (2020) and Ali et al. (2016) provide an overview of BCI applications, including but not limited to neuroprosthesis control, rehabilitation, diagnostics and emotion analysis. This technology captures the activity of the brain and translates it into commands, as shown in McFarland (2011).

One of the simplest, fastest and most available methods of obtaining the signals of brain activity is electroencephalogram (EEG), registering the spontaneous electrical activity of the brain over a certain period of time, using a semi-spherical grid of electrodes placed on the patient's head. These signals have been successfully decoded for a variety of classification tasks with the help of convolutional neural networks (CNNs). Generally, the leading solutions in the field involve CNN-based models. Gao et al. (2019), for instance, interprets EEG signals as images in the space of electrode indices and time and applies two-dimensional convolutional filters to them. A different approach is described in Mattioli et al. (2022), where signals from different electrodes are independently processed by the model via one-dimensional filters. Yet another method involves performing 2D convolution in the space of electrodes alone by approximating the spatial arrangement of the electrodes with a planar grid, as in Zhang et al. (2018).

These solutions, while having proven to be effective for particular decoding tasks, appear to share the same drawback: they disregard to different extents the three-dimensional spatial structure of the brain signals, partially accounted for by the spherical arrangement of EEG electrodes. However, human brain is an incredibly complex dynamic system, constantly exchanging information between different inter-connected regions. These regions, called functional activity groups of the brain and described by Power et al. (2011), form a network with hierarchical, spatial and structural organisation. We believe that this functional group network can be essential for our general understanding of the brain and, more importantly, for the purpose of obtaining informative low-dimensional features of human brain signals. As will be shown in the following sections, capturing this spatial structure is possible with graph representation of the signal, subsequently processed by a graph neural network (GNN). What is more, with such representation we are capable of outperforming conventional CNN methods on emotion recognition tasks, while also preserving interpretability.

2 Materials and Methods: problem statement

2.1 Connectivity map of the functional groups of the brain

The EEG signal is given in a matrix form:

$$\underline{\mathbf{X}} = \big[\mathbf{X}_m\big]_{m=1}^{M},\ \mathbf{X}_m \in \mathbb{R}^{E\times N},$$

where N corresponds to the time measurements within one experiment, E is the number of electrodes, recording the signal, , and M is the number of experiments. We are also provided with a matrix of electrode coordinates $\mathbf{Z} \in \mathbb{R}^{E\times 3}$, defined by the chosen standard of electrode placement.

We propose a representation of EEG signal with a non-oriented dynamic graph

$$\mathcal{G}(m,t) = \Big(\mathcal{V}(m,t), \mathcal{E}(m,t), \mathbf{A}_{\underline{\mathbf{X}},\mathbf{Z}}(m,t)\Big),$$

capturing irregular spatial and functional connections between electrodes. Vertices of this graph $\mathcal{V}(m,t)$ are the electrodes themselves, attaining values of the signal at the moment t in the experiment m. The edges of this graph $\mathcal{E}(m,t)$ are defined by a connectivity matrix $\mathbf{A}_{\underline{\mathbf{X}},\mathbf{Z}}(m,t)$.

Our goal is to explore different methods of constructing such a connectivity matrix

$$\mathbf{A}_{\underline{\mathbf{X}},\mathbf{Z}}(m,t) = \big[a_{ij}(m,t)\big]_{i,j=1}^{E,E}$$

in a form of a symmetric real matrix with non-negative elements $a_{ij}(m,t) \geqslant 0$ — weights of the edges between vertices i и j.

Obtaining a connectivity matrix of the graph $\mathbf{A}_{\underline{\mathbf{X}},\mathbf{Z}}(m,t)$ for the given indices of time and experiment is equivalent to finding a connectivity map, a function of the following form:

$$\mathbf{A}_{\underline{\mathbf{X}},\mathbf{Z}}(m,t) : M \times T' \to \mathbb{R}_{+}^{E\times E},\ T' \subseteq T,\ T = \{t_n\}_{n=1}^{N}. \tag{1}$$

2.2 Decoding the signals of the human brain: classification task

A sample of human brain activity $\mathfrak{D} = (\underline{\mathbf{X}}, \mathbf{Z}, \mathbf{y})$ for the classification task is given, where

$$\begin{gathered}
\underline{\mathbf{X}} = \big[\mathbf{X}_m\big]_{m=1}^{M} \text{ — EEG signals,}\\
\mathbf{X}_m = [\mathbf{x}_t]_{t\in T} \text{ — multidimensional signal, obtained in the } m\text{-th experiment,}\\
\mathbf{x}_t \in \mathbb{R}^E \text{ — measurements of the multidimensional signal at the moment } t,\\
\mathbf{Z} = [\mathbf{z}_k]_{k=1}^{E}, \mathbf{z}_k \in \mathbb{R}^3 \text{ — electrode coordinates,}\\
\mathbf{y} = [y_m]_{m=1}^{M} \text{ — values of the target,}\\
y_m \in \{1,\dots C\} \text{ — class index,}\\
T = \{t_n\}_{n=1}^{N} \text{ — sequence of the time measurements,}\\
E = 62 \text{ — total number of the electrodes,}\\
N \text{ — total number of time measurements in one signal.}
\end{gathered} \tag{2}$$

The following penalty function is applied to the connectivity matrix to ensure the stability of the soluton:

$$\underline{\mathbf{A}}^{*}_{\underline{\mathbf{X}},\mathbf{Z}} = \arg\min_{\underline{\mathbf{A}}_{\underline{\mathbf{X}},\mathbf{Z}}} \Big|\,||\underline{\mathbf{A}}_{\underline{\mathbf{X}},\mathbf{Z}}||_1 - p\Big|, \tag{3}$$

where p is a parameter corresponding to the degree of matrix sparsity.

For the final task of signal decoding a graph recurrent neural network model is employed:

$$h_{\boldsymbol{\theta}} : (\underline{\mathbf{X}}, \underline{\mathbf{A}}^*_{\underline{\mathbf{X}},\mathbf{Z}}) \to \mathbf{y}. \tag{4}$$

Multi-class cross-entropy is the natural choice for the loss-function:

$$\begin{gathered} \mathcal{L} = -\frac{1}{M}\sum_{m=1}^{M}\left[\sum_{c=1}^{C} \mathbb{1}(y_m = c)\log(p_m^c)\right], \text{ where} \\ p_m^c = h_{\boldsymbol{\theta}}\Big(\mathbf{X}_m, \underline{\mathbf{A}}^*_{\underline{\mathbf{X}},\mathbf{Z}}(m)\Big) \quad \text{— probability of the class } c \\ \text{for } \mathbf{X}_m \text{ with a connectivity matrix } \underline{\mathbf{A}}^*_{\underline{\mathbf{X}},\mathbf{Z}}(m). \end{gathered} \tag{5}$$

The optimal parameters of the model are obtained through optimization of the loss function:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}, \mathbf{X}, \underline{\mathbf{A}}^*_{\underline{\mathbf{X}},\mathbf{Z}}). \tag{6}$$

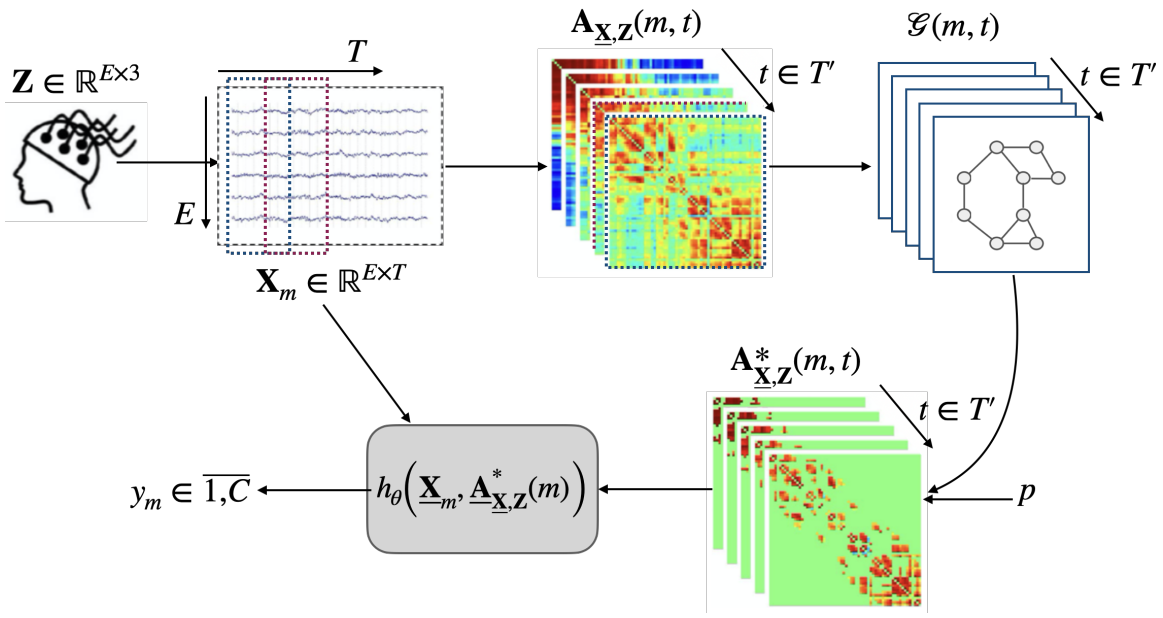

Fig. 1 Solving the human brain activity classification task with a connectivity map of the functional groups of the brain. The signals of the EEG are used to obtain a set of connectivity maps $\mathbf{A}_{\underline{\mathbf{X}},\mathbf{Z}}(m,t)$ graphs $\mathcal{G}(m,t)$ for all m and t. Optimal matrices $\mathbf{A}^*_{\underline{\mathbf{X}},\mathbf{Z}}$ in terms of the introduced penalty are then found. These matrices and the initial signals $\mathbf{X}_m$ finally serve as input to the graph neural network, outputting the probabilities of the classes.

3 Materials and methods: obtaining the connectivity map

The connectivity matrix determines the topological structure of the EEG signal and is a necessary component for training the graph neural network. In this section we describe several methods of constructing this matrix through the estimation of multidimensional signal component correlation. Both linear and nonlinear approaches are considered. Beside the classical methods, such as Euclidean distance and Pearson correlation coefficient, we also investigate spectral coherence, partial directed coherence and phase synchronization methods for capturing of potential nonlinear dependencies in the data.

3.1 Euclidean distance and linear correlation

Euclidean distance: the coordinates of electrodes are constant, therefore we denote $d_{ij}(m,t) = d_{ij}$, where

$$d_{ij} = ||\mathbf{z}_i - \mathbf{z}_j||_2^2,\ i,j \ \text{— electrode indices.} \tag{7}$$

We obtain the connectivity matrix as follows:

$$\mathbf{A}^*_{\underline{\mathbf{X}},\mathbf{Z}}(m,t) = \big[a_{ij}\big] \in {\mathbb{R}_+}^{E\times E},\ a_{ij} = \begin{cases} d_{ij}, & \text{if } d_{ij} \leqslant \rho(p) \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

Linear Pearson correlation: we denote the rows of the matrix $\mathbf{X}_m$, corresponding to signals on i-th и j-th electrodes on a time segment $[t_n - T_w,\ t_n]$ in the m-th experiment as $\mathbf{x} = \mathbf{x}_{mi}$ and $\mathbf{y} = \mathbf{x}_{mj}$. The correlation coefficient is then given by the following expression:

$$\tilde{r}_{ij}(m,t_n) = \frac{\sum\limits_{k=t_n-T_w}^{t_n}(x_k - \overline{\mathbf{x}})(y_k - \overline{\mathbf{y}})}{\sqrt{s^2_{\mathbf{x}} s^2_{\mathbf{y}}}},\ \text{where } x_k = (\mathbf{x})_k,\ y_k = (\mathbf{y})_k,$$
$$\overline{\mathbf{x}}, \overline{\mathbf{y}}, s^2_{\mathbf{x}}, s^2_{\mathbf{y}} \ \text{— sample average and sample variance}$$
$$\text{for } i\text{-th and } j\text{-th electrodes respectively. Finally,}$$
$$\tilde{r}_{ij}(m,t_n) = \frac{\sum\limits_{k=t_n-T_w}^{t_n}(x_k - \overline{\mathbf{x}})(y_k - \overline{\mathbf{y}})}{\sqrt{\sum\limits_{k=t_n-T_w}^{t_n}(x_k - \overline{\mathbf{x}})^2 \sum\limits_{k=t_n-T_w}^{t_n}(y_k - \overline{\mathbf{y}})^2}} \ \text{and} \tag{9}$$
$$\mathbf{A}^*_{\underline{\mathbf{X}},\mathbf{Z}}(m,t) = \big[a_{ij}(m,t)\big] \in {\mathbb{R}_+}^{E\times E},\ a_{ij}(m,t) = \begin{cases} r_{ij}(m,t), & \text{if } r_{ij}(m,t) \geqslant \rho(p) \\ 0, & \text{otherwise,} \end{cases}$$
$$\text{where } r_{ij}(m,t) = |\tilde{r}_{ij}(m,t)|.$$

3.2 Spectral coherence

We denote the continuous stochastic processes with realizations $\mathbf{x}$ and $\mathbf{y} = \mathbf{x}_{mj}$ as $x(t)$ and $y(t)$, where $\mathbf{x} = \mathbf{x}_{mi}$ and $\mathbf{y}$ are again the rows of the matrix $\mathbf{X}_m$, corresponding to signals on i-th и j-th electrodes on a time segment $[t_n - T_w,\ t_n]$ in the m-th experiment.

Following Santos Filho (2005), we define spectral coherence between two signals as follows:

$$\gamma_{ij}(m,t_n,f) = \frac{|S_{xy}(t_n,f)|^2}{S_{xx}(t_n,f)S_{yy}(t_n,f)}, \tag{10}$$

where the spectral density functions S are Fourier transforms of the respective correlation functions, according to the Wiener-Khinchin theorem (1989).

In particular, the auto-spectral density function:

$$S_{xx}(t_n, f) = \int\limits_{(t_n - T_w)}^{t_n} \left(1 - \frac{\tau}{T_w}\right) R_{xx}(\tau) e^{-i2\pi f\tau} d\tau. \tag{11}$$

Similarly, the cross-spectral density function:

$$S_{xy}(t_n, f) = \int\limits_{(t_n - T_w)}^{t_n} \left(1 - \frac{\tau}{T_w}\right) R_{xy}(\tau) e^{-i2\pi f\tau} d\tau. \tag{12}$$

The correlation function is given by

$$R_{xy}(\tau) = \frac{1}{T_w} \int\limits_{(t_n - T_w)}^{t_n} x(t)\overline{y(t-\tau)} dt, \text{ where } \tau \text{ is the correlation lag.} \tag{13}$$

Finally, we integrate the spectral coherence over a frequeny range $[f_1, f_2]$ and obtain the connectivity map:

$$\gamma_{ij}(m, t_n) = \int\limits_{f_1}^{f_2} \gamma_{ij}(m, t_n, f) df,$$
$$\mathbf{A}^*_{\underline{\mathbf{X}},\mathbf{Z}}(m,t) = \left[a_{ij}(m,t)\right] \in {\mathbb{R}_+}^{E\times E},\ a_{ij}(m,t) = \begin{cases} \gamma_{ij}(m,t), & \text{if } \gamma_{ij}(m,t) \geqslant \rho(p) \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

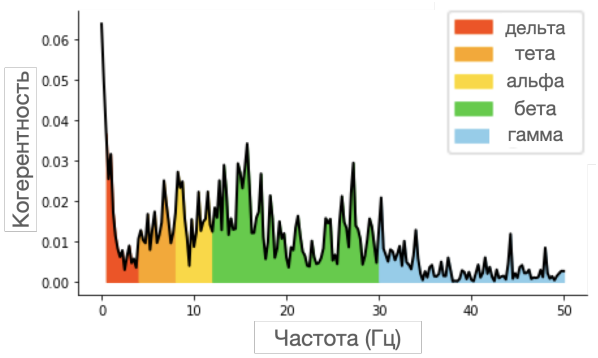

Fig. 2 Typical signal spectral coherence over several frequency ranges. [comment: the figure labels are not translated in the draft]

3.3 Partial directed coherence

...[comment: section omitted in the draft]...

3.4 Phase synchronization

...[comment: section omitted in the draft]...

4 Materials and methods: constructing the decoding model

For simplicity in this section we omit the experiment index m and introduce the following notation at the time step t:

$$\mathcal{G}(m,t) = \mathcal{G}_t,\ \mathcal{V}(m,t) = \mathcal{V}_t,\ \mathcal{E}(m,t) = \mathcal{E}_t,$$
$$\mathbf{A}^*_{\underline{\mathbf{X}},\mathbf{Z}}(m,t) = \mathbf{A}^*_t,$$
$$\mathbf{x}_{mt} = \mathbf{x}_t.$$

4.1 GCN LSTM model

For aggregating the spatial information in the data and capturing the temporal dependencies of the signal we consider a composition of a graph convolution network and an LSTM recurrent

neural network. A detailed overview of the graph convolution is presented in the works of Defferrard et al. (2016) and Wu et al. (2020). Similarly, a description of LSTM model is provided in Sundermeyer et al. (2012).

The input $\mathbf{x}_t$ is processed by the model as follows:

$$\begin{aligned}
\mathbf{x}_t^{\mathrm{GCN}} &= \mathrm{GConv}(\mathbf{x}_t),\\
\mathbf{i}_t &= \sigma(\mathbf{W}_i\mathbf{x}_t^{\mathrm{GCN}} + \mathbf{V}_i\mathbf{h}_{t-1} + \mathbf{w}_i \odot \mathbf{c}_{t-1} + \mathbf{b}_i),\\
\mathbf{f}_t &= \sigma(\mathbf{W}_f\mathbf{x}_t^{\mathrm{GCN}} + \mathbf{V}_f\mathbf{h}_{t-1} + \mathbf{w}_f \odot \mathbf{c}_{t-1} + \mathbf{b}_f),\\
\mathbf{c}_t &= \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tanh(\mathbf{W}_c\mathbf{x}_t^{\mathrm{GCN}} + \mathbf{V}_c\mathbf{h}_{t-1} + \mathbf{b}_c),\\
\mathbf{o}_t &= \sigma(\mathbf{W}_o\mathbf{x}_t^{\mathrm{GCN}} + \mathbf{V}_o\mathbf{h}_{t-1} + \mathbf{w}_o \odot \mathbf{c}_t + \mathbf{b}_o),\\
\mathbf{h}_t &= \mathbf{o}_t \odot \tanh(\mathbf{c}_t),
\end{aligned} \tag{15}$$

where $\mathbf{x}_t \in \mathbb{R}^{E\times D}$ is the observed signal of the dynamic system at the moment t, the system is described by a graph $\mathcal{G}_t$, the vectors $\mathbf{i}_t, \mathbf{f}_t, \mathbf{o}_t$ are the parameters of LSTM , and $\mathbf{h}_{t-1},\ \mathbf{c}_{t-1} \in \mathbb{R}^{E\times H}$ is the hidden state of the LSTM cell at the moment $t-1$. These hidden states are then finally transformed into the output of the classifier.

4.2 Signal features

Instead of the raw values of the signal we utilize the differential entropy over the preset ranges of frequency, corresponding to different rhythms of human brain activity, as proposed in Duan et al. (2013): delta (1 – 3Hz), theta (4 – 7Hz), alpha (8 – 13Hz), beta (14 – 30Hz), gamma (31 – 50Hz):

$$\mathrm{DE}(Y) = -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \log\Big(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}\Big) dy, \tag{16}$$

$$Y \in \mathcal{N}(\mu, \sigma^2) \text{ is the temporal series.}$$

And so at the moment t the graph signal, with the values of the signal associated with the vertices of the graph, has the dimensionality $\mathbf{x}_t \in \mathbb{R}^{62\times 5}$, where 62 is the number of electrodes in the EEG grid, and 5 is the number of the considered brain rhythms.

5 Results

5.1 The hypothesis of the experiment

Accounting for the spatial and functional structure of the signal increases the predictive capability of the human emotion classifiers.

5.2 The objectives of the experiment

1. Obtain the signal connectivity maps with different methods.
2. Estimate the quality of the proposed spatio-temporal predictive model utilizing the connectivity maps.

5.3 The dataset

The dataset considered is taken from the work of Duan et al. (2013). It contains the results of an experiment of registering human emotional states. In the experiment, 15 participants were watching 4 minute long video fragments meant to invoke certain emotions. These emotions are

Table 1 Emotion class

№	Class index	Name of the fragment
1	Negative	Back to 1942
2	Negative	Earthquake in Tan-Shan
3	Positive	A flirting scientist
4	Positive	Lost in Tailand
5	Positive	Just another Pandora's Box
6	Neutral	World legacy in China

divided into three subgroups, serving as targets for the classification task: positive, neutral and negative. For each participants, three experiments with 15 attempts each were held.

EEG signal was measured by 62 electrodes arranged accoding to the 10-20 system with discretization frequency of 1 kHz. A frequency filter with a range of 0.3 – 50 Hz was applied to the signal in order to reduce noise and remove outliers.

5.4 Obtaining the connectivity matrix

Matrices obtained for different experiments and participants were averaged in order to avoid overfitting. The results are presented in the Figure 3.

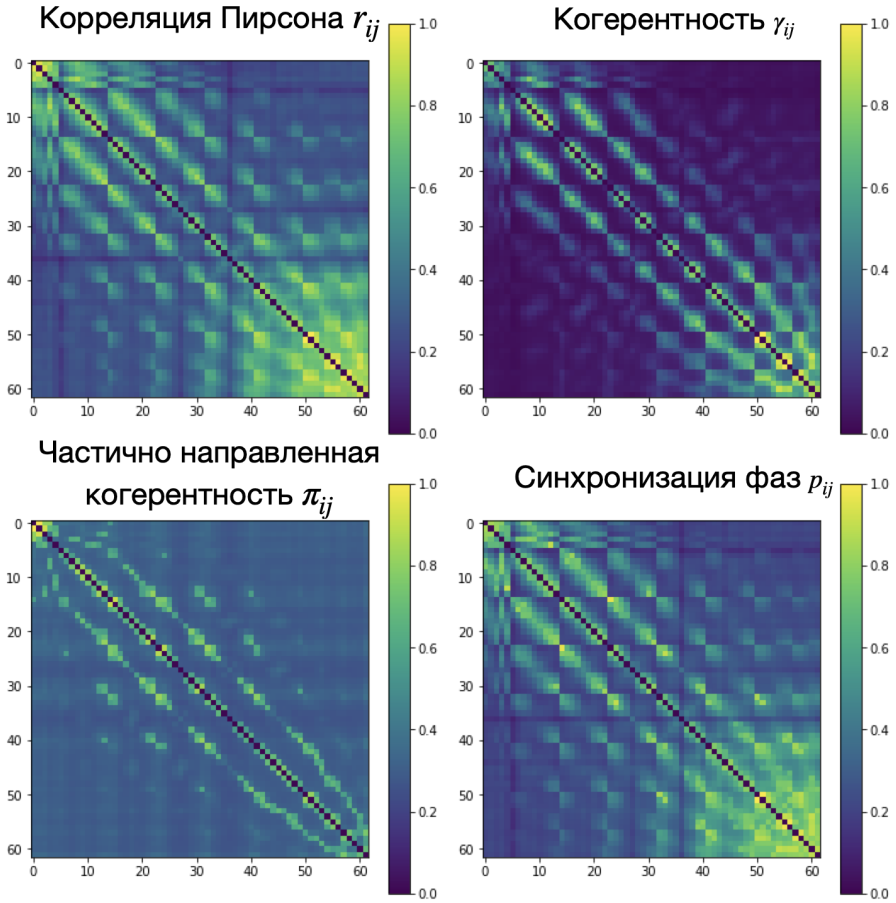

Fig. 3 Comparison of different connectivity map construction methods. The matrices are represented as heatmaps, and the axis correspond to electrode indices. [comment: the figure labels are not translated in the draft]

Obtained estimated of the electrode connectivity correspond to each other with a clearly visible common structure. The phase synchronization method produces a result that is particularly similar to the Pearson correlation method, which might indicate that nonlinear dependencies are not present in the signal.

5.5 Signal decoding results

In the Table 2 we observe gains in terms of classification accuracy of the proposed methods over vanilla LSTM without the graph representation of the signal, and the gains are obtained regardless of the connectivity map construction method used. The best performance is obtained with the linear Pearson correlation method and the Phase synchronization method.

Table 2 Comparison between LSTM and GCN LSTM

Model	Accuracy	Loss
LSTM	0.869 ± 0.010	0.268 ± 0.014
GCN LSTM: d_{ij}	0.894 ± 0.013	0.220 ± 0.012
GCN LSTM: r_{ij}	$\mathbf{0.914 \pm 0.011}$	$\mathbf{0.183 \pm 0.009}$
GCN LSTM: γ_{ij}	0.898 ± 0.010	0.214 ± 0.013
GCN LSTM: π_{ij}	0.898 ± 0.007	0.213 ± 0.012
GCN LSTM: p_{ij}	$\mathbf{0.925 \pm 0.008}$	$\mathbf{0.173 \pm 0.014}$

The difference between the connectivity matrices obtained via Euclidean distance and Phase Synchronization methods is illustrated in the Figure 4 below. A particular group of electrodes having the strongest connections between each other is highlighted. It turns out that these electrodes do correspond to the occipital lobe of the human brain, responsible for processing visual stimuli, and the emotional centers of the brain marked by blue ellipses.

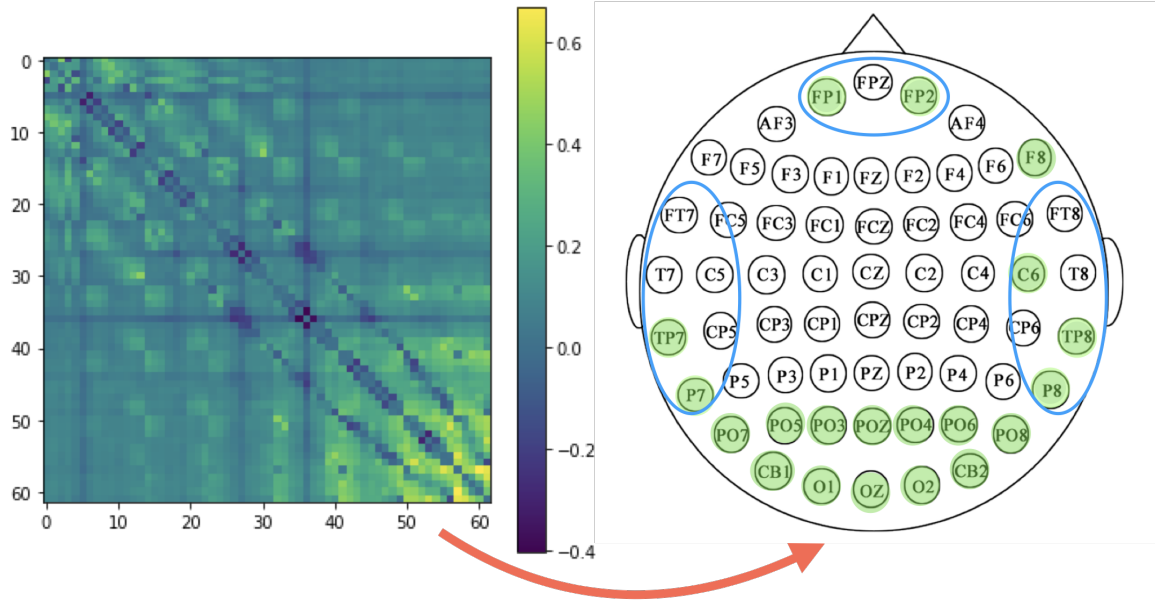

Fig. 4 The difference between the connectivity matrices obtained via Euclidean distance and Phase Synchronization methods (on the left), and the corresponding electrode positions on the EEG grid (on the right). The electrodes demonstrating the strongest connections are situated in the occipital lobe of the brain and in the emotional centers, marked by blue ellipses.

6 Discussion

6.1 Neurobiological interpretation

In the Figure 5 below the connectivity matrices obtained with phase synchronization method are illustrated. The electrodes are arranged in a circle, with the higher part of circle corresponding to the frontal lobe of the brain, and the lower part – to the occipital lobe, and the connections are represented as arcs between the electrodes. We observe a cluster of connections in the frontal lobe for the delta frequency range for the neutral class, and in the parietal and occipital parts for the negative class, which corresponds to the findings of Zheng et al. (2017). Similarly, in alpha range, responsible for relaxed activity and corresponding to the positive class, we observe synchronization in the frontal part of the brain, which is also described in Lee (2014). Another interpretable aspect of the obtained connectivity maps is the activity in side and parietal regions for the neutral and the negative classes, precisely as described in Saarimaki (2016) and Zheng et al. (2015).

7 Conclusion

In this paper graph representation of the EEG signal has been investigated as an alternative to classic convolutional approaches, unable to properly capture the spatio-temporal structure of the human brain functional groups. A comparative analysis has been carried out, demonstrating the superiority of the proposed graph recurrent predictive model. At the same time different methods of the connectivity map construction have been studied. According to the results, the best accuracy is attained with the usage of the Phase Synchronization method, and the connectivity maps provided by this method correspond to neurobiological studies concerned with functionality of different regions of the brain. More advanced methods of applying graph signal representation to the tasks of BCI, such as GRAND, Bronstein et al.(2021) will be studied in the future.

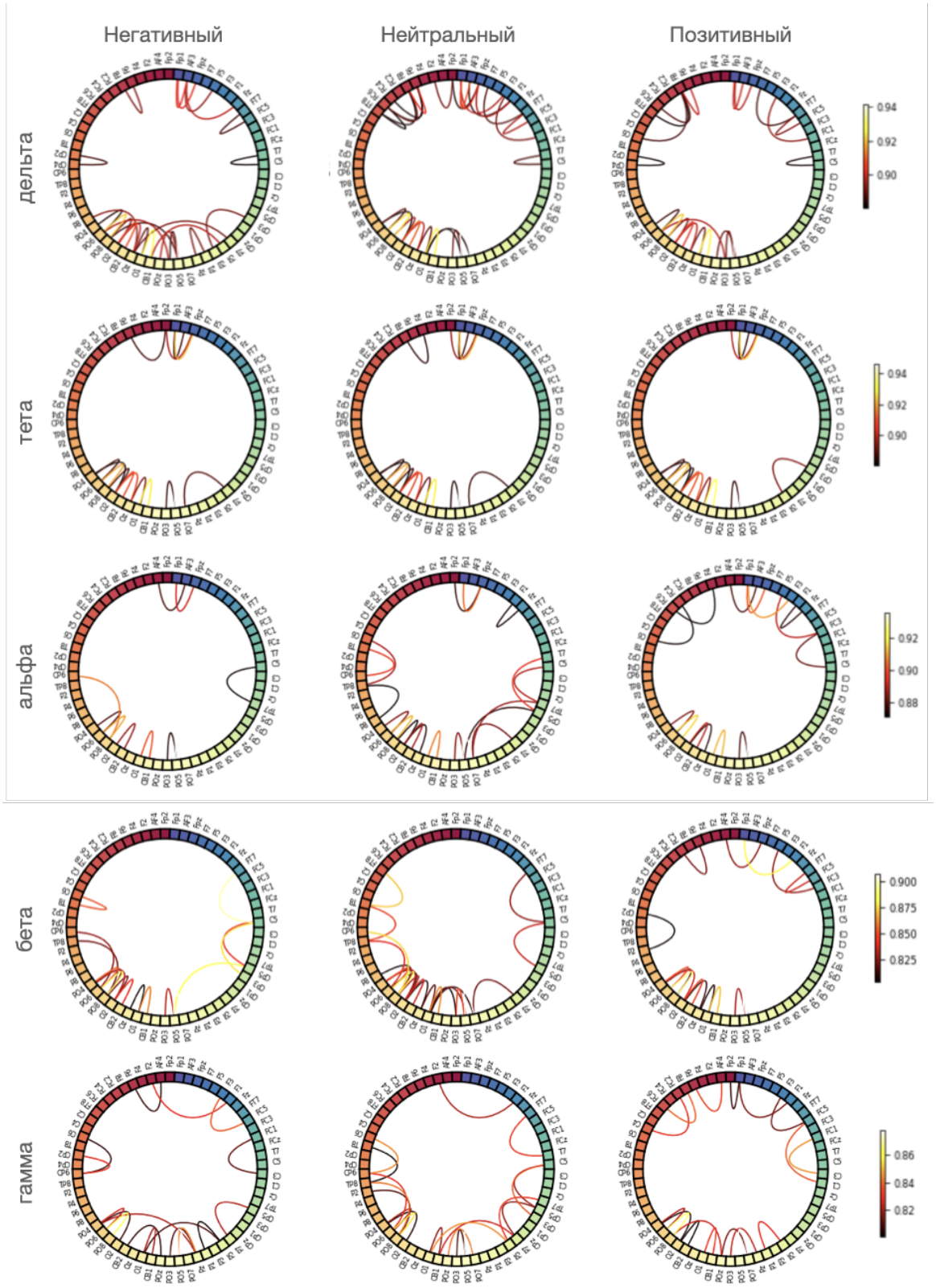

Fig. 5 Phase Synchronization connectivity maps for different frequency ranges and for different classes. Electrodes are arranged in a circle, and arcs correspond to connections between them. [comment: the figure labels are not translated in the draft]