Restoring Functional Brain Groups Using Graph Diffusion Models

A. M. Astakhov, S. K. Panchenko, V. V. Strijov
astakhov.am@phystech.edu; panchenko.sk@phystech.edu; strijov@phystech.edu

This paper addresses the problem of classifying a multivariate time series representing a human brain electroencephalogram (EEG). Standard approaches using two-dimensional convolutions fail to account for the spatial structure of the signal, since the sensors collecting the data are located on a spherical surface. As a solution, we propose using a graph-based representation of functional brain groups and modeling with neural diffusion.

Brain, EEG, Graph Neural Networks, Diffusion Models

1 Introduction

Emotions play a key role in human perception, decision-making, and social interaction. Their automatic classification based on neurophysiological data, such as electroencephalography (EEG), opens up new opportunities in psychology, medicine, affective computing, and human-computer interaction. However, despite significant advances in machine learning and neuroscience, accurate and reliable emotion classification using EEG remains a challenging task. This is due to high individual variability in signals, the nonlinear nature of emotional processes, and the limitations of existing preprocessing and classification methods.

This paper reviews modern approaches to EEG-based emotion recognition, analyzing their advantages and disadvantages, and proposes ways to improve classification accuracy. Special attention is given to signal processing methods, informative feature extraction, and the application of deep learning algorithms. The results of this research can be useful in developing more effective affective interaction systems, neurorehabilitation tools, and psychophysiological studies. The subject of our study is a signal obtained through electroencephalographic examination of the human brain, interpreted as a multivariate time series, where each dimension corresponds to a specific sensor on the subject's head. EEG research is technically constrained by the method's low spatial resolution and high sensitivity to artifacts. As shown in [3], eye movements and muscle activity can significantly distort the signal. Moreover, individual differences in brain activity patterns among subjects greatly reduce the effectiveness of universal classifiers.

Existing approaches are primarily based on two theoretical emotion models: the locationist (basic emotions) and the dimensional (valence-arousal-dominance, VAD) models [10]. However, most modern methods do not take into account spatial relationships between electrodes, which limits their effectiveness.

Various feature extraction methods are used in research:

Temporal features: In [7], six statistical EEG parameters were used followed by channel selection using PCA and ReliefF, achieving an accuracy of 81.87% on the DEAP dataset. However, the authors did not consider spatial correlations between electrodes.

Frequency features: The study [1] demonstrates the effectiveness of PCA for dimensionality reduction, followed by classification using SVM (accuracy of 85.85% on SEED). Similarly, [8] compares various features, showing that statistical characteristics combined with KNN yield an accuracy of 77.54–79%.

A major limitation of these works is that the analysis was performed on each electrode separately, without considering spatial interactions between different brain regions. This is

especially important since emotional states are known to be associated with coordinated activity of distributed neural networks [6].

The goal of our research is to utilize spatial connections between sensors to improve classification quality. We propose treating the time series as a dynamic graph, where edges represent spatial or statistical interconnections between sensors. We believe that accounting for these factors will allow for building a more accurate and robust classification model. We will explore approaches to constructing such connections and assess how they influence classification outcomes. Model performance will be evaluated using the open SEED IV dataset. As the classification model, we propose using DCGRU, which has shown strong results in the related problem of classifying epileptic seizures from EEG data [9].

2 Problem Statement

2.1 Construction of the Adjacency Matrix

The original EEG signal is given as a tensor $\mathbf{X} = [\mathbf{X}_m]_{m=1}^M$, where $\mathbf{X}_m \in \mathbb{R}^{E\times N}$, N corresponds to the number of time samples in the signal, E is the number of electrodes capturing the signal, and M is the number of trials. Additionally, the coordinate matrix of the electrodes is provided as $\mathbf{Z} \in \mathbb{R}^{E\times 3}$, determined by the EEG electrode placement standard used during recording. In this work, we propose to interpret the signal as an undirected dynamic graph:

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

to address the problem of modeling spatial relationships between electrodes on the subject's head. The set of vertices $\mathcal{V}(m,t)$ corresponds to the electrodes, with signal values at time t assigned to each vertex. The set of edges $\mathcal{E}(m,t)$ is defined by the graph adjacency matrix $\mathbf{A}_{\mathbf{X},\mathbf{Z}}(m,t)$.

2.2 Basic Definitions

We are given a dataset $\mathfrak{D} = (\mathbf{X}, \mathbf{Z}, \mathbf{y})$ of brain activity, where:

- $\mathbf{X} = [\mathbf{X}_m]_{m=1}^M$ — set of EEG signals;
- $\mathbf{X}_m = [\mathbf{x}_t]_{t\in T}$ — signal recorded in the m-th trial;
- $\mathbf{x}_t \in \mathbb{R}^E$ — signal observations at time t;
- $\mathbf{Z} = [\mathbf{z}_k]_{k=1}^E$, $\mathbf{z}_k \in \mathbb{R}^3$ — coordinates of electrodes;
- $\mathbf{y} = [y_m]_{m=1}^M$ — target variable;
- $y_m \in \{1, \ldots, C\}$ — class label;
- $T = \{t_n\}_{n=1}^N$ — set of time steps;
- $E = 62$ — number of electrodes;
- N — number of observations in a signal segment.

To solve the decoding problem, we consider a model from the class of graph recurrent diffusion neural networks:

$$h_\theta : (\mathbf{X}, \mathbf{\Delta}^*_{\mathbf{X},\mathbf{Z}}) \to \mathbf{y}. \tag{1}$$

The cross-entropy function is chosen as the loss function:

$$\mathcal{L} = -\frac{1}{M}\sum_{m=1}^{M}\left[\sum_{c=1}^{C}\mathbf{I}(y_m = c)\log(p_m^c)\right], \tag{2}$$

where $p_m^c = h_\theta\left(\mathbf{X}_m, \mathbf{\Delta}^*_{\mathbf{X},\mathbf{Z}}(m)\right)$ is the probability of class c for input $\mathbf{X}_m$ with adjacency matrix $\mathbf{\Delta}^*_{\mathbf{X},\mathbf{Z}}(m)$.

The parameter optimization problem is defined as:

$$\hat{\theta} = \arg\max_\theta \mathcal{L}(\theta, \mathbf{X}, \mathbf{\Delta}^*_{\mathbf{X},\mathbf{Z}}). \tag{3}$$

3 Adjacency Matrix Construction

This section describes methods for constructing the adjacency matrix by estimating the relationships between time series corresponding to the electrodes. We focus on phase synchronization of the signals.

3.1 Phase Synchronization of Signals

Phase synchronization is an approach for analyzing potential nonlinear dependencies and focuses on the phases of signals. It is assumed that two dynamic systems may exhibit phase synchronization even if their amplitudes are independent. Let $x(t)$ and $y(t)$ denote the dynamic systems corresponding to signal observations $\mathbf{x}_{mi}$ and $\mathbf{x}_{mj}$ in the time interval $[t_n - T_w, t_n]$ of the m-th trial. Phase synchronization is defined as:

$$|\varphi_x(t) - \varphi_y(t)| = \text{const}. \tag{4}$$

To estimate the phase, the analytic signal representation is computed using the Hilbert transform:

$$H(t) = x(t) + i\dot{x}(t), \tag{5}$$

where

$$\dot{x}(t) = \frac{1}{\pi}\text{v.p.}\int_{-\infty}^{\infty}\frac{x(t')}{t-t'}dt' \quad \text{— Hilbert transform of the signal } x(t), \tag{6}$$

with v.p. denoting the Cauchy principal value of the integral.

The phase of the analytic signal is defined as:

$$\varphi(t) = \arctan\left(\frac{\dot{x}(t)}{x(t)}\right). \tag{7}$$

For two signals $x(t)$ and $y(t)$ of equal duration T_w with phases $\varphi_x(t)$ and $\varphi_y(t)$, the phase locking value (PLV) [2] is computed as:

$$p_{ij}(m, t_n) = \left|\frac{1}{T_w}\sum_{k=1}^{T_w}\exp\left(i(\varphi_x(k\Delta t) - \varphi_y(k\Delta t))\right)\right|, \tag{8}$$

where Δt is the time step, and $i = \sqrt{-1}$.

The adjacency matrix is defined as:

$$\mathbf{A}^*_{\mathbf{X},Z}(m,t) = [a_{ij}(m,t)] \in {\mathbb{R}_+}^{E\times E}, \quad a_{ij}(m,t) = \begin{cases} p_{ij}(m,t), & \text{if } p_{ij}(m,t) \geqslant \rho(p), \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

4 Classification Model

To solve the classification task, we propose using the **DCGRU** model [5], which has shown strong performance in EEG-based epileptic seizure classification [9]. We argue that diffusion allows information to propagate across distant graph vertices, improving classification accuracy and making the model more robust to noise—an important property given the highly individual nature of EEG data.

Graph diffusion is modeled via the spectral convolution

$$X_{:,p} \star_{\mathcal{G}} f_\theta = \Phi\, F(\theta)\, \Phi^\top X_{:,p},$$

where

- $L = \Phi\Lambda\Phi^\top$ is the spectral decomposition of the graph Laplacian,
- $F(\theta) = \sum_{k=0}^{K-1} \theta_k \Lambda^k$ is a polynomial filter,
- p is the vertex-feature index.

For an undirected graph $\mathcal{G}$ this operation is equivalent (up to a similarity transform) to diffusion convolution on the graph [5].

The core of the model is defined by

$$\begin{aligned}
r^{(t)} &= \sigma\big(\Theta_r \star_g [X^{(t)}, H^{(t-1)}] + b_r\big),\\
u^{(t)} &= \sigma\big(\Theta_u \star_g [X^{(t)}, H^{(t-1)}] + b_u\big),\\
C^{(t)} &= \tanh\big(\Theta_C \star_g [X^{(t)}, r^{(t)} \odot H^{(t-1)}] + b_c\big),\\
H^{(t)} &= u^{(t)} \odot H^{(t-1)} + \big(1 - u^{(t)}\big) \odot C^{(t)},
\end{aligned}$$

where

- $X^{(t)}$, $H^{(t)}$ are the input and hidden state at time step t,
- $r^{(t)}$, $u^{(t)}$ are the reset and update gates,
- Θ_r, Θ_u, Θ_C are learnable filter parameters,
- $\star_g$ denotes the diffusion convolution operator,
- $\odot$ is element-wise multiplication,
- σ is the logistic sigmoid.

5 Feature Representation

As node features we use the *differential entropy* values computed for the following EEG rhythm bands:

- delta (1–3 Hz),
- theta (4–7 Hz),
- alpha (8–13 Hz),
- beta (14–30 Hz),
- gamma (31–50 Hz).

For a normally distributed random variable $Y \sim \mathcal{N}(\mu, \sigma^2)$, the differential entropy is

$$DE(Y) = -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}\right) dy. \tag{10}$$

Hence, at each time instant t the graph signal has shape

$$x_t \in \mathbb{R}^{62\times 5},$$

where 62 is the number of electrodes and 5 is the number of frequency bands.

6 Computational Experiment Plan

Hypothesis: accounting for the spatial and functional structure of the EEG signal and using diffusion-based methods improves the accuracy of human emotion classification.

Objectives:

1. Construct adjacency matrices between electrodes using different methods.
2. Evaluate the performance of the proposed spatio-temporal model on the resulting graphs.

The study employs the dataset described in [4], aimed at analysing affective states. Fifteen participants meeting the required medical criteria took part after providing informed consent and being briefed on the protocol.

Visual stimuli consisted of video clips from four categories. Selection criteria included

- limited duration to avoid fatigue,
- clear content without extra explanation,
- ability to evoke well-defined emotions.

Each clip lasted about two minutes and was edited to enhance emotional impact.

The experiment comprised three sessions of 24 trials each; the clip order prevented consecutive presentation of the same category. After every clip, participants filled out a questionnaire describing their experienced emotions.

EEG was recorded with 62 electrodes placed according to the standard montage, at a sampling rate of 1 kHz.

Pre-processing included:

- band-pass filtering in 0.3–50 Hz to remove noise and artifacts,
- down-sampling to 200 Hz.

7 Results

The experiment compared the performance of two recurrent neural network architectures: one with a single recurrent layer and one with two recurrent layers. The effect of input sequence length on learning quality was also studied by using sequences of 12 and 17 time steps. All model configurations were trained on an NVIDIA Tesla T4 GPU, with training time for each setup approximately 15 minutes.

The best results were achieved by the model with two recurrent layers trained on sequences of 17 elements. As the plots show, the GRU baseline performed worse than DCGRU. It is also worth noting that models with only one recurrent layer tended to overfit more quickly.

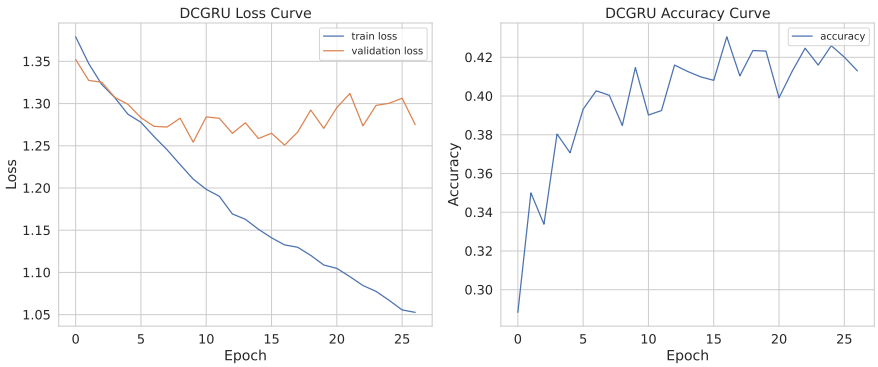

Рис. 1 Model with one recurrent layer, 12-element input sequence

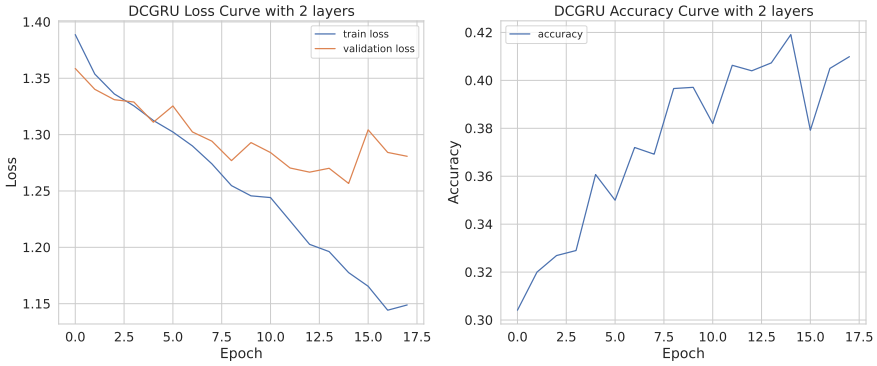

Рис. 2 Model with two recurrent layers, 12-element input sequence

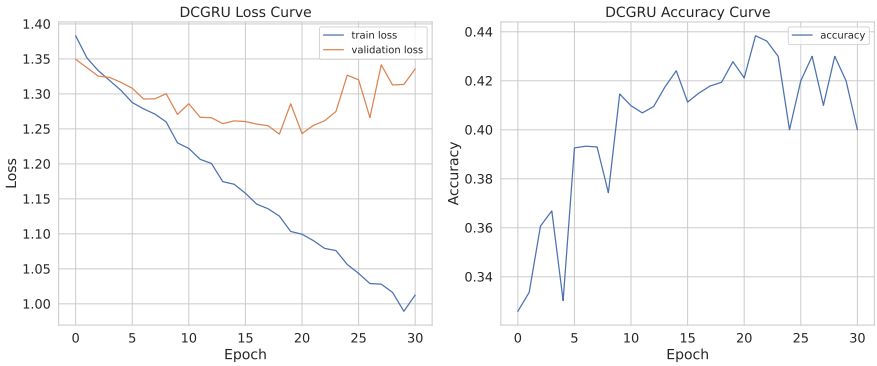

Рис. 3 Model with one recurrent layer, 17-element input sequence

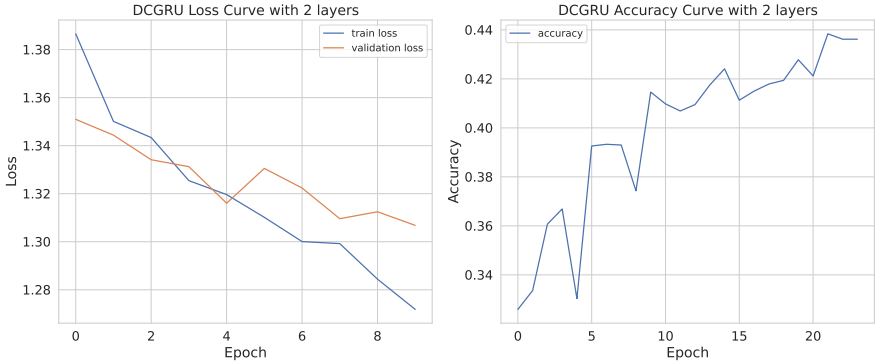

Рис. 4 Model with two recurrent layers, 17-element input sequence

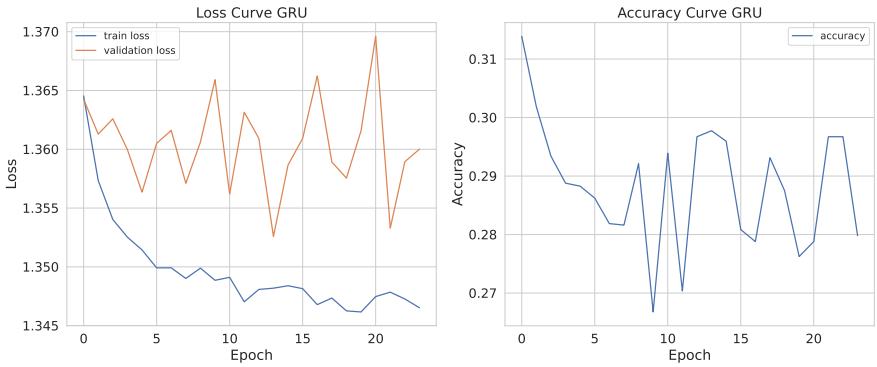

Рис. 5 GRU baseline model

8 Conclusion

This study proposes an approach to human emotion classification based on EEG data that takes into account both spatial and functional brain structure. The EEG signal is interpreted as a dynamic graph, where the vertices represent electrodes and the edges represent connections derived via phase synchronization.

Using this graph representation, we implemented a classification model based on the DCGRU graph recurrent neural network, which can capture both temporal and spatial dependencies among EEG channels.

The proposed approach achieved higher classification accuracy compared to baseline models with similar parameter counts that do not account for electrode interaction structure. This supports the hypothesis that spatio-temporal organization of brain activity plays an important role in EEG signal analysis. The results demonstrate the potential of graph neural networks and diffusion-based models in neurophysiological analysis and emotion recognition tasks, opening avenues for further research and applications in affective computing, neurorehabilitation, and psychophysiological diagnostics.

Литература

[1] Md. Asadur Rahman, Md. Foisal Hossain, Mazhar Hossain, and Rasel Ahmmed. Employing pca and t-statistical approach for feature extraction and classification of emotion from multichannel eeg signal. *Egyptian Informatics Journal*, 21(1):23–35, March 2020.

[2] Sergul Aydore, Dimitrios Pantazis, and Richard M. Leahy. A note on the phase locking value and its properties. *NeuroImage*, 74:231–244, July 2013.

[3] Arnaud Delorme and Scott Makeig. Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21, March 2004.

[4] Ruo-Nan Duan, Jia-Yi Zhu, and Bao-Liang Lu. Differential entropy feature for eeg-based emotion classification. In *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, page 81–84. IEEE, November 2013.

[5] Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting, 2017.

[6] Kristen A. Lindquist, Tor D. Wager, Hedy Kober, Eliza Bliss-Moreau, and Lisa Feldman Barrett. The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences*, 35(3):121–143, May 2012.

[7] Yishu Liu and Guifang Fu. Emotion recognition by deeply learned multi-channel textual and eeg features. *Future Generation Computer Systems*, 119:1–6, June 2021.

[8] Rab Nawaz, Kit Hwa Cheah, Humaira Nisar, and Vooi Voon Yap. Comparison of different feature extraction methods for eeg-based emotion recognition. *Biocybernetics and Biomedical Engineering*, 40(3):910–926, July 2020.

[9] Siyi Tang, Jared A. Dunnmon, Khaled Saab, Xuan Zhang, Qianying Huang, Florian Dubost, Daniel L. Rubin, and Christopher Lee-Messer. Self-supervised graph neural networks for improved electroencephalographic seizure analysis. 2021.

[10] Miroslaw Wyczesany and Tomasz S. Ligeza. Towards a constructionist approach to emotions: verification of the three-dimensional model of affect with eeg-independent component analysis. *Experimental Brain Research*, 233(3):723–733, November 2014.