

Восстановление функциональных групп головного мозга с помощью графовых диффузных моделей

А. М. Астахов, С. К. Панченко, В. В. Стрижов

astakhov.am@phystech.edu; panchenko.sk@phystech.edu; strijov@phystech.edu

В данной работе рассматривается задача классификации многомерного временного ряда, представляющего собой электроэнцефалограмму головного мозга человека. Стандартные подходы, использующие двухмерные свертки, не могут учесть пространственную структуру сигнала, поскольку датчики, считывающие показатели, находятся на сферической поверхности. В качестве решения предлагается использовать графовое представление функциональных групп, а для моделирования использовать нейронную диффузию.

Головной мозг, ЭЭГ, Графовые нейронные сети, диффузионные модели

1 Введение

Эмоции играют ключевую роль в человеческом восприятии, принятии решений и социальном взаимодействии. Их автоматическая классификация на основе нейрофизиологических данных, таких как электроэнцефалография (ЭЭГ), открывает новые возможности в психологии, медицине, аффективных вычислениях и человеко-машинном взаимодействии. Однако, несмотря на значительный прогресс в области машинного обучения и нейронаук, точная и надёжная классификация эмоций по ЭЭГ остаётся сложной задачей. Это связано с высокой индивидуальной вариабельностью сигналов, нелинейной природой эмоциональных процессов, а также ограничениями существующих методов предобработки и классификации.

В данной статье рассматриваются современные подходы к распознаванию эмоций по ЭЭГ, анализируются их преимущества и недостатки, а также предлагаются пути улучшения точности классификации. Особое внимание уделяется методам обработки сигналов, выделения информативных признаков и использованию алгоритмов глубокого обучения. Результаты исследования могут быть полезны для разработки более эффективных систем аффективного взаимодействия, нейрореабилитации и психофизиологических исследований. Объектом нашего исследования является сигнал, полученный путём электроэнцефалографического исследования человеческого мозга, воспринимаемый как многомерный временной ряд, где каждая размерность отвечает конкретному датчику на голове испытуемого. Технические ограничения ЭЭГ-исследований включают низкое пространственное разрешение метода и высокую чувствительность к артефактам. Как показано в работе [3], движения глаз и мышечная активность могут существенно искажать сигнал. Более того, индивидуальные различия в паттернах мозговой активности между испытуемыми значительно снижают эффективность универсальных классификаторов.

Существующие подходы преимущественно основаны на двух теоретических моделях эмоций: локационистской (базовые эмоции) и многомерной (валентность-возбуждение-доминантность, VAD) [10]. Однако большинство современных методов не учитывают пространственные взаимосвязи между электродами, что ограничивает их эффективность.

В исследованиях используются различные методы извлечения признаков:

Временные характеристики: В работе [7] применены шесть статистических параметров ЭЭГ с последующим отбором каналов методами PCA и ReliefF, что позволило достичь точности 81,87% на датасете DEAP. Однако авторы не рассматривали пространственные корреляции между электродами.

Частотные характеристики: Исследование [1] демонстрирует эффективность РСА для сокращения размерности признаков с последующей классификацией методом SVM (точность 85,85% на SEED). Аналогично, в [8] сравнение различных признаков показало, что статистические характеристики в сочетании с KNN дают точность 77,54–79%.

Важное ограничение этих работ заключается в том, что анализ проводился для каждого электрода отдельно, без учёта пространственных взаимодействий между различными областями мозга. Это особенно существенно, поскольку эмоциональные состояния, как известно, связаны с скоординированной активностью распределённых нейронных сетей [6].

Задача нашего исследования – использовать пространственные связи между датчиками для улучшения качества классификации. Мы предлагаем рассматривать временной ряд как динамический граф, где рёбра представляют взаимосвязи между датчиками в пространстве или статистически. Мы считаем, что учёт этих факторов позволит построить более качественную и устойчивую модель классификации. Мы рассмотрим подходы построения подобных связей и посмотрим, как они влияют на результат классификации. Проводить оценку качества модели мы будем на открытом датасете SEED IV. В качестве модели мы предлагаем использовать DCGRU, которая хорошо себя показала в схожей проблеме классификации эпилептических припадков по данным электроэнцефалограммы [9].

2 Постановка задачи

2.1 Построение матрицы связности

Исходный сигнал ЭЭГ задан в виде тензора $\mathbf{X} = [\mathbf{X}_m]_{m=1}^M$, $\mathbf{X}_m \in \mathbb{R}^{E \times N}$, где N соответствует числу отсчётов времени при измерении сигнала, E — число электродов, считываемый сигнал, M — число испытаний. Также дана матрица координат электродов $\mathbf{Z} \in \mathbb{R}^{E \times 3}$, определяемая выбранным при снятии электроэнцефалограммы стандартом размещения. В данной работе предлагается рассматривать сигнал в качестве неориентированного динамического графа:

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

для решения проблемы моделирования пространственной взаимосвязи между электродами на голове испытуемого. В качестве множества вершин $\mathcal{V}(m, t)$ мы рассматриваем электроды, а значениями в вершинах будут значения сигнала в момент t . Множество рёбер $\mathcal{E}(m, t)$ задаётся матрицей связности графа $\mathbf{A}_{\mathbf{X}, \mathbf{Z}}(m, t)$.

2.2 Основные определения

Дана выборка $\mathcal{D} = (\mathbf{X}, \mathbf{Z}, \mathbf{y})$ активности головного мозга, где:

- $\mathbf{X} = [\mathbf{X}_m]_{m=1}^M$ — набор сигналов;
- $\mathbf{X}_m = [\mathbf{x}_t]_{t \in T}$ — сигнал, полученный в m -м испытании;
- $\mathbf{x}_t \in \mathbb{R}^E$ — наблюдения сигнала в момент времени t ;
- $\mathbf{Z} = [\mathbf{z}_k]_{k=1}^E$, $\mathbf{z}_k \in \mathbb{R}^3$ — координаты электродов;
- $\mathbf{y} = [\mathbf{y}_m]_{m=1}^M$ — целевая переменная;
- $\mathbf{y}_m \in \{1, \dots, C\}$ — метка класса;
- $T = \{t_n\}_{n=1}^N$ — набор временных отсчётов;
- $E = 62$ — число электродов;
- N — число наблюдений в одном отрезке сигнала.

Для решения задачи декодирования рассматривается модель из класса графовых рекуррентных диффузионных нейронных сетей:

$$h_\theta : (\mathbf{X}, \Delta_{\mathbf{X}, \mathbf{Z}}^*) \rightarrow \mathbf{y}. \quad (1)$$

В качестве функции ошибки выбрана кросс-энтропия:

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

где $p_m^c = h_\theta(\mathbf{X}_m, \Delta_{\mathbf{X}, \mathbf{Z}}^*(m))$ — вероятность класса c для \mathbf{X}_m с матрицей $\Delta_{\mathbf{X}, \mathbf{Z}}^*(m)$.

Задача поиска оптимальных параметров имеет следующий вид:

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

3 Построение матрицы связности

В этом разделе описываются методы построения матрицы смежности через оценку взаимосвязи между временными рядами, соответствующими электродам. Мы рассматриваем построение матрицы смежности на основе синхронизации фаз сигналов.

3.1 Синхронизация фаз сигналов

Фазовая синхронизация представляет собой подход к анализу возможных нелинейных взаимозависимостей и фокусируется на фазах сигналов. Предполагается, что две динамические системы могут иметь синхронизацию фаз, даже если их амплитуды независимы. Обозначим $x(t)$, $y(t)$ — динамические системы, соответствующие наблюдениям сигнала \mathbf{x}_{mi} и \mathbf{x}_{mj} в отрезке времени $[t_n - T_w, t_n]$ в m -м испытании. Синхронизация фаз понимается как:

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

Для оценки фазы вычисляется аналитическое представление сигнала с использованием преобразования Гильберта:

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

где

$$\dot{x}(t) = \frac{1}{\pi} \text{v.p.} \int_{-\infty}^{\infty} \frac{x(t')}{t - t'} dt' \quad \text{— преобразование Гильберта сигнала } x(t), \quad (6)$$

где v.p. — главное значение интеграла по Коши.

Фаза аналитического сигнала определяется как:

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

Таким образом, для двух сигналов $x(t)$, $y(t)$ равной продолжительности T_w с фазами $\varphi_x(t)$, $\varphi_y(t)$ значение синхронизации фаз (phase locking value) [2] задаётся уравнением:

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

где Δt — шаг по времени, $i = \sqrt{-1}$.

Матрица связности определяется как:

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

4 Модель классификации

Для решения задачи классификации мы предлагаем использовать модель **DCGRU** [5], хорошо показавшую себя в задаче классификации эпилептических припадков по ЭЭГ в работе [9]. Мы считаем, что использование диффузии позволит учитывать удалённые вершины графа, что улучшит качество классификации, а также сделает модель более устойчивой к шуму, что важно, поскольку ЭЭГ-данные крайне индивидуальны.

Для моделирования диффузии на графе используется спектральная свёртка, определяемая как:

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

где:

- $L = \Phi \Lambda \Phi^\top$ — спектральное разложение лапласиана графа;
- $F(\theta) = \sum_{k=0}^{K-1} \theta_k \Lambda^k$ — полиномиальный фильтр;
- p — индекс признака вершины.

Эта операция эквивалентна диффузионной свёртке на графе с точностью до преобразования подобия, если граф \mathcal{G} является неориентированным [5].

Ядро модели описывается следующими уравнениями:

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

где:

- $X^{(t)}, H^{(t)}$ — вход и выход на временном шаге t ;
- $r^{(t)}, u^{(t)}$ — вентиль сброса и вентиль обновления;
- $\Theta_r, \Theta_u, \Theta_C$ — обучаемые параметры фильтров;
- \star_g — оператор диффузионной свёртки;
- \odot — поэлементное умножение;
- σ — сигмоидальная функция активации.

5 Признаковое описание

В качестве признаков используются значения дифференциальной энтропии для следующих диапазонов ритмов головного мозга:

- дельта (1–3 Гц);
- тета (4–7 Гц);
- альфа (8–13 Гц);
- бета (14–30 Гц);

– гамма (31–50 Гц).

Формула дифференциальной энтропии для нормально распределённой случайной величины $Y \sim \mathcal{N}(\mu, \sigma^2)$ имеет вид:

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

Таким образом, в каждый момент времени t графовый сигнал имеет размерность:

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

где:

- 62 — количество электродов;
- 5 — количество частотных диапазонов.

6 План вычислительного эксперимента

Гипотеза: учёт пространственной и функциональной структуры ЭЭГ-сигнала, а также использование диффузионных методов повышает качество классификации эмоциональных состояний человека.

Цели эксперимента:

1. Построить матрицы связей между электродами с использованием различных методов.
2. Оценить качество работы предложенной пространственно-временной модели на основе построенных графов.

В исследовании использовался набор данных [4], предназначенный для анализа аффективных состояний человека. В эксперименте участвовали 15 испытуемых, соответствующих необходимым медицинским требованиям. Все участники подписали информированное согласие и были ознакомлены с протоколом исследования.

В качестве визуальных стимулов применялись видеоролики из шести категорий. Критерии отбора видеофрагментов включали:

- ограничение по длительности, чтобы избежать утомления;
- ясность содержания без дополнительных пояснений;
- способность вызывать чётко определённые эмоции.

Каждое видео длилось около 4 минут и было специально отредактировано для усиления эмоционального воздействия.

Эксперимент проводился в 3 сессии по 15 попыток. Порядок видео исключал последовательную демонстрацию фрагментов одной категории. После каждого видео участники заполняли опросник с описанием пережитых эмоций.

ЭЭГ записывалась с использованием 62 электродов, расположенных по системе 10–20. Частота дискретизации составляла 1 кГц.

На этапе предобработки применялись:

- полосовая фильтрация в диапазоне 0.3–50 Гц для удаления шумов и артефактов;
- понижение частоты дискретизации до 200 Гц.

7 Результаты

В ходе эксперимента было проведено сравнение производительности двух архитектур рекуррентных нейронных сетей: модели с одним рекуррентным слоем и модели с двумя

рекуррентными слоями. Исследование также включало анализ влияния длины входной последовательности на качество обучения — для этого использовались данные с последовательностями длиной 12 и 17 элементов. Обучение всех вариантов моделей проводилось на видеокарте NVIDIA Tesla T4, при этом время обучения для каждой конфигурации составило примерно 15 минут.

Наилучшие результаты продемонстрировала модель с двумя рекуррентными слоями, обученная на последовательностях из 17 элементов. Как видно из графиков качество GRU модели хуже качества DCGRU. Также стоит отметить, что модели с одним рекуррентным слоем быстрее начинают переобучаться.

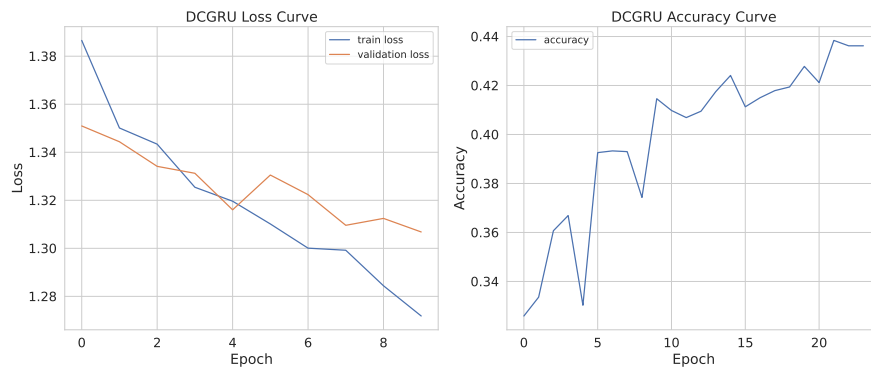


Рис. 1 модель с одним рекуррентным слоем и группой в 12 элементов

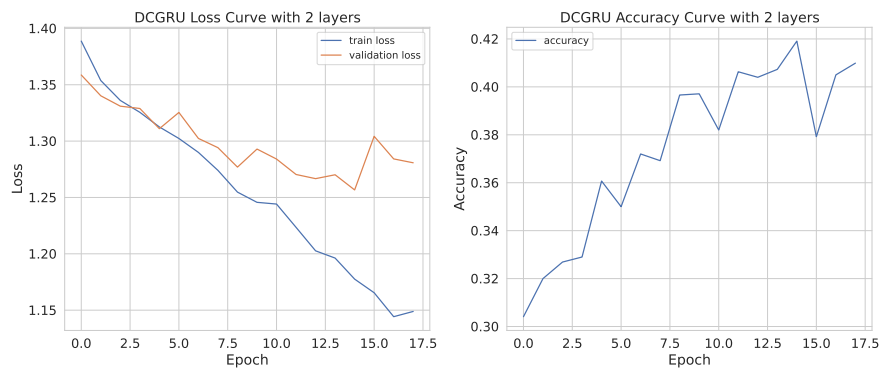


Рис. 2 модель с двумя рекуррентным слоем и группой в 12 элементов

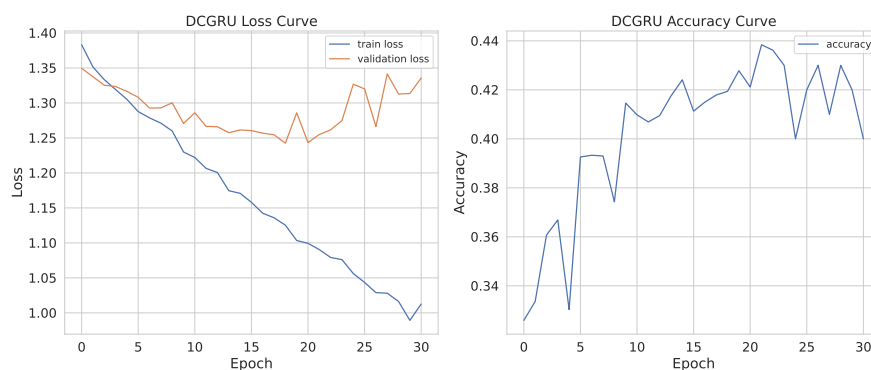
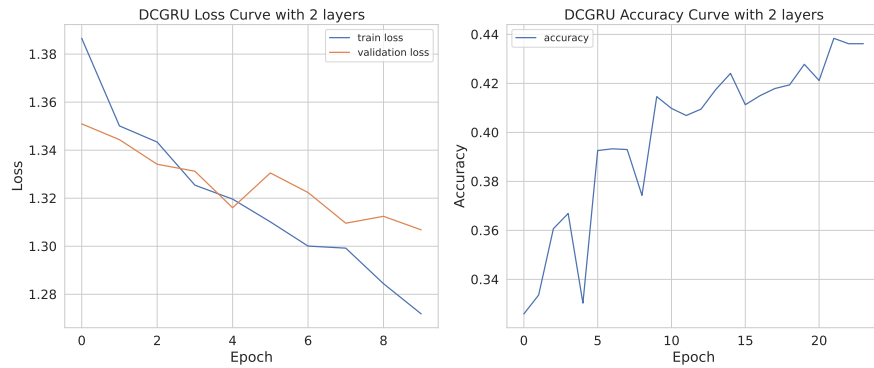
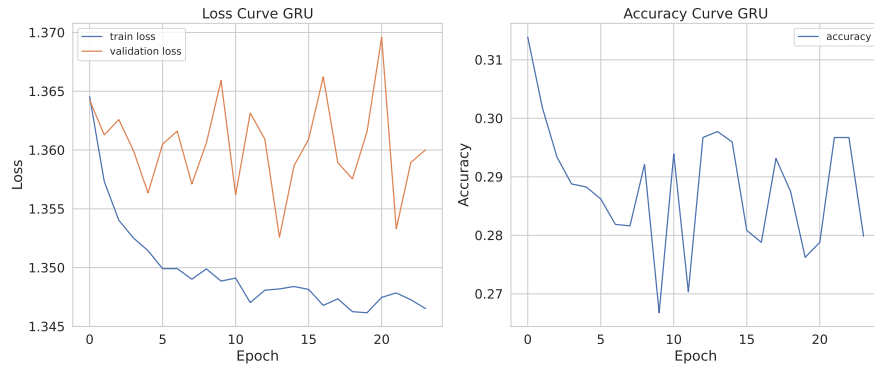


Рис. 3 модель с одним рекуррентным слоем и группой в 17 элементов**Рис. 4** модель с двумя рекуррентным слоем и группой в 17 элементов**Рис. 5** GRU

8 Заключение

В данной работе предложен подход к классификации эмоциональных состояний человека по данным электроэнцефалографии, учитывающий пространственную и функциональную структуру мозга. Для этого сигнал ЭЭГ интерпретировался как динамический граф, вершинами которого являются электроды, а рёбрами — связи, определяемые на основе фазовой синхронизации сигналов. На основе построенного графа была реализована модель классификации с использованием графовой рекуррентной нейронной сети DCGRU, способной моделировать как временные, так и пространственные зависимости между каналами ЭЭГ.

Предложенный подход позволил достичь более высокой точности классификации по сравнению с базовыми моделями, содержащими примерно такое же количество параметров, но не учитывающими структуру взаимодействий между электродами. Это подтверждает гипотезу о важности учёта пространственно-временной организации мозговой активности при анализе ЭЭГ-сигналов. Результаты эксперимента демонстрируют потенциал графовых нейронных сетей и диффузионных моделей в задачах нейрофизиологического анализа и распознавания эмоций, открывая возможности для дальнейших исследований и применения в системах аффективного взаимодействия, нейрореабилитации и психофизиологической диагностики.

Литература

- [1] Md. Asadur Rahman, Md. Faisal 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] Mirosław 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.