
ВОССТАНОВЛЕНИЕ СНИМКОВ фМРТ ПО ПРОСМАТРИВАЕМОМУ ВИДЕОРЕЯДУ

Киселев Никита
kiselev.ns@phystech.edu

Грабовой Андрей
grabovoy.av@phystech.edu

2 апреля 2023 г.

АННОТАЦИЯ

Исследуется проблема восстановления зависимости между показаниями датчиков фМРТ и восприятием внешнего мира человеком. Проводится анализ зависимости между последовательностью снимков фМРТ и видеорядом, просматриваемым человеком. На основе исследования зависимости предлагается метод аппроксимации показаний фМРТ по просматриваемому видеоряду. Для анализа предложенного метода проводится вычислительный эксперимент на выборке, полученной при томографическом обследовании большого числа испытуемых.

Ключевые слова: нейровизуализация · фМРТ · видеоряд · зависимость между данными

1 Введение

Совокупность методов, визуализирующих структуру и функции человеческого мозга, называется *нейровизуализацией*. Методы нейровизуализации, такие как ЭКГ, КТ, МРТ и фМРТ, используются для изучения мозга, а также для обнаружения заболеваний и психических расстройств.

Функциональная магнитно-резонансная томография или *фМРТ* (англ. *fMRI*) является разновидностью магнитно-резонансной томографии и основана на изменениях в токе крови, вызванных нейронной активностью мозга [4]. Эти изменения происходят не моментально, а с некоторой задержкой. Она возникает из-за того, что сосудистая система достаточно долго реагирует на потребность мозга в глюкозе [5]. Изображения, получаемые с помощью фМРТ, показывают, какие участки мозга активированы при выполнении испытуемым определенных заданий.

Настоящая работа посвящена восстановлению зависимости между снимками фМРТ и видеорядом. Используется предположение, что такая зависимость существует. Кроме

того, предполагается, что между снимком и видеорядом есть постоянная задержка во времени [5]. Проверяется зависимость снимка фМРТ от одного изображения. Время задержки выступает в качестве гиперпараметра модели. На основе анализа зависимости предлагается метод аппроксимации показаний фМРТ по просматриваемому видеоряду.

Метод фМРТ играет большую роль в нейровизуализации, однако имеет ряд важных ограничений. В работах [6, 7] рассматриваются временное и пространственное разрешения фМРТ. Временное разрешение является существенным недостатком данного метода. Другой недостаток фМРТ — неизбежно возникающие шумы, связанные с движением объекта в сканере, сердцебиением и дыханием человека, тепловыми флуктуациями самого прибора и т. д. В работе [8] предлагаются методы подавления вышеперечисленных шумов на основе графов и демонстрируется их эффективность в задаче выявления эпилепсии и депрессии.

Обобщением уже естественных для обработки изображений 2D сверток в CNN являются 3D свертки [9]. Они агрегируют информацию как по времени, так и по пространству. Однако это приводит к сильному увеличению количества используемых параметров. В настоящей работе используется наиболее современная архитектура — Transformer. Впервые она была предложена в статье [10]. Не так давно появилась адаптация архитектуры Transformer для работы с видео [11]. Данная архитектура состоит из кодировщика и декодировщика, каждый из которых в свою очередь состоит из отдельных слоев. Использование механизма Attention [10] позволяет значительно повысить качество работы модели.

Данные, на которых проводятся проверка гипотезы зависимости и демонстрация работы построенного метода, представлены в работе [2]. Этот набор данных был получен при обследовании группы из 63 испытуемых. Тридцать из них проходили обследование фМРТ. Им предлагалось выполнить одно и то же задание — просмотреть короткий аудиовизуальный фильм. Для него в рассматриваемой работе были сгенерированы аннотации, содержащие в том числе информацию о времени появления и исчезновения отдельных слов, объектов и персонажей. Методы аудио- и видеоаннотирования подробно излагаются в [3] и [1].

2 Постановка задачи

Задана частота кадров $\nu \in \mathbb{R}$ и продолжительность $t \in \mathbb{R}$ видеоряда. Задан видеоряд

$$\mathbf{P} = [\mathbf{p}_1, \dots, \mathbf{p}_{\nu t}], \quad \mathbf{p}_i \in \mathbb{R}^{W_{\mathbf{P}} \times H_{\mathbf{P}} \times C_{\mathbf{P}}}, \quad (1)$$

с шириной, высотой и числом каналов изображения $W_{\mathbf{P}}$, $H_{\mathbf{P}}$ и $C_{\mathbf{P}}$ соответственно.

Обозначим частоту снимков фМРТ $\mu \in \mathbb{R}$. Задана последовательность снимков

$$\mathbf{S} = [\mathbf{s}_1, \dots, \mathbf{s}_{\mu t}], \quad \mathbf{s}_i \in \mathbb{R}^{W_{\mathbf{S}} \times H_{\mathbf{S}} \times D_{\mathbf{S}}}, \quad (2)$$

где $W_{\mathbf{S}}$, $H_{\mathbf{S}}$ и $D_{\mathbf{S}}$ — размерности воксельного изображения.

Задача состоит в построении отображения, которое бы учитывало задержку Δt между снимком фМРТ и видеорядом, а также предыдущие томографические показания.

Формально, необходимо найти такое отображение g , что

$$g(\mathbf{p}_1, \dots, \mathbf{p}_{k_i - \nu \Delta t}; \mathbf{s}_1, \dots, \mathbf{s}_{i-1}) = \mathbf{s}_i, \quad i = 1, \dots, \mu t, \quad (3)$$

где для i -го снимка фМРТ номер соответствующего изображения k_i определяется по формуле

$$k_i = \frac{i \cdot \nu}{\mu}. \quad (4)$$

3 Вычислительный эксперимент

3.1 Цель эксперимента

Целью вычислительного эксперимента является выбор оптимального метода аппроксимации отображения (3). Данные для эксперимента взяты из работы [2]. Набор данных содержит результаты обследования 63 испытуемых. Для тридцати из них известны показания фМРТ.

Последовательность снимков для испытуемого представляется в виде тензора. Каждый снимок представляется тензором третьего ранга с размерностями

$$W_S = 40, H_S = 64, D_S = 64.$$

Частота снимков $\mu = 1,64$ кадров в секунду, продолжительность $t = 390$ с, число снимков $N_S = \mu \cdot t = 641$. Таким образом,

$$\mathbf{S} = [\mathbf{s}_1, \dots, \mathbf{s}_{641}] \in \mathbb{R}^{40 \times 64 \times 64 \times 641}.$$

Продолжительность видео также $t = 390$ с, частота кадров $\nu = 25$, число изображений $N_P = \nu \cdot t = 9750$. Каждое изображение представляется в виде тензора третьего ранга с размерностями

$$W_P = 640, H_P = 480, C_P = 3.$$

Тензор видеоряда

$$\mathbf{P} = [\mathbf{p}_1, \dots, \mathbf{p}_{9750}] \in \mathbb{R}^{640 \times 480 \times 3 \times 9750}.$$

Согласно (4), номер изображения, соответствующего снимку фМРТ, $k_i = 15 \cdot i$. Тогда суммарное число пар (изображение, снимок) равно $N = N_S - \mu \Delta t = 641 - \mu \Delta t$.

3.2 Базовый метод

Предположим, что каждый снимок фМРТ зависит только от одного изображения. Тогда соответствующее отображение

$$g(\mathbf{p}_{k_i - \nu \Delta t}) = \mathbf{s}_i, \quad i = 1, \dots, \mu t. \quad (5)$$

Рассмотрим снимок $\mathbf{s}_\ell = [v_{ijk}^\ell] \in \mathbb{R}^{40 \times 64 \times 64}$, где $v_{ijk}^\ell \in \mathbb{R}_+$ — значение соответствующего вокселя. Для каждого изображения из видеоряда был получен вектор признакового описания размерности $d = 2048$:

$$\mathbf{x}_i = [x_1^i, \dots, x_d^i]^\top \in \mathbb{R}^d, \quad i = 1, \dots, 9750.$$

Использовалась архитектура нейронной сети ResNet152 без последнего линейного слоя.

Таким образом, для каждого вокселя задана выборка

$$\mathfrak{D}_{ijk} = \{(\mathbf{x}_\ell, v_{ijk}^\ell), \ell = 1, \dots, N\}.$$

Поставлена задача восстановления регрессии

$$y_{ijk} : \mathbb{R}^d \rightarrow \mathbb{R}. \quad (6)$$

Используется линейная модель с вектором параметров

$$\begin{aligned} \mathbf{w}_{ijk} &= [w_1^{ijk}, \dots, w_d^{ijk}]^\top \in \mathbb{R}^d : \\ f_{ijk}(\mathbf{x}, \mathbf{w}_{ijk}) &= \sum_{q=1}^d w_q^{ijk} x_q. \end{aligned} \quad (7)$$

Для модели f_{ijk} с соответствующим ей вектором параметров $\mathbf{w}_{ijk} \in \mathbb{R}^d$ определим квадратичную функцию потерь:

$$\mathcal{L}_{ijk}(\mathbf{w}_{ijk}, \Delta t) = \sum_{\ell=1}^{N_S - \mu \Delta t} (f_{ijk}(\mathbf{x}_\ell, \mathbf{w}_{ijk}) - v_{ijk}^\ell)^2. \quad (8)$$

Требуется найти параметры, доставляющие минимум функционалу потерь $\mathcal{L}_{ijk}(\mathbf{w}_{ijk}, \Delta t)$ при заданном гиперпараметре Δt :

$$\hat{\mathbf{w}}_{ijk} = \arg \min_{\mathbf{w}_{ijk}} \mathcal{L}_{ijk}(\mathbf{w}_{ijk}, \Delta t). \quad (9)$$

Минимум функции потерь находится методом наименьших квадратов. Определим матрицу плана

$$\mathbf{X} = [\mathbf{x}_1^\top, \dots, \mathbf{x}_N^\top]^\top = [x_j^i] \in \mathbb{R}^{N \times d} \quad (10)$$

и вектор, компонентами которого являются значения одного и того же вокселя в разных снимках,

$$\mathbf{v}_{ijk} = [v_{ijk}^1, \dots, v_{ijk}^N]^\top \in \mathbb{R}^N. \quad (11)$$

Решение можно записать в виде

$$\mathbf{w}_{ijk} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{v}_{ijk} = \mathbf{X}^+ \mathbf{v}_{ijk}, \quad (12)$$

где \mathbf{X}^+ — псевдообратная матрица для \mathbf{X} .

Получим формулу для восстановленных снимков фМРТ. Введем матрицу весов

$$\mathbf{W} = [\mathbf{w}_1^\top, \dots, \mathbf{w}_{W_S \cdot H_S \cdot D_S}^\top]^\top = [w_j^i] \in \mathbb{R}^{W_S \cdot H_S \cdot D_S \times d}. \quad (13)$$

Тогда строки матрицы

$$\mathbf{X} \cdot \mathbf{W}^\top \quad (14)$$

содержат значения вокселей соответствующего снимка. После преобразования в тензор третьего ранга получаем восстановленный снимок \mathbf{s}_ℓ .

Размерность снимка фМРТ предварительно понижается с помощью сверточного слоя MaxPool3D. Рассматриваются коэффициенты сжатия 1, 2, 4 и 8. Значения вокселей нормализуются процедурой MinMaxScale. Производится разделение выборки на обучающую и тестовую. Качество рассматриваемого метода оценивается с помощью метрики MSE. Усреднение берется по всем снимкам фМРТ в тестовой выборке и по всем вокселям снимка.

3.2.1 Предварительные результаты

Получены восстановленные снимки фМРТ без предварительного сжатия. На Рис. 1 представлены срезы истинного и восстановленного снимков из тестовой выборки. На Рис. 1.(с) можно наблюдать разность между ними.

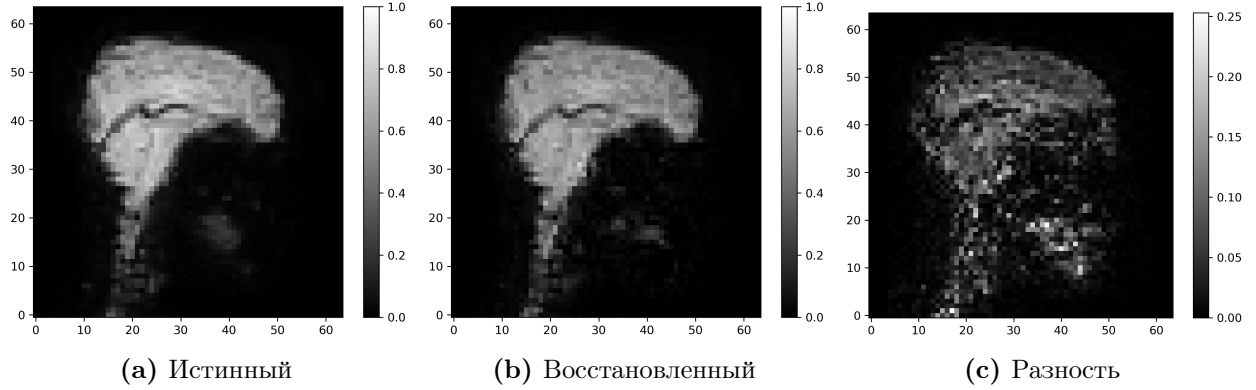


Рис. 1: Срез снимка фМРТ из тестовой выборки

Проанализирована зависимость качества восстановления снимков фМРТ от гиперпараметра Δt . Использовалось предварительное 8-кратное сжатие снимка. Зависимость метрики MSE от гиперпараметра Δt представлена на Рис. 2

В рассматриваемом методе не учитывается зависимость значений соседних вокселей. Это является его главным недостатком, что приводит к большим выбросам в восстановленных снимках и малой контрастности. Несмотря на это, на восстановленных снимках отчетливо видны границы активных областей.

3.3 Основной метод

Предположим, что каждый снимок фМРТ зависит только от одного изображения и предыдущего снимка. Тогда соответствующее отображение

$$g(\mathbf{p}_{k_i - \nu \Delta t}; \mathbf{s}_{i-1}) = \mathbf{s}_i, \quad i = 1, \dots, \mu t. \quad (15)$$

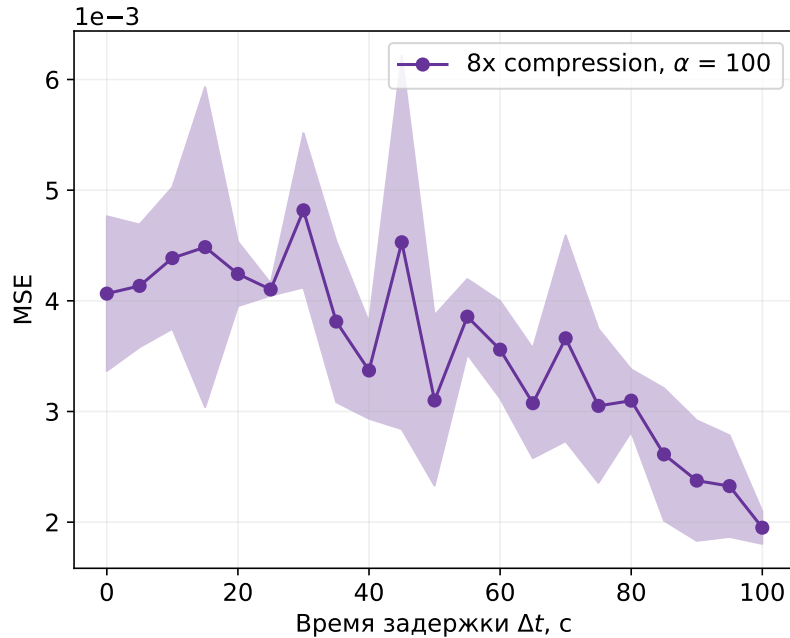


Рис. 2: Зависимость метрики MSE от гиперпараметра Δt на снимке из тестовой выборки

4 Анализ ошибки

5 Заключение

Список литературы

- [1] Julia Berezutskaya, Zachary V. Freudenburg, Luca Ambrogioni, Umut Güçlü, Marcel A. J. van Gerven, and Nick F. Ramsey. Cortical network responses map onto data-driven features that capture visual semantics of movie fragments. *Scientific Reports*, 10(1), July 2020.
- [2] Julia Berezutskaya, Mariska J. Vansteensel, Erik J. Aarnoutse, Zachary V. Freudenburg, Giovanni Piantoni, Mariana P. Branco, and Nick F. Ramsey. Open multimodal iEEG-fMRI dataset from naturalistic stimulation with a short audiovisual film. *Scientific Data*, 9(1), March 2022.
- [3] Paul Boersma and David Weenink. Praat: doing phonetics by computer [computer program]. version 6.0. 37. Retrieved February, 3:2018, 2018.
- [4] Gary H. Glover. Overview of functional magnetic resonance imaging. *Neurosurgery Clinics of North America*, 22(2):133–139, April 2011.
- [5] Nikos K. Logothetis. The underpinnings of the BOLD functional magnetic resonance imaging signal. *The Journal of Neuroscience*, 23(10):3963–3971, May 2003.
- [6] Nikos K Logothetis. What we can do and what we cannot do with fmri. *Nature*, 453(7197):869–878, 2008.
- [7] Ravi S Menon and Seong-Gi Kim. Spatial and temporal limits in cognitive neuroimaging with fmri. *Trends in cognitive sciences*, 3(6):207–216, 1999.

- [8] Maxim Sharaev, Alexander Andreev, Alexey Artemov, Alexander Bernstein, Evgeny Burnaev, Ekaterina Kondratyeva, Svetlana Sushchinskaya, and Renat Akzhigitov. fmri: preprocessing, classification and pattern recognition, 2018.
- [9] Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, and Manohar Paluri. Learning spatiotemporal features with 3d convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December 2015.
- [10] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017.
- [11] Shen Yan, Xuehan Xiong, Anurag Arnab, Zhichao Lu, Mi Zhang, Chen Sun, and Cordelia Schmid. Multiview transformers for video recognition, 2022.