
FITS: fMRI temporal and sparse relations

A Preprint

Galina Boeva
MIPT
Antiplagiat
Skoltech

Nikita Kornilov
MIPT
Skoltech

Marat Khusainov
MIPT

Bair Mihailov
MIPT
Skoltech

12 декабря 2023 г.

Abstract

В данном исследовании предлагается рассмотреть пространственно-временную модель визуализации временного ряда данных fMRI (Функциональная магнитно-резонансная томография) для предсказания на основе целевого фрагмента fMRI видео. В качестве решения мы предлагаем новую технику глубокого обучения основанную на механизме внимания, примененного как и к пространственной составляющей данных, так и к временной компоненте. Итоговая модель обладает высокой скоростью работы и высокими метриками качества предсказаний.

Keywords First keyword · Second keyword · More

1 Введение

В данном исследовании предлагается рассмотреть пространственно-временную модель визуализации временного ряда данных fMRI (Функциональная магнитно-резонансная томография) для предсказания на основе целевого фрагмента fMRI видео. При просмотре видеороликов в интернете, фильмов, TikTok, наш мозг реагирует, что приводит в возбуждение разные отделы мозга. Например, средний мозг реагирует на зрительные раздражения, в свою очередь задний отдел реагирует на мимику человека. fMRI позволяет определить активность мозга в тот или иной момент времени, так как при влиянии раздражителей мы можем выявить зависимость реакции конкретной части головного мозга.

В данной области основной проблемой является решение трудностей с огромной размерностью данных. В ряде статей Rezaei et al. [2023, 2020] предлагается применение модели декодера для многомерных наблюдений, используя теорему Байеса и условие марковости для входных данных. Также использовался байесовский подход и в статье Yousefi et al. [2019], авторы изучают когнитивные процессы, которые в свою очередь тоже являются многомерным временным рядом. Предлагается использовать структуру моделирования, в которой когнитивный процесс определяется как низкоразмерная динамическая латентная переменная, называемая когнитивным состоянием. Для моделирования всех этих взаимосвязей применяется архитектура энкодер-декодер.

Сопутствующей к большой размерности данных проблемой является высокое время обработки видео и fMRI. В работе Rezaei et al. [2021] авторы предложили вместо схемы энкодер-декодер использовать смесь гауссиан для моделирования отклика каждой клетки fMRI, тем самым сократив время вычислений до приемлемого на практике. Для обучения использовался принцип максимизации вариационной нижней оценки ELBO. Однако в статье рассматривались только низкоразмерные данные без пространственной составляющей, что является плохим приближением fMRI.

Анализ данных fMRI проводился многими авторами. Интересным подходом является применение методов, которые могут учитывать как пространственную, так и временную компоненту в данных. Временная компонента это распределение каждого отдельного признака, отвечающего за конкретный

отдел мозга, по временной шкале. Пространственная в свою очередь отвечает за связь различных признаков определенных отделов мозга друг с другом, которые могут описывать разные моменты времени. Например, в статье Azevedo et al. [2022] авторами рассматривается метод, основанный на двух блоках. Первый - это сверточная временная сетка для работы с временной компонентой данных. Вторая - это граф для связи признаков, которые распределены в пространстве, то есть учет пространственной компоненты. В работе Malkiel et al. [2022a] используется трансформер, обученный в 2 этапа: сначала заполнение пропусков в видео fMRI, затем предсказание целевой величины. Другим примером анализа данных fMRI является работа Albuquerque et al. [2021]. Авторами предложен гибридный подход, объединяющий глубокие порождающие модели и гауссовские процессы в обобщенной аддитивной модели.

Так как использования гауссовских процессов в моделях вариационных автокодировщиков может играть одну из ключевых ролей в нашей задаче, стоит отметить работу Dai et al. [2016], которая по сути является пионерской в данной области. Авторами была предложена модель, которая состоит из нескольких слоев скрытых переменных и использует гауссовские процессы для отображения между последовательными слоями. Несмотря на то, что в статье рассматривается реализация данной модели на низкоразмерных датасетах, она послужила фундаментом для развития данной области исследования. Гауссовские процессы нашли свое применение не только в моделях вариационных автокодировщиков. Так, в работах Jankowiak et al. [2020] и Wang [2022] рассматривается их применение в задаче регрессии. В случае регрессии часто прогнозируемая дисперсия преобладает над шумом во входных данных. Кроме того, оценки неопределенности, практически не используют неопределенность функции, зависящей от входных данных. Обе работы решают эту проблему и, таким образом, дают существенно улучшенные оценки неопределенности прогнозирования.

В работе Matsubara et al. [2019] рассматриваются подходы, которые помогают работать с сигналами, не интересующие нас при предсказании. Так как данные многомерны, то и признаком порождается в много раз больше, что позволяет утверждать, что нужно проводить фильтрацию. Предложена генеративная модель на данных fMRI в состоянии покоя. Предлагаемая модель обусловлена предположением о состоянии субъекта и оценивает апостериорную вероятность состояния субъекта с учетом данных визуализации, используя теорему Байеса.

Генерация fMRI по видео является востребованной задачей. В частности, в 2021 году соревнование The Algonauts project было посвящено именно этой задаче Cichy et al. [2021]. В работе Yang et al. [2021] победителей использовался ансамбль моделей, каждая из которых фокусировалась на различных областях видео: движение, границы объектов, звук и т.п.

2 Математическая постановка

2.1 Формулировка

В этой главе мы представляем математическую формулировку нашей задачи. Пусть дан временной ряд из T элементов fMRI и видео, а именно

$$\{(\mathbf{X}_t, \mathbf{V}_t)\}_{t=1}^T,$$

где переменная $\mathbf{X}_t \in \mathbb{R}^{H \times W \times D}$ отвечает за fMRI снимок размера $H \times W \times D$ в момент времени t , а переменная $\mathbf{V}_t \in \mathbb{R}^{H_v \times W_v \times C}$ за кадр видео размера $H_v \times W_v$ с C каналами. Будем считать, что видеоряд и fMRI являются согласованными, то есть реакции \mathbf{X}_t происходит на фрагмент видео \mathbf{V}_t в тот же самый момент времени.

Задача заключается в прогнозировании fMRI ряда на N моментов времени наперёд. Для этого мы будем обучать модель M_θ с параметрами θ , которая будет по входу $\{\mathbf{X}_t\}_{t=1}^{K-1}, \{\mathbf{V}_t\}_{t=1}^K$ произвольной длины K предсказывать следующий снимок \mathbf{X}_K . Также введём функцию различия двух кадров fMRI $\mathcal{L} : \mathbb{R}^{H \times W \times D} \times \mathbb{R}^{H \times W \times D} \rightarrow \mathbb{R}$, и будем минимизировать среднюю ошибку по всем доступным моментам времени

$$\min_{\theta} \frac{1}{T} \sum_{k=1}^T \mathcal{L}(\mathbf{X}_k, M_\theta(\{\mathbf{X}_t\}_{t=1}^{k-1}, \{\mathbf{V}_t\}_{t=1}^k)). \quad (1)$$

Оптимизацию этого функционала мы будем проводить градиентными методами первого порядка по типу Adam Kingma and Ba [2014], RMSProp.

2.2 Описание данных

Как мы упоминали раньше данные fMRI, представляют из 4-мерные тензоры, в которых 3 размерности пространственные и 1 размерность - временная. На рисунках ниже представлены разные размерности одного конкретного сэмпла из обучающей выборки. Цвет пикселя соответствует интенсивности отклика мозга.

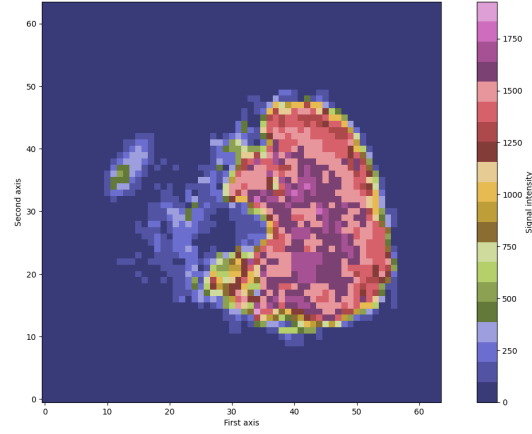


Рис. 1: Срез головного мозга

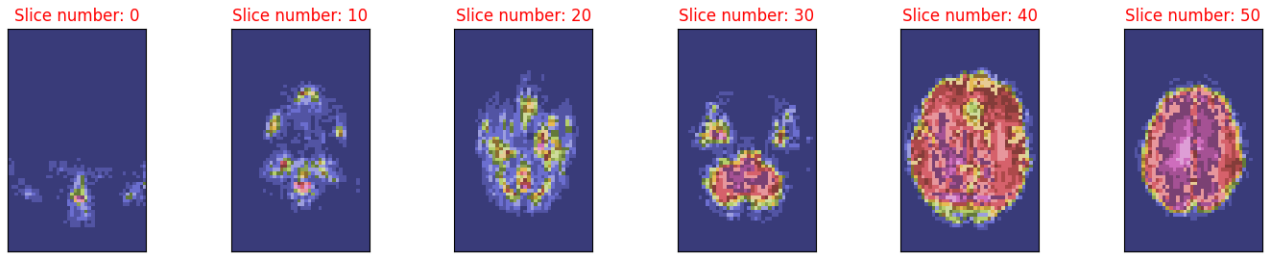


Рис. 2: Пространственная компонента

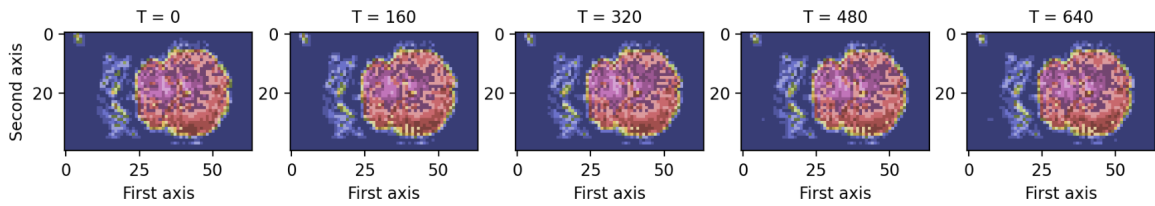


Рис. 3: Временная компонента при фиксированном горизонтальном срезе ($W = const$) головного мозга

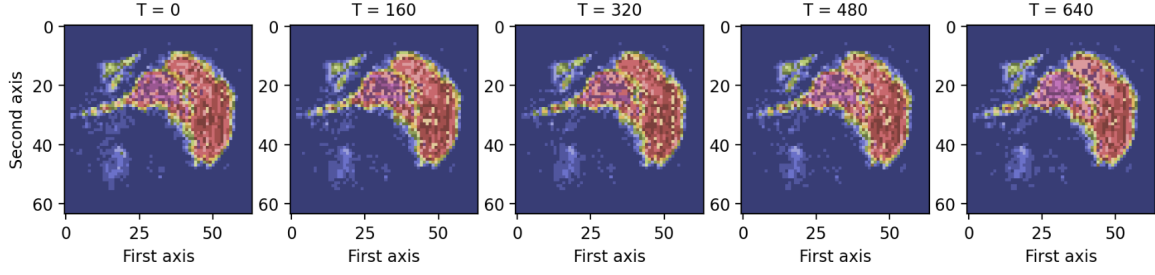


Рис. 4: Временная компонента при фиксированном вертикальном срезе ($H = const$) головного мозга

В данной работе мы предполагаем следующую гипотезу относительно наших данных: fMRI снимки, генерируемые реальным человеком, обладают свойством марковости, т.е.

$$P(\mathbf{X}_t | \mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \dots, \mathbf{X}_1) = P(\mathbf{X}_t | \mathbf{X}_{t-1}).$$

Такую гипотезу мы вводим ввиду того, что в процессе просмотра видеоролика, fMRI изменяется крайне мало под действием конкретных кадров. Таким образом, в самой модели мы можем обрабатывать не полный ряд $\{\mathbf{X}_t\}_{t=1}^T$, а лишь два последовательных кадра $\mathbf{X}_T, \mathbf{X}_{T-1}$.

2.3 Описание модели

В данной главе будет рассмотрена архитектура модели используемой в работе. Ее схема представлена на Рис. 5. Основной составляющей модели является нейросетевая архитектура TFF для работы с данными fMRI (верхний блок), которая была предложена в работе Malkiel et al. [2022a].

Модель состоит из трансформера, который работает с векторами признаков, извлеченных энкодером с использованием трехмерных сверток. На первом этапе происходит предобработка данных fMRI. А именно, происходит глобальная и воксельная нормализация кадров видео. После этого, энкодер оперирует отдельно с трехмерным кадром fMRI, отображая каждый кадр видео последовательности в вектор. Далее, модель агрегирует вектора последовательных кадров в единую последовательность, и она подается в трансформер. Затем, выход трансформера подается в декодер, который восстанавливает оригинальные данные. Для контроля качества восстановленного сигнала используются трехкомпонентная функция потерь:

$$\mathcal{L}_{rec} = \mathcal{L}_1 + \mathcal{L}_1^b + \mathcal{L}_p$$

, где \mathcal{L}_1 и \mathcal{L}_1^b являются функциями потерь средней абсолютной ошибки между значениями пикселей восстановленного изображения и оригинального изображения после глобальной и воксельной нормализаций соответственно; \mathcal{L}_p так же представляет собой функцию потерь средней абсолютной ошибки, однако на признаковом пространстве между сгенерированным кадром и оригинальным.

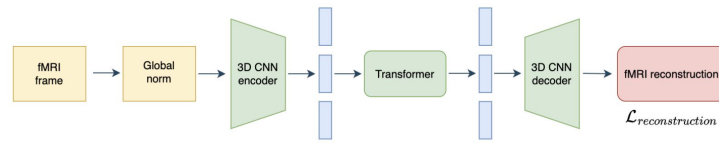


Рис. 5: Схема модели.

3 Эксперименты

График обучения модели и результаты её работы представлены на рисунках 6а и ?? соответственно. Основными метриками качества являются reconstruction loss, средняя норма отклонения от истинного ответа, и intensity. О них и поговорим подробнее ниже.

Начнем с общей метрики оценки качества нашего подхода, мы ссылаемся на статью Malkiel et al. [2022b].

В данной работе используется трехкомпонентная функция потерь, отвечающая за реконструкцию:

$$L_{reconstruction} = L_1 + L_{perceptual} + Loss_{intensity}$$

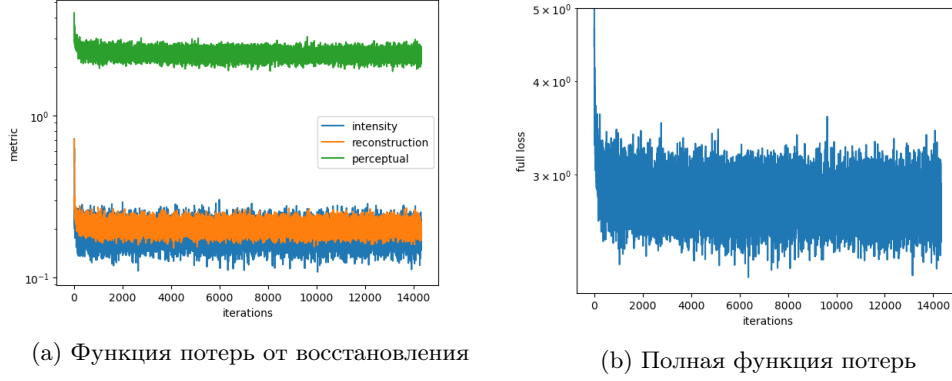


Рис. 6: Процесс обучения модели

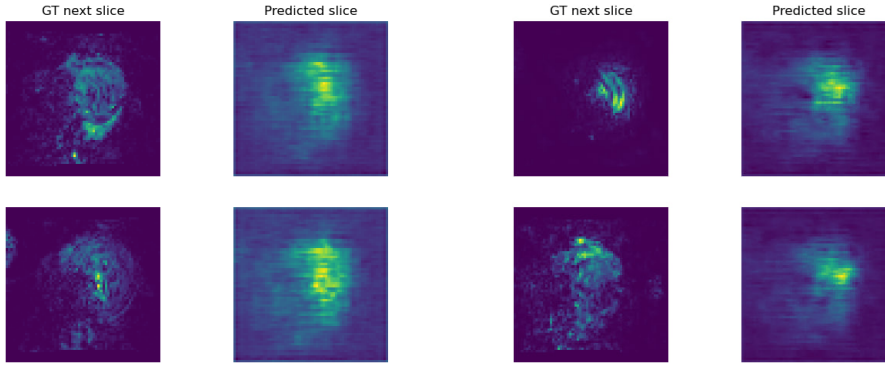


Рис. 7: Результаты работы модели

$L_{perceptual}$ отвечает за минимизацию расстояния L_2 между картами признаков первого и второго слоев, извлеченными из восстановленных данных и входных кадров. L_1 это стандартная функция потерь, применяемая между выходом декодера и глобально нормализованными кадрами $X_n = \frac{X - \mu}{\sigma}$, это μ - среднее и σ - дисперсия для исходного потока кадров X . $Loss_{intensity}$ - эта потеря основана на L_1 , применяемом к подмножеству вокселей, связанных с локальными значениями интенсивности, которые, скорее всего, представляют соответствующий сигнал. Более конкретно, учитывая полное сканирование (x_1, \dots, x_n) , мы выводим нормализацию вокселя X_n^v , мы считаем среднее и дисперсию для каждого вокселя и нормализуем его. Затем для каждого нормализованного по вокселу кадра $x_i^v \in X^v$, мы устанавливаем $x_i^v = 0$, если $|\hat{x}_i^v| < b$ и $x_i^v = \hat{x}_i^v$ в противном случае, где b - настроенное пороговое значение 80%-ный квантиль абсолютных значений, нормализованных по вокселу. Мотивация, стоящая за устранением вокселей, связанных с 80% значений, которые ближе к 0, заключается в том, что эти значения типичны для многих кадров и, следовательно, вряд ли будут представлять собой отличительный сигнал.

На рисунке 7 изображены реконструированные изображения. Что можно заметить, что достаточно много шума. Но в свою очередь область выявлена верно, но с большим разбросом значений. В будущем стоит использовать дополнительную нормализацию для изображений, и регуляризацию при обучении. На рисунке 6 изображены функции потерь модели, наблюдаем, что модель достаточно быстро сходится, то есть обучение происходит корректно и переобучение отсутствует. Конечно, результаты далеки от истинных, но работа достаточно нетривиальная и тут большая область исследований.

4 Выводы

В данной работе произведено исследование по анализу видео ряда fMRI данных. Используя подход, основанный на методе TFF, мы обучили модель и получили семплы fMRI снимков в определенные моменты времени, зависящие от истории. Область высокой интенсивности мы определяем верно, но с достаточно большой дисперсией. У нас есть несколько идей по работе с данными недостатками,

как, например, регуляризация или нормализация. Также основной идеей в будущем стала применение модели SWIN, которая будет получать вектор отдельного кадра, выявляя тем самым дополнительную взаимосвязь признаков в картинке. Что в будущем, может помочь повысить результат полученных семплов.

Список литературы

- Mohammad Rezaei, Milos Popovic, Uri Eden, Milad Lankarany, and Ali Yousefi. Deep direct discriminative decoders for high-dimensional time-series data analysis. *Neurons, Behavior, Data analysis, and Theory*, pages 1–22, 2023.
- Mohammad R Rezaei, Alex E Hadjinicolaou, Sydney S Cash, Uri T Eden, and Ali Yousefi. Bayesian decoder models with a discriminative observation process. *bioRxiv*, pages 2020–07, 2020.
- Ali Yousefi, Ishita Basu, Angelique C Paulk, Noam Peled, Emad N Eskandar, Darin D Dougherty, Sydney S Cash, Alik S Widge, and Uri T Eden. Decoding hidden cognitive states from behavior and physiology using a bayesian approach. *Neural computation*, 31(9):1751–1788, 2019.
- Mohammad Reza Rezaei, Kensuke Arai, Loren M Frank, Uri T Eden, and Ali Yousefi. Real-time point process filter for multidimensional decoding problems using mixture models. *Journal of neuroscience methods*, 348: 109006, 2021.
- Tiago Azevedo, Alexander Campbell, Rafael Romero-Garcia, Luca Passamonti, Richard AI Bethlehem, Pietro Lio, and Nicola Toschi. A deep graph neural network architecture for modelling spatio-temporal dynamics in resting-state functional mri data. *Medical Image Analysis*, 79:102471, 2022.
- Itzik Malkiel, Gony Rosenman, Lior Wolf, and Talma Hendler. Self-supervised transformers for fmri representation, 2022a.
- Daniela Albuquerque, Jack Goffinet, Rachael Wright, and John Pearson. Deep generative analysis for task-based functional mri experiments. *Proceedings of machine learning research*, 149:146–175, 01 2021.
- Zhenwen Dai, Andreas Damianou, Javier González, and Neil Lawrence. Variational auto-encoded deep gaussian processes, 2016.
- Martin Jankowiak, Geoff Pleiss, and Jacob R. Gardner. Parametric gaussian process regressors, 2020.
- Jie Wang. An intuitive tutorial to gaussian processes regression, 2022.
- Takashi Matsubara, Tetsuo Tashiro, and Kuniaki Uehara. Deep neural generative model of functional mri images for psychiatric disorder diagnosis. *IEEE Transactions on Biomedical Engineering*, 66(10):2768–2779, 2019.
- Radoslaw Martin Cichy, Kshitij Dwivedi, Benjamin Lahner, Alex Lascelles, P. Iamshchinina, Monika Graumann, Alex Andonian, N. A. R. Murty, K. Kay, Gemma Roig, and Aude Oliva. The algonauts project 2021 challenge: How the human brain makes sense of a world in motion. *CoRR*, abs/2104.13714, 2021.
- Huzheng Yang, Shanghang Zhang, Yifan Wu, Yuanning Li, and Shi Gu. Effective ensemble of deep neural networks predicts neural responses to naturalistic videos. *bioRxiv*, 2021. doi:10.1101/2021.08.24.457581.
- Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- Itzik Malkiel, Gony Rosenman, Lior Wolf, and Talma Hendler. Self-supervised transformers for fmri representation. In *International Conference on Medical Imaging with Deep Learning*, pages 895–913. PMLR, 2022b.