

Извлечение признаков из магнитно-резонансной томографии с использованием атласа мозга для задач классификации

Неверов В.А., СПбГУ, Санкт-Петербург vasenkaneverov@mail.ru,
Липкович М.М., ИПМаш РАН, Санкт-Петербург lipkovich.michail@gmail.com

Аннотация

Современные исследования в области нейронаук требуют глубокого понимания функциональных связей между регионами мозга, что важно для диагностики и лечения заболеваний. Функциональная магнитно-резонансная томография является одним из инструментов для анализа этих связей. В данной работе проведено исследование с использованием методов машинного обучения для анализа связей между различными регионами.

Введение

Существующие подходы исследований в области нейронаук, такие как корреляционный анализ и трактография [1], обеспечивают определенные результаты в понимании функциональных и структурных связей между регионами мозга, однако остаются нерешенные вопросы оптимизации анализа и интерпретации данных. В данной работе рассматривается построение графа мозговых связей на основе функциональной магнитно-резонансной томографии (фМРТ) с использованием атласа Mutli-Modal Brain Atlas (MSDL) [2].

Основное направление исследования заключается в извлечении значимых признаков фМРТ, что позволит более точно оценить взаимодействия между различными регионами мозга [3]. В ходе работы была проведена интеграция методов машинного обучения для анализа данных, что может углубить понимание функциональной архитектуры мозга [4].

Задача состоит в применении фМРТ для анализа связей в мозге с использованием атласа, обучении модели для классификации данных испытуемых на предмет различий между здоровыми и больными пациентами. Цель работы заключается в построении графа мозговых связей, что позволит оценить взаимодействия между различными регионами мозга и выявить их особенности.

Ход работы

В ходе данного исследовательского проекта была проведена комплексная работа, направленная на анализ и обработку функциональных магнитно-резонансных томографических (фМРТ) данных с применением современных методов нейровизуализации и машинного обучения. Для реализации проекта использовался язык Python, а также ряд библиотек, включая `nilearn` для работы с фМРТ данными [6], `numpy` и `pandas` для манипуляций с массивами данных, `networkx` для построения графов и `scikit-learn` для задач машинного обучения.

Для анализа была загружена карта мозговых регионов из атласа MSDL с помощью функции `fetch_atlas_msdl()` [2]. Атлас содержит информацию о местоположении различных функциональных областей мозга, что важно для последующего анализа. На основе карт атласа была создана маска `NiftiMapsMasker`, позволяющая извлекать временные ряды данных фМРТ из заданных мозговых областей, стандартизируя результаты и обеспечивая необходимую нормализацию данных.

Основным компонентом разработки являлась функция `extract_features_from_subject`, обеспечивающая извлечение характерных признаков из данных фМРТ для каждого испытуемого. Процесс начинается с загрузки временных рядов, после чего строится матрица корреляции, описывающая взаимодействия между различными регионами мозга. Используемая метрика корреляции позволила учитывать нормированные значения и исключить влияние выбросов.

Для получения информации о каждом участнике исследования использовались данные из файла `participants.tsv`, который насчитывает 155 испытуемых (89 здоровых пациентов и 66 больных), загруженного с помощью библиотеки `pandas` [6]. Каждый участник был классифицирован как здоровый или пациент на основе пометок из набора данных, что было важно для дальнейшего анализа.

На следующем этапе был осуществлен сбор всех извлеченных признаков и соответствующих меток в массивы X и y . Признаки, полученные из графового анализа, были объединены в одном массиве, что позволяет эффективно использовать их для обучения модели. Данные были сохранены с помощью библиотеки `pickle` для обеспечения быстрого доступа к собранной информации.

Признаки, собранные на этапе графового анализа, представляют собой количественные характеристики, которые описывают взаимодействия между различными регионами мозга. В процессе извлечения признаков из фМРТ данных для каждого испытуемого было рассчитано несколько ключевых мет-

рик, отражающих структурные и функциональные свойства когнитивных процессов. Полученные признаки включали:

- **Степень центральности** (degree centrality): показывает количество связей, имеющих у каждого узла (региона мозга), что позволяет выявить наиболее активные области.
- **Промежуточная центральность** (betweenness centrality): измеряет значимость узла в качестве “моста” между другими узлами, что помогает выявлять ключевые регионы, влияющие на передачу информации между различными частями мозга.
- **Собственная центральность** (eigenvector centrality): определяет важность узла в сети, принимая во внимание не только количество соединений, но и значимость соседних узлов.
- **Коэффициент кластеризации** (clustering coefficient): характеризует степень, с которой узлы группируются в плотные кластеры, что может указывать на функциональные ансамбли, работающие совместно.

Эти измерения предоставляют глубокое понимание функционирования мозга и позволяют выявить паттерны, характерные для различных когнитивных состояний.

Разделение данных на обучающую и тестовую выборки: Для разделения данных мы использовали метод, который случайным образом делит весь набор данных на две части: обучающую выборку и тестовую выборку. В нашем случае размер обучающей выборки составил 124 испытуемых, что составляет примерно 80% от общего объема данных, в то время как тестовая выборка включает 31 испытуемого, что составляет около 20%. Такое соотношение позволяет обеспечить достаточное количество данных для обучения модели, сохраняя при этом достаточно данных для проверки ее производительности. Важно отметить, что при разделении данных мы учитывали случайное распределение, чтобы избежать смещения в выборках. Это означает, что каждая запись в наборе данных имела равные шансы попасть как в обучающую, так и в тестовую выборку.

Для предсказательного анализа была выбрана модель Random Forest, обеспечивающая хорошую производительность даже при наличии большого количества признаков. Мы использовали GridSearch чтобы найти оптимальные параметры для RandomForest (например, сколько деревьев использовать и какой порог для классификации лучше). Это помогает сделать модель точнее. Обучили модель на обучающей выборке с использованием кросс-валидации(5 фолдов). Разбили обучающие данные на 5 частей (фолдов), поочередно обучали модель на 4 частях и проверяли на 1, чтобы убедиться, что

модель хорошо работает на новых данных и не переобучается. Для каждого фолда мы перебирали разные пороги классификации, вычисляли метрики (точность, F1-скор, ROC-AUC) и выбирали порог, при котором модель лучше всего предсказывает данные. Финальным этапом работы стало использование обученной модели для проведения предсказаний на тестовых данных, что позволило оценить точность модели и выявить возможные области для улучшения.

Результаты работы

В результате анализа фМРТ, модель продемонстрировала точность 83.12% на тестовой выборке, что подтверждает её способность различать классы участников. Значения метрик указаны на рисунке 1.

Final evaluation on test set:					
Best threshold: 0.30					
	precision	recall	f1-score	support	
0	0.88	0.78	0.82	18	
1	0.73	0.85	0.79	13	
accuracy			0.81	31	
macro avg	0.80	0.81	0.80	31	
weighted avg	0.82	0.81	0.81	31	
ROC-AUC: 0.8312					

Рис. 1: Значения метрик классификации

В ходе работы была выполнена визуализация графов (рисунок 2) и корреляционных матриц.

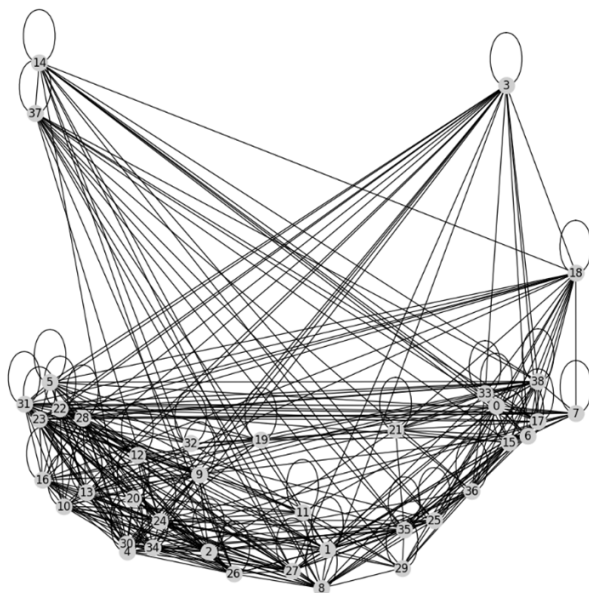


Рис. 2: Визуализация графа, где 1–38 — регионы мозга

Граф строится, чтобы визуализировать и анализировать связи между регионами мозга как сеть, где узлы — это регионы, а ребра — их взаимодействия [3].

Заключение

В ходе проведенного исследования с использованием фМРТ и атласа MSDL удалось достичь результатов в анализе связей между различными регионами мозга. Построение графа мозговых связей, основанного на данных, позволило выявить ключевые взаимодействия и особенности функциональной архитектуры мозга. Полученная модель показала точность 83.12%. Дальнейший анализ с использованием методов машинного обучения нацелен на предсказание разнообразных состояний пациентов и может стать инструментом для улучшения диагностики и лечения [7].

Список литературы

- [1] Doyen S. What is tractography? [Электронный ресурс]. — 2022. URL: <https://www.o8t.com/blog/tractography>
- [2] Sughrue M. What are the brain maps? [Электронный ресурс]. — 2022. URL: <https://www.o8t.com/blog/brain-maps>
- [3] Sughrue M. What are brain networks? [Электронный ресурс]. — 2022. URL: <https://www.o8t.com/blog/brain-networks>
- [4] Doyen S. How we build personalized brain maps [Электронный ресурс]. — 2022. URL: <https://www.o8t.com/blog/structural-connectivity-atlas>
- [5] Sughrue M. How Personalized Mapping of Brain Networks aids Neurosurgery [Электронный ресурс]. — 2022. URL: <https://www.o8t.com/blog/how-personalized-mapping-of-brain-networks-aids-neurosurgery>
- [6] OpenNeuro Dataset [Электронный ресурс]. URL: <https://openneuro.org/datasets/ds005366/versions/1.2.0>
- [7] Sughrue M. The trouble with calling brain regions 'eloquent' [Электронный ресурс]. — 2023. URL: <https://www.o8t.com/blog/the-trouble-with-calling-brain-regions-eloquent>