

Обнаружение скачкообразных смещений головы в сеансе функциональной МРТ на основе машинного обучения

А.С Плиско¹, П.Г. Серафимович^{1,2}, А.В. Никоноров^{1,2}, Ю.А. Ковш³

¹Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34А, Самара, Россия, 443086

²Институт систем обработки изображений РАН - филиал ФНИЦ «Кристаллография и фотоника» РАН, Молодогвардейская 151, Самара, Россия, 443001

³Yale University, 300 Cedar Street, New Haven, CT, 06519, USA

Аннотация. Обнаружение артефактов в данных функциональной МРТ (фМРТ) это актуальная проблема для исследовательских и клинических приложений. Движение головы пациента является одним из основных источников артефактов в фМРТ данных, а также остаточные артефакты в обработанных фМРТ данных остаются взаимосвязанными с антропоморфическими, поведенческими и клиническими факторами. Точная характеристика артефактов движения головы остается неисследованной. В настоящей работе предлагается алгоритм обнаружения скачкообразных изменений положения головы в сеансе фМРТ на основе классификаторов k-ближайших соседей и метода опорных векторов (SVM). Использованы данные движения головы шести человек в виде наборов из шести стандартных параметров движения, полученных при помощи алгоритма корегистрации фМРТ данных. Разработан программный модуль полуавтоматической разметки обучающей выборки для поиска искомых аномалий. При разметке используется статистический метод обнаружения искомых аномалий на основе скользящего окна, часть скачкообразных артефактов размечается вручную. Далее, размеченная выборка используется для обучения и проверки классификаторов. Предложенный метод идентификации скачкообразных артефактов на базе классификатора k-ближайших соседей показал высокую точность детектирования и мог бы использоваться для определения артефактов заданного типа, вызванных движением головы в фМРТ данных.

1. Введение

Данная статья посвящена выделению движения заданного типа во временных рядах параметров движения головы пациента во время проведения функциональной магнитно-резонансной томографии (фМРТ) [1], [11]. Обнаружение артефактов в данных фМРТ актуальная проблема для исследовательских и клинических приложений. Движение головы пациента является одним из основных источников таких артефактов, однако остается неисследованным то как определенные типы движения головы влияют на качество сигнала

фМРТ. Кроме того, в работах [6] и [7] были рассмотрены артефакты фМРТ данных типа выброс (spike) и тренд (drift). Алгоритмы компенсации этих артефактов могут также терять точность при резких значительных смещениях головы типа скачок.

Интеллектуальный анализ данных, используется для выделения новой значимой информации из большого объема данных. В условиях постоянного увеличения этих объемов, а также возрастающей значимости результатов их анализа вопрос идентификации имеющихся в них аномалий и артефактов стоит особенно остро. Результаты анализа без предварительного исключения аномальных экземпляров данных могут быть значительно искажены [8], [9].

Для предварительного обнаружения артефактов или закономерностей заданного типа в данных может быть использован целый ряд статистических методов [10]. Несмотря на свою универсальность, данные методы сложно адаптировать для конкретного класса данных, кроме того, для большинства из них требуется выполнять подбор параметров для конкретной реализации временного ряда. Методы машинного обучения напротив, могут быть адаптированы к специфике конкретного набора данных, и не требуют дальнейшего подбора параметров [9]. Однако для обучения классификатора аномалий требуется наличие размеченного набора данных. Для рассматриваемой нами задачи обнаружения движения заданного типа в параметрах движения головы пациента в сеансе фМРТ такие размеченные данные отсутствуют.

В настоящей работе предлагается использовать статистические методы обнаружения совместно с методами машинного обучения. Мы использовали данные движения головы здоровых волонтеров во время стандартных фМРТ сессий отдыха с закрытыми глазами. Статистические методы обнаружения были использованы для детектирования аномалий на обучающем наборе данных и создания предварительной разметки обучающей выборки. Далее, часть скачкообразных артефактов была размечена вручную, при помощи созданного программного модуля в Python. На размеченном таким образом наборе данных было выполнено обучение и проверка классификаторов, для обнаружения скачкообразных изменений в данных.

2. Статистические методы обнаружения скачкообразных изменений

В статистике и обработке сигналов обнаружение скачкообразных изменений (сглаживание или обнаружение скачков) – это процесс обнаружения резких изменений (шагов, скачков, сдвигов) на среднем уровне временного ряда или сигнала. Пример скачкообразных изменений приведен на рисунке 1. Задача обнаружения скачков может быть решена как частный случай статистического подхода к обнаружению изменений в сигнале или обнаружения точки изменения сигнала. При небольшой амплитуде скачка и зашумлении временного ряда проблема обнаружения усложняется, поскольку скачок может быть скрыт шумом.

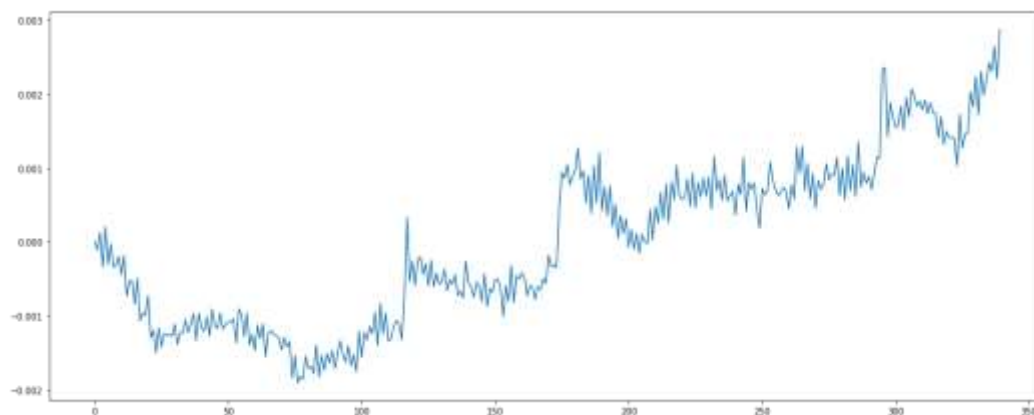


Рисунок 1. Временной ряд одного из параметров движения головы в сеансе МРТ сканирования с характерными скачкообразными изменениями.

В настоящей работе был выбран алгоритм оконной сегментации для обнаружения аномалий [10]. Алгоритм использует два скользящих окна, которые двигаются по временному ряду как показано на рисунке 2. Мера расхождения для некоторой заданной функции стоимости $c(\cdot)$ выглядит следующим образом:

$$d(y_{u..v}, y_{v..w}) = c(y_{u..w}) - c(y_{u..v}) - c(y_{v..w}).$$

где $\{y_t\}_t$ – входной сигнал, а $u < v < w$ индексы.

Если оба окна $u..v$ и $v..w$ попадают в сегмент, их статистические свойства аналогичны, и расхождение между первым окном и вторым окном является низким. Если окна делятся на два разнородных сегмента, расхождение значительно выше, что позволяет предположить, что граница между окнами является точкой изменения (т.е., искомой аномалией). Такая искомая аномалия может соответствовать идентификации параметра смещения головы типа скачок. Мы использовали наименьшее квадратичное отклонение в качестве функции стоимости. Последовательный поиск точек изменения был выполнен на кривой несоответствия, чтобы обнаружить искомые аномальные значения. На рисунке 2 представлен пример сигнала и поиск скачкообразных изменений алгоритмом скользящего окна.

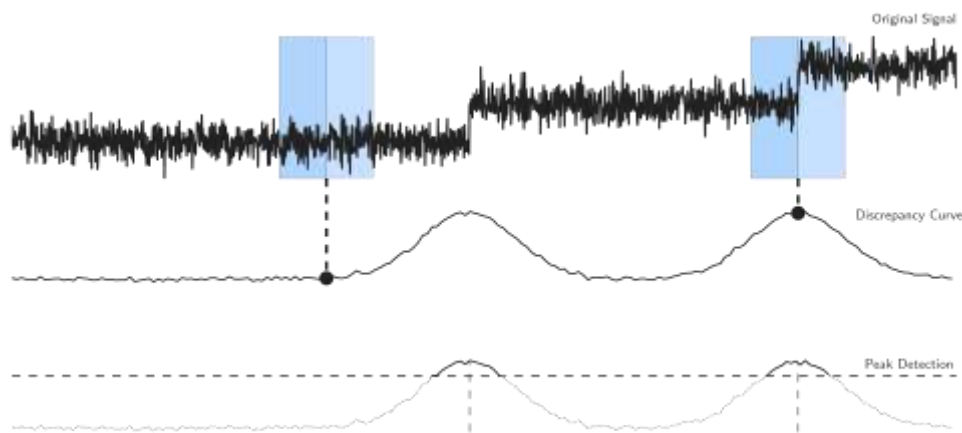


Рисунок 2. Поиска скачкообразных изменений алгоритмом скользящего окна.

На рисунке 3 показаны примеры сегментации при различных предполагаемых количествах искомым аномалий на участке. Однако точное количество скачкообразных точек не известно, что является одним из главных препятствий при поиске аномальных значений.

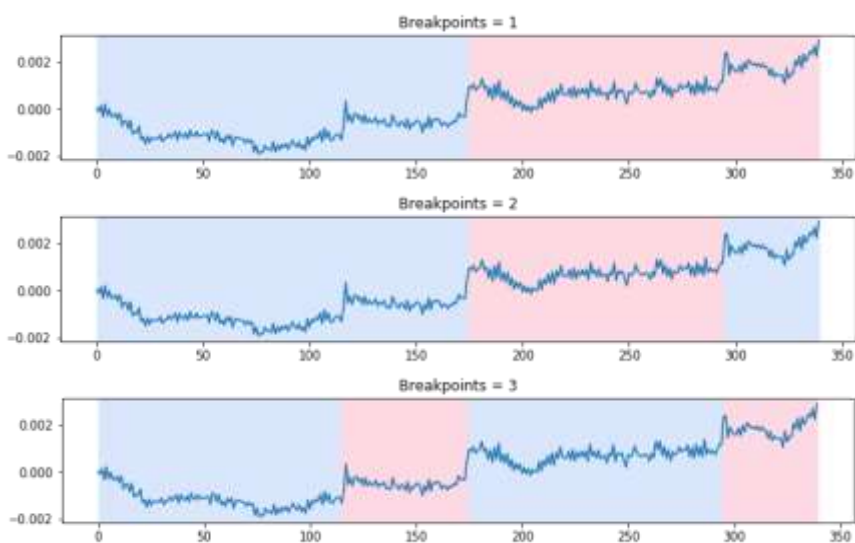


Рисунок 3. Результат работы при выборе алгоритма скользящего окна.

В ряде случаев при неизвестном количестве аномальных точек могут быть использованы различные ограничения, такие как линейные или относительные штрафные функции, рассчитываемые при оценке качества обнаружения. Согласно [10] значение штрафа, связано с амплитудой обнаруженных изменений: при небольшом штрафе (по сравнению со значениями функции несоответствия) обнаруживается множество точек изменения, даже те, которые являются результатом шума. Большое значение штрафа выявляет только самые существенные изменения или же их отсутствие. Параметр штрафа позволяет повысить точность обнаружения, однако также является неизвестным для произвольного сигнала.

3. Программный модуль уточнения разметки данных

Поскольку для корректной работы статистических алгоритмов необходимо для каждого конкретного сигнала указывать параметры обнаружения, количество искомых аномалий или величину штрафа, была реализована программа, которая позволяет пользователю уточнить разметку данных вручную, подобрав требуемые значения параметров. Изначально каждый из экземпляров данных предварительно обрабатывался с помощью уже известного алгоритма скользящего окна. Сигнал отображался с предполагаемыми участками аномалий. Далее, в зависимости от приемлемости полученных результатов, аномалии типа скачок были проверены визуально и промаркированы на графике. Статус найденной аномалии также может быть изменен на противоположный. Для реализации программного модуля был использована библиотека PyQt5 [3]. Реализация алгоритма скользящего окна была взята из библиотеки Ruptures, пакета для анализа и сегментации нестационарных сигналов [10]. Интерфейс разработанного программного модуля изображен на рисунке 4.



Рисунок 4. Интерфейс программного модуля.

4. Построение и обучение модели машинного обучения

Мы использовали данные движения головы 6 здоровых волонтеров во время стандартных фМРТ сессий отдыха с закрытыми глазами (5 мин, 323 фМРТ скана головного мозга, $120 \times 120 \times 48$ объем, 1.8мм iso воксели, TR = 1сек). Данные смещения головы были получены в виде наборов из шести стандартных параметров движения (три смещения вдоль осей x y z и вращения вокруг них) в результате стандартных процедур корегистрации фМРТ сессии по отношению к среднему объему этой сессии [12].

Для обнаружения аномальных значений типа скачок мы использовали методы машинного обучения. Размеченные данные для тренировочной и тестовой выборок были получены с помощью созданного программного модуля для уточнения разметки данных. Для обнаружения искомых аномалий мы разделили отрезки временных данных на два класса: первый класс – аномальные, и нулевой класс – не аномальные.

Для классификации были использованы два алгоритма машинного обучения, k -ближайших соседей [4] и SVM. Метод k -ближайших соседей – метрический классификатор, основанный на оценивании сходства объектов. Классифицируемый объект относится к тому классу, которому принадлежат ближайшие к нему объекты обучающей выборки [5]. SVM – метод обучения, используемый в задачах классификации и регрессии, строящий оптимальную разделяющую гиперплоскость между векторами разных классов. Были использованы линейные, полиномиальные ядра SVM, а также ядра на основе радиальных базисных функций (RBF) [12].

При обучении классификатора была использована кросс-валидация по K блокам. В этом случае исходный набор данных разбивается на K одинаковых по размеру блоков. Из K блоков один оставляется для тестирования модели, а остающиеся $K-1$ блока используются как тренировочный набор. Процесс повторяется K раз, и каждый из блоков используется один раз как тестовый набор.

5. Результаты классификации обученной модели

Для экспериментальной классификации был создан набор данных, состоящий из 1076 отрезков временных рядов, полученных при разметке (538 аномальных и 538 не аномальных). Длина каждого отрезка 24 отсчетов.

В таблице 1 и таблице 2 показаны результаты оценки алгоритмов k -ближайших соседей и SVM методом кросс-валидации на размеченной выборке, соответственно. Из таблицы 1 видно, что лучшую точность (0.703) метод k -ближайших соседей показывает при трёх ближайших соседях и 10 блоках валидации. В тоже время, SVM дает максимальную точность при выборе ядра RBF и 5 блоков валидации – 0.543 (таблица 2), что гораздо хуже, чем результат метода k -ближайших соседей.

Таблица 1. Точность для алгоритма ближайших соседей.

Число соседей	3	5	10	15	25
Число блоков					
3	0.674	0.684	0.639	0.638	0.577
5	0.684	0.695	0.656	0.636	0.599
7	0.691	0.692	0.648	0.636	0.598
10	0.703	0.688	0.657	0.640	0.607

Таблица 2. Точность для метода опорных векторов.

Число блоков	Ядро	Linear	Poly	RBF
3		0.506	0.507	0.539
5		0.506	0.503	0.543
7		0.493	0.508	0.538
10		0.507	0.513	0.533

6. Заключение

В рамках данной работы был реализован программный модуль для уточнения разметки данных коррекции параметров движения при снятии данных фМРТ, позволяющий отметить искомые аномальные и не аномальные участки сигналов движения головы типа скачок. При разработке был использован алгоритм скользящих окон из библиотеки Ruptures для предварительной разметки аномальных данных.

Были разработаны два бинарных классификатора, которые обучаются с применением методов k -ближайших соседей и опорных векторов. Был выбран оптимальный классификатор и параметр – число ближайших соседей (3 соседа) с применением техники кросс-валидации данных для метода k -ближайших соседей. Точность такой модели составила 0.703.

Предложенный метод идентификации скачкообразных артефактов на базе классификатора k-ближайших соседей мог бы использоваться для определения артефактов заданного типа, вызванных движением головы в фМРТ данных.

7. Благодарности

Данное исследование выполнено при поддержке грантов РФФИ (проекты 16-29-1-1744-офи-м, № 16-29-09528-офи-м, № 17-29-03112-офи-м, № 18-07-01390-А, № 16-07-00729-А, № 18-37-00457-мол_а), а также в рамках госзадания ИСОИ РАН - филиал ФНИЦ "Кристаллография и Фотоника" РАН (соглашение № 007-ГЗ/Ч3363/26).

8. Литература

- [1] Dosenbach, N.U.F. Real-time motion analytics during brain MRI improve data quality and reduce costs / N.U.F. Dosenbach // *NeuroImage*. – 2017. – Vol. 161. DOI: 10.1016/j.neuroimage.2017.08.025.
- [2] Koush, Y. Real-time fMRI data for testing OpenNFT functionality / Y. Koush, J. Ashburner, E. Prilepin, R. Sladky, P. Zeidman, S. Bibikov, F. Scharnowski, A. Nikonorov, D. Van De Ville // *Data in Brief*. – 2017. – Vol. 14. – P. 344-347.
- [3] Harwani, B.M. Qt5 Python GUI Programming Cookbook: Building responsive and powerful cross-platform applications with PyQt / B.M. Harwani – Packt Publishing, 2018. – P. 373.
- [4] Liao, Y. Use of K-Nearest Neighbor classifier for intrusion detection / Y. Liao, V.R. Vemuri // *Computers & Security*. – 2002. – Vol. 21(5). – P. 439-448.
- [5] Lin, W-C. An intrusion detection system based on combining cluster centers and nearest neighbors / W-C. Lin, S-W. Ke, C-F. Tsai // *Knowledge-Based Systems*. – 2015. – Vol. 78. – P. 13-21.
- [6] Koush, U. Signal quality and Bayesian signal processing in neurofeedback based on real-time fMRI / U. Koush, M. Zvyagintsev, M. Dyck, K. Mathiak, K. Mathiak // *NeuroImage*. – 2012. – Vol. 59. DOI: 10.1016/j.neuroimage.2011.07.076.
- [7] Kopel, R. No time for drifting: comparing performance and applicability of signal detrending algorithms for real-time fMRI / R. Kopel, R. Sladky, P. Laub, Y. Koush, F. Robineau, C. Hutton, N. Weiskopf, P. Vuilleumier, D. Van De Ville, F. Scharnowski // *NeuroImage*. – 2019. – Vol. 191. DOI: <https://doi.org/10.1016/j.neuroimage.2019.02.058>.
- [8] Chandola, V. Anomaly detection: A survey / V. Chandola, A. Banerjee, V. Kumar // *ACM Computing Surveys (CSUR)*. – 2009. – Vol. 15. DOI: 10.1145/1541880.1541882.
- [9] Chalapathy, R. Deep learning for anomaly detection: a survey / R. Chalapathy, S. Chawla // Cornell University, 2019. – Preprint.
- [10] Truong, C. Selective review of offline change point detection methods / C. Truong, L. Oudre, N. Vayatis // Cornell University, 2019. – Preprint.
- [11] Koush, Y. OpenNFT: An open-source Python/Matlab framework for real-time fMRI neurofeedback training based on activity, connectivity and multivariate pattern analysis / Y. Koush, J. Ashburner, E. Prilepin, R. Sladky, P. Zeidman, S. Bibikov, F. Scharnowski, A. Nikonorov, D. VanDe Ville // *NeuroImage*. – 2017. – Vol. 156. DOI: 10.1016/j.neuroimage.2017.06.039.
- [12] Zhiliang, L. Kernel Parameter Selection for Support Vector Machine Classification / L. Zhiliang, X. Hongbing // *Journal of Algorithms & Computational Technology*. – 2013. – Vol. 8. – P. 163-177.

Detection of step-like head displacements in fMRI head motion data based on machine learning

A.S. Plisko¹, P.G. Serafimovich^{1,2}, A.V. Nikonorov^{1,2}, Y. Koush³

¹Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086

²Image Processing Systems Institute of RAS - Branch of the FSRC "Crystallography and Photonics" RAS, Molodogvardejskaya street 151, Samara, Russia, 443001

³Yale University, 300 Cedar Street, New Haven, CT, 06519, USA

Abstract. Accurate artifacts detection in functional magnetic resonance imaging (fMRI) data is important in clinical applications and research. Subjects head movement remains the major source of fMRI artifacts, and retained head motion in the pre-processed fMRI data could also predict anthropomorphic, behavioral and clinical factors. However, an accurate characterization of subject head motion artifacts is lacking. We searched for step-like displacements of the subject's head in fMRI head motion data using k-nearest neighborhood classification and support vector machine (SVM). Head motion data of six subjects were defined using conventional six head motion parameters as produced by fMRI realignment procedure. We created semi-automatic markup tool for preparing training head motion data-set for classification. The semi-automatic markup was done using the sliding-window statistical anomaly detection and manual mark up of a few characteristic step-like artifacts. Marked up training data set was used to train the k-nearest neighborhood and SVM classifiers and to classify the step-like head displacements. Our approach based on k-nearest neighborhood classification showed the high accuracy in detection of step-like head displacements, which could be used for an accurate detection of specific fMRI artifacts associated with head motions.