



Факультет компьютерных наук

Образовательная программа
«Программная инженерия»

Москва
2024

Индивидуальный исследовательский курсовой проект на тему:

Нахождение скрытых функциональных состояний по ЭЭГ, основываясь на топологических признаках

Finding Hidden Functional States from EEG Based on Topological Features

Исполнитель:

Студент группы БПИ213
Абрамов Александр Сергеевич

Руководитель:

Доцент департамента математики
факультета экономических наук НИУ ВШЭ,
кандидат физико-математических наук
Михайлец Екатерина Викторовна

Консультант:

Доцент департамента больших данных и
информационного поиска ФКН НИУ ВШЭ,
кандидат физико-математических наук
Чернышев Всеволод Леонидович



Предметная область

➔ Нахождение границ функциональных состояний по ЭЭГ **непрерывного** процесса:

Невозможность использования временных ориентиров

Неизвестное количество
искомых состояний

Необходимость
автоматизации процесса

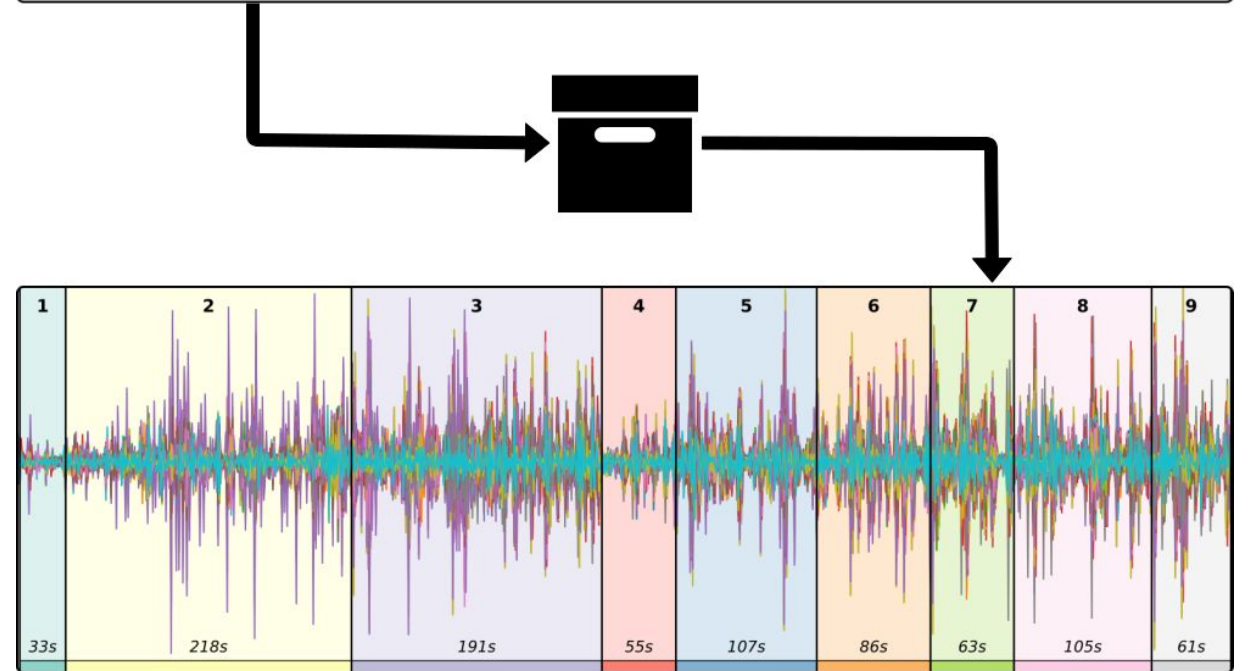
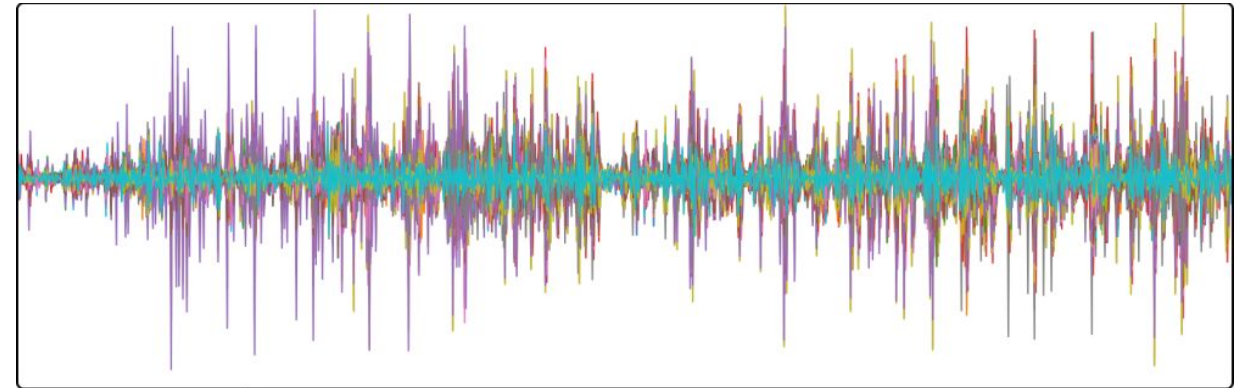
➔ **Важно** для проведения исследований во многих областях нейробиологии и медицины:

Сомнология

Диагностика эпилепсии

Самоиндуцированные
функциональные
состояния

Отвлечение внимания и
усталость





State-Detecting Algorithm

Guhyasamaja Tantra – традиционный процесс медитации, производимый в соответствии со строгими принципами, закреплёнными священными писаниями буддизма.

Сбор данных

- Регистрация сигналов с помощью системы NVX-52
- Частота дискретизации записываемого сигнала: 500 Гц
- Аналоговая полосовая фильтрация от 0,1 до 200 Гц
- Режекторный фильтр на частоте 50 Гц для удаления артефактов ЛЭП

Предобработка данных

- Применение полосовой фильтрации на частотах 0,9 - 40 Гц
- Разделение сигнала путём анализа независимых компонент
- Разбиение на эпохи по времени длительностью 1 секунда
- Удаление 10 – 15% эпох, наиболее отклоняющихся от среднего

SDA

- Традиционное признаковое описание разм. 765: PSD, PLV и Coherence
- Стандартизация признаков и выделение 15 главных компонент
- Агломеративный метод иерархической кластеризации Уорда
- Поиск кандидатов итоговых границ состояний методом k-средних
- Вычисление метрик качества кластеризации
- Выбор ответа экспертным решением с учётом метрик качества



Цель и задачи

Цель работы:

Оценить применимость признакового описания ЭЭГ, полученного методами топологического анализа данных, к задаче нахождения скрытых функциональных состояний по ЭЭГ непрерывного процесса



Изучить принципы работы с
нейрофизиологическими данными



Изучить методы топологического
анализа данных



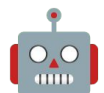
Изучить устройство и принципы
работы алгоритма SDA



Реализовать обработку ЭЭГ
топологическими методами



Извлечь топологические признаки
из ЭЭГ медитации трёх монахов



Запустить SDA на полученном
признаковом пространстве



Выявить наиболее
информационно-ценные признаки



Сделать выводы и зафиксировать
результаты работы в отчёте

Топологический анализ данных

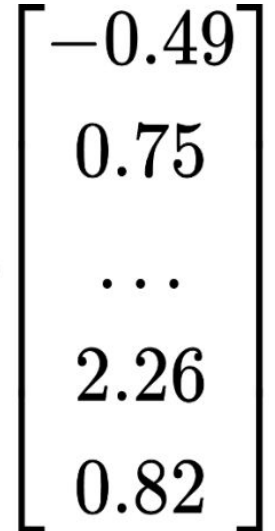
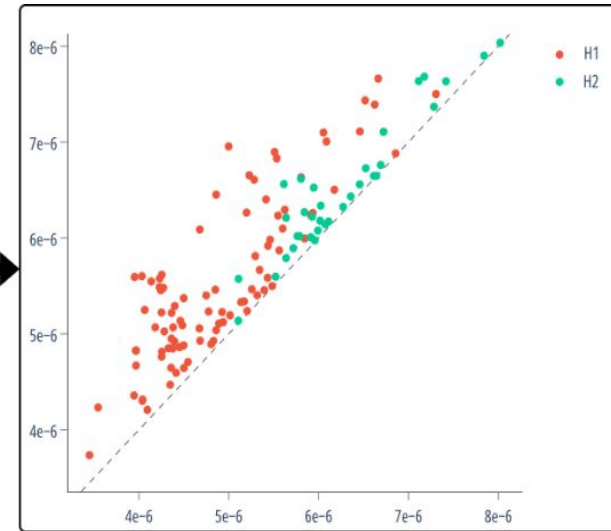
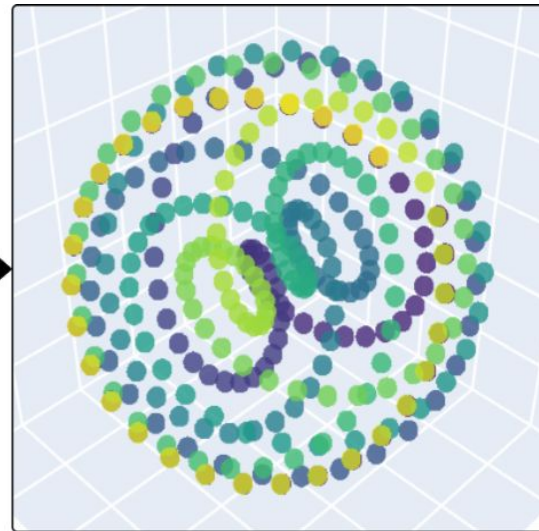
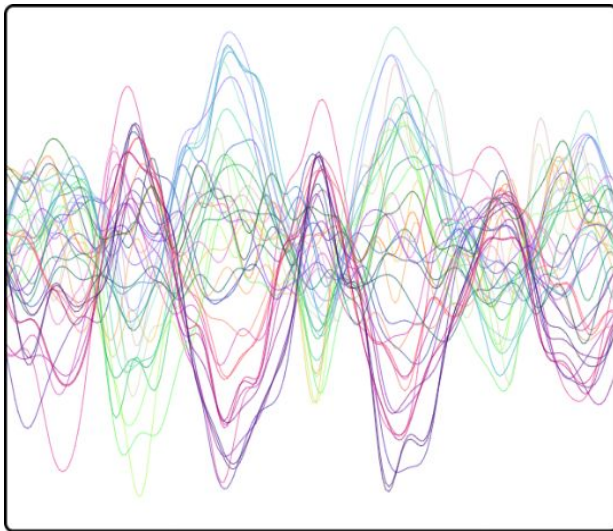
Современный подход к анализу данных различного рода, основанный на понижении размерности без значительной потери информации путём изучения пространственных характеристик данных (формы, структуры и др.) методами алгебраической топологии.

giotto-tda (Python) [2]

Ripser (C++)

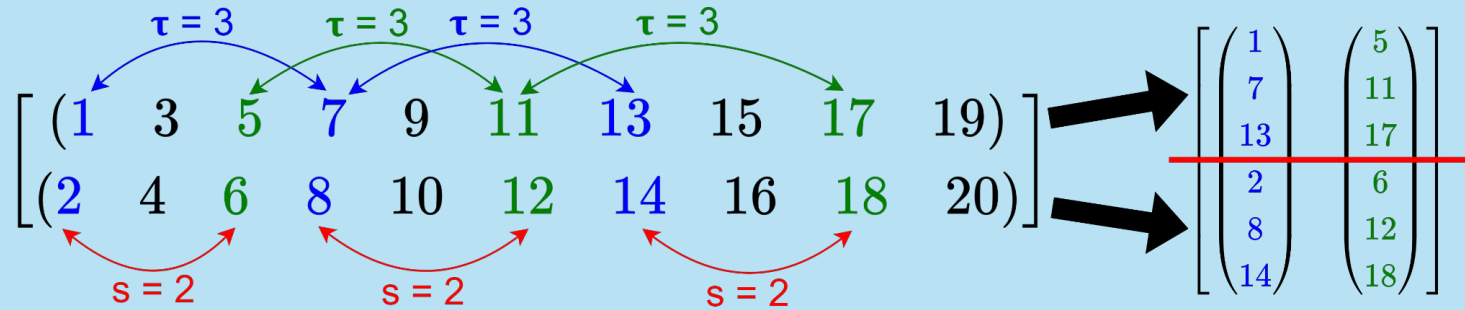
GUDHI (C++)

JavaPlex (Java)



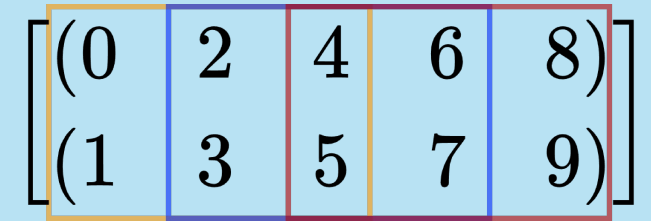
Преобразование временного ряда в метрическое пространство

➤ Алгоритм Такенса ($d = 3, \tau = 3, s = 2$)



➤ Скользящее окно

Частный случай алг. Такенса при $\tau = 1$

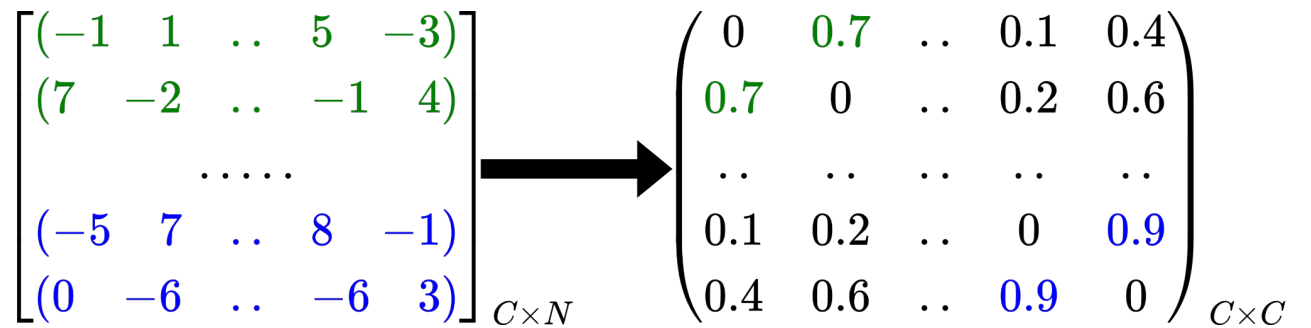


➤ Пирсоновское несходство

$$C_{XY} = \frac{1}{N} \sum_{i=1}^N \left((X_i - \frac{1}{N} \sum_{j=1}^N X_j) (Y_i - \frac{1}{N} \sum_{j=1}^N Y_j) \right)$$

$$R_{XY} = \frac{C_{XY}}{\sqrt{C_{XX} \cdot C_{YY}}}$$

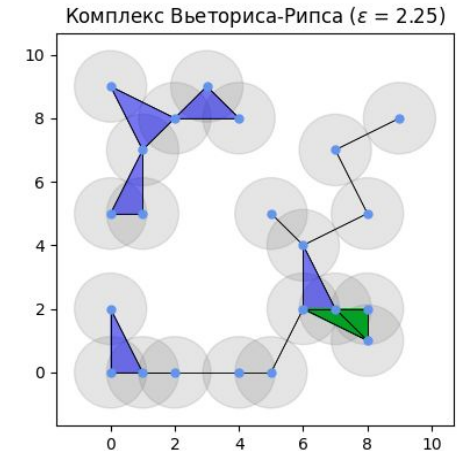
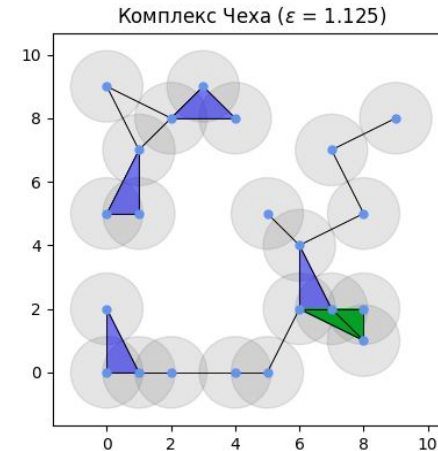
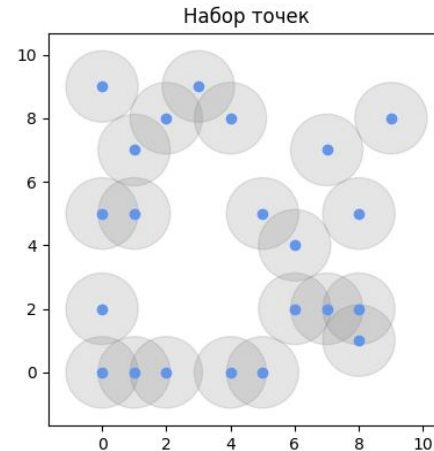
$$D_{XY} = \frac{1 - R_{XY}}{2}$$



Построение диаграмм устойчивости

Комплекс Чеха

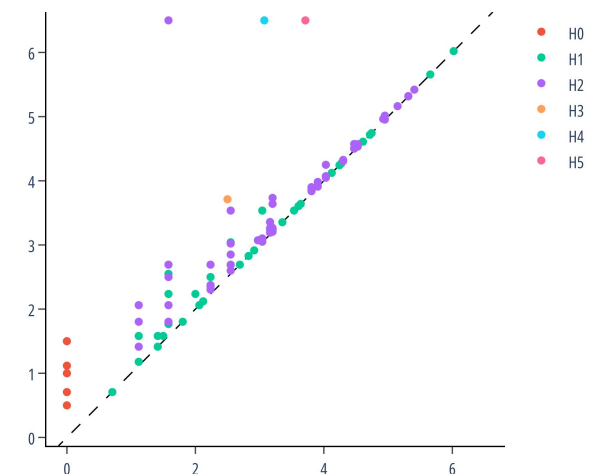
$$\check{C}_\epsilon(X, d) = \left\{ [x_1, \dots, x_n] \mid \bigcap_{i=1}^n B_{d, \epsilon}(x_i) \neq \emptyset \right\}$$



Комплекс Вьеториса – Рипса

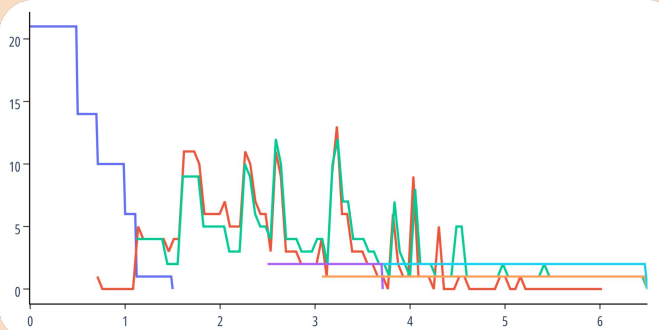
$$VR_\epsilon(X, d) = \left\{ [x_1, \dots, x_n] \mid \forall i, j \ d(x_i, x_j) \leq \epsilon \right\}$$

Диаграмма устойчивости – диаграмма, каждая точка на которой соответствует некоторому симплексу, а её координаты равны минимальному и максимальному значениям ϵ , при которых симплекс входит в соответствующий комплекс.

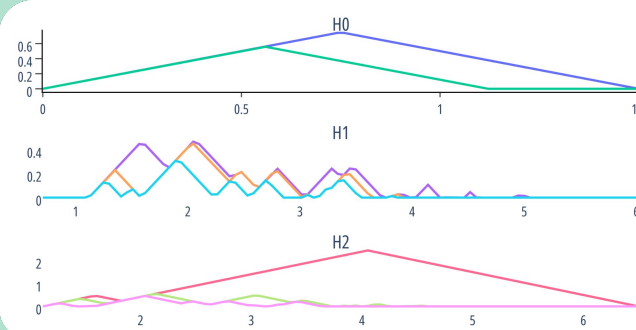


Векторизация диаграмм устойчивости

Числа и кривые Бетти



Ландшафт устойчивости



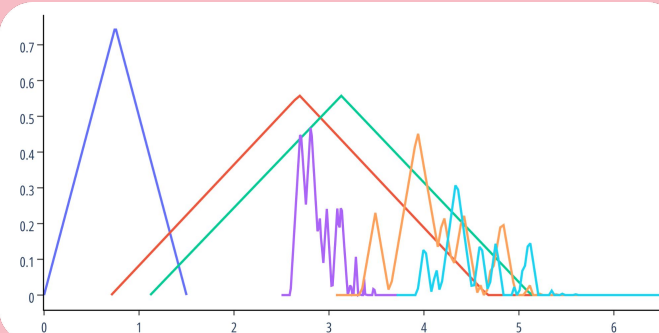
Амплитуда устойчивости

Расстояние Васерштейна

$$d_\gamma = \sqrt[p]{\sum_{x \in D_1 \cup \Delta} \|x - \gamma(x)\|_\infty^p}$$

$$W(D_1, D_2) = \min_{\gamma: D_1 \cup \Delta \rightarrow D_2 \cup \Delta} (d_\gamma)$$

Силуэт устойчивости



Энтропия устойчивости

$$p_i = \frac{d_i - b_i}{\sum_{j=1}^n d_j - b_j}$$

$$E(D) = - \sum_{i=1}^n p_i \cdot \ln(p_i)$$

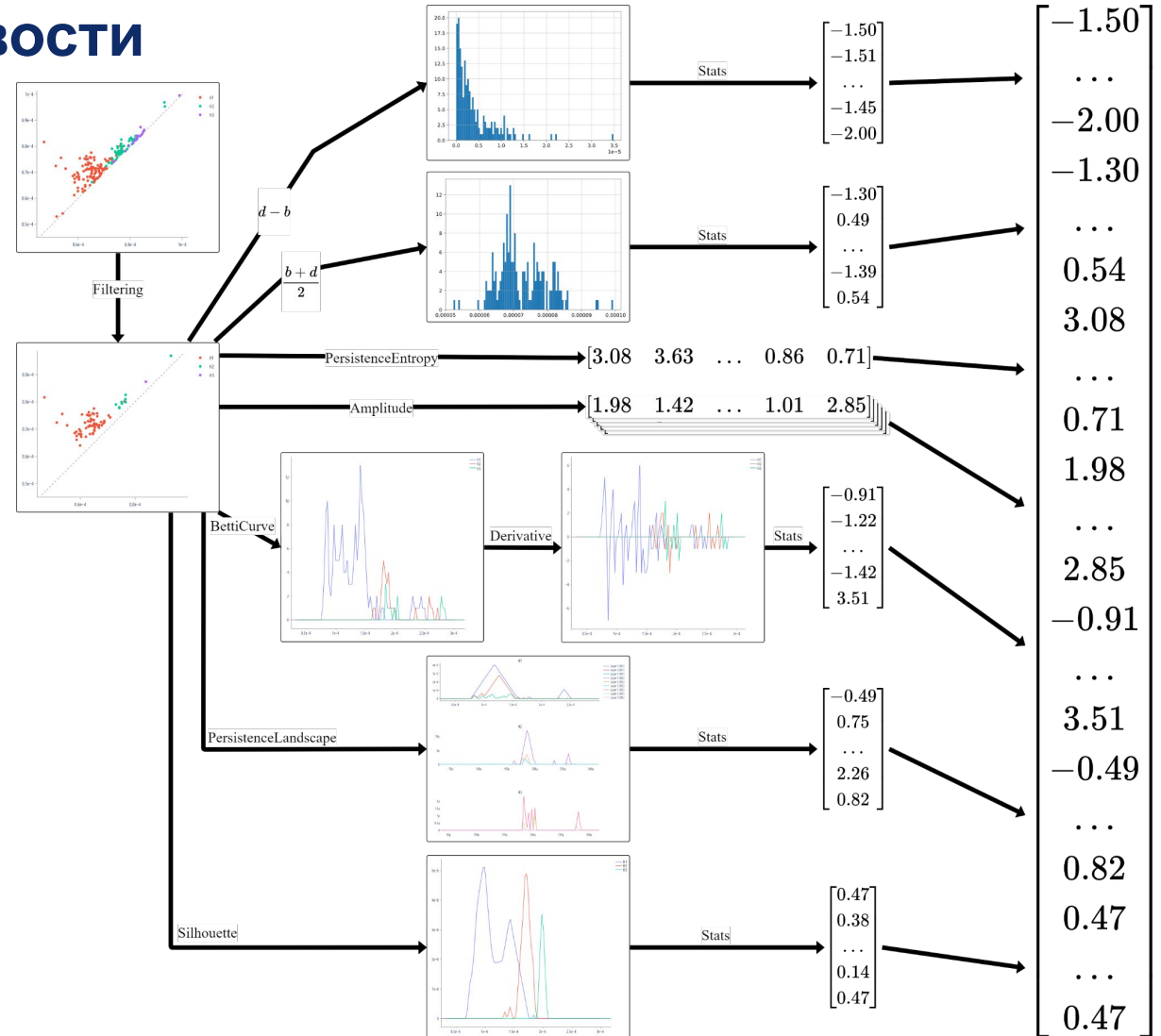
L^p и ℓ^p – расстояния

$$L^p(f, g) = \sqrt[p]{\int_U |f(x) - g(x)|^p dx}$$

$$L^p(f, g) \sim \ell^p(F, G) = \|F - G\|_p$$

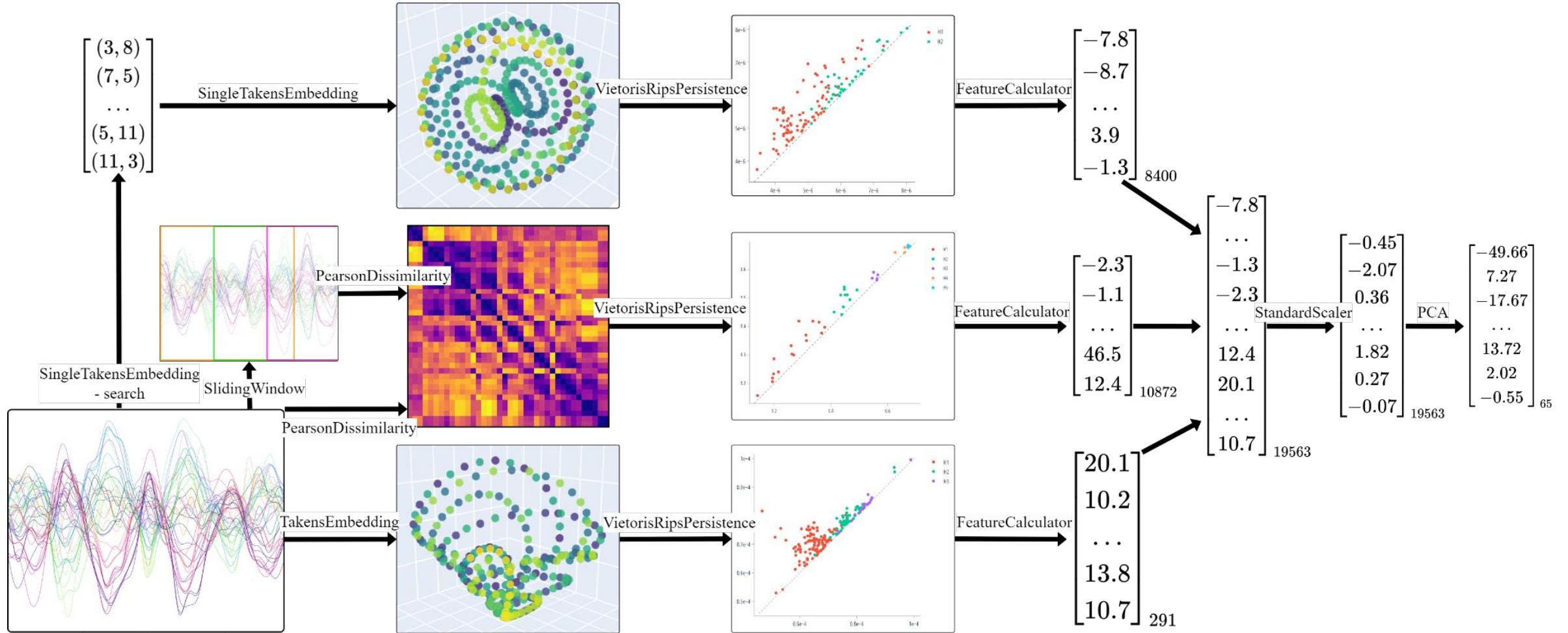
Векторизация диаграмм устойчивости

- Статистические характеристики последовательности времён жизни симплексов
- Статистические характеристики последовательности средних значений между моментами появления и исчезновения симплексов
- Энтропия устойчивости
- Набор амплитуд устойчивости и их векторные нормы первого и второго порядков
- Статистические характеристики первых производных кривых Бетти
- Статистические характеристики первых уровней ландшафтов устойчивости
- Статистические характеристики силуэтов устойчивости степеней 1 и 2





Признаковое описание ЭЭГ



Метрики оценки качества

➤ Внутренняя оценка

Достаточно ли похожи объекты в одном кластере? Сильно ли отличаются объекты в разных кластерах?

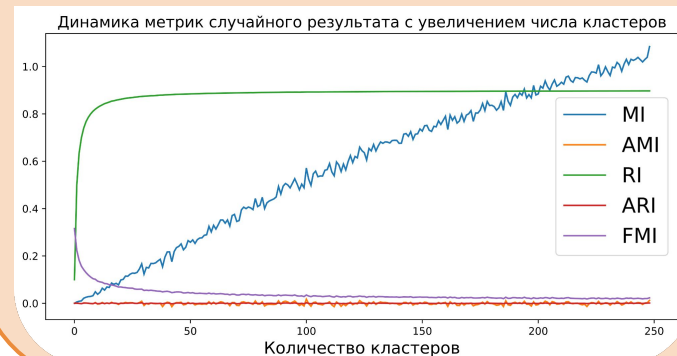
Результат усредняется по всем парам соседних функциональных состояний.

- └─ Расстояние Уорда
- └─ Центроидное расстояние
- └─ Коэффициент силуэта
- └─ Индекс Калински – Харабаса
- └─ Индекс Дэвиса – Болдина

➤ Внешняя оценка

Насколько похожи два результата?
Близки ли они к “правильному”?

- └─ Скорректированный коэфф. взаимной информации
- └─ Скорректированный индекс Рэнда
- └─ Индекс Фаулкса-Маллоуса



➤ Информационная ценность признаков

Насколько признак влияет на результат бинарной классификации категориальной переменной?

- └─ Количественные признаки разбивают на заданное количество контейнеров
- └─ Многоклассовую классиф. рассматривают как набор бинарных
- └─ Кластеризация эквивалентна многоклассовой классификации



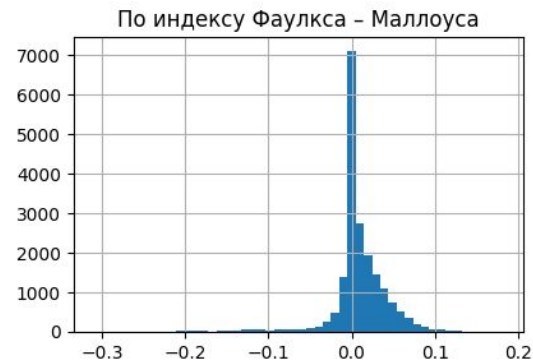
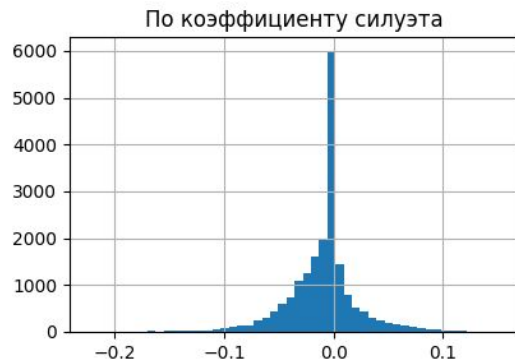
Результат для объекта 1

	Ответ на традиционных признаках	По коэффициенту силуэта	По индексу Фаулкса-Маллоуса	Ответ при объединении признаков
Расстояние Уорда	24208	149649	145861	42752
Центроидное расстояние	21.38	52.64	48.35	28.08
Коэффициент силуэта	0.199	0.070	0.055	0.234
Индекс Калински – Харабаса	69.90	19.67	19.16	85.13
Индекс Дэвиса – Болдина	1.637	3.348	3.594	1.582
Коэффициент взаимной информации		0.869	0.878	0.920
Индекс Рэнда		0.791	0.795	0.832
Индекс Фаулкса-Маллоуса		0.823	0.824	0.857

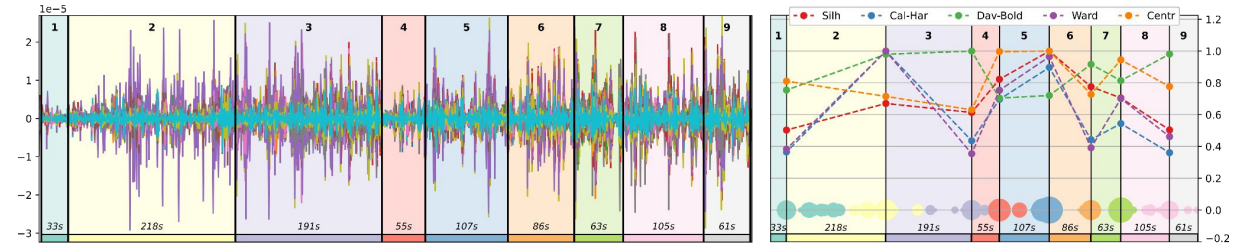


Результат для объекта 1

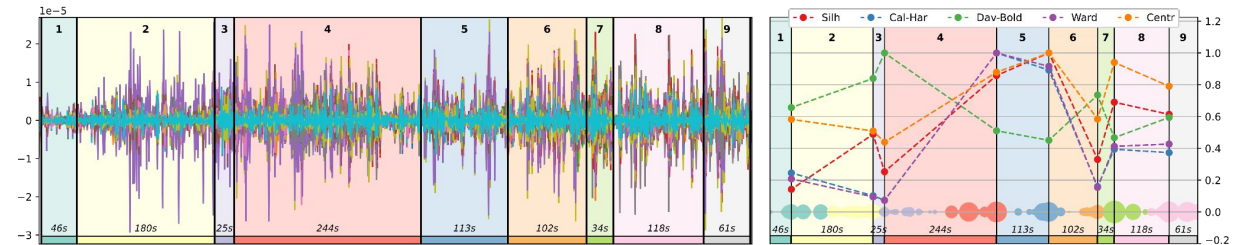
- ➔ Прослеживаются общие черты между полученными и ранее известными ответами
- ➔ Границы трудноразличимых состояний были смещены или не найдены вовсе
- ➔ При объединении признаков пространств улучшаются значения внутренних метрик
- ➔ **Высокое** сходство значений информационной ценности относительно различных результатов



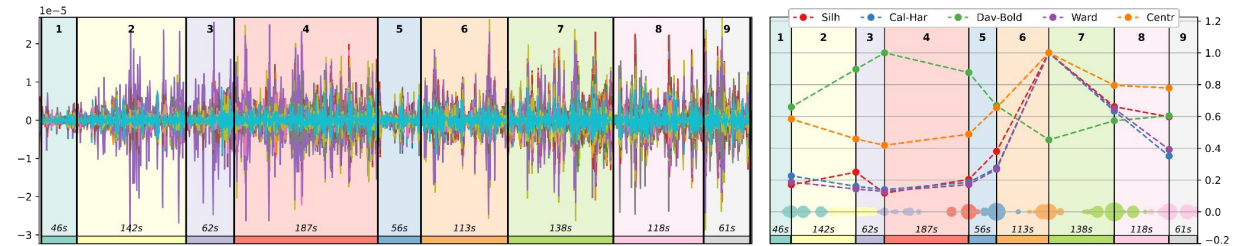
Результат на традиционных признаках



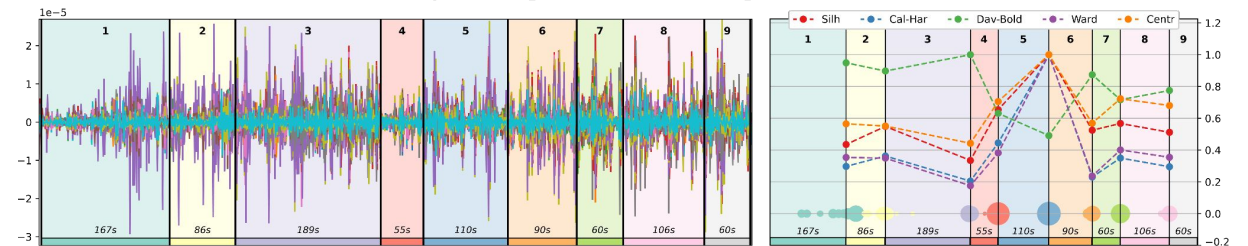
Лучший результат по коэффициенту силуэта



Лучший результат по индексу Фаулкса - Маллоуса



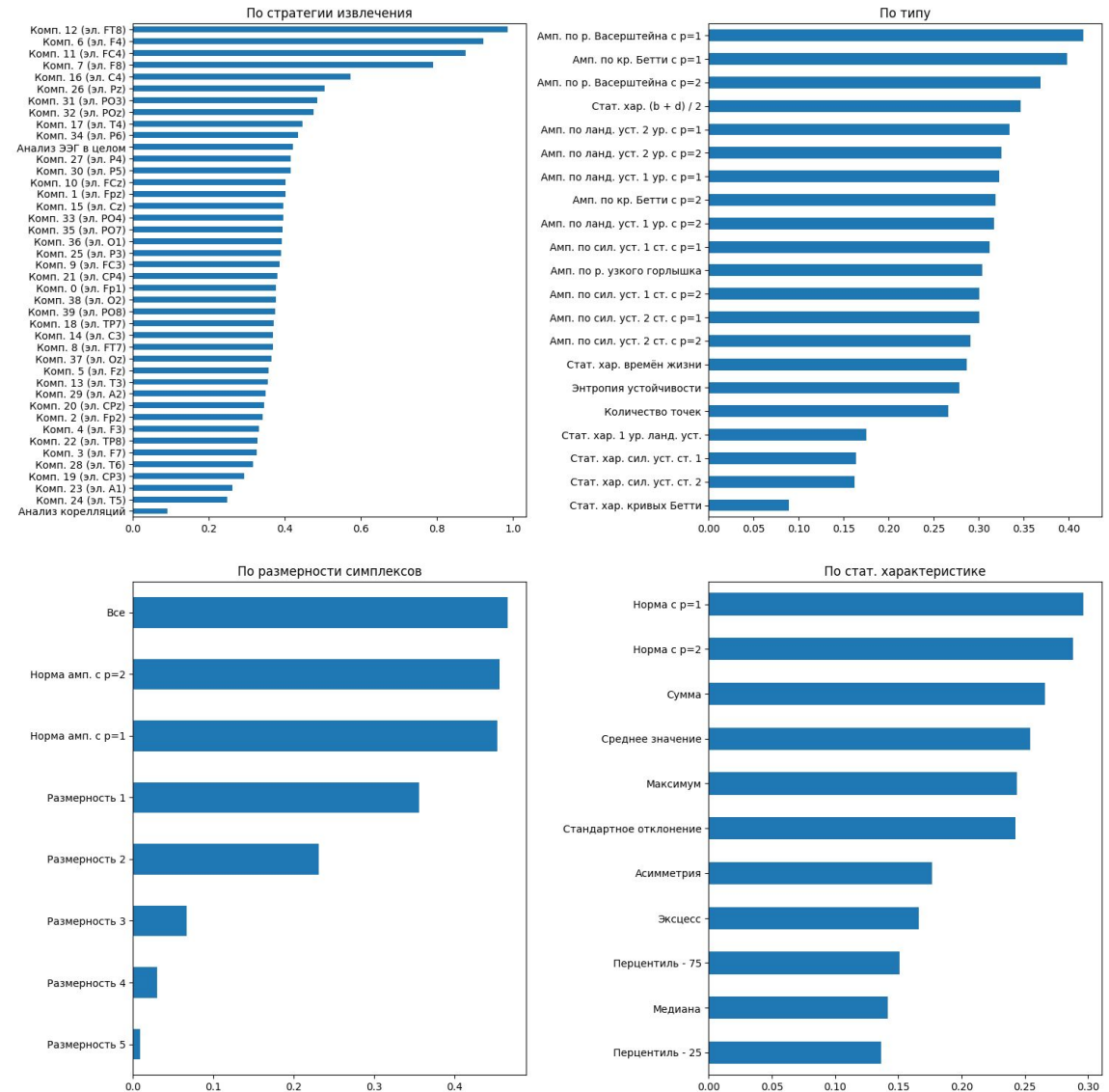
Результат при объединении признаков





Результат для объекта 1

Наиболее ценные	Наименее ценные
Фронтальная и правая части мозга	Левая и затылочная части мозга, корреляции между переменными
Статистические характеристики последовательностей $(b + d) / 2$, амплитуды по различным метрикам	Статистические характеристики кривых Бетти и силуэтов устойчивости
Совокупность симплексов всех размерностей, особенно симплексы размерностей 1 и 2	—
Манхэттенская и евклидова нормы	Перцентили (25, 50 и 75), эксцесс и асимметрия





Направления дальнейшей работы

Отбор наиболее информативных признаков

- Удаление наиболее чувствительных к “шуму” признаков
- Понижение размерности итогового пространства
- Повышение доли объясняемой дисперсии
- Сбор новых записей ЭЭГ процессов медитации других монахов

1

Анализ совместного поведения признаков

- Выявление общих закономерностей
- Исследование ключевых отличий
- Определение наборов признаков, наилучшим образом дополняющих друг друга при решении поставленной задачи

2

Оценка информационного вклада признаков

Построение и анализ метрик качества моделей многоклассовой классификации (метод опорных векторов, логистическая регрессия, XGBoost и др.), предсказывающих номер кластера по признаковому описанию эпохи

3

Привлечение экспертов

Оценка и сравнение получаемых результатов не только на основании метрик качества кластеризации, но и с учётом существующих знаний о мозговой активности человека во время изучаемого процесса, выявление присутствующих закономерностей

4



Список использованных источников

- 1) SDA: a data-driven algorithm that detects functional states applied to EEG of Guhyasamaja meditation [Электронный ресурс] / E. V. Mikhaylets, A. R. Razorenova, V. L. Chernyshev, N. V. Syrov, L. V. Yakovlev, J. A. Boytsova, E. V. Kokurina, Y. S. Zhironkina, S. V. Medvedev and A. Y. Kaplan. – Front. Neuroinform., 29 January 2024. – URL: <https://doi.org/10.3389/fninf.2023.1301718>. (дата обращения: 01.02.24).
- 2) giotto-tda: A Topological Data Analysis Toolkit for Machine Learning and Data Exploration [Электронный ресурс] / M. Rucco, F. Castiglione, E. Merelli, M. Pettini. – arXiv, 5 Mar 2021. – URL: <https://doi.org/10.48550/arXiv.2004.02551>. (дата обращения: 01.02.24).
- 3) Stability of persistence diagrams [Электронный ресурс] / D. Cohen-Steiner, H. Edelsbrunner and J. Harer. – Discrete & Computational Geometry: электрон. журн., vol. 37 (2007), pp. 103 – 120. – Springer, 12 December 2006. – URL: <https://doi.org/10.1007/s00454-006-1276-5>. (дата обращения: 01.02.24).
- 4) Topological Persistence and Simplification [Электронный ресурс] / H. Edelsbrunner, D. Letscher and A. Zomorodian. – Discrete & Computational Geometry: электрон. журн., vol. 28 (2002), pp. 511 – 533. – Springer, 01 November 2002. – URL: <https://doi.org/10.1007/s00454-002-2885-2>. (дата обращения: 01.02.24).
- 5) Computing persistent homology [Электронный ресурс] / A. Zomorodian, G. Carlsson. – Discrete & Computational Geometry: электрон. журн., vol. 33 (2005), pp. 249 – 274. – Springer, 19 November 2004. – URL: <https://doi.org/10.1007/s00454-004-1146-y>. (дата обращения: 01.02.24).
- 6) A Method for Comparing Two Hierarchical Clusterings [Электронный ресурс] / E. B. Fowkles and C. L. Mallows. – Journal of the American Statistical Association: vol. 78 (1983), pp. 553 – 569. – Taylor & Francis, 12 Mar 2012. – URL: <https://doi.org/10.1080/01621459.1983.10478008>. (дата обращения: 01.02.24).
- 7) Osteyee, D. B. [Электронный ресурс]: Information, Weight of Evidence. The Singularity Between Probability Measures and Signal Detection. / D. B. Osteyee, I. J. Good. – Lecture Notes in Mathematics: vol. 376. – Springer, 1974. – URL: <https://doi.org/10.1007/BFb0064126>. (дата обращения: 16.03.24).

Спасибо за внимание!



Абрамов Александр Сергеевич

asabramov@edu.hse.ru

Москва 2024



Факультет компьютерных наук

Образовательная программа
«Программная инженерия»

Москва
2024

Индивидуальный исследовательский курсовой проект на тему:

Нахождение скрытых функциональных состояний по ЭЭГ, основываясь на топологических признаках

Finding Hidden Functional States from EEG Based on Topological Features

Исполнитель:

Студент группы БПИ213
Абрамов Александр Сергеевич

Руководитель:

Доцент департамента математики
факультета экономических наук НИУ ВШЭ,
кандидат физико-математических наук
Михайлец Екатерина Викторовна

Консультант:

Доцент департамента больших данных и
информационного поиска ФКН НИУ ВШЭ,
кандидат физико-математических наук
Чернышев Всеволод Леонидович