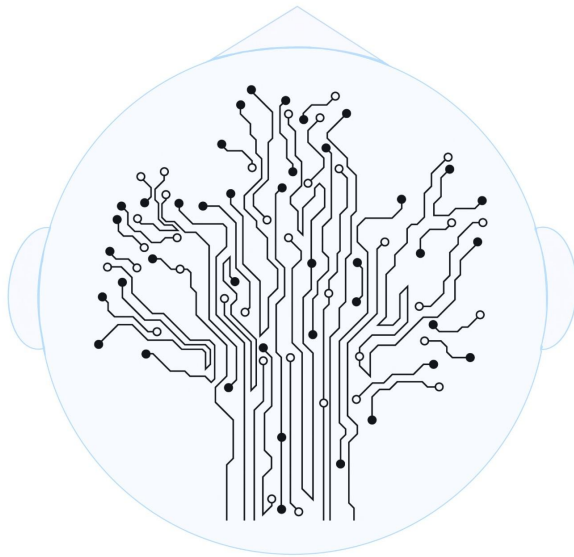


Исследовательский проект на тему: Автоматизированное определение стадий сна / Automated detection of sleep stages



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Яндекс



Выполнила:

Маркович Анна Александровна
студентка ОП ПМИ группы 219

Руководитель:

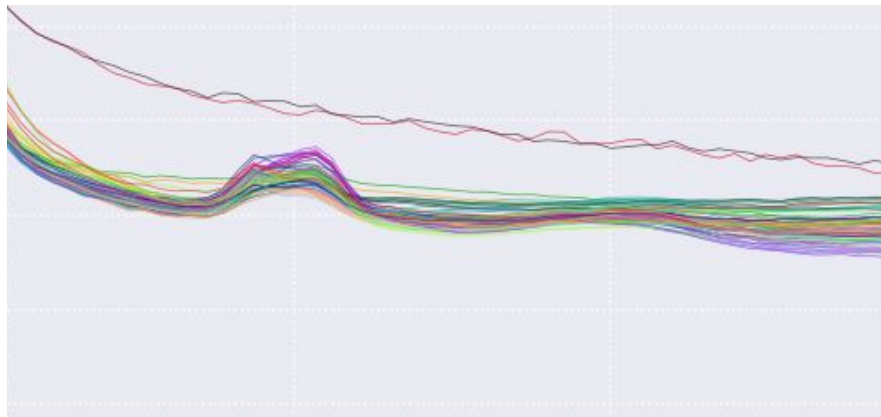
Чернышев Всеволод Леонидович
Кандидат физико-математических наук;
доцент департамента больших данных и информационного
поиска факультета компьютерных наук НИУ ВШЭ

Область исследования и актуальность

Анализ мозговой активности в нейронауке, психологии и смежных областях.

Сферы применения:

- Исследование заболеваний, связанных с мозгом
- Исследование сна
- Исследование когнитивных процессов личности в разных возрастных категориях



Проблема и потенциальное решение

Проблема: нехватка алгоритмов определения скрытых функциональных состояний в записи ЭЭГ, которые:

- учитывают непрерывность этих состояний во времени
- не требуют информации о пространстве признаков, наилучшим образом описывающих процессы
- не требуют истинной разметки стадий активности мозга

Потенциальное решение: State-detecting algorithm (SDA) – новый подход к выявлению скрытых функциональных состояний активности мозга в записи ЭЭГ, разработанный В. Л. Чернышевым, Е. В. Михайлец и А. М. Разореновой.

Цель и задачи



Цель – оценить оптимальность, результативность, эффективность SDA на данных по сну.

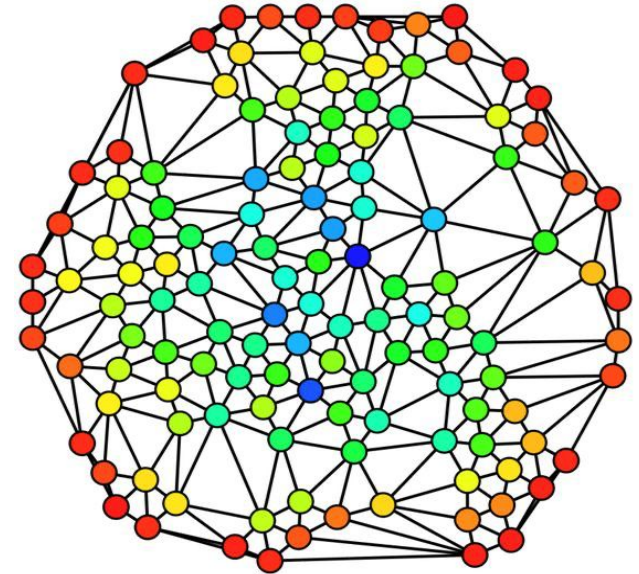
Задачи:

1. Обработать запись ЭЭГ (reference, filtering, ICA)
2. Подсчитать описательные признаки сигналов
3. Применить алгоритм кластеризации и получить разбиение записи на непрерывные стадии
4. Проанализировать поведение признаков на выделенных стадиях
5. Сравнить результаты с проверенным методом

Идея алгоритма и его структура

SDA – state detecting algorithm – метод поиска функциональных состояний в записях ЭЭГ, то есть обнаружения точек изменения в пространстве признаков.

- ◆ Основан на методе иерархической кластеризации (метод Варда)
- ◆ Учитывает непрерывность данных во времени с помощью матрицы смежности для соседних эпох
- ◆ Не требует знаний об истинной разметке стадий (no ground truth)



Структура связного неориентированного графа в пространстве эпох (пары из эпох с расстоянием, не превышающим k секунд соединены ребрами).

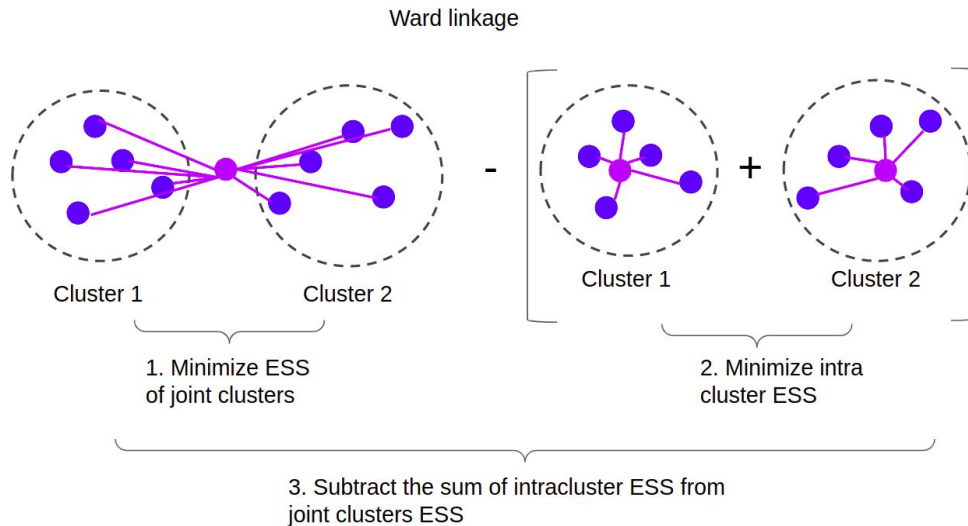
Идея алгоритма и его структура

Метод Варда – метод агломеративной иерархической кластеризации, который минимизирует дисперсию внутри кластеров.

$$D(X, Y) = ESS(X \cup Y) - (ESS(X) + ESS(Y)) = \frac{N_X * N_Y}{N_X + N_Y} \|\bar{m}_X - \bar{m}_Y\|^2$$

ESS - error sum of squares
(сумма квадратов ошибок)

N_x и N_Y – размеры, \bar{m}_X и \bar{m}_Y – центры кластеров.



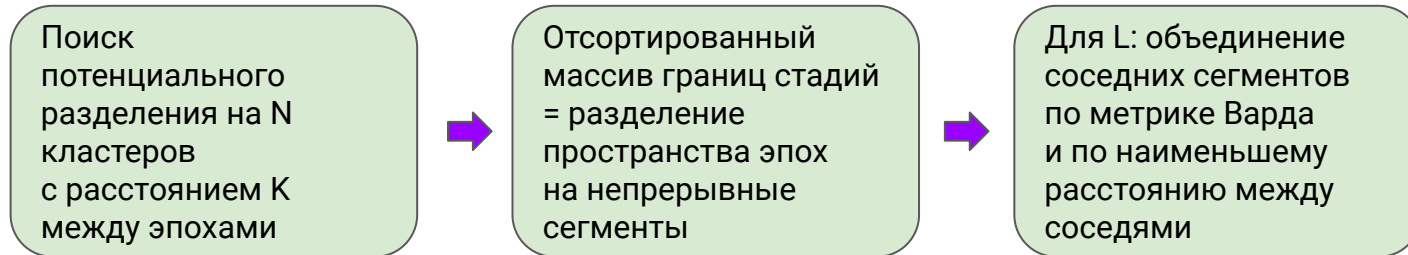
Идея алгоритма и его структура

N – количество кластеров в методе Варда

K – расстояние между эпохами для матрицы смежности

L – длина стадии для объединения соседних сегментов

Для каждой пары параметров (N, K):



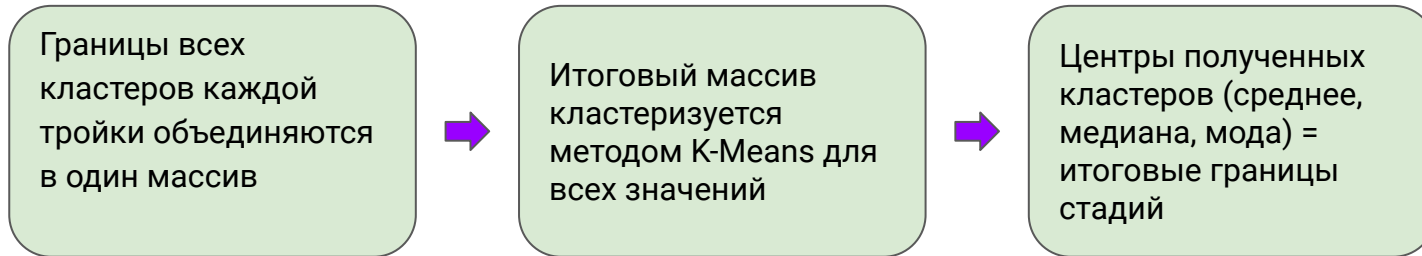
В итоге для каждой тройки параметров получается свое разделение на кластеры.

Идея алгоритма и его структура

Nmax – максимальное количество кластеров в итоговом массиве границ кластеров

Kmax – максимальное расстояние между эпохами в итоговом массиве границ кластеров

Для каждой тройки параметров (N_{\max} , K_{\max} , L) за одну итерацию перебор всех троек вида ($N \leq N_{\max}$, $K \leq K_{\max}$, L):






В итоге на основе метрик кластеризации выбирается итоговое разделение пространства эпох на стадии сна.

Идея алгоритма и структура

Метрики качества кластеризации:

- **Коэффициент Силуэта (Silhouette Coefficient)**, измеряющий, насколько объекты похожи на свой собственный кластер по сравнению с другими кластерами в среднем.
- **Индекс Калински–Харабаша (Calinski–Harabasz Index)**, также известный как критерий отношения дисперсии, характеризующий отношение дисперсии между кластерами к дисперсии внутри кластеров.
- **Индекс Дэвиса–Боулдина (Davies–Bouldin Index)**, который измеряет разделение между кластерами, сравнивая расстояние между кластерами с размером самих кластеров.

$[-1, 1]$	
$[0, \infty)$	
$[0, \infty)$	

Частотные диапазоны и классификация каналов для подсчета признаков

Все признаки вычисляются усреднением на 5 частотных диапазонах для 16 групп каналов.

Частотные диапазоны:

Delta (0.9–4 Hz),

Theta (4–8 Hz)

Alpha (8–14 Hz)

Beta (14–25 Hz)

Gamma (25–40 Hz).

Группы каналов:

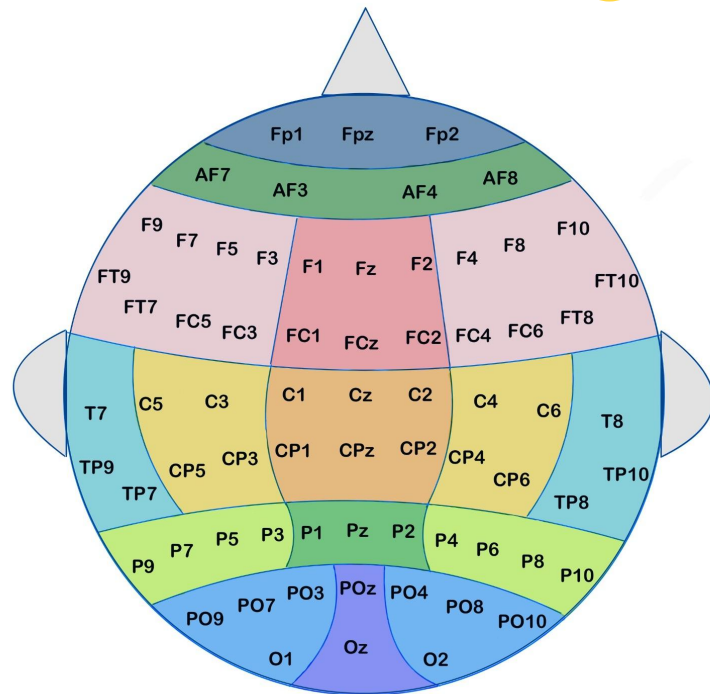
Frontal (prefrontal, in-between, left, midline, right)

Central (left, midline, right)

Temporal (left, right)

Parietal (left, midline, right)

Occipital (left, midline, right)



Признаки

- **Power spectral density** – спектральная плотность мощности, по сути мера силы сигнала. Действительное неотрицательное число.
- **PSD-индексы** – соотношения мощности сигналов на следующих частотах: theta/delta, alpha/ delta, alpha/theta, alpha/(delta+theta), beta/delta, beta/theta, beta/alpha, beta/(delta+theta) и некоторых других (всего 16).
- **Когерентность** – это мера синхронизации между двумя сигналами ЭЭГ. Для сигналов x и y для одной эпохи и частотного диапазона когерентность варьируется от 0 до 1.
- **Phase-locking value** – мера фазовой синхронизации сигналов. Для сигналов x и y и конкретной эпохи и частотной группы варьируется от 0 до 1.
- **Coherence Index и PLV Index** канала x с порогом P – это количество каналов y, $y \neq x$, где $\text{Coh}(x, y) > P$ и $\text{PLV}(x, y) > P$ соответственно, для $P \in \{0.6, 0.7, 0.8\}$.

$$\text{Coh}_{xy} = \frac{|E[S_{xy}]|}{\sqrt{E[S_{xx}]E[S_{yy}]}}$$

PSD

$$\text{PLV}_{xy} = \left| E \left[\frac{S_{xy}}{|S_{xy}|} \right] \right|$$

CSD (cross-PSD)

Использование признаков в кластеризации



Пространство признаков – векторы из посчитанных значений PSD, PSD Index, Coh Index, PLV Index для каждой эпохи



Размерность этого пространства слишком **велика**



Метод главных компонент (PCA) - понижение размерности путем SVD

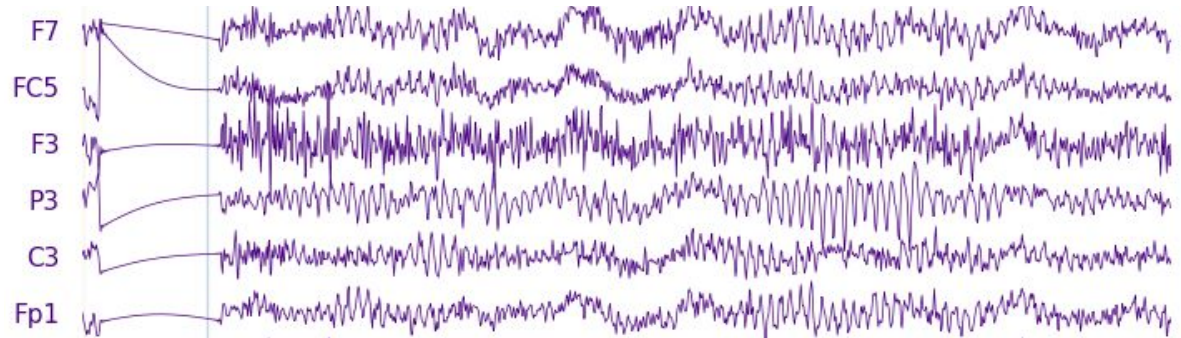


Каждая **эпоха** представима **вектором** фиксированной длины (кол-во главных компонент) в пространстве признаков для использования в алгоритме кластеризации



Испытание SDA на данных

Предобработка



● Референс
(average referencing)

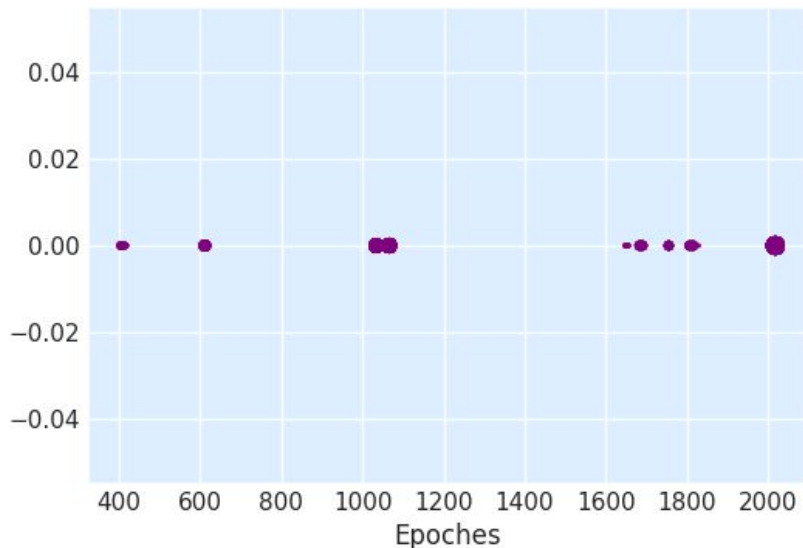
● Фильтрация
0.9 - 40 Hz

● Очистка от компонентов,
не связанных
с активностью мозга

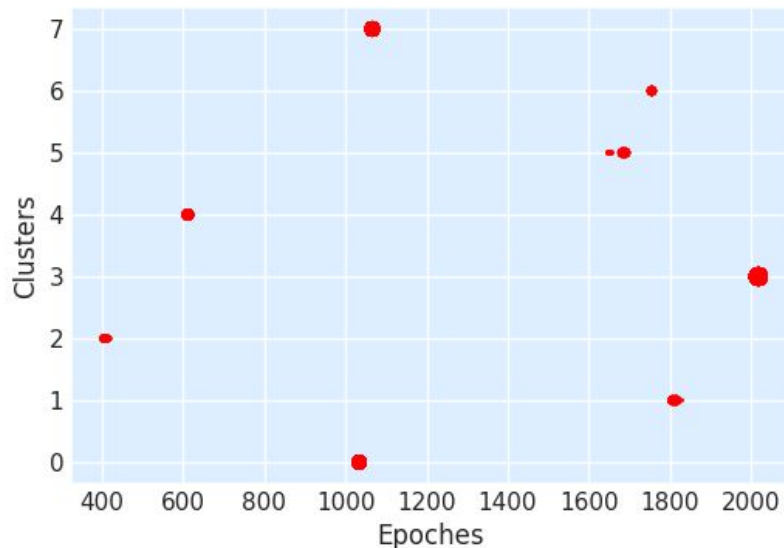
Результат кластеризации для 9 стадий

- границы полученных кластеров (большое количество итераций для разных значений параметров) объединяются в итоговый массив
- массив сортируется и кластеризуется методом k-means

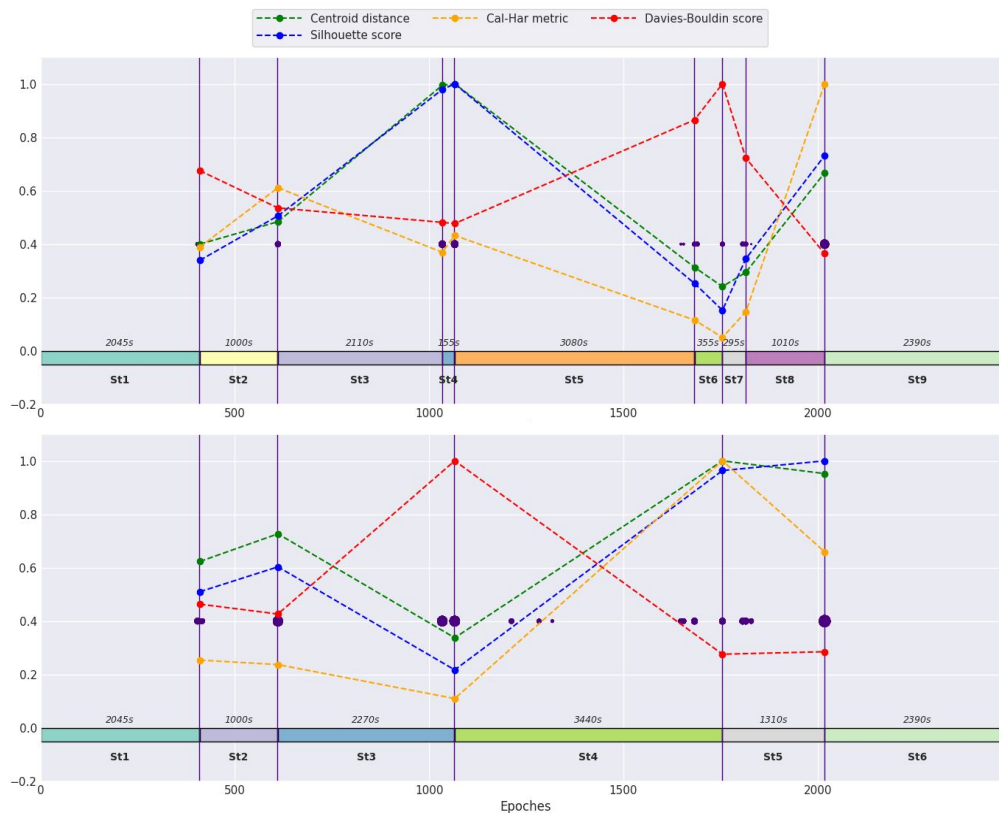
Stage edges merged for all parameters



Stage edge clusters



Итоговая разметка



Точечная диаграмма (на уровне 0.4) представляет собой объединенный массив потенциальных границ стадий.

Пунктирные линии показывают показатели качества кластеризации, рассчитанные для соседних этапов:

Centroid distance – **зеленая** линия

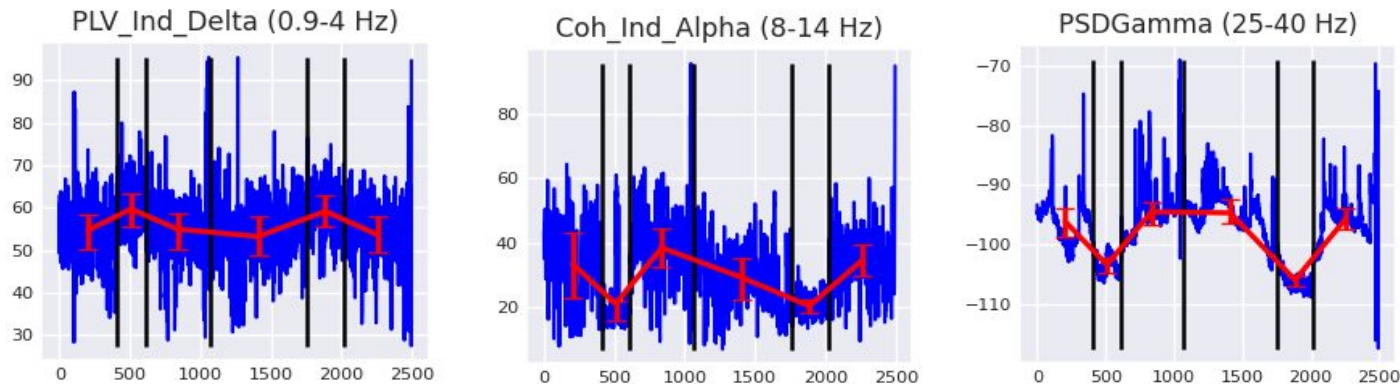
Silhouette – **синяя** линия

Calinski– Harabasz – **оранжевая** линия

Davies–Bouldin – **красная** линия

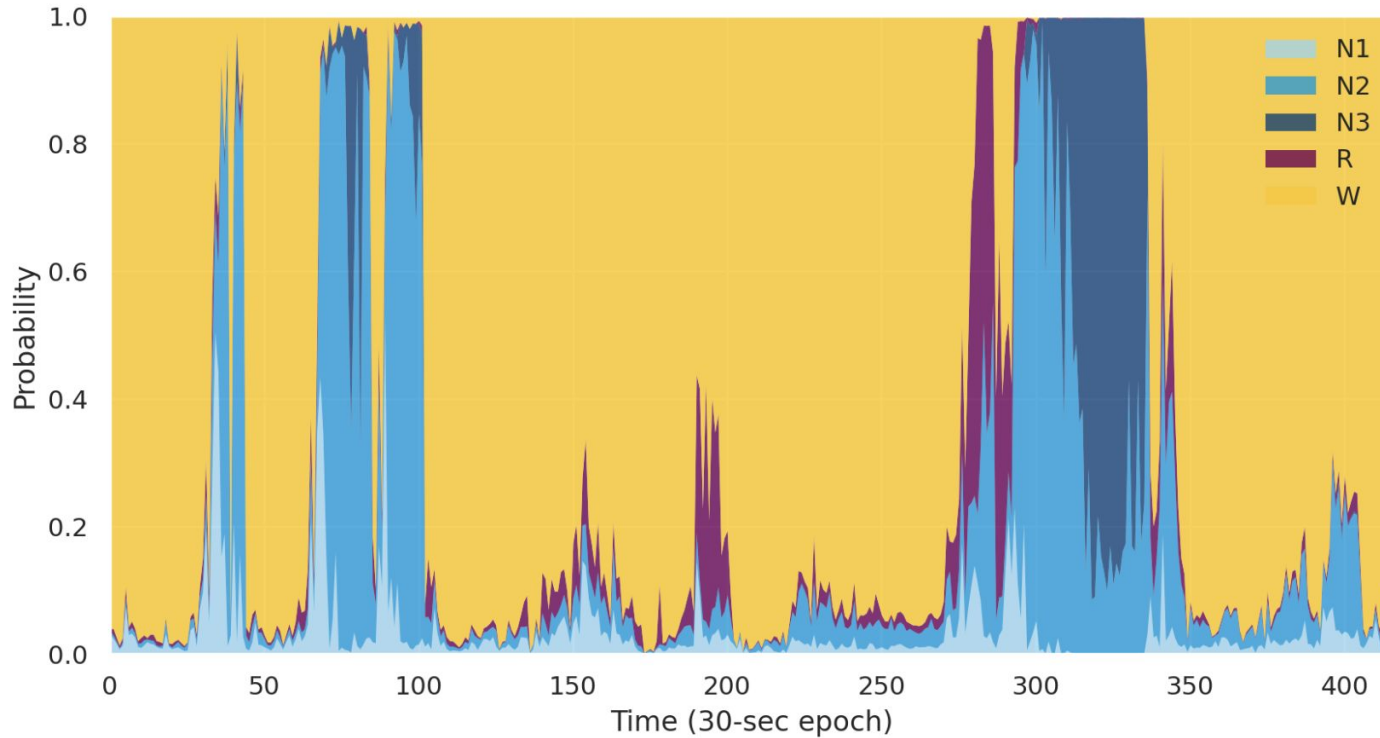


Анализ поведения признаков на выделенных стадиях

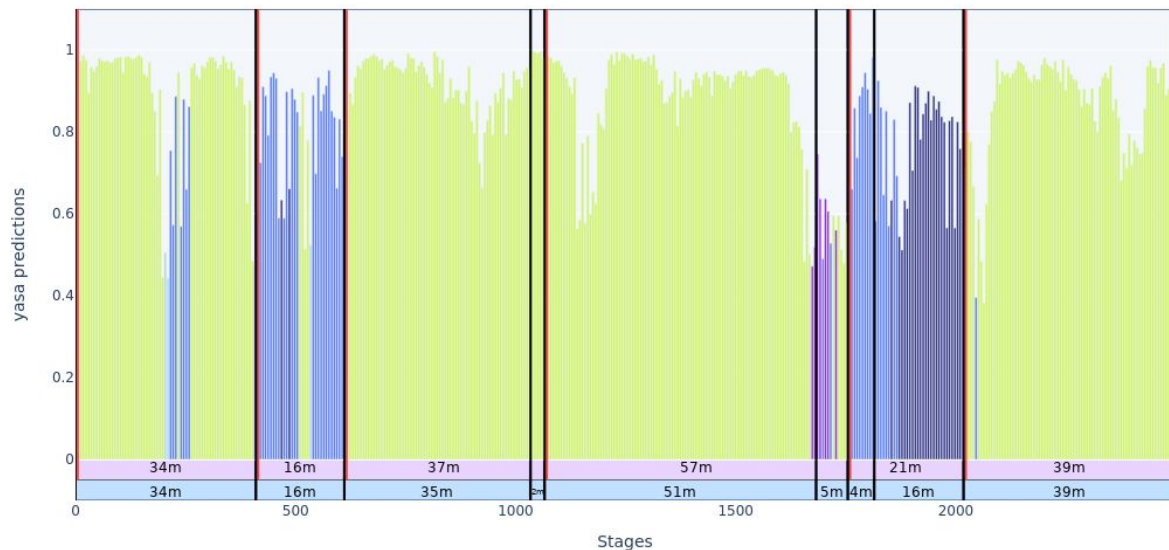


- Горизонтальная ось обозначает индекс эпохи вдоль записи.
- Вертикальные черные линии – это границы стадий.
- Синими линиями обозначена динамика характеристик ЭЭГ на протяжении записи.
- Красные линии представляют медиану \pm межквартильный размах для выбранного объекта на каждом этапе.

Высота “пиков” - вероятность принадлежности эпохи к указанной стадии сна.



Сравнение результатов



Гистограмма – предсказания стадий сна от yasa.

- **Салатовый** цвет – бодрствование
- **Светло-голубой** – засыпание
- **Синий** – медленный сон
- **Темно-синий** – глубокий сон
- **Фиолетовый** – быстрый сон

Вертикальные линии – границы стадий по SDA.

Красные линии и **сиреневая** полоса внизу – разметка 6 стадий, **черные** линии и **голубая** полоса – разметка 9 стадий.

В прямоугольниках внизу указана длительность стадии в минутах.

Выводы



Адаптируемость к задаче определения стадий сна



Высокая разделяющая способность



Самовалидация



Предоставление дополнительной информации о специфике каждой стадии



Перспективы к усовершенствованию

Спасибо за внимание!