

Бинарный классификатор сон / не сон по EEG

Задача: Задача классификации фрагментов (5 секунд) EEG записей с двух каналов (Fpz, Pz) на два состояния: сон или не сон. [SleepDataset](#)

Данные сформированы из открытого датасета: <https://physionet.org/content/sleep-edfx/1.0.0/>. В качестве тестовой части используются записи с телеметрии (ST). В приложенных данных в обучающей части выборки сохранены все фрагменты, включающие в себя как состояние бодрствования, так и различные стадии сна, тестовые данные включают в себя только два состояния: бодрствование и глубокая стадия сна (Sleep stage 4), вы можете как использовать для обучения все стадии, так и отфильтровать данные. Данные представлены в виде массивов .npy со следующими полями: label - категория сна, person - уникальный айди человека, с которого велась запись, time - время начала фрагмента в секундах относительно всей записи данного человека, value - непосредственный массив значений исходного EEG сигнала. Данные дополнительно не обработаны, не отфильтрованы и не очищены.

У вас полная свобода в обработке данных: использование сырого сигнала, преобразование в другие формы. Также нет ограничений на способы классификации/кластеризации, supervised/unsupervised, нейронные сети или классический мл.

Ориентировочный бейзлайн: precision = [0.67650478, 0.85736434], recall = [0.92153996, 0.51696047], labels = ['Sleep stage 4', 'Sleep stage W'].

На что мы обращаем внимание при оценке результата:

1 Код

- структурированный хорошо комментированный и понятный (отделены: предобработка, train/test),
- есть краткое readme: как запустить и какие дополнительные данные использовали

2 Описание результатов:

- Анализ данных
- Выбор метода (Как выбирали?),
- Описание экспериментов
- Анализ результатов и аргументированные предложения по дальнейшему улучшению

Важно: мы оцениваем ваше знания и навыки в ЦОС, ML/DL, программировании и культуре проведения экспериментов. Мы НЕ рекомендуем фокусироваться на больших датасетах и использовании многих GPU. Важнее не показать state-of-the-art результат в плане точности/качества решения, а продемонстрировать решение на разумном датасете и продемонстрировать адекватное решение (при условии заданного датасета и вычислительных мощностей), раскрывающее навыки в постановке задачи обучения, постановке эксперимента и анализе его результатов.

О нас: Информацию о проектах лаборатории искусственного интеллекта CBG AI можно посмотреть в наших видеороликах с конференций например тут: [Видеоролик SPECOM2021](#)