

Оптимизация сложности моделей глубокого обучения в условиях нехватки данных

Черных Владимир Юрьевич
Научный руководитель: Рябенко Е. А.

Московский физико-технический институт
Факультет управления и прикладной математики
Кафедра «Интеллектуальные системы» ВЦ РАН

29 июня 2016 г.

Цели исследования

Построить процедуру выбора оптимальной модели для анализа ЭЭГ в зависимости от конкретных данных

Проблема

- Сейчас структура моделей глубокого обучения выбирается в основном исходя из предыдущих задач или неформализуемого опыта эксперта и лишь косвенно учитывает структуру данных
- Традиционные методы анализа ЭЭГ сложны и многоступенчаты

Решение

Поиск оптимальной структуры нейронной сети

$\mathcal{D} = \{(X_i, \mathbf{y}_i)\}_{i=1}^N = \mathcal{D}_I \cup \mathcal{D}_v$, где $X_i \in \mathcal{X} = \mathbb{R}^n \times \mathbb{R}^c$, $\mathbf{y}_i \in \mathcal{Y} = [0; 1]^k$, c — число каналов, n — длина ряда, k — число классов.

$\mathcal{F}_{\mathbf{h}} = \{f : \mathcal{X} \mapsto \mathcal{Y}\}$ — множество классификаторов фиксированной структуры $\mathbf{h} \in \mathcal{H}$.

Выбор модели внутри класса — минимизация эмпирического риска:

$$\hat{f} = \arg \min_{f \in \mathcal{F}} \frac{1}{|\mathcal{D}_I|} \sum_{\mathcal{D}_I} \mathcal{L}(\mathbf{y}_i, f(X_i)) = \arg \min_{\mathbf{w} \in \mathcal{W}} Q(\mathbf{w}, \mathcal{D}_I, \mathcal{L}).$$

Выбор класса моделей — кросс-валидация:

$$\hat{\mathbf{h}} = \arg \min_{\mathbf{h} \in \mathcal{H}} \frac{1}{q} \sum_{i=1}^q Q(f_{CV}, \mathcal{D}_I^i, \mathcal{L}),$$

$$f_{CV} = \arg \min_{f \in \mathcal{F}_{\mathbf{h}}} Q(f, \mathcal{D}_I \setminus \mathcal{D}_I^i, \mathcal{L}).$$

Традиционный анализ ЭЭГ:

- Частотная фильтрация
- Удаление физиологических артефактов
- Понижение размерности
- Устранение зависимости от субъекта
- Сегментация сигнала
- Классификация

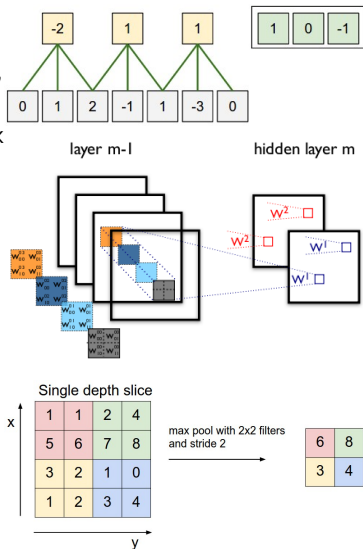
Предлагаемый подход:

- Глубокая сверточная сеть с исходным сигналом на входе
- +
- Подбор оптимальной структуры исходя из количества данных
- Спектральное преобразование сигнала и сверточная нейронная сеть

Сверточные нейронные сети

Различают несколько видов слоев:

- Сверточные — выделение признаков
 - Локальная зона видимости нейрона, т.е. отсутствие полносвязности
 - Имеется набор фильтров — входных весов нейронов
 - Линейная функция активации
- Пулинг-слои — снижение размерности/трансляционная инвариантность признаков
 - Локальная зона видимости нейрона
 - Все входные веса 1
 - Обычно нелинейная функция активации, например максимум
- Полносвязные — классификация



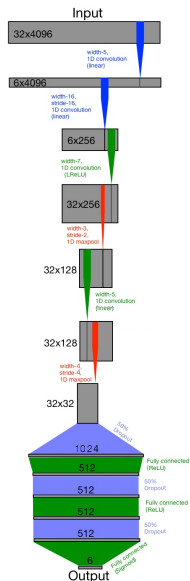
Особенности дизайна для непрерывных данных:

- Отсекаем часть данных с конца для валидации
- Для уменьшения размера датасета отрезаем точки с начала
- Выборка генерируется с помощью окна заданной ширины n

Структура	Непрерывная		Попытками	
Название	Kaggle EEG	BBCI IV-I	Ewan Nurse	BBCI II-IV
Год	2014	2007	2015	2002
Размер	Попыток	~ 250	~ 400	~ 1350
	Точек	~ $1.5 \cdot 10^6$	~ $2 \cdot 10^6$	~ $1350 \cdot 100$
Частота	500	1000	250	100
Каналов	32	59	10 – 62	28
Классов	6	3	3	2

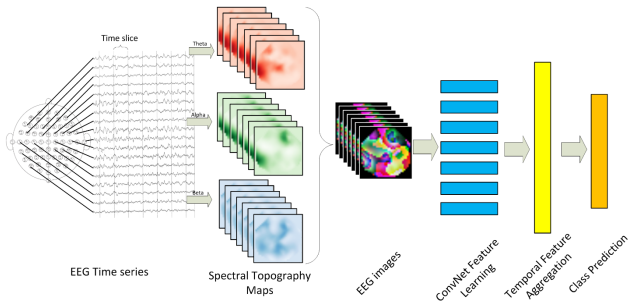
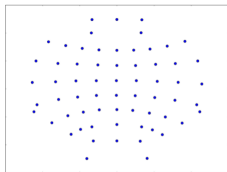
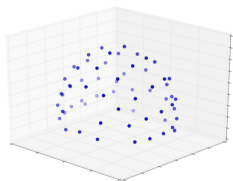
Исходная архитектура:

- Первые два сверточных слоя — понижение размерности по обоим измерениям: признаковому и временному
- Следующие 4 слоя — стандартный стекинг слоев сверточный/пулинг для генерации нового представления
- Завершающие три полносвязных слоя — классификация



Сверточная сеть со спектральным преобразованием

- Переход в спектральную область: $\mathfrak{D}_\omega = \{(\text{FFT}(X_i), \mathbf{y}_i)\}_{i=1}^N$
- Создание трехканального изображения суммированием мощностей в трех (α, β, θ) диапазонах, проектированием и интерполяцией
- Выделение из каждой картинки набора признаков с помощью сверточной сети



Рассматриваемые задачи:

- Выбор оптимальной модели
- Перенос архитектуры модели с больших данных на маленькие

Исходную модель в том же виде нельзя перенести на другие данные из-за проблемы переобучения.

Гипотеза 1

Модель плохо переносится из-за недостаточного количества данных

- Прореживание больших данных до размеров маленькой выборки

Гипотеза 2

Непереносимость обусловлена излишней сложностью модели

- Параллельное упрощение модели

Выбор модели

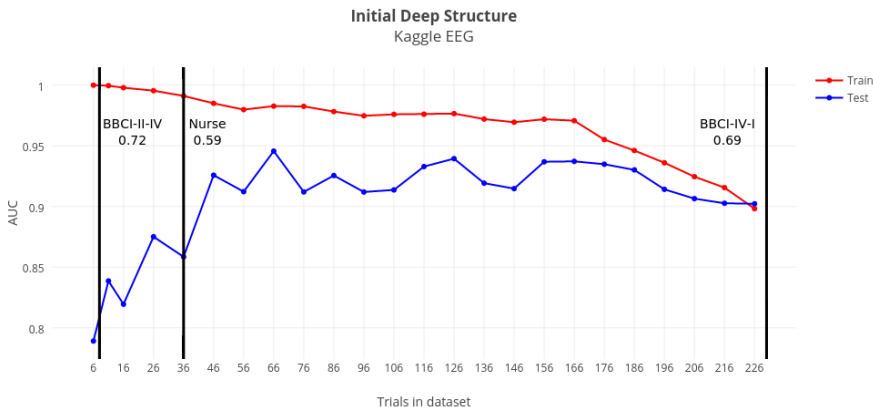
Кросс-валидация. Пространство \mathcal{H} задается с помощью параметров:

- Количество нейронов в полносвязных слоях
- Вероятность Dropout
- Ширина окна n

Перенос архитектуры

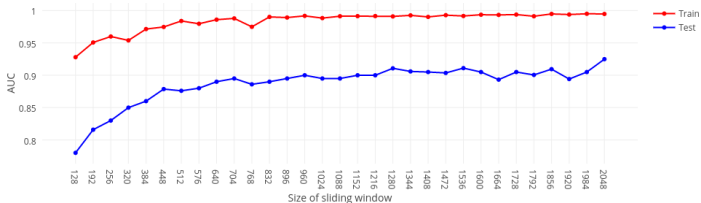
- Берем хорошую, но сложную модель на больших данных
- Уменьшаем объем данных, пока он не сравняется с объемом маленькой выборки
- Выбираем лучшую архитектуру из пространства \mathcal{H}

Гипотеза 1

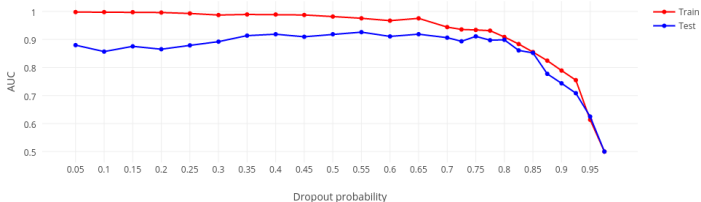


Качество на "своей" выборке с уменьшением объема данных падает, но не так сильно, как при переносе на другие данные.

Window size optimization
Kaggle EEG, 36 trials in dataset



Dropout probability optimization
Kaggle EEG, 36 trials in dataset



Структура	Непрерывная		Попытками		
Название	Kaggle EEG	BBCI IV-I	Ewan Nurse	BBCI II-IV	
AUC	Начальная	0.93	0.69	0.59	0.72
	Простая	0.89	0.66	0.58	0.72
	Спектральная	0.78	0.70	0.52	—
Метрика	AUC	RMSE	Acc	Acc	
Начальная	0.93	5.04	0.45	0.64	
Простая	0.89	4.54	0.44	0.65	
Спектральная	0.78	4.51	0.44	—	
Авторская	0.98	0.42	0.58	0.84	

- AUC — средний по колонкам AUC в случае многих классов
- RMSE — средняя квадратичная ошибка
- Acc — точность

Результаты

Предложены:

- Метод классификации ЭЭГ с подбором локально-оптимальной модели по конкретным данным
- Способ переноса архитектуры сети на меньшие выборки

Показана:

- Плохая переносимость моделей между данными из-за существенных различий в выборках

Направления исследований

- Способ параметризации всего пространства сверточных сетей
- Формализация понятия сложности структуры (аналог VC-теории)
- Применение эвристических алгоритмов выбора модели, а не переборных