

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

15 июня 2016 г.

Цели исследования

Построить процедуру выбора оптимальной модели для анализа данных ЭЭГ в зависимости от их количества.

Проблема

- Сейчас структура моделей глубокого обучения выбирается в основном исходя из экспертных знаний или из предыдущих задач и только интуитивно учитывает размер выборки.
- 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.

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

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:

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

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

Традиционный анализ ЭЭГ:

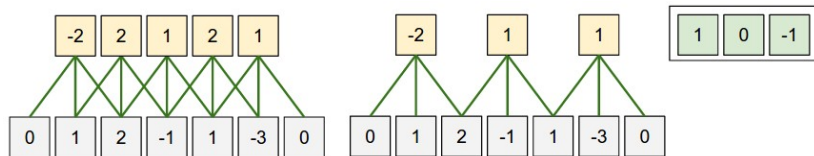
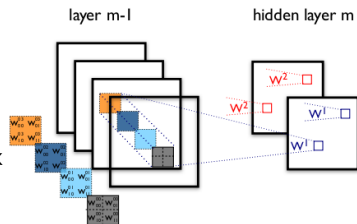
- Частотная фильтрация
- Удаление физиологических артефактов
- Понижение размерности
- Устранение зависимости от субъекта
- Сегментация сигнала
- Классификация

Предлагаемый подход:

- Глубокая сверточная сеть с исходным сигналом на входе
- Подбор оптимальной структуры исходя из количества данных

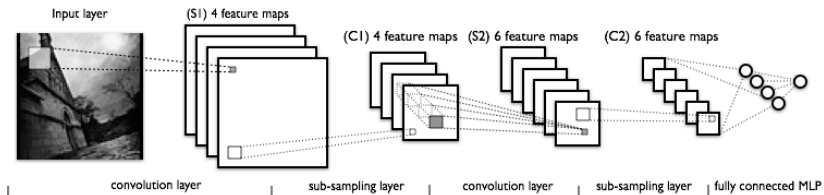
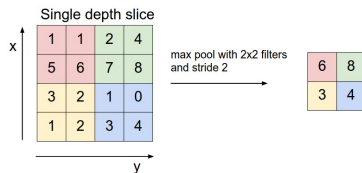
Различают несколько видов слоев:

- Конволюционные — выделение признаков
 - Локальная зона видимости нейрона, т.е. отсутствие полносвязности
 - Имеется набор фильтров — входных весов нейронов
 - Линейная функция активации



Сверточные нейронные сети

- Пулинг-слои — снижение размерности/трансляционная инвариантность признаков
 - Локальная зона видимости нейрона
 - Все входные веса 1
 - Нелинейная функция активации, обычно max
- Полносвязные — классификация



Варьируются следующие параметры:

- Размер обучающей выборки
- Число фильтров и степень сжатия в первом слое
- Ширина окна, т.е. длина временного ряда n
- Количество нейронов и параметр dropout в полносвязных слоях

Основная задача — выбор оптимальной модели. Исследуется также еще одна проблема — перенос типа модели с больших данных на маленькие.

- Берем большие данные и выкидываем куски до уравнивания с маленькой выборкой
- Параллельно упрощаем модель во избежание переобучения

Гипотеза

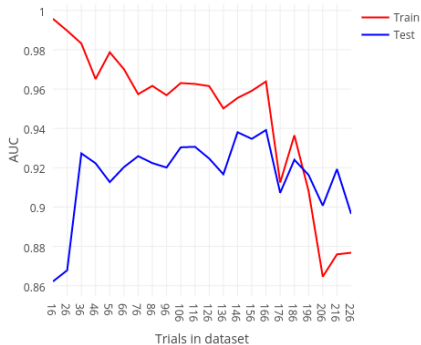
При определенном размере обучающей выборки можно перенести структуру модели на меньшие данные без сильной потери качества.

Особенности дизайна для непрерывных данных:

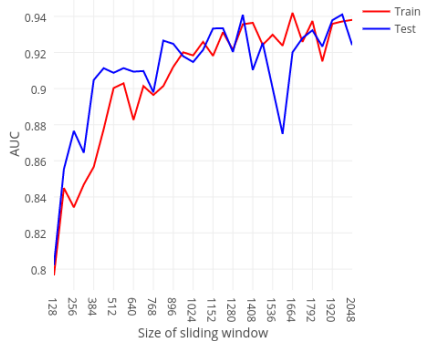
- Отсекаем часть данных с конца для валидации
- Для уменьшения длины отрезаем точки с начала ряда
- Выборку генерируем с помощью окна заданной ширины

| Structure | Continious | | Trials | |
|-----------|------------|-------------------------|-----------------------|--------------|
| Name | Kaggle EEG | BBCI IV-I | Ewan Nurse | BBCI II-IV |
| Year | 2014 | 2007 | 2015 | 2002 |
| Size | Trials | ~ 250 | ~ 400 | ~ 1350 |
| | Points | ~ 1.5 * 10 ⁶ | ~ 2 * 10 ⁶ | ~ 1350 * 100 |
| Frequency | 500 | 1000 | 250 | 100 |
| Channels | 32 | 59 | 10 — 62 | 28 |
| Classes | 6 | 3 | 3 | 2 |

512 size of sliding window,
256 units in mlp, 0.5 dropout probability

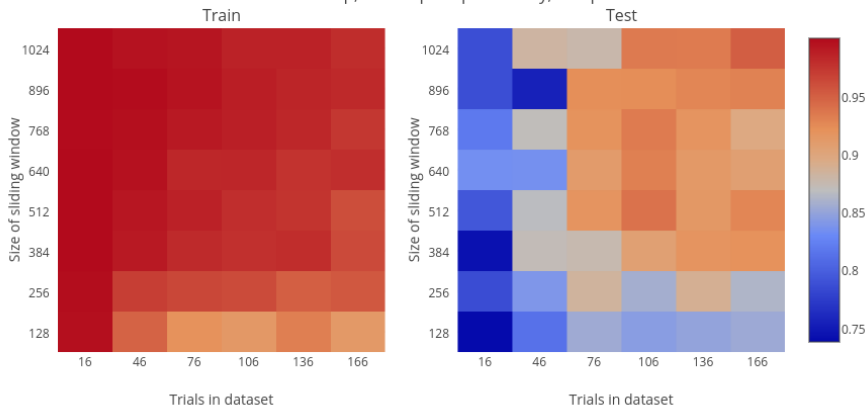


226 trials in dataset,
0.5 dropout probability, 256 units in mlp

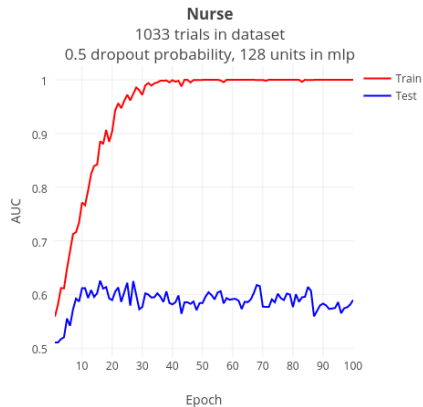
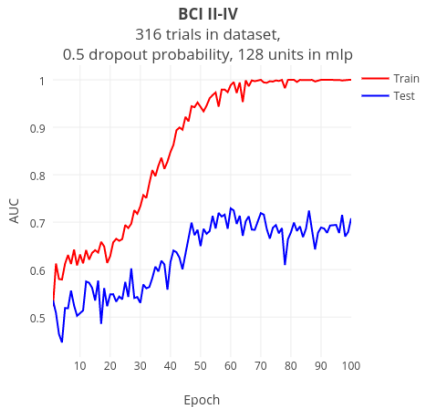


AUC

256 units in mlp, 0.1 dropout probability, 10 epochs



Trial datasets



Результаты

Предложены:

- Алгоритм анализа ЭЭГ с выбором локально-оптимальной модели на основе данных
- Метод переноса типа модели на меньшие выборки

Направления исследований

- Способ параметризации всего пространства сверточных сетей
- Формализация понятия сложности структуры (аналог VC—теории)