

Исследование коннектом и когнетом

Мокров Никита

Московский Физико-Технический Институт

Факультет Радиотехники и Кибернетики

mokrov@frtk.ru

Введение

Постановка задачи

Термин «коннектом» предложен в 2005 году независимо двумя исследователями Олафом Спюрсом и Патриком Хэгмэнном по аналогии с «геномом» (полное описание всех генов) и «протеомом» (полное описание строения и функций всех белков). Сегодня под «коннектомом» понимают полное описание связей в нервной системе того или иного организма.

Как известно, главной задачей нейронаук является понимание того, как из работы материальных и доступных изучению приборами элементов нервной системы получается неуловимая работа психики. Создание сетевых моделей формальных элементов мозга (когнитомы) — всего лишь один из этапов. Наш соотечественник Константин Анохин вводит новый термин «ког» — элемент психического опыта, связанный с работой какого-то участка нейронной сети. Из множества связанных друг с другом когов строится когнитом — сеть психики, внутренний мир животного или человека, которому принадлежит данная нейронная сеть.

С точки зрения математики процесса, мы можем представить человеческий мозг в виде взвешенного графа взаимодействия отдельных его участков. Этот факт дает широкие возможности для большого спектра задач.

Одна из таких - задача машинного обучения. Как по коннектому и когнетому человека, определить определенную болезнь человека на ранней стадии.

Методы решения и новые подходы

На данный момент эта задача решается широким кругом лиц из всех стран мира. Исследование в данном отчете основывается на работе по повышению классификации путем сочетания различных нормировок[1][2].

В данной работе приведены следующие методы нормализации:

- Оригинальная нормализация
- Бинарная нормализация
- Геометрическая нормализация
- Топологическая нормализация

Данный подход смог учесть различные характеристики графа и помог улучшить результат.

Альтернативный подход

В данной статье использовались данные людей здоровых и больных аутизмом (UCLA Multimodal Connectivity Database), которые содержат порядка 100 матриц смежности размера $(264, 264)$ для графов, которые и представляют когнетом пациента.

В связи с этим возникает проблема большой размерности, вследствие чего, большого числа признаков. В данной работе была проверена теория о том, что исходные данные сильно зашумлены. Во-первых, это следует из того, что каждая нервная клетка обновляет силу связи между смежными нейронными, и поэтому мы строим модель для определенного момента времени и состояния человека. Во-вторых, данная теория появилась вследствие того, что исходный графы, а значит и матрицы, были получены путем сшивания послойных участков мозга вручную.

Для решения этой проблемы, воспользуемся спектральным разложением (Или что тоже самое для симметричной матрицы - сингулярное разложение).

$$A = V\Lambda V^{-1}$$

После чего, будет использовать вместо матрицы A , матрицу A'

$$A' = V\Lambda$$

Дальше, убирая k столбцов матрицы V , соответствующие наименьшим собственным значениям, и соответсвенно в матрице Λ сами собственные значения, мы получаем из матрицы размера (N, N) , матрицу размера $(N, N-k)$. И эту новую матрицу будем использовать как целевую переменную.

Результаты

Так для чистых данных, которые никак не нормализуются и не перевзвешиваются, значение по метрике ROC AUC для 100 валидаций выглядит следующим образом.

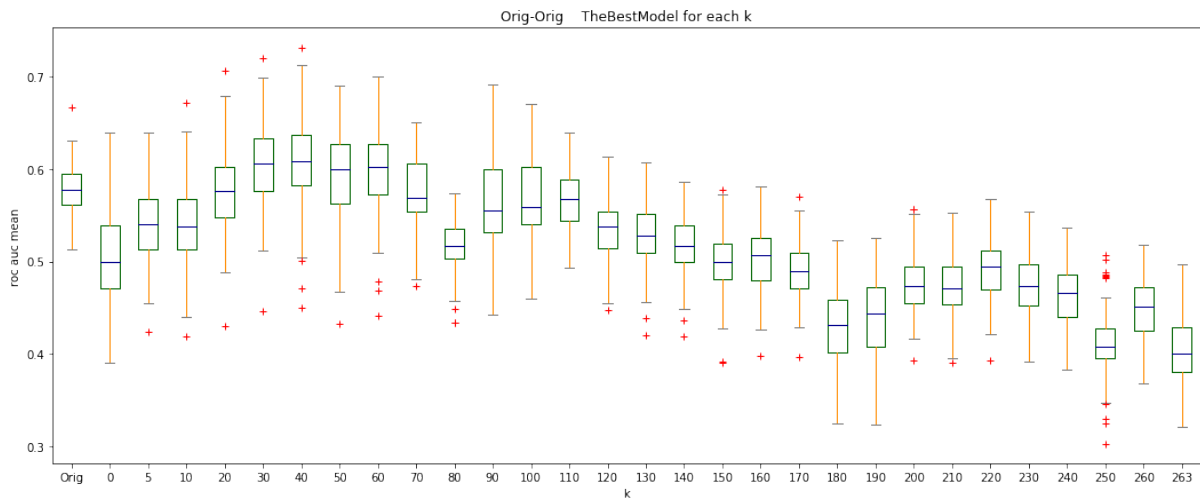


Рис. 1: Boxplot для оригинальной матрицы в зависимости от k

Чаще всего так получается, что лучшее решение заключается в объединении двух или более других. Перед тем как понижать размерность матрицы A , можно по-разному ее нормализовать. Если воспользоваться бинарной нормировкой, то результат смотрится гораздо лучше.

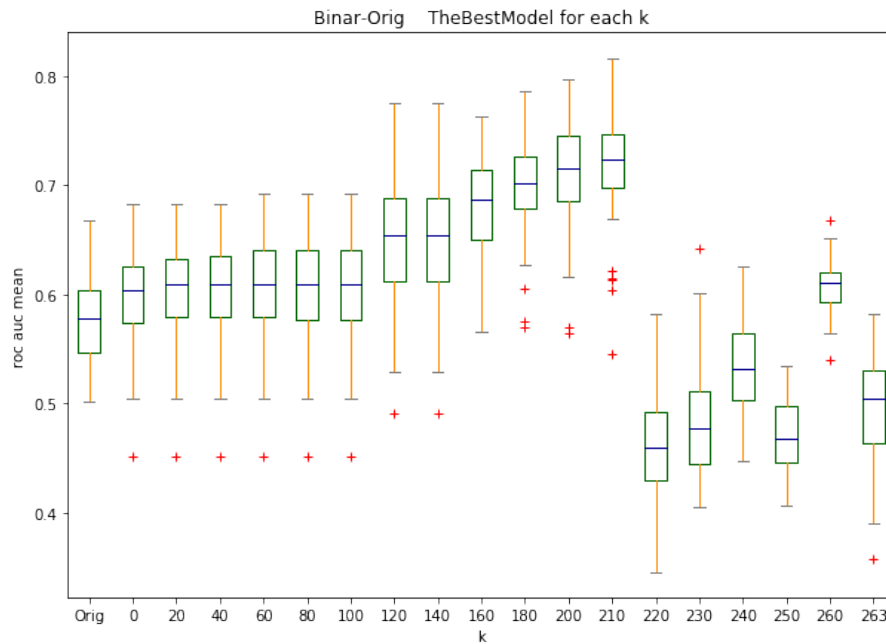


Рис. 2: Boxplot для бинарной матрицы в зависимости от k

Данный метод был опробован на других данных (АРОЕ-4), которые содержат пары (матрица размером (68, 68) и метка класса). Проведем тот же алгоритм на бинарной нормировке и получим такие значения для 100 валидаций.

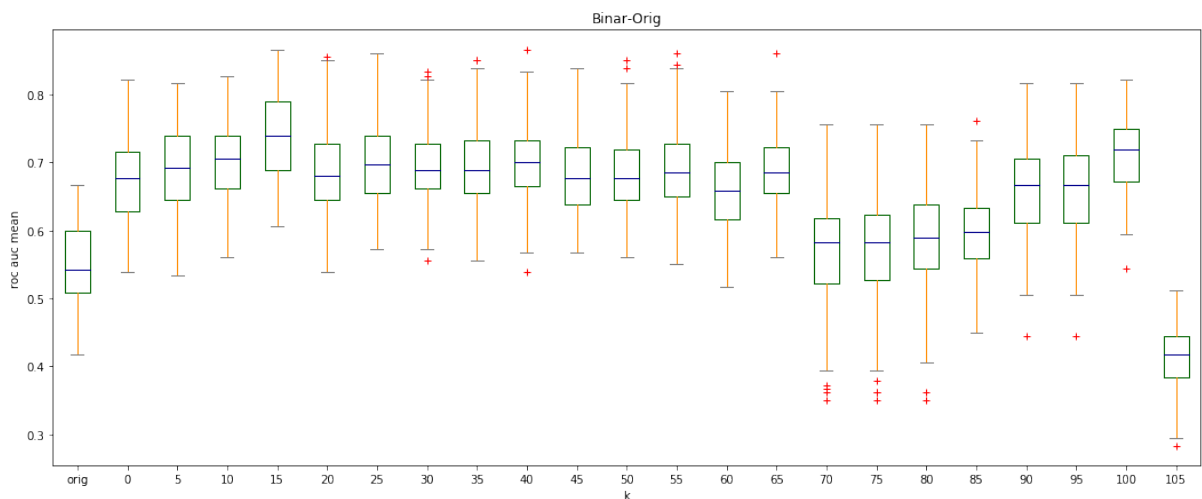


Рис. 3: Boxplot для бинарной матрицы в зависимости от k

Заключение

Как можно заметить, используя только более информативные признаки матрицы (собственные вектора, умноженные на собственные значения), результат может улучшиться для правильной первоначальной нормировки данных.

В заключении хотелось бы отметить, что по ходу исследования возникали новые вопросы, на которые пока нет четких ответов.

Используемая литература

Список литературы

- [1] Boosting Connectome Classification via Combination of Geometric and Topological Normalizations by Dmitry Petrov, Yulia Dodonova, Leonid Zhukov and Mikhail Belyaev
- [2] Machine Learning Application to Human Brain Network Studies: A Kernel Approach by Anvar Kurmukov, Yulia Dodonova and Leonid Zhukov