

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 Спюрнсом 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, 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 определить ту или иную болезнь человека на ранней стадии.

Методы решения и подходы

На данный момент эта задача решается широким кругом лиц из всех стран мира. Исследования ведутся в различных направлениях: применяются как классические методы машинного обучения, так и используются новые методы. В данном отчете исследование основывается на работе по повышению классификации путем сочетания различных нормировок[1].

В данной работе приведены следующие методы нормализации:

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

С помощью данного подхода, можно учесть многие характеристики графа. Основываясь на данном методе были получены существенные результаты[1] для области машинного обучения в нейронауках.

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.

Из-за этого мы сталкиваемся с проблемой большого количества входных признаков. В этой работе была проверена теория¹ 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, обработка данных проводилась вручную, путем сравнения послойных участков.

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'

$$A' = V\Lambda$$

Теперь, вместо того, чтобы использовать все собственные вектора и значения, возьмем только k самых больших собственных значений и векторов, соответствующие им. Таким образом, вместо матрицы размером (N, N) , мы будем использовать матрицу размера $(N, N-k)$.

Результаты

Обучив логистическую регрессию на новой матрице для 100 различных валидаций, мы получим зависимость значения по метрике Roc Auc от параметра k , который указывает на количество выкинутых собственных значений и собственных векторов.

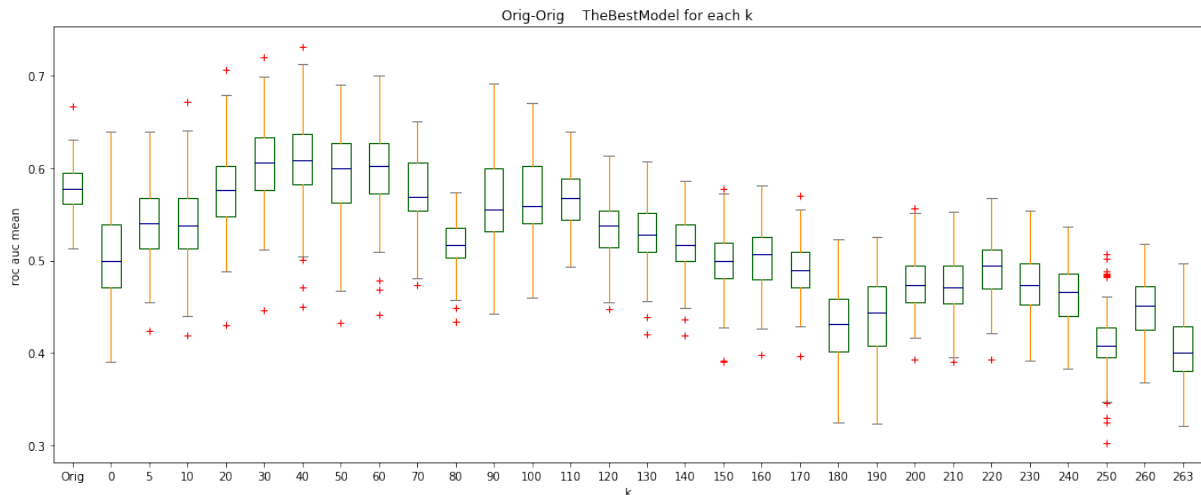


Рис. 1: Boxplot для оригинальной матрицы в зависимости от k

¹ Данная теория принадлежит Максиму Панову

Чаще всего получается так, что лучшее решение является объединением двух или более других методов. Перед тем как понижать размерность матрицы A , можно по-разному ее нормализовать. Если воспользоваться бинарной нормировкой, то результат смотрится гораздо лучше и видна более четкая зависимость.

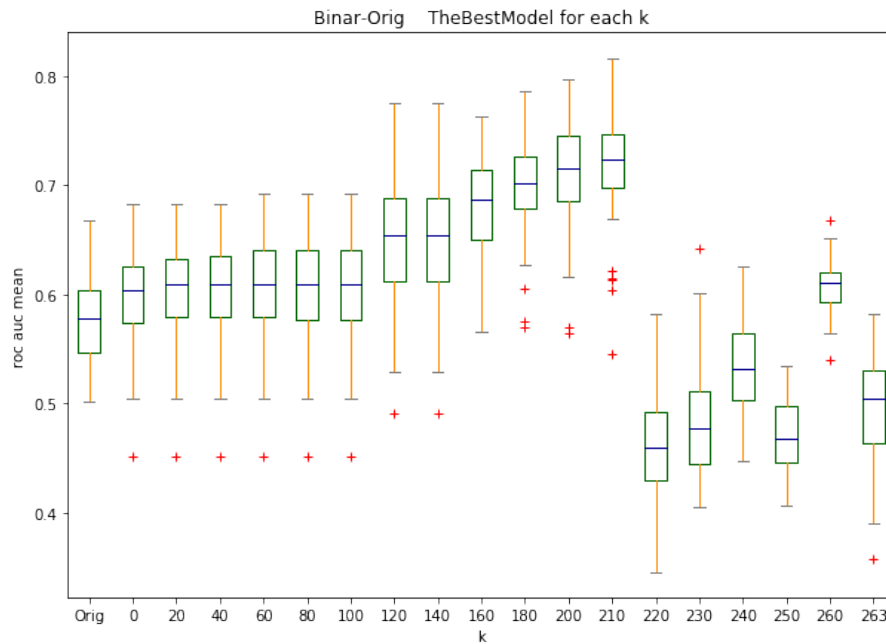


Рис. 2: Boxplot для бинарной матрицы в зависимости от k

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 различных валидаций.

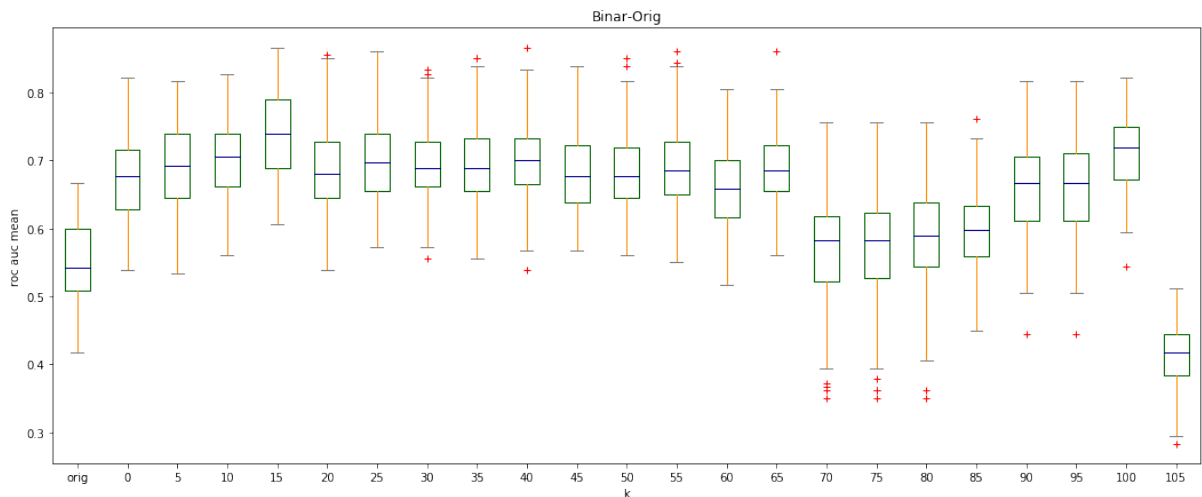


Рис. 3: Boxplot для бинарной матрицы в зависимости от k

Стоит отметить, что как в работе[1], так в работе[2], в которой применялся подход ядра, бинарная нормировка не давала преимущества перед другими. Напротив, в данном исследовании, такой тип предобработки данных был более удачен. С помощью этого метода

удалось увеличить результат для обоих данных, в сравнении с использованием оригинальной матрицы, как целевой переменной.

Данные	Результат для A	Результат для A'
UCLAbaseline	0.58	0.72
APOE-4	0.54	0.73

Заключение

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

Возникает вопрос, почему при использовании оригинальной матрицы результат существенно отличается, если использовать матрицу с бинарной нормировкой? Другими словами, нам не так важно, как сильно взаимодействуют два конкретных участка мозга. А важно нам только наличие или отсутствие этой связи.

При более обширных исследованиях, возникают новые вопросы такого рода, на которых нет четских ответов и правильных рассуждений.

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