

Классификация коннектомов на основе локальных метрик на стохастических матрицах

Иванов Александр, Дмитрий Петров
alexander.radievich@gmail.com, to.dmitry.petrov@gmail.com

Аннотация

Графовые метрики – популярный подход для классификации структурных коннектомов, графов описывающих структурные связи между различными участками мозга. В нашей работе мы предлагаем считать эти метрики на стохастических матрицах случайных блужданий этих графов. При этом часть этих метрик мы предлагаем считать на логарифмах элементов матриц, чтобы сохранить физический смысл вероятностей перехода между вершинами (поскольку вероятности переходов перемножаются, а не складываются). Используя этот прием, мы сгенерировали признаки на основе метрик, использующих расстояние, для задачи классификации нормы и людей с расстройствами аутистического спектра методами машинного обучения. В итоге на новых признаках были получены результаты (площадь под ROC-кривой – 0.75) на уровне ранее опубликованных работ по этой теме.

Данные

Были использованы публично доступные данные по расстройствам аутистического спектра (UCLA). Данные включают матрицы 51 коннектомов с расстройством аутистического спектра и 43 коннектомов нормы. Каждый коннектом представлен матрицей смежности 264 на 264.

Взвешивания

Мы использовали 6 различных схем взвешивания. Во-первых, мы использовали оригинальные веса коннектомов: a_{ij}^{origw} . Во-вторых, мы использовали бинарные веса: a_{ij}^{binar} . В-третьих, взвешивание на квадрат расстояния между вершинами: $a_{ij}^{wbysqdist} = a_{ij}/l_{ij}^2$. В-четвертых, корень от предыдущего взвешенного результата: $a_{ij}^{rootwbydist} = \sqrt{a_{ij}/l_{ij}}$. В-пятых, корень оригинальных весов: $a_{ij}^{sqrtdist} = \sqrt{a_{ij}}$. В-шестых, обратные веса: $a_{ij}^{invdist} = 1/a_{ij}$.

Нормировки

Путем комбинирования 6 методов взвешивания и 2 видов нормировок (отсутствие нормировки и спектральная нормировка) мы получаем 12 вариантов множества коннектомов.

Спектральная нормировка :

$$w_{ij}^{spectral} = \frac{a_{ij}}{\sqrt{d_i d_j}}, \quad (1)$$

где w_{ij} — веса ребер, соединяющих вершины i и j ; d_i — взвешенная степень вершины i .

Общие метрики

Мы делим метрики на два типа. Во-первых, те, которые **считаются на весах графов**.

Входящие степени вершин. $k_i^{in} = \sum_{j \in V} p_{ji}$.

Pagerank. Pagerank приблизительно рассчитывает вероятность, что некто, случайно переходящий по ссылкам в интернете, перейдет к определенной странице (вершине).

Hubs and authorities. Это иной алгоритм оценки важности вершин в направленных графах: Авторитеты оценивают вершины в терминах входящих ребер, и хабы оценивают вершины в плане соединений к другим вершинам.

Аддитивность весов

Второй тип метрик основан на предположении, что **веса графа представляют расстояния между вершинами, которые можно складывать**. Но для матриц случайных блужданий, которые мы предлагаем использовать, это не так. Если мы хотим использовать метрики, основанные на расстоянии, мы должны использовать аддитивные веса. Мы решили эту проблему использованием отрицательных логарифмов весов ребер:

$$w_{ij}^p = -\ln p_{ij}, \quad (2)$$

Метрики на длинах кратчайших путей

На матрице W мы вычисляем матрицу кратчайших путей $D = d_{ij}$, используя алгоритм Дейкстры. Здесь d_{ij}^W — взвешенная длина кратчайшего пути между вершинами i и j , n — количество вершин.

Эффективность e^W и характеристическая длина пути l^W :

$$e_i^W = \frac{\sum_{j \in V, j \neq i} (d_{ij}^W)^{(-1)}}{n-1}, \quad l_i^W = \frac{\sum_{j \in V, j \neq i} d_{ij}^W}{n-1}. \quad (3)$$

Эксцентриситет ecc^W и центральность близости $l^{W(-1)}$:

$$ecc_i^W = \max_{j \in V, j \neq i} d_{ij}^W, \quad l_i^{W(-1)} = \frac{n-1}{\sum_{j \in V, j \neq i} d_{ij}^W}. \quad (4)$$

Промежуточная центральность

$$b_i = \frac{2}{(n-1)(n-2)} \sum_{h, j \in V, h \neq j, h \neq i, j \neq i} \frac{\rho_{hj}(i)}{\rho_{hj}}, \quad (5)$$

где ρ_{hj} — число взвешенных кратчайших путей между h и j , и $\rho_{hj}(i)$ — число взвешенных кратчайших путей между h и j , которые проходят через i .

Результаты

В итоге на данных по расстройствам аутистического спектра были опробованы **785 комбинаций взвешиваний, нормировок, признаков и моделей**. Для каждой комбинации были найдены оптимальные параметры моделей. В качестве моделей были использованы **логистическая регрессия, стохастический градиентный спуск, метод опорных векторов, случайный лес и градиентный бустинг деревьев**.

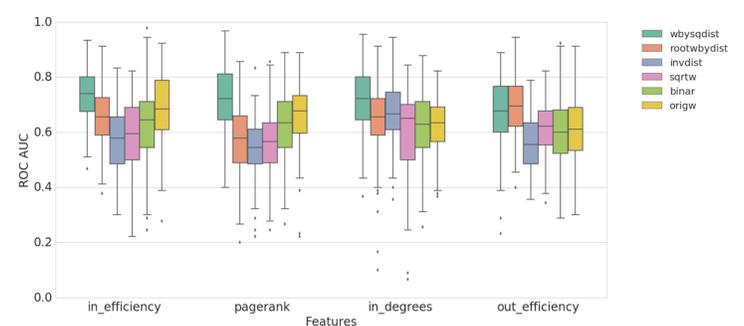


Рис. 1: Результаты для выбранных признаков в зависимости от взвешиваний.

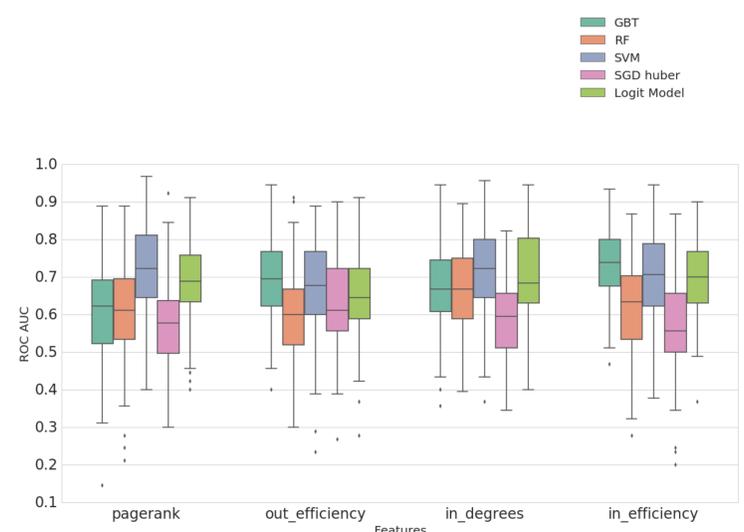


Рис. 2: Результаты для выбранных признаков в зависимости от классификаторов.

Мы выяснили, что признаками, показавшими наилучшую метрику качества, оказались pagerank, in-efficiency и in-degrees. Для этих признаков получены результаты на уровне работ на эти данные в районе 0.75. Таким образом, мы получили интерпретируемые признаки, которые дают результаты не хуже существующих.