

PCA vs ICA in EEG Modeling

Independent Component Analysis

Independent Component Analysis (ICA) — метод восстановления наблюдаемых сигналов на основе их линейных смесей, используя предположение о их независимости и не имея никакой другой априорной информации. Формально, метод извлечения статистически независимых сигналов, генерирующих одновременно наблюдаемые последовательности данных. (Известный пример с вечеринкой и надобностью раскрасить речь кучи говорящих людей.)

Предполагается, что каждый наблюдаемый сигнал — линейная комбинация исходных:

$$x_i(t) = a_{i1}s_1(t) + a_{i2}s_2(t) + \dots + a_{in}s_n(t)$$

$$\mathbf{x}(t) = A\mathbf{s}(t) \forall t \in T$$

$$\mathbf{s}(t) = [s_1(t), \dots, s_n(t)]^T$$

— исходные сигналы, которые надо восстановить. Считаем, что они независимы, и используем, если возможно, информацию об их распределении. Задача можно сформулировать, как вычисление такой разделяющей (separating) матрицы W , что:

$$\mathbf{u}(t) = W\mathbf{x}(t)$$

— есть оценка исходных сигналов с точностью до константы и перестановки.

На практике, ICA находит такую матрицу W , что полученные сигналы статистически независимы настолько, насколько это возможно.

Для задачи ICA имеем плотность функции вероятности (n-мерную):

$$p(s) = \prod_{i=1}^n p_i(s_i(t)), \text{ where } p_i(s_i) \text{ is the density of the } i\text{-th source}$$

В мат статистике две случайные величины X и Y независимы статистически, если и только если:

$$E(X^p Y^q) - E(X^p) E(Y^q) = 0 \forall p, \quad q > 0$$

Это расширение стандартного определения независимости.

В теории информации два сигнала независимы, если их совместная

информация равна 0. Информация p -мерного случайного вектора X :

$$I(X) = \int p(x) \log \frac{p(x)}{\prod_{i=1}^n p_i(x)} dx$$

Отсюда, если источники (исходные сигналы) независимы, их информация в точности равна 0.

На практике, ICA вычисляет разделяющую матрицу W , используя подходы:

- количества информации: минимизируя информацию Wx
- энтропию

Энтропия p -мерного случайного вектора X с плотностью вероятности $p(x)$:

$$H(X) = H(p(x)) = - \int p(x) \log p(x) dx$$

Т.е. чем менее предсказуема случайная величина, тем выше энтропия. Самую большую энтропию имеет гауссовская величина среди всех случайных величин равной дисперсии, иначе говоря, распределение Гаусса самое случайное.

Совместная энтропия двух случайных величин:

$$H(p(x, y)) = - \int \int p(x, y) \log p(x, y) dx dy$$

Ее можно интерпретировать как информацию, которую X и Y могут совместно закодировать, потому что она задает среднюю неопределенность совместного измерения двух переменных.

Тогда, минимизируя совместную информацию, максимизируем информацию, которую вместе величины могут закодировать, те совместную энтропию.

Условная энтропия — измеряет неопределенность в X при условии знаний Y :

$$H(X | Y) = - \int \int p(x, y) \log p(x | y) dx dy$$

Связь энтропии и совместной информации:

$$I(X, Y) = H(X) - H(X | Y)$$

PCA vs ICA

Гауссовский микс исходных сигналов не сепарабелен в алгоритме ICA. В этом случае пользуются PCA.

Есть два основных отличия PCA от ICA:

- PCA использует ортогональные направления, на которые проецирует данные. Это не всегда хорошо. Там пример с выборкой 2Д, представленной в виде параллелограмма. PCA берет базис, параллельный стандартной системе координат, а ICA — векторы с

углом < 90 .

- Упорядочение компонент. В PCA первой компонентой выбирается та, что описывает максимальную дисперсию в данных, следующая компонента — максимальную дисперсию в оставшихся данных и тд. Это можно использовать, чтобы уменьшить размерность задачи. В ICA нужно изначально выбрать число исходных сигналов.

Dimensionality reduction

Мы предполагаем, что наблюдаемые сигналы — линейные комбинации исходных, т.е. главных компонент или независимых источников. Верно и обратное: исходные можно линейно выразить через наблюдаемые. Можно выразить через все найденные, а можно только через часть. Это есть понижение размерности.

Выбор размерности

Часто проще выбрать количество компонент для PCA. Мера того, насколько классно первые p компонент описывают данные, вычисляется как:

$$\phi_n = 1 - \frac{\sum_{i=1}^n \lambda_i}{\sum_{i=1}^p \lambda_i}$$

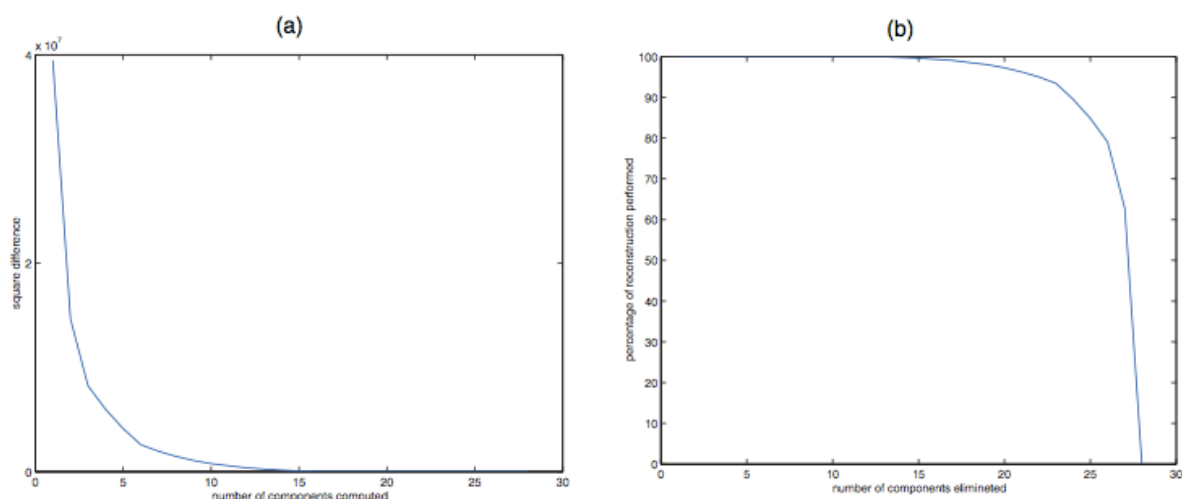
Где лямбды — собственные значения ковариационной матрицы

$$X^T X$$

Для ICA все несколько сложнее. Потому что дисперсия независимых компонент не связана с собственными значениями матрицы ковариации. Однако, когда ICA вычисляет столько же источников, сколько электродов, наблюдаемые сигналы восстанавливаются однозначно. В ином случае — с ошибкой.

Чуваки проводили эксперимент на 720 ЭЭГ с 28 электродами: как изменяется точность с уменьшением количества источников.

Использовали квадрат разности по электродам и времени.



Больше 7 компонент — точность стремится к 0. Больше 80% данных описывается уже 8 компонентами.

Есть классный вывод, что количество компонент, нужных для описания больше 80% данных, больше чем количество, нужное для описания 80% дисперсии. И вообще, PCA работает быстрее ICA, поэтому сначала можно запускать его, находить нужное количество компонент, а потом с этим значением запускать ICA.

Применение в ERP

Тут утверждается, что для анализа ERP систем больше подходит ICA, потому что с его помощью были извлечены нейрофизиологические феномены. Они усредняли значения на электродах и на компонентах. В итоге у главных компонент, только одна была сильно скорректирована с ожидаемым эффектом, у независимых компонент две особенно были классными — обе описывали эффект, но со смещением. Это короче связано с тем, что отклик головного мозга на стимул идет не одновременно по всей площади, а сначала в префронтальной зоне, а потом в сторону темени. И наблюдаемый отклик на стимул как раз примерно содержит сигналы, полученные как независимые компоненты.

– JADE алгоритм — PCA