17021231 – Đỗ Thành Đạt

16020046 – Nguyễn Văn Mạnh

Huấn luyện mô hình HMM

**Xây dựng mô hình HMM:**

Sử dụng hàm GMMHMM của gói thư viện hmmlean để huấn luyện mô hình HMM. Với mỗi từ được chọn, sử dụng hàm GMMHMM với các thông số:

* n\_mix = 4 (Số state trong GMM)
* n\_iter = 1000 (Số lần lặp tối đa của một mô hình)
* params = “mctw” (Cho phép huấn luyện means, covars, transmat, GMM mixing weight)
* init\_params “mct” (Cho phép mô hình tự khởi tạo means, covars, transmat)
* Ngoài ra còn có các thông số riêng cho từng từ:
  + n\_components: Số state của từng từ đã chọn
  + startprob\_: Khởi tạo Ma trận phân phối ban đầu

Khởi tạo là ma trận có dạng 1xN (N là số states):

[ 1., 0., 0., ……, 0.]

* + transmat\_: Khởi tạo Ma trận chuyển trạng thái giữa các states

Khởi tạo là ma trận có dạng NxN (N là số states), là ma trận đơn vị có đường chéo chính là 1 còn lại là 0:

[[ 1., 0., 0., ……, 0.],

[ 0., 1., 0., ……, 0.],

[ 0., 0., 1., ……, 0.],

……

[ 0., 0., 0., ……, 1.]]

**5 từ đã chọn:**

* **Y tế [** i˧˧ te˧˥ **]**

Tham khảo: [*Y tế (Wikitionary)*](https://vi.wiktionary.org/wiki/y_t%E1%BA%BF#Ti%E1%BA%BFng_Vi%E1%BB%87t)

Tham số:

* States: 9
* Kết quả dự đoán:

Trên tập thử nghiệm: 18/20

Thu âm trực tiếp: 16/20

* **Thành phố [** tʰa̤jŋ˨˩ fo˧˥ **]**

Tham khảo: [*Thành phố (Wikitionary)*](https://vi.wiktionary.org/wiki/th%C3%A0nh_ph%E1%BB%91#Ti%E1%BA%BFng_Vi%E1%BB%87t)

Tham số:

* States: 15
* Kết quả dự đoán:

Trên tập thử nghiệm: 17/20

Thu âm trực tiếp: 16/20

* **Nhà [** ɲa̤ː˨˩ **]**

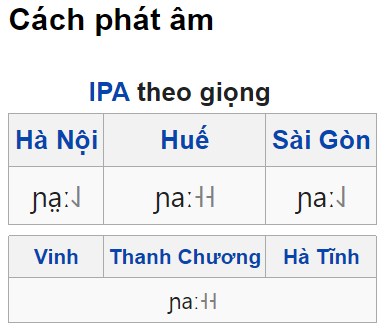
Tham khảo: [*Nhà (Wikitionary)*](https://vi.wiktionary.org/wiki/nh%C3%A0#Ti%E1%BA%BFng_Vi%E1%BB%87t)

Tham số:

* States: 6
* Kết quả dự đoán:

Trên tập thử nghiệm: 20/20

Thu âm trực tiếp: 19/20

**Mẹ [** mɛ̰ʔ˨˩ **]**

Tham khảo: [*Mẹ (Wikitionary)*](https://vi.wiktionary.org/wiki/m%E1%BA%B9#Ti%E1%BA%BFng_Vi%E1%BB%87t)

Tham số:

* States: 6
* Kết quả dự đoán:

Trên tập thử nghiệm: 15/20

Thu âm trực tiếp: 10/20

* **Học [** ha̰ʔwk˨ **]**

Tham khảo: [*Học (Wikitionary)*](https://vi.wiktionary.org/wiki/h%E1%BB%8Dc#Ti%E1%BA%BFng_Vi%E1%BB%87t)

Tham số:

* States: 9
* Kết quả dự đoán:

Trên tập thử nghiệm: 18/20

Thu âm trực tiếp: 18/20