Nguyễn Minh Dương - 17021227

Lê Quốc Khánh - 17021272

UET-VNU | INT3411\_20

Xử lý tiếng nói

Bài tập 2: nhận dạng tiếng nói – huấn luyện mô hình hmm

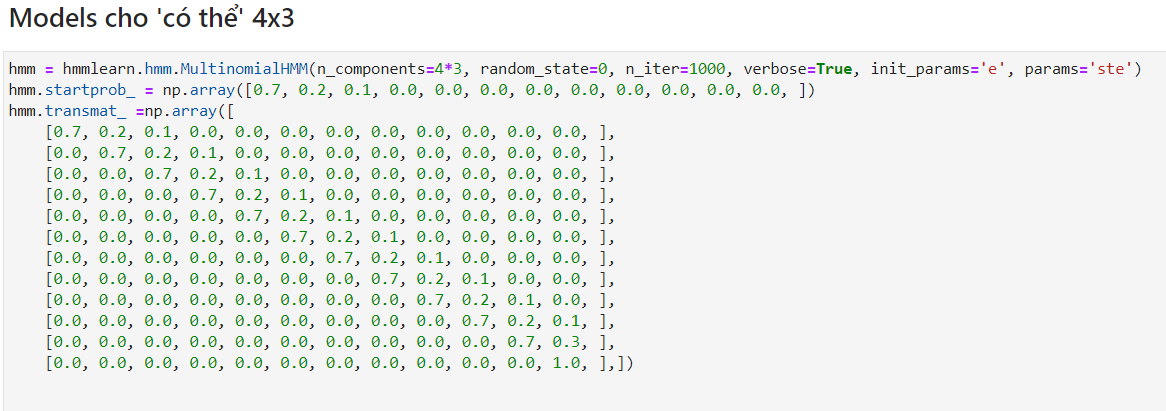
**Dataset**: 100 - 150 file wav cho các từ “có thể”, “chúng ta”, “được”, “người”, “trong”. Chia tập dữ liệu thành tập train và test theo tỉ lệ 80-20.

**Thư viện sử dụng**: MultinomialHMM của hmmlearn.

**Kmeans**: 14 cluster cho kết quả tốt nhất.

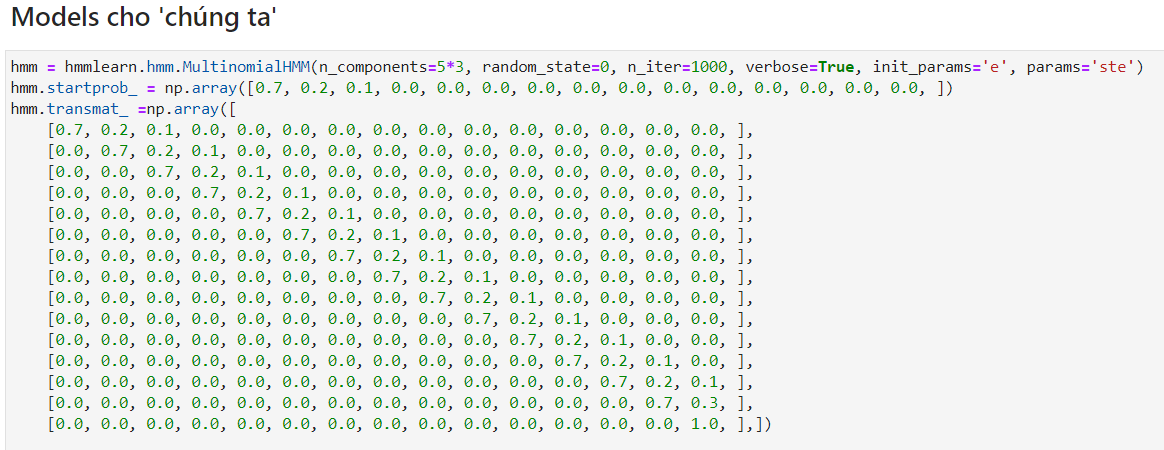
**Các mô hình hmm sử dụng kĩ thuật subdividing phones(3 trạng thái cho mỗi âm vị) và 1 skip connection**:

* “có thể”: có tổng 4 âm vị không tính dấu cách nên sử dụng mô hình có 12 trạng thái:



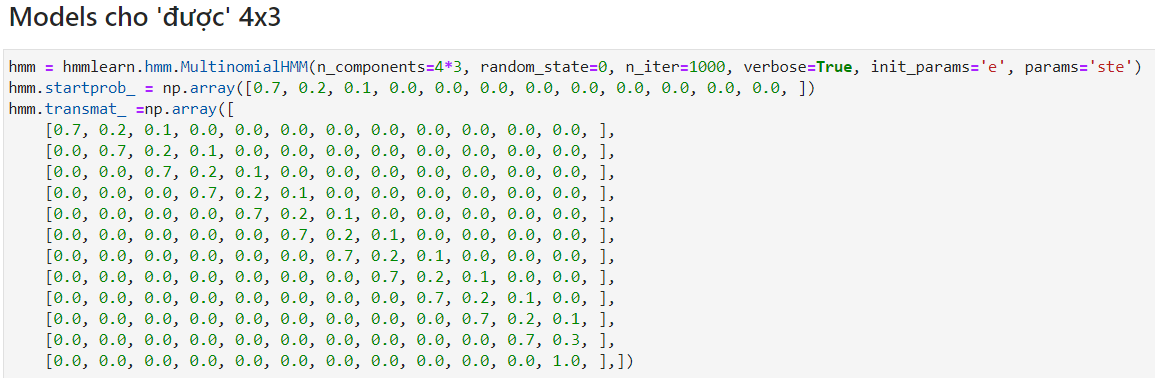
Hình 1: khởi tạo cho hmm "có thể".

* “chúng ta”: có 5 âm vị không trính dấu cách, mô hình có 15 trạng thái:



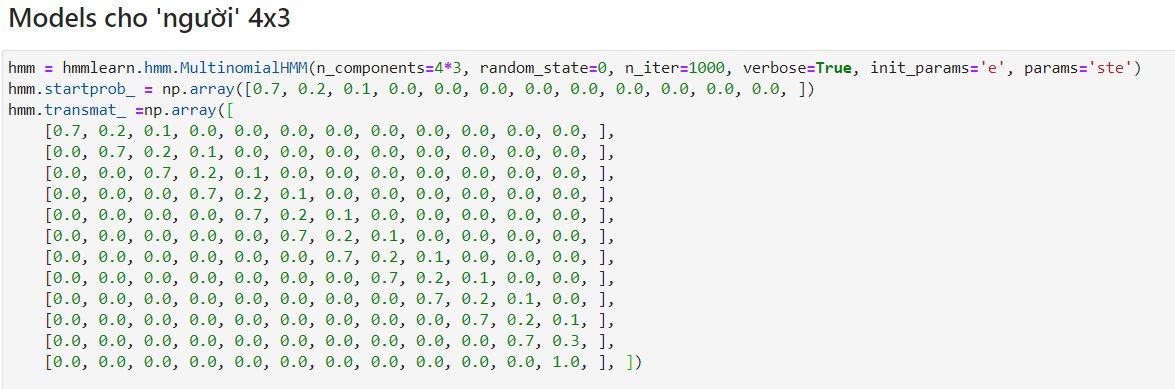
Hình 2: khởi tạo cho hmm "chúng ta".

* “được”: có 4 âm vị, mô hình có 12 trạng thái:



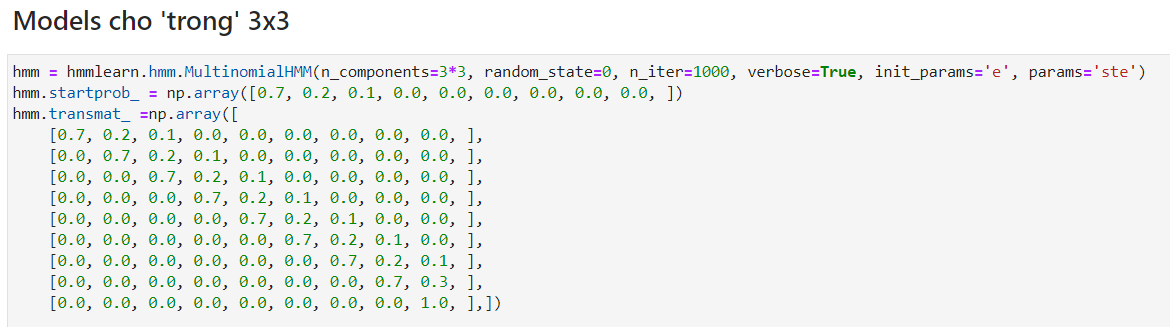
Hình 3: khởi tạo cho hmm "được".

* “người”: có 4 âm vị, mô hình có 12 trạng thái:



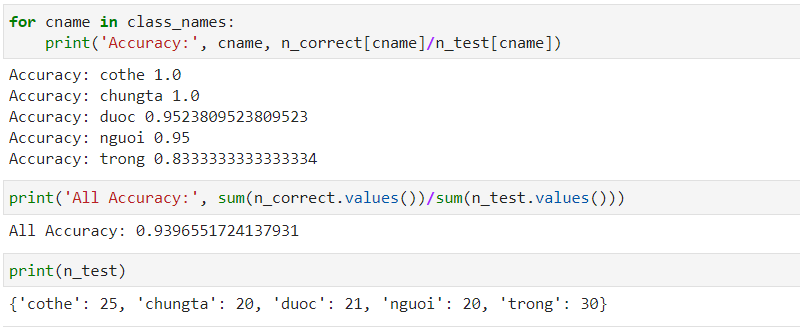
Hình 4: khởi tạo cho hmm "người".

* “trong”: có 3 âm vị, mô hình có 9 trạng thái:



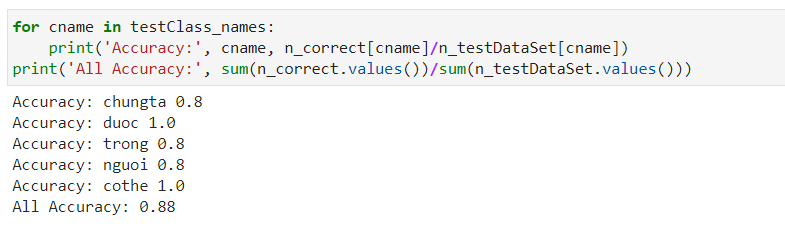
Hình 5: khởi tạo cho hmm "trong".

**Kết quả trên tập test:**



Hình 6: độ chính xác trên tập test.

**Kết quả trên tập các file wav tự phát âm các từ qua micro:**



Hình 7: độ chính xác trên tập tự phát âm qua micro.

**Code & bộ tham số của các mô hình**: <https://github.com/quockhanh21299/LanguageProcessing_MidTerm2/blob/master/hmms_train.ipynb>