BÁO CÁO BÀI TẬP 2: XÂY DỰNG MÔ HÌNH HMM CHO NHẬN DIỆN TIẾNG NÓI

Thành viên nhóm: Phạm Lê Việt Anh - 17021180

Nguyễn Huy Hoàng - 17021191

Nguyễn Tuấn Anh - 15020971

1. Mô tả bài toán

Xây dựng 05 mô hình HMM để nhận dạng 05 từ tiếng Việt, trong đó có ít nhất một từ ghép.

2. Dữ liệu

Dữ liệu đầu vào của mô hình HMM là các file âm thanh (file .wav) chứa các từ đơn hoặc từ ghép được cắt ra từ các file ghi âm câu ở bài tập 1.

Ở bài tập này, nhóm sử dụng 05 từ là: “tôi”, “học”, “nhà”, “nhân viên”, “hà nội”. Đầu tiên, mỗi từ được cắt 100 file làm dữ liệu huấn luyện và thử nghiệm. Hai tập dữ liệu này được chia ngẫu nhiên theo tỷ lệ 80 - 20. Sau khi huấn luyện mô hình xong, một tập dữ liệu thử nghiệm mới được các thành viên nhóm thu âm trực tiếp cũng được sử dụng để đánh giá độ chính xác của mô hình. Số lượng file cụ thể cho tập huấn luyện, tập thử nghiệm và tập thử nghiệm khi thu trực tiếp của mỗi từ như sau:

Từ	Tập huấn luyện	Tập thử nghiệm cắt	Tập thử nghiệm thu trực tiếp
tôi	80	20	40
học	80	20	40
nhà	80	20	40
nhân viên	80	20	40
hà nội	80	20	40

3. Xây dựng mô hình HMM

Trích xuất đặc trưng MFCC

Mỗi file dữ liệu âm thanh được xử lý để lấy 12 đặc trưng mfcc với win_length = 25ms và hop_length = 10ms. Sau đó các đặc trưng mfcc này được chuẩn hóa bằng cách trừ đi giá trị mean của chúng. Sau đó lấy giá trị mfcc sau trừ đi trước để được delta1. Lấy giá trị delta1 sau trừ đi trước để có được delta2. Nối các đặc trưng trên lại, ta có:

12 mfcc + 12 delta1 + 12 delta2 (delta của delta) = 36 đặc trưng

Từ đây, ta nhận được một ma trận X có cỡ T*36 (với T là số frame) để đưa vào huấn luyện mô hình hmm.

Xây dựng mô hình bằng GMMHMM

Để xây dựng mô hình hmm, nhóm sử dụng GMMHMM trong gói hmmlearn. Với mỗi từ, nhóm xây dựng một mô hình hmm từ trái sang phải với các parameter chung:

- n_mix = 4. Thể hiện 4 miền giọng nói khác nhau trong dữ liệu
- random_state = 42.
- n_iter = 1000. Số lần lặp tối đa
- verbose = true.
- params = ‘mctw’. Cho phép huấn luyện m: means; c: covars; t: transmat; w: GMM mixing weights.
- init_params = ‘mct’. Cho phép mô hình tự khởi tạo m, c và t.

Các parameter riêng:

Từ	Theo âm vị	n_components	startprob_	transmat_
tôi	|t||o||i|	9	[1.0,0.0,0.0, 0.0,0.0,0.0,0.0,0.0,0.0,0.0]	[0.7,0.3,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0], [0.0,0.7,0.3,0.0,0.0,0.0,0.0,0.0,0.0,0.0], [0.0,0.0,0.7,0.3,0.0,0.0,0.0,0.0,0.0,0.0], [0.0,0.0,0.0,0.7,0.3,0.0,0.0,0.0,0.0,0.0], [0.0,0.0,0.0,0.0,0.7,0.3,0.0,0.0,0.0,0.0], [0.0,0.0,0.0,0.0,0.0,0.7,0.3,0.0,0.0,0.0], [0.0,0.0,0.0,0.0,0.0,0.0,0.7,0.3,0.0,0.0], [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.7,0.3], [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,1.0],
học	|h||ɔ||k|	9	[1.0,0.0,0.0, 0.0,0.0,0.0,0.0,0.0,0.0,0.0]	[0.7,0.3,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0], [0.0,0.7,0.3,0.0,0.0,0.0,0.0,0.0,0.0,0.0], [0.0,0.0,0.7,0.3,0.0,0.0,0.0,0.0,0.0,0.0], [0.0,0.0,0.0,0.7,0.3,0.0,0.0,0.0,0.0,0.0], [0.0,0.0,0.0,0.0,0.7,0.3,0.0,0.0,0.0,0.0], [0.0,0.0,0.0,0.0,0.0,0.7,0.3,0.0,0.0,0.0], [0.0,0.0,0.0,0.0,0.0,0.0,0.7,0.3,0.0,0.0], [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.7,0.3], [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,1.0],

nhà	|ɲ||a|	6	[1.0,0.0,0.0, 0.0,0.0,0.0]	[0.7,0.3,0.0,0.0,0.0,0.0], [0.0,0.7,0.3,0.0,0.0,0.0], [0.0,0.0,0.7,0.3,0.0,0.0], [0.0,0.0,0.0,0.7,0.3,0.0], [0.0,0.0,0.0,0.0,0.7,0.3], [0.0,0.0,0.0,0.0,0.0,1.0],
nhân viên	|ɲ||ɤ̆||n| |v||ie||n|	12	[1.0,0.0,0.0, 0.0,0.0,0.0,0.0,0.0,0. 0,0.0,0.0, 0.0,0.0]	[0.7,0.3,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0], [0.0,0.7,0.3,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0], [0.0,0.0,0.7,0.3,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0], [0.0,0.0,0.0,0.7,0.3,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0], [0.0,0.0,0.0,0.0,0.7,0.3,0.0,0.0,0.0,0.0,0.0,0.0,0.0], [0.0,0.0,0.0,0.0,0.0,0.7,0.3,0.0,0.0,0.0,0.0,0.0,0.0], [0.0,0.0,0.0,0.0,0.0,0.0,0.7,0.3,0.0,0.0,0.0,0.0,0.0], [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.7,0.3,0.0,0.0,0.0,0.0], [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.7,0.3,0.0,0.0,0.0], [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.7,0.3,0.0,0.0], [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.7,0.3,0.0], [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.7,0.3], [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,1.0],
hà nội	|h||a| |n||o||i|	10	[1.0,0.0,0.0, 0.0,0.0,0.0,0.0,0.0,0. 0,0.0,0.0]	[0.7,0.3,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0], [0.0,0.7,0.3,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0], [0.0,0.0,0.7,0.3,0.0,0.0,0.0,0.0,0.0,0.0,0.0], [0.0,0.0,0.0,0.7,0.3,0.0,0.0,0.0,0.0,0.0,0.0], [0.0,0.0,0.0,0.0,0.7,0.3,0.0,0.0,0.0,0.0,0.0], [0.0,0.0,0.0,0.0,0.0,0.7,0.3,0.0,0.0,0.0,0.0], [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.7,0.3,0.0,0.0], [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.7,0.3,0.0], [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.7,0.3], [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,1.0],

4. Kết quả thử nghiệm

Kết quả thử nghiệm trên bộ dữ liệu cắt từ ở bài tập 1 và dữ liệu ghi trực tiếp như sau:

Từ	Tập thử nghiệm cắt		Tập thử nghiệm thu trực tiếp	
	Correct Predict	Accuracy	Correct Predict	Accuracy
tôi	20/20	1.0	34/40	0.85
học	20/20	1.0	40/40	1.0
nhà	19/20	0.95	27/40	0.675
nhân viên	20/20	1.0	40/40	1.0
hà nội	20/20	1.0	31/40	0.775