

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	60
học	80	20	60
nhà	80	20	60
nhân viên	80	20	60
hà nội	80	20	60

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:

$$12 \text{ mfcc} + 12 \text{ delta1} + 12 \text{ delta2 (delta của delta)} = 36 \text{ đặc trưng}$$

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]	[0.7,0.3,0.0,0.0,0.0,0.0,0.0,0.0,0.0], [0.0,0.7,0.3,0.0,0.0,0.0,0.0,0.0,0.0], [0.0,0.0,0.7,0.3,0.0,0.0,0.0,0.0,0.0], [0.0,0.0,0.0,0.7,0.3,0.0,0.0,0.0,0.0], [0.0,0.0,0.0,0.0,0.7,0.3,0.0,0.0,0.0], [0.0,0.0,0.0,0.0,0.0,0.7,0.3,0.0,0.0], [0.0,0.0,0.0,0.0,0.0,0.0,0.7,0.3,0.0], [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.7,0.3], [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,1.0],
học	h ۆ k	9	[1.0,0.0,0.0, 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], [0.0,0.7,0.3,0.0,0.0,0.0,0.0,0.0,0.0], [0.0,0.0,0.7,0.3,0.0,0.0,0.0,0.0,0.0], [0.0,0.0,0.0,0.7,0.3,0.0,0.0,0.0,0.0], [0.0,0.0,0.0,0.0,0.7,0.3,0.0,0.0,0.0], [0.0,0.0,0.0,0.0,0.0,0.7,0.3,0.0,0.0], [0.0,0.0,0.0,0.0,0.0,0.0,0.7,0.3,0.0], [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.7,0.3], [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 ɣ 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]	[0.7,0.3,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0], [0.0,0.7,0.3,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0], [0.0,0.0,0.7,0.3,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0], [0.0,0.0,0.0,0.7,0.3,0.0,0.0,0.0,0.0,0.0,0.0,0.0], [0.0,0.0,0.0,0.0,0.7,0.3,0.0,0.0,0.0,0.0,0.0,0.0], [0.0,0.0,0.0,0.0,0.0,0.7,0.3,0.0,0.0,0.0,0.0,0.0], [0.0,0.0,0.0,0.0,0.0,0.0,0.7,0.3,0.0,0.0,0.0,0.0], [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.7,0.3,0.0,0.0,0.0], [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.7,0.3,0.0,0.0], [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.7,0.3,0.0], [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.7,0.3], [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,1.0],
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]	[0.7,0.3,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0], [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], [0.0,0.0,0.0,0.0,0.0,0.0,0.7,0.3,0.0,0.0,0.0], [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], [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,1.0],

4. Giảm nhiễu và cắt âm thanh thừa

Nhiễu và âm thanh thừa có thể ảnh hưởng lớn tới độ chính xác của quá trình nhận diện. Để tránh những sai sót này, nhóm sử dụng giảm nhiễu và cắt âm thanh thừa khỏi các file âm thanh.

Để tìm ra các đoạn nhiễu, nhóm sử dụng bộ lọc median của thư viện scipy với cỡ kernel là 3. Sau đó cắt các đoạn âm thanh thừa ở đầu và cuối file để được file chỉ chứa từ cần nhận diện.

5. Kết quả thử nghiệm

Kết quả thử nghiệm trên bộ dữ liệu cắt từ ở bài tập 1 và dữ liệu ghi trực tiếp khi chưa khử nhiễu và cắt âm thanh thừa như sau:

Từ	Tập thử nghiệm cắt		Tập thử nghiệm thu trực tiếp	
	Correct Predict	Accuracy	Correct Predict	Accuracy
tôi	20/20	1.0	43/60	0.716
học	20/20	1.0	60/60	1.0

nhà	19/20	0.95	51/60	0.85
nhân viên	20/20	1.0	60/60	1.0
hà nội	20/20	1.0	46/60	0.76

Sau khi đã khử nhiễu và cắt âm thanh thừa:

Từ	Tập thử nghiệm cắt		Tập thử nghiệm thu trực tiếp	
	Correct Predict	Accuracy	Correct Predict	Accuracy
tôi	20/20	1.0	58/60	0.966
học	20/20	1.0	60/60	1.0
nhà	19/20	0.95	59/60	0.983
nhân viên	20/20	1.0	60/60	1.0
hà nội	20/20	1.0	60/60	1.0

6. Đóng góp của các thành viên

Nhìn chung các thành viên trong nhóm đều có ý thức tham gia đóng góp cho bài tập.

Đóng góp cụ thể của từng thành viên như sau:

Công việc	Cắt dữ liệu	Ghi âm dữ liệu test	Làm app	Hiệu chỉnh mô hình	Khử nhiễu và cắt âm thừa	Viết báo cáo
Phạm Lê Việt Anh	✓	✓	✓	✓	✓	✓
Nguyễn Huy Hoàng	✓	✓		✓		✓
Nguyễn Tuấn Anh		✓			✓	✓