Thành viên:

Nguyễn Duy Quang– 16020272

Nguyễn Quốc Cường- 17021213

Báo cáo bài tập môn xử lý tiếng nói

Dataset: ~100 file wav cho các từ “không" ”, “chúng ta”, “người”, “tôi,"cách ly ”.

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:

- “*Không*”: có tổng 3 âm vị không tính dấu cách nên sử dụng mô hình có 9 trạng thái:

```
hmm.startprob_ = np.array([0.7, 0.2, 0.1, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ])
hmm.transmat_ =np.array([
    [0.7, 0.2, 0.1, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ],
    [0.0, 0.7, 0.2, 0.1, 0.0, 0.0, 0.0, 0.0, 0.0, ],
    [0.0, 0.0, 0.7, 0.2, 0.1, 0.0, 0.0, 0.0, 0.0, ],
    [0.0, 0.0, 0.0, 0.7, 0.2, 0.1, 0.0, 0.0, 0.0, ],
    [0.0, 0.0, 0.0, 0.0, 0.7, 0.2, 0.1, 0.0, 0.0, ],
    [0.0, 0.0, 0.0, 0.0, 0.0, 0.7, 0.2, 0.1, 0.0, ],
    [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.7, 0.2, 0.1, ],
    [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, ],])
```

Figure 1: khởi tạo cho hmm "Không".

- “*Tôi*”: có 3 âm vị không tính dấu cách, mô hình có 9 trạng thái:

```
hmm.startprob_ = np.array([0.6,0.3,0.1,0.0,0.0,0.0,0.0,0.0,0.0])
hmm.transmat_ = np.array([
        [0.5,0.3,0.2,0.0,0.0,0.0,0.0,0.0,0.0],
        [0.0,0.5,0.3,0.2,0.0,0.0,0.0,0.0,0.0],
        [0.0,0.0,0.5,0.3,0.2,0.0,0.0,0.0,0.0],
        [0.0,0.0,0.0,0.5,0.3,0.2,0.0,0.0,0.0],
        [0.0,0.0,0.0,0.0,0.5,0.3,0.2,0.0,0.0],
        [0.0,0.0,0.0,0.0,0.0,0.5,0.3,0.2,0.0],
        [0.0,0.0,0.0,0.0,0.0,0.0,0.5,0.3,0.2],
        [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.6,0.4],
        [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,1.0],
    ])
```

Figure 2: khởi tạo cho hmm "tôi".

- “*Chúng ta*”: có 5 âm vị không tính dấu cách, mô hình có 15 trạng thái:

```
hmm.startprob_ = np.array([0.7, 0.2, 0.1, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ])
hmm.transmat_ =np.array([
    [0.7, 0.2, 0.1, 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, 0.7, 0.2, 0.1, 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, 0.7, 0.2, 0.1, 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, 0.7, 0.2, 0.1, 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, 0.7, 0.2, 0.1, 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, 0.7, 0.2, 0.1, 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, 0.7, 0.2, 0.1, 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, 0.7, 0.2, 0.1, 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, 0.7, 0.2, 0.1, 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, 0.7, 0.2, 0.1, 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, 0.7, 0.2, 0.1, 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, 0.7, 0.2, 0.1, 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, 0.7, 0.2, 0.1, ],
    [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, 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, 0.0, 0.0, 1.0, ],])
```

Figure 3: khởi tạo cho hmm "chúng ta".

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

```
hmm.startprob_ = np.array([0.7, 0.2, 0.1, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ])
hmm.transmat_ =np.array([
    [0.7, 0.2, 0.1, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ],
    [0.0, 0.7, 0.2, 0.1, 0.0, 0.0, 0.0, 0.0, 0.0, ],
    [0.0, 0.0, 0.7, 0.2, 0.1, 0.0, 0.0, 0.0, 0.0, ],
    [0.0, 0.0, 0.0, 0.7, 0.2, 0.1, 0.0, 0.0, 0.0, ],
    [0.0, 0.0, 0.0, 0.0, 0.7, 0.2, 0.1, 0.0, 0.0, ],
    [0.0, 0.0, 0.0, 0.0, 0.0, 0.7, 0.2, 0.1, 0.0, ],
    [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.7, 0.2, 0.1, ],
    [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, ],])
```

Figure 4: khởi tạo cho hmm "người".

- “*Cách ly*”: có 5 âm vị không tính dấu cách, mô hình có 15 trạng thái:

```
hmm.startprob_ = np.array([0.7,0.2,0.1,0.0,0.0,0.0, 0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0])
hmm.transmat_ =np.array([
    [0.5,0.3,0.2,0.0,0.0,0.0, 0.0, 0.0, 0.0,0.0, 0.0, 0.0,0.0, 0.0,0.0],
        [0.0,0.5,0.3,0.2,0.0,0.0, 0.0, 0.0, 0.0,0.0, 0.0, 0.0,0.0, 0.0,0.0],
        [0.0,0.0,0.5,0.3,0.2,0.0, 0.0, 0.0, 0.0,0.0, 0.0, 0.0,0.0, 0.0,0.0],
        [0.0,0.0,0.0,0.5,0.3,0.2,0.0, 0.0, 0.0,0.0, 0.0, 0.0,0.0, 0.0,0.0],
        [0.0,0.0,0.0,0.0,0.5,0.3,0.2,0.0,0.0,0.0, 0.0, 0.0,0.0, 0.0,0.0],
        [0.0,0.0,0.0,0.0,0.0,0.5,0.3,0.2,0.0,0.0, 0.0, 0.0,0.0, 0.0,0.0],
        [0.0,0.0,0.0,0.0,0.0,0.0,0.5,0.3,0.2,0.0,0.0, 0.0,0.0, 0.0,0.0],
        [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.5,0.3,0.2,0.0,0.0,0.0, 0.0,0.0],
        [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.5,0.3,0.2,0.0,0.0, 0.0,0.0],
        [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.5,0.3,0.2,0.0, 0.0,0.0],
        [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.5,0.3,0.2,0.0,0.0],
        [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.5,0.3,0.2,0.0],
        [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.5,0.3,0.2],
        [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.5,0.5],
        [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,1.0],])
```

Figure 5: khởi tạo cho hmm "Cách ly".

Kết quả trên tập test:

```
Accuracy: khong 0.8666666666666667
Accuracy: tôi 0.6842105263157895
Accuracy: cachly 0.9565217391304348
Accuracy: chungta 0.9333333333333333
Accuracy: nguoi 0.90625
```

Kết quả trên tập thu:

```
Accuracy: khong 0.8387096774193549
Accuracy: tôi 0.7
Accuracy: cachly 1.0
Accuracy: chungta 1.0
Accuracy: nguoi 0.8214285714285714
```

Code & bộ tham số của các mô hình:

https://github.com/quoccuong205/XLTN