**Thành viên:**

Nguyễn Thành Nam – 17021305

Nguyễn Trung Hiếu – 17021246

**Báo cáo bài tập môn xử lý tiếng nói**

**Dataset**: ~100 file wav cho các từ “bệnh nhân”, “chúng ta”, “có thể”, “người”, “Việt Nam”. Chia tập dữ liệu thành tập train và test theo tỉ lệ 70-30.

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

**Kmeans**: sau khi thử số cluster trong khoảng [10,20] thì nhận thấy 14 cluster cho kết quả tốt nhất trên tập test.

**Các mô hình hmm sử dụng kĩ thuật subdividing phones(3 trạng thái cho mỗi âm vị) và 1 skip connection. Khởi tạo ai,I = 0.7, ai+1,I = 0.2, ai+2,I = 0.1**:

* “bệnh nhân” /b-ệ-nh nh-â-n/: có tổng 6 âm vị không tính dấu cách nên sử dụng mô hình có 18 trạng thái.
* “chúng ta” /ch-ú-ng t-a/: có 5 âm vị không trính dấu cách, mô hình có 15 trạng thái.
* “có thể” /c-ó th-ể/: có 4 âm vị không tính dấy cách, mô hình có 12 trạng thái.
* “người” /ng-ười-i/: có 3 âm vị, mô hình có 9 trạng thái.
* “Việt Nam” /v-iệ-t n-a-m/: có 6 âm vị không tính dấu cách, mô hình có 18 trạng thái.

**Kết quả trên tập test:**

|  |  |  |
| --- | --- | --- |
| Từ | Số lượng file trong tập test | Accuracy |
| Bệnh nhân | 30 | 90% |
| Chúng ta | 30 | 93.33% |
| Có thể | 29 | 96.55% |
| Người | 32 | 84.37% |
| Việt Nam | 30 | 96.67% |
| Tổng | | 92.05% |

Table : kết quả dự đoán của mô hình hmm trên tập test.

**Kết quả khi thu âm trực tiếp:**

**Code & bộ tham số của các mô hình**: https://github.com/not-nam-or-am-i/voice-processing-hmm/blob/master/hmms.ipynb