*Họ và tên: Ngô Quang Minh   
Lớp: 64TTNT2  
MSV: 2251262618*

**ỨNG DỤNG HMM VÀ KALMAN TRONG PHÂN TÍCH CHO BÀI TOÁN NHẬN DẠNG TỪ VỰNG QUA TIẾNG NÓI**

**1. Mô tả công việc đã làm**

Trong báo cáo này, em đã thực hiện việc so sánh hiệu suất giữa hai mô hình nhận dạng từ vựng: Hidden Markov Model (HMM) đơn thuần và HMM kết hợp với bộ lọc Kalman. Cụ thể:

* Sử dụng bộ dữ liệu Free Spoken Digit Dataset (FSDD) - một bộ dữ liệu âm thanh gồm các bản ghi âm của các chữ số nói (0-9) ở định dạng WAV 8kHz
* Trích xuất đặc trưng MFCC (Mel-Frequency Cepstral Coefficients) từ các file âm thanh
* Xây dựng hai mô hình: HMM đơn thuần và HMM kết hợp với bộ lọc Kalman
* So sánh hiệu suất của hai mô hình trong việc nhận dạng chữ số nói

Việc này giúp kiểm chứng hiệu quả của bộ lọc Kalman trong việc xử lý đặc trưng MFCC trước khi đưa vào mô hình HMM, từ đó đánh giá liệu việc kết hợp Kalman có cải thiện hiệu suất nhận dạng hay không

**2. Mô tả quá trình huấn luyện**

**Chuẩn bị dữ liệu**

* Clone bộ dữ liệu FSDD từ GitHub
* Cài đặt các thư viện cần thiết: python\_speech\_features, hmmlearn, filterpy
* Trích xuất đặc trưng MFCC từ mỗi file âm thanh để sử dụng cho việc huấn luyện
* Chia dữ liệu thành tập huấn luyện (70%) và tập kiểm thử (30%)

**Xây dựng và huấn luyện mô hình**

1. **Mô hình HMM đơn thuần**:
   * Tạo một mô hình GMMHMM riêng cho mỗi chữ số từ 0 đến 9
   * Sử dụng 5 trạng thái ẩn (n\_components=5) và 3 thành phần hỗn hợp Gaussian (n\_mix=3)
   * Huấn luyện mỗi mô hình trên các đặc trưng MFCC của từng chữ số tương ứng
2. **Mô hình HMM+Kalman**:
   * Thiết lập bộ lọc Kalman để lọc đặc trưng MFCC
   * Cấu hình bộ lọc với ma trận chuyển trạng thái (F), ma trận quan sát (H), ma trận nhiễu đo lường (R) và ma trận nhiễu hệ thống (Q)
   * Lọc đặc trưng MFCC qua bộ lọc Kalman trước khi đưa vào huấn luyện HMM
   * Sử dụng cùng cấu hình HMM như mô hình đơn thuần (5 trạng thái, 3 thành phần hỗn hợp)

**Đánh giá**

* Dự đoán nhãn cho tập kiểm thử bằng cả hai mô hình
* Tính toán độ chính xác và so sánh hiệu suất giữa hai phương pháp
* Vẽ biểu đồ trực quan hóa kết quả so sánh

4. Mô tả kết quả đạt được

Sau khi huấn luyện và đánh giá hai mô hình trên bộ dữ liệu FSDD, tôi đạt được kết quả như sau:

* **Độ chính xác của HMM đơn thuần**: 93.33%
* **Độ chính xác của HMM+Kalman**: 92.67%
* **Sự thay đổi**: -0.67% (giảm nhẹ)

Kết quả cho thấy, trong trường hợp cụ thể này, việc sử dụng bộ lọc Kalman không cải thiện hiệu suất của mô hình HMM, mà ngược lại còn làm giảm nhẹ độ chính xác. Điều này có thể do:

1. Bộ dữ liệu FSDD là bộ dữ liệu tương đối sạch, với ít nhiễu nền, nên việc sử dụng bộ lọc Kalman không mang lại nhiều lợi ích.
2. Các tham số của bộ lọc Kalman (ma trận Q, R) có thể chưa được tối ưu hóa cho bài toán cụ thể này.
3. Mô hình HMM đơn thuần đã đủ mạnh để mô hình hóa dữ liệu âm thanh trong bài toán nhận dạng chữ số đơn giản.

*Link github:*