



HUST

ĐẠI HỌC BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

ONE LOVE. ONE FUTURE.



ĐẠI HỌC
BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY
OF SCIENCE AND TECHNOLOGY

SPEECH TECHNOLOGY

Automatic Speech Recognition Challenge
Group 12

ONE LOVE. ONE FUTURE.

Giới thiệu vấn đề

Nhận dạng tiếng nói tự động (ASR) là công nghệ chuyển đổi âm thanh tiếng nói thành văn bản.

Ứng dụng: trợ lý ảo, chuyển đổi giọng nói thành văn bản, hỗ trợ người khuyết tật, nhập liệu nhanh,...

Tiêu chí đánh giá: Word Error Rate (WER)

⇒ Mục tiêu: Xây dựng hệ thống nhận dạng tiếng nói tiếng Việt với độ chính xác cao (WER thấp).

Tổng quan pipeline

Các bước chính của pipeline:

- Cài đặt môi trường và thư viện
- Tiền xử lý dữ liệu audio và transcript
- Chuẩn bị đặc trưng và tokenizer
- Xây dựng, huấn luyện, hậu xử lý và đánh giá mô hình

Môi trường và thư viện

Cài đặt môi trường

- transformers: Cho mô hình Whisper
- datasets[audio]: Xử lý dữ liệu âm thanh
- torchaudio: Xử lý âm thanh
- deepfilternet: Lọc nhiễu
- evaluate, jiwer: Đánh giá mô hình
- bitsandbytes: Tối ưu bộ nhớ
- PEFT: Cho fine-tuning hiệu quả

Tổng hợp thông tin

Mô hình sử dụng:

- Whisper (OpenAI) phiên bản openai/whisper-medium
- Là mô hình pre-trained (huấn luyện trước) trên tập dữ liệu đa ngôn ngữ lớn.

Tập dữ liệu:

- Dữ liệu tiếng Việt từ VLSP 2020 (ví dụ: /kaggle/input/vlsp-trainset/vlsp2020_train_set_02/)
- Số lượng mẫu sử dụng: 10.000 (có thể điều chỉnh)
- Chia train/val/test: 80%/10%/10%

Sử dụng DeepFilterNet để xử lý âm thanh đầu vào, loại bỏ tạp âm

- Lý do: Tăng chất lượng tín hiệu, giúp mô hình nhận dạng chính xác hơn trong môi trường thực tế

Áp dụng n-gram hoặc các kỹ thuật language model để sửa lỗi đầu ra

- Lý do: Giảm lỗi ngữ pháp, tăng độ tự nhiên của kết quả chuyển đổi

- Kết hợp mô hình mạnh (Whisper) với tiền xử lý âm thanh giúp tăng độ chính xác
- Phù hợp với dữ liệu thực tế nhiều nhiễu, đặc biệt cho tiếng Việt
- Dễ mở rộng, tích hợp thêm các bước xử lý khác nếu cần

Nhóm đã đạt được WER trên baseline

11	kythemc	14	06/06/25	40.99 (10)
----	---------	----	----------	------------



HUST

THANK YOU !