

Tổng hợp những gì em đã đạt được theo chủ đề tìm hiểu của tuần 5

Với chủ đề thử nghiệm bài toán asr cho tiếng Việt, em đã thực hiện:

- Tự thực nghiệm những mô hình end-to-end neural asr (CTC-based CNN-RNN và Attention-based), train từ đầu
- Test thử những mô hình open source trên Hugging face

Khó khăn gặp phải:

- Do lần đầu xử lý tín hiệu giọng nói, em gặp rất nhiều khó khăn ngay trong quá trình xử lý tín hiệu:
 - + Tuy chỉ là gọi thư viện để transform wave thành dạng logMel hoặc MFCCs nhưng do tính chất tập dữ liệu em tìm được đều là các câu ngắn chỉ 2 - 10s nên không có kinh nghiệm trong việc chỉnh tham số cho biến đổi nên đã ảnh hưởng rất nhiều đến hiệu quả.

Kết quả thu được:

- + Với mô hình CTC-based, kết quả thu được khá quan nhưng vẫn còn kém những mô hình mở trên hugging face là tương đối, có thể do dữ liệu xử lý chưa được tốt, data chưa đủ nhiều chỉ tầm 3000 đoạn.
- + Với mô hình attention-based, em đang xử dụng phương pháp đoán character tiếp theo như a, b, c. Hiệu quả thì chưa được cao, em đang không biết là có thể do cách làm sai, ví dụ như vẫn là dạng đoán cả token tiếp theo như 1 từ "chào" nó có hiệu quả hơn không do em chưa thử.