MỤC LỤC

DANH MỤC CÁC TỪ VIẾT TẮT VÀ THUẬT NGỮ

Viết tắt	**Tiếng Anh**	**Tiếng Việt**
DNN-HMM	Deep Neural Network-Hidden Markov Models	Mô hình DNN-HMM
ASR	Automatic Speech Recognition	Bài toán nhận dạng tiếng nói tự động
DNN	Deep Neural Network	Mạng nơ ron sâu
LDA	Linear Discriminant Analysis	Mô hình LDA
GMM	Gaussian Mixture Model	Mô hình hỗn hợp Gausian
LSTM	Long Short - Term Memory	Mô hình LSTM
RNN	Recurent Neural Network	Mô hình mạng truy hồi RNN
HMM	Hidden Markov Model	Mô hình HMM
BERT	Bidirectional Encoder Representation from Transformers	Mô hình biểu diễn mã hóa hai chiều từ Transformer - Mô hình BERT
MFCC	Mel Frequency Cepstral Coefficients	Đặc trưng MFCC
NLP	Natural Language Processing	Xử lý ngôn ngữ tự nhiên
IDFT	Invert Discrete Fourier transform	phương pháp IDFT
VAD	Voice Activity Detector	Công cụ phát hiện tiếng nói
GMM-SAT	GMM with Speaker Adaptive Training	Mô hình GMM với huấn luyện phù hợp người nói
TDNN	Time Delay Deep Neural Network	Mạng TDNN
WER	Word Error Rate	Độ đo WER
MLLT	Maximum Likelihood Linear Transformation	Mô hình MLLT
SER	Syllable Error Rate	Độ đo SER
MMI	Maximum Mutual Information	Mô hình MMI
RLAT	Rapid Language Adaptation Tool	Công cụ RLAT

RNNLM	Recurrent Neural Network Language Model	Mô hình ngôn ngữ RNN
BLSTM	Bi-directional Long-Short Term Memory	Mô hình LSTM hai chiều
CTC	Connectionist Temporal Classification	Mô hình CTC

DANH MỤC HÌNH VẼ

DANH MỤC KÝ HIỆU TOÁN HỌC

Ký hiệu	Ý nghĩa
x, y, N, k	In nghiêng, chữ thường hoặc hoa, là các số vô hướng
$\mathbf{x}$, $\mathbf{y}$	In đậm, chữ thường, là các véc-tơ
x_i	Phần tử thứ i của véc tơ $\mathbf{x}$
$\mathbf{A}$, $\mathbf{B}$	In đậm, chữ hoa, là các ma trận
$\mathbf{A}^T$	chuyển vị của ma trận $\mathbf{A}$
$\mathbf{A}^{-1}$	Ma trận nghịch đảo của ma trận $\mathbf{A}$
$\|\mathbf{x}\|$	Chuẩn của véc tơ $\mathbf{x}$
$\circledcirc$	Phép toán với từng phần tử element-wise
$\mathbb{R}$	Tập hợp các số thực
$\mathbb{N}$	Tập hợp các số tự nhiên
$\mathbb{R}^n$	Không gian véc tơ số thực n chiều
$\in$	Thuộc về
$\log(x)$	*logarit* tự nhiên của số thực dương x
$\exp(x)$	Hàm mũ e^x

MỞ ĐẦU

Trong phần này trình bày cách huấn luyện mô hình End-to-End (cụ thể là mô hình wav2vec2) cho bài toán nhận dạng giọng nói. Trong mô hình lai DNN-HMM phải sử dụng rất nhiều phương pháp để trích rút ra đặc trưng của tiếng nói thì mô hình End-to-End hướng tới sự đơn giản mà không cần sử dụng các kỹ thuật trích rút đặc trưng trong pha riêng biệt của hệ thống nhận dạng. End-to-End learning trong bối cảnh AI và ML là một kỹ thuật mà trong đó mô hình học tất cả các bước từ Input đầu vào của bài toán và cho ra output đầu ra cuối cùng mà không cần tách thành các thành phần xử lý riêng biệt. Mô hình End-to-End là một quá trình học sâu trong đó tất cả các phần khác nhau được học đồng thời thay vì thực hiện một cách tuần tự.

Các ưu điểm của mô hình End-to-End[1]:

- Các mô hình End-to-End thường cho hiệu năng cao hơn trên các độ đo Precision và recall
- Tính đơn giản: Các mô hình end-to-end tránh được vấn đề hóc búa là xác định thành phần nào cần thiết để thực hiện một tác vụ và cách các thành phần đó tương tác. Trong các hệ thống dựa trên thành phần (Component-based system), nếu định dạng đầu ra của một thành phần bị thay đổi, th cần phải sửa đổi định dạng đầu vào của các thành phần khác.
- Giảm nỗ lực: Các mô hình end-to-end được cho là cần ít công việc để tạo hơn so với các hệ thống dựa trên thành phần. Các hệ thống dựa trên thành phần yêu cầu số lượng thiết kế lớn hơn
- có khả năng áp dụng cho các nhiệm vụ mới: Các mô hình end-to-end có khả năng hoạt động cho một nhiệm vụ mới chỉ đơn giản bằng cách huấn luyện lại sử dụng dữ liệu mới. Trong khí đó, Các hệ thống dựa trên thành phần cần tái thiết kế lại các modul một cách đáng kể cho các nhiệm vụ mới.

[1]https://www.capitalone.com/tech/machine-learning/pros-and-cons-of-end-to-end-models/

- Khả năng tận dụng dữ liệu tự nhiên: Các mô hình end-to-end có thể được huấn luyện dựa trên dữ liệu hiện có, chẳng hạn như dịch từ ngôn ngữ này sang ngôn ngữ khác. Trong khi hệ thống dựa trên thành phần có thể cần tạo dữ liệu được gắn nhãn mới để huấn luyện trên từng thành phần.
- Tối ưu hóa: Các mô hình end-to-end được tối ưu hóa cho toàn bộ hệ thống của bài toán. Trong khi đó, hệ thống dựa trên thành phần lại Rất khó để tối ưu hóa hệ thống. Lỗi tích lũy xảy ra giữa các thành phần, với một sai sót trong một thành phần ảnh hưởng đến các thành phần phía sau. Thông tin từ các thành phía sau không thể thông báo cho các thành phần trước.
- Mức độ phụ thuộc thấp hơn vào các chuyên gia về chủ đề: Các mô hình end-to-end có thể được đào tạo dựa trên dữ liệu tự nhiên, giúp giảm nhu cầu về kiến thức ngôn ngữ và miền chuyên biệt. Nhưng chuyên môn về mạng nơ-ron sâu lại cần yêu cầu.
- Khả năng tận dụng hoàn toàn việc học máy: Các mô hình end-to-end đưa ý tưởng về máy học lên giới hạn.

Các thách thức của mô hình End-to-End[2]:

- Thiếu khả năng giải thích: Có thể khó hiểu tại sao một mô hình end-to-end lại tạo ra một kết quả nhất định (mặc dù điều này cũng xảy ra ở mức độ thấp hơn đối với các hệ thống dựa trên thành phần).
- Thiếu khả năng dự đoán: Có thể khó dự đoán cách một mô hình end-to-end sẽ hoạt động như thế nào trong nhiều tình huống khác nhau (mặc dù trường hợp này cũng xảy ra ở mức độ thấp hơn đối với các hệ thống dựa trên thành phần).
- Thiếu khả năng chẩn đoán: Trong hệ thống dựa trên thành phần, có thể xác định thành phần nào chịu trách nhiệm cho lỗi, điều này có thể (hoặc có thể không) cho phép một thành phần sửa đổi thành phần để tránh các lỗi tương tự, đồng thời duy trì hiệu suất của hệ thống trong các trường hợp. Trong một hệ thống end-to-end, việc xác định nguồn gốc của lỗi có thể không dễ dàng như vậy. Giải quyết các lỗi trong hệ

[2]https://www.capitalone.com/tech/machine-learning/pros-and-cons-of-end-to-end-models/

thống end-to-end thường liên quan đến việc sửa đổi các tham số mô hình, kiến trúc mô hình hoặc dữ liệu huấn luyện.

- Cường độ dữ liệu: Các mô hình end-to-end yêu cầu một lượng dữ liệu lớn để huấn luyện.
- Huấn luyện và suy diễn tốn kém: Việc huấn luyện và áp dụng các mô hình end-to-end có thể tiêu tốn rất nhiều tài nguyên của máy tính hoặc thậm chí là khó sửa chữa.
- Giới hạn không xác định: Bởi vì các mô hình end-to-end tương đối mới, các giới hạn của chúng chưa được hiểu rõ và cần nghiên cứu thêm về chúng. Một số ứng dụng có thể không phù hợp với các mô hình end-to-end và có thể yêu cầu tạo các hệ thống dựa trên thành phần.

Chính vì lý do trên mà mô hình End-to-End được sử dụng vào bài toán nhận dạng tiếng nói và cho kết quả tốt hơn so với các mô hình dựa vào thành phần hoặc mô hình lai. Nội dung tiếp theo được trình bày như sau: Chương 1: BÀI TOÁN NHẬN DẠNG TIẾNG NÓI TIẾNG VIỆT; Chương 2: MÔ HÌNH WAV2VEC2; Phần cuối cùng là chương 3. XÂY DỰNG ỨNG DỤNG DEMO SỬ DỤNG END-TO-END WAV2VEC2