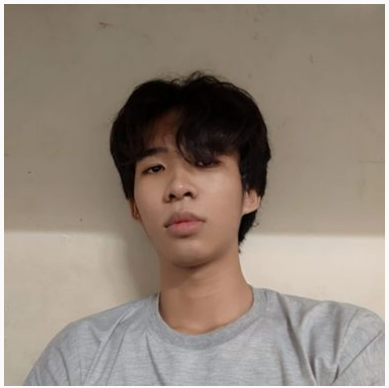
ALMT-VI: PHƯƠNG PHÁP TIẾP CẬN MULTIMODAL SENTIMENT ANALYSIS CHO TIẾNG VIỆT

Phạm Quốc Việt- 21522792

Tóm tắt

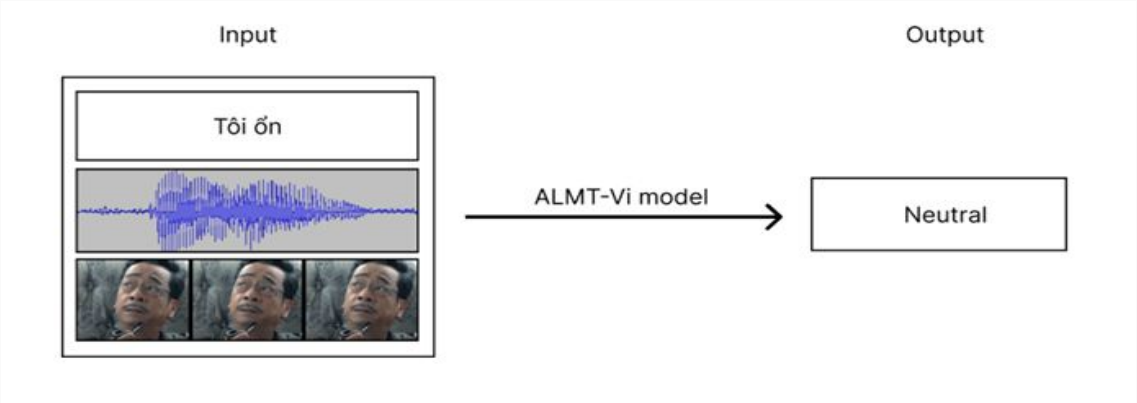
- Lớp: CS519.O11
- Link Github của nhóm: https://github.com/HatakaCder/CS519.O11
- Link YouTube video: https://youtu.be/F37RiPlgGoA



Phạm Quốc Việt - 21522792

Giới thiệu

Vấn đề : Vì mô hình MSA khá mới ở Việt Nam thế nên việc áp dụng MSA với dữ liệu tiếng Việt còn hạn chế. Có cách nào để tăng hiệu suất cho các mô hình MSA hiện nay cho việc xử lý ngôn ngữ Tiếng Việt?



Mục tiêu

- Nghiên cứu mô hình ALMT [1], tích hợp các mô hình ngôn ngữ xử lý văn bản tiếng Việt [2, 3] và các mô hình xử lý âm thanh và video [4, 5], tạo thành một mô hình mới có tên là ALMT-Vi.
- Xây dựng tập dữ liệu UIT-ConversationsMSA50k gồm 50000 đoạn video kết hợp âm thanh và văn bản của người nói.
- Xây dựng ứng dụng minh hoạ nhận biết cảm xúc thông qua các cuộc trò chuyện thoại dựa vào mô hình được nghiên cứu ALMT-Vi.

Nội dung và Phương pháp

Nội dung:

- Nghiên cứu mô hình ALMT [1] cho việc phân tích cảm xúc đa phương thức.
- Tích hợp mô hình ALMT [1] và các đơn mô hình (văn bản, âm thanh, video) [2, 3, 4, 5].
- Tự xây dựng bộ dữ liệu mới UIT-ConversationsMSA50k.
- Huấn luyện mô hình ALMT-Vi và mô hình thuần ALMT[1] trên bộ dữ liệu UIT-ConversationsMSA50k đã thu thập, so sánh và đánh giá từng loại mô hình.
- Xây dựng chương trình ứng dụng minh hoạ.

Nội dung và Phương pháp

Phương pháp:

- Tìm hiểu cấu trúc mô hình ALMT [1].
- Lần lượt áp dụng các mô hình PhoBERT[2]/ViDeBERTa[3], Librosa[4], OpenFace[5] cho việc phân tích xử lý từng loại dữ liệu.
- Tìm hiểu cách đánh giá một mô hình MSA bằng độ đo seven classification accuracy (Acc-7), F1-Score và mean absolute error (MAE).
- Tạo một bộ dữ liệu mới có tên UIT-ConversationsMSA50k gồm 50000 mẫu dữ liệu từ các kênh Youtube với video từ nhiều chủ đề khác nhau.

Nội dung và Phương pháp

- Huấn luyện các mô hình ALMT-Vi, ALMT [1] trên bộ dữ liệu đã được thu thập, so sánh và đánh giá kết quả dựa trên độ đo Acc-7, F1, MAE.
- Xây dựng một chương trình ứng dụng trên nền web cho phép người dùng đăng tải video và dự đoán cảm xúc của người nói theo thời gian thời gian thực.

Kết quả dự kiến

- Tập dữ liệu UIT-ConversationsMSA50k gồm 50000 đoạn video kết hợp âm thanh và văn bản của người nói.
- Báo cáo các phương pháp và kỹ thuật của trong mô hình ALMT-Vi được sử dụng với tập dữ liệu UIT-ConversationsMSA50k. Kết quả thực nghiệm, so sánh và đánh giá với phương pháp thuần ALMT.
- Chương trình minh hoạ việc dự đoán cảm xúc của người nói trong cuộc trò chuyện thoại dựa trên mô hình ALMT-Vi đã được nghiên cứu.

Tài liệu tham khảo

[1]. Haoyu Zhang, Yu Wang, Guanghao Yin, Kejun Liu, Yuanyuan Liu, Tianshu Yu: Learning Language-guided Adaptive Hyper-modality Representation for Multimodal Sentiment Analysis. EMNLP 2023: 756-767

[2]. Dat Quoc Nguyen, Anh Tuan Nguyen: PhoBERT: Pre-trained language models for Vietnamese. EMNLP (Findings) 2020: 1037-1042

[3]. Cong Dao Tran, Nhut Huy Pham, Anh Nguyen, Truong Son Hy, Tu Vu: ViDeBERTa: A powerful pre-trained language model for Vietnamese. EACL (Findings) 2023: 1041-1048

[4]. Brian McFee, Colin Raffel, Dawen Liang, Daniel P. W. Ellis, Matt McVicar, Eric Battenberg, Oriol Nieto: librosa: Audio and Music Signal Analysis in Python. SciPy 2015: 18-24

[5]. Tadas Baltrusaitis, Peter Robinson, Louis-Philippe Morency: OpenFace: An open source facial behavior analysis toolkit. WACV 2016: 1-10