

BÁO CÁO ĐỒ ÁN CUỐI KỲ

Môn học: CS2205 - PHƯƠNG PHÁP LUẬN NCKH

Lớp: CS2205.APR2023

GV: PGS.TS. Lê Đình Duy

Trường ĐH Công Nghệ Thông Tin, ĐHQG-HCM



Đề tài

HƯỚNG ĐẾN TÓM LƯỢC NỘI DUNG NÓI THÀNH VĂN BẢN TỪ ĐẦU ĐẾN CUỐI

(TOWARDS END-TO-END SPEECH-TO-TEXT SUMMARIZATION)

Nguyễn Hoa - 220101044

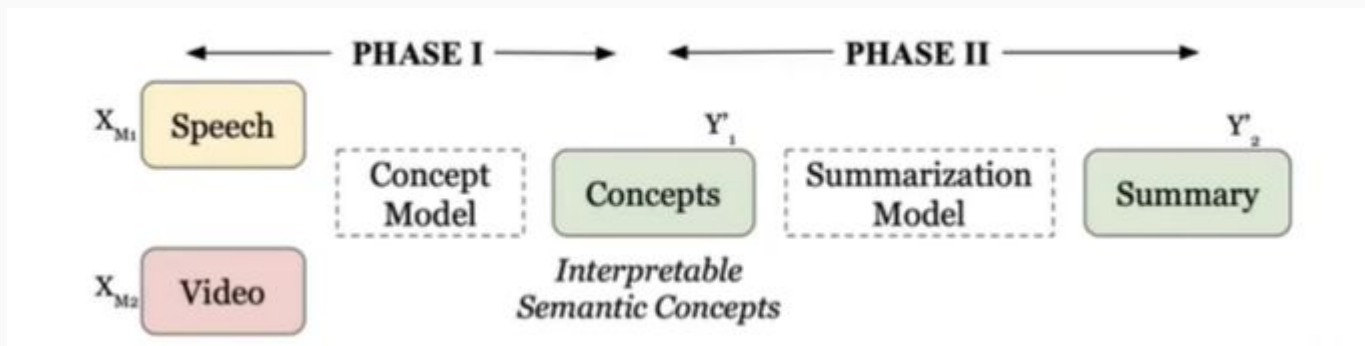
Tóm tắt

- Lớp: CS2205.CH1702-KHMT
- Link Github:
- Link slides:
<https://github.com/akaihachu/CS2205.APR2023/blob/main/slide.pdf>
- <https://github.com/akaihachu/CS2205.APR2023/blob/main/proposal.pdf>
- <https://github.com/akaihachu/CS2205.APR2023/blob/main/brochure.pdf>
- Link YouTube video: <https://youtu.be/-SOdWQcY0yg>
- Họ và Tên: Nguyễn Hoa
- MSSV: 220101044

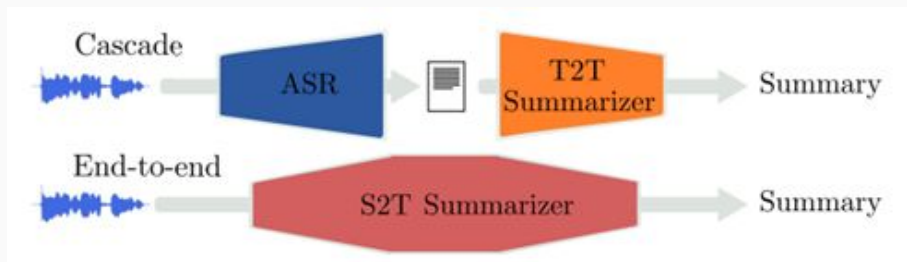


Giới thiệu

- Tin tức phát sóng, podcast, bài giảng, ... được phát hành với số lượng lớn
- Cần tóm lược nội dung S2T thành một văn bản vắn tắt.
- Các hệ thống SOTA tóm lược nội dung S2T sử dụng phương pháp *ghép tầng ASR/T2T* vẫn còn chứa một số lỗi không thể khắc phục, do không thể tận dụng các thông tin ngoài lời thoại và do cách train dữ liệu.



Giới thiệu



- Nhiều mô hình ngôn ngữ khác đã được đề xuất: End to End (E2E).
- Lượng dữ liệu khan hiếm, sự phức tạp và đa dạng của ngôn ngữ, chi phí cao ...
- Đề xuất 02 mô hình E2E: sử dụng transformer và mô hình học chuyển giao (transfer learning) + 01 mô hình ASR/T2T

Do đó cần hướng đến tóm lược nội dung nói thành văn bản từ đầu đến cuối với

- **INPUT:** Một bản tin phát sóng audio.
- **OUTPUT:** Một đoạn văn bản tóm lược nội dung

Mục tiêu

- Nghiên cứu và đánh giá các phương pháp ASR/T2T thông dụng hiện nay.
- Đề xuất phương pháp E2E và so sánh hiệu quả.
- Thực nghiệm bộ dữ liệu có nhãn có số lượng khan hiếm và đánh giá khả năng mở rộng.

Nội dung và Phương pháp

Nội dung

- **Cơ sở lý thuyết**
- **Các công trình liên quan**
- **Phương pháp đề xuất**
- **Thực nghiệm và đánh giá:** Trình bày các thiết lập thực nghiệm, các chỉ số đánh giá, các kết quả thực nghiệm
- **Kết luận và hướng phát triển:** Tổng kết lại các kết quả đạt được, nhận xét về ưu nhược điểm và đưa ra các hướng phát triển trong tương lai.

Nội dung và Phương pháp

Phương pháp

- **Phương pháp dựa trên transformer:** sử dụng một kiến trúc encoder-decoder với restricted self-attention
- **Phương pháp dựa trên transfer learning:** từ mô hình BART, được hoán cải bằng cách thêm vào một lớp *tích chập 1 chiều* (1D convolution) ở đầu của encoder nhằm giảm số chiều

Kết quả dự kiến

- Xây dựng được các mô hình S2T theo phương pháp E2E
- Cải thiện được hiệu suất của tóm lược nội dung S2T
- Đánh giá khả năng mở rộng dữ liệu

Tài liệu tham khảo

- [1] Dana Rezazadegan, Shlomo Berkovsky, Juan C. Quiroz, Ahmet Baki Kocaballi, Ying Wang, Liliana Laranjo, Enrico W. Coiera: *Automatic Speech Summarisation: A Scoping Review*. CoRR abs/2008.11897 (2020)
- [2] Yusen Zhang, Ansong Ni, Tao Yu, Rui Zhang, Chenguang Zhu, Budhaditya Deb, Asli Celikyilmaz, Ahmed Hassan Awadallah, Dragomir R. Radev: *An Exploratory Study on Long Dialogue Summarization: What Works and What's Next*. CoRR abs/2109.04609 (2021)
- [3] Solène Evain, Ha Nguyen, Hang Le, Marcelly Zanon Boito, Salima Mdhaffar, Sina Alisamir, Ziyi Tong, Natalia A. Tomashenko, Marco Dinarelli, Titouan Parcollet, Alexandre Allauzen, Yannick Estève, Benjamin Lecouteux, François Portet, Solange Rossato, Fabien Ringeval, Didier Schwab, Laurent Besacier: *Task Agnostic and Task Specific Self-Supervised Learning from Speech with LeBenchmark*. NeurIPS Datasets and Benchmarks 2021
- [4] Atsunori Ogawa, Tsutomu Hirao, Tomohiro Nakatani, Masaaki Nagata: *ILP-based Compressive Speech Summarization with Content Word Coverage Maximization and Its Oracle Performance Analysis*. ICASSP 2019: 7190-7194
- [5] Shi-Yan Weng, Tien-Hong Lo, Berlin Chen: *An Effective Contextual Language Modeling Framework for Speech Summarization with Augmented Features*. CoRR abs/2006.01189 (2020)
- [6] Sharma, R., Palaskar, S., Black, A.W., Metze, F.: *End-to-end speech summarization using restricted self-attention*. In: ICASSP 2022: 8072–8076.
- [7] Kohei Matsuura, Takanori Ashihara, Takafumi Moriya, Tomohiro Tanaka, Atsunori Ogawa, Marc Delcroix, Ryo Masumura: *Leveraging Large Text Corpora for End-to-End Speech Summarization*. CoRR abs/2303.00978 (2023)
- [8] Akos Tundik, M., Kaszas, V., Szaszak, G.: *Assessing the Semantic Space Bias Caused by ASR Error Propagation and its Effect on Spoken Document Summarization*. In: Proc. Interspeech 2019/1333–1337 (2019).