

20127674 – Lê Đức Đạt

Nhóm đầu tiên trong ngày: Ứng dụng AI tutor học ngoại ngữ

1/ Related work, methods: Dựa vào bài báo Attention is all you need (2017)

Giới thiệu kiến trúc transformer, Bark, Whisper và Gemini.

Flow ứng dụng:

Người dùng sẽ bắt đầu->voice profile (ghi âm bằng chữ cái)->máy chiết xuất đặc trưng->người dùng chọn AI prompt->...->speech-to-text->tính toán điểm phản hồi->gửi lại phản hồi cho mình, đó là 1 vòng lặp

2/Kiến trúc Transformer – Bark

Input->tokenizing->nhúng->encoder->attention(tách riêng, nằm trong cả encode, decode)->decode->output

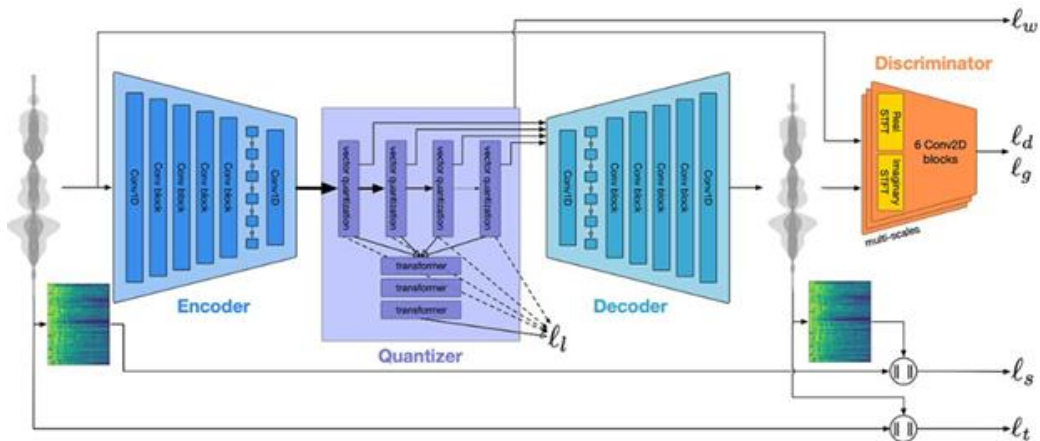
Tokenization: tách theo dấu và tách theo từ, cuối cùng là EOS.

Cơ chế Masked Self-Attention

Self-Attention cho decoder và encoder

Bark: audio -> HuBERT->semantic tokens

Bark Architecture:



3/Result: Demo khá hay

4/Summary: Tóm lược những gì đã làm được và những thứ cần cải thiện (code chạy chậm).

1. Tóm tắt nội dung của nhóm: tập trung vào việc phát triển ứng dụng AI Tutor nhằm hỗ trợ học ngoại ngữ, sử dụng các công nghệ tiên tiến như Transformer, Bark, Whisper, và Gemini. Ứng dụng được thiết kế với một luồng hoạt động phức tạp: người dùng sẽ khởi động bằng cách tạo một hồ sơ giọng nói (ghi âm bằng chữ cái), sau đó hệ thống sẽ chiết xuất các đặc trưng âm thanh, người dùng lựa chọn prompt AI, và tiếp theo là quá trình chuyển đổi giọng nói thành văn bản. Dựa trên văn bản đó, hệ thống sẽ tính toán và gửi phản hồi nhằm cải thiện khả năng học ngoại ngữ của người dùng, tạo nên một vòng lặp học tập hiệu quả. Kiến trúc Transformer, nền tảng cho ứng dụng, đã được giới thiệu thông qua các bước từ mã hóa đầu vào, nhúng, cơ chế attention, và giải mã để tạo ra đầu ra. Cơ chế attention bao gồm self-attention và masked self-attention, được áp dụng cả ở encoder và decoder. Công nghệ Bark sử dụng HuBERT để chuyển đổi âm thanh thành các

token ngữ nghĩa. Mặc dù kết quả demo của hệ thống khá ấn tượng, một số cải tiến vẫn cần được thực hiện, đặc biệt là về tốc độ xử lý mã.

## 2. Những điều đã học được từ chủ đề này:

- Kiến trúc Transformer: Hiểu về cách hoạt động của kiến trúc Transformer, từ việc mã hóa dữ liệu đầu vào, sử dụng các cơ chế attention cho phép hệ thống học các mối quan hệ trong dữ liệu, đến việc giải mã và tạo ra đầu ra.
- Công nghệ Bark: Tìm hiểu về cách Bark xử lý âm thanh bằng cách sử dụng HuBERT để chuyển đổi dữ liệu âm thanh thành các token ngữ nghĩa, giúp cải thiện khả năng xử lý và hiểu ngôn ngữ tự nhiên của hệ thống.
- Ứng dụng AI trong học ngôn ngữ: Khám phá luồng làm việc phức tạp của ứng dụng AI Tutor, từ quá trình tạo hồ sơ giọng nói, xử lý và phản hồi, đến việc cải thiện khả năng học ngoại ngữ một cách tự động và thông minh.

## 3. Đề xuất cải tiến nếu nó trở thành đề tài tốt nghiệp:

- Nghiên cứu sâu hơn về hiệu quả của AI Tutor

Phân tích hiệu quả học ngôn ngữ của AI Tutor: Cần tiến hành các nghiên cứu và thử nghiệm thực tế để đánh giá khả năng cải thiện quá trình học ngôn ngữ của AI Tutor. Việc so sánh AI Tutor với các phương pháp học truyền thống sẽ giúp chứng minh tính hiệu quả của hệ thống.

Đánh giá kết quả học tập: Cần thu thập và phân tích dữ liệu phản hồi từ người dùng để xem AI Tutor có thể giúp cải thiện khả năng phát âm, từ vựng và ngữ pháp của người học như thế nào.

Làm rõ sự khác biệt: Nghiên cứu có thể chỉ ra sự khác biệt giữa việc học ngoại ngữ với AI Tutor và các ứng dụng học ngôn ngữ khác (ví dụ như

Duolingo, Babbel, v.v.), từ đó chỉ ra ưu điểm của hệ thống AI Tutor trong việc phát triển kỹ năng giao tiếp thực tế.

- Tối ưu hóa hiệu suất mã

Giảm độ trễ trong quá trình chuyển đổi giọng nói thành văn bản: Để cải thiện trải nghiệm người dùng, việc giảm độ trễ giữa việc ghi âm giọng nói và nhận phản hồi từ hệ thống là rất quan trọng. Các kỹ thuật tối ưu hóa mã có thể được áp dụng, ví dụ như sử dụng các thuật toán mã hóa hiệu quả hơn hoặc tinh chỉnh các mô hình nhận dạng giọng nói để tăng tốc độ xử lý.

Cải thiện tốc độ xử lý của Bark và Whisper: Các mô hình Bark và Whisper có thể được tối ưu hóa để giảm thiểu thời gian tính toán, từ đó giúp hệ thống phản hồi nhanh hơn.

- Phát triển đánh giá khách quan

Thiết kế công cụ đánh giá chính xác: Cần phát triển các công cụ đánh giá chính xác khả năng tiếp thu của người học. Điều này có thể bao gồm việc xây dựng các bài kiểm tra từ vựng, ngữ pháp, và phát âm dựa trên dữ liệu thu thập được từ AI Tutor.

Phân tích phản hồi AI: Sử dụng các phân tích dữ liệu để tạo ra một mô hình đánh giá khách quan và chi tiết về sự tiến bộ của người học qua thời gian. Hệ thống sẽ đưa ra các lời khuyên cụ thể để cải thiện kỹ năng ngôn ngữ.

- Ứng dụng AI đa phương thức

Tích hợp các mô hình như Gemini: Để tăng cường khả năng hiểu ngôn ngữ qua nhiều phương thức khác nhau, có thể tích hợp thêm các mô hình

như Gemini, giúp AI Tutor không chỉ xử lý giọng nói và văn bản, mà còn kết hợp với các phương thức khác như hình ảnh và video. Điều này sẽ tạo ra một môi trường học ngôn ngữ phong phú và toàn diện hơn.

Khả năng học qua hình ảnh và ngữ cảnh: Bằng cách sử dụng các mô hình đa phương thức, AI Tutor có thể học từ hình ảnh, video, và thậm chí là ngữ cảnh văn hóa, giúp người học dễ dàng tiếp cận và học ngôn ngữ trong các tình huống thực tế.

- **Nâng cao khả năng tùy chỉnh AI Tutor**

Cá nhân hóa quá trình học: Phát triển các thuật toán tùy chỉnh dựa trên mức độ tiến bộ của người học, giúp hệ thống tự động điều chỉnh các bài học theo nhu cầu và trình độ của từng cá nhân. Ví dụ, hệ thống có thể thay đổi các chủ đề học hoặc điều chỉnh độ khó của các bài học dựa trên hiệu suất của người học.

Đo lường mức độ tiến bộ của người học: Cần tạo ra các phương pháp đánh giá và theo dõi sự tiến bộ của người học trong thời gian dài, giúp người học và hệ thống điều chỉnh phương pháp học phù hợp.