TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI

ĐỒ ÁN TỐT NGHIỆP

Chuyển đổi cuộc gọi sang đoạn hội thoại tiếng Việt

Khuất Ngọc Sơn
son.kn204602@sis.hust.edu.vn

Ngành: Khoa học máy tính

Giảng viên hướng dẫn: PGS.TS. Lê Thanh Hương _____

Chữ kí GVHD

Khoa: Khoa hoc máy tính

Trường: Công nghệ Thông tin và Truyền thông

HÀ NỘI, 01/2025

LỜI CAM KẾT

Họ và tên sinh viên: Khuất Ngọc Sơn
Điện thoại liên lạc: 0379135639
Email: son.kn204602@sis.hust.edu.vn
Lớp: Khoa học máy tính 03-K65
Hệ đào tạo: Hệ Kỹ sư chính quy

Tôi – *Khuất Ngọc Sơn* – cam kết Đồ án Tốt nghiệp (ĐATN) là công trình nghiên cứu của bản thân tôi dưới sự hướng dẫn của *PGS.TS. Lê Thanh Hương*. Các kết quả nêu trong ĐATN là trung thực, là thành quả của riêng tôi, không sao chép theo bất kỳ công trình nào khác. Tất cả những tham khảo trong ĐATN – bao gồm hình ảnh, bảng biểu, số liệu, và các câu từ trích dẫn – đều được ghi rõ ràng và đầy đủ nguồn gốc trong danh mục tài liệu tham khảo. Tôi xin hoàn toàn chịu trách nhiệm với dù chỉ một sao chép vi phạm quy chế của nhà trường.

Hà Nội, ngày 06 tháng 01 năm 2025

Tác giả ĐATN

Sơn

Khuất Ngọc Sơn

LỜI CẢM ƠN

Đồ án này được em hoàn thành tại Trường Công nghệ Thông tin và Truyền thông, Đại học Bách Khoa Hà Nội dưới sự hướng dẫn của giáo viên hướng dẫn của PGS.TS. Lê Thanh Hương.

Lời cảm ơn đầu tiên em xin gửi tới PGS.TS. Lê Thanh Hương, người đã giúp đỡ và hướng dẫn em tận tình trong khoảng thời gian vừa qua để em có thể hoàn thành môn học cuối cùng này.

Tiếp theo, em xin được gửi lời cảm ơn chân thành đến đại học Bách Khoa Hà Nội, nơi chấp cánh ước mơ của bao thế hệ sinh viên và giúp chúng em nên người và trở thành những công dân tốt và có ích cho nước nhà.

Em xin được cảm ơn quý giảng viên đã và đang làm việc tại đây, đặc biệt là những thầy cô giảng viên trong hơn 4 năm qua đã bảo ban và truyền đạt cho em bao nhiêu kiến thức hay và bổ ích để em có thể phát triển bản thân mình.

Và cuối cùng không thể thiếu, em xin cảm ơn gia đình nơi điểm tựa tinh thần lớn lao để em bước tiếp trên con đường của chính mình. Cảm ơn bạn bè đã đồng hành cùng em trong những năm học Đại học và tương lai xa nữa.

Trong quá trình hoàn thành báo cáo và đồ án khó có thể tránh khỏi sai sót, em mong thầy cô và bạn đọc có thể góp ý để em hoàn thiện hơn nữa đồ án tốt nghiệp này.

TÓM TẮT NỘI DUNG ĐỒ ÁN

Trong thời đại công nghệ số phát triển mạnh mẽ, việc tự động hóa các quy trình trong doanh nghiệp ngày càng trở nên quan trọng. Một trong những lĩnh vực nổi bật là chuyển đổi dữ liệu từ các cuộc gọi telesale sang văn bản và trích xuất thông tin quan trọng nhằm tối ưu hóa hoạt động kinh doanh. Hiện nay, mặc dù có nhiều giải pháp xử lý âm thanh và nhận diện giọng nói, việc áp dụng chúng trong ngữ cảnh cụ thể như telesale vẫn còn nhiều thách thức. Các phương pháp truyền thống như ghi âm và phân tích thủ công không chỉ tốn thời gian mà còn dễ dẫn đến sai sót. Các giải pháp tự động, dù đã đạt được một số tiến bộ, vẫn gặp hạn chế về độ chính xác, đặc biệt trong việc nhận diện giọng nói vùng miền hoặc trong các đoạn hội thoại phức tạp.

Để giải quyết vấn đề này, nghiên cứu lựa chọn hướng tiếp cận bằng sử dụng các mô hình xử lý âm thanh tiên tiến như Whisper, Wav2Vec hay Whisperx. Lý do lựa chọn hướng này là khả năng mở rộng và tiềm năng cải thiện độ chính xác khi áp dụng vào các bài toán thực tiễn.

Giải pháp đề xuất bao gồm ba giai đoạn chính: (i) chuyển đổi âm thanh cuộc gọi sang văn bản bản thông qua các mô hình nhận diện giọng nói, (ii) hậu xử lý đoạn văn bản hội thoại và rút được văn bản tóm tắt, và (iii) sử dụng mô hình GPT để trích xuất thông tin quan trọng như tên khách hàng, sản phẩm quan tâm, và kết quả cuộc gọi. Hệ thống được thiết kế với khả năng tùy chỉnh theo nhu cầu của từng doanh nghiệp.

Đóng góp chính của đề tài là phát triển một quy trình tích hợp từ nhận diện giọng nói đến trích xuất thông tin, giúp tăng hiệu suất và giảm chi phí xử lý thủ công. Kết quả thử nghiệm cho thấy hệ thống đạt độ chính xác cao trong cả hai bước nhận diện và trích xuất, hứa hẹn mang lại giá trị thực tiễn cao trong lĩnh vực telesale và chăm sóc khách hàng.

Sinh viên thực hiện

ABSTRACT

In the era of rapidly advancing digital technology, automating business processes has become increasingly important. One prominent area is the transformation of telesales call data into text and the extraction of key information to optimize business operations. Although numerous solutions for audio processing and speech recognition are currently available, their application in specific contexts such as telesales still presents many challenges. Traditional methods, such as recording and manual analysis, are not only time-consuming but also prone to errors. While automated solutions have made some progress, they still face limitations in accuracy, especially when dealing with regional accents or complex conversations.

To address this issue, this study adopts an approach based on advanced audio processing models like Whisper, Wav2Vec, and Whisperx. These models are chosen for their scalability and potential to improve accuracy when applied to real-world problems.

The proposed solution comprises three main phases: (i) converting call audio into text using speech recognition models, (ii) post-processing the conversational text to generate summaries, and (iii) employing GPT models to extract critical information such as customer names, products of interest, and call outcomes. The system is designed to be customizable to meet the specific needs of different businesses.

The primary contribution of this study lies in developing an integrated process that spans from speech recognition to information extraction, enhancing efficiency while reducing manual processing costs. Experimental results demonstrate high accuracy in both recognition and extraction steps, promising significant practical value in telesales and customer service domains.

MỤC LỤC

DANH MỤC HÌNH VẼ

DANH MỤC BẢNG BIỂU

DANH MỤC THUẬT NGỮ VÀ TỪ VIẾT TẮT

Thuật ngữ	**Ý nghĩa**
ĐATN	Đồ án tốt nghiệp
AI	Trí tuệ nhân tạo (Artificial Intelligence)
API	Giao diện lập trình ứng dụng (Application Programming Interface)
ASR	Nhận dạng tiếng nói tự động (Automatic Speech Recognition)
CER	Tỷ lệ lỗi ký tự (Character Error Rate))
DTW	Dynamic Time Warping)
NLP	Xử lý ngôn ngữ tự nhiên (Natural Language Processing)
OCR	Nhận dạng ký tự quang học (Optical Character Recognition))
VAD	Phát hiện hoạt động giọng nói (Voice Activity Detection)