# Giới Thiệu

Hiện nay, công nghệ thông tin ngày càng phát triển và đóng góp to lớn trong sự phát triển của ngành công nghiệp, một trong những ngành quan trọng, lâu đời, cần thiết và được áp dụng hầu hết các lĩnh vực khác nhau như quân sự, hàng không, giao thông, công nghiệp, giáo dục,..đặc biệt là trong cuộc sống của chúng ta.

Hiện nay, công nghệ thông tin đang ngày càng phát triển và đóng góp rất to lớn trong sự phát triển của ngành công nghiệp. Một trong những chuyên ngành quan trọng, lâu đời và được áp dụng trong nhiều lĩnh khác nhau như y học, quân sự, giao thông… đó là xử lý ảnh. Hầu hết mọi người cảm nhận thông tin hoặc hình ảnh một cách trực quan. Trong xử lý ảnh, việc nhận dạng và phân lớp các đối tượng khác nhau cần trải qua nhiều giai đoạn cũng như nhiều thao tác. Phát hiện biên là một trong những giai đoạn rất quan trọng trong quá trình xử lý hình ảnh để đánh dấu những điểm trong một ảnh số mà có sự thay đổi đột ngột về độ xám, tập hợp nhiều điểm biên tạo nên một đường bao quanh ảnh được gọi là đường biên. Từ đó có thể phân biệt được giữa đối tượng và nền, giữa vùng này và vùng kia để nhận dạng các đối tượng đã được định vị. Nhưng trên thực tế có rất nhiều kỹ thuật phát hiện biên hiện đang được sử dụng trong 20 năm trở lại đây, mỗi kỹ thuật thường được áp dụng với một loại đường biên cụ thể. Bài viết này tổng quát về các phương pháp phát hiện biên, từ các kỹ thuật dò biên để đưa ra các nhận xét, đánh giá về các phương pháp phát hiện biên.

# Khái quát về ngành

Ngành trí tuệ nhân tạo (TTNT) là một ngành nghề mới trong lĩnh vực khoa học máy tính. Trí tuệ nhân tạo với tên tiếng anh Artificial Intelligence hay còn được viết tắt AI, được định nghĩa một cách đơn giản, dễ hiểu là trí thông minh được thể hiện bằng máy móc do con người tạo ra. Điều này ngược lại với trí thông minh của con người hoặc trí thông minh của một vài loài động vật ngoài tự nhiên khác. Nghiên cứu và phát triển trí tuệ nhân tạo là làm cho những cỗ máy "vô tri, vô giác" có được khả năng của trí tuệ và xử lý thông minh của con người. Cụ thể hơn là biết suy nghĩ, lập luận đưa ra cách giải quyết vấn đề;  có giao tiếp (hiểu, phản ứng) với ngôn ngữ tự nhiên của con người (tiếng nói, chữ viết); có khả năng học và tự thích nghi với những môi trường và điều kiện khác nhau.

Mục đích của việc nghiên cứu AI là để sao chép lại quá trình học tập, tư duy và suy nghĩ của con người sau đó áp dụng cho máy học. Quá trình nghiên cứu, phát triển trí tuệ nhân tạo bao gồm học tập (tự thu thập, học hỏi những thông tin xung quanh và tìm ra quy tắc sử dụng thông tin), sau đó lập luận (lập luận và tư duy thông tin theo những quy tắc từ trước, để cho ra kết quả) và khả năng tự sửa lỗi.

Khái niệm về công nghệ AI được xuất hiện đầu tiên trên thế giới vào năm 1956 tại hội nghị The Dartmouth. Người đã đề xuất cho ý tưởng này là nhà khoa học máy tính người MỹJohn McCarthy. Ngày nay, khi nhắc đến AI tức là bao gồm tất cả mọi thứ từ quá trình tự động hoá robot đến người máy thực tế và nhiều ứng dụng thông minh khác.

# Khái quát về đề tài

Nhận dạng giọng nói đã được biết đến hàng thập kỷ, tại sao chỉ đến bây giờ, công nghệ mới thực sự bùng nổ? Theo wikipedia, khó khăn cơ bản của nhận dạng giọng nói đó là tiếng nói luôn biến thiên theo thời gian và có sự khác biệt lớn giữa tiếng nói của những người nói khác nhau, tốc độ nói, ngữ cảnh và môi trường âm học khác nhau. Sự ra đời của Deep Learning đã giúp nhận diện giọng nói chính xác, thậm chí ở ngoài môi trường phòng lab.

Ví dụ, đối với Google Assistant, bạn có thể dễ dàng tìm kiếm chỉ với giọng nói câu lệnh từ bạn. Nó là một phần của việc chuyển đổi giọng nói thành văn bản. Ở một mức độ cao hơn, Google Assistant có thể hiểu được câu nói của bạn và phản hồi lại với một kết quả có thể nói là gần như hoàn hảo. Để có thể có được một mức độ thông minh như vậy thì AI cần nguồn dữ liệu lớn để học hỏi, quá trình này do người dùng cung cấp cũng như do chính bạn tương tác thường xuyên với Google Assistant.

Để AI thông minh thì cần phải có dữ liệu để huấn luyện cho nó, cả về nhận diện hình ảnh, văn bản, giọng nói. Google có hàng tỷ người dùng với công cụ tìm kiếm, nó có thể biết được trong khoảng thời gian nào, trong từng thời điểm người dùng quan tâm từ khóa nào, lĩnh vực nào. Đó là một cách người dùng tự tạo dữ liệu cho AI. Cũng còn một cách là người dùng trực tiếp cung cấp dữ liệu cho AI.

### Cách thức xây dựng công nghệ Nhận dạng, giả lập giọng nói

Vậy người ta áp dụng công nghệ giọng nói vào phần mềm như thế nào? Thông thường một bộ máy giọng nói sẽ có hai phần. Phần thứ nhất gọi là Speech synthesizer (còn gọi là Text to Speech hay TTS). Đây là một trình tổng hợp giọng nói và thiết bị hoặc ứng dụng xài để tương tác với người dùng, ví dụ: đọc văn bản trên màn hình, thông báo về tiến độ chạy một tác vụ nào đó. Phần thứ hai là một công nghệ nhận dạng cho phép app biết được người dùng đang nói gì, từ đó chuyển thể thành lệnh để thiết bị thực thi hoặc chuyển đổi thành các kí tự nhập liệu. Nói cách khác, đây là thứ thay thế cho bàn phím của chúng ta. Một ứng dụng nhận dạng giọng nói lý tưởng sẽ bao gồm cả hai bộ phận nói trên, nhưng một số app chỉ xài một cái rồi từ từ nâng cấp sau.

Thoạt nhìn thì việc triển khai công nghệ nhận dạng giọng nói khá đơn giản, nhưng thực chất thì không phải như thế.

Thứ nhất, các nhà phát triển phải xây dựng nên một công nghệ có thể lắng nghe, phân tích và phiên dịch một cách chính xác giọng nói của người dùng. Nếu không thì làm sao app biết bạn đang nói gì, còn nếu độ chính xác không cao thì cũng như không.

Thứ hai, vấn đề bản địa hóa (localization) cũng là một chuyện làm đau đầu các lập trình viên. Mỗi quốc gia sẽ có ngôn ngữ của riêng mình, vấn đề đó là làm thế nào để có thể hỗ trợ càng nhiều ngôn ngữ càng tốt.

Có một kĩ thuật được nhắc đến nhiều trong thời gian gần đây, đó là Xử lý ngôn ngữ tự nhiên (Natural Language Processing – NLP). Nó là tập hợp của nhiều thuận toán phức tạp nhằm phân tích mệnh lệnh của người dùng nhưng không bắt buộc họ phải nói theo một cấu trúc câu định sẵn. Nhiều năm trước khi muốn điều khiển bằng giọng nói, bạn chỉ có thể nói những thứ như “Mở bản đồ”, “Nhắn tin cho vợ”, “Báo thức lúc 5 giờ sáng”. Còn bây giờ thì nhờ có NLP, chúng ta có thể nói các câu như “Siri, vui lòng nhắn tin cho vợ của tôi là tôi sẽ về trễ nhé”, hay như “Hãy đánh thức tôi lúc 5 giờ sáng ngày mai”.

# Ngoại vi và nội vi

Khái niệm định nghĩa

Phân tích

Cấu trúc