|  |  |
| --- | --- |
| **BỘ GIÁO DỤC VÀ ĐÀO TẠO** | **BỘ QUỐC PHÒNG** |

**HỌC VIỆN KỸ THUẬT QUÂN SỰ**

**\_\_\_\_\_\_\_\_\_\_\_\_\_**

**NGUYỄN PHÚC TÂN**

**NGHIÊN CỨU CÔNG NGHỆ NHẬN DẠNG GIỌNG NÓI VỚI MÔ HÌNH HIDDEN MARKOV MODEL VÀ ỨNG DỤNG TRONG VIỆC ĐIỀU KHIỂN TRÌNH DUYỆT CHROME**

**CHUYÊN NGÀNH: Công Nghệ Phần Mềm**

**ĐỀ CƯƠNG MÔN HỌC PHƯƠNG PHÁP NGHIÊN CỨU IT**

**HÀ NỘI – NĂM 2015**

|  |  |
| --- | --- |
| **BỘ GIÁO DỤC VÀ ĐÀO TẠO** | **BỘ QUỐC PHÒNG** |

**HỌC VIỆN KỸ THUẬT QUÂN SỰ**

**\_\_\_\_\_\_\_\_\_\_\_\_\_**

**ĐỀ CƯƠNG MÔN HỌC PHƯƠNG PHÁP NGHIÊN CỨU IT**

**Chuyên ngành:** Công Nghệ Phần Mềm

**Mã số:**

**Ngày giao đồ án:**

**Ngày nộp đồ án:**

**Tên đề tài:**

NGHIÊN CỨU CÔNG NGHỆ NHẬN DẠNG GIỌNG NÓI VỚI MÔ HÌNH HIDDEN MARKOV MODEL VÀ ỨNG DỤNG TRONG VIỆC ĐIỀU KHIỂN TRÌNH DUYỆT CHROME

**Học viên thực hiện:** Nguyễn Phúc Tân

**Lớp:** Công Nghệ phần mềm 10

**Cán bộ hướng dẫn:**

**Họ và tên:** Ngô Hữu Phúc

**Cấp bậc:**

**Học hàm, học vị:**

**Đơn vị:** BM Khoa học máy tính – Khoa CNTT – HVKTQS

**HÀ NỘI – NĂM 2015**

**ĐỀ CƯƠNG MÔN HỌC PHƯƠNG PHÁP NGHIÊN CỨU IT**

Tên đề tài: **NGHIÊN CỨU CÔNG NGHỆ NHẬN DẠNG GIỌNG NÓI VỚI MÔ HÌNH HIDDEN MARKOV MODEL VÀ ỨNG DỤNG TRONG VIỆC ĐIỀU KHIỂN TRÌNH DUYỆT CHROME**

Chuyên ngành: Công Nghệ Phần Mềm

Thời gian thực hiện: 4 tháng

1. **Cơ sở khoa học và tính thực tiễn của đề tài**

* Cơ sở khoa học

+ Lý thuyết về âm học, ngữ âm học, đặc điểm trong tiếng Việt

+ Các kiến thức cơ bản để xây dựng và sử dụng 1 hệ nhận dạng tiếng nói

+ Cở sở về rút trích đặc trưng

+ Lý thuyết về mô hình Hidden Markov Model (HMM)

* Tính thực tiễn

*Lịch sử phát triển:*

*Trên thế giới:*

+ Nghiên cứu nhận dạng tiếng nói tự động (Automatic speech recognition - ASR) đã trải qua quá trình 70 năm phát triển

+ Những năm 50: Các nỗ lực nghiên cứu ASR được tiến hành với ý tưởng dựa trên ngữ âm. Do khả năng của máy tính còn giới hạn nên các hệ thống nhận dạng lúc đó chỉ tập trung khai thác đặc trưng phổ cộng hưởng đối với các nguyên âm của tín hiện, sau khi đi qua bộ lọc tương tự.

Các hệ thống đáng chú ý như: hệ thống nhận dạng ký tự số rời rạc của Bell-lab(1952), bộ nhận dạng 13 âm vị của trường đại học College-Anh(1958) [1, p. 8]

+ Những năm 60: tác giả người Nga: Vintsyuk đề xuất phương pháp nhận dạng tiếng nói dựa trên quy hoạch động theo thời gian [2, p. 1]

Reddy ở trường đại học CMU ( Mỹ) đề xuất ý tưởng về nhận dạng tiếng nói liên tục bằng kỹ thuật đánh dấu đường đi và truy vết lùi tìm kết quả [2, p. 2]

+ Những năm 70: nghiên cứu về nhận dạng tiếng nói bước đầu thu được kết quả khích lệ.

Các hệ thống đáng chú ý: Harpy và Hearsay-II của trường CMU (Mỹ), hệ thống HWIM của BBN .. [2, p. 2]

+ Những năm 80: Chuyển dịch từ cách tiếp cận đối sánh mẫu sang cách tiếp cận sử dụng mô hình thống kê ( ngày nay hầu hết các hệ thống ASR đều dựa trên mô hình này)

Các phát minh quan trọng: mô hình Markov ẩn (Hidden Markov Model - HMM), ngoài ra còn 2 đề xuất quan trọng khác là tổ hợp Cepstrum và mô hình ngôn ngữ.

Các hệ thống ra đời trong giai đoạn này: Sphinx của trường đại học CMU, Byblos của công ty BBN...

+Những năm 90: ghi nhận 1 số kết quả nghiên cứu mới trong lĩnh vực phân lớp mẫu.

Các ứng dụng được phát triển: hệ thống trả lời thông tin tự động cho các chuyến bay (Air Travel Information Service - ATIS), hệ thống ghi lại các bản tin phát thanh (Broadcast News Tráncription System)..

+ Các năm đầu thế kỷ 21: Các nghiên cứu tập trung vào việc nâng cao kết quả nhận dạng tiếng nói, thông qua chương trình có tên EARS (Effective Affordable Reuseable Speech-to-Text) [2, p. 3]

Hiện nay các công trình nghiên cứu với các ngôn ngữ phổ biến như Anh, Pháp, Tây Ban Nha đã thu được những kết quả tốt, có nhiều ứng dụng thực tiễn được triển khai

*Trong nước:*

Tại Việt Nam, có 2 nhóm nghiên cứu chính về ASR:

+ Nhóm đầu tiên thuộc Viện Công nghệ Thông tin do PGS. Lương Chi Mai đứng đầu với phương pháp ANN và công cụ CSLU [3]

+ Nhóm thứ 2 thuộc trường Đại học Khoa học Tự nhiên thành phố Hồ Chí Minh do PGS. Vũ Hải Quân đứng đầu với phương pháp HMM và công cụ HTK, các nghiên cứu tập trung vào bài toán truy vấn thông tin tiếng Việt, nhận dạng tiếng nói, hệ thống giao tiếp giữa người và máy, tìm kiếm bằng giọng nói...

+ Gần đây có nghiên cứu của LIG (Laboratoire Inforatique de Grenoble) hợp tác với phòng thí nghiệm MICA ở Hà Nội về sự khả chuyển của các mô hình ngữ âm.

*Ứng dụng:*

Trong nhiều lĩnh vực như kinh tế, thương mại, nghiên cứu khoa học ..

*Các ứng dụng nổi bật như:*

+ Hệ thông trả lời thông tin tự động các chuyến bay(Air travel information service - ATIS),

+ Hệ thống ghi lại các bản tin phát thanh(Broadcast News Transcription System) ,

+ Hệ thống hỏi đáp thông tin tự động qua điện thoại,

+ Hệ thống thông dịch xuyên ngôn ngữ tự động,

+ Hệ thống kiểm soát điều khiển sử dụng giọng nói

* Kết quả nghiên cứu trong đề tài này có thể dùng để:

+ Giảng dạy, nghiên cứu

+ Kinh tế

+ Kỹ thuật

1. **Mục tiêu đề tài:**

* Tìm hiểu các khái niệm có liên qua đến hệ nhận dạng tiếng nói để làm rõ hơn 1 số yếu tố quan trọng trong viêc sử dụng công cụ hỗ trợ
* Tìm hiểu phương pháp cài đặt công cụ hỗ trợ xây dựng hệ nhận dạng tiếng nói
* Tìm hiểu xây dựng mô hình âm học, mô hình ngôn ngữ thích hợp cho tiếng Việt
* Xây dựng chương trình mô phỏng áp dụng nghiên cứu vào việc điều khiển trình duyệt Chrome

1. **Phương pháp nghiên cứu:**

* Về lý thuyết:

+ Lý thuyết về âm học, ngữ âm học, đặc điểm trong tiếng Việt

+ Các kiến thức cơ bản để xây dựng và sử dụng 1 hệ nhận dạng tiếng nói

+ Cở sở về rút trích đặc trưng

+ Lý thuyết về mô hình Hidden Markov Model (HMM)

* Về thực nghiệm:

+ Giới thiệu về 2 framework hỗ trợ để xây dựng 1 hệ nhận dạng tiếng nói phổ biến hiện nay: HTK, Sphinx4.

+ Trình bày cách cài đặt Sphinx 4 , thu âm, xây dựng bộ huấn luyện, tiến hành huấn luyện, đánh giá hệ thống

1. **Nội dung nghiên cứu:**

* Thu thập tài liệu phục vụ nghiên cứu:

+ Thu thập các tài liệu liên quan đến nhận dạng giọng nói

+ Thu thập kho dữ liệu giọng nói làm cơ sở để nghiên cứu, thực nghiệm.

* Nghiên cứu tổng quan:

+ Lý thuyết về âm học, ngữ âm học, đặc điểm trong tiếng Việt

+ Cơ sở rút trích đặc trưng

* Nghiên cứu của tác giả:

+ Lý thuyết về mô hình Hidden Markov Model (HMM)

+ Ứng dụng mô hình HMM vào bài toán nhận diện giọng nói để điều khiển trình duyệt Chrome

* Kết quả đạt được:

+ Ứng dụng mô hình HMM vào bài toán nhận dạng giọng nói để điều khiển trình duyệt Chrome thành công với lệnh điều khiển gồm 20 từ, tỷ lệ xác suất nhận dạng đúng ~ 80%.

**LỜI GIỚI THIỆU**

**ĐẶT VẤN ĐỀ**

**Chương 1:** Trình bày tổng quan: giới thiệu về đề tài, xác định mục tiêu những vấn đề cần giải quyết, giới hạn pahmj vi đề tài, xác định phương pháp giải quyết và chỉ ra những điểm nổi bật của đề tài.

**Chương 2:** Trình bày cơ sở lý thuyết bao gồm: lý thuyết cơ bản về âm học, ngữ âm học, đặc điểm trong tiếng Việt, các kiến thức cơ bản về xây dung và sử dụng 1 hệ nhận dạng tiếng nói. Cơ sở lý thuyết về rút trích đặc trưng, 1 trong những khái niệm quan trọng trong các hệ nhận dạng tiếng nói. Lý thuyết về mô hình Hidden Markov Model (HMM). Bao gồm khái niệm, các thuật toán liên quan, ý nghĩa của HMM trong 1 hệ nhận dang tiếng nói.

**Chương 3:** Giới thiệu các khái niệm cơ bản, quan trọng trong 2 framework hỗ trợ xây dựng 1 hệ nhận dạng tiếng nói phổ biến hiện này là HTK và Sphinx 4. Đồ án này sử dụng Sphinx 4 để xây dựng chương trình demo thực nghiệm.

**Chương 4:** Trình bày chi tiết qúa trình cài đặt Sphinx 4, thu âm, xây dựng bộ huấn luyện, tiến hành huấn luyện, giải thích kết quả huấn luyện, thực nghiệm thử nghiệm, xây dựng chương trình demo

**Chương 5:** Nêu kết luận, trình bày kết quả đạt được, những điểm hạn chế, kinh nghiệm rút ra, nêu các hướng cải thiện, nghiên cứu phát triển.

1. **Tài liệu tham khảo:**

[1] B.H. Juang, Lawrence R. Rabiner. “Automatic Speech Recognition – A Brief History of the Technology”.

[2] S. Furui. “50 years of progress in speech and speaker recognition”.

[3] [Online]. Available: <http://www.cslu.ogi.edu/toolkit/>.

1. **Dự kiến kế hoạch thực hiện:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **STT** | **Nội dung công việc** | **Thời gian thực hiện** | **Địa điểm** | **Ghi chú** |
| 1 | Tổng quan về nhận dạng giọng nói, thu thập số liệu | 2 tuần |  |  |
| 2 | Mô hình HMM | 4 tuần |  |  |
| 3 | Xây dựng demo | 6 tuần |  |  |
| 4 | Kiểm thử, đánh giá kết quả và hoàn thiện mô hình | 4 tuần |  |  |
| 5 | Viết báo cáo | 3 tuần |  |  |
| 6 | Hoàn thiện, kiểm tra, đóng quyển | 1 tuần |  |  |

1. **Các cơ quan, đơn vị cần liên hệ:**

- Khoa CNTT – Học viện kỹ thuật quân sự

1. **Kinh phí thực hiện đề tài:**

|  |  |
| --- | --- |
|  | Ngày tháng năm 2015 |
| **Chủ nhiệm bộ môn** | **Người lập đề cương**  Nguyễn Phúc Tân |
| **Chủ nhiệm khoa** | **Giáo viên hướng dẫn** |