**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN**

**KHOA CÔNG NGHỆ THÔNG TIN**

**🙠🕮🙢**

**XỬ LÝ NGÔN NGỮ NÓI**

**BÀI TẬP 4: TỔNG HỢP TIẾNG NÓI**

GVHD: PGS. TS. Vũ Hải Quân

Nhóm thực hiện:

Ngành Khoa Học Máy Tính – Cao học khóa 23

1. Đỗ Đặng Minh - 13 11 015
2. Huỳnh Công Toàn - 13 11 026
3. Hồ Văn Tấn - 13 11 058

*Tp. Hồ Chí Minh, tháng 06 năm 2014*

# Thông tin nhóm

|  |  |  |  |
| --- | --- | --- | --- |
| **MSHV** | **Họ tên** | **Số điện thoại** | **E-mail** |
| 13 11 015 | Đỗ Đặng Minh | 0168-993-5242 | [masterminh219@gmail.com](mailto:masterminh219@gmail.com) |
| 13 11 026 | Huỳnh Công Toàn | 0121-516-1090 | [alex7huynh@gmail.com](mailto:alex7huynh@gmail.com) |
| 13 11 058 | Hồ Văn Tấn | 090-290-9334 | [tanhv90@gmail.com](mailto:tanhv90@gmail.com) |

Mục lục

[Thông tin nhóm 1](#_Toc390289445)

[I. Cơ sở lý thuyết 2](#_Toc390289446)

[1) Trích chọn đặc trưng 2](#_Toc390289447)

[2) Mô hình Markov ẩn (HMM) 2](#_Toc390289448)

[3) Các bài toán với mô hình Markov ẩn 3](#_Toc390289449)

[II. Thực nghiệm 3](#_Toc390289450)

[1) Dữ liệu thực nghiệm 3](#_Toc390289451)

[2) HTK 4](#_Toc390289452)

[3) Phương pháp thực hiện 5](#_Toc390289453)

[4) Các trường hợp chuẩn hóa dữ liệu đầu vào 7](#_Toc390289454)

[a) Các ký hiệu đặc biệt 7](#_Toc390289455)

[b) Các ký hiệu toán học 7](#_Toc390289456)

[c) Tên địa danh 7](#_Toc390289457)

[d) Khoảng cách 7](#_Toc390289458)

[e) Nhiệt độ 7](#_Toc390289459)

[f) Tiền tệ 7](#_Toc390289460)

[g) Số 7](#_Toc390289461)

[h) Ngày tháng 7](#_Toc390289462)

[5) Nhận xét 7](#_Toc390289463)

[III. Tài liệu tham khảo 7](#_Toc390289464)

# Cơ sở lý thuyết

1. Trích chọn đặc trưng

Các đặc trưng mang thông tin về tiếng nói là các tần số formant, tần số âm cơ bản và năng lượng.

Các đặc trưng mang thông tin về đường bao phổ:

* Hệ số dự đoán tuyến tính (LPC).
* Các hệ số cepstrum.
* Các hệ số cepstrum dự đoán tuyến tính (LPCC).
* Các hệ số cepstrum tần số mel (MFCC).

Trích chọn đặc trưng trong xử lý ngôn ngữ nói là biến đổi từ giọng nói thô thành những đặc trưng của người nói. Đặc trưng được sử dụng trong bài này là đặc trưng MFCC (Mel Frequency Cepstral Coefficients), đặc trưng MFCC thể hiện hình thái của cuống họng, … khi ta nói.

Các bước trính chọn đặc trưng MFCC:

* Phân khung (Frame blocking): chia tín hiệu đầu vào thành các đoạn nhỏ khoảng 20ms­30ms. Phân khung tín hiệu mỗi khung N mẫu, hai khung kề nhau lệch M mẫu.
* Lấy cửa sổ (Windowing): giảm sự gián đoạn của tín hiệu ở đầu và cuối mỗi khung vừa được chia.
* Biến đổi FFT (Fast Fourier Transform): chuyển đổi mỗi khung với N mẫu từ miền thới gian sang miền tần số.
* Biến đổi sang thang đo mel (Mel-frequency wrapping).
* Hệ số cepstrum (cepstral coefficients).

1. Mô hình Markov ẩn (HMM)

Mô hình Markov ẩn có nền tảng của mô hình thống kê được sử dụng trong hệ thống nhận dạng giọng nói hiện đại. Các biến thể của HMMs được sử dụng rộng rãi và thường được coi là thành công nhất.

Cho *µ=(S, A, Π)* gọi *X = (X1, X2, …, XT)* là một dãy các biến ngẫu nhiên với giá trị nằm trong tập *S={s1, s2, s3, …, sN}* - tập không gian trạng thái, thỏa các tính chất sau:

* Xác xuất chuyển trạng thái độc lập với thời gian.
* Xác suất chuyển trạng thái (lưu vào ma trận A): và .
* Xác suất trạng thái khởi động: và .

µ được mô tả như trên là một mô hình Markov.

Một HMM là một bộ 5 (S, K, Π, A, B):

* Tập trạng thái: *S={s1, s2, s3, …, sN}*
* Tập ký tự xuất: *K={k1, k2, k3, …, kM} = {1, …, M}*
* Xác suất chuyển trạng thái ban đầu:
* Ma trận xác suất chuyển trạng thái:
* Xác suất sinh ký hiệu:
* Dãy trạng thái:
* Dãy xuất:

1. Các bài toán với mô hình Markov ẩn

Bài toán đánh giá: cho trước chuỗi quan sát tương ứng với mô hình , xác định xác suất tương ứng có điều kiện .

Bài toán so khớp: cho trước chuỗi quan sát tương ứng với mô hình , xác định chuỗi trạng thái nghiên cứu tương ứng sao cho xác suất tương ứng đạt cực đại.

Bài toán huấn luyện: cho trước chuỗi quan sát , xác định các tham số phù hợp cho mô hình sao cho xác suất tương ứng co điều kiện là cực đại.

# Thực nghiệm

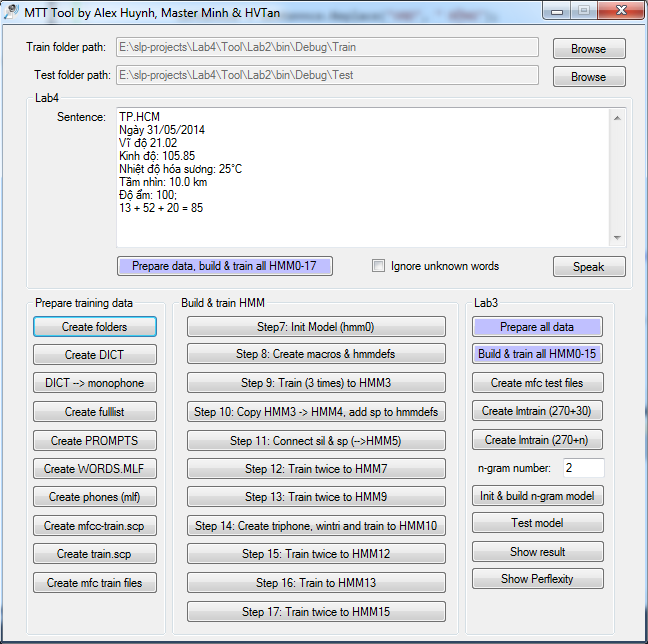
1. Dữ liệu thực nghiệm

Dữ liệu huấn luyện là 300 tập tin wav (mỗi tập tin có độ dài 2-4 giây). Mỗi tập tin wav đi kèm với một tập tin txt ghi nhận lại câu đã nói.

Mục tiêu của chương trình là phát ra tiếng nói dựa trên một câu do người dùng nhập vào.

Công cụ sử dụng chính là HTK (Hidden Markov Model Toolkit).

Ngoài ra để thuận tiện cho việc tạo lập các tập tin chính, nhóm đã viết một công cụ hỗ trợ MTT Tool bằng ngôn ngữ C# để thao tác nhanh hơn. Công cụ này có thể thực hiện tự động tạo ra các tập tin cần thiết, chạy tất cả các lệnh HTK để cuối cùng phát ra tiếng nói.



1. HTK

HTK cung cấp các công cụ cho việc xây dựng một mô hình Markov ẩn (HMMs). HMMs có thể dùng cho các bài toán với dữ liệu là một chuỗi theo thời gian, và HTK cũng có thể sử dụng cho các bài toán dạng này. Tuy nhiên, HTK được thiết kế chính là phục vụ cho các bài toán về tiếng nói. Dưới đây là một vài công cụ cơ bản thường được dùng:

* HCopy: copy và thức hiện việc rút trích đặc trưng tùy thuộc vào file cấu hình.
* HCompV: công cụ khởi tạo tham số cho các hàm phân bố xác suất có trong từng trạng thái của mô hình Markov ẩn tương ứng với dữ liệu huấn luyện.
* HRest: công cụ huấn luyện mô hình bằng việc tính toán, ước lượng lại các tham số của mô hình.
* HParse: công cụ chuyển đổi từ dữ liệu đầu vào thành dữ liệu theo format HTK.
* HVite: công cụ so khớp/nhận dạng với đầu vào là dữ liệu cần nhận dạng, và trả về mô hình tốt nhất/khớp nhất mà nó thuộc vào.
* HResults: công cụ thống kê, đánh giá kết quả nhận dạng

1. Phương pháp thực hiện
2. tạo từ điển phiên âm DICT

* Từ điển này giúp máy biết cách phát âm từng từ.
* Sử dụng công cụ MTT Tool để tạo.

1. tạo PROMPTS

* Lưu đường dẫn và câu tương ứng.
* Sử dụng công cụ MTT Tool để tạo.

1. tạo WORDS.MLF

* Danh sách các từ trong một câu và nhãn của chúng.
* Sử dụng công cụ MTT Tool để tạo.

1. tạo monophones0 và monophones1

* Thống kê các từ.
* Sử dụng công cụ MTT Tool để tạo.
* Ở hai tập tin đều có sil. monophones1 khác với monophones0 là có thêm sp ở cuối tập tin. sp là khoảng ngưng giữa các từ, sil là khoảng lặng ở đầu/cuối tập tin ghi âm.

1. tạo phones0.mlf và phones1.mlf

* Thống kê các từ.
* Dùng lệnh HTK:

*HLEd -l \* -d dict -i phones0.mlf mkphone0.led words.mlf*

*HLEd -l \* -d dict -i phones1.mlf mkphone1.led words.mlf*

* mkphones0.led và mkphones1.led là hai tập tin cho trước dùng để cấu hình, có các từ khóa EX (expand), IS (thêm khoảng lặng đầu cuối), DE (xóa tất cả khoảng nghỉ).

1. tạo mfcc.scp, rút trích đặt trưng

* Danh sách tập tin wav và tên mfc tương ứng.
* Sử dụng công cụ MTT Tool để tạo.
* Rút trích đặc trưng bằng lệnh HTK:

*HCopy -T 1 -C config.hcopy -S mfcc.scp*

* HTK sẽ tạo ra các tập tin \*.mfc trong thư mục MFC.

1. khởi tạo mô hình

* Khởi tạo mô hình bằng lệnh HTK:

*HCompV -C config -f 0.01 -m -S train.scp -M hmm0 proto*

* proto là tập tin được cung cấp trước cho biết rằng có 39 đặc trưng MFCC\_0\_D\_A.
  + 0: 12 MFCC features và (1) total energy in the frame
  + D: 13 “Delta coefficients” (đạo hàm bậc 1 của MFCC)
  + A: 13 “Acceleration coefficients” (đạo hàm bậc 2 của MFCC)
  + Mỗi một âm tiết là một mô hình Markov ẩn.
  + Số trạng thái: 5.

1. Từ hmm0, tạo mô hình chuẩn

* Tạo tập tin macros: kết hợp đoạn ~o của tập tin hmm0\proto và ~v của tập tin hmm0\ vFloors.
* Tạo tập tin hmmdefs: với mỗi dòng trong monophones0, kết hợp thêm đoạn <BEGINHMM> và <ENDHMM> của hmm\proto.
* macros và hmmdefs là mô hình chuẩn.

1. Training 3 vòng để ra hmm3

* Huấn luyện bằng lệnh HTK:

*HERest -C config -I phones0.mlf -S train.scp -H hmm0/macros -H hmm0/hmmdefs -M hmm1 monophones0*

*HERest -C config -I phones0.mlf -S train.scp -H hmm1/macros -H hmm1/hmmdefs -M hmm2 monophones0*

*HERest -C config -I phones0.mlf -S train.scp -H hmm2/macros -H hmm2/hmmdefs -M hmm3 monophones0*

* macros và hmmdefs đã được sửa lại, lưu vào hmm3.

1. Thêm sp vào danh sách mô hình

* Tạo hmm4, copy mô hình của hmm3 vào hmm4.
* Sửa hmmdefs trong hmm4: bổ sung sp, chỉnh sửa NUMSTATES thành 3, chỉ giữ lại state 2, TRANSP xóa bỏ 2 cột và 2 dòng cuối.

1. Liên kết sil và sp

* Tạo hmm5, thực hiện lệnh HTK:

*HHEd -H hmm4/macros -H hmm4/hmmdefs -M hmm5 sil.hed monophones1*

1. Training 2 vòng để ra hmm7

* Huấn luyện bằng lệnh HTK:

*HERest -C config -I phones0.mlf -S train.scp -H hmm5/macros -H hmm5/hmmdefs -M hmm6 monophones1*

*HERest -C config -I phones0.mlf -S train.scp -H hmm6/macros -H hmm6/hmmdefs -M hmm7 monophones1*

1. Training 10 vòng để ra hmm17

* Huấn luyện bằng lệnh HTK:

*HERest -C config -I phones0.mlf -S train.scp -H hmm7/macros -H hmm7/hmmdefs -M hmm8 monophones1*

* Dùng câu lệnh tương tự, huấn luyện lên hmm17.

1. Nhận dạng tập huấn luyện

* Chạy lệnh HTK:

*HVite -l \* -o S -b silence -C config\\config -a -H hmm17/macros -H hmm17/hmmdefs -i recout.mlf -t 250.0 -I words.mlf -S train.scp dict monophones1*

* Kết quả có được tập tin recout.mlf, ghi lại vị trí của từng từ (nano giây) trong từng tập tin âm thanh.

1. Tổng hợp tiếng nói

* Chuẩn hóa câu văn bản đầu vào (lớp *Standardizer*), sau đó chuyển về dạng Telex (lớp *WordConversion*).
* So sánh từng từ của câu văn bản với từ trong recout.mlf, dựa vào từ đứng trước và từ đứng sau (lớp *MyWord*).
* Ghép âm thanh tất cả các từ lại và phát ra.

1. Các trường hợp chuẩn hóa dữ liệu đầu vào

Được cài đặt chi tiết trong lớp Standardizer.

* 1. Các ký hiệu đặc biệt

Xử lý được các ký hiệu đặc biệt như: *dấu phẩy, dấu hai chấm, dấu chấm phẩy, dấu mở ngoặc tròn, dấu đóng ngoặc tròn, dấu mở ngoặc vuông, dấu đóng ngoặc vuông, dấu chấm cuối câu.*

* 1. Các ký hiệu toán học

Xử lý được các ký toán học như: *@, #, %, +, ^, =, ->*.

* 1. Tên địa danh

Xử lý được các tên địa danh viết tắt như: *BRVT, TPHCM, TP.HCM, DNA, ĐNA*.

* 1. Khoảng cách

Xử lý được các ký hiệu khoảng cách như: *km, nm, ft*.

* 1. Nhiệt độ

Xử lý được các ký hiệu nhiệt độ như: *°F, °C*.

* 1. Tiền tệ

Xử lý được các ký hiệu tiền tệ như: *$, USD, €, EURO, £, GBP, penny, pence, ¥, JPY, VND, VNĐ*.

* 1. Số

Xử lý được số nguyên và số thập phân.

* 1. Ngày tháng

Xử lý được ngày tháng dạng: *ngày/tháng/năm*.

1. Nhận xét

* Chương trình thỏa mãn các yêu cầu cơ bản về tổng hợp tiếng nói.
* Xử lý được nhiều ký hiệu phổ biến.
* Chất lượng tiếng nói tổng hợp ra còn thấp một phần do giọng phát âm chưa rõ ràng, một phần do dữ liệu huấn luyện rất ít, nên âm thanh chưa thể phủ hết các từ trong câu nhập vào.
* Chưa xử lý được các từ nước ngoài.

# Tài liệu tham khảo

Tiếng Anh:

1. Daniel Jurafsky & James H. Martin, *Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition*, 2006, pp 192-236.
2. Huỳnh Diệp Tân, et al., *Mô hình Markov*, 2010
3. Nguyễn Bá Công, *Nhận dạng tiếng nói rời rạc*, 2009
4. Steve Young et al., *The HTK Book*, 2009

Website:

1. <http://en.wikipedia.org/wiki/Hidden_Markov_model>
2. <http://htk.eng.cam.ac.uk/>