#### 1311055 \_ Đậu Phạm Hữu Tài

Hướng dẫn chạy HTK để nhận dạng tiếng nói

Các bước cơ bản

Lời nói đầu: mình thấy các bạn làm vất vả theo HTK book và file hướng dẫn của anh Phong đã hơn chục năm, phát sinh nhiều lỗi, nay mình tổng hợp theo kinh nghiệm chạy rất nhiều rất nhiều lần của mình để soạn lại tài liệu nhằm giúp các bạn không mắc lỗi trong quá trình chạy. Version hôm nay là 1.2\_06/05/2014, tài liệu sẽ còn chỉnh sửa tiếp trong tương lai khi các bạn có thắc mắc. Đính kèm tài liệu là labsimple.zip, được co rút từ folder simple trong tài liệu của các thầy thôi.

Đậu Tài - 06/05/2014

Chúc cẩn thận và may mắn :-D

# Chuẩn bị

* Tạo thư mục “htklab” để chứa project của bạn
* Copy các thư mục sau từ “labsimple.zip” (đính kèm) vào thư mục “htklab”:
  + “bin32”: các lệnh thực thi của HTK.
  + “cfg”: các file config dùng cho các lệnh của HTK.
  + “scripts”: các mã perl để tạo một số file tự động thay vì làm thủ công.
* Tạo thư mục “wav/train”: các file âm thanh chuẩn bị cho việc training.
* Mở Command Prompt để chạy lệnh chuyển đến thư mục “htklab”.

# Thực hiện

Ghi chú: các file được khuyến cáo copy từ “labsimple.zip” chỉ là file mẫu, các bạn theo format đó mà sửa theo dữ liệu của mình, khuyến khích dùng Notepad++.

## 1. Training

**Bước 1**

Tạo file gram.txt với nội dung sau (có thể copy từ “labsimple.zip” vào “htklab”)

Lưu ý: lưu encode dạng ANSI, nếu lưu Unicode hoặc UTF-8 sẽ bị lỗi

$speaker = tai | minh;($speaker)

**Bước 2**

Tạo file wdnet bằng lệnh dưới đây:

>bin32\HParse gram.txt wdnet

**Bước 3**

Tạo file dict với nội dung sau (có thể copy từ “labsimple.zip” vào “htklab”):

“tai [phuong] t a i

minh [toan] m i nh

silence sil

“

Lưu ý: có dấu enter cuối file và nhớ sắp xếp theo thứ tự anphabet

**Bước 4**

Tạo file thu âm.

>bin32\HSGen -l -n 1000 wdnet dict >train

1000 là số file ghi âm sẽ phải thu sau (thứ tự random)

Sau đó tiến hành thu âm theo mẫu trong file train, mỗi dòng là 1 file .wav, lưu hết vào thư mục wav/train.

**Bước 5**

Tạo file “listwavmfccTrain.scp”, danh sách file âm thanh để rút trích đặc trưng MFCC và các file MFCC tương ứng.

Chạy script perl để tạo file tự động từ danh sách file âm thanh trong thư mục “wav\train”.

>scripts\listwavmfc.pl wav\train listwavmfccTrain.scp

**Bước 6**

Trích đặt trưng MFCC

>bin32\HCopy -T 1 -C cfg\HCopy.cfg -S listwavmfccTrain.scp

Lúc này trong thư mục “wav\train” sẽ có các file mfc tương ứng với các file wav.

**Bước 7**

Tạo file train.scp, danh sách các file mfc vừa tạo ở bước 6.

Chạy script perl sau để tạo tự động (có thể tạo bằng tay)

>scripts\mkTrainFile.pl wav\train train.scp

**Bước 8**

Tạo thư mục hmm\hmm0

Tạo thư mục hmm\hmm1

Tạo thư mục hmm\hmm2

…

Đến thư mục 22

Copy file proto mẫu vào thư mục hmm\hmm0

Có thể tạo file proto có nội dung theo mẫu của htkbook lưu với encoding là ANSI (nếu ko sẽ lỗi), (file proto mẫu tạo sẵn có đính kèm thèo tài liệu này).

Lưu ý các dòng 0.0 và 1.0 có số 39 tức là yêu cầu copy ra 39 số 0.0 (hoặc 1.0) ở dấu … (thay dấu … bằng 36 số 0.0 hoặc 1.0)

**Bước 9**

Khởi tạo giá trị huấn luyện

>bin32\HCompV -C cfg\HCompV.cfg -f 0.01 -m -S train.scp -M hmm\hmm0 hmm\hmm0\proto

**Bước 10**

Tạo file lable.mlf

>scripts\prompts2mlf.pl lable.mlf train

Tạo file wlist

>scripts\prompts2wlist.pl train wlist

**Bước 11**

Tạo file mkphone0.led với nội dung :

“EX

IS sil sil

DE sp

”

Đặt ở đâu thì nhớ lấy để bước sau sẽ lấy

Lưu ý: có dấu enter

**Bước 12**

Tạo file phones0.mlf

>bin32\HLed -l '\*' -d dict -i phones0.mlf mkphone0.led lable.mlf

Tạo xong mở file phones0.mlf lên và sửa nội dung trong ngoặc kép thành đường dẫn đầy đủ

Ví dụ:

"'\*'/1..lab" ----->> “C:\Users\Bean\Desktop\htklab\wav\train\01.lab”

**Bước 13**

Tạo file lexicon với nội dung sau (có thể copy từ “labsimple.zip” vào “htklab” nhưng nhớ bỏ dòng thứ 3 “silent …” gì đó đi):

Hoặc tạo file lexicon với nội dung sau:

tai [tai] t a i

minh [minh] m i nh

nhớ xếp theo thứ tự anphabet

**Bước 14**

Tạo file monophones0

>bin32\HDMan -A -D -T 1 -m -w wlist -n monophones0 -i -l dlog dict lexicon

Báo lỗi 1 missing, ko vấn đề gì

Rồi sắp xếp lại monophones0 theo thứ tự anphabet

**Bước 15**

Tạo file hmmdefs bằng mã perl

>scripts\mkHmmdefsFile.pl hmm\hmm0\proto monophones0 hmm\hmm0\hmmdefs

**Bước 16**

Tạo file macros bằng mã perl

>scripts\mkMacrosFile.pl hmm\hmm0\vFloors hmm\hmm0\macros

**Bước 17**

Sửa file phone0.mlf như sau:

* Bỏ nguyên dòng chứa sil
* Sửa dấu “\” thành “/”

Ước lượng thông số

>bin32\HERest -C cfg\HERest.cfg -I phones0.mlf -t 250.0 150.0 1000.0 -S train.scp -H hmm\hmm0\macros -H hmm\hmm0\hmmdefs -M hmm\hmm1 monophones0

>bin32\HERest -C cfg\HERest.cfg -I phones0.mlf -t 250.0 150.0 1000.0 -S train.scp -H hmm\hmm1\macros -H hmm\hmm1\hmmdefs -M hmm\hmm2 monophones0

Tiếp như vậy tới hmm22.

Ghi chú: các thông số chưa nắm rõ, cần xem lại trong htkbook

**Training cơ bản đã xong, bây giờ là test kết quả**

## 2. Test kết quả

Tạo thư mục “kq”

Copy file HmmList từ “labsimple.zip” vào “htklab”

Mở HmmList lên và bỏ dòng chữ “sil” lưu lại (nếu có)

Hoặc có thể tạo HmmList như sau:

Ví dụ:

"a"

"i"

"t"

Thực hiện thu âm và rút trích đặc trưng mfc của các file thu âm test như sau:

Tạo thư mục wav/test, bỏ các file thu âm wav cần test vào đó.  
Tạo listwavmfccTest.scp

>scripts\listwavmfc.pl wav\test listwavmfccTest.scp

Trích đặt trưng MFCC

>bin32\HCopy -T 1 -C cfg\HCopy.cfg -S listwavmfccTest.scp

Tạo file test.scp:

>scripts\mkTrainFile.pl wav\test test.scp

Xong chạy test:

>bin32\HVite -H hmm\hmm3\macros -H hmm\hmm3\hmmdefs -S test.scp -i kq\recout.mlf -w wdnet -p 0.0 -s 5.0 dict HmmList

## 3. Đánh giá kết quả test

Tạo file phones0test.mlf có nội dung sau (có thể tham khảo từ labsimple.zip)

“#!MLF!#

"F:/labsimple/mfcc/test/41.lab"

toan

.

"F:/labsimple/mfcc/test/42.lab"

phuong

.

“

Rồi thực hiện lệnh:

HResults -I phones0test.mlf Hmmlist kq\recout.mlf

Lấy file recout.mlf ra đọc và thấy kết quả :D