7. gyakorlat

“Flat start” tanítás HTK-ban

A számfelismeréses feladatnál rövid tanítófájlok voltak, mindegyikben egy-egy szóval. Ráadásul minden tanító hangfájl fonetikai szinten fel volt szegmentálva és címkézve, azaz meg volt adva minden egyes hang helye és típusa. Folytonos felismerési feladatnál általában a tanító fájlok is hosszabbak (egy vagy több mondatot tartalmaznak), és fonetikai átiratunk sincsen, csak ortografikus (betű szerinti). Megnézzük, hogy ilyen esetben hogyan történik a tanítás a HTK-ban. Nyilván főleg a kezdeti lépések fognak eltérni a korábban látottaktól, ezt hívja a HTK „flat start” tanításnak.

A tanító adatbázis: egy hangoskönyvet fogunk használni, amelyben Krúdy Szindbád-történeteit olvassa fel egy színész. Az eredeti hangoskönyv 3 CD-ből állt, mindegyiken 3-4 történettel. Mi most csak az 1. CD 1. track-jén fogunk tanítani, ezt 8 darabra kellett vágnunk (túl nagy volt a tanításhoz). Tesztelni a 3. CD 3. track-jén, valamint a 2. CD 4. track-jén fogunk (ezeket nem vágtam szét kisebb darabokra).

Erőforrások:

- mfc jellemzőfájlok az mfc/ folderben

- szószintű átiratok az ort/ folderben (megj: írásjelek nélkül)

- kiejtési szótár a tanító adatbázis szavaihoz: network/szindbadall.dic

- kiejtési szótár a teszteléshez: network/krudy62962.dic (Krúdy összes művei - teszt)

- nyelvi modell: network/trigram.lattice (statisztikai modell Krúdy összesből - teszt)

Eltérések a számoknál látott tanításhoz képest:

1. Nincs fonetikai átiratunk – erre a kiejtési szótár alapján fogjuk becslést adni

2. Nincsenek adva beszédhang-határok: ezért beágyazott tanítást fogunk csinálni, flat start inicializálással, sok újraillesztéssel

3. A szavak között lehetnek szünetek: „short pause” modell fogunk rakni a szavak végére, a HTK kézikönyv által javasolt módon

Néhány technikai eltérés:

1. a lab file-okat egyben fogjuk kezelni. Erre a HTK-nak van egy egyszerű formátuma, az ún. „master label file”, .mlf kiterjesztéssel

2. A hmm modelleket is egy fájlban fogjuk tárolni, ún. hmm „makró” fájlban

3. A felismeréshez a hvite.exe helyett a hdecode.exe parancsot fogjuk használni, mert ez gyorsabb. A paraméterezése majdnem ugyanaz, de

- Csak trifón akusztikus modellel és statisztikai nyelvi modellel működik

- A felvétel eleji és végi csendet kötelező sil-nek hívni, a szavak köztit pedig sp-nek

- a szótárba nem kell beleírni a szóvégi sp-ket (ezért van két változat a szótárból: krudy62962.dic és krudy62962-sp.dic)

És akkor lássuk (a parancsok a flattrain.lepesek fájlban is megtalálhatók):

1. Az ort fájlokat összefűzzük egyetlen mlf fájllá (az mlf fájl formátuma egyszerű: gyakorlatilag konkatenálni kell a fájlokat, az elején a fájlnévvel, a végén egy ponttal):

- concatlabfiles.vbs trainlist.lab trainall-word.mlf

2. A szószintű átiratból megpróbáljuk kitalálni a fonetikai átiratot. Ehhez

a) feltesszük, hogy minden fájl elején és végén csend (sil) van

b) a szavak fonetikai átiratát a kiejtési szótárból helyettesítjük be

c) Mivel a szavak közt ritkábban tartunk szünetet, mint amikor nem. ezért egyelőre jobb híján feltesszük, hogy SEHOL sincs szünet a szavak közt

A szükséges átalakítást az alábbi parancs végzi el.

- HLEd.exe -l ../mfc -d ..\network\szindbadall.dic -i trainall-phone.mlf makephones0.led trainall-word.mlf

trainall-word.mlf: a szószintű átirat

trainall-phone: a fonetikai átirat (kimenet)

szindbadall.dic: a kiejtési szótár

makephones0.led: ebben az átalakítási parancsok:

EX – exchange (szavak 🡪fonetikai átiratok)

IS sil sil – insert (sil a fájl elejére-végére)

DE sp – delete sp (egyelőre kihagyjuk a szóvégi szüneteket)

„Flat start” beszédhang-modelleket készítünk. Ehhez Gauss-eloszlást illesztünk az ÖSSZES tanítóvektorra, és ezzel inivializáljuk az összes modell össze állapotában a kibocsátási valószínűséget.

3. Globális statisztika számítása az össezs trainfájlból:

- HCompV.exe -T 1 -m -C nobyteswap.config -S trainlist.mfc -f 0.01 -M . -o new ..\hmm\proto3st1g

proto3st1g: prototípus modell (3 állapot, 1 gauss), ennek mintájára készül a „flat” modell

new: a kimenő flat modell

-M: a kimenő modell helye (az aktuális könyvtár)

-f: kimenetként létre fog jönni egy vFloors nevű fájl is, ebben valamiféle küszöbök vannak kapott szórásértékekre, de a pontos működését nem tudom (valószínűleg segít kiszűrni a gyanúsan kicsit szórásokat, ami hibás működést eredményezne)

Nézzünk bele a kapott „flat” modellbe, láthatjuk hogy mindhárom állapothoz ugyanazok az eloszlási paraméterek tartoznak!

4. Létrehozzuk az összes beszédhang modelljét; ehhez gyakorlatilag lemásoljuk a flat modellt minden beszédhang esetére

- BuildFlatModel.vbs new vFloors ..\hmm\hmmlist52 ..\hmm\hmm0\newMacros

hmmlist52: beszédhangok listája

hmm0\newMacros: ebbe a fájlba kerülnek a modellek (a könyvtárnak léteznie kell!)

5. A fent kapott kiindulási modellekkel elindulhat a beágyazott tanítás:

- HERest.exe -T 1 -C nobyteswap.config -I trainall-phone.mlf -S trainlist.mfc -H ..\hmm\hmm0\newMacros -M ..\hmm\hmm1 ..\hmm\hmmlist52

trainall-phone.mlf: tanító címkefálj

trainlist.mfc: tanító jellemzőfájlok listája

hmm0\newMacros: kiindulási modell

hmm1: ide kerül az új modell (a könyvtárnak léteznie kell!)

hmmlist52: a beszédhangok (betanítandó modellek) listája

A fenti parancsot ismételjük meg többször (mondjuk négyszer), persze a kiindulási modell könyvtárát és a kimeneti modell könyvtárát mindig átírva (hmm0🡪hmm1, hmm1🡪hmm2, hmm2🡪hmm3, hmm3🡪hmm4, …)

Figyeljük meg, hogy a tanítás során a logprob értékek növekednek

6. Most, hogy a modelljeink már valamennyire betanultak, létrehozzuk a rövid szünet (short pause, sp) modellt. Mivel a rövid szünet spektrálisan pont úgy néz ki, mint a hosszú, ezért a sil modellből fogjuk átmásolni a kibocsátási paramétereket.

Az sp modell létrehozására a HTK nem ad eszközt, kézzel kell csinálnunk ☹

a) a hmm4/newModel fájlt másoljuk át a hmm5 könyvtárba, majd nyissuk meg szövegszerkesztőben.

b) Keressük meg a „sil” modellt, és másoljuk át a fájl végére (az ábrán jobb oldalon).

c) A másolat módosításával kaphatjuk meg az sp modellt (bal oldalon, a különbségek pirossal).

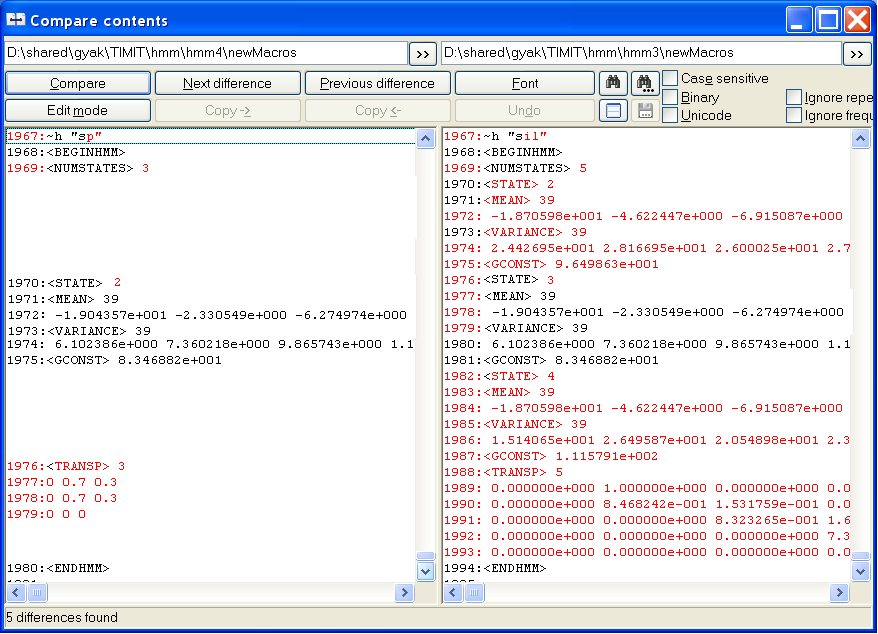
Írjuk át a nevet (sil helyett sp)

Az állapotok száma 5 helyett 3 lesz

Töröljük ki a 2 és 4 állapotokat

State 3-at módosítsuk State 2-re

Csináljunk új állapotátmenet-mátrixot 3 állapottal, az ábrán látható értékekkel



7. A kapott új modellekkel kényszerített illesztést végzünk a HVite paranccsal a tanítófájlokon. Ez majdnem ugyanaz, mint egy felismerés, de ebben az esetben megadjuk a helyes átiratot, tehát az illesztés során a Viterbi-algoritmus csak a határok tologatásával tud játszani)

-HVite.exe -T 1 -o W -b sil -C nobyteswap.config -a -H ..\hmm\hmm5\newMacros -i trainall-realigned.mlf -m -t 250.0 -I trainall-word.mlf -S trainlist.mfc ..\network\szindbadall.dic ..\hmm\hmmlist52sp

trainall-word.mlf: az illesztett szószintű átirat

trainall-realigned.mlf: itt kapjuk meg az újraillesztett fonetikai címkéket

-a: aligment mód, tehát nem kell nyelvi modell

-o W: szószintű címkéket nem kérünk a kimenetben, csak beszédhang-szintűeket

-b sil: a mondatok elején-végén sil van

-t 250.0: beam width, ld. Viterbi beam search az előadáson

hmmlist52sp: modellek listája – vigyázat, most már sp modellünk is van!

Hasonlítsuk össze a trainall-phone.mlf és a trainall-realigned címkéit! Most már vannak sp modellek minden szó végén, de legtöbbször 0 hosszúságúak.

14600000 14900000 A -269.687775

14900000 14900000 **sp** -1.203973

14900000 15200000 h -246.945969

15200000 15500000 A -252.145432

15500000 15800000 j -243.959793

15800000 31500001 oo -10149.295898

31500001 35800001 sz -3460.163818

35800001 36600001 **sp** -538.669373

8. Ezután persze tanácsos újabb HeREst köröket futtatni, hogy – most már az opcionális szóközi csendekkel – tovább tanítsuk a modelleket. A HeRest parancs majdnem változtatás nélkül használható, de:

- Ne felejtsük el, hogy most már a kibővített hmm-listát (hmmlist52sp) és az újraillesztett címkefájlt (trainall-realigned.mlf) kell használnunk!

- Illetve fog kelleni egy -X rec kapcsoló, mert a hvite „.lab” helyett „.rec” kiterjesztéssel mentette el az újraillesztett címkefájlokat.

Én csak egy HeRest kört mentem az alábbi paranccsal:

HERest.exe -T 1 -X rec -C nobyteswap.config -I trainall-realigned.mlf -S trainlist.mfc -H ..\hmm\hmm5\newMacros -M ..\hmm\hmm6 ..\hmm\hmmlist52sp

Ezután jöhet a trifón modellek létrehozása, majd a gauss-komponensek számának növelése…

8. gyakorlat

Trifón modellek létrehozása és tanítása

A parancsok a trifon.lepesek fájlban is megtalálhatók!

1. A trainall-realigned fáj címkéit trifónokká konvertáljuk az alábbi paranccsal:

HLed -n traintriphones -l ../mfc -i trainall-tri.mlf mktri.led trainall-realigned.mlf

trainall-realigned.mlf: az átalakítandó monofón címkefájl

trainall-tri.mlf: az új, trifón címkefájl

mktri.led: a konvertálást elvégző parancsok vannak benne

traintriphones: egyúttal kigyűjti az össze előforduló trifónt ebbe a listába

Hasonlítsuk össze a trainall-realigned.mlf és a trainall-tri.mlf fájlokat!

Láthatjuk, hogy

1. minden hang trifónná lett alakítva, kivéve a sil és sp hangokat
2. a trifónok „átmennek” az sp-n, azaz az sp bal-jobb szomszédként sem szerepel.Ez azon a feltevésen alapul, hogy az sp nagyon rövid szüneteket jelöl, amiken a koartikuláció átlép.

2. Legyártjuk az összes lehetséges trifón listáját az alábbi paranccsal. Erre a listára majd a döntési fa építésénél lesz szükség, mivel a tanító adatbázisban nem szereplő trifónokra is meg kell tudni majd mondani, hogy melyik modellel helyettesítsük a felismerés során.

alltriphones.vbs ..\hmm\hmmlist52 alltriphones

Láthatjuk, hogy az sp nélküli listából indulunk ki, mivel az sp nem szerepel trifónokban.

A fájlméretekből az is látszik, hogy az összes lehetséges trifón listája jóval bővebb a tanítóadatban előforduló trifónok listájánál.

3. A következő lépés a monofón modellek trifón modellekké konvertálása lesz. Ehhez előkészítésként legyártjuk a mktri.hed fájlt. Ebben a modellek klónozására vonatkozó parancs kerül, illetve a létrehozás után rögtön össze is kapcsoljuk minden hang összes trifón változatának az állapotátmenet-mátrixait (csak a kibocsátási eloszlásokra nézve akarunk trifónokat képezni).

maketrihed.vbs ..\hmm\hmmlist52 traintriphones mktri.hed

5. A legutóbbi órán a hmm\hmm6 könyvtárban hagytuk a legutolsó monofón modellünket. Ezt most másoljuk át a hmm\hmm-tri0 könyvtárba, ugyanis a következő prancs felül fogja írni a fájlt.

6. Most végre létrehozzuk a trifón modelleket. Kezdeti értékként mindegyik trifón a megfelelő monofón modell paramétereit örökli:

HHed -C nobyteswap.config -H ..\hmm\hmm-tri0\newMacros mktri.hed ..\hmm\hmmlist52sp

7. Végezzünk legalább egy tanítást a trifón modelleinken a szokásos HERest paranccsal!

HERest -T 1 -t 500.0 -m 0 -C nobyteswap.config -S trainlist.mfc -I trainall-tri.mlf -H ..\hmm\hmm-tri0\newMacros -M ..\hmm\hmm-tri1 -s stats traintriphones

hmm-tri0\newMacros: ez a kiindulási modellünk

hmm-tri1\newMacros: ide kerül az új modell

trainall-tri.mlf: a trifón címkéket használjuk a tanításhoz

traintriphones: a hmmlist52sp lista helyett most már a trifónok listája kell!

-s stats: csinál egy statisztikát a modellekről. Belenézve láthatjuk, hogy a legtöbb trifón csak egyszer fordul elő.

8. Másoljuk át a hmm\hmm-tri1\newMacros modellt a hmm\hmm-tri2\ könyvtárba, mert a következő parancs megint helyben fog dolgozni, felülírja az eredeti fájlt.

9. Most következik végre a döntési fa alapú állapotkapcsolás! Ezt az alábbi parancs végzi el:

HHed -H ..\hmm\hmm-tri2\newMacros tree.hed traintriphones

A tree.hed fájlban találhatók a döntési fa építését vezérlő kérdések, és az összevonási parancsok. A fa méretét szabályozó paraméteek közül az egyik az R0, itt lehet megadni a minimális példaszámot, ami alatt egy állapot garantáltan össze lesz vonva (jelen esetben 20). A másik a TB parancs, az utána álló szám adja meg a továbbosztáshoz elvárt minimális likelihood-csökkenést (jelen esetben 300).

Jelen esetben az eredeti 9009 különböző állapotból 440 kapcsolt állapotot tart meg, a modellfájl mérete is töredékére csökken.

Előáll továbbá egy „tiedlist”, ezt kell majd használni a felismerés során során a hmmlist helyett. Ez a fájl egyúttal azt is megadja, hogy a tanulá során nem látott trifónokat melyik modellel kell helyettesíteni felismeréskor: az egy sorban levő modellek közül az első a létező modell, a mellette levők pedig a vele helyettesítendő modellek.

Ezután következne a gauss-komponensek számának növelése. A HHEd megfelelő parancsával lehet újabb Gauss-komponenseket hozzáadni a modellekhez. A kézikönyv javaslata szerint érdemes fokozatosan egyszerre csak néhány Gauss-komponenst hozzávéve bővíteni a modelleket, minegyik után újabb tanítási iterációkat futtatva. Ezt nem fogjuk végigcsinálni, mivel elméleti szempontból nem túl érdekes, és sok ideig tart. Viszont felraktam egy 28 Gauss-komponenst tartalmazó, a teljes adatbázison tanított modellt hmm\teljesmodell-28-4 néven (a Gauss-komponensek számát 28-ig vittem fel kettesével, mindegyik bővítés után 4 iterációt tanítva). Kipróbálni a runhvitew.bat paranccsal, kiértékelni a runhresults.bat paranccsal tudjuk.