Pokročilé metody rozpoznávání řeči

Přednáška 6

Další úlohy analýzy řeči a metody jejich řešení

Časové zarovnání textového přepisu (1)

Úloha:

K dispozici je nahrávka řeči a její přepis (automatický, ruční, stenozáznam, atd.) Jak k sobě přesně přiřadit audio a text?

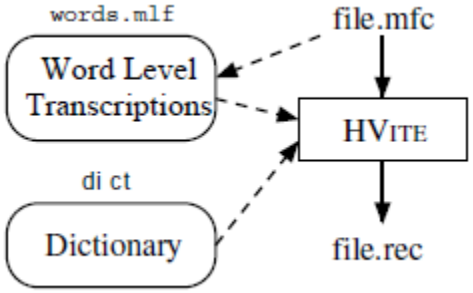
Řešení: „Vynucené zarovnání“ (**Forced Alignment**)

Idea:
Donutit rozpoznávač, aby pracoval pouze se slovy obsaženými v textu a pouze v tomto pořadí.

Časové zarovnání textového přepisu (2)

Řešení v HTK (HTKbook str. 207)

- Je potřeba slovník, soubor s přepisem ve formátu MLF, zparametrizovaný audiosoubor



Úlohu vyřeší program HVite se switchem –a

Výsledkem jsou přesné časové značky u každého slova

```
 7500000  8700000 f   -1081.604736 FOUR 30.000000
 8700000  9800000 ao   -903.821350
 9800000 10400000 r    -665.931641
10400000 10400000 sp     -0.103585
10400000 11700000 s   -1266.470093 SEVEN 22.860001
11700000 12500000 eh   -765.568237
12500000 13000000 v    -476.323334
13000000 14400000 n   -1285.369629
14400000 14400000 sp     -0.103585
```

Časové zarovnání textového přepisu (3)

Použití:

1. **Tvorba „prolinkovaných“ multimediálních databází**
2. **Full-text search v multimediálních databázích**
   (hledá se v textu a z textu vedou časové značky do audio stopy)
3. **Upřesňování výslovnosti v přepisech**
   máme-li slovník s více výslovnostmi u některých slov, rozpoznávač si v režimu nuceného rozpoznávání vybere tu nejsprávnější (použití při upřesňování trénovací databáze)
4. **Nalezení pozic jednotlivých fonémů v nahrávce**
   (např. pro účely vývoje fonémového TTS systému, trénování neuronových sítí, atd.)

Fonémový rozpoznávač (1)

Úloha:

Zjistit nejpravděpodobnější sekvenci fonémů v nahrávce řeči

Účel:

- identifikovat/upřesnit fonémy v přepisu nahrávky
- pokus o fonetický přepis promluvy v cizím jazyce
- metoda (omezeně) použitelná pro vyhledávání klíčových slov

Idea řešení:

Slovník je tvořen všemi fonémy (+ticho a hluky), gramatika umožňuje libovolný přechod mezi nimi

Fonémový rozpoznávač (2)

Problém:

rozpoznávač mívá tendenci vkládat více hlásek, často opakovaně za sebou (místo „a je to“ může rozpoznat „aaa jee too, místo šumů v tichu některé hlásky, atd.)

Řešení:

V HVite specifikovat parametry
-s (násobitel vlivu jazykového modelu)
-p (penále za každé rozpoznané slovo/znak)

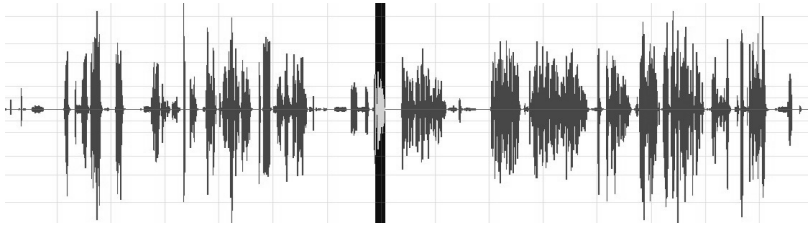
Přesnost fonémového rozpoznávače:

- výrazně nižší než při rozpoznávání řeči (schází kontext)
- cca 60 až 70 % (u nejnovějších systémů s NN vyšší)

Detekce klíčových slov (1)

Úloha:

Nalézt (*pokud možno rychle*) pouze vybraná klíčová slova v záznamu řeči – angl. zkratka KWS (Key-Word Spotting)



Řešení:

1. **Provést rozpoznávání (s dostatečně rozsáhlým slovníkem)** a u každého slova určit časové značky. Pak provést full-search a najít časovou pozici hledaných slov.

   Nevýhody: pomalé, výpočetně náročné,

   hledaná slova nemusí být ve slovníku

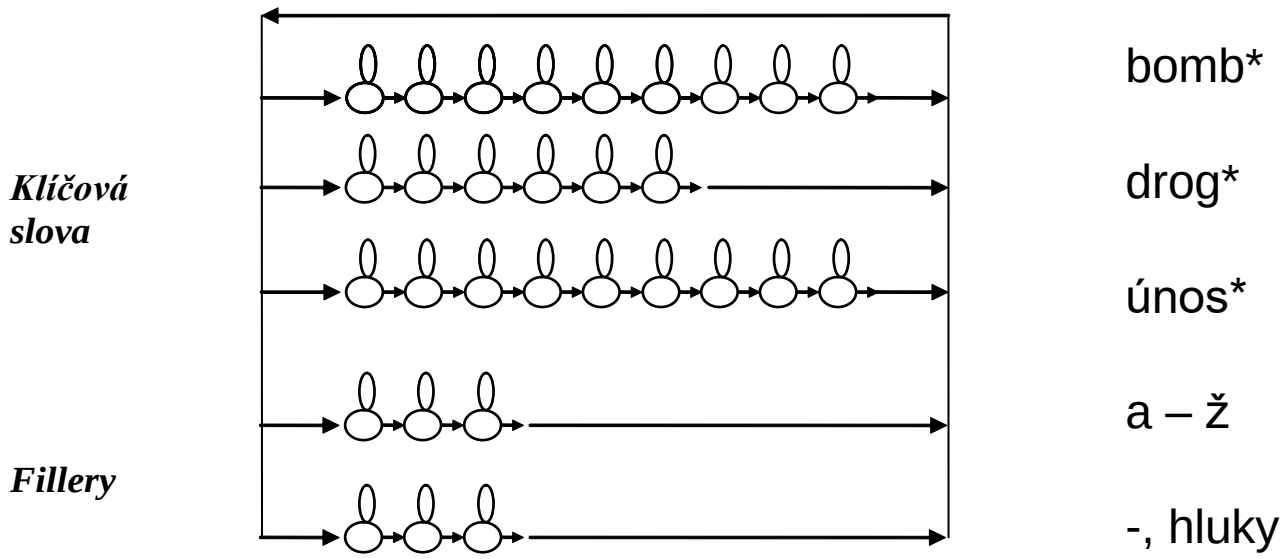
   pro spontánní řeč nemusí platit použitý jazykový model ani standardní výslovnost

2. **Provést rozpoznávání pouze s hledanými slovy a s fillery** (výplňovými jednotkami)

Detekce klíčových slov (2)

Jednoduchý KWS s fillery

Slovník systému obsahuje hledaná slova či jejich části (často raději s více výslovnostmi) a jako fillery slouží všechny fonémy + ticho a hluky)



Detekce klíčových slov (3)

Problém:

Jak docílit, že místo hledaných slov nebudou nalezeny příslušné sekvence fonémů?

Řešení: nastavení parametrů –s a –p v programu Hvite

Pomocí penále –p docílíme, že fonémový model slova bude mít vyšší skóre než stejná sekvence (v níž je každý foném zatížen penálem)

Příliš velké penále zase způsobí, že slova jsou nacházena i tam, kde řeč zní podobně

Vhodnou hodnotu penále je třeba vyladit experimentálně

Omezení:

Metoda nebere v úvahu kontext.

Např. slovo „auto“ může být nalezeno i ve větě „j**au to** bolí“

slovo „drogu“ „já**dro gu**my“

Rozpoznávání mluvčího (1)

Úloha: Rozpoznat identitu mluvící osoby

Typy úloh:

- **verifikace mluvčího** (je hlas přisuzovaný osobě A opravdu od ní?)
- **identifikace mluvčího z uzavřené sady** (1 z N osob)
- **identifikace mluvčího z otevřené sady** (1 nebo žádná z N osob)
-

Přístupy:

- **textově závislé rozpoznávání mluvčího (lze použít např. DTW)**
- **textově nezávislé rozpoznávání mluvčího (používají se GMM, DNN)**

Rozpoznávání mluvčího (2)

Jednoduché rozpoznávání mluvčích v HTK

Řešení:

- natrénovat pro každou osobu její HMM (stačí 1 stav, naopak je třeba hodně mixtur)
- jednotkami „abecedy“ zde nebudou fonémy, ale symboly mluvčích
- pro rozpoznávání se použije Hvite, slovníkem budou symboly mluvčích

```
#!MLF!#
„train-jan-1.lab"
jan
.
„train-jan-2.lab"
jan
.
„train-petr-1.lab"
petr
.
„train-petr-2.lab"
petr
.
```

Do konce tohoto týdne

Vyřešit úlohu rozpoznávání mluvčího z uzavřené sady

- Použijte data pro trénování HMM (vaše záznamy + PMR + sadu KCJ + Stanek) – tj. 95 mluvčích
- Malou část nahrávek (cca 10 od každého mluvčího) vyjměte z trénování a použijte jako testovací set
- Natrénujte modely všech mluvčích (32, 64, 128 mixtur)
- Proveďte rozpoznávací test mluvčích a vyhodnoťte skóre.
- Použijte polovinu dat z trénovací sady a zopakujte test.
- Použijte čtvrtinu dat z trénovací sady a zopakujte test.
- Výsledky mi pošlete do konce tohoto týdne.

Do konce příštího týdne

Vyzkoušet si úlohu vyhledávání klíčových slov

- Stáhněte si z e-learningu soubor Interview (cca 20-minutový záznam rozhovoru) a rozdělte si ho do úseků dlouhých 4 minuty (kvůli omezením HTK)
- Připravte si podle přednášky (a HTKbook) skript pro řešení úlohy KWS (key-word spotting)
- V dodaném souboru najděte (a poslechově ověřte) výskyty dvou slov „komentátor“ (4x) a „prezident“. Ve výsledné tabulce uveďte časy jejich výskytu (od-do, formát MM:SS).
- Pro co nejlepší funkci si musíte vyladit optimální hodnotu přepínače –p.
- Výsledky mi pošlete do konce příštího týdne.