

# **SPEECH RECOGNITION SYSTEM FOR RADIOLOGIC APPLICATION**

Kamil CHALUPNÍČEK, Master Degree Programme (5)

Dept. of Radio Electronics, FEEC, BUT

E-mail: k.chalupnicek@email.cz

Supervised by: Ing. Pavel Matějka

## **ABSTRACT**

This work deal with building speech recognizer for radiodiagnostic applications. Main part is concern in proposing structure of this recognizer and preprocessing text diagnosis for building specific language model.

## **1 ÚVOD**

Hlavní myšlenkou projektu je usnadnit lékařské obci zaznamenávání zjištěných diagnostických dat do elektronické podoby. Doslova to tedy znamená, místo ručního psaní textu na klávesnici „jen“ diktování diagnózy přes mikrofon připojený ke zvukové kartě počítače a následné softwarové vyhodnocení mluvené řeči s textovým výstupem na obrazovku či do souboru.

Řešení vyžaduje reálné zápisy diagnóz, tak jak byly zaznamenány samotnými lékaři, vhodné nástroje pro rozpoznávání mluvené řeči, dostatek záznamů mluvené řeči pro natrénování akustických modelů, případně již natrénovaný akustický model. Jednotlivé zápisy diagnóz je potřeba upravit pro další zpracování (kapitola 2). Následně z nich vytvořit jazykový model (kapitola 3.1) a slovník všech v textu užitých slovních tvarů. Poté vytvořit rozpoznávač (založený na skrytých Markovových modelech - kapitola 3.3), do kterého se implementuje jazykový model.

## **2 PŘEDZPRACOVÁNÍ TEXTU PRO JAZYKOVÉ MODELOVÁNÍ**

Zápisy diagnóz jsou z letmého pohledu jen text jako každý jiný. Nicméně obsahují pro nás důležité informace o užívaných slovních spojeních a použitých slovních tvarech v tomto oboru. Aby bylo možné tyto informace z textu získat, musí se na něm provést předzpracování.

Každý reálný text velkého objemu musí zákonitě obsahovat mnoho překlepů, chybně interpretovaných slov, zkratk, atd. S takto „poškozenými“ daty nelze přímo pracovat. Musí se odhalit překlepy, nahradit chybná slova správnými a zrozsáť zkratky. Takto interpretované

to zní lehce, nicméně celý postup je náročný. Pro konkrétní představu, vstupní text obsahuje cca 73000 různých slovních tvarů. Z toho je asi 35000 chybně interpretovaných a přibližně 1000 zkratk. Ruční oprava všeho je časově náročná a neakceptovatelná, automatická zase není v současné době tak dokonalá. Musí se zvolit vhodný kompromis mezi výtěžností použitelných dat z textu a pracností oprav. V praxi to znamená provést analýzu kolik zůstane dat vyhodí-li se méně často užívané slovní tvary se svým okolím, které po odstranění tohoto slova ztrácí pro další zpracování význam. Tímto krokem lze dojít na rozumný kompromis, z textu se vyhodí věty obsahující slova, která se v celém textu vyskytují méně než 20x a zbude 60 % textu obsahujícího 8403 slovních tvarů. Nyní je vhodná chvíle zapojit do projektu i kolegy zabývající se lingvistikou a češtinou vůbec, jelikož mají lepší prostředky i znalosti na automatické opravení chybných slov. Nebo, jako to bylo v tomto případě jen vyhodnotí, která slova jsou zaručeně správně a která nikoli. S touto problematikou byl nápomocen Radek Sedláček z Fakulty informatiky Masarykovy univerzity v Brně. Dostáváme se na 6116 slov, která jsou zaručeně správně a 2287 nerozpoznaných slov. Z oněch 2287 slov je možno ještě odfiltrovat za využití pomocného medicínského slovníku [3] dalších přibližně 900 specializovaných tvarů slov. Po popsání procedurách získáme „pouze“ 1400 slov a ty již je možno ručně projít a opravit.

Další potřebnou věcí je rozpis zkratek použitých v textu do plného tvaru. Zde se naráží na problém kontextové závislosti. Zkratka má totiž v různých místech textu různou koncovku. Automatický rozpis do správných tvarů je bez prostředků lingvistické analýzy okolního textu nemožný. Za předpokladu, že existuje jeden nejčastější tvar rozpisu dané zkratky, lze provést v celém textu náhrada této zkratky tímto nejčastějším tvarem. Jednou z komplikací, která si vyžádá práci navíc, jsou rozpisy zkratek s větším množstvím nepříbuzných tvarů rozpisu.

### **3 KONCEPCE ROZPOZNÁVAČE**

Rozpoznávač pro medicínské aplikace je složen z akustických modelů natrénovaných na dané řeči a speciálního jazykového modelu přizpůsobeného pro daný úkol.

#### **3.1 PARAMETRIZACE ŘEČOVÉHO SIGNÁLU**

Při generování řeči člověkem je hlasové ústrojí během krátkého časového intervalu (řádu milisekund) v jednom z konečného počtu stavů artikulačního ústrojí. Často používaná hodnota tohoto intervalu je 20 ms. Následně se na každém vzorku provede standardní MFCC parametrizace, neboli jsou vytvořeny Mel-frekvenční keprstrální koeficienty.

#### **3.2 SKRYTÉ MARKOVOVY MODEL (HIDDEN MARKOV MODELS)**

Algoritmy skrytých Markovových modelů (HMM) jsou založeny na statistice. HMMs vychází z představy procesu, který přechází mezi stavy a po každém přechodu vygeneruje vektor pozorování. Matematický model je popsán maticí přechodových pravděpodobností  $A$ , kde prvek  $a_{ij}$  udává pravděpodobnost přechodu ze stavu  $i$  do stavu  $j$ , a pravděpodobnostními rozloženími generování vektorů pozorování stavy  $b_j(o_t)$ . Pro řeč jsou používány tzv. levo-pravé modely, kdy matice přechodových pravděpodobností umožňuje pouze přechody do vyšších stavů nebo setrvání ve stavu stávajícím.

### 3.3 DEKODÉR

Spojením několika skrytých Markovových modelů vznikne v globálním pohledu mnoho možností jeho průchodu. Z cesty, pro kterou jsou zjištěny největší hodnoty podobnosti jednotlivých vygenerovaných vektorů pozorování s vektory navzorkovaného parametrizovaného řečového signálu je určeno o jakou posloupnost „znaků“ se v promluvě jednalo. Výsledná přesnost je závislá na kvalitě vytvořeného jazykového a akustického modelu.

### 3.4 JAZYKOVÝ MODEL

Jazykový model je v podstatě seznam kontextových závislostí. Znamená to, jaká slova a jak často se vedle sebe vyskytují. Modely mohou být obecně n-gramové, používají se však nejčastěji bigramové a trigramové.

Bigramový model obsahuje pouze závislosti dvou sousedních slov. Pro přiblížení jeden příklad: Pro slovo „štítina“ existuje nejčastěji sousední slovo „žláza“, naopak je nepravděpodobné (a taky se to v textu nikdy nevyskytne), aby sousední slovo bylo „gáza“. Trigramový model obsahuje informace o možnostech slova před daným slovem a také o variantách slova za ním. Tím, že rozpoznávači pomocí jazykového modelu dáme informace o bigramech, či trigramech zvýšíme úspěšnost rozpoznání pro danou specifickou úlohu.

Jak již bylo zmíněno, jazykový model se získá z předzpracovaného textu.

## 4 ZÁVĚR

Byla navržena koncepce hlasového rozpoznávače pro medicínské účely. Jeho praktické využití je reálné, jelikož v dnešní době je pro real-time rozpoznávače s velkým slovníkem únosných 20000 slov ve slovníku (v tomto případě je jich pouhých 8403). Nicméně, tento nízký počet je dvousečný, protože samotný rozpoznávač je schopen rozpoznat jen slova, která má ve svém slovníku. Ostatní slova, byť by se užívat mohla, nemůže rozpoznat. To lze vyřešit dodáním chybějících slov do slovníku.

Tento systém může být při jednoduchých modifikacích použit i u jiných vstupních dat, než radiodiagnostických. Pokud je ovšem požadováno zpracování real-time je nutné dodržet podmínku menšího slovníku rozpoznatelných slov, případně provozovat aplikaci založenou na této bázi na výkonnějším počítači.

## LITERATURA

- [1] Young, S.: The HTK book
- [2] Psutka, J.: Komunikace s počítačem mluvenou řečí
- [3] Medicínský slovník online: <http://www.hzp.cz/main/slovník.php>