AUTOMATICKÁ KOREKCE FONETICKÉ SEGMENTACE ŘEČI

Martin MATURA[1]

1 ÚVOD

Automatická fonetická segmentace zahrnuje řadu úloh a metod, jež si kladou za cíl automaticky nalézt co nejpřesnější hranice řečových jednotek v souvislém signálu řeči. Úkol přesného vymezení hranic je však velmi obtížný a automaticky nalezené hranice se často odlišují od toho, jak by je určil lidský expert. Práce se tedy zabývá návrhem techniky, která by vedla k přesnějšímu vymezení řečových jednotek a tím pádem i k vyšší kvalitě syntetizované řeči vytvořené **konkatenační metodou**.

Podstatou konkatenační metody je vytváření řeči řetězením (konkatenací) řečových jednotek, kde pojem řečové jednotky zahrnuje množství různých realizací hlásek. Řečové jednotky se vhodně vybírají a řetězí se za sebou a z toho vyplývá, že čím přesněji jsou určeny hranice řečových jednotek, tím lépe na sebe při řetězení jednotky navazují. Výsledná syntetizovaná řeč je potom kvalitnější.

2 ZPŮSOB ŘEŠENÍ

Řešení této úlohy spočívá ve vytvoření programu, který automaticky opravuje automaticky nalezené hranice mezi řečovými jednotkami. Hranice je definována jako časový okamžik v řečovém signálu oddělující dvě hlásky. Princip programu spočívá ve výběru kandidátů na správnou hranici a v následné činnosti **regresního SPM** (*score predictive model*), který každého kandidáta ohodnotí. Na základě ohodnocení je potom vybírána správná hranice.

Regresní SPM, k jehož vytvoření byla použita **metoda podpůrných vektorů** (*SVM* z anglického *support vector machine*), je velmi důležitou součástí práce a podle něj můžeme program v zásadě rozdělit na dvě části:

- ***natrénování regresního SPM***
  - z trénovacích (ručně nasegmentovaných) dat se okolo hranic vyberou kandidáti
  - pro kandidáty je spočtena sada vlastností (příznaků) a skóre, které určuje jejich kvalitu, tj. čím vyšší tím blíže je kandidát k ručně určené (správné) hranici
  - natrénování modelu za využití nástrojů z LIBSVM práce Changa a Lina (2001)
- ***korekce automaticky nalezených hranic pomocí SPM***
  - z automaticky nasegmentovaných dat se okolo hranic vyberou kandidáti
  - pro každého kandidáta jsou spočteny příznaky
  - na základě příznaků ohodnotí SPM kandidáta příslušným skóre
  - kandidát s největším skóre se stává novou hranicí

[1] Martin Matura, student navazujícího studijního programu Aplikované vědy a informatika, obor Kybernetika a řídící technika, e-mail: mate221@students.zcu.cz

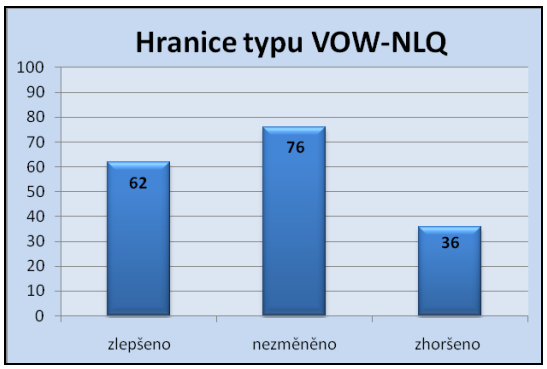
Příznaky počítané pro jednotlivé kandidáty se značně liší pro různé typy hranic mezi řečovými jednotkami. Typ hranice je určen podle toho, jaké skupiny hlásek odděluje. Hlásky jsou totiž rozděleny do skupin (samohlásky, nosovky, frikativy, aj.) podle jejich akustických vlastností.

3 VÝSLEDKY

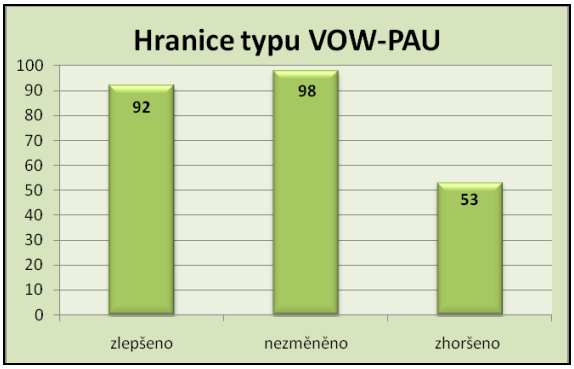
Činnost celého programu a úspěšnost korekce v podstatě závisí na správném natrénování SPM, který by úspěšně predikoval nejvyšší skóre pro správnou hranici. Toho je však obtížné dosáhnout a tak při korekci dochází ke třem různým případům:

1) dojde k úplné opravě hranice nebo k jejímu zlepšení (přiblížení ke správné hranici)
2) po korekci zůstane hranice beze změny
3) dojde ke zhoršení hranice (oddálení od správné hranice)

Práce je zatím zaměřena pouze na korekci hranic typu **samohláska - nosovka (VOW - NLQ)** a **samohláska - pauza (VOW - PAU)**. Pro každou z těchto hranic byl natrénován SPM, díky jehož činnosti dochází k částečnému opravení nasegmentovaného souboru. Na grafech 1. a 2. je přehledně znázorněn počet hranic, které se po opravě zlepšily, zhoršily nebo zůstaly nezměněny.



Graf 1: Hranice VOW-NLQ



Graf 2: Hranice VOW-PAU

Jak je vidět, počet zlepšených hranic skoro dvojnásobně přesahuje počet hranic, u kterých dojde ke zhoršení a to v obou případech. Regresní SPM tedy pracuje správně a je velice pravděpodobné, že dalšími úpravami programu je možné tyto výsledky ještě vylepšit.

4 ZÁVĚR

Automatická korekce fonetické segmentace řeči pomocí regresního skóre prediktivního modelu vytvořeného metodou support vector machine za pomocí nástrojů z LIBSVM, se jeví jako nadějný způsob pro vylepšení kvality syntetizované řeči konkatenační metodou, která je v dnešní době nejpoužívanější přístupem pro vytváření řeči počítačem. Při korekci byl poměr zlepšených a zhoršených hranic téměř dvě ku jedné. Průměrné zlepšení u hranic VOW_NLQ je 9,51 ms a průměrné zhoršení 6,11 ms. U VOW-PAU je průměrné zlepšení 7,66 ms a průměrné zhoršení 4,97 ms, což je dobrý výsledek. Do budoucna je možné provést ještě další změny v programu, především v oblasti hledání parametrů modelu potřebných k jeho natrénování, které mohou vést k ještě lepším výsledkům.

LITERATURA

Chih-Chung Chang and Chih-Jen Lin, LIBSVM : a library for support vector machines, 2001. Available at *http://www.csie.ntu.edu.tw/cjlin/libsvm*