

Tvorba systému syntézy řeči z limitované oblasti

Markéta Jůzová¹

1 Úvod

Syntéza řeči, tj. převod textu na řeč, používá k vytváření syntetizované řeči krátkých úseků skutečné řeči, tzv. *řečových jednotek* (např. difóny, trifóny), které se vybírají z řečového korpusu (tzv. *konkatenační metoda*, [Psutka J., Matoušek J., Müller L., Radová V. (2006)]).

Postup obecné syntézy je následující: Syntetizovaná věta se rozloží na jednotky, poté se pro každou jednotku v databázi najdou všichni reprezentanti, každý se ohodnotí kritériální funkcí, z těchto ohodnocených reprezentantů se hledá optimální posloupnost. Po nalezení nejlepší posloupnosti se vybraní reprezentanti, tj. reprezentanti tvořící optimální posloupnost, nakonec zřetězí za sebe.

2 Systém syntézy řeči z limitované oblasti

Limitovanou (omezenou) oblastí syntézy řeči může být např. automat podávající informace o počasí či odjezdech a příjezdech vlaků, automatická předpověď počasí nebo mluvící hodinky. Tvorba korpusu pro limitovanou oblast je popsána v [Jůzová (2011)].

Oproti obecné syntéze umí systém syntézy řeči z limitované oblasti syntetizovat jen omezenou množinu textu. Při syntéze se předpokládá, že se budou syntetizovat věty, které jsou (alespoň částečně) obsaženy v řečovém korpusu.

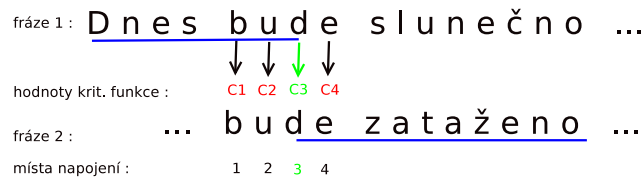
2.1 Hledání nejdelsích frází

Jednou z hlavních myšlenek přístupu popsaném v [Donovan R.E., Franz M., Sorensen J.S., Roukos S. (1999)] je hledání nejdelsích úseků syntetizované věty v korpusu, za nejkratší řečovou jednotku je považováno slovo. Touto metodou nám ze syntetizované věty vznikne několik typů úseků (podrobně v [Jůzová (2013)]) - úseky, které lze „vyříznout“ z původních promluv, úseky, kde je nutné hledat vhodné místo napojení mezi dvěma promluvami, a úseky, které se v limitovaném korpusu nevyskytují a je třeba je syntetizovat obecným způsobem.

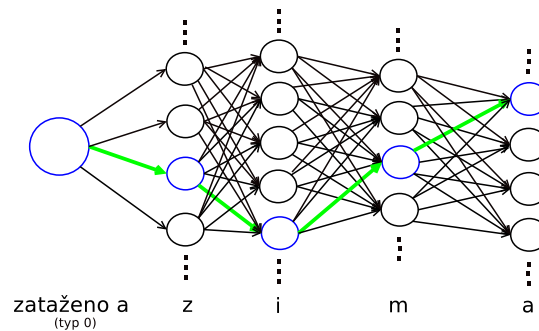
2.2 Syntéza jednotlivých úseků

- **Syntéza frází z limitované oblasti:** Jelikož je fráze přímo obsažena v řečovém korpusu, společně s ostatními „dlouhými“ řečovými jednotkami, které odpovídají slovům, frázím a větám, stačí při syntéze použít tento úsek jako celek.
- **Syntéza překryvů:** V tomto případě potřebujeme najít vhodné místo na přechod od jedné fráze ke druhé frázi. Vhodné je použít kritériální funkci obecné syntézy, která zde bude hledat takové místo napojení, pro které je její hodnota minimální. Hledání místa napojení pro úsek *bude* je znázorněno na následující ukázce:

¹ studentka navazujícího studijního programu Aplikované vědy a informatika, obor Kybernetika a řídicí technika, specializace Umělá inteligence, e-mail: juzova@students.zcu.cz



- **Syntéza úseku mimo limitovanou oblast:** Pro syntézu úseků, které v korpusu pro limitovanou oblast nejsou, lze použít klasický algoritmus obecné syntézy (viz 1), který je ale výpočetně náročnější, neboť každá jednotka může mít několik desítek až stovek reprezentantů. Nicméně se tento algoritmus použije jen pro části syntetizované věty, kdy počáteční a koncový úsek bude v korpusu přímo obsažen.



2.3 Syntéza celé věty

Po nalezení optimální posloupnosti řečových jednotek pro všechny úseky je můžeme zřetězit za sebe a použít standardní algoritmus na spojování jednotek.

3 Shrnutí

Použitím tohoto postupu došlo k výraznému snížení výpočetní náročnosti syntézy řeči, neboť se používají delší řečové jednotky (slova, fráze), případně se pouze hledá vhodné místo napojení dvou frází. Obecný, výpočetně náročný algoritmus se tedy používá jen na slova, která do limitované oblasti nepatří. Kvalita výsledné syntetizované řeči je srovnatelná s výstupem obecného syntetizéru, v mnoha případech i lepší.

Poděkování: Příspěvek byl podpořen grantovým projektem SGS-2013-032.

Literatura

- Psutka, J., Matoušek, J., Müller, L., and Radová, V., 2006 *Mluvíme s počítačem česky*. Academia, Praha.
- Jůzová, M., 2011. *Tvorba textového korpusu pro syntézu řeči z limitované oblasti*. ZČU, Plzeň.
- Jůzová, M., 2011. *Systém syntézy řeči z limitované oblasti*. ZČU, Plzeň.
- Donovan, R.E., Franz, M., Sorensen, J.S., and Roukos, S., 1999 *Phrase splicing and variable substitution using the trainable speech synthesis system*. IBM T. J. Watson Research Center, NY, USA