



Zohlednění variability v textech z omezené oblasti

Markéta Jůzová¹

1 Syntéza řeči z limitované oblasti

Pojmem *syntéza řeči* rozumíme převod (libovolného) textu na řeč. V případě syntézy z *limitované oblasti* (*Limited Domain Speech Synthesis*) se ale předpokládá, že syntetizovaný text pochází z dané oblasti. Příkladem takové omezené oblasti mohou být např. mluvící hodinky pro nevidomé nebo automat podávající informace o odjezdech a příjezdech vlaků či o předpovědi počasí. Dalším příkladem je ATC (*Air Traffic Control*), což je komunikace letiště s letadly s jasně definovanou frazeologií.

Protože syntéza řeči *konkatenační metodou*, která se dnes velmi používá, spočívá v řetězení úseků skutečné řeči uložených v databázi, nazývejme je dále *řečovými jednotkami*, zdá se být vhodné, aby tato databáze co nejvíce pokrývala danou doménu. Zároveň je ale nutné připomenout i časové a finanční náklady nahrávání řečové databáze, proto je vždy cílem vytvořit databázi řečových jednotek co nejmenší.

2 Texty z limitované oblasti

Ačkoliv se může zdát, že výběr textů pro nahrávání řečové databáze nebude v případě limitované oblasti složitý (vyberou se často používané věty), není tomu tak. Příkladem je již zmíněná omezená oblast ATC, jež by se měla řídit frazeologií, přesto je však v reálných datech vidět určitá variabilita.

Např. při potvrzování pilotem klesání letadla na určitou letovou hladinu je správná struktura promluvy následující:

```
označení_letadla descending flight levelight level číslo  
nebo  
descending flight levelight level číslo označení_letadla
```

V reálných datech pořizovaných na letišti během skutečné komunikace letadla s letištěm lze najít následující věty. První dvě přesně odpovídají správné struktuře promluvy, pouze mají prohozené pořadí slov, ostatní promluvy ale obsahují slova navíc (pozdravy, poděkování apod.)

```
AirBerlin 2 1 2 L descending flight level 2 9 0  
descending flight level 8 0 AirFrance 1 0 8 2  
descending flight level 1 0 0 díky CSA 8 7 9  
dzien dobry Lufthansa 9 V K descending 2 9 0  
now descending flight level 8 0 Aerosvit 2 0 9 thank you  
thank you descending level 2 9 0 Atlas Jet 1 1 9 7  
descending flight level 9 0 1 1 9 0 bye WizzAir 4 9 9
```

¹ studentka doktorského studijního programu Aplikované vědy a informatika, obor Kybernetika, specializace Umělá inteligence, e-mail: juzova@kky.zcu.cz

Pro zajištění kvality syntetizované řeči bychom měli připravovat texty pro nahrávání tak, aby obsahovaly všechna častá slova a fráze. Pak by bylo možné při syntéze spojovat delší řečové jednotky, ideálně takové, které se překrývají (lze pak vybrat vhodné místo pro řetězení).

3 Algoritmus výběru vět

Předpokládejme, že máme k dispozici reálná data z dané limitované oblasti a chceme vybrat množinu větných úseků pro nahrávání řečové databáze. Pro výběr vět jsme navrhli následující algoritmus (podrobně bude prezentován na konferenci na podzim tohoto roku, [1]):

1. určení četnosti výskytu všech unigramů (slov) a bigramů (dvojic slov) v textu
2. určení četnosti výskytu všech různých unigramů a bigramů pro každou větu s
3. ohodnocení každé věty s pomocí skóre $\sigma(s)$ na základě četností
4. výběr věty s nejvyšším skóre ($s^* = \operatorname{argmax}_s \sigma(s)$)
5. rozdělení všech vět na větné úseky tak, aby neobsahovaly bigramy z vybrané věty s^*
6. pokud nebyl dosažen požadovaný počet vět, zpět na 2

4 Výsledky

Následující tabulka 1 ukazuje procentuální pokrytí slov vyskytujících se v datech, která jsme měli k dispozici, při použití popsaného algoritmu pro výběr určitého počtu větných úseků. V tabulce jsou dále vypsány i počty a maximální četnosti slov, která nejsou ve vybraných větách.

	data	vybraná množina vět				
počet vět/větných úseků	45,551	1,000	2,000	3,000	4,000	5,000
celkový počet slov	391,429	6,919	12,276	15,455	19,567	23,431
počet různých slov	2,012	938	1197	1347	1513	1617
pokrytí různých slov	100%	46.6%	59.5%	66.9%	75.1%	80.4%
počet chybějících slov	–	1,074	815	665	499	395
maximální četnost chybějících slov	–	18	8	4	3	2

Tabulka 1: Porovnání původní množiny vět s množinami vybraných větných úseků.

Při výběru vět dává algoritmus přednost větám obsahujícím časté unigramy a bigramy. Slova, která do omezené oblasti přímo nepatří a vyskytují se tedy v datech jen ojediněle, jako např. pozdravy a poděkování, se díky své nízké četnosti do výběru vět možná vůbec nedostanou (závisí na počtu vybraných vět), což ale kvalitu syntézy ovlivní jen zřídka. Navržený algoritmus řeší i problém variability pořadí slov ve větě (viz prohození pořadí označení letadla se zbytkem fráze v ukázkách).

Poděkování

Příspěvek byl podpořen z prostředků NTIS.

Literatura

- [1] Jůzová, M., Tihelka D. *Minimum Text Corpus Selection for Limited Domain Speech Synthesis*. Submitted to TSD 2014