



## Detekce hranic frází klasifikačními přístupy

Markéta Jůzová<sup>1</sup>

### 1 Úvod

Pojmem *frázování* rozumíme dělení věty na kratší úseky vyslovené bez diskontinuit v prozodii. Palková (1974) uvádí, že frázování souvisí se syntaktickou strukturou věty a také s pozicí větne interpunkce (čárek). Vkládání hranic frází (pauz) do souvislé mluvené řeči přispívá k lepší srozumitelnosti promluvy, jedním z důvodů je také potřeba člověka se nadechnout. Přestože systémy převodu textu na řeč (TTS systémy) nepotřebují „dýchat“, častější výskyt hranic frází v syntetické řeči (spolu s odpovídající koncovou prozodií jednotlivých úseků) vede ke zvýšení její přirozenosti. Velmi dlouhé věty zní nepřirozeně, jsou náročnější na pozornost posluchačů a navíc je zde větší riziko vzniku rušivého nepřirozeného artefaktu.

Problémem úlohy frázování je, že většinou neexistuje pouze jedno správné frázování dané věty. Četnost vkládání hranic závisí na konkrétním řečníkovi (příp. anotátorovi), na stylu a rychlosti mluvení apod. – toto komplikuje i vyhodnocování.

TTS systém *ARTIC* vyvíjený na naší katedře využívá na určení hranic frází jednoduchý algoritmus založený na výskytu interpunkce. Přestože čeština používá čárky poměrně hodně, vznikají tak často velmi dlouhé fráze. Tento příspěvek prezentuje porovnání jednoduchého, čárky využívajícího frázovacího algoritmu s několika klasifikačními přístupy, které pro každou pozici mezi dvěma slovy rozhodují, zda se jedná o hranici fráze, či nikoliv.

### 2 Trénovací data

Klasifikační přístupy vyžadují velké množství trénovacích dat, ruční anotace hranic frází by ale byly zdlouhavé. Proto jsem se rozhodla využít velké řečové korpusy nahrané profesionálními řečníky pro účely TTS, které jsou již anotovány a segmentovány. Pauzy a nádechy, spolu s pozicí čárek, jsem pak použila jako informaci o hranici fráze.

Při návrhu experimentu jsem vycházela ze studií používajících klasifikační přístupy pro frázování angličtiny a použila jsem tedy sadu příznaků podobných těm, které autoři těchto studií často používají: předchozí a následující slovo a jejich morfologické tagy, délku věty, vzdálenost od nejbližší předchozí/následující čárky, zda předcházející slovo má čárku apod.

### 3 Výsledky klasifikace

Prováděla jsem experimenty s frázováním na řečníkovi závislým i nezávislým (generalizovaným) a používala následující přístupy:

- *Comma* – vkládá hranici fráze jen a pouze za každou čárku (používaný v TTS *ARTIC*)
- *LogReg* – logistická regrese
- *SVM* – Support Vector Machines (lineární)
- *CRF* – Conditional Random Fields

<sup>1</sup> studentka doktorského studijního programu Aplikované vědy a informatika, obor Kybernetika, specializace Počítačová syntéza řeči, e-mail: juzova@kky.zcu.cz

*LogReg* a *SVM* jsou zástupci klasifikačních přístupů, které vyhodnocují každou pozici zvlášť a neberou tedy v potaz výsledek klasifikace v jejím okolí. Oproti tomu *CRF* představuje sekvenci modelování a vyhodnocuje tedy celou větu najednou – což by mohlo vést k lepším výsledkům, neboť člověk dělá v řeči pauzy v určitých intervalech.

Tabulka 1 srovnává výsledky predikce hranic frází pro vybrané dva řečové korpusy a jasně ukazuje výhody použití klasifikačních přístupů v porovnání se základním frázovačem. Výhodou frázování natrénovaném na konkrétním řečnickovi je jeho konzistence se stylem frázování v nahraných datech.

		<i>A</i>	<i>P</i>	<i>R</i>	<i>F1</i>			<i>A</i>	<i>P</i>	<i>R</i>	<i>F1</i>
korpus1	<i>Comma</i>	0.954	1.000	0.743	0.852	korpus2	<i>Comma</i>	0.990	1.000	0.938	0.968
	<i>LogReg</i>	0.972	0.870	0.960	0.912		<i>LogReg</i>	0.991	0.947	0.995	0.970
	<i>SVM</i>	0.971	0.873	0.954	0.912		<i>SVMn</i>	0.991	0.947	0.991	0.969
	<i>CRF</i>	0.973	0.895	0.944	<b>0.919</b>		<i>CRF</i>	0.991	0.949	0.994	<b>0.971</b>

**Tabulka 1:** Výsledky detekce hranic frází prezentované na dvou vybraných řečových korpusech. Na trénování bylo vždy použito 80 % náhodně vybraných vět, 20 % na testování.

Ne vždy máme k dispozici dostatečně velký řečový korpus, a tak jsem vytvořila generalizovaný frázovací model, který by bylo možné použít při syntéze „malých“ hlasů. K natrénování takového modelu jsem použila všech šest velkých korpusů nahraných pro účely TTS. Z výsledků v tabulce 2 je vidět, že úspěšnost je nižší než v předchozím případě, ale i tak klasifikační přístupy předčili základní algoritmus.

	<i>A</i>	<i>P</i>	<i>R</i>	<i>F1</i>
<i>Comma</i>	0.959	1.000	0.767	0.868
<i>LogReg</i>	0.960	0.813	0.953	0.877
<i>SVM</i>	0.959	0.809	0.946	0.872
<i>CRF</i>	0.963	0.829	0.953	<b>0.887</b>

**Tabulka 2:** Výsledky detekce hranic frází pro spojený korpus (20 % náhodně vybraných vět z tohoto spojeného korpusu bylo opět použito pro testování).

Podrobným zkoumáním chyb klasifikátorů jsem zjistila, že ne všechny chybně detekované/nedetekované hranice jsou opravdu chybami – někdy se jedná jen o jinou možnost frázování dané věty. Bylo zjištěno také několik „nesmyslně“ umístěných hranic frází (pauz, nádechů) v řečových korpusech, což bylo pravděpodobně způsobené „zadrhnutím“ řečníka či nádechem z důvodu nedostatku dechu před koncem věty.

Podrobné výsledky všech experimentů, včetně vyhodnocení na manuálně připravených datech, jsou popsány v Jůzová (2017a) a Jůzová (2017b).

## Poděkování

Příspěvek byl podpořen grantovým projektem číslo SGS-2016-039.

## Literatura

- Palková, Z.(1974) *Rytmická výstavba prozaického textu*. Studia ČSAV, Academia, 1974.
- Jůzová, M. (2017) *Prosodic Phrase Boundary Classification Based on Czech Speech Corpora*. Přijato na TSD 2017.
- Jůzová, M. (2017) *CRF-based Phrase Boundary Detection Trained on Large-Scale TTS Speech Corpora*. Odesláno na SPECOM 2017.