Západočeská univerzita v Plzni, fakulta aplikovaných věd, katedra kybernetiky
Posudek na disertační práci k získání titulu doktor v oboru Kybernetika
Ing. Jiřího Kaly

Optimalizace rychlosti výběru řečových jednotek v konkatenační syntéze řeči

Ing. Jiří Kala se ve své disertační práci zabývá jednou z úloh zpracování řeči. Práce se váže k tématu syntézy řeči. Vývoj metody i experimenty jsou orientovány na konkatenační syntézu Češtiny pro syntezátor vyvinutý na pracovišti disertanta.

Předkládaná práce je členěna do 11 kapitol a je doplněna rozsáhlým seznamem prostudované literatury (obsahuje 76 položek), seznamem vlastních publikací nebo publikací, jejichž je spoluautorem (8 položek), dále abstraktem (v češtině, angličtině a němčině), obsahem, seznamem obrázků (41), tabulek (36) a seznamem zkratek a symbolů. Práci dokreslují tři přílohy. V prvních čtyřech kapitolách uvádí autor obecný rozbor problematiky syntézy řeči. Jsou to aktuální trendy a způsoby využití syntetizérů, konkatenační syntéza a syntéza výběrem jednotek. Pátá kapitola je věnována cílům disertační práce. Zbylé kapitoly jsou pro práci stěžejní. Konkrétně se jedná o přehled optimalizací procesu výběru řečových jednotek, způsob hodnocení algoritmů, které jsou dále popsány spolu s experimenty, o popis modifikace zaměřené na odstranění artefaktů a o hodnocení algoritmů nad redukovanou databází řečových segmentů. V závěru se disertant vyjadřuje k cílům práce a k jejich splnění, věnuje se závěrečnému zhodnocení výsledků a zamýšlí se nad návrhy na další práci. Přílohy se týkají množiny testovacích promluv, hodnot parametrů použitých v experimentech a syntetizovaných promluv (a to těch, které byly použity v poslechových testech, ukázkách artefaktů a v dalších promluvách).

Téma zvolené pro disertační práci je velmi aktuální. Syntéza řeči je potřebná v mnoha oblastech lidské činnosti. Velmi užitečná je pak především syntéza z textu. Do popředí zájmu výzkumných pracovníků se znovu dostává konkatenační syntéza, která je náplní předložené disertační práce. Důvodem jsou její možné aplikace v mobilních telefonech. Vedle dnes už nutného požadavku na srozumitelnost je jedním z hlavních požadavků přirozenost syntetické řeči a její produkce v reálném čase. O důležitosti této oblasti svědčí i to, že se touto oblastí zpracování signálu zabývají velmi intenzivně mnohá výzkumná pracoviště na celém světě. K velkému množství příspěvků na nejrůznějších konferencích, seminářích a workshopech přidal svůj díl i autor předkládané disertační práce.

Hlavním cílem práce je zrychlení vyhledání optimální sekvence řečových segmentů, vedlejším pak snížení rizika vzniku artefaktů ve vygenerovaných promluvách.

Navržené řešení je komplexní, použité metody jsou moderní. Jsou založeny na rozsáhlých znalostech v mnoha oborech, jsou podepřeny velmi dobrým matematickým zázemím a zkušenostmi s řešením otázek zpracování řečového signálu, které měl autor možnost získat na jednom předních pracovišť v tomto oboru. Velký význam celé práce vidím nejen ve velkém množství experimentů s ověřováním a porovnáváním výsledků získaných popisovanými metodami, ale také v tom, že práce je součástí projektu vývoje a realizace původního českého syntezátoru založeného na TTS syntéze. Popisované výsledky experimentů, které jsou dokresleny ukázkami a poslechovými testy na přiloženém CD, jednoznačně ukazují správnost volby tématu, dobře zvolený postup a náplň prací. Původním přínosem disertanta jsou vytvořené algoritmy, které umožní zvýšit rychlost výběru řečových jednotek a zároveň zachovávají kvalitu syntetické řeči, což potvrdily nejen matematické výsledky, ale i poslechové testy. To považuji za stěžejní výsledek.

Po formální a technické stránce je předkládaná práce na velmi dobré úrovni. Je psána přehledně, autor prokázal schopnost pracovat tvůrčím způsobem. Také logická stavba práce je na velmi dobré úrovni. Oceňuji používání anglických termínů u pojmů vysvětlovaných v češtině. Tento způsob pomáhá lepší orientaci čtenáře v zahraniční literatuře.

Odpovídající počet prací (8), jichž je autorem či spoluautorem, publikovaných na prestižních mezinárodních konferencích dokazuje nejenom autorovu schopnost vědecky pracovat, ale i schopnost informovat o dosažených výsledcích. Jedná se vesměs o sborníky z mezinárodních konferencí. Chybí mi však časopisecká publikace.

Mám několik připomínek a dotazů. Mezi připomínky patří např. konstatování, že:

- Na str. 54 textu pod vztahem (7.5) má být označení $N^C(i)$ pro celkový počet všech bodů řetězení v promluvě (v souladu se vztahem).
- Na str. 60 na obr. 8.1. chybí jedna uzavírací závorka.
- Graf na obr. 8.9. na str. 73 a text pod obrázkem si vzájemně neodpovídají. Popis obrázku je nejasný.
- Na str. 118 při hodnocení výkonnosti algoritmu ZCCVIT hodnotíte míru shody s referenčním algoritmem jako „velmi vysokou", a to u 6-ti promluv z 20. To je 30%. Taková shoda se mi nezdá jako „velmi vysoká".

Dotazy k práci mám následující:

- Definujte fonetický, spektrální a prozodický kontext, o kterých se v práci zmiňujete.
- Je podstatný rozdíl mezi výběrem jednotek pro různé národní jazyky nebo skupiny jazyků (indoevropské, tónové). Ověřoval jste popisované postupy pro angličtinu i u češtiny?
- Text na str. 58-59 budí dojem, že základní Viterbiův algoritmus VITORIG je hodnocen jinak, než ostatní popisované algoritmy. V čem je rozdíl?
- Na str. 73 píšete „o náhodně vybraných promluvách z novinových článků". Kolika mluvčími byl text namluven? Jednalo se o muže i ženu?
- Str. 144, závěr odstavce „Algoritmus VITBASE". Je možné vysvětlit, proč nejsou tvořeny žádné dvě promluvy stejnými řečovými segmenty u redukované databáze v porovnání s úplnou databází?

Na závěr konstatuji, že i přes uvedené připomínky, které nejsou zásadního rázu, Ing. Jiří Kala projevil schopnost samostatně vědecky pracovat. Vytčené cíle byly splněny. Proto mohu konstatovat, že **disertační práci doporučuji k obhajobě.**

Jana Tučková

V Praze, 14.11.2014

Prof. Ing. Jana Tučková, CSc.
Katedra teorie obvodů
FEL ČVUT v Praze

ÚSTAV MERANIA

SLOVENSKÁ AKADÉMIA VIED

Dúbravská cesta 9, 841 04 Bratislava

Tel.: 02/ 5477 4033, Fax: 02/ 5477 5943

Email: umersekr@savba.sk, Web: http://www.um.sav.sk

Oponentský posudek disertační práce Ing. Jiřího Kaly

Optimalizace rychlosti výběru řečových jednotek v konkatenační syntéze řeci

Posuzovaná disertační práce představuje monotématickou studii, kde na téměř 200 stranách autor nejprve provádí rozbor základních metod syntézy řeči a tvorby TTS systémů, dále se pak detailněji zabývá konkrétním popisem TTS systému ARTIC, resp. metodami výběru řečových jednotek pro konkatenační syntézu a problémům s tím spojenými. V praktické části textu jsou poměrně obsáhle popisovány provedené výpočetní a srovnávací experimenty a návrhy úprav algoritmů, které jsou v závěru hodnoceny pomocí poslechových testů. Práce obsahuje v úvodní části jasně formulované cíle a v závěrečném shrnutí pak popis dosažených výsledků – vedoucích k splnění stanovených cílů. Z popisu navržené metodiky a postupu řešení je jasné, že se jedná o velmi rozsáhlou komplexní problematiku, kdy bylo zapotřebí provést velké množství práce, analýz a výpočtů. V neposlední řadě to znamenalo rovněž vytvoření přípravných operací a pomocných programových nástrojů včetně návrhu a realizace poslechových testů k ohodnocení kvality syntetické řeči generované s použitím posuzovaných algoritmů výběru řečových jednotek. To vše jako celek přesahuje možnosti jednoho člověka a je samozřejmě i nad rámec jediné disertační práce. Přílohy A-C vhodně demonstrují a doplňují postup celého průběhu řešení včetně dosáhnutých výsledků a hodnocení kvality poslechovými testy. Ukázky generovaných promluv na přiloženém CD se vyznačují vysokou kvalitou syntetické řeči, dokazují obtížnost jejich subjektivního hodnocení (velmi malé rozdíly v posuzovaných větách), přesto je však výskyt artefaktů (způsobených nesprávným výběrem nebo skokovou změnou energie či základního tónu F0) dobře patrný.

Ve smyslu požadavků kladených na disertační práci mohu konstatovat:

- Z formálního pohledu je provedení práce na výborné úrovni – po grafické a lexikální stránce prakticky téměř bez chyb. Práci lze vytknout jedině tabulku zkratek a symbolů v úvodu, která nepokrývá použitou matematickou symboliku, naopak redundantní jsou definice zkratek opakující se následně v textu.
- Práce je dobře strukturovaná a logicky členěná, což napomáhá dobré čitelnosti a srozumitelnosti jinak hutného odborného textu.
- Citační odkazy jsou používány správně, relativně rozsáhlý výběr použitých zdrojů literatury je vhodný pro tento typ práce a plně pokrývá celou oblast řešené problematiky.
- V teoretické části práce autor prokázal, že se dobře orientuje ve stavu řešené problematiky na světové i domácí úrovni. Z experimentální části práce vyplývá, že disponuje širokými znalostmi z oblasti zpracování signálů a především tvorby TTS systémů. Navíc dostatečně

ovládá matematický aparát potřebný k pochopení, návrhu a realizaci jednotlivých analytických a statistických úloh pro řešení stanovených cílů práce.

- Z rozboru publikačních aktivit disertanta uvedených na konci práce vyplývá, že jádro předložené disertační práce bylo publikováno na dostatečné úrovni a že disertant představuje osobnost uchazeče s vědeckou erudicí a s vysokými realizačními a aplikačními schopnostmi.

Dotazy do diskuse v průběhu obhajoby:

- V práci prakticky chybí popis základních podmínek a průběhu provádění poslechových testů. Text se zde omezuje pouze na seznam dosažených výsledků a jejich zpracování. Výsledky takovýchto subjektivních hodnocení výrazně závisí na osobách posluchačů-hodnotitelů. Proto bych uvítal doplňující informace o věkové struktuře (pohlaví), povolání (studenti/kolegové, odborníci v oblasti řečové problematiky), poslechových podmínkách (sluchátka/reproduktory) a motivaci jednotlivých hodnotitelů. Prováděla se nějaká klasifikace, z níž by bylo možné určit jejich „spolehlivost" – možnost vyřadit takového posluchače, jak bývá obvyklé např. při labelování řečových korpusů?
- Výsledky ABX poslechových testů byly následně statisticky zpracovány pomocí testů hypotézy, zda se jedná o statisticky významnou odchylku. Zde opět chybí bližší vysvětlení, jaký konkrétní typ testu byl použit (*t*-test, *F*-test, chí-kvadrát, variance apod.) – v textu je zmiňován pouze „znaménkový test" (str. 124-125).
- Jako ekvivalent k subjektivní metodě poslechových testů se naskýtá možnost aplikace GMM, HMM, NN nebo jiných přístupů používaných v systémech ASR. Získané skóre úspěšného rozpoznání originálního řečníka, z jehož nahrávek byl vytvořen řečový korpus pro daný TTS systém, lze obecně považovat za objektivní parametr, který je možné dále numericky porovnávat a statisticky vyhodnocovat. Zajímal by mne disertantův názor o jejich využitelnosti při řešení této úlohy.

Závěr:

Předložená disertační práce vyhovuje podmínkám kladeným na tvůrčí vědeckou práci, proto ji **doporučuji k obhajobě**. Stanovené cíle práce byly beze zbytku splněny, současně však existuje řada problémů, které je třeba v budoucnu ještě vyřešit. Zejména se jedná o dosažení potlačení artefaktů v generovaných syntetických promluvách a v neposlední řadě také snížení výpočetní a paměťové náročnosti s cílem aplikace TTS systému ARTIC i pro platformu mobilních elektronických zařízení.

V Bratislavě, 27.10. 2014.

.....

Dr. Ing. Jiří PŘIBIL