POSUDEK NA DISERTAČNÍ PRÁCI

Téma práce: Adaptace akustického modelu v úloze s malým množstvím adaptačních dat.

Doktorand: Ing. Zbyněk ZAJÍC

Posudek vypracoval: Doc. Ing. Petr POLLÁK, CSc.
ČVUT FEL K13131, Technická 2, 166 27 Praha 6

Předložená práce ing. Zajíce se věnuje aktuální problematice adaptace akustického modelu v systémech rozpoznávání řeči, zajišťující přizpůsobení rozpoznávače konkrétním podmínkám, nejčastěji aktuálnímu mluvčímu. Tato adaptace je nezbytnou součástí systémů rozpoznávání spojité řeči s velkým slovníkem, neboť se významnou měrou podílí na dosažení co nejvyšší přesnosti rozpoznávání. Uvedená problematika je dlouhodobě řešena v celosvětovém měřítku, avšak výzkum v dané oblasti je stále vysoce aktuální zejména s ohledem na nové dílčí úlohy, jakou je i adaptace s malým množstvím dat řešená v rámci předložené práce. Cíle předložené disertační práce jsou přehledně uvedeny hned v úvodní kapitole a lze je stručně shrnout následovně:

1. popsat a analyzovat známé metody generativní i diskriminativní adaptace s užším zaměřením na problematiku malého množství dat,
2. navrhnout a experimentálně ověřit modifikace metod na bázi lineárních transformací s cílem dosáhnout lepší účinnosti pro malé množství dat,
3. navržené metody implementovat do existujícího systému rozpoznávání mluvené řeči.

Takto stanovené cíle jsou disertabilní a autor tyto cíle v předložené práci naplnil. Hlavní přínosy předložené práce jsou shrnuty v následujících bodech.

- V první řadě autor naplňuje první cíl, podat přehled používaných metod adaptace a adaptačního trénování, což je obsahem kapitol 3-6. Uvedený souhrn je zpracován velmi přehledně a srozumitelně. Autor odkazuje nejvýznamnější práce, kde jsou dané metody popsány podrobněji, avšak trochu širší diskuse zahrnující i výsledky dosažené citovanými autory by tento přehled jistě vhodně doplnila. Toto srovnání jsem trochu postrádal i v samotném závěru práce.
- Vlastním přínosem autora jsou pak adaptační metody navrhované pro danou úlohu, a to jmenovitě kombinace základních adaptačních technik, ze kterých bylo dosaženo nejlepších výsledků při kombinaci diskriminativních verzí MLLR adaptace na úrovni příznaků a MAP adaptace akustických modelů, a zejména pak modifikované metody, jejichž hlavním obecným cílem je snížení počtu odhadovaných parametrů pro úlohu s velmi malým množstvím adaptačních dat. Zde je potřeba vyzdvihnout zejména techniky založené na inicializaci matic lineárních transformací z trénovacích dat či adaptaci založenou na transformační matici vytvořené kombinací bázových matic určených opět z trénovací databáze. Zajímavá je metoda redukující chybu adaptace pomocí neuronových sítí typu bottleneck. V této metodě při trénování neuronové sítě pracuje autor s chybně a korektně odhadnutou adaptací získanou ze SpeechDat-E korpusu. Je v tomto případě však dostupné dostatečné množství dat pro korektně odhadnutou adaptaci? A je dostatečné i vlastní množství dat pro natrénování sítě? Mluvíte o 20 vstupních vektorech pro 700 mluvčích (to představuje 14000 trénovacích párů), počet parametrů sítě pro daný počet neuronů je však vyšší.
- Z aplikačního hlediska je nepochybným přínosem, že vybrané popisované techniky byly implementovány také v rámci reálného systému pro on-line rozpoznávání mluvené řeči vyvinutém pro účely automatického titulkování na školícím pracovišti autora.
- Schopnost úspěšné adaptace pomocí popisovaných technik byla potvrzena v experimen-

tech, provedených ve většině případů na telefonních databázích, a to na privátní databázi CzT vytvořené na pracovišti autora a na databázi z řady SpeechDat-E, nakonec také i ve zmiňovaném on-line systému titulkování přenosů z Parlamentu České republiky. Dosažené výsledky potvrdily schopnost adaptace i pro velmi malé množství adaptačních dat.
K popisu experimentů bych však měl několik drobných výhrad. Pro lepší srovnání přínosů jednotlivých adaptačních technik by bylo vhodné používat nejenom absolutní míry (Acc resp. WER), ale také relativní míru snížení chyby na úrovni slov (WERR), vztaženou vždy k SI modelům v dané úloze. To by umožňovalo i srovnání s výsledky prezentovanými jinými autory (viz poznámka výše) či přehlednější srovnání výsledků dosažených na různých datech (tj. pro experimenty se signály různé kvality a pro různá nastavení rozpoznávače). Pro úplnost popisu experimentů by bylo také dobré doplnit i některé další detaily, např. při srovnání času adaptace na jednoho řečníka průměrnou dobu trvání adaptačních promluv, u experimentů na korpusu SpeechDat-E pak není jasné, jaké promluvy byly v experimentech použity. Autor mluví o větách, z textu v kapitole 7.1.2 však může jít i o obsažené jednoslovní povely. Jaká konkrétní data tedy byla použita v pokusech na SpeechDat-E korpusu? Jednalo se pro jednotlivé mluvčí vždy o stejné položky či byly promluvy náhodně vybírány? Překvapila mne i poznámka o neexistujícím referenčním přepisu. Tato databáze totiž obsahuje jak ortografický tak fonetický přepis všech promluv v položkách LBO a EPI dostupného souboru s anotací.

K práci bych měl také připomínku k obsahu a rozsahu jednotlivých kapitol. Část věnovaná popisu známých technik je spíše rozsáhlejší (kapitoly 2-6), zatímco popisu výsledků realizovaných analýz i popisu vlastních modifikací je věnováno prostoru méně. To je vše součástí kapitoly 7, kde jsou však popisovány jak základní nastavení experimentů, tak výsledky experimentů se známými metodami, a potom i navržené modifikace adaptačních technik s dosaženými výsledky. Práci by celkově jistě prospělo, kdyby navržené modifikace a vlastní přínosy byly popsány odděleně a místy i podrobněji na obecné úrovni, a nejenom jako součást popisu realizovaných experimentů. Autor také ve své práci popisuje podrobně jednotlivé techniky, avšak chybí systematická informace o použitém softwaru realizujícím tyto techniky (výjimkou jsou malé poznámky o implementaci fMLLR techniky do on-line systému, o konstrukci regresního stromu pro MLLR pomocí HTK či o nástroji pro výpočet ICA). Byl pro navrhované metody autorem vytvořený či modifikovaný vlastní software či byly použity některé dostupné sady nástrojů? Tato informace je zásadní i z toho pohledu, že mnohé autorem navržené modifikace využívají relativně pokročilé algoritmy a techniky, které nejsou obvyklou součástí dostupných nástrojů.

Po formální stránce je předložená práce na dobré úrovni, s výjimkou výše uvedených obsahových připomínek. Text je psaný srozumitelnou češtinou, grafická úprava je velmi dobrá. Nalezl jsem jen několik závažnějších gramatických chyb a několik drobných formálních prohřešků (např. používání anglické terminologie i v případě, kdy je zavedený český překlad resp. míchaní anglických a českých výrazů či vytváření novotvarů jako "unsupervised úloha, transkribovány, zrobustnění", častý je výskyt jednopísmenných předložek na konci řádku).

Celkový přínos práce je přes připomínky výše neodiskutovatelný a o originálním přínosu autora v dané oblasti není pochyb. To nakonec potvrzuje i kvalitní publikační činnost autora, která zahrnuje 15 publikací (v 8 případech jako první autor), přičemž je nutné vyzdvihnout zejména 4 příspěvky na prestižních konferencích řady Interspeech, kde procento přijatých příspěvků je typicky velmi nízké.

Na základě výše uvedených skutečností a přes výše zmíněné připomínky lze však jistě konstatovat, že předložená práce popisuje originální výsledky vědecké práce, a proto práci **doporučuji** k obhajobě za účelem získání vědecké hodnosti doktora na Západočeské univerzitě v Plzni.

V Praze dne 9. října 2012

Ing. Petr Červa, Ph.D.
Technická univerzita v Liberci
Ústav informačních technologií a elektroniky
Studentská 2, 461 17 Liberec

Oponentský posudek disertační práce Ing. Zbyňka Zajíce:

„Adaptace akustického modelu v úloze s malým množstvím adaptačních dat"

Disertační práce Ing. Zbyňka Zajíce se týká problematiky automatického rozpoznávání řeči, speciálně pak oblasti adaptace akustického modelu. Je zaměřena zejména na úlohu adaptace s malým množstvím adaptačních dat, která patří už z principu k těm nejobtížnějším a je zároveň velice zajímavá z hlediska možností praktického využití v různých systémech převádějících řeč na text.

Při řešení cílů práce disertant vychází z velmi podrobné znalosti současného stavu problematiky. To dokládá nejen velice rozsáhlý seznam použité literatury, ale i podrobně zpracované kapitoly číslo 3 až 6, v rámci kterých je podán ucelený přehled a vysvětlen princip v současné době nejpoužívanějších adaptačních technik. Zároveň jsou zde popsány i problémy spojené s úlohou adaptace v online režimu a principy metod zaměřené na adaptaci s malým množstvím adaptačních dat. V současné době mi není známa žádná práce v českém jazyce, která by se danou problematikou zabývala ve větším rozsahu. Za přínosné také považuji, že práce se věnuje klasickým i diskriminativním adaptačním metodám a dále také způsobem uplatnění těchto metod v rámci trénovaní akustických modelů.

Samotné řešení cílů práce je popsáno v kapitole číslo 7, která popisuje i dosažené experimentální výsledky. Kladně hodnotím volbu testovacích množin, na nichž byly jednotlivé metody ověřovány. Použité sady obsahovaly dostatečně velký počet různých mluvčích a i celkové množství použitých dat má dostatečně velkou vypovídající hodnotu. Kromě toho je plusem, že disertant bere v potaz i statistickou významnost provedených experimentů, nejen pouhý rozdíl dosažených skóre. Naopak určitou nevýhodu vidím v tom, že experimenty byly provedeny jen za situace, kdy bylo použito pro trénování výchozích akustických modelů jen malé množství dat a dané systémy pracovaly bez jazykového modelu nebo jen s malým slovníkem (to se netýká systému pro přepis parlamentních debat). Základní chybovost systémů před adaptací byla z těchto důvodů poměrně nízká. Ačkoli jsou tak dobře vidět rozdíly mezi jednotlivými vyhodnocovanými metodami, což je v práci uvedeno jako logický důvod, není na druhou stranu zase zřejmé, jaké by byly výsledky metod v praxi, kdy je většina systémů vybavena co nejlepším výchozím akustickým a jazykovým modelem.

Co se týká dosažených výsledků, jsou podle mého názoru v souladu se zadáním. Autorem navržené přístupy pro robustní adaptaci vedou k dosažení lepších výsledků než v současné době běžně používaná základní varianta metody fMLLR, i když nedávají ve většině případů lepší výsledky než existující metoda založená na lineární kombinaci bázových matic získaných pomocí ML odhadu. Ku prospěchu disertanta je ale třeba podotknout, že navržené metody jsou založené na netriviálních technikách používaných pro řešení jiných úloh. Autor tak musel pro jejich pochopení a použití vynaložit další dodatečné úsilí a nápad s jejich aplikací je originální a zajímavý. V tomto světle ovšem považuji za nešťastné, že použití těchto technik není názorněji popsáno. Vhodné by bylo například doplnit obecný text o ICA na straně 86 konkrétním popisem, jak vypadá vstup a výstup do této metody v daném případě použití.

V oblasti online adaptace pak vidím největší přínos práce v tom, že navržená metoda byla koncipována s ohledem na skutečné požadavky kladené na úlohu online přepisu řeči a jistě tak bude prakticky využívána v celé řadě systémů vyvíjených na ZČU. Za přínosné pro celou komunitu dále považuji, že autor v rámci práce za stejných a jasně definovaných podmínek experimentálně porovnal celou řadu různých metod - ať již klasických tak i moderních pro robustní adaptaci.

Celkově hodnotím práci po obsahové i formální stránce pozitivně. Publikační činnost autora je také vyhovující, neboť Ing. Zbyněk Zajíc je uveden jako autor či spoluautor u 15 prací, přičemž většina z nich je zahrnuta v obecně uznávaných citačních indexech a v několika případech se jedná o články na prestižních světových konferencích z oboru.

Práci Ing. Zbyňka Zajíce proto **doporučuji** k obhajobě.

Do diskuze navrhuji několik otázek:

1) Jaké by byly výsledky nejlepších ověřovaných metod robustní adaptace v případě, že by byly použity v rámci systému, který by dosahoval nižší základní chybovosti?
2) V kapitole 3.2 a 7.8.1 je uvedeno, že pro akumulaci parametrů z adaptační promluvy je třeba mít data zarovnána ke stavům modelu pomocí pevného zarovnání - *force alignmentu* (správněji má být zřejmě *forced alignmentu*). Ve skutečnosti je ovšem možné použít i pomalejší ale přesnější forward-backward algoritmus, který přiřazuje data ke stavům modelu pomocí pravděpodobnosti respektive bere v úvahu všechny možné varianty zarovnání. V případě velmi malého množství dat by tato skutečnosti mohla hrát určitou roli. Byly prováděny experimenty i za použití jiného než pevného zarovnání?
3) V případě malého množství adaptačních dat je možné hledat transformační matici v rámci fMLLR i jako diagonální či blokově diagonální. Množství odhadovaných parametrů je pak nižší a výsledná transformace by měla být robustnější. Vyzkoušel jste, jak by navržené přístupy pro robustní adaptaci dopadly při porovnání se zmíněnými variantami fMLLR?
4) V kapitole 7.6.2 je v rámci online adaptace navržen postup využívající informace o jistotě rozpoznaných slov. Přitom není uvedeno, která konkrétní metoda je pro výpočet CF použita a není zřejmé, jaký je vliv této informace na výsledky adaptace, protože prezentovány jsou pouze výsledky za použití CF. Jak by dopadla adaptace v online režimu, pokud by CF nebyl využit?

V Liberci, 4.10.2012

Ing. Petr Červa, Ph.D.