

POSUDEK NA DISERTAČNÍ PRÁCI

Téma práce: Generativní a diskriminativní klasifikátory v úlohách textově nezávislého rozpoznávání a diarizace mluvcích.

Doktorand: Ing. Jan SILOVSKÝ

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

Předložená práce ing. Silovského se věnuje problematice rozpoznávání a diarizace mluvcích, která nachází v současné době mnohé aplikace v systémech hlasových technologií. Autorův výzkum je prováděn především v souvislosti s výzkumným zaměřením jeho pracoviště v oblasti automatického přepisu rozhlasových a televizních pořadů, kde je úloha diarizace mluvcích důležitou částí celého systému. Uvedená problematika je velmi aktuální a je intenzivně řešena v celosvětovém měřítku, což je rovněž zřejmé z autorem provedeného přehledu stavu problematiky a z rozsáhlého seznamu citovaných pramenů.

Cíle předložené disertační práce jsou přehledně uvedeny po stručném úvodu do dané problematiky v první kapitole a lze je stručně shrnout následovně:

1. teoretickým cílem disertace je analýza a návrh modifikovaných metod pro rozpoznávání a diarizaci mluvcích s užším zaměřením na generativní a diskriminativní klasifikátory,
2. praktickým cílem pak je navrhnout vhodný způsob vyhodnocování a navržené metody v experimentální části otestovat na vytvořených evaluačních databázích či v rámci aplikací v reálných systémech.

Výše stanovené cíle jsou jistě disertabilní a autor tyto cíle v předložené práci naplnil. Předložená práce představuje jednoznačný přínos v dané problematice na národní i mezinárodní úrovni a nejvýznamnější dosažené výsledky lze shrnout v následujících bodech.

- V první řadě autor naplňuje první ze svých stanovených cílů, podat přehledný a ucelený pohled na danou problematiku. Kapitoly 4 a 5 obsahují ucelený a velmi detailní popis používaných klasifikačních algoritmů v úlohách rozpoznávání mluvcího. Většinou se jedná o známé a standardně používané metody (i v jiných úlohách), ale jejich podrobný souhrnný popis je velmi užitečný nejen pro další výklad v předložené práci, ale významně rozšiřuje základní popis této problematiky dostupný v pramenech v českém jazyce.
- Vlastním přínosem autora jsou pak optimalizované metody a systémy rozpoznávání mluvcího, které jsou popsány v kapitolách 6 a 7, kde konkrétní přínosy jsou především v analýze a optimalizovaném nastavení základních systémů pro rozpoznávání mluvcího na bázi UBM-GMM, GMM-SVM, MLLR-GMM, JFA i na bázi i-vektorů.
- Ve vlastní úloze diarizace mluvcích popisované v kapitole 8 vychází autor ze standardně používané struktury pro tuto úlohu a přináší vlastní řešení především v úloze shlukování mluvcích na bázi faktorové analýzy, která jsou založena na shlukování založeném na BIC kritériu, na základě kosinové vzdálenosti i-vektorů a na bázi PLDA modelování. Nakonec autor předkládá návrh kombinovaného dvoufázového shlukování na bázi BIC v prvním kroku následované metodami založenými na i-vektorech v kroku druhém.
- Za důležitý výsledek považuji návrh komplexní metodiky hodnocení systémů identifikace a verifikace mluvcích, která je popsána v kapitole 2. Autor popisuje soubor kritérií zahrnujících kromě vyhodnocení rozlišovacích schopností analyzovaných detektorů i vyhodnocení kvality jejich kalibrace.

- Pro objektivní vyhodnocování navrhovaných systémů má velký význam i vytvoření vlastních evaluačních databází ze záznamů televizních a rozhlasových pořadů.
- Významným výsledkem je účast autora v evaluačních kampaních NIST SRE v letech 2008 a 2010, které umožňují srovnání navrhovaných metod a systémů v celosvětovém měřítku. Zde autor prezentoval vlastní optimalizace systémů na bázi UBM-GMM, JFA a i-vektorů.
- Z aplikačního hlediska je přínosná skutečnost, že navrhované metody nejsou analyzovány pouze v simulovaných experimentech, ale že byly implementovány a experimentálně ověřeny v rámci reálných systémů vyvíjených na školícím pracovišti autora, zejména pak v systémech pro automatickou transkripci televizních a rozhlasových pořadů. Ve výsledném systému bylo s rozpoznáváním mluvčích a následným výběrem modelů pro daného mluvčího dosaženo snížení chyby na úrovni slov (WER) až o 22%.

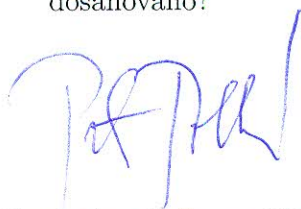
Po formální stránce je předložená práce na vynikající úrovni, nenalezl jsem žádné závažné prohřešky. Je členěna přehledně, text je psaný velmi pěknou češtinou, grafická úprava je na velmi vysoké úrovni.

Nezpochybnitelný přínos autora v dané oblasti potvrzuje i výjimečně obsáhlá a kvalitní publikační činnost autora, která zahrnuje 28 publikací (v 10 případech jako první autor), přičemž je nutné vyzdvihnout 4 články v impaktovaných časopisech a 4 příspěvky na prestižních konferencích řady Interspeech. I ostatní publikace ovšem představují příspěvky z kvalitních mezinárodních konferencí a workshopů.

Na základě výše uvedených skutečností lze tedy jistě konstatovat, že předložená práce popisuje originální výsledky vědecké práce. Z předložené práce je zřejmý jak teoretický, tak i praktický nadhled autora nad danou problematikou. Práci hodnotím celkově jako velmi kvalitní a **doporučuji** ji k obhajobě za účelem získání vědecké hodnosti doktora na Technické univerzitě v Liberci.

Do diskuse bych měl v tomto okamžiku následující 2 otázky:

- V práci jsem trochu postrádal v kapitole 7 širší diskusi nad výsledky ostatních účastníků evaluačních kampaní a vzájemné srovnání dosažených výsledků. Můžete stručně nastínit, jaké výsledky byly přibližně dosahovány dalšími účastníky NIST SRE evaluací?
- V práci zmiňujete také aplikace v kriminalistické praxi a v justiční oblasti. Jsou v této souvislosti Vámi navrhované metody přímo aplikovatelné, především pak s ohledem na často výrazně nižší kvalitu signálů v těchto situacích? Jakých výsledků bylo orientačně dosahováno?



V Praze dne 2. února 2012

OPONENTSKÝ POSUDEK DISERTAČNÍ PRÁCE

Kandidát: Ing. Jan Silovský, Technická univerzita v Liberci
Název: Generativní a diskriminativní klasifikátory v úlohách textově nezávislého rozpoznávání a diarizace mluvčích
Recenzent: Doc. Dr. Ing. Jan Černocký, FIT VUT v Brně

Předložená disertační práce Ing. Jana Silovského má 170 stran včetně příloh a obsahuje 9 kapitol. Tento posudek se v sekci 1 nejprve zabývá jednotlivými kapitolami včetně poznámek a v sekci 2 obsahuje zhodnocení technické stránky práce. Sekce 3 hodnotí formální stránku a sekce 4 obsahuje závěr, celkové zhodnocení práce a doporučení. Posudek je doplněn otázkami k obhajobě.

1 Obsah práce a poznámky ke kapitolám

Práce se věnuje velmi aktuální problematice rozpoznávání mluvčího a diarizace mluvčích v proudě neoznačovaných audio dat. Svým zaměřením je na hranici mezi elektrotechnikou a informačními technologiemi, obsahuje netriviální matematický komponent a podle mého názoru je vhodně začleněna do aktivit fakulty, která má ve svém názvu „mezioborová studia“.

V první kapitole je uveden úvod do problematiky a autor přehledným způsobem shrnuje „state of the art“ v oboru. Příkladná pozornost je věnována převodu terminologie, která dosud existovala pouze v anglickém jazyce, do češtiny – o jazykové otázce více níže. Kapitola by si zasloužila uvedení obecného schématu systému pro rozpoznávání mluvčího (výpočet příznaků, vyhodnocení věrohodností, kalibrace, prahování, atd.), které by ulehčilo další výklad – znalému čtenáři je toto schéma známo, mělo by ovšem být doplněno, pokud bude práce dále sloužit k pedagogickým účelům. Závěr kapitoly definuje cíle práce.

Kapitola 2 obsahuje studii způsobů vyhodnocení systémů pro ověřování mluvčího a vynikajícím způsobem shrnuje klasické metody (jako je DET křivka, EER, a podobně) i moderní přístupy aplikačně nezávislého vyhodnocení, definované v posledních letech N. Brummerem a dalšími. Jedná se o výborný text, který bude jistě hojně využíván dalšími generacemi výzkumníků v našem i souvisejících oborech. Z hlediska systematickosti celé práce bych se pouze přimlouval, aby bylo do této kapitoly přesunuto i vyhodnocení systémů pro diarizaci, které je nyní v experimentální kapitole na konci práce.

Kapitola 3 se zabývá kalibrací a fúzí systémů, která, pokud je mi známo, nebyla souhrnně nikdy česky zpracována, jedná se tedy opět o velmi záslužný čin.

Kapitola 4 tvoří teoretické jádro práce a popisuje generativní klasifikátory pro rozpoznávání mluvčího. Začíná klasickým přístupem adaptace GMM modelu na jednotlivé mluvčí, pokračuje moderními metodami sdružené faktorové analýzy (JFA) a končí u velmi progresivní metody iVektorů, která byla definována v několika posledních letech. Autor si dal značnou práci s konsolidací poznatků roztržitých v několika desítkách článků (především od Patricka Kennyho a jeho studentů), výsledkem je velmi kompaktní text, který pokrývá teoretické i praktické aspekty JFA a iVektorových systémů. Kapitola je navíc doplněna přílohami obsahujícími detailní odvození vztahů a matematickou „kuchařku“ pro implementaci systémů. Podobně jako u předchozích kapitol se jedná o záslužné dílo, drobné poznámky mám pouze k absenci některých standardních anglických termínů, například u definice lineární diskriminační analýzy LDA (sekce 4.4.3 - „within-class a across-class covariance matrices...“)

Kapitola 5 s názvem „Diskriminativní klasifikátory“ se zabývá především jednou třídou těchto algoritmů, a to SVM. Metoda je opět detailně zpracována, včetně technik, které jsou používány pro kompenzaci nepříznivých vlivů přenosového kanálu. V rámci kompletnosti mohla být kapitola doplněna bodovým přehledem dalších možností použití diskriminativního trénování v rozpoznávání řečníka, jako jsou pokusy

o diskriminativní trénování JFA, iVektorů, případně o použití neuronových sítí. Je také škoda, že SVM klasifikátory byly použity pouze v jedné rozpoznávací úloze (záznamy televizních a rozhlasových pořadů, TVR) a nebylo provedeno srovnání s GMM, JFA a iVektorovými systémy na NIST datech.

Kapitola 6 představuje experimenty rozpoznávání mluvčích v TVR pořadech. Byla definována vlastní evaluační databáze, která je pro danou úlohu reprezentativní a dokládá zkušenost autora a jeho týmu i v tvorbě dat. Popis vytvořených dat by bylo vhodné doplnit přehlednou tabulkou složení jednotlivých evaluačních úloh (textový popis není příliš přehledný). Bylo by také vhodné doplnit rozsah nebo histogram počtu nahrávek na jednoho mluvčího, což je pro specifikaci dat poměrně důležité. V implementační části bych doporučoval přesun metody ECA do teoretické části (mohla by být vložena jako předstupeň JFA), a konečně, není jasné, na čem byly trénovány parametry logistické regrese pro kalibraci. V sekci o MLLR-SVM systému jsem postrádal detailnější popis rozpoznávacího systému osahujícího MLLR adaptaci – je to plný rozpoznávací systém s jazykovým modelem nebo pouze fonémový rozpoznávač? Popsané výsledky systémů jsou velice dobré, ale v případě několika systémů je vždy nasnadě otázka: „jak dopadla fúze“ (především když ji autor v teoretické části popisuje).

Kapitola 7 uvádí popis systémů pro NIST SRE evaluace 2008 a 2010. Zde je především nutné ocenit úsilí autora – účast v evaluacích představuje značnou zátěž, především ve výběru trénovacích a vývojových dat, je zřejmé, že tato etapa byla úspěšně vyřešena. Pro českého čtenáře jsou opět velmi kvalitně shrnuty principy NIST evaluací včetně kompozice datových sad, evaluačních podmínek, atd. Byly postaveny 3 experimentální systémy: UGM/GMM, JFA a iVektorový, které jsou detailně popsány, chybí však detailnější popis detektoru řečové aktivity (VAD), který je pro funkcionalitu všech systémů klíčový. V závěru kapitoly postrádám detailnější hodnocení přesnosti jednotlivých systémů, především iVektorového, který za ostatními pokulhával – je to způsobeno absencí PLDA pro skórování iVektorů?

Kapitola 8 uvádí do velmi zajímavé problematiky diarizace pomocí klasického BIC přístupu a moderní techniky využívající opět iVektory. Kapitola je opět zpracována velmi kvalitně, i když by si zasloužila rozdělení mezi teoretickou a experimentální část. Techniky diarizace byly testovány na databázi projektu COST, která byla opět velmi pečlivě připravena, a vyhodnoceny pomocí standardní metriky. Velmi kladně hodnotím definici dvoufázové metody (nejprve BIC, pak iVektory), která je dobře zdůvodněna a poskytuje velmi slibné výsledky.

Kapitola 9 uzavírá práci, shrnuje její obsah, přínosy vědnímu oboru i přínosy pro praxi. Doporučoval bych pouze doplnění sekce „Budoucí práce“, kterou by bylo vhodné strukturovat do okamžitých a lehce dosažitelných cílů („low-hanging fruit“) a do cílů dlouhodobých či vizionářských.

2 Zhodnocení technické stránky práce

Práce dokládá, že kandidát pronikl do problematiky, byl schopen nastudovat velmi složitou teorii, naimplementovat množství systémů, provést množství experimentů, jejich výsledky porovnat mezi sebou a především je velmi zasvěceně diskutovat – analýza výsledků je v práci provedena (až na některé zmíněné mezery) velmi podrobně a fundovaně. Jeho výsledky jsou srovnatelné se světovým state-of-the-art v naší oblasti. S obdivem hodnotím systematickост autora při popisu teorie, kdy se byl schopen zorientovat v množství (často vzájemně se doplňujících) publikací a převést je do lineárního a velmi logického textu. Velmi kladně také hodnotím matematickou rigoróznost, včetně detailního odvození a uvádění potřebných pouček z lineární algebry pod čarou.

3 Formální stránka

Práce je psána výbornou češtinou – v dobách, kdy kvalita češtiny (i u vysokoškolských studentů) upadá, mi bylo čtení práce skutečným jazykovým potěšením. V některých místech bych však uvítal uvádění

standardních anglických termínů (zminěná sekce 4.4.3). Práce je logicky strukturována (až na některé prohršky, kdy je teorie míchána v experimenty, viz výše), a kvalita matematické sazby, obrázků a tabulek je prvotřídní. V práci jsem nenašel jediný překlep a objevil jedinou(!) chybějící čárku, což jsem ještě nikdy nezažil – autorovi patří dík za velmi pečlivou ediční práci.

4 Celkové zhodnocení

Předložená disertační práce je jednou z nejlepších, které jsem měl kdy k dispozici na posudek i čtení. Jejím kladem i záparem (současně) je čeština: na jedné straně značně omezuje mezinárodní dosah autorovy práce (jedná se o dílo, které by bylo uznávané i ve světovém měřítku), na straně druhé byla vytvořena dokonalá učebnice, která poslouží mnoha lokálním studentům a praktikům.

Doktorand publikoval jádro disertační práce v několika člancích na respektovaných zahraničních konferencích (Interspeech), je spoluautorem publikace v respektovaném časopise (SpeechCom) a prvním autorem dvou článků v nyní již impaktovaném Radioengineeringu. Jeho publikační činnost hodnotím jako rozsáhlou a kvalitní.

Z předložené práce, seznamu publikací i z mých osobních zkušeností s kandidátem vyplývá, že se jedná o pracovníka s velkým smyslem pro vědeckou práci: od teorie, přes implementaci až po provedení experimentů a diskusi jejich výsledků. Oceňuji také jeho otevřenost a schopnost zapojit se do aktivit lokální i mezinárodní vědecké komunity.

Závěr

Vzhledem k uvedeným skutečnostem prohlašuji, že podle mého názoru disertační práce uchazeče **odpovídá obecně uznávaným požadavkům k udělení akademického titulu Ph.D., doporučuji tuto práci k obhajobě, a pokud je to na TUL zvykem, doporučuji ji k udělení vyznamenání či zvláštního ocenění.**

Pro obhajobu navrhuji následující otázky:

1. Uveďte stručný popis LVCSR systému, který byl použit pro generování MLLR matic v MLLR-SVM systému (strana 98).
2. Byla pro česká data (kapitola 6) provedena fúze ? S jakými výsledky ?
3. Komentujte, proč nebyla použita v NIST iVektorovém systému PLDA, odhadněte výsledky s ní.
4. Uveďte, čím by Vaše práce mohla pokračovat v krátkodobém a dlouhodobém horizontu.

V Brně, 17.2. 2012



Doc. Dr. Ing. Jan Černocký
Speech@FIT, Ústav počítačové grafiky a multimédií
Vysoké učení technické v Brně, Fakulta informačních technologií
Božetěchova 2, 612 66 Brno
Tel: +420 5 41141284 Fax: +420 5 41141270, <mailto:cernocky@fit.vutbr.cz>
<http://www.fit.vutbr.cz/~cernocky> <http://www.fit.vutbr.cz/speech>

