21 V 039 707/2012

28 -11- 2012

Posudek na doktorskou disertační práci Ing. Lukáše Machlici
Vysokodimenzionální prostory a modelování
v úloze rozpoznávání řečníka

Ing. Lukáš Machlica předkládá disertační práci, která se věnuje problematice parametrizace signálů pro automatické rozpoznávače řečníka. Práce má 118 stran textu a přílohy.

Disertační práce se věnuje aktuálnímu tématu. Autor se zabývá algoritmy zpracování signálů, které generují vektory příznaků, použitelné při rozpoznávání řečníka. Jedná se o aktuální problematiku zpracovávanou v řadě publikací. Stále je oprávněná snaha ze samotného akustického signálu řeči vytěžit pro potřeby identifikace mluvčího více, než umožňují dosavadní postupy, resp., hledají se postupy efektivnější, jak z hlediska úspěšnosti rozpoznávání, tak z hlediska výpočetní náročnosti.

První kapitola práce uvádí do problematiky. Stručně jsou vymezeny základní pojmy z oblasti identifikace a verifikace mluvčích, a to ve všech specifických podmínkách (identifikace, verifikace, atd.). Jsou uvedeny algoritmické stránky procedur, jakými jsou parametrizace, modelování a klasifikace. Hned v této kapitole jsou popsány cíle práce, resp. proponované přínosy k problematice.

Takto pojatá první kapitola signalizuje, jaký koncept autor ve svém výkladu zvolil. Text nemá standardní strukturu, ve které se souhrnně popíše současný stav problematiky, ve zvláštní kapitole popíší cíle práce tak, aby mohly být v dalších statích naplněny, shrnuty a zhodnoceny. Setkáváme se tak s textem, který probírá téměř všechny známé postupy, které jsou dosud na cestě k rozpoznávání mluvčího používány. Jakoby „na pochodu" jsou jednotlivé procedury hodnoceny a po dílčích krocích zdokonalovány, a to včetně implementačních specifik (např. implementací v grafické procesorové jednotce).

Takže od druhé kapitoly má čtenář možnost sledovat jednotlivé procedury tvořící konstrukci systému. Autor důsledně zapisuje náročným matematickým aparátem všechny algoritmy, které považuje za relevantní pro svoji úlohu. Druhá kapitola je tak věnována konstrukci klasifikátorů. Bez hlubšího hodnocení zůstala klasifikace založená na neuronových sítích, nicméně její význam je v závěru kapitoly zmíněn. Třetí kapitola je věnována adaptačním algoritmům aplikovaným na parametrické popisy signálů. Hodnocení je motivováno hledáním optimálního popisu univerzálního modelu pozadí (UBM). Dále jsou popsány procedury mapování, normalizace a faktorové analýzy. Zájem je orientován směrem k vektorům založeným na všeobecně používaných modelech GMM a odvozeným supervektorům.

Sedmá kapitola je věnována experimentům. Je prezentována rozsáhlá dokumentace experimentů založených na identifikaci mluvčího z databáze promluv zaznamenaných v telefonní kvalitě. Z uvedených grafů je přehledně patrný efekt nejrůznějších variant výpočtů aplikovaných na data rozpoznávaného mluvčího v definovaném prostředí ostatních mluvčích. Shromážděné výsledky jsou vzájemně dokonale porovnatelné a snadno lze vysledovat, co v algoritmech přináší efekt.

Pozoruhodná je osmá kapitola, která popisuje implementaci některých výpočtů v grafickém procesoru. Autor ukazuje významné zrychlení výpočtů, pokud jsou implementovány ve vhodné hardwarové struktuře. Disertant by mohl při obhajobě stručně

popsat konkrétní hardwarovou konstrukci kooperujícího procesoru a způsob softwarové obsluhy.

V hodnocení výsledku autorova řešení implementace použitých algoritmů je uvedeno srovnání s jiným autorem jen co do rychlosti výpočtu. Bylo by vhodné ukázat v grafu DET obdobném k některému z kapitoly 7 srovnání s některým známým a dosud publikovaným výsledkem jiných autorů. Je zřejmé, že nalezení publikace, která by prezentovala skutečně plně srovnatelná data, může být obtížné. Nicméně alespoň odstavec textu s adekvátním popisem jinde dosažených výsledků by měl být v odkazu na práce zahraničních autorů věnován. Čtenář nemusí být vždy dokonale informován o obsahu srovnatelných publikací.

K charakteristice práce a jejího autora lze doplnit, že seznam literatury, kterou při studiu a zpracování tématu použil má 114 odkazů a ve vlastní bibliografii uvádí autorství a spoluautorství 17 kvalitních publikací.

Z uvedeného plyne, že recenzent jen obtížně hledal důvod k výhradám. Disertace dokumentuje samostatnou vědeckou práci. Autor předkládá vyzrálé vědecké zpracování velmi dobře vymezené problematiky. Vytvořil celou řadu přísně recenzovaných prací, které uspěly v konkurenci světové odborné komunity. Svými publikacemi prokázal nejen vědeckou erudici, ale i závažnost zpracovaného tématu.

Ing. Lukáš Machlica prokázal schopnost samostatné vysoce kvalifikované vědecké práce a naplnil podmínky nutné k tomu, aby mohl po úspěšné obhajobě obdržet titul doktora (Ph.D.). Práci doporučuji k obhajobě.

V Praze 22. listopadu 2012

Prof. Ing. Jan Uhlíř, CSc.

OPONENTSKÝ POSUDEK DISERTAČNÍ PRÁCE

Kandidát: Ing. Lukáš Machlica, Západočeská univerzita v Plzni
Název: **Vysokorozměrné prostory pro modelování v oblasti rozpoznávání mluvčího**
Recenzent: Ing. Pavel Matějka, PhD, FIT VUT v Brně

Předložená disertační práce Ing. Lukáše Machlici má 136 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 v proudu neoznačkovaných audio dat. Svým zaměřením je na hranici mezi elektrotechnikou a informačními technologiemi a obsahuje netriviální matematický komponent.

V první kapitole je uveden úvod do problematicky a autor přehledným způsobem shrnuje „state of the art" v oboru. Závěr kapitoly popisuje obsahy jednotlivých kapitol a definuje cíle práce.

Kapitola 2 obsahuje teoretický úvod a porovnání generativních a diskriminativních klasifikátorů hojně používaných v oblasti rozpoznávání mluvčích zejména Gausovské modely (GMM) a Support Vector Machines (SVM).

Kapitola 3 a 4 je více méně rozšíření či plynulé pokračování částí z kapitoly 2. Jedná se o adaptační techniky pro GMM a kernel funkce pro SVM. Velice hodnotím porovnání několika různých technik na jednom místě a jejich porovnání mezi sebou z teoretického hlediska.

Kapitola 5 je věnována normalizaci příznaků v různé hloubce systému. Začíná na normalizaci řečových příznaků a končí u po-zpracování statistik z GMM. V této kapitole se poprvé objevují techniky pro kompenzaci nepříznivých vlivů přenosového kanálu na rozpoznávání mluvčích. Jedná se o Nuisance Atribute Projection (NAP), která byla vyvinuta zhruba před 6ti lety.

Kapitola 6 obsahuje teoretické jádro práce a navazuje na kapitolu 2 složitějšími modelovacími technikami. Začíná u Faktorové Analýzy (FA) končí u velmi progresivní metody iVektorů následovaných PLDA (Probabilistic Linear Discriminant Analysis), která byla definována v několika posledních letech. Autor si dal značnou práci s konsolidací poznatků roztříštěných v několika desítkách článků (především od Patricka Kennyho, který s těmito technikami v rozpoznávání mluvčích začal), výsledkem je velmi kompaktní text, který pokrývá teoretické i praktické aspekty JFA a iVektorových systémů. Kapitola je navíc doplněna matematickým odvozením vztahů, porovnání mezi jednotlivými technikami a „kuchařkou" pro implementaci systémů. Autor navrhuje několik vylepšení hlavně na rychlost trénování systému, což v dnešní době, kde tyto techniky jsou tzv. „hladové po datech" nebo-li čím víc jim předhodíme tím lepší výsledky dostáváme, je opravdu velmi důležité.

Kapitola 7 obsahuje podkapitoly od popisu databází přes extrakci řečových příznaků a modelování po výsledky a porovnání jednotlivých metod. Velice oceňuji, že autor používá dnes už standardní data z NIST evaluací (pořádaných Národním Institutem pro Standardy a Technologie, USA), protože to umožňuje jednoduché porovnání s ostatními vývojovými skupinami na světě.
Zde postrádám výsledky systému založeném na JFA, který považuji jako vývojový krok mezi GMM a iVektory a které byly dopodrobna popsány v teoretické části. Dále v kapitolách 7.7.1 a 7.7.2 autor uvádí experimenty s různými databázemi použitých pro trénování PLDA, o kterých si myslím že měli být

sloučeny do jedné kapitoly. Navíc mohli být uvedeny podrobnější výsledky pro každou databázi zvlášť. Zajímavá je také analýza kolik opravdu mluvčích je potřeba k saturaci výsledků pro PLDA, je toto číslo stejné pro jednotlivé databáze? Kapitola 7.8 porovnává všechny zástupce důležitých systémů zmíněných v této práci, mohli zde být i výsledky na stejných datových setech publikovaných na mezinárodních konferencích, čímž autor nevyužil výhodu použití NIST dat k porovnání s ostatními skupinami pracujícími ve stejném oboru.

Kapitola 8 pojednává o implementaci některých částí kódu na grafické karty od jejich návrhu po testování. Testy ukazují zrychlení několik řádů. Tato oblast se dostává velice do popředí, protože například nyní pro NIST SRE 2012 zpracováváme pro každý experiment půl milionu souborů.

Kapitola 9 uzavírá práci, shrnuje její obsah, přínosy vědnímu oboru i přínosy pro praxi i nastiňuje budoucí směry. Jeden z těchto směrů je opravdu důležitý – máme obrovské množství dat a slepě na něm trénujeme GMM, které nám obecně popíše tento prostor místo toho abychom se snažili nějak popsat prostor s informací o mluvčím a tím hned na začátku extrahovat informaci užitečnou k rozpoznávání mluvčích, ale jak to udělat?

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. S obdivem hodnotím systematičnost 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.

3 Formální stránka

Hodnotím velice kladně, že práce je psána anglicky, protože si ji může přečíst více lidí než 10 (z toho školitel a dva oponenti), kteří se touto tématikou v zabývají v České republice.
Práce je logicky strukturována, a kvalita matematické sazby, obrázků a tabulek je prvotřídní. Mám jen jedinou výtku a to je nadměrné používání zkratek.

4 Celkové zhodnocení

Předložená disertační práce se opravdu dobře četla a jednotlivé kapitoly na sebe logicky navazovali.
V experimentální části mohl jít autor v některých ohledech ještě více do hloubky (viz výše).

Doktorand publikoval jádro disertační práce v několika článcích na respektovaných zahraničních konferencích jako je Interspeech, ICASSP, Odyssey a další. 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.**

Pro obhajobu navrhuji následující otázky:

1. Proč jste zvolil pro trénování kompenzace kanálu u NAP databáze Switchboard a ne NIST data, o kterých by se dalo předpokládat, že budou blíže cílovým datům?
2. Ve Vaší práci jsem nenašel mnoho o kalibraci skóre. Je sice obecně známo, že výsledky s PLDA jsou relativně dobře kalibrované, mohl byste přesto vysvětlit proč chceme v praxi dobře kalibrované skóre?

V Brně, 17.10. 2012

Ing. Pavel Matějka PhD
Speech@FIT, Ústav počítačové grafiky a multimédií
Vysoké učení technické v Brně, Fakulta informačních technologií
Božetěchova 2, 612 66 Brno
Tel: +420 5 41141283 Fax: +420 5 41141270, mailto:matejkap@fit.vutbr.cz
http://www.fit.vutbr.cz/~matejkap http://www.fit.vutbr.cz/speech