



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. Hichama Atassiho

ROZPOZNÁNÍ EMOČNÍHO STAVU Z HRANÉ A SPONTANNÍ ŘEČI

Předložená disertační práce představuje poměrně rozsáhlou komplexní studii v anglickém jazyku v rozsahu 140 stran včetně příloh a autora odborného životopisu. Nosné téma práce – detekce a klasifikace emocionálního stavu mluvčího z řečového signálu – je velmi aktuální a perspektivní a to jak z hlediska současného vývoje a stavu vědeckého poznání v této oblasti, tak především z pohledu praktické aplikovatelnosti. Rozpoznávání emočního stavu člověka představuje významnou součást komunikačního rozhraní člověk-stroj hlasových informačních systémů a to především při tvorbě vhodných dialogových strategií s cílem maximálního přiblížení se lidskému projevu (přirozenost, uživatelská příjemnost a i rovněž efektivita vedeného dialogu). Praktickou využitelnost vedle popisovaného vyhodnocování provozu v call-centrech, spatřuji rovněž v systémech monitorování, detekce a předcházení krizových situací na linkách integrovaného záchranného systému, a v neposlední řadě také v oblasti kompenzačních pomůcek pro nevidomé a slabozraké.

Posuzovaná práce se kromě úvodu a závěrečného shrnutí skládá ze šesti kapitol mapujících současný stav problematiky v oblasti detekce a klasifikace emocionální řeči, metodiky návrhu a výběru klasifikačních příznaků, analýzy a zpracování řečového signálu obecně. Kapitoly 7-9 představují realizační část: autor se nejprve věnuje popisu použitých databází emocionální řeči, dále pak práce obsahuje presentaci a diskusi dosažených výsledků z provedených experimentů. Za jádro práce považuji provedený popis aplikovaných technik detekce spolu s rozsáhlejším rozбором hodnocení výkonnosti jednotlivých algoritmů včetně analýzy a diskuse nepříznivých vlivů působících na výslednou účinnost a stabilitu. V závěru následuje shrnutí dosažených výsledků a nástin dalších možností výzkumu a vývoje. Přílohy A-B obsahují základní popis používaných SW nástrojů pro platformu Matlab k určování a výpočtu řečových příznaků, v příloze C je uveden stručný popis autonomního systému pro vyhodnocování provozu v call-centru pracujícího v reálném čase. Práce bohužel neobsahuje doprovodný datový nosič s ukázkami základních nahrávek z obou používaných řečových databází ani jiný programový produkt (např. DEMO verze programů uváděných v přílohách), takže nebylo možné provést posouzení jejich funkčnosti a ověřit tak výsledky uváděné autorem v experimentální části.

Z formálního pohledu je provedení práce na velmi dobré úrovni – po grafické a lexikální stránce obsahuje relativně málo překlepů a nepřesností. Práce je dobře strukturovaná a logicky členěná, což napomáhá relativně dobré čitelnosti a srozumitelnosti jinak hutného

odborného textu. Autor pracoval celkem s 111 prameny literatury včetně svých publikací pokrývající celou řešenou problematiku. Citační odkazy jsou používány správně, výběr použitých zdrojů literatury je vhodný pro tento typ práce. V teoretické části práce autor prokázal že se dobře orientuje ve stavu řešené problematiky na světové i domácí úrovni. Navíc ovládá potřebný matematický aparát k pochopení a realizaci jednotlivých analytických a statistických úloh pro řešení stanovených cílů práce.

Mám však několik připomínek týkajících se způsobu provedení práce:

- V úvodu je správně začleněna tabulka zkratk, bohužel však chybí tabulka matematických symbolů přestože práce obsahuje značné množství rovnic a definičních vztahů. Na druhou stranu uvádění některých obecně známých definic (např. rovnice 3.1, 3.2) považuji za zbytečné.
- Za větší nedostatek považuji fakt, že si autor zavádí vlastní symboliku (zkratky), které jsou rozdílné od obecně používaných v vědecké literatuře z tohoto oboru (např. BDES vs. EmoDB na str. 48 a hlavně pak zavedení pojmu „high-level feature“ v kontextu supra-segmentálních příznaků na str. 73) což spolu s vícenásobnou re-definicí některých zkratk je dosti zavádějící. Za zcela matoucí pak považuji popisku tabulky 7.12 na str. 74 – prosím o objasnění jak byly ve skutečnosti jednotlivé řečové příznaky definovány a vypočítávány.
- Text dále neobsahuje dostatečně jasnou specifikaci toho, co byla vlastní práce disertanta a co bylo provedeno společně – zejména v kontextu s autorovou publikační a realizační činností. To se týká i popisovaných řečových databází (vytvářených ve spolupráci s firmou Retia) – podílel se Ing. Atassi rovněž na jejich tvorbě? Dále nelze jednoznačně určit zda autor používal již hotové SW nástroje „Hila“ a „GoEmotionally“, případně se i sám podílel na jejich vývoji.
- V úvodní části práce pak postrádám přehledně definované cíle a v závěru popis jejich konkrétního splnění; ten obsahuje jen obecné shrnutí dosažených výsledků úspěšnosti klasifikace/detekce.

Z přiložených doprovodných materiálů vyplývá že disertant začal postgraduální studium v 2007, přičemž presentované publikační výstupy jsou z období 2008-2013. V tomto kontextu je třeba vyzdvihnout, že se systematicky věnoval problematice zpracování a analýzy emocionální řeči. V letech 2009-10 absolvoval dlouhodobou stáž na Univerzitě ve Stirlingu a dále 3 krátkodobé zahraniční stáže: 2008 a 2009 na International Institute of Advanced Scientific Studies v Salernu, a v roce 2010 na Univerzitě Pierre et Marie Curie v Paříži. Z pohledu publikačních aktivit lze rovněž kladně hodnotit, že se disertant podílel celkem na 22 publikacích – s výjimkou 2 článků se jedná o příspěvky na zahraničních i domácích konferencích a workshopech (včetně 4 článků pro internetový časopis Elektrověv) – vše ale výhradně ve spoluautorství. To je z hlediska bodového výsledného hodnocení více než dostačující, přesto bych však ocenil větší zastoupení zahraničních, čistě časopiseckých publikací. Podle citační databáze SCOPUS má disertant celkem 16 ohlasů na 15 publikací (h index=3), pomocí databáze ISI WEB of Knowledge se mi podařilo nalézt celkem 12 citačních ohlasů. Ing. Atassi je dále spoluautorem celkem 6 SW produktů a jednoho patentu,

participoval na řešení 7 vědeckých projektů převážně z oblasti nosné problematiky zpracování řeči a navíc byl sám odpovědným řešitelem (nositelem) jednoho grantu pro zavedení problematiky rozpoznávání vzoru do pedagogické výuky. Z uvedeného výčtu a rozboru publikačních a tvůrčích aktivit disertanta lze proto konstatovat, že jádro předložené disertační práce bylo publikováno na dostatečné úrovni a že disertant představuje osobnost uchazeče s vysokou vědeckou erudicí a především s realizačními a aplikačními schopnostmi.

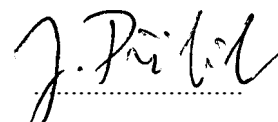
Dotazy do diskuse v průběhu obhajoby:

- Z popisu multi-linguální databáze spontánní emociální řeči v úvodu kap. 7 vyplývá, že subjektivního hodnocení slovenštiny se zúčastnili pouze 3 posluchači (hodnotitelé) muži a dokonce jen 2 ženy. To považuji ze statistického hlediska za zcela nepostačující, odpovídají tomu i uváděné výsledky ve formě konfúzní matice v tabulce 7.3, kdy průměrná úspěšnost byla dosažena pouze 35 %. V následně provedených experimentech byly použity nahrávky „ve slovanských jazycích“ a bylo uvedeno tvrzení, že takto získané výsledky je možné generalizovat pro všechny ostatní databáze (viz úvod sekce 7.2). V dalším textu je pak konkretizováno, že se jednalo o experimenty pouze s nahrávkami v češtině, polštině a ruštině (str. 75). Prosím o ozřejmění těchto nejasností = jak se prováděl výběr analyzovaného jazyka, podle jakých kritérií?
- Kapitola 8 je věnována analýze vlivu emocionálního stavu mluvčího na úspěšnost klasifikace jeho pohlaví (muž-žena). Zajímalo by mne zda také věk (stáří mluvčího) má vliv na správnost určení emocionálního stavu řečového projevu, zejména v kontextu s popisovaným systémem pro automatickou detekci úspěšných hovorů v call-centrech, který obsahuje též blok detekce věku volajícího (obsluhy).

Závěr:

Přes dílčí připomínky předložená disertační práce splňuje podmínky kladené na tvůrčí vědeckou práci. Za hlavní její přínos považuji provedení návrh, vývoj a úspěšné otestování nové metody pro mapování diskrétních emočních stavů do dvou dimenzionálního prostoru, dále komplexní analýzu výběru řečových příznaků používaných pro klasifikační algoritmy k rozpoznávání emocí v „hrané“ řeči a určování emocí ze spontánní řeči (z nahrávek dialogů pořízených v call-centrech). Proto práci **doporučuji k obhajobě** a zároveň **navrhuji** udělit Ing. Hichamu Atassimu akademicko-vědecký titul **doktor** v oboru TELEINFORMATIKA.

V Bratislavě, 29.9. 2014.


Dr. Ing. Jiří PŘIBIL