dr hab. Zygmunt Mazur, prof. PWr.
Instytut Informatyki
Politechniki Wrocławskiej
Wyb. Wyspiańskiego 27
50–370 Wrocław
tel./fax: (+71) 320 42 23
e–mail: zygmunt.mazur@pwr.wroc.pl

Wrocław, 25 sierpnia 2014 r.

RECENZJA ROZPRAWY DOKTORSKIEJ

Tytuł rozprawy:

„Rozpoznawanie mowy na podstawie mowy naturalnej"

Autor rozprawy: **mgr inż. Dorota Kamińska**

Promotor rozprawy: **dr hab. inż. Adam Pelikant, prof. nzw. P.Ł.**

1. Charakterystyka wyboru tematu i przedmiot rozprawy

Oprócz informacji przekazywanych w mowie naturalnej są także przekazywane dodatkowe informacje dotyczące stanu emocjonalnego mówcy. Tekst mówiony posiada ponadto także pewien niejawny przekaz, taki jak emocje mówcy. Ważnym elementem każdej naturalnej konwersacji jest także ocena stanu emocjonalnego rozmówcy, który możemy również z niej odczytać. W rozmowie naturalnej wydobycie obu rodzajów informacji znacznie poprawia jakość komunikacji interpersonalnej.

W wyniku dynamicznego wzrostu mocy komputerów komputery, które stały się częścią naszego codziennego życia, poszukuje się nowych rozwiązań mających na celu polepszenie komunikacji człowiek–komputer / człowiek–robot (HCI/HRI). Coraz więcej aplikacji obsługujących różne urządzenia sterowanych jest głosem, niekoniecznie trzeba nadal używać pilota czy klawiatury. Dlatego też powstają nowoczesne technologie rozpoznawania ludzkiej mowy.

Analiza emocji znajduje zastosowanie w syntezatorach głosu oraz w systemach wspomagających rozpoznawanie mowy. Dodatkowo istotną dziedziną zastosowań jest medycyna – diagnoza zaburzeń psychologicznych i neurologicznych, objawiających się nieprawidłową percepcją i ekspresją emocji (autyzm, schizofrenia, depresja, stres) oraz wspomaganie terapii behawioralnej.

Tematyka rozprawy doktorskiej dotyczy rozpoznawania stanów emocjonalnych wyrażanych głosem. Przedstawiona w rozprawie tematyka badań jest obecnie bardzo ważna i celowa. Doktorantka skupiła się głównie na mowie spontanicznej,

przedstawiła nowatorski sposób klasyfikacji emocji spontanicznych stosując przy tym zarówno powszechnie znane deskryptory sygnału mowy jak i percepcyjne współczynniki hybrydowe, dotychczas niewykorzystywane do opisu emocji.

2. Zakres i tematyka rozprawy

We wstępie rozprawy zostały przedstawione tezy badawcze, ściśle powiązane z zagadnieniami klasyfikacji mowy emocjonalnej i brzmią one następująco (cytat ze str. 7):

Teza 1: Wykorzystanie hybrydowych współczynników percepcyjnych w procesie klasyfikacji dokonywanym przy użyciu komitetu klasyfikatorów pozwala na uzyskanie wysokiej skuteczności rozpoznawania emocji na podstawie mowy naturalnej.

Teza 2: Opierając sie na teorii emocji Plutchika można dokonać klasyfikacji emocji wtórnych.

Teza 3: Istnieje możliwość zwiększenie dokładności rozpoznawania emocji zawartych w głosie poprzez minimalizację cech osobniczych mówcy.

Przedstawione tezy zostały zweryfikowane w rozprawie doktorskiej.

3. Konstrukcja rozprawy

Recenzowana praca składa się z czterostronicowego wstępu, siedmiu rozdziałów, pięciostronicowego zakończenia oraz bibliografii liczącej 90 pozycji. Rozprawa liczy 107 stron.

Rozprawa doktorska została podzielona na trzy podstawowe części. Pierwszą z nich stanowią rozdziały, które zawierają teoretyczne wprowadzenie do analizowanego tematu oraz przegląd aktualnego stanu wiedzy. Drugą część stanowią rozdziały opisujące zaproponowane przez doktorantkę rozwiązanie zagadnienia. Trzecia część, to rozdziały prezentujące weryfikacje zaprezentowanych autorskich algorytmów, przeprowadzone badania oraz wnioski i propozycje kontynuacji prac nad rozpoznawaniem emocji na podstawie mowy naturalnej.

W rozdziale 2 opisano taksonomię stanów emocjonalnych, ich źródeł i korelatów, rozdział jest opisem stanów emocjonalnych. Zaprezentowano zarówno mechanizm powstawania emocji jak i jego korelację z procesami zachodzącymi w ludzkim ciele, ze szczególnym uwzględnieniem sygnału mowy. Dodatkowo zaprezentowano psychologiczne modele emocji oraz ich taksonomię. W rozdziale 3 przedstawiono automatyczne rozpoznawanie emocji na podstawie sygnału mowy – opisano aktualny stan wiedzy, rozdział jest podsumowaniem prac badawczych dotyczących analizowanego tematu. Skrótowo opisane zostały szeroko eksploatowane korpusy mowy emocjonalnej, metody minimalizacji cech osobniczych, parametryzacja mowy oraz jej klasyfikacja. Rozdział 4 zawiera podstawy teoretyczne niezbędne do zrozumienia opisywanego zagadnienia, a więc metody analizy sygnału mowy, jego parametryzację oraz algorytmy ekstrakcji i selekcji cech. W rozdziale 5 doktorantka

opisała autorską bazę próbek mowy emocjonalnej stworzonej na potrzeby badan. Przedstawiono kolejne etapy tworzenia korpusu: gromadzenie próbek, etykietowanie, tworzenie zbioru treningowego i testowego. Dodatkowo zaprezentowano korpus mowy odegranej wykorzystywany w niniejszych badaniach w celach porównawczych. Wskazano różnice i podobieństwa pomiędzy mową spontaniczną i odegraną. W rozdziale 6 zaprezentowano proces tworzenia algorytmu rozpoznawania, którego bazę stanowi Podstawowy Algorytm Klasyfikacji Emocji PAKEmo. Następnie, w trakcie badan, uzupełniano go o dodatkowe elementy (podproblemy), mające na celu podniesienie jakości klasyfikacji. Ostateczny algorytm stanowi wielopoziomowy komitet klasyfikatorów, budowany poprzez dodawanie kolejnych poziomów podproblemów, wykorzystujący prosty klasyfikator bazowy w węźle. Taki mechanizm umożliwia budowę silnego klasyfikatora poprzez dodawanie kolejnych poziomów podproblemów, które w tym przypadku stanowią, między innymi: rozpoznawanie płci, budowa profili emocjonalnych, zależność emocji od długości wypowiedzi oraz badanie natężenia emocji. Wszystkie wyżej wymienione zadania zostały kolejno zaprezentowane i szczegółowo omówione. W rozdziale 7 przedstawiono wyniki doświadczeń przeprowadzonych na puli cech powszechnie wykorzystywanej w rozpoznawaniu emocji (częstotliwość podstawowa, formanty, energia sygnału, współczynniki MFCC, PLP, RASTA PLP oraz LPC). Dodatkowo, póle tę poszerzono o współczynniki percepcyjne, takie jak BFCC, HFCC oraz RPLP, szeroko stosowane w badaniach nad rozpoznawaniem mowy, natomiast pomijane w rozpoznawaniu emocji. Przy użyciu selekcji cech wyłoniono najbardziej dyskryminatywne atrybuty. Rozdział 8 przedstawia wyniki klasyfikacji poszczególnych kroków algorytmu, jak i całej struktury hierarchicznej. W celach porównawczych eksperymenty przeprowadzono przy użyciu dwóch korpusów mowy emocjonalnej: autorskiej bazie mowy spontanicznej oraz bazie mowy odegranej. Otrzymane wyniki przeanalizowano i podsumowano. Rozdział 9 stanowi konkluzję przeprowadzonych badań, prac projektowych, implementacyjnych oraz wyników wykonanych eksperymentów. Przedstawione zostały wady i zalety proponowanego rozwiązania oraz dalsze plany autorki związane z rozwojem i możliwością zastosowania stworzonego systemu.

4. Ocena merytoryczna rozprawy

Praca doktorska ma charakter eksperymentalny, zdefiniowano problemy naukowe i przedstawiono tezy rozprawy. Precyzyjnie zaplanowano szereg badań eksperymentalnych pozwalających udowodnić wykreowane we wstępie do rozprawy tezy naukowe. Planowanie eksperymentów i analiza ich wyników pomagają odkryć rzeczywiste związki. W wyniku zebrania opisów wielu obiektów rodzą się pytania naukowe dotyczące wytłumaczenia, „dlaczego" opisywany obiekt jest taki a nie inny, jak wytłumaczyć jego "strukturę", jak funkcjonuje, jak się zmienia itd. Przeprowadzenie samych eksperymentów nie byłoby możliwe bez wcześniejszych badań opisowych i nie byłoby później co testować.

Głównym założeniem zaprezentowanych w rozprawie wyników badań eksperymentalnych była realizacja systemu informatycznego pozwalającego na automatyczne rozpoznawanie stanów emocjonalnych na podstawie mowy naturalnej. W tym celu przygotowano polską bazę emocji spontanicznych. Ponadto w celach

porównawczych dokonano również analizy emocji odegranych przez profesjonalnych aktorów.

Ilościowy opis problemu stanowią powszechnie używane w tego typu badaniach deskryptory mowy, które zestawiono z hybrydowymi współczynnikami percepcyjnymi. Jak wykazały badania, atrybuty te okazały się silnie dyskryminatywne, co uzasadnia ich użycie. W trakcie klasyfikacji porównano algorytm k-NN z autorskim podejściem opartym na zbiorze klasyfikatorów (komitecie), mającym zapewnić lepsze wyniki rozpoznawania. Analiza wyników potwierdziła początkowe założenia doktorantki.

Wyniki z analizy badań wpływu korpusu mowy spontanicznej na rozpoznawanie emocji przeprowadzonych na dwóch bazach, porównujące emocje odegrane z naturalnymi, wykazały jak duży wpływ na wyniki klasyfikacji mają korpusy tworzące wzorce, co istotnie utrudnia porównywanie skuteczności różnych, zaproponowanych dotychczas podejść. Duże znaczenie ma przede wszystkim liczność wzorców. Odpowiednia liczba i różnorodność przykładów próbek w znacznym stopniu zwiększa jakość rozpoznawania. W przypadku mowy spontanicznej różnorodność próbek (płeć oraz wiek mówcy) ma wpływ na lepsze wyniki klasyfikacji. Poprzez wykorzystanie wypowiedzi różnego typu, ograniczany jest wpływ cech osobniczych na rozpoznawanie. Aby uwzględnić różnice w sposobie ekspresji emocji przez kobiety i mężczyzn, doktorantka wprowadziła moduł rozpoznawania płci. W przypadku mowy spontanicznej spowodowało to poprawę wyników rozpoznawania. Obniżenie wydajności klasyfikatora w przypadku mowy odegranej może wiązać się z ograniczeniem liczności wzorców po podziale na płeć.

W badaniach dokonano również klasyfikacji natężeń mowy emocjonalnej. Zaprezentowano autorski algorytm określenia intensywności danej emocji na podstawie stopnia jej podobieństwa do mowy neutralnej. Zadanie to również wydaje sie być istotnym: rozróżnienie, czy mówca jest lekko podirytowany, czy tez mocno rozwścieczony, ma znaczenie, a w szczególności w zastosowaniach aplikacyjnych. Biorąc pod uwagę rozmyte granice między konkretnymi natężeniami danej emocji, w przyszłych badaniach należałoby przetestować różnego typu funkcje przynależności do rozpoznania konkretnego natężenia.

Przeprowadzona analiza otrzymanych wyników pokazuje złożoność mechanizmów powstawania emocji, ich percepcji i ekspresji. W naturalnym środowisku mówca w tym samym momencie może być pod wpływem różnych emocji, a słuchacz może różnie odbierać wysyłane sygnały. Wzorce mowy odegranej mogą nie sprawdzać się przy klasyfikacji emocji w warunkach naturalnych. Otrzymane wyniki badań wskazują, że zaproponowany algorytm radzi sobie również z próbkami niejednoznacznie określonymi.

Proces wyznaczania odpowiednich atrybutów, które trafnie opisują przedmiot analizy, ma ogromne znaczenie w zadaniach rozpoznawania wzorców. Algorytm klasyfikacji musi być poprzedzony procesem doboru wydajnych zestawów cech oraz procesem ich selekcji. Badania przeprowadzone na konkretnych podzbiorach cech wskazują, ze w przypadku obu korpusów najwyższe wyniki rozpoznawania osiągana są właśnie przy użyciu zaproponowanych atrybutów. Podano wartości wielu współczynników potwierdzających poprawę rozpoznawania emocji.

Proces klasyfikacji opiera sie na znanych już standardowych narzędziach rozpoznawania, wraz z rosnąca złożonością żądań pojawia się potrzeba projektowania nowych rozwiązań, mających na celu zapewnienie lepszej skuteczności. W tym celu tworzone są całkowicie nowe klasyfikatory, metody hybrydowe, łączące poszczególne algorytmy, a także metody usprawniające istniejące już rozwiązania. W pracy zaproponowano algorytm oparty o teorię komitetów, które pracując wspólnie, osiągają wyniki lepsze niż pojedyncze modele Przedstawione w rozprawie rozwiązanie jest całkowicie innowacyjnym podejściem. Modele oparte na atrybutach wybranych do reprezentacji emocji popełniają różne bledy dla nowych danych, a zatem można mówić o różnorodności komitetu. Dodatkowo w trakcie badań doktorantka zauważyła, że dla określonych podzbiorów najlepsze wyniki osiągane są przy użyciu rożnych wartości liczby k algorytmu k-NN. Rozbicie pojedynczego modelu na zbiór klasyfikatorów, z którego każdy dokonuje rozpoznawania na podstawie innego podzbioru atrybutów, a ostateczna decyzja podejmowana jest na podstawie głosowania, prowadzi do zwiększenia jakości rozpoznawania.

W związku z tym, ze każdy z podzbiorów atrybutów ma inny wkład w rozpoznawanie, następnym krokiem było zastąpienie głosowania równoważnego ważonym. Wagi dobrano na podstawie bledu konkretnego modelu, a ich wprowadzenie do algorytmu głosowania ostatecznie uzasadnia użycie zaproponowanego rozwiązania.

Doktorantka doskonale zdaje sobie sprawę ze złożoności analizowanych problemów badawczych, a otrzymane wyniki w rozprawie nie rozwiązują wszystkich zagadnień w zakresie rozpoznawania mowy naturalnej. Doktorantka wyraźnie określiła aktualny stan swoich badań i wskazała kierunki kontynuacji dalszych prac. Naturalnym kierunkiem badań jest użycie innych algorytmów klasyfikacji i sprawdzenie ich możliwości, jako modeli bazowych komitetu, dopasowując odpowiedni algorytm rozpoznawania do konkretnego podzbioru cech. W przyszłości można testować inne kombinacje komitetu stosując różne warunki podziału modelu bazowego. Kolejnym kierunkiem rozwoju algorytmu jest określenie dodatkowych cech opisujących przedmiot analizy. Badany algorytm można poszerzyć o dodatkowe modele, bazujące na prozodiach sygnału (tempo, pauzy) czy tez atrybutach wyznaczanych na podstawie opisu sygnału metodami zaczerpniętymi z analizy układów nieliniowych. Można także poszerzyć korpus o próbki wypowiedzi dzieci czy rozpoznawanie stanów niejednoznacznie określonych.

Potencjalne możliwości rozwoju pozwalają na dalsze prace nad rozpoznawaniem stanów emocjonalnych na podstawie sygnału mowy. Obiecującym krokiem może być tworzenie dodatkowych modeli bazujących na innych sygnałach: obraz (mimika oraz gesty), sygnały EEG czy analiza obrazu w podczerwieni. Komitet stworzony na podstawie dodatkowych przesłanek może w znacznym stopniu poprawić klasyfikację.

5. Uwaga krytyczna

W rozdziale I. "Wprowadzenie", na stronie 7 rozprawy doktorantka sformułowała trzy tezy badawcze określające podstawowe cele pracy doktorskiej. W zakończeniu rozprawy brak konkluzji o zrealizowaniu celu rozprawy, wykazaniu

prawdziwości tez, z odwołaniami do odpowiednich fragmentów rozprawy. Końcowa konkluzja jest zbyt lakoniczna. W tekście rozprawy konkluzja doktorantki sprowadziła się tylko do jednego podsumowującego, krótkiego, jednozdaniowego stwierdzenia "*Badania udowodniły tezę postawioną przez autorkę*" (Rozdział "Podsumowanie", 92 str., 9 wiersz od góry).

Powstaje pytanie: o którą tezę chodzi?, a co z pozostałymi dwoma?
W tekście rozprawie brakuje formalnych dowodów trzech tez rozprawy.

6. Podsumowanie i wniosek końcowy

W podsumowaniu recenzji stwierdzam, że przedstawiona do oceny rozprawa doktorska zawiera oryginalne i wartościowe wyniki rozwiązania problemu naukowego, stanowiące znaczący wkład do nowoczesnej inżynierii projektowania systemów informatycznych w zakresie systemów rozpoznawania mowy.

Tematyka rozprawy doktorskiej dotyczy rozpoznawania stanów emocjonalnych wyrażanych głosem w mowie naturalnej. Przedstawiona w rozprawie tematyka badań jest bardzo ważna, ciekawa i celowa. Doktorantka skupiła się głównie na mowie spontanicznej, przedstawiła nowatorski sposób klasyfikacji emocji spontanicznych stosując przy tym zarówno powszechnie znane deskryptory sygnału mowy jak i percepcyjne współczynniki hybrydowe, dotychczas niewykorzystywane do opisu emocji. Przeprowadziła wiele badań eksperymentalnych. Na drodze badań empirycznych wykazała poprawność zaproponowanych metod rozpoznawania emocji.

Doktorantka wykazała się dużą wiedzą teoretyczną, bardzo dobrą znajomością literatury oraz wiedzą praktyczną w tej dyscyplinie naukowej, w zakresie rozpoznawania mowy i wykazała się umiejętnościami do samodzielnego prowadzenie pracy naukowej.

W tekście rozprawy doktorantka przedstawiła wyniki przeprowadzonych badań, częściowo zaprezentowanych już na konferencjach i czasopismach naukowych (5 współautorskich publikacji).

Strona edytorska i szata graficzna rozprawy nie budzą zastrzeżeń.

Z uwagi na osiągnięte wyniki w rozprawie doktorskiej i spójną jej zawartość stwierdzam, że rozprawa doktorska pt.: „Rozpoznawanie emocji na podstawie mowy natyralnej" spełnia wymagania stawiane rozprawom doktorskim określonym w art. 13, ust. 1 i ust. 2 Ustawy z dnia 14 marca 2003 r. o stopniach i tytule naukowym i wnoszę o dopuszczenie mgr inż. Dorotę Kamińską do dalszych etapów przewodu doktorskiego w dziedzinie nauk technicznych, w dyscyplinie informatyka.

Zygmunt Mazur