

Prof. dr hab. inż. Adam Dąbrowski

Politechnika Poznańska

Wydział Informatyki

Instytut Automatyki i Sterowania

Zakład Układów Elektronicznych

i Przetwarzania Sygnałów

OCENA

rozprawy doktorskiej pt.: *„System automatycznego rozpoznawania mowy oparty na analizie cepstralnej sygnału mowy i modelach mieszanin gaussowskich”*

Pana **mgra inż. Kamila Kamińskiego**

1 Ocena doboru tematu oraz celu i zakresu przeprowadzonych badań (w odniesieniu do tez rozprawy)

Tematem ocenianej rozprawy doktorskiej jest opracowanie oprogramowania do rozpoznawania mowy (w skrócie ASR od ang. automatic speaker recognition) na podstawie głosów zarejestrowanych z jakością telefoniczną. Doktorant zajął się w szczególności wielokryterialnym ulepszaniem (w tym optymalizacją) wcześniejszych wariantów oprogramowania opracowanego w Wojskowej Akademii Technicznej jak i przeprowadzeniem licznych testów przy wykorzystaniu zarówno komercyjnych jak i lokalnie nagranych baz wypowiedzi.

Celem podjętych badań, jak pisze Autor rozprawy już w pierwszym zdaniu jej streszczenia, „było opracowanie systemu automatycznego rozpoznawania mowy, pozwalającego na skuteczną identyfikację głosów w utrudnionych warunkach rejestracji, zbliżonych do warunków transmisji telefonicznej”.

Teza naukowa, sformułowana w ostatnim zdaniu streszczenia oraz we wprowadzeniu na str. 9, jest następująca: „możliwe jest takie przetworzenie sygnału mowy, w wyniku którego można utworzyć oszczędne pamięciowo modele głosów, pozwalające na identyfikację mowy z wysoką skutecznością”.

Mimo, iż rozważne w rozprawie zagadnienia, należą do już dobrze rozpoznanych, to tematykę rozprawy zwłaszcza w zakresie ulepszania i testowania systemu rozpoznawania mowy na podstawie nagrań o jakości telefonicznej, bez zakładania ustalonych tekstów, oceniam wysoko. Uważam bowiem, że to jest nadal aktualny problem naukowy o dużym znaczeniu aplikacyjnym.

Pan mgr inż. Kamil Kamiński podjął się więc opracowania ciekawego i ważnego problemu naukowego. Zarówno określony w rozprawie cel badań jak i sformułowaną tezę naukową oceniam bardzo pozytywnie. Zakres przeprowadzonych i przedstawionych w ocenianej rozprawie doktorskiej badań jest moim zdaniem dobrany właściwie i adekwatnie do postawionego celu i do podanej tezy naukowej.

2 Charakter rozprawy w tym ocena tekstu i sposobu przedstawienia wyników

Recenzowana rozprawa została napisana w języku polskim. Jej tekst zawiera 119 stron i oprócz umieszczonych na samym początku: dedykacji, spisu treści, streszczeń (w języku polskim i w języku angielskim) oraz wprowadzenia, składa się z 8-miu rozdziałów stanowiących główną część pracy, a także z podsumowania, spisu literatury i z trzech załączników technicznych zawierających informacje nt.: zastosowanych baz głosów, kodeków fonicznych i implementacji opracowanego systemu ASR w środowisku Matlab.

We wprowadzeniu Doktorant uzasadnił istotność podjętych badań, określił ich przedmiot, sformułował, jak już podałem, tezę naukową oraz omówił strukturę tekstu rozprawy.

W tekście można wyróżnić cztery części. Pierwszą z nich jest sam rozdział 1 pt. „Istota i zastosowania systemów rozpoznawania mowy”, w którym Doktorant omówił podstawowe zagadnienia związane z problematyką biometrii głosu ludzkiego wraz z najważniejszymi zastosowaniami.

Część druga zawiera omówienie poszczególnych etapów przetwarzania sygnału mowy i architektury opracowanego systemu ASR. Ta część składa się z rozdziałów 2–6 zatytułowanych kolejno: „Akwizycja i wstępne przetwarzanie sygnału mowy”, „Generacja cech osobniczych”, „Selekcja cech osobniczych”, „Modelowanie głosu mówcy”, „Reguła decyzyjna”.

Główną część rozprawy, określającą jej projektowo-eksperymentalny charakter, w której Pan mgr inż. Kamil Kamiński przedstawił oryginalne wyniki przeprowadzonych badań optymalizacyjnych i eksperymentów, stanowią: rozdział 7 pt. „Badania ewaluacyjne i optymalizacje systemu ASR”, poświęcony projektowaniu i ulepszaniu opracowanego systemu ASR oraz rozdział 8 zatytułowany „Badania weryfikacyjne systemu”, który zawiera szczegółowy opis i wyniki przeprowadzonych eksperymentów.

Ostatnią częścią pracy są trzy wymienione już wcześniej załączniki, uzupełniające rozważania zawarte w poprzednich częściach o informacje techniczne nt. wykorzystanych baz danych, kodeków fonicznych i przygotowania oprogramowania w środowisku MATLAB.

3 Sposób przeprowadzenia analizy źródeł i sformułowania wynikających z niej wniosków

Spis literatury zamieszczony na stronach 107–111 zawiera 85 pozycji. Nie jest to lista wyczerpująca pełen przegląd literatury światowej dotyczącej rozpoznawania mówców, ale podane pozycje literaturowe zostały wystarczająco dobrze dobrane z punktu widzenia prawidłowego zilustrowania zagadnień przeanalizowanych i opracowanych przez Doktoranta. Doktorant, rozwinął, uzupełnił i zoptymalizował metody rozpoznawania mówców opracowane przez Promotora rozprawy Pana Profesora Andrzeja Dobrowolskiego oraz Panią dr inż. Ewelinę Majdę. Doktorant prawidłowo zacytował odpowiednie źródła i przeprowadził ich poprawną analizę.

Spis literatury zawiera 9 publikacji Doktoranta, w tym jedną publikację samodzielną w Zeszytach Naukowych Politechniki Gdańskiej. Cztery współautorskie artykuły zostały opublikowane w Przeglądzie Elektrotechnicznym.

4 Rozwiązanie postawionych zadań, poprawność przyjętych metod i założeń

Do najważniejszych zadań podjętych i wykonanych przez Doktoranta należy zaliczyć:

- opracowanie ulepszonych metod do automatycznego rozpoznawania mówców w trudnych warunkach narzucanych przez łącza telefoniczne
- optymalizację tych metod na podstawie oryginalnych pomysłów Doktoranta związanych z wykorzystaniem algorytmów genetycznych
- ulepszenie procedury tworzenia modeli głosów z uwzględnieniem modelu uniwersalnego głosu za pomocą mieszanin gaussowskich
- opracowanie zoptymalizowanego oprogramowania do automatycznego rozpoznawania mówców rejestrowanych z jakością telefoniczną
- przeprowadzenie wszechstronnej weryfikacji eksperymentalnej opracowanego oprogramowania.

Na podstawie osiągniętych bardzo dobrych wyników rozpoznawania mówców (skuteczność 97,24 % w bazie 1160 głosów) zaproponowane, zoptymalizowane metody rozpoznawania i opracowane na ich podstawie algorytmy oraz oprogramowanie oceniam pozytywnie (jako skuteczne i w pełni poprawne).

5 Ocena uzyskanych wyników, w tym oryginalność rozprawy, samodzielny dorobek Autora i pozycja rozprawy w stosunku do stanu wiedzy

Zamieszczone w rozprawie wyniki prac, w których upatruję istotny wkład jej Autora wynikają bezpośrednio z wymienionych w poprzednim punkcie podjętych przez Niego zadań badawczych. Najistotniejsze z nich to:

- przeprowadzenie wielokryterialnej optymalizacji poszczególnych etapów automatycznego rozpoznawania mówców na podstawie głosów o jakości telefonicznej
- zastosowanie algorytmów genetycznych do ulepszonej selekcji cech głosów
- wykorzystanie mieszanin gaussowskich do tworzenia zoptymalizowanych modeli głosów.

Wyniki te, na podstawie przedstawionych eksperymentów, oceniam wysoko w odniesieniu do dotychczasowego stanu wiedzy na temat tworzenia systemów ASR. Uważam, że są one ważnymi i wartościowymi składnikami samodzielnie uzyskanego, znacznego dorobku naukowego Doktoranta.

6 Poprawność przedstawienia uzyskanych wyników

Język rozprawy jest w ogólności poprawny i zrozumiały, choć zawiera dość liczne błędy gramatyczne i stylistyczne, świadczące o pewnej niestaranności Doktoranta. Niektóre z nich wymieniłem w punkcie pt.: „Wybrane uwagi szczegółowe”.

Rozważania matematyczne na ogół są przeprowadzone z wymaganą starannością, chociaż niektóre sformułowania budzą wątpliwości bądź nie są w pełni poprawne.

We wzorach (2.9) i (2.10) na str. 32 występują niewyjaśnione indeksy.

Kolejność linii we wzorze (3.12) na str. 42 powinna być odwrotna.

Indeksy i oraz j we wzorze (4.1) na str. 46 są przypisane różnym cechom. Czy nie chodzi jednak o ten sam wektor cech (te same cechy) dla różnych obiektów (klas) i oraz j ?

Na str. 47 po wyrażeniu (4.3) jest mowa o „nieujemnej” macierzy. Nie ma takiego pojęcia. Na prawdę chodzi o macierz dodatniookreśloną (lub ogólniej — dodatnioosemiokreśloną) i nie potrzeba tego zaznaczać, bo macierze autokorelacji mają tę własność.

Warunek (4.6) na str. 47 to warunek ortonormalności a nie tylko ortogonalności jak podaje Autor.

Wyrażenie (4.12) na str. 49 definiuje pojęcie informacji wzajemnej zmiennych losowych. Zdanie poniżej: „Dla zmiennych całkowicie zależnych informacja wzajemna wynosi 1.” jest niezrozumiałe. Czy Autorowi chodzi o przypadek zmiennych losowych w pełni związanych deterministycznie, dla których $H(X) = H(Y) = H(X, Y) = I(X, Y)$?

Co oznacza zapis

$$(x_t - \mu_i)' \Sigma_i^{-1} (x_t - \mu_i)$$

we wzorze (5.2) na str. 54? Dlaczego macierz Σ_i występuje w wykładniku?

Reasumując, kompozycja rozprawy jest poprawna. Podobnie jej tekst, poza pewnymi niedociągnięciami, które jednak nie zmniejszają ogólnego pozytywnego wrażenia. Sposób przedstawienia wyników zasługuje więc na ocenę pozytywną.

7 Słabe strony rozprawy

Istotną wadą zmatematyzowanego w wielu miejscach tekstu rozprawy jest brak, bardzo w takim przypadku przydatnego, wykazu zastosowanych oznaczeń oraz skrótów. Brak wykazu zastosowanych oznaczeń oraz wykazu skrótów utrudnił czytanie tekstu i śledzenie myśli Doktoranta, tym bardziej, że nie utrzymał On rygoru stosowania precyzyjnie tych samych oznaczeń w całym tekście pracy.

Mimo, iż same rozważania matematyczne są na ogół przedstawione z wymaganą starannością, to jednak wyjątki omówione powyżej oraz w punkcie pt. „Wybrane uwagi szczegółowe” należy uznać za słabą stronę pracy.

8 Przydatność rozprawy dla nauk technicznych, przemysłu i obronności kraju

Zaprezentowane w rozprawie badania dotyczące rozwoju biometrii głosu, dotyczące opracowania oprogramowania do skutecznego rozpoznawania mówcy w trudnych warunkach, do których należy zakwalifikować telefoniczną jakość rejestracji głosu, mają duże znaczenie w technikach wojskowych a zatem są istotne w obronności kraju. Wynika to stąd, że głos jest dobrą cechą

biometryczną i jest coraz częściej stosowany jako jeden z głównych testów do rozpoznawania osób w systemach z kontrolą dostępu.

9 Wybrane uwagi szczegółowe

Poniżej zebrałem niektóre uwagi szczegółowe, które nasunęły mi się podczas czytania tekstu rozprawy:

- str. 6 (wiersz 11^g): sformułowanie „Rozprawa składa się z czterech logicznych części ...” jest dziwne; czy mogłaby składać się z części nielogicznych?
- str. 6 (wiersz 5_d): sformułowanie „... przewidywane dalszej obszary pracy rozwojowych...” jest niepoprawne i niezrozumiałe
- str. 8 (wiersze 6 i 7^g): sformułowanie „... dopiero próba ... ukazują skalę złożoności przy jej przetwarzaniu” jest niepoprawne
- str. 9 (wiersz 16^g): zamiast „W niniejszym rozprawie” powinno być „W niniejszej rozprawie”; w tekście jest znacznie więcej podobnych błędów
- str. 18: nie wiadomo dlaczego niektóre nazwy np. „(ang. *Text-Prompted Recognition*)” są pisane wielkimi literami a inne np. „(ang. *radio-frequency identification*)” — małymi; w tekście jest znacznie więcej podobnych niekonsekwencji
- str. 20 (wiersz 7_d): zamiast „Kolejny systemu” powinno być „Kolejny system”
- str. 21 (wiersz 11_d): zamiast „Wpłynęła to” powinno być „Wpłynęło to”
- str. 22 (wiersz 7^g): zamiast „,[79]:.” powinno być „,[79]:”
- str. 29: wielkości Δt i τ zdefiniowane wzorami (2.6.b) nie występują we wzorze (2.6.a), a więc fraza „przy czym” jest nie na miejscu
- str. 32: indeksy występujące we wzorach (2.9) i (2.10) nie zostały wyjaśnione
- str. 37: używane są zamiennie dwa określenia: „czas cepstralny” i „pseudoczas” — należy wybrać jedno z nich
- str. 38 (wiersz 8^g): zamiast „(głoska „i”)” powinno być „(głoska „s”)”
- str. 43 (wiersz 18_d): zamiast „system ASR” powinno być „systemu ASR”
- str. 43 (wiersz 13_d): zamiast „zachować tą” powinno być „zachować tę”
- str. 54 (wiersz 9_d): zamiast „,[59][58]” powinno być „,[58], [59]”
- str. 56 i str. 64: we wzorach (5.12) i (6.3) zamiast

$$\sum_{t=1}^T -\log p(x_t|\lambda_k)$$

powinno być

$$\sum_{t=1}^T \left(-\log p(x_t|\lambda_k) \right)$$

- str. 69 (wiersz 19_d): zamiast „cech cepstralne” powinno być „cechy cepstralne”
- str. 71 (wiersz 20_d): zamiast „Kolejne wariant” powinno być „Kolejny wariant”
- str. 71 (wiersz 18_d): zamiast „przedstawiany” powinno być „przedstawiony”
- str. 71 (wiersz 13_d): zamiast „w elementy” powinno być „o elementy”
- str. 78 (wiersz 5^g): zamiast „Wpływ liczebność” powinno być „Wpływ liczebności”
- str. 106 (wiersz 8^g): zamiast „uzyskań skuteczność” powinno być „uzyskać skuteczność”
- str. 106 (wiersz 6_d): zamiast „działa systemu” powinno być „działania systemu”
- str. 106 (wiersz 5_d): zamiast „w pasma” powinno być „w pasmach”
- str. 116 (wiersz 3_d): zamiast „[18]” powinno być „[18].”.

10 Konkluzja

Podsumowując powyższą charakterystykę recenzowanej rozprawy doktorskiej stwierdzam, że Pan mgr inż. Kamil Kamiński zamieścił w niej bardzo wartościowe, oryginalne wyniki prac przeprowadzonych przez Niego pod kierunkiem Promotora rozprawy Pana Profesora dra hab. inż. Andrzeja Dobrowolskiego.

Oceniam, że Doktorant osiągnął zakładane cele badawcze i wykazał prawdziwość stawianej tezy naukowej. Wymienione przeze mnie w poprzednich punktach uwagi krytyczne mają znaczenie drugorzędne i nie zmniejszają wartości osiągniętych i przedstawionych w rozprawie wyników.

Na zakończenie stwierdzam, że przedłożona praca spełnia wymagania stawiane przez stosowne przepisy rozprawom doktorskim. Uważam zatem, że Doktorant powinien być dopuszczony do dalszych etapów procedury doktorskiej i do publicznej obrony.

