



Prof. dr hab. inż. Jacek ŁĘSKI
Instytut Elektroniki, Zakład Elektroniki Biomedycznej
Politechnika Śląska w Gliwicach

Gliwice, dn. 09.06.2014 r.

RECENZJA

rozprawy doktorskiej mgra Ireneusza CODELLO

pt.: „Komputerowe rozpoznawanie niepełności mowy

z zastosowaniem transformaty falkowej i sztucznych sieci neuronowych”

Promotor: dr hab. Wiesława KUNISZYK-JÓŹKOWIAK, prof.nadzw.

Przedstawiona do recenzji rozprawa doktorska jest ulokowana w nurcie cyfrowego przetwarzania sygnałów biomedycznych, a dokładniej dotyczy komputerowo wspomaganego rozpoznawania niepełności na podstawie analizy sygnału mowy. Automatyzacja tego procesu ma niezwykle istotne znaczenie w diagnozowaniu zaburzeń mowy, gdyż prowadzi do obiektywizacji, zmniejszenia kosztu badania poprzez skrócenie czasu opisu tych sygnałów przez specjalistę, a co za tym idzie, szerszej dostępności tego badania dla pacjentów. Uznane jako klasyczne metody analizy sygnałów zazwyczaj słabo nadają się do realizacji zadań związanych z przetwarzaniem sygnałów mowy z punktu widzenia rozpoznawania ich niepełności. Spowodowane jest to obecnością zakłóceń oraz z trudnością w matematycznym modelowaniu różnego rodzaju niepełności, które byłyby niezależne od pacjenta i typu wypowiedzi. Jeżeli zauważyć dodatkowo, że automatyzacja procesu wykrywania niepełności mowy dostarcza specjalistę do odsłuchiwanie tylko skończoną liczbę „podejrzanych” fragmentów sygnału, powodując, że w jednostce czasu może on diagnozować większą liczbę pacjentów, to należy uznać, że Autor podjął się trudnego problemu, którego rozwiązanie jest jak najbardziej uzasadnione zarówno z praktycznego jak i społecznego punktu widzenia. Opracowane nowe metody testowane były na bazie rzeczywistych sygnałów mowy z niepełnościami, które, co istotne, zawierały powtórzenia głosek i sylab oraz przedłużenia dla różnych mówców.

W celu rozwiązania problemu komputerowo wspomaganego wyodrębniania fragmentów sygnałów zawierających niepełności mowy wykorzystano opracowane w ramach rozprawy nowe metody stosujące oryginalne połączenie przekształcenia falkowego oraz sztucznych sieci neuronowych (Kohonena i wielowarstwowego perceptronu). W konstrukcji nowych metod analizy sygnałów mowy zastosowano autorski pakiet oprogramowania „WaveBlaseter”, który umożliwił między innymi analizę widmową sygnałów, zastosowanie predykcji liniowej, sieci Kohonena oraz wizualizację wyników. Wybór metodologii i narzędzi programistycznych

stosowanych przez Autora do osiągnięcia celu postawionego w rozprawie jest zgodny z najnowszymi trendami panującymi w cyfrowym przetwarzaniu informacji.

Tezę, jaką postawiono w rozprawie, którą można streścić, że połączenie zastosowania transformaty falkowej ze skalami barkowymi, odpowiednio zmodyfikowanej sieci Kohonena, metod korelacyjnych i sztucznych sieci neuronowych pozwala na skuteczne rozpoznawanie niepełności w mowie ciągłej jest oryginalna i istotna z naukowego punktu widzenia, a opracowane dla uprawdopodobnienia tej tezy metody i algorytmy będą z pewnością mieć duże znaczenie dla praktyki wykrywania niepełności mowy. Przy formułowaniu tez Autor nadużywa słowa 'optymalny', nie sprawdza przecież wszystkich możliwości detekcji niepełności mowy. Ponadto, słowo 'możliwy' sugeruje błędnie, że Autor nie zamierza podać metody rozwiązania nurtujących Go w rozprawie problemów.

Przedstawiona do recenzji rozprawa obejmuje 109 stron druku. Zawiera wstęp, sześć rozdziałów, wnioski końcowe oraz 93 trafnie dobrane i cytowane w rozprawie pozycje bibliograficzne. Na pozytywne podkreślenie zasługuje uwzględnienie w bibliografii 15 prac naukowych współautorstwa Kandydata dotyczących tematyki rozprawy. Istotne jest także i to, że Autor informacje znane z literatury przedstawił tylko w zakresie koniecznym dla zrozumienia nowych metod prezentowanych w rozprawie. W pracy znajdują się dwa dodatki obejmujące: opis stworzonego w ramach pracy oprogramowania „WaveBlaseter” oraz potwierdzenie uzyskania nagrody na międzynarodowej konferencji „CORES 2013” za pracę współautorstwa Kandydata, dotyczącą tematyki rozprawy.

Analizując zawartość rozprawy można mieć zastrzeżenia co do jej układu. Bardziej odpowiednie byłoby umieszczenie treści rozdziału trzeciego, jako kontynuację wstępu. Wówczas cele i tezy pracy znajdowałyby się po wprowadzeniu w zagadnienia niepełności mowy oraz przeglądzie metod ich wykrywania wraz z krytyczną analizą. Zawarty w rozdziale siódmym, liczący jedną stronę, opis oprogramowania powinien być dołączony do innego rozdziału, jako podrozdział. Dla pozostałych rozdziałów następstwo jest właściwe, informacje znane z literatury zamieszczono głównie w rozdziałach 3 i 4, ograniczając się jedynie do przypomnienia podstawowych pojęć dotyczących detekcji niepełności mowy i metod przetwarzania sygnałów cyfrowych. Kolejne rozdziały, oznaczone, jako 5 i 6, poświęcone są wprowadzonym nowym metodom i algorytmom oraz testowaniu ich przydatności dla rzeczywistych zapisów mowy. Pracę kończą właściwie sformułowane wnioski końcowe.

Do szczególnie wartościowych elementów rozprawy o charakterze oryginalnym należy zaliczyć:

1. Opracowanie nowej metody detekcji tzw. przedłużeń w sygnale mowy. Metoda opiera się na sprawdzaniu czasu trwania zwycięstwa neuronu w sieci Kohonena z jednym neuronem

przypisanym do „braku fonacji”. Wspomniana sieć służyła do redukcji wymiaru wektorów reprezentujących cechy sygnału mowy, a dokładniej będącymi jej częstotliwościowymi reprezentacjami uzyskanymi za pomocą transformaty falkowej z falką Morleta i skalami barkowymi. Jako materiał badawczy zastosowano 373 przedłużenia zapisane u dziesięciu osób. Badania przeprowadzono dla sieci Kohonena o różnych strukturach, czułości łączenia neuronów podobnych i funkcji sąsiedztwa oraz minimalnej długości wykrywanej sekwencji oraz długości jej otoczenia.

2. Zaproponowanie metody detekcji powtórzeń głosek w sygnale mowy. Metoda podobnie jak poprzednia bazuje na przekształceniu falkowym i zastosowaniu sieci Kohonena, ale jest dodatkowo wzbogacona o sztuczną sieć neuronową o strukturze wielowarstwowego perceptronu, wykrywającej powtórzenia głosek. Jako materiał badawczy zastosowano 294 zapisy niepełności u dziesięciu osób. Badania przeprowadzono dla różnych struktur perceptronu i różnych metod jego uczenia, jak i poziomu odcięcia szumu oraz czasu trwania przerwy pomiędzy wyrazami.
3. Opracowanie nowej metody detekcji powtórzeń sylab w sygnale mowy. Metoda podobnie jak poprzednio omówione bazuje na przekształceniu falkowym i skalach barkowych, ale do wykrywania powtórzeń zastosowano współczynnik korelacji wzajemnej wyznaczany dla wyniku przekształcenia falkowego w sąsiadujących oknach dla fragmentów mowy spełniających określone wymagania czasowe (patrz str. 73). Dodatkowo wartości graniczne współczynnika korelacji uzależniono od czasu trwania sylaby. Jako materiał badawczy zastosowano 106 zapisów niepełności u pięciu osób. Badania przeprowadzono dla różnych poziomów odcięcia szumu, różnych liczb uwzględnianych skal barkowych oraz różnych częstotliwości środkowych falki Morleta, struktur perceptronu i różnych metod jego uczenia, jak i poziomu odcięcia szumu.
4. Zaprojektowanie i wykonanie aplikacji „WaveBlaster” implementującej wprowadzone w rozprawie algorytmy detekcji niepełności mowy, a także pozwalającej wykonywać inne analizy sygnału mowy. Na podkreślenie zasługuje ergonomiczność zaprojektowanych interfejsów graficznych oraz możliwość realizacji badań za pomocą skryptów 'xml'.

Uzyskane wyniki stanowią potwierdzenie osiągnięcia celów jaki Autor postawił we wstępie rozprawy oraz są uprawdopodobnieniem tezy rozprawy.

Uwagi ogólne i komentarze jakie nasuwają się przy studiowaniu recenzowanej rozprawy doktorskiej dotyczą następujących zagadnień:

1. Na rys. 6.1:1 przedstawiono tzw. kontur wartości uzyskanych dla przekształcenia falkowego. Rysunek sugeruje obliczanie obwiedni oryginalnych wartości. W tekście rozprawy napisano,

że wyznacza się kontur oscylacji. Formuły matematyczne sugerują zastosowanie wartości średniej. Jak w rzeczywistości „wygładzano” wartości przekształcenia falkowego?

2. Na str. 49 Autor pisze: „...najlepiej dopasować algorytmy rozpoznawania do rodzaju wykrywanych cech...”. Oczywiście z punktu widzenia wniosków wynikających ze statystycznej teorii uczenia (wywodzącej się od „brzytwy Ockhama”) nie chodzi przy układach podejmujących decyzje o dopasowanie do zbioru uczącego, ale o uzyskanie dobrej zdolności uogólniania, rozumianej jako zdolność do podejmowania dobrych decyzji dla „nowych” danych wejściowych. Dla wielowarstwowego perceptronu Autor stosował prawidłową metodologię polegającą na podziale danych na zbiory uczący, weryfikujący i testujący. Czy podobna metodologia powinna być stosowana przy uczeniu innych etapów badanych algorytmów?
3. Dla zaproponowanej w rozprawie metody rozpoznawania powtórzeń sylab zastosowano wartość współczynnika korelacji zależną od czasu trwania sylaby. W pracy na str. 72 napisano: „...postanowiono również wyznaczyć przybliżoną optymalną linię graniczną...”. Pomijając nieprawidłowe użycie słowa 'optymalny', powstaje pytanie jak wyznaczano prostą regresji?

Uwagi szczegółowe:

- 1 — Str.6 – stopniowanie jest dziwne w tym przypadku, 'bardzo praktyczny'.
- 2 — Str. 6 i nast. – pomiędzy odnośnikami powinny być przecinki.
- 3 — Str. 6 – raczej 'ekstrakcja cech'.
- 4 — Str. 11 i nast. – po polsku powinno być 'ręcznie'; błąd literowy w słowie 'zatem'.
- 5 — Str. 13 – dziwny zwrot 'optymalny dystans';
- 6 — Str. 14 – chyba filtry o trójkątnej odpowiedzi impulsowej; błąd literowy w słowie 'współczynników'.
- 7 — Str. 17 – raczej informacje o częstotliwościach sinusoid.
- 8 — Str. 18 i nast. – dziwna metoda numeracji formuł matematycznych; 'i' objaśnione dopiero na str. 20.; częstotliwość zazwyczaj oznaczana jako 'f'; brak indeksu k w (4.1.3).
- 9 — Str. 19 – zbyt skomplikowany system numerowania rysunków.
- 10 — Str. 20 – powinny być odnośniki do literatury przed (4.2.2) i (4.2.3).
- 11 — Str. 22 – błędny indeks sumowania w (4.3.1).
- 12 — Str. 25 – brak kropki.
- 13 — Str. 26 – brak odnośnika do literatury przed (4.3.1.3).
- 14 — Str. 28 – w (4.3.2.1) powinno być $k=0, \dots, K-1$; brak odnośnika do literatury przed (4.3.2.2).
- 15 — Str. 29 – niejednorodny system oddzielania miejsc dziesiętnych.
- 16 — Str. 37 – nieprawidłowo określony wektor wag.
- 17 — Str. 39 – inne oznaczenie wektorów; niezgodna liczba neuronów na rys. 5.1.4 i 5.1.5.
- 18 — Str. 40 – raczej rozkład 'jednostajny'.
- 19 — Str. 52 – raczej 'bardziej odległego' niż 'słabszego'.
- 20 — Str. 71 – rys. 6.5:13 nie jest 'ostry'.

- 21 — Str. 72 – fatalnie użyte słowo 'najoptymalniej' i zwrot 'przybliżona optymalna linia'.
22 — Str. 83 – zwrot 'udało się' może sugerować, że osiągnięcia pracy były przypadkiem.
23 — Str. 84 i nast. – dziwny system numerowania pozycji literaturowych; w niektórych pozycjach nie podano stron, a w innych nie wymieniono wszystkich autorów; w [38] błąd literowy.
24 — Str. 102 – co to jest 'post-procesowanie'?

Reasumując uważam, że mgr Ireneusz Codello wykazał się wiedzą oraz umiejętnościami wymaganymi do uzyskania stopnia doktora nauk technicznych rozwiązując samodzielnie istotne zadanie naukowe z dyscypliny naukowej informatyka. Pozytywną ocenę pracy nie zmniejszają wyszczególnione uprzednio błędy oraz uwagi o charakterze polemicznym. Przedstawiona rozprawa doktorska spełnia wymagania stawiane pracom doktorskim przez odnośne przepisy. Wnioskuje zatem o dopuszczenie mgra Ireneusza Codello do jej publicznej obrony.



Handwritten signature