Prof. dr hab. inż. Bożena Kostek, czł. koresp. PAN
Politechnika Gdańska,
Wydział Elektroniki, Telekomunikacji i Informatyki
Lab. Akustyki Fonicznej

27.12.2024 r.

Opinia nt. rozprawy doktorskiej mgra inż. Mariusza Klecia

pt.: "**Efektywna reprezentacja danych w systemach przetwarzania sygnałów dźwiękowych**", wykonanej pod kierunkiem dr hab. Alicji Wieczorkowskiej oraz promotora pomocniczego dra hab. inż. Krzysztofa Szklannego

a. **Tytuł rozprawy doktorskiej**

Przedmiotem recenzji jest rozprawa doktorska mgra inż. **Mariusza Klecia** pt.: **"Efektywna reprezentacja danych w systemach przetwarzania sygnałów dźwiękowych"**.

b. **Ocena układu rozprawy doktorskiej, w tym informacje o jej poszczególnych częściach składowych**

Układ pracy jest typowy dla prac mających charakter zestawienia cyklu publikacji. Rozprawa doktorska zawiera w początkowym fragmencie typowe elementy, jak streszczenie w j. polskim i angielskim, oświadczenie, rozdział wprowadzający, który dotyczy motywacji prowadzonych prac badawczych. W dalszej kolejności pojawia się przegląd literatury w tematyce rozprawy doktorskiej (rozdział 2), obejmujący odniesienie do splotowych sieci neuronowych stosowanych w ekstrakcji cech sygnałów dźwiękowych, sieci rekurencyjnych wykorzystywanych w przetwarzaniu czasowym, zastosowań transformerów w różnych obszarach, m.in. w zadaniach klasyfikacji oraz separacji mówców, wydobywania cech sygnałów fonicznych w sposób nienadzorowany, np. architektura autoenkoderów, wykorzystania surowych danych w uczeniu głębokim, inżynierii cech w systemach rekomendacji muzyki, analizy nagrań fizjologicznych oraz możliwości tworzenia i wykorzystania własnych zbiorów danych. Pewien niedosyt sprawia w tym przeglądzie odniesienie do nielicznych źródeł z lat 2023 i 2024.

Rozdział trzeci odnosi się do cyklu prac w postaci krótkiego ich przeglądu, zawierającego cel danej pracy, najważniejsze (bądź podane w formie syntetycznej) wyniki oraz wnioski bądź podsumowanie. Dotyczy to następujących prac:

[1] Kleć, M., Szklanny, K. & Wieczorkowska, A. (2024). Developing a Corpus for Polish Speech Enhancement by Reducing Noise, Reverberation, and Disruptions. In B. Marcinkowski, A. Przybylek, A. Jarzębowicz, N. Iivari, E. Insfran, M. Lang, H. Linger, & C. Schneider (Eds.), Harnessing Opportunities: Reshaping ISD in the post-COVID-19 and Generative AI Era (ISD2024 Proceedings). Gdańsk, Poland: University of Gdańsk. ISBN: 978-83 972632-0-8. https://doi.org/10.62036/ISD.2024.37.

[2] Kleć, M., & Korzinek, D. (2015). Pre-trained deep neural network using sparse autoencoders and scattering wavelet transform for musical genre recognition. Computer Science, 16 (2), 133–144,
[3] Kleć, M. (2018). Early Detection of Heart Symptoms with Convolutional Neural Network and Scattering Wavelet Transformation. In: Ceci, M., Japkowicz, N., Liu, J., Papadopoulos, G., Raś, Z. (eds) Foundations of Intelligent Systems. ISMIS 2018. Lecture Notes in Computer Science, vol 11177. Springer, Cham. https://doi.org/10.1007/978-3-030-01851-1_3 http://dx.doi.org/10.7494/csci.2015.16.2.133
[4] Kleć, M., Wieczorkowska, A., Szklanny, K., & Strus, W.: Beyond the Big Five personality traits for music recommendation systems. J. Audio Speech Music Proc. 2023, 4 (2023). https://doi.org/10.1186/s13636-022-00269-0
[5] Kleć, M., Wieczorkowska, A. (2021). Music Recommendation Systems: A Survey. In: Ras, Z.W., Wieczorkowska, A., Tsumoto, S. (eds) Recommender Systems for Medicine and Music. Studies in Computational Intelligence, vol 946. Springer, Cham. https://doi.org/10.1007/978-3-030-66450-3_7

Rozdział 4 stanowi podsumowanie, które jest istotne w kontekście udowodnienia postawionych w pracy hipotez badawczych i odniesienia ich do poszczególnych prac z przedstawionego cyklu publikacji. Rozdział 5 jest zestawieniem źródeł literatury (46 pozycji), wykorzystanych w przeglądzie literatury w rozdziale 2. W rozdziale 6. dołączono prace stanowiące cykl publikacji przedstawionych do oceny.

Typowo w rozprawach doktorskich, szczególnie takich, które mają charakter monograficzny, można również znaleźć spis najważniejszych oznaczeń, symboli i skrótów, jak również wykazy rysunków i tabel. Jest zrozumiałe, że te ostatnie nie miałaby większego sensu, ale chyba jednak warto by było zawrzeć w pracy najważniejsze oznaczenia.

W ogólności, recenzowana rozprawa doktorska ma charakter eksperymentalny, dlatego rozdział 2. ma charakter uzasadnienia wyboru użytych narzędzi, a nie typowego przeglądu literatury. Z kolei, w Podsumowaniu warto było wyraźniej odnieść się do **pozycji uzyskanych wyników w stosunku do aktualnego stanu wiedzy oraz planów na przyszłość.**

c. **Ocena celu pracy kandydata**

Główne cele rozprawy – chociaż nie zostały podane w sposób jednoznaczny – dotyczą metod poprawy skuteczności systemów przetwarzających sygnał dźwiękowy poprzez włączenie zdarzeń dźwiękowych do korpusów mowy,

wygenerowanych razem z pogłosem i dźwiękami tła, wykorzystanie różnej długości ramek czasowych rozproszonej transformaty falkowej oraz wykorzystanie modelu stanowiącego rozszerzenie standardowego modelu tzw. Wielkiej Piątki do reprezentacji osobowości użytkownika w przetwarzaniu danych muzycznych. Cele te pojawiają się w formie rozszerzonej jako hipotezy sformułowane przez doktoranta i stanowią jednocześnie odniesienie do cyklu publikacji, które zostały zawarte w rozprawie doktorskiej.

Hipotezy badawcze:

1. Włączenie zdarzeń dźwiękowych do korpusów mowy, wygenerowanych razem z pogłosem i dźwiękami tła, poprawia odporność modeli przetwarzających mowę na występowanie tych zakłóceń.
2. Wykorzystanie różnej długości ramek czasowych rozproszonej transformaty falkowej oraz splotowych sieci neuronowych przyczynia się do poprawy rozpoznawania wczesnych objawów chorób serca.
3. Wykorzystanie modelu stanowiącego rozszerzenie standardowego modelu tzw. Wielkiej Piątki do reprezentacji osobowości użytkownika przyczynia się do redukcji błędu popełnionego przez systemy rekomendacji muzycznej, w porównaniu modelem standardowym.

Zarówno podana motywacja prac badawczych, jak i cele należy ocenić jako wartościowe i aktualne. Jednak podanie hipotez w formie mieszanej, tj. jakościowo-ilościowej wzbogaciłoby uzyskane wnioski z prowadzonych eksperymentów.

d. **Ocena zastosowanych metod badawczych**

Aby zrealizować założony cel, doktorant przeprowadził prace badawcze w schemacie, jak poniżej:

– W proponowanym rozwiązaniu (praca [1]) w procesie trenowania wykorzystano kilka modeli, tj. (1) separujących pojedynczego mówcę od dźwięku tła, (2) separujących dwóch mówców od siebie, (3) separujących dwóch mówców od tła szumów i zakłóceń, ale działających na surowych danych dźwiękowych. W tym celu zastosowana została architektura sieci Conv-TasNet, działająca w dziedzinie czasu, zaproponowana przez Luo i in. W 2019 r. Modele separacji porównano z wynikami uzyskanymi przez transformer SepFormer wytrenowanym na niezaszumionych danych. Uzyskane wyniki świadczą o wyższości podejścia proponowanego przez doktoranta, zwłaszcza, że odnoszą się do eksperymentów z różnymi bazami danych. W ramach eksperymentów autorski model został porównany z modelem CTNoar, który został pierwotnie wytrenowany na miksie dwóch i trzech sygnałów mowy w celu wyodrębnienia jednego sygnału i umieszczenia pozostałych w oddzielnym kanale. W założeniu model CTNoar potrafi zidentyfikować jednego mówcę i oddzielić szum w osobnym kanale

podczas testów na zaszumionym zestawie danych. Jednak to założenie nie sprawdziło się w rzeczywistej ocenie. Przedstawione autorskie wyniki pozwoliły udowodnić hipotezę nr 1, tj. włączenie zdarzeń dźwiękowych do korpusów mowy, wygenerowanych razem z pogłosem i dźwiękami tła, poprawia odporność modeli przetwarzających mowę na występowanie tych zakłóceń;

– Doktorant w drugiej z prezentowanych prac [2] zaproponował metodę treningu autoenkoderów z wykorzystaniem rozproszonej transformacji falkowej (SWT, Scattering Wavelet Transform) do rozpoznawania gatunku muzycznego i ocenił skuteczność tej metody. W pracy zastosowano dwa podejścia do treningu autoenkoderów (Sparse Autoencoder, SAE), pierwsze z wykorzystaniem parametryzacji SWT, w drugim – każdy kolejny autoenkoder był trenowany na podstawie cech wydobytych z poprzednich autoenkoderów. Wagi wytrenowanych autoenkoderów posłużyły do inicjalizacji wag sieci neuronowej (Deep Neural Network, DNN). Należy zauważyć, że zaproponowana metodologia ma charakter oryginalny w kontekście wykorzystania podejścia SWT, a uzyskane wyniki dokładności (błąd rozpoznania) rozpoznawania gatunku muzycznego na poziomie 90% (praca została opublikowana w 2015 r.) zgodne ze stanem wiedzy. Jednocześnie – należy podkreślić – były też obiecujące, gdyż baza nagrań w eksperymentach była kilkukrotnie większa niż typowo wykorzystana przez innych badaczy. Obecny system analizuje dźwięk w oparciu o pojedyncze ramki przedstawiające widmową zawartość sygnału w danym momencie, niejako „ignorując" aspekty czasowe, czyli zmiany częstotliwości w czasie. Doktorant podaje wniosek, że analiza wielu próbek jednocześnie może umożliwić rozpoznawanie wzorców czasowych , również w zastosowaniu splotowych sieci neuronowych, które dobrze sprawdzają się w analizie reprezentacji sygnałów dwuwymiarowych. Dlatego też doktorant w kolejnej pracy [3] rozwinął zaproponowane podejście również w kontekście modelowania sekwencji kilku ramek jednocześnie za pomocą sieci splotowej CNN;

– W rozwinięciu [3] wcześniej podanego rozwiązania, wykorzystano różnej długości ramki SWT oraz szerokości filtra splotowego w treningu splotowych sieci neuronowych, co przyczyniło się do uzyskania poprawy skuteczności wczesnego rozpoznawania chorób serca. W podejściu zastosowano różne typy filtrów (Gabora i Morleta) w kolejnych warstwach sieci, ale – jak wspomniano wcześniej – też z różną długością. Warto zauważyć, że w otrzymane wyniki zostały uśrednione po dziesięciokrotnym przeprowadzeniu eksperymentu, następnie porównane ze stanem wiedzy. Istotą tej metody jest nie tylko wysoka dokładność w przypadku niektórych wad serca, identyfikowanych za pomocą sygnałów dźwiękowych, ale również wysoka precyzja w rozpoznawaniu artefaktów w nagraniu, chociaż eksperymenty dotyczą baz sygnałów z niewielką ilością danych. Przedstawione wyniki pozwoliły udowodnić hipotezę nr 2;

– W kolejnej pracy [4] doktorant zbadał wpływ rozszerzenia danych na wejściu systemu rekomendującego o reprezentację cech osobowości słuchacza. W tym

celu zastosował model tzw. Wielkiej Piątki (Big Five) odnoszący się do pięciu głównych cech osobowości, tj. neurotyczności, ekstrawersji, otwartości na doświadczenia, ugodowości i sumienności. Model ten został rozszerzony o dodatkowe piętnaście aspektów osobowościowych (cech drugorzędnych), po trzy dodatkowe dla każdej z pięciu głównych cech. Przygotowana aplikacja pozwoliła zebrać dane od ok. 300 osób, które wzięły udział w testach odsłuchowych, oceniających dany utwór i jednocześnie określająca uczestnika testów w oparciu o rozszerzony formularz reprezentacji cech osobowości. Powstała baza jest wartością dodaną prowadzonych badań. Eksperymenty polegały na przeprowadzeniu predykcji oceny utworów przez użytkownika. Funkcja predykcyjna uwzględniała pomiar podobieństwa zarówno użytkowników, jak i utworów. Eksperymenty opierały się na poszukiwaniu optymalnego zbioru cech osobowości, które w wyniku pomiaru podobieństwa między użytkownikami skutkują najmniejszym błędem predykcji. Wyniki eksperymentów pokazały, iż zastosowanie cech osobowościowych niższego rzędu zwraca mniejszy błąd predykcji niż uzyskany w eksperymentach z wykorzystaniem cech Wielkiej Piątki bez cech niższego rzędu. Pozwoliło to na udowodnienie hipotezy badawczej nr 3;

– Ostatnia pracy [5] z cyklu publikacji przedstawionego w ramach rozprawy doktorskiej stanowi przegląd systemów rekomendacji muzyki w kontekście ich personalizacji. Na podstawie dokonanego przeglądu doktorant odnosi się do możliwych kierunków badań w obszarze przetwarzania danych w platformach streamingowych.

Przyjęta metodologia jest poprawna, a przedstawienie wyników w poszczególnych pracach – przekonujące. Niewątpliwie niektóre z prac wnoszą duży walor oryginalności w obszarze poddanym analizie, tj. rozpoznawanie sygnałów biomedycznych czy w ogólności systemy rekomendacji muzycznej.

Przedstawione w rozprawie wyniki można uznać za obiecujące w kontekście skutecznej (wysokie wartości miar) rozpoznawania gatunków muzycznych przy zastosowaniu rozproszonej transformacji falkowej. Autor rozprawy w pracach z cyklu zawartego w rozprawie (oraz częściowo w Podsumowaniu) wskazuje też na ograniczenia, co jest niewątpliwie ważnym wnioskiem z przeprowadzonych badań. W ogólności, stwierdzam, że wynik prowadzonych przez doktoranta badań stanowi ważny wkład w nauki inżynieryjno-techniczne oraz dyscyplinę naukową, jaką jest informatyka techniczna i telekomunikacja.

<u>Uwagi do dyskusji:</u>

1. Warto było w przeglądzie literatury przywołać więcej źródeł z 2023 i 2024 r. − łatwiej by było wtedy odnieść się do aktualnego stanu wiedzy w niektórych obszarach. Czy doktorant może podać kilka nowszych źródeł do czterech pierwszych prac z cyklu?

2. W Podsumowaniu doktorant mógłby przedstawić w sposób bardziej wyrazisty mocne punkty rozprawy (poszczególnych prac cyklu), które świadczą też o oryginalności przeprowadzonych badań. Zabrakło odniesienia do tzw. „planów na przyszłość", które zostały sformułowane w pracach z cyklu publikacji. Dlatego poniżej jest prośba do doktoranta o odniesienie się do wybranych zagadnień, przedstawionych w pracach cyklu jako „rozwinięcie eksperymentów".
3. Czy doktorant mógłby odnieść się do końcowych uwag zawartych w publikacji [1], tj.:
   - „Dalsze badania mogą być prowadzone w kierunku analizy wpływu danej warstwy zawierającej szum na skuteczność zadań separacji, poprawy i rozpoznawania mowy."
4. Czy doktorant mógłby odnieść się do końcowych uwag zawartych w publikacji [2], tj.:
   - „Analiza wielu próbek jednocześnie może umożliwić rozpoznawanie wzorców czasowych. Warto również rozważyć zastosowanie Splotowych Sieci Neuronowych (CNN), które dobrze sprawdzają się w analizie sygnałów dwuwymiarowych".

   czy takie próby zostały podjęte?
5. Czy doktorant mógłby odnieść się do końcowych uwag zawartych w publikacji [3], tj.:
   - „Autor planuje zastosować metodę CSWT do analizy danych, a także przeprowadzić dodatkowe eksperymenty w celu pełnej weryfikacji wyników oraz uwzględnienia porównań statystycznych z uwzględnieniem dużej bazy danych PhysioNet/Computing in (CinC)[5], udostępnionej w ramach konkursu Cardiology Challenge 2016".
   - „Dane z urządzeń zewnętrznych, takich jak urządzenia mobilne, wearables czy elektroniczne stetoskopy, mogą być silnie skorelowane z danymi z tradycyjnego sprzętu medycznego (np. systemów EKG). Autor uważa, że podejście oparte na urządzeniach zewnętrznych zasługuje na dalsze badania, ponieważ może wspierać medycynę prewencyjną, umożliwiając monitorowanie pracy serca bez zakłócania codziennych czynności".

   - czy takie próby zostały podjęte?
6. W kolejnej publikacji [4] jest mowa o ograniczeniach, które są typowe dla eksperymentów, w których dane są zbierane w ramach badań (uwaga: to nie jest uwaga krytyczna!). Czy te prace są rozwijane?
7. Rozdział 2 może lepiej by było zatytułować: Uzasadnienie wyboru stosowanych narzędzi (lub podobnie), gdyż ten rozdział ma częściowo

charakter założeń badawczych − jest to oczywiście uwaga „porządkująca”, a nie zarzut.

e. **Ocena informacje o ewentualnych nieprawidłowościach i niesłusznych sformułowaniach w ocenianej - brak**

f. **Ocena, czy rozprawa doktorska stanowi oryginalne rozwiązanie problemu naukowego**

Postawione hipotezy badawcze mają wyraźny walor oryginalności. Przedstawione autorskie wyniki pozwoliły udowodnić te hipotezy. Najbardziej wartościowe wydaje się włączenie zdarzeń dźwiękowych do korpusów mowy, wygenerowanych razem z pogłosem i dźwiękami tła, co poprawiło w przedstawionych eksperymentach odporność modeli przetwarzających mowę na występowanie tych zakłóceń. Walor oryginalności można również odnaleźć w obszarach poddanych analizie, tj. rozpoznawaniu sygnałów biomedycznych czy w ogólności systemy rekomendacji muzycznej.

g. **Konkluzje co do końcowej oceny rozprawy doktorskiej**

W konkluzji stwierdzam, że ocena przedłożonej mi do recenzji rozprawy doktorskiej p. **Mariusza Klecia jest pozytywna** i **spełnia wymagania** stawiane rozprawom doktorskim w aktualnie obowiązującej Ustawie o stopniach i tytule naukowym. W związku z tym wnoszę **o dopuszczenie rozprawy doktorskiej p. mgra inż. Mariusza Klecia do publicznej obrony.**

Uzasadnienie wyróżnienia

Jednocześnie wnoszę o wyróżnienie rozprawy doktorskiej. W uzasadnieniu chciałabym się odnieść do pracy opublikowanej na renomowanej konferencji ISD (lista MEiN, 140 punktów), w której doktorant jest wiodącym autorem. Referat ten został wyróżniony przez Organizatorów konferencji.

Dlatego zgodnie z przyjętymi zasadami wyróżniania rozpraw doktorskich w Polsko-Japońskiej Akademii Technik Komputerowych w dyscyplinie naukowej Informatyka techniczna i telekomunikacja – **wnoszę o wyróżnienie rozprawy doktorskiej.**

prof. dr hab. inż. Bożena Kostek