POLITECHNIKA OPOLSKA

Opole, 1.02.2024 r.

Recenzja pracy doktorskiej mgr. Mikołaja Pudo pt.: „Methods of optimizing models and algorithms for automatic speech recognition in mobile applications (Metody optymalizacji modeli i algorytmów automatycznego rozpoznawania mowy pod kątem działania na urządzeniach mobilnych)" przygotowanej pod kierunkiem promotora: dr hab. inż. Artura Janickiego, prof. uczelni, w dyscyplinie: Informatyka Techniczna i Telekomunikacja

Recenzja została sporządzona w związku z powołaniem przez Radę Naukową Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Warszawskiej w dniu 19.09.2023 roku do pełnienia funkcji recenzenta w postępowaniu o nadanie stopnia naukowego doktora nauk technicznych panu mgr. Mikołajowi Pudo.

Niniejsza recenzja ma za zadanie zgodnie z Art. 13 ust. 1 ustawy z dnia 14 marca 2003 r. o stopniach i tytule naukowym oraz o stopniach i tytule naukowym w zakresie sztuki (t.j. Dz.U. 2017 poz. 1789) ocenić, czy rozprawa doktorska stanowi oryginalne rozwiązanie problemu naukowego oraz wkład w dyscyplinę, zgodnie z art. 175 ust. 1 Przepisów wprowadzających ustawę – Prawo o szkolnictwie wyższym i nauce z 3.7.2018 r. (Dz.U. 2018 r. poz. 1669).

W ramach przeprowadzonej recenzji zostaną ocenione następujące punkty:

1. Tematyka pracy doktorskiej i jej wkład w dyscyplinę.

Ad. 1. Temat pracy doktorskiej brzmi: „Methods of optimizing models and algorithms for automatic speech recognition in mobile applications".

Niniejszy doktorat ma charakter wdrożeniowy i powstał przy współpracy z firmą Samsung R&D Institute Poland, przy wsparciu Dr Bożeny Łukasiak. Tematyka pracy dotyczy natomiast interfejsów głosowych, które mają bardzo szerokie zastosowania w interfejsach człowiek-maszyna, ze szczególnym uwzględnieniem inteligentnych asystentów.

W ramach realizacji doktoratu, autor skupił się na automatycznym rozpoznawaniu mowy, a także udostępnił publicznie utworzony wcześniej zestaw testowy o nazwie Multilingual Open Custom Keyword Spotting Testset (MOCKS), który powstał z otwartych zbiorów danych, takich jak LibriSpeech i Mozilla Common Voice. Zestaw ten zawiera prawie 50 000 słów kluczowych w pięciu językach: angielskim, francuskim, niemieckim, włoskim i hiszpańskim.

W mojej ocenie – wybrany temat rozprawy, jest aktualny i stanowi bardzo istotny wkład w dyscyplinę naukową **Informatyka Techniczna i Telekomunikacja**.

2. Zagadnienia naukowe rozprawy – cel i teza pracy.

Ad. 2. W pracy zostały postawione aż trzy tezy:

- The performance and accuracy of a keyword spotting model can be significantly improved by using a unigram language model and audio recordings generated by a text-to-speech system (*Wydajność i dokładność modelu rozpoznawania słów kluczowych można znacznie poprawić, korzystając z modelu języka unigramowego i nagrań dźwiękowych generowanych przez system zamiany tekstu na mowę*).
- The accuracy of an end-of-speech detection model can be effectively improved by training the model with the proposed loss function (*Dokładność modelu wykrywania końca mowy można skutecznie poprawić poprzez uczenie modelu z proponowaną funkcją straty*).
- Semi-supervised learning methods can be effectively used to adapt acousticmodels even with small datasets (*Metody uczenia się częściowo nadzorowanego można skutecznie wykorzystać do adaptacji modeli akustycznych nawet w przypadku małych zbiorów danych*).

Na podstawie wyników przedstawionych w rozprawie, uważam, że tezy zostały potwierdzone. Oceniając otrzymane przez doktoranta wyniki, można stwierdzić, że niniejsza rozprawa doktorska spełnia wszystkie standardy obowiązujące w przypadku prac doktorskich oraz, iż w znacznym stopniu przyczynia się do rozwoju dyscypliny naukowej jaką jest **Informatyka Techniczna i Telekomunikacja**.

3. **Struktura pracy**

Ad. 3. W ramach realizacji niniejszej pracy doktorskiej podjęto się tematyki związanej z rozpoznawaniem mowy. Autor doktoratu skoncentrował się na automatycznym rozpoznawaniu mowy w urządzeniach mobilnych. Praca składa się z **90** stron (wliczając stronę tytułową, podziękowania, streszczenie w języku angielskim i polskim, spisy tabel, rysunków, symboli i skrótów, a także opis osiągnięć autora, załączniki - **A** i **B** oraz wykaz cytowanej literatury). Rozprawa została napisana w całości w języku angielskim i składa się ona z **8** rozdziałów (wliczając nieponumerowaną literaturę), **30** rysunków, **11** tabel, **4** wzory, **3** listingi kodu oraz **118** pozycji literaturowych.

Pierwszy rozdział pracy stanowi **Introduction**, czyli wprowadzenie, w którym zawarto teoretyczny wstęp do pracy. Rozdział **drugi** to przegląd literaturowy. Rozdział **trzeci**, zatytułowany **Contribution of this Thesis** stanowi opis wkładu pracy doktorskiej w dyscyplinę, przedstawiono zdefiniowane tezy badawcze. W rozdziale **czwartym**, zatytułowanym **Custom Keyword Spotting** (***Niestandardowe Wykrywanie Słów Kluczowych***) opisano próby rozwiązania problemu wyszukiwania słów kluczowych (KWS) oraz szczegółowo opisano nowatorskie rozwiązanie w trybie innym niż przesyłanie

strumieniowe, oparte na architekturze modelu akustycznego (AM). Rozdział ten posiada także dyskusję oraz podsumowanie. W rozdziale **piątym End-of-speech Detection** (***Wykrywanie Końca Mowy***) opisano min. metody wykrywania granic mowy (SBD), opisano wykorzystany model wraz z funkcją straty, a także przestawiono zestaw wykorzystany w badaniach, wyniki eksperymentalne oraz wnioski. Rozdział **szósty** (Semi-supervised Learning for Speech Recognition) zawiera opis częściowo nadzorowanego uczenie się rozpoznawania mowy, wraz z proponowanymi ulepszeniami (trzema), opis stanowiska badawczego (architektura modelu, dane, metryki), a także wyniki eksperymentalne oraz wyniki dla dodatkowych danych oznaczonych etykietami czy dla treningu wstępnego z danymi oznaczonymi etykietami. Dla tego rozdziału zawarto także dyskusję oraz wnioski. W rozdziale **siódmym** zawarto podsumowanie pracy wraz z osiągniętymi wynikami. Po nim autor zamieścił załączniki **A** oraz **B**, gdzie w załączniku **A** zawarł swoje osiągnięcia, a w załączniku **B** - niestandardowe wykrywanie słów kluczowych oraz dodatkowe rysunki/wykresy. Rozdział **ósmy** (nieponumerowany) stanowi literatura, na którą składa się **118** pozycji.

4. Uwagi redakcyjne, krytyczne oraz pytania do pracy.

Ad. 4.

- Praca ma bardzo nietypowy układ, gdzie poszczególne rozdziały posiadają swoje osobne wnioski, co nieco utrudnia jej odbiór.
- W pracy brakuje szczegółowej dyskusji, w której można byłoby omówić napotkane trudności.
- Brakuje podrozdziału dotyczącego dalszych planów badawczych, rozwoju.
- Praca napisana jest zbyt potocznym jak na pracę doktorską językiem.
- W pracy występują drobne błędy językowe, czy została ona sprawdzona przez tłumacza/"native speakera"?
- Cytowanie Wikipedii ([1]) jest na tym etapie kształcenia niedopuszczalne.
- Często występuje w pracy forma "we" (my), czy praca nie była realizowana samodzielnie?
- Bardzo krótka dyskusja.
- Jako, że praca ta ma charakter wdrożeniowy - jak wygląda wdrożenie prototypu proponowanego rozwiązania?
- Jakie jest największe osiągnięcie tej pracy?
- Który język był najtrudniejszy spośród wymienionych w pracy?

5. Podsumowanie.

Ad. 5. Uważam, że Doktorant z powodzeniem udowodnił postawione tezy, w której bardzo dobrze opisał próby rozwiązania problemu wyszukiwania słów kluczowych (KWS) oraz szczegółowo przedstawił nowatorskie rozwiązanie w trybie innym niż przesyłanie strumieniowe, oparte na architekturze modelu akustycznego (AM).

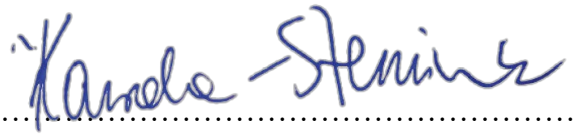
Autor zdefiniował także aż trzy tezy badawcze, które udowodnił w pracy.

Na szczególną uwagę zasługuję również ogólna działalność naukowa doktoranta, który jest współautorem **8** publikacji naukowych (wg Google Scholar) oraz **jednego** patentu zarejestrowanego w Stanach Zjednoczonych. Był także laureatem nagrody im. Profesora Zdzisława Pawlaka w kategorii współpracy z przemysłem za artykuł: *M. Pudo, M.Wosik, and A. Janicki, "Open vocabulary keyword spotting with small-footprint ASR-based architecture and language models", in 18th Conference on Computer Science and Intelligence Systems (FedCSIS 2023), 2023, Warsaw, Poland.*

Ponadto doktorant brał udział w **7** konferencjach naukowych (**5** międzynarodowych oraz **2** krajowych), jest także współautorem **2** rozdziałów monografii. Jak już zostało to wcześniej wspomniane jest także współautorem **1** patentu (amerykańskiego) oraz laureatem **1** nagrody.

Moja ocena pracy **mgra inż. Mikołaja Pudo** jest **pozytywna**. Moim zdaniem niniejsza praca prezentuje cenne wyniki badań i jest znaczącym osiągnięciem naukowym w dyscyplinie naukowej **Informatyka Techniczna i Telekomunikacja.** Spełnia ona również w mojej ocenie wszystkie wymogi zawarte w aktualnie obowiązującej Ustawie z dnia 20 lipca 2018 roku "Prawo o szkolnictwie wyższym i nauce" w sprawie warunków i trybu przeprowadzania przewodów doktorskich i może być przedmiotem publicznej obrony.

Wnioskuję do Rady Naukowej Dyscypliny **Informatyka Techniczna i Telekomunikacja** o dopuszczenie pana **mgra inż. Mikołaja Pudo** do dalszych etapów przewodu doktorskiego, a także o **wyróżnienie rozprawy**. Stworzenie całego korpusu MOCKS oraz posiadanie amerykańskiego patentu stanowi w mojej ocenie przesłanki do wyróżnienia całej rozprawy.



.....

Dr hab. Inż. Aleksandra Kawala-Sterniuk, prof. uczelni
Wydział Elektrotechniki, Automatyki i Informatyki
Politechnika Opolska
ul. Prószkowska 76
45-758 Opole
a.kawala-sterniuk@po.edu.pl