Łódź, 27.01.2024

Prof. dr hab. inż. **Krzysztof Ślot**
Instytut Informatyki Stosowanej
Politechnika Łódzka

Recenzja rozprawy doktorskiej Pana magistra inżyniera

Mikołaj Pudo

pt.

Methods of optimizing models and algorithms for automatic speech recognition in mobile applications

1. Tematyka, kompozycja i kontekst rozprawy

Przedstawiona do recenzji rozprawa doktorska dotyczy, jak dość precyzyjnie informuje jej tytuł, tematyki doskonalenia algorytmów automatycznego rozpoznawania mowy, ukierunkowanego na ich implementację w urządzeniach mobilnych. Zasadniczą trudnością autonomicznej (dokonywanej bez odwoływania się do zewnętrznych serwerów) realizacji zadania rozpoznawania mowy są dwa specyficzne uwarunkowania rozważanego kontekstu: ograniczone zasoby sprzętowe, wykluczające możliwość używania istniejących, skutecznych, ale zbyt dużych modeli obliczeniowych, oraz konieczność uwzględnienia nieuchronnego występowania czynników istotnie zakłócających proces akwizycji głosu użytkownika. Obydwa przedstawione aspekty podjętej przez Doktoranta tematyki: złożoność podejmowanego problemu i praktyczna przydatność prac (przedstawiona praca ma bardzo utylitarny wymiar), odpowiadają wymaganiom stawianym przed rozprawą doktorską, a merytoryczny zakres prac lokuje się w obszarze informatyki technicznej, w której realizowany jest przewód doktorski.

Przyjęta przez Doktoranta struktura prezentacji treści nie budzi zastrzeżeń – po przedstawieniu zakresu tematycznego prac, omówieniu stanu wiedzy w rozważanym obszarze i określeniu celów oraz przyjętej metodyki, szczegółowo omawiane są trzy różne zaproponowane przez Niego pomysły, zmierzające do osiągnięcia postawionego celu. Architektury obliczeniowe wykorzystywane przez Doktoranta to zaawansowane i złożone modele uczenia głębokiego, pretrenowane z użyciem obszernych baz nagrań mowy i dostrajane w proponowanych przez Niego scenariuszach z użyciem właściwie dobranego, dodatkowego materiału eksperymentalnego. Na zdecydowanie uznanie zasługuje obszerność i metodyczna poprawność ewaluacji opracowanych algorytmów: są one rzetelnie i w dobrze przemyślany sposób konfrontowane z rozwiązaniami referencyjnymi.

Doktorant posiada świetną orientację w aktualnym stanie wiedzy w obszarze rozpoznawania mowy, w szczególności, realizowanego w warunkach posiadania jedynie ograniczonych zasobów obliczeniowych, wynikające w istotnej mierze z jego aktywności zawodowej (dział badawczo-rozwojowy firmy Samsung), która jest zbieżna z podjętymi przez Niego pracami naukowymi. Niewątpliwym atutem łączenia wątków działalności zawodowej i naukowej jest, wynikająca z dostępu do aktualnych rozwiązań technologicznych, możliwość opracowywania i testowania rozwiązań realnych do implementacji w istniejącym sprzęcie. Minusem wspomnianej kombinacji jest ograniczona swoboda publikowania swoich osiągnięć, co przekłada się na ograniczony dorobek naukowy Doktoranta.

2. Cele i tezy rozprawy

Poszukując rozwiązań umożliwiających adaptację istniejących metod rozpoznawania mowy, zdominowanych przez algorytmy i architektury uczenia głębokiego, do ograniczeń narzucanych przez możliwości obliczeniowe urządzeń mobilnych, Doktorant skupia swoją uwagę na trzech różnych wątkach tematycznych, formułując trzy hipotezy badawcze, których wykazanie stanowi przedmiot jego prac:

1. 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
2. The accuracy of an end-of-speech detection model can be effectively improved by training the model with the proposed loss function.
3. Semi-supervised learning methods can be effectively used to adapt Acoustic Models even with small datasets.

Przedstawienie rozwiązań zaproponowanych przez Doktoranta jest poprzedzone prezentacją aktualnego stanu prac dla każdego z rozważanych wątków, stanowiącą kompetentne kompendium wiedzy (chociaż w odniesieniu do pierwszego z wątków, dla zwiększenia czytelności prezentacji należało według mnie wydzielić dodatkowo podrozdział dotyczący omówienia zbiorów danych stosowanych w uczeniu modeli). Fundamentem algorytmów zaproponowanych przez Doktoranta w każdym z przedstawionych wątków tematycznych jest wykorzystanie głębokich modeli neuronowych, stanowiących na chwilę obecną bezkonkurencyjne narzędzia analizy danych rzeczywistych. Istotą prac Doktoranta stało się więc poszukiwanie sposobów zapewnienia akceptowalnego kompromisu między wymuszonym przez możliwości implementacyjne stopniem redukcji złożoności architektur głębokich a poprawnością realizacji analizy.

3. Merytoryczna ocena pracy

Problematyka redukcji złożoności głębokich modeli neuronowych jest, z uwagi na znaczenie praktyczne, niezmiernie rozległym obszarem prac badawczych, zwieńczonym obszerną paletą metod o różnym stopniu ogólności. Skupienie przez Doktoranta uwagi na szczególnym obszarze aplikacyjnym, pozwoliło na wykorzystanie jego specyfiki i sformułowanie zestawu nowych, oryginalnych propozycji algorytmów obliczeniowych, dedykowanych realizacji wybranych analiz sygnału mowy. Przedstawione przez Doktoranta pomysły stanowią poprawne metodycznie, oryginalne sposoby rozwiązania problemów naukowych, pozwalające na uzyskanie użytecznych z

punktu widzenia praktycznego efektów, lepszych, jak wykazał On w drodze przeprowadzanych przez siebie eksperymentów, od możliwości oferowanych przez podejścia alternatywne.

W szczególności, Doktorant wykazał, że wymagana przez docelowe środowiska implementacji konieczność redukcji rozmiarów stosowanych z powodzeniem w analizie mowy rozbudowanych modeli uczenia głębokiego, zaproponowanymi dotychczas metodami nie przynosi oczekiwanych efektów, a jednym ze sposobów rozwiązania tego problemu jest wprowadzenie zaproponowanych przez Niego rozszerzeń i modyfikacji bazowych algorytmów. Przedmiotem kolejnych trzech podrozdziałów będzie ocena wkładu Doktoranta w doskonalenie metod analizy sygnału mowy, uwzględniających ograniczoność zasobów obliczeniowych dostępnych dla wykonania algorytmu.

3.1. Detekcja ustalonych poleceń słownych

W pierwszym obszarze tematycznym podjętym w rozprawie – detekcji występowania w mowie zadanych fraz, Doktorant rozważa dwa scenariusze stanowiące podstawę komunikacji użytkownika z inteligentnymi asystentami głosowymi, znacząco różniące się jakością informacji dostarczanych algorytmowi w procesie uczenia. Pierwszy z nich, określany jako Query-by-Example (QbyE) zakłada bezpośrednią prezentację algorytmowi wypowiedzi użytkownika mających podlegać wykrywaniu, co stanowi naturalne źródło wiedzy referencyjnej, używanej później podczas funkcjonowania systemu. Drugi, o wiele trudniejszy scenariusz, zakłada brak istnienia wypowiedzi referencyjnych i korzysta z tekstu jako jedynej formy identyfikacji frazy podlegających detekcji (Query-by-Text, QbyT).
Narzucającym się obszarem aplikacji detekcji wystąpienia fraz kluczowych, zaprezentowanych systemowi wcześniej w postaci referencyjnych wypowiedzi użytkownika jest, jak na wstępie wskazał Doktorant, korzystanie z usług różnego rodzaju inteligentnych ‘asystentów’, którym użytkownik przedstawia w fazie ‘douczania’ algorytmu zbiór komend, jakie mają inicjować wykonywanie określonych funkcji. W drugim rozważanym przypadku zastosowaniem, w którym algorytm ma poprawnie rozpoznawać frazy sformułowane w postaci tekstowej jest, jak podejrzewam, kontekst wirtualnego asystenta działającego w warunkach ‘niespersonalizowanych’, np. w miejscach publicznych. Szkoda, że Doktorant nie wskazał jawnie obszaru, który jest adekwatny dla rozważanego kontekstu, bo być może moje wyobrażenie o potencjalnych scenariuszach wykorzystania rozważanego schematu jest mocno ograniczone. Na marginesie należy zauważyć niepokojącą implikację doskonalenia metod detekcji dowolnej wypowiedzi, określonej formalnie tekstem (lub jego fonetycznym odpowiednikiem), dostarczającą lepszych narzędzi o charakterze szpiegowskim, gdzie oprogramowanie (np. urządzenia mobilnego) może podlegać szybkiej adaptacji w celu rejestrowania interesujących treści w monitorowanym otoczeniu.

Mimo deklaracji Doktoranta o podjęciu prac w zakresie obydwu scenariuszy detekcji fraz (zarówno QbyE jak i QbyT), przedstawione nowe koncepcje ograniczają się do drugiego, trudniejszego wątku naukowego. Doktorant w istocie prześlizguje się nad tematyką rozpoznawania w scenariuszu QbyE, traktując go raczej jako kontekst dla ewaluacji utworzonej przy swoim współudziale bazy nagrań, niż jako obszar merytorycznego wkładu do dziedziny. Przedstawiona przez Niego architektura algorytmu detekcji – uczenie kontrastowe modelu rekurencyjnego zaczerpniętego z literatury, jest schematyczna i nie wnosi w mojej opinii zauważalnego wkładu naukowego. Postulowana przez Niego dodatkowa faza dostrajania parametrów ekstraktora cech reprezentujących frazę,

wykorzystująca dodatkowy zbiór danych, na pewno jest pomysłem ciekawym, ale nie stanowi pomysłu o naukowo znaczącym kalibrze.

Główną wartością części rozprawy poświęconej schematowi QbyE rozpoznawania jest obszerna prezentacja opracowanej przy współudziale Doktoranta, bardzo wartościowej bazy nagrań, dedykowanej problemowi detekcji zadanych fraz w mowie swobodnej. Na uznanie zasługuje w szczególności sformułowanie zbioru warunków, jakie według Autorów powinny być spełnione przez zbiory stosowane do ewaluacji algorytmów detekcji wystąpienia zadanych fraz, z których za najistotniejszy można uznać wymaganie zapewnienia odpowiedniego stopnia trudności opracowanego materiału. Środkiem do osiągnięcia tego celu jest umieszczenie zbioru ‘negatywnych’ przykładów cechujących się maksymalnym możliwym podobieństwem do przykładów pozytywnych, gdzie jako miara oceny podobieństwa proponowana jest odległość edycyjna między fonetycznymi reprezentacjami przykładów obydwu grup. Implementacja przedstawionej zasady doprowadziła do uzyskania zbioru uczącego, sprawiającego standardowo stosowanym algorytmom analizy znacząco większą trudność niż istniejące bazy nagrań, stosowane w dotychczasowych pracach.

Daleko cenniejszym fragmentem rozdziału jest część poświęcona tworzeniu algorytmu detekcji zadanych fraz w mowie swobodnej dla drugiego z rozważanych scenariuszy rozpoznawania, czyli QbyT. Aby umożliwić satysfakcjonującą realizację zadania w architekturach sprzętowych urządzeń mobilnych, Doktorant formułuje dwie propozycje rozbudowy ścieżki przetwarzania sygnału mowy, bazującej na zredukowanym metodą destylacji wiedzy ‘głębokim’ bazowym modelu akustycznym (AM). Pierwszym pomysłem jest dodanie prostego modelu językowego jako elementu ważącego fonetyczne hipotezy generowane przez AM. Drugim, jest włączenie AM do budowy zbioru referencyjnych realizacji rozważanych w detekcji fraz. Doktorant w pomysłowy sposób wzbogaca wiedzę modułu podejmowania decyzji odnośnie oczekiwanej formy tekstowej zadanej wypowiedzi o dodatkowe modele akustyczne zadanych fraz, budowane na podstawie zbioru próbek mowy wygenerowanych przez moduł Text-To-Speech i przetworzonych przez moduł AM. Wyznaczone w ten sposób reprezentacje uzupełniają informację referencyjną, stosowaną w weryfikacji hipotez generowanych przez rozważany algorytm dla sygnałów pozyskiwanych w fazie działania algorytmu.

Architektura obliczeniowa zastosowana przez Doktoranta do realizacji zadania detekcji zadanych fraz jest dobrana w sposób właściwy: ścieżka przekształcania wejściowych sygnałów rozpoczyna się realizowaną przez model AM identyfikacją zawartych w sygnale jednostek fonetycznych, które są następnie agregowane w najbardziej prawdopodobne sekwencje z użyciem metody przeszukiwania Beam Search i konfrontowane z referencyjnym słownikiem w celu podjęcia decyzji co do wystąpienia lub nie poszukiwanej frazy. Mankamentem przedstawionego opisu jest jego nadmierna powierzchowność utrudniająca śledzenie rozważań. Czytelność ścieżki przetwarzania danych zyskałaby np. dzięki identyfikacji przykładowych reprezentantów zbioru generowanych przez używany przez niego model akustyczny ‘sub-słów’ (czy są to fonemy trójfony czy inne jednostki fonetyczne?). Doktorant nie informuje o szerokości ‘wiązki’ (beam) przyjętej w procedurze przeszukiwania. Nie wiadomo również, jaka metoda destylacji wiedzy została przez Niego użyta w celu redukcji rozmiaru AM i jakie są ilościowe różnice wskaźników oceny jakości analizy między modelem nauczyciela i modelem studenta. Wreszcie, Doktorant jedynie wzmiankuje o wykorzystanych w treningu modelu AM funkcjach straty, będących kluczowymi dla uzyskania

jego optymalnego działania. Powierzchowność komunikacji nie pozwala na dokładne zrozumienie tej procedury: połączenie funkcji oceny CTC z entropią skrośną (CE) nie jest oczywiste, również nie wiadomo jak zdefiniowana jest funkcja oceny jakości ‘destylacji’ wiedzy.
Zdecydowanie bardziej obszerna i klarowna jest eksperymentalna weryfikacja przedstawianych przez Doktoranta wariantów procedury. Doktorant przedstawia wyniki oceny poprawności detekcji w funkcji przyjętego dopuszczalnego progu odstępstwa hipotezy od wzorca i konfrontuje efekty stosowania zaproponowanych modyfikacji z wynikami uzyskanymi dla modelu bazowego. Przeprowadzona na końcu dyskusja otrzymanych wyników, która sprowadza się do zebrania najistotniejszych zaobserwowanych faktów, mogłaby być ciekawsza, gdyby zawierała próby identyfikacji lub wyjaśnienia odpowiedzialnych za nie mechanizmów (co nie jest oczywiście proste).

3.2. Detekcja końca wypowiedzi

Drugim obszarem prac Doktoranta była analiza możliwości poprawy realizacji zadania detekcji końca wypowiedzi, stanowiącej istotny element interfejsów komunikacji z wirtualnymi asystentami (przedwczesna reakcja uniemożliwia poprawną reakcję systemu, zaś zbyt długa zwłoka czyni interakcję trudną do zaakceptowania). Przedstawione zadanie wbrew pozorom nie jest proste, bowiem określenie właściwego momentu zakończenia wypowiedzi w warunkach istnienia rzeczywistego tła akustycznego oraz indywidualnej i kontekstowej zmienności tempa i stylu wypowiedzi, nie pozwala na zastosowanie narzucającej się metody ‘detekcji ciszy’ o czasie trwania przekraczającym założony próg. Narzędziem realizacji zadania stała się ponownie głęboka, rekurencyjna sieć neuronowa, a istotą pomysłu Doktoranta jest różnicowanie podczas jej uczenia istotności fragmentów przedziału czasu zawierającego faktyczny koniec wypowiedzi, dokonywane poprzez wprowadzenie w kryterium BCE (binarnej entropii skrośnej) oceny poprawności klasyfikacji współczynników ważących istotność ramek sygnału mowy dla podejmowania decyzji. Zaproponowany schemat ważenia znaczenia ramek: intuicyjne dążenie do całkowitej eliminacji błędów przedwczesnej detekcji i karania za wydłużającą się zwłokę w podejmowaniu decyzji, nie budzi zastrzeżeń, podobnie jak użyte przez Doktoranta narzędzie klasyfikacji.
Metryki zastosowane w ocenie jakości działania algorytmu zostały dobrane w sposób zapewniający jej właściwą ewaluację, pozwalając na wykazanie zasadniczej przewagi zaproponowanego podejścia nad (właściwie dobranym) rozwiązaniem referencyjnym, ujawniającej się szczególnie wyraźnie podczas testów wytrenowanego algorytmu na wymagających zbiorach danych, zawierających próbki mowy z silnym tłem akustycznym (własna baza IN-HOUSE i LibriSpeech). Uzyskane wyniki pozwoliły jednocześnie na optymalizację parametrów algorytmu: estymację liczby ramek poddawanych analizie (optymalna szerokość horyzontu analizy) i przeprowadzenie analizy wpływu wartości pozostałych parametrów (wag) stosowanych w kryterium uczenia sieci. Efektem prac w zakresie przedstawionego wątku tematycznego jest potwierdzenie przydatności zaproponowanej funkcji kryterialnej, pozwalające na poprawę dokładności detekcji końca wypowiedzi swobodnych, dla warunków uwzględniających realistyczne poziomy i rodzaje tła akustycznego, co stanowi osiągnięcie o istotnym znaczeniu praktycznym.

Niejasnym fragmentem opisu sposobu przygotowania próbek do analizy jest informacja o przedłużaniu nagrań o dwusekundowe fragmenty ‘ciszy’ w odniesieniu do części nagrań. Nie wiem,

jaka była dokładna realizacja ‘ciszy’ dodawanej do nagrania, ale wydaje się, że ten zabieg wprowadza nienaturalne uwarunkowanie dla zadania detekcji końca wypowiedzi.

3.3. Częściowo nadzorowane uczenie algorytmów rozpoznawania mowy

Trzeci wątek prac Doktoranta dotyczył poszukiwania sposobów kompensacji spadku poprawności rozpoznawania mowy w modelach upraszczanych na potrzeby ich implementacji w urządzeniach mobilnych, w drodze nadzorowanego i częściowo-nadzorowanego (Semi-Supervised Learning - SSL) douczania pretrenowanych modeli bazowych. Pierwszym pomysłem na zwiększenie poprawności rozpoznawania przy użyciu modelu o rozmiarach realnych do jego efektywnego uruchomienia w warunkach ograniczonych zasobów, jest jego dotrenowanie z wykorzystaniem niewielkiego etykietowanego zbioru przykładów, niepowiązanego ze zbiorem użytym do treningu zbioru bazowego. Skorzystanie z nowego źródła informacji ma na celu wskazanie modelowi bazowemu potencjalnych odstępstw od struktury danych stosowanych w jego treningu, co powinno pozwolić na przystosowanie algorytmu do pracy w warunkach docelowych, tym lepsze, im bliższe docelowym są charakterystyki zbioru stosowanego w dostrajaniu (np. specyficzne warunki i parametry toru akwizycji dźwięku, inne tło akustyczne oraz inny charakter zbioru użytkowników).
Drugim pomysłem na zwiększenie precyzji rozpoznawania mowy przez algorytmy działające w środowiskach mobilnych jest użycie mechanizmu uczenia częściowo-nadzorowanego w iteracyjnej procedurze dostrajania modelu bazowego. Podobnie jak poprzednio, źródłem informacji jest zbiór danych niewykorzystywany w budowie modelu bazowego, ale tym razem, jest on dużo obszerniejszy i zawiera zarówno próbki etykietowane (nieliczne) jak i próbki bez etykiet. Istota zaproponowanej modyfikacji jest nieco rozproszona w opisach ‘mechaniki’ jej stosowania – intencją Doktoranta jest naprzemienna predykcja właściwych etykiet (realizowana na podstawie aktualnej postaci modelu) i douczanie modelu, przy czym argumentem douczania może być w każdej iteracji wyjściowy model bazowy (efektywnie dokonywana jest tylko jedna modyfikacja jego parametrów, następująca w ostatniej epoce uczenia), albo model stopniowo ewoluujący w miarę postępu uczenia, startujący w pierwszej iteracji z modelu bazowego i poddawany sekwencji kolejnych modyfikacji. Doktorant rozważa dwa scenariusze procedury douczania: w pierwszym z nich do korygowania modelu wykorzystuje wyłącznie próbki nieetykietowane, zaś w drugim, zarówno próbki nieetykietowane, jak i próbki posiadające oryginalne etykiety.

Przedstawione propozycje, mimo przekonywującej intuicji, mają czysto heurystyczny charakter, więc ciężar wykazania słuszności przedstawionych pomysłów spoczywa na wynikach eksperymentalnej weryfikacji ich przydatności. Przedstawiona staranna, obszerna i dobrze przemyślana metodyka testowania zaproponowanych rozwiązań pozwala na sformułowanie przekonywujących wniosków. Punktem wyjścia testów jest dobór właściwej architektury bazowej (ponownie, odpowiednio zredukowanej głębokiej sieci neuronowej, tym razem o strukturze rekurencyjnego kodera-dekodera), zapewniającej zadowalającą jakość rozpoznawania mowy, a zarazem realnej z punktu widzenia jej implementacji w urządzeniach mobilnych. Dla oceny porównawczej działania algorytmu bazowego bez i z wprowadzonymi przez Niego modyfikacjami Doktorant używa szeregu zbiorów uczących o bardzo zróżnicowanych charakterystykach. Przedmiotem porównań jest konfrontacja jakości działania różnych wariantów algorytmu, ocenianych za pomocą współczynnika błędnej identyfikacji słów (Word Error Rate – WER), prowadząca do wykazania przydatności zaproponowanych przez Doktoranta pomysłów, przy czym

wydaje się, że najbardziej przydatnym wariantem jest dotrenowywanie modelu bazowego w schemacie pojedynczej modyfikacji z wykorzystaniem zarówno etykietowanych jak i nieetykietowanych przykładów. W dyskusji uzyskanych wyników Doktorant przedstawia trafne spostrzeżenia dotyczące związków między bogactwem słownikowym dodatkowego zbioru danych i możliwą do uzyskania poprawą działania modelu. Podsumowując ocenę przedstawionego materiału chcę stwierdzić, że prace Doktoranta stanowią interesującą, oryginalną i praktycznie zweryfikowaną propozycję metodyki przydatnej w poprawie jakości realizacji analiz z użyciem uproszczonych modeli głębokich sieci neuronowych, o zakresie stosowalności wykraczającym poza rozważany wątek tematyczny.

4. Uwagi szczegółowe

W pracy znajdują się bardzo nieliczne niejasności wymagające doprecyzowania lub drugorzędne usterki o różnym charakterze, które przedstawiam w formie poniższej, być może niekompletnej listy:

17. *„Those errors propagate to NLG, which will fail to parse the user's command."*
Prawdopodobnie miało być NLU

20-22. Niepotrzebne powtórzenia: dwukrotna prezentacja tego samego zbioru danych GSC z wyjaśnianiem akronimu, odwoływaniem do literatury i charakterystyką; dwukrotne odesłanie do pozycji [6], anonsowane tą samą informacją.

32. … *contains approximately 500 k test cases*
Czy 'test-case' to para przykładów?

33. … *We pre-trained the encoder model as a part of the Listen, Attend, Spell model …*
Szkoda, że nie ma tu więcej wyjaśnień – pozwoliłoby to na uzyskanie lepszej, konkretnej orientacji zarówno co do celów treningu, jak i do roli używanego przez Doktoranta modułu rozpoznawania w użytej architekturze.

36. Z jakiej definicji WER korzysta Doktorant? - jakiej normalizacji podlegają podawane przez Niego miary (wartości 16 czy 30.9 nic nie mówią o jakości systemu). Być może przytoczone wartości to procenty?

49. 0 + 1+
Doktorant odwołuje się do nieistniejącego kanonu symboliki definiowania wyrażeń regularnych. Czy znak '+' oznacza ciąg takich samych znaków o dowolnej długości, a więc podane wyrażenie oznacza sekwencję etykiet przypisanych kolejnym ramkom, najpierw nie spełniających kryterium końca wypowiedzi, a następnie, spełniających?

50. Użyty we wzorze (3) symbol nawiasu domkniętego od dołu oznacza, jak rozumiem, funkcję 'floor' - to powinno być jawnie wyjaśnione.

60. *We justify this approach by the fact that however good the labels generated by the model are, they can still contain errors.*
Przedstawione zdanie ma wyjaśnić intuicję stojącą za 'resetowaniem' modelu po każdej epoce (tak wynika z algorytmu 3). Jeśli taki miał być zamiar, to przedstawione tłumaczenie niewiele wyjaśnia.

63. … *known as Word Error Rate (WER) in the ASR research.*

WER pojawiało się w tekście już kilka razy, więc anonsowanie co to jest w tym miejscu jest niedopatrzeniem.

5. Wniosek końcowy

Podsumowując opinię o rozprawie doktorskiej Pana magistra Mikołaja Pudo, chcę jednoznacznie stwierdzić, że prezentuje ona zbiór trzech metod w oryginalny sposób rozwiązujących trudne problemy naukowe dotyczące wybranych aspektów metodyki rozpoznawania mowy, ukierunkowane na implementację w urządzeniach o istotnie ograniczonych zasobach obliczeniowych. W ten sposób uzyskane przez Doktoranta wyniki dostarczają rozwiązań o wyraźnym wymiarze praktycznym, co stanowi niewątpliwy dodatkowy walor rozprawy. Osiągnięcia Doktoranta wypełniają według mnie wymagania stawiane rozprawom doktorskim przez odnośne przepisy, dlatego też **wnioskuję o dopuszczenie jej Autora do dalszych etapów przewodu doktorskiego**.

Signed by /
Podpisano przez:

Krzysztof Maciej
Ślot
Politechnika Łódzka

Date / Data: 2024-
01-29 14:13