Łódź, 15. 06. 2020 r.

Prof. dr hab. inż. Danuta Rutkowska
Instytut Technologii Informatycznych
Społeczna Akademia Nauk w Łodzi

RECENZJA

rozprawy doktorskiej

p.t. „New approaches in speech recognition from isolated words to practical solutions"

Autor: mgr inż. Mohammad Kheir Nammous

Niniejsza recenzja została przygotowana w odpowiedzi na pismo otrzymane z Politechniki Warszawskiej, informujące o powołaniu mnie przez Radę Wydziału Matematyki i Nauk Informacyjnych na recenzenta tej rozprawy doktorskiej, w dniu 27 czerwca 2019 r, oraz na ponowną prośbę zawartą w piśmie z dnia 15 maja 2020, po otrzymaniu poprawionej wersji pracy Doktoranta.

1. PROBLEMATYKA I CEL ROZPRAWY

Głównym celem tej rozprawy doktorskiej jest rozwinięcie podejścia automatycznej identyfikacji mówcy (osoby) poprzez rozpoznawanie mowy (głosu), na podstawie próbek pochodzących z dużych baz danych, przy wykorzystaniu możliwie najmniejszej ilości informacji, co – według Doktoranta – upraszcza proces rozpoznawania.

Uważam, że to ważny problem, by nie wykorzystywać nadmiarowych danych, wszystkich informacji dostępnych w tych bazach, lecz tylko to co jest najistotniejsze, a jednocześnie wystarczające do poprawnego rozpoznania. W zagadnieniach klasyfikacji wybór najbardziej istotnych atrybutów stanowi etap wstępnego przetwarzania danych (preprocessing) w celu uproszczenia zadania i redukcji złożoności obliczeniowej.

Uproszczenie rozważanego problemu na etapie preprocessingu Doktorant traktuje jako jeden z oryginalnych rezultatów swojej rozprawy doktorskiej.

W różnych zadaniach klasyfikacji - i ogólnie *Data Mining* - sam problem redukcji wymiarowości oraz wyboru najistotniejszych atrybutów nie jest niczym nowym, jednakże konkretne metody i algorytmy pozwalające rozwiązać ten problem i uniknąć tzw. „przekleństwa wymiarowości" są stosowane i udoskonalane.

Przedłożona do recenzji rozprawa doktorska mgra inż. Mohammada K. Nammousa dotyczy biometrii, w szczególności identyfikacji osób na podstawie cechy biometrycznej, jaką jest ludzki głos. Jest to problem związany z *Pattern Recognition*, czyli ogólnie rozpoznawaniem wzorców. Metody stosowane do rozwiązywania zadań z tego zakresu, również te prezentowane w recenzowanej pracy doktorskiej to często tzw. metody sztucznej inteligencji *(Artificial Intelligence)*, czy inteligencji obliczeniowej *(Computational Intelligence)*. Znanym przykładem tej grupy metod są sztuczne sieci neuronowe *(Neural Networks)*, wykorzystywane przez Doktoranta w swojej rozprawie doktorskiej. Są to oczywiście zagadnienia i metody zaliczane do informatyki.

2. FORMA REALIZACJI ROZPRAWY

Recenzowana obecnie rozprawa doktorska mgra inż. Mohammada K. Nammousa została napisana w języku angielskim i wydana jako Ph.D. Thesis Wydziału Matematyki i Nauk Informacyjnych Politechniki Warszawskiej, w Warszawie (Warsaw) 2020; podobnie jak poprzednia wersja (z roku 2019). Praca ta liczy 174 strony i została rozszerzona w stosunku do poprzedniej wersji (122 strony) o ponad 50 stron. Składa się teraz z 8-miu rozdziałów i dwóch dodatków (poprzednio było 9 rozdziałów). Ponadto zawiera „*Abstract*" w języku angielskim oraz odpowiadające mu „Streszczenie" po polsku, a także spis treści (*Table of Contents*), wykaz rysunków (*List of Figures*), wykaz tabel (*List of Tables*), jak również „*List of Listings*", „*Preface*", czyli przedmowę. To wszystko znajduje się na pierwszych stronach, przed „*Chapter 1*". Natomiast na ostatnich stronach (po „*Chapter 8*") Autor zamieścił dwa dodatki („*Appendix A*" i „*Appendix B*"). Pierwszy z nich przedstawia przykłady kodu programu komputerowego, wykorzystywanego przez Doktoranta m.in. do uczenia sieci neuronowej w środowisku programistycznym MATLAB. W istocie Dodatek A zawiera 4 listingi kodu komputerowego, wykazane na początku w „*List of Listings*". W drugim dodatku Doktorant prezentuje listę swoich publikacji (z uaktualnionymi danymi odnośnie pierwszej publikacji na tej liście – artykuł w druku), Ponadto Autor zamieszcza listę stosowanych w pracy akronimów, co jest bardzo przydatne podczas studiowania tej rozprawy. Warto zauważyć, że ta lista akronimów została znacznie wydłużona w aktualnej wersji monografii. Całą książkę kończy wykaz bibliografii, liczący 183 pozycje, do których Autor odwołuje się w swojej pracy. Bibliografia została rozszerzona o ponad 30 pozycji cytowanych w tej rozprawie doktorskiej, w stosunku do poprzedniej wersji (152 pozycje).

Treść poszczególnych rozdziałów, w obecnej postaci rozprawy doktorskiej mgra inż. Mohammada K. Nammousa, została uzupełniona i uaktualniona. Wobec tego znacznie rozszerzona jest też lista rysunków (List of *Figures*) oraz tabel (*List of Tables*). Zmieniono też „*Abstract*" oraz odpowiadające mu „Streszczenie" w języku polskim i oczywiście spis treści (*Table of Contents*). Należy wyraźnie zaznaczyć, że zmiany dokonane w spisie treści, a więc w strukturze pracy są bardzo duże. Spis treści został znacznie wydłużony, gdyż pojawiły się nowe podrozdziały. Zmianie uległo też wiele tytułów rozdziałów i podrozdziałów.

Rozdział 1 stanowi wprowadzenie do tematyki rozprawy. Został on znacznie uzupełniony w stosunku do poprzedniej wersji pracy. Mimo, że część materiału przesunięto do rozdziału 3, jest on bardziej obszerny i zawiera nowe treści, m.in. przedstawiające miary wykorzystywane do oceny działania stosowanych metod. Istotnym uzupełnieniem jest też sformułowanie przez Doktoranta celów realizowanej rozprawy doktorskiej. Na końcu tego rozdziału, podobnie jak poprzednio, Autor wskazuje, które rozdziały zawierają oryginalny wkład w tematykę rozprawy; tym razem są to rozdziały od 4 do 7, czyli *Chapter 4 - Chapter 7*. Konkluzje i dyskusja znajdują się, tak jak wersji poprzedniej, w rozdziale ostatnim, czyli teraz w *Chapter 8*.

Rozdział 2 jest zatytułowany podobnie jak w poprzedniej wersji monografii, czyli „*The State of the Art*" i również prezentuje stan wiedzy dotyczący tematyki rozprawy Jednak jest on teraz znacznie dłuższy i rozbudowany o wprowadzone podrozdziały. Pierwszy z nich stanowi przegląd metod stosowanych do rozpoznawania głosu. Podobnie jak w poprzedniej wersji pracy, Doktorant najwięcej uwagi poświęca sztucznym sieciom neuronowym. W obecnym wydaniu monografii pojawiły się ważne podrozdziały, pod wspólnym tytułem „*Speaker Recognition Systems*", czyli systemy rozpoznawani mówcy. Cztery podrozdziały w tej części rozprawy odpowiadają rozdziałom 4-7, które dotyczą oryginalnego wkładu Doktoranta. Znalazły one też odzwierciedlenie w tabeli (*Table 3.1*) wprowadzonej do rozdziału 3 w tej wersji pracy. Ostatni podrozdział opisuje wkład Doktoranta w prezentowaną rozprawę i jednocześnie podsumowuje Jego dorobek publikacyjny.

Rozdział 3 zawiera opis różnych metod, stosowanych m.in. w przetwarzaniu sygnałów, preprocessingu, ekstrakcji cech, jak też klasyfikacji. Autor prezentuje tu głównie metody, które także wykorzystuje w swojej pracy. W nowej wersji monografii zamieszczono tu również fragment dotyczący opisu narządu mowy w aspekcie biologicznym, przeniesiony z rozdziału 1. Przede wszystkim w rozdziale 3 dodano tabelę (*Table 3.1*) odnoszącą się do rozdziałów 4-7, zawierających oryginalny wkład Doktoranta. Tabela ta przedstawia informacje o zastosowanych danych, proporcjach podziału na dane uczące i testujące, zastosowanych metodach ekstrakcji cech i klasyfikacji, a także wskazuje cele – wyszczególnione w rozdziale 1.3 – w odniesieniu do każdego rozważanego przypadku rozpoznawania mówcy, analizowanego w rozdziałach 4-7. Ponadto dokonano drobnych uzupełnień przy opisie niektórych metod, w szczególności tych najbardziej istotnych z punktu widzenia realizacji rozprawy. Pod względem redakcyjnym - wyeksponowano podane wzory oraz niektóre wypunktowane treści, nadano wyrazistości rysunkom. W największym stopniu uzupełniono podrozdział dotyczący macierzy Toeplitza, co jest bardzo ważne z punktu widzenia tej pracy i związanych z nią publikacji, a czego brakowało w poprzedniej wersji rozprawy. Dodano też istotny fragment w podrozdziale na temat probabilistycznych sieci neuronowych. Mówi on o architekturze tej sieci oraz parametrach, odwołując się do zastosowania tej sieci w rozprawie Doktoranta, nawiązując do wcześniej przedstawionych metod. W podrozdziale na temat sieci RBF dokonano też modyfikacji, przede wszystkim poprawiono błędy (m.in językowe) występujące w poprzedniej wersji pracy. Ponadto, podobnie jak dla sieci probabilistycznych, dodano fragment na temat architektury sieci

radialnych (RBF, które Autor alternatywnie oznacza jako RNN, czyli *Radial Neural Networks*).

Rozdziały od 4 do 7 zawierają zagadnienia omawiane w rozdziałach 4-8 w poprzedniej wersji monografii, czyli m.in. opis używanych baz danych, przeprowadzane eksperymenty i otrzymane rezultaty. Jednak w obecnym wydaniu ta część pracy została rozszerzona, uzupełniona i przeorganizowana pod względem redakcyjnym.

Rozdział 4 zmienił nazwę w nowym wydaniu pracy. Dodano też kilka podrozdziałów, zachowując zasadniczą część z wersji poprzedniej.

Aktualnie rozdział 5 obejmuje zagadnienia rozdziałów 5 i 6 z pierwszego wydania. Dlatego monografia liczy teraz 8 rozdziałów (a nie 9, jak w wersji poprzedniej).

Rozdział 6, który jest odpowiednikiem rozdziału 7 w starej wersji, nosi obecnie podobny tytuł, lecz zawiera znacznie więcej stron i dużo nowych treści.

Rozdział 7, o takim samym tytule jak rozdział 8 w wersji poprzedniej, także został uzupełniony o kilka stron.

Rozdział 8, czyli ostatni, ma tytuł taki jak rozdział 9 w pierwszym wydaniu, czyli „*Conclusions and Discussion*". Zawiera podsumowanie, które odnosi się do najważniejszych aspektów rozprawy doktorskiej.

Aktualna wersja rozprawy prezentuje się jako nowe wydanie monografii - poprawione, uzupełnione i uaktualnione.

3. UWAGI REDAKCYJNE

Pod względem redakcyjnym praca jest napisana starannie, przejrzyście, z odpowiednio umieszczonymi podpisami pod rysunkami i nagłówkami nad tabelami oraz odsyłaczami do nich w tekście. Usterki redakcyjne (niezbyt liczne) zauważone w poprzedniej wersji monografii zostały w większości poprawione. Mimo to wciąż można zauważyć np. literówki, niepotrzebny przecinek, rozpoczynanie zdanie od skrótu „Eg." (zamiast „For example"), czy „Where" pod wzorami pisane dużą literą zamiast małą. Słowo „equation" czasami jest w całości, natomiast w innym miejscu w postaci skrótu „eq." Można wskazać jeszcze kilka drobnych usterek redakcyjnych ale tego typu uwagi są nieistotne i nie mają wpływu na ocenę pracy.

4. ZAGADNIENIA PREZENTOWANE W ROZPRAWIE DOKTORSKIEJ

W rozdziale 1 Doktorant wyjaśnia, że w szeroko pojętej tematyce rozpoznawania mowy (*speech recognition*) skupia się głównie na rozpoznawaniu mówcy w kontekście praktycznych zastosowań i bezpieczeństwa, pokazując jak różne czynniki wpływają na proponowane rozwiązania. W tym zakresie w swojej rozprawie doktorskiej przedstawia przypadki analizowane z punktu widzenia dokładności działania. Jako rozszerzenie w stosunku do poprzedniej wersji rozprawy Autor prezentuje różne, znane w literaturze, miary oceny działania systemów, w tym dokładności. W rozdziale tym Doktorant ilustruje też jak atrybut głosu jest umiejscowiony na tle innych cech biometrycznych. Ponadto krótko omawia zastosowane w swojej pracy metody i realizowane cele szczegółowe oraz ogólną strukturę monografii.

W rozdziale 2 (*The State of the Art*) Doktorant wspomina o znanych współcześnie aplikacjach dotyczących rozpoznawania głosu, takich jak *Google Voice Search*, czy też *Siri* firmy *Apple* oraz *Alexa* firmy *Amazon*. Podkreśla jednak, że wciąż istnieje przestrzeń do rozwoju badań w tym obszarze zastosowań. W swojej pracy doktorskiej skupia się na aspektach rozpoznawania mówcy, w szczególności na algorytmach *text-dependent* i *text-independent*, weryfikacji mówcy, a także rozpoznawania języka oraz płci osoby mówiącej. W rozdziale tym Doktorant dokonuje przeglądu metod znanych w literaturze, potencjalnych zastosowań i zbiorów danych dotyczących rozpoznawania mowy. W odniesieniu do swojej rozprawy doktorskiej, rozważa aspekty rozpoznawania mówcy, kładąc nacisk na wykorzystanie małej porcji danych do uczenia modelu.

W rozdziale 3 (*Foundation of the Applied Methods*) Doktorant przedstawia podstawowe koncepcje konieczne do zrozumienia zastosowanych w rozprawie technik, metod i algorytmów. Wyjaśnia, że z naukowego punktu widzenia rozpoznawanie głosu (*voice recognition*) jest obszarem badań informatyki (*Computer Science*), zależnym od oprogramowania (*computer software*) i algorytmów wykorzystywanych do identyfikacji różnych cech, zarówno mowy, jak też mówców. Autor omawia te cechy, sposoby ich ekstrakcji i kodowania oraz metody przetwarzania sygnałów głosowych.

Rozdział 4 (*Text-Dependent Speaker Identification*) dotyczy systemów identyfikacji mówcy. Rozważane są systemy, które potrafią rozpoznać osobę mówiącą na podstawie analizy sygnałów mowy. Takie systemy wymagają wypowiadania dokładnie słów, co jest wykorzystywane zarówno w procesie uczenia, jak też rozpoznawania. Systemy typu *text-dependent* są łatwiejsze w implementacji niż *text-independent* i dają większy poziom bezpieczeństwa. Doktorant w rozdziale 4 proponuje *text-dependent* system identyfikacji mówcy w oparciu o dane zawierające wypowiadane głosem cyfry arabskie. W tym przypadku do klasyfikacji stosowane są zarówno klasyczne metody, jak też sztuczne sieci neuronowe. Autor rozpoczyna od prostszego przypadku, czyli rozpoznawania pojedynczych słów (wypowiadanych cyfr: 0, 1, 2), a następnie prezentuje wielopoziomowy system i dokonuje weryfikacji mówcy.

Rozdział 5 (*Text-Independent Speaker Identification*) dotyczy systemów identyfikacji mówcy w przypadku, gdy nie ma znaczenia jakie słowa są wypowiadane. Próbki danych uczących i testujących mogą stanowić na przykład zarejestrowany materiał z przekazu wiadomości telewizyjnych, czy radiowych. Systemy typu *text-independent* są trudniejsze w implementacji ale mają zastosowanie w weryfikacji osób na podstawie głosu. Doktorant przeprowadził eksperymenty dla odpowiednio przygotowanych danych, w różnych wariantach, wykorzystując prezentowane wcześniej metody. Dokonał porównania otrzymanych rezultatów. Zrealizował szczegółowe cele postawione w swojej rozprawie doktorskiej, m. in. dotyczące zadania weryfikacji mówcy.

Rozdział 6 (*Language, Gender and Speaker Identification*) dotyczy systemów identyfikacji mówcy, gdzie rozpoznawany jest również język osoby mówiącej oraz płeć. Autor prezentuje opis zbioru danych oraz warunki eksperymentu i rezultaty oraz dyskusję otrzymanych wyników. Warto przypomnieć, że treść tego rozdziału została znacznie wzbogacona i rozszerzona w stosunku do poprzedniej wersji tej pracy. Jest to szczególnie widoczne w odniesieniu do eksperymentów z zastosowaniem sieci neuronowych.

Rozdział 7 (*Large-Scale Speaker Identification*) dotyczy systemów identyfikacji mówcy, przy czym rozważa się problem w znacznie większej skali. Doktorant wykorzystuje bazę zawierającą dane pochodzące od ponad 4 tysięcy osób. Autor opisuje ten zbiór oraz swoje podejście i wyniki.

Rozdział 8 zawiera konkluzje i dyskusję w ramach podsumowania pracy. Autor porównuje systemy identyfikacji mówcy w przypadkach „*text-dependent*" oraz „*text-independent*". Wymienia też nowatorskie punkty swojego podejścia.

5. ODNIESIENIA DO UWAG W RECENZJI POPRZEDNIEJ WERSJI PRACY

Należy podkreślić, że w nowej wersji swojej pracy doktorskiej Autor uwzględnił następującą uwagę zawartą w recenzji rozprawy w wydaniu poprzednim.

Uwaga 1:

W rozdziale 1, w odniesieniu do tabeli 1.1, Autor pisze, że rozpoznawanie głosu ma przewagę, czyli jest korzystniejsze niż identyfikacja na podstawie innych cech biometrycznych, takich jak odcisk palca (fingerprint), czy tęczówka oka (iris). Jednakże w tej tabeli widać przewagę jedynie w aspekcie kosztu (cost) i prostoty (simplicity). Natomiast widoczna jest słabość tej cechy przede wszystkim w aspekcie „age stability", a też pod względem „emotional stability" i „accuracy" nie jest lepiej.

„Age stability" dotyczy zmiany danej cechy biometrycznej wraz z wiekiem, co utrudnia identyfikację (rozpoznanie osoby na podstawie tej cechy). W przeciwieństwie do głosu, tęczówka oka człowieka oraz odcisk palca pozostają niezmienne.

„Emotional stability", czyli stabilność emocjonalna, nie ma znaczenia przy rozpoznawaniu osób na podstawie tęczówki oka lub odcisków palców, natomiast głos nie jest emocjonalnie stabilny. Z drugiej jednak strony warto zaznajomić się z pracami wielu autorów dotyczącymi rozpoznawania emocji w ludzkim głosie. Doktorant w zasadzie pominął ten aspekt bardzo ciekawych badań, choć wspomina o takich pracach i niektóre cytuje.

„Accuracy" jest z pewnością bardzo ważnym wskaźnikiem skuteczności rozpoznawania na podstawie danej cechy biometrycznej. W tabeli 1.1. widzimy, że dla wszystkich wymienionych cech, za wyjątkiem geometrii dłoni i właśnie głosu, „accuracy" nie jest najlepsze. Zatem analizując tę tabelę, trudno zgodzić się ze stwierdzeniem Doktoranta o korzystnej przewadze głosu nad innymi cechami biometrycznymi.

Mimo to uważam, że warto podejmować badania dotyczące analizy ludzkiego głosu jako cechy biometrycznej. Bogata literatura pokazuje, że istnieje wiele prac na ten temat i Doktorant niewątpliwie wpisuje się w ten nurt działalności naukowo-badawczej.

Nienajlepsze „accuracy" może być inspiracją do badań zmierzających do poprawy tego wskaźnika i Autor o tym wspomina.

Aktualny komentarz: W nowej wersji swojej rozprawy doktorskiej Autor dodał wyjaśnienie uzasadniające dlaczego, mimo zalet i wad różnych cech biometrycznych, głos jest przedmiotem zainteresowania biometrii, wskazując na rozwój badań w zakresie rozpoznawania mowy. Cała strona takiego uzupełnienia w odniesieniu do tabeli 1.1. pojawiła się w nowym wydaniu monografii. Została również rozszerzona lista cytowanych publikacji.

Warto też pokazać jak Doktorant potraktował następującą uwagę z poprzedniej recenzji:

Uwaga 2:

Doktorant w swojej rozprawie doktorskiej wielokrotnie podkreśla, że głównym zagadnieniem w odniesieniu do zbioru danych oraz stosowanych sieci neuronowych jest odpowiedni stosunek wielkości ciągu uczącego do testującego. W rozdziale 1, na str. 24, Autor pisze, że w większości przypadków, w eksperymentach prowadzonych przez różnych badaczy, wybiera się znacznie większą część zbioru danych do uczenia i niewielką do testowania. Dlatego, co Doktorant wyraźnie zaznacza, w Jego podejściu rozważa jak najmniej (tak mało jak to możliwe) danych do uczenia rożnych modeli sieci neuronowych. W swoich eksperymentach stosuje krótkie fragmenty sygnałów głosowych, chcąc uzyskać wysoki wskaźnik „accuracy". Skupia się na wspomnianym stosunku danych uczących do testujących (training to testing ratios), traktując to jak niezwykle istotny faktor, wyzwanie w realizacji celu, jakim jest osiągnięcie wysokiego „accuracy" przy wykorzystaniu jak najmniej danych, o czym przypomina też np. w rodziale 8.

Brakuje mi uzasadnienia dla takiego podejścia. Postępowanie inne niż powszechnie stosowane zdecydowanie wymaga odpowiedniego wyjaśnienia. Oczywiście wybór możliwie najmniejszego (ale wystarczającego) zbioru danych jest usprawiedliwiony, o czym wspominam na początku tej recenzji. Jednak wiadomo, że sieci neuronowe uczą się tym lepiej im więcej danych dostarcza się im do uczenia, w postaci tzw. ciągu uczącego (training data). Zwykle, zwłaszcza przy ograniczonej liczbie danych, wybiera się ich jak najwięcej do uczenia,

pozostawiając niewielką część do testowania. Ważny jest jednak nie stosunek danych uczących do testujących, lecz wielokrotny podział zbioru danych w takim stosunku, czyli „cross-validation". Rezultaty badań Doktoranta, otrzymane dla znacznie większych zbiorów testujących niż uczących wymagają odniesienia do podobnych eksperymentów, realizowanych według klasycznego postępowania.

Aktualny komentarz: W poprzedniej wersji swojej pracy Doktorant pisał, że sieci neuronowe są dominującą metodą klasyfikacji, wykorzystywaną do rozpoznawania mowy. Jednocześnie, wielokrotnie podkreślał, że w odróżnieniu od innych badaczy, którzy więcej danych używają do uczenia a znacznie mniej do testowania, On w swojej pracy postępuje odwrotnie. Brakowało mi i nadal brakuje wyraźnego uzasadnienia dla takiego podejścia. Jednakże rozszerzona wersja rozprawy pokazuje nieco szerszy kontekst zagadnienia i pozwala zrozumieć dlaczego w przypadku stosowanych przez Doktoranta metod można tak postępować i daje to dobre wyniki.

Z drugiej strony, Doktorant wskazuje podejścia innych autorów w przypadku ograniczonej liczby danych. Należą do nich np. „transfer learning", „data augmentation", czy „domain expertize". Nie stosuje żadnego z nich w swojej pracy, ale wspomina o zamiarze wykorzystania metody powiększania zbioru danych („data augmentation"). Nasuwa się więc pytanie: Skoro Doktorant widzi taką potrzebę, by w przyszłości zastosować to podejście, to czy z tego nie wynika słabe uzasadnienie dla eksperymentów z małą liczbą danych uczących w stosunku do testujących?

Istotną kwestią w tym temacie jest też fakt, że Doktorant – rozważając i stosując sieci neuronowe – pomija problem podziału danych na uczące i testujące w odniesieniu do różnych typów sieci neuronowych. Warto porównać z tego punktu widzenia probabilistyczne sieci neuronowe, sieci radialne, a także najbardziej popularne sieci typu MLP z uczeniem metodą wstecznej propagacji błędów.

W rozdziale 1.2 Doktorant, odnosząc się ogólnie do uczenia maszynowego (*machine learning*), stwierdza że w praktyce wykorzystuje się większość posiadanych danych do uczenia i ograniczoną ich część do testowania. To jest oczywiście prawdą. Autor zdaje sobie sprawę z problemu ograniczonego zbioru danych, pisząc o potrzebie skalowania i użyteczności algorytmów w rzeczywistych zastosowaniach. Doktorant uważa jednak, że jest możliwe rozpoznawanie osób na podstawie ich głosu pomimo wykorzystania niewielu próbek sygnałów mowy. Oznacza to, że można zbudować system rozpoznawania - na podstawie ograniczonego zbioru próbek głosu – i testować go na większym zbiorze danych, uzyskując zadowalające rezultaty dokładności. Ma to oczywiście sens. Z punktu widzenia sztucznej inteligencji warto zastanowić się, czy do rozpoznawania głosu mówcy istotnie potrzebujemy ogromnego zbioru danych (pochodzących od różnych osób), skoro małe dziecko bez tak wielu przykładów rozpoznaje głos swojej mamy.

Na uwagę zasługuje fakt, że Doktorant w nowym wydaniu swojej rozprawy doktorskiej, w rozdziale 1.2 (*Motivation and Problem Definition*), analizuje kwestię ograniczonego zbioru danych uczących, rozważając kilka związanych z tym podproblemów, formułując odpowiednie pytania w ramach postawienia problemu. Pierwsze z tych pytań dotyczy właśnie kwestii, czy mały zbiór danych sygnałów mowy może być wykorzystany do uczenia modelu, który pozwoli na rozwiązanie problemu rozpoznawania mówcy w większej skali. Drugie z tych pytań, niezwykle istotne, to – jaki wektor cech najlepiej reprezentuje sygnały mowy w zadaniach klasyfikacji. Trzecie pytanie – czy można efektywnie zredukować kroki wstępnego przetwarzania? Doktorant postawił 5 tego typu pytań, traktując je jako przedmiot swoich badań. W tym kontekście Autor sformułował szczegółowe cele swojej dysertacji (w rozdziale 1.3.1).

6. INNE UWAGI DO AKTUALNEJ WERSJI PRACY DOKTORSKIEJ

Nie jest prawdą, co pisze Doktorant, że w sieci neuronowej RBF liczba neuronów radialnych jest równa liczbie przykładów (danych) z ciągu uczącego. Tak jest w probabilistycznych sieciach neuronowych ale nie w sieciach RBF, gdzie tych neuronów jest zwykle znacznie mniej. Oczywiście zastosowanie sieci RBF z tak dużą liczbą neuronów też może dać właściwy rezultat (w przypadku małej liczby danych) ale wiąże się to z niepotrzebnie dużym nakładem obliczeniowym i traci się zdolność generalizacji.

7. ORYGINALNE REZULTATY

Doktorant w ostatnim rozdziale monografii, podsumowującym rezultaty pracy, podkreśla najważniejsze punkty swojej rozprawy doktorskiej. Dotyczą one przede wszystkim uproszczenia preprocesingu, skalowalności oraz odpowiednich proporcji danych uczących do testujących. Są to aspekty widoczne w prezentacji całej dysertacji.

Warto zwrócić uwagę na fakt, że Doktorant może się pochwalić listą własnych publikacji, zamieszczonych w Dodatku B. Wykaz ten zawiera 9 publikacji współautorskich Doktoranta z Promotorem. Należy podkreślić, że w 5-ciu z tych artykułów Pan M. K. Nammous występuje jako pierwszy autor, zwłaszcza w najnowszych. Natomiast w 2007 r. ukazała się Jego współautorska publikacja, p.t. „*A speech and speaker identification system: Feature extraction, description, and classification of speech-signal image*", w czasopiśmie „*IEEE Transactions on Industrial Electronics*". Trzeba przyznać, że nie każdy doktorant posiada publikację tak wysokiej rangi w swoim dorobku naukowym. Poza tym istotny jest też aspekt aplikacyjny prezentowanej działalności naukowej. Należy również odnotować fakt przyjęcia niedawno do druku artykułu Doktoranta w czasopiśmie „*Journal of King Saud University – Computer and Information Sciences*", w roku 2020.

Wymienione publikacje Doktoranta związane są tematycznie z Jego rozprawą doktorską. Tytuł tej pracy - „Nowe podejścia w rozpoznawaniu mowy od pojedynczych

słów do praktycznych rozwiązań" - mógłby być wspólnym tytułem dla wszystkich Jego publikacji, gdyby potraktować je jako spójny tematycznie zbiór prac. Taki cykl publikacji, mimo że są to prace współautorskie, można byłoby przyjąć jako podstawę do uzyskania stopnia doktora. Warto dodać, że obecnie coraz rzadziej wymaga się publikacji indywidualnych. Artykuły współautorskie świadczą o umiejętności współpracy naukowej, co w dzisiejszych czasach, gdy istnieje konieczność pracy zespołowej, szczególnie w informatyce, jest niezwykle istotne.

Autor recenzowanej rozprawy doktorskiej postawił sobie za cel rozwinięcie podejścia automatycznej identyfikacji mówcy (osoby) poprzez rozpoznawanie mowy (głosu), na podstawie próbek pochodzących z dużych baz danych, przy wykorzystaniu możliwie najmniejszej ilości informacji. Główny nacisk został tu położony na te dwa, zasadnicze zdaniem Doktoranta, elementy – czyli: 1) duży zbiór danych (znacznie większy niż stosowany w Jego wcześniejszych badaniach, 2) wykorzystanie jak najmniejszej ilości tych danych. Realizacja tego celu stanowi rozwinięcie podejścia Doktoranta, będącego przedmiotem cyklu Jego publikacji. Autor wykazał się znajomością wiedzy teoretycznej w reprezentowanej dyscyplinie oraz umiejętnością samodzielnego rozwiązywania problemów naukowo-badawczych.

Doktorant opracował system identyfikacji osób na podstawie głosu. Udoskonalał ten system, wzbogacając go o nowe elementy, poprawiając skuteczność działania i rozszerzając jego funkcjonalność. Wymagało to zarówno własnego wkładu naukowego, jak też nowych pomysłów w przeprowadzaniu eksperymentów, a także wykorzystania umiejętności programistycznych. Zdobył odpowiednie doświadczenie naukowo-badawcze, by kontynuować rozpoczętą działalność naukową i w praktyce stosować swoje rozwiązania

Mgr inż. Mohammad K. Nammous posiada znaczący dorobek publikacyjny i jest autorem rozprawy doktorskiej, wydanej w formie monografii Politechniki Warszawskiej, jako „Ph.D. Thesis".

8. UWAGI KOŃCOWE

Zgodnie z art. 13 ust. 1 ustawy z dnia 14.03.2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki: „Rozprawa doktorska, przygotowywana pod opieką promotora, powinna stanowić oryginalne rozwiązanie problemu naukowego lub artystycznego oraz wykazywać ogólną wiedzę teoretyczną kandydata w danej dyscyplinie naukowej lub artystycznej, a także umiejętność samodzielnego prowadzenia pracy naukowej lub artystycznej." Te, podkreślone tu, najistotniejsze elementy pozostają w tym samym brzmieniu po dokonaniu późniejszych zmian w cytowanej ustawie.

Z pełnym przekonaniem mogę stwierdzić, że **Doktorant wykazuje ogólną wiedzę teoretyczną w dyscyplinie naukowej, związanej z przedłożoną do recenzji rozprawą doktorską.** Ponadto z pewnością **posiada umiejętność samodzielnego prowadzenia pracy naukowej.** Czy recenzowana rozprawa doktorska stanowi oryginalne rozwiązanie problemu naukowego?

Uważam, że wyraźnie widoczne są aspekty nowatorskiego, oryginalnego podejścia Doktoranta do rozwiązywanego problemu, co zostało zauważone w Jego publikacjach. Moim zdaniem spełnia wymagania stawiane rozprawom doktorskim.

9. KONKLUZJA

Uwzględniając wszystkie wymienione w recenzji uwagi redakcyjne, a przede wszystkim merytoryczne, wyrażam swoją opinię w formie następującej konkluzji:

Przedłożona do recenzji rozprawa doktorska p.t. „New approaches in speech recognition from isolated words to practical solutions", której autorem jest mgr inż. Mohammad Kheir Nammous, spełnia wymagania stawiane przez odpowiednią ustawę, co pozwala mi wnioskować o dopuszczenie jej do dalszych etapów przewodu doktorskiego i obrony tej pracy przez Doktoranta. Zatem podpisuję się pod takim wnioskiem.