Politechnika Warszawska
Wydział Elektroniki i Technik Informacyjnych

dr hab. inż. Artur Janicki, prof. uczelni
Instytut Telekomunikacji
Warszawa, 10.10.2024

RECENZJA ROZPRAWY DOKTORSKIEJ

mgr. **MICHAŁA JUNCZYKA**
pt. „*Application of speech datasets management methods for the evaluation of Automatic Speech Recognition systems for Polish*", przedłożonej Radzie Naukowej Informatyki Uniwersytetu im. Adama Mickiewicza w Poznaniu

PRZEDMIOT ROZPRAWY, GŁÓWNE CELE PRACY

Praca Pana **mgr. Michała Junczyka** pt. „*Application of speech datasets management methods for the evaluation of Automatic Speech Recognition systems for Polish*" dotyczy zagadnienia przygotowania głosowych baz danych (korpusów mowy) do skutecznej ewaluacji systemów automatycznego rozpoznawania mowy (ARM, ang. *automatic speech recognition*, ASR) dla języka polskiego oraz przygotowania systemu do oceny porównawczej (ang. *benchmark*) systemów ASR dla języka polskiego.

Przedstawiona rozprawa jest pracą badawczą z bardzo wyraźnym komponentem praktycznym. W swojej rozprawie Autor za cel nadrzędny stawia zaprojektowanie i implementację systemu/struktury (ang. *framework*), który umożliwi wykorzystanie istniejących głosowych baz danych (korpusów) w celu ewaluacji systemu ASR. Doktorant formułuje przy tym następującą tezę badawczą:

- *Utworzenie rozbudowanej struktury zarządzania danymi pozwoli na rzetelną i obiektywną ocenę systemów ASR dostępnych dla języka polskiego.*

Następnie wymienia sześć celów badawczych, do których przypisane są łącznie 23 szczegółowe pytania badawcze. Te sześć celów badawczych (ang. *research objectives*, RO) Doktorant w swojej rozprawie sformułował następująco:

- RO1: Przegląd korpusów mowy ASR dla języka polskiego
- RO2: Zaprojektowanie i zarządzanie korpusem mowy dla języka polskiego
- RO3: Przegląd ocen porównawczych (benchmarków) ASR dla języka polskiego
- RO4: Projekt i wdrożenie systemu do oceny porównawczej (benchmarku) systemów ASR
- RO5: Wykorzystanie opracowanego zbioru danych do oceny porównawczej systemów ASR dla języka polskiego
- RO6: Organizacja otwartego konkursu dla społeczności ASR

Nowowiejska 15/19,
00-665 Warszawa
tel.: 22 234 77 22
e-mail:
Artur.Janicki
@pw.edu.pl

Doktorant systematycznie przywołuje ww. przedstawione cele badawcze w poszczególnych rozdziałach rozprawy. Rozprawa została napisana w języku angielskim, składa się z 6 rozdziałów oraz 1 załącznika, zajmuje łącznie aż 251 stron, nie licząc spisu treści, spisu rysunków i tabel.

Rozdział 1. wprowadza w tematykę rozprawy, przedstawia hipotezę, cele i pytania badawcze.

Rozdział 2. zawiera przegląd literatury związanej z tematyką rozprawy.

Rozdział 3. prezentuje zaproponowaną metodologię badawczą z podziałem dla poszczególnych celów badawczych.

Rozdział 4. przedstawia wyniki prac, prowadzących do osiągnięcia ww. celów badawczych.

Rozdział 5. zawiera dyskusję otrzymanych wyników.

Rozdział 6. prezentuje wnioski z całości pracy, podsumowuje osiągnięcia i wskazuje potencjalne dalsze kierunki prac.

W załączniku Autor umieścił miedzy innymi szczegóły ankiet dot. analizowanych korpusów oraz benchmarków, a także szczegółowe dane statystyczne poszczególnych podzbiorów danych.

MOCNE STRONY ROZPRAWY

Za najmocniejszą stronę rozprawy uważam kompleksowe podejście Doktoranta do tematu rozprawy, tj. przygotowania korpusu mowy do rzetelnej ewaluacji systemów automatycznego rozpoznawania mowy dla języka polskiego. Doktorant w sposób niezwykle systematyczny prowadzi Czytelnika przez cały proces badawczy. Rozpoczyna od szczegółowego przeglądu literatury, analizy stanu wiedzy w tej dziedzinie i identyfikacji istniejących braków. Następnie dla każdego celu badawczego umiejętnie proponuje odpowiednią metodologię. W kolejnym kroku prezentuje wyniki swoich prac, również w rozbiciu na poszczególne cele badawcze. Następnie dla każdego celu badawczego przeprowadza dyskusję osiągniętych rezultatów, by na końcu wszystko podsumować, wykazując poprawność hipotezy postawionej na wstępie.

Za najważniejsze osiągnięcie Doktoranta uważam przygotowanie korpusów mowy BIGOS oraz PELCRA for BIGOS służących do ewaluacji systemów ASR dla języka polskiego, z wykorzystaniem 24 otwartych korpusów audio, takich jak Mozilla Common Voice, Multilingual LibriSpeech, Clarin Studio, Clarin Mobile, Google FLEURS, SpokesMix czy PolEval 22 Diabiz. Korpusy te zawierają zarówno mowę czytaną jak i spontaniczną o

Nowowiejska 15/19,
00-665 Warszawa
tel.: 22 234 77 22
e-mail:
Artur.Janicki
@pw.edu.pl

różnej jakości. Doktorant przeprowadził szereg operacji, takich jak selekcja materiału, unifikacja formatu plików, normalizacja głośności, ujednolicenie formatu transkrypcji, podział na zbiory train/dev/test itd. Wynikowe zbiory zawierają prawie 400 tys. nagrań wraz z transkrypcjami o łącznej długości sygnału mowy ponad 800 godzin. Zasoby te, wraz ze skryptami, są dostępne dla wszystkich chętnych, mogą więc stanowić ważne narzędzie dla badaczy zajmujących się systemami ASR dla języka polskiego, a także dla specjalistów z przemysłu.

Do znaczących osiągnięć Doktoranta należy zaprojektowanie i wdrożenie systemu do oceny porównawczej (benchmarku) systemów ASR. Proces oceny obejmuje m.in. inicjalizację poszczególnych systemów ASR wraz z odpowiednią konfiguracją, generację hipotez ASR, obliczanie wartości metryk, analizę oraz wizualizację wyników.

Przy pomocy opracowanego systemu Doktorant przeprowadził porównanie siedmiu różnych systemów ASR (Google STT, Azure STT, Whisper, AssemblyAI, NeMo, MMS i Wav2Vec) pracujących łącznie z 25 modelami. Autor porównał te systemy, wyznaczając wartości odpowiednich metryk: zdaniowej, wyrazowej oraz znakowej stopy błędów (ang. SER, WER, CER, dodatkowo także *Match Error Rate* – MER). Przeprowadził szereg analiz, porównując np. dokładność rozpoznawania otwartych i komercyjnych systemów ASR czy dokładność rozpoznawania dla mowy męskiej lub żeńskiej. Na mój stan wiedzy jest to aktualnie najbardziej kompleksowa ocena porównawcza współczesnych systemów ASR dla języka polskiego.

Na uznanie zasługuje również przygotowanie otwartego konkursu (ang. *challenge*) dla społeczności ASR dla języka polskiego. Ponieważ konkurs był jeszcze otwarty w czasie składania rozprawy, brakuje informacji nt. wyników tego konkursu.

Potwierdzeniem osiągnięć Doktoranta są publikacje, co warto podkreślić, samodzielne: 1 publikacja w czasopiśmie Poznan Studies in Contemporary Linguistics (70 pkt. MNiSW) zawierająca przegląd korpusów ASR dla języka polskiego oraz 1 publikacja na międzynarodowej konferencji FedCSIS 2023, opisująca korpus BIGOS. Oprócz tego, Doktorant przygotował także samodzielną publikację opisującą studium przypadku: cały system (framework) do analizy porównawczej systemów ASR na przykładzie języka polskiego. Publikacja jest od lipca 2024 umieszczona w repozytorium arXiv i, jak się domyślam, jest obecnie w procesie recenzyjnym jakiegoś czasopisma bądź konferencji (nie znalazłem o tym informacji). Oprócz tego, Doktorant jest też współautorem publikacji na renomowanej konferencji LREC 2020.

Uważam, że rozprawa doktorska bez wątpienia prezentuje oryginalne rozwiązania w zakresie zarządzania korpusami mowy służącymi do ewaluacji systemów ASR. Jest napisana w sposób bardzo systematyczny i przystępny dla Czytelnika. Układ pracy jest bardzo czytelny, także sposób prezentacji wyników jest bardzo klarowny. Doktorant wykazał się umiejętnością samodzielnego prowadzenia pracy naukowej, sprawnym posługiwaniem warsztatem badawczym, wykazał też biegłą znajomość teoretyczną z dziedziny informatyki. Zaproponowana metodologia zarządzania korpusami do

Nowowiejska 15/19,
00-665 Warszawa
tel.: 22 234 77 22
e-mail:
Artur.Janicki
@pw.edu.pl

ewaluacji systemów ASR w rozprawie została zaprezentowana dla języka polskiego, ale z powodzeniem może zostać wykorzystana dla innych języków, czemu mogą sprzyjać też udostępnione skrypty.

SŁABE STRONY ROZPRAWY

Słabych stron rozprawy nie znajduję wiele, poza drobnymi uchybieniami opisanymi poniżej. Rozprawę czyta się na ogół bardzo dobrze, choć czasem przeszkadza nagromadzenie nagłówków, list numerowanych i punktowanych – odniosłem wrażenie, że Doktorant nie przepada za jednolitym, ciągłym tekstem.

Szkoda, że Doktorant w podsumowaniu rozprawy nie wraca do postawionej na wstępie hipotezy badawczej. Powinno się ją ponownie przywołać i stwierdzić, czy została ona udowodniona lub przynajmniej przeprowadzić dyskusję na ten temat. Ciekaw jestem opinii Doktoranta w tej kwestii.

Mam również pewne wątpliwości dot. kolejności rozpatrywanych zagadnień (patrz także Rys. 3.1): dlaczego Doktorant najpierw zajmował się celem badawczym RO2: „Zaprojektowanie i zarządzanie korpusem mowy dla języka polskiego”, a dopiero potem celem badawczym RO3 „Przegląd ocen porównawczych (benchmarków) ASR dla języka polskiego"? Czy z przeglądu benchmarków ASR nie wypłynęłyby wnioski, które byłyby przydatne przy tworzeniu własnego korpusu?

UWAGI SZCZEGÓŁOWE

- Temat oceny porównawczej systemów uczenia maszynowego (rozdział 2.2) wydaje się mi być tematem pobocznym względem ewaluacji systemów ASR. Nie rozumiem, dlaczego został mieszczony w rozprawie. W dodatku uczenie maszynowe to zagadnienie bardzo szerokie (obejmuje np. zagadnienia klasyfikacji obrazów, modele regresji dla różnych typów danych itd.), a w rozprawie dziedzina ta została zawężona do wybranych zagadnień przetwarzania języka naturalnego (NLP).
- W Załączniku, w Tab. 7.1 brakuje niektórych danych.
- Jak wyliczane były przedziały ufności wyrazowej stopy błędów (WER) na Rys. 4.14 – 4.17?
- Autor pisze, że ww. wykresy nie uwzględniają wartości odstających (*outliers*). Jak te wartości odstające były identyfikowane?
- Nie znalazłem wykresu pokazującego rozkład długości nagrań w wynikowych korpusach – myślę, że niosłoby to ciekawe informacje.
- Tab. 4.63: skąd wzięła się różnica -0.92 p.p. mediany WER dla głosów męskich?
- Dlaczego ASR assembly_nano nie radzi sobie zupełnie z głosami męskimi z korpusu BIGOS, a dla głosów żeńskich wychodzi znacznie lepiej, podczas gdy dla nagrań z bazy PELCRA sytuacja jest odwrotna?

Nowowiejska 15/19,
00-665 Warszawa
tel.: 22 234 77 22
e-mail:
Artur.Janicki
@pw.edu.pl

- Skąd taki duży spadek średniej wartości wyrazowej stopy błędów WER dla grupy wiekowej 60+? Czy została tu przeprowadzona analiza błędu pomiaru?
- Zastanawia mnie listing wyjściowego pliku `out.tsv` (str. 226) z przykładową hipotezą ASR podaną w ogłoszeniu o konkursie „2024 Polish ASR challenge": pojawia się tam słowo „hyacynt" zapisane według dawnych zasad ortograficznych. Czy transkrypcja zbiorów audio była sprawdzana pod kątem poprawności ortograficznej, zgodnej ze współczesnymi zasadami?

STRONA EDYCYJNA PRACY

Rozprawa doktorska **mgr. Michała Junczyka** jest napisana bardzo starannie. Doktorant sprawnie posługuje się językiem angielskim i pisze stylem naukowym o wysokiej jakości. Tym niemniej zauważam pewne niedociągnięcia natury językowej i typograficznej, m.in.:

- Skrót „*p.p.*" rozumiem jako punkty procentowe i jest to czytelne dla osób znających polską notację. Tymczasem dla odbiorcy zagranicznego określenie „*percentage points*" często jest niezrozumiałe. Stosuje się wtedy raczej „*% relative*".
- W wersji drukowanej rozprawy brakuje adresów różnych zasobów, np. PolEval 2024 dates, AMU ASR Leaderboard. W wersji PDF są tam hiperłącza, więc zasoby daje się zlokalizować i łatwo do nich przejść. Warto jednak byłoby pamiętać także o czytelnikach papierowej wersji rozprawy.
- „*Appendix*" powinien być nienumerowanym rozdziałem na końcu pracy, nie zaś Rozdziałem 7.
- Niekonsekwentne umieszczanie tytułów pod lub nad tabelami (np. str. 102-103).
- Przypis dolny 3) na stronie xx jest pusty, jego treść chyba wskoczyła do głównego tekstu.
- Często brakuje spacji między tekstem a odsyłaczem do pozycji literatury (np. str. 2, 3, 5 itd).
- Niektóre akronimy są niepotrzebnie wyjaśnione wielokrotnie (np. WER, CER, WIL).
- Niespójna pisownia małych i dużych liter (np. *word error rate, machine learning*).
- Liczby poniżej 10 występujące w tekście powinny być pisane słownie, np. „*3 commercial ASR systems*" => „*three commercial ASR systems*" (str. 9)
- Czasem występują prawe znaki cudzysłowu zamiast lewych (np. str. 18: '*dog vs. cat classification*' zamiast '*dog vs. cat classification*').
- Odsyłacz [53] kieruje do repozytorium GitHub, a nie do czasopisma, jak napisano na str. 60.
- Tytuły tabel 4.5-4.7 są zbyt skrótowe. Poza tym „*recording devices*" byłoby bardziej precyzyjne niż „*audio devices*".
- Fig. 4.4-4.12 wg mnie zawierają tabele.
- Podpisy Fig. 4.28 i 4.30 nie zmieściły się na stronie.

Nowowiejska 15/19,
00-665 Warszawa
tel.: 22 234 77 22
e-mail:
Artur.Janicki
@pw.edu.pl

Politechnika Warszawska
Wydział Elektroniki i Technik Informacyjnych

- Etykiety powinny być w jednym wierszu razem z numerem (np. Appendix~7.1.5 na str.75).
- Brak numeracji wzorów matematycznych (np. str. 33, 35).
- Występują nieliczne błędy literowe i interpunkcyjne (np. str. 34: *pass* => *past*, str. 125: *et. al* => *et al.*).
- Błędne użycie podwójnego minusa we wzorach na dole str. 34.
- Nierówne złączenia linii na Rys. 2.4.
- Zbędne przedimki „*the*”, np. „*the Appendix 7.1.1*” => „*Appendix 7.1.1*”.
- Niespójna czcionka w przypisach dolnych na str. 59.
- Przypisy dolne pojawiające się w środku strony (str. 125).
- Nie wiem, czy jest sens umieszczania w tabeli kolumny, której wartość jest w każdym wierszu stała (np. Tab. 4.42, Tab. 4.43).
- Zmienna precyzja wartości w Tab. 4.46 (np. 7.5 vs. 7.02).
- Pozycja literatury [135] zawiera 450 autorów – może nie trzeba było drukować wszystkich nazwisk na trzech stronach? ;) Np. ArXiv pokazuje „tylko” 50, a następnie dodaje „et al.” :)

Wyżej wymienione uchybienia edycyjne w żadnym stopniu nie umniejszają jednak osiągnieć Doktoranta.

WNIOSKI KOŃCOWE

W podsumowaniu stwierdzam, że cele badawcze postawione w rozprawie **mgr. Michała Junczyka** zostały osiągnięte. Doktorant szczegółowo opracował zagadnienie zarządzania korpusami mowy dla celów ewaluacji systemów ASR, stworzył takie korpusy, zaimplementował system do oceny porównawczej systemów ASR, a następnie przeprowadził szczegółową ocenę porównawczą dostępnych systemów ASR dla języka polskiego.

Stwierdzam, że przedstawiona rozprawa doktorska spełnia warunki określone w art. 187 ustawy z dn. 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz.U. z 2022 r. poz. 574 z późn. zm.), dlatego niniejszym wnioskuję o **dopuszczenie** Doktoranta, Pana **mgr. Michała Junczyka**, do publicznej obrony jego rozprawy doktorskiej. Ze względu na wyjątkowo staranne systematyczne podejście Doktoranta do rozwiązania problemu badawczego, a także wysoką jakość samej rozprawy składam również wniosek o **wyróżnienie rozprawy**.

Nowowiejska 15/19,
00-665 Warszawa
tel.: 22 234 77 22
e-mail:
Artur.Janicki
@pw.edu.pl

dr hab. inż. Artur Janicki, prof. uczelni
Politechnika Warszawska

PODPIS ZAUFANY
ARTUR
JANICKI
30.10.2024 11:06:38 [GMT+1]
Dokument podpisany elektronicznie
podpisem zaufanym