



Politechnika Warszawska

Wydział Elektroniki i Technik Informatycznych

dr hab. inż. Artur Janicki, prof. uczelni
Instytut Telekomunikacji
Warszawa, 12.01.2024

RECENZJA ROZPRAWY DOKTORSKIEJ

mgr. **TOMASZA ZIĘTKIEWICZA**

pt. „Zaprojektowanie oraz implementacja systemu automatycznej korekty błędów i normalizacji wyjścia z systemu rozpoznawania mowy”, przedłożonej Radzie Naukowej Informatyki Uniwersytetu im. Adama Mickiewicza w Poznaniu

PRZEDMIOT ROZPRAWY, GŁÓWNE TEZY PRACY

Praca Pana **mgr. Tomasza Ziętkiewicza** pt. „Zaprojektowanie oraz implementacja systemu automatycznej korekty błędów i normalizacji wyjścia z systemu rozpoznawania mowy” dotyczy usprawniania pracy systemów rozpoznawania mowy (ang. *automatic speech recognition*, ASR) poprzez opracowanie i implementację dwóch metod przetwarzania końcowego (ang. *post-processing*): korygującej wyjście modułu ASR, a także przywracającej znaki interpunkcyjne.

Przedstawiona rozprawa jest pracą badawczą z wyraźnym komponentem wdrożeniowym. W swojej rozprawie Autor nie formułuje konkretnych tez, natomiast realizuje dwa cele, sformułowane następująco:

- Opracowanie metody korekty błędów systemu rozpoznawania mowy;
- Opracowanie metod normalizacji wyjścia z systemu rozpoznawania mowy, w szczególności przywracania znaków interpunkcyjnych.

Dodatkowo, oba cele miały spełniać wymogi stawiane przez wdrożeniowy kontekst badań (rozdział 2), to jest działać w sposób efektywny i umożliwiający precyzyjną kontrolę nad ich działaniem.

Rozprawa została napisana w języku polskim, składa się z 7 rozdziałów i zajmuje 119 stron.

Rozdział 1. wprowadza w tematykę rozprawy i przedstawia motywację prowadzonych badań.

W rozdziale 2. Doktorant prezentuje uwarunkowania badań wynikające z faktu ich wdrażania w firmie Samsung.

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



Politechnika Warszawska

Wydział Elektroniki i Technik Informatycznych

W rozdziale 3. Autor wprowadza w tematykę uczenia maszynowego i metod ewaluacji stosowanych w zadaniach klasyfikacji, w szczególności w zadaniach związanych z algorytmami ASR.

Rozdział 4. dotyczy metod automatycznej korekty błędów rozpoznawania mowy. W tym rozdziale Doktorant opisuje zaproponowaną metodę „Otaguj i popraw”, a także jej ewaluację dla wielu języków.

W rozdziale 5. Autor proponuje metodę rekonstrukcji znaków interpunkcyjnych nazwaną „Otaguj i przywróć” i również przeprowadza jej ewaluację.

Rozdział 6. prezentuje zrealizowane wdrożenia opracowanych metod, w tym wpływ zmniejszenia błędów ASR na poprawę pracy algorytmów rozumienia języka naturalnego (ang. *natural language understanding*, NLU).

W rozdziale 7. Autor przedstawia skrótowo podsumowanie swojej pracy.

MOCNE STRONY ROZPRAWY

Mocną stroną rozprawy jest zaproponowanie i implementacja metody korekcji wyjścia modułu ASR „Otaguj i popraw” oraz metody rekonstrukcji interpunkcji „Otaguj i przywróć”, które wykazały się dużą skutecznością, a przy tym nie wymagały wielkich zbiorów uczących. Nie potrzebują one także dużej mocy obliczeniowej podczas inferencji, czyli właściwej pracy algorytmów w docelowym środowisku. Dodatkowo, przy pracy tych algorytmów nie ma ryzyka halucynacji, tak jak przy innych metodach wykorzystujących wielkie modele językowe (LLM).

Dużą zaletą zaproponowanych i zaimplementowanych metod jest też właśnie ich relatywna prostota. Fakt, że są one bardzo intuicyjne (np. poprzez użycie czytelnych dla człowieka szablonów do edycji w metodzie „Otaguj i popraw”), umożliwia łatwą ich interpretację, a przez to łatwe zarządzanie nimi, ich modyfikację czy adaptację do nowego środowiska.

Bardzo mocną stroną rozprawy jest jej aspekt wdrożeniowy. Zaproponowane metody „Otaguj i popraw” oraz „Otaguj i przywróć”, po ich szczegółowym opracowaniu i przebadaniu, zostały następnie wdrożone w firmie Samsung. Zaowocowało to obniżeniem liczby błędów ASR o ponad 20%, co jest wynikiem bardzo znaczącym.

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

Na uznanie zasługuje też czytelnie zaprezentowana w Rozdziale 1.3, a następnie zastosowana w pracy metodologia przeprowadzonych badań, składająca się z 1) etapu pozyskiwania danych, 2) etapu stworzenia i wyuczenia modeli z wykorzystaniem tych danych, 3) etapu ewaluacji opracowanych modeli i algorytmów na wydzielonym zbiorze



Politechnika Warszawska

Wydział Elektroniki i Technik Informacyjnych

testowym. Bardzo słusznie Doktorant przykłada duże znaczenie do zastosowanych danych, wiedząc, że od nich zależy w dużej mierze powodzenie prac badawczych.

Potwierdzeniem znaczenia osiągnięć Doktoranta są publikacje, których jest on współautorem. Autor opublikował 8 publikacji na międzynarodowych konferencjach, w tym na konferencji EMNLP 2023 w Singapurze, zaliczanej do kategorii CORE A+ (200 pkt. MNiSW). Świadczy to o wysokim poziomie prezentowanych badań i ich wyników.

Doktorant przyczynił się także do rozwoju badań w dziedzinie przetwarzania języka naturalnego w polskim środowisku naukowym. Opracowane przez niego dane stały się podstawą organizacji otwartych wyzwań: wyzwania OCCESRS dotyczącego korekty błędów ASR oraz CAICCAIC, dotyczącego opracowywania modeli NLU odpornych na błędy.

Uważam, że rozprawa doktorska prezentuje oryginalne rozwiązanie w zakresie zastosowania wyników własnych badań naukowych w sferze gospodarczej (opracowane metody zostały praktycznie wykorzystane w firmie Samsung). Doktorant wykazał również sprawne posługiwanie się warsztatem badawczym oraz biegłą znajomość teoretyczną z dziedziny informatyki.

SŁABE STRONY ROZPRAWY

Do słabych stron rozprawy zaliczam brak porównania proponowanej metody „Otaguj i popraw” z innymi stosowanymi w tej dziedzinie metodami, opisanymi zresztą przez Doktoranta, takimi jak stosowanie modeli end-to-end: np. neuronowych sieci rekurencyjnych lub sieci neuronowych typu Transformer. Rozumiem, że nie byłoby to łatwe (Doktorant właściwie wskazuje wady innych podejść w Rozdziale 4.2.1), jednak brakuje nawet informacji, czy taka próba była w ogóle rozważana lub podjęta.

Niedociągnięciem tej rozprawy jest też fakt, że w czasie testów zaproponowanej metody rekonstrukcji interpunkcji Doktorant przez pomyłkę użył (przypuszczalnie na swoją niekorzyść) innego, mniejszego zbioru uczącego, niż pozostali uczestnicy wyzwania PolEval 2021 Zadanie 1. Ponieważ Doktorant zauważył swoją pomyłkę, mógł powtórzyć eksperymenty, i wtedy porównanie z wynikami innych uczestników wyzwania byłoby bardziej miarodajne. Także wyniki i wyciągnięte wnioski mogłyby się trochę różnić od przedstawionych w Rozdziałach 5.4.3 oraz 5.5.

Pewną słabością badań dotyczących przywracania interpunkcji jest wg mnie źródło pochodzenia danych. Pochodzą one z publikacji na WikiNews i WikiTalks, które następnie były korygowane przez „rodzimy użytkowników języka polskiego” (str. 71). Wg mnie szkoda, że nie wykorzystano tutaj tekstów literackich, które zazwyczaj przechodzą przez bardzo wnikliwą korektę językową i dlatego zawierają z dużym



Politechnika Warszawska

Wydział Elektroniki i Technik Informacyjnych

prawdopodobieństwem poprawną interpunkcją. Rozumiem oczywiście, że dobór źródeł danych należał do autorów wyzwania PolEval. Ciekawe wg mnie byłoby jednak wypróbowanie działania zaproponowanej metody na tekście literackim.

Inne, drobniejsze uwagi:

- W pracy występują pewne niedociągnięcia formy – o czym więcej w następnym rozdziale.
- Temat rozprawy, a dokładniej podkreślony fragment „*Zaprojektowanie oraz implementacja systemu automatycznej korekcji błędów i normalizacji wyjścia z systemu rozpoznawania mowy*” obiecuje więcej, niż rozprawa zawiera. W zakresie normalizacji tekstu rozprawa koncentruje się tylko (aż!) na zagadnieniu rekonstrukcji interpunkcji. Według mnie właśnie to hasło powinno się pojawić w temacie rozprawy zamiast dość szerokiego pojęcia normalizacji tekstu.
- Autor używa czasem określenia „normalizacja tekstu” w sytuacjach, gdy według mnie właściwsze byłoby wyrażenie „odwrotna normalizacja tekstu”, np. str. 8: „*dostosowują format danych wyjściowych do formatu oczekiwanego na wyjściu całego systemu (dokonują jego normalizacji)*”.
- Str. 24: „*Pierre Curie wykryła dwa pierwiastki promieniotwórcze: polon i rad.*” – czy chodziło o jego żonę, Marię? Choć podobno odkryli je wspólnie...
- Str. 45: czy zbiór zaproponowanych klas operacji edycyjnych dla celów korekty rozpoznawania, umieszczony w Tabeli 4.1, wystarczy do dowolnego rodzaju korekty? Co np. w wypadku, gdy w słowie są dwa identyczne podciągi, a my chcemy zamienić tylko jeden z nich? Jak zachowa się wówczas klasa „`sreplace_{s}_{r}`”?
- Str. 45: jaka jest celowość używania wg mnie nadmiarowych klas operacji edycyjnych „`split_on_first_{c}`” oraz „`split_on_last_{c}`”, skoro te same operacje można by wykonać z wykorzystaniem klasy „`split_after_{n}`”?
- Str. 53: nie wyjaśniono, co oznacza „WER wyroczeni”.
- Str. 57: w Tabeli 4.5 stosuje się naprzemiennie metryki WRR oraz WER. Wg mnie czytelniej byłoby trzymać się jednej z ww. metryk.

STRONA EDYCYJNA PRACY

Rozprawa doktorska **mgr. Tomasza Ziętkiewicza** jest napisana starannie, jej format jest zgodny z ogólnie przyjętymi standardami dla prac naukowych w dziedzinie nauk technicznych. Układ pracy jest logiczny, bardzo przejrzysty i przyjazny dla czytelnika. Rysunki i tabele są staranne.

W pracy występują jednak pewne niedociągnięcia natury językowej i typograficznej. Są to m.in.:

- Liczne braki w interpunkcji lub nadmiarowe przecinki.



Politechnika Warszawska

Wydział Elektroniki i Technik Informatycznych

- W języku polskim przecinek jest znakiem dziesiętnym, tymczasem Doktorant używa kropki, jak w notacji anglosaskiej (na szczęście robi to konsekwentnie).
- Błędy językowe, np. „*Trenowanie i adaptacja modeli ASR jest kosztowna obliczeniowo*” (str. 35).
- Słowo „*studio*” współcześnie się odmienia (str. 51).
- Czasami styl zbyt potoczny, np. „*różnice (...) nie podpadają pod schemat*” (str. 46).
- Błędy literowe, np. „*spejcalnym*” (str. 26), „*precision*” (str. 48), „*asystema*” (str. 51), „*różnić*” zamiast „*różnic*” (str. 54).
- Prawidłowy zapis nazw konkursów to PolEval oraz SemEval.
- Nieużywanie polskich odpowiedników, które są już od lat stosowane w języku polskim, np. wyrazowa stopa błędów (odpowiednik ang. WER), zdaniowa stopa błędów (odpowiednik ang. SER).
- „*Recall*” jest częściej tłumaczony jako „*czułość*”.
- „*Named entity*” jest niespójnie tłumaczone jako „*encja nazwana*” lub „*jednostka nazwana*”.
- Częste stosowanie krótszych łączników „-” zamiast dłuższych myślników „-”.
- Niespójnie stosowanie kursywy (np. przy CRF).
- Niespójne używanie znaku %, np. w Tabeli 4.5.
- Ponowne wyjaśnianie akronimów wyjaśnionych wcześniej (np. WER).
- Niespójne podawanie wartości WER, np. 78 WER (str. 56), WER 24.7 (str. 58).
- Notoryczne stosowanie prawego cudzysłowu zamiast lewego (np. „słowo” zamiast „słowo” lub „słowo”).
- Zagnieżdżanie nawiasów w tekście.
- W przykładzie operacji „*point*” („*wstaw kropkę*”) jest pokazane wstawienie przecinka, a nie kropki (str. 74).
- Etykiety obiektów na niektórych rysunkach (np. Rysunek 4.4) nachodzą na kontury obiektów.

Uchybienia edycyjne nie umniejszają jednak w żadnym stopniu dorobku Doktoranta.

WNIOSKI KOŃCOWE

W podsumowaniu stwierdzam, że problemy badawcze postawione w rozprawie zostały opracowane i rozwiązane. Doktorant zaproponował rozwiązanie problemu automatycznej korekty wyjścia systemu rozpoznawania mowy, a także w znacznym stopniu problemu odtwarzania interpunkcji. Rozprawa doktorska mgr. Tomasza Ziętkiewicza prezentuje oryginalne rozwiązanie w zakresie zastosowania wyników własnych badań naukowych w sferze gospodarczej.

Stwierdzam, że przedstawiona rozprawa doktorska spełnia warunki określone w art. 187 ustawy Prawo o szkolnictwie wyższym i nauce (Dz.U. z 2022 r. poz. 574 z późn. zm.),



Politechnika Warszawska
Wydział Elektroniki i Technik Informatycznych

dlatego niniejszym wnioskuję o dopuszczenie Doktoranta, Pana **mgr. Tomasza Ziętkiewicza**, do publicznej obrony jego rozprawy doktorskiej.

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

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