Recenzja

Rozprawy doktorskiej mgr. Michała Junczyka

Pt „Application of speech datasets management methods for the evaluation of Automatic Speech Recognition systems for Polish”

1. Tematyka rozprawy

Automatyczne rozpoznawanie mowy (ASR) to technologia, która przekształca mowę mówioną w tekst. Proces automatycznego rozpoznawania mowy składa się z kilku kluczowych kroków: przetwarzania akustycznego sygnału mowy, ekstrakcji cech, modelowania fonemów i używania modeli językowych do przewidywania sekwencji słów. Systemy ASR napotykają wyzwania, takie jak radzenie sobie z różnicami w akcentach, szumem tła i homofonami. W przypadku języka polskiego ASR napotyka wyjątkowe wyzwania ze względu na specyficzne cechy języka polskiego, takie jak złożona gramatyka, natura fleksyjna i bogata morfologia.

Przedstawiona do oceny rozprawa doktorska skupia się na bardzo technicznych zagadnieniach związanych danymi do trenowania modeli ASR. W tym kontekście należy zauważyć, że istniejące publiczne zestawy danych dotyczących mowy są często niedostatecznie wykorzystywane z powodu problemów z ich wykrywalnością i interoperacyjnością, a ograniczony dostęp do danych ewaluacyjnych utrudnia weryfikację jakości systemów ASR. Przedstawiona praca doktorska jest poświęcona stworzeniu kompleksowego, dostępnego i rozszerzalnego zestawu zbiorów danych dla ASR, oraz opracowaniu procesu oceny jakości ASR.

2. Ocena treści rozprawy i wkładu oryginalnego

2.1. Treść rozprawy

Rozprawa została przygotowana w języku angielskim, składa się ona z 6 rozdziałów oraz Załącznika (ang. Appendix) i ma objętość 225 stron bez bibliografii.

Rozdział 1 zaczyna się z krótkiego wprowadzenia w dziedzinę ASR oraz opisu na czym polega specyficzność problemów automatycznego rozpoznawania mowy w języku polskim (polskie ASR). Następnie Doktorant przedstawił cele oraz motywacje prowadzonych badań. Głównym celem było zwiększenie użyteczności dostępnych zbiorów danych z nagraniami mowy oraz zaproponowanie uniwersalnej metody oceny jakości tych zbiorów.

Dalsza część Rozdziału 1 poświęcona jest wyjaśnieniom dotyczącym roli zbiorów danych w trenowaniu i ewaluacji systemów uczenia maszynowego w ogóle, oraz systemów automatycznego rozpoznawania mowy w szczególności. Omawiane są wyzwania stojące przed badaczami w tej dziedzinie. Przede wszystkim szczegółowo omówione zostały wyzwania związane z ewaluacją ASR. Są to przede wszystkim (i) brak tak zwanego „ground truth” oraz (ii) specyfika dziedziny (na przykład słownictwo, stosowane w różnych kontekstach). Biorąc pod uwagę, że praca doktorska jest poświęcona systemom rozpoznawania mowy w języku polskim,

przedstawiona została również sytuacja dotycząca zbiorów danych oraz ich ewaluacją, dla „polskiego” ASR.

Rozdział zamykają sekcje omawiające: (1) hipotezę badawczą, która (w tłumaczeniu na język polski) brzmi następująco:

Stworzenie rozbudowanego frameworku do zarządzania danymi, który pozwoli na rzetelną i obiektywną ocenę systemów ASR w języku polskim;

(2) sześciu celi badawczych:

- Przegląd zbiorów danych dla polskiego ASR
- Stworzenie i utrzymanie (zarządzanie) zbioru danych dla języka polskiego
- Stworzenie systemu do badań wydajnościowych (ang. benchmarking) systemów ASR
- Wykorzystanie przygotowanego zbioru danych do badań wydajnościowych systemów ASR dla języka polskiego
- Organizacja otwartych konkursów dla „wspólnoty” ASR

(3) pytań badawczych, związanych z każdym z wymienionych powyżej celów

(4) możliwe ograniczenia przeprowadzonych badań (specyficzność języka, problemy z doborem zbiorów danych, dostępność zasobów itp.)

(5) przyjętej metodologii badań, która właściwie powtarza wymianę postawionych celów badawczych, oraz

(6) podsumowanie dokonań.

Tak więc rozdział pierwszy jest wprowadzeniem, zawierającym podstawowe formalne definicje, konieczne dla dalszych badań.

Rozdział drugi jest poświęcony przeglądowi literatury, przedstawiającej aktualny stan wiedzy i badań w następujących dziedzinach:

- wyzwania w wyznaczaniu standardów (benchmarking) uczenia maszynowego oraz systemów automatycznego rozpoznawania mowy (ASR)

- wyzwania, metody i narzędzia do zarządzania zbiorami danych dla ASR

- zbiory ASR oraz tak standardy (benchmarki) dla języka polskiego.

Trzeba podkreślić, że wszystkie wymienione powyżej kroki zostały opisane bardzo szczegółowo, zaczynając z wyzwań technicznych, prezentacji metryk, stosowanych do mierzenia wydajności, możliwych scenariuszy takiego sprawdzania, kończąc wspomnieniem o ewentualnych problemach, które czekają na badaczy.

Na takim samym poziomie szczegółowości i dokładności zostały omówione, w dalszej części rozdziału, metody oraz narzędzia do zarządzania zbiorami danych dla ASR. Następnie, Doktorant przeszedł do omawiania zbiorów danych, przeznaczonych do benchmarkowania narzędzi do ASR w języku polskim oraz do ewaluacji ASR.

Trzeba tu jednak zauważyć, że przeglądając spis treści i później czytając pracę ma się wrażenie bardzo nietrafnego dobrania nazewnictwa rozdziałów/ podrozdziałów / … pracy. Na przykład Rozdział 3, który zatytułowany został: Methodology, ma w pierwszym podrozdziale sekcję „Research methodology", która zawiera podsekcję „Overview of methodology". Podobnie jest w innych podrozdziałach tego rozdziału. Powstaje pytanie, co dokładnie rozumie Doktorant pod pojęciem „methodology" oraz dlaczego metodologia jest rozważana „rekurencyjnie" z trzema poziomami zagłębienia (metodologia, metodologii, metodologii). Zwłaszcza, że rozdział jest tak naprawdę poświęcony tworzeniu frameworku do benchmarkowania systemów polskiego ASR.

Owszem, po wnikliwym przeczytaniu całej pracy i chwili zastanowienia, można domyślić się co autor miał na myśli, ale struktura sekcji i ich nazewnictwo stanowi bardzo poważny mankament pracy doktorskiej.

Ogólnie, trzeba zaznaczyć, że praca napisana w sposób bardzo „techniczny" i „kwestionariuszowy" (pytania-odpowiedzi). Tak więc Rozdział 3 zawiera wyłącznie informacje jak będą szukane odpowiedzi na postawione pytania badawcze związane z sześcioma celami badawczymi, a w następnym Rozdziale (Rozdział 4 „Results") zostały wylistowane odpowiedzi na pytania badawcze, natomiast dopiero w Rozdziale 5 „Discussion" można znaleźć omówienie powstałych zbiorów jak i niektórych wyników. Taka struktura pracy i zastosowany język, utrudniają czytanie pracy i docenienie wkładu Doktoranta.

Wracając więc do Rozdziału 3, warto zauważyć, że rozważania dotyczące designu frameworku są bardzo ciekawe i przedstawiają obraz bardzo szeroko zakrojonych badań dotyczących zbiorów danych umożliwiających ocenę wydajności systemów ASR dla języka polskiego. Została tutaj stworzona taksonomia takich zbiorów danych i ich cech. Ponadto sporą uwagę Doktorant poświęcił istniejącym narzędziom do zarządzania takimi zbiorami danych, w celu przygotowania tych zbiorów do sprawdzań wydajnościowych. Tu chciałabym zaznaczyć, że bardzo ciężko znaleźć w języku polskim odpowiednik dla angielskiego „curate". Tak więc będę tu i dalej używała najbliższego (w mojej opinii) tłumaczenia „przygotować".

Rozdział Czwarty, zatytułowany został Results. Zawiera on wyniki badań i eksperymentów, zaanonsowanych w Rozdziale 3. Po przeczytaniu tego rozdziału wrażenie sprawia ilość pracy wykonanej przez Doktoranta, wyniki której przedstawione zostały w wielu tabelach i na wielu wykresach. Szkoda, że zgodnie z planem Doktoranta, omawianie tych wyników ma miejsce dopiero w Rozdziale 5. Jeśli chodzi o zawartość, to rozdział ten przypomina kwestionariusz, czyli jest pisany w stylu pytanie odpowiedź. Co więcej, często odpowiedziami są pojedyncze zdania. Jeżeli chodzi o pytania to są to wciąż te same pytania badawcze, które zostały wymienione (po raz pierwszy) w Rozdziale 1 i powtarzane w kolejnych rozdziałach i podrozdziałach. Lekkie zdumienie

wywołują również sekcje, zawierające wyłącznie informację, gdzie ta lub inna treść, zbiór danych, lub framework do badań wydajnościowych, zostały udostępnione (dla potencjalnych użytkowników).

Rozdział 6, „Conclusions", zgodnie z „tradycją" ocenianej pracy doktorskie, ponownie zawiera pytania badawcze, ale odpowiedzi zawierają podsumowania osiągnięć i dokonań Doktoranta.

Kolejny, ostatni, Rozdział 7, „Appendix" zawiera listy cech zbiorów danych dla ASR oraz do benchmarkowania systemów ASR, informacje techniczne i dostępowe zbiorów danych publicznie dostępnych i dostępnych odpłatnie i inne informacje technicznego charakteru.

2.2. Wkład oryginalny

Najważniejsze samodzielne i oryginalne osiągnięcia Doktoranta to:

1. Gruntowna analiza zbiorów danych dotyczących mowy, skatalogowanie oraz ocena 53 zbiorów danych dla systemów automatycznego rozpoznawania mowy w języku polskim.
2. Przygotowanie zbioru danych benchmarkowych, który zawiera próbki audio z różnych źródeł mowy czytanej i spontanicznej. W tym kontekście przeprowadzona została analiza, standaryzacja oraz integracja źródeł danych, w celu zapewnienia spójności i niezawodności. Powstały zbiór danych został udostępniony do ogólnego używania i uzupełniania.
3. Zbadanie istniejących polskich standardów ASR, metod katalogowania zestawów danych i systemów używanych w poprzednich ocenach. Na podstawie wyników tych badań zidentyfikowanie kluczowych luk. W oparciu o tę wiedzę opracowanie bardziej kompleksowych ram porównawczych dla ASR.
4. Stworzenie frameworku, do badań wydajnościowych systemów ASR, który obsługuje różne zestawy danych, systemy i metryki, zapewniając ocenę ASR spójną ze standardowymi protokołami. Framework może być użyty również do wykonania testów porównawczych dla innych zestawów danych lub języków.
5. Korzystając z przygotowanego zestawu danych, porównanie 7 systemów ASR i 25 modeli, zarówno komercyjnych, jak i open-source. Odkrycie znaczących różnic między systemami, zestawami danych i danymi demograficznymi mówców. Zakwestionowanie wyższej wydajności systemów Azure i Google, przez lepsze wyniki nowszych systemów, takich jak Whisper i Assembly AI.

Warto podkreślić, że wszystkie te zbiory, analizy, katalogi itp. Zostały udostępnione dla społeczności badaczy ASR. Co więcej, w celu zaangażowania tej społeczności został zorganizowany otwarty konkurs. Przedstawione materiały sugerują, że wyniki tego konkursu powinny być dostępne w najbliższym czasie.

2.3. Ocena zawartości pracy

Przedstawiona do oceny praca doktorska jest skupiona na bardzo technicznych zagadnieniach, takich jak przegląd dostępnych zbiorów danych do trenowania systemów automatycznego

rozpoznawania mowy w języku polskim, analiza ogólnie dostępnych frameworków do badań wydajnościowych systemów ASR oraz na ewaluacji systemów ASR w języku polskim.

Należy w tym kontekście zdecydowanie zaznaczyć tu, że mimo technicznego brzmienia celów prowadzonych badań i pytań stawianych przez Doktoranta, praca doktorska stanowi zdecydowanie wkład w Informatykę. Uporządkowanie, skatalogowanie oraz standaryzacja i udostępnianie zbiorów danych dla systemów ASR, przygotowanie zbiorów oraz scenariuszy do eksperymentów wydajnościowych i ewaluacji systemów ASR jest bardzo ważne dla dalszego rozwoju systemów ASR dla języka polskiego.

3. Konkluzja końcowa

Oceniając całościowy wkład Doktoranta, należy podkreślić ogrom i systematyczność wykonanej pracy. Ponadto ważnym jest fakt, że wszystkie wyniki tej olbrzymiej pracy zostały udostępnione dla społeczności zajmującej się ASR dla języka polskiego. Co więcej, nie tylko zostały przygotowane, ale również podjęte działania dosyć rzadkie w przypadku prac doktorskich – czyli promowanie i komunikowanie wyników. W tym przypadku za pomocą zorganizowania konkursu bezpośrednio wykorzystującego wyniki podjętych badań. Podsumowując, uważam, że cele badawcze, sformułowane przez autora rozprawy, zostały zdecydowanie potwierdzone.

Uważam więc, że rozprawa doktorska mgr. Michała Junczyka spełnia warunki stawiane przez ustawę „Prawo o szkolnictwie wyższym i nauce” w odniesienie do rozpraw doktorskich, a zatem powinna być dopuszczona do publicznej obrony o co wnoszę do Rady Dyscypliny Naukowej Matematyka i Informatyka Uniwersytetu Adama Mickiewicza.

Maria Ganzha