

**RADA NAUKOWA DYSCYPLINY
INFORMATYKA TECHNICZNA I TELEKOMUNIKACJA POLITECHNIKI WARSZAWSKIEJ**

zaprasza na
OBRONĘ ROZPRAWY DOKTORSKIEJ

mgr. inż. Marcina Leszka Sowańskiego

która odbędzie się w dniu **8 kwietnia 2024 roku**, o godzinie **11:00** w trybie hybrydowym (zdalnym i stacjonarnym)

Temat rozprawy:

Multilingual Machine Translation System for Dialogue Agents

Promotor: dr hab. inż. Artur Janicki , prof. uczelni – Politechnika Warszawska

Recenzenci: prof. dr hab. Krzysztof Jassem– Uniwersytet Adama Mickiewicza w Poznaniu

dr hab. inż. Rafał Kozik, prof. uczelni – Politechnika Bydgoska

prof. dr hab. inż. Piotr Nawrocki – Akademia Górniczo-Hutnicza w Krakowie

Obrona odbędzie się zdalnie na platformie MS Teams oraz w Sali nr 478. Osoby zainteresowane uczestnictwem w obronie proszone są o zgłoszenie chęci uczestnictwa w formie elektronicznej na adres sekretarza komisji: dr hab. inż. Grzegorz Stępiak, prof. uczelni – email: grzegorz.stepniak@pw.edu.pl, do dnia 5 kwietnia godz. 12:00.

Z rozprawą doktorską i recenzjami można zapoznać się w Czytelni Biblioteki Głównej Politechniki Warszawskiej, Warszawa, Plac Politechniki 1.

Streszczenie rozprawy doktorskiej i recenzje są zamieszczone na stronie internetowej: <https://www.bip.pw.edu.pl/Postepowania-w-sprawie-nadania-stopnia-naukowego/Doktoraty/Wszczete-po-30-kwietnia-2019-r/Rada-Naukowa-Dyscypliny-Informatyka-Techniczna-i-Telekomunikacja/mgr-inz.-Marcin-Sowanski>

Przewodniczący Rady Naukowej Dyscypliny
Informatyka Techniczna i Telekomunikacja
Politechniki Warszawskiej
dr hab. inż. Jarosław Arabas, prof. uczelni

Wielojęzyczny system tłumaczenia maszynowego dla agentów dialogowych

Streszczenie

W rozprawie przedstawiono sposób w jaki tłumaczenie maszynowe (MT) może być wykorzystane do tłumaczenia zasobów uczących i ewaluacyjnych dla modeli rozumienia języka naturalnego (NLU), które są używane między innymi w inteligentnych asystentach wirtualnych (IVA). Celem tej pracy jest udowodnienie, że MT może być efektywnym narzędziem do lokalizacji językowej w procesie rozwijania IVA dla nowych języków. Jako przykład przemysłowego wdrożenia tego konceptu w pracy użyto asystenta wirtualnego Bixby rozwijanego przez firmę Samsung Electronics. Cel rozprawy został osiągnięty i opisany szczegółowo w tej pracy.

Wszystkie modele, zbiory danych i kod źródłowy opisane w tej rozprawie, z wyłączeniem zasobów użytych podczas prac wdrożeniowych, zostały udostępnione, aby wspierać dalsze badania w tej dziedzinie.

Pomysł wykorzystania modeli MT do tłumaczenia zbiorów treningowych agentów dialogowych jest dobrze opisany w literaturze, ale brakuje dostępnych modeli MT adaptowanych do domeny IVA. Jakość nieadaptowanych modeli MT jest niewystarczająca i, co najważniejsze, nie pozwala na przenoszenie semantycznych anotacji używanych w zasobach NLU z języka wyjściowego do języka docelowego. Współczesne modele NLU wymagają wielu, różnorodnych przykładów uczących dla każdej domeny IVA, którą obsługują, co prowadzi do kolejnego problemu, ponieważ MT zwykle zwraca to samo tłumaczenie dla różnych bliskoznacznych zdań źródłowych. Rozwiązanie tych problemów pozwoliłoby na tańszy i łatwiejszy rozwój agentów dialogowych dla nowych języków. Ponadto pozwoliłoby to na korzystanie z produktów opartych na AI sterowanych głosem przez większą liczbę użytkowników. Jest to istotne, ponieważ w chwili obecnej większość narzędzi AI dostępnych jest jedynie dla języka angielskiego.

W pierwszej części tej pracy omówiono, jakie zasoby są potrzebne do stworzenia MT zaadaptowanego do domeny IVA. Obecnie dostępne zbiory danych NLU i MT są niewystarczające pod względem ilości dostępnych domen a także różnorodności intencji i slotów. W rozprawie zaproponowano nowy zbiór danych o nazwie Leyzer, który rozwiązuje wymienione problemy. Zbiór ten został zaprojektowany do badania jakości modeli NLU i MT. Leyzer obejmuje 18 domen z 186 intencjami w językach angielskim, polskim i hiszpańskim. Jedną z wyróżniających cech tego zbioru danych jest przypisywanie każdemu zdaniu poziomu naturalności oraz wzorca czasownikowego, do którego należy. Ta cecha, niewystępująca w innych tego typu zasobach, pozwala nam śledzić inklinacje (ang. bias) modeli MT oraz lepiej określać jakość tłumaczeń.

W drugiej części tej pracy przedstawiona jest technika adaptacji domenowej MT dla domeny IVA. Przeprowadzone eksperymenty pokazują, jak adaptowanie modeli za pomocą fine-tuningu pozwala poprawić wyniki MT. Stworzone modele mogą przenosić semantyczne anotacje używane w modelach NLU, nazywane slotami, co rozwiązuje jeden z trzech głównych problemów zdefiniowany w tej pracy. Zaadaptowany model MT uzyskał lepsze

wyniki niż linia bazowa, uzyskując +17,21 punktu BLEU na zestawie testowym IVA. Osiągnął wynik F1-score na poziomie 87,54% dla zdań z jednym typem slotu (encji nazwanej) i 65,47% dla zdań z wieloma slotami.

W trzeciej części tej pracy przedstawione jest rozwiązanie problemu braku różnorodności w tłumaczeniach zwracanych przez modele MT. Po dogłębnej analizie ośmiu korpusów NLU w celu zidentyfikowania najczęściej występujących czasowników, opracowana została ontologia czasowników. Ontologia wykorzystuje bazy językowe WordNet i VerbNet, które w połączeniu ze zaadaptowanymi modelami MT umożliwiają generowanie wielu wariantów tłumaczenia. Rozwiązanie pozwala lepiej uchwycić niuanse języka naturalnego, a także umożliwia ulepszyć IVA. Zaprezentowany model zwiększył skuteczność klasyfikacji intencji o 3,8% w stosunku do modelu tłumaczącego na jeden wariant.

Po omówieniu trzech kluczowych komponentów niezbędnych do stworzenia adaptowanego MT, w niniejszej rozprawie omówiono wdrożenia przemysłowe. Każdy element wymienionych wcześniej badań został zastosowany komercyjnie. Pozwala to sprostać wyzwaniom biznesowym związanym z lokalizacją zasobów NLU dla asystenta Bixby, IVA opracowanego przez firmę Samsung Electronics.

Rozprawa kończy się listą moich osiągnięć naukowych, w tym artykułów naukowych, patentów i prezentacji, które wygłosiłem. Wszystkie wymienione elementy przyczyniły się do powstania tej pracy lub są z nią tematycznie związane.

Multilingual Machine Translation System for Dialogue Agents

Abstract

The dissertation presents how machine translation (MT) can be used to translate training and evaluation resources for natural language understanding (NLU) models that are, among others, used in intelligent virtual assistants (IVA). The goal of this thesis is to prove that MT can be used as an efficient tool for language localization in the process of developing IVAs. Samsung's virtual assistant, Bixby, has been provided here as an example of the industrial implementation of this concept. The goal has been met and described in detail in this work.

All models, datasets, and the source code described in this dissertation, excluding the resources used in industrial development, have been released to foster further research on this topic.

The idea of using MT models to translate the training set of dialog agents is well described in the literature, but there are no open-source MT models available that are adapted to the IVAs. The quality of not adapted MT models is insufficient and, most importantly, does not transfer semantic annotations used in NLU resources from source to target. State-of-the-art NLU models require various examples for each IVA domain, which causes another problem as MT tends to return the same translation for different source sentences. Solving all these problems would allow the development of dialogue agents for new languages to be cheaper and easier. Moreover, it would let more users use voice-based AI products that currently are mostly available in English.

The first part of this work discusses what resources are needed to build MT adapted to IVA. Available NLU and MT datasets are insufficient regarding domain coverage and the diversity of intents and slots. A new dataset called Leyzer is proposed that addresses that. The dataset is designed to be used as a benchmark for NLU and MT models. Leyzer covers 18 domains with 186 commands across English, Polish, and Spanish. One of the distinguishing features of the dataset is assigning naturalness level and verb patterns to each sentence. This novelty allows us to track the biases of MT and check the quality of translations.

In the second part of this work, the MT domain adaptation technique for the domain of IVA is presented. The performed experiments show how adapting the models with fine-tuning helps improve the results of MT. Created models can transfer semantic annotations used in NLU models, called slots, which solves the fundamental problem of this thesis. The adapted MT model outperformed the baseline with a +17.21 BLEU point gain on an IVA test set. It achieved an F1-score of 87.54% for single-slot sentences and 65.47% for multi-slot sentences.

In the third part of this work, a solution for the absence of variability in MT outputs is presented. Following an in-depth analysis of eight different NLU corpora to identify the most frequently occurring verbs, a verb ontology is developed. This ontology, grounded in WordNet and VerbNet, when integrated with IVA-adapted Machine Translation models, enables the generation of multiple translation variants. This advancement not only captures the nuances of human language but also enriches the user experience in Intelligent Virtual Assistants. The presented model increased intent classification accuracy by 3.8% relative when compared to single-best translation.

Following the discussion of the three key components required for customizing MT, the study delves into its industrial implementation. Each element of this research has been applied commercially to address business challenges associated with localizing Natural Language Understanding (NLU) resources for Bixby, an IVA developed by Samsung Electronics.

This dissertation ends with a list of my academic achievements, including research articles, patents, and presentations I gave. All of these items contributed to this work or are thematically connected with it.

.....
Marcin Souwanski

(Podpis)

Recenzja pracy doktorskiej mgr. Marcina Sowańskiego „Multilingual Machine Translation System for Dialogue Agents”

Krzysztof Jassem

20 grudnia 2023

1 Wstęp

Celem niniejszej recenzji jest stwierdzenie, czy rozprawa doktorska mgr. Marcina Sowańskiego „Multilingual Machine Translation System for Dialogue Agents” spełnia wymagania Ustawy o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. 2003 Nr 65, poz. 595, ze zm.).

2 Ocena merytorycznej strony rozprawy

Recenzowana rozprawa jest wynikiem doktoratu wdrożeniowego, który z założenia powinien dotyczyć badań związanych z wdrożeniem w podmiocie gospodarczym. W opisywanym przypadku wdrożeniem, któremu podporządkowana jest rozprawa, jest wielojęzyczny system komunikacji między człowiekiem a wirtualnym asystentem, realizowany w firmie Samsung. Charakterystyczną cechą dialogów w tymże systemie jest lakoniczność wypowiedzi oraz ich ukierunkowanie na wykonanie określonej akcji przez system informatyczny. Dodatkowym założeniem jest wielojęzyczność – w rozprawie Autor koncentruje się na trzech językach: polskim, angielskim i hiszpańskim, natomiast wyniki badań mają na celu ich uogólnienie na większą liczbę języków.

Rozwiązanie przyjęte w rozprawie zakłada przetwarzanie sekwencyjne wypowiedzi użytkownika: rozpoznawanie mowy, rozumienie wypowiedzi, wygenerowanie reakcji i wykonanie akcji (lub synteza reakcji głosowej systemu). Aby ułatwić realizację najbardziej złożonego modułu – rozumienia wypowiedzi – Autor proponuje metodę języka pośrednika, do którego tłumaczone są wypowiedzi. W tej sytuacji kluczowym zagadnieniem staje się poprawne tłumaczenie wypowiedzi z oraz na język pośredni – angielski.

W dziedzinie tłumaczenia automatycznego notuje się w ostatnich latach stały postęp nie tylko w jakości translacji (co dokumentowane jest coraz lepszymi wynikami rozwiązań na organizowanych konkursach), lecz również w uniwersalności rozwiązań, co realizowane jest przez publiczny dostęp do modeli umożliwiających tłumaczenie między dziesiątkami języków. Niszą w tym zagadnieniu, której wypełnienie jest intencją rozprawy, jest tłumaczenie krótkich wypowiedzi, które mają na celu przekazanie instrukcji postępowania.

Autor realizuje tę intencję poprzez realizację trzech zadań badawczych: wytrenowanie systemu translacji dopasowanego do języka asystentów głosowych, dostosowanie systemu do transferu slotów (czyli elementów wypowiedzi najbardziej istotnych dla zrozumienia instrukcji postępowania) oraz

opracowanie metody augmentacji danych uczących poprzez tłumaczenie z wariantami czasowników.

W kolejnych częściach recenzji ocenię realizację poszczególnych, wyżej wymienionych zadań.

2.1 Wytrenowanie systemu translacji dostosowanego do języka asystentów głosowych

Standardową metodą dostosowania (dostrojenia) modelu języka do konkretnej dziedziny tłumaczenia automatycznego jest dostarczenie korpusu specyficznych przykładów tłumaczenia. Celem Autora było opracowanie korpusu wypowiedzi, który byłby zrównoważony pod względem wyszczególnionych domen dialogów oraz zdefiniowanego zbioru intencji. Autor postanowił wygenerować taki korpus automatycznie na podstawie manualnie opracowanych gramatyk.

W tym celu Autor opracował 20 wzorców gramatycznych dla generowania wypowiedzi języka angielskiego, z których każdy stał się podstawą do wygenerowania zestawu zdań o różnych stopniach naturalności i zawierających różne intencje.

Podobną pracę tę wykonano dla języków hiszpańskiego i polskiego, uzyskując korpus wypowiedzi w trzech językach.

Łza się w oku kręci. Przypominają się lata 80., kiedy za pomocą ręcznie opracowanych gramatyk próbowano ogarnąć ogrom konstrukcji języka polskiego. Na podstawie reguł tego typu autor niniejszej recenzji zmagał się z problemem tłumaczenia z języka polskiego na język angielski...

...z efektem podobnym do wyniku raportowanego w rozprawie. W stosunku do modelu bazowego (publicznego modelu M2M100) uzyskano w ramach rozprawy co prawda znaczący postęp wyrażony w metryce BLUE i nieco mniejszy w metryce BLEURT, jednak zanotowano bardzo wyraźny spadek jakości na zdaniach spoza dziedziny. Fakt ten stawia pod znakiem zapytania odporność metody na wypowiedzi wykraczające poza zdefiniowaną gramatykę.

Podsumowując tę część recenzji, stwierdzam, że uzyskane wyniki nie przekonują mnie o wyższości zastosowanej metody pozyskiwania danych nad alternatywną metodą – crowd sourcingu – której efektem byłoby pozyskanie danych bardziej zbliżonych do autentycznych.

2.2 Dostosowanie systemu do transferu slotów

W zadaniu transferu slotów Autor dostrajał model bazowy (M2M100) z wykorzystaniem nie tylko korpusu autorskiego, ale również innych dostępnych korpusów dialogów z wirtualnym asystentem. Wynik eksperymentu wydaje się pozytywny: nastąpił prawie dwukrotny wzrost jakości tłumaczenia w stosunku do modelu bazowego, a miara F1 określająca poprawność lokalizacji slotów w tekście docelowym wynosiła ponad 65%.

Z drugiej strony pozostają pewne niedopowiedzenia:

1. Dlaczego w pomiarze jakości tłumaczenia zastosowano wyłącznie wychodzącą z użycia metrykę BLEU?

2. Do jakiego rozwiązania można porównać jakość lokalizacji slotów?

Podsumowując tę część recenzji, chciałbym podkreślić, że metoda zastosowana przez Autora pozwoliła na osiągnięcie wyraźnego postępu w stosunku do metody bazowej. Podane przez Autora dane nie pozwalają jednak czytelnikowi na wyciągnięcie wniosków dotyczących odporności metody oraz porównawczej oceny jakości lokalizacji slotów w tekście docelowym.

2.3 Opracowanie metody augmentacji danych trenujących poprzez tłumaczenie z wariantami czasowników

W mojej ocenie trzeci podrozdział pracy jest najciekawszy, gdyż łączy kompetencje ludzkie ze stosunkowo nowymi metodami tłumaczenia leksykalnego przez system neuronowy. Kompetencje ludzkie niezbędne są do opracowania klasyfikacji czasowników stosowanych w instrukcjach podawanych do asystentów wirtualnych – do jednej klasy mają należeć czasowniki o podobnym znaczeniu w określonym kontekście. Klasyfikacja ta pomaga wygenerować kilka tłumaczeń o podobnym znaczeniu różniących się wyłącznie tłumaczeniem czasownika. Warto zauważyć, że metodologia ingerowania w rezultaty tłumaczenia neuronowego za pomocą leksykonów jest ciekawym wyzwaniem, szczególnie dla języków fleksyjnych.

Niezwykle trudne do analizy są wyniki ewaluacji tych eksperymentów. Okazuje się bowiem, że tłumaczenie z leksykonem okazuje się gorsze niż bez niego. Jest to jednak fakt ogólnie znany i potwierdzony w wielu eksperymentach (z tego powodu leksykony stosuje się tylko w bardzo specyficznych przypadkach tłumaczenia). Optylizmem napawa fakt, że tłumaczenie z trzema wariantami daje lepsze wyniki niż tłumaczenie z jednym wariantem.

Pozytywnie oceniam pomysł przedstawiony w tej części rozprawy. Jedyne zastrzeżenie mam do tego, że nie wyjaśniono, dlaczego wynik niższy od bazowego można uznać w tym wypadku za pozytywny.

2.4 Zastosowania

W mojej ocenie najwyższą wartością rozprawy jest pozytywne wdrożenie jej wyników do gospodarki. Jeżeli przyjmiemy, że najbardziej obiektywną miarą jakości rozwiązania informatycznego jest ewaluacja zewnętrzna, czyli "money in the bank" lub wysoki stopień zadowolenia użytkowników, to szósty rozdział pracy wykazuje, że cel ten został spełniony. Rozwiązanie zostało wdrożone w systemie Bixby, stosowanym przez miliony użytkowników, którzy pozytywnie oceniają swoje doświadczenia.

3 Ocena formalnej strony rozprawy

3.1 Język

Praca napisana jest w języku angielskim. Autor posługuje się językiem obcym bardzo swobodnie. Niezwykle trudno jest dostrzec jakiegokolwiek błędy językowe. Nie ma w rozprawie zbędnych zawiłych konstrukcji zdaniowych – praca napisana jest w sposób klarowny.

3.2 Ocena układu pracy

Układ pracy jest zgodny z przyjętymi standardami. Wstęp zawiera dokładnie tyle informacji, ile jest niezbędnych dla zrozumienia dalszych części. Nie zauważyłem odwołań do niezdefiniowanych wcześniej pojęć.

3.3 Strona estetyczne

Stronę estetyczną oceniam bardzo wysoko. Format wydruku jest czytelny, praca jest ładnie złożona. Czytelność znacząco poprawiają diagramy przepływu danych, które są czytelne i nieskomplikowane.

3.4 Spójność opisu

W ocenie spójności opisu biorę pod uwagę dwa kryteria: jednorodność stosowanej terminologii oraz spójność raportowanych danych. W ramach pierwszego kryterium nie mam większych zastrzeżeń poza nazewnictwem slotów. W tytule rozdziału 4. mowa jest o tłumaczeniu i transferze jednostek (Entity Translation and Transfer), w hipotezie T2 (str. 24) mowa jest o tłumaczeniu jednostek nazwanych, a w większości pracy mowa jest o transferze slotów.

Pod względem spójności raportowanych danych rozprawa pozostawia wiele do życzenia: Liczba domen tłumaczenia, która niewątpliwie powinna być stała w całej pracy, waha się od wartości 18 (wstęp) do 20 (tabela 1) i 21 (tabela 2). Liczba intencji wynosi albo 186 (wstęp), albo 187 (tabela 2), albo nawet 193 (str. 41).

4 Podsumowanie recenzji

W podsumowaniu recenzji odwołuję się do poszczególnych ustępów ustawy o rozprawie doktorskiej.

1. ust. 1. „Rozprawa doktorska prezentuje ogólną wiedzę teoretyczną kandydata w dyscyplinie albo dyscyplinach oraz umiejętność samodzielnego prowadzenia pracy naukowej lub artystycznej.”

Autor posiada ogólną wiedzę teoretyczną w dyscyplinie, co wykazuje w rozdziale 1. rozprawy.

Główne wyniki rozprawy, dotyczące przygotowania korpusu Leyzer, transferu slotów oraz translacji wielowariantowej zostały również opisane w trzech publikacjach międzynarodowych

- w każdej z nich Doktorant jest wymieniony jako pierwszy Autor. Ponadto, Doktorant był liderem zespołu informatycznego, który wdrażał wyniki rozprawy do obszaru gospodarczego. Fakty te niepodważalnie świadczą o umiejętności samodzielnego prowadzenia pracy badawczej przez Doktoranta.

2. ust. 2. „Przedmiotem rozprawy doktorskiej jest oryginalne rozwiązanie problemu naukowego, oryginalne rozwiązanie w zakresie zastosowania wyników własnych badań naukowych w sferze gospodarczej lub społecznej albo oryginalne dokonanie artystyczne.”

Recenzowana rozprawa nie wnosi moim zdaniem nowatorskiego rozwiązania problemu naukowego. Autor w sposób niezwykle umiejętny stosuje i modyfikuje zastany aparat badawczy do rozwiązania problemu praktycznego, czym moim zdaniem wypełnia postulat „zastosowania wyników własnych badań naukowych w sferze gospodarczej”.

3. ust. 3. „Rozprawę doktorską może stanowić praca pisemna, w tym monografia naukowa, zbiór opublikowanych i powiązanych tematycznie artykułów naukowych, praca projektowa, konstrukcyjna, technologiczna, wdrożeniowa lub artystyczna, a także samodzielna i wyodrębniona część pracy zbiorowej.”

W omawianym przypadku rozprawę stanowi praca pisemna w postaci monografii naukowej.

Stwierdzam zatem, że recenzowana rozprawa spełnia wymagania stawiane pracom doktorskim. Rekomenduję dopuszczenie Doktoranta do kolejnych etapów przewodu.

Krzysztof Janiak

Bydgoszcz, 02.01.2024

dr hab. inż. Rafał Kozik, profesor uczelni
Politechnika Bydgoska
im. Jana i Jędrzeja Śniadeckich w Bydgoszczy,
Wydział Telekomunikacji, Informatyki i
Elektrotechniki,
Al. prof. S. Kaliskiego 7,
85-796 Bydgoszcz

Rada Naukowa Dyscypliny
INFORMATYKA TECHNICZNA
I TELEKOMUNIKACJA
Sekretariat
Data wpływu.....*12.01.2024*.....
Numer.....

Recenzja rozprawy doktorskiej pt.:

**„Multilingual Machine Translation System for Dialogue Agents
(Wielojęzyczny system tłumaczenia maszynowego dla agentów
dialogowych)”,**

której Autorem jest Pan mgr inż. Marcin Sowański

1) Jaki jest problem naukowy (teza) rozprawy? Czy został on trafnie i jasno sformułowany?

Rozprawa, której Autorem jest Pan mgr inż. Marcin Sowański, dotyczy złożonych zagadnień tłumaczenia maszynowego danych treningowych, walidacyjnych i testowych przy opracowywaniu modeli rozumienia języka naturalnego. W szczególności, przedstawione i opisane w rozprawie zagadnienia badawcze skupione są na inteligentnych wirtualnych asystentach (IVA).

Przedstawione badania i wyniki eksperymentów mają na celu udowodnić, że tłumaczenie maszynowe jest efektywnym sposobem rozszerzania możliwości inteligentnych wirtualnych asystentów o nowe języki. W szczególności Autor rozprawy porusza aspekty związane z usprawnieniem i automatyzacją procesu tworzenia asystentów.

Pan mgr inż. Marcin Sowański definiuje trzy hipotezy, które (podobnie jak problem naukowy rozprawy) są jasno zdefiniowane. Pierwsza teza zakłada, iż tłumaczenie maszynowe stanowi efektywne narzędzie do tzw. lokalizacji modeli rozumienia języka naturalnego. Druga teza stanowi, iż tłumaczenie maszynowe musi zachować i odpowiednio przetłumaczyć (oraz wskazać położenie) nazwanych jednostek (ang. NERs), aby przetłumaczyć dane treningowe na potrzeby zrozumienia języka naturalnego. Ostatnia teza zakłada, że tworzenie wielu wariantów danych treningowych dla wirtualnych asystentów wirtualnych poprawia dokładność zrozumienia naturalnego języka.

Autor niniejszej rozprawy, również jasno podkreśla istotność i wagę podjętego problemu naukowego. W szczególności Pan mgr inż. Marcin Sowański podkreśla, iż aktualnie brakuje modeli tłumaczenia maszynowego, które byłyby dostosowane do inteligentnych wirtualnych asystentów. Ponadto, Autor rozprawy, przytacza argument, iż jakość istniejących modeli tłumaczenia maszynowego, które nie są dostosowane, jest niewystarczająca.

Podsumowując, w mojej ocenie, problem naukowy rozprawy jest trafnie i jasno sformułowany przez Pan mgr inż. Marcina Sowańskiego.

2) Czy autor rozwiązał postawiony problem? Czy użył do tego właściwych metod dowodząc, że posiadał umiejętności związane z metodyką i metodologią prowadzenia badań naukowych?

W mojej ocenie, Pan mgr inż. Marcin Sowański w sposób odpowiedni rozwiązał problem, którego dotyczy rozprawa oraz użył do tego właściwych metod. W szczególności zagadnienie badawcze zostało podzielone na trzy hipotezy, których słuszność oraz sposób ich dowodzenia został szczegółowo opisany w rozdziałach 3, 4 oraz 5.

W przypadku pierwszej hipotezy (T1), autor słusznie podkreślił istotność posiadania wielojęzycznych zbiorów danych dla trenowania modeli tłumaczenia maszynowego. Autor rozprawy słusznie wskazał, iż istniejące zbiory danych NLU są angielskojęzyczne i mają ograniczone zastosowanie. Dlatego prace na zbiorze danych o nazwie Leyzer wydają się być jak najbardziej celowe. Dodatkowo jest to uzasadnione przez Autora rozprawy tym, że opracowany korpus pozwala na bardziej precyzyjną ocenę agentów dialogowych. Dowodząc drugiej hipotezy (T2), Autor rozprawy badał szczegóły transferu slotów w kontekście

tłumaczenia maszynowego dla inteligentnych wirtualnych asystentów (IVAs). Używając odpowiednio przygotowanych eksperymentów, Pan mgr inż. Marcin Sowański wykazał, iż proponowany mechanizm anotowania slotów i ich transfer pozwala uzyskać tłumaczenia wysokiej jakości. W celu zweryfikowania ostatniej hipotezy (T3), Autor rozprawy zbadał istotność generowania wielu poprawnych wariantów tłumaczeń. Poprzez zastosowanie tej techniki zaobserwowana została poprawa klasyfikacji intencji, zwłaszcza podczas tłumaczenia angielskiego na język polski.

Przy weryfikowaniu wszystkich hipotez użyte zostały właściwe narzędzia i metody. W szczególności Autor rozprawy użył popularnych miar jakości klasyfikacji czy oceny skuteczności systemów tłumaczenia maszynowego. Tym samym Autor dowiódł, że posiadał umiejętności związane z metodyką i metodologią prowadzenia badań. Niemniej, Autor rozprawy mógł porównać więcej (popularnych w literaturze [67]) metod/technik wykorzystywanych w procesie tłumaczenia maszynowego.

3) Czy tematyka rozprawy jest aktualna i dostatecznie ważna?

Na aktualność i wagę tematyki niniejszej rozprawy wpływa kilka czynników. W szczególności, jak zauważył Autor rozprawy, pomysł wykorzystania modeli tłumaczenia maszynowego do adaptowania zestawów uczących dla agentów dialogowych jest opisany w literaturze, ale aktualnie można zauważyć brak takich modeli, które dostosowane byłyby do potrzeb budowania inteligentnych wirtualnych asystentów. Ponadto, niewystarczająca jest jakość tłumaczenia otrzymana w oparciu o modele, które nie są odpowiednio adaptowane. Jak wskazał Pan mgr inż. Marcin Sowański, tłumaczenia takie nie pozwalają przenieść anotacji używanych w zasobach NLU. W tym świetle, rozwiązania proponowane przez Autora rozprawy otwierają możliwość tańszego i łatwiejszego tworzenia agentów dialogowych dla nowych języków. Jednocześnie umożliwia to większej liczbie użytkowników na korzystanie z produktów opartych o sztuczną inteligencję obsługiwaną głosem.

4) Na czym polega oryginalny dorobek autora i jakie jest jego znaczenie poznawcze lub przydatność praktyczna dla nauki bądź techniki?

Na oryginalność dorobku Pana mgr inż. Marcina Sowańskiego wpływa wiele elementów, które powstały w trakcie prac nad niniejszą rozprawą. W szczególności Autor zaproponował nowy zbiór danych o nazwie Leyzer, który rozwiązuje ograniczenia istniejących zbiorów z punktu widzenia różnorodności intencji i slotów. Na podkreślenie zasługuje fakt, iż Leyzer jest publicznie dostępny dla innych badaczy i może być wykorzystywany w niekomercyjnych

zastosowaniach. W mojej ocenie, istotne z punktu widzenia oryginalności dorobku autora, jest też technika adaptacji domenowej MT (przedstawiona w rozdziale drugim) oraz (przedstawiona w rozdziale trzecim) propozycja rozwiązania braku różnorodności w tłumaczeniach zwracanych przez modele MT. O przydatności praktycznej dla nauki i techniki świadczy wdrożenie przemysłowe, które zostało szeroko opisane w rozdziale 6 niniejszej rozprawy.

5) Czy rozprawa świadczy o dostatecznej wiedzy autora, wiedzy o zaawansowanym poziomie, o charakterze podstawowym dla dziedziny nauk technicznych oraz o charakterze szczegółowym, odpowiadającej obszarowi prowadzonych badań naukowych?

Nie mam wątpliwości, iż Autor posiada dużą wiedzę dot. zagadnień związanych z tematyką rozprawy. W szczególności, Pana mgr inż. Marcin Sowański wykazał się umiejętnościami projektowania rozwiązań informatycznych oraz techniczną wiedzą w zakresie budowania inteligentnych wirtualnych asystentów. Autor rozprawy wykazał się również umiejętnościami dotyczącymi prowadzenia badań naukowych. Wielokrotnie w przeprowadzonych eksperymentach przedstawione zostaje autorskie osiągnięcie na tle innych rozwiązań dostępnych w literaturze. W procesie tym wykorzystywane są znane i stosowane w środowisku wskaźniki jakości takie jak BLEU czy F1-score. Na podkreślenie zasługuje także istotny dorobek naukowy Pana mgr inż. Marcina Sowańskiego. W szczególności, w rozdziale 7 niniejszej rozprawy, Autor przedstawił listę powiązanych z tematem rozprawy publikacji oraz udzielony patent, który dotyczy systemów IVA.

6) Czy Autor wykazał umiejętności poprawnego i przekonującego przedstawienia uzyskanych przez siebie wyników?

Zrealizowane przez Autora rozprawy eksperymenty są zgodne z dobrymi praktykami w zakresie pomiaru skuteczności systemów klasyfikacji i detekcji. Autor w tym zakresie użył znanych metryk (np. BLEU czy F1-score), co pozwala odnieść się do innych rozwiązań dostępnych w literaturze. W szczególności Autor rozprawy demonstrując wyniki jasno pokazał, iż proponowany model MT pozwala uzyskać (na zestawie testowym) więcej punktu BLEU niż linia bazowa. Ponadto, demonstrując wyniki, Autor rozprawy klarownie zademonstrował przydatność proponowanej techniki transferu jednostek oraz podejścia opartego wielowariantowe tłumaczenie.

7) Układ i poprawność redakcyjna rozprawy.

Rozprawa składa się z ośmiu rozdziałów. W pierwszej części Autor nakreśla tło, problem badawczy oraz motywację dla swoich badań. W tym świetle określony zostaje cel badawczy niniejszej rozprawy. W kolejnych trzech rozdziałach Pan mgr inż. Marcin Sowański przedstawia autorskie propozycje, które dowodzą, że specjalizowane modele tłumaczenia maszynowego mogą usprawnić proces tworzenia asystentów dialogowych dla nowych języków. W dalszej części rozprawy Autor przedstawia szczegóły na temat wdrożenia oraz swoje główne osiągnięcia w postaci wystąpień, publikacji naukowych oraz patentu.

Rozprawa napisana jest w sposób profesjonalny. Niemniej, w tekście można wskazać kilka niedociągnięć natury edytorskiej lub nieściśłości:

- W tekście nawiązującym i opisującym Figure 11 opisane są element (“ creating English grammar, creating grammars for Spanish and Polish, slot expansion and splitting data into train-, dev- and test sets), których nie widać na obrazie.
- Fig. 12. Czarne kropki nie mają numeracji (cyfry wydają się być przycięte).
- Figure 6 (który podobny jest do Fig.7), może mylnie sugerować, iż m2m100 jest modelem opartym o sieć rekurencyjną. W moim odczuciu, zacytowanie “Beyond English-Centric Multilingual Machine Translation” (Angela Fan et al.) i przedstawienie ich architektury lepiej poparłoby stwierdzenie „sequence-to-sequence (Seq2Seq) is the most popular architecture in NLU”.
- Nazwy korpusów językowych, w szczególności różnych wersji WTM, mogłyby być lepiej opisane i opatrzone odpowiednią referencją do źródła.

8) Jakie są wady i słabe strony rozprawy?

Rolą recenzenta jest zauważenie ewentualnych niedociągnięć i mankamentów przedstawianej pracy oraz zgłoszenie uwag, które mogą być pomocne i przydatne w dalszych pracach. Dlatego w mojej ocenie:

- Wyniki otrzymane w rozdziale 4 mogłyby być bardziej rozbudowane, aby uwypuklić wpływ pewnych parametrów (np. chociażby wielkości modelu, czasu uczenia) na jakość otrzymywanych wyników (np. wartość BLEU).

- Eksperymenty w rozdziale 3 i 4 uwzględniają wyłącznie porównanie do modelu bazowego. Inne modele (poza m2m100) wykorzystywane w tłumaczeniu maszynowym wydają się być pominięte w eksperymentach i analizach (np. facebook/wmt21*, czy facebook/mbart-large*).
- W literaturze można wskazać metody, które wykorzystują różne podejścia do rozwiązywania problemów związanych z IC i SL. W szczególności wspomniana przez Autora rozprawa analiza [67], która dotyczy WMT22, zawiera całe spektrum różnych podejść związanych pośrednio lub bezpośrednio z tłumaczeniem maszynowym. Zastanawiające jest to, czy któreś z technik mogłyby być wykorzystane w przyszłych badaniach Autora rozprawy. Chociażby „back-translation” pojawia się wielokrotnie w [67].
- W rozdziale 1.1.4 Autor rozprawy mógł bardziej szczegółowo ocenić przewidywany wpływ rozwoju dużych modeli językowych na zagadnienia poruszane w pracy. We wrześniu 2023 roku OpenAI ogłosiło możliwość interakcji z modelem poprzez głos, co czyni go inteligentnym asystentem głosowym. Pytanie jak znaczący wpływ ma to na aktualną architekturę IVA oraz na rozwiązania proponowane w niniejszej rozprawie.

Pomimo przedstawionych powyżej uwag, rozprawa Pana mgra inż. Marcina Sowańskiego napisana jest w sposób profesjonalny i posiada wiele mocnych stron. Niewątpliwie recenzowana praca dowodzi dużej wiedzy Autora w tematach dotyczących zagadnień związanych z przetwarzaniem języka naturalnego.

9) Wniosek

Biorąc pod uwagę przedstawioną przez Doktoranta rozprawę stwierdzam, że spełnia ona wymagania stawiane rozprawom doktorskim przez obowiązujące przepisy i wnioskuję o dopuszczenie jej do publicznej obrony.



Kraków, 21.12.2023

Dr hab. inż. Piotr Nawrocki, prof. AGH
Wydział Informatyki
Akademia Górniczo-Hutnicza w Krakowie
Al. Mickiewicza 30
30-059 Kraków
e-mail: piotr.nawrocki@agh.edu.pl

RECENZJA

rozprawy doktorskiej mgr. inż. Marcina Sowańskiego pt.
„Multilingual Machine Translation System for Dialogue Agents”

1. Oświadczenia

1. Nie mam wspólnych prac z doktorantem mgr. inż. Marcinem Sowańskim, nie prowadziłem z nim wspólnych prac badawczych, nie oceniałem jego dorobku wydawniczego ani naukowego i nie pozostajemy w żadnej zależności służbowej.
2. Niniejsza recenzja została opracowana na podstawie pisma dr hab. inż. Jarosława Arabasa, prof. uczelni, Przewodniczącego Rady Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Warszawskiej z dnia 03.11.2023 r.
3. Recenzja jest opracowana zgodnie z wymaganiami określonymi w ustawie „Prawo o szkolnictwie wyższym i nauce” z dnia 20 lipca 2018 r., Dz.U. 2022 poz. 574, z późn. zm.

2. Ocena czy rozprawa doktorska stanowi oryginalne rozwiązanie problemu naukowego

Rozprawa mgr. inż. Marcina Sowańskiego dotyczy wykorzystania tłumaczenia maszynowego (MT) do tłumaczenia zasobów uczących i ewaluacyjnych dla modeli NLU (Natural Language Understanding) używanych między innymi w inteligentnych wirtualnych asystentach (IVA). Zagadnienia poruszane w pracy dotyczą obszarów informatyki i lingwistyki. Istnieją prace dotyczące metod wykorzystywania modeli MT do tłumaczenia zbiorów treningowych agentów dialogowych jednakże brakuje badań dotyczących modeli MT adaptowanych dla inteligentnych asystentów wirtualnych. Brak takich modeli MT nie pozwala na przenoszenie semantycznych anotacji używanych w

zasobach NLU z języka wyjściowego do języka docelowego. Rozwiązanie tego problemu pozwoli na lokalizację językową w procesie rozwijania IVA dla nowych języków z wykorzystaniem tłumaczenia maszynowego.

Zagadnienia rozważane w rozprawie doktorskiej mgr. inż. Marcina Sowańskiego mogą być przedmiotem badań na gruncie dyscypliny „Informatyka Techniczna i Telekomunikacja”.

Mgr inż. Marcin Sowański sformułował następujący cel swojej rozprawy doktorskiej „*wykazanie, że wyspecjalizowane modele tłumaczenia maszynowego mogą znacząco zautomatyzować proces tworzenia asystentów dialogowych dla nowych języków*” oraz trzy tezy doprecyzowujące cel rozprawy:

1. Tłumaczenie maszynowe, po dostosowaniu do języka inteligentnych wirtualnych asystentów (IVA), służy jako skuteczne narzędzie do lokalizacji modeli rozumienia języka naturalnego.
2. W celu przetłumaczenia zasobów treningowych rozumienia języka naturalnego, które obejmują anotacje semantyczne, tłumaczenie maszynowe musi odpowiednio przetłumaczyć lokalizacje nazwanych jednostek.
3. Generowanie wielu wariantów podczas tłumaczenia danych szkoleniowych dla inteligentnych wirtualnych asystentów (IVA) poprawia dokładność rozumienia języka naturalnego.

Pytanie dlaczego mgr inż. Marcin Sowański zdecydował się na określenie aż trzech tez rozprawy, które według niego doprecyzowują cel rozprawy. Wydaje się że bez problemu można było określić jedną tezę, która obejmowałaby wszystkie poruszane w rozprawie zagadnienia. W sytuacji kiedy w rozprawie doktorskiej pojawia się teza i cel, zdecydowanie bardziej istotna z punktu widzenia naukowego jest teza. Dlatego uważam za niefortunne jednoznaczne określenie jedynie celu badań, bez określenia głównej tezy rozprawy doktorskiej.

Mgr inż. Marcin Sowański rozwiązał postawiony problem opracowując oryginalny rozwiązanie tłumaczenia maszynowego zaadaptowanego do domeny IVA. W tym celu zaproponował nowy zbiór danych o nazwie Leyzer rozwiązujący problem różnorodności i intencji slotów. Zbiór Leyzer został opracowany w celu badania jakości modeli NLU i MT i obejmuje 18 domen z 186 intencjami w językach angielskim, polskim i hiszpańskim. Dodanie do tego zbioru poziomu naturalności oraz wzorca czasownikowego, pozwoliło na śledzenie inklinacji modeli MT oraz lepsze określanie jakości tłumaczeń. Drugim istotnym elementem opracowanego rozwiązania jest technika adaptacji domenowej MT dla domeny IVA za pomocą fine-tuningu, która pozwoliła poprawić wyniki MT. Wyniki eksperymentów potwierdziły skuteczność modelu MT w tworzeniu wielojęzycznego NLU. Ostatnim ważnym elementem opracowanego rozwiązania jest ontologia czasowników. Rozwiązuje ona problem braku różnorodności w tłumaczeniach zwracanych przez modele MT. Zaproponowana ontologia, wykorzystująca bazy językowe WordNet i VerbNet, została opracowana na podstawie analizy ośmiu korpusów NLU w celu zidentyfikowania najczęściej występujących czasowników. Opracowane rozwiązanie pozwala na lepsze uchwycenie niuansów języka naturalnego co prowadzi do ulepszenia inteligentnych wirtualnych asystentów.

W rozprawie brakuje jednak wyraźnego określenia co jest osiągnięciem mgr. inż. Marcina

Sowańskiego na poziomie opracowanego rozwiązania programowego, które pozwala na spełnienie tezy pracy. Czy są to osobne trzy elementy uzupełniające istniejący model tłumaczenia maszynowego (MT) czy jest to nowo opracowany mechanizm tłumaczenia maszynowego (MT) zawierający te trzy elementy. Brakuje również w rozprawie rysunku pokazującego architekturę opracowanego rozwiązania na którym można by przedstawić autorskie rozwiązania mgr. inż. Marcina Sowańskiego na tle wykorzystanych narzędzi i mechanizmów. W kontekście przeprowadzonych eksperymentów brakuje punktu, który podsumowywałby wyniki wszystkich testów. Istnieje podsumowanie w rozdziale 6 dotyczącym wdrożenia przemysłowego ale nie wiadomo jak opisane tam wyniki korespondują z wynikami przedstawionymi we wcześniejszych rozdziałach.

Rozprawa jest czytelnie zredagowana i napisana poprawnym językiem. Strona edycyjna rozprawy jest na bardzo wysokim poziomie. Brakuje jednak spisu rysunków, tabel oraz wykazu skrótów co pomogłoby w czytelności rozprawy.

3. Ocena czy rozprawa doktorska prezentuje ogólną wiedzę teoretyczną mgr. inż. Marcina Sowańskiego w dyscyplinie Informatyka Techniczna i Telekomunikacja

Rozprawa doktorska mgr. inż. Marcina Sowańskiego prezentuje jego szeroką wiedzę teoretyczną i praktyczną w różnych obszarach informatyki. Mgr inż. Marcin Sowański wykazał się wiedzą w zakresie tłumaczenia maszynowego, inteligentnych wirtualnych asystentów, modeli rozumienia języka naturalnego oraz ontologii. Ponadto wykazał się wiedzą z zakresu lingwistyki wraz z analizą intencji. Mgr inż. Marcin Sowański wykazał się również umiejętnościami praktycznymi pozwalającymi na przemysłowe wdrożenie tłumaczenia maszynowego dla NLU wykorzystywanego w wirtualnym asystencie Bixby. Warty podkreślenia jest umiejętność mgr. inż. Marcina Sowańskiego opracowania ontologii czasowników wykorzystującej bazy językowe WordNet i VerbNet i pozwalającej wraz z modelami tłumaczenia maszynowego na generowanie wielu wariantów tłumaczenia.

4. Ocena czy rozprawa doktorska wykazuje umiejętność samodzielnego prowadzenia pracy naukowej przez mgr. inż. Marcina Sowańskiego

Rozprawa doktorska mgr. inż. Marcina Sowańskiego wykazuje jego umiejętność samodzielnego prowadzenia pracy naukowej w szczególności sformułowania problemu badawczego w obszarze informatyki technicznej i telekomunikacji, wypracowania oryginalnego rozwiązania postawionego problemu przy użyciu właściwych metod naukowych, wykazania poprawności opracowanego rozwiązania poprzez realizację odpowiednich eksperymentów oraz właściwej redakcji tekstu rozprawy doktorskiej.

5. Konkluzja

Biorąc pod uwagę aktualność tematyki doktoratu oraz osiągnięte wyniki badawcze stwierdzam, że rozprawa doktorska mgr. inż. Marcina Sowańskiego pomimo pewnych uwag krytycznych spełnia wymagania obowiązującej ustawy „Prawo o szkolnictwie wyższym i nauce” i wnoszę o jej dopuszczenie do publicznej obrony.

Nawrodti

