

Recenzja rozprawy doktorskiej
mgr Tomasza Dwojaka,
pt. “Efficient Algorithms for Hybrid Neural
Machine Translation”
(„Wydajne algorytmy dla hybrydowego modelu
neuronowego tłumaczenia maszynowego”)
Promotor: prof. dr hab. Krzysztof Jassem

dr hab. inż. Maciej Piasecki, profesor PWr.
Katedra Inteligencji Obliczeniowej, Politechnika Wrocławska,
maciej.piasecki@pwr.edu.pl

25 maja 2021

1 Podstawowe informacje o rozprawie

Rozprawa, która została przedstawiona do recenzji ma formę zbioru 6 prac naukowych, których Doktorant był współautorem. Grono pozostałych współautorów jest szerokie i obejmuje (pominięte zostały stopnie i tytuły naukowe, liczba wspólnych prac podana została w nawiasach):

4 Marcin Junczys-Dowmunt,

3 Hieu Hoang,

2 Aleksandra Birch, Kenneth Heafield,

1 Alham Frikri Aji, Nikolay Bogoychev, Ulrich Germann, Roman Grundkiewicz, Krzysztof Jassem, Phillip Koehn, Rihards Krislauk, Andre FT Martins, Tom Neckermann, Marta Nadejde, Siva Reddy, Frank Seide, Rico Sennrich, Daniel Torregrosa.

Bardzo duża liczba współautorów wynika z uczestnictwa Doktoranta w dwóch dużych naukowych projektach europejskich i stażach w znanych zagranicznych ośrodkach badawczych, co jest dobrym sygnałem umiejętności współpracy badawczej i doświadczenia.

Zwraca również uwagę duża liczba prac napisanych we współpracy z dr Marcinem Junczysem-Dowmuntem, który jednak został określony mianem



„promotora pomocniczego” w autoreferacie (nie została ta kwestia potwierdzona w dokumentach przesłanych do recenzji wraz z rozprawą).

W autoreferacie do osiągnięć badawczych zostało zaliczonych i opisanych 9 prac, ale do cyklu stanowiącego rozprawę zostało włączonych ostatecznie 6 z nich, dlatego tylko te 6 prac zostało dalej omówionych w recenzji. Wszystkie 6 prac są recenzowanymi pracami o zasięgu międzynarodowym. Wśród nich najważniejszy referat opublikowany na konferencji ACL 2018 (CORE A*). Ponadto cykl obejmuje 1 artykuł w czasopiśmie z IF, referat na warsztatach w ramach konferencji ACL 2018 oraz 3 referaty w ramach światowych konferencji *Conference on Machine Translation* oraz *International Workshop on Spoken Language Translation*. Podsumowując, wszystkie prace z cyklu zostały opublikowane na dobrym i wymagającym poziomie.

Zgodnie z przedłożonymi oświadczeniami i opisem w autoreferacie (który rozszerza krótkie notki w oświadczeniach współautorów) udział Doktoranta w pracach z cyklu był istotny. W artykule z czasopisma udział był bardzo znaczący. Warto również podkreślić, że Doktorant był głównym autorem (50% przy 5 współautorach) pracy [P5], która dotyczy sedna tematu rozprawy.

Przedstawiony zbiór prac został również opatrzony 17 stronicowym Autoreferatem, w którym Doktorant przedstawił tematykę poszczególnych publikacji włączonych do zbioru. Przyjęta forma rozprawy w postaci zbioru publikacji z lat 2016–2019 (choć zasadnicza część cyklu powstała w latach 2016–2018, praca [P6] z roku 2019 ma już charakter popularyzatorsko-aplikacyjny) z definicji nie ułatwia uchwycenia programu badań i jego realizacji. Poszczególne publikacje konferencyjne mają z konieczności ograniczoną długość oraz są napisane dość technicznym i hermetycznym językiem.

Niestety, autoreferat nie jest tak pomocny w zapoznaniu się z programem badań Doktoranta, jakby tego można było oczekiwać. Pierwsze pięć stron autoreferatu to bardzo ogólne wprowadzenie do tłumaczenia maszynowego, wręcz zbliżające się do poziomu popularyzacji nauki i jedynie punkt 1.3 dotyczący złożoności obliczeniowej wprowadza kilka bardziej szczegółowych i istotniejszych informacji. W pozostałej części autoreferatu poszczególne sześć prac tworzących zbiór są omówione każda z osobna i brakuje osi narracyjnej, która spajałaby te opisy. Dopiero w bardzo krótkim Podsumowaniu (osiem zdań) można znaleźć kilka zdań wyjaśniających główne zagadnienia badawcze spinające poszczególne publikacje ze zbioru.

2 Zagadnienie naukowe rozpatrzone w rozprawie

Główną osią rozprawy jest łączenie metod statystycznych i neuronowych w ramach tłumaczenia maszynowego, poprawa efektywności neuronowego tłumaczenia maszynowego (praca z największym wkładem własnym Doktoranta) oraz konstrukcja platform programowych do neuronowego tłumaczenia

maszynowego. Ponadto pojawiają się również zagadnienia wykorzystania dodatkowej wiedzy, np. składniowej, w poprawie jakości działania metod tłumaczenia. Doktorant w „Podsumowaniu” Autoreferatu określił, że skupił się na „hipotezie badawczej o zasadności podejścia łączącego obie metody”, tj. statystyczną i neuronową oraz podkreślił współdziałanie z zespołami z czołowych światowych ośrodków nad „polepszeniem jakości i wydajności dominującej metody neuronowej”.

3 Zaproponowane rozwiązania i zrealizowane zadania

W pracy [P1] dokonano porównania jakości i szybkości tłumaczenia maszynowego przy pomocy metod statystycznych i neuronowych. W przypadku tych drugich zaproponowano oryginalne rozwiązanie w części dekodującej tłumaczenie (tj. po stronie języka docelowego) – system *AmuNMT*, do którego budowy w istotny sposób przyczynił się Doktorant. Zaproponowane rozwiązanie zostało porównane w szeroko zakrojonych eksperymentach dla trzydziestu języków z system statystycznego tłumaczenia maszynowego reprezentującym aktualny stan badań. W konstrukcji zastosowano kombinację kilku znanych technik poprawy efektywności oraz przebadano eksperymentalnie wpływ poszczególnych parametrów i ich kombinacji, np. metody doboru podstłownika i wielkości wiązki dekodowania. *AmuNMT* osiągnął szczególnie dobrą wydajność w przypadku działania na zwykłych procesorach (CPU), dzięki technikom dobrego wykorzystania wielowątkowości. Poczynione obserwacje o większej szybkości tłumaczenia z wykorzystaniem procesorów graficznych były raczej do przewidzenia już w roku 2016. Główną wartością poznawczą było eksperymentalne potwierdzenie przewagi metod neuronalnych. Zbudowany system *AmuNMT* okazał się cennym osiągnięciem w zakresie technologii i został później wykorzystany w wielu pracach badawczo-rozwojowych.

W pracy [P2] zaproponowane zostało mieszane podejście, w którym na etapie dekodowania „neuronowy model z atencją” został zintegrowany z „statystycznym modelem opartym na frazach”. Było to pierwsze podejście do integracji tego typu modeli. Zastosowano algorytmy wsadowe dekodowania oraz zaproponowano dwukrotnego przejścia przez stos hipotez, dzięki czemu poprawiono efektywność działania. Ponadto skonstruowano szybką implementację w języku C++. Zbudowane rozwiązanie uplasowało się w na czołowych pozycjach w konkursie systemów tłumaczących.

Praca [P3], opublikowana w ramach prestiżowej konferencji ACL, przedstawia system *Marian NMT* stanowiący bardzo wydajną platformę do implementacji neuronowych systemów tłumaczenia maszynowego. Do tej pory system doczekał się dziesiątek zastosowań, w tym wdrożeń w bardzo znanych, dużych firmach, nie tylko w ramach tłumaczenia, ale również, np., w ramach korekty gramatycznej. System *Marian NMT* to niewątpliwie bardzo



znaczące osiągnięcie w dziedzinie technologii językowych i bardzo zauważalny polski wkład w rozwój tłumaczenia maszynowego.

Praca [P4] przedstawia ciekawe poszukiwania w zakresie wzbogacania reprezentacji tekstu na potrzeby poprawy jakości tłumaczenia. Postawiono przekonującą hipotezę, że sieci neuronowe nie radzą sobie ze śledzeniem niektórych rodzajów zależności pomiędzy tekstowo oddalonymi elementami struktury zdania. W tym celu zaproponowano dwa sposoby wzbogacania reprezentacji tekstu na etapie treningu o wyniki automatycznej analizy składniowej. Zadanie systemu tłumaczącego zostało rozszerzone o przewidywanie metadanych reprezentujących strukturę składniowe. Praca opublikowana na najbardziej znaczącej światowej konferencji z dziedziny tłumaczenia maszynowego wykazała uzyskaną poprawę jakości tłumaczenia. Przeprowadzono bogatą analizę wyników oraz analizę wybranych klas błędów, która pokazała słuszność sformułowanej hipotezy i celność zaproponowanego rozwiązania. Niemniej, uważna lektura pracy przynosi pewne wątpliwości, które nie zostały całkowicie rozwiane przez przedstawione badania empiryczne. Autorzy zastosowali automatyczny parser składniowy, który z natury jest obciążony błędem i działa w oparciu o pewien wytrenowany model. Mogło to w pewien sposób znormalizować wyniki tłumaczenia, chociaż ocena jego jakości jest pozytywna. Pytanie, jak system zachowa się, gdy rodzaj tłumaczonego tekstu będzie bardziej odległy od modelu działania parsera. W pracy możemy też odnaleźć stwierdzenie (str. 72), że dłuższe zdania, z którymi parser sobie nie radził zostały usunięte zarówno ze zbioru treningowego jak i testowego. Brakuje jednak informacji ile było takich zdań oraz analizy wpływu tego obciążenia na działanie całości systemu.

Praca [P5] wydaje się być najbardziej znacząca w dorobku Doktoranta, ponieważ została opublikowana na bardzo dobrym światowym forum i udział Doktoranta w niej jest przeważający (50% przy 5 współautorach). Została w całości poświęcona zagadnieniu poprawy efektywności neuronowego tłumaczenia maszynowego, które wydaje się jednak głównym osiągnięciem Doktoranta w ramach prac stanowiących rozprawę. W pracy [P5] wykorzystano autorski algorytm ograniczonego przetwarzania wsadowego na etapie dekodowania (ang. *mini-batching*), pojawił się on już w pracy [P2], oraz autorskiego efektywnego algorytmu szacowania wartości prawdopodobieństwa (ang. *softmax*) w połączeniu ze śledzeniem k -najlepszych hipotez (ang. *k-best*). Eksperymentalny system został zbudowany jako rozszerzenie wspomnianego już wcześniej systemu *AmuNMT*. Pewnym niedostatkiem pracy jest to, że przedstawione porównanie efektywności jest ograniczone jedynie do systemu *Marian NMT*, co prawda jednego z czołowych na świecie. Brakuje w autoreferacie informacji o wpływie rozwiązań zaproponowanych w pracy [P5] na dalszy rozwój dziedziny.

Ostatnia z prac włączonych do zbioru, tj. [P6], zawiera ciekawe porównanie statystycznego i neuronowego systemu tłumaczenia maszynowego zastosowanych w ograniczonej i wyspecjalizowanej dziedzinie tekstów dotyczących

przemysłu stoczniowego. Warto podkreślić, że to druga z prac, w której jest bardzo znaczący udział Doktoranta, tj. 50%, tym razem przy dwóch współautorach. Porównano standardowy statystyczny model z systemem *Marian NMT*. Obydwa systemy zostały wytrenowane i ocenione na pozyskanym korpusie dziedzinowym obejmującym ponad 1,2 mln zdań i równoważników (co daje 14 mln tokenów poziomu wyrazowego). Korpus został podzielony stale na część treningową i bardzo niewielką część testową (trochę ponad 22 tys. tokenów). Testy całego procesu na tym dziedzinowym korpusie pokazały nieznaczną przewagę jakości systemu statystycznego, ale za to przewagę jakościową systemu neuronowego w ręcznej ocenie. Niestety nie obliczono statystycznej istotności zaobserwowanych różnic. Ciekawostką jest fakt, iż porównanie obu wytrenowanych systemów z komercyjnym system Google wykazało ich lepszą jakość i to według zarówno miary automatycznej jak i ręcznej. Szkoda, że w pracy nie przeanalizowano na ile wpływ na ten wynik miał fakt, że wykorzystany korpus dziedzinowy pochodzi z pamięci tłumaczeniowej systemu wspomagającego pracę tłumaczy-ludzi przez kilka lat, więc można się spodziewać dużej zgodności i systematyczności tłumaczenia na poziomie fraz. Tłumaczy to przewagę systemu statystycznego, co w pracy [P6] nie jest rozważone. Zabrakło również porównania z innymi metodami i współczesnymi systemami poza Google Translate. Praca [P6] zawiera również kilka dalszych eksperymentów z porównaniem obu systemów: statystycznego i neuronowego na kilku publicznie dostępnych zbiorach danych, w tym również języka lotewskiego. Przedstawiono ciekawe analizy jakościowe wybranych błędów.

Każda z prac włączonych do zbioru, opublikowanych w bardzo dobrych miejscach, zawiera porównania z metodami z literatury.

4 Osiągnięcie celów rozprawy i wkład własny Doktoranta

W trakcie pracy nad publikacjami składającymi się na rozprawę (przyjmując tu okres lat 2016–2019) Doktorant był zaangażowany w badania realizowane przez czołowe światowe grupy badawcze w dziedzinie tłumaczenia maszynowego. Brał udział w realizacji dużych i bardzo dobrych europejskich projektów badawczych. Jest praktycznie niemożliwością, aby system tłumaczący zbudować w pojedynkę, a dopiero kompletny system stanowi właściwą weryfikację proponowanych rozwiązań. Stąd jest w pełni zrozumiałe, że wszystkie publikacje Doktoranta są publikacjami wieloautorskimi.

We „Podsumowaniu” Autoreferatu, Doktorant podkreślił swoje prace nad porównaniem statystycznego i neuronowego podejścia do tłumaczenia maszynowego. W ramach prac ze zbioru wątek analizy porównawczej pojawia się, np. w [P1], przedstawiono kilka badań porównawczych, studium przypadku w pracy [P6] jest w dużym stopniu poświęcone takiej analizie, ale wyciąga-

ne wnioski są dość ograniczone, przede wszystkim do konkretnej dziedziny i porównywanych systemów. Podobnie wcześniejsze analizy porównawcze, np. w pracy [P1], zostały przeprowadzone z ogólnej perspektywy: systemy statystyczne vs. neuronowe. Brakuje analizy odnoszącej się do różnych metod w obrębie tych paradygmatów i analizy wpływu poszczególnych cech i rozwiązań.

Doktorant wskazuje również na prace nad łączeniem metod neuronowych i statystycznych, które w opisanych w pracach eksperymentach przyniosły dobre wyniki. Brakuje w Autoreferacie spojrzenia w perspektywie czasu – a upłynęło już kilka lat od momentu opublikowania prac ze zbioru – jak proponowane wtedy rozwiązania wpłynęły na rozwój metod w dziedzinie lub jak sytuują się wobec aktualnych metod.

Z analizy prac ze zbioru i wkładu w nie Doktoranta, najważniejszym osiągnięciem wydają się prace nad poprawą efektywności działania neuronowego dekodera w systemie tłumaczenia maszynowego. Szczegółowe rozwiązania zaproponowane przez Doktorant w zakresie wspomnianych już wsadowego przetwarzania i efektywnego algorytmu szacowania wartości prawdopodobieństwa (ang. *softmax*) w połączeniu ze śledzeniem *k*-najlepszych hipotez (ang. *k-best*) są wartym zauważenia w rozwój algorytmów tłumaczenia.

Niewątpliwie najważniejszym dokonaniem Doktoranta jest istotny udział w budowie i rozwoju unikatowych systemów *AmuNMT* i *Marian NMT*, gdzie szczególnie wpływ tego drugiego na rozwój dziedziny jest bardzo istotny, a wdrożenia szerokie i znaczące.

5 Redakcja rozprawy

Przedstawienie wyniku realizacji prac badawczych w postaci zbioru publikacji jest wbrew pozorom formą trudną. Autoreferat został za bardzo w swojej strukturze zdominowany przez uwagę ukierunkowaną na poszczególne publikacje i ich wartość z osobna. Brakuje spojrzenia na całość, a szczególnie na relację osiągniętych wyników do rozwoju dziedziny. Pierwsza część Autoreferatu jest na zbyt ogólnym poziomie i dlatego nie wnosi nic istotnego. Mogłaby być z powodzeniem zastąpiona przedstawieniem planu badań, nawet w perspektywie retrospektywnej, i być przewodnikiem po przedstawionych dalej pracach. Również trochę mylące jest włączenie do tabeli publikacji w autoreferacie prac, które nie składają się na rozprawę, tj. ostatnich trzech.

Zgadając się z Doktorantem, że wszystkie prace ze zbioru zostały opublikowane na bardzo dobrych światowych forach, od strony formalnej, należy jednak zauważyć, że zaproponowana ich ekstrapolacyjna ocena punktowa może być trochę myląca.



6 Zastosowania badawcze i wdrożeniowe wyników

Doktorant miał szczęście pracować na realizacją swoich badań w najlepszych zespołach naukowych na świecie, przy realizacji bardzo dużych i znaczących projektów, a przede wszystkim nad opracowaniem i zbudowaniem bardzo istotnych systemów w dziedzinie tłumaczenia maszynowego, w tym systemu *Marian NMT*. Przełożyło się to na bardzo wysoką liczbę cytowań większości prac ze zbioru (jak zostało to sprawdzone niezależnie od Doktoranta, w tej liczbie ilość autocytowań przez kogokolwiek ze współautorów jest niewielka) oraz na imponujący zakres wdrożeń systemu *Marian NMT* i jego wpływ na dziedzinę. Oczywiście Doktorant był członkiem zespołu odpowiedzialnego jako całość za te osiągnięcia, ale tego typu przedsięwzięcia jest niemożliwym realizować w pojedynkę.

Odnosić wielu osiągnięć szczegółowych, jak już było wspomniane, brakuje w Autoreferacie pokazania ich wpływu na aktualny stan dziedziny (poza licznymi cytowaniami), jednak wszystkie zostały opublikowane w ramach wymagających konferencji i warsztatów naukowych.

7 Konkluzja

Badania przedstawione w rozprawie dotyczą rozwoju metod tłumaczenia maszynowego, zarówno w odniesieniu do ich efektywności, architektur systemów tłumaczenia maszynowego, jak również efektywnych sposobów reprezentacji wyrażeń językowych i w pełni mieszczą się w zakresie tematycznym dyscypliny informatyka techniczna i telekomunikacja.

Praca spełnia w wystarczającym stopniu formalne i zwyczajowe wymagania stawiane rozprawom doktorskim.

Wnoszę o dopuszczenie Doktoranta do dalszych faz przewodu doktorskiego.

