RADA NAUKOWA DYSCYPLINY
INFORMATYKA TECHNICZNA I TELEKOMUNIKACJA POLITECHNIKI WARSZAWSKIEJ

zaprasza na

OBRONĘ ROZPRAWY DOKTORSKIEJ

mgr. inż. Mateusza Klimaszewskiego

która odbędzie się w dniu **13 kwietnia 2026 roku,** o godzinie **12:00** w trybie stacjonarnym

Temat rozprawy:

"Multi-objective modularity in Natural Language Processing"

Promotor: dr hab. inż. Tomasz Gambin, prof. uczelni – Politechnika Warszawska

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

dr hab. Agnieszka Mykowiecka, profesor instytutu – Instytut Podstaw Informatyki Polskiej Akademii Nauk

dr hab. Piotr Pęzik, prof. uczelni – Uniwersytet Łódzki

Obrona odbędzie się stacjonarnie w Audytorium Centralnym Wydziału Technik Informacyjnych Politechniki Warszawskiej w Warszawie.

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.-Mateusz-Klimaszewski

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

Wielokryterialna modularność w przetwarzaniu języka naturalnego

Dziedzina przetwarzania języka naturalnego (NLP) odnotowała w ostatnich latach błyskawiczny postęp dzięki ciągłemu skalowaniu głębokich sieci neuronowych (DNN). W rezultacie, pojawienie się modeli podstawowych znacząco zmieniło krajobraz NLP, gdzie wstępnie wytrenowane modele językowe i duże modele językowe (LLM) osiągają wyjątkowe rezultaty w licznych zadaniach.

Trwający proces skalowania głębokich sieci neuronowych doprowadził do wzrostu popularności modularnego uczenia głębokiego (MDL). Metody MDL prezentują tańszą alternatywę do całkowitego dostrajania modeli poprzez parametrycznie efektywny odpowiednik, który redukuje wymagania obliczeniowe modyfikacji modelu. W szczególności, w przypadku sztandarowych modeli LLM, które swoim rozmiarem przekraczają 100 miliardów parametrów, nawet predykcja wymaga znacznej infrastruktury sprzętowej, pomijając całkowicie koszty treningu takiego modelu. Co więcej, moduły MDL mogą wzbogacać modele podstawowe dodając im różne umiejętności, często wykorzystując jednocześnie wiele z nich i łącząc je w zależności od potrzeb zadania docelowego.

Niniejsza praca skupia się na modularnym uczeniu głębokim i bada wielokryterialne aspekty metod modularnych w formie serii czterech publikacji. W pierwszych trzech z nich stawiane jest pytanie badawcze, które rozważa konkretny problem w istniejących aspektach: wielozadaniowości, wielodomenowości i wielojęzyczności. Owe prace skupiają się na jednym, konkretnym przypadku wielokryteriowości na raz, bez modyfikacji pozostałych (np. podczas eksploracji konfiguracji zadań wielozadaniowych zachowujemy niezmienioną domenę i język(i)). Każda z pierwszych trzech publikacji przedstawia nowy system lub metodę, która ma na celu usunięcie ograniczeń obecnych rozwiązań w danym aspekcie. W ostatniej, czwartej publikacji, został sformułowany nowy aspekt, nazwany wielomodelowym. Praca testuje istniejące metody modularne w nowych warunkach i ukazuje ograniczenia obecnych metod modularnych.

Słowa kluczowe: Modularne uczenie głębokie, Parametrycznie efektywne dostrajanie, Przetwarzania języka naturalnego

Warszawa, 9.12.2025

dr hab. Agnieszka Mykowiecka
Instytut Podstaw Informatyki PAN
Jana Kazimierza 5, Warszawa

Recenzja rozprawy doktorskiej mgr. Mateusza Klimaszewskiego

Multi-objective modularity in Natural Language Processing

Recenzja rozprawy doktorskiej mgr. Mateusza Klimaszewskiego, zrealizowanej pod opieką prof. Tomasza Gambina i promotora pomocniczego dr. Piotra Andruszkiewicza, wykonana została na zlecenie Rady Naukowej Politechniki Warszawskiej dyscypliny informatyka techniczna i telekomunikacja.

Na zasadniczą treść pracy doktorskiej składają się 4 artykuły opublikowane na bardzo dobrych międzynarodowych konferencjach. Praca ma postać spójnego tekstu o ujednolicone j formie edycyjnej, w którym artykuły poprzedzone zostały wprowadzeniem zwierającym przedstawienie ogólnego zamysłu pracy i jej kontekst. Pracę zamyka uwspólniona bibliografia. Ze względu na fakt, że zasadniczą treścią pracy są artykuły opublikowane na znaczących konferencjach, co oznacza, że przeszły one przez proces recenzyjny, w swojej recenzji skupię się na ogólnej charakterystyce rozprawy, nie wchodząc w analizę każdego detalu poszczególnych artykułów.

Zawartość pracy

Przedstawionym we wstępie tematem spajającym składające się na rozprawę artykuły jest sprawdzenie efektywności modeli językowych w konfiguracjach innych niż realizacja jednego zadnia. Testowane warianty uogólnień to wiele-zadań, wiele-dziedzin, wiele-języków i wiele modeli. Tak postawiony problem jest bardzo aktualny gdyż wielość możliwych zastosowań sprawia, że z jednej strony naturalnym zdaje się budowanie konfiguracji zdolnych rozwiązywać więcej niż jeden wariant postawionego zadania, a z drugiej strony wykorzystanie danych przeznaczonych dla innych zadań może pomóc rozwiązać takie, dla którego mamy mało danych treningowych i/lub testowych. W ostatnich latach zyskał on odrębną nazwę – modularne uczenie głębokie (MDL) – i spore zainteresowanie.

Pytania sformułowane przez doktoranta to:

1. Jak zbudować wielozadaniowy system NLP, w którym użytkownik może wybrać odpowiednie zadania końcowe?
2. Czy można wykorzystać zewnętrzny wielodomenowy model nauczyciela ?
3. Czy możemy wykorzystać wiedzę z wielu modułów specyficznych dla danego języka bez zwiększania kosztów wnioskowania?
4. Czy możemy ponownie wykorzystać wstępnie wytrenowane moduły z różnych modeli podstawowych (*Foundation Models*)? Jakie są ograniczenia obecnych metod modułowych podczas przenoszenia między modelami?

Pierwszy artykuł, z roku 2021, gdy nie mówiono jeszcze o pojęciu uczenia modularnego, opisuje opracowany pod kierunkiem dr Aliny Wróblewskiej, ale w dużej mierze samodzielnie, system COMBO, implementujący podstawowe narzędzia pozwalające na wstępna analizę zdań w języku polskim. Jest on tagerem, analizatorem morfologicznym, lematyzatorem i

parserem, budującym reprezentujące strukturę syntaktyczną zdania drzewa zależnościowe. Można przy tym wybrać, które z wymienionych zadań mają być realizowane.

Zaproponowana architektura wykorzystuje szereg cech słów, takich jak embedding słowa, część mowy oraz wartości cech morfologicznych. Cechy te są uwzględniane przez model wtedy, gdy wybrane zadanie wymaga ich do trenowania. Autor zaproponował metodę kodowania wartości cech w sposób kontekstowy poprzez przepuszczenie wartości dla poszczególnych słów przez sieć BILSTM. Dla rozwiązywania poszczególnych zadań zdefiniowane są odrębne architektury korzystające z kontekstowej reprezentacji cech i zewnętrznych wektorów reprezentacji słów. Przykładowo, predykcja łuków zależnościowych odbywa się w sposób następujący: Dwie pojedyncze warstwy FC przekształcają globalne wektory cech dla elementów głównych i zależnych. Reprezentacją zależności jest macierz sąsiedztwa, której elementy są iloczynami skalarnymi wszystkich par osadzeń głównych i zależnych (iloczyn skalarny określa pewność krawędzi między dwoma słowami). Funkcja softmax zastosowana do każdego wiersza macierzy przewiduje sąsiednie pary głównych i zależnych. Takie podejście nie gwarantuje jednak, że wynikowa macierz sąsiedztwa jest prawidłowo zbudowanym drzewem zależności. W ostatnim etapie przewidywania stosuje się zatem algorytm Chu-Liu-Edmonds [39, 56].

Wartością zaproponowanego rozwiązania jest dostarczenie narzędzi do trenowania własnych modeli, co jest możliwe dzięki przyjęciu standardowego modelu wejścia i formalizmu Universal Dependencies jako formy reprezentowania składni języka. Modele wytrenowane przez autora wykazały się bardzo dobrą jakością w porównaniu do najlepszych w danym momencie rozwiązań (Stanza i spaCy) i bardzo dobrą efektywnością na etapie trenowania, przy wolniejszym przebiegu predykcji.

Druga praca – Gated Adapters for Multi-Domain Neural Machine Translation – dotyczy poprawy wyników tłumaczenia maszynowego poprzez dodanie informacji o dziedzinie, z której pochodzą teksty. Zaproponowana architektura pozwala na pozostawienie bez zmian początkowego modelu. Trenowane są tylko mniejsze moduły dodatkowe (Adapters). W standardowym ustawieniu neuronowego tłumaczenia maszynowego adapter (AD) przetwarza ukryty stan transformera i składa się warstwy normalizacji oraz dwóch warstw liniowych: warstwy zwężającej i warstwy rozszerzającej, z funkcją aktywacji RelU. Zaproponowane rozwiązanie, nazwane Gated Adapters, polega na możliwości łączenia informacji z wielu adapterów w miejsce wyboru jednego. Mechanizm ten jest dołączony na każdym z poziomów transformera i daje w wyniku ważoną średnią wyniku wszystkich adapterów z tego poziomu. Wynik działania każdego adaptera jest mnożony przez wartość prawdopodobieństwa dostarczoną przez zewnętrzny moduł podający stopień, w jakim zdanie należy do domeny reprezentowanej przez adapter.

Możliwe jest wykorzystanie zaproponowanej architektury bez znajomości a priori, który z adapterów powinien zostać wykorzystany, gdyż klasyfikacja odbywa się on-line z przetwarzaniem tekstu. W rozpatrywanym zadaniu uczenia maszynowego adaptery odpowiadały za dostosowanie się systemu do tematyki tekstu. Łączenie adapterów pozwala na lepsze dostosowane się do nieznanej tematyki, zbliżonej do więcej niż jednej z tych ustalonych a priori. Ewaluacji dokonano na tłumaczeniu z angielskiego na polski i z angielskiego na grecki, przy 6 zakresach tematycznych dla obu par. Ewaluacja wyników maszynowego tłumaczenia zawsze nastręcza trudności, gdyż tekst może być przetłumaczony na różne sposoby, a ocena jakości tłumaczeń jest subiektywna. Trudno zatem polegać do końca na metodach automatycznych, ale w pracy zastosowano najbardziej zaawansowane z nich. W cytowanych wynikach widać poprawę wyników, choć nie jest to poprawa bardzo widoczna. Zacytowany przykład, zapewne staranie dobrany, ale jednak rzeczywisty, pokazuje jak istotne jest w niektórych przypadkach prawidłowe rozpoznanie kontekstu tłumaczenia.

Trzecia praca – No Train but Gain: Language Arithmetic for training-free Language Ad-

apters enhancement – dotyczy ulepszenia architektury wykorzystującej adaptery językowe. Autor proponuje zamiast binarnej selekcji języka składanie wektorów reprezentujących różne języki. W środowisku wielojęzycznym pozwala to na dostosowanie modelu do jednego z wielu języków oraz uwzględnienie informacji o językach pokrewnych. Wykorzystując adaptery językowe i zadaniowe, arytmetyka językowa umożliwia przetwarzanie bez konieczności dodatkowego trenowania w dwóch przypadkach użycia: (i) zero-shot, gdzie adapter językowy dla języka docelowego nie został wytrenowany lub (ii) w celu ulepszenia istniejących adapterów językowych poprzez arytmetykę z językiem pokrewnym lub językiem, na którym został wytrenowany adapter zadania.Adaptery językowe są tu rozłączne od adapterów związanych z konkretnym zadaniem. Autorzy przeprowadzili eksperymenty z danymi w 13 językach sumarycznie w kontekście trzech zadań: rozpoznawania nazw własnych (Named Entity Recognition, NER), wnioskowania (Natural Language Inference, NLI) i odpowiadnaia na pytaia (Question Answering, QA). Wykorzystano przy tym ogólnodostępne zasoby danych. Eksperymenty przeprowadzono z użyciem dwóch wielojęzycznych modeli: XLM-R i mBERT. Uzyskana poprawa wyników to do 3 punktów F1 bez konieczności wykonywania dodatkowego trenowania. Nieco zaskakujące są generalnie niższe wyniki dla zadania NER niż NLI.

Ostatnia część rozprawy – Is Modularity Transferable? A Case Study through the Lens of Knowledge Distillation – poświęcona jest zbadaniu możliwości transferu wiedzy (rozumianego jako transfer parametrów) z jednego modelu do drugiego. Rozpatrywane są dwa schematy. W pierwszym oba modele różnią się tylko głębokością sieci, w drugim cała struktura, w tym wielkość reprezentacji, może być różna. Główną inspiracją dla tego pomysłu było umożliwienie trenowania mniejszych modeli w wykorzystaniem już dokonanego treningu modeli większych. Przeprowadzono eksperymenty z modelami mBert, DistilBert, XLM-Base i XLM-Large w kontekście trzech zadań: rozpoznawania nazw własnych, tworzenia parafraz oraz wnioskowania o zależnościach między zdaniami. Potwierdziły one wstępne założenia o możliwości transferu wiedzy w przypadku modeli o tym samym wymiarze reprezentacji. W przypadku przenoszenia informacji z modelu głębszego do posiadającego mniejszą liczbę warstw metodą skuteczniejszą od uśredniania wag okazało się kopiowanie wag co którąś warstwę. Niestety przenoszenie informacji między modelami różniącymi się wielkością reprezentacji w niektórych wariantach nie przyniosło oczekiwanych rezultatów. Ten problem wymaga dalszych badań.

Ocena

Przedstawiona mi do oceny rozprawa zawiera cztery artykuły, których wspólnym tematem jest uelastycznienie lub ulepszanie rozwiązań pozwalające na zmniejszenie kosztów trenowania lub/i poprawienie wyników w konfiguracji multi-językowej lub/i multi-zadaniowej. Podjęty temat różnie rozumianej modularności rozwiązań jest bardzo aktualny wobec dużych kosztów trenowania modeli i bardzo różnorodnej gamy możliwych zastosowań oznaczającej konieczność wytrenowania bardzo dużej ich liczby. Obecnie powszechne wykorzystywanie LLM nieco zmieniło punkt ciężkości w tym obszarze, co autor zauważa w kończącym rozprawę podsumowaniu, jednak wiele problemów wciąż nie ma zadowalającego rozwiązania, a zatem i tej eksplorowanej w pracy ścieżki nie można uznać za zamkniętą.

Zaproponowane w pracy modyfikacje istniejących rozwiązań to typowa droga postępu w nauce, gdzie tylko od czasu do czasu mamy okazję obserwować znaczący przełom. Publikowanie wyników prac na najważniejszych konferencjach dowodzi, że podejmowane tematy były uznane za aktualne i ważne przez społeczność międzynarodową.

Oczywiście rozprawa będąca zbiorem artykułów ma standardowe niedostatki związane z tą koncepcją. Zaprezentowane rozwiązania budowane na osiągnięciach innych nie są bardzo szczegółowo opisane. Struktura sieci przedstawiona jako schematy blokowe, skądinąd dobrze skonstruowane i cenne, to trochę mało dokładny poziom opisu. W szczególności dla mnie nie jest dostatecznie dobrze opisane jak tworzone są adaptery językowe. Oczywiście można zajrzeć do repozytoriów i tam prześledzić rozwiązanie, jest to jednak już poza samą rozprawą. Podobnie praca nie zawiera opisów wykorzystywanych zbiorów danych. Jest to zrozumiałe w artykułach, ale stanowi pewien brak rozprawy.

Uzyskiwane dzięki zaproponowanym modyfikacjom rezultaty nie były na ogół znacząco różne od tych uzyskiwanych innymi metodami. Niewielka poprawa w przypadkach gdy wyniki te są dość niskie, nie wydaje się mieć dużego praktycznego znaczenia. Jednak wobec ograniczeń kosztu rozwiązania nawet niewielkie pogorszenie wyników może okazać się opłacalne. Jako osobie zajmującej się przetwarzaniem danych tekstowych brakuje mi w pracy przykładów lingwistycznych, choć rozumiem że przy skali i liczbie eksperymentów trudno w artykule umieścić sensowną ilustrację przykładami tego, co zostało poprawione (tu jest jeden przykład z tłumaczeniem) lub uległo zmianie na gorsze po wprowadzeniu zmian. W rozszerzonej wersji pracy głębsza analiza jakościowa wyników mogłaby jednak znaleźć miejsce.

Bardzo pozytywnie oceniam stronę edycyjną rozprawy. Zebranie prac w spójny tekst rozszerzony o dosyć obszerny wstęp jest bardzo dobrym pomysłem na zaprezentowanie rozprawy niemającej postaci spójnej monografii. Dobrym ruchem było też uwspólnienie bibliografii, choć jej długość (204 pozycje) budzi pewne wątpliwości, czy naprawdę została przeczytana. Język pracy jest dobry, praktycznie nie zauważa się żadnych błędów językowych czy pomyłek literowych. Jednym zauważonym przeze mnie miejscem, w którym błąd wpływa na znaczenie am jest użycie 'prove' w zdaniu (str 70): "In the following Sections 5.3.2-5.3.2, we present the framework ... and (ii) an enhancement case, where we prove existing language adapters ..." Czy to miało być 'improve'?

Wniosek końcowy

Stwierdzam, że przedłożona mi do recenzji rozprawa mgr. Mateusza Klimaszewskiego zawiera opis istotnych osiągnięć w dziedzinie badania architektur sieci rozwiązujących problemy NLP ze szczególnym uwzględnieniem modularności rozwiązań. Doktorant wykazał się dużą wiedzą w tematyce związanej z rozprawą i zaproponował oraz przetestował na ogólnodostępnych danych własne rozszerzenia istniejących metod. Propozycje te zyskały uznanie społeczności międzynarodowej skupionej wokół najważniejszych dziedzinowych konferencji. Moim zdaniem rozprawa spełnia wymagania stawiane rozprawom doktorskim i wnoszę o dopuszczenie magistra Mateusza Klimaszewskiego do publicznej obrony.

A. Tykocinski

Recenzja pracy doktorskiej mgr. Mateusza Klimaszewskiego „Multi-objective modularity in deep learning for natural language processing"

autor: Krzysztof Jassem
26 października 2025

1. Wstęp

Niniejsza recenzja dotyczy rozprawy doktorskiej mgr. Mateusza Klimaszewskiego pt. *„Multi-objective Modularity in Deep Learning for Natural Language Processing"*. Rozprawa, napisana w języku angielskim, ma charakter cyklu czterech publikacji naukowych. Prace te zostały zebrane w formie spójnej monografii, poprzedzonej wprowadzeniem oraz omówieniem tła teoretycznego dotyczącego zastosowania modularnego uczenia głębokiego (Modular Deep Learning, MDL) w przetwarzaniu języka naturalnego (Natural Language Processing, NLP).

2. Ocena merytorycznej strony rozprawy

2.1. Zakres badań i pytania badawcze

W rozdziale 1. Autor rozprawy formułuje cztery pytania badawcze:

Wielozadaniowość: Jak zaprojektować system NLP, który umożliwia użytkownikowi swobodny wybór zestawu realizowanych zadań?

Wielodziedzinowość: W jaki sposób, w ramach jednego zadania, skutecznie połączyć wiedzę specjalistyczną pochodzącą z różnych domen?

Wielojęzyczność: Czy możliwa jest efektywna integracja wiedzy z modeli generatywnych obsługujących różne języki naturalne?

Wielomodelowość: Czy można w sposób skuteczny scalić informacje pozyskiwane z wielu modeli generatywnych?

Zaznaczam, że powyższe sformułowania problemów badawczych stanowią parafrazę dokonaną przez Recenzenta.

Oceniam, że postawione pytania badawcze mają charakter aktualny i odnoszą się do zagadnień istotnych dla rozwoju współczesnego przetwarzania języka naturalnego. Odnoszę wrażenie, że wynikają one z realnych potrzeb praktycznych, a nie z perspektywy teoretycznej, co podkreśla ich znaczenie aplikacyjne.

2.2. Wkład badawczy

Podrozdział 1.2 przedstawia wkład badawczy Autora, potwierdzony publikacjami w prestiżowych konferencjach (EMNLP, LREC, ECAI, COLING). Warto podkreślić, że w większości prac udział Doktoranta jest dominujący – tak przynajmniej wynika z Jego oświadczenia. Nie znalazłem oświadczeń współautorów prac, ale w każdej pracy wchodzących w cykl Kandydat znajduje się na pierwszym miejscu.

Wkład Autora rozprawy przedstawia się następująco:

- opracowanie modularnej metody analizy morfosyntaktycznej oraz jej implementacja w ogólnodostępnym systemie *Combo*,

- zastosowanie mechanizmu *Gated Adapters* w zadaniu wielodziedzinowego tłumaczenia automatycznego,

- opracowanie koncepcji *language arithmetic*, będącej adaptacją idei *task arithmetic* do środowiska wielojęzycznego,

- wprowadzenie i eksperymentalna weryfikacja koncepcji *transferu modułów zadaniowych* pomiędzy różnymi modelami językowymi.

Przedstawione osiągnięcia mają zarówno znaczenie teoretyczne, jak i praktyczne. Autor wprowadza nowe rozwiązania w obszarze modularnego uczenia głębokiego, poszerzając możliwość adaptacji modeli językowych do wielu zadań, domen i języków.

2.3. Umiejscowienie w literaturze

Rozdział 2 dostarcza przeglądu badań nad modularnym uczeniem głębokim oraz jego zastosowaniami w NLP. Autor wykazuje staranność w oznaczaniu referencji oraz osadzenie własnych badań w szerszym kontekście. Dzięki temu rozprawa ukazuje nie tylko oryginalny wkład Autora, ale i pełne zrozumienie miejsca, jakie ten wkład zajmuje w rozwoju dziedziny.

2.4. Wartość merytoryczna poszczególnych publikacji

COMBO: State-of-the-Art Morphosyntactic Analysis (EMNLP 2021, System Demonstrations)

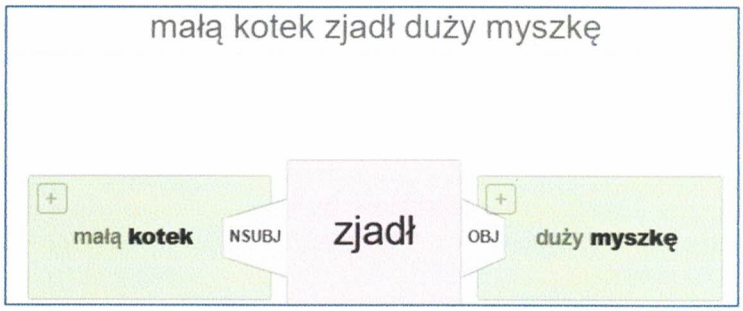
Publikacja przedstawia system *COMBO*, który jest modułową architekturą do analizy morfosyntaktycznej języka naturalnego. System wspiera następujące zadania:

- tokenizacja,
- analiza morfologiczna,
- lematyzacja,
- parsing zależnościowy.

Według Autora architektura systemu pozwala użytkownikowi końcowemu (ang. *end-user*) na elastyczne definiowanie modułów wejściowych (np. wykorzystanie cech morfologicznych lub znaczników POS) oraz wyjściowych (np. ograniczenie się tylko do parsera).

Dla mnie nieco mylące jest użycie w pracy terminu *end-user* w odniesieniu do programisty, który ma możliwość wytrenowania nowego modelu na swoich danych i zastosowania go w systemie *COMBO*. Takiej możliwości (zgodnie z moimi eksperymentami) nie ma bowiem użytkownik końcowy, który korzysta z systemu dostępnego na stronie https://combo-demo.nlp.ipipan.waw.pl/combo-eng.

Odporność narzędzi do analizy zależnościowej języka polskiego ciekawie bada się na przykładach krótkich zdań o strukturze nieciągłej (z którymi dobrze radzą sobie parsery regułowe). Autorowi rozprawy pozostawiam analizę rezultatów systemu *COMBO* dla zdań „Małą kotek zjadł duży myszkę" (rysunek poniżej) oraz „Która jest godzina?" (w porównaniu z wynikiem parsingu dla pytania „Która godzina jest?").



Podsumowując, *COMBO* jest narzędziem inżynieryjnym, które obrazuje praktyczne zalety modułowości (jeszcze przed pojawieniem się pojęcia *Modular Deep Laearning*). System zyskał uznanie międzynarodowe – znalazł się w czołówce konkursu IWPT 2021.

Gated Adapters for Multi-Domain Neural Machine Translation (ECAI 2023)

Artykuł dotyczy zastosowania metody wykorzystującej dane pochodzące z różnych domen w zadaniu tłumaczenia automatycznego. W tym celu opracowano architekturę *Gated Adapters,* w której komponent pełniący rolę routera decyduje o sposobie łączenia wyjść pochodzących z różnych adapterów domenowych.

Praca adresuje realny problem: w sytuacji, gdy tekst wejściowy nie może zostać jednoznacznie przypisany do jednej domeny, zamiast ograniczać się do pojedynczego źródła danych treningowych, możliwe jest efektywne wykorzystanie danych pochodzących z wielu domen równocześnie.

Na szczególną uwagę zasługuje *aplikacyjny charakter* metody, gdyż zaproponowane rozwiązanie może być stosowane w rzeczywistych systemach tłumaczenia automatycznego.

No Train but Gain: Language Arithmetic for training-free Language Adapters enhancement

W tej publikacji wprowadza się koncepcję *language arithmetic,* dla której inspiracją jest idea *task arithmetic*, przeniesiona tutaj na poziom reprezentacji językowych. Badania eksperymentalne pokazują, że możliwe jest połączenie adapterów wytrenowanych dla różnych języków (np. angielskiego i francuskiego) w celu uzyskania nowego adaptera dla języka trzeciego (np. hiszpańskiego), co pozwala osiągnąć poprawę jakości w scenariuszu *zero-shot* (czyli bez dodatkowych przykładów uczących).

Praca ukazuje zdolność autorów do twórczej adaptacji istniejących paradygmatów badawczych nowych obszarów zastosowań.

Wyniki mogą znaleźć praktyczne znaczenie szczególnie dla języków o mniejszych zasobów (*low-resource languages*).

Is Modularity Transferable? A Case Study through the Lens of Knowledge Distillation

Destylacja wiedzy ma na celu zwiększenie efektywności modeli językowych poprzez zmniejszenie ich rozmiarów. W omawianej pracy weryfikuje się, czy w procesie destylacji możliwe jest skuteczne przeniesienie modułów zadaniowych z modelu większego (nauczyciela) do modelu mniejszego (ucznia). Przeprowadzone eksperymenty wykazują, że jest to możliwe w scenariuszu, w którym model ucznia powstał jako destylowana wersja modelu nauczyciela, natomiast skuteczność tego podejścia istotnie maleje w przypadku modeli trenowanych niezależnie.

Badania zrealizowano na danych wielojęzycznych, obejmujących korpusy w 20 językach, oraz dla trzech reprezentatywnych zadań NLP: rozpoznawania nazw własnych (NER), identyfikacji parafrazy oraz wnioskowania w języku naturalnym (NLI). Projekt eksperymentalny potwierdza zatem uniwersalność prowadzonych analiz.

Wartość merytoryczna pracy polega na krytycznej ocenie możliwości oraz ograniczeń metod modularnego uczenia głębokiego.

2.5. Replikowalność wyników

W ocenie prac badawczych z dziedziny informatyki istotnym aspektem jest replikowalność uzyskanych wyników. W tym kontekście należy podkreślić, że wszystkie wyniki badań Autora zostały udostępnione w postaci pakietów oprogramowania o otwartym dostępie, co umożliwia ich niezależne odtworzenie i ponowne wykorzystanie. Takie podejście zwiększa transparentność badań oraz znacząco ułatwia dalszy rozwój prac w dziedzinie rozprawy.

3. Ocena formalnej strony rozprawy

3.1. Język

Praca napisana jest w języku angielskim na wysokim poziomie. Styl jest precyzyjny i pozbawiony zbędnych komplikacji. Autor unika zdań złożonych!

3.2. Układ pracy

Struktura rozprawy jest logiczna i przejrzysta. Rozdziały 1 i 2 dostarczają zarówno wprowadzenia teoretycznego, jak i kontekstu dla własnych badań. Odwołania do literatury są poprawne i spójne.

3.3. Estetyka

Rozprawa jest dobrze zredagowana i czytelna. Schematy rysunkowe tabele są klarowne i ułatwiają odbiór treści.

4. Wpływ pozostałych osiągnięć doktoranta

Rozprawa doktorska obejmuje cykl publikacji, z których trzy zostały przygotowane we współpracy z zespołami krajowymi, natomiast jedna powstała w ramach współpracy międzynarodowej. Na uwagę zasługuje pozostały dorobek publikacyjny Doktoranta, realizowany we współpracy z badaczami zagranicznymi. Świadczy to o ugruntowanej pozycji międzynarodowej Autora.

Za jedno z najbardziej znaczących osiągnięć Autora uważam jego udział w konsorcjum międzynarodowym, które opracowało wielojęzyczny model językowy *EuroLLM*. Model ten stanowi istotny krok naprzód w zakresie przetwarzania języka naturalnego dla języków europejskich. Co istotne, jego wysoka jakość potwierdza się w zastosowaniach praktycznych, o czym miałem możliwość przekonać się osobiście, stosując go w komercyjnym systemie tłumaczenia automatycznego.

5. Podsumowanie recenzji

Stwierdzam, że rozprawa doktorska magistra Mateusza Klimaszewskiego spełnia wymagania stawiane pracom doktorskim i rekomenduję dopuszczenie Doktoranta do dalszych etapów przewodu doktorskiego.

Ponadto wnioskuję o nadanie rozprawie wyróżnienia, co uzasadniam następującymi, obiektywnie potwierdzonymi osiągnięciami Doktoranta:

- wszystkie publikacje składające się na rozprawę są ocenione na liście MNiSW na poziomie 140 punktów,

- Autor rozprawy jest pierwszym autorem każdej z tych publikacji,

- Kandydat jest autorem kodów źródłowych udostępnionych w formie sześciu otwartych pakietów programistycznych,

- Doktorant prowadził cztery własne projekty badawcze oraz uczestniczył w realizacji trzech kolejnych, w tym dwóch o charakterze międzynarodowym,

- Kandydat jest jedynym przedstawicielem Polski w międzynarodowym konsorcjum opracowującym wielojęzyczny model językowy EuroLLM.

Krzysztof Jassem

dr hab. Piotr Pęzik, prof. UŁ
Wydział Filologiczny, Uniwersytet Łódzki
ul. Pomorska 171/173, 90-236, Łódź

Łódź, 1.12.2025

Recenzja rozprawy doktorskiej Pana mgr. inż. Mateusza Klimaszewskiego pt. "*Multi-objective modularity in Natural Language Processing*" napisanej pod kierunkiem prof. PW, dr hab. inż. Tomasza Gambina.

Przedstawiona rozprawa doktorska składa się z dwóch rozdziałów wstępnych, części zasadniczej *w postaci czterech spójnych tematycznie publikacji,* których pierwszym autorem jest Doktorant, rozdziału podsumowującego oraz bibliografii obejmującej 204 pozycje.

We wstępie do scalonej wersji rozprawy Autor wprowadza paradygmat *modularnego uczenia głębokiego* (MDL z ang. Modular Deep Learning) jako ważny obszar badań nad modelami językowymi opartymi na głębokich sieciach neuronowych zwłaszcza w kontekście przetwarzania języka naturalnego. Należy w tym miejscu podkreślić wyjątkową aktualność tej tematyki. Po początkowym zachwycie nad możliwościami ogólnego przeznaczenia generatywnych Dużych Modeli Językowych jako uniwersalnej formy sztucznej inteligencji działającej niezależnie od domeny czy konkretnego zadania, powszechne przekonanie o rychłym nadejściu AGI (generycznej sztucznej inteligencji) wyraźnie osłabło. Generatywne modele językowe operują w spektrum memoryzacji, interpolacji i ekstrapolacji wiedzy i zdolności. Innymi słowy, wiedza i umiejętności, które nie znalazły się w docelowej postaci w pre-treningu i post-treningu bazowych modeli

językowych (ang. foundation models) są mniej lub bardziej udanie interpolowane na podstawie danych z innych języków, pokrewnych domen czy podobnych zadań. Jak podkreśla autor adaptacja poprzez parametrycznie efektywne dostrajanie (PEFT) modeli bazowych może skutkować znaczną optymalizacją obliczeniową w konkretnych zastosowaniach. Można wykazać, że celowana adaptacja językowa, domenowa i zadaniowa modeli pre-trenowanych np. poprzez dostrajanie PEFT na dedykowanych danych może skutkować jakościową poprawą działania nawet największych modeli generatywnych. Często jest to poprawa, która decyduje o przydatności modelu w warunkach produkcyjnych, a nie jedynie demonstracyjnych.

Rozdział drugi nakreśla teoretyczne podstawy modularnego uczenia głębokiego, które Autor wywodzi z badań nad uczeniem transferowych: rozwój metod MDL wynikał z jednej strony z potrzeby strony ograniczenia negatywnego transferu zadaniowego czy językowego, a z drugiej z konieczności znacznej optymalizacji procesu dostrajania modeli głębokich do konkretnych domen czy zadań. Osobnym aspektem MDL poruszonym w tej części wstępu są metody scalania modeli (model merging), których celem jest m. in. 'arytmetyka językowa' (termin wprowadzony w publikacji P3), czyli wzmacnianie kompetencji językowych w różnych konfiguracjach adapterów modelu nauczyciela i studenta.

Można uznać, że w tych krótkich wstępnych rozdziałach Autor klarownie wykazuje spójność tematyczną przedstawionych w dalszej części publikacji jako prac nad modularnością głębokich modeli językowych w aspekcie wielozadaniowości, wielodziedzinowości, wielojęzyczności oraz wielomodelowości. Najzwięźlej tę spójność wyrażają jasno sformułowane na

stronie 12 cztery główne pytania badawcze, z którymi autor zmierzył się w zgłoszonych publikacjach.

Rozdział 3 to pierwsza publikacja z cyklu zatytułowana *COMBO: State-of-the-Art Morphosyntactic Analysis* przedstawiony na konferencji EMNLP w 2021 r. Udział doktoranta w tej publikacji określono na 60%. Artykuł przedstawia system przetwarzania języka naturalnego COMBO, którego najważniejszą, choć nie jedyną funkcją jest analiza składniowo-zależnościowa tekstów w ponad 40 językach. Wyróżnikiem tego rozwiązania jest wysoka skuteczność w zakresie predykcji rozkładów zależnościowych (SOTA w 2021 r.), optymalizacja procesu trenowania, dostępność reprezentacji wektorowych dla poszczególnych warstw strukturalnych oraz modularność użytkowa. Osobne głowice predykcyjne COMBO obsługują różne zadania z zakresu analizy morfologicznej, morfosyntaktycznej i składniowej-zależnościowej. W niezależnej kontynuacji tej pracy powstał również moduł ekstrakcji jednostek nazewniczych. Praca ta zasługuje z pewnością na bardzo wysoką ocenę a raportowane w niej wyniki można uznać za poziom światowy NLP w 2021 r. (być może z wyjątkiem podzadania lematyzacji). Po lekturze publikacji nasuwa się pytanie o aktualność tego typu rozwiązań w 2025 r. Sama przydatność badawcza parserów składniowych choćby w językoznawstwie komputerowym, czy w humanistyce cyfrowej nie budzi wątpliwości recenzenta. W części wstępnej brakuje natomiast w moim odczuciu kilkuzdaniowego odniesienia do skuteczności dużo większych modeli klasy LLM w tym obszarze. Czy oparty na modelach reprezentacyjnych parser COMBO nadal jest nie tylko wielokrotnie wydajniejszy, ale też dokładniejszy od LLM-ów na tych zadaniach? Czy może adaptacja LLM-ów np. poprzez pełne lub adaptacyjne dostrojenie modeli bazowych na tym zadaniu daje porównywalne wyniki?

Rozdział 4 stanowi druga publikacja z cyklu, zatytułowana *Gated Adapters for Multi-Domain Neural Machine Translation,* przedstawioną na konferencji European Conference on Artificial Intelligence (ECAI 2023). Udział Doktoranta w tej pracy określono na poziomie 65%. Zaproponowano w niej system tłumaczenia maszynowego złożonego z tzw. mieszaniny ekspertów (ang. MoE, mixture of experts), a dokładniej modułów modelu językowego pre-trenowanych na jednej z góry założonych domen, m.in. prawo, medycyna, informatyka. W odróżnieniu od standardowych adapterów domenowych, rozwiązanie to nie wymaga odrębnego klasyfikatora dziedzinowego, który może spowalniać działanie całego systemu, a dodatkowo nie dokonuje sztywnego przypisania próbki tekstu do jednej domeny. Zaproponowane podejście polega na użyciu modelu routera destylowanego z zewnętrznego modelu RoBERTa, który pozwala na 'wieloetykietową' klasyfikację domenową tłumaczonego tekstu. Ewaluacja na zadaniu tłumaczenia maszynowego wykazuje przewagę tego rozwiązania nie tylko nad systemami z zewnętrznym klasyfikatorem domenowym, ale też (w przypadkach niejednorodnych tekstów) nawet nad systemami z tzw. 'wyrocznią dziedzinową', tzn. takimi, które znają prawdziwą domenę tłumaczonego tekstu. Praca ta stanowi doskonały przykład zalet modularności domenowej głęboki modeli językowych. Można też stwierdzić, że stanowi udaną próbę wzmocnienia pozytywnego transferu między domenami poprzez równoległa aktywację adapterów. Potencjalną wadą tego rozwiązania może być konieczność zapewnienia stosunkowo dużych ilości danych domenowych dla tłumaczonych języków.

Rozdział 5 stanowi publikacja pt. *No Train But Gain: Language Arithmetic for training-free Language Adapters Enhancement*, przedstawiona na konferencji COLING 2025 z 70% udziałem Doktoranta. Tematem pracy jest scalanie pretrenowanych adapterów językowych jako beztreningowa metoda dopasowania

językowego modeli. Podejście to rozpatrywane jest w dwóch głównych aspektach: a) scenariusz 'zero-shot', czyli adaptacyjne aproksymacje języków, na których nie dokonywano treningu bazowym oraz b) wzmacnianie kompetencji języków podobnie lub słabiej reprezentowanych w pre-treningu. Ewaluację przeprowadzono na docelowych zadaniach wykrywania jednostek nazewniczych, inferencji i odpowiadania na pytania. Wyniki eksperymentów są szczególnie przekonujące dla scenariusza zero-shot (braku adaptera dla danego języka) oraz dla języków słabo reprezentowanych (ang. low-resource languages). Publikacja ta stanowi przykład doskonałej znajomości tematyki adaptacji językowej głębokich modeli. Na marginesie tej pracy nasuwa się pewna wątpliwość terminologiczna, dotycząca trafności rozróżnienia między PLM-ami (Pre-trained Language Models) a LLM-ami (Large Language Models). Desygnaty tych terminów są w kontekście publikacji jasne: PLM-y to zdaje się przede wszystkim transformery reprezentacyjne, a LLM-y to nowszej generacji transformery typu text-to-text o znacznie większej liczbie parametrów. Jednakże każdy LLM jest w dosłownym znaczeniu tego terminu również PLM-em, w związku z czym to rozróżnienie (obecne również w podsumowaniu rozprawy) może być mylące bez dokładnej definicji.

Czwarta publikacja z cyklu zawarta w **Rozdziale 6** pt. *Is Modularity Transferable? A Case Study through the Lens of Knowledge Distillation* zaprezentowana została podczas konferencji LREC-COLING 2024 z 80% udziałem Doktoranta. Praca prezentuje badanie przenośności modułów wydajnego dostrajania (PEFT) między modelami nauczyciela i studenta w dwóch scenariuszach, tj. 'łatwym' i 'trudnym'. Łatwość scenariusza pierwszego polega na tym, że do modelu studenta przenoszone są adaptery o kompatybilnej architekturze (zgodność między modelami *mBERT* i *distilbert-base-multilingual-cased*). W trudniejszym

scenariuszu architektury modeli są niekompatybilne. W trakcie ewaluacji udaje się potwierdzić, że przeszczepienie warstw nauczyciela skutkuje lepszym wykonywaniem szeregu standardowych zadań NLP. Co ciekawe, w trakcie eksperymentów autorzy potwierdzają hipotezę o specjalizacji warstwowej modeli: podejście SKIP polegające na transferze co n-tej warstwy modelu nauczyciela daje lepsze wyniki niż uśrednianie sąsiadujących warstw. Praca ta niewątpliwe stanowi ważne badanie modularności modelowej głębokiego uczenia.

Wnioski i konkluzje z recenzji

Przedstawiona do recenzji rozprawa doktorska zasługuje na jednoznacznie pozytywną ocenę i jednocześnie ***spełnia wymogi stawiane w ustawie Prawo o szkolnictwie wyższym i nauce , Dz. U. 2023, poz. 742 z późn. zm***. Dodatkowo w ocenie Recenzenta ***zasługuje na wyróżnienie***. Autor proponuje i pomyślnie waliduje oryginalne rozwiązania problemu naukowego jakim jest modularność głębokiego uczenia w obszarze przetwarzania języka naturalnego. Przedstawiony cykl publikacji, których pierwszym i głównym autorem jest Doktorant potwierdza szeroką wiedzę w dyscyplinie informatyki technicznej i telekomunikacji oraz bezsprzeczne umiejętności samodzielnego prowadzenia pracy naukowej. Zaproponowane w rozprawie podejścia w sposób kompleksowy i wieloaspektowy traktują modularność głębokiego uczenia. Doktorant projektuje i przeprowadza eksperymenty na zróżnicowanych zbiorach danych osiągając np. w zadaniu analizy składniowej języka naturalnego wyniki na poziomie światowym. Wywód przeprowadza w klarowny i konsekwentny sposób z zachowaniem poprawności metodologicznej właściwej dla dyscypliny i obszaru badań. Wobec powyższego ***wnioskuję o dopuszczenie Doktoranta do dalszych etapów przewodu doktorskiego***.

Piotr Jeziorek