Dr hab. inż. Maciej Piasecki, prof. PWr.
Katedra Sztucznej Inteligencji
Politechnika Wrocławska
maciej.piasecki@pwr.edu.pl

8 II 2026

Recenzja rozprawy doktorskiej

mgr. Katarzyny Krasnowskiej-Kieraś

zatytułowanej:

Automatyczne tworzenie składnikowo-zależnościowych rozbiorów wypowiedzeń z wykorzystaniem głębokich sieci neuronowych

1. Rozprawa od strony formalnej

W myśl Ustawy z dnia 20 lipca 2018 r. „Prawo o szkolnictwie wyższym i nauce", przedłożoną rozprawę stanowi „praca pisemna, w tym monografia naukowa". Rozprawa ma prawidłową strukturę. Składa się z czterech rozdziałów i trzech dodatków. Dwa pierwsze rozdziały poświęcone są omówieniu dziedziny problemu w zarysie z koncentracją uwagi na najważniejszych pojęciach. Praca nie jest obszerna i przede wszystkim wprowadzenie do problemu i porównanie proponowanych rozwiązań ze stanem badań jest bardzo skondensowane. W omówieniu kluczowe elementy są wybrane i przedstawione trafnie, ale narracja ma charakter raczej typowy dla krótkich referatów konferencyjnych, niż obszerniejszej pracy, gdzie nie ma limitu stron.

Własny wkład twórczy, począwszy od celu, planu badawczego poprzez zaproponowane metody i rozwiązania technologiczne, a skończywszy na wnikliwych badaniach eksperymentalnych jest omówiony w bardzo dobry i ciekawy sposób. Przedstawiona praca pokazuje wielką zaletę tradycyjnej rozprawy, która pozwala na systematyczne i dogłębne przedstawienie rezultatów wieloletnich badań. Należy bardzo docenić trud Doktorantki włożony w napisanie kompletnej rozprawy.

2. Problem badawczy i jego znaczenie

Celem badań przedstawionych w rozprawie było opracowanie „algorytmu automatycznego parsowania dla wypowiedzeń w języku naturalnym, operującym na hybrydowych strukturach składniowych zawierających w sobie kompletne, w pełni ze sobą spójne reprezentacje składnikową i zależnościową." Od samego początku celem było opracowanie metody opartej na ogólnym schemacie uczenia nadzorowanego (str. 39), co jest uzasadnione jednak biorąc pod uwagę bogate zbiory danych treningowych dla języka polskiego w tym zadaniu. Postawiony problem ma wymiar zarówno badawczy jak i praktyczny.

Przede wszystkim cel rozprawy kontynuuje długą i cenną tradycję pracy nad opracowaniem dla języka polskiego systemu podstawowych narzędzi językowych o wysokiej jakości działania pozwalającej na praktyczne zastosowania. W tym aspekcie w ramach prac nad rozprawą opracowano nowy parser dla języka polskiego, który wykazuje się nie tylko lepszymi parametrami jakościowymi w stosunku do stanu badań i technologii, ale też poszerza zakres analizy poprzez połączenie reprezentacji zależnościowej z elementami reprezentacji składnikowej.

Z punktu widzenia badawczego, cel rozprawy wpisuje się w nurt badań nad automatyczną analizą składniową (czyli parsowaniem) traktującą łącznie reprezentację zależnościową i składnikową. Nie wchodząc w szczegóły, ta pierwsza opisuje struktury składniowe jako grafy relacji binarnych pomiędzy segmentami wyrazowymi, a ta druga jako rodzaj hierarchicznej agregacji wyrażeń językowych. Połączenie obu reprezentacji daje pełniejszy obraz struktury składniowej wypowiedzeń. Zaproponowano przynajmniej kilka podejść do tego problemu w literaturze w ciągu ostatnich lat. Podejście zaproponowane w rozprawie uzupełnia częściowo luki w dotychczasowym stanie badań.

Metoda parsowania zaproponowana w rozprawie przyczyniła się do empirycznej weryfikacji bardzo cennych badań nad rozwojem językoznawczych metod opisu polskiej składni oraz dużych gramatycznych zasobów językowych. Metoda ta dobrze wykorzystała własności istniejących zasobów zbudowanych przez kilka lat przy znacznych nakładach pracy. Warto również podkreślić, że w ramach zaproponowanego algorytmu parsowania zintegrowano również skutecznie bardzo cenne polskie zasoby morfologiczne, między innymi poprzez rzutowanie wyjścia parsera na opisy morfologiczne ze analizatora morfologicznego Morfeusz. Zbudowany parser ułatwi dalszy rozwój połączonych polskich zasobów składniowych: składnikowo-zależnościowych.

W ten sposób przeszliśmy do znaczenia praktycznego rozprawy, które jest bardzo duże. Doktorantka opracowała kompletne rozwiązanie: parser o bardzo dobrej skuteczności wraz z procesem jego trenowania na polskich zasobach językowych oraz wybranych neuronowych modelach językowych dla języka polskiego. Zbudowany parser został już wykorzystanych w badaniach korpusowych, i jak było to już wspomniane, stanowi cenne wsparcie dla dalszego rozwoju zasobów składniowych dla języka polskiego.

3. Oryginalny wkład w rozwój dyscypliny

Doktorantka twierdzi (str. 10), że „hybrydowa postać struktur produkowanych przez [jej] parser […] jest nowością na gruncie automatycznego przetwarzania polszczyzny" i należy się z tym całkowicie zgodzić. Cennym wkładem pracy jest nie tylko zaproponowanie nowej metody analizy składniowej, ale też zrealizowana przy jej pomocy empiryczna weryfikacja prac z dziedziny językoznawstwa korpusowego w postaci koncepcji i budowy dużego zasobu składniowego dla języka polskiego. Tym samym rozprawa wnosi cenny wkład w rozwój językoznawstwa korpusowego w odniesieniu do języka polskiego.

Bardzo ważnym osiągnięciem w ramach rozprawy jest również jej praktyczne wdrożenie w postaci kompletnego rozwiązania – pełnej implementacji parsera rozszerzonego o moduły ujednoznaczniania morfosyntaktycznego (tagowania) oraz rozpoznawania i klasyfikacji wystąpień nazw własnych.

W stosunku do dotychczasowego stanu prac nad parserami dla języka polskiego, w ramach rozprawy zaproponowano pewne uproszczenie architektury neuronowej parsera z wykorzystaniem wstępnie wytrenowanego modelu językowego opartego na sieci neuronowej typu Transformer (dokładniej HerBERT). Przyniosło to zaobserwowaną poprawę jakości działania w dostępnych testach, chociaż zabrakło dokładnego sprawdzenie statystycznej istotności obserwowanych różnic, jakkolwiek systematycznie obecnych w różnych testach.

W literaturze zaproponowano szereg formalizmów do połączenie reprezentacji składnikowej i zależnościowej, w rozprawie głównie omówiono i przyjęto jako podstawę jeden konkretny zdefiniowany na potrzeby polskiego zasobu składniowego o nazwie Zależnica. Pokrótce przedstawiono również zasób składniowy TIGER dla języka niemieckiego o budowie podobnej do Zależnica. Natomiast pominięto kilka innych zasobów na różne sposoby odnoszących do siebie reprezentacje zależnościową i składnikową.

Definicję formalizmu i budowę zasobu należy traktować jako zadane czy też przyjęte z założenia w rozprawie i będące poza zakresem rozprawy. Warto jednak podkreślić, że zagadnienia te zostały przedstawione w bardzo przejrzysty sposób.

W odniesieniu do samej metody parsowania, zaproponowano rozwiązanie oparte na głębokiej sieci neuronowej, której jądrem jest struktura do oszacowania rozkładu prawdopodobieństwa binarnych relacji zależnościowych i która została rozszerzona o trzy klasyfikatory (warstwy neuronowe) określające własności danego segmentu jako elementu struktury składnikowej.

W przypadku relacji zależnościowych, zaproponowane rozwiązanie jest uproszczeniem rozwiązania znanego z polskiego parsera zależnościowego COMBO. Zmiany w architekturze sieci neuronowej są niewielkie i obejmują uproszczony mechanizm klasyfikacji relacji i funkcja straty oraz, przede wszystkim, wykorzystanie wstępnie trenowanego modelu językowego opartego na sieci typu BERT i jego dostrojenie do zadania wraz z całym rozwiązaniem. Przyniosło ono jednak bardzo dobre skutki. Zastosowany model językowy, szczególnie HerBERT, wstępnie wytrenowany na dużych zbiorach polskich tekstów, dostarczył bogatej wstępnej wiedzy o własnościach leksykalno-strukturalnych polskich tekstów. Zaowocowało to wyższą skutecznością parsera Doktorantki w testach w porównaniu do COMBO w przypadku parsowania zależnościowego, pomimo uproszczenia rozwiązania w stosunku do COMBO.

Wprowadzenie dodatkowych trzech klasyfikatorów do rozpoznawania niezbędnej informacji o strukturze składnikowej, w tym tzw. kręgosłupów składniowych, czyli ścieżek w grafie wyznaczających rodzaj elementów głównych, trenowanych w oparciu o tę samą wyjściową reprezentację tekstu, ale w dużej mierze niezależnie od pozostałych komponentów, wydaje się być rozwiązaniem dosyć bezpośrednim i konserwatywnym w pewnym sensie. Brakuje próby ich sprzężenia razem w ramach struktury sieci neuronowej, np. w procesie uczenia. Estymacje prawdopodobieństwa decyzji generowane przez te komponenty oraz przez komponent zależnościowy są łączone razem dopiero poprzez heurystyczny i deterministyczny algorytm konstrukcji hybrydowego drzewa.

Również dodatkowe moduły do segmentacji poziomu wyrazowego, tagowania i lematyzacji oraz rozpoznawania wystąpień nazw własnych zostały zaproponowane w postaci dodatkowych warstw klasyfikujących dodanych do wspólnej bazy reprezentacji tekstu w oparciu o sieć typu BERT.

Podsumowując, w rozprawie zaproponowano dosyć klasyczne rozwinięcie znanej już architektury parsera zależnościowego o dodatkowe warstwy klasyfikacji, co doprowadziło jednak do satysfakcjonujących rezultatów. Zarówno w parsowaniu jak i w większości zadań częściowych są to wyniki przewyższającego dotychczasowy poziom dla języka polskiego. W tym kontekście pewna prostota tego rozwiązania jest jego zaletą.

Skuteczność, kompleksowość i systematyczność zaproponowanego rozwiązania stanowią o jego wartości jako wkładu w rozwój dziedziny. Warto też zauważyć, że rozwiązanie to pokazało też swoją wartość w łatwym dopasowaniu do analizy tekstów dawnych, gdzie ilość danych treningowych jest ograniczona.

W porównaniu do stanu badań na świecie zaproponowane metody wpisują się w znane schematy, jednak duża wartość rozprawy leży w sukcesie kompletnego rozwiązania jakie dostarcza.

4. Poprawność i jakość badań

Rozprawa prezentuje bardzo dobry poziom warsztatu naukowego. Przyjęto właściwy plan prac badawczo-rozwojowych. Punktem wyjścia była analiza dostępnych anotowanych danych i stanu badań w zakresie metod parsowania oraz modelowania języka polskiego. Zaproponowano dostosowanie

istniejącego algorytmu parsowania zależnościowego do innych modeli językowych oraz jego uproszczenie i rozszerzenie o szereg dodatkowych zadań. Następnie bardzo dobrze zaplanowano i przeprowadzono szereg badań eksperymentalnych, w tym na korpusach tekstów dawnych. Całość procesu badawczo-implementacyjnego robi bardzo dobre wrażenie i osiągnięte wyniki są dobrze poparte badaniami eksperymentalnymi. Niemniej, przy wnikliwej analizie można sformułować kilka uwag, które mogą być pomocne w dalszym toku prac badawczych i rozwoju parsera i całego systemu przetwarzania.

Praca jest napisana dobrze, czyta się ją z zainteresowanie, ale brakuje precyzji w opisie zaproponowanych rozwiązań. Przedstawiane są one w ogromnej większości przypadków w sposób opisowy, a co gorsza nawet w oparciu o wybrane przykłady. Świetnie, że przykłady są, ale brakuje formalnego podsumowania, np. precyzyjnej specyfikacji algorytmów oraz formalnego przedstawienia struktury sieci neuronowej. W tym drugim przypadku, brakuje dobrego diagramu całościowej struktury systemu przetwarzania, czyli parsera wraz z poszczególnymi modułami. Brakuje precyzyjnej specyfikacji procesu trenowania całości rozwiązania. Na przykład, bardzo mało uwagi poświęca się kwestii definicji funkcji straty, która często nie jest nawet jawnie wyspecyfikowana. Tymczasem jest to podstawowy mechanizm do ukierunkowania procesu uczenia. To wszystko sprawia, że rodzi się szereg pytań, na które trudno znaleźć odpowiedź w tekście.

Czy dla wszystkich zadań i dodawanych modułów był dotrenowywany wykorzystany podstawowy model językowy, tzn. w jakim zakresie jego wagi były odmrożone podczas trenowania całości?

Czy moduł segmentacji był (podrozdz. 3.4.1) był trenowany razem z parserem i tagerem, a może też z innymi modułami? Jeżeli nie, to jaki inny schemat treningu został przyjęty?

Podobnych pytań można sformułować więcej. Oczywiście są udostępnione kody źródłowe i całość rozwiązania jest otwarta – co jest bardzo cennym osiągnięciem rozprawy – ale kody mogą się różnić w szczegółach od głównych idei, a znajomość przyjętych założeń pomaga zrozumieć implementację.

Ten brak rygoru w opisie metod utrudnia replikowalność badań.

Zaproponowane podejście jest pewną realizacją schematu uczenia wielozadaniowego (ang. Multi-task Learning). Niestety jest to obecne jedynie śladowo. Brakuje zarówno jawnego odniesienia do tego schematu i literatury przedmiotu, jak również głębszej eksploracji możliwości jakie on daje w odniesieniu do trenowania rozwiązania w oparciu o głęboką sieć neuronową i wzajemnych powiązań pomiędzy zadaniami cząstkowymi.

W wyraźny sposób brakuje dobrego porównania zaproponowanej metody parsowania do aktualnego stanu badań na świecie. Zaprezentowane porównanie w zakresie algorytmu koncentruje się na jedynie na porównaniu do parsera COMBO, który był w pewnym stopniu punktem wyjścia. Jednak i to porównanie jest dosyć powierzchniowe, bo w analizie wyników nie wzięto pod uwagę wielu różnic w zakresie konstrukcji obu rozwiązań, np. jaki jest wpływ uproszczeń w stosunku do oryginalnej konstrukcji COMBO, szczególnie jak to się objawia w zakresie dodatkowych zadań.

Poza jednak COMBO, na świecie zaproponowano kilka innych podejść do parsowania hybrydowego i nawet na etapie przeglądu literatury zostały one potraktowane bardzo skrótowo. W odniesieniu do poszczególnych zadań cząstkowych, jak parsowanie zależnościowe czy tagowanie, dokonano eksperymentalnego porównania, ale raczej z popularnymi bibliotekami, czyli rozwiązaniami technologicznymi. Z drugiej strony w kilku testach przyjęto bardzo dobrą praktykę własnoręcznego przetestowania poszczególnych rozwiązań i zweryfikowania ich skuteczności w stosunku do wartości podawanych w literaturze.

Bardzo brakuje porównania z technologią opartą na dużych generatywnych modelach językowych (GMJ, ang. LLM). Nie można ich ignorować, stały się obecnie podstawową technologią. Pewnym wyzwaniem jest fakt, iż wszystkie dane użyte w eksperymentach są publicznie dostępne i jest duże prawdopodobieństwo, że zostały użyte do treningów GMJ (ang. LLM). Z drugiej strony podobny problem mamy w przypadku niektórych modeli typu BERT, np. podczas trenowania HerBERT-a użyto znacznej części danych, na których były prowadzone eksperymenty w rozprawie. Fakt ten nie został uwzględniony w analizie.

Szkoda, że w rozprawie pominięto badanie niezawodności i efektywności zbudowanego systemu, podczas gdy aspekt praktyczny jest bardzo pozytywnym wyróżnikiem przedstawionych osiągnięć. Na przykład wspominany parser COMBO zawsze miał problemy z efektywnością obliczeniową.

Podczas ewaluacji przeprowadzono szereg interesujących testów porównawczych, np. dotyczących wykorzystywanych modeli lub kombinacji zbiorów danych, jednak praktycznie zabrakło analizy błędów. Analiza wyników skupia się na przedstawieniu wartości miar (chociaż nigdzie nie zweryfikowano statystycznej istotności różnic). Brakuje próby spojrzenia na to, co się za różnicami w wynikach kryje.

Przeprowadzono bardzo interesujące eksperymenty z zastosowaniem systemu do tekstów dawnych, co pokazało wysoką wartość rozwiązania, ale zabrakło testów, szczególnie parsera, na tekstach pochodzących ze swobodnej komunikacji ludzi (ang. CMC – Computer Mediated Communication), w tym z mediów społecznościowych. Teksty takie są pełne języka niestandardowego, zanieczyszczonego oraz błędów.

Rozprawa prezentuje wysoka jakość od strony językowej i edycyjnej. Występuje tylko kilka drobnych błędów wskazanych w następnej sekcji, jako pomoc w ewentualnej dalszej edycji tekstu.

W niektórych częściach rozprawy, np. str. 60, Doktorantka zmienia formę na częste użycie pierwszej osoby. Nie jest to oczywiście błąd, a jedynie jest odbiega od typowego stylu.

5. Wiedza Kandydatki

W ramach przeglądu literatury, rozdziały dotyczące metod parsowania, np. parsowania składnikowego są bardzo syntetyczne. Szczególnie jest to widoczne w przypadku omówienia podejść łączących parsowanie składnikowe i zależnościowe w podrozdziale 2.3, składającego się z kilku zdań, gdzie są wzmiankowane aktualne prace z literatury światowej.

Kilka merytorycznych ograniczeń pracy zostało również wskazanych we wcześniejszych częściach recenzji.

Za to Kandydatka wykazała się bogatą wiedzą w wielu aspektach językoznawstwa komputerowego i korpusowego. Również rozprawa jako całość jest bardzo dobrze zaplanowana i napisana.

Można uznać, że ogólna wiedza Kandydatki w dziedzinie prezentuje się dobrze.

6. Komentarze szczegółowe

Str. 13: Przytoczona definicja składni jest bardzo skrótowa i np. nie próbuje określić relacji pomiędzy składnią i semantyką, co w dalszej części rozdz. 1.1 nie zostało również podjęte.

Str. 43: „na wyuczaniu modeli zdolnych do produkowania ‘’na bieżąco’’ reprezentacji słów” – w odniesieniu do modeli typu ELMo jest to mylące lub dosyć metaforyczne stwierdzenie, nie są to modele generatywne

Str. 58, 3.2.1: brak jasnej specyfikacji wejścia, postać danych wejściowych nie „jest prosta", czy są segmenty to kwestia wyboru, Tab. 3.1 to tylko jeden przykład, a nie specyfikacja, itd.
Str. 58, Tab. 3.1: format wygląda na rozszerzenie formatu CoNNL – brak cytowania.
Str. 59: „Etykieta węzła" – w tym momencie to nie jest zdefiniowane pojęcie.
Str. 59: „losowo usuniętych" – raczej zamaskowanych, pozycje nie zostały usunięte.
Str. 59: „pozbawić ostatnich warstw" – dlaczego tylko ostatnich, to kwestia wyboru strategii dostrajania, można również rozszerzyć o dodatkowe warstwy.
Str. 60: „modele językowe" – brakuje definicji (sic!), dalej zawężone tylko do „architektury typu encoder" – nieprecyzyjne i dodatkowo nadmierne.
Str. 60: „odczytaniu napisu" – delikatnie mówiąc, bardzo metaforyczne określenie.
Str. 60: „prompcie" – anglicyzm.
Str. 60: „Skala i specyfikacja dużych modeli językowych
Str. 61: rys. 3.3 nie przedstawia dobrze struktury sieci, pomijając już kwestię braku modułów dodatkowych; dalsza część jest na rys. 3.4, ale oba nie są dobrze sparowane i nadal brakuje istotnego połączenia.
Str. 64: czy reguły przedstawione opisowo zostały zapisane w sposób deklaratywny, czy też zostały zaimplementowane bezpośrednio w kodzie programu?
Str. 65: brakuje poprawnego cytowania TensoFlow (por. https://www.tensorflow.org/about/bib?hl=pl)
Str. 66: Entropia krzyżowa jako funkcja straty to tylko jedna najbardziej podstawowa opcja, w całej pracy brakuje głębszej uwagi dla zagadnienia funkcji straty, szczególnie w kontekście uczenia wielozadaniowego.
Str. 67: „sekwencyj" – zamierzone?
Str. 68: „Wszystkie wybrane moduły […], trenowana są łącznie […]" – sugeruje, że model bazowy był w pełni dotrenowywany wraz z całością sieci, ale ta uwaga pojawia się dopiero tutaj i brakuje jej precyzji.
Str. 71: bardzo podobny zapis reguł lematyzacji pojawił się już w narzędziu Odgadywacz (Piasecki i Radziszewski, 2008) do predykcji opisów morfologicznych.
Str. 76: atomowa reprezentacja zagnieżdżonych nazw własnych to tylko jedna z możliwości, brakuje odniesienia do literatury, np. Sławomir Dadas przedstawił w doktoracie ciekawe rozwiązanie.
Str. 79: „Sprowadzenie parsowania", „taką etykiety" – literówki.
Str. 83: „można spodziewać się, że parser będzie „znał'' schemat odmiany" – należałoby to zmierzyć eksperymentalnie
Str. 86: 4.2.1 Miary – ponownie definicje poprzez przykład i opisowo.
Str. 97: w wielu miejscach, ale szczególnie tutaj brakuje analizy statystycznej istotności zaobserwowanych różnic w wynikach.
Str. 140: „polish – comparison" – !

7. Konkluzja

Biorąc pod uwagę opinie zaprezentowane w poprzednich punktach i wymagania zdefiniowane przez art. 187 Ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (z późniejszymi zmianami) [1] moja ocena rozprawy pod względem trzech podstawowych kryteriów jest następująca:

[1] http://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20190000276

- rozprawa jako całość i w zakresie wkładu Kandydatki prezentuje oryginalne rozwiązanie problemu naukowego,
- Kandydatka posiada ogólną wiedzę teoretyczną w dyscyplinie informatyka i telekomunikacja,
- oraz Kandydatka wykazała się umiejętnością samodzielnego prowadzenia pracy naukowej, w szczególności w zakresie określenia celów naukowych i sformułowania planu badawczego.

Podsumowując, rozprawa spełnia wszystkie formalne i zwyczajowe wymagania stawiane rozprawom doktorskim i wnioskuję o dopuszczenie Doktorantki do dalszych faz przewodu doktorskiego.

Podpis