

Recenzja rozprawy doktorskiej mgr Jakuba Pokrywki pt. „Optimization and Evaluation in Machine Learning Challenges” – polski tytuł „Optymalizacja i ewaluacja w wyzwaniach uczenia maszynowego”

1. Uwagi ogólne nt. problematyki rozprawy i jej celów

Niniejsza recenzja jest sporządzona w odpowiedzi na pismo Przewodniczącego Rady Dyscypliny Informatyka Uniwersytetu Adama Mickiewicza w Poznaniu, profesora Krzysztofa Dyczkowskiego.

Rozprawa dotyczy problematyki stosowania metod uczenia maszynowego (w większości wykorzystujących głębokie uczenie się) oraz przetwarzania wstępnego danych tekstowych lub obrazowych w zakresie rozwiązywania konkursów ogłaszanych przy różnych konferencjach. Przedstawiona rozprawa jest zbiorem 7 publikacji konferencyjnych (trzy samodzielne oraz cztery współautorskie) z bardzo krótkim wstępem – formę i struktury tego „oprawionego maszynopisu” omawiam w kolejnym punkcie recenzji.

Odnosząc się do wymagań ustawowych oraz tzw. zwyczajowych (regulowanych przez niektóre Rady Dyscyplin lub Naukowych) rozprawa doktorska nie musi być typowym maszynopisem lecz może być przedstawiona w formie powiązanego tematycznie zbioru artykułów (opublikowanych lub przyjętych do druku w czasopiśmie naukowych, określonych przez Ministra właściwego do spraw nauki) – art. 187. Ponadto inny artykuł ustawy stwierdza, że doktor powinien mieć choć jeden artykuł opublikowany w czasopiśmie naukowym lub w recenzowanych materiałach z konferencji międzynarodowej, które w roku opublikowania artykułu w ostatecznej formie były ujęte w wykazie sporządzonym przez Ministra – co jest w przypadku recenzowanej rozprawy spełnione. W momencie pisania niniejszej recenzji nie są mi znane inne specjalistyczne wymagania lokalnej Rady Dyscypliny (np. niektóre Rady doprecyzowują wymagania do umieszczenia w cyklu także minimalnej liczby publikacji w czasopiśmie naukowych z tzw. listy ministerialnej).

Niewątpliwie przedstawione prace konferencyjnie spełniają wymóg powiązania tematycznego – są rodzajem opisu udziału doktoranta w różnych konkursach uczenia maszynowego (w zakresie analizy tekstów lub rozpoznawania obrazów) organizowanych w ramach kilku międzynarodowych konferencji. Głównym hasłem przewodnim tej rozprawy jest konkurs uczenia maszynowego (z ang. machine learning challenge). Jak sam pan mgr Pokrywka go definiuje (str 1 wstępu) jest to wg. Niego wybrany zbiór danych, wskazane miary oceny oraz zdefiniowanie zasad testowania rozwiązania. Dalej doktorant doprecyzowuje, że w przypadku zbiorów danych jest niezwykle ważne rozróżnianie podziałów danych na części uczące, testowe i ew. inne udostępnione do tzw. dodatkowego dostrajania modelu uczącego się. W przypadku zdefiniowania zasad konkursu doktorant wskazuje na reguły ograniczeń stosowania dodatkowych danych lub innego rodzaju ograniczeń charakterystycznych dla wybranego konkursu.

Zastanawiając się nad sensownością zdefiniowania problematyki badawczej wokół udziału oraz także projektowania konkursów uczenia maszynowego w powyższym

zakresie mogą sformułować argumenty za i przeciw – co wpływa także na późniejszą ocenę dorobku doktoranta.

Niewątpliwie takie konkursy (jak bym w odróżnieniu od doktoranta, który moim zdaniem błędnie używa zbyt często słowa wyzwania) stały się od jeszcze poprzedniego wieku popularne w środowisku uczenia maszynowego, sztucznej inteligencji oraz pokrewnych dziedzin, w szczególności wśród tzw. młodych badaczy. Były i są organizowane przy konferencjach naukowych (np. historycznie ACM KDD) lecz także w specjalizowanych platformach jak Kaggle lub także oferowane przez tzw. środowisko biznesowe. Sam pan mgr J. Pokrywka w wstępie do swojej rozprawy (strona 1 oraz 2) odnosi się do różnych ciekawych konkursów z zakresu języka naturalnego lub rozpoznawania obrazów.

Ja osobiście zwróciłbym uwagę, że pozwala to bardziej „praktycznie” sprawdzić się badaczom w rozwiązywaniu mniej lub bardziej trudnych i często dość praktycznych zadań (tak jak w typowym konkursie w dowolnej dziedzinie), w tym osiągnięcie pewnych celów, takich jak np.:

- Budowanie tzw. portfolio projektów (często związanych z dość praktycznymi, rzeczywistymi problemami zastosowań), w którym można zaprezentować swoje umiejętności.
- Porównywanie swojego rozwiązania z innymi, potencjalnie dość konkurencyjnymi propozycjami.
- Porównywanie swoich umiejętności z innymi osobami w danym środowisku (często rozważany problem wymaga tzw. dość specyficznych operacji na danych lub eksperymentalnego strojenia parametrów, które jest raczej rodzajem sztuki albo masowego testowania różnych parametrów lub konfiguracji proponowanego rozwiązania a mniej systematycznych podejść).
- Wykorzystanie swoich umiejętności do rozwiązywania rzeczywistych problemów związanych z nauką o danych lub zdobywanie nowych umiejętności w obszarze danej dziedziny.
- Potencjalnie wygranie nagród (czasami honorowych, czasami pieniężnymi lub ułatwiającym udział w danej konferencji),

Chciałbym zacytować wypowiedź p. Ray’a Heberera sprzed paru lat, który podsumowywał takie konkursy jako „in machine learning competition (..) population of algorithms compete against one another, mixing and refining characteristics of the most successful of their predecessors, and occasionally introducing novel approaches.” co jest może kontrowersyjnym stwierdzeniem, lecz równocześnie wskazuje to na pewną wątpliwość – na ile w rozwiązywaniu konkursów pojawiają się rzeczywiście nowatorskie propozycje metodologicznie, a na ile dominujące działania i osiągnięcia są jednak w mniejszym stopniu typową działalnością badawczą -- jak próbowałem scharakteryzowałem powyżej.

Odnosząc się do tekstu wprowadzenia do maszynopisu (str 2 dół) pan mgr J. Pokrywka stwierdził wprost, że jego praca ma “umożliwienie postępu w dziedzinie uczenia maszynowego (ML) poprzez rozwój metodologii przygotowywania wyzwań ML i opracowywania dla nich rozwiązań (z j. ang. enable progress in the ML field by advancing the methodology of preparing ML challenges and developing solutions for them)”. Podobnie w punkcie 1.2 doktorant opisał zakres swojej dysertacji – siedem artykułów dotyczy przygotowywania konkursów ML i tworzenia ich rozwiązań.

W mojej opinii jest to zbyt szeroko i niestety ryzykownie sformułowana motywacja oraz ew. cele (choć doktorant jawnie ich nie podał) rozprawy. Ponadto w dalszej dyskusji wyrażę wątpliwości co do możliwości zaproponowania tzw. metodologii naukowej.

Sam tekst istotnego wprowadzenia jest raczej napisany niejasno i bardziej opisuje konkretne konkursy i prace z tym związane. Osobiście oczekiwałem precyzyjnie zdefiniowanego problemu naukowego (najlepiej jednego problem, czego oczekuje się od rozpraw doktorskich), motywacji krytycznych (tzw. braku – z j. ang. gap - w aktualnym stanie wiedzy), hipotez oraz celów bardziej konkretnych do osiągnięcia.

2. Ocena formy i struktury rozprawy

Przekazany maszynopis rozprawy jest napisany w języku angielskim i obejmuje kopie 7 artykułów (materiałów konferencyjnych), które stanowią osiągnięcie doktoranta. Nie wymieniam dokładnie ich tytułów oraz miejsc publikacji – są one zastawione na stronie 3 maszynopisu rozprawy wraz z informacjami o tzw. punktach z listy ministerialnej. Wszystkie teksty umieszczone są w materiałach konferencji z lat 2022 oraz 2023 – w tym dwóch dominujących – 2022 IEEE Inst. Conference on Big Data oraz FedCSIS. Każda z tych prac związana jest z konkretnym konkursem organizowanym przy tych konferencjach – co zbiorczo opisane jest w rozdziale 2 zatytułowanym „research papers overview”. Ponadto maszynopis zawiera krótki rozdział wstępny (tzw. wprowadzanie wraz z listą odwołań do literatury), spis treści, streszczenia polsko i anglojęzyczne, podziękowania dla współpracowników oraz na końcu kopie deklaracji wkładu do 4 prac współautorskich podpisanych przez współautorów. Ten ostatni składnik jest ważny, gdyż specyfikuje co wniósł każdy z autorów, choć można było też podać udział procentowy tak jak w przypadku dokumentacji habilitacji – wtedy ułatwiło by to ocenę w których publikacjach p. mgr Jakub Pokrywka był wiodącym autorem. W przypadku pozostałych 3 artykułów doktorant był jedynym autorem. Z opisu zawartego w rozdziale drugim wynika, że większość z przedstawionych prac miała status tzw. ustnej prezentacji – przy czym należałoby dokładniej sprawdzić materiały programu danej konferencji, po to aby sprawdzić w jakiej formie się to odbyło.

W mojej ocenie taka organizacja maszynopisu może być zastosowana dla rozprawy doktorskiej, która jest cyklem powiązanych ze sobą opublikowanych artykułów.

Niestety nie jestem zadowolony z obecnego kształtu rozdziału wstępnego, który jest zdecydowanie zbyt krótki i niejasno definiujący problem badawczy oraz cele rozprawy. Powrócę to tych wątpliwości w punkcie 4tym mojej recenzji. Muszę jednak stwierdzić, że nie są mi znane dokładne wytyczne dla dyscypliny naukowej informatyka na UAM wobec takich prac, więc najprawdopodobniej doktorant i jego promotor mieli dużą wolność twórczą w takiej konstrukcji. Mogę tylko stwierdzić, że w przypadku Rady Dyscypliny informatyki technicznej i telekomunikacji (i historycznie poprzedniej dyscypliny w ramach dawnej Rady Wydziału Informatyki Politechniki Poznańskiej) formułowano bardziej szczegółowe wymagania dla doktoratów zestawionych jako cykl powiązanych publikacji i oczekiwania do wstępu oraz ew. rozdziału pokazującego własne dokonania doktoranta na tle aktualnego stanu wiedzy, literatury w danym obszarze. Są one jednak precyzyjniejsze, obszerniejsze i recenzowany maszynopis ich niestety nie spełnia.

W związku z tym, że artykuły były już recenzowane – choć selektywność zależy od konkursu i rodzaju konferencji – nie zamierzam oceniać ich formy i języka angielskiego. W ogólności sposób ich pisanie jest typowy oraz forma w miarę czytelna.

Należy jednak zwrócić uwagę, że wg. serwisów dblp są to wszystkie dotychczas opublikowane prace doktoranta (dokładnie 7dem) lecz serwis ResearchGate może wskazywać, że już po złożeniu manuskryptu pojawiły się na ostatniej wrześniowej konferencji FedCSIS nowe prace współautorskie – także częściowo związane z konkursami.

3. Ocena dorobku doktoranta

Po zapoznaniu się z maszynopisem, w szczególności przeczytaniu wskazanych artykułów sądzę, że osiągnięcia pana mgr Jakuba Pokrywki można w ogólności określić jako przedstawienie różnych technik przetwarzania wstępnego oraz analizy danych w konkretnym konkursie, doboru różnym modeli uczenia maszynowego, często wykonywania intensywnych eksperymentów w zakresie dostrojenia ich parametrów dla osiągnięcia możliwie wysokich wartości wskazanych miar oceny w danym konkursie. Ponadto czasami wykonano eksperymenty porównawcze różnych metod uczenia maszynowego.

Cztery pierwsze artykuły dotyczą konkursów związanych z analizą danych tekstowych i przetwarzania języka naturalnego a trzy ostatnie obrazów (klasyfikacji, rozpoznawania lub analizy kodowania dla prywatności). W większości rozważanych propozycji doktorant (czasami ze współpracownikami) rozważał modele językowe (najczęściej neuronowe sieci transformerów) dla zadań tekstowych lub lingwistycznych oraz sieci neuronowe typu YOLO dla zadań obrazowych. Tylko w ostatniej pracy wykorzystano drzewa wzmacniane gradientowo. Mogę ogólnie stwierdzić, że doktorant wykorzystuje nowoczesne i popularne metody uczenia głębokich sieci neuronowych na ogół połączone z mniej lub bardziej typową inżynierią cech, lub przetwarzaniem danych, w tym tzw. rozszerzaniem danych obrazowych (ang. data augmentation) – co pokazuje jego znajomość współczesnych rozwiązań w zakresie uczenia maszynowego oraz chyba sprawne wykorzystywanie i rozszerzanie ich dostępnych implementacji.

Z punktu widzenia konkursów – w niektórych z nich doktorant (lub jego zespół) osiągał wysokie miejsca.

Z drugiej strony konkretne rozwiązania są bardzo dedykowane pod zadania konkretnego konkursu – czyli w praktyce wybranego artykułu i nie przenoszą się pomiędzy nimi. Poniżej bardzo krótko charakteryzuje kolejne z nich:
Pierwsza z prac zatytułowana z j. ang. „Challenging America Modeling language in longer time scales” opisuje bardzo ciekawy problem konkursowy związany z analizą wielkiej kolekcji amerykańskich tekstów prasowych z długiego horyzontu czasowego – trzy różne podzadania – tutaj zgodnie z tekstem artykułu widać też wysiłek zespołu autorów w odpowiednie wykorzystanie tych danych (ze specjalizowanym przetwarzaniem wstępnym oraz odpowiednimi podziałami zbioru na części) na ich autorskiej platformie konkursowej Gonito. W zależności od rodzaju zadania trzeba było dobrać odpowiednią miarę oceny, co dyskutuje się w artykule. Z uwagi na zbadanie różnych metod uczenia maszynowego wskazuje się na dobre dostrojenie modelu językowego sieci transformerów RoBERTa, zwłaszcza w jego wersji z

aspektem czasu. Modele językowe są standardowe, ale ich dotrenowanie i dostrojenie są wkładem autorów. Może to stanowić tzw. silny punkt odniesienia (z j. ang. baseline) dla porównywania się z innymi potencjalnie proponowanymi metodami.

Drugi z artykułów opisuje ciekawy problem kategoryzacji irlandzkich –angielskich tekstów prasowych z uwzględnieniem aspektu czasowego. Zadania są ciekawe, a dokonania własne autorów artykułu to odpowiednie stworzenie / dostosowanie lub redefinicja korpusów tekstów pochodzących z oryginalnego źródła – danych do zadań konkursu. Metodologicznie autorzy rozważali różne sposoby wprowadzanie elementu czasowego i włączenie go do uczenia neuronowych modeli językowych typu RoBERTa - czyli odmiany czasowej sieci transformerów poprzez wprowadzanie tokenów (rok, miesiąc, dzień lub pora tygodnia) - co wiązało się także z jej eksperymentalnym dostrajaniem parametrów. Praca zawiera wiele wyników eksperymentów badających tego typu modele w różnych wersjach wprowadzenie takich elementów do modelu.

Trzeci artykuł opisuje bardzo nietypowe zastosowanie tzw. kart powtórkowych w systemach uczenia się języka obcego w środowiskach tzw. Supermemo. Znowu wybrane zastosowanie jest ciekawe i powiązane z bardzo praktycznym problemem oszacowania czasu potrzebnego do powtarzania materiału, po to aby przyspieszyć efektywność uczenia, oryginalnie w Supermemo stosuje się wykładniczą krzywą zapominania – opis w artykule jest ciekawy. Autorzy artykułu opisali eksperymenty porównawcze kilku znanych algorytmów uczenia (od nietypowej regresji, różnych sieci neuronowych do modelu sieci LSTM rozszerzonego o wykładnicze zapominanie). Z punktu widzenia lansowanej miary oceny Macro- uśrednianie błędów predykcji MacroAvgMAE ten ostatni model działał najsukuteczniej.

Ostatni z lingwistycznych artykułów, tj. Using transformer models for gender attribution in Polish, opisuje zadanie wykrywania płci autorów tekstów. Podobnie jak w niektórych z powyższych artykułów widać wysiłek autorów w stworzenie odpowiedniej wersji tego korpusu tekstów do konkursów, przygotowanie anotacji przez ludzkich ekspertów, dobór miary oceny – i to jest ta ciekawsza część tekstu. Pozostała zawiera wyniki eksperymentów porównawczych wielu klasyfikatorów ze wskazaniem na model językowy RoBERTa dla języka polskiego.

Kolejne trzy artykuły omawiają trzy różne konkursy związane z rozpoznawaniem obrazów – dwa pierwsze mają odniesienie do tzw. zastosowań drogowych i wykorzystania sieci splotowych typu YOLOv5.

W przypadku piątego artykułu – zadania polegają na wykryciu na zdjęciach samochodu, przewidywaniu jego klasy (typu) oraz orientacji (czy jest przodem, tyłem). Specyfika tego konkursu z konferencji IEEE Big Data wiązała się też z uczeniem się z danych symulowanych, a testowanie na danych rzeczywistych. Wybrane rozwiązanie jest zespołem dość znanych neuronowych sieci obrazowych YOLOv5l6 i YOLO5x6 z różnymi zestawami parametrów.

Artykuł szósty dotyczy rozpoznawania uszkodzeń dróg na podstawie zdjęć z komórek – smartfonów zamontowanych w samochodach. Zbiór danych jest silnie niezbalansowany, co w tym konkursie prowadzi do optymalizacji odmiany miary F1. Dość oryginalnym acz prostym koncepcyjnie osiągnięciem doktoranta jest

zapropozowanie znowu procesu uczenia zespołu sieci neuronowych YOLO i powiązania ich z rozszerzaniem danych, aby ograniczać koszty uczenia metod.

Ostatnia z przedstawionych prac rozważa bardzo nietypowy i ciekawy problem konkursowy IEEE Big Cup badania stopnia zaszyfrowania danych obrazowych (czy zakodowany obraz i jego odpowiedni oryginalne są podobne czy nie). W tekście artykułu autor zawarł dobre uzasadnienie praktyczne oraz ilustracje. Tutaj mamy zupełnie inną koncepcję rozwiązania – sensowne wykorzystanie analizy histogramów kolorów i stopni szarości, ekstrakcji cech. Jest to później przekazywane do standardowego algorytmu uczenia drzew wzmacnianych gradientów GBoost (implementacji CatBoost) – wyniki eksperymentalne są dość dobre - co w rezultacie pozwoliło na zajęcie wysokiego miejsca w tym konkursie.

W pierwszych artykułach wspomina się o wykorzystywaniu platformy Gonito (rozwijanej przez zespół lingwistyczny UAM pod kierownictwem promotora tj. profesora Gralińskiego) – niestety brak informacji o ewentualnym wkładzie doktoranta w jej rozwój a to akurat mogłoby być elementem oryginalnym, gdyż z opisu dostępnego w internecie by wynikało, że ta platforma i repozytorium typu github ma na celu wspomaganie realizacji konkursów.

4. Uwagi krytyczne i polemiczne

Pomimo interesujących propozycji rozwiązywania całkiem nietrywialnych zadań konkursowych (opisywanie problemy w większości przypadków mi się bardzo spodobały) formułuję kilka uwag krytycznych do treści rozprawy oraz pytania o charakterze dyskusji polemicznej z autorem, które powinny być przedmiotem dalszej dyskusji podczas samej obrony doktoratu:

Moja najważniejsza wątpliwość o charakterze metodologii prac naukowych dotyczy dość niejasnego i zbyt szerokiego ew. celu rozprawy. We wstępie do maszynopisu nie jest to precyzyjnie wyrażone. Na str 2 doktorant pisze, cyt z j.ang. "the motivation for this work is to enable progress in the ML field by advancing the methodology of preparing ML challenges and developing solution for them". W dwóch innych miejscach (str 3) pisze następnie (znowu cyt. z j. angielskiego "this work covers selected aspects of evaluation and optimization techniques for machine learning challenges". A w streszczeniu mamy wyraźnie napisane że "niniejsza praca doktorska opisuje aspekty metodyki tworzenia wyzwań uczenia maszynowego oraz techniki opracowania ich rozwiązań". Niestety wstęp jest tak napisany, że nie ma podanego jawnie celu pracy – ale sądzę że to są cele tej rozprawy. Obawiam się, że jako czytelnik nie zidentyfikowałem w przedstawionej rozprawie zaproponowania klarownej i silnej metodologii tworzenia i rozwiązywania konkursów uczenia maszynowego. Przypomnę definicję mówiącą, iż metodologia to nauka o metodach działalności naukowej i stosowanych w nauce procedurach badawczych; w aspekcie teoretycznym – nauka o elementach i strukturze systemów naukowych [Encyklopedia PWN]; Moglibyśmy to zawęzić to wąskiej metodologii w określonym podobszarze badań – nadal jednak powinniśmy oczekiwać szerszego i ogólniejszego schematu prowadzenia badań, czego ewidentnie nie zaprezentowano w tej rozprawie. Nawet jeśli ograniczymy się do metodyki postępowania to rozumiemy najczęściej jako zbiór zasad dotyczących sposobów postępowania przy wykonywaniu pewnej pracy lub zmierzając do danego celu (definicja podana przez K. Borzęckiego), to metodyka nie

zawiera rozważań teoretycznych (na określonej poddziedziną badawczą) lecz bardziej praktyczne dyrektyw i ew. objaśnienia, jak je poprawnie i skutecznie stosować w określonej klasie problemów naukowych – i przypuszczam że to rozumienie będzie bliższe treści rozprawy – lecz znowu niestety nie dostrzegam tego w recenzowanej rozprawie.

Dlatego mam pytanie do doktoranta – w jakim stopniu przedstawił w swojej rozprawie metodologię albo choć w miarę ogólną metodykę postępowania naukowego? – to jest związane z moimi dalszymi wątpliwościami, które precyzuje w dalszych punktach.

Czytając zawartość rozprawy doszedłem do przekonania że jest ona w zasadzie zbiorem dość niezależnych artykułów powiązanych wątkiem konkursów. Natomiast ich opisy są dość jednostkowe, oraz zastosowane rozwiązania zarówno do konstrukcji danych, jak ich wyboru metod są mocno dedykowane pod konkretne zadania i jako zbyt szczegółowe podejścia nie mają takiej siły uogólnienia, aby mówić o właśnie bardziej metodycznym nowatorskim podejściu, gdzie wyraźnie doktorant przedstawia ogólniejsze koncepcje i zasady postępowania.

Ponadto wątpliwości są powiązane ze zbyt ograniczoną nowatorskością omawianych metod uczenia maszynowego. W małej ilości artykułów dostrzegłem wyraźnie nowy innowacyjny pomysł w zakresie wprowadzenia nowych metod. Oczywiście dostrzegam pewne ciekawe własne pomysły, np. jak zrobić stopniowe rozszerzanie obrazów w iteracyjnej budowie zespołu sieci neuronowych YOLO w 8 artykule, które może prowadzić do zmniejszenia czasów obliczeń na kartach GPU albo sprytnego triku analizy histogramów obrazów zakodowanych w ostatnim z konkursów (artykuł 9ty). Lecz czytając pozostałe artykuły odniosłem wrażenie (przy zastrzeżeniu, że nie jest mocnym ekspertem w zadaniach NLP), że przedstawione sposoby rozwiązywania zadania wykorzystują znane poprzednio algorytmy, czasami chytrze połączone z elementami przetwarzania danych albo wplecenia do modelu informacji kontekstowej albo czasowej mocno związanej z konkretnym zadaniem. Wysiłek doktoranta albo jego współautorów jest wprawdzie widoczny w częściach eksperymentalnych lub dostrajania parametrów, zwłaszcza modeli neuronowych ale nie widzę tego jako mocnego i oryginalnego osiągnięcia naukowego i trudno mi się zgodzić z terminem że to jest osiągnięcie w zakresie optymalizacji. Co więcej wiele z nich są bardzo specyficzne dla konkretnego, często nietypowego zadania konkursowego i nie widzę prób uogólniania tego na coś bardziej metodycznie przydatnego dla szerszej klasy zadań. W tej chwili widzę tą rozprawę jako zbyt indywidualne spojrzenie na wiele specyficznych przypadków zadań, które się nie uogólniają. Wiele z tzw. kroków przygotowanie danych do konkursu z uwzględnianiem, np., czasu czy innych informacji przy podziale zbioru danych na części uczące lub testowe jest wcześniej znane albo w miarę naturalne. Dotyczy to także doboru wybranych miar oceny pod specyfikę danych (np. średnia geometryczna jest od dziesiątków lat stosowana dla oceny klasyfikatorów uczonych z niezbalansowanych danych) albo zadania – także nie zauważyłem nowatorskich propozycji. Prosiłbym doktoranta, aby spróbował określić co jest oryginalną i mocno nowatorską propozycją metod, algorytmów, którą wprowadził w rozprawie. Co jest w jego opinii tzw. mocnym wkładem metodologicznym do dziedziny informatyki, w zakresie uczenia maszynowego?

Niestety w mojej opinii rozdział wstępny nie jest dobrze napisany oraz wydaje się być zdecydowanie kontrowersyjny z metodologicznego punktu widzenia. W obecnej formie jest to dość krótkie wprowadzenie do idei konkursów w różnych poddziedzinach informatyki, w szczególności inżynierii lingwistycznej a później charakterystyka poszczególnych prac w cyklu, niestety w stylu „czego dotyczy konkurs i co autor w nim zrobił”. Zdecydowanie brakuje zarysowania czegoś jak motywacji krytycznej do badań naukowych (czego brakuje w obecnych zasadach realizacji oraz rozwiązywania konkursów, gdzie są nowe wyzwania – w prawdziwym tego słowa znaczeniu – ale badawcze). Doktorant nie napisał wyraźnie co jest jego problem naukowym, który zamierza rozwiązać i dlaczego? Jaki jest obecnie stan wiedzy na ten temat. Brakuje sformułowania hipotez naukowych do sposobu postępowania, i co jest najbardziej irytujące czytelnika jak ja, nie ma jasnych celów tej rozprawy.

Stylistycznie mam zastrzeżenia do części z tekstów w rozdziale 2gim. O ile jego sam zamysł jest dobry i pożyteczny dla czytelnika – zawiera podstawowe informacje o wybranym konkursie, artykule oraz często odnośniki do źródeł internetowych albo repozytoriów danych. O tyle opis dokonań doktoranta czasami stylistycznie budzi wątpliwości – zbyt często autor pisze że jego dokonaniem jest implementacja znanych już wcześniej metod + ew. przygotowanie danych – czasami jest to wszystko w opisie danego artykułu. Rodzi to wątpliwości czy pojawia się osiągnięcie naukowe w danej pracy – przynajmniej na miejscu autora bym bardziej starał się przekonać do „własnych propozycji”.

Większość z przedstawionych prac nie ma silnego podłoża teoretycznego. Jest silnie oparte na eksperymentach obliczeniowych, zwłaszcza przy doborze tzw. hiperparametrów sieci neuronowych lub porównywaniu różnych algorytmów lub ich wariantów. Niestety nie w każdym artykule zawarte są obszerne raporty z eksperymentów – często są to wybrane wyniki (oczywiście istnieją pewnie ograniczenia stron dla danych konferencji). Prawie we wszystkich artykułach brakuje informacji o kosztach obliczeniowych lub pamięciowych. Jest wprowadzone podane jawnie w zakończeniu artykułu ósmego (Efficient GPU ..) lecz też nie mamy tam podanego punktu odniesienia innych metod. Prosiłbym doktoranta o rozszerzenie informacji o przykładach rzeczywistych kosztów obliczeniowych choć niektórych z jego rozwiązań.

Być może wynika to ze specyfiki publikacji dla konkursu danej konferencji – lecz nie zawsze w tekście mogłem odnaleźć informacje o zestawieniu hiperparametrów badanych modeli (sieci neuronowych albo zespołów klasyfikatorów). Nie sądzę, abyśmy mieli dostateczne szczegółowe informacje dla odtwarzalności (tzw. ML reproducibility) prowadzonych eksperymentów pomimo, że jest to chyba jedna z motywacji ciekawej platformy Gonito rozwijanej na UAM.

W zakresie języka polskiego – które jest obecne prawie wyłącznie w streszczeniu rozprawy występują wyraźne uchybienia lub nieakceptowalne slangowe kalki językowe, np. „pozyskanie datasetu” albo „przygotowanie baseline”.

5. Konkluzja końcowa

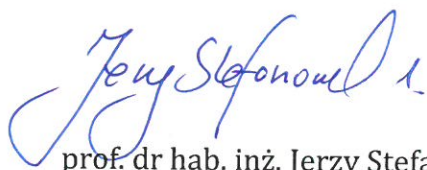
W przypadku tej rozprawy mam dość niejednoznaczne odczucia. Jak powyżej napisałem mam dużo zastrzeżeń metodologicznych oraz brak przekonania o tzw.

wkładzie własnym tj. zaproponowaniu przed doktoranta odpowiednio wiele i bardziej oryginalnych nowych metod rozwiązania wybranych problemów naukowych.

Częściowo może to wynikać z kontrowersyjnej specyfiki wyboru tematyki dość zróżnicowanych konkursów uczenia maszynowego gdzie zawartość prac to po prostu opis jak rozwiązano zadanie w danym konkursie (patrz też moje uwagi z punktu pierwszego recenzji). Przez to opisane dokonania doktoranta są bardzo szczegółowe i często dość specyficzne wykorzystujące mniej lub bardziej złożone połączenia na ogół znanych już wcześniej rozwiązań. Część tzw. osiągnięć w ramach hasła optymalizacji to intensywne obliczeniowo poszukiwanie dogodnych parametrów wybranych modeli – co także się nie uogólnia na szerszą metodologię rozwiązywania wybranej podklasy konkursów. Ponadto w mojej ocenie można mieć zastrzeżenia do niezbyt poprawnego przedstawienia swoich dokonań w zbyt ogólnikowym wstępie pracy – tutaj mój trudny proces decyzyjny wiązał się z zastanawianiem się, czy nie cofnąć maszynopisu do tzw. poprawek edycyjnych, aby przynajmniej wstęp do maszynopisu był napisany metodologicznie i treściowo dużo lepiej.

Z drugiej strony widać duży wysiłek autora w jak najsprawniejsze rozwiązanie konkretnych zadań w ramach wybranego konkursu – w niektórych przypadkach udokumentowanym odpowiednim miejscem osiągniętym w rozstrzygnięciu konkursu. Można stwierdzić, że w tych pracach nabył on szeroką wiedzę oraz umiejętności stosowania współczesnych metod uczenia maszynowego dla przetwarzania języka naturalnego oraz obrazów z wykorzystaniem głębokich sieci neuronowych. W części z artykułów doktorant wskazywał na duży wysiłek implementacyjny – choć nie powinno to być głównym elementem oceny dokonania naukowego w przypadku rozprawy doktorskiej. Ponadto sądzę, że niektóre z zaproponowanych specjalizowanych podejść mogą być przydatne w zastosowaniach praktycznych

Dlatego ostatecznie zdecydowałem się dać szansę do dalszej dyskusji naukowej i mam nadzieję na pojawienie się choć częściowej poprawy w prezentacji na obronie. W konsekwencji tych argumentów postuluję, że rozprawa pana mgra Jakuba Pokrywki może być dopuszczona do dalszych etapów postępowania, w tym publicznej obrony.



prof. dr hab. inż. Jerzy Stefanowski