Politechnika Łódzka
Instytut Informatyki

Łódź, 2 września 2024 roku

prof. dr hab. inż. Adam Wojciechowski
Instytut Informatyki
Wydział Fizyki Technicznej, Informatyki i Matematyki Stosowanej
Politechnika Łódzka
Al. Politechniki 8, 93-590 Łódź

RECENZJA ROZPRAWY DOKTORSKIEJ

Tytuł rozprawy: **Strategie postepowania w przypadku małej liczby obserwacji uczących w problemie klasyfikacji danych nieustrukturyzowanych**

Autor rozprawy: **mgr Dominik Lewy**

Promotor rozprawy: **prof. dr hab. inż. Jacek Mańdziuk**

Jakie zagadnienie naukowe jest rozpatrywane w pracy (teza rozprawy) i czy zostało ono dostatecznie jasno sformułowane przez Autora? Czy tematyka rozprawy jest aktualna lub dostatecznie ważna? Jaki charakter ma rozprawa (teoretyczny, doświadczalny, inny)?

Pan mgr Dominik Lewy podjął w swojej pracy doktorskiej zagadnienie wzbogacania (augmentacji) zbiorów danych w sytuacji niedoboru lub braku danych dla potrzeb uczenia maszynowego modeli klasyfikatorów. Rozwiązania dedykował w głównej mierze danym nieustrukturyzowanym obejmującym dwie modalności: tekstową i obrazową. Zagadnienie naukowe zostało bardzo precyzyjnie i klarownie przedstawione w rozprawie. Doktorant koncentruje się na strategiach wzbogacania danych inspirowanych mieszaniem, jak również na zastąpieniu mieszania poprzez wykorzystanie danych pozyskanych z Internetu. Szczegółowym celem badawczym było sprawdzenie czy algorytmy mieszania można przenosić na inne modalności, wykorzystując mechanizmy specyficzne dla danego rodzaju danych (np. korzystanie z mechanizmu uwagi dla przetwarzania tekstu naturalnego). W badaniach postawiono hipotezę, że metody wzbogacania zdjęć poprzez mieszanie są na tyle elastyczne, iż mogą być zastosowane w ramach wąskiego problemu jakim jest uczenie federacyjne jak i problemu klasyfikacji dla innej modalności. W kontekście ilościowym, zastosowanie tych metod wzbogacania powinno przyczynić się do wzrostu skuteczności modeli uczonych wzbogaconymi danymi, względem modeli uczonych bez wzbogacania. Hipoteza zakłada też alternatywizację wzbogacania danych poprzez zastosowanie dużej ilości zaszumionych danych pobranych z Internetu.

Tematyka rozprawy jest bardzo aktualna, ważna i dobrze postawiona. Dynamiczny rozwój sztucznej inteligencji oraz jej zastosowań jest w głównej mierze uzależniony od danych, do których dostęp jest bardzo trudny – zarówno pod względem wolumenów danych, jak i ograniczeń formalnych w ich przetwarzaniu. Tym samym strategie uczenia maszynowego w przypadku małej liczby nieustrukturyzowanych obserwacji stanowią wyzwanie dla współczesnego świata.

Rozprawa ma charakter teoretyczno-eksperymentalny, ale nosi w sobie istotny potencjał praktyczny. Przedstawione w pracy rozważania koncentrują się na systemowym wzbogacaniu danych lub ich suplementacji z Internetu, celem podniesienia skuteczności modeli klasyfikacyjnych. Przedstawione analizy są mocno pogłębione i bardzo profesjonalne. Dodatkowo, analiza wyników jest mocno inspirująca i otwiera wiele kontekstów dla kolejnych badań. Praca z racji podjętej problematyki,

zakresu oraz metodologii badawczej jest pełnoprawnym osiągnięciem o charakterze naukowym w dziedzinie nauk inżynieryjno-technicznych, w dyscyplinie Informatyka Techniczna i Telekomunikacja.

Czy w rozprawie przeprowadzono w sposób właściwy analizę źródeł (w tym literatury światowej i stanu zagadnień w przemyśle) świadczącą o dostatecznej wiedzy Autora? Czy wnioski z przeglądu źródeł sformułowano w sposób jasny i przekonywujący?

W przedstawionej do oceny rozprawie doktorskiej, odniesienia do literatury pojawiają się głównie w rozdziale drugim, trzecim i czwartym. Początkowo Doktorant usystematyzował kontekst metod i eksperymentów oraz wprowadził kluczową terminologię sprofilowaną na klasyfikację zbalansowanych, lecz nieustrukturyzowanych zbiorów danych o różnych modalnościach. W dalszej kolejności (rozdział 3) przybliżył wybrane strategie rozszerzania danych, aby ostatecznie przeanalizować aktualny stan literatury wzbogacania danych (rozdział 4), w kontekście dwóch podstawowych modalności: danych obrazowych i danych tekstowych. Przegląd jest merytoryczny i aktualny. Metody literaturowe zostały poprawnie omówione oraz uzupełnione cennymi schematami, które pozwalają lepiej zrozumieć wartość dodaną rozwiązań Doktoranta przedstawionych w dalszej części pracy.

Bibliografia jest obszerna i liczy 119 pozycji, w tym 3 prace opublikowane z udziałem Doktoranta oraz jeden *preprint*. Prace ukazały się w latach 2022-2023, a Doktorant jest ich pierwszym autorem – w przeważającej większości wyłącznie we współautorstwie z Promotorem. Są to zarówno prace w renomowanych czasopismach z listy *JCR* tj. *Artificial Intelligence Review* (IF=10.7), *Journal of Artificial Intelligence and Soft Computing Research* (IF=3.3), jak i publikacja na uznanej konferencji międzynarodowej *International Conference on Neural Information Processing (ICONIP)*. Całokształt dokonań publikacyjnych Doktoranta jest pokaźny i również stanowi cenne źródło odniesień do literatury światowej w przedmiocie badań.

Czy Autor rozwiązał postawione zagadnienie, czy użył właściwej do tego metody i czy przyjęte założenia są uzasadnione?

W obszarze podjętej problematyki, Doktorant opracował zbiór autorskich metod i przeprowadził serię eksperymentów, każdorazowo rzetelnie dokumentując prowadzone badania poprawnie opracowanymi wynikami i odpowiednią dyskusją. Docenienia wymaga schemat opisu eksperymentów i forma dyskusji, która obejmuje pseudokod i/lub schemat algorytmu, metodologię eksperymentu, pytania badawcze i dyskusję wyników w kontekście tych pytań. Wnioski są spójne z wynikami eksperymentu i wzmacniają kontrybucje Doktoranta.

W pierwszej kolejności Doktorant zaprezentował metodę o akronimie *AttentionMix*, która jest rozwiązaniem inspirowanym mieszaniem obrazów, ale zaadaptowanym do pracy z tekstem, dzięki zastosowaniu wektorowej reprezentacji słów oraz wykorzystaniu mechanizmu atencji. Wprowadzeniu metody towarzyszy hipoteza, która sugeruje, że wykorzystanie informacji płynącej z mechanizmu uwagi, stosowanej bliżej oryginalnych wektorów reprezentacji słów, przed ich zakodowaniem, może poprawić skuteczność uczonych modeli klasyfikatorów. Autor zamieścił w pracy formalny opis autorskiego pomysłu obliczania współczynników mieszania wektorowych reprezentacji słów oraz wektorów etykiet. Pomysł został przetestowany na trzech uznanych zbiorach danych poświęconych ocenie sentymentu i porównany do trzech wybranych rozwiązań referencyjnych. Wyniki eksperymentów pokazały jednoznacznie, że zaproponowane mieszanie danych pozwala uzyskać wyższą skuteczność klasyfikacji, ale otworzyły również dyskusję na temat selektywnego wyboru poziomu, na którym wykorzystywana jest informacja z mechanizmu uwagi: poziom warstwy (wszystkich głów w danej warstwie) i poziom wybranej głowy w warstwie. Ogólne wnioski wskazują, że zamiast podejścia uśrednionego, które implementuje podstawowa wersja metody, warto rozważać

pojedyncze głowy, a wybierając głowy należy kierować uwagę na warstwy bliższe wejściu do sieci, gdyż w początkowych warstwach jest więcej głów dających lepsze wyniki. Doktorant bardzo dojrzale zwrócił uwagę na istotność semantycznego wyjaśnienia istotności poszczególnych głów, co pozwoliło na postawienie kolejnej hipotezy, że w analizie sentymentu (tego dotyczą zbiory danych), kluczowe znaczenie mogą mieć wybrane części mowy, tj. przymiotniki, przysłówki, czasowniki. Dodatkowe eksperymenty, jak demonstruje to Doktorant na wykresach, potwierdziły częściowo, że dla testowanych zbiorów danych, wśród sześciu najlepiej radzących sobie głów uwagi znajduje się wskazanie na trzy postulowane części mowy. Tym samym sekwencja eksperymentów jest wkładem w aktualny stan wiedzy dotyczący wykorzystania mechanizmu atencji we wzbogacaniu danych uczących.

Druga metoda, o akronimie *StatMix*, osadzona jest bezpośrednio w wyzwaniach, które narzuca uczenie federacyjne modeli. Zamiast przesyłania gradientów na lokalnych danych lub wag lokalnych modeli, Doktorant postuluje autorski sposób dzielenia się uśrednionymi parametrami lokalnych danych obrazowych. Bazując na wartościach średnich kolorów (każdego kanału koloru obrazu osobno) oraz odchyleniach standardowych kolorów pikseli obrazów, rozpropagowanych pomiędzy węzłami, odbywa się wzbogacanie danych obrazowych. Proces wzbogacania danych w węzłach odbywa się na zasadzie podobnej do transferu stylu pomiędzy obrazami. Eksperymenty przeprowadzone na dwóch zbiorach danych (CIFAR-10 i CIFAR-100) wykazały, że zastosowanie metody *StatMix* wpływa pozytywnie na skuteczność trenowanych modeli, a wpływ rośnie również wraz ze wzrostem liczby węzłów biorących udział w uczeniu federacyjnym – w przypadku zbioru CIFAR-100 nie dla każdego testowanego przypadku zasada ta została udowodniona. Ciekawym spostrzeżeniem jest fakt, że tradycyjne metody augmentacji (np. odwracanie, przycinanie) pogarszają skuteczność modeli w scenariuszach uczenia federacyjnego. W ramach badań uzupełniających wykazano również, że stosowanie metody *StatMix* powinno być ograniczone prawdopodobieństwem z przedziału [0.1;0.8] dla zbioru CIFAR-10 lub przedziału [0.1;0.4] dla zbioru CIFAR-100, aby można było zaobserwować poprawę skuteczności trenowanego modelu. Eksperymenty pokazują, że mechanizm wzbogacania danych powinien być stosowany bardzo subtelnie i zweryfikowany eksperymentalnie, aby korzystnie wpływał na skuteczność modeli klasyfikacyjnych.

Trzecie rozwiązanie, zaproponowane przez Doktoranta, jest odpowiedzią na powszechny niedobór oznaczonych danych treningowych w wielu współczesnych procesach biznesowych. W takiej sytuacji pozostaje możliwość posiłkowania się danymi dostępnymi w Internecie, przy zastrzeżeniu ich ograniczonej jakości. Aby tworzyć modele przy użyciu takich danych, Doktorant postuluje rozwiązanie, w którym wyszukane, zgodnie z zadanym kluczem, dane są czyszczone automatycznie i wykorzystane w procesie uczenia wybranych architektur klasyfikatorów. Doktorant dowodzi, że nawet korzystając z takich niedoskonałych danych można skutecznie nauczyć klasyfikator rozwiązywania wybranych problemów, a ostateczna skuteczność klasyfikacji zależy od reprezentatywności danych testowych. Eksperymenty przeprowadzone na trzech zbiorach danych, w tym jednym autorskim, które obejmowały zdjęcia 10 miast, 10 kategorii posiłków i 20 kategorii obiektów, przy zachowaniu odpowiedniego wolumenu danych, mogą być zastosowane do efektywnego douczenia popularnych splotowych architektur klasyfikatorów (VGG16, ResNet50). Klasyfikatory uczone na zbiorach danych zaszumionych są w stanie osiągnąć dobre rezultaty na testowych danych reprezentatywnych. Co więcej, dodatkowe eksperymenty, pozwoliły wykazać, że modele uczone na małych zbiorach danych reprezentatywnych mogą mieć gorszą skuteczność niż modele uczone na dużych zbiorach danych pobranych z Internetu, a stopniowe zwiększanie wolumenu danych pobranych z Internetu poprawia skuteczność klasyfikatora w coraz mniejszym stopniu. Zademonstrował również fakt, że inicjacja modelu klasyfikatora losowymi wagami pozwala osiągnąć podobne rezultaty co inicjalizacja wagami z *ImageNet*, chociaż zbieżność modelu jest nieco wolniejsza.

Dysertację wieńczy rozdział podsumowujący wkład w dyscyplinę naukową oraz zalety i wady zaproponowanych rozwiązań, z którymi w pełni się zgadzam. Hipotezy badawcze, które zarówno w formule ogólnej, sformułowanej na początku pracy, jak również w postaci hipotez cząstkowych, zamieszczonych w poszczególnych rozdziałach, należy uznać za udowodnione. Doktorant zaproponował również perspektywy dalszych badań, co jest o tyle wartościowe, że duża część wniosków nie jest oczywista i skłania do przemyśleń.

Podsumowując, przedstawione rozwiązania, eksperymenty oraz wyniki są ciekawe i nowatorskie. Szczególnie inspirujący, choć nieoczywisty, jest algorytm mieszania wykorzystujący mechanizm atencji, który pozwala lepiej organizować dane podczas uczenia modeli klasyfikatorów. W pewnym sensie otwierające dyskusję są też eksperymenty czerpiące dane masowo z Internetu do nauki klasyfikatorów, które osiągają bardzo zbliżone rezultaty do scenariuszy bazujących na wycyzelowanych zbiorach danych obrazowych.

Na czym polega oryginalność rozprawy, co stanowi samodzielny i oryginalny dorobek Autora, jaka jest pozycja rozprawy w stosunku do stanu wiedzy, czy poziomu techniki reprezentowanych przez literaturę światową?

Oryginalność rozprawy doktorskiej leży w zbiorze metod, które stosując różne strategie rozszerzania nieustrukturyzowanych danych uczących pozwalają zwiększyć skuteczność wybranych modeli klasyfikatorów. O wartości rozprawy stanowi również alternatywne podejście do zwiększania wolumenu danych, gdzie zamiast technik mieszania, wykorzystuje się dane pozyskane bezpośrednio z Internetu. W szczególności najważniejszymi, moim zdaniem, osiągnięciami są:

- metoda wzbogacania danych, o akronimie *AttentioMix*, polegająca na mieszaniu obserwacji z użyciem mechanizmu uwagi, która została dedykowana danym tekstowym;

- metoda wzbogacania danych, o akronimie *StatMix*, polegająca na mieszaniu zdjęć ze statystykami innego zdjęć, która była inspirowana transferem stylu i została dedykowana uczeniu federacyjnemu klasyfikatorów;

- metoda opierająca proces uczenia o automatycznie przetworzone dane pobrane z Internetu, bez konieczności ich etykietowania;

Wymienione osiągnięcia stanowią bezsprzecznie oryginalny wkład Doktoranta w rozwój dyscypliny Informatyka Techniczna i Telekomunikacja.

Czy Autor wykazał umiejętność poprawnego i przekonywującego przedstawienia uzyskanych przez siebie wyników (zwięzłość, jasność, poprawność redakcyjna rozprawy)?

Wszystkie opracowane przez Doktoranta metody zostały udokumentowane w sposób poprawny, zgodnie z przyjętymi w dziedzinie badań zasadami oraz należytą drobiazgowością. Praca jest zredagowana w spójny i logiczny sposób. Miejscami komentarz mógłby być bardziej kompletny, przykładowo, czy rzeczywiście w ramach zbioru danych IMDB wydzielono zbiór testowy liczący aż 25000 obserwacji, jak sugeruje tabela 5.1? W pracy, przy każdej kluczowej metodzie, umieszczone zostały cenne schematy, które stanowiły istotne uzupełnienie pseudokodów. Zarówno scenariusz eksperymentów jak i opracowanie statystyczne wyników, było zrealizowane bardzo rzetelnie i z rzadko spotykaną drobiazgowością. Na uwagę zasługuje również schemat prowadzonych rozważań. Przy każdym rozwiązaniu stawiane były hipotezy badawcze, pod kątem których prowadzone były eksperymenty, a opracowane statystycznie wyniki opatrzone zostały należytą dyskusją i odpowiedziami na postawione wcześniej pytania. Praca została zredagowana bardzo starannie, zawiera odpowiednie spisy, odwołania i odsyłacze do literatury.

Jakie są słabe strony rozprawy i jej główne wady?

Rozprawa doktorska nie zawiera istotnych uchybień, a poniższe uwagi mają przede wszystkim charakter dyskusyjny. Przykładowe kwestie, które mogłyby zostać przedyskutowane, lecz nie ujmują pozytywnej oceny rozprawy, dotyczą:

- Czy metoda mieszania *AttentionMix*, czerpiąca informację z mechanizmu uwagi, mogłaby być zastosowana w augmentacji przywracającej zbalansowanie zbioru danych? Czy można świadomie zarządzać parametrami określającymi liczbę głów uwagi oraz warstw uwagi, jak również ich wartościami, aby skuteczniej interpretować jakość zbioru danych uczących?

- Inspirując się wynikami zamieszczonymi w tabeli 6.2 i 6.3 warto podjąć dyskusję dotyczącą zakresu stosowania metody *StatMix*; Jakie powinny być proporcje pomiędzy liczbą klas, wielkością zbioru i poziomem augmentacji oraz być może złożonością architektury modelu klasyfikatora, aby realne było zwiększenie skuteczności klasyfikatorów, a nie osłabienie skuteczności klasyfikacji?

- W kontekście metody zaprezentowanej w rozdziale 7 przedstawiono wyniki, które pokazują, że niezależnie czy model był uczony/douczany na zbiorze wyczyszczonym, czy zbiorze zaszumionym, zaczerpniętym z Internetu, skuteczność klasyfikacji obrazów reprezentatywnych jest podobna i lepsza niż skuteczność klasyfikacji danych niereprezentatywnych. Czy w przypadku modeli uczonych danymi reprezentatywnymi można zaobserwować podobną tendencję?

Konkluzja

Uważam, że cele rozprawy doktorskiej zostały w pełni zrealizowane, a postawione hipotezy bezsprzecznie udowodnione. Doktorant przedstawił w rozprawie nowe i oryginalne strategie radzenia sobie z niedoborem obserwacji uczących lub nawet ich całkowitym brakiem. Uzyskane przez Doktoranta wyniki uważam za oryginalne, wartościowe i ciekawe poznawczo. Zakres prowadzonych badań był szeroki, analiza opracowanych rozwiązań wszechstronna, a warsztat badawczy Doktoranta zasługuje na wyróżnienie. Wyróżniająca jest również działalność publikacyjna Doktoranta, która potwierdza jakość opracowanych metod. Doktorant może poszczycić się dwoma publikacjami w czasopismach z listy JCR, o wysokim współczynniku wpływu, jedną publikacją konferencyjną z listy CORE i jednym tekstem jeszcze nieopublikowanym.

Tym samym rozprawa prezentuje wartościowe osiągnięcia naukowe w obszarze dyscypliny Informatyka Techniczna i Telekomunikacja oraz potwierdza umiejętność samodzielnego prowadzenia przez Doktoranta pracy naukowej.

Bez najmniejszej wątpliwości stwierdzam, że recenzowana rozprawa doktorska Pana mgra Dominika Lewego spełnia z wyraźnym nadmiarem wymagania stawiane rozprawom doktorskim, przez obowiązującą ustawę. Wnoszę o jej przyjęcie i dopuszczenie rozprawy do publicznej obrony. Wnoszę również o wyróżnienie rozprawy, argumentując swój wniosek wartością naukową opracowanych rozwiązań, zakresem i wnikliwością przeprowadzonych eksperymentów badawczych oraz znamienitym dorobkiem publikacyjnym.

Adam Wojciechowski