

Dr hab. Marek Sikora, prof. PŚ
Katedra Sieci i Systemów Komputerowych
Politechnika Śląska
ul. Akademicka 16
44-100 Gliwice
Email: marek.sikora@polsl.pl

Recenzja rozprawy doktorskiej

Tytuł rozprawy: Big data w systemach rozpoznania wojskowego

Autor rozprawy: mgr Kamila Matela

Promotor rozprawy: dr hab. inż. Andrzej Stachurski

Dziedzina: nauki inżynieryjno-techniczne

Dyscyplina: informatyka techniczna i telekomunikacja

1. Temat i cel rozprawy

Tematyka rozprawy obejmuje zagadnienia zastosowania metody aktywnego uczenia do oznaczania (tagowania) przykładów w problemach klasyfikacji obrazów. Cele pracy zdefiniowane przez autorkę to między innymi: zaprezentowanie wielokryterialnej strategii wyboru przykładów do oznaczania – Autorka bazuje w tym zakresie głównie na publikacji, jaka ukazała się w 2019 roku w uznanym czasopiśmie Knowledge Based Systems, przedstawienie modyfikacji tej strategii poprzez zastosowanie metody optymalizacji wielokryterialnej, bazującej na metodzie punktu odniesienia oraz weryfikacja przedstawionej propozycji na podstawie analizy dwóch benchmarkowych zbiorów danych. Weryfikacja obejmuje głównie badanie sprawność klasyfikacji klasyfikatorów trenowanych na oznaczanych zbiorach przykładów. Badanie zawiera analizę jak zmienia się ta sprawność w kolejnych rundach oznaczania.

Tłem dla prowadzonych badań są rozważania na temat systemów rozpoznania wojskowego oraz roli klasyfikacji obrazów w tego typu systemach. Główny nacisk położono tutaj na rozpoznawanie obiektów, chociaż Autorka wspomina także o monitorowaniu zmian i prognozowaniu zjawisk.

Uzasadnienie wyboru tematu nie budzi wątpliwości, Autorka dosyć trafnie uzasadnia celowość podjęcia badań opisanych w rozprawie. Dysponowanie dobrze oznaczonymi (otagowanymi) zbiorami danych jest niezwykle ważne z punktu widzenia trenowania systemów klasyfikacji i nie dotyczy jedynie klasyfikacji obrazów. W przypadku dużych zbiorów danych, dobra metoda wyboru „reprezentatywnej” próbki poprawnie oznaczonych danych pozwala zazwyczaj na znaczne przyspieszenie budowy klasyfikatora o dobrych zdolnościach klasyfikacji. Ponadto, ograniczenie liczby przykładów do tych najistotniejszych powoduje, że wymagana jest mniejsza liczba ekspertów oznaczających (tagujących) dane.

2. Zawartość i charakter rozprawy

Rozprawa składa się z czterech rozdziałów oraz spisu literatury.

We wstępie przedstawiono krótko motywację podjęcia badań wraz z głównymi celami pracy. Wstęp zawiera również zdefiniowaną przez Doktorantkę tezę.

Rozdział pierwszy obejmuje tematycznie szeroki wachlarz zagadnień. W rozdziale tym przedstawiono definicję i zastosowania koncepcji Big Data, omówiono zagadnienia związane z wizją komputerową oraz algorytmami klasyfikacji i miarami oceny jakości klasyfikatorów. Tematyka rozdziału jest szeroka. Powoduje to, że poruszane w nim zagadnienia omawiane są dosyć lakonicznie. Najwięcej uwagi poświęcono algorytmom klasyfikacji, przy czym Autorka koncentruje się na standardowych algorytmach, które poza SVM, niekoniecznie dedykowane

są do celu klasyfikacji obrazów. Dobrze przedstawione są zagadnienia weryfikacji jakości klasyfikacji, przy czym w rozdziale w ogóle nie poruszono kwestii metodyk oceny eksperymentalnej (podziałów na zbiory treningowe, testowe i ew. walidacyjne). Nie wspomniano w ogóle o metodach porównania klasyfikatorów, chociaż zagadnienie to jest ważne dla wyników prezentowanych w dalszej części pracy.

Rozdział drugi koncentruje się na metodach aktywnego uczenia. Autorka przedstawia kilka strategii aktywnego uczenia i na końcu tego przeglądu prezentuje pracę *Multi-criteria active deep learning for image classification* (MCADL), która ukazała się w 2019 w czasopiśmie Knowledge Based Systems. Autorka tę właśnie pracę i opisaną w niej metodę wzięła jako punkt odniesienia do swoich badań. W rozdziale przedstawiono szczegółowo zasady działania metody MCADL. Autorka chcąc upewnić się, że wyniki osiągnięte przez metodę są rzeczywiście konkurencyjne lub lepsze do znanych metod, przeprowadza szereg eksperymentów obliczeniowych związanych z zastosowaniem własnej implementacji MCADL do benchmarkowego zbioru danych MNIST. Uzyskane wyniki nie są jednoznaczne, co uczciwie Autorka spostrzega. Metoda MCADL na tle losowej inicjalizacji i wyboru przykładów czasem nie wykazuje swojej przewagi zarówno w szybkości dochodzenia do wysokiej sprawności (dokładności) klasyfikacji końcowego modelu, jak również w trakcie kolejnych rund uczenia. Autorka nie przedstawia pogłębionej analizy przyczyn takiego stanu rzeczy. Dziwi również, że Autorka nie przedstawia wyników MCADL na tle innych znanych metod aktywnego uczenia, o których wspomina w rozdziale. W ogólnie dostępnych bibliotekach programistycznych metody takie są dostępne.

Rozdział III poświęcony jest zagadnieniom analizy i optymalizacji wielokryterialnej. Sekcja 1 zawiera opis/przegląd wybranych zagadnień związanych z tą tematyką. Sekcja 2 jest najważniejszą częścią ocenianej rozprawy, gdyż zawiera opis modyfikacji MCADL proponowanych przez Autorkę. Ponieważ tę część pracy uważam za największe i najwartościowsze osiągnięcie, opiszę je szczegółowo w sekcji Oryginalne wyniki i ich znaczenie. W początkowej części rozdziału III – przed przedstawieniem oryginalnych metod proponowanych przez Autorkę – zamieszczono również opis dwóch benchmarkowych zbiorów danych, jakich użyto w eksperymentach obliczeniowych. Paradoksalnie, zbiory te nie mają wiele wspólnego z obszarem wojskowości. Pewną wskazówkę dotyczącą takiego stanu rzeczy jest umieszczona na początku rozprawy wzmianka o tym, że postępowanie na dostawę elementów systemu rozpoznania wojskowego, w ramach którego zamierzano zgromadzić również dane niezbędne do realizacji ocenianej pracy został odwołany. Sytuacja taka jest niezależna od Doktorantki, a zakres i temat rozprawy realizowanej w trybie doktoratu wdrożeniowego niezwykle trudno zmienić, recenzent zna tę sytuację jako promotor z autopsji. Doktorantka rozważa dwa zbiory danych: MNIST (użycie go można wytłumaczyć chęcią odniesienia się do wyników metody MCADL – wyniki te dla oryginalnej implementacji autorów prezentowane są w przywoływanej już w recenzji publikacji) oraz CAF AR-10. Ten drugi z użytych zbiorów danych dedykowany jest do klasyfikacji – rozpoznania – obiektów

różnych typów (nie są to jednak obiekty wojskowe). W rozdziale przedstawiono również architektury sieci głębokich, jakie stosowane były do klasyfikacji obiektów. Sieci trenowane były na danych dostarczanych/wybieanych przez metodę aktywnego uczenia.

Rozdział IV składa się z dwóch części. W pierwszej części prezentowane są wyniki działania – w kontekście dokładności klasyfikacji na tle liczby rund uczenia – propozycji przedstawionych przez Doktorantkę. Punktem odniesienia jest metoda MCADL, stosowana jest metodyka train-and-test, wnioski definiowane są na podstawie porównań pomiędzy dokładnością poszczególnych metod. Wobec przyjętej metodyki testowania nie sposób oczywiście sprawdzić istotności różnic pomiędzy metodami lub metodą uruchamianą z różnymi wartościami hiperparametrów za pomocą testu statystycznego. Przed prezentacją wyników swojej metody Autorka przytacza jeszcze raz wyniki metody MCADL, w szczególności przedstawia interesującą moim zdaniem analizę – wizualizację – polegającą na oznaczeniu na wykresie 2d przykładów z poszczególnych klas wybranych przez metodę MCADL do uczenia. Niestety czytelnik nie dowiaduje się bezpośrednio jak zrealizowano redukcję wymiarów – można to wywnioskować pośrednio z nazwy pliku, w jakim przechowywane są prezentowane grafiki. Analiza – jak pisze Autorka – ma na celu „zbadanie zachowania różnych kryteriów” stosowanych w MCADL, w celu znalezienia wzorca pozwalającego na zdefiniowanie poziomów referencji dla kryteriów stosowanych w MCADL. Wzorzec ten zostałby wykorzystany w metodzie punktów odniesienia. Autorka konkluduje jednak, że wzorców takich nie znaleziono. Druga część rozdziału IV przedstawia aspekty wdrożeniowe systemu rozpoznania wojskowego. Jest to opis bardzo ogólny, nie przedstawiający architektury systemu rozpoznania, jaki w dalszym ciągu planowany jest do wdrożenia w praktyce. Interesujące w tej części rozdziału jest uzasadnienie konieczności szybkiego oznaczania (tagowania) danych (w tym obrazów) dla celów trenowania klasyfikatorów, będących elementami większych systemów wspomagania podejmowania decyzji dedykowanych dla zastosowań wojskowych. Na zakończenie Autorka opisuje działania Polskiej Grupy Zbrojeniowej, mające na celu doprowadzenie do budowy Systemu Analiz Obrazowych. Na tym tle wskazuje miejsce potencjalnego zastosowania prezentowanego przez nią rozwiązania – szerzej, systemów wspomagających oznaczenie przykładów.

Ostatni rozdział zawiera podsumowanie wyników oraz odniesienie się do tezy rozprawy. W podsumowaniu Autorka nie przedstawiła możliwych kierunków dalszych prac. Stwierdzam to z ubolewaniem, bo czytelnik czuje pewien niedosyt, zwłaszcza w zakresie przedstawienia bardziej szczegółowej architektury planowanego systemu rozpoznania wojskowego i miejsca, jakie w tej architekturze (w elemencie/module klasyfikacji obiektów) zajmie rozwiązanie proponowane przez autorkę.

Teza rozprawy

We wstępie Doktorantka przedstawia kilka celów rozprawy i na tym tle stawia następującą tezę:

Zastosowanie punktów odniesienia w algorytmie Multi-criteri active deep learning for image classification, zwiększa dokładność analizy obrazu i wykazuje wyższość nad wykorzystywaną w tymże algorytmie metodą ważenia kryteriów.

Teza została postawiona jasno i czytelnie, a badania opisane w rozprawie pokazują dążenie do jej weryfikacji empirycznej. Uzyskane wyniki nie pozwalają w sposób wysoce prawdopodobny na wnioskowanie o prawdziwości tej tezy. Recenzent odniesie się do tego aspektu w sekcji Słabe strony i uwagi krytyczne/dyskusyjne. Ważne jest, że w pracy przedstawiono cele rozprawy, teza jest tutaj moim zdaniem mniej istotna, gdyż w pracach związanych

z metodami wnioskowania indukcyjnego powinniśmy raczej definiować cele niż tezy, których prawdziwości lub nieprawdziwości zależy silnie od rozważanych danych.

Pewne zdziwienie budzi fakt, że teza rozprawy w luźny sposób nawiązuje do jej tytułu. Z treści rozprawy wnioskujemy, że główny cel badań i związana z nim teza stanowi jedynie niewielki element postulowanego systemu rozpoznania wojskowego.

3. Analiza źródeł i zastany stan wiedzy

Bibliografia recenzowanej rozprawy doktorskiej składa się ze 109 pozycji. Autorka cytuje je w odpowiednim kontekście. Źródła te dobrze przedstawiają bieżący stan wiedzy w obszarze tematyki rozprawy. Liczba pozycji literatury jest w mojej opinii odpowiednia i wystarczająca. Tematyka przywoływanych prac dotyczy zarówno aktywnego uczenia, jak również klasyfikacji obrazów, optymalizacji i systemów wsparcia/rozpoznania wojskowego. Przywoływane prace są w dużej części nowe, opublikowane w okresie ostatnich dziesięciu lat. Autorka przywołuje również swoje publikacje związane z tematyką pracy – głównie z systemami rozpoznania wojskowego.

W spisie literatury jest dużo odnośników do publikacji zamieszczonych w serwisie arxiv. Zastrzeżenia budzi sposób przywoływania części z tych prac, zamiast podawania tytułu pracy i listy autorów, Autorka podaje jedynie link do dokumentu źródłowego. Taki sposób prezentowania referencji utrudnia czytelnikowi analizę źródeł.

4. Oryginalne wyniki i ich znaczenie

Doktorantka podejmuje ważny problem poprawy wyników metody – strategii – aktywnego uczenia poprzez wielokryterialną, stosowaną iteracyjnie, analizę przykładów kandydujących

do kolejnych rund uczenia. Doktorantka adaptuje metodę punktów odniesienia do metody MCADL i porównuje uzyskane wyniki z bazową metodą MCADL oraz strategią losową. Doktorantka wykazuje wiele zalet takiego podejścia, z których za najważniejsze uważam wsparcie w procesie ustalania granic pomiędzy przykładami. Autorka proponuje również pewien systematyczny sposób wyboru próbki inicjującej – zamiast wyboru losowego bazuje on na wynikach grupowania danych wykonywanego na przekształconym (metoda PCA) zbiorze przykładów i zastosowaniu algorytmu k-średnich.

Punkty odniesienia ustalane są dla czterech kryteriów – gęstości podobieństwa, niepewności i, jak to definiuje Autorka, kryterium opartego na etykietach przykładów. W zakresie wpływu wartości kryteriów na wyniki trenowania modelu Doktorantka przeprowadza dosyć szczegółową analizę eksperymentalną.

Dobrze oceniam też krytyczną analizę zbalansowania wyników (dokładności klasyfikacji) na zbiorach treningowych i walidacyjnych oraz próbę ich lepszego zbalansowania (znielowania nadmiernych różnic). Doktorantka stosuje metodę margin sampling, modyfikując proces wyboru próbek – metoda margin sampling bazuje na próbkach wybranych uprzednio przez metodę wielokryterialną.

Mając na uwadze aspekt wdrożeniowy, ważnym wynikiem prowadzonych prac jest przygotowana implementacja zarówno metody odniesienia, jak i proponowanych przez Autorkę innowacji.

5. Redakcja rozprawy i prezentacja wyników

Rozprawa zredagowana jest w sposób poprawny. Doktorantka nie ustrzegła się jednak mniejszych i większych błędów – zwłaszcza w zakresie prezentacji wyników eksperymentów. Wyniki prezentowane są głównie w postaci wykresów dokładności klasyfikacji na tle kolejnych rund uczenia. Interpretacja tak prezentowanych wyników jest trudna, nie zawsze wiadomo czy wykres przedstawia wynik uzyskane na danych treningowych, czy testowych. W pracy brakuje także bardziej formalnego opisu używanych metod, w szczególności metod proponowanych przez Autorkę.

Pozostałe rozdziały zredagowane są poprawnie, indeksy, spisy, tabele, etc. ułatwiają czytanie i zrozumienie prezentowanego materiału. W podpisach pod rysunkami Autorka w miarę dokładnie opisuje ich (rysunków) zawartość. W przypadku reprodukcji zaczerpniętych z prac innych autorów (sporadyczne przypadki) Doktorantka wskazuje wyraźnie prace źródłową.

Kolejne rozdziały tworzą logiczną całość – wyjątkiem są: sekcja 2 rozdziału VI, która moim zdaniem powinna być umieszczona w początkowej części pracy oraz część sekcji 3 rozdziału II, zawierające krótki opis zasady działania sztucznej sieci neuronowej – ta część pracy powinna znaleźć się w rozdziale I obok opisów innych metod klasyfikacji.

6. Słabe strony i uwagi krytyczne/dyskusyjne

Uwagi krytyczne dotyczą dwóch głównych aspektów: weryfikacji metody oraz osadzenia rozwiązania w systemie rozpoznania wojskowego. Poniżej przedstawiono najważniejsze z uwag krytycznych.

1. W doktoracie brak jest wystarczającego uzasadnienia, że podejście wielokryterialne ma znaczenie dla rozważanego obszaru zastosowania. Wybór jednej konkretnej metody odniesienia był ryzykowany, zwłaszcza w kontekście przedstawionych uczciwie przez Doktorantkę badań weryfikacyjnych.
 - a. Dlaczego Doktorantka kontynuowała prace nad poprawą wyników działania metody, skoro Jej badania weryfikujące wykazały, że metoda nie posiada zdecydowanych przewag nad metodami znanymi?
 - b. Dlaczego oparto się na własnej implementacji, a nie na implementacji udostępnionej w repozytorium github?
 - c. Skąd wynika konieczność stosowania metod wielokryterialnych LUB jakie teoretycznie korzyści metody te mogą przynieść w zadaniu active learning?
2. Dlaczego Autorka nie porównała wyników swoich badań z innymi – co najmniej jedną – znanymi metodami Active Learning?
3. Dlaczego nie zastosowano lepszej metodyki testowania np. wielokrotny trian-and-test lub walidacja krzyżowa? Czy zdaniem Autorki pozwoliłoby to na lepszą/dokładniejszą weryfikację jakości porównywanych metod?
4. W sekcji 2.4.3 rozdziału II przedstawiono warunek stopu algorytmu dla konkretnego zbioru danych. Wydaje się, że propozycja ta po pierwsze jest bardzo ograniczona utylitarnie (a moim zdaniem ma potencjał praktyczny), po drugie z rysunku 37 wynika, że możemy mieć do czynienia z wyciekami danych – ostatni etap losowania przykładów ze zbioru walidacyjnego do procesu uczenia – a co za tym idzie, prezentowane wyniki są zbyt optymistyczne.
 - a. Czy metoda może być również stosowana do innych zbiorów danych?
 - b. Czy rzeczywiście dobrze identyfikują wyciek danych?
5. Warstwa prezentacji wyników jest uboga, bazuje jedynie na dosyć, moim zdaniem, nieczytelnych wykresach. Wyniki eksperymentów rozmieszczone są trochę chaotycznie w rozdziałach III i IV. Brakuje zestawienia tabelarycznego z rozbiciem na rundy oraz wyniki na zbiorze treningowym i testowym. Trudno jednoznacznie wywnioskować czy badania w świetle postawionej tezy się powiodły, czy nie (przy czym wynik negatywny też uznawałbym za cenny).

6. Jedynie z części wykresów dowiemy się czy wyniki dotyczą zbiorów treningowego czy testowego/walidacyjnego.
7. Nie doszukałem się jednoznacznego opisu przyjętej metodyki eksperymentów – ile przykładów było prezentowanych algorytmowi w kolejnych rundach?
8. Autorzy oryginalnej metody piszą, że sprawuje się ona dobrze dla danych niezbalansowanych, Autorka w ogóle nie adresuje tego problemu. Moim zdaniem zbiory danych, na których przeprowadzono eksperymenty nie są niezbalansowane, co może tłumaczyć brak jednoznacznych konkluzji dotyczących efektywności metody.

W aspekcie wdrożeniowym – jak już pisałem – brakuje mi uzasadnienia, że proces oznaczania danych ma znaczenie w rozważanej dziedzinie zastosowania oraz że wybór wielokryterialny jest tutaj ważny – niekoniecznie szczególnie ważny, ale ważny. Brak mi również – pomimo tego, że wiem, iż realizacja prac wdrożeniowych w dużej mierze uzależniona była od wyników przetargu – propozycji ogólnej architektury systemu wsparcia rozpoznania z wyszczególnieniem elementu rozpoznania obiektów.

Mniej istotne uwagi odnoszą się do pewnej liczby błędów gramatycznych, jakie zauważano w pracy oraz bardzo mało sformalizowanego opisu. Brakuje mi trochę wzorów, algorytmów, schematów blokowych, etc.

7. Podsumowanie i wniosek końcowy

Po analizie rozprawy mogę stwierdzić, że zamieszczone w niej rezultaty prezentują oryginalne rozwiązanie w zakresie zastosowania wyników badań naukowych Doktorantki w obszarze udoskonalania metod active learning dla celów wyboru przykładów dla klasyfikacji obrazów. Rozwiązanie to może być stosowane również w systemach rozpoznania wojskowego. Nie lekceważąc dosyć poważnych uwag krytycznych i dyskusyjnych, jakie podniosłem w niniejszej recenzji, stwierdzam jednak, że unikalne osiągnięcie przedstawione w rozdziale III.2 skłania mnie do umiarkowanie pozytywnej oceny pracy. Reasumując, moja konkluzja dotycząca rozprawy na obecnym etapie postępowania jest pozytywna. Z niecierpliwością czekam na możliwość dyskusji z Doktorantką podczas publicznej obrony. Warto nadmienić, że pewną część wątpliwości dotyczących ocenianej pracy rozwiązałem podczas spotkania seminaryjnego z Doktorantką.

Stwierdzam, że recenzowana rozprawa pt. „Big data w systemach rozpoznania wojskowego” przygotowana przez mgr Kamilę Matelę spełnia wymagania i warunki określone w ustawie z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (jednolity tekst Dz. U. z 2023 r. z późn. zm.) i wnoszę o jej dopuszczenie do publicznej obrony i dalszych etapów postępowania doktorskiego.

