

Moł. RDITT-25.05.2023v.  
M. Skory

prof. dr hab. Piotr Porwik  
Uniwersytet Śląski, Katowice  
Wydział Nauk Ścisłych i Technicznych  
Instytut Informatyki  
ul. Będzińska 39  
41-200 Sosnowiec

Katowice 12.05.2023

## RECENZJA ROZPRAWY DOKTORSKIEJ

*”ENSEMBLES OF SUPPORT VECTOR MACHINES WITH  
EVOLUTIONARILY OPTIMIZED HYPERPARAMETERS AND  
TRAINING SETS”*

mgra inż. Wojciecha Dudzika

Recenzja została sporządzona na podstawie pisma prof. dr hab. inż. Andrzeja Polańskiego, Przewodniczącego Rady Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Śląskiej.

*Promotorem rozprawy doktorskiej jest Pan dr hab. inż. Michał Kawulok, prof. PŚ, przy współdziałaniu promotora pomocniczego, Pana dr hab. inż. Jakuba Nalepy, prof. PŚ.*

### **Ocena doboru tematu**

Rozprawa doktorska napisana została w języku angielskim i opisuje sposób budowy zespołu maszyn wektorów nośnych, dla których parametry optymalizowano metodami ewolucyjnymi z równoczesnym optymalizowaniem zbioru uczącego.

Tematyka pracy jest bardzo aktualna w świetle najnowszych osiągnięć związanych z Informatyką i sztuczną inteligencją. Powody obecnego zainteresowania tematyką uczenia maszynowego są wieloaspektowe. Z jednej strony można analizować dane z różnych dziedzin, z drugiej strony nowoczesne algorytmy sztucznej inteligencji pozwalają na badanie dużych zbiorów danych. Takie zbiory można oczywiście klasyfikować różnymi metodami - klasyfikatorami pojedynczymi czy zespołami klasyfikatorów. Każdorazowo należy jednak dobrać zestawy odpowiednich parametrów lub hiperparametrów klasyfikatora, co jest zadaniem nietrywialnym. Autor rozprawy podejmuje się rozwiązać te problemy. Prowadzi to w efekcie do praktycznych stosowań algorytmów sztucznej inteligencji w wielu obszarach

ludzkiej działalności: medycynie, bioinformatyce, bankowości, rozwiązaniach biometrycznych, wojskowości i wielu innych. Ma to więc duże znaczenie dla poprawy dobrostanu społeczeństw i gospodarek różnych krajów. Podejście proponowane przez Doktoranta pozwala na lepsze zrozumienie mechanizmów doboru klasyfikatorów SVM oraz ich podstawowych parametrów, dla których uzyskiwana predykcja wyników jest najlepsza. Z naukowego punktu widzenia jest to działanie pożyteczne, pozwalające na planowanie nowych eksperymentów.

### **Charakterystyka rozprawy i jej zakres**

Praca jest mocno związana z Informatyką teoretyczną, a rozwiązania algorytmiczne potwierdzone są metodami eksperymentalnymi i statystycznymi. Proponowane w pracy doktorskiej rozwiązania mogą być oczywiście z powodzeniem stosowane w praktyce, co jest domeną Informatyki Technicznej.

Należy również podkreślić, że praca została opracowana bardzo starannie i prezentuje zarówno wysoki poziom edytorski, jak i dobry poziom merytoryczny.

Praca podzielona została na kilka uzupełniających się części. Składa się z 5 rozdziałów, dwóch dodatków oraz spisu literatury. Na stronie 4, w rozdziale 1.1 Autor stawia dwie hipotezy badawcze, które będą potwierdzone w dalszej części rozprawy:

- prowadzenie jednoczesnej optymalizacji zbioru uczącego oraz hiperparametrów SVM poprawia czas uczenia w porównaniu z innymi adekwatnymi metodami klasyfikowania danych,
- utworzone przy użyciu metod ewolucyjnych zespoły klasyfikatorów SVM zapewniają lepszą wydajność klasyfikacji w porównaniu z innymi algorytmami bazującymi również na maszynach wektorów nośnych (SVM).

W pierwszej hipotezie lepszym zamiennikiem wyrazu "poprawia" byłoby użycie wyrazu "zmniejsza" lub "skraca" czas uczenia, bo właśnie o to chodzi Autorowi i jest bardziej precyzyjne.

Zrozumienie drugiej hipotezy jest trudniejsze, gdyż Autor zapewnia w niej, że nastąpi polepszenie "wydajności klasyfikacji". Kontekst użycia wyrazu "wydajność" jest tutaj pojęciem wyjątkowo rozmytym i nie wyjaśnia o jaką wydajność Autorowi chodzi. Objasnienie tych terminów znajdują się dopiero w kolejnych akapitach zawierających komentarze Autora.

W tym samym rozdziale Autor podaje, w Tabeli 1.1, listę opublikowanych artykułów, gdzie jest autorem lub współautorem. Na liście znajduje się 7 prac. W Tabeli 3.3. natomiast jest już zacytowanych 11 prac własnych (w tym jedna jeszcze nie opublikowana). Z czego wynika ta różnica?

Obszerny rozdział 2, o charakterze przeglądowym, przedstawia taksonomię związaną z wielowariantową klasyfikacją SVM oraz algorytmami ewolucyjnymi wraz z objaśnieniem współczynników używanych w uczeniu maszynowym do porównywania różnych metod klasyfikacji. Rozdział został napisany kompetentnie, z dbałością o szczegóły i znacząco pomaga w zrozumieniu materiałów zawartych w kolejnych rozdziałach rozprawy. Maszyny SVM są jedną z popularnych metod przeprowadzania klasyfikacji binarnej. Działają poprzez dopasowanie hiperpłaszczyzny, która oddziela dwie klasy punktów danych, jednocześnie maksymalizując margines między hiperpłaszczyzną, a najbliższymi punktami danych. Te najbliższe punkty są często wybierane jako wektory nośne, które wyznaczają granicę decyzyjną (hiperpłaszczyznę) klasyfikatora SVM i są kluczowe dla procesu klasyfikacji nowych punktów danych. Dla wyjaśnienia tych zależności Doktorant posługuje się często przykładami graficznymi, pokazującymi rozkład wektorów nośnych (SV) na tle granic decyzyjnych klasyfikatora dla różnych jąder SVM. W wielu miejscach Autor odwołuje się również do bogatej literatury przedmiotu, zamieszczając własne komentarze związane z ograniczeniami niektórych rozwiązań. W praktyce, nie są to działania łatwe, szczególnie wtedy, gdy są przetwarzane duże zbiory danych. Dla takich przypadków wydłuża się czas obliczeń, uwidacznia się problem doboru cech wielowymiarowych próbek, wyznaczania hiperparametrów SVM oraz wzrasta zapotrzebowanie na zasoby pamięciowe.

Rozdział 3 ma wyraźny kontekst teoretyczny i zawiera szczegółowy opis autorskich, wielowariantowych algorytmów optymalizujących dla: ewolucyjnego doboru cech próbek, naprzemiennej oraz równoczesnej genetycznej optymalizacji hiperparametrów SVM, a także algorytm łączenia różnych typów jąder dla mapowania danych wejściowych na przestrzeń cech w maszynie SVM. Autor zauważa, że dane wejściowe podlegające klasyfikacji mają często niejednorodny charakter - mogą być homo lub heterogeniczne. Homogeniczne dane wejściowe są w postawionej hipotezie uznawane za łatwe (pewne) do klasyfikacji, podczas gdy dane heterogeniczne traktowane są jako trudne (niepewne). Prowadzi to w efekcie do utworzenia eksperckiego modelu klasyfikatorów zespołowych działających w zaprogramowanych dla nich obszarach kompetencji, co poprawia globalną skuteczność klasyfikatora zespołowego. Proponowane rozwiązania są korzystne w sensie Pareto ustalając optymalny front równowagi pomiędzy możliwą do osiągnięcia wysoką skutecznością klasyfikatora, a czasem jego uczenia.

Rozdział 4 zawiera eksperymentalną ocenę algorytmów opisanych w rozdziale 3. Do walidacji algorytmów Autor zastosował zarówno sztuczne dane o znanych charakterystykach, jak i dane rzeczywiste, gdzie charakterystyki rozkładu danych są nieznane. Część tych danych posiada również, stworzoną przez Autora, reprezentację graficzną. Jakościowa oraz ilościowa ocena poszczególnych algorytmów odbywają się a) na podstawie obrazów graficznych z zaznaczonymi regionami rozkładu danych na różnych etapach klasyfikacji

SVM b) w postaci wartości numerycznych, za pomocą odpowiednich współczynników jakości. Graficzna prezentacja wyników jest ciekawa szczególnie w obliczeniach ewolucyjnych, dla lepszego zrozumienia przypadków szczególnych w poszczególnych etapach obliczeń ewolucyjnych. Na rysunkach można zaobserwować zmiany położenia wektorów nośnych w zależności od etapu obliczeń. Regiony te zapewniają różne "poziom trudności" klasyfikacji. Regiony trudne (niejednorodne) gorzej się klasyfikują, podczas gdy regiony jednorodne są łatwe i proste do poprawnego sklasyfikowania. Wyszkolony klasyfikator zespołowy SVM jest wykorzystywany do dzielenia przestrzeni wejściowej na części pewne i niepewne, gdzie tzw. pewne regiony zapewniają lepszą jakość klasyfikacji. Region pewny rozumiany jest tutaj jako region, w którym występuje bezbłędna klasyfikacja próbek treningowych. Autor przedstawił także w tabelach 4.3. oraz 4.4 uśrednione wyniki skłarnych wskaźników jakości (ACC, F1, Precision, Recall oraz MCC) uzyskane dla różnych zbiorów danych wejściowych. Ile było tych zbiorów, czy te zbiory były zrównoważone czy nie? Czy to były te same zbiory danych jak wymienione w Tabelach 4.1 i 4.2? Z treści rozprawy wynika, że niektóre ze zbiorów danych były niezbalansowane, dla takich przypadków wskaźniki ACC i F1 są mało wiarygodne.

Rozdział 5 zawiera podsumowanie uzyskanych wyników badań. Po nim zamieszczone są dwa Dodatki: A i B. Ważniejszy dodatek B, przedstawia dowody, że proponowane w rozprawie rozwiązania są lepsze i wykazują statystycznie istotną różnicę w stosunku do metod referencyjnych, znanych z literatury. Wykazano to w oparciu o wyniki testu Friedmana i *post-hoc* testu Conovera.

### **Ocena przyjętych metod rozwiązania zagadnienia**

Metodologia przyjętych rozwiązań jest poprawna, a eksperymenty są wiarygodne i wielowariantowe. Pseudokody algorytmów pozwalają innym eksperymentatorom na przeprowadzenie badań porównawczych. Rozwiązania autorskie wykazują istotną przewagę nad innymi metodami opisanymi w samej rozprawie. Eksperymenty zostały zaprojektowane wielowariantowo, co świadczy o dobrym przygotowaniu Autora do prowadzenia badań naukowych, a także o bardzo dobrej znajomości podejmowanej w pracy tematyki badawczej i metod statystycznych. Lektura rozprawy, a w szczególności Rozdział 4, uzasadniają stwierdzenie, że tezy zostały poprawnie zweryfikowane, a wytyczone cele zostały osiągnięte. Mam jednak kilka uwag szczegółowych.

Tabela 4.4 zawiera ranking metod. Używanie w niej tych samych opisów (ACC, F1,...) jak w tabeli 4.3 jest mylące, chociaż, po chwili namysłu można się zorientować o co Autorowi chodziło. W tym samym rozdziale 4 Doktorant porównuje za pomocą wskaźnika MCC i wykresu pudełkowego różnice dotyczących jakości własnych algorytmów. Z wykresu pudełkowego wynika, że opisywane w pracy algorytmy są podobne (nie widać statystycznie istotnych różnic), ze względu na wspólne obszary rozstępów międzykwartylowych w różnych pudełkach (1 pudełko = 1 metoda), co wynika z podobnych wskaźników korelacji

MCC, uzyskanych na podstawie pewnej liczby zestawów danych wejściowych. Wskaźnik MCC jest tutaj o wiele lepszym wyborem niż współczynniki skalarne ACC lub F1. Wykres na rys. 4.10 nie jest opisany. Pomiędzy kwartylem dolnym i górnym jest prawdopodobnie mediana (linia ciągła), co jednak oznacza linia przerywana? Jak należy interpretować punkty leżące obok pudełek? Autor zastosował chyba najprostrzy wariant, gdzie dolny wąs wyznacza najmniejszą wartość w zbiorze, natomiast górny największą.

Inne testy statystyczne (Friedmana i Conovera) rozstrzygające o statystycznej istotności różnic algorytmów własnych i algorytmów referencyjnych znajdują się dopiero dalej, w Dodatku B. Różnice wyznaczone były dla różnych wartości  $p$  – *value*. Autor wykonał najpierw test Friedmana, aby ustalić czy występują różnice globalne w analizowanych metodach, a potem jeśli odpowiedź testu była twierdząca, wykonany został test *post – hoc* dla ustalenia, które grupy parami dają istotną różnicę. Autor nie podał dla jakich wartości  $p$  ustalona została istotność testu Friedmana. Dałoby to zbiorczy obraz istotności różnic we wszystkich metodach bez wskazywania konkretnej, najlepszej. Testy wykonano dla wykazania różnic we własnych algorytmach, a także dla wykazania różnic pomiędzy własnymi rozwiązaniami, a propozycjami znanymi z literatury. Przyjęta konwencja prezentacji wyników w Dodatku B jest mało czytelna i zawiera niepotrzebnie pola o kolorze różowym dla par algorytmów, które nie były (część słusznie, część nie) porównywane. Wskazywałem na to w tekście powyżej. Niestety Autor nie zadbał o jakikolwiek komentarz wyników, co bardzo wydłuża czas ekspertyzy, gdyż zmusza do żmudnej analizy obrazów graficznych poszczególnych przypadków. W tym miejscu mam również kłopot z interpretacją wyników z rys. 4.10 i rys. 3 z Dodatku B. W obu przypadkach analizowana jest istotność współczynnika MCC na tych samych zbiorach danych surowych. Z rys. 4.10 wynika, że np. dla metod ECE-SVM oraz CE-SVM brak jest podstaw do wykazania statystycznie istotnej różnicy pomiędzy tymi metodami. Dlaczego zatem na rys. 3 pominięto to wyliczenie. Dla innych danych (rys. 4 z Dodatku D) porównanie dla tych metod zostało przeprowadzone. Dotyczy to także innych porównań metod parami, dla ustalenia ich ewentualnych statystycznie istotnych różnic. Proszę o wyjaśnienie tej niekonsekwencji.

### **Analiza źródeł literaturowych**

W rozprawie znajdują się odwołania do 145 prac, wśród których znajduje się 7 prac, gdzie Doktorant jest współautorem artykułu. Są to prace związane z tematyką rozprawy. Popularna platforma WoS (*Web of Science*) wykazuje jednak 12 indeksowanych prac Autora rozprawy z liczbą cytowań 73. Są to prace wartościowe zarówno konferencyjne i publikowane w dobrych czasopismach. Baza WoS odnotowuje indeks Hirscha  $h = 5$ , co jest wartością dobrą. Aktywność naukową Doktoranta należy docenić. Źródła literaturowe dotyczą w większości publikacji z obszaru uczenia maszynowego i metod klasyfikacji, w tym klasyfikatorów SVM. Zestaw pozycji literaturowych jest reprezentatywny i uwzględnia również doniesienia najnowsze, z roku 2023. Wnioski i polemiki wypływające z szerokiego

przeglądu literatury zostały sformułowane przekonująco i należy je uznać za poprawne. Autor kompetentnie dokonał przeglądu i oceny literatury związanej z tematyką rozprawy.

**Ocena wyników przedstawionych w rozprawie** Opracowane przez Doktoranta algorytmy lub ich modyfikacje zostały zweryfikowane w wyniku badań eksperymentalnych i badań porównawczych. Autor konfrontował własne metody z metodami znanymi z literatury oraz metodami opracowanymi przez środowisko kolegów, z którymi pracuje. Przeprowadzona analiza statystyczna uzyskanych wyników klasyfikacji SVM wskazuje, że hipotezy badawcze zostały potwierdzone. Opisane w rozprawie metody równoczesnej optymalizacji zbiorów uczących i hiperparametrów klasyfikatorów SVM przyczyniają się do poprawy fazy treningu klasyfikatora. Szczególnie ważnym osiągnięciem jest wykazanie, że zaproponowane w rozprawie doktorskiej modyfikacje klasyfikatorów SVM zapewniają lepszą jakość klasyfikacji, mierzoną różnymi wskaźnikami skalarnymi i metodami statystycznymi, w porównaniu z adekwatnymi metodami znanymi z literatury.

#### **Wnioski końcowe**

Uważam, że Doktorant, Pan mgr inż. Wojciech Dudzik posiada predyspozycje do samodzielnego prowadzenia prac badawczych oraz wykazał się dobrą znajomością obszarów związanych z uczeniem maszynowym i konstruowaniem algorytmów.

Biorąc pod uwagę wysoki poziom merytoryczny rozprawy oraz dobry dorobek publikacyjny, składam wniosek o wyróżnienie rozprawy. Moje uwagi mają charakter polemiczny i mam nadzieję, że Doktorant odniesie się do nich w trakcie obrony.

Pod względem merytorycznym i formalnym praca spełnia wymagania stawiane rozprawom doktorskim i jest zgodna z wytycznymi ustawy o stopniach naukowych i tytułach naukowych oraz o stopniach i tytułach w zakresie sztuki.

**Wnoszę o dopuszczenie opiniowanej dysertacji do dalszych etapów przewodu doktorskiego, w tym do jej publicznej obrony.**

