

**Recenzja rozprawy doktorskiej mgr Krzysztofa Sopyły:  
“Metody akceleracji procesu klasyfikacji z wykorzystaniem  
Support Vector Machines”<sup>1</sup>**

## 1 Wstęp

Rozprawa doktorska pana magistra Krzysztofa Sopyły poświęcona jest zadaniu klasyfikacji przy użyciu metody wektorów wspierających. Metoda ta, w skrócie SVM od angielskiego terminu *Support Vector Machine*, jest techniką uczenia maszynowego pozwalającą na analizowanie danych i rozpoznawanie wzorców w celu klasyfikacji lub regresji. Istotą metody SVM jest konstrukcja hiperpłaszczyzny, której zadaniem jest rozdzielenie danych, należących do dwóch przeciwnych klas, z możliwie największym marginesem zaufania. Chociaż metoda SVM została zapoczątkowana pracami z lat sześćdziesiątych poprzedniego stulecia, jest ona nadal intensywnie studiowana przez wielu badaczy. Algorytmy oparte na SVM zostały z powodzeniem zastosowane w rozmaitych dziedzinach, takich jak klasyfikacja tekstów, bioinformatyka, nauki rolnicze, medycyna, fizyka, i wielu innych.

## 2 Układ pracy

Recenzowana rozprawa doktorska jest podzielona na dwie części: "Wprowadzenie teoretyczne i przegląd literatury" oraz "Algorytmy i eksperymenty".

Pierwsza część pracy ma charakter teoretyczny i stanowi wprowadzenie do problematyki maszynowego uczenia się przy użyciu metody SVM. Autor omawia tu również architekturę CUDA (ang. *Compute Unified Device Architecture*), umożliwiającą programowanie równoległe na procesorach wielordzeniowych, głównie kartach graficznych. Architektura ta odgrywa kluczową rolę w dalszej części rozprawy.

---

<sup>1</sup>Recenzja przygotowana dla Rady Wydziału Informatyki Polsko-Japońskiej Wyższej Szkoły Technik Komputerowych.

Pierwsza część pracy nie zawiera wyników własnych autora. Pokazuje jednak, że pan mgr Sopyła posiada szeroką wiedzę dotyczącą rozważanej problematyki.

Druga część rozprawy, zawierająca oryginalne wyniki autora, ma charakter implementacyjno-eksperymentalny. Składa się ona z dwóch rozdziałów.

Pierwszy rozdział części drugiej (Rozdział 5) poświęcony jest problemowi uczenia klasyfikatora opartego na metodzie SVM z jądrem nieliniowym. Problem ten jest obliczeniowo trudny i dlatego prowadzi do bardzo czasochłonnych obliczeń w przypadku dużych danych. Możliwym rozwiązaniem tej trudności jest wykorzystanie architektury pozwalającej za zrównoleglenie obliczeń. Przykładem takiej architektury, wykorzystanej przez autora recenzowanej rozprawy, są procesory graficzne, w skrócie GPU (ang. *Graphics Processing Units*).

Wykorzystanie procesora graficznego w procesie uczenia algorytmu SVM nie jest pomysłem nowym i było opisane w wielu pracach. Autorskim pomysłem pana magistra Sopyły jest użycie rzadkich formatów macierzy do obliczeń wartości funkcji jądra. Zabieg ten umożliwił znaczne zwiększenie rozmiarów przetwarzanych danych przy zachowaniu dotychczas osiąganych współczynników przyspieszenia. Dodatkowym wkładem autora jest modyfikacja algorytmu SMO (ang. *Sequential Minimal Optimization*), pełniącego kluczową rolę w metodzie wektorów wspierających. Modyfikacja ta pozwoliła na łatwe implementowanie nowych funkcji jądra oraz na zastosowanie różnych formatów macierzy rzadkich. Oryginalnym wynikiem recenzowanej pracy są dwie nowe metody mnożenia macierzy rzadkiej przez gęsty wektor. Wszystkie zaproponowane rozwiązania zostały przetestowane na ogólnie dostępnych zbiorach danych, co pozwoliło na porównanie prezentowanych metod z dotychczasowymi podejściami.

Drugi rozdział części drugiej (Rozdział 6) poświęcony jest metodom uczenia wersji liniowej metody SVM. Wkładem autora jest tu rozszerzenie metody SGD (ang. *Stochastic Gradient Descent*), co umożliwiło zmniejszenie wrażliwości jakości procesu uczenia od wyboru parametrów początkowych. W rozdziale przedstawiono eksperymenty porównujące zaproponowane metody z metodami już istniejącymi. W odróżnieniu od poprzedniego rozdziału, algorytmy tu przedstawione są algorytmami sekwencyjnymi. Wynika to z faktu, że w przypadku wersji liniowej istnieją metody treningu działające w czasie liniowym i ich zrównoleglenie niewiele by dało.



### 3 Ocena uzyskanych wyników

Bardzo istotnym elementem recenzowanej rozprawy doktorskiej jest część eksperymentalna, mająca na celu oszacowanie jakości przedstawionych rozwiązań.

W części eksperymentalnej związanej z procesorami graficznymi (Rozdział 5) autor wykorzystał publicznie dostępne dane znajdujące się w bibliotece LibSVM. Jako punkt odniesienia do zaproponowanych przez siebie algorytmów, użył dwóch podstawowych implementacji metody wektorów wspierających na procesorach graficznych: cuSVM oraz GTSVM. Uzyskane wyniki są bardzo obszerne i dlatego nie będę ich wszystkich przytaczał. Najważniejszym wnioskiem wynikającym z przeprowadzonych eksperymentów jest fakt, że zastosowanie rzadkich formatów danych pozwala na przetwarzanie znacznie większych zbiorów niż porównywane metody cuSVM i GTSVM.

W części eksperymentalnej związanej z jądrem liniowym (Rozdział 6) położono nacisk na porównanie zaproponowanych rozwiązań z rozwiązaniami znanymi z literatury ze względu na dwa aspekty:

- Wpływ parametrów początkowych na jakość procesu klasyfikacji.
- Porównanie poziomu zbieżności zaproponowanych metod.

Tak jak poprzednio, eksperymenty przeprowadzono na publicznie dostępnych danych. Okazało się, że w obu aspektach algorytmy przedstawione w pracy nie odbiegają, a często są lepsze od tych znanych z literatury.

### 4 Uwagi końcowe

Recenzowaną rozprawę oceniam jako ciekawą i oryginalną. Przedstawione w niej wyniki, stanowią niewątpliwą postępowanie w dziedzinie metody wektorów wspierających. Wyniki te mają duże znaczenie praktyczne. Nowe algorytmy przedstawione w pracy zostały przez autora zaimplementowane i przetestowane na ogólnie dostępnych danych. Pozwoliło to na ich szczegółowe porównanie z algorytmami już istniejącymi.

Autor wykazał się bardzo dobrą znajomością prezentowanej problematyki. Poszczególne rozdziały zawierają dokładny opis aktualnego stanu wiedzy, której dotyczą. Cytowana literatura jest obszerna i zawiera wszystkie najbardziej istotne pozycje związane z dziedziną metody wektorów wspierających.

Silną stroną pracy jest jej tematyka. Uczenie maszynowe, a w szczególności metoda wektorów wspierających, to obecnie bardzo ważne dziedziny

sztucznej inteligencji. Dlatego jestem przekonany, że wyniki pracy nie stanowią zamkniętego etapu, ale będą mogły być dalej rozwijane.

Słabym punktem pracy jest język. W szczególności chodzi mi o tendencję do nadużywania słów angielskich. W polskim tekście naukowym, który pretenduje do tekstu poważnego, nie powinny pojawiać się słowa typu 'solver', czy 'listing'. Bardzo niefortunne jest użycie słów języka angielskiego w tytule rozprawy. Ponadto pewne wyrazy obce, akceptowalne w języku polskim, mają dużo lepsze polskie odpowiedniki. Przykładem jest używane nagminnie przez autora słowo 'akceleracja', które ma znacznie lepszy polski odpowiednik 'przyspieszenie'.

Reasumując, uważam, że recenzowana praca spełnia wymogi stawiane rozprawom doktorskim i wnioskuję o dopuszczenie autora do dalszych etapów przewodu doktorskiego.

