

Prof. dr hab. inż. Stanisław Osowski

Politechnika Warszawska

Warszawa, 6.10.2013

Ocena poprawionej wersji rozprawy doktorskiej mgr Wiesława Chmielnickiego  
pt. **„Efektywne metody selekcji cech i rozwiązywania problemu  
wieloklasowego w nadzorowanej klasyfikacji danych”**

Po zapoznaniu się z poprawioną wersją rozprawy doktorskiej mgr W. Chmielnickiego **podtrzymuję moją końcową pozytywną ocenę rozprawy i dopuszczam ją do obrony.** Poprawiona wersja zawiera ustosunkowanie się do większości moich uwag, choć nie wszystkich.

W szczególności nie znalazłem w pracy żadnej krzywej ROC klasyfikatorów zaproponowanych przez doktoranta (uwaga ósma). Pole pod tą krzywą świadczy o jakości proponowanego rozwiązania. Spodziewam się na obronie prezentacji takiej krzywej przynajmniej dla jednego przypadku.

Nie znalazłem w pracy odpowiedzi na uwagi nr 10 i 17

10. *Na stronie 63 autor omawia heurystyczny algorytm generacji kodów za pracą Dietricha [34] twierdząc, że jest skuteczny dla dużych  $N$ . Aż prosi się, aby pokazać wynik takiej generacji dla wzmiankowanej w pracy liczby klas równej 1000.*

17. *Szkoda, że wyniki badań ograniczono do liczby klas równej 50. Z taką liczbą klas metoda OVO radzi sobie również zupełnie przyzwoicie i nie ma realnej potrzeby stosowania innych rozwiązań. Wyższości rozwiązań proponowanych przez autora można byłoby spodziewać się dopiero przy bardzo dużej liczbie klas. Tego jednak autor rozprawy nie sprawdził eksperymentalnie i nie udowodnił.*

W tabeli 5.4 zestawiono parametry „optymalne” zastosowanych klasyfikatorów. Intrygujące są dla mnie wartości parametru gamma. Są one bardzo małe, co oznacza zastosowanie bardzo szerokich funkcji jądra. W praktyce oznacza to, że optymalny klasyfikator jest zbliżony do liniowego.

Nie znalazłem odpowiedzi na uwagę nr 19:

19. Mam wątpliwości co do selekcji cech. Jak ten rozdział ma się do poprzedniego? Czy w rozdziale 3 (dotyczącym klasyfikacji wieloklasowej) nie zastosowano optymalnego zestawu cech? Jak w takim razie wytłumaczyć lepsze wyniki klasyfikacji przy braku selekcji w problemie białek? W tabeli 4.1 najlepsza dokładność dla białek to 54,2% przy najlepszej selekcji, natomiast bez selekcji(?) w rozdziale 3 w tabeli 3.10 mamy 57,1%. Co zatem daje selekcja cech?

**Podsumowując, podtrzymuję mój sformułowany w pierwotnej opinii wniosek o dopuszczenie mgr inż. Wiesława Chmielnickiego do publicznej obrony rozprawy doktorskiej.**

