

Dr hab. Zbysław Tabor, prof. PK

Kraków, 15 października 2016

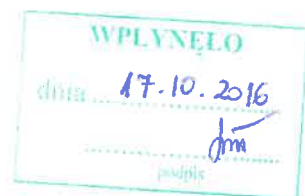
Instytut Teleinformatyki

Wydział Fizyki, Matematyki i Informatyki

Politechnika Krakowska

ul. Warszawska 24, 31-155 Kraków

ztabor@pk.edu.pl



Recenzja

rozprawy doktorskiej Pana mgr inż. Pawła Pławiaka pt.:

**„AUTOMATYCZNE DIAGNOZOWANIE DYSFUNKCJI MIĘŚNIA SERCOWEGO NA PODSTAWIE
ANALIZY SYGNAŁU ELEKTROKARDIOGRAFICZNEGO (EKG) PRZY WYKORZYSTANIU
SYSTEMU EWOLUCYJNO-NEURONOWEGO”**

wykonanej pod kierunkiem prof. dra hab. inż. Ryszarda Tadeusiewicza
na Wydziale Elektrotechniki, Automatyki, Informatyki i Inżynierii Biomedycznej
Akademii Górniczo-Hutniczej w Krakowie

I. Ogólna charakterystyka podjętego przez Doktoranta problemu badawczego

Recenzowana rozprawa dotyczy zagadnienia klasyfikacji dysfunkcji serca na podstawie analizy sygnału EKG, pochodzącego z odprowadzenia MLII. W celu rozwiązania tego zagadnienia Doktorant zaimplementował i wytrenował systemy oparte na metodach uczenia maszynowego z nadzorem. Ponieważ systemy tego typu "uczą" się klasyfikacji na podstawie próbek danych sklasyfikowanych przez ekspertów, pierwszym etapem rozwiązania postawionego przez Doktoranta problemu było zbudowanie bazy danych. Na tym etapie zostały podjęte pierwsze decyzje projektowe - Doktorant założył, że analizowane będą sygnały EKG z jednego odprowadzenia (MLII) oraz, że systemy klasyfikujące będą podejmowały decyzje na podstawie 10 sekundowych zapisów EKG. Nie są to założenia kluczowe, a wynikają z charakterystyki danych udostępnionych w otwartej bazie MIT-BIH Arrhythmia, z której Doktorant zdecydował się skorzystać. Obserwując aktualne trendy w środowiskach zajmujących się przetwarzaniem różnego rodzaju sygnałów uważam, że jest to pomysł dobry ponieważ znakomicie ułatwia obiektywne porównanie wyników

uzyskiwanych przez różne grupy badawcze. Ostatecznie baza danych użyta w badaniach Doktoranta zawierała 744 dziesięciosekundowe fragmenty sygnału EKG pochodzące od 29 pacjentów. Sygnały zostały przyporządkowane do 17 klas (15 rodzajów zaburzeń, rytm normalny oraz rytm z rozrusznika), przy czym dla niektórych klas sygnały pochodziły tylko od jednego pacjenta, a liczba fragmentów w klasie była nie mniejsza, niż 10.

Zaprojektowana metoda przetwarzania danych jest poprawnie dobrana do zagadnienia i obejmuje typowe dla tematyki uczenia maszynowego etapy tzn. wstępne przetwarzanie, ekstrakcja cech, redukcja wymiarowości danych, trening i testowanie powiązane - w celu wyeliminowania przeuczenia - w procedurze walidacji krzyżowej. Na etapie przetwarzania wstępnego zarejestrowane sygnały są ewentualnie przeskalowane liniowo. Etap ekstrakcji cech to przejście z dziedziny czasu do dziedziny częstotliwości. Etap redukcji wymiarowości realizowany jest przy zastosowaniu algorytmu genetycznego sprzężonego z etapem uczenia maszynowego. W jego efekcie z sygnału w dziedzinie częstotliwości wycinane są pewne częstotliwości - kryterium wycięcia jest polepszenie skuteczności systemu klasyfikującego. Ostatni etap jest realizowany przez sześć różnych systemów klasyfikujących, którymi są albo standardowe klasyfikatory, jak SVM, albo złożone, wielowarstwowe zbiory klasyfikatorów noszące cechy systemów wykorzystywanych w metodach "deep learning".

Stosując coraz bardziej złożone systemy klasyfikujące, Doktorant demonstruje ich coraz większą skuteczność w problemie klasyfikacji "zorientowanym na pacjenta", używając sformułowania Doktoranta. To "zorientowanie na pacjentów" definiuje obszar zastosowań, dla którego opracowane przez Doktoranta metody oraz uzyskane wyniki są szczególnie ważne. "Zorientowanie na pacjenta" to metoda tworzenia zbiorów treningowych i testowych w walidacji krzyżowej, zgodnie z którą w zbiorze treningowym zawierającym próbki zaburzenia X są dane od pacjenta Y, ale w zbiorze testowym takich próbek już być nie musi. Oznacza to, że klasyfikator powinien rozpoznawać zaburzenie X u pacjenta Z nawet wtedy, jeśli uczył się na zaburzeniu X pacjenta Y. Obszarem zastosowań metodyki opracowanej przez Doktoranta są zatem badania przesiewowe zaburzeń serca, z czułością jako kryterium użyteczności systemu klasyfikującego. W tym kontekście oparcie analizy na jednym odprowadzeniu jest zaletą, ponieważ urządzenia używane w testach przesiewowych powinny być jak najprostsze. Podobnie, użycie w analizie stosunkowo krótkich odcinków sygnału nie nakłada zbyt dużych ograniczeń na moc obliczeniową urządzeń, na których przetwarzanie miałyby być wykonywane, co znowu jest korzystne dla zastosowań w testach przesiewowych.

Podjęty temat badawczy oceniam jako ważny i doskonale wpisujący się we współczesne trendy rozwoju biocybernetyki i inżynierii biomedycznej, w szczególności w obszar wykorzystania uczenia maszynowego w diagnostyce medycznej.

II. Ogólna charakterystyka rozprawy

Rozprawa składa się z siedmiu rozdziałów z których pierwszy przedstawia przyczyny, które skłoniły Doktoranta do podjęcia badań w zakresie automatycznej diagnostyki dysfunkcji serca, a ostatni zawiera podsumowanie uzyskanych wyników. Układ rozprawy jest przejrzysty i odpowiada standardowemu układowi publikacji naukowych, dzięki czemu przechodząc do lektury kolejnych rozdziałów czytelnik ma jasność co do intencji, które towarzyszyły Doktorantowi w kolejnych etapach badań. W drugim rozdziale Doktorant przedstawia aktualny stan wiedzy w obszarze podjętej tematyki oraz przeprowadza dyskusję wyników uzyskanych przez inne grupy badawcze, podkreślając mocne i słabe strony metodologii zastosowanych w innych badaniach. W rozdziale trzecim Doktorant szczegółowo prezentuje narzędzia używane do zbudowania własnych systemów klasyfikujących, poczynając od EKG, poprzez metody przetwarzania sygnału, kończąc na metodach oceny „jakości” systemu klasyfikującego. Prezentacja nie skupia się na konkretnych zastosowaniach, które są przedmiotem rozprawy, ale buduje szeroką perspektywę z której można wyciągnąć wnioski o możliwościach prezentowanych narzędzi. Rozdział czwarty to kontynuacja prezentacji metodyki badań: tym razem Doktorant skupia się na opisie przeprowadzonych eksperymentów i wyjaśnia szczegółowo w jaki sposób „cegiełki” zaprezentowane w rozdziale trzecim służą do zbudowania w pełni funkcjonalnych systemów klasyfikujących. W rozdziale czwartym zaprezentowane są zatem pseudokody zaprojektowanych systemów oraz diagramy blokowe – uważna lektura tej części daje pełne wyobrażenie o sposobie funkcjonowania zaprojektowanych systemów. Rozdział piąty to prezentacja wyników, począwszy od surowych danych, przez etap przetwarzania wstępnego, po wyniki klasyfikacji. Rozdział szósty to dyskusja uzyskanych wyników, która przede wszystkim dowodzi skuteczności zaprojektowanych systemów klasyfikujących, z których systemy zaprezentowane w Eksperymentach 4 i 6 uważam za oryginalne osiągnięcie Doktoranta, które nadają pracy walor rozprawy doktorskiej. Zaprojektowane systemy osiągnęły bardzo wysoką czułość, przy aż 17 klasach sygnału EKG.

III. Uwagi ogólne do rozprawy

Praca jest zredagowana starannie, ale przy dużej (może zbyt dużej) objętości Doktorant nie ustrzegł się błędów językowych. W opisie eksperymentów jest sporo powtórzeń. Z pewnością opis metodologii jest dzięki temu bardziej precyzyjny, ale też – niestety – bardziej nużący. W części pracy, w której Doktorant referuje stan wiedzy jest sporo cytatów. Niektóre strony rozprawy są wręcz w całości wypełnione cytataми. Dowodzi to bez wątpienia, że Doktorant ma spory zasób wiedzy i informacje o źródłach wiedzy niezbędne do prowadzenia badań w tematyce rozprawy. Nie jest jednak jasne dlaczego Doktorant nie stara się przekazać tej wiedzy swoimi słowami.

W rozprawie Doktorant przeciwstawia sobie metodę klasyfikacji „zorientowaną na pacjenta” i metodę „zorientowaną na zaburzenia” przy czym stara się wykazać, że zastosowana w rozprawie metoda „zorientowana na pacjenta” jest bardziej poprawna. Według mnie nie jest to podejście słuszne. Metoda „zorientowana na pacjenta”, użyta przez Doktoranta jest odpowiednia do badań przesiewowych. Metoda „zorientowana na zaburzenie” na podstawie danych pacjenta X jest uczona rozpoznawania zaburzeń pacjenta X, a więc ma zastosowanie w medycynie spersonalizowanej. Obie metody nie stoją więc w opozycji, a są komplementarne – pacjent u którego w badaniach przesiewowych zdiagnozowano zaburzenia stosując metodę „zorientowaną na pacjenta” powinien być kontrolowany w toku terapii przez urządzenie z zaimplementowaną metodą „zorientowaną na zaburzenia”.

Doktorant stosuje redukcję wymiarowości opartą o algorytm genetyczny (np. Ref. 175 i Ref. 202), uzyskując dwukrotną redukcję liczby cech. Czy Doktorant rozważał inne metody redukcji wymiarowości?

Nie jest jasne, jak tworzone są zbiory uczące i testowe dla genetycznych algorytmów selekcji cech i dla głębokich zespołów klasyfikatorów. Poprawnie zbiory uczące powinny być konstruowane z odpowiedzi udzielanych przez pierwszą warstwę po podaniu na jej wejście danych uczących i z opisu algorytmów wynika, że tak właśnie jest. Tymczasem w tekście Doktorant pisze np. na stronie 114 „...Zadaniem algorytmu genetycznego było odrzucenie błędnych głosów [...] na podstawie błędów w zbiorach testowych...”. Ta – w mojej opinii – sprzeczność powinna być wyjaśniona.

III. Uwagi szczegółowe do rozprawy

1. Na stronie 18 Doktorant wyraża dość zdecydowany, ale nie wiadomo jak uzasadniony sąd o wynikach pracy innych grup badawczych „...artykułów badawczych [...] jest bardzo mało i nie stoją one na wysokim poziomie...”.

-
2. Na stronie 24 Doktorant pisze, nie popierając swoich słów odnośnikiem do literatury „...większość naukowców uważa, że najlepiej w tej dziedzinie sprawdza się transformata falkowa...”.
 3. Również na stronie 24 pojawia się niejasne dla mnie zdanie „Tak wyliczone cechy mają wielką zaletę, ponieważ są odporne na zmiany oznaczeń punktów odniesienia”. O jakie punkty odniesienia chodzi?
 4. Na stronie 26 Doktorant wymienia klasteryzację, jako jedną z metod klasyfikacji. Klasteryzacja jest typowo rozumiana jako metoda uczenia maszynowego bez nadzoru, a klasyfikacja jako uczenie z nadzorem.
 5. Na stronie 30 Doktorant pisze, że w rozprawie współczynnik ACC jest równy współczynnikowi SEN, czego nie potwierdzają wyniki przedstawione w rozdziale piątym.
 6. Na stronie 55: etap normalizacji i standaryzacji są poprzedzone redukcją wzmocnienia i redukcją składowej stałej. Wydaje się, że w obu przypadkach są to operacje nadmiarowe.
 7. Na stronie 57 Doktorant pisze, że „...Gęstość widmowa opisuje, jak dużo sygnału [...] występuje na jednostkę szerokości pasma...”, co jest definicją dość nieprecyzyjną.
 8. Na stronie 58 w algorytmie 3.1 używany jest „okienkowany periodogram”, nigdzie nie zdefiniowany.
 9. Na stronie 64 na liście metod kodowania brakuje kodowania liczbami zmiennoprzecinkowymi, używanego przez Doktoranta. Na liście metod krzyżowania brakuje krzyżowania arytmetycznego, używanego w rozprawie.
 10. Na stronie 65 Doktorant pisze, że „...Macierze pomyłek dla [...] SVM i KNN są wektorami...Natomiast dla sieci neuronowych macierze [...] są 17 wymiarowe...” Wektor może być macierzą, ale macierz nie jest wektorem. Dlaczego macierze dla sieci neuronowych są 17-wymiarowe (tzn. mają 17 indeksów)?
 11. Na stronie 78 Doktorant cytuję opinię: „...Jakość [...] klasyfikatorów składowych musi być znacząco lepsza od klasyfikatorów losowych...” AdaBoost jest kontrprzykładem – zespół słabych klasyfikatorów, niewiele lepszych, niż losowe może być bardzo skutecznym klasyfikatorem.
 12. Na stronie 80 w algorytmie 3.3 symbol K jest niezdefiniowany.
 13. Na stronie 81: warunek przyporządkowania systemu klasyfikującego do zbioru metod „deep learning” oparty na liczbie ścieżek CAP, która dla metod „głębokich powinna być większa od 2. Czy nie chodzi o długość ścieżek CAP, a nie ich liczbę?
-

14. Na stronie 85 Doktorant pisze o kryterium krzywej ROC. Krzywa ROC służy do wyboru progu dyskryminacji. Nie jest jasne, w którym momencie i jak ten próg (np. dla SVM) jest wybierany. Jeśli rzeczywiście analiza krzywej ROC została przeprowadzona, w pracy należałoby zaprezentować konkretny przykład.
15. Na stronie 91 Doktorant podaje listę wielu parametrów, które nie są zdefiniowane w pracy – należałoby podać odnośnik do dokumentacji Matlaba.
16. Z opisu przedstawionego przez Doktoranta wynika, że każdy klasyfikator pierwszej warstwy w zespole zaprojektowanym w Eksperyment 4 udziela 17 odpowiedzi, z których tylko jedna ma wartość '1', a pozostałe mają wartość '0'. Tymczasem na Rysunku 4.4 klasyfikator SVM_1 ma na wyjściu dwie wartości '1'.
17. Nie jest jasne, co chciał Doktorant wykazać, przeprowadzając Eksperyment 5.
18. Na stronie 140 Doktorant pisze o zespole klasyfikatorów, który ma dwie warstwy wyjściowe. Rozumiem, że warstwa wyjściowa jest w rzeczywistości tylko jedna.
19. Nałożenie wykresów na Rysunkach 5.5 do 5.9 nie ułatwia ich interpretacji.

V. Wnioski

Oceniając część rozprawy składającą się z rozdziałów 2 i 3 stwierdzam, że Doktorant przeprowadził analizę źródłową podjętego tematu badawczego w zakresie świadczącym o ogólnej wiedzy teoretycznej Doktoranta w dyscyplinie biocybernetyki i inżynierii biomedycznej.

Oceniając część rozprawy składającą się z rozdziałów 4-7 stwierdzam, że Doktorant przedstawił oryginalne rozwiązanie problemu badawczego oraz zademonstrował umiejętność samodzielnego prowadzenia pracy naukowej.

Podsumowując, stwierdzam, że rozprawa doktorska Pana mgr inż. Pawła Pławiaka pt.: „Automatyczne diagnozowanie dysfunkcji mięśnia sercowego na podstawie analizy sygnału elektrokardiograficznego (EKG) przy wykorzystaniu systemu ewolucyjno-neuronowego” spełnia warunki określone w art. 13 ust. 1 Ustawy o stopniach i tytule naukowym i wnioskuję o dopuszczenie jej do publicznej obrony.

Biorąc pod uwagę zakres oraz szczegółowość zaplanowanych i przeprowadzonych przez Doktoranta eksperymentów wnioskuję również o wyróżnienie rozprawy.

Taber Wibista