



UNIwersytet
JAGIELLOŃSKI
W KRAKOWIE

dr hab. Paweł F. Góra, prof. UJ
Wydział Fizyki, Astronomii i Informatyki
Stosowanej UJ
+48 (12) 664 4566
e-mail: pawel.gora@uj.edu.pl

Kraków, 29 września 2024

Wydział

Fizyki

Astronomii

i Informatyki

Stosowanej

Recenzja pracy doktorskiej pana mgr. inż. Mateusza Ozimka

Zastosowanie wybranych metod entropowych w badaniach EKG

przygotowanej pod opieką prof. dr. hab. Jana Jacka Żebrowskiego

i dr. hab. Teodora Buchnera, prof. PW

Podsumowanie: Recenzowana rozprawa, choć nie jest wolna od błędów, spełnia warunki stawiane pracom doktorskim i może być podstawą dopuszczenia doktoranta do dalszych etapów przewodu.

Wstęp

Praca pana mgr. inż. Mateusza Ozimka poświęcona jest zastosowaniu metod entropowych do analizy i kwalifikacji szeregów czasowych pochodzących z zapisów EKG. Jest to typowa praca z zakresu zastosowania metod wywodzących się z fizyki statystycznej do opisu i analizy zjawisk fizjologicznych (oraz patologicznych w znaczeniu klinicznym). Ponieważ choroby układu krążenia stanowią współcześnie jedną z najważniejszych przyczyn zgonów, tematyka tej pracy jest ważna, gdyż badania takie - potencjalnie - mogą doprowadzić do opracowania nowych testów diagnostycznych.

Praca ma charakter oddzielnej rozprawy, choć autor niejednokrotnie odwołuje się do swoich poprzednich prac na ten temat (referencje [14], [100]).

Treść pracy

Rozdział 1 przedstawia cel i motywacje rozprawy.

Rozdział 2 stanowi wstęp, w którym autor tłumaczy, co to jest EKG, jakie są charakterystyczne elementy sygnału EKG i jakie mają one znaczenie, a także

ul. prof. Stanisława

Łojasiewicza 11

PL 30-348 Kraków

tel. +48(12) 664-48-90

fax +48(12) 664-49-05

e-mail:

wydzial.fais@uj.edu.pl

opisuje różne choroby serca manifestujące się w zapisie EKG. Muszę przyznać, że rozdział ten, a zwłaszcza podrozdziały opisujące poszczególne patologie, są trudne do zrozumienia dla nie-lekarza. Głównym wnioskiem, jaki fizyk może wyciągnąć z tego rozdziału, jest konstatacja, iż samo badanie EKG, tradycyjnie interpretowane, choć znane i stosowane od ponad stu lat, nie jest wystarczające do zdiagnozowania pewnych chorób serca. Stanowi to **dobre** uzasadnienie dla rozwijania metod będących zasadniczą częścią niniejszej rozprawy.

W Rozdziale 3 autor wprowadza poszczególne miary, jakich użyje do badania szeregów czasowych, w tym miary oparte na entropii Gibbsa-Shannona. (Trochę mnie martwi, że autor nazywa tę entropię “entropią Shannona”. Tak mówią informatycy, ale my, fizycy, powinniśmy wiedzieć, że Josiah Willard Gibbs znalazł identyczne wyrażenie na entropię układów w stanie równowagi na ponad 40 lat przed Claude’em Shannonem.) Duży nacisk położony jest na dekompozycję entropii na entropię własną i “nową informację”, możliwą do odczytania po zarejestrowaniu każdego ostatniego (w danym momencie) elementu szeregu czasowego oraz na transfer informacji pomiędzy oddziałującymi szeregami czasowymi (znak transferu wskazuje, co wpływa na co i pozwala interpretować powiązania w obrębie postulowanej sieci fizjologicznej).

Nie rozumiem, dlaczego autor poprzedza podawane definicje zastrzeżeniem “definicje podane poniżej [...] przyjmują na wejściu interwały NN”, natomiast “analizy wykonane w rozprawie odnosiły się do interwałów RR” (str. 12). Przecież podane definicje są sformułowane abstrakcyjnie, są słuszne dla dowolnych szeregów czasowych, bez względu na to, jak w praktyce uzyskano ich kolejne elementy. Dla badań prowadzonych przez autora nie ma również znaczenia, że analizowane szeregi dotyczą procesów fizjologicznych nierównomiernie próbkowanych - to mogłoby mieć znaczenie, gdyby autor badał charakterystyki częstotliwościowe swoich szeregów, np. widmo mocy, lub też próbował interpolacji fourierowskiej, ale przecież tego nie robi.

W dalszej części tego rozdziału autor wprowadza metody oparte na entropii próby, nieaddytywnej entropii Tsallisa, wreszcie omawia metody klasyfikacji oparte o lasy losowe i *Support Vector Machines* (nawet nie wiem, jaka jest poprawna polska nazwa tej metody). Zabrakło mi tu wyjaśnienia *dlaczego* autor postanowił użyć entropii Tsallisa. Jest wiele obiektów matematycznych, których tu i ówdzie można użyć, ale za każdym razem powinno się uzasadniać, czemu to się robi, o ile rzecz nie jest oczywista. Można by na przykład spekulować, że entropia Gibbsa-Shannona jest właściwa dla szeregów stacjonarnych, czyli w pewnym sensie dla układów w stanie równowagi, a ponieważ żywy człowiek w stanie równowagi nie jest, użycie wielkości nierównowagowych może, choć nie musi, mieć sens (nb, to by być może wyjaśniało, dlaczego w przypadku entropii Tsallisa lepiej sprawdzały się dłuższe okna czasowe).

W Rozdziale 4, moim zdaniem jednym z najważniejszych w całej pracy, autor omawia dane medyczne i sposób ich przygotowania. Doktorant posługiwał się szeregami czasowymi z publicznie dostępnych (po zarejestrowaniu) baz danych oraz szeregami dostarczonymi przez Instytut Kardiologii w Aninie. Te trzy grupy surowych danych różniły się między sobą, wymagały więc wstępnego opracowania i zestandaryzowania, aby w jednolity, zautomatyzowany sposób mogły być poddane testom interesującym autora. W pewnym miejscu (str. 72) doktorant wprost zauważa, że taka sama analiza, jak w jego wcześniejszej publikacji, ale przeprowadzona “na innym podzbiorze danych i po innym ich przygotowaniu”, dała w recenzowanej rozprawie inne wyniki, niż w tamtej publikacji. Jak w bardzo wielu ludzkich przedsięwzięciach, interesujące autora szeregi czasowe uzyskane zostały w sposób heterogeniczny, miały różny format i były opatrzone różnymi informacjami dodatkowymi. Trzeba je więc pre-procesować, aby nadawały się do dalszej analizy. Na pierwszy rzut oka nie jest to zagadnienie, któremu warto poświęcić cały rozdział pracy doktorskiej **z fizyki**. Jest to wrażenie błędne. Wręcz przeciwnie, w czasach rosnącego znaczenia uczenia maszynowego i Sztucznej Inteligencji, jednolitość formatu i jakość danych często stanowią o sukcesie lub porażce całego przedsięwzięcia. Należy pochwalić autora za zwrócenie należytej uwagi na ten punkt.

Rozdział ten zawiera też sporo informacji “technicznych” odnośnie do wyboru odprowadzenia aparatu EKG i metod wyznaczania poszczególnych charakterystycznych elementów zapisu EKG, omówionych w Rozdziale 2. Autor testował dwie procedury numeryczne, aby ostatecznie zdecydować się na jedną z nich. Bardzo dużo uwagi autor poświęca problemowi optymalnego doboru okna czasowego, tak, aby z jednego okna można było wyciągnąć dostatecznie dużo informacji, a żeby przy tym w oknie nie manifestowała się niestacjonarność badanego szeregu. Rozdział zawiera też szereg innych informacji potrzebnych przy konstrukcji metod entropowych.

Rozdziały 5, 6, 7 prezentują główne ilościowe wyniki pracy, czyli efekty zastosowania poszczególnych testów entropowych do (wstępnie przygotowanych) szeregów czasowych z analizowanych baz danych. Wyników szczegółowych jest tak dużo, że nie sposób je tu wymieniść, a nie-lekarzowi trudno wypowiadać się na temat znaczenia klinicznego poszczególnych wyników. Autor zdaje sobie sprawę z tego nadmiaru, gdyż w podsumowaniu pisze, że “może to być swoista klęska urodzaju, gdyż nie jest łatwo wskazać jeden konkretny parametr, który umożliwi wskazanie czym różni się konkretne schorzenie” (str. 109).

Rozdział 8 zawiera podsumowanie uzyskanych wyników, Rozdział 9 przedstawia możliwości dalszego rozwoju przeprowadzonych przez doktoranta badań.

O ile dobrze interpretuję wyniki recenzowanej pracy, przeprowadzone przed doktoranta badania wciąż nie pozwalają na opracowanie testu klinicznego służącego do zaklasyfikowania pojedynczego nowego (spoza analizowanych baz!) pacjenta do grupy zdrowych lub cierpiących na jedno bądź drugie, bądź trzecie schorzenie. Autor dowodzi co prawda, na zasadzie *Proof of Concept* (str. 75), że konstrukcja takiego klasyfikatora jest możliwa, ale do osiągnięcia “zdolności operacyjnej” jest wciąż jeszcze bardzo daleko.

Krytyka

Omawiając treść pracy sformułowałem już kilka uwag krytycznych o mniejszym znaczeniu. Teraz chciałbym przedstawić dwie poważniejsze uwagi.

1. W zastosowaniu metod entropowych kluczowym elementem jest wyliczanie (przybliżonych) wartości poszczególnych prawdopodobieństw i prawdopodobieństw warunkowych na podstawie posiadanych szeregów czasowych. Niestety, opis, jak doktorant to robił, jest bardzo pobieżny. Na str. 17 czytamy, że “wybrano estymator oparty na metodzie histogramowania” i że będzie on “szerzej opisany” w rozdziale 4.4. Jednak jedyną nową rzeczą, jakiej można się stamtąd dowiedzieć, jest że “liczba przedziałów histogramu ustawiona była na 10”. Autor parokrotnie podkreśla, że “metoda histogramowania” jest obliczeniowo znacznie szybsza, niż inne metody estymacji prawdopodobieństw, nie porównuje jednak wyników działania różnych metod, choć jest świadom, że metody histogramowe mają swoje wady, co “może mieć przełożenie na otrzymywane wyniki” (str. 39). Moim zdaniem należało porównać wyniki działania różnych metod estymacji prawdopodobieństw przynajmniej dla kilku wybranych serii danych. To, że jakaś metoda jest szybsza od innych, jest oczywiście ważkim argumentem na rzecz jej stosowania, pod warunkiem jednak, że wyniki, jakie ona daje, są co najmniej porównywalne z wynikami innych metod. Czy tak jednak jest, tego z recenzowanej rozprawy nie sposób się dowiedzieć.

2. Idąc dalej tym tropem, autor dużo uwagi poświęca optymalnemu doborowi rozmiaru okna, na jakie dzieli analizowane szeregi czasowe, by uniknąć niestacjonarności. Rozumiem, że dla każdego okna, prawdopodobieństwa oraz miary entropowe estymowane są niezależnie, ale co dalej? Czy jeśli szereg dla jednego pacjenta daje się podzielić na, powiedzmy, pięć okien, to czy w ostatecznej analizie są one traktowane jako pięć niezależnych serii czasowych, czy też miary entropowe z różnych okien dla tego samego pacjenta są jakoś łączone, a jeśli tak, to jak? W rozprawie nie jest to wyjaśnione.

Oczekuję, że doktorant **szczegółowo i dogłębnie** wyjaśni te kwestie w trakcie obrony.

Słaba prezentacja

Praca nie jest też wolna od licznych błędów w zakresie prezentacji. Powierzchność opisów zastosowanych procedur, poza tym, że prowadzi do niepewności odnośnie do meritum poszczególnych elementów pracy, jest **także** błędem prezentacyjnym. Innym przykładem tego rodzaju jest niezwykle zagmatwany opis drzew decyzyjnych i lasów losowych zamieszczony na str. 25-26, a przykład drzewa decyzyjnego zamieszczony na str. 26 doprawdy jest zupełnie niezrozumiały. Jest to bardzo dziwne, gdyż drzewa decyzyjne i lasy losowe są powszechnie dziś stosowanymi elementami uczenia maszynowego i Sztucznej Inteligencji i naprawdę nietrudno znaleźć ich klarowne, a jednocześnie wyczerpujące opisy.

Skandalicznym wręcz błędem w prezentacji są zamieszczone rysunki, kompletnie pozbawione objaśnień co oznaczają ich poszczególne elementy. Cóż te rysunki właściwie mówią? Na Rysunkach 4.6-5.50 i 6.1-7.9 różne serie danych reprezentowane są przez prostokąty o różnych kolorach, to jedno jest jasne. Ale co mają oznaczać pozostałe elementy wykresów? Dlaczego prostokąty mają różne wysokości i co te wysokości oznaczają? Co oznaczają "granice błędu"? Co wreszcie oznaczają symbole rombów (diamentów) znajdujące się na niektórych rysunkach? W pracy [14] przeczytałem, że pozioma linia w poprzek każdego prostokąta oznacza medianę dla danej grupy. Można się domyślać, że to samo znaczy ona w recenzowanej pracy doktorskiej, ale doktorant nie uznał za stosowne wyjaśnić czytelnikom, co poszczególne elementy oznaczają (nawiasem mówiąc, w pracy [14] elementy rysunków, poza oznaczeniami średniej i mediany, też nie są objaśnione). W tej sytuacji wymienione rysunki można uznać za graficzne elementy o charakterze dekoracyjnym i znikomej zawartości informacyjnej. Równie dobrze mogłoby ich nie być. Informacja opisowa bądź podana w tabelach zupełnie by wystarczyła; ufam, że doktorant prawidłowo przeprowadził wszystkie testy statystyczne, a to, że postanowił

przedstawić swoje wyniki w sposób graficznie zupełnie niezadawalający jest przykre, ale merytorycznie ma mniejsze znaczenie.

Problemem być może mniejszej wagi, ale bolesnym dla czytelnika, jest język, styl, jakim posługuje się autor pracy. Podam kilka najbardziej uderzających przykładów:

- “Pełna analiza ryzyka wymaga prospektywnego badania, a takie dane nie są częste” (str. 9). Po pierwsze nie wiem, co w tym kontekście miałyby oznaczać przymiotnik “prospektywny”, a po drugie, jaki jest logiczny i składniowy związek pomiędzy “badaniem” a “takimi danymi”, o których mowa w tym zdaniu?
- “Głównym powodem była potrzeba uniknięcia i tak niewielkiej, liczby wykorzystanych zapisów” (str. 30 u dołu). Nawet pomijając nieprawidłowo postawiony przecinek, kompletnie nie rozumiem tego zdania.
- Długie zdanie na początku rozdziału 4.6, od słów “W celu analizy wpływu” aż do “przedstawiony na rysunku 4.6” nie ma orzeczenia.
- Na str. 34 czytamy, że “opracowane zostały dwa algorytmy w języku programowania *Python*”. Mogę zapewnić autora, że algorytmy opracowuje się w głowie lub na kartce papieru. W języku programowania można je co najwyżej implementować.
- U góry str. 39 czytamy: “Z kolei dla okien coraz dłuższych zwiększał się procent okien, dla których spełnione było kryterium stacjonarności”. Biorąc pod uwagę całą uprzednią dyskusję, doktorant chyba chciał napisać “dla których **nie spełnione** było kryterium stacjonarności”?

Jest też wiele innych niezręczności składniowych, które pomijam. Poza tym praca pełna jest terminów nieledwie żargonowych, które zapewne są oczywiste dla badaczy na codzien zajmujących zagadnieniami związanymi z tematyką recenzowanej rozprawy, więc odruchowo ich używają, ja sądzę jednak, że powinno się ich unikać w pracy potencjalnie przeznaczanej dla szerszego grona czytelników. Zastosowana interpunkcja pozostawia wiele do życzenia. Te same wielkości zapisywane są raz czcionką prostą, innym razem kursywą. W tej samej

tabeli (np. Tabela 4.8) widzimy notację "0,0002", a zaraz obok "10⁻⁴". Wszystko to, niestety, dowodzi bardzo małej staranności, jaką włożono w przygotowanie tekstowej i szerzej, prezentacyjnej, części pracy.

Podsumowanie i wnioski

Przedłożona praca doktorska, choć nie jest wolna od wad, porusza ważne i aktualne zagadnienia, pozwalające za pomocą metod wywodzących się z fizyki statystycznej lepiej zrozumieć fizjologię pracy serca i dyskryminować pomiędzy **grupami** pacjentów zdrowych i z różnymi schorzeniami. W przyszłości może doprowadzić to do opracowania opartych na metodach entropowych testów klinicznych, pozwalających zakwalifikować **pojedynczego** pacjenta do danej grupy szybciej i lepiej, niż istniejące metody, jednak w tym zakresie najwyraźniej jesteśmy dopiero na początku drogi. Zastosowanie entropii nierównowagowych (entropii Tsallisa) nie przyniosło wyników istotnie lepszych od zastosowania dekompozycji entropii Gibbsa-Shannona. Można powiedzieć, że wynik tej części pracy jest negatywny, co najwyżej słabo pozytywny, ale to też jest ważna obserwacja. Autor wykazał się przy tym dogłębną znajomością literatury przedmiotu w zakresie metod entropowych, metod statystycznych i istniejących analiz szeregów czasowych pochodzących z zapisów EKG i medycznych wniosków z nich płynących. Za najsilniejszą stroną pracy uważam mocne zwrócenie uwagi na konieczność standaryzacji surowych danych i ujednoczenie protokołu, który do tego prowadzi. Jest to, moim zdaniem, obserwacja ważniejsza, niż szczegółowe wyniki dotyczące poszczególnych serii danych. Być może nie umiem docenić znaczenia tej grupy wyników, gdyż jest ono ukryte za bardzo słabą prezentacją.

Biorąc wszystko powyższe pod uwagę, stwierdzam, że w moim odczuciu rozprawa pana mgr. inż. Mateusza Ozimka spełnia ustawowe i zwyczajowe wymogi stawiane pracom doktorskim i wnoszę o dopuszczenie pana Mateusza Ozimka do dalszych etapów przewodu doktorskiego.

A handwritten signature in blue ink, appearing to read "P.F. Góra". The signature is written in a cursive style with a large initial 'P' and 'G'.

P.F.Góra