



Prof. dr hab. Daniel Krzysztof Wójcik

Warszawa, 13.1.2022

Pracownia Neuroinformatyki

Instytut Biologii Doświadczalnej

im. Marcelego Nenckiego PAN

02-093 Warszawa, ul. Pasteura 3

Recenzja rozprawy doktorskiej mgr Katarzyny Stępień pt.

„Analysis of selected physiological networks using information theory methods”.

Magister Katarzyna Stępień w swojej rozprawie doktorskiej analizuje jednoczesne rejestracje EKG i EEG z wykorzystaniem metod teorii informacyjnych. Badania są rozszerzeniem wcześniejszych prac Luki Faesa i wsp. Autorka rozszerzyła wcześniejsze podejście o analizę parametrów wykorzystywanych w opisie klinicznym aktywności serca i zastosowała zaproponowane metody w celu porównania trzech grup osób: zdrowych, z zaburzeniami zachowania w czasie fazy snu REM, oraz z chorobą Parkinsona. Wyniki analizy są spójne z obecną wiedzą. Autorka postuluje możliwość wykorzystania swoich metod w diagnostyce obu chorób w praktyce klinicznej.

Recenzowana rozprawa zawiera prawie 100 stron maszynopisu w ośmiu rozdziałach, odnosi się do prawie 100 publikacji i jest bogato ilustrowana około 60 rysunkami. Praca jest przygotowana schludnie, pod względem wizualnym prezentuje się dobrze, natomiast została przygotowana przy założeniu druku w formacie A4. Ostateczny druk w mniejszym formacie utrudnia czytanie pracy, zwłaszcza analizę niektórych rysunków i wzorów. Rozprawa jest dobrze językowo napisana po angielsku, co potencjalnie zwiększa jej oddziaływanie w środowisku międzynarodowym. Jest napisana bardzo starannie, nie znalazłem żadnej literówki i jedynie dwa drobne błędy językowe.

Autorka rozpoczyna od przedstawienia motywacji swojej pracy i głównych hipotez. Autorka zauważa, że nie ma dobrych metod diagnostycznych choroby Parkinsona, która prowadzi do zmian zarówno w aktywności mózgu jak i serca. Stąd pomysł wykorzystania związków między różnymi sygnałami kardio- i neurologicznymi jako biomarkerów tej choroby. Z kolei zaburzenia zachowania w czasie fazy snu REM często poprzedzają chorobę Parkinsona stąd chorzy na tą chorobę są naturalną grupą do zbadania, poza osobami zdrowymi a chorymi na chorobę Parkinsona. Motywacja pracy jest rozsądna. Autorka wybrała technikę analizy opracowaną kilka lat temu, gdzie tworzone są formalne grafy, których węzłom przypisuje się sygnały niosące pewną informację fizjologiczną, a następnie bada transfer informacji pomiędzy tymi zmiennymi.

Rozprawa ma tradycyjny układ wprowadzenia, metod (rozdziały 2-4), wyników (rozdziały 5-6) i dyskusji (rozdział 7). W rozdziale drugim autorka przedstawia elementy fizjologii mózgu i serca istotne dla zrozumienia

pracy, w szczególności omawia sygnały EEG i ECG, na podstawie których konstruuje sygnały pochodne, które będzie badać. Pochodne EEG to sygnał EEG filtrowany w pięciu pasmach, w przypadku ECG wybrano sześć zmiennych wykorzystywanych w analizie zmienności rytmu serca. Omówiono też podstawowe fazy snu, polisomnografię, a także chorobę Parkinsona, zaburzenia zachowania w czasie fazy snu REM, które są prekursorem choroby Parkinsona, oraz zmiany w EEG i zmienności rytmu serca obserwowane u pacjentów z chorobą Parkinsona. W rozdziale trzecim autorka omawia wybrane miary teorii informacyjnej, które wykorzystuje w późniejszej analizie. Ten rozdział był dla mnie nieczytelny i brakowało w nim definicji badanych pojęć, piszę o tym więcej dalej.

W rozdziale 4 autorka wprowadza zmienne wyznaczone na podstawie EEG, których szeregi czasowe analizuje, oraz pasma EEG, w których filtruje ten sygnał do analizy. Omówiono dokładnie szczegóły przetwarzania wstępnego i przygotowania danych do analizy. Autorka w analizie zakłada, że badane sygnały mają łączny rozkład gaussowski i pokazuje w jaki sposób wzrokowo to ocenia. Następnie autorka wskazuje w jaki sposób definiuje badane sieci oraz jak je wizualizuje dla pojedynczych zestawów danych i dla populacji, po czym omawia stosowane testy statystyczne.

W rozdziale piątym autorka testuje na danych symulacyjnych kilka metod pozwalających na wykrywanie połączeń między węzłami sieci – direct transfer, transfer entropy, binned transfer entropy – i pokazuje, że przekaz bezpośredni (ang. *direct transfer*) najlepiej odtwarza połączenia w wybranych przykładowo układach. Autorka testuje w wielu symulacjach jaka długość sygnału jest konieczna do poprawnego wykrycia połączenia w danych przy konkretnej sile połączenia. Zbadano również wpływ błędnej identyfikacji załamek R na identyfikację siły połączeń w sieciach definiowanych aktywnością różnych funkcji zmienności rytmu serca.

Rozdział szósty stanowi najważniejszą część rozprawy. Autorka stosuje w nim przetestowane wcześniej miary do sygnałów zarejestrowanych u pacjentów z trzech grup osób: zdrowych, z zaburzeniami zachowania w czasie fazy snu REM, oraz z chorobą Parkinsona. Rozdział rozpoczyna definicja przyjętego potoku analitycznego który składa się z podstawowej analizy wybranych szeregów czasowych, po których następuje ich analiza teorii informacyjnej, zakończona analizą sieciową zestawów zmiennych. Następnie autorka omawia bazę danych, które analizuje, sprawdzanie jakości danych i wstępne przetwarzanie w tym sprawdzenie łącznej gaussowskości rozkładów otrzymanych zmiennych. W rozdziale 6.3 autorka analizuje zmienne charakteryzujące zmienność rytmu serca oraz poszczególne pasma w sygnale EEG poprzez średnie sygnałów w poszczególnych grupach oraz poprzez podstawowe miary teorii informacyjnej. Ostatni rozdział z wynikami, 6.4, omawia aspekty sieciowe badanych sygnałów w tym walkę autorki z bogactwem miar w celu uzyskania sumarycznych charakterystyk, które mogłyby wskazać istotne różnice między badanymi grupami.

Ostatni rozdział rozprawy stanowi obszerna dyskusja, w której autorka obszernie omawia zarówno napotkane wyzwania, decyzje, które musiała podjąć (jak na przykład wybór zmiennych), jak i uzyskane wyniki i ich znaczenie. Rozdziały 6 i 7 zawierające wyniki i dyskusję są dobrze napisane i jasno wskazują, że autorka jest dojrzałą badaczką, która radzi sobie z trudnymi wyzwaniami analitycznymi, ale jest też świadoma ograniczeń swoich metod i wyzwań, jakie stawiają przed nią dane, z którymi musi pracować.



Główne wyniki tej pracy to:

- wykonanie szeregu analiz teoriiinformacyjnych sygnałów otrzymanych na podstawie ECG i EEG, które mogą informować o interakcji układu krwionośnego i nerwowego;
- wskazanie postępujących zmian tych interakcji w grupie pacjentów z zaburzeniami zachowania w czasie fazy snu REM oraz z chorobą Parkinsona, w szczególności zmniejszenie transferu informacji przy postępującej chorobie;
- wskazanie, że takie zmiany nie są widoczne, kiedy poszczególne sygnały są analizowane oddzielnie, natomiast z perspektywy sieciowej widać różnice.

Praca jest istotnie interdyscyplinarna i zrobiła na mnie ogólnie dobre wrażenie. Autorka wykazała się znajomością zagadnień z teorii informacji, analizy szeregów czasowych, fizyki, elektrofizjologii, kardiologii. Mam jednak zastrzeżenia, głównie jeśli chodzi o elementy szeroko rozumianego wprowadzenia i definicji używanych obiektów. W zasadzie cały rozdział 3 jest dla mnie nieczytelny. Na przykład, czytając rozprawę nie jest dla mnie jasne, co to jest H_Y w równaniu 1 – wiem, co to jest entropia rozkładu, wiem też jak policzyć różne entropie z szeregów czasowych, np. techniką zanurzeniową albo bezpośrednio. Ale nie wiem, co autorka ma w tym miejscu na myśli. Podany obiekt jest niejednoznaczny. Moje pierwsze skojarzenie kiedy czytam "[] *information content of Y as the information obtained while sampling the process up to the present time n*" jest takie, że bierzemy Y_1 do Y_n i traktujemy je jako próbkę z jakiegoś rozkładu. Trzeba powiedzieć, w jakich binach itd. Czy o to chodzi autorce? Czy może raczej zakładamy, że sygnał jest stacjonarnym procesem gaussowskim, z danych szacujemy parametry i liczymy entropię analitycznie zgodnie ze wzorem 18?

Dalej, Y_n^- w równaniu 2 jest również niezdefiniowane, co utrudnia zrozumienie dalszych definicji. Co prawda podane referencje a także lektura rozdziału 3 do końca pomagają częściowo zrozumieć podane definicje, ale czytelnik, a zwłaszcza recenzent, nie powinien się domyślać z jakiej definicji akurat teraz autorka chce korzystać. Ten rozdział był dla mnie napisany niejasno. Uważam, że powinien być rozszerzony, informacje z końca rozdziału 3.5, gdyby je podać na początku rozdziału 3, pokazywać po drodze przykłady explicite, to może byłoby czytelniejsze. Na pewno definicje wszystkich obiektów powinny być podane odpowiednio wcześniej. **Oczekuję, że na obronie autorka wprowadzi te wszystkie definicje i wyjaśni dokładnie co to jest u niej $H(y)$, co to jest i jak liczy $H(Y_n)$, oraz kolejne definicje w rozdziale 3.**

W rozdziale 4.1.2 nie rozumiem pomysłu testowanie łącznej normalności rozkładu przez wybór wartości kombinacji liniowej z czterech przedziałów. Zauważmy, że jeżeli mamy dowolny ciąg liczb x_n , a następnie przemnożymy każdy element przez tą samą liczbę k , to rozkład x_n i kx_n ma taki sam charakter, gaussowski albo nie. Z powodzeniem można było przyjąć więzy $a_1+\dots+a_6=1$ i liczyć 16 powtórzeń przy tym założeniu. Równomierny rozkład powinien próbować na $n-1$ wymiarowej sferze. Poza tym prawdopodobnie lepiej byłoby skorzystać z rozkładu na składowe niezależne, przynajmniej w tym kierunku bym szukał niegaussowskości.

Rozdział 5 zaczyna się od wprowadzenia dwóch symulacji, nazwanych przez autorkę A i B. Niestety, nie są one dobrze zdefiniowane. Nie wiadomo, co to znaczy 'directed coupling between the nodes'. Nawet przegląd cytowanej literatury nie wskazuje jednoznacznie, co autorka mogła mieć na myśli. Na przykład w cytowanej



pracy [69] autorzy podają różne przykłady, sprzężonych odwzorowań Henona albo procesów gaussowskich sprzężonych bardzo konkretnymi równaniami. Tutaj w ogóle nie wiemy, z jakimi układami dynamicznymi mamy do czynienia. **Oczekuję, że na obronie autorka zdefiniuje układy, które tutaj symulowała.**

Nie przekonuje mnie zaproponowana metoda oceny łącznej normalności rozkładów, nie jest też dla mnie jasne, że jest potrzebna. Jak autorka pokazała w rozdziale 5.2.3 (rys. 20), dane z jej symulacji B nie mają łącznego rozkładu Gaussa, bo wymagałoby to w szczególności, żeby każdy z pokazanych rozkładów był Gaussowski, a tak oczywiście nie jest. Autorka pisze, że podany rozkład jest bliski do rozkładu Gaussa pokazując, że przechodzi on test autorki wzrokowego podobieństwa do Gaussa według rys. 22. Moim zdaniem doktorantka nie pokazała, że ma zmienne rozłożone prawie gaussowsko. Raczej pokazała, że 1) jej test łącznej normalności nie działa, ale też, że 2) chociaż ma układ zmiennych, które nie mają łącznego rozkładu Gaussa, to pomimo tego metoda detekcji połączeń może być stosowana (Fig. 25). Jest to częsty przypadek sytuacji, kiedy twierdzenie działa nawet wtedy, gdy założenia nie są spełnione. Niestety, nie wiadomo wtedy, jakie konkretnie warunki powinny być spełnione, żeby twierdzenie jeszcze działało i dalej nie pozwala stwierdzić, czy zastosowanie metody przekazu bezpośredniego do danych doświadczalnych jest czy nie jest uprawnione. Niestety, jest to typowa sytuacja w większości analiz trudnych danych doświadczalnych.

Przyjęta kolorystyka prezentacji wyników na wykresach sieciowych (rys. 12, 13 itp) mnie nie przekonuje. W sytuacji kiedy „siła” połączenia zmienia się od 0 do wartości dodatniej, wybrałbym natężenie jednego koloru do opisu tej wartości, np. niebieskiego. Przejście od niebieskiego do żółtego przez zielony nie pozwala łatwo odczytać tej siły. Procent znaczących połączeń na wykresach sieciowych jest pokazywany grubością strzałek, ale różnice między 1% a 100% są minimalne, sugerowałbym znaczące zwiększenie tych różnic w grubości.

Pomimo przedstawionych powyżej uwag krytycznych moja ogólna ocena pracy pani mgr Katarzyny Stępień jest pozytywna. Uważam, że wyniki i dyskusja są bardzo solidne, same analizy obszerne i wykonane poprawnie. Praca stanowi ona istotny przyczynek do badania potencjalnych biomarkerów choroby Parkinsona. **Uważam, że rozprawa doktorska Pani mgr Katarzyny Stępień spełnia zwyczajowe i formalne warunki stawiane rozprawom doktorskim i dlatego wnoszę do Rady Naukowej Dyscypliny Nauki Fizyczne Politechniki Warszawskiej o przyjęcie rozprawy i dopuszczenie autorki do dalszych etapów przewodu doktorskiego.**

Prof. dr hab. Daniel K. Wójcik