

## Recenzja rozprawy doktorskiej mgra Pawła Piaseckiego pt. „Application of selected artificial intelligence methods to time series analysis”

### 1. Uwagi ogólne nt. problematyki rozprawy i jej celów

Niniejsza recenzja jest sporządzona w odpowiedzi na pismo Dziekana Wydziału Matematyki i Informatyki Uniwersytetu im. Adama Mickiewicza w Poznaniu prof. UAM dr hab. Krzysztofa Dyczkowskiego z lipca br. Rozprawa jest zatytułowana „Application of selected artificial intelligence methods to time series analysis” – z polskim odpowiednikiem pt. Zastosowanie wybranych metod sztucznej inteligencji do analizy szeregów czasowych i jest kolekcją 6 prac naukowych (artykułów w czasopismach lub prac konferencyjnych) uzupełnionych autorskim wprowadzeniem. Wszystko jest napisane w języku angielskim.

Tak jak pisze sam autor jego rozprawa porusza kilka problemów badawczych dotyczących analizy danych w postaci szeregów czasowych oraz innych złożonych reprezentacji danych. Rozważane metody dotyczą zarówno statystyki, w tym dość zaawansowanych metod analizy funkcjonalnej i nietypowej regresji w analizie spektrografów, jak i uczenia maszynowego w zakresie klasyfikatorów, zespołów modeli predykcyjnych, oceny ważności cech i sieci neuronowych. Jest to szeroki obszar zainteresowania badawczego. W moim zrozumieniu w większości ta tematyka może być powiązana z tzw. eksploracją złożonych reprezentacji danych (odpowiedniki terminu angielskiego data mining), lecz także związana z współczesnymi metodami sztucznej inteligencji. Podsumowując poruszana tematyka mieści się w obszarze wskazanej **dyscypliny informatyka**.

Krótko komentując skupienie uwagi na problemach związanych z danymi w postaci szeregów czasowych (w zasadzie jest to dobrze widoczne w 4-ech pierwszych pracach tego cyklu) – pomimo, że tradycyjnie jest to typ danych rozważany w statystyce oraz wielu powiązanych obszarach zastosowań, są one również często rozważane w eksploracji danych, uczeniu maszynowym lub szerzej sztucznej inteligencji jako dane uporządkowane czasowo, często równocześnie wielowymiarowe. Są one rozważane w zadaniach nie tylko predykcji wartości zmiennej (co jest bardziej typowe dla klasycznej analizy szeregów czasowych w zakresie ang. terminu time series forecasting), lecz także klasyfikacji, grupowania, wyszukiwania anomalii w danych, poszukiwania wzorców występujących tylko we fragmentach danych, segmentacji szeregów, indeksacji danych, porównywania i badania korelacji zmiennych wielowymiarowych i szeregów. Mogą też występować w kontekście strumieni danych i innych złożonych typów danych rozważanych obecnie w zastosowaniach systemów uczących się. Jest w widoczne w wielu pozycjach z zakresu tzw. data mining i uczenia maszynowego. Można to dostrzec w zagranicznych książkach z powyższego zakresu jak i artykułach przeglądowych (patrz np. silnie cytowana pozycja Esling P. Agon V. Time series data mining. ACM Computing Surveys). Zgodnie z moją najlepszą wiedzą takie szersze i nowoczesne spojrzenie jest też obecne w badaniach prowadzonych przez zespół statystyki matematycznej i analizy danych UAM.

Pan mgr Paweł Piasecki w wstępnym rozdziale rozprawy nie sformułował jawnie tzw. hipotezy badań. Ponadto tematyka wybranych publikacji jest trochę „szeroko rozrzucona”. Można jednak za rozdziałem 1.2 (tytuł ang. objectives of the thesis) powtórzyć, że największa część badań poświęcona jest analizie porównawczej wielu miar podobieństwa szeregów czasowych i przeprowadzenia bardzo obszernych studiów eksperymentalnych porównujących ich przydatność do klasyfikacji szeregów. To zadanie jest ważne i rozważane we współczesnej literaturze. Kolejny artykuł ma na celu wprowadzenie nowej miary współzależności wielowymiarowych danych funkcjonalnych – konkretnie specjalnego uogólnienia dyskretnego współczynnika Prokrusta na punkty wielowymiarowe w funkcjonalnej analizie danych – co jest także stosunkowo nowym obszarem współczesnej analizy danych. Ponadto doktorant przedstawił prace z zakresu zastosowań – statystyczna analiza chemometryczna danych z spektroskopii oceny jakości soków jabłkowych oraz końcową propozycję rozszerzania zespół klasyfikatorów, o nazwie RANNs, wykorzystującego losowe przetworzenie działania wielu sieci neuronowych – co jest mało związane z motywem przewodnim szeregów czasowych, lecz mieści się w zakresie współczesnych metod uczenia maszynowego.

Podsumowując uważam, że powyższa tematyka i cele bardziej szczegółowe są interesujące poznawczo i są warte prowadzenie badań oraz realizacji doktoratu.

## 2. Ocena formy i struktury rozprawy

Recenzowana rozprawa jest napisana w języku angielskim. Została przedstawiona jako oprawiony maszynopis obejmujące cykl 6 współautorskich artykułów, rozdziału wprowadzającego (5 stron), dwóch streszczeń (angielskie i polskie) oraz kopii oświadczeń współautorów w zakresie udziału w przygotowaniu każdego z artykułu (co jest wymogiem formalnym). Łącznie liczy ona 122 stron + powyższe oświadczenia. Ponadto dostarczono autoreferat pt. Summary of academic accomplishments obejmującego podsumowanie aktywności naukowej doktoranta, informacje o jego wykształceniu, studiach doktoranckich i opis zainteresowań badawczych (mocno zgodnych z motywacją i celami rozprawy) oraz listę projektów badawczych, w których doktorant uczestniczył.

Cykl publikacji obejmuje następujące pozycje:

1. Górecki T., Piasecki P.: An Experimental Evaluation of Time Series Classification Using Various Distance Measures *Archives of Data Science, Series A*, 5(1), 2018
2. Górecki T., Piasecki P.: A Comprehensive Comparison of Distance Measures for Time Series Classification. In: Steland A., Rafajłowicz E., Okhrin O. (eds) *Stochastic Models, Statistics and Their Applications. Springer Proceedings in Mathematics & Statistics*, vol 294, pages 409–428, 2019.
3. Tomasz Górecki, Maciej Łuczak, Paweł Piasecki: An exhaustive comparison of distance measures in the classification of time series with 1NN method. *Expert systems with applications* (w recenzji)
4. Górecki T., Piasecki P. Generalization of the Procrustes coefficient to functional data. *Communications in Statistics - Simulation and Computation*, 3(49): 808–816, 2019
5. Włodarska K., Piasecki P., Lobo-Prieto A., Pawlak-Lemańska K., Górecki T., Sikorska E. Rapid screening of apple juice quality using ultraviolet, visible, and near infrared spectroscopy and chemometrics: A comparative study. *Microchemical Journal*, 164, 2021.
6. Górecki T., Piasecki P. RANdom Neural Networks (RANNs): a new general classifier inspired by Random Forest. *PP-RAI'2019 Conference Proceedings*. Department of Systems and Computer Networks, Faculty of Electronics, Wrocław University of Science and Technology, 68–72, 2019.

W maszynopisie umieszczone są kopie tych publikacji jako kolejne rozdziały. Według zamieszczonych oświadczeń współautorów udział Pana mgr P. Piaseckiego jest pomiędzy 20% do 80% (w pracach dwuautorskich z promotorem), czyli często jest tzw. wiodącym / pierwszym autorem – co w mojej ocenie jest ważne przy tej formie złożenia rozprawy doktorskiej, gdyż oczekiwałbym właśnie dużego wkładu własnego doktoranta.

Na ogół ciągle jesteśmy przyzwyczajeni, że rozprawa doktorska to spójnie napisany autorski zwarty maszynopis. Tutaj mamy odmienną sytuację – doktorant przedstawił tzw. spójny cykl publikacji. Osobiście spotkałem się z tą formą głównie zagranicą, przede wszystkim w krajach skandynawskich, lecz jest to oczywiście od pewnego czasu dopuszczalne w Polsce. Zgodnie z wymogami rozprawę doktorską może stanowić spójny tematycznie cykl prac oryginalnych opublikowanych w recenzowanych czasopismach naukowych, jak i rozdziałach w książkach.

W przypadku przedstawionego cyklu, doktorant we wstępie starał się przedstawić taki spójny obraz tematyki prac – tj. wykorzystanie metod sztucznej inteligencji w analizie danych w postaci szeregów czasowych. Spełnia to wymóg powiązania tematycznego prac, o którym mówi stosowna ustawa. Przy czym osobiście sędzę, że bezpośrednio o danych w postaci szeregów czasowych możemy mówić w stosunku do czterech pierwszych prac. Pracę 5 odbieram raczej jako ciekawy przykład zastosowania, gdzie była potrzebna bardziej zaawansowana analiza statystyczna. Wtedy główne dane, to są zapisy spektrografów z urządzeń laboratoryjnych czyli sygnału widma, który jest raczej uporządkowany jednostkami długości fal, częstotliwości itd. Stwierdzenie, że to jest rodzaj szeregu czasowego jest dla mnie mocno „naciągane”. Ostatnia szósta praca nie dotyczy danych w postaci szeregów czasowych,

lecz typowych współczesnych danych obrazowych dla nowoczesnych głębokich sieci neuronowych, a dokładniej ocena eksperymentalna jest wykonana na bardzo popularnej bazie zdjęć różnych przedmiotów MINST, co nie ma nic wspólnego z szeregami czasowymi. Oczywiście można twierdzić, że przedstawiona koncepcja losowego zespołu sieci to rozwiązanie w ramach metod uczenia nadzorowanego w uczeniu maszynowym i tak szukać tzw. spójnego tematycznie powiązania prac.

Patrząc na same miejsca publikacji, to są one różnej jakości. *Microchemical journal* wydawany przez Elsevier ma wysoki impact factor 3.5 z tzw. 70 pkt min. – przy czym jest on raczej w obszarze zastosowań innych dyscyplin. *Communications in statistics - simulation and computation* jest międzynarodowym czasopismem z IF około 1.1 i tzw. 40 pkt. *Archives of Data Science* – niemieckie czasopismo hybrydowe powiązane z Niemieckim Towarzystwem Klasyfikacyjnym (GfKL) bez impact factor. Druga praca w serii Springer'a to rozdział konferencyjny z 14th Workshop on Stochastic Models, Statistics and Their Applications, Drezno. Szósta praca jest materiałach głównej polskiej konferencji sztucznej inteligencji PP-RAI (są recenzowane, co jako jeden z głównych przewodniczących tej konferencji mogę potwierdzić). Artykuł w *Expert Systems with Applications* (Elsevier, impact factor 6.9 bardzo dobrze rozpoznawalny w środowisku sztucznej inteligencji) – lecz tekst jest ciągle w recenzji (na podstawie własnego doświadczenia jest to wyjątkowo trudna redakcja z strasznie długim okresem oczekiwania).

W przypadku mojego macierzystego Wydziału nie ma zbyt wielu przypadków przedstawienia cyklu publikacji jako rozprawy doktorskiej, lecz nie do końca precyzyjnie i formalnie oczekuje się, że cykl obejmuje co najmniej 3 artykuły naukowe wydane w czasopiśmie z listy *Journal Citations Reports (JCR)*, ew. prace z konferencji CORE A. Lecz Wydział Matematyki i Informatyki UAM nie przedstawił mi swoich szczegółowych wymagań – dlatego nie powinienem brać tego pod uwagę. Patrząc do rekordu bibliometrycznego pana Piaseckiego w *Google Scholar* nie są one zbyt widzialne – 3 obce cytowania.

Nie jestem drobiazgowy w analizie i zastanowieniu się czy wszystkie prace cyklu odnoszą się do oczekiwań, że są to czasopisma lub rozdziały w monografiach oceniane jako te, które znajdują się na odpowiednich ministerialnych listach. W przypadku niektórych wydziałów, można spotkać się z dodatkowymi wymaganiami do rodzaju czasopism, łącznej liczby punktów (np. wg. mojego macierzystego Wydziału zaleca się wyższą łączną liczbę punktów). W zleceniu recenzji nie przekazano mi podobnych wytycznych WMI UAM, dlatego wole skupić się na ocenie merytorycznej zawartości prac i uznać, że liczba publikacji jest wystarczającą. Poza tym ocena doktoratu ma zupełnie inną charakterystykę niż ocena dorobku habilitacyjnego. Ponadto cytując artykuł z ustawy prawo o szkolnictwie wyższym i nauce, od kandydata na stopień doktora powinniśmy oczekiwać „minimum 1 artykuł naukowy opublikowany w czasopiśmie naukowym lub w recenzowanych materiałach z konferencji międzynarodowej, które w roku opublikowania artykułu w ostatecznej formie były ujęte w wykazie“ (...) ministerialnym, a to Pan mgr Paweł Piasecki spełnia.

Powracając do spojrzenia na strukturę maszynopisu mam mieszane odczucia wobec rozdziału wstępnego. Osobiście bym wolał bardziej obszerniejszy tekst zawierający silniejsze motywacje do prowadzonych badań (mogę je dopiero odnaleźć w poszczególnych artykułach) i przede wszystkim szerszego kontekstu tzn. osadzenie własnych propozycji w aktualnym stanie najbardziej podobnych prac w światowej literaturze przedmiotu i lepszego opisanie powiązań 4 różnych wątków tematycznych w ramach cyklu. Szczęśliwie, pomimo braku takiego „parasola metodycznego” cyklu, same cele są dość jasno (lecz krótko) wyrażone w podrozdziale 1.2.

Jestem także mocno zaskoczony zawartością kolejnego podrozdziału 1.3. (zatyłowanego *research methodology*) – jest to w zasadzie opis zastosowanego sprzętu komputerowego, stworzonego oprogramowania, itd. W mojej opinii to jest ewidentnie bardzo krótki opis warunków eksperymentów a nie jakakolwiek metodologia badawcza – patrz prace z metodologii prowadzenia badań naukowych. Dlatego przynajmniej powinien być inaczej zatyłowany. Niewątpliwie „prawdziwy” rozdział nt. stosowanych metod w pracach i uzasadnienia ich wyboru mógłby być napisany.

Ponadto podczas czytania maszynopisu zrodziła mi się wątpliwość do podobieństwa trzech pierwszych artykułów. Wszystkie one dotyczą porównania miar podobieństwa / odległości dla szeregów czasowych oraz użycia ich dla klasyfikacji z wykorzystaniem klasyfikatora k-NN dla dużej liczby tzw. benchmarkowych danych szeregów czasowych z repozytorium UCR. Ogólnie mówiąc

artykuły różnią się one liczbą rozważanych miar, danych oraz pewnymi różnicami w analizie zbiorczych wyników (trochę inne testy statystyczne, rodzaje wykresów, itd.) a wiele części po prostu się powtarza. W przypadku pisania tradycyjnego maszynopisu – całość by była dużo bardziej zwartym i spójnym jednym rozdziałem i czytelnik nie byłby tak zmęczony.

Mam też odczucie braku krótkiego syntetycznego podsumowania rozprawy – odpowiednik rozdziału Podsumowanie i wnioski. Myślę, że nawet dla cyklu artykułów można było oczekiwać autorskiego spojrzenia dotyczące przedstawionego w rozprawie problemu badawczego oraz syntetyczne, wypunktowane wnioski i komentarz, co do osiągnięć w zakresie czterech różnych tematycznie kierunków badawczych. Przynajmniej mam takie doświadczenie z doktoratów zagranicznych, gdzie własne rozdziały wokół cyklu prac są obszerniejsze.

Na koniec dodam także, że streszczenie polskojęzyczne jest zdecydowanie za krótkie. W moim odczuciu należało w nim opisać dokładniej własne propozycje metod oraz rozszerzyć informacje o głównych wnioskach z oceny eksperymentalnej, propozycji metod, itd.

W dalszym punkcie recenzji przechodzę do dokładniejszego omówienia głównych osiągnięć pana mgra Pawła Piaseckiego przedstawionych w omawianych 6 artykułach.

### 3. Osiągnięcia badawcze doktoranta

Cykl publikacji pokazuje, że Pan mgr Paweł Piasecki prowadził badania w kilku grupach problemów:

1. Trzy pierwsze artykuły dotyczą porównania różnych miar odległości szeregów czasowych i pokazania różnic ich przydatności w zadaniu klasyfikacji szeregów
2. Zaproponowania nowej miary zależności wielowymiarowych danych funkcjonalnych, tj. uogólnienia współczynnika Prokrusta (ang. Procrustes coefficient) – artykuł nr 4
3. Pokazanie wykorzystania metod statycznych, głównie przedziałowej wersji PLS regresji, w spektrometrii do oceny które parametry chemiczne wpływają jak jakość soku jabłkowego – artykuł nr 5
4. Wprowadzenia zespołu losowych sieci neuronowych do klasyfikacji złożonych danych – artykuł nr 6

W ramach pierwszej grupy tematycznej obejmujących trzy kolejne artykuły (o numerach 1-3) doktorant interesował się zagadnieniem oceny podobieństwa dwóch szeregów czasowych. Jak słusznie wskazuje sprowadza się to do wyboru właściwej miary odległości pomiędzy dwoma szeregami. Zgodnie w przeglądami miar jest prawie 60 i są na ogół podzielone na cztery kategorie (autorzy przedstawiony prac podążają za dość przekonującą klasyfikacją Eslinga i Agona z cytowanego już przez mnie artykułu w ACM Comput. Surveys). Są one dość zróżnicowane i samo porównanie ich definicji nie jest trywialne. Ponadto jak wskazują doktorant i promotor brakowało eksperymentalnych prac porównawczych. Pomimo, że miary odległości mogą być wykorzystane do różnych zdań, np. grupowania szeregów, wyszukiwania anomalii, zdecydowano się na rozważenie zdania klasyfikacji, gdzie trzeba przypisać pewne etykiety znanych kategorii do nowo prezentowanych szeregów. Patrząc na literaturę eksploracji danych jest to niewątpliwie popularne zadanie. Doktorant zdecydował się rozważyć klasyfikator minimalno-odległościowy KNN w wersji ( $K=1$ ) pierwszego najbliższego sąsiada. Zgadzałbym się z przedstawionymi uzasadnieniami w artykułach, że jest to klasyfikator bezpośrednio zależny od wybranej miary odległości oraz dość skuteczny także w porównaniu z innymi klasyfikatorami jak SVM, zespoły klasyfikatorów lub specjalizowane sieci neuronowe. Prace kilku autorów (np. Rodrigez i Kuncheva 2007) wskazują, że 1NN połączony z elastycznymi odległościami wykorzystującymi ideę DTM (Dynamic Time Warping – gdzie porównujemy szeregi o podobnej strukturze, ale przesunięte w czasie co wymaga transformacji i dopasowania) są bardzo skutecznymi klasyfikatorami. Inni autorzy podkreślali popularność tego rodzaju klasyfikatora w literaturze eksploracji danych. Dlatego można się zgodzić, że taki decyzje doktoranta są uzasadnione.

Zgodnie z zamieszczonymi przeglądami literatury liczba prac porównującymi eksperymentalne użycie różnych miar z tym klasyfikatorem jest bardzo ograniczona (4 prace wg. zestawienia w artykule doktoranta nr 3), a liczba rozważanych miar i benchmarkowych danych szeregów czasowych dość ograniczona. Dlatego doktorant wraz z promotorem zdecydowali się na przeprowadzenie dużo

obszerniejszych badań eksperymentalnych na publicznie dostępnych szeregach z repozytorium UCR (wykorzystywanego przez wielu badaczy). Autorzy chcą w pełniejszy sposób dokonać podziału znanych miar na podgrupy ze względu na ich skuteczność w zadaniu klasyfikacji szeregów oraz dogłębniej przeanalizować różnice pomiędzy nimi i przedstawić rekomendacje do ich wyboru. Trzy kolejne artykuły zawarte w recenzowanej rozprawie różnią się przede wszystkim zakresem prowadzonych eksperymentów. Pierwszy obejmuje porównanie 26 miar na 34 zbiorach testowych, drugi z artykułów 30 miar na 47 zbiorach a ostatni 55 miar na 128 zbiorach (obejmujących ostatnie kolekcje szeregów w repozytorium UCR). Dane – szeregi z UCR mają gotowy podział na część uczącą i testową, a schemat postępowania w badaniach w każdym z artykułów jest w zasadzie taki sami: obliczenie błędu predykcji (zestawienie w tabelach), porównanie ich w postaci zbiorczych wykresów (na ogół pudełkowe box plots), następnie przeprowadzenie globalnych rangowych testów statystycznych (Friedmana) z odpowiednią post-hoc analizą. Takie postępowanie jest w miarę standardowe w obecnych pracach z zakresu klasyfikacji w uczeniu maszynowym. W zależności od artykułu autorzy zastosowali dodatkowe środki analizy, bardzo często podział miar poprzez analizę skupisk (czasami przeprowadzoną trochę subiektywnie, np. za dużo ciekawszy i uzasadniony praktycznie uważam wersję dwukryterialną ranga błędu / czas rys 4 z pierwszego artykułu niż grupowanie po dwóch spojrzeniach na błąd - rys 7 w trzecim artykule) oraz dodatkowe analizy różnic, np. procedury Bergmanna i Hommela w ostatnim artykule. Ten typ postępowania jest zgodny z rekomendacjami literaturowymi, widać że doktorat podążał zwłaszcza ze pracami zespołu Garcia, Herrera z uniwersytetu w Granadzie. Rezultaty analizy wskazują na dominującą przewagę miary odległości wykorzystujących różne modyfikacje DTW. O ile przekonanie o skuteczności miary DTW jest już obecne w literaturze, to tak szeroki eksperyment z wskazaniem na wysoką ocenę innych bardziej zaawansowanych modyfikacji wyprowadzonych z DTW jest oryginalnym osiągnięciem. Przeszukując zasoby internetowe nie znalazłem innych tak obszernych prac eksperymentalnych, choć pewnym ograniczeniem jest powtarzalność schematu i stylu prezentacji oraz wręcz języka pisanie w kolejnych trzech artykułach.

Kolejna rozważana problematyka w recenzowanej rozprawie w artykule o numerze 4 jest ściśle związana w wielowymiarową analizą danych funkcjonalnych. Jest to jeden z działów statystyki zajmujący się obiektami które nie są punktami dyskretnymi lecz funkcjami ciągłymi, w którym dokonał się intensywny rozwój w ostatnich kilkunastu latach i ma ciekawe odniesienia praktyczne. Oprócz rozwijania nowych propozycji matematycznych ma on silne powiązanie do zaawansowanych obliczeń numerycznych i związki z eksploracją danych w ujęciu informatycznym. Przynajmniej mam taką opinię patrząc na to z boku jako informatyk, a nie jako statystyk. Ponadto badania nad wielowymiarową analizą funkcjonalną są od wielu lat z powodzeniem prowadzone w Zakładzie Statystyki Matematycznej UAM, w ramach których zaproponowano wiele uogólnień funkcjonalnych metod statystycznych. W szczególności jak pokazuje to doktorant dotyczyć to może badania współzależności dwóch zbiorów zmiennych wielowymiarowych. We wstępie zamieszonego artykułu zamieszczono sprawnie napisany przegląd literatury związanej z współczynnikami korelacji oraz bardziej zaawansowanych miar współzależności. W rezultacie pokazano potrzebę nowych propozycji miar lub uogólnień istniejących dla wielowymiarowych danych funkcjonalnych – lecz w momencie pisanie tego artykułu liczba propozycji była ograniczonych. Promotor prof. T.Górecki wraz z profesorami M.Krzyśko oraz W.Wołyńskim wcześniej proponowali uogólnienia współczynnika  $\rho_V$  oraz zasad jego testowania a także wprowadzali nowy funkcjonalny współczynnik odległościowy oraz uogólnienia współczynników HSIC i KTA dla lepszego działania w przypadku zależności nieliniowych. Na tym tle doktorant wraz z promotorem proponują nowe uogólnienie współczynnika Prokrusta (ang. Procrustes coefficient). Artykuł nr 4 ma charakter dokonania teoretyczno-eksperymentalnego dla wielowymiarowej analizy statystycznej. Tekst zawiera najpierw dwa typowe rozdziały: opis przekształcenia procesu stochastycznego z wykorzystaniem wybranej liczby tzw. funkcji bazowych. Autorzy rozważają przekształcenie Fouriera oraz tzw. krzywych B-sklejanych różnego stopnia. Na podstawie mojej ograniczonej wiedzy jest to dość typowe podejście do przekształcenia wielowymiarowych szeregów do postaci funkcjonalnej i późniejszego działania na parametrach zapisu macierzowego lub wektorach współczynników. Rozdział 3ci opisuje podstawowe dyskretne sformułowanie współczynnika Prokrusta jako metody oceny współzależności dwóch macierzy. Nie ukrywam, że rozszerzając trochę moją wiedzę o tej klasie metod widzę to jako specjalny współczynnik wywodzący się z uogólnionej analizy Prokrusta, gdzie oryginalne zadanie

polega na właściwym nakładaniu złożonych obiektów w wielowymiarowej przestrzeni, a następnie przesuwanie, obracanie i skalowanie, aby zmaksymalizować kryterium dopasowania. Tutaj RLS(X,Y) jest on prostszy polegający na wykonaniu wielu obliczeń na porównywanych dwóch macierzach i jak pokazali autorzy ma prostą interpretację zakresu swoich wartości. Główne oryginalne dokonanie doktoranta jest przedstawione w podrozdziale 3.2 gdzie wprowadzono wzór na uogólniony współczynnik wykorzystujący całkowanie przekształconych złożeniach funkcji bazowych i ich parametrów. Jak wskazują autorzy jego obliczenie nie jest łatwe (poprzednie uogólnienia współczynników często wykorzystywały tylko wektory parametrów w składanych funkcjach) i licznik ułamka nie może być uproszczony i musi być obliczany numerycznie. Kolejnym dokonaniem doktoranta jest pokazanie praktycznego wykorzystania propozycji współczynnika w badaniu wzrostu twarzoczaszki, których celem jest analiza zmian kształtu poprzez rejestrowanie pozycji charakterystycznych w różnym wieku – wybrano przykład publicznie dostępnych danych nt. zmiany kształtu czaszki samca *Macaca nemestrina* (czyli orientalnych małp Makak) w różnym wieku. Opis analizy (dla dwóch różnych transformacji oryginalnych) jest interesujący i wsparty odpowiednimi permutacyjnymi testami wskazuje na statystycznie znaczącą wartość współzależności wg. nowego współczynnika Prokrusta (znajduję także w tekście artykułu informacje, że taki wynik jest oczekiwany z wcześniejszych analiz tego problemu, czyli ten przykład jest rodzajem oceny skuteczności przedstawionej propozycji). Ciekawą obserwacją są uwagi nt. zmian wartości samego współczynnika w zależności od doboru liczb elementów składowych funkcji bazowych.

Podsumowując te dwa rodzaje problemów badawczych i sposobów ich rozwiązanie – oceniam, że oba są bardzo interesujące oraz wnoszą wkład oryginalnych propozycji do dziedziny eksploracji danych. Ponadto zwłaszcza w pierwszych trzech artykułach widać duży wysiłek doktoranta w implementacji metod oraz przeprowadzenie niezwykle obszernych eksperymentów oraz wyczerpującej statystycznej analizy otrzymanych wyników. Poza tym są bardzo dobrze związane z wskazywaną przez doktoranta tematyką wykorzystania nowych metod wywodzących się z sztucznej inteligencji w eksploracji danych w postaci szeregów czasowych.

Dwa kolejne artykuły zamieszczone w recenzowanej rozprawie, trochę odbiegają od tego ogólnego zakresu. Praca nr 5 jest silnie ukierunkowana w stronę zastosowania metod statystycznych w spektrometrii. Bardziej precyzyjnie jest artykuł wielo-autorski (z wieloma autorami specjalizującymi się ocenie jakości żywności). Celem pracy była ocena i porównanie potencjału różnych technik spektroskopii optycznej do oceny jakości soków jabłkowych. Przedstawiono uzasadnienie prowadzonych badań oraz środowiska przyrządów laboratoryjnych – dwa rodzaje spektrometrów do nieniszczącego wyznaczania parametrów jakościowych soku jabłkowego na podstawie widm. Ze strony metodologicznej wybrano tzw. przedziałowe kalibracyjne modele regresji metodą cząstkowych najmniejszych kwadratów (PLS). Na ile dobrze spojrziałem do źródeł internetowych właśnie ta metoda była często rozważana przez innych autorów do badań chemometrycznych w spektroskopii w bliskiej podczerwieni (widm NIR) – co jest także rozważanie w tym artykule. Zgodnie z moim zrozumieniem tekstu doktorant brał udział we właściwym opracowaniu modeli PLS regresji oraz intensywnej ich kalibracji. Pokazano także wykorzystanie tych modeli do predykcji różnych ważnych parametrów badanych soków, jak np. rozpuszczalna zawartość substancji stałych SSC, kwasowości. Opis (bardzo szczegółowy i specjalistyczny) przeprowadzonej analizy widm wskazuje, że optymalizacja kalibracji parametrów oraz zwłaszcza wybór zmiennych ma duży wpływ na poprawę predykcji modeli na danych testowych. Odebrałem ten artykuł jako wskazanie użyteczności wybranych metod statystycznych do praktycznego problemu analizy danych żywnościowych. Można go także zinterpretować jako pokazanie wiedzy doktoranta w pewnych działach statystycznej analizy danych oraz umiejętności ich praktycznego stosowania oraz udziału w interpretacji wyników.

Ostatnia z prac (nr 6) zawarta w recenzowanej rozprawie doktorskiej dotyczy uczenia maszynowego – uczenia zespołu klasyfikatorów z nową autorską propozycją losowej sieci neuronowej, tj. Random Neural Networks (RANNs). Doktorant przedstawia inspiracje płynące z sukcesu konstruowania zespołów klasyfikatorów, zwłaszcza tzw. bagging (wprowadzonego przez Leo Breimana w 1996 roku) i jego uogólnień do lasów losowych (ang. random forests), które w stabilizuje działanie klasyfikacyjne (np. mniejsza podatność na przeuczenie), zwiększenia zróżnicowania klasyfikatorów składowych oraz działania dla wielu cech, poprzez odpowiednie losowanie podzbiorów cech. Same klasyfikatory tej rodziny nie są aż tak często stosowane z klasycznymi

sieciami neuronowymi (postulat braku stabilności w odpowiedzi na pewne perturbacje w zbiorach uczących). Lecz doktorant we wstępie odnosi do danych wysokowielowymiarowej oraz większego wprowadzanie losowości do struktury i uczenia sieci jako modeli składowych w zespole bagging. W propozycji pana mgra Piaseckiego występują trzy poziomy losowości:

1. Losowanie obecności przykładów do tzw. bootstrap-owych zbiorów uczących (standardowo obecne w konstrukcji bagging)
2. Losowanie cech – wg. zapisu algorytmu 1- dodatkowo realizowane na każdej próbie bootstrapowej (także znane już z literatury)
3. Losowanie wewnątrz konstrukcji samej sieci neuronowej (w eksperymencie realizowane jako losowanie struktury warstw w sieci (3 wersje liczby neuronów ReLU w każdej z dwóch warstw ukrytych)

Największa oryginalność propozycji dotyczy losowej modyfikacji architektury sieci. Samo losowanie cech (drugi poziom) był już rozważane w pracach nad uogólnieniem zespołów klasyfikatorów. W najprostszej wersji Ho zaproponowała go w tzw. feature ensemble jako metoda random subspace, następnie belgijscy badacze P. Latinne, O. Debeir, Ch. Decaestecker w pracy *Mixing Bagging and Multiple Feature Subsets to Improve Classification Accuracy of Decision Tree Combination* (2000) [1] rozważali już połączenie 1 i 2 poziomu losowości. Ponadto piszący tą recenzję także przedstawiał inną koncepcję (poziom 1 połączony z innymi strategiami wyboru cech) - praca *Improving Bagging by Feature Selection with Dynamic Integration of Sub-classifiers*. Dlatego element losowości budowy sieci jest najciekawszy i jego wpływ mógłby być jeszcze dokładniej przebadany. Kolejnym własnym dokonaniem doktoranta jest przeprowadzanie oceny eksperymentalnej własnej propozycji na dużej popularnej kolekcji obrazów (zbiór MNIST i jego wersja Fashion-MNIST), często wykorzystywanej do oceny głębokich sieci neuronowych. Z uwagi na wielkość zbioru danych dokonano jednego podziału na część uczącą i testową oraz porównania z prostszymi sieciami neuronowymi i popularnymi klasyfikatorami, w tym zespołami (XgBoost, GradientBoosting i Random Forest) oceniając trafność predykcji przykładów testowych. Zamieszone wyniki (Tabela nr 1) pokazują, że propozycja RANNs jest dokładniejsza niż porównywane algorytmy. Ten artykuł pokazuje następujący oryginalny wkład doktoranta -- własny nowy algorytm i jego ocena – w obszarze uczenia maszynowego i danych obrazowych. Jest napisany poprawnie w stylu wielu prac z uczenia nadzorowanego, choć równocześnie niektóre elementy sprawiają wrażenie prac w toku, które można by dalej prowadzić. Tematycznie mieści się on w obszarze sztucznej inteligencji (tytuł i zakres rozprawy), choć nie dotyczy bezpośrednio danych w postaci szeregów czasowych.

Podsumowując moją prezentację dokonań wraz z jej dyskusją i oceną, sądzę że najciekawsze pod względem poznawczym i najbardziej związane z tematem rozprawy są dwa pierwsze obszary tematyczne: eksperymentalna ocena porównawcza wielu miar podobieństwa w zadaniu klasyfikacji szeregów czasowych oraz wprowadzenie nowego uogólnienia współczynnika Prokrusta dla danych funkcjonalnych. Dwie pozostałe prace wydają się być nie aż tak oryginalne (patrz powyższe uwagi), lecz równocześnie pokazują wiedzę, umiejętności i pracowitość autora. Wszystkie artykuły są napisane z dobrą organizacją i prezentacją (choć praca nr 5 jest dość trudna z uwagą na terminologię dziedziny zastosowania). Podsumowując uważam, że omawiane osiągnięcia pozwalają uznać, że założone cele pracy doktorskiej zostały osiągnięte.

#### 4. Uwagi krytyczne i dyskusyjne

Pomimo wielu interesujących propozycji oraz faktu, że prace przeszły już proces recenzyjny, były więc sprawdzone, ciągle można rozważać pewne wątpliwości oraz otwarte pytania lub uwagi do dalszych badań. Poniżej formułuję uwagi krytyczne do treści rozprawy oraz pytania o charakterze dyskusji polemicznej z autorem oraz zachęty do dalszej dyskusji podczas samej obrony doktoratu:

- Eksperymenty nad klasyfikacją szeregów wykonano ze względu na błąd klasyfikowania (i jego rangi). Przedstawione rezultaty wskazują na przewagę miar budowanych z wykorzystaniem dopasowania DTW. Równocześnie obliczanie dopasowania punktów w DTW jest kosztowne obliczeniowo w stosunku do większości innych miar. Widać to także w pierwszym artykule na rysunku 4, gdzie pokazano średnie czasy obliczeń. Jeśli dobrze interpretuje krótkie komentarze w dalszych artykułach, to wersję uogólnień jak DTWc, DDTW i podobne mają redukować te

koszty. Wyniki z artykułu trzeciego pokazują dobre predykcje dla miar z rodziny Derivative Distances DD(with DTW,LCSS), co do których nie ma dyskusji jak kosztowne czasowo jest ich obliczanie. Ponadto w pracach 2 i 3 nie uwzględniano tego w eksperymentach. Chciałbym prosić doktoranta o komentarz rozszerzający na temat czasochłonności obliczeń tych elastycznych wykorzystujących ideę DTW oraz ew. dyskusję podczas obrony innych ograniczeń wskazywanych dominujących miar. Tego brakuje mi w przedstawionych tekstach.

- Interesujące byłoby także pytanie, na ile przedstawione wyniki przenosiłby się na inne zadania eksploracji strumieni niż wyłączenie klasyfikacja. Na przykład czy rekomendacje co do grup miar mogą być także przydatne w grupowaniu?
- W przypadku uogólnienia współczynnika Prokrusta brakuje mi pewnej dyskusji na temat różnic w stosunku do innych wcześniej wprowadzonych funkcjonalnych miar współzależności ( $\rho_V$ , nowy funkcjonalny współczynnik odległościowy, uogólnienia współczynników HSIC i KTA, itd.). Także z punktu widzenia praktycznej stosowalności – kiedy rekomenduje się stosowanie której miary z tych propozycji.
- Na temat pracy nt. analizy danych chemometrycznych z badań nad jakością soków jabłkowych, doktorant pisze na początku rozprawy, że dane pochodzące ze spektrogramów mogą mieć postać szeregów czasowych (patrz także polskojęzyczne streszczenie). Nie jestem przekonany do tej analogii, zwłaszcza jak patrzę na przedstawione rysunku widma pochodzących ze spektrometru.
- W przypadku oceny eksperymentalnej zespołu z losowymi sieciami RANNs brakuje mi dokładniejszej analizy składników całego zespołu, w szczególności jaki jest wpływ każdego z poziomów losowości na działanie całego klasyfikatora. Ponadto w pracach porównawczych powinno się użyć pojedynczej sieci o architekturze zbliżonej do użytej w RANNs, gdyż sieć MLP jest zbyt prosta do analizy danych obrazowych z MNIST. Eksperymenty powinny być szersze, obejmować więcej danych, nie tylko obrazowych, oraz dodatkowo zawierać sprawdzenie dokładniej liczby komponentów (klasyfikatorów składowych – brakuje rekomendacji jak je dobrać), różnych sposobów losowania zarówno w topologii warstw sieci, jak w wyborze cech.
- Można też sformułować pytanie jak dostosować koncepcję losowych sieci RANNs do przetwarzania szeregów czasowych.
- Wyraziłem także w punkcie 2 uwagi do konstrukcji maszynopisu, i zbyt ograniczonego wstępu oraz braku podsumowania.

Mam nadzieję, że doktorant będzie mógł się odnieść do moich pytań w dyskusji podczas obrony.

## 5. Konkluzja końcowa

Podsumowując, pomimo powyższych uwag polemicznych sądzę, że przedstawiona rozprawa charakteryzuje się dobrym poziomem merytorycznym i zawiera interesujące rezultaty badawcze, które mogą być przydatne dla efektywniejszej realizacji zadań eksploracji danych, w szczególności dla szeregów czasowych, oraz użyte w zastosowaniach praktycznych. W mojej opinii doktorant wykazał się umiejętnościami prowadzenia badań naukowych, rozwiązywania problemów oraz pracowitością.

Dlatego uważam, że rozprawa pana mgra Piotra Piaseckiego spełnia warunki stawiane przez ustawę prawo o szkolnictwie wyższym i nauce w odniesieniu do rozpraw doktorskich i może być dopuszczona do publicznej obrony.



Prof. dr hab. inż. Jerzy Stefanowski