Wydział Automatyki, Elektroniki i Informatyki
Katedra Inżynierii i Analizy Eksploracyjnej Danych

Prof. dr hab. inż.
Joanna Polańska
Kierownik Katedry

Gliwice, 26 stycznia 2026 r.

Recenzja rozprawy doktorskiej

Autor: **mgr Mateusz Staniak**

Tytuł: **Metody statystyczne dla danych proteomicznych ze strukturą wielokrotnej przynależności pozyskanych przy pomocy spektrometrii mas** (ang. *Statistical Methods for Mass Spectrometry Proteomics Data with Multiple Membership Structure*)

Promotorzy: **Prof. dr hab. Małgorzata Bogdan**

Prof. dr Tomasz Burzykowski

1. Ogólna charakterystyka rozprawy

Przedłożona do recenzji rozprawa doktorska jest napisana w języku angielskim, liczy 133 strony tekstu, składa się z pięciu rozdziałów oraz bibliografii. Praca ma charakter interdyscyplinarny i ogniskuje się na problematyce analizy wyników eksperymentów spektroskopii masowej białek. Jest to obecnie bardzo szeroki obszar badawczy, obejmujący różnorodne techniki eksperymentalne, a także wymagający zastosowania odpowiednich modeli matematycznych oraz konstrukcji oprogramowania dla interpretacji wyników pomiarowych. Większość metod eksperymentalnych spektroskopii białek bazuje na trawieniu łańcuchów białkowych na mniejsze fragmenty – peptydy. Sekwencje białkowe, a także stężenia białek w badanych próbkach są np. oceniane przez wykorzystanie pomiarów czasów przelotów oraz zliczeń liczb zjonizowanych cząstek peptydów w detektorach. Istnieje wiele wariantów / scenariuszy obliczeniowych, dopasowanych do różnych technik eksperymentalnych, a także szereg pakietów programistycznych oferowanych przez dostawców sprzętu pomiarowego, jak również powstałych w szeregu zespołów badawczych. Wiele z technik obliczeniowych wykorzystuje wersje statystycznych modeli liniowych lub uogólnionych modeli liniowych w powiązaniu z procedurami optymalizacji wskaźników kwadratowych jakości lub innych funkcji, staranniej dopasowanych do wymagań odpornościowych oraz do wiedzy biologicznej.

Politechnika Śląska
Wydział Automatyki, Elektroniki i Informatyki
ul. Akademicka 16, 44-100 Gliwice
+48 32 237 2144 / +48 601 891 245
joanna.polanska@polsl.pl

W recenzowanej rozprawie doktorskiej badania skupione są na specjalnym wariancie problemu analizy danych eksperymentów proteomicznych, w których materiał badawczy/pomiarowy stanowi najczęściej materiał biologiczny o złożonej kompozycji molekularnej, co wiąże się ze strukturą niejednoznacznej przynależności peptydów do analizowanych w eksperymentach profili białkowych. Taka sytuacja bardzo często występuje w badaniach biomedycznych, gdzie analizie podlega całkowity profil molekularny np. krwi chorych, a ocena jakościowa i ilościowa proteomu lub jego wybranych fragmentów jest niezwykle ważna z punktu widzenia diagnostyki czy oceny wpływu/skuteczności terapii. Należy tutaj również podkreślić, że poprawna analiza takich danych wymaga opracowania dodatkowych, wysoko specjalizowanych modeli i technik obliczeniowych. Dlatego wybór tematyki rozprawy doktorskiej należy uznać z bardzo trafny i interesujący.

Rozdział 1 pracy jest omówieniem całości jej zawartości. Stwierdza się na wstępie, że praca koncentruje się na problematyce wielokrotnej przynależności peptydów w eksperymentach spektrometrii proteomicznej oraz wymienia się dwie główne motywacje i kierunki badań. Pierwszy kierunek to problem oceny stężeń białek w próbkach, w sytuacji występowania wielokrotnej przynależności peptydów w pomiarach. Drugi obszar zastosowania to spektrometria masowa wymiany wodoru i deuteru (HDX-MS). W następnej kolejności przedstawiona jest zawartość dalszych rozdziałów 2 - 5. Wymienione są też publikacje, związane z pracą, których Doktorant jest współautorem.

Rozdział 2 rozprawy składa się z 5 podrozdziałów i stanowi formę przeglądu literatury oraz istniejących rozwiązań, w którym znajdziemy odwołanie się do wyników znanych w literaturze, istotnych i wykorzystywanych w pracy. Ma, tak jak całość pracy, charakter interdyscyplinarny i zawiera omówienie metod matematycznych, opisy eksperymentalnych technik proteomicznych, a także przedstawia algorytmikę (a raczej elementy algorytmiki) analizy danych proteomicznych.

Podrozdział 2.1 poświęcony jest metodom optymalizacji. W podrozdziale tym przedstawia się problem optymalizacji funkcji wielu zmiennych, gdzie zmienne można podzielić na dwa podwektory. Dla takiego podziału funkcja jest wypukła ze względu na każdy z podwektorów. Dla takiej funkcji używa się angielskiego terminu bi-convex function (polskie tłumaczenie mogłoby brzmieć funkcja dwuwypukła). Przedstawia się algorytm (Algorithm 1) sekwencyjnej optymalizacji takiej funkcji i za literaturą stwierdza się zbieżność tego algorytmu. Algorytm ten jest istotny i wykorzystywany w dalszych rozdziałach pracy.

Podrozdział 2.2 poświęcony jest metodom dopasowania statystycznych modeli regresji liniowej i nieliniowej do danych na bazie maksymalizacji funkcji wiarygodności. W tym kontekście Doktorant przedstawia także specyfikę danych proteomicznych ze strukturą wielokrotnej przynależności (nazywanych także danymi ze strukturą mieszanej przynależności) oraz cytuje kilka, związanych z analizą takich danych, pozycji literatury. W podrozdziale tym Doktorant przedstawia także grafy

Politechnika Śląska
Wydział Automatyki, Elektroniki i Informatyki
ul. Akademicka 16, 44-100 Gliwice
+48 32 237 1775 / +48 601 891 245
joanna.polanska@polsl.pl

dwudzielne jako narzędzie do reprezentacji i wizualizacji struktury danych o wielokrotnej przynależności.

Podrozdział 2.3 omawia techniki eksperymentalne proteomiki i odnosi je do problemów budowy odpowiednich algorytmów analizy danych. Jako podstawowe podejście wymienia sposób „od dołu do góry" analizy składu i stężeń białek przez rozdzielenie / trawienie na peptydy i badanie ich spektrów w jedno lub wieloetapowych procesach pomiarowych. Omawia także problemy obwiedni izotopowych, techniki multipleksowania, znakowania izobarycznego. Przedstawia problematykę analizy post-translacyjnej oraz spektrometrię wymiany wodoru i deuteru. Wspomina także o istniejących bazach danych proteomicznych. Przy omawianiu technik proteomicznych w tym podrozdziale Doktorant zwraca także uwagę na specyfikę danych o wielokrotnej przynależności peptydów, ich modelowanie i traktowanie w rozwijaniu metodologii analizy.

W podrozdziale 2.4 Doktorant omawia procedury analizy danych proteomicznych, ich dopasowanie do specyfiki przeprowadzonego eksperymentu. Wymienia najważniejsze, związane z tym zagadnieniem pozycje literaturowe, istniejące narzędzia programistyczne, istniejące, dostępne literaturowo zbiory danych. Jednym z najczęściej używanych narzędzi algorytmicznych i programistycznych jest pakiet MSstats, w którego tworzeniu i rozwijaniu brał udział Doktorant. Charakteryzuje także różne warianty problemów analizy danych proteomicznych, analizę ilościową składu proteomicznego, analizę różnicową. W tym aspekcie omawia także dokładniej problem analizy danych proteomicznych wymiany wodoru i deuteru.

W podrozdziale 2.5 Doktorant wymienia i charakteryzuje dokładniej zbiory danych proteomicznych, które są wykorzystywane w jego pracy. Jest to 5 dużych zbiorów danych eksperymentalnych uzyskanych różnymi technikami, które stanowią bardzo dobrą ilustrację dla studiowania i porównywania metod analizy danych proteomicznych.

Rozdział 3 poświęcony jest dokładnemu przedstawieniu algorytmu analizy danych proteomicznych uzyskiwanych technikami spektrometrii masowej, ze znakowaniem izobarycznym w sytuacji gdy cechy peptydowe wykazują charakter wielokrotnej przynależności. Na wstępie formułowane są zależności bilansowe (3.1) - (3.2), które charakteryzują problemy oceny zmiennych. Dyskutowana jest także postać funkcji celu (3.4). Zadanie oceny struktury i (względnych) stężeń białek, przez optymalizację funkcji celu (uogólnionej funkcji wiarygodności), w przypadku występowania wielokrotnej przynależności peptydów staje się trudniejsze. Konieczne jest wprowadzenie wag, opisujących podział współdzielonych peptydów pomiędzy różne białka, które w modelu tworzą iloczyny ze stężeniami białek. Dla rozwiązania takiego problemu proponuje się algorytm 3, optymalizacji dwuwypukłej, w którym wartości czynników iloczynów są na przemian „zamrażane". W dalszej części rozdziału przedstawione są wyczerpujące wyniki badań jakości zaproponowanego algorytmu, zarówno dla danych symulowanych jak i rzeczywistych.

Politechnika Śląska
Wydział Automatyki, Elektroniki i Informatyki
ul. Akademicka 16, 44-100 Gliwice
+48 32 237 1775 / +48 601 891 245
joanna.polanska@polsl.pl

Rozdział 4 poświęcony jest rozwijaniu algorytmów oceny prawdopodobieństw wymiany wodoru i deuteru w spektrometrii masowej HDX-MS. Jako bazę do rozwijania algorytmów oceny tych prawdopodobieństw przyjmuje się wielomianowy model logit (4.4). Z modelem tym wiąże się dwa typy wskaźników jakości i proponuje się metody numeryczne ich optymalizacji. Podobnie jak w poprzednim rozdziale proponowana strategia optymalizacji jest dwuetapowa, przedstawiona jest przez algorytm 4 i jego pseudokod na stronie 87. Proponowana metoda analizy jest dalej przebadana na szeregu danych symulacyjnych i rzeczywistych.

Rozdział 5 poświęcony jest przedstawieniu narzędzi programistycznych opracowanych przy współudziale Doktoranta.

2. Ocena pracy

Praca jest napisana szczegółowo i dość precyzyjnie. Ma bardzo szeroki zakres jeśli chodzi o różnorodność opisywanych technik eksperymentalnych oraz analizowanych zbiorów danych. Praca ma logiczną konstrukcję, dobrze dopasowaną do przedstawianej problematyki.

Praca bazuje na zacytowanej szerokiej literaturze. Widać, że Doktorant ma bardzo dobrą orientację w literaturze naukowej dotyczącej wszystkich aspektów jego badań.

Praca ma bardzo silny walor interdyscyplinarny. Metody praktyczne proteomiki eksperymentalnej są przedstawione na bardzo wysokim poziomie, z dużą szczegółowością. Są one powiązane z metodami modelowania matematycznego, dobranymi w odpowiedni sposób i wszechstronnie przetestowanymi, a także z technikami konstrukcji odpowiednich narzędzi programistycznych. Także metody modelowania i programowania są opisane wyczerpująco i na wysokim poziomie.

Doktorant bardzo dobrze panuje nad wszystkimi tymi aspektami, co wydaje się także potwierdzać jego dotychczasowy dorobek naukowy. Doktorant wykazuje się szeroką wiedzą i opanowaniem zarówno podstaw biologicznych jak też technik eksperymentalnych, aspektów modelowania, obliczeń, konstrukcji oprogramowania na bardzo szczegółowym i zaawansowanym poziomie. Posiadany przez Doktoranta warsztat naukowy ma bardzo szerokie zastosowania.

Za najważniejsze oryginalne osiągnięcie naukowe pracy uważam sformułowanie i rozwiązanie szeregu wariantów modeli matematycznych w zastosowaniu do badań w obszarze proteomiki masowej białek, w szczególnej sytuacji gdy występuje sytuacja wielokrotnej przynależności mierzonych peptydów. Dla tego problemu wszystkie zbudowane modele matematyczne zostały zaimplementowane i dokonano wszechstronnych analiz statystycznych.

HR EXCELLENCE IN RESEARCH
Politechnika Śląska
Wydział Automatyki, Elektroniki i Informatyki
ul. Akademicka 16, 44-100 Gliwice
+48 32 237 1775 / +48 601 891 245
joanna.polanska@polsl.pl

Z wynikami przedstawionymi w rozprawie doktorskiej powiązane są bardzo wartościowe publikacje naukowe, których współautorem jest Doktorant, wydane w bardzo prestiżowych czasopismach naukowych.

Doktorant posiada już duży i wartościowy dorobek naukowy, który ponadto podsumowuje się znaczącymi wartościami wskaźników bibliometrycznych.

3. Uwagi krytyczne i dyskusyjne

W pracy bardzo przydałyby się wykazy oznaczeń, skrótów, rysunków i tabel. Byłyby one bardzo użyteczne do prześledzenia dość złożonych wyników przedstawianych w pracy.

Praca w swojej konstrukcji trochę odbiega od standardowej formy. Brak jest jawnego sformułowania tez rozprawy, jak również rozdziału stanowiącego podsumowanie całości badań wraz z odniesieniem się do wcześniej sformułowanych tez. Każdy z rozdziałów ma wprawdzie swoje podsumowanie (dyskusję), jednak wyciągnięcie jednolitych wniosków z całości przeprowadzonych prac byłoby bardzo wskazane.

Prezentacje części z wyników jest trochę niekonsekwentna, Doktorant z jednej strony w rozdziale 2 definiuje bardzo podstawowe pojęcia, jak ciągłość funkcji, pochodne, hesjan (trochę nawet brakuje tu precyzji matematycznej), a z drugiej strony posługuje się pojęciami, takimi jak *maximum likelihood with nuisance parameters*, *pseudolikelihood approach*, z dość zaawansowanej monografii Davidian, Marie and David M Giltinan (1995). Nonlinear Models for Repeated Measurement Data. Vol. 62. CRC Press bez ich niezależnego zdefiniowania.

Nie potrafię w Internecie znaleźć przytoczonej na stronie 6 pracy: "Mateusz Staniak, Jürgen Claesen, Tomasz Burzykowski, Estimation of peptide-segment-level kinetic exchange rates based on the isotope distributions of overlapping peptides, 2025", dla której brak jest szczegółowych danych bibliograficznych

4. Pytania

- Czy w opinii Doktoranta proponowane metoda dwuetapowej optymalizacji ma we wszystkich sytuacjach przewagę nad metodą usuwania wspólnych peptydów, czy też istnieją przypadki, gdy druga jest lepsza?
- Czy mogę prosić o szczegółowe omówienie „Delta method" wspomnianej na stronie 88 ?
- Czy we wzorze (4.4) n_j^ex jest oceniane czy uważane za znane ?

Politechnika Śląska
Wydział Automatyki, Elektroniki i Informatyki
ul. Akademicka 16, 44-100 Gliwice
+48 32 237 1775 / +48 601 891 245
joanna.polanska@polsl.pl

- Na początku strony 85 wspomina się o metodach optymalizacji globalnej (*global optimization methods*). O jakie metody tu chodzi? Metody optymalizacji wymienione dalej na dole strony 86 są to wszystko metody optymalizacji lokalnej. Czy przez „*global optimization*" rozumie się tutaj przegląd wartości funkcji jednej zmiennej?

- Jakie są czasy obliczeń dla algorytmów z rozdziału 4? Jak one zależą od rozmiaru problemu?

5. Konkluzja

Praca stanowi podsumowanie oryginalnych, interdyscyplinarnych badań naukowych, w których formułowano i weryfikowano oryginalne hipotezy badawcze. Dokumentuje wiedzę i kompetencje Doktoranta. Osiągnięcia i oryginalne elementy pracy są na pewno wystarczające do jej ogólnej pozytywnej oceny. Stwierdzam, że rozprawa spełnia warunki formalne, jak i zwyczajowe stawiane pracom doktorskim i wnioskuję o jej dopuszczenie do publicznej obrony.

Ponadto, pomimo uwag krytycznych, biorąc pod uwagę poziom naukowy pracy, opanowania przez Doktoranta bardzo szerokiego zakresu tematycznego oraz dorobek publikacyjny Doktoranta wnioskuję o wyróżnienie rozprawy.

HR EXCELLENCE IN RESEARCH
Politechnika Śląska
Wydział Automatyki, Elektroniki i Informatyki
ul. Akademicka 16, 44-100 Gliwice
+48 32 237 1775 / +48 601 891 245
joanna.polanska@polsl.pl