

prof. dr hab. inż. Małgorzata Bogdan
Instytut Matematyki
Uniwersytet Wrocławski

RECENZJA ROZPRAWY DOKTORSKIEJ

Tytuł rozprawy: Computational methods for leading-edge mass spectrometry techniques

Autor rozprawy: mgr Piotr Radziński

Promotorzy rozprawy: prof. dr hab. Anna Gambin i dr hab. Jakub Karasiński

Zasadnicza część rozprawy poświęcona jest modelowaniu i analizie danych pochodzących z eksperymentów spektrometrii mas. Analiza tego rodzaju danych ma ogromne znaczenie, ponieważ umożliwia identyfikację i ilościowe oznaczanie związków chemicznych oraz biomolekuł, co odgrywa kluczową rolę w biologii, medycynie, farmacji i naukach środowiskowych. Jest to jednak zadanie niezwykle wymagające – nowoczesne instrumenty generują bardzo obszerne, złożone i szasumione zbiory danych, których poprawna interpretacja wymaga zastosowania zaawansowanych metod obliczeniowych i statystycznych. Osiąganie istotnych wyników w tej dziedzinie wymaga interdyscyplinarnej wiedzy obejmującej zarówno chemię i zasady działania współczesnych urządzeń spektrometrycznych, jak i znajomość metod statystycznych oraz algorytmicznych. Opracowywanie nowych algorytmów wiąże się dodatkowo z koniecznością pogodzenia precyzji z efektywnością obliczeniową, co wymaga umiejętnego kompromisu.

Rozprawa składa się z sześciu rozdziałów. Rozdział 1 zawiera wprowadzenie oraz krótkie omówienie struktury pracy. Rozdział 2 prezentuje propozycje nowych algorytmów do wyznaczania mas izotopowych sygnałów białkowych i oligonukleotydowych. W części pierwszej opisano algorytm Envemind dla sygnałów białkowych, który działa dwuetapowo: (1) identyfikuje teoretyczne spektrum najlepiej dopasowane do widma uzyskanego w pomiarze MS, a następnie (2) predykuje masę monoizotopową za pomocą modeli liniowych wytrenowanych na danych wygenerowanych algorytmem IsoSpec (Łacki i in., 2017). Kluczowym elementem metody jest estymacja parametrów kraty, czyli dyskretnego zbioru wartości możliwych do przyjęcia przez masy izotopowych izomorfów. Nowością w stosunku do wcześniejszych podejść jest wykorzystanie nie tylko średniej masy proteiny, lecz

także wariancji jej spektrum. Wyniki tych badań opublikowano w *Journal of the American Society of Mass Spectrometry* (2022), gdzie Pan Radziński jest pierwszym autorem.

Druga część rozdziału 2 dotyczy szacowania masy monoizotopowej oligonukleotydów. Zaproponowano algorytm, który najpierw wyznacza najbardziej intensywny pik w funkcji średniej masy, a następnie estymuje offset – typową różnicę między tym pikiem a rzeczywistą masą monoizotopową. Był to pierwszy algorytm automatyzujący wyznaczanie mas izotopowych oligonukleotydów. Został on wdrożony m.in. w ścieżkach analitycznych Janssen Pharmaceuticals (części Johnson & Johnson Pharmaceutical R&D). Wyniki opublikowano w *Analytical Chemistry* (2024) w pracy wieloautorskiej, w której Pan Radziński jest drugim autorem, co podkreśla jego istotny wkład.

Rozdział 3 przedstawia nowy algorytm kompresji obrazów spektrometrycznych (MSI – mass spectrometry imaging) z wykorzystaniem konwolucyjnych sieci neuronowych w architekturze kodera-dekodera. Metoda ta umożliwia znaczną redukcję rozmiaru danych, co ułatwia ich dalszą analizę. Wykazano również, że kompresja sprzyja eliminacji szumu i poprawia dokładność segmentacji obrazów MSI. Wyniki badań opublikowano w *Analytical Chemistry*, gdzie Pan Radziński jest pierwszym autorem.

Rozdział 4 zawiera propozycję nowego algorytmu szacowania niepewności w metodzie kalibracji naprzemiennej stosowanej w dokładnych pomiarach stosunków izotopowych przy użyciu MC-ICP-MS. Wyniki opublikowano w *Analytical Chemistry* (2024), gdzie Pan Radziński jest trzecim autorem. Jego wkład dotyczył oceny niepewności w kalibracji stosunków izotopowych, w szczególności estymacji parametru δ , opisującego różnicę między wartościami w analizie i w standardzie odniesienia.

Rozdział 5 stanowi podsumowanie wyników badań medycznych, w których Autor odpowiadał za analizy statystyczne. Publikacje te są wysoko cytowane, co dodatkowo podkreśla znaczenie prac.

Tematyka rozprawy jest niezwykle istotna, a przedstawione artykuły zawierają oryginalne rozwiązania i implementacje algorytmów o dużym znaczeniu praktycznym dla analizy danych spektrometrycznych. Wszystkie publikacje są wieloautorskie, co jest typowe dla interdyscyplinarnej dziedziny, jaką jest bioinformatyka. We wszystkich pracach włączonych do rozprawy Pan Radziński pełnił rolę jednego z głównych autorów, co dowodzi jego zasadniczego wkładu.

Pewnym mankamentem rozprawy jest mało precyzyjny język i brak ścisłości w opisie metod matematycznych oraz informatycznych. Prowadzi to czasami do usterek, a niektóre fragmenty są trudne do formalnej oceny. Szczególnie widoczne jest to w rozdziale 4, który należy uznać za najsłabszy element pracy. Proces statystycznego szacowania niepewności wymaga precyzyjnego modelu matematycznego, czego tu brakuje. Na przykład interpretacja części wyników z tabel 4.1 i 4.2 jest niejasna. W szczególności wyniki podano

dla różnych prób, bez wskazania, co w danej próbie pełni rolę standardu, a co analitu. Ponadto przeprowadzone testy Studenta sugerują obecność systematycznych błędów w trzech z ośmiu prób, co wymaga pogłębionej analizy i może być główną przyczyną utraty pokrycia przedziału podanego we wzorze (4.7). Proponowany algorytm Monte Carlo został opracowany dość ad hoc i nie posiada solidnych podstaw statystycznych. W szczególności wprowadzenie efektu losowego modelującego błąd wynikający z propagacji błędów pomiarowych stosunków izotopowych w analizie i standardzie nie znajduje uzasadnienia, gdyż błąd ten uwzględniony jest już w różnicy między wyliczonym parametrem δ a jego prawdziwą wartością.

Pewne braki precyzji widoczne są również w rozdziale 2. Brakuje w nim formalnych definicji pojęcia *averagine* czy parametru ρ . Ponadto opis wyznaczenia parametru ρ na stronie 25 jest niezrozumiały – nie jest jasne, jak zależy on od dwóch rozwiązań problemu optymalizacyjnego wyspecyfikowanego w Lemacie 1. Paragraf poprzedzający Tabelę 2.1 również jest niejasny. Wektor średnich mas a ma wszystkie elementy dodatnie, zatem równanie $\langle a, x \rangle = 0$ implikuje, że niektóre elementy wektora x są ujemne. W tej sytuacji iloczyn skalarny $\langle v, x \rangle$ zawiera składniki ujemne i trudno go interpretować jako wariancję, jak sugeruje tekst rozprawy. Nie jest również jasne, w jaki sposób wzór (2.13) wiąże się z Lematem 1.

W rozdziale 3 opis architektury sieci neuronowych jest niepełny – brakuje szczegółów dotyczących rozmiaru jąder konwolucyjnych czy celu i konstrukcji warstwy liniowej. Ponadto wzór na stratę kontrastową jest niekompletny, gdyż uwzględnia jedynie składniki odpowiadające pozytywnym parom. Nie podano również wzoru na całkowitą funkcję straty. Pewne wątpliwości budzi również niejasny opis segmentacji za metodą iteracyjnego k-means, którego konstrukcja nie została właściwie uzasadniona. W części 3.3.1 nie podano również liczby klas wybranych do segmentacji. W rozdziale 5 metody statystyczne opisano jedynie szkicowo – brakuje formalnego sformułowania i uzasadnienia modeli wykorzystanych w analizach. Na przykład w części 5.1.3 nie jest jasne, jaką cechę modelowano, tj. jaką zmienną opisywały predyktory z tabeli 5.1.

Pomimo powyższych krytycznych uwag moja ocena rozprawy doktorskiej jest pozytywna. Autor wykazał się szeroką wiedzą z zakresu chemii, biologii, statystyki i uczenia maszynowego. Rozprawa zawiera szereg oryginalnych rozwiązań matematyczno-informatycznych, które w istotny sposób usprawniają analizę danych spektrometrycznych. Dowodzi to umiejętności samodzielnego prowadzenia badań naukowych przez Doktoranta. Wskazane wyżej braki w precyzji opisu części matematycznej wynikają prawdopodobnie z konieczności komunikowania wyników naukowcom z innych dziedzin i nie wpływają na ogólnie pozytywną ocenę całości.

Podsumowując – w świetle przedstawionych osiągnięć oraz realizacji ce-

łów badawczych uważam, że rozprawa spełnia wymagania stawiane pracom doktorskim i uzasadnia dopuszczenie mgr. Radzińskiego do dalszych etapów postępowania w sprawie nadania stopnia doktora w dziedzinie nauk ścisłych i przyrodniczych, w dyscyplinie informatyka.

Z wyrazami szacunku,

Monika Bogdan

