

Prof. dr hab. Piotr Bała
Interdyscyplinarne Centrum Modelowania
Matematycznego i Komputerowego
Uniwersytet Warszawski
bala@icm.edu.pl

Toruń, 24 października 2023

RECENZJA ROZPRAWY DOKTORSKIEJ MGR MARCINA SOBIERAJA DLA RADY
NAUKI DYSCYPLINY FIZYKA UNIWERSYTETU WARSZAWSKIEGO

Tytuł rozprawy: Nowe metody analizy mechanizmów funkcjonowania układów biomolekularnych w oparciu o symulacje dynamiki molekularnej

Autor rozprawy: mgr Marcin Sobieraj

1. Jakie zagadnienie naukowe jest rozpatrywane w pracy (teza rozprawy) i czy zostało one dostatecznie jasno sformułowane przez autora? Jaki charakter ma rozprawa (teoretyczny, doświadczalny, inny)?

Rozprawa doktorska mgr Marcina Sobieraja dotyczy opracowania i wykorzystania metod opartych o teorię informacji i analizę przyczynowości Grangera do opisu czasowo-zależnych właściwości makromolekuł biologicznych. Celem pracy było zaproponowanie nowych rozwiązań metodologicznych, pozwalających na uzyskanie wglądu w złożone zależności ukryte w obszernych zbiorach danych generowanych w oparciu o wysokowydajne, współczesne symulacje układów biomolekularnych. Cel pracy został przez autora jasno sprecyzowany w Rozdziale 1.2 rozprawy. Praca ma charakter teoretyczny, zawiera opis zaproponowanej metody, jej teoretyczne uzasadnienie oraz weryfikację poprzez zastosowanie do analizy procesu zwijania małego białka.

Rozprawa jest obszerna i liczy 194 strony. Jest podzielona na 5 rozdziałów. Rozdział 1. przedstawia badanie zagadnienie i definiuje cel pracy. Rozdział 2. zawiera opis metod redukcji wymiarowości zagadnienia przy minimalnej stracie informacji. Rozdział 3. przedstawia analizę Grangera i jej przeniesienie na grunt modelowania molekularnego a Rozdział 4 przedstawia rozszerzenie metody Grangera. Rozdział 5. przedstawia możliwe kierunki dalszego rozwoju rozwiązań opisanych w rozprawie. Przedstawiono w nim podejścia, które uogólniają standardową analizę Grangera, tak aby uwzględniała ona nieliniowe korelacje i efekty wielociałowe/wielokanałowe.

Praca uzupełniona jest o 4 dodatki zawierające między innymi szczegółowe wyprowadzenia oraz rozszerzoną analizę procesu zwijania Chignoliny.

2. Czy w rozprawie przeprowadzono w sposób właściwy analizę źródeł (w tym literatury światowej, stanu wiedzy i zastosowań w przemyśle) świadczący o dostatecznej wiedzy autora? Czy wnioski z przeglądu źródeł sformułowano w sposób jasny i przekonujący?

Początkowe rozdziały pracy zawierają wprowadzenie do tematyki związane z pracą i zawierają opis zastosowanych algorytmów, narzędzi programistycznych i sprzętu wykorzystywanego do analizy dużych danych. Rozdziały te przedstawiają aktualny stan wiedzy w zakresie dotyczącym tematyki badań. Ta część rozprawy jest podbudowana odnośnikami do literatury.

Bibliografia zamieszczona w rozprawie jest rozbudowana i obejmuje 172 pozycje. Literatura dobrana jest poprawnie i nie budzi zastrzeżeń. Należy zaznaczyć, że istotny element pracy opiera się na analizie trajektorii zwijania niewielkiego układu biomolekularnego, która została uzyskana przez innych autorów. Podejście takie jest w pełni uzasadnione ze względu na wielkość symulacji oraz czas i zasoby obliczeniowe potrzebne do jej wygenerowania.

3. Czy autor rozwiązał postawione zagadnienia, czy użył właściwej do tego celu metody i czy przyjęte założenia są uzasadnione?

Postawione w rozdziale 1.2 zagadnienia zostały przez autora rozwiązane. Pierwsza część pracy dotyczy sieci strukturalnych białek (ang. Protein structure network) zdefiniowanych w oparciu o zbiór węzłów (aminokwasów) połączonych kontaktami, opisywanymi przez długość kontaktu. Dążąc do określenia sieci zapewniających optymalną reprezentację struktur białek w pracy skupiono się na znalezieniu właściwej odległości odcięcia kontaktu, tj. długości powyżej której jest on uznawany za zerwany. Stosując podejście oparte na teorii informacji, w ramach którego poszukiwano reprezentacji charakteryzującej się możliwie dużą zawartością informacji przypadającej na kontakt, autor pokazał że odległość odcięcia zapewniająca spełnienie tego kryterium dla szerokiej klasy białek jest taka sama, niezależnie od rodzaju drugorzędowej struktury i wielkości białka. Autor wprowadził i zastosował miarę średniej korelacji, która w zastosowaniu do sygnałów gaussowskich równoważna jest mierze informacji wzajemnej, określającej ilość redundantnej informacji w sygnale.

W drugiej części pracy wykorzystano model sieci kontaktów uzyskany dla trajektorii dynamiki molekularnej białka chignoliny oraz formalizm analizy przyczynowości Grangera, do określania powiązań czasowych w procesie odwracalnego przechodzenia białka ze struktury rozwiniętej do natywnej struktury β -spinki do włosów. Wprowadził szereg deskryptorów pozwalających na uzyskanie prostej interpretacji macierzy współczynników Grangera. Przeprowadzona analiza pokazała który spośród dwóch głównych mechanizmów zwijania się chignoliny w strukturę β -spinki jest realizowany w ramach badanej symulacji. Należy zauważyć, że autor posłużył się trajektoriami uzyskanymi przez innych autorów i dostępnymi w literaturze.

W ostatniej części pracy autor zaproponował wydajny numerycznie formalizm oparty o algebrę kwaternionów, pozwalający na zastosowanie analizy przyczynowości Grangera do badania powiązań czasowych pomiędzy wektorami trójwymiarowymi. Formalizm taki pozwala na stosowanie analizy powiązań czasowych w odniesieniu do wektorowych wielkości fizycznych, m.in. powszechnie występujących w symulacjach układów (bio)molekularnych. Jego poprawne działanie zostało zweryfikowane w oparciu o symulacje modelowego układu fizycznego.

4. Na czym polega oryginalność rozprawy, co stanowi samodzielny i oryginalny dorobek autora, jaka jest pozycja rozprawy w stosunku do stanu wiedzy czy poziomu techniki reprezentowanych przez literaturę światową?

Oryginalność rozprawy polega na opracowaniu oryginalnej wersji formalizmu analizy przyczynowości Grangera w zastosowaniu do badania procesu zmiany konformacji układu biomolekularnego. Opracowany formalizm został zastosowany do analizy trajektorii Chignoliny w wyniku czego można było ustalić współrzedną reakcji oraz przeanalizować kolejne etapy procesu. Należy podkreślić, że jest to bardzo istotny wynik otwierający drogę do zautomatyzowania analizy zmian konformacyjnych. Do tej pory wyznaczenie ścieżki reakcji czy analiza dynamiki dużych układów molekularnych była trudna i wymagała dużego doświadczenia. Zaproponowane rozwiązania pozwalają na zobjektywizowanie i automatyzację procesu co jest szczególnie ważne w przypadku dużych układów molekularnych.

5. Czy autor wykazał umiejętność poprawnego i przekonującego przedstawienia uzyskanych przez siebie wyników (zwięzłość, jasność, poprawność redakcyjna rozprawy)?

Wyniki przedstawione w rozprawie są udokumentowane, przedstawione w poprawny sposób. W szczególności wyprowadzenia matematyczne są jasne i czytelne, wskazują na dużą swobodę w posługiwaniu się aparatem matematycznym i pokazują znajomość teorii informacji. Zastosowania zaproponowanych metod do przykładowych układów molekularnych jest dobrze i czytelnie opisane. Podstawowe wyniki autor umieścił w pracy, w dodatku rozszerzoną analizę. Od strony redakcyjnej praca jest napisana poprawnie, zgodnie z obowiązującymi zasadami.

6. Jakie są słabe strony rozprawy i jej główne wady?

Pewien niedosyt, nie umniejszający jednak znaczenie otrzymanych wyników związany jest z zastosowaniem zaproponowanych metod do tylko jednego, i to stosunkowo niedużego układu biomolekularnego. Chciałoby się zobaczyć skuteczność metody dla większych układów molekularnych, zwłaszcza takich w których zmiany konformacyjne nie są tak wyraźne czy oczywiste. Przydałoby się również zaprezentowanie wyników dla trajektorii w których położenia atomów dostępne są w większych odstępach czasu, jak to się dzieje dla długich symulacji układów molekularnych. Prawdopodobnie rzadsze próbkowanie położenia nie zmieni znacząco wyników, jednak warto byłoby się o tym przekonać.

W pracy autor kilkakrotnie informuje, że ze względów na wydajność numeryczną wybrano konkretne rozwiązanie. Są to stwierdzenia nie poparte głębszym uzasadnieniem, nie wiadomo jakie algorytmy były rozważane i jaki był czas ich wykonania. Idąc dalej przydałaby się informacja jakie systemy obliczeniowe były wykorzystywane i generalnie trochę więcej informacji na temat obliczeniowej strony zaproponowanego rozwiązania.

7. Jaka jest przydatność rozprawy dla fizyki?

Rozprawa podejmuje istotny temat przyczynowości w układach biomolekularnych oraz związanych z tym problemem kwestii takich jak wyznaczenie współrzędnej zmian konformacyjnych, określenie istotnych dla procesu konformacji układu i kierunku zmian. Zaproponowane rozwiązanie pozwala na algorytmizację analizy i tym samym ułatwia dokonanie analiz zachowania dużych układów biomolekularnych. W efekcie pozwala na obiektywizację analizy i uniezależnia ją od doświadczenia. Nie bez znaczenia jest także rozwój metody Grangera i dostosowanie jej do nowej klasy układów i zjawisk.

8. Do której kategorii recenzent zalicza rozprawę?

Po zapoznaniu się z rozprawą doktorską Marcina Sobieraja, a także po zapoznaniu się z pozostałym dorobkiem naukowym i osiągnięciami mgra Sobieraja uważam, że **rozprawa doktorska magistra mgr Marcina Sobieraja spełnia wymagania ustawowe i zwyczajowe stawiane rozprawom doktorskim.**

Biorąc pod uwagę rozprawę oraz związane z nią publikacje opublikowane w znaczących czasopismach, wnoszę o wyróżnienie rozprawy.