

Recenzja rozprawy doktorskiej

Autor: Mgr Mateusz Banach

Tytuł: Metody obliczeniowe jedno- i wielokryterialnej optymalizacji rojem cząstek.
Zastosowanie w bioinformatyce.

Promotor: Prof. dr hab. Irena Roterman – Konieczna

Charakterystyka rozprawy

Przedłożona do recenzji rozprawa doktorska składa się ze 298 stron tekstu, obejmuje streszczenie po polsku oraz po angielsku, 9 rozdziałów, bardzo obszerną bibliografię składającą się z 428 pozycji, spis rysunków, tabel, równań definicji oraz skrótów.

Praca poświęcona jest pewnym analizom zastosowania metody rozmytej kropli oliwy (FOD, fuzzy oil drop). Metoda FOD opracowana i rozwijana przez szereg lat przez zespół naukowy, z udziałem Promotor pracy i Doktoranta, jest rozwinięciem metody (idei) Kauzmana opisu wpływu środowiska wodnego na dynamikę molekularną białek i peptydów bazującego na własnościach hydrofobowości i hydrofilowości reszt aminokwasowych. Model FOD wprowadza do metody Kauzmana ciągłe, trójwymiarowe rozkłady prawdopodobieństwa Gaussa. Zastępuje się nimi binarną dychotomię, hydrofobowe jądro – hydrofilowy płaszcz. Tematykę rozprawy należy uznać za bardzo ciekawą. Metoda Kauzmana, sformułowana wiele lat temu (w 1959 roku) była tematem kontrowersji i dyskusji. Jest tematem współczesnych badań. Pojawiają się interesujące publikacje dyskutujące idee tej metody (np. prace [135-137] cytowane przez Doktoranta).

Praca jest skomponowana w postaci pięciu rozdziałów.

Pierwszy rozdział to wprowadzenie. Zawiera trzy podrozdziały opisujące kolejno motywację do podjęcia badań, cele pracy oraz dotychczasowy stan wiedzy. Trzeci z tych podrozdziałów jest najobszerniejszy. Doktorant opisuje w nim podstawowe podejścia do

zagadnienia zwijania białek, najważniejsze pola sił, metody dynamiki molekularnej, oddziaływanie reszt aminokwasowych z otoczeniem, hydrofobowość i hydrofilowość oraz teorię Kauzmanna. W dalszej części skupia się na zagadnieniu „kompleksowania” białek oraz na przeglądzie technik optymalizacji, w tym zwłaszcza optymalizacji wielokryterialnej.

Drugi rozdział zatytułowany jest „Materiał i metody”. Obejmuje pięć podrozdziałów. Pierwszy z nich „Baza danych białek homodimerycznych” opisuje sposób wyboru 200 homodimerów z bazy PDB, których struktury są użyte do eksperymentów obliczeniowych przeprowadzonych w ramach pracy. Drugi, „Pole wewnętrzne, pole siłowe ECEPP/3”, przedstawia znane z literatury pole sił międzyatomowych, używane w pracy do liczenia energii łańcucha polipeptydowego. Trzeci, „Pole zewnętrzne, model FOD” opisuje ważny dla pracy model rozmytej kropli oliwy FOD, stosowany do oceny energii oddziaływań pomiędzy dwoma rozdzielonymi łańcuchami białkowymi analizowanych homodimerów. Czwarty „Optymalizacja rojem cząstek” opisuje stosowane w pracy techniki optymalizacji globalnej. Wreszcie piąty, „Pozostałe algorytmy” zawiera opisy kilku innych, literaturowych algorytmów używanych w pracy.

Trzeci rozdział „Wyniki” zawiera siedem podrozdziałów. Pierwsze dwa podrozdziały poświęcone są oryginalnemu algorytmowi MOSF opracowanemu przez Doktoranta. Najpierw przedstawia się ideę algorytmu (podrozdział pierwszy) a potem wyniki jego porównania z kilkoma innymi, dla kilku literaturowych funkcji testowych. Algorytm wykazuje bardzo obiecujące własności. Kolejny, trzeci podrozdział poświęcony jest autorskiej modyfikacji modelu FOD, polegającej na zastosowaniu analizy PCA dla odpowiedniego określenia orientacji analizowanej cząstki białka. Podrozdziały od piątego do siódmego poświęcone są opisowi i przedstawieniu wyników eksperymentów obliczeniowych „kompleksowania” białek.

Czwarty rozdział „Dyskusja i wnioski” zawiera omówienie wyników uzyskanych w rozdziale trzecim. Kolejno dyskutowane są zaobserwowane własności algorytmu MOSF (podrozdział pierwszy), zalety dokonanej przez Doktoranta modyfikacji modelu FOD (podrozdział drugi), własności analizowanego zbioru 200 białek o strukturze homodimerów (podrozdział trzeci), omówienie wyników „kompleksowania” białek (podrozdział czwarty) oraz elementy innowacyjne zawarte w pracy (podrozdział piąty).

Piąty rozdział stanowi podsumowanie całej pracy oraz zawiera omówienie kilku możliwych kierunków dalszych badań.

Ocena rozprawy, najważniejsze osiągnięcia i elementy oryginalne rozprawy

Recenzowana praca stanowi oryginalne osiągnięcie naukowe. W cytowanej literaturze znajduje się kilkanaście artykułów naukowych, których współautorem jest

Doktorant. Ich tematyka jest zbieżna z tematyką recenzowanej pracy doktorskiej. Duża część z nich ukazała się w prestiżowych czasopismach naukowych.

Należy podkreślić interdyscyplinarność tematyki pracy. Obejmuje ona zarówno zagadnienia z obszaru biochemii jak też informatyki i bioinformatyki. Badania naukowe w zakresie takiej tematyki wymagały od Doktoranta dużej wiedzy i erudycji w zakresie różnych dyscyplin naukowych.

Na duże uznanie zasługuje głębokie i dociekliwe podejście doktoranta do badanych zagadnień. Żaden ze stosowanych algorytmów nie jest wykorzystywany na zasadzie „czarnej skrzynki”, bez refleksji co do jego budowy, własności, przydatności. Każda ze stosowanych metod jest przez Doktoranta wyczerpująco przedstawiana, dyskutowana, często analizowane są możliwości jej modyfikacji. Także, dobre wrażenie robi orientacja Doktoranta w biologicznych, biochemicznych aspektach obecnej w pracy problematyki.

Najważniejsze oryginalne osiągnięcia rozprawy to przeprowadzone i przeanalizowane eksperymenty obliczeniowe procesów „kompleksowania” białek oraz opracowanie nowych algorytmów optymalizacji z wykorzystaniem roju cząstek, użytecznych zwłaszcza w aspekcie optymalizacji wielokryterialnej.

Praca jest bardzo obszerna, dodatkowo odniesiona do bardzo szerokiej literatury (428 pozycji). Praca ma przy tym dobrą, logiczną strukturę. Zwraca uwagę staranna edycja. Wszystkie pozycje literatury są cytowane w tekście. Nie mam wątpliwości co do poprawności cytowań.

Uwagi dyskusyjne i krytyczne

W pracy nie formułuje się żadnych tez. W tym sensie praca nie ma standardowej struktury zwykle występującej w rozprawach doktorskich. Formułuje się za to cele pracy (strona 7), pierwszy to sprawdzenie założeń modelu FOD, drugi to użycie algorytmów roju cząstek do przewidywania struktury czwartorzędowej 200 białek homodimerycznych, trzeci to opracowanie autorskiego algorytmu optymalizacji wielokryterialnej, bazującego na metodzie roju cząstek, czwarty (poboczny) to modyfikacja modelu FOD.

Nasuwają się pewne pytania dotyczące celu pierwszego. Jakie są założenia modelu FOD? W jakim sensie wykonane eksperymenty obliczeniowe „kompleksowania” białek prowadzą do sprawdzenia założeń modelu FOD? Czy gdyby zamiast modelu FOD użyć oryginalnego modelu Kauzmanna to wyniki „kompleksowania” byłyby inne? Czy do realizacji eksperymentów „kompleksowania” można użyć algorytmów i dostępnego oprogramowania dynamiki molekularnej, dokowania białek? Jakie byłyby wyniki?

Co do celu drugiego i trzeciego, to przedstawione wyniki są bardzo interesujące i obiecujące. Jednak dla dobrego udokumentowania użyteczności opracowanych algorytmów

od strony metodologicznej potrzebne byłyby bardziej wyczerpujące badania. Można na przykład spróbować przeanalizować duże zbiory danych (lub ich fragmenty) publikowane w ramach serii konkursów CEC (Congress on Evolutionary Computation, CEC Competition on Evolutionary Multi-task Optimization). Przeprowadzenie bardziej szczegółowych badań dałoby szansę na opublikowanie opracowanych metod w czasopiśmie o profilu informatycznym.

Mimo, że istotnym dorobkiem pracy są algorytmy i oprogramowanie, do pracy nie są dołączone żadne pliki czy materiały, które pozwoliłyby przetestować funkcjonalność opracowanego przez Doktoranta oprogramowania.

Praca zawiera trochę nieściśłych, niejasnych oraz żargonowych określeń, terminów i sformułowań. Obecność nieściśłości oraz żargonowych określeń utrudnia zrozumienie wyników uzyskanych w pracy. Kilka przykładów podanych jest poniżej. W pracy jest ich więcej.

Strona 61: „*Nowy wektor prędkości każdej cząstki jest sumą trzech czynników. Pierwszy z nich stanowi jej obecna trajektoria*” – Powinno się napisać „sumą trzech składników”. Trajektoria nie może być ani czynnikiem ani składnikiem. Chodzi raczej o wektor styczny do trajektorii lub o wektor prędkości cząstki poruszającej się po trajektorii.

Strona 82: „*Algorytm autorstwa Wolfganga Kabscha służy do wyznaczania przekształceń liniowych minimalizujących pierwiastek średniej kwadratów różnicy (root mean square deviation, RMSD) dwóch równolicznych zbiorów wektorów*”. Algorytm Kabscha służy do oceny macierzy obrotu. Co prawda obrót jest szczególnym przypadkiem przekształcenia liniowego, ale jednak zdanie takie wprowadza zamieszanie. Dalej w tym samym zdaniu jest: „*różnicy ... dwóch równolicznych zbiorów*”. Nie chodzi tu jednak o różnicę zbiorów tylko o różnice współrzędnych wektorów.

Strona 82: „*Jeżeli punkty należące do zbiorów X i Y są zapisane odpowiednio w macierzach X i Y*” – sformułowanie żargonowe.

Strona 83: „*S jest macierzą korygującą – odmianą macierzy jednostkowej o rozmiarze $d \times d$* ” – nie istnieją odmiany macierzy jednostkowej.

Strona 83: „*Na koniec, do obliczenia pozostaje tylko optymalna macierz translacji T*” – Powinno być „wektor translacji”.

Strona 133: „*... wariacje atomów efektywnych*” – termin żargonowy

Strona 134: „*... rzutu atomów efektywnych na płaszczyznę YZ*” – termin żargonowy

Strona 135: „*Rzutowanie na nie całego zbioru atomów efektywnych powoduje obrót białka w taki sposób, że ich wariacja staje się największa w pierwszym wymiarze przestrzeni, potem w drugim, a na końcu – w trzecim.*” – Dlaczego rzutowanie powoduje obrót białka? „Wariacja atomów” – tak jak powyżej, niezrozumiałe, żargonowe określenie.

Konkluzja

Mimo uwag dyskusyjnych i krytycznych, osiągnięcia i oryginalne elementy rozprawy są na pewno wystarczające do jej ogólnej bardzo pozytywnej oceny. Bardzo wartościowe jest, że praca zawiera wszystkie składniki twórczego studium w dziedzinie nauk ścisłych, badania teoretyczne, obliczenia oraz użycie danych eksperymentalnych. Sposób prezentacji wyników potwierdza dużą erudycję Doktoranta oraz jego bardzo staranne i dociekliwe podejście do prowadzonych badań. Na bardzo duże uznanie zasługuje także duży dorobek publikacyjny Doktoranta, powiązany tematycznie z badaniami opisanymi w rozprawie.

Stwierdzam, że rozprawa spełnia warunki odpowiedniej Ustawy i wnioskuję o jej dopuszczenie do publicznej obrony.

Biorąc pod uwagę bardzo poważny dorobek publikacyjny doktoranta powinno się w mojej opinii pomyśleć o przyznaniu rozprawie wyróżnienia. W przypadku jeśli Komisja Doktorska będzie taką propozycję rozważała będę ją akceptował. Jednak biorąc pod uwagę fakt, że duża część dorobku ma znaczenie w biologii i biochemii, a ja oceniałem raczej stronę bioinformatyczną rozprawy, nie czuję się kompetentny aby wystąpić w mojej recenzji z formalnym wnioskiem o wyróżnienie.

