

Wrocław, 8 marca 2018 r.

Dr hab. inż. Małgorzata Kotulska, prof. PWr
Katedra Inżynierii Biomedycznej
Wydział Podstawowych Problemów Techniki
Politechnika Wroclawska
E-mail: malgorzata.kotulska@pwr.edu.pl

Recenzja rozprawy doktorskiej

mgr inż. Tomasza Żoka

zatytułowanej

Algorithmic Aspects of RNA Structure Similarity Analysis

1. Charakterystyka rozprawy

Przedstawiona mi do oceny praca doktorska pana Tomasza Żoka powstała w Instytucie Informatyki Politechniki Poznańskiej pod kierunkiem dr hab. inż. Marty Szachniuk. Rozprawa została napisana w języku angielskim, zawarto w niej krótkie streszczenia w języku polskim i angielskim. Wraz z bibliografią część główna pracy doktorskiej liczy 102 strony i składa się z 5 rozdziałów, spisu literatury zawierającego 72 pozycje, a także spisów ilustracji, tabel i algorytmów. Ponadto, co jest raczej nietypowe dla pracy doktorskiej, która nie jest zbiorem publikacji, do pracy włączono także obszernie omówienie 4 najważniejszych artykułów związanych z problematyką doktoratu, a także same artykuły. Jako odrębny dokument dołączona została też wersja elektroniczna 10-stronnicowego streszczenia pracy doktorskiej w języku polskim.

2. Problem badawczy i jego znaczenie

Bioinformatyczne modelowanie struktur molekularnych ma coraz większe znaczenie w badaniach biologicznych i biomedycznych. Nie tylko obniża koszty i podnosi dostępność badań, ale też umożliwia racjonalną analizę wyników, odkrywanie wzorców i lepsze zrozumienie praw natury na większą skalę. Komputery jednak nie działają bez inteligentnego wkładu człowieka. Błyskotliwa metoda potrafi wydrzeć naturze największe tajemnice. Metody bioinformatyczne muszą bazować na zrozumieniu biologii, ale kluczowe jest zastosowanie najnowszej wiedzy informatycznej do budowy konkretnych algorytmów i ich implementacji. Tylko zaawansowana metodyka informatyczna gwarantuje poprawne i szybkie obliczenia.

Problem jakim zajął się doktorant w swojej rozprawie wpisuje się w nurt najnowszych badań naukowych na świecie. RNA okazało się molekułą o wielu obliczach. Nie jest już pogardliwie uważana za etap pośredni w przekazywaniu informacji od DNA do białek, ujawniła się jako cząsteczka o ogromnym znaczeniu fizjologicznym, a w dodatku o skomplikowanej strukturze. Modelowanie budowy RNA może być bardzo trudne, gdyż inaczej niż to ma miejsce w modelowaniu białek, nie da się łatwo przewidzieć jego struktury drugorzędowej. Struktura drugorzędowa RNA, która jest istotnym stopniem pośrednim w modelowaniu pełnej struktury 3D, może przyjmować bardzo wielorakie formy. Społeczność naukowa RNA pracuje nad metodami i narzędziami, które umożliwią przewidywanie struktury, a stąd także funkcji RNA. Tematyka, której podjął się doktorant jest bardzo ważna i aktualna, na wyniki nowych metod oczekują badacze z całego świata i różnych nauk biomedycznych. Co istotne, rozwiązywanie problemów tego typu zwykle wymaga rozwoju metod podstawowych, w tym przypadku z dziedziny informatyki.

3. Wkład autora

Praca ma charakter aplikacyjny ukierunkowany na bioinformatykę. Autor opracował nowe metody i algorytmy do analizy podobieństwa struktur 2D i 3D RNA. W szczególności, w odniesieniu do struktur 2D, zaproponował nowy model grafowy oraz algorytmy PMMG i IMMIG pozwalające wyszukiwać fragmenty struktur spełniające cechy zadanego motywu strukturalnego. Przy braku dokładnego dopasowania algorytm IMMIG znajduje wynik najbardziej podobny, na czym polega innowacyjne podejście autora. Natomiast w odniesieniu do struktur 3D autor wprowadził trygonometryczną reprezentację struktur, gdzie wykorzystał zaproponowaną przez siebie nową metrykę kątową. Wykazał też, że średnia MCQ oraz mediana MedCQ pozwalają na dobre porównanie podobieństwa struktur. Wkład w rozwój metodologii badań struktury RNA jest udokumentowany współautorstwem doktoranta w 9 publikacjach w czasopiśmie z bazy JCR, z tego w 2 jako pierwszy autor. Spora część z tych czasopism należy do czołówki w swojej dziedzinie.

Efektom badań autora jest zarówno opracowanie nowych metod, jak również ich implementacja w postaci gotowych narzędzi. Zaproponowane metody znalazły zastosowanie w programach do modelowania RNA, rozwijanych w zespole w którym powstała omawiana praca doktorska. Efektywność procesu modelowania zależy od skutecznego znajdowania najbardziej podobnych motywów strukturalnych z biblioteki. RNAComposer, który przewiduje strukturę trzeciorzędową dużych cząsteczek RNA, korzysta z informacji o potwierdzonych strukturach fragmentów RNA, zebranych w bazie RNA FRABASE. Algorytmy pozwalające

na lepsze porównanie struktur molekularnych, a w szczególności takie, które umożliwiają znalezienie struktur podobnych jednak nieidentycznych, mogą bardzo poprawić skuteczność modelowania. Wkład autora w rozwój narzędzi typu RNAComposer i RNA FRABASE, które są często wykorzystywane przez zespoły naukowe zajmujące się badaniem RNA, uważam za bardzo istotne osiągnięcie.

Ponadto, miary zaproponowane przez autora zostały wykorzystane w międzynarodowym konkursie RNA-Puzzles, co wskazuje na ich wysoką skuteczność i oryginalność.

4. Poprawność

Autor przekonująco wykazuje zalety zaproponowanych przez siebie metod do oceny podobieństwa cząstek RNA, zarówno w odniesieniu do ich struktury drugorzędowej, jak i trzeciorzędowej. W części merytorycznej pracy nie napotkałam żadnych istotnych błędów, jakkolwiek w czasie czytania nasunęły mi pewne uwagi i pytania:

- Na str. 35 autor opisuje czym jest mieszany graf. Byłoby dużo lepiej, gdyby przedstawił graficznie jego reprezentację na tym samym przykładzie co poprzednio (rys.3.2). Pokazałoby to różnice pomiędzy podejściem autora a dotychczas stosowanymi formatami.
- Czy jakakolwiek rozszerzona reprezentacja kątowa struktury cząsteczki była już wcześniej stosowana przez innych autorów i czy jej skuteczność ogranicza się tylko do RNA? Chociaż prawdopodobnie nie ma ona zastosowania w białkach, których struktura ogranicza swobodę kątów torsyjnych, jednak być może ułatwiłaby modelowanie innych molekuł. Jeśli tak, to do jakiego typu związków innych niż RNA mogłaby być przydatna i czy autor widzi potencjał włączenia swojej metody do innych technik modelowania?
- Brakuje mi porównania efektywności zastosowania miar MCQ i MedCQ, które stratnie kompresują informację trygonometryczną, z zastosowaniem pełnej informacji o wszystkich kątach. Jak dużo jest np. wskazań fałszywie pozytywnych wynikających z uśredniania?
- Nie jest dla mnie jasna tabela 3.5 ze str. 54. Jak, w sposób automatyczny, zostaje wybrana metoda korekty przewidywania (czyli wybór konkretnego wiersza z tabeli) i jak dokładnie przebiega procedura modyfikacji?
- Jak rozumieć tabelę 4.2 na stronie 72? Autor nie wyjaśnia znaczenia wartości zawartych w niej liczb z zakresu 0-3.

Oprócz merytorycznej poprawności proponowanego rozwiązania problemu badawczego, bardzo istotne jest klarowne jego przedstawienie. Pracę rozpoczyna informatywny i ciekawy wstęp przedstawiający tematykę badań oraz aktualny stan wiedzy oraz przegląd aktualnie stosowanych technik i ich ograniczenia. W sposób bardzo płynny, od strony 34 autor przechodzi do własnych rozwiązań przedstawionych problemów. W tym miejscu chciałabym też zwrócić uwagę na niezwykle staranność edycyjną przedstawionego mi dokumentu. Mimo, że praca została napisana w języku angielskim, poprawność językowa jest na bardzo wysokim poziomie.

W pracy są też oczywiście drobne mankamenty techniczne. Przykładowo:

- rysunki 4.3 (str. 59) i 4.8 (str.71) są zbyt małe, przez co są nieczytelne,
- nie wszystkie symbole są opisane, np. *dm* na str. 57 (distance measure?),
- nie jest podany sposób obliczenia wielkości INF. Skoro autor podaje definicje znacznie bardziej rozpowszechnionego RMSD, powinien też konsekwentnie zdefiniować pozostałe wielkości, nie poprzestając na cytowaniu publikacji,
- format bibliografii jest dosyć nietypowy, gdyż cytując prace według nazwisk autorów, w bibliografii powinno się raczej rozpocząć od nazwiska autora.

5. Wiedza kandydata

Kandydat wykazał się wiedzą z dziedziny bioinformatyki, jak również ogólnej wiedzy informatycznej, którą skutecznie zastosował do rozwiązywanego problemu. Wiedzę ogólną z dyscypliny przedstawił w rozdziale 2 i 3, gdzie dokonał przeglądu metod, odwołując się do poprawnie wybranej i aktualnej bibliografii. Omówione zostały zarówno podstawy biologiczne problemu, jak i odniósł się do reprezentacji struktur za pomocą grafów. Co ważne, potrafił nabytą wiedzę w sposób twórczy i oryginalny wykorzystać do rozwoju dyscypliny, którą reprezentuje.

6. Inne uwagi

Mimo, że włączanie do pracy doktorskiej wybranych publikacji jest raczej nietypowe i może być uważane za kontrowersyjne, jednak pomaga to recenzentowi osiągnąć lepszy ogólny obraz całości dorobku doktoranta. Uważam, że w tym wypadku zabieg ten spełnił swoje zadanie.

7. Podsumowanie

Uważam, że mgr inż. Tomasz Żok zaprezentował bardzo rzetelną pracę doktorską, potwierdzoną publikacjami w najlepszych czasopismach z dziedziny, a także praktyczną implementacją opracowanych metod, która wzbogaciła narzędzia modelujące strukturę RNA. Co ważne, nowa metodyka została też zaadoptowana do znanego międzynarodowego konkursu.

Biorąc pod uwagę opinie zaprezentowane w poprzednich punktach i wymagania zdefiniowane przez artykuł 13 Ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym (z późniejszymi zmianami) moja ocena rozprawy pod względem trzech podstawowych kryteriów jest następująca:

A. Czy rozprawa zawiera oryginalne rozwiązanie problemu naukowego?

X				
Zdecydowanie TAK	Raczej TAK	Trudno powiedzieć	Raczej NIE	Zdecydowanie NIE

B. Czy po przeczytaniu rozprawy zgadzasz się, że kandydat posiada ogólną wiedzę teoretyczną w dyscyplinie Informatyka lub Automatyka i Robotyka?

X				
Zdecydowanie TAK	Raczej TAK	Trudno powiedzieć	Raczej NIE	Zdecydowanie NIE

C. Czy kandydat posiada umiejętność samodzielnego prowadzenia pracy naukowej?

X				
Zdecydowanie TAK	Raczej TAK	Trudno powiedzieć	Raczej NIE	Zdecydowanie NIE

Ponadto, biorąc pod uwagę wysoki poziom rozprawy doktorskiej, duże znaczenie wyników dla dziedziny modelowania RNA oraz imponujący dorobek publikacyjny kandydata rekomenduję wyróżnienie rozprawy doktorskiej.


Małgorzata Kotulska