

Prof. dr hab. inż. Franciszek Seredyński
Uniwersytet Kardynała Stefana Wyszyńskiego w
Warszawie
Wydział Matematyczno-Przyrodniczy.SNS
Instytut Informatyki
f.seredyński@uksw.edu.pl

Warszawa, 25.01.2018

Recenzja rozprawy doktorskiej

mgr Natalii Szóstak

zatytułowanej:

Bioinformatyczne metody modelowania i weryfikacji hipotezy Świata RNA

1. Problem badawczy i jego znaczenie

Pytania dotyczące pojawienia się życia na Ziemi jak też hipotezy próbujące wyjaśnić jego źródła zaczęły pojawiać się w różnej formie prawdopodobnie wraz z pojawieniem się na Ziemi gatunku homo sapiens. Milowymi krokami w tym kierunku dokonanymi już na gruncie nowożytnej biologii było odkrycie cząsteczki DNA, później poznanie jej struktury, a zupełnie niedawno poznanie struktury genomu. Rosnąca wiedza biologiczna jak też osiągnięcia w zakresie chemii pozwalały na formułowanie licznych hipotez wyjaśniających powstanie życia na Ziemi. Do najbardziej uznawanych hipotez jest hipoteza Świata RNA, która zakłada, że to samoreplikujące się cząsteczki RNA, a nie białka były prekursorem życia na Ziemi. Hipoteza ta jest przedmiotem badań doktorantki, która postawiła sobie za cel jej weryfikację z użyciem metod bioinformatycznych.

Historia informatyki pokazuje, że zagadnienie powstania życia i jego istota jest również przedmiotem jej zainteresowania. Kwestia ta pojawiła się już na etapie poprzedzającym pojawienie się pierwszych komputerów. Automaty komórkowe zaproponowane przez John von Neumann zainspirowanego pomysłami Stanisława Ulama pojawiły się w pierwszej kolejności, aby zrealizować koncepcję samoreprodukcji - jednej z najważniejszych cech żywych organizmów. W badaniach dotyczących modelu automatu komórkowego określanego jako *Gra w Życie* zaproponowanego w 1970 r przez Johna Conway'a zastanawiano się czy w ramach tego systemu może powstać życie. Rok 1987 to data powstania interdyscyplinarnego kierunku badań określonego jako *Sztuczne życie* zorientowanego na zrozumienie i wykorzystanie istoty życia, którego pomysłodawcą był matematyk i informatyk Christopher Langton,

Pojawienie się na styku biologii, matematyki i informatyki nowej dziedziny wiedzy – bioinformatyki, było więc naturalnym procesem, ze względu na szereg łączących je wspólnych zainteresowań i nowych możliwości oferowanych oddzielnie przez każdą z tych dziedzin. Rosnące tempo badań w zakresie biologii otwierało nowe problemy teoretyczne, w tym problemy o dużej złożoności obliczeniowej. Matematyka dawała możliwości tworzenia i badania prostych i eleganckich modeli zjawisk biologicznych. Informatyka oferowała moce obliczeniowe i metodologię badań symulacyjnych tam, gdzie dokładność modeli matematycznych nie była zadowalająca. To dlatego właśnie doktorantka postanowiła na gruncie bioinformatyki dokonać weryfikacji istniejących modeli i dowodów na poparcie hipotezy Świata RNA, jak też skonfrontować je z własnymi propozycjami.

Hipoteza Świata RNA to złożony naukowy problem badawczy mający interdyscyplinarny charakter niezwykle istotny zarówno dla biologii, ale też w powodów wyżej przedstawionych dla informatyki. Na pierwszy rzut oka problem ten wydaje się być problemem czysto naukowym i tak być może jest z perspektywy nauk biologicznych. Obserwując jednak dynamiczny rozwój bioinformatyki i jego wpływ na rozwój nowych technologii jestem przekonany, że każdy postęp w jego rozwiązaniu będzie skutkował pojawieniem się nowych metodologii i technologii informatycznych.

Mimo istnienia wielu badań mających charakter dowodów na poprawność tej hipotezy można stwierdzić, że droga do jej pełnego udowodnienia nie jest prosta. Dzieje się tak dlatego, że problem badawczy związany z Hipotezą Świata RNA ma złożony wielowymiarowy charakter, a tworzone przez badaczy modele są w stanie, ze względu na stosowane dostępne metodologie, uwzględnić tylko niektóre aspekty problemu. Tworzenie w ramach modeli odpowiednich powiązań między elementami ma charakter roboczych hipotez, które w ramach prowadzonych badań są pomyślnie bądź negatywnie weryfikowane. Doktorantka przystępując do badań bardzo dobrze rozpoznała istniejący stan wiedzy w zakresie proponowanych modeli rozwiązań badanego problemu, ich zalet jak też wad i w ramach swojej pracy zaproponowała oryginalną bioinformatyczną metodologię badań i uzyskała wyniki, które w znaczący sposób wzbogacają wiedzę o mechanizmach powstania życia na Ziemi w ramach hipotezy Świata RNA.

2. Wkład autorki

Doktorantka w swojej rozprawie dotyczącej potwierdzenia hipotezy Świata RNA uzyskała szereg oryginalnych i znaczących dla współczesnej nauki wyników tworzących logiczną całość.

Wśród nich chciałbym wskazać trzy następujące główne wyniki.

Pierwszym z nich jest zaproponowany w Rozdziale 5 bioinformatyczny model początków życia na Ziemi, w którym wyróżniono cztery poziomy organizacji prebiotycznego świata, a następnie zdefiniowano na najwyższym poziomie organizacyjnym za pomocą języka ML-Rules hipotezę Świata RNA. Kluczowym elementem tego najwyższego poziomu organizacyjnego jest interakcja (model RP) między cząsteczkami typu pasożyt modelującymi łańcuchy RNA nie posiadające zdolności katalizowania replikacji, a cząsteczkami typu replikaza modelującymi łańcuchy RNA, które są zdolne do przeprowadzania reakcji replikazy.

Drugim bardzo interesującym wynikiem pracy również przedstawionym w Rozdziale 5 jest zaproponowanie koncepcji rozwiązania problemu powstania życia na Ziemi. Koncepcja ta zakłada realizację i badanie modelu poprzez sekwencję czterech metodologii: ewolucyjną teorię gier, równania różniczkowe, automaty komórkowe oraz systemy wieloagentowe. Metodologie te mają charakter kolejnych filtrów umożliwiających tworzenie coraz bardziej realistycznych modeli z zachowaniem właściwych głównych parametrów każdego następnego modelu.

Trzecim osiągnięciem pracy to wyniki prac teoretyczno-eksperymentalnych uzyskanych z użyciem wspomnianych czterech metodologii i wykazanie możliwości stabilności systemu jak też warunków tej stabilności, a w szczególności:

- Z użyciem ewolucyjnej teorii gier wykazano możliwość koegzystencji w uproszczonym modelu systemu składającym się z graczy- pasożytów i replikaz, jak też określono warunki brzegowe istnienia rozwiązania (Rozdział 6)

- Zrewidowano znany w literaturze model RP wykorzystujący równania różniczkowe i zbadano go przy zmodyfikowanych założeniach oraz określono wartości analizowanych parametrów systemu zapewniające jego stabilność jak też pokazano ograniczenia tego podejścia (Rozdział 7)
- Zaproponowano przestrzenny 2-wymiarowy model symulacyjny RP wykorzystujący automaty komórkowe, który umożliwił obejście ograniczeń równań różniczkowych, jak też głównego ograniczenia ewolucyjnej teorii gier – braku mutacji parametrów. Pokazano, że dla określonych parametrów modelowanego systemu możliwa jest stabilna koegzystencja pasożytów i replikaz, jak też wykazano istotną rolę dla stabilności systemu aspektu przestrzennego oferowanego przez automaty komórkowe i zjawiska mutacji (Rozdział 8)
- Zaproponowano symulacyjny model systemu RP wykorzystujący metodologię systemów wieloagentowych, który bazuje na modelu opartym o automaty komórkowe, ale w odróżnieniu od poprzedniego podejścia umożliwia przemieszczanie się cząsteczek w przestrzeni ciągłej. Model ten umożliwia również modelowanie kilku typów reakcji chemicznych. Badania modelu pokazały, że w określonych sytuacjach system pasożytów i replikaz jest stabilny, ale znacząco zawężono warunki tej stabilności. Było to możliwe dzięki zastosowaniu metodologii systemów wieloagentowych, które umożliwiają znacznie bardziej realistyczne modelowanie systemów biologicznych (Rozdział 9).

Ważnym osiągnięciem rozprawy jest też opracowanie środowiska symulacyjnego niezbędnego do przeprowadzenia eksperymentalnych badań oraz szeregu algorytmów niezbędnych do analizy i weryfikacji hipotezy Świata RNA.

Na specjalne uznanie zasługuje Rozdział 10 rozprawy, w którym doktorantka w sposób obszerny i szczegółowy konfrontuje swoje wyniki z wynikami opublikowanymi do tej pory. Rozdział ten jest bardzo cenny i praktycznie gotowy do publikacji w dobrym czasopiśmie. Recenzując różne prace doktorskie widzę po raz pierwszy tak rzeczowe podejście do badań naukowych u młodego naukowca, co świadczy o dużej dojrzałości doktorantki w zakresie prowadzenia badań naukowych.

Podsumowując, główny wynik uzyskany w rozprawie można ująć w sposób następujący. Doktorantka wykazała, że aby życie, które aktualnie znamy mogło wyewoluować, musiało nastąpić na wczesnych etapach ewolucji rozdzielenie funkcji pełnionych przez pierwotne replikazy na dwie oddzielne cząsteczki, z których jedna pełniła funkcję magazynu informacji, a druga funkcję katalizatora powielania tej informacji.

3. Poprawność

Badania symulacyjne, które są bardzo ważnym elementem rozprawy i dzięki którym uzyskano główne wyniki mogą być potencjalnie poddane krytyce ze względu np. na mało realistyczne warunki symulowanego zjawiska. Zdecydowanie można powiedzieć, że tak nie jest w przypadku recenzowanej rozprawy i to z kilku powodów.

Doktorantka podeszła do rozwiązywanego problemu kompleksowo rozpoczynając swoje rozważania od pierwszego formalnego matematycznego modelu, przechodząc następnie do innego formalnego modelu, a dalej stworzyła dwa kolejne modele symulacyjne różniące się możliwościami modelowania rzeczywistości. Modele formalne były potwierdzeniem faktu dotyczącego możliwości modelowania zjawiska życia z użyciem pewnych parametrów. Były też swoistymi filtrami

parametrów i ich wartości ze względu na modelowane zjawisko i uświadamiały ograniczone możliwości tych metod ze względu na nieuwzględnienie przez nie pewnych oczywistych parametrów. Doktorantka już na tym etapie badań z użyciem metod formalnych weryfikowała uzyskiwane wyniki z wynikami istniejącymi w literaturze i dzięki temu była w stanie znaleźć i poprawić błąd w modelu znanym z literatury. Tak więc przejście do modeli symulacyjnych następowało na etapie zakończonej weryfikacji modelu życia i ustalenia pewnych parametrów i ich wartości i było świadomym krokiem wynikającym z obserwacji ograniczonych możliwości rozpatrywanych formalnych modeli i konieczności rozszerzenia modelu życia w celu bardziej dokładnego opisu zjawiska życia.

Tworząc modele symulacyjne zjawiska ewolucji życia, modelując proces ewolucji życia drogą symulacji komputerowej i obserwując uzyskiwane wyniki doktorantka na każdym z tych etapów konfrontowała swoje decyzje z istniejącymi w literaturze wynikami. Ta konfrontacja zapewniała rygorystyczne warunki poprawności tworzonych modeli symulacyjnych i dużą wiarygodność uzyskiwanych wyników. Jestem przekonany, że stopień wiarygodności uzyskanych wyników jest co najmniej taki sam jak wyników prezentowanych aktualnie w literaturze przez elitarną grupę naukowców zajmujących się problematyką pojawienia się życia, bądź nawet większy, mając na uwadze odnalezione przez doktorantkę błędy w opublikowanych już modelach innych autorów.

Koronnym dowodem poprawności uzyskanych w rozprawie wyników jest prezentacja i akceptacja tych wyników na szeregu krajowych i międzynarodowych konferencjach naukowych, publikacje tych wyników w szeregu renomowanych czasopismach, w tym w czasopiśmie z listy filadelfijskiej PLOS ONE.

Przyglądając się opisom badań symulacyjnych zwróciły moją uwagę dwie następujące kwestie, które z mojej perspektywy warte byłyby dodatkowego komentarza doktorantki:

- W eksperymentach symulacyjnych zarówno z użyciem automatów komórkowych jak i systemów wieloagentowych stosowano dwa ustawienia początkowe pasożytów i replikaz: ustawienie w postaci koła (półkoła) oraz ustawienie losowe; ustawienie losowe dla rozpatrywanego modelu życia wydaje się naturalne, a ustawienie w postaci koła raczej wydaje się być specyficzne; na str. 128 rozprawy znajdujemy informację, że w przypadku systemu wieloagentowego w jednym z eksperymentów z użyciem losowego ustawienia agentów reprezentujących pasożyty i replikazy system w większości przypadków nie był w stanie zachować stabilnej koegzystencji między replikazami i pasożytami, natomiast było to możliwe w przypadku zastosowania kołowego ustawienia dwóch typów cząsteczek RNA; w jaki sposób można wyjaśnić takie zachowanie się modelu ?
- Model wykorzystujący systemy wieloagentowe, w odróżnieniu od modelu wykorzystującego automaty komórkowe daje możliwość traktowania czasu i przestrzeni jako wartości ciągłych; jak rozumiem, w swoim modelu doktorantka wykorzystywała tę własność w odniesieniu do przestrzeni, natomiast czas traktowała jako wielkość dyskretną; co było przyczyną tego i czy to mogło mieć wpływ na wyniki symulacji ?

4. Wiedza kandydatki

Wiedza zaprezentowana przez doktorantkę w rozprawie jest imponująca. W Rozdziale 1, sekcja 1.1 doktorantka dokonuje przeglądu literatury dotyczącej dowodów istnienia hipotezy Świata RNA oraz metod modelowania złożonych systemów biologicznych. W Rozdziale 2, w sekcjach 2.2, 2.3 oraz 2.4 dokonuje krótkich przeglądów literaturowych związanych ze stosowanymi w rozprawie metodami wykorzystującymi odpowiednio ewolucyjną teorię gier, automaty komórkowe oraz systemy wieloagentowe.

W Rozdziale 3 w sekcji 3.1 znajduje się przegląd literaturowy znanych definicji życia z różnych perspektyw, a w sekcji 3.2 znajdujemy szereg odnośników literaturowych związanych z poziomami organizacji przestrzennej cząsteczek biologicznych DNA, RNA i białek. W sekcji 3.3 doktorantka analizuje aktualny stan wiedzy dotyczący ewolucji, a w sekcji 3.4 analizuje opisane w literaturze eksperymenty związane z hipotezą świata RNA. Również takie zagadnienia jak quasi-gatunek (sekcja 3.6), hipercykl (sekcja 3.7) oraz zbiór autokatalityczny (sekcja 3.8) są dyskutowane w oparciu o bieżącą literaturę. Rozdział 4 prezentujący opis stosowanych przez doktorantkę metod modelowania biologicznego, ze wskazaniem ich zalet i wad został również zredagowany z wykorzystaniem literatury przedmiotu.

Założenia do formułowanych modeli matematycznych i symulacyjnych oraz wyniki badań własnych doktorantki przedstawione w Rozdziałach 6, 7, 8 i 9 są na bieżąco konfrontowane z aktualnymi pozycjami bibliograficznymi, a Rozdział 10 zawiera całościową dyskusję nt. uzyskanych w rozprawie wyników na tle wyników znanych w literaturze.

Cytowana w rozprawie literatura obejmująca 145 pozycji bibliograficznych jest obszerna i wyczerpująca. Przeprowadzone analizy literaturowe potwierdzają dogłębną wiedzę doktorantki w zakresie bioinformatyki jak też samej informatyki.

5. Inne uwagi¹

Rozprawa posiada przemyślaną logiczną strukturę rozdziałów oraz sekcji, a struktury językowe są precyzyjne i nie budzą zastrzeżeń. Wyniki badań przedstawiane są w formie wykresów, tabel oraz zrzutów ekranowych. Tworzy to razem estetyczną formę i dzięki temu czyta się ją z przyjemnością i zrozumieniem, mimo tego, że dotyka ona szerokiego spektrum wiedzy obejmującej biologię, matematykę i informatykę.

6. Podsumowanie

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)² moja ocena rozprawy pod względem trzech podstawowych kryteriów jest następująca:

¹ Opcjonalnie

² http://www.nauka.gov.pl/g2/oryginal/2013_05/b26ba540a5785d48bee41aec63403b2c.pdf

A. Czy rozprawa zawiera oryginalne rozwiązanie problemu naukowego? (wybierz jedną opcję stawiając znak 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 ?

Zdecydowanie
TAK

Raczej TAK

Trudno
powiedzieć

Raczej NIE

Zdecydowanie
NIE

C. Czy kandydat posiada umiejętność samodzielnego prowadzenia pracy naukowej?

Zdecydowanie
TAK

Raczej TAK

Trudno
powiedzieć

Raczej NIE

Zdecydowanie
NIE

Ponadto, biorąc pod uwagę takie czynniki jak ważność i oryginalność tematyki badawczej rozprawy, wielostronne podejście do rozwiązania postawionego problemu, oryginalność uzyskanych wyników potwierdzonych publikacjami w renomowanych czasopismach z listy filadelfijskiej jak też wzorowy sposób analizy uzyskanych wyników i porównania ich z aktualnymi opublikowanymi wynikami innych autorów rekomenduję wyróżnienie rozprawy doktorskiej³.



Podpis

³ Oczywiście to zdanie jest opcjonalne.