

RECENZJA

rozprawy doktorskiej mgr inż. Ruby Khan pt. „Development of Methods for Identifying Key Variables in Complex Mathematical Models of Biological Systems”

Promotor: dr hab. inż. Krzysztof Puszyński, prof. Pol. Śl.

I. Problematyka naukowa oraz przedmiot rozprawy

Imponujące postępy nauk biologicznych, które można obserwować w ciągu ostatnich trzech dekad, były i są możliwe w dużej mierze dzięki zastosowaniu odpowiednich metod matematycznych oraz informatycznych. Jedną z przyczyn takiego stanu rzeczy jest pojawienie się dużych zbiorów danych biologicznych, zapoczątkowane dynamicznym rozwojem metod sekwencjonowania DNA w połowie lat 90. XX wieku, których analiza w praktyce jest niemożliwa bez wykorzystania metod algorytmicznych. Inną przyczyną jest fakt, że organizmy żywe składają się ze złożonych systemów występujących na różnych poziomach ich organizacji (organy, tkanki, komórki itp.). Wzajemne oddziaływania elementów tych systemów są kluczowe dla ich funkcjonowania, a zatem ich dokładne poznanie wydaje się być niezbędne dla dogłębnego zrozumienia zasad, na których oparte jest działanie świata ożywionego. Oddziaływania te są jednak bardzo złożone, stąd ich analiza bez zastosowania odpowiednich metod matematycznych oraz informatycznych jest praktycznie nierealizowalna. Spostrzeżenia te stały się impulsem do stosowania metod pochodzących z nauk systemowych do badania złożonych zjawisk biologicznych oraz rozwoju biologii systemów – nauki funkcjonującej na pograniczu biologii, matematyki, informatyki i obszarów pokrewnych, na gruncie której badane są systemy biologiczne. Warto zauważyć, iż systemy takie posiadają swoją specyfikę, która powoduje, że bezpośrednie zastosowanie do ich analizy metod zaprojektowanych pierwotnie do badania systemów innego rodzaju może być w wielu przypadkach niewystarczające. Rodzi to potrzebę opracowywania metod uwzględniających tę specyfikę. Konstrukcja tego rodzaju metod oraz ich zastosowanie stanowi obecnie bardzo intensywnie rozwijający się nurt badawczy.

Jednym z wielu obszarów, w których metody systemowe są szczególnie istotne, jest badanie procesów związanych z powstawaniem i rozwojem chorób nowotworowych. Wiadomo od dawna, iż choroby te są bardzo złożone – na ich powstawanie i przebieg ma wpływ wiele czynników oddziałujących ze sobą i ze środowiskiem, w którym rozwija się nowotwór, w bardzo skomplikowany sposób. Nie ułatwia to oczywiście projektowania skutecznych terapii przeciwnowotworowych. Niezbędne do tego jest dokładne poznanie procesów nowotworowych, co ze względu na wspomnianą ich dużą złożoność stanowi bardzo istotne wyzwanie.

Magister Ruby Khan w swojej rozprawie doktorskiej zajęła się problemem

zaprojektowania metod analizy modeli matematycznych złożonych systemów biologicznych. Modele takie są podstawą analizy prowadzonej za pomocą ścisłych metod, a ze względu na charakter oddziaływań między elementami systemów biologicznych szczególnie przydatne są modele grafowe, czy szerzej, sieciowe. Doktorantka w swojej rozprawie proponuje zastosowanie zestawu metod, za pomocą których można badać wybrane parametry tego rodzaju modeli i na tej podstawie formułować pewne wnioski dotyczące własności analizowanych systemów biologicznych. Zaproponowany zestaw metod został przetestowany na danych dotyczących m. in. raka jajnika. Uzyskane wyniki wskazują, iż za pomocą zaprojektowanego zestawu metod możliwe jest wykrywanie kluczowych elementów badanych systemów biologicznych.

II. Analiza treści rozprawy oraz uzyskanych wyników

1. Treść rozprawy

Rozprawa doktorska mgr inż. Ruby Khan napisana jest w języku angielskim i składa się z pięciu rozdziałów oraz bibliografii, poprzedzonych streszczeniem w językach angielskim i polskim, spisem treści oraz listą publikacji i wystąpień konferencyjnych Autorki. Na końcu rozprawy zamieszczone zostały dodatki.

Rozdział pierwszy zawiera wprowadzenie do tematyki rozprawy, sformułowanie celu badawczego oraz postawionych hipotez, a także bardzo zwięzłe omówienie zastosowanej metodologii. Głównym celem sformułowanim przez Doktorantkę jest opracowanie i weryfikacja innowacyjnych technik obliczeniowych służących do identyfikacji kluczowych zmiennych w złożonych systemach biologicznych.

W rozdziale drugim przedstawiony został przegląd literatury oraz historia stosowania metod matematycznych do modelowania i analizy złożonych systemów biologicznych. Zostały tu również krótko omówione wybrane zagadnienia związane z analizą modeli systemów biologicznych oraz wskazane są metody, które były stosowane w badaniach opisanych w rozprawie.

Rozdział trzeci poświęcony jest metodologii zastosowanej przez Doktorantkę do przeprowadzenia badań będących przedmiotem rozprawy. Została w nim przedstawiona struktura zaprojektowanego przez Doktorantkę potoku obliczeniowego oraz omówione zostały poszczególne jego składniki, tj. pobieranie oraz integracja danych niezbędnych do zbudowania modelu analizowanego systemu biologicznego, implementacja bazy danych, konstrukcja modelu sieciowego, algorytm PageRank, algorytm spacerów losowych, modelowanie boolowskie oraz rekurencyjne splotowe sieci neuronowe. Wraz z przedstawieniem wspomnianych składników zostało też zarysowane ich zastosowanie do badania analizowanego systemu biologicznego.

Rozdział ten składa się z bardzo wielu krótkich podrozdziałów, których zawartość ma w wielu przypadkach charakter bardzo techniczny.

Rozdział piąty zawiera przedstawienie uzyskanych wyników. Omówiono w nim zastosowane metody pozyskiwania danych biologicznych z różnych baz danych, implementację bazy danych wykorzystywaną do przechowywania pozyskanych danych, konstrukcję modelu sieciowego na podstawie wspomnianych danych, wyniki analizy własności strukturalnych skonstruowanego modelu, wizualizacje uzyskanej sieci, wyniki analiz przeprowadzonych za pomocą algorytmu PageRank oraz algorytmu spacerów losowych, a także modelowania boolowskiego. Ponadto, w rozdziale tym przedstawione zostały wyniki dodatkowych eksperymentów mających potwierdzić przydatność

zaprojektowanego potoku obliczeniowego również w przypadku innych systemów biologicznych niż ten związany z rakiem jajnika, który był przedmiotem głównych analiz. Doktorantka skonstruowała i przeanalizowała modele cyklu komórkowego oraz sieci sygnałowej MAPK. Wyniki tych analiz zamieszczone zostały w omawianym rozdziale.

Podobnie jak rozdział trzeci, również i ten rozdział składa się z bardzo wielu krótkich podrozdziałów, z których duża część ma charakter przywodzący na myśl raport techniczny.

Rozdział piąty zawiera omówienie uzyskanych wyników oraz wnioski końcowe. Doktorantka omówiła w nim m. in. zgodność uzyskanych wyników biologicznych z wiedzą na temat analizowanych procesów dostępną w literaturze oraz wskazała na osiągnięcie celu rozprawy, a także pozytywną weryfikację postawionych hipotez badawczych.

Rozprawę zamyka spis literatury zawierający 130 pozycji oraz dodatki, na które składają się kody programów w języku Python, a także dodatkowe tabele.

2. Uwagi merytoryczne i redakcyjne

Generalnie rozprawa mgr inż. Ruby Khan jest interesująca i dotyczy ważnego, zarówno z teoretycznego, jak i praktycznego punktu widzenia, zagadnienia. Jak było wspomniane w początkowej części niniejszej recenzji, analiza złożonych systemów biologicznych to bardzo istotne zagadnienie i wciąż potrzebne są bardziej efektywne i dokładne metody, za pomocą których taką analizę można by prowadzić (wiele wskazuje na to, że w dziedzinie modelowania i analizy złożonych systemów biologicznych jesteśmy dopiero na początku drogi). Warto też wspomnieć, iż wyniki badania systemów biologicznych mogą znaleźć bezpośrednie zastosowanie w wielu obszarach, z których medycyna jest jednym z najważniejszych. A zatem, nie powinno być wątpliwości, iż Doktorantka podjęła bardzo istotny temat i wyznaczyła ważny cel w ramach swojej rozprawy doktorskiej.

Tytuł rozprawy mgr inż. Ruby Khan dobrze odpowiada jej zawartości, gdyż Doktorantka zaproponowała metodę, mającą postać potoku obliczeniowego, służącą do analizy modeli systemów biologicznych wyrażonych w postaci sieci, w szczególności do określenia pewnych istotnych cech charakterystycznych tego typu modeli.

Układ rozprawy doktorskiej jest zasadniczo poprawny – zawiera ona wstęp, w którym Doktorantka przedstawiła przedmiot rozprawy, cele i hipotezy, po nim następują rozdziały zawierające przegląd literatury, omówienie zastosowanej metodologii oraz uzyskanych wyników, a rozprawa kończy się rozdziałem poświęconym dyskusji tychże wyników i wnioskowi końcowemu. Pewne wątpliwości i zastrzeżenia budzi jednak podział rozdziałów trzeciego i czwartego na bardzo dużą liczbę krótkich (w wielu przypadkach bardzo krótkich) podrozdziałów. Ich zawartość, ma często charakter techniczny, czy wręcz implementacyjny (dotyczy to zwłaszcza rozdziału trzeciego), a informacje podane są nierzadko w sposób hasłowy. Powoduje to, że główna myśl będąca przedmiotem rozprawy ginie w gąszczu niepotrzebnych, jak się wydaje, szczegółów. W rozprawie doktorskiej najbardziej istotne są główne idee badawcze, a nie szczegóły techniczne i to one powinny być przede wszystkim wyeksponowane. Zastosowany przez Doktorantkę układ rozprawy nie ułatwia dostrzeżenia i docenienia tych idei.

Doktorantka w prawidłowy sposób korzysta z literatury, choć pewien niedosyt budzi liczba odwołań do literatury dotyczącej biologicznych aspektów modelowanych procesów. Jest to jednak bezpośrednio związane z faktem, że w rozprawie nie ma omówienia biologicznych procesów, które były modelowane za pomocą zaproponowanej metody. Ten brak budzi pewne zdziwienie, gdyż omówienie takie wydaje się być czymś naturalnym i pozwoliłoby lepiej zrozumieć efekty zastosowanych metod, zwłaszcza w zakresie tworzenia

modeli. Jeżeli Doktorantka wybrała pewne procesy biologiczne do przetestowania zaproponowanego przez siebie podejścia, to procesy te powinny być, przynajmniej w pewnym zakresie, opisane.

Zaproponowana przez Doktorantkę metoda, mająca postać potoku obliczeniowego, składa się ze znanych algorytmów, zastosowanych w odpowiedni sposób, umożliwiające osiągnięcie założonych przez Autorkę celów. Zatem można stwierdzić, że algorytmy te zostały dobrane i zastosowane w poprawny sposób. Otrzymane wyniki zostały porównane z wiedzą dotyczącą analizowanych procesów biologicznych zawartą w literaturze, potwierdzając skuteczność zastosowanego podejścia. Porównanie to zostało przedstawione w rozdziale piątym. Warto podkreślić (co było już wspomniane), że oprócz przetestowania zaproponowanej metody na procesie biologicznym związanym z rakiem jajnika, który był głównym przedmiotem zainteresowania Doktorantki od strony biologicznej, metoda ta została przetestowana także na dwóch dodatkowych procesach, tj. cyklu komórkowym oraz szlaku sygnałowym MAPK, we wszystkich przypadkach dając zadowalające wyniki. Choć wyniki te nie przynoszą istotnych nowych spostrzeżeń biologicznych, to jednak wskazują na potencjał opracowanej metody do dokonywania odkryć tego rodzaju za jej pomocą. Oznacza to, że metoda ta mogłaby znaleźć zastosowanie w badaniach różnych procesów i systemów biologicznych, przede wszystkim szlaków sygnałowych i metabolicznych.

Oprócz wspomnianych powyżej zastrzeżeń w rozprawie można dostrzec pewną liczbę błędów, czy usterek, przy czym wiele z nich ma charakter redakcyjny. Wśród nich są m. in.:

- w rozprawie znajduje się wiele odwołań do kodów programów zamieszczonych w dodatku, które prawdopodobnie miały wyjaśniać, jak konkretnie przebiegają poszczególne fazy modelowania i analizy, jednak to powinno być wprost wyjaśnione w zasadniczej części rozprawy, gdyż nie wszystkie aspekty, zarówno procesu tworzenia modeli, jak i ich analizy, są oczywiste i nie powinno się oczekiwać od czytelnika, że część istotnych informacji wydedukuje samodzielnie na podstawie zamieszczonych kodów programów;
- na str. 16 na końcu drugiego akapitu znajduje się niedokończone zdanie „Using iterative modeling”;
- w tekście rozprawy brakuje bezpośrednich odwołań do niektórych rysunków i tabel (poprzez ich numer), np. do tabel 3.1, 4.7, 4.10, 4.12, 4.13, 4.14, 4.20, 4.21 oraz rysunków 1 (rysunek ten powinien mieć numer 3.1), 4.7, 4.8, 4.12, 4.31, 4.32;
- podpis dotyczący rysunku 4.1 powinien znajdować się pod rysunkiem, a nie nad nim;
- pod rysunkiem 4.4 znajdują się trzy podpisy, a sam rysunek ma trzy numery, tj. 4.2, 4.3 oraz 4.4, mimo iż jest to jeden rysunek;
- na str. 65 znajdują się błędne odwołania do rysunków 1 i 2 – w rozdziale czwartym nie ma rysunków o takich numerach – odwołania te powinny dotyczyć odpowiednich części rysunku 4.4;
- w ostatniej linii na str. 71 znajduje się tekst „Listing 4.1: Network Filtering Implementation” – nie wiadomo, do czego się on odnosi;
- rysunek 4.7 jest zupełnie nieczytelny i bardzo trudno jest go zinterpretować;
- na str. 72 znajduje się wzór (4.1), który nie został w żaden sposób skomentowany i nie wiadomo, co oznaczają występujące w nim symbole oraz jak należy go interpretować;
- podrozdziały 4.2.3 oraz 4.2.4 mają ten sam tytuł, przy czym drugi z nich zawiera tylko jedno zdanie kończące się dwukropkiem, po którym rozpoczyna się kolejny podrozdział;
- podrozdziały 4.4.3 oraz 4.4.5 mają ten sam tytuł;
- podrozdział 4.4.4 jest pusty;
- brakuje omówienia rysunku 4.12 – czytelnik samodzielnie musi się domyślać, dlaczego znalazł się on w rozprawie i co jest na nim przedstawione;

- brakuje omówienia sieci pokazanych na rysunkach 4.13 i 4.14 – czym różnią się przedstawione na nich reprezentacje sieci, jakich istotnych informacji dostarczają i jak one mają się do sieci pokazanych na rysunkach 4.10 i 4.11;
- z rysunkiem 4.19 związanych jest pięć podpisów i pięć numerów rysunków (4.15 – 4.19) – jeżeli jest to jeden rysunek, to powinien on mieć jeden numer i powinien znajdować się jeden podpis wyjaśniający zawartość poszczególnych jego części, jeżeli natomiast są to cztery odrębne rysunki, to nie powinny one mieć wspólnego numeru ani wspólnego podpisu;
- na wykresach zamieszczonych na rysunkach 4.19, 4.23 i 4.24 powinno być wyjaśnione, co oznaczają poszczególne krzywe (na każdym z nich znajdują się dwie krzywe);
- na rysunkach 4.25 – 4.29 znajdują się wizualizacje analizowanej sieci otrzymane za pomocą różnych metod, ale brakuje wyjaśnienia, czym te różne wizualizacje się różnią, jakie są ich cechy charakterystyczne i w jakim zakresie okazały się one przydatne w toku prowadzonych analiz;
- rysunki 4.26 – 4.29 są nieczytelne;
- nie wiadomo co oznaczają żółte okręgi na rysunku 4.29;
- w tabeli 4.18 znajdują się informacje tylko o trzech wierzchołkach, mimo iż na str. 90 jest napisane, że powinny tam być informacje o wszystkich wierzchołkach;
- podrozdział 4.11 „Rndom Walk Analysis” powinien być zawarty w podrozdziale 4.10 „Results of Random Walk and Network Analysis”;
- w tabeli 4.20 kolejność wierzchołków Wip1, PIP3 i ATM jest inna niż na rysunku 4.31 (dlaczego?);
- nie jest jasne, czym różnią się informacje przedstawione na wykresach zamieszczonych na rysunkach 4.33 i 4.34;
- nie zostały podane wartości współczynnika C (równanie (4.3)) dla grup wierzchołków opisanych w tabeli 4.22 – mogłyby być one zamieszczone w tej tabeli;
- nie jest jasne, co zostało przedstawione na rysunku 4.35;
- zamiast TNFa, IKK α , IKK β powinno być TNF α oraz IKK α (w kilku miejscach);
- na rysunku 4.46 przedstawione są wartości znormalizowane liczby odwiedzin wierzchołków, ale są one z przedziału od 0 do 12, w związku z czym nasuwa się pytanie, w jaki sposób była przeprowadzona normalizacja?;
- podrozdział 5.0.1 powinien mieć numer 5.0;
- Dodatek A jest zatytułowany „Supplementary Tables”, jednak zawarte są w nim przede wszystkim kody programów, więc tytuł nie odzwierciedla jego zawartości;
- w rozprawie brakuje odniesienia do/porównania z innymi metodami obliczeniowymi służącymi do wykonywania podobnych analiz, których istnieje wiele. Porównanie takie pozwoliłoby lepiej ocenić i docenić znaczenie metody zaproponowanej przez Doktorantkę.

Powyższe uwagi nie wpływają istotnie na pozytywną ocenę rozprawy. Doktorantka zaproponowała w jej ramach metodę obliczeniową umożliwiającą przeprowadzanie konstrukcji modeli złożonych systemów biologicznych oraz ich analizę w sposób w dużym stopniu zautomatyzowany. Metoda ta może być wartościowym narzędziem służącym do prowadzenia badań różnego rodzaju złożonych systemów biologicznych, m. in. związanych z procesami chorobowymi zachodzącymi w organizmie ludzkim, może więc być potencjalnie jednym z narzędzi wykorzystywanych do projektowania nowych, bardziej skutecznych terapii różnych chorób, np. nowotworowych. Lektura treści rozprawy skłania do przekonania, że Doktorantka bardzo dobrze orientuje się obszarze metod wykorzystywanych do modelowania i analizy złożonych systemów biologicznych oraz że posiada umiejętność samodzielnego prowadzenia pracy badawczej.

3. Podsumowanie

Moim zdaniem mgr inż. Ruby Khan przedstawiła w swojej rozprawie doktorskiej interesujące wyniki dotyczące modelowania i analizy złożonych systemów biologicznych. Zaproponowany przez nią potok obliczeniowy stanowi ciekawe połączenie wielu znanych metod, umożliwiające prowadzenie analiz skomplikowanych procesów biologicznych w sposób w dużym stopniu zautomatyzowany. Metoda ta stanowi ważny wkład w obszarze biologii systemów i może zarówno stanowić cenne narzędzie wykorzystywane w badaniach systemów biologicznych, jak i być punktem wyjściowym do dalszych prac nad metodami wykorzystywanymi do takich badań. Doktorantka wykazała się zarówno wiedzą z zakresu inżynierii biomedycznej, jak i umiejętnościami samodzielnego prowadzenia pracy naukowej. Warto też zauważyć, że część wyników badań prowadzonych przez Doktorantkę, związanych z tematyką rozprawy, została opublikowana w dwóch czasopismach z listy JCR, tj. *Genes* oraz *Scientific Reports*, co wskazuje na docenienie wyników jej pracy badawczej przez środowisko naukowe.

III. Konkluzja

Rozprawa doktorska mgr inż. Ruby Khan zawiera oryginalne i interesujące wyniki naukowe dotyczące metod modelowania i analizy złożonych systemów biologicznych. Moim zdaniem przedstawione w rozprawie wyniki stanowią istotny wkład do dyscypliny inżynieria biomedyczna. Uważam, że wymagania stawiane rozprawom doktorskim przez przepisy określone w Ustawie z dnia 20 lipca 2018 r. - Prawo o szkolnictwie wyższym i nauce, Dz. U. z 2024 r. Poz. 1571, z późn. zm., zostały spełnione. Wnoszę zatem o dopuszczenie mgr inż. Ruby Khan do dalszych etapów postępowania w sprawie nadania stopnia naukowego doktora.