

RECENZJA

rozprawy doktorskiej mgr. Tomasza Jetki pt.
„Quantitative methods to describe information flow in biochemical signalling pathways”

Promotorzy: prof. dr hab. Janusz Szczepański
dr hab. Michał Komorowski

I. Problematyka naukowa oraz przedmiot rozprawy

Dynamiczny postęp mający od wielu lat miejsce na gruncie nauk biologicznych zaowocował wieloma fascynującymi odkryciami, ale spowodował również, iż stało się jasne, że do jego kontynuowania niezbędne jest stosowanie w badaniach nad zjawiskami zachodzącymi w organizmach żywych metod matematycznych oraz informatycznych. Było to przyczyną powstania i szybkiego rozwoju w ostatnich 20 - 30 latach na pograniczu biologii, informatyki i matematyki dziedzin takich jak biologia obliczeniowa czy biologia systemowa, na gruncie których odpowiednie metody matematyczne oraz informatyczne nie tylko są stosowane do analizy zjawisk biologicznych, ale również (a może przede wszystkim) metody te są rozwijane.

W szczególności, w ciągu ostatnich dwudziestu lat coraz bardziej stawało się jasne, że organizmy żywe, a także ich funkcjonalne bloki, takie jak narządy, tkanki, komórki itp. składają się z pewnych elementarnych składników, które połączone są gęstą siecią wzajemnych oddziaływań. Okazuje się, że struktura tej sieci ma istotny wpływ (być może dominujący) na funkcjonowanie organizmów. Wynika stąd potrzeba odmiennego, niż dominujące do niedawna, postrzegania i badania organizmów żywych, tzn. postrzegania ich jako złożone systemy. Pociąga to za sobą oczywiście konieczność stosowania w badaniach układów biologicznych metod właściwych do analizy systemów. Do pewnego stopnia można wykorzystać metody opracowane w ramach nauk systemowych do badania systemów innego rodzaju (np. technicznych), jednak systemy biologiczne mają swoją specyfikę, więc takie podejście ma pewne ograniczenia i często konieczne jest opracowywanie metod lepiej dostosowanych do analizy systemów biologicznych.

Metody systemowe do badania układów biologicznych stosowane są na gruncie wspomnianej wcześniej biologii systemowej. Ta młoda dziedzina rozwija się w ostatnich latach bardzo szybko, m. in. ze względu na coraz powszechniejsze przekonanie, że podejście systemowe jest niezbędne do dogłębnego poznania natury świata ożywionego.

Jedną z podstawowych własności systemów jest przepływ informacji między ich składnikami. Przepływ ten jest kluczowy dla funkcjonowania systemu. Dotyczy to oczywiście nie tylko systemów biologicznych. Zagadnienia zapisywania i przesyłania informacji badane są na gruncie teorii informacji, której podwaliny w połowie XX w. położył Claude E. Shannon. Od tamtego czasu koncepcje teorii informacji są szeroko stosowane m. in. w wielu obszarach nauk technicznych do analizy systemów różnego rodzaju (zwłaszcza teleinformatycznych), jednak w dużo mniejszym stopniu były one dotąd stosowane w badaniach systemów biologicznych. Nie oznacza to, że teoria informacji jest mniej przydatna na gruncie biologii systemowej niż innych

nauk systemowych. Przyczyną jest raczej wspomniana wcześniej specyfika systemów biologicznych i związana z nią pewna trudność zastosowania znanych koncepcji teorii informacji do badania procesów przesyłania informacji w organizmach żywych. Istnieje w tym obszarze wiele wyzwań naukowych, których podjęcie mogłoby się przyczynić do istotnego postępu na drodze do dogłębnego zrozumienia mechanizmów rządzących światem ożywionym. Część z tych wyzwań zdecydował się podjąć w ramach swojej rozprawy doktorskiej mgr Tomasz Jetka. Rozprawa ta dotyczy rozważań związanych z potraktowaniem szlaków sygnałowych jako kanałów, przez które przekazywane są w komórkach informacje, a nawet szerzej - jako systemów, w których informacja nie tylko jest przesyłana, ale również przetwarzana.

Rozprawa doktorska mgr. Tomasza Jetki oparta jest na cyklu czterech publikacji. Składają się na nią następujące artykuły:

- [1] T. Jetka, K. Nieniałowski, S. Filippi, M. Stumpf, M. Komorowski. An information-theoretic framework for deciphering pleiotropic and noisy biochemical signalling. *Nature Communications*, 9(1), 4591, 2018.
- [2] T. Jetka, T. Winarski, K. Nieniałowski, S. Błoński, M. Komorowski. Information-theoretic analysis of multivariate single-cell signalling responses using SLEMI. *PLOS Computational Biology*, 15(7), e1007132, 2019.
- [3] U. Billing, T. Jetka, L. Nortmann, N. Wundrack, M. Komorowski, S. Waldherr, F. Schaper, A. Dittrich. Robustness and information transfer within IL-6-induced JAK/STAT signalling. *Communications Biology*, 2(1), 27, 2019.
- [4] T. Jetka, A. Charzyńska, A. Gambin, M. Stumpf, M. Komorowski. StochDecomp - Matlab package for noise decomposition in stochastic biochemical systems. *Bioinformatics*, 30(1), 137-138, 2014.

II. Analiza treści rozprawy oraz uzyskanych wyników

1. Treść rozprawy

Wspomniane powyżej artykuły w rozprawie doktorskiej mgr. Tomasza Jetki uzupełnione są o streszczenie, obszernie wprowadzenie, w którym Autor wyjaśnia podstawowe pojęcia związane z zagadnieniami rozważanymi w rozprawie, sformułowanie celów rozprawy, opis zastosowanej metodologii, przegląd uzyskanych wyników, wnioski, a także spis literatury obejmujący ponad 300 prac.

Sformułowany przez Autora cel rozprawy obejmuje opracowanie nowych metod analizy procesów przetwarzania informacji w sieciach sygnałowych uwzględniających stochastyczną naturę odpowiedzi komórkowych i umożliwiających analizę złożonych systemów wielowymiarowych. Ten ogólny cel został doprecyzowany przez sformułowanie trzech celów szczegółowych, tj.

- A) opracowanie efektywnej metodologii oszacowania pojemności kanału informacyjnego o ciągłym wejściu i ciągłym wyjściu, które mogłyby być wykorzystane w badaniach symulacyjnych ścieżek sygnałowych;
- B) opracowanie efektywnych metod oszacowania informacji wzajemnej oraz pojemności kanału informacyjnego o dyskretnym wejściu i ciągłym wyjściu umożliwiających analizę przepływu informacji przez ścieżki sygnałowe na podstawie danych pozyskanych z eksperymentów;
- C) zbadanie stochastycznej natury ścieżek sygnałowych za pomocą odpowiednich metod opartych na teorii informacji.

Rozdział pierwszy stanowi wstęp do rozprawy. Doktorant zwraca w nim uwagę na fakt, że stochastyczna natura ścieżek sygnałowych jest jedną z ich zasadniczych cech i wskazuje na jej źródła. Opisuje on też w tym rozdziale stosowane obecnie metody analizy uwzględniające tę stochastyczną naturę procesów komórkowych. W dalszej części tego rozdziału opisane są podstawy teorii informacji, których znajomość jest niezbędna do zrozumienia wyników zamieszczonych w publikacjach składających się na rozprawę doktorską. Autor zaczyna od podstawowych koncepcji, takich jak kanał komunikacyjny, entropia, informacja wzajemna oraz bit, by przejść do rozważań na temat matematycznego opisu przesyłania informacji przez kanał komunikacyjny, zarówno w przypadku, gdy zmienne losowe opisujące wejście oraz wyjście kanału są dyskretne, jak i ciągłe. W dalszej części wstępu przedstawiony został krótki przegląd wybranych zastosowań teorii informacji w biologii molekularnej oraz obecnie stosowanych w tej dziedzinie metod.

W rozdziale drugim sformułowane zostały cele rozprawy, o których mowa była powyżej.

Z kolei w rozdziale trzecim Doktorant przedstawił stosowane przez niego metody analizy dynamiki sygnałów pojawiających się w ścieżkach sygnałowych, krótko opisał wybrane zagadnienia związane z problemem klasyfikacji oraz procedury przetwarzania danych uzyskanych za pomocą mikroskopu konfokalnego.

Rozdział czwarty zawiera przegląd wyników zamieszczonych w publikacjach będących podstawą rozprawy. Jest on podzielony na trzy podrozdziały, zgodnie ze sformułowanymi wcześniej celami szczegółowymi.

Rozdział piąty zawiera podsumowanie oraz zarysowanie perspektyw związanych z jej tematyką i uzyskanymi wynikami.

Artykuł [1] związany jest z celem szczegółowym A rozprawy, tj. dotyczy on postrzegania ścieżek sygnałowych jako kanałów komunikacyjnych z ciągłym zarówno wejściem, jak i wyjściem, a także z celem C. Zaproponowana w nim została metoda analizy modeli z wieloma wejściami i wieloma wyjściami. Polega ona na zastosowaniu podejścia opartego na macierzy informacyjnej Fishera oraz rozkładu apriorycznego Jeffreysa. Metoda ta została zastosowana do badania procesu przesyłania przez ścieżkę sygnałową JAK/STAT informacji dotyczących cząsteczek INF- α oraz INF- λ .

Wyniki przedstawione w artykule [2] dotyczą głównie celu szczegółowego B rozprawy, czyli określenia informacji wzajemnej i pojemności informacyjnej ścieżki sygnałowej postrzeganej jako kanał komunikacyjny w przypadku, gdy wejście tego kanału ma charakter dyskretny, a wyjście ciągłe, a także celu C. Badania te dotyczyły analizy statystycznej danych eksperymentalnych. W ich ramach opracowany został przez Doktoranta pakiet SLEMI napisany w języku R, który umożliwia również osobom bez odpowiedniego przygotowania matematycznego zastosowanie zaproponowanej metodologii. Pozwala ona na analizę szlaków sygnałowych z wielowymiarowymi wyjściami oraz dużą liczbą wejść, na co na ogół nie pozwalają w praktyce metody znane z literatury. Zaproponowana metoda oparta jest na obliczaniu, inaczej niż inne znane metody, wyjściowego rozkładu warunkowego $P(X|Y=y)$ (gdzie X reprezentuje wejście, a Y wyjście kanału komunikacyjnego). Metoda ta została zastosowana do analizy szlaku sygnałowego TNF- α -NF- κ B. Jej dokładność została zbadana przez porównanie z metodą KNN. Wyniki przeprowadzonego eksperymentu obliczeniowego wskazują, że zaproponowane przez Doktoranta podejście daje lepsze oszacowania pojemności informacyjnej.

W artykule [3], który dotyczy celu szczegółowego C rozprawy mgr. Jetki, opracowane przez niego metody wyznaczania pojemności kanału komunikacyjnego (SLEMI) zastosowane zostały do badania przepływu sygnałów w ścieżce JAK/STAT indukowanego interleukiną 6.

Z kolei w artykule [4] opisany został pakiet do Matlaba o nazwie StochDecomp, służący do identyfikacji szumu w sieciach sygnałowych. Artykuł ten związany jest z celem szczegółowym C.

2. Uwagi merytoryczne i redakcyjne

Rozprawa doktorska mgr Tomasza Jetki napisana jest bardzo starannie, z dużą dbałością zarówno o jej stronę językową, jak i matematyczną. Główne idee, na których oparte są badania Doktoranta, przedstawione zostały w jasny sposób we wstępie, co powinno ułatwić ich zrozumienie również osobom, dla których teoria informacji nie jest podstawowym obszarem działalności naukowej. Warto zwrócić uwagę na dużą liczbę publikacji, na które w swojej rozprawie powołuje się Doktorant (302 pozycje w spisie literatury) oraz na fakt, że w toku prac nad rozprawą, oprócz uzyskanych wyników teoretycznych oraz przeprowadzonych eksperymentów, powstały trzy pakiety oprogramowania (StochDekomp, StochSens i SLEMI), umożliwiające korzystanie z zaproponowanych metod.

Autor w ramach rozprawy podjął ważny i w niewielkim stopniu zbadany dotąd problem analizy procesów przesyłania informacji w biochemicznych ścieżkach sygnałowych, a w szczególności określenia pojemności informacyjnej biochemicznych kanałów komunikacyjnych. Dokładne zrozumienie tych procesów wydaje się być kluczowe dla możliwości dogłębnego poznania mechanizmów, na których oparte jest funkcjonowanie organizmów żywych. Zastosowanie metod teorii informacji do badania tych mechanizmów to dziedzina, w której jest bardzo wiele otwartych pól badawczych. Zaproponowane w rozprawie doktorskiej mgr. Tomasza Jetki metody i narzędzia szacowania pojemności informacyjnej ścieżek sygnałowych stanowią cenny wkład w tę dziedzinę umożliwiającą innym naukowcom prowadzenie badań za pomocą metod bardziej efektywnych niż te, z których mogli korzystać do tej pory. Przedstawione przez Doktoranta wyniki są też naturalnym punktem wyjściowym do prowadzenia dalszych poszukiwań naukowych. W tym kontekście interesująca byłaby odpowiedź na pytanie, jakie problemy pozostające do rozwiązania w toku ewentualnej kontynuacji swoich badań doktorant uważa za najważniejsze? Interesujące jest również, co można powiedzieć o złożoności obliczeniowej metod opisanych w artykułach [2] i [4]? W konkluzjach Doktorant napisał, że zaproponowany przez niego algorytm mógłby być połączony z różnymi metodami uczenia maszynowego. Czego można by się spodziewać po takim połączeniu i dlaczego?

3. Podsumowanie

Moim zdaniem mgr Tomasz Jetka przedstawił w swojej rozprawie doktorskiej interesujące wyniki naukowe dotyczące zastosowań teorii informacji w badaniach procesów komórkowych. Rozważane przez Doktoranta problemy są bardzo ważne dla biologii systemowej i możliwości dalszego istotnego postępu na drodze do dogłębnego poznania mechanizmów rządzących światem ożywionym. Zastosowane przez niego metody badawcze są właściwe i odwołuje się on do odpowiedniej literatury. Warto też zauważyć, że wyniki składające się na rozprawę mgr Jetki zostały opublikowane w czterech artykułach, z czego trzy ukazały się w bardzo dobrych czasopismach z listy JCR, tj. *Nature Communications*, *PLOS Computational Biology* oraz *Bioinformatics*. Świadczy to o tym, że działalność badawcza Doktoranta została już doceniona przez środowisko naukowe.

III. Konkluzja

Rozprawa doktorska mgr. Tomasza Jetki zawiera oryginalne i interesujące wyniki naukowe dotyczące zastosowania teorii informacji do badania procesów komórkowych. Uważam, że wymagania stawiane rozprawom doktorskim przez obowiązujące przepisy zostały spełnione. Wnoszę zatem o dopuszczenie mgr. Tomasza Jetki do dalszych etapów postępowania o nadanie stopnia naukowego doktora. Ponadto, ze względu na wagę uzyskanych wyników oraz fakt opublikowania ich w trzech bardzo dobrych czasopismach z listy JCR, wnioskuję o wyróżnienie rozprawy.

