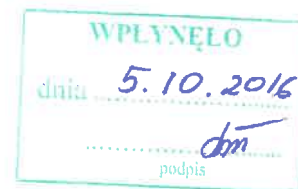


Prof. dr hab. Inż. Maciej Ogorzałek
Zakład Technologii Informatycznych
Wydział Fizyki, Astronomii i Informatyki Stosowanej
Uniwersytetu Jagiellońskiego
Ul. Prof. S. Łojasiewicza 11
30-348 Kraków



RECENZJA
Roprawy doktorskiej
mgra inż. Krzysztofa Sarapaty
pt. "Model biocybernetyczny w analizie sekwencji genomu
człowieka"

Niniejsza recenzja została wykonana na zlecenie Dziekana Wydziału Elektrotechniki, Automatyki, Informatyki i Inżynierii Biomedycznej Akademii Górniczo-Hutniczej im. Stanisława Staszica w Krakowie dra hab. Antoniego Cieśli (pismo z dnia 8 lipca 2016 L.Dz. WEAlIB-b/510-4.2/13).

1. Tematyka rozprawy

Przedstawiona do recenzji praca dotyczy niezwykle istotnej i burzliwie rozwijającej się dziedziny badań interdyscyplinarnych – zastosowań zaawansowanych metod matematycznych oraz narzędzi obliczeniowych do analizy struktur i mechanizmów rządzących najbardziej elementarnymi funkcjami komórek na poziomie biochemicznym. Metody statystyczne stosowane są w biologii i medycynie od dość dawna natomiast modelowanie bio-cybernetyczne oraz zastosowanie metod przetwarzania sygnałów, analizy danych czy uczenia maszynowego znalazły się w centrum zainteresowania dopiero w ostatnich latach. Podjęcie tematu jest ze wszechmiar uzasadnione wielkim zainteresowaniem badaniami tego typu wśród specjalistów a także w szerszym aspekcie wśród szerszego kręgu badaczy różnych aspektów budowy i funkcjonowania organizmów żywych.

2. Teza pracy

Na podstawie analizy literatury przedmiotu Autor sformułował ciekawe z punktu widzenia naukowego zagadnienie badawcze, które zawarł w następującej tezie pracy:

Możliwe jest rozszerzenie modelu TargetScore do postaci modelu biocybernetycznego biTargetScore, zbudowanie takiego modelu oraz jego walidacja na podstawie danych eksperymentalnych dotyczących sekwencji genomu człowieka uzyskanych z materiału klinicznego

W pracy po nakreśleniu tła rozważań w oparciu o źródła literaturowe Autor w sposób przekonujący udowodnił tę tezę budując oryginalny model oraz jego implementację a w końcu przeprowadzając w oparciu o bazę danych eksperymentalnych walidację tego modelu.

W mojej opinii organiczę się do oceny części bioinformatycznej pracy – oceny aspektów biologii molekularnej oraz wartości opracowania dla tej dziedziny nie mogę się podjąć nie będąc specjalistą.

1. Struktura pracy

Na początku po typowym wstępie do rozprawy w rozdziale 2 Autor przedstawił "Biologiczne podstawy regulacji genów" – wprowadzając czytelnika do tematyki oraz opisując szczegóły z punktu widzenia biologii molekularnej. Opisano jedną z eksperymentalnych technik wysokoprzepustowych oraz wskazano nowe kierunki poszukiwań. W rozdziale 3 zatytułowanym "Modele w analizie interakcji miRNA/mRNA" zawarty jest opis oryginalnych rozwiązań zaproponowanych przez Autora, poprawiających jakość predykcji procesu regulacji. Koncepcja nowego rozwiązania opisana jest szczegółowo w rozdziale 4 pt. "Definiowanie modelu biocybernetycznego", natomiast rozdział 5 zawiera "Opis implementacji" – opis narzędzi informatycznych wykorzystanych w analizie i przeprowadzonym wnioskowaniu. Rozdział 6 "Walidacja opracowanego modelu" opisuje testowanie wprowadzonego rozwiązania i porównanie go z innymi narzędziami. Rozdział 7 "Podsumowanie" zawiera wnioski z przeprowadzonych badań oraz przedstawia perspektywy rozwoju i kontynuacji badań.

Praca zawiera także obszerny spis literatury (188 pozycji!), świadczący o doskonałej orientacji Autora w dziedzinie prowadzonych badań. Autor przedstawił w pracy bardzo dobrą analizę tych źródeł, przedstawił aktualny stan wiedzy a wnioski wynikające z tej analizy dotyczące rozwiązania postawionego problemu są sformułowane w sposób jasny i przekonujący.

Na końcu pracy zamieszczono także bardzo pomocne dodatki A-E pomagające potencjalnemu czytelnikowi nie mającemu głębszego przygotowania w zaawansowanych zagadnieniach rachunku prawdopodobieństwa, w zrozumieniu całości tekstu i interpretacji przeprowadzonych badań.

Praca zawiera bardzo dobrze przygotowany zestaw tabl, rysunków i ilustracji na każdym etapie prowadzonych rozważań, jest bardzo dobrze opracowana graficznie. Język jest kwiecisty ale jasny i zrozumiały.

2. Oryginalne osiągnięcia autora rozprawy

Celem pracy jest zarówno teoretyczne opracowanie nowego, poprawionego modelu biocybernetycznego jak i praktyczna realizacja narzędzia bioinformatycznego. Model ma na celu identyfikację (rozpoznawanie) dupleksów: transkryptów oraz miRNA - cząsteczek nukleotydowych w konkretnym biologicznym procesie regulacji genów na podstawie danych eksperymentalnych pochodzących z wysokoprzepustowych technik eksperymentalnych. Do budowy modelu przyjęto sensowne/właściwe założenia oraz przyjęto klasyczną metodologię spośród zaawansowanych metod statystycznych.

Najważniejsze samodzielne i oryginalne osiągnięcie Doktoranta to zbudowanie ulepszanego modelu biocybernetycznego nazwanego **biTargetScore**, dla którego zrealizowano implementację, a następnie przeprowadzono jego walidację.

W zaproponowanym modelu typowania par miRNA/mRNA wykorzystano bibliotekę TargetScore implementującą probabilistyczny model Bayesa połączony z wnioskowaniem wariacyjnym występujący pod nazwą *Variational Bayesian-Gaussian Mixture Model* (VB-GMM).

Autor do analizy danych eksperymentalnych zastosował dwa podejścia: sekwencyjne i biocybernetyczne. Podejście sekwencyjne polega na przeprowadzeniu a priori zdefiniowanej sekwencji obliczeniowej. Zakłada ona, że rozpoznanie dupleksów można dokonać jedynie na podstawie informacji uzyskanej z wcześniejszego etapu, startując od informacji o sekwencji nukleotydowej. Drugie podejście bazuje na analizie "sygnału wyjściowego" (przyjętego w pracy jako poziom ekspresji regulowanych genów) dla rozbudowy istniejącego modelu rozpoznawania dupleksów.

Praca pokazała także możliwości adaptacji narzędzi i metod informatyczno-matematycznych w szczególności podejścia *Variational Bayesian-Gaussian Mixture Model* (VB-GMM), wykorzystanego w budowie modelu. Zaproponowany model uwzględnia profil ekspresji transkryptów oraz miRNAs a z punktu widzenia informatycznego integruje dane kontekstowe i filogenetyczne

Model **biTargetScore** został porównany z innymi narzędziami predykcji, wykazując nad nimi przewagę, wyrażającą się przez poprawę jakości rozpoznania targetów. Do weryfikacji poprawności budowanych modeli wykorzystano oryginalne dane eksperymentalne, zgromadzone w Zakładzie Biologii Molekularnej i GenetykiKlinicznej CMUJ w ramach eksperymentu dotyczącego oszacowania wpływu eksperymentalnego zakażenia rinowirusem HRV16 komórek pierwotnych linii nabłonka oddechowego pobranych od osób z rozpoznaną astmą oraz od osób zdrowych. Baza danych wydaje się być dostatecznie duża by można było formułować sensowne wnioski z przeprowadzonych badań. Przedstawione wyniki potwierdzają, że stworzony przez Doktoranta model daje prawidłowe wyniki w sensie statystycznym, uzyskując lepsze parametry czułości i swoistości.

Autor podjął też próbę wyjaśnienia przyczyn i wskazania trudności przetwarzania danych surowych, pochodzących z eksperymentów mikromacierzowych, które charakteryzuje stochastyczność. Przedstawił także analizę stosowanych uproszczeń i uogólnień.

Realizacja celu pracy stanowi istotny przyczynek do poznania elementarnych mechanizmów natury na poziomie molekularnym, może w przyszłości przyczynić się np. do powstania nowych terapii różnych chorób.

Połączenie metod i podejść stosowanych w kilku różnych dziedzinach pozwoliło na poprawę wyników w stosunku do bieżącego stanu wiedzy w dziedzinie. Transkrypty wyszukane w oparciu o model **biTargetScore** nie były wykrywane przy użyciu znanych pakietów oprogramowania ani korelacji statystycznych opartych na analizie składników głównych.

Uwagi dyskusyjne

W czasie czytania tak obszernej i wielowątkowej rozprawy nasuwa się szereg pytań, które mogą stanowić element dyskusji w trakcie obrony pracy a nie decydują o jednoznacznie pozytywnej ocenie samej pracy. Jednym z takich pytań jest szersze

zastosowanie opracowanego modelu, jego przydatność do badań dla innych baz danych – czy prowadzone były inne testy? Czy wyniki są równie dobre?
Metodyka i procedury TargetScore doczekała się szerokiej dostępności (patrz <https://www.bioconductor.org/packages/release/bioc/html/TargetScore.html>) - czy przewidywane jest szersze udostępnienie metod/oprogramowania?

Jako profesora nauk technicznych interesuje mnie także wspomniana w zakończeniu możliwość dalszych rozszerzeń modelu i zastosowań proponowanego podejścia także w naukach technicznych.

Podsumowanie

Oceniając wartość naukową w części dotyczącej opracowanego modelu biocybernetycznego, jego implemtacji i walidacji stwierdzam, że przedstawiona przez mgra Krzysztofa Sarapatę praca stanowi istotny wkład w dyscyplinie naukowej biocybernetyka i inżynieria biomedyczna i może być podstawą do nadania stopnia doktora w tej dyscyplinie.

Przedstawiona praca spełnia jednoznacznie wymagania stawiane rozprawom doktorskim przez odnośne przepisy. Wnoszę o dopuszczenie kandydata do dalszych etapów przewodu doktorskiego.



Kraków, dnia 12 września 2016