

RAU	Biuro Dziekana	
	Wpłynęło dnia	23.08.19
	Nr	461 / zał.

Dr hab. January Weiner
Berlin Institute of Health
Anna-Louisa-Karsch-Str. 2
10178 Berlin

Szanowny Pan Prof. dr hab. inż. Adam Czornik, Dziekan
Wydział Automatyki, Elektroniki i Informatyki

Recenzja rozprawy doktorskiej Anny Papież

Tytuł rozprawy: Integrative data analysis methods in multi-omics molecular biology studies for disease of affluence biomarker research.

Autor rozprawy: Anna Papież

Promotor rozprawy: Prof. Joanna Polańska

Streszczenie: Rozprawa doktorska pani Anny Papież porusza niezwykle ważne zagadnienie z dziedziny biologii systemów: integracji wysokowymiarowych danych (*high throughput data*) pochodzących z różnych źródeł oraz kohort. Ogólnie mówiąc, dane wysokowymiarowe – takie jak dane transkryptomiczne albo metaboliczne – pozyskiwane są obecnie stosunkowo niewielkim kosztem, czego wynikiem jest olbrzymia liczba produkowanych zestawów danych, wykładniczo rosnąca z każdym rokiem. Jednak zadanie polegające na *zestawieniu* danych, które zostały pozyskane w różnych badaniach, z wykorzystaniem różnych technik analitycznych nie jest trywialne.

Zagadnienie to nadal wymaga wyjątkowej dbałości o szczegóły i nie ma ogólnego rozwiązania, jednak praca (a właściwie: prace) p. Papież stanowią niebalny przyczynek w tej trudnej dziedzinie, proponując nowe podejścia i algorytmy.

Przedstawienie pracy: P. Papież zebrała w swojej rozprawie szereg badań z dziedziny integracji wysokowymiarowych zestawów danych. Po pierwsze, opracowała i zaimplementowała algorytm do identyfikacji systematycznego efektu specyficznego dla zestawu danych (*batch effect*). Algorytm opiera się na programowaniu dynamicznym i umożliwia korektę danych nawet w sytuacji, gdy nie znana jest struktura danych (tzn. umożliwia rozdzielanie próbek na grupy obciążone specyficznym efektem). Jest to szczególnie ciekawa część pracy, ponieważ, jak pokazała autorka, brak odpowiedniej korekcji błędów może w efekcie doprowadzić do braku wyników analizy funkcjonalnej (rozdz. 4.1.2).

Metody integracji wieloplatformowej zostały wykorzystane do analizy szeregu oryginalnych zestawów danych z dziedziny radiologii i onkologii. Analizy doczekały się publikacji w recenzowanych pracach naukowych. Na przykład, korzystając z integracji danych z różnych platform dokonano analizę biomarkerów służących

do dyskryminacji między pacjentami wrażliwymi na promieniowanie jonizujące (w terapii) a niewrażliwymi (rozdz. 4.2). W pracy tej wykorzystano nowe podejście do identyfikacji grup genów wykazujących różny rodzaj trendów w zależności od otrzymanej dawki promieniowania. W tym celu zastosowano test Jonckheere-Tepstry – znakomity pomysł, który przyniósł też dobre wyniki.

W dalszej części pracy autorka dokonała analizy i integracji danych uzyskanych metodami transkryptomiki oraz proteomiki (rozdz. 4.3). W tym celu wykorzystywała integrację wartości p metodą Fishera, wzmacniając w ten sposób wyniki uzyskane tylko w oparciu o analizę transkryptomu. Na koniec (rozdz. 4.4) autorka dokonała integracji danych pochodzących z różnych tkanek czy typów komórek.

W pracy położono nacisk na systematyczną kontrolę stosowanych metod i analizę porównawczą różnych rozwiązań. Wyliczenie wszystkich zastosowanych i porównanych metod wzbudziło mój podziw: autorka wykonała olbrzymią pracę (która czasem jest w rozprawie zbyta jednym zdaniem!).

Opinia: Należy przede wszystkim podkreślić, że rozprawa doktorska przedstawia badania, które zostały opublikowane już recenzowanych czasopismach naukowych. Liczba i jakość publikacji naukowych wyrasta znacząco powyżej europejskiej średniej: prace p. Papiież ukazały się w czterech czasopismach, w tym w *Bioinformatics*, jednym z najlepszych czasopism z dziedziny bioinformatyki.

W swojej pracy naukowej i rozprawie p. Papiież wykazała się bardzo szeroką znajomością statystyki i bioinformatyki, a także biologii; myślę, że właśnie umiejętność nie tylko programowania i rozumowania algorytmicznego, ale również rozumowania statystycznego i biologicznego (obce wielu adeptom bioinformatyki) zadecydowała o sukcesie prezentowanych metod. P. Papiież swobodnie posługuje się zarówno modelowaniem liniowym, jak i uczeniem maszynowym (np. z wykorzystaniem maszyn wektorów nośnych, *support vector machines*) czy funkcjonalną analizą z dziedziny biologii systemów. Rozwiązania zaprezentowane w pracy są oryginalne, pomysłowe i świeże; autorka poszukiwała (skutecznie) nowych dróg, a czasem wykorzystywała bardzo proste rozwiązania (na które jednak jednak inni nie wpadli) żeby uzyskać znakomite efekty.

Główny punkt mojej krytyki to niedostateczne uwypuklenie swojego dorobku naukowego w rozprawie. Uważam, że zbędne jest rozpisywanie się na tematy, które zostały już szczegółowo przedstawione w recenzowanych czasopismach, ale brak wystarczających odniesień do tych prac może sprawić wrażenie, że rozprawa doktorska jest trywialna. To nieprawda: streszcza ona olbrzymią pracę, jednak by to w pełni docenić, należy, niestety, sięgnąć po publikacje.

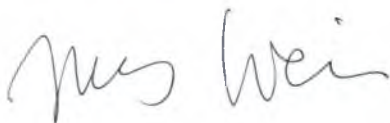
To, że p. Papiież ma na swoim koncie już szereg prac naukowych miało też inny pozornie negatywny efekt dla rozprawy: chociaż w pracy sformułowano centralne cele (rozdział 1.2 oraz początek rozdz. 5), spora jej część dotyczy narzędzi, które trzeba było najpierw wykonać. Myślę, że było to trudne do uniknięcia, zwłaszcza w pracy *par excellence* bioinformatycznej, co w pełni usprawiedliwia też brak wyraźnej biologicznej hipotezy. Efektem ubocznym jest też nieco zagmatwana

struktura rozdziału “Wyniki i Dyskusja”, gdzie czasem ciężko się zorientować, o której ze swoich prac autorka akurat pisze.

Dorobek p. Papież stanowi nietrywialny przyczynek do pracy z wysokowymi-
arowymi zestawami danych. Szczególnie BatchI, algorytm służący do iden-
tyfikacji efektu specyficznego dla zestawu danych (*batch effect*) zaimplementowany
w języku R wydaje się bardzo obiecującym nowym narzędziem, które sam
zamierzam wykorzystać w swojej pracy. Również w dziedzinie biomarkerów
wykazała nowatorskie podejście do analizy danych transkryptomowych i proteo-
mowych.

Uwagi techniczne dotyczące wykonania rozprawy: Praca jest wykonana
bardzo starannie. Ilustracje są czytelne i właściwie opisane; jedyne zastrzeżenie
to używanie ilustracji łączących czerwony i zielony kolor, co może być problemem
dla czytelników cierpiących na ślepotę barw. Podobnie, wzory matematyczne są
czytelne i dobrze opisane. Rozdział między wstępnym omówieniem problemu,
metodami oraz wynikami nie budzi zastrzeżeń. Dyskusja została połączona
z wynikami, co jest częstym zabiegiem w pracach, w których trzeba omówić
szereg uzyskanych wyników. W dziedzinach, które mogę ocenić (bioinformatyka,
transkryptomika, proteomika, biomarkery), nie pominięto żadnych istotnych
publikacji, p. Papież wykazała się dobrym przygotowaniem merytorycznym.
Język pracy (pomijając drobne lapsusy) jest zwięzły, poprawny i czytelny.

Rekomendacja: Dopuszczenie rozprawy doktorskiej do publicznej obrony i
jej wyróżnienie. **Uzasadnienie:** dorobek p. Anny Papież w pełni uzasadnia
przyznanie jej stopnia doktora; wykonana praca leży znacznie powyżej przeciętnej
europejskiej zarówno pod względem obszernego ujęcia tematu, jak i osiągniętych
publikacji naukowych.



January Weiner