

Hasselt, 28 marca 2023

**Recenzja rozprawy doktorskiej****Tytuł rozprawy:** „Algorithms and Computational Models in Chemical Analysis”**Autor:** Grzegorz Skoraczyński

Recenzowana rozprawa doktorska składa się z trzech części. Pierwszą stanowi rozdział zawierający krótkie wprowadzenie do problemów, których dotyczą metody przedstawione w rozprawie. Druga składa się z trzech rozdziałów (2-4) zawierających szczegółowy opis wyników dotyczących nowej metody rozwiązania problemu uliniowania czasu retencji (RT) w zautomatyzowanej analizie danych z zakresu chromatografii cieczowej ze spektrometrią mas (LC-MS). Wyniki te stanowiły podstawę publikacji: Grzegorz Skoraczyński, Anna Gambin, and Błażej Miasojedow, “Alignstein: Optimal Transport for Improved LC-MS Retention Time Alignment.” *GigaScience* 2022, 11, 1-9 (DOI: 10.1093/gigascience/giac101). Trzecia część rozprawy składa się z dwóch rozdziałów (5-6) i rozważa modele uczenia maszynowego konstruowane w celu predykcji syntetyzowalności cząsteczek. W szczególności zawiera empiryczne porównanie różnych modeli, włącznie z modelami zaproponowanymi przez doktoranta, przy użyciu specjalnie w tym celu stworzonej bazy danych związków o różnym stopniu syntetyzowalności. Wyniki tego porównania stanowią podstawę publikacji: Grzegorz Skoraczyński, Mateusz Kitlas, Błażej Miasojedow, and Anna Gambin, “Critical assessment of synthetic accessibility scores in computer-assisted synthesis planning.” *Journal of Cheminformatics* 2023, 15, 6 (DOI: 10.1186/s13321-023-00678-z). Warto zauważyć, że publikacje ukazały się w czasopismach naukowych ze wskaźnikiem cytowań (za 2021 rok) wyższym niż 7.5 (J of Cheminformatics 8.489, GigaScience 7.658).

Rozprawa dotyczy wyników badań interdyscyplinarnych. Interdyscyplinarność jest dużym wyzwaniem, wymaga bowiem łączenia wiedzy z różnych dziedzin. W przypadku ocenianej rozprawy, doktorant musiał połączyć wiedzę z zakresu matematyki, statystyki i informatyki ze zrozumieniem procesu generacji i specyfiki danych otrzymywanych przy pomocy LC-MS, jak również zagadnień syntetyzowalności cząsteczek. Dokonał tego w satysfakcjonujący sposób, który umożliwił zaproponowanie interesujących rozwiązań w dziedzinach nie będących jego podstawowym obszarem badań.

Badania interdyscyplinarne stanowią również duże wyzwanie z punktu widzenia komunikacji osiągniętych wyników. Celem komunikacji powinno być przedstawienie metod i rezultatów w taki sposób, by stały się zrozumiałe (lub przynajmniej intepretowalne) przez badaczy z różnych dziedzin. W mojej ocenie rozprawa, niestety, nie osiąga tego celu, co uznaję za jej główny mankament.

Z tego punktu widzenia za zmarnowaną szansę uznaję część pierwszą rozprawy. Ogranicza się ona do siedmiu stron. W rozszerzonej formie mogłaby natomiast zostać wykorzystana do szerszego przybliżenia technologii używanych w chromatografii i MS czytelnikom (np. matematykom, statystykom, informatykom) nie zajmującym się tą dziedziną, co ułatwiłoby im zrozumienie części drugiej rozprawy. Podobny komentarz dotyczy problemu predykcji syntetyzowalności cząstek. W istocie, jestem przykładem czytelnika rozprawy, dla którego syntetyzowalności cząstek była kompletnie nieznanym zagadnieniem, i który to stan nie uległ zasadniczej zmianie po lekturze rozdziału 1. Siłą rzeczy, w tej sytuacji próba zrozumienia materiału zawartego w części trzeciej rozprawy była dla mnie wyzwaniem.

Z drugiej strony, np. opis proponowanej metody uliniowania czasu retencji, zawarty w rozdziale 2, jest wysoce teoretyczny i, moim zdaniem, kompletnie niedostępny dla specjalistów z zakresu MS. Również dla osób rozumiejących aparat matematyczny wspomniany opis zawiera fragmenty, które budzą wątpliwości dotyczące dokonywanych wyborów, jak w poniższych przykładach:

Strona 22: “For our application, the total variation divergence achieved the best results” – najlepsze rezultaty z punktu widzenia jakiego kryterium?

Strona 27: “In the beginning, features are denoised by removing signals of the lowest intensity (about 1 % of signal) and normalized by its total ion current.” – dlaczego przyjęto taki poziom odcięcia dla mierzonej intensywności?

Strona 28: “centroid clustering is done in two steps: firstly, we split the whole space into several smaller pieces using the Mini-batch K-Means algorithm [116], then we do final precise clustering using hierarchical clustering [117].” – jakimi kryteriami kierowano się przy wyborze tych metod?

Dyskusja zalet i ograniczeń algorytmu Alignstein w rozdziale 4 pomija kwestie technologii, do których algorytm może mieć zastosowanie. W rozprawie skoncentrowano się na chromatografii cieczowej (LC), a walidację przeprowadzono na zbiorach danych uzyskanych, jak można się dowiedzieć ze wskazanych publikacji źródłowych, technikami wykorzystującymi jonizację za pomocą elektrorozpyłania (ESI). Czy algorytm może być również stosowany np. w przypadku połączenia LC-APCI? Czy możliwe jest jego użycie w przypadku chromatografii gazowej (GS)? Tego rodzaju dyskusja mogłaby ułatwić zrozumienie własności algorytmu badaczom zajmującym się MS.

Rozprawa została napisana po angielsku. Generalnie, językowo jest poprawna – w trakcie lektury zauważyłem pojedyncze błędy językowe i stylistyczne. Redakcja rozprawy jest zadowolająca. Usterki, które to w tej materii zwracają uwagę, to

- niepoprawne odniesienie do ryciny 3.2 jako 5.6 (str. 25),
- użycie terminu „binomial” zamiast „bimodal” w opisie ryciny 5.5,
- użycie terminu „randomized from” zamiast (prawdopodobnie) „randomly sampled from” na początku sekcji 5.2.2,
- brak konsekwencji w stosowaniu terminu „cheminformatics” („cheminformatics” str. 45).

Pomijając kwestię zbyt technicznego opisu, utrudniającego zrozumienie prezentowanych wyników specjalistom z dziedzin, w których wyniki te mogą mieć zastosowanie, zawartość merytoryczną ocenianej rozprawy uważam za interesującą i w wielu aspektach nowatorską. Pokazuje ona, że doktorant posiada szeroką wiedzę w zakresie modelowania matematycznego, statystycznego i informatyki. Potwierdza również, że doktorant zgłębił problematykę analizy danych LC-MS i syntetyzowalności cząstek. Zawartość rozprawy wskazuje, że doktorant potrafi zaproponować nowe, interesujące metody rozwiązania problemów w różnych dziedzinach oraz skutecznie je zaimplementować w postaci narzędzi informatycznych, które pozwalają na użycie metod w praktyce. Wobec powyższego stwierdzam, że recenzowana praca spełnia wymagania stawiane rozprawom doktorskim przez obowiązujące przepisy i wnoszę o dopuszczenie magistra Grzegorza Skoraczyńskiego do dalszych etapów przewodu doktorskiego.



Tomasz Burzykowski  
 Profesor Uniwersytetu Hasselt