

Katowice, 16 marca 2024 r.

Prof. dr hab. inż. Adam Smoliński
Główny Instytut Górnictwa - PIB
Plac Gwarków 1, 40-166 Katowice
asmolinski@gig.eu

Recenzja rozprawy doktorskiej

kpt. mgr inż. Bartłomieja Fliszkiewicza

**pt. „Właściwości kwantowe fragmentów cząsteczek jako
deskryptory molekularne”**

Podstawą opracowania niniejszej recenzji jest pismo Dziekana Wydziału Nowych Technologii i Chemii Wojskowej Akademii Technicznej w Warszawie, prof. dr hab. inż. Krzysztofa Czupryńskiego z dnia 19 lutego 2024 roku. Promotorem rozprawy jest dr hab. inż. Marcin Sajdak.

1. Tematyka rozprawy doktorskiej i trafność jej wyboru

Wykorzystanie metod komputerowych w chemii pozwala na zwiększenie efektywności przetwarzania danych, konstruowanie lepszych modeli, a tym samym uzyskiwanie dokładniejszych wyników. Innymi słowy metody komputerowe wykorzystują matematykę in silico, polegającą na przetwarzaniu przez maszynę liczącą dużej ilości prostych operacji. Biorąc pod uwagę kwestie wykorzystania obliczeń komputerowych w chemii można wyodrębnić trzy obszary takie jak chemia kwantowa, chemometria i chemoinformatyka. W

przypadku chemii kwantowej przedmiotem badań są atomy lub małe cząsteczki chemiczne, w przypadku chemometrii są to statystyczne i numeryczne metody eksploracji i modelowania danych, natomiast w przypadku chemoinformatyki są to badania dużych „systemów chemicznych”. Jak słusznie kpt. mgr inż. Fliszkiewicz zauważył zastosowanie metod chemoinformatycznych pozwala w łatwy i szybki sposób uzyskiwać wyniki. Doktorant ma świadomość, że takie podejście wiąże się z mniejszą dokładnością, jednak pozwala na uzyskiwanie rezultatów w miarę szybko, co jest bardzo pomocne np. w metodach przesiewowych i ostatecznie pozwala na skonstruowanie rozwiązania typu „czarna skrzynka”, które może być wykorzystane przez osoby nie posiadające specjalistycznej wiedzy w zakresie chemoinformatyki. Z kolei wspomniane obliczenia kwantowo-chemiczne, wymagające specjalistycznej, zaawansowanej wiedzy pozwalają na uzyskiwanie rezultatów charakteryzujących się wysoką dokładnością. Są to jednak obliczenia niezwykle czasochłonne i wymagające zastosowania dużych mocy obliczeniowych.

Tematyka pracy wpisuje się w cele i kierunki badań prowadzonych od wielu lat na świecie ukierunkowanych na poszukiwanie zależności między strukturą a właściwościami związku chemicznego oraz tworzenie nowych, ulepszonych metod chemoinformatycznych. Swoimi badaniami Doktorant wpisał się w ten światowy nurt badań ukierunkowując się na poszukiwanie połączenia rozwiązań chemoinformatycznych z metodami obliczeń kwantowo-chemicznych, wykorzystując zalety obu tych podejść. W swoich badaniach postawił i udowodnił tezę o możliwości budowy prostych i dokładnych modeli predykcyjnych lub klasyfikacyjnych, w oparciu o właściwości kwantowe fragmentów cząsteczek chemicznych pochodzących z wcześniej zdefiniowanych zbiorów fragmentów opisujących ich właściwości kwantowe. W swojej pracy Doktorant podjął się zdefiniowania nowych deskryptorów molekularnych opartych o właściwości kwantowe fragmentów cząsteczek chemicznych wraz z krytycznym spojrzeniem na ograniczenia tego podejścia oraz wystąpienie możliwych błędów. Do tego celu wykorzystywał dostępne bazy danych właściwości kwantowych i eksperymentalnych w celu budowy modeli QSAR/QSPR.

Reasumując uważam, że zaproponowany przez Doktoranta temat i zakres pracy doktorskiej wpisują się bez wątpienia w obszar ważnych i aktualnych problemów z obszaru obliczeń kwantowo-chemicznych oraz chemoinformatyki, w szczególności stworzenia nowych narzędzi wzbogacających paletę dostępnych metod chemoinformatycznych umożliwiających optymalizację wykonywania dużej liczby obliczeń kwantowych. Podjęcie tematu należy uznać za w pełni uzasadnione.

2. Ogólna charakterystyka rozprawy

Rozprawa doktorska zawiera 113 stron, 11 rozdziałów (w tym rozdział 11 stanowi opis części aplikacyjnej dostępnej dla pracowników laboratorium służącej do przewidywania długości fali maksimum emisji oraz interfejsu do generowania fragmentarycznych deskryptorów kwantowych i kwantowych par atomów), 41 rysunków, 12 tabel i 107 pozycji bibliograficznych, wszystkie w języku angielskim, obejmujących najważniejsze pozycje literatury przedmiotu z ostatnich lat.

Rozprawę doktorską podzielono na dwie części: część I obejmującą przegląd stanu wiedzy (rozdziały 1-3) oraz część II dotyczącą badań eksperymentalnych. Część dotyczącą przeglądu stanu wiedzy Doktorant rozpoczął od przedstawienia w rozdziale 1 różnorodności sposobów przedstawiania cząsteczki chemicznej, omawiając kolejno notacje liniowe, postać macierzową, tablice połączeń, jak również wskazuje na problemy związane z brakiem unikatowości (kanonikalizacja), stosowaniem innych metody opisu cząsteczek związków chemicznych w chemoinformatyce oraz stosowanych formatów plików zawierających struktury chemiczne. W rozdziale przedstawiono w sposób przejrzysty spotykane i stosowane w chemoinformatyce sposoby reprezentacji, przechowywania i przetwarzania informacji dotyczących struktur chemicznych w systemach informatycznych. Doktorant zwraca uwagę na brak standardowego formatu stosowanego w chemoinformatyce, co więcej słusznie wskazuje na problem związany z faktem, że część rozwiązań i algorytmów nie jest publicznie dostępnych. Pewne wątpliwości budzi stwierdzenie Doktoranta we wstępie rozdziału 1, że przedstawiony w rozdziale przegląd metod przedstawiania struktur chemicznych można poszerzyć „np. o artykuł Wigh i in. [1].” W części teoretycznej rozprawy doktorskiej powinno się przedstawić całościowo stan wiedzy a nie odsyłać czytelnika do innych pozycji literaturowych tzn. najważniejsze konkluzje wynikające z artykułu Wigh i in. powinny pojawić się w opisie stanu wiedzy w niniejszej pracy. Omawianie sposobów przedstawiania struktur Doktorant rozpoczyna od najbardziej popularnych notacji liniowych. W omówieniu ciągu SMILES wyjaśnienia wymaga błędne sformułowanie, że „atomy reprezentowane są przez symbole ich związków chemicznych”. W zapisie SMILES atomy są reprezentowane przez standardowe symbole pierwiastków a nie związków chemicznych.

Bardzo dobrze w części teoretycznej pracy opisano kwestie przedstawiania struktur związków chemicznych w formie grafów molekularnych. Doktorant omówił możliwości przedstawiania tychże grafów w formie macierzy, omawiając kolejno ich rodzaje (macierz

sąsiedztwa, incydencji, odległości, wiązań oraz wiązań i elektronów). Zostało to dobrze zobrazowane licznymi przykładami. Omawiając macierz incydencji zwrócono słusznie uwagę, że w teorii grafów istnieje jeszcze wartość -1, w przypadku grafów skierowanych, co nie występuje w grafach molekularnych. To jednak nie zostało zobrazowane na przykładzie i wymaga dodatkowego wyjaśnienia podczas publicznej obrony pracy. W rozdziale 2 pracy omówiono różne grupy deskryptorów molekularnych oraz cechy je charakteryzujące. Kpt. mgr inż. Fliszkiewicz szczegółowo omówił wymagania jakie stawiane są deskryptorom molekularnym, które mają kodować informacje dotyczące struktury związków chemicznych. W opisie Doktorant wskazał, że w podziale deskryptorów molekularnych „granice między wymaganą reprezentacją do wygenerowania danego deskryptora zaliczanego do deskryptorów 0D, 1D i 2D nie są ostre”. Doktorant próbował wyjaśnić to stwierdzenie tym, że deskryptory 0D i 1D tworzą wspólną grupę deskryptorów konstrukcyjnych (uwzględniających tylko skład atomów i rodzaj wiązań) natomiast deskryptory 2D są deskryptorami topologicznymi, czyli wymagają grafu molekularnego i uwzględniają wzajemne połączenia i ustawienia atomów, jednak wymaga to moim zdaniem dodatkowego komentarza podczas publicznej obrony pracy.

W podrozdziale 2.3 Doktorant wspomniał o wykorzystaniu uczenia maszynowego jako alternatywy dla deskryptorów molekularnych. Jest to bardzo atrakcyjna opcja, w której zamiast stosowania skomplikowanych operacji matematycznych w celu kodowania zjawisk i zależności strukturalnych w cząsteczkach, wykorzystuje się podejście, w którym komputer samodzielnie konstruuje reprezentację cząsteczki na potrzeby danego zadania, tzw. poznawanie reprezentacji. To podejście określane jest jako metoda sterowania danymi, i wykorzystuje się tu np. metody grafowych sieci neuronowych. Proszę tu jednak o komentarz podczas publicznej obrony na temat możliwości wystąpienia błędu (np. przeuczenie się sieci, ekstrema lokalne, itp.).

W rozdziale 3 kpt. mgr inż. B. Fliszkiewicz krótko scharakteryzował aplikacje zarówno komercyjne jak i niekomercyjne stosowane do obliczania wartości deskryptorów molekularnych.

Na koniec przedstawionego opisu stanu wiedzy zabrakło krótkiego, syntetycznego podsumowania części teoretycznej pracy, uzasadniającego cel prowadzenia podjętych badań i stanowiącego nawiązanie do kolejnej części rozprawy, przedstawiającej metodykę i część badawczą.

Opis badań własnych Doktorant rozpoczął od scharakteryzowania stosowanej metodyki oraz przedstawienia obiektu badań (rozdział 4). Szczegółowo przedstawił typy i parametry techniczne wykorzystanych w obliczeniach komputerów; wskazał, że podstawowym narzędziem pracy był język Python wraz ze wszystkimi wykorzystanymi bibliotekami/modułami Python'a spoza biblioteki. Jako algorytmy uczenia maszynowego w badaniach wykorzystano regresję liniową lub logistyczną. Dodatkowo w pracy wykorzystano metody złożone oparte o drzewa decyzyjne. Doktorant wskazał, że w celu jak najlepszej oceny i porównania zdolności predykcyjnych budowanych modeli uczenia maszynowego zastosował metryki oceniające. Czy w określaniu błędu przewidywania podzielono zbiór danych na zbiór uczący i testowy czy też zastosowano krosvalidację? Omawiając prowadzone badania Doktorant „dokonał trzech różnych definicji proponowanych deskryptorów kwantowych”. Prosiłbym o doprecyzowanie tego stwierdzenia podczas publicznej obrony pracy.

W rozdziale 5 Doktorant omówił fragmentaryczne deskryptory kwantowe. Co Doktorant miał na myśli pisząc, że „w niniejszej pracy doktorskiej podjęto wysiłek ku wzbogaceniu dostępnych metod opisu struktur związków chemicznych na podstawie podstruktur, które można w nich wyróżnić”? Omawiając potencjalne ograniczenia opisu struktury cząsteczki związku przez jej podstruktury, Doktorant wskazał, że istnieją takie podstruktury, których nie można było w przeprowadzonych badaniach i analizach wyróżnić. Proszę również o komentarz do tego stwierdzenia podczas publicznej obrony.

Bardzo dobrze przedstawiono w rozdziale 6 kwestie predykcji długości fali dla maksimum emisji związków optycznie czynnych. Stworzenie modelu wiążącego znane wartości parametrów fizycznych ze strukturą związków chemicznych umożliwia opracowanie narzędzi informatycznych, pozwalających na wstępną ocenę projektowanych związków lub wykorzystanie do badań przesiewowych wielu struktur i wytypowania kandydatów do dalszych badań. To pozwoliło Doktorantowi wykorzystać fragmentaryczne deskryptory kwantowe do zbudowania modelu regresji wybranej właściwości optycznej (tj. długości fali, przy której następuje maksimum emisji) związków organicznych typu OLED. Dodatkowy komentarz przydałby się w części, w której Doktorant przedstawił analizę wyników i dyskusję (podrozdział 6.3), gdzie wskazał, że wyniki średniego błędu kwadratowego miały podobny charakter do wyników średniego błędu bezwzględnego, przy czym model, w którym zastosowano tylko nowe deskryptory kwantowe osiągnął wartości tej miary ponad dwukrotnie większe od wyników modeli, dla których zastosowano tradycyjne deskryptory molekularne.

W rozdziale 7 kpt. mgr. inż. B. Fliszkiewicz przedstawił wyniki swoich badań nad rozszerzeniem bazy danych właściwości kwantowych QM9 o związki zawierające w swojej strukturze inne halogeny niż tylko fluor. Takie podejście pozwoliło na mniejsze ograniczenie bazy danych eksperymentalnych, a jednocześnie w celu zachowania spójności między bazą rozszerzoną a pierwotną, obliczenia wykonano metodą B3LYP, która jest jedną z metod teorii funkcjonału gęstości elektronowej (DFT), w której zastosowany jest korelacyjno-wymienny funkcjonał hybrydowy Becke'go, Lee-Yang Parr'a. W efekcie należy stwierdzić, że rozbudowana baza danych właściwości kwantowych umożliwia poszerzenie domeny aplikowalności fragmentarycznych deskryptorów kwantowych w pierwszej kolejności o związki chemiczne zawierające atomy chloru a następnie bromu. To należy uznać moim zdaniem za duże osiągnięcie recenzowanej pracy doktorskiej. Te wyniki posłużyły jako wsad do dalszych badań w ramach pracy doktorskiej opisanych w rozdziale 8, dotyczących klasyfikacji związków chemicznych ze względu na bioakumulację w organizmach żywych. W wyniku badań stwierdzono jednak, że fragmentaryczne deskryptory kwantowe nie wywołały polepszenia skuteczności klasyfikacji ze względu na zdolność do bioakumulacji w stosunku do znanych już deskryptorów molekularnych. Jednocześnie nasuwa się pytanie w kwestii metodyki przedstawionej w podrozdziale 8.1 w podpunkcie dotyczącym brakujących danych. Dlaczego rozważano zastępowanie brakujących elementów zerami. Co jest szczególnie w zerze. Czy to nie zaburzyło struktury danych?

Ostatnim rozdziałem dotyczącym badań własnych jest rozdział 9, w którym kpt. mgr. inż. B. Fliszkiewicz zajął się tematyką kwantowo informowanych par atomów, czyli podejściem, w którym nie definiuje się nowych deskryptorów kwantowych od podstaw, ale skupia się na „ukwantowaniu” metody już istniejącej. Na podstawie przeprowadzonych badań Doktorant stwierdził (głównie w odniesieniu do modeli regresyjnych), że metoda opisu struktur związków chemicznych poprzez właściwości kwantowe par atomów pozwala na zwiększenie skuteczności modelowania w stosunku do opisu przy pomocy standardowych par atomów.

Część dotyczącą badań własnych kończy rozdział 10, w którym Doktorant w sposób zwięzły przedstawił wnioski z przeprowadzonych badań własnych, jak również wskazał możliwe dalsze kierunki rozwoju tychże badań.

Czytając rozprawę doktorską zauważono drobne niedociągnięcia o charakterze redakcyjnym. Wybrane z nich przedstawiono poniżej:

1. Brakuje numeracji poszczególnych rozdziałów w tekście.
2. W pracy nie ustrzeżono się drobnych błędów stylistycznych (np. str.6 jest „Obecnie najbardziej popularna notacja liniowa.”, powinno być „Obecnie najbardziej popularna jest notacja liniowa.”; str. 20/21 jest „Ponadto pożądanym jest, aby deskryptor molekularny: ...8. można było go zastosować...” powinno być ...8. Można było zastosować...”);
3. Cytując prace innych autorów powinno pisać się Nazwisko i in. a nie „Nazwisko i wsp.”, np. str.6 jest Wigh i wsp., powinno być Wigh i in.; str. 9 jest Kernn i wsp. Powinno być Kernn i in.;
4. Opisy na rysunkach powinny być w języku polskim, np. Rys.1.2.; Rys.6.7;
5. Nomenklatura IUPAC powinna być stosowana w całej pracy, np. na str. 63 jest „dwutlenek węgla” powinno być „ditlenek węgla”, etc.
6. W Tabeli 8.1 i 8.2 na str. 73 powinno zastąpić się kropki przecinkami (wartości dziesiętne zapisujemy w języku polskim po przecinku). Ta sama uwaga dotyczy wartości przedstawionych w „dodatkach” (powinno być załącznikach) do rozdziału 8 – patrz Tabele B1-B4.
7. Brak konsekwencji w spisie cytowanej literatury. Raz wymieniani są wszyscy autorzy, raz tylko jeden autor i in., raz dwóch autorów i in. Ponadto raz podawane jest imię i nazwisko autora pracy, raz tylko inicjały.

Wniosek końcowy

Jednoznacznie stwierdzam, że rozprawa doktorska Pana kpt. mgr. inż. Bartłomieja Fliszkiewicza stanowi oryginalne rozwiązanie problemu naukowego. Doktorant, na podstawie przeprowadzonych badań i analiz, potwierdził postawioną tezę, iż możliwe jest zbudowanie prostych i dokładnych modeli predykcyjnych lub klasyfikacyjnych, w oparciu o właściwości kwantowe fragmentów cząsteczek chemicznych pochodzących z wcześniej zdefiniowanego zbioru fragmentów opisującego ich właściwości kwantowe. Uzyskane wyniki mają wysoką wartość naukową i poznawczą. Doktorant wykazał się dobrą umiejętnością formułowania problemów naukowych, prowadzenia badań oraz dokonywania

analiz wyników. Przedstawiona do recenzji rozprawa doktorska dotyczy dyscypliny naukowej Nauki Chemiczne.

Stwierdzam, że praca doktorska kpt. mgr. inż. Bartłomieja Fliszkiewicza pt. „*Właściwości kwantowe fragmentów cząsteczek jako deskryptory molekularne*” w pełni odpowiada warunkom określonym w ustawie z dnia 20 lipca 2018 Prawo o szkolnictwie wyższym i nauce i wnoszę o dopuszczenie jej przez Radę Dyscypliny Naukowej - Nauki Chemiczne Wojskowej Akademii Technicznej do publicznej obrony.