

prof. dr hab. Tomasz Puzyn
Kierownik Pracowni Chemoinformatyki Środowiska

RECENZJA

rozprawy doktorskiej kpt. mgr. inż. Bartłomieja Fliszkiewicza
pt. "Właściwości kwantowe fragmentów cząsteczek jako deskryptory molekularne"

1. Wstęp

Przedstawiona mi do recenzji rozprawa doktorska kpt. mgr. inż. Bartłomieja Fliszkiewicza pt. "Właściwości kwantowe fragmentów cząsteczek jako deskryptory molekularne" wykonana została na Wydziale Nowych Technologii i Chemii Wojskowej Akademii Technicznej, pod kierunkiem dr. hab. inż. Marcina Sajdaka z Katedry Ochrony Powietrza na Wydziale Inżynierii Środowiska i Energetyki Politechniki Śląskiej.

2. Ocena formalna

Rozprawa liczy 113 stron i ma formę monografii o układzie typowym dla tego rodzaju prac. Składa się z dwóch głównych części zatytułowanych: „Przegląd stanu wiedzy” (28 stron) oraz „Część eksperymentalna” (60 stron). Elementami rozprawy są również streszczenia w języku polskim i angielskim, prezentacja tezy i celów pracy oraz dodatki do rozdziałów, lista publikacji i wystąpień konferencyjnych oraz wykaz użytej bibliografii. Formalna strona rozprawy spełnia zatem wymogi opisane w stosownej ustawie.

3. Ocena merytoryczna

Wraz z nadejściem czwartej rewolucji przemysłowej metody uczenia maszynowego i sztucznej inteligencji znajdują coraz szersze zastosowanie w wielu dziedzinach nauki i przemysłu. Jednym z pionierskich obszarów tym zakresie są metody badania zależności pomiędzy strukturą chemiczną związku a jego właściwościami lub aktywnością biologiczną, określane w języku angielskim nazwą odpowiednio: *Quantitative Structure-Property Relationships* (QSPR) lub *Quantitative Structure-Activity Relationships* (QSAR). Metody modelowania QSPR/QSAR rozwijane są już od lat sześćdziesiątych ubiegłego wieku. Zmiennymi objaśniającymi w modelach QSPR/QSAR są tzw.

deskryptory molekularne odzwierciedlające w sposób numeryczny różnice pomiędzy poszczególnymi związkami chemicznymi w modelowanej grupie. Natomiast zmienną zależną – interesująca badacza właściwość fizykochemiczna, aktywność biologiczna (np. terapeutyczna) lub toksyczność związku.

W powyższy obszar wpisują się badania kpt. mgr. inż. Bartłomieja Fliszkiewicza. Autor zaproponował i zweryfikował ciekawą hipotezę, w myśl której: „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”. Dodatkowo założył realizację trzech celów pośrednich: (i) „dokonanie rozszerzenia bazy danych właściwości kwantowych”; (ii) „optymalizacja wykonywania dużej liczby obliczeń kwantowych” oraz (iii) „stworzenie oprogramowania pozwalającego na obliczanie deskryptorów kwantowych opracowanych na potrzeby tej pracy”.

W części I rozprawy Doktorant przeprowadził przegląd istniejącego stanu wiedzy koncentrując się na różnych typach reprezentacji cząsteczki oraz różnych rodzajach deskryptorów. Rozdział kończy się przeglądem oprogramowania powszechnie wykorzystywanego do obliczania wartości deskryptorów. Ta część, pomimo zwartej formy, zawiera dość kompleksowy i poprawnie przygotowany opis tematu. Autor udowadnia, że posiada ogólną wiedzę teoretyczną w dyscyplinie, a wiedzę szczegółową – w obszarze, którego dotyczą badania własne.

Zabrakło mi jednak tutaj pokazania szerszej perspektywy kierunków badań, które prowadzone są obecnie nad nowymi rodzajami deskryptorów popartego odniesieniami do najnowszego piśmiennictwa (np. deskryptory dla związków o budowie jonowej, czy deskryptory uwzględniające strukturę supramolekularną).

Na str. 21 Autor wspomina o wykorzystaniu logarytmu współczynnika podziału n-oktanol/woda ($\log P$) jako deskryptora eksperymentalnego. Chciałbym przy tej okazji spytać, czy $\log P$ może być również deskryptorem teoretycznym (obliczonym)? Jeśli tak, to w jaki sposób można tę wartość uzyskać?

Część eksperymentalna stanowi zapis przeprowadzonych badań, które można byłoby określić serią studiów przypadku przy wykorzystaniu różnych zbiorów danych. W badaniach tych Autor testował dwie zaproponowane przez siebie metody kwantowych reprezentacji związków chemicznych: (i) reprezentacje oparte o właściwości kwantowe wykrytych fragmentów cząsteczek związków chemicznych oraz (ii) kwantowo zmodyfikowane pary atomów. Do najważniejszych osiągnięć Doktoranta należą:

- Pokazanie, że w przypadku zastosowania fragmentarycznych deskryptorów kwantowych otrzymywano lepsze wyniki niż ustalenie predykcji jako wartości średniej w przypadku regresji lub przypisanie losowe w przypadku klasyfikacji.
- Zweryfikowanie użyteczności fragmentarycznych deskryptorów kwantowych w kontekście potrzebnego czasu i zasobów obliczeniowych.
- Udowodnienie, że kwantowe pary atomów w przypadku części modelowanych właściwości poprawiły zdolności predykcyjne w stosunku do swoich pierwowzorów.

Dodatkowo kpt. mgr inż. Bartłomiej Fliszkiewicz stworzył i udostępnił oprogramowanie umożliwiające wyznaczenie wartości deskryptorów badanych w pracy.

Po lekturze tej części rozprawy nie mam najmniejszych wątpliwości, że Doktorant posiada umiejętność samodzielnego prowadzenia pracy naukowej, a także, że jej przedmiotem było oryginalne rozwiązanie problemu naukowego. Z obowiązku recenzenta chciałbym jednak poddać pod dyskusję na kilka kwestii merytorycznych, które nie umniejszają jednak mojego pozytywnego odbioru rozprawy:

1. W przypadku przewidywania długości fali dla maksimum emisji związków optycznie czynnych zastanowiła mnie liczba używanych deskryptorów oraz sposób ich przygotowania (*preprocessing*). Po pierwsze, wraz ze zwiększaniem się liczby zmiennych w macierzy deskryptorów rośnie prawdopodobieństwo uzyskania korelacji przypadkowej (ang. *correlation-by-chance*) [Topliss i Costello (1972) *J. Med. Chem.* 15 (10), 1066]. Jakie algorytmy redukcji/wyboru zmiennych zastosował Autor w tym przypadku? Po drugie, duża korelacja pomiędzy deskryptorami może prowadzić do przeuczenia (ang. *overfitting*) modelu. Jakie metody ortogonalizacji macierzy deskryptorów lub eliminacji deskryptorów skorelowanych zostały zastosowane?
2. Czy w związku z długim czasem obliczeń DFT oraz zapotrzebowaniem na zasoby możliwe byłoby wyznaczenie brakujących wartości deskryptorów dla związków w bazie QM9 w oparciu o stworzenie odpowiedniego modelu uczenia maszynowego? Alternatywnie czy uzasadnione mogłoby być podjęcie próby wyznaczenia tych wartości przy wykorzystaniu funkcjonałów typu DM21 [Kirkpatrick et al. (2021) *Science* 374, 6573]?
3. Jak zdefiniowane zostały klasy bioakumulacji (str. 66)? Czy użyto tutaj kryteriów, które mogłyby mieć znaczenie dla regulacji obrotu związkami chemicznymi wg. tzw. Rozporządzenia REACH [Rozporządzenie nr 1907/2006 Parlamentu Europejskiego i Rady z dnia 18 grudnia 2006 r.]?

4. Czy w przypadku brakujących wartości deskryptorów kwantowych używano wartości „0”? Jeśli tak, proszę o komentarz, jaki wpływ na ocenę właściwości związku mogłaby mieć informacja, że jego moment dipolowy jest równy zero?
5. Istotną cechą modelu, która zwiększa jego wiarygodność z punktu widzenia wykorzystania potencjalnych przewidywań do celów regulacyjnych, jest możliwość interpretacji użytych deskryptorów w kontekście wyjaśnienia mechanizmu odpowiedzialnego za zmianę obserwowanej właściwości w zbiorze modelowanych związków [(Q)SAR Assessment Framework, OECD, Series on Testing and Assessment No. 386]. Proszę o komentarz, w jaki sposób interpretuje się fragmentaryczne deskryptory kwantowe i ich wpływ na właściwości całej cząsteczki?

4. Ocena strony edytorskiej

Zwyczajowym obowiązkiem recenzenta jest również ocena edytorskiej strony dysertacji. Rozprawa sformatowana jest bardzo estetycznie dzięki użyciu procesora LaTeX. Napisana jest poprawnym językiem, chociaż czytelnika razić mogą zwroty z języka potocznego i żargonu branżowego, np. „maszyna posiadała 32 procesory” (str. 35), „policzenia nowego rodzaju deskryptorów” (str. 41), „Za pomocą skryptu Pythona generowano pliki wsadowe do obliczeń” (str. 57). Zdarzają się pojedyncze literówki (np. w rozwinięciu skrótów QSAR i QSPR w „Spisie skrótów” na początku pracy). Zalecałbym również zmianę sposobu narracji użytego w rozprawie na sposób niepozostawiający wątpliwości dotyczących tego, że Doktorant przeprowadził badania samodzielnie.

5. Konkluzja

W mojej ocenie rozprawa stanowi oryginalne rozwiązanie istotnego problemu naukowego oraz potwierdza ogólną wiedzę teoretyczną Kandydata w dyscyplinie. Potwierdza również umiejętność samodzielnego prowadzenia pracy naukowej. Drobne potknięcia, które wymieniłem w niniejszej recenzji nie wpływają na wysoką ocenę merytoryczną przeprowadzonych badań. Dlatego z pełnym przekonaniem stwierdzam, że rozprawa kpt. mgr. inż. Bartłomieja Fliszkiewicza spełnia wymagania zwyczajowe i formalne stawiane rozprawom doktorskim. Wnioskuje zatem o dopuszczenie Kandydata do dalszych etapów przewodu doktorskiego.

Gdańsk, 24 maja 2024 r.

