

dr hab. Piotr Wnuk-Lipiński, prof. UWr
Zakład Inteligencji Obliczeniowej
Instytut Informatyki
Uniwersytet Wrocławski
ul. F. Joliot-Curie 15
50-383 Wrocław

Wrocław, dnia 10 stycznia 2024 r.

Recenzja rozprawy doktorskiej
mgr Agnieszki Wojtuch
"Adapting explainable artificial intelligence to support drug design"

Rozprawa doktorska dotyczy zastosowania algorytmów wyjaśnialnej sztucznej inteligencji, zwłaszcza wyjaśniania grafowych sieci neuronowych, do wsparcia projektowania związków chemicznych. Skupia się głównie na dwóch problemach:

- doborze efektywnej reprezentacji związków chemicznych w modelach grafowych sieci neuronowych (w tym wykorzystaniu wyjaśnień modeli grafowych sieci neuronowych do określenia istotności poszczególnych cech związków chemicznych) oraz
- wykorzystaniu wyjaśnień modeli uczenia maszynowego do konstrukcji reguł umożliwiających optymalizację właściwości molekularnych związków chemicznych (optymalizacja kierowana wyjaśnialnością).

Rozprawa doktorska ma formę monografii, ale Autorka wskazuje 4 publikacje naukowe, jedną opublikowaną na międzynarodowej konferencji naukowej poświęconej sieciom neuronowym (IJCNN), dwie opublikowane i jedną złożoną w czasopismach naukowych poświęconych zastosowaniom informatyki w chemii i biotechnologii (Journal of Cheminformatics, Computational and Structural Biotechnology Journal), które zawierają część wyników z rozprawy doktorskiej.

1. Oryginalny wkład Autorki zawarty w rozprawie doktorskiej

Najważniejsze osiągnięcia Autorki zawarte w rozprawie doktorskiej dotyczą:

- porównania reprezentacji atomów w kontekście ich użyteczności w grafowych sieciach neuronowych oraz transformerach (uzyskane wyniki pokazują, że skuteczność modelu zależy od wybranej reprezentacji danych, ale optymalna reprezentacja danych zależy od konkretnego zadania i konkretnej architektury, dodatkową wartością jest też propozycja kilkunastu reprezentacji danych o dość wysokiej skuteczności w testowych zadaniach),
- wykorzystania wyjaśnień modeli grafowych sieci neuronowych do określenia istotności poszczególnych cech związków chemicznych (wartością tych badań jest sam pomysł ekstrakcji wiedzy o istotności cech z generowanych wyjaśnień oraz wybór konkretnych cech dla testowych zadań),

Pl

- propozycji metodologii wykorzystującej wyjaśnienia modeli uczenia maszynowego do konstrukcji reguł umożliwiających optymalizację właściwości molekularnych związków chemicznych (optymalizacja kierowana wyjaśnialnością, może być stosowana także do innych zagadnień niż rozważane w niniejszej rozprawie),
- opracowania modeli predykcji i optymalizacji stabilności metabolicznej związków chemicznych.

Uzyskane wyniki uważam za wartościowe, zarówno w kontekście samego uczenia maszynowego, gdyż wprowadzają nowe metody wykorzystywania wiedzy z generowanych wyjaśnień modeli uczenia maszynowego, jak i zastosowań uczenia maszynowego w analizie związków chemicznych czy projektowaniu leków, gdyż dostarczają nowych modeli uczenia maszynowego dla problemów rozpatrywanych w rozprawie doktorskiej.

2. Treść rozprawy doktorskiej

Rozdział 1 wprowadza do tematyki rozprawy doktorskiej, przedstawia kontekst prowadzonych badań naukowych oraz znaczenie postawionych przez Autorkę problemów. Kolejne sekcje zawierają krótki przegląd literatury dotyczącej komputerowo wspomaganego projektowania leków i wyjaśnialności sztucznej inteligencji oraz streszczenie opublikowanych prac Autorki. Rozdział ten potwierdza, że Autorka ma świadomość znaczenia prowadzonych przez siebie badań naukowych, potrafi powiązać je z pracami prowadzonymi przez innych naukowców oraz wpisać we współczesne kierunki rozwoju studiowanej tematyki.

Rozdział 2 przedstawia najważniejsze metody sztucznej inteligencji związane z dalszą treścią pracy. Pierwsza część krótko opisuje podstawowe modele grafowych sieci neuronowych stosowane do przetwarzania danych chemicznych, takie jak Message Passing Neural Networks, Graph Convolutional Neural Networks, Directed Message Passing Neural Networks oraz Molecule Attention Transformer. Sam opis jest bardzo zwięzły – pozbawiony wyjaśnienia ogólnej architektury opisywanych modeli i jej poszczególnych elementów, przez co zrozumiały jedynie dla czytelnika z doświadczeniem w omawianej tematyce – co samo w sobie nie jest wadą, ale pewnym brakiem jest pominięcie dyskusji znaczenia i różnic między omawianymi czterema modelami grafowych sieci neuronowych. Druga część poświęcona jest wyjaśnialnej sztucznej inteligencji i opisuje popularne podejścia, takie jak SHAP (ogólne podejście do wyjaśniania modeli sztucznej inteligencji) i GNNExplainer (podejście dedykowane do wyjaśniania modeli grafowych sieci neuronowych). Opis utrzymany jest w zwięzłym stylu, ale tutaj wystarczającym do zrozumienia idei przedstawianych podejść. Kolejna część opisuje zbiory danych używane w eksperymentach obliczeniowych przedstawionych w dalszej części pracy. Opis jest dość techniczny i szczegółowy, co pozwala zrozumieć charakterystykę zbiorów danych i metody ich wstępnego przetwarzania. Ostatnia część przedstawia popularne reprezentacje danych molekularnych.

Rozdziały 3 i 4 dotyczą sposobów reprezentacji związków chemicznych w grafowych sieciach neuronowych: Rozdział 3 zawiera analizę wybranych reprezentacji danych i ich wpływu na skuteczność wybranych metod przetwarzania danych. Rozdział 4 skupia się na wyjaśnialnej sztucznej inteligencji i wykorzystaniu otrzymywanych wyjaśnień do inżynierii cech atomów, m.in.

określaniu istotności poszczególnych cech i ich przydatności w modelach klasyfikacji czy regresji oraz wyborze stosownych cech atomów.

Autorka słusznie zauważa, że skuteczność modeli sztucznej inteligencji zależy nie tylko od trafnego wyboru architektury, co jest ostatnio mocno eksploatowanym kierunkiem badań w głębokim uczeniu maszynowym, ale także od stosownej reprezentacji danych i stosownego wyboru cech obserwacji. W Rozdziale 3 przedstawia swoje badania dotyczące porównania reprezentacji atomów w kontekście ich użyteczności w grafowych sieciach neuronowych oraz transformerach. Eksperymenty obliczeniowe dotyczą wybranych 16 reprezentacji danych (12 własnych i 4 znanych z literatury), modeli GCN, MPNN, D-MPNN i MAT oraz czterech testowych zadań predykcji właściwości molekularnych na zbiorach danych ESOL, QM9, HUMAN i RAT. Badania zostały zaplanowane i przeprowadzone rzetelnie, w sposób powszechnie przyjęty dla podobnych prac w dziedzinie uczenia maszynowego, pewnym brakiem wydaje się jedynie pominięcie dyskusji parametryzacji algorytmów uczenia, optymalizacji hiperparametrów, itp. (częściowo przeniesionej do dodatków). Uzyskane wyniki pokazują, że skuteczność modelu zależy od wybranej reprezentacji danych, ale optymalna reprezentacja danych zależy od konkretnego zadania i konkretnej architektury. Chociaż sam wniosek nie jest szczególnie odkrywczy, wartością przeprowadzonych badań jest rzetelne potwierdzenie wpływu reprezentacji danych na skuteczność modeli w studiowanych problemach, a także (a może przede wszystkim) opracowanie propozycji 12 reprezentacji danych o dość wysokiej skuteczności w testowych zadaniach. Kontynuacją jest Rozdział 4, w którym Autorka omawia badania nad wykorzystaniem wyjaśnień modeli grafowych sieci neuronowych generowanych przez GNNExplainer do określenia istotności poszczególnych cech. Wartością tych badań jest sam pomysł ekstrakcji wiedzy o istotności cech z generowanych wyjaśnień oraz wybór konkretnych cech dla testowych zadań.

Rozdział 5 poświęcony jest optymalizacji kierowanej wyjaśnialnością (ang. Explanation-driven Optimisation, EDO). Autorka proponuje metodologię wykorzystującą wyjaśnienia modeli uczenia maszynowego do konstrukcji reguł umożliwiających optymalizację właściwości molekularnych związków chemicznych. Podejście opiera się na prostych modelach naiwnych klasyfikatorów bayesowskich, SVM czy klasyfikatorów drzewiastych i ich wyjaśnieniach przez SHAP oraz zaproponowanym przez Autorkę algorytmie konstrukcji reguł, wykorzystującym wyjaśnienia SHAP i występujące w nim cechy. Rozdział bardzo szczegółowo opisuje proponowane podejście, ilustrując je licznymi przykładami. Pewnym brakiem jest pominięcie dyskusji własności informatycznych proponowanego podejścia, złożoności obliczeniowej, możliwych problemów numerycznych, itp. Niemniej, zaproponowane podejście jest bardzo ciekawe i wartościowe, czego dowodzi Rozdział 7, może być także stosowane do innych zagadnień niż rozważane w niniejszej rozprawie.

Rozdziały 6 i 7 dotyczą stabilności metabolicznej, predykcji i optymalizacji, odpowiednio.

Rozdział 6 przedstawia modele służące do przewidywania stabilności metabolicznej (dla dwóch wybranych zbiorów danych), oparte na naiwnym klasyfikatorze bayesowskim, SVM i klasyfikatorach drzewiastych (pierwszy do klasyfikacji, dwa kolejne zarówno do klasyfikacji jak i regresji). Opracowane modele zostały porównane z podstawowym podejściem opartym na klasyfikatorze KNN. Badania zostały przeprowadzone rzetelnie (na uwagę zasługuje staranny

proces optymalizacji hiperparametrów modeli, który często jest słabą stroną oceny skuteczności algorytmów uczenia maszynowego), są opisane starannie, część szczegółów jest wyjaśniona dokładniej w dodatku B. Pewne wątpliwości budzi wybór bardzo prostego podejścia opartego na KNN jako podejścia referencyjnego do oceny wyników proponowanych modeli oraz przedstawienie jedynie wartości metryk ewaluacyjnych AUROC OVO i MSE bez bardziej wnikliwej analizy otrzymywanych wyników klasyfikacji i regresji (ciekawe wydaje się sprawdzenie dla jakich próbek danych modele popełniają największe błędy, jaki jest rozkład tych błędów, itp., a same metryki pokazują jedynie ostateczne uśrednione wartości).

Rozdział 7 poświęcony jest zastosowaniu optymalizacji kierowanej wyjaśnialnością, opisanej w Rozdziale 5, do optymalizacji właściwości molekularnych związków chemicznych. Autorka przedstawia metodologię wykorzystania wiedzy pochodzącej z wyjaśnień modeli uczenia maszynowego (klasyfikacji, regresji) do konstrukcji reguł pozwalających na optymalizację właściwości związków chemicznych. Jest to interesujące, oryginalne i wartościowe podejście. W przeprowadzonych eksperymentach obliczeniowych, Autorka sprawdza kilka możliwych wariantów zastosowań proponowanej metodologii i porównuje je z podstawowym podejściem opartym na regułach losowych (można mieć pewne wątpliwości do poziomu skuteczności takiego podejścia referencyjnego, niemniej ma ono wystarczający poziom do potwierdzenia, że proponowana metodologia dostarcza wartościowych wyników). Eksperymenty są starannie opisane, część szczegółów jest wyjaśniona dokładniej w dodatku C. Ciekawym eksperymentem jest zastosowanie reguł uczonych na jednym zbiorze obserwacji (niska stabilność) do innego zbioru obserwacji (średnia stabilność) i uzyskanie wysokiej skuteczności takich reguł. Niektóre eksperymenty pozostawiają pole do dalszych badań, m.in. konstrukcja jednego uniwersalnego zbioru reguł dla wszystkich obserwacji czy konstrukcja reguł w oparciu o wyjaśnienia błędnych predykcji. Autorka omawia też praktyczne znaczenie przeprowadzonych badań, m.in. pokazuje ilustrację działania reguł na związkach chemicznych i sposób interpretacji spodziewanych zysków w odniesieniu do pewnych podstruktur cząsteczek.

Rozdział 8 zawiera obszerną dyskusję przeprowadzonych badań, z uwzględnieniem ich praktycznego znaczenia w analizie związków chemicznych czy projektowaniu leków. Podkreśla to wiedzę i doświadczenie Autorki nie tylko w dyscyplinie informatyki, w zakresie uczenia maszynowego i sztucznej inteligencji, ale także w dziedzinie chemii / biotechnologii, w której opracowane podejścia informatyczne Autorka stosuje.

3. Redakcja rozprawy doktorskiej

Rozprawa doktorska składa się z 8 rozdziałów i 4 dodatków. Układ pracy jest bardzo przejrzysty. Struktura podziału treści odpowiada zawartości rozprawy doktorskiej - poszczególne rozdziały dotyczą kolejnych tematów studiowanych przez Autorkę. Dodatki uzupełniają główną treść pracy i dostarczają szczegółów technicznych dotyczących przeprowadzonych eksperymentów obliczeniowych.

Praca jest zrozumiała (choć wymaga trochę wysiłku od czytelnika niezaznajomionego z tematyką zastosowań uczenia maszynowego w chemii / biotechnologii) i napisana w sposób staranny.

Szczególnie starannie i wyczerpująco opisane są przeprowadzone eksperymenty obliczeniowe dotyczące optymalizacji kierowanej wyjaśnialnością do optymalizacji właściwości molekularnych związków chemicznych, także praktyczne znaczenie takich badań.

4. Konkluzja

Rozprawę doktorską mgr Agnieszki Wojtuch oceniam bardzo pozytywnie. Sugestie i uwagi krytyczne występujące w mojej recenzji, dotyczące pewnych fragmentów pracy, są naturalnym elementem dyskusji naukowej i nie zmieniają mojej pozytywnej opinii o całości pracy.

Uważam, że rozprawa doktorska mgr Agnieszki Wojtuch spełnia wymagania stawiane rozprawom doktorskim w dyscyplinie naukowej informatyka, stanowi oryginalne rozwiązanie problemu naukowego, potwierdza wiedzę i umiejętności kandydatki w dyscyplinie informatyki oraz spełnia warunki określone w *Ustawie z dnia 14 marca 2003 o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki* (Dz.U. 2003, nr 65, poz. 595, z późniejszymi zmianami). Wnoszę więc o dopuszczenie rozprawy doktorskiej mgr Agnieszki Wojtuch do publicznej obrony.

Piotr Lipiński