

Adaptacja Architektur Głębokiego Uczenia do Procesu Odkrywania Leków

mgr Łukasz Maziarka

Streszczenie

Proces opracowywania nowych leków jest złożony, kosztowny i wymaga dużych zasobów. Jednocześnie znaczna część potencjalnych kandydatów na leki nie przechodzi pomyślnie późnych etapów badań klinicznych, co podkreśla potrzebę stworzenia lepszych modeli predykcyjnych i bardziej efektywnych metod optymalizacji związków chemicznych, używanych już na wczesnych etapach rozwoju leków. Niniejsza rozprawa doktorska pokazuje, że architektury głębokiego uczenia mogą być specjalnie dostosowane i zoptymalizowane do zadań cheminformatycznych, co pozwoli sprostać wyzwaniom związanym z reprezentacją molekuł oraz zwiększyć efektywność i dokładność różnych etapów odkrywania leków.

Celem tej rozprawy doktorskiej jest zbadanie możliwości adaptacji głębokich sieci neuronowych do kluczowych wyzwań cheminformatycznych, ze szczególnym uwzględnieniem komputerowo wspomaganego projektowania leków. Badania przedstawione w tej pracy skupiają się na dwóch głównych obszarach: przewidywaniu właściwości molekularnych oraz generowaniu nowych związków chemicznych. W rozprawie przedstawiono nowatorskie modele głębokiego uczenia, które wykorzystują konwolucyjne sieci grafowe (GCN), mechanizmy uwagi (attention mechanism) i architektury transformerów, w celu zwiększenia dokładności i efektywności przewidywania właściwości molekuł. Dodatkowo zaprezentowano innowacyjne metody generowania i optymalizacji cząsteczek, w tym wykorzystanie sieci generatywnych przeciwstawnych (GAN) oraz modeli opartych na przepływach (flow-based models) do warunkowego generowania molekularnego.

Wkład tej pracy obejmuje opracowanie architektury *Set Aggregation Network (SAN)* do ulepszonej agregacji cech w modelach grafowych, wprowadzenie *Spatial Graph Convolutional Network (SGCN)*, która integruje trójwymiarowe informacje przestrzenne do procesu konwolucji grafowej, oraz stworzenie architektur *Molecule Attention Transformer (MAT)* i jego wariantu, *Relative Molecule Attention Transformer (R-MAT)*, wykorzystujących mechanizmy uwagi do uchwycenia złożonych interakcji molekularnych pomiędzy atomami. Ponadto, rozprawa przedstawia *HuggingMolecules*, otwarto-źródłową bibliotekę, która ułatwia korzystanie z wytrenowanych już modeli do przewidywania właściwości molekularnych. Rozprawa porównuje również różne reprezentacje atomowe stosowane w grafowych sieciach

neuronowych, w celu identyfikacji optymalnych cech wejściowych oraz wprowadza *EmBERT* - adaptację architektury BERT do uczenia wielozadaniowego.

W zakresie generowania związków chemicznych, rozprawa wprowadza *Mol-CycleGAN*, metodę optymalizacji istniejących molekuł z wykorzystaniem sieci GAN, oraz *PluGeN*, opartą na przepływach sieć neuronową, która przekształca przestrzeń ukrytą istniejących modeli generatywnych w celu kontrolowanego generowania molekuł. Modele te pokazują potencjał głębokiego uczenia nie tylko w zakresie przewidywania właściwości związków chemicznych, ale także w zadaniu generowania nowych molekuł o pożądanych cechach, co przyspiesza proces odkrywania leków.

Zaprezentowane w tej rozprawie odkrycia i modele zostały szeroko uznane w społeczności cheminformatycznej, przyczyniając się do rozwoju metod głębokiego uczenia w projektowaniu leków, czego dowodem jest znacząca liczba cytowań prac kandydata, przekraczająca 800 cytowań według Google Scholar (stan na grudzień 2024 roku). Badania te zaowocowały kilkoma publikacjami o dużym wpływie oraz stanowią fundament przyszłych prac nad zastosowaniem sztucznej inteligencji w procesie odkrywaniu leków.

Podsumowując, niniejsza rozprawa doktorska skupia się na wykazaniu potencjału adaptacji architektur głębokiego uczenia do cheminformatyki, ze szczególnym naciskiem na komputerowo wspomagane projektowanie leków.

Na podstawie przeprowadzonych badań opublikowano dziewięć artykułów. Dwa z nich na konferencjach rangi A* zgodnie z klasyfikacją CORE, cztery na konferencjach rangi A, jeden na sesji warsztatowej konferencji rangi A* oraz dwa w renomowanym czasopiśmie cheminformatycznym. W pięciu z nich doktorant był pierwszym autorem. Ponadto, doktorant był kierownikiem projektu NCN Preludium, ściśle związanego z tymi badaniami.

Słowa kluczowe: głębokie uczenie, cheminformatyka, projektowanie leków, przewidywanie właściwości związków chemicznych, generowanie związków chemicznych.