

INFORMACJA O PROPONOWANEJ DO OTWARCIA ROZPRAWIE DOKTORSKIEJ

Głębokie modele generujące w zastosowaniu do projektowania leków

Doktorant: **Joanna KACZMAR**

Promotor: **prof. dr hab. inż. Jerzy ŚWIĄTEK**

Promotor pomocniczy: **dr Maciej ZIĘBA**

Uzasadnienie podjęcia tematu

Zastosowanie uczenia maszynowego do projektowania nowych leków jest ważnym tematem badań, umożliwiającym skrócenie czasu i zmniejszenie kosztów produkcji leków. W [1] do generowania nowych molekuł zostały użyte procedury oparte na mutacjach i krzyżowaniu podobne do algorytmów genetycznych. Takie przeszukiwanie przestrzeni jest jednak nieefektywne ze względu na liczbę możliwości i to, że tylko mały procent z zaprojektowanych w ten sposób molekuł jest syntezowalny.

Mimo dynamicznego rozwoju szczególnie gałęzi uczenia maszynowego, uczenia głębokiego, techniki z tej dziedziny zostały zastosowane do projektowania nowych leków dopiero ok. dwóch lat temu. Dotychczas zaproponowane modele [2, 3] operują na uproszczonym zapisie molekuł (SMILES), który pozbawiony jest istotnych informacji wpływających na wiązanie się cząsteczek leku (ligandu) z białkiem chorobotwórczym. Dodatkowo w żadnym z opisanych modeli nie ma możliwości warunkowania wyjścia białkiem chorobotwórczym, aby projektować leki dedykowane dla danego białka.

Praca związana jest z projektem POIR.01.01.01-00-1083/15, pt. "Opracowanie narzędzi bioinformatycznych do projektowania leków".

Cel rozprawy

Celem rozprawy jest zbudowanie modelu opartego na metodach uczenia głębokiego, który przyjmując na wejściu część białka chorobotwórczego odpowiedzialną za reakcję (tzw. pocket) generować będzie ligandy (cząsteczki leku) dostosowane do tego białka przy ograniczeniu, że ligandy takie są syntezowalne.

Metodyka badań

Zbudowany model zostanie zbadany pod kątem syntezowalności generowanych molekuł przy użyciu biblioteki rdkit, posiadającej zaimplementowaną walidację związków, na podstawie własności fizycznych i chemicznych atomów. Sprawdzona eksperymentalnie zostanie także zdolność wiązania się wygenerowanych molekuł z podanym białkiem. W tym celu użyty zostanie opracowany wcześniej model służący do skanowania wirtualnego, który mając na wejściu białko chorobotwórcze i potencjalny ligand na wyjściu odpowiada przybliżonym prawdopodobieństwem wiązania się tych dwóch cząsteczek.

Zakres rozprawy

- Wybór zbiorów benchmarkowych do badań
- Zbudowanie modelu GAN dedykowanego do problemu generowania ligandów.
- Wyuczenie modelu na zbiorze danych PDB-bind i DUD-e
- Wykonanie eksperymentów mających na celu sprawdzenie jakości (syntezowalności) wygenerowanych ligandów
- Porównanie otrzymanych wyników z wynikami literaturowymi.
- Zbudowanie warunkowego modelu GAN do generowania ligandów łączących się z zadaniem białkiem
- Wyuczenie warunkowego modelu GAN do generowania ligandów łączących się z zadaniem białkiem.
- Wykonanie eksperymentów sprawdzających skuteczność łączenia się wygenerowanych ligandów z pocketami.
- W przypadku niewystarczającej pojemności modelu GAN połączenie modelu DNC z modelem GAN.
- Powtórzenie eksperymentu dotyczącego skuteczności łączenia się wygenerowanych ligandów z pocketami dla tak rozszerzonego modelu.

Uzyskane wyniki

Wraz z zespołem projektowym udało się nam zbudować model służący do skanowania wirtualnego[4, 5], który mając na wejściu białko chorobotwórcze i potencjalny ligand na wyjściu odpowiada prawdopodobieństwem związania się tych dwóch cząsteczek, który będzie używany do weryfikacji wyników w proponowanej pracy.

Literatura

1. Virshup A. M., Contreras-García J., Wipf P., Yang W., Beratan D. N., Stochastic Voyages into Uncharted Chemical Space Produce a Representative Library of All Possible Drug-Like Compounds, *Journal of American Chemical Society, ACS*, 135 (19), 2013, pp 7296–7303
2. Gómez-Bombarelli R., Duvenaud D., Hernández-Lobato J. M., Aguilera-Iparraguirre J., Hirzel T. D., Adams R. P., Aspuru-Guzik A. Automatic chemical design using a data-driven continuous representation of molecules, <https://arxiv.org/pdf/1610.02415.pdf>
3. Kusner M.J., Paige B., Hernández-Lobato J. M., Grammar Variational Autoencoder, <https://arxiv.org/pdf/1703.01925.pdf>
4. Gonczarek A., Tomczak J., Zaręba S., Kaczmar J., Dąbrowski P., Walczak M., Learning Deep Architectures for Interaction Prediction in Structure-based Virtual Screening, *NIPS Workshop 2016*)
5. Gonczarek A., Tomczak J., Zaręba S., Kaczmar J., Dąbrowski P., Walczak M., Interaction Prediction in Structure-based Virtual Screening using Deep Learning, *Computers in Biology and Medicine* [w recenzji]