

多視点画像と CNN を用いたバーチャルスクリーニング

釘本 晃佑[†] 古田 諒佑^{††} 入江 豪^{†††} 青木 伸^{††††} 谷口行信[†]

[†] 東京理科大学工学部情報工学科

^{††} 東京大学生産技術研究所

^{†††} NTT コミュニケーション科学基礎研究所

^{††††} 東京理科大学薬学部

1. はじめに

新規治療薬の開発現場では、化学実験の時間的、人的コスト削減のため、標的タンパク質に結合する化合物(リガンド)を計算により絞り込むバーチャルスクリーニングが注目されている。本稿では、3次元物体認識技術を応用することによるバーチャルスクリーニングの高速化について検討する。

バーチャルスクリーニングは2段階の処理から構成される: (i)標的タンパク質とリガンドの結合部位(ポケット)を検出する, (ii)ポケットに対してリガンドがどのようなポーズで結合するのかを化学的に計算する(ドッキングシミュレーション)。しかし、ドッキングシミュレーションには、膨大な時間がかかるという問題がある。本研究は、ドッキングシミュレーションを、ポケットと類似した3次元形状を持つ候補リガンドを探す3次元物体検索問題とみなしてその高速化を図る。

2. 多視点画像ベースのバーチャルスクリーニング

提案法は2段階の処理で構成される: (1)特徴抽出器の学習, (2)リガンド検索。タンパク質のポケット検出には、モルフロジー演算を応用した手法[2]を用いた。

2.1 特徴抽出器の学習。特徴抽出器は、多視点画像を用いて3次元物体認識を行うMVCNN[1]を応用したものである。学習フローを図1に示す: (i)タンパク質のポケット, リガンドの3次元モデルを囲む正十二面体を考え、各頂点に配置した仮想カメラで撮影した多視点画像を生成する。(ii)多視点画像をCNNに入力して特徴マップを算出する。(iii)特徴マップを最大プーリングしたものを特徴ベクトル x, y とする。(iv)特徴ベクトル x, y 間のコサイン類似度によりポケット, リガンドの結合親和性(類似度)を予測する。この類似度が実際に結合するポケット, リガンドのペアであれば出力が1に、そうでなければ-1に近づくようにCNNの重みを学習する。CNNは、ImageNetで学習済みのAlexNetをファインチューニングする。

2.2 リガンド検索。標的タンパク質に結合するリガンドを検索する処理のフローを図2に示す。前節で学習したモデルのうち、View poolingまでを特徴抽出器FeatureExtractor1, 2とする。標的タンパク質のポケットをFeatureExtractor2に入力することでポケットの特徴ベクトルを抽出する。一方、あらかじめリガンド候補からFeatureExtractor1を用いて抽出した特徴ベクトルをデータベースに保存しておく。ポケッ

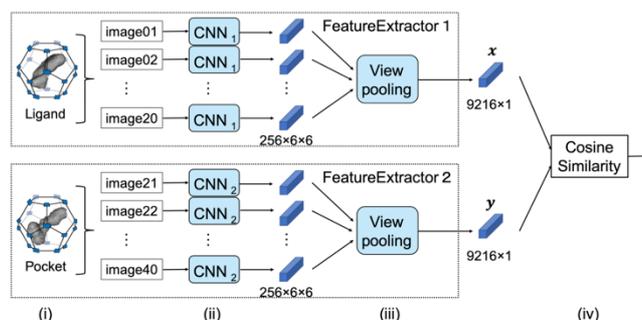


図1 提案手法の特徴抽出器の学習フロー

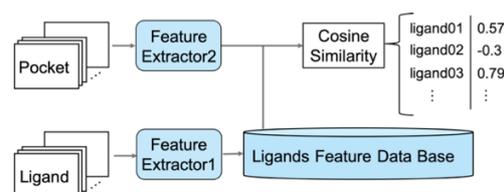


図2 提案手法のリガンド検索フロー

表1 リガンド検索の精度と処理時間

	平均 Top 10 正解率	平均処理時間(s)
比較手法[3]	46.5%	2906
提案手法	32.7%	8

トとリガンドの特徴ベクトルのコサイン類似度が大きい順に出力する。

3. 実験結果

58種類のタンパク質について、58種類のリガンドを対象に検索を行った実験結果を表1に示す。学習データとして50種類のタンパク質とリガンドを用い、15エポックの学習を行った。比較手法には、公開ソフトウェア AutoDock Vina[3](<http://vina.scripps.edu/>)を用いた。平均 Top10 正解率は比較手法[3]より低くなったが、平均処理時間については大幅に短縮した。

4. 今後の課題

今後は、タンパク質のポケットをより高精度に検出し、画像データの質の向上による検索精度の向上を目指す。

参考文献

- [1] Su, H., *et al.*, ICCV, vol. 1, pp. 945–953, 2015.
- [2] Kawabata, T., *Proteins*, vol.78, no.5, pp.1195–1211, 2010.
- [3] Trott, O., *et al.*, *Computational Chemistry*, vol. 31, no.2, pp.455–461, 2010.