

論文内容の要旨

博士論文題目 機械学習を用いた説明可能なアクティビティ・クリフ予測手法の開発

Developing explainable machine learning models for activity cliff prediction

氏 名 田村 峻佑

(論文内容の要旨)

創薬開発における低分子構造最適化過程において、分子構造における側鎖や一部の官能基などを構造変化させた時に、その分子の活性が劇的に変化する「アクティビティ・クリフ (Activity Cliff, AC)」と呼ばれる現象が報告されている。本学位論文では、AC を分子構造から予測するモデル開発と予測モデル解釈手法を提案している。

本学位論文は、第1章から第4章で構成されており、第1章では標的マクロ分子の立体情報を使用しない「リガンドベース」からのAC予測についての既往の研究と標的タンパクの立体情報を利用する「ストラクチャーベース」のAC予測についてレビューをし、研究スコープを明確にしている。加えて、入力となる化学構造が類似し、出力である活性が大きく変化する機械学習が苦手とするACに対し予測を可能にした先行研究の分子表現と機械学習手法を紹介し、その限界を議論する形で本論文の背景と目的を述べる。

第2章では機械学習を用いて幅広いターゲットに適用可能な高精度なAC予測手法を検討する。公共データベースから抽出した100種類の標的タンパク質に対するAC予測精度を計測した。ここでは、入力表現と方法論的複雑さが異なる様々な機械及び深層学習を用いた9種類のAC予測モデルを検討した。結果として、サポートベクターマシンや単純な最近傍法によるモデルが高精度となった。深層学習によるモデルの精度は十分ではなかった。

第3章ではAC予測モデルの解釈方法の検討と、解釈のための新規手法を考案する。サポートベクターマシンを用いたモデルに用いるカーネル関数の出力を、各部分構造の有無に対する重み付き線形和として近似する手法を提案した。さらにその結果を予測対象の化合物ペアの化学構造上に可視化する方法を提案する。本解釈手法を適用することでそれら重要な部分構造を解釈し得ることを示した。

第4章では、本研究で提案したAC予測及び解釈手法の限界と、今後の展望について述べる。特に、モデル構築に利用できるデータ数が数百程度と深層学習を利用するためのデータ数が少ない状況下で、別の標的マクロ分子に対しての活性データを利用する転移学習等の手法を適用する展望を述べている。

(論文審査結果の要旨)

創薬開発における低分子構造最適化プロセスにおいて、分子構造における側鎖や一部の官能基などを構造変化させた時に、その分子の活性が劇的に変化する「アクティビティ・クリフ」と呼ばれる現象が報告されている。この現象は、「類似の分子構造は類似の活性を示す」という類似性の原則に反するため、合理的な構造（合成）展開の妨げとなっている。その一方で、標的とするマクロ分子に対して鍵となる相互作用を表しており、「構造活性相関」に関する情報を多く有する。田村 峻佑氏は、アクティビティ・クリフを分子構造から予測するモデル開発と予測モデル解釈に関する研究を提出した。

本論文では、アクティビティ・クリフを分子構造から予測する機械学習モデル開発と大規模データセットを用いた予測精度の検証、構築したモデルを解釈するためのアルゴリズム考案をおこなっている。本論文の主要な結果は以下の通りである。

1. 大規模データセットを利用したアクティビティ・クリフ予測モデル構築手法の検証。公共データベースから取得した100種類の標的マクロ分子に対する活性低分子データから、アクティビティ・クリフ予測モデルを構築し、様々な分子表現とモデリング手法を検討した。結果、サポートベクターマシンと「骨格と側鎖に適応したカーネル関数」の組み合わせが安定して高い予測精度を発見した。また、データ数と予測精度の高さには相関関係が見られず、アクティビティ・クリフ予測は「データ数が足りないので精度が低い」という問題ではないことをレトロスペクティブに実証した。現在までのアクティビティ・クリフに関する関連研究では、二、三種類の標的タンパク質に対して新規なモデリング手法を報告するケースが多く、本研究成果が、アクティビティ・クリフ予測アプローチに対するベンチマークとして利用されることが期待される。

2. アクティビティ・クリフ予測モデルの解釈手法の開発。分子構造のペアから予測した活性の大きな差を解釈することができれば、活性を発現するために重要となる側鎖（もしくは分子骨格）を特定することができる。特定した部分構造は構造展開における重要な部分として利用することができる。しかし、サポートベクターマシンと「骨格と側鎖に適応したカーネル関数」の組み合わせによるモデルでは、積の形のカーネル関数を利用しているため、既往の解釈手法を適用することができない。田村 峻佑氏は主骨格と側鎖の積からなるカーネル関数の出力値を当分割することにより、モデルの予測値を分子構造に反映させる手法を新規開発した。開発した手法は、複数の分子ペアに対して適用され、既往の共結晶構造から推測される相互作用を表現したものであるとの確認がなされていた。

以上、本論文ではアクティビティ・クリフの予測や解釈手法の構築などについてまとめられている。学術的にも大きな意義があり、審査委員一同は本論文が博士（工学）の学位論文として価値のあるものと認めた。