

先端科学技術研究科 修士論文要旨

所属研究室 (主指導教員)	計算システムズ生物学 (金谷 重彦 (教授))		
学籍番号	2311078	提出日	令和 7年 1月 20日
学生氏名	奥野 智也		
論文題目	酵素インフォマティクスに基づく反応予測最適化のためのタンパク質言語モデル の特徴量の開発		
要旨			
<p>大規模なアミノ酸配列データセットを用いて事前学習されたタンパク質言語モデルの発展に伴い、酵素-基質相互作用の予測問題において、タンパク質言語モデルを応用する研究が進められている。このようなタンパク質言語モデルの利用は、特定の酵素や酵素ファミリーの枠を超えた、より一般的な酵素化合物間相互作用の理解を予測モデルにもたらす可能性がある。しかし、実験データを基にした酵素-基質相互作用の予測問題ではデータに含まれる酵素の多様性が乏しいことが多く、タンパク質言語モデルを追加の調整なしで適用した場合、過学習を引き起こす懸念がある。本研究では、事前学習済みのタンパク質言語モデルから、教師データが限られた酵素-基質相互作用の予測問題において有用な文章レベルの特徴量を獲得することを目的とし、多様な酵素のアミノ酸配列とその酵素が触媒する反応のデータを用いた自己教師あり学習による特徴表現学習タスクを提案する。3種類の酵素活性スクリーニング実験データセットを用いた活性分類タスクでの評価実験の結果、提案手法を適用した言語モデルで生成した埋め込みは、特定の酵素データセットにおいてのみ、活性部位の情報を用いた既存のプーリング手法を上回る予測性能を示すことを確認した。この結果から、中間タスクと下流タスクに含まれる配列の同一性が下流タスクの性能に影響する可能性を指摘し、今後の研究の方向性を示した。</p>			