

2023 年度 大阪大学蛋白質研究所 拠点事業

## 研究成果報告書

(1) 事業名 (下記より該当事業名を選択し、ほかは削除して下さい。)

共同研究員

(2) 研究代表者

氏名：村上洋一

所属機関名・部局名・職名：東京情報大学・総合情報学部総合情報学科・準教授

(3) 研究課題名 (申請時に記載したものと同一課題名を記入して下さい。)

ゲノムスケールでの蛋白質間相互作用予測

(4) 蛋白質研究所受入担当教員

教員名：水口賢司 (研究室名：計算生物学研究室)

(5) 研究成果の概要 (公開)

\*背景および目的, 方法と結果, について, 公開して差し支えない範囲で 1 ページ以内で記載.

生体内における蛋白質間の相互作用 (Protein-Protein Interactions, PPI) ネットワークを明らかにすることは, 生命現象をより深く理解する上で重要であるが, 過渡的現象のような物理化学的な要因, 翻訳後修飾, 実験条件等によって, 他の蛋白質との相互作用を実験的に明らかにすることが困難な蛋白質もある. そのような課題に対して, 近年, 自然言語処理の埋め込み法を, PPI 予測に応用する研究が報告されている. 埋め込み法とは, 単語や文などの自然言語の情報を保持しつつ, コンピュータが扱える数値に変換する手法である. すなわち, 蛋白質の配列を文字列として扱い, その文脈情報を保持しつつ変換されたベクトルを用いて, PPI を予測する手法が提案されており, その有効性が報告されている.

しかしながら, 蛋白質の配列の文脈情報が, 埋め込み法によってどの程度保持されているのかについて十分な解析がされていない. そこで, 文の埋め込み手法である, Doc2Vec によって変換された蛋白質のベクトルと構造との関係性について解析を行った. その結果, 蛋白質のベクトル類似度と構造類似度との間に弱い相関 ( $-0.577$ ) があることがわかった. また, 蛋白質の埋め込みモデルと, 蛋白質の進化的な情報を組み合わせた PPI 予測手法は提案されていないことから, Doc2Vec と位置特異的なスコア行列 (PSSM) を組み合わせた PPI 予測手法を開発した. その結果, Doc2Vec に基づく PPI 予測モデルと PSSM に基づく予測モデルのそれぞれの分類結果を, 単純な voting 法によるアンサンブル学習によって予測を行った結果, それぞれの予測モデルよりも高い性能 ( $F$  値 =  $85.9\%$ ,  $MCC = 0.71$ ) を示すことができた. さらにベンチマークテストでは, 先行研究よりも高い性能を示すことができた. 以上のことから, Doc2Vec を用いて変換された蛋白質の埋め込み表現は, 配列から構造の文脈情報を効果的に捉えている可能性が示唆された. また, Doc2Vec に基づく埋め込みモデルと PSSM を組み合わせることで, PPI 予測性能の向上が図られることが示唆された.