

ダイバーシティ事業 国際共同研究PI養成プログラム
報告書

報告日：2020年3月5日

派遣者所属名	システム情報学研究科
派遣者氏名	全 昌勤
研究タイトル	計算知能学及び健康情報システムへの応用
研究目的	<p>本研究目的は、統計学や深層学習などの機械学習に関する計算知能（CI）の方法論を研究し、ビッグデータから有用な知見を発見する基礎理論を構築して、健康・医療分野への応用を試みることである。具体的な応用課題として、PubMedから生物医学に関係論文の概要を取り出し、テキストマイニングを用いて関連性のあるタンパク質名を抽出して、各タンパク質間の関係性を明らかにする。</p>
研究報告	<p>本研究は、深層学習を代表とする機械学習に関する計算知能（CI）の方法論を探索し、ビッグデータから如何に有用な知見を発見するかに関する基礎理論を構築して、健康・医療分野への応用を試みる。</p> <p>生物医学文献は大量に発表されている今日、単に人の手で処理することで文献から各タンパク質間の相互関連（PPI）を発見することは極めて困難となってきた。そのため、生物学データベースの更新が遅れている。PPI抽出の最新のテキストマイニング手法は、主に深層学習（ディープラーニングDL）モデルに基づいている。DLベースの手法の学習結果は、主にDLモデルのアーキテクチャや特徴の埋め込み手法に大いに影響を受ける。</p> <p>本研究では、畳み込みニューラルネットワーク（CNN）、長期短期記憶（LSTM）、およびハイブリッドモデルを含むDLモデルなどのさまざまなアーキテクチャを比較し、PPI抽出のための双方向LSTM + CNNモデルのハイブリッドアーキテクチャを提案した。ここでは、学習モデルで特徴情報をより正確に抽出できるように、事前学習済みの単語埋め込みと最短依存パス（sdp）埋め込みを統合的に扱うようにしている。計算実験の結果として、PPI抽出に関して本研究で提案されたモデルは、非ハイブリッドDLモデルおよびハイブリッドCNN + 双方向LSTMモデルよりも優れていることを示した。そして、本手法では、特徴の次元を減らし、視覚化することが可能となり、異なるDLモデルで学習された特徴の比較により、提案モデルの有効性が確認された。</p>
今後の研究の見通し	<p>提案されたハイブリッドDLモデルをそのほかの自然言語処理（NLP）タスクに適用してその有用性を調査し、より複雑なネットワークアーキテクチャ（例：より多くの層を持つディープハイブリッドモデル、またはディープ強化学習モデル）を検討して、学習性能の向上を一層探求する。</p>
研究成果の発表予定	<p>・筆頭著者として国際共著による学術論文や国際会議での発表予定 2020.4 論文「A Hybrid Deep Learning Model for Protein-protein Interactions Extraction from Biomedical Literature」を学術雑誌</p>

	「International Journal of Advanced Intelligence」に投稿予定
--	---

海外派遣終了後の研究の進捗状況（2020年3月現在）

帰国後も共同研究者と連絡をとり、来年3月に開催する予定の国際シンポジウムについて打ち合わせを行っている。また、シンポジウムのテーマに関連する研究に取り組んでいる国際文化科学研究科の教員および韓国の研究者にも協力を依頼し、シンポジウム案について協議を行っている。1月に韓国・釜山で開かれた国際シンポジウムでは、別の韓国の研究者にも同シンポジウムについて打診し、これからシンポジウムの詳細を固めたうえで参加を求める予定である。また、派遣期間中に執筆した論文を学会誌（『美学』）に投稿した（現在査読進行中）。