

## 論文内容の要旨

博士論文題目 自然言語と生物配列のための再帰的ニューラルネットワーク

氏名 椿真史

(論文内容の要旨)

近年、画像認識、自然言語処理、バイオインフォマティクスなどの様々な研究分野において、機械学習は最も重要な技術の一つとなっている。本論文では、(i) 自然言語やタンパク質などの系列データを扱うことに焦点を当て、(ii) 自然言語処理やバイオインフォマティクスにおける幾つかの問題を解決する機械学習手法を提案する。

特に近年、深層ニューラルネットワーク (DNN) が、音声認識や機械翻訳などにおいて高い性能を達成している。本論文では DNN の中でも主に、再帰的ニューラルネットワーク (RNN) を用いる。RNN は、自然言語やタンパク質などの任意長の系列データに適した DNN である。そして我々は、RNN をベースとしたモデルを用いて、自然言語処理における「文の意味構成」とバイオインフォマティクスにおける「タンパク質構造予測」の問題を解決する。

提案法で我々はまず、long short-term memory (LSTM) を持つ RNN を用いる。LSTM は、系列データ内の長期的な依存関係の情報を保持することができる。我々は LSTM が、上記の自然言語とタンパク質の両問題に対して効果的であることを示す。そして我々は LSTM の上に、自然言語やタンパク質のデータが持つ様々な性質を捉えるために、カーネル法、畳み込みニューラルネットワーク、アテンションモデルなど他の機械学習技術を組み合わせたモデルを提案する。実験では、様々な既存研究と比較して、提案法の有効性を実証する。

氏名	椿真史
----	-----

(論文審査結果の要旨)

平成28年12月19日に開催した公聴会の結果を参考に平成29年2月15日に本博士論文の審査を行った。以下のとおり、本博士論文は、提案者が独立した研究者として、研究活動を続けていくための十分な素養を備えていることを示すものと認める。

椿真史は、本博士論文において、再帰ニューラルネットワークに基づいて、自然言語表現の非構成的な意味の問題、および、蛋白質の構造予測問題を解く手法を提案した。また、それらの手法を既存のデータに適用し、性能評価によりその有効性を示した。本論文の貢献は、以下のようにまとめることができる。

1. 自然言語表現の意味は、個々の単語の意味から全体の意味を構成的に得ることができない場合があり、そのような場合に非構成的な意味をどのように求めるかが困難な問題になっている。本論文は、再帰ニューラルネットワークとカーネル法を組み合わせた手法を提案し、文の意味的な類似性を予測する問題において高い性能が達成可能であることを示した。
2. 蛋白質の折り畳み構造予測問題に対して、LSTMを用いた手法を2つ提案し、ベンチマークデータに対して高い性能を示した。
3. 蛋白質の立体構造予測、特に、塩基配列間の近接関係の問題に対して、LSTMとCNNを組み合わせた手法を提案し、細かい素性選択なしに有効な結果を達成可能なことを示した。

再帰ニューラルネットを利用し、自然言語文の非構成的意味の問題と蛋白質の構造解析に対して有効な結果を示した本研究は、独創性が高く、実用的であり、系列データに対する意味的な構造解析に関する機械学習分野において高い貢献があると評価する。

よって、本論文は、博士(工学)の学位論文として価値あるものと認める。