

論文審査の要旨

博士の専攻分野の名称	博 士 (情報科学)	氏名	唐 惠東
学位授与の要件	学位規則第 4 条第 1 項・2 項該当		
論 文 題 目 Sequence Studies for Text Data Augmentation and Molecular Generation (テキストデータの拡張と化合物分子の人工生成のためのデータベースシーケンスの研究)			
論文審査担当者 主 査 教 授 森本 康彦 審査委員 教 授 江口 浩二 審査委員 准教授 亀井 清華 審査委員 教 授 ALAM, Kazi MD. Rokibul (Khulna University of Engineering & Technology)			
〔論文審査の要旨〕 本論文（本研究）は、 <ul style="list-style-type: none"> ・テキストデータのロバストネスを強めるためのデータ拡張技術 ・創薬に利用することを想定した化合物分子の人工生成技術 に関する研究成果をまとめたものである。分野の異なる研究成果の様に見えるが、両者ともにデータベースシーケンスを扱う人工知能技術という共通点を持ち、技術としての本質はシーケンスデータの知能科学であり、本論文は成果をその視点でまとめたものである。 第 1 章では、テキストデータや分子構造の線形表現である Simplified Molecular Input Line Entry System (SMILES) などのシーケンスデータとして蓄積されるデータと、そのデータを利用した知能科学の概要とその重要性を論じた。 第 2 章では、本論文内で利用される敵対的生成ネットワーク (GAN)、強化学習 (RL)、長短記憶 (LSTM) などの人工知能技術や関連する用語や定義について概説した。 第 3 章では、機械学習における学習データとして利用できるテキストデータの量を増やすための技術について提案した。 テキストデータに関連する代表的な問題の一つにテキスト分類がある。これはテキストを 1 つ以上のクラスに分類することを目的としているが、分類結果は、テキストに対し軽微な摂動操作を加えることにより容易に変わるなどロバストネスの面で大きな課題が残っている。本論文では、この課題に対処するため、同義語、反意語、および感情的に関連する単語を組み合わせた単語置換に基づく 3 つのデータ拡張方法を提案した。我々は、敵対的攻撃手法を使用して 4 種類のデータセットで提案手法を評価し、その堅牢性の向上を検証した。			

第4章では、創薬に利用することを想定した化合物分子の人工生成技術を提案した。

化合物の分子構造は、SMILES と呼ばれるテキストデータのような記号系列で表現することができる。本研究では、SMILES 系列を人工生成する手法での化合物分子の生成手法を提案した。テキストデータの生成手法を単純に適用すると SMILES の持つべき性質との互換性を失ってしまう上に、時間もかかってしまう。本研究では、情報エントロピーの最大化とともに、モーメント報酬とグローバル報酬を組み合わせた効率的な報酬関数を導入することで、問題を解決し、広範な評価実験により、その有効性を確認した。

最後に、第5章で、シーケンスデータの知能科学としての本研究の貢献をまとめた。

口頭試問において、審査委員から、「評価指標の妥当性」、「実験結果の評価」、「提案した手法のロバストネス」などに関して詳しい説明を求められ、申請者はそれぞれに対し、的確かつ丁寧に回答することができた。

以上、審査の結果、学位申請者は博士（情報科学）の学位を授与されるに十分な資格があるものと認められる。

備考：審査の要旨は、1,500 字以内とする。