

厚生労働科学研究費補助金
(政策科学総合研究事業 (臨床研究等 ICT 基盤構築・人工知能実装研究事業))
分担研究報告書

課題名 : ツール物質効率的探索のためのアルゴリズムの開発と実装
研究分担者名 : 田部井 靖生
国立研究開発法理化学研究所 革新知能統合研究センター
圧縮情報処理ユニット ユニットリーダー

研究要旨

膨大な数の化合物の構造データや実験データを効率的に処理するアルゴリズムを開発する。代表者が整備する化合物の大規模データを高速かつメモリ効率良く処理するため簡潔データ構造の技術の高度化を行う。また機械学習の手法とデータ圧縮技術を組み合わせて、大規模な化合物データから機械学習の予測モデルを高速に学習するためのアルゴリズムの実装を行う。特発性肺線維症や肺がんの創薬標的分子に対して、膨大な化合物のインシリコスクリーニングを行うための予測モデルの学習の際に、実装したアルゴリズムを用いる。

A. 研究目的

山西グループとの共同研究のため、これまで開発した大規模化合物データベースの類似度検索アルゴリズム b-Bit Sketch Trie (bST) と大規模タンパク質配列分類アルゴリズム SFMEDM の実装の改善を行った。

B. 研究方法

bST と SFMEDM は C++ 言語により実装されている。プログラムを効率化し、メモリーリークをなくすとともに実行速度の改善を行う。

(倫理面への配慮)

人及び動物を研究対象としていない研究であるため、倫理面の問題はないと判断した。

C. 研究結果

メモリーリークや速度面の改善を行うことができた。

D. 考察

プログラムは改善をすることができたが、大規模データセット上で効率的に運用するためには、さらなるプログラムの効率が必要と考えられる。

E. 結論

今後は、プログラムのさらなる効率化を行う。

F. 研究発表

1. 論文発表
なし
2. 学会発表
なし

G. 知的財産権の出願・登録状況 (予定を含む。)

1. 特許取得
なし

2. 実用新案登録
なし
3. その他
なし