

プロジェクト名(タイトル): 深層学習などによる NMR データおよび生体高分子の解析

利用者氏名: 小林直宏

理研における所属研究室名: 生命機能科学研究センター先端 NMR 開発・応用研究チーム

1. 本プロジェクトの研究の背景、目的、関係するプロジェクトとの関係

HOKUSAI のような大規模 PC クラスタの利用は、NMR のデータ解析の自動化および高度化、深層学習ツールの開発などへの応用を目的とした。現状では NMR 解析法は他の構造解析法として X 線結晶解析やクライオ電顕解析などに比べて大きく遅れを取っている。研究代表者は HOKUSAI に搭載されている 40~200 Intel-Core を利用することで大規模な深層学習を試験的に N 実行し、その一方で NMR 信号解析の自動化を多数実行した。

2. 具体的な利用内容、計算方法

NMR 信号自動帰属ツールである FLYA (40-core), 分子動力学計算ツールである Amber (20-40 core), NMR 信号データより構造をモデリングする CS-ROSETTA (40-200 core), 自身で開発した構造から化学シフトを予測する深層学習器の試作版を実行利用した。HOKUSAI には GPU が搭載されていないため、深層学習訓練計算は自身で構築した GPU workstation で実行し、大規模ストレージとコア数を必要とする計算を HOKUSAI によって実行可能であった。

3. 結果

構造から化学シフト予測する方法論は以前より公開されており、既存のツールでも浅めの機械学習は、SPARTA+, UCFS-Shift でもある程度の実現はされている。代表者により開発された新規深層学習器によるベンチマークはまだ微妙なラインであり、特に大規模にデータを収集できたタンパク質主鎖信号に関して一定の成果を得た。また、従来法では実現できていなかったタンパク質側鎖信号や RNA など核酸といった高分子への応用の実現性を示すことができた。

4. まとめ

代表者により開発された深層学習器の予測性能は既存のツール、SPARTA+, UCFS-Shift との大きな差別化には成功しなかったが、収集できる実験データの規模とクオリティ次第で十分な性能を発揮できることを示せたと言える。現在、関連アルゴリズムに関する論文を鋭意執筆中である。

5. 今後の計画・展望

代表者により開発された深層学習器は、改良の余地はあるものの搭載されている複数のアルゴリズムやルーティンの新規性は高いものと考えられる。その理由から単なる化学シフト予測のみならず、NOE 強度マッピングなど相対的な原子間距離依存による NMR 実験データへの自動解析も含め、多岐にわたる応用も可能と考えており、継続的な開発を続けたい。

6. 利用がなかった場合の理由