

(Graduate School of Science, Kyoto University) ○Yuki Yamamoto

RMSD は化学構造類似性の最も一般的なメトリックの一つである。化学構造類似性は化学反応性や物理特性、生物活性に関連し、データ化学において重要である。RMSD は広く普及しているが、原子マッピング最適化を伴う RMSD は未だ原子数に対して多項式時間では解けない難しい問題である。我々は分子多量体に対して定式化された厳密な RMSD (図 a) を効率的に求める分枝限定法アルゴリズム mobbRMSD を紹介する。mobbRMSD は分子内の原子数に関係なく、分子数に対して平均指数時間で実行できる (図 b)。更に我々は mobbRMSD に基づいた構造クラスタリングアルゴリズムを紹介する。この方法はデータ数の増加に従って漸近的に平均多項式時間のアルゴリズムであり、分子数に対する指数スケールアップの問題を解消する。

Figure 1 is a log-linear plot showing the scaling of the proposed algorithm. The y-axis represents the wall-clock time in seconds per run, ranging from 10^{-5} to 10^5 . The x-axis represents the number of molecules, ranging from 2 to 20. Three data series are plotted: mobbRMSD (Exact) (orange squares), IsometryOpt (Exact) (purple squares), and LP-RMSD (Heuristic) (black crosses). The LP-RMSD series shows a constant wall-clock time of approximately $10^{-4.5}$ sec/run. The mobbRMSD and IsometryOpt series show an increasing wall-clock time, with mobbRMSD reaching approximately $10^{4.5}$ sec/run and IsometryOpt reaching approximately $10^{4.2}$ sec/run at 20 molecules. A legend indicates that the orange square represents the upper quartile, the purple square represents the median, and the black cross represents the lower quartile. A shaded region around the median line indicates the interquartile range.

