

デジタルアニーラによる分子類似度評価を用いた高精度クラスタリング

High-precision clustering using molecular similarity evaluation with Digital Annealer

富士通 [○]丸尾 昭人, 實宝 秀幸

Fujitsu LIMITED, [○]Akito Maruo and Hideyuki Jippo

E-mail: maruo.akito@fujitsu.com

構造類似度が高い物質は特性も類似すると期待されるため、化合物の探索において、分子構造の類似度に着目した手法がよく用いられる。我々は 2 分子間の最大共通部分構造を組合せ最適化問題として、デジタルアニーラで解き類似度を評価する方法を提案してきた。同じ種類の香りを持つ香料分子の抽出問題においてデジタルアニーラを用いた類似度評価手法は、従来のフィンガープリントを用いた手法より、精度よく目的の分子を抽出できることを示した[1]。しかし、本手法では、クラスタリングの精度が十分でないという問題があった。そこで、今回は記述子の相対誤差の情報をを用いることで、目的特性に対するクラスタリング精度を向上できるか検証した。

本研究では、balsam、citrus、vanilla の香りを持つ 29 個の香料分子を対象とした。まず、デジタルアニーラを用いてそれぞれの分子間の構造類似度を計算した。次に、RDKit を用いて 208 個の 2 次元記述子を計算し、ランダムフォレストの変数重要度に基づく変数選択手法である Boruta と相関分析を用いて記述子を 22 個まで絞込み、残った重要記述子に対して記述子の相対誤差を計算した。そして、記述子の相対誤差を重み付けして構造類似度に加えることで、新たな類似度評価値とした。ここで、記述子の重みは、クラスタリング精度の指標である調整ランド指数とシルエットスコアが最大となるようにベイズ最適化を用いて定めた。算出した類似度を距離に直し、多次元尺度構成法を用いて分子を 2 次元マップ上に可視化した結果を図 1 に示す。フィンガープリント類似度と比べ、デジタルアニーラ類似度のみを用いた従来法は、クラスがはっきりと分かれているが、balsam と vanilla の分類ができていなかった。一方で、提案手法は balsam と vanilla の分類もできていることがわかる。提案手法において最も重要視された記述子は PEOE_VSA11(ある所定の範囲に収まる電荷を持つ原子の表面積の和)であり、分子の構造情報に加えて、静電相互作用の情報まで考慮したため、提案手法では香料分子を高精度にクラスタリングできたと考えられる。

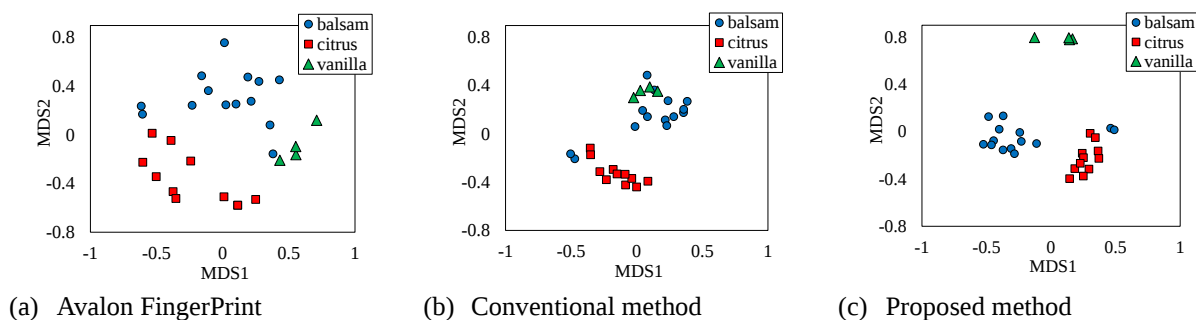


Fig. 1: Multidimensional scaling results

[1] 寺島千絵子ほか, 第 80 回応用物理学会秋季学術講演会 講演予稿集, 19a-B01-4, 2019.