

物性最適化のためのデータ駆動型有機分子設計

Data-Driven Organic Molecule Design for Optimization of Physical Properties

岡山大¹, 広工大² ○ 佐藤 郷介¹, 大村 訓史², 鶴田 健二¹Okayama Univ.¹, Hiroshima Inst. Tech.² ○ Kyosuke Sato¹, Satoshi Ohmura², Kenji Tsuruta¹

E-mail: tsuruta@okayama-u.ac.jp

近年, マテリアルズ・インフォマティクスと呼ばれる情報科学を用いて望ましい物性を持つ材料の探索を効率的に行う手法が注目されている. このようなデータ駆動による材料探索技術は新規材料の発見や, 材料の機能向上のために重要である. 我々は機能性有機分子の効率的な探索を目指している.

分子の構造式は SMILES や MOL 等の形式で表現することができる. しかし, 構造式から直接的に計算することができるのは logP 値の予測や drug-likeness 等の構造式の部分構造の有無で評価される値である. より様々な物性について新規分子を設計するためには分子の立体構造が必要となる. 近年, SMILES 記法で記述された分子構造の文字列を距離幾何学法で簡易的に原子配置を計算し, 経験的な力場や第一原理計算を用いて構造や物性値を計算して, スクリーニングによる分子設計を行うことが可能であることが, Rafael らにより示されている [1]. スクリーニングでは得られる分子がデータベースに依存するため, 本研究では SMILES を最適化するというアプローチをとることにした. しかし, SMILES 記法のような離散表現では組み合わせ最適化問題となるため分子設計が困難である. そこで, Variational Autoencoder(VAE) という機械学習法を用いることで, SMILES を連続変数へ変換する手法 [2] に着目した. この手法を用いることで, 物性値に関する組み合わせ最適化問題から, 連続な多変数の最適化問題へ置き換えることができる. VAE はエンコーダとデコーダの2つのニューラルネットで構成される. 学習は, エンコーダが SMILES を連続変数へ変換し, デコーダが連続変数を元の SMILES へ変換するように行われる. 学習後のエンコーダとデコーダを用いて, 図1に示す計算を行う. 例として, 分子の重心から各原子へ向かうベクトルの総和のノルムの最小化を行った結果得られた分子を図2に示す. このとき, SMILES は 56 変数の実数で表されており, ランダム探索を行った. 講演では, 物性探索の例も報告する.

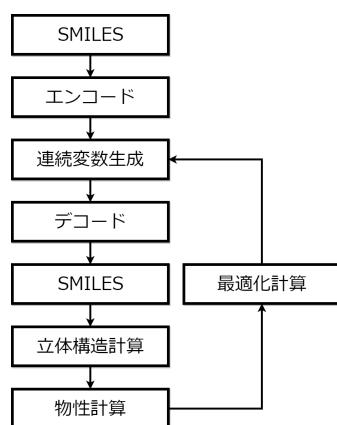


図 1: データ駆動型分子探索

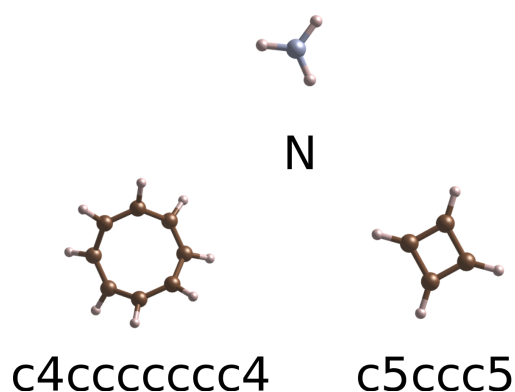


図 2: 最適化で得られた分子

参考文献

[1] Nat. Mater. 2016, **15**, 1120-1127[2] ACS Cent. Sci. 2018, **4**, 268-276