# 拡散確率モデルによる分子グラフ構造の潜在変数表現の学習

氏　　名　高下 大貴

研 究 室 名　計算システムズ生物学研究室

主指導教員名（論文博士の場合は推薦教員名）金谷 重彦

内 容 梗 概（1ページ目に収めること）

ケモインフォマティクスにおける深層学習や機械学習を用いた分子構造の探索および分子の特性予測において，化学構造を表現する適切な記述子を設計することは，非常に重要な課題である．近年では，Graph Neural Networkのような深層学習モデルによって，化学構造式をグラフ化した分子グラフから，その潜在変数表現として記述子を設計するアプローチが盛んに研究されている．これらのアプローチは数万件以上の教師データ(分子特性と分子グラフが対になったデータ)を学習に必要とするため，現実の問題に実用的な方法と言えない．

Variational Autoencoder (VAE) のような深層生成モデルは，潜在変数の学習に入力データ(分子の化学構造)のみを使用するため，深層学習を用いた分子記述子の設計において非常に有効なアプローチであると考えられる．しかし，これらのアプローチは潜在変数空間上での分子特性の滑らかさについて十分に考慮しておらず，学習した分子グラフの潜在変数同士の近さが，分子の特性の近さを保証するとは限らない．

この研究では，分子グラフの潜在変数空間上での分子特性の滑らかさを評価する指標として Widely applicable Bayesian Information Criterion (WBIC)を導入し，分子の特性が滑らかに配置されるような潜在変数空間を設計する新たな深層生成モデルの構築を試みた．我々が提案するこのモデルは，Transformer を用いたGraph Autoencoder と拡散確率モデル(DDPM)を統合した深層生成モデルである．

実際に，学習した分子の潜在変数表現を用いて HOMO エネルギーなどの量子化学的な特性，水溶性などの物理化学的特性，活性などの生化学的特性に対する滑らかさを WBIC によって評価したところ，提案手法は VAE や Normalizing flow を用いた従来の深層生成モデルと比較して，滑らかな潜在変数空間をモデル化できていることが分かった．種類の異なる分子特性同士は，何かしらの相関関係を持つことが多いため，本実験で評価した物性値や活性値以外にも，提案手法は滑らかな潜在変数表現を学習できていることが期待できる．