

## 論文内容の要旨

博士論文題目 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

(Smooth Embedding of Molecular Graphs with Graph Latent Diffusion Autoencoder)

氏 名 高下 大貴

(論文内容の要旨)

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、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。しかし、特定のタンパク質の活性などにデータセットを限定した場合、得られるサンプル数は高々数百件の場合が多いため、実用性のあるアプローチとは言えない。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深層生成モデルの構築を試みた。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。提案手法は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。

|     |       |
|-----|-------|
| 氏 名 | 高下 大貴 |
|-----|-------|

(論文審査結果の要旨)

令和6年1月12日に開催した公聴会の結果を参考に、令和6年2月5日に本博士論文の審査を実施した。本博士論文は、本学位申請者が、独立した研究者として研究開発活動が続けていくために必要な素養を備えていることを示すものである。本論文は、博士（工学）の学位論文としての価値があるものと認める。