

という課題があった。さらに、画像生成などでは、様々なバリエーションが求められ、必要に応じてGANの学習を行う必要があるため、高品質画像を高速に生成することは困難だった。しかし、2017年のICLR(International Conference on Learning Representations)では、 1024×1024 というメガピクセルレベルの画像を高速に生成するProgressive growing of GANs [7] が発表されている。Progressive growing of GANsでは、低い解像度から学習を始め、段階的に解像度を上げながら学習を行うことで学習時間を短縮した。

マテリアルズ・インフォマティクス／創薬への応用

化学分野では、統計学の応用として、化合物の特徴を表す記述子と活性値との関係を統計的にモデリングする定量的構造活性相関(Quantitative Structure-Activity Relationships; QSAR)が存在する。一般的なQSARは、化学構造から特徴を抽出した記述子SMILES(Simplified Molecular Input Line Entry System)を計算し、記述子と活性値との間に統計モデルを構築するという2段階のステップに基づいて行われる。構築した統計モデルを利用することで化学構造から活性値を予測することができる。

また予測の逆問題である合成問題、すなわち、望ましい活性を示すような化学構造を得ることはinverse-QSARと呼ばれる。

まず、inverse-QSARに、深層ニューラルネットを使う方法を示す。SMILESは文字列であるので、SIMLESを入出力として文字列の「変分オートエンコーダー(VAE)」を構築すれば、化合物構造から、「圧縮された特徴量」＝潜在変数を事前学習できる。そして、所望の活性値を持つ化学構造を得るために、学習した潜在変数特徴量と活性値との間で統計モデルを構築し、所望の活性値に対応する特徴量を逆算する。この特徴量を、VAEの事前学習で得られたデコーダーを通せば、SMILESで記述された、求めたい化学構造を得ることができる [8]。

もう一つのアプローチは、ベイズ推定による合成である。ベイズ推定による分子設計では化合物Sが原因、特性Yが結果、Uが望ましい特性とすると、確率分布pを考える、ベイズの式により、目標の特性($Y \in U$)をもつ、化合物Sに関する条件付き分布は、

$$p(S | Y \in U) \sim p(Y \in U | S)p(S)$$

と書き下すことができる(「 $A \sim B$ 」は、Aが確率分布Bに従うことを指す)。ここで $p(Y \in U | S)$ は予測特性と理想特性の差から計算でき、 $p(S)$ は、既存の化合物のデータから学習した確率モデル、化学式の場合は、化学構造の文字列表記SMILESのパターンを学習した確率言語モデルとして得ることができる。あとはベイズの式に従って $p(S | Y \in U)$ を、モンテカルロ法を使ってサンプリングすれば、目標の特性を持つ具体的な設計化合物を得ることができる [9]。

また、化合物のグラフ構造を直接入力としたグラフ畳込みニューラルネットワーク(Graph Convolutional Neural Networks)を導入して、化合物からの特徴量抽出に用いる方法もあり、得られた特徴ベクトルとタンパク質の結合を判別し、薬としての活性や副作用を予測することができる。このような道具立てがあれば、計算機上で、大量のグラフ(化合物のモデル)を生成し、効果の評価を繰り返すことで所望の新薬の候補を得ることができる^{※69}。

参考文献

- [1] Kingma, Diederik P., and Max Welling. "Auto-encoding variational bayes." arXiv preprint arXiv:1312.6114 (2013).
- [2] Goodfellow, Ian, et al. "Generative adversarial nets." Advances in neural information processing systems. 2014.
- [3] Hinton, Geoffrey E., and Ruslan R. Salakhutdinov. "Reducing the dimensionality of data with neural networks." science 313.5786 (2006): 504-507.
- [4] Zhu, Jun-Yan, et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks." Proceedings of the IEEE international conference on computer vision. 2017.
- [5] Xu, Tao, et al. "AttnGAN: Fine-grained text to image generation with attentional generative adversarial networks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.
- [6] A.Brock, "Large Scale GAN Training for High Fidelity Natural Image Synthesis", 2018, arXiv:1809.11096
- [7] Karras, Tero, et al. "Progressive growing of gans for improved quality, stability, and variation." arXiv preprint arXiv:1710.10196 (2017).
- [8] 鈴木 他, "深層生成モデルを利用した新規医薬品構造提案手法の開発", ケモインフォマティクス討論会, 2018.
- [9] 池端 他, "言語モデルベースの化学構造生成方法の提案と生体活性化分子をターゲットにしたInverse-QSARモデルへの適用", 情報処理学会研究報告, 2016-BIO-45.

※69 化学、生物学分野のための深層学習ライブラリChainer Chemistry公開、2017、Preferred Networks