

## 논문 요약 보고서

논문 제목: Conditional Molecular Design with Deep Generative Models

저자명: Seokho Kang, Kyunghyun Cho

출판 정보: J. Chem. Inf. Model. 2019, 59, 1, 43-52 Publication Date: July 17, 2018

### ■ 소개 배경

본 논문에서는 원하는 성질을 가진 새로운 분자 생성을 위한 conditional 분자설계기법을 제안합니다. 분자 설계의 1차적인 목표는 원하는 성질을 만족시키는 새로운 분자를 제안하는 것인데, 그 동안은 큰 chemical space를 효율적으로 탐구하는 데에 어려움을 겪어 왔습니다. property 예측과 분자 생성을 동시에 수행하는 본 모델은 semisupervised VAE 형태로 설계되어, 다양한 조건을 충족하는 새로운 분자를 효율적으로 생성해냄과 동시에 라벨에 unlabeled molecule을 이용해서 property 예측을 더욱 향상시킵니다. 이와 같은 모델의 생성 능력과 예측 능력은 drug-like 분자를 통해 평가 및 입증됩니다.

### ■ 결과

Baseline 모델인 SSVEA의 regression 버전을 사용하여 분자를 효율적이고 정확하게 생성하기 위한 새로운 접근법을 제시했습니다. 일부 labeled된 training dataset을 가지고 모델을 설계하고 훈련시켜, 원하는 특성을 가진 새로운 분자는 특성의 target 조건이 주어진 SSVEA 모델의 생성(generation) 분포로부터 생성되었습니다. 우리는 ZINC database으로부터 랜덤하게 수집한 310,000개의 drug-like molecules SMILES를 수집합니다. 이 중 300,000개는 train set, 10,000개는 test set으로 사용합니다. chemical property로는 MolWt, LogP, QED를 rdkit으로부터 추출하여 사용하였습니다. 분자 생성 모델과 관련된 초기 연구에서는 분자를 fingerprint로 변형시켜 property 예측을 위한 input으로 이용했으나, 최근 연구들은 분자의 그래프 표현 방법에서 직접적으로 feature를 추출하여 input으로 사용하거나 더 최근에는 분자 그래프 구조를 직접 생성하는 모델도 제안되었습니다. 본 논문에서는 그 동안 이루어진 연구들 (RNN, VAE, GAN)에서 대부분 사용한 SMILES를 사용하여 분자를 표현하였으며, conditional 분자 설계 방법을 이용하여 별도의 최적화(optimization) 절차 없이 conditional generative distribution(생성 분포)에서 분자를 직접 sampling하여 분자 설계를 수행할 수 있게 됩니다. 최적화 과정이 사라지고 decoder network만 사용하는 경우에는 소요 시간이 대폭 감소하게 되어 모델 효율성 측면에서 큰 이점을 가지게 됩니다. 뿐만 아니라 본 논문이 제시한 모델은 semi supervised model이기 때문에 특히 unlabeled data를 효과적으로 다룰 수 있게 됩니다.

### ■ 시사점

본 논문에서 제안된 접근 방식은, 우리가 지정한 목표 조건을 가진 새로운 화학 구조를 설계하는 효율적인 도구로 다양한 방면으로 사용될 수 있습니다. 다만, 본 연구에서는 'SMILES 문자열'을 분자 표현 방법으로 사용했기 때문에 chemical space가 다소 제한적이라고 볼 수 있고, 이 space를 더 넓히기 위해서는 SMILES 외에 다른 분자 표현 방법을 사용해 보는 것이 더 좋은 결과를 낼 가능성이 있습니다.