

논문 요약 보고서

논문 제목: Inverse molecular design using machine learning: Generative models for matter engineering

저자명: Benjamin Sanchez-Lengeling, Alán Aspuru-Guzik

출판 정보: Science 27 Jul 2018 Vol. 361, Issue 6400, pp. 360-365

■ 소개 배경

Inverse design(역설계)이란 원하는 기능에서 시작하여 이상적인 분자구조를 찾아냄으로써 기존패러다임을 뒤집는 개념이다. Inverse design을 통한 화학적 공간(chemical space)에서 발견되지 않은 region에서 새로운 재료의 발견은 엄청난 사회적, 기술적 발전을 가져올 수 있기에 이 논문은 원하는 특정 기능에서 맞춤 재료를 발견하는 것을 목표로 하는 inverse design method를 달성하는 것을 목표로 한다.

■ 결과

역설계 모델은 다양한 접근법이 존재하는데, 가장 중요한 것은 어떠한 접근법을 사용하느냐보다도 분자의 표현 방식이다. 분자 시스템을 정확하게 모델링하기 위해서 이산(텍스트), 연속(벡터), 그리고 그래프방식으로 분자가 표현되며, 가장 natural한 표현은 그래프 기반 표현 방법이다.

Inverse design은 첫번째로 생성 모델 (generative model)의 접근법으로 접근할 수 있다. 생성 모델의 목표는 대량의 데이터에 대한 모델을 교육하고 이와 같은 데이터를 생성하려고 시도함으로써 데이터 분포를 모델링하는 것이다. 이때 손실 함수는 유사성의 개념을 암호화하여 경험적으로 관찰된 두 분포와 생성된 분포 간의 차이를 측정한다. inverse design에서는 이러한 '생성' 외에도 생성과정을 통제하거나 bias시켜 최적화를 달성해야 한다. VAE를 사용하면 잠재 변수를 통해 데이터 생성을 제어하여 연 속적인 표현을 통해 property optimization이 수행되며, 일반적으로 강화학습(RL)과 함께 생성 과정을 bias시킴으로써 property의 최적화를 달성할 수 있다. 생성 모델을 구축하는 또 다른 방법은 GAN 프레임워크에 따른 적대적 훈련이다. 생성 프로세스를 GAN과 RNN으로 편향시키려면 네트워크의 최적화를 desired property로 이끌기 위한 gradient가 필요하다. 이는 NN을 통해 모델링이 가능하며, RL을 통해 GAN에서 발생하는 여러 문제를 해결할 수 있다. 앞서 언급한 접근방식은 배타적이지 않으며 각 접근방식은 서로 섞여서 장점을 얻을 수 있다 (ex. druGAN, ORGANIC). 생성 모델의 결과는 대부분 제약 맥락에서 사용되었고, 물의 용해성, 용해 온도, 합성성, 특정 하부 구조의 유무와 같은 잠재적 약물과 관련된 특성을 최적화했다.

■ 시사점

역설계는 물질을 빠른 속도로 설계하는 프레임워크에 있어서 필요한 중요한 구성요소다. Generative models(생성 모델)은 후보 분자를 생성하며, 이러한 후보자들의 물리적 실현은 자동화된 high-throughput 노력을 필요로 한다. chemical, biochemical, medicinal, and materials sciences 분야에서 위 방법론이 사용 가능하다. 아직까지 이 분야에서 machine learning을 사용할 때는 가장 중요한 '분자 표현 (molecule representation)'이 발전해야 할 영역으로 남아있는데, 그 중에서도 분자의 그래프와 계층적 표현은 더 많은 연구가 필요한 영역이다.