

논문 요약 보고서

논문 제목: druGAN: An Advanced Generative Adversarial Autoencoder Model for de Novo Generation of New Molecules with Desired Molecular Properties in Silico

저자명: Artur Kadurin, Sergey Nikolenko, Kuzma Khrabrov, Alex Alipez, and Alex Zhavoronkov

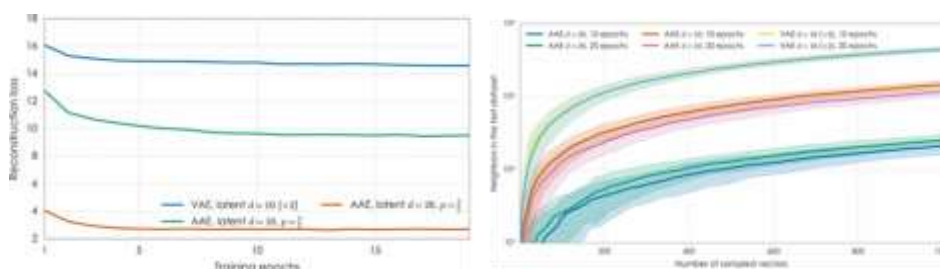
출판 정보: Mol. Pharmaceutics 2017, 14, 9, 3098–3104

■ 소개 배경

현재 가장 인기 있는 생성 모델인 VAE와 GAN, 그리고 AAE는 최근 약물 발견과 생물지표 개발 분야에서 사용된다. 특히 AAE 구현을 통해 항암 특성을 가진 새로운 molecular fingerprint를 찾아내는 기술이 사용된다. 본 논문에서 소개하는 발전된 AAE 모델인 DruGAN은, 분자 fingerprints를 생성하면서 큰 데이터셋 처리를 가능하게 하고 unsupervised pretraining 회귀 모델을 진행했을 때 efficiency에 있어서 우수하다는 것을 입증한다.

■ 결과

본 연구의 목표는 AAE 분자 생성 모델의 기존 결과를 개선하고 train 과정을 안정화하는 것이다. 제안된 기법에 대한 정성적인 평가를 위해 새로운 AAE를 VAE와 비교하였다.



첫번째로 두 모델을 autoencoder로 비교하였고 AAE가 VAE보다 월등히 우수하다는 것이 나타났다. 두 번째 실험에서는 VAE와 AAE를 generative model로 비교하였고, AAE 아키텍처는 이전에 발표한 작업에 비해 상당한 개선을 나타내지만, 다양한 fingerprint를 샘플링하는 경우 VAE가 AAE를 능가하는 적용 분야가 있을 수 있다는 게 한계점이다. 세 번째 실험에서는 생성 모델을 unsupervised feature추출에 사용하였으며, 본 모델의 autoencoder는 모든 회귀 문제에 대한 만족스러운 값을 만들어내고, pretrained network는 기존보다 훨씬 빠르게 작동했다.

■ 시사점

기존 AAE모델을 통해, 기존에 sampling되지 않았던 분자들이 잠재적인 항암제로서 연구할 가치가 있을 수 있다는 것을 알 수 있었다. 하지만 기존 연구에서는 상대적으로 작은 데이터셋을 사용하였고 그렇기에 제한적이었다. 본 논문에서는 DruGAN 모델을 더 발전시켜, 발전된 AAE 모델을 제안한다. 새롭게 제안된 모델은 Deep generative model을 사용하여 특정 항암 특성(property)을 가진 새로운 분자 개발을 할 때의 용량(capacity)과 효율성(efficiency)을 크게 향상시키며, 이는 약물 발견과 생물지표 개발 분야와 같은 제약 연구 개발 분야에서 활용될 수 있다.