

논문 요약 보고서

논문 제목: End-to-End Representation Learning for Chemical-Chemical Interaction Prediction

저자명: Sunyoung Kwon, Sungroh Yoon

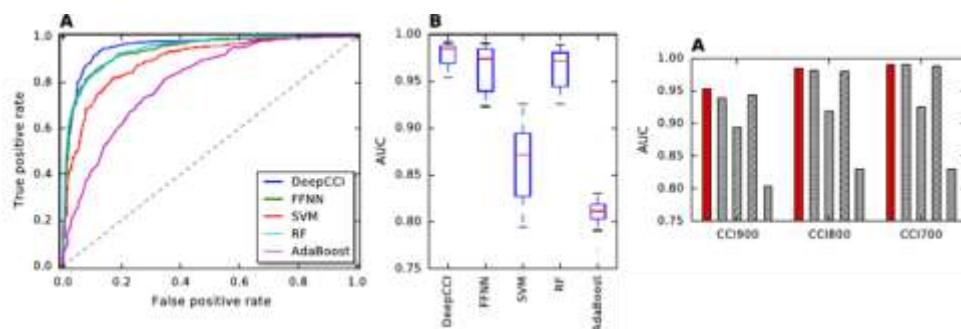
출판 정보: IEEE/ACM Trans Comput Biol Bioinform. Sep-Oct 2019;16(5):1436-1447.

■ 소개 배경

최근 Chemical-chemical interaction (이하 CCI)가 발전을 이루며 화학적 독성 또는 부작용 예측 연구, 치료 정확성 예측 연구 등 다양한 분야에서의 연구가 CCI를 통해 진행되고 있다. 다른 약물 분석에서는 최근 딥러닝이 활발히 활용되고 있긴 하지만, CCI는 예서는 아직까지 학습기반 예측은 진행된 적이 없었다. 이 논문에서는 특징 추출 및 학습 기반 접근법을 포함하는 'DeepCCI'를 제안하며, 12가지의 fingerprint와 method 조합을 이용하여 DeepCCI의 성능을 비교한다.

■ 결과

CCI dataset의 단일 화합물 data만을 사용하였다. SMILES 문자열은 논문에서 제안된 학습 방법인 DeepCCI의 Input으로 사용되었으며, 1D-CNN을 사용하여 SMILES에 숨겨진 sequential 학습 방법을 이용했다. 성능 비교를 위해 ECFP, MACCS fingerprint, PubChem, SMILES네 가지의 Input이 SVM, Random Forest, AdaBoost, FFNN에 학습되었다. 평가지표로는 AUC, ACC, MCC, TPR을 사용하여 성능을 비교하였다.



DeepCCI, FFNN, SVM, RF, AdaBoost의 AUC의 평균값은 각각 0.98, 0.86, 0.96, 0.86, 0.80으로 DeepCCI가 가장 AUC값이 큰 것으로 나타났다. 데이터 집합 크기가 커짐에 따라 전체 성능도 따라 증가했으며 AUC의 평균은 0.88에서 0.92로 향상되었다.

■ 시사점

본 논문에서 제안한 DeepCCI는 사용되는 대부분의 평가 지표에서 최상의 성능을 보여준다. CCI는 약물, 독성, 치료 효과, 생물학적 기능을 예측하는 등 넓은 연구 분야에서 사용될 수 있다. 더불어 그동안의 약물 분석에서는 neural network가 classifier의 역할만 주로 해온 점을 감안하면 이 새로운 학습 방법을 통해 숨겨진 feature를 찾아내어 모델의 성능을 높이는 학습 방법을 고안해낸 것도 또 하나의 의미 있는 발견이다.