심화 학습 기반의 약물-표적 단백질 간 상호작용 예측 시스템 및 그 방법

김홍기 교수
서울대학교 치과대학 치의학과

기술 내용

- 인공 신경망 기반의 다중채널 구조를 구축하여 약물 후보 화합물이 표적 단백질을 억제 또는 상승 등의 작용을 할 수 있는지 판단 및 예측하는 기술임
- 표적 단백질을 조절할 수 있는 약물 후보물을 도출할 때 실험을 실제로 진행할 경우 높은 비용이 소모됨
- 가상 선별에 뛰어난 성능을 보이는 인공지능 모델을 도출하여 신약 개발에 소요되는 비용을 현저히 줄이고 높은 정확률로 약물과 표적 단백질 간의 상호작용 여부를 예측함. 이를 통해 양질의 신약을 개발하고 기존 약물의 용도 확장을 기대할 수 있음

기술 개발 단계

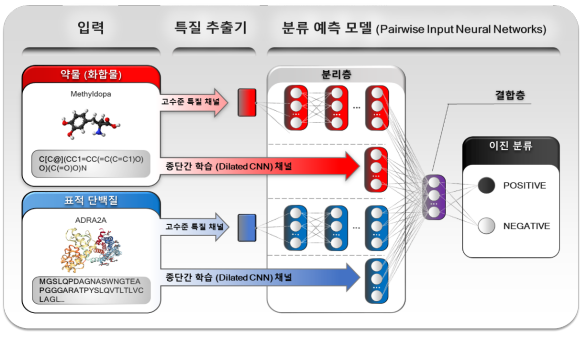
- TRL4

기술 개발 배경

신약 개발을 위해 전 임상단계에서 일반적으로 수행되는 3가지 단계는 질병을 야기하는 표적 단백질을 정의하고 표적 단백질을 조절할 수 있는 약물 후보물 도출하며 마지막으로 도출된 약물의 안정화 및 최적화 단계로 진행함. 표적 단백질을 조절할 수 있는 약물 후보물 도출하기 위해 실험을 실제로 진행할 경우 많은 시간과 장비 그리고 인력이 소요되며 컴퓨터 기반의 가상 선별 실험법이 대안으로 제시됨

기술 특장점

핵심기술요소	특장점
word2vec 기반의 표적 단백질과 약물 후보 화합물의 고차원 특질 추출 채널	• 타겟 단백질과 약물 후보 화합물을 벡터화 • 보다 적은 차원의 벡터로 표현력의 풍부함과 연산과 메모리의 효율성 획득 • 표적 단백질과 약물 후보 물질을 실수 차원으로 표현 • 아미노산 서열과 약물 그래프의 세부적인 (local) 영역의 특질 추출이 우수함
Dilated CNN을 활용한 종단간 학습을 활용한 채널	• 원 데이터에서 표적 단백질과 약물 후보물질의 특질 직접 가공 • 적은 공간 차원 손실 • 연산 효율 향상 • 아미노산 서열 및 화학식 특질 추출에 적합 • 아미노산 서열과 약물 그래프의 세부적인 (local) 영역의 특질 추출이 우수함
PINN 기반으로 다중채널을 입력 받는 예측모델	• 연산과 학습 효율성 증대 • 분리층을 통해 고도화된 특질들을 결합층에서 연결하여 각 특질이 상호보완적으로 표적 단백질과 약물 후보 화합물의 연결성 예측함



약물-표적 단백질 간 상호작용 예측 시스템 도면

기존 기술 현황

- 기존의 인공 신경망 기반의 예측 모델의 경우 인공 신경망의 기능을 분류기의 기능에 한정하여 다양한 기능을 활용하기 어려우며 표적 단백질과 약물 후보 화합물의 특질이 인간 전문가의 한정된 경험에 기반하여 가공되고 예측 모델의 입력 데이터로 사용됨. 또한 인공 신경망의 층 구조가 일괄적으로 모든 노드의 연산으로 구성되어 예측 모델의 복잡성이 과도하게 높아지고 연산의 비효율성과 과적합에 취약해지는 경향이 있음

기존 기술 대비 차별성

종래 기술의 한계	본 발명의 동작/구성	본 발명의 효과 및 이점
인공 신경망의 기능이 분류기에 한정	인공 신경망의 3가지 기능들 (특질 추출기, 종단간 학습기, 분류기)를 유기적으로 활용함	표적 단백질과 약물 후보 화합물의 다양한 수준의 특질이 각각의 채널로써 세부적이고 전체적인 특질이 상호보완적으로 활용되며 예측율을 높임
표적 단백질과 약물 후보 화합물의 특질이 인간 전문가의 한정된 경험에 기반하여 가공되고 그 정보를 예측 모델의 입력 데이터로 사용함	표적 단백질과 약물 후보 화합물의 특질을 각각 약 55만 종의 단백질과 1,990만 개의 구분하고 약물 후보군을 기반으로 기계가 스스로 유의미한 특질을 학습함	인간 전문가가 아직 발견하지 못한 단백질 또는 화합물에 대해서도 다양한 특질을 추출하고 예측 모델에 활용함
인공 신경망의 층 구조가 일괄적으로 모든 노드의 연산으로 구성되어 있기에 예측 모델의 복잡성이 과도하게 높아지고 연산의 비효율성과 과적합에 취약함	분리층과 결합층으로 이루어진 인공신경망의 층 구조는 분리층의 독립된 학습구조를 지원하며 고도화된 분리층의 특질은 결합층에서 상호보완적으로 활용됨	기존의 인공 신경망 모델에 비해서 동일한 성능으로 연산에 보다 효율적이고 과적합에 강인한 효과를 보임

기술 활용 분야

- 의생물 관련 데이터를 기반으로 생물 활동성(bioactivity)을 활용하는 제약 분야 및 생물 정보 분야

지식재산권 현황

No.	명칭	국가	상태	출원번호(출원일)	등록번호(등록일)	권리자
1	RISC를 이용한 폴리뉴클레오타이드의 검출방법	대한민국	출원	10-2019-0050716 (2019.04.30.)	-	서울대학교 산학협력단

기술 문의처

- 서울대학교 산학협력단 이한용 변리사 | 02-880-2026 | boribob@snu.ac.kr

약물-표적 단백질 간 상호작용 예측 시스템 및 그 방법

김홍기 교수
서울대학교 치과대학 치의학과

기술 내용

- 인공 신경망 기반으로 예측 모델을 구축하여 해당 약물이 주어진 단백질을 조절할 수 있는지를 판단 및 예측하는 기술임
- 표적 단백질을 조절할 수 있는 약물 후보물을 도출할 때 실험을 실제로 진행할 경우 높은 비용이 소모됨
- 약물의 화학적 구조와 단백질의 아미노산 서열을 벡터화하여 신경망의 입력 레이어의 입력으로 활용하며 인공 신경망을 통한 벡터화된 단백질 및 약물 간의 상호작용 여부를 예측을 수행할 수 있음

기술 개발 단계

- TRL4

기술 개발 배경

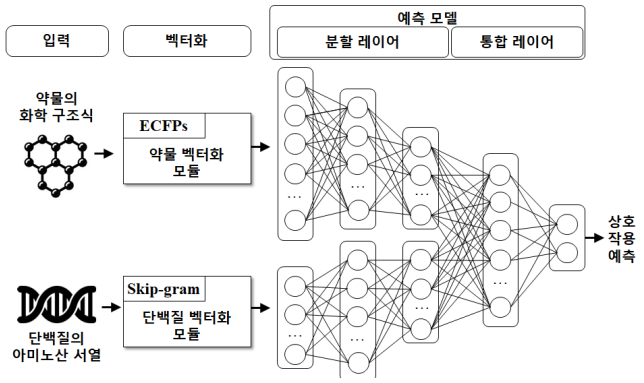
- 신약 개발을 위해 전 임상단계에서 일반적으로 수행되는 3가지 단계는 질병을 야기하는 표적 단백질을 정의하고 표적 단백질을 조절할 수 있는 약물 후보물 도출하며 마지막으로 도출된 약물의 안정화 및 최적화 단계로 진행함. 표적 단백질을 조절할 수 있는 약물 후보물 도출하기 위해 실험을 실제로 진행할 경우 많은 시간과 장비 그리고 인력이 소요되며 컴퓨터 기반의 가상 선별 실험법이 대안으로 제시됨

기술 특장점

핵심기술요소	특장점
Skip-gram 모델을 통한 단백질의 벡터화	• 중심 단어에서 주변 단어를 예측 • 3가지 아미노산을 하나의 단어로 아미노산 서열을 문장으로 취급 • 단백질의 아미노산 서열 특질을 반영하여 벡터화 • 단백질을 다양한 기계 학습 및 통계 기법의 데이터로 활용가능
Extended-Connectivity Fingerprints(이하 ECFPs) 방법론을 통한 약물의 벡터화	• 약물의 화학적 구조를 반영하여, 각 약물 별 이진 벡터를 식별자로 할당 • 약물을 화학 구조 기반으로 벡터화 • 약물을 기계 학습 및 통계 기법의 데이터로 활용 가능 • 순수 데이터 기반으로 예측 모델의 입력 값을 생성하여 기계 학습 모델의 자가 학습 가능
인공 신경망을 통한 벡터화된 단백질 및 약물 간의 상호작용 여부 예측 모델 도출	• 기존 인공신경망의 경우 신경망을 이루는 각 레이어의 모든 노드들이 엣지로 연결되어 통합 레이어로 구성됨 – 최초 n번째 레이어까지는 약물 노드 간에만 엣지 설계 – 최초 n번째 레이어까지는 단백질 노드 간에만 엣지 설계 – n+1 번째 레이어부터 출력 레이어까지는 통합 레이어로 구성
	• 약물-표적 단백질 간의 상호작용 예측 시 편향성 완화 • 각 요소의 특질을 기계가 자동적으로 추출 → 높은 정확도를 갖는 약물-표적 단백질 상호작용 예측 모델 구축

기존 기술 현황

- 기존의 인공 신경망 및 기계 학습을 기반으로 제안된 기술의 경우 단백질 및 약물의 속성 일부만을 도메인 전문가의 선험적 경험으로 일부분만 발췌하여 입력데이터로 활용하며 예측 모델을 위한 신경망의 구조가 일괄된 통합 레이어로 구성되어 있음. 또한 약물만의 특성을 학습에 반영하는 형식을 취하게 됨
- 따라서 특질 추출의 단계에서 단백질과 약물의 정보가 알려지지 않은 경우 기존 기법은 양질의 특질을 추출하기가 어려움. 또한 예측 모델을 구성할 때 전통적인 완전 연결형 뉴럴 네트워크 모델을 사용하기에 매우 높은 모델 복잡도에서 기인하는 과적합의 위험성이 높음



약물-표적 단백질 간 상호작용 예측 시스템 도면

기존 기술 대비 차별성

종래 기술의 한계	본 발명의 동작/구성	본 발명의 효과 및 이점
도메인 전문가의 한정된 지식에 의존하여 단백질 및 약물 속성의 일부분만을 활용하여 주요한 특성 누락	단백질 및 약물의 모든 속성을 고려하여 약물-표적 단백질 상호작용 여부 예측에 활용	약물 및 표적의 모든 속성을 고려하여 약물-표적 단백질 상호작용에 주요한 영향을 끼치는 특질을 추출하고 기계의 자가 학습을 지원함
통합 레이어 집합으로 구성된 인공 신경망 구조를 활용하여 입력 데이터의 표현력에 의한 간섭에 영향을 받음	분할된 레이어를 인공 신경망의 전방 레이어에 위치시키고 후방 레이어를 통해 통합	입력 데이터의 표현력에 의한 간섭을 축소 시키고, 약물-표적 단백질 상호작용에 주요한 영향을 끼치는 특질을 각 개체 별로 추출함
약물의 특성만을 고려하고 고정된 단백질 집합에 대해서만 동작하는 예측 모델	약물 및 단백질의 쌍(Pair)을 입력으로 취함으로써 두 종류의 개체 속성들을 모두 반영	학습 시 활용되지 않았던 임의의 단백질에 대해, 이를 조절하는 약물 도출 지원

기술 활용 분야

- 의생물 데이터의 전처리 기술, 특질 추출 기술, 예측 모델 설계 기술 등의 모든 파이프라인에 적용 가능함

지식재산권 현황

No.	명칭	국가	상태	출원번호(출원일)	등록번호(등록일)	권리자
1	심화 학습 기반의 약물-표적 단백질 간 상호작용 예측 시스템 및 그 방법	대한민국	출원	10-2017-0142389(2017.110.30.)	-	서울대학교 산학협력단
		대한민국	등록	10-2018-0130090(2018.10.29.)	10-2220653(2021.02.22.)	

기술 문의처

- 서울대학교 산학협력단 이한용 변리사 | 02-880-2026 | boribob@snu.ac.kr