

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국

(43) 국제공개일
2023년 3월 9일 (09.03.2023)

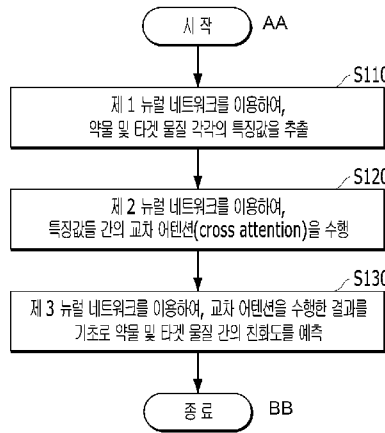


(10) 국제공개번호
WO 2023/033281 A1

- (51) 국제특허분류: *G16C 20/30* (2019.01) *G16C 20/40* (2019.01)
G16C 20/60 (2019.01) *G06N 3/08* (2006.01)
G16C 20/70 (2019.01) *G06N 3/04* (2006.01)
- (21) 국제출원번호: PCT/KR2022/003582
- (22) 국제출원일: 2022년 3월 15일 (15.03.2022)
- (25) 출원언어: 한국어
- (26) 공개언어: 한국어
- (30) 우선권정보: 10-2021-0115509 2021년 8월 31일 (31.08.2021) KR
- (71) 출원인: 디어젠 주식회사 (DEARGEN INC.) [KR/KR]; 34051 대전광역시 유성구 문지로 193, 602호, Daejeon (KR).
- (72) 발명자: 김예찬 (KIM, Yeachan); 13617 경기도 성남시 분당구 미금로 246, 603동 901호, Gyeonggi-do (KR). 신
- 봉근 (SHIN, Bonggun); 33124 충청남도 부여군 규암면 충절로 2250-11, Chungcheongnam-do (KR).
- (74) 대리인: 이대호 등 (LEE, Dae Ho et al.); 06120 서울특별시 강남구 강남대로 476 11층, Seoul (KR).
- (81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.
- (84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): ARIPO (BW, GH, GM, KE,

(54) Title: METHOD FOR PREDICTING AFFINITY BETWEEN DRUG AND TARGET SUBSTANCE

(54) 발명의 명칭: 약물 및 타겟 물질 간의 친화도를 예측하는 방법



S110 ... Extract feature value of each of drug and target substance by using first neural network

S120 ... Perform cross attention between feature values by using second neural network

S130 ... Predict affinity between drug and target substance on basis of result of performing cross attention by using third neural network

AA ... Start

BB ... End

(57) Abstract: According to some embodiments of the present disclosure, disclosed is a method for predicting the affinity between a drug and a target substance that is performed by a computing device including at least one processor. The method for predicting the affinity between a drug and a target substance may comprise the steps of: extracting a feature value of each of the drug and the target substance by using a first neural network; performing cross attention between the feature values by using a second neural network; and predicting the affinity between the drug and the target substance on the basis of the result of performing the cross attention by using a third neural network.



WO 2023/033281 A1

LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 유라시아 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 유럽 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

규칙 4.17에 의한 선언서:

- 신규성을 해치지 아니하는 개시 또는 신규성 상실의 예외에 관한 선언 (규칙 4.17(v))

공개:

- 국제조사보고서와 함께 (조약 제21조(3))

(57) 요약서: 본 개시의 몇몇 실시예에 따른, 적어도 하나의 프로세서를 포함하는 컴퓨팅 장치에 의해 수행되는 약물 및 타겟 물질 간의 친화도(affinity)를 예측하는 방법이 개시된다. 상기, 약물 및 타겟 물질 간의 친화도를 예측하는 방법은 제 1 뉴럴 네트워크를 이용하여, 상기 약물 및 상기 타겟 물질 각각의 특징값을 추출하는 단계; 제 2 뉴럴 네트워크를 이용하여, 상기 특징값들 간의 교차 어텐션(cross attention)을 수행하는 단계; 및 제 3 뉴럴 네트워크를 이용하여, 상기 교차 어텐션을 수행한 결과를 기초로 상기 약물 및 상기 타겟 물질 간의 친화도를 예측하는 단계를 포함할 수 있다.

명세서

발명의 명칭: 약물 및 타겟 물질 간의 친화도를 예측하는 방법 기술분야

- [1] 본 개시는 약물 및 타겟 물질 간의 친화도를 예측하는 방법에 관한 것으로, 구체적으로 약물 및 타겟 물질 간의 교차 어텐션을 수행하여 친화도를 예측하는 방법에 관한 것이다.

배경기술

- [2] 근래에는 약물-타겟의 상호작용(drug-target interaction, DTI)을 예측하기 위해 다양한 방법의 인실리코(in silico) 방법이 개발되고 있다. 여기서, 인실리코는 가상실험에서의 컴퓨터 프로그래밍을 뜻하는 생명정보학 용어일 수 있다. 특히, 실제 실험(in-vitro)(또는, 임상실험)을 기반으로 약물들과 타겟 물질의 상호작용의 여부를 진행할 경우 많은 시간과 더불어 높은 비용이 소모되므로, 다수의 연구자들은 인실리코 기반의 가상실험을 통해 약물과 타겟 물질 간의 상호작용을 예측하는 것을 선호할 수 있다.
- [3] 약물-타겟의 상호작용을 예측하기 위한 기술 중 하나로, 그래프 컨볼루션 네트워크(graph convolutional networks)에 기반하여 약물-타겟 간의 친화도를 예측하는 기술 “GraphDTA: prediction of drug-target binding affinity using graph convolutional networks”이 종래에 공지되었다. 그러나, 해당 기술은 예측 결과의 해석 가능성 및 약물과 타겟 물질 간의 상호 작용에 대한 고려가 미비하여, 결과에 대한 신뢰도에 한계가 존재할 수 있다.

발명의 상세한 설명

기술적 과제

- [4] 본 개시는 전술한 배경기술에 대응하여 안출된 것으로, 약물 및 타겟 물질 간의 교차 어텐션을 수행하여 친화도를 예측하기 위한 방법을 제공하고자 한다.
- [5] 본 개시의 기술적 과제들은 이상에서 언급한 기술적 과제로 제한되지 않으며, 언급되지 않은 또 다른 기술적 과제들은 아래의 기재로부터 당업자에게 명확하게 이해될 수 있을 것이다.

과제 해결 수단

- [6] 전술한 바와 같은 과제를 해결하기 위한 적어도 하나의 프로세서를 포함하는 컴퓨팅 장치에 의해 수행되는 약물 및 타겟 물질 간의 친화도(affinity)를 예측하는 방법이 개시된다. 상기 약물 및 타겟 물질 간의 친화도를 예측하는 방법은, 제 1 뉴럴 네트워크를 이용하여, 상기 약물 및 상기 타겟 물질 각각의 특징값을 추출하는 단계; 제 2 뉴럴 네트워크를 이용하여, 상기 특징값들 간의 교차 어텐션(cross attention)을 수행하는 단계; 및 제 3 뉴럴 네트워크를 이용하여, 상기 교차 어텐션을 수행한 결과를 기초로 상기 약물 및 상기 타겟 물질 간의 친화도를 예측하는 단계; 를 포함할 수 있다.

- [7] 또한, 상기 제 1 뉴럴 네트워크를 이용하여, 상기 약물 및 상기 타겟 물질 각각의 특징값을 추출하는 단계는, 상기 제 1 뉴럴 네트워크에 포함된 제 1 서브 네트워크를 이용하여, 상기 약물의 특징값을 추출하는 단계; 및 상기 제 1 뉴럴 네트워크에 포함된 제 2 서브 네트워크를 이용하여, 상기 타겟 물질의 특징값을 추출하는 단계; 를 포함할 수 있다.
- [8] 또한, 상기 제 2 뉴럴 네트워크를 이용하여, 상기 특징값들 간의 교차 어텐션(cross attention)을 수행하는 단계는, 상기 제 2 뉴럴 네트워크에 포함된 제 1 서브 네트워크를 기초로 상기 약물의 특징값과 상기 타겟 물질의 특징값을 정렬시키는 교차 어텐션을 수행하여, 상기 약물에 대한 상기 타겟 물질의 특징값을 계산하는 단계; 및 상기 제 2 뉴럴 네트워크에 포함된 제 2 서브 네트워크를 기초로 상기 약물의 특징값과 상기 타겟 물질의 특징값을 정렬시키는 교차 어텐션을 수행하여, 상기 타겟 물질에 대한 상기 약물의 특징값을 계산하는 단계; 를 포함할 수 있다.
- [9] 또한, 상기 제 2 뉴럴 네트워크에 포함된 제 1 서브 네트워크를 기초로 상기 약물의 특징값과 상기 타겟 물질의 특징값을 정렬시키는 교차 어텐션을 수행하여, 상기 약물에 대한 상기 타겟 물질의 특징값을 계산하는 단계는, 상기 약물의 상기 특징값을 기초로, 상기 약물에 대한 제 1 키 및 제 1 밸류를 생성하고, 상기 타겟 물질의 상기 특징값을 기초로, 상기 타겟 물질에 대한 제 2 쿼리를 생성하는 단계; 상기 제 2 쿼리 및 상기 제 1 키를 기초로 교차 어텐션 벡터를 생성하는 단계; 및 상기 교차 어텐션 벡터를 상기 제 1 밸류에 적용하여, 상기 약물에 대한 상기 타겟 물질의 특징값을 계산하는 단계; 를 포함할 수 있다.
- [10] 또한, 상기 제 2 쿼리 및 상기 제 1 키를 기초로 교차 어텐션 벡터를 생성하는 단계는, 상기 타겟 물질의 전체 시퀀스에 대한 상기 제 2 쿼리 및 상기 제 1 키를 기초로 하는 확률값을 계산하여 상기 교차 어텐션 벡터를 생성하는 단계; 를 포함할 수 있다.
- [11] 또한, 상기 확률값은, sparsemax 함수에 기초하여 결정될 수 있다.
- [12] 또한, 상기 교차 어텐션 벡터를 상기 제 1 밸류에 적용하여, 상기 약물에 대한 상기 타겟 물질의 특징값을 계산하는 단계는, 상기 교차 어텐션 벡터와 상기 제 1 밸류 간의 요소 별(element-wise) 연산을 수행하는 단계; 및 상기 요소 별 연산을 수행한 결과를 기초로 상기 약물에 대한 상기 타겟 물질의 특징값을 계산하는 단계; 를 포함할 수 있다.
- [13] 또한, 상기 제 2 뉴럴 네트워크에 포함된 제 2 서브 네트워크를 기초로 상기 약물의 특징값과 상기 타겟 물질의 특징값을 정렬시키는 교차 어텐션을 수행하여, 상기 타겟 물질에 대한 상기 약물의 특징값을 계산하는 단계는, 상기 타겟 물질의 상기 특징값을 기초로, 상기 타겟 물질에 대한 제 2 키 및 제 2 밸류를 생성하고, 상기 약물의 상기 특징값을 기초로, 상기 약물에 대한 제 1 쿼리를 생성하는 단계; 상기 제 1 쿼리 및 상기 제 2 키를 기초로 교차 어텐션 벡터를 생성하는 단계; 및 상기 교차 어텐션 벡터를 상기 제 2 밸류에 적용하여,

- 상기 타겟 물질에 대한 상기 약물의 특징값을 계산하는 단계; 를 포함할 수 있다.
- [14] 또한, 상기 제 1 쿼리 및 상기 제 2 키를 기초로 교차 어텐션 벡터를 생성하는 단계는, 상기 약물의 전체 시퀀스에 대한 상기 제 1 쿼리 및 상기 제 2 키를 기초로 하는 확률값을 계산하여 상기 교차 어텐션 벡터를 생성하는 단계; 를 포함할 수 있다.
- [15] 또한, 상기 확률값은, sparsemax 함수에 기초하여 결정될 수 있다.
- [16] 또한, 상기 교차 어텐션 벡터를 상기 제 2 밸류에 적용하여, 상기 타겟 물질에 대한 상기 약물의 특징값을 계산하는 단계는, 상기 교차 어텐션 벡터와 상기 제 2 밸류 간의 요소 별 연산을 수행하는 단계; 및 상기 요소 별 연산을 수행한 결과를 기초로 상기 타겟 물질에 대한 상기 약물의 특징값을 계산하는 단계; 를 포함할 수 있다.
- [17] 또한, 컴퓨터 판독가능 저장 매체에 저장된 컴퓨터 프로그램으로서, 상기 컴퓨터 프로그램은 하나 이상의 프로세서에서 실행되는 경우, 약물 및 타겟 물질 간의 친화도를 예측하기 위한 방법을 수행하도록 하며, 상기 방법은: 제 1 뉴럴 네트워크를 이용하여, 상기 약물 및 상기 타겟 물질 각각의 특징값을 추출하는 단계; 제 2 뉴럴 네트워크를 이용하여, 상기 특징값들 간의 교차 어텐션(cross attention)을 수행하는 단계; 및 제 3 뉴럴 네트워크를 이용하여, 상기 교차 어텐션을 수행한 결과를 기초로 상기 약물 및 상기 타겟 물질 간의 친화도를 예측하는 단계; 를 포함할 수 있다.
- [18] 또한, 약물 및 타겟 물질 간의 친화도를 예측하는 컴퓨팅 장치로서, 적어도 하나의 코어를 포함하는 프로세서; 및 약물 및 타겟 물질과 관련된 정보를 저장하는 저장부; 를 포함하고, 상기 프로세서는, 제 1 뉴럴 네트워크를 이용하여, 상기 약물 및 상기 타겟 물질 각각의 특징값을 추출하고, 제 2 뉴럴 네트워크를 이용하여, 상기 특징값들 간의 교차 어텐션(cross attention)을 수행하고, 그리고 제 3 뉴럴 네트워크를 이용하여, 상기 교차 어텐션을 수행한 결과를 기초로 상기 약물 및 상기 타겟 물질 간의 친화도를 예측할 수 있다.

[19]

- [20] 본 개시에서 얻을 수 있는 기술적 해결 수단은 이상에서 언급한 해결 수단들로 제한되지 않으며, 언급하지 않은 또 다른 해결 수단들은 아래의 기재로부터 본 개시가 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

발명의 효과

- [21] 본 개시의 몇몇 실시예에 따르면, 예측 결과의 해석 가능성을 제공하고 또한 신뢰도가 높은 약물 및 타겟 물질 간의 친화도를 예측하는 방법을 제공할 수 있도록 한다.
- [22] 본 개시에서 얻을 수 있는 효과는 이상에서 언급한 효과로 제한되지 않으며, 언급하지 않은 또 다른 효과들은 아래의 기재로부터 본 개시가 속하는

기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

- [23] 다양한 양상들이 이제 도면들을 참조로 기재되며, 여기서 유사한 참조 번호들은 총괄적으로 유사한 구성요소들을 지칭하는데 이용된다. 이하의 실시예에서, 설명 목적을 위해, 다수의 특정 세부사항들이 하나 이상의 양상들의 총체적 이해를 제공하기 위해 제시된다. 그러나, 그러한 양상(들)이 이러한 특정 세부사항들 없이 실시될 수 있음은 명백할 것이다. 다른 예시들에서, 공지의 구조들 및 장치들이 하나 이상의 양상들의 기재를 용이하게 하기 위해 블록도 형태로 도시된다.
- [24] 도 1은 본 개시의 몇몇 실시예에 따른 컴퓨팅 장치의 일례를 설명하기 위한 블록 구성도이다.
- [25] 도 2는 본 개시의 몇몇 실시예에 따른 컴퓨팅 장치가 약물 및 타겟 물질 간의 친화도를 예측하는 방법의 일례를 설명하기 위한 흐름도이다.
- [26] 도 3은 본 개시의 몇몇 실시예에 따른 컴퓨팅 장치가 특징값들 간의 교차 어텐션을 수행하는 방법의 일례를 설명하기 위한 흐름도이다.
- [27] 도 4는 본 개시의 몇몇 실시예에 따른 컴퓨팅 장치가 약물에 대한 타겟 물질의 특징값을 계산하는 방법의 일례를 설명하기 위한 흐름도이다.
- [28] 도 5는 본 개시의 몇몇 실시예에 따른 컴퓨팅 장치가 타겟 물질에 대한 약물의 특징값을 계산하는 방법의 일례를 설명하기 위한 흐름도이다.
- [29] 도 6은 본 개시의 몇몇 실시예에 따른 약물 및 타겟 물질 간의 친화도를 예측하는 방법을 구현하기 위한 프레임워크의 일례를 설명하기 위한 도면이다.
- [30] 도 7은 본 개시의 몇몇 실시예에 따른 본 발명과 종래의 기술과의 비교를 위한 도면이다.
- [31] 도 8은 본 개시내용의 실시예들이 구현될 수 있는 예시적인 컴퓨팅 환경에 대한 일반적인 개략도를 도시한다.

발명의 실시를 위한 형태

- [32] 다양한 실시예들 및/또는 양상들이 이제 도면들을 참조하여 개시된다. 하기 설명에서는 설명을 목적으로, 하나 이상의 양상들의 전반적 이해를 돕기 위해 다수의 구체적인 세부사항들이 개시된다. 그러나, 이러한 양상(들)은 이러한 구체적인 세부사항들 없이도 실행될 수 있다는 점 또한 본 개시의 기술 분야에서 통상의 지식을 가진 자에게 감지될 수 있을 것이다. 이후의 기재 및 첨부된 도면들은 하나 이상의 양상들의 특정한 예시적인 양상들을 상세하게 기술한다. 하지만, 이러한 양상들은 예시적인 것이고 다양한 양상들의 원리들에서의 다양한 방법들 중 일부가 이용될 수 있으며, 기술되는 설명들은 그러한 양상들 및 그들의 균등물들을 모두 포함하고자 하는 의도이다. 구체적으로, 본 명세서에서 사용되는 "실시예", "예", "양상", "예시" 등은 기술되는 임의의 양상 또는 설계가 다른 양상 또는 설계들보다 양호하다거나, 이점이 있는 것으로

해석되지 않을 수도 있다.

- [33] 이하, 도면 부호에 관계없이 동일하거나 유사한 구성 요소는 동일한 참조 번호를 부여하고 이에 대한 중복되는 설명은 생략한다. 또한, 본 명세서에 개시된 실시예를 설명함에 있어서 관련된 공지 기술에 대한 구체적인 설명이 본 명세서에 개시된 실시예의 요지를 흐릴 수 있다고 판단되는 경우 그 상세한 설명을 생략한다. 또한, 첨부된 도면은 본 명세서에 개시된 실시예를 쉽게 이해할 수 있도록 하기 위한 것일 뿐, 첨부된 도면에 의해 본 명세서에 개시된 기술적 사상이 제한되지 않는다.
- [34] 비록 제 1, 제 2 등이 다양한 소자나 구성요소들을 서술하기 위해서 사용되나, 이들 소자나 구성요소들은 이들 용어에 의해 제한되지 않음은 물론이다. 이들 용어들은 단지 하나의 소자나 구성요소를 다른 소자나 구성요소와 구별하기 위하여 사용하는 것이다. 따라서, 이하에서 언급되는 제 1 소자나 구성요소는 본 발명의 기술적 사상 내에서 제 2 소자나 구성요소 일 수도 있음은 물론이다.
- [35] 다른 정의가 없다면, 본 명세서에서 사용되는 모든 용어(기술 및 과학적 용어를 포함)는 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 공통적으로 이해될 수 있는 의미로 사용될 수 있을 것이다. 또 일반적으로 사용되는 사전에 정의되어 있는 용어들은 명백하게 특별히 정의되어 있지 않는 한 이상적으로 또는 과도하게 해석되지 않는다.
- [36] 더불어, 용어 "또는"은 배타적 "또는"이 아니라 내포적 "또는"을 의미하는 것으로 의도된다. 즉, 달리 특정되지 않거나 문맥상 명확하지 않은 경우에, "X는 A 또는 B를 이용한다"는 자연적인 내포적 치환 중 하나를 의미하는 것으로 의도된다. 즉, X가 A를 이용하거나; X가 B를 이용하거나; 또는 X가 A 및 B 모두를 이용하는 경우, "X는 A 또는 B를 이용한다"가 이들 경우들 어느 것으로도 적용될 수 있다. 또한, 본 명세서에 사용된 "및/또는"이라는 용어는 열거된 관련 아이템들 중 하나 이상의 아이템의 가능한 모든 조합을 지칭하고 포함하는 것으로 이해되어야 한다.
- [37] 또한, "포함한다" 및/또는 "포함하는"이라는 용어는, 해당 특징 및/또는 구성요소가 존재함을 의미하지만, 하나 이상의 다른 특징, 구성요소 및/또는 이들의 그룹의 존재 또는 추가를 배제하지 않는 것으로 이해되어야 한다. 또한, 달리 특정되지 않거나 단수 형태를 지시하는 것으로 문맥상 명확하지 않은 경우에, 본 명세서와 청구범위에서 단수는 일반적으로 "하나 또는 그 이상"을 의미하는 것으로 해석되어야 한다.
- [38] 더불어, 본 명세서에서 사용되는 용어 "정보" 및 "데이터"는 종종 서로 상호교환 가능하도록 사용될 수 있다.
- [39] 어떤 구성 요소가 다른 구성 요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성 요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성 요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성 요소가 다른 구성 요소에 "직접 연결되어" 있다거나

"직접 접속되어" 있다고 언급된 때에는, 중간에 다른 구성 요소가 존재하지 않는 것으로 이해되어야 할 것이다.

[40] 이하의 설명에서 사용되는 구성 요소에 대한 접미사 "모듈" 및 "부"는 명세서 작성의 용이함만이 고려되어 부여되거나 혼용되는 것으로서 그 자체로서 구별되는 의미 또는 역할을 갖는 것은 아니다.

[41] 본 개시의 목적 및 효과, 그리고 그것들을 달성하기 위한 기술적 구성들은 첨부되는 도면과 함께 상세하게 후술되어 있는 실시예들을 참조하면 명확해질 것이다. 본 개시를 설명하는데 있어서 공지 기능 또는 구성에 대한 구체적인 설명이 본 개시의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략할 것이다. 그리고 후술되는 용어들은 본 개시에서의 기능을 고려하여 정의된 용어들로써 이는 사용자, 운용자의 의도 또는 관례 등에 따라 달라질 수 있다.

[42] 그러나 본 개시는 이하에서 개시되는 실시예들에 한정되는 것이 아니라 서로 다른 다양한 형태로 구현될 수 있다. 단지 본 실시예들은 본 개시가 완전하도록 하고, 본 개시가 속하는 기술분야에서 통상의 지식을 가진 자에게 개시의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 개시는 청구항의 범주에 의해 정의될 뿐이다. 그러므로 그 정의는 본 명세서 전반에 걸친 내용을 토대로 내려져야 할 것이다.

[43]

[44] 본 개시에서 컴퓨팅 장치는 약물 및 타겟 물질 간의 친화도(affinity)를 예측할 수 있다. 여기서, 친화도는 약물과 타겟 물질 간에 작용하는 결합력 또는 힘을 나타낼 수 있다. 타겟 물질은 예를 들어 단백질 등일 수 있다. 컴퓨팅 장치는 친화도를 예측하기 위해 뉴럴 네트워크(neural network)를 이용하여 약물 및 타겟 물질 각각의 특징값을 추출할 수 있다. 특징값이 추출된 경우, 컴퓨팅 장치는 특징값들 간의 교차 어텐션(Gated Cross Attention)을 수행하고, 교차 어텐션을 수행한 결과를 기초로 약물 및 타겟 물질 간의 친화도를 예측할 수 있다. 여기서, 어텐션은 딥러닝 모델이 특정 벡터에 주목하게 만들어 모델의 성능을 높이는 기법일 수 있다. 본 개시에서는 어텐션 기법에 기반하여, 약물 및 타겟 물질 간의 특징값들 간의 교차 어텐션을 수행함으로써 보다 높은 신뢰도를 갖는 친화도를 예측할 수 있다. 이하, 도 1 내지 도 7을 통해 본 개시에 따른 컴퓨팅 장치가 약물 및 타겟 물질 간의 친화도를 예측하는 방법에 대해 설명한다.

[45]

[46] 도 1은 본 개시의 몇몇 실시예에 따른 컴퓨팅 장치의 일례를 설명하기 위한 블록 구성도이다.

[47] 도 1을 참조하면, 컴퓨팅 장치(100)는 프로세서(110) 및 저장부(120)를 포함할 수 있다. 다만, 상술한 구성 요소들은 컴퓨팅 장치(100)를 구현하는데 있어서 필수적인 것은 아니어서, 컴퓨팅 장치(100)는 위에서 열거된 구성요소들 보다 많거나, 또는 적은 구성요소들을 가질 수 있다.

- [48] 컴퓨팅 장치(100)는 예를 들어, 마이크로프로세서, 메인프레임 컴퓨터, 디지털 프로세서, 휴대용 디바이스 또는 디바이스 제어기 등과 같은 임의의 타입의 컴퓨터 시스템 또는 컴퓨터 디바이스를 포함할 수 있다.
- [49] 프로세서(110)는 통상적으로 컴퓨팅 장치(100)의 전반적인 동작을 처리할 수 있다. 프로세서(110)는 컴퓨팅 장치(100)에 포함된 구성요소들을 통해 입력 또는 출력되는 신호, 데이터, 정보 등을 처리하거나 저장부(120)에 저장된 응용 프로그램을 구동함으로써, 사용자에게 적절한 정보 또는 기능을 제공 또는 처리할 수 있다.
- [50] 본 개시에서, 프로세서(110)는 약물 및 타겟 물질 각각의 특징값을 추출할 수 있다. 일례로, 프로세서(110)는 제 1 뉴럴 네트워크를 이용하여, 약물 및 타겟 물질 각각의 특징값(feature)을 추출할 수 있다. 여기서, 제 1 뉴럴 네트워크는 종래의 컨볼루션 뉴럴 네트워크(Convolutional Neural Networks, CNN) 동일 수 있다. 또는 제 1 뉴럴 네트워크는 약물 및 타겟 물질 각각의 특징값을 추출하기 위한 특징값 추출기(feature extractor)일 수도 있다. 프로세서(110)는 특징값들이 추출된 경우, 제 2 뉴럴 네트워크를 이용하여, 특징값들 간의 교차 어텐션(Gated Cross Attention)을 수행할 수 있다. 본 개시에서, 교차 어텐션은 약물의 특징값과 타겟 물질의 특징값을 정렬시킴으로써, 약물에 대한 타겟 물질의 특징값을 계산하는 어텐션일 수 있다. 또는, 교차 어텐션은 약물의 특징값과 타겟 물질의 특징값을 정렬시킴으로써, 타겟 물질에 대한 약물의 특징값을 계산하는 어텐션일 수 있다. 일례로, 프로세서(110)는 약물에 대한 타겟 물질의 특징값을 계산하기 위해, 약물에 대한 키(key) 및 타겟 물질에 대한 쿼리(query)를 기초로 교차 어텐션 벡터를 생성할 수 있다. 여기서, 교차 어텐션 벡터는 약물에 대한 타겟 물질의 특징값을 계산하기 위한 게이팅 함수(gating function)일 수 있다. 프로세서(110)는 교차 어텐션 벡터를 약물에 대한 밸류(value)에 적용하여, 약물에 대한 타겟 물질의 특징값을 계산할 수 있다. 환언하자면, 프로세서(110)는 약물에 대한 타겟 물질의 특징값을 계산함에 있어서, 약물의 특징값 뿐만 아니라, 타겟 물질의 특징값을 교차로 이용할 수 있다. 프로세서(110)는 약물에 대한 타겟 물질의 특징값이 계산되고, 타겟 물질에 대한 약물의 특징값이 계산된 경우, 제 3 뉴럴 네트워크를 이용하여 약물 및 타겟 물질 간의 친화도를 예측할 수 있다. 따라서, 프로세서(110)는 보다 높은 신뢰도를 갖는 약물 및 타겟 물질 간의 친화도를 예측할 수 있다. 이하, 본 개시에 따른 프로세서(110)가 약물 및 타겟 물질 간의 친화도를 예측하는 일례는 도 2 내지 도 5를 통해 설명한다.
- [51] 저장부(120)는 메모리 및/또는 영구저장매체를 포함할 수 있다. 메모리는 플래시 메모리 타입(flash memory type), 하드디스크 타입(hard disk type), 멀티미디어 카드 마이크로 타입(multimedia card micro type), 카드 타입의 메모리(예를 들어 SD 또는 XD 메모리 등), 램(Random Access Memory, RAM), SRAM(Static Random Access Memory), 롬(Read-Only Memory, ROM), EEPROM(Electrically Erasable Programmable Read-Only Memory),

PROM(Programmable Read-Only Memory), 자기 메모리, 자기 디스크, 광디스크 중 적어도 하나의 타입의 저장매체를 포함할 수 있다.

[52] 본 개시에서, 저장부(120)는 복수의 약물 및 복수의 타겟 물질과 관련된 정보를 데이터베이스로 저장할 수 있다. 일례로, 저장부(120)는 간략화된 분자 입력 라인-엔트리 시스템(simplified molecular input line entry system: SMILES)으로 표현된 약물 후보 화합물의 원 데이터를 저장하고 있을 수 있다. 여기서, SMILES는 short ASCII 문자열을 사용하여 화학 종의 구조를 설명하기 위한 라인 표기법의 규격일 수 있다. 일례로, 저장부(120)는 통신부(미도시)를 통해 사용 가능한 화합물에 대한 무료 데이터 베이스인 ZINC 데이터 베이스로부터 수신된 약물 후보 화합물의 원 데이터를 저장하고 있을 수 있다. 또한, 저장부(120)는 단백질의 구성원인 아미노산의 서열을 나타내는 아미노산 서열(amino acid sequence) 데이터를 저장하고 있을 수 있다.

[53] 상술한 구성에 따르면, 컴퓨팅 장치(100)는 약물에 대한 타겟 물질의 특징값을 계산함에 있어서, 타겟 물질의 특징값을 이용할 수 있다. 또한, 컴퓨팅 장치(100)는 타겟 물질에 대한 약물의 특징값을 계산함에 있어서, 약물의 특징값을 이용할 수 있다. 컴퓨팅 장치(100)는 계산된 약물에 대한 타겟 물질의 특징값 및 계산된 타겟 물질에 대한 약물의 특징값에 기초하여, 약물 및 타겟 물질 간의 친화도를 예측할 수 있다. 따라서, 컴퓨팅 장치(100)는 보다 높은 신뢰도를 갖는 약물 및 타겟 물질 간의 친화도를 예측할 수 있다. 이하, 도 2를 통해 본 개시에 따른 컴퓨팅 장치(100)가 약물 및 타겟 물질 간의 친화도를 예측하는 구체적인 방법에 대해 설명한다.

[54]

[55] 도 2는 본 개시의 몇몇 실시예에 따른 컴퓨팅 장치가 약물 및 타겟 물질 간의 친화도를 예측하는 방법의 일례를 설명하기 위한 흐름도이다.

[56] 도 2를 참조하면, 컴퓨팅 장치(100)의 프로세서(110)는 제 1 뉴럴 네트워크를 이용하여, 약물 및 타겟 물질 각각의 특징값을 추출할 수 있다(S110).

[57] 일례로, 프로세서(110)는 제 1 뉴럴 네트워크에 포함된 제 1 서브 네트워크를 이용하여, 약물의 특징값을 추출할 수 있다. 여기서, 제 1 뉴럴 네트워크 및 제 1 서브 네트워크 각각은 컨볼루션 뉴럴 네트워크(Convolutional Neural Networks, CNN)일 수 있으나, 이에 한정되는 것은 아니고 ANN(Artificial Neural Network), DNN(Deep Neural Network) 또는 RNN(Recurrent Neural Network) 등의 네트워크 모델이 사용될 수도 있다. 프로세서(110)는 제 1 서브 네트워크를 이용하여, 저장부(120)에 저장된 약물로부터 특징값을 추출할 수 있다. 일례로, 프로세서는 하기의 수학적 식을 통해 약물의 특징값을 추출하여 결정할 수 있다.

[58] [수식1]

$$r_d = \text{pooling}(d_{1:m_d})$$

[59] 여기서,

pooling(·)

는 글로벌 맥스 풀링(global max pooling) 또는 평균 풀링(average pooling) 등일 수 있다.

n_d

는 약물의 특징값의 순차 길이이고 아래 첨자 1은 약물의 특징값의 순차 범위를 나타낼 수 있다.

[60] 프로세서(110)는 제 1 뉴럴 네트워크에 포함된 제 2 서브 네트워크를 이용하여, 타겟 물질의 특징값을 추출할 수 있다. 여기서, 제 2 서브 네트워크는 컨볼루션 뉴럴 네트워크일 수 있으나, 이에 한정되는 것은 아니고 ANN, DNN 또는 RNN 등의 네트워크 모델이 사용될 수도 있다. 프로세서(110)는 제 2 서브 네트워크를 이용하여, 저장부(120)에 저장된 타겟 물질로부터 특징값을 추출할 수 있다. 일례로, 프로세서는 하기의 수학적식을 통해 타겟 물질의 특징값을 추출하여 결정할 수 있다.

[61] [수식2]

$$r_p = pooling(p_{1:n_p})$$

[62] 여기서,

pooling(·)

는 글로벌 맥스 풀링(global max pooling) 또는 평균 풀링(average pooling) 등일 수 있다.

n_p

는 타겟 물질의 특징값의 순차 길이이고 아래 첨자 1은 타겟 물질의 특징값의 순차 범위를 나타낼 수 있다.

[63] 컴퓨팅 장치(100)의 프로세서(110)는 제 2 뉴럴 네트워크를 이용하여, 특징값들 간의 교차 어텐션(Gated Cross Attention)을 수행할 수 있다(S120). 여기서, 교차 어텐션의 수행은 약물의 특징값과 타겟 물질의 특징값을 정렬시킴으로써, 약물에 대한 타겟 물질의 특징값을 계산하는 동작일 수 있다. 또는, 교차 어텐션의 수행은 약물의 특징값과 타겟 물질의 특징값을 정렬시킴으로써, 타겟 물질에 대한 약물의 특징값을 계산하는 동작일 수 있다.

[64] 일례로, 프로세서(110)는 제 2 뉴럴 네트워크에 포함된 제 1 서브 네트워크를 기초로 약물의 특징값과 타겟 물질의 특징값 간의 교차 어텐션을 수행하여, 약물에 대한 타겟 물질의 특징값을 계산할 수 있다. 구체적으로, 프로세서(110)는 제 1 뉴럴 네트워크를 이용하여 추출된 약물의 특징값을 기초로, 약물에 대한 제 1 키(key) 및 제 1 밸류(value)를 생성할 수 있다. 프로세서(110)는 제 1 뉴럴 네트워크를 이용하여 추출된 타겟 물질의 특징값을 기초로, 타겟 물질에 대한 제 2 쿼리(query)를 생성할 수 있다. 프로세서(110)는 제 2 쿼리 및 제 1 키를 기초로

교차 어텐션 벡터를 생성하고, 생성된 교차 어텐션 벡터를 제 1 벨류에 적용하여, 약물에 대한 타겟 물질의 특징값을 계산할 수 있다. 여기서, 교차 어텐션 벡터는 약물에 대한 타겟 물질의 특징값을 계산하기 위한 게이팅 함수(gating function)일 수 있다.

[65] 프로세서(110)는 제 2 뉴럴 네트워크에 포함된 제 2 서브 네트워크를 기초로 타겟 물질의 특징값과 약물의 특징값 간의 교차 어텐션을 수행하여, 타겟 물질에 대한 약물의 특징값을 계산할 수 있다. 구체적으로, 프로세서(110)는 제 1 뉴럴 네트워크를 이용하여 추출된 타겟 물질의 특징값을 기초로, 타겟 물질에 대한 제 2 키 및 제 2 벨류를 생성할 수 있다. 프로세서(110)는 제 1 뉴럴 네트워크를 이용하여 추출된 약물의 특징값을 기초로, 약물에 대한 제 1 쿼리를 생성할 수 있다. 프로세서(110)는 제 1 쿼리 및 제 2 키를 기초로 교차 어텐션 벡터를 생성하고, 생성된 교차 어텐션 벡터를 제 2 벨류에 적용하여, 타겟 물질에 대한 약물의 특징값을 계산할 수 있다. 이하, 본 개시에 따른 프로세서(110)가 약물에 대한 타겟 물질의 특징값을 계산하는 방법은 도 3 및 도 4를 통해 설명한다. 또한, 본 개시에 따른 프로세서(110)가 타겟 물질에 대한 약물의 특징값을 계산하는 방법은 도 3 및 도 5를 통해 설명한다.

[66] 본 개시의 몇몇 실시예에 따르면, 컴퓨팅 장치(100)의 프로세서(110)는 하기의 수학적식을 통해 약물에 대해 타겟 물질의 교차 어텐션이 수행된 결과를 정의할 수 있다.

[67] [수식3]

$$d'_{1:m_d} = g_{p \rightarrow d}(d_{1:m_d}, p_{1:m_p}; \theta_{p \rightarrow d})$$

[68] 여기서,

$$g(\cdot)$$

는 상대 정보(타겟 물질 정보)를 고려하는 어텐션 함수일 수 있다. 예를 들어,

$$g_{p \rightarrow d}$$

는 학습 가능한 파라미터

$$\theta_{p \rightarrow d}$$

를 사용하여 약물에 대한 특징값

$$d_{1:m_d}$$

에 타겟 물질에 기반한 어텐션을 반영하기 위해, 타겟 물질의 특징값

$$p_{1:m_p}$$

을 사용할 수 있다. 그리고,

$$g_{p \rightarrow d}$$

는

$$p_{1:m_p}$$

사용하여

$$d_{1:m_d}$$

의 어텐션된 특징값인

$$d'_{1:m_d}$$

을 반환할 수 있다.

[69] 또한, 컴퓨팅 장치(100)의 프로세서(110)는 하기의 수학식을 통해 타겟 물질에 대해 약물의 교차 어텐션이 수행된 결과를 정의할 수 있다.

[70] [수식4]

$$p'_{1:m_p} = g_{d \rightarrow p}(p_{1:m_p}, d_{1:m_d}; \theta_{d \rightarrow p})$$

[71] 여기서,

$$g(\cdot)$$

는 상대 정보(약물 정보)를 고려하는 어텐션 함수일 수 있다. 예를 들어,

$$g_{d \rightarrow p}$$

는 학습 가능한 파라미터

$$\theta_{d \rightarrow p}$$

를 사용하여 타겟 물질에 대한 특징값

$$pd_{1:m_p}$$

에 약물에 기반한 어텐션을 반영하기 위해, 약물의 특징값

$$d_{1:m_d}$$

을 사용할 수 있다. 그리고,

$$g_{d \rightarrow p}$$

는

$$d_{1:m_d}$$

을 사용하여

$$p_{1:m_p}$$

의 어텐션된 특징값인

$$p'_{1:m_p}$$

을 반환할 수 있다.

- [72] 본 개시의 몇몇 실시예에 따르면, 교차 어텐션은 멀티 헤드 어텐션(Multi-head Attention 또는 Multi-head Gated attention)일 수도 있다. 여기서, 멀티 헤드 어텐션은 복수개의 교차 어텐션을 병렬적으로 수행하여, 약물에 대한 타겟 물질의 특징값 및 타겟 물질에 대한 약물의 특징값을 계산하기 위한 어텐션일 수 있다.
- [73] 컴퓨팅 장치(100)의 프로세서(110)는 제 3 뉴럴 네트워크를 이용하여, 교차 어텐션을 수행한 결과를 기초로 약물 및 타겟 물질 간의 친화도를 예측할 수 있다(S130).
- [74] 실시예에 따라, 프로세서(110)가 교차 어텐션을 수행한 약물 및 타겟 물질 간의 상호 작용은 명시적으로 구성될 수 있다. 예를 들어, 프로세서(110)는 명시적 상호 작용의 결과물로서 어텐션 맵(attention map)을 생성할 수 있다. 어텐션 맵은 예측 결과에 대한 해석 가능한 요소의 역할을 수행할 수 있고, 프로세서(110)는 어텐션 맵을 이용하여, 약물 및 타겟 물질 간의 친화도를 예측할 수 있다.
- [75] 본 개시의 몇몇 실시예에 따르면, 컴퓨팅 장치(100)의 프로세서(110)는 예측된 친화도에 대한 스코어를 생성할 수 있다. 일례로, 프로세서(110)는 하기의 수학적식을 통해 스코어를 생성할 수 있다.

[76] [수식5]

$$\hat{y} = \phi(r_d, r_p; \theta_i)$$

[77] 여기서,

\hat{y}

는 친화도 점수일 수 있다.

$\phi(\cdot)$

는 제 3 뉴럴 네트워크일 수 있고,

θ_i

는 제 3 뉴럴 네트워크에 의해 사용되는 파라미터일 수 있다.

- [78] 상술한 구성에 따르면, 컴퓨팅 장치(100)는 교차 어텐션을 기반으로 약물 및 타겟 물질 간의 친화도를 예측할 수 있다. 일례로, 프로세서(110)는 약물의 특징값과 타겟 물질의 특징값을 정렬시키는 교차 어텐션을 수행하여, 약물에 대한 타겟 물질의 특징값을 계산할 수 있다. 또한, 프로세서(110)는 약물의 특징값과 타겟 물질의 특징값을 정렬시키는 교차 어텐션을 수행하여, 타겟 물질에 대한 약물의 특징값을 계산할 수 있다. 프로세서(110)는 약물 및 타겟 물질 각각에 대한 특징값을 계산함에 있어서, 상대 정보를 고려하기 때문에 보다 높은 신뢰도를 갖는 친화도를 예측할 수 있다. 이하에서는 컴퓨팅 장치(100)의 프로세서(110)가 제 2 뉴럴 네트워크를 이용하여, 특징값들 간의 교차 어텐션을 수행하는 방법에 대해 설명한다.

[79]

[80] 도 3은 본 개시의 몇몇 실시예에 따른 컴퓨팅 장치가 특징값들 간의 교차 어텐션을 수행하는 방법의 일례를 설명하기 위한 흐름도이다. 도 4는 본 개시의 몇몇 실시예에 따른 컴퓨팅 장치가 약물에 대한 타겟 물질의 특징값을 계산하는 방법의 일례를 설명하기 위한 흐름도이다. 도 5는 본 개시의 몇몇 실시예에 따른 컴퓨팅 장치가 타겟 물질에 대한 약물의 특징값을 계산하는 방법의 일례를 설명하기 위한 흐름도이다.

[81] 도 3을 참조하면, 컴퓨팅 장치(100)의 프로세서(110)는 제 2 뉴럴 네트워크에 포함된 제 1 서브 네트워크를 기초로 약물의 특징값과 타겟 물질의 특징값을 정렬시키는 교차 어텐션을 수행할 수 있다. 그리고, 프로세서(110)는 교차 어텐션을 수행한 결과를 기초로 약물에 대한 타겟 물질의 특징값을 계산할 수 있다(S121).

[82] 구체적으로, 도 4를 참조하면, 프로세서(110)는 제 1 뉴럴 네트워크에 포함된 제 1 서브 네트워크를 이용하여 추출된 약물의 특징값을 기초로, 약물에 대한 제 1 키(key) 및 제 1 벨류(value)를 생성할 수 있다. 그리고, 프로세서(110)는 제 1 뉴럴 네트워크에 포함된 제 2 서브 네트워크를 이용하여 추출된 타겟 물질의 특징값을 기초로, 타겟 물질에 대한 제 2 쿼리(query)를 생성할 수 있다(S1211). 좀 더 구체적으로, 프로세서(110)는 동일한 입력

$$d_{1:n_d}$$

를 가중치 행렬

$$W_K$$

,

$$W_V$$

로 변환하여 제 1 키 및 제 1 벨류를 구성할 수 있다. 여기서,

$$n_d$$

는 약물의 특징값의 순차 길이이고 아래 첨자 1은 약물의 특징값의 순차 범위를 나타낼 수 있다. 프로세서(110)는 생성된 가중치 행렬을 각각

$$K_d$$

,

$$V_d$$

로 표시할 수 있고,

$$K_d$$

,

V_d

는 입력

$d_{1:m_d}$

과 동일한 차원을 가질 수 있다. 프로세서(110)는 동일한 입력

$p_{1:m_p}$

를 가중치 행렬

W_Q

로 변환하여 제 2 쿼리를 구성할 수 있다. 여기서,

n_p

는 타겟 물질(protein)의 특징값의 순차 길이이고 아래 첨자 1은 타겟 물질의 특징값의 순차 범위를 나타낼 수 있다. 프로세서(110)는 생성된 가중치 행렬을 각각

Q_p

로 표시할 수 있고,

Q_p

는 입력

$d_{1:m_d}$

과 동일한 차원을 가질 수 있다.

[83] 프로세서(110)는 제 2 쿼리 및 제 1 키를 기초로 교차 어텐션 벡터를 생성할 수 있다(S1212).

[84] 구체적으로, 프로세서(110)는 타겟 물질의 전체 시퀀스에 대한 제 2 쿼리 및 제 1 키를 기초로 하는 확률값을 계산하여 교차 어텐션 벡터를 생성할 수 있다. 환언하자면, 프로세서(110)는 토큰 레벨 어텐션 대신 전체 시퀀스에 대한 컨텍스트 레벨 어텐션을 수행하여 컨텍스트 수준의 교차 어텐션 벡터를 생성할 수 있다.

[85] 일례로, 프로세서(110)는 하기의 수학적식을 통해 약물에 대한 타겟 물질의 교차 어텐션 벡터를 생성할 수 있다.

[86] [수식6]

$$a = \frac{1}{n_p} \sum_{i=1}^{n_p} \text{softmax}\left(\frac{Q_p K_d^T}{\sqrt{b}}\right)_{:,i}$$

[87] 여기서,

a

는 약물에 대한 타겟 물질의 교차 어텐션 벡터일 수 있다.

n_p

는 타겟 물질의 특징값의 순차 길이일 수 있다.

softmax

는 출력값에 대한 정규화를 수행하는 함수일 수 있다.

b

는 약물에 대한 특징값의 차원일 수 있다.

K_d

는 제 1 키를 나타내고,

Q_p

는 제 2 쿼리를 나타낼 수 있다.

- [88] 본 개시의 몇몇 실시예에 따르면, 확률값은 *sparsemax* 함수에 기초하여 결정될 수 있다. 실시예에 따라, 교차 어텐션 벡터는 전체 시퀀스에 대해 확률값을 계산할 수 있다. 시퀀스 대부분의 확률값은 0은 아니지만, 작은 값을 가지기 때문에 중요한 시퀀스에 대한 어텐션이 약화될 수 있다. 따라서, 본 개시에서는 제 2 뉴럴 네트워크가 중요한 시퀀스에 대해 어텐션을 수행할 수 있도록 어텐션 벡터를 희소하게 생성할 수 있다. 이를 위해, 프로세서(110)는 *softmax* 함수 대신 *sparsemax* 함수를 이용하여, 확률값을 결정할 수 있다. 여기서, *sparsemax* 함수는 행렬의 값이 대부분 0이 되도록 값을 생성하는 함수로서 작은 가중치는 중요하지 않은 것으로 간주하기 위한 함수일 수 있다. 따라서, 프로세서(110)는 *sparsemax* 함수를 이용하여, 하기의 수학적식을 통해 약물에 대한 타겟 물질의 교차 어텐션 벡터를 생성할 수 있다.

- [89] [수식7]

$$a = \frac{1}{n_p} \sum_{i=1}^{n_p} \text{sparsemax}\left(\frac{Q_p K_d^T}{\sqrt{b}}\right)_{:,i}$$

- [90] 프로세서(110)는 생성된 교차 어텐션 벡터를 제 1 밸류에 적용하여, 약물에 대한 타겟 물질의 특징값을 계산할 수 있다(S1213).

- [91] 구체적으로, 프로세서(110)는 교차 어텐션 벡터와 제 1 밸류 간의 요소별(element-wise) 연산을 수행할 수 있다. 여기서, 요소별 연산은 차원이 매칭되지 않을 경우, 브로드캐스팅 방식으로 연산을 수행하는 방법일 수 있다. 일례로, 요소별 연산은 행렬이나 다차원 배열을 처리할 수 있는 파이썬의 라이브러리 *Numpy*를 이용할 수 있다. 다만, 이에 한정되는 것은 아니다. 프로세서(110)는 요소 별 연산을 수행한 결과를 기초로 약물에 대한 타겟 물질의 특징값을 계산할 수 있다. 일례로, 프로세서(110)는 하기의 수학적식을 통해 교차 어텐션 벡터와 제 1

밸류 간의 요소별 연산을 수행하여 약물에 대한 타겟 물질의 특징값을 계산할 수 있다.

[92] [수식8]

$$V'_d = \text{softmax}(a) \odot V_d$$

[93] 여기서,

V_d

는 약물에 대한 타겟 물질의 특징값일 수 있고,

a

는 교차 어텐션 벡터일 수 있다. 그리고,

\odot

는 요소별 연산을 나타내는 기호이고,

V_d

는 제 1 밸류일 수 있다.

[94] 본 개시의 몇몇 실시예에 따르면, 프로세서(110)는 제 2 뉴럴 네트워크가 최소하게 어텐션 맵을 생성할 수 있도록 softmax 함수 대신 sparsemax 함수를 이용하여, 약물에 대한 타겟 물질의 특징값을 결정할 수 있다. 따라서, 프로세서(110)는 sparsemax 함수를 이용하여, 하기의 수학식을 통해 약물에 대한 타겟 물질의 특징값을 계산할 수 있다.

[95] [수식9]

$$V'_d = \text{sparsemax}(a) \odot V_d$$

[96] 상술한 과정을 통해 컴퓨팅 장치(100)의 프로세서(110)는 제 2 뉴럴 네트워크에 포함된 제 1 서브 네트워크를 기초로 약물의 특징값과 타겟 물질의 특징값을 정렬시키는 교차 어텐션을 수행할 수 있다. 그리고, 프로세서(110)는 교차 어텐션을 수행한 결과를 기초로 약물에 대한 타겟 물질의 특징값을 계산할 수 있다.

[97] 한편, 다시 도 3을 참조하면, 컴퓨팅 장치(100)의 프로세서(110)는 제 2 뉴럴 네트워크에 포함된 제 2 서브 네트워크를 기초로 약물의 특징값과 타겟 물질의 특징값을 정렬시키는 교차 어텐션을 수행할 수 있다. 그리고, 프로세서(110)는 교차 어텐션을 수행한 결과를 기초로 타겟 물질에 대한 약물의 특징값을 계산할 수 있다(S122).

[98] 구체적으로, 도 5를 참조하면, 프로세서(110)는 제 1 뉴럴 네트워크에 포함된 제 2 서브 네트워크를 이용하여 추출된 타겟 물질의 특징값을 기초로, 타겟에 대한 제 2 키 및 제 2 밸류를 생성할 수 있다. 그리고, 프로세서(110)는 제 1 뉴럴 네트워크에 포함된 제 1 서브 네트워크를 이용하여 추출된 약물의 특징값을

기초로, 약물에 대한 제 1 쿼리를 생성할 수 있다(S1221). 좀 더 구체적으로, 프로세서(110)는 동일한 입력

$P_{1:m_p}$

를 가중치 행렬

W_K

,

W_V

로 변환하여 제 2 키 및 제 2 벨류를 구성할 수 있다. 여기서,

n_p

는 타겟 물질의 특징값의 순차 길이이고 아래 첨자 1은 타겟 물질의 특징값의 순차 범위를 나타낼 수 있다. 프로세서(110)는 생성된 가중치 행렬을 각각

K_p

,

V_p

로 표시할 수 있고,

K_p

,

V_p

는 입력

$P_{1:m_p}$

과 동일한 차원을 가질 수 있다. 프로세서(110)는 동일한 입력

$d_{1:n_d}$

를 가중치 행렬

W_O

로 변환하여 제 2 쿼리를 구성할 수 있다. 여기서,

n_d

는 약물의 특징값의 순차 길이이고 아래 첨자 1은 약물의 특징값의 순차 범위를 나타낼 수 있다. 프로세서(110)는 생성된 가중치 행렬을 각각

Q_d

로 표시할 수 있고,

Q_d

는 입력

$P_{1:m_p}$

과 동일한 차원을 가질 수 있다.

[99] 프로세서(110)는 제 1 쿼리 및 제 2 키를 기초로 교차 어텐션 벡터를 생성할 수 있다(S1212).

[100] 구체적으로, 프로세서(110)는 약물의 전체 시퀀스에 대한 제 1 쿼리 및 제 2 키를 기초로 하는 확률값을 계산하여 교차 어텐션 벡터를 생성할 수 있다. 환언하자면, 프로세서(110)는 토큰 레벨 어텐션 대신 전체 시퀀스에 대한 컨텍스트 레벨 어텐션을 수행하여 컨텍스트 수준의 교차 어텐션 벡터를 생성할 수 있다.

[101] 일례로, 프로세서(110)는 하기의 수학식을 통해 타겟 물질에 대한 약물의 교차 어텐션 벡터를 생성할 수 있다.

[102] [수식10]

$$a = \frac{1}{n_d} \sum_{i=1}^{n_d} \text{softmax}\left(\frac{Q_d K_p^T}{\sqrt{b}}\right)_{:,i}$$

[103] 여기서,

a

는 타겟 물질에 대한 약물의 교차 어텐션 벡터일 수 있다.

n_d

는 약물의 특징값의 순차 길이일 수 있다.

softmax

는 출력값에 대한 정규화를 수행하는 함수일 수 있다.

b

는 타겟 물질에 대한 특징값의 차원일 수 있다.

K_p

는 제 2 키를 나타내고,

Q_d

는 제 1 쿼리를 나타낼 수 있다.

[104] 본 개시의 몇몇 실시예에 따르면, 확률값은 sparsemax 함수에 기초하여 결정될 수 있다. 실시예에 따라, 프로세서(110)는 sparsemax 함수를 이용하여, 하기의 수학식을 통해 타겟 물질에 대한 약물의 교차 어텐션 벡터를 생성할 수 있다.

[105] [수식11]

$$a = \frac{1}{n_d} \sum_{i=1}^{n_d} \text{sparsemax} \left(\frac{Q_d K_p^T}{\sqrt{b}} \right)_{:,i}$$

[106] 프로세서(110)는 생성된 교차 어텐션 벡터를 제 2 밸류에 적용하여, 타겟 물질에 대한 약물의 특징값을 계산할 수 있다(S1223).

[107] 구체적으로, 프로세서(110)는 교차 어텐션 벡터와 제 2 밸류 간의 요소 별 연산을 수행할 수 있다. 일례로, 프로세서(110)는 행렬이나 다차원 배열을 처리할 수 있는 파이썬의 라이브러리 Numpy를 이용하여 요소별 연산을 수행할 수 있다. 다만, 이에 한정되는 것은 아니다. 프로세서(110)는 요소 별 연산을 수행한 결과를 기초로 타겟 물질에 대한 약물의 특징값을 계산할 수 있다. 일례로, 프로세서(110)는 하기의 수학식을 통해 교차 어텐션 벡터와 제 2 밸류 간의 요소별 연산을 수행하여 타겟 물질에 대한 약물의 특징값을 계산할 수 있다.

[108] [수식12]

$$V'_p = \text{softmax}(a) \odot V_p$$

[109] 여기서,

$$V'_p$$

는 타겟 물질에 대한 약물의 특징값일 수 있고,

$$a$$

는 교차 어텐션 벡터일 수 있다. 그리고,

$$\odot$$

는 요소별 연산을 나타내는 기호이고,

$$V_p$$

는 제 2 밸류일 수 있다.

[110] 본 개시의 몇몇 실시예에 따르면, 프로세서(110)는 제 2 뉴럴 네트워크가 희소하게 어텐션 맵을 생성할 수 있도록 softmax 함수 대신 sparsemax 함수를 이용하여, 약물에 대한 타겟 물질의 특징값을 결정할 수 있다. 따라서, 프로세서(110)는 sparsemax 함수를 이용하여, 하기의 수학식을 통해 타겟 물질에 대한 약물의 특징값을 계산할 수 있다.

[111] [수식13]

$$V'_p = \text{sparsemax}(a) \odot V_p$$

[112] 상술한 과정을 통해 컴퓨팅 장치(100)의 프로세서(110)는 제 2 뉴럴 네트워크에 포함된 제 2 서브 네트워크를 기초로 약물의 특징값과 타겟 물질의 특징값을 정렬시키는 교차 어텐션을 수행할 수 있다. 그리고, 프로세서(110)는 교차

어텐션을 수행한 결과를 기초로 타겟 물질에 대한 약물의 특징값을 계산할 수 있다.

[113] 이하에서는 본 개시에 따른 약물 및 타겟 물질 간의 친화도를 예측하는 방법을 구현하기 위한 구체적인 프레임워크를 설명한다.

[114]

[115] 도 6은 본 개시의 몇몇 실시예에 따른 약물 및 타겟 물질 간의 친화도를 예측하는 방법을 구현하기 위한 프레임워크의 일례를 설명하기 위한 도면이다.

[116] 도 6을 참조하면, 약물 및 타겟 물질 간의 친화도를 예측하는 방법을 구현하기 위한 프레임워크(200)는 제 1 뉴럴 네트워크(210), 제 2 뉴럴 네트워크(220) 및 제 3 뉴럴 네트워크(230)를 포함할 수 있다. 다만, 이에 한정되는 것은 아니다.

[117] 실시예에 따라, 제 1 뉴럴 네트워크(210)는 제 1 서브 네트워크(211) 및 제 2 서브 네트워크(212)를 포함할 수 있다. 그리고, 프로세서(110)는 제 1 서브 네트워크(211)를 이용하여, 약물의 특징값을 추출하고(예를 들어,

$$d_1, d_2, d_3 \dots d_{n_d}$$

), 제 2 서브 네트워크(212)를 이용하여, 타겟 물질의 특징값(예를 들어,

$$p_1, p_2, p_3 \dots p_{n_p}$$

)을 추출할 수 있다.

[118] 제 2 뉴럴 네트워크(220)는 제 1 서브 네트워크(221) 및 제 2 서브 네트워크(222)를 포함할 수 있다. 그리고, 프로세서(110)는 제 1 서브 네트워크(221)를 이용하여, 약물에 대한 타겟 물질의 특징값을 정렬시켜(예를 들어,

$$d'_1, d'_2, d'_3 \dots d'_{n_d}$$

) 특징값을 계산할 수 있다. 프로세서(110)는 제 2 서브 네트워크(222)를 이용하여, 타겟에 대한 약물의 특징값을 정렬시켜(예를 들어,

$$p'_1, p'_2, p'_3 \dots p'_{n_p}$$

) 특징값을 계산할 수 있다.

[119] 본 개시의 몇몇 실시예에 따르면, 제 2 뉴럴 네트워크(220)는 멀티 헤드 어텐션(2221)을 수행할 수 있다. 여기서, 멀티 헤드 어텐션(2221)은 복수개의 교차 어텐션을 병렬적으로 수행하여, 약물에 대한 타겟 물질의 특징값 및 타겟 물질에 대한 약물의 특징값을 계산하기 위한 어텐션일 수 있다.

[120] 제 3 뉴럴 네트워크(230)는 교차 어텐션을 수행한 결과를 기초로 약물 및 타겟 물질 간의 친화도를 예측할 수 있다. 구체적으로, 컴퓨팅 장치(100)의 프로세서(110)는 제 3 뉴럴 네트워크(230)를 이용하여, 교차 어텐션을 수행한 결과를 기초로 약물 및 타겟 물질 간의 친화도를 나타내는 친화도 점수

ŷ

를 결정할 수 있다.

[121] 종래의 딥 러닝 기반 약물-타겟 물질의 상호작용(drug-target interaction, DTI) 예측 방법은 비용 측면에서 유망한 결과를 보여 왔다. 그러나, 종래의 기술은 예측 결과의 해석 가능성과 더불어 약물과 타겟 물질 간의 특징값 수준으로 상호작용이 고려되지 못할 수 있다. 반면, 본 발명의 프레임워크에 따르면, 약물 및 타겟 물질 간의 명시적인 상호 작용을 구성하여, 약물 및 타겟 물질에 교차 어텐션이 수행될 수 있다. 또한, 본 개시에 따른 프레임워크는 약물에 대한 타겟 물질의 특징값을 계산하기 위한 게이팅 함수인 교차 어텐션 벡터를 통해 약물 및 타겟 물질의 전체 시퀀스에 대해 확률값을 계산할 수 있다. 더하여, 명시적 상호작용의 결과물로서 생성되는 어텐션 맵은 해석 가능한 요소로 작용될 수 있다. 또한, 본 발명에 따른 프레임워크는 돌연변이에 대해서도 민감한 반응을 나타낼 수 있다. 환언하자면, 본 발명에 따른 프레임 워크는 돌연변이 단백질을 타겟으로 하는 새로운 약물에 대한 친화도도 계산될 수 있다. 이하에서는 종래의 기술과의 비교를 통해 본 발명에 따른 약물 및 타겟 물질 간의 친화도를 예측하는 방법에 대해 설명한다.

[122]

[123] 도 7은 본 개시의 몇몇 실시예에 따른 본 발명과 종래의 기술과의 비교를 위한 도면이다.

[124] 도 7의 (a)는 종래의 기술을 사용하여 약물

(d, d')

및 타겟 물질

(p, p')

간의 쌍별(pair-wise) 유사성을 나타낸 결과일 수 있다. 여기서, 종래의 기술은 디코더 어텐션(Decoder Attention)일 수 있다. 도 7의 (a)에 도시된 그래프의 x축은 트레이닝 에폭(Training epochs)을 나타낼 수 있고, y 축은 코사인 유사도(Cosine Similarity)를 나타낼 수 있다. 제 1 선(310)은 약물 특징값

(d)

및 교차 어텐션이 수행되어 정렬된 약물 특징값

(d')

간의 코사인 유사도를 나타낼 수 있고, 제 2 선(320)은 약물 특징값

(d)

및 교차 어텐션이 수행되어 정렬된 타겟 물질 특징값

(p')

간의 코사인 유사도를 나타낼 수 있다. 제 3 선(330)은 타겟 물질 특징값

(p)

및 교차 어텐션이 수행되어 정렬된 타겟 특징값

(p')

간의 코사인 유사도를 나타낼 수 있고, 제 4 선(340)은 타겟 물질 특징값

(p)

및 교차 어텐션이 수행되어 정렬된 약물 특징값

(d')

간의 코사인 유사도를 나타낼 수 있다. 제 1 선(310) 내지 제 4 선(340)을 참조하면, 제 1 선(310) 내지 제 4 선(340)은 기울기 등이 차이가 크게 나지 않는 것을 확인할 수 있다. 이는 종래의 기술을 이용하여 어텐션을 수행하는 경우 약물 및 타겟 물질 간의 구별되는 특징값이 생성되지 못한다는 것을 나타낼 수 있다.

[125] 반면, 도 7의 (b)는 본 발명에 따른 약물 및 타겟 물질 간의 친화도를 예측하는 방법에 따라 약물

(d, d')

및 타겟 물질

(p, p')

간의 쌍별 유사성을 나타낸 결과일 수 있다. 도 7의 (b)에 도시된 그래프의 x축은 트레이닝 에폭을 나타낼 수 있고, y 축은 코사인 유사도를 나타낼 수 있다. 제 1 선(410)은 약물 특징값

(d)

및 교차 어텐션이 수행되어 정렬된 약물 특징값

(d')

간의 코사인 유사도를 나타낼 수 있고, 제 2 선(420)은 약물 특징값

(d)

및 교차 어텐션이 수행되어 정렬된 타겟 물질 특징값

(p')

간의 코사인 유사도를 나타낼 수 있다. 제 3 선(430)은 타겟 물질 특징값

(p)

및 교차 어텐션이 수행되어 정렬된 타겟 특징값

(p')

간의 코사인 유사도를 나타낼 수 있고, 제 4 선(440)은 타겟 물질 특징값

(p)

및 교차 어텐션이 수행되어 정렬된 약물 특징값

(d')

간의 코사인 유사도를 나타낼 수 있다. 제 1 선(410) 내지 제 4 선(440)을 참조하면, 제 1 선(410) 내지 제 4 선(440)은 도 7의 (a)에 도시된 제 1 선(310) 내지 제 4 선(440)에 비해 기울기 등이 차이가 크게 나는 것을 확인할 수 있다. 이는 본 발명에 따른 방법을 이용하여, 교차 어텐션을 수행하는 경우 약물 및 타겟 물질 간의 구별되는 특징값이 생성된다는 것을 나타낼 수 있다. 이는 컴퓨팅 장치(100)의 프로세서(110)가 제 3 뉴럴 네트워크를 이용하여 약물 및 타겟 물질 간의 친화도를 예측함에 있어서, 교차 어텐션을 통해 약물 및 타겟 물질 간의 구별되는 특징값을 이용할 수 있다는 의미일 수 있다. 따라서, 프로세서(110)에 의해 예측되는 친화도는 높은 신뢰도를 가질 수 있다.

[126]

[127] 도 8은 본 개시내용의 실시예들이 구현될 수 있는 예시적인 컴퓨팅 환경에 대한 일반적인 개략도를 도시한다.

[128] 본 개시내용이 일반적으로 하나 이상의 컴퓨터 상에서 실행될 수 있는 컴퓨터 실행가능 명령어와 관련하여 기술되었지만, 당업자라면 본 개시내용 기타 프로그램 모듈들과 결합되어 및/또는 하드웨어와 소프트웨어의 조합으로서 구현될 수 있다는 것을 잘 알 것이다.

[129] 일반적으로, 본 명세서에서의 모듈은 특정의 태스크를 수행하거나 특정의 추상 데이터 유형을 구현하는 루틴, 프로시저, 프로그램, 컴포넌트, 데이터 구조, 기타 등등을 포함한다. 또한, 당업자라면 본 개시의 방법이 단일-프로세서 또는 멀티프로세서 컴퓨터 시스템, 미니컴퓨터, 메인프레임 컴퓨터는 물론 퍼스널 컴퓨터, 핸드헬드 컴퓨팅 장치, 마이크로프로세서-기반 또는 프로그램가능 가전 제품, 기타 등등(이들 각각은 하나 이상의 연관된 장치와 연결되어 동작할 수 있음)을 비롯한 다른 컴퓨터 시스템 구성으로 실시될 수 있다는 것을 잘 알 것이다.

[130] 본 개시의 설명된 실시예들은 또한 어떤 태스크들이 통신 네트워크를 통해 연결되어 있는 원격 처리 장치들에 의해 수행되는 분산 컴퓨팅 환경에서 실시될 수 있다. 분산 컴퓨팅 환경에서, 프로그램 모듈은 로컬 및 원격 메모리 저장 장치 둘다에 위치할 수 있다.

[131] 컴퓨터는 통상적으로 다양한 컴퓨터 판독가능 매체를 포함한다. 컴퓨터에 의해 액세스 가능한 매체로서, 휘발성 및 비휘발성 매체, 일시적(transitory) 및 비일시적(non-transitory) 매체, 이동식 및 비-이동식 매체를 포함한다. 제한이 아닌 예로서, 컴퓨터 판독가능 매체는 컴퓨터 판독가능 저장 매체 및 컴퓨터

판독가능 전송 매체를 포함할 수 있다.

- [132] 컴퓨터 판독가능 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보를 저장하는 임의의 방법 또는 기술로 구현되는 휘발성 및 비휘발성 매체, 일시적 및 비-일시적 매체, 이동식 및 비이동식 매체를 포함한다. 컴퓨터 판독가능 저장 매체는 RAM, ROM, EEPROM, 플래시 메모리 또는 기타 메모리 기술, CD-ROM, DVD(digital video disk) 또는 기타 광 디스크 저장 장치, 자기 카세트, 자기 테이프, 자기 디스크 저장 장치 또는 기타 자기 저장 장치, 또는 컴퓨터에 의해 액세스될 수 있고 원하는 정보를 저장하는 데 사용될 수 있는 임의의 기타 매체를 포함하지만, 이에 한정되지 않는다.
- [133] 컴퓨터 판독가능 전송 매체는 통상적으로 반송파(carrier wave) 또는 기타 전송 메커니즘(transport mechanism)과 같은 피변조 데이터 신호(modulated data signal)에 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터등을 구현하고 모든 정보 전달 매체를 포함한다. 피변조 데이터 신호라는 용어는 신호 내에 정보를 인코딩하도록 그 신호의 특성들 중 하나 이상을 설정 또는 변경시킨 신호를 의미한다. 제한이 아닌 예로서, 컴퓨터 판독가능 전송 매체는 유선 네트워크 또는 직접 배선 접속(direct-wired connection)과 같은 유선 매체, 그리고 음향, RF, 적외선, 기타 무선 매체와 같은 무선 매체를 포함한다. 상술된 매체들 중 임의의 것의 조합도 역시 컴퓨터 판독가능 전송 매체의 범위 안에 포함되는 것으로 한다.
- [134] 컴퓨터(1102)를 포함하는 본 개시의 여러가지 측면들을 구현하는 예시적인 환경(1100)이 나타내어져 있으며, 컴퓨터(1102)는 처리 장치(1104), 시스템 메모리(1106) 및 시스템 버스(1108)를 포함한다. 시스템 버스(1108)는 시스템 메모리(1106)(이에 한정되지 않음)를 비롯한 시스템 컴포넌트들을 처리 장치(1104)에 연결시킨다. 처리 장치(1104)는 다양한 상용 프로세서들 중 임의의 프로세서일 수 있다. 듀얼 프로세서 및 기타 멀티프로세서 아키텍처도 역시 처리 장치(1104)로서 이용될 수 있다.
- [135] 시스템 버스(1108)는 메모리 버스, 주변장치 버스, 및 다양한 상용 버스 아키텍처 중 임의의 것을 사용하는 로컬 버스에 추가적으로 상호 연결될 수 있는 몇 가지 유형의 버스 구조 중 임의의 것일 수 있다. 시스템 메모리(1106)는 판독 전용 메모리(ROM)(1110) 및 랜덤 액세스 메모리(RAM)(1112)를 포함한다. 기본 입/출력 시스템(BIOS)은 ROM, EPROM, EEPROM 등의 비휘발성 메모리(1110)에 저장되며, 이 BIOS는 시동 중과 같은 때에 컴퓨터(1102) 내의 구성요소들 간에 정보를 전송하는 일을 돕는 기본적인 루틴을 포함한다. RAM(1112)은 또한 데이터를 캐싱하기 위한 정적 RAM 등의 고속 RAM을 포함할 수 있다.
- [136] 컴퓨터(1102)는 또한 내장형 하드 디스크 드라이브(HDD)(1114)(예를 들어, EIDE, SATA)—이 내장형 하드 디스크 드라이브(1114)는 또한 적당한 새시(도시 생략) 내에서 외장형 용도로 구성될 수 있음—, 자기 플로피 디스크

드라이브(FDD)(1116)(예를 들어, 이동식 디스켓(1118)으로부터 판독을 하거나 그에 기록을 하기 위한 것임), 및 광 디스크 드라이브(1120)(예를 들어, CD-ROM 디스크(1122)를 판독하거나 DVD 등의 기타 고용량 광 매체로부터 판독을 하거나 그에 기록을 하기 위한 것임)를 포함한다. 하드 디스크 드라이브(1114), 자기 디스크 드라이브(1116) 및 광 디스크 드라이브(1120)는 각각 하드 디스크 드라이브 인터페이스(1124), 자기 디스크 드라이브 인터페이스(1126) 및 광 드라이브 인터페이스(1128)에 의해 시스템 버스(1108)에 연결될 수 있다. 외장형 드라이브 구현을 위한 인터페이스(1124)는 예를 들어, USB(Universal Serial Bus) 및 IEEE 1394 인터페이스 기술 중 적어도 하나 또는 그 둘 다를 포함한다.

- [137] 이들 드라이브 및 그와 연관된 컴퓨터 판독가능 매체는 데이터, 데이터 구조, 컴퓨터 실행가능 명령어, 기타 등등의 비휘발성 저장을 제공한다. 컴퓨터(1102)의 경우, 드라이브 및 매체는 임의의 데이터를 적당한 디지털 형식으로 저장하는 것에 대응한다. 상기에서의 컴퓨터 판독가능 저장 매체에 대한 설명이 HDD, 이동식 자기 디스크, 및 CD 또는 DVD 등의 이동식 광 매체를 언급하고 있지만, 당업자라면 zip 드라이브(zip drive), 자기 카세트, 플래쉬 메모리 카드, 카트리지, 기타 등등의 컴퓨터에 의해 판독가능한 다른 유형의 저장 매체도 역시 예시적인 운영 환경에서 사용될 수 있으며 또 임의의 이러한 매체가 본 개시의 방법들을 수행하기 위한 컴퓨터 실행가능 명령어를 포함할 수 있다는 것을 잘 알 것이다.
- [138] 운영 체제(1130), 하나 이상의 애플리케이션 프로그램(1132), 기타 프로그램 모듈(1134) 및 프로그램 데이터(1136)를 비롯한 다수의 프로그램 모듈이 드라이브 및 RAM(1112)에 저장될 수 있다. 운영 체제, 애플리케이션, 모듈 및/또는 데이터의 전부 또는 그 일부분이 또한 RAM(1112)에 캐싱될 수 있다. 본 개시가 여러가지 상업적으로 이용가능한 운영 체제 또는 운영 체제들의 조합에서 구현될 수 있다는 것을 잘 알 것이다.
- [139] 사용자는 하나 이상의 유선/무선 입력 장치, 예를 들어, 키보드(1138) 및 마우스(1140) 등의 포인팅 장치를 통해 컴퓨터(1102)에 명령 및 정보를 입력할 수 있다. 기타 입력 장치(도시 생략)로는 마이크, IR 리모콘, 조이스틱, 게임 패드, 스타일러스 펜, 터치 스크린, 기타 등등이 있을 수 있다. 이들 및 기타 입력 장치가 종종 시스템 버스(1108)에 연결되어 있는 입력 장치 인터페이스(1142)를 통해 처리 장치(1104)에 연결되지만, 병렬 포트, IEEE 1394 직렬 포트, 게임 포트, USB 포트, IR 인터페이스, 기타 등등의 기타 인터페이스에 의해 연결될 수 있다.
- [140] 모니터(1144) 또는 다른 유형의 디스플레이 장치도 역시 비디오 어댑터(1146) 등의 인터페이스를 통해 시스템 버스(1108)에 연결된다. 모니터(1144)에 부가하여, 컴퓨터는 일반적으로 스피커, 프린터, 기타 등등의 기타 주변 출력 장치(도시 생략)를 포함한다.
- [141] 컴퓨터(1102)는 유선 및/또는 무선 통신을 통한 원격 컴퓨터(들)(1148) 등의 하나 이상의 원격 컴퓨터로의 논리적 연결을 사용하여 네트워크화된 환경에서

동작할 수 있다. 원격 컴퓨터(들)(1148)는 워크스테이션, 서버 컴퓨터, 라우터, 퍼스널 컴퓨터, 휴대용 컴퓨터, 마이크로프로세서-기반 오락 기기, 피어 장치 또는 기타 통상의 네트워크 노드일 수 있으며, 일반적으로 컴퓨터(1102)에 대해 기술된 구성요소들 중 다수 또는 그 전부를 포함하지만, 간략함을 위해, 메모리 저장 장치(1150)만이 도시되어 있다. 도시되어 있는 논리적 연결은 근거리 통신망(LAN)(1152) 및/또는 더 큰 네트워크, 예를 들어, 원거리 통신망(WAN)(1154)에의 유선/무선 연결을 포함한다. 이러한 LAN 및 WAN 네트워킹 환경은 사무실 및 회사에서 일반적인 것이며, 인트라넷 등의 전사적 컴퓨터 네트워크(enterprise-wide computer network)를 용이하게 해주며, 이들 모두는 전세계 컴퓨터 네트워크, 예를 들어, 인터넷에 연결될 수 있다.

- [142] LAN 네트워킹 환경에서 사용될 때, 컴퓨터(1102)는 유선 및/또는 무선 통신 네트워크 인터페이스 또는 어댑터(1156)를 통해 로컬 네트워크(1152)에 연결된다. 어댑터(1156)는 LAN(1152)에의 유선 또는 무선 통신을 용이하게 해줄 수 있으며, 이 LAN(1152)은 또한 무선 어댑터(1156)와 통신하기 위해 그에 설치되어 있는 무선 액세스 포인트를 포함하고 있다. WAN 네트워킹 환경에서 사용될 때, 컴퓨터(1102)는 모뎀(1158)을 포함할 수 있거나, WAN(1154) 상의 통신 서버에 연결되거나, 또는 인터넷을 통하는 등, WAN(1154)을 통해 통신을 설정하는 기타 수단을 갖는다. 내장형 또는 외장형 및 유선 또는 무선 장치일 수 있는 모뎀(1158)은 직렬 포트 인터페이스(1142)를 통해 시스템 버스(1108)에 연결된다. 네트워크화된 환경에서, 컴퓨터(1102)에 대해 설명된 프로그램 모듈들 또는 그의 일부분이 원격 메모리/저장 장치(1150)에 저장될 수 있다. 도시된 네트워크 연결이 예시적인 것이며 컴퓨터들 사이에 통신 링크를 설정하는 기타 수단이 사용될 수 있다는 것을 잘 알 것이다.
- [143] 컴퓨터(1102)는 무선 통신으로 배치되어 동작하는 임의의 무선 장치 또는 개체, 예를 들어, 프린터, 스캐너, 데스크톱 및/또는 휴대용 컴퓨터, PDA(portable data assistant), 통신 위성, 무선 검출가능 태그와 연관된 임의의 장비 또는 장소, 및 전화와 통신을 하는 동작을 한다. 이것은 적어도 Wi-Fi 및 블루투스 무선 기술을 포함한다. 따라서, 통신은 종래의 네트워크에서와 같이 미리 정의된 구조이거나 단순히 적어도 2개의 장치 사이의 애드혹 통신(ad hoc communication)일 수 있다.
- [144] Wi-Fi(Wireless Fidelity)는 유선 없이도 인터넷 등으로의 연결을 가능하게 해준다. Wi-Fi는 이러한 장치, 예를 들어, 컴퓨터가 실내에서 및 실외에서, 즉 지지국의 통화권 내의 아무 곳에서도 데이터를 전송 및 수신할 수 있게 해주는 셀 전화와 같은 무선 기술이다. Wi-Fi 네트워크는 안전하고 신뢰성 있으며 고속인 무선 연결을 제공하기 위해 IEEE 802.11(a,b,g, 기타)이라고 하는 무선 기술을 사용한다. 컴퓨터를 서로에, 인터넷에 및 유선 네트워크(IEEE 802.3 또는 이더넷을 사용함)에 연결시키기 위해 Wi-Fi가 사용될 수 있다. Wi-Fi 네트워크는 비인가 2.4 및 5 GHz 무선 대역에서, 예를 들어, 11Mbps(802.11a) 또는 54

Mbps(802.11b) 데이터 레이트로 동작하거나, 양 대역(듀얼 대역)을 포함하는 제품에서 동작할 수 있다.

[145] 본 개시의 기술 분야에서 통상의 지식을 가진 자는 여기에 개시된 실시예들과 관련하여 설명된 다양한 예시적인 논리 블록들, 모듈들, 프로세서들, 수단들, 회로들 및 알고리즘 단계들이 전자 하드웨어, (편의를 위해, 여기에서 "소프트웨어"로 지칭되는) 다양한 형태들의 프로그램 또는 설계 코드 또는 이들 모두의 결합에 의해 구현될 수 있다는 것을 이해할 것이다. 하드웨어 및 소프트웨어의 이러한 상호 호환성을 명확하게 설명하기 위해, 다양한 예시적인 컴포넌트들, 블록들, 모듈들, 회로들 및 단계들이 이들의 기능과 관련하여 위에서 일반적으로 설명되었다. 이러한 기능이 하드웨어 또는 소프트웨어로서 구현되는지 여부는 특정한 애플리케이션 및 전체 시스템에 대하여 부과되는 설계 제약들에 따라 좌우된다. 본 개시의 기술 분야에서 통상의 지식을 가진 자는 각각의 특정한 애플리케이션에 대하여 다양한 방식들로 설명된 기능을 구현할 수 있으나, 이러한 구현 결정들은 본 개시의 범위를 벗어나는 것으로 해석되어서는 안 될 것이다.

[146] 여기서 제시된 다양한 실시예들은 방법, 장치, 또는 표준 프로그래밍 및/또는 엔지니어링 기술을 사용한 제조 물품(article)으로 구현될 수 있다. 용어 "제조 물품"은 임의의 컴퓨터-판독가능 장치로부터 액세스 가능한 컴퓨터 프로그램 또는 매체(media)를 포함한다. 예를 들어, 컴퓨터-판독가능 저장 매체는 자기 저장 장치(예를 들면, 하드 디스크, 플로피 디스크, 자기 스트림, 등), 광학 디스크(예를 들면, CD, DVD, 등), 스마트 카드, 및 플래쉬 메모리 장치(예를 들면, EEPROM, 카드, 스틱, 키 드라이브, 등)를 포함하지만, 이들로 제한되는 것은 아니다. 용어 "기계-판독가능 매체"는 명령(들) 및/또는 데이터를 저장, 보유, 및/또는 전달할 수 있는 무선 채널 및 다양한 다른 매체를 포함하지만, 이들로 제한되는 것은 아니다.

[147]

[148] 제시된 실시예들에 대한 설명은 임의의 본 개시의 기술 분야에서 통상의 지식을 가진 자가 본 개시를 이용하거나 또는 실시할 수 있도록 제공된다. 이러한 실시예들에 대한 다양한 변형들은 본 개시의 기술 분야에서 통상의 지식을 가진 자에게 명백할 것이며, 여기에 정의된 일반적인 원리들은 본 개시의 범위를 벗어남이 없이 다른 실시예들에 적용될 수 있다. 그리하여, 본 개시는 여기에 제시된 실시예들로 한정되는 것이 아니라, 여기에 제시된 원리들 및 신규한 특징들과 일관되는 최광의의 범위에서 해석되어야 할 것이다.

청구범위

- [청구항 1] 적어도 하나의 프로세서를 포함하는 컴퓨팅 장치에 의해 수행되는 약물 및 타겟 물질 간의 친화도(affinity)를 예측하는 방법으로서, 제 1 뉴럴 네트워크를 이용하여, 상기 약물 및 상기 타겟 물질 각각의 특징값을 추출하는 단계; 제 2 뉴럴 네트워크를 이용하여, 상기 특징값들 간의 교차 어텐션(cross attention)을 수행하는 단계; 및 제 3 뉴럴 네트워크를 이용하여, 상기 교차 어텐션을 수행한 결과를 기초로 상기 약물 및 상기 타겟 물질 간의 친화도를 예측하는 단계; 를 포함하는, 약물 및 타겟 물질 간의 친화도를 예측하는 방법.
- [청구항 2] 제 1 항에 있어서, 상기 제 1 뉴럴 네트워크를 이용하여, 상기 약물 및 상기 타겟 물질 각각의 특징값을 추출하는 단계는, 상기 제 1 뉴럴 네트워크에 포함된 제 1 서브 네트워크를 이용하여, 상기 약물의 특징값을 추출하는 단계; 및 상기 제 1 뉴럴 네트워크에 포함된 제 2 서브 네트워크를 이용하여, 상기 타겟 물질의 특징값을 추출하는 단계; 를 포함하는, 약물 및 타겟 물질 간의 친화도를 예측하는 방법.
- [청구항 3] 제 1 항에 있어서, 상기 제 2 뉴럴 네트워크를 이용하여, 상기 특징값들 간의 교차 어텐션(cross attention)을 수행하는 단계는, 상기 제 2 뉴럴 네트워크에 포함된 제 1 서브 네트워크를 기초로 상기 약물의 특징값과 상기 타겟 물질의 특징값을 정렬시키는 교차 어텐션을 수행하여, 상기 약물에 대한 상기 타겟 물질의 특징값을 계산하는 단계; 및 상기 제 2 뉴럴 네트워크에 포함된 제 2 서브 네트워크를 기초로 상기 약물의 특징값과 상기 타겟 물질의 특징값을 정렬시키는 교차 어텐션을 수행하여, 상기 타겟 물질에 대한 상기 약물의 특징값을 계산하는 단계; 를 포함하는, 약물 및 타겟 물질 간의 친화도를 예측하는 방법.
- [청구항 4] 제 3 항에 있어서, 상기 제 2 뉴럴 네트워크에 포함된 제 1 서브 네트워크를 기초로 상기 약물의 특징값과 상기 타겟 물질의 특징값을 정렬시키는 교차 어텐션을 수행하여, 상기 약물에 대한 상기 타겟 물질의 특징값을 계산하는 단계는, 상기 약물의 상기 특징값을 기초로, 상기 약물에 대한 제 1 키 및 제 1

벨류를 생성하고, 상기 타겟 물질의 상기 특징값을 기초로, 상기 타겟 물질에 대한 제 2 쿼리를 생성하는 단계;
 상기 제 2 쿼리 및 상기 제 1 키를 기초로 교차 어텐션 벡터를 생성하는 단계; 및
 상기 교차 어텐션 벡터를 상기 제 1 벨류에 적용하여, 상기 약물에 대한 상기 타겟 물질의 특징값을 계산하는 단계;
 를 포함하는,
 약물 및 타겟 물질 간의 친화도를 예측하는 방법.

[청구항 5] 제 4 항에 있어서,
 상기 제 2 쿼리 및 상기 제 1 키를 기초로 교차 어텐션 벡터를 생성하는 단계는,
 상기 타겟 물질의 전체 시퀀스에 대한 상기 제 2 쿼리 및 상기 제 1 키를 기초로 하는 확률값을 계산하여 상기 교차 어텐션 벡터를 생성하는 단계;
 를 포함하는,
 약물 및 타겟 물질 간의 친화도를 예측하는 방법.

[청구항 6] 제 5 항에 있어서,
 상기 확률값은,
 sparsemax 함수에 기초하여 결정되는,
 약물 및 타겟 물질 간의 친화도를 예측하는 방법.

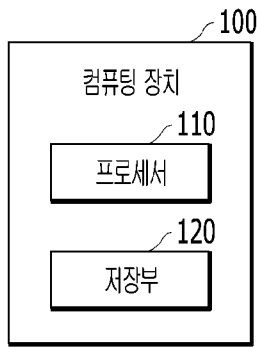
[청구항 7] 제 4 항에 있어서,
 상기 교차 어텐션 벡터를 상기 제 1 벨류에 적용하여, 상기 약물에 대한 상기 타겟 물질의 특징값을 계산하는 단계는,
 상기 교차 어텐션 벡터와 상기 제 1 벨류 간의 요소 별(element-wise) 연산을 수행하는 단계; 및
 상기 요소 별 연산을 수행한 결과를 기초로 상기 약물에 대한 상기 타겟 물질의 특징값을 계산하는 단계;
 를 포함하는,
 약물 및 타겟 물질 간의 친화도를 예측하는 방법.

[청구항 8] 제 3 항에 있어서,
 상기 제 2 뉴럴 네트워크에 포함된 제 2 서브 네트워크를 기초로 상기 약물의 특징값과 상기 타겟 물질의 특징값을 정렬시키는 교차 어텐션을 수행하여, 상기 타겟 물질에 대한 상기 약물의 특징값을 계산하는 단계는,
 상기 타겟 물질의 상기 특징값을 기초로, 상기 타겟 물질에 대한 제 2 키 및 제 2 벨류를 생성하고, 상기 약물의 상기 특징값을 기초로, 상기 약물에 대한 제 1 쿼리를 생성하는 단계;
 상기 제 1 쿼리 및 상기 제 2 키를 기초로 교차 어텐션 벡터를 생성하는 단계; 및
 상기 교차 어텐션 벡터를 상기 제 2 벨류에 적용하여, 상기 타겟 물질에

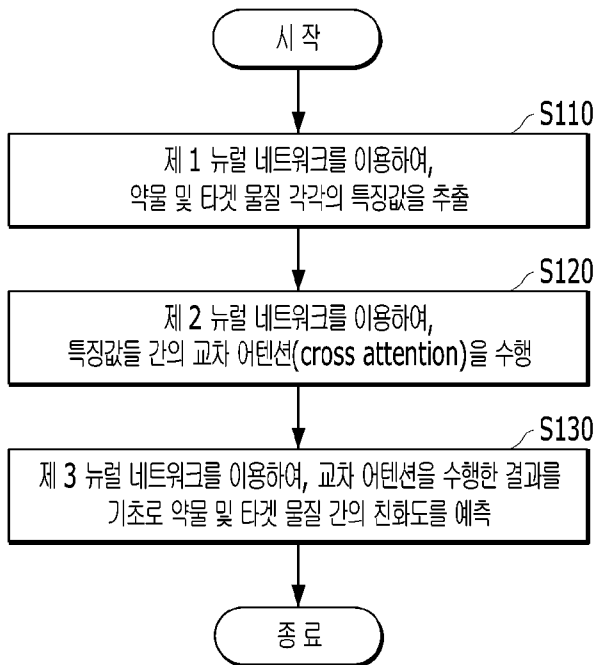
- 대한 상기 약물의 특징값을 계산하는 단계;
 를 포함하는,
 약물 및 타겟 물질 간의 친화도를 예측하는 방법.
- [청구항 9] 제 8 항에 있어서,
 상기 제 1 쿼리 및 상기 제 2 키를 기초로 교차 어텐션 벡터를 생성하는 단계는,
 상기 약물의 전체 시퀀스에 대한 상기 제 1 쿼리 및 상기 제 2 키를 기초로 하는 확률값을 계산하여 상기 교차 어텐션 벡터를 생성하는 단계;
 를 포함하는,
 약물 및 타겟 물질 간의 친화도를 예측하는 방법.
- [청구항 10] 제 9 항에 있어서,
 상기 확률값은,
 sparsemax 함수에 기초하여 결정되는,
 약물 및 타겟 물질 간의 친화도를 예측하는 방법.
- [청구항 11] 제 8 항에 있어서,
 상기 교차 어텐션 벡터를 상기 제 2 밸류에 적용하여, 상기 타겟 물질에 대한 상기 약물의 특징값을 계산하는 단계는,
 상기 교차 어텐션 벡터와 상기 제 2 밸류 간의 요소 별 연산을 수행하는 단계; 및
 상기 요소 별 연산을 수행한 결과를 기초로 상기 타겟 물질에 대한 상기 약물의 특징값을 계산하는 단계;
 를 포함하는,
 약물 및 타겟 물질 간의 친화도를 예측하는 방법.
- [청구항 12] 컴퓨터 판독가능 저장 매체에 저장된 컴퓨터 프로그램으로서, 상기 컴퓨터 프로그램은 하나 이상의 프로세서에서 실행되는 경우, 약물 및 타겟 물질 간의 친화도를 예측하기 위한 방법을 수행하도록 하며, 상기 방법은:
 제 1 뉴럴 네트워크를 이용하여, 상기 약물 및 상기 타겟 물질 각각의 특징값을 추출하는 단계;
 제 2 뉴럴 네트워크를 이용하여, 상기 특징값들 간의 교차 어텐션(cross attention)을 수행하는 단계; 및
 제 3 뉴럴 네트워크를 이용하여, 상기 교차 어텐션을 수행한 결과를 기초로 상기 약물 및 상기 타겟 물질 간의 친화도를 예측하는 단계;
 를 포함하는,
 컴퓨터 판독가능 저장 매체에 저장된 컴퓨터 프로그램.
- [청구항 13] 약물 및 타겟 물질 간의 친화도를 예측하는 컴퓨팅 장치로서,
 적어도 하나의 코어를 포함하는 프로세서; 및
 약물 및 타겟 물질과 관련된 정보를 저장하는 저장부;

를 포함하고,
상기 프로세서는,
제 1 뉴럴 네트워크를 이용하여, 상기 약물 및 상기 타겟 물질 각각의
특징값을 추출하고,
제 2 뉴럴 네트워크를 이용하여, 상기 특징값들 간의 교차 어텐션(cross
attention)을 수행하고, 그리고
제 3 뉴럴 네트워크를 이용하여, 상기 교차 어텐션을 수행한 결과를
기초로 상기 약물 및 상기 타겟 물질 간의 친화도를 예측하는,
장치.

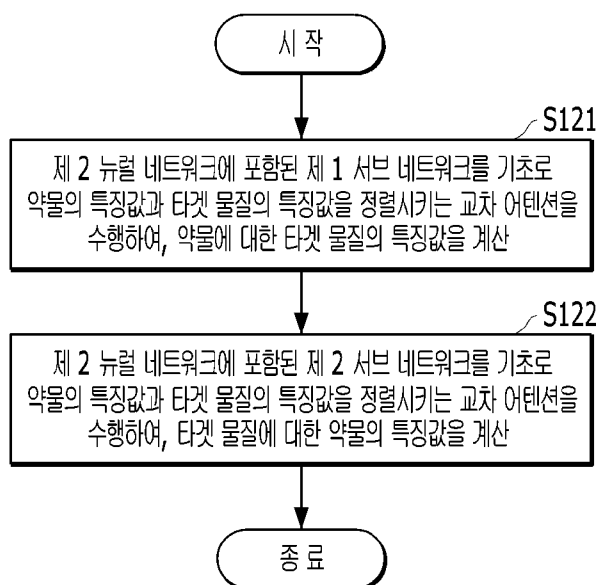
[도 1]



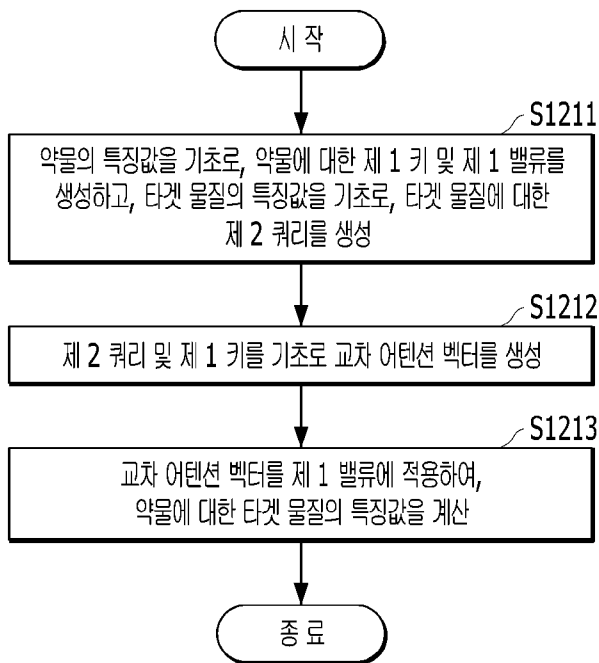
[도 2]



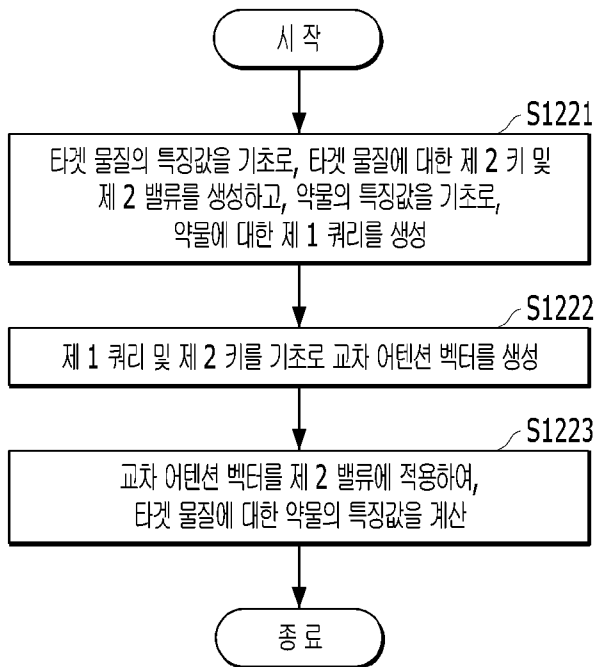
[도 3]



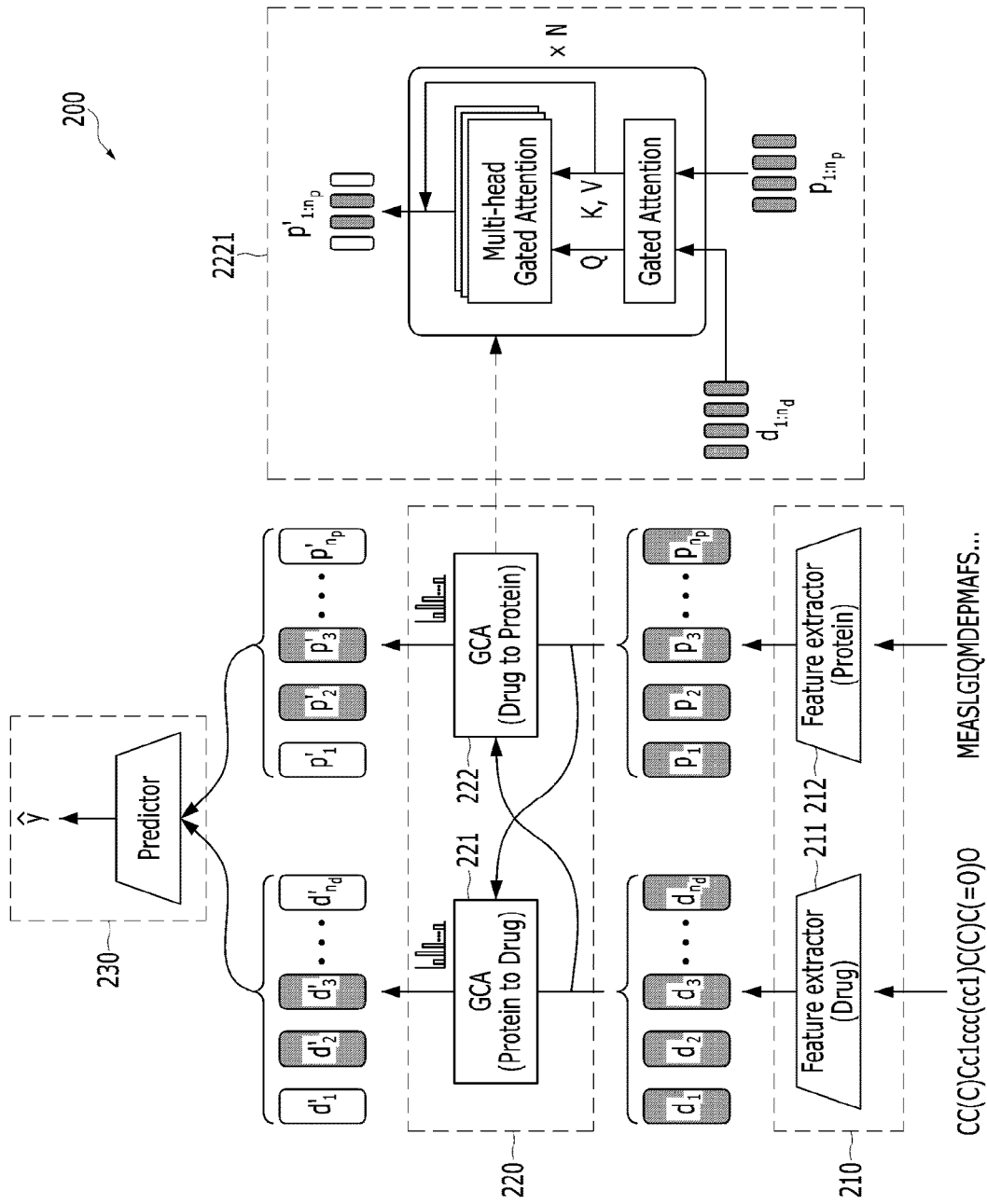
[도4]



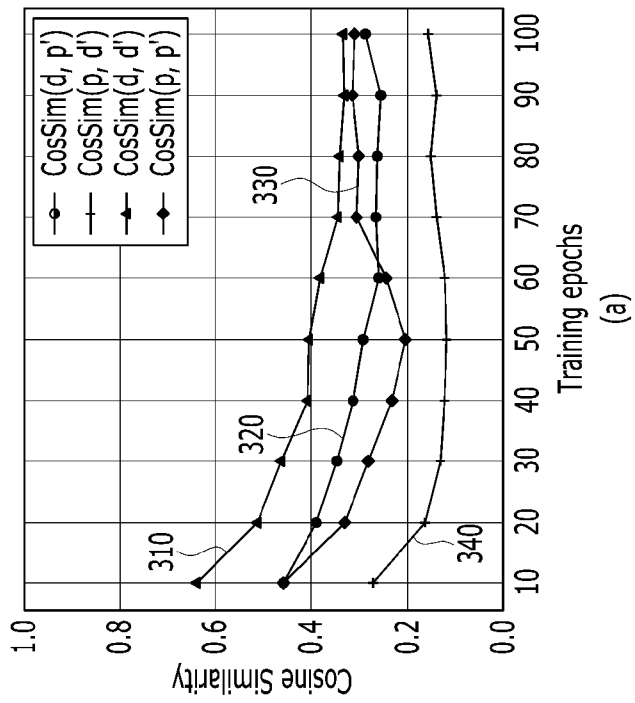
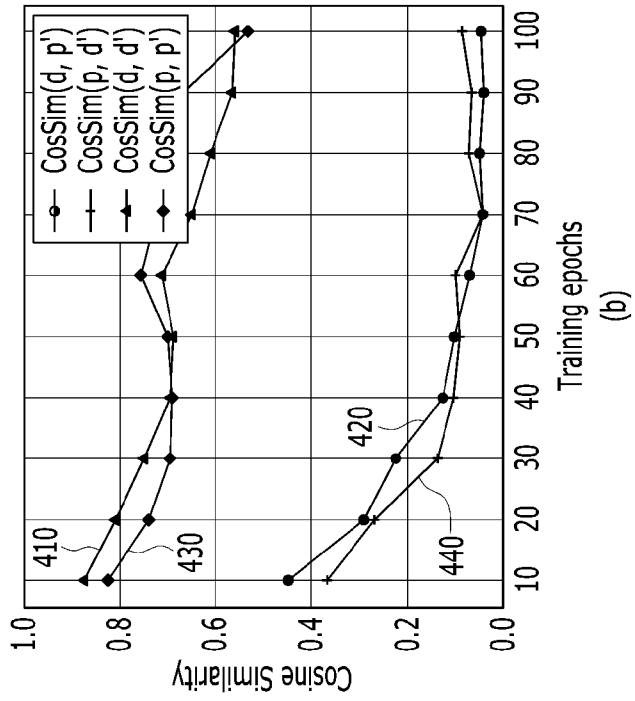
[도5]



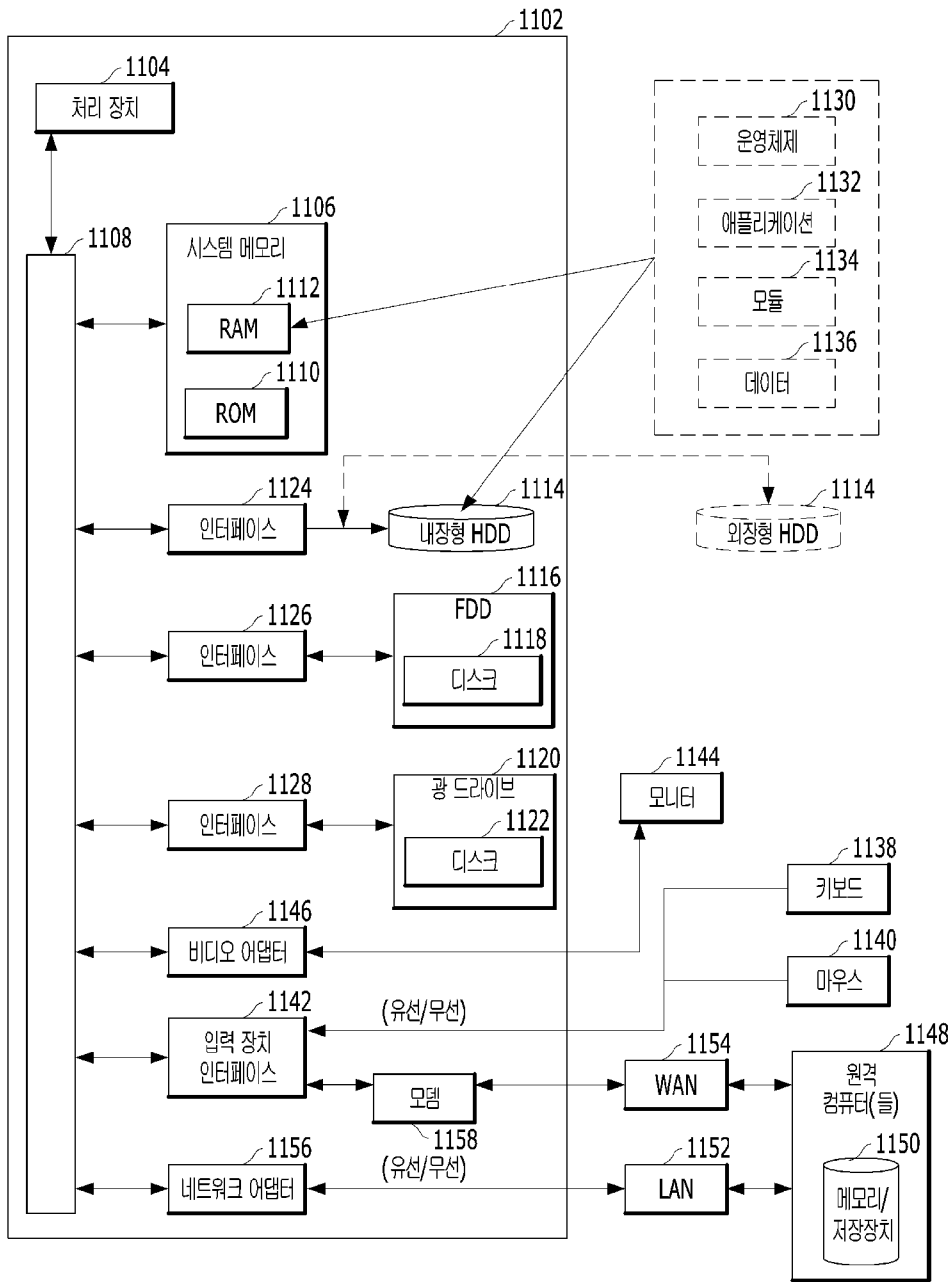
[도6]



[도7]



[도8]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2022/003582

A. CLASSIFICATION OF SUBJECT MATTER		
G16C 20/30(2019.01)i; G16C 20/60(2019.01)i; G16C 20/70(2019.01)i; G16C 20/40(2019.01)i; G06N 3/08(2006.01)i; G06N 3/04(2006.01)i		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols) G16C 20/30(2019.01); G06N 3/00(2006.01); G16B 25/10(2019.01); G16B 40/00(2019.01); G16B 5/00(2019.01)		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Korean utility models and applications for utility models: IPC as above Japanese utility models and applications for utility models: IPC as above		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) eKOMPASS (KIPO internal) & keywords: 친화도(affinity), 교차 어텐션(cross attention), 뉴럴 네트워크(neural network)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X A	KOYAMA, K. et al. Cross Attention DTI: Drug-Target Interaction Prediction with Cross Attention module in the Blind Evaluation Setup. BIODDD. 24 August 2020. Retrieved from https://biokdd.org/biokdd20/camera_ready/17CrossAttentionDTI_BIODDD2020.pdf . See abstract; 1. INTRODUCTION; 2.2 Machine Learning Based DTI Prediction and 4. EXPERIMENT; and figures 2-3.	1,12-13 2-11
A	KR 10-2020-0017653 A (GWANGJU INSTITUTE OF SCIENCE AND TECHNOLOGY) 19 February 2020 (2020-02-19) See entire document.	1-13
A	CN 112037856 A (HUAZHONG AGRICULTURAL UNIVERSITY) 04 December 2020 (2020-12-04) See entire document.	1-13
A	KR 10-2020-0126720 A (SEOUL NATIONAL UNIVERSITY R&DB FOUNDATION) 09 November 2020 (2020-11-09) See entire document.	1-13
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "D" document cited by the applicant in the international application "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 21 June 2022		Date of mailing of the international search report 21 June 2022
Name and mailing address of the ISA/KR Korean Intellectual Property Office Government Complex-Daejeon Building 4, 189 Cheongsaro, Seo-gu, Daejeon 35208 Facsimile No. +82-42-481-8578		Authorized officer Telephone No.

C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Retrieved from the Internet. Machine Learning for Healthcare. `37. Interpretable Frameworks for Drug-Target Interaction with Gated Cross Attention. 01 August 2021. Retrieved from https://www.youtube.com/watch?v=aljeI74WY9w . See entire document. * This document is a known document declaring exceptions to lack of novelty by the applicant.	1-13
X	KIM, Y. et al. An Interpretable Framework for Drug-Target Interaction with Gated Cross Attention. Proceedings of Machine Learning Research. 17 September 2021, vol. 149, pp. 1-16. See abstract; pages 3-6; and figures 1-2. * This document is a known document declaring exceptions to lack of novelty by the applicant.	1-13

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/KR2022/003582

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
KR 10-2020-0017653 A	19 February 2020	KR 10-2213670 B1	08 February 2021
CN 112037856 A	04 December 2020	None	
KR 10-2020-0126720 A	09 November 2020	KR 10-2022-0015479 A	08 February 2022
		KR 10-2355489 B1	24 January 2022

A. 발명이 속하는 기술분류(국제특허분류(IPC)) G16C 20/30(2019.01)i; G16C 20/60(2019.01)i; G16C 20/70(2019.01)i; G16C 20/40(2019.01)i; G06N 3/08(2006.01)i; G06N 3/04(2006.01)i		
B. 조사된 분야 조사된 최소문헌(국제특허분류를 기재) G16C 20/30(2019.01); G06N 3/00(2006.01); G16B 25/10(2019.01); G16B 40/00(2019.01); G16B 5/00(2019.01) 조사된 기술분야에 속하는 최소문헌 이외의 문헌 한국등록실용신안공보 및 한국공개실용신안공보: 조사된 최소문헌란에 기재된 IPC 일본등록실용신안공보 및 일본공개실용신안공보: 조사된 최소문헌란에 기재된 IPC 국제조사에 이용된 전산 데이터베이스(데이터베이스의 명칭 및 검색어(해당하는 경우)) eKOMPASS(특허청 내부 검색시스템) & 키워드: 친화도(affinity), 교차 어텐션(cross attention), 뉴럴 네트워크(neural network)		
C. 관련 문헌		
카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
X A	KOYAMA. K. 등, 'Cross Attention DTI: Drug-Target Interaction Prediction with Cross Attention module in the Blind Evaluation Setup', BIOKDD, 2020.08.24, https://biokdd.org/biokdd20/camera_ready/17CrossAttentionDTI_BIOKDD2020.pdf 초록; 1. INTRODUCTION; 2.2 Machine Learning Based DTI Prediction; 4. EXPERIMENT; 도면 2-3	1,12-13 2-11
A	KR 10-2020-0017653 A (광주과학기술원) 2020.02.19 전체 문헌	1-13
A	CN 112037856 A (HUAZHONG AGRICULTURAL UNIVERSITY) 2020.12.04 전체 문헌	1-13
A	KR 10-2020-0126720 A (서울대학교산학협력단) 2020.11.09 전체 문헌	1-13
<input checked="" type="checkbox"/> 추가 문헌이 C(계속)에 기재되어 있습니다. <input checked="" type="checkbox"/> 대응특허에 관한 별지를 참조하십시오.		
* 인용된 문헌의 특별 카테고리: "A" 특별히 관련이 없는 것으로 보이는 일반적인 기술수준을 정의한 문헌 "D" 본 국제출원에서 출원인이 인용한 문헌 "E" 국제출원일보다 빠른 출원일 또는 우선일을 가지나 국제출원일 이후에 공개된 선출원 또는 특허 문헌 "L" 우선권 주장에 의문을 제기하는 문헌 또는 다른 인용문헌의 공개일 또는 다른 특별한 이유(이유를 명시)를 밝히기 위하여 인용된 문헌 "O" 구두 개시, 사용, 전시 또는 기타 수단을 언급하고 있는 문헌 "P" 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌 "T" 국제출원일 또는 우선일 후에 공개된 문헌으로, 출원과 상충하지 않으며 발명의 기초가 되는 원리나 이론을 이해하기 위해 인용된 문헌 "X" 특별한 관련이 있는 문헌. 해당 문헌 하나만으로 청구된 발명의 신규성 또는 진보성이 없는 것으로 본다. "Y" 특별한 관련이 있는 문헌. 해당 문헌이 하나 이상의 다른 문헌과 조합하는 경우로 그 조합이 당업자에게 자명한 경우 청구된 발명은 진보성이 없는 것으로 본다. "&" 동일한 대응특허문헌에 속하는 문헌		
국제조사의 실제 완료일 2022년06월21일(21.06.2022)		국제조사보고서 발송일 2022년06월21일(21.06.2022)
ISA/KR의 명칭 및 우편주소 대한민국 특허청 (35208) 대전광역시 서구 청사로 189, 4동 (둔산동, 정부대전청사) 팩스 번호 +82-42-481-8578		심사관 전선애 전화번호 +82-42-481-8150

C. 관련 문헌		
카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
X	인터넷 자료, Machine Learning for Healthcare, `37. Interpretable Frameworks for Drug-Target Interaction with Gated Cross Attention`, 2021.08.01, https://www.youtube.com/watch?v=aljel74WY9w 전체 문헌 * 위 문헌은 출원인이 신규성 상실의 예외로서 선언한 공지문헌임.	1-13
X	KIM, Y. 등, `An Interpretable Framework for Drug-Target Interaction with Gated Cross Attention`, Proceedings of Machine Learning Research, 2021.09.17, 149권, 페이지 1-16 초록; 페이지 3-6; 도면 1-2 * 위 문헌은 출원인이 신규성 상실의 예외로서 선언한 공지문헌임.	1-13

국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
KR 10-2020-0017653 A	2020/02/19	KR 10-2213670 B1	2021/02/08
CN 112037856 A	2020/12/04	없음	
KR 10-2020-0126720 A	2020/11/09	KR 10-2022-0015479 A	2022/02/08
		KR 10-2355489 B1	2022/01/24