



공개특허 10-2021-0010505



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0010505  
(43) 공개일자 2021년01월27일

- (51) 국제특허분류(Int. Cl.)  
*G06N 3/04* (2006.01) *G06K 9/62* (2006.01)  
*G16H 30/40* (2018.01) *G16H 50/70* (2018.01)
- (52) CPC특허분류  
*G06N 3/0454* (2013.01)  
*G06K 9/6259* (2013.01)
- (21) 출원번호 10-2020-7035725
- (22) 출원일자(국제) 2019년05월08일  
심사청구일자 없음
- (85) 번역문제출일자 2020년12월11일
- (86) 국제출원번호 PCT/US2019/031255
- (87) 국제공개번호 WO 2019/221985  
국제공개일자 2019년11월21일
- (30) 우선권주장  
62/671,068 2018년05월14일 미국(US)  
62/678,074 2018년05월30일 미국(US)

- (71) 출원인  
퀀텀-에스아이 인코포레이티드  
미국 06437 코네티컷주 길포드 올드 화이트필드  
스트리트 530
- (72) 발명자  
로스버그, 조나단, 엠.  
미국 06437 코네티컷주 길포드 언카스 포인트 로  
드 215  
서, 우듯  
미국 02420 매사추세츠주 렉싱턴 코트야드 플레이  
스 8  
메이어, 마이클  
미국 06437 코네티컷주 길포드 시사이드 애비뉴  
93
- (74) 대리인  
양영준, 김연송, 백만기

전체 청구항 수 : 총 43 항

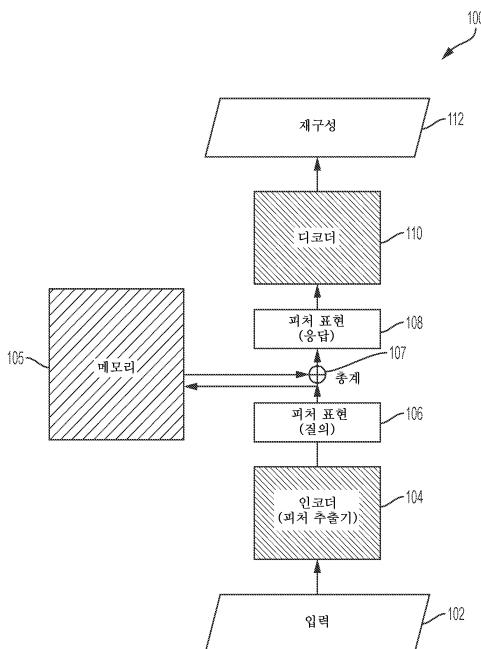
(54) 발명의 명칭 상이한 데이터 모달리티들에 대한 통계적 모델들을 단일화하기 위한 시스템들 및 방법들

### (57) 요약

제1 모달리티로부터의 입력 데이터, 및 제1 모달리티와는 상이한 제2 모달리티로부터의 입력 데이터를 포함하는 복수의 모달리티로부터의 입력 데이터를 수신하도록 구성된 다중-모달 통계적 모델을 이용하여 예측 태스크를 수행하기 위한 기술들이 개시된다. 이 기술들은, 다중-모달 통계적 모델의 복수의 컴포넌트 – 복수의 컴포넌트는,

(뒷면에 계속)

대 표 도 - 도1



제1 및 제2 모달리티에 대한 입력 데이터를 각각 처리하기 위한 제1 및 제2 인코더, 제1 및 제2 모달리티 임베딩, 합동-모달리티 표현, 및 예측기를 포함함— 각각의 파라미터들의 값들을 포함하는 다중-모달 통계적 모델을 명시하는 정보를 획득하는 단계; 제1 데이터 모달리티에 대한 제1 입력 데이터를 획득하는 단계; 제1 피처 벡터를 생성하기 위해 제1 입력 데이터를 제1 인코더에 제공하는 단계; 합동-모달리티 표현, 제1 모달리티 임베딩 및 제1 피처 벡터를 이용하여 제2 피처 벡터를 식별하는 단계; 및 예측기 및 제2 피처 벡터를 이용하여 예측 테스크에 대한 예측을 생성하는 단계를 포함한다.

## (52) CPC특허분류

*G06K 9/6269* (2013.01)

*G06N 3/0445* (2013.01)

*G06N 3/084* (2013.01)

*G06N 3/088* (2013.01)

*G16H 30/40* (2018.01)

*G16H 50/70* (2018.01)

---

## 명세서

### 청구범위

#### 청구항 1

제1 모달리티로부터의 입력 데이터 및 상기 제1 모달리티와는 상이한 제2 모달리티로부터의 입력 데이터를 포함한 복수의 모달리티로부터의 입력 데이터를 수신하도록 구성된 다중-모달 통계적 모델을 훈련시키기 위한 방법으로서,

상기 제1 모달리티에 대한 무라밸 훈련 데이터 및 상기 제2 모달리티에 대한 무라밸 훈련 데이터를 포함하는 무라밸 훈련 데이터에 액세스하는 단계;

상기 제1 모달리티에 대한 라벨링된 훈련 데이터 및 상기 제2 모달리티에 대한 라벨링된 훈련 데이터를 포함하는 라벨링된 훈련 데이터에 액세스하는 단계;

상기 다중-모달 통계적 모델을 2개의 스테이지로 훈련시키는 단계 – 상기 다중-모달 통계적 모델은 상기 제1 및 상기 제2 모달리티에 대한 입력 데이터를 각각 처리하기 위한 제1 및 제2 인코더, 제1 및 제2 모달리티 임베딩, 합동-모달리티 표현, 및 예측기를 포함한 복수의 컴포넌트를 포함하고, 상기 훈련시키는 단계는 :

자체 감독형 학습 기술 및 상기 무라밸 훈련 데이터를 이용하여 상기 제1 및 제2 모달리티 임베딩, 및 합동-모달리티 표현의 파라미터들에 대한 값을 추정함으로써 적어도 부분적으로 제1 훈련 스테이지를 수행하는 단계, 및

감독형 학습 기술 및 상기 라벨링된 훈련 데이터를 이용하여 상기 예측기의 파라미터들에 대한 값을 추정함으로써 적어도 부분적으로 제2 훈련 스테이지를 수행하는 단계를 포함함 –; 및

상기 다중-모달 통계적 모델의 상기 복수의 컴포넌트의 파라미터들에 대한 상기 추정된 값을 저장함으로써 상기 다중-모달 통계적 모델을 적어도 부분적으로 명시하는 정보를 저장하는 단계

를 포함하는 방법.

#### 청구항 2

제1항에 있어서, 상기 다중-모달 통계적 모델을 훈련시키는 단계는 상기 제1 훈련 스테이지에 앞서 상기 제1 및 제2 인코더의 파라미터들에 대한 값을 추정하는 단계를 더 포함하는, 방법.

#### 청구항 3

제2항에 있어서, 상기 다중-모달 통계적 모델을 훈련시키는 단계는, 상기 제1 훈련 스테이지에 앞서, 각각, 상기 제1 및 제2 모달리티에 대한 제1 및 제2 디코더의 파라미터들에 대한 값을 추정하는 단계를 더 포함하는, 방법.

#### 청구항 4

제1항 또는 임의의 다른 선행하는 항에 있어서, 상기 다중-모달 통계적 모델을 훈련시키는 단계는, 상기 제1 훈련 스테이지 동안, 상기 합동-모달리티 표현의 파라미터들의 값을 추정하는 단계와 함께 상기 제1 및 제2 인코더의 파라미터들에 대한 값을 추정하는 단계를 더 포함하는, 방법.

#### 청구항 5

제4항 또는 임의의 다른 선행하는 항에 있어서, 상기 다중-모달 통계적 모델을 훈련시키는 단계는, 상기 제1 훈련 스테이지 동안, 상기 제1 모달리티에 대한 제1 디코더 및 상기 제2 모달리티에 대한 제2 디코더의 파라미터들에 대한 값을 추정하는 단계를 더 포함하는, 방법.

#### 청구항 6

제1항 또는 임의의 다른 선행하는 항에 있어서, 상기 제1 훈련 스테이지를 수행하는 단계는 :

상기 제1 모달리티에 대한 무라벨 훈련 데이터의 제1 데이터 입력에 액세스하는 단계;  
 제1 피쳐 벡터를 생성하기 위해 상기 제1 데이터 입력을 상기 제1 인코더에 제공하는 단계;  
 상기 합동-모달리티 표현, 상기 제1 모달리티 임베딩 및 상기 제1 피쳐 벡터를 이용하여 제2 피쳐 벡터를 식별하는 단계; 및  
 제1 데이터 출력을 생성하기 위해 상기 제2 피쳐 벡터를 제1 디코더에 입력으로서 제공하는 단계  
 를 포함하는, 방법.

#### 청구항 7

제6항 또는 임의의 다른 선행하는 항에 있어서,  
 상기 제1 데이터 출력을 상기 제1 데이터 입력과 비교하는 단계; 및  
 상기 비교의 결과들에 기초하여 상기 합동-모달리티 표현의 적어도 하나의 파라미터의 적어도 하나의 값을 업데이트하는 단계  
 를 더 포함하는 방법.

#### 청구항 8

제1항 또는 임의의 다른 선행하는 항에 있어서, 상기 제1 훈련 스테이지를 수행하는 단계는 :  
 상기 제1 모달리티에 대한 상기 무라벨 훈련 데이터의 제1 입력에 액세스하는 단계;  
 제1 피쳐 벡터를 생성하기 위해 상기 제1 입력 데이터를 상기 제1 인코더에 제공하는 단계;  
 상기 합동-모달리티 표현, 상기 제2 모달리티 임베딩 및 상기 제1 피쳐 벡터를 이용하여 제2 피쳐 벡터를 식별하는 단계; 및  
 제2 출력 데이터를 생성하기 위해 상기 제2 피쳐 벡터를 상기 제2 모달리티에 대한 제2 디코더에 입력으로서 제공하는 단계  
 를 포함하는 방법.

#### 청구항 9

제6항 또는 임의의 다른 선행하는 항에 있어서, 상기 제1 인코더는 d차원 벡터를 출력하도록 구성되고, 상기 합동-모달리티 표현은 N개의 m차원 벡터를 포함하고, 상기 제1 모달리티 임베딩은 m\*d개의 가중치를 포함하는, 방법.

#### 청구항 10

제9항에 있어서, 상기 합동-모달리티 표현, 상기 제1 피쳐 벡터, 및 상기 제1 모달리티 임베딩을 이용하여 상기 제2 피쳐 벡터를 식별하는 단계는 :

N개의 d차원 벡터를 획득하기 위해 상기 제1 모달리티 임베딩을 이용함으로써 상기 합동-모달리티 표현을 상기 제1 모달리티의 공간에 투사하는 단계;

유사성 메트릭에 따라, 상기 합동-모달리티 표현의 N개의 d차원 벡터 중에서, 상기 제1 피쳐 벡터와 가장 유사한 제3 피쳐 벡터를 식별하는 단계; 및

상기 제1 피쳐 벡터를 상기 제3 피쳐 벡터와 총계(aggregate)함으로써 상기 제2 피쳐 벡터를 생성하는 단계  
 를 포함하는, 방법.

#### 청구항 11

제9항 또는 임의의 다른 선행하는 항에 있어서, 상기 합동-모달리티 표현, 상기 제1 피쳐 벡터, 및 상기 제1 모달리티 임베딩을 이용하여 상기 제2 피쳐 벡터를 식별하는 단계는 :

N개의 d차원 벡터를 획득하기 위해 상기 제1 모달리티 임베딩을 이용함으로써 상기 합동-모달리티 표현을 상기

제1 모달리티의 공간에 투사하는 단계;

상기 N개의 d차원 벡터 중 적어도 일부와 상기 제1 피쳐 벡터 사이의 유사성에 따라 상기 합동-모달리티 표현의 상기 N개의 d차원 벡터 중 적어도 일부에 대한 가중치들을 계산하는 단계; 및

상기 계산된 가중치들에 의해 가중된 상기 N개의 d차원 벡터 중 적어도 일부의 가중 합과 상기 제1 피쳐 벡터를 총계함으로써 상기 제2 피쳐 벡터를 생성하는 단계

를 포함하는, 방법.

## 청구항 12

제1항 또는 임의의 다른 선행하는 항에 있어서, 상기 다중-모달 통계적 모델은 제1 및 제2 태스크 임베딩(task embedding)을 더 포함하고, 상기 다중-모달 통계적 모델을 훈련시키는 단계는, 상기 제2 훈련 스테이지 동안, 상기 예측기의 파라미터들의 값들을 추정하는 단계와 함께 상기 제1 및 제2 태스크 임베딩의 파라미터들에 대한 값들을 추정하는 단계를 더 포함하는, 방법.

## 청구항 13

제1항 또는 임의의 다른 선행하는 항에 있어서, 상기 제1 인코더는 신경망을 포함하는, 방법.

## 청구항 14

제13항 또는 임의의 다른 선행하는 항에 있어서, 상기 신경망은 콘볼루션 신경망인, 방법.

## 청구항 15

제13항 또는 임의의 다른 선행하는 항에 있어서, 상기 신경망은 순환 신경망인, 방법.

## 청구항 16

제1항 또는 임의의 다른 선행하는 항에 있어서, 상기 제1 훈련 스테이지는 스토캐스틱 경사 하강 기술(stochastic gradient descent technique)을 이용하여 상기 합동-모달리티 표현의 파라미터들에 대한 값들을 추정하는 단계를 포함하는, 방법.

## 청구항 17

제16항 또는 임의의 다른 선행하는 항에 있어서, 상기 제1 훈련 스테이지는 상기 스토캐스틱 경사 하강 기술을 이용하여 상기 제1 및 제2 모달리티 임베딩의 파라미터들에 대한 값들을 추정하는 단계를 더 포함하는, 방법.

## 청구항 18

제1항 또는 임의의 다른 선행하는 항에 있어서, 상기 제1 모달리티에 대한 무라벨 훈련 데이터는 이미지들을 포함하는, 방법.

## 청구항 19

제18항 또는 임의의 다른 선행하는 항에 있어서, 상기 제2 모달리티에 대한 무라벨 훈련 데이터는 텍스트를 포함하는, 방법.

## 청구항 20

제1항 또는 임의의 다른 선행하는 항에 있어서, 상기 제1 모달리티에 대한 무라벨 훈련 데이터는 단백질 서열 데이터를 포함하는, 방법.

## 청구항 21

제20항 또는 임의의 다른 선행하는 항에 있어서, 상기 제2 모달리티에 대한 무라벨 훈련 데이터는, 단백질 패밀리 데이터, 생물학적 프로세스 온톨로지 데이터(ontology data), 분자 기능 온톨로지 데이터, 세포 성분 온톨로지 데이터, 또는 분류학적 종들 패밀리 데이터(taxonomic species family data)를 포함하는, 방법.

## 청구항 22

제1항 또는 임의의 다른 선행하는 항에 있어서,

제3 모달리티에 대한 무라밸 훈련 데이터에 액세스하는 단계;

상기 제3 모달리티에 대한 라벨링된 훈련 데이터에 액세스하는 단계;

상기 제3 모달리티에 대한 제3 인코더 및 제3 모달리티 임베딩을 포함하도록 상기 다중-모달 통계적 모델을 증강하는 단계;

자체 감독형 학습 기술 및 상기 제3 모달리티에 대한 상기 무라밸 훈련 데이터를 이용하여 제3 모달리티 임베딩 및 합동-모달리티 표현의 파라미터들에 대한 값들을 업데이트하는 단계; 및

감독형 학습 기술 및 상기 제3 모달리티에 대한 라벨링된 훈련 데이터를 이용하여 상기 예측기의 파라미터들에 대한 값들을 업데이트하는 단계

에 의해 상기 다중-모달 통계적 모델을 업데이트하는 단계

를 더 포함하는, 방법.

## 청구항 23

제1항 또는 임의의 다른 선행하는 항에 있어서, 상기 다중-모달 통계적 모델은 상기 제1 및 제2 모달리티와는 상이한 제3 모달리티로부터 입력 데이터를 수신하도록 구성되고, 제3 모달리티 임베딩을 더 포함하며,

상기 무라밸 훈련 데이터에 액세스하는 단계는, 상기 제3 모달리티에 대한 무라밸 훈련 데이터에 액세스하는 단계를 포함하고;

상기 라벨링된 훈련 데이터에 액세스하는 단계는, 상기 제3 모달리티에 대한 라벨링된 훈련 데이터에 액세스하는 단계를 포함하고;

상기 제1 훈련 스테이지를 수행하는 단계는, 상기 제3 모달리티에 대한 무라밸 훈련 데이터에도 추가로 기초하여 상기 제3 모달리티 임베딩의 파라미터들에 대한 값들을 추정하는 단계를 더 포함하고;

상기 제2 훈련 스테이지를 수행하는 단계는 상기 제3 모달리티에 대한 라벨링된 훈련 데이터에도 추가로 기초하여 상기 예측기의 파라미터들에 대한 값들을 추정하는 단계를 포함하는, 방법.

## 청구항 24

시스템으로서,

적어도 하나의 컴퓨터 하드웨어 프로세서; 및

상기 적어도 하나의 컴퓨터 하드웨어 프로세서에 의해 실행될 때, 상기 적어도 하나의 컴퓨터 하드웨어 프로세서로 하여금, 제1 모달리티로부터의 입력 데이터 및 상기 제1 모달리티와는 상이한 제2 모달리티로부터의 입력 데이터를 포함하는 복수의 모달리티로부터의 입력 데이터를 수신하도록 구성된 다중-모달 통계적 모델을 훈련시키기 위한 방법을 수행하게 하는 프로세서-실행가능한 명령어들을 저장한 적어도 하나의 비일시적인 컴퓨터-판독가능한 저장 매체를 포함하고, 상기 방법은 :

상기 제1 모달리티에 대한 무라밸 훈련 데이터 및 상기 제2 모달리티에 대한 무라밸 훈련 데이터를 포함하는 무라밸 훈련 데이터에 액세스하는 단계;

상기 제1 모달리티에 대한 라벨링된 훈련 데이터 및 상기 제2 모달리티에 대한 라벨링된 훈련 데이터를 포함하는 라벨링된 훈련 데이터에 액세스하는 단계;

상기 다중-모달 통계적 모델을 2개의 스테이지로 훈련시키는 단계 – 상기 다중-모달 통계적 모델은 상기 제1 및 상기 제2 모달리티에 대한 입력 데이터를 각각 처리하기 위한 제1 및 제2 인코더, 제1 및 제2 모달리티 임베딩, 합동-모달리티 표현, 및 예측기를 포함한 복수의 컴포넌트를 포함하고, 상기 훈련시키는 단계는 :

자체 감독형 학습 기술 및 상기 무라밸 훈련 데이터를 이용하여 상기 제1 및 제2 모달리티 임베딩, 및 합동-모달리티 표현의 파라미터들에 대한 값들을 추정함으로써 적어도 부분적으로 제1 훈련 스테이지를 수행하는 단계,

및

감독형 학습 기술 및 상기 라벨링된 훈련 데이터를 이용하여 상기 예측기의 파라미터들에 대한 값들을 추정함으로써 적어도 부분적으로 제2 훈련 스테이지를 수행하는 단계를 포함함—; 및

상기 다중-모달 통계적 모델의 상기 복수의 컴포넌트의 파라미터들에 대한 상기 추정된 값들을 저장함으로써 상기 다중-모달 통계적 모델을 적어도 부분적으로 명시하는 정보를 저장하는 단계

를 포함하는 시스템.

#### 청구항 25

제24항에 있어서, 상기 합동-모달리티 표현은 복수의 ॥차원 벡터를 포함하고, 상기 시스템은 상기 복수의 ॥차원 벡터를 저장하도록 구성된 메모리를 더 포함하는, 시스템.

#### 청구항 26

적어도 하나의 컴퓨터 하드웨어 프로세서에 의해 실행될 때, 상기 적어도 하나의 컴퓨터 하드웨어 프로세서로 하여금, 제1 모달리티로부터의 입력 데이터 및 상기 제1 모달리티와는 상이한 제2 모달리티로부터의 입력 데이터를 포함하는 복수의 모달리티로부터의 입력 데이터를 수신하도록 구성된 다중-모달 통계적 모델을 훈련시키기 위한 방법을 수행하게 하는 프로세서-실행가능한 명령어들을 저장한 적어도 하나의 비일시적인 컴퓨터-판독가능한 저장 매체로서, 상기 방법은 :

상기 제1 모달리티에 대한 무라밸 훈련 데이터 및 상기 제2 모달리티에 대한 무라밸 훈련 데이터를 포함하는 무라밸 훈련 데이터에 액세스하는 단계;

상기 제1 모달리티에 대한 라벨링된 훈련 데이터 및 상기 제2 모달리티에 대한 라벨링된 훈련 데이터를 포함하는 라벨링된 훈련 데이터에 액세스하는 단계;

상기 다중-모달 통계적 모델을 2개의 스테이지로 훈련시키는 단계 —상기 다중-모달 통계적 모델은 상기 제1 및 상기 제2 모달리티에 대한 입력 데이터를 각각 처리하기 위한 제1 및 제2 인코더, 제1 및 제2 모달리티 임베딩, 합동-모달리티 표현, 및 예측기를 포함한 복수의 컴포넌트를 포함하고, 상기 훈련시키는 단계는 :

자체 감독형 학습 기술 및 상기 무라밸 훈련 데이터를 이용하여 상기 제1 및 제2 모달리티 임베딩, 및 합동-모달리티 표현의 파라미터들에 대한 값들을 추정함으로써 적어도 부분적으로 제1 훈련 스테이지를 수행하는 단계, 및

감독형 학습 기술 및 상기 라벨링된 훈련 데이터를 이용하여 상기 예측기의 파라미터들에 대한 값들을 추정함으로써 적어도 부분적으로 제2 훈련 스테이지를 수행하는 단계를 포함함—; 및

상기 다중-모달 통계적 모델의 상기 복수의 컴포넌트의 파라미터들에 대한 상기 추정된 값들을 저장함으로써 상기 다중-모달 통계적 모델을 적어도 부분적으로 명시하는 정보를 저장하는 단계

를 포함하는, 적어도 하나의 비일시적인 컴퓨터-판독가능한 저장 매체.

#### 청구항 27

제1 모달리티로부터의 입력 데이터 및 상기 제1 모달리티와는 상이한 제2 모달리티로부터의 입력 데이터를 포함한 복수의 모달리티로부터의 입력 데이터를 수신하도록 구성된 다중-모달 통계적 모델을 이용하여 예측 태스크를 수행하기 위한 방법으로서,

상기 다중-모달 통계적 모델의 복수의 컴포넌트 —상기 복수의 컴포넌트는, 상기 제1 및 상기 제2 모달리티에 대한 입력 데이터를 각각 처리하기 위한 제1 및 제2 인코더, 제1 및 제2 모달리티 임베딩, 합동-모달리티 표현, 및 예측기를 포함함— 각각의 파라미터들의 값들을 포함하는 다중-모달 통계적 모델을 명시하는 정보를 획득하는 단계;

상기 제1 데이터 모달리티에 대한 제1 입력 데이터를 획득하는 단계;

제1 피쳐 벡터를 생성하기 위해 상기 제1 입력 데이터를 상기 제1 인코더에 제공하는 단계;

상기 합동-모달리티 표현, 상기 제1 모달리티 임베딩 및 상기 제1 피쳐 벡터를 이용하여 제2 피쳐 벡터를 식별

하는 단계; 및

상기 예측기 및 상기 제2 피쳐 벡터를 이용하여 상기 예측 태스크에 대한 예측을 생성하는 단계를 포함하는 방법.

### 청구항 28

제27항에 있어서,

상기 제2 데이터 모달리티에 대한 제2 입력 데이터를 획득하는 단계;

제3 피쳐 벡터를 생성하기 위해 상기 제2 입력 데이터를 상기 제2 인코더에 제공하는 단계; 및

상기 합동-모달리티 표현, 상기 제2 모달리티 임베딩 및 상기 제3 피쳐 벡터를 이용하여 제4 피쳐 벡터를 식별하는 단계

를 더 포함하고,

상기 예측 태스크에 대한 예측을 생성하는 단계는 상기 제4 피쳐 벡터를 이용하여 수행되는, 방법.

### 청구항 29

제28항 또는 임의의 다른 선행하는 항에 있어서, 상기 다중-모달 통계적 모델은 상기 제1 및 제2 모달리티에 대한 제1 및 제2 태스크 임베딩을 포함하고, 상기 예측 태스크에 대한 예측을 생성하는 단계는 :

상기 제1 태스크 임베딩을 이용하여 상기 제2 피쳐 벡터를 가중하는 단계;

상기 제2 태스크 임베딩을 이용하여 상기 제4 피쳐 벡터를 가중하는 단계; 및

상기 가중된 제2 및 제4 피쳐 벡터 및 예측기를 이용하여 상기 예측 태스크에 대한 예측을 생성하는 단계를 더 포함하는 방법.

### 청구항 30

제29항 또는 임의의 다른 선행하는 항에 있어서,

상기 가중된 제2 및 제4 피쳐 벡터를 상기 예측기에게 제공하는 단계를 더 포함하는 방법.

### 청구항 31

제27항 또는 임의의 다른 선행하는 항에 있어서, 상기 제1 인코더는 d차원 벡터를 출력하도록 구성되고, 상기 합동-모달리티 표현은 N개의 m차원 벡터를 포함하고, 상기 제1 모달리티 임베딩은 m × d개의 가중치를 포함하는, 방법.

### 청구항 32

제31항 또는 임의의 다른 선행하는 항에 있어서, 상기 합동-모달리티 표현, 상기 제1 피쳐 벡터, 및 상기 제1 모달리티 임베딩을 이용하여 상기 제2 피쳐 벡터를 식별하는 단계는 :

N개의 d차원 벡터를 획득하기 위해 상기 제1 모달리티 임베딩을 이용함으로써 상기 합동-모달리티 표현을 상기 제1 모달리티의 공간에 투사하는 단계;

유사성 메트릭에 따라, 상기 합동-모달리티 표현의 N개의 d차원 벡터 중에서, 상기 제1 피쳐 벡터와 가장 유사한 제3 피쳐 벡터를 식별하는 단계; 및

상기 제1 모달리티 임베딩에서의 가중치들을 이용하여 상기 제3 피쳐 벡터의 차원들을 가중함으로써 상기 제2 피쳐 벡터를 생성하는 단계를 포함하는, 방법.

### 청구항 33

제31항 또는 임의의 다른 선행하는 항에 있어서, 상기 합동-모달리티 표현, 상기 제1 피쳐 벡터, 및 상기 제1 모달리티 임베딩을 이용하여 상기 제2 피쳐 벡터를 식별하는 단계는 :

N개의 d차원 벡터를 획득하기 위해 상기 제1 모달리티 임베딩을 이용함으로써 상기 합동-모달리티 표현을 상기 제1 모달리티의 공간에 투사하는 단계;

유사성 메트릭에 따라, 상기 합동-모달리티 표현의 N개의 d차원 벡터 중에서, 상기 제1 피처 벡터와 가장 유사한 제3 피처 벡터를 식별하는 단계; 및

상기 제1 피처 벡터와 상기 제3 피처 벡터를 총계(aggregate)함으로써 상기 제2 피처 벡터를 생성하는 단계를 포함하는, 방법.

#### **청구항 34**

제32항 또는 임의의 다른 선행하는 항에 있어서, 상기 합동-모달리티 표현, 상기 제1 피처 벡터, 및 상기 제1 모달리티 임베딩을 이용하여 상기 제2 피처 벡터를 식별하는 단계는 :

N개의 d차원 벡터를 획득하기 위해 상기 제1 모달리티 임베딩을 이용함으로써 상기 합동-모달리티 표현을 상기 제1 모달리티의 공간에 투사하는 단계;

상기 N개의 d차원 벡터 중 적어도 일부와 상기 제1 피처 벡터 사이의 유사성에 따라 상기 합동-모달리티 표현의 상기 N개의 d차원 벡터 중 적어도 일부에 대한 가중치들을 계산하는 단계; 및

상기 계산된 가중치들에 의해 가중된 상기 N개의 d차원 벡터 중 적어도 일부의 가중 합으로서 상기 제2 피처 벡터를 생성하는 단계를 포함하는, 방법.

#### **청구항 35**

제27항 또는 임의의 다른 선행하는 항에 있어서, 상기 제1 인코더는 신경망을 포함하는, 방법.

#### **청구항 36**

제35항 또는 임의의 다른 선행하는 항에 있어서, 상기 신경망은 콘볼루션 신경망인, 방법.

#### **청구항 37**

제35항 또는 임의의 다른 선행하는 항에 있어서, 상기 신경망은 순환 신경망인, 방법.

#### **청구항 38**

제27항 또는 임의의 다른 선행하는 항에 있어서, 상기 제1 모달리티에 대한 입력 데이터는 적어도 하나의 이미지를 포함하는, 방법.

#### **청구항 39**

제27항 또는 임의의 다른 선행하는 항에 있어서, 상기 제2 모달리티에 대한 입력 데이터는 텍스트를 포함하는, 방법.

#### **청구항 40**

제39항 또는 임의의 다른 선행하는 항에 있어서, 상기 제1 모달리티에 대한 입력 데이터는 단백질 서열 데이터를 포함하는, 방법.

#### **청구항 41**

제40항 또는 임의의 다른 선행하는 항에 있어서, 상기 제2 모달리티에 대한 입력 데이터는, 단백질 패밀리 데이터, 생물학적 프로세스 온톨로지 데이터(ontology data), 분자 기능 온톨로지 데이터, 세포 성분 온톨로지 데이터, 또는 분류학적 종들 패밀리 데이터(taxonomic species family data)를 포함하는, 방법.

#### **청구항 42**

제1 모달리티로부터의 입력 데이터 및 상기 제1 모달리티와는 상이한 제2 모달리티로부터의 입력 데이터를 포함한 복수의 모달리티로부터의 입력 데이터를 수신하도록 구성된 다중-모달 통계적 모델을 이용하여 예측 태스크를 수행하기 위한 시스템으로서,

적어도 하나의 컴퓨터 하드웨어 프로세서; 및

프로세서 실행가능한 명령어들을 저장한 적어도 하나의 비일시적인 컴퓨터-판독가능한 저장 매체를 포함하고,

상기 프로세서 실행가능한 명령어들은, 상기 적어도 하나의 컴퓨터 하드웨어 프로세서에 의해 실행될 때, 상기 적어도 하나의 컴퓨터 하드웨어 프로세서로 하여금 :

상기 다중-모달 통계적 모델의 복수의 컴포넌트 –상기 복수의 컴포넌트는, 상기 제1 및 상기 제2 모달리티에 대한 입력 데이터를 각각 처리하기 위한 제1 및 제2 인코더, 제1 및 제2 모달리티 임베딩, 합동-모달리티 표현, 및 예측기를 포함함— 각각의 파라미터들의 값들을 포함하는 다중-모달 통계적 모델을 명시하는 정보를 획득하는 단계;

상기 제1 데이터 모달리티에 대한 제1 입력 데이터를 획득하는 단계;

제1 피처 벡터를 생성하기 위해 상기 제1 입력 데이터를 상기 제1 인코더에 제공하는 단계;

상기 합동-모달리티 표현, 상기 제1 모달리티 임베딩 및 상기 제1 피처 벡터를 이용하여 제2 피처 벡터를 식별하는 단계; 및

상기 예측기 및 상기 제2 피처 벡터를 이용하여 상기 예측 태스크에 대한 예측을 생성하는 단계

를 수행하게 하는, 시스템.

### 청구항 43

프로세서 실행가능한 명령어들을 저장한 적어도 하나의 비일시적인 컴퓨터-판독가능한 저장 매체로서, 상기 프로세서 실행가능한 명령어들은, 상기 적어도 하나의 컴퓨터 하드웨어 프로세서에 의해 실행될 때, 상기 적어도 하나의 컴퓨터 하드웨어 프로세서로 하여금 :

다중-모달 통계적 모델의 복수의 컴포넌트 –상기 복수의 컴포넌트는, 제1 및 제2 모달리티에 대한 입력 데이터를 각각 처리하기 위한 제1 및 제2 인코더, 제1 및 제2 모달리티 임베딩, 합동-모달리티 표현, 및 예측기를 포함함— 각각의 파라미터들의 값들을 포함하는 다중-모달 통계적 모델을 명시하는 정보를 획득하는 단계;

상기 제1 데이터 모달리티에 대한 제1 입력 데이터를 획득하는 단계;

제1 피처 벡터를 생성하기 위해 상기 제1 입력 데이터를 상기 제1 인코더에 제공하는 단계;

상기 합동-모달리티 표현, 상기 제1 모달리티 임베딩 및 상기 제1 피처 벡터를 이용하여 제2 피처 벡터를 식별하는 단계; 및

상기 예측기 및 상기 제2 피처 벡터를 이용하여 상기 예측 태스크에 대한 예측을 생성하는 단계

를 수행하게 하는, 적어도 하나의 비일시적인 컴퓨터-판독가능한 저장 매체.

### 발명의 설명

#### 기술 분야

##### 관련 출원들의 상호참조

[0001] 본 출원은, 그 전체내용이 참조에 의해 각각 본 명세서에 포함되는, 2018년 5월 14일에 출원된, 발명의 명칭이 "Systems and Methods for Multi-Modal Prediction"인 미국 가출원 제62/671,068호, 및 2018년 5월 30일 출원된, 발명의 명칭이 "Systems and Methods for Unifying Statistical Models for Different Data Modalities"인 미국 가출원 제62/678,074호에 대해 35 U.S.C. § 119(e) 하에서 우선권을 주장한다.

### 배경 기술

[0003] 머신 학습 기술은 종종 복수의 모달리티의 데이터가 이용가능한 경우의 문제에 적용된다. 데이터는 상이한 취득 프레임워크들을 이용하여 수집될 수 있고, 각각의 프레임워크는, 각각의 데이터 소스, 데이터 유형, 데이터 수집 기술, 센서, 및/또는 환경에 의해 특성규정될 수 있다. 하나의 모달리티와 연관된 데이터는, 상이한 모달리티와 연관된 데이터를 수집하는데 이용되는 취득 프레임워크와는 상이한 취득 프레임워크를 이용하여 수집될 수

있다. 예를 들어, 한 유형의 센서 또는 실험 기술에 의해 수집된 데이터는, 또 다른 유형의 센서 또는 실험 기술에 의해 수집된 데이터와는 상이한 모달리티를 갖는다. 또 다른 예로서, 한 유형의 데이터(예를 들어, 이미지 데이터)는 또 다른 유형의 데이터(예를 들어, 텍스트 데이터)와 동일한 모달리티가 아니다.

[0004] 특정한 모달리티에 대한 데이터를 처리하기 위한 많은 종래의 통계적 모델이 있다. 예를 들어, 이미지에 나타난 물체를 식별하는 문제를 해결하기 위해 콘볼루션 신경망이 이미지에 적용될 수 있다. 또 다른 예로서, 음성 인식을 위해 음성 데이터에 순환 신경망이 적용될 수 있다.

[0005] 그러나, 복수의 상이한 데이터 모달리티의 데이터를 효과적으로 활용할 수 있는 통계적 학습 모델을 훈련하고 이용하는 것은 더 어렵다. 이러한 다중-모달 통계적 학습 모델은 의학 및 생물학을 포함한 다양한 분야에서 광범위하게 적용할 수 있고, 이 경우 관심대상 문제(예를 들어, 환자가 특정한 약물 치료에 반응할지를 예측하는 것)를 감당할 수 있는 많은 수의 이종 데이터 소스(예를 들어, 환자의 DNA, RNA 및 단백질 발현 데이터, 하나 이상의 모달리티의 환자의 의료 이미지, 환자의 병력, 환자가 가질 수 있는 질병에 관한 정보 등)가 있다.

## 발명의 내용

### 해결하려는 과제

### 과제의 해결 수단

[0006] 일부 실시예는, 제1 모달리티로부터의 입력 데이터 및 제1 모달리티와는 상이한 제2 모달리티로부터의 입력 데이터를 포함하는 복수의 모달리티로부터 입력 데이터를 수신하도록 구성된 다중-모달 통계적 모델을 훈련시키기 위한 방법을 포함한다. 이 방법은 : 제1 모달리티에 대한 무라벨 훈련 데이터 및 제2 모달리티에 대한 무라벨 훈련 데이터를 포함하는 무라벨 훈련 데이터에 액세스하는 단계; 제1 모달리티에 대한 라벨링된 훈련 데이터 및 제2 모달리티에 대한 라벨링된 훈련 데이터를 포함하는 라벨링된 훈련 데이터에 액세스하는 단계; 다중-모달 통계적 모델을 2개의 스테이지로 훈련시키는 단계 -다중-모달 통계적 모델은 제1 및 제2 모달리티에 대한 입력 데이터를 각각 처리하기 위한 제1 및 제2 인코더, 제1 및 제2 모달리티 임베딩, 합동-모달리티 표현, 및 예측기를 포함한 복수의 컴포넌트를 포함하고, 상기 훈련시키는 단계는, 자체 감독형 학습 기술 및 무라벨 훈련 데이터를 이용하여 상기 제1 및 제2 모달리티 임베딩, 및 합동-모달리티 표현의 파라미터들에 대한 값들을 추정함으로써 적어도 부분적으로 제1 훈련 스테이지를 수행하는 단계, 및 감독형 학습 기술 및 라벨링된 훈련 데이터를 이용하여 예측기의 파라미터들에 대한 값들을 추정함으로써 적어도 부분적으로 제2 훈련 스테이지를 수행하는 단계를 포함함-; 및 다중-모달 통계적 모델의 복수의 컴포넌트의 파라미터들에 대한 추정된 값들을 저장함으로써 다중-모달 통계적 모델을 적어도 부분적으로 명시하는 정보를 저장하는 단계를 포함한다.

[0007] 일부 실시예는, 적어도 하나의 컴퓨터 하드웨어 프로세서; 및 적어도 하나의 컴퓨터 하드웨어 프로세서에 의해 실행될 때, 적어도 하나의 컴퓨터 하드웨어 프로세서로 하여금, 제1 모달리티로부터의 입력 데이터 및 제1 모달리티와는 상이한 제2 모달리티로부터의 입력 데이터를 포함하는 복수의 모달리티로부터의 입력 데이터를 수신하도록 구성된 다중-모달 통계적 모델을 훈련시키기 위한 방법을 수행하게 하는 프로세서-실행 가능한 명령어들을 저장한 적어도 하나의 비일시적인 컴퓨터-판독 가능한 저장 매체를 포함하는 시스템을 포함한다. 이 방법은 : 제1 모달리티에 대한 무라벨 훈련 데이터 및 제2 모달리티에 대한 무라벨 훈련 데이터를 포함하는 무라벨 훈련 데이터에 액세스하는 단계; 제1 모달리티에 대한 라벨링된 훈련 데이터 및 제2 모달리티에 대한 라벨링된 훈련 데이터를 포함하는 라벨링된 훈련 데이터에 액세스하는 단계; 다중-모달 통계적 모델을 2개의 스테이지로 훈련시키는 단계 -다중-모달 통계적 모델은 제1 및 제2 모달리티에 대한 입력 데이터를 각각 처리하기 위한 제1 및 제2 인코더, 제1 및 제2 모달리티 임베딩, 합동-모달리티 표현, 및 예측기를 포함한 복수의 컴포넌트를 포함하고, 상기 훈련시키는 단계는, 자체 감독형 학습 기술 및 무라벨 훈련 데이터를 이용하여 상기 제1 및 제2 모달리티 임베딩, 및 합동-모달리티 표현의 파라미터들에 대한 값들을 추정함으로써 적어도 부분적으로 제1 훈련 스테이지를 수행하는 단계, 및 감독형 학습 기술 및 라벨링된 훈련 데이터를 이용하여 예측기의 파라미터들에 대한 값들을 추정함으로써 적어도 부분적으로 제2 훈련 스테이지를 수행하는 단계를 포함함-; 및 다중-모달 통계적 모델의 복수의 컴포넌트의 파라미터들에 대한 추정된 값들을 저장함으로써 다중-모달 통계적 모델을 적어도 부분적으로 명시하는 정보를 저장하는 단계를 포함한다.

[0008] 일부 실시예는, 적어도 하나의 컴퓨터 하드웨어 프로세서에 의해 실행될 때, 적어도 하나의 컴퓨터 하드웨어 프

로 세서로 하여금, 제1 모달리티로부터의 입력 데이터 및 제1 모달리티와는 상이한 제2 모달리티로부터의 입력 데이터를 포함하는 복수의 모달리티로부터의 입력 데이터를 수신하도록 구성된 다중-모달 통계적 모델을 훈련시키기 위한 방법을 수행하게 하는 프로세서-실행가능한 명령어들을 저장한 적어도 하나의 비일시적인 컴퓨터-판독가능한 저장 매체를 포함한다. 이 방법은 : 제1 모달리티에 대한 무라벨 훈련 데이터 및 제2 모달리티에 대한 무라벨 훈련 데이터를 포함하는 무라벨 훈련 데이터에 액세스하는 단계; 제1 모달리티에 대한 라벨링된 훈련 데이터 및 제2 모달리티에 대한 라벨링된 훈련 데이터를 포함하는 라벨링된 훈련 데이터에 액세스하는 단계; 다중-모달 통계적 모델을 2개의 스테이지로 훈련시키는 단계 – 다중-모달 통계적 모델은 제1 및 제2 모달리티에 대한 입력 데이터를 각각 처리하기 위한 제1 및 제2 인코더, 제1 및 제2 모달리티 임베딩, 합동-모달리티 표현, 및 예측기를 포함한 복수의 컴포넌트를 포함하고, 상기 훈련시키는 단계는, 자체 감독형 학습 기술 및 무라벨 훈련 데이터를 이용하여 상기 제1 및 제2 모달리티 임베딩, 및 합동-모달리티 표현의 파라미터들에 대한 값들을 추정함으로써 적어도 부분적으로 제1 훈련 스테이지를 수행하는 단계, 및 감독형 학습 기술 및 라벨링된 훈련 데이터를 이용하여 예측기의 파라미터들에 대한 값들을 추정함으로써 적어도 부분적으로 제2 훈련 스테이지를 수행하는 단계를 포함함 –; 및 다중-모달 통계적 모델의 복수의 컴포넌트의 파라미터들에 대한 추정된 값들을 저장함으로써 다중-모달 통계적 모델을 적어도 부분적으로 명시하는 정보를 저장하는 단계를 포함한다.

- [0009] 일부 실시예에서, 다중-모달 통계적 모델을 훈련시키는 단계는, 제1 훈련 스테이지에 앞서 제1 및 제2 인코더의 파라미터들에 대한 값들을 추정하는 단계를 더 포함한다.
- [0010] 일부 실시예에서, 다중-모달 통계적 모델을 훈련시키는 단계는, 제1 훈련 스테이지에 앞서 각각 제1 및 제2 모달리티에 대한 제1 및 제2 디코더의 파라미터들에 대한 값들을 추정하는 단계를 더 포함한다.
- [0011] 일부 실시예에서, 다중-모달 통계적 모델을 훈련시키는 단계는, 제1 훈련 스테이지 동안, 합동-모달리티 표현의 파라미터들의 값들을 추정하는 단계와 함께 제1 및 제2 인코더들의 파라미터들에 대한 값들을 추정하는 단계를 더 포함한다.
- [0012] 일부 실시예에서, 다중-모달 통계적 모델을 훈련시키는 단계는, 제1 훈련 스테이지 동안, 제1 모달리티에 대한 제1 디코더 및 제2 모달리티에 대한 제2 디코더의 파라미터들에 대한 값들을 추정하는 단계를 더 포함한다.
- [0013] 일부 실시예에서, 제1 훈련 스테이지를 수행하는 단계는 : 제1 모달리티에 대한 무라벨 훈련 데이터의 제1 데이터 입력에 액세스하는 단계; 제1 피쳐 벡터를 생성하기 위해 제1 데이터 입력을 제1 인코더에 제공하는 단계; 합동-모달리티 표현, 제1 모달리티 임베딩 및 제1 피쳐 벡터를 이용하여 제2 피쳐 벡터를 식별하는 단계; 및 제1 데이터 출력을 생성하기 위해 제2 피쳐 벡터를 입력으로서 제1 디코더에 제공하는 단계를 포함한다.
- [0014] 일부 실시예에서, 이 방법은 제1 데이터 출력을 제1 데이터 입력과 비교하는 단계; 및 비교의 결과들에 기초하여 합동-모달리티 표현의 적어도 하나의 파라미터의 적어도 하나의 값을 업데이트하는 단계를 더 포함한다.
- [0015] 일부 실시예에서, 제1 훈련 스테이지를 수행하는 단계는 : 제1 모달리티에 대한 무라벨 훈련 데이터의 제1 입력에 액세스하는 단계; 제1 피쳐 벡터를 생성하기 위해 제1 입력 데이터를 제1 인코더에 제공하는 단계; 합동-모달리티 표현, 제2 모달리티 임베딩 및 제1 피쳐 벡터를 이용하여 제2 피쳐 벡터를 식별하는 단계; 및 제2 출력 데이터를 생성하기 위해 제2 피쳐 벡터를 제2 모달리티에 대한 제2 디코더에 입력으로서 제공하는 단계를 포함한다.
- [0016] 일부 실시예에서, 제1 인코더는 d차원 벡터를 출력하도록 구성되고, 합동-모달리티 표현은 N개의 m차원 벡터를 포함하고, 제1 모달리티 임베딩은 m\*d개의 가중치를 포함한다.
- [0017] 일부 실시예에서, 합동-모달리티 표현, 제1 피쳐 벡터 및 제1 모달리티 임베딩을 이용하여 제2 피쳐 벡터를 식별하는 단계는, N개의 d차원 벡터를 획득하기 위해 제1 모달리티 임베딩을 이용함으로써 합동-모달리티 표현을 제1 모달리티의 공간에 투사하는 단계; 및 유사성 메트릭에 따라, 합동-모달리티 표현의 N개의 d차원 벡터 중에서, 제1 피쳐 벡터와 가장 유사한 제3 피쳐 벡터를 식별하는 단계; 및 제1 피쳐 벡터를 제3 피쳐 벡터와 총계(aggregate)함으로써 제2 피쳐 벡터를 생성하는 단계를 포함한다.
- [0018] 일부 실시예에서, 합동-모달리티 표현, 제1 피쳐 벡터, 및 제1 모달리티 임베딩을 이용하여 제2 피쳐 벡터를 식별하는 단계는 : N개의 d차원 벡터를 획득하기 위해 제1 모달리티 임베딩을 이용함으로써 합동-모달리티 표현을 제1 모달리티의 공간에 투사하는 단계; N개의 d차원 벡터 중 적어도 일부와 제1 피쳐 벡터 사이의 유사성에 따라 합동-모달리티 표현의 N개의 d차원 벡터 중 적어도 일부에 대한 가중치들을 계산하는 단계; 및 계산된 가중치들에 의해 가중된 N개의 d차원 벡터 중 적어도 일부의 가중 합과 제1 피쳐 벡터를 총계함으로써 제2 피쳐 벡터를 계산하는 단계를 포함한다.

터를 생성하는 단계를 포함한다.

- [0019] 일부 실시예에서, 다중-모달 통계적 모델은 제1 및 제2 태스크 임베딩(task embedding)을 더 포함하고, 여기서, 다중-모달 통계적 모델을 훈련시키는 단계는, 제2 훈련 스테이지 동안, 예측기의 파라미터들의 값들을 추정하는 단계와 함께 제1 및 제2 태스크 임베딩의 파라미터들에 대한 값들을 추정하는 단계를 더 포함한다.
- [0020] 일부 실시예에서, 제1 인코더는 신경망을 포함한다. 일부 실시예에서, 신경망은 콘볼루션 신경망이다. 일부 실시예에서, 신경망은 순환 신경망이다.
- [0021] 일부 실시예에서, 제1 훈련 스테이지는 스토캐스틱 경사 하강 기술을 이용하여 합동-모달리티 표현의 파라미터들에 대한 값들을 추정하는 단계를 포함한다. 일부 실시예에서, 제1 훈련 스테이지는 스토캐스틱 경사 하강 기술을 이용하여 제1 및 제2 모달리티 임베딩의 파라미터들에 대한 값들을 추정하는 단계를 더 포함한다.
- [0022] 일부 실시예에서, 제1 모달리티에 대한 무라벨 훈련 데이터는 이미지들을 포함한다. 일부 실시예에서, 제2 모달리티에 대한 무라벨 훈련 데이터는 텍스트를 포함한다. 일부 실시예에서, 제1 모달리티에 대한 무라벨 훈련 데이터는 단백질 서열 데이터를 포함한다. 일부 실시예에서, 제2 모달리티에 대한 무라벨 훈련 데이터는, 단백질 패밀리 데이터, 생물학적 프로세스 온톨로지 데이터(ontology data), 분자 기능 온톨로지 데이터, 세포 성분 온톨로지 데이터, 또는 분류학적 종들 패밀리 데이터를 포함한다.
- [0023] 일부 실시예에서, 이 방법은 : 제3 모달리티에 대한 무라벨 훈련 데이터에 액세스하는 단계; 제3 모달리티에 대한 라벨링된 훈련 데이터에 액세스하는 단계; 제3 모달리티에 대한 제3 인코더 및 제3 모달리티 임베딩을 포함하도록 다중-모달 통계적 모델을 증강하는 단계; 및 자체 감독형 학습 기술 및 제3 모달리티에 대한 무라벨 훈련 데이터를 이용하여 제3 모달리티 임베딩 및 합동-모달리티 표현의 파라미터들에 대한 값을 업데이트하는 단계, 및 감독형 학습 기술 및 제3 모달리티에 대한 라벨링된 훈련 데이터를 이용하여 예측기의 파라미터들에 대한 값을 업데이트하는 단계에 의해 다중-모달 통계적 모델을 업데이트하는 단계를 더 포함한다.
- [0024] 일부 실시예에서, 다중-모달 통계적 모델은 제1 및 제2 모달리티와는 상이한 제3 모달리티로부터 입력 데이터를 수신하도록 구성되고, 제3 모달리티 임베딩을 추가로 포함하며, 여기서 : 무라벨 훈련 데이터에 액세스하는 단계는 제3 모달리티에 대한 무라벨 훈련 데이터에 액세스하는 단계를 포함하고; 라벨링된 훈련 데이터에 액세스하는 단계는 제3 모달리티에 대한 라벨링된 훈련 데이터에 액세스하는 단계를 포함하고; 제1 훈련 스테이지를 수행하는 단계는, 제3 모달리티에 대한 무라벨 훈련 데이터에도 추가로 기초하여 제3 모달리티 임베딩의 파라미터들에 대한 값을 추정하는 단계를 더 포함하고; 제2 훈련 스테이지를 수행하는 단계는 제3 모달리티에 대한 라벨링된 훈련 데이터에도 추가로 기초하여 예측기의 파라미터들에 대한 값을 추정하는 단계를 포함한다.
- [0025] 일부 실시예는, 제1 모달리티로부터의 입력 데이터, 및 제1 모달리티와는 상이한 제2 모달리티로부터의 입력 데이터를 포함하는 복수의 모달리티로부터의 입력 데이터를 수신하도록 구성된 다중-모달 통계적 모델을 이용하여 예측 태스크를 수행하기 위한 방법을 포함한다. 이 방법은, 다중-모달 통계적 모델의 복수의 컴포넌트 – 복수의 컴포넌트는, 제1 및 제2 모달리티에 대한 입력 데이터를 각각 처리하기 위한 제1 및 제2 인코더, 제1 및 제2 모달리티 임베딩, 합동-모달리티 표현, 및 예측기를 포함함 – 각각의 파라미터들의 값을 포함하는 다중-모달 통계적 모델을 명시하는 정보를 획득하는 단계; 제1 데이터 모달리티에 대한 제1 입력 데이터를 획득하는 단계; 제1 피처 벡터를 생성하기 위해 제1 입력 데이터를 제1 인코더에 제공하는 단계; 합동-모달리티 표현, 제1 모달리티 임베딩 및 제1 피처 벡터를 이용하여 제2 피처 벡터를 식별하는 단계; 및 예측기 및 제2 피처 벡터를 이용하여 예측 태스크에 대한 예측을 생성하는 단계를 포함한다.
- [0026] 일부 실시예는, 제1 모달리티로부터의 입력 데이터, 및 제1 모달리티와는 상이한 제2 모달리티로부터의 입력 데이터를 포함하는 복수의 모달리티로부터의 입력 데이터를 수신하도록 구성된 다중-모달 통계적 모델을 이용하여 예측 태스크를 수행하기 위한 시스템을 포함한다. 이 시스템은 : 적어도 하나의 컴퓨터 하드웨어 프로세서; 및 프로세서-실행 가능한 명령어들을 저장한 적어도 하나의 비일시적인 컴퓨터-판독 가능한 저장 매체를 포함하고, 이 프로세서-실행 가능한 명령어들은, 적어도 하나의 컴퓨터 하드웨어 프로세서에 의해 실행될 때, 적어도 하나의 컴퓨터 하드웨어 프로세서로 하여금 : 다중-모달 통계적 모델의 복수의 컴포넌트 – 복수의 컴포넌트는, 제1 및 제2 모달리티에 대한 입력 데이터를 각각 처리하기 위한 제1 및 제2 인코더, 제1 및 제2 모달리티 임베딩, 합동-모달리티 표현, 및 예측기를 포함함 – 각각의 파라미터들의 값을 포함하는 다중-모달 통계적 모델을 명시하는 정보를 획득하는 단계; 제1 데이터 모달리티에 대한 제1 입력 데이터를 획득하는 단계; 제1 피처 벡터를 생성하기 위해 제1 입력 데이터를 제1 인코더에 제공하는 단계; 합동-모달리티 표현, 제1 모달리티 임베딩 및 제1 피처 벡터를 이용하여 제2 피처 벡터를 식별하는 단계; 및 예측기 및 제2 피처 벡터를 이용하여 예측 태

크에 대한 예측을 생성하는 단계를 수행하게 한다.

[0027] 일부 실시예는, 프로세서-실행가능한 명령어들을 저장한 적어도 하나의 비일시적인 컴퓨터-판독가능한 저장 매체를 포함하고, 이 프로세서-실행가능한 명령어들은, 적어도 하나의 컴퓨터 하드웨어 프로세서에 의해 실행될 때, 적어도 하나의 컴퓨터 하드웨어 프로세서로 하여금 : 다중-모달 통계적 모델의 복수의 컴포넌트 – 복수의 컴포넌트는, 제1 및 제2 모달리티에 대한 입력 데이터를 각각 처리하기 위한 제1 및 제2 인코더, 제1 및 제2 모달리티 임베딩, 합동-모달리티 표현, 및 예측기를 포함함 – 각각의 파라미터들의 값들을 포함하는 다중-모달 통계적 모델을 명시하는 정보를 획득하는 단계; 제1 데이터 모달리티에 대한 제1 입력 데이터를 획득하는 단계; 제1 피쳐 벡터를 생성하기 위해 제1 입력 데이터를 제1 인코더에 제공하는 단계; 합동-모달리티 표현, 제1 모달리티 임베딩 및 제1 피쳐 벡터를 이용하여 제2 피쳐 벡터를 식별하는 단계; 및 예측기 및 제2 피쳐 벡터를 이용하여 예측 태스크에 대한 예측을 생성하는 단계를 수행하게 한다.

[0028] 일부 실시예에서, 이 방법은 : 제2 데이터 모달리티에 대한 제2 입력 데이터를 획득하는 단계; 제3 피쳐 벡터를 생성하기 위해 제2 입력 데이터를 제2 인코더에 제공하는 단계; 합동-모달리티 표현, 제2 모달리티 임베딩 및 제3 피쳐 벡터를 이용하여 제4 피쳐 벡터를 식별하는 단계를 더 포함하고, 여기서, 예측 태스크에 대한 예측을 생성하는 단계는 제4 피쳐 벡터를 이용하여 수행된다.

[0029] 일부 실시예에서, 다중-모달 통계적 모델은 제1 및 제2 모달리티에 대한 제1 및 제2 태스크 임베딩을 포함하고, 여기서, 예측 태스크에 대한 예측을 생성하는 단계는 : 제1 태스크 임베딩을 이용하여 제2 피쳐 벡터를 가중하는 단계; 제2 태스크 임베딩을 이용하여 제4 피쳐 벡터를 가중하는 단계; 및 가중된 제2 및 제4 피쳐 벡터 및 예측기를 이용하여 예측 태스크에 대한 예측을 생성하는 단계를 더 포함한다.

[0030] 일부 실시예에서, 이 방법은 가중된 제2 및 제4 피쳐 벡터를 예측기에게 제공하는 단계를 더 포함한다.

[0031] 일부 실시예에서, 제1 인코더는 d차원 벡터를 출력하도록 구성되고, 여기서, 합동-모달리티 표현은 N개의 m차원 벡터를 포함하고, 제1 모달리티 임베딩은  $m \times d$ 개의 가중치를 포함한다.

[0032] 일부 실시예에서, 합동-모달리티 표현, 제1 피쳐 벡터 및 제1 모달리티 임베딩을 이용하여 제2 피쳐 벡터를 식별하는 단계는, N개의 d차원 벡터를 획득하기 위해 제1 모달리티 임베딩을 이용함으로써 합동-모달리티 표현을 제1 모달리티의 공간에 투사하는 단계; 및 유사성 메트릭에 따라, 합동-모달리티 표현의 N개의 d차원 벡터 중에서, 제1 피쳐 벡터와 가장 유사한 제3 피쳐 벡터를 식별하는 단계; 및 제1 모달리티 임베딩의 가중치들을 이용하여 제3 피쳐 벡터의 차원들을 가중함으로써 제2 피쳐 벡터를 생성하는 단계를 포함한다.

[0033] 일부 실시예에서, 합동-모달리티 표현, 제1 피쳐 벡터 및 제1 모달리티 임베딩을 이용하여 제2 피쳐 벡터를 식별하는 단계는, N개의 d차원 벡터를 획득하기 위해 제1 모달리티 임베딩을 이용함으로써 합동-모달리티 표현을 제1 모달리티의 공간에 투사하는 단계; 합동-모달리티 표현의 N개의 d차원 벡터들 중에서, 유사성 메트릭에 따라 제1 피쳐 벡터와 가장 유사한 제3 피쳐 벡터를 식별하는 단계; 및 제1 피쳐 벡터와 제3 피쳐 벡터를 총계함으로써 제2 피쳐 벡터를 생성하는 단계를 포함한다.

[0034] 일부 실시예에서, 합동-모달리티 표현, 제1 피쳐 벡터, 및 제1 모달리티 임베딩을 이용하여 제2 피쳐 벡터를 식별하는 단계는 : N개의 d차원 벡터를 획득하기 위해 제1 모달리티 임베딩을 이용함으로써 합동-모달리티 표현을 제1 모달리티의 공간에 투사하는 단계; N개의 d차원 벡터 중 적어도 일부와 제1 피쳐 벡터 사이의 유사성에 따라 합동-모달리티 표현의 N개의 d차원 벡터 중 적어도 일부에 대한 가중치들을 계산하는 단계; 및 계산된 가중치들에 의해 가중된 N개의 d차원 벡터 중 적어도 일부의 가중 합으로서 제2 피쳐 벡터를 생성하는 단계를 포함한다.

[0035] 일부 실시예에서, 제1 인코더는 신경망을 포함한다. 일부 실시예에서, 신경망은 콘볼루션 신경망이다. 일부 실시예에서, 신경망은 순환 신경망이다.

[0036] 일부 실시예에서, 제1 모달리티에 대한 입력 데이터는 적어도 하나의 이미지를 포함한다. 일부 실시예에서, 제2 모달리티에 대한 입력 데이터는 텍스트를 포함한다. 일부 실시예에서, 제1 모달리티에 대한 입력 데이터는 단백질 서열 데이터를 포함한다. 일부 실시예에서, 제2 모달리티에 대한 입력 데이터는, 단백질 패밀리 데이터, 생물학적 프로세스 온톨로지 데이터, 문자 기능 온톨로지 데이터, 세포 성분 온톨로지 데이터, 또는 분류학적 종들 패밀리 데이터를 포함한다.

[0037] 전술된 개념들 및 이하에서 더 상세히 논의되는 추가의 개념들의 모든 조합은 (이러한 개념들이 서로 불일치하지 않는다면) 본 명세서에서 개시된 발명 주제의 일부로서 간주된다는 것을 이해하여야 한다.

## 도면의 간단한 설명

[0038]

본 기술의 다양한 비제한적인 실시예가 이하의 도면들을 참조하여 설명될 것이다. 도면들은 반드시 축척비율대로 그려진 것은 아니라는 점을 이해해야 한다.

도 1은, 본 명세서에 설명된 기술의 일부 실시예에 따른, 자체 감독형 학습 기술을 이용하는 단일 모달리티에 대한 통계적 모델을 위한 지식 베이스(knowledge base)의 훈련을 나타내는 다이어그램이다.

도 2a는, 본 명세서에 설명된 기술의 일부 실시예에 따른, 자체 감독형 학습 기술을 이용한 다중-모달 통계적 모델의 제1 훈련 스테이지를 나타내는 다이어그램이다.

도 2b는, 본 명세서에 설명된 기술의 일부 실시예에 따른 감독형 학습 기술을 이용한 다중-모달 통계적 모델의 제2 훈련 스테이지를 나타내는 다이어그램이다.

도 3은, 본 명세서에서 설명된 기술의 일부 실시예에 따른, 2-스테이지 훈련 절차를 이용하여 다중-모달 통계적 모델을 훈련시키기 위한 예시적인 프로세스의 플로차트로서, 제1 스테이지는 자체 감독형 학습을 포함하고 제2 스테이지는 감독형 학습을 포함한다.

도 4는, 본 명세서에서 설명된 기술의 일부 실시예에 따른, 예측 태스크를 위한 다중-모달 통계적 모델을 이용하기 위한 예시적인 프로세스(400)의 플로차트이다.

도 5는, 본 명세서에서 설명된 기술의 일부 실시예에 따른, 종래 기술에 비교한 예측 태스크에 관한 다중-모달 통계적 모델의 성능을 나타낸다.

도 6a 및 도 6b는 본 명세서에 설명된 기술의 일부 실시예에 따른 인코더 및 디코더를 나타낸다.

도 7은 본 명세서에서 설명된 기술의 일부 실시예가 구현될 수 있는 예시적인 컴퓨터 시스템의 컴포넌트들을 나타낸다.

## 발명을 실시하기 위한 구체적인 내용

[0039]

복수의 모달리티로부터 데이터를 입력으로서 수신하고 처리하도록 구성된 통계적 모델은 다중-모달 통계적 모델이라고 지칭될 수 있다. 본 발명자들은, 다중-모달 통계적 모델을 생성하기 위해, 각각이 상이한 각각의 모달리티의 데이터를 처리하도록 설계된, 복수의 개개의 통계적 모델을 통합하기 위한 새로운 기술을 개발함으로써 새로운 클래스의 다중-모달 통계적 모델을 개발했다. 본 명세서에서 설명된 기술은, 상이한 모달리티 및/또는 임의의 기타 적절한 유형의 통계적 모델에 대해 훈련된 복수의 딥 러닝 모델을 통합하는데 이용될 수 있다. 본 발명자들에 의해 개발된 기술은 다중-모달 통계적 모델을 구축하기 위한 종래 기술의 단점을 해결한다. 이를 단점을 해결함으로써, 발명자들은 종래의 머신 학습 시스템과 이를 구현하는데 이용되는 컴퓨터 기술을 개선하는 기술을 개발했다.

[0040]

다중-모달 통계적 모델을 훈련시키기 위한 종래의 머신 학습 기술은, 복수의 모달리티 각각으로부터의 링크된 데이터 –각각의 훈련 데이터는, 통계적 모델이 처리하도록 훈련받고 있는 모달리티들 각각으로부터의 데이터를 포함함– 를 이용하여 다중-모달 통계적 모델이 "동기식으로" 훈련될 것을 요구한다. 이러한 동시 훈련에 대한 요건은 심각한 제한사항이며 적은 수(예를 들어, 2 또는 3개)보다 많은 모달리티로부터의 데이터를 수신하고 처리할 수 있는 다중-모달 통계적 모델의 설계를 방해하지만, 특히 의학 및 생물학 등의 분야에서 훨씬 많은 수의 데이터 모달리티로부터의 입력을 처리할 수 있는 다중-모달 통계적 모델이 필요하다.

[0041]

동기식 훈련은 링크된 데이터의 수집을 요구하기 때문에 심각한 제한사항이다 – 하나의 모달리티에 대한 각각의 훈련 데이터는, 다중-모달 통계적 모델이 처리하도록 훈련받고 있는 다른 모든 모달리티에 대응하는 훈련 데이터를 가져야 한다. 이러한 훈련 데이터를 수집하는 것은 엄청나게 비싸고 시간 소모적이므로, 데이터를 수집하고 라벨링하는데 수백 또는 수천 인시(man hours)가 요구된다. 동시 훈련이 가능하고 2개의 데이터 모달리티에 대해 링크된 데이터가 이용 가능하더라도, 나중에 또 다른 데이터 모달리티에 대한 새로운 데이터가 획득된다면, 그 새로운 데이터는 기존의 데이터에 링크되어야 하고(다시 한번, 시간 소모적이고 비용이 많이 듈다), 전체의 통계적 모델이 재훈련되어야 할 것이다. 요약하면, 동기식 훈련은 적은 수(즉, 2개 또는 3개)보다 많은 모달리티에 대한 다중-모달 통계적 모델을 생성하고 업데이트하는 것이 비실용적이고, 사실상 거의 불가능하게 만든다.

[0042]

본 발명자들에 의해 개발되고 본 명세서에서 설명된 기술은, 통계적 모델이 처리하도록 훈련받고 있는 복수의

모달리티 각각으로부터의 링크된 데이터를 이용하여 훈련이 동기식으로 수행될 것을 요구하지 않고 다중-모달 통계적 모델의 효율적 생성 및 업데이트를 허용한다. 종래 기술과는 대조적으로, 본 발명자들은 다중-모달 통계적 모델의 비동기식 훈련 및 업데이트를 허용하는 접근법을 개발했다. 비동기식 훈련은 본 명세서에서 설명된 혁신적인 공유된 코드북 아키텍처에 의해 가능케된다. 이 아키텍처에서, 각각의 모달리티의 데이터를 처리하도록 이전에 훈련된 개개의 통계적 모델은, 각각의 잠재적 표현을 합동-모달리티 표현에 결합함으로써 통합되어, 개개의 모델들 사이에서 정보가 공유되는 것을 허용한다.

[0043] 본 발명자들은 개개의 통계적 모델을 통합하기 위한 혁신적인 아키텍처를 개발했을 뿐만 아니라, 복수의 모달리티 각각으로부터의 훈련 데이터를 이용하여 이 아키텍처의 컴포넌트들을 비동기식으로 훈련시키고 추가 데이터가 이용가능하게 될 때 훈련된 컴포넌트의 파라미터를 업데이트하기 위한 새로운 알고리즘을 명시했다. 본 명세서에서 설명된 기술은, 임의의 적절한 수의 데이터 모달리티(예를 들어, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 등)에 대한 데이터를 수신하고 처리하도록 다중-모달 통계적 모델을 훈련시키는데 적용될 수 있다. 도 5를 참조하여 아래에서 설명된 바와 같이, 본 발명자들은, 종래 기술로는 가능하지 않은, (단백질 구조 예측의 문제를 위한) 6개의 상이한 모달리티에서 발생하는 데이터를 처리하기 위한 다중-모달 통계적 모델을 생성하기 위해 새로운 기술을 이용했다.

[0044] 비동기식 훈련을 이용하는 것은, 처음으로, 임의의 적절한 수의 데이터 모달리티에 대한 다중-모달 통계적 모델의 생성을 가능케한다는 점에서 종래 기술에 비해 개선을 제공할 뿐만 아니라, 이러한 머신 학습 시스템을 훈련시키고 배치하는데 이용되는 컴퓨터 기술도 개선한다. 특히, 본 명세서에서 설명된 다중-모달 통계적 모델은 더 적은 훈련 데이터로 훈련될 수 있어서(모든 모달리티에 걸친 링크된 훈련 데이터 인스턴스들이 요구되는 것은 아니기 때문에), 결국 이것은, 이러한 모델을 훈련시키고 배치하는데 더 적은 컴퓨팅 자원이 이용될 필요가 있다는 것을 의미한다. 구체적으로는, 더 적은 프로세서 전력과 시간이 필요하고, 더 적은 메모리가 필요하며, 이러한 데이터의 전송을 위해 더 적은 네트워킹 자원(예를 들어, 네트워크 대역폭)이 필요하고, 이를 모두는 컴퓨터 기능을 직접적으로 개선한다.

[0045] 본 발명자들에 의해 개발된 기술은, 본 발명자들에 의해 개발되고 본 명세서에서 설명되는 다중-모달 통계적 모델을 이용하고 이 훈련 기술을 통해 상이한 데이터 모달리티에 대한 통계적 모델들의 효율적인 단일화를 허용하기 때문에, 때때로 "유니티(UNITY)" 프레임워크라고도 지칭된다.

[0046] 따라서, 일부 실시예는, 제1 모달리티로부터의 입력 데이터 및 제1 모달리티와는 상이한 제2 모달리티로부터의 입력 데이터를 포함하는 복수의 모달리티로부터 입력 데이터를 수신하도록 구성된 다중-모달 통계적 모델을 훈련하기 위한 기술을 제공한다. 이 기술은: (1) 제1 모달리티에 대한 무라벨 훈련 데이터 및 제2 모달리티에 대한 무라벨 훈련 데이터를 포함하는 무라벨 훈련 데이터에 액세스하는 단계; (2) 제1 모달리티에 대한 라벨링된 훈련 데이터 및 제2 모달리티에 대한 라벨링된 훈련 데이터를 포함하는 라벨링된 훈련 데이터에 액세스하는 단계; (3) 2개의 스테이지에서 다중-모달 통계적 모델 –다중-모달 통계적 모델은, 제1 및 제2 모달리티에 대한 입력 데이터를 각각 처리하기 위한 제1 및 제2 인코더, 제1 및 제2 모달리티 임베딩, 합동-모달리티 표현, 및 예측기를 포함하는 복수의 컴포넌트를 포함함– 을 훈련시키는 단계를 포함하고, –상기 훈련시키는 단계는 : (A) 자체 감독형 학습 기술 및 무라벨 훈련 데이터를 이용하여 제1 및 제2 모달리티 임베딩의 파라미터들에 대한 값들 및 합동-모달리티 표현을 추정함으로써 적어도 부분적으로 제1 훈련 스테이지를 수행하는 단계; (B) 감독형 학습 기술 및 라벨링된 훈련 데이터를 이용하여 예측기의 파라미터들에 대한 값들을 추정함으로써 적어도 부분적으로 제2 훈련 스테이지를 수행하는 단계를 포함함–, (4) 다중-모달 통계적 모델의 복수의 컴포넌트의 파라미터들에 대한 추정된 값들을 저장함으로써 다중-모달 통계적 모델을 적어도 부분적으로 명시하는 정보를 저장하는 단계를 포함한다.

[0047] 일부 실시예에서, 제1 및 제2 인코더의 파라미터는 다중-모달 통계적 모델의 제1 훈련 스테이지에 앞서 추정될 수 있다. 이것은, 통합되는 개개의 통계적 모델이 이전에 훈련되었고 각각의 인코더의 파라미터가 추정된 경우 일 수 있다. 다른 실시예에서, 인코더의 파라미터는 다중-모달 통계적 모델의 훈련 동안 처음으로 추정되거나 및/또는 업데이트될 수 있다. 유사하게, 제1 및 제2 디코더는 다중-모달 통계적 모델의 훈련에 앞서 또는 훈련 동안에 훈련될 수 있다.

[0048] 일부 실시예에서, 합동-모달리티 표현은 N개의 차원 벡터를 포함하는 코드북일 수 있다. 통합되는 각각의 개개의 통계적 모델은 : 입력의 잠재적 표현을 생성하고 이 잠재적 표현을 이용하여 합동-모달리티 표현에서 유사한 벡터 또는 벡터들을 식별하도록 구성될 수 있다. 차례로, 식별된 벡터(들)는 예측 태스크에 이용될 수 있는 새로운 피쳐 세트를 생성하는데 이용될 수 있다. 이러한 방식으로, 하나의 모달리티에 대해 생성된 피쳐는, 공통

코드북의 이용을 통해, 상이한 모달리티에서 수집된 정보를 반영하도록 업데이트될 수 있다.

- [0049] 일부 실시예에서, 제1 훈련 스테이지를 수행하는 단계는 다음을 포함한다 : (A) 제1 모달리티에 대한 무라벨 훈련 데이터의 제1 데이터 입력에 액세스하는 단계; (B) 제1 피처 벡터를 생성하기 위해 제1 데이터 입력을 제1 인코더에 제공하는 단계; (C) 합동-모달리티 표현, 제1 모달리티 임베딩 및 제1 피처 벡터를 이용하여 제2 피처 벡터를 식별하는 단계; 및 (D) 제1 데이터 출력을 생성하기 위해 제2 피처 벡터를 입력으로서 제1 디코더에 제공하는 단계. 그 다음, 제1 데이터 출력은 제1 데이터 입력과 비교될 수 있고, 합동-모달리티 표현의 하나 이상의 파라미터 값은 비교의 결과들에 기초하여 업데이트될 수 있다(예를 들어, 스토캐스틱 경사 하강을 이용).
- [0050] 일부 실시예에서, 합동-모달리티 표현, 제1 피처 벡터 및 제1 모달리티 임베딩을 이용하여 제2 피처 벡터를 식별하는 단계는 다음을 포함한다 : (A) N개의 d차원 벡터를 획득하기 제1 모달리티 임베딩을 이용함으로써 합동-모달리티 표현을 제1 모달리티의 공간에 투사하는 단계;
- [0051] (B) N개의 d차원 벡터 중 적어도 일부와 제1 피처 벡터 사이의 유사성에 따라 합동-모달리티 표현의 N개의 d차원 벡터 중 적어도 일부에 대한 가중치들을 계산하는 단계; 및 (C) 제1 피처 벡터를 계산된 가중치들에 의해 가중된 N개의 d차원 벡터 중 적어도 일부의 가중 합과 총계함으로써 제2 피처 벡터를 생성하는 단계.
- [0052] 일부 실시예에서, 훈련되는 다중-모달 통계적 모델은 제1 및 제2 태스크 임베딩을 더 포함하고, 여기서, 다중-모달 통계적 모델을 훈련시키는 단계는, 제2 훈련 스테이지 동안, 예측기의 파라미터들의 값들을 추정하는 단계와 함께 제1 및 제2 태스크 임베딩의 파라미터들에 대한 값들을 추정하는 단계를 더 포함한다.
- [0053] 일부 실시예에서, 제1 인코더는, 콘볼루션 신경망, 순환 신경망, 또는 임의의 기타 적절한 유형의 통계적 모델 등의 신경망일 수 있다.
- [0054] 일부 실시예에서, 제1 모달리티에 대한 무라벨 훈련 데이터는 이미지들을 포함하고 제2 모달리티에 대한 무라벨 훈련 데이터는 텍스트를 포함한다. 일부 실시예에서, 제1 모달리티에 대한 무라벨 훈련 데이터는 단백질 서열 데이터를 포함하고, 제2 모달리티에 대한 무라벨 훈련 데이터는, 단백질 패밀리 데이터, 생물학적 프로세스 온톨로지 데이터, 문자 기능 온톨로지 데이터, 세포 성분 온톨로지 데이터, 또는 분류학적 종들 패밀리 데이터를 포함한다.
- [0055] 일부 실시예는, 제1 모달리티로부터의 입력 데이터, 및 제1 모달리티와는 상이한 제2 모달리티로부터의 입력 데이터를 포함하는 복수의 모달리티로부터의 입력 데이터를 수신하도록 구성된 다중-모달 통계적 모델을 이용하여 예측 태스크를 수행하기 위한 기술을 포함한다. 이 기술은: (A) 다중-모달 통계적 모델의 복수의 컴포넌트 – 복수의 컴포넌트는, 제1 및 제2 모달리티에 대한 입력 데이터를 각각 처리하기 위한 제1 및 제2 인코더, 제1 및 제2 모달리티 임베딩, 합동-모달리티 표현, 및 예측기를 포함함 – 각각의 파라미터들의 값들을 포함하는 다중-모달 통계적 모델을 명시하는 정보를 획득하는 단계; (B) 제1 데이터 모달리티에 대한 제1 입력 데이터를 획득하는 단계; (C) 제1 피처 벡터를 생성하기 위해 제1 입력 데이터를 제1 인코더에 제공하는 단계; (D) 합동-모달리티 표현, 제1 모달리티 임베딩 및 제1 피처 벡터를 이용하여 제2 피처 벡터를 식별하는 단계; 및 (E) 예측기 및 제2 피처 벡터를 이용하여 예측 태스크에 대한 예측을 생성하는 단계를 포함한다.
- [0056] 일부 실시예에서, 이 기술은 : (A) 제2 데이터 모달리티에 대한 제2 입력 데이터를 획득하는 단계; (B) 제3 피처 벡터를 생성하기 위해 제2 입력 데이터를 제2 인코더에 제공하는 단계; (C) 합동-모달리티 표현, 제2 모달리티 임베딩 및 제3 피처 벡터를 이용하여 제4 피처 벡터를 식별하는 단계를 더 포함한다. 예측 태스크에 대한 예측을 생성하는 단계는, 제2 피처 벡터 및 제4 피처 벡터를 이용하여 수행될 수 있다.
- [0057] 일부 실시예에서, 다중-모달 통계적 모델은 제1 및 제2 모달리티에 대한 제1 및 제2 태스크 임베딩을 포함하고, 예측 태스크에 대한 예측을 생성하는 단계는 : 제1 태스크 임베딩을 이용하여 제2 피처 벡터를 가중하는 단계; 제2 태스크 임베딩을 이용하여 제4 피처 벡터를 가중하는 단계; 및 가중된 제2 및 제4 피처 벡터 및 예측기를 이용하여 예측 태스크에 대한 예측을 생성하는 단계를 포함할 수 있다.
- [0058] 앞서 소개하고 나중에 더 상세히 논의된 기술은 임의의 특정한 구현 방식으로 제한되지 않기 때문에, 이 기술은 수많은 방식들 중 임의의 방식으로 구현될 수 있다는 것을 이해해야 한다. 구현의 상세사항의 예는 오로지 예시 목적을 위해 여기서 제공된다. 더욱이, 본 명세서에서 설명된 기술의 양태는 임의의 특정한 기술 또는 기술들의 조합의 이용으로 제한되지 않기 때문에, 본 명세서에서 개시된 기술은 개별적으로 또는 임의의 적절한 조합으로 이용될 수 있다.
- [0059] 도 1은, 본 명세서에 설명된 기술의 일부 실시예에 따른, 자체 감독형 학습 기술을 이용하는 단일 모달리티에

대한 통계적 모델(100)을 위한 지식 베이스의 훈련을 나타내는 다이어그램이다. 통계적 모델(100)은, 인코더(104), 디코더(110) 및 지식 베이스를 나타내는 메모리(105)를 포함하는 각각의 파라미터를 갖는 복수의 컴포넌트를 포함한다.

[0060] 이 예에서, 인코더(104) 및 디코더(110)는 좌측으로부터 우측으로 아래로 연장되는 대각선을 갖는 채우기 패턴에 의해 도시된 바와 같이 이전에 훈련되었고, 메모리(105)는 좌측으로부터 우측으로 위로 연장되는 대각선을 갖는 채우기 패턴에 의해 도시된 바와 같이 아직 훈련되지 않았다고 가정한다. 그러나, 아래에서 더 상세히 설명되는 바와 같이, 일부 실시예에서, 개개의 통계적 모델은 처음으로 훈련되거나 적어도 다중-모달 통계적 모델의 훈련 동안 업데이트될 수 있다.

[0061] 일부 실시예에서, 인코더(104)는 입력을 수신하고 (입력 데이터의 차원성보다 낮은 차원성을 가질 수 있는) 잠재적 표현을 출력하도록 구성될 수 있고, 제1 디코더는 잠재적 표현으로부터 입력 데이터를 재구성하도록 구성될 수 있다. 일부 실시예에서, 인코더 및 디코더는 자동 인코더의 일부일 수 있다. 일부 실시예에서, 통계적 모델(100)은 신경망 모델일 수 있고, 인코더(104) 및 디코더(110)는, 인코더(104) 및 디코더(110)의 파라미터가 각각의 신경망 계층에 대한 가중치를 포함하도록 하나 이상의 신경망 계층을 포함할 수 있다. 인코더(104) 및 디코더(110)는 신경망인 것으로 제한되지 않고 임의의 기타 적절한 유형의 통계적 모델일 수 있다는 것을 이해해야 한다.

[0062] 일부 실시예에서, 메모리(105)에 대한 파라미터의 값들은, 통계적 모델(100)의 출력이 가능한 한 가깝게 통계적 모델(100)에 대한 입력을 재현하도록, 자체 감독형 학습을 이용하여 추정될 수 있다. 따라서, 일부 실시예에서, 훈련 동안, 통계적 모델(100)의 출력은 입력과 비교되고, 메모리(105)의 파라미터 값은, 스토캐스틱 경사 하강(인코더와 디코더가 신경망일 때 역 전파를 이용하여 경사가 계산됨) 또는 임의의 기타 적절한 훈련 알고리즘을 이용하여, 입력과 출력 사이의 거리 측정치에 기초하여 반복적으로 업데이트된다.

[0063] 예를 들어, 일부 실시예에서, 훈련 데이터는 입력(102)으로서 제1 인코더(104)에 제공될 수 있다. 인코더(104)는, 입력(102)에 기초하여, 제1 피처 표현(106)을 생성한다. 피처 표현(106)은, 메모리(105)를 이용하여, 제2 피처 표현(108)을 획득하는데 이용된다. 일부 실시예에서, 메모리(105)는 피처 표현(106)의 것과 동일한 차원성을 갖는 복수의 벡터를 저장할 수 있다. 예를 들어, 피처 표현(108)은 d차원 벡터일 수 있고 메모리(105)는 N개의 d차원 벡터를 저장할 수 있다. 일부 실시예에서, 제2 피처 표현(108)은, 메모리(105) 내의 벡터들 중에서 제1 피처 표현(106)과 가장 유사한(코사인 유사성, 유클리드 거리 등의 적절한 유사성 측정치에 따라) 벡터를 선택하고, (합산, 곱셈, 산술 평균, 기하학적 평균, 또는 임의의 기타 적절한 연산일 수 있는) 총계 연산(107)을 통해 선택된 벡터를 피처 표현(106)에 추가함으로써 획득될 수 있다. 일부 실시예에서, 제2 피처 표현(108)은, 메모리(105) 내의 벡터들의 가중된 선형 조합을 피처 표현(106)과 총계함으로써 생성될 수 있고, 여기서 각각의 벡터에 대한 가중치는 그 벡터와 피처 표현(106) 사이의 거리에 비례한다. 제2 피처 표현은 디코더(110)에 대한 입력으로서 제공된다. 차례로, 디코더(110)의 출력은 인코더(104)에 제공된 입력과 비교될 수 있고, 메모리(105)의 파라미터 값들 중 적어도 일부는, 인코더(104)에 대한 입력과 디코더(110)의 출력 사이의 차이에 기초하여 업데이트될 수 있다.

[0064] 도 1을 참조하여 설명된 실시예에서는, 인코더(104) 및 디코더(110)가 훈련된 것으로 가정되었지만, 다른 실시예에서, 인코더(104) 및 디코더(110)의 파라미터 값은 처음으로 추정되거나 및/또는 메모리(105)의 파라미터 값이 추정되는 것과 동시에 업데이트될 수 있다.

[0065] 도 1의 예시적인 예는, 이전에 훈련된 복수의 통계적 모델을 단일 다중-모달 통계적 모델로 통합하기 위해 본 발명자들에 의해 개발된 기술을 이해하는데 도움이 된다. 특히, 본 명세서에서 설명된 바와 같이, 다중-모달 통계적 모델은 합동-모달리티 표현을 통해 상이한 모달리티들간의 정보의 공유를 허용한다. 단일 모달리티에 대한 통계적 모델(100)의 훈련 및 이용 동안에 액세스되는 메모리(105)처럼, 합동-모달리티 표현(예를 들어, 도 2a 및 도 2b에 도시된 지식 베이스(230))은 본 명세서에서 설명된 다중-모달 통계적 모델(예를 들어, 모델 250)의 훈련 및 이용 동안에 액세스된다.

[0066] 본 명세서에 설명된 바와 같이, 합동-모달리티 표현이 액세스되어 하나의 모달리티에 대한 계산을 수행할 때, 그 내용은 먼저, 모달리티 임베딩을 이용하여 그 모달리티에 투사될 수 있다. 이러한 모달리티 투사는 본 명세서에서 설명된 다중-모달 통계적 모델의 일부를 형성한다.

[0067] 도 1과 관련하여 설명된 바와 같이, 단일-모달리티 통계적 모델(100)은, 이전에 훈련된 인코더(104), 디코더(110) 및 (임의의 분류 태스크에 관하여 라벨링될 필요가 없는) 훈련 데이터를 이용한 자체 감독형 학습을 이용

하여 훈련될 수 있는 메모리(105)를 포함한다. 본 발명자들에 의해 개발된 다중-모달 통계적 모델(예를 들어, 다중-모달 통계적 모델 250)은 합동-모달리티 표현(예를 들어, 지식 베이스(230)) 및 복수의 모달리티 임베딩(예를 들어, 모달리티 임베딩(232))을 포함하며, 이것은, 도 2a, 도 2b 및 도 3의 참조를 포함한 본 명세서에서 설명된 바와 같이 자체 감독형 학습을 이용하여 훈련될 수 있고, 도 2b 및 도 4의 참조를 포함한 본 명세서에서 설명된 바와 같이 예측에 이용될 수 있다.

[0068] 일부 실시예에서, 본 발명자들에 의해 개발된 다중-모달 통계적 모델은 2-스테이지 훈련 절차를 이용하여 훈련될 수 있다. 제1 훈련 스테이지는, 자체 감독형 훈련 기술을 이용하여 수행되며 합동-모달리티 표현 및 모달리티 임베딩의 파라미터 학습을 포함한다. 제2 스테이지는 감독형 훈련 기술을 이용하여 수행되며 (적절한 예측 태스크를 위한) 예측기 및 태스크 임베딩의 파라미터 학습을 포함한다. 도 2a 및 도 2b는, 일부 실시예에서, 이들 2개의 스테이지 각각에서 다중-모달 통계적 모델의 어떤 컴포넌트가 학습되는지를 나타낸다.

[0069] 도 2a는, 본 명세서에 설명된 기술의 일부 실시예에 따른, 자체 감독형 학습 기술을 이용한 다중-모달 통계적 모델의 제1 훈련 스테이지를 나타내는 다이어그램이다. 도 2a에 도시된 바와 같이, 통계적 모델은, 제1 모달리티에 대한 인코더(204), 제2 모달리티에 대한 인코더(214), 지식 베이스(230), 및 제1 및 제2 모달리티 각각에 대한 모달리티 임베딩을 포함한 모달리티 임베딩(232)을 포함하는, 각각의 파라미터를 갖는 다수의 컴포넌트를 포함한다. 추가적으로, 도 2a에 도시된 바와 같이, 훈련 환경(200)은 제1 모달리티에 대한 디코더(210) 및 제2 모달리티에 대한 디코더(220)를 포함한다. 이들 디코더는 다중-모달 통계적 모델의 일부가 아니며, 오히려, 디코더들은 자체 감독형 훈련 스테이지에서 다중-모달 통계적 모델을 훈련시키는데 이용된다; 이들은 도 2b로부터 알 수 있는 바와 같이 예측에 이용되지 않는다.

[0070] 도 2a에 나타낸 실시예에서, 인코더(204 및 214) 및 디코더(210 및 220)는 좌측으로부터 우측으로 아래로 연장되는 대각선을 갖는 채우기 패턴에 의해 도시된 바와 같이 이전에 훈련되었고, 지식 베이스(230) 및 모달리티 임베딩(232)은 좌측으로부터 우측으로 위로 연장되는 대각선을 갖는 채우기 패턴으로 나타낸 바와 같이 아직 훈련되지 않았다고 가정한다. 그러나, 본 명세서의 일부 실시예에서 설명된 바와 같이, 인코더들 및 디코더들 중 하나 이상은 처음으로 훈련되거나 적어도 다중-모달 통계적 모델의 훈련 동안 업데이트될 수 있다.

[0071] 일부 실시예에서, 인코더(204), 인코더(214), 디코더(210) 및 디코더(220) 각각은, 하나 이상의 신경망 계층을 포함하는 상응하는 신경망일 수 있다. 계층은, 하나 이상의 콘볼루션 계층, 하나 이상의 폴링 계층, 하나 또는 서브샘플링 계층, 하나 이상의 완전 접속된 계층, 및/또는 임의의 기타 적절한 계층(들)을 포함할 수 있다. 그러나, 인코더(204 및 214) 및 디코더(210 및 220) 중 어느 것도 신경망 모델로 제한되지 않으며, 본 명세서에서 설명된 기술의 양태는 이 점에서 제한되지 않기 때문에 임의의 기타 적절한 유형의 통계적 모델(들)일 수 있다.

[0072] 일부 실시예에서, (합동-모달리티 표현의 한 예인) 지식 베이스(230)는 N개의 m차원 벡터를 포함할 수 있다. 본 명세서에서 설명된 기술의 양태는 이 점에서 제한되지 않기 때문에, 이들 벡터는 행렬(예를 들어,  $N \times m$  행렬) 또는 임의의 기타 적절한 데이터 구조(들)를 이용하여 저장 및/또는 표현될 수 있다.

[0073] 일부 실시예에서, 각각의 모달리티 임베딩은 지식 베이스(230)를 각각의 모달리티 공간에 투사하도록 구성될 수 있다. 예를 들어, 일부 실시예에서, 제1 모달리티에 대한 (모달리티 임베딩들(232) 중의) 모달리티 임베딩은, 지식 베이스(230)의 제1 모달리티 뷰(238)를 획득하기 위해 투사 연산(237)을 이용하여 지식 베이스(230)를 제1 모달리티에 투사하는데 이용될 수 있다. 투사 연산은 제1 모달리티에 대한 모달리티 임베딩의 가중치(234) 부분을 이용할 수 있다. 또 다른 예로서, 일부 실시예에서, 제2 모달리티에 대한 (모달리티 임베딩들(232) 중의) 모달리티 임베딩은, 지식 베이스(230)의 제2 모달리티 뷰(240)를 획득하기 위해 투사 연산(239)을 이용하여 지식 베이스(230)를 제2 모달리티에 투사하는데 이용될 수 있다. 투사 연산은 제2 모달리티에 대한 모달리티 임베딩의 가중치(236) 부분을 이용할 수 있다.

[0074] 일부 실시예에서, 각각의 모달리티 임베딩은, 투사된 지식 베이스의 벡터의 차원성이 그 모달리티 공간에서의 잠재적 표현의 차원성과 정합하도록 지식 베이스(230)를 각각의 모달리티 공간으로 투사하도록 구성될 수 있다. 예를 들어, 지식 베이스(230)이  $N = 512$  및  $m = 64$ 인 N개의 m차원 벡터를 포함하고, 제1 모달리티에 대해 인코더에 의해 생성된 잠재적 표현이  $d = 10$ 인 d차원 벡터라고 가정한다. 이 예에서, 제1 모달리티에 대한 모달리티 임베딩은,  $512 \times 64$  지식 베이스(230)에 적용될 때 제1 모달리티에 대한 지식 베이스(230)의  $512 \times 10$ 개의 뷰를 생성하는  $m \times d$ ( $64 \times 10$ ) 행렬일 수 있다. 제2 모달리티에 대해 인코더에 의해 생성된 잠재적 표현은  $p = 12$ 인 p차원 벡터라고 더 가정한다. 그 다음, 제1 모달리티에 대한 모달리티 임베딩은,  $512 \times 64$  지식 베이스(230)에 적용될 때 제2 모달리티에 대한 지식 베이스(230)의  $512 \times 12$ 개의 뷰를 생성하는  $m \times p$ ( $64 \times 12$ ) 행렬일 수 있다. 전술된 예로부터 이해할 수 있는 바와 같이, 모달리티 임베딩은, 무엇보다도, 잠재적 표현의 차원성이 동일

하지 않은 상황(예를 들어, 하나의 모달리티에서는 10차원이고 또 다른 모달리티에서는 12차원)에서 상이한 모달리티에 대한 통계적 모델의 통합을 허용한다.

[0075] 다중-모달 통계적 모델의 제1(자체 감독형) 훈련 스테이지의 양태는 도 3을 참조하여 아래에서 상세히 설명된다.

[0076] 도 2b는, 본 명세서에 설명된 기술의 일부 실시예에 따른, 감독형 학습 기술을 이용한 다중-모달 통계적 모델(250)의 제2 훈련 스테이지를 나타내는 다이어그램이다. 도 2b에 도시된 바와 같이, 다중-모달 통계적 모델(250)은 예측 테스크(256) 및 테스크 임베딩(254)에 대한 예측기(252)를 포함한다.

[0077] 도 2b에 나타낸 실시예에서, 인코더(204 및 214), 디코더(210 및 220), 지식 베이스(230), 및 모달리티 임베딩(232)은, 좌측으로부터 우측으로 아래로 연장되는 대각선을 갖는 채우기 패턴에 의해 도시된 바와 같이 이전에 훈련되었고, 예측기(252) 및 테스크 임베딩(254)은 좌측으로부터 우측으로 위로 연장되는 대각선을 갖는 채우기 패턴에 의해 도시된 바와 같다고 가정한다. 그러나, 본 명세서에 설명된 바와 같이, 일부 실시예에서, 인코더, 디코더, 모달리티 임베딩 및 합동-모달리티 표현 중 하나 이상은 처음으로 훈련되거나 적어도 다중-모달 통계적 모델의 훈련의 제2 스테이지 동안 업데이트될 수 있다.

[0078] 일부 실시예에서, 예측기(252)는 입력 피처를 출력(예를 들어, 분류기의 경우 불연속 라벨 또는 회귀기의 경우 연속 변수)에 맵핑하는 임의의 적절한 유형의 통계적 모델일 수 있다. 예를 들어, 본 명세서에서 설명된 기술의 양태는 이 점에서 제한되지 않기 때문에, 예측기(252)는, 선형 모델(예를 들어, 선형 회귀 모델), 일반화된 선형 모델(예를 들어, 로지스틱 회귀, 프로빗 회귀), 신경망 또는 기타의 비선형 회귀 모델, Gaussian 혼합 모델, 지원 벡터 머신, 결정 트리 모델, 랜덤 포레스트 모델, Bayesian 계층 모델, Markov 랜덤 필드, 및/또는 임의의 기타 적절한 유형의 통계적 모델을 포함할 수 있다.

[0079] 일부 실시예에서, 테스크 임베딩(254)은, 연산(256 및 258)을 통해, 제1 및 제2 모달리티로부터의 피처의 기여도를 가중하는데 이용될 수 있다. 예를 들어, 도 2b에 도시된 바와 같이, 피처 표현(208)은 제1 모달리티에 대한 테스크 임베딩을 이용하여 연산(256)을 통해 가중될 수 있고, 피처 표현(218)은 제2 모달리티에 대한 테스크 임베딩을 이용하여 연산(258)을 통해 가중될 수 있다. 이들 가중된 피처 표현은 예측기(252)에 대한 입력을 생성하기 위해 연산(260)을 통해 (예를 들어, 가중된 합 또는 곱으로서) 총계될 수 있다. 테스크 임베딩에 의해 피처 표현에 유도된 가중은 포인트별 곱셈 가중(예를 들어, Hadamard 곱)일 수 있다.

[0080] 다중-모달 통계적 모델의 제2(감독형) 훈련 스테이지의 양태가 도 3을 참조하여 아래에서 상세히 설명된다.

#### 다중-모달 통계적 모델의 훈련

[0082] 도 3은, 본 명세서에서 설명된 기술의 일부 실시예에 따른, 2-스테이지 훈련 절차를 이용하여 다중-모달 통계적 모델을 훈련시키기 위한 예시적인 프로세스(300)의 플로차트로서, 제1 스테이지는 자체 감독형 학습을 포함하고 제2 스테이지는 감독형 학습을 포함한다. 프로세스(300)는 임의의 적절한 컴퓨팅 디바이스(들)에 의해 수행될 수 있다. 본 명세서에서 설명된 기술의 양태는 이 점에서 제한되지 않기 때문에, 예를 들어, 프로세스(300)는, 하나 이상의 그래픽 처리 유닛(GPU), 클라우드 컴퓨팅 서비스에 의해 제공되는 하나 이상의 컴퓨팅 디바이스(들), 및/또는 임의의 기타 적절한 컴퓨팅 디바이스(들)에 의해 수행될 수 있다.

[0083] 도 3에 나타내고 아래에서 설명되는 실시예에서, 프로세스(300)는 2개의 모달리티 – 제1 모달리티 및 제2 모달리티로부터 입력을 수신하도록 구성된 다중-모달 통계적 모델을 훈련시키는데 이용된다. 그러나, 본 명세서에서 설명된 기술의 양태는 이 점에서 제한되지 않기 때문에, 프로세스(300)는 임의의 적절한 수의 모달리티(예를 들어, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 등)로부터 입력을 수신하도록 구성된 다중-모달 통계적 모델을 훈련시키는데 이용될 수 있다는 것을 이해해야 한다.

[0084] 이 예에서, 프로세스(300)의 시작에 앞서, 개개의 통계적 모델이 제1 및 제2 모달리티에 대해 훈련되었다고 가정한다. 특히, 제1 인코더 및 제1 디코더를 포함하는 제1 통계적 모델은 제1 모달리티에 대해 훈련되었고 제2 인코더 및 제2 디코더를 포함하는 제2 통계적 모델은 제2 모달리티에 대해 훈련되었다고 가정한다. 제1 통계적 모델은 제1 모달리티의 데이터를 이용하여 훈련된 자동 인코더 유형의 통계적 모델일 수 있다. 제2 통계적 모델은 제2 모달리티의 데이터를 이용하여 훈련된 자동 인코더 유형의 통계일 수 있다. 그러나, 아래에서 더 상세히 설명되는 바와 같이, 일부 실시예에서, 개개의 통계적 모델은 처음으로 훈련되거나 적어도 다중-모달 통계적 모델의 훈련 동안 업데이트될 수 있다.

[0085] 일부 실시예에서, 프로세스(300)의 실행 동안 훈련되는 다중-모달 통계적 모델은 각각의 모달리티에 대한 인코

더 컴포넌트, 합동-모달리티 표현 컴포넌트, 각각의 모달리티에 대한 모달리티 임베딩 컴포넌트, 예측기 컴포넌트, 및 각각의 모달리티에 대한 태스크 임베딩 컴포넌트를 포함할 수 있고, 프로세스(300)는 하나 이상의 이들 컴포넌트 각각에 대한 파라미터 값을 추정하는데 이용될 수 있다. 예를 들어, 도 2b의 다중-모달 통계적 모델(250)은, 인코더(204), 인코더(214), 지식 베이스(230), 모달리티 임베딩(232), 예측기(252) 및 태스크 임베딩(254)을 포함하고, 컴포넌트(230, 232, 252, 254)의 파라미터는 프로세스(300)의 일부로서 추정될 있다. (통합되는 개개의 통계적 모델의 일부일 수 있는) 복수의 모달리티 각각에 대한 디코더는 다중-모달 통계적 모델의 일부가 아닐 수 있다는 것을 이해해야 한다. 그럼에도 불구하고, 이러한 디코더는 아래에서 더 상세히 설명되는 바와 같이 자체 감독형 학습 스테이지에서 다중-모달 통계적 모델을 훈련시키는데 이용될 수 있다.

[0086] 프로세스(300)는, 제1 모달리티에 대한 제1 훈련된 통계적 모델의 파라미터 및 제2 모달리티에 대한 제2 훈련된 통계적 모델의 파라미터가 액세스되는 동작 302에서 시작한다. 파라미터는, 로컬 스토리지로부터 액세스되거나, 네트워크를 통해 원격 스토리지로부터 액세스되거나, 임의의 기타 적절한 소스로부터 액세스될 수 있다.

[0087] 일부 실시예에서, 제1 훈련된 통계적 모델은 자동 인코더를 포함할 수 있고, 제1 인코더 및 제1 디코더를 포함할 수 있으며, 각각은 동작 302에서 액세스될 수 있는 상응하는 파라미터 세트를 갖는다. 제1 인코더는, 제1 모달리티를 갖는 데이터를 입력으로서 수신하고 (입력 데이터의 차원성보다 낮은 차원성을 가질 수 있는) 잠재적 표현을 출력하도록 구성될 수 있고, 제1 디코더는 잠재적 표현으로부터 입력 데이터를 재구성하도록 구성될 수 있다. 일부 실시예에서, 제1 훈련된 통계적 모델은 신경망(예를 들어, 피드포워드 신경망, 콘볼루션 신경망, 순환 신경망, 완전 접속된 신경망 등)일 수 있고, 제1 인코더 및 제1 디코더는, 제1 인코더 및 디코더의 파라미터가 각각의 신경망 계층에 대한 가중치를 포함하도록 하나 이상의 신경망 계층을 포함할 수 있다. 그러나, 제1 훈련된 통계적 모델은 신경망인 것으로 제한되지 않고 임의의 기타 적절한 통계적 모델일 수 있다는 것을 이해해야 한다.

[0088] 일부 실시예에서, 제2 훈련된 통계적 모델은 자동 인코더를 포함할 수 있고, 제2 인코더 및 제2 디코더를 포함할 수 있으며, 각각은 동작 302에서 액세스될 수 있는 상응하는 파라미터 세트를 갖는다. 제2 인코더는, 제2 모달리티를 갖는 데이터를 입력으로서 수신하고 (입력 데이터의 차원성보다 낮은 차원성을 가질 수 있는) 잠재적 표현을 출력하도록 구성될 수 있고, 제2 디코더는 잠재적 표현으로부터 입력 데이터를 재구성하도록 구성될 수 있다. 일부 실시예에서, 제2 훈련된 통계적 모델은 신경망(예를 들어, 피드포워드 신경망, 콘볼루션 신경망, 순환 신경망, 완전 접속된 신경망 등)일 수 있고, 제2 인코더 및 제2 디코더는, 제2 인코더 및 디코더의 파라미터가 각각의 신경망 계층에 대한 가중치를 포함하도록 하나 이상의 신경망 계층을 포함할 수 있다. 그러나, 제2 훈련된 통계적 모델은 신경망인 것으로 제한되지 않고 임의의 기타 적절한 통계적 모델일 수 있다는 것을 이해해야 한다.

[0089] 일부 실시예에서, 제1 인코더 및 제2 인코더는 상이한 모달리티의 데이터를 수신하도록 구성되기 때문에 서로 상이하다. 이러한 실시예에서, 제1 및 제2 디코더는 서로 상이하다. 일부 이러한 실시예에서, 인코더들 각각이 신경망으로서 구현될 때, 인코더들에 대한 신경망 아키텍처는 상이하다(예를 들어, 상이한 수의 계층, 상이한 유형의 계층, 상이한 차원성의 계층, 상이한 비선형성 등). 한 예로서, 제1 인코더는 입력으로서 이미지를 수신하고 이미지의 잠재적 표현을 생성하도록 구성될 수 있고, 제2 인코더는 입력으로서 텍스트를 수신하고 텍스트의 잠재적 표현을 생성하도록 구성될 수 있다. 또 다른 예로서, 제1 인코더는 단백질 서열 데이터의 잠재적 표현을 수신 및 생성하도록 구성될 수 있고, 제2 인코더는 단백질 패밀리 데이터의 잠재적 표현을 수신 및 생성하도록 구성될 수 있다. 역시 또 다른 예로서, 제1 인코더는 제1 유형(예를 들어, 초음파)의 의료 이미지의 잠재적 표현을 수신 및 생성하도록 구성될 수 있고, 제2 인코더는 제1 유형과는 상이한 제2 유형(예를 들어, MRI 이미지)의 의료 이미지의 잠재적 표현을 수신 및 생성하도록 구성될 수 있다.

[0090] 일부 실시예에서, 제1 인코더의 출력에서 생성된 잠재적 표현은 제2 인코더의 출력에서 생성된 잠재적 표현과 동일한 차원성을 가질 수 있다. 예를 들어, 아래에서 더 상세히 설명되는 바와 같이, 제1 인코더는 단백질 서열(예를 들어, 20 x 1024 원-핫-인코딩된(one-hot-encoded) 단백질 서열)의 표현을 입력으로서 수신하고 10x1 잠재적 표현을 반환할 수 있다. 동일한 이 예에서, 제2 인코더는 입력으로서 (예를 들어, 24,937-차원 벡터로서 원-핫-인코딩될 수 있는) 생물학적 프로세스 입력을 수신하고 10x1 잠재적 표현을 반환할 수 있다. 그러나, 상이한 모달리티 임베딩을 이용하는 것은 추가적인 유연성을 제공하므로 잠재적 표현들이 동일한 차원이 될 것이 요구되지 않아서, 상이한 모달리티에 대한 잠재적 표현은 상이한 차원을 갖는다.

[0091] 도 2a는 동작 302에서 액세스될 수 있는 파라미터의 한 예를 나타낸다. 특히, 인코더(204)(제1 인코더), 디코더(210)(제1 디코더), 인코더(214)(제2 인코더) 및 디코더(218)의 파라미터는 동작 302에서 액세스될 수 있다.

[0092]

그 다음, 프로세스(300)는, 제1 및 제2 모달리티 각각에 대해 무라벨 훈련 데이터가 액세스되는 동작 303으로 진행한다. 동작 303에서 액세스된 무라벨 훈련 데이터는, 동작 306에서 자체 감독형 학습을 이용하여 다중-모달 통계적 모델을 훈련시키는 제1 스테이지에 이용될 수 있다. 제1 훈련 스테이지의 일부로서, 무라벨 훈련 데이터는 다중-모달 통계적 모델의 하나 이상의 컴포넌트의 파라미터를 추정하는데 이용될 수 있으며, 그 컴포넌트들은 (그 파라미터가 동작 302에서 액세스되었던) 제1 및 제2 통계적 모델을 통합하는 것을 허용한다. 예를 들어, 다중-모달 통계적 모델(예를 들어, 도 2b에 도시된 모델 250)은 합동-모달리티 표현(예를 들어, 지식 베이스 230), 제1 모달리티 임베딩(예를 들어, 모달리티 임베딩들 232 중의 일부) 및 제2 모달리티 임베딩(예를 들어, 모달리티 임베딩들 232 중의 일부)을 포함할 수 있고, 무라벨 훈련 데이터는, 동작 306 동안 합동-모달리티 표현, 제1 모달리티 임베딩, 및 제2 모달리티 임베딩의 파라미터를 추정하는데 이용될 수 있다.

[0093]

동작 303에서 액세스된 무라벨 훈련 데이터는 제1 및 제2 모달리티 각각에 대한 훈련 데이터를 포함하지만, 이들 데이터는 동기식으로 또는 조율된 방식으로 수집될 필요가 없다는 것을 이해해야 한다. 제1 모달리티에 대한 무라벨 훈련 데이터는 제2 모달리티에 대한 무라벨 훈련 데이터와는 독립적으로 생성되었을 수 있다. 상이한 모달리티들에 대한 무라벨 훈련 데이터는 상이한 엔티티들에 의해 상이한 시간들에 생성되거나 및/또는 상이한 데이터베이스들에 저장될 수 있다. 제1 모달리티에 대한 훈련 데이터가 제2에 대한 것보다 많거나 그 반대일 수 있다. 제1 및 제2 모달리티에 대한 훈련 데이터는 쌍을 이를 필요가 없다 – 일대일 대응이 있을 필요가 없다. 일부 실시예에서, 동작 303에서 획득된 훈련 데이터는 라벨링될 수 있지만, 동작 306에서 훈련 데이터가 제1 훈련 스테이지 동안 이용될 때 라벨은 폐기되거나 무시될 수 있다.

[0094]

그 다음, 프로세스(300)는 제1 및 제2 모달리티에 대해 라벨링된 훈련 데이터가 액세스되는 동작 304로 진행한다. 동작 304에서 액세스된 라벨링된 훈련 데이터는, 동작 308에서 감독형 학습을 이용하여 다중-모달 통계적 모델을 훈련시키는 제2 스테이지에 이용될 수 있다. 제2 훈련 스테이지의 일부로서, 라벨링된 훈련 데이터는 다중-모달 통계적 모델의 하나 이상의 컴포넌트의 파라미터를 추정하는데 이용될 수 있으며, 그 컴포넌트들은 (그 파라미터가 동작 302에서 액세스되었던) 제1 및 제2 통계적 모델을 통합하고 이를 모델을 이용하여 예측 테스크를 수행하는 것을 허용한다. 예를 들어, 다중-모달 통계적 모델(예를 들어, 도 2b에 도시된 모델 250)은 예측기(예를 들어, 예측기(252)), 제1 태스크 임베딩(예를 들어, 태스크 임베딩들(254) 중의 일부) 및 제2 모달리티 임베딩(예를 들어, 태스크 임베딩들(254) 중의 일부)을 포함할 수 있고, 라벨링된 훈련 데이터는, 동작 308 동안, 예측기, 제1 태스크 임베딩 및/또는 제2 모달리티 임베딩의 파라미터를 추정하는데 이용될 수 있다.

[0095]

동작 304에서 액세스된 라벨링된 훈련 데이터는 제1 및 제2 모달리티 각각에 대한 훈련 데이터를 포함하지만, 이들 데이터는 동기식으로 또는 조율된 방식으로 수집될 필요가 없다. 제1 모달리티에 대한 라벨링된 훈련 데이터는 제2 모달리티에 대한 라벨링된 훈련 데이터와는 독립적으로 생성되었을 수 있다. 상이한 모달리티들에 대한 라벨링된 훈련 데이터는 상이한 엔티티들에 의해 상이한 시간들에 생성되거나 및/또는 상이한 데이터베이스들에 저장될 수 있다. 제1 모달리티에 대한 훈련 데이터가 제2에 대한 것보다 많거나 그 반대일 수 있다. 제1 및 제2 모달리티에 대한 훈련 데이터는 쌍을 이를 필요가 없다 – 일대일 대응이 있을 필요가 없다.

[0096]

그 다음, 프로세스(300)는 다중-모달 통계적 모델이 2-스테이지 절차를 이용하여 훈련되는 동작 305로 진행한다. 처음에, 동작 306에서, 동작 303에서 획득된 무라벨 데이터는, 자체 감독형 학습 기술을 이용하여 다중-모달 통계적 모델의 하나 이상의 컴포넌트에 대한 파라미터 값을 추정하는데 이용된다. 그 다음, 동작 308에서, 동작 304에서 획득된 라벨링된 데이터는 감독형 학습 기술을 이용하여 다중-모달 통계적 모델의 하나 이상의 추가 컴포넌트에 대한 파라미터 값을 추정하는데 이용된다. 이들 동작들 각각은 아래에서 더 상세히 설명된다.

[0097]

일부 실시예에서, 동작 306은 자체 감독형 학습 기술을 이용하여 다중-모달 통계적 모델의 하나 이상의 컴포넌트에 대한 파라미터 값을 추정하는 단계를 포함할 수 있다. 일부 실시예에서, 합동-모달리티 표현(예를 들어, 도 2b의 예에서 지식 베이스(230))의 파라미터는 동작 306에서 추정될 수 있다. 추가적으로, 일부 실시예에서, 하나 이상의 모달리티 임베딩(예를 들어, 모달리티 임베딩들(232) 중의 하나 이상)의 파라미터가 동작 306에서 추정될 수 있다.

[0098]

일부 실시예에서, 동작 306의 일부로서 추정된 파라미터 값을 자체 감독형 학습을 이용하여 추정될 수 있다. 자체 감독형 학습을 이용하여 통계적 모델을 훈련시키는 것은 출력에서 입력을 재현하기 위해 통계적 모델을 훈련시키는 것을 포함할 수 있다. 따라서, 일부 실시예에서, 특정한 데이터가 통계적 모델에 대한 입력으로서 제공될 수 있고 통계적 모델의 출력은 바로 그 동일한 특정한 데이터와 비교될 수 있다. 차례로, 통계적 모델의 파라미터의 하나 이상의 값을 통계적 모델의 출력과 통계적 모델에 제공된 특정한 데이터 사이의 차이에 기초하여

(예를 들어, 스토캐스틱 경사 하강 또는 임의의 기타 적절한 훈련 알고리즘을 이용하여) 업데이트될 수 있고, 그 차이는, 통계적 모델의 출력이, 그 현재 세트의 파라미터 값들과 함께 동작될 때 입력을 얼마나 잘 재현하는지의 척도를 제공한다.

[0099] 일부 실시예에서, 동작 303에서 액세스된 무라벨 훈련 데이터는 다중-모달 통계적 모델에서 합동-모달리티 표현 및 모달리티 임베딩의 파라미터 값을 추정하는데 이용될 수 있다. 파라미터 값은, 예를 들어 스토캐스틱 경사 하강 등의 반복 학습 알고리즘을 이용하여 추정될 수 있다. 반복 학습 알고리즘은, 다중-모달 통계적 모델의 인코더에 대한 입력으로서 무라벨 훈련 데이터의 적어도 일부를 제공하고, 각각의 디코더를 이용하여 출력을 생성하고, 입력을 생성된 출력과 비교하고, 입력과 출력 사이의 차이에 기초하여 합동-모달리티 표현 및/또는 모달리티 임베딩의 파라미터 값을 업데이트하는 것을 포함할 수 있다.

[0100] 예를 들어, 일부 실시예에서, 제1 모달리티의 훈련 데이터는 제1 모달리티에 대한 제1 인코더(예를 들어, 인코더(204))로의 입력으로서 제공될 수 있다. 제1 인코더의 출력(예를 들어, 피처 표현(206)), 합동-모달리티 표현(예를 들어, 지식 베이스(230)), 및 제1 모달리티 임베딩(예를 들어, 모달리티 임베딩들(232) 중 하나)은, 제1 모달리티에 대한 제1 디코더(예를 들어, 디코더(210))에 대한 입력(예를 들어, 피처 표현(208))을 생성하는데 이용될 수 있다. 차례로, 디코더(210)의 출력은 제1 인코더에 제공된 입력과 비교될 수 있고, 합동-모달리티 표현 및/또는 제1 모달리티 임베딩의 파라미터 값을 중 적어도 일부는, 제1 인코더로의 입력과 제1 디코더의 출력 사이의 차이에 기초하여 업데이트될 수 있다.

[0101] 이 예에서, 제1 인코더의 출력으로부터 제1 디코더에 대한 입력을 생성하는 단계는 다음을 포함할 수 있다 : (1) 복수의 투사된 벡터를 획득하기 위해 합동-모달리티 표현을 제1 모달리티의 공간에 투사하는 단계; (2) 복수의 투사된 벡터 각각과 제1 인코더의 출력 사이의 거리(예를 들어, 코사인 거리 및/또는 임의의 기타 적절한 유형의 거리 측정치)를 계산하고 이를 거리를 이용하여 투사된 벡터에 대한 가중치를 (예를 들어, 소프트-맥스 가중을 이용함으로써) 계산하는 단계; 및 (3) 계산된 가중치들에 의해 가중된 투사된 벡터들의 가중 합을 제1 인코더의 출력과 총계함으로써 제1 디코더에 대한 입력을 생성하는 단계. 예를 들어, 합동-모달리티 표현은, ( $N \times m$  행렬로 표현되거나 및/또는 저장될 수 있는)  $N$ 개의  $m$ 차원 벡터를 포함할 수 있고,  $m \times d$ 로서 표현될 수 있는 제1 모달리티 투사를 이용하여 합동-모달리티 표현을 제1 모달리티에 투사하는 것은, ( $N \times d$  행렬로서 표현될 수 있는)  $N$ 개의  $d$ 차원 벡터를 생성할 수 있다. 제1 인코더의 출력(예를 들어, 도 2a에 도시된 피처 표현(206))과  $N$ 개의  $d$ 차원 벡터 각각 사이의 거리가 계산되고  $N$ 개의  $d$ 차원 벡터 각각에 대한 가중치를 획득하는데 이용될 수 있다. 그 다음, 제1 디코더에 대한 입력(예를 들어, 피처 표현(208))은  $N$ 개의  $d$ 차원 벡터의 가중 합과 피처 표현(206)의 총계(207)(예를 들어, 합, 곱, 산술 평균, 기하 평균)로서 계산될 수 있으며, 여기서, 벡터들은 계산된 가중치들에 의해 가중된다. 다른 실시예에서, 본 명세서에서 설명된 기술의 양태가 이 점에서 제한되지 않기 때문에, 제1 디코더에 대한 입력은, 투사된 합동-모달리티 표현의 복수의  $d$ 차원 벡터의 가중 평균이 아니라 적절하게 선택된 거리 측정치(예를 들어, 코사인 거리)에 따라 제1 인코더의 출력과 제1 인코더의 출력에 가장 가까운  $N$ 개의  $d$ 차원 벡터의 합일 수 있다. 역시 다른 실시예에서, 제1 디코더에 대한 입력은, (위에서 논의된 바와 같이 계산되는)  $N$ 개의  $d$ 차원 벡터의 가중 합이거나 (전술된 바와 같이 식별되는) 제1 인코더의 출력과 가장 유사한 벡터일 수 있지만, 제1 인코더의 출력과 총계되지 않는다.

[0102] 또 다른 예로서, 일부 실시예에서, 제2 모달리티의 훈련 데이터는 제2 모달리티에 대한 제2 인코더(예를 들어, 인코더(214))에 대한 입력으로서 제공될 수 있다. 제2 인코더의 출력(예를 들어, 피처 표현(216)), 합동-모달리티 표현(예를 들어, 지식 베이스(230)), 및 제2 모달리티 임베딩(예를 들어, 모달리티 임베딩들(232) 중의 하나)은, 총계 연산(217)을 이용하여 제2 모달리티에 대한 제2 디코더(예를 들어, 디코더(220))로의 입력(예를 들어, 피처 표현(218))을 생성하는데 이용될 수 있다. 차례로, 디코더(220)의 출력은 제2 인코더에 제공된 입력과 비교될 수 있고, 합동-모달리티 표현 및/또는 제2 모달리티 임베딩의 파라미터 값을 중 적어도 일부는, 제2 인코더로의 입력과 제2 디코더의 출력 사이의 차이에 기초하여 업데이트될 수 있다.

[0103] 일부 실시예에서, 동작 308은 감독형 학습 기술을 이용하여 다중-모달 통계적 모델의 하나 이상의 컴포넌트에 대한 파라미터 값을 추정하는 단계를 포함할 수 있다. 일부 실시예에서, 예측기(예를 들어, 도 2b의 예에서 예측기(252))의 파라미터는 동작 308에서 추정될 수 있다. 추가적으로, 일부 실시예에서, 하나 이상의 태스크 임베딩(예를 들어, 태스크 임베딩들(254) 중의 하나 이상)의 파라미터가 동작 308에서 추정될 수 있다.

[0104] 일부 실시예에서, 동작 306의 일부로서 추정된 파라미터 값을 동작 304에서 액세스된 라벨링된 훈련 데이터에 기초하여 감독형 학습을 이용하여 추정될 수 있다. 일부 실시예에서, 특정한 데이터가 통계적 모델에 대한 입력으로서 제공될 수 있고 통계적 모델의 출력은 그 특정한 데이터에 대한 라벨과 비교될 수 있다. 차례로, 통계적

모델의 파라미터의 하나 이상의 값은 통계적 모델의 출력과 통계적 모델에 대한 라벨과 제공된 특정한 데이터에 대한 라벨 사이의 차이에 기초하여 (예를 들어, 스토캐스틱 경사 하강 또는 임의의 기타 적절한 훈련 알고리즘을 이용하여) 업데이트될 수 있고, 그 차이는, 통계적 모델의 출력이, 그 현재 세트의 파라미터 값들과 함께 동작될 때 제공된 라벨을 얼마나 잘 재현하는지의 척도를 제공한다.

[0105] 일부 실시예에서, 제2 훈련 스테이지 동안 이용되는 손실(또는 비용) 함수는, 다중-모달 통계적 모델의 예측기 컴포넌트가 훈련되고 있는 태스크의 유형에 따라 선택될 수 있다. 예를 들어, 태스크가 다중-라벨 배타적 분류를 포함한다면, 교차-엔트로피 손실이 이용될 수 있다. 또 다른 예로서, 태스크가 연속 분포의 예측을 포함한다면, Kullback-Leibler 밸런이 손실 함수에서 이용될 수 있다.

[0106] 일부 실시예에서, 제2 스테이지의 수행 동안, 제1 훈련 스테이지 동안 추정된 파라미터 값은 고정될 수 있다. 예를 들어, 제1 훈련 스테이지 동안에 합동-모달리티 표현 및 모달리티 임베딩에 대한 파라미터 값이 추정된 후, 이를 값은 제2 훈련 스테이지 동안에 고정된 상태로 유지될 수 있지만, 예측기 및 태스크 임베딩에 대한 파라미터 값은 제2 훈련 스테이지 동안에 추정되고 있다.

[0107] 동작 308이 완료되어 동작 305를 완료한 후, 훈련된 다중-모달 통계적 모델은 후속 이용을 위해 동작 310에서 저장될 수 있다. 훈련된 다중-모달 통계적 모델을 저장하는 것은, 다중-모달 통계적 모델의 하나 이상의 컴포넌트에 대한 파라미터 값을 저장하는 것을 포함한다. 일부 실시예에서, 훈련된 다중-모달 통계적 모델을 저장하는 것은 다음과 같은 컴포넌트들 중 하나 이상에 대해 동작 305 동안 추정된 파라미터 값들을 저장하는 것을 포함한다 : 합동-모달리티 표현, 제1 모달리티 임베딩, 제2 모달리티 임베딩, 예측기, 제1 태스크 임베딩, 및 제2 태스크 임베딩. 본 명세서에서 설명된 기술의 양태가 이 점에서 제한되지 않기 때문에, 파라미터 값은 임의의 적절한 포맷으로 저장될 수 있다. 파라미터 값은 하나 또는 복수의 컴퓨터-관독가능한 저장 매체(예를 들어, 하나 이상의 메모리)를 이용하여 저장될 수 있다.

[0108] 프로세스(300)는 예시적이며 변형이 있다는 것을 이해해야 한다. 예를 들어, 프로세스(300)가 2개의 모달리티를 갖는 입력을 수신하도록 구성된 다중-모달 통계적 모델을 훈련시키는 것을 참조하여 설명되었지만, 프로세스(300)는 2개보다 많은 모달리티(예를 들어, 3, 4, 5, 6, 7, 8, 9, 10개 등의 모달리티)로부터의 입력을 수신하도록 구성된 다중-모달 통계적 모델을 훈련시키도록 수정될 수 있다. 이러한 일부 실시예에서, 복수의 모달리티 각각에 대한 합동-모달리티 표현 및 모달리티 임베딩은 자체 감독형 학습 스테이지 동안 학습될 것이다(동작 306). 복수의 모달리티 각각에 대한 예측기와 태스크 임베딩은 감독형 학습 스테이지 동안 학습된다(동작 308).

[0109] 전술된 바와 같이, 일부 실시예에서, 모달리티들 각각에 대한 인코더 및 디코더는 프로세스(300)의 실행에 앞서 학습될 수 있다. 그러나, 일부 실시예에서, 인코더(들) 및/또는 디코더(들) 중 하나 이상은, 그들의 파라미터 값이 처음으로 추정되거나 및/또는 프로세스(300) 동안 업데이트되도록 프로세스(300) 동안 학습될 수 있다.

[0110] 다중-모달 통계적 모델을 훈련시키기 위한 기술의 추가적인 양태는 자체 감독형 및 감독형 훈련 스테이지에 대한 이하의 논의에서 이해될 수 있다.

[0111] 자체 감독형 훈련 국면

[0112]  $x_i \in X_i$  이 모달리티  $i$ 에 대한 입력 데이터 포인트이고,  $t_i \in T_i$  이 다음과 같도록  $x_i$ 의 압축된 표현이라고 하자 :

[0113]  $t_i = \psi_i(x_i)$ ,

[0114] 여기서,  $\psi_i$  는  $i$ 번째 모달리티에 대한 인코더를 나타내는 인코딩 함수이다. (본 명세서에서는 때때로 지식 베이스라고도 하는) 합동-모달리티 표현이  $n \times m$  행렬  $M$ 이라고 하자, 여기서  $n$ 은 합동-모달리티 표현에서의 엔트리의 수를 나타내고  $m$ 은 각각의 엔트리의 차원성을 나타낸다. 합동-모달리티 표현은, 모달리티 임베딩  $E_i$ (자체 감독형 훈련 국면 동안에 학습되는  $m \times d_i$  행렬)를 이용함으로써  $i$ 번째 모달리티의 표현 공간에 선형으로 투사될 수 있다 :

[0115]  $\tilde{M}_i = M E_i$ .

[0116] 그 다음, 표현  $t_i$ 와 투사된 합동-모달리티 표현  $\tilde{M}_i$ 의 행들 사이의 코사인 유사성은, 합동-모달리티 표현의 각각의 엔트리(예를 들어, 메모리 행렬의 각각의 행)에 관한 유사성 점수를 제공하며, 이것은, 다음에 따라

$p(m|t)$  를 근사화하는, 소프트맥스 함수를 이용하여 확률로 변환될 수 있다 :

$$s_i = \tilde{M}_i t_i$$

$$[0117] p(m_k|t_i) = \frac{e^{s_{ik}/\tau}}{\sum_k e^{s_{ik}/\tau}}$$

[0118] 여기서,  $\tau$  는 온도 변수이고, 분포의 선명성/엔트로피(sharpness/entropy)를 감안한다. 그 다음, 투사된 합동-모달리티 표현 행렬 엔트리들  $\tilde{t}_i$  의 각종 평균이  $i$  번째 모달리티 디코더  $\Phi_i$ 에 입력으로서 제공된다 :

$$\tilde{t}_i = \sum_k p(m_k|t_i) m_k$$

$$[0119] \hat{x}_i = \Phi_i(\tilde{t}_i)$$

[0120] 네트워크 파라미터들 중 적어도 일부(예를 들어, 인코더, 디코더, 합동-모달리티 표현 및 모달리티 임베딩의 파라미터 값들의 일부 또는 전부)에 관한 재구성 손실의 경사가 역 전파되고 파라미터들은 스토캐스틱 경사 하강 알고리즘을 통해 업데이트된다 :

$$[0121] \theta_{j,t+1} = \theta_{j,t} + \lambda \nabla_{\theta_{j,t}} \ell(\hat{x}_i, x_i) + \mu_t$$

[0122] 여기서,  $\theta_{j,t}$  는 시간  $t$ 에서의  $j$  번째 파라미터이다;  $\lambda$  및  $\mu$ 는 각각 학습률 및 모멘텀 파라미터이며,  $\ell$ 은 손실 함수이다. 본 명세서에서 설명된 기술의 양태는 이 점에서 제한되지 않으므로, 손실 함수는 교차-엔트로피, Kullback-Liebler 발산, L1 거리, L2 거리(Euclidean 거리), 및/또는 임의의 기타 적절한 손실 함수일 수 있다.

[0123] 감독형 훈련 스테이지

[0124] 우리는 태스크를  $y \in Y$  로 표기되는 예측 라벨 또는 값으로서 정의한다. 데이터 쌍  $(x_i, y_j)$  의 존재하에, 우리는 상기 도시된 방정식에 도시된 바와 같이 표현  $t_i \in T_i$  를 생성하기 위해 자체 감독형 학습 스테이지 동안에 훈련 된 합동-모달리티 표현 및  $x_i \in X_i$  에 대한 인코더  $\Psi_i(x_i)$  를 이용한다. 그 다음, 다음에 따라 피처 표현  $\tilde{t}_i$  와 태스크 임베딩  $U_j$  사이에서 Hadamard 곱이 수행된다:

$$[0125] \tilde{u}_{ij} = u_j \odot \tilde{t}_i$$

[0126] 마지막으로, 정방향 패스(forward pass)를 위해, 우리는 투사된 표현을 태스크 예측기에 제공한다

$$[0127] \Pi_j \hat{y}_j = \Pi_j(\tilde{u}_{ij})$$

[0128] 태스크 유형에 대해 적절한 손실 함수가 선택된다. 예를 들어, 태스크가 다중-라벨 베타적 분류라면, 교차-엔트로피 손실이 이용될 수 있다. 또 다른 예로서, 태스크가 연속 분포의 예측이라면, 손실 함수로서 Kullback-Leibler 발산 등의 정보 이론적 측정이 이용될 수 있다. 손실 함수의 선택에 관계없이, 태스크 예측기  $\Pi_j$  및 태스크 임베딩  $U_j$  의 파라미터에 관한 손실의 경사는 위에서 도시된 스토캐스틱 경사 하강 방정식에서 설명된 바와 같이 계산되고 역 전파될 수 있다.

[0129] 예측을 위한 다중-모달 통계적 모델의 이용

[0130] 도 4는, 본 명세서에서 설명된 기술의 일부 실시예에 따른, 예측 태스크를 위한 다중-모달 통계적 모델을 이용하기 위한 예시적인 프로세스(400)의 플로차트이다. 프로세스(400)는 임의의 적절한 컴퓨팅 디바이스(들)에 의해 수행될 수 있다. 본 명세서에서 설명된 기술의 양태는 이 점에서 제한되지 않기 때문에, 예를 들어, 프로세스(400)는, 하나 이상의 그래픽 처리 유닛(GPU), 클라우드 컴퓨팅 서비스에 의해 제공되는 하나 이상의 컴퓨팅 디바이스(들), 및/또는 임의의 기타 적절한 컴퓨팅 디바이스(들)에 의해 수행될 수 있다.

- [0131] 이 예에서, 프로세스(400)의 시작 전에, 적어도 2개의 상이한 모달리티에 대한 입력을 수신하도록 구성된 다중-모달 통계적 모델이 훈련되었고 그 파라미터가 저장되었다고 가정한다. 예를 들어, 프로세스(400)의 시작에 앞서, 다중-모달 통계적 모델이 본 명세서에서 설명된 2-스테이지 훈련 프로세스(300)를 이용하여 훈련되었을 수 있다.
- [0132] 프로세스(400)는, 이전에 훈련된 다중-모달 통계적 모델을 명시하는 정보가 액세스되는 동작 402에서 시작한다. 본 명세서에서 설명된 기술의 양태는 이 점에서 제한되지 않기 때문에, 다중-모달 통계적 모델을 명시하는 정보는 임의의 적절한 포맷일 수 있고, 로컬 스토리지로부터 액세스되거나, 네트워크를 통해 원격 스토리지로부터 액세스되거나, 임의의 기타 적절한 소스(들)로부터 액세스될 수 있다. 정보는 다중-모달 통계적 모델의 파라미터에 대한 값을 포함할 수 있다. 다중-모달 통계적 모델은 파라미터를 갖는 컴포넌트를 포함할 수 있고, 모델을 명시하는 정보는 이를 컴포넌트 중 하나 이상의 각각에 대한 파라미터의 파라미터 값을 포함할 수 있다. 예를 들어, 다중-모달 통계적 모델은, 합동-모달리티 표현, 예측기, 및 복수의 모달리티 각각에 대해, 각각의 인코더, 각각의 모달리티 임베딩, 및 각각의 태스크 임베딩을 포함할 수 있으며, 동작 402에서 액세스된 정보는 이를 컴포넌트에 대한 값을 포함할 수 있다.
- [0133] 도 4를 참조하여 설명된 실시예에서, (그 파라미터가 액세스되는) 다중-모달 통계적 모델이, 2개의 모달리티 – 제1 모달리티 및 제2 모달리티로부터 입력을 수신하도록 구성된다고 가정한다. 그러나, 본 명세서에서 설명된 기술의 양태는 이 점에서 제한되지 않기 때문에, 다른 실시예에서, 다중-모달 통계적 모델은 임의의 적절한 수의 모달리티(예를 들어, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 등)으로부터 입력을 수신하도록 구성될 수 있다는 것을 이해해야 한다.
- [0134] 그 다음, 프로세스(400)는, 제1 데이터 모달리티(예를 들어, 단백질 서열 데이터)에 대한 입력 데이터가 획득되는 동작 404로 진행한다. 일부 실시예에서, 입력 데이터는, 제1 모달리티에 대한 인코더에 제공하기에 적합한 표현으로 변환되거나 기타의 방식으로 전처리될 수 있다. 예를 들어, 범주형 데이터는 제1 모달리티에 대한 인코더에 제공되기 전에 원-핫-인코딩될 수 있다. 또 다른 예로서, 이미지 데이터는 제1 모달리티에 대한 인코더에 제공되기 전에 크기조정될 수 있다. 그러나, 다른 실시예에서, 어떠한 변환 및/또는 전처리도 요구되거나 수행되지 않을 수 있다.
- [0135] 그 다음, 프로세스(400)는 동작 406으로 진행하고, 여기서 입력 데이터는, 출력으로서 제1 피쳐 벡터를 생성하는 제1 인코더에 대한 입력으로서 제공된다. 예를 들어, 도 2b를 참조하여 도시된 바와 같이, 모달리티 "A"에 대한 입력(202)은 모달리티 "A"에 대한 인코더(204)에 대한 입력으로서 제공되고, 인코더(204)는 제1 피쳐 벡터(예를 들어, 출력으로서 피쳐 표현(206))를 생성한다.
- [0136] 그 다음, 프로세스(400)는, 동작 406에서 (제1 인코더의 출력에서) 생성된 제1 피쳐 벡터가 합동-모달리티 표현 및 제1 모달리티 임베딩과 함께 이용되어 제2 피쳐 벡터를 생성하는 동작 408로 진행한다. 예를 들어, 도 2b를 참조하여 도시된 바와 같이, 제1 피쳐 벡터(예를 들어, 피쳐 표현(206))는 제2 피쳐 벡터(예를 들어, 피쳐 표현(208))를 식별(예를 들어, 생성 또는 선택)하기 위해 모달리티 임베딩들(232) 및 지식 베이스(230) 중 하나와 함께 이용될 수 있다.
- [0137] 제2 피쳐 벡터는 본 명세서에 설명된 임의의 방식으로 식별될 수 있다. 예를 들어, 일부 실시예에서, 제2 피쳐 벡터를 식별하는 것은 다음을 포함할 수 있다 : (1) 복수의 투사된 벡터를 획득하기 위해 합동-모달리티 표현(예를 들어, 지식 베이스(230))을 제1 모달리티의 공간에 투사하는 단계; (2) 복수의 투사된 벡터 각각과 제1 피쳐 벡터(예를 들어, 피쳐 표현(206)) 사이의 거리(예를 들어, 코사인 거리 및/또는 임의의 기타 적절한 유형의 거리 측정치)를 계산하고, 이를 거리를 이용하여 투사된 벡터에 대한 가중치를 (예를 들어, 소프트-맥스 가중을 이용하여) 계산하는 단계; 및 (3) 계산된 가중치들에 의해 가중된 투사된 벡터들의 가중 합으로서 제2 피쳐 벡터를 생성하는 단계. 예를 들어, 합동-모달리티 표현은, ( $N \times m$  행렬로 표현되거나 및/또는 저장될 수 있는)  $N$ 개의  $m$ 차원 벡터를 포함할 수 있고,  $m \times d$ 로서 표현될 수 있는 제1 모달리티 투사를 이용하여 합동-모달리티 표현을 제1 모달리티에 투사하는 것은, ( $N \times d$  행렬로서 표현될 수 있는)  $N$ 개의  $d$ 차원 벡터를 생성할 수 있다. 제1 인코더에 의해 출력된 제1 피쳐 벡터(예를 들어, 도 2a에 도시된 피쳐 표현(206))과  $N$ 개의  $d$ 차원 벡터 각각 사이의 거리가 계산되고  $N$ 개의  $d$ 차원 벡터 각각에 대한 가중치를 획득하는데 이용될 수 있다. 그 다음, 제2 피쳐 벡터(예를 들어, 피쳐 표현(208))는  $N$ 개의  $d$ 차원 벡터의 가중 합으로서 계산될 수 있고, 벡터들은 계산된 가중치들에 의해 가중된다. 다른 실시예에서, 제2 피쳐 벡터는,  $N$ 개의  $d$ 차원의 투사된 벡터 중에서, 투사된 합동-모달리티 표현의 복수의  $d$ 차원 벡터의 가중 평균이 아니라, 적절하게 선택된 거리 측정치(예를 들어, 코사인 거리)에 따라 제1 인코더에 의해 생성된 제1 피쳐 벡터에 가장 가까운 벡터를 선택함으로써 식별될 수 있다.

- [0138] 그 다음, 프로세스(400)는 동작 410으로 진행하고, 여기서 제2 피처 벡터는 예측기를 이용하여 예측 태스크에 대한 예측을 생성하고 제1 모달리티에 대한 태스크 임베딩을 생성하는데 이용된다(이들 양쪽 모두는 다중-모달 통계적 모델의 컴포넌트임). 이것은 임의의 적절한 방식으로 이루어질 수 있다. 예를 들어, 제1 모달리티에 대한 태스크 임베딩은 제2 피처 벡터의 것과 동일한 차원성을 가질 수 있다. 이 예에서, 태스크 임베딩의 가중치는, 예측기에 대한 입력을 생성하기 위해 (예를 들어, Hadamard 곱의 의미에서) 제2 피처 벡터의 값들을 포인트 별로 곱하는데 이용될 수 있다. 차례로, 예측기는 이 입력에 기초하여 태스크에 대한 예측을 출력할 수 있다. 예를 들어, 도 2b에 도시된 바와 같이, 제2 피처 벡터(예를 들어, 표현(208))는 태스크 임베딩들(254) 중의 제1 태스크 임베딩에 의해 포인트별로 수정(예를 들어, 곱해짐)될 수 있고 예측 태스크(256)에 대한 출력을 생성하기 위해 예측기(252)에 입력으로서 제공될 수 있다.
- [0139] 프로세스(400)의 상기 설명으로부터 이해할 수 있는 바와 같이, 다중-모달 통계적 모델은 단일 모달리티로부터의 입력만을 이용하여 태스크에 대한 예측을 생성하는데 이용될 수 있다. 이것은, 상이한 시간들에서 복수의 상이한 모달리티로부터의 입력들이 이용가능할 때, 이들은 이용가능하게 될 때, 다중-모달 통계적 모델에 대한 입력으로 비동기식으로 제공될 수 있다는 것을 의미한다.
- [0140] 일부 실시예에서, 다중-모달 통계적 모델은 동기식으로 동작될 수 있고, 2개의 모달리티로부터의 쌍을 이루는 입력 또는 2개보다 많은 모달리티로부터의 링크된 입력을 처리하는데 이용될 수 있다. 예를 들어, 제1 모달리티에 대한 제1 입력(예를 들어, 입력(202))은 제1 피처 벡터(예를 들어, 피처 표현(206))를 생성하기 위해 제1 모달리티에 대한 인코더(예를 들어, 인코더(204))에 대한 입력으로서 제공될 수 있고, 제1 피처 벡터는, 합동-모달리티 표현(예를 들어, 지식 베이스(230)) 및 (예를 들어, 모달리티 표현들(232) 중의) 제1 모달리티 표현과 함께, 제2 피처 벡터(예를 들어, 피처 표현(208))를 식별(예를 들어, 생성 또는 선택)하는데 이용될 수 있다. 이 예에서, 제1 모달리티에 대한 제1 입력(예를 들어, 입력(202))은 제2 모달리티에 대한 제1 입력(예를 들어, 입력(212))은 제3 피처 벡터(예를 들어, 피처 표현(216))를 식별(예를 들어, 생성 또는 선택)하기 위해 제2 모달리티에 대한 인코더(예를 들어, 인코더(214))에 대한 입력으로서 제공될 수 있고, 제1 피처 벡터는, 합동-모달리티 표현(예를 들어, 지식 베이스(230)) 및 (예를 들어, 모달리티 표현들(232) 중의) 제2 모달리티 표현과 함께, 제4 피처 벡터(예를 들어, 피처 표현(218))를 생성하는데 이용될 수 있다. 차례로, 제2 및 제4 피처 벡터는, 제1 및 제2 모달리티에 대한 태스크 임베딩에 의해 수정될 수 있고, 그 결과는 (예를 들어, 좌표별 추가(260)에 의해) 결합되어 예측기(예를 들어, 예측기(252))에 대한 입력으로 제공되어 태스크(256)에 대한 예측을 제공할 수 있다.
- [0141] 예: 단백질 구조 예측
- [0142] 이 섹션에서는, 본 명세서에서 설명된 상이한 데이터 모달리티들에 대한 딥 러닝 통계적 모델들을 단일화하기 위한 기술이 단백질 구조 예측의 예시적인 문제에 대해 예시된다. 문자 생물학에서 예측 모델을 구축하는 종래 기술은 종종 부족하여, 그 결과적인 모델은 원하는 성능 특성(예를 들어, 정밀도)이 부족하다.
- [0143] 이용 가능한 다양한 유형의 생물학적 데이터에 대한 예측 모델링을 수용하기 위한 공통 프레임워크를 구축하는 것은 다음과 같은 것을 포함하는 다양한 이유로 매우 어렵다 :
- [0144] · 소스 이종성(Source Heterogeneity): 잠재적으로 수천개의 상이한 문자 엔티티들이 프로브될 수 있으며 우리가 취득하는 데이터는 다양한 형태 또는 모달리티로 나온다.
  - [0145] · 높은 차원성(High Dimensionality) : 관찰된 데이터는 입력 공간의 모든 가능한 구성을 드물게 샘플링한다. 따라서, 대부분의 경우 이용 가능한 데이터가 희소하고 불충분하다.
- [0146] · 실험 노이즈 : 생물학적 데이터 취득에서는, 종종 노이즈가 많고 실험적 편견과 특이성 문제를 겪는다.
  - [0147] · 정합되지 않는 모달리티 및 불완전성 : 실험 및 관찰은 한 번에 한 쌍의 모달리티만으로 제한되므로, 데이터가 매우 불완전하다.
- [0148] 이러한 까다로운 모델링 정황에서 고품질의 예측 모델을 구축하는 종래의 접근법은, 깊은 도메인-수준의 전문성과 지식을 표현하는 강력한 선행물에 의존한다. 그러나, 이러한 선행물을 명시하는 우리의 능력은 이용 가능한 도메인-수준의 지식의 양에 의해 제한된다. 예를 들어, 광범위한 도메인 지식 없이, BLAST 쿼리(가장 가까운 알려진 시퀀스 찾기)를 수행하고 최상위 히트로부터 기능적 할당을 전송함으로써 새로운 발견된 종의 단백질 서열이 기능적으로 주석부기될 수 있다. 그러나, 이 접근법은, 특히 관심대상 단백질이 어떤 생물학적 프로세스에

관여하는지를 식별할 때 매우 좋지 않고 오도하는 것으로 보고되고 있다. 더 양호하게 수행하는 모델은, 단백질, 아미노산 모티프, 생물학적 프로세스에 대한 이들의 참여 등에 관한 수년간의 누적된 도메인 지식을 요구한다.

[0149] 본 명세서에 설명된 기술은, (각각의 데이터 포인트가 복수의 상이한 모달리티 각각으로부터의 기여도를 포함하면서) 훈련 데이터가 완전히 정합되는 데이터 포인트를 포함할 것을 요구하지 않고 복수의 모달리티를 수용함으로써 전술된 과제를 해결한다. 본 명세서에서 설명된 합동-모달리티 표현은, 교차-모달리티 피처 추출을 위한 데이터-주도형 선행물을 제공하여, 개개의 모델을 정규화하고 추가 압축을 완화한다. 추가 압축의 모든 비트는 라벨링된 데이터의 2배를 갖는 것과 균등하다.

[0150] 본 명세서에서 설명된 기술이 단백질 기능 예측 태스크에 관해 아래에서 예시된다. 먼저, 우리는 554,452개의 단백질을 포함하는 Swiss-Prot 데이터베이스를 다운로드하여, 다음을 포함하는 6개의 상이한 데이터 모달리티를 선택했다 : (1) 단백질 서열; (2) pfam 도메인; (3) 생물학적 프로세스 온톨로지; (4) 분자 기능 온톨로지; (5) 세포 성분 온톨로지; 및 (6) 종의 분류군. 기능적 주석(온톨로지)은 매우 불완전하고 아마도 노이즈가 있을 수 있다. 두 번째 CAFA2(Critical Assessment of Functional Annotation) 컨소시엄에 대한 테스트 세트로서 정의된 단백질은 결과 평가를 용이화하기 위해 배제되었다.

#### [0151] 구현 상세사항

[0152] 기능적 온톨로지 예측이 우리의 태스크이지만, 우리는 이를 온톨로지를 개개의 모달리티로서 취급했다. 본 명세서에서 설명된 기술을 기능적 온톨로지 예측 태스크에 적용하기 위해, 우리는, 인코더, 디코더, 합동-모달리티 표현, 모달리티 임베딩, 및 태스크 임베딩의 양태들을 명시할 필요가 있다.

#### [0153] 인코더

[0154] 이 예시적인 예에서, 단백질 서열 입력을 위한 인코더는, 크기가 20인 10개의 필터가 있는 1D 콘볼루션을 각각 포함하는 4개의 콘볼루션 블록을 포함하고, 후속해서, 계층 정규화, 스트라이드(stride) 3과 함께 크기 3의 1차원 맥스-풀링, 및 ReLU(rectified linear unit) 활성화가 이어졌다. 4개의 콘볼루션 블록 후에, 인코더는 크기 11의 커널 10개 및 크기 1에 대한 적응형 1d 맥스-풀링이 있는 또 다른 콘볼루션 계층을 포함한다. 그 결과, 단백질 서열 인코더는  $10 \times 1024$  원-핫-인코딩된 단백질 서열 입력을 취하고(서열이 1024보다 짧은 경우, 입력은 모두 0으로 패딩됨)  $10 \times 1$  잠재적 표현을 반환한다.

[0155] 우리는 범주형 데이터 소스의 인코더로서 임베딩 사전(embedding dictionary)을 이용했다. 임베딩 사전을 인덱싱하는 것은, 바이어스 항(bias term) 없이 선형 계층에 원-핫-인코딩된 입력 데이터를 포워딩하는 것과 균등하지만, 입력이 매우 희소하므로 계산적으로 훨씬 더 효율적이다. 첫 번째 엔트리는 항상 미지의 범주 또는 패딩 인덱스 용으로 예약되어 있으므로 임베딩 사전의 크기는 각각의 모달리티의 범주의 개수보다 하나 더 많다. 실험에 이용된 실제 크기는, 생물학적 프로세스, 분자 기능, 세포 성분, 분류학적 패밀리, 및 pfam 도메인에 대해 각각 24937, 9572, 3185, 1779 및 11679이다. 임베딩의 차원은 10으로 선택된다.

#### [0156] 디코더

[0157] 단백질 서열 디코더는 6개의 연속된 계층의 디콘볼루션 블록들을 포함한다. 각각의 블록은, 필터 수가 128개, 필터 크기가 6, 스트라이드가 3, 양쪽 말단이 1로 패딩된 디콘볼루션 연산을 포함하고, 후속해서, 계층 정규화 및 기울기 0.1의 누설 ReLU 활성화가 이어진다.

[0158] 범주형 모달리티의 디코더는, 크기  $10 \times N$ 인 완전히 접속된 선형 계층으로 선택되고, 이것은, 합동-모달리티 표현(예를 들어, 지식 베이스)으로부터 반환된 표현을 취하여 모든 클래스에 대한 시그모이드 활성화된 점수(sigmoid-activated score)를 반환한다( $N$ 은 각각의 모달리티의 클래스의 수이다).

#### [0159] 합동-모달리티 표현 및 모달리티 투사

[0160] 합동-모달리티 표현은 차원 64의 512개 벡터를 포함한다. 이 예에서, 이를 벡터는 크기  $512 \times 64$ 인 행렬에 저장될 수 있다. 행들은, 각각의 업데이트 후에 L2 정규화된다. 이 예에서는 6개의 모달리티가 있기 때문에, 6개의 모달리티 임베딩이 있으며, 그 각각은 크기  $64 \times 10$ 인 행렬을 이용하여 표현된다. 각각의 모달리티 임베딩은 합동-모달리티 표현을 각각의 모달리티의 표현 공간으로 투사한다.

#### [0161] 손실 함수

[0162] 서열 재구성을 위해, 우리는, 서열의 모든 아미노산 잔기(amino acid residue)에 대해 20개보다 많은 가능한 아

미노산에 관한 확률 분포에 대해 계산된 교차 엔트로피 손실을 이용했다. 우리는 패딩된 영역을 배제했다. 3개의 온톨로지 모달리티 및 pfam 도메인 모달리티에 대해, 우리는, 네거티브 샘플링 절차와 마진 값 1을 이용하여 맥스-마진 손실을 이용했다. 분류학적 패밀리 모달리티에 대해, 우리는 교차 엔트로피를 이용했다.

### [0163] 훈련

[0164] 우리는, 학습률이  $10^{-3}$ 이고 뒷음 크기(batch size)가 25인 "Adam"이라는 SGD 최적화 프로그램의 한 변형을 이용했다. 우리는 2개의 상이한 시나리오 양쪽 모두를 테스트했다 : (1) 쌍을 이룬 데이터를 이용한 동기식 훈련; 및 (2) 쌍을 이루지 않은 데이터를 이용한 비동기식 훈련.

[0165] 쌍을 이룬 데이터로 훈련할 때, 다른 모든 파라미터와 유사하게, 모든 모달리티에 걸쳐, 모든 재구성 손실로부터 기인한 경사들의 합과 관련하여 합동-모달리티 표현 가중치가 업데이트된다.

[0166] 비동기식으로 훈련할 때, 각각의 모달리티에 대한 파라미터는 합동-모달리티 표현을 질의함으로써 하나씩 훈련된다. 합동-모달리티 표현 가중치는, 모달리티가 그 자신의 재구성 목표와 함께 훈련될 때마다 업데이트된다. 우리는 모든 모달리티를 3번씩 수행했고 매번 완전히 수렴될 때까지 훈련했다. 우리는, 모달리티가 훈련될 때마다 합동-모달리티 표현 파라미터에 관한 학습률을 감소시켰다.

### [0167] 결과

[0168] 도 5에 도시된 바와 같이, 초기 실험은, 단백질의 기능적 주석부기에 대해 전술된 다중-모달 통계적 모델이, 광범위한 피처 공학을 모두 요구하는 경쟁하는 종래의 다른 접근법의 모델들보다 훨씬 더 잘 수행한다는 것을 나타낸다는 것을 보여준다. 도 5에 도시된 바와 같이, 전술된 다중-모달 통계적 모델에 대한 평균 AUROC(area under receiver operating characteristic curve)는 경쟁하는 종래의 접근법보다 높다. 도 5에 도시된 경쟁하는 접근법의 성능은, 참조에 의해 그 전체내용이 본 명세서에 포함되는, 2016년 9월 7일 Genome Biology, volume 17, page 184에 게시된, "An expanded evaluation of protein function prediction methods shows an improvement in accuracy"라는 제목의 논문에서 추가로 논의된다.

### [0169] 이론적 토대

[0170] 본 명세서에서, 설명된 다중-모달 통계적 모델의 추가 양태는 이하의 논의로부터 이해될 수 있다.

### [0171] 관련 정보 추출

[0172]  $X$ 는 고정된 확률 측정치  $\rho(x)$ 를 갖는 신호(메시지) 공간을 나타내고,  $T$ 는 그 양자화된 코드북 또는 압축된 표현을 나타낸다고 하자.

[0173] 각각의  $x \in X$ 에 대해, 우리는, 조건부 확률 밀도 함수(pdf; probability density function)  $p(t|x)$ 를 특징으로 하는 코드북  $t \in T$  내의 대표 또는 코드워드에 대한 아마도 스토캐스틱 맵핑을 추구한다. 이 맵핑은, 각각의 블록이 확률  $p(t|x)$ 를 갖는 코드북 요소  $t \in T$  와 연관되는  $X$ 의 소프트 파티셔닝(soft partitioning)을 유도한다. 코드워드  $t \in T$ 의 총 확률 다음과 같이 주어진다:

$$[0174] p(t) = \sum_x p(t|x)p(x).$$

[0175] 동일한 코드워드에 맵핑되는  $X$ 의 요소들의 평균 볼륨은  $2^{H(X|T)}$ 이고, 여기서,

$$[0176] H(X|T) = \sum_t p(t)H(X|T=t) = -\sum_t p(t)\sum_x p(x|t)\log(p(x|t))$$

[0177] 양자화의 품질은 혼동없이 코드북 내의 요소를 명시하는데 필요한 "레이트" 또는 "메시지 당 평균 비트"에 의해 결정된다.  $X$ 의 요소 당 이 숫자는 상호정보에 의해 아래와 같이 제한된다

$$[0178] I(X; T) = \sum_{x \in X} \sum_{t \in T} p(x, t) \log\left(\frac{p(x|t)}{p(t)}\right)$$

[0179] 이 방정식은, 평균 파티션의 볼륨에 대한 X의 볼륨의 비율에 의해 주어지는 X의 파티셔닝의 평균 카디널리티(average cardinality)로서 간주될 수 있다, 즉,

$$K = \frac{2^{H(X)}}{2^H(X|T)} = 2^{I(X;T)}$$

[0180] 정보 병목

[0182] 궁극적으로, 임의의 예측 태스크의 경우, 우리는, 입력 공간 X로부터 예측(라벨) 공간 Y에 대한 관련 정보만 유지하는 표현 공간 T로의 맵핑  $p(t|x)$ 을 배우기를 원한다. 즉, 우리는, 맵핑  $p(t|x)$ 과 관련하여 다음과 같은 함수를 최소화함으로써 캡처될 수 있는 T와 Y 사이의 상호정보를 최대화하면서 X와 T 사이의 상호정보를 최소화하기를 원한다:

$$L[p(t|x)] = I(X;T) - \beta I(Y;T)$$

[0184] 여기서,  $\beta$ 는 절충 파라미터(tradeoff parameter)이다.

[0185] 입력 압축 한계

[0186] 최상의 예측 성능을 위해, 우리의 목표는 데이터 처리 불평등으로 인해 상한  $I(T;Y) \leq I(X;Y)$ 로 제약되는  $I(T;Y)$ 의 최대화이다. 우리가 X와 Y에서 무제한의 데이터량을 갖는다면, 우리는 합동 분포  $p(x,y)$ 를 임의로 근사할 수 있으므로, X의 컴팩트 표현이 필요하지 않을 것이다. 그러나, 데이터의 양은 종종 제한되기 때문에,  $p(x|y)$ 는 충분히 양호하게 추정될 수 없다. 따라서, 우리는 입력을 압축함으로써 모델을 정규화해야 한다 –  $I(X;T)$ 를 최소화함으로써 복잡성을 감소시킨다.

[0187]  $\hat{I}$ 는 제한된 샘플들의 상호정보에 대한 경험적 추정치를 나타낸다고 하자. 일반화 한계는 다음과 같은 것으로 나타났다 :

$$I(T;Y) \leq \hat{I}(T;Y) + O\left(\frac{|T||Y|}{\sqrt{n}}\right) \quad \text{및}$$

$$I(T;X) \leq \hat{I}(T;X) + O\left(\frac{|T|}{\sqrt{n}}\right)$$

[0189] 특히, 상한은 표현  $K = |T|2^{I(T;X)}$ 의 카디널리티에 의존한다. 즉, 추가 압축의 추가적인 단일 비트는 동일한 일반화 캡에 대해 데이터 크기를 2배로 늘리는 것과 균등하다.

[0191] 다중-모달 예측을 위한 압축

[0192] 간단한 교차-모달리티 예측 설정을 고려하자, 여기서, 모달리티  $X_1$ 과  $X_2$ 는 각각  $X_2$ 와  $X_1$ 을 예측해야 하는  $T_1$ 과  $T_2$  표현으로 압축된다. 도 6a에 도시된 바와 같이, 관찰된 변수  $X_1$ 과  $X_2$ 는,  $X_1$ 과  $X_2$ 의 압축된 표현인 잠재적 랜덤 변수  $T_1$ 과  $T_2$ 로 표현된다. 제1 및 제2 모달리티에 대한 잠재적 랜덤 변수  $T_1$ 과  $T_2$ 는 각각 제1 및 제2 모달리티에 대한 인코더의 출력으로서 정의될 수 있다. 도 6b에 도시된 바와 같이, 잠재적 랜덤 변수  $T_1$ 과  $T_2$ 는 변수  $X_1$ 과  $X_2$ 를 예측하는데 이용될 수 있다. 제1 및 제2 모달리티에 대한 디코더들은, 각각, 잠재적 표현  $T_1$ 과  $T_2$ 로부터 변수  $X_1$ 과  $X_2$ 를 예측하는데 이용될 수 있다.

[0193] 이 상황에서, 최소화할 Lagrangian은 다음과 같이 주어진다 :

$$L = I(T_1;X_1) + I(T_2;X_2) - \gamma I(T_1;T_2)$$

[0194] 따라서, 우리가 압축하는 동안, 우리는 또한, 압축된 표현  $T_1$ 과  $T_2$ 가 가능한 한 서로에 대해 정보를 제공하도록

확실히 하기를 원한다. 이 방정식은, 우리가,  $X_1$ ,  $T_1$ 과  $X_2$ ,  $T_2$  사이의 상호정보를 최소화하면서  $T_1$ 과  $T_2$  사이의 상호정보(상관)를 최대화함으로써  $X_1$ 과  $X_2$ 를 최대로 압축해야 함을 나타낸다. 본 명세서에서 설명된 프레임워크에서,  $T_1$ 과  $T_2$  사이의 상호정보를 최대화하는 것은, 각각의 인코딩된 입력이 코드북 내의 코드워드들의 단일의 또는 가중 평균 – 합동-모달리티 표현(예를 들어, 지식 베이스(230))과 정합되도록 강제함으로써 달성될 수 있다; 정합된 엔트리는 자체 감독형 훈련 단계 동안 디코더에 입력으로 제공된다.

[0196] 직관적으로, 교차-모달리티 기반의 압축 표현을 학습함으로써, 우리는 많은 모달리티에 걸쳐 라벨링된(또는 쌍을 이루는) 데이터를 활용하고 있고, 이것은 일반화 캡을 감소시킨다.

[0197] 본 명세서에 제공된 개시내용의 임의의 실시예와 관련하여 이용될 수 있는 컴퓨터 시스템(700)의 예시적인 구현이 도 7에 도시되어 있다. 컴퓨터 시스템(700)은, 하나 이상의 컴퓨터 하드웨어 프로세서(700), 및 비일시적인 컴퓨터-판독가능한 저장 매체(예를 들어, 메모리(720) 및 하나 이상의 비휘발성 저장 디바이스(730))를 포함하는 하나 이상의 제품을 포함할 수 있다. 프로세서(들)(710)는 임의의 적절한 방식으로 메모리(720) 및 비휘발성 저장 디바이스(들)(730)로의 데이터 기입 및 그로부터 데이터 판독을 제어할 수 있다. 본 명세서에서 설명된 임의의 기능을 수행하기 위해, 프로세서(들)(710)는, 프로세서(들)(710)에 의한 실행을 위한 프로세서-실행가능한 명령어들을 저장하고 있는 비일시적인 컴퓨터-판독가능한 저장 매체로서 역할할 수 있는 하나 이상의 비일시적인 컴퓨터-판독가능한 저장 매체(예를 들어, 메모리(720))에 저장된 하나 이상의 프로세서-실행가능한 명령어들을 실행할 수 있다.

[0198] 용어 "프로그램" 또는 "소프트웨어"는, 앞서 논의된 바와 같은 실시예들의 다양한 형태를 구현하기 위해 컴퓨터 또는 기타의 프로세서(물리적 또는 가상적)를 프로그램하도록 재용될 수 있는 임의의 유형의 컴퓨터 코드 또는 프로세서-실행가능한 명령어 세트를 지칭하는 일반적 의미로 본 명세서에서 사용된다. 추가적으로, 한 형태에 따르면, 실행될 때 본 명세서에서 제공된 본 개시내용의 방법들을 수행하는 하나 이상의 컴퓨터 프로그램은 단일의 컴퓨터 또는 프로세서 상에 존재할 필요는 없고, 본 명세서에서 제공된 본 개시내용의 다양한 형태를 구현하기 위해 다수의 상이한 컴퓨터 또는 프로세서들 사이에서 모듈식으로 분산될 수도 있다.

[0199] 프로세서-실행가능한 명령어들은, 하나 이상의 컴퓨터 또는 기타의 디바이스에 의해 실행되는 프로그램 모듈 등의, 많은 형태로 존재할 수 있다. 일반적으로, 프로그램 모듈은, 특정 태스크들을 수행하거나 특정 추상 데이터 유형들을 구현하는 루틴, 프로그램, 객체, 컴포넌트, 데이터 구조 등을 포함한다. 전형적으로, 프로그램 모듈의 기능은 결합되거나 분산될 수도 있다.

[0200] 또한, 데이터 구조는 임의의 적절한 형태의 하나 이상의 비일시적인 컴퓨터-판독가능한 저장 매체에 저장될 수 있다. 예시의 간소화를 위해, 데이터 구조는 데이터 구조 내의 위치를 통해 관련된 필드들을 갖는 것으로 도시될 수 있다. 이러한 관계는 마찬가지로, 필드들 사이의 관계를 전달하는 비일시적인 컴퓨터-판독가능한 매체 내의 위치들을 갖는 필드들에 대해 저장장소를 할당함으로써 달성될 수 있다. 그러나, 데이터 구조의 필드들 내의 정보간의 관계를 설정하기 위해, 포인터, 태그 또는 데이터 요소들 사이의 관계를 설정하는 다른 메커니즘의 이용을 포함한, 임의의 적절한 메커니즘이 이용될 수 있다.

[0201] 다양한 발명적 개념들은 하나 이상의 프로세스로서 구현될 수 있고, 그 예가 제공되었다. 각각의 프로세스의 일부로서 수행되는 동작들은 임의의 적절한 방식으로 정렬될 수 있다. 따라서, 예시적 실시예에서는 순차적 동작들로서 도시되어 있더라도, 소정 동작들을 동시에 수행하는 것을 포함한, 예시된 것과는 상이한 순서로 동작들이 수행되는 실시예들이 구성될 수 있다.

[0202] 여기 본 명세서와 청구항에서 사용될 때, 하나 이상의 요소의 열거를 참조한 구문 "적어도 하나"는, 요소들의 목록 내의 요소들 중 임의의 하나 이상으로부터 선택된 적어도 하나의 요소를 의미하지만, 요소들의 목록 내의 구체적으로 열거된 각각의 및 모든 요소 중 적어도 하나를 반드시 포함할 필요는 없고 요소들의 목록 내의 요소들의 임의의 조합을 배제하지 않는 것으로 이해되어야 한다. 이 정의는 또한, 구문 "적어도 하나"가 참조하는 요소들의 목록 내에서 구체적으로 식별되는 요소들 이외의 요소들이, 구체적으로 식별되는 요소들과 관련되든 관련되지 않든, 선택사항적으로 존재할 수 있다는 것을 허용한다. 따라서, 예를 들어, "A와 B 중 적어도 하나"(또는, 등가적으로, "A 또는 B 중 적어도 하나", 또는 등가적으로 "A 및/또는 B 중 적어도 하나")는, 한 실시예에서는, B가 존재하지 않는(및 선택사항으로서 B 이외의 요소를 포함함), 선택사항으로서는 하나보다 많은 A를 포함하는, 적어도 하나의 A를 지칭할 수 있고; 또 다른 실시예에서는, A가 존재하지 않는(및 선택사항으로서 A 이외의 요소를 포함함), 선택사항으로서는 하나보다 많은 B를 포함하는, 적어도 하나의 B를 지칭할 수 있고; 역시 또 다른 실시예에서는, 선택사항으로서 하나보다 많은 A를 포함하는, 적어도 하나의 A와, 선택사항으로서는

하나보다 많은 B(및 선택사항으로서 다른 요소들을 포함함)를 포함하는, 적어도 하나의 B를 지칭할 수 있다; 기타 등등.

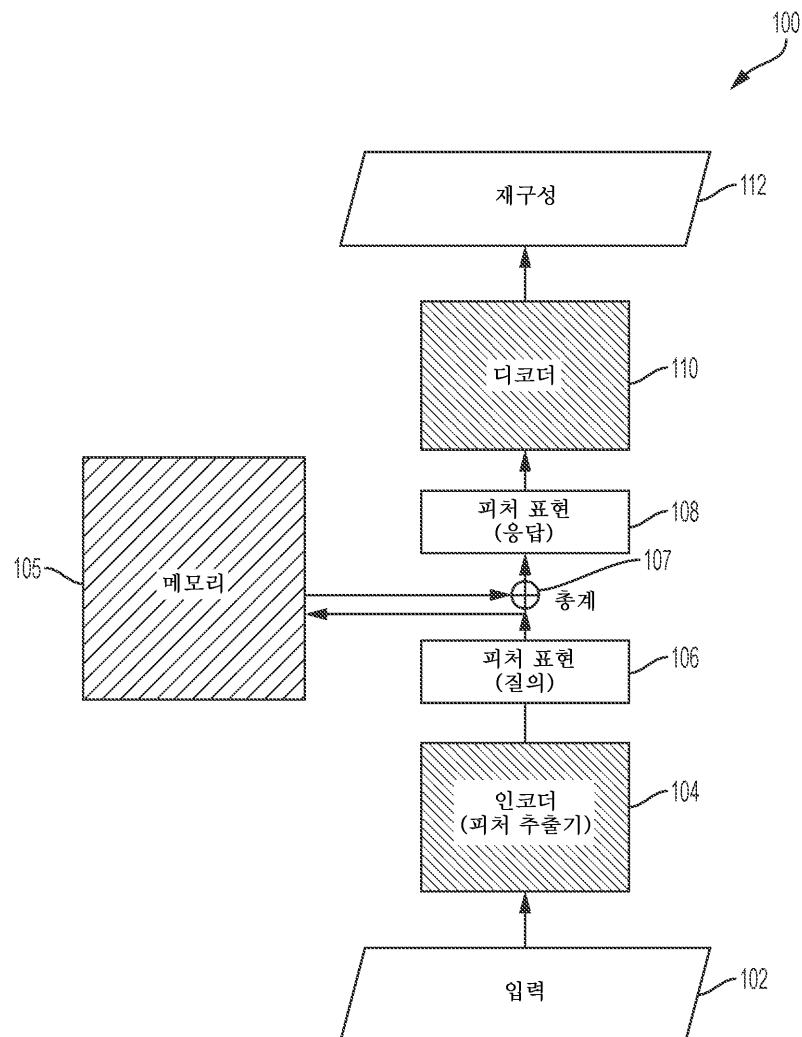
[0203] 구문 "및/또는"은, 여기 본 명세서 및 청구항에서 사용될 때, 이와 같이 결합된 요소들, 즉, 어떤 경우에는 결합되어 존재하고 다른 경우에는 분리되어 존재하는 요소들의 "어느 하나 또는 양쪽 모두"를 의미하는 것으로 이해되어야 한다. "및/또는"에 의해 열거된 복수의 요소들은, 동일한 방식으로, 즉, 이와 같이 결합된 요소들의 "하나 이상"으로서 해석되어야 한다. "및/또는"에 의해 구체적으로 식별되는 요소들 이외의 다른 요소들이, 구체적으로 식별된 이들 요소들과 관련되든 관련되지 않든, 선택사항으로서 존재할 수 있다. 따라서, 비제한적 예로서, "A 및/또는 B"에 대한 언급은, "포함하는" 등의 개방형 용어와 연계하여 사용될 때, 한 실시예에서는, (선택사항으로서 B 이외의 다른 요소를 포함한) A만을 지칭할 수 있고; 또 다른 실시예에서는, (선택사항으로서 A 이외의 요소를 포함한) B만을 지칭할 수 있으며; 역시 또 다른 실시예에서는, (선택사항으로서 다른 요소를 포함한) A와 B 양쪽 모두를 지칭할 수 있다; 기타 등등.

[0204] 청구항 요소를 수정하는 청구항 내의 "제1", "제2", "제3" 등의 서수적 용어의 사용은, 그 자체로, 한 청구항 요소의 또 다른 청구항 요소에 대한 어떠한 우선순위, 우선권, 또는 순서나, 방법의 동작들이 수행되는 시간적 순서를 암시하는 것은 아니다. 이러한 용어들은 (서수적 용어의 사용을 제외하고는) 단순히 소정의 명칭을 갖는 한 청구항 요소를 동일한 명칭을 갖는 또 다른 요소와 구분하기 위한 라벨로서 사용된다. 또한, 여기서 사용되는 어법과 용어는 설명의 목적을 위한 것이며 제한적인 것으로 간주되어서는 안 된다. 여기서 "포함하는 (including)", "포함하는(comprising)", 또는 "갖는(having)", "포함하는(containing)", "포함하는 (involving)" 및 그 파생어들의 사용은, 그 이후에 열거되는 항목들과 추가 항목들을 포괄하는 것을 의미한다.

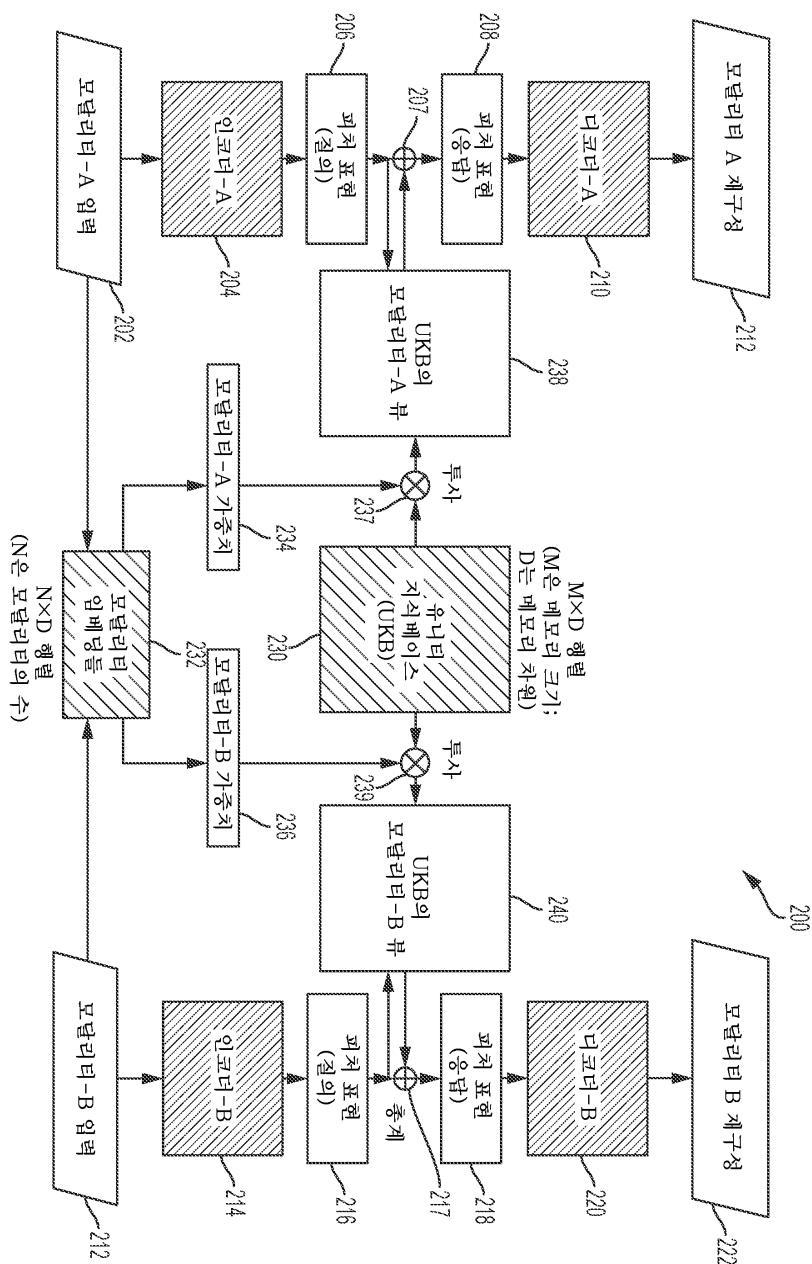
[0205] 본 명세서에서 설명된 기술은 몇 가지 실시예들이 상세히 설명되었지만, 다양한 수정, 및 개선이 본 기술분야의 통상의 기술자에 의해 용이하게 이루어질 수 있다. 이러한 수정 및 개선은 본 개시내용의 사상과 범위 내에 포함시키고자 한다. 따라서, 상기의 설명은 단지 예일 뿐이고 제한하고자 함이 아니다. 본 기술은, 오직 이하의 청구항들과 그 균등률들에 의해 정의된 바와 같이 제한된다.

## 도면

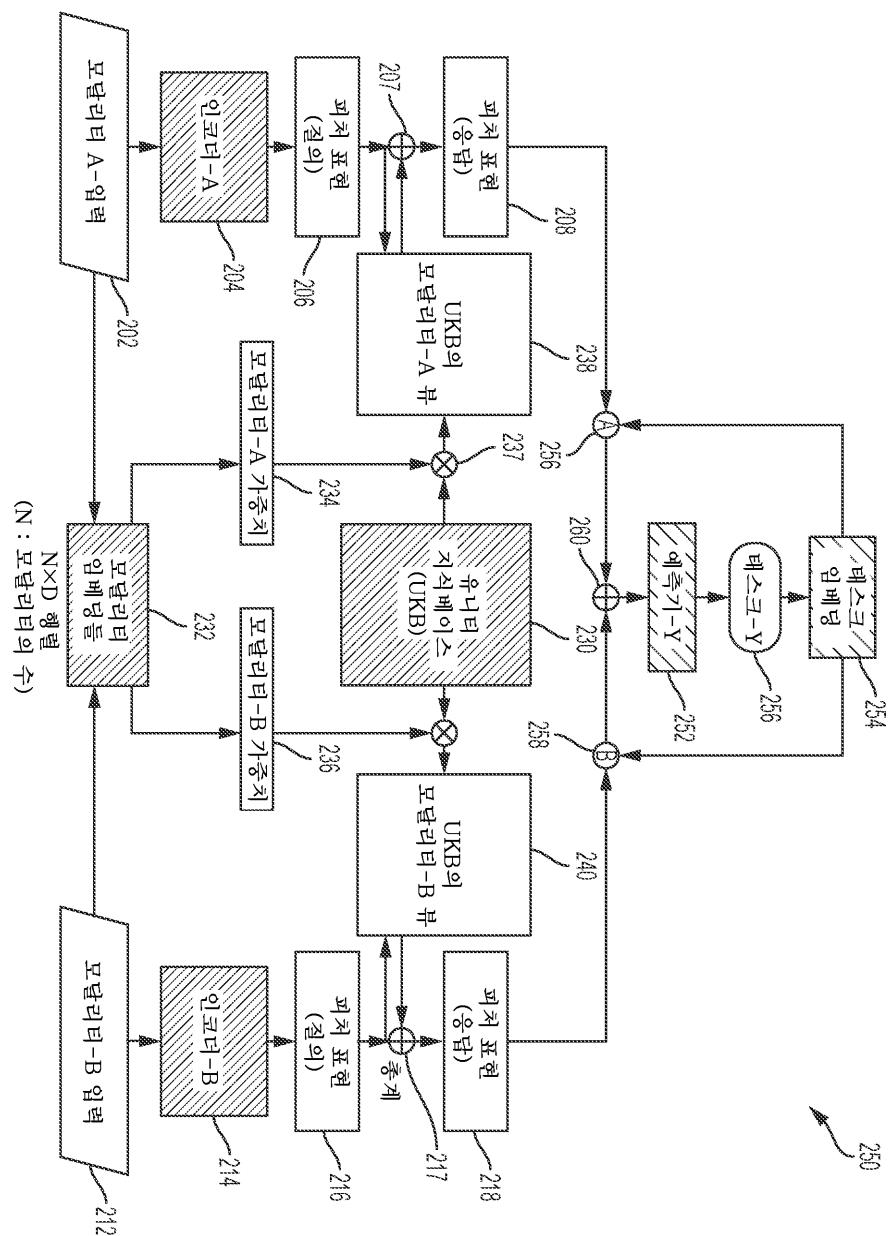
## 도면1



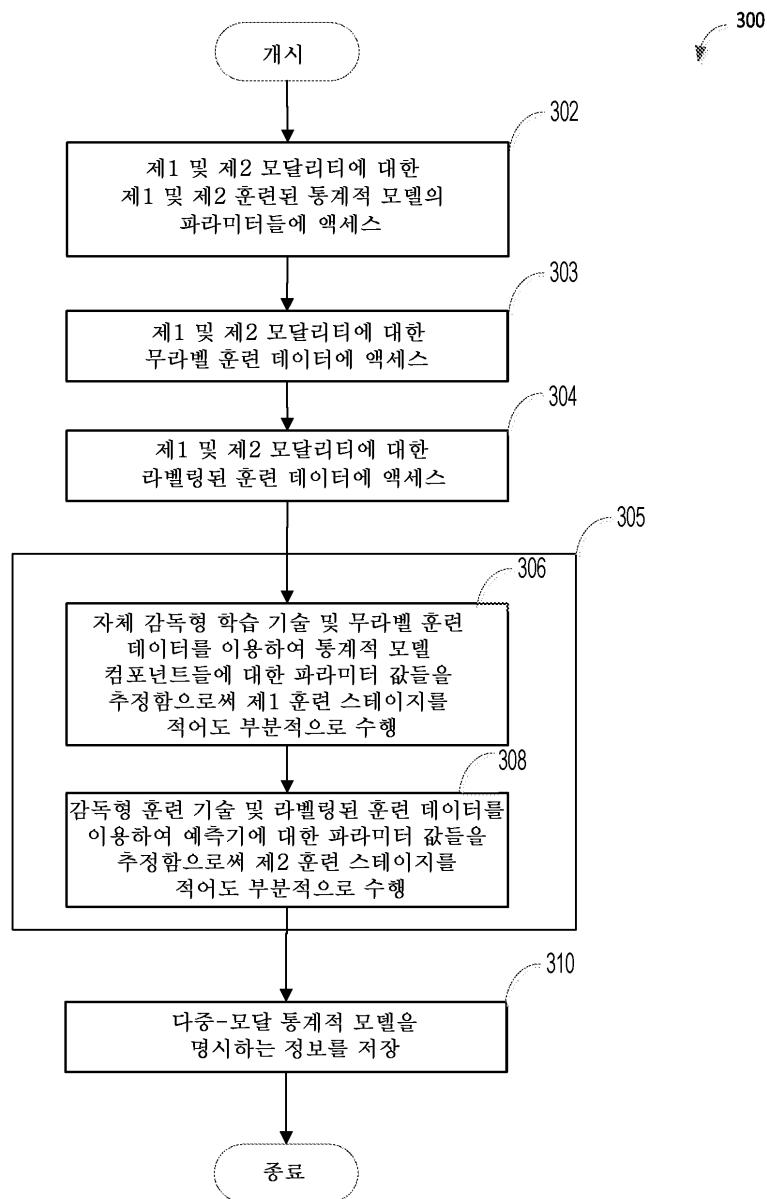
도면2a



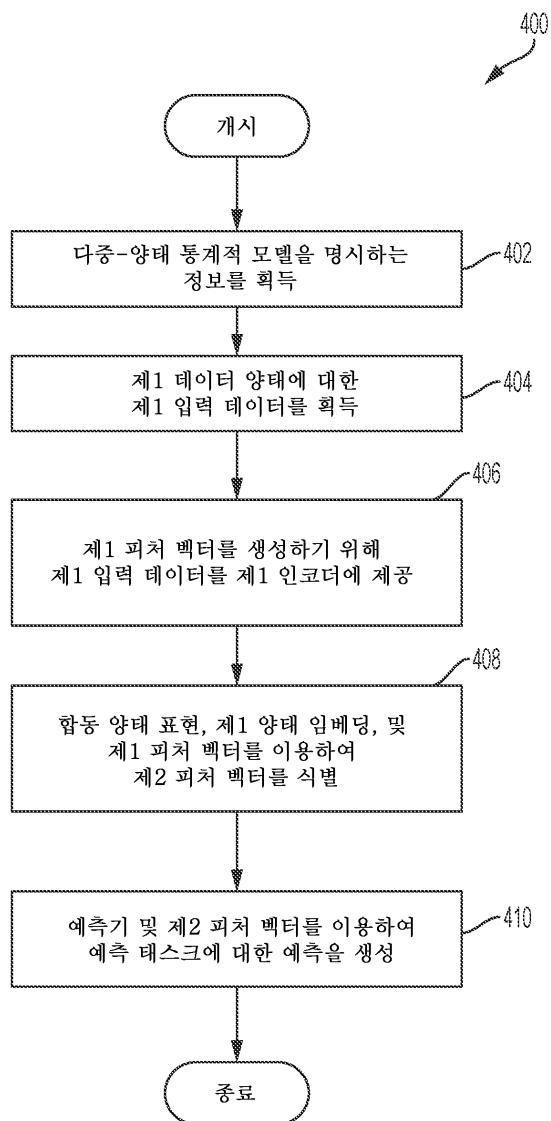
도면2b



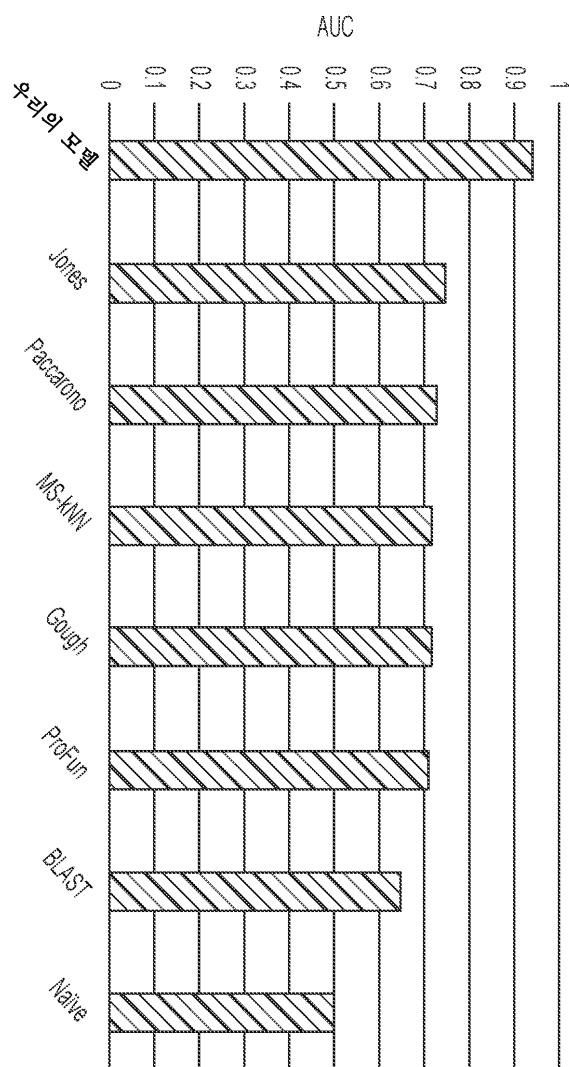
## 도면3



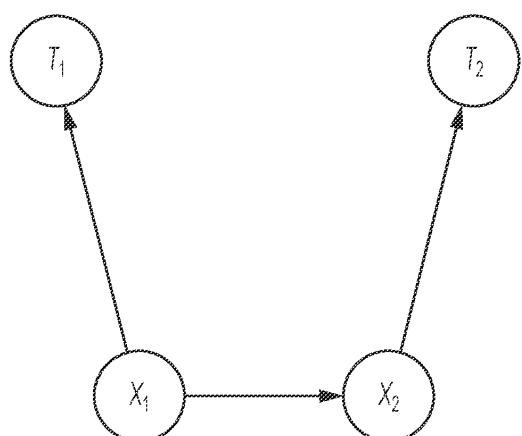
## 도면4



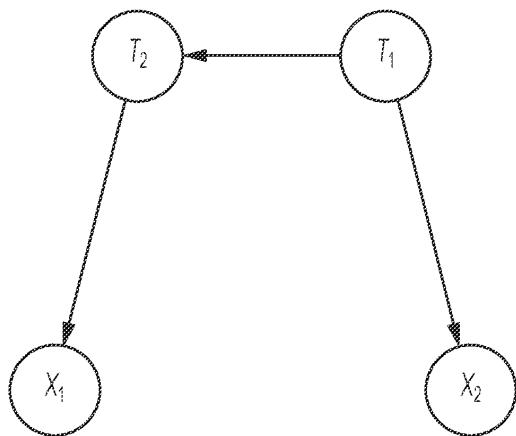
## 도면5



## 도면6



도면6b



도면7

