

(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0056449

(43) 공개일자 2025년04월28일

(51) 국제특허분류(Int. Cl.)

G06F 16/532 (2019.01) G06F 16/538 (2019.01)

G06F 16/583 (2019.01) G06F 16/9032 (2019.01)

G06F 16/9038 (2019.01) G06N 20/00 (2019.01)

(52) CPC특허분류

G06F 16/532 (2019.01)

G06F 16/538 (2019.01)

(21) 출원번호 10-2023-0140183

(22) 출원일자 2023년10월19일

심사청구일자 2023년10월19일

(71) 출원인

국립창원대학교 산학협력단

경상남도 창원시 의창구 창원대학교 20(퇴촌동)

(72) 발명자

차정원

경상남도 창원시 성산구 동산로 115, 105동 1407호 (상남동, 대동아파트)

오진영

경상남도 창원시 의창구 대봉로 27, 301동 1303호 (봉림동, 엘에이치피닉스포레)

(74) 대리인

김종선, 이형석

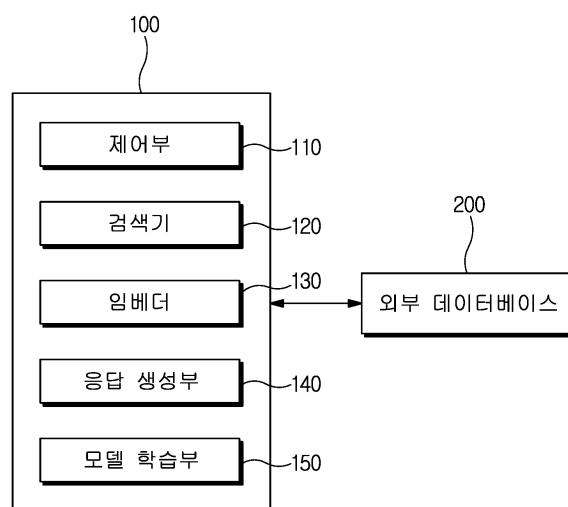
전체 청구항 수 : 총 8 항

(54) 발명의 명칭 멀티모달 모델의 분산 표현 학습 방법 및 장치

(57) 요약

본 발명은 멀티모달 모델의 분산 표현 학습 방법 및 장치에 관한 것으로, 입력 문장을 수신하는 단계, 상기 입력 문장 중 외부 데이터베이스와 매칭 가능한 적어도 하나의 구간을 검색하는 단계, 검색된 적어도 하나의 구간을 상기 외부 데이터베이스의 이미지 정보와 매칭하는 단계, 상기 매칭된 구간 및 매칭되지 않은 구간을 분산 표현으로 변환하는 단계, 상기 분산 표현을 기반으로 문장을 생성하는 단계 및 생성된 문장과 상기 입력 문장을 비교하여, 학습 모델을 업데이트하는 단계를 포함하는 것을 특징으로 한다.

대표도 - 도1



(52) CPC특허분류

G06F 16/5846 (2019.01)
G06F 16/5854 (2019.01)
G06F 16/90332 (2019.01)
G06F 16/9038 (2019.01)
G06N 20/00 (2021.08)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711192744
과제번호	00187238
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	한국어대형인공지능언어모델기술개발
연구과제명	효율적 사전학습이 가능한 한국어 대형 언어모델 사전학습 기술 개발
기 여 율	1/1
과제수행기관명	창원대학교 산학협력단
연구기간	2022.08.01 ~ 2025.12.31

명세서

청구범위

청구항 1

입력 문장을 수신하는 단계;

상기 입력 문장 중 외부 데이터베이스와 매칭 가능한 적어도 하나의 구간을 검색하는 단계;

검색된 적어도 하나의 구간을 상기 외부 데이터베이스의 이미지 정보와 매칭하는 단계;

상기 매칭된 구간 및 매칭되지 않은 구간을 분산 표현으로 변환하는 단계;

상기 분산 표현을 기반으로 문장을 생성하는 단계; 및

생성된 문장과 상기 입력 문장을 비교하여, 학습 모델을 업데이트하는 단계를 포함하는 것을 특징으로 하는 멀티모달 모델의 분산 표현 학습 방법.

청구항 2

제1항에 있어서,

상기 분산 표현으로 변환하는 단계에서는,

상기 매칭된 구간은 매칭된 이미지 정보에 기초하여 이미지 분산 표현으로 변환하는 것을 특징으로 하는 멀티모달 모델의 분산 표현 학습 방법.

청구항 3

제2항에 있어서,

상기 분산 표현으로 변환하는 단계에서는,

상기 매칭되지 않은 구간은 텍스트 분산 표현으로 변환하는 것을 특징으로 하는 멀티모달 모델의 분산 표현 학습 방법.

청구항 4

제3항에 있어서,

상기 매칭된 구간 및 상기 매칭되지 않은 구간을 분산 표현으로 변환하기 전에 상기 입력 문장의 적어도 하나의 구간을 변경하는 단계를 더 포함하는 것을 특징으로 하는 멀티모달 모델의 분산 표현 학습 방법.

청구항 5

제4항에 있어서,

상기 학습 모델을 업데이트하는 단계에서는,

상기 생성된 문장과 변경된 구간의 예측을 비교하여 학습 모델을 업데이트하는 것을 특징으로 하는 멀티모달 모델의 분산 표현 학습 방법.

청구항 6

입력 문장 중 외부 데이터베이스와 매칭 가능한 적어도 하나의 구간을 검색하는 검색기;

상기 검색된 구간을 상기 외부 데이터베이스의 이미지 정보와 매칭하는 제어부;

상기 매칭된 구간 및 매칭되지 않은 구간을 텍스트 분산 표현으로 변환하는 임베더;

상기 변환된 분산 표현을 기반으로 문장을 생성하는 응답 생성부; 및

상기 생성된 문장과 입력 문장을 비교하여 멀티모달 모델을 학습하는 모델 학습부를 포함하는 것을 특징으로 하

는 멀티모달 모델의 분산 표현 학습 장치.

청구항 7

제6항에 있어서,

상기 임베더는,

상기 매칭된 구간은 매칭된 이미지 정보에 기초하여 이미지 분산 표현으로 변환하는 것을 특징으로 하는 멀티모달 모델의 분산 표현 학습 장치.

청구항 8

제7항에 있어서,

상기 임베더는,

상기 매칭되지 않은 구간은 텍스트 분산 표현으로 변환하는 것을 특징으로 하는 멀티모달 모델의 분산 표현 학습 장치.

발명의 설명

기술 분야

[0001] 본 발명은 이미지, 텍스트의 혼합 문맥을 이용한 멀티모달 모델의 분산 표현 학습 방법 및 장치에 관한 것이다.

배경 기술

[0002] 최근 딥러닝 모델의 성능이 높아지는 가운데 여러 도메인(텍스트, 이미지, 비디오 등)을 이용한 딥러닝 모델에 대한 연구가 활발하게 이루어지고 있다.

[0003] 초기에는 여러 도메인에 대한 딥러닝 모델을 학습하기 위해 먼저 하나의 도메인에 대한 딥러닝 모델의 데이터를 구축하고 모델을 구성하여 학습한 뒤, 각 도메인 결과를 합하는 방식을 이용하였다. 이와 같은 방식은 각 모델의 오류를 제어할 수 없는 문제점이 있다. 또한 이는 결과 기반이므로 학습에 대한 높은 성능을 기대하기 어렵다.

[0004] 인간과 비슷한 추론을 하기 위해서는 여러 도메인이 통합된 멀티모달(MultiModal, CrossModal) 모델의 필요성이 높아졌고, 통합 초기 모델은 학습하고자 하는 여러 도메인의 데이터가 각 샘플에 포함되어 있는 구조로 학습하였다.

[0005] 예를 들어, 이미지와 텍스트 통합 딥러닝 모델의 경우 하나의 샘플은 이미지와 연관된 텍스트로 구성되어야 한다. ‘사과가 테이블에 있다’는 문장과 해당 문장이 표현된 이미지가 쌍(pair)로 구성되어 학습한다. 이와 같이 쌍으로 구성된 데이터를 대용량으로 구축하는 것에 한계가 있고 이를 위해 비용 및 시간이 많이 소요되는 단점이 있다.

[0006] 그러므로 이미지와 텍스트가 혼합되어 구성된 문맥을 보다 개선된 성능으로 학습하기 위한 모델의 필요성이 점차 증대되고 있다.

[0007] 이에 본 발명은, 이미지, 텍스트의 혼합 문맥을 이용한 멀티모달 모델의 분산 표현 학습 방법 및 장치를 제안한다.

선행기술문헌

특허문헌

[0008] (특허문헌 0001) 한국공개특허 10-2023-0100534(2023. 07.05)

발명의 내용

해결하려는 과제

- [0009] 본 발명은 상기와 같은 문제점을 해결하기 위하여 안출된 것으로서, 이미지와 텍스트가 혼합되어 구성된 문맥에 대한 분산 표현을 이용하여 학습하는 방법을 제공하는 것이다.
- [0010] 본 발명의 다른 목적은, 입력된 문장에 대하여 이미지와 텍스트의 데이터 쌍을 구성할 필요 없이 이미지 및 텍스트로 구성된 혼합 문맥에 대해 각각의 분산 표현을 이용하여 멀티모달 모델의 학습 방법을 제공하는 것이다.
- [0011] 본 발명의 또 다른 목적은, 혼합 문맥에 대해 각각의 분산 표현을 이용하여 학습을 위한 서버의 용량, 비용 및 시간을 절감하기 위한 멀티모달 모델의 분산 표현 학습 방법을 제공하는 것이다.
- [0012] 본 발명의 기술적 과제들은 이상에서 언급한 기술적 과제로 제한되지 않으며, 언급되지 않은 또 다른 기술적 과제들은 아래의 기재로부터 당업자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

- [0013] 이와 같은 목적을 달성하기 위한 본 발명의 멀티모달 모델의 분산 표현 학습 방법은, 입력 문장을 수신하는 단계, 상기 입력 문장 중 외부 데이터베이스와 매칭 가능한 적어도 하나의 구간을 검색하는 단계, 검색된 적어도 하나의 구간을 상기 외부 데이터베이스의 이미지 정보와 매칭하는 단계, 상기 매칭된 구간 및 매칭되지 않은 구간을 분산 표현으로 변환하는 단계, 상기 분산 표현을 기반으로 문장을 생성하는 단계 및 생성된 문장과 상기 입력 문장을 비교하여, 학습 모델을 업데이트하는 단계를 포함하는 것을 특징으로 한다.
- [0014] 상기 분산 표현으로 변환하는 단계에서는, 상기 매칭된 구간은 매칭된 이미지 정보에 기초하여 이미지 분산 표현으로 변환하는 것을 특징으로 한다.
- [0015] 상기 분산 표현으로 변환하는 단계에서는, 상기 매칭되지 않은 구간은 텍스트 분산 표현으로 변환하는 것을 특징으로 한다.
- [0016] 상기 매칭된 구간 및 상기 매칭되지 않은 구간을 분산 표현으로 변환하기 전에 상기 입력 문장의 적어도 하나의 구간을 변경하는 단계를 더 포함하는 것을 특징으로 한다.
- [0017] 상기 학습 모델을 업데이트하는 단계에서는, 상기 생성된 문장과 변경된 구간의 예측을 비교하여 학습 모델을 업데이트하는 것을 특징으로 한다.
- [0018] 본 발명의 다른 실시 예에 따른 본 발명의 멀티모달 모델의 분산 표현 학습 장치는, 입력 문장 중 외부 데이터베이스와 매칭 가능한 적어도 하나의 구간을 검색하는 검색기, 상기 검색된 구간을 상기 외부 데이터베이스의 이미지 정보와 매칭하는 제어부, 상기 매칭된 구간 및 매칭되지 않은 구간을 텍스트 분산 표현으로 변환하는 임베더, 상기 변환된 분산 표현을 기반으로 문장을 생성하는 응답 생성부 및 상기 생성된 문장과 입력 문장을 비교하여 멀티모달 모델을 학습하는 모델 학습부를 포함하는 것을 특징으로 한다.
- [0019] 상기 임베더는, 상기 매칭된 구간은 매칭된 이미지 정보에 기초하여 이미지 분산 표현으로 변환하는 것을 특징으로 한다.
- [0020] 상기 임베더는, 상기 매칭되지 않은 구간은 텍스트 분산 표현으로 변환하는 것을 특징으로 한다.
- [0021] 상기 제어부는, 상기 매칭된 구간 및 상기 매칭되지 않은 구간을 분산 표현으로 변환하기 전에 상기 입력 문장의 적어도 하나의 구간을 변경하는 것을 특징으로 한다.
- [0022] 상기 모델 학습부는, 상기 생성된 문장과 변경된 구간의 예측을 비교하여 학습 모델을 업데이트하는 것을 특징으로 한다.

발명의 효과

- [0023] 이와 같은 본 발명에 따르면, 본 발명은 이미지와 텍스트가 혼합되어 구성된 문맥에 대해 분산 표현을 이용하여 개선된 멀티모달 모델의 학습하는 방법을 제공할 수 있다.
- [0024] 한편 본 발명에 따른 멀티모달 모델의 분산 표현 학습 방법 및 장치는 이미지 딥러닝 모델을 텍스트 추론에 활용하거나 또는 텍스트 모델을 이미지 추론에 활용할 수 있는 방법을 제공할 수 있다.
- [0025] 한편 본 발명에 따른 멀티모달 모델의 분산 표현 학습 방법 및 장치는, 이미지에 대한 텍스트 질문이 주어졌을

때 텍스트로 답변(Visual Question Answering, VQA)을 제공할 수 있다.

[0026] 나아가 본 발명에 따른 멀티모달 모델의 분산 표현 학습 방법 및 장치는, 이미지에 대한 질문의 답변과 함께 답변에 대한 이유를 생성(Visual Commonsense Reasoning, VCR)할 수 있다.

도면의 간단한 설명

[0027] 도 1은 본 발명에 따른 멀티모달 모델의 분산 표현 학습 장치를 설명하기 위한 개념도이다.

도 2는 본 발명에 따른 멀티모달 모델의 분산 표현 학습 방법을 설명하기 위한 흐름도이다.

도 3은 본 발명에 따른 이미지, 텍스트의 혼합 문맥을 이용한 멀티모달 모델을 설명하기 위한 개념도이다.

발명을 실시하기 위한 구체적인 내용

[0028] 본 발명의 목적 및 효과, 그리고 그것들을 달성하기 위한 기술적 구성들은 첨부되는 도면과 함께 상세하게 후술되어 있는 실시 예들을 참조하면 명확해질 것이다. 본 발명을 설명함에 있어서 공지 기능 또는 구성에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략할 것이다.

[0029] 그리고 후술되는 용어들은 본 발명에서의 기능을 고려하여 정의된 용어들로서 이는 사용자, 운용자의 의도 또는 관례 등에 따라 달라질 수 있다.

[0030] 그러나 본 발명은 이하에서 개시되는 실시 예들에 한정되는 것이 아니라 서로 다른 다양한 형태로 구현될 수 있다. 단지 본 실시 예들은 본 발명의 개시가 완전하도록 하고, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 발명의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 발명은 청구항의 범주에 의해 정의될 뿐이다. 그러므로 그 정의는 본 명세서 전반에 걸친 내용을 토대로 내려져야 할 것이다.

[0031] 따라서, 본 발명의 사상은 설명된 실시 예에 국한되어 정해져서는 아니되며, 후술하는 특허청구범위뿐 아니라 이 특허청구범위와 균등하거나 등가적 변형이 있는 모든 것들은 본 발명 사상의 범주에 속한다고 할 것이다.

[0032] 본 발명은 멀티모달 모델의 분산 표현 학습 방법 및 장치에 관한 것으로서, 이하에서 도면을 참조하여 본 발명을 보다 구체적으로 설명하도록 한다.

[0033] 도 1은 본 발명에 따른 멀티모달 모델의 분산 표현 학습 장치를 설명하기 위한 개념도이고, 도 2는 본 발명에 따른 멀티모달 모델의 분산 표현 학습 방법을 설명하기 위한 흐름도, 도 3은 본 발명에 따른 이미지, 텍스트의 혼합 문맥을 이용한 멀티모달 모델의 분산 표현 학습 방법을 설명하기 위한 개념도이다.

[0034] 먼저 본 발명에 멀티모달 모델의 분산 표현 학습 방법 및 장치의 개념을 도 1을 참조하여 설명하도록 한다.

[0035] 본 발명의 멀티모달 모델의 학습 방식은, 입력에 여러 도메인 정보를 혼합하여 사용하는 방식으로, 하나의 입력에 대해 분산표현의 크기가 100이라고 한다면 멀티모달 방식으로 학습하여도 크기가 증가하지 않는다. 그러나 기존의 학습 방식은, 사용하는 도메인의 개수만큼 분산 표현의 크기가 더해져 학습을 위한 물리 서버의 용량(또는 요구사항)이 높아진다. 예를 들어, 학습을 위한 입력 데이터가 텍스트와 이미지로 구성된 경우 텍스트의 분산 표현의 크기 100, 이미지의 분산 표현의 크기 200이 포함되어 입력에 대해 300의 분산표현의 크기가 사용되므로 서버의 용량이 높아짐으로써 부하가 발생할 수 있다.

[0036] 본 발명에 따른 멀티모달 모델의 분산 표현 학습 장치의 입력 문장이 텍스트로 구성된 문장이고, 텍스트에 표현된 이미지 키워드 구간(예를 들어, 나무, 벤치 등)을 해당 이미지로 매칭(또는 연결, 또는 연동)하여 학습하므로 쌍(pair)의 데이터가 필요하지 않다. 따라서 본 발명에 따른 멀티모달 모델의 분산 표현 학습 장치의 입력은, 문장과 이미지가 혼합된 형태로 문맥이 구성될 수 있다. 이와 같이 본 발명에 따른 멀티모달 모델의 분산 표현 학습 장치의 학습 모델을 통해 이미지와 텍스트의 분산 표현이 유사해지게 됨으로써 응용 모델의 성능을 향상시킬 수 있다.

[0037] 도 1에 도시된 것과 같이, 본 발명에 따른 멀티모달 모델의 분산 표현 학습 장치(100, 이하에서는 '학습 장치'로 명명함)는, 제어부(110), 검색기(120), 임베더(130), 응답 생성부(140) 및 모델 학습부(150) 중 적어도 하나를 포함하도록 구성될 수 있다. 이때 본 발명에 따른 학습 장치(100)는 상술한 구성 요소로 제한되는 것은 아니며, 본 명세서에 설명에 따른 기능과 동일 또는 유사한 역할을 수행하는 구성요소를 더 포함할 수 있다.

[0038] 한편, 본 발명에 따른 학습 장치(100)는 외부 데이터베이스(200)와 필요에 따라 연동될 수 있다. 본 발명에 따

른 학습 장치(100)는 통신 모듈을 통해 외부 데이터베이스(200)로부터 필요한 데이터를 송수신할 수 있다. 구체적으로, 본 발명은 통신 모듈(예를 들어, 이동통신 모듈, 근거리 통신 모듈, 무선 인터넷 모듈, 위치정보 모듈, 방송 수신 모듈 등)의 구성을 이용하여, 적어도 하나의 외부 서버(예를 들어, 데이터베이스 서버(또는 외부 저장소) 또는 클라우드 서버 등)와 데이터를 송수신할 수 있다.

[0039] 즉, 통신 모듈은, 학습 장치(100)와 외부 데이터베이스 사이, 또는 학습 장치(100)와 외부 서버 사이의 무선 또는 유선 통신을 가능하도록 하나 이상의 모듈을 포함할 수 있다. 또한, 상기 통신 모듈은, 학습 장치(100)를 하나 이상의 네트워크에 연결하는 하나 이상의 모듈을 포함할 수 있다.

[0040] 이러한 통신 모듈은, 사용자에게 의해 입력된 입력 문장을 수신할 수 있다. 또한 통신 모듈은, 외부 데이터베이스로부터 이미지 정보, 비디오 정보, 오디오 정보, 텍스트 정보 등을 수신할 수 있다.

[0041] 한편, 저장부는, 학습 장치(100)에 구비된(또는 삽입된) 메모리 또는 외부 서버(데이터베이스 서버(또는 외부 저장소) 또는 클라우드 서버 등)를 이용하여, 입력 문장에 매칭되는 데이터 정보를 저장할 수 있다. 여기서 저장부는, 데이터베이스(Database, DB)라고도 명명될 수 있다.

[0042] 한편, 제어부(110)는, 학습을 수행하기 위한 데이터를 조절할 수 있다. 또한 제어부(110)는, 입력 문장(또는 입력 데이터)을 검색기(120)에 전달할 수 있다.

[0043] 나아가 제어부(110)는, 본 발명에 따른 학습 장치(100)의 전반적인 동작을 제어할 수 있다.

[0044] 한편, 검색기(120)는 입력 문장(또는 입력 데이터)에서 외부 데이터베이스와 매칭(또는 연결)될 수 있는 구간을 검색(또는 서칭)할 수 있다. 즉, 검색기(120)는, 입력 문장 중 외부 데이터베이스와 매칭 가능한 적어도 하나의 구간을 검색할 수 있다. 이때 제어부(110)는, 검색된 구간을 상기 외부 데이터베이스의 이미지 정보와 매칭할 수 있다.

[0045] 또한 검색기(120)는, 검색된 구간을 상기 외부 데이터베이스의 이미지 정보와 매칭하여 임베더(130)에 전달한다.

[0046] 한편, 임베더(130)는 입력 문장 및 매칭된 이미지 정보를 기반으로 분산 표현으로 변환할 수 있다. 구체적으로 임베더(130)는, 입력 문장과 외부 데이터베이스 간에 매칭된 구간 및 매칭되지 않은 구간을 텍스트 분산 표현으로 변환할 있다.

[0047] 한편, 제어부(110)는, 매칭된 구간 및 매칭되지 않은 구간을 분산 표현으로 변환하기 전에 입력 문장의 적어도 하나의 구간을 변경할 수 있다. 이는 모델 학습부(150)를 통해 생성된 문장과 변경된 구간의 예측을 비교하여 학습 모델을 업데이트하기 위함이다.

[0048] 임베더(130)는, 매칭된 구간을 매칭된 이미지 정보에 기초하여 이미지 분산 표현으로 변환할 수 있다. 또한 임베더(130)는, 상기 매칭되지 않은 구간을 텍스트 분산 표현으로 변환할 수 있다.

[0049] 한편, 응답 생성부(140)는, 입력 문장 및 매칭된 이미지 정보의 분산 표현을 기반으로 문장을 생성할 수 있다. 즉, 응답 생성부(140)는, 입력 문장과 외부 데이터베이스 간에 매칭된 구간 및 매칭되지 않은 구간에 대한 변환된 분산 표현을 기반으로 문장을 생성할 수 있다.

[0050] 모델 학습부(150)는, 상기 응답 생성부(140)에 의해 생성된 문장과 입력 문장을 비교하여 업데이트를 하는 방식으로 멀티모달 모델에 대한 학습을 수행할 수 있다.

[0051] 이상에서는 본 발명에 따른 혼합 문맥의 분산 표현을 이용한 학습 장치(100) (100)의 구성요소를 살펴보았다.

[0052] 이하에서는 본 발명에 따른 혼합 문맥의 분산 표현을 이용한 학습 장치(100)를 통해, 학습하기 위한 방법을 도 2를 참고하여 구체적으로 살펴보도록 한다.

[0053] 제어부(110)는, 통신 모듈을 통해 적어도 하나의 입력 문장을 수신할 수 있다(S210). 제어부(110)는, 수신된 입력 문장을 검색기(120)로 전달할 수 있다.

[0054] 검색기(120)는, 입력 문장에서 외부 데이터베이스에 매칭 가능한 적어도 하나의 구간을 검색할 수 있다(S220). 여기서 외부 데이터베이스에 매칭되는 구간은, 입력 문장에 외부 데이터베이스에 포함된 이미지 정보와 매칭되는 구간으로 여러 어절을 포함할 수 있다. 예를 들어, 도3에 도시된 것과 같이, 입력 문장이 ‘식당에는 현대식 나무 탁자와 의자가 있다’인 경우, ‘나무 탁자’와 ‘의자’가 외부 데이터베이스와 매칭되는 구간이 될 수 있다.

- [0055] 제어부(110)는, 입력 문장에서 검색된 적어도 하나의 구간을 외부 데이터베이스의 이미지 정보와 매칭할 수 있다(S230).
- [0056] 한편, 제어부(110)는, 입력 문장 및 매칭된 이미지 정보를 임베더(130)로 전달한다. 임베더(130)는, 텍스트 및 이미지 혼합 문맥으로 구성된 문장에 대해 딥러닝 학습을 수행하기 위한 분산 표현으로 변환할 수 있다.
- [0057] 한편, 제어부(110)는, 임베더(130)가 입력 문장을 분산 표현으로 변환하기 전에, 멀티모달 모델 학습을 위해 입력 문장의 일부 구간을 다른 표현으로 교체할 수 있다(S240). 즉 제어부(110)는, 입력 문장의 적어도 하나의 구간을 변경할 수 있다. 이때, 제어부(110)는, 변경된 구간을 예측하는 방식으로 멀티모달 모델 학습을 수행할 수 있다. 이러한 학습 방식은, 사전 언어 학습 모델인 BERT, GPT 등에서 사용된다.
- [0058] 한편, 임베더(130)는, 입력 문장에서 외부 데이터베이스와 매칭된 구간 및 매칭되지 않은 구간을 분산 표현으로 변환할 수 있다(S250). 구체적으로, 임베더(130)는, 외부 데이터베이스에 매칭된 구간은 매칭된 이미지 정보에 기초하여 이미지 분산 표현으로 변환할 수 있다. 또한 임베더(130)는, 외부 데이터베이스에 매칭된 구간 이외의 구간, 즉 외부 데이터베이스에 매칭되지 않은 구간은 텍스트 분산 표현으로 변환할 수 있다. 따라서 본 발명에 따른 멀티모달 모델의 분산 표현 학습 방법은, 이미지와 텍스트가 혼합된 혼합 문맥에 대한 분산 표현을 이용하여 학습 모델에 대한 학습을 수행할 수 있다.
- [0059] 한편, 응답 생성부(140)는, 외부 데이터베이스에 매칭된 구간 및 매칭되지 않은 구간에 대한 분산 표현을 기반으로 문장을 생성할 수 있다(S260). 즉, 응답 생성부(140)는, 입력 문장의 분산 표현을 학습 모델에 입력함으로써 예측 문장을 생성할 수 있다.
- [0060] 한편, 제어부(110)는, 응답 생성부(140)에 의해 생성된 문장과 입력 문장을 비교하고(S270), 학습 모델을 업데이트 할 수 있다(S280). 또한 제어부(110)는, 생성된 문장과 변경된 구간의 예측을 비교하여 학습 모델을 업데이트할 수 있다. 제어부(110)는, 생성된 문장과 변경된 구간의 예측을 비교 판단한 결과, 일치하지 않는 경우, 예측에 오류가 발생한 것으로 이에 대한 피드백을 수행할 수 있다. 즉 제어부(110)는, 생성된 문장과 변경한 부분의 일치 여부를 반영하여 학습 모델을 업데이트함으로써 멀티모달 모델에 대한 학습을 수행할 수 있다. 따라서 본 발명에 따른 혼합 문맥의 분산 표현을 이용한 학습 방법은 이러한 피드백, 즉 업데이트를 통해 보다 정확도 높은 학습 모델을 구성할 수 있다.
- [0061] 위에서 살펴본 것과 같이, 본 발명은 이미지와 텍스트가 혼합되어 구성된 문맥에 대해 분산 표현을 이용하여 개선된 학습하는 방법을 제공할 수 있다.
- [0062] 한편 본 발명에 따른 혼합 문맥의 분산 표현을 이용한 학습 방법 및 장치는 이미지 딥러닝 모델을 텍스트 추론에 활용하거나 또는 텍스트 모델을 이미지 추론에 활용할 수 있는 방법을 제공할 수 있다.
- [0063] 한편 본 발명에 따른 혼합 문맥의 분산 표현을 이용한 학습 방법 및 장치는, 이미지에 대한 텍스트 질문이 주어졌을 때 텍스트로 답변(Visual Question Answering, VQA)을 제공할 수 있다.
- [0064] 나아가 본 발명에 따른 혼합 문맥의 분산 표현을 이용한 학습 방법 및 장치는, 이미지에 대한 질문의 답변과 함께 답변에 대한 이유를 생성(Visual Commonsense Reasoning, VCR)할 수 있다.
- [0066] 이상과 같이 본 발명의 도시된 실시 예를 참고하여 설명하고 있으나, 이는 예시적인 것들에 불과하며, 본 발명이 속하는 기술 분야의 통상의 지식을 가진 자라면 본 발명의 요지 및 범위에 벗어나지 않으면서도 다양한 변형, 변경 및 균등한 타 실시 예들이 가능하다는 것을 명백하게 알 수 있을 것이다. 따라서 본 발명의 진정한 기술적 보호 범위는 첨부된 청구범위의 기술적인 사상에 의해 정해져야 할 것이다.

부호의 설명

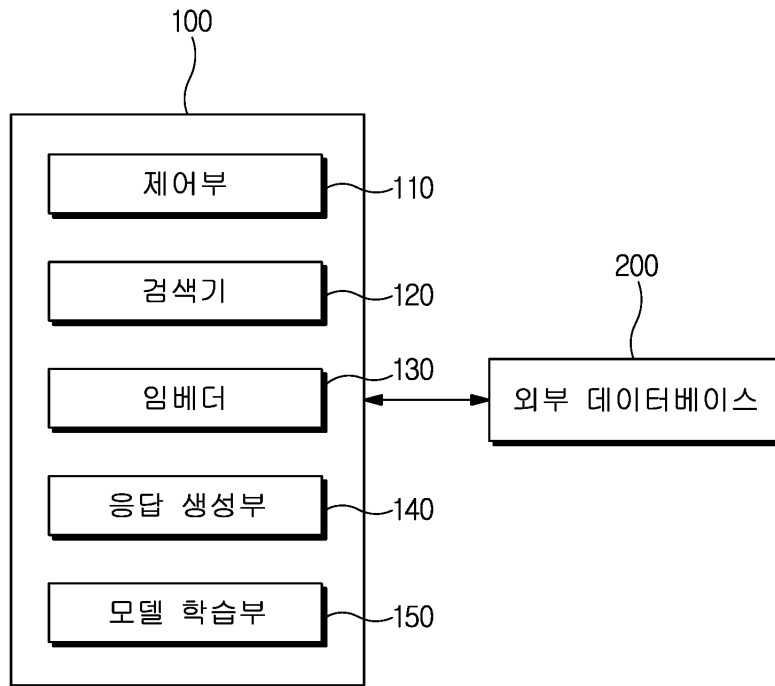
- [0067] 100: 본 발명의 학습장치
- 110: 제어부
- 120: 검색기
- 130: 임베더
- 140: 응답 생성부

150: 모델 학습부

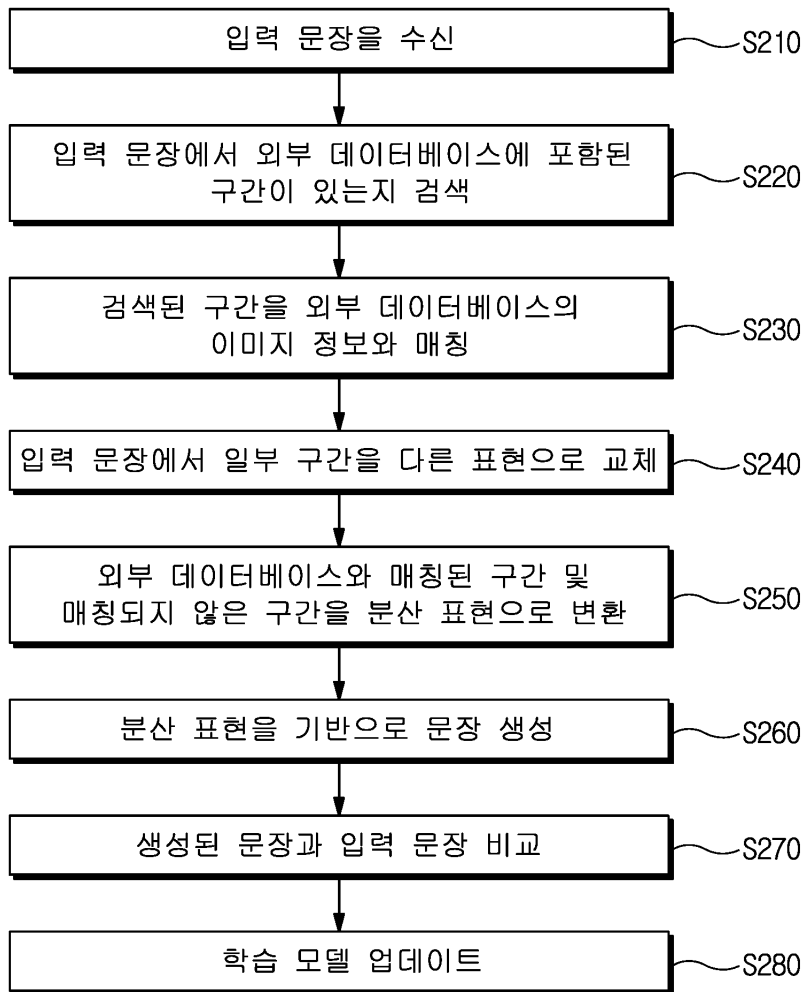
200: 데이터베이스

도면

도면1



도면2



도면3

