(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0071791
(43) 공개일자 2022년05월31일

(51) 국제특허분류(Int. Cl.)
G06F 16/332 (2019.01) *G06F 16/33* (2019.01)
G06F 40/284 (2020.01) *G06N 3/04* (2006.01)
G06N 3/08 (2006.01)
(52) CPC특허분류
G06F 16/3329 (2019.01)
G06F 16/3347 (2019.01)
(21) 출원번호 10-2020-0159325
(22) 출원일자 2020년11월24일
심사청구일자 2020년11월24일

(71) 출원인
경희대학교 산학협력단
경기도 용인시 기흥구 덕영대로 1732 (서천동, 경희대학교 국제캠퍼스내)
(72) 발명자
박성배
경기도 용인시 기흥구 덕영대로 1732 전자정보대학 318호
김소언
경기도 용인시 기흥구 덕영대로 1732 전자정보대학 436호
(뒷면에 계속)
(74) 대리인
특허법인지원

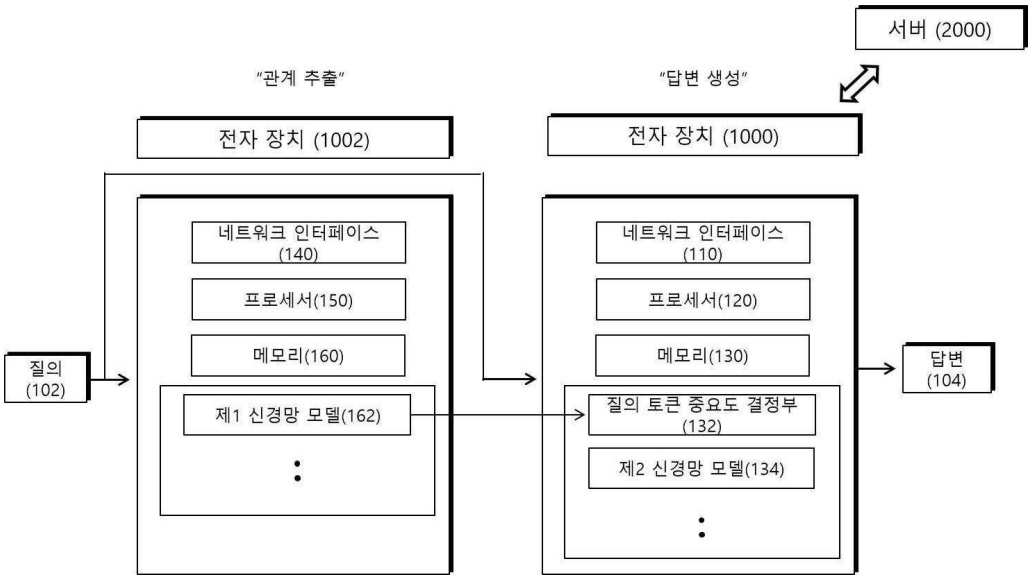
전체 청구항 수 : 총 17 항

(54) 발명의 명칭 **응답과 질의 토큰의 관계에 기초하여 상기 질의에 대한 답변을 제공하는 방법 및 전자 장치**

(57) 요 약

본 개시는 사용자 질의에 대한 답변을 제공하는 방법 및 이를 수행하는 전자 장치에 관한 것이다. 일 실시 예에 의하면, 전자 장치가 사용자 질의에 대한 답변을 제공하는 방법은 상기 사용자로부터 질의를 획득하는 단계; 상기 질의가 입력되면, 상기 질의의 토큰 별 중요도를 결정하고, 상기 결정된 상기 질의의 토큰 별 중요도에 기초하여, 상기 질의 및 상기 질의에 대한 전체 답변의 유사도를 출력하는 제1 신경망 모델로부터 상기 질의의 토큰 별 중요도를 획득하는 단계; 및 상기 질의 및 상기 획득된 질의의 토큰 별 중요도가 입력되면 상기 질의에 대한 답변을 출력하는 제2 신경망 모델에 상기 질의 및 상기 획득된 질의의 토큰 별 중요도를 입력함으로써, 상기 질의에 대한 응답을 생성하는 단계; 를 포함할 수 있다.

대 표 도



(52) CPC특허분류
G06F 40/284 (2020.01)
G06N 3/0454 (2013.01)
G06N 3/08 (2013.01)

(72) 발명자
박형준
경기도 용인시 기흥구 덕영대로 1732 전자정보대학 436호
박규민
경기도 용인시 기흥구 덕영대로 1732 전자정보대학 436호
윤수환
경기도 용인시 기흥구 덕영대로 1732 전자정보대학 436호
이준범
경기도 용인시 기흥구 덕영대로 1732 전자정보대학 436호
임연수
경기도 용인시 기흥구 덕영대로 1732 전자정보대학 436호
김아영
경기도 용인시 기흥구 덕영대로 1732 전자정보대학 436호
카이발리나 나지라
경기도 용인시 기흥구 덕영대로 1732 전자정보대학 436호
권성구
경기도 용인시 기흥구 덕영대로 1732 전자정보대학 436호

이 발명을 지원한 국가연구개발사업

과제고유번호	1711103335
과제번호	2013-2-00109-008
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	혁신성장동력프로젝트(R&D)(과기정통부)(1311101501551100115841l)
연구과제명	(엑소브레인-2세부) WiseKB: 빅데이터 이해 기반 자가학습형 지식베이스 및 추론 기술 개발
기 여 율	1/1
과제수행기관명	주식회사 솔트룩스
연구기간	2020.01.01 ~ 2020.12.31

명 세 서

청구범위

청구항 1

전자 장치가 사용자의 질의에 대한 답변을 제공하는 방법에 있어서,

상기 사용자로부터 질의를 획득하는 단계;

상기 질의가 입력되면, 상기 질의의 토큰 별 중요도를 결정하고, 상기 결정된 상기 질의의 토큰 별 중요도에 기초하여, 상기 질의 및 상기 질의에 대한 전체 답변의 유사도를 출력하는 제1 신경망 모델로부터 상기 질의의 토큰 별 중요도를 획득하는 단계; 및

상기 질의 및 상기 획득된 질의의 토큰 별 중요도가 입력되면 상기 질의에 대한 답변을 출력하는 제2 신경망 모델에 상기 질의 및 상기 획득된 질의의 토큰 별 중요도를 입력함으로써, 상기 질의에 대한 응답을 생성하는 단계; 를 포함하는, 방법.

청구항 2

제1항에 있어서, 상기 방법은

상기 생성된 응답을 상기 사용자의 질의에 대한 답변으로 제공하는 단계; 를 더 포함하는, 방법.

청구항 3

제1항에 있어서, 상기 제1 신경망 모델은

소정의 질의 및 상기 질의에 대한 답변을 포함하는 대화쌍들이 저장된 데이터 베이스로부터, 상기 질의를 획득하고, 상기 질의를 인코딩함으로써 상기 질의에 대한 표현 벡터를 생성하는 질의 표현 벡터 인코더; 및

상기 질의에 대한 답변을 획득하고, 상기 획득된 답변 전체를 인코딩함으로써 상기 답변 전체에 대한 표현 벡터를 생성하는 답변 전체 표현 벡터 인코더; 를 포함하는 것인, 방법.

청구항 4

제3항에 있어서, 상기 질의 표현 벡터 인코더는

상기 질의를 토큰화함으로써 질의에 대한 적어도 하나의 토큰을 생성하고, 상기 생성된 토큰 별 표현 벡터를 생성하는 질의 토큰 표현 벡터 생성부; 및

상기 생성된 토큰 별 표현 벡터의 중요도를 결정하는 질의 토큰 중요도 결정부; 를 포함하고,

상기 질의 표현 벡터 인코더는 상기 결정된 질의 토큰 중요도에 기초하여 상기 토큰 별 표현 벡터를 가중합합으로써 상기 질의에 대한 표현 벡터를 생성하는 것인, 방법.

청구항 5

제4항에 있어서,

상기 질의 표현 벡터 인코더 및 상기 답변 전체 표현 벡터 인코더는, 상기 질의에 대한 표현 벡터 및 상기 답변 전체에 대한 표현 벡터가 서로 유사해지도록 미리 학습되고,

상기 질의에 대한 표현 벡터 및 상기 답변 전체에 대한 표현 벡터가 서로 유사해지도록 상기 질의 표현 벡터 인코더가 학습됨에 따라 상기 질의 표현 벡터 인코더 내 상기 질의 토큰 중요도 결정부는 상기 답변 전체에 대한 내용에 기초하여 상기 질의의 토큰 별 중요도를 결정하게 되는 것을 특징으로 하는, 방법.

청구항 6

제1항에 있어서, 상기 제2 신경망 모델은

상기 사용자로부터 질의가 입력되면 상기 질의를 인코딩하는 인코더; 및

상기 인코더의 출력 값 및 상기 인코더의 출력 값에 상기 질의의 토큰 별 중요도를 반영함으로써 결정되는 인코더 출력 변환 값을 획득하고, 상기 획득된 인코더의 출력 값 및 상기 인코더 출력 변환 값을 디코딩하는 디코더; 를 포함하는 것을 특징으로 하는, 방법.

청구항 7

제6항에 있어서, 상기 디코더는

상기 인코더의 출력 값 및 상기 인코더 출력 변환 값을 순차로 또는 함께 획득하는 것을 특징으로 하는, 방법.

청구항 8

제1항에 있어서, 상기 제2 신경망 모델은

시퀀스 투 시퀀스 구조의 신경망 모델인 것을 특징으로 하는, 방법.

청구항 9

사용자의 질의에 대한 답변을 제공하는 전자 장치에 있어서,

네트워크 인터페이스;

하나 이상의 인스트럭션을 저장하는 메모리; 및

상기 하나 이상의 인스트럭션을 실행하는 적어도 하나의 프로세서; 를 포함하고,

상기 적어도 하나의 프로세서는

상기 사용자로부터 질의를 획득하고,

상기 질의가 입력되면, 상기 질의의 토큰 별 중요도를 결정하고, 상기 결정된 상기 질의의 토큰 별 중요도에 기초하여, 상기 질의 및 상기 질의에 대한 전체 답변의 유사도를 출력하는 제1 신경망 모델로부터 상기 질의의 토큰 별 중요도를 획득하고,

상기 질의 및 상기 획득된 질의의 토큰 별 중요도가 입력되면 상기 질의에 대한 답변을 출력하는 제2 신경망 모델에 상기 질의 및 상기 획득된 질의의 토큰 별 중요도를 입력함으로써, 상기 질의에 대한 응답을 생성하는, 전자 장치.

청구항 10

제9항에 있어서, 상기 적어도 하나의 프로세서는

상기 생성된 응답을 상기 사용자의 질의에 대한 답변으로 제공하는, 전자 장치.

청구항 11

제9항에 있어서, 상기 제1 신경망 모델은

소정의 질의 및 상기 질의에 대한 답변을 포함하는 대화쌍들이 저장된 데이터 베이스로부터, 상기 질의를 획득하고, 상기 질의를 인코딩함으로써 상기 질의에 대한 표현 벡터를 생성하는 질의 표현 벡터 인코더; 및

상기 질의에 대한 답변을 획득하고, 상기 획득된 답변 전체를 인코딩함으로써 상기 답변 전체에 대한 표현 벡터를 생성하는 답변 전체 표현 벡터 인코더; 를 포함하는 것인, 전자 장치.

청구항 12

제11항에 있어서, 상기 질의 표현 벡터 인코더는

상기 질의를 토큰화함으로써 질의에 대한 적어도 하나의 토큰을 생성하고, 상기 생성된 토큰 별 표현 벡터를 생성하는 질의 토큰 표현 벡터 생성부; 및

상기 생성된 토큰 별 표현 벡터의 중요도를 결정하는 질의 토큰 중요도 결정부; 를 포함하고,

상기 질의 표현 벡터 인코더는 상기 결정된 질의 토큰 중요도에 기초하여 상기 토큰 별 표현 벡터를 가중합함으로써 상기 질의에 대한 표현 벡터를 생성하는 것인, 전자 장치.

청구항 13

제12항에 있어서,

상기 질의 표현 벡터 인코더 및 상기 답변 전체 표현 벡터 인코더는, 상기 질의에 대한 표현 벡터 및 상기 답변 전체에 대한 표현 벡터가 서로 유사해지도록 미리 학습되고,

상기 질의에 대한 표현 벡터 및 상기 답변 전체에 대한 표현 벡터가 서로 유사해지도록 상기 질의 표현 벡터 인코더가 학습됨에 따라 상기 질의 표현 벡터 인코더 내 상기 질의 토큰 중요도 결정부는 상기 답변 전체에 대한 내용에 기초하여 상기 질의의 토큰 별 중요도를 결정하게 되는 것을 특징으로 하는, 전자 장치.

청구항 14

제9항에 있어서, 상기 제2 신경망 모델은

상기 사용자로부터 질의가 입력되면 상기 질의를 인코딩하는 인코더; 및

상기 인코더의 출력 값 및 상기 인코더의 출력 값에 상기 질의의 토큰 별 중요도를 반영함으로써 결정되는 인코더 출력 변환 값을 획득하고, 상기 획득된 인코더의 출력 값 및 상기 인코더 출력 변환 값을 디코딩하는 디코더; 를 포함하는 것을 특징으로 하는, 전자 장치.

청구항 15

제14항에 있어서, 상기 디코더는

상기 인코더의 출력 값 및 상기 인코더 출력 변환 값을 순차로 또는 함께 획득하는 것을 특징으로 하는, 전자 장치.

청구항 16

제9항에 있어서, 상기 제2 신경망 모델은

시퀀스 투 시퀀스 구조의 신경망 모델인 것을 특징으로 하는, 전자 장치.

청구항 17

전자 장치가 사용자의 질의에 대한 답변을 제공하는 방법에 있어서,

상기 사용자로부터 질의를 획득하는 단계;

상기 질의가 입력되면, 상기 질의의 토큰 별 중요도를 결정하고, 상기 결정된 상기 질의의 토큰 별 중요도에 기초하여, 상기 질의 및 상기 질의에 대한 전체 답변의 유사도를 출력하는 제1 신경망 모델로부터 상기 질의의 토큰 별 중요도를 획득하는 단계; 및

상기 질의 및 상기 획득된 질의의 토큰 별 중요도가 입력되면 상기 질의에 대한 답변을 출력하는 제2 신경망 모델에 상기 질의 및 상기 획득된 질의의 토큰 별 중요도를 입력함으로써, 상기 질의에 대한 응답을 생성하는 단계; 를 포함하는, 방법을 수행하도록 하는 프로그램이 저장된 컴퓨터로 읽을 수 있는 기록 매체.

발명의 설명

기 술 분 야

[0001] 본 개시는 질의에 대한 답변을 제공하는 방법 및 장치에 관한 것이다. 보다 상세하게는 신경망 기반의 모델을 이용하여 질의에 대한 답변을 제공하는 방법 및 장치에 관한 것이다.

배 경 기 술

[0002] 인터넷 상에 존재하는 데이터의 양이 기하급수적으로 증가함에 따라 오늘날 사용자가 원하는 정보를 제공하기 위한 데이터 처리 기술 개발이 활발하게 연구되고 있다. 이에 따라 사용자의 질의와 관련된 응답을 제공하는 기

술 역시 활발하게 진행되고 있다.

[0003] 최근 시퀀스 투 시퀀스(sequence to sequence) 모델을 이용하여 질의에 대한 응답을 제공하기 위한 기술들이 연구되고 있다. 시퀀스 투 시퀀스 모델을 이용한 답변 생성기는 어느 질의에도 답변이 될 수 있는 일반적인 답변의 생성 확률이 높은 한계가 있다. 종래의 시퀀스 투 시퀀스 모델은 토큰 단위로 디코딩 시 질의에 대한 답변 전체 내용에 대한 정보가 반영되지 않는 한계가 있기 때문이다.

[0004] 따라서, 답변 전체에 대한 내용에 대한 정보를 반영하여 질의에 대한 답변을 효과적으로 제공하기 위한 기술개발이 요구되고 있다.

선행기술문헌

특허문헌

[0005] (특허문헌 0001) 한국공개특허 제2020-0023620호

발명의 내용

해결하려는 과제

[0006] 일 실시 예에 의하면, 질의에 대한 답변을 제공하는 방법 및 이를 수행하는 전자 장치가 제공될 수 있다.

[0007] 보다 상세하게는 질의의 토큰 별 중요도에 기초하여, 질의에 대한 답변을 제공하는 방법 및 이를 수행하는 전자 장치가 제공될 수 있다.

과제의 해결 수단

[0008] 상술한 기술적 과제를 달성하기 위한 기술적 수단으로서, 일 실시 예에 의하면, 전자 장치가 사용자의 질의에 대한 답변을 제공하는 방법은 상기 사용자로부터 질의를 획득하는 단계; 상기 질의가 입력되면, 상기 질의의 토큰 별 중요도를 결정하고, 상기 결정된 상기 질의의 토큰 별 중요도에 기초하여, 상기 질의 및 상기 질의에 대한 전체 답변의 유사도를 출력하는 제1 신경망 모델로부터 상기 질의의 토큰 별 중요도를 획득하는 단계; 및 상기 질의 및 상기 획득된 질의의 토큰 별 중요도가 입력되면 상기 질의에 대한 답변을 출력하는 제2 신경망 모델에 상기 질의 및 상기 획득된 질의의 토큰 별 중요도를 입력함으로써, 상기 질의에 대한 응답을 생성하는 단계; 를 포함할 수 있다.

[0009] 상술한 기술적 과제를 달성하기 위한 기술적 수단으로서 또 다른 실시 예에 의하면, 사용자의 질의에 대한 답변을 제공하는 전자 장치에 있어서, 네트워크 인터페이스; 하나 이상의 인스트럭션을 저장하는 메모리; 및 상기 하나 이상의 인스트럭션을 실행하는 적어도 하나의 프로세서; 를 포함하고, 상기 적어도 하나의 프로세서는 상기 사용자로부터 질의를 획득하고, 상기 질의가 입력되면, 상기 질의의 토큰 별 중요도를 결정하고, 상기 결정된 상기 질의의 토큰 별 중요도에 기초하여, 상기 질의 및 상기 질의에 대한 전체 답변의 유사도를 출력하는 제1 신경망 모델로부터 상기 질의의 토큰 별 중요도를 획득하고, 상기 질의 및 상기 획득된 질의의 토큰 별 중요도가 입력되면 상기 질의에 대한 답변을 출력하는 제2 신경망 모델에 상기 질의 및 상기 획득된 질의의 토큰 별 중요도를 입력함으로써, 상기 질의에 대한 응답을 생성하는, 전자 장치가 제공될 수 있다.

[0010] 상술한 기술적 과제를 달성하기 위한 기술적 수단으로서 또 다른 실시 예에 의하면, 전자 장치가 사용자의 질의에 대한 답변을 제공하는 방법에 있어서, 상기 사용자로부터 질의를 획득하는 단계; 상기 질의가 입력되면, 상기 질의의 토큰 별 중요도를 결정하고, 상기 결정된 상기 질의의 토큰 별 중요도에 기초하여, 상기 질의 및 상기 질의에 대한 전체 답변의 유사도를 출력하는 제1 신경망 모델로부터 상기 질의의 토큰 별 중요도를 획득하는 단계; 및 상기 질의 및 상기 획득된 질의의 토큰 별 중요도가 입력되면 상기 질의에 대한 답변을 출력하는 제2 신경망 모델에 상기 질의 및 상기 획득된 질의의 토큰 별 중요도를 입력함으로써, 상기 질의에 대한 응답을 생성하는 단계; 를 포함하는, 방법을 수행하도록 하는 프로그램이 저장된 컴퓨터로 읽을 수 있는 기록 매체가 제공될 수 있다.

발명의 효과

[0011] 일 실시 예에 의하면 질의에 대한 답변을 효과적으로 제공할 수 있다.

[0012] 일 실시 예에 의하면 질의에 대한 답변 전체 내용에 대한 정보를 반영하는, 질의에 대한 답변을 제공하는 방법이 제공될 수 있다.

도면의 간단한 설명

[0013] 도 1은 일 실시 예에 따라 전자 장치가 사용자 질의에 대한 답변을 제공하는 과정을 개략적으로 설명하기 위한 도면이다.

도 2는 일 실시 예에 따른 전자 장치가 사용자의 질의에 대한 답변을 제공하는 방법의 흐름도이다.

도 3은 또 다른 실시 예에 따라 전자 장치가 복수의 신경망 모델을 이용하여 사용자 질의에 대한 답변을 제공하는 과정을 설명하기 위한 도면이다.

도 4는 일 실시 예에 따른 전자 장치의 블록도이다.

발명을 실시하기 위한 구체적인 내용

[0014] 본 명세서에서 사용되는 용어에 대해 간략히 설명하고, 본 개시에 대해 구체적으로 설명하기로 한다.

[0015] 본 개시에서 사용되는 용어는 본 개시에서의 기능을 고려하면서 가능한 현재 널리 사용되는 일반적인 용어들을 선택하였으나, 이는 당 분야에 종사하는 기술자의 의도 또는 판례, 새로운 기술의 출현 등에 따라 달라질 수 있다. 또한, 특정한 경우는 출원인이 임의로 선정한 용어도 있으며, 이 경우 해당되는 발명의 설명 부분에서 상세히 그 의미를 기재할 것이다. 따라서 본 개시에서 사용되는 용어는 단순한 용어의 명칭이 아닌, 그 용어가 가지는 의미와 본 개시의 전반에 걸친 내용을 토대로 정의되어야 한다.

[0016] 명세서 전체에서 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있음을 의미한다. 또한, 명세서에 기재된 "...부", "모듈" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어 또는 소프트웨어로 구현되거나 하드웨어와 소프트웨어의 결합으로 구현될 수 있다.

[0017] 아래에서는 첨부한 도면을 참고하여 본 개시의 실시 예에 대하여 본 개시가 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 상세히 설명한다. 그러나 본 개시는 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시 예에 한정되지 않는다. 그리고 도면에서 본 개시를 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

[0018] 도 1은 일 실시 예에 따라 전자 장치가 사용자 질의에 대한 답변을 제공하는 과정을 개략적으로 설명하기 위한 도면이다.

[0019] 일 실시 예에 의하면, 전자 장치(1000)는 질의(102)를 획득하고, 획득된 질의(102)에 대한 응답으로써 답변(104)을 제공할 수 있다. 일 실시 예에 의하면, 전자 장치(1000)는 질의에 대한 답변을 생성하기 위해 서버(2000) 또는 다른 전자 장치(1002)와 연동함으로써 질의에 대한 답변(104)으 제공할 수도 있다. 일 실시 예에 의하면, 전자 장치(1000)는 적어도 하나의 신경망 모델을 이용하여 질의(102)에 대한 답변(104)을 제공할 수 있다. 일 실시 예에 의하면 전자 장치(1000)는 전자 장치(1002)로부터 제1 신경망 모델의 출력 값 중 일부를 획득하고, 획득된 제1 신경망 모델의 출력 값을 이용하여 질의(102)에 대한 답변(104)을 제공할 수 있다.

[0020] 그러나 또 다른 실시 예에 의하면, 전자 장치(1000)는 전자 장치(1002)로부터 제1 신경망 모델(162) 또는 제1 신경망 모델(162)의 출력 값 중 일부를 획득하고, 획득된 제1 신경망 모델(162) 또는 제1 신경망 모델(162)의 출력 값 및 제2 신경망 모델(134)을 이용하여 질의(102)에 대한 답변(104)을 제공할 수도 있다. 또 다른 실시 예에 의하면, 전자 장치(1000)는 서버(2000)로부터 제1 신경망 모델(162) 또는 제1 신경망 모델(162)의 출력 값 중 적어도 하나를 획득하고, 서버(2000)로부터 획득된 제1 신경망 모델에 대한 정보를 이용하여 질의(102)에 대한 답변(104)을 제공할 수도 있다.

[0021] 일 실시 예에 의하면, 전자 장치(1000)가 이용하는 신경망 모델은 시퀀스 투 시퀀스 모델 구조의 신경망 모델일 수 있다. 일 실시 예에 의하면, 전자 장치(1000)가 이용하는 신경망 모델은 심층 신경망(DNN:Deep Neural Network)을 포함할 수도 있다. 예를 들어, 전자 장치(1000)가 질의에 대한 답변을 제공하기 위해 이용하는 신경망 모델은 CNN (Convolutional Neural Network), DNN (Deep Neural Network), RNN (Recurrent Neural Network), RBM (Restricted Boltzmann Machine), DBN (Deep Belief Network), BRDNN(Bidirectional Recurrent

Deep Neural Network) 또는 심층 Q-네트워크 (Deep Q-Networks) 등이 있으나, 역시 이에 한정되는 것은 아니다.

[0022] 일반적으로 종래의 시퀀스 투 시퀀스 모델은, 어느 질의에도 답변이 될 수 있는 일반적인 답변(복수의 질의에 대한 공통 답변이 될 수 있는 응답)을 제공하는 한계가 있다. 이는 일반적인 답변 생성기의 경우, 디코딩 시 토큰 단위 답변 토큰 생성을 통하여, 매 순간 마다 정답 답변 전체 내용에 대한 정보가 반영되지 않기 때문인데, 따라서, 질의에 대한 답변 토큰 생성 시, 매 순간마다 정답 답변 전체에 대한 내용 반영을 통하여 질의에 대한 답변을 효과적으로 생성하기 위한 기술 개발이 요구되고 있다.

[0023] 본 개시에 따른 전자 장치(1000)는 전자 장치(1002)로부터 질의(102)에 대한 토큰 별 중요도를 획득하고, 획득된 질의의 토큰 별 중요도 값을 이용하여 질의에 대한 답변을 제공함으로써, 질의에 대한 답변(104)을 효과적으로 제공할 수 있다. 예를 들어, 전자 장치(1002)는 네트워크 인터페이스(140), 프로세서(150) 및 메모리(160)를 포함할 수 있다. 전자 장치(1002)내 프로세서(150)는 메모리(160)내 하나 이상의 인스트럭션을 수행함으로써, 질의(102)와 예상 답변 사이의 관계를 추출할 수 있다.

[0024] 보다 상세하게는, 프로세서(150)는 메모리(160)에 저장된 제1 신경망 모델(162)을 이용하여 질의 토큰 별 중요도를 결정할 수 있다. 일 실시 예에 의하면, 제1 신경망 모델은 질의 토큰 별 중요도에 기초하여 질의 및 상기 질의에 대한 전체 답변의 유사도를 출력하도록 학습된 신경망 모델일 수 있다. 전자 장치(1002)는 소정의 대화쌍 데이터를 포함하는 데이터 베이스를 이용하여 제1 신경망 모델을 미리 학습시킬 수 있다. 전자 장치(1000)는 학습된 제1 신경망 모델에 질의가 입력됨에 따라, 상기 제1 신경망 모델로부터 질의의 토큰 별 중요도를 획득할 수 있다. 전자 장치(1002) 내 프로세서(150)는 제1 신경망 모델로부터 결정된 질의의 토큰 별 중요도에 대한 정보를, 네트워크 인터페이스(140)를 통하여 전자 장치(1000)로 전송할 수 있다.

[0025] 또 다른 실시 예에 의하면, 전자 장치(1002)는 제1 신경망 모델(162)에 대한 정보 전체를 전자 장치(1000)로 전송할 수도 있다. 이 경우, 전자 장치(1000)는 질의(102)를 획득하고, 획득된 질의(102)를 제1 신경망 모델(162)에 입력하며, 제1 신경망 모델(162)로부터 질의의 토큰 별 중요도 값을 획득하고, 획득된 질의의 토큰 별 중요도 값을 이용하여 질의에 대한 답변을 효과적으로 제공할 수 있다.

[0026] 일 실시 예에 의하면, 전자 장치(1000)는 전자 장치(1002)에 대응되는 구성을 포함할 수 있다. 예를 들어, 전자 장치(1000)는 네트워크 인터페이스(110), 프로세서(120) 및 메모리(130)를 포함할 수 있다. 프로세서(120)는 메모리(130)내 하나 이상의 인스트럭션을 수행함으로써, 질의에 대한 답변을 제공하기 위한 방법을 수행한다. 예를 들어, 프로세서(120)는 네트워크 인터페이스(110)를 제어함으로써, 전자 장치(1002)로부터 제1 신경망 모델(162)에 대한 정보 또는 제1 신경망 모델(162)로부터 질의의 토큰 별 중요도에 대한 정보를 획득할 수 있다. 프로세서(120)는 전자 장치(1002)로부터 획득되는 제1 신경망 모델(162) 또는 제1 신경망 모델(162)로부터 획득 가능한 질의의 토큰 별 중요도 값을 이용하여 질의에 대한 답변을 제공할 수 있다.

[0027] 보다 상세하게는, 전자 장치(1000)는 질의의 토큰 별 중요도에 대한 정보를 전자 장치(1000)로부터 획득하고, 질의(102) 및 질의의 토큰 별 중요도 값을 제2 신경망 모델(134)에 입력함으로써, 질의에 대한 답변(104)을 생성할 수 있다. 그러나 또 다른 실시 예에 의하면, 전자 장치(1000)는 전자 장치(1002)로부터 질의의 토큰 별 중요도 값을 획득하는 것이 아니고, 전자 장치(1000)가 학습 시킨 제1 신경망 모델에 대한 정보를 획득한 후, 전자 장치(1000)에 저장된 제1 신경망 모델을 이용하여, 질의에 대한 토큰 별 중요도 값을 직접 획득할 수도 있다.

[0028] 또 다른 실시 예에 의하면, 전자 장치(1000)는 제1 신경망 모델(162)의 구성 중 질의 토큰 중요도 결정부(132)에 대한 정보만을 획득함으로써, 제1 신경망 모델(162)의 일부 구성에 대한 모델만을 전자 장치(1000)에 저장해 둘 수도 있다. 전자 장치(1000)는 메모리(130)에 미리 저장된 질의 토큰 중요도 결정부(132) 및 제2 신경망 모델(134)을 이용하여 질의에 대한 답변(104)을 효과적으로 제공할 수도 있다.

[0029] 일 실시 예에 의하면, 전자 장치(1000)는 소정의 네트워크를 통하여 서버(2000)와 연결될 수 있다. 일 실시 예에 의하면, 전자 장치(1000)는 소정의 네트워크를 통하여 전자 장치(1002) 또는 서버(2000)와 연결됨으로써, 사용자 질의에 대한 답변을 제공하기 위해 필요한 신경망 모델, 질의 토큰 정보, 질의의 토큰 별 중요도 정보 등을 서로 송수신할 수도 있다.

[0030] 즉, 전자 장치(1002)는 질의 토큰 및 전체 답변 과의 관계를 추출하도록, 제1 신경망 모델을 학습시키고, 제1 신경망 모델을 이용하여 질의 및 전체 답변 과의 관계(예컨대 질의 토큰 별 중요도)를 추출할 수 있다. 전자 장치(1002)는 질의 토큰 별 중요도 값을 전자 장치(1000)로 전송할 수도 있지만, 제1 신경망 모델 자체에 대한 정

보 또는 제1 신경망 모델 내 질의 토큰 중요도 결정부와 같은 일부 모델에 대한 정보를 직접 전자 장치(1000)로 전송할 수도 있다. 전자 장치(1000)는 전자 장치(1002)로부터 획득된 질의 토큰 별 중요도 값 또는 전자 장치(1002)로부터 획득된 질의 토큰 중요도 결정부(134) 또는 제1 신경망 모델(162)을 직접 운용함으로써 스스로 결정한 질의 토큰 별 중요도 값을 이용하여 질의에 대한 답변을 제공할 수 있다.

[0031] 도 2는 일 실시 예에 따른 전자 장치가 사용자의 질의에 대한 답변을 제공하는 방법의 흐름도이다.

[0032] S210에서, 전자 장치(1000)는 사용자로부터 질의를 획득할 수 있다. 일 실시 예에 의하면, 전자 장치(1000)는 사용자 입력에 기초하여 텍스트 형태(예컨대 단어 또는 문장)로 질의를 획득할 수 있다. 그러나 또 다른 실시 예에 의하면, 전자 장치(1000)는 이미지 또는 오디오 형태로 질의를 획득할 수도 있음은 물론이다.

[0033] S220에서, 전자 장치(1000)는 질의가 입력되면, 질의의 토큰 별 중요도를 결정하고, 상기 결정된 질의의 토큰 별 중요도에 기초하여 질의 및 상기 질의에 대한 전체 답변의 유사도를 출력하는 제1 신경망 모델로부터 질의 토큰 별 중요도를 획득할 수 있다. 예를 들어, 전자 장치(1000)는 제1 신경망 모델을 이용하여 질의의 토큰 별 중요도를 결정할 수 있다. 토큰은 질의에 대하여 문법적으로 더 이상으로 나눌 수 없는 언어 요소를 의미할 수 있다. 전자 장치(1000) 또는 전자 장치(1002)는 제1 신경망 모델을 이용하여 질의를 토큰화함으로써 질의 별 토큰을 획득할 수 있다.

[0034] 일 실시 예에 의하면, 제1 신경망 모델은 소정의 질의 및 상기 질의에 대한 답변을 포함하는 대화쌍들이 저장된 데이터 베이스로부터, 상기 질의를 획득하고, 상기 질의를 인코딩함으로써 상기 질의에 대한 표현 벡터를 생성하는 질의 표현 벡터 인코더 및 상기 질의에 대한 답변을 획득하고, 상기 획득된 답변 전체를 인코딩함으로써 상기 답변 전체에 대한 표현 벡터를 생성하는 답변 전체 표현 벡터 인코더를 포함할 수 있다. 즉, 본 개시에 따른 제1 신경망 모델은 2이상의 인코더를 포함할 수 있다.

[0035] 일 실시 예에 의하면, 제1 신경망 모델은 질의에 대한 표현 벡터인 질의 표현 벡터와 답변 전체에 대한 표현 벡터인 답변 전체 표현 벡터가 서로 유사해지도록 학습될 수 있다. 일 실시 예에 의하면, 질의 표현 벡터 및 답변 전체 표현 벡터가 유사하면(예컨대 유사도가 1에 가까우면) 질의 표현 벡터에 대응되는 질의 및 답변 전체 표현 벡터에 대응되는 답변은 언어적으로 의미 있는 질의 및 답변 쌍(예컨대 실제 사람들이 나누는 유의미한 실제 대화쌍)을 나타낼 수 있다.

[0036] 그러나 질의 표현 벡터 및 답변 전체 표현 벡터가 유사하지 않을 경우, 질의 표현 벡터에 대응되는 질의 및 답변 전체 표현 벡터에 대응되는 답변은 언어적으로 의미가 없는 질의 및 답변 쌍(예컨대 실제 사람들이 나누는 대화가 아닌 거짓 대화쌍으로, 언어적으로 무의미한 대화쌍)을 나타낼 수 있다.

[0037] 또한, 일 실시 예에 의하면, 상기 제1 신경망 모델 내 질의 표현 벡터 인코더는 상기 질의를 토큰화함으로써 질의에 대한 적어도 하나의 토큰을 생성하고, 상기 생성된 토큰 별 표현 벡터를 생성하는 질의 토큰 표현 벡터 생성부 및 상기 생성된 토큰 별 표현 벡터의 중요도를 결정하는 질의 토큰 중요도 결정부를 포함할 수 있다. 예를 들어, 질의 토큰 표현 벡터 생성부는 질의를 토큰화하고, 질의의 토큰들을 인코딩함으로써 표현 벡터를 생성할 수 있다. 일 실시 예에 의하면, 질의 토큰 중요도 결정부는 토큰 별 표현 벡터의 중요도를 결정할 수 있다.

[0038] 보다 상세하게는, 질의 표현 벡터 인코더는 결정된 질의 토큰 중요도에 기초하여 토큰 별 표현 벡터를 가중합함으로써 질의에 대한 표현 벡터를 생성할 수 있다. 제1 신경망 모델은 상술한 방법에 따라 생성된 질의 표현 벡터가 답변 전체 표현 벡터가 유사해지도록 신경망 모델 내 레이어들 및 상기 레이어들 사이의 연결 강도에 관한 가중치를 수정 및 갱신함으로써 학습될 수 있다. 제1 신경망 모델이 질의 표현 벡터가 답변 전체 표현 벡터가 유사해지도록 학습이 완료되면, 전자 장치(1000)는 제1 신경망 모델 내 질의 토큰 중요도 결정부에서 출력되는 질의 토큰 별 중요도 값에 기초하여 질의에 대한 답변 생성에 이용할 수 있다.

[0039] 상술한 제1 신경망 모델 내 상기 질의 표현 벡터 인코더 및 상기 답변 전체 표현 벡터 인코더가, 상기 질의에 대한 표현 벡터 및 상기 답변 전체에 대한 표현 벡터가 서로 유사해지도록 미리 학습되면, 질의 토큰 중요도 결정부는, 답변 전체에 대한 내용이 충분히 반영된, 토큰 별 중요도를 결정할 수 있게 된다. 일 실시 예에 의하면, 제1 신경망 모델 내 질의 토큰 중요도 결정부가 결정하는 중요도는 질의 토큰 별 전체 답변에 대한 유사도를 의미할 수도 있다. 상술한, 제1 신경망 모델의 학습 과정은 전자 장치(1000)에 의해서 수행될 수 있다. 그러나 또 다른 실시 예에 의하면, 제1 신경망 모델의 학습은 전자 장치(1000)와 연결된 서버(2000) 또는 전자 장치(1000)와 연결된 다른 전자 장치(1002)에 의해서도 수행될 수 있음은 물론이다.

[0040] S230에서, 전자 장치(1000)는 질의 및 상기 획득된 질의의 토큰 별 중요도가 입력되면 상기 질의에 대한 답변을 출력하는 제2 신경망 모델에 상기 질의 및 상기 획득된 질의의 토큰 별 중요도를 입력함으로써, 상기 질의에 대

한 응답을 생성할 수 있다. 일 실시 예에 의하면, 제2 신경망 모델은 시퀀스 투 시퀀스 모델 구조의 신경망 모델로써 인코더 및 디코더를 포함할 수 있다. 전자 장치(1000)가 이용하는 제2 신경망 모델은, 제1 신경망 모델 내 일부의 신경망 구조(예컨대 질의 토큰 중요도 결정부)로부터 획득되는 질의 토큰 별 중요도 값 및 질의가 입력되면, 답변 전체에 대한 질의 토큰 별 중요도가 반영된, 답변을 생성할 수 있다.

[0041] 일 실시 예에 의하면, 상기 제2 신경망 모델은 상기 사용자로부터 질의가 입력되면 상기 질의를 인코딩하는 인코더 및 상기 인코더의 출력 값 및 상기 인코더의 출력 값에 상기 질의의 토큰 별 중요도를 반영함으로써 결정되는 인코더 출력 변환 값을 획득하고, 상기 획득된 인코더의 출력 값 및 상기 인코더 출력 변환 값을 디코딩하는 디코더를 포함할 수 있다. 일 실시 예에 의하면 제2 신경망 모델 내 디코더는 인코더의 출력 값 및 상기 인코더 출력 변환 값을 순차로 또는 함께 획득할 수도 있다.

[0042] 예를 들어, 인코더는 질의가 입력되면 질의를 인코딩함으로써 인코더 출력 값을 생성할 수 있다. 제1 신경망 모델은 인코더의 출력 값에, 질의 토큰 별 중요도 값을 토큰 별로 스칼라배 하여, 전체 답변과의 관계가 반영된 중요성이 곱해진 인코더 출력 변환 값을 결정할 수 있다. 제1 신경망 모델은 상술한 인코더의 출력 값 및 인코더 출력 변환 값을 디코더에 입력함으로써 디코딩하고, 디코딩 결과에 기초하여 질의에 대한 답변을 생성할 수 있다.

[0043] 도 2에는 도시되지 않았지만, 전자 장치(1000)는 생성된 응답을 사용자의 질의에 대한 답변으로 제공할 수 있다. 일 실시 예에 의하면, 전자 장치(1000)는 전자 장치에 포함된 스피커 또는 디스플레이를 이용하여, 생성된 답변을 이미지, 비디오 또는 오디오 형태로 출력할 수 있다.

[0044] 도 3은 또 다른 실시 예에 따라 전자 장치가 복수의 신경망 모델을 이용하여 사용자 질의에 대한 답변을 제공하는 과정을 설명하기 위한 도면이다.

[0045] 도 3을 참조하여 전자 장치(1000)가 복수의 신경망 모델을 이용하여 질의에 대한 답변을 제공하는 과정을 설명하기로 한다. 일 실시 예에 의하면, 전자 장치(1000)는 제1 신경망 모델(310) 및 제2 신경망 모델(330)을 이용하여 질의 토큰 별 중요도에 기초하여, 질의에 대한 답변을 제공할 수 있다. 일 실시 예에 의하면, 제1 신경망 모델(310)은 질의와 답변의 관계를 추출하는 관계 추출기, 제2 신경망 모델은, 관계 추출기로부터 추출된 질의 및 답변 사이의 관계에 기초하여, 질의에 대한 답변을 제공하는 답변 생성기에 대응될 수 있다.

[0046] 일 실시 예에 의하면, 전자 장치(1000)가 이용하는 제1 신경망 모델(310) 및 제2 신경망 모델(330)은 전자 장치의 메모리에 저장될 수 있다. 그러나 또 다른 실시 예에 의하면, 제1 신경망 모델(310)내 질의 토큰 중요도 결정부 및 제2 신경망 모델에 대한 정보가 전자 장치의 메모리에 저장될 수도 있다. 그러나 또 다른 실시 예에 의하면, 전자 장치(1000)의 메모리에는 제2 신경망 모델(330)에 대한 정보만이 저장될 수도 있다.

[0047] 일 실시 예에 의하면, 제1 신경망 모델(310)은 소정의 대화쌍들이 저장된 데이터 베이스(312), 질의 토큰 표현 벡터 생성기(314), 질의 토큰 중요도 계산기(316), 답변 전체 표현 벡터 생성기(318) 및 유사도 검사부(320)를 포함할 수 있다. 일 실시 예에 의하면, 데이터 베이스(312)는 제1 신경망 모델(310)에 포함되지 않고, 제1 신경망 모델에 대화쌍 학습 데이터를 제공하기 위한 외부에 마련될 수도 있다.

[0048] 일 실시 예에 의하면, 질의 토큰 표현 벡터 생성기(314)는 인코더로써 상술한 질의 토큰 표현 벡터 생성부에 대응될 수 있고, 질의 토큰 중요도 계산기는 상술한 질의 토큰 중요도 결정부에 대응될 수 있다. 또한, 일 실시 예에 의하면, 답변 전체 표현 벡터 생성기(318)은 상술한 답변 전체 표현 벡터 인코더에 대응될 수 있다. 전자 장치(1000)가 이용하는 제1 신경망 모델(310)은 전자 장치(1000) 또는 전자 장치와 연결된 전자 장치(1002)에 의해 학습될 수 있다. 일 실시 예에 의하면, 제1 신경망 모델 내 질의 토큰 표현 벡터 생성기(314)는 대화쌍 데이터 베이스(312)로부터 질의 학습 데이터를 획득하고, 답변 전체 표현 벡터 생성기(318)은 대화쌍 데이터 베이스(312)로부터 학습 답변 데이터를 획득한다.

[0049] 질의 토큰 표현 벡터 생성기(314)는 질의 학습 데이터를 인코딩함으로써 질의 표현 벡터를 생성하고, 답변 전체 표현 벡터 생성기(318)은 학습 답변 데이터를 인코딩함으로써 답변 전체에 대한 표현 벡터를 생성한다. 질의 토큰 중요도 계산기(316)은 질의 토큰 별 표현 벡터에 대한 중요도를 결정하고, 결정된 중요도를 가중치로써, 질의 토큰 별 표현 벡터에 반영한다. 질의 토큰 표현 벡터 생성기(314)은 질의 토큰 표현 벡터에 질의 토큰 별 중요도를 가중치로 반영함으로써 전체 질의에 대한 표현 벡터를 생성할 수 있다.

[0050] 유사도 검사부(320)는 질의 표현 벡터 및 답변 전체 표현 벡터의 유사도를 비교할 수 있다. 일 실시 예에 의하면, 유사도 검사부(320)는 질의 표현 벡터 및 답변 전체 표현 벡터의 유사도 값이 1인 경우, 질의 표현 벡터 및 답변 전체 표현 벡터가 유사한 것으로 결정하고, 현재 입력된 질의 및 답변은 실제 대화쌍을 구성하는 것으로

결정할 수 있다. 그러나 또 다른 실시 예에 따라 유사도 검사부(320)가 질의 표현 벡터 및 답변 전체 표현 벡터의 유사도 값을 0으로 식별하는 경우, 현재 제1 신경망 모델로 입력된 질의 및 답변은 거짓 대화쌍을 구성하는 것으로 결정될 수도 있다.

[0051] 제1 신경망 모델(310)은 질의 표현 벡터 및 답변 전체 표현 벡터가 유사해지도록 학습됨에 따라, 각 질의의 각 토큰의 중요도를 계산할 때, 실제 정답 답변과 유사한 표현 벡터가 최종적으로 생성될 수 있도록, 질의 표현 벡터 및 유사도 검사부가 학습될 수 있다. 제1 신경망 모델(310)의 학습 과정이 완료되면, 이후에는 답변이 입력되지 않고, 질의에 대한 정보만이 입력되더라도, 답변이 고려된 각 질의의 토큰 별 중요도가 출력될 수 있다.

[0052] 제1 신경망 모델(310)에서 결정된 질의 토큰 별 중요도는 제2 신경망 모델(330)로 입력될 수 있다. 예를 들어, 제2 신경망 모델(330)은 제1 신경망 모델(310)에서 출력된 질의 토큰 별 중요도(334) 및 사용자 질의에 대응되는 사용자 입력 문장(332)을 획득할 수 있다. 제2 신경망 모델(330)은 사용자 입력 문장(332) 및 질의 토큰 중요도(334)를 답변 생성기(336)에 입력한다.

[0053] 일 실시 예에 의하면, 답변 생성기(336)는 제1 신경망 모델에 대응될 수 있다. 일 실시 예에 의하면 답변 생성기(336)는 인코더 및 디코더를 포함할 수 있다. 예를 들어, 답변 생성기(336)내 인코더는 질의를 인코딩함으로써 제1 인코딩 출력 값을 결정할 수 있다. 또한, 답변 생성기(336)는 제1 인코딩 출력 값에, 상기 질의의 토큰 별 중요도를, 토큰 별로 스칼라배 함으로써 인코더 출력 변환 값(예컨대 제2 인코딩 출력 값)을 결정할 수 있다. 답변 생성기(336)는 제1 인코딩 출력 값을 디코더의 제1 입력으로, 제2 인코딩 출력 값을 디코더의 제2 입력으로 사용할 수 있다.

[0054] 답변 생성기(336)내 디코더는 제1 인코딩 출력 값 및 제2 인코딩 출력 값을 디코딩함으로써, 질의에 대한 답변 문장(338)을 생성할 수 있다. 일 실시 예에 의하면, 답변 생성기(336)내 디코더는 제1 인코딩 출력 값 및 제2 인코딩 출력 값을 순차로 획득할 수 있다. 그러나 또 다른 실시 예에 의하면, 답변 생성기(336)내 디코더는 제1 인코딩 출력 값 및 제2 인코딩 출력 값을 함께 획득할 수도 있다.

[0055] 도 4는 일 실시 예에 따른 전자 장치의 블록도이다.

[0056] 일 실시 예에 의하면, 전자 장치(1000)는 프로세서(1300), 네트워크 인터페이스(1500) 및 메모리(1700)를 포함할 수 있다. 그러나 상술한 구성에 한정되는 것은 아니고, 전자 장치(1000)는 응답 내지 답변을 출력하기 위한 스피커, 디스플레이 등을 더 포함할 수도 있다. 일 실시 예에 의하면, 프로세서(1300)는 메모리(1700)에 저장된 하나 이상의 인스트럭션을 실행함으로써 질의에 대한 답변을 제공하는 과정을 수행한다.

[0057] 일 실시 예에 의하면, 네트워크 인터페이스(1500)는 전자 장치(1000)가 다른 장치(미도시) 및 서버(미도시)와 통신을 하게 하는 하나 이상의 구성요소를 포함할 수 있다. 다른 장치(미도시)는 전자 장치(1000)와 같이 질의에 대한 응답을 제공하는 장치로써, 제1 신경망 모델을 학습시키기 위한 장치일 수 있다. 네트워크 인터페이스(1500)는 다른 전자 장치(1002)로부터 제1 신경망 모델에 대한 정보, 제1 신경망 모델의 질의 토큰 중요도 결정부에 대한 정보, 질의의 토큰 별 중요도 값을 획득할 수 있다.

[0058] 메모리(1700)는, 프로세서(1300)의 처리 및 제어를 위한 프로그램을 저장할 수 있고, 전자 장치(1000)로 입력되는 질의, 전자 장치(1000)로부터 출력되는 데이터로써 응답 또는 정답 정보를 저장할 수도 있다. 또한, 메모리(1700)는 상술한 제1 신경망 모델, 제2 신경망 모델, 제1 신경망 모델 내 질의 토큰 중요도 결정부에 대한 정보, 질의 토큰 중요도 값에 대한 정보 등을 더 저장할 수도 있다. 또한, 메모리(1700)는 각 신경망 모델을 학습시키기 위한 데이터 베이스, 데이터 베이스에 저장될 학습 데이터에 대한 정보를 더 저장할 수도 있다.

[0059] 메모리(1700)는 복수의 신경망 모델 내 레이어들 및 상기 레이어들의 연결 강도에 관한 가중치(WEIGHT) 값들을 저장할 수도 있다. 메모리(1700)에 저장된 가중치 값들은 미리 설정된 주기에 따라 수정 및 갱신될 수 있다. 즉, 메모리(1700)에는 전자 장치(1000)가 다른 전자 장치(1002)로부터 획득된 데이터(예컨대 질의 토큰 별 중요도)에 기초하여 새로 추가한 질의 토큰 중요도 결정부, 갱신된 제2 신경망 모델에 대한 정보, 또는 갱신된 제1 신경망 모델에 대한 정보들이 추가 및 갱신될 수 있다.

[0060] 메모리(1700)는 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(RAM, Random Access Memory) SRAM(Static Random Access Memory), 롬(ROM, Read-Only Memory), 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.

[0061] 프로세서(1300)는 메모리에 저장된 하나 이상의 인스트럭션을 수행함으로써, 질의를 획득하고, 제1 신경망 모델 또는 제2 신경망 모델 중 적어도 하나를 이용하여 질의에 대한 답변을 결정한다. 예를 들어, 프로세서는 상기 질의의 토큰 별 중요도를 결정하고, 상기 결정된 상기 질의의 토큰 별 중요도에 기초하여, 상기 질의 및 상기 질의에 대한 전체 답변의 유사도를 출력하는 제1 신경망 모델로부터 상기 질의의 토큰 별 중요도를 획득하고, 상기 질의 및 상기 획득된 질의의 토큰 별 중요도가 입력되면 상기 질의에 대한 답변을 출력하는 제2 신경망 모델에 상기 질의 및 상기 획득된 질의의 토큰 별 중요도를 입력함으로써, 상기 질의에 대한 응답을 생성할 수 있다.

[0062] 또한, 상기 적어도 하나의 프로세서(1300)는 상기 생성된 응답을 상기 사용자의 질의에 대한 답변으로 제공할 수 있다. 또한, 프로세서(1300)는 도 1 내지 3에 상술한 질의에 대한 답변을 제공하는 방법을 수행하기 위해 네트워크 인터페이스(1500) 및 메모리(1700)의 전반적인 동작을 제어한다.

[0063] 일 실시 예에 의하면, 프로세서(1300)는 제1 신경망 모델이 소정의 질의 및 상기 질의에 대한 답변을 포함하는 대화쌍들이 저장된 데이터 베이스로부터, 상기 질의를 획득하고, 상기 질의를 인코딩함으로써 상기 질의에 대한 표현 벡터를 생성하도록 제어할 수 있다. 또한, 프로세서(1300)는 제1 신경망 모델이, 상기 질의에 대한 답변을 획득하고, 상기 획득된 답변 전체를 인코딩함으로써 상기 답변 전체에 대한 표현 벡터를 생성하도록 제어할 수 있다.

[0064] 또한, 프로세서(1300)는 제1 신경망 모델이 상기 질의를 토큰화함으로써 질의에 대한 적어도 하나의 토큰을 생성하고, 상기 생성된 토큰 별 표현 벡터를 생성하도록 하고, 상기 생성된 토큰 별 표현 벡터의 중요도를 결정하며, 상기 결정된 질의 토큰 중요도에 기초하여 상기 토큰 별 표현 벡터를 가중합함으로써 상기 질의에 대한 표현 벡터를 생성하도록, 제1 신경망 모델을 제어할 수 있다.

[0065] 또한, 프로세서(1300)는 상기 질의에 대한 표현 벡터 및 상기 답변 전체에 대한 표현 벡터가 서로 유사해지도록 질의 표현 벡터 인코더 및 상기 답변 전체 표현 벡터 인코더를 학습시킬 수 있다. 또한, 프로세서(1300)는 상기 질의에 대한 표현 벡터 및 상기 답변 전체에 대한 표현 벡터가 서로 유사해지도록 상기 질의 표현 벡터 인코더가 학습됨에 따라 상기 질의 표현 벡터 인코더 내 상기 질의 토큰 중요도 결정부가 상기 답변 전체에 대한 내용에 기초하여 상기 질의의 토큰 별 중요도를 결정하도록, 제1 신경망 모델을 학습시킬 수 있다.

[0066] 또한, 프로세서(1300)는 사용자로부터 질의가 입력되면 상기 질의를 인코딩하도록 제2 신경망 모델 내 인코더를 제어하고, 상기 인코더의 출력 값 및 상기 인코더의 출력 값에 상기 질의의 토큰 별 중요도를 반영함으로써 결정되는 인코더 출력 변환 값을 획득하며, 상기 획득된 인코더의 출력 값 및 상기 인코더 출력 변환 값을 디코딩하도록 제2 신경망 모델 내 디코더를 제어할 수 있다.

[0067] 일 실시예에 따른 전자 장치가 사용자의 질의에 대한 답변을 제공하는 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 본 발명을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다.

[0068] 또한, 전자 장치가 사용자의 질의에 대한 답변을 제공하는 방법을 수행하도록 하는 프로그램이 저장된 기록 매체를 포함하는 컴퓨터 프로그램 장치가 제공될 수 있다.

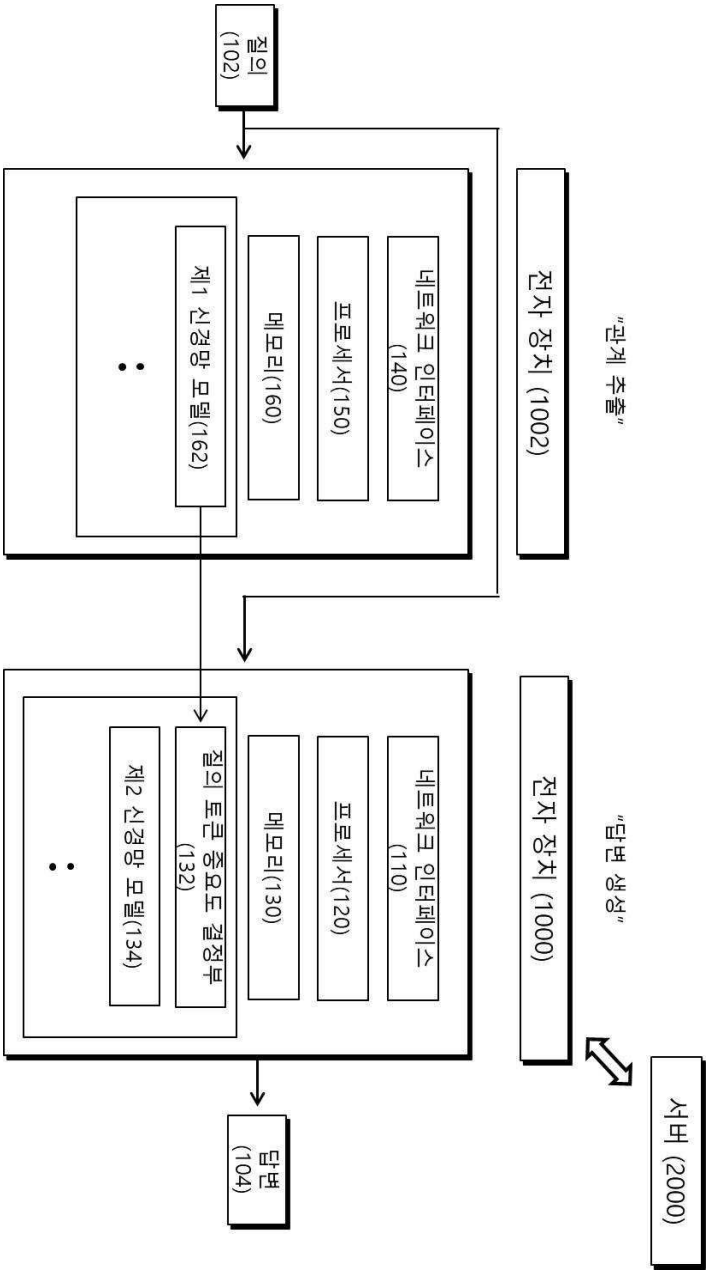
[0069] 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.

[0070] 일부 실시예는 컴퓨터에 의해 실행되는 프로그램 모듈과 같은 컴퓨터에 의해 실행 가능한 명령어를 포함하는 기록 매체의 형태로도 구현될 수 있다. 컴퓨터 판독 가능 매체는 컴퓨터에 의해 액세스될 수 있는 임의의 가용 매체일 수 있고, 휘발성 및 비휘발성 매체, 분리형 및 비분리형 매체를 모두 포함한다. 또한, 컴퓨터 판독가능 매체는 컴퓨터 저장 매체 및 통신 매체를 모두 포함할 수 있다. 컴퓨터 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현된 휘발성 및 비휘발성, 분리형 및 비분리형 매체를 모두 포함한다. 통신 매체는 전형적으로 컴퓨터 판독가능 명령어,

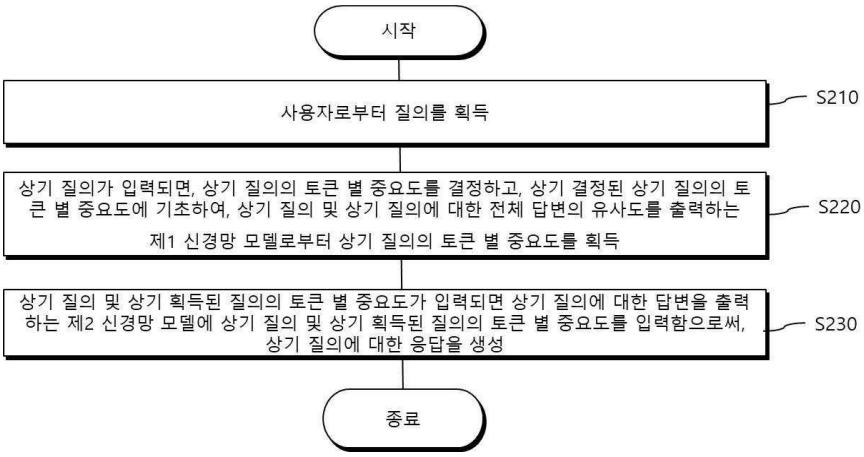
데이터 구조, 프로그램 모듈, 또는 반송파와 같은 변조된 데이터 신호의 기타 데이터, 또는 기타 전송 메커니즘을 포함하며, 임의의 정보 전달 매체를 포함한다. 또한, 일부 실시예는 컴퓨터에 의해 실행되는 컴퓨터 프로그램과 같은 컴퓨터에 의해 실행 가능한 명령어를 포함하는 컴퓨터 프로그램 또는 컴퓨터 프로그램 제품(computer program product)으로도 구현될 수 있다. 이상에서 본 발명의 실시예에 대하여 상세하게 설명하였지만 본 발명의 권리범위는 이에 한정되는 것은 아니고 다음의 청구범위에서 정의하고 있는 본 발명의 기본 개념을 이용한 당업자의 여러 변형 및 개량 형태 또한 본 발명의 권리범위에 속한다.

도면

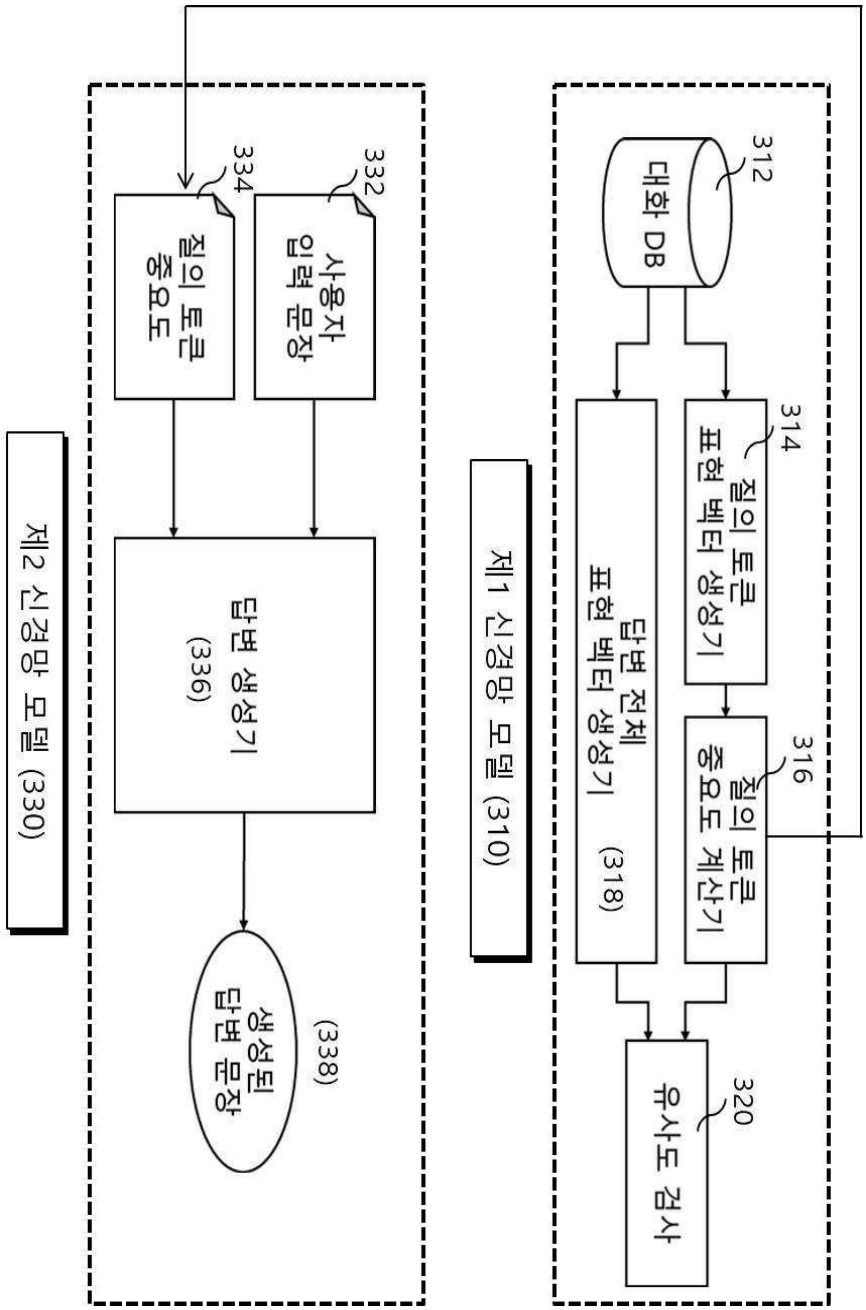
도면1



도면2


시작
사용자로부터 질의를 획득
S210
상기 질의가 입력되면, 상기 질의의 토큰 별 중요도를 결정하고, 상기 결정된 상기 질의의 토큰 별 중요도에 기초하여, 상기 질의 및 상기 질의에 대한 전체 답변의 유사도를 출력하는 제1 신경망 모델로부터 상기 질의의 토큰 별 중요도를 획득
S220
상기 질의 및 상기 획득된 질의의 토큰 별 중요도가 입력되면 상기 질의에 대한 답변을 출력하는 제2 신경망 모델에 상기 질의 및 상기 획득된 질의의 토큰 별 중요도를 입력함으로써, 상기 질의에 대한 응답을 생성
S230
종료

도면3


312
대화 DB
314
질의 토큰
표현 벡터 생성기
316
질의 토큰
중요도 계산기
답변 전체
표현 벡터 생성기
(318)
320
유사도 검사
제1 신경망 모델 (310)
332
사용자
입력 문장
334
질의 토큰
중요도
답변 생성기
(336)
생성된
답변 문장
(338)
제2 신경망 모델 (330)

도면4

