



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0106947
(43) 공개일자 2024년07월08일

<p>(51) 국제특허분류(Int. Cl.) G06F 16/332 (2019.01) G06F 16/31 (2019.01) G06F 16/33 (2019.01) G06F 40/157 (2020.01) G06F 40/279 (2020.01) G06N 3/0455 (2023.01) G06N 3/08 (2023.01)</p> <p>(52) CPC특허분류 G06F 16/3329 (2019.01) G06F 16/316 (2019.01)</p> <p>(21) 출원번호 10-2023-0138086 (22) 출원일자 2023년10월16일 심사청구일자 2023년10월16일</p> <p>(30) 우선권주장 1020220188021 2022년12월28일 대한민국(KR)</p>	<p>(71) 출원인 포항공과대학교 산학협력단 경상북도 포항시 남구 청암로 77 (지곡동)</p> <p>(72) 발명자 한옥신 경상북도 포항시 남구 청암로 77 김현지 경상북도 포항시 남구 청암로 77 강혁규 경상북도 포항시 남구 청암로 77</p> <p>(74) 대리인 특허법인이상</p>
--	--

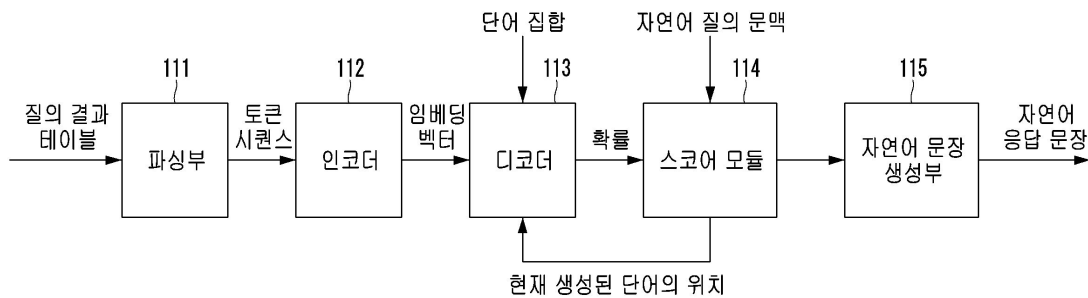
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 질의 검색 결과 테이블을 자연어 응답 문장으로 변환하는 장치, 이를 위한 방법, 및 상기 장치를 포함하는 대화형 인터페이스 장치

(57) 요약

프로세서를 포함하고, 상기 프로세서는, 자연어 질의 문장에 대한 검색 결과를 나타내는 테이블 데이터를 언어모델의 입력으로 입력 받아 임베딩 벡터들을 생성하고, 상기 임베딩 벡터들을 이용하여 생성될 자연어 응답 문장의 임의의 위치에 미리 결정된 단어들에 배치될 각각의 확률을 계산하고, 상기 계산된 확률들로 구성되는 확률분포, 및 상기 자연어 질의 문장을 이용하여 상기 확률분포를 조절하고, 상기 자연어 응답 문장의 모든 위치에 있어서, 상기 모든 위치 중 임의의 위치에 배치될 단어를, 상기 임의의 위치와의 쌍으로 생성되는 확률분포에서 가장 큰 확률을 갖는 단어로 결정하고, 그리고 상기 모든 위치에서 결정된 단어들에 나열하여 상기 테이블의 변환 결과의 문장인, 상기 자연어 응답 문장을 생성하는, 질의 검색 결과 테이블을 자연어 응답 문장으로 변환하는 장치를 공개한다.

대표도



(52) CPC특허분류

- G06F 16/3334 (2019.01)
- G06F 16/3337 (2019.01)
- G06F 16/3346 (2019.01)
- G06F 16/3347 (2019.01)
- G06F 40/157 (2020.01)
- G06F 40/279 (2020.01)
- G06N 3/0455 (2023.01)
- G06N 3/08 (2023.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711159740
과제번호	2018-0-01398-005
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	SW컴퓨팅산업원천기술개발(R&D, 정보화)
연구과제명	(SW 스타랩) 대화 가능하고 자동으로 튜닝하는 DBMS의 개발
기 여 율	1/1
과제수행기관명	포항공과대학교 산학협력단
연구기간	2022.01.01 ~ 2022.12.31

명세서

청구범위

청구항 1

프로세서;

를 포함하고,

상기 프로세서는,

자연어 질의 문장에 대한 검색 결과를 나타내는 테이블 데이터를 언어모델의 입력으로 입력 받아 임베딩 벡터들을 생성하고,

상기 임베딩 벡터들을 이용하여 생성될 자연어 응답 문장의 임의의 위치에 미리 결정된 단어들에 배치될 각각의 확률을 계산하고,

상기 계산된 확률들로 구성되는 확률분포, 및 상기 자연어 질의 문장을 이용하여 상기 확률분포를 조절하고,

상기 자연어 응답 문장의 모든 위치에 있어서, 상기 모든 위치 중 임의의 위치에 배치될 단어를, 상기 임의의 위치와의 쌍으로 생성되는 확률분포에서 가장 큰 확률을 갖는 단어로 결정하고, 그리고

상기 모든 위치에서 결정된 단어들에 나열하여 상기 테이블의 변환 결과의 문장인, 상기 자연어 응답 문장을 생성하는,

질의 검색 결과 테이블을 자연어 응답 문장으로 변환하는 장치.

청구항 2

제1항에 있어서,

상기 테이블 데이터를 입력 받아 임베딩 벡터들을 생성하기 이전에,

상기 프로세서가 상기 테이블 데이터를 문자열로 변환하는,

질의 검색 결과 테이블을 자연어 응답 문장으로 변환하는 장치.

청구항 3

제1항에 있어서, 상기 프로세서가 상기 확률분포를 조절하는 프로세스를, 상기 생성될 자연어 응답 문장의 모든 위치마다 수행하는, 질의 검색 결과 테이블을 자연어 응답 문장으로 변환하는 장치.

청구항 4

제1항에 있어서,

상기 프로세서가 상기 확률을 계산할 때, 현재 생성할 단어의 위치에 대한 값을 더 입력 받아 계산하는, 질의 검색 결과 테이블을 자연어 응답 문장으로 변환하는 장치.

청구항 5

제1항에 있어서, 상기 프로세서가 상기 확률분포를 조절할 때,

상기 프로세서는,

상기 자연어 질의 문장에 포함된 토큰들의 제1집합을 생성하고,

상기 제1집합의 각 토큰에 대한 임베딩 벡터, 상기 미리 결정된 단어들의 각 임베딩 벡터, 및 상기 테이블 데이터의 각 임베딩 벡터간의 코사인 유사도를 계산하고, 그리고

상기 계산된 코사인 유사도에 소정의 값을 곱한 결과값을 상기 생성된 확률분포에 더하여 상기 조절된 확률분포를 생성하는,

질의 검색 결과 테이블을 자연어 응답 문장으로 변환하는 장치.

청구항 6

제5항에 있어서, 상기 소정의 값은 상기 생성될 자연어 문장의 길이에 따라 조절되는 값이며,

상기 소정의 값은 상기 미리 결정된 단어들 중에서 상기 자연어 질의 문장에 포함된 단어들과 유사도가 높은 단어들에 가중치를 주기 위한 값인,

질의 검색 결과 테이블을 자연어 응답 문장으로 변환하는 장치.

청구항 7

제6항에 있어서, 상기 소정의 값은 상기 자연어 질의 문장의 단어들로 이용되는 나열된 문장 또는 단어의 길이가 소정의 길이보다 크지 않도록 조절된 값인, 질의 검색 결과 테이블을 자연어 응답 문장으로 변환하는 장치.

청구항 8

제1항에 있어서, 상기 계산된 확률들로 구성되는 확률분포는 상기 생성될 자연어 응답 문장의 모든 위치 중 가장 왼쪽의 위치에 대한 확률분포부터 차례대로 생성되는, 질의 검색 결과 테이블을 자연어 응답 문장으로 변환하는 장치.

청구항 9

자연어 질의 문장에 대한 검색 결과를 나타내는 테이블 데이터를 언어모델의 입력으로 입력받아 임베딩 벡터들을 생성하는 단계;

상기 임베딩 벡터들을 이용하여 생성될 자연어 응답 문장의 임의의 위치에 미리 결정된 단어들이 배치될 각각의 확률을 계산하는 단계;

상기 계산된 확률들로 구성되는 확률분포, 및 상기 자연어 질의 문장을 이용하여 상기 확률분포를 조절하는 단계;

상기 자연어 응답 문장의 모든 위치에 있어서, 상기 모든 위치 중 임의의 위치에 배치될 단어를, 상기 임의의 위치와의 쌍으로 생성되는 확률분포에서 가장 큰 확률을 갖는 단어로 결정하는 단계; 및

상기 모든 위치에서 결정된 단어들을 나열하여 상기 테이블의 변환 결과의 문장인, 상기 자연어 응답 문장을 생성하는 단계;

를 포함하는,

질의 검색 결과 테이블을 자연어 응답 문장으로 변환하는 방법.

청구항 10

제9항에 있어서,

상기 확률을 계산하는 단계는, 현재 생성할 단어의 위치에 대한 값을 더 입력 받아 실행되는, 질의 검색 결과 테이블을 자연어 응답 문장으로 변환하는 방법.

청구항 11

제9항에 있어서, 상기 결정하는 단계는,

상기 자연어 질의 문장에 포함된 토큰들의 제1집합을 생성하는 단계;

상기 제1집합의 각 토큰에 대한 임베딩 벡터, 상기 미리 결정된 단어들의 각 임베딩 벡터, 및 상기 테이블 데이터의 각 임베딩 벡터간의 코사인 유사도를 계산하는 단계; 및

상기 계산된 코사인 유사도에 소정의 값을 곱한 결과값을 상기 생성된 확률분포에 더하여 상기 조절된 확률분포를 생성하는 단계;

를 포함하는,

질의 검색 결과 테이블을 자연어 응답 문장으로 변환하는 방법.

청구항 12

제11항에 있어서, 상기 소정의 값은 상기 생성될 자연어 문장의 길이에 따라 조절되는 값이며,

상기 소정의 값은 상기 미리 결정된 단어들 중에서 상기 자연어 질의 문장에 포함된 단어들과 유사도가 높은 단어들에 가중치를 주기 위한 값인,

질의 검색 결과 테이블을 자연어 응답 문장으로 변환하는 방법.

청구항 13

제12항에 있어서, 상기 소정의 값은 상기 자연어 질의 문장의 단어들로 이용되는 나열된 문장의 길이가 소정의 길이보다 크지 않도록 조절된 값인, 질의 검색 결과 테이블을 자연어 응답 문장으로 변환하는 방법.

청구항 14

제9항에 있어서, 상기 계산된 확률들로 구성되는 확률분포는 상기 생성될 자연어 응답 문장의 모든 위치 중 가장 왼쪽의 위치에 대한 확률분포부터 차례대로 생성되는, 질의 검색 결과 테이블을 자연어 응답 문장으로 변환하는 방법.

청구항 15

프로세서;

를 포함하고,

상기 프로세서는,

사용자에 의하여 입력된 자연어 질의 문장을 수신하고,

상기 자연어 질의 문장에 기반하여 구조화된 질의를 생성하고,

상기 구조화된 질의에 기반하여 데이터베이스에서 검색된 검색 결과를 나타내는 테이블 데이터를 언어모델의 입력으로 입력받아 임베딩 벡터들을 생성하고,

상기 임베딩 벡터들을 이용하여 생성될 자연어 응답 문장의 임의의 위치에 미리 결정된 단어들이 배치될 각각의 확률을 계산하고,

상기 계산된 확률들로 구성되는 확률분포, 및 상기 자연어 질의 문장을 이용하여 상기 확률분포를 조절하고,

상기 자연어 응답 문장의 모든 위치에 있어서, 상기 모든 위치 중 임의의 위치에 배치될 단어를, 상기 임의의 위치와의 쌍으로 생성되는 확률분포에서 가장 큰 확률을 갖는 단어로 결정하고, 그리고

상기 모든 위치에서 결정된 단어들을 나열하여 상기 테이블의 변환 결과의 문장인, 상기 자연어 응답 문장을 생성하여 제공하는,

대화형 인터페이스 장치.

청구항 16

제15항에 있어서,

상기 프로세서가 상기 확률을 계산할 때, 현재 생성할 단어의 위치에 대한 값을 더 입력 받아 계산하는, 대화형 인터페이스 장치.

청구항 17

제15항에 있어서, 상기 프로세서가 상기 확률분포를 조절할 때,

상기 프로세서는,

상기 자연어 질의 문장에 포함된 토큰들의 제1집합을 생성하고,

상기 제1집합의 각 토큰에 대한 임베딩 벡터, 상기 미리 결정된 단어들의 각 임베딩 벡터, 및 상기 테이블 데이터의 각 임베딩 벡터간의 코사인 유사도를 계산하고, 그리고

상기 계산된 코사인 유사도에 소정의 값을 곱한 결과값을 상기 생성된 확률분포에 더하여 상기 조절된 확률분포를 생성하는,

대화형 인터페이스 장치.

청구항 18

제17항에 있어서, 상기 소정의 값은 상기 생성될 자연어 문장의 길이에 따라 조절되는 값이며,

상기 소정의 값은 상기 미리 결정된 단어들 중에서 상기 자연어 질의 문장에 포함된 단어들과 유사도가 높은 단어들에 가중치를 주기 위한 값인,

대화형 인터페이스 장치.

청구항 19

제18항에 있어서, 상기 소정의 값은 상기 자연어 질의 문장의 단어들로 이용되는 나열된 문장의 길이가 소정의 길이보다 크지 않도록 조절된 값인, 대화형 인터페이스 장치.

청구항 20

제15항에 있어서, 상기 계산된 확률들로 구성되는 확률분포는 상기 생성될 자연어 응답 문장의 모든 위치 중 가장 왼쪽의 위치에 대한 확률분포부터 차례대로 생성되는, 대화형 인터페이스 장치.

발명의 설명

기술 분야

[0001] 본 발명은 질의 검색 결과 테이블을 자연어 응답 문장으로 변환하는 장치, 이를 위한 방법, 및 상기 장치를 포함하는 대화형 인터페이스 장치에 관한 기술로서, 특히 자연어 질의 문장과 질의 결과 테이블을 이용하여 자연어 문장을 요약 및/또는 변환하는 기술에 관한 것이다.

배경 기술

[0002] 최근 인터넷의 보급으로 사용자는 대단히 방대한 정보에 노출되고 있다. 이러한 방대한 정보 중 자신에게 필요한 정보만을 획득하기 위해서는 효과적인 검색이 필요하다.

[0003] 검색은 질의에 대한 응답으로 이루어진다. 일반적인 질의 응답 시스템은 질문에 대한 결과로 해답을 제시한다. 원하는 결과를 얻기 위해서는 질의가 규격을 갖춘 구조화된 질의일 것이 요구된다.

[0004] 대화형 인터페이스 모듈은 크게 자연어 질의를 구조화된 질의(예컨대, SQL 질의)로 번역하는 모듈과 생성된 질의 결과(예컨대, 테이블 데이터)를 자연어 문장으로 요약하는 모듈로 구분될 수 있다. 대한민국 공개특허(10-2023-0079729)의 문헌은 ‘자연어 질의의 SQL 변환 방법 및 장치’를 공개하고 있다. 그리고 대한민국 등록특허(10-2240303)의 문헌은 ‘데이터 테이블로부터의 자연어 생성 방법 및 데이터 테이블로부터의 자연어 생성을 수행하는 전자 장치’를 공개하고 있다.

[0005] 이때, 상기 번역 모듈과 호환이 가능한 질의 결과(예컨대, 테이블 데이터)를 자연어 문장으로 요약하는 모듈이 요구된다.

발명의 내용

해결하려는 과제

[0006] 본 발명에서는 자연어 질의를 구조화된 질의로 번역하는 모듈과 호환이 가능한, 질의 결과를 자연어 문장으로 요약/변환/번역하는 모듈을 제공하고자 한다.

[0007] 또한, 본 발명에서는 질의 검색 결과 테이블을 자연어 응답 문장으로 요약/변환/번역하는 장치를 제공하고자 한다. 즉, 본 발명에서는 자연어 질의 문장과 상기 질의 검색 결과 테이블만을 이용하여 자연어 응답 문장을 생성

하는 장치를 제공하고자 한다.

- [0008] 또한, 본 발명에서는 질의 검색 결과 테이블을 자연어 응답 문장으로 변환하는 방법을 제공하고자 한다.
- [0009] 또한, 본 발명에서는, 상기 변환하는 장치를 이용하는 대화형 인터페이스 장치를 제공하고자 한다.

과제의 해결 수단

- [0010] 본 발명의 일 관점에 따라 제공되는 질의 검색 결과 테이블을 자연어 응답 문장으로 변환하는 장치는, 프로세서를 포함할 수 있다. 이때, 상기 프로세서는, 자연어 질의 문장(문맥)에 대한 검색 결과를 나타내는 테이블 데이터를 언어모델의 입력으로 입력 받아 임베딩 벡터들을 생성하고, 상기 임베딩 벡터들을 이용하여 생성될 자연어 응답 문장의 임의의 위치에 미리 결정된 단어들 이 배치될 각각의 확률을 계산하고, 상기 계산된 확률들로 구성되는 확률분포, 및 상기 자연어 질의 문장을 이용하여 상기 확률분포를 조절하고, 상기 자연어 응답 문장의 모든 위치에 있어서, 상기 모든 위치 중 임의의 위치에 배치될 단어를, 상기 임의의 위치와의 쌍으로 생성되는 확률분포에서 가장 큰 확률을 갖는 단어로 결정하고, 그리고 상기 모든 위치에서 결정된 단어들 을 나열하여 상기 테이블의 변환 결과의 문장인, 상기 자연어 응답 문장을 생성할 수 있다.
- [0011] 이때, 상기 테이블 데이터를 입력 받아 임베딩 벡터들을 생성하기 이전에, 상기 프로세서는 상기 테이블 데이터를 문자열(토큰 시퀀스)로 변환할 수 있다.
- [0012] 이때, 상기 프로세서는 상기 확률분포를 조절하는 프로세스를, 상기 생성될 자연어 응답 문장의 모든 위치마다 수행할 수 있다.
- [0013] 이때, 상기 프로세서는 상기 확률을 계산할 때, 현재 생성할 단어의 위치에 대한 값을 더 입력 받아 계산할 수 있다.
- [0014] 이때, 상기 프로세서가 상기 확률분포를 조절할 때, 상기 프로세서는 상기 자연어 질의 문장에 포함된 토큰들의 제1집합을 생성하고, 상기 제1집합의 각 토큰에 대한 임베딩 벡터, 상기 미리 결정된 단어들의 각 임베딩 벡터, 및 상기 테이블 데이터의 각 임베딩 벡터간의 코사인 유사도를 계산하고, 그리고 상기 계산된 코사인 유사도에 소정의 값(λ)을 곱한 결과값을 상기 생성된 확률분포에 더하여 상기 조절된 확률분포를 생성할 수 있다.
- [0015] 이때, 상기 소정의 값(λ)은 상기 생성될 자연어 문장의 길이에 따라 조절되는 값이며, 상기 소정의 값은 상기 미리 결정된 단어들 중에서 상기 자연어 질의 문장에 포함된 단어들과 유사도가 높은 단어들에 가중치를 주기 위한 값일 수 있다.
- [0016] 이때, 상기 소정의 값은 상기 자연어 질의 문장의 단어들로 이용되는 나열된 문장 또는 단어의 길이가 소정의 길이보다 크지 않도록 조절된 값일 수 있다.
- [0017] 이때, 상기 계산된 확률들로 구성되는 확률분포는 상기 생성될 자연어 응답 문장의 모든 위치 중 가장 왼쪽의 위치에 대한 확률분포부터 차례대로 생성될 수 있다.
- [0018] 본 발명의 일 관점에 따라 제공되는 질의 검색 결과 테이블을 자연어 응답 문장으로 변환하는 방법은, 자연어 질의 문장(문맥)에 대한 검색 결과를 나타내는 테이블 데이터를 언어모델의 입력으로 입력받아 임베딩 벡터들을 생성하는 단계; 상기 임베딩 벡터들을 이용하여 생성될 자연어 응답 문장의 임의의 위치에 미리 결정된 단어들 이 배치될 각각의 확률을 계산하는 단계; 상기 계산된 확률들로 구성되는 확률분포, 및 상기 자연어 질의 문장 을 이용하여 상기 확률분포를 조절하는 단계; 상기 자연어 응답 문장의 모든 위치에 있어서, 상기 모든 위치 중 임의의 위치에 배치될 단어를, 상기 임의의 위치와의 쌍으로 생성되는 확률분포에서 가장 큰 확률을 갖는 단어로 결정하는 단계; 및 상기 모든 위치에서 결정된 단어들 을 나열하여 상기 테이블의 변환 결과의 문장인, 상기 자연어 응답 문장을 생성하는 단계;를 포함할 수 있다.
- [0019] 이때, 상기 확률을 계산하는 단계는, 현재 생성할 단어의 위치에 대한 값을 더 입력 받아 실행될 수 있다.
- [0020] 이때, 상기 결정하는 단계는, 상기 자연어 질의 문장에 포함된 토큰들의 제1집합을 생성하는 단계; 상기 제1집합의 각 토큰에 대한 임베딩 벡터, 상기 미리 결정된 단어들의 각 임베딩 벡터, 및 상기 테이블 데이터의 각 임베딩 벡터간의 코사인 유사도를 계산하는 단계; 및 상기 계산된 코사인 유사도에 소정의 값(λ)을 곱한 결과값 을 상기 생성된 확률분포에 더하여 상기 조절된 확률분포를 생성하는 단계;를 포함할 수 있다.
- [0021] 이때, 상기 소정의 값(λ)은 상기 생성될 자연어 문장의 길이에 따라 조절되는 값이며, 상기 소정의 값은 상기 미리 결정된 단어들 중에서 상기 자연어 질의 문장에 포함된 단어들과 유사도가 높은 단어들에 가중치를 주기

위한 값일 수 있다.

- [0022] 이때, 상기 소정의 값은 상기 자연어 질의 문장의 단어들로 이용되는 나열된 문장의 길이가 소정의 길이보다 크지 않도록 조절된 값일 수 있다.
- [0023] 이때, 상기 계산된 확률들로 구성되는 확률분포는 상기 생성될 자연어 응답 문장의 모든 위치 중 가장 왼쪽의 위치에 대한 확률분포부터 차례대로 생성될 수 있다.
- [0024] 본 발명의 일 관점에 따라 제공되는 대화형 인터페이스 장치는 프로세서를 포함할 수 있다. 상기 프로세서는, 사용자에게 의하여 입력된 자연어 질의 문장을 수신하고, 상기 자연어 질의 문장에 기반하여 구조화된 질의를 생성하고, 상기 구조화된 질의에 기반하여 데이터베이스에서 검색된 검색 결과를 나타내는 테이블 데이터를 언어 모델의 입력으로 입력받아 임베딩 벡터들을 생성하고, 상기 임베딩 벡터들을 이용하여 생성될 자연어 응답 문장의 임의의 위치에 미리 결정된 단어들에 배치될 각각의 확률을 계산하고, 상기 계산된 확률들로 구성되는 확률 분포, 및 상기 자연어 질의 문장을 이용하여 상기 확률분포를 조절하고, 상기 자연어 응답 문장의 모든 위치에 있어서, 상기 모든 위치 중 임의의 위치에 배치될 단어를, 상기 임의의 위치와의 쌍으로 생성되는 확률분포에서 가장 큰 확률을 갖는 단어로 결정하고, 그리고 상기 모든 위치에서 결정된 단어들에 나열하여 상기 테이블의 변환 결과의 문장인, 상기 자연어 응답 문장을 생성하여 제공할 수 있다.
- [0025] 이때, 상기 프로세서가 상기 확률을 계산할 때, 현재 생성할 단어의 위치에 대한 값을 더 입력 받아 계산할 수 있다.
- [0026] 이때, 상기 프로세서가 상기 확률분포를 조절할 때, 상기 프로세서는, 상기 자연어 질의 문장에 포함된 토큰들의 제1집합을 생성하고, 상기 제1집합의 각 토큰에 대한 임베딩 벡터, 상기 미리 결정된 단어들의 각 임베딩 벡터, 및 상기 테이블 데이터의 각 임베딩 벡터간의 코사인 유사도를 계산하고, 그리고 상기 계산된 코사인 유사도에 소정의 값(λ)을 곱한 결과값을 상기 생성된 확률분포에 더하여 상기 조절된 확률분포를 생성할 수 있다.
- [0027] 이때, 상기 소정의 값(λ)은 상기 생성될 자연어 문장의 길이에 따라 조절되는 값이며, 상기 소정의 값은 상기 미리 결정된 단어들 중에서 상기 자연어 질의 문장에 포함된 단어들과 유사도가 높은 단어들에 가중치를 주기 위한 값일 수 있다.
- [0028] 이때, 상기 소정의 값은 상기 자연어 질의 문장의 단어들로 이용되는 나열된 문장의 길이가 소정의 길이보다 크지 않도록 조절된 값일 수 있다.
- [0029] 이때, 상기 계산된 확률들로 구성되는 확률분포는 상기 생성될 자연어 응답 문장의 모든 위치 중 가장 왼쪽의 위치에 대한 확률분포부터 차례대로 생성될 수 있다.

발명의 효과

- [0030] 본 발명에 따르면, 자연어 질의를 구조화된 질의로 번역하는 모듈과 호환이 가능한, 질의 결과를 자연어 문장으로 요약/변환/번역하는 모듈을 제공할 수 있다.
- [0031] 또한, 본 발명에 따르면, 질의 검색 결과 테이블을 자연어 응답 문장으로 요약/변환/번역하는 장치를 제공할 수 있다. 즉, 본 발명에 따르면, 자연어 질의 문장과 상기 질의 검색 결과 테이블만을 이용하여 자연어 응답 문장을 생성하는 장치를 제공할 수 있다.
- [0032] 또한, 본 발명에 따르면, 질의 검색 결과 테이블을 자연어 응답 문장으로 변환하는 방법을 제공할 수 있다.
- [0033] 또한, 본 발명에 따르면, 상기 변환하는 장치를 이용하는 대화형 인터페이스 장치를 제공할 수 있다.
- [0034] 또한, 본 발명에 따르면, 변환장치 및/또는 대화형 인터페이스 장치를 통해 데이터베이스를 잘 모르는 일반 사람들도 대화를 통해 데이터베이스에 저장된 정보를 검색하고 이해할 수 있게 하는 효과가 있다.

도면의 간단한 설명

- [0035] 도 1은 본 발명의 일 실시예에 따라 질의 검색 결과 테이블을 자연어 응답 문장으로 변환하는 장치를 설명하기 위한 도면이다.
 도 2는 본 발명의 다른 실시예에 따라 자연어 질의 문맥을 얻기 위한 전처리부를 포함하는 경우에 대해 설명하기 위한 도면이다.

도 3은 본 발명의 일 실시예에 따른 변환장치의 구성도를 나타낸 것이다.

도 4는 본 발명의 일 실시예에 따른 테이블 데이터와 토큰 시퀀스를 나타낸 것이다.

도 5는 본 발명의 일 실시예에 따른 확률분포를 설명하기 위한 도면이다.

도 6은 본 발명의 일 실시예에 따라 자연어 질의 문맥을 임베딩 벡터로 변환하는 과정을 설명하기 위한 도면이다.

도 7은 본 발명의 일 실시예에 따라 질의 검색 결과 테이블을 자연어 문장으로 변환하는 방법을 설명하기 위한 순서도이다.

도 8은 본 발명의 일 실시예에 따라 스코어 모듈이 확률분포를 조절하는 과정을 설명하기 위한 순서도이다.

도 9는 본 발명의 일 실시예에 따라 코사인 유사도를 계산하는 방법을 설명하기 위한 도면이다.

도 10은 본 발명의 일 실시예에 따라 단계(S340) 및 단계(S350)를 설명하기 위한 도면이다.

도 11은 본 발명의 일 실시예에 따른 대화형 인터페이스 장치를 설명하기 위한 도면이다.

도 12은 본 발명의 일 실시예에 따른 대화형 인터페이스 장치가 자연어 응답 문장을 생성하여 제공하는 과정을 설명하기 위한 순서도이다.

도 13은 도 1 내지 도 12의 과정의 적어도 일부를 수행할 수 있는 일반화된 변환장치 또는 대화형 인터페이스 장치 또는 컴퓨팅 시스템의 예시를 도시하는 개념도이다.

발명을 실시하기 위한 구체적인 내용

- [0036] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세하게 설명하고자 한다. 그러나, 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.
- [0037] 제1, 제2 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 본 발명의 권리 범위를 벗어나지 않으면서 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다. 및/또는 이라는 용어는 복수의 관련된 기재된 항목들의 조합 또는 복수의 관련된 기재된 항목들 중의 어느 항목을 포함한다.
- [0038] 본 출원의 실시예들에서, "A 및 B 중에서 적어도 하나"는 "A 또는 B 중에서 적어도 하나" 또는 "A 및 B 중 하나 이상의 조합들 중에서 적어도 하나"를 의미할 수 있다. 또한, 본 출원의 실시예들에서, "A 및 B 중에서 하나 이상"은 "A 또는 B 중에서 하나 이상" 또는 "A 및 B 중 하나 이상의 조합들 중에서 하나 이상"을 의미할 수 있다.
- [0039] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소가 다른 구성요소에 "직접 연결되어" 있다거나 "직접 접속되어" 있다고 언급된 때에는, 중간에 다른 구성요소가 존재하지 않는 것으로 이해되어야 할 것이다.
- [0040] 본 출원에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 출원에서, "포함하다" 또는 "가지다" 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.
- [0041] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가지고 있다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥 상 가지는 의미와 일치하는 의미를 가진 것으로 해석되어야 하며, 본 출원에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.
- [0042] 이하, 첨부한 도면들을 참조하여, 본 발명의 바람직한 실시예를 보다 상세하게 설명하고자 한다. 본 발명을 설명함에 있어 전체적인 이해를 용이하게 하기 위하여 도면상의 동일한 구성요소에 대해서는 동일한 참조부호를

사용하고 동일한 구성요소에 대해서 중복된 설명은 생략한다.

- [0043] 도 1은 본 발명의 일 실시예에 따라 질의 검색 결과 테이블을 자연어 응답 문장으로 변환하는 장치를 설명하기 위한 도면이다.
- [0044] 변환장치(110)는 사용자(10)에 의하여 자연어 질의 문장(예컨대, show year, name, price in dollar of cars manufactured after 2020)을 수신하고, 상기 자연어 질의 문장의 검색 결과를 나타내는 테이블 데이터(131)를 입력 받아 자연어 응답 문장(예컨대, In 2022, the model of the genesis was 8900 in dollar)을 생성할 수 있다. 이때, 자연어 질의 문장의 입력 방식은 키보드/키패드를 경유하는 텍스트 입력일 수도 있다.
- [0045] 본 발명의 변환장치(110)는 질의 검색 결과 테이블을 자연어 응답 문장으로 변환(요약)하는 장치로서, 관계형 데이터베이스에 대화형 인터페이스를 개발할 때 적용될 수 있다. 대화형 인터페이스를 위해서는 데이터베이스 질의 결과로 생성된 테이블(131)을 요약해서 사용자에게 자연어 문장으로 제공해 주어야 한다. 이때, 본 발명의 변환장치(110)가 이용될 수 있다. 이때, 상기 자연어 문장의 제공의 의미는 시각적으로 디스플레이 하거나, 음성으로 출력하거나, 그리고/또는 점자로 표현하여 출력하는 것을 의미할 수 있으며 이에 제한되지 않는다.
- [0046] 도 1에서 변환장치(110)에 입력되는 테이블이 하나인 것으로 예를 들었지만, 복수 개의 테이블들을 입력 받을 수도 있다. 본 발명의 변환장치(110)는 복수 개의 테이블들을 입력 받는 경우에도 자연스러운 질의 응답 문장을 생성할 수 있다.
- [0047] 도 2는 본 발명의 다른 실시예에 따라 자연어 질의 문맥을 얻기 위한 전처리부를 포함하는 경우에 대해 설명하기 위한 도면이다.
- [0048] 이하, 도 1 및 도 2를 함께 참조하여 설명한다.
- [0049] 예컨대, 사용자(10)로부터 자연어 질의 문장 텍스트가 아니라 자연어 질의 음성을 입력 받을 수도 있다. 이때, 변환장치(110)에 자연어 질의 문장을 입력하기 위해, 전처리부(140)는 상기 자연어 질의 음성을 인식하고 인식된 결과를 자연어 질의 문장(문맥)으로 변환할 수 있다.
- [0050] 도 3은 본 발명의 일 실시예에 따른 변환장치의 구성도를 나타낸 것이다.
- [0051] 도 4는 본 발명의 일 실시예에 따른 테이블 데이터와 토큰 시퀀스를 나타낸 것이다.
- [0052] 도 5는 본 발명의 일 실시예에 따른 확률분포를 설명하기 위한 도면이다.
- [0053] 도 5에서는 확률분포가 생성되는 시간(t)에 따른 확률분포(P)를 도시했으며, 각 확률분포(P)에 대응하는 조절된 확률분포(C_P), 상기 조절된 확률분포(C_P)의 확률들 중 가장 큰 확률을 갖는 토큰(T01~T0k), 그리고 상기 토큰(T01~T0k)이 위치할, 즉 생성될 자연어 응답 문장에 포함될 단어의 위치(TOP1~TOPk)를 도시했다. 도 5의 도시 방법은 설명의 편의를 위한 것이다.
- [0054] 이하, 도 3 내지 도 5를 함께 참조하여 설명한다.
- [0055] 변환장치(110)는 파싱부(111), 인코더(112), 디코더(113), 스코어 모듈(114), 및 자연어 문장 생성부(115)를 포함할 수 있다.
- [0056] 파싱부(111)는 도 4와 같이, 질의 결과 테이블(131)을 입력 받아 파싱하여 토큰 시퀀스(문자열)(132)를 생성할 수 있다. 즉, 테이블의 각 요소들이 토큰화될 수 있다.
- [0057] 인코더(112)는 토큰 시퀀스(132)를 인코더(112)에 포함된 언어모델의 입력으로 입력 받아 토큰 시퀀스(132)의 토큰들 각각에 대해 임베딩 벡터를 생성할 수 있다. 이때, 상기 언어모델은 사전 학습된 언어 모델일 수 있다.
- [0058] 디코더(113)는 임베딩 벡터들을 입력 받아 생성될 자연어 응답 문장의 임의의 위치에 미리 결정된 단어들이 배치될 각각의 확률을 계산할 수 있다. 디코더(113)는 사전 학습된 언어모델을 포함할 수 있으며, 상기 언어모델을 이용하여 상기 확률을 계산할 수 있다.
- [0059] 디코더(113)는 미리 결정된 단어들의 임베딩 벡터들의 집합을 더 입력 받을 수 있다. 이때, 상기 미리 결정된 단어들의 임베딩 벡터들의 집합은 학습 집합(또는 단어 집합, 예컨대, 단어, 알파벳, 토큰 등 가능한 모든 토큰들의 집합)을 의미할 수 있다. 예컨대, 도 5를 참조하면, 상기 임의의 위치가 첫번째 단어(또는 토큰)이 들어갈 위치(TOP1)인 것으로 가정할 수 있다. 이때, 첫번째 단어가 들어갈 위치에 상기 미리 결정된 단어들 각각이 선택될 확률을 계산할 수 있다. 예컨대, 상기 미리 결정된 단어들 중 '오늘'이라는 단어가 상기 첫번째 단어가 들어갈 위치에 배치될 확률은 P1이고, 상기 미리 결정된 단어들 중 '이름'이 상기 첫번째 단어가 들어갈 위치에

배치될 확률은 P2이고, 그리고 상기 미리 결정된 단어들 중 '연도'가 상기 첫번째 상기 첫번째 단어가 들어갈 위치에 배치될 확률은 P3일 수 있다. 이와 같이, 상기 미리 결정된 단어들 모두에 대해 상기 임의의 위치에 배치될 확률을 각각 계산할 수 있다. 이때, 상기 임의의 위치의 범위는 상기 자연어 응답 문장의 첫번째 단어의 위치부터 마지막 단어의 위치까지일 수 있다.

- [0060] 예컨대, 디코더(113)에서 첫번째 단어가 들어갈 위치에 대한 상기 미리 결정된 단어들 모두의 확률이 계산되면, 상기 계산된 확률들로 구성되는 확률분포(P)가 생성될 수 있다. 예컨대, 상기 확률분포는 생성될 자연어 응답 문장의 모든 위치 중 가장 왼쪽의 위치에 대한 확률분포부터 차례대로 생성될 수 있다.
- [0061] 스코어 모듈(114)은 디코더(113)에서 계산된 확률들을 입력 받아 확률분포를 생성하고, 자연어 질의 문맥을 입력 받아 상기 확률분포를 조절할 수 있다. 또는 스코어 모듈(114)은 이미 생성된 확률분포를 입력 받을 수도 있다. 스코어 모듈(114)은 상기 확률분포의 확률들에 대한 스코어들을 생성할 수 있으며, 생성한 스코어들을 상기 자연어 질의 문맥을 이용하여 갱신된 스코어들을 생성할 수 있다. 이때, 상기 갱신된 스코어들 중 가장 큰 스코어를 갖는, 임베딩 벡터에 대응하는 토큰이 결정될 수 있다.
- [0062] 이때, 상기 자연어 질의 문맥은 자연어 질의 문장인 텍스트 그 자체를 의미할 수도 있고, 텍스트를 이용하여 생성한 다른 구성요소일 수도 있다.
- [0063] 이때, 스코어 모듈(114)이 상기 자연어 질의 문맥을 입력 받기 위해 상기 자연어 질의 문맥은 임베딩 벡터로 변환될 수 있다.
- [0064] 이때, 도 3, 및 도 5를 함께 참조하면, '생성될 자연어 응답 문장'의 모든 위치마다 확률분포가 생성되고 생성된 확률분포가 조절될 수 있다.
- [0065] 다시 도 3을 참조하면, 자연어 문장 생성부(115)는 자연어 응답 문장의 모든 위치마다 스코어 모듈(50)에서 조절된 확률분포에서 가장 큰 확률을 갖는 단어들(즉, 스코어가 가장 큰 단어들)을 나열하여 자연어 응답 문장을 생성할 수 있다.
- [0066] 도 6은 본 발명의 일 실시예에 따라 자연어 질의 문맥을 임베딩 벡터로 변환하는 과정을 설명하기 위한 도면이다.
- [0067] 이하, 도 3, 도 5, 및 도 6을 함께 참조하여 설명한다.
- [0068] 변환장치(110)는 제2 파싱부(116), 및 제2 인코더(117)를 더 포함할 수 있다.
- [0069] 제2 파싱부(116)는 자연어 질의 문맥을 입력 받아 파싱하여 토큰 시퀀스를 생성할 수 있다. 상기 토큰 시퀀스의 토큰들의 집합을 제1집합이라고 지칭할 수 있다.
- [0070] 제2 인코더(117)는 상기 토큰 시퀀스를 입력 받아 토큰들 각각에 대해 임베딩 벡터를 생성할 수 있다. 이를 위하여, 제2 인코더(70)는 사전에 학습된 언어 모델을 포함할 수 있다.
- [0071] 스코어 모듈(114)은 상기 임베딩 벡터, 및 확률분포(P)를 이용하여 조절된 확률분포(C_P)를 생성할 수 있다. 즉, 스코어 모듈(114)은 상기 임베딩 벡터를 이용하여 생성될 자연어 응답 문장의 단어들이 상기 자연어 질의 문맥의 단어들과 유사한 순서대로 배치될 수 있도록 할 수 있다.
- [0072] 예컨대, 도 1, 도 3, 도 5 및 도 6을 함께 참조하면, 디코더(113)에서 생성된 확률로서, 미리 결정된 단어들 중 상기 '첫번째 단어가 들어갈 위치'에 배치될 확률이 가장 높은 단어가 '이름(name)' 이라고 가정할 수 있다. 이때, 스코어 모듈(114)은 상기 제2 인코더(117)에서 생성된 임베딩 벡터를 이용하여 자연어 질의 문맥(예컨대, show year, name, price in dollar of cars manufactured after 2020)의 단어 배치 순서와 유사하게, 상기 '이름' 보다는 '연도'가 배치될 확률이 더 높아지도록 상기 확률분포를 조절할 수 있다. 이때, 도 5를 참조하면, '첫번째 단어가 들어갈 위치' (TOP1)에 대하여, 조절된 확률분포(C_P)에서 가장 큰 확률(P2a)을 갖는 토큰(T01)은 '연도'가 될 수 있으며, '연도'가 상기 '첫번째 단어가 들어갈 위치'에 배치되는 단어로 결정될 수 있다.
- [0073] 스코어 모듈(114)은 생성될 문장에 포함될 단어의 위치에 대한 정보를 생성할 수 있으며, 생성된 정보는 디코더(113)에 제공될 수 있다. 도 3을 참조하면, 디코더(113)는 스코어 모듈(114)에서 현재 시점(예컨대, t1)에서 생성한 단어의 위치(TOP1)를 입력 받아 다음 위치(TOP2)를 결정하고, 상기 다음 위치(TOP2)에 배치될 단어들의 확률을 계산할 수 있다. 또는 스코어 모듈(114)에서 현재 시점(예컨대, t1)에 생성한 단어의 위치에 대한 값(예컨대, TOP=1)에 소정의 값(예컨대, 1)을 더하여 다음 위치에 대한 값(TOP=2)을 생성할 수 있다. 디코더(113)는 상

기 다음 위치에 대한 값(TOP=2)을 입력 받아 상기 다음 위치에 배치될 단어들의 확률을 계산할 수 있다.

- [0074] 도 7은 본 발명의 일 실시예에 따라 질의 검색 결과 테이블을 자연어 문장으로 변환하는 방법을 설명하기 위한 순서도이다.
- [0075] 도 7에 도시한 질의 검색 결과 테이블을 자연어 문장으로 변환하는 방법의 각 단계들은 상술한 내용들을 정리한 것이다.
- [0076] 단계(S310)에서, 자연어 질의 문장(문맥)에 대한 검색 결과를 나타내는 테이블 데이터를 언어모델의 입력으로 입력 받아 임베딩 벡터들을 생성할 수 있다.
- [0077] 단계(S320)에서, 상기 임베딩 벡터들을 이용하여 생성될 자연어 응답 문장의 임의의 위치에 미리 결정된 단어들이 배치될 각각의 확률을 계산할 수 있다. 이때, 상기 미리 결정된 단어들은 단어, 알파벳, 토큰 등 가능한 모든 토큰들을 의미할 수 있다.
- [0078] 단계(S330)에서, 상기 계산된 확률들로 구성되는 확률분포, 및 상기 자연어 질의 문장을 이용하여 상기 확률분포를 조절할 수 있다.
- [0079] 단계(S340)에서, 상기 자연어 응답 문장의 모든 위치에 있어서, 상기 모든 위치 중 임의의 위치에 배치될 단어를, 상기 임의의 위치(예컨대, 도 5의 TOPk)와의 쌍으로 생성되는 확률분포(C_P=k)에서 가장 큰 확률을 갖는 단어로 결정할 수 있다.
- [0080] 단계(S350)에서, 상기 모든 위치에서 결정된 단어들을 나열하여 상기 테이블의 변환 결과의 문장인, 상기 자연어 응답 문장을 생성할 수 있다.
- [0081] 이때, 상기 모든 위치에 배치될 단어들을 결정하기 위해서, 단계(S320) 및 단계(S330)은 하나의 프로세스로서 상기 생성될 자연어 응답 문장의 모든 위치마다 반복하여 실행될 수 있다.
- [0082] 도 8은 본 발명의 일 실시예에 따라 스코어 모듈이 확률분포를 조절하는 과정을 설명하기 위한 순서도이다.
- [0083] 도 9는 본 발명의 일 실시예에 따라 코사인 유사도를 계산하는 방법을 설명하기 위한 도면이다.
- [0084] 이하, 도 1, 도 3, 및 도 5 내지 9를 함께 참조하여 설명한다.
- [0085] 스코어 모듈(114)이 확률분포를 조절하는 단계(S330)는 이하의 단계(S331) 및 단계(S332)를 포함할 수 있다. 이때, 단계(S331) 이전에 자연어 질의 문장에 대한 임베딩 벡터를 생성하기 위한 단계들(S410~S420)이 더 실행될 수 있다. 순서대로 살펴보면 다음과 같다.
- [0086] 단계(S410)에서, 자연어 질의 문장에 포함된 토큰들의 제1집합을 생성할 수 있다. 예컨대, 도 1을 참조하면, 상기 제1집합은 {show, year, name, price, in, dollar, of, cars, manufactured, after, 2020}일 수 있다.
- [0087] 단계(S420)에서, 상기 제1집합의 각 토큰에 대한 임베딩 벡터를 생성할 수 있다.
- [0088] 단계(S331)에서, 상기 제1집합의 각 토큰에 대한 임베딩 벡터, 미리 결정된 단어들의 각 임베딩 벡터(단계(S320) 참조), 및 테이블 데이터의 각 임베딩 벡터간의 코사인 유사도를 계산할 수 있다. 예컨대, 도 9의 표(510)는 자연어 질의의 각 임베딩 벡터에 번호(1~e)를 매칭한 것이고, 표(520)은 테이블 데이터의 각 임베딩 벡터에 번호(1~z)를 매칭한 것이며, 그리고 표(530)은 상술한 미리 결정된 단어들의 각 임베딩 벡터에 번호(1~g)를 매칭한 것이다. 예컨대, 표(540)을 참조하면, 테이블 데이터의 각 임베딩 벡터에 대하여 e*g 개수만큼의 코사인 유사도 값이 계산될 수 있다. 이때, 각 계산 조합의 내부 순서는 {테이블 데이터의 임베딩 벡터 번호, 자연어 질의 임베딩 벡터 번호, 미리 결정된 단어들의 임베딩 벡터 번호}로 나타냈다. 이는 설명의 편의를 위한 것이며, 이에 제한되지 않는다. 이때, 테이블 데이터의 임의의 임베딩 벡터(예컨대, 2번 임베딩 벡터)에 대하여 e*g 개수만큼의 코사인 유사도들 중 가장 큰 코사인 유사도(예컨대, 조합{2, e, 7}의 코사인 유사도)를 상기 임의의 임베딩 벡터의 최종 코사인 유사도라고 지칭할 수 있다. 이때, 표(540)은 테이블 데이터의 각 임베딩 벡터를 기준으로 작성한 것이며, 상기 미리 결정된 단어들의 각 임베딩 벡터를 기준으로 할 수도 있다. 이 경우, 상기 최종 코사인 유사도는 미리 결정된 단어들의 각 임베딩 벡터에 대한 것일 수 있다.
- [0089] 단계(S332)에서, 상기 계산된 최종 코사인 유사도에 소정의 값(λ)을 곱한 결과값을 생성된 확률분포에 더하여 조절된 확률분포를 생성할 수 있다. 다시 말하면, 상기 생성된 확률분포를 구성하는 확률들 중 임의의 확률(예컨대, 테이블 데이터의 임베딩 벡터 번호 2의 확률)(스코어, score)에, 상기 임의의 확률에 대응하는 최종 코사인 유사도(예컨대, 테이블 데이터의 임베딩 벡터 번호 2의 최종 코사인 유사도)에 소정의 값(λ)을 곱한 결과값

을 더하여 조절된 확률(스코어, score')이 생성될 수 있다. 상기 조절된 확률분포는 조절된 확률들을 포함할 수 있다.

[0090] 예컨대, 조절된 확률분포를 생성하는 수식 1은 아래와 같을 수 있다.

[0091] [수식 1]

$$score'(y_t | w, y_{<t}) = score(y_t | y_{<t}) + \lambda \cdot \max(0, \cos(\gamma(y_t), \gamma(w)))$$

[0092]

[0094] 이때, y_t 는 t 시점(time step)에서 생성될 토큰을 의미하고, $y_{<t}$ 는 t 시점 이전에 생성된 토큰들을 의미할 수 있다. w는 스코어 수정에 사용되는 가이드 단어(guide word) 집합으로 도 6의 제2 과정부(116)에서 과성된 토큰들의 제1집합을 의미할 수 있다. λ 는 가중치(weight) 값을 의미할 수 있다. 그리고 γ 는 단어가 입력으로 주어졌을 때 임베딩 벡터를 출력하는 함수를 의미할 수 있다.

[0095] 이때, 상기 t시점에서 추가 점수 비율을 조절하기 위한 가중치 값인 상기 소정의 값(λ)을 결정하기 위한 수식 2는 아래와 같을 수 있다.

[0096] [수식 2]

$$\lambda_t = \lambda_0 \cdot \exp\left(\frac{ct}{T - |W|}\right)$$

[0097]

[0099] 이때, λ_t 는 t시점에서의 스코어를 수정할 가중치 값을 의미하고, λ_0 는 초기 가중치 값을 의미할 수 있다.

T 는 사전에 정의되는 생성될 문장의 최대 토큰 개수를 의미할 수 있다. W 는 t시점에서 아직 생성에 사용되지 않은 가이드 단어들의 집합을 의미할 수 있다. 그리고 c는 비율을 조절하는 하이퍼파라미터(hyperparameter)일 수 있다.

[0100] 예컨대, 상기 소정의 값(λ)은 상기 생성될 자연어 문장의 길이(예컨대, 윈도우의 길이)에 따라 조절되는 값일 수 있다. 이때, 상기 소정의 값(λ)은 상기 미리 결정된 단어들 중에서 상기 자연어 질의 문장에 포함된 단어들과 유사도가 높은 단어들에 가중치를 주기 위한 값일 수 있다. 이때, 상기 유사도가 높음을 판단하는 기준은 미리 설정된 기준치에 따라 판단될 수 있다. 이때, 상기 소정의 값(λ)은 상기 자연어 질의 문장의 단어들로 이용되는 나열된 문장 또는 단어의 길이가 소정의 길이보다 크지 않도록 조절된 값일 수 있다. 이러한 이유는 자연어 질의 문장이 너무 길면 확률분포에 너무 큰 영향을 미치게 되기 때문이다.

[0101] 상술한 소정의 값을 이용하여 확률분포를 조절하는 과정을 통해 변환장치(110)가 상황에 따라 다양한 형태로 자연스러운 자연어 응답 문장을 생성할 수 있다.

[0102] 도 10은 본 발명의 일 실시예에 따라 단계(S340) 및 단계(S350)를 설명하기 위한 도면이다.

[0103] 도 10의 표는 단계(S330)의 반복 순서(즉, 생성될 자연어 응답 문장의 위치 순서)와 상기 반복 순서에 대응하는 순서(예컨대, 2번째)에 생성된 조절된 확률분포의 확률들 중 가장 큰 확률을 갖는 토큰(예컨대, T02)의 쌍을 나타낸 것이다.

[0104] 이때, 상술한 확률분포는 생성될 자연어 응답 문장의 모든 위치 중 가장 왼쪽의 위치에 대한 확률분포부터 차례대로 생성됨을 확인할 수 있다.

[0105] 예컨대, 첫번째 위치의 토큰(T01)은 'in', 두번째 위치의 토큰(T02)은 '2022', 세번째 위치의 토큰(T03)은 'the', ..., 그리고 마지막 위치의 토큰(T0k)은 'dollar'일 수 있다. 각 위치마다 결정된 토큰들을 나열하면 우측의 자연어 응답 문장(예컨대, In 2022, the ... dollar)이 생성될 수 있다.

[0106] 도 11은 본 발명의 일 실시예에 따른 대화형 인터페이스 장치를 설명하기 위한 도면이다.

[0107] 도 11의 실시예는 본 발명의 대화형 인터페이스 장치에서 자연어 응답 문장을 생성하는 예이다.

- [0108] 대화형 인터페이스 장치(100)는 상술한 변환장치(110), 질의 번역기(120), 및 질의 결과 테이블 생성기(130)를 포함할 수 있다.
- [0109] 질의 번역기(120)는 자연어 질의 문장의 입력을 구조화된 질의로 번역할 수 있다. 질의 번역기(120)는 구조화된 질의를 질의 결과 테이블 생성기(130)로 전달할 수 있다.
- [0110] 질의 결과 테이블 생성기(130)는 데이터베이스의 테이블들 중 구조화된 질의에 대응하는 테이블들(131)을 결정할 수 있다.
- [0111] 변환장치(110)는 상기 구조화된 질의에 대응하는 테이블들(131)을 입력으로서 전달받고 상기 자연어 질의 문장을 이용하여 자연어 응답(요약) 문장을 생성할 수 있다.
- [0112] 최근 연구된 NLProv 기술은 생성된 질의 결과의 기원(provenance)을 바탕으로 질의 결과가 어떻게 생성되었는지를 파악하고, 주어진 자연어 질의에서 각 단어와 SQL 질의의 연결성을 활용해서 질의 결과를 요약하는 문장을 생성하는 기술이다.
- [0113] 데이터베이스의 대화형 인터페이스 장치를 개발하는 데 있어서, 상술한 NLProv 기술은 자연어 질의를 SQL 질의로 번역하는 모듈(Text-to-SQL 모듈)과의 호환성 문제가 있다. NLProv 기술은 상기 Text-to-SQL 모듈에서 자연어 질의 문장의 단어들에 SQL 질의와 어떻게 연결이 되는지에 대한 정보가 주어지면 동작이 가능한 기술이다. 그러나 최근 Text-to-SQL 문제에서 높은 성능 향상을 이끌어온 인공 신경망 기반의 방법들은 상기 어떻게 연결이 되는지에 대한 정보를 생성하지 않는다. 따라서 NLProv 기술은 Text-to-SQL 모듈과 호환이 불가능하다.
- [0114] NLProv 기술은 단순 결합 질의(simple conjunctive query)에 대해서만 질의 결과를 요약하는 자연어 문장을 생성할 수 있다. 하지만 데이터베이스가 실제로 사용될 때는 집합 연산자(Aggregation)를 비롯한 다양한 질의에 대한 지원이 필요하다. 또한, NLProv 기술은 주어진 질의에 따라서 자연어 문장을 생성하는 시간이 많이 차이날 수 있다. NLProv 기술은 연산이 많이 필요한 질의에 대해서는 크게 10초까지 시간이 걸린다. 사용자에게 즉각적으로 피드백을 보여줘야 하는 시스템으로써 10초는 긴 시간이다.
- [0115] 한편, NLProv에서 생성되는 자연어 문장 형태는 사람이 지정한 문장 템플릿으로 고정되어 있다. 자연스러운 대화형 인터페이스 개발에 있어서 시스템이 생성하는 자연어 문장이 정해진 형태가 아닌, 상황에 따라 다양한 형태로 생성되는 것은 중요한 문제이다.
- [0116] 본 발명은 상술한 NLProv 기술의 문제들을 해결할 수 있는 발명으로서, Text-to-SQL 모듈에서 자연어 질의 문장의 단어들에 SQL 질의와 어떻게 연결이 되는지에 대한 정보 없이도 질의 결과 테이블을 이용하여 자연어 응답 문장을 제공할 수 있는 변환장치에 관한 것이다. 본 발명의 변환장치는 Text-to-SQL 모듈과 호환이 가능한 생성된 질의 결과(즉, 테이블)를 자연어 문장으로 요약하는 모듈(Table-to-Text)일 수 있다. 따라서 본 발명의 변환장치는 Text-to-SQL 모듈과는 독립적으로 존재할 수 있다. 본 발명의 변환장치는 질의 결과를 요약할 때, 자연어 응답 문장을 생성할 때 질의 결과 테이블 자체만을 사용하고 해당 결과(즉, 테이블)가 어떤 식으로 생성되었는지 고려하지 않기 때문에 모든 형태의 질의에 대해서 지원이 가능하다. 또한, 본 발명의 변환장치를 이용하면 질의에 따른 속도 저하 문제가 발생하지 않는다. 또한, 본 발명은 테이블을 인코딩하고 자연어 문장을 생성하는 데 있어서 사전 학습된 언어 모델을 기반으로 하고 있다. 본 발명은 NLProv 기술의 제한적인 자연어 문장이 생성된다는 단점을 없애고 상황에 따라서 다양하고 매끄러운 문장을 생성할 수 있다(즉, 자연어의 유창성 문제를 해결한다)는 장점이 있다.
- [0117] 도 12는 본 발명의 일 실시예에 따른 대화형 인터페이스 장치가 자연어 응답 문장을 생성하여 제공하는 과정을 설명하기 위한 순서도이다.
- [0118] 단계(S610)에서, 사용자에게 의하여 자연어 질의 문장을 수신할 수 있다.
- [0119] 단계(S620)에서, 상기 자연어 질의 문장에 기반하여 구조화된 질의를 생성할 수 있다.
- [0120] 단계(S630)에서, 상기 구조화된 질의를 데이터베이스에 검색하여 검색 결과로서 하나 이상의 테이블 데이터를 획득할 수 있다.
- [0121] 단계(S640)에서, 상기 테이블 데이터를 언어모델의 입력으로 입력받아 임베딩 벡터들을 생성할 수 있다.
- [0122] 이하의 단계(S650)은 도 7의 단계(S320)에, 단계(S660)은 단계(S330)에, 그리고 단계(S670)은 단계(S340)에 각각 대응될 수 있다.

- [0123] 이때, 단계(S680)은 도 7의 단계(S350)에 대응할 수 있으며, 더 나아가 단계(S350)에서 생성된 자연어 응답 문장을 사용자에게 제공하는 과정을 더 포함하는 개념일 수 있다. 이때, '제공'의 방식은 사용자가 확인할 수 있도록 시각적으로 디스플레이, 음성으로 표현, 및/또는 점자로 표현하는 방식과 같을 수 있으며 이에 제한되지 않는다.
- [0124] 도 13은 도 1 내지 도 12의 과정의 적어도 일부를 수행할 수 있는 일반화된 변환장치 또는 대화형 인터페이스 장치 또는 컴퓨팅 시스템의 예시를 도시하는 개념도이다.
- [0125] 본 발명의 일 실시예에 따른 질의 검색 결과 테이블을 자연어 응답 문장으로 변환하는 방법, 및 대화형 인터페이스 장치가 자연어 응답 문장을 생성하여 제공하는 방법의 적어도 일부의 과정은 도 13의 컴퓨팅 시스템(1000)에 의하여 실행될 수 있다.
- [0126] 도 13을 참조하면, 본 발명의 일 실시예에 따른 컴퓨팅 시스템(1000)은, 프로세서(1100), 메모리(1200), 통신 인터페이스(1300), 저장 장치(1400), 입력 인터페이스(1500), 출력 인터페이스(1600) 및 버스(bus)(1700)를 포함하여 구성될 수 있다.
- [0127] 본 발명의 일 실시예에 따른 컴퓨팅 시스템(1000)은, 적어도 하나의 프로세서(processor)(1100) 및 상기 적어도 하나의 프로세서(1100)가 적어도 하나의 단계를 수행하도록 지시하는 명령어들(instructions)을 저장하는 메모리(memory)(1200)를 포함할 수 있다. 본 발명의 일 실시예에 따른 방법의 적어도 일부의 단계는 상기 적어도 하나의 프로세서(1100)가 상기 메모리(1200)로부터 명령어들을 로드하여 실행함으로써 수행될 수 있다.
- [0128] 프로세서(1100)는 중앙 처리 장치(central processing unit, CPU), 그래픽 처리 장치(graphics processing unit, GPU), 또는 본 발명의 실시예들에 따른 방법들이 수행되는 전용의 프로세서를 의미할 수 있다.
- [0129] 메모리(1200) 및 저장 장치(1400) 각각은 휘발성 저장 매체 및 비휘발성 저장 매체 중에서 적어도 하나로 구성될 수 있다. 예를 들어, 메모리(1200)는 읽기 전용 메모리(read only memory, ROM) 및 랜덤 액세스 메모리(random access memory, RAM) 중에서 적어도 하나로 구성될 수 있다.
- [0130] 또한, 컴퓨팅 시스템(1000)은, 무선 네트워크를 통해 통신을 수행하는 통신 인터페이스(1300)를 포함할 수 있다.
- [0131] 또한, 컴퓨팅 시스템(1000)은, 저장 장치(1400), 입력 인터페이스(1500), 출력 인터페이스(1600) 등을 더 포함할 수 있다.
- [0132] 또한, 컴퓨팅 시스템(1000)에 포함된 각각의 구성 요소들은 버스(bus)(1700)에 의해 연결되어 서로 통신을 수행할 수 있다.
- [0133] 본 발명의 컴퓨팅 시스템(1000)의 예를 들면, 통신 가능한 데스크탑 컴퓨터(desktop computer), 랩탑 컴퓨터(laptop computer), 노트북(notebook), 스마트폰(smart phone), 태블릿 PC(tablet PC), 모바일폰(mobile phone), 스마트 워치(smart watch), 스마트 글래스(smart glass), e-book 리더기, PMP(portable multimedia player), 휴대용 게임기, 네비게이션(navigation) 장치, 디지털 카메라(digital camera), DMB(digital multimedia broadcasting) 재생기, 디지털 음성 녹음기(digital audio recorder), 디지털 음성 재생기(digital audio player), 디지털 동영상 녹화기(digital video recorder), 디지털 동영상 재생기(digital video player), PDA(Personal Digital Assistant) 등일 수 있다.
- [0134] 본 발명의 실시예에 따른 방법의 동작은 컴퓨터로 읽을 수 있는 기록매체에 컴퓨터가 읽을 수 있는 프로그램 또는 코드로서 구현하는 것이 가능하다. 컴퓨터가 읽을 수 있는 기록매체는 컴퓨터 시스템에 의해 읽힐 수 있는 정보가 저장되는 모든 종류의 기록장치를 포함한다. 또한 컴퓨터가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어 분산 방식으로 컴퓨터로 읽을 수 있는 프로그램 또는 코드가 저장되고 실행될 수 있다.
- [0135] 또한, 컴퓨터가 읽을 수 있는 기록매체는 롬(rom), 램(ram), 플래시 메모리(flash memory) 등과 같이 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치를 포함할 수 있다. 프로그램 명령은 컴파일러(compiler)에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터(interpreter) 등을 사용해서 컴퓨터에 의해 실행될 수 있는 고급 언어 코드를 포함할 수 있다.
- [0136] 본 발명의 일부 측면들은 장치의 문맥에서 설명되었으나, 그것은 상응하는 방법에 따른 설명 또한 나타낼 수 있고, 여기서 블록 또는 장치는 방법 단계 또는 방법 단계의 특징에 상응한다. 유사하게, 방법의 문맥에서 설명된

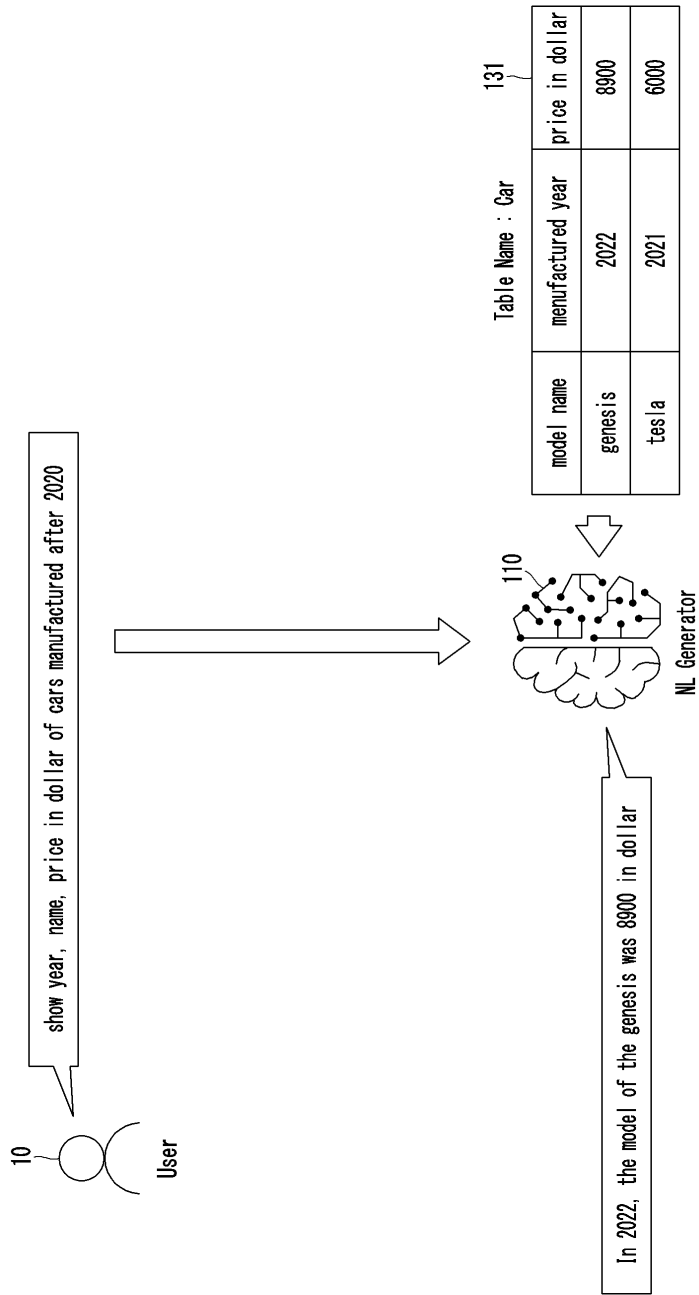
측면들은 또한 상응하는 블록 또는 아이템 또는 상응하는 장치의 특징으로 나타낼 수 있다. 방법 단계들의 몇몇 또는 전부는 예를 들어, 마이크로프로세서, 프로그램 가능한 컴퓨터 또는 전자 회로와 같은 하드웨어 장치에 의해(또는 이용하여) 수행될 수 있다. 몇몇의 실시 예에서, 가장 중요한 방법 단계들의 적어도 하나 이상은 이와 같은 장치에 의해 수행될 수 있다.

[0137] 실시예들에서, 프로그램 가능한 로직 장치(예를 들어, 필드 프로그래머블 게이트 어레이)가 여기서 설명된 방법들의 기능의 일부 또는 전부를 수행하기 위해 사용될 수 있다. 실시예들에서, 필드 프로그래머블 게이트 어레이(field-programmable gate array)는 여기서 설명된 방법들 중 하나를 수행하기 위한 마이크로프로세서(microprocessor)와 함께 작동할 수 있다. 일반적으로, 방법들은 어떤 하드웨어 장치에 의해 수행되는 것이 바람직하다.

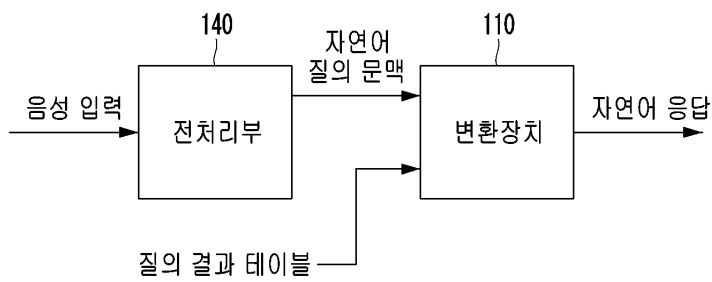
[0138] 이상 본 발명의 바람직한 실시 예를 참조하여 설명하였지만, 해당 기술 분야의 숙련된 당업자는 하기의 특허 청구의 범위에 기재된 본 발명의 사상 및 영역으로부터 벗어나지 않는 범위 내에서 본 발명을 다양하게 수정 및 변경시킬 수 있음을 이해할 수 있을 것이다.

도면

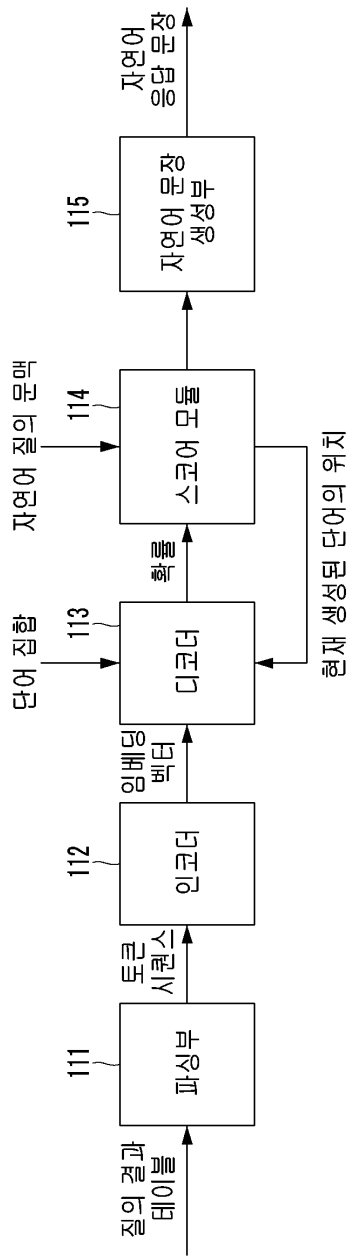
도면1



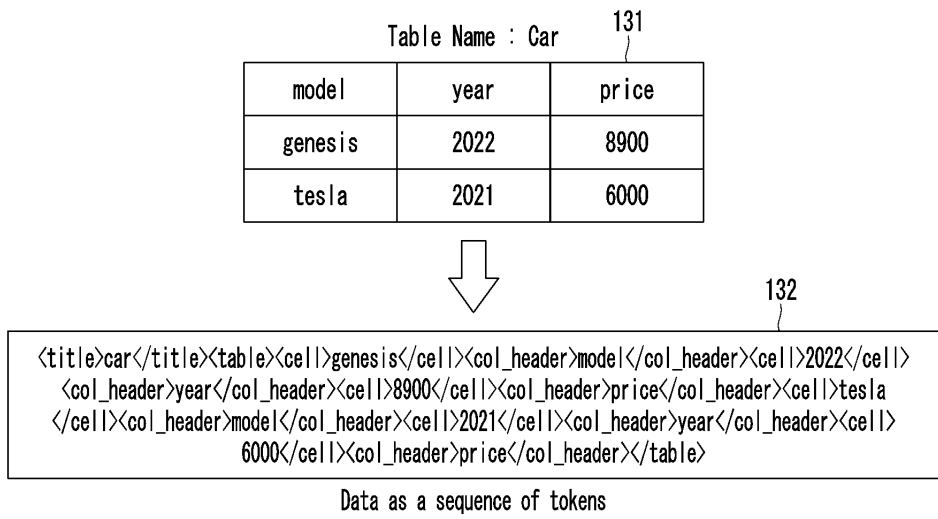
도면2



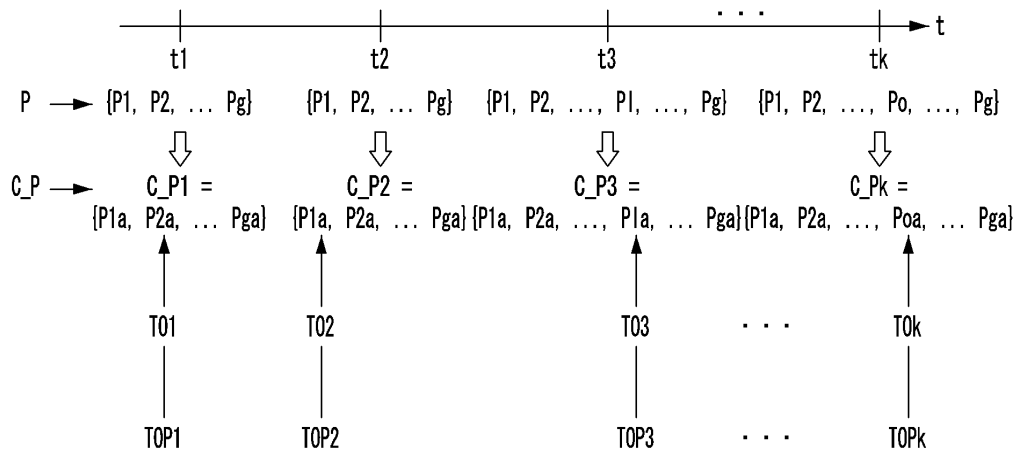
도면3



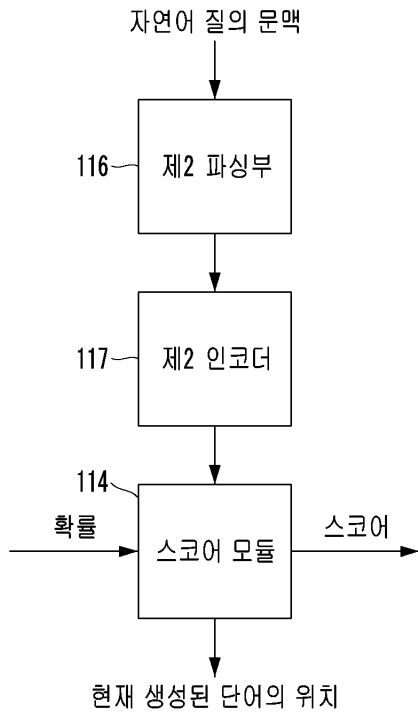
도면4



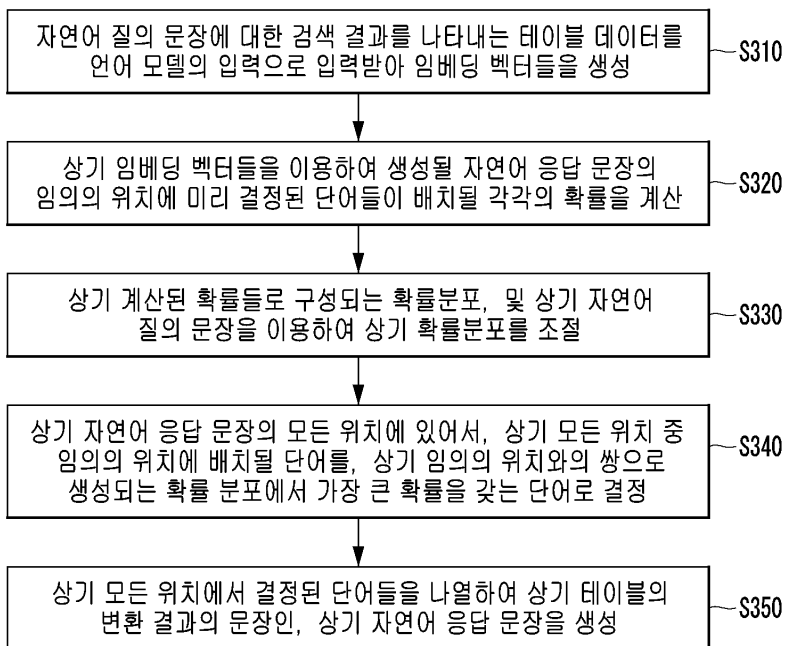
도면5



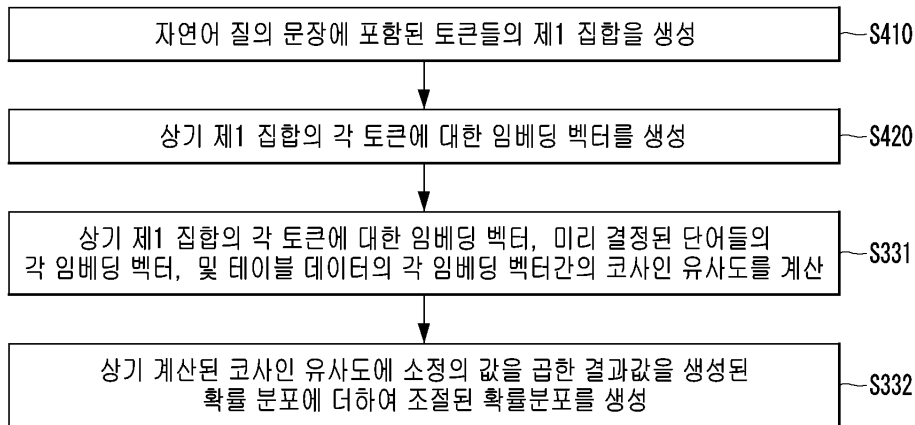
도면6



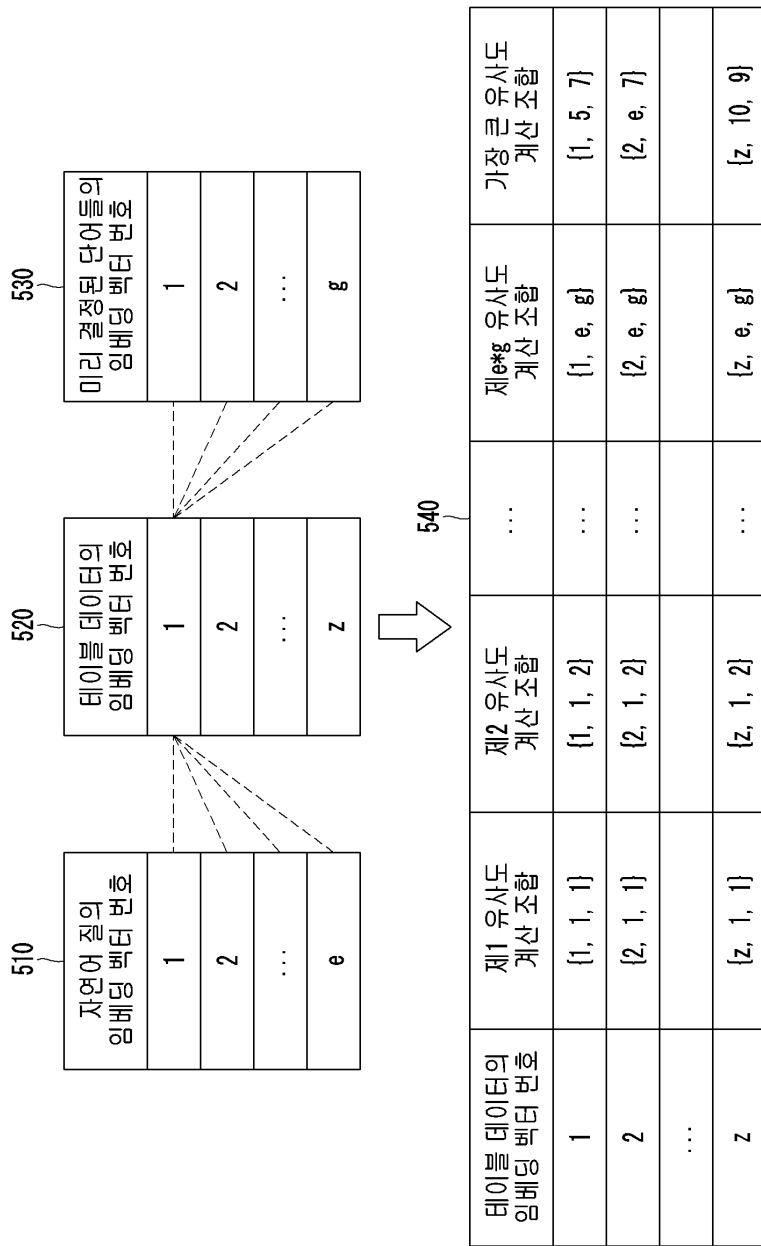
도면7



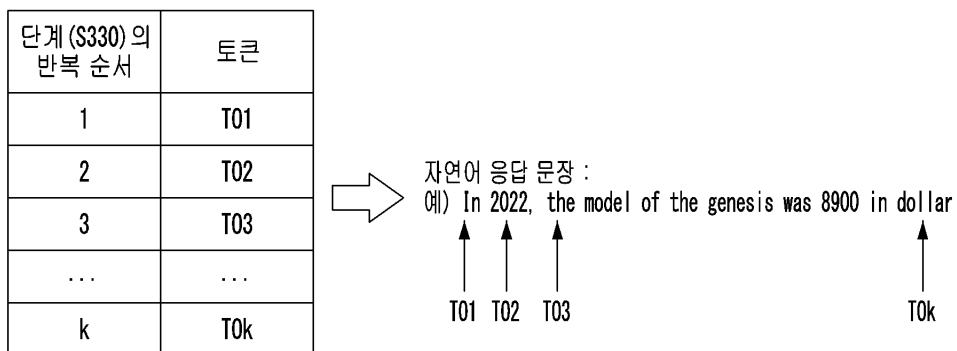
도면8



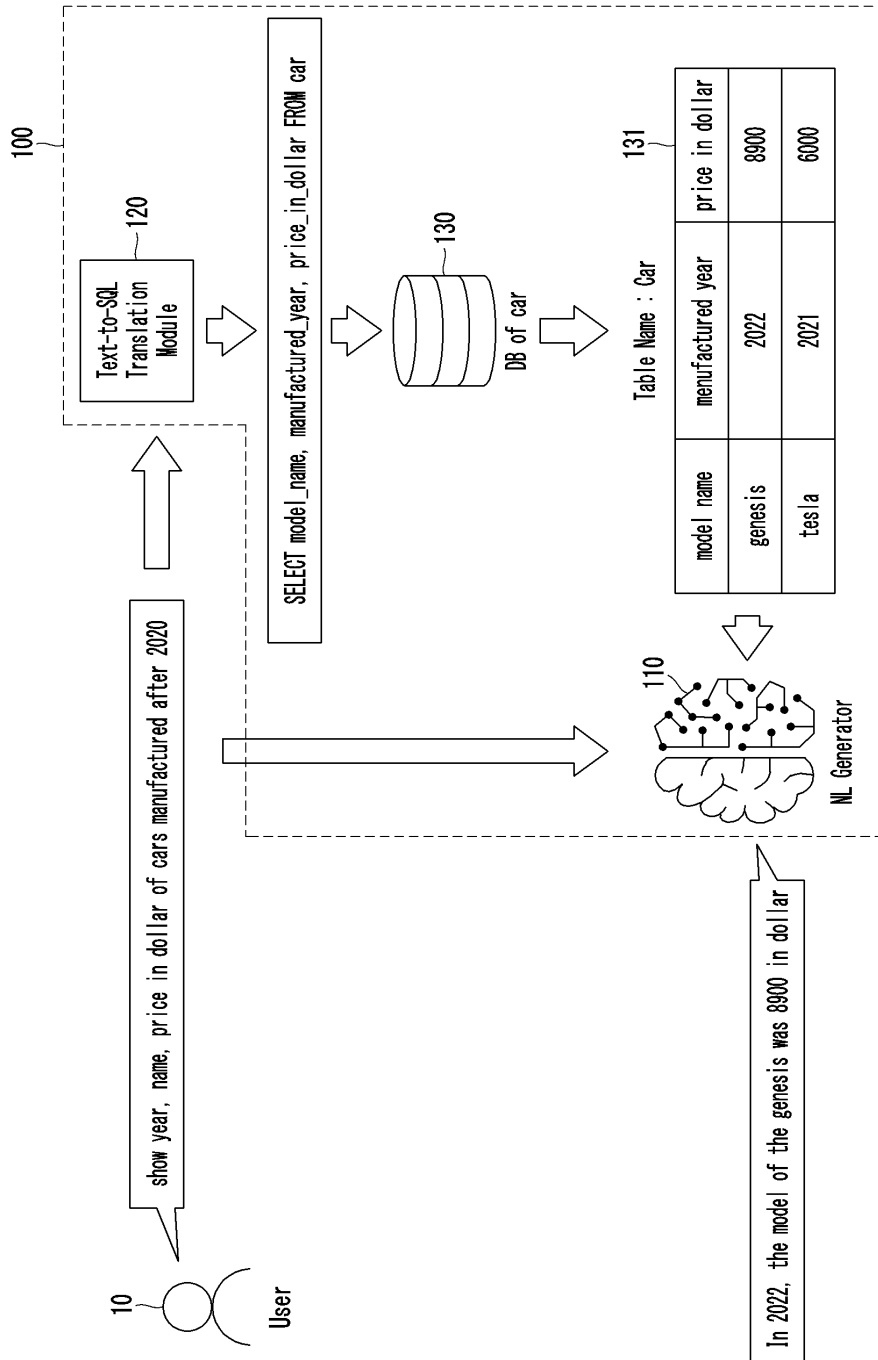
도면9



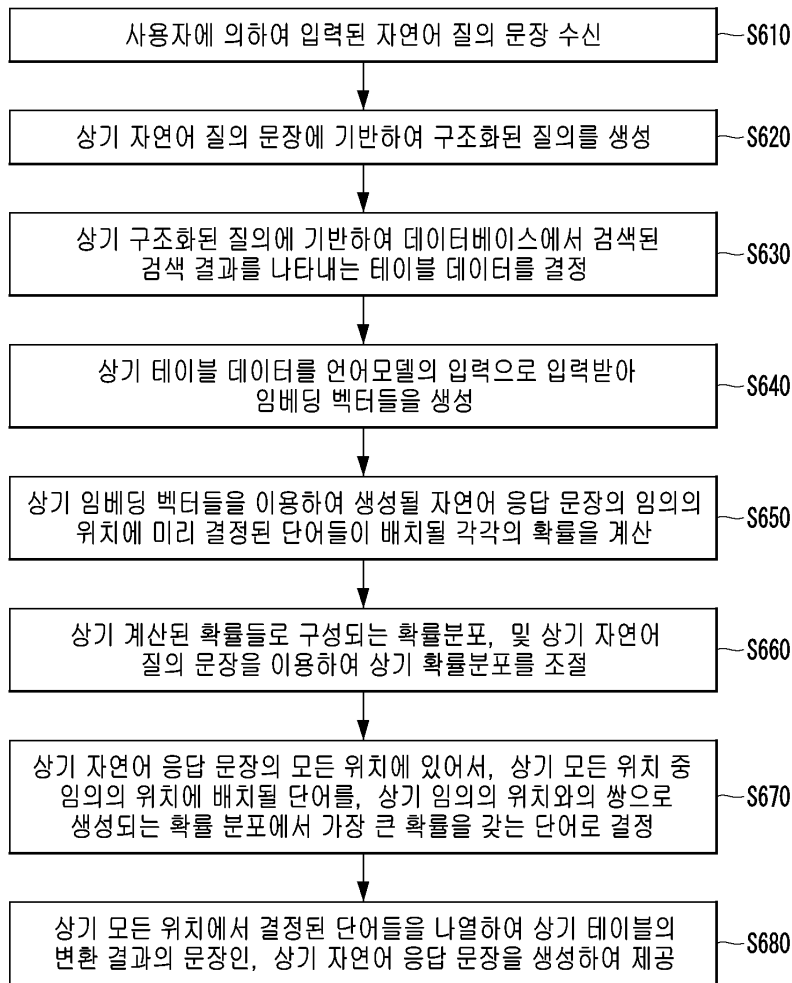
도면10



도면11



도면12



도면13

