



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0166613
(43) 공개일자 2023년12월07일

(51) 국제특허분류(Int. Cl.)
G06F 16/332 (2019.01) G06F 16/338 (2019.01)
G06F 3/16 (2018.01) G06N 20/00 (2019.01)
G06N 3/04 (2023.01) G10L 15/22 (2006.01)
(52) CPC특허분류
G06F 16/3329 (2019.01)
G06F 16/338 (2019.01)
(21) 출원번호 10-2022-0066802
(22) 출원일자 2022년05월31일
심사청구일자 2022년05월31일

(71) 출원인
경희대학교 산학협력단
경기도 용인시 기흥구 덕영대로 1732 (서천동, 경희대학교 국제캠퍼스내)
(72) 발명자
박성배
서울특별시 송파구 오금로 396, 1402호 (가락동, 에스케이허브파크)
박규민
경기도 수원시 영통구 청명남로12번길 13, 401호 (영통동)
(뒷면에 계속)
(74) 대리인
유미특허법인

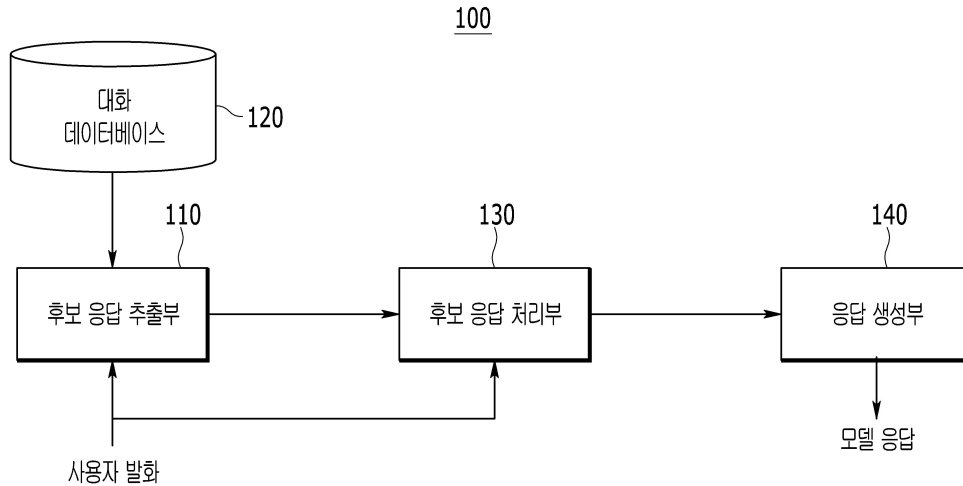
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 하이브리드 대화 생성 시스템 및 이를 이용한 후보 응답 검색 방법

(57) 요약

적어도 하나의 프로세서에 의해 동작하는 하이브리드 대화 생성 시스템으로서, 질의와 응답이 쌍을 이루는 복수의 질의-응답 쌍들이 저장되어 있는 대화 데이터베이스, 그리고 프로세서를 포함하고, 프로세서는 입력된 사용자 발화에 대한 유망 응답을 생성하여 사용자 발화와 유망 응답을 쌍으로 하는 사용자 발화 대화 쌍을 생성하고, 상기 복수의 질의-응답 쌍들과 상기 사용자 발화 대화 쌍의 유사도를 기초로 복수의 후보 대화 쌍들을 검색하며, 검색한 후보 대화 쌍들로부터 생성된 복수의 후보 응답 쌍들을 기초로 상기 사용자 발화에 대응하는 응답을 생성한다.

대표도



(52) CPC특허분류

G06F 3/167 (2013.01)
G06N 20/00 (2021.08)
G06N 3/04 (2023.01)
G10L 15/04 (2013.01)
G10L 15/22 (2013.01)
G10L 2015/225 (2013.01)
Y10S 715/978 (2013.01)

김소연

경기도 용인시 기흥구 서천동로21번길 6-1, 301호
 (서천동)

(72) 발명자

이준범

경기도 용인시 기흥구 서천동로21번길 6-1, 301호
 (서천동)

김봉민

경기도 용인시 기흥구 서그내로53번길 3-17, 304호
 (서천동, 서천빌리지)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711159756
과제번호	20220147
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	혁신성장동력프로젝트(과기정통부)
연구과제명	WiseKB: 빅데이터 이해 기반 자가학습형 지식베이스 및 추론 기술 개발
기 여 율	1/1
과제수행기관명	주식회사 솔트룩스
연구기간	2022.01.01~2023.02.28

명세서

청구범위

청구항 1

적어도 하나의 프로세서에 의해 동작하는 하이브리드 대화 생성 시스템으로서,

질의와 응답이 쌍을 이루는 복수의 질의-응답 쌍들이 저장되어 있는 대화 데이터베이스, 그리고

프로세서

를 포함하고,

상기 프로세서는,

입력된 사용자 발화에 대한 유망 응답을 생성하여 사용자 발화와 유망 응답을 쌍으로 하는 사용자 발화 대화 쌍을 생성하고, 상기 복수의 질의-응답 쌍들과 상기 사용자 발화 대화 쌍의 유사도를 기초로 복수의 후보 대화 쌍들을 검색하며, 검색한 후보 대화 쌍들로부터 생성된 복수의 후보 응답 쌍들을 기초로 상기 사용자 발화에 대응하는 응답을 생성하는, 하이브리드 대화 생성 시스템.

청구항 2

제1항에 있어서,

상기 프로세서는,

상기 입력된 사용자 발화를 유망 응답 생성 모델로 입력하여, 상기 사용자 발화에 대한 유망 응답을 생성하는, 하이브리드 대화 생성 시스템.

청구항 3

제2항에 있어서,

상기 유망 응답 생성 모델은 사용자 발화가 입력되면 응답이 출력되도록 발화와 응답이 쌍으로 구성된 학습 데이터로 학습된, 하이브리드 대화 생성 시스템.

청구항 4

제3항에 있어서,

상기 프로세서는,

상기 복수의 후보 대화 쌍들에 포함되어 있는 복수의 질의-응답들에서, 상기 복수의 질의 위치에 상기 사용자 발화를 삽입하여 상기 복수의 후보 응답 쌍들을 생성하는, 하이브리드 대화 생성 시스템.

청구항 5

제3항에 있어서,

상기 프로세서는,

상기 입력된 사용자 발화를 검색 모델로 입력하여 상기 복수의 후보 응답 쌍들을 생성하고, 상기 복수의 후보 응답 쌍들을 생성 모델로 입력하여 상기 사용자 발화에 대응하는 응답을 생성하는, 하이브리드 대화 생성 시스템.

청구항 6

제4항에 있어서,

상기 생성 모델은, 상기 복수의 후보 응답 쌍들에 포함된 복수의 답변들을 기초로 상기 사용자 발화에 대응하는 응답을 생성하는 하이브리드 대화 생성 시스템.

청구항 7

하이브리드 대화 생성 시스템의 동작 방법으로서,

사용자 발화를 입력 받아, 학습된 유망 응답 생성기를 이용하여 상기 사용자 발화에 대한 유망 응답을 생성하는 단계,

상기 사용자 발화와 상기 유망 응답이 쌍을 이루는 사용자 발화 대화 쌍을 대화 데이터베이스에 기 등록된 질의-응답 쌍들과 유사도를 계산하여, 복수의 후보 대화 쌍들을 검색하는 단계,

각 후보 대화 쌍의 질의 부분을 상기 사용자 발화로 교체하여 복수의 후보 응답 쌍들을 생성하는 단계, 그리고

상기 후보 응답 쌍들을 학습된 생성 모델을 이용하여 상기 질의에 대한 응답을 생성하는 단계

를 포함하는, 동작 방법.

청구항 8

제7항에 있어서,

상기 유망 응답을 생성하는 단계 이전에,

학습 질의와 응답이 쌍으로 구성된 학습 데이터를 기초로, 상기 유망 응답 생성기에 상기 학습 질의가 입력되면 상기 쌍으로 구성된 응답이 출력되도록 상기 유망 응답 생성기를 학습시키는 단계

를 포함하는, 동작 방법.

청구항 9

제7항에 있어서,

상기 입력된 사용자 발화는 검색 모델로 입력되어 상기 복수의 후보 응답 쌍들을 생성하고,

상기 복수의 후보 응답 쌍들을 생성 모델로 입력하여, 상기 복수의 후보 응답 쌍들에 포함된 복수의 답변들을 기초로 상기 사용자 발화에 대응하는 응답을 생성하는, 동작 방법.

청구항 10

제7항에 있어서,

상기 생성 모델은 BART(Bidirectional Encoder Representation from Transformer) 모델인, 동작 방법.

발명의 설명

기술 분야

[0001] 본 발명은 하이브리드 대화 생성 모델 내의 검색 기반 모델이 양질의 후보 응답을 검색하는 하이브리드 대화 생성 시스템 및 이를 이용한 후보 응답 검색 방법에 관한 것이다.

배경 기술

[0002] 검색 기반의 대화 모델은 사용자 발화가 입력되면 대화 데이터베이스 안에 있는 여러 개의 질의-응답 쌍에서, 사용자 발화와 가장 유사성이 높은 질의와 쌍으로 이어지는 응답을 모델의 출력으로 제공한다. 기존의 검색 방법은 크게 두 가지로 볼 수 있다.

[0003] 첫 번째 방법은 사용자 발화가 입력되면, 데이터베이스 내의 후보 질의-응답 쌍에서 사용자 발화와 후보 질의와의 유사성을 계산하고, 높은 값을 갖는 후보 질의에 이어지는 응답을 모델의 출력으로 사용하는 방법이다. 이는 사용자 발화와 후보 질의가 거의 일치하면 양질의 후보 응답을 검색할 수 있다. 그러나, 후보 질의가 사용자 발화의 다른 부분을 해결하려 할 때, 검색된 응답이 사용자 발화에 대한 응답과 매우 달라지는 문제점을 갖는다.

[0004] 두 번째 방법은 사용자 발화와 데이터베이스 내의 후보 응답과의 유사성을 계산하는 방법이다. 이 검색 방법은 사용자 발화와 후보 질의 사이의 오버랩이 많아질수록, 양질의 응답을 검색할 가능성이 증가할 것이라 가정하여 검색하는 방법이다.

[0005] 그러나 둘 사이의 오버랩이 많다는 것이 항상 양질의 응답 검색을 보장하지는 않는다. 그렇기에 상술한 두 가지 방법 모두 데이터베이스를 구성하고 있는 후보 질의-응답 쌍에 검색 성능이 의존적이라는 문제점을 갖는다.

발명의 내용

해결하려는 과제

[0006] 따라서, 본 발명은 검색 기반의 대화 모델과 생성 기반의 대화 모델을 함께 사용하는 하이브리드 대화 생성 모델을 이용하고, 모델의 학습 단계에서 사용 가능한 실제 응답을 활용하여 양질의 후보 응답을 검색하는 하이브리드 대화 생성 시스템 및 이를 이용한 후보 응답 검색 방법을 제공한다.

과제의 해결 수단

[0007] 상기 본 발명의 기술적 과제를 달성하기 위한 본 발명의 하나의 특징인 적어도 하나의 프로세서에 의해 동작하는 하이브리드 대화 생성 시스템으로서,

[0008] 질의와 응답이 쌍을 이루는 복수의 질의-응답 쌍들이 저장되어 있는 대화 데이터베이스, 그리고 프로세서를 포함하고, 상기 프로세서는, 입력된 사용자 발화에 대한 유망 응답을 생성하여 사용자 발화와 유망 응답을 쌍으로 하는 사용자 발화 대화 쌍을 생성하고, 상기 복수의 질의-응답 쌍들과 상기 사용자 발화 대화 쌍의 유사도를 기초로 복수의 후보 대화 쌍들을 검색하며, 검색한 후보 대화 쌍들로부터 생성된 복수의 후보 응답 쌍들을 기초로 상기 사용자 발화에 대응하는 응답을 생성한다.

[0009] 상기 프로세서는, 상기 입력된 사용자 발화를 유망 응답 생성 모델로 입력하여, 상기 사용자 발화에 대한 유망 응답을 생성할 수 있다.

[0010] 상기 유망 응답 생성 모델은 사용자 발화가 입력되면 응답이 출력되도록 발화와 응답이 쌍으로 구성된 학습 데이터로 학습될 수 있다.

[0011] 상기 프로세서는, 상기 복수의 후보 대화 쌍들에 포함되어 있는 복수의 질의-응답들에서, 상기 복수의 질의 위치에 상기 사용자 발화를 삽입하여 상기 복수의 후보 응답 쌍들을 생성할 수 있다.

[0012] 상기 프로세서는, 상기 입력된 사용자 발화를 검색 모델로 입력하여 상기 복수의 후보 응답 쌍들을 생성하고, 상기 복수의 후보 응답 쌍들을 생성 모델로 입력하여 상기 사용자 발화에 대응하는 응답을 생성할 수 있다.

[0013] 상기 생성 모델은, 상기 복수의 후보 응답 쌍들에 포함된 복수의 답변들을 기초로 상기 사용자 발화에 대응하는 응답을 생성할 수 있다.

[0014] 상기 본 발명의 기술적 과제를 달성하기 위한 본 발명의 또 다른 특징인 하이브리드 대화 생성 시스템의 동작 방법으로서,

[0015] 사용자 발화를 입력 받아, 학습된 유망 응답 생성기를 이용하여 상기 사용자 발화에 대한 유망 응답을 생성하는 단계, 상기 사용자 발화와 상기 유망 응답이 쌍을 이루는 사용자 발화 대화 쌍을 대화 데이터베이스에 기록된 질의-응답 쌍들과 유사도를 계산하여, 복수의 후보 대화 쌍들을 검색하는 단계, 각 후보 대화 쌍의 질의 부분을 상기 사용자 발화로 교체하여 복수의 후보 응답 쌍들을 생성하는 단계, 그리고 상기 후보 응답 쌍들을 학습된 생성 모델을 이용하여 상기 질의에 대한 응답을 생성하는 단계를 포함한다.

[0016] 상기 유망 응답을 생성하는 단계 이전에, 학습 질의와 응답이 쌍으로 구성된 학습 데이터를 기초로, 상기 유망 응답 생성기에 상기 학습 질의가 입력되면 상기 쌍으로 구성된 응답이 출력되도록 상기 유망 응답 생성기를 학습시키는 단계를 포함할 수 있다.

[0017] 상기 입력된 사용자 발화는 검색 모델로 입력되어 상기 복수의 후보 응답 쌍들을 생성하고, 상기 복수의 후보 응답 쌍들을 생성 모델로 입력하여, 상기 복수의 후보 응답 쌍들에 포함된 복수의 답변들을 기초로 상기 사용자 발화에 대응하는 응답을 생성할 수 있다.

[0018] 상기 생성 모델은 BART(Bidirectional Encoder Representation from Transformer) 모델일 수 있다.

발명의 효과

[0019] 본 발명에 따르면, 하이브리드 대화 생성 모델의 검색 기반 모델이 학습 단계에서 사용 가능한 실제 응답을 활용하여 후보 응답을 검색함으로써, 양질의 후보 응답을 검색할 수 있고 대화 생성 모델의 성능을 높일 수 있다.

도면의 간단한 설명

[0020] 도 1은 본 발명의 실시예에 따른 하이브리드 대화 생성 시스템의 예시도이다.

도 2는 본 발명의 실시예에 따른 후보 응답 추출부의 예시도이다.

도 3은 본 발명의 실시예에 따라 하이브리드 대화 생성 시스템이 후보 응답을 검색하는 방법에 대한 흐름도이다.

도 4는 본 발명의 실시예에 따른 컴퓨팅 장치의 구조도이다.

발명을 실시하기 위한 구체적인 내용

[0021] 아래에서는 첨부한 도면을 참고로 하여 본 발명의 실시예에 대하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

[0022] 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 있는 것을 의미한다.

[0023] 이하, 도면을 참조로 하여 본 발명의 실시예에 따른 하이브리드 대화 생성 시스템 및 이를 이용한 후보 응답 검색 방법에 대해 상세히 설명한다.

[0024] 도 1은 본 발명의 실시예에 따른 하이브리드 대화 생성 시스템의 예시도이다.

[0025] 도 1에 도시된 바와 같이, 하이브리드 대화 생성 시스템(100)은 후보 응답 추출부(110), 대화 데이터베이스(120), 후보 응답 처리부(130), 그리고 응답 생성부(140)를 포함한다. 여기서, 후보 응답 추출부(110)는 검색 모델로 구현되고, 응답 생성부(140)는 생성 모델로 구현되어, 하이브리드 대화 생성 모델로 표현된다.

[0026] 후보 응답 추출부(110)는 사용자 발화가 입력되면, 사용자 발화에 대응하는 유망 응답을 생성한다. 후보 응답 추출부(110)는 사용자 발화를 질의로, 생성한 유망 응답을 답변으로 하는 사용자 질의-답변 쌍(이하, 설명의 편의를 위하여, '사용자 발화 대화 쌍'이라 지칭함)을 생성한다.

[0027] 여기서, 후보 응답 추출부(110)에는 사용자 발화에 대응하는 유망 응답을 생성하는 유망 응답 생성기가 포함되어 있다. 유망 응답 생성기는 사용자 발화에 대한 유망 응답을 생성하도록 학습 데이터로 기 학습되어 있다. 이때, 학습 데이터는 질의와 정답 답변의 쌍으로 구성되며, 이에 대해서는 이후 상세히 설명한다.

[0028] 또한, 후보 응답 추출부(110)는 입력된 사용자 발화를 기초로, 대화 데이터베이스(120)에 저장되어 있는 복수의 질의-답변 쌍들 중 사용자 발화와 유사한 후보 질의들과 각 후보 질의와 쌍들로 연결된 답변들을 후보 질의-답변 쌍(이하, 설명의 편의를 위하여 '후보 대화 쌍'이라 지칭함)들로 추출한다.

[0029] 후보 응답 처리부(130)는 후보 응답 추출부(110)에서 추출한 복수의 후보 대화 쌍들에서 후보 질의 부분을 사용자 발화로 교체하여, 사용자 발화와 후보 답변을 연결하여 복수의 사용자 발화-답변 쌍(이하, 설명의 편의를 위하여 '후보 응답 쌍'이라 지칭함)들을 생성한다. 본 발명의 실시예에서는 후보 응답 처리부(130)를 별도의 구성요소로 표현하였으나, 후보 응답 추출부(110)에서 후보 응답 처리부(130)의 기능을 수행할 수도 있다.

[0030] 응답 생성부(140)는 후보 응답 처리부(130)에서 생성한 복수의 후보 응답 쌍들을 수신하고, 수신한 복수의 후보 응답 쌍들에서 사용자에게 제공할 응답을 생성한다. 즉, 후보 응답 추출부(110)에서 검색된 후보 대화 쌍들은 후보 응답 처리부(130)에서 사용자 발화와 연결되어, 응답 생성부(140)의 인코더로 입력된다. 이렇게 입력된 후보 응답 쌍들에서, 응답 생성부(140)의 디코더는 응답을 생성하도록 학습되어 있다.

[0031] 예를 들어, 사용자 발화와 연결된 후보 응답 쌍들을, 응답 생성부(140)의 인코더가 벡터로 생성한다. 디코더는 인코더가 생성한 벡터를 사용하여 다음에 나올 단어가 무엇인지 확률적으로 계산함으로써, 복수의 답변들을 확

용하여 사용자 발화에 대한 응답을 생성한다. 응답 생성부(140)가 복수의 후보 응답 쌍들 중 복수의 답변들을 그대로 사용하지 않고 해당 답변들을 기초로 사용자에게 제공할 응답을 새로 생성하는 방법은 다양한 방식으로 이미 알려진 기술이므로, 본 발명의 실시예에서는 어느 하나의 방법으로 한정하지 않는다. 그리고 응답 생성부(140)를 구성하는 생성 모델은 BART(Bidirectional Encoder Representation from Transformer) 모델을 사용하는 것을 예로 하여 설명하며, 생성 모델은 검색된 후보 응답을 고려하여 다양한 응답 패턴을 학습할 수 있다.

[0032] 이와 같은 하이브리드 대화 생성 시스템(100)에서, 검색 모델로 구현된 후보 응답 추출부(110)에 대해 도 2를 참조로 설명한다.

[0033] 도 2는 본 발명의 실시예에 따른 후보 응답 추출부의 예시도이다.

[0034] 도 2에 도시된 바와 같이, 후보 응답 추출부(110)는 크게 사용자 응답을 사용할 수 있는 교사 모델(111)과 그렇지 않은 학생 모델(112)로 나뉜다. 교사 모델(111)은 사용자 발화와 사용자 응답 모두를 임베딩 레이어와 트랜스포머 인코더에 넣어 벡터로 만든다.

[0035] 즉, 사용자 발화와 대화 데이터베이스(120)에 등록된 복수의 질의들의 유사도를 계산하기 위해, 입력된 사용자 발화를 구성하는 토큰들을 단위, 위치, 속성을 기초로 임베딩하고, 토큰별로 임베딩된 벡터들을 합한다. 그리고, 트랜스포머 인코더를 기반으로 임베딩된 벡터에서 중요 정보들을 선택한다. 여기서, 임베딩 레이어와 트랜스포머 인코더의 기능은 이미 알려진 것으로, 본 발명의 실시예에서는 상세한 설명을 생략한다.

[0036] 한편, 학생 모델(112)의 경우는 사용자 발화만을 입력으로 받는다. 이때, 학생 모델(112)에는 임베딩 레이어(114)와 선형 레이어(115)로 구성된 유망 응답 생성기(113)가 포함되어 있다. 유망 응답 생성기(113)는 사용자 발화에 따라 유망한 사용자 응답 벡터를 생성하도록 학습된다.

[0037] 즉, 교사 모델(111)의 출력은 {사용자 발화 벡터, 사용자 응답 벡터}이고, 학생 모델(112)의 출력은 {사용자 발화 벡터, 유망 응답 벡터}이다. 각각의 출력은 임베딩된 후보 대화 쌍들과의 내적을 통해, 어느 후보 쌍이 각 모델의 출력 임베딩과 가장 유사도가 높은지 분포를 구할 수 있다.

[0038] 학생 모델(112)은 사용자 질의와 가장 유사한 쌍을 대화 데이터베이스(120)에 저장되어 있는 복수의 질의-답변 쌍들 중에서 검색하도록 학습되어 있다. 동시에 교사 모델(111)의 출력 분포를 답도록 학습되어 있다.

[0039] 즉, 후보 대화 쌍들 중 입력된 사용자 질의와 가장 관련성이 높은 쌍이 무엇인지 기 레이블링 되어 있다. 교사 모델(111)과 학생 모델(112)의 출력은 각 후보 대화 쌍들이 사용자 질의와 얼마나 유사한지에 대한 점수들로 표현된다. 따라서, KL 다이버전스(kullback leibler divergence)를 통해 학생 점수 분포가 교사 점수 분포를 답도록 학생 모델(112)을 학습시킨다.

[0040] 학생 모델(112)이 교사 모델(111)의 분포를 답도록 학습하는 과정에서 유망 응답 생성기(113)는 사용자 질의에 그럴듯한 사용자 응답 벡터를 생성하도록 학습된다. 학생 모델(112)에 대한 학습이 완료되면, 학생 모델(112)만을 후보 응답 추출부(110)로 사용한다.

[0041] 이와 같이 후보 응답 추출부(110)인 검색 모델이 학습되면, 사용자 발화가 입력될 때 후보 응답 추출부(110)은 대화 데이터베이스(120)에서 k개의 후보 대화 쌍들의 응답들을 검색할 수 있다.

[0042] 다음은 상술한 하이브리드 대화 생성 시스템이 후보 응답을 검색하여 사용자에게 제공하는 방법에 대해 도 3을 참조로 설명한다.

[0043] 도 3은 본 발명의 실시예에 따라 하이브리드 대화 생성 시스템이 후보 응답을 검색하는 방법에 대한 흐름도이다.

[0044] 본 발명의 실시예에 따른 검색 방법을 설명하기 앞서, 기존 검색 모델의 검색 방법을 본 발명의 실시예와 비교하기 위해, 다음 표 1과 같이 사용자 발화와 후보 질의, 후보 응답이 있다고 가정한다.

표 1

[0045]

사용자 발화	불고기피자 한 판 얼마입니까?
후보 질의와의 비교 (①)	
후보 질의	불고기피자 한 판 팔면 얼마 남아?
후보 응답	불고기피자 얼마 안 남아.
후보 응답과의 비교 (②)	

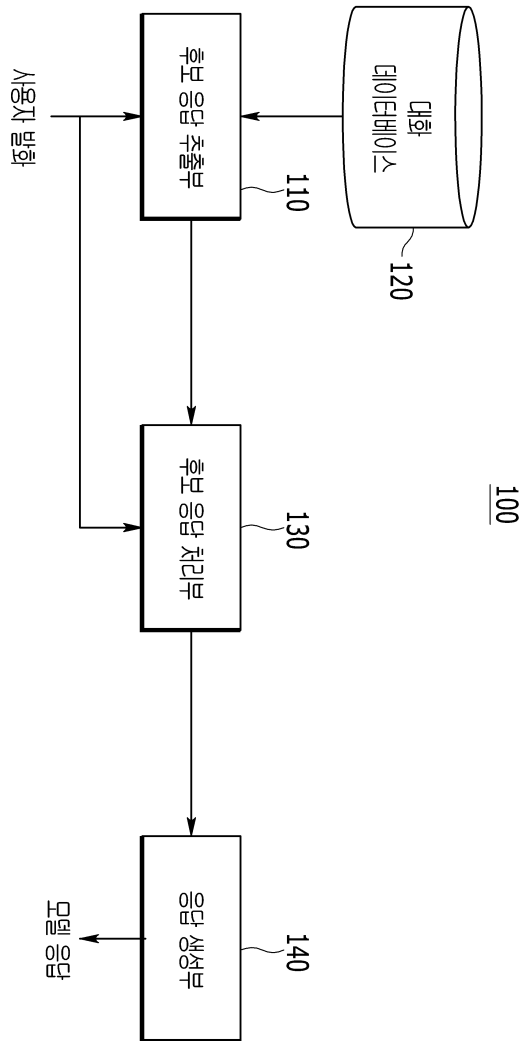
후보 질의 후보 응답	게임 한 판 하는데 얼마입니까? 한 판에 천 얼마 정도해요.
후보 질의 및 응답 모두 비교 (③)	
후보 질의 후보 응답	불고기피자 비싼가요? 비싸진 않고 만원 정도 합니다.

- [0046] 기존의 검색 방법은 입력된 사용자 발화에 대해 데이터베이스 내의 있는 후보 질의(①) 또는 후보 응답(②)과 비교한다. 즉, '불고기피자 한 판 얼마입니까?'라는 사용자 발화가 입력되면, 기존의 검색 방법 중 후보 질의와의 비교(①) 방법을 이용할 경우, '불고기피자 한 판 팔면 얼마 남아?'와 같은 후보 질의에 높은 유사도 점수를 준다. 그러나, 후보 질의가 사용자 발화의 다른 부분을 해결하려 하기에, 상관없는 후보 응답인 '불고기피자 얼마 안 남아.'를 검색하게 된다.
- [0047] 한편, 기존의 검색 방법 중 후보 응답과의 비교(②) 방법을 사용할 경우, '한 판에 천 얼마 정도해요.'라는 후보 응답에 높은 유사도 점수가 부여된다. 이 또한, 단어간 중복이 많더라도 잘못된 응답을 검색할 가능성이 있다.
- [0048] 본 발명의 실시예에서는 후보 질의와 응답을 모두 비교(③)함으로써, 기존의 검색 방법에서 발생할 수 있는 문제점을 보완하여, 알맞은 후보 응답을 검색할 수 있다.
- [0049] 즉, 도 3에 도시된 바와 같이, 하이브리드 대화 생성 시스템(100)은 먼저 사용자의 발화가 입력되면 유망 응답이 출력되도록 학습 데이터로 유망 응답 생성기(113)를 학습시킨다(S100). 이때, 교사 모델(111)의 출력 점수 분포와 학생 모델(112)의 점수 분포가 유사해지도록 학습 데이터로 유망 응답 생성기(113)를 학습시킨다.
- [0050] 하이브리드 대화 생성 시스템(100)은 사용자 발화를 수신하면(S110), 사용자 발화에 대한 유망 응답을 유망 응답 생성기(113)를 통해 생성한다(S120). 하이브리드 대화 생성 시스템(100)은 사용자 발화와 S120 단계에서 생성한 유망 응답을 쌍으로 하는 사용자 발화 대화 쌍을 생성한다(S130).
- [0051] 하이브리드 대화 생성 시스템(100)은 S130 단계에서 생성한 사용자 발화 대화 쌍과 대화 데이터베이스(120)에 등록된 복수의 질의-응답 쌍들의 유사도를 계산하고, 유사도가 높은 복수의 후보 대화 쌍들을 검색한다(S140). 하이브리드 대화 생성 시스템(100)의 후보 응답 처리부(130)는 후보 응답 추출부(110)가 검색한 복수의 후보 대화 쌍들에서 질의 부분을 S110 단계에서 수신한 사용자 발화로 교체하여 복수의 후보 응답 쌍들로 생성한다(S150).
- [0052] 하이브리드 대화 생성 시스템(100)의 응답 생성부(140)는 S150 단계에서 생성된 복수의 후보 응답 쌍들을 기초로 S110 단계에서 수신한 사용자 발화에 대한 응답을 생성한다(S160). 이를 위해, 응답 생성부(140)는 BART 모델을 사용하여 검색된 복수의 후보 응답 쌍들에 포함된 후보 응답들을 고려하여 다양한 응답 패턴을 생성하도록 학습되어 있다.
- [0053] 도 4는 본 발명의 실시예에 따른 컴퓨팅 장치의 구조도이다.
- [0054] 도 4에 도시된 바와 같이, 하이브리드 대화 생성 시스템(100)은 적어도 하나의 프로세서에 의해 동작하는 컴퓨팅 장치(200)에서, 본 발명의 동작을 실행하도록 기술된 명령들(instructions)이 포함된 프로그램을 실행한다.
- [0055] 컴퓨팅 장치(200)의 하드웨어는 적어도 하나의 프로세서(210), 메모리(220), 스토리지(230), 통신 인터페이스(240)를 포함할 수 있고, 버스를 통해 연결될 수 있다. 이외에도 입력 장치 및 출력 장치 등의 하드웨어가 포함될 수 있다. 컴퓨팅 장치(200)는 프로그램을 구동할 수 있는 운영 체제를 비롯한 각종 소프트웨어가 탑재될 수 있다.
- [0056] 프로세서(210)는 컴퓨팅 장치(200)의 동작을 제어하는 장치로서, 프로그램에 포함된 명령들을 처리하는 다양한 형태의 프로세서(210)일 수 있고, 예를 들면, CPU(Central Processing Unit), MPU(Micro Processor Unit), MCU(Micro Controller Unit), GPU(Graphic Processing Unit) 등 일 수 있다.
- [0057] 메모리(220)는 본 발명의 동작을 실행하도록 기술된 명령들이 프로세서(210)에 의해 처리되도록 해당 프로그램을 로드한다. 메모리(220)는 예를 들면, ROM(read only memory), RAM(random access memory) 등 일 수 있다. 스토리지(230)는 본 발명의 동작을 실행하는데 요구되는 각종 데이터, 프로그램 등을 저장한다. 통신 인터페이스(240)는 유/무선 통신부일 수 있다.
- [0058] 이상에서 본 발명의 실시예에 대하여 상세하게 설명하였지만 본 발명의 권리범위는 이에 한정되는 것은 아니고

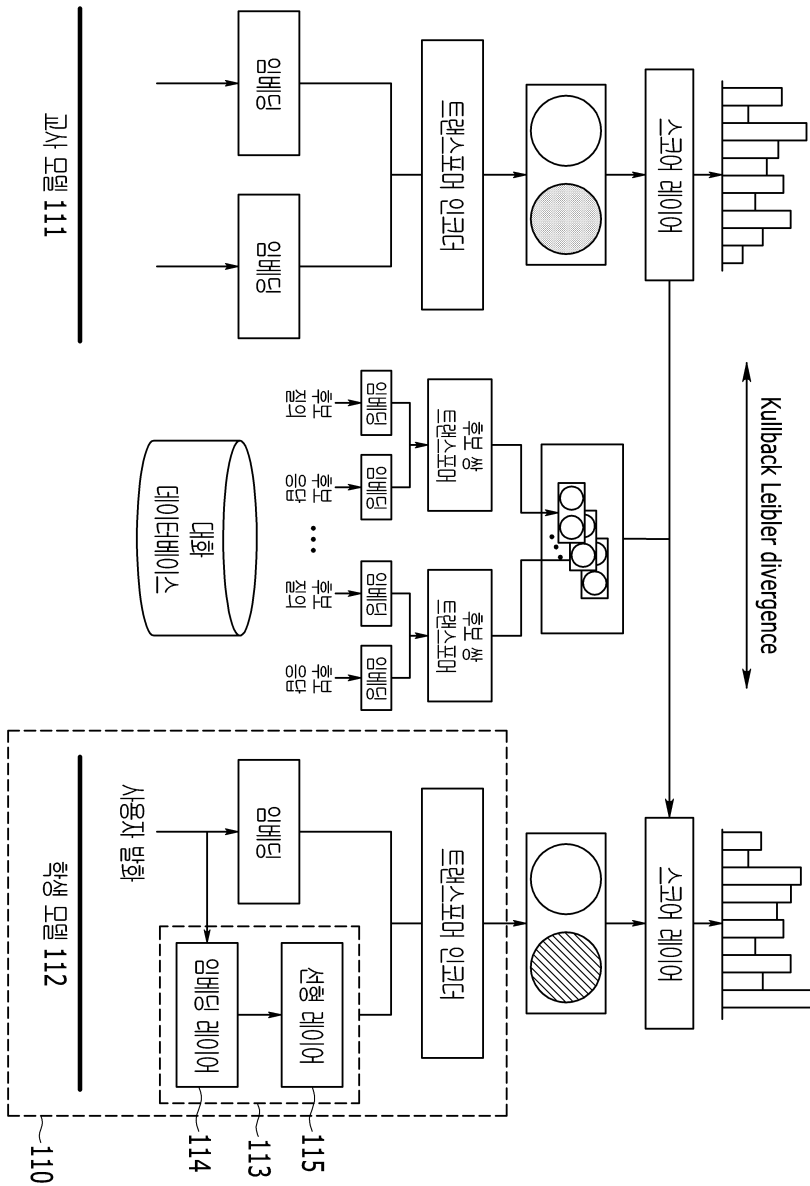
다음의 청구범위에서 정의하고 있는 본 발명의 기본 개념을 이용한 당업자의 여러 변형 및 개량 형태 또한 본 발명의 권리범위에 속하는 것이다.

도면

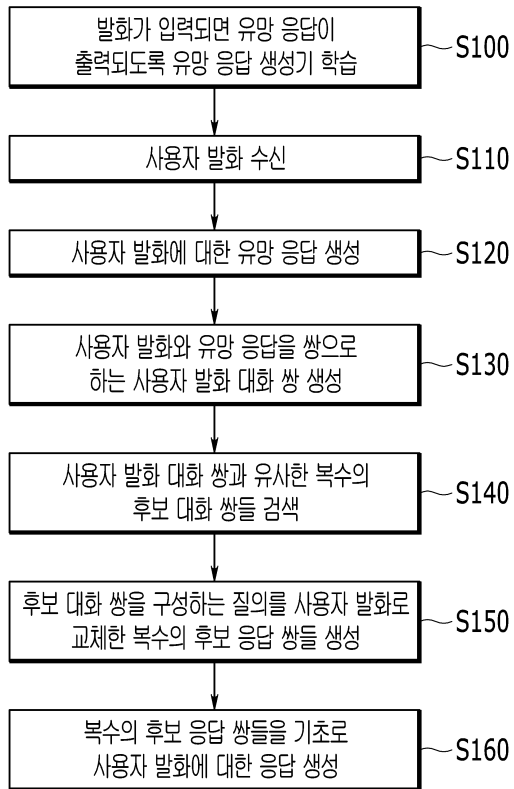
도면1



도면2



도면3



도면4

