



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0000216
(43) 공개일자 2020년01월02일

(51) 국제특허분류(Int. Cl.)
G06F 16/00 (2019.01)
(52) CPC특허분류
G06F 16/3346 (2019.01)
G06F 16/3329 (2019.01)
(21) 출원번호 10-2018-0072234
(22) 출원일자 2018년06월22일
심사청구일자 2018년06월22일

(71) 출원인
서강대학교산학협력단
서울특별시 마포구 백범로 35 (신수동, 서강대학교)
(72) 발명자
구명완
서울특별시 양천구 목동동로 100, 1312동 1304호 (신정동, 목동13단지아파트)
김병재
서울특별시 마포구 대흥동 264-3 106호
서정연
서울특별시 서초구 서초중앙로 260 B동 203호 (반포동, 삼호가든3차)
(74) 대리인
이지연

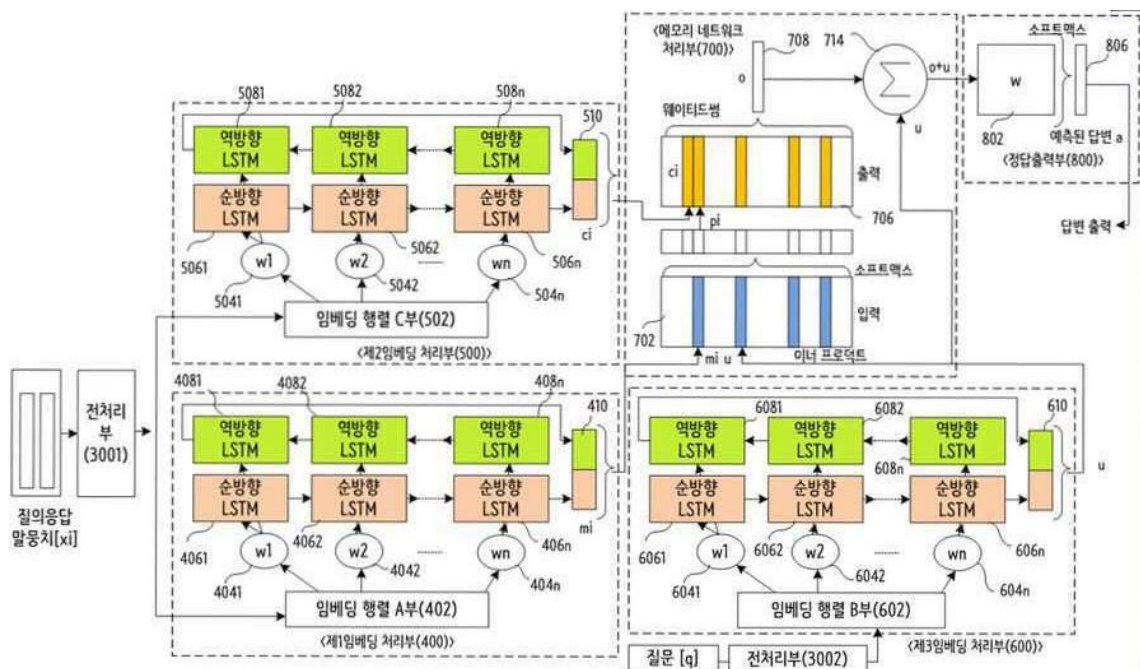
전체 청구항 수 : 총 5 항

(54) 발명의 명칭 단어자질을 강화한 음성 대화 방법 및 시스템

(57) 요약

본 발명에 따르는 단어자질을 강화한 음성 대화 방법은, 다수의 문장으로 구성되는 질의 응답 말뭉치를 입력받아 다수의 문장 각각에 대해 단어들로 구성되는 제1메타 데이터를 생성하는 단계; 상기 다수의 문장 각각에 대한 제1메타 데이터를 입력받아, 상기 제1메타 데이터의 단어들 각각에 대해 제1임베딩 벡터를 부가하여 제1단어벡터들 (뒷면에 계속)

대표도



을 생성하고, 상기 제1단어벡터들을 양방향 LSTM 처리하여 제1문장 임베딩 정보를 생성하고, 상기 제1메타 데이터의 단어들 각각에 대해 제2임베딩 벡터를 추가하여 제2단어벡터들을 생성하고, 상기 제2단어벡터들을 양방향 LSTM 처리하여 제2문장 임베딩 정보를 생성하는 단계; 질문 문장에 대해 단어들로 구성되는 제2메타 데이터를 생성하고, 상기 제2메타 데이터의 단어들 각각에 대해 제3임베딩 벡터를 추가하여 제3단어벡터들을 생성하고, 상기 제3단어벡터들을 양방향 LSTM 처리하여 질문문장 임베딩 정보를 생성하는 단계; 상기 다수의 문장들 각각에 대한 제1문장 임베딩 정보와 상기 질문문장 임베딩 정보를 입력받아, 상기 다수의 문장들 각각에 대한 제1문장 임베딩 정보와 상기 질문 임베딩 정보 사이의 관련성을 나타내는 제1정보를 생성하는 단계; 상기 다수의 문장들 각각에 대한 제1정보와 상기 다수의 문장들 각각에 대한 제2문장 임베딩 정보의 웨이트드 썸을 이행하여 제2정보를 생성하는 단계; 상기 다수의 문장들 각각에 대해 제2정보와 상기 질문정보를 가산하여 제3정보를 출력하는 단계; 및 상기 제3정보를 통해 상기 다수의 문장과 질문정보 사이의 확률을 산출하고, 가장 높은 확률을 가진 문장을 답변 문장으로 예측하여 출력하는 단계;를 포함하는 것을 특징으로 한다.

(52) CPC특허분류

G06F 16/3347 (2019.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	10063424
부처명	산업통상자원부
연구관리전문기관	한국산업기술평가관리원
연구사업명	산업핵심기술개발사업
연구과제명	실내용 음성대화 로봇을 위한 원거리 음성인식 기술 및 멀티 태스크 대화처리 기술 개발
기 여 율	1/1
주관기관	KAIST
연구기간	2016.05.01 ~ 2020.04.30

명세서

청구범위

청구항 1

단어자질을 강화한 음성 대화 방법에 있어서,

다수의 문장으로 구성되는 질의 응답 말뭉치를 입력받아 다수의 문장 각각에 대해 단어들로 구성되는 제1메타 데이터를 생성하는 단계;

상기 다수의 문장 각각에 대한 제1메타 데이터를 입력받아, 상기 제1메타 데이터의 단어들 각각에 대해 제1임베딩 벡터를 부가하여 제1단어벡터들을 생성하고, 상기 제1단어벡터들을 양방향 LSTM 처리하여 제1문장 임베딩 정보를 생성하고, 상기 제1메타 데이터의 단어들 각각에 대해 제2임베딩 벡터를 부가하여 제2단어벡터들을 생성하고, 상기 제2단어벡터들을 양방향 LSTM 처리하여 제2문장 임베딩 정보를 생성하는 단계;

질문 문장에 대해 단어들로 구성되는 제2메타 데이터를 생성하고, 상기 제2메타 데이터의 단어들 각각에 대해 제3임베딩 벡터를 부가하여 제3단어벡터들을 생성하고, 상기 제3단어벡터들을 양방향 LSTM 처리하여 질문문장 임베딩 정보를 생성하는 단계;

상기 다수의 문장들 각각에 대한 제1문장 임베딩 정보와 상기 질문문장 임베딩 정보를 입력받아, 상기 다수의 문장들 각각에 대한 제1문장 임베딩 정보와 상기 질문 임베딩 정보 사이의 관련성을 나타내는 제1정보를 생성하는 단계;

상기 다수의 문장들 각각에 대한 제1정보와 상기 다수의 문장들 각각에 대한 제2문장 임베딩 정보의 웨이트드를 곱하여 제2정보를 생성하는 단계;

상기 다수의 문장들 각각에 대해 제2정보와 상기 질문정보를 가산하여 제3정보를 출력하는 단계; 및

상기 제3정보를 통해 상기 다수의 문장과 질문정보 사이의 확률을 산출하고, 가장 높은 확률을 가진 문장을 답변 문장으로 예측하여 출력하는 단계;를 포함하는 것을 특징으로 하는 단어자질을 강화한 음성 대화 방법.

청구항 2

제1항에 있어서,

상기 제1 내지 제3메타 데이터의 단어들은 데이터 타입과 식별번호로 치환되며,

상기 데이터 타입과 식별번호에 대응되게 치환된 단어들은 예측된 답변 문장을 구성하는 치환된 단어들을 원래의 단어로 치환함을 특징으로 하는 단어자질을 강화한 음성 대화 방법.

청구항 3

제1항에 있어서,

다수의 문장들 각각에 대한 제1문장 임베딩 정보와 상기 질문 임베딩 정보 사이의 관련성을 나타내는 제1정보는 수학적식 5에 따라 산출됨을 특징으로 하는 단어자질을 강화한 음성 대화 방법.

수학적식 5

$$p_i = \text{Softmax}(u \cdot m_i)$$

$$\text{softmax} = \frac{e^{z_i}}{\sum_{k=1}^K e^{z_k}}$$

$$z_i = u \cdot m_i$$

상기 수학적식 5에서 p_i 는 제1문장 임베딩 정보와 상기 질문 임베딩 정보의 관련정도를 나타내는 제1정보이고, u 는 입력문장정보이고, m_i 는 질문 임베딩 정보이고, i 는 제1문장 임베딩 정보의 식별정보이고, k 는 소프트맥스

수식에 입력되는 구성요소(element)들의 총 개수이고, $\cdot$ 은 행렬곱(matrix multiplication)을 나타냄.

청구항 4

제3항에 있어서,

상기 다수의 문장들 각각에 대한 제1정보와 상기 다수의 문장들 각각에 대한 제2문장 임베딩 정보의 웨이트드를 수학적 식 6에 따라 산출함을 특징으로 하는 단어자질을 강화한 음성 대화 방법.

수학적 식 6

$$o = \sum_i p_i c_i$$

상기 수학적 식 6에서, 상기 p_i 는 입력문장정보와 질문문장정보의 관련정도를 나타내는 제1정보이고, 상기 c_i 는 상기 제2문장 임베딩 정보임.

청구항 5

제4항에 있어서,

상기 제3정보를 통해 상기 다수의 문장과 질문정보 사이의 확률을 수학적 식 7에 따라 산출하고 가장 높은 확률을 가진 문장을 답변 문장으로 예측하여 출력함을 특징으로 하는 단어자질을 강화한 음성 대화 방법.

수학적 식 7

$$a = \text{softmax}(W(o+u))$$

상기 수학적 식 7에서, a 의 차원은 전체 후보(global candidate) 개수와 같으며 상기 다수의 문장들인 후보 문장 각각에 대한 확률값들을 나타냄.

발명의 설명

기술 분야

[0001] 본 발명은 자연어 처리 기술에 관한 것으로, 더욱 상세하게는 질의응답 말뭉치를 구성하는 문장들과 질문 문장에 대해 문장을 구성하는 단어의 자질을 강화하여 자연어인 질문 문장에 대해 적절한 답변 문장을 생성할 수 있게 하는 단어자질을 강화한 음성 대화 시스템 및 방법에 관한 것이다.

배경 기술

[0002] 자연어 질의응답기술은 자연어인 질문 문장이 사용자로부터 입력되면, 상기 사용자의 질문 의도에 맞는 답변들을 검색하고, 상기 답변들 중 어느 하나를 선택하여 사용자에게 반환하는 기술이다.

[0003] 이러한 자연어 질의응답기술은 모바일 시대의 도래와 함께 모바일 기기의 하드웨어적인 한계, 즉 작은 화면으로 검색 결과를 제시하여야 하는 한계를 극복하기 위한 기술로 다시 각광을 받고 있다. 예를들어, 아이폰의 경우에는 '시리'라는 기술을 제안하여 음성으로 사용자가 질문을 하면 그 질문에 응답하는 기술을 선보인 바 있다.

[0004] 이러한 기술로는 Sainbayar Sukhbaatar, Arthur Szlam, Jason Weston, Rob Fergus. "End-To-End Memory Networks." Proceedings of the NIPS2015 Symposium. Advances in Neural Information Processing Systems, 2015.에 개시된 엔드 투 엔드 메모리 네트워크가 있다. 상기 엔드 투 엔드 메모리 네트워크는 질의응답 문제에 대해 탁월한 성능을 보이는 모델로 문맥과 질문이 주어질 때 문맥 안에서 그에 맞는 적절한 답을 찾아내 주는 프로그램이다. 이 프로그램에 따르면 답은 한 단어로 되어 있으며, 불필요한 정보가 들어 있는 문맥에서도 필요한 정보만을 뽑아내어 적절한 답을 유추할 수 있었다.

[0005] 그러나 상기한 엔드 투 엔드 메모리 네트워크는 비교적 간단한 질의응답만 가능하였으며 답이 단어로 된 것에서는 잘 동작하였으나, 정보의 수가 많은 문맥이 복잡한 질의문에 대해서는 성능이 많이 감소되는 문제가 있었다. 즉, 제시되는 정보의 개수가 3개일 때, 에러율이 21.9 퍼센트이다.

[0006] 또한 상기한 엔드 투 엔드 메모리 네트워크는 단어의 의미를 잘 반영하지 못하는 문제가 있었다. 예를 들어, four apples와 six apples가 있으면 four와 six를 구분하지 못하고 그냥 숫자로만 인식하였다. 이에 사과가 몇

개 있는가에 대한 질문에 가짜 정보가 추가되면 제대로 답을 하지 못하였다.

- [0007] 또한 상기 엔드 투 엔드 메모리 네트워크는 훈련 데이터에 포함되지 않은 고유명사에 대해서는 전혀 대비되지 않아, 훈련 데이터에 포함되지 않은 고유명사가 질문에 포함되는 경우에는 대응할 수 없는 문제가 있었다.

선행기술문헌

특허문헌

- [0008] (특허문헌 0001) 대한민국 특허공개 제1020160060247호
(특허문헌 0002) 대한민국 특허공개 제1020150041908호
(특허문헌 0003) 대한민국 특허공개 제1020160114668호
(특허문헌 0004) 대한민국 특허공개 제1020170096282호

발명의 내용

해결하려는 과제

- [0009] 본 발명은 질의응답 말뭉치를 구성하는 문장들과 질문 문장에 대해 문장을 구성하는 단어의 자질을 강화하여 자연어인 질문 문장에 대해 적절한 답변 문장을 생성할 수 있게 하는 단어자질을 강화한 음성 대화 시스템 및 방법을 제공하는 것을 목적으로 한다.
- [0010] 또한 본 발명의 다른 목적은 질의응답 말뭉치에 포함되지 않았던 고유명사가 포함된 질문문장에 대해서도 적절한 답변 문장을 검출할 수 있는 단어자질을 강화한 음성 대화 시스템 및 방법을 제공하는 것이다.

과제의 해결 수단

- [0011] 상기한 목적을 달성하기 위한 본 발명에 따르는 단어자질을 강화한 음성 대화 방법은, 다수의 문장으로 구성되는 질의 응답 말뭉치를 입력받아 다수의 문장 각각에 대해 단어들로 구성되는 제1메타 데이터를 생성하는 단계; 상기 다수의 문장 각각에 대한 제1메타 데이터를 입력받아, 상기 제1메타 데이터의 단어들 각각에 대해 제1임베딩 벡터를 부가하여 제1단어벡터들을 생성하고, 상기 제1단어벡터들을 양방향 LSTM 처리하여 제1문장 임베딩 정보를 생성하고, 상기 제1메타 데이터의 단어들 각각에 대해 제2임베딩 벡터를 부가하여 제2단어벡터들을 생성하고, 상기 제2단어벡터들을 양방향 LSTM 처리하여 제2문장 임베딩 정보를 생성하는 단계; 질문 문장에 대해 단어들로 구성되는 제2메타 데이터를 생성하고, 상기 제2메타 데이터의 단어들 각각에 대해 제3임베딩 벡터를 부가하여 제3단어벡터들을 생성하고, 상기 제3단어벡터들을 양방향 LSTM 처리하여 질문문장 임베딩 정보를 생성하는 단계; 상기 다수의 문장들 각각에 대한 제1문장 임베딩 정보와 상기 질문문장 임베딩 정보를 입력받아, 상기 다수의 문장들 각각에 대한 제1문장 임베딩 정보와 상기 질문 임베딩 정보 사이의 관련성을 나타내는 제1정보를 생성하는 단계; 상기 다수의 문장들 각각에 대한 제1정보와 상기 다수의 문장들 각각에 대한 제2문장 임베딩 정보의 웨이트드 썸을 이행하여 제2정보를 생성하는 단계; 상기 다수의 문장들 각각에 대해 제2정보와 상기 질문 정보를 가산하여 제3정보를 출력하는 단계; 및 상기 제3정보를 통해 상기 다수의 문장과 질문정보 사이의 확률을 산출하고, 가장 높은 확률을 가진 문장을 답변 문장으로 예측하여 출력하는 단계;를 포함하는 것을 특징으로 한다.

발명의 효과

- [0012] 상기한 본 발명은 질의응답 말뭉치를 구성하는 문장들과 질문 문장에 대해 문장을 구성하는 단어의 자질을 강화하여 자연어인 질문 문장에 대해 적절한 답변 문장을 생성할 수 있게 하는 효과를 야기한다.
- [0013] 또한 본 발명은 질의응답 말뭉치에 포함되지 않았던 고유명사가 포함된 질문문장에 대해서도 적절한 답변 문장을 검출할 수 있는 효과를 야기한다.

도면의 간단한 설명

- [0014] 도 1은 본 발명의 바람직한 실시예에 따르는 단어자질을 강화한 음성 대화 시스템의 구성도.

도 2는 도 1의 단어자질을 강화한 음성 대화 처리부의 구성도.

도 3은 도 2의 전처리부의 처리절차도.

도 4는 도 2의 답변문장 출력부의 처리절차도.

발명을 실시하기 위한 구체적인 내용

- [0015] 본 발명은 질의응답 말뭉치를 구성하는 문장들과 질문 문장에 대해 문장을 구성하는 단어의 자질을 강화하여 자연어인 질문 문장에 대해 적절한 답변 문장을 생성한다.
- [0016] 또한 본 발명은 질의응답 말뭉치에 포함되지 않았던 고유명사가 포함된 질문문장에 대해서도 적절한 답변 문장을 검출한다.
- [0017] 상기한 본 발명에 따르는 단어자질을 강화한 음성 대화 시스템의 구성 및 동작을 도면을 참조하여 상세히 설명한다.
- [0018] <단어자질을 강화한 음성 대화 시스템의 구성>
- [0019] 도 1은 본 발명의 바람직한 실시예에 따르는 단어자질을 강화한 음성 대화 시스템의 구성도이다. 상기 도 1을 참조하면, 상기 음성 대화 서비스 시스템(100)은 음성 대화 서비스 서버(102)와 데이터베이스(104)로 구성된다.
- [0020] 상기 음성 대화 서비스 서버(102)는 본 발명의 바람직한 실시예에 따라 단어자질을 강화한 음성 대화 서비스를 사용자 단말기(200)에 제공한다.
- [0021] 상기 데이터베이스(104)는 상기 음성 대화 서비스 서버(102)에서 요구되는 다양한 정보를 저장하며, 특히 상기 데이터베이스(104)는 다수의 문장들로 구성되는 질의응답 말뭉치와 질문문장에 대해 문장을 구성하는 단어들을 치환하기 위한 참조정보와, 상기 참조정보를 토대로 단어들을 치환한 치환된 단어정보와, 상기 치환된 단어정보에 대응되는 원래의 단어정보를 저장한다.
- [0022] 상기 사용자 단말기(200)는 네트워크를 통해 상기 음성 대화 서비스 시스템(100)에 접속하여 사용자에게 본 발명의 바람직한 실시예에 따르는 단어 자질을 강화한 음성 대화 서비스를 제공한다. 상기한 사용자 단말기(200)는 음성 대화 서비스를 위해 사용자 단말기(200)의 각부를 제어하는 제어장치(202)와, 상기 제어장치(202)의 처리 프로그램을 포함하는 다양한 정보를 저장하는 메모리부(204)와, 상기 제어장치(202)의 제어에 따르는 오디오 및 비디오를 출력하는 오디오 및 비디오 출력부(210)와, 상기 제어장치(202)와 음성 대화 서비스 서버(102) 사이의 통신을 담당하는 통신부(212)를 구비한다.
- [0023] 상기한 본 발명의 바람직한 실시예에서는 음성 대화 서비스 시스템(100)에 접속한 사용자 단말기(200)를 통해, 상기 음성 대화 서비스 시스템(100)이 사용자에게 음성 대화 서비스를 제공하는 것만을 개시하였으나, 음성 대화 서비스 실행을 위한 프로그램 및 데이터베이스를 사용자 단말기에 구축하여 독립적으로 음성 대화 서비스를 실행할 수도 있으며, 이는 본 발명에 의해 당업자에게 자명하다.
- [0024] 이제 본 발명의 바람직한 실시예에 따르는 단어자질을 강화한 음성 대화 서비스 방법을 설명한다.
- [0025] <단어자질을 강화한 음성 대화 서비스 프로그램의 구성>
- [0026] 도 2는 본 발명의 바람직한 실시예에 따르는 단어자질을 강화한 음성 대화 서비스 서버에 구축되는 프로그램의 구성도를 도시한 것이다. 상기 도 2를 참조하면, 상기 음성 대화 서비스 프로그램은 제1 및 제2전처리부(3001,3002)와 제1임베딩 처리부(400)와 제2임베딩 처리부(500)와 제3임베딩 처리부(600)와 메모리 네트워크 처리부(700)와 정답출력부(800)로 구성된다.
- [0027] 이러한 음성 대화 서비스 프로그램에서 사용되는 말뭉치 데이터는 DSTC 6 경진대회에서 제공한 레스토랑 예약 관련 질의 응답 말뭉치 데이터가 사용될 수 있으며, bAbI보다 약간 더 난이도 있는 문제이며 답변은 문장으로 구성되며, 본 발명은 10개의 후보문장 중에서 정답문장을 선택하여 출력한다.
- [0028] <제1 및 제2전처리부(3001,3002)>
- [0029] 상기한 본 발명에 따르는 제1 및 제2전처리부(3001,3002)의 처리절차는 동일하므로, 이하 제1전처리부(3001)의 처리 절차만을 도 3을 참조하여 설명한다.
- [0030] 상기 제1전처리부(3001)는 다수의 문장들로 구성되는 질의응답 말뭉치 데이터가 입력되면(304단계), 상기 질의

응답 말뭉치를 구성하는 다수의 문장 각각에 대해 단어들로 구성되는 메타 데이터로 변환하고(306단계), 데이터 베이스(104)의 참조정보를 토대로 상기 다수의 문장 각각에 대한 메타 데이터의 단어들을 데이터 타입정보와 식별번호로 치환하여 치환된 단어정보를 생성하고, 상기 치환된 단어정보와 원래의 단어정보를 대응시켜 상기 데이터베이스(104)에 기록한다(308단계).

[0031] 상기 제1전처리부(3001)는 상기 다수의 문장 각각에 대해 상기 치환된 단어정보들을 결합하여 단어열 정보를 생성하고, 상기 단어열 정보의 단어수가 미리 정해진 수, 즉 가장 긴 문장이라고 임의 결정한 문장에 속한 단어들의 수보다 작으면, 상기 미리 정해진 수보다 부족한 단어에 대해서는 널(null)을 패딩하여 출력 메타 데이터를 생성하여 출력한다(310단계). 여기서, 상기 널은 임베딩을 거치면 벡터 0으로 치환한다.

[0032] 이로서 상기 제1전처리부(3001)는 다수의 문장들 각각에 대해 해당 문장을 구성하는 단어들 각각을 데이터 타입정보와 식별번호로 치환하여 치환된 단어정보를 생성하고, 상기 치환된 단어정보를 결합하여 단어열 정보를 생성하고, 상기 단어열 정보의 단어수가 미리 정해진 수보다 부족하면 부족한 단어에 대해서는 널을 패딩하여 출력 메타 데이터를 생성하여 제1 내지 제2임베딩 처리부(400,500)로 입력한다.

[0033] 그리고 상기 제2전처리부(3002)는 질문 문장을 입력받아 메타 데이터로 변환하고, 상기 메타 데이터에 속한 단어들 각각을 타입정보와 식별정보로 치환하고, 상기 타입정보로 치환된 단어들을 결합하여 단어열 정보를 생성하며, 상기 단어열의 단어수가 미리 정해진 수보다 작으면 그 수보다 부족한 단어에 대해서는 널을 패딩하여 출력 메타 데이터를 생성하여 제3임베딩 처리부(600)로 입력한다.

[0034] <제1임베딩 처리부(400)>

[0035] 상기 제1임베딩 처리부(400)는 임베딩 행렬 A부(402)와 제1순방향 LSTM부들(4061~406n)과 제1역방향 LSTM부들(4081~408n)과 제1LSTM 셀 추가부(410)로 구성된다.

[0036] 상기 임베딩 행렬 A부(402)는 상기 제1전처리부(3001)로부터 질의응답 말뭉치 데이터를 구성하는 다수의 문장 각각에 대한 출력 메타 데이터의 단어들 각각에 대해 제1임베딩 벡터를 부가하여 제1단어벡터들 $w_1 \sim w_n$ (4041~404n)를 생성하여 출력한다. 상기 제1순방향 LSTM부들(4061~406n)은 순방향으로 연결된 셀들로 구성되어 상기 제1단어벡터들 $w_1 \sim w_n$ (4041~404n) 각각에 대해 단어위치에 따라 순방향으로 정렬된 제1정보를 생성한다. 상기 제1역방향 LSTM부들(4081~408n)은 역방향으로 연결된 셀들로 구성되어 상기 단어벡터들 $w_1 \sim w_n$ (4041~404n) 각각에 대해 단어위치에 따라 역방향으로 정렬된 제2정보를 생성한다. 여기서, 상기 제1순방향 LSTM부들(4061~406n)과 제1역방향 LSTM부들(4081~408n)에 구비되는 셀의 개수는 문장 단어열의 생성시에 미리 정해진 수, 즉 가장 긴 문장이라고 임의 결정한 문장에 속한 단어들의 수에 대응되게 결정된다.

[0037] 상기 제1LSTM 셀 추가부(410)는 상기 제1순방향 LSTM부들(4061~406n)의 마지막 부분에 히든 스테이트 벡터를 부가하고, 상기 제1역방향 LSTM부들(4081~408n)의 첫 부분에 히든 스테이트 벡터를 부가하여, 제1문장 임베딩 정보 m_1 를 생성하여 출력한다. 여기서, 상기 히든 스테이트 벡터의 벡터값은 0으로 초기화된다.

[0038] 이와 같이 제1임베딩 처리부(400)는 질의 응답 말뭉치를 구성하는 다수의 문장 각각에 대한 출력 메타 데이터를 입력받아, 상기 출력 메타 데이터의 단어들 각각에 대해 제1임베딩 벡터를 부가하여 제1단어벡터들 $w_1 \sim w_n$ (4041~404n)을 생성하고, 상기 제1단어벡터들 $w_1 \sim w_n$ (4041~404n)을 양방향 LSTM 처리하여 제1문장 임베딩 정보 m_1 를 생성하여 출력한다.

[0039] <제2임베딩 처리부(500)>

[0040] 상기 제2임베딩 처리부(500)는 임베딩 행렬 C부(502)와 제2순방향 LSTM부들(5061~506n)과 제2역방향 LSTM부들(5081~508n)과 제2LSTM 셀 추가부(510)로 구성된다.

[0041] 상기 임베딩 행렬 C부(502)는 상기 제1전처리부(3001)로부터 질의응답 말뭉치 데이터를 구성하는 다수의 문장 각각에 대한 출력 메타 데이터의 단어들 각각에 대해 임베딩 벡터를 부가하여 단어벡터들 $w_1 \sim w_n$ (5041~504n)을 생성하여 출력한다. 상기 제2순방향 LSTM부들(5061~506n)은 순방향으로 연결된 셀들로 구성되어 상기 단어벡터들 $w_1 \sim w_n$ (5041~504n) 각각에 대해 단어위치에 따라 순방향으로 정렬된 제3정보를 생성한다. 상기 제2역방향 LSTM부들(5081~508n)은 역방향으로 연결된 셀들로 구성되어 상기 단어벡터들 $w_1 \sim w_n$ (5041~504n) 각각에 대해 단어위치에 따라 역방향으로 정렬된 제4정보를 생성한다. 여기서, 상기 제2순방향 LSTM부들(5061~506n)과 제2역방향 LSTM부들(5081~508n)에 구비되는 셀의 개수는 문장 단어열의 생성시에 미리 정해진 수, 즉 가장 긴 문장이라고 임의 결정한 문장에 속한 단어들의 수에 대응되게 결정된다.

[0042] 상기 제2LSTM 셀 추가부(510)는 상기 제2순방향 LSTM부들(5061~506n)의 마지막 부분에 히든 스테이트 벡터를 부

가하고, 상기 제2역방향 LSTM부들(5081~508n)의 첫 부분에 히든 스테이트 벡터를 부가하여 제2문장 임베딩 정보 ci를 생성하여 출력한다. 여기서, 상기 히든 스테이트 벡터의 벡터값은 0으로 결정된다.

[0043] 이와 같이 제2임베딩 처리부(500)는 질의 응답 말뭉치를 구성하는 다수의 문장 각각에 대한 출력 메타 데이터를 입력받아, 상기 출력 메타 데이터의 단어들 각각에 대해 제2임베딩 벡터를 부가하여 제2단어벡터들 w1~wn(5041~504n)을 생성하고, 상기 제2단어벡터들을 양방향 LSTM 처리하여 제2문장 임베딩 정보 ci를 생성하여 출력한다.

[0044] <제3임베딩 처리부(600)>

[0045] 상기 제3임베딩 처리부(600)는 상기 임베딩 행렬 B부(602)와 제3순방향 LSTM부들(6061~606n)과 제3역방향 LSTM부들(6081~608n)과 제4LSTM 셀 추가부(610)로 구성된다.

[0046] 상기 임베딩 행렬 B부(602)는 상기 제2전처리부(3002)로부터 질문 문장에 대한 출력 메타 데이터의 단어들 각각에 대해 제3임베딩 벡터를 부가하여 제3단어벡터들 w1~wn(6041~604n)을 출력한다. 상기 제3순방향 LSTM부들(6061~606n)은 순방향으로 연결된 셀들로 구성되어 상기 제3단어벡터들 w1~wn(6041~604n)을 문장내 단어의 위치에 따라 순방향으로 정렬된 제5정보를 생성한다. 상기 제3역방향 LSTM부들(6081~608n)은 역방향으로 연결된 셀들로 구성되어 상기 제3단어벡터들 w1~wn(6041~604n)을 문장내 단어의 위치에 따라 역방향으로 정렬된 제6정보를 생성한다. 여기서, 상기 제3순방향 LSTM부들(6061~606n)과 제3역방향 LSTM부들(6081~608n)을 구성하는 셀의 개수는 문장 단어열의 생성시에 미리 정해진 수, 즉 가장 긴 문장이라고 임의 결정한 문장에 속한 단어들의 수에 대응되게 결정된다.

[0047] 상기 제3LSTM 셀 추가부(610)는 상기 제3순방향 LSTM부들(6061~606n)의 마지막 부분에 히든 스테이트 벡터를 부가하고, 상기 제3역방향 LSTM부들(6081~608n)의 첫 부분에 히든 스테이트 벡터를 부가하여 질문 임베딩 정보 u를 생성하여 출력한다. 여기서, 상기 히든 스테이트 벡터의 벡터값은 0으로 결정된다.

[0048] 이와 같이 제3임베딩 처리부(600)는 질문 문장에 대한 출력 메타 데이터를 입력받아, 상기 출력 메타 데이터의 단어들 각각에 대해 제3임베딩 벡터를 부가하여 제3단어벡터들 w1~wn(6041~604n)을 생성하고, 상기 제3단어벡터들 w1~wn(6041~604n)을 양방향 LSTM 처리하여 질문문장 임베딩 정보 u를 생성하여 출력한다.

[0049] 이러한 제1 내지 제3임베딩 처리부(400,500,600)는 단어의 의미가 훼손되는 문제를 해결하기 위해서 단어의 위치정보를 반영하도록 양방향 LSTM을 적용하며, 양방향 LSTM의 양쪽 종단에 히든 스테이트 벡터를 부가하여 더 풍부한 벡터 차원으로 문장을 나타내고 뒤에 나오는 단어와 앞에 나오는 단어의 의미를 균형있게 반영할 수 있게 한다.

[0050] 특히 상기한 양방향 LSTM이 적용되는 부분은 입력 메모리부와 출력 메모리부에 저장될 문장 임베딩들과 질문문장에 대한 임베딩에 적용되며 LSTM의 각 셀마다 들어가는 입력은 문장을 이루는 단어들이다. 여기서, 하나의 단어는 하나의 셀에 들어가며 단어는 해당 단어에 해당하는 임베딩 벡터로 치환되어 입력된다. 상기 임베딩 벡터는 초기에는 랜덤으로 결정되나 훈련과정을 통해 최적의 값으로 튜닝된다. 이와같이 문장 임베딩을 구성하면 문장내에 똑같이 숫자를 나타내는 단어가 있더라도 서로 다른 벡터값으로 임베딩되어 의미가 구분된다.

[0051] 또한 상기 양방향 LSTM으로 학습을 진행하게 되면 학습을 통해 축약된 정보를 가진 벡터가 LSTM의 출력으로 나오기 때문에 단어 자질이 강화된다. 상기 자질 중에 특히 문장내에서 단어의 위치정보가 강화되며, 이는 LSTM의 특성에 따라 시간적 및 위치적 정보가 반영되기 때문이다. 즉 i번째 문장 xi는 임베딩 행렬 A부(402)와 임베딩 행렬 C부(502)를 거치며, 두 행렬에는 각 단어들과 매칭되는 벡터들이 저장된다. 이러한 행렬을 거쳐서 문장에 있는 단어들은 각 단어들에 매칭되는 벡터 값으로 변환된다. 상기 입력 메모리(702)에 들어갈 mi를 생성하는 양방향 LSTM(406n,408n,410)에는 임베딩 행렬 A부(402)를 거친 벡터 값이 입력으로 제공되고, 출력 메모리(706)에 제공되는 ci를 생성하는 양방향 LSTM(5061~506n,5081~508n,510)에는 임베딩 행렬 C부(502)를 거친 벡터 값이 입력으로 제공된다. 그리고 질문 문장에 대해서는 임베딩 행렬 B부(302)를 거쳐 위와 같은 방법으로 u를 생성한다.

[0052] 상기한 임베딩 행렬 A부(402)와 임베딩 행렬 C부(502)와 임베딩 B부(602)는 각기 별도의 학습을 진행한다.

[0053] 즉, 상기 임베딩 행렬 A부(402)와 임베딩 행렬 C부(502)는 초기에 랜덤하게 정의된 값을 가지나 학습을 진행함에 따라 각기 다른 값을 가지게 된다. 이에따라 상기 제1 및 제2임베딩 처리부(400,500)의 LSTM의 내부의 히든 스테이트(hidden state) 값도 다르게 학습되어, mi와 ci의 값이 상이하게 된다.

[0054] 이와 같이 제1 및 제2임베딩 처리부(400,500)가 서로 상이하게 학습하는 것은 입력 및 출력 메모리의 용도가 다

르기 때문이다. 즉, 입력 부분은 히스토리 문장들과 질문 문장의 관련성을 구하기 위한 것으로, 입력 메모리에서 질문 문장과 관련이 깊은 문장일수록 웨이트 소프트맥스 벡터에서 큰 값을 출력하도록 하는 것이다. 그리고 출력 부분은 질문 문장과 관련성을 띄는 부분이 부각된 출력정보 o를 생성하도록 하는 것이다.

[0055] 그리고 마지막으로 임베딩 행렬 B부(302)는 질문 문장에 대해 임베딩을 하면서 학습을 이행하며, 입력 및 출력 메모리(702,704)에 쓰이는 임베딩과는 용도가 다르게 질문 문장을 벡터화하는 데에 특화된 학습을 이행한다.

[0056] <메모리 네트워크 처리부(700)>

[0057] 상기 메모리 네트워크 처리부(700)는 입력 메모리부(702)와 출력 메모리부(706)와 출력정보 저장부(708)와 결합부(710)로 구성된다.

[0058] 상기 입력 메모리부(702)는 제1임베딩 처리부(400)가 출력하는 제1문장 임베딩 정보 m_i 및 제3임베딩 처리부(600)가 출력하는 질문문장 임베딩 정보 u 를 입력받아, 히스토리 문장들 각각과 질문 문장 사이의 관련성을 나타내는 제1정보 p_i 를 생성하여 출력하고,

[0059] 상기 출력 메모리부(706)는 상기 제2임베딩 처리부(500)가 출력하는 제2문장 임베딩 정보 c_i 및 상기 히스토리 문장들 각각과 질문 문장 사이의 관련성을 나타내는 제1정보 p_i 를 제공받아 질문 문장에 대한 답변 후보 문장들을 추출하고, 상기 답변 후보 문장들 각각과 질문문장 임베딩 정보 u 를 가산하여 출력한다.

[0060] 좀더 설명하면, 상기 입력 메모리부(702)는 제1임베딩 처리부(400)가 출력하는 m_i 및 제3임베딩 처리부(600)가 출력하는 질문문장정보 u 를 입력받아 수학식 1에 따라 처리한다.

수학식 1

$$p_i = \text{Softmax}(u \cdot m_i)$$

$$\text{softmax} = \frac{e^{z_i}}{\sum_{k=1}^K e^{z_k}}$$

$$z_i = u \cdot m_i$$

[0061]

[0062] 상기 수학식 1에서 p_i 는 상기 제1문장 임베딩 정보 m_i 와 질문문장 임베딩 정보 u 의 관련정도를 나타내는 정보이고, i 는 입력문장의 식별정보이고, k 는 소프트맥스 수식에 입력되는 구성요소(element)들의 총 개수이고, $\cdot$ 은 행렬곱(matrix multiplication)을 나타낸다.

[0063] 여기서, 일반적으로 자연어 처리에서 임베딩 값이 비슷하면 관련성이 있다고 판단하며 그 비슷함을 구할때 코사인 시밀러리티(cosine similarity)를 사용한다. 상기 코사인 시밀러리티 수식을 간단화하면 이너 프로덕트와 같아지며, 히스토리의 문장 벡터인 m_i 와 질문문장 벡터인 u 가 이너 프로덕트 된 것들이 관련성을 나타낸다.

[0064] 상기 출력 메모리부(408)는 수학식 2에 따라 상기 제2임베딩 처리부(200)의 출력 c_i 와 상기 입력 메모리부(702)의 처리결과 p_i 를 제공받아 웨이트드 썸을 이행하고 그 결과정보인 o 를 출력한다.

수학식 2

$$o = \sum_i p_i c_i$$

[0065]

[0066] 상기 수학식 2에서 p_i 는 입력 메모리부(702)의 출력이고, 상기 c_i 는 상기 제2임베딩 처리부(500)가 출력하는 제2문장 임베딩 정보로서, 답변후보 문장이 아닌 히스토리 문장에 대해 출력(output)용으로 정보전달을 위해 따로 임베딩한 것이다. 그리고 o 는 답변 후보문장 중에서 정답을 고르기 위하여 히스토리 문장에서 적절한 문장 정보

를 골라 출력하는 역할을 이행한다.

[0067] 상기 웨이트드 썸 결과정보 o 는 결합부(714)에 제공된다.

[0068] 상기 결합부(714)는 상기 웨이트드 썸 결과정보 o 와 질문정보 u 를 수학식 3에 따라 가산하여 출력한다.

수학식 3

[0069] $\text{결합결과} = o + u$

[0070] 상기 수학식 3에서 상기 웨이트드 썸 결과정보 o 와 질문정보 u 는 같은 차원의 벡터이기 때문에 가산하는 것이 가능하다.

[0071] 상기 결합부(714)는 상기 웨이트드 썸 결과정보 o 와 질문정보 u 를 결합하여 정답출력부(800)로 제공한다.

[0072] <정답출력부(800)>

[0073] 상기 정답출력부(800)는 w 부(802)와 예측된 답변 a 저장부(506)로 구성된다.

[0074] 상기 w 부(802)의 처리 절차는 도 4에 도시한 바와 같다. 상기 도 4를 참조하면, 상기 w 부(802)는 수학식 4에 따라 전체 후보 문장에 대해 가장 높은 확률을 가지는 문장부터 탐색하여 후보 문장과 일치하는 문장이 탐색되면, 그것을 답변 문장으로 예측한다(900단계).

수학식 4

[0075] $a = \text{softmax}(W(o+u))$

[0076] 상기 수학식 4에서 a 의 차원은 전체후보(global candidate) 개수와 같고 소프트맥스로 정규화했기 때문에 a 는 총합 1인 각 후보에 대한 확률값들을 나타낸다. 여기서 가장 높은 확률을 가지는 후보 문장부터 해당 대화에 있는 10개의 후보 문장 중에 일치하는 것이 있는지를 탐색하고, 후보 문장에 해당 대화에 있는 10개의 후보 문장이 있는 문장을 최종 정답인 답변 문장으로 선택한다.

[0077] 상기 w 부(802)는 상기 답변 문장이 예측되면, 데이터베이스(104)에 저장된 단어와 치환정보를 이용하여 예측된 답변의 메타 데이터들을 원래의 단어로 치환하여 완성된 답변 문장을 출력한다(902,904단계).

[0078] 이러한 본 발명은 훈련 데이터에 등장할 가능성이 적은 고유명사 등에 대비하여 위해 데이터베이스를 참조하여 메타 데이터 타입으로 단어를 변환하여 학습한다. 이러한 방식은 데이터베이스에 등록된 단어들에 대해 대비가 되어 성능을 향상시킨다. 또한 본 발명은 변환된 단어를 복구하기 위해 원형을 별도 저장한다.

[0079] 만약 고유명사를 그대로 둔 채로 학습을 진행하면 단어 종류의 폭이 매우 크게 늘어나게 되며, 테스트 단계에서 사용되지 않은 단어가 나올 경우에는 모르는 단어로 인식되어 제대로 된 답변을 찾기가 어렵다. 이러한 상황을 방지하기 위해 본 발명은 주어진 데이터베이스를 참고하여 등록되어 있는 고유명사 등에 해당하는 단어들을 특정 형식으로 치환한다. 여기서, 치환하는 종류에는 레스토랑 이름, 전화번호, 가격, 인원수, 음식 종류, 음식점의 장소 등이 있다.

[0080] 도 5는 상기 메타 데이터 타입의 변환을 예시한 것이다. 상기 도 5를 참조하여 치환되는 형식을 설명하면, ‘<R_cuisine.1>’과 같이 뒤에 ‘.’과 숫자가 부여되며, 이는 한 대화 안에서 음식 관련 단어가 여러 개 나올 수 있으므로 그것들을 구분하기 위한 식별정보이다. 즉, 첫 번째 음식은 ‘<R_cuisine.1>’으로 두 번째 음식 단어는 ‘<R_cuisine.2>’로 나타내게 된다.

[0081] 또한 본 발명은 메타 데이터 타입으로 변환된 단어를 원형으로 복구할 수 있도록 단어들을 사전 형식으로 저장하여 원문손상을 미연에 방지한다.

[0082] 이와 같이 본 발명은 단어를 정해진 형식으로 치환을 함으로써 단어의 종류 폭이 좁아져서 원활한 학습이 이루어지고, 테스트 단계에서 처음 보는 단어도 데이터베이스에 존재만 하면 음식인지, 장소인지 등을 알 수 있어 적절한 정답 문장을 찾아 치환된 단어들을 원래 단어들로 복원하여 제대로 된 답변 문장을 출력할 수 있게

한다.

[0083] 이러한 본 발명에 따르는 치환은 파이썬 코드를 통해 자동적으로 이루어지며 입력된 데이터베이스에 존재하는 고유명사들에 대해서 치환작업을 하게 된다. 또한 각 대화와 후보문장에는 고유한 식별정보(ID)가 붙어 있기 때문에 치환되었더라도 고유 식별정보(ID)를 통해 알맞은 치환 정보를 찾을 수가 있다. 그리고 최종적으로 선택된 고유 식별정보(ID)와 답변 문장에 대해 고유 식별정보(ID)를 통해 찾은 치환 정보를 이용하여 치환 이전의 원래 문장으로 복구한다.

부호의 설명

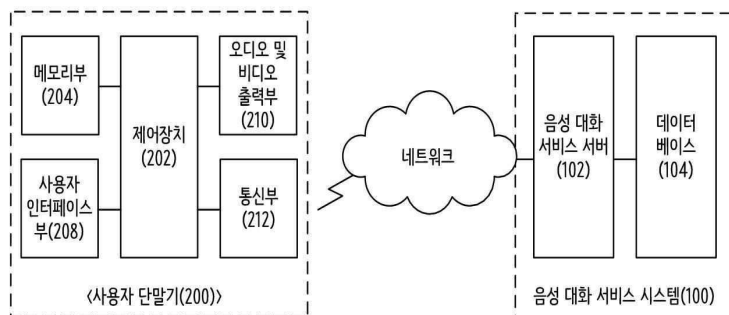
[0084] 100 : 음성 대화 서비스 시스템

102 : 음성 대화 서비스 서버

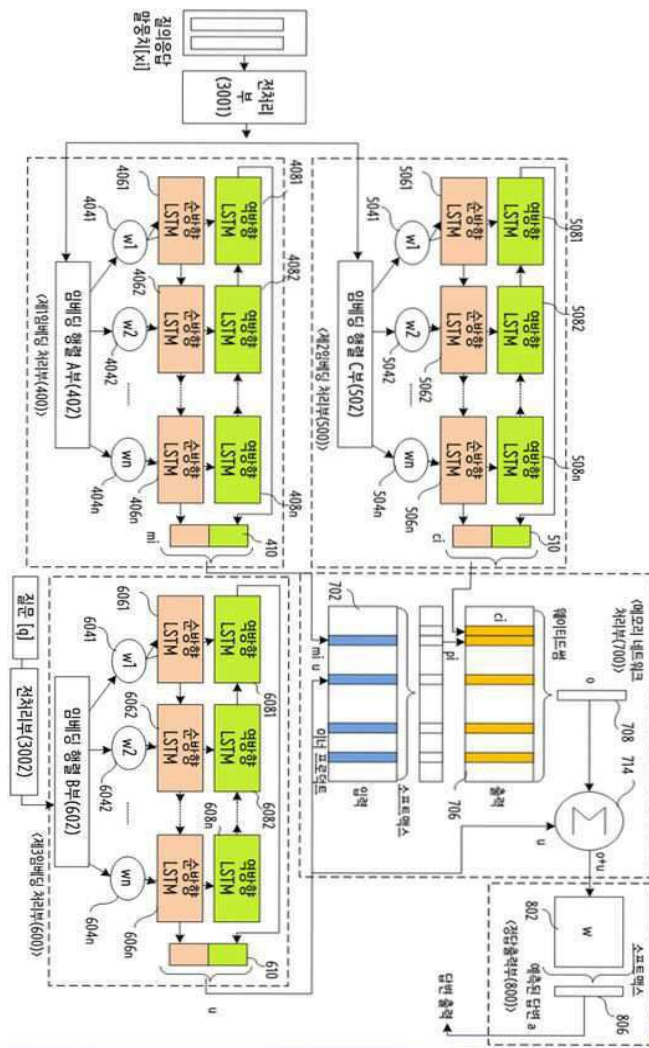
104 : 데이터베이스

도면

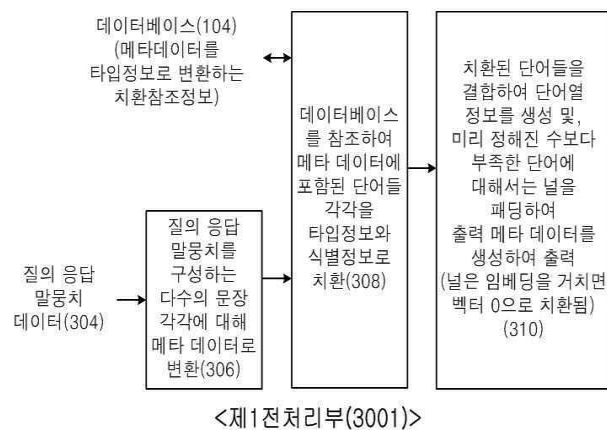
도면1



도면2



도면3



도면4

