



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2021년11월11일
(11) 등록번호 10-2324829
(24) 등록일자 2021년11월04일

- (51) 국제특허분류(Int. Cl.)
G10L 21/02 (2006.01) G06N 20/10 (2019.01)
G06N 3/08 (2006.01) G10L 15/01 (2013.01)
G10L 15/02 (2006.01) G10L 15/06 (2006.01)
G10L 15/10 (2006.01)
- (52) CPC특허분류
G10L 21/02 (2021.08)
G06N 20/10 (2021.08)
- (21) 출원번호 10-2019-0158218
(22) 출원일자 2019년12월02일
심사청구일자 2019년12월02일
- (65) 공개번호 10-2021-0068776
(43) 공개일자 2021년06월10일
- (56) 선행기술조사문헌
JP2016110082 A*
KR1020130128329 A*
*는 심사관에 의하여 인용된 문헌
- (73) 특허권자
주식회사 엘지유플러스
서울특별시 용산구 한강대로 32(한강로3가)
- (72) 발명자
이종언
서울특별시 용산구 한강대로 32 LG유플러스 빌딩
이보람
서울특별시 용산구 한강대로 32 LG유플러스 빌딩
- (74) 대리인
김함곤, 안광석, 박영일

전체 청구항 수 : 총 11 항

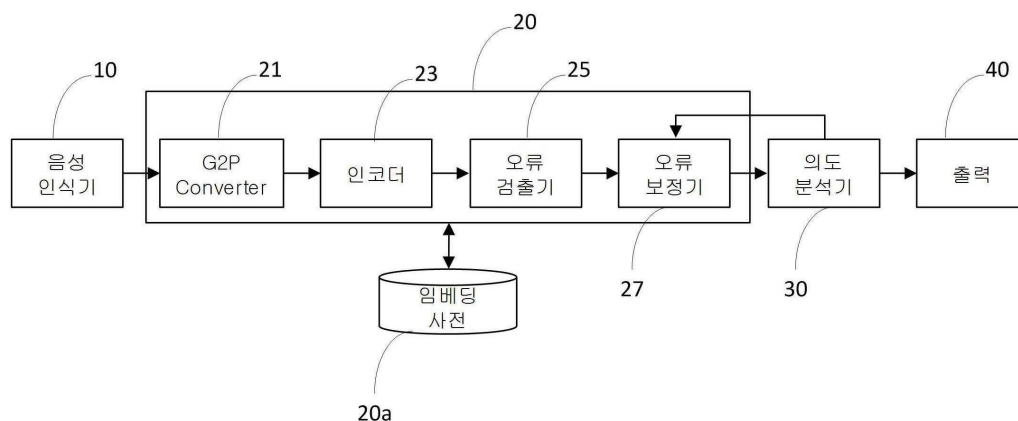
심사관 : 정성윤

(54) 발명의 명칭 음성 인식 오류 보정 장치 및 방법

(57) 요약

본 발명의 예시적인 실시예는 음성 인식 오류를 자연어처리 이전에 검출하고 보정하여 다양한 서비스도메인에서 최적의 효율성으로 자연어처리의 의도분석 성공률을 높이기 위한 음성 인식 오류 보정 장치에 관한 것으로, 정답 발화 토큰 및 해당 정답발화 토큰에 대응하는 하나 이상의 오류발화 토큰을 모두 유사한 벡터값을 가지는 정답 토큰별 밀집 벡터 공간안에 위치하도록 임베딩하기 위한 인코더; 상기 임베딩된 정답 토큰별 밀집 벡터 공간안의 정답발화 토큰을 기준으로 기 설정된 제1 맨하탄(Manhattan) 거리 내에 있는 하나 이상의 발화 토큰을 오류발화 토큰으로 검출하기 위한 오류검출기; 및 상기 임베딩된 정답 토큰별 밀집 벡터 공간 내에서 상기 검출된 오류발화 토큰으로부터 맨하탄 거리 기준으로 가장 가까이 있는 정답발화 토큰을 정답으로 추출하여 오류를 보정하기 위한 오류보정기를 포함할 수 있다.

대표도



(52) CPC특허분류

G06N 3/08 (2013.01)
G10L 15/01 (2013.01)
G10L 15/02 (2013.01)
G10L 15/063 (2013.01)
G10L 15/10 (2013.01)

명세서

청구범위

청구항 1

정답발화 토큰 및 해당 정답발화 토큰에 대응하는 하나 이상의 오류발화 토큰을 모두 유사한 벡터값을 가지는 정답 토큰별 밀집 벡터 공간안에 위치하도록 임베딩하기 위한 인코더;

상기 임베딩된 정답 토큰별 밀집 벡터 공간안의 정답발화 토큰을 기준으로 기 설정된 제1 맨하탄(Manhattan) 거리 내에 있는 하나 이상의 발화 토큰을 오류발화 토큰으로 검출하기 위한 오류검출기; 및

상기 임베딩된 정답 토큰별 밀집 벡터 공간 내에서 상기 검출된 오류발화 토큰으로부터 맨하탄 거리 기준으로 가장 가까이 있는 정답발화 토큰을 정답으로 추출하여 오류를 보정하기 위한 오류보정기;를 포함하고,

상기 오류보정기는 상기 임베딩된 정답 토큰별 밀집 벡터 공간 내에서 상기 검출된 오류발화 토큰으로부터 기 설정된 제2 맨하탄 거리 내에 있는 하나 이상의 정답발화 토큰을 정답 후보 보정군으로 추출하고, 상기 정답 후보 보정군 중 상기 검출된 오류발화 토큰으로부터 맨하탄 거리 기준으로 가장 가까이 있는 정답발화 토큰을 정답으로 추출하는 것을 특징으로 하는 음성 인식 오류 보정 장치.

청구항 2

삭제

청구항 3

제1항에 있어서,

정답으로 추출된 발화에 대한 의도분석 실패 시, 상기 오류보정기는 상기 추출된 정답 후보 보정군 중 이전 추출된 정답발화 토큰을 제외하고 상기 검출된 오류발화 토큰으로부터 맨하탄 거리 기준으로 가장 가까이 있는 정답발화 토큰을 정답으로 추출하는 것을 특징으로 하는 음성 인식 오류 보정 장치.

청구항 4

제1항에 있어서,

상기 인코더는 SNN(Spiking Neural Network) 기반 학습 모델로 구축된 것을 특징으로 하는 음성 인식 오류 보정 장치.

청구항 5

제4항에 있어서,

상기 SNN 기반 학습 모델은 LSTM(Long Short Term Memory), CNN(Convolutional Neural Network), 및 SVM(Supportive Vector Machine) 중 하나의 모델을 포함하는 것을 특징으로 하는 음성 인식 오류 보정 장치.

청구항 6

(a) 정답발화 토큰 및 해당 정답발화 토큰에 대응하는 하나 이상의 오류발화 토큰을 모두 유사한 벡터값을 가지는 정답 토큰별 밀집 벡터 공간안에 위치하도록 임베딩하기 위한 인코더를 구축하는 단계;

(b) 상기 임베딩된 정답 토큰별 밀집 벡터 공간안의 정답발화 토큰을 기준으로 기 설정된 제1 맨하탄(Manhattan) 거리 내에 있는 하나 이상의 발화 토큰을 오류발화 토큰으로 검출하기 위한 단계; 및

(c) 상기 임베딩된 정답 토큰별 밀집 벡터 공간 내에서 상기 검출된 오류발화 토큰으로부터 맨하탄 거리 기준으로 가장 가까이 있는 정답발화 토큰을 정답으로 추출하여 오류를 보정하기 위한 단계;를 포함하고,

상기 단계 (c)는 상기 임베딩된 정답 토큰별 밀집 벡터 공간 내에서 상기 검출된 오류발화 토큰으로부터 기 설정된 제2 맨하탄 거리 내에 있는 하나 이상의 정답발화 토큰을 정답 후보 보정군으로 추출하고, 상기 정답 후보 보정군 중 상기 검출된 오류발화 토큰으로부터 맨하탄 거리 기준으로 가장 가까이 있는 정답발화 토큰을 정답으로

로 추출하는 것을 특징으로 하는 음성 인식 오류 보정 방법.

청구항 7

삭제

청구항 8

제6항에 있어서,

상기 단계 (c)는, 정답으로 추출된 발화에 대한 의도분석 실패 시, 상기 추출된 정답 후보 보정군 중 이전 추출된 정답발화 토큰을 제외하고 상기 검출된 오류발화 토큰으로부터 맨하탄 거리 기준으로 가장 가까이 있는 정답 발화 토큰을 정답으로 추출하는 것을 특징으로 하는 음성 인식 오류 보정 방법.

청구항 9

제6항에 있어서,

상기 인코더는 (2개 이상의 동일한 서브 네트워크를 갖는) SNN(Spiking Neural Network) 기반 학습 모델로 구축된 것을 특징으로 하는 음성 인식 오류 보정 방법.

청구항 10

제9항에 있어서,

상기 SNN 기반 학습 모델은 LSTM(Long Short Term Memory), CNN(Convolutional Neural Network), 및 SVM(Supportive Vector Machine) 중 하나의 모델을 포함하는 것을 특징으로 하는 음성 인식 오류 보정 방법.

청구항 11

제6항 및 제8항 내지 제10항 중 어느 한 항의 상기 음성 인식 오류 보정 방법을 컴퓨터에서 실행시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록 매체.

청구항 12

제6항 및 제8항 내지 제10항 중 어느 한 항의 상기 음성 인식 오류 보정 방법을 하드웨어와 결합하여 실행시키기 위하여 컴퓨터로 읽을 수 있는 기록 매체에 저장된 애플리케이션.

청구항 13

제6항 및 제8항 내지 제10항 중 어느 한 항의 상기 음성 인식 오류 보정 방법을 컴퓨터에서 실행시키기 위하여 컴퓨터로 읽을 수 있는 기록 매체에 저장된 컴퓨터 프로그램.

발명의 설명

기술 분야

[0001] 본 발명은 자연어처리 의도분석 오류 중 음성인식 오류로 인해 발생한 오류를 자연어처리 전에 검출하여 선 보정하는 장치 및 방법에 관한 것이다.

배경 기술

[0002] 일반적으로, 음성기반 서비스는 발화자의 음성을 입력으로 받아 자연어텍스트를 출력하는 음성인식 과정 및 출력된 자연어텍스트를 분석하여 발화자의 의도, 감성, 메타데이터 등을 뽑아내는 자연어처리 과정을 포함하여 이루어진다. 이 후 서비스에 따라 답변을 생성하여 텍스트로 제공하거나 음성합성을 거쳐 음성 답변을 제공한다.

[0003] 음성인식 시스템은 음향 모델 및 언어 모델로 이루어져 있으며, 음향모델은 음성 신호의 특징을 분석하고, 언어 모델은 인식된 음소 및 단어의 앞뒤 요소들간의 발생 확률을 통해 해당 단어 및 음소를 예측한다.

[0004] 음향모델과 언어모델은 음성인식 정확도를 높이기 위해 음성인식 시스템 내에서 지속적인 학습 및 유사발음사전 구축을 통한 성능 향상이 필요하다. 하지만 향상된 음성인식 성능에도 불구하고 발음변이 및 예외발음법칙 등에

의해 음성인식 오류는 지속적으로 발생한다.

[0005] 음성인식 성능 보완을 위해 발음 사전을 구축하기도 하지만 지속적으로 생성되는 서비스용어, 신조어 및 구어체들을 계속 사전에 업데이트 하는 것은 한계가 있다.

[0006] 즉, 음성인식은 자체 언어모델 및 음향모델을 통해 선별된 텍스트를 배출하지만, 그렇게 선별되어 배출한 텍스트 또한 오류가 발생한다. 이러한 음향 및 발음기반 음성인식(엔진)의 오류는 신조어 및 복합명사가 빈번한 상품명 및 서비스명 도메인에서 빈번하게 발생한다(ex. 알뜰광광무제한요금제). 이때마다 음성인식엔진의 음향모델 및 언어모델을 재학습 시키거나, 유사발음사전을 수작업으로 업데이트하여 보완하는 방법으로 서비스를 진행하고 있으나, 이는 인력 및 비용 측면에서 비효율적인 문제점이 있다.

선행기술문헌

특허문헌

[0007] (특허문헌 0001) 공개특허공보 제10-2013-0128329호(2013.11.26.)

발명의 내용

해결하려는 과제

[0008] 본 발명은 전술한 종래의 문제점을 해결하기 위한 것으로, 그 목적은 음성 인식 오류를 자연어처리 이전에 검출하고 보정하여 다양한 서비스도메인에서 최적의 효율성으로 자연어처리의 의도분석 성공률을 높이기 위한, 음성 인식 오류 보정 장치 및 방법을 제공하는 것이다.

과제의 해결 수단

[0009] 전술한 목적을 달성하기 위하여 본 발명의 일 측면에 따른 음성 인식 오류 보정 장치는, 정답발화 토큰 및 해당 정답발화 토큰에 대응하는 하나 이상의 오류발화 토큰을 모두 유사한 벡터값을 가지는 정답 토큰별 밀집 벡터 공간안에 위치하도록 임베딩하기 위한 인코더; 상기 임베딩된 정답 토큰별 밀집 벡터 공간 안의 정답발화 토큰을 기준으로 기 설정된 제1 맨하탄(Manhattan) 거리 내에 있는 하나 이상의 발화 토큰을 오류발화 토큰으로 검출하기 위한 오류검출기; 및 상기 임베딩된 정답 토큰별 밀집 벡터 공간 내에서 상기 검출된 오류발화 토큰으로부터 맨하탄 거리 기준으로 가장 가까이 있는 정답발화 토큰을 정답으로 추출하여 오류를 보정하기 위한 오류보정기를 포함할 수 있다.

[0010] 상기 오류보정기는 상기 임베딩된 정답 토큰별 밀집 벡터 공간 내에서 상기 검출된 오류발화 토큰으로부터 기 설정된 제2 맨하탄 거리 내에 있는 하나 이상의 정답발화 토큰을 정답 후보 보정군으로 추출하고, 상기 정답 후보 보정군 중 상기 검출된 오류발화 토큰으로부터 맨하탄 거리 기준으로 가장 가까이 있는 정답발화 토큰을 정답으로 추출할 수 있다.

[0011] 정답으로 추출된 발화에 대한 의도분석 실패 시, 상기 오류보정기는 상기 추출된 정답 후보 보정군 중 이전 추출된 정답발화 토큰을 제외하고 상기 검출된 오류발화 토큰으로부터 맨하탄 거리 기준으로 가장 가까이 있는 정답발화 토큰을 정답으로 추출할 수 있다.

[0012] 전술한 목적을 달성하기 위하여 본 발명의 다른 측면에 따른 음성 인식 오류 보정 방법은, (a) 정답발화 토큰 및 해당 정답발화 토큰에 대응하는 하나 이상의 오류발화 토큰을 모두 유사한 벡터값을 가지는 정답 토큰별 밀집 벡터 공간안에 위치하도록 임베딩하기 위한 인코더를 구축하는 단계; (b) 상기 임베딩된 정답 토큰별 밀집 벡터 공간안의 정답발화 토큰을 기준으로 기 설정된 제1 맨하탄(Manhattan) 거리 내에 있는 하나 이상의 발화 토큰을 오류발화 토큰으로 검출하기 위한 단계; 및 (c) 상기 임베딩된 정답 토큰별 밀집 벡터 공간 내에서 상기 검출된 오류발화 토큰으로부터 맨하탄 거리 기준으로 가장 가까이 있는 정답발화 토큰을 정답으로 추출하여 오류를 보정하기 위한 단계를 포함할 수 있다.

[0013] 상기 단계 (c)는 상기 임베딩된 정답 토큰별 밀집 벡터 공간 내에서 상기 검출된 오류발화 토큰으로부터 기 설정된 제2 맨하탄 거리 내에 있는 하나 이상의 정답발화 토큰을 정답 후보 보정군으로 추출하고, 상기 정답 후보 보정군 중 상기 검출된 오류발화 토큰으로부터 맨하탄 거리 기준으로 가장 가까이 있는 정답발화 토큰을 정답으로

로 추출할 수 있다.

[0014] 본 발명의 또 다른 측면에 따르면, 정답으로 추출된 발화에 대한 의도분석 실패 시, 상기 단계 (c)는 상기 추출된 정답 후보 보정군 중 이전 추출된 정답발화 토큰을 제외하고 상기 검출된 오류발화 토큰으로부터 맨하탄 거리 기준으로 가장 가까이 있는 정답발화 토큰을 정답으로 추출할 수 있다.

[0015] 진술한 목적을 달성하기 위하여 본 발명의 또 다른 측면에 따르면, 상기 음성 인식 오류 보정 방법을 컴퓨터에서 실행시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록 매체가 제공될 수 있다.

[0016] 진술한 목적을 달성하기 위하여 본 발명의 또 다른 측면에 따르면, 상기 음성 인식 오류 보정 방법을 하드웨어와 결합하여 실행시키기 위하여 컴퓨터로 읽을 수 있는 기록 매체에 저장된 애플리케이션이 제공될 수 있다.

[0017] 진술한 목적을 달성하기 위하여 본 발명의 또 다른 측면에 따르면, 상기 음성 인식 오류 보정 방법을 컴퓨터에서 실행시키기 위하여 컴퓨터로 읽을 수 있는 기록 매체에 저장된 컴퓨터 프로그램이 제공될 수 있다.

발명의 효과

[0018] 이상에서 설명한 바와 같이 본 발명의 다양한 측면에 따르면, 음성 인식 오류를 자연어처리 이전에 검출하고 보정하여 다양한 서비스도메인에서 최적의 효율성으로 자연어처리의 의도분석 성공률을 향상시킬 수 있는 효과가 있다.

[0019] 즉, 본 발명을 통해 서비스 도메인 별로 다른 버전의 음성인식기 및 유사어사전을 지속적으로 업데이트 할 필요 없이 의도분석 전단계에서 오류검출 및 오류보정을 할 수 있도록 함으로써 의도분석 정확도 향상과 그에 따른 서비스 성공률을 높일 수 있다. 또한, 본 발명의 다양한 측면에 따르면 오류 검출 시 의도분석 성공 여부를 반영하는데, 5G의 도입에 따라 혁신적으로 높아진 데이터 전송속도에 의해 (추가 소요되는) 의도분석 데이터 전송 시간을 최소화 하면서 의도분석 결과를 활용하여 명확한 오류 검출이 가능하다.

[0020] 본 발명을 통해 음성인식의 음향모델 및 언어모델에서 자체 보정되지 못한 음성인식 오류를 의도분석 전단계에서 추가 보정하여 의도분석 정확도와 서비스 성공률을 높일 수 있고, 음향 모델에서 예외처리 되는 신조어 및 상품명 등을 보정할 수 있는 효과가 있다.

도면의 간단한 설명

[0021] 도 1은 본 발명의 예시적인 실시예에 따른 음성 인식 오류 보정 장치의 구성도,

도 2는 G2P 변환 과정을 설명하기 위한 예시도,

도 3은 한국어 음소셋을 나타낸 도면,

도 4는 P2V 변환 과정을 설명하기 위한 예시도,

도 5는 SNN 기반의 bi-directional LSTM 학습 모델을 설명하기 위한 예시도,

도 6은 임베딩된 정답 토큰별 밀집 벡터 공간의 예시도,

도 7은 본 발명의 예시적인 실시예에 따른 음성 인식 오류 보정 방법의 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0022] 이하, 첨부도면을 참조하여 본 발명의 실시예에 대해 구체적으로 설명한다. 각 도면의 구성요소들에 참조부호를 부가함에 있어서 동일한 구성요소들에 대해서는 비록 다른 도면상에 표시되더라도 가능한 한 동일한 부호를 가지도록 한다. 또한, 본 발명의 실시예에 대한 설명 시 관련된 공지 구성 또는 기능에 대한 구체적인 설명이 본 발명의 요지를 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명은 생략한다.

[0023] 도 1은 본 발명의 예시적인 실시예에 따른 음성 인식 오류 보정 장치(20)의 구성도로, 동 도면에 도시된 바와 같이, 음성 인식 오류 보정 장치(이하, '보정장치'라 약칭함)(20)는 G2P 컨버터(21), 인코더(23), 오류검출기(25), 및 오류보정기(27)를 포함할 수 있다.

[0024] G2P(Grapheme to Phoneme) 컨버터(21)는 음성인식기(10)로부터 입력된 음성인식(STT: Speech to Text) 전사 데이터의 문자소(Grapheme: 의미를 나타내는 글자 최소 단위)를 음소(Phoneme)로 변환한다.

[0025] G2P 컨버터(21)는 도 2-3에 예시된 바와 같이 자소에 따른 한국어 발음법이 지정되어 있기 때문에 자소와 음소

에 따라 지정된 발음 규칙에 의해 변환을 수행한다. G2P 컨버터(21)에서 변환된 음소(phoneme)는 P2V(Phoneme to Vector) 컨버터(미도시)를 통해 도 4에 예시된 바와 같이 음소단위별 벡터로 임베딩(특정 언어단위별로 복수 차원으로 벡터화하여 맵핑시키는 것)할 수 있다.

- [0026] 인코더(23)는 정답발화 토큰 및 해당 정답발화 토큰에 대응하는 하나 이상의 오류발화 토큰을 모두 유사한 벡터 값을 가지는 정답 토큰별 밀집 벡터 공간안에 위치하도록 임베딩하기 위한 것으로, 예를 들어, G2P 컨버터(21)에서 변환된 음소(phoneme)를 P2V(Phoneme to Vector) 컨버터(미도시)를 통해 음소단위별 벡터로 임베딩하고, 음소단위별 임베딩된 벡터를 토큰별 벡터 공간으로 매핑함으로써, 서로 음소 유사성을 갖는 정답발화 토큰과 오류발화 토큰 모두를 정답 토큰별 밀집 벡터 공간안에 위치하도록 임베딩할 수 있다.
- [0027] 본 실시예에서 인코더(23)는 SNN(Spiking Neural Network) 기반의 학습을 통해 구축될 수 있으며, 일 예로 도 5에 도시된 바와 같이 SNN(Spiking Neural Network) 기반의 bi-directional LSTM(Long Short Term Memory)을 이용하여 학습 및 모델링 할 수 있다.
- [0028] 즉, 본 실시예에 따른 SNN 기반 인코더(23)는 오류발화(또는 오류발화 토큰이라 칭함)의 입력 시 정답발화(또는 정답발화 토큰이라 칭함)과 비슷한 벡터공간 위치(이를, 정답 토큰별 밀집 벡터 공간안의 위치라 칭한다)에 매핑(임베딩)되도록 임베딩 패턴을 구축할 수 있다.
- [0029] SNN은 인공신경망 구조 클래스로 2개 이상의 서로 동일한 서브네트워크를 갖는다. 복수개의 서브네트워크들은 동일한 파라미터와 가중치를 가지며 서로 동기화되어 업데이트한다.
- [0030] SNN 기반 인코더(23)의 구축은 하나 이상의 오류발화(예를 들어, 도 5에서 '트리아스')와 이에 대응하는 정상발화(예를 들어, 도 5에서 '트와이스')를 각 서브네트워크에 입력하여 오류발화와 정상발화를 모두 비슷한 벡터 공간(즉, 정답 토큰별 밀집 벡터 공간)안에 위치하도록 각 서브네트워크 내의 파라미터들을 학습시킨다. 학습이 완료된 서브네트워크들은 서로 동일한 파라미터를 가지게 되며 최종적으로는 복수개의 동일한 인코더 모델이 구축된다(도 5 참조).
- [0031] SNN을 기반으로 인코더(23)의 구축 시 핵심 인코더 모델은 도 5의 예시와 같은 LSTM(Long Short Term Memory) 외에도 CNN(Convolutional Neural Network) 또는 SVM(Supportive Vector Machine) 등 다양한 확률기반 모델이 가능하다.
- [0032] 본 실시예에서는 SNN 학습 기반의 bi-directional LSTM를 활용한 인코더(23)를 구축한다.
- [0033] 인코더(23)는 학습되지 않은 새로운 토큰이 입력되었을 때 해당 토큰을 정답 토큰별 밀집 벡터 공간으로 임베딩하며, 이후 오류 검출 및 보정은 해당 임베딩 벡터 공간(즉, 임베딩된 토큰별 밀집 벡터 공간)(도 6에 예시된 임베딩된 토큰별 밀집 벡터 공간)을 참조하여 이루어진다.
- [0034] 본 실시예에서 임베딩된 토큰별 밀집 벡터 공간은 도 6의 예시와 같이 임베딩사전(20a)에 데이터베이스로 구축될 수 있다.
- [0035] 오류검출기(25)는 임베딩된 정답 토큰별 밀집 벡터 공간안의 정답발화 토큰을 기준으로 기 설정된 제1 맨하탄(Manhattan) 거리 내에 있는 하나 이상의 발화 토큰을 오류발화 토큰으로 검출하기 위한 것으로, 예를 들어, 도 6에 예시된 토큰별 밀집 벡터 공간에서 정답으로 태깅된 정답발화 토큰을 '방탄소년단'이라 전제하면 해당 정답 발화 토큰 '방탄소년단'을 기준으로 기 설정된 제1 맨하탄(Manhattan) 거리 내의 정답 태깅 없는 하나 이상의 발화 토큰인 '달탐사소년단'등을 오류발화 토큰으로 검출할 수 있다.
- [0036] 오류검출기(25) 또한 인코딩된 토큰 벡터를 활용하여 학습 및 분류가 진행된다. 임베딩 공간의 정답발화를 기준으로 특정 맨하탄(Manhattan) 거리 내에 있는 발화들을 오류발화로 지정하고, 오류발화와 정답발화를 이용하여 오류 검출기(25)를 모델링한다. 본 실시예에서는 CNN(Convolutional Neural Network) 기반의 오류검출기(25) 모델을 구축할 수 있다.
- [0037] 오류보정기(27)는 임베딩된 정답 토큰별 밀집 벡터 공간 내에서 오류검출기(25)를 통해 검출된 오류발화 토큰으로부터 맨하탄 거리 기준으로 가장 가까이 있는 정답발화 토큰을 정답으로 추출하여 오류를 보정하기 위한 것으로, 예를 들어, 도 6에 예시된 토큰별 밀집 벡터 공간에서 발화 토큰 '달탐사소년단'이 오류발화 토큰으로 검출된 경우 해당 오류발화 토큰 '달탐사소년단'으로부터 맨하탄 거리 기준으로 가장 가까이 있는 (태깅된) 정답발화 토큰 '방탄소년단'을 정답으로 추출하여 오류를 보정 즉, '달탐사소년단'을 '방탄소년단'으로 보정할 수 있다.

- [0038] CNN을 기반으로 학습된 오류검출기(25)를 통해 검출된 오류는 오류보정기(27)로 넘어간다. 오류보정기(27)는 학습 기반 보정 모델을 따로 구축하지 않는다. 오류보정기(27)는 임베딩 된 정답 토큰별 밀집 벡터 공간내에서 해당 오류와 가까이(맨하탄거리 기준 일정 거리 이내에) 있는 후보 보정군을 추출하며 그 중 가장 가까이 있는 정답 발화를 반환한다(도 6 참조).
- [0039] 오류보정기(27)를 통해 추출된 정답발화는 의도분석기(30)에서 의도분석 후 출력(40)하는데, 의도분석기(30)에서 의도분석이 되지 않는 즉, 실패한 의도분석에 해당되는 발화는 오류 보정기(27)를 통해 오류 재보정을 진행한다.
- [0040] 오류보정기(27)는 임베딩된 토큰별 밀집 벡터 공간내에서 첫 번째 반환된 정답발화를 제외하고 해당 오류발화와 가장 가까운 거리 내의 정답발화를 추출하여 오류 재보정을 수행할 수 있다.
- [0041] 즉, 오류보정기(27)는 임베딩된 정답 토큰별 밀집 벡터 공간 내에서 오류검출기(25)를 통해 검출된 오류발화 토큰으로부터 기 설정된 제2 맨하탄 거리 내에 있는 하나 이상의 정답발화 토큰을 정답 후보 보정군으로 추출하고, 정답 후보 보정군 중 해당 오류발화 토큰으로부터 맨하탄 거리 기준으로 가장 가까이 있는 정답발화 토큰을 정답으로 추출하는데, 만약 정답으로 추출된 발화에 대한 의도분석 실패 시, 오류보정기(27)는 추출된 정답 후보 보정군 중 이전 추출된 정답발화 토큰을 제외하고 해당 오류발화 토큰으로부터 맨하탄 거리 기준으로 가장 가까이 있는 정답발화 토큰을 정답으로 추출할 수 있다.
- [0042] 도 7은 본 발명의 예시적인 실시예에 따른 음성 인식 오류 보정 방법의 흐름도로, 전술한 도 1의 장치(20)에 적용되므로 해당 장치(20)의 동작과 병행하여 설명한다.
- [0043] 먼저, 도 5에 도시된 바와 같이 SNN(Spiking Neural Network) 기반의 bi-directional LSTM(Long Short Term Memory)을 이용하여 학습 및 모델링 된 인코더(23)를 구축하고, 해당 인코더(23)는 G2P 컨버터(21)를 통해 음성 인식(STT) 전사 데이터의 문자소(Grapheme)가 음소(Phoneme)로 변환되어 입력되면 G2P 컨버터(21)에서 변환된 음소(phoneme)를 P2V(Phoneme to Vector) 컨버터(미도시)를 통해 음소단위별 벡터로 임베딩하고, 음소단위별 임베딩된 벡터를 토큰별 벡터 공간으로 매핑함으로써, 음소 유사성을 갖는 정답발화 토큰과 오류발화 토큰 모두를 도 6의 예시와 같이 정답 토큰별 밀집 벡터 공간안에 위치하도록 임베딩 한다(S701).
- [0044] 이어, 오류검출기(25)는 단계 S701에서 인코더(23)를 통해 임베딩된 정답 토큰별 밀집 벡터 공간안의 정답발화 토큰을 기준으로 기 설정된 제1 맨하탄(Manhatan) 거리 내에 있는 하나 이상의 발화 토큰을 오류발화 토큰으로 검출한다(S703). 예를 들어, 단계 S703에서 오류검출기(25)는 도 6에 예시된 토큰별 밀집 벡터 공간에서 정답으로 태깅된 정답발화 토큰 '방탄소년단'을 기준으로 기 설정된 제1 맨하탄(Manhatan) 거리 내의 정답 태깅 없는 하나 이상의 발화 토큰 '달탐사소년단', '홍탄소년단' 등을 오류발화 토큰으로 검출할 수 있다.
- [0045] 이어, 오류보정기(27)는 단계 S701에서 인코더(23)를 통해 임베딩된 정답 토큰별 밀집 벡터 공간 내에서 오류검출기(25)를 통해 검출된 오류발화 토큰으로부터 맨하탄 거리 기준으로 가장 가까이 있는 정답발화 토큰을 정답으로 추출하여 오류를 보정한다(S705). 예를 들어, 단계 S705에서 오류보정기(27)는 도 6에 예시된 정답 토큰별 밀집 벡터 공간에서 발화 토큰 '달탐사소년단'이 오류발화 토큰으로 검출된 경우 해당 오류발화 토큰 '달탐사소년단'으로부터 맨하탄 거리 기준으로 가장 가까이 있는 (태깅된) 정답발화 토큰 '방탄소년단'을 정답으로 추출하여 오류를 보정 즉, '달탐사소년단'을 '방탄소년단'으로 보정한다(S705).
- [0046] 마지막으로, 오류보정기(27)를 통해 추출/보정된 정답발화는 의도분석기(30)에서 의도분석 후 출력하는데, 오류보정기(27)는 의도분석기(30)에서의 의도분석 성공 여부를 판단하고(S707), 단계 S707에서 의도분석 성공으로 판단되면 첫 번째 추출/보정된 정답발화를 그대로 정상 출력하도록 하며(S709), 단계 S707에서 의도분석 실패로 판단되면 단계 S705의 과정으로 돌아가 실패한 의도분석에 해당되는 발화에 대해 오류 재보정을 진행한다(S707-S705).
- [0047] 오류보정기(27)에서의 재보정 과정은 임베딩된 토큰별 밀집 벡터 공간내에서 첫 번째 반환된 정답발화를 제외하고 해당 오류발화와 가장 가까운 거리 내의 정답발화를 추출하여 오류 재보정을 수행할 수 있다. 즉, 오류보정기(27)는 임베딩된 정답 토큰별 밀집 벡터 공간 내에서 오류검출기(25)를 통해 검출된 오류발화 토큰으로부터 기 설정된 제2 맨하탄 거리 내에 있는 하나 이상의 정답발화 토큰을 정답 후보 보정군으로 추출하고, 정답 후보 보정군 중 해당 오류발화 토큰으로부터 맨하탄 거리 기준으로 가장 가까이 있는 정답발화 토큰을 정답으로 추출하는데, 만약 정답으로 추출된 발화에 대한 의도분석 실패 시, 오류보정기(27)는 추출된 정답 후보 보정군 중 이전 추출된 정답발화 토큰을 제외하고 해당 오류발화 토큰으로부터 맨하탄 거리 기준으로 가장 가까이 있는 정답발화 토큰을 정답으로 추출할 수 있다.

- [0048] 전술한 본 발명의 예시적인 실시예에 따르면, SNN을 기반으로 학습 및 모델링 된 인코더(23)를 구축하되, 해당 인코더(23)를 통하여 자연어처리의 전처리 단계에서 발화 텍스트(토큰)를 임베딩 할 때, 음소 유사성을 갖는 정답발화 토큰과 오류발화 토큰 모두를 정답 토큰별 밀집 벡터 공간안에 위치하도록 임베딩 할 수 있다.
- [0049] 통상적인 자연어전처리 단계에서의 임베딩은 오류 발화 가능성을 고려하지 않은 채로 음소별, 단어별, 어구별로 인코딩하는 반면, 전술한 본 발명의 예시적인 실시예에 따르면 자연어 전처리 단계에서 음성인식 오류 발화와 정답 발화 사이의 유사도를 반영하여 임베딩하고, 이에 의해 임베딩된 정답 토큰별 밀집 벡터 공간을 기초로 특정 발화 토큰들의 오류 유사 후보군들과 반대로 오류 발화 토큰들의 정답 발화 후보군을 추출할 수 있으며, 또한 해당 인코더의 구축 시 의도분석 결과가 함께 학습데이터로 활용함으로써, 통상적인 음성인식의 음향모델 및 언어모델에서 자체 보정되지 못한 음성인식 오류를 의도분석 전단계에서 보정하여 의도분석 정확도와 서비스 성공률을 높일 수 있고, 음향 모델에서 예외처리 되는 신조어 및 상품명 등을 보정할 수 있다.
- [0050] 한편, 전술한 음성 인식 오류 보정 방법에 따르면 해당 방법을 컴퓨터에서 실행시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록 매체를 구현할 수 있다.
- [0051] 또 한편, 전술한 음성 인식 오류 보정 방법에 따르면 해당 방법을 하드웨어와 결합하여 실행시키기 위하여 컴퓨터로 읽을 수 있는 기록 매체에 저장된 애플리케이션을 구현할 수 있다.
- [0052] 또 다른 한편, 전술한 음성 인식 오류 보정 방법에 따르면 해당 방법을 컴퓨터에서 실행시키기 위하여 컴퓨터로 읽을 수 있는 기록 매체에 저장된 컴퓨터 프로그램을 구현할 수 있다.
- [0053] 예를 들어, 전술한 바와 같이 본 발명의 예시적인 실시예에 따른 음성 인식 오류 보정 방법은 다양한 컴퓨터로 구현되는 동작을 수행하기 위한 프로그램 명령을 포함하는 컴퓨터 판독가능 기록 매체 또는 이러한 기록 매체에 저장된 애플리케이션으로 구현될 수 있다. 상기 컴퓨터 판독 가능 기록 매체는 프로그램 명령, 로컬 데이터 파일, 로컬 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 기록 매체는 본 발명의 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체, CD-ROM, DVD와 같은 광기록 매체, 플롭티컬 디스크와 같은 자기-광 매체, 및 롬, 램, 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 포함할 수 있다.
- [0054] 이상의 설명은 본 발명의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 다양한 수정 및 변형이 가능할 것이다. 따라서, 본 발명에 개시된 실시예들은 본 발명의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이고, 이러한 실시예에 의하여 본 발명의 기술 사상의 범위가 한정되는 것은 아니다. 본 발명의 보호 범위는 아래의 청구범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 발명의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

부호의 설명

- [0055] 10: 음성인식기
20: 음성 인식 오류 보정 장치
20a: 임베딩 사전
21: G2P 컨버터
23: 인코더
25: 오류검출기
27: 오류보정기
30: 의도분석기

도면3

한글	NC	한글	NC	한글	NC
ㅏ	a	ㅓ	ɛE	ㅗ	wv
ㅕ	E	ㅛ	o	ㅜ(ㅟ)	we
ㅑ	ja	ㅕ	wa	ㅛ	wi
ㅓ	je	ㅓ	wE	ㅛ	ju
ㅗ	v	ㅓ(ㅟ)	we	ㅛ	U
ㅓ	e	ㅓ	jo	ㅓ	Wi
ㅗ	jv	ㅓ	u	ㅓ	i

place of articulation Manner of articulation		Labial	Alveolar	Postalveolar	Velar	Glottal
Stops(파열음) & Affricates (마찰음)	Plain/voiced	ㅍ b	ㄸ d	ㅌ ɸ	ㅊ g	
	Tense	ㅍ B	ㄸ D	ㅌ Z	ㅊ G	
	Aspirated	ㅍ p	ㄸ t	ㅌ c	ㅊ k	
	implosive	ㅍ bx	ㄸ dx		ㅊ gx	
Fricatives (마찰음)	Plain/voiced		ㅅ s			ㅎ h
	Tense	ㅍ m	ㅅ S			
Nasal(비음)			ㄴ n		ㅇ o	Nx
Liquids(유음)			ㄹ r			

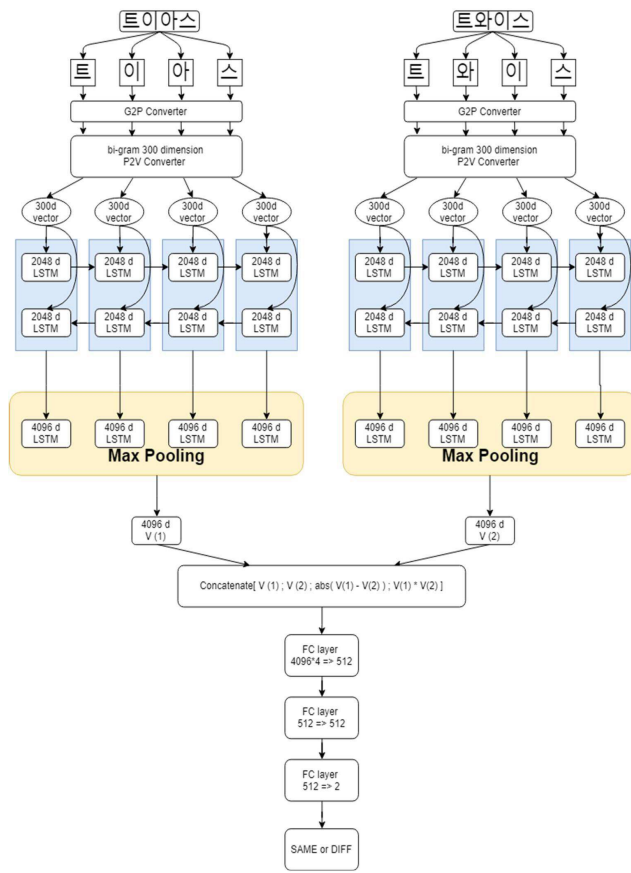
도면4

Grapheme	Phoneme	Alphabet
굳이 gud-i	구지 gu- ji	ㄱ ㅌ ㅈ ㅏ
밭에 bat-e	바테 ba- te	ㅂ ㅏ ㅓ ㅓ

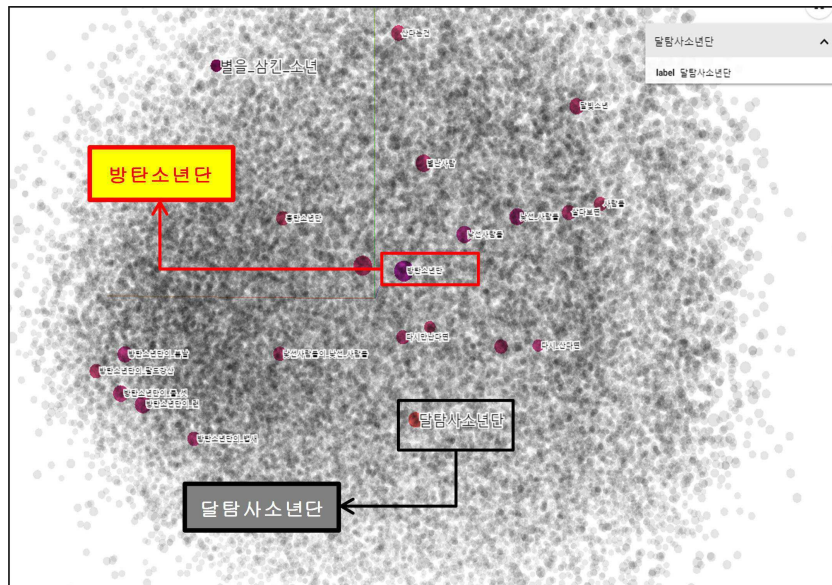
ㄱ ㅌ ㅈ ㅏ
ㅂ ㅏ ㅓ ㅓ

D-mensional vector
[32,45,52,11,0,9,7,10,82,71,4,8,12,...]

도면5



도면6



도면7

