

(19) 대한민국특허청(KR)
(12) 공개특허공보(A)(11) 공개번호 10-2022-0061329
(43) 공개일자 2022년05월13일

(51) 국제특허분류(Int. Cl.)
G06F 40/30 (2020.01) *G06F 40/126* (2020.01)
G06F 40/279 (2020.01) *G06N 3/08* (2006.01)
 (52) CPC특허분류
G06F 40/30 (2020.01)
G06F 40/126 (2020.01)
 (21) 출원번호 10-2020-0147029
 (22) 출원일자 2020년11월05일
 심사청구일자 2021년11월12일

(71) 출원인
 기초과학연구원
 대전광역시 유성구 엑스포로 55(도룡동)
 (72) 발명자
 김재현
 경상남도 창원시 의창구 평산로 159번길 46
 원동희
 대전광역시 유성구 엑스포로339번길 320, 10동
 104호 (원촌동, 싸이언스빌)
 차미영
 대전 유성구 가정로 43 삼성한울아파트 107-1101
 (74) 대리인
 특허법인 고려

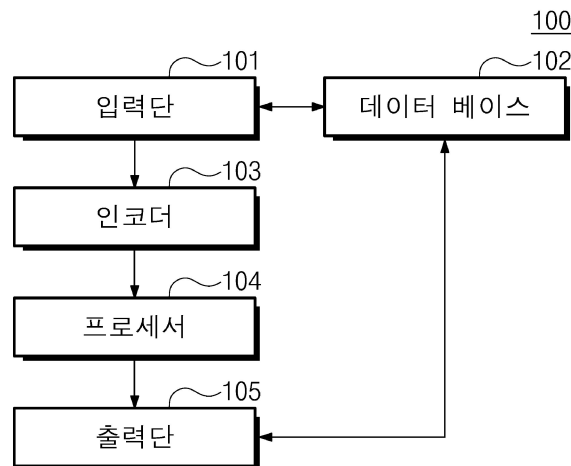
전체 청구항 수 : 총 16 항

(54) 발명의 명칭 **협오 표현을 처리하기 위한 언어 처리 장치 및 방법**

(57) 요약

개시된 일 측면에 언어 처리 장치는 따른 글자와 이모트 또는 이모지가 포함된 입력 단어를 입력받고, 상기 입력 단어를 글자 단위로 분리하는 입력단; 분리된 상기 글자들을 인코딩하여 벡터화하는 인코더; 및 상기 분리된 글자들의 인코딩된 벡터들을 미리 정해진 LSTM(Long Short Term Memory) 알고리즘에 기초하여 학습하고, 학습 결과 상기 이모트 또는 이모지에 대응되는 문자를 판단하고, 판단결과에 기초하여 상기 이모트 또는 이모지에 대응하는 문자를 예측하는 프로세서를 포함한다.

대표도 - 도2



(52) CPC특허분류

G06F 40/279 (2020.01)

G06N 3/08 (2013.01)

G10L 15/26 (2013.01)

명세서

청구범위

청구항 1

이모트 또는 이모지가 포함된 입력 단어를 입력 받고, 상기 입력 단어를 문자 단위로 분리하는 입력단;
분리된 상기 문자들을 인코딩하여 벡터화하는 인코더;및

상기 분리된 문자들의 인코딩된 벡터들을 미리 정해진 LSTM(Long Short Term Memory) 알고리즘에 기초하여 학습하고, 학습 결과 상기 이모트 또는 이모지에 대응되는 문자를 판단하고, 판단결과에 기초하여 상기 이모트 또는 이모지에 대응하는 문자를 예측하는 프로세서를 포함하는 언어 처리 장치.

청구항 2

제1 항에 있어서,
상기 프로세서는,
상기 학습의 결과를 피드백하고, 피드백 결과를 계속하여 저장하는 언어 처리 장치.

청구항 3

제2 항에 있어서,
상기 프로세서는,
상기 학습 결과 입력 단어의 의미를 미리 설정된 기준에 기초하여 혐오 표현, 공격성 언어 또는 일반 언어 중 어느 하나로 분류하는 언어 처리 장치.

청구항 4

제3 항에 있어서,
상기 프로세서는,
상기 혐오 표현은 성적 표현 또는 인종 차별적인 표현을 포함하는지 판단하는 언어 처리 장치.

청구항 5

제1 항에 있어서,
상기 프로세서는,
예측된 상기 이모트 또는 이모지에 대응하는 문자에 기초하여, 입력 단어를 복원하는 것을 더 포함하는 언어 처리 장치.

청구항 6

제1 항에 있어서,
상기 프로세서는,
온라인 스트리밍 화면 상에 표시되는 단어를 상기 입력 단어로 결정하는 언어 처리 장치.

청구항 7

제1 항에 있어서,
상기 입력단은,
외부로부터 음성신호를 입력 받고 입력된 음성 신호를 문자화하고,

상기 프로세서는,
입력된 상기 음성신호에 기초하여 문자를 생성하는 언어 처리 장치.

청구항 8

제7 항에 있어서,
상기 프로세서는,
생성된 상기 문자에 기초에 기초하여 상기 음성신호에 혐오 표현 또는 공격성 언어 중 적어도 하나가 포함되어 있는지 판단하는 언어 처리 장치.

청구항 9

글자와 이모트 또는 이모지가 포함된 입력 단어를 입력받고, 상기 입력 단어를 글자 단위로 분리하고,
분리된 상기 글자들을 인코딩하여 벡터화하고,
상기 분리된 글자들의 인코딩된 벡터들을 미리 정해진 LSTM(Long Short Term Memory) 알고리즘에 기초하여 학습하고, 학습 결과 상기 이모트 또는 이모지에 대응되는 문자를 판단하고, 판단결과에 기초하여 상기 이모트 또는 이모지에 대응하는 문자를 예측하는 것을 포함하는 언어 처리 방법.

청구항 10

제9 항에 있어서,
상기 LSTM(Long Short Term Memory) 알고리즘에 기초하여 학습하는 것은,
상기 학습의 결과를 피드백하고, 피드백결과를 계속하여 저장하는 언어 처리 방법.

청구항 11

제10 항에 있어서,
상기 피드백 결과는,
상기 학습 결과 입력 단어의 의미를 미리 설정된 기준에 기초하여 혐오 표현, 공격성 언어 또는 일반 언어 중 어느 하나로 분류하는 언어 처리 방법.

청구항 12

제11 항에 있어서,
상기 LSTM(Long Short Term Memory) 알고리즘에 기초하여 학습하는 것은,
상기 혐오 표현은 성적 표현 또는 인종 차별적인 표현을 포함하는지 판단하는 언어 처리 방법.

청구항 13

제9 항에 있어서,
예측된 상기 이모트 또는 이모지에 대응하는 문자에 기초하여, 입력 단어를 복원하는 것을 더 포함하는 언어 처리 방법.

청구항 14

제9 항에 있어서,
상기 이모트 또는 이모지가 조합된 입력 단어는
온라인 스트리밍 화면 상에 표시되는 언어 처리 방법.

청구항 15

제9 항에 있어서,
외부로부터 음성신호를 입력 받는 것;을 더 포함하고,
입력된 상기 음성신호에 기초하여 문자를 생성하는 언어 처리 방법.

청구항 16

제15 항에 있어서,
생성된 상기 문자에 기초에 기초하여 상기 음성신호에 혐오 표현 또는 공격성 언어 중 적어도 하나가 포함되어 있는지 판단하는 언어 처리 방법.

발명의 설명

기술 분야

[0001] 본 발명은 자연 언어 처리 장치 및 자연 언어 처리 방법에 관한 것으로, 더욱 상세하게는 텍스트 및 이모트(Emote) 또는 이모지(Emoji)를 포함하는 그림 형태의 문자로부터 혐오 표현을 감지하고, 이를 처리하는 방법 및 장치에 관한 것이다.

배경 기술

[0002] 소셜 네트워크 서비스(SNS: Social Network Service)가 발달하면서, 인터넷 상에서 제공되는 콘텐츠의 양이 증가하고 있다. 인터넷 상에서 제공되는 콘텐츠의 양이 증가함에 따라서, 콘텐츠 수요자들은 실시간으로 제공되는 콘텐츠를 평가할 수 있게 되었다. 여기서, 인터넷 상에서 제공되는 콘텐츠는 온라인 스트리밍 서비스, 동영상 제공 서비스, SNS상의 사진 또는 콘텐츠 게재 등이 있으나 이에 한정되는 것은 아니다.

[0003] 그런데, 일부 수요자들은 욕설, 인종 차별성 언행, 성적 수치심이 드는 언행 또는 외모 차별성 언행 등 혐오스러운 표현을 사용하여 제공되는 콘텐츠를 평가한다. 이를 방지하기 위하여, 인터넷상에서 사용되는 텍스트의 의미를 분석하는 기술이 개발되고 있지만, 이모티콘, 사진 또는 이미지 등을 통하여 이루어지는 혐오 표현에 대하여는 그 의미를 분석하고 처리하는 기술이 존재하지 않는 문제점이 존재한다.

발명의 내용

해결하려는 과제

[0004] 상술한 문제점을 해결하기 위하여 본 발명은 인터넷상에서 대화자가 사용하는 언어 및 이모트(Emote) 또는 이모지(Emoji)를 포함하는 그림 형태의 문자를 분석하고, 분석결과 대화자가 사용하는 언어 및 이모트 또는 이모지를 포함하는 그림 형태의 문자에 혐오 표현이 포함된 것으로 판단되면, 이를 처리하는 방법 및 장치를 제공한다.

과제의 해결 수단

[0005] 상술한 기술적 과제를 달성하기 위한 기술적 수단으로서, 일 측면에 언어 처리 장치는 이모트 또는 이모지가 포함된 입력 단어를 입력받고, 상기 입력 단어를 글자 단위로 분리하는 입력단; 분리된 상기 글자들을 인코딩하여 벡터화하는 인코더; 및 상기 분리된 글자들의 인코딩된 벡터들을 미리 정해진 LSTM(Long Short Term Memory) 알고리즘에 기초하여 학습하고, 학습 결과 상기 이모트 또는 이모지에 대응되는 문자를 판단하고, 판단결과에 기초하여 상기 이모트 또는 이모지에 대응하는 문자를 예측하는 프로세서를 포함한다.

[0006] 또한, 상기 프로세서는, 상기 학습의 결과를 피드백하고, 피드백 결과를 계속하여 저장할 수 있다.

[0007] 또한, 상기 프로세서는, 상기 학습 결과 입력 단어의 의미를 미리 설정된 기준에 기초하여 혐오 표현, 공격성 언어 또는 일반 언어 중 어느 하나로 분류할 수 있다.

[0008] 또한, 상기 프로세서는, 상기 혐오 표현은 성적 표현 또는 인종 차별적인 표현을 포함하는지 판단하는 언어 처리 장치.

- [0009] 또한, 상기 프로세서는, 예측된 상기 이모트 또는 이모지에 대응하는 문자에 기초하여, 입력 단어를 복원할 수 있다.
- [0010] 또한, 상기 프로세서는, 온라인 스트리밍 화면 상에 표시되는 단어를 상기 입력 단어로 결정하는 언어 처리 장치.
- [0011] 또한, 상기 입력단은, 외부로부터 음성신호를 입력 받고 입력된 음성 신호를 문자화하고, 상기 프로세서는, 입력된 상기 음성신호에 기초하여 문자를 생성할 수 있고, 상기 프로세서는, 생성된 상기 문자에 기초에 기초하여 상기 음성신호에 혐오 표현 또는 공격성 언어 중 적어도 하나가 포함되어 있는지 판단할 수 있다.
- [0012] 개시된 다른 측면에 따른 언어 처리 방법은 글자와 이모트 또는 이모지가 포함된 입력 단어를 입력받고, 상기 입력 단어를 글자 단위로 분리하고, 분리된 상기 글자들을 인코딩하여 벡터화하고, 상기 분리된 글자들의 인코딩된 벡터들을 미리 정해진 LSTM(Long Short Term Memory) 알고리즘에 기초하여 학습하고, 학습 결과 상기 이모트 또는 이모지에 대응되는 문자를 판단하고, 판단결과에 기초하여 상기 이모트 또는 이모지에 대응하는 문자를 예측하는 것을 포함한다.
- [0013] 또한, 상기 LSTM(Long Short Term Memory) 알고리즘에 기초하여 학습하는 것은, 상기 학습의 결과를 피드백하고, 피드백결과를 계속하여 저장하는 언어 처리 방법.
- [0014] 또한, 상기 피드백 결과는, 상기 학습 결과 입력 단어의 의미를 미리 설정된 기준에 기초하여 혐오 표현, 공격성 언어 또는 일반 언어 중 어느 하나로 분류하는 것을 포함할 수 있다.
- [0015] 또한, 상기 LSTM(Long Short Term Memory) 알고리즘에 기초하여 학습하는 것은,
- [0016] 상기 혐오 표현은 성적 표현 또는 인종 차별적인 표현을 포함하는지 판단하는 것을 포함할 수 있다.
- [0017] 또한, 예측된 상기 이모트 또는 이모지에 대응하는 문자에 기초하여, 입력 단어를 복원하는 것을 더 포함하는 언어 처리 방법.
- [0018] 또한, 상기 이모트 또는 이모지가 조합된 입력 단어는 온라인 스트리밍 화면 상에 표시되는 언어 처리 방법.
- [0019] 또한, 개시된 언어 처리 방법은 외부로부터 음성신호를 입력 받는 것;을 더 포함하고,
- [0020] 입력된 상기 음성신호에 기초하여 문자를 생성할 수 있다.
- [0021] 또한, 생성된 상기 문자에 기초에 기초하여 상기 음성신호에 혐오 표현 또는 공격성 언어 중 적어도 하나가 포함되어 있는지 판단할 수 있다.

발명의 효과

- [0022] 상술한 구성을 포함함으로써, 본 발명은 이모트 또는 이모지를 포함하는 그림 형태의 문자가 포함된 표현에서 표현의 의미를 분석하고, 혐오 표현을 추출할 수 있는 효과가 존재한다.

도면의 간단한 설명

- [0023] 도 1은 본 발명에 따른 언어 처리 장치의 입력과 출력을 도시한 것이다.
- 도 2는 본 발명에 따른 언어 처리 장치의 구성을 도시한 블록도이다.
- 도 3은 본 발명에 따른 프로세서를 도시한 것이다.
- 도 4은 본 발명에 따른 언어 처리 장치가 이모트에 대응되는 문자를 예측하는 과정을 도시한 순서도이다.
- 도 5는 본 발명에 따른 언어 처리 장치가 외부의 음성에 대응하는 문자열을 생성하는 과정을 도시한 순서도이다.
- 도 6는 본 발명에 따른 학습 모델을 적용한 경우, 본 개시가 기존 언어 처리 방법과 비교하여 더 많은 혐오 표현을 검출한 결과를 나타낸 것이다.

발명을 실시하기 위한 구체적인 내용

- [0024] 명세서 전체에 걸쳐 동일 참조 부호는 동일 구성요소를 지칭한다. 본 명세서가 실시예들의 모든 요소들을 설명하는 것은 아니며, 본 발명이 속하는 기술분야에서 일반적인 내용 또는 실시예들 간에 중복되는 내용은 생략한

다. 명세서에서 사용되는 '부, 모듈, 부재, 블록'이라는 용어는 소프트웨어 또는 하드웨어로 구현될 수 있으며, 실시예들에 따라 복수의 '부, 모듈, 부재, 블록'이 하나의 구성요소로 구현되거나, 하나의 '부, 모듈, 부재, 블록'이 복수의 구성요소들을 포함하는 것도 가능하다.

- [0025] 명세서 전체에서, 어떤 부분이 다른 부분과 "연결"되어 있다고 할 때, 이는 직접적으로 연결되어 있는 경우뿐 아니라, 간접적으로 연결되어 있는 경우를 포함하고, 간접적인 연결은 무선 통신망을 통해 연결되는 것을 포함한다.
- [0026] 또한 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.
- [0027] 제 1, 제 2 등의 용어는 하나의 구성요소를 다른 구성요소로부터 구별하기 위해 사용되는 것으로, 구성요소가 전술된 용어들에 의해 제한되는 것은 아니다.
- [0028] 단수의 표현은 문맥상 명백하게 예외가 있지 않는 한, 복수의 표현을 포함한다.
- [0029] 각 단계들에 있어 식별부호는 설명의 편의를 위하여 사용되는 것으로 식별부호는 각 단계들의 순서를 설명하는 것이 아니며, 각 단계들은 문맥상 명백하게 특정 순서를 기재하지 않는 이상 명기된 순서와 다르게 실시될 수 있다.
- [0030] 본 명세서에서 개시되는 일부 실시 예들에서, 언어 처리 장치 및 방법에 이용되는 그림 형태의 문자로서 이모트(Emote)가 이용되는 것으로 개시되나, 이는 발명을 쉽게 설명하기 위한 일부 실시 예일 뿐, 본 발명에 이용되는 문자의 종류를 한정하는 것은 아니다. 예로서, 본 발명에 따른 언어 처리 장치 및 방법에 이용되는 문자는 이모지(Emoji)를 포함하는 그림 형태의 문자일 수 있다. 이하 첨부된 도면들을 참고하여 본 발명의 작용 원리 및 실시예들에 대해 설명한다.
- [0031] 도 1은 본 발명에 따른 언어 처리 장치(100)의 입력과 출력을 도시한 것이다.
- [0032] 도 1을 참조하면, 본 발명에 따른 언어 처리 장치(100)는 복수의 문자들(문자 1, 문자 2, 문자 3, ..., 문자 n)과 복수의 이모트들(이모트 1, 이모트 2, 이모트 3, ..., 이모트n)을 입력 받고, 문자 또는 이모트 각각에 대응하는 출력들(출력1, 출력2, 출력3, ..., 출력n)을 생성한다. 구체적으로, 문자는 영어 알파벳, 한글의 자음과 모음 및 글자 또는 기타 외국어 문자를 포함할 수 있다. 또한, 이모트는 도안화된 문자, 이모티콘 또는 이미지를 포함하나 이에 한정되지 아니하고, 문자가 아닌 표현으로 의미를 전달할 수 있는 다른 수단을 더 포함할 수 있다. 언어 처리 장치(100)의 입력이 문자인 경우, 출력은 입력된 문자가 그대로 출력된다. 그러나, 언어 처리 장치(100)의 입력이 이모트인 경우, 출력은 입력된 이모트에 대응하는 문자로 변환되어 생성된다.
- [0033] 예를들어, 이모트 1이 영문 알파벳 B, 이모트 3이 영문 알파벳 b 및 이모트 4는 영문 알파벳 y이고, 이모트 2는 사람 모양의 이미지인 경우, 이모트 1, 3 및 4에 대응하는 B, b, y로부터 이모트 2에 대응하는 알파벳을 a로 결정하고, Baby라는 단어를 생성할 수 있다. 또한, 영문 알파벳을 예로 들어 설명하였으나, 이모트들은 한글, 한자 기타 문자 또는 이미지로 표현될 수 있는 것들을 포함할 수 있다.
- [0034] 도 2은 본 개시에 따른 언어 처리 장치(100)의 구성을 도시한 블록도이다.
- [0035] 도 2를 참조하면, 본 발명에 따른 언어 처리 장치(100)는 입력단(101), 데이터 베이스(102), 인코더(103), 프로세서(104) 및 출력단(105)를 포함한다. 다만, 실시 예에 따라서, 데이터 베이스(102)는 미리 저장된 언어 처리 결과 및 언어 학습 결과를 저장하기 위한 것으로써, 언어 처리 장치(100)의 구성이 아닌 것으로 여겨질 수도 있다.
- [0036] 입력단(101)은 외부로부터 이모트가 포함된 문자열을 입력 받는다. 여기서 이모트는 이모티콘, 이미지 또는 사진을 포함하나 이에 한정되는 것은 아니고, 글자로 구성되지 않은 다른 의미 전달 표현을 포함할 수 있다. 또한, 문자는 단어 및 복수의 단어들로 구성된 문장을 포함한다. 입력단(101)은 이모트가 포함된 단어를 입력 받고, 입력된 단어를 글자 단위로 분리한다. 또한, 입력단(101)은 분리된 글자들을 데이터 베이스(102) 및 인코더(103)에 입력한다.
- [0037] 데이터 베이스(102)는 입력단(101)에 미리 저장된 언어 처리 결과 및 언어 학습 결과를 입력 할 수 있다. 여기서 미리 저장된 언어 처리 결과 및 언어 학습 결과는 프로세서(104)의 언어 처리 결과 및 언어 학습 결과이다. 프로세서(104)가 입력된 문자 또는 이모트에 대하여 학습하고, 언어를 처리 하는 과정은 프로세서(104)의 설명에서 상세히 서술한다.

- [0038] 인코더(103)는 입력 단어를 학습하기 위하여 인코딩을 수행한다. 일 실시예에 따른 언어 처리 방법은 입력 단어들에 대하여 정수 인코딩(Integer Encoding)을 수행하고, 정수화된 문자들에 대하여 원-핫 인코딩(One-hot encoding)을 수행할 수 있으나, 인코딩 방법은 이에 한정되는 것은 아니고, 언어 처리에 사용되는 다른 인코딩 방법이 사용될 수 있다.
- [0039] 구체적으로 정수 인코딩이란 입력 단어에 정수를 부여하는 방법이다. 정수 인코딩 방법을 사용하면, 인코더(103)는 단어 이모트 또는 글자를 빈도수 순으로 정렬한 단어 집합(vocabulary)을 만들고, 빈도수가 높은 순서대로 차례로 낮은 숫자부터 정수를 부여할 수 있다. 정수 인코딩은 원-핫 인코딩을 수행하기 전 단계로 수행될 수 있다. 정수 인코딩을 수행한 결과, 인코더(103)는 단어 집합(vocabulary)을 생성한다. 여기서 단어 집합은 텍스트의 모든 단어를 중복을 허용하지 않고 모아놓은 것을 의미한다. 예를 들어, 입력된 문자열에 입력 단어가 총 5,000개가 존재한다면, 단어 집합의 단어들마다 1번부터 5,000번까지 인덱스를 부여하고, 단어 집합의 크기는 5,000이 된다. 이러한 단어 집합에 book, dog, love 또는 books가 포함되어 있는 경우, book은 150번, dog는 171번, love는 192번, books는 212번과 같이 정수가 부여될 수 있다.
- [0040] 원-핫 인코딩은 컴퓨터에서 언어 처리를 원활하게 하기 위하여 문자를 숫자로 표현하는 방법이다. 구체적으로, 원-핫 인코딩은 정수 인코딩에 의해 생성된 단어 집합의 크기를 벡터의 차원으로 하고, 표현하고 싶은 단어 또는 글자의 인덱스에 1의 값을 부여하고, 다른 인덱스에는 0을 부여하는 단어의 벡터 표현 방식이다. 이렇게 표현된 벡터를 원-핫 벡터(One-hot vector)라고 한다. 원-핫 인코딩은 후술하는 프로세서(104)의 단어 학습을 위한 전처리 과정이다. 개시된 언어 처리 방법은 기본적으로 26개의 영문 알파벳 각각의 글자 및 문자부호에 대하여 원-핫 벡터를 생성하므로, 원-핫 인코딩 결과 총 $32 * 1$ 개의 벡터를 생성할 수 있다.
- [0041] 프로세서(104)는 인코더(103)가 생성한 벡터들에 대응하여 각각의 입력에 대응하는 출력을 생성한다. 또한, 프로세서(104)는 LSTM(Long Short Term Memory)알고리즘에 기초하여 학습하고, 학습 결과 입력된 이모트에 대응되는 문자를 판단하고, 판단결과에 기초하여 입력된 이모트에 대응하는 문자를 예측할 수 있다. 여기서 학습 결과는 이모트가 포함된 문자, 이모트가 포함되지 않은 문자 및 입력 단어들에 대응되는 의미일 수 있다.
- [0042] 또한, 프로세서(104)는 데이터 베이스(102)에 학습 결과를 미리 설정된 기준에 기초하여 혐오 표현, 공격성 언어 또는 일반 언어 중 어느 하나로 분류하여 저장할 수 있다. 여기서 미리 설정된 기준은 공지된 표준어 사전 및 공지된 혐오표현 또는 공격성 언어에 의하여 정해질 수 있다. 즉, 미리 설정된 기준은 입력 단어가 혐오 표현 또는 공격성 언어를 포함하고 있는지 필터링 하는 기준이 되고, 이는 공지된 표준어 사전 및 공지된 혐오 표현 또는 공격성 언어를 필터링하는 것으로 정해질 수 있다. 또한, 혐오 표현은 성적 표현 또는 인종 차별적인 표현을 포함할 수 있다. 여기서 성적 표현 및 인종 차별적인 표현은 일반인의 상식에 비추어 수치심이나 모욕감을 들게 하는 언어를 의미한다. 또한, 공격성 언어는 욕설 등 일반인의 상식에 비추어 수치심이나 모욕감을 들게 하는 언어를 의미한다. 또한, 프로세서(104)는 데이터 베이스(102)에 학습 결과의 피드백 결과를 저장할 수 있다. 여기서 학습 결과의 피드백 결과는 동일한 문자적 의미에 대응하는 이모트의 변형일 수 있으나 이에 한정되지 않는다.
- [0043] 또한, 프로세서(104)는 데이터 베이스(102)에 저장된 입력 단어의 학습 결과 및 피드백 결과에 기초하여 새로 입력된 단어가 데이터 베이스(102)에 존재하지 여부를 검색할 수 있다. 프로세서(104)는 새로 입력된 단어가 데이터 베이스(102)에 존재하면, 단어의 학습 과정을 거치지 않고 곧바로 이모트가 포함된 문자에 대응되는 문자를 예측할 수 있다.
- [0044] 출력단(105)은 프로세서(104)의 출력을 외부로 전송하거나 출력한다. 구체적으로, 출력단(105)은 복원된 문자를 외부에 마련된 표시 장치(미도시)로 전송할 수 있다. 여기서, 표시 장치(미도시)는 외부에 마련된 모니터 또는 사용자 단말을 포함할 수 있다. 사용자 단말(미도시)은 네트워크를 통해 (장치)에 접속할 수 있는 컴퓨터나 휴대용 단말기로 구현될 수 있다. 여기서, 컴퓨터는 예를 들어, 웹 브라우저(WEB Browser)가 탑재된 노트북, 데스크톱(desktop), 랩톱(laptop), 태블릿 PC, 슬레이트 PC 등을 포함하고, 휴대용 단말기는 예를 들어, 휴대성과 이동성이 보장되는 무선 통신 장치로서, PCS(Personal Communication System), GSM(Global System for Mobile communications), PDC(Personal Digital Cellular), PHS(Personal Handyphone System), PDA(Personal Digital Assistant), IMT(International Mobile Telecommunication)-2000, CDMA(Code Division Multiple Access)-2000, W-CDMA(W-Code Division Multiple Access), WiBro(Wireless Broadband Internet) 단말, 스마트 폰(Smart Phone) 등과 같은 모든 종류의 핸드헬드(Handheld) 기반의 무선 통신 장치와 시계, 반지, 팔찌, 발찌, 목걸이, 안경, 콘택트 렌즈, 또는 머리 착용형 장치(head-mounted-device(HMD)) 등과 같은 웨어러블 장치를 포함할 수 있다.

- [0045] 도 3은 본 발명에 따른 프로세서(104)를 도시한 것이다.
- [0046] 도 3을 참조하면, 프로세서(104)는 인코딩된 글자의 입력 벡터(Input)(401)를 입력받고, 미리 설정된 LSTM 알고리즘을 거쳐 출력 벡터(Output)(405)를 생성한다. 프로세서(104)는 제1 레이어(Forward Layer)(402) 및 제2 레이어(Backward Layer)(403)를 포함한다. 제1 레이어(402) 및 제2 레이어(403)은 프로세서(104)에 의해 실행되는 LSTM 알고리즘의 레이어이다. 여기서 제1 레이어(402)는 입력된 문자를 정방향으로 배치한 글자 벡터의 조합이 입력되는 레이어이고, 제2 레이어(403)는 입력된 문자를 역방향으로 배치한 글자 벡터의 조합이 입력되는 레이어이다.
- [0047] 예를 들어, 입력된 문자가 apple인 경우, 제1 레이어는 "a, p, p, l, e"에 해당하는 벡터가 순서대로 입력되고, 제2 레이어에는 "e, l, p, p, a"순서로 벡터가 입력된다. 개시된 프로세서(104)는 각각 LSTM을 수행하는 제1 레이어 및 제2 레이어를 포함함으로써, 이모트가 단어 중간에 삽입되어 있는 경우에도, 입력된 단어의 정확한 의미를 도출할 수 있다.
- [0048] 프로세서(104)는 인코더(103)로부터 글자의 입력 벡터(401)를 입력 받는다. 여기서 입력 벡터는 X_{t-1} , X_t , X_{t+1} 와 같이 표현된다. 다만, 입력벡터는 X_{t-1} , X_t , X_{t+1} 에 한정되는 것은 아니고, 단어가 포함하는 글자수에 따라 복수개의 벡터를 포함할 수 있다.
- [0049] 입력 벡터(401)가 입력되면, 프로세서(104)는 제1 레이어(402) 및 제2 레이어(403)에 각각의 입력 벡터를 인식시킨다. 도 2를 참조하면, 제1 레이어는 X_{t-1} , X_t , X_{t+1} 순서로 LSTM 학습을 수행하고, 제2 레이어(403)는 X_{t+1} , X_t , X_{t-1} 순서로 LSTM 학습을 수행한다. 프로세서(104)는 제1 레이어(402) 및 제2 레이어(403)에서 LSTM 학습이 완료되면, 활성화 값(activation)(404)을 생성한다. 여기서 제1 레이어(402) 및 제2 레이어(403)에서 생성한 활성화 값이 같은 것으로 판단되면, 프로세서(104)는 입력단어의 정방향 분석 및 역방향 분석이 일치하는 것으로 판단하고, 출력(Output)(405)을 생성한다. 출력(405)은 Y_{t-1} , Y_t , Y_{t+1} 로 표현되지만, 이에 한정되지 아니하고 출력되는 글자의 수와 일치하는 개수의 벡터를 생성할 수 있다.
- [0050] 여기서 LSTM 알고리즘은 단일 데이터 포인트 (예: 이미지)뿐만 아니라 전체 데이터 시퀀스 (예: 음성 또는 비디오)도 처리할 수 있고, 입력되는 데이터들을 피드백할 수 있다. 예를 들어 LSTM 알고리즘은 분할되지 않은 연결된 필기를 인식할 수 있다. 또한, LSTM 알고리즘은 입력된 음성을 인식할 수 있다. 구체적으로, 개시된 알고리즘은 혐오 표현을 빅데이터화 하여 학습하고, 이를 인코딩한다. 여기서 빅데이터화 되고, 인코딩 되는 혐오 표현은 문자 표현, 이모트 표현 및 문자와 이모트가 결합된 표현을 모두 포함한다.
- [0051] 도 2 및 도 3에 도시된 언어 처리 장치(100) 및 프로세서(104)의 구성 요소들의 성능에 대응하여 적어도 하나의 구성요소가 추가되거나 삭제될 수 있다. 또한, 구성 요소들의 상호 위치는 시스템의 성능 또는 구조에 대응하여 변경될 수 있다는 것은 당해 기술 분야에서 통상의 지식을 가진 자에게 용이하게 이해될 것이다.
- [0052] 한편, 도 2 및 도 3에서 도시된 각각의 구성요소는 소프트웨어 및/또는 하드웨어 구성요소를 의미한다.
- [0053] 도 4는 본 발명에 따른 언어 처리 장치(100)가 이모트에 대응되는 문자를 예측하는 과정을 도시한 순서도이다.
- [0054] 도 4를 참조하면, 개시된 언어 처리 방법은 이모트를 포함한 문자를 입력 함으로써 시작된다(S114).
- [0055] 이모트를 포함한 문자가 입력되면, 입력단(101)은 입력된 문자를 인코더(103)으로 전달하고, 인코더(103)는 입력된 문자를 인코딩 한다(S124). 여기서 인코딩 방식은 상술한 바와 같이 원-핫인코딩 방식일 수 있으나 이에 한정되지 않는다. 인코딩 결과, 입력 단어의 글자에 해당하는 벡터가 생성될 수 있다.
- [0056] 입력 단어에 대한 인코딩이 완료되면, 프로세서(104)는 LSTM 모델을 이용한 입력 단어의 학습을 수행한다. 입력 단어의 학습은 문자열에 포함된 각각의 글자에 대응되는 벡터를 추출하고, 추출된 벡터를 정방향 및 역방향으로 배치하고, 문자열의 순서에 기초하여 입력 단어에 포함된 이모트에 대응되는 글자를 예측하는 과정이다. 예를 들어, APPLE 이라는 단어에 존재하는 세번째 글자 P가 사과의 이미지(이모트)로 표현된 경우, 프로세서(104)는 정방향으로 AP<이모트>LE, 역방향으로 LE<이모트>PA로 표현된 문자임을 판단하고, 이모트에 대응되는 글자가 P인 것으로 판단할 수 있다.
- [0057] LSTM모델을 통한 학습이 완료되면, 프로세서(104)는 이모트에 대응되는 문자를 예측하고(S144), 이모트가 포함된 입력 단어의 문자열을 생성할 수 있다(S154).
- [0058] 도 5는 본 개시의 일 측면에 따른 프로세서(104)가 외부의 음성에 대응하는 문자열을 생성하는 과정을 도시한

순서도이다. 도 5를 참조하면 사용자가 단어의 일부 또는 문장의 일부를 표준어가 아닌 단어 또는 혐오 표현을 비유적으로 변형하여 발화하는 경우, 개시된 언어 처리 방법은 사용자의 음성을 문자화하고, 사용자의 음성에 대응되는 문자열을 생성할 수 있다.

[0059] 구체적으로, 개시된 언어 처리 장치(100)는 사용자의 음성을 인식한다(S114-1). 사용자의 음성은 외부에 마련된 마이크를 비롯한 입력 장치에 의하여 입력된 신호에 기초하여 인식될 수 있다. 사용자의 음성이 입력되면, 프로세서(104)는 입력된 음성을 문자화한다(S124-1). 입력된 음성이 문자화 되면, 프로세서(104)는 생성된 문자에 대하여 LSTM 모델을 통한 학습을 수행한다(S134-1). LSTM학습은 전술한 바와 같이 수행된다. 예를 들어, 영어 단어 apple을 사용자가 abble이라고 발음한 경우, 프로세서(104)는 a??le 및 el??a로부터 사용자가 apple을 말하려는 것임을 예측할 수 있다. 여기서 "?"는 입력 단어의 원래의 표현이 아닌 변형된 표현을 의미한다. 생성된 문자의 LSTM 학습이 완료되면, 프로세서(104)는 사용자의 음성에 대응하는 의미를 갖는 문자열을 예측하고, 생성한다(S144-1).

[0060] 도 6는 본 개시에 따른 학습 모델을 적용한 경우, 본 개시가 기존 언어 처리 방법과 비교하여 더 많은 혐오 표현을 검출한 결과를 나타낸 것이다.

[0061] 도 6을 참조하면, 본 개시의 효과를 검증하기 위하여 온라인 스트리밍 서비스에서 수집한 15,290,530개의 이모트를 포함한 채팅(Chats include emotes)에 대하여 혐오 표현을 감지하는 학습을 수행한 결과 종래 기술(Previous Model)은 혐오 표현(Hate Speech)를 694,738개 감지하였으나, 본 개시에 따른 LSTM 모델은 혐오 표현을 845,203개 감지하였다. 도 6을 참조하면, 본 개시는 이모트가 포함된 입력 단어에 혐오 표현이 포함되어 있는지에 대하여 종래 기술보다 혐오 표현을 현저히 많이 감지할 수 있는 효과가 존재한다.

[0062] 한편, 개시된 실시예들은 컴퓨터에 의해 실행 가능한 명령어를 저장하는 기록매체의 형태로 구현될 수 있다. 명령어는 프로그램 코드의 형태로 저장될 수 있으며, 프로세서에 의해 실행되었을 때, 프로그램 모듈을 생성하여 개시된 실시예들의 동작을 수행할 수 있다. 기록매체는 컴퓨터로 읽을 수 있는 기록매체로 구현될 수 있다.

[0063] 컴퓨터가 읽을 수 있는 기록매체로는 컴퓨터에 의하여 해독될 수 있는 명령어가 저장된 모든 종류의 기록 매체를 포함한다. 예를 들어, ROM(Read Only Memory), RAM(Random Access Memory), 자기 테이프, 자기 디스크, 플래쉬 메모리, 광 데이터 저장장치 등이 있을 수 있다.

[0064] 이상에서와 같이 첨부된 도면을 참조하여 개시된 실시예들을 설명하였다. 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자는 본 발명의 기술적 사상이나 필수적인 특징을 변경하지 않고도, 개시된 실시예들과 다른 형태로 본 발명이 실시될 수 있음을 이해할 것이다. 개시된 실시예들은 예시적인 것이며, 한정적으로 해석되어서는 안 된다.

부호의 설명

[0065] 100: 언어 처리 장치

101: 입력단

102: 데이터 베이스

103: 인코더

104: 프로세서

105: 출력단

402: 제1 레이어

403: 제2 레이어

402a, 403a: LSTM모델

402a-1: 삭제 게이트

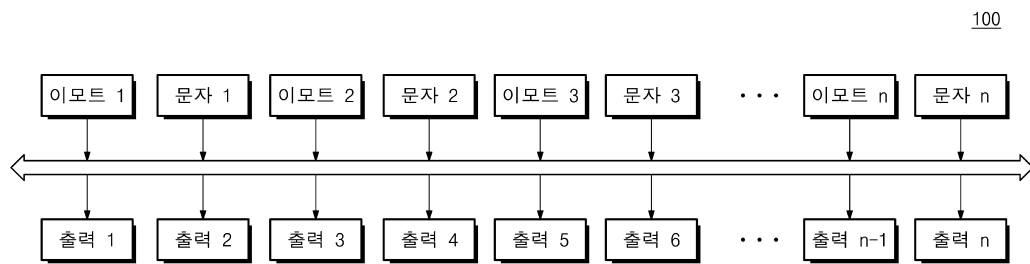
402a-2, 402a-3: 입력 게이트

402a-4, 402a-5: 출력 게이트

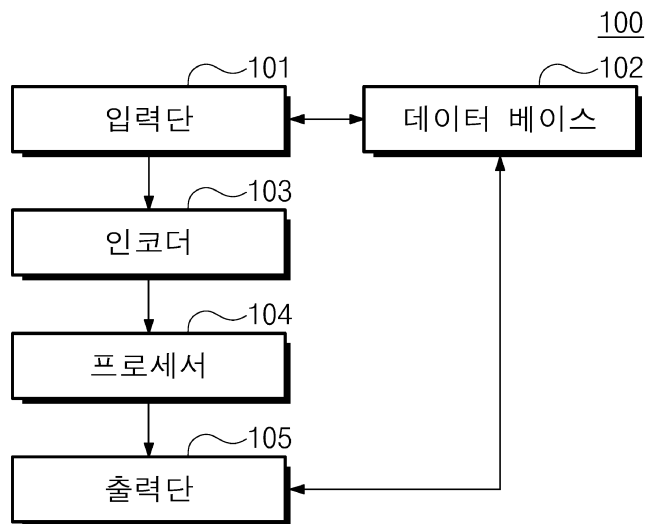
402a-6: 셀 상태

도면

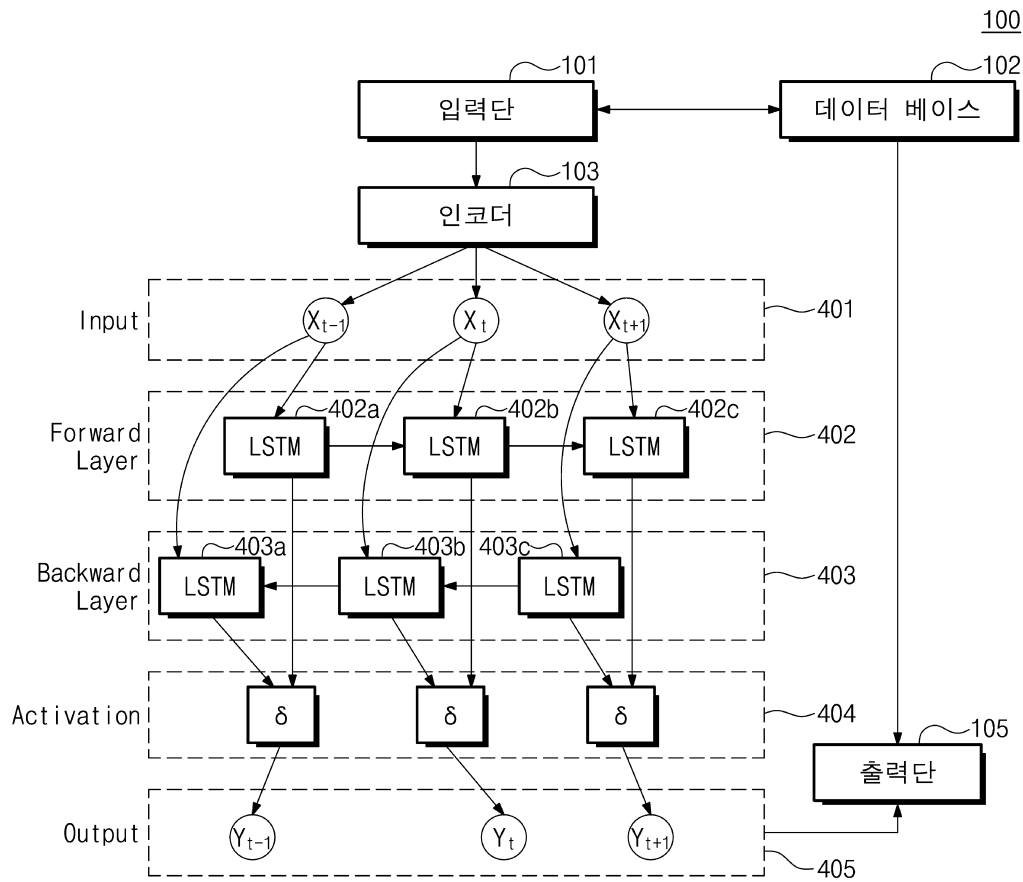
도면1



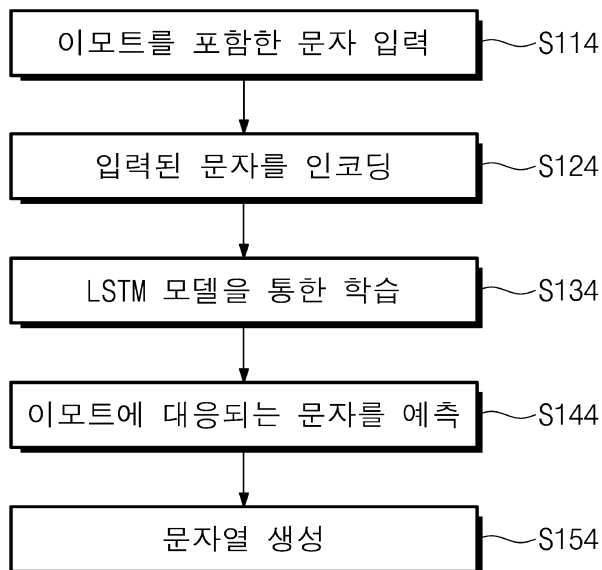
도면2



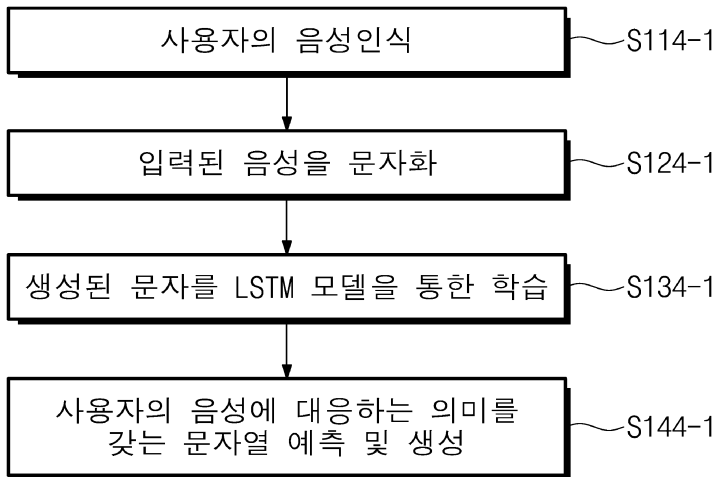
도면3



도면4



도면5



도면6

Chats include emotes	Hate Speech Detection (Previous Model)	Hate Speech Detection (Restored Model)
15,290,530	694,783(4.55%)	845,203(5.53%)