



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0118614  
(43) 공개일자 2020년10월16일

(51) 국제특허분류(Int. Cl.)  
G06F 40/20 (2020.01)

(52) CPC특허분류  
G06F 40/284 (2020.01)  
G06F 40/205 (2020.01)

(21) 출원번호 10-2019-0040734

(22) 출원일자 2019년04월08일  
심사청구일자 2019년04월08일

(71) 출원인

연세대학교 산학협력단

서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)

(72) 발명자

한요섭

서울특별시 은평구 진관1로 77-8, 403동 204호 (진관동, 은평뉴타운폭포동아파트)

박준우

경기도 부천시 소향로 11, 20층 B동 2002호(상동, 코오롱이데아폴리스)

이주형

서울특별시 서초구 신반포로 45, 70동 306호(반포동, 반포아파트)

(74) 대리인

민영준

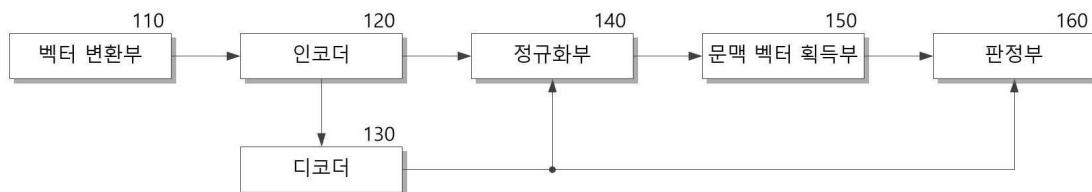
전체 청구항 수 : 총 8 항

(54) 발명의 명칭 학습 기반 비윤리어 탐지 장치 및 방법

(57) 요약

본 발명은 문장을 문자 단위로 구분함으로써, 띄어쓰기나 맞춤법과 같은 문법상의 오류에도 강인하게 비윤리어를 탐지할 수 있도록 하며, 문자별 특징 추출 시에 이전 특징이 누적 반영되도록 하여, 문자 단위로 구분됨으로 인해 문장의 의미 분석이 어려움에도 정확하게 비윤리어의 포함 여부를 판별할 수 있는 비윤리어 탐지 장치 및 방법을 제공할 수 있다.

대표도



(52) CPC특허분류

**G06F 40/242** (2020.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	2016-0-00562
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기술진흥센터(NIPA산하)
연구사업명	정보통신방송연구개발사업
연구과제명	[이지바로][주관/한국과학기술원] 상대방의 감성을 추론, 판단하여 그에 맞추어 대
화하고 대응할 수 있는 감성 지능 연구개발(3/5)	
기 여 율	1/1
과제수행기관명	한국과학기술원
연구기간	2018.07.01 ~ 2019.04.30

---

## 명세서

### 청구범위

#### 청구항 1

문장을 인가받아 문자 단위로 구분하고, 구분된 다수의 문자 각각을 문자 벡터로 변환하는 전처리부;

미리 학습된 패턴 추정 방식에 따라 다수의 문자 벡터를 순차적으로 인코딩하여 다수의 특징 벡터와 다수의 히든 벡터를 추출하되, 이전 문자 벡터를 인코딩할때 획득된 히든 벡터를 현재 문자 벡터와 함께 인코딩하는 인코더;

상기 다수의 히든 벡터 중 최종 히든 벡터를 인가받아 미리 학습된 패턴 인식 방식에 따라 디코딩하여 디코딩 벡터를 획득하는 디코더;

상기 다수의 특징 벡터 각각과 상기 디코딩 벡터 사이의 유사도를 비교하여 다수의 가중치를 획득하고, 획득된 다수의 가중치 각각을 대응하는 특징 벡터에 반영하여 다수의 정규화 벡터를 획득하는 정규화부;

상기 정규화 벡터를 미리 지정된 방식으로 결합하여 구문 벡터를 획득하는 구문 벡터 획득부; 및

미리 학습된 패턴 추정 방식에 따라 상기 구문 벡터의 특징을 추출하고, 추출된 구문 벡터의 특징을 분류하여 비윤리어 포함 여부를 판별하는 판정부; 를 포함하는 비윤리어 탐지 장치.

#### 청구항 2

제1 항에 있어서, 상기 인코더는

상기 다수의 문자 벡터 중 대응하는 문자 벡터와 이전 획득된 히든 벡터를 인가받아 문자 벡터와 히든 벡터의 특징을 추출하여 특징 벡터와 히든 벡터를 각각 출력하는 다수의 인코딩 셀을 포함하고,

상기 다수의 인코딩 셀 중 최초 인코딩 셀은 대응하는 문자 벡터와 함께 미리 설정된 더미 히든 벡터를 인가받아 특징 벡터와 히든 벡터를 출력하는 비윤리어 탐지 장치.

#### 청구항 3

제2 항에 있어서, 상기 다수의 히든 벡터와 상기 다수의 특징 벡터는 동일한 값을 갖는 비윤리어 탐지 장치.

#### 청구항 4

제2 항에 있어서, 상기 인코더와 상기 디코더 및 상기 판정부는

비속어 포함 여부만이 레이블링된 문장을 학습용 데이터로 이용하여 판별된 비윤리어 포함 여부에 대한 오차가 역전파되는 약지도 학습된 인공 신경망으로 구현되는 비윤리어 탐지 장치.

#### 청구항 5

제4 항에 있어서, 상기 인코더는

미리 학습된 LSTM(Long Short Term Memory)로 구현되고, 상기 다수의 인코딩 셀 각각은 대응하는 문자 벡터와 이전 획득된 히든 벡터 중 적어도 하나를 인가받아 문자 벡터와 히든 벡터를 각각 출력하는 LSTM 셀로 구현되는 비윤리어 탐지 장치.

#### 청구항 6

문장을 인가받아 문자 단위로 구분하고, 구분된 다수의 문자 각각을 문자 벡터로 변환하는 단계;

미리 학습된 패턴 추정 방식에 따라 다수의 문자 벡터를 순차적으로 인코딩하여 다수의 특징 벡터와 다수의 히든 벡터를 추출하되, 이전 문자 벡터를 인코딩할때 획득된 히든 벡터를 현재 문자 벡터와 함께 인코딩하는 단계;

상기 다수의 히든 벡터 중 최종 히든 벡터를 인가받아 미리 학습된 패턴 인식 방식에 따라 디코딩하여 디코딩

벡터를 획득하는 단계;

상기 다수의 특징 벡터 각각과 상기 디코딩 벡터 사이의 유사도를 비교하여 다수의 가중치를 획득하고, 획득된 다수의 가중치 각각을 대응하는 특징 벡터에 반영하여 다수의 정규화 벡터를 획득하는 단계;

상기 정규화 벡터를 미리 지정된 방식으로 결합하여 구문 벡터를 획득하는 단계; 및

미리 학습된 패턴 추정 방식에 따라 상기 구문 벡터의 특징을 추출하고, 추출된 구문 벡터의 특징을 분류하여 비윤리어 포함 여부를 판별하는 단계; 를 포함하는 비윤리어 탐지 방법.

## 청구항 7

제6 항에 있어서, 상기 인코딩하는 단계는

상기 다수의 문자 벡터 중 최초 문자 벡터를 미리 설정된 더미 히든 벡터와 함께 인코딩하여 특징 벡터와 히든 벡터를 추출하는 단계; 및

이후 나머지 문자 벡터 중 대응하는 문자 벡터와 이전 획득된 히든 벡터를 함께 인코딩하여 특징 벡터와 히든 벡터를 각각 추출하는 단계; 를 포함하는 비윤리어 탐지 방법.

## 청구항 8

제6 항에 있어서, 상기 비윤리어 탐지 방법은

학습 시에, 비속어 포함 여부만이 레이블링된 문장을 학습용 데이터로 이용하여 판별된 비윤리어 포함 여부에 대한 오차가 역전파되어 약지도 학습되는 비윤리어 탐지 방법.

## 발명의 설명

### 기술 분야

[0001] 본 발명은 비윤리어 탐지 장치 및 방법에 관한 것으로, 학습 기반 비윤리어 탐지 장치 및 방법에 관한 것이다.

### 배경 기술

[0002] 현재 각종 온라인 커뮤니티, SNS, 게시판 등에서는 다수의 사용자들 간의 대화, 정보 전달, 댓글 등과 같이 다양한 방식으로 언어가 사용되고 있다.

[0003] 이러한 온라인 환경에서는 익명성으로 인해 많은 비속어가 사용되고 있으며, 또한 빠르게 전파되는 특징을 갖는다. 이에 온라인 상에서 비속어를 탐지하여 필터링하기 위한 다양한 방법이 연구되어 오고 있다.

[0004] 그러나 온라인에서는 다양한 비윤리적 언어 또한 빈번하게 사용되고 있다. 현재 비윤리적 언어를 탐지하여 필터링하기 위한 연구는 매우 빈약한 실정이며, 이러한 필터링의 취약함을 이용하여 비속어를 포함하지 않고 비윤리적 언어를 포함한 문장을 이용하여 다른 사용자에게 불쾌감을 유발하는 일이 빈번하게 발생하고 있다. 즉 비윤리적 언어는 비속어에 못지 않게 사용자들 사이의 분쟁을 유발할 뿐만 아니라 사회 물의를 일으킬 수 있다. 따라서 비윤리적 언어 또한 필터링 할 수 있도록 탐지하는 기술이 요구된다.

## 선행기술문헌

### 특허문헌

[0005] (특허문헌 0001) 한국 공개 특허 제10-2017-0131291호 (2017.11.29 공개)

## 발명의 내용

### 해결하려는 과제

[0006] 본 발명의 목적은 비윤리적 언어가 포함된 문장을 용이하게 탐지할 수 있는 비윤리어 탐지 장치 및 방법을 제공하는데 있다.

[0007] 본 발명의 다른 목적은 비윤리적 언어가 포함된 문장을 용이하게 탐지할 수 있는 비윤리어 탐지 장치 및 방법을 제공하는데 있다.

### 과제의 해결 수단

[0008] 상기 목적을 달성하기 위한 본 발명의 일 실시예에 따른 비윤리어 탐지 장치는 문장을 인가받아 문자 단위로 구분하고, 구분된 다수의 문자 각각을 문자 벡터로 변환하는 전처리부; 미리 학습된 패턴 추정 방식에 따라 다수의 문자 벡터를 순차적으로 인코딩하여 다수의 특징 벡터와 다수의 히든 벡터를 추출하되, 이전 문자 벡터를 인코딩할때 획득된 히든 벡터를 현재 문자 벡터와 함께 인코딩하는 인코더; 상기 다수의 히든 벡터 중 최종 히든 벡터를 인가받아 미리 학습된 패턴 인식 방식에 따라 디코딩하여 디코딩 벡터를 획득하는 디코더; 상기 다수의 특징 벡터 각각과 상기 디코딩 벡터 사이의 유사도를 비교하여 다수의 가중치를 획득하고, 획득된 다수의 가중치 각각을 대응하는 특징 벡터에 반영하여 다수의 정규화 벡터를 획득하는 정규화부; 상기 정규화 벡터를 미리 지정된 방식으로 결합하여 구문 벡터를 획득하는 구문 벡터 획득부; 및 미리 학습된 패턴 추정 방식에 따라 상기 구문 벡터의 특징을 추출하고, 추출된 구문 벡터의 특징을 분류하여 비윤리어 포함 여부를 판별하는 판정부; 를 포함한다.

[0009] 상기 인코더는 상기 다수의 문자 벡터 중 대응하는 문자 벡터와 이전 획득된 히든 벡터를 인가받아 문자 벡터와 히든 벡터의 특징을 추출하여 특징 벡터와 히든 벡터를 각각 출력하는 다수의 인코딩 셀을 포함하고, 상기 다수의 인코딩 셀 중 최초 인코딩 셀은 대응하는 문자 벡터와 함께 미리 설정된 더미 히든 벡터를 인가받아 특징 벡터와 히든 벡터를 출력할 수 있다.

[0010] 상기 인코더와 상기 디코더 및 상기 판정부는 비속어 포함 여부만이 레이블링된 문장을 학습용 데이터로 이용하여 판별된 비윤리어 포함 여부에 대한 오차가 역전과되는 약지도 학습될 수 있다.

[0011] 상기 인코더는 미리 학습된 LSTM(Long Short Term Memory)로 구현되고, 상기 다수의 인코딩 셀 각각은 대응하는 문자 벡터와 이전 획득된 히든 벡터 중 적어도 하나를 인가받아 문자 벡터와 히든 벡터를 각각 출력하는 LSTM 셀로 구현될 수 있다.

[0012] 상기 목적을 달성하기 위한 본 발명의 다른 실시예에 따른 비윤리어 탐지 방법은 문장을 인가받아 문자 단위로 구분하고, 구분된 다수의 문자 각각을 문자 벡터로 변환하는 단계; 미리 학습된 패턴 추정 방식에 따라 다수의 문자 벡터를 순차적으로 인코딩하여 다수의 특징 벡터와 다수의 히든 벡터를 추출하되, 이전 문자 벡터를 인코딩할때 획득된 히든 벡터를 현재 문자 벡터와 함께 인코딩하는 단계; 상기 다수의 히든 벡터 중 최종 히든 벡터를 인가받아 미리 학습된 패턴 인식 방식에 따라 디코딩하여 디코딩 벡터를 획득하는 단계; 상기 다수의 특징 벡터 각각과 상기 디코딩 벡터 사이의 유사도를 비교하여 다수의 가중치를 획득하고, 획득된 다수의 가중치 각각을 대응하는 특징 벡터에 반영하여 다수의 정규화 벡터를 획득하는 단계; 상기 정규화 벡터를 미리 지정된 방식으로 결합하여 구문 벡터를 획득하는 단계; 및 미리 학습된 패턴 추정 방식에 따라 상기 구문 벡터의 특징을 추출하고, 추출된 구문 벡터의 특징을 분류하여 비윤리어 포함 여부를 판별하는 단계; 를 포함한다.

### 발명의 효과

[0013] 따라서, 본 발명의 실시예에 따른 비윤리어 탐지 장치 및 방법은 문자 단위로 벡터화된 문장에서 학습을 기반으로 비윤리어를 탐색함으로써, 띄어쓰기와 맞춤법 등이 제대로 지켜지지 않은 문장에서도 정확하게 비윤리어를 탐지할 수 있다. 또한 기존에 연구된 비속어 탐지 기법과 결합되어 비속어와 비윤리어를 용이하게 탐지 및 필터링하여 건전한 온라인 환경을 조성할 수 있다.

### 도면의 간단한 설명

[0014] 도 1은 본 발명의 일 실시예에 따른 비윤리어 탐지 장치의 개략적 구조를 나타낸다.

도 2는 도 1의 비윤리어 탐지 장치의 상세 구조를 나타낸다.

도 3은 본 발명의 일 실시예에 따른 비윤리어 탐지 방법을 나타낸다.

### 발명을 실시하기 위한 구체적인 내용

[0015] 본 발명과 본 발명의 동작상의 이점 및 본 발명의 실시예에 의하여 달성되는 목적을 충분히 이해하기 위해서는 본 발명의 바람직한 실시예를 예시하는 첨부 도면 및 첨부 도면에 기재된 내용을 참조하여야만 한다.

- [0016] 이하, 첨부한 도면을 참조하여 본 발명의 바람직한 실시예를 설명함으로써, 본 발명을 상세히 설명한다. 그러나, 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며, 설명하는 실시예에 한정되는 것이 아니다. 그리고, 본 발명을 명확하게 설명하기 위하여 설명과 관계없는 부분은 생략되며, 도면의 동일한 참조부호는 동일한 부재임을 나타낸다.
- [0017] 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. 또한, 명세서에 기재된 "...부", "...기", "모듈", "블록" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.
- [0018] 도 1은 본 발명의 일 실시예에 따른 비윤리어 탐지 장치의 개략적 구조를 나타내고, 도 2는 도 1의 비윤리어 탐지 장치의 상세 구조를 나타낸다.
- [0019] 도 1 및 도 2를 참조하면, 본 실시예에 따른 비윤리어 탐지 장치는 전처리부(110), 인코더(120), 디코더(130), 정규화부(140), 구문 벡터 획득부(150) 및 판정부(160)를 포함한다.
- [0020] 전처리부(110)는 비윤리어를 탐지할 문장을 인가받아 문자 단위로 구분하고, 구분된 문자를 벡터화하여 다수의 문자 벡터( $V_1, V_2, \dots, V_{i-1}, V_i$ )를 출력한다. 문장을 문자 단위로 구분하여 벡터화하면, 단어 단위의 의미를 분석하기에 부적합하지만 문장의 전체 의미를 파악하기에는 더욱 효율적이다. 특히 온라인 상에서는 띄어쓰기나 맞춤법이 제대로 지켜지지 않기 때문에 기존에 문장 분석 시에 일반적으로 이용되는 word2vec 기법 등을 이용하여 문장을 단어 단위로 구분하는 경우, 비윤리어를 판별하기가 매우 어렵다. 이에 본 실시예에서는 전처리부(110)가 문장을 문자 단위로 구분하여 다수의 문자 벡터( $V_1, V_2, \dots, V_{i-1}, V_i$ )를 획득한다.
- [0021] 여기서 전처리부(110)는 일예로 미리 학습된 인공 신경망으로 구현될 수 있으나, 인공 신경망이 아닌 방식으로 구현되어도 무방하다. 그리고 전처리부(110)는 일예로 다수의 문자 벡터( $V_1, V_2, \dots, V_{i-1}, V_i$ ) 각각을 기지정된 길이(여기서는 일예로  $n$ )를 갖는 1차원 벡터로 획득할 수 있다.
- [0022] 인코더(120)는 미리 학습된 인공 신경망으로 구현되어, 다수의 문자 벡터( $V_1, V_2, \dots, V_{i-1}, V_i$ )를 학습된 방식으로 순차적으로 인코딩하여 다수의 특징 벡터( $f_1, f_2, \dots, f_{i-1}, f_i$ )와 히든 벡터( $h_1, h_2, \dots, h_{i-1}, h_i$ )를 출력한다.
- [0023] 인코더(120)는 학습된 패턴 추정 방식에 따라 다수의 문자 벡터( $V_1, V_2, \dots, V_{i-1}, V_i$ ) 각각을 인코딩하여 다수의 문자 벡터( $V_1, V_2, \dots, V_{i-1}, V_i$ )의 특징을 추출함으로써, 다수의 특징 벡터( $f_1, f_2, \dots, f_{i-1}, f_i$ )를 획득한다.
- [0024] 이때 인코더(120)는 단순히 다수의 문자 벡터( $V_1, V_2, \dots, V_{i-1}, V_i$ ) 각각을 개별적으로 인코딩하지 않고, 이전 문자 벡터( $V_1, V_2, \dots, V_{i-1}, V_i$ )에 대한 특징 벡터( $f_1, f_2, \dots, f_{i-1}$ )를 획득할 때 함께 획득되는 히든 벡터( $h_1, h_2, \dots, h_{i-1}$ )를 인가받아 다음 문자 벡터를 인코딩할 때 함께 반영하여 인코딩을 수행함으로써, 이전 문자 벡터( $V_1, V_2, \dots, V_{i-1}$ )들의 특징이 다음 특징 벡터( $f_1, f_2, \dots, f_{i-1}, f_i$ )를 획득할 때 반영될 수 있도록 한다.
- [0025] 본 실시예에서 인코더(120)는 전처리부(110)에서 문장을 문자 단위로 구분하여 문자 벡터( $V_1, V_2, \dots, V_{i-1}, V_i$ )로 변환함에 따라 문장 특징을 추출하기 어렵다는 단점을 보완하기 위해, 이전 획득된 히든 벡터( $h_1, h_2, \dots, h_{i-1}$ )를 함께 반영하여 문장 특징이 이후 추출되는 특징 벡터( $f_1, f_2, \dots, f_{i-1}, f_i$ )에 포함될 수 있도록 한다.
- [0026] 이에 인코더(120)는 도 2에 도시된 바와 같이 다수의 인코딩 셀( $EN_1, EN_2, \dots, EN_{i-1}, EN_i$ )을 포함하여 구성될 수 있으며, 다수의 인코딩 셀( $EN_1, EN_2, \dots, EN_{i-1}, EN_i$ ) 각각은 다수의 문자 벡터( $V_1, V_2, \dots, V_{i-1}, V_i$ ) 중 대응하는 문자 벡터와 이전단의 인코딩 셀( $EN_1, EN_2, \dots, EN_{i-1}, EN_i$ )에서 출력되는 다수의 히든 벡터( $h_1, h_2, \dots, h_{i-1}$ )를 인가받아 인코딩하여 특징 벡터( $f_1, f_2, \dots, f_{i-1}, f_i$ )를 출력할 수 있다.
- [0027] 여기서 히든 벡터( $h_1, h_2, \dots, h_{i-1}, h_i$ )는 이후 인코딩되는 문자 벡터( $V_1, V_2, \dots, V_{i-1}, V_i$ )로부터 특징 벡터( $f_1, f_2, \dots, f_{i-1}, f_i$ )를 추출할 때 추가적으로 반영하고자 하는 이전 문자 벡터( $V_1, V_2, \dots, V_{i-1}$ )들의 특징이

다.

- [0028] 다만 다수의 인코딩 셀( $EN_1, EN_2, \dots, EN_{i-1}, EN_i$ ) 중 맨 앞단의 인코딩 셀( $EN_1$ )의 경우, 인가받을 수 있는 히든 벡터가 존재하지 않는다. 이에 최초 문자 벡터( $V_1$ )를 인코딩하기 위한 맨 앞단의 인코딩 셀( $EN_1$ )은 미리 설정된 더미 히든 벡터( $h_0$ )를 인가받도록 미리 설정될 수 있다.
- [0029] 그리고 인코더(120)에서 획득되는 다수의 특징 벡터( $f_1, f_2, \dots, f_{i-1}, f_i$ )는 다수의 문자 벡터( $V_1, V_2, \dots, V_{i-1}, V_i$ )와 동일한 크기 및 차원을 갖는 벡터로 획득될 수 있으나, 이에 한정되지 않는다. 즉 다수의 특징 벡터( $f_1, f_2, \dots, f_{i-1}, f_i$ )는 다수의 문자 벡터( $V_1, V_2, \dots, V_{i-1}, V_i$ )와 다른 크기 또는 다른 차원을 갖는 벡터로 획득될 수 있다. 그러나 다수의 특징 벡터( $f_1, f_2, \dots, f_{i-1}, f_i$ )와 다수의 히든 벡터( $h_1, h_2, \dots, h_{i-1}, h_i$ )는 동일한 크기 및 차원을 갖도록 획득된다.
- [0030] 그리고 다수의 히든 벡터( $h_1, h_2, \dots, h_{i-1}, h_i$ )는 다수의 특징 벡터( $f_1, f_2, \dots, f_{i-1}, f_i$ )와 상이하게 추출될 수 있으나, 특징 벡터( $f_1, f_2, \dots, f_{i-1}, f_i$ )와 동일하게 추출될 수도 있다. 즉 특징 벡터( $f_1, f_2, \dots, f_{i-1}$ )가 히든 벡터( $h_1, h_2, \dots, h_{i-1}$ )로서, 이후 문자 벡터( $V_1, V_2, \dots, V_{i-1}, V_i$ )를 인코딩할 때 함께 반영되도록 구성될 수도 있다. 여기서는 일례로 특징 벡터( $f_1, f_2, \dots, f_{i-1}, f_i$ )가 히든 벡터( $h_1, h_2, \dots, h_{i-1}, h_i$ )가 동일한 것으로 가정하였다.
- [0031] 한편 본 실시예에서는 인코더(120)가 일례로 미리 학습된 LSTM(Long Short Term Memory)로 구현될 수 있다. LSTM은 인공 신경망의 하나로서, 순환 신경망(Recurrent Neural Network: RNN)이 장기간(Long Term) 특징을 반영할 수 있도록 개선한 구조를 갖는다. 인코더(120)가 LSTM으로 구현되는 경우, 인코더(120)는 직전 획득된 히든 벡터( $h_1, h_2, \dots, h_{i-1}$ ) 뿐만 아니라, 그 이전 획득된 적어도 하나의 히든 벡터( $h_1, h_2, \dots, h_{i-1}$ )를 함께 인가받아 문자 벡터( $V_1, V_2, \dots, V_{i-1}, V_i$ )와 함께 인코딩할 수 있다.
- [0032] 인코더(120)가 LSTM으로 구현되는 경우, 다수의 인코딩 셀( $EN_1, EN_2, \dots, EN_{i-1}, EN_i$ ) 각각은 LSTM 셀로 구현될 수 있다.
- [0033] 디코더(130)는 인코더(120)에서 획득된 다수의 히든 벡터( $h_1, h_2, \dots, h_{i-1}, h_i$ ) 중 최종 히든 벡터( $h_i$ )를 인가받고, 기지정된 방식으로 최종 히든 벡터( $h_i$ )를 디코딩하여 디코딩 벡터( $d_i$ )를 출력한다.
- [0034] 상기한 바와 같이 인코더(120)는 이전 문자 벡터( $V_1, V_2, \dots, V_{i-1}$ )를 인코딩하여 특징 벡터( $f_1, f_2, \dots, f_{i-1}, f_i$ )를 획득함과 동시에 히든 벡터( $h_1, h_2, \dots, h_{i-1}, h_i$ )를 획득하여 다음 문자 벡터( $V_2, V_3, \dots, V_{i-1}, V_i$ )를 인코딩할 때 획득된 히든 벡터( $h_1, h_2, \dots, h_{i-1}$ )를 함께 반영한다. 따라서 최종 히든 벡터( $h_i$ )에는 이전 문자 벡터( $V_1, V_2, \dots, V_{i-1}$ )들의 특징이 누적되어 포함된 것으로 볼 수 있다.
- [0035] 이에 디코더(130)는 최종 히든 벡터( $h_i$ )를 인가받아 디코딩함으로써, 다수의 문자 벡터( $V_1, V_2, \dots, V_{i-1}, V_i$ ) 전체의 특징을 디코딩 벡터( $d_i$ )로 획득할 수 있다.
- [0036] 여기서 디코더(130) 또한 미리 학습된 인공 신경망으로 구현될 수 있으며, 인코더(120)와 동일한 인공 신경망으로 구현될 수 있으나, 인코더(120)와 상이한 인공 신경망(예를 들면, 컨볼루션 신경망(CNN))으로 구현되어도 무방하다. 다만 디코더(130)가 인코더(120)와 동일한 인공 신경망으로 구현되고, 인코더(120)가 상기한 바와 같이 LSTM으로 구현되는 경우, 디코더(130)는 기지정된 더미 입력인 더미 최종 히든 벡터(미도시)를 더 인가받을 수 있다.
- [0037] 또한 디코딩 벡터( $d_i$ )의 크기 및 차원은 다수의 특징 벡터( $f_1, f_2, \dots, f_{i-1}, f_i$ )의 크기 및 차원과 동일하게 획득될 수 있다.
- [0038] 정규화부(140)는 인코더(120)에서 획득된 다수의 특징 벡터( $f_1, f_2, \dots, f_{i-1}, f_i$ ) 각각과 디코더(130)에서 획득된 디코딩 벡터( $d_i$ )를 유사도 비교하여 다수의 특징 벡터( $f_1, f_2, \dots, f_{i-1}, f_i$ ) 각각에 대한 가중치( $w_1, w_2, \dots, w_{i-1}, w_i$ )를 획득한다. 그리고 획득된 가중치( $w_1, w_2, \dots, w_{i-1}, w_i$ ) 각각을 다수의 특징 벡터( $f_1, f_2, \dots,$



$f_{i-1}, f_i$ )에 적용하여 다수의 특징 벡터( $f_1, f_2, \dots, f_{i-1}, f_i$ )를 정규화하여 정규화 벡터( $c_1, c_2, \dots, c_{i-1}, c_i$ )를 획득한다.

[0039] 정규화부(140)는 다수의 비교기( $CM_1, CM_2, \dots, CM_{i-1}, CM_i$ )를 포함할 수 있으며, 다수의 비교기( $CM_1, CM_2, \dots, CM_{i-1}, CM_i$ ) 각각은 다수의 특징 벡터( $f_1, f_2, \dots, f_{i-1}, f_i$ ) 중 대응하는 특징 벡터와 디코딩 벡터( $d_i$ ) 사이의 유사도 비교하여 가중치( $w_1, w_2, \dots, w_{i-1}, w_i$ )를 획득하고, 획득된 가중치를 다수의 특징 벡터( $f_1, f_2, \dots, f_{i-1}, f_i$ )에 적용하여, 정규화 벡터( $c_1, c_2, \dots, c_{i-1}, c_i$ )를 출력한다.

[0040] 여기서 다수의 가중치( $w_1, w_2, \dots, w_{i-1}, w_i$ )는 일예로 각각 0 ~ 1 사이의 값을 갖고, 가중치( $w_1, w_2, \dots, w_{i-1}, w_i$ )의 합이 1이 되도록 획득될 수 있다.

[0041] 정규화부(140) 또한 미리 학습된 인공 신경망으로 구현될 수 있으나, 이에 한정되지 않는다.

[0042] 구문 벡터 획득부(150)는 다수의 정규화 벡터( $c_1, c_2, \dots, c_{i-1}, c_i$ )를 기지정된 방식으로 결합(concatenate)하여 구문 벡터(context vector)(cv)를 획득한다. 일예로 구문 벡터 획득부(150)는 다수의 정규화 벡터( $c_1, c_2, \dots, c_{i-1}, c_i$ )를 단순 연결하여 구문 벡터(cv)를 획득할 수도 있다.

[0043] 판정부(160)는 패턴 추정 방식이 미리 학습된 인공 신경망으로 구현될 수 있으며, 구문 벡터 획득부(150)에서 획득된 구문 벡터(cv)와 디코더(130)에서 획득된 디코딩 벡터( $d_i$ )를 인가받아 결합하고, 결합된 벡터에 대해 미리 학습된 패턴 추정 방식에 따라 구문 특징을 추출하고 추출된 구문 특징을 분류하여 비윤리어가 포함되었는지 여부를 판별한다.

[0044] 여기서 판정부(160)는 일예로 컨볼루션 신경망(CNN) 등으로 구현될 수 있으며, 적어도 하나의 완전 연결 레이어(fully connected layer)가 포함된 인경 신경망으로 구현되어 구문 특징을 비윤리어 포함 클래스와 비윤리어 비포함 클래스로 분류할 수 있다.

[0045] 상기한 비윤리어 탐지 장치는 온라인 상에서 수집된 다수의 한글 문장 데이터에 비윤리어의 포함 여부만이 간략하게 레이블링된 학습 데이터를 이용하여, 약지도 학습(weakly-supervised learning) 방식으로 학습될 수 있다. 즉 비윤리어 탐지 장치가 학습 데이터에 대해 분류한 결과와 학습 데이터에 레이블링된 결과를 비교하여, 오차를 역전파함으로써 학습될 수 있다.

[0046] 이와 같이 약지도 학습 방식으로 학습을 수행하는 경우, 비윤리어 각각에 대한 레이블이 필요하지 않으므로, 용이하게 대량의 학습 데이터를 획득할 수 있다. 뿐만 아니라, 비윤리어를 개별적으로 지정하여 학습시키지 않고 비윤리어 탐지 장치가 학습 데이터로부터 스스로 비윤리어를 식별하도록 학습되므로, 사람의 주관적 견해가 급적 배제된 다양한 형태의 비윤리어를 탐지할 수 있다.

[0047] 도 3은 본 발명의 일 실시예에 따른 비윤리어 탐지 방법을 나타낸다.

[0048] 도 1 및 도 2를 참조하여 도 3의 비윤리어 탐지 방법을 설명하면, 우선 비윤리어가 포함되었는지 여부를 판정하기 위한 탐색 문장을 획득한다(S11). 그리고 획득된 문장을 문자 단위로 구분하고, 기지정된 방식으로 구분된 다수의 문자들 각각을 문자 벡터( $V_1, V_2, \dots, V_{i-1}$ )로 변환한다(S12).

[0049] 그리고 미리 학습된 패턴 추정 방식에 따라 다수의 문자 벡터( $V_1, V_2, \dots, V_{i-1}$ ) 각각을 순차적으로 인코딩하여 특징 벡터( $f_1, f_2, \dots, f_{i-1}, f_i$ )를 추출한다(S13). 이때, 각 문자 벡터( $V_1, V_2, \dots, V_{i-1}$ )는 이전 문자 벡터를 인코딩할 때 특징 벡터( $f_1, f_2, \dots, f_{i-1}, f_i$ )와 함께 획득되는 히든 벡터( $h_1, h_2, \dots, h_{i-1}, h_i$ )와 함께 인코딩되어, 특징 벡터와 함께 히든 벡터( $h_1, h_2, \dots, h_{i-1}, h_i$ )를 획득하고, 획득된 히든 벡터( $h_1, h_2, \dots, h_{i-1}, h_i$ )를 다시 다음 문자 벡터의 인코딩할 반영한다.

[0050] 즉 히든 벡터( $h_1, h_2, \dots, h_{i-1}, h_i$ )는 문장에서 다수의 문자 벡터의 인코딩 시에 획득되는 특징들이 누적되어 다음 특징 벡터( $f_1, f_2, \dots, f_{i-1}, f_i$ )를 추출할 때 반영되도록 한다. 여기서 히든 벡터( $h_1, h_2, \dots, h_{i-1}, h_i$ )는 특징 벡터( $f_1, f_2, \dots, f_{i-1}, f_i$ )와 동일하게 획득될 수도 있다. 여기서 문장의 최초 문자에 대해서는 미리 설정된 더미 히든 벡터( $h_0$ )를 이용하여 특징 벡터를 추출할 수 있다.



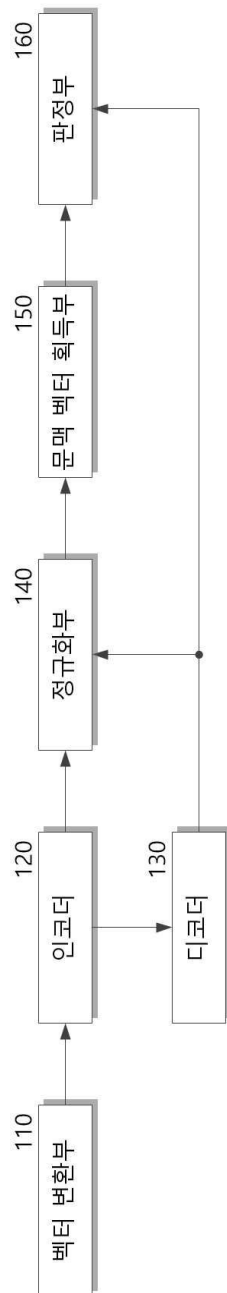
- [0051] 그리고 문장의 모든 문자에 대해 특징 벡터( $f_1, f_2, \dots, f_{i-1}, f_i$ )와 히든 벡터( $h_1, h_2, \dots, h_{i-1}, h_i$ )가 추출되면, 다수의 히든 벡터( $h_1, h_2, \dots, h_{i-1}, h_i$ ) 중 최종 히든 벡터( $h_i$ )를 미리 학습된 방식으로 디코딩하여 디코딩 벡터( $d_i$ )를 획득한다(S14).
- [0052] 디코딩 벡터( $d_i$ )가 획득되면, 다수의 특징 벡터( $f_1, f_2, \dots, f_{i-1}, f_i$ ) 각각과 디코딩 벡터( $d_i$ ) 사이의 유사도 비교를 통해 가중치( $w_1, w_2, \dots, w_{i-1}, w_i$ )를 획득하고, 획득된 가중치를 다수의 특징 벡터( $f_1, f_2, \dots, f_{i-1}, f_i$ )에 적용하여, 특징 벡터( $f_1, f_2, \dots, f_{i-1}, f_i$ )를 정규화 벡터( $c_1, c_2, \dots, c_{i-1}, c_i$ )로 정규화한다(S15).
- [0053] 그리고 다수의 정규화 벡터( $c_1, c_2, \dots, c_{i-1}, c_i$ )를 기지정된 방식으로 결합하여 구문 벡터( $cv$ )를 획득한다(S16). 구문 벡터( $cv$ )가 획득되면 미리 학습된 패턴 추정 방식에 따라 구문 벡터( $cv$ )에서 구문 특징을 추출하고(S17), 추출된 구문 특징을 분류하여 비윤리어가 포함되었는지 여부를 판별한다(S18).
- [0054] 결과적으로 본 실시예에 따른 비윤리어 탐지 방법은 문장을 문자 단위로 구분함으로써, 띄어쓰기나 맞춤법과 같은 문법상의 오류에도 강인하게 비윤리어를 탐지할 수 있도록 하며, 문자별 특징 추출 시에 이전 특징이 누적 반영되도록 하여, 문자 단위로 구분됨으로 인해 문장의 의미 분석이 어려움에도 정확하게 비윤리어의 포함 여부를 판별할 수 있다.
- [0055] 본 발명에 따른 방법은 컴퓨터에서 실행 시키기 위한 매체에 저장된 컴퓨터 프로그램으로 구현될 수 있다. 여기서 컴퓨터 판독가능 매체는 컴퓨터에 의해 액세스 될 수 있는 임의의 가용 매체일 수 있고, 또한 컴퓨터 저장 매체를 모두 포함할 수 있다. 컴퓨터 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현된 휘발성 및 비휘발성, 분리형 및 비분리형 매체를 모두 포함하며, ROM(판독 전용 메모리), RAM(랜덤 액세스 메모리), CD(컴팩트 디스크)-ROM, DVD(디지털 비디오 디스크)-ROM, 자기 테이프, 플로피 디스크, 광데이터 저장장치 등을 포함할 수 있다.
- [0056] 본 발명은 도면에 도시된 실시예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 것이다.
- [0057] 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 청구범위의 기술적 사상에 의해 정해져야 할 것이다.

### 부호의 설명

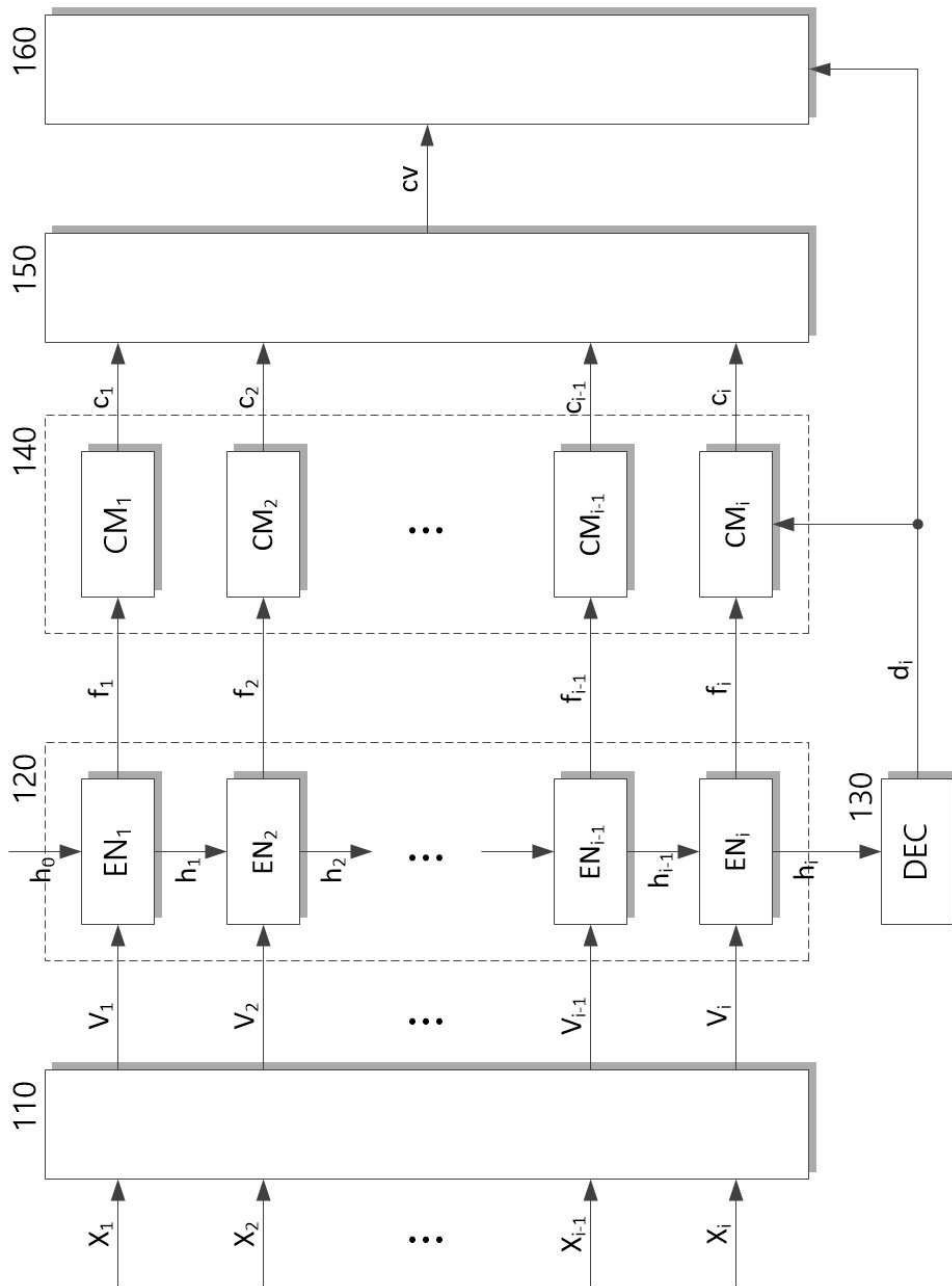
- [0058] 110: 전처리부      120: 인코더  
130: 디코더      140: 정규화부  
150: 구문 벡터 획득부      160: 판정부

도면

도면1



도면2



도면3

