(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0072226
(43) 공개일자 2022년06월02일

(51) 국제특허분류(Int. Cl.)
G06N 3/08 (2006.01) *G06N 3/04* (2006.01)
G06N 5/02 (2006.01) *G06N 7/00* (2022.01)
(52) CPC특허분류
G06N 3/08 (2013.01)
G06N 3/04 (2013.01)
(21) 출원번호 10-2020-0159601
(22) 출원일자 2020년11월25일
심사청구일자 2020년11월25일

(71) 출원인
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
(72) 발명자
한요섭
서울특별시 은평구 진관1로 77-8, 403동 204호 (진관동,은평뉴타운폭포동아파트)
이주형
서울특별시 서초구 신반포로 45, 70동 306호(반포동, 반포아파트)
(뒷면에 계속)
(74) 대리인
민영준

전체 청구항 수 : 총 4 항

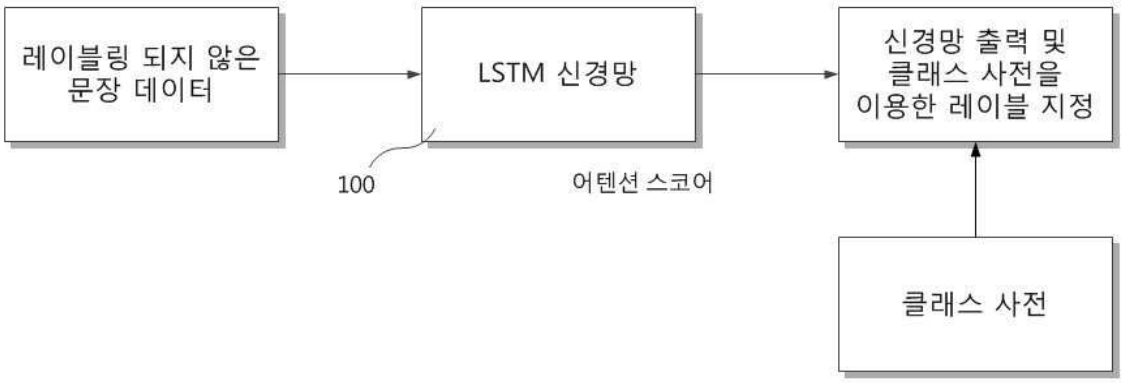
(54) 발명의 명칭 **문장 데이터 클래스 분류 및 클래스 사전 생성을 위한 신경망 학습 방법**

(57) 요 약

문장 데이터 클래스 분류 및 클래스 사전 생성을 위한 신경망 학습 방법이 개시된다. 개시된 방법은, 클래스가 레이블링된 문장 데이터를 LSTM 분류 신경망에 입력하여 클래스 레이블과의 차이값을 이용한 초기 학습을 수행하는 단계(a); 상기 초기 학습이 이루어진 후 클래스가 레이블링된 문장 데이터를 LSTM 분류 신경망에 입력하여 중

(뒷면에 계속)

대 표 도 - 도7



기 학습을 수행하고, LSTM 신경망으로부터 획득되는 상기 문장을 구성하는 단어들의 어텐션 스코어에 기초하여 클래스 사전을 갱신하는 단계(b); 클래스 레이블이 없는 문장 데이터를 LSTM 분류 신경망에 입력하여 출력되는 클래스 확률과 상기 클래스 레이블이 없는 문장을 구성하는 단어들 중 상기 생성된 클래스 사전에 있는 단어들과 매칭되는 매칭 단어의 수를 카운팅한 후 상기 클래스 확률 및 상기 매칭 단어의 수에 기초하여 상기 클래스 레이블이 없는 문장 데이터에 대해 클래스 레이블을 부여하는 단계(c); 및 상기 클래스 레이블이 부여된 문장 데이터를 이용하여 상기 LSTM 분류 신경망에 대한 학습을 수행하고 상기 클래스 레이블이 부여된 문장을 구성하는 단어들의 어텐션 스코어에 기초하여 클래스 사전을 갱신하는 단계(d)를 포함한다. 개시된 방법은 레이블이 부여되지 않은 문장 데이터를 학습에 사용할 수 있어 신경망의 성능을 향상시킬 수 있는 장점이 있다.

(52) CPC특허분류
G06N 5/025 (2019.01)
G06N 7/005 (2013.01)

(72) 발명자
천현준
서울특별시 서대문구 연세로 50, 연세대학교 제4공학관 709호(신촌동)

손재만
서울특별시 서대문구 연세로 50, 연세대학교 제4공학관 709호(신촌동)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711102904
과제번호	2018-0-00276-003
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원(한국연구재단부설)
연구사업명	정보통신방송연구개발사업
연구과제명	딥러닝 기반 악성코드 패턴 룰셋 생성 자동화 원천 기술 개발 (3/5)
기 여 율	1/1
과제수행기관명	연세대학교 산학협력단
연구기간	2020.01.01 ~ 2020.12.31

명 세 서

청구범위

청구항 1

클래스가 레이블링된 문장 데이터를 LSTM 분류 신경망에 입력하여 클래스 레이블과의 차이값을 이용한 초기 학습을 수행하는 단계(a);

상기 초기 학습이 이루어진 후 클래스가 레이블링된 문장 데이터를 LSTM 분류 신경망에 입력하여 중기 학습을 수행하고, LSTM 신경망으로부터 획득되는 상기 문장을 구성하는 단어들의 어텐션 스코어에 기초하여 클래스 사전을 갱신하는 단계(b);

클래스 레이블이 없는 문장 데이터를 LSTM 분류 신경망에 입력하여 출력되는 클래스 확률과 상기 클래스 레이블이 없는 문장을 구성하는 단어들 중 상기 생성된 클래스 사전에 있는 단어들과 매칭되는 매칭 단어의 수를 카운팅한 후 상기 클래스 확률 및 상기 매칭 단어의 수에 기초하여 상기 클래스 레이블이 없는 문장 데이터에 대해 클래스 레이블을 부여하는 단계(c); 및

상기 클래스 레이블이 부여된 문장 데이터를 이용하여 상기 LSTM 분류 신경망에 대한 학습을 수행하고 상기 클래스 레이블이 부여된 문장을 구성하는 단어들의 어텐션 스코어에 기초하여 클래스 사전을 갱신하는 단계(d)를 포함하는 것을 특징으로 하는 클래스 분류 및 클래스 사전 생성을 위한 신경망 학습 방법.

청구항 2

제1항에 있어서,

상기 단계(b)는 상기 LSTM 신경망의 출력값에 대한 신뢰도가 미리 설정된 제1 경계값 이상인 문장에 대해서만 문장을 구성하는 단어들 중 어텐션 스코어가 미리 설정된 제2 경계값 이상인 단어들을 상기 클래스 사전에 추가하는 것을 특징으로 하는 클래스 분류 및 클래스 사전 생성을 위한 신경망 학습 방법

청구항 3

제1항에 있어서,

상기 단계(c)는 상기 클래스 확률이 미리 설정된 제3 경계값 이하이고, 상기 매칭 단어의 수가 미리 설정된 제4 경계값 이상일 경우 상기 클래스 레이블이 없는 문장 데이터에 대해 레이블을 부여하는 것을 특징으로 하는 클래스 분류 및 클래스 사전 생성을 위한 신경망 학습 방법.

청구항 4

제3항에 있어서,

상기 단계(c)는 상기 클래스 확률이 미리 설정된 제3 경계값 이상이고, 상기 매칭 단어의 수가 미리 설정된 제5 경계값 이상일 경우 상기 클래스 레이블이 없는 문장 데이터에 대해 레이블을 부여하고, 상기 제5 경계값은 상기 제4 경계값에 비해 작게 설정되는 것을 특징으로 하는 클래스 분류 및 클래스 사전 생성을 위한 신경망 학습 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 신경망 학습 방법에 관한 것으로서, 더욱 상세하게는 텍스트 데이터 분류 및 클래스 사전 생성을 위한 신경망 학습 방법에 관한 것입니다.

배 경 기 술

[0003] 특정 문장에 대한 클래스 분류는 주로 텍스트 마이닝을 통해 이루어졌다. 문장을 구성하는 단어들의 의미와 단어들간의 관계를 분석해서 문장의 클래스를 분류하는 것이다.

[0004] 그러나, 텍스트 마이닝만으로 문장의 클래스(예를 들어, 문장이 비윤리 문장인지 또는 윤리 문장인지 여부)를 분류 정확도를 확보하는 데에는 한계가 있었으며, 근래에는 신경망을 이용하여 문장의 클래스를 분류하는 시도가 이루어졌다.

[0005] 신경망을 이용한 문장 클래스 분류는 텍스트 마이닝에 비해 높은 정확도를 확보할 수는 있으나 신경망의 학습에 레이블이 부여된 대량의 학습 데이터가 요구된다는 문제점이 있었다. 신경망이 정확성이 담보되려면 충분한 학습이 이루어져야 하나, 학습을 위한 레이블이 부여된 문장 데이터는 많지 않기에 신경망의 성능을 극대화시키기 용이하지 않은 문제점이 있었다.

발명의 내용

해결하려는 과제

[0008] 본 발명은 레이블이 부여되지 않은 문장 데이터를 학습에 사용할 수 있는 클래스 분류 및 클래스 사전 생성을 위한 신경망 학습 방법을 제안한다.

과제의 해결 수단

[0010] 상기 목적을 달성하기 위해 본 발명의 일 측면에 따르면, 클래스가 레이블링된 문장 데이터를 LSTM 분류 신경망에 입력하여 클래스 레이블과의 차이값을 이용한 초기 학습을 수행하는 단계(a); 상기 초기 학습이 이루어진 후 클래스가 레이블링된 문장 데이터를 LSTM 분류 신경망에 입력하여 중기 학습을 수행하고, LSTM 신경망으로부터 획득되는 상기 문장을 구성하는 단어들의 어텐션 스코어에 기초하여 클래스 사전을 갱신하는 단계(b); 클래스 레이블이 없는 문장 데이터를 LSTM 분류 신경망에 입력하여 출력되는 클래스 확률과 상기 클래스 레이블이 없는 문장을 구성하는 단어들 중 상기 생성된 클래스 사전에 있는 단어들과 매칭되는 매칭 단어의 수를 카운팅한 후 상기 클래스 확률 및 상기 매칭 단어의 수에 기초하여 상기 클래스 레이블이 없는 문장 데이터에 대해 클래스 레이블을 부여하는 단계(c); 및 상기 클래스 레이블이 부여된 문장 데이터를 이용하여 상기 LSTM 분류 신경망에 대한 학습을 수행하고 상기 클래스 레이블이 부여된 문장을 구성하는 단어들의 어텐션 스코어에 기초하여 클래스 사전을 갱신하는 단계(d)를 포함하는 클래스 분류 및 클래스 사전 생성을 위한 신경망 학습 방법이 제공된다.

[0011] 상기 단계(b)는 상기 LSTM 신경망의 출력값에 대한 신뢰도가 미리 설정된 제1 경계값 이상인 문장에 대해서만 문장을 구성하는 단어들 중 어텐션 스코어가 미리 설정된 제2 경계값 이상인 단어들을 상기 클래스 사전에 추가한다.

[0012] 상기 단계(c)는 상기 클래스 확률이 미리 설정된 제3 경계값 이하이고, 상기 매칭 단어의 수가 미리 설정된 제4 경계값 이상일 경우 상기 클래스 레이블이 없는 문장 데이터에 대해 레이블을 부여한다.

[0013] 상기 단계(c)는 상기 클래스 확률이 미리 설정된 제3 경계값 이상이고, 상기 매칭 단어의 수가 미리 설정된 제5 경계값 이상일 경우 상기 클래스 레이블이 없는 문장 데이터에 대해 레이블을 부여하고, 상기 제5 경계값은 상기 제4 경계값에 비해 작게 설정된다.

발명의 효과

[0015] 본 발명의 클래스 분류 및 클래스 사전 생성을 위한 신경망 학습 방법은 레이블이 부여되지 않은 문장 데이터를

학습에 사용할 수 있어 신경망의 성능을 향상시킬 수 있는 장점이 있다.

도면의 간단한 설명

[0017] 도 1은 레이블링된 데이터를 이용하여 초기 학습을 수행하는 구조를 나타낸 도면.

도 2는 레이블링된 데이터를 이용하여 사전을 생성하는 중기 학습을 수행하는 구조를 나타낸 도면.

도 3은 레이블링된 데이터의 신경망 출력값에 대한 신뢰도의 일례를 나타낸 도면.

도 4는 본 발명의 일 실시예에 따른 사전 생성 방법을 나타낸 순서도.

도 5는 본 발명의 일 실시예에 따른 LSTM 분류 신경망으로부터 출력되는 어텐션 스코어의 일례를 나타낸 도면.

도 6은 본 발명의 일 실시예에 따라 생성되는 사전의 일례를 나타낸 도면.

도 7은 레이블링이 없는 없는 데이터에 대해 레이블을 부여하고 사전을 생성하는 구조를 나타낸 도면.

도 8은 레이블링이 없는 데이터에 대해 레이블을 부여하는 방법을 나타낸 순서도.

도 9는 본 발명의 일 실시예에 따른 레이블 부여 방법을 설명하기 위한 예시 도면.

발명을 실시하기 위한 구체적인 내용

[0018] 본 발명과 본 발명의 동작상의 이점 및 본 발명의 실시에 의하여 달성되는 목적을 충분히 이해하기 위해서는 본 발명의 바람직한 실시예를 예시하는 첨부 도면 및 첨부 도면에 기재된 내용을 참조하여야만 한다.

[0019] 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.

[0020] 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.

[0021] 본 발명은 문장의 클래스 분류 방법에 관한 것으로서, 문장의 클래스 분류는 다양한 용도로 활용될 수 있다. 예를 들어, 기사의 댓글, 게임 내 주고받는 메시지 등에서 특정 문장이 비윤리 문장인지 또는 윤리 문장인지 여부를 판단하여 필터링 여부를 판단하는데 사용될 수 있다. 또한, 특정 문장이 정치, 경제, 가회 중 어느 분야와 관련된 문장인지 판단하여 사용자가 요구하는 정보를 제공하는데에도 활용될 수 있을 것이다.

[0022] 근래에 들어 이와 같은 문장의 클래스 분류는 신경망을 통해 많이 이루어지고 있으나, 클래스 분류의 높은 정확도를 담보하려면 많은 학습 데이터가 필요하나, 충분한 수의 학습 데이터를 확보하는 것은 용이하지 않다.

[0023] 여기서, 학습 데이터는 레이블링이 있는 학습 데이터를 의미하며, 예를 들어, 윤리/비윤리 클래스를 분류하는 경우, 특정 문장이 윤리 문장인지 또는 비윤리 문장인지 여부가 미리 레이블로 설정되어 있는 데이터를 의미한다.

[0024] 그러나, 이러한 레이블링이 있는 학습 데이터는 충분히 확보하기 어려우며, 이에 높은 정확도를 가지는 문장 분류 신경망을 학습시키는 것 역시 용이하지 않게 된다.

[0025] 본 발명은 이와 같은 문제를 해결하기 위해 레이블링이 없는 문장 데이터를 이용하여 학습을 수행할 수 있는 방법을 제안하며, 레이블링이 없는 문장 데이터에 레이블을 부여하기 위해 클래스별 사전을 생성하고 생성된 클래스별 사전을 이용하여 레이블을 부여하는 방법을 제안한다.

[0026] 본 발명의 전체적인 학습 방법은 초기/중기/후기로 구분될 수 있다. 초기와 중기는 레이블링이 있는 학습 데이터를 이용하여 신경망을 학습시키는 단계이다. 후기 학습 단계는 레이블링이 없는 학습 데이터를 이용하여 신경망을 학습시키는 단계이다.

[0027] 본 발명의 문장 데이터 분류는 다양한 클래스 분류에 사용될 수 있을 것이나, 설명의 편의를 위해 본 실시예에

서는 윤리 문장과 비윤리 문장을 구분하는 경우를 예로 하여 설명하기로 한다.

[0028] 도 1은 레이블링된 데이터를 이용하여 초기 학습을 수행하는 구조를 나타낸 도면이다.

[0029] 도 1을 참조하면, 레이블링이 있는 문장 데이터가 LSTM 분류 신경망(100)으로 입력된다. 초기 단계의 LSTM 분류 신경망(100)은 아직 학습이 충분히 이루어지지 않은 신경망이다. 이에, 초기 단계에서 LSTN 분류 신경망(100)이 출력하는 결과는 신뢰도가 높다고 보기 어려우며, 초기 단계는 요구되는 문장 데이터 분류를 위해 신경망이 어느 정도의 성능을 가지도록 학습하는 단계로 정의할 수 있을 것이다.

[0030] 초기 단계의 학습은 신경망의 출력과 문장 데이터에 대해 부여된 레이블을 비교하여 그 차이를 역전파하여 LSTM 분류 신경망(100)의 가중치를 변경하는 일반적인 학습이다.

[0031] 예를 들어, LSTM 분류 신경망(100)이 윤리/비윤리 문장에 대한 클래스 분류를 수행하도록 설정된 경우, LSTM 분류 신경망(100)은 문장 데이터를 입력받아 윤리 문장일 확률과 비윤리 문장일 확률을 각각 출력한다.

[0032] 비윤리로 레이블된 문장 데이터를 LSTM 분류 신경망(100)에 입력하고, LSTM 분류 신경망(100)이 윤리 확률 0.2/비윤리 확률 0.8을 출력한다고 가정하자. 이때 레이블이 윤리 확률 1/비윤리 확률 0이므로 LSTM 신경망은 이러한 손실을 역전파하여 학습을 수행한다. 요컨대, 윤리 확률이 1에 근접하고 비윤리 확률이 0에 근접하도록 LSTM 신경망의 가중치를 갱신하는 것이다.

[0033] 도 2는 레이블링된 데이터를 이용하여 사전을 생성하는 중기 학습을 수행하는 구조를 나타낸 도면이다.

[0034] 초기 학습을 통해 LSTM 분류 신경망(100)이 학습된 후 레이블링된 데이터를 이용하여 학습을 지속하면서 클래스 사전을 생성하는 과정이 함께 이루어진다.

[0035] 중기 학습에서도 레이블링된 문장 데이터가 LSTM 분류 신경망(100)에 입력되고, 레이블과 LSTM 분류 신경망(100) 출력과의 손실을 역전파하여 학습을 수행하는 구조는 동일하다.

[0036] 중기 학습 단계에서는 LSTM 분류 신경망(100)의 출력값과 레이블 값을 비교하여 신뢰도를 연산한다.

[0037] 도 3은 레이블링된 데이터의 신경망 출력값에 대한 신뢰도의 일례를 나타낸 도면이다.

[0038] 도 3을 참조하면, 데이터별 예측 클래스(윤리 또는 비윤리)와 예측값에 대한 신뢰도가 표시되어 있다.

[0039] 신뢰도는 레이블링 값과 신경망 출력값의 비에 의해 연산될 수 있을 것이다. 예를 들어, 비윤리인 경우 1 윤리인 경우 0으로 레이블 값이 정해져 있을 때, 비윤리일 확률이 0.95로 LSTM 분류 신경망에서 출력될 경우 신뢰도는 0.95로 연산된다.

[0040] 중기 학습 단계에서는 신경망 출력값에 대한 신뢰도에 기초하여 클래스 사전을 생성한다. 클래스 사전은 각 클래스와 연관이 있는 단어를 저장하고 있는 사전이다.

[0041] 도 4는 본 발명의 일 실시예에 따른 사전 생성 방법을 나타낸 순서도이다.

[0042] 도 4를 참조하면, 우선 입력된 문장을 LSTM 분류 신경망(100)에 입력하여 신경망 출력값의 신뢰도를 획득한다(400).

[0043] 획득한 신뢰도가 미리 설정된 제1 경계값 이상인지 여부를 판단한다(402).

[0044] 획득한 신뢰도가 미리 설정된 제1 경계값 이하일 경우, 입력된 문장으로부터는 사전을 생성하거나 갱신하지 않는다(단계 404).

[0045] 획득한 신뢰도가 미리 설정된 제1 경계값 이상일 경우, LSTM 분류 신경망(100)으로부터 입력된 문장을 구상하는 단어들에 대한 어텐션 스코어(attention score)를 획득한다(단계 406).

[0046] 도 5는 본 발명의 일 실시예에 따른 LSTM 분류 신경망으로부터 출력되는 어텐션 스코어의 일례를 나타낸 도면이다.

[0047] 도 5를 참조하면, LSTM 분류 신경망은 입력된 문장을 구성하는 각 단어별로 어텐션 스코어를 출력한다. 초기 학습 단계에서도 어텐션 스코어를 획득할 수 있으나, 초기 학습 단계에서 획득되는 어텐션 스코어는 사용되지 아니하며, 중기 학습 단계에서는 신뢰도가 미리 설정된 제1 경계값 이상인 경우 어텐션 스코어가 이용된다.

[0048] 도 5를 참조하면, 다른 단어에 비해 “stupid”와 “bad”는 다른 단어들에 비해 높은 어텐션 스코어를 가지는 것을 확인할 수 있다. 어텐션 스코어가 높다는 것은 해당 단어가 LSTN의 분류 신경망의 클래스 판단에 주요한

원인이 되는 단어라는 것을 의미한다.

[0049] 입력된 문장의 각 단어에 대한 어텐션 스코어가 획득되면, 미리 설정된 제2 경계값 이상의 어텐션 스코어를 가지는 단어를 클래스 사전에 추가하여 사전을 생성하거나 갱신한다(단계 408).

[0050] 도 6은 본 발명의 일 실시예에 따라 생성되는 사전의 일례를 나타낸 도면이다.

[0051] 도 6을 참조하면, 윤리 클래스 사전과 비윤리 클래스 사전이 구분되어 있다. 앞서 설명한 바와 같이, LSTM 분류 신경망에서 문장의 클래스가 비윤리로 판단되고, 입력된 문장 데이터에서 어텐션 스코어가 높은 단어가 비윤리 클래스에 추가되어 도 6과 같은 비윤리 클래스 사전이 생성된다.

[0052] 또한, LSTM 분류 신경망에서 클래스가 윤리로 판단되고, 입력된 문장 데이터에서 어텐션 스코어가 높은 단어가 윤리 클래스에 추가되어 도 6과 같은 윤리 클래스 사전이 생성된다.

[0053] 도 7은 레이블링이 없는 없는 데이터에 대해 레이블을 부여하고 사전을 생성하는 구조를 나타낸 도면이다.

[0054] 중기 학습이 완료되고 클래스 사전이 생성되면, 레이블링이 없는 문장 데이터를 이용하여 학습을 지속한다. LSTM 분류 신경망은 레이블링이 있어야만 학습이 가능한 구조이다. 이에, 본 발명은 레이블링이 없는 문장에 레이블을 부여하고 부여된 레이블을 이용하여 학습이 지속되도록 한다.

[0055] 도 8은 레이블링이 없는 데이터에 대해 레이블을 부여하는 방법을 나타낸 순서도이다.

[0056] 도 8을 참조하면, 우선 레이블링이 없는 문장 데이터를 학습된 LSTM 분류 신경망에 입력하여, LSTM 분류 신경망의 출력값을 확인한다(단계 800). LSTM 신경망은 학습된 가중치에 기반하여 클래스 판단 및 판단된 클래스 확률을 출력한다.

[0057] LSTM 분류 신경망에서 출력되는 클래스 확률이 미리 설정된 제3 경계값 이상인지 여부를 판단한다(단계 802).

[0058] LSTM 분류 신경망에서 출력되는 클래스 확률이 미리 설정된 제3 경계값 이하일 경우, LSTM 분류 신경망으로 입력된 문장의 단어 중 LSTM 분류 신경망에서 판단된 클래스 사전에 있는 단어들과 매칭되는 매칭 단어의 수를 카운팅한다(단계 804).

[0059] 클래스 확률이 미리 설정된 제3 경계값 이하이나, 입력된 문장을 구성하는 단어들 중 클래스 사전에 있는 단어들과 매칭되는 매칭 단어의 수가 미리 설정된 제4 경계값 이상일 경우, LSTM 분류 신경망에서 판단한 클래스로 입력된 문장에 대한 레이블을 부여한다(단계 806). 예를 들어, 입력된 문장에 대해 LSTM 분류 신경망에서 비윤리 클래스로 판단하고, 비윤리 클래스 확률이 제3 경계값 이하이나 매칭 단어의 수가 미리 설정된 제4 경계값 이상일 경우에는 입력된 문장의 클래스가 '비윤리'라고 레이블을 부여하는 것이다.

[0060] 그러나, 매칭 단어의 수가 미리 설정된 제4 경계값 이하일 경우에는 입력된 문장에 대해 레이블을 부여하지 않고 입력된 문장에 대한 학습을 수행하지 않는다(단계 808).

[0061] 클래스 확률이 미리 설정된 제3 경계값 이상일 경우, LSTM 분류 신경망으로 입력된 문장의 단어 중 LSTM 분류 신경망에서 판단된 클래스 사전에 있는 단어들과 매칭되는 매칭 단어의 수를 카운팅한다(단계 810)

[0062] 클래스 확률이 미리 설정된 제3 경계값 이상이고 입력된 문장을 구성하는 단어들 중 클래스 사전에 있는 단어들과 매칭되는 매칭 단어의 수가 미리 설정된 제5 경계값 이상일 경우, LSTM 분류 신경망에서 판단한 클래스로 입력된 문장에 대한 레이블을 부여한다(단계 812). 여기서, 제5 경계값은 제4 경계값에 비해 낮게 설정된다. 예를 들어, 클래스 확률이 제3 경계값 이하일 경우 레이블을 부여하기 위한 매칭 단어 수 경계값(제4 경계값)이 4라면, 클래스 확률이 제3 경계값 이사일 경우 레이블을 부여하기 위한 매칭 단어 수 경계값(제5 경계값)은 3으로 설정될 수 있다.

[0063] 클래스 확률이 미리 설정된 제3 경계값 이상이나 입력된 문장을 구성하는 단어들 중 클래스 사전에 있는 단어들과 매칭되는 매칭 단어의 수가 미리 설정된 제5 경계값 이하일 경우 입력된 문장에 대해 레이블을 부여하지 않고 입력된 문장에 대한 학습을 수행하지 않는다(단계 814).

[0064] 입력된 문장에 대해 레이블이 부여되면 입력된 문장에 대한 신경망 연산 결과에 대한 오차를 역전파하여 학습을 수행하고, 입력된 문장에서 어텐션 스코어가 제2 경계값 이상인 단어들은 클래스 사전에 추가한다(단계 816).

[0065] 도 9는 본 발명의 일 실시예에 따른 레이블 부여 방법을 설명하기 위한 예시 도면이다.

[0066] 클래스 확률에 대한 제3 경계값은 0.7, 제4 경계값은 4, 제5 경계값은 3이라고 가정한다.

[0067] 도 9에서, Data1(문장 데이터)은 클래스 확률이 0.7 이상이고 매칭 단어의 수가 3 이상이므로 Data1에 대해서는 레이블을 부여한다.

[0068] Data2는 클래스 확률이 0.7 이하이나, 매칭 단어의 수가 4이상이므로 Data2에 대해서도 레이블이 부여된다.

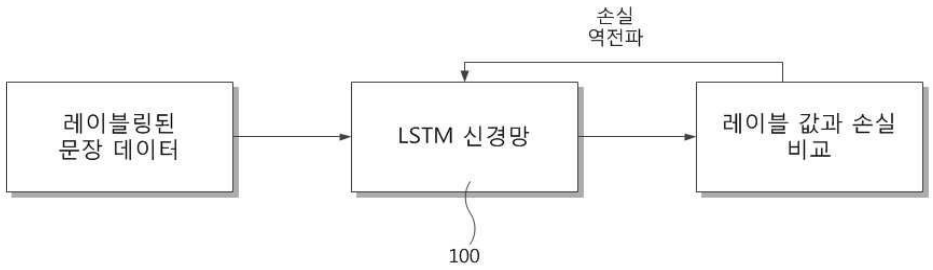
[0069] 그러나, Data3는 클래스 확률이 0.7 이상이지만, 매칭 단어의 수가 3 이하이므로 레이블이 부여되지 않는다.

[0070] 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.

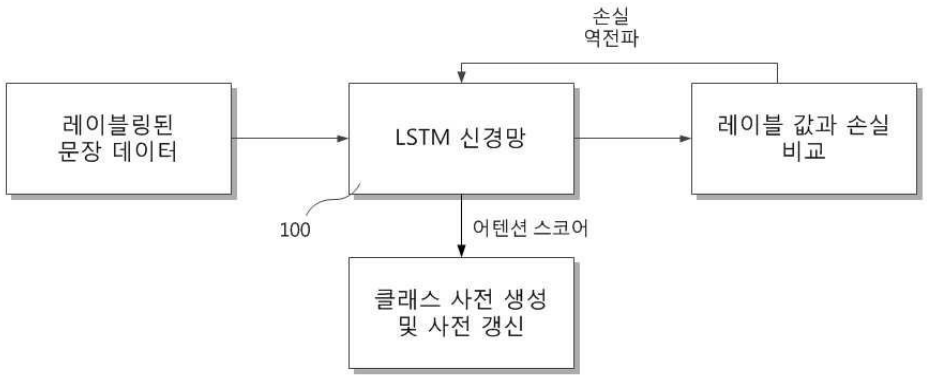
[0071] 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 청구범위의 기술적 사상에 의해 정해져야 할 것이다.

도면

도면1



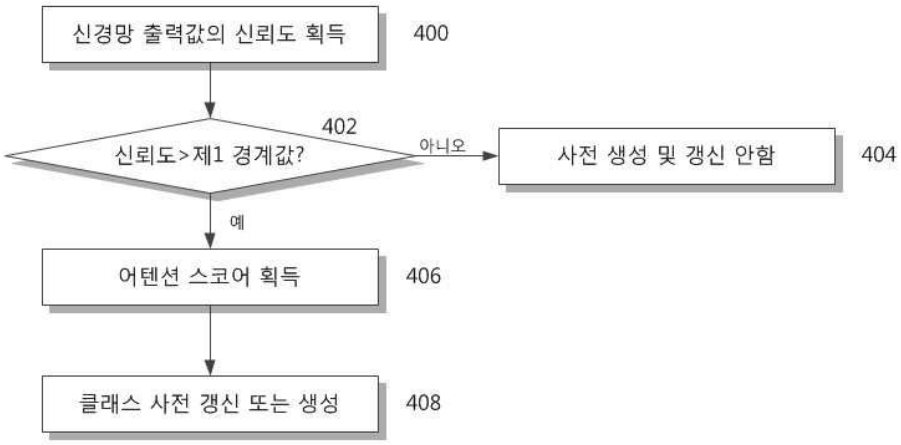
도면2



도면3

Data	$Data_0$	$Data_1$	$Data_2$	$Data_3$	$Data_4$	$Data_5$	$Data_6$	...	$Data_{20000}$
예측값	비윤리	비윤리	윤리	비윤리	윤리	비윤리	비윤리	...	윤리
신뢰도	0.98	0.54	0.68	0.97	0.90	0.99	0.98	...	0.75

도면4



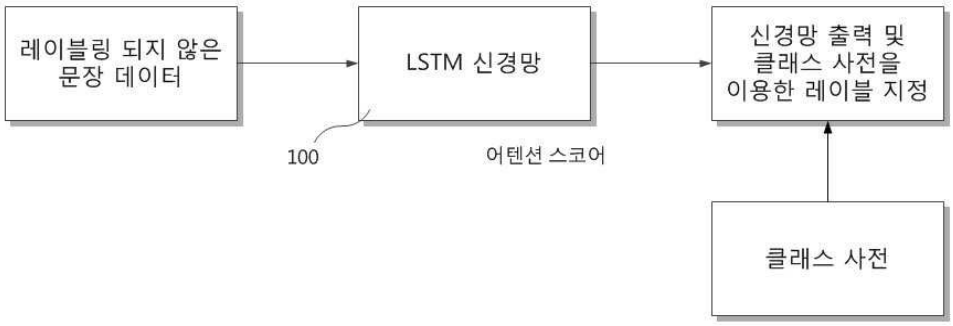
도면5

Words	Attention Score
This	0.024
movie	0.023
is	0.003
stupid	0.4
...	...
bad	0.35

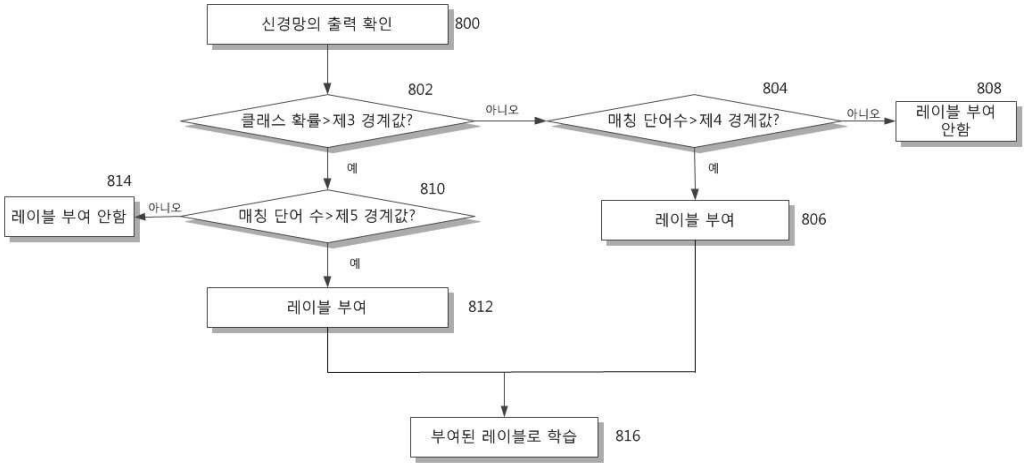
도면6

$MetaLexicon_1$	$MetaLexicon_2$
{stupid, bad, ··· }, {worst, horrible, ··· }, ··· {awful, bad, ··· }	{excellent, good, ··· }, {great, superb, ··· }, ··· {great, usual, ··· }

도면7



도면8



도면9

Data	모델 예측값 (신뢰도)	사전 예측 레이블 (사전 맞춘 개수)
$Data_1$	비윤리 (0.99)	비윤리 (3개)
$Data_2$	윤리 (0.65)	비윤리 (4개)
$Data_3$	윤리 (0.76)	비윤리 (1개)
...	...	...
$Data_{20000}$	비윤리 (0.95)	비윤리 (4개)