

(11) 공개번호 10-2022-0060435
(43) 공개일자 2022년05월11일

(51) 국제특허분류(Int. Cl.)		(71) 출원인	
G06F 40/295 (2020.01)	G06F 40/205 (2020.01)	포항공과대학교 산학협력단	
G06F 40/242 (2020.01)	G06F 40/268 (2020.01)	경상북도 포항시 남구 청암로 77 (지곡동)	
G06F 40/30 (2020.01)		(72) 발명자	
(52) CPC특허분류		이근배	
G06F 40/295 (2020.01)		경상북도 포항시 남구 지곡로 155,9동 1003호	
G06F 40/205 (2020.01)		한상도	
(21) 출원번호	10-2020-0146419	경기도 군포시 광정로 119, 731동 302호	
(22) 출원일자	2020년11월04일	(74) 대리인	
심사청구일자	없음	특허법인이상	

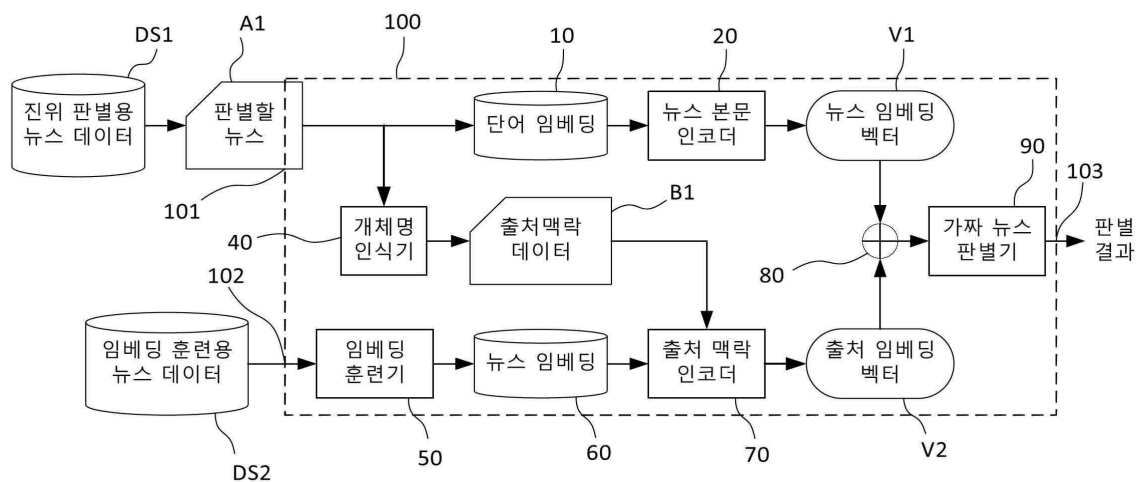
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 출처 정보를 활용하는 가짜 뉴스 탐지 방법 및 시스템

(57) 요약

출처 정보를 활용하여 기계학습 기반으로 가짜 뉴스를 효율적으로 탐지하는 방법 및 시스템이 개시된다. 가짜 뉴스 탐지 시스템은, 판별할 뉴스를 포함한 진위 판별용 뉴스 데이터를 입력 받아 판별할 뉴스의 단어를 벡터와 매칭시키는 단어 임베딩 사전과, 단어 임베딩 사전에서 사상된 판별한 뉴스의 단어들을 하나의 뉴스 임베딩 벡터로 (뒷면에 계속)

대표도 - 도1



변환시키는 뉴스 본문 인코더와, 판별할 뉴스에 존재하는 개체명을 추출하고 추출한 개체명이 어떤 속성에 해당하는지를 판별하여 개체명으로 이루어진 벡터나 개체명만으로 작성된 문서를 포함하는 출처 맥락 데이터를 생성하는 개체명 인식기와, 임베딩 훈련용 뉴스 데이터의 문서 말뭉치를 입력 받아 단어 임베딩을 훈련하는 임베딩 훈련기와, 임베딩 훈련기에서 훈련된 임베딩 훈련용 뉴스 데이터로 뉴스 임베딩을 훈련하는 뉴스 임베딩 모듈과, 출처 맥락 데이터와 뉴스 임베딩을 활용하여 출처 임베딩 벡터를 생성하는 출처 맥락 인코더와, 뉴스 임베딩 벡터와 출처 임베딩 벡터에 기초하여 가짜 뉴스를 판별하는 가짜 뉴스 판별기를 포함한다.

(52) CPC특허분류

G06F 40/242 (2020.01)

G06F 40/268 (2020.01)

G06F 40/30 (2020.01)

G06N 20/10 (2021.08)

G06N 3/08 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711108630
과제번호	2019-0-01906-002
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신방송혁신인재양성
연구과제명	인공지능대학원지원(포항공과대학교)
기 여 율	1/2
과제수행기관명	포항공과대학교 산학협력단
연구기간	2020.01.01 ~ 2020.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711120087
과제번호	2020-0-01789-001
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신방송혁신인재양성
연구과제명	High Performance Knowledge System 개발 및 인력양성
기 여 율	1/2
과제수행기관명	동국대학교산학협력단
연구기간	2020.07.01 ~ 2020.12.31

명세서

청구범위

청구항 1

판별할 뉴스를 포함한 진위 판별용 뉴스 데이터를 입력 받아 상기 판별할 뉴스의 단어를 벡터와 매칭시키는 단어 임베딩 사전;

상기 단어 임베딩 사전에서 사상된 판별한 뉴스의 단어들을 하나의 뉴스 임베딩 벡터로 변환시키는 뉴스 본문 인코더;

상기 판별할 뉴스에 존재하는 개체명을 추출하고 추출한 개체명이 어떤 속성에 해당하는지를 판별하여 개체명으로 이루어진 벡터나 개체명만으로 작성된 문서를 포함하는 출처 맥락 데이터를 생성하는 개체명 인식기;

임베딩 훈련용 뉴스 데이터의 문서 말뭉치를 입력 받아 단어 임베딩을 훈련하는 임베딩 훈련기;

상기 임베딩 훈련기에서 훈련된 임베딩 훈련용 뉴스 데이터로 뉴스 임베딩을 훈련하는 뉴스 임베딩 모듈;

상기 출처 맥락 데이터와 상기 뉴스 임베딩을 활용하여 출처 임베딩 벡터를 생성하는 출처 맥락 인코더; 및

상기 뉴스 임베딩 벡터와 상기 출처 임베딩 벡터에 기초하여 가짜 뉴스를 판별하는 가짜 뉴스 판별기;를 포함하는 가짜 뉴스 탐지 시스템.

청구항 2

청구항 1에 있어서,

상기 단어 임베딩 사전은 의미 추정 기법이나 인공신경망으로 언어 형태소를 다수의 벡터공간 상에 하나씩 매핑하는, 가짜 뉴스 탐지 시스템.

청구항 3

청구항 1에 있어서,

상기 출처 맥락 데이터는 상기 판별할 뉴스에 존재하는 개체명을 추출하는 작성한 데이터로서 개체명만으로 작성된 문서이거나 개체명으로 이루어진 벡터를 포함하는, 가짜 뉴스 탐지 시스템.

청구항 4

청구항 3에 있어서,

상기 출처 맥락 데이터는 상기 판별할 뉴스에서 추출하는 개체명은 '인물'이나 '회사'에 집중되는, 가짜 뉴스 탐지 시스템.

청구항 5

청구항 1에 있어서,

상기 출처 맥락 인코더는 상기 출처 맥락 데이터를 개체명이 적힌 문서로 구축하고, 구축한 문서의 문장을 인코딩하는, 가짜 뉴스 탐지 시스템.

청구항 6

청구항 5에 있어서,

상기 출처 임베딩 벡터는 상기 판별할 뉴스의 출처 정보를 압축하고 있는 벡터 형태의 데이터인, 가짜 뉴스 탐지 시스템.

청구항 7

청구항 1에 있어서,

상기 뉴스 본문 인코더로부터 상기 뉴스 임베딩 벡터를 받고 상기 출처 맥락 인코더로부터 상기 출처 임베딩 벡터를 받고, 상기 뉴스 임베딩 벡터와 상기 출처 임베딩 벡터를 합치는 벡터 결합부를 더 포함하는, 가짜 뉴스 탐지 시스템.

청구항 8

청구항 7에 있어서,

상기 벡터 결합부는 벡터간 결합 및 인공신경망 기반 결합을 활용하는, 가짜 뉴스 탐지 시스템.

청구항 9

청구항 7에 있어서,

상기 가짜 뉴스 판별기는 상기 벡터 결합부로부터 최종 벡터를 받고 상기 최종 벡터에 기초하여 상기 판별할 뉴스의 진위 여부를 판별하는, 가짜 뉴스 탐지 시스템.

청구항 10

청구항 9에 있어서,

상기 가짜 뉴스 판별기는 서포트 벡터 머신 또는 피드포워드 신경망을 포함한 기계 학습을 활용하는, 가짜 뉴스 탐지 시스템.

청구항 11

판별할 뉴스를 포함한 진위 판별용 뉴스 데이터를 입력 받아 상기 판별할 뉴스의 단어를 단어 임베딩을 통해 벡터와 매칭시키는 단계;

상기 단어 임베딩으로 사상된 판별한 뉴스의 단어들을 하나의 뉴스 임베딩 벡터로 변환시키는 단계;

상기 판별할 뉴스에 존재하는 개체명을 추출하고 추출한 개체명이 어떤 속성에 해당하는지를 판별하여 개체명으로서 이루어진 벡터나 개체명만으로 작성된 문서를 포함하는 출처 맥락 데이터를 생성하는 단계;

임베딩 훈련용 뉴스 데이터의 문서 말뭉치를 입력 받아 단어 임베딩을 훈련하는 단계;

상기 임베딩 훈련용 뉴스 데이터로 뉴스 임베딩을 훈련하는 단계;

상기 출처 맥락 데이터와 상기 뉴스 임베딩을 활용하여 출처 임베딩 벡터를 생성하는 단계; 및

상기 뉴스 임베딩 벡터와 상기 출처 임베딩 벡터에 기초하여 가짜 뉴스를 판별하는 단계;를 포함하는 가짜 뉴스 탐지 방법.

청구항 12

청구항 11에 있어서,

상기 매칭시키는 단계는 의미 추정 기법이나 인공신경망으로 언어 형태소를 다수의 벡터공간 상에 하나씩 매핑하는, 가짜 뉴스 탐지 방법.

청구항 13

청구항 11에 있어서,

상기 출처 맥락 데이터는 상기 판별할 뉴스에 존재하는 개체명을 추출하는 작성한 데이터로서 개체명만으로 작성된 문서이거나 개체명으로 이루어진 벡터를 포함하는, 가짜 뉴스 탐지 방법.

청구항 14

청구항 13에 있어서,

상기 출처 맥락 데이터는 상기 판별할 뉴스에서 추출하는 개체명은 '인물'이나 '회사'에 집중되는, 가짜 뉴스

탐지 방법.

청구항 15

청구항 11에 있어서,

상기 출처 임베딩 벡터를 생성하는 단계는 상기 출처 맥락 데이터를 개체명이 적힌 문서로 구축하고, 구축한 문서의 문장을 인코딩하는, 가짜 뉴스 탐지 방법.

청구항 16

청구항 15에 있어서,

상기 출처 임베딩 벡터는 상기 판별할 뉴스의 출처 정보를 압축하고 있는 벡터 형태의 데이터인, 가짜 뉴스 탐지 방법.

청구항 17

청구항 11에 있어서,

상기 가짜 뉴스를 판별하는 단계는 상기 뉴스 임베딩 벡터와 상기 출처 임베딩 벡터를 합치는 단계를 포함하는, 가짜 뉴스 탐지 방법.

청구항 18

청구항 17에 있어서,

상기 합치는 단계는 벡터간 결합 및 인공신경망 기반 결합을 활용하는, 가짜 뉴스 탐지 방법.

청구항 19

청구항 17에 있어서,

상기 가짜 뉴스를 판별하는 단계는 상기 합치는 단계를 거쳐 받은 최종 벡터에 기초하여 상기 판별할 뉴스의 진위 여부를 판별하는 단계를 포함하는, 가짜 뉴스 탐지 방법.

청구항 20

청구항 19에 있어서,

상기 판별할 뉴스의 진위 여부를 판별하는 단계는 서포트 벡터 머신 또는 피드포워드 신경망을 포함한 기계 학습을 활용하는, 가짜 뉴스 탐지 방법.

발명의 설명

기술 분야

[0001] 본 발명은 기계학습 기반의 가짜 뉴스 탐지 기술에 관한 것으로, 보다 상세하게는 출처 정보를 활용하여 인공신경망 기반으로 가짜 뉴스를 효율적으로 탐지하는 방법 및 시스템에 관한 것이다.

배경 기술

[0002] 가짜 뉴스는 일반적으로 정확한 출처를 알 수 없으나 자극적이고 흥미로운 소재를 갖는 정보로서 대중의 관심에 편승하여 대중 속에 빠르고 넓게 퍼져나갈 수 있다. 최근에서 모바일 기기의 대중화와 소셜 네트워크 서비스 등의 정보 전달 플랫폼의 대중화로 인하여 가짜 뉴스가 더 빠르고 넓게 확산되는 현상을 나타내고 있다.

[0003] 예컨대, 기존의 가짜 뉴스 탐지 기술은 뉴스를 대하는 사람들의 반응을 소셜 네트워크 서비스(SNS)에서 추출하여 탐지하거나, 뉴스의 본문을 인코딩하여 탐지하는 방법을 사용한다. 그러나 SNS를 활용하는 방법은 가짜 뉴스를 탐지하고 반응하는 데 시간이 걸린다는 단점이 있다. 그리고 뉴스의 본문을 인코딩하는 방법은 현재로서는 성능이 낮아 활용하기 어렵다는 단점이 있다.

[0004] 한편, 가짜 뉴스를 탐지하기 위한 연구는 2016년 미국 대통령 선거 이후 본격적으로 진행되기 시작했다. 그것은

인공지능을 기반으로 한 자연어 생성 알고리즘을 이용하여 정치적 허위 정보 등을 대량으로 생성할 수 있게 되었기 때문이다. 예를 들어, 수많은 형태의 자연어 텍스트를 생성할 수 있는 인공지능의 다수개(예컨대 수천억개)의 매개변수를 활용하면 사람이 만드는 것과 같은 텍스트를 거의 아무런 제한없이 창작해 낼 수 있게 되었기 때문이다.

[0005] 이와 같이 인공지능으로 가짜 뉴스를 만들어 낼 수 있다. 하지만, 역으로 인공지능이나 기계학습을 사용하여 가짜 뉴스를 탐지할 수 있으며, 이와 관련하여 다양한 연구 개발이 진행되고 있다.

[0006] 하지만, 인공지능이나 기계학습 기반의 가짜 뉴스 탐지 기술은 그 보급이 미미하여 새로운 가짜 뉴스 탐지 방안이 여전히 요구되고 있는 실정이다.

발명의 내용

해결하려는 과제

[0007] 본 발명은 전술한 종래 기술의 요구에 부응하기 위한 것으로, 본 발명의 목적은 뉴스의 출처 정보가 뉴스의 진위 여부를 판단하는데 중요한 근거가 된다는 점에 착안한, 출처 정보를 활용하는 가짜 뉴스 탐지 방법 및 시스템을 제공하는데 있다.

과제의 해결 수단

[0008] 상기 기술적 과제를 해결하기 위한 본 발명의 일 측면에 따른 가짜 뉴스 탐지 시스템은, 판별할 뉴스를 포함한 진위 판별용 뉴스 데이터를 입력 받아 상기 판별할 뉴스의 단어를 벡터와 매칭시키는 단어 임베딩 사전; 상기 단어 임베딩 사전에서 사상된 판별한 뉴스의 단어들을 하나의 뉴스 임베딩 벡터로 변환시키는 뉴스 본문 인코더; 상기 판별할 뉴스에 존재하는 개체명을 추출하고 추출한 개체명이 어떤 속성에 해당하는지를 판별하여 개체명으로 이루어진 벡터나 개체명만으로 작성된 문서를 포함하는 출처 맥락 데이터를 생성하는 개체명 인식기; 임베딩 훈련용 뉴스 데이터의 문서 말뭉치를 입력 받아 단어 임베딩을 훈련하는 임베딩 훈련기; 상기 임베딩 훈련기에서 훈련된 임베딩 훈련용 뉴스 데이터로 뉴스 임베딩을 훈련하는 뉴스 임베딩 모듈; 상기 출처 맥락 데이터와 상기 뉴스 임베딩을 활용하여 출처 임베딩 벡터를 생성하는 출처 맥락 인코더; 및 상기 뉴스 임베딩 벡터와 상기 출처 임베딩 벡터에 기초하여 가짜 뉴스를 판별하는 가짜 뉴스 판별기;를 포함한다.

[0009] 상기 기술적 과제를 해결하기 위한 본 발명의 또 다른 측면에 따른 가짜 뉴스 탐지 방법은, 판별할 뉴스를 포함한 진위 판별용 뉴스 데이터를 입력 받아 상기 판별할 뉴스의 단어를 단어 임베딩을 통해 벡터와 매칭시키는 단계; 상기 단어 임베딩으로 사상된 판별한 뉴스의 단어들을 하나의 뉴스 임베딩 벡터로 변환시키는 단계; 상기 판별할 뉴스에 존재하는 개체명을 추출하고 추출한 개체명이 어떤 속성에 해당하는지를 판별하여 개체명으로 이루어진 벡터나 개체명만으로 작성된 문서를 포함하는 출처 맥락 데이터를 생성하는 단계; 임베딩 훈련용 뉴스 데이터의 문서 말뭉치를 입력 받아 단어 임베딩을 훈련하는 단계; 상기 임베딩 훈련용 뉴스 데이터로 뉴스 임베딩을 훈련하는 단계; 상기 출처 맥락 데이터와 상기 뉴스 임베딩을 활용하여 출처 임베딩 벡터를 생성하는 단계; 및 상기 뉴스 임베딩 벡터와 상기 출처 임베딩 벡터에 기초하여 가짜 뉴스를 판별하는 단계;를 포함한다.

발명의 효과

[0010] 전술한 출처 정보를 활용하는 가짜 뉴스 탐지 방법 및 시스템을 사용하는 경우에는, 뉴스의 본문을 활용하여 출처 정보를 추정하고 추정한 출처 정보에 기초하여 가짜 뉴스 여부를 효과적으로 탐지할 수 있다.

[0011] 또한, 본 발명에 의하면, 가짜 뉴스에 대한 빠른 대응이 가능하며 더욱 정확한 탐지가 가능한 가짜 뉴스 탐지 방법 및 시스템을 제공할 수 있다.

도면의 간단한 설명

[0012] 도 1은 본 발명의 일실시예에 따른 출처 정보를 활용하는 가짜 뉴스 탐지 시스템의 구조도이다.

도 2는 도 1의 가짜 뉴스 탐지 시스템에 채용할 수 있는 벡터 결합부에 대한 블록도이다.

도 3은 본 발명의 다른 실시예에 따른 출처 정보를 활용하는 가짜 뉴스 탐지 방법의 흐름도이다.

발명을 실시하기 위한 구체적인 내용

- [0013] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세하게 설명하고자 한다. 그러나, 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.
- [0014] 제1, 제2 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 본 발명의 권리 범위를 벗어나지 않으면서 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다. 및/또는 이라는 용어는 복수의 관련된 기재된 항목들의 조합 또는 복수의 관련된 기재된 항목들 중의 어느 항목을 포함한다.
- [0015] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소가 다른 구성요소에 "직접 연결되어" 있다거나 "직접 접속되어" 있다고 언급된 때에는, 중간에 다른 구성요소가 존재하지 않는 것으로 이해되어야 할 것이다.
- [0016] 본 출원에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 출원에서, "포함하다" 또는 "가지다" 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.
- [0017] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가지고 있다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥 상 가지는 의미와 일치하는 의미를 가진 것으로 해석되어야 하며, 본 출원에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.
- [0018] 이하, 본 발명의 바람직한 실시예를, 첨부한 도면들을 참조하여 보다 상세하게 설명한다.
- [0019] 도 1은 본 발명의 일실시예에 따른 출처 정보를 활용하는 가짜 뉴스 탐지 시스템(이하 간략히 '가짜 뉴스 탐지 시스템'이라 한다)의 구조도이다. 그리고 도 2는 도 1의 가짜 뉴스 탐지 시스템에 채용할 수 있는 벡터 결합부에 대한 블록도이다.
- [0020] 도 1을 참조하면, 본 실시예에 따른 가짜 뉴스 탐지 시스템은 제1 데이터베이스 시스템(DS1), 제2 데이터베이스 시스템(DS2) 및 시스템 본체(100)를 포함한다. 시스템 본체(100)는 단어 임베딩 사전(10), 뉴스 본문 인코더(20), 객체명 인식기(40), 임베딩 훈련기(50), 뉴스 임베딩 모듈(60), 출처 맥락 인코더(70), 벡터 결합부(80) 및 가짜 뉴스 판별기(90)를 구비한다.
- [0021] 시스템 본체(100)는 제1 데이터베이스 시스템(DS1)으로부터 진위 판별용 뉴스 데이터가 '판별한 뉴스'로 입력되는 제1 입력부(101)와, 제2 데이터베이스 시스템(DS2)으로부터 임베딩 훈련용 뉴스 데이터가 입력되는 제2 입력부(102)와, 가짜 뉴스 판별기(90)의 판별 결과를 출력하는 출력부(103)를 구비한다.
- [0022] 좀더 구체적으로 각 구성요소를 살펴보면, 제1 데이터베이스 시스템(DS1)은 진위 판별용 뉴스 데이터를 저장하고 관리한다. 진위 판별용 뉴스 데이터는 가짜 뉴스인지 아닌지의 진위 여부가 표시된 뉴스 데이터이다. 진위 판별용 뉴스 데이터는 가짜 뉴스를 판별해주는 웹 서비스로부터 자동 추출할 수 있다.
- [0023] 진위 판별용 뉴스 데이터는 서울대학교 언론정보연구소의 팩트체크(SUN factcheck), 포인터 연구소(The Point Insititute)의 폴리티팩트(POLITIFACT), 미국의 팩트체크기구(FactCheck.org), 팩트체커(FactChecker), 영국의 위키텐트리븐, 프랑스의 크로스체크(cross check), 노르웨이의 팩티스크, 덴마크의 디텍토, 아르헨티나의 체케아도(Chequeado), 독일의 코렉티브(Corrective), 국제 팩트체크 네트워크(IFCN, international fact-checking network) 등의 현재 이용가능한 웹 서비스를 통해 추출할 수 있다.
- [0024] 팩트체크기구가 제시하는 가짜 뉴스 판별법을 보면, 첫째 정보 제공처를 확인하고, 둘째 제목 너머를 읽고, 셋째 글쓴이를 확인하고, 넷째 근거 자료를 링크하고, 다섯째 날짜를 확인하고, 여섯째 농담인지를 판단하고, 일곱째 독자가 편견이 있는지를 확인하고, 여덟째 전문가에게 물어볼 것을 제시한다.
- [0025] 또한, 제1 데이터베이스 시스템(DS1)은 진위 판별용 뉴스 데이터로써 소셜 네트워크 서비스(social network service)에서 활발히 공유되는 뉴스를 추출하는 방법 등 다양한 방법들 중 적어도 어느 하나를 이용하여 추출된

진위 판별용 뉴스 데이터를 저장하고 관리할 수 있다.

- [0026] 또한, 제1 데이터베이스 시스템(DS1)은 시스템 본체(100)에 판별할 뉴스(A1)를 제공한다. 판별할 뉴스(A1)는 가짜 뉴스 판별기(95)가 최종적으로 판별하고자 하는 뉴스 데이터 중 하나의 샘플이다.
- [0027] 제2 데이터베이스 시스템(DS2)은 임베딩 훈련용 뉴스 데이터를 저장하고 관리한다. 이 뉴스 데이터는 뉴스 임베딩을 구축하기 위한 임베딩 훈련 데이터로, 진위 판별용 뉴스 데이터와 같을 필요는 없다. 임베딩 훈련용 뉴스 데이터는 뉴스 임베딩을 충분히 훈련할 수 있을 정도의 규모가 확보된 뉴스 데이터이어야 한다. 단, 뉴스가 시간이나 토픽에 민감한 데이터인 만큼, 진위 판별용 뉴스 데이터와 비슷하게 구축될 필요가 있을 수 있다.
- [0028] 제1 데이터베이스 시스템(DS1)으로부터 '판별할 뉴스'를 포함한 진위 판별용 뉴스 데이터가 입력되면, 단어 임베딩 사전(10)은 '판별할 뉴스'에 포함된 단어를 벡터와 매칭시킨다. 매칭되는 벡터는 단어의 의미를 의미공간에 하나의 벡터로 표현된 것이다. 단어 임베딩 사전(10)은 단어를 벡터와 매칭시키는 사전 모듈이다. 단어 임베딩 사전(10)은 word2vec, GloVe(global vectors) 등의 의미 추정 기법을 이용하여 인공신경망 기반으로 한국어 형태소 등의 언어 형태소를 수천 차원의 벡터공간(vector space) 상에 하나씩 매핑시킬 수 있다. 이 경우, 비슷한 맥락을 갖는 단어들은 가까운 벡터를 지니게 되고, 분산 시맨틱스 가정에 기초하여 벡터끼리 시맨틱 연산을 수행할 수 있다.
- [0029] 뉴스 본문 인코더(20)는 단어 임베딩 사전(10)의 출력단에 연결되고 단어 임베딩에 기초하여 판별할 뉴스(A1)를 하나의 벡터로 변환시킨다. 뉴스 본문 인코더(20)는 단어 임베딩으로 사상된 판별할 뉴스(A1)의 단어들을 하나의 벡터로 표현한다. 뉴스 본문 인코더(20) RNN(Revolutional Neural Network), CNN(Convolutional Neural Network), BERT(Bidirectional Encoder Representations from Transformer), Transformer(one of maching learning model), GPT(Generative Pre-trained Transformer) 등의 일반적인 인코딩 모델의 형식을 취할 수 있다.
- [0030] 객체명 인식기(40)는 입력된 문장 내에 존재하는 개체명을 추출하고, 해당 개체명이 어떤 속성에 해당하는지를 판별한다. 예를 들어, '트럼프는 미국의 대통령이다'라는 문장의 개체명은 '트럼프, 미국, 대통령'이고, 그 속성은 '인물, 국가, 직업'에 해당한다.
- [0031] 개체명 인식기(40)는 추출한 개체명과 이 개체명에 대하여 판별한 속성 정보를 포함하는 출처 맥락 데이터를 생성하여 뉴스 임베딩 모듈(60)에 출처 맥락 데이터(B1)를 제공한다.
- [0032] 출처 맥락 데이터(B1)는 판별한 뉴스 데이터를 개체명 인식기(40)에 적용하여 뉴스에 존재하는 개체명을 추출하여 작성한 데이터이다. 이 데이터는 개체명만으로 새로 작성한 문서나 개체명으로 이루어진 벡터 등을 포함한다. 특히 본 실시예에서 출처 정보를 활용하고자 하는 만큼, '판별한 뉴스'에서 추출하는 개체명은 '인물'이나 '회사'에 집중될 수 있다.
- [0033] 임베딩 훈련기(50)는 제2 데이터베이스 시스템(DS2)로부터 문서 말뭉치를 입력 받고 단어 임베딩을 훈련한다. 임베딩 훈련기(50)에서 활용하는 훈련기는 word2vec, GloVe(global vectors), 패스트텍스트(fasttext) 등을 포함할 수 있다.
- [0034] 패스트텍스트의 경우, 한글로 된 코퍼스(corpus)를 자소 분리하여 패스트텍스트의 입력으로 전달함으로 각각의 하나의 문자에 대하여 n-gram을 하도록 한다. 여기서, n-gram은 n(임의의 자연수)개 어절/음절을 연속적으로 분류해 그 빈도를 따지는 것으로, 간단히 말해서 입력한 문자열을 N개의 기준 단위로 절단하는 것을 말한다.
- [0035] 뉴스 임베딩 모듈(60)은 문서 말뭉치에 포함된 임베딩 훈련용 뉴스 데이터로 훈련한 '뉴스 단어 임베딩'이다. 이러한 본 실시예의 뉴스 임베딩은 일반적인 '단어 임베딩'과는 다르며, 추출한 뉴스의 시기적, 이슈적 정보가 단어 임베딩보다 잘 반영된 임베딩이다.
- [0036] 출처 맥락 인코더(70)는 출처 맥락 데이터(B1)와 뉴스 임베딩을 활용하여 출처 임베딩 벡터를 만드는 모듈이다. 간단한 일례로는 출처 맥락 데이터를 개체명이 적힌 문서로 구축하고, 이를 문장 인코딩하는 인코더라고 볼 수 있다. 한편, 본 실시예에서 출처 맥락 인코더(70)의 입력은 상황에 따라 다양한 방식의 인코딩 방식을 포함할 수 있고, RNN, CNN, BERT, GPT, 기계학습 모델의 일종인 Transformer 등의 일반적인 인코딩 모델의 형식을 취할 수 있다.
- [0037] 출처 맥락 인코더(70)의 출력 벡터인 출처 임베딩 벡터는 판별할 뉴스의 출처 정보를 압축하고 있는 벡터 형태의 데이터이다.

- [0038] 벡터 결합부(80)는 뉴스 임베딩 벡터(V1)와 출처 임베딩 벡터(V2)를 다양한 방식으로 합친다. 이때 합치는 방식은 도 2에 도시한 바와 같이 벡터간 결합 및 인공신경망 기반 결합을 포함한다. 즉, 벡터 결합부(80)는 벡터간 결합부(81)와 인공신경망 기반 결합부(82)를 구비할 수 있다. 벡터 결합부(80)에서 합쳐진 최종 벡터는 가짜 뉴스 판별기(90)의 입력으로 전달된다.
- [0039] 다시 도 1을 참조하면, 가짜 뉴스 판별기(90)는 최종 벡터에 포함된 뉴스 임베딩 벡터(V1)와 출처 임베딩 벡터(V2)를 활용하여 판별할 뉴스의 진위 여부를 최종 판별한다. 가짜 뉴스 판별기(90)는 서포트 벡터 머신(support vector machine, SVM), 피드포워드 신경망(feed forward neural network 혹은 feed forward network) 등의 기계 학습 방식을 활용할 수 있다.
- [0040] 여기서, 가짜 뉴스 판별기(90)에 활용되는 서포트 벡터 머신은 뉴스 임베딩 벡터(V1)와 출처 임베딩 벡터(V2)의 분류와 회귀 분석으로 판별할 뉴스의 진위 여부를 판별하도록 구현될 수 있다.
- [0042] 도 3은 본 발명의 다른 실시예에 따른 출처 정보를 활용하는 가짜 뉴스 탐지 방법의 흐름도이다.
- [0043] 도 3을 참조하면, 본 실시예에 따른 가짜 뉴스 탐지 방법은, 먼저 제1 데이터베이스 시스템으로부터 판별할 뉴스를 포함한 진위 판별용 뉴스 데이터가 입력되고, 제2 데이터베이스 시스템으로부터 임베딩 훈련용 뉴스 데이터가 입력되면, 이 데이터를 토대로 뉴스 임베딩 벡터와 출처 임베딩 벡터를 생성하고 이 벡터들에 기초하여 판별한 뉴스의 진위 여부를 판별한다.
- [0044] 좀더 구체적으로, 가짜 뉴스 탐지 방법은, 먼저 단어 임베딩 사전에서 판별할 뉴스를 포함한 진위 판별용 뉴스 데이터를 입력 받아 판별할 뉴스의 단어를 단어 임베딩을 통해 벡터와 매칭시킨다(S31). 매칭시키는 단계는 의미 추정 기법이나 인공신경망으로 언어 형태소를 다수의 벡터공간 상에 하나씩 매핑하도록 구현될 수 있다.
- [0045] 다음, 단어 임베딩으로 사상된 판별한 뉴스의 단어들을 하나의 뉴스 임베딩 벡터로 변환시킨다(S32).
- [0046] 또한, 판별할 뉴스에 존재하는 개체명을 추출하고 추출한 개체명이 어떤 속성에 해당하는지를 판별하여 개체명으로 이루어진 벡터나 개체명만으로 작성된 문서를 포함하는 출처 맥락 데이터를 생성한다(S33). 출처 맥락 데이터는 상기 판별할 뉴스에 존재하는 개체명을 추출하는 작성한 데이터로서 개체명만으로 작성된 문서이거나 개체명으로 이루어진 벡터를 포함한다. 또한, 본 실시예의 출처 맥락 데이터는 상기 판별할 뉴스에서 추출하는 개체명은 '인물'이나 '회사'에 집중될 수 있다.
- [0047] 또한, 위의 진위 판별용 뉴스 데이터와 병렬적으로 입력되는 임베딩 훈련용 뉴스 데이터의 문서 말뭉치로 단어 임베딩을 훈련한다(S34).
- [0048] 다음, 단어 임베딩 훈련이 된 임베딩 훈련용 뉴스 데이터로 뉴스 임베딩을 훈련한다(S35).
- [0049] 다음, 출처 맥락 데이터와 뉴스 임베딩을 활용하여 출처 임베딩 벡터를 생성한다(S36). 본 단계는 상기 출처 맥락 데이터를 개체명이 적힌 문서로 구축하고, 구축한 문서의 문장을 인코딩하도록 구현될 수 있다. 출처 임베딩 벡터는 상기 판별할 뉴스의 출처 정보를 압축하고 있는 벡터 형태의 데이터를 포함한다.
- [0050] 다음으로, 뉴스 임베딩 벡터와 출처 임베딩 벡터에 기초하여 가짜 뉴스를 판별한다(S37). 본 단계는 뉴스 임베딩 벡터와 상기 출처 임베딩 벡터를 합치는 단계와 합치는 단계를 거쳐 받은 최종 벡터에 기초하여 상기 판별할 뉴스의 진위 여부를 판별하는 단계를 포함할 수 있다. 여기서, 합치는 단계는 벡터간 결합 및 인공신경망 기반 결합을 활용하고, 판별할 뉴스의 진위 여부를 판별하는 단계는 서포트 벡터 머신 또는 피드포워드 신경망을 포함한 기계 학습을 활용하도록 구현될 수 있다.
- [0051] 전술한 본 실시예에 의하면, 뉴스의 본문을 활용하여 출처 정보를 효과적으로 추정함으로써 추정한 출처 정보에 이용하여 판별할 뉴스의 진위 여부를 탐지함으로써 매우 정확한 가짜 뉴스의 탐지가 가능하다.
- [0052] 한편, 본 실시예에 따른 가짜 뉴스 탐지 방법은 다양한 컴퓨터 수단을 통해 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 컴퓨터 판독 가능 매체에 기록되는 프로그램 명령은 본 발명을 위해 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수 있다.
- [0053] 컴퓨터 판독 가능 매체의 예에는 롬(rom), 램(ram), 플래시 메모리(flash memory) 등과 같이 프로그램 명령을

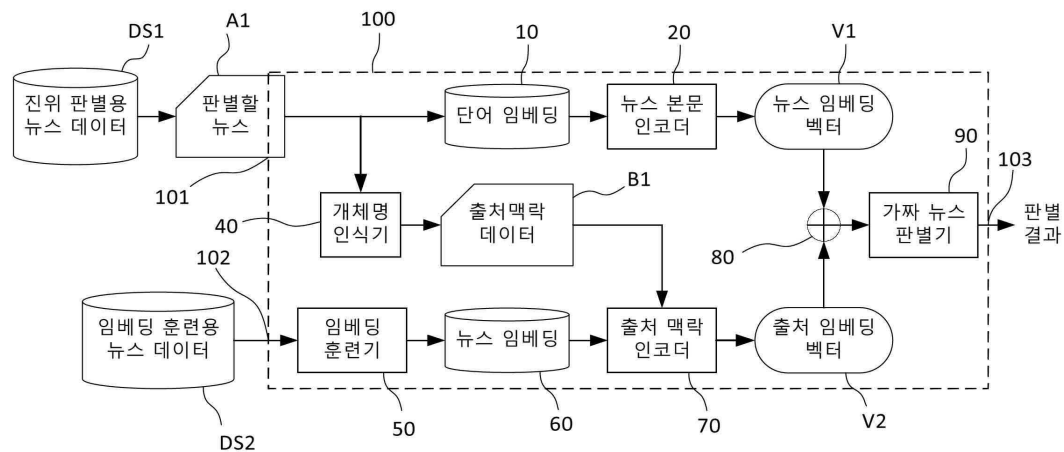
저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러(compiler)에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터(interpreter) 등을 사용해서 컴퓨터에 의해 실행될 수 있는 고급 언어 코드를 포함한다. 상술한 하드웨어 장치는 본 발명의 동작을 수행하기 위해 적어도 하나의 소프트웨어 모듈로 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

[0054]

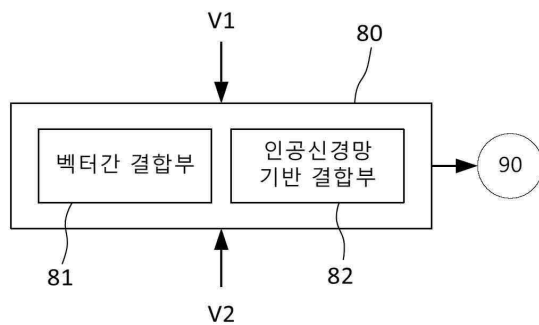
이상과 같이 실시예들을 참조하여 설명하였지만, 해당 기술 분야의 숙련된 당업자는 하기의 청구범위에 기재된 본 발명의 사상 및 영역으로부터 벗어나지 않는 범위 내에서 본 발명을 다양하게 수정 및 변경시킬 수 있음을 이해할 수 있을 것이다.

도면

도면1



도면2



도면3

