

(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0064884

(43) 공개일자 2020년06월08일

- (51) 국제특허분류(Int. Cl.)
G06F 40/20 (2020.01) G06N 3/08 (2006.01)
- (52) CPC특허분류
G06F 40/205 (2020.01)
G06F 40/284 (2020.01)
- (21) 출원번호 10-2019-0098775
- (22) 출원일자 2019년08월13일
심사청구일자 2019년08월13일
- (30) 우선권주장
1020180151181 2018년11월29일 대한민국(KR)

- (71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
- (72) 발명자
정창성
서울특별시 노원구 노원로 214, 6동 905호 (하계동, 삼익선경아파트)
- 서영경
서울특별시 동작구 동작대로29가길 15 (사당동)
- (74) 대리인
특허법인엠에이피에스

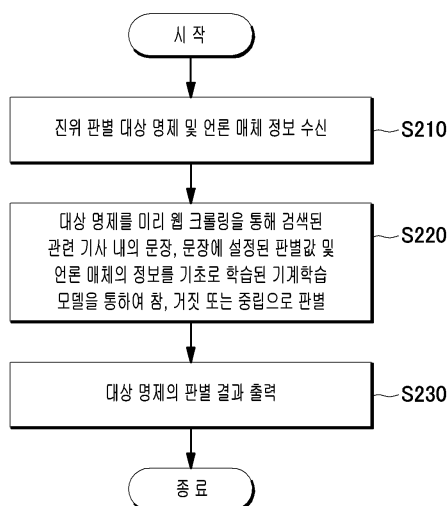
전체 청구항 수 : 총 15 항

(54) 발명의 명칭 언론 매체의 신뢰성을 반영한 기사 판별 시스템 및 명제의 진위 판별 방법

(57) 요약

본 발명의 일 실시예에 따른 언론 매체의 신뢰성을 반영한 기사 판별 시스템에 의해 수행되는 기사 판별을 위한 기계학습 모델을 학습시키는 방법으로서, 진위 판별을 위한 대상 명제를 수신하는 단계; 수신한 대상 명제를 기초로 웹 크롤링을 수행하여 관련 기사를 검색하고, 대상 명제와 검색된 관련 기사에 포함된 문장의 유사도를 산출하는 단계; 산출된 문장의 유사도에 따라 문장을 참, 거짓 또는 중립으로서 판별하여 판별값을 설정하는 단계; 및 유사도가 산출된 문장, 문장에 각각 설정된 판별값 및 문장을 포함하는 관련 기사를 제공한 언론 매체의 정보를 학습데이터로서 기계학습 모델에 입력하여 학습을 수행하는 단계를 포함하는, 기사 판별 기계학습 모델의 학습 방법을 제공하고자 한다.

대표도 - 도2



(52) CPC특허분류

G06N 3/08 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호 1711073495

부처명 과학기술정보통신부

연구관리전문기관 정보통신기술진흥센터

연구사업명 정보통신방송연구개발사업

연구과제명 가짜뉴스 판독을 위한 딥러닝 기반 자연어 콘텐츠 가치 판별 기술 개발

기 여 율 1/1

주관기관 주식회사 누아

연구기간 2018.04.01 ~ 2018.12.31

명세서

청구범위

청구항 1

언론 매체의 신뢰성을 반영한 기사 판별 시스템에 의해 수행되는 기사 판별을 위한 기계학습 모델을 학습시키는 방법으로서,

진위 판별을 위한 대상 명제를 수신하는 단계;

상기 수신한 대상 명제를 기초로 웹 크롤링을 수행하여 관련 기사를 검색하고, 상기 대상 명제와 상기 검색된 관련 기사에 포함된 문장의 유사도를 산출하는 단계;

상기 산출된 문장의 유사도에 따라 상기 문장을 참, 거짓 또는 중립으로서 판별하여 판별값을 설정하는 단계; 및

상기 유사도가 산출된 문장, 상기 문장에 각각 설정된 판별값 및 상기 문장을 포함하는 상기 관련 기사를 제공한 언론 매체의 정보를 학습데이터로서 기계학습 모델에 입력하여 학습을 수행하는 단계를 포함하는, 기사 판별 기계학습 모델의 학습 방법.

청구항 2

제1항에 있어서

상기 웹 크롤링은,

상기 대상 명제와 웹페이지 상의 문장을 비교하여 코사인 유사도(Cosine Similarity)가 높은 문장을 검색하는 것을 포함하는, 기사 판별 기계학습 모델의 학습 방법.

청구항 3

제1항에 있어서,

상기 판별값을 설정하는 단계는,

상기 검색 대상 명제와 상기 관련 기사에 포함된 문장의 유사도가 50%이상이면 상기 문장은 참이고, 상기 유사도가 30%이상 50%미만이면 거짓이고, 상기 유사도가 30% 미만이면 중립으로 설정되는, 기사 판별 기계학습 모델의 학습 방법.

청구항 4

제1항에 있어서,

상기 기계학습 모델은 CNN(Convolution Neural Network) 모델을 이용한 것인, 기사 판별 기계학습 모델의 학습 방법.

청구항 5

언론 매체의 신뢰성을 반영한 기사 판별 시스템에 의해 수행되는 명제의 진위를 판별하는 방법에 있어서,

진위를 판별하고자 하는 대상 명제를 수신하는 단계;

상기 수신한 대상 명제를 미리 웹 크롤링을 통해 검색된 상기 명제와 관련된 관련 기사 내의 문장, 상기 문장에 설정된 판별값 및 상기 관련 기사를 제공한 언론 매체의 정보를 기초로 미리 학습된 기계학습 모델을 통하여 참, 거짓 또는 중립으로 판별하는 단계; 및

상기 대상 명제의 판별 결과를 출력하는 단계를 포함하는, 언론 매체의 신뢰성을 반영한 명제의 진위 판별 방법.

청구항 6

제5항에 있어서,

상기 대상 명제를 수신하는 단계에서, 상기 대상 명제가 추출된 언론 매체의 정보를 더 수신하고,

상기 판별 단계에서, 상기 기계학습 모델은 상기 언론 매체 정보에 더 기초하여 상기 대상 명제의 진위를 판별하는, 언론 매체의 신뢰성을 반영한 명제의 진위 판별 방법.

청구항 7

제5항에 있어서,

상기 웹 크롤링은 상기 대상 명제와 웹페이지 상의 문장을 비교하여 코사인 유사도(Cosine Similarity)가 높은 문장을 검색하는 것을 포함하는, 언론 매체의 신뢰성을 반영한 명제의 진위 판별 방법.

청구항 8

제5항에 있어서,

상기 설정된 판별값은,

상기 검색 대상 명제와 상기 관련 기사에 포함된 문장의 유사도가 50%이상이면 상기 문장은 참이고, 상기 유사도가 30%이상 50%미만이면 거짓이고, 상기 유사도가 30% 미만이면 중립으로 설정된 것인, 언론 매체의 신뢰성을 반영한 명제의 진위 판별 방법.

청구항 9

제6항에 있어서,

상기 기계학습 모델은 CNN(Convolution Neural Network) 모델을 이용한 것인, 언론 매체의 신뢰성을 반영한 명제의 진위 판별 방법.

청구항 10

언론 매체의 신뢰성을 반영한 기사 판별 시스템에 있어서,

언론 매체의 신뢰성을 반영하여 기사를 판별하는 프로그램이 저장된 메모리; 및

상기 메모리에 저장된 프로그램을 실행하는 프로세서를 포함하고,

상기 프로세서는 상기 프로그램의 수행에 따라,

진위 판별을 위한 대상 명제를 수신하고, 상기 수신한 대상 명제를 기초로 웹 크롤링을 수행하여 관련 기사를 검색하고, 상기 대상 명제와 상기 검색된 관련 기사에 포함된 문장의 유사도를 산출하고, 상기 산출된 문장의 유사도에 따라 상기 문장을 참, 거짓 또는 중립으로서 판별하여 판별값을 설정하고, 상기 유사도가 산출된 문장과 상기 문장에 각각 설정된 판별값을 학습데이터로서 기계학습 모델에 입력하여 학습을 수행하여 학습된 기계학습 모델을 통해 기사의 진위를 판별하는, 언론 매체의 신뢰성을 반영한 기사 판별 시스템.

청구항 11

제10항에 있어서,

상기 기계학습 모델을 학습시킨 다음에,

진위를 판별하고자 하는 대상 명제를 수신하고, 상기 수신한 대상 명제를 상기 미리 학습된 기계학습 모델을 통하여 참, 거짓 또는 중립으로 판별하고, 상기 대상 명제의 판별 결과를 출력하는, 언론 매체의 신뢰성을 반영한 기사 판별 시스템.

청구항 12

제11항에 있어서,

상기 대상 명제의 진위 판별을 위해와 상기 대상 명제가 추출된 언론 매체의 정보를 더 수신하고,

상기 기계학습 모델은 상기 수신한 언론 매체 정보에 더 기초하여 상기 대상 명제의 진위를 판별하는, 언론 매체의 신뢰성을 반영한 기사 판별 시스템.

청구항 13

제10항에 있어서,

상기 웹 크롤링은 상기 대상 명제와 웹페이지 상의 문장을 비교하여 코사인 유사도(Cosine Similarity)가 높은 문장을 검색하는 것을 포함하는, 언론 매체의 신뢰성을 반영한 기사 판별 시스템.

청구항 14

제10항에 있어서,

상기 설정된 판별값은,

상기 검색 대상 명제와 상기 관련 기사에 포함된 문장의 유사도가 50%이상이면 상기 문장은 참이고, 상기 유사도가 30%이상 50%미만이면 거짓이고, 상기 유사도가 30% 미만이면 중립으로 설정된 것인, 언론 매체의 신뢰성을 반영한 기사 판별 시스템.

청구항 15

제10항에 있어서,

상기 기계학습 모델은 CNN(Convolution Neural Network) 모델을 이용한 것인, 언론 매체의 신뢰성을 반영한 기사 판별 시스템.

발명의 설명

기술 분야

[0001] 본 발명은 언론 매체의 신뢰성을 반영한 기사 판별 시스템 및 명제의 진위 판별 방법에 관한 것으로서, 언론 매체의 신뢰성을 평가하고 이를 반영하여 기사를 판별하는 시스템과 기사에 기재된 명제의 진위를 판별하는 방법에 관한 것이다.

배경 기술

[0002] 현재, 각종 언론 매체에서 인공지능을 이용하여 쓰여진 다량의 기사를 제공하고 있다. 이렇게 제공되는 기사의 양이 많아짐에 따라 언론 매체별로 제공하는 기사의 진위를 판별하는 것이 힘들어지고 있다.

[0003] 언론 매체의 수가 늘어나면서, 정확한 정보를 제공하는 기사를 파악하는 것이 중요해지고 있으며, 정확도가 부족한 잘못된 정보를 전달하는 페이크 뉴스(Fake News)를 찾아내는 것은 필수적인 기술로 떠오르고 있다.

[0004] 페이크 뉴스 탐색 시스템은 다양한 종류가 있으며, 그 중에 BS 탐색기(BS detector)는 모질(Mozil)사에서 개발된 플러그인으로서 가짜 뉴스 웹사이트를 경고하는 기능을 가지는데, 페이스북(Facebook)에서 사용되고 있다. 이 시스템은 사람들에게 의해 측정된 가짜 뉴스에 플래그를 표시하는 서비스를 제공한다. 미국에서는 정치 뉴스에 대해 진위 여부를 확인하여 폴리팩트(PolitiFact)라는 웹사이트에 개시한다. 여기에서는 뉴스의 진위를 여섯 단계로 평가하는 방법을 사용한다. 하지만 이러한 서비스에서 진위를 확인할 수 있는 뉴스의 종류는 정치 뉴스에 한정되어 있으며, 뉴스의 진위가 사람에 의해 확인되는 것은 마찬가지이다. 또한, 가짜 뉴스를 탐색하는 시스템으로서 플록 페이크 뉴스 탐색기(Flock Fake News Detector)라는 시스템이 있으며, 이는 통계적으로 채팅 서비스 사용자들이 뉴스에 순위를 매겨 진위 여부를 판별할 수 있는 구성을 가진다.

[0005] 상술한 페이크 뉴스 탐색 시스템은 사람에 의해 뉴스의 거짓이 확인되어 결과를 확인하는데 오랜 시간이 걸리고, 데이터셋이 부족하다는 결정적인 단점이 있다.

[0006] 한편, 데이터 과학 영역에 있어서 이전에는 데이터베이스에 데이터가 구조적인 형태로 저장되어 시스템이 데이터베이스에 쿼리를 전송하여 응답을 얻는 방식의 질의 응답 시스템(Question Answering System)이 존재했다. 다양한 딥러닝 모델이 질의 응답 시스템(Question Answering System)에 적용되었으나, 딥러닝 모델로 언어 모델

을 구현하는 경우, 히든 유닛 벡터가 바로 이전의 정보만을 저장하고 있기 때문에 의미 있는 긴 문장을 적용하여 언어 모델을 구현하는 것에는 어려움이 있었다.

[0007] 예를 들어, 시암FC(SiamFC)와 같은 질의 응답 시스템(Question Answering System)은 템플릿 매칭(Template-Matching) 방법을 사용하여 학습하고 정답을 판단하는 데 빠른 시간이 걸리지만, 타겟 문장(Target Sentence)가 다양한 형태를 가지고 있을 경우에는 정확도가 낮아지는 문제점이 있다. 이 때문에, 긴 문장들이 히든 벡터에 저장될 때 손실이 생길 수 있다. 긴 문장 형태를 가지는 입력의 손실을 방지하기 위해, 단어들을 전방향, 후방향으로 인코딩하고 이 벡터들을 합쳐 문장의 손실을 줄이는 바이 디렉셔널 RNN(Bi-directional RNN)을 사용하는 경우가 있는데, 이는 메모리 네트워크를 두 개의 모듈로 나누어 적용할 수 있다. 이들은 각각 문장 또는 단어들을 메모리에 저장하는 모듈과 질문과 가장 관련 있는 문장을 찾는 모듈이다. 이를 통해 이 모델은 타겟 문장의 형태가 변화하는 문제를 해결할 수 있다. 일 예로, 페이스북에서 질의 응답 시스템에 바이 디렉셔널 RNN 모델을 적용시켜 91퍼센트의 정확도를 구현하였다.

[0008] 상술한 두 개의 모델들은 지도학습이 필요하기 때문에, 학습할 때 적절한 가중치 값들이 정해져 있어야 한다. 엔드 투 엔드(End-to-End) 메모리 네트워크는 이러한 모델들을 보완하여 개발되었다. 이 모델은 문제와 문장의 유사성을 가중치합(Weighted Sum)으로 계산하여 최종 정답을 생성하고, 이를 메모리에 업데이트할 수 있다. 메모리의 적은 용량을 극복하기 위해, 다이내믹 메모리 네트워크(Dynamic Memory Network)가 개발되어 입력 문장을 일화 기억(Episodic Memory) 형태로 저장하여 관련된 정답을 생성할 때 사용할 수 있다.

발명의 내용

해결하려는 과제

[0009] 본 발명에 따른 매체의 신뢰성을 고려한 가짜 뉴스 탐색 모델은, 딥 러닝 모델인 CNN을 변형해 설계되었으며, 각 매체의 진실 여부와 명제를 입력 데이터셋(Input Dataset)으로 학습하여 다양한 매체의 신뢰성을 반영할 수 있다.

[0010] 또한, 이 모델은 데이터 증강(data augmentation), 배치 사이즈 변화, 모델 수정을 통해 높은 정확도를 보일 수 있도록 설계되었으므로, 높은 정확도를 가지는 가짜 뉴스 탐색 모델을 제공할 수 있다.

과제의 해결 수단

[0011] 본 발명의 일 실시예에 따른 언론 매체의 신뢰성을 반영한 기사 판별 시스템에 의해 수행되는 기사 판별을 위한 기계학습 모델을 학습시키는 방법으로서, 진위 판별을 위한 대상 명제를 수신하는 단계; 수신한 대상 명제를 기초로 웹 크롤링을 수행하여 관련 기사를 검색하고, 대상 명제와 검색된 관련 기사에 포함된 문장의 유사도를 산출하는 단계; 산출된 문장의 유사도에 따라 문장을 참, 거짓 또는 중립으로서 판별하여 판별값을 설정하는 단계; 및 유사도가 산출된 문장, 문장에 각각 설정된 판별값 및 문장을 포함하는 관련 기사를 제공한 언론 매체의 정보를 학습데이터로서 기계학습 모델에 입력하여 학습을 수행하는 단계를 포함하는, 기사 판별 기계학습 모델의 학습 방법을 제공하고자 한다.

[0012] 본 실시예에 있어서, 웹 크롤링은, 대상 명제와 웹페이지 상의 문장을 비교하여 코사인 유사도(Cosine Similarity)가 높은 문장을 검색하는 것을 포함하는, 기사 판별 기계학습 모델의 학습 방법을 제공할 수 있다.

[0013] 본 실시예에 있어서, 판별값을 설정하는 단계는, 검색 대상 명제와 관련 기사에 포함된 문장의 유사도가 50%이상이면 문장은 참이고, 유사도가 30%이상 50%미만이면 거짓이고, 유사도가 30% 미만이면 중립으로 설정되는, 기사 판별 기계학습 모델의 학습 방법을 제공할 수 있다.

[0014] 본 실시예에 있어서, 기계학습 모델은 CNN(Convolution Neural Network) 모델을 이용한 것인, 기사 판별 기계학습 모델의 학습 방법을 제공할 수 있다.

[0015] 본 발명의 일 실시예에 따른 언론 매체의 신뢰성을 반영한 기사 판별 시스템에 의해 수행되는 명제의 진위를 판별하는 방법에 있어서, 진위를 판별하고자 하는 대상 명제를 수신하는 단계; 수신한 대상 명제를 미리 웹 크롤링을 통해 검색된 명제와 관련된 관련 기사 내의 문장, 문장에 설정된 판별값 및 관련 기사를 제공한 언론 매체의 정보를 기초로 미리 학습된 기계학습 모델을 통하여 참, 거짓 또는 중립으로 판별하는 단계; 및 대상 명제의 판별 결과를 출력하는 단계를 포함하는, 언론 매체의 신뢰성을 반영한 명제의 진위 판별 방법을 제공하고자 한다.

- [0016] 본 실시예에 있어서, 대상 명제를 수신하는 단계에서, 대상 명제가 추출된 언론 매체의 정보를 더 수신하고, 판별 단계에서, 기계학습 모델은 언론 매체 정보에 더 기초하여 대상 명제의 진위를 판별하는, 언론 매체의 신뢰성을 반영한 명제의 진위 판별 방법을 제공할 수 있다.
- [0017] 본 실시예에 있어서, 웹 크롤링은 대상 명제와 웹페이지 상의 문장을 비교하여 코사인 유사도(Cosine Similarity)가 높은 문장을 검색하는 것을 포함하는, 언론 매체의 신뢰성을 반영한 명제의 진위 판별 방법을 제공할 수 있다.
- [0018] 본 실시예에 있어서, 설정된 판별값은, 검색 대상 명제와 관련 기사에 포함된 문장의 유사도가 50%이상이면 문장은 참이고, 유사도가 30%이상 50%미만이면 거짓이고, 유사도가 30% 미만이면 중립으로 설정된 것인, 언론 매체의 신뢰성을 반영한 명제의 진위 판별 방법을 제공할 수 있다.
- [0019] 본 실시예에 있어서, 기계학습 모델은 CNN(Convolution Neural Network) 모델을 이용한 것인, 언론 매체의 신뢰성을 반영한 명제의 진위 판별 방법을 제공할 수 있다.
- [0020] 본 발명의 일 실시예에 따른 언론 매체의 신뢰성을 반영한 기사 판별 시스템에 있어서, 언론 매체의 신뢰성을 반영하여 기사를 판별하는 프로그램이 저장된 메모리; 및
- [0021] 메모리에 저장된 프로그램을 실행하는 프로세서를 포함하고, 프로세서는 프로그램의 수행에 따라, 진위 판별을 위한 대상 명제를 수신하고, 수신한 대상 명제를 기초로 웹 크롤링을 수행하여 관련 기사를 검색하고, 대상 명제와 검색된 관련 기사에 포함된 문장의 유사도를 산출하고, 산출된 문장의 유사도에 따라 문장을 참, 거짓 또는 중립으로서 판별하여 판별값을 설정하고, 유사도가 산출된 문장과 문장에 각각 설정된 판별값을 학습데이터로서 기계학습 모델에 입력하여 학습을 수행하여 학습된 기계학습 모델을 통해 기사의 진위를 판별하는, 언론 매체의 신뢰성을 반영한 기사 판별 시스템을 제공하고자 한다.
- [0022] 본 실시예에 있어서, 기계학습 모델을 학습시킨 다음에, 진위를 판별하고자 하는 대상 명제를 수신하고, 수신한 대상 명제를 미리 학습된 기계학습 모델을 통하여 참, 거짓 또는 중립으로 판별하고, 대상 명제의 판별 결과를 출력하는, 언론 매체의 신뢰성을 반영한 기사 판별 시스템을 제공할 수 있다.
- [0023] 본 실시예에 있어서, 대상 명제의 진위 판별을 위해와 대상 명제가 추출된 언론 매체의 정보를 더 수신하고, 기계학습 모델은 수신한 언론 매체 정보에 더 기초하여 대상 명제의 진위를 판별하는, 언론 매체의 신뢰성을 반영한 기사 판별 시스템을 제공할 수 있다.
- [0024] 본 실시예에 있어서, 웹 크롤링은 대상 명제와 웹페이지 상의 문장을 비교하여 코사인 유사도(Cosine Similarity)가 높은 문장을 검색하는 것을 포함하는, 언론 매체의 신뢰성을 반영한 기사 판별 시스템을 제공할 수 있다.
- [0025] 본 실시예에 있어서, 설정된 판별값은, 검색 대상 명제와 관련 기사에 포함된 문장의 유사도가 50%이상이면 문장은 참이고, 유사도가 30%이상 50%미만이면 거짓이고, 유사도가 30% 미만이면 중립으로 설정된 것인, 언론 매체의 신뢰성을 반영한 기사 판별 시스템을 제공할 수 있다.
- [0026] 본 실시예에 있어서, 기계학습 모델은 CNN(Convolution Neural Network) 모델을 이용한 것인, 언론 매체의 신뢰성을 반영한 기사 판별 시스템을 제공할 수 있다.

발명의 효과

- [0027] 본 발명의 일 실시예에 따른 언론 매체의 신뢰성을 고려한 기사 판별 시스템은 명제의 참, 거짓 또는 중립 여부와 각 언론 매체의 정답 여부를 포함하는 입력 데이터셋을 학습하여 다양한 매체의 신뢰도를 반영하여 통계적인 방법보다 더 높은 정확도를 가지는 효과가 있다.
- [0028] 본 발명의 일 실시예에 따른 언론 매체의 신뢰성을 고려한 기사 판별 시스템은 딥러닝 모델을 기초로 하여 가짜 뉴스를 탐색할 뿐만 아니라 다양한 매체의 신뢰성을 평가하는 데 효율적으로 사용될 수 있다.
- [0029] 본 발명의 일 실시예에 따른 언론 매체의 신뢰성을 고려한 기사 판별 시스템을 더욱 많은 양의 데이터셋에 적용하여 정확도를 높일 수 있으며, 분산 병렬 환경에서 구현하면 실시간으로 다양한 매체의 가짜 뉴스 탐색을 빠르게 처리할 수 있다.

도면의 간단한 설명

- [0030] 도 1은 본 발명의 일 실시예에 따른 기사 판별 시스템의 구성을 보여주는 구성도이다.
- 도 2는 본 발명의 일 실시예에 따른 명제의 진위 판별 방법을 보여주는 순서도이다.
- 도 3은 본 발명의 일 실시예에 따른 기사 판별 기계학습 모델의 학습 방법을 보여주는 순서도이다.
- 도 4는 본 발명의 일 실시예에 따른 기사 판별 시스템의 구성을 보여주는 개략도이다.
- 도 5는 본 발명의 일 실시예에 따른 신뢰성 확인 모델의 구성을 보여주는 개략도이다.
- 도 6은 본 발명의 일 실험에 따른 테스트용 입력 데이터셋을 보여주는 도면이다.
- 도 7은 본 발명의 일 실험예에 따른 가짜 뉴스 탐색 시스템의 데이터 증강(Data Augmentation)을 적용하지 않은 정확도를 보여주는 도면이다.
- 도 8은 본 발명의 일 실시예에 따른 기사 판별 시스템이 학습하는 데이터셋의 배치 사이즈(Batch Size) 별로 정확도를 보여주는 도면이다.
- 도 9는 본 발명의 일 실험예에 따른 기사 판별 시스템의 컨볼루션 레이어(Convolutional layer)의 크기를 달리 하여 정확도를 측정한 결과를 보여주는 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0031] 아래에서는 첨부한 도면을 참조하여, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 본 발명의 실시예를 상세히 설명한다. 그러나, 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고, 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.
- [0032] 명세서 전체에서, 어떤 부분이 다른 부분과 "연결"되어 있다고 할 때, 이는 "직접적으로 연결"되어 있는 경우, 그 중간에 다른 소자를 사이에 두고 "전기적으로 연결"되어 있는 경우도 포함한다. 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.
- [0033] 먼저 본 발명의 일 실시예에 따른 언론 매체의 신뢰성을 반영한 기사 판별 시스템(100)에 대하여 설명하기로 한다.
- [0034] 도 1은 본 발명의 일 실시예에 따른 기사 판별 시스템(100)의 구성을 보여주는 구성도이다.
- [0035] 도 1을 참조하면, 본 발명의 일 실시예에 따른 기사 판별 시스템(100)은 통신 모듈(110), 메모리(120), 프로세서(130), 데이터베이스(DB: 140)를 포함할 수 있다.
- [0036] 통신 모듈(110)은 통신망과 연동하여 기사 판별 시스템(100)에 통신 인터페이스를 제공하는데, 사용자 단말(미도시)로부터 진위 판별이 필요한 기사 또는 명제에 대한 데이터를 송수신하는 역할을 수행할 수 있다. 여기서, 통신 모듈(110)은 다른 네트워크 장치와 유무선 연결을 통해 제어 신호 또는 데이터 신호와 같은 신호를 송수신하기 위해 필요한 하드웨어 및 소프트웨어를 포함하는 장치일 수 있다.
- [0037] 메모리(120)는 언론 매체의 신뢰성을 반영한 기사 판별 프로그램이 기록된 저장 매체일 수 있다. 또한, 메모리(120)는 프로세서(130)가 처리하는 데이터를 일시적 또는 영구적으로 저장하는 기능을 수행할 수 있다. 여기서, 메모리(120)는 휘발성 저장 매체(volatile storage media) 또는 비휘발성 저장 매체(non-volatile storage media)를 포함할 수 있으나, 본 발명의 범위가 이에 한정되는 것은 아니다.
- [0038] 프로세서(130)는 기사 판별 시스템(100)에서 언론 매체의 신뢰성을 반영한 기사 판별을 위한 프로그램이 수행하는 전체 과정을 제어할 수 있다. 프로세서(130)가 수행하는 과정의 각 단계에 대해서는 도 2 내지 도 8을 참조하여 후술하기로 한다.
- [0039] 여기서, 프로세서(130)는 프로세서(processor)와 같이 데이터를 처리할 수 있는 모든 종류의 장치를 포함할 수 있다. 여기서, '프로세서(processor)'는, 예를 들어 프로그램 내에 포함된 코드 또는 명령으로 표현된 기능을 수행하기 위해 물리적으로 구조화된 회로를 갖는, 하드웨어에 내장된 데이터 처리 장치를 의미할 수 있다. 이와 같이 하드웨어에 내장된 데이터 처리 장치의 일 예로서, 마이크로프로세서(microprocessor), 중앙처리장치(central processing unit: CPU), 프로세서 코어(processor core), 멀티프로세서(multiprocessor),

ASIC(application-specific integrated circuit), FPGA(field programmable gate array) 등의 처리 장치를 망라할 수 있으나, 본 발명의 범위가 이에 한정되는 것은 아니다.

- [0040] 데이터베이스(140)는 언론 매체의 신뢰성을 반영한 기사 판별을 위한 기초 데이터가 되는 언론 매체의 신뢰성 데이터, 언론 매체에서 제공하는 기사에 포함된 명제 데이터, 명제의 진위 판별을 기계학습하기 위한 학습데이터 등이 저장되는 것일 수 있다.
- [0041] 아래에서는 본 발명의 일 실시예에 따른 기사 판별 시스템(100)에 의하여 언론 매체의 신뢰성을 반영한 기사 판별 방법에 대하여 설명하기로 한다.
- [0042] 도 2는 본 발명의 일 실시예에 따른 명제의 진위 판별 방법을 보여주는 순서도이다. 도 3은 본 발명의 일 실시예에 따른 기사 판별 기계학습 모델의 학습 방법을 보여주는 순서도이다.
- [0043] 도 2 및 도 3을 참조하면, 언론 매체의 신뢰성을 반영한 기사 판별 방법은 기사에 포함된 명제의 진위를 판별하여 수행될 수 있다.
- [0044] 먼저, 기사 판별 시스템(100)은 진위 판별 대상인 기사에 포함된 진위를 판별하고자 하는 대상 명제를 수신하는 단계(S210)를 수행할 수 있다.
- [0045] 여기에서, 기사 판별 시스템(100)은 대상 명제와 함께 대상 명제가 추출된 언론 매체의 정보를 더 수신할 수 있다. 예를 들어, 기사 판별 시스템(100)은 명제 'a=b'가 추출된 기사를 제공한 언론 매체 'A'의 명칭, 웹사이트 주소, 신뢰성 등의 정보를 더 수신할 수 있다.
- [0046] 단계(S210) 다음으로, 기사 판별 시스템(100)은 수신한 대상 명제를 미리 웹 크롤링을 통해 검색된 명제와 관련된 관련 기사 내의 문장, 문장에 설정된 판별값 및 관련 기사를 제공한 언론 매체의 정보를 기초로 미리 학습된 기계학습 모델을 통하여 참, 거짓 또는 중립으로 판별하는 단계(S220)를 수행할 수 있다.
- [0047] 여기에서 웹 크롤링은 대상 명제와 웹페이지 상의 문장을 비교하여 코사인 유사도(Cosine Similarity)가 높은 문장을 검색하는 것일 수 있다.
- [0048] 또한, 상술한 기계학습 모델은 상기 언론 매체 정보에 더 기초하여 대상 명제의 진위를 판별할 수 있다. 예를 들어, 미리 평가된 언론 매체의 신뢰성을 반영하여 대상 명제의 진위를 판별할 수 있다. 언론 매체의 신뢰성이 낮은 경우 대상 명제가 거짓으로 판별될 확률이 높아질 수 있다.
- [0049] 한편, 기계학습 모델을 학습시키는 방법으로서, 먼저 진위 판별 대상 명제를 수신하는 단계(S310)가 수행될 수 있다.
- [0050] 단계 다음으로, 기사 판별 시스템(100)이 수신한 대상 명제를 기초로 웹 크롤링을 수행하여 관련 기사를 검색하고, 대상 명제와 검색된 관련 기사에 포함된 문장의 유사도를 산출하는 단계(S320)가 수행될 수 있다.
- [0051] 단계(S320) 다음으로, 산출된 문장의 유사도에 따라 문장을 참, 거짓 또는 중립으로서 판별하여 판별값을 설정하는 단계(S330)가 수행될 수 있다.
- [0052] 단계(S330) 다음으로, 유사도가 산출된 문장, 문장에 각각 설정된 판별값 및 문장을 포함하는 관련 기사를 제공한 언론 매체의 정보를 학습데이터로서 기계학습 모델에 입력하여 학습을 수행하는 단계(S340)가 수행될 수 있다.
- [0053] 또한, 기계학습 모델은 CNN(Convolution Neural Network) 모델을 이용한 것일 수 있다. 한편, 딥러닝 모델 중에서 CNN 모델은 일반적으로 데이터를 분류할 때 사용된다. 특히, 이미지 데이터셋에서 특징을 추출하여 적합한 클래스로 분류하는 데 적용된다. 이미지넷의 경우, 이미지 데이터를 다양한 레벨의 컨볼루션 층(Convolutional Layer)으로 이미지를 분류할 수 있다. 이미지넷과 마찬가지로, 다양한 언론 매체의 정답 데이터셋에 CNN 모델이 적용되어 참, 거짓, 또는 중립의 세 클래스로 분류될 수 있다. 본 발명의 일 실시예에 따른 기사 판별 시스템(100)에 적용된 기계학습 모델은 이미지넷을 기초로 하였으나 층의 수와 크기는 데이터셋의 정확도를 높이는 데 적절하도록 변형되었다.
- [0054] 예를 들어, 기계학습 모델이 학습되는 데이터의 배치 사이즈의 크기를 변화시킬 수 있다. 기계학습에서 학습을 위한 데이터셋의 크기와 모양에 따라 적절한 배치 사이즈를 정하는 것이 중요하기 때문이다. 또한, 파인튜닝(fine-tune) 딥러닝을 사용하여 오버피팅을 막을 수 있다. CNN 모델은 일반적으로 초기 가중치를 랜덤으로 설정하고 학습하며 가중치를 업데이트한다. 파인튜닝은 학습된 가중치를 사용해 학습하는 방법으로서 CNN 모델의 정확

도를 향상시킬 수 있다. 원샷러닝(One-shot learning)은 클래스의 유사성을 계산하여 다양한 클래스로 적은 양의 데이터셋을 분류하는 방법으로서, 클래스의 수가 많을 때 적절하게 적용될 수 있다. 다음으로, 데이터 증강(Data augmentation)은 이미지의 각도를 변형시키고, 크기를 바꾸고, 수직 또는 수평으로 바꾸며 학습 데이터셋의 수를 늘리는 방법으로서, 이미지 데이터셋뿐만 아니라 기사 판별 시스템(100)의 페이크 뉴스 탐색 모델에서 사용되는 매체 데이터셋의 특징을 변형시키는 데 사용될 수 있다.

- [0055] 한편, 상술한 웹 크롤링은 대상 명제와 웹페이지 상의 문장을 비교하여 코사인 유사도(Cosine Similarity)가 높은 문장을 검색하는 것을 포함하여 수행될 수 있다.
- [0056] 또한, 상술한 문장에 설정된 판별값은 검색 대상 명제와 상기 관련 기사에 포함된 문장의 유사도가 50%이상이면 상기 문장은 참이고, 상기 유사도가 30%이상 50%미만이면 거짓이고, 상기 유사도가 30% 미만이면 중립으로 설정된 것일 수 있다.
- [0057] 단계(S220) 다음으로, 기사 판별 시스템(100)은 대상 명제의 판별 결과를 출력하는 단계(S230)를 수행할 수 있다.
- [0058] 아래에서는 본 발명의 일 실시예에 따른 언론 매체의 신뢰성을 반영한 기사 판별 시스템(100) 및 명제의 진위 판별 방법에 대해서 더 구체적으로 설명하기로 한다.
- [0059] 도 4는 본 발명의 일 실시예에 따른 기사 판별 시스템(100)을 보여주는 개략도이다.
- [0060] 본 발명의 일 실시예에 따른 기사 판별 시스템(100)은 두 부분으로 나누어져 있다. 이들은 각각 언론 매체의 페이크 기사를 탐색하는 부분(Fake news detection model)과 딥러닝 모델을 사용해 언론 매체의 신뢰성을 확인하는 부분(Reliability checking model)일 수 있다. 언론 매체의 페이크 기사를 탐색하는 부분은 각 언론 매체가 해당 명제에 대해 참으로 주장하는지, 거짓으로 주장하는지, 또는 중립으로 주장하는지 탐색할 수 있다. 언론 매체의 신뢰성을 확인하는 부분은 CNN 모델을 변형한 딥러닝 모델을 이용하여 기사들의 명제에 대하여 수집된 데이터를 학습하며 각 언론 매체의 신뢰성을 판별하고, 기사 판별 시스템(100)은 이를 기초하여 기사가 참인지, 거짓인지 또는 중립인지 결정할 수 있다.
- [0061] 즉, 일반적인 페이크 뉴스 탐색 모델과 달리, 본 발명의 일 실시예에 따른 기사 판별 시스템(100)의 페이크 뉴스 탐색 부분에서는 다양한 언론 매체의 명제에 대한 정답 데이터를 추출하고, 신뢰성 확인 부분에서는 변형된 CNN 모델에 추출된 정답 데이터를 입력하여 정확도를 높일 수 있다. 여기에서, 기계학습 모델의 입력 데이터셋의 가중치를 지속적으로 업데이트하여 언론 매체의 신뢰성을 반영하여 더욱 정확한 가짜 뉴스 탐색 결과를 가질 수 있다.
- [0062] 본 발명의 일 실시예에 따른 기사 판별 시스템(100)은 CNN 모델을 기초로 하였으며 다양한 언론 매체로부터 얻은 데이터셋을 학습하며 작동한다. 또한, 기사 판별 시스템(100)은 데이터 증강(data augmentation), 배치 사이즈 변화, 모델 수정을 통해 높은 정확도를 보일 수 있도록 설계될 수 있다. 그 결과로서, 언론 매체의 명제에 대한 정답의 빈도수를 측정한 통계적인 방법을 사용하여 기사를 탐색한 모델과 수집한 데이터셋을 학습하여 수행한 기사 판별 방법이 정확도 높은 결과를 보였다.
- [0063] 도 5는 본 발명의 일 실시예에 따른 신뢰성 확인 모델의 구성을 보여주는 구성도이다.
- [0064] 도 5를 참조하면, 신뢰성 확인 모델은 CNN 모델을 변형하여 레이어(Layer)들로 구성되어 있다: 입력 레이어(Input layer), 컨볼루션 레이어(Convolutional layer), 맥스풀링 레이어(Max-pooling layer), 두 개의 드롭아웃 레이어(Dropout layer), 풀리 커넥티드 레이어(Fully-connected layer), 소프트맥스 레이어(Softmax layer) 그리고 출력 레이어(Output layer)이다. 입력 레이어는 각각 4 x 4 x 1 크기이며, 출력 레이어는 클래스들인 참, 거짓, 중립 중 하나를 생성한다. 컨볼루션 레이어에서는 12 정규화가 사용되어 가중치를 일정하게 분배하고 노이즈를 줄이고 적절한 크기의 필터를 사용해 정확한 결과를 얻을 수 있도록 하였다. 신뢰성 확인 모델은 적은 양의 데이터를 더욱 효과적으로 사용하기 위해 복잡하지 않은 단순한 구조로 설계되었다. 또한, 신뢰성 확인 모델은 드롭아웃 값(Dropout value)을 비교적 큰 값으로 설정하여 불필요한 노드를 제거하고 레이어들의 변수를 줄였다. 또한, 신뢰성 확인 모델은 추가적인 드롭아웃 레이어를 사용하여 적은 양의 데이터에서 높은 정확도를 보일 수 있도록 하였다. 또한, 신뢰성 확인 모델은 소프트맥스 레이어에서, Softmax-with-Loss 방법을 사용하여 효율적인 분류를 위해 가중치를 적절하게 업데이트할 수 있도록 하였다.
- [0065] 본 발명의 일 실시예에 따른 기사 판별 시스템(100)은 1 CPU(Intel® Core® CPU @ 3.50GHz)과 8GB 메모리를 가지고 파이참(Pycharm) 환경에서 세 종류의 데이터셋을 활용하여 실험을 진행하였다. 또한, 딥러닝 모델을 구

측하기 위해 케라스 라이브러리(Keras library)를 사용하였다.

[0066] 본 발명의 일 실시예에 따른 신뢰성 평가 모델의 성능을 평가하기 위해 세 가지 데이터셋을 사용하였다. 첫번째 데이터셋은 MNIST 데이터셋으로 28 x 28 크기의 숫자 데이터셋이다. MNIST 원본 데이터셋은 0에서 9까지 10개의 클래스로 나누어져 빨강, 초록 그리고 파랑의 3개 채널로 구성되어 있다. 신뢰성 평가 모델은 참, 거짓 그리고 중립의 세 클래스로 구성되어 있으므로, 고정된 크기의 0, 1, 2 이미지 클래스의 수를 줄여 사용하였다.

[0067] 도 6은 본 발명의 일 실험에 따른 테스트용 입력 데이터셋을 보여주는 도면이다.

[0068] 도 6을 참조하면, 실제 데이터셋을 사용하기 전에 테스트를 하기 위해 정제된 데이터를 임의로 만들어 사용했다. 각 데이터는 16개의 정답 데이터로 4 x 4 크기의 matrix로 구성되어 있으며, 거짓, 참, 그리고 중립의 여부는 각각 0, 1, 2로 표현될 수 있다. 신뢰성 평가 모델의 효율적인 학습을 위해 다양한 매체 중에서 한 매체는 100%의 정확도로, 또 다른 한 매체는 0%의 정확도로 그리고 나머지 매체는 랜덤으로 데이터를 구성하였다.

[0069] 마지막 데이터셋은 두 번째 데이터셋과 비슷하지만 16개의 매체에서 얻은 실제 뉴스의 데이터이다. 아래의 [표 1]은 첫번째 데이터셋을 보여주는 표이고, [표 2]는 두번째 데이터셋을 보여주는 표이고, [표 3]은 세번째 데이터셋을 보여주는 표이다.

[0070] 가짜 뉴스 탐색 시스템의 성능을 평가하기 위해 아래의 3개의 표에 따른 3 개의 데이터셋을 사용하였다. 첫 번째 데이터셋은 MNIST 데이터셋으로 28 x 28 크기의 숫자 데이터셋이다. 두 번째 데이터셋은 실제 데이터셋을 사용하기 전에 테스트를 하기 위해 정제된 데이터를 임의로 만들어 사용했다. 마지막 데이터셋은 두 번째 데이터셋과 비슷하지만 16개의 매체에서 얻은 실제 뉴스의 데이터이다.

표 1

Attribute type	Size
Data size	28 x 28
Number of channels	1
Number of classes	3
Number of training dataset	7,728
Number of validation dataset	1,472
Number of test dataset	6,389

[0071]

표 2

Attribute type	Size
Data size	4 x 4
Number of channels	1
Number of classes	3
Number of training dataset	4,149
Number of validation dataset	790
Number of test dataset	244

[0072]

표 3

Attribute type	Size
Data size	4 x 4
Number of channels	1
Number of classes	3
Number of training dataset	84
Number of validation dataset	18
Number of test dataset	21

[0073]

[0074] 본 발명에 따른 가짜 뉴스 탐색 시스템은 위의 [표 1], [표 2], [표 3]에 따른 세 개의 데이터셋으로 실험되었다. 또한 정확도를 높이기 위해 데이터 증강(Data augmentation)을 사용하고, 배치 사이즈를 변경하고 모델 레이어를 변형해 모델을 보완하였다.

[0075] 도 7은 본 발명의 일 실시예에 따른 가짜 뉴스 탐색 시스템의 데이터 증강(Data Augmentation)을 적용하지 않은 정확도를 보여주는 도면이다.

[0076] 도 7을 참조하면 먼저, 매체 데이터셋에 대하여 데이터 증강이 학습에 도움이 되는지 확인하는 실험을 한 것을 볼 수 있다. 구체적으로, 데이터 증강이 적용된 그래프(with)와 데이터 증강이 적용되지 않은 그래프(without)를 각각 볼 수 있다. 매체 데이터의 양은 MNIST 데이터셋과 정제된 데이터셋의 양보다 훨씬 적었다. MNIST 데이터셋과 정제된 데이터는 높은 정확도를 보이는 반면 매체 데이터셋은 낮은 정확도를 보이고 오버피팅을 보였다. 이러한 문제를 해결하기 위해 데이터 증강을 적용하여 더 높은 정확도를 확보했다.

[0077] 여기서, 데이터 증강은 다양하게 이미지 데이터셋을 변형하는 사용되는 기술로 이미지 처리에서 다양한 특징을 추출할 때 사용된다. 적은 양의 매체 데이터셋으로 정확도를 높이기 위해 매체 데이터셋에 적용하여 다양한 특징을 더욱 효율적으로 추출하도록 하였다. 데이터 증강이 적용된 매체 데이터는 각각 왼쪽과 오른쪽 다양한 각도로 변형하여 실험하였고 왼쪽과 오른쪽 방향으로 5도씩 움직여 매체 데이터셋의 적은 양을 보완하였다. 데이터 증강이 적용되었을 때와 적용하지 않았을 때의 정확도 그래프는 도 7의 'with' 그래프와 같이 오버피팅이 발생하지 않는다는 것을 보여준다.

- [0078] 도 8은 본 발명의 일 실시예에 따른 기사 판별 시스템(100)이 학습하는 데이터셋의 배치 사이즈(Batch Size) 별로 정확도를 보여주는 도면이다.
- [0079] 도 8을 참조하면, 기사 판별 시스템(100)이 학습할 때의 배치 사이즈를 4, 16, 32, 64로 변화시켜 최적의 배치 사이즈를 찾기 위해 실험한 것을 볼 수 있다. 이 실험에서 MNIST, 정제된 데이터셋 그리고 매체 데이터셋에서 학습 배치 사이즈가 64일 때 가장 높은 정확도를 보였다.
- [0080] 도9는 본 발명의 일 실시예에 따른 기사 판별 시스템(100)의 컨볼루션 레이어(Convolutional layer)의 크기를 달리하여 정확도를 측정한 결과를 보여주는 도면이다.
- [0081] 도 9를 참조하면 세번째로, 본 발명에 따른 가짜 뉴스 탐색 시스템의 컨볼루션 레이어의 크기를 1,2,3으로 나누어 실험한 것을 볼 수 있다. 레이어의 개수가 더 많을 때 상대적으로 더 많은 양으로 구성된 MNIST 데이터셋과 정제된 데이터셋은 매체 데이터셋보다 더 높은 정확도를 보인 반면에, 레이어의 개수가 더 많을 때 매체 데이터는 더욱 낮은 정확도를 보였다. 도 8에서와 같이, 가짜 뉴스 탐색 시스템은 매체 데이터셋이 더 많은 개수의 레이어로 구성됐을 때보다 1개의 컨볼루션 레이어로 구성된 간단한 모델일 때 더욱 높은 정확도를 보였다. 그러므로, 기사 판별 시스템(100)은 하나의 컨볼루션 레이어로 구성되는 것이 효과적이다. 또한, 이 실험에서 드롭아웃 레이어(Dropout layer)가 추가되었을 때 더욱 높은 정확도를 보였다.

표 4

Model	MNIST dataset	Refined dataset	Media dataset
기사 판별 시스템(100)	100%	97.7%	75.6%
통계 시스템	88%	86%	72%

- [0082]
- [0083] [표 4]는 MNIST 데이터셋, 정제된 데이터셋, 매체 데이터셋에 대한 기사 판별 시스템(100)과 통계 시스템의 정확도를 비교한 표이다. 여기에서, 기사 판별 시스템(100)을 각 매체의 정답 수를 측정하여 통계적인 방법으로 비교했다. 통계적인 방법은 정답을 잘 찾아내지 못했으며 가짜 뉴스 탐색 시스템이 통계적인 방법보다 3% 이상 더 잘 찾아냈다는 것을 확인할 수 있다.
- [0084] 정리하자면, 본 발명의 일 실시예에 따른 기사 판별 시스템(100)은 데이터 증강, 배치 사이즈 변화, 모델 수정을 통해 높은 정확도를 보일 수 있도록 설계했다. 데이터 증강은 낮은 정확도와 오버피팅의 문제를 해결하기 위해 매체 데이터셋에 사용되었다. 그리고 MNIST, 정제된 데이터셋, 그리고 매체 데이터셋에서 가장 높은 정확도를 보이는 최적의 배치사이즈를 찾기 위해 실험하였다. 또한, 가장 높은 정확도를 보이는 최적의 모델을 찾기 위해 컨볼루션 레이어의 개수를 바꾸어 다양한 개수에서 실험하며 기사 판별 시스템(100)을 평가하였다.
- [0085] 이로써, 본 발명의 일 실시예에 따른 기사 판별 시스템(100)은 명체의 참, 거짓 또는 중립 여부와 각 언론 매체의 정답 여부를 포함하는 입력 데이터셋을 학습하여 다양한 매체의 신뢰도를 반영하여 통계적인 방법보다 더 높은 정확도를 가지는 효과가 있다.
- [0086] 이상으로 설명된 본 발명에 따른 매체의 신뢰성을 고려한 기사 판별 시스템(100)은 딥러닝 모델을 기초로 하여 가짜 뉴스를 탐색할 뿐만 아니라 다양한 매체의 신뢰성을 평가하는 데 효율적으로 사용될 수 있다. 기사 판별 시스템(100)을 더욱 많은 양의 데이터셋에 적용하여 정확도를 높일 수 있으며, 분산 병렬 환경에서 구현하면 실시간으로 다양한 매체의 가짜 뉴스 탐색을 빠르게 처리할 수 있다.
- [0087] 한편, 본 발명의 일 실시예에 따른 기사 판별 기계학습 모델의 학습 방법 및 다른 실시예에 따른 언론 매체의 신뢰성을 반영한 명체의 진위 판별 방법은 컴퓨터에 의해 실행되는 프로그램 모듈과 같은 컴퓨터에 의해 실행 가능한 명령어를 포함하는 기록 매체의 형태로도 구현될 수 있다. 컴퓨터 판독 가능 매체는 컴퓨터에 의해 액세스될 수 있는 임의의 가용 매체일 수 있고, 휘발성 및 비휘발성 매체, 분리형 및 비분리형 매체를 모두 포함한다. 또한, 컴퓨터 판독가능 매체는 컴퓨터 저장 매체를 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 포함한다. 본 발명의 방법 및 시스템은 특정 실시예와 관련하여 설명되었지만, 그것들의 구성 요소 또는 동작의 일부 또는 전부는 범용 하드웨어 아키텍처를 갖는

컴퓨터 시스템을 사용하여 구현될 수 있다.

[0088] 이상의 설명은 본 발명의 기술적 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.

[0089] 따라서, 본 발명에 개시된 실시예는 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 기술적 사상은 본 발명의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

부호의 설명

[0090] 100: 언론 매체의 신뢰성을 반영한 기사 판별 시스템

110: 통신 모듈

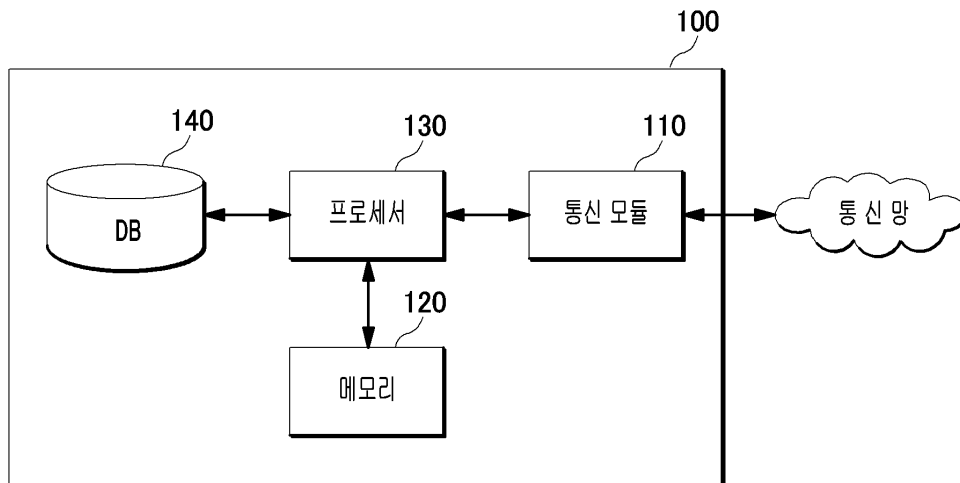
120: 메모리

130: 프로세서

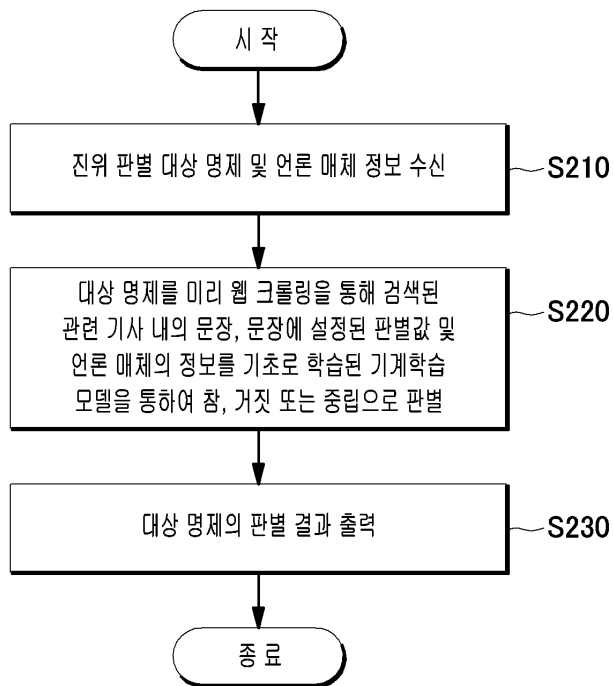
140: 데이터베이스

도면

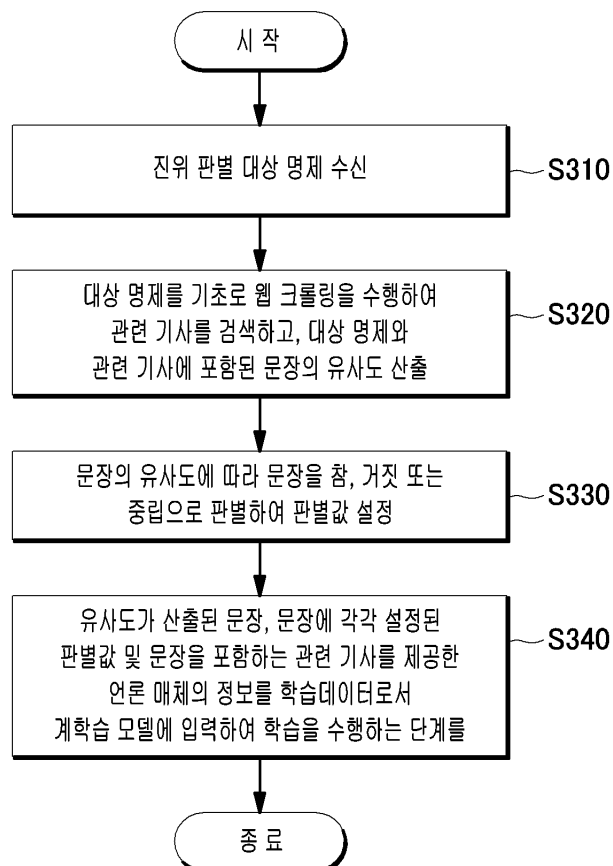
도면1



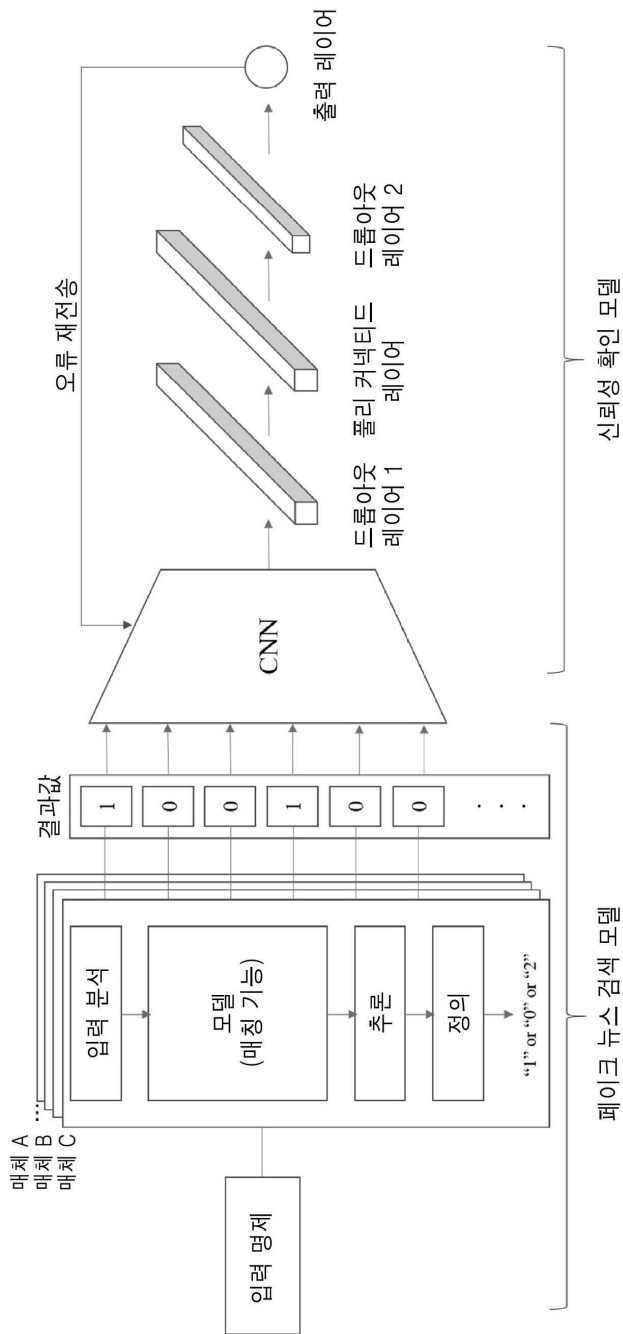
도면2



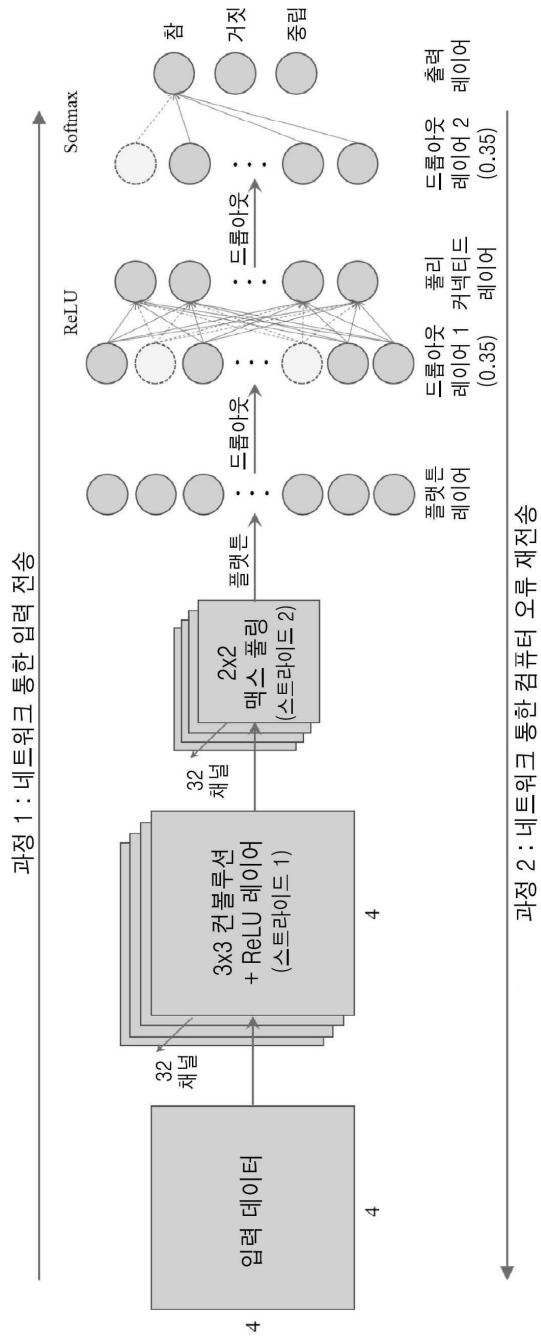
도면3



도면4



도면5



도면6

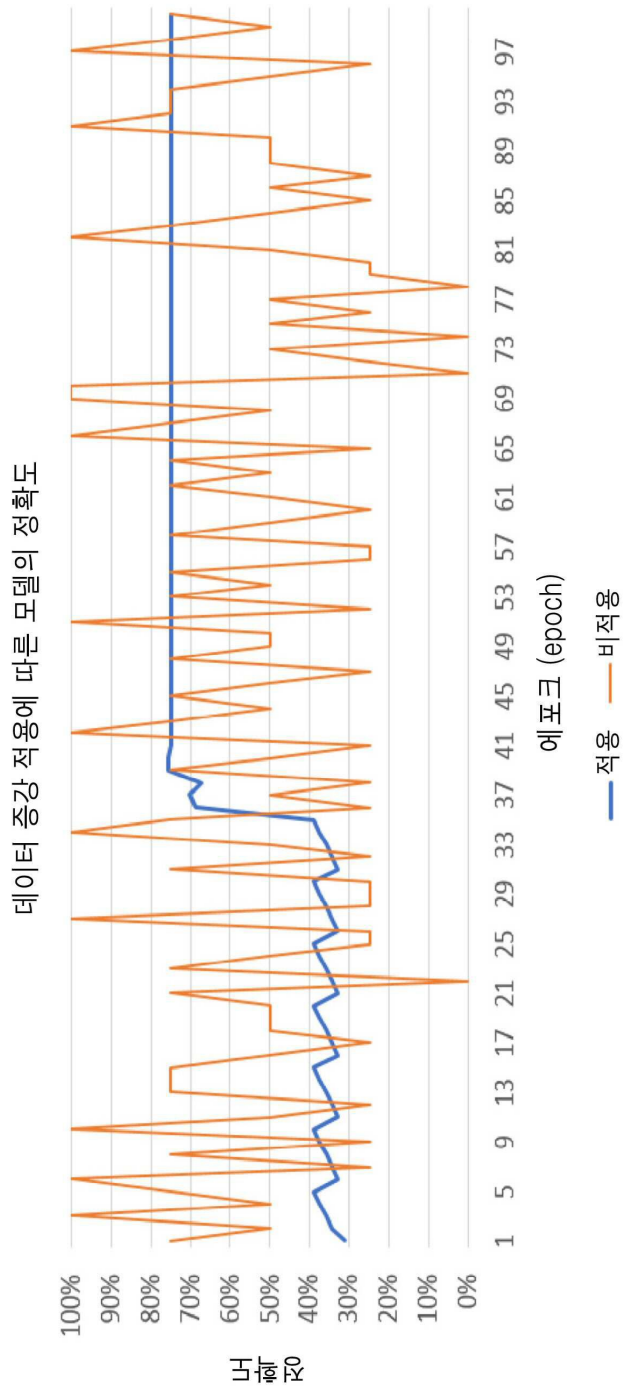
매체 A	매체 B	매체 C	매체 D
매체 E	매체 F	매체 G	매체 H
매체 I	매체 J	매체 K	매체 L
매체 M	매체 N	매체 O	매체 P

참

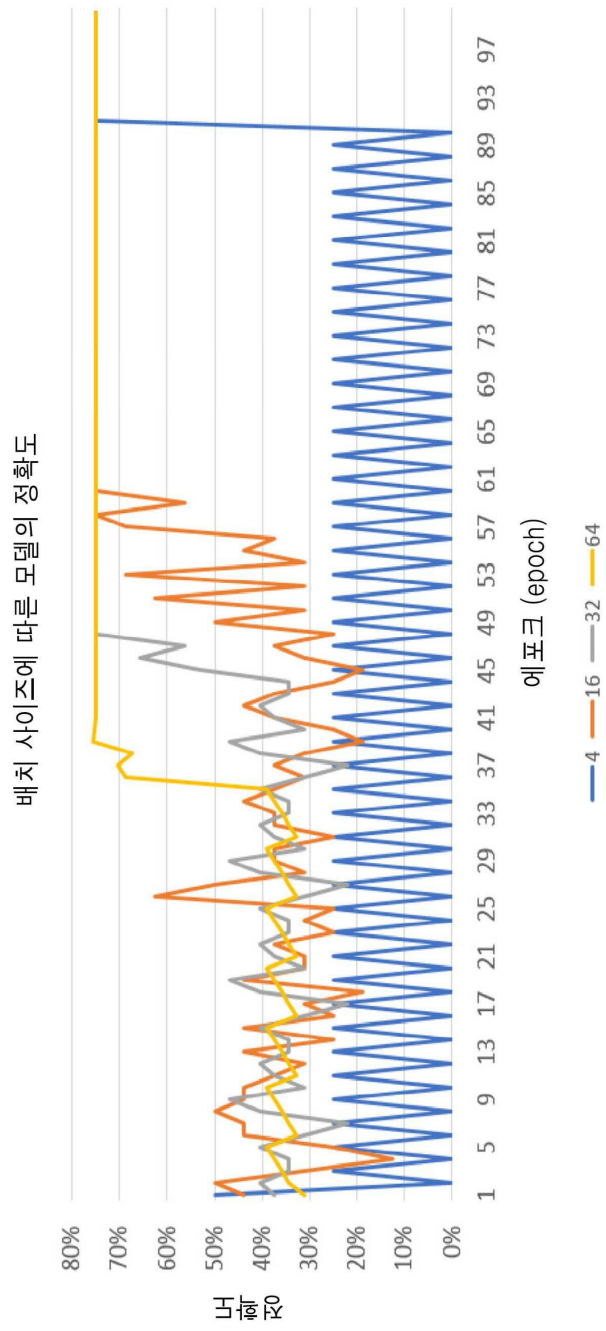
거짓

중립

도면7



도면8



도면9

레이어 수에 따른 모델의 정확도

